www.koreaai.org

제2회 한국 인공지능 학술대회

The 2nd Korea Artificial Intelligence Conference

Proceedings

일자 2021년 9월 29일(수) ~ 10월 1일(금)

장소 신라스테이 제주 [온/오프라인 하이브리드 개최]

주최 한국통신학회

주관 한국통신학회 인공지능소사이어티

후원 국민대학교 에너지인터넷연구센터, 국민대 텔레매틱스연구소,
경북대학교 ICT·자동차융합연구센터

KICS
한 국 통 신 학 회

준비위원

▣ 자문위원

강충구(고려대) 김동인(성균관대) 김영한(숭실대)
김형준(한국전자통신연구원) 노종선(서울대) 박세웅(서울대)
박종현(한국전자통신연구원) 박현제(소프트웨어정책연구소) 방승찬(한국전자통신연구원)
이규복(한국전자기술연구원) 이윤근(한국전자통신연구원) 정일영(한국외국어대)
조용수(중앙대) 조유제(경북대)

▣ 조정위원장

장영민(국민대)

▣ 조정위원

강신각(한국전자통신연구원) 김대진(전남대) 문영준(한국교통연구원)
송홍엽(연세대) 신요안(숭실대) 여 현(순천대)
이수인(한국전자통신연구원) 이인규(고려대) 정 송(한국과학기술원)
정성호(한국외국어대) 허 준(고려대) 홍인기(경희대)

▣ 운영위원장

한동석 (한국통신학회 인공지능소사이어티 회장)

▣ 운영위원

- 총무

김동균(경북대) 김동성(금오공과대) 김정구(부산대)
김홍국(광주과학기술원)

- ManuscriptLink

한연희(한국기술교육대)

- 등록

조성래(중앙대) 최준원(한양대) 박경준(대구경북과학기술원)

- 섭외

김중헌(고려대) 이현우(한국전자통신연구원) 최계원(성균관대)

- 재무

김재일(경북대) 최윤호(부산대)

- 출판

김상철(국민대) 김덕경(인하대) 김정곤(한국산업기술대)
석준희(고려대)

준비위원

• 지역협조

김도현(제주대) 김선옥(한라대) 송왕철(제주대)
좌정우(제주대)

• 홍보

강승택(인천대) 김남수(서울대) 김동호(서울과학기술대)
남해운(한양대) 백명선(ETRI) 이동명(동명대)
이예훈(서울과학기술대) 정연호(부경대) 황승훈(동국대)
황인태(전남대) 심동규(광운대) 한종기(세종대)

• 특별세션

김종원(광주과학기술원) 박상준(한국전자통신연구원) 최지웅(대구경북과학기술원)

▣ 프로그램위원장

손인수(동국대) 한연희(한국기술교육대) 신석주(조선대)

▣ 프로그램위원

고석주(경북대) 고정길(연세대) 권태수(서울과학기술대)
김광주(한국전자통신연구원) 김근영(한국전자통신연구원) 김대중(한국정보통신기술협회)
김수민(한국산업기술대) 김용환(한국과학기술정보연구원) 김원태(한국기술교육대)
김준수(한국산업기술대) 김찬명((주)씽크온웹) 김태운(한림대)
김평수(한국산업기술대) 김호원(부산대) 박기웅(세종대)
박대진(경북대) 박정훈(경북대) 박준구(경북대)
박현희(명지대) 박형곤(이화여대) 박혜영(경북대)
소재우(서강대) 송윤정(한국전자통신연구원) 신수용(금오공과대)
신오순(숭실대) 심병효(서울대) 유준혁(대구대)
윤주상(동의대) 이규만(건양대) 이상웅(가천대)
이승형(광운대) 이원철(숭실대) 이재호(덕성여자대)
이종혁(세종대) 이주현(한양대) 임길택(한국전자통신연구원)
임민중(동국대) 임완수(금오공대) 임유진(숙명여대)
장혜령(동국대) 전상운(한양대) 전찬준(조선대)
정방철(충남대) 정준영(한국전자통신연구원) 조봉균(KIAPI)
조오현(충북대) 최동준(한국전자통신연구원) 최우열(조선대)
최 창(가천대) 최현호(한경대) 허재두(한국전자통신연구원)
Alwin Poulose(경북대) Ullha Ihsan(한국기술교육대)

진행시간표

2021년 9월 29일(수요일)

시 간	주제 / 강사 / 내용				
11:50-12:20	등 록				
12:20-13:20 (60“)	**특별 세션 1** ITRC 인공지능 I 좌장: 고려대 한재호 교수	**특별 세션 2** ICT 자동차 융합 좌장: 경북대 한동석 교수	**특별 세션 3** ETRI대경권 연구센터 좌장: ETRI 백장운 박사	**특별 세션 4** (ONLINE) AI데이터컴퓨팅 기술 좌장: 경북대 박혜영 교수	**특별 세션 5** (ONLINE) ITRC 복합지능ICT 좌장: 경북대 정호영 교수
13:20-13:30 (10“)	Session Break				
13:30-14:30 (60“)	**초청 강연** ETRI 김흥묵 본부장 경북대 한동석 교수				
14:30-14:40 (10“)	Session Break				
14:40-15:40 (60“)	**특별 세션 6** ITRC 인공지능 II 좌장: 고려대 한재호 교수	**특별 세션 7** 차세대AI 컴퓨팅 기술 좌장: 경북대 김재일 교수	**특별 세션 8** 지능형 사회문제해결기술 좌장: ETRI 권은정 박사	**특별 세션 9** (ONLINE) 모빌리티 플랫폼 좌장: KIAPI 조봉균 박사	**논문 발표A-1** (ONLINE) 로봇 제어 좌장: 경북대 김민영 교수
15:40-15:50 (10“)	Session Break				
15:50-16:20 (30“)	**Top Conference 초청 세션 I** DGIST 임성훈 교수 좌장: 동국대 손인수 교수			**논문발표 A-2** 인공지능 응용 좌장: 세종대 한종기 교수	**논문발표 A-3** (ONLINE) 지능형 통신 I 좌장: ETRI 백명선 박사
16:20-16:50 (30“)	**Top Conference 초청 세션 II** ETRI 손정우 박사 좌장: 동국대 손인수 교수				
16:50-17:00 (10“)	Session Break				
17:00-18:00 (60“)	**논문발표 B-1** 의료 인공지능I 좌장: 경북대 김재일 교수	**논문발표 B-2** 지능형 차량 I 좌장: 건양대 이규만 교수	**논문발표 B-3** 인공지능 응용 II 좌장: 고려대 홍민석 교수	**논문발표 B-4** (ONLINE) 지능형 통신 II 좌장: 고려대 박중석 교수	

1. 오프라인 세션은 시간표 순서대로 미팅룸 1, 2, 3에서 개최됩니다.
2. 초청강연, Top Conference 초청 세션 및 튜토리얼은 미팅룸 1과 2에서 개최됩니다.

진행시간표

2021년 9월 30일(목요일)

시 간	주제 / 강사 / 내용			
08:30-08:50	등 록			
08:50-09:00 (10“)	개회식 좌장: 동국대 손인수 교수			
09:00-10:00 (60“)	**논문발표 C-1** 의료 인공지능II 좌장: 전남대 황인태 교수	**논문발표 C-2** 지능형 차량 II 좌장: 명지대 박현희 교수	**논문발표 C-3** 인공지능 응용 III 좌장: 동국대 장혜령 교수	**논문발표 C-4** (ONLINE) 영어논문 세션 I 좌장: 인천대 강승택 교수
10:00-10:10 (10“)	Session Break			
10:10-11:10 (60“)	**튜토리얼 I** GIST 전해곤 교수 좌장: 한밭대 김은경 교수			
11:10-12:10 (60“)	**논문발표 D-1** 지능형 통신 III 좌장: 금오공대 신수용 교수	**논문발표 D-2** 지능형 보안 I 좌장: 명지대 박현희 교수	**논문발표 D-3** 인공지능 응용 IV 좌장: 서울과기대 권태수 교수	**논문발표D-4** (ONLINE) 학부생 세션 I 좌장: 인천대 강승택 교수
12:10-13:00 (50“)	Lunch			
13:00-14:00 (60“)	**튜토리얼 II** 고려대 석준희 교수 좌장: 한밭대 김은경 교수			
14:00-14:30 (30“)	**Top Conference 초청 세션 III** 고려대 김중헌 교수 좌장: 동명대 이동명 교수			
14:30-14:40 (10“)	Session Break			
14:40-15:10 (30“)	**Top Conference 초청 세션 IV** 서울대 이경한 교수 좌장: 한국기술교육대 한연희 교수			
15:10-16:10 (60“)	**논문발표 E-1** 지능형 보안 II 좌장: 경북대 김동균 교수	**논문발표 E-2** 영상처리 I 좌장: 한국산업기술대 김정곤 교수	**논문발표 E-3** 학부생 세션 II 좌장: 동의대 윤주상 교수	**논문발표 E-4** (ONLINE) 스마트 산업 I 좌장: 연세대 고정길 교수
16:10-16:20 (10“)	Session Break			
16:20-16:50 (30“)	**Top Conference 초청 세션 V** 고려대 김진규 교수 좌장: 한국기술교육대 한연희 교수			
16:50-17:50 (60“)	**논문발표 F-1** 인공지능 응용 V 좌장: 한국산업기술대 김평수 교수	**논문발표 F-2** 영상처리 II 좌장: 덕성여대 이재호 교수	**논문발표 F-3** 학부생 세션 III 좌장: 국민대 김상철 교수	**논문발표 F-4** (ONLINE) 스마트 산업 II 좌장: 조선대 신석주 교수

1. 오프라인 세션은 시간표 순서대로 미팅룸 1, 2, 3에서 개최됩니다.
2. 초청강연, Top Conference 초청 세션 및 튜토리얼은 미팅룸 1과 2에서 개최됩니다.

진행시간표

2021년 10월 1일(금요일)

<table>
<tr><th>시 간</th><th colspan="4">주제 / 강사 / 내용</th></tr>
<tr><td>08:30-09:00</td><td colspan="3">등 록</td><td rowspan="2">논문 발표G-1
(ONLINE)
스마트 산업 III
좌장: 조선대 최우열 교수</td></tr>
<tr><td>9:00-9:30
(30“)</td><td colspan="3">Top Conference 초청 세션 VI
동국대학교 장혜령 교수
좌장: 국민대 김상철 교수</td></tr>
<tr><td>09:30-10:30
(60“)</td><td>논문발표 G-2
영어논문 세션 II
좌장: 금오공대 김동성 교수</td><td>논문발표 G-3
영상처리 III
좌장: 광운대 심동규 교수</td><td>논문발표 G-4
영어논문 세션 III
좌장: 금오공대 임완수 교수</td><td>논문발표G-5
(ONLINE)
스마트 산업 IV
좌장: 이화여대 박형곤 교수</td></tr>
<tr><td>10:30-10:40
(10“)</td><td colspan="4">Session Break</td></tr>
<tr><td>10:40-11:40
(60“)</td><td colspan="4">튜토리얼 III
Naver AI 김현수 박사
좌장: 금오공대 임완수 교수</td></tr>
</table>

1. 오프라인 세션은 시간표 순서대로 미팅룸 1, 2, 3에서 개최됩니다.
2. 초청강연, Top Conference 초청 세션 및 튜토리얼은 미팅룸 1과 2에서 개최됩니다.

목 차

특별세션 1. ITRC 인공지능 I

2021년 9월 29일 (수요일) 12:20 ~ 13:20 (60분) | 좌장: 고려대 한재호 교수

S-1-1. Trends for Artificial Intelligence in Ophthalmology: A focused study for retinal images and diagnosis 002
정연우(고려대), 홍민석(고려대), 윤인식(고려대), 박중석(고려대), 한재호(고려대)

S-1-2. 시퀀스 생성을 통한 제로샷 행동 인식 해결방법 003
이제욱(연세대), 박성호(연세대), 변혜란(연세대)

S-1-3. 사전 학습된 이미지 생성 모델을 이용한 Semantic Segmentation 복원 005
안천수(고려대), 설상훈(고려대)

S-1-4. 객체의 형태 추정을 활용한 단일 객체 추적 007
박원희(고려대), 진동권(고려대), 김창수(고려대)

S-1-5. A study on Video Generation with Self-supervised Learning 008
홍기범(연세대), 문철현(연세대), 변혜란(연세대)

S-1-6. 전전두엽 뇌파 신호 기반 예측 디코딩 연구 010
홍을석(고려대), 홍기환(고려대), 민병경(고려대)

특별세션 2. ICT 자동차 융합

2021년 9월 29일 (수요일) 12:20 ~ 13:20 (60분) | 좌장: 경북대 한동석 교수

S-2-1. Equilibratory Data Augmenting Machine for Speech Emotion Recognition 013
김정환(경북대), Alwin Poulose(경북대), 한동석(경북대)

S-2-2. Motion Classifier Using Deep Learning Architecture 016
Alwin Poulose(경북대), 김정환(경북대), 한동석(경북대)

S-2-3. 영상기반 드론 인지를 통한 편대비행 및 장애물 회피 알고리즘 개발 019
장지호(경북대), 이선일(경북대), 이현범(경북대)

S-2-4. KS-FER: Kernel Sharpening Based Facial Emotion Recognition 021
Alwin Poulose(경북대), Chinthala Sreya Reddy(경북대), 김정환(경북대), 한동석(경북대)

S-2-5. MTF 기반 해상력 성능에 따른 보행자 검출 성능평가 방법에 관한 연구 024
김희강(한국건설생활환경시험연구원), 손준우(한국건설생활환경시험연구원),
김지연(한국건설생활환경시험연구원), 조태식(한국건설생활환경시험연구원), 한동석(경북대)

목 차

특별세션 3. ETRI 대경권 연구센터

2021년 9월 29일 (수요일) 12:20 ~ 13:20 (60분) | 좌장: ETRI 백장운 박사

S-3-1. 클래스 정보를 고려한 샴 네트워크 기반의 추적 방법 026
이준구(한국전자통신연구원), 오지용(한국전자통신연구원), 남승우(한국전자통신연구원)

S-3-2. 딥러닝 기반의 수정조기경보지수(Modified Early Warning Score) 예측 방법에 관한 연구..... 028
진용식(한국전자통신연구원), 김범휘(한국전자통신연구원), 김규형(한국전자통신연구원),
조성현(한국전자통신연구원), 신우상(한국생산기술연구원)

S-3-3. 중증 환자 욕창 예방을 위한 유연인쇄압력센서 기반 자세 추정 기법 연구..... 030
박찬은(한국전자통신연구원), 문애경(한국전자통신연구원)

S-3-4. 조제 자동화 시스템 적용을 위한 AI 기반 의약품 비전 검수용 데이터 확장 및 학습 방법..... 032
김휘강(한국전자통신연구원), 이상범(한국전자통신연구원), 김광용(한국전자통신연구원),
김규형(한국전자통신연구원), 권혁주(경북대), 이성학 (경북대)

S-3-5. CART 기반의 변수랭킹을 이용한 곱셈형 고장의 고장변수식별..... 035
유정원(한국전자통신연구원), 김진홍(한국전자통신연구원), 이영재(한국전자통신연구원),
임길택(한국전자통신연구원)

특별세션 4(ONLINE). AI데이터컴퓨팅 기술

2021년 9월 29일 (수요일) 12:20 ~ 13:20 (60분) | 좌장: 경북대 박혜영 교수

S-4-1. 한국노인노쇠코호트사업 데이터 분석과 활용 – 기계학습과 딥러닝 적용, 확산을 위하여..... 037
구동준(경북대), 이아라(경북대), 이은주(경북대), 김일곤 (경북대) [오프라인 현장 발표 희망]

S-4-2. 이미지 인페인팅을 위한 효과적인 제거 영역 선택 기법 039
윤상빈(경북대), 정순기(경북대)

S-4-3. Understanding Resource Usage of NoSQL Databases Through System call Trace..... 041
서창호(경북대), 채윤창(경북대), 이재륜(경북대), 서의성(성균관대), 탁병철(경북대)

S-4-4. 연안 파랑 관측 및 분석을 위한 비지도 학습 기법 연구 043
김재일(경북대), 김태경(한국해양과학기술원), 허동(DGIST), 박성준(경북대),
김진아(한국해양과학기술원)

S-4-5. 복부암 다중 오믹스 유전체 데이터 기반의 생물학기작 활성화 정도 분석 방법..... 045
전재민(경북대), 정인욱(경북대)

목 차

특별세션 5(ONLINE). ITRC 복합지능 ICT

2021년 9월 29일 (수요일) 12:20 ~ 13:20 (60분) | 좌장: 경북대 정호영 교수

S-5-1. 단기 시계열 데이터에 대한 장기간 다단계 예측을 위한 학습 방법 연구 048
김동현(경북대), 도주성(경북대), 정호영(경북대)

S-5-2. Lightweight-VGG based Facial Landmarks Detection with Inception-ResNet Module 050
Savina Jassica Colaco(경북대), MyungChul Park(경북대), DaeWoong Cha(경북대), ChaeHyun Lee(경북대), Dong Seog Han(경북대)

S-5-3. 금속 표면 결함 탐지 및 영상 분할을 위한 딥러닝 기반의 다중 클래스 결함 학습 052
박재현(경북대), 최우제(경북대), 전세호(경북대), 이동규(경북대)

S-5-4. 음성인식 오류를 포함한 텍스트 대상 언어모델 학습 방법 및 이를 이용한 음성언어 의도 분류 성능 개선에 관한 연구 054
김준우(경북대), 윤혜경(경북대), 정호영(경북대)

S-5-5. Attention Guided Feature Enhancement for Age-Related Macular Degeneration Classification in Optical Coherence Tomography 056
조정래(경북대), 박동호(경북대), 정성문(경북대)

특별세션 6. ITRC 인공지능 II

2021년 9월 29일 (수요일) 14:40 ~ 15:40 (60분) | 좌장: 고려대 한재호 교수

S-6-1. 동작 상상 뇌파 분류를 위한 합성곱 신경망 기반의 전이학습에 관한 연구 058
송하윤(고려대), 김동규(고려대), 김영탁(고려대), 김학승(고려대), 김동주(고려대)

S-6-2. 효율적인 수면단계 분류를 위한 딥러닝 모델 압축 연구 060
허재욱(고려대), 이철희(고려대), 김학승 (고려대), 김동주(고려대)

S-6-3. 효율적인 언어모델 사전학습을 위한 집중학습기법 062
박준형(고려대), 이상근(고려대)

S-6-4. 마비말장애 중증도 자동 분류를 위한 퓨샷 러닝 064
여은정(서울대), 정민화(서울대)

S-6-5. 어텐션 쌍방향 LSTM 신경망을 활용한 억양 기반 한국어 방언 자동 식별 066
이주영(서울대), 김경화(대검찰청), 정민화(서울대)

목 차

특별세션 7. 차세대 AI 컴퓨팅 기술

2021년 9월 29일 (수요일) 14:40 ~ 15:40 (60분) | 좌장: 경북대 김재일 교수

S-7-1. 부동소수점 형식에 따른 스파이킹 신경망의 학습 및 정확도 분석 068
곽명진(경북대), 서효주(경북대), 김용태(경북대)

S-7-2. 랜덤 가중치를 가지는 오류 역전파 학습의 성능 분석 070
이소하(경북대), 박혜영(경북대)

S-7-3. 건조 중인 LNG 화물창 시스템의 안정적 환경 유지를 위한 기계학습 기반 화물창 내부 습도 예측 073
김재성(경북대), 옥진성(대우조선해양), 서영균(경북대)

S-7-4. 연합학습 기법을 활용한 IoT와 Edge Computing 기반의 현장중심의 재난 탐지 및 대응에 대한 연구..... 075
우아밍(경북대), 안재광(기상청), 권영우(경북대)

특별세션 8. 지능형 사회문제해결기술

2021년 9월 29일 (수요일) 14:40 ~ 15:40 (60분) | 좌장: ETRI 권은정 박사

S-8-1. 경찰의 사건현장 대응 지원을 위한 긴급출동 의사결정 지원 시스템 077
박현호(한국전자통신연구원), 변성원(한국전자통신연구원), 권은정(한국전자통신연구원),
장동만(한국전자통신연구원), 정의석(한국전자통신연구원)

S-8-2. 시공간 범죄 위험도 추론 기술 기반의 스마트 치안 플랫폼 개발 및 검증 079
백명선(한국전자통신연구원), 박원주(한국전자통신연구원), 이희조(모비젠), 이용태(한국전자통신연구원)

S-8-3. KorBERT와 능동 학습 기법을 활용한 한국어 기반 유형 분류 모델에 관한 경험적 연구..... 081
박원주(한국전자통신연구원), 김연수(이화여대), 백명선(한국전자통신연구원), 이용태(한국전자통신연구원)

S-8-4. 저품질 영상의 초고화질 변환을 위한 실시간/비실시간 처리 시스템 설계 083
라상중(한국전자통신연구원), 조숙희(한국전자통신연구원), 김나연(과학기술연합대학원대학교)

S-8-5. 기계학습 기반의 범죄 유형 및 범죄위험스코어 예측 기술 성능 향상 연구..... 085
안다영(한양대학교), 백명선(한국전자통신연구원), 박원주(한국전자통신연구원), 이용태(한국전자통신연구원)

S-8-6. 저화질 미디어의 감정인식 문제 해결을 위한 고품질 변환 기술 현황과 성능분석 087
김나연(과학기술연합대학원대학교), 조숙희(한국전자통신연구원), 배병준(한국전자통신연구원)

목 차

특별세션 9(ONLINE). 모빌리티 플랫폼

2021년 9월 29일 (수요일) 14:40 ~ 15:40 (60분) | 좌장: KIAPI 조봉균 박사

S-9-1. 딥러닝 기반의 탑승자 머리 상해도 예측 알고리즘 개발 089
유시현(순천향대), 이지은(순천향대), 박성근(순천향대) [오프라인 현장 발표 희망]

S-9-2. 바이너리 영상과 DWA 퓨전을 통한 자율주행 시스템 090
황요섭(씨랩), 이재오(씨랩), 안종우(씨랩), 이동혁(씨랩)

S-9-3. 교통소외지역 사회문제 해결을 위한 자율주행 기반 서비스 어플리케이션에 관한 연구 092
조봉균(지능형자동차부품진흥원), 이윤화(지능형자동차부품진흥원),
김태형(지능형자동차부품진흥원), 윤경수(지능형자동차부품진흥원)

S-9-4. 머신러닝 기반 전기차 배터리 제어 관련 최신 기술 동향 093
정선호(지능형자동차부품진흥원), 윤윤기(지능형자동차부품진흥원), 강병도(지능형자동차부품진흥원)

세션 A-1(ONLINE). 로봇 제어

2021년 9월 29일 (수요일) 14:40 ~ 15:40 (60분) | 좌장: 경북대 김민영 교수

A-1-1. 물리적 정보 기반 강화학습을 이용한 로봇 팔 제어 연구 096
김성현(한국전자통신연구원), 장인국(한국전자통신연구원), 김현석(한국전자통신연구원),
박찬원(한국전자통신연구원), 박준희(한국전자통신연구원)

A-1-2. 심층강화학습을 이용한 Cellular-V2X 요구사항 적응형 혼잡제어 알고리즘 098
양우열(한밭대), 조한신(한밭대)

A-1-3. 다관절 로봇팔을 위한 모방학습 기반 역기구학 해석 기술에 관한 연구 100
최진철(한국전자통신연구원), 박찬원(한국전자통신연구원), 박준희(한국전자통신연구원)

A-1-4. 딥러닝 기반 가짜뉴스 탐지 시스템에서 뉴스 카테고리의 영향 분석 101
한윤진(동의대학교), 김성한(한국전자통신연구원), 김근형(동의대학교)

A-1-5. 길 안내 로봇의 신호등 없는 횡단보도 보행 안내를 위한 차량 인식 및 거리 측정 모듈 개발 103
최인훈(경희대), 홍승준(경희대), 김서현(경희대), 정홍주(경희대), 박수용(경희대), 허의남(경희대)

목 차

세션 A-2. 인공지능 응용 I

2021년 9월 29일 (수요일) 15:50 ~ 16:50 (60분) | 좌장: 세종대 한종기 교수

A-2-1. Time-GAN을 활용한 트레이딩 알고리즘 선택 105
이재윤(인하대), 이주홍 (인하대), 최범기(큐햇지), 송재원(밸류파인더스)

A-2-2. 학습자 정밀 평가를 위한 개인 맞춤형 테스트셋 생성 시스템 107
김연찬(태그하이브), 운봉영, (태그하이브), Agarwal Pankaj(태그하이브)

A-2-3. 국민수요반응을 위한 기계학습 예측 모델 기반 고객기준부하 산정 방법 109
이현용(한국전자통신연구원), 고석갑(한국전자통신연구원), 이병탁(한국전자통신연구원)

A-2-4. 인공지능 기초 의사결정나무학습을 위한 창의적 교육 프로그램 개발 111
최은선(제주대), 정유진(제주대), 현도익(제주대), 주연수(제주대), 김미진(제주대), 박남제(제주대)

A-2-5. 디리클레 분포를 이용한 트레이딩 알고리즘의 앙상블에 관한 연구 113
정재용(인하대), 이주홍(인하대), 최범기(큐햇지), 송재원(밸류파인더스)

세션 A-3(ONLINE). 지능형 통신 I

2021년 9월 29일 (수요일) 15:50 ~ 16:50 (60분) | 좌장: ETRI 백명선 박사

A-3-1. 제한된 프런트홀의 셀탈피 MIMO 시스템을 위한 최적화 학습 기법 115
유대성(전북대), 이훈(부경대), 홍승은(한국전자통신연구원), 박석환(전북대)

A-3-2. 톰슨 샘플링 기반 동적 프레임 슬롯 알로하 116
조영제(한밭대), 황경호(한밭대)

A-3-3. OFDM 시스템에서 딥 러닝 기반 변조 식별 118
신정완(한양대), 송건호(한양대), 장민규(한양대), 윤동원(한양대)

A-3-4. RaPP 기반 네트워크 이상탐지 시스템 120
김미르(숭실대), 계효선(숭실대), 권민혜(숭실대)

A-3-5. Nature-Inspired Microstrip Antennas For High-Directivity: Fractals and Genetics 122
서예준(인천대), 전문수(인천대), 조정현(인천대), 이창형(파동에너지극한제어연구단),
Jaume Anguera(Universitat Ramon Llull), 강승택(인천대)

목 차

세션 B-1. 의료인공지능 I

2021년 9월 29일 (수요일) 17:00 ~ 18:00 (60분) | 좌장: 경북대 김재일 교수

B-1-1. Sparse CNN을 이용한 위조 안저영상 포렌식 모델에 관한 연구 125
송호중(건양대), 한주혁(건양대), 김웅식(건양대)

B-1-2. 인공지능 기반 COVID-19 진단 시스템 보안에 관한 연구 127
손인수(동국대)

B-1-3. 멀티센서를 이용한 CNN 기반 낙상감지 시스템 129
이영욱(금오공과대), 신수용(금오공과대)

B-1-4. 독거노인 돌봄을 위한 전력데이터 기반 기계학습 이상진단 분석 131
김에덴(한국전자통신연구원), 고석갑(한국전자통신연구원), 이병탁(한국전자통신연구원)

B-1-5. 심전도 신호의 양자화 방법에 대한 심혈관 질환 분류시스템의 성능에 관한연구 133
이희창(인제대), 윤태영(인제대), 여채윤(인제대), 오현영(인제대), 지예빈(인제대), 심성우(인제대), 강대성(인제대)

세션 B-2. 지능형 차량 I

2021년 9월 29일 (수요일) 17:00 ~ 18:00 (60분) | 좌장: 건양대 이규만 교수

B-2-1. 자원이 제한된 자율주행 단말의 실시간 지능형 서비스를 위한 엣지 컴퓨팅 기반 이미지 처리 및 분석 시스템 개발 135
홍승준(경희대), 최인훈(경희대), 박수용(경희대), 정홍주(경희대), 김서현(경희대), 허의남(경희대)

B-2-2. LSTM 기반의 개선된 UAV 경로 알고리즘 137
김훈희(금오공과대), 이만희(금오공과대), 신수용(금오공과대)

B-2-3. 한국 밤시간대의 실시간 차선인식 기술 연구 139
고석현(동서울), 한지은(동서울), 임예원(동서울), 이성진(동서울)

B-2-4. 심층강화학습 기반 자율주행차량의 차선 변경 정책 안정성 평가 142
이동수(숭실대), 권민혜(숭실대)

B-2-5. 고정밀 지도를 이용한 모방 학습 기반 종단간 자율 주행 145
한승태(KAIST), 조상재(KAIST), 정명근(KAIST), 김태선(KAIST), 공승현(KAIST)

목 차

세션 B-3. 인공지능 응용 II

2021년 9월 29일 (수요일) 17:00 ~ 18:00 (60분) | 좌장: 좌장: 고려대 홍민석 교수

B-3-1. 고급 챗봇 서비스를 위한 인공지능 마이크로서비스풀 공통 플랫폼 148
손승철(한국전자통신연구원), 고석갑(한국전자통신연구원), 이병탁(한국전자통신연구원)

B-3-2. SuMul 모듈을 이용한 퓨-샷 객체 검출 방법 150
임동주(경북대), 정순기(경북대)

B-3-3. LSTM 순환신경망을 이용한 비선형 시변 전력데이터 예측 및 분석기술에 관한 연구 152
최성수(한국전기연구원), 이효동(한국전기연구원)

B-3-4. 연합학습에서 이상치 데이터가 미치는 영향 평가 154
정성우(대구대), 유준혁(대구대)

B-3-5. 시각장애인 길 안내 서비스를 위한 딥러닝 기반 횡단보도 가이드 솔루션 개발 156
김서현(경희대), 최인훈(경희대), 박수용(경희대), 정홍주(경희대), 홍승준(경희대), 허의남(경희대)

세션 B-4(ONLINE). 지능형 통신 II

2021년 9월 29일 (수요일) 17:00 ~ 18:00 (60분) | 좌장: 고려대 박중석 교수

B-4-1. 다양한 혼잡도에 적응하는 DQN 기반 Cellular V2X 분산혼잡제어 158
이은화(한국교통대), 정구선(한국교통대), 최주영(연세대), 문철(한국교통대)

B-4-2. k-NN 및 PSO 기반 실내 위치 측위 개선 방안 160
오성현(한국산업기술대), 김정곤(한국산업기술대)

B-4-3. 무선D2D 네트워크를 위한 자원 인지 및 위치 기반 클러스터링 연합학습 162
정준표(아주대), 고영배(아주대), 임성화(남서울대)

B-4-4. SDN(Software-Defined Network) Network상에서 인공지능을 활용한 네트워크장비의 장애예측 및 대응시스템구현에 관한 연구 164
함대훈(에스비정보기술)

B-4-5. 군 물류 스마트화를 위한 IoT 적용 방안 166
하원용(육군), 한충식(육군)

목 차

세션 C-1. 의료인공지능 II

2021년 9월 30일 (목요일) 9:00 ~ 10:00 (60분) | 좌장: 전남대 황인태 교수

C-1-1. 빅데이터 분석을 통한 개인정보자기결정권 영향 요인 모델 연구 169
이연수(숙명여대), 이욱한(숙명여대)

C-1-2. 전이학습 기반 사시 진단을 위한 안구운동영상 수집 및 소스 데이터 구성에 따른 성능 비교..... 171
주재한(부산대), 김석찬(부산대), 김동환(부산대), 최건호(부산대)

C-1-3. 유방 촬영술을 이용한 딥러닝 기반 병리학적 완전관해율 (pCR) 예측 연구 173
신호경(경북대), 김원화(경북대병원), 김혜정(경북대병원), 김찬호(경북대), 김재일(경북대)

C-1-4. 유방 초음파 진단 일반화 성능 향상을 위한 캡슐 네트워크 기반 분류 모델..... 175
김찬호(경북대), 김원화(경북대병원), 김혜정(경북대병원), 김재일(경북대)

세션 C-2. 지능형 차량 II

2021년 9월 30일 (목요일) 9:00 ~ 10:00 (60분) | 좌장: 명지대 박현희 교수

C-2-1. 어린이 보호구역 내 사고 상황에서의 AEB 기능 효과성에 관한 실차 실험 연구..... 177
장우영(지능형자동차부품진흥원), 최형수(지능형자동차부품진흥원), 이학주(지능형자동차부품진흥원), 김홍익(지능형자동차부품진흥원), 박지수(지능형자동차부품진흥원), 이태희(지능형자동차부품진흥원)

C-2-2. 신경망 탐색을 이용한 최적의 운전자 행위 분류 모델 181
성재호(경북대), 한동석(경북대)

C-2-3. 시각장애인 길 안내를 위한 실외 자율주행 카트 개발 183
박수용(경희대), 홍승준(경희대), 최인훈(경희대), 정홍주(경희대), 김서현(경희대), 허의남(경희대)

C-2-4. 다중 에이전트 강화학습 기반 무인운반기 최적 경로 제어..... 185
최호빈(한국기술교육대), 김주봉(한국기술교육대), 김찬명(씽크온웹), 한연희(한국기술교육대)

C-2-5. 자율주행차를 위한 실시간 거리예측 기술 연구..... 187
한지은(동서울대), 고석현(동서울대), 이성진(동서울대)

목 차

세션 C-3. 인공지능 응용 III

2021년 9월 30일 (목요일) 9:00 ~ 10:00 (60분) | 좌장: 동국대 장혜령 교수

C-3-1. 인공지능 기반 스마트 수주/발주 처리를 위한 플랫폼 연구 189
오암석(동명대학교)

C-3-2. 딥러닝 기반 온습도 예측 모델을 통한 화재 조기 감지 시스템 구축..... 191
박치원(경희대), 홍기대(경희대), 허의남(경희대)

C-3-3. 화재 빅데이터 분석을 통한 화재 대응 위험도 모델링에 대한 연구 193
권은정(한국전자통신연구원), 변성원(한국전자통신연구원), 박현호(한국전자통신연구원)

C-3-4. K-Means 군집화 알고리즘을 이용한 주기 변동 랜덤 신호의 정렬 및 이벤트 추출 195
변성훈(선박해양플랜트연구소)

C-3-5. LDAPS 예측 및 AWS 관측 데이터를 이용한 CNN-BLSTM 기반의 단기 기온 예측..... 197
전지민(광주과학기술원), 김홍국(광주과학기술원)

세션 C-4(ONLINE). 영어논문 세션 I

2021년 9월 30일 (목요일) 9:00 ~ 10:00 (60분) | 좌장: 인천대 강승택 교수

C-4-1. On the Performance Gains of Decentralized Edge Caching Schemes in 5G and Beyond..... 199
Lilian Mutalemwa(조선대), 신석주(조선대)

C-4-2. Optimal Batch Size for Fine-Tuning Pre-Trained Deep Learning Models 202
아흐마드 이자즈(조선대), 신석주(조선대)

C-4-3. Evaluating Artificial Intelligence Mitigation Techniques in the Security Attack of a Smart Factory SCADA Network..... 206
Love Allen Chijioke Ahakonye(금오공대), Cosmas Ifeanyi Nwakanma(금오공대),
이재현(금오공대), 이재민(금오공대), 김동성(금오공대)

C-4-4. Depression Group Prediction with Passive Sensing: Prediction Accuracy and Window Sizes..... 208
사비나콘(인하대), 노영태(인하대)

C-4-5. Deep Learning Utilized Semantic Alignment Evaluation for Data Interoperability Provisioning: Use Case of Web Objects based Depressive Disorder Ontologies..... 210
정일영(디엘정보기술), 강현국(고려대)

목 차

세션 D-1. 지능형 통신 III

2021년 9월 30일 (목요일) 11:10 ~ 12:10 (60분) | 좌장: 금오공대 신수용 교수

D-1-1. 그래프 신경망을 이용한 관망에서의 신호 소스 위치 예측 방법 214
고석갑(한국전자통신연구원), 김에덴(한국전자통신연구원), 이병탁(한국전자통신연구원)

D-1-2. 시각장애인 길 안내 서비스를 위한 Image Segmentation 기반 경로 보정 모듈 개발 217
정홍주(경희대), 최인훈(경희대), 홍승준(경희대), 김서현(경희대), 박수용(경희대), 허의남(경희대)

D-1-3. 학습기반 측위DB를 이용한 미수집지점에서의 측위 인프라 유사도 분석 219
지명인(한국전자통신연구원), 조영수(한국전자통신연구원) [오프라인 현장 발표 희망]

D-1-4. 딥러닝 기반 자동 변조 분류 위한 CNN 필터 성능 분석,딥러닝 기반 자동 변조 분류 위한 CNN 모델 성능 분석 221
김승환(ICT융합특성화연구센터), 김동성(금오공대), 권기협(ICT융합특성화연구센터)

D-1-5. 링크 이용률 향상을 위한 DQN 기반 TCP 혼잡제어 알고리즘 223
서상진(경북대), 조유제(경북대)

세션 D-2. 지능형 보안 I

2021년 9월 30일 (목요일) 11:10 ~ 12:10 (60분) | 좌장: 명지대 박현희 교수

D-2-1. 동형암호를 이용한 프라이버시 보호 얼굴 검증 시스템 225
심준석(부산대), 조현진(부산대), 홍윤영(부산대), 김호원(부산대)

D-2-2. 완전 동형 암호화된 데이터를 위한 근사 합성곱 신경망의 근사 범위에 따른 성능에 관한 연구 227
이정현(서울대), 노종선(서울대)

D-2-3. 네트워크 플로우 데이터 기반 이상징후 탐지 인공지능 모델 성능 비교 230
박태정(세종대), 최태정(세종대), 서성관(세종대), 손배훈(세종대), 윤주범(세종대)

D-2-4. 이종데이터 연계 기반 계층적 범죄 네트워크 분석 232
지종호(아주대), 김명준(아주대), 박명건(아주대), 신현정(아주대)

D-2-5. 탈중앙화 신원증명 기반 인증 시뮬레이션 모델에 관한 연구 234
김호윤(동명대), 신승수(동명대)

목 차

세션 D-3. 인공지능 응용 IV

2021년 9월 30일 (목요일) 11:10 ~ 12:10 (60분) | 좌장: 서울과기대 권태수 교수

D-3-1. 변동성 라벨링을 이용한 앙상블 머신러닝 트레이딩 시스템 개발 236
박명석(순천향대), 유성주(순천향대), 김재윤(순천향대)

D-3-2. 오토인코더를 이용한 불량 탐지에 관한 연구 238
김지섭(서울대), 김준한(서울대), 김상태(서울대), 심병효(서울대)

D-3-3. 해밍부호 기반 이진인공신경망 학습 240
이선용(서울대), 노종선(서울대)

D-3-4. 스마트 트렁크 오프너 시스템용 딥러닝과 레이더 시그너쳐를 이용한 발동작 인식 기술 242
송승언(DGIST), 김봉석(DGIST), 김상동(DGIST), 이종훈(DGIST)

D-3-5. 기계학습에 사용되는 테스트 케이스 생성에 관한 고찰 244
정재호(서울대), 노종선(서울대)

세션 D-4(ONLINE). 학부생 세션 I

2021년 9월 30일 (목요일) 11:10 ~ 12:10 (60분) | 좌장: 인천대 강승택 교수

D-4-1. 심전도에서 심장 박동의 특징을 추출하는 방법에 관한 연구 246
윤수진(동국대), 문소연(동국대), 권나영(동국대), Kim Donald Dong(동국대)

D-4-2. Gradient boosting Classifier와 Feature importance를 이용한 EEG 신호처리에 따른 Hyponogram예측 모델에 관한 연구 248
류진수(아주대), 김규형(아주대), 조위덕(아주대)

D-4-3. 알츠하이머형 치매 진행단계 구분을 위한 딥러닝 성능 비교 연구 250
이사랑(동의대), 송은채(동의대), 강재환(동의대), 김준석(동의대)

D-4-4. 심리상담 내담자의 심리변화 시각화시스템 252
손우정(부산대), 손동현(부산대), 김주성(부산대), 강수환(라인플러스)

D-4-5. 기상 IoT 센서 데이터를 활용한 서리 예측 연구 254
서재하(한국전자통신연구원), 안은지(국립안동대), 문애경(한국전자통신연구원), 박주영(한국전자통신연구원)

목 차

세션 E-1. 지능형 보안 II

2021년 9월 30일 (목요일) 15:10 ~ 16:10 (60분) | 좌장: 경북대 김동균 교수

E-1-1. 치안현장 의사결정 지원을 위한 상황 대응 분류 모델 257
권은정(한국전자통신연구원), 변성원(한국전자통신연구원), 박현호(한국전자통신연구원), 장동만(한국전자통신연구원), 정의석(한국전자통신연구원)

E-1-2. IoT센서와 영상 학습을 통한 조기 이상 탐지 시스템 설계 259
홍기대(경희대), 박치원(경희대), 허의남(경희대)

E-1-3. 로지스틱회귀 모델을 이용한 익명처리 데이터의 유용성 측정 261
성민경(한국정보통신기술협회), 이진규(한국정보통신기술협회), 정성룡(한국정보통신기술협회), 김윤중(이지서티), 김동례(이지서티)

E-1-4. 딥러닝 및 이미지화를 통한 악성코드 분류 시스템 263
조현진(부산대), 홍윤영(부산대), 김호원(부산대)

E-1-5. 연합학습 기반 오토인코더 모델을 이용한 네트워크 이상탐지 시스템에 관한 연구 265
계효선(숭실대), 권민혜(숭실대)

세션 E-2. 영상처리 I

2021년 9월 30일 (목요일) 15:10 ~ 16:10 (60분) | 좌장: 산기대 김정곤 교수

E-2-1. 메타버스 플랫폼에서 인터랙션을 위한 단일 RGB 카메라 기반 경량화 자세 추정 모델에 관한 연구 267
남성현(한국전자기술연구원), 송하영(한국전자기술연구원), 김영원(한국전자기술연구원)

E-2-2. 로블록스 플랫폼을 활용한 메타버스·AI 융합기반 인터랙티브 콘텐츠 제안 269
송하영(한국전자기술연구원), 남성현(한국전자기술연구원), 김영원(한국전자기술연구원)

E-2-3. Group Attention U-Net을 이용한 단일 영상 연무 제거 271
홍찬의(전남대), 박진효(전남대), 최현덕(전남대)

E-2-4. 딥 러닝 기반 지게차 충돌 방지 시스템 273
박재한(금오공과대), 신수용(금오공과대)

목 차

세션 E-3. 학부생 세션 II

2021년 9월 30일 (목요일) 15:10 ~ 16:10 (60분) | 좌장: 동의대 윤주상 교수

E-3-1. 다중 에이전트 강화학습을 이용한 UAV-자원 할당 최적화에 대한 기술 동향 275
조현희(성균관대), 김중헌(고려대)

E-3-2. 무선 이미지 전송을 위한 딥러닝 기반의 효율적인 결합 소스-채널 부호화 277
임수빈(순천향대), 김정현(순천향대), 송홍엽(연세대)

E-3-3. 독거노인 고독사 방지를 위한 다양한 실시간 행동인식 방법 279
정우진(경희대)

E-3-4. 전동킥보드 불법행위 판별을 위한 딥러닝 시스템 281
전하용(경남대), 최예원(경남대), 김찬욱(경남대), 이현성(경남대)

E-3-5. 고정지역 감시 엣지 AI 단말용 자동 DB 생성 도구 283
허재윤(영남대), 이준구(한국전자통신연구원), 최윤원(한국전자통신연구원),
백장운(한국전자통신연구원)

세션 E-4(ONLINE). 스마트 산업 I

2021년 9월 30일 (목요일) 15:10 ~ 16:10 (60분) | 좌장: 연세대 고정길 교수

E-4-1. 이기종 공작기계 상태진단을 위한 제조 및 센서 데이터 통합 관리에 관한 연구 285
이정환(한국전자통신연구원), 이문영(한국전자통신연구원), 윤성재(한국전자통신연구원),
나중찬(한국전자통신연구원), 이성희(한국전자통신연구원)

E-4-2. 멀티태스크 러닝을 이용한 심층신경망 기반 기계 소리 이상 진단 287
최지훈(KAIST), 최정우(KAIST) [오프라인 현장 발표 희망]

E-4-3. Bi-LSTM을 활용한 제주도 전력 수요 중요 변수 선정에 관한 연구 289
김진실(충북대), 전유정(충북대), 양여진(충북대), 정병진(충북대), 김재성(충북대)

E-4-4. CCTV 영상 기반 신경망을 이용한 유기물 탐지 291
박민성(마크애니), 손창대(마크애니), 최성진(마크애니), 이현지(마크애니)

목 차

세션 F-1. 인공지능 응용 V

2021년 9월 30일 (목요일) 16:50 ~ 17:50 (60분) | 좌장: 산기대 김평수 교수

F-1-1. Reverse Low-High 라벨링 방법을 이용한 XGBoost 트레이딩 시스템 개발 294
한예찬(순천향대), 박태신(순천향대), 김재윤(순천향대)

F-1-2. 디노이징 셀프 어텐션 네트워크 기반 정형 데이터 결측값 보간 기법 296
이도훈(서울시립대), 김한준(서울시립대)

F-1-3. 감염병 전파 과정을 위한 마르코프 체인 기반 SEIVR 모델 연구 298
안수진(숭실대), 권민혜(숭실대)

F-1-4. 다중안테나 통신시스템 성능 분석을 위한 회귀 기반의 확률 밀도 함수 근사 방법론 300
이영석(충남대), 이기훈(충남대), 정방철(충남대)

F-1-5. 양산라인 실시간 품질 검사를 위한 딥러닝 기반 엣지컴퓨팅 비전 검사 장치 302
김량수(한국전자통신연구원), 유학(한국전자통신연구원), 김근용(한국전자통신연구원),
김성창(한국전자통신연구원) [오프라인 현장 발표 희망]

세션 F-2. 영상처리 II

2021년 9월 30일 (목요일) 16:50 ~ 17:50 (60분) | 좌장: 덕성여대 이재호 교수

F-2-1. 한국지형에서의 카메라 기반 실시간 3차원 객체인식 기술 연구 304
한지은(동서울대), 고석현(동서울대), 임예원(동서울대), 이성진(동서울대)

F-2-2. 생성적 적대 신경망 기반 Image-to-Image translation 기술 연구 동향 306
황도연(부산대), 김재석(부산대), 최윤호(부산대)

F-2-3. 이미지 노이즈 제거를 위한 다중 오토인코더 기반 점진적 복원에 관한 연구 308
김형선(인하대), 허청환(인하대), 이한음(인하대), 황현택(인하대), 강상길(인하대)

F-2-4. 하이브리드 트랜스포머 기반 비지도 단안 카메라 영상 깊이 추정 방법 310
황승준(한국항공대), 박성준(한국항공대), 김규민(한국항공대), 백중환(한국항공대)

F-2-5. STFT를 활용한 비가시영역 형상 복원에 관한 실험적 연구 312
신의현(성균관대), 장승우(성균관대), 김광수(성균관대)

목 차

세션 F-3. 학부생 세션 III

2021년 9월 30일 (목요일) 16:50 ~ 17:50 (60분) | 좌장: 국민대 김상철 교수

F-3-1. 등간격 평면배열 안테나에서의 Deep Neural Network 기반 방위각 및 고각 탐지 314
지오근(한양대), 정현진(한양대), 서형욱(한양대), 김선우(한양대)

F-3-2. 뉴스 기반의 주가 변동폭 예측 시스템에 관한 연구 316
류재학(국민대), 조성혁(국민대), Md Habibur Rahman(국민대), 장영민(국민대)

F-3-3. Deep CNN을 통한 손 제스처 인식 및 ASL-파벳 변환 318
이건(국민대), 윤리나(국민대), Ida Bagus Krishna Yoga Utama(국민대), 장영민(국민대)

F-3-4. 행동감지를 이용한 운동자세 교정 시스템 설계 및 구현 320
이철민(동명대), 이동명(동명대)

F-3-5. 합성곱 신경망을 이용한 자동 변조 분류 기법 설계와 성능 분석 322
오영우(조선대), 최우열(조선대)

F-3-6. 다수의 병렬 가중 모델 기반 스마트팩토리 에너지 사용량 예측 알고리즘에 관한 연구 324
여현영(이화여대), 심서영(이화여대), 홍수민(이화여대), 차채연(이화여대), 박형곤(이화여대)

세션 F-4(ONLINE). 스마트 산업 II

2021년 9월 30일 (목요일) 16:50 ~ 17:50 (60분) | 좌장: 조선대 신석주 교수

F-4-1. 최대 풀링 레이어를 위한 저면적 재구성 가능한 하드웨어 구조 326
김은서(한국전자기술연구원), 신동엽(한국전자기술연구원), 임용석(한국전자기술연구원)

F-4-2. 심층 신경망 가속을 위한 유연한 곱셈 누산 연산기 기반 가변 필터 처리 시스톨릭 어레이 하드웨어 328
임현서(한국전자기술연구원), 신동엽(한국전자기술연구원), 권대길(한국전자기술연구원),
임용석(한국전자기술연구원)

F-4-3. 반도체 식각 공정에서 커널 밀도 추정 기반의 식각 종점 탐지 330
최준규(숭실대), 김보배(숭실대), 심영준(숭실대), 임성빈(숭실대)

F-4-4. 플라즈마 식각 공정에서 CNN과 KDE의 반도체 식각 종점 탐지 비교 332
김보배(숭실대), 최준규(숭실대), 심영준(숭실대), 임성빈(숭실대)

목 차

세션 G-1(ONLINE). 스마트 산업 III

2021년 10월 1일 (금요일) 8:30 ~ 9:30 (60분) | 좌장: 조선대 최우열 교수

G-1-1. Partial Isolation Type Saddle-FinFET (Pi-FinFET)s 에서 Single Charge Trap에 의한 누설전류 분석 335
박진효(전남대), 김건(전남대), 김동영(전남대), 김수연(전남대), 박제원(전남대), 이명진(전남대)

G-1-2. 유도 전동기에서 다변량 커널 밀도 추정 기반의 이상치 탐지 337
심영준(숭실대), 최준규(숭실대), 김보배(숭실대), 임성빈(숭실대)

G-1-3. 그래프 신경망을 이용한 전력 거래 파트너 매칭 서비스 339
강효은(부산대), 김용수(부산대), 홍윤영(부산대), 김호원(부산대)

G-1-4. 스마트글라스 및 경량 OCR 기반 축산동물 귀표 식별 시스템 341
전광명(인트플로우), 류인철(인트플로우), 김누리(인트플로우), 임채준(인트플로우), 오영우(조선대), 전찬준(조선대), 최우열(조선대)

G-1-5. 딥러닝을 이용한 풍력 블레이드의 상태 구분 344
김용빈(한남대), 김용민(한남대), 류상우(한남대), 김영민(한남대), 최인식(한남대)

세션 G-2. 영어논문 세션 II

2021년 10월 1일 (금요일) 9:30 ~ 10:30 (60분) | 좌장: 금오공대 김동성 교수

G-2-1. CaRFE: Candidates Recursive Feature Elimination for Black-Box Model 345
송영재(성균관대), 김광수(성균관대)

G-2-2. Performance analysis of Clipping based PAPR reduction technique in UFMC systems 347
Esmot Ara Tuli(금오공대), 이재현(금오공대), 김동성(금오공대), 이재민(금오공대)

G-2-3. Comparisons of Domain-Removal Losses in Multi-Domain Pedestrian-View Intersection Classification 349
Marcella Astrid(UST), Muhammad Zaigham Zaheer(UST), 이승익(한국전자통신연구원)

G-2-4. Deep Learning for Corrupted Image-Text Matching 351
김선우(서울대), 신성욱(서울대), 심병효(서울대)

G-2-5. Anti-Drone Systems Design: Safeguarding Airspace through Real-Time Trustworthy AI Paradigm 353
Simeon Okechukwu Ajakw(금오공대), Rubina Akter(금오공대), 김다혜(금오공대), Golam Mohatsin(금오공대), 김동성(금오공대), 이재민(금오공대)

목 차

세션 G-3. 영상처리 III

2021년 10월 1일 (금요일) 9:30 ~ 10:30 (60분) | 좌장: 광운대 심동규 교수

G-3-1. Transformer를 이용한 Image Captioning 355
김동훈(서울대), Xu Jiajie(서울대), 심병효(서울대)

G-3-2. HSV 색공간을 기반한 사람 및 차량 객체의 실시간 색상 추출 357
장시예(마크애니)

G-3-3. 특징 피라미드 네트워크의 특징 압축을 위한 레벨 간 특징 예측 방법 359
김민섭(광운대), 이종석(광운대), 심동규(광운대)

G-3-4. 초분광 데이터 분석용 인공지능 플랫폼 연구 361
김거식(한국전자통신연구원), 김계은(한국전자통신연구원), 김정은(한국전자통신연구원), 강현서(한국전자통신연구원)

세션 G-4. 영어논문 세션 III

2021년 10월 1일 (금요일) 9:30 ~ 10:30 (60분) | 좌장: 금오공대 임완수 교수

G-4-1. A Design of Traffic Accident Analysis Algorithm based on Scene Recognition 363
Min Htet Thar(동명대), 이철민(동명대), 이동명(동명대)

G-4-2. Motor imagery 디코딩을 위한 Residual Dense Network 366
데니퍼마나(성균관대), 최계원(성균관대)

G-4-3. YOLOv5 Inference Performance Analysis on Edge with different Machine Learning Frameworks 368
James Rigor Camacho(금오공대), Angela Caliwag(금오공대), 임완수(금오공대)

G-4-4. Reinforcement Learning based resource scheduling for network slicing in vehicular networks 373
Muhammad Ashar Tariq(경북대), Malik Muhammad Saad(경북대), Mahmudul Islam(경북대), Muhammad Toaha Raza Khan(경북대), Junho Seo(경북대), Dongkyun Kim(경북대)

G-4-5. A Lightweight Facial Emotion Recognition System Using Network Optimization Methods 376
Erick Valverde(금오공대), Angela Caliwag(금오공대), 임완수(금오공대)

목 차

세션 G-5(ONLINE)]. 스마트 산업 IV

2021년 10월 1일 (금요일) 9:30 ~ 10:30 (60분) | 좌장: 이화여대 박형곤 교수

G-5-1. 시계열 패턴 군집분석을 기반으로 LSTM을 이용한 단기 전력 수요 예측 384
오민지(충북대), 김주하(충북대), 엄지성(충북대), 최은선(충북대), 김재성(충북대), 조완섭(충북대)

G-5-2. 모바일 애드혹 네트워크 구축을 위한 Q-learning 기반 노드 위치 결정 전략 연구 387
김나영(이화여대), 권민혜(숭실대), 박형곤(이화여대)

G-5-3. 레벨링과 분류기를 통한 자율주행에서의 차량인식 389
이준연(동명대), 안재오(동명대)

G-5-4. 사용자의 선호도를 고려한 콘텐츠 캐싱을 위해 추천시스템을 활용하여 연속적인 공간에서 콘텐츠를 캐싱하는 심층강화학습 모델 391
이우빈(동국대), 임민중(동국대)

G-5-5. 궤적에서 객체의 중심점과 변위 추정을 통한 영상 객체 탐지 연구 393
손하영(성균관대), 이유진(성균관대), 최계원(성균관대)

제2회 한국 인공지능 학술대회

The 2nd Korea Artificial Intelligence Conference

특별세션

안과에서의 인공지능 동향: 망막 영상과 진단 연구

정연우[1], 홍민석[2], 윤인식[2], 박중석[2], 한재호[1*]
[1]고려대학교 뇌공학과, [2]고려대학교 ITRC 인공지능연구센터
*hanjaeho@korea.ac.kr

Trends for Artificial Intelligence in Ophthalmology: A focused study for retinal images and diagnosis

Yeonwoo Jeong[1], Minsuk Hong[2], Insig Yun[2], Joongsuk Park[2], Jae-Ho Han[1]
[1]Dept. Brain and Cognitive Eng., Korea Univ., [2]ITRC AI Research Center, Korea Univ.

요 약

본 논문에서는 문헌 조사를 통해 다양한 임상뿐만 아니라 기초적인 안과 질환을 감지, 선별, 진단, 모니터링하기 위한 머신 러닝, 딥 러닝 등 인공 지능 기술 활용의 최근 진전과 추세를 살펴보았다. 특별히, 심층학습 아키텍처를 적용하여 안과 질환을 인식하는 방법을 학습함으로써 임상적으로 수용 가능한 성능으로 진단률을 높일 수 있음을 확인하였으며, 인공지능은 환자와 의사 모두에게 안전장치 역할을 하고 결과를 신속하게 판단할 수 있는 보조 도구 역할과 의료 전문가 지원을 효과적으로 할 수 있음을 소개하고자 한다.

Ⅰ. 서 론

4차 산업 혁명의 소용돌이 속에서 인공 지능 (Artificial Intelligence, AI)의 사용이 의료계로 급속히 확산되고 있는 가운데 안과에서의 인공 지능 사용은 시신경 질환을 비롯한 다양한 안과 질환의 진단에 주목을 받고 있다[1-5]. 특히, 안저 검사(Funduscopy, Ophthalmoscopy), 광간섭단층촬영(Optical Coherence Tomography, OCT)을 통해 획득한 데이터로 학습한 AI 알고리즘은, 안구 및 망막 질환 검출에서 높은 정확도를 보여줌으로써 고정밀 진단에 도움이 될 수 있다. 본 발표의 목표는 다양한 임상 실습뿐만 아니라 다양한 안과 질환을 감지, 선별, 진단 및 모니터링하기 위한 머신 러닝 및 딥 러닝과 같은 AI 기술 활용의 최근 진전과 추세를 강조하는 것이다.

Ⅱ. 본론

망막 이미지와 영상을 통해 출혈, 황반 이상 (예 : 드루젠) 맥락막 이상, 망막 혈관 이상, 신경 섬유층 결함, 녹내장 시신경 유두 변화 등 다양한 질병을 관찰할 수 있다. 따라서, 딥러닝을 적용하여 안과 질환을 인식하는 방법을 학습함으로써 임상적으로 수용 가능한 성능으로 진단률을 높일 수 있다. 즉, AI는 환자와 의사 모두에게 안전장치 역할을 하고 진단 결과를 신속하게 판단할 수 있는 보조 도구 역할을 할 수 있다. 처음에 발생할 수 있는 오진 가능성을 방지하고 치료 효율성을 제공하며 환자의 신뢰도를 높여서, 결과적으로 AI는 미래에 안과에서 시행되는 진단 방식을 혁신할 수 있는 잠재력을 갖는다.

Ⅲ. 결론

본 발표에서는 문헌 조사를 통해 안과 영상에서 AI는 일차 진료 및 지역사회 환경에서 주요 안과 질환 환자를 선별, 진단 및 모니터링이 가능한 솔루션으로 사용할 수 있음을 확인할 수 있었다.

ACKNOWLEDGMENT

본 연구는 정보통신기획평가원(IITP)의 대학ICT연구센터육성지원사업과 (IITP-2021-2016-0-00464) 한국연구재단을 통해 과학기술정보통신부의 기초연구사업으로부터 지원받아 수행되었음 (NRF-2020R1A4A1018309).

참 고 문 헌

[1] Atalie, C., Alessandro, A., Felipe, A., “A Review of Deep Learning for Screening, Diagnosing, and Detection of Glaucoma Progression” Transl. Vis. Sci. Technol, vol. 9, issue 2, art. no. 42, 2020.

[2] Jiang, Z., Huang, Z., Qiu, B., Meng, X., You, Y., Liu, X., Liu, G., Zhou, C., Yang, K., Maier, A., Ren, Q., Lu, Y., “Comparative study of deep learning models for optical coherence tomography angiography,” Biomed. Opt. Express, vol. 11, no. 3, pp. 1580-1597, 2020.

[3] Yan, T., Wei, L., Yue, Y., Yin, S, “Application of machine learning in ophthalmic imaging modalities” Eye and Vision, vol. 7, no. 22, 2020.

[4] Ting, D.S.W., Peng, L., Varadarajan, A.V., Keane, P.A., Burlina, P.M., Chiang, M.F., Schmetterer, L., Pasquale, L.R., Bressler, N.M., Webster, D.R., Abramoff, M., Wong, T. Y., "Deep learning in ophthalmology: The technical and clinical considerations," Prog. Retin. Eye Res., vol. 72, art. no. 100759, 2019.

[5] Muthu, R., U., Rajendra, Hamido, F., Jen H., Chua, K., Sulatha, V., Augustinus, L., Louis, T., “Application of different imaging modalities for diagnosis of Diabetic Macular Edema: A review” Comput. Biol. Ved., vol. 66, no. 1, pp. 295-315, 2015.

시퀀스 생성을 통한 제로샷 행동 인식 해결방법

이제욱, 박성호, 변혜란
연세대학교

(hooraid, qkrtjdgh18, hrbyun)@yonsei.ac.kr

Multiple Features Generative Approach to Zero-Shot Learning

Lee Jewook, Park Sungho, Byun Hyeran
Yonsei Univ.

(hooraid, qkrtjdgh18, hrbyun)@yonsei.ac.kr

요 약

Generating an averaged feature vectors on the unseen video is intuitive way for zero-shot learning, but it crushes temporal organization in a video. To address this problem, we propose the new paradigm, which generate not a averaged feature, but features for each video. Our networks is composed of three modules, attribute encoder, generator and two discriminator.

Ⅰ. 서 론

Recently, Zero-shot Learning has been a popular topic in computer science due to it is a more practical scenario. In detail, the Zero-shot Learning scenario is to classify unseen classes which are not included in the training dataset. In zero-shot image classification, feature generating for training unseen classes is considered a powerful but simple approach.

Similarly, in zero-shot action recognition, existing methods[1,2,3] synthesize a single feature even though the video has temporal information. So we propose a new novel approach to generate multiple features for a video. Our network is composed of three modules, attribute encoder, generator, and two discriminators.

Ⅱ. 본론

For generating multiple features from a single latent vector and attribute, we first design an attribute encoder. By using Recurrent Neural Networks, the encoder can increase the number of features. Also, For ensuring diversity of latent space, we utilize the re-parametrization trick with KL divergence. It helps to augment condition and latent by sampling from the gaussian distribution.

Our network is based on Generative Adversarial Networks(GAN). The proposed network has a generator and discriminator. It is enough for existing methods because they generate a single feature from a condition. However, we aim to generate multiple features. Therefore, the network should consider the nature of temporal flow among generated features. For achieving the goal, we propose a new discriminator to scrutinize whole features.

We evaluate our proposed networks by the GZSL metric[5], which is the harmonic mean of the accuracy of seen actions and unseen actions. We set a baseline as f-CLSWGAN[6] with adaption from image classification to action recognition. This represents the model which generates an averaged feature. Though baseline records 31.4±3.2, our full model, which contains the attribute encoder and two discriminators, achieves the superior performance by 35.1±3.2.

Ⅲ. 결론

Through several experiments, we verified effectiveness of a sequence generative approach. Also, we conclude that sequence generating with enough fidelity could outperform than averaged feature generating.

ACKNOWLEDGMENT

"This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2021-2016-0-00464) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation)"

참 고 문 헌

[1] Christoph H. Lampert, Hannes Nickisch, and Stefan Harmeling. Learning to detect unseen object classes by between-class attribute transfer. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 951- 958, 2009.

[2] Zeynep Akata, Scott E. Reed, Daniel Walter, Honglak Lee, and Bernt Schiele. Evaluation of output embeddings for fine-grained image classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2927- 2936, 2015.

[3] Long Chen, Hanwang Zhang, Jun Xiao, Wei Liu, and Shih-Fu Chang. Zero-shot visual recognition using semantics-preserving adversarial embedding networks. In The IEEE Conference on Computer Vision and Pattern Recognition(CVPR), June 2018.

[4] Yashas Annadani and Soma Biswas. Preserving semantic relations for zero-shot learning. In The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June2018.

[5] Y. Xian, B. Schiele, Z. Akata, Zero-shot learning - the good, the bad and the ugly, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3077- 30

[6] Y. Xian, T. Lorenz, B. Schiele, Z. Akata, Feature generating networks for zero-shot learning, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018

사전 학습된 이미지 생성 모델을 이용한 Semantic Segmentation 복원

안천수, 설상훈
고려대학교 전기전자공학부

csahn@mpeg.korea.ac.kr, sull@korea.ac.kr

Semantic Segmentation Restoration Using Pretrained Image Synthesis Model

CheonSu Ahn, Sanghoon Sull
School of Electrical Engineering, Korea University

요 약

최근 pixel 단위로 annotation 된 semantic segmentation mask 를 이용해 이미지를 생성하는 다양한 모델이 제시되고 있다. 하지만 사전 학습된 모델을 사용하면 주어진 mask 가 불완전할 경우 제대로 된 이미지를 생성하지 못하는 문제점이 있다. 본 논문은 mask 로부터 이미지를 생성하는 사전 학습된 모델을 활용하여 품질이 떨어지는 불완전한 segmentation mask 를 완전한 mask 로 복원하는 모델을 제시한다. 이는 사전 학습된 이미지 생성 모델에 대한 추가 학습 없이 mask 생성 모델에 대한 학습을 통해서만 이루어진다. 학습 과정에서는 생성한 mask 와 완전한 segmentation mask 를 비교하고 복원한 mask 를 기반으로 생성한 이미지와 실제 이미지를 비교하여 mask 복원 성능을 높일 수 있도록 했다. 도로 주행 영상 데이터 셋인 Cityscapes 에 대해 제시한 모델을 이용하여 sparse 한 mask 와 sampling 을 통해 8 배만큼 저화질로 만든 mask 를 실제 완전한 mask 로 복원하는 실험을 진행하였다.

Ⅰ. 서 론

최근 semantic segmentation mask 가 입력으로 주어졌을 때 실제와 같은 이미지를 생성하는 다양한 모델이 제시되고 있다[1,2]. 그중에서 대표적으로 SPADE[3] 구조를 활용한 모델이 있다.

이런 semantic mask 를 입력으로 받아 이미지를 생성하는 모델에서는 학습할 때 입력으로 받는 segmentation mask 를 고화질의 pixel 단위의 annotation 으로 사용하는 경우가 많다. 이를 실제 모델을 활용할 때 불완전한 mask 에 대해서는 이미지를 제대로 생성하지 못하는 단점이 있다. 이것을 해결하기 위해서는 불완전한 mask 에 대해서 네트워크를 다시 처음부터 학습시켜야 한다. 하지만 이미지 생성 모델은 대체로 복잡하기에 처음부터 학습시키기 위해서는 컴퓨터 연산이 많이 필요하고 학습 시간이 길다는 문제점이 발생한다.

이런 문제점을 해결하기 위해 본 논문에서는 사전 학습된 이미지 생성 모델을 학습 과정에 활용하여 불완전한 mask 를 완전한 mask 로 복원하는 mask 생성 모델을 제안한다. 학습 시 이미지 생성 모델은 loss 에 대한 gradient 만 전파할 뿐 업데이트를 하지 않아 컴퓨터 연산량과 소요 시간을 절감할 수 있다. 학습 과정을 통해서 전체 모델은 실제와 유사한 복원된 mask 를 만들 수 있게 되고 이를 통해 사전 학습된 이미지 생성 모델이 실제와 같은 이미지를 생성할 수 있게 해준다.

Ⅱ. 본론

2.1 모델 구조

전체 모델 구조는 그림 1 과 같은 구조를 사용하였다. Baseline 모델로 SPADE[3]를 이용한 이미지 생성 모델과 같은 형태로 이미지와 mask 생성 모델을 각각 구성하였으며 학습 과정에서 이미지 생성 모델은 사전에 학습된 모델을 사용하여 업데이트 하지 않고 loss 에 대한 gradient 만 전파하며 mask 생성 모델만 업데이트 하게 된다.

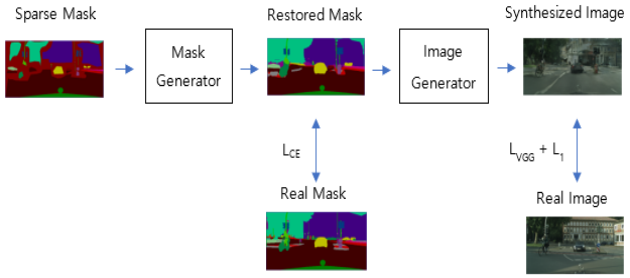


그림 1. 전체 모델 구조

Mask 에 관한 loss 는 생성한 mask 와 실제 mask 의 cross entropy loss 를 사용하였으며 복원한 mask 를 r, 실제 mask 를 x 라고 할 때 $L_{mask} = L_{CE}(r,x)$로 표현된다. 이미지 생성에 관한 loss 는 기존 [3]과 유사하게 VGG loss 와 추가적으로 L1 loss 를 사용하였으며 생성한 이미지를 s, 실제 이미지를 y 라고 할 때 $L_{image} = L_{VGG}(s,y) + L_1(s,y)$로 표현된다. 따라서 학습 과정에서 사용되는 전체 loss 는 $L_{total} = L_{mask} + L_{image}$ 가 된다.

앞에서 제시한 loss 를 통해 학습이 이루어지면 mask 생성 모델은 실제 mask 와 유사하면서도 이미지 생성 모델로 하여금 실제 이미지와 유사한 이미지를 만들도록 하는 mask 를 얻을 수 있다.

2.2 실험 결과

모델 학습과 평가를 위해 도로 주행 데이터 셋인 Cityscapes[4]를 사용하였다. Semantic segmentation mask 는 35 개의 class 로 구분되어 있고 3000 개의 학습 데이터와 500 개의 검증 데이터로 구성되어 있다.

모델을 평가하기 위해 데이터셋에 존재하는 sparse mask 를 pixel 단위로 완전히 annotation 된 완전한 mask 로 복원하는 실험과 8 배만큼 저화질로 만든 mask 를 다시 완전한 mask 로 복원하는 실험을 진행하였다. 각 실험에 대한 주요 결과는 그림 2 와 같다.

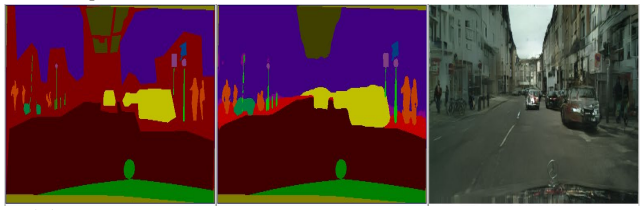
표 1 에 평가 결과를 정리하였다. 생성 이미지와 실제 이미지 분포의 차이를 측정하는 Frechet Inception Distance(FID)는 낮을수록 좋고 실제 mask 와의 유사도를 측정하는 pixel accuracy (accu), mean Intersection-over-Union (mIoU)는 높을수록 좋다.

먼저 sparse mask 를 입력으로 받아 실제 mask 와 비슷하게 복원하는 실험을 진행하였다. 제안한 모델을 사용할 경우 복원 과정을 거치지 않을 경우에 비해 FID 가 낮아지며 mask 에 대한 accu 와 mIoU 도 크게 증가함을 확인할 수 있었다. Loss 를 변경하며 결과를 비교해 보면 L_{CE} 에 L_{VGG} 와 L1 loss 를 적용했을 때 전체적으로 가장 좋은 성능을 보였다.

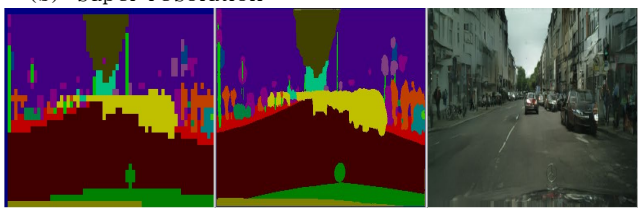
그리고 nearest 방식으로 8 배만큼 저화질로 만든 mask 를 다시 원래의 고화질 mask 로 복원하는 실험을 진행하였다. 복원 과정을 거치면서 FID 는 낮아지고 accu 와 mIoU 는 높아지는 것을 확인할 수 있었다. Loss 를 변경하며 결과를 비교해 보면 L_{CE} 에 L1 을 추가했을 때 mIoU 가 가장 높았으며 전체 loss 를 적용했을 때 FID 가 가장 낮았다.

전체 실험 결과를 볼 때 단순히 mask 에 대한 loss 인 L_{CE} 만 적용하는 것 보다 이미지에 대한 loss 인 L_{VGG} 와 L1 loss 를 적절하게 사용하면 전체 loss 를 최소화하는 과정에서 성능에 도움이 됨을 확인할 수 있었다.

(a) sparse→fine densification



(b) super resolution



(c) Real mask & image

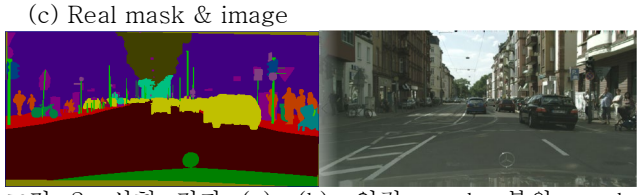


그림 2. 실험 결과 (a), (b): 입력 mask, 복원 mask, 생성 이미지 (c): 실제 mask 및 이미지

Task	Loss	FID	accu	mIoU
synthesis	-	62.9	-	-
sparse→fine densification	-	157.7	68.5	39.4
	CE+ VGG+ L1	68.4	**86.5**	**52.2**
	CE	70.7	86.1	42.1
	CE+ VGG	71.1	85.8	37.9
	CE+ L1	**68.2**	86.2	44.3
super-resolution	-	78.1	84.2	57.6
	CE+ VGG+ L1	**65.4**	**93.8**	65.2
	CE	69.2	93.3	69.5
	CE+ VGG	65.5	93.7	65.1
	CE+ L1	66.4	93.7	**70.6**

표 1. Cityscapes 데이터 셋 평가 결과. Synthesis 에서는 실제 온전한 mask 로 이미지를 생성했다. Loss 가 존재하지 않는 경우는 mask 복원 과정 없이 이미지를 생성하고 실제 이미지 및 mask 와 비교한 경우이다.

Ⅲ. 결론

본 논문에서는 사전 학습된 이미지 생성 모델을 활용하여 불완전한 semantic mask 가 주어졌을 때 완전한 mask 로 복원하는 모델을 제시하였다. 실험을 통해 mask 생성 모델을 학습할 때 실제 mask 와 비교할 뿐만 아니라 복원한 mask 로 생성한 이미지와 실제 이미지를 비교하는 방식을 이용하면 mask 를 더 잘 복원할 수 있을 뿐만 아니라 사전 학습된 이미지 생성 모델이 더 사실과 같은 이미지를 생성할 수 있음을 확인할 수 있었다. semantic segmentation task 와 같이 데이터가 부족한 상황에서 data augmentation 이 필요할 때 sparse 하거나 저화질의 semantic mask 만 있더라도 본 논문에서 제시한 방법을 통해 실제와 유사한 데이터 셋을 만들어 낼 수 있을 것이다.

ACKNOWLEDGMENT

"본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT 연구센터육성지원사업의 연구결과로 수행되었음" (IITP-2021-2016-0-00464)

참 고 문 헌

[1] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, “High-resolution image synthesis and semantic manipulation with conditional gans.” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018

[2] Xihui Liu, Guojun Yin, Jing Shao, Xiaogang Wang, et al, “learning to predict layout-to-image conditional convolutions for semantic image synthesis,” Advances in Neural Information Processing Systems (NeurIPS), 2019.

[3] Park, T., Liu, M.Y., Wang, T.C., Zhu, J.Y, “Semantic image synthesis with spatially-adaptive normalization,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019

[4] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, “The cityscapes dataset for semantic urban scene understanding,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016

객체의 형태 추정을 활용한 단일 객체 추적

박원희, 진동권, 김창수
고려대학교

whpark@mcl.korea.ac.kr, dkjin@mcl.korea.ac.kr, changsukim@korea.ac.kr

요 약

단일 객체 추적은 컴퓨터비전의 한 분야로써, 최근의 영상 내 객체 추적 기술들은 Siamese 구조의 convolutional neural network (CNN)을 통해 추출된 타겟 객체와 탐색 영역의 특징 간 유사도를 추정함으로써 이루어진다. 본 논문에서는 객체의 형태 계수화, 형태 계수 추정 네트워크를 구성, 그리고 추정된 형태를 이용하여 기존 특징정보를 개선해 객체를 추적하는 기법을 제안한다. 실험결과를 통해 제안하는 객체 형태 추정이 객체 추적 성능을 향상시킴을 확인한다.

Ⅰ. 서론

객체 추적은 컴퓨터비전의 한 분야로써 보안 감시, 자율 주행을 비롯한 응용 기술에 활용된다. 단일 객체 추적의 목표는 영상 첫 프레임의 객체 위치가 주어졌을 때, 다음 프레임에서의 객체 위치를 예측하는 것이다.

최근 객체 추적 기술은 Siamese 구조의 CNN 을 통해 추출된 타겟 객체와 탐색 영역 간의 유사도를 추정하여 이루어진다. 하지만, 네트워크의 입력 영상은 객체와 배경을 포함한 직사각형 형태이기 때문에 불필요한 특징을 포함하여 객체 추적 성능을 저하시킬 수 있다.

본 논문에서는 타겟 영상에서 객체의 형태를 계수화 하는 방법을 제안하고, 형태를 추정하고, 활용하여 객체의 특징정보를 개선하는 방법을 제안한다. 실험을 통해 객체 추적에 적용할 수 있음을 보인다.

Ⅱ. 본론

Ⅱ-1. 제안 기법

기존의 객체 추적 기법[1]은 Siamese 구조의 CNN 을 통해 추출한 타겟 객체와 탐색 영역의 특징 간 유사도를 추정함으로써 이루어진다. 하지만, 타겟 객체의 직사각형 형태의 입력은 불필요한 특징정보를 포함한다.

이 과정에서 타겟 객체의 형태를 추정 및 활용하기 위해 먼저 형태를 수치화 한다. 훈련 데이터의 객체를 객체 중심으로부터 객체 경계까지의 거리를 일정 간격 샘플링 하여 형태 벡터로 표현한다. 그리고 형태 벡터들을 특이값 분해(SVD)하여 낮은 순위 근사를 한다. 다양한 형태의 객체들을 적은 개수의 계수로 표현 가능해진다. 형태 계수는 CNN 에 형태 추정 단을 구성하여 추정한다. 학습시에는 평균 제곱 오차 손실 함수로 형태 추정 단을 학습한다. 추정된 형태 계수로는 형태 마스크를 구성하고, 이를 특징정보와의 곱 연산 후, 다시 기존 특징정보와의 합 연산을 수행한다. 이 과정은 그림 1 에서 도시 된다. 변환된 타겟 객체의 특징정보는 기존 기법[1]과 동일하게 객체 추적을 수행한다.

Ⅱ-2. 실험 결과 및 분석

단일 객체 추적을 위한 데이터셋인 VOT2018 데이터셋에서 기존 기법[1]과 제안하는 기법의 정량적 성능을 비교한다. 훈련 데이터셋은 동일하며, 실험 데이터셋에서는 '사람' 객체의 동영상만 이용해 실험을 진행한다. 표 1 은 정량적 성능의 결과이다. 기존 기법[1]과 비교하여 제안 기법을 통한 개선된 특징 정보가 객체 추적의 성능을 향상시킴을 확인할 수 있다.

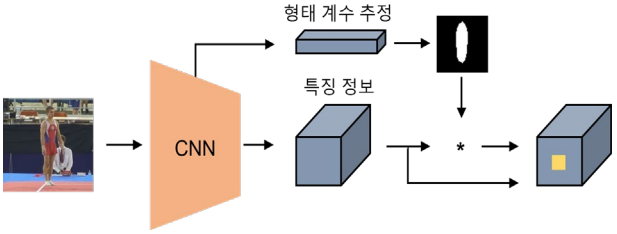


그림 1. 제안 기법의 형태 추정 및 활용

표 1 제안 기법 성능 평가

	기존 기법[1]	제안 기법
정확도(↑)	0.646	0.637
강인성(↓)	0.683	0.544
EAO (↑)	**0.235**	**0.268**

Ⅲ. 결론

본 논문은 단일 객체 추적 과정에서 객체의 형태를 활용하여 특징정보를 개선하는 기법을 제안하였다. 제안하는 기법은 객체의 형태를 수치화하고, CNN 에서 출력된 특징정보로 형태 계수를 추정하는 네트워크를 구성하였다. 추정한 형태를 활용하여 기존 특징정보와의 연산을 통해 특징정보를 개선하여 학습하였다. 실험 결과는 제안하는 기법이 객체 형태 활용이 객체 추적에 유용함을 보였다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT 연구센터육성지원사업의 연구결과로 수행되었음 (IITP-2021-2016-0-00464)

참 고 문 헌

[1] Q. Wang, L. Zhang, L. Bertinetto, W. Hu, and P.H.S. Torr, "Fast Online Object Tracking and Segmentation: A Unifying Approach", In *Proc. CVPR 2019.*

자가 학습을 통한 비디오 생성 성능 향상 연구

홍기범, 문철현, 변혜란
연세대학교

(cha2068, cheolhyunmun, hrbyun)@yonsei.ac.kr

A Study on Video Generation with Self-supervised Learning

Hong Kibeom, Mun Cheolhyun, Byun Hyeran
Yonsei Univ.

(cha2068, cheolhyunmun, hrbyun)@yonsei.ac.kr

요 약

Training the video generative models is even more sophisticated than image generative ones due to the complexity of dimension on videos. To solve this problem, we propose the novel framework to generate plausible videos with self-supervised learning. To this end, we introduce the Arrow of Time (AoT) as an additional self-superivsory task with generative adversarial networks. Finally, our networks train not only the video generation but also the direction of video plays. Our methods achieve the improvement of performance for video generation with three baselines.

Ⅰ. Introduction

Though recent studies of image generations produce real-like images with or without guidance, generating a video is more sophisticated than generating a single image since the dimension of videos is more complex. Previous methods [1,2,3] employ the GAN [4] framework for video generation. Nevertheless, they still struggle from the difficulty in generating plausible videos. To overcome this difficulty, we introduce the novel framework that critics the discriminators with better understanding about the properties of video, time. Furthermore, we explore categorical video generation with recent techniques in image generation, namely conditional batch normalization [5], spectral normalization [6], projection discriminator [7], and mode-seeking regularizer [8]. Our extensive experiments demonstrate that our framework improve the video generation performance based on three baselines.

Ⅱ. Method

Our intuition is to teach discriminators and generators to have sense of Arrow of Time (AoT) as humans do. We propose the classifying AoT as an auxiliary task for self-supervised learning and explain how the generators benefit from it.

We first induce our discriminator to distinguish real and fake videos as well as the direction of video plays (i.e., forward, and backward videos). We expect our generator G to learn that the generated videos should run forward in time as an inductive bias. We note that G only produces forward videos, but D receives both forward and backward generated videos by reversal, to encourage them to resemble real videos. We apply our novel framework on previous three baselines [1,2,3].

Furthermore, we add recent techniques in conditional image generation to gain improved performance for categorical video generation. First, we extend conditional batch normalization [5] to embed class label into the generator well. Second, the vanilla discriminator is replaced by the projection discriminator [7]. Lastly, we employ the spectral normalization layers [6] at each layer of discriminators and use mode-seeking loss [8] for stable adversarial training. We utilize three datasets: Weizmann action, UCF-sports and UCF-101 for training and evaluation.

Ⅲ. Conclusion

We proposed the self-supervised video discriminator with arrow of time. Our video discriminator not only critic the generated samples but also learn the properties of videos without additional labeled data. By doing so, our networks generate the various and realistic videos. As a results, our proposed framework improves the performance of video generation over all baselines and all datasets in terms of IS and FVD metrics.

Also, we have succeeded in generating conditional videos based on the projection method to mitigate intra-class mode dropping. Specifically, our categorical video generative method achieves the state-of-the-art IS of 28.97 and the FVD value of 26.68 on UCF 101 which is a benchmark for video generation.

ACKNOWLEDGMENT

"This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2021-2016-0-00464) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation)"

Reference

[1] Carl Vondrick, Hamed Pirsiavash, and Antonio Torralba. Generating videos with scene dynamics. In Advances In Neural Information Processing Systems, pages 613- 621, 2016.

[2] Masaki Saito, Eiichi Matsumoto, and Shunta Saito. Temporal generative adversarial nets with singular value clipping. In Proceedings of the IEEE International Conference on Computer Vision, pages 2830- 2839, 2017.

[3] Sergey Tulyakov, Ming-Yu Liu, Xiaodong Yang, and Jan Kautz. Mocogan: Decomposing motion and content for video generation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1526-1535, 2018.

[4] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, DavidWarde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Advances in neural information processing systems, pages 2672- 2680, 2014.

[5] Harm De Vries, Florian Strub, J´er´emie Mary, Hugo Larochelle, Olivier Pietquin, and Aaron C Courville. Modulating early visual processing by language. In Advances in Neural Information Processing Systems, pages 6594- 6604, 2017.

[6] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral normalization for generative adversarial networks. In International Conference on Learning Representations, 2018.

[7] Takeru Miyato and Masanori Koyama. cGANs with projection discriminator. In International Conference on Learning Representations, 2018.

[8] Qi Mao, Hsin-Ying Lee, Hung-Yu Tseng, Siwei Ma, and Ming-Hsuan Yang. Mode seeking generative adversarial networks for diverse image synthesis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1429- 1437, 2019.

전전두엽 뇌파 신호 기반 예측 디코딩 연구

홍을석, 홍기환, 민병경*
고려대학교

hong_es@korea.ac.kr, hongkh@korea.ac.kr, *min_bk@korea.ac.kr

Prefrontal EEG-based decoding of expectancy

Hong Eul-Seok, Hong Ki-Hwan, Min Byoung-Kyong*
Korea Univ.

요 약

뇌파는 인지 기능을 반영한다. 자극이 제시되기 전에 뇌파를 통해서 앞으로 나올 자극을 예측한다면 뇌파를 사용한 인지 도구의 활용을 고급화할 수 있다. 본 연구에서는 물체가 가까워지거나 멀어지는 것을 예측할 때와 예측하지 못할 때의 뇌파를 시간적 일반화 (temporal generalization) 방법으로 디코딩하였다. 선행연구에서 보고된 전전두엽의 뇌파 신호가 인지 기능에 중요하다는 결과와 베타파가 하향식 정보 처리에 관여한다는 연구 결과를 토대로 전전두엽의 베타 대역 뇌파가 예측 기능 디코딩에서 중요한 역할을 할 것이라는 가설을 수립하고, 실험을 통해 검증하고자 하였다. 뇌 영역별 디코딩 정확성을 구분하여 분석한 결과, 뇌의 후두엽이나 두정엽보다 전두엽에서 유의하게 디코딩 정확도가 증가하였으며, 전전두엽에서 가장 높은 결과를 보였다. 따라서, 전전두엽의 고주파 베타파 뇌파 신호가 뇌의 예측 기능을 최적으로 반영한다는 점을 보여준다.

Ⅰ. 서 론

뇌파(EEG)는 인간의 다양한 인지 기능을 반영한다. 그 중에 어떤 자극이 등장하는 것을 미리 예측하는 인지 기능이 자극이 등장하기 전부터 뇌파에 반영되어 있다는 점이 선행 연구에서 보고 되었다.[1] 더욱이, 전전두엽(PFC: prefrontal cortex) 지역은 고차원적 인지기능을 담당한다고 알려져 있어서 [2], 본 연구를 통해서, 예측에 대한 인지적 속성이 전전두엽 지역의 뇌파에서 특별히 뚜렷하게 반영될 것이라는 가설을 검증하고자 하였다.

뇌파는 주파수대역에 따라, 델타파(DC~4Hz), 세타파(4~8Hz), 알파파(8~13Hz), 베타파(13~30Hz), 감마파(30Hz 이상)로 나눌 수 있고, 일반적으로 주파수 대역별로 반영하는 해당 인지 기능이 연결되어 있다. 선행 연구에서 보고된 결과에 따르면, 자극 예측 기간에 베타파 활성이 향상되는 것으로 알려졌다.[3] 그리고, 우리의 선행연구 결과에서는 전전두엽 기반의 뇌파 신호가 고차원적인 인지 기능의 디코딩에 효과적으로 사용될 수 있음을 밝혀냈다.[2] 따라서, 우리는 본 연구를 통해서 전전두엽에서 베타파 대역의 뇌파 신호를 통해 자극의 변화에 대한 예측을 디코딩하고, 전전두엽의 베타파 뇌파가 예측 기능의 디코딩에 가장 중추적인 역할을 한다는 것을 검증해보고자 한다.

Ⅱ. 본론

본 논문에서는 동심원 자극 크기의 점진적 발산과 점진적 수렴을 통해 다음 자극을 예측할 때에 발생하는 뇌파를 사용해서 해당 자극 크기 변화의 예측을 디코딩하고자 하였다. 건강한 참가자 21 명이 해당 과제를 수행하는 동안의 뇌파를 측정하였다. 우리는 자극의 크기가 점진적으로 변화하는 것을 예측할 때와 그러지 못할 때의 뇌파가 다를 것이라 가정하였다.

본 실험에 사용된 하나의 시각 자극 세트는 4 개의 동심원으로 구성되었고, 각 자극 세트의 시작마다 500ms 동안의 새로운 자극 세트의 시작을 알리는 지침이 표시되었으며, 1000ms~1200ms 의 고정 십자가(fixation cross) 표시 후 동심원 자극이 700ms 동안 제시되었다. 자극 사이 간격(ISI: interstimulus interval)은 900ms~1100ms 범위로 사용되었고, 다음 자극 세트와의 간격은 900ms 로 설정되었다. 시각자극은 표적자극과 비표적자극으로 구성되었고, 표적자극은 동심원의 크기가 순차적으로 발산하거나 수렴했으며, 비표적자극은 동심원의 크기가 무작위로 변했다. 자극 세트가 표적자극인지 비표적자극인지에 대해서는 동심원의 색깔에 의해 피험자에게 표지가 되었고, 뇌파 결과에 색깔에 대한 효과를 없애기 위해, 사용된 두 가지 색인 적색과 녹색의 뇌파 결과를 최종 분석에서는 평균하여 처리하였다. 표적자극과 비표적자극은 각각 100 개의 자극 세트로 구성되었다. 또한, 작업 수행에 사용된 모든 시각자극은 맹점에 상이 맺히지 않는 시야각 안 쪽으로 제시되었다.

우선, 작업을 수행하는 참가자의 뇌파가 다음 시각자극을 예측하는지를 분석하기 위해 표적자극과 비표적자극에 대한 ISI 구간의 뇌파 데이터를 사용하였다. 해당 뇌파 신호를 SVM(Support Vector Machine)과 시간적 일반화(temporal generalization) 방법을 이용하여 디코딩하였다.[4] SVM 은 많은 분야에서 높은 정확도를 보이며, 널리 사용되고 있는 기계학습 모델이고,[5] 시간적 일반화 방법은 한 시점에서 데이터를 학습한 다음, 다른 모든 시점에서 이를 디코딩 테스트하여 상호 관계를 이용한 뇌의 정보처리 단계를 관찰할 수 있다.[4]

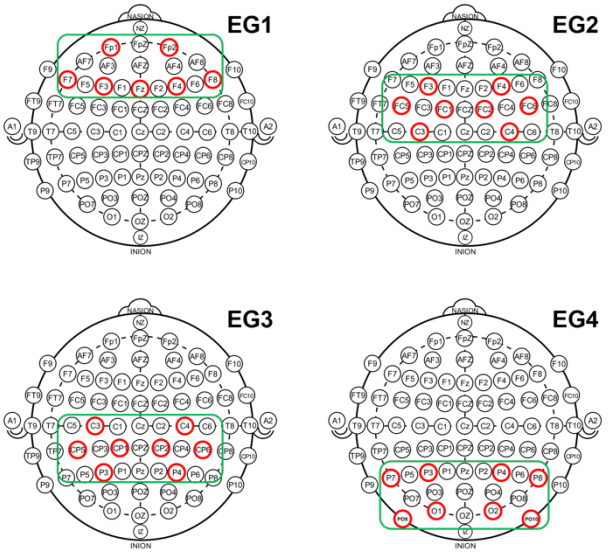


그림 1. 영역 조건별 전극 조합 분포도 (EG1 영역: Fp1, Fp2, Fz, F3, F4, F7, F8; EG2 영역: F3, F4, FC1, FC2, FC5, FC6, C3, C4; EG3 영역: C3, C4, CP1, CP2, CP5, CP6, P3, P4; EG4 영역: P3, P4, P7, P8, O1, O2, PO9, PO10)

예측과 관련된 인지 처리가 전전두엽에서 일어나는지를 검증하기 위해 뇌파 측정 전극을 [그림 1]과 같이 네 영역으로 나누었고, 예측이 베타파와 관련되어 있다는 선행연구의 결과를 바탕으로 베타 대역에 집중하여 예측 시간대에서 디코딩을 진행하였다.

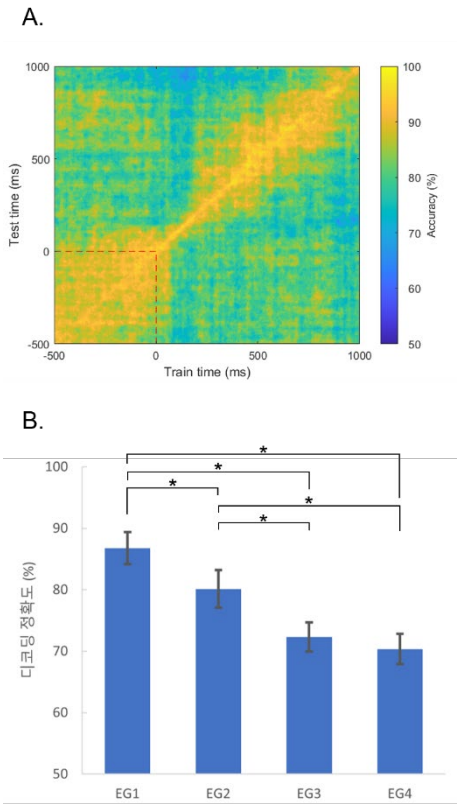


그림 2. SVM 의 디코딩 결과. A. 베타파 영역(4~30Hz)의 뇌파를 디코딩한 모든 피험자에 대한 평균 결과로 빨간색 점선은 자극 제시 이전의 예측 기간을 표시함. B. 베타 대역의 전극 조합별 SVM 의 디코딩 정확도 (평균 ± 표준오차 %)로 EG1: 86.8±2.60%, EG2: 80.1±3.06%, EG3: 72.3±2.38%, EG4: 70.3±2.46%.

결과적으로, [그림 2A]의 시각 자극 이전에 나타나는 사각형태의 높은 디코딩 정확도를 얻었고, 이는 예측의 유무에 따라 서로 다른 정신적 활동이 유지된다는 것을 의미한다.[4] 뇌 영역별 구분 디코딩 정확도 분석을 통해 [그림 2B]의 결과를 얻었고, 전전두엽에 해당하는 EG1 에서 두정엽이나 후두엽보다 유의하게 높은 디코딩 정확도가 확인되었다. 각 조건이 정규분포를 이루고 있어 대응표본 t-검정을 수행하였고, 디코딩의 정확도는 뇌의 다른 지역보다 전두엽 영역에서 유의하게 높았으며, 전전두엽에서 가장 두드러졌다. 아래 [표 1]의 결과는 q<0.05 의 FDR(false discovery rate)로 다중 비교가 보정되었고,[6] 본 논문의 모든 분석과 통계 처리는 MATLAB (ver. R2021a)에서 수행되었다.

표 1. 실험 조건 조합별 대응표본 t-검정

대응 전극 조합	t (20)	p-value
EG1-EG2	2.860	0.010*
EG1-EG3	5.982	0.000*
EG1-EG4	5.153	0.000*
EG2-EG3	3.000	0.007*
EG2-EG4	2.728	0.013*
EG3-EG4	0.766	0.453

*FDR-보정 유의성

Ⅲ. 결론

본 연구의 실험 결과를 바탕으로 살펴보건대, 예측을 디코딩하는 뇌파가 전전두엽 지역에서 우세하게 존재하고 있고, 특별히 뇌파의 베타파 영역이 이를 반영한다는 점을 알아내었다. 본 결과는 베타파가 예측에 관여한다는 선행연구 결과와 맥락을 같이 하고,[3] 전전두엽의 뇌파 신호가 인지 과제의 디코딩에 효과적으로 사용될 수 있다는 우리의 선행연구의 결과와도 방향을 같이 한다.[2] 본 연구 결과를 활용하면, 자극이 제시되기 전에 전전두엽의 뇌파만으로 앞으로 발생할 일을 예측하는 정보를 알아낼 수 있다. 따라서, 전전두엽의 뇌파가 고차원적인 인지 기능을 담당하고, 뇌-기계 인터페이스 분야를 포함해서, 뇌파 기반 인지 도구 제어의 다양한 분야에 중추적으로 사용될 수 있는 가능성과 유효성을 제시한 결과라고 할 수 있다. 본 연구를 발전시켜, 후속 연구로 좀더 구체적이고 다양한 인지 기능을 전전두엽의 뇌파를 사용하여, 정교하게 디코딩하는 방법을 살펴보고자 한다.

ACKNOWLEDGMENT

"This work was supported by the Convergent Technology R&D Program for Human Augmentation (grant number 2020M3C1B8081319), and the Information Technology Research Center (ITRC) support program (grant number IITP-2021-2016-0-00464), which are funded by the Korean government (MSICT) through the National Research Foundation of Korea."

참 고 문 헌

[1] B.-K. Min et al. “Prestimulus EEG Alpha activity Reflects Temporal Expectancy” Neuroscience Letters, 438 (3), June 2008, pp. 270-274

[2] B.-K. Min et al. “Electrophysiological Decoding of Spatial and Color Processing in Human Prefrontal Cortex” NeuroImage, 237, 118163, Aug. 2021.

[3] HajiHosseini, Azadeh, and Clay B. Holroyd. "Reward feedback stimuli elicit high-beta EEG oscillations in human dorsolateral prefrontal cortex." Scientific Reports 5.1 (2015): 1-8.

[4] J-R. King et al. "Characterizing the dynamics of mental representations: the temporal generalization method" Trends in Cognitive Sciences, 18(4), pp. 203-210, Apr. 2014.

[5] Quitadamo, L. R., et al. "Support vector machines to detect physiological patterns for EEG and EMG-based human–computer interaction: a review." Journal of neural engineering 14.1 (2017): 011001.

[6] Benjamin & Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing” Journal of the Royal Statistical Society: Series B (Methodological), 57(1), pp.289-300, 1995.

Equilibratory Data Augmenting Machine for Speech Emotion Recognition

Jung Hwan Kim, Alwin Poulose, Woo-Sung Son and Dong Seog Han*
School of Electronics and Electrical Engineering, Kyungpook National University
Republic of Korea
jkim267@knu.ac.kr, alwinpoulosepalatty@knu.ac.kr, sonws1230@knu.ac.kr, dshan@knu.ac.kr*

Abstract

The speech emotion recognition (SER) system detects a driver's emotion to prevent road rage related-accidents. The SER system handles the audio signals to recognize the driver's emotion. While building the SER system's foundation, we obtained the publicly available SER's dataset, called the Ryerson audio-visual database of emotion speech and song (RADVESS). The reliable SER system is supposed to detect the driver's emotion even in noisy environments. Nevertheless, we discovered that even simple noise could reduce the SER system's performance. In this paper, we propose an equilibratory data augmenting machine (E-DAM) on the RADVESS dataset. The proposed E-DAM could marginally improve the SER system's performance by adding noise and magnify the amount of RADVESS's samples. The performance of the SER system originally reached 71%, but it increased by approximately 3% after applying E-DAM.

I. Introduction

The speech emotion recognition (SER) system detects a driver's voice and emotion. The benefit of researching the SER dataset could detect the driver emotion more accurately, accompanying our previous research, the facial emotion recognition (FER) system [1-4]. Detecting a driver's emotion via speech is considerably challenging for many beginning researchers since it requires solid knowledge of signal processing. The driver's speech could contain environmental noises such as raindrops, car engines, and passengers' speech. Those noises could disrupt detecting a driver's emotion, although the sound is audible for our ears.

In addition, building the foundation of the SER system requires any SER dataset, so we obtained the Ryerson audio-visual database of emotion speech and song (RADVESS) [5]. The RADVESS contains samples of video, audio, and songs. Our SER system is trained with only 1,440 speech samples to detect a driver's emotion as the infuriated driver does not sing a song to threaten other drivers. Each audio wave file's name represents the properties of the audio file. The third numerical characters of the audio file name represent the speech emotion classification and have 8 different emotions: neutral, calm, happy, sad, angry, fear, disgust, and surprise. The length of each audio is 3 seconds. We merged the neutral and calm audio samples since the SER's classifications must match the FER's emotion classifications, which are 7 emotions: angry, disgust, fear, happiness, neutral, sadness, and surprise. We installed and applied Librosa [6] to load and pre-process the raw audio file. Then, all audio signals are

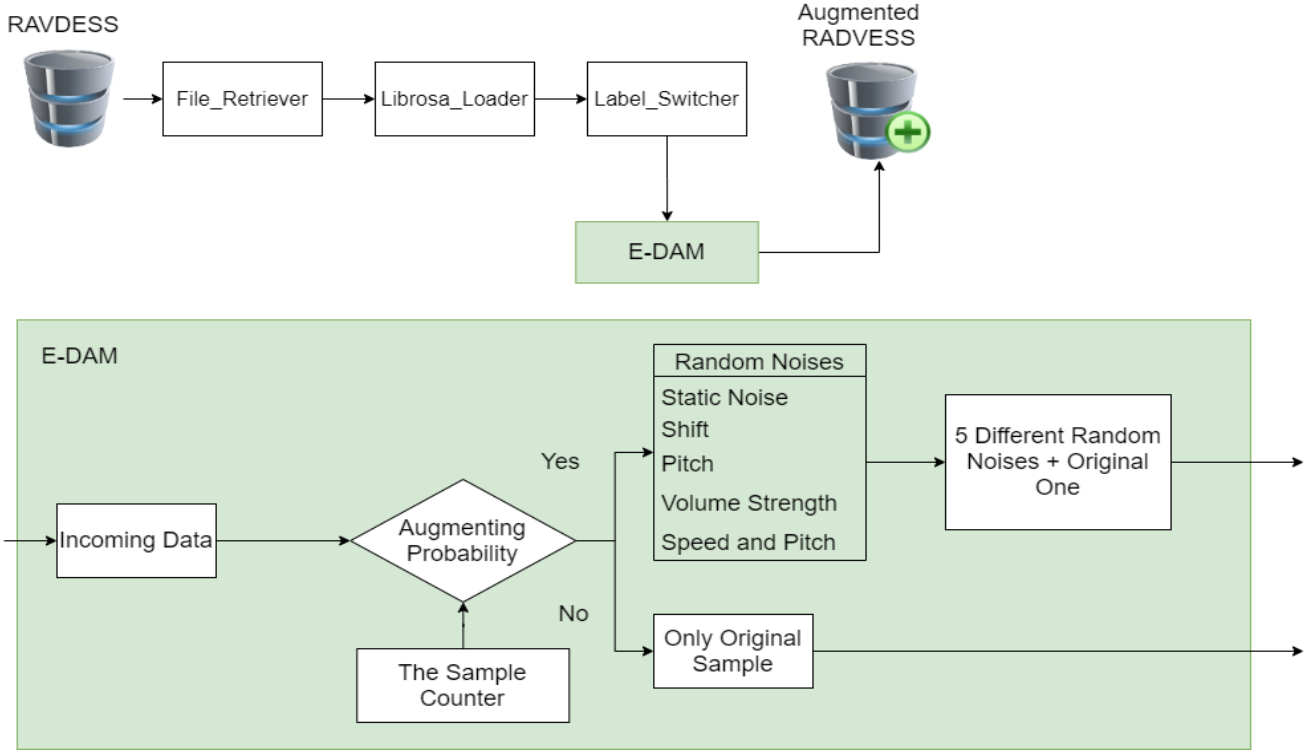
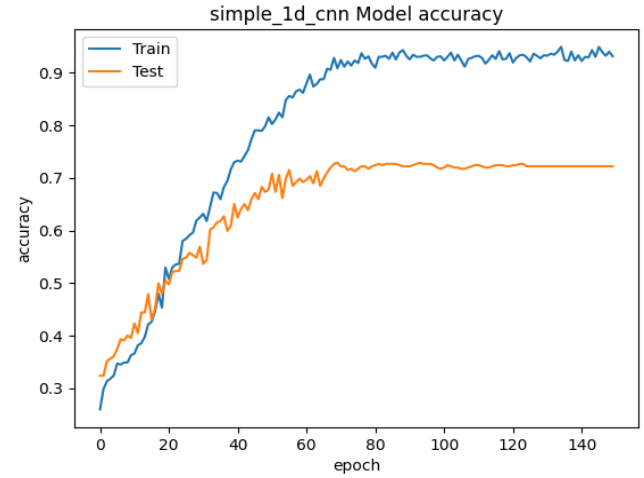


Fig. 1 Diagram of the proposed E-DAM.

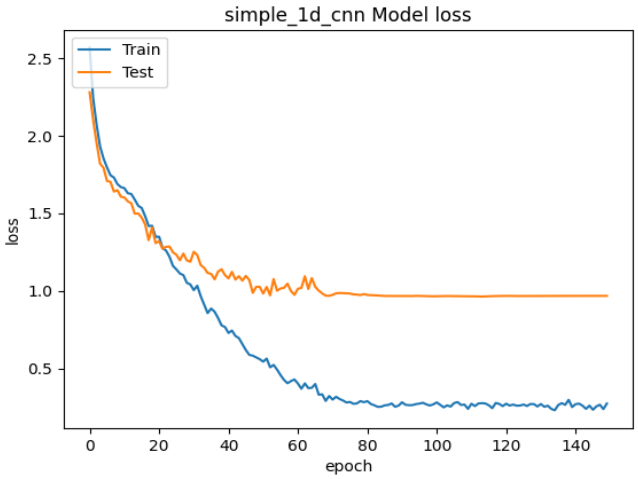
stored in the comma separable values (CSV) file.

In this paper, we propose an equilibratory data augmenting machine (E-DAM) to detect the driver's emotion in a noisy environment. E-DAM generates 5 different random noise additional audio samples per one audio sample from RADVESS. Besides, the proposed E-DAM can generate additional audio samples to balance the distribution of each class's samples.

II. The EDAM



(a)

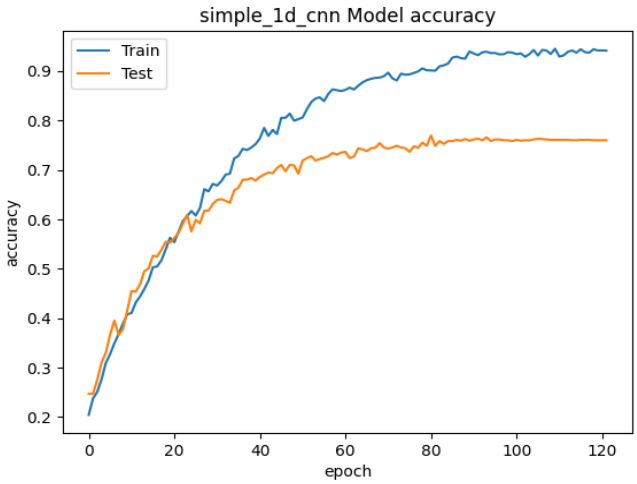


(b)

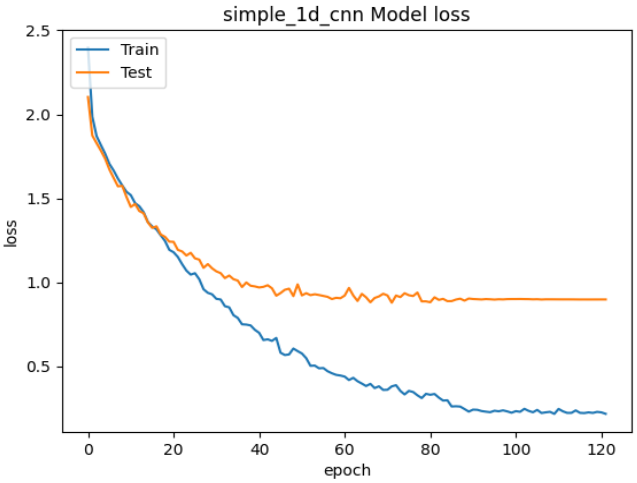
Fig. 2 The performance of the simple 1-dimentional CNN's architecture with RADVESS's training. (a) The validating accuracy. (b) The validating loss.

Fig 1. demonstrates E-DAM's structure. E-DAM consists augmenting probability, the sample counter, and random noises after all audio samples are loaded via the file retriever, Librosa loader, and label switcher.

Before entering E-DAM, the file retriever locates the audio file in the RADVESS. Librosa loader [6], which loads the file, contains tools for pre-processing raw audio signals and converts from time-sequential signals into discrete signal values. The label switcher switches the order of labels matching our previous FER's system. Before the discrete values convert into a Mel spectrogram, E-DAM generates 5 different random noises from the



(a)



(b)

Fig. 3 The performance of the simple 1-dimentional CNN's architecture with E-DAM and RADVESS's training. (a) The validating accuracy. (b) The validating

original discrete values and increases the number of audio samples per each classification. The sample counter handles increasing the samples in each classification in an equilibratory fashion. E-DAM contains 5 different random noises such as white noise, shift, speed, and strength over the original sound. After including the random noises, Librosa [6] can generate the Mel spectrogram, then converted them into decibel values. After pre-processing from the Librosa module, the dimensional channel size is reduced by calculating the decibel average from different channels. Calculating the decibel average of the different channels helps to smooth the model training. Finally, every 3 seconds of the audio file becomes 128 × 1 sized decibels, stored in the CSV file.

III. Test Results

The conditions for our experiment are 10^{-3} of the initial learning, 10^{-7} of Adam's epsilon, 10^{-2} of L2-regularizer, 150 epochs, and 50 of the patients. The software that we use is Python V. 3.8 and TensorFlow-GPU 2.2. Our computer hardware is Intel R CoreTM i5 10600k CPU @ 4.10 GHz, 32GB RAM, and GeForce RTX 2070.

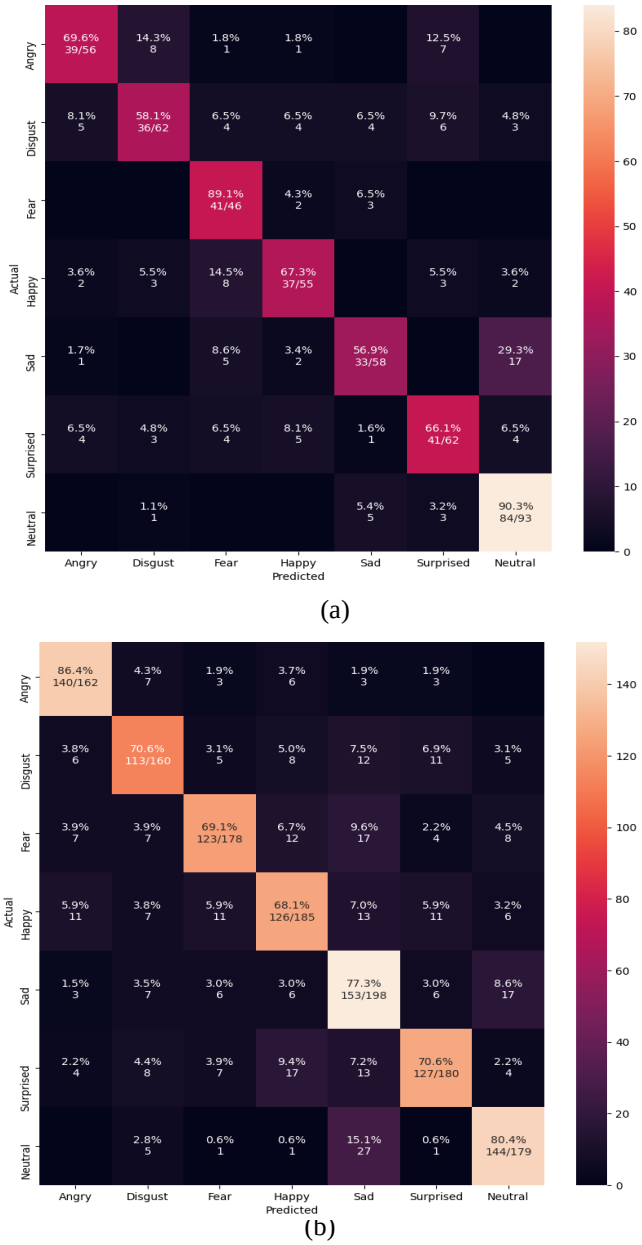


Fig. 4 The confusion matrices of the simple 1-dimentional CNN's architecture with RADVESS's training. (a) Without E-DAM. (b) With E-DAM.

As shown in Figs. 2 and 3, the simple convolution neural network (CNN) [7] for a 1-dimensional dataset reached 71% of the validating accuracy and 0.9639 of the validating loss on the original RADVESS training. From Fig. 3, E-DAM adds the random noises and increases the number of samples on the original RADVESS dataset to alleviate the overfitting problem as we previously applied on the FER dataset [1]. Applying E-DAM achieves 74% of the validating accuracy and 0.8793 of validating loss. Fig. 4 displays how the pre-trained 1-D CNN's architecture more precisely classifies speech emotion samples after E-DAM's approach. From Table I, the precision, recall, and F1-score summarize the experimental results from Fig 4. After applying E-DAM on the RADVESS dataset, the overall performance of the SER's system increased by about 3%.

IV. Conclusion

In this paper, we proposed EDAM to have improved the detection of the driver's emotions. The proposed E-DAM would solve performance problems in many disruptive environments. Increasing the number of the original SER dataset and adding the random noise deliberately for training the SER's model improves the SER system reliability. A reliable SER system precisely detects a driver's emotion even in a disruptive environment and saves more lives from preventable road rage-related accidents.

Table I. The confusion matrix summarizes the performance graph from Fig. 4.

Approaches	Precision	Recall	F1 Score
Without E-DAM	72.088 %	71.9907 %	71.4754 %
With E-DAM	**74.9147 %**	**74.5571 %**	**74.5485 %**

ACKNOWLEDGMENT

This research was supported by MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2021-2020-0-01808) supervised by the IITP (Institute of Information & Communications Technology Planning & Evaluation).

References

[1] J. H. Kim, A. Poulose, and D. S. Han, "The Extensive Usage of the Facial Image Threshing Machine for Facial Emotion Recognition Performance," Sensors, vol. 21, p. 2026, 2021.
[2] J. H. Kim and D. S. Han, "Data Augmentation & Merging Dataset for Facial Emotion Recognition," in Proceedings of the Symposium of the 1st Korea Artificial Intelligence Conference, Jeju, Korea, 2020, pp. 12-16.
[3] J. H. Kim, A. Poulose, and D. S. Han, "Facial Image Threshing Machine for Collecting Facial Emotion Recognition Dataset," The Journal of Korean Institute of Communications and Information Sciences, pp. 67-68, 2020.
[4] J. H. Kim and D. S. Han, "Data Augmentation & Merging Dataset for Facial Emotion Recognition," in Proceedings of the Symposium of the 1st Korea Artificial Intelligence Conference, Jeju, Korea, 2020, pp. 12-16.
[5] S. R. Livingstone and F. A. Russo, "The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American English," PloS one, vol. 13, p. e0196391, 2018.
[6] B. McFee, C. Raffel, D. Liang, D. P. Ellis, M. McVicar, E. Battenberg, et al., "Librosa: Audio and music signal analysis in python," in Proceedings of the 14th python in science conference, 2015, pp. 18-25.
[7] S. Malek, F. Melgani, and Y. Bazi, "One-dimensional convolutional neural networks for spectroscopic signal regression," Journal of Chemometrics, vol. 32, p. e2977, 2018.

Motion Classifier Using Deep Learning Architecture

Alwin Poulose, Jung Hwan Kim and Dong Seog Han*

School of Electronic and Electrical Engineering, Kyungpook National University, Daegu,

Republic of Korea

alwinpoulosepalatty@knu.ac.kr, jkim267@knu.ac.kr, dshan@knu.ac.kr*

Abstract

Taekwondo is a Korean martial art characterized by different types of blocks, punches, and kicks. This martial art plays a significant role in Korean culture, and most Koreans learn Taekwondo from their childhood. Amidst the covid-19 pandemic, Taekwondo martial art faces many challenges, and the Taekwondo players had to change the offline training to online classes. During the online training classes, the players are at home, and the master gives the instructions from his/her house. In such situations, practical Taekwondo training is complex, and it reduces the proper Taekwondo forms. To solve this issue during the online training classes, in this paper, we propose a Taekwondo motion classifier that classifies the Taekwondo motions and can assist the master for better Taekwondo training. The proposed motion classifier uses a visual geometry group (VGG)-13 model as the deep learning model and classifies the Taekwondo blocks and kicks. The experiment result and analysis show that the VGG model-based motion classifier archives 90.51% classification accuracy and 0.255 model loss for Taekwondo online training.

Ⅰ. Introduction

Taekwondo is the most popular martial art in Korea that provides much value to a person. Taekwondo is not only a physical skill; but also, a way of living with respect, discipline, concentration, and self-interaction [1]. The main advantages of Taekwondo include flexibility, strength, posture, stamina, and stress relieves. In addition to the physical health benefits, Taekwondo also improves a person's mental health through increased confidence, self-esteem, focus, and concentration levels [2]. During the covid-19 pandemic in our community, Taekwondo training is a challenging task, and most Taekwondo academies shut down for a long time. The covid-19 situation is not completely over and most of the Taekwondo academies started their Taekwondo training through online portals. Taekwondo training using online portals is a short time solution and the trainer faces many challenges during the online classes. The trainer faces difficulties in correcting the forms of the Taekwondo player. When the Taekwondo player performs a wrong move, it is not easy to correct the motion with correct instructions [3]. We propose a motion classifier that can assist the trainer in his/her online classes to address this issue. The proposed motion classifier has a Taekwondo motion dataset with a deep learning model that classifies the Taekwondo blocks and kicks. The motion classifier can notify the trainer when the Taekwondo player performs a wrong motion. The results from the motion classifier can help the Taekwondo trainer to effectively correct the form of the player. The experiment results from our proposed motion classifier show the significance of the proposed classifier for the online tutoring of Taekwondo.

In this paper, we proposed a motion classifier that classifies the Taekwondo blocks and kicks. The classifier uses a Taekwondo motion dataset from Taekwondo videos and this dataset is used for training the model. The proposed system utilizes a VGG-13 architecture [4] for Taekwondo motion classification. During the Taekwondo online training classes, the motion classifier differentiates different Taekwondo motions such as low block, inner block, face blocks, front snap kick, roundhouse kick, sidekick, and back kick. The experiment results from our proposed classifier show that the VGG-13 based Taekwondo motion classifier is a practical approach to assist the Taekwondo trainers for their online training classes. We also analyzed the impact of the different deep learning models such as Xception, and convolutional neural network (CNN) for our Taekwondo motion dataset and compared the performance of each model in terms of accuracy, precision, recall, and F1 score.

The rest of the paper is organized as follows. Section II presents the proposed Taekwondo motion classifier using a deep learning approach, and its experiment results and analysis are discussed in Section III. Section IV concludes our research with future research directions.

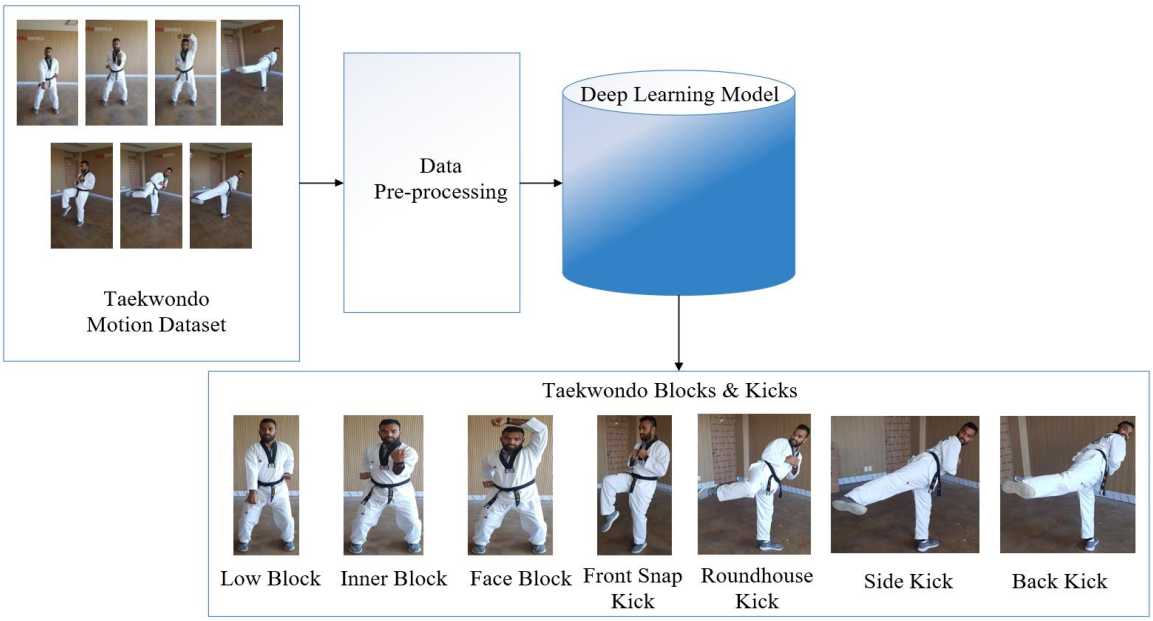


Fig. 1 Proposed Taekwondo motion classifier.

Ⅱ. Proposed Taekwondo Motion Classifier Using Deep Learning

The proposed Taekwondo motion classifier takes advantage of the VGG-13 architecture for its blocks and kicks classification. Fig. 1 shows the proposed Taekwondo motion classifier.

The proposed system consists of a Taekwondo motion dataset, data preprocessing, and VGG-13 architecture as the deep learning model. In the Taekwondo motion dataset, we created a dataset that includes different Taekwondo blocks and kicks. In the data preprocessing stage, the system performs image cropping followed by a resizing process [5]. After the cropping and resizing process, the dataset is ready for the model training. A VGG-13 architecture is used as the deep learning model, and it can classify the Taekwondo motions. The classification labels of the VGG-13 architecture include the low block, inner block, face blocks, front snap kick, roundhouse kick, sidekick, and back kick as the Taekwondo motions. Fig. 2 shows the VGG-13 model used in the proposed motion classifier.

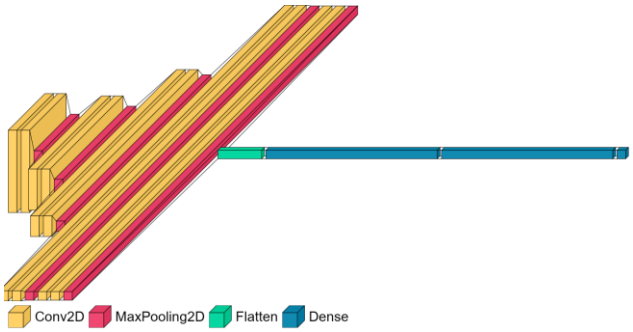


Fig. 2 VGG-13 model used in the motion classifier.

The VGG-13 model uses the input Taekwondo motion images of size 48 × 48 pixels. It uses stochastic gradient descent (SGD) as the optimizer. The color images are converted into greyscale before feeding them to the deep learning model thus reducing the number of input channels to one. This, in turn, reduces the computational complexity for training the model. The learning rate of the model is set to 0.02 with a batch size of 128. The model uses 25 epochs for training with an early stopping approach.

Ⅲ. Experiment and Result Analysis

To validate our proposed Taekwondo motion classifier for online training classes, we started with data collection. We recorded a 60-seconds video of each Taekwondo block and kick from different players and created a dataset for training the model. Fig. 3 shows our Taekwondo motion dataset.



Fig. 3 Taekwondo motion dataset.

Our dataset has 4,336 images from the low block, inner block, face blocks, front snap kick, roundhouse kick, sidekick, and back kick. The VGG-13 architecture from the motion classifier takes these Taekwondo images for its training and testing. The Taekwondo motion dataset is divided in the ratio 7:3 for training and testing respectively. Fig. 4 shows the result in the form of a confusion matrix from the VGG-13 model for the Taekwondo motion dataset.

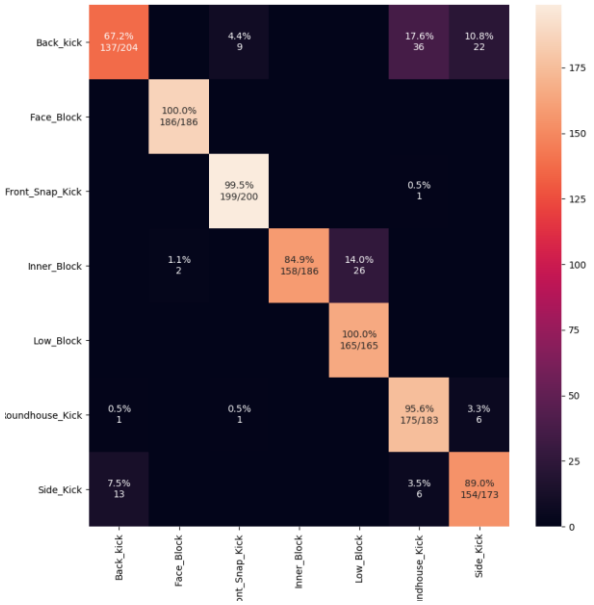


Fig. 4 Confusion matrix result from VGG-13 model.

As shown in Fig. 4, the proposed Taekwondo motion classifier achieved a classification accuracy of 90.51% and a model loss of 0.255. To validate our proposed classifier further, we also tested our Taekwondo motion dataset with the Xception and CNN deep learning models. Table 1 summarizes the performance of these deep learning models along with the VGG-13 architecture.

Table 1: Performance of deep learning models.

VGG Models	Accuracy (%)	Precision	Recall	F1 Score
VGG-13	90.51	91.08	90.51	90.26
Xception	61.21	84.77	61.21	57.37
CNN	67.61	70.47	67.61	65.54

From Table 1, we can see that the VGG-13 model shows accurate classification results for our Taekwondo motion dataset. The accuracy, precision, recall, and F1 score from the VGG-13 model have better classification results than the Xception and CNN models. The results from the CNN architecture show better classification results than the Xception model with a better model loss. The experiments and results analysis of the VGG-13 based Taekwondo motion classifier shows that the proposed approach is suitable for online Taekwondo classes. The Taekwondo trainer can easily use the advantages of the proposed motion classifier for his/her online Taekwondo classes.

IV. Conclusion

In this paper, we proposed a Taekwondo motion classifier using a deep learning architecture. The Taekwondo motion classifier can assist the Taekwondo master during his/her online training class. The classifier can effectively differentiate the Taekwondo blocks and kicks, and the Taekwondo master can quickly correct the students' blocks and kicks during an online class. The experiment and results show that the proposed approach classifies different blocks and kicks such as low block, inner block, face blocks, front snap kick, roundhouse kick, sidekick, and back kick with accurate classification results. In future work, we can create a Taekwondo training model which will be able to classify all physical activities during Taekwondo training.

Acknowledgment

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2021-2020-0-01808) supervised by the IITP (Institute of Information & Communications Technology Planning & Evaluation) and the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2021R1A6A1A03043144).

References

[1]H. Barnamehei, M. Razaghi, S. Panahi, M. Modabberibejandi, M. Lashgari, M. A. Safaei, et al., "Identification and quantification of modular control during Roundhouse kick executed by elite Taekwondo players," in 2018 25th National and 3rd International Iranian Conference on Biomedical Engineering (ICBME), 2018, pp. 1-6.

[2]U. Rahardja, M. Hardini, A. L. Al Nasir, and Q. Aini, "Taekwondo sports test and training data management using blockchain," in 2020 Fifth International Conference on Informatics and Computing (ICIC), 2020, pp. 1-6.

[3]X. Pang, "Taekwondo Video Event Detection Considering Integrated Integration Valuation Method And Its Application In Physical Education Teaching," in 2019 International Conference on Robots & Intelligent System (ICRIS), 2019, pp. 386-390.

[4]K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[5]J. H. Kim, A. Poulose, and D. S. Han, "The extensive usage of the facial image threshing machine for facial emotion recognition performance," Sensors, vol. 21, p. 2026, 2021.

영상기반 드론 인지를 통한 편대비행 및 장애물 회피 알고리즘 개발

장지호[1], 이선일[1], 이현범[1]*

[1]경북대학교

twa04047@naver.com, lseon0392@gmail.com, *hbeomlee@knu.ac.kr

Drone Formation Flight and Obstacle Avoidance Algorithm using on Vision-Based Drone Detection

Ji-Ho Jang[1], Seon-il Lee[1], Hyeonbeom Lee[1]*

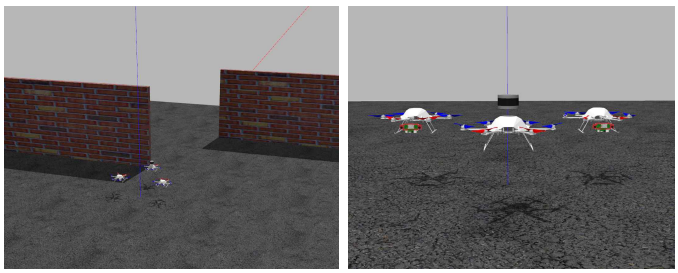
[1]Kyungpook National University, Korea

요 약

본 논문은 숲이나 고층 빌딩 옆 등 GPS 수신이 어려운 환경에서 영상 기반으로 드론의 위치를 추정하여 편대비행을 하는 장애물 회피 알고리즘을 제안한다. 리더 드론은 LiDAR 센서를 사용하여 장애물을 피해 안전한 경로를 생성 및 주행한다. 팔로워 드론은 깊이카메라를 사용하여 주행하는 리더 드론을 대상으로 객체 인식 및 추적한 뒤, 자신의 목표지점을 찾는다. 이후 저주파 통과 필터를 적용하여 부드러운 경로를 생성한다. 이 연구는 통신이 힘든 환경에서는 물론, 장애물 회피 주행을 위해 모든 드론이 고가의 센서와 컴퓨터를 필요로 하지 않아 다양한 분야에서 활용될 수 있을 것이라 기대된다.

Ⅰ. 서 론

최근, 여러 대의 드론이 대형을 맞추어 주행하는 연구[1]가 지속되고 있다. 드론 편대 비행이라 불리는 연구는 무게가 무거운 물체를 옮기기 위한 협업 운송[2], 효율적인 지역 탐사 등 다양한 분야에 사용되고 있다. 대부분의 군집 드론은 GPS 등 통신을 이용하여 현재 좌표와 경로를 공유하여 주행한다. 하지만 통신은 도심의 빌딩 옆이나 숲에서 오차가 심해져 제 기능을 못할 수가 있다. 이러한 한계를 극복하기 위해 본 논문은 통신이 없는 삼각편대 비행 방법을 제시한다.



[Fig. 1] (a) GAZEBO 시뮬레이션에서 확인한 실제 주행 환경 (b) VLP-16을 탑재한 리더 드론과 VI-Sensor를 탑재한 팔로워 드론

Ⅱ. 본론

본 논문에서는 3대의 드론이 서로간의 통신 없이 삼각 대형을 유지하며 장애물을 회피하는 시뮬레이션 기반 연구를 소개한다. 본 연구는 다음과 같은 단계로 실행된다.

1) 시뮬레이션 환경

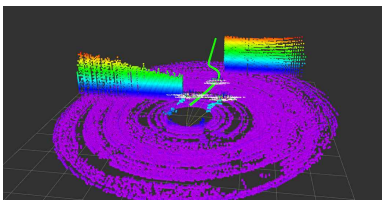
시뮬레이션 환경으로는 Ubuntu 18.04 OS에서 ROS Melodic을 사용한다. 드론 시뮬레이션 툴은 RotorS Simulator[3]를 사용한다.

2) 리더 드론의 경로생성

선두에 있는 리더 드론은 LiDAR를 기반으로 SLAM을 수행한다. 드론 상단에 Velodyne LiDAR VLP-16을 탑재한 후 포인트 클라우드 데이터를 사용하여 지도를 작성한다. 그 후 FCL을 활용하여 따라오는 팔로워 드론의 반경까지 고려했고, RRT* 기반의 경로생성 알고리즘을 사용하여 장애물을 회피하며 주행한다. 여기서 사용된 경로계획법으로는 OMPL(Open Motion Planning Library)[4]이 적용되었다.

3) 리더 드론 탐색 및 추적

리더 드론이 주행하는 동안 뒤에서 따라가는 드론은 팔로워 드론이라

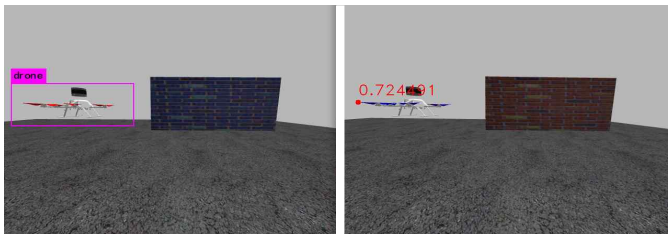


[Fig. 2] Rviz에서 확인한 LiDAR의 포인트 클라우드와 RRT*를 통해 생성된 경로

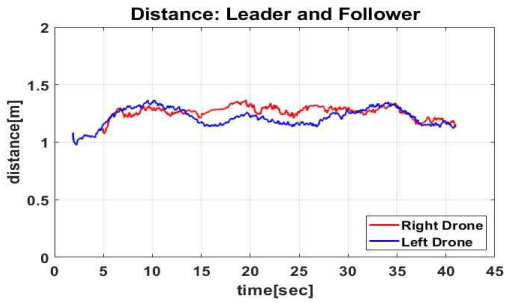
불린다. 팔로워 드론 전방에 깊이카메라인 VI-Sensor를 탑재한다. 리더 드론을 탐색 및 추적을 수행하기 위해 darknet_ros package[5]를 사용한다. 해당 패키지는 드론에 대한 학습가중치가 없기 때문에 YOLO v4로 드론만을 학습시킨 가중치를 사용해 리더 드론을 인식하여 bounding box를 생성한다.

4) 팔로워 드론의 목표좌표 추정

이어서 팔로워 드론의 목표좌표를 찾기 위한 방법을 소개한다. 우선 팔로워 드론을 원점으로 한 리더 드론의 상대 좌표를 찾는다. 리더 드론은 크기가 작고 중간에 비어있는 부분이 많기 때문에 bounding box의 중앙 좌표를 이용할 경우 깊이 정보를 추출할 때 잡음이 심해진다. 이를



[Fig. 3] (a) YOLO v4를 사용하여 리더 드론을 인식하고, Bounding Box를 생성 (b) Bounding Box 내에서 깊이의 최솟값에 해당하는 픽셀 좌표 지점



[Fig. 4] 리더 드론과 팔로워 드론 사이 거리를 나타낸 그래프

해결하기 위해 bounding box 내에서 깊이의 최솟값을 구한다.

이어 깊이의 최솟값에 해당하는 픽셀 좌표를 찾는다. 픽셀 좌표를 3차원 카메라 좌표계로 변환하는 식은 다음과 같다.

$$X^C = Z^C \times \frac{x - p_x}{f_x} \qquad Y^C = Z^C \times \frac{y - p_y}{f_y} \tag{1}$$

X^C, Y^C는 카메라 좌표계의 값에 해당하며, Z^C는 카메라 좌표계의 깊이 정보, x, y는 픽셀값, f_x, f_y는 카메라의 초점거리, p_x, p_y는 카메라의 중심 좌표를 뜻한다. 그 후 카메라 좌표계를 관성좌표계로 변환한다. 식은 다음과 같다.

$$[X^I \; Y^I \; Z^I]^T = R[X^C \; Y^C \; Z^C]^T \tag{2}$$

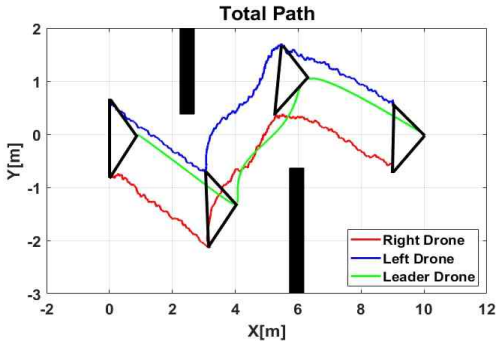
X^I, Y^I, Z^I는 관성좌표계로 변환한 리더 드론의 상대좌푯값, R은 회전 행렬이다. 리더 드론의 상대좌표에 팔로워 드론의 절대좌표를 더하여 리더 드론의 절대좌표를 구한다. 그 후 삼각 대형을 형성하기 위한 팔로워 드론의 좌표를 구하기 위해 각 좌푯값에 일정 상수를 더한다. 이 좌표를 팔로워 드론에 목표지점으로 보내어 삼각 대형을 유지한다. 이를 통해 [Fig. 4]에서 리더 드론과 팔로워 드론의 거리가 약 1.25m로 유지되는 것을 확인할 수 있다. 또한 [Fig. 5]에서 삼각대형이 유지된 채로 장애물을 피해가는 것을 확인할 수 있다.

5) 저주파 통과 필터

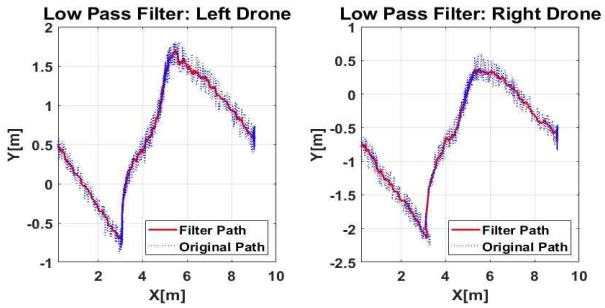
드론이 주행할 때 불안정한 흔들림으로 인해 bounding box에서 추출한 point 좌표는 잡음이 심해진다. [Fig. 6]의 파란색 그래프가 기존에 생성되었던 팔로워 드론의 경로다. 잡음을 줄이고 부드러운 경로를 생성하기 위해 저주파 통과 필터를 적용한다. 계산된 값을 다시 팔로워 드론에 목표지점으로 보내면 [Fig. 6]의 빨간색 그래프와 같이 부드러운 경로를 생성하며 주행할 수 있다.

Ⅲ. 결론

본 논문에서는 통신 없이 깊이카메라 기반으로 리더 드론의 위치를 추정하여 안전하게 장애물을 피하는 삼각편대 비행을 소개하였다. 통신이 불가한 환경에서 활용되는 것은 물론이며, 장애물 회피 주행을 위해 모든



[Fig. 5] 장애물을 회피해 주행하는 3대의 드론 주행 경로로, 삼각 대형을 유지하고 있다.



[Fig. 6] 파란색 점선은 저주파 통과 필터를 적용하기 전 팔로워 드론의 목표좌표, 빨간색 실선은 저주파 통과 필터를 적용한 후 팔로워 드론의 목표좌표

드론에 고가의 센서와 컴퓨터를 탑재할 필요가 없어 상업용으로 활용될 가능성도 높다. 이러한 연구는 향후 물류창고 내 물류운반, 동굴과 같은 환경에서의 탐사 등 다양한 분야에 응용될 것이다.

ACKNOWLEDGMENT

이 논문은 2021년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(2021R1A6A1A03043144)

참 고 문 헌

[1] F. Schilling, F. Schiano, and D. Floreano, “Vision-based Drone Flocking in Outdoor Environments,” IEEE Robotics and Automation Letters (RA-L), vol. 6, no. 2, pp. 2954-2961, 2021.

[2] H. Lee, H. Kim, W. Kim and H. J. Kim, “An Integrated Framework for Cooperative Aerial Manipulators in Unknown Environments,” in IEEE Robotics and Automation Letters, vol. 3, no. 3, pp. 2307-2314, July 2018.

[3] F. Furrer, M. Burri, M. Achtelik, & R. Siegwart, “RotorS—A modular gazebo MAV simulator framework.” In Robot operating system (ROS) (pp. 595-625). Springer, Cham, 2016.

[4] I. A. Sucan, M. Moll and L. E. Kavraki, “The Open Motion Planning Library,” in IEEE Robotics & Automation Magazine, vol. 19, no. 4, pp. 72-82, Dec. 2012.

[5] M. Bjelonic. “YOLO ROS: Real-time object detection for ROS.” leggedrobotics, 2018.

KS-FER: Kernel Sharpening Based Facial Emotion Recognition

Alwin Poulose[1], Chinthala Sreya Reddy[2,3], Jung Hwan Kim[1] and Dong Seog Han[1, *]

[1]School of Electronic and Electrical Engineering, Kyungpook National University, Daegu, Republic of Korea

[2]School of Computer Science and Engineering, Kyungpook National University, Daegu, Republic of Korea

[3]Department of Computer Science, CHRIST University, Bangalore, India

alwinpoulosepalatty@knu.ac.kr[1], sreyareddy2000@gmail.com[2,3], jkim267@knu.ac.kr[1], dshan@knu.ac.kr[1*]

Abstract

This paper proposes a kernel sharpening-based facial emotion recognition (KS-FER) system using the deep learning Xception architecture. The conventional FER systems use raw facial images from a camera for emotion recognition. However, the raw images contain background noise which reduces the performance of the FER system. The deep learning models in the FER systems have challenges extracting accurate facial features from the raw input images if the background noise ratio is large. To address the background noise in the facial images, in this paper, we propose a kernel sharpening technique for facial images, and this approach enhances the facial features of the input images. The kernel sharpening accentuates the edges of facial images and adds contrast to them. The kernel sharpening-based facial images provide more accurate input features to the deep learning model than the raw facial input images by removing the background noise from the facial images. Our KS-FER system's experiment and result analysis show that the proposed approach has a 16.54% improvement in accuracy than the conventional FER approach. The proposed KS-FER approach also reduces the loss of the Xception model from 0.58 to 0.027 for emotion recognition.

Ⅰ. Introduction

The facial emotion recognition (FER) system classifies the emotions and assists the driver monitoring systems (DMS) for safe driving [1]. In FER, the system uses facial images as input and identifies the user's emotion based on his/her facial expressions [2]. The results from the FER system are helpful for safe driving as it can reduce road accidents in autonomous driving systems (ADS) [3]. The conventional FER systems use facial images from the cameras and it classifies them into seven user emotions: happiness, sadness, anger, surprise, disgust, fear, and neutrality. When the conventional FER systems use the raw facial images from the camera, the input images are not free from the background noise. The deep learning model in the FER system fails to classify the user emotions accurately due to the lack of proper input features. The background noise in the facial images reduces the classification accuracy, and the classification results from the deep learning model have a higher degree of errors. To address the background noise in the facial images, in this paper, we propose a kernel sharpening-based facial emotion recognition (KS-FER) system for emotion recognition. The KS-FER system reduces the background noise in the facial images and improves the features that the deep learning model can effectively use. The results from the KS-FER system show that this approach is an effective method for emotion recognition.

In this paper, we propose a KS-FER system that can classify seven user emotions with lesser errors. In KS-FER, the kernel sharpening highlights the edges of the facial images and makes the transitioning of facial features more significant to the deep learning model. The kernel sharpening emphasizes the differences in adjacent pixel values that make the facial images more vivid. The final images after kernel sharpening are enhanced facial images with lesser background noise. The deep learning model uses these images as its input to classifies the user emotion. Our FER system uses the Xception architecture as the deep learning model, and the kernel sharpened images provide accurate facial features to this model. The classification results from our KS-FER show that the kernel sharpening-based FER approach reduces the classification error and the Xception model effectively uses the kernel sharpened facial images.

The rest of the paper is organized as follows. Our proposed FER system using kernel sharpening is presented in Section II with a detailed explanation of the KS-FER system. Section III discusses our FER experiments and results with a comparative analysis of proposed KS-FER and

conventional FER approaches. Section IV concludes this paper with future research directions.

Ⅱ. Proposed FER System Using Kernel Sharpening

The proposed KS-FER system effectively utilizes foreground extraction [4], kernel filtering, and the facial image threshing machine (FIT) [5] for image preprocessing. Fig. 1 shows the proposed KS-FER system.

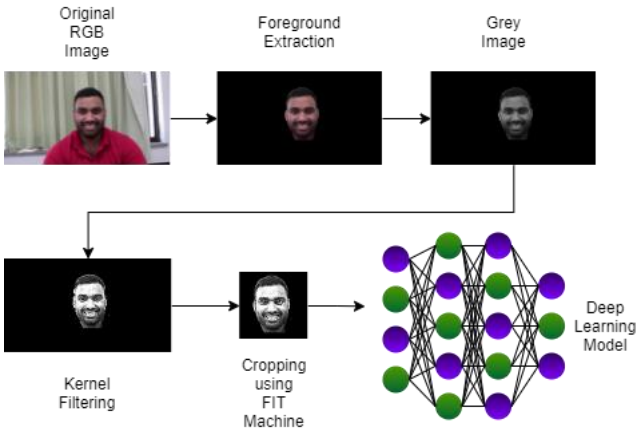


Fig. 1 Proposed KS-FER system.

In the proposed KS-FER system, we start our FER data collection from different users using a camera. The FER dataset contains seven facial expressions which are processed by a foreground extraction technique. The foreground extraction technique removes the background of the images from the raw images and retains the face of the person. After foreground extraction, we convert the images into greyscale which reduces the computational complexity of the system. A kernel sharpening technique is applied to the greyscale images, which enhances the edges of the facial images with significant features. Fig. 2 shows the raw images along with their kernel sharpened images of the different user emotions.

As shown in Fig. 2, the kernel sharpening process highlights the emotions of the user and improves the features for better classification. A FIT machine is used after the kernel sharpening process which resizes and crops the facial images. The output from the FIT machine is a suitable image to feed the deep learning model, and the model can efficiently train with these images. An Xception architecture is used as the deep learning model, which classifies the seven user emotions. Fig. 3 shows the Xception architecture used in our KS-FER system.

As shown in Fig. 3, the Xception model uses an image size of 48×48 as its input with Adam as its optimizer. The model uses 0.02 as its learning rate with a single channel and a batch size of 128. The model is set to train for 100 epochs with the early stopping method. The proposed KS-FER and conventional FER approaches use the same model configuration except for the size of input images. The conventional FER approach uses 88×50 as the input image size as we maintain the aspect ratio of the original image.

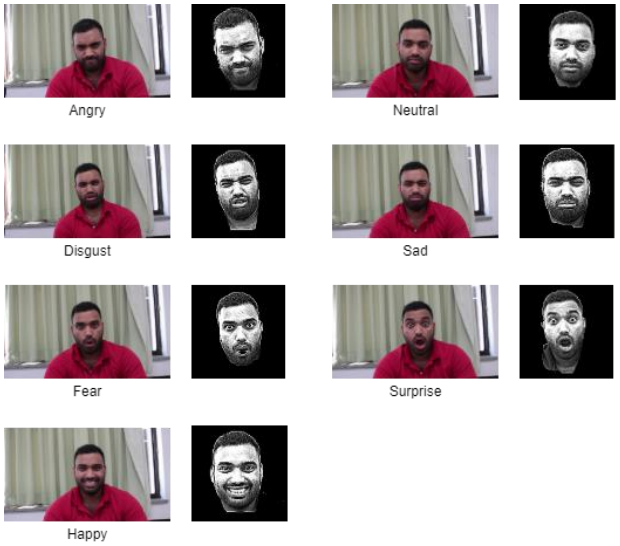


Fig. 2 Kernel sharpening on user emotions.

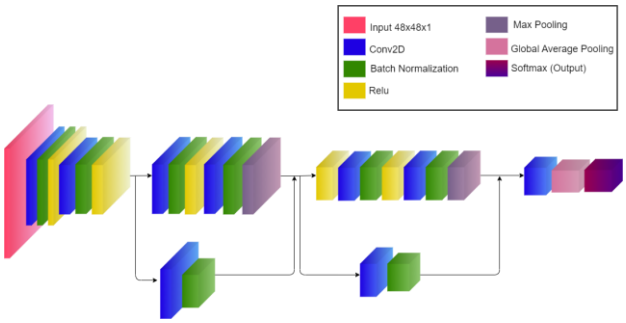


Fig. 3 The Xception model used in the KS-FER system.

Ⅲ. Experiment and Result Analysis

To validate our KS-FER system for emotion recognition, we collected images of facial expressions made by 9 people for the following emotions: happiness, sadness, angry, surprise, disgust, fear, and neutrality. Our FER dataset consists of 9,000 facial images, and we used 60% of the facial images for model training, 20% for validation, and 20% for testing the model. The Xception model used in our KS-FER system uses this dataset, and the confusion matrix results from proposed and conventional approaches are shown in Fig. 4.

As shown in Fig. 4, the conventional FER system uses the raw images without foreground extraction or kernel sharpening and has a higher classification error. The conventional approach achieved an accuracy of 82.59% with a model loss of 0.583. The KS-FER approach achieved 99.1% model accuracy with a model loss of 0.027. These results indicate that the proposed approach has an improvement of 16.54% in accuracy when compared to the conventional FER approach. The performance of our KS-FER and conventional FER approach is summarized in Table 1.

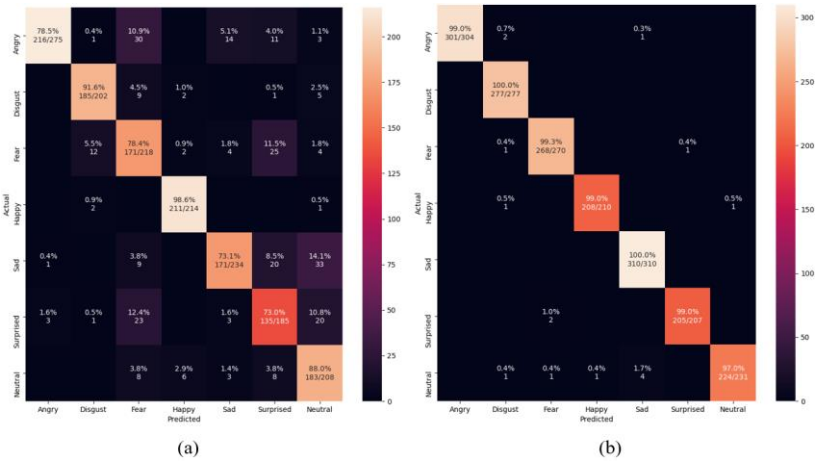


Fig. 4 Performance of the FER approaches. (a) Conventional FER approach. (b) Proposed KS-FER approach.

Table 1: Performance comparison of proposed and conventional FER approaches.

FER Approach	Accuracy (%)	Precision	Recall	F1 Score
Proposed KS-FER	99.13%	99.12	99.11	99.11
Conventional FER	82.59%	84.13	82.81	83.02

From Table 1, we can see that the KS-FER system outperforms than conventional FER approach in terms of accuracy, precision, recall, and F1 score. The kernel sharpening approach improves the features of the facial images, and the deep learning model can easily interpret the expressions in the facial images. The foreground extraction technique reduces the background noise. The kernel sharpening and foreground extraction together enhance the facial features in the images for accurate classification. From the experiment and result analysis, it is clear that the KS-FER approach has a significant improvement in facial emotion recognition systems.

IV. Conclusion

In this paper, we proposed a KS- FER system for emotion recognition. The proposed approach effectively utilizes the foreground extraction and kernel sharpening of the facial images and improves the input features learnt to the deep learning model. The system uses the Xception architecture to classify seven user emotions. The experiments and results show that the KS-FER system improves the classification accuracy and reduces the model loss for emotion recognition. We can add Gaussian blur and unsharp mask image processing techniques to the facial images for better FER performance in future work.

Acknowledgment

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2021-2020-0-01808) supervised by the IITP (Institute of Information & Communications Technology Planning & Evaluation) and the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2021R1A6A1A03043144).

References

[1] D. Y. Choi and B. C. Song, “Facial Micro-Expression Recognition Using Two-Dimensional Landmark Feature Maps," IEEE Access, vol. 8, pp. 121549-121563, 2020.

[2] J. H. Kim, A. Poulose, and D. S. Han, " Facial Image Threshing Machine for Collecting Facial Emotion Recognition Dataset,” in *Proceedings of the Symposium of the Korean Institute of communications and Information Sciences (KICS) Fall Conference*, Seoul, Korea, 13 November 2020; pp. 67- 68.

[3] J. H. Kim, R. Mutegeki, A. Poulose, and D. S. Han, "A Study of a Data Standardization and Cleaning Technique for a Facial Emotion Recognition System," In proceedings of the Symposium of the Korean Institute of communication and Information Science (KICS) Summer Conference, pp. 1193-1195, 2020.

[4] C. Rother, V. Kolmogorov, and A. Blake, “GrabCut interactive foreground extraction using iterated graph cuts,” ACM transactions on graphics (TOG) vol. 23, pp. 309-314, 2004.

[5] J. H. Kim, A. Poulose, and D. S. Han, " The extensive usage of the facial image threshing machine for facial emotion recognition performance,” Sensors, vol. 21, p. 2026, 2021.

MTF 기반 해상력 성능에 따른 보행자 검출 성능평가 방법에 관한 연구

김희강, 손준우, 김지연, 조태식, 한동석*

한국건설생활환경시험연구원, *경북대학교

heekangkim@kcl.re.kr, joonsohn@kcl.re.kr, jiyeon0311@kcl.re.kr, tscho@kcl.re.kr, *dshan@knu.ac.kr

A Study on the Pedestrian Detection Performance Evaluation Method According to MTF-based Resolution Performance

Heekang Kim, Joonwoo Sohn, Jiyeon Kim, Tae-Sik Cho, Dong Seog Han*

Korea Conformity Laboratories, *Kyungpook National Univ.

요 약

인공지능 기술이 고도화되고 경량화 되면서 CCTV 카메라의 기능이 단순한 관제와 움직임 인지를 넘어 객체 인식 및 행위 인식을 통해 CCTV 관제사들이 더욱 효율적으로 관제할 수 있도록 인공지능 기반의 지능형솔루션 또는 엣지형 카메라가 지능형CCTV 개발되어지고 있다. 이에 따라 지능형CCTV를 평가하는 방법으로 단순 특정 시나리에오 따른 인식률, 검출률에 대한 지표로만 평가가 되고 있으나, 인공지능 알고리즘은 영상 품질에 의해 많은 영향을 받고 있음에도 영상의 품질에 따른 성능을 검증하는 방법이 부재한 상황이다. 본 논문에서는 객체의 선명도를 평가할 수 있는 영상 품질 지표인 MTF 기반의 해상력 성능에 따른 보행자 검출 성능평가 방법을 제시하며, 지능형CCTV는 단순 알고리즘 평가 뿐 아닌, CCTV 카메라의 하드웨어 성능을 함께 평가해야함을 제시한다.

Ⅰ. 서 론

지능형CCTV는 기존에 단순 방범용, 교통관제용 단계의 용도가 아닌 스스로 인식하고 판단할 수 있는 솔루션으로써 다양한 분야에서 개발되어지고 사용되어지고 있다. 특히 방범을 위한 다양한 보행자 검출 기술이 발전되고 있다[1].

하지만 지능형CCTV 평가에 대해서는 단순 알고리즘 평가를 위해 학습DB와 평가DB를 구분하여 검출/분류/인식 성능에 대한 평가가 주로 이루어지고 있다. 하지만, CCTV 카메라 자체의 하드웨어 성능이 낮게 되면 원 소스(Original source) 데이터에 대한 품질이 낮아지므로 지능형CCTV의 알고리즘의 성능이 낮아질 수 있다.

또한, 국내에 민간부분에서는 많은 CCTV 카메라는 하드웨어 성능에 대한 검증 없이 무분별하게 설치되고 있어 저가의 저품질 CCTV가 많이 설치되고 있어 기존 설치된 CCTV 인프라에서 지능형솔루션이 탑재된다 하더라도 낮은 하드웨어의 성능 때문에 지능형CCTV의 높은 성능을 기대할 수 없다. 그러므로 지능형CCTV는 단순 알고리즘 평가에 그치지 않고 CCTV 카메라 자체의 하드웨어 성능도 함께 평가하여 복합적이고 정량적인 성능평가를 진행해야한다.

기존 연구들은 픽셀 크기(해상도)에 따른 객체 검출 성능 비교[2] 등으로 단순히 영상 내에서 객체의 픽셀 크기에 따른 검출 성능 정량적 평가를 하였으나, 객체의 위치, 카메라의 화각 등의 변수로 인해 객관적인 평가가 어렵다.

이에 따라, 본 논문에서는 지능형CCTV의 객관적이고 정량적인 성능평가를 위해 하드웨어의 실환경 성능 평가를 위해 야외 Slanted-Edge 차트를 개발하였으며, 해당 차트를 이용한 해상력 성능에 따른 지능형솔루션 성능을 평가하여 분석한다.

하드웨어 성능은 MTF(Modulation Transfer Function) 기반 해상력을 측정[3]하며, 검출하는 객체는 보행자이며, 사용 알고리즘은 YOLO v3를 사용하였다. 변수를 최소화하기 위해 원 영상에서 Blur 효과로 해상력과 보행자 검출 성능을 비교하여 정량적 상관관계를 제시한다.

Ⅱ. 본론

실험은 한국건설생활환경시험연구원 안전융합센터 부지에서 진행하였으며, 카메라는 옥상에 설치된 PTZ 카메라를 이용하였다. 카메라의 높이는 약 55m이며, 차트와 카메라 간 수평 거리는 약 90m 이다.



Fig. 1 부지 촬영 화면 및 차트 위치

동일한 DB에서 동일한 보행자의 움직임을 비교하기 위해 다음과 같이 실험 흐름도를 계획하였다.

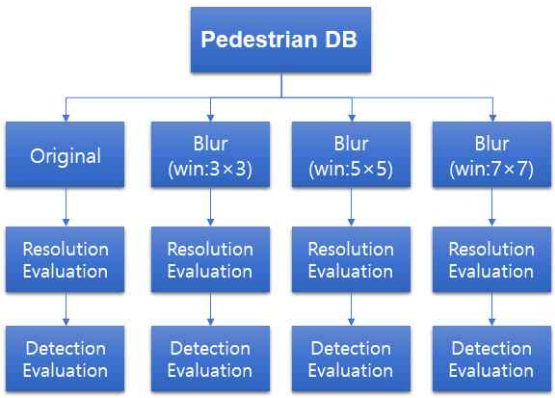


Fig. 2 실험 계획 흐름도

보행자 비디오 클립 DB(20초 이내)을 확보하여, 모든 프레임을 추출하고, 각 프레임 별 Blur 처리를 한다. Window size는 3×3, 5×5, 7×7으로 설정하고 Filter는 평균으로 설정한다.

Slanted-Edge MTF 해상력 평가는 흑백의 시그모이드 모양의 프로파일을 미분 후 푸리에 변환을 통해 MTF로 변환한다.

MTF 곡선은 아래 그림과 같으며, MTF 0.5에 위치할 때의 LW/PH 수치가 측정하고자 하는 해상력의 성능 지표이며, MTF50P 기준으로 측정하였다. 본 실험에서는 IMATEST 5.0 버전을 사용하여 측정하였다.

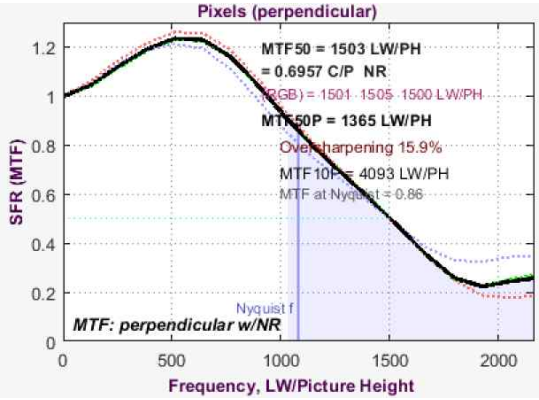


Fig. 3 MTF 곡선

동일 DB에서 원 영상 및 Blur 영상에 대한 해상력 평가를 통해 아래와 같은 해상력 결과를 확인할 수 있다.

DB명	해상력 (MTF50P)	원영상 대비 성능
원영상	1368 LW/PH	100 %
Blur 3×3	597 LW/PH	43.6 %
Blur 5×5	341 LW/PH	24.9 %
Blur 7×7	243 LW/PH	17.7 %

보행자 검출 알고리즘은 YOLO v3 모델을 활용하였으며[4], 본 모델은 YOLO 후속 버전인 YOLO v2의 아키텍처를 그대로 사용하였으나 아키텍처는 Darknet-53으로 변경하고, FPN처럼 다양한 크기의 해상도 Feature Map을 이용해 Bounding Box로 Output을 나타내었다.



Fig. 6 보행자 검출 박스 처리

DB명	전체 프레임 수	검출 프레임 수	검출률
원영상	350	298	85.1 %
Blur 3×3	350	284	81.1 %
Blur 5×5	350	255	72.8 %
Blur 7×7	350	227	64.8 %

보행자가 최초 등장하는 위치에서의 보행자 크기는 세로 60픽셀이며, 최종 등장하는 위치에서의 보행자 크기는 세로 80픽셀이다.

보행자와 카메라 간 직선거리는 약 75 m이며, 해상력을 측정했던 블러 효과를 통해 총 4개의 보행자 DB의 검출 성능을 평가한다.

CCTV카메라의 해상력이 낮아질수록 검출률은 낮아지는 상관관계를 확인할 수 있으며, 이는 지능형CCTV의 목적에 따라 거리별, 객체 크기별 검출이 필요할 때 하드웨어의 성능 기준치를 마련하는데 충분한 가능성을 확인할 수 있다.

Ⅲ. 결론

본 논문은 지능형CCTV의 정량적이고 복합적으로 평가하기 위해 MTF 기반 해상력 성능에 따른 보행자 검출 성능평가를 수행하였다.

해상력 평가를 위해 실환경 대형 Slanted-Edge 차트를 개발하였으며, 해상력의 변화를 주기 위해 획득한 DB를 블러처리하여 해상력 차이별 보행자 검출 성능을 확인하였다. 보행자 검출 알고리즘은 YOLO v3을 사용하여 비교하였으며, 해상력 저하에 따라 검출률이 저하되는 것을 확인할 수 있었다.

이에 따라 지능형CCTV 성능을 보장하기 위해 해상력 성능이 검증된 CCTV 카메라를 채택하고, 알고리즘 평가를 위해 이런 실험 방법으로 지능형CCTV의 평가를 할 수 있을 것으로 판단된다.

본 논문의 연구를 바탕으로 지능형CCTV에 대한 성능 기준치를 단순 알고리즘의 성능 뿐 아니라 카메라 자체의 하드웨어 부분의 성능에 대한 기준을 마련하여 하드웨어와 소프트웨어 성능을 모두 고려한 복합 성능에 대한 평가를 할 수 있으며, 국내에 설치하는 CCTV의 성능 기준치를 제시할 수 있으며, 향후 연구에서는 현장에서 요구하는 객체 크기, 인식 거리에 따른 MTF 기반 해상력 성능을 비교하여 지능형CCTV를 위한 해상력 기준을 제시할 것이다.

ACKNOWLEDGMENT

이 논문은 산업통상자원부 국가기술표준원에서 시행한 산업표준개발사업의 결과를 활용한 연구임

참 고 문 헌

[1] Antonio, J. A., & Romero, M. (2018, December). Pedestrians' Detection Methods in Video Images: A Literature Review. In 2018 International Conference on Computational Science and Computational Intelligence (CSCI) (pp. 354-360). IEEE.

[2] Shermeyer, J., & Van Etten, A. (2019). The effects of super-resolution on object detection performance in satellite imagery. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (pp. 0-0).

[3] ISO 12233:2017 - Photography – Electronic still picture imaging

[4] Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.

클래스 정보를 고려한 샴 네트워크 기반의 추적 방법

이준구, 오지용, 남승우

한국전자통신연구원

leejg01679@etri.re.kr, jiyongoh@etri.re.kr, swnam@etri.re.kr

A Tracking Method Based on Siamese Network with Class Information

Joon-Goo Lee, Jiyong Oh, Seung-Woo Nam

Electronics and Telecommunications Research Institute (ETRI)

요 약

본 논문은 휴먼 추종을 위한 샴(Siamese) 네트워크 기반 추적기에 객체의 클래스 정보를 함께 고려하는 방법을 제안한다. 샴 네트워크 기반의 추적기는 클래스의 구분 없이 선택된 대상으로부터 추출된 특징만을 이용하기 때문에 추적 중인 클래스의 객체와 유사한 특징을 갖는 다른 클래스의 객체를 혼동하기 쉽다. 따라서 객체의 클래스 정보를 추출할 수 있는 검출기를 결합하여 추적 대상의 초기화를 수행하고 추적 중인 객체가 동일한 클래스인지 주기적으로 비교한다. 실험에서 제안하는 방법은 기존의 방법보다 유사한 클래스가 존재하는 경우에 대해 더 강인하게 작동하였다.

Ⅰ. 서 론

영상기반 추적 기술은 생활/방범, 교통 감시, 물류 로봇 등의 다양한 산업 분야에서 응용되어 필요성이 점차 증가하고 있으며, 특히 물류 로봇 분야에서는 휴먼 추종과 관련된 기술 수요가 증가하고 있다. 최근 딥러닝 기술의 사용이 보편화 되면서 추적 기술에도 다양하게 접목되고 있으며, 샴(Siamese) 네트워크 기반 추적 기술은 높은 성능과 빠른 처리 속도로 인해 효율이 매우 우수하다[1]. Zhang은 샴 네트워크 기반의 추적기에 기존의 검출 목적인 백본 네트워크 대신 추적에 적합한 백본 네트워크를 제안하였으며, 변형 가능한 합성곱(deformable convolution)을 적용하여 객체의 경계 상자를 더 정확하게 검출하였다[2-3]. 하지만 샴 네트워크 기반의 추적기는 실시간으로 입력되는 영상에 대한 초기화 방법이 적절하지 않고, 사용자가 초기에 설정한 대상으로부터 추출된 특징만을 고려하기 때문에 추적 중 대상이 혼동되는 경우가 발생한다.

본 논문에서는 객체의 클래스 정보를 추출할 수 있는 검출기를 샴 네트워크 기반의 추적기와 결합하여 추적 대상의 초기화를 수행하고, 추적 중인 객체가 동일한 클래스인지 추가적으로 비교하여 추적 성공률을 높이는 방법에 대하여 제안한다.

Ⅱ. 본론

제안하는 방법은 그림 1과 같이 기존의 추적기에 검출기를 결합한 형태로 구성된다. 기존의 추적기는 이미지를 입력받아 추적하고자 하는 대상과 검색 이미지를 생성하고, 동일한 가중치(weights)를 공유하는 CNN(Convolution Neural Network)을 통해 특징을 추출하여 경계 상자 예측 및 특징 분류를 수행한다. 결합된 검출기는 입력 이미지에 대해 검출을 수행하고, 검출 결과로 출력된 객체의 경계 상자와 클래스 정보를 추적 결과와 함께 고려하여 초기화 및 추적을 수행한다. 여기서 추적기는 Ocean 추적기를 사용하였으며[3], 검출기는 Yolov4-tiny를 사람만을 대상으로 검출하는 1-클래스 검출기로 변형하여 사용하였다[4].

먼저 기존의 추적기 초기화 방법은 저장된 영상의 첫 프레임에 대하여 마우스 드래그 또는 미리 설정된 좌표를 이용해 추적 대상의 경계 상자를 입력하였다. 하지만 실시간 영상에서는 초기 프레임의 경계 상자를 사전에 추출할 수 없으며, 마우스 드래그 조작시 객체의 움직임으로 인해 정확한 경계 상자를 설정하기 어렵다. 따라서 그림 2와 같이 검출기를 이용해 실시간으로 입력되는 프레임의 객체들을 추출하여 추적 대상의 상자를 선택하거나 화면의 특정 위치에 일정 시간 이상 존재하는 대상의 상자를 자

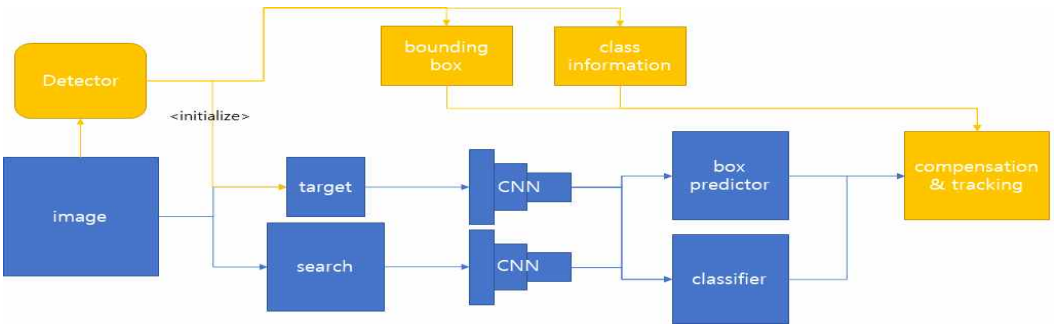


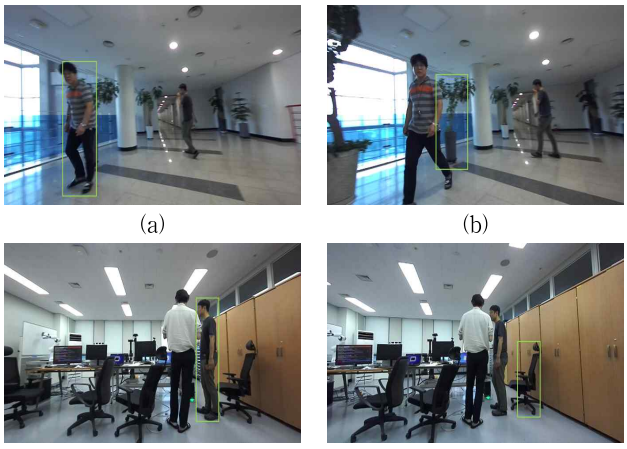
그림 1. 제안 방법의 구조

동으로 선택하여 추적 대상을 초기화한다. 여기서 녹색 상자는 검출기가 검출한 객체의 경계 상자이고, person은 클래스명, (354, 235)는 경계 상자의 중심이다. 1.64는 사전에 설정한 추적 시작 점과의 유클리드 거리로 임계값 10 미만으로 가까워지면 추적을 수행하도록 설계하였다.



그림 2. 초기화 방법

제안하는 방법의 성능 비교를 위해 1분 단위의 영상 5개를 선정하여 테스트를 수행하였으며, 기존의 방법은 5개 영상 중 2가지 영상에 대하여 추적에 실패하였다. 기존의 샴 네트워크 기반의 추적기는 클래스 정보를 고려하지 않고 초기 설정된 대상의 특징만을 사용하기 때문에 형태나 컬러가 비슷한 다른 클래스와 혼동되기 쉽다. 그림 3은 이러한 추적 오류의 예를 보여주고 있으며, 여기서 녹색 경계 상자는 추적기가 정상적으로 추적 중인 대상을 나타낸다. (a), (c)는 초기 설정한 사람을 대상으로 추적하고 있었으나 카메라와 대상이 움직이면서 (b), (d)와 같이 유사한 형태 또는 유사한 컬러의 다른 클래스의 객체로 추적 대상이 변경되었으며, 추적 신뢰도 역시 90% 이상으로 잘못된 대상을 계속 정상적으로 추적하였다.

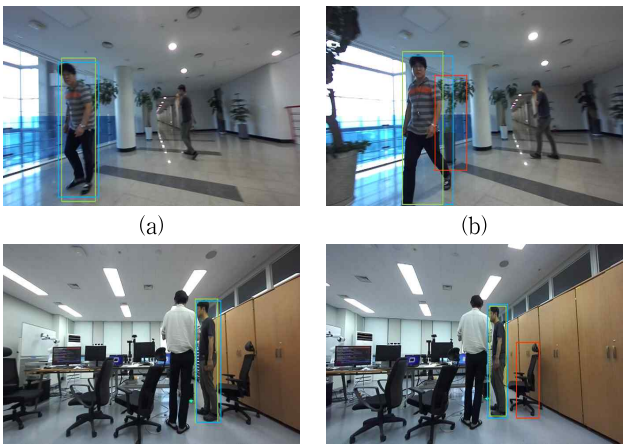


(a) (b)

(c) (d)

그림 3. 추적 오류 예

제안하는 방법은 초기에 검출기로부터 설정된 클래스 정보와 현재 프레임의 검출기로부터 검출된 경계 상자를 함께 고려하여 문제점을 보완하였으며, 테스트 영상 모두 추적에 성공하였다. 그림 4는 기존의 방법으로 추적에 실패한 영상에 대하여 제안하는 방법을 수행한 결과를 보여준다. (a), (c)의 녹색 경계 상자는 추적기를 통해 추적 중인 대상이며, 파란색 경계 상자는 검출기를 통해 검출된 객체의 경계 상자들 중에 IoU가 가장 가까운 객체이다. 초기에 추적 대상으로 설정된 클래스와 현재 검출되어 매칭된 경계 상자의 클래스가 같기 때문에 정상으로 판단하여 추적을 진행한다. (b), (d)의 빨간색 경계 상자는 추적기를 통해 추적 중인 대상이고, 녹색 경계 상자는 제안하는 방법을 통해 보정된 결과이다. 초기에 추적 대상으로 설정된 클래스와 현재 검출되어 매칭된 경계 상자의 클래스가 다르기 때문에 비정상으로 판단하고, 이전 프레임에서 정상으로 추적하던 대상과 현재 프레임에서 검출된 객체들 중에 IoU가 가장 높은 경계 상자를 선택하여 재 초기화 후 다시 추적을 수행하였다.



(a) (b)

(c) (d)

그림 4. 제안 방법 결과 예

Ⅲ. 결론

본 논문에서는 기존의 샴 네트워크 기반의 추적기와 검출기를 결합하여 다른 클래스와 혼동되는 상황에서도 강인한 추적 방법을 제안하였다. 또한 추적기의 초기화에도 검출기를 사용하여 실
시간으로 입력되는 영상에 초기화가 가능하도록 하였다. 실험에서 제안하는 방법은 기존의 방법보다 우수한 성능을 나타내었다.

추후 RGB만 사용하는 기존의 추적 방법에 Depth 정보를 함께 고려하여 동일한 클래스 내에서 혼동되는 경우에도 강인하게 추적할 수 있는 방법을 연구할 예정이다.

ACKNOWLEDGMENT

본 연구 논문은 한국전자통신연구원 연구운영지원사업의 일환으로 수행되었음. [21ZD1130, 지능제어기반 스마트기계 및 로봇 기술 개발]

참 고 문 헌

[1] Berinetto L., et al., “Fully-convolutional siamese networks for object tracking,” European conference on computer vision, pp. 850-865, 2016.

[2] Z. Zhang, et al., “Deeper and wider Siameses networks for real-time visual tracking,” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4591-4600, 2019.

[3] Z. Zhang, et al., “Ocean: Object-aware Anchor-free Tracking,” Proceedings of 16th European Conference on Computer Vision, pp. 771-787, 2020.

[4] Bochkovskiy A., et al., “Yolov4:Optimal speed and accuracy of object detection,” arXiv preprint arXiv:2004. 10934. 2020.

딥러닝 기반의 수정조기경보지수(Modified Early Warning Score) 예측 방법에 관한 연구

진용식 [1*], 김범휘 [1], 김규형 [1], 조성현 [1], 신우상 [2]
[1] 한국전자통신연구원, [2] 한국생산기술연구원, *교신 저자

{yongsik, bumhwi.kim, jaykim, seonghyeonjo}@etri.re.kr, welldone@kitech.re.kr

A Study on deep learning-based methods for MEWS (Modified Early Warning Score) prediction

Yongsik Jin[1*], Bumhwi Kim[1], Kyu hyung Kim[1], Seonnghyeon Jo[1], Woo Sang Shin[2]
[1]Electronics and Telecommunications Research Institute (ETRI), [2]Korea Institute of Industrial Technology (KITECH), *Corresponding author

요 약

본 논문은 수정조기경보지수(Modified Early Warning Score)를 기반으로 환자의 위험 상태를 예측하기 위한 딥러닝 기반의 알고리즘을 제안한다. 효율적인 예측 방법을 개발하기 위하여 Convolutional neural networks, Long short-term memory 기반의 방법들이 주어진 데이터에 대해 성능이 평가된다. 특히 Long short-term memory 의 성능을 개선하기 위하여 Attention 알고리즘과 입력 데이터를 변형하는 방법들이 제안된다. 본 논문의 실험 결과는 수정조기경보지수(Modified Early Warning Score)를 기반으로 환자의 위험 상태를 예측하기 위한 모델들의 효율성을 보여주고 추후의 연구 방향을 제시한다.

Ⅰ. 서 론

수정조기경보지수(MEWS: Modified Early Warning Score)는 환자의 위험 징후를 효율적으로 나타낼 수 있는 지표 중 하나로 위험 환자들의 모니터링 방법에 사용되고 있다 [1]. MEWS 는 환자의 위험 추이를 예측하기 위하여 혈압, 맥박, 호흡률, 체온, 환자의 의식 상태를 통계적으로 분석하여 개발된 점수로서 주로 MEWS 를 4 개의 위험 단계 (안정, 주의, 경고, 위험)로 나누어 환자의 건강 상태를 적절하게 관리한다. 미리 환자의 상태를 예측한다는 점에서 환자의 생존률을 높이고 의료자원을 효율적으로 운영할 있다는 장점을 가지고 있다.

환자 위험 상태 예측과 관련하여 최근 딥러닝 기반의 알고리즘들이 많은 관심을 받고 있다. 특히 뷰노에서 연구되어진 딥러닝 기반의 조기 경보 시스템은 임상 시험을 통해 효율성을 증명하였고 실용화를 위해 실제 의료 현장에서 시험을 진행 중이다.

본 논문에서는 환자의 MEWS 를 기반으로 환자의 위험 상태를 한 시간 미리 예측하는 딥러닝 기반의 알고리즘을 제안한다.

Ⅱ. 본론

본론에서는 1 시간 후의 MEWS 예측에 사용된 방법들이 소개된다. 시계열 데이터 처리를 위한 convolutional neural network (CNN)와 Long short-term memory (LSTM), Attention 알고리즘, onehot encoding, MEWS 환산식을 활용한 입력 데이터 처리 기법들이 사용된다. MEWS 계산 식은 아래의 표와 같이 주어진다.

표 1. MEWS 계산 테이블(SBP: Systolic Blood Pressure, HR: Heart Rate, RR: Respiratory rate, BT: Body Temperature, AVPU: Alert, Voice, Pain, Unresponsive)

Socore	0	1	2	3
SBP (mmHG)	101-199	81-100	70-80, ≥ 200	<70
HR (bpm)	51-100	40-50, 101-110	<40, 111-129	≥ 130
RR (bpm)	9-14	15-20	21-29, <9	≥ 30
BT (℃)	35-38.4	-	<35, ≥ 38.5	-
AVPU score	Alert	Reacting to Voice	Reacting to Pain	Unresponsive

Ⅱ-1. 실험 데이터

본 논문의 실험에서는 영남대학교 병원에서 후향적 임상 시험으로 획득되어진 데이터를 사용한다. 환자의 MEWS 를 계산하기 위해 1 시간 마다 수축기 혈압, 맥박, 체온, 호흡률을 측정하였다. 하나의 데이터 단위는 총 9 시간동안 측정한 9 개의 수축기 혈압, 맥박, 체온, 호흡률로 구성되며 딥러닝 모델의 입력으로는 앞의 8 시간 동안

의 8 개의 수축기 혈압, 맥박, 체온, 호흡률 값들이 사용된다. 딥러닝 모델의 목표 출력 값으로는 마지막 시점인 9 번째 시점의 MEWS 으로부터 계산된 환자 위험 상태가 주어진다. MEWS 값에 따라 환자 위험 상태는 다음과 같이 주어진다.

표 2. MEWS 점수별 위험 단계 구분 기준

위험 단계	점수 분포	위험 단계
0	MEWS≤ 1	안정
1	2≤ MEWS≤ 4	주의
2	5≤ MEWS≤ 6	경고
3	7≤ MEWS	위험

시험에 사용되어진 모델들의 평가를 위해 학습 데이터 30000 개, 검증 데이터 15000 개, 테스트 데이터 15000 개를 하나의 학습, 검증, 테스트 셋으로 구성하였다. 총 5 개의 학습, 검증, 테스트 셋에 대하여 학습 결과를 검증한다.

Ⅱ-2. 학습 방법 및 모델

딥러닝 모델들은 Pytorch Python 라이브러리를 사용하여 구현되었고 손실함수는 Cross entropy, 최적화를 위해 Adam optimizer 를 사용하였다. 모델 구조에 대한 설명을 다음과 같이 간략하게 요약하였다.

- CNN 구조:

Layer 1
(cnn_1): Conv1d(4, 8, kernel_size=(3,), stride=(1,), padding=(1,)), (relu_1): ReLU()
Layer 2
(cnn_2): Conv1d(8, 8, kernel_size=(3,), stride=(1,), padding=(1,)), (relu_2): ReLU()
Layer 3
(cnn_3): Conv1d(8, 8, kernel_size=(3,), stride=(1,), padding=(1,)), (relu_3): ReLU()
Layer 4 (dense): Linear(in_features=64, out_features=4, bias=True)

- LSTM 구조:

Layer 1
(lstm): LSTM(4, 120, num_layers=3, batch_first=True)
Layer 2
(dense): Linear(in_features=120, out_features=12, bias=True)

- LSTM 구조 (Onehot encdoing 을 사용하는 경우):

Layer 1
(lstm_1): LSTM(4, 100, num_layers=2, batch_first=True)
(lstm_2): LSTM(4, 100, num_layers=2, batch_first=True)
(lstm_3): LSTM(4, 100, num_layers=2, batch_first=True)
(lstm_4): LSTM(4, 100, num_layers=2, batch_first=True)
Layer 2
(dense): Linear(in_features=100, out_features=4, bias=True)

* lstm_1, lstm_2, lstm_3, lstm_4 의 출력 값들을 더한 후 dense layer 에 입력함

- Input attention:

아래의 수식에서 계산된 a_t^k를 입력 데이터에 곱하여 새로운 입력 데이터를 계산한다 [3].

$$e_t^k = V_e^T \tanh\left(W_e\left[h_{t-1}; s_{t-1}\right] + U_e x^k\right), a_t^k = \frac{\exp(e_t^k)}{\sum_{i=1}^{n}\exp(e_t^i)}$$

- EWS encoding:

입력 데이터를 MEWS 환산 테이블을 이용하여 0~3 사이의 정수값으로 변환한다. 예) 75mmHg 값을 가지는 SBP 는 2 점으로 변환된다.

* 최종 입력으로는 3 으로 나누어 0~1 사이의 값으로 변환되어 사용되어진다.

- Onehot encoding:

EWS encoding 에서 변환된 점수 값들을 onehot encoding 된 벡터 형태의 입력으로 변환한다.

Ⅲ. 실험 결과

MEWS 예측에 효율적인 모델을 찾기 위해 본론에서 설명된 모델에 대하여 5 개의 실험 데이터 군에 대해서 실험을 수행하였다. 실험 결과는 다음과 같이 주어진다.

표 3. 5개의 데이터셋에 대한 예측 정확도 (단위: %)

데이터셋/모델	1	2	3	4	5
EWS encoding+ CNN	76.3	71.0	76.7	76.2	76.9
EWS encoding+ LSTM	77.6	64.1	77.2	75.9	77.4
EWS encoding + Attention+ LSTM	76.3	68.2	76.2	75.8	76.3
Onehot encoding + LSTM	90.5	87.1	90.6	90.9	91.0

Ⅲ. 결론

본논문에서는 효율적인 MEWS 예측 알고리즘을 개발하기 위하여 다양한 모델들의 성능을 비교하였다. 실험 결과에서 입력 데이터들을 onehot encoding 하여 LSTM 을 기반으로 예측 알고리즘을 구성하는 것이 가장 효율적임을 볼 수 있었다.

ACKNOWLEDGMENT

This work was supported by Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korean government. [21ZD1140, Development of ICT Convergence Technology for Daegu-Gyeongbuk Regional Industry].

참 고 문 헌

[1] Subbe, Christian P., et al"Validation of a modified Early Warning Score in medical admissions,"QJM, pp. 521-526, Oct. 2001.

[2] Kwon, Joon-myoung, et al. "An algorithm based on deep learning for predicting in-hospital cardiac arrest," Journal of the American Heart Association, Jun. 2018.

[3] Qin, Yao, et al. "A dual-stage attention-based recurrent neural network for time series prediction,"arXiv preprint arXiv:1704.02971, 2017.

중증 환자 욕창 예방을 위한 유연인쇄압력센서 기반 자세 추정 기법 연구

박찬은*, 문애경
한국전자통신연구원

*chaneun@etri.re.kr, akmoon@etri.re.kr

Patient pose estimation for bedsore prevention by using flexible pressure sensors

Chan-eun Park, Aekyung Moon

요 약

본 논문은 중증 환자의 욕창 예방을 위한 자세 추정 알고리즘을 제안한다. 유연인쇄기법으로 제작된 28 개의 압력 센서를 배치하여 압력 매트를 구상하였고, 이를 병상 침대에 배치하여 중증 환자의 신체에 가해지는 압력 정보를 수집하였다. 1Hz 주기로 전달되는 압력 데이터를 환자 자세가 입력되는 주기인 2 시간 간격에 맞추어 분할 후 최종 자세 정보를 라벨링 해주었다. 환자 특성상 자세 기록 시점으로부터 5 분 이내에는 자세 변화가 없다는 가정을 할 수 있었고, 이를 통해 총 32400 개의 자세 학습 데이터를 구축하였다. Random forest 알고리즘을 통해 자세 분류기를 구현하였고, 이를 전체 압력 데이터에 적용 후 후처리를 통해 환자의 자세 변화를 유추할 수 있다.

Ⅰ. 서 론

욕창은 신체의 특정 부위에 지속적으로 가해지는 압력 등의 이유로 발생하는 피부 및 기저 조직의 국부적인 손상으로, 장시간 누워있는 중증 환자의 등, 천골, 엉덩이와 같은 돌출부에 발생할 확률이 높다고 알려 져 있다 [1]. 욕창은 다른 질병의 원인으로 발전 할 가능성이 높으므로 무엇보다 사전 예방이 중요하다. 현재 욕창 예방은 간병인이나 의료인이 환자의 자세를 주기적으로 변경시켜주는 방식으로, 많은 인력과 비용을 발생시킨다.

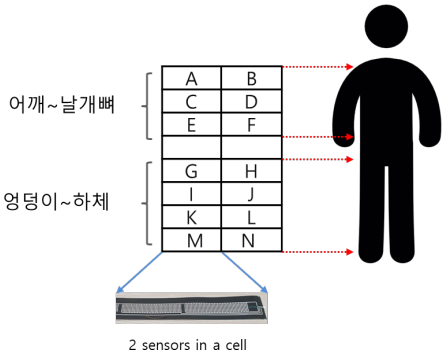
욕창 예방을 위한 연구의 일환으로, 압력 매트리스 등을 활용한 자세 추정 및 제어 연구가 진행되고 있다. 압력 센서를 활용하여 추출되는 신체의 압력에 기반한 자세 추정 연구가 진행되고 있으며 [2,3], 보급성을 위하여 저비용의 센서를 활용한 연구가 많은 관심을 받고 있다.

유연인쇄기법으로 제작되는 유연압력센서의 경우 생산 비용 절감 및 소재의 유연성을 충족시키기 때문에 신체 압력측정에 용이하다. 본 연구에서 사용되는 압력 센서는 8Hz 주기로 압력 정보를 전송한다. 이 중 자세 판단을 위해서는 1Hz 주기로 측정되는 압력 정보만 활용하였다. 최종적으로 28 개의 압력센서를 통해 5 명의 환자로부터 신체 압력 데이터를 수집하였다. 환자의 자세는 2 시간 간격으로 간병인에 의해 수기로 측정되었다. 환자의 자세와 압력이 모두 기록된 구간의 데이터와 다중 분류 알고리즘을 활용하여 자세 판별 알고리즘을 구성한다. 다중 분류 알고리즘을 자세가 측정되지 않는 구간의 압력 데이터에 적용하여 자세를 추정하고, 후처리를 통해 환자의 자세를 추정하는 알고리즘을 제시한다.

Ⅱ. 유연압력센서 기반 데이터 수집

본 논문에서는 유연압력소자로 구성된 압력센서매트를 구성, 이를 병상의 침대에 배치하여 누워있는 환자의 신체에 가해지는 압력을 측정한다. 2 개의 압력 센서로 구성된 압력 셀을 7 행 2 열의 구조로 배치하여, 총 28 개의 압력 센서로 구성된 압력 매트를 구성하였다. 상/하체에 각각 3 개와 4 개의 행을 배치하여 아래 그림 1 과 같은 압력 센서 매트를 병상의 침대에 배치한다. 압력 센서는 8Hz 주기로 압력을 측정하며, 3 초에 한번씩 데이터를 전송한다.

판별하고자 하는 자세는 Supine, Left lateral, Right lateral, Head elevation 으로 총 네 가지이다. 병실의 중증 환자가 사용하는 침대에 압력 매트를 설치하여 총 5 명의 환자를 대상으로 데이터를 수집하였다. 환자의 자세는 2 시간 간격으로 측정되었다. 6 일간 수집한 데이터를 자세 측정 시점 간격으로 분리하였고, 데이터 누락이 발생한 구간을 제외하고 총 360 개 구간에서의 압력 데이터를 획득하였다.



[그림 1. 유연압력센서로 구성한 압력 매트]

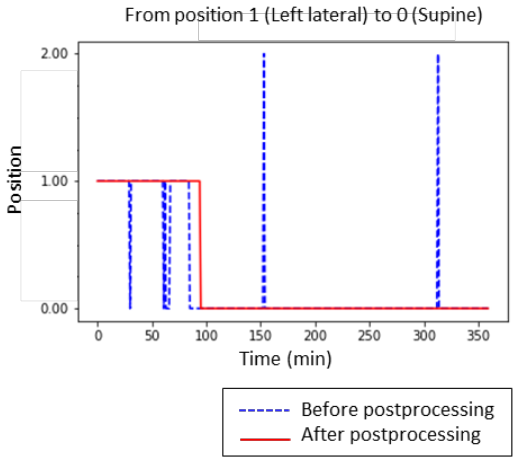
Ⅲ. 압력데이터 기반 자세 추정

본 연구에서는 압력/자세 데이터가 수집되는 과정에서 환자의 자세가 변경되는 시점은 알 수 없는 문제가 존재한다. 따라서 자세 분류 알고리즘을 위하여, 환자의 자세가 측정되는 전후 5 분간은 자세 변동이 없었을 것이라고 가정하였다. 즉, 자세가 측정되는 시점 (각 데이터셋의 마지막 지점)으로부터 5 분 이내의 데이터에 대하여 측정된 압력 데이터는 모두 측정된 자세의 라벨을 부여하였다.

자세 판별을 위한 압력 신호는 1 분간 측정되는 압력의 평균값을 사용하였다. 측정된 데이터 구간에서 여러 개의 학습 데이터를 추출하기 위하여 측정 구간의 마지막 지점으로부터 5 분 이내 데이터 중 랜덤하게 총 100 개의 압력 신호를 추출한다. 결측 치를 제외하고 총 32400 개의 압력-자세 데이터셋을 구성하였고, 4 개의 자세 라벨링에 대하여 약 4:2:2:1 의 비율을 구성한다. 환자 자세 변경 패턴을 고려하였을 때 실제 움직임과 유사한 비율을 구성하므로 별도의 augmentation 없이 학습에 사용하였다.

자세 분류를 위하여 support vector machine, decision tree, artificial neural network, convolutional neural network, random forest 기법을 고려였고, 가장 좋은 성능을 보이는 random forest 알고리즘 (N_features = 100, max_features = 10)을 사용하였다. Test data 에 대하여 accuracy score = 0.9971, f1 score = 0.9944 의 성능을 확인하였다. 해당 분류기를 자세 라벨링이 되지 않은 구간에서의 압력 데이터에 적용하여 자세를 추정하였다. 라벨링이 되어 있지 않은 구간의 reference data 가 없는 관계로, 각 데이터 구간에서 시작 자세(이전 데이터 구간의 마지막 자세)와 마지막 자세 사이에 최소한의 변화만 발생했을 것이라고 가정하였다.

자세 분류기를 통해 예측한 자세 변화는 아래 그림 2 의 형태를 보인다. 압력 데이터를 1 분간 저장 후 평균값으로 검출한 예상 자세가 파란 점선으로 표시되어 있다. 국소적인 시간에 다른 자세로 예측되는 경우는 잘못된 자세 예측으로 의심되는 경우이다. 이를 보정하기 위한 후처리 기법으로 moving horizon window 를 활용하였다. 현재 시점으로부터 과거 L 개의 자세 검출을 저장하는 window 를 구성하여 window 내부에서 가장 많이 저장된 자세를 현재 자세로 반영하였다. 이 후처리 방법은 자세 변화 시점 검출에 delay 를 발생시킬 수도 있지만 튀는 값에 대한 보정이 비교적 간편하다. 또한 환자의 자세 이동이 빈번하게 발생하지 않는 데이터의 특성에 적합하다.



[그림 2. 자세 추정 및 후처리 결과]

Ⅳ. 분석 및 결론

본 연구는 적은 개수의 센서를 통해 수집한 압력 데이터를 토대로 환자의 자세를 판단할 수 있다. 자세 라벨링의 경우 개인정보 문제상 수기로 기록되므로 충분한 데이터 확보가 어렵다. 본 연구는 라벨링이 되어 있지 않은 구간의 자세 추정에 활용될 수 있다는 점에서 의미가 있다. 하지만 자세 변경 시점에 대한 데이터가 없는 상황에서 본 알고리즘의 정확도를 판단하기 어려운 문제가 있다. 데이터의 특성을 고려하여 semi-supervised learning 등의 알고리즘을 통해 추가 실험 예정에 있다. 추후 압력센서매트를 통한 실험을 통해 자세 변경 시점 검출에 대한 성능 평가를 진행 예정이다.

ACKNOWLEDGMENT

This work is supported by the Korea Innovation Foundation (INNOPOLIS) grant funded by the Korea government (MSIT) (2020-DD-UP-0278)

참 고 문 헌

[1] Tsuji S, Ichioka S, Sekiya N, Nakatsuka T. Analysis of ischemia-reperfusion injury in a microcirculatory model of pressure ulcers. Wound repair and regeneration. 2005 Mar;13(2):209-15.

[2] 김진규, 심명훈, 배은경, 문영진, 최재순, “체압 분포 영상 초고해상도 보정을 이용한 욕창 예방을 위한 환자 자세 추정”, 제어로봇시스템학회논문지 27(5) pp. 342-348, May. 2021

[3] M. B. Pouyan, J. Birjandtalab, M. Nourani, and M. D. M. Pompeo, “Automatic limb identification and sleeping parameters assessment for pressure ulcer prevention,” Computers in Biology and Medicine, vol. 75, pp. 98-108, 2016.

조제 자동화 시스템 적용을 위한 AI 기반 의약품 비전 검수용 데이터 확장 및 학습 방법

김휘강 [1*], 이상범 [1], 김광용 [1], 김규형 [1], 권혁주 [2], 이성학 [2]
[1] 한국전자통신연구원, [2] 경북대학교

{hwigangkim*, sblee230, kimky, jaykim}@etri.re.kr, {olin1223, shak2}@ee.knu.ac.kr

Data augmentation and learning method for AI-based vision inspection of automation tablet dispensing and packaging system

Hwi-Gang Kim[1], Sangbeom Lee[1], Kwang Yong Kim[1], Kyu Hyung Kim[1],
Hyuk-Ju Kwon[2], Sung-Hak Lee[2]

[1]Electronics and Telecommunications Research Institute,
[2]School of Electronics Engineering, Kyungpook National University

요 약

본 논문은 딥러닝 기반의 의약품 비전 검수 시스템 개발을 위한 데이터 확장 및 학습 방법과 관련한 건으로, 전자동 약품 조제 시스템의 사용에 있어 조제 완료된 포장과 처방의 일치 여부를 확인하기 위한 AI 기반의 비전 검수 모듈 개발 및 검수 모듈을 적용한 조제 자동화 시스템의 개발을 위한 것이다. 이와 관련하여 본 논문에서는 조제 자동화 시스템 적용을 위한 학습용 영상 데이터 확장 방법 및 데이터 확장 시 알약 정보 레이블링 방법, 확장한 데이터를 활용하여 실제 조제 환경에 적합한 데이터 학습 방법을 제안하였으며, 제안한 방법들을 통하여 학습을 수행한 뒤 대표적인 객체 검출 딥러닝 모델인 YOLOv3 를 사용하여 그 성능을 확인하였다.

Ⅰ. 서 론

정제 등 고형 의약품의 조제와 관련하여 약국 및 병원 등에서는 조제 업무의 자동화를 위한 전자동 정제 분배 및 포장 시스템을 사용하고 있다. 관련 의료진들은 이를 이용하여 처방에 따라 약품을 자동으로 분류하고 포장한 뒤 환자에게 전달하게 된다. 이 때, 만일의 경우 발생할 수 있는 오조제를 방지하기 위해 환자에게 전달 전 관련 의료진들에 의한 약품 검수가 반드시 수행되어야 한다. 다만 환자가 많거나 조제에 사용되는 취급 약품이 많은 대형병원 또는 약품의 장기 복용이 필요한 만성질환자를 대상으로 한 조제의 경우, 의료진들에 의한 전량 직접 검수가 불가능한 상황이며 이에 따른 시간 및 인력 낭비 역시 크다는 문제가 있다.

이와 관련하여 최근 포장이 완료된 약품들을 대상으로 한 검수 방법에 대한 연구와 더불어 포장 영상을 촬영한 뒤 이를 바탕으로 한 비전 기반의 자동 검수 시스템의 개발이 활발히 진행되고 있다. 그러나 비전 기반의 약품 검수를 위하여, 포장 내의 약품 검출 및 인식에 기존의 영상처리 기술들을 활용할 경우, 비슷한 크기, 모양, 색상 등을 가진 알약들을 각각 검출하기에는 한계가 있으며, 포장이 완료된 포장지 내 여러 정의 알약은 매우 다양한 형태로 위치(이동, 회전, 기울어짐, 접함, 겹침 등)하므로 이를 정확히 인식하는 것 또한 불가능하다 [1].

최근 아주 많은 양의 영상 데이터를 학습하여 영상 내 객체들을 검출하고 인식하는 인공지능 기술 기반 딥러닝 모델들이 각광받고 있으며, 이러한 딥러닝 기술을 통하여 위 언급한 기존의 영상처리 기술들을 활용한 약품 검수 방법의 한계를 극복할 수 있을 것으로 보인다. 다만 현재 딥러닝 기술의 사용을 위해서는 학습을 위한 아주 많은 양의 데이터 획득(영상 촬영)과 획득 데이터 내 알약들에 대한 정보(이름 및 위치)들의 라벨을 붙여주는 레이블링 작업이 필요한데, 이와 관련하여 조제 관련 기업 및 약국, 병원 등에서 취급하는 수많은 종의 알약들을 실제 조제 환경처럼 처방된 알약의 조합으로 모두 표현할 수 없을 뿐 아니라, 학습 데이터 획득을 위하여 수많은 알약들을 모두 포장하는 것도 불가능하다. 설령 포장이 가능하다 해도 포장 내 각각의 알약들을 임의의 다양한 형태로 위치하게 하여 촬영하는 것 역시 불가능하며, 최종적으로 각 영상 내 객체(알약)들의 정보를 일일이 레이블링 하는 것 또한 불가능하다.

이에 따라 본 논문에서는 약품 조제 자동화를 위한 AI 기반의 의약품 비전 검수 기술 개발에 있어서, 위 기술한 어려움들을 극복하기 위한 데이터 확보 (확장) 및 학습 방법을 제안하고자 한다.

Ⅱ. 본 론

우선 본 방법을 적용할 약품 비전 검수 장비 내 촬영 환경(배경 및 조명 등)은 항상 동일하므로 단일 알약을 각각 촬영한 경우 그림 1 (a)와 같고, 처방에 포함된 알약이 6 개일 때, 포장 후를 가정하여 모두를 포함한 뒤 촬영한 영상은 그림 1 (b)와 같이 약품 6 알이 다양한 형태로 위치(이동, 회전, 접함 등)하게 된다. 이 때 실제 조제 환경 및 상황 등을 고려해보면, 그림 1 (a)와 같이 포장 전 한 알의 알약을 각각 촬영하여 데이터베이스화 하여 이를 대상으로 학습을 수행한 뒤, 그림 1 (b)와 같이 한 영상에 다수의 알약이 다양한 형태로 위치한 포장 후 영상을 대상으로 알약의 위치를 검출하고 약의 종류를 인식하는 방법으로 검수를 진행하여야 한다.



(a) (b)

그림 1. 알약 영상의 예 ((a) 포장 전 단일 알약 영상, (b) 포장 후 복합 알약 영상)

일반적으로 딥러닝 기술을 활용하기 위해서는 아주 많은 양의 학습 데이터가 필요하며, 학습에는 최종 대상 영상과 유사한 형태의 영상이 많이 사용될수록 학습 모델의 성능이 좋아지게 되는데, 실제 조제 환경에서는 그림 1 (a)와 같이 하나의 알약 정도를 확보하여 촬영할 수 있을 뿐, 여러 개의 알약을 한꺼번에 임의의 다양한 형태로 위치시켜 촬영하는 것이 불가능하다. 또한 이 때 각각의 알약에 대한 정보를 직접 레이블링 하는 것 역시 불가능하다.

이런 경우 딥러닝 분야에서는 학습 데이터를 충분하게 확보하기 위하여 데이터를 확장하여 학습에 사용하는데 [2], 대표적인 데이터 확장 방법으로 영상의 환경 변경 (해상도, 조명 등) 또는 객체 변화 (색, 크기, 회전, 위치 등) 등을 들 수 있다. 본 논문에서 사용할 대상 영상의 경우, 영상의 촬영 환경은 고정되어 있기 때문에 데이터 확장 시 알약의 회전 및 위치 변화만 고려하여 데이터를 확장하고 학습하면 된다. 또한 실제 조제 및 검수 환경을 고려했을 때 한 포장 내 여러 알의 알약이 함께 들어가 있는 상태일 것이므로, 한 영상 내 여러 알약이 다양한 형태로 위치할 수 있도록 영상 데이터를 확장해야 한다.

우선 그림 2 (a)와 같이 앞/뒤 (정제류 약품 두 면의 형상이 다르거나 인쇄가 다른 경우, 캡슐류 약품 내 인쇄의 유/무 면 경우 등) 측면 (정제류 원형, 정방형 등 측면으로 서 있을 수 있는 경우 등) 등 알약이 검수 영상 내 존재할 수 있는 형태의 모든 면을 한 장씩 촬영하고, 알약들이 위치하게 될 배경 영상(그림 2 (b))을 준비한다.

다음으로 그림 3 과 같이 각 알약에 대하여 임의 면에 해당하는 영상을 선택해 임의의 각도로 회전한 뒤, 배경 영상 내 임의의 픽셀 위치에 합성하면 원하는 수만큼의 영상을 만들어 학습 영상을 확장할 수 있게 된다. 그림 3 과 같이 6 종의 알약에 대한 처방이 각 알약을 1 알씩 포함하는 경우, 영상 확장을 위한 모든 경우의 수는 수식 1 과 같다. 6 종의 알약을 A 에서 F 로 표현했을 때, 각 알약의 앞/뒤/측면 영상을 한 집합으로 보고 각 집합에서 하나씩을 선택하는 조합의 곱으로 경우의 수(수식 (1))를 계산하고, 여기에 각도의 변화를 1°로 설정했을 때 360,



(a) (b)

그림 2. 준비 영상의 예 (6 종 대상, (a) 단일 약품 영상, (b) 배경 영상)

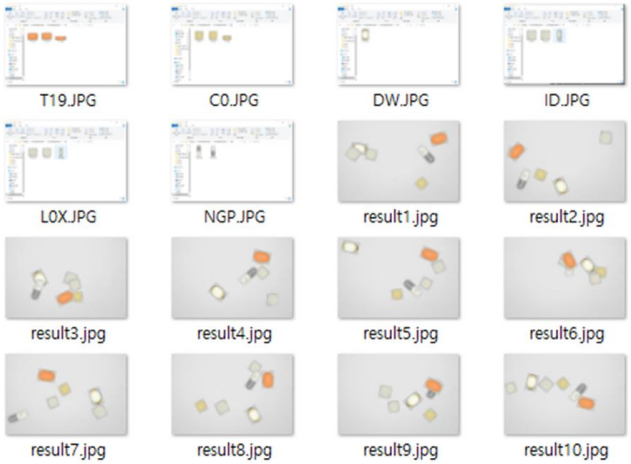


그림 3. 영상 확장의 예 (6 종 대상)

및 배경 영상의 픽셀 개수(가로×세로, 단일 알약 영상이 배경을 벗어나는 외곽 영역 제외)를 곱하면 모든 경우의 수를 표현할 수 있다. 단지 이렇게만 가정해도 엄청나게 많은 양의 데이터를 생성할 수 있는데 본 방법의 경우, 데이터 확장 시 배경 영상 내 각 알약을 합성하기 위한 픽셀의 위치를 알약의 위치로 사용하므로 알약의 이름과 함께 확장 영상 내 각 알약에 대한 자동 레이블링 또한 가능하게 된다.

$$A=\{a_f, a_b, a_{s1}, a_{s2}\},\quad D=\{d\}$$
$$B=\{b_f, b_b, b_s\},\quad E=\{e_f, e_b\}$$
$$C=\{c_f, c_b, c_s\},\quad F=\{f_f, f_b, f_s\}$$
$$,\quad {}_4\mathbf{C}_1 \times {}_3\mathbf{C}_1 \times {}_3\mathbf{C}_1 \times 1 \times 2 \times {}_3\mathbf{C}_1 \quad (1)$$

where f is front, b is back, s is side

최종적으로 단일 알약 영상에 대한 회전 및 합성 위치 등을 적절히 샘플링하여 무분별하게 많은 데이터 확장을 방지하고 이렇게 확장한 데이터를 학습에 사용하면, 그림 1 (b)와 같이 하나의 영상 내 여러 알약이 다양한 형태로 위치하는 경우에도 각 알약에 대한 보다 정확한 검출 및 인식이 가능하게 된다.

본 논문에서는 제안한 복합 알약 대상 데이터 확장 및 학습 방법의 효과를 검증하기 위하여 그림 4 와 같이 한 영상 내 여러 알의 알약을 다양한 형태로 위치하게 하여 시험 영상을 구성한 뒤, 시험 영상 내 각 알약들에 대한 검출 및 인식 결과를 확인하였다. 이와 함께 보다 정확한 성능 평가를 위하여 단일 알약 영상을 대상으로 각도 및 위치를 수동으로 바꿔가며 직접 촬영한 영상 및 이를 이용하여 각도와 위치만을 자동으로 변경하여 확장한 영상으로 학습을 수행한 뒤, 동일한 시험 영상들에 대한 결과를 비교하였다. 본 논문에서 사용한 딥러닝 모델은 YOLOv3 로, 학습에 사용한 영상은 총 2,000 장이며 4:1 의 비율로 학습 및 검증 세트를 구성하였다. 또한 배치 사이즈(Batch Size)는 64 로, 학습률(Learning Rate)은 0.001 로 설정하여 학습을 수행하였다.



그림 4. 시험 영상 (6 종 대상, 시험 영상 ① ~ ⑧)

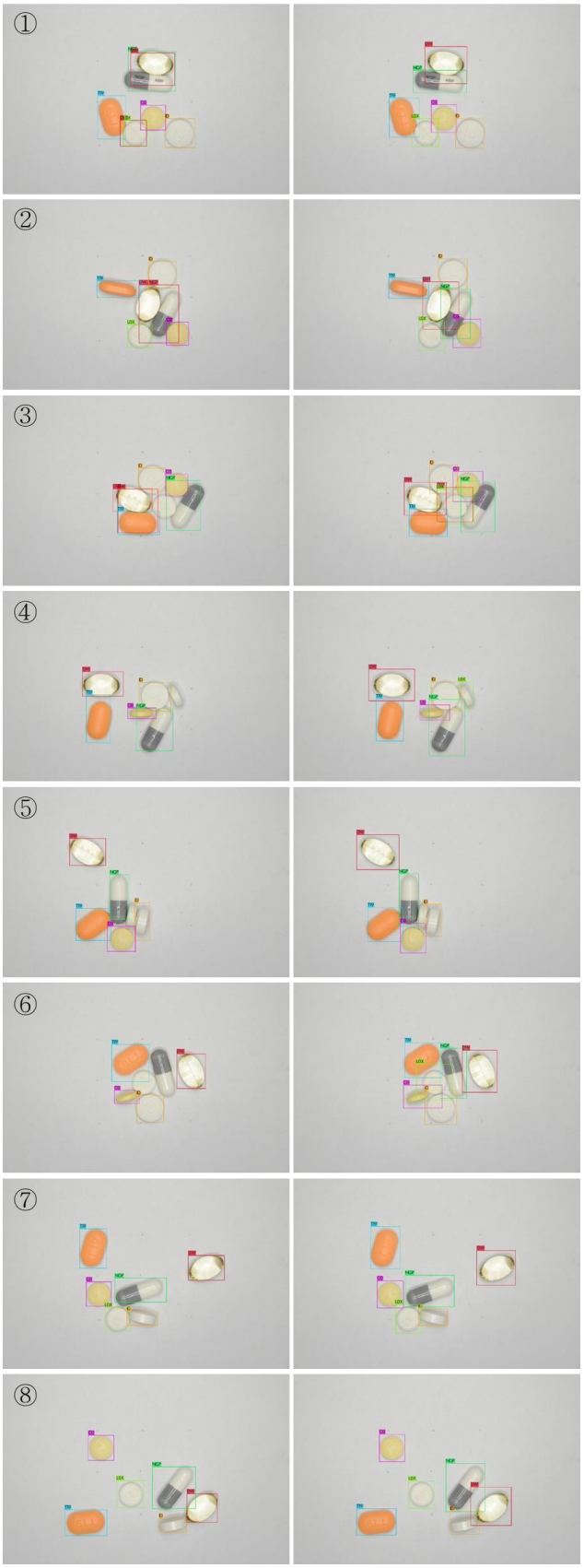


그림 5. 시험 결과 영상 (시험 영상 ① ~ ⑧, (a) 단일 알약 대상 확장 및 학습에 대한 결과, (b) 복합 알약 대상 확장 및 학습에 대한 결과)

시험 결과 그림 5 와 같이 한 영상 내 각 알약에 대한 검출 및 인식 성능을 확인할 수 있었다. 특별히 시험 영상 ①, ②의 경우, 그림 5 (a)에서 단일 알약 영상 대상의 확장 및 학습 시 인접한 알약을 하나의 알약으로 인식한 것을 볼 수 있지만, 그림 5 (b)와 같이 제안한 복합 알약 영상 대상의 확장 및 학습 시 이를 제대로 각각의 알약으로 인식하는 것을 볼 수 있다. 또한 시험 영상 ③, ④의 경우, 인접한 알약 중 하나를 검출하지 못하는 것을 볼 수 있지만 (그림 5 (a)), 그림 5 (b)와 같이 복합 알약 영상 대상의 확장 및 학습 결과에서 이를 제대로 검출하고 인식하는 것을 볼 수 있다. 물론 시험 영상 ⑤의 결과와 같이 두 방법 모두 인접 알약을 검출하지 못하는 경우도 있었지만 다른 시험 영상 ⑥, ⑦, ⑧ 모두에서 제안한 방법을 사용한 알약 검출 및 인식 결과가 단일 알약 대상의 확장 및 학습에 대한 결과와 비교하여 동등 혹은 우수한 검출 및 인식 성능을 보이는 것을 확인할 수 있다.

이러한 결과들을 바탕으로 본 논문에서 제안한 복합 알약 영상 대상의 데이터 확장 및 학습 방법이 의약품 비전 검수 장치 및 실제 조제 자동화 시스템 적용에 더 적합한 방법이라 판단할 수 있다. 다만 본 논문의 시험 결과들이, 학습 및 시험에 사용한 데이터가 충분히 많지 않고 알약의 종이 실제 약국 또는 병원에서 사용하는 것보다 현저히 작아 아주 신뢰할 만한 결과라 할 수 없고, 영상 확장 및 학습에 사용할 최적의 경우의 수 및 영상 장 수를 제안한 것이 아니라 실제 현장에 바로 적용할 수 있을 만한 결과는 아니지만, 본 방법을 통하여 앞서 설명한 AI 기반의 비전 검수 기술 개발 및 이를 조제 자동화 시스템에 적용할 경우 예상되는 어려움들을 어느 정도 해결할 수 있을 것으로 보인다.

Ⅲ. 결 론

본 논문에서는 약품 조제 자동화를 위한 AI 기반의 의약품 비전 검수 기술 개발에 있어, 실제 조제 및 검수 환경을 고려한 복합 알약 영상 대상의 데이터 확장 및 학습 방법을 제안하고 이에 따른 효과를 확인하였다. 결론적으로 본 논문에서 제안한 방법을 통하여 의약품 비전 검수 기술의 개발에 있어 AI 기반의 기술을 훨씬 효과적으로 적용할 수 있을 것으로 보이며, 이와 더불어 AI 기반의 비전 검수 기술을 적용한 조제 자동화 시스템 상용화를 통한 전문 의료진들의 업무 효율 상승 및 관련 기업들의 추가 이윤 창출이 가능할 것으로 기대된다.

ACKNOWLEDGMENT

본 연구 논문은 한국전자통신연구원 연구운영지원사업의 일환으로 수행되었음. [21ZD1140, 지능형 의료.보건 산업 실용화 기술 개발]

참 고 문 헌

[1] 김휘강, 이상범, 김광용, 김규형, “조제 자동화 시스템을 위한 딥러닝 기반 약품 검수 방법에 관련 연구,” 한국통신학회 학술대회논문집, pp. 108-110, 2020.

[2] Y. Wang, J, Ribera, C. Liu, S. Yarlagadda, and F. Zhu, “PILL RECOGNITION USING MINIMAL LABELED DATA” *2017 IEEE Third International Conference on Multimedia Big Data*, pp. 346-353, 2017.

CART 기반의 변수랭킹을 이용한 곱셈형 고장의 고장변수식별

유정원, 김진홍, 이영재, 임길택
한국전자통신연구원 대경권연구센터
{gardenyoo, jinhong, lyj4295, ktl}@etri.re.kr

Faulty variable isolation of multiplicative fault using CART-based variable ranking

Jungwon Yu, Jinhong Kim, Youngjae Lee and Kil-Taek Lim
Electronics and Telecommunications Research Institute

요 약

고장변수식별(faulty variable isolation)은 대상공정에서 발생 된 고장(fault)과 가장 밀접하게 연관된 공정변수를 식별하는 작업이다. 복잡한 산업공정의 고장변수를 정확하게 식별하는 작업은 운전원과 엔지니어가 고장을 정확하게 진단(diagnosis)할 수 있도록 도움을 준다. 본 논문에서는, classification and regression tree (CART) 기반의 변수랭킹 방법을 곱셈형 고장(multiplicative fault)의 고장변수식별을 위해 적용한 결과를 제시한다. 실험결과는 CART 기반의 변수랭킹 방법이 되묻음 효과(smearing effect) 없이, 고장변수를 정확하게 식별할 수 있음을 보여준다.

Ⅰ. 서 론

복잡한 산업공정에서, 적시에 정확하게 발생 된 고장(fault)과 연관된 공정변수(process variable)를 식별하는 작업은 안전하고, 효율적이며, 경제적인 공정운영을 위해서 필수적이다. 여기서 '고장'이란 공정의 운전 중에 발생하는 비정상적인 이벤트를 의미한다. 고장변수(faulty variable)를 식별하기 위해서 가장 널리 사용되는 방법은 기여도 분석(contribution analysis)이다. 기여도 분석은 고장탐지(fault detection)를 위해서 계산된 탐지지수(detection index)에 대한 각 변수의 기여도 값을 계산한 후, 계산된 기여도 값이 높은 변수를 고장변수로 식별한다. 기여도 분석은 발생 가능한 고장유형(fault type)에 대한 사전지식과 과거 고장데이터 없이, 빠르고 손쉽게 고장변수를 식별할 수 있지만, 되묻음 효과(smearing effect)로 인해 공정 운전원이나 엔지니어에게 혼란을 초래할 수 있다. '되묻음 효과'란, 고장이 탐지된 이후에 수행되는 기여도 분석에서 고장변수의 기여도 값 뿐만 아니라 정상변수의 기여도 값 역시 변화되는 현상이다. 복잡한 고장이 발생했을 경우, 기여도 분석에 의한 고장변수식별은 상당히 애매모호한 진단(diagnosis)결과로 이어질 수 있다.

본 논문에서는, 참고문헌 [1]에서 제안된 classification and regression tree (CART) 기반의 변수랭킹(variable ranking) 방법을 곱셈형 고장(multiplicative fault)의 고장변수식별(faulty variable isolation)을 위해서 적용한 결과를 제시한다. 적용결과는 CART 기반의 변수랭킹 방법이 정상변수로의 되묻음 없이 고장변수를 정확하게 식별할 수 있음을 보여준다.

Ⅱ. CART 기반의 변수랭킹 방법

참고문헌 [1]에서 제안된 CART 기반의 변수랭킹을 이용한 고장변수식별 방법에서는, 고장샘플(faulty sample)과 정상샘플(normal sample)로 구성된 학습데이터에 CART 알고리즘을 적용하여 이진분류트리(binary classification tree)를 건설한 후, 최종트리(final tree)로부터 각 입력변수의 중요도(variable importance) 값을 추출하여 변수랭킹을 수행한다. 정상 클래스가 할당된 정상샘플은 대상공정이 정상적으로 동작할 때 수집된 공정데이터이고, 고장 클래스가 할당된 고장샘플은 고장탐지기법에서 탐지지수가 문턱값(threshold)을 위배한 시간구간에 해당하는 공정데이터이다. 계산된 중요도 값이 높은 변수들은 원 입력공간(original input space)에서 정상샘플과 고장샘플을 분류하기 위한 중요한 변수들이다. 즉, 클래스를 구분을 위한 설명력(explanatory power)이 높은 변수들이다.

참고문헌 [1]에는 CART 기반의 변수랭킹을 이용한 고장변수식별을 위해서, 두 가지 접근방법이 사용되었다. 첫 번째 접근방법('방법 1')은 전체 고장구간에 대하여 발생된 고장과 가장 밀접하게 연관된 고장변수를 식별하기 위함이고, 두 번째 접근방법('방법 2')은 time window sliding을 이용하여 고장이 어떻게 파급되는지를 관찰하기 위함이다.

Ⅲ. 실험결과

모니터링 변수의 공분산 구조(covariance structure)에 이상이 생기는 고장을 '곱셈형 고장'이라고 정의한다 [2]. 곱셈형 고장의 경우, 곧바로 시스템실패(system failure)로 이어지지 않을 수도 있지만, 대상 설비의 수명과 밀접한 관련이 있다. 본 논문에서는, 참고문헌 [2]에서 사용된 곱셈형 고장과 관련된 벤치마크(benchmark) 문제에 CART 기반의 변수랭킹방

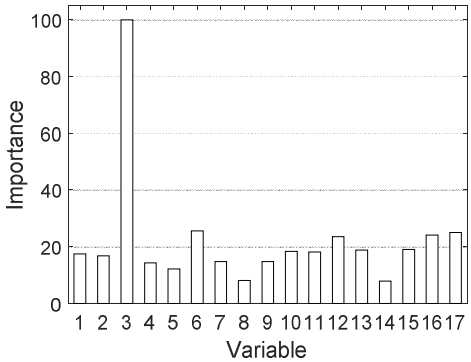


그림 1. '방법 1'에 의해서 전체 고장구간(k = 101,..., 500)에서 계산된 변수 중요도 값 [3].

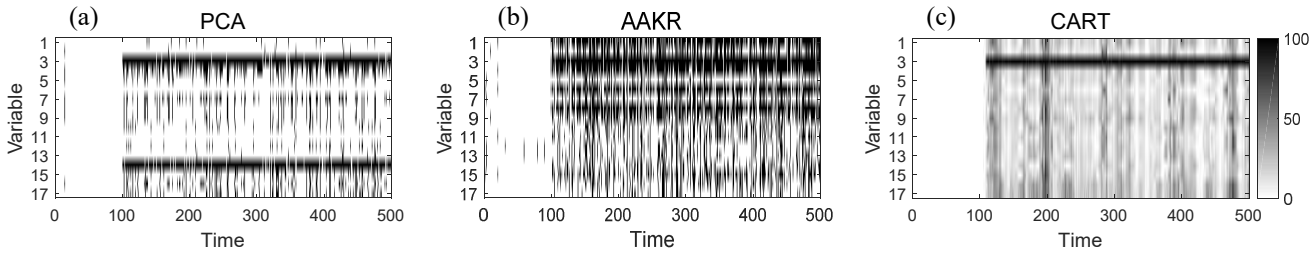


그림 2. 비교방법 및 '방법 2'에 의한 고장변수식별 결과: (a) 주성분분석 기반의 기여도 분석; (b) AAKR 기반의 기여도 분석; (c) CART 기반의 변수랭킹을 이용한 고장변수 식별 [3].

법을 적용하여 고장변수식별 작업을 수행한다. 시뮬레이션을 위한 데이터는 아래의 식[2]에 의해서 생성된다.

$$\begin{aligned}\mathbf{y}(k) &= \mathbf{W}\mathbf{x}(k) + \boldsymbol{\xi}(k) \\ \boldsymbol{\theta}(k) &= \boldsymbol{\Psi}\mathbf{y}(k) + \mathbf{b} + \boldsymbol{\eta}(k)\end{aligned} \tag{1}$$

여기서 $\mathbf{x}(k) \in R^3$, $\mathbf{y}(k) \in R^{15}$, $\boldsymbol{\theta}(k) \in R^2$, $\mathbf{W} = rand(15,3) \in R^{15\times 3}$, $\boldsymbol{\Psi} = 1 + rand(2,15) \in R^{2\times 15}$, $\mathbf{b} = [1055, 1825]^T$, $\mathbf{x} \sim MVN_3(0, diag[4, 1.96, 1.44])$, $\boldsymbol{\xi} \sim MVN_{15}(0, 10^{-16}\mathbf{I}_{15})$, $\boldsymbol{\eta} \sim MVN_2(0, 10^{-8}\mathbf{I}_2)$이고, 행렬 $rand(\cdot,\cdot)$의 각 성분은 0에서 1 사이의 균등분포를 따른다. 식 (1)로부터, 17개의 공정변수 $[\mathbf{y}^T(k), \boldsymbol{\theta}^T(k)]^T$로 구성된 500개의 정상샘플을 생성하여, 다변량데이터 기반의 고장탐지 모델의 학습을 위해 사용하며, 이 정상데이터에는 분류트리를 건설하기 위해 정상 클래스가 할당된다. 고장변수식별의 성능을 검증하기 위해서, 고장데이터는 다음과 같은 시나리오에 의해서 생성된다. 먼저 식 (1)로부터 500개의 정상샘플을 생성한 후에, 곱셈형 고장이 발생한 시간구간(k = 101,..., 500)에 대해 식 $\mathbf{y}(k) = \mathbf{F}\mathbf{y}(k)$에 의해서 곱셈형 고장효과(fault effect)가 주입된다. 여기서 행렬 F은 고장구간에서는 $\mathbf{F}(3,6) = \mathbf{I}_{15}(3,6) + 5$로 정의되며, 정상구간에서는 $\mathbf{F} = \mathbf{I}_{15}$를 만족한다.

먼저 '방법 1'을 이용한 전체 고장구간(k = 101,..., 500)에서의 고장변수식별결과에 대해서 살펴보자. 그림 1[3]은 위에서 설명한 과정에 의해서 생성된 곱셈형 고장데이터에 '방법 1'을 적용하여 계산된 변수중요도이다. 그림 1에서 확인할 수 있듯이, 고장변수 y_3의 중요도 값은 100이며, 나머지 정상변수의 중요도 값은 30보다 작다.

그림 2[3]는 주성분분석(principal component analysis, PCA) 및 auto-associative kernel regression(AAKR)에 기반한 기여도 분석과 '방법 2'를 곱셈형 고장데이터에 적용하여 얻어진 고장변수식별 결과를 보여준다. 주성분분석 기반의 기여도 분석은 고장변수인 y_3뿐만 아니라, 정상변수인 y_{14} 역시 고장변수로 식별한다. 즉, y_{14}에 되묻음이 발생하였다. AAKR에 기반한 기여도 분석에서는, 되묻음 효과가 더 심하게 발생하며, 정상변수 y_1, y_4, y_6, y_8에서 되묻음이 발생함을 관찰할 수 있다. '방법 2'의 경우, 고장변수 y_3의 중요도 값은 항상 100이다. y_3를 제외한 나머지 정상변수의 중요도 값은 시간이 지남에 따라 변화가 있지만, 그 크기는 고장변수의 중요도보다 항상 작다.

IV. 결 론

CART 기반의 변수랭킹을 이용한 고장변수식별 방법[1]은 원 공간에서 정상샘플과 고장샘플을 구분하기 위해서 설명력이 높은 공정변수를 고장변수로 식별한다. 즉, 원 공간에서 결정경계(decision boundary)를 설정하는 분류트리에 기반하여 고장변수식별이 수행되므로, 되묻음 효과가 발생하지 않는다. 본 연구에서는, CART 기반의 변수랭킹을 이용한 방법을 곱셈형 고장에 적용한 결과를 제시하였다. 실험결과로부터 CART에 기반한 방법이 비교방법인 주성분분석 및 AAKR에 기반한 기여도 분석에 비해서 더 명확하게 곱셈형 고장의 고장변수를 식별할 수 있음을 확인할 수 있다.

ACKNOWLEDGMENT

This work was supported by Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korean government. [21ZD1120, Development of ICT Convergence Technology for Daegu-GyeongBuk Regional Industry]

참 고 문 헌

[1] J. Yu, J. Jang, J. Yoo, J. H. Park, and S. Kim, "A Fault Isolation Method via Classification and Regression Tree-Based Variable Ranking for Drum-Type Steam Boiler in Thermal Power Plant. Energies, vol. 11, no. 5, pp. 1-19, 2018.

[2] H. Hao, K. Zhang, S. X. Ding, Z. Chen, and Y. Lei, "A data-driven multiplicative fault diagnosis approach for automation processes," ISA Trans., vol. 53, no. 5, pp. 1436-1445, Sep. 2014.

[3] J. Yu, "Data-Driven Fault Detection and Isolation of Complex Industrial Processes Using Hybrid of Multivariate Statistical Techniques," Ph.D. dissertation, Dept. of Elect. and Comput. Eng., Pusan Nat. Univ., Busan, South Korea, 2018.

한국노인노쇠코호트사업 데이터 분석과 활용 - 기계학습과 딥러닝 적용, 확산을 위하여

구동준, 이아라, 이은주, 김일곤

경북대학교

dongjunkoo02@gmail.com, eara0367@gmail.com, jewelee@knu.ac.kr, ikkim@knu.ac.kr

Data Analysis and Utilization in Korean Frailty and Aging Cohort Study - Toward Adaptation and Expansion of Machine Learning and Deep Learning

Dong Jun Koo, Ah Ra Lee, Eunjoo Lee, Il Kon Kim

Kyungpook National University

요 약

노쇠는 인간수명의 증가와 노화에 따른 필수적으로 수반되는 증상 중 하나이다. 그리하여 건강 수명 및 삶의 질 향상을 위해서는 노쇠의 발현을 예방하고 지연시키는 것이 매우 중요하다. 최근 급속히 증가하는 빅데이터를 활용하여 노쇠의 발현을 조기에 예측할 수 있다면, 노쇠 진단을 위해 수행되는 다양한 검사와 그에 따르는 비용과 노력을 경감시킬 수 있을 것이다. 본 연구에서는 노쇠코호트 데이터에 근거하여 가장 잘 알려진 Fried 노쇠 지표 이외에 머신러닝을 활용하여 노쇠 판단에 도움이 될 수 있는 관련 요소들을 발굴 및 분석하고자 한다. 이러한 과정을 통해 궁극적으로 노쇠를 보다 효과적으로 예측할 수 있는 방안을 모색하고자 한다.

Ⅰ. 서 론

노쇠란 세포의 노화로 인해 신체 능력이 저하되면서, 저하되기 이전 상태로 다시 돌아가지 못하는 상황이 반복되다가, 언젠가는 영영 회복 불능의 상태로 있는 것을 의미하며, 인간에게 있어 불가피한 현상 중 하나이다 [1]. 노쇠로 노인들은 행동 제어가 힘들어지고 균형감을 상실함으로써 다양한 낙상사고가 발생하게 된다. 노인들의 낙상사고는 골절과 밀접한 연관이 있으며, 이로 인해 외상 상태에서 헤어나오지 못하고 결국 사망에까지 이르게 되는 경우가 종종 발생한다. 따라서 노쇠를 미리 예측하고, 노인들이 자신의 노쇠 정도를 인지할 수 있도록 한다면, 낙상을 예방할 수 대책을 마련할 수 있으므로 노쇠를 예측하는 것은 노인들의 삶의 질 뿐 건강 수명연장에도 매우 중요하다.

현재 우리는 빅데이터 시대에 살고 있다. 기술의 발전과 함께 더욱더 많은 데이터가 매일 생성되고 있으며 진보된 의료장비를 통한 더욱더 다양한 신체정보 수집이 가능해졌다 [2]. 또한, 스마트워치, 바이오센서와 같은 일상생활에서도 측정이 가능한 장비를 통해 사용자의 생활 방식 또한 수집되어 분석이 가능해져 이러한 데이터를 잘 활용하는 것이 매우 중요하다.

실제로, KFACS(Korean Frailty and Aging Cohort Study, 한국 노인노쇠코호트사업단, https://www.kfacs.kr/html/) [3] 에서 개발되어 수집되어 온 노쇠 코호트 데이터 셋은 설문 항목 이외에도 검사장비를 통해 수집한 다양한 건강 관련 자료가 수집되어 있다. 이렇듯, 노쇠와 관련된 데이터들이 기하급수적으로 늘어나고 있지만 이를 활용하여 노인들의 건강관리 및 사고를 예방하는 대책을 마련하는 데는 아직 거의 연구가 수행되고 있지 않다. 즉, 어떤 데이터가 노쇠 판별에 유용한지를 수집된 데이터에 근거하여 분석하기보다는 설문 문항에 의존하여 노인들의 노쇠과정을 추적, 관찰하는 데에만 집중하였기 때문이다. 이에, 본 논문에서는 기계학습 더 나아가 딥러닝으로 노쇠 코호트 데이터의 효과를 증가시킬 수 있는 방안을 제시하고자 한다.

Ⅱ. 본론

노쇠를 측정하기 위한 방법에 있어 Fried와 Rockwood의 노쇠 지표가 많이 활용되고 있다. Fried 노쇠 지표에서는 약 5가지 항목(체중 감소, 체력 고갈, 행동 저하, 보행 속도, 나약함)을 사용하고 [4], Rockwood 지표에서는 70가지 항목을 사용한다 [5]. 이들은 기준 항목 중 해당 항목이 일정 개수 이상 포함되면 노쇠로 판별하는 방식이어서 측정이 쉽고 빠르다는 장점이 있다. 이를 통한, 단기간에 많은 노쇠 결과 도출이 가능하지만, 지표 항목이 고정적임으로 새로운 데이터에 대한 연계가 부족함으로 머신러닝을 활용하여 기존의 지표를 보완한 새로운 요소를 찾고자 한다.

본 연구에서는 노쇠와 새로운 요소의 적용 가능성을 설립하기 위하여 그림 1과 같이 기계학습 위에 현존하는 노쇠 지표를 적용하여 노쇠와 노인들의 일상에 관한 다양한 요인들의 관계성을 확보하는 맞춤형 노쇠 예측 표준 프레임워크를 제안하고자 한다. 그림 1에서 보는 바와 같이, 지표 기반 Korean Instrumental Activities of Daily Living (K-IADL)에서의 머신러닝 분류 예시에서 드러난 것처럼, Frail 측정에 활용 가능한 지표를 머신러닝을 적용하여 분류 학습을 통한 노쇠와 정상 두 가지의 분류 학습률을 비교·분석하여 발견하고 자하였다. 또한 Frail 측정에 유용한 지표가 전문가들이 가진 주관적 판단에 의해 한 단계 더 높은 수준의 유효성을 갖게 되면, 이를 노쇠 관련 중요 요소로 등록하여 활용할 수 있도록 할 수 있다. 이렇게 투입된 정보와 검사결과가 함께 통합되어 진다면, 연령, 위치 장소, 성별, 등과 같은 개인의 특성을 고려하여 다양한 데이터 모델이 개발되어 질 수 있으며, 개인맞춤 판단도 가능해진다. 또한 설문 조사에 있어 반복된 학습과 여러 가지 방면에서의 데이터를 확보함으로써, 특이점 보완과 함께, 보다 정밀한 분석을 통해 수준 높은 노쇠 예측이 가능하고, 이러한 예측결과가 국제표준으로 확립된다면, 딥러

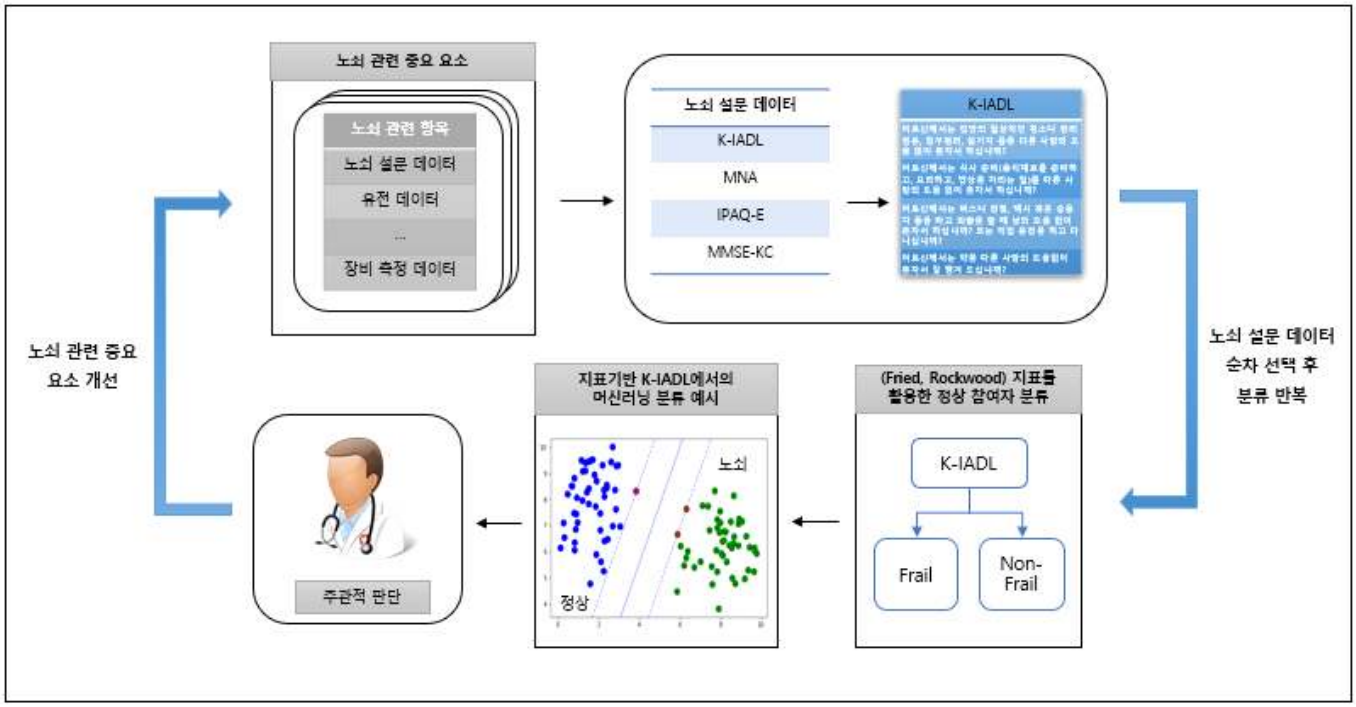


그림 1. 머신러닝 분류 기반 노쇠 관련 중요 요소 개선도

닝 기법의 적용이 가능할 수 있도록 맞춤형 노쇠 설문지 개발 표준 프레임워크 개발에도 회환되어 질 수 있을 것이다.

Ⅲ. 결론

본 논문에서는 다양한 정보수집이 가능해지는 초연결사회에서, 노인들의 건강 수명에 있어 매우 중요한 요소 중 하나인 노쇠 판별을 위해서 어떤 새로운 요소를 추가하는 것이 노쇠 예측을 더 빨리 정확하게 할 수 있는지를 확인하고자 하였다. 이를 위해, 기계학습, 전문가 의견 수집을 모두 포함할 수 있는 방안을 제안하였다. 또한, 노쇠 판별에 사용되는 지표 이외에 생활 환경에서 자연스럽게 수집되는 데이터와 연계를 통해 개인 맞춤형 노쇠 예측이 가능해짐으로 조기에 노쇠를 예측하고 노쇠로 인해 발생하는 다양한 낙상사고를 예방하고자 한다. 다만 일상생활 중에 발생하는 다양한 데이터들이 얼마나 정확하게 노쇠를 예측할 수 있는지에 대한 검증은 추후 연구에서 수행되는 것이 필요하다고 제안하는 바이다.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2019R1I1A3A01058112, 2019R1I1A3A01060579), the Technology Innovation Program (20011061, Personalized digital health framework for infectious diseases and frailty prevention) funded By the Ministry of Trade, Industry & Energy(MOTIE, Korea), and a grant of the Korea Health Technology R&D Project through the Korean Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HI15C3153)

참 고 문 헌

[1] John Y. "Frailty – what it means and how to keep well over the winter months," 2013, (https://www.england.nhs.uk/blog/frailty/).

[2] UNIVERSITY OF ILLINOIS CHICAGO "Big Data and Wearable Health Monitors: Harnessing the Benefits and Overcoming Challenges," 2019, (https://healthinformatics.uic.edu/blog/big-data-and-wearable-health-monitors-harnessing-the-benefits-and-overcoming-challenges/).

[3] Won, Chang Won et al. “Korean frailty and aging cohort study (KFACS): cohort profile,” Submission to BMJ, April 2020.

[4] Fried, L P et al. “Frailty in older adults: evidence for a phenotype.”Submission to The journals of gerontology. Series A, Biological sciences and medical sciences, March 2001.

[5] Rockwood, Kenneth et al. “A global clinical measure of fitness and frailty in elderly people,” Submission to CMAJ : Canadian Medical Association journal, August 2005.

이미지 인페인팅을 위한 효과적인 제거 영역 선택 기법

윤상빈, 정순기
경북대학교 컴퓨터학부
kor135@knu.ac.kr, skjung@knu.ac.kr

Efficient Removal Region Selection for Image Inpainting

Sang Bin Yun, Soon Ki Jung
School of Computer Science and Engineering, Kyungpook National University

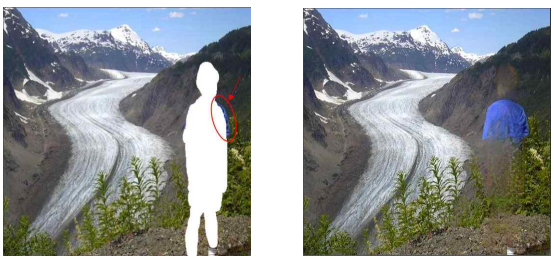
요 약

이미지 인페인팅은 이미지에서 특정 영역을 제거하고 매끄럽게 채우는 작업을 의미한다. 이미지 인페인팅에서 물체 제거 대상 영역을 설정하는 방법은 사람이 수작업을 하거나 사각형 박스로 설정한다. 수작업의 경우 많은 시간과 비용이 소모되고 사각형 박스의 경우 물체만 제거하기 위해 필요한 영역 이외에 사각형 박스 안의 나머지 부분도 새로 생성해내기 때문에 불필요한 정보의 손실이 발생한다. 따라서 본 논문에서는 물체를 제거하는 영상 작업에서 의미론적 영상분할 기술을 이용하여 물체 제거에 불필요한 영역 설정을 최소화하여 수작업의 방법에 비해서는 시간과 비용적 문제를 최소화하였고 사각형 박스보다는 자연스러운 이미지 생성하는 결과를 보였다.

1. 서 론

본 논문은 이미지 인페인팅[1] 분야 중 이미지에서 물체를 제거하고 물체를 지운 영역을 주변 배경과 자연스럽게 채우는 것에 중점을 두어, 지우고자 하는 대상 물체의 설정 과정에서 수작업을 최소화함과 동시에 불필요한 정보손실을 최소화함을 목표로 한다. 현재 대부분의 이미지 인페인팅 모델들은 목표 물체를 제거하기 위한 영역 설정을 위해 사각형 박스(bounding box) 또는 자유롭게 채색(freeform)하는 것으로 이루어져 있다. 사각형 박스의 경우 관심 물체 이외의 영역을 포함하여 제거하고 새로운 이미지를 생성하기 때문에 필요 이상의 정보가 손실되는 문제가 있다. 자유롭게 채색하는 경우 수작업으로 진행해야 하기 때문에 시간, 비용적 요소가 소모된다. 또한 지우고자 하는 물체를 모두 포함하지 않은 경우 해당 영역이 영향을 주게 되어 아래 (그림 1)과 같이 자연스러운 배경 생성을 하는 과정에 좋지 못한 영향을 주게 된다.

따라서 본 논문에서는 이러한 문제를 해결하기 위해 의미론적 영상분할(semantic segmentation)을 통해 지우고자 하는 물체의 영역을 정확하게 구별하여 물체 이외의 정보 손실을 최소화하여 영상 영역을 제거하고 자연스럽게 배경 영상을 생성한다.



관심 영역 결과 이미지

(그림 1)

2. 관련 연구

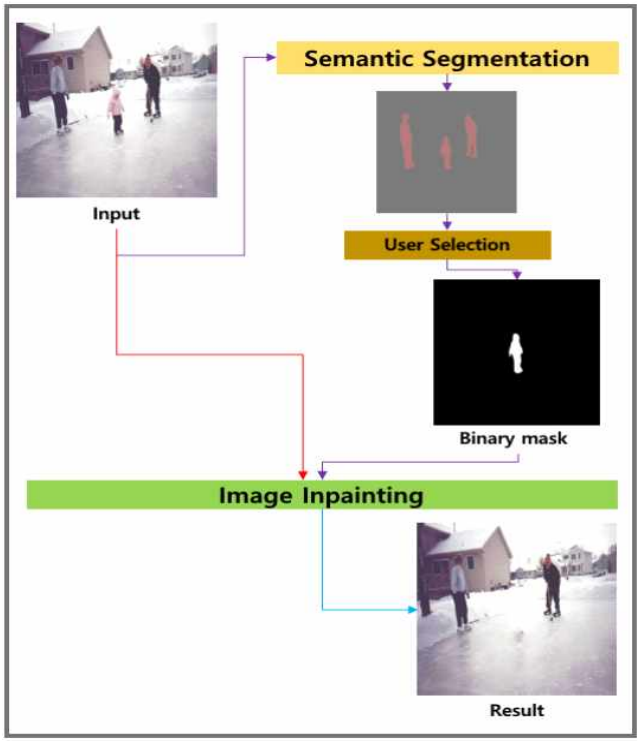
2.1 이미지 인페인팅(Image Inpainting)

이미지 인페인팅은 이미지에서 지워진 영역을 채우는 작업을 의미한다. 딥러닝 이전의 이미지 인페인팅 방식은 확산 기반(difussion based) 방법[1]과 패치 기반(patch based) 방법[2]으로 해당 영역을 채우는 것으로 해당 문제를 해결해왔다. 확산 기반 방법[1]은 경계 주변의 픽셀 값의 미분방정식을 이용해 채우기 때문에 채워야 할 영역이 클 경우 블러링(blurring)되는 문제가 있고 패치 기반 방법[2]은 이미지 내의 패치만을 이용해서 채우기 때문에 새로운 이미지를 생성해 내야 하는 작업에서는 부자연스러운 느낌이 드는 이미지를 생성해낸다. 딥러닝 이후의 이미지 인페인팅 방식[3]은 특히 적대적 생성 모델[4] 이후 이미지 인페인팅의 성능이 크게 향상되었다. 기존의 방식은 현재 가지고 있는 영상의 정보만을 이용해서 빈 공간을 채우는 방식이기 때문에 새로운 이미지를 만들어내는 것에 한계가 있었지만 적대적 생성 모델은 딥페이크, 워터마크 제거, 물체 제거 등 다양한 응용분야에 이용되고 있다.

2.2 의미론적 영상분할(Semantic Segmentation)

의미론적 영상분할은 이미지 내에 있는 객체들을 의미있는 단위로 분할하는 작업을 말한다. 즉 영상을 여러 개의 픽셀 집합으로 나누는 과정이며 단순히 이미지를 분류(classification)하고 객체를 탐지(detection)하는 것을 넘어서 객체의 경계 영역까지 정확하게 분류하는 문제이다. 의미론적 영상 분할 모델은 FCN[5] 이후로 딥-러닝 기반의 모델로 확장되어 다양한 응용분야에 이용되고 있다.

3. 의미론적 영상분할을 이용한 물체 제거



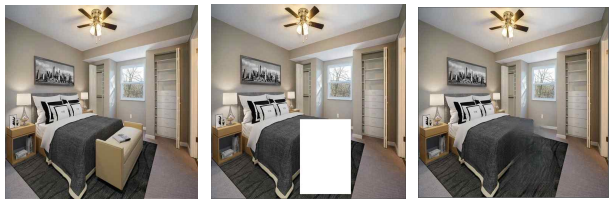
(그림 2) 제안된 물체제거 구조

위 그림 2의 구조에서 입력 이미지는 의미론적 영상분할 모델을 통과하여 물체 영역이 분할된다. 분할된 이미지에서 목표로 하는 물체를 유저 선택 모듈을 통해 원하는 객체를 이진 마스크 영상으로 만든 뒤 입력 이미지와 함께 이미지 인페인팅 모델의 입력으로 들어가 최종 이미지가 출력된다.

4. 실 험

본 실험에서는 기존의 물체 제거 작업 중 사각형 박스를 이용한 영역 설정 방법과 제안된 모델을 비교하였다. 제안된 모델에서 의미론적 영상분할은 ADE20k 데이터셋을 기반으로 학습한 HRNetV2[6] 모델을 이용하였고 인페인팅 모델은 Places2 데이터 셋으로부터 학습된 AOT-GAN[7]를 이용하여 실험을 진행하였다.

본 실험의 결과는 아래 (그림 3)에 나타나있다. (그림 3a)에서 사각형 박스의 경우 지우고자 하는 물체 이외의 정보가 함께 삭제되는 문제가 발생한다. 특히 물체의 모양이 사각형과 차이가 나거나 대상 물체와 다른 물체에 겹쳐있을 경우 (그림 3a)의 침대 모서리 부분처럼 불필요한 정보의 삭제로 인한 부자연스러움이 발생할 수 있다. 이 경우 제안된 모델을 활용할 경우 (그림 3b)의 결과와 같이 불필요한 정보가 삭제되는 것을 방지할 수 있고 해당 물체만을 삭제하여 주변 환경에 맞게 재생성이 가능하다.



(a) 사각형 박스



(b) 제안된 모델

(그림 3) 영역설정에 따른 인페인팅 결과 이미지

5. 결론 및 향후 연구

본 논문에서는 이미지 생성 작업 중 물체 제거에서 기존의 물체 영역을 선택하는 방법의 문제를 해결하기 위해 의미론적 분할 모델을 이용하여 물체 영역을 선택할 수 있는 모델을 제안하였다. 제안된 모델을 활용하면 인적, 시간적 요소를 최소화하고 또한 불필요한 영역이 선택되는 문제를 최소화할 수 있다. 또한 제안된 모델은 비디오에서의 물체를 제거하는 응용 기술로 사용될 수 있다. 향후 물체의 그림자 영역을 포함한 의미론적 영상분할 모델로 확장하여 다양한 물체에 적용이 가능하도록 연구가 진행될 예정이다.

Acknowledgement

본 논문(저서)은 교육부 및 한국연구재단의 4단계 BK21 사업(경북대학교 컴퓨터학부 지능융합 소프트웨어 교육연구단)으로 지원된 연구임 (4199990214394)

참 고 문 헌

[1]. M. Bertalmio, et al., “Image inpainting,” in SIGGRAPH, 2000.
[2]. A. Criminisi, P. P´erez, and K. Toyama, “Region filling and object removal by exemplar-based image inpainting,” in TIP, 2004.
[3]. Ian J. Goodfellow, et al., “Generative Adversarial Networks”, in NIPS ,2014.
[4]. A. Radford, et al. “Unsupervised representation learning with deep convolutional generative adversarial networks”, in ICLR , 2016.
[5]. J. Long, et al., “Fully convoutional networks for semantic segmentation”, in CVPR, 2015.
[6]. Ke Sun, et al., “High-Resolution Representations for Labeling Pixels and Regions”, in arXiv:1511.06434, 2019.
[7]. Yanhong Zeng, et al., “Aggregated Contextual Transformations for High-Resolution Image Inpainting”, in arxiv.org/abs/2104.01431, 2021.

Understanding Resource Usage of NoSQL Databases Through System call Trace

Changho Seo, Yunchang Chae, Jaeryun Lee, Euiseong Seo*, and Byungchul Tak
Kyungpook National University, *Sungkyunkwan University
[changho.seo, cyc, jrlee, bctak]@knu.ac.kr, *euiseong@skku.ac.kr

Abstract

NoSQL databases have quickly become an indispensable component for big data processing and AI applications. However, it is very challenging to select the best NoSQL database that fits the performance and scalability requirements of the big data and AI application at hand. To address these challenges, we investigate the feasibility of the approach that tries to build accurate resource consumption models of NoSQL databases from low-level system call information. Accurate resource consumption models would enable us to reason about the expected performance and the scalability capacities for given workloads and hardware specifications. Our study reveals that NoSQL databases exhibit unique resource consumption behaviors even for functionally the same operations. Also, we verify that it is feasible to build accurate resource consumption models with only a small number of experiments.

Ⅰ. Introduction

NoSQL databases are established as the essential component for big data and AI applications. Selecting the right NoSQL database as the backend storage for the big data and AI is critical for the performance and efficiency. However, it is challenging to select the most fitting one for the type of applications to be run. Static information, such as software features or capabilities from various document sources can provide only the general information. Key information about the actual performance characteristics and scalability under the expected workloads are not easily derivable from such documentations.

Performance benchmarking is one way to build performance models and infer the scalability [1], [2]. Unfortunately, these black-box methods have several shortcomings. Benchmarking approaches can cover only a fraction of vast search space made of workload types, configurations, and environment preferences. Furthermore, it is difficult to explain the observed performance outcomes and reason for various scenarios, such as different workload intensity and mix.

In this work, we address these challenges by investigating the gray box techniques for building accurate resource consumption models of NoSQL databases. We use system call activities directly related to the resource consumption such as read, write, send, recv and futex system calls. These system calls decorated with CPU cycles, memory bandwidth, and I/O byte accounting information from the basis of our resource model.

In our grey-box technique, system call-level information is used because they offer several advantages. First, developing a technique does not require us to have application-specific knowledge such as internal function descriptions or software architecture. Second, system calls directly provide I/O resource accounting data through return values. Third, source code is not required since we do not intend to instrument the code. Lastly, the amount of traced data we need to process is significantly less than those of function-level or instruction-level traces.

We were able to develop a set of novel algorithms to handle all challenges mentioned above and built the desired resource consumption models of three key operations (insert, select, delete) for Redis, Memcached and MongoDB NoSQL databases.

Ⅱ. Design

We describe the detailed architectural design of our system for resource consumption modeling in Figure 1. We divide the functionality into three nodes – Experiment, Analysis, and Control node.

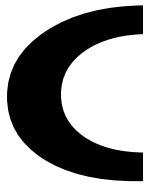


Figure 1. System Architecture

Our target events for tracing are the entry and exit point of all system calls. We capture the metrics at the entry and exit of a system call invocation, and calculates the elapsed time. This elapsed time indicates the number of CPU cycles used during the system call execution. Similarly, the difference between the exit of one system call and the entry of the next system call gives us the CPU cycles consumed by the application (i.e. NoSQL database of choice) during the user mode. The memory bandwidth usage can be obtained similarly. The network and disk I/O bandwidth usage is obtained from the return values of these system calls.

There are a few known tools available for tracing the system calls from target applications and gather resource usage information.

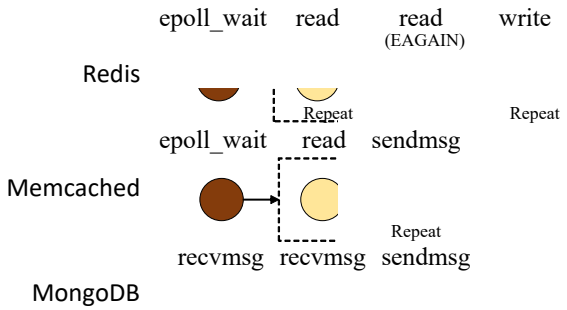


Figure 2. System call sequence model of Redis, Memcached, and MongoDB

Figure 3. Resource consumption result of Insert operation

Figure 4. Resource consumption result of Select operation

Figure 5. Resource consumption result of Delete operation

SystemTap, one of those, is a tool that traces the Linux kernel activities without recompiling the kernel. It allows users can specify various actions for events and even add custom helper functions. In the current implementation, SystemTap is selected as a tool for system call tracing and resource usage accounting. We have added the CPU and memory accounting logic into the SystemTap module.

Ⅲ. Evaluation

We chose three NoSQL databases (Redis, Memcached and MongoDB) as the targets of resource model construction. Depending on our key objectives of achieving the explainability, we present here the modeling results of three NoSQL databases produced by our framework and provide descriptions of internal behaviors. Figure 2 shows the system call sequences we discovered for three NoSQL databases. We have applied three operations (insert, delete, and select) to all three. However, the system call patterns were identical for three operations within the same database. Figure 3, 4, and 5 describes the resource consumption amount for unhalted cycle and retired instruction of each three operations constructed by our technique.

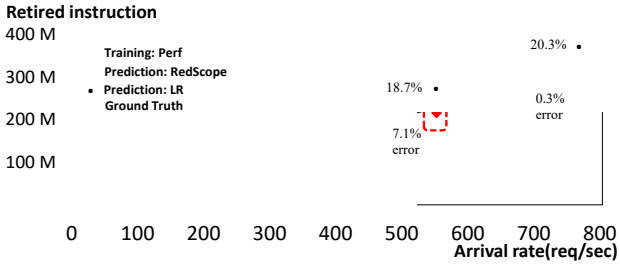


Figure 6. Accuracy of our technique and the LR(Linear Regression) model on MongoDB insert operation

Figure 6 shows the instruction consumption according to each operation predicted by ours and their actual observations. The experiment was designed to predict the resource usage used by the target NoSQL database when a series of requests are issued by varying the arrival rate from the client. For comparison, we set up a simple linear regression model based on the values collected from the Perf tool. The retired instruction model for MongoDB insert operation was evaluated. In the graph, dotted lines show trends obtained from linear regression model. In MongoDB, as the request frequency increases, the inclination of resource usage becomes less stiff, unlike the linear regression result.

Ⅳ. Conclusion

In this work, we have demonstrated the feasibility of the grey-box approach based on the system call traces to building the resource consumption models of NoSQL databases. Using the proposed framework, we investigated three NoSQL databases to gain new insights and analyzed their scalability potentials. Our findings revealed that NoSQL databases all have unique and complex internal operations that affect their overall scalability.

REFERENCES

[1] P. Malakar, P. Balaprakash, V. Vishwanath, V. Morozov, and K. Kumaran. Benchmarking machine learning methods for performance modeling of scientific applications. In 2018 IEEE/ACM Performance Modeling, Benchmarking and Simulation of High Performance Computer Systems (PMBS), pages 33-44, 2018.

[2] C. Witt, M. Bux, W. Gusew, and U. Leser. Predictive performance modeling for distributed batch processing using black box monitoring and machine learning. Information Systems, 82:33-52, 2019.

연안 파랑 관측 및 분석을 위한 비지도 학습 기법 연구

김재일, 김태경*, 허동**, 박성준, 김진아*
경북대학교, *한국해양과학기술원, **대구경북과학기술원
threeyears@gmail.com, paperrune@naver.com, hd901210@dgist.ac.kr,
sjp611@naver.com, jakim@kiost.ac.kr

A Study on Unsupervised Learning Methods for Coastal Wave Monitoring and Analysis

Jaeil Kim, Tae Kyung Kim*, Dong Huh**, Sung Joon Park, Jinah Kim*
Kyungpook National Univ., *Korea Institute of Ocean Science and Technology,
**Daegu Gyeongbuk Institute of Science and Technology

요 약

연안에서의 파랑 관측 및 분석은 파랑의 동역학적 모델링과 연안 재해 예측에 필수적으로 요구되고 있으나, 파랑 현상의 비선형성과 날씨 변화 등에 의한 데이터의 왜곡, 장시간 수집되는 방대한 데이터에 대한 레이블링 어려움 등에 의해 인공지능 기술의 적용이 어려운 분야이다. 본 논문은 연안 데이터로부터 파랑의 정확한 관측 및 분석을 위한 비지도 학습 기반 신경망 기술과, 실제 국내 연안에서 수집된 연안 관측 비디오 데이터에 대해 수행한 분석 실험을 소개한다.

Ⅰ. 서 론

연안은 바다와 육지가 맞닿아 서로 밀접한 영향을 주는 공간으로 이용 집약도가 높은 지역이지만 동시에 해일, 월파 등의 파랑에 의한 연안 재해에도 취약한 지역이기도 하다. 이에 따라 정확한 파랑 관측과 모델링을 통해 파랑의 동역학적 특징을 이해하고, 나아가 기후나 해양 환경 변화에 따른 파랑 변화를 예측하고자 하는 연구가 활발히 진행되고 있다. 하지만 연안 구조에 의한 파랑 현상의 높은 비선형성과 센서를 이용한 직접 관측의 어려움에 의해 연안에서의 파랑 관측과 모델링은 여전히 어려운 과제이다 [1,2].

연안의 연속적 파랑 관측을 위하여 CCTV 비디오나 레이더 영상 등 원격 탐사 기술을 이용한 연구가 활발히 진행되고 있는데, 넓은 연안 영역에 대한 장시간의 시공간 데이터 확보가 유리하고, 관측 기간 중 유지 보수가 용이하기 때문이다. 하지만 비디오나 레이더 등 시공간 데이터가 매우 방대하고, 날씨 변화 등 외부 요인에 따른 데이터 왜곡이 발생하여 정확한 관측을 어렵게 한다. 연안 비디오 관측 데이터를 이용한 파랑 관측을 위하여, 기존 연구에서는 Optical Current Meter[3], Particle Image Velocimetry[4] 등 영상 처리 기법을 이용한 연구들이 제안되었으나, 배경에 대한 사전 정의가 요구되는 등 실제 파랑 영상 분석에 적용되기에 제한점을 가지고 있다.

보다 최근에는 학습 기반 기술을 이용하여 파쇄 지점을 검출하는 기법[5]이 제안되었으나, 이에 대한 레이블을 요구하는 지도학습 기반이어서 학습 데이터 구축에 대한 비용에 의해 다양한 데이터를 학습하기 어렵다는 단점이 있다. 본 논문은 방대한 연안 데이터로부터 파랑 움직임을 추적하고, 이를 분석하기 위한 비지도 학습 기반 파랑 분석 기술을 소개한다. 특히, 연안 CCTV 관측에서 흔하게 발생할 수 있는 비디오 영상 왜곡 현상을 해결하기 위한 화질 개선 기술[6], 파랑의 움직임 영역만 분리하기 위한 동적 영역 분리와 시간에 따른 파랑 움직임을 연속적으로 추적하기 위한 영상 정합을 수행하는 신경망 [1] 등을 소개한다. 제안 기술들은 국내 연안에서 수집된 실제 관측 영상을 대상으로 기존 기술에 비해 파랑 추적에서 높은 정확도를 보였다.

Ⅱ. 본론

본 논문에서 소개하는 연안 파랑 관측 및 분석 기술은 연안 CCTV 영상 왜곡 보정을 위한 적대적 학습 기반 신경망과 연안 파랑 추적을 위한 장면 분리 및 프레임 간 영상 정합 신경망으로 구성된다. 먼저, CCTV 왜곡 보정 신경망[6]은 빗물 등에 의한 CCTV 렌즈의 부분적 왜곡을 보정하고 배경에 있는 파랑 모습을 복원하기 위하여, 왜곡 영상과 깨끗한 화질의 정상 영상을 동시에 학습하여 정상 영상처럼 복원하는 Generator과 생성 영상과 정상 영상을 구분하는 Discriminator로 구성된다. 특히, Generator는 왜곡 영역에서만 새롭게 배경을 복원하기 위하여, 정상 영상과 다른 특징을 갖는 부분을 표시하는 Mask 영상을 동시에 생성하며, Mask를 이용해 입력과 생성 영상을 합성하여 최종 생성 영상으로 제공한다. Generator는 Discriminator와의 적대적 학습을 통해 정상 영상 데이터와 생성 영상 간 분포를 동일하게 하도록 학습함으로써 왜곡 보정 영상을 생성한다.

연안 파랑 추적 기술[1,2]은 연속된 비디오 입력 영상으로부터 움직임이 없는 배경 영상과 움직임을 포함하는 전경 영상을 분리하는 GAN 모델과 전경을 나타내는 연속된 Mask 영상을 시간 순서에 맞추어 정합하여 파랑 움직임을 표현하는 Displacement Vector Map으로 생성하는 영상 정합 모델로 구성된다. 이후 파속 측정, 파쇄 분석 등을 위해 영상 위에 배치한 Particle을 연속된 Displacement Vector에 따라 변화된 위치를 계산하여 시간에 따른 파랑의 움직임을 추적한다. 이러한 방법을 통해 파랑 관측을 위해 장기간 획득되는 CCTV 영상에서 별도의 파랑 위치에 대한 레이블링 없이, 연안 영역 전체에 대한 파랑 움직임을 쉽게 정량화할 수 있다. 제안 기술은 수조 실험에서 수집된 CCTV 영상을 대상으로 한 실험에서 센서를 이용해 측정된 파속에 대응되는 높은 정확도를 보였으며, 실제 안목항에서 수집된 CCTV를 이용한 파랑 분석에서도 관측 자료와 대응되는 파속 측정 결과를 보였다 [1,7].

Ⅲ. 결론

본 논문에서는 CCTV 왜곡 보정과 영상 분리 및 정합 기술을 이용한 연안 파랑 관측 기술을 소개하였다. 해당 기술은 날씨 변화 등에 대해 보다 강직하게 파랑의 연속적인 움직임을 정량화할 수 있으며, 이를 통해 연안에서의 파랑 변화에 대한 정확한 관측 결과를 제시함으로써 연안 환경 변화에 대한 관측 및 예측에 대한 연구 기반이 되고 있다.

ACKNOWLEDGMENT

이 연구는 한국해양과학기술원 주요사업인 'AI기반 파랑기인 연안재해 모델링 플랫폼 및 해무 예측기술 개발' 과제와 2020년도 정부(교육부)의 재원으로 한국연구재단 기초연구사업의 지원(2020R1I1A3074639)으로 수행되었습니다.

참 고 문 헌

[1] Jinah Kim, et al. "Wave-tracking in the surf zone using coastal video imagery with deep neural networks." Atmosphere 11.3 (2020): 304.

[2] Jinah Kim, and Jaeil Kim. "Estimation of Water Surface Flow Velocity in Coastal Video Imagery by Visual Tracking with Deep Learning." Journal of Coastal Research 95.SI (2020): 522-526.

[3] C. Chickadel, R. A. Holman, and M. H. Freilich, "An optical technique for the measurement of longshore currents," Journal of Geophysical Research: Oceans, vol. 108, no. C11, 2003.

[4] K. Holland, J. Puleo, and T. Kooney, "Quantification of swash flows using video-based particle image velocimetry," Coastal Engineering, vol. 44, no. 2, pp. 65 - 77, 2001.

[5] Stringari, C. E., D. L. Harris, and H. E. Power. "A novel machine learning algorithm for tracking remotely sensed waves in the surf zone." Coastal Engineering 147 (2019): 149-158.

[6] Dong Huh, et al. "Raindrop-aware GAN: Unsupervised learning for raindrop-contaminated coastal video enhancement." Remote Sensing 12.20 (2020): 3461.

[7] 김진아, 김재일, and 신성원. "비디오 영상의 비지도 영상 정합을 이용한 파랑 이동 속도 추정." 정보과학회논문지 46.12 (2019): 1296-1303.

복부암 다중 오믹스 유전체 데이터 기반의 생물학기작 활성화 정도 분석 방법

전재민, 정인욱*
IT 대학 컴퓨터학부

경북대학교

jaemin1479@knu.ac.kr, inukjung@knu.ac.kr*

Multi Omics Pathway Activation Analysis

Jae Min Jeon, Inuk Jung*
School of Computer Science and Engineering
Kyungpook National University

요 약

본 논문에서는 PARAFAC 을 이용한 텐서 분해기법을 활용하여 다중 오믹스 유전체 데이터 기반 유전자 패스웨이 활성화 정도를 측정하기 위한 방법을 제안한다. 단일 생물체로부터 다양한 유전체 정보가 존재하는데 단일 오믹스에 비해 다중(또는 복수) 오믹스를 활용했을 때 복부암의 기작을 더 잘 설명할 수 있음을 보였다. 또한, 본 방법을 활용하여 복부암 환자들의 임상 변수를 라벨로 정의하여 이들에 대한 분류 정확도가 다른 방법들에 비해 높은 성능을 가졌음을 보였다.

Ⅰ. 서 론

바이오의학 기술의 발전으로 생물체로부터 다양한 종류의 오믹스(omics) 데이터와 세부정보를 수집할 수 있게 되었다. 오믹스는 유전체 정보를 뜻하며 여기에는 유전자 (e.g., DNA, RNA), 단백질, 대사체 등의 다양한 생체물질을 포괄한다. 기존에는 단일 오믹스를 통하여 생물학적 복잡성을 설명하였다면 다중 오믹스를 사용했을 때는 보다 다양한 관점에서 이를 설명 할 수 있게 된다. 환자들의 생물학적 복잡성의 정보를 패스웨이(pathway)를 통하여 설명을 할 수 있다. 오믹스 데이터를 통해 환자들의 Gene set analysis (GSA)는 오믹스 데이터를 통하여 특정 유전자들의 특징을 발굴하고 패스웨이를 통하여 분석하는 기법이다. 패스웨이는 생체 활동에서 중요한 경로에 해당되는 유전자들을 각각 모아 놓은 유전자 집합이라고 할 수 있다. 기존에는 하나의 오믹스를 사용한 single sample Gene Set Enrichment Analysis (ssGSEA)[1], Gene Set Variation Analysis(GSVA)[2], Zscore 등의 기법이 있다. 다중 오믹스를 사용하여 패스웨이에 대하여 하나의 샘플 당 스코어를 구하는 경우는 Multi Omics Gene set Analysis(MOGSA)[3]가 있다. Multi Factor Analysis(MFA)[4]를 사용한 MOGSA는 다중 오믹스를 합하여 패스웨이 점수를 계산한다. 그러나 MOGSA 에서는 패스웨이 스코어를 통하여 subtype 분류 성능이 부족하였고 PCA (Principal component analysis) 사용하여 설명을 잘하는 component 만 뽑아서 하나의 오믹스에 편중화 되는 경향이 있다. 본 논문에서는 mRNA 발현량, microRNA 발현량, DNA methylation level 의 오믹스 데이터를 사용하여 패스웨이 활성화를 분석한다. 본 논문은 3 가지 오믹스를 텐서화를 통해 유전자들을 서열화해서 각 환자의 패스웨이 활성화를 분석한다. 이 활성화 점수를 통해 각 임상정보마다 분류 정확도와 클러스터링 정도를 기존 기법들과 비교하여 성능을 보여준다.

Ⅱ. 본론

2.1 데이터 및 정규화

데이터는 암유전체지도 TCGA (The Cancer Genome Atlas)[5] 프로젝트에 있는 복부암(Stomach adenocarcinoma, STAD)에 해당되는 환자 유전자 정보를 사용하였다. 복부암 환자에 대한 7 가지 임상 타입 데이터를 사용한다 (표 1).

표 1. 복부암 임상타입에 따른 데이터

임상타입	샘플 수	아형	유전자 개수
Hypermethylation category	311	[CIMP-H, CIMP_EBV, GEA_CIMP-L, Non-CIMP]	14,454
Hypermutated	312	[0,1]	
Molecular subtype	305	[MSI-H,MSI-L,MSS]	
Pathologic_N	312	NO,N1,N2,N3,N3a	
Pathologic_M	299	M0,M1,Mx	
Stage	310	'I', 'II', 'III', 'IV'	
ABSOLUTE	396	['AC', 'AF', 'SNP']	

오믹스 데이터는 mRNA, miRNA, methylation level 오믹스를 사용한다. 유전자 집합 데이터는 Reactome Pathway Database[6]에서 암과 관련된 300 개 데이터를 사용한다. 3 개의 오믹스 원시 데이터를 quantile 정규화를 이용하여 정규화를 한다

2.2 다중 오믹스 패스웨이 분석 모델

본 논문에서는 텐서 분해를 통한 다중 오믹스 유전자 집합 스코어 모델을 제안한다. 이 모델은 크게 텐서 분해와 그 텐서를 사용한 패스웨이 별 각 샘플의 점수를 계산하는 과정으로 구성한다.

$$\mathrm{T} = \sum_{f=1}^{R} g_f \otimes p_f \otimes o_f \qquad (1)$$

(1)번 수식에서의 유전자 텐서 g_f, 샘플 텐서 p_f, 오믹스 텐서 o_f, 3 개의 텐서로 하나의 T 가 구성되는 것을 볼 수 있다. $\otimes$ 수식은 행렬 곱셈이다. 3 개의 정보가 섞인 값으로 분해된 각 텐서를 이용하여 패스웨이 활성화 점수를 구한다. 앞의 수식에서 만든 텐서들과 패스웨이 바이너리 데이터를 입력 값으로 받는다. 본 모델 에서 오믹스별 텐서값은 유전자 $g_{(i)} = \{g_{ji}, \dots, g_{ji}\}$, 환자 텐서 $s_{(k)} = \{s_{jk}, \dots, s_{jk}\}$으로 표시한다. j 는 랭크, i는 유전자, k 는 샘플을 뜻한다. $g_{(i)}$는 평균 0 표준편차 1 의 표준화를 한다. 그리고 음수 값을 제거하기 위하여 전체 텐서에서 최소값을 뺀 모든 텐서 값이 양수가 되도록 한다. 환자 텐서 값에서 각 랭크마다 높은 값을 가지는 환자들을 뽑는다. 이 기준은 가우시안 커널 함수를 통하여 최빈값을 가지는 샘플 이상의 샘플 텐서들을 선택한다. 한 환자는 여러 랭크에 속하는 것을 볼 수 있고 (2)번과 같이 각 랭크에 해당되는 유전자 텐서 값과 환자 텐서 값을 곱하여 더한다. (2) 수식으로 다중 오믹스를 통합한 환자에 따른 유전자 중요도가 구해지며 이를 유전자 서열화를 하는데 사용 한다.

$$r_{(j)} = \sum S_{(j)} \times g_{(j)} \qquad (2)$$

유전자 서열화 데이터 $r_{(j)}$ 를 GSVA 에 사용된 Kolmogorov-Smirnov (KS) like random walk statistic[1]를 변형하여 각 환자 당 패스웨이의 스코어를 계산한다. 기존의 KS 에서는 $r_{(j)}$ 의 순서를 사용하였지만 본 모델에서는 $r_{(j)}$ 값 자체의 차이를 이용한다.

$$E_{jt} = \frac{\sum_{j=1}^{l} r_{(j)} I(G_{(j)} \in p_{(t)})}{\sum_{j=1}^{p} r_{(j)} I(G_{(j)} \in p_{(t)})} - \frac{\sum_{j=1}^{l} r(j) I(G_{(j)} \notin p_{(t)})}{|G_{(j)}| - |p(t)|} \qquad (3)$$

$p_{(t)}$ 는 유전자 집합이고 $G_{(j)}$ 는 입력 값이 되는 유전자들이다. $I(G_{(j)} \in p_{(t)})$ 은 지시 함수로 패스웨이에 해당 유전자가 포함되는지를 표시하는 함수이다. 속한다면 1, 아니면 0 의 값을 가지게 된다.

$$\mathrm{diff} = \left|E_{jt}^{+}\right| - \left|E_{jt}^{-}\right| \qquad (4)$$

(3)번 수식에서 나오는 값들 중에 최대값 (E_{jk}^{+}) 과 최소값(E_{jk}^{-})의 차이를 (4)번 수식에 대입을 한다. + 와 - 는 패스웨이에 속하는 유전자들이 아닌 유전자들 보다 높은 값을 가지는 여부에 대한 수식이다. (4)번 수식의 결과가 한 유전자 집합의 한 샘플의 스코어이다. 스코어의 값은 -1~1 사이며 높을수록 유전자 집합에 해당되는 유전자의 중요도가 높음을 알 수 있다.

2.3 실험 결과

멀티 오믹스를 활용한 패스웨이 스코어가 실제 데이터에서도 아형에 따른 분류를 도와주는지 알아보았다. 이를 위해 복부암의 7 가지 임상타입에 따른 데이터를 사용하여 분석을 하였다.
그림 1 은 TCGA 에서 제공하는 STAD 의 7 가지 clinical data 에 따라 활성화 스코어가 얼마나 잘 구분하는지를 레이더 차트로 표시를 하였다.

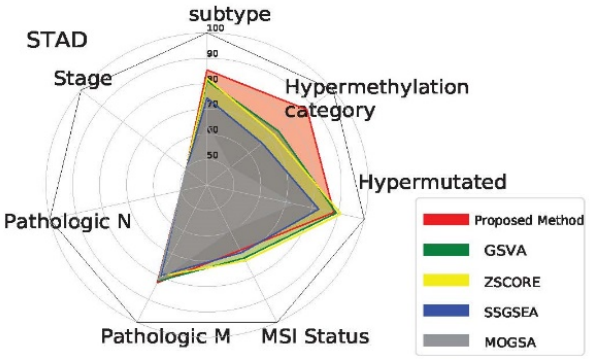


그림 1. STAD 임상정보에 따른 모델들의 MLP f1-score

Multilayer perceptron(MLP) 분류기를 사용하여 10 교차검증을 하였고 테스트 샘플과 훈련 데이터를 9:1 로 나누어서 f1-score 계산을 하였다. f1-score 를 사용한 이유는 아형 마다 샘플 수의 차이가 많이 차이 났다. 본 모델이 다른 모델들과 비슷한 정확도를 가지지만 subtype, Hypermethylation category 임상에서는 f1-score 가 높은 것을 볼 수 있다. 새로운 모델은 methylation 오믹스를 사용하였기 때문에 mRNA 만 사용하는 기법들보다는 높은 분류 성능을 보였다. 그러나 같은 오믹스들을 사용한 MOGSA 보다도 정확도가 높은 것을 볼 수 있다. 그림 2 의 (B) hypermethylation category 에서 클러스터링이 어느정도 되는지를 ARI(Adjusted Rand Index)와 AMI(Adjusted Mutual information)을 사용하여 보여준다.

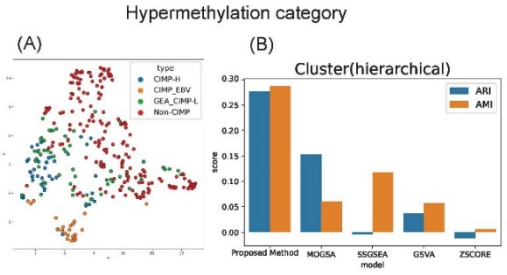


그림 2. Hypermethylation category 에 대한 (A) tsne_plot (B) 클러스터링 스코어

본 논문에서 제안하는 모델이 다른 기존 모델 보다 높은 점수를 가지는 것을 볼 수 있다. 그리고 본 본문의 모델의 활성화 스코어의 TSNE-plot 으로 얼마나 잘 구분되는지를 확인한다.

Ⅲ. 결론

본 논문에서는 질병을 가진 환자들의 패스웨이 분석을 하기 위하여 텐서 기반 다중 오믹스 패스웨이 활성화 기법을 소개한다. 각 환자에 따라 값이 높은 텐서를 추출하고 패스웨이에 따른 생물학 기작을

활성화 정도를 구하기 때문에 아형끼리 비슷한 활성화 정도를 가지며 클러스터링 성능이 뛰어났다. 이를 복부암 환자 데이터셋을 사용하여 성능 평가한 결과 기존 기법보다 높은 성능을 보이는 것을 볼 수 있었다.

ACKNOWLEDGMENT

이 논문은 2020 년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의지원을 받아 수행된 기초연구사업임(No. 한국연구재단에서 부여한 과제번호: 2020M3C9A5085604)

참 고 문 헌

[1] Barbie, David A., et al. “Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1,” *Nature* 462.7269(2009):108-112.

[2] Hanzelmann,Sonja,Robert Castelo,and Justin Guinney.”GSVA : gene set variation analysis for microarray and RNA-seq data,” *BMC bioinformatics* 14.1 (2013): 1-15

[3] Meng, Chen, et al. “MOGSA: integrative single sample gene-set analysis of multiple omics data,” *Molecular &Cellular Proteomics*, 18(8), S153-S168

[4] Abdi,Herve,Lynne J. Williams, and Domininique Valentin. “Multiple factor analysis: principal component analysis for multitable and multiblock data sets,” *Wiley Interdisciplinary reviews: computational statistics* 5.2(2013):149-179

[5] Tomczak,Katarzyna,Patycja Czerwiriska,and Maciej Wizenerowicz.”The Cancer Genome Atlas(TCGA): an immeasurable source of knowledge,” *Contemporary oncology* 19.1A(2015).A68

[6]Fabregat, Antonio,et al.”The reactome pathway Knowledgebase,”*Nucleic acids research*, 46.D1(2018):D649-D655.

단기 시계열 데이터에 대한 장기간 다단계 예측을 위한 학습 방법 연구

김동현, 도주성, 정호영*
경북대학교

kdh1477@knu.ac.kr, jsdo@knu.ac.kr, *hoyjung@knu.ac.kr

A study on learning method for multi-step long-term forecasting using short-term time series data

Dong-Hyun Kim, Ju-Seong Do, Ho-Young Jung*
Kyungpook National Univ.

요 약

본 논문은 기존 시계열 데이터 예측의 문제점을 해결하기 위해 단기간 (1 일)의 데이터를 활용하여 장기간 (1 달)의 데이터를 예측하는 연구에 대해 소개한다. 기존의 시계열 데이터 예측에서 단기간에 대해서는 준수한 성능을 보이지만, 장기간은 거의 예측하지 못하는 문제점이 있다. 이를 해결하기 위해, 본 논문에서는 단기간 예측 성능이 우수하다는 점을 활용하여 이를 연속으로 수행함으로써 보다 나은 장기간 예측이 가능한 방법을 제안한다. 장기간 예측에서 제안된 방법이 기존의 예측 방법보다 목푯값을 더 잘 추적하는 경향을 보인다. 본 논문은 제안된 학습 방법이 단기간의 시계열 데이터를 활용한 데이터 예측 모델 개발에 응용될 수 있을 것으로 기대한다.

Ⅰ. 서론

최근 인공지능을 기반으로 시계열 데이터 예측을 위한 기술들이 개발되고 있다. 기존의 시계열 데이터 예측에 사용된 모델은 순환 신경망인 RNN (Recurrent Neural Network)이다. RNN 의 구조를 가지고 있는 대표적인 모델에는 LSTM[1] (Long Short-Term Memory), GRU[2] (Gated Recurrent Units)등이 있다. 위 모델들은 장기간 데이터 학습에 어려움이 있고 학습 시간이 오래 걸리는 등의 단점이 있다. 이후 등장한 Transformer[3] 모델은 시계열 데이터의 병렬처리로 RNN 의 학습 시간 문제를 해결하였다.

본 논문에서는 Transformer 모델 기반의 Informer[4] 모델을 LSTF (Long Sequence time-series forecasting) 태스크에 활용한다. Informer 모델은 기존의 시계열 데이터 예측 모델들보다 더 긴 sequence 를 예측하는 성과가 있지만, 매우 긴 sequence 를 예측하는 데는 기존의 모델과 비슷한 성능을 보인다.

위의 문제를 해결하기 위해, 본 논문에서는 단기간 예측이 우수하다는 점에 주목하여 장기간을 예측하는데 단기간 예측을 그 다음 단기간 예측에 활용하여 연속적으로 추론하는 방법을 시도한다. 이러한 장기간 다단계 예측을 학습에도 적용하고 개선된 장기간 예측의 결과를 보여준다.

Ⅱ. 본론

본 장에서는 실험에 사용된 데이터셋 (ETDataset), 모델 (Informer), 장기간 다단계 예측을 위한 학습 방법에 대해 소개한다.

Ⅱ-Ⅰ. ETDataset (Electricity Transformer Dataset)

본 논문에서 LSTF 태스크에 사용한 학습데이터는 ETDataset 으로 실제 중국에서 2 년간 2 곳의 Electricity Transformers 에서 수집된 데이터이다. ETDataset 은 1 시간 단위인 {ETTh1, ETTh2}와 15 분 단위인 {ETTm1, ETTm2}로 나뉘어 있다. 각 데이터의 변수는 총 7 개로 부하특성인 HUFL (High UseFul Load), HULL (High UseLess Load), MUFL (Middle UseFul Load), MULL (Middle UseLess Load), LUFL (Low UseFul Load), LULL (Low UseLess Load)와 목푯값인 OT (Oil Temperature)로 구성되어 있다.[1]

Ⅱ-Ⅱ. Informer

Informer 모델은 기존의 Transformer 모델에서 사용하던 self-attention 을 ProbSparse self-attention 으로 바꿔 연산량을 줄여 좀 더 빠른 학습이 가능하게 한다. Transformer 의 self-attention 은 key 와 query 의 내적값을 모두 이용하여 연산을 수행하는데, Informer 의 ProbSparse self-attention 은 key 와

[1] https://github.com/zhouhaoyi/ETDataset

query의 내적값 중 큰 값만을 가려내어 연산을 수행한다. 또한, Informer 는 Transformer 와 달리 convolution layer 를 추가하여 데이터의 특징을 좀 더 잘 찾아내도록 한다. 결과적으로, Transformer 가 N^2 만큼의 시간과 메모리를 소비하며 학습할 때, Informer는 Nlog(N)만큼의 시간과 메모리를 소비하며 학습한다. (N은 sequence 의 길이를 의미)

Ⅱ-Ⅲ. 장기간 다단계 예측을 위한 학습 방법

본 논문은 단기간 미래 예측 모델을 연속적으로 적용하여 장기간 미래 예측시스템을 위한 WFO (Walk-Forward Optimization) 기반의 학습 방법을 제안한다. 이를 위해 Informer의 criterion function을 개선한다. 식 (1)의 L_1 은 기존의 Informer 에 사용된 loss 이다. 즉, 실제 과거 1 일의 데이터로 미래 1 일의 데이터를 찾는 모델의 loss 이다. WFO 를 위한 L_2 는 임의의 1 일 데이터에 대한 미래 1 일 예측값을 다시 입력으로 하여 다음 미래 1 일 데이터를 예측하는 criterion 이다.

$$\mathbf{Loss} = \boldsymbol{\lambda L_1 + (1-\lambda) L_2} \quad \textbf{(1)}$$

실험에 사용된 데이터셋은 ETDataset 의 ETTh1 의 첫 번째 변수인 HUFL 이다. 실험은 NVIDIA Quadro RTX 24GB GPU 로 진행되었으며, 학습률은 1e-4, batch size 는 24, λ는 0.5, 옵티마이저는 Adam 을 사용하였다. 평가는 각 모델에 대해 30 번씩 수행되었으며 실험 결과에 기술된 MSE (Mean Square Error)는 30 번 실험의 평균치이다.

Ⅱ-Ⅳ. 실험 결과

제안하는 학습 방법의 성능을 평가하기 위해, 기존의 방법으로 학습된 모델의 추론 결과와 제안된 방법으로 학습된 모델의 추론 결과를 비교한다.

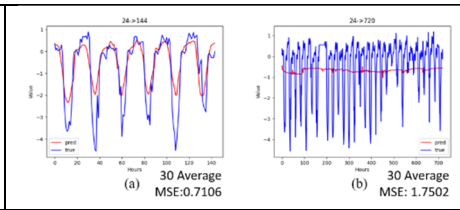
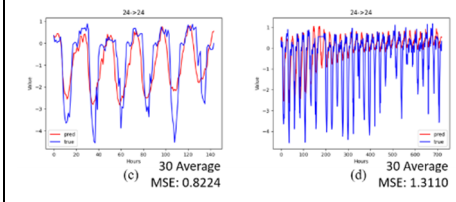
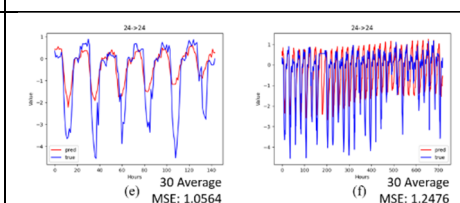
과거 1 일 →미래 N 일 예측 모델	
과거 1 일 →미래 1 일 예측 모델 (N 일 다단계 적용) 기존 학습법 적용	
과거 1 일 →미래 1 일 예측 모델 (N 일 다단계 적용) 제안된 학습법 적용	

그림 1 : 추론 결과 비교.

그림 1 의 (a, b)는 sequence 길이에 따른 Informer 성능 비교 결과이다. 그림 1 의 (a)는 1 일치 데이터로 미래의 6 일치 데이터를 예측한 결과이고, (b)는 1 일치 데이터로 미래의 30 일치 데이터를 예측한 결과이다. 그림 1 의 (c, d)는 1 일치 데이터로 다음 1 일치 데이터를 예측하고 예측한 데이터로 그 다음 1 일치 데이터를 연속적으로 예측한 결과이다. 그림 1 의 (e, f)는 장기간 다단계 예측 방법을 학습에도 적용하고 새롭게 학습된 모델로 장기간을 다단계 예측하는 LSTF 결과이다. 즉, 그림 1 의 (a, c, e)는 총 6 일의 기간을 예측한 것이고, 그림 1 의 (b, d, f)는 총 30 일의 기간을 예측한 것이다.

6 일의 기간을 예측할 때는 그림 1 의 (a, c, e)에서 볼 수 있듯이 기존의 모델로 한 번에 6 일을 예측한 결과와 장기간 다단계 예측 모델로 6 일 동안 예측한 결과가 유사하다. 30 일의 기간을 예측할 때는 그림 1 의 (b)와 같이 기존의 모델로 한 번에 30 일을 예측한 결과는 평균값을 추적하는 경향이 있지만, 장기간 다단계 예측 모델로 30 일 동안 예측한 결과는 그림 1 의 (d, f)에서 보이듯이 변화를 지속적으로 추적하고 있음을 확인할 수 있다.

Ⅲ. 결론

본 논문에서는 시계열 데이터의 장기간 예측에 집중하였다. 이를 위해 기존 연구와 달리 시계열 데이터의 연속적인 추론을 통해서 LSTF 태스크를 수행하는 다양한 실험을 하였다. 30 일의 장기간을 예측할 때 기존의 예측 모델은 평균값을 추적하는 경향이 있다. 반면, 제안된 장기간 다단계 예측법은 변화를 따라감과 동시에 이를 학습에 적용하면 오차 또한 줄어드는 것을 확인할 수 있다. 이러한 결과는 시간의 흐름을 가진 모든 데이터에 충분히 적용될 수 있을 뿐만 아니라 주가 예측, 설비 예지보전, 날씨 예측 등에도 적용될 수 있을 것으로 기대된다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT 연구센터지원사업의 연구결과로 수행되었음 (IITP-2021-2020-0-01808*)

참 고 문 헌

[1] Hochreiter, Sepp, and Jürgen Schmidhuber. "Long short-term memory." Neural computation 9.8 (1997): 1735-1780.

[2] Cho, Kyunghyun, et al. "Learning phrase representations using RNN encoder-decoder for statistical machine translation." arXiv preprint arXiv:1406.1078 (2014).

[3] Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems. 2017.

[4] Zhou, Haoyi, et al. "Informer: Beyond efficient transformer for long sequence time-series forecasting." Proceedings of AAAI. 2021.

[5] Chung, Junyoung, et al. "Empirical evaluation of gated recurrent neural networks on sequence modeling." arXiv preprint arXiv:1412.3555 (2014).

Lightweight-VGG based Facial Landmarks Detection with Inception-ResNet Module

Savina Jassica Colaco, MyungChul Park, DaeWoong Cha, ChaeHyun Lee and Dong Seog Han*
School of Electronics and Electrical Engineering
Kyungpook National University
Daegu, Republic of Korea
savinacolaco@knu.ac.kr, mcpark@knu.ac.kr, dwcha@knu.ac.kr, hyeu330@knu.ac.kr, dshan@knu.ac.kr*

***Abstract*—Facial landmark detection has achieved great performance by learning discriminative features from the rich deformation of face shapes and poses. The landmarks such as eye centres, nose centres, jawlines, etc are localized to give vital information to computer vision-related applications. The paper proposes a modified light-weight VGG based model for predicting facial landmarks on the detected faces from digital images or video.**

***Index Terms*—facial landmarks, convolutional neural networks, inception-ResNet**

I. Introduction

Facial landmarks detection locates a set of vital points in the digital images of faces taken in unconstraint conditions. This important information is used in different applications such as 3D facial structure estimation, attributes estimation or facial expression recognition [1]. Since convolutional neural networks (CNNs) outperform conventional methods in different applications, CNN also shows significant performance in the detection of facial landmarks. Recent facial landmark detection approaches mainly focus on features from face shapes, poses, different expressions, occlusions and other conditions. A simple facial landmark detection model uses CNN to detect and predict landmarks simultaneously. In this paper, simple CNN is trained to predict 68 facial landmarks and then mapped onto the detected faces in real-time.

II. Experiment

A. Facial landmarks detection model

The VGG16 [2] architecture acts as a baseline model for predicting 68 facial landmarks. The convolution layers are replaced with depthwise convolution layers which apply a single convolutional filter to each input channel. An inception-ResNet module is introduced between the depthwise layers to provide a different level of feature extraction at different scales. The Inception-ResNet , as shown in Figure 1, is highly adjustable giving several possibilities to change the number of filters in the layers. The different number of filters such as 1×1, 3×3 and dilated filters are used to extract features. This helps to focus on different regions of the face images to predict facial landmarks. A skip connection is presented to perform identity mapping where the original input features are added to the outputs of the stacked layers. Each layer is followed by batch normalization and Hswish activation. The Hswish activation function replaces the expensive sigmoid with its piece-wise linear in swish which could be a disadvantage for mobile devices. Figure 2 depicts the model design of facial landmarks detection.

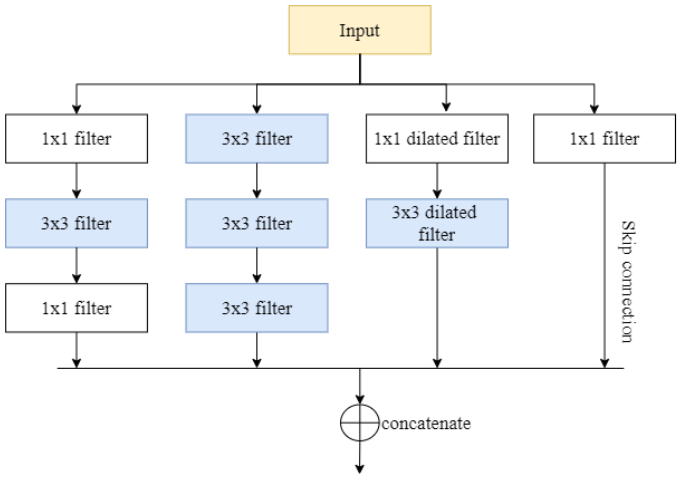


Fig. 1. Inception-ResNet module.

B. Implementation details

The facial landmarks model is trained with a combined dataset of 300W [3] and 300VW [4]–[6] with a total number of 112,111 images. The 300W dataset comprises AFW, HELEN, LFPW, XM2VTS, and IBUG datasets where images are annotated with 68 landmarks. The images in the dataset are resized to 112×112 resolution in grayscale. The Keras framework is used for model implementation and trained with a batch size of 32 and epochs of 50. The model is continuously optimized with the Adam optimization technique with a learning rate of 10^{-3}. For the model training, mean squared error (MSE) is used between the ground truth keypoint coordinate vector and the predicted one.

C. Results

The faces are detected with the ResNet-single-shot face detector (SSD) from images or video. The SSD [7] is faster than Faster R-CNN since it does not need an initial object proposals

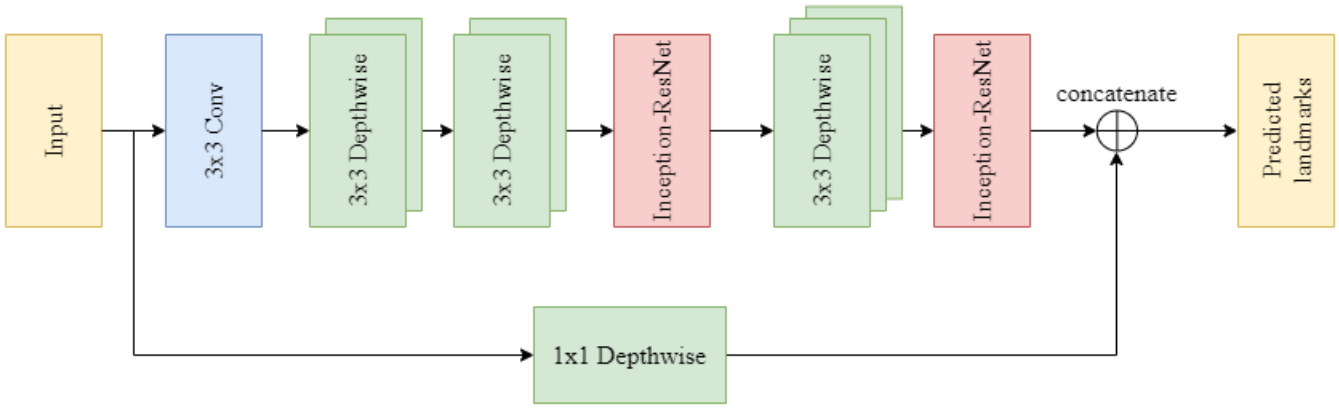


Fig. 2. Facial landmarks detection model design.

generation step. The model achieves a prediction accuracy of 71% with 16M parameters. The predicted landmarks are detected approximately well for frontal face and small head pose images. The facial landmarks are not detected well with extreme head poses such as yaw, pitch and roll conditions. The landmarks are detected well around the eye and nose area due to the inception-ResNet module that concentrates on important regions of the face. This information can be useful in the application such as driver monitoring and assistance systems that require driver's status information. The normalized mean error (NME) on the 300W dataset for the model achieves 6.5 on the full-set, 18.33 on the challenging subset and 7.2 on the common subset. The overall NME is greater than the challenging subset with contains extreme conditions of head poses, illumination and occlusion. Figure 3 shows the real-time facial landmarks detection with various head poses.



Fig. 3. Facial landmarks detection in real-time.

III. Conclusion

This paper predicts facial landmarks using VGG16 as a baseline model. It is replaced with the depthwise convolutions layers and an inception-resnet module to make the model light and extract a different level of features. The model still suffers from extreme head poses and illumination conditions. The future work is focused on improving the model to make it robust to different imaging conditions.

Acknowledgment

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2021-2020-0-01808) supervised by the IITP (Institute of Information & Communications Technology Planning & Evaluation) and Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education(2019R1D1A3A0310384913).

References

[1] Shi, S., Facial Keypoints Detection. ArXiv 2017, abs/1710.05279.
[2] Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. CoRR 2015, abs/1409.1556.
[3] Sagonas, C.; Tzimiropoulos, G.; Zafeiriou, S.; Pantic, M. 300 faces in-the-wild challenge: The first facial landmark localization challenge. In Proceedings of the IEEE International Conference on Computer Vision Workshops,Sydney, Australia, December 2 – December 8 2013 2013; pp. 397-403.
[4] Chrysos, G. G.; Antonakos, E.; Zafeiriou, S.; Snape, P. Offline deformable face tracking in arbitrary videos. In Proceedings of the IEEE international conference on computer vision workshops,Santiago, Chile, December 7 – December 13 2015; pp. 1-9.
[5] Shen, J.; Zafeiriou, S.; Chrysos, G. G.; Kossaifi, J.; Tzimiropoulos, G.; Pantic, M. The first facial landmark tracking in-the-wild challenge: Benchmark and results. In Proceedings of the IEEE international conference on computer vision workshops,Santiago, Chile, December 7 – December 13 2015; pp. 50-58.
[6] Tzimiropoulos, G. Project-out cascaded regression with an application to face alignment. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,Boston, MA, USA, June 7 – June 12 2015; pp. 3659-3667.
[7] Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A. C. Ssd: Single shot multibox detector. In Proceedings of the European conference on computer vision, Amsterdam, The Netherlands, October 8 – October 16 2016; pp. 21-37.

금속 표면 결함 탐지 및 영상 분할을 위한 딥러닝 기반의 다중 클래스 결함 학습

박재현 [1], 최우제 [2], 전세호 [1], 이동규 [1,2*]

경북대학교 융합학부 [1]
경북대학교 인공지능학과 [12*]

{aurumstar24, jeya9811, jsh4945, dglee}@knu.ac.kr

Deep learning based Multi-class Defect Learning for Steel Defect Detection and Segmentation

Jae-Hyun Park, Woo-Je Choi, Sae-Ho Jeon, and Dong-Gyu Lee*
Kyungpook National University*.

요 약

본 논문은 금속 표면 결함 탐지 및 영상 분할을 위한 다중 클래스 학습 방법을 제안한다. 기존의 U-Net 기반 금속 결함 탐지 연구에서는 영상을 정상/비정상으로 구분하였을 경우 다양한 결함의 종류에 효과적으로 대응하지 못하는 단점이 있다. 이러한 문제를 해결하기 위해서 결함의 종류를 세분화하여 모델을 학습시키고 이를 결함 탐지에 활용하여, 최종적으로 강인한 금속 결함 모델을 학습시킬 수 있다. 제안하는 방법은 공개 데이터셋에서 우수한 성능을 보임으로써 그 효용성을 입증하였다.

Ⅰ. 서 론

금속 제작 공정에 있어서 금속 표면 검사 및 결함 탐지는 생산된 금속의 품질을 보장하기 위해 필수적인 공정이다. 표면 결함은 철강의 부식을 촉진시키기 때문에 최종적으로 생산된 제품의 품질 및 안정성에 큰 영향을 끼칠 수 있다 [1]. 통상적인 금속 제조 공정에서 표면 결함 탐지는 사람에 의해 수행되어 왔으나, 이러한 방법은 시간이 많이 소요되고 작업자에 따라 불량 선별 기준이 변할 수 있기 때문에 동일한 품질의 결과물을 보장하기 어렵다는 단점이 있다. 이러한 문제점을 해결하기 위해, 카메라로 촬영된 영상으로부터 금속 표면을 검사하여 결함을 자동으로 탐지하고 그 위치를 특정하는 기술이 금속 산업 분야에서 많은 주목을 받고 있다 [2].

금속 표면의 결함 탐지를 위한 영상 분석 기술로 category-aware object detection network (CADN) [3], defect detection network (DDN) [4], U-Net [5] 과 같은 convolutional neural network (CNN) 기반의 영상 특징 추출 및 영상 분할 기법이 주로 활용되고 있다. 기존의 금속 결함 탐지는 결함의 종류와는 별개로 결함의 유무에 따라 결함의 위치를 판별하는 형식으로 학습을 진행한다. 이때 영상 내에 다른 종류의 결함이 동시에 존재하는 경우, 이것을 모두 결함으로 학습을 하지 않고 결함의 종류별로 구분하여 별도로 학습하게 된다. 각 결함의 유무를 판단하여 반복적으로 학습시킴으로써 다른 종류의 결함 인식 성능 및 학습 효율 저하가 발생한다.

본 논문에서는 금속 표면에 발생 가능한 결함 영역을 파악하고 정확한 결함 영역의 분할 제시가 가능한 심층 신경망 모델을 제안한다. 기존의 방법은 금속의 표면을 정상/결함 영역으로 나누어 분할하였지만, 제안하는 방법에서는 결함의 종류에 따른 패턴을 구분하여 학습함으로써, 한 장의 영상에 두가지 이상의 결함이 존재할 때 모든 결함을 효율적으로 탐지 및 분할 가능하다.

Ⅱ. 본론

2.1 금속 결함 탐지 및 분할을 위한 네트워크 구조

본 연구에서는 금속 결함 탐지 및 결함 영역 분할을 위해서, 우리는 영상 분할에 우수한 성능을 보이는 U-Net [5] 구조를 활용한다. U-Net 은 영상 분할을 목적으로 제안된 end-to-end 방식의 fully-convolutional network 기반 모델로 수축 경로 (contracting path)와 확장 경로 (expansive path)로 구성되어 있다. 수축 경로에서 발생하는 입력 영상의 공간적 상관관계의 문맥 정보의 손실을 최소화하기 위하여, 동일 단계의 확장 경로 해당 레이어의 정보를 직접 전달함으로써, 정교한 영상 분할이 가능하다.

2.2 모델 학습 방법

본 연구에서는 모델의 학습을 위해 데이터셋의 ground-truth mask 를 활용하여, 각각의 영상에 대해 금속 표면의 정상 상태 및 4 종류의 결함을 5 개 채널에 one-hot vector 형태로 인코딩 하여 처리한다.

각각의 픽셀에 부여된 라벨을 토대로 한 최종예측 값은 소프트 맥스(softmax) 함수를 통해 확률분포 형태로 출력한다. 예측값의 오차 계산은 크로스 엔트로피(cross entropy) loss 를 측정하며, 라벨을 p, 예측 값을 q 로 표기할때 최종적인 오차 계산의 수식은 아래 수식과 같다.

$$H(p,q) = -\sum_{x} p(x)\log q(x)$$

Ⅲ. 실험 및 분석

3.1 데이터셋

제안하는 금속 결함 탐지 프레임워크를 학습 및 평가하기 위해 Severstal: Steel Defect Detection [6] 데이터셋을 활용한다. 해당 데이터셋은 kaggle 에서 2019 년 competition 활용된 데이터셋으로, train split 12600 장과 test split 5506 장으로 구성되어 있다. 한장의 영상은 256×1600 크기이며, 모든 영상은 픽셀 단위의 결함 마스크가 총 5 종류의 라벨(정상, 결함 1, 결함 2, 결함 3, 결함 4)로 제공된다.

3.2 구현 환경 상세

본 논문에서는 수행된 실험은 다음과 같은 환경에서 수행되었다. 제안하는 모델을 학습시키기 위하여 learning rate = 1×10^{-4}, batchsize = 16 으로 설정하였다. 학습 단계에서 AdamOptimizer 를 활용하여 40 epoch 를 학습하였다. 모든 실험은 Google Colab 환경에서 Tensorflow 를 사용하여 구현되었다.

3.3 실험결과 및 분석

본 실험에서 우리는 입력 영상의 결함 탐지 및 영상 내 결함 영역 분할 성능 측정을 위해 mean Intersection-over-Union (mIoU)를 사용하였다. 또한, 연구의 목적이 결함이 발생한 금속을 탐지하는데 있기 때문에, 학습단계에서는 정상 및 4 종의 결함 클래스로 학습하고, 최종 결함 탐지 성능 측정단계에서는 결함/비결함을 분류하여 성능을 측정하였다

방법	성능
Vanilla CNN	75.9
U-Net [5]	91.3
제안하는 방법	94.1

표 1 Steel Defect Detection 데이터셋 에서의 결함 탐지 성능 비교

표 1 은 금속 결함 탐지 성능을 기존 방법과 비교한 것이다. 제안하는 방법은 94.1 의 성능을 보여, 동일 환경에서 학습시킨 기본 CNN 구조 및 U-Net 구조에 비하여 2.8 높은 결함 탐지 결과를 보임으로써, 제안하는 방법이 효율적이고 유효함을 보였다.

Ⅳ. 결론

본논문에서는 금속 결함 탐지를 위하여 U-Net 기반 멀티 클래스 학습 및 영상 분할 방법을 제안한다. 다양한 종류의 결함을 학습 단계에서 구분하여 별도의 클래스로 인식하도록 학습시키고, 최종적으로 분류할 때 멀티 클래스를 하나의 결함으로 인식하도록 함으로써, 금속 표면의 상태 비정상 인식 성능을 향상시켰다. 공개 데이터셋에서의 금속 결함 탐지 성능은 기존의 방법을 능가하면서 결함 종류에 따른 멀티 클래스 학습이 금속 결함 탐지에 효과적임을 보였다.

.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT 연구센터지원사업(IITP-2021-2020-0-01808)과 교육부의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2021R1C1C1012590).

참 고 문 헌

[1] S. Jinjie and J. Ming, "Influence of defects at the steel-mortar interface on the corrosion behavior of steel," Construction and Building Materials, 136, (2017): 118-125.

[2] L. Qiwu, X. Fang, L. Liu, C. Yang, and Y. Sun, "Automated visual defect detection for flat steel surface: A survey," IEEE Transactions on Instrumentation and Measurement, 69, no. 3 (2020): 626-644.

[3] Z. Jiabin, H. Su, W. Zou, X. Gong, Z. Zhang, and F. Shen, "CADN: A weakly supervised learning-based category-aware object detection network for surface defect detection," Pattern Recognition, 109, (2021): 107571.

[4] Y. He, K. Song, Q. Meng, and Y. Yan, "An end-to-end steel surface defect detection approach via fusing multiple hierarchical features," IEEE Transactions on Instrumentation and Measurement, 69, no. 4 (2019): 1493-1504.

[5] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," In International Conference on Medical image computing and computer-assisted intervention, pp. 234-241. Springer, Cham, 2015.

[6] https://www.kaggle.com/c/severstal-steel-defect-detection/data.

음성인식 오류를 포함한 텍스트 대상 언어모델 학습 방법 및 이를 이용한 음성언어 의도 분류 성능 개선에 관한 연구

김준우, 윤혜경, 정호영 *
경북대학교

kaen2891@knu.ac.kr, yhk04150@knu.ac.kr, *hoyjung@knu.ac.kr

A study on training language model of speech recognized text and improvement of spoken language intent classification

Kim June-Woo, Yoon Hyekyung, Jung Ho-Young*
Kyungpook National Univ.

요 약

본 논문은 음성인식 오류로 인해 실제 사용 환경에서 발생하는 사용자 의도 분류 성능 저하를 개선하고자 사전 학습된 BERT 언어모델을 바탕으로 오류를 포함한 음성으로부터 인식된 텍스트와 전사 데이터로 각각 학습을 진행하고 의도 분류 성능을 평가한다. 실제 환경에서 음성인식 오류에 강인한 모델 학습법에 대해 제안하고 실험 결과를 통해 본 방법의 효과에 대해 검증한다.

Ⅰ. 서 론

인공지능 기술이 발전함에 따라 다양한 분야에 이를 적용하는 추세이며, 특히 인공지능 스피커에 대한 수요가 해마다 증가하고 있다. 하지만 시중의 인공지능 스피커들은 사용자의 요청을 이해하는데 여전히 불안정한 성능을 보이고 있다. 사용 환경에서의 주변 소음과 음성인식 시스템을 학습할 때 사용하는 데이터셋에서 희박한 노인과 어린이의 음성, 그리고 불확실한 발음 등이 음성인식 시스템의 결과인 텍스트에 오류를 유발하고, 이 오류들이 그대로 언어모델로 전해지면서 사용자 의도 분류 시 오류가 발생한다.

자연어처리 도메인에서 Transformer [1], 그리고 해당 모델을 기반으로 한 BERT [2] 모델이 널리 사용되고 있다. 특히 BERT 는 대규모 텍스트 데이터셋으로 self-supervised learning 을 통해 representation 추출 및 적은 수의 데이터로 fine-tuning 이 가능한 강력한 모델이다. 하지만 해당 모델을 사용하여 음성인식 오류, 즉 잡음이 포함된 텍스트에 대해서 강력한 효과를 보이지 못하는 단점이 존재한다.

위의 문제를 해결하기 위해, 본 논문에서는 음성인식 오류가 포함된 recognized text 를 추가적으로 얻고, 원본 그대로인 전사 데이터 labeled text 를 함께 활용하는 방법을 제안한다. 이를 위해, Fluent Speech Commands (FSC) [3] 데이터셋과 audio-SNIPS [4] 음성 데이터셋을 open 음성인식 시스템인 wav2vec 2.0 [5] 에 입력으로 사용하여 recognized text 셋을 확보한다. 이후, labeled text 셋과 함께 사전 학습된 언어모델 BERT 2 개를 활용하여 이중 학습을 진행한다.

평가 단계에서는 학습된 target-BERT 모델만 이용하여 recognized, labeled 모두에 대해 평가를 진행한다. 우리의 실험 결과는 recognized text 셋에 개선된 의도 분류 정확도를 보여줌과 동시에 labeled text 셋에도 큰 성능 저하가 없음을 보여준다. 이를 통해 제안하는 방법이 효과적임을 보인다.

Ⅱ. 본론

본 장에서는 제안하는 이중 학습법에 대해 소개한다.

Ⅱ-Ⅰ. 이중 BERT 모델

언어모델 학습을 위해 대규모 텍스트 데이터셋으로 사전 학습된 BERT 를 활용한다. 우리의 방법은 2 개의 BERT 를 이중 공동 학습 방법이며, labeled text 로 학습을 진행할 C-BERT 와 recognized text 로 학습을 진행할 N-BERT 로 구성되어 있다. 이 두 모델을 함께 배치하여 동시에 의도 분류 이중 학습을 진행하며, 학습 중 C-BERT 의 gradient 들이 N-BERT 에 흐르게 하여 N-BERT 가 labeled text 의 정보를 함께 학습할 수 있도록 한다. 이를 통해 정형화된 데이터셋 및 잡음에 강인한 모델이 만들어질 것으로 예상된다. 평가시에는 recognized text 와 labeled text 모두로부터 학습된 N-BERT 만 활용하여 테스트를 진행한다.

Ⅱ-Ⅱ. 실험 세팅

실험을 위해 spoken language understanding 분야의 벤치마크 데이터셋 중 짧은 명령어가 녹음된 FSC 데이터셋과 인공지능 음성 합성기인 Amazon Polly 에 의해 생성된 음성 파일로 구성된 Audio-SNIPS 데이터셋을 활용한다. FSC 데이터셋은 총 30,043 개의 음성파일과 labeled text 로 구성되어 있으며, Audio-SNIPS 데이터셋에는 총 16 개의 서로 다른 AI 음성 합성기가 생성한 각각 2,700 개의 음성 파일과 그에 해당하는 labeled text 가 포함되어 있다.

음성인식을 통해 얻어진 오류가 있는 recognized text 를 확보하기 위하여 self-supervised 기반의 대규모 음성 데이터셋으로 학습된 wav2vec 2.0 모델을 활용한다.

실험은 NVIDIA TITAN RTX 24GB GPU 로 진행되었으며, 학습률은 1e-4, batch 사이즈는 64, AdamW 옵티마이저를 사용하였다. 모든 fine-tuning 은 5 epoch 으로 하였다.

Ⅱ-Ⅲ. 실험 결과

제안하는 모델 성능을 평가하기 위해 사전 학습된 기존 BERT 와 N-BERT 의 성능을 비교한다.

Model	평가 데이터셋: X	
	FSC	A-SNIPS
	의도 분류 성능(%)	
BERT		
X	100	99
$\hat{X}$	98.86	97.97
N-BERT		
$X + \hat{X}$	100	98.86

Table 1: Labeled 테스트셋에 대한 의도 분류 결과

Table 1 은 각각 단일 BERT 를 labeled text X, recognized text $\hat{X}$로 학습시킨 뒤 X의 테스트셋에 대한 평가 결과이다. 단일 BERT 로 학습된 두 결과 모두 비슷한 성능을 나타내고 있으며, 이는 사전 학습된 BERT 가 대규모 text 데이터셋으로부터 언어모델을 충분히 표현하기 때문에 fine-tuning 에 잡음이 추가된 데이터가 오더라도 좋은 퍼포먼스를 보여주는 것을 알 수 있다. 마찬가지로 본 논문에서 제안한 N-BERT 도 두 데이터셋에 대해 높은 성능을 내고 있다.

Model	평가 데이터셋: $\hat{X}$	
	FSC	A-SNIPS
	의도 분류 성능(%)	
BERT		
X (100%)	89.27	90.79
X (100%) + $\hat{X}$(10%)	95.36	97.62
X (100%) + $\hat{X}$(99.9 %)	97.92	98.16
N-BERT		
$X + \hat{X}$	98.49	98.54

Table 2: Recognized text 셋에 대한 각 모델의 실험 결과

Table 2 는 recognized text 의 테스트셋에 대한 실험 결과이다. FSC, Audio-SNIPS 두 데이터셋에 대하여 recognized text 의 비율이 높아질수록 단일 BERT 의 정확도개 개선된다. 하지만, 단일 BERT 로 사용한 의도 분류 성능은 labeled text 셋에 비해 성능 저하가 일어나는 것을 알 수 있다.

반면에 제안한 N-BERT 는 recognized 테스트셋에 대해 각각 98.49%, 98.45%의 결과를 보였으며, Table 1 에서 볼 수 있듯이, N-BERT 는 labeled text 에서도 성능 저하가 크게 일어나지 않는 것을 알 수 있다.

Ⅲ. 결론

본 논문에서는 음성인식 오류로부터 강인한 언어모델 연구에 초점을 맞추었다. 이를 위해 음성인식 시스템을 통과하여 오류를 포함한 recognized text 셋으로 N-BERT 를 학습하고, 원본 그대로인 전사 데이터 labeled text 셋으로 C-BERT 를 학습하는 이중 학습 방법을 제안하였다. N-BERT 와 C-BERT 를 학습한 언어모델의 테스트 성능은 오류가 포함된 입력과 더불어 오류가 포함되지 않은 입력에 대해서도 강인한 성능을 보여주었다. 이를 통해 제안하는 방법이 실제 사용 환경에서 발생하는 음성인식 오류들에 대하여 강인한 언어모델을 제공하여 사용자 음성언어 대상 의도 파악 시스템을 개선할 수 있을 것으로 기대된다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT 연구센터지원사업의 연구결과로 수행되었음 (IITP-2021-2020-0-01808*)

참 고 문 헌

[1] Vaswani, Ashish, et al. "Attention is all you need." Advances in Neural Information Processing Systems. 2017.

[2] Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).

[3] Lai, Cheng-I., et al. "Towards semi-supervised semantics understanding from speech." arXiv preprint arXiv:2011.06195 (2020).

[4] Lugosch, Loren, et al. "Speech Model Pre-Training for End-to-End Spoken Language Understanding" Interspeech 2019-818.

[5] Baevski, Alexei, et al. "wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations." Advances in Neural Information Processing Systems 33 (2020).

Attention-Guided Feature Enhancement for Age-Related Macular Degeneration Classification in Optical Coherence Tomography

Jung Rae Cho[1†] , Dong Ho Park[2†] and Sung Moon Jeong[3*]
[1]ICT-connected Composite Intelligence Research Center, Kyungpook National University, Daegu, Republic of Korea
[2]Department of Ophthalmology, School of Medicine, Kyungpook National University, Kyungpook National University Hospital, Daegu, Republic of Korea
[3]Department of Medical Informatics, School of Medicine, Kyungpook National University, Daegu, Republic of Korea

Abstract.

Age-related macular degeneration (AMD) is a vision-threatening chorioretinal disease, which is one of the most common causes of adult blindness. To diagnose AMD automatically and accurately, convolutional neural network (CNN)-based various deep learning models are applied to classify which optical coherence tomography (OCT) scans have AMD lesions or not. A conventional approach was to use a pair of input and output data without specific domain knowledge, which means OCT scans are given as input to the network and it is propagated convolutional layers to classify OCT scans called end-to-end learning method. To improve the accuracy of AMD classification, attention-guided feature enhancement method was applied in CNN by directly supervising to network for attending and intensifying feature maps of AMD biomarkers. Pre-trained AMD biomarker detector produces attention maps of biomarkers, and the maps are combined with several feature maps of CNN classifier to diagnose AMD types. The proposed approach has better classification accuracy than simple end-to-end baseline learning model.

Ⅰ. Introduction

There have been many researches to diagnose or classify ophthalmologic diseases including age-macular degeneration (AMD) and diabetic macular edema (DME) from optical coherence tomography (OCT). Naohiro Motozawa et al. (2019) [1] have proposed OCT-based convolutional neural networks (CNNs) with transfer learning to classify normal and AMD with and without exudate change, and have achieved high sensitivity, specificity, and accuracy. Ali Serener et al. [2] (2019) has suggested CNN to diagnose dry and wet AMD from OCT images. Several researches have utilized region-based CNN (R-CNN) to detect AMD biomarkers in OCT scans. M. Suchetha et al. (2021) [3] have tried to utilize Faster R-CNN [4] for semantic segmentation of AMD and DME. This approach has focused on localization and segmentation of AMD and DME. Most of approaches have usually used a pair of data called end-to-end learning, which consists of OCT images as an input and AMD lesions or types as an output, respectively. During training phase, latent model is expected to automatically tune by propagating a data through the convolutional layers to indirectly find relationships between inputs and outputs.

In this paper, we introduce a new AMD diagnosis algorithm by directly enhancing multiple CNN feature maps through the supervisions to the deep network. There are two stages in the proposed approach: (1) AMD lesion localizer is pre-trained and it can give an attention map of AMD biomarkers to supervise the network in the next classification stage. (2) The attention map is directly used for supervising feature maps of CNN classifier for AMD types, i.e. normal, dry and wet AMD. Based on these two stages, the proposed model can directly understand relationships between inputs and outputs to classify AMD types and it has shown that the classification accuracy has much improvement by applying attention guided feature enhancement method from the end-to-end baseline learning method.

Ⅱ. Method

2.1 Dataset

In this paper, we have collected OCT volumes from 99 eyes from controlled groups, 96 and 238 eyes from dry and wet AMD patients, respectively. From each OCT volume, OCT scans close to macular have been annotated by drawing bounding box and labeling the name of biomarker type, e.g. drusen, geometric atrophy for AMD biomarker localization. For each OCT scan, annotators have also labeled type of AMD, i.e. normal, dry AMD, and wet AMD for classification. Finally, 2,153 OCT scans have been collected and annotated. Ophthalmologists have participated in annotation of OCT scans. Fig. 1 shows examples of different types of OCT scans.

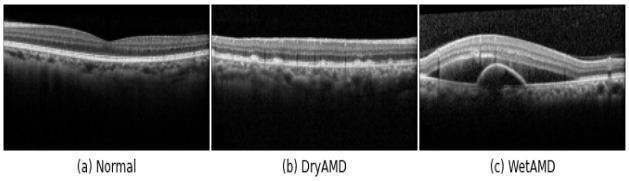


Fig. 1 Examples of different type of OCT scans

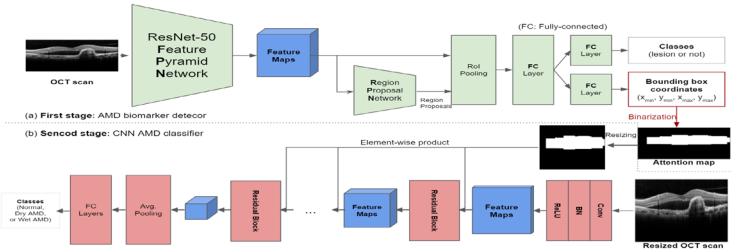


Fig. 2 Overview of the proposed method

2.2 Method

We propose two-stage AMD classification method. Fig 2. shows overview of the proposed method. In the first stage, Faster R-CNN (green blocks in Fig. 2) is used to detect AMD biomarkers. We regard the class of AMD biomarkers as just binary class 'lesion' to reduce complexity of training models and focus on localization itself. To use localization of trained detector, we generate binary map as shown in Fig. 2. The foreground and background values are set to 1.5 and 1.0 to attend location of AMD biomarkers in feature maps of CNN in the second stage.

The second stage is focused on classification of AMD type. In this work, we choose ResNet-18 architecture [5] as CNN AMD classifier as described in Fig. 2. Once trained AMD biomarker detector produces an attention map, OCT scan and attention map are fed into CNN concurrently. The attention map is down-sampled and multiplied in an element-wise manner to each feature map of convolutional layers. Hence, the region-attended feature maps are propagated layer by layer as shown in Fig. 2.

2.3 Experiments

We set 395 OCT scans as test set. We evaluate the performance of AMD biomarker detector and AMD classifier respectively. Since we focus on the ability of localization not specific class of biomarker, we calculate mean IoU (mIoU) of attention map with ground-truth binary map. For test set, mIoU records 0.55. We also evaluate classification accuracy, sensitivity, and, specificity for each class. We experiment 5-fold cross validation, and result is presented in table. 1. Table. 1 shows that proposed method enhances overall performances compared to the baseline model.

Ⅲ. Conclusion

In this paper, we propose two-stage AMD classification method to enhance biomarkers' features in CNN classifier. Pre-trained AMD biomarker detector generates attention map and it is fed into CNN to attend and intensify multiple feature maps of convolutional layers, and it leads to enhance overall performances of classification of AMD types. The proposed approach shows that clinical domain knowledge like biomarkers' feature in OCT scans is useful a supervising signal to directly enhance feature maps for deep neural networks.

Model	Class	Accuracy	Sensitivity	Specificity
ResNet-18 (baseline)	Normal	0.985	0.963	0.98
	Dry AMD	0.967	**0.965**	0.967
	Wet AMD	0.98	0.967	0.995
Proposed Model	Normal	**0.989**	**0.996**	**0.986**
	Dry AMD	**0.981**	0.919	**0.994**
	Wet AMD	**0.99**	**0.991**	**0.989**

Table. 1 Evaluation of classification of AMD

ACKNOWLEDGMENT

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2021-2020-0-01808) supervised by the IITP (Institute of Information & Communications Technology Planning & Evaluation). DP was financially supported by the Basic Science research Program of the National Research Foundation of Korea (NRF), funded by the Korean government (Ministry of Science and ICT; 2019R1A2C1084371).

References

[1] Motozawa, Naohiro, et al. "Optical coherence tomography-based deep-learning models for classifying normal and age-related macular degeneration and exudative and non-exudative age-related macular degeneration changes." Ophthalmology and therapy 8.4 (2019): 527-539.

[2] Serener, Ali, and Sertan Serte. "Dry and wet age-related macular degeneration classification using oct images and deep learning." 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT). IEEE, 2019.

[3] Suchetha, M., Rajiv Raman, and Edwin Dhas. "Region of Interest based Predictive Algorithm for Subretinal Hemorrhage Detection using Faster R-CNN." (2021).

[4] Ren, Shaoqing, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks." Advances in neural information processing systems 28 (2015): 91-99.

[5] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

동작 상상 뇌파 분류를 위한 합성곱 신경망 기반의 전이학습에 관한 연구

송하윤 [1], 김동규 [1], 김영탁 [1], 김학승 [1], 김동주 [1,2,3*]
[1] 고려대학교 뇌공학과, [2] 고려대학교 의과대학 신경과, [3] 고려대학교 인공지능학과

{hayoonsong, dien200318, kyt3426, mkhsm, dongjookim}@korea.ac.kr

A Study on the Transfer Learning with Convolution Neural Network for Electroencephalogram Based Motor Imagery Classification

Ha-Yoon Song[1], Dong-Kyu Kim[1], Young-Tak Kim[1], Hakseung Kim[1], Dong-Joo Kim[1,2,3*]
[1]Dept. of Brain and Cognitive Engineering, Korea University,
[2]Dept. of Neurology, College of Medicine, Korea University,
[3]Dept. of Artificial Intelligence, Korea University

요 약

뇌-컴퓨터 인터페이스(Brain-Computer Interface, BCI)는 뇌활성도를 통해 외부기기를 움직이거나 제어하는 기술로, 신체적 기능에 제한이 있는 노약자 및 장애인에게 새로운 움직임 보조수단이 될 수 있다. 최근 딥러닝 기술은 BCI 분야에서 안정적이고 높은 수준의 뇌파(Electroencephalogram, EEG) 특징 추출 및 분류 성능을 보여주었으나, 데이터가 충분하지 않을 경우 학습하는 과정에서 과적합 문제를 초래할 수 있다. 본 논문에서는 합성곱 신경망 기반의 전이 학습을 활용하여 BCI 동작 상상 EEG 를 분류하는 기법을 제안한다. 이를 위해 BCI competition IV-2a 데이터셋에서 대상 사용자의 EEG 데이터를 제외한 나머지 데이터로 합성곱 신경망 모델을 생성하고 대상 사용자의 EEG 데이터를 이용하여 전이학습을 수행하였다. 제안하는 기법은 소규모 데이터만으로 생성된 모델보다 평균 10.8%의 분류 정확도 향상을 이끌어내었다 (64.7%).

Ⅰ. 서 론

뇌-컴퓨터 인터페이스(Brain-Computer Interface, BCI)는 뇌활동을 통해 외부기기를 움직이거나 제어할 수 있는 새로운 기술이다. 그 중 뇌파(Electroencephalogram, EEG) 기반의 동작 상상(Motor Imagery)을 분류하는 기술은 신체장애가 있는 사람에게 새로운 움직임 보조수단으로 각광받고 있다[1].

기계학습 모델 중 합성곱 신경망(Convolutional Neural Network, CNN)은 이미지 및 음성인식 분야에서 특징 추출 및 분류 알고리즘으로 우수한 성능을 보이면서, 최근 BCI 분야에서도 효과적인 EEG 의 특징 추출 및 분류를 위해 사용되고 있다[2]. CNN 을 학습하기 위해서는 충분한 양의 데이터가 필요하다. 그러나 개인 차의 영향을 받는 BCI 분야에서 대규모 EEG 데이터를 구축하기에는 한계가 있다[3]. 이러한 데이터 부족을 해결하거나 비용 절감을 위한 하나의 방법으로 전이학습(Transfer learning)이 사용된다[4].

전이학습은 사전에 학습이 완료된 모델의 가중치를 초기 가중치로 설정하여, 목적하는 새로운 모델로 가중치를 전이하여 학습하는 기법이다[5]. 전이학습은 비용 절감 뿐만 아니라, 사전 학습된 네트워크의 지식이 새로운 데이터로 전이되어 적은 데이터로도 좋은 성능을 나타낸다는 장점이 있다[6].

본 논문은 대상 사용자를 제외한 나머지 데이터로 학습된 CNN 을 대상 사용자의 데이터로 전이학습한 모델과, 대상 사용자의 데이터만으로 학습한 CNN 모델 사이의 성능을 비교하였다.

Ⅱ. 본 론

본 논문에서는 비교를 위해 BCI Competition IV-2a 오픈 데이터셋을 사용하였다[7]. EEG 데이터는 9 명의 사용자로부터 22 개의 EEG 채널과 3 개의 안전도(Electrooculography, EOG) 채널로 총 25 채널을 사용하여 기록되었고 250Hz 로 샘플링 되었다. 총 4 클래스(왼손, 오른손, 혀, 양발)로 구성되어 있으며 각 클래스 마다 72 trials 가 제공된다. 이 연구에서는 cue 가 주어진 후 0.5 초부터 2 초간의 EEG 데이터가 사용되었다.

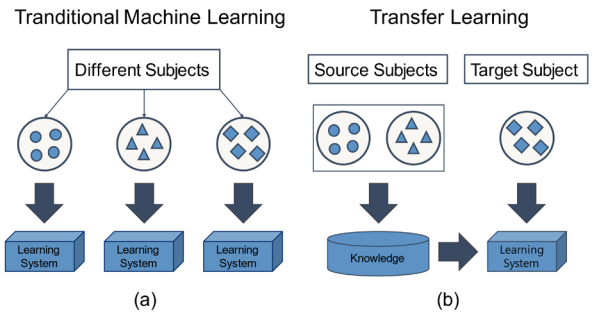


그림 1. 모델의 학습 방법: (a) 일반적인 기계학습 방법 (b) 전이학습 방법

*Corresponding author

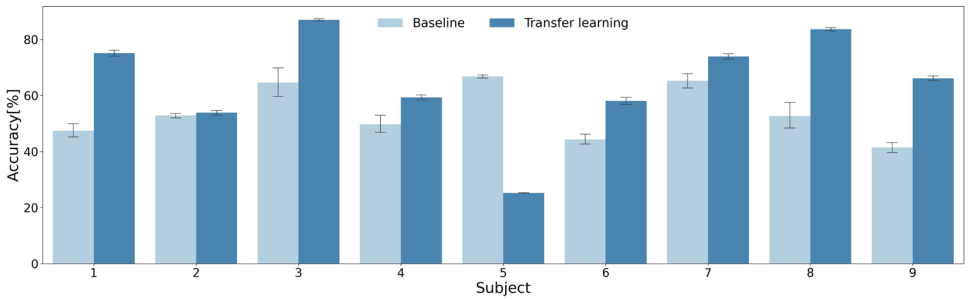


그림 2. 사용자에 따른 분류 정확도

그림 1 은 일반적인 기계학습과 전이학습에서의 모델 학습 방법을 나타낸다. 전이학습의 성능을 비교하기 위하여, 그림 1 (a)와 같이 각 사용자마다 구축된 모델을 기준점으로 설정하였다. CNN 기반의 전이학습은 그림 1 (b)와 같이 대상 사용자를 제외한 나머지 사용자들의 데이터를 이용하여 사전 모델을 구축하고, 구축된 사전 모델의 지식 이전(Knowledge transfer)과 함께 대상 사용자의 데이터만을 활용하여 모델을 학습하였다. 이 때, 미세조정을 통해 사전 학습된 CNN 의 모든 층의 가중치를 미세하게 재조정하였다[8].

본 논문에서 사용된 모델인 EEGNet 은 EEG 기반 분류 문제를 위한 경량형 CNN 으로, 적은 데이터셋에서도 좋은 성능을 보여주었다[9]. 최적화 알고리즘으로 Adam(Adaptive moment)을 사용하였으며, 학습률은 0.001 으로 설정하였다. 배치사이즈(Batch Size)와 반복 횟수(Epoch)를 각각 128, 500 으로 고정하였고, 학습 조기 종료(Early stopping) 함수의 patience 를 50 으로 지정하였고, 각 모델별로 10-fold 교차 검증을 수행하였다.

그림 2 는 전이학습을 여부에 따른 동작상상 분류 모델의 정확도를 보여준다. 전이학습의 효과를 평가하기 위한 기준으로, 대상 사용자의 데이터만을 이용하여 학습된 EEGNet 은 평균 53.89%의 분류 정확도를 보였다. 반면에 다른 사용자의 데이터로 사전 학습하고 대상 사용자의 데이터로 전이학습 시킨 EEGNet 의 경우 평균 64.69 %로 기준점보다 높은 정확도를 보였다.

Ⅲ. 결 론

본 논문에서는 동작 상상 뇌파 분류의 정확도를 향상시키기 위해 CNN 기반의 전이학습을 제안하였다. 제안하는 방법은 사전 학습 없이 학습된 CNN 모델에 비해 10.8% 이상의 성능 증가를 보였다.

ACKNOWLEDGMENT

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program (IITP-2021-2016-0-00464) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation); IITP grant funded by the Korea government (MSIT) (No. 2017-0-00432, Development of non-invasive integrated BCI SW platform to control home appliances and external devices by user's thought via AR/VR interface); National Research Foundation of Korea (NRF) grant funded by the Korean government (Ministry of Science and ICT, MSIT) (No. 2020R1C1C1006773).

참 고 문 헌

[1] P. Gaur, R. B. Pachori, H. Wang, and G. J. E. S. w. A. Prasad, "A multi-class EEG-based BCI classification using multivariate empirical mode decomposition based filtering and Riemannian geometry," vol. 95, pp. 201-211, 2018.

[2] A. Craik, Y. He, and J. L. J. J. o. n. e. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: a review," vol. 16, no. 3, p. 031001, 2019.

[3] Z. Wan, R. Yang, M. Huang, N. Zeng, and X. J. N. Liu, "A review on transfer learning in EEG signal analysis," vol. 421, pp. 1-14, 2021.

[4] H. Lee, J. Kim, J. Yu, Y. Jeong, and S. J. J. o. K. I. o. I. S. Kim, "Multi-class Classification using Transfer Learning based Convolutional Neural Network," vol. 28, no. 6, pp. 531-537, 2018.

[5] S. Chung, M. G. J. J. o. I. C. Chung, and Services, "Pedestrian Classification using CNN's Deep Features and Transfer Learning," vol. 20, no. 4, pp. 91-102, 2019.

[6] D.-K. Kim, Y.-T. Kim, H.-R. Jung, H. Kim, and D.-J. Kim, "Sequential Transfer Learning via Segment After Cue Enhances the Motor Imagery-based Brain-Computer Interface," in *2021 9th International Winter Conference on Brain-Computer Interface (BCI)*, 2021, pp. 1-5: IEEE.

[7] C. Brunner, R. Leeb, G. Müller-Putz, A. Schlögl, and G. J. I. f. K. D. Pfurtscheller, Graz University of Technology, "BCI Competition 2008- Graz data set A," vol. 16, pp. 1-6, 2008.

[8] H.-W. Ng, V. D. Nguyen, V. Vonikakis, and S. Winkler, "Deep learning for emotion recognition on small datasets using transfer learning," in *Proceedings of the 2015 ACM on international conference on multimodal interaction*, 2015, pp. 443-449.

[9] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. J. J. o. n. e. Lance, "EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces," vol. 15, no. 5, p. 056013, 2018.

효율적인 수면단계 분류를 위한 딥러닝 모델 압축 연구

허재욱 [1], 이철희 [1], 김학승 [1], 김동주 [1,2,3*]
[1] 고려대학교 뇌공학과, [2] 고려대학교 의과대학 신경과, [3] 고려대학교 인공지능학과

{kinik33, dlcjfgmlnasa28, mkhsm, dongjookim}@korea.ac.kr

Efficient Sleep Stage Classification Based on Deep Learning Using Model Compression

Jae-Wook Heo[1], Choelhui Lee[1], Hakseung Kim[1], Dong-Joo Kim[1,2,3*]
[1]Dept. of Brain and Cognitive Engineering, Korea University,
[2]Dept. of Neurology, College of Medicine, Korea University,
[3]Dept. of Artificial Intelligence, Korea University

요 약

수면단계는 수면장애의 진단에 있어 중요한 지표로 사용된다. 수면단계 진단의 검사자 의존성 및 많은 시간소요를 해결하기 위해 딥러닝 연구가 활발하게 이루어지고 있지만, 아직까지 생체신호 수집의 낮은 접근성으로 인해 많은 사람이 수면장애를 진단받지 못한 채로 남아있으며, 딥러닝 모델의 경량화를 통한 웨어러블 디바이스의 탑재가 이에 대한 해결방안으로 떠오르고 있다. 본 논문은 수면단계 분류를 위한 딥러닝 모델의 경량화를 위해 두 단계의 모델 압축 기법을 제안한다. 해당 압축기법을 통해 높은 성능을 유지하며 모델의 파라미터 수를 획기적으로 줄여 모델을 경량화 시킬 수 있다. 실험 결과, 압축을 통해 모델의 파라미터 수를 약 85.6% 줄일 수 있었으며, 압축된 모델은 기존 모델에 비해 약 12%의 분류 속도 향상을 보였다.

Ⅰ. 서 론

수면단계는 수면장애의 진단에 있어 중요한 지표로 사용되지만, 수면다원검사를 통한 수면단계 분류는 시간이 많이 걸리고 힘든 과정일 뿐만 아니라 평가자의 숙련도 및 주관에 따라 영향을 받을 수 있다는 문제가 있다. [1] 최근에는 이러한 문제점을 해결하고자, 딥러닝 방법론을 적용한 수면단계 분류 연구가 활발하게 이루어지고 있지만, 여전히 생체신호 수집의 낮은 접근성으로 인해 많은 사람들이 수면장애를 진단받지 못한 채로 남아있다.

모델압축 기법은 많은 파라미터를 가지는 딥러닝 모델의 경량화를 통해 웨어러블 디바이스 탑재를 가능케 하여 수면단계 분류의 접근성을 크게 높일 수 있다. 그럼에도 불구하고, 아직 수면단계 분류를 위한 딥러닝 모델을 압축하려는 시도는 관련 연구가 부족한 상태이다. [2]

본 논문에서는 분류 정확도가 높지만, 모델의 크기가 큰 딥러닝 모델인 DeepSleepNet 에 두 단계의 압축 기법을 적용하여 수면단계를 분류하였으며, 압축정도에 따른 분류 성능과 모델 크기에 대한 고찰을 진행하였다. [3]

Ⅱ. 본론

본 논문에서는 수면단계 분류를 위한 딥러닝 모델인 DeepSleepNet 이 CNN(Convolutional neural network) 층으로 이루어진 representation learning 부분과 LSTM(Long short-term memory) 층으로 이루어진 sequential residual learning 부분으로 나뉘어진 점에 착안하여 다음과 같은 두 단계의 압축기법을 제안한다. 1 단계에서는 채널 가지치기 기법을 이용하여 representation learning 부분의 각 층에서 중요도가 낮은 채널의 가중치를 제거하여 파라미터 수를 줄인 후 재학습을 하여 가중치의 미세조정을 진행한다. 2 단계에서는 1 단계에서 학습된 모델의 LSTM 층의 hidden state 수를 줄여가며 재학습을 하여 파라미터 수를 줄이고, 가중치의 미세조정을 진행한다. [4]

모델의 성능평가를 위해 PhysioNet 의 Sleep-EDF-39 데이터베이스를 wake, REM(Rapid eye movement), N1, N2, N3 의 5 클래스로 나누어 분류를 진행하였다. [5, 6] 데이터는 100Hz 로 기록되었으며, Fpz-Cz 채널을 학습에 사용하였다. 해당 데이터셋을 subject 단위로 train set, validation set, test set 을 나누어 학습을 진행하였으며, 사용된 최적화 기법은 AdamW 이고, 학습률은 0.0001 으로 설정하였다. 1 단계 압축의 가지치기는 각 층별로 L1 norm 최소화를 통해 5% 단위로 가지치기를 진행하였다. 2 단계 압축은 1 단계 압축의 압축률이 0%, 20%, 40%, 60%, 80% 인 모델들에 대하여 진행되었으며, LSTM 층은 hidden state 수를 128 씩 줄이며 학습을 진행하였다.

*Corresponding author

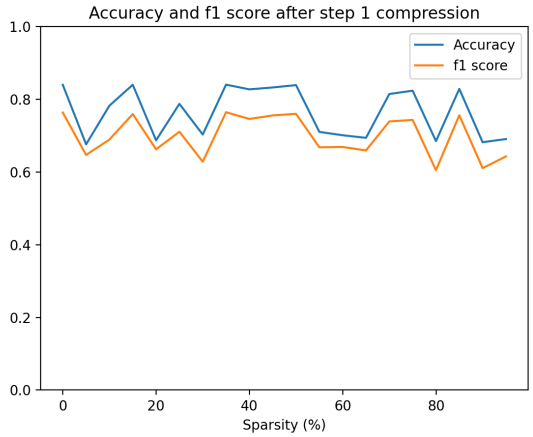


그림 1. 1 단계 압축 정도에 따른 정확도 및 f1 점수

그림 1 은 1 단계 압축을 진행한 모델의 압축 정도와 성능 사이의 상관관계를 나타낸다. 압축의 진행에 따른 눈에 띄는 성능저하가 발견되지 않는 것을 확인할 수 있다. 1 단계 압축 정도가 0%, 20%, 40%, 60%, 80%인 모델들에 대하여 2 단계 압축을 진행하였다.

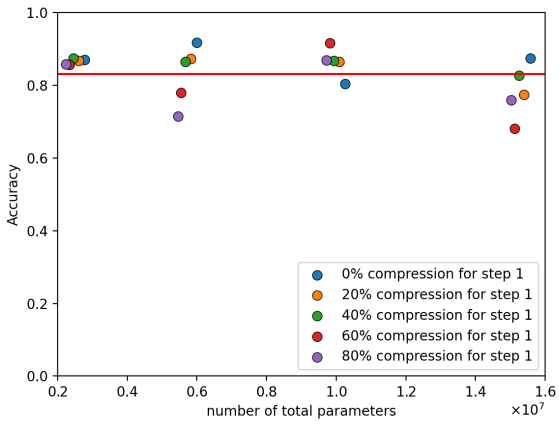


그림 2. 압축모델의 파라미터 수에 따른 정확도의 변화. 붉은 선은 기존 모델의 95% 정확도를 나타낸 선이다.

그림 2 는 2 단계 압축까지 진행한 모델들의 파라미터 수에 따른 정확도를 나타낸다. 전체 압축 모델 중 5% 이상의 성능 저하가 있는 모델은 7 개에 불과하며, 특히 압축 정도가 가장 높은 모델의 경우, 압축이 진행되지 않은 기존모델에 비해 1.8% 정도의 정확도 하락만이 발견되었다.

Ⅲ. 결론

본 논문에서는 수면단계 분류를 위한 딥러닝 모델에 두 단계의 모델 압축 기법의 적용을 제안하였다.

압축 정도가 가장 높은 모델과 기존모델 사이의 분류 속도 및 모델의 파라미터 수를 비교한 결과, 기존모델에 비해 압축모델의 분류 속도가 12% 정도 빠르며 모델의 파라미터 수를 약 85.6% 줄일 수 있음을 확인하였다.

수면단계 분류 연구에서 모델 압축은 모델 성능을 유지하며 모델의 크기를 획기적으로 줄일 수 있는 것으로 보이며, 더 많은 데이터 및 다양한 압축 조건에서 모델 압축의 효용성을 검증하고, 최적의 압축 조건을 찾는 후속연구가 필요하다.

ACKNOWLEDGMENT

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program (IITP-2021-2016-0-00464) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation); IITP grant funded by the Korea government (MSIT) (No. 2017-0-00432, Development of non-invasive integrated BCI SW platform to control home appliances and external devices by user's thought via AR/VR interface); National Research Foundation of Korea (NRF) grant funded by the Korean government (Ministry of Science and ICT, MSIT) (No. 2020R1C1C1006773).

참 고 문 헌

[1] 김형욱, 이영록, and 박동규, "안전도, 뇌파도, 근전도 분석을 통한 수면 단계 분류," 멀티미디어학회논문지, vol. 22, no. 12, pp. 1491-1499, 2019

[2] R. Duggal, S. Freitas, C. Xiao, D. H. Chau, and J. Sun, “REST: Robust and efficient neural networks for sleep monitoring in the wild,” in Proceedings of The Web Conference 2020, 2020.

[3] A. Supratak, H. Dong, C. Wu, and Y. Guo, “DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG,” arXiv [stat.ML], 2017.

[4] H. Li, A. Kadav, I. Durdanovic, H. Samet, and H. P. Graf, “Pruning Filters for Efficient ConvNets,” arXiv [cs.CV], 2016.

[5] A. L. Goldberger et al., “PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals: Components of a new Research Resource for Complex Physiologic Signals,” Circulation, vol. 101, no. 23, pp. E215-20, 2000.

[6] B. Kemp, A. H. Zwinderman, B. Tuk, H. A. Kamphuisen, and J. J. Oberyé, “Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG,” IEEE Trans. Biomed. Eng., vol. 47, no. 9, pp. 1185-1194, 2000.

효율적인 언어모델 사전학습을 위한 집중학습기법

박준형[1], 이상근[2]
[1]고려대학교 컴퓨터학과, [2]고려대학교 인공지능학과
{irish07, yalphy}@korea.ac.kr

Concentrated Learning for Efficient Pre-training of Language Models

Jun-Hyung Park[1], SangKeun Lee[2]
[1]Department of Computer Science and Engineering, Korea University
[2]Department of Artificial Intelligence, Korea University

요 약

사전학습된 언어모델은 다양한 자연어 처리 작업에서 뛰어난 성공을 거두고 있다. 그러나 성능 향상을 위해 사전학습에서 사용하는 말뭉치의 크기와 계산량이 계속해서 증가하고 있다. 우리는 효율적인 사전학습을 위해 말뭉치에서 많은 지식을 함유하는 텍스트를 추출하여 집중적으로 학습하는 집중학습 기법을 제안한다. BERT 모델을 사용한 실험을 통해 제안 기법이 일반적인 사전학습 기법보다 더 효율적으로 더 높은 자연어 이해 성능을 달성하는 것을 확인한다.

Ⅰ. 서 론

언어모델 사전학습은 자연어 추론, 감정 분류, 질의응답 등 다양한 자연어 처리 작업 성능을 향상시킬 수 있는 효과적인 수단이다 [1]. 사전학습 과정에서 언어모델은 말뭉치로부터 자연어의 구조적, 의미적 정보와 지식에 대해 학습할 수 있고, 이를 자연어 처리 작업에 활용하여 높은 성능을 달성할 수 있다. 특히 사전학습된 Transformer [2] 기반 언어모델들은 자연어 이해 벤치마크(GLUE [3] 등)에서 인간을 초월하는 점수를 달성했다. 2018년 이래로 BERT [1], RoBERTa [4] 등 사전학습된 Transformer 기반 언어모델이 계속해서 제안되어오고 있으며 다양한 자연어 처리 작업에서 최고 성능을 갱신하고 있다.

최근 연구들은 사전학습에서 사용하는 말뭉치의 크기가 커질수록 언어모델의 자연어 처리 작업의 성능이 향상된다는 것을 전제로 받아들이고 있다 [4, 5]. RoBERTa는 BERT와 비교하여 열 배 수준의 말뭉치로 사전학습되어 자연어 이해 벤치마크와 질의응답 벤치마크에서 성능을 크게 향상시키는 결과를 보여주었다. 그리고 GPT-3 [5]는 BERT와 비교하여 33배의 말뭉치로 사전학습되어 인간이 구분하지 못할 정도로 자연스러운 기사를 생성해냈고, 다양한 자연어 처리 벤치마크에서 미세조정 없이 언어모델만으로 경쟁력 있는 성능을 보여주었다.

그러나 언어모델은 전체 말뭉치를 일정 회 이상 반복적으로 학습해야 효과적인 성능을 보이기 때문에 [5], 말뭉치의 크기가 증가함과 함께 사전학습에 필요한 계산 비용 또한 증가하고 있다. 최신 언어모델들의 사전학습에 필요한 계산을 모두 수행하기 위해서 고성능의 NVIDIA V100 GPU 1대로도 1년 이상의 시간이 소요된다 [5]. 이로 인해 수백 대 이상의 GPU/TPU로 구성된 클러스터 없이는 최신 언어모델의 사전학습이 어렵고, 이는 자연어 처리 연구에 큰 걸림돌이 되고 있다.

본 연구에서 우리는 앞서 언급한 전제가 Transformer 기반 언어모델 사전학습에 있어 항상 성립하는 것이 아님을 보인다. 다양한 자연어 처리 작업에서 지식은 유용하게 사용될 수 있다. 따라서 우리는 많은 지식을 함유하는 텍스트를 추출하여 집중적으로 학습하는 집중학습기법을 제안한다. 우리는 BERT를 활용한 집중학습기법의 실증적인 평가를 통해 기존 전제와 모순되는 두 가지 현상을 관측했다. 첫째로, 우리는 사전학습 말뭉치에서 지식을 많이 함유하는 극히 일부의 텍스트만으로도 BERT의 성능을 효과적으로 향상시킬 수 있음을 발견했다. 둘째로, 우리는 사전학습 말뭉치에서 BERT 성능 향상에 도움이 되지 않는 텍스트가 있음을 발견했다. 이러한 발견들을 기반으로 우리는 큰 말뭉치와 계산 비용을 사용하여 언어모델을 사전학습하는 현재의 접근법을 재고해야 함을 주장한다. 본 연구는 단순히 대규모의 사전학습 말뭉치를 사용하기보다 소규모의 풍부한 지식을 담고 있는 텍스트를 선별하여 사전학습을 진행하는 것이 더 효과적, 효율적으로 성능을 향상시킬 수 있음을 보여준다. 본 연구의 기여점을 요약하면 다음과 같다.

- 우리는 언어모델 사전학습에서 많은 지식을 담고 있는 텍스트를 추출하여 집중적으로 학습하는 집중학습기법을 제안한다.
- 우리는 더 큰 말뭉치가 언어모델 성능을 향상시킨다는 기존 전제와는 달리 소규모의 말뭉치로도 사전학습된 언어모델의 성능을 향상시킬 수 있음을 입증한다.
- 우리는 집중학습기법을 통해 사전학습된 BERT 모델이 자연어 이해 벤치마크에서 뛰어난 성능을 달성함을 보인다.

Ⅱ. 집중학습기법

이 장에서 우리는 집중학습기법의 설계와 세부 진행 과정에 관해 설명한다. 우선 지식을 많이 함유하는 텍스트를 추출하는 기법에 관해 설명하고, 추출한 텍스트를 활용한 언어모델 사전학습 기법에 관해 설명한다.

Ⅱ-1. 지식 함유 텍스트 추출

언어모델 사전학습에 사용되는 말뭉치는 과학, 역사, 정치 등 다양한 분야와 백과사전, 소설, 기사 등 다양한 갈래의 글을 포함하고 있다. 마찬가지로, 텍스트 속에 들어있는 지식의 분야와 표현 형태 또한 다양하게 나타날 것이다. 텍스트 속에 들어있는 다양한 분야와 형태의 지식을 감지하기 위해서 우리는 3200만 개의 트리플렛으로 구성된 대규모 지식 그래프

ConceptNet [6]을 활용했다. ConceptNet은 $G \subset C \times R \times C$ 구조의 지식 그래프로 C는 자연어로 표현된 개념을 나타내고 R은 개념 사이의 관계성을 나타낸다. 하나의 트리플렛은 (c_1, r, c_2)로 구성되어 c_1이 c_2와 r의 관계성을 갖는다는 지식을 나타낼 수 있다. 예를 들어 트리플렛(river, AtLocation, waterfall)은 "강은 폭포가 있는 위치에 있다."는 지식을 표현할 수 있다.

우리는 distant supervision [7]을 바탕으로 ConceptNet을 활용하여 텍스트에 들어있는 지식을 계측한다. 다음 일련의 수식을 통해 텍스트에 들어있는 지식의 양을 측정하고 많은 지식을 포함하는 텍스트를 선별한다.

$$f(x, G) = \left|\{r \mid c_1 \subseteq x, c_2 \subseteq x, (c_1, r, c_2) \in G\}\right|$$

$$T_{extracted} = \{t \mid f(t) > \alpha, t \in T\}$$

위 수식에서 T는 사전학습 말뭉치에 들어있는 문장들의 집합을 나타내며 α는 판별 기준을 설정하는 초파라미터다. 사용하는 모든 개념과 문장들은 언어모델 학습에 필요한 방식으로 토큰화된다.

Ⅱ-2. 언어모델 사전학습

언어모델 사전학습은 기존 BERT 모델에서 사용하는 마스크 언어 모델링 방식을 사용한다. 토큰화된 문장에서 15%의 토큰을 마스킹하며 이 중 80%는 특별 토큰 중 하나인 [MASK] 토큰으로, 10%는 임의의 토큰 중 하나로 대체하며, 나머지 10%는 원본 토큰을 변경 없이 그대로 사용한다. 문맥 표현을 학습하기 위해 여러 개의 연속된 문장을 이어붙여 한 번에 최대 128 또는 512개의 토큰을 학습에 사용하는 기존 언어모델 방법론과는 달리, 우리의 방법론은 문장 단위로 학습하기 때문에 최대 64개의 토큰을 사용하여 학습을 진행한다.

Ⅲ. 실험

우리는 본 논문에서 제안한 집중학습기법으로 학습된 모델과 일반사전학습으로 학습된 모델, 그리고 지식 함유 텍스트를 추출하지 않고 임의의 문장을 추출하여 학습하는 임의추출학습 기법으로 학습된 모델을 비교하여 자연어 이해 성능 향상의 효율성을 측정했다. 자연어 이해 성능을 측정하기 위해 GLUE 벤치마크 [3]를 사용하며 GLUE 점수와 사용한 말뭉치의 크기, 계산 비용을 통해 성능 향상의 효율성을 비교했다. 사전학습된 언어모델로 Hugging Face[1] 라이브러리에서 배포한 BERT 모델을 사용했으며 배치 사이즈 128, 학습률 3e-5를 사용했다. 다른 학습 관련 초파라미터는 기존 BERT 모델과 동일하게 설정했다. 집중학습을 위한 초파라미터 α는 1.0으로 설정했다. 텍스트 추출과 사전학습 시 BERT에서 사용한 말뭉치를 사용했다. 실험은 Pytorch 라이브러리를 통해 구현되었으며 NVIDIA RTX 3090 네 대를 사용하여 이루어졌다.

실험 결과는 표 1, 2에 보고되어 있다. 실험 결과에 따르면 BERT-base와 BERT-large 두 개의 모델에서 우리가 제안한 집중학습기법이 다른 기법들을 뛰어넘는 성능을 달성했다. 특히 집중학습기법은 일반사전학습 대비 0.3% 수준의 말뭉치만 사용해서 더 높은 성능을 달성했다. 이 결과로부터 집중학습기법이 말뭉치 크기 측면에서 더 효율적으로 성능을 향상시킨다는 것을 확인할 수 있다. 특히 일반사전학습으로 20배의 계산량(FLOPs)을 사용하여 모델을 학습시켰을 때에도 집중학습기법이 더 높은 성능을 달성했다. 이로부터 우리가 제안한 집중학습기법이 일반사전학습 기법보다 더 효율적인 기법임을 확인할 수 있다. 그리고 지식 함유 텍스트가 아니라 임의의 문장들을 선별해서 학습했을 때 다른 조건이 모두 같음에도 GLUE 점수가 가장 떨어지는 현상을 관측했다. 이로부터 우리가 선별한 지식 함유 텍스트가 사전학습 말뭉치에 들어있는 텍스트보다 평균적으로 성능 향상에 더 기여한다는 것을 알 수 있다.

학습 기법	말뭉치 크기	FLOPs	GLUE 점수
일반사전학습	16GB	9.6e15	82.0
일반사전학습 20x	16GB	1.9e17	82.8
임의추출학습	50MB	9.6e15	81.7
집중학습	50MB	9.6e15	**83.1**

표 1 BERT-base GLUE 점수

학습 기법	말뭉치 크기	FLOPs	GLUE 점수
일반사전학습	16GB	2.9e16	83.8
일반사전학습 20x	16GB	5.7e18	84.3
임의추출학습	50MB	2.9e16	82.2
집중학습	50MB	2.9e16	**84.5**

표 2 BERT-large GLUE 점수

Ⅲ. 결론

본 논문에서는 많은 지식을 담고 있는 텍스트를 집중적으로 사전학습하는 집중학습 기법을 제안하여 말뭉치가 클수록 언어모델 성능이 좋아진다는 기존 전제가 성립하지 않는 경우를 확인했다. 본 논문의 결과는 현재 계속해서 늘어나고 있는 사전학습 말뭉치의 크기와 사전학습 계산량을 크게 줄일 수 있다는 가능성을 보여준다. 추후 우리는 집중학습 기법을 더 다양한 모델과 지식 그래프, 사전학습 말뭉치에 적용할 수 있도록 확장할 계획이다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터육성지원사업의 연구결과로 수행되었음 (IITP-2021-2016-0-00464)

참 고 문 헌

[1] Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).

[2] Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems. 2017.

[3] Wang, Alex, et al. "GLUE: A multi-task benchmark and analysis platform for natural language understanding." arXiv preprint arXiv:1804.07461 (2018).

[4] Liu, Yinhan, et al. "Roberta: A robustly optimized bert pretraining approach." arXiv preprint arXiv:1907.11692 (2019).

[5] Brown, Tom B., et al. "Language models are few-shot learners." arXiv preprint arXiv:2005.14165 (2020).

[6] Speer, Robyn, Joshua Chin, and Catherine Havasi. "Conceptnet 5.5: An open multilingual graph of general knowledge." Thirty-first AAAI conference on artificial intelligence. 2017.

[7] Mintz, Mike, et al. "Distant supervision for relation extraction without labeled data." Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP. 2009.

1) https://huggingface.co/

마비말장애 중증도 자동 분류를 위한 퓨샷 러닝

여은정, 정민화*
서울대학교

ej.yeo@snu.ac.kr, *mchung@snu.ac.kr

Few-shot learning for automatic severity classification of dysarthric speech

Yeo Eun Jung, Chung Minhwa*
Seoul National Univ.

요 약

마비말장애 중증도 자동 분류 기술은 객관적이고 일관적인 결과를 산출함으로써 언어재활사의 평가를 보조할 수 있다. 그러나 마비말장애 음성은 대량의 균형 코퍼스 수집이 어렵기 때문에 딥러닝 모델 사용에 제한이 있다. 따라서 본 논문에서는 다양한 소규모 이미지 데이터 분류 과제에서 좋은 성능을 보이고 있는 퓨샷 러닝 방법론을 소규모 음성 데이터 분류 과제인 마비말장애 중증도 자동 분류에 적용할 것을 제안한다. 우선, 본 논문은 음성을 시각화한 스펙트로그램을 사용하여 장애 중증도 분류를 시도한다. 더 나아가, 마비말장애 중증도 분류에 도움이 되는 것으로 알려진 음질 특징과 운율 특징을 함께 사용할 것을 제안한다. 실험 결과, 퓨샷 러닝을 사용한 모델이 베이스 모델인 Resnet 또는 파인 튜닝한 모델보다 더 높은 분류 정확도를 보였다. 또한 스펙트로그램을 단독으로 사용했을 때보다 스펙트로그램과 음질 특징, 운율 특징을 함께 사용했을 때 더 높은 성능이 나타났다. 본 연구는 마비말장애 중증도 자동 분류를 위한 퓨샷 러닝의 초기 연구로서, 장애 음성 분류에 대한 퓨샷 러닝 방법론의 적용 가능성과 방향성을 탐구한 것에 그 의의가 있다.

Ⅰ. 서 론

마비말장애는 신경계 손상으로 말소리 산출과 관련된 근육이 마비되거나 약해지면서 발생하는 말운동장애이다. 마비말장애는 선천적 뇌손상, 후천적 뇌손상, 노화에 따른 퇴행성 신경질환 등 넓은 범위에서 나타나며, 호흡, 발성, 공명, 조음, 운율 등 다양한 말 기능에 영향을 미친다. 언어재활사는 마비말장애 중증도 평가를 통해 화자가 의사소통 상황에서 겪는 어려움의 정도를 확인하고, 화자의 상황에 알맞은 중재 방안을 모색한다. 그러나 임상 현장에서 주로 사용되는 말 명료도 평가는 청지각적 평가로, 주관적일 뿐만 아니라 많은 시간과 노력이 소요된다. 따라서 숙련된 언어재활사의 평가 결과와 유사한 결과를 산출하는 마비말장애 중증도 자동 분류 기술은 객관적이고 일관적인 결과를 바탕으로 언어재활사의 평가에 유용하게 사용될 수 있다.

최근 딥러닝 기술이 발전함에 따라 마비말장애 중증도 자동 분류 연구에서도 딥러닝 방법론이 시도되고 있다. 그러나 장애 발화는 대규모 코퍼스 수집이 어렵기 때문에 딥러닝 기술 적용이 쉽지 않은 실정이다. 소규모 데이터를 딥러닝 모델 학습에 사용할 경우, 과적합(overfitting)이 발생하기 쉽기 때문이다. 따라서 딥러닝 모델을 사용한 대부분의 선행연구들은 대개 비장애 음성과 장애 음성만을 구분하는 장애 음성 검출 연구에 그치고 있다.[1][2]

소규모 데이터로 인한 과적합 문제에 대한 해결 방안으로는 우선 사전 학습(pre-trained)된 모델의 층 일부를 훈련하는 파인 튜닝(fine-tuning)이 있다. 훈련하는 모델 층의 개수는 훈련 데이터의 양과 종류에 따라 다르게 설정할 수 있는데, 장애 발화에 대한 파인 튜닝은 과적합 문제에서 자유롭지 못하다. 대부분의 장애 발화 실험의 경우 학습 데이터가 사전 훈련에 사용된 데이터와 속성이 다르면서 분량 역시 매우 적기 때문이다.

두 번째 해결책으로는 퓨샷 러닝(few-shot learning)이 있는데, 모델에게 데이터를 구분하도록 학습시킴으로써 단 몇 개의 데이터만으로도 분류 과제를 시행할 수 있는 방법론이다. 모델의 파라미터가 아닌 데이터 간의 거리(유사도) 또는 관계를 학습하기 때문에 과적합의 우려가 다소 적다. 최근 들어 데이터 라벨링이 어려운 소규모 데이터 분류 과제를 위해 다양한 퓨샷 러닝 접근법이 제안되고 있지만 대개 이미지 분류에 치우쳐 보고되었다. 본 연구에서는 퓨샷 러닝 접근법을 소량의 음성 데이터 분류 과제인 마비말장애 중증도 분류에 적용하여 그 효과를 살펴보고자 한다.

Ⅱ. 본 론

i) 모델 구조

그림 1 은 본 연구에서 제안하는 모델의 구조도이다. 퓨샷 러닝을 위한 다양한 모델 중 프로토타입 네트워크(Prototypical Network)[3]를 사용하였는데, 서포트 데이터의 평균 위치에 있는 프로토타입을 사용해 장애 발화처럼 불균형한 데이터셋에도 좋은 성능을 보이는 것으로 보고되기 때문이다.

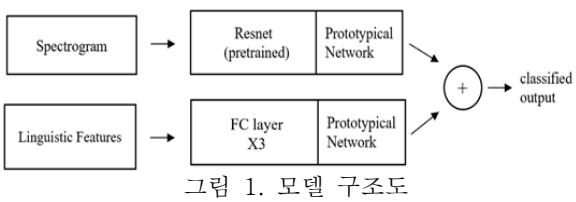


그림 1. 모델 구조도

모델은 크게 두 갈래로 구성된다. 첫 번째 갈래는 이미지를 사용한 분류로, 각 음성을 스펙트로그램으로 시각화해 입력 값으로 사용한다. 이미지 분류를 위해서는 베이스 모델(backbone model)로 ImageNet 데이터셋으로 사전 학습된 Resnet 모델을 사용하고, flatten 한 Resnet 의 마지막 층에 프로토타입 네트워크를 잇는다. 두 번째 갈래는 언어학적 특징을 이용한 분류로, 각 음성의 음질 특징, 운율 특징을 입력 값으로 사용한다. 베이스 모델로는 3 개의 완전연결층(fully-connected layer)을 쌓은 MLP 를 사용하고, 후단에 프로토타입 네트워크를 쌓는다.

각 갈래의 출력값은 프로토타입과 쿼리 데이터 간의 거리로, 두 출력값의 합이 마비말장애 중증도 분류에 사용된다. 모델은 음성을 5 분류하며, 장애 중증도는 비장애, 경도, 경도-중등도, 중등도-중도, 중도로 구분된다.

ii) 사용 특징

본 연구에서 사용된 특징은 스펙트로그램과 언어학적 특징으로 구분된다. 우선 스펙트로그램은 시간에 따른 주파수와 진폭에 대한 변화를 시각화한 그림으로, 음성의 발음 정보를 담고 있다. librosa 0.8.를 통해 멜 필터뱅크(mel filterbanks) 개수 400 개, 최고 주파수 8000Hz 로 추출하였다.

마비말장애 중증도 분류에는 발음 정보뿐만 아니라 음질 정도와 운율 정보도 함께 고려되어야 한다. 음질 특징은 음성 장애를 진단할 때 사용되는 jitter, shimmer, HNR (Harmonics to Noise Ratio)와 화자의 유성음 유지 능력을 확인할 수 있는 voice breaks 의 개수와 정도를 포함한다. 모든 음질 특징은 Praat 6.1.16 을 사용하여 추출하였다. 운율 특징은 음높이 특징과 발화 속도 특징으로 구분된다. 음높이 특징은 F0 의 평균값, 표준편차, 최솟값, 최댓값, 중앙값, 25 분위수, 75 분위수를, 발화 속도 특징은 전체 길이, 말 길이, 말 속도, 조음 속도를 포함한다. 특징들은 각각 Praat 6.1.16 과 Parselmouth 0.4.0 를 사용하여 추출하였다.

iii) 코퍼스

본 연구에서는 뇌성마비 마비말장애인 대상 음성인식기 개발을 위해 구축된 한국어 코퍼스인 QoLT 를 사용하였다. 코퍼스는 고립단어와 문장으로 구성되지만, 본 연구에서는 음질 특징, 운율 특징, 발음 특징 모두를 종합적으로 평가에 반영하기 위해 문장만을 사용하였다. 화자는 비장애인 화자 10 명(남 5 명, 여 5 명)과 마비말장애 화자 70 명(남 45 명, 여 25 명)으로 구성되며, 모든 화자는 5 개의 문장을 2 회 반복 발화하였다. 화자의 장애 중증도는 2 년 이상의 임상 경험이 있는 언어재활사 5 명이 모든 문장을 듣고 판단한 것으로, 각 장애 중증도 별 화자 수는 경도, 경도-중등도, 중등도-중도, 중도 순으로 각각 14 명, 14 명, 9 명, 4 명이다. 분류 실험에는 훈련 화자, 검증 화자, 테스트 화자를 중증도 집단 별 6:2:2 로 구성하였으며, 각 과정에서 화자는 겹치지 않았다.

Ⅲ. 실 험 및 실 험 결 과

본 모델은 5-way 5-shot 7-query 태스크로 훈련되었다. 훈련과 검증은 각 에피소드마다 서포트 데이터와 쿼리 데이터를 각 중증도 별로 랜덤하게 선택하였으며, 이 과정을 훈련과 검증 단계에서 각 5000 회를 반복하였다. 옵티마이저(optimizer)로는 Adam 을 사용하였으며, learning rate 는 0.01 이었다.

실험 결과, 퓨샷 러닝을 적용했을 때 분류 정확도 67.42%로 베이스라인인 사전 학습된 Resnet 의 33.52% 또는 파인 튜닝의 52.94%보다 더 높은 분류 정확도를 보였다. 더 나아가, 스펙트로그램만을 입력값으로 사용했을 때와 스펙트로그램과 음질, 운율 특징을 함께 사용했을 때의 분류 정확도를 비교했을 때, 각각 62.39%, 67.42%로 이미지와 음향학적 특징을 함께 사용했을 때 8.06%의 상대적 증가율을 보였다.

Ⅵ. 결 론

본 연구는 마비말장애 중증도 자동 분류를 위해 스펙트로그램과 음향학적 특징(음질 특징, 운율 특징)을 입력값으로 사용하는 퓨샷 러닝 모델을 제안하였다. 본 연구의 최고 성능은 67.42%로, QoLT 코퍼스를 사용한 선행연구[4]의 SVM 분류기 최고 성능인 80.15% 또는 MLP 분류기의 최고 성능인 78.09%보다 낮은 성능이었다. 이는 본 연구에서 사용하지 않은 음소 단위의 발음 특징을 두 실험에서는 포함했기 때문이라고 추측된다. 향후 연구에서는 음소 단위의 발음 특징도 함께 특징으로 포함하여 그 효과를 살펴보아야 할 것이다. 반면, 본 연구와 동일한 범주인 프레임 단위의 특징(MFCCs 또는 스펙트로그램), 음질 특징, 운율 특징만을 사용한 연구 중에서는 본 논문이 가장 높은 성능을 보였으며, 이는 본 선행연구[5]의 MLP 성능보다 8.61% 상대적 증가한 값이다.

본 연구는 기존 소규모 이미지 데이터 분류 과제에서 좋은 성능을 보이는 퓨샷 러닝 방법론을 소규모 음성 데이터 분류 과제인 마비말장애 중증도 자동 분류 과제에 적용한 초기 연구로서 소규모 장애 음성 분류에 대한 퓨샷 러닝 방법론의 적용 가능성과 방향성을 탐구한 것에 그 의의가 있다.

ACKNOWLEDGMENT

"본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT 연구센터육성지원사업의 연구결과로 수행되었음" (IITP-2021-2016-0-00464)

참 고 문 헌

[1] Bhat, C., & Strik, H. (2020). Automatic assessment of sentence-level dysarthria intelligibility using BLSTM. *IEEE Journal of Selected Topics in Signal Processing, 14*(2), 322-330.

[2] Millet, J., & Zeghidour, N. (2019, May). Learning to detect dysarthria from raw speech. " *IEEE International Conference on Acoustics, Speech and Signal Processing*, 5831-5835.

[3] Snell, J., Swersky, K., & Zemel, R. (2017, December). Prototypical networks for few-shot learning. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 4080-4090.

[4] Yeo, E. J., Kim, S., & Chung, M. (2021). Automatic severity classification of dysarthria using voice quality, prosody, and pronunciation features. *Phonetics and Speech Sciences, 13*(2), 57-66.

[5] Hernandez, A., Kim, S., & Chung, M. (2020). Prosody-based measures for automatic severity assessment of dysarthric speech. *Applied Sciences, 10*(19), 6999.

어텐션 쌍방향 LSTM 신경망을 활용한 억양 기반 한국어 방언 자동 식별

이주영[1], 김경화[2], 정민화[1]
서울대학교[1], 대검찰청[2]
excalibur12@snu.ac.kr, savoix@spo.go.kr, mchung@snu.ac.kr

Korean Dialect Identification based on Intonation using an Attention BiLSTM network

Lee Joo Young[1], Kim Kyungwha[2], Chung Minhwa[1]
Seoul National University[1], Supreme Prosecutor's Office[2]

요 약

한국어 방언 식별 연구는 소규모 데이터를 대상으로 방언학과 실험음성학에서 질적 연구로서 꾸준히 이어져 왔으나, 데이터가 부족하였기 때문에 기계학습 방법론을 적용한 방언 자동 식별 연구는 이루어지지 않았다. 본 논문에서는 대검찰청 보유 '한국인 표준 음성 DB'와 국립국어원 방언 DB의 대용량 데이터, 한국어 방언 식별 표지인 억양 특징, 그리고 방언 구간을 감지하여 학습하는 시계열 신경망 어텐션 쌍방향 LSTM 모델을 활용하여 한국어 방언 자동 식별 연구를 진행한다. 음의 높낮이 변화로 표현되는 억양은 F0(기본주파수)를 중심으로 프레임 단위의 여러 음향 특징을 LSTM 모델에서 학습하는 형태로 실현하였고, 연령대 및 발화유형별로 식별 성능을 확인하였다. 성능으로는 두 방언끼리 비교하는 방식으로 학습한 제안 모델이 베이스라인에 비해 평균 25.34%p 성능 향상을 보였다.

Ⅰ. 서론

방언 식별(dialect identification)은 음성으로부터 추출한 음향 패턴을 통해 화자의 방언을 자동으로 추정하는 연구 분야이다. 방언 식별을 언어 식별(language identification)의 하위 분야로 볼 수 있으나, 방언은 동일 언어의 음운 체계 안에서 유사한 언어적 특징을 보이기 때문에 언어 식별보다 연구가 더 까다롭다.

한국어 방언 식별 역시 연구가 까다로운 분야이다. 방언의 특징을 언어학적으로 기술하는 연구는 전통 방언학과 실험음성학에서 지속적으로 이루어져 왔으나 수집된 데이터의 규모가 작아 특정 지역을 대상으로 여러 층위(음운론, 형태론, 통사론 등)의 언어학적 분석을 하거나 인접한 두 지역 방언의 특징 차이를 비교하는 등 소규모의 질적 연구가 중심이었다. 이처럼 데이터가 많지 않았기 때문에 딥러닝 모델을 활용한 자동 식별은 실현이 어려웠다.

음성은 크게 분절적(segmental) 요소와 초분절적(suprasegmental) 요소로 나눌 수 있다. 분절적 요소는 개별 음절(syllable)의 모음과 자음에 해당하고 초분절적 요소는 음성 전체에 얹히는 운율에 해당한다. 운율 특징은 다시 억양(intonation), 리듬(rhythm), 강세(stress)로 나눌 수 있으며, 이 중 억양 특징은 한국어 방언을 구분하는 척도로 사용할 수 있다 [1]. 억양은 음의 높낮이 변화로, 방언 식별 연구에서 실현하기 위해서는 높낮이 변화를 감지하고 학습할 수 있는 모델이 필요하다.

본 연구에서는 억양 정보와 LSTM 계열 인공신경망을 활용하여 한국어 방언 식별 실험을 진행한다. 실험에 사용하는 음성 데이터는 대검찰청의 대용량 코퍼스인 '한국인 표준 음성 데이터베이스'와 국립국어원의 방언 코퍼스이다. 억양 정보를 표현하기 위해 음향 특징으로는 음성의 프레임 단위에서 추출하는 F0, MFCC, 에너지, zero crossing rate 및 이들의 1차 미분값을 합한 30차원 특징 벡터를 사용하며, 음향 특징 벡터를 어텐션 쌍방향 LSTM 모델에 입력하여 학습하는 방식으로 진행한다. 이때, 어텐션 기법은 발화 내 방언 특징이 두드러진 구간을 학습하기 위해 사용한다. 모델의 예측 단계에서는 하나의 모델이 6개 방언을 분류하는 방법, 특정 방언과 나머지 방언으로 나누어 분류하는 방법(One-versus-Rest; OvR), 그리고 두 방언끼리 비교 분류하는 방법(One-versus-One; OvO)으로 나누어 결과를 확인한다.

논문의 나머지 구성은 다음과 같다. Ⅱ장에서는 실험 코퍼스로 사용하는 두 개 데이터셋을 소개하고, Ⅲ장에서는 억양 기반 한국어 방언 식별 실험 설계 과정에 대해 설명한다. Ⅳ장에서는 실험 결과에 대한 분석을 하고, Ⅴ장에서 본 연구를 정리하고 결론을 내린다.

Ⅱ. 실험 코퍼스

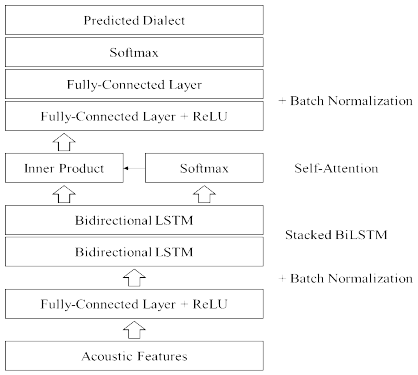
본 연구에서 사용하는 음성 데이터는 크게 대검찰청에서 3년간 수집한 '한국인 표준 음성 데이터베이스'(이하 KSS DB) [2]와 국립국어원에서 수집한 데이터베이스(이하 NIKL DB)이다.

KSS DB는 2,500명 이상의 한국인 화자로 구성된 대용량 음성 코퍼스로 화자별 성, 연령, 지역 정보가 표시되어 있다. 성은 남성과 여성 2 분류, 연령은 20대, 30대, 40대, 50대, 60대 이상의 5 분류, 지역은 수도권, 경남권, 경북권, 전남권, 전북권, 충남권, 충북권, 강원권, 제주권의 9 분류이다. 또한, 화자마다 낭독발화와 자유발화 음성이 있어 다양한 실험이 가능하여, 본 연구에서는 연령대와 발화 유형에 따른 식별 실험에 사용한다.

NIKL DB는 약 150개 음성 파일로 구성된 40시간의 자유발화 방언 데이터로 70세 이상의 고령층 화자로 구성되어 있다. 고령층 화자로 이루어진 코퍼스인 만큼 녹음된 발화에서 드러나는 방언 특색이 매우 짙으며, 따라서 식별 실험에서는 NIKL DB 추가 여부에 따른 식별 성능의 차이를 알아보는 용도로 사용한다.

Ⅲ. 억양 기반 한국어 방언 식별 실험 설계

본 연구의 대상 방언의 수는 6개이다. 이는 경상권, 전라권, 충청권 방언



〔그림 1〕 Attn-BiLSTM 모델 구조도

권의 남북 지역을 각각 합한 것이며, 남북 지역의 방언적 차이가 존재함에도 방언별 데이터 확보와 식별 성능 향상을 위해 분류 가지 수를 9개 방언권에서 6개 방언권으로 축소하였다.

[그림 1]은 한국어 방언 식별에 사용한 어텐션 쌍방향 LSTM (Attn-BiLSTM) 모델 구조도이다. 본 구조의 기본 틀은 감정 분류 실험을 한 [3]에 바탕을 두고 있으나, 자체 실험을 통해 [그림 1]과 같은 형태의 Stacked BiLSTM 층과 Batch Normalization 층을 추가하였다. 어텐션 층은 Stacked BiLSTM 층 바로 위에 배치하여 LSTM 층에서 출력된 벡터에 자기자신의 softmax 가중치 값을 적용하는 self-attention 형태로 만들었다.

음향 특징 또한 [3]를 참고하여 25ms 프레임 단위 F0, MFCC, 에너지, zero crossing rate와 이들의 1차 미분값(first-order derivatives)을 합한 30차원 벡터로 구성하였고, [3]의 음향 특징 셋과 달리 성능에는 영향을 주지 않고 학습 시간만 늘리는 voicing probability는 제외하였다.

[표 1]은 KSS DB와 NIKL DB 데이터셋의 학습과 테스트 데이터 분포이다. 표에서 Utt.와 Spon.은 각각 발화(utterance)와 자유발화(spontaneous speech)를 의미하며, 발화당 길이는 평균 3-5분이다.

실험에 반영한 연령대는 크게 20-40대로 구성된 젊은 층 그룹(young)과 40대 이상의 장노년층 그룹(old)으로 나누었으며, 40대가 양 그룹에 속해 있는 것은 실험 과정에서 40대를 배제하거나 한쪽 그룹에만 속해 있을 때보다 높은 성능을 보였기 때문이다.

프레임 단위 모델 학습 후 테스트 데이터에 대한 평가는 발화 단위로 하였다. 원본 데이터가 3-5분 길이이기 때문에 이를 30초 구간씩 분절하였다. 따라서 실제 테스트 데이터 샘플 수는 [표 1]의 테스트 발화(3-5분 발화) 수보다 많다.

Attn-BiLSTM 모델은 기본적으로 2분류 방식으로 실험하였는데, '특정 방언' vs '나머지 방언'으로 학습한 6개 모델을 사용한 OvR 방식과 '방언 1' vs '방언 2'로 학습한 15개 모델을 사용한 OvO 방식으로 하였다.

〔표 1〕 실험 데이터 분포

Dataset		Train	Test
KSS-DB	# of Utt.	6,218	1,808
Read	Size (hrs)	309.8	67.4
KSS-DB	# of Utt.	2,049	438
Spon.	Size (hrs)	114.8	24.7
NIKL-DB	# of Utt.	151	-
Spon.	Size (hrs)	40.1	-

Ⅳ. 결과 및 분석

[표 2]는 베이스라인 1 [4], 베이스라인 2, OvR 제안 모델, OvO 제안 모델을 각각 사용하였을 때 연령대 및 발화유형별 식별 정확도를 나타낸 것이다. 베이스라인 1은 [4]의 단순 형태의 6분류 LSTM 모델이며 음향 특징은 F0와 MFCC만을 사용했다. 베이스라인 2는 본 연구에서 진행한 실험 중 6분류 실험을 Attn-BiLSTM 모델로 한 것이다. [표 2]의 모델별, 연령대 + 발화유형별 결과를 비교하면, 모든 경우에서 OvO 제안 모델의 성능이 월등히 좋았다. 이는 두 개 방언만 비교하면 되는 모델들로 구성되어 있어 방언 간 분류를 극대화할 수 있었기 때문으로 본다. 이에 반해 OvR 제안 모델은 '나머지 방언'의 데이터 양의 비율이 지나치게 높았기 때문에 모델이 '나머지 방언'으로만 학습하는 경향이 생겨 성능이 오히려 베이스라인들보다 낮게 나타났다. 또한, Attn-BiLSTM 모델이 포함된 실험들은 국립국어원 NIKL DB를 포함한 경우의 정확도도 확인하였는데, 베이스라인 2, OvR 제안 모델, OvO 제안 모델 모두 NIKL DB를 포함하지 않았을 때의 성능보다 포함하였을 때의 성능이 더 높았다. 데이터 양이 추가된 점도 있으나 방언 특징이 짙은 데이터가 추가된 점도 성능 향상의 이유가 될 수 있다. 전체적으로 가장 성능이 좋은 OvO 제안 모델은 베이스라인 1에 비해 평균 25.34%p 정확도가 높았다.

〔표 2〕 한국어 방언 식별 실험 결과

평가 방식	연령대 + 발화유형	발화 단위 정확도(%)
베이스라인 1 [4] 6분류 단순 LSTM	(1) Young + Read	42.06
	(2) Old + Read	46.19
	(3) Young + Spon.	25.85
	(4) Old + Spon.	39.17
베이스라인 2 6분류 Attn-BiLSTM	(1) Young + Read	36.79
	(2) Old + Read	32.71
	(3) Young + Spon.	31.95
	(4) Old + Spon.	29.97
	(5) (4) + NIKL-DB	31.38
제안 모델 Attn-BiLSTM + OvR	(1) Young + Read	25.62
	(2) Old + Read	30.38
	(3) Young + Spon.	26.10
	(4) Old + Spon.	33.82
	(5) (4) + NIKL-DB	35.23
제안 모델 Attn-BiLSTM + OvO	**(1) Young + Read**	**62.34**
	(2) Old + Read	**66.21**
	(3) Young + Spon.	**58.74**
	(4) Old + Spon.	**67.33**
	(5) (4) + NIKL-DB	**68.51**

Ⅴ. 결론

본 연구에서는 한국어 방언 표지인 억양을 프레임 단위 음향 특징과 어텐션 쌍방향 LSTM 학습 방식을 통해 모델링 하여, 대용량 음성 코퍼스에서 한국어 방언 자동 식별 실험을 진행하였다. 발화 내 방언 특징 구간을 잘 학습하기 위해 어텐션 기법을 사용하였고, 연령대 및 발화 유형별로 실험을 나누어 진행하였으며, 예측 방식으로 OvR과 OvO 방식을 Attn-BiLSTM 모델과 함께 사용하는 것을 제안하였다. 모델 성능으로는 OvO 기반 Attn-BiLSTM 모델이 가장 우수하였고, 베이스라인에 비해 평균 25%p 성능이 향상하였다.

ACKNOWLEDGMENT

"본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터육성지원사업의 연구결과로 수행되었음" (IITP-2021-2016-0-00464)

참 고 문 헌

[1] 이병운. (1998). 중부방언, 경남방언, 전남방언의 억양에 대한 비교 연구. 우리말연구, 8.

[2] 신지영, 김경화. (2017). 한국인 표준 음성 DB 구축 (II). 말소리와 음성과학, 9(2), 9-22.

[3] Mirsamadi, S., Barsoum, E., Zhang, C. (2017, March). Automatic speech emotion recognition using recurrent neural networks with local attention. In 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 2227-2231). IEEE.

[4] Lee, J., Kim, K., Lee, K., Chung, M. (2019). Gender, Age, and Dialect Identification for Speaker Profiling. In 2019 22nd Conference of the Oriental COCOSDA International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques (O-COCOSDA).

부동소수점 형식에 따른 스파이킹 신경망의 학습 및 정확도 분석

곽명진, 서효주, 김용태*
경북대학교 컴퓨터학부

*yongtae@knu.ac.kr

Training and Accuracy Analysis of Spiking Neural Network using Floating-Point Formats

Myeongjin Kwak, Hyoju Seo, and Yongtae Kim*
School of Computer Science and Engineering, Kyungpook National University

요 약

본 논문은 에너지 효율적인 스파이킹 신경망(spiking neural network) 구현을 위해 다양한 부동소수점 형식(floating-point format)의 정밀도에 따른 학습의 정확도를 체계적으로 분석한다. IEEE 754 기반의 32 비트 단정밀도 부동소수점 및 비트폭이 축소된 부동소수점 형식을 스파이킹 신경망의 Leaky Integrate-and-Fire (LIF) 뉴런에 적용하여 축소된 부동소수점 형식이 스파이킹 신경망에 미치는 영향을 분석한다. 손글씨 이미지 MNIST 데이터를 분류하기 위한 두 개의 레이어로 구성된 스파이킹 신경망에 다양한 정밀도의 부동소수점 형식을 적용하여 비지도 학습을 수행하고 결과를 분석한다. 지수비트가 4 비트 이상인 16 비트폭의 부동소수점 형식은 32 비트 단정밀도 부동소수점 형식과 유사한 정확도로 학습할 수 있다. 또한, 신경망의 학습을 위해서는 최소 지수 4 비트와 가수 6 비트가 필요하며 이를 하드웨어로 구현할 경우 다른 부동소수점 형식에 비해 면적, 전력, 및 에너지 효율이 우수하다.

Ⅰ. 서 론

인공 신경망(artificial neural network) 발전으로 다양한 분야에 기계 학습(Machine Learning) 기술이 적용되고 있다. 이를 처리해야 할 데이터도 급격하게 증가하였고 필요한 컴퓨팅 자원과 전력/에너지 소모도 지속적으로 증가하고 있다. 이러한 문제를 해결하기 위한 방법 중의 하나로 뉴로모픽 컴퓨팅(Neuromorphic Computing) 기술이 있다. 이는 포유류의 뇌를 모방하는 기술로 스파이킹 신경망(spiking neural network)으로 구현할 수 있다. 생물학적 뇌는 데이터 처리에 대하여 매우 우수한 전력 및 에너지 효율성을 보여주고 이러한 신경구조를 활용한 뉴로모픽 컴퓨팅 기술은 전력/에너지 효율적인 컴퓨팅을 위한 해결책이 될 수 있다 [1].

스파이킹 신경망의 뉴런은 Hodgkin-Huxley, Izhikevich, 그리고 Leaky-Integrated and Fire (LIF) 모델 등을 사용하여 구현할 수 있다. 이러한 뉴런 모델을 적용한 신경망은 막 전위(membrane potential)가 문턱 전압(threshold voltage)보다 높을 때 스파이크(Spike)로 반응이 전달되는 이벤트 기반 방식이기 때문에 높은 전력 효율을 가진다. 이는 문자 인식, 컴퓨터 비전, 그리고 음성인식 등 다양한 분야에서 사용될 수 있다.

스파이킹 신경망은 GPU (Graphic Processing Unit) 등을 사용하여 32-비트 단정밀도부동소수점 형식(single precision floating-point format)의 연산을 수행하여 동작한다 [2]. 이때 부동소수점 연산과 표현을 위해 많은 하드웨어 자원이 필요하다. 32-비트 단정말도 부동소수점보다 낮은 정밀도를 가지는 부동소수점 형식과 연산으로 학습이 가능하다면 하드웨어 자원을 크게 절감할 수 있다. 따라서 스파이킹 신경망의 전력/에너지 효율성을 위해 다양한 정밀도의 부동소수점 연산을 스파이킹 신경망에 적용하여 성능을 평가하는 것은 매우 중요하다.

본 연구에서는 전력/에너지 효율적인 스파이킹 신경망 프로세서를 구현을 위해 다양한 정밀도의 부동소수점 형식을 스파이킹 신경망에 적용하고 정확도에 미치는 영향을 분석한다. LIF 뉴런 모델을 사용한 스파이킹 신경망에서 대표적인 비지도 학습 방법인 STDP (Spike Timing Dependent Plasticity)를 사용하여 손글씨 이미지 MNIST 데이터를 학습하였다. 실험 결과로 학습을 위해 적어도 4-비트 지수(exponent)와 6-비트 가수(mantissa)가 필요한 것을 확인하였다.

Ⅱ. 부동 소수점 형식

현대의 컴퓨터에서 사용되는 부동소수점 형식은 IEEE-754 에 정의되어 있으며, 부호, 지수, 그리고 가수 부분으로 구성된다. 단정밀도 부동소수점(FP32)은 부호, 지수, 그리고 가수가 각각 1-비트, 8-비트, 그리고 23-비트로 구성된다 [3]. 표 1 은 다양한 부동소수점 형식과 각 부분의 비트 수를 나열한다. Tensorflow-32 (TF32) 형식과 Brain floating-point (BF16) 형식은 FP32 와 지수 비트의 크기는 같으나 가수 비트는 각각 10-비트 그리고 7-비트로 구성된다 [4]. 반정밀도 부동소수점(FP16)과 DLFloat(DLF16) 형식은 지수와 가수부분에서 1 비트의 차이가 있다 [5]. 또한, EFloat16 Average (EF16_A) 형식은 지수 3-비트, 가수 12-비트로 구성되며, 8-비트 부동소수점 형식인 Minifloat(FP8)은 지수 4-비트, 가수 3-비트로 구성된다 [6].

표 1. 비트폭에 따른 다양한 부동 소수점 형식.

Format	*sign*	*exponent*	*mantissa*	*total*
FP32	1	8	23	32
TF32	1	8	10	19
FP16	1	5	10	16
BF16	1	8	7	16
DLF16	1	6	9	16
EF16_A	1	3	12	16
FP8	1	4	3	8

Ⅲ. 실험 결과

다양한 부동소수점 형식을 적용하기 위한 학습 모델로 [7]에서 제시된 스파이킹 신경망의 구조를 사용하였다. 첫번째 레이어는 입력 레이어로 784 개의 뉴런으로 구성되며, 두번째 레이어는 흥분성(excitatory) 뉴런과 억제성(inhibitory) 뉴런으로 구성되며 각각의 뉴런들의 개수는 동일하다. 두번째 레이어는 400 개와 1600 개의 흥분성 뉴런과 억제성 뉴런으로 구성하여 실험하였다. LIF 뉴런 모델을 사용하였고 해당 연산에 다양한 정밀도의 부동소수점 형식을 적용하였다. 학습은 훈련 데이터 6 만개와 테스트 데이터 1 만개를 사용하여 진행하였다.

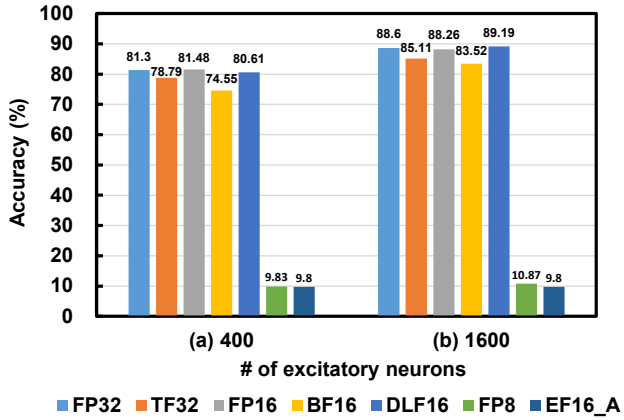


그림 1. 부동소수점 형식에 따른 스파이킹 신경망의 정확도; (a) 흥분성 뉴런 400 개, (b) 흥분성 뉴런 1600 개 일 때 정확도.

그림 1 은 다양한 정밀도를 가지는 부동소수점 형식을 적용하여 학습한 스파이킹 신경망의 정확도를 보여준다. FP32 는 뉴런이 400 개, 1600 개 일 때 각각 81.3%, 88.6%의 정확도를 가진다. FP8 과 EF16_A 를 제외한 부동소수점 형식에서는 FP32 와 비슷한 정확도를 가진다. TF32, FP16, BF16, 그리고 DLF16 을 적용한 뉴런이 400 개인 스파이킹 신경망에서 각각 78.8%, 81.5%, 74.6%, 그리고 80.6%의 정확도를 가지며, 1600 개 일 때 각각 85.1%, 88.3%, 83.5%, 89.2%의 정확도를 가진다. FP8 과 EF16_A 를 적용하였을 경우의 정확도는 뉴런이 400 개 일 때 모두 9.8%, 1600 개 일 때 각각 10.9%, 9.8%로 정확도가 크게 감소하여, 정상적으로 손글씨 분류가 되지 않는 것을 확인할 수 있다.

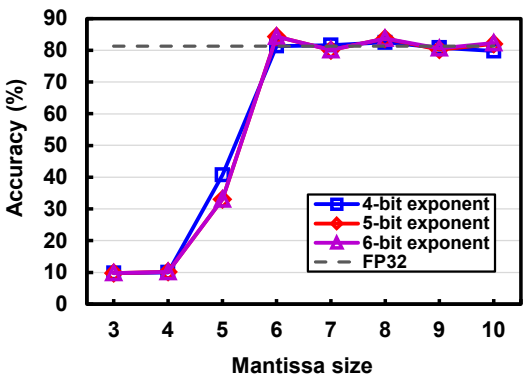


그림 2. 부동소수점 비트폭에 따른 스파이킹 신경망의 정확도.

스파이킹 신경망의 학습 정확도를 유지하기 위한 최소의 지수와 가수 비트 크기를 결정하기 위해 지수의 크기를 4, 5, 그리고 6 비트로 고정하고 가수의 크기를 변경하며 학습의 정확도를 측정하였다. 그림 2 는 흥분성 뉴런이 400 개로 구성된 스파이킹 신경망에서 지수와 가수의 크기에 따른 정확도를 보여준다. 지수가 4-비트 이상일 때 가수가 6-비트 이상이면 FP32 형식과 유사한 정확도를 보인다. 하지만, 가수가 5-비트 이하 일 때 정확도가 급격하게 감소함을 확인할 수 있다. 따라서, LIF 뉴런을 적용한 스파이킹 신경망에서 정확도의 영향 없이 학습하기 위해서 최소 4-비트 지수와 6-비트 가수가 필요한 것을 확인할 수 있다.

표 2. 다양한 부동소수점 형식의 하드웨어 비용.

Format	*Area (μm^2)*	*Power (μW)*	*Energy (fJ)*
FP32	310.08	95.43	257.76
TF32	193.60	58.31	79.58
FP16	162.88	48.32	67.27
BF16	166.72	49.75	55.15
DLF16	164.80	49.32	61.57
FP11	88.96	25.10	16.58

실험에서 적용한 다양한 부동소수점 형식의 하드웨어 비용을 산출하고 비교하기 위해 부동소수점 연산의 핵심 부분인 지수 감산기와 가수 가산기를 Verilog HDL 로 설계하여 *65-nm* CMOS 공정으로 합성하였다. 표 2 는 구현 결과이며, 학습에 실패한 형식은 제외하였고, FP11 은 지수 4-비트, 가수 6-비트를 가지는 형식이다. FP32 는 면적, 파워, 및 에너지를 가장 많이 소모하며. FP16, BF16, 그리고 DLF16 은 FP32 와 비교하면 면적과 파워는 거의 절반을 소비하고 에너지 효율은 4 배 우수하다. FP11 은 면적, 파워, 및 에너지 측면에서 가장 효율적이고, 구체적으로 FP32 와 비교하여 각각 71.3%, 73.7%, 93.6% 우수함을 확인할 수 있다.

Ⅳ. 결론

스파이킹 신경망의 LIF 뉴런에 적용가능한 다양한 정밀도를 가지는 부동소수점 형식을 분석하였다. TF32, FP16, BF16, 그리고 DLF16 형식을 스파이킹 신경망의 LIF 뉴런에 적용하여 정확도의 감소 없이 적용 가능함을 확인했다. 또한 LIF 뉴런에 적용가능한 최적 비트 크기를 분석한 결과 지수 비트가 4 비트 이상이고 가수 비트가 6 비트 이상일 때 정확도의 영향 없이 학습이 가능하다. 따라서 FP32 보다 작은 비트로 학습이 가능하고 이를 통해 하드웨어 리소스 사용량을 감소시킬 수 있다.

ACKNOWLEDGMENT

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded bythe Ministry of Education (NRF-2019R1I1A3A01061266).

참 고 문 헌

[1] Y. Kim et al., “A Reconfigurable Digital Neuromorphic Processor with Memristive Synaptic Crossbar for Cognitive Computing,” *ACM J. Emerg. Tech. Comput. Syst.*, 2015

[2] Y. Wang et al., “Efficient Spiking Neural Network Training and Inference with Reduced Precision Memory and Computing,” *IET Comput. Dig. Tech.*, 2019.

[3] “IEEE Standard for Floating-Point Arithmetic,” IEEE Std. 754-2008, 2008.

[4] N. Burgess et al., “Bfloat16 Processing for Neural Networks,” *IEEE Symp. Comput. Arith. (ARITH)*, 2019.

[5] A. Agrawal et al., “DLFloat: A 16-b Floating Point Formatn Designed for Deep Learning Training and Inference,” *IEEE Symp. Comput. Arith. (ARITH)*, 2019.

[6] R. Bordawekar et al., “EFloat: Entropy-coded Floating Point Format for Deep Learning,” arXiv:2102.02705, 2021

[7] P. U. Diehl et al., “Unsupervised Learning of Digit Recognition using Spike-Timing-Dependent Plasticity,” Front. Comput. Neurosci., 2015.

랜덤 가중치를 가지는 오류 역전파 학습의 성능 분석

이소하, 박혜영
경북대학교 컴퓨터학부
kiyo91012@gmail.com, hypark@knu.ac.kr

Analysis on the Performance of Error Backpropagation Learning with Random Weights

Lee Soha, Park Hyeyoung
School of Computer Science and Engineering, KyungPook National Univ.

요 약

인공신경망의 학습을 위해 제안된 오류 역전파 알고리즘은 최근까지도 딥러닝 모델 핵심 학습 알고리즘으로 사용되고 있다. 하지만 오류 역전파 학습법에서 최상위층의 오류 신호가 가중치를 따라 하위층으로 순차적으로 전달되는 가중치 전송(weight transport)방식에 대해 정보 전달의 효율성과 생물학적 타당성에 대한 논의는 계속되어왔다. 그 대안을 찾는 연구의 하나로, 역방향 오류 신호 전달 과정에서 랜덤 가중치를 이용하는 알고리즘들이 최근에 제안되었다. 이러한 방법들은 간단한 딥러닝 모델을 사용하여 실험을 통해 그 가능성을 입증하였으나, 아직 실험 초기 단계에 머물고 있다. 본 논문은 랜덤 가중치 학습법에 다양한 랜덤 가중치 생성 방식을 결합하였을 때 그 성능 변화에 대하여 실험을 통해 분석한다.

Ⅰ. 서 론

인공신경망들은 최근 다양한 응용 분야에서 성공적으로 사용되고 있으며 이런 인공신경망을 학습시키는 알고리즘인 오류역전파(Backpropagation, BP) 학습법은 개발된 이후 현재까지 대부분의 딥러닝 모델 학습에 활용되고 있다[1]. 오류 역전파 알고리즘의 학습은 하위층의 가중치를 수정하기 위해 상위층들의 가중치 값을 활용하여 오류 신호를 순차적으로 전달받으면서 수행된다. 이런 오류 역전파 알고리즘의 역전파를 통한 오류 기반 학습 특성 때문에 가중치 전송(weight transport) 문제와 같은 정보 전달의 비효율성과 생물학적 타당성 결여와 같은 한계가 존재한다[2]. 특히 출력값을 계산하는 순방향 가중치 값이 오류 신호를 전달하는 역방향 계산에서도 동일하게 사용되는 가중치 대칭(weight symmetry)의 특성은 실제 뇌에서 일어나는 정보 전달 방식이라고 보기 힘들며, 이를 해결하기 위한 시도들이 이뤄지고 있다[2-5].

피드백 할당(feedback alignment, FA) 알고리즘은 역방향 오류 신호 전달 단계에서 순방향 가중치에 의존하지 않고 고정된 랜덤 가중치를 사용하여 네트워크의 학습을 수행하는데 오류 역전파 알고리즘과 비교하였을 때 충분히 비슷한 학습 성과를 낼 수 있음을 확인할 수 있었다[2]. 피드백 할당 알고리즘을 확장하여 학습 병렬 처리를 통한 정보 전달의 효용성 측면에서 그 가능성을 확인한 직접 피드백 할당(direct feedback alignment, DFA) 알고리즘과 스파스 연결을 직접 피드백 할당 알고리즘에 결합하여 학습 중 계산 복잡성을 개선한 알고리즘도 제안되었다[3, 4]. 한편 가중치 대칭성의 문제를 완화하는 또 다른 방법으로 순방향 가중치와 부호만 일치시킨 랜덤 가중치를 사용하는 부호 대칭(sign symmetry, SS) 알고리즘이 제안되어, 역방향 학습에서 가중치의 크기는 성능에 영향을 크게 미치지 않음과 부호 일치성이 학습에 중요한 역할을 함을 증명하였다[5].

그러나 이러한 역방향 랜덤 가중치를 이용하는 연구들은 아직 초기 단계로 비교적 간단한 실험 결과만을 제시하고 있어, 그 특성을 이해하기 위해서는 보다 많은 분석이 필요하다. 본 연구에서는 피드백 할당 알고리즘에서 역방향 랜덤 가중치를 생성할 때 각 층의 가중치 사이 관계를 고려하여 다양한 가중치 생성 방법[6-8]을 적용하여 그 효과를 분석한다. 기존의 피드백 할당 알고리즘에서는 역방향 랜덤 가중치를 생성할 때 카이밍 균등 분포 초기화 방법을 적용하였지만 본 연구에서는 [6-8]과 같은 연구를 활용하여 역방향 가중치의 랜덤 샘플링 방식이 학습 성능에 미치는 영향에 대해서 분석한다.

이 논문의 구성은 다음과 같다. 2에서는 FA 방법의 설명과 학습 성능 향상을 위한 학습 전략 제안, 3절에서는 실험 결과를 제시한 후 4절에서는 결론과 향후 연구 계획에 대하여 설명한다.

Ⅱ. 랜덤 가중치를 가지는 학습법을 이용한 학습 성능 개선 전략

2.1. 랜덤가중치를이용한학습알고리즘

먼저 기본적인 오류 역전파 학습법(backpropagation, BP)에 대하여 기술한 후, 이로부터 변형된 랜덤 가중치 학습법에 대하여 기술한다. 설명의 편의성을 위해 다층 퍼셉트론을 기반으로 설명하나, 합성곱 신경망 등의 다른 모델들도 같은 방식으로 적용된다. l번째 층의 i번째 노드의 출력 $X_i^{(l)}$는 $l-1$번째 층의 k번째 노드의 출력 $X_k^{(l-1)}$와 이들을 연결하는 가중치 $W_{ki}^{(l)}$, i번째 노드로 들어오는 입력들의 가중합 U_i와 활성화 함수 $f(.)$에 의해 다음과 같이 계산된다.

$$X_i^{(l)} = f\left(U_i^{(l)}\right), \quad U_i^{(l)} = \sum_k W_{ki}^{(l)} X_k^{(l-1)} \qquad (1)$$

순방향 계산 과정에서 입력층에서 출력층으로 순차적으로 노드들의 출력값이 계산되면, 신경망의 최종 출력에 대하여 오차값 E

가 얻어진다. 이 오류를 감소시키기 위한 오류 역전파 학습법을 적용하면, 노드($X_i^{(l)}$)에서의 오류 신호는 다음과 같이 상위층의 역전파 신호를 이용하여 계산된다.

$$\frac{\partial E}{\partial X_i^{(l)}} = \sum_j W_{ij}^{(l+1)} f'\left(U_j^{(l+1)}\right) \frac{\partial E}{\partial X_j^{(l+1)}} \quad (2)$$

이 식에서 순방향 계산에 사용되는 가중치 $W_{ij}^{(l+1)}$ 값이 그대로 적용되는 가중치 대칭 문제가 존재함을 알 수 있다. 이 오류 신호를 이용한 가중치 수정 식은 다음과 같이 얻어진다.

$$\frac{\partial E}{\partial W_{ki}^{(l)}} = \frac{\partial E}{\partial X_i^{(l)}} f'\left(U_i^{(l)}\right) X_k^{(l-1)} \quad (3)$$

$$W_{ki}^{(l)} \quad W_{ki}^{(l)} - \eta \frac{\partial E}{\partial W_{ki}^{(l)}} \quad (4)$$

랜덤 가중치를 이용한 학습 알고리즘[2-5]들은 역방향 계산을 수행시에 순방향 가중치 W 대신 랜덤 가중치 V를 적용한다. 즉, 노드($X_i^{(l)}$)에서의 오류 신호는 다음과 같이 계산된다.

$$\frac{\partial E}{\partial X_i^{(l)}} = \sum_j V_{ij}^{(l+1)} f'\left(U_j^{(l+1)}\right) \frac{\partial E}{\partial X_j^{(l+1)}} \quad (5)$$

오류 역전파 이론과 달리, 이러한 방법들은 순방향 가중치(W)와 역방향 가중치(V)가 같을 필요가 없으며, 학습시에 각 노드는 상위층의 가중치 정보를 계속 유지할 필요가 없다. 이러한 방식을 신경망 학습에 처음 적용한 피드백 할당(FA) 알고리즘[2]에서는 랜덤하게 선택한 V를 그대로 학습에 적용하였으며, 이를 통해서도 오류역전파 학습법과 유사한 성능을 얻을 수 있음을 보였다.

기존의 랜덤 가중치를 이용한 학습 연구들에서는 가중치 생성 방법에 대해서는 별도로 고려하지 않고 기본적인 카이밍 균등 분포 방법과 같은 특정 방법으로 고정한 분석을 수행하였다. 그러나 앞에서 기술한 바와 같이, 랜덤 가중치를 생성하는 방식과 초기 가중치와 학습 동적 특성 간의 관계를 통해 랜덤 가중치를 이용하는 알고리즘의 성능을 보완할 수 있을 것이라 기대되며 이러한 효과를 실험적으로 분석할 필요가 있다. 본 연구에서는 이를 위해 피드백 할당 알고리즘에 가중치 생성 방법을 결합한 학습 전략들에 대하여 정리하고 분석하고자 한다.

2.2. 역박향 랜덤 가중치의 설정

여기서는 다양한 가중치 초기화 방법을 역방향 가중치 초기화에 적용하여 그 성능을 분석한다. 인공신경망에서 순방향 가중치(W)를 생성할 때 같은 크기의 랜덤 가중치(V)를 생성하고 이러한 랜덤 가중치 생성에 여러 가중치 초기화 방법을 적용한다. 첫 번째, 기존 방법인 카이밍 균등 분포 초기화 방법[6]을 사용하면 활성화 함수에 따라 범위를 계산하고 이 범위 내에서 랜덤 가중치 텐서들이 균등하게 생성된다. 두 번째로, 직교 가중치 초기화 방법[7]을 통해 직교 가중치 텐서들로 랜덤 가중치 행렬을 생성하면 정방 행렬이 아닌 경우 차원이 큰 쪽으로 벡터 직교가 성립하게끔 생성된다. 세 번째, 희소 초기화 방법[8]은 가중치 텐서의 크기를 받아와서 먼저 정규 분포로 생성된 난수를 채운 후에 열의 수 만큼 임의의 위치에 원소들을 0으로 변환시켜 희소 초기화 텐서를 생성한다.

본 논문에서는 단순히 하나의 가중치 초기화 방법을 이용하여 성능 향상을 노리는 것 외에 입력층과 은닉층에서 직교 초기화 방법을 사용하고 은닉층과 출력층 사이에는 카이밍 균등 분포나 희소 초기화 방법을 사용한 가중치 초기화 방법을 제안한다. 출력층에서 은닉층으로의 가중치 사이 차이가 클수록 성능에 영향이 있다는 연구[9]에 기반하여, 벡터 공간에서는 직교할 때 차이가 가장 크고 또 역전파의 방향성에 따른 직교를 성립시키기 위해 출력층을 제외한 은닉층의 가중치 설정에 직교 초기화 방법을 적용하였다. 역시 가중치 직교 방향을 유지하기 위해 마지막 은닉층과 출력층의 사이 가중치 설정은 다른 초기화 방법으로 카이밍 균등 분포 초기화와 희소 초기화 방법을 적용하였다.

다음 절에서는 랜덤 가중치를 이용한 학습 방법에 가중치 초기화 방법을 다르게 주었을 경우 어떤 효과가 있는지 실험을 통해 분석하고 이를 바탕으로 효과적인 랜덤 가중치 학습 전략을 제안한다.

Ⅲ. 실험 분석

3.1. 실험 조건

랜덤 가중치 설정 방법에 따른 학습 효과를 확인하기 위하여 MNIST 데이터 세트를 이용하여 다층 퍼셉트론과 합성곱 신경망 구조에 대한 학습 실험을 수행하였다. 학습 에포크는 100번을 수행하였으며 학습률은 0.0001로 설정하였다. 실험에서 사용한 모델은 다층 퍼셉트론 네트워크에서는 완전 연결 층이 1개부터 4개까지의 경우가 있으며 합성곱 신경망 모델은 합성곱 블록이 1개부터 2개인 경우로 구성하였다. 모든 실험의 오차 함수로는 크로스 엔트로피 함수를 사용하였으며 활성화 함수는 하이퍼 탄젠트 함수를 사용하였고 최적화 알고리즘으로는 Adam을 사용하였다. 실험 결과는 테스트 오차를 백분율로 나타냈으며 각 학습 시도에서 가장 낮은 테스트 오차율을 집계하였고, 초기 가중치를 달리하는 10번의 실험에 대한 평균과 표준편차를 얻었다.

랜덤 가중치 설정 방법에 카이밍 균등 분포 초기화 방법을 사용한 FA 방법을 UFA라 명명하였다. 다음으로 직교 초기화 방법을 사용한 FA 방법은 OFA 그리고 이 논문에서 제안한 학습 전략으로 직교 초기화 방법을 은닉층 역전파 가중치에 사용하고 마지막 출력층에 대해 균등 분포와 희소 초기화 방법을 사용한 것을 각각 OUFA와 OSFA로 명명하였다.

3.2. 실험 결과

MNIST 데이터 세트에서 각 가중치 초기화 방법을 랜덤 가중치 설정에 적용하여 학습한 성능을 [표 1]에 비교 정리하였다. 각 모델 구조에서 가장 성능이 좋은 결과는 굵은 표시와 밑줄로 표시하였고, 두 번째로 성능이 좋은 가중치 설정 방법은 굵은 글씨로만 표시하였다.

[표 1]에 정리한 것을 살펴보면 신경망 구조가 단순할 때에 각 가중치 설정 방법들의 성능 차이는 크게 보이지 않는 것으로 나타났다. 하지만 전체적으로 봤을 때 구조가 복잡해질수록 직교 초기화와 희소 초기화를 결합한 OSFA와 직교 초기화와 균등 분포 초기화를 결합한 OUFA의 오차율이 낮은 경향을 보이는 것을 확인할 수 있다. 또한 직교 초기화만 사용한 방식인 OFA의 경우에도 오차율이 낮은 경향을 보이는 것을 확인할 수 있다. 이를 통해 직교 초기화 방식과 희소 초기화 방식을 이용하였을 때 노드 간 가중치의 차이를 크게 만들어 랜덤 가중치를 사용하는 학습 알고리즘의 성능 향상을 노릴 수 있다는 가능성을 확인하였다.

[표 1] 랜덤 가중치 설정 방법에 따른 MNIST 성능 비교

가중치 설정 / 모델	평균 오차율(%)±표준편차			
	UFA	OFA	OUFA	OSFA
FC1-1000	1.75±0.06	1.77±0.04	1.75±0.07	1.75±0.05
FC2-1000	1.67±0.05	1.64±0.05	1.67±0.03	1.62±0.03
FC3-800	1.60±0.06	1.58±0.04	1.58±0.04	1.57±0.05
FC4-800	1.69±0.05	1.65±0.04	1.66±0.04	1.68±0.06
CONV1-5x5	1.24±0.02	1.25±0.04	1.21±0.06	1.23±0.04
CONV2-5x5	0.82±0.06	0.80±0.06	0.76±0.04	0.78±0.06

본 실험에서는 랜덤 가중치 생성 방식을 다양하게 적용하였을 때 그 효과에 대하여 조사하였다. 학습 다이나믹스 측면에서 기존의 균등 분포를 통해 가중치를 생성한 것보다 직교 초기화 방법을 통해 초기 가중치 간 거리를 멀게 하였을 때 성능이 조금 더 향상되는 것을 확인할 수 있었다. 이러한 결과를 바탕으로 랜덤 가중치를 이용한 학습에 적절한 랜덤 가중치 설정 방법을 적용함으로써 가중치 전송 문제를 가지는 기존 BP와 비슷한 성능을 얻을 수 있음을 확인하였다.

Ⅳ. 결론

본 논문에서는 오류 역전파 학습법이 가지는 가중치 전송 문제를 해결하기 위해 개발된 랜덤 가중치 학습 알고리즘의 성능에 랜덤 가중치 설정 방법이 미치는 영향을 조사하였다. 기울기 역학에서 가중치의 관계를 통해 성능을 개선할 수 있다는 점에 착안하여 랜덤 가중치를 설정할 때 다양한 가중치 초기화 방법을 적용하여 그 차이를 분석하였다. 기존 가중치 생성 방법인 균등 분포 초기화 방법을 사용하였을 때 간단한 구조에서는 성능이 가장 좋지만 구조가 복잡해질수록 직교 초기화 방법을 사용하였을 경우가 성능이 더 좋은 것으로 나타났다. 이러한 직교 초기화 방법을 균등 분포 초기화 방법과 희소 초기화 방법에 결합한 방법은 기존의 균등 분포 초기화 방법과 비교하였을 때 더 낮은 학습 오차를 기록하는 등의 성능 개선을 실험을 통해 그 효과를 확인하였다. 향후에는 랜덤 가중치 초기화 설정 시 가중치 벡터 사이의 관계에 대한 심층적 분석과 구현 방법을 모색할 것이다.

ACKNOWLEDGMENT

* 본 연구는 2020년도 산업통상자원부의 재원으로 한국에너지기술평가원(KETEP)의 에너지인력양성사업으로 지원받아 수행한 인력양성 성과입니다. (No. 20204010600060)

* 이 논문은 2020년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. NRF-2020R1A2C1010020)

참 고 문 헌

[1]Rumelhart D.E., Hinton G.E. and Williams R.J., "Learning representations by back-propagating errors", Nature, 323(9):533-536, 1986.

[2]Lillicrap T.P., Cownden, D., Tweed, D.B. & Akerman, C. J., "Random synaptic feedback weights support error back-propagation for deep learning", Nature Communications, 1-10., 2016.

[3]Nøkland A., "Direct Feedback Alignment Provides Learning in Deep Neural Networks", 30th Conference on Neural Information Processing Systems, 2016

[4]Crafton B., Parihar A., Gebhardt E. & Raychowdhury A., "Direct Feedback Alignment With Sparse Connections for Local Learning", Front Neurosci. 2019; 13: 525., 2019.

[5]Liao Q., Leibo, Z.L. & Poggio, T., How Important Is Weight Symmetry in Backpropagation?", 30th AAAI Conference on Artificial Intelligence, 2016.

[6]He K., Zhang X., Ren S. and Sun J., "Delving Deep into rectifiers: Surpassing human-level performance on imagenet classification", In Proceedings of the IEEE international conference on computer vision, 2015.

[7]Saxe A.M., McClelland J.L. and Ganguli S.,"Exact solutions to the nonlinear dynamics of learning in deep linear neural networks", In Proceedings of the Second International Conference on Learning Representations, 2014.

[8]Martens J.,"Deep learning via Hessian-free optimization", Proceedings of the 27th International Conference on Machine Learning, 2010.

[9]Park H.-Y., Inoue M. and Okada M., "Online learning dynamics of multilayer perceptrons with unidentifiable parameters", Journal of Physics A: Mathematical and General, Vol. 36, pages 11753-11764, 2003.

건조 중인 LNG 화물창 시스템의 안정적 환경 유지를 위한 기계학습 기반 화물창 내부 습도 예측

김재성[1], 옥진성[1,2], 서영균[1,*]

[1]경북대학교, [2]대우조선해양

horsequake@knu.ac.kr, jsok@dsme.co.kr, *yksuh@knu.ac.kr

Machine Learning Based Indoor Humidity Forecasting for Steady Environment Maintenance of LNG Cargo Containment System under Construction

Jaeseong Kim[1], Jin-Sung Ok[1,2], and Young-Kyoon Suh[1,*]

[1]Kyungpook National University, [2]Daewoo Shipbuilding & Marine Engineering

요 약

LNG 화물창 시스템과 같은 특수한 환경은 기존의 데이터 기반 환경 정보 예측 모델을 있는 그대로 적용하기 어렵다. 본 논문에서는 건조 중인 LNG 화물창 시스템의 안정적인 내부 환경 유지를 위한 기계 학습 기반의 환경 정보 예측 방법론을 제안한다. 또한 제안된 방법론에 따라 실제 공기조화 시스템에서 수집한 데이터를 이용하여 학습된 다양한 기계 학습 모델의 내부 습도 예측 성능을 평가하고 최적의 예측 모델을 탐색한다. 실험 결과, 본 논문에서 제안한 방법론이 LNG 화물창 환경 정보 예측에 효과적이었으며 XGBoost Regressor의 성능이 가장 우수함을 알 수 있었다.

Ⅰ. 서 론

최근 액화천연가스 운반선(LNGC) 발주량이 폭발적으로 증가하고 있다[1]. 이는 석유 및 석탄 사용량이 감소하고 탄소 배출량을 줄이기 위해 천연 가스로의 전환이 증가했기 때문이다. LNG 운송선의 핵심 구성요소인 화물창(CCS)은 액화된 천연가스를 저장하기 때문에 화물창의 내부 환경을 안정적으로 유지하는 것이 운송능력과 직결된다. 따라서 건조중인 화물창 내부의 단열재가 녹과 이슬 등으로 인해 오염되지 않도록 화물창의 내부 환경 온도와 습도를 안정적으로 유지하는 것이 매우 중요하다. 이에 따라 안정적인 화물창 내부 환경 유지를 목적으로 대형 공기조화(HVAC) 시스템을 24시간 가동하고 있다. 그러나 이로 인해 화물창의 건조 공정이 진행되는 동안 엄청난 양의 전력이 낭비되고 있다. 그래서 전력 소모를 줄이기 위한 HVAC 시스템의 자동 제어에 대한 요구가 증가하고 있다. 이러한 수요를 충족시키기 위해서는 예측한 환경 데이터를 바탕으로 HVAC 시스템을 자동으로 제어할 수 있도록 화물창의 상태에 영향을 미치는 습도 및 온도와 같은 환경 정보를 정확하게 예측하는 것이 중요하다.

HVAC 시스템의 자동 제어에 적용할 수 있는 데이터 기반 환경정보 예측에 대한 기존 연구[2, 3]가 진행되었으나, 대부분의 접근 방식은 온도, 습도, 강수량, 일조량 등과 같이 일반적인 환경 변수만 고려했다. 따라서 기존 접근 방식은 화물창과 같이 특수한 환경에 적용하기 어렵다. 또한 정확한 예측을 위해서 모델은 더 복잡해지고 계산 자원을 많이 소모하게 되었다.

이와 관련하여 본 논문에서는 화물창 내부 환경을 예측하기 위해 적합한 특징(feature)을 탐색하고 상대적으로 구조가 간단한 기계학습 모델로 중요한 환경 정보 중 하나인 습도를 예측할 수 있음을 실험을 통해 보인다. 본 논문에서 적용한 모델은 특징 선택, 최적의 윈도우(window) 크기 결정, 최적의 시계열 기계학습 모델 탐색 단계를 거치도록 구성된다. 예측의 정확성을 향상시키기 위해 우리는 워크-포워드 검증(walk-forward validation)[4]을 활용하여 학습 시 예측 값과 실제 값을 재활용한다. 본 연구의 타당성을 보이기 위해 실제 건조 중인 화물창 및 HVAC 시스템에 부착된 IoT 센서로부터 수집된 데이터를 전처리 후, 정제된 데이터로 5종의 기계학습 모델을 훈련하고 모델의 성능을 평가한다.

본 논문에서는 실험을 통해 XGBoost regressor[5]가 0.5%의 낮은 오차율을 산출함으로써 다른 모델들 보다 성능이 우수하고 비교적 정확하게 환경 정보를 예측할 수 있음을 보인다.

본 논문의 기여점은 다음과 같이 요약할 수 있다.

- 건조중인 LNG 운송선 화물창 내부의 실내 습도 예측을 위한 데이터 기반 방법론을 제안한다.
- 제안한 방법론의 타당성을 조선소의 실제 HVAC 데이터로 평가한다.
- 5종의 기계학습 모델의 예측 성능에 대한 평가 결과를 제시한다.

Ⅱ. 제안하는 방법론

본 장에서는 화물창 내부의 습도 예측을 위한 방법론을 제안하고 설명한다. 제안하는 방법론은 3단계로 진행된다. 첫 번째 단계는 학습에 사용할 입력 특징을 선택하는 단계이다. 실험에 사용한 데이터셋은 화물창 내·외부에 설치한 센서와 HVAC 시스템에서 분 단위로 2020년 6월부터 9월까지 수집된 데이터이다. 일부 결측값이 포함되어 총 94,435 행으로 이루어진 데이터 셋에 IQR을 적용하여 센서나 네트워크상의 오류로 인해 발생한 이상치를 제거하였고 칼만 필터[6]를 적용하여 노이즈를 제거하였다. 수치형 데이터와 범주형 데이터가 섞여있는 기존 데이터셋에서 입력 특징을 선택하기 위해 피어슨(Pearson) 상관계수와 양분점(Point-biserial) 상관계수를 구하여 특징 간 상관분석을 수행하였다. 총 80개의 특징 중 연관성이 큰 14개의 특징을 선택하여 학습 데이터로 사용하였다.

다음 단계에서 학습을 위한 최적의 시간 윈도우 크기를 결정한다. 각 모델은 데이터셋을 일정 크기의 윈도우를 학습시키고 예측한 데이터를 다시 학습데이터로 사용하는 워크-포워드 검증 방식을 적용하여 데이터를 학

표 1 윈도우 크기에 따른 기계학습 모델의 습도 예측 오차

Humidity(%) Forecasting Error												
Windows Size	1 Day			7 Days			14 Days			1 Month		
Model \ Error	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
XGB Regressor	**0.260**	**0.309**	**0.6%**	**0.251**	**0.299**	**0.5%**	**0.214**	**0.259**	**0.5%**	0.261	0.311	**0.5%**
XGB Random Forest	0.377	0.432	0.8%	0.391	0.440	0.8%	0.411	0.454	0.9%	0.534	0.579	1.1%
Linear Regressor	0.343	0.396	0.8%	0.271	0.317	0.6%	0.264	0.307	0.6%	**0.260**	**0.302**	**0.5%**
KNN	0.554	0.648	1.2%	0.569	0.682	1.2%	0.539	0.648	1.2%	0.530	0.637	1.1%
SVM	0.467	0.519	1.1%	0.411	0.452	0.9%	0.395	0.343	0.9%	0.387	0.426	0.8%

습하였다. 본 연구에서는 윈도우 크기를 1일, 1주, 2주, 1달로 나누어 각각에 대해 성능을 비교하였다.

마지막 단계에서 화물창 내부의 습도 값을 예측하고 시간의 변화에 따른 가장 최적의 모델을 탐색한다.

Ⅲ. 실험 결과

본 논문에서는 XGBoost Regressor, XGBoost Random Forest[4], Linear Regressor, KNN, SVM 총 5가지 기계학습 모델을 적용하여 화물창 내부의 습도를 예측하고 각 모델의 성능을 비교하였다. Mean Absolute Error(MAE), Root Mean Squared Error(RMSE), Mean Absolute Percentage Error(MAPE)를 평가지표로 하여 기계학습 모델의 화물창 내부 습도 예측 성능을 평가하였다. 그 결과는 표 1과 같다. 윈도우 크기에 관계없이 XGB Regressor가 전반적으로 좋은 성능을 보였다. 그에 반해 KNN 모델은 성능이 가장 낮은 것을 확인할 수 있다. 이는 입력 특징들의 국지적인 변화에 민감하여 예측 값의 추세가 비정상적으로 나타난 것으로 보인다. 윈도우 크기가 한 달일 경우에는 Linear Regressor 모델의 성능이 가장 좋았으나 XGB Regressor의 성능은 더 낮아지는 경향을 보였다. 이는 장기간의 데이터를 학습할 경우 화물창 건조 공정과 외부 환경이 점차 변화하는 것을 반영하는 것이 어렵다는 것을 보여준다. 최적의 윈도우 크기에 대한 추가적인 검증이 필요하지만 실험을 통해 습도 예측에 대해서는 2주의 윈도우 크기가 가장 적합한 것을 확인하였다. 그림 1은 윈도우 크기가 2주일 때 특정 날짜의 습도 예측 값을 그래프로 나타낸 것이다. XGB Regressor는 실제 측정된 값에 가장 가깝게 근사하였고 KNN은 불안정한 곡선을 그리는 것을 볼 수 있다. 이러한 결과를 바탕으로 계산 비용이 적은 기계학습 모델로 충분히 환경정보를 예측할 수 있음을 확인하였다.

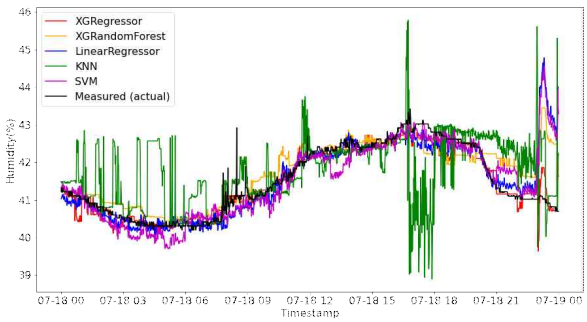


그림 1 일일 습도예측 그래프

Ⅳ. 결론 및 향후연구

본 논문에서는 실제 건조중인 LNG 수송선의 화물창에서 수집된 데이터를 기반으로 기계학습 모델을 적용하여 화물창 내부 습도를 예측하였다. 실험을 통해 윈도우 크기가 2주일 때 XGBoost Regressor 모델의 성능이 가장 우수함을 보였고, 계산비용이 적은 기계학습 모델로 충분히 실제 값에 가깝게 예측할 수 있음을 확인하였다. 향후에는 보다 더 정확하게 환경정보를 예측하는 모델과 최소한의 전력을 사용하여 화물창 내부 환경을 자동으로 제어하는 강화학습 모델을 개발하는 연구를 진행할 예정이다.

ACKNOWLEDGMENT

본 논문은 2021년 경북대학교-대우조선해양㈜ 산학협력 과제(과제명: 에너지 절감(전기 사용량)을 위한 건조 중인 화물창의 최적 환경 제어 모델 연구)의 지원으로 수행된 연구임.

참 고 문 헌

[1] Andi Case and Jeff Woyda, “Clarkson Full Year Results 2020,” 2021, (https://www.clarksons.com/media/1242925/20210308_-_clarkson_full_year_results_2020.pdf)

[2] Attoue N, Shahrour I and Younes R, “Smart Building: Use of the Artificial Neural Network Approach for Indoor Temperature Forecasting,” Energies, vol. 11, no. 2: 395, Feb. 2018.

[3] Kaicheng Zhang, Akhil Guliani, Seda Ogrenci-Memik, Gokhan Memik, Kazu-tomo Yoshii, Rajesh Sankaran, and Pete Beckman, “Machine Learning-Based Temperature Prediction for Runtime Thermal Management Across System Components,” IEEE Transactions on Parallel and Distributed Systems, vol. 29, no. 2, pp. 405-419, Feb. 2018.

[4] Stein and Roger M, “Benchmarking default prediction models: Pitfalls and remedies in model validation,” Moody’s KMV, New York, vol. 20305, 2002.

[5] Guestrin Chen, Tianqi and Carlos, “XGBoost: A Scalable Tree Boosting System,” In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016.

[6] Greg Welch and Gary Bishop, “An introduction to the Kalman filter,” Proc. Siggraph Course, vol. 8, Jan. 2006.

연합학습 기법을 활용한 IoT 와 Edge Computing 기반의 현장중심의 재난 탐지 및 대응에 대한 연구

우아밍, 안재광*, 권영우
경북대학교, *기상청

wuaming@knu.ac.kr, propjk@korea.kr, ywkwon@knu.ac.kr,

Wu Aming, Jae-Kwang Ahn*, Young-Woo Kwon

Kyungpook National University, *Korea Meteorological Administration

요 약

최근 소형 센서 기술과 인공지능의 발전으로 인하여 이를 활용한 재난 탐지 및 대응에 대한 연구가 증가하고 있다. 지진조기경보 분야에서는 기존의 네트워크 기반의 지진조기경보 체계에서 발생하는 경보 공백역을 보완하기 위하여 소형 가속도 센서를 활용하여 지진을 탐지하는 현장경보에 대한 연구가 활발히 이루어지고 있다. 현장경보는 전국 단위의 네트워크 경보와는 달리 데이터가 수집되는 현장에서 즉시 지진을 판단하고 경보를 발령하는 기술로 신속한 지진 감지가 가능하지만 오경보의 가능성이 존재한다. 이를 보완하기 위하여 소형 가속도 센서를 보조 수단으로 활용하는 연구가 증가하고 있다. 본 논문에서는 소형가속도 센서로 구성된 엣지 컴퓨팅 기반의 현장경보 체계와 연합학습 기반의 지진 감지 알고리즘을 소개한다.

Ⅰ. 서 론

최근 소형 센서 기술과 인공지능의 발전으로 인하여 이를 활용한 재난 탐지 및 대응에 대한 연구가 증가하고 있다. 지진조기경보 분야에서도 이처럼 소형 센서와 인공지능 기술의 도입이 활발히 이루어지고 있는데, 기존의 지구 물리학 기반의 지진 감지가 아닌 소형 센서(IoT 센서)에서 수집되는 데이터 기반 지진 감지 기술이 최근 연구의 주를 이루고 있다. 지진조기경보는 지진이 발생하였을 때 P 파를 감지하여 S 파가 도달하기 전에 경보를 전파함으로서 피해를 줄이는 기술로 기상청은 전국에 약 300 개소의 지진관측소를 하나의 네트워크로 연결하여 지진조기경보 체계를 운영하고 있다[1].

현재의 네트워크 기반의 지진조기경보 체계는 데이터 전송 시간과 관측소 간의 거리로 인하여 경보 공백역의 발생이 불가피하다. 경보 공백역은 지진발생 시 지진 경보를 수신하기 전에 지진이 먼저 도달하는 지역으로 조밀한 관측소 구축을 통하여 어느정도 해결할 수 있지만 중앙집중식 데이터 처리로 인하여 경보 시간의 단축에는 한계가 있다. 이러한 한계를 극복하기 위하여 최근 연구에서는 현장에서 수집되는 데이터를 즉시 처리하여 지진을 감지하고 경보를 발생시키는 현장경보 기술에 대한 연구가 활발히 이루어지고 있다. 현장경보 기술은 단일 관측소에서 수집되는 데이터로 지진을 판단하므로 신속한 경보가 가능하지만 오경보 발생 가능성이 높다는 문제가 있다. 그러므로 본 논문에서는 3 개 이상의 IoT 센서를 사용하여 지진을 감지하고 경보를 발생시킬 수 있는 엣지 컴퓨팅 기반의 현장경보 체계에 대해서 소개한다. 지진 감지를 위하여 연합학습 기반의 지진 감지 알고리즘을 개발하여 현장에서 수집되는 진동의고유한 특징을 반영하여 모델을 학습하고 이를 통해 지진 감지 성능을 더욱 향상시키고자 하였다.

본 논문은 다음과 같이 구성되어 있다. 2 장에서 지진 감지 모델을 소개하고 3 장에서 실험을 통해 지진 감지 모델의 성능을 평가한다 4 장에서 본 논문을 마무리한다.

Ⅱ. 본론

선행 연구를 통해 약 8,000 개의 IoT 센서를 활용한 지진 관측망을 구축하였으며 2021 년 1 월 이후에 국내에서 발생한 규모 2.5 이상의 지진을 모두 감지하고 있다[2]. 이를 통해 IoT 센서의 효용성을 검증하였으며 본 연구에서는 이러한 IoT 센서를 활용한 지진 현장경보 체계를 구축한다.

그림 1 은 지진 현장경보체계의 구성을 보여주고 있다. 최소 3 개 이상의 가속도 센서가 포함된 IoT 장치를 설치하고 연합학습을 위한 엣지서버를 배치한다. IoT 센서는 초당 100 개의 가속도 데이터를 수집하고 이를 엣지서버로 전송하면 엣지서버에서는 지진을 판단하는 알고리즘을 실행하여 지진 판정을 한다. 최종 지진으로 판정되면 경보 방송을 통해 행동 요령을 전파하여 안전하게 대피할 수 있도록 안내한다.

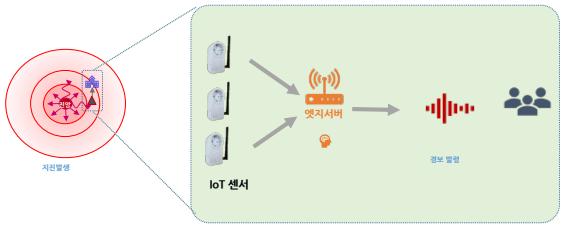


Figure 1. 연합학습 기반의 지진 감지 모델.

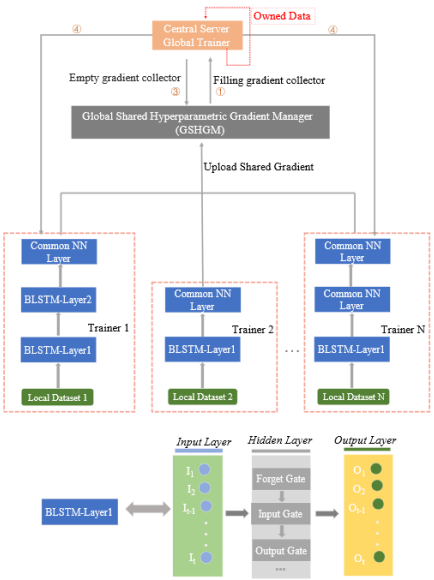


Figure 2. 연합학습 기반의 지진 감지 모델.

그림 2 는 지진 판단을 위한 연합학습 기반의 지진 감지 모델의 구조를 보여주고 있다. 지진 감지 성능은 센서가 설치된 환경에 매우 의존적임을 지난 연구에서 보인바 있다. 네트워크 기반의 지진조기경보 체계에서는 수 많은 센서를 동시에 활용할 수 있으므로 전체 센서를 대상으로 한 글로벌 지진 감지 모델을 사용하는 것이 효율적이지만 현장경보 체계에서는 센서가 설치된 환경정보를 반영하여 모델을 최적화해야한다.

Algorithm: SHEFDML

Assume current step per trainer: T_i,episode per trainer: $episode_i$, maximum step per trainer T_{mi}

Initialization: $\{T_i \leftarrow 0\}_{i=1}^{M}$=0,$\{episode_i\}_{i=1}^{M}$, $\{T_{mi}\}_{i=1}^{M}$, $1 \le i \le M$/*M the count of local networks*/

// Assume local and global shared parameters ρ

// Assume local gradient weight $\{W_i \leftarrow 1\}_{i=1}^{M}$

// Assume the datasets per trainer $\{Trainer_i \leftarrow D_i\}_{i=1}^{M}$, $D_1 \cap D_2 \cap \cdots \cap D_M = \emptyset$

// Assume global gradient collector Q with size q

1. Initialize all the network parameters α, ρ, ϑ with Normal distribution
2. Create one global network with parameters ρ,ρ be shared
3. Create M local network with parametersρ,ρ be shared
4. Create one global gradient collector with size q
5. **repeat**
6. Asynchronous run all the local trainers with none parameters update
7. Collect gradient ∇_{θ_i} for current $trainer_i$, $\nabla_{\theta_i} = \sum_j^h \sum_k^c p_{jk} ln_{p_{jk}} \times (-1) \times W_i$
8. $Q \leftarrow \nabla_{\theta_i}$
9. **If** Q is full
10. Update shared parameters ρ with weighted local gradients $\{\nabla_{\theta_i}\}_{i=1}^{q}$
11. Clear global gradient collector Q
12. random choice the m-th trainer then
13. $trainer_{m,\rho} \leftarrow global_{\rho}$ /*broadcast global parameters to local */
14. **If** one local trainer i converged or $T_i \ge T_{mi}$
15. Stop trainer$_i$ training
16. **until** T % Episode == 0
17. **until** $T > T_{max}$

Figure 3.지진 감지 알고리즘.

III. 실험 및 평가

제안하는 연합학습 기반의 지진감지모델은 각 센서에서 수집된 데이터를 사용하여 로컬 모델을 각자 학습 시키고 그 결과를 중앙에 위치한 Gradient Manager 에 전송한다. 그리고 전송된 각각의 Gradient 에 가중치를 적용하여 글로벌 모델을 업데이트하고 이를 반복함으로써 새롭게 발생되는 잡음 데이터에 강건한 지진감지 모델로 발전할 수 있게 된다. 그림 3 은 이와 같은 과정을 의사코드로 표현하고 있다.

본 논문에서 제안하는 지진 감지 모델을 테스트하기 위하여 실제 지진 파형을 대상으로 실험을 하였다.일본의 NIED 데이터를 활용하여 지진 데이터를 수집하였고 다양한 비지진 데이터를 IoT 센서를 사용하여 수집하였다. 다음 그림 3 은 모델에 의해 감지된 지진 파형을 보여주고 있다. 지진 파형은 가속도의 벡터합으로 표현되고 있으며 0.9 이상의 임계치를 넘을 경우 지진으로 판단한다.

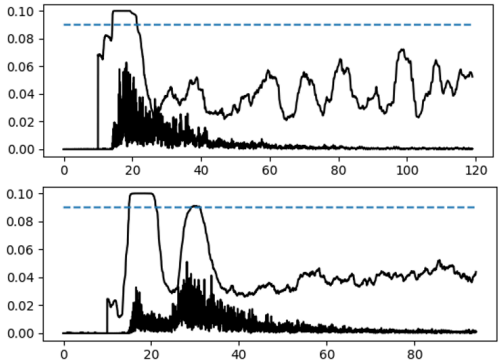


Figure 4. 연합학습 기반의 지진 감지 모델.

IV. 결론

엣지 컴퓨팅과 연합학습에 기반한 지진 현장경보 체계는 네트워크 인프라가 제공되지 않는 산간지역 및 국가 단위의 네트워크 지진 관측망을 구축하기 어려운 저개발 국가에 적용이 가능하며, 이를 통해 지진으로 인한 피해를 크게 경감시킬 수 있을 것으로 기대한다.

ACKNOWLEDGMENT

본 연구는 지진·지진해일·화산 감시 및 예측기술 개발(Development of earthquake, tsunami, volcano monitoring and prediction technology) (NTIS:1365003423) 과제로 수행되었습니다.

참 고 문 헌

[1] 기상청, "국가 지진관측망도," 2021, (https://www.kma.go.kr/aboutkma/biz/earthquake_volcano_02.jsp).

[2] Xin Huang, Jangsoo Lee, Young-Woo Kwon, and Chul-Ho Lee, "CrowdQuake: A Networked System of Low-Cost Sensors for Earthquake Detection via Deep Learning," in *Proc. ACM SIGKDD*, pp. 3261-3271, 2020.

경찰의 사건현장 대응 지원을 위한 긴급출동 의사결정 지원 시스템

박현호, 변성원, 권은정, 장동만, 정의석
한국전자통신연구원

hyunhopark@etri.re.kr, swbyon@etri.re.kr, ejkwon@etri.re.kr, dmjang@etri.re.kr, esjung@etri.re.kr

Emergency Dispatch Decision Support System for Supporting the Police's Response to Incident Scenes

Hyunho Park, Sungwon Byon, Eunjung Kwon, Dong Man Jang, Eui-Suk Jung
Electronics and Telecommunications Research Institute

요 약

사건현장으로 출동한 경찰은 그들의 경험과 지식에 따라 사건현장에 대한 대응 능력은 차이가 나며, 대응능력 차이는 사건현장 대응 시간 차이와 피해 규모의 차이로 이어진다. 본 논문은 사건현장에서의 경찰의 대응을 지원하기 위한 긴급출동 의사결정 지원 시스템을 제안한다. 본 논문의 시스템은 사건현장에 출동한 경찰에게 신고내용에 대한 정보를 제공해줄 뿐만이 아니라, 사건과 관련 있는 법률, 지침, 유사한 이전 사례를 참고하여 사건현장 대응 정보를 생산하여 제공한다. 긴급출동 의사결정지원 시스템을 통하여 출동 경찰의 신속하고 정확한 사건현장 대응을 지원하여, 사건현장 피해를 줄일 수 있을 것이다.

Ⅰ. 서 론

경찰이 사건현장 관련한 법률, 지침, 이전 유사 사례에 대한 지식이나 경험을 가지기는 쉽지 않다. 지식이나 경험이 갖춰지지 않은 경찰은 사건현장에 적절하게 대응하기 어려워, 사건현장에 대응하는 시간이 길어지고 사건현장의 피해를 악화시킬 수 있다. 뿐만 아니라, 부적절한 사건현장 대응은 경찰관의 생명을 위험에 처하게 할 수 있다. 경찰청 통계[1]에 따르면, 매년 13.8 명의 경찰이 순직하고, 1,700 명의 경찰이 부상을 당하는 데, 부적절한 사건현장의 대응이 이러한 경찰의 피해에 영향을 주었을 것이다.

경찰의 업무를 지원할 수 있는 연구나 개발은 꾸준히 진행되어 오고 있다[2-4]. 로스앤젤레스 경찰국은 발생할 수 있는 범죄의 시간, 장소를 예측하고 그에 따른 순찰 경로를 제공할 수 있는 PredPol(Predictive Policing) 시스템을 도입하였다[3]. Microsoft 는 뉴욕 경찰국을 위해 요주의 인물을 지속적으로 감시할 수 있는 시스템인 DAS(Domain Awareness System)를 개발하였다[4]. 사건 현장에 출동한 경찰을 즉각적으로 지원할 수 있는 시스템에 대한 연구도 필요하다.

본 논문은 경찰의 사건현장 대응 지원을 위한 긴급출동 의사결정 지원 시스템을 제안한다. 제안하는 시스템은 경찰이 사건현장에 도착하기 전에는 사건현장에 대한 사전 정보를 제공하고, 신고 정보와 사건현장 경찰의 음성 정보를 통한 보고를 통한 사건현장을 분석하고, 사건현장에 적합한 정보를 제공한다. 긴급출동 의사결정 지원 시스템을 통해 피해를 줄일 뿐만 아니라 사건현장으로부터 경찰관의 안전을 지킬 수 있을 것이다.

Ⅱ. 긴급출동 의사결정 지원 시스템

본 섹션에서는 긴급출동 의사결정 지원 시스템의 아키텍처(architecture)와 기능적 동작에 대해 설명한다. 그림 1 은 긴급출동 의사결정 지원 시스템의 아키텍처이다. 본 시스템은 112 신고처리 시스템과 현장에 출동하는 출동 경찰의 경찰 휴대 장치 및 차량 거치 기기와 상호작용을 통해 동작한다. 112 신고처리 시스템은 112 신고 전화를 받아 신고 내용을 정리한 긴급전화 정보를 생산하여 긴급출동 의사결정 지원시스템에 전달한다. 경찰 휴대 장치는 경찰이 휴대한 스마트폰과 같은 기기이며, 차량 거치 기기는 경찰차에 탑재된 타블렛 기기이며, 경찰은 경찰 휴대 장치와 차량 거치 기기를 통하여 긴급출동 의사결정 지원 시스템에 사건 현장상황 보고 및 현장상황 대응에 대한 문의를 할 수 있다. 그리고 경찰 휴대 장치와 차량 거치 기기는 출동 정보를 긴급출동 의사결정 지원 시스템에 전달할 수 있다.

긴급출동 의사결정 지원 시스템은 출동업무지원 모듈, 음성인식 모듈, 의사결정 모듈, 빅데이터지식베이스 모듈로 구성된다. 출동업무지원모듈은 112 신고처리 시스템, 경찰 휴대 장치, 차량 거치 기기, 음성인식 모듈, 의사결정 모듈과 데이터를 교환한다. 출동업무지원 모듈은 112 신고처리 시스템으로부터 긴급전화 정보를 수집하고, 경찰 휴대 장치와 차량 거치 기기로부터는 출동 정보와 경찰 음성 데이터를 수집한다. 출동업무지원 모듈은 음성인식 모듈로 경찰 음성 데이터를 전달하고, 긴급전화 정보와 출동 정보를 의사결정 모듈에 전달한다. 그리고, 출동업무지원 모듈은 의사결정 모듈로부터 사건현장의 경찰이 대응하는 데에 도움이 되는 정보인 대응 정보를 받아서, 경찰 휴대 장치와 차량 거치 기기에 전달한다.

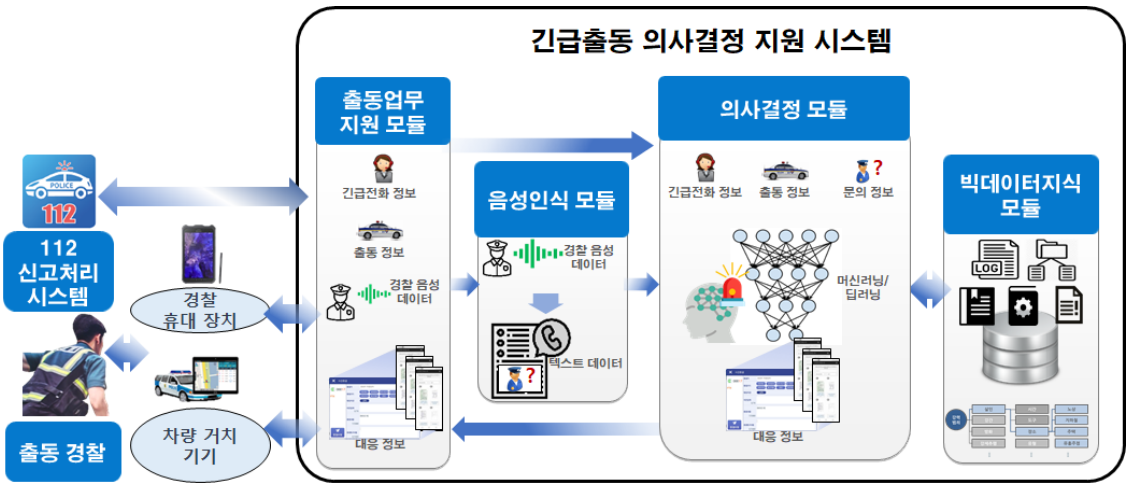


그림. 1. 긴급출동 의사결정 지원 시스템의 아키텍처

음성인식 모듈은 경찰 음성 데이터를 STT(Speech-To-Text) 기술을 활용하여 텍스트 데이터로 변환시킨다. 경찰 음성 데이터에는 사건 현장에 대한 대응 질문들을 포함될 수 있으며, 질문에는 사건과 관련 있는 법률, 지침, 유사한 이전 사례에 대한 질문과 같은 문의 정보도 포함이 된다. 음성인식 모듈에 사용되는 STT 기술에 BLSTM-HMM(Bidirectional Long Short-Term Memory Hidden Markov Model)과 같은 딥러닝(deep-learning) 모델을 사용할 수 있다.

의사결정 모듈은 대응 정보를 생성시키는 모듈이다. 의사결정 출동업무지원 모듈로부터 받은 긴급전화 정보와 출동 정보를 이용하고 음성인식 모듈에서 경찰관의 문의 정보를 머신러닝(machine learning)이나 딥러닝 기술을 이용하여 사건현장 상황을 분석할 수 있다. 사건현장의 상황을 분석할 때 NLP(Natural Language Processing)을 이용한 분석을 하고, 사건현장의 위험도를 분석한다. 사건현장의 위험도를 분석하기 위해 연구 [5]의 위험도 분석방법이 쓰일 수 있다. 그리고, 또한 경찰의 문의 정보와 분석된 사건현장 상황을 기반으로 출동 경찰에게 적절한 대응 정보를 생성할 수 있다. 대응 정보를 생성하기 위해서 빅데이터지식베이스 모듈로부터 법률, 지침, 이전 사례에 대한 정보를 요청하고 받을 수 있다.

빅데이터지식베이스 모듈은 법률, 지침, 이전 사례에 대한 정보를 저장하고 관리하는 빅데이터베이스(big data base)이다. 빅데이터지식베이스 모듈은 저장된 데이터를 정형화하고 사건종류별로 분류한다. 의사결정 모듈이 빅데이터지식베이스 모듈에 문의 정보를 전달하면, 문의 정보에 대응하는 법률, 지침, 사례에 대한 정보를 의사결정 모듈에 전달한다.

Ⅲ. 결론

본논문에서는 경찰의 사건현장 대응 지원을 위한 긴급출동 의사결정 지원 시스템을 제안하였다. 긴급출동 의사결정 지원 시스템은 112 신고처리 시스템으로부터 긴급전화 정보를 받고, 출동 경찰의 경찰 휴대 장치 및 차량 거치 기기로부터 출동 정보 및 경찰 음성 데이터를 받아서 사건현장의 상황을 분석한다. 그 후에 긴급출동 의사결정 지원 시스템은 사건현장의 상황에 적합한 대응 정보를 제공한다. 현장에 출동한 경찰은 긴급출동 의사결정 지원 시스템으로부터 받은 대응 정보를 바탕으로 적절히 현장상황에 대응하여 사건현장의 피해를 줄일 수 있을 것이다. 향후에는 긴급출동 의사결정 지원 시스템을 위한 구체적인 알고리즘에 대한 연구를 진행할 것이다.

ACKNOWLEDGMENT

이 논문은 2021 년도 정부(경찰청)의 재원으로 지원받아 수행된 연구결과임 [내역사업명: 112 긴급출동 의사결정 지원 시스템/ 연구개발과제번호: PR08-03-000-21]

참 고 문 헌

[1] 경찰통계자료, 2019, (https://www.police.go.kr/www/open/publice/publice0208.jsp)

[2] 방준성 외, “지능형 치안 서비스 기술 동향”, 전자통신동향분석, 제 34 권, 1 호, pp. 111-122, 2019.02.

[3] Z. Friend, “Predictive policing: using technology to reduce crime,” FBI Law Enforcement Bulletin, February 8, 2018.

[4] C. Francescani , “NYPD expands surveillance net to fight crime as well as terrorism,” Reuters, June 21, 2013.

[5] M-S Baek et al., “Design and performance evaluation of crime type and crime risk score estimation technique for fast and efficient response of severe crimes,” 2020 International Conference on Information and Communication Technology Convergence (ICTC), pp. 1826-1828, October 2020.

시공간 범죄 위험도 추론 기술 기반의 스마트 치안 플랫폼 개발 및 검증

백명선, 박원주, 이희조*, 이용태
한국전자통신연구원, *㈜모비젠

sabman@etri.re.kr, wjpark@etri.re.kr, *heejo@mobigen.com, ytlee@etri.re.kr

Development and Verification of Smart Policing Platform based on Space-Time Crime Risk Score Prediction Technology

Myung-Sun Baek, Wonjoo Park, Heejo Lee*, Yong-Tae Lee
ETRI, *Mobigen Co., Ltd.

요 약

본 논문은 시공간 범죄 위험도 추론 기술 기반의 스마트 치안 플랫폼을 개발하고 성능을 검증한다. 개발된 플랫폼은 시공간에 따른 범죄 위험도를 시각화 하여 지도에 표출하는 것이 가능하며, 시공간 치안 관련 다양한 비교/분석 기능을 제공한다. 따라서 개발된 시공간 범죄 위험도 추론 기반 스마트 치안 플랫폼을 활용하면 시공간에 따른 고위험지역을 분석하여 순찰 인력 배치 효율화, CCTV 추가 설치 지역 선별 등등 치안 현장 관리를 보다 효과적으로 수행할 수 있다.

Ⅰ. 서 론

최근 인공지능 및 빅데이터 기술을 발판으로 과학기술은 비약적으로 성장하고 있다. 국민의 생명과 재산을 지키고 안전한 삶을 보장하기 위한 치안 분야에서도 고도화된 치안기술 개발을 위한 인공지능/빅데이터 적용이 활발하게 연구되고 있다[1]. 가장 주목받고 있는 인공지능/빅데이터 기반의 치안 기술은 경찰의 누적된 데이터를 활용한 스마트 치안기술이다 [1].

본 논문에서는 시공간에 따른 위험도를 추론할 수 있는 스마트 치안 플랫폼을 개발하고 성능을 검증한다. 시공간 위험도 추론을 위해 11년간 누적된 치안데이터를 활용하였다. 누적된 치안데이터를 분석하여, 범죄 죄종 및 범죄 발생빈도 기반의 '범죄 발생 강도'와 피해자의 범죄 피해기반의 '피해점수'를 산출하여 범죄 위험도를 생성하였다. 또한 기계학습 기반으로 시공간에 따른 범죄 위험도를 추론/분석할 수 있는 기술을 개발하였다. 이러한 기술을 바탕으로 전국단위의 시간과 공간에 따른 범죄 위험도를 추론 및 표출할 수 있는 스마트 치안 플랫폼을 개발하였다. 개발된 스마트 치안 플랫폼은 직관적인 GUI 형태로 구현되어 현장 요원이 어려움없이 사용할 수 있도록 하였다. 또한 지역기반 시공간 위험도 표출 뿐만 아니라 각 지역별 범죄 위험도 비교 분석 기능을 지원하여, 범죄 위험도 추론과 관련된 다양한 분석에 활용될 수 있도록 개발되었다.

Ⅱ. 시공간 범죄 위험도 추론 기술 개발

시공간 범죄 위험도는 범죄 피해점수와 범죄 발생 강도를 기반으로 개발되었다. 범죄 피해점수는 피해자가 범죄에 얼마나 취약하며 범죄를 통해 신체/재산상 어떠한 피해를 입었는지를 나타내는 지표이다. 아래는 범죄 피해점수를 산출하는 산출식이다.

$$\text{범죄피해점수} = \text{신체피해점수} \times \text{성별가중치} \times \text{연령별가중치} \quad (1)$$

신체, 성별 및 연령별 피해의 가중치는 기존에 수립한 범죄위험스코어의 가중치를 활용하였다 [2]. 이와 더불어 범죄 발생 강도를 산출하는 산출식은 월별, 요일별, 시간대별 발생 강도로 구성되며 다음과 같다.

$$\text{월별 발생 강도} = \frac{\textit{해당 지역 월 발생 건수}}{\textit{전국 월 평균 범죄 발생건수}} \quad (2)$$

$$\text{요일별 발생 강도} = \frac{\textit{해당 지역 요일 발생 건수}}{\textit{전국 요일평균 범죄 발생건수}} \quad (3)$$

$$\text{시간별 발생 강도} = \frac{\textit{해당지역 시간 발생 건수}}{\textit{전국 시간평균 범죄 발생건수}} \quad (4)$$

상기 식을 활용하여 시간 및 공간 별 범죄 위험도는 다음과 같이 산출할 수 있다.

범죄 위험도 = α · 범죄피해점수 + β · 범죄발생강도 (5)

상기 식에서 α 와 β 는 조절 계수이다. 조절 계수는 범죄 발생강도와 범죄 피해점수 사이의 상관도와 영향도를 분석하여 설정되었다.

이렇게 산출된 범죄 위험도는 기계학습 기반으로 시공간에 따라 추론값을 도출하여 지도에 표출한다.

Ⅲ. 시공간 범죄 위험도 추론 기술 기반 스마트 치안 플랫폼 및 동작 검증

그림 1 은 시공간 범죄 위험도 추론 기술 및 표출기술이 포함된 스마트 치안 플랫폼이다. 개발된 플랫폼은 대표적인 3 종의 범죄 유형인 강력, 폭력, 절도에 따른 시공간 범죄위험도, 범죄피해점수 및 범죄발생강도를 나타낼 수 있다. 시간의 경우 월, 요일(일), 시간 별 위험도를 보여주는 것이 가능하며, 공간의 경우 전국 및 지방 단위의 위험도를 나타낼 수 있다.

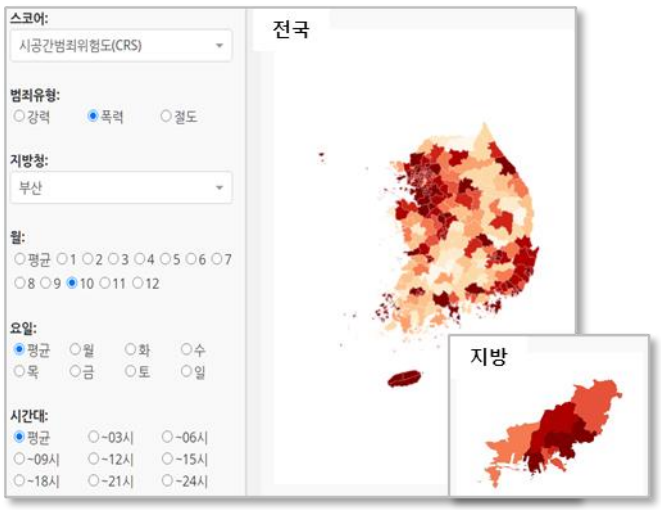
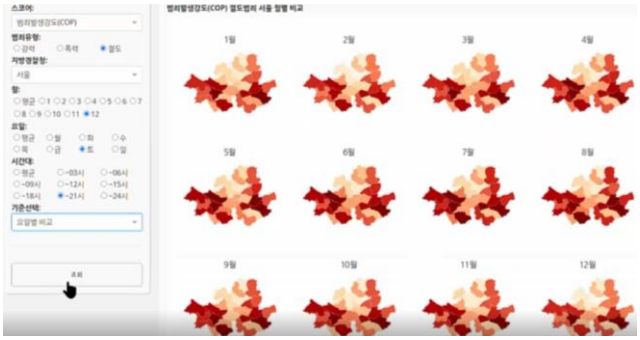


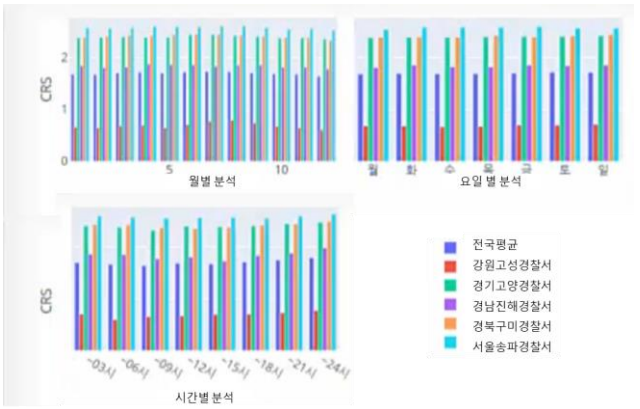
그림 1. 시공간 범죄 위험도 추론 기반 스마트 치안 플랫폼

또한 개발된 플랫폼은 다양한 분석 기능을 제공할 수 있다. 그림 2 는 해당 스마트 치안 플랫폼이 제공하는 분석 기능을 보여준다. 그림 2-(a)는 각 지방의 월별 범죄위험도를 비교 분석할 수 있는 기능을 보여준다. 해당 기능은 월에 따른 분석 뿐만 아니라, 요일별, 시간별 분석도 가능하다. 그림 2-(b)는 지방 경찰서 별 범죄 위험도를 비교 분석할 수 있는 기능을 보여준다. 따라서 특정 시간 및 기간에 지역별 위험수준을 한번에 비교분석 할 수 있다. 제공되는 분석 플랫폼이 인구수 정보, 유흥업소 밀집지역 정보, 유동 인구 정보, 날씨 및 기후 정보 등 다양한 공공 통계데이터와 접목하여 분석하게 되면 인구 수별, 유동인구 별, 유흥업소 밀집 지역별, 유동 인구 별 및 기상 별 위험도를 추론하는 것도 가능하다.

따라서 상기 개발된 시공간 범죄 위험도 추론 기반 스마트 치안 플랫폼을 활용하면 시공간에 따른 고위험 지역을 분석하여 순찰 인력 배치 효율화, CCTV 추가 설치 지역 선별 등등 치안 현장관리를 보다 효과적으로 수행할 수 있다.



(a) 각 지방의 월별 범죄 위험도 표출 및 분석 기능



(b) 지방 별 범죄 위험도, 범죄피해점수, 범죄발생강도 비교 분석 기능

그림 2. 스마트 치안 플랫폼이 제공하는 치안데이터 기반의 분석 기능

Ⅲ. 결론

본논문에서는 시공간 범죄 위험도 추론 기술 기반의 스마트 치안 플랫폼을 개발하고 동작을 검증하였다. 개발된 플랫폼은 직관적인 GUI 를 제공하고 있어, 현장요원들이 쉽게 해당 플랫폼을 사용하는 것이 가능하며, 시공간 위험도 추론 기능 뿐만 아니라 다양한 분석기능을 제공할 수 있다. 따라서 개발된 스마트 치안 플랫홈을 활용하면 보다 효과적으로 치안현장을 관리하는 것이 가능하다.

ACKNOWLEDGMENT

이 논문은 2021 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2018-0-00440, 위험 상황 초기 인지를 위한 ICT 기반의 범죄 위험도 예측 및 대응 기술 개발).

참 고 문 헌

[1] 백명선외 5 명, “효율적인 초동 수사를 위한 지도/준지도 학습 기반의 범죄 사건 정보 예측 및 용의자 후보군 특정 기술 연구,” 한국통신학회 하계학술대회, 2021, 6.

[2] M. -S. Baek et al., "Design and Performance Evaluation of Crime Type and Crime Risk Score Estimation Technique for Fast and Efficient Response of Severe Crimes," 2020 International Conference on Information and Communication Technology Convergence (ICTC), 2020, pp. 1826-1828.

KorBERT 와 능동 학습 기법을 활용한 한국어 기반 유형 분류 모델에 관한 경험적 연구

박원주, 김연수*, 백명선, 이용태
한국전자통신연구원 국방·안전 ICT 연구단, *이화여자대학교 컴퓨터공학과
[wjpark, sabman, ytlee]@etri.re.kr, *wx12348@i.ewha.ac.kr

An Empirical Study on the Korean Text Classification based on Active Learning with KorBERT

Wonjoo PARK, Yeon-Soo Kim*, Myung-Sun Baek, Yong-Tae Lee
ETRI, *Ewha Womans University

요 약

능동 학습 기법은 지도 학습에서 데이터 레이블링을 위한 비용과 시간을 절감하기 위하여 널리 사용되고 있다. 최근 BERT 는 다양한 자연어 처리 태스크에서 높은 성능을 보여주고 있다. 본 논문은 ETRI KorBERT 모델과 능동 학습 기법을 활용하여 한국어 텍스트의 유형 분류 모델의 성능을 개선한 경험적 연구 결과를 제안한다. 특히 전문 지식이 필요한 치안 도메인에 다양한 능동 학습 기법을 적용하여 데이터 구축 비용 등 예산 절감 효과를 제시하고, 실제 데이터와 최신 딥러닝 모델을 결합한 실험 결과를 제공한다.

Ⅰ. 서 론

텍스트 유형 분류 태스크는 자연어 처리 분야에서 오랫동안 연구되고 있고 좋은 성능을 보여주고 있는 분야이다. 특히 구글의 BERT 모델과 코드가 공개된 이후 자국의 언어를 활용한 사전학습모델을 생성하고, 미세조정한 후, 다양한 과업에 적용하여 좋은 성능을 제시하고 있다. 한국어는 ETRI 에서 KorBERT 모델을 공개한 이후, SKT 의 KoBERT, 투블럭 AI 의 HanBERT 와 같은 한국어 사전학습모델이 공개되고 있다. 지도학습모델인 KorBERT 를 텍스트 유형 분류 태스크에 적용하기 위해 데이터 레이블링 작업이 선행되어야 한다. 그러나 데이터 레이블링은 많은 비용과 시간을 요구하고, 특히나 실제 환경에서는 도메인 전문 지식이 필요하다[1]. 이 때 능동 학습은 정보량이 풍부한 데이터, 즉 성능 향상에 효과적인 인스턴스들을 추출하여 데이터 레이블링을 수행하되, 적은 데이터로 최적의 성능을 달성하는 것을 목적으로 한다[2]. 본 논문에서는 치안도메인에서 다양한 능동 학습 기법을 적용하여 데이터 레이블링 비용과 시간을 절감한 효과를 제시함으로써, 실제 데이터와 KorBERT 모델을 결합한 실험 결과를 살펴본다.

Ⅱ. 능동 학습과 KorBERT 를 결합한 사건 유형 추론 모델 학습 및 검증

본 연구는 과거의 112 신고 내용을 학습하여 신규 신고 접수 시 사건 유형을 추론하는 기술 개발을 목적으로 하고 있다. 이 때, 사건 유형은 사회가 변화함에 따라 신규 유형이 추가되기도 하고, 경찰의 적절한 대응을 위하여 보다 세분화되기도 한다. 실제 현장에서는, 일반적으로 적절한 시스템 운영을 위하여 과거의 유형의 삭제는 최소화하고, 신규 유형을 추가하도록 설계되어야 한다. 예를 들어 '놀이터에서 청소년들이 크게 노래를 불러서 시끄럽다. 순찰 요망'이라는 신고가 접수되었을 때, 이 신고는 과거에는 '기타경범'이라는 유형으로 처리하였으나, 소음에 대한 신고량이 많아지고, 소음 발생 상황 등에 따라 적절한 대응이 필요하여 '소음'이라는 유형이 추가될 수 있다. 그러나 과거의 신고내용 중 소음 관련된 신고내용은 기타경범으로 이미 분류된 사례가 적지 않을 것이다. 이렇게 포괄적으로 운영되던 분류의 일부가 세분화되고, 이를 포함하여 전체 클래스를 지도 학습하여 유형을 분류하는 모델을 생성하기 위하여 본 연구는 포괄적 유형에 능동 학습을 적용한다. 112 의 대표적인 포괄적 유형은 '기타형사범'과 '기타경범'이다. '기타형사범'은 동물학대, 마약, 주거침입, 재물손괴 등 형사 사건 관련 유형과 신고 시 두개 이상의 범죄행위가 발생한 경우를 포함하고 '기타경범'은 소음, 교통불편, 보호조치, 노상방뇨, 무단투기 등 질서유지가 필요한 유형을 포함하고 있다. 본 시스템은 전체 112 신고사건 중 기타형사범(others1)과 기타경범(others2)으로 분류된 인스턴스만을 추출하여 능동 학습을 적용함으로써 포괄적 유형으로 분류된 인스턴스 중 정보량이 풍부한 데이터를 선별하여, 재레이블링을 수행하고, 이 인스턴스를 포함하여 전체 클래스를 분류하는 모델을 학습하여 사건 유형 분류 성능을 향상시키고자 한다. 그림 1 은 능동 학습을 사용하여 질의전략에 따른 데이터 샘플링과 레이블링 작업 프로세스를 보여준다.

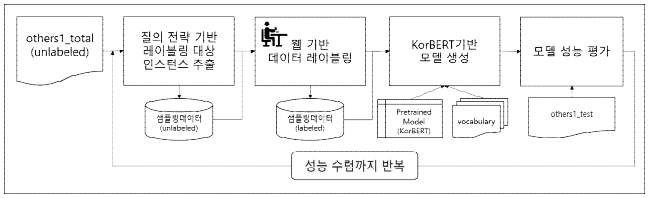


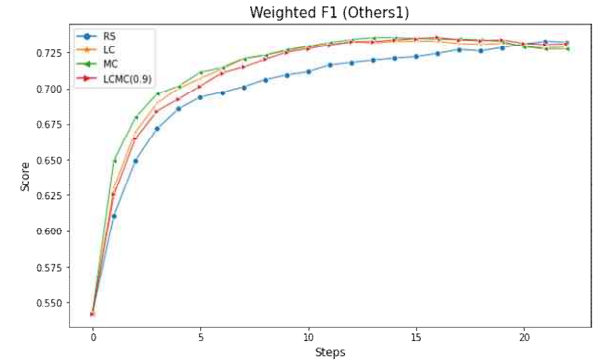
그림 1. 능동 학습 기반 데이터 구축 프로세스

능동 학습을 적용하기 위하여 대상 인스턴스를 샘플링하는 질의 전략은 다양하다. 본 연구는 베이스로 RS(Random Sampling)을 선정한 후, 능동 학습에서 많이 사용하는 LC(Least Confidence), MC (Margin Confidence)를 사용한다. LC 는 모델이 예측한

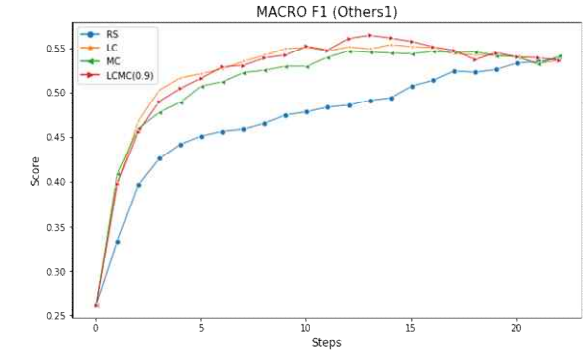
확률값이 가장 낮게 측정한 인스턴스를 우선 사용하는 방법으로, 확신도가 낮은 인스턴스를 우선 선택하는 질의 전략이다. MC 는 LC 전략이 모델에서 하나의 확률값만 사용하는 것을 보완하기 위해, 클래스 확률이 가장 큰 값과 두번째로 큰 값의 차이가 가장 작은 인스턴스를 우선 선택하는 전략이다[3]. 이것은 두 라벨의 차이가 적어 모호함을 많이 포함하는 인스턴스를 위주로 데이터 레이블링을 구축하려는 목적이다. 또한 본 논문은 LC 질의 전략이 확신도가 낮은 인스턴스를 우선 선택하여 모델을 학습시키는 경우, 일반성이 떨어지고 특성 인스턴스에 모델이 과적합될 수 있어, 확신도가 낮은 인스턴스를 샘플링하는 비율(α)과 확신도가 높은 인스턴스를 샘플링하는 비율(β)을 결합하여 질의하는 전략인 LCMC(Least Confidence & Most Confidence)를 제안한다. α 와 β 의 합은 1 이고, 사용자가 조정한다. 표 1 은 기타형사범(others1)과 기타경범(others2)에 대하여 능동 학습을 진행한 실험 환경이고, 그림 2 는 실험 결과이다.

표 1. 능동학습 실험 환경

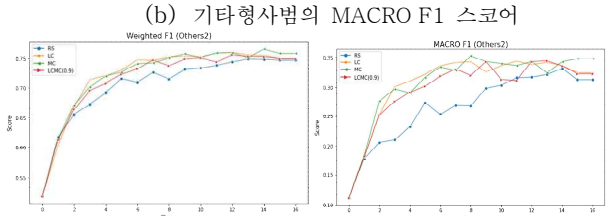
	기타형사범	기타경범
전체 인스턴스의 수	399,459	33,093
스텝당 인스턴스의 수	3,000	400
배치크기(batch size)	32	8
에포크(epoch)	3	3
학습률(learning rate)	2e-5	2e-5
테스트 인스턴스의 수	19,000	1,600



(a) 기타형사범의 Weighted F1 스코어



(b) 기타형사범의 MACRO F1 스코어



(c) 기타경범의 Weighted F1 (d) 기타경범의 MACRO F1

그림 2. 각 질의 전략에 따른 테스트 실험 결과

실험 결과는 유형별 데이터 인스턴스의 불균형을 고려하여 Weighted F1, MACRO F1 두 지표로 평가하였다. 결과를 분석하면, 레이블링을 위한 인스턴스 선택 시 무작위로 선택하는 RS 에 비해 LC, MC, LCMC(α :0.9) 등 능동 학습에 따른 인스턴스를 선택하는 기법이 성능 수렴에 빠르게 도달한다는 것을 알 수 있다. 기타형사범의 경우, RS 전략으로 60,00 인스스턴스를 학습해야 Weighted F1 스코어가 73%에 도달하는 데에 비해 LC, MC, LCMC 질의 전략으로 33,000~36,000 인스턴스를 사용하여 유사한 성능을 달성할 수 있다. 특히 MACRO F1 스코어의 경우 LCMC 질의 전략을 적용한 경우 가장 높은 결과를 보임에 따라, LC 질의 전략이 가질 수 있는 과적합을 보완할 수 있음을 확인하였다. 112 사건 유형 중 포괄적 유형으로 분류되었던 기타형사범과 기타경범을 능동 학습에 따라 재분류함으로써, 전체 신고데이터에 대한 사건 유형 분류 성능은 표 2 와 같이 MACRO F1 스코어가 약 11.6% 향상되었다.

표 2. 112 사건 유형 분류 성능

	AL 적용 전	AL 적용 후
Weighted F1	80.5%	85.4%
MACRO F1	67.0%	75.8%

Ⅲ. 결 론

본 논문은 과거의 신고내용 텍스트를 학습하여 신규 사건에 대한 유형을 추론하는 모델을 학습할 때, 기존 신고 사건에 대한 유형 레이블링의 품질을 보완하기 위한 능동 학습 기법을 개발하고 검증하였다. 특히 기존의 능동 학습 기법의 단점을 보완하기 위하여, 확신도가 낮은 인스턴스뿐만 아니라 확신도가 높은 인스턴스를 샘플링하여 모델을 학습시킴으로써, 과적합에 건강한 질의 전략을 새롭게 제안하고, 실험 결과로 검증하였다. 또한 능동 학습 기법으로 선택된 인스턴스를 사건 유형 추론 모델 학습에 활용하여 사건 유형 추론 성능을 향상시켰다. 향후 이 모델이 치안현장에서 112 신고 접수 시 경찰관이 사건 유형을 결정할 때, 의사지원 솔루션으로 활용될 것을 기대한다.

ACKNOWLEDGMENT

이 논문은 2021 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2018-0-00440, 위험 상황 초기 인지를 위한 ICT 기반의 범죄 위험도 예측 및 대응 기술 개발).

참 고 문 헌

[1] 정동호, 허민강, 김형철, 박상원, “능동 학습 기법을 활용한 한국어 금융 도메인 개체명 인식 데이터 구축”, 한글 및 한국어 정보처리학술대회 논문집(2020), pp.82-85

[2] Liat Ein Dor, Alon Halfon, Ariel Gera, Eyal Shnarch, Lena Dankin, Leshem Choshen, Marina Danilevsky, Ranit Aharonov, Yoav Katz, and Noam Slonim, ”Active Learning for BERT: An Empirical Study”, EMNLP2020, .pp.7949-7962, Nov. 2020

[3] Burr Setteles, “Active Learning Literature Survey”, University of Wisconsin-Madison. 2009

저품질 영상의 초고화질 변환을 위한 실시간/비실시간 처리 시스템 설계

라상중, 조숙희, 김나연*
한국전자통신연구원, *과학기술연합대학원대학교

sjna@etri.re.kr

Real-time/Non-real-time Processing System Design for Ultra-High-Definition Conversion of Low-Quality Videos

Sang-Jung Ra, Sukhee Cho, Na-Yeon Kim*
Electronics and Telecommunications Research Institute, *University of Science & Technology

요 약

본 논문에서는 낮은 해상도의 저품질 영상을 초고화질 영상으로 변환하기 위한 실시간/비실시간 처리 시스템에 대한 구조 및 설계 방안에 대하여 제안한다. 우선 낮은 해상도의 영상을 초고화질 영상으로 변환하는데 필요한 핵심요소기술에 대하여 정의하고, 이러한 요소기술이 적용된 시스템의 실시간/비실시간 처리를 위한 구조설계를 제안한다. 이를 통해 최종적으로 상용화 수준의 통합 시스템을 구현하는 것을 목표로 한다.

Ⅰ. 서 론

최근 딥러닝 기술을 적용하여 낮은 해상도의 저품질 영상을 초고화질로 변환하기 위한 기술에 대한 수요 증가에 따라 관련 연구가 많이 수행되고 있으나, 각각의 요소기술에 대한 연구에 편중되어 성능향상에 한계를 보이고 있다. 이러한 문제점을 극복하기 위해 변환 요소들을 복합적/적응적으로 적용하여 초고화질 영상으로 변환할 수 있는 기술 개발 및 이를 통한 성능의 고도화가 필요하다.

본 논문에서는 이러한 환경을 고려하여 낮은 해상도의 저품질 영상을 초고화질 영상으로 변환하기 위한 실시간/비실시간 처리 시스템에 대한 구조설계를 제안한다. 구체적으로 낮은 해상도의 영상을 초고화질로 변환하는데 필요한 핵심요소기술에 대해 설명하고, 실시간/비실시간 처리를 위한 통합 시스템의 구현방안을 제시한다.

Ⅱ. 저품질 영상의 초고화질 영상 변환 기술

저품질 영상이라 함은 과거 아날로그 수준부터 최대 60p 수준의 SD 품질 영상 콘텐츠를 의미한다. 이러한 영상은 다양한 과거 방송 콘텐츠부터 구형 디지털카메라 및 CCTV 등을 통해 녹화된 형태까지 다양한 형태로 제작된 영상을 포함할 수 있다.

이러한 낮은 해상도의 저품질 영상을 초고화질로 변환하기 위해서는 더욱 넓은 색영역을 지원하기 위한 WCG(Wide Color Gamut)와 더욱 섬세하게 음영을 처리할 수 있는 HDR(High Dynamic Range) 및 기존보다 초당 전송되는 프레임율을 높여 화면 전환이 더욱 용이한 HFR(High Frame Rate) 기술이 필수적으로 지원되어야 한다[1][2]. 이러한 핵심요소기술에 대하여 각각의 알고리즘을 딥러닝 기반으로 구현한 후 알고리즘간 결합을 통해 초고화질 영상으로 변환하는데 따른 통합 성능을 얻을 수 있다.



그림 1. 저품질 영상의 초고화질 변환을 위한 요소기술

Ⅲ. 초고화질 영상 변환을 위한 실시간/비실시간 처리 시스템 설계

낮은 해상도의 저품질 영상을 초고화질의 영상으로 변환하기 위한 시스템은 동작환경에 따라 실시간 및 비실시간 처리 시스템으로 구분할 수 있다.

우선 비실시간 처리 시스템의 경우 사용자의 요구사항에 따른 선택적 맞춤형 변환을 지원한다. 이

시스템은 분석된 원본 영상의 특성에 따라 최적의 핵심요소기술에 대한 조합을 적용한 후, 이를 사용자의 의도에 따라 동적으로 조정할 수 있도록 하여 사용자가 원하는 최적의 화질 개선을 수행할 수 있도록 한다.

실시간 처리 시스템의 경우에는 비실시간 처리 시스템에 비해 고속화 및 경량화 된 핵심요소기술을 적용하여 영상의 화질 개선을 수행할 수 있도록 한다.



그림 2. 초고화질 변환을 위한 실시간/비실시간 변환 시스템 개념

초고화질 변환을 위한 실시간/비실시간 시스템의 구현과 관련하여 스탠드얼론 형태의 시스템은 비실시간 적용 환경만을 고려하는 시스템이다. 별도의 데이터베이스를 통해 보관중인 저품질의 원본 콘텐츠에 대한 시간 및 공간 복잡도를 분석하고, 분석 결과에 따라 시스템에서 최적의 핵심요소기술을 적용한 후 사용자의 의도를 추가로 반영하여 화질 개선을 수행한다.

또다른 시스템의 구현 형태로 클라우드 기반의 변환 솔루션을 고려한다. 클라우드 기반의 변환 솔루션은 실시간 및 비실시간 적용 환경을 모두 고려하는 형태의 솔루션이며, 실시간 적용의 경우 알고리즘을 경량화 및 고속화하여 동작을 수행하도록 한다.

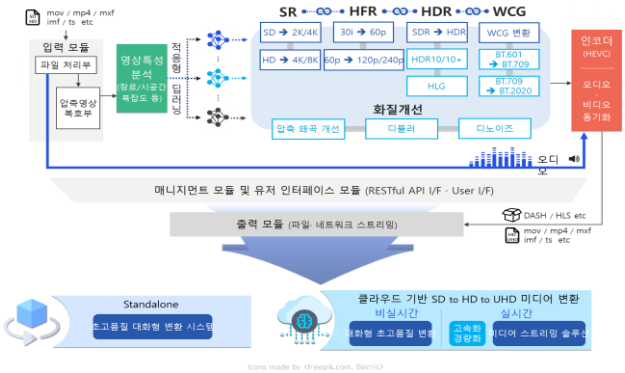


그림 3. 초고화질 변환을 위한 실시간/비실시간 변환 시스템 구조 설계

클라우드 기반의 변환 솔루션을 구현하기 위한 기술로 영상 변환을 위하여 MPEG 표준에서 정의하고 있는 NBMP(Network-based Media Processing) 규격을 고려하고 있다[3]. NBMP 표준은 저지연 환경에서 동영상 콘텐츠를 처리하기 위한 규격이며, 5G MEC 환경과의 연동 등에 많이 활용되고 있는 기술이다.

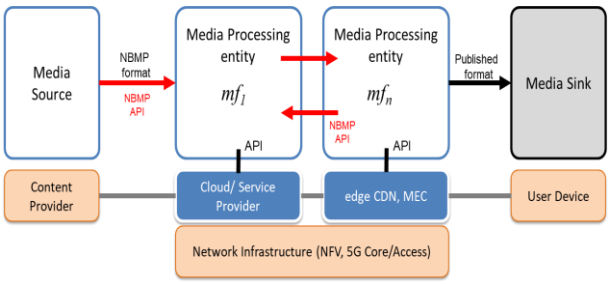


그림 3. NBMP 규격의 개념도

Ⅳ. 결 론

본 논문에서는 저품질 영상의 초고화질 영상 변환을 위한 시스템을 구현하는데 필요한 핵심요소기술을 정의하고 각 요소기술의 연동을 통한 실시간/비실시간 처리 시스템을 설계하는 것을 목적으로 한다. 향후 이러한 연구결과를 활용하여 상용화 수준의 통합 시스템을 구현하여 낮은 해상도의 저품질 동영상에 대한 화질 개선을 수행하여 구작 방송 콘텐츠의 복원 및 저화질 CCTV 분석을 통한 범죄 예방 등 다양한 분야에서 중요한 역할을 하기를 기대한다.

ACKNOWLEDGMENT

"이 논문은 2021 년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No.2021-0-00087, SD/HD 급 저화질 미디어의 고품질 변환 기술 개발)"

참 고 문 헌

[1] Woonsung Park, Munchurl Kim, "Deep Predictive Video Compression using Mode-Selective Uni- and Bi-directional Predictions based on Multi-frame Hypothesis", IEEE Access., vol. 9, pp. 72-85, Dec. 2020

[2] Sehwan Ki, Jeonghyeok Do, Munchurl Kim, "Learning-based JND-directed HDR Video Preprocessing for Perceptually Lossless Compression with HEVC", IEEE Access., vol. 8, pp. 228605-228618, Dec. 2020

[3] ISO/IEC 23090-8:2020, Information technology — Coded representation of immersive media — Part 8: Network based media processing, 2020.

기계학습 기반의 범죄 유형 및 범죄위험스코어 예측 기술 성능 향상 연구

안다영, 백명선*, 박원주*, 이용태*
한양대학교 ERICA, *한국전자통신연구원

helloda0@hanyang.ac.kr, *{sabman, wjpark, ytlee}@etri.re.kr

A Study on the Performance Improvement of Crime Risk Score Estimation Technique based on Machine Learning

Dayoung Ahn, Myung-Sun Baek*, Wonjoo Park*, Yong-Tae Lee*
Hanyang Univ. ERICA, *ETRI

요 약

본 논문은 기존 심각 범죄의 효과적 대응을 위한 기계학습 기반 범죄유형 및 범죄위험스코어 예측 기술의 성능 향상을 위한 새로운 모델을 제안한다. 기계 학습을 통해 범죄유형을 예측하고 그 결과 값을 범죄위험스코어 예측에 이용한다. 예측 기술 성능의 향상으로 오류율을 낮춰 더 높은 신뢰성을 확보할 수 있고, 이 기술의 연구 목적인 심각 범죄의 효과적 대응에 기여하는 것이 가능하다.

Ⅰ. 서 론

심각 범죄의 빈도가 증가하면서 효과적인 대응을 위해 빅데이터/인공지능 기반 ICT 기술을 활용한 지능형 치안 관련 기술의 개발이 활발히 진행중에 있다[1]-[2]. 그의 일환으로 기계학습 기반의 범죄 유형 및 범죄위험스코어 예측 기술에 관한 연구[3]가 진행되었다. 이 연구에서는 범죄 위험 심각도를 수치화하여 범죄위험스코어를 설계하였다. 또한 신고 접수 단계에서 범죄유형, 범죄위험스코어를 예측하고 제공하는 시스템을 개발하였다. 개발된 시스템을 활용하여 신고 접수 경찰은 즉시 출동 여부/인력 배치 등 효과적인 초동 대응을 수행할 수 있다.

본 연구는 이전 연구 기술의 성능을 향상시키는 것에 목적을 두고 진행되었다. 모델 두 개를 이용하여 범죄위험스코어를 예측하고, 모델의 오류율을 줄인다. 이로써 시스템의 더 높은 신뢰성을 확보하는 것이 가능하다.

Ⅱ. 구 조

그림 1 은 성능 향상을 위해 제안된 범죄유형 및 범죄위험스코어 예측 시스템의 구조도이다. 기존과 마찬가지로 텍스트 형식의 Reference data 로부터 WordRank[4], TF-IDF[5]를 이용한 Keyword 를 추출하여 Dictionary 를 구축한다. 구축된 Dictionary 와 Reference data 를 이용하여 Training dataset 과 Test dataset 을 생성한다. Training dataset 을 바탕으로 범죄 유형을 예측하는 Model1 을 생성하고 범죄위험스코어를 예측하는 Model2 를 생성한다. Model1 과 Model2 가 서로 독립적이었던 이전 연구와는 다르게 본 연구에서는 Model1 에서 예측한 결과를 범죄위험스코어 예측에 활용한다. Model1 을 이용하여 범죄유형을 예측하고 해당 결과를 Model2 로 입력하여, Model2 를 위한 새로운 Feature 로 사용한다.

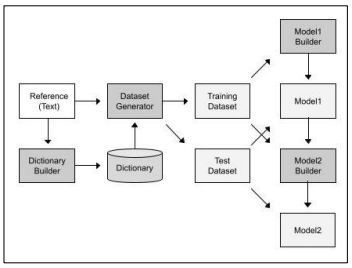


그림 1. 범죄유형/범죄 위험스코어 예측 시스템 구조도

2.1 범죄유형 예측을 위한 SVM

지도 학습 모델인 서포트 벡터 머신(Support Vector Machine, SVM)을 활용하여 범죄유형 예측 모델을 생성한다. SVM 중에서도 분류 문제를 위한 SVC(C-Support Vector Classification)를 이용한다. SVM 은 기존에 범죄유형 예측을 위해 사용되던 CNN 모델에 비해 계산량이 현저하게 적다는 특징이 있다. 개발된 모델에서는 마진을 최대로 하여 일반화 능력을 향상시켰으며 K-Fold 교차 검증과 Grid Search 방법을 적용하여 최적의 하이퍼파라미터를 찾는다. SVM 은 C 값이 10 인 Linear Kernel 로 구현되었다.

2.2 범죄위험스코어 예측을 위한 CNN

합성곱 신경망(Convolution Neural Network, CNN)을 이용하여 범죄위험스코어 예측 모델을 생성한다. 이 모델은 그림 2 처럼 Convolution layer, Dense layer, Flatten layer 와 Dropout layer 로 구성된다. Convolution layer 와 Dense layer 에서는 활성화 함수로 Relu 가 사용되었고 과대적합을 방지하기 위하여 Dropout layer 를 이용한다. Optimizer 로 Rmsprop 가 사용되었고 Loss 함수로 MSE(Mean Squared Error)가 사용되었다.

표 1. 모델 별 범죄유형 예측 정확도(Accuracy)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	전체
DNN	0.98	**0.95**	0.97	0.97	0.86	0.79	0.80	0.93	0.66	0.50	0.91	**0.94**	0.73	0.69	0.47	0.50	**1.00**	0.40	0.88	**1.00**	**1.00**	0.88
CNN	0.95	0.89	**0.98**	0.96	**0.98**	0.74	0.73	**0.97**	**1.00**	0.71	0.93	0.86	0.71	**0.70**	**0.57**	**0.57**	0.89	0.75	**1.00**	**1.00**	**1.00**	0.90
M1	**0.99**	0.89	0.97	**1.00**	0.93	**0.87**	**0.88**	0.94	**1.00**	**0.89**	**0.99**	0.93	**0.82**	0.67	0.40	**0.57**	0.89	**0.80**	**1.00**	0.89	0.91	**0.94**

표 2. 모델 별 범죄위험스코어 예측 오류율(MAPE)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	전체
DNN	13.6	19.8	36.1	21.6	26.0	13.9	10.9	17.5	24.0	41.0	37.9	32.4	25.2	25.0	35.1	39.5	**36.3**	22.9	22.8	**19.4**	87.6	24.6
CNN	8.3	**11.9**	58.9	13.3	15.6	8.7	7.8	12.3	42.3	34.8	28.3	32.8	18.4	17.8	21.3	34.0	64.8	17.6	11.4	31.9	75.4	21.0
M2	**4.7**	15.9	**11.4**	**6.4**	**7.9**	**6.8**	**6.7**	**7.8**	**21.3**	**8.7**	**10.4**	**11.5**	**9.0**	**10.1**	**13.9**	**24.3**	46.4	**12.8**	**5.6**	33.0	**66.0**	**9.1**

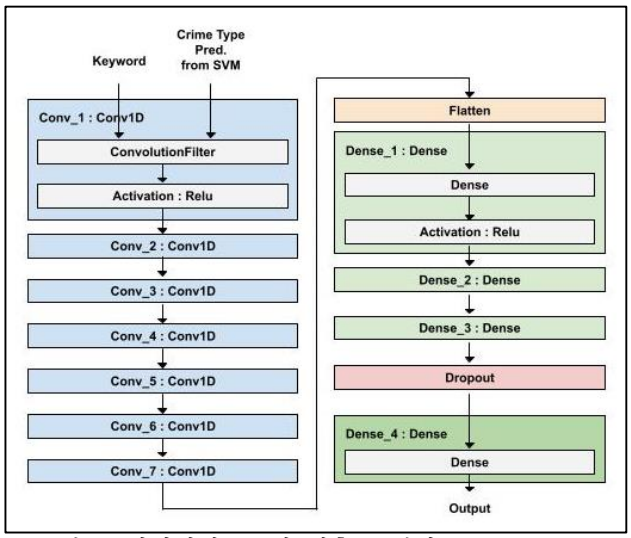


그림 2. 범죄위험스코어 예측 모델인 CNN Layer 구조

앞서 구현한 SVM 모델의 범죄유형 예측 결과를 Input Dataset 의 새로운 Feature 로 생성한다. 이때 예측 결과는 One-Hot Encoding 이 적용된 형태로 삽입된다.

Ⅲ. 모의실험 및 결과

3.1 데이터

범죄유형, 범죄 내용(텍스트)이 담긴 KICS 가상 범죄 데이터를 Reference data 로 이용한다. Training, Test dataset 은 범죄유형, 키워드 벡터, 범죄위험스코어로 이루어져있다. 범죄위험스코어는 이전 연구에서 설계된 범죄위험스코어 계산식을 통해 데이터가 생성된다[3].

3.2 실험 결과

표 1 과 표 2 는 이전 연구에서 구현한 DNN(Deep Neural Network), CNN 모델과 본 연구에서 구현한 모델의 성능을 비교하기 위한 표이다. 제 1 행의 1 ~ 21 은 절도, 체포감금, 손괴 등 21 개의 범죄 유형을 나타낸다. 표 1 에서 M1 은 Model1 이고, 표 2 에서 M2 는 Model2 를 뜻한다.

분류 모델인 Model1 은 성능을 평가하기 위한 지표로 Accuracy 를 사용한다. 표 1 에서와 같이 제안된 추정 모델이 대부분의 범죄 유형에 대해서 기존의 CNN/DNN 모델보다 우수한 성능을 달성할 수 있다. 샘플 전체의 평균 정확도 또한 Model1 이 0.94 로 우수하다.

회귀 모델인 Model2 는 MAPE(Mean Absolute Percentage Error)로 성능을 평가한다. MAPE 는 평균 절대 백분율 오차로 0 에 가까울수록 오류율이 낮다고 할 수 있다. 표 2 에서도 표 1 과 마찬가지로 대부분의 범죄 유형에 대하여 Model2 가 우수한 성능을 달성할 수 있으며, 전체 샘플의 평균 MAPE 또한 기존 DNN 과 CNN 대비 약 12 퍼센트 정도 우수한 성능을 제공하는 것이 가능하다.

Ⅳ. 결 론

본 연구에서는 기존 범죄유형과 범죄위험스코어의 예측 성능을 향상시킬 수 있는 새로운 모델을 제안하였다. 새롭게 제안된 모델은 기존 모델보다 우수한 성능을 제공할 수 있으며, 특히 범죄 유형을 예측하는 Model1 의 예측 결과를 범죄위험스코어를 추정하는 Model2 에 적용하여 효과적으로 Model2 의 성능을 향상시켰다. 이를 통해 범죄위험스코어를 예측 시 Model1 의 예측 값을 Model2 에 활용하는 것은 효과적이라는 것을 확인할 수 있다. 또한 이 기술의 연구 목적인 심각 범죄의 효과적 대응에 기여할 수 있다.

범죄위험스코어 예측 오류율이 범죄유형마다 격차가 있는 것을 확인할 수 있는데 향후 Dataset 이 충분히 늘어나면 지금보다 성능이 더 좋아질 것으로 기대된다.

ACKNOWLEDGMENT

이 논문은 2021 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2018-0-00440, 위험 상황 초기 인지를 위한 ICT 기반의 범죄 위험도 예측 및 대응 기술 개발)

참 고 문 헌

[1] 방준성, 박원주, 윤상연, 신지호, 이용태, “지능형 치안 서비스 기술 동향”, [ETRI] 전자통신동향분석, vol. 34, no. 1, pp. 111-122, 2019. 2.

[2] 이상윤, 염경록, 황재원, 박원주, 장광호, 이용태, “치안정보 분석을 위한 데이터뱅크 설계에 관한 연구”, 대한전자공학회 학술대회, pp. 697-698, 2019. 11.

[3] 백명선, 이용태, 장광호, 이희조, 박원주, “심각범죄의 효과적 대응을 위한 기계학습 기반의 범죄유형 및 범죄위험스코어 예측 기술 연구”, 한국통신학회 종합 학술 발표회 (하계) 2020, pp.512-513, 2020. 8.

[4] 김현중, 조성준, 강필성, “KR-WordRank; WordRank 를 개선한 비지도학습 기반 한국어 단어 추출 방법”, 대한산업공학회지, vol. 40, no. 1, pp. 18-33, 2014. 2.

[5] 박대서, 김화종, “TF-IFD 기반 키워드 추출에서의 의미적 요소 반영을 위한 결합벡터 제안,” 한국정보기술학회논문지, vol. 16, no. 2, pp. 1-16, 2018

저화질 미디어의 감정인식 문제 해결을 위한 고품질 변환 기술 현황과 성능분석

김나연*, 조숙희**, 배병준*, **
*과학기술연합대학원대학교, **한국전자통신연구원
{boboss, shee, 1080i}@etri.re.kr

Status and Performance of the High-Quality Conversion Technology for the Problem of Facial Expression Recognition in Low-Quality Media

Nayeon Kim, Sukhee Cho, Byungjun Bae
*Univ. of Sci. and Tech., **Electronics and Telecommunications Research Institute

요 약

본 논문에서는 지하철 CCTV 영상과 같은 저화질 미디어에서 사람의 표정으로 감정을 인식할 때 성능을 향상시킬 수 있는 딥러닝 기반 미디어 고품질 변환 기술에 대해 알아보고, 성능 평가 결과를 제공하여 범죄 예방에 활용할 수 있는 기초자료를 제공하는 것을 목표로 한다. 성능 평가는 공개된 감정인식 데이터셋을 일반적인 쌍선형 보간법으로 해상도를 8 배한 이미지와 GAN 모델을 기반으로 8 배한 이미지를 ResNet 으로 각각 학습하여 수행되었으며, 학습, 검증, 평가 실험 모두에서 GAN 기반 초해상도 이미지로 학습한 모델이 더 높은 감정인식 정확도가 나타난 것을 확인하였다.

Ⅰ. 서 론

표정은 감정을 인식하는 데 중요한 역할을 한다. 특히, 저화질 미디어에서의 표정인식은 범죄취약지역의 방범용 CCTV 나 지하철 등의 저화질 CCTV 처럼 얼굴 식별이 어려운 미디어에 적용하여 범죄 예방에 활용될 수 있다.

표정을 통한 감정 인식은 컴퓨터 비전 분야에서 오랫동안 연구되어 왔으나, 딥러닝의 발전 이후에는 딥러닝을 기반으로 표정을 인식하는 연구들이 활발하게 진행되고 있다. 이러한 이미지 작업의 성능은 이미지의 해상도에 큰 영향을 받기 때문에, 해상도가 낮고 저품질인 얼굴 이미지에서 감정인식 성능의 저하를 방지하기 위한 다양한 방법들이 함께 제시되었으며[1][2], 최근에는 새로운 데이터 생성에 사용되는 딥러닝 모델 GAN(Generative Adversarial Network)을 통해 이미지를 고품질화하여 감정인식 성능을 높이는 연구가 시도되고 있다[3].

이에 본 논문에서는 감정인식 성능을 높이는 데 활용할 수 있는 GAN 기반 고품질 변환 기술에 대해 파악하고, 기본적인 이미지 보간 방법인 쌍선형 보간법(ilinear)과 GAN 기반 얼굴 이미지 생성 모델[8] 사이의 감정인식 성능을 비교 및 분석하여 저화질 미디어에서의 감정인식 성능 향상을 위한 고품질 변환 기술 설계 및 개발의 기초자료로 제공하고 사회 문제 해결에 기여하고자 한다.

Ⅱ. GAN 기반 얼굴 이미지 고품질 변환 기술

얼굴 이미지 고품질 변환은 단일 이미지에서 임의의 장면 대신 얼굴 이미지만 고려하기 때문에 랜드마크나 히트맵과 같은 얼굴의 지역 특징들이 강력한 사전 지식으로써 활용되고 있다[4]-[7]. 이러한 사전 정보를 충분히 활용하기 위해 고품질 변환의 지역화 및 정렬 과정에서 왜곡없이 랜드마크 및 얼굴 구성 요소 맵을 정확하게 추정하기 위한 방법 또한 제안되었다[8].

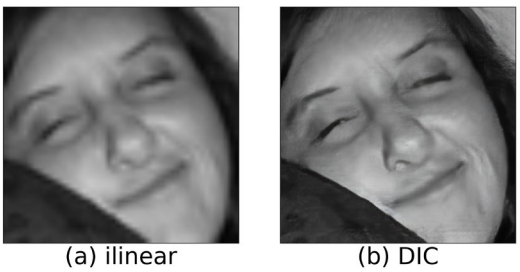

(a) ilinear (b) DIC

그림 1. FER 데이터셋(48×48)의 쌍선형 보간법과 DIC 적용 결과(384x384)

그림 1 은 감정인식 데이터셋 FER2013 에 쌍선형 보간법과 [8]에서 제안된 GAN 기반 초해상도 기술을 적용한 결과를 나타낸다.

Ⅲ. 고품질 변환 이미지의 감정인식 성능 평가

성능 평가는 감정인식 데이터셋인 FER2013 에서 잘못된 감정 라벨을 다시 지정한 FER plus 를 사용하여 수행되었다. FER 데이터셋에 쌍선형 보간법과 고품질 변환을 적용하여 ResNet[9]을 각각 학습시킨 후 감정인식 정확도를 평가했고, 고품질 변환에 사용한 얼굴 초해상도 기술로는 [8]에서 제안된 얼굴 이미지 복구

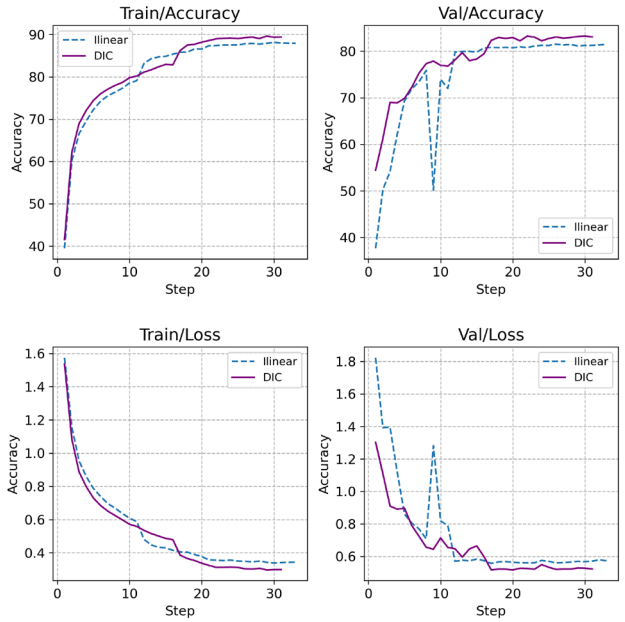


그림 2. 쌍선형 보간법과 DIC 를 적용한 감정인식 학습 및 검증 실험의 정확도와 Loss 그래프

네트워크와 정확한 랜드마크 추정 네트워크가 반복 협업하여 이미지를 생성하는 DIC (Deep Iterative Collaboration)를 사용하였다.

그림 2 성능 평가 결과 그래프로, 학습 및 검증 시 정확도 변화와 Loss 변화를 나타낸다. 쌍선형 보간법을 적용한 실험의 경우 학습 정확도 87.84%, 검증 정확도 81.52%, 평가 정확도 83.45%가 가장 높은 정확도로 나타났으며, DIC 실험은 학습 정확도 88.99%, 검증 정확도 83.24%, 평가 정확도 83.52%가 가장 높은 정확도로 나타났다. 세 가지 분리된 데이터셋 모두에서 DIC 를 적용한 실험이 더 높은 정확도를 보였다.

Ⅳ. 결론

본 논문에서는 저화질 미디어에서의 표정을 통한 감정인식 기술들을 살펴보고, 그 중 저화질 미디어에 초해상도 기술을 적용했을 때 감정인식 성능을 평가 및 비교하였다. 저화질 미디어의 감정인식 기술은 얼굴식별이 어려운 미디어에 적용하여 범죄취약지역의 범죄 혹은 디지털 성범죄 예방에 활용될 수 있다. 본 논문에서 수행한 평가 결과가 이러한 사회 문제 해결을 위한 기술 개발의 기초자료로 활용될 것을 기대한다. 그러나, 본 논문에서 사용된 FER2013 데이터셋은 구글 이미지 검색 API 를 사용하여 감정, 성별, 나이 등의 키워드로 검색된 결과 이미지들로 구성되어 있기 때문에 실제 환경에서 감정인식 성능을 보기 위해서는 조명, 각도, 가림 등 현실의 다양한 상황을 반영한 데이터셋을 이용하는 성능 평가 또한 필요하다.

ACKNOWLEDGMENT

“이 논문은 2021 년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No.2021-0-00087, SD/HD 급 저화질 미디어의 고품질 변환 기술 개발)”

참 고 문 헌

[1] CHENG, Bowen, et al. Robust emotion recognition from low quality and low bit rate video: A deep learning approach. In: 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII). IEEE, 2017. p. 65-70.

[2] H. Jung, S. Lee, J. Yim, S. Park and J. Kim, "Joint Fine-Tuning in Deep Neural Networks for Facial Expression Recognition," 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 2983-2991

[3] HU, Xiaobin, et al. Face super-resolution guided by 3d facial priors. In: European Conference on Computer Vision. Springer, Cham, 2020. p. 763-780.

[4] Yu, Xin, et al. "Super-resolving very low-resolution face images with supplementary attributes." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

[5] KIM, Deokyun, et al. Progressive face super-resolution via attention to facial landmark. arXiv preprint arXiv:1908.08239, 2019.

[6] Xin Yu, Basura Fernando, Bernard Ghanem, Fatih Porikli and Richard Hartley, "Face super-resolution guided by facial component heatmaps", ECCV, pp. 217-233, 2018.

[7] Yu Chen, Ying Tai, Xiaoming Liu, Chunhua Shen and Jian Yang, "Fsrnet: End-to-end learning face super-resolution with facial priors", CVPR, pp. 2492-2501, 2018.

[8] C. Ma, Z. Jiang, Y. Rao, J. Lu and J. Zhou, "Deep Face Super-Resolution With Iterative Collaboration Between Attentive Recovery and Landmark Estimation," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 5568-5577

[9] He, Kaiming, et al. "Identity mappings in deep residual networks." European conference on computer vision. Springer, Cham, 2016.

딥러닝 기반의 탑승자 머리 상해도 예측 알고리즘 개발

유시현, 이지은, 박성근
순천향대학교

qjhsj@sch.ac.kr, kgm5004@sch.ac.kr, keiny@sch.ac.kr

Development of Deep Learning-Based HIC Prediction Algorithm for Passengers

Si Hyeon Yu, Ji Eun Lee, Seongkeun Park
SoonChunHyang Univ.

요 약

본 논문에서는 자동차 충돌 시험 시 얻어지는 도면 이미지 데이터를 기반으로 딥러닝 기법인 CNN 을 활용하여 머리 상해 지수인 HIC(d) 예측하는 알고리즘을 제안한다. 자동차 안전성 평가를 위해 실차 충돌 시험 시, 많은 비용과 시간이 소모된다. 이에 본 논문에서는 실차 충돌 시험의 대체 방안으로서, 충돌 시의 더미의 단면 이미지 데이터를 기반으로 운전자의 HIC(d)를 예측하는 모델을 제안한다.

Ⅰ. 서 론

자동차 제조사들은 NCAP 기준에 따라 신차 출시 전 자동차 안전성을 평가하며 이중 하나가 충돌 시험이다. 충돌 시험 시 더미 머리에 가해진 충격량을 HIC(Head Injury Criterion)이라고 한다. 이 충돌 시험에는 많은 시간과 비용이 소모된다.

이에 본 논문에서는 실제 충돌 시험의 대체 방안으로서 딥러닝 기반의 HIC 예측 기법을 제안한다.

Ⅱ. 본론

본 논문에서는 충돌 시험 시 얻어지는 CATIA 도면 중 머리 더미를 Vertical 방향으로 자른 단면을 입력으로 사용하였다. 충돌 시 가속도 센서로부터 얻어지는 머리 더미의 속도 값(HIC)을 학습 Label 로 설정하였다. 데이터는 32 개 차종으로 총 706 장으로 구성된다.

차체 소재 및 Headform 의 CATIA 단면을 다음과 같은 색상으로 설정하여 소재를 구분한다.

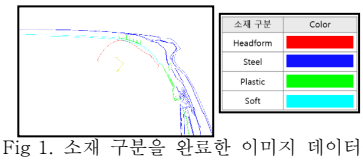


Fig 1. 소재 구분을 완료한 이미지 데이터(좌)와 소재에 따른 색상 구분(중)

도면 상에서 headform 과 roof 사이의 거리(distance 를 추출하여 Standard scaler 를 적용 후 모델의 두 번째 입력으로 사용하였다.

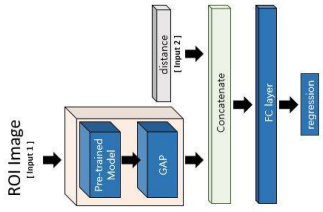


Fig 2. 2Input Model Architecture

ImageNet 데이터로 사전학습 된 DenseNet201 구조를 사용하였다.

모델의 전체구조는 위과 같다. (Fig 2)

Ⅲ. 결론

본논문에서는 HIC 예측을 위한 최적의 Model 을 제안한다.

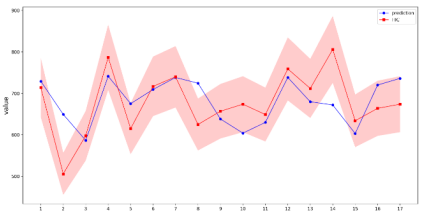


Fig 3. Prediction Graph

성능은 오차율로 판단하며, 오차율은 실제 HIC 와 예측한 HIC 의 차이를 백분율로 나타낸 것이다.

평균 오차율 7.5%의 성능을 도출하였다.

ACKNOWLEDGMENT

이 논문은 2021 년도 정부(과학기술정토봉신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No, 2019-0-01343, 융합보안핵심인재양성)

참 고 문 헌

[1] HUANG, Gao, et al. “Densely Connected Convolutional Networks”. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.p.4700~4708.

바이너리 영상과 DWA 퓨전을 통한 자율주행 시스템

황요섭, 이재오, 안종우, 이동혁
(주)씨랩
{mmx001, jaeoh, jwan, donghyuk}@cilab.kr

Autonomous driving system Using Binary image and DWA fusion

Hwng Yo-Seop, Lee Jae-Oh, An Jong-woo, Dong-Hyuk Lee
CILAB Pte.

요 약

최근, 센서 기술의 발전에 따라 라이다, GPS, 카메라 등 다중센서융합 연구가 활발히 진행되고 있다. 이처럼 센서 융합 기술의 발전은 자율주행차량의 상용화를 가속화하는 주요 요인으로 볼 수 있다. 하지만, 이와 같은 다중센서융합은 많은 데이터 연산이 필요하기 때문에 임베디드 시스템에서는 성능적 제한이 존재한다. 이러한 문제를 해결하기 위해서, 자율주행 개발업체들은 임베디드 시스템에 최적화 하기위한 연구를 진행하고 있다. 본 논문은 이러한 추세에 맞추어 자율주행차량에서 카메라의 영상정보만을 이용하여, 임베디드 시스템에서 주행이 가능한 알고리즘을 제안한다. 영상 정보를 바이너리로 변환하여 Dynamic Window Approach(DWA) 알고리즘에 적용, 주행 차선에 장애물이 없을 때는, 차선 중앙을 유지하는 기능과, 차선에 장애물이 있을 경우, 장애물과 차선 사이로 경로를 만들어 주행하게 하였다. 기존 경로 계획보다 연산량이 적고, 임베디드 시스템과 같이 제한된 상황에서 최적의 성능을 낼 수 있다.

키워드: Camera, DWA(Dynamic Window Approach), Autonomous Vehicles, Estimate, Obstacle

Ⅰ. 서 론

최근 자율주행이 가능한 로봇에 대한 관심이 증가하면서 다양한 로봇 제품들이 개발되고 있으며, 자율주행은 4차 산업 혁명의 핵심 키워드로써 자율주행 자동차, 배달 로봇, 무인 경비 로봇 등 적용분야가 광범위한 핵심 기술이다[1]. 특히 자율주행 자동차 산업의 발전으로 인하여, 첨단자동차가 궁극적으로 자율주행자동차로 발전하기 위해서는 레이더, 카메라, 라이다, 초음파와 같은 외부 환경을 탐지하는 센서들과 외부와 무선 통신하는 기능 등이 필수적으로 요구된다[2-3]. 경로 계획 및 모션 제어를 성공적으로 달성하려면 자율주행 로봇이 피해야 할 장애물을 감지하여 안전하게 목표에 도달할 수 있어야 한다. 자율주행차량에서는 일반적으로 센티미터 단위의 정확도를 가지는 정밀 지도를 기반으로 차선 정보 및 차량 주변 정보를 이용하여 경로를 생성하는데 이와 같은 경로 생성에는 전역 경로 계획과 지역 경로 계획이 필요하다[4]. 이에 자율주행자동차의 특성을 고려하여 전역 경로 계획을 추종하기 위한, 지역 경로를 생성하는 알고리즘도 연구되고 있다[5].

본 논문에서는, 소형으로 제작한 자율주행플랫폼에서, 카메라 영상만을 이용하여, 차선을 유지 및 장애물을 회피하기 위한 연구를 진행 하였으며, 본론에서 바이너리 기반의 DWA 알고리즘 적용한 실험 및 성능을 검증하였다. 마지막으로 결론에서 추후 활용 방안을 작성하였다.

Ⅱ. 본론

DWA는 로봇의 기구학 및 동역학 제약조건을 고려하고 있으며, 주변 환경의 장애물과의 충돌을 피하기 위해 예측 기능을 가지고 있다. 또한 자율주행 자동차에서 카메라를 통해서 받은 영상을 격자로 구성하여 주변 장애물을 회피하고, 자율주행 자동차의 주행 경로를 생성하기 때문에 실시간 연산에 유리한 알고리즘이다.

또한 DWA에서의 로봇의 위치를 기준으로, 목표지점에 대한 이동 방향, 주행 속도 및 충돌 관련된 함수들로 구성이 되어 있으며, 관련 수식은 다음과 같다.

$$O = \gamma_{head}\omega_{head} + \gamma_{speed}\omega_{speed} + \gamma_{obstacle}\omega_{obstacle} \quad (1)$$

위 수식은 DWA의 목적함수를 구하기 위함이며, ω_{head} 는 로봇의 위치를 기준으로 목표지점에 대한 이동방향을 수치화한 함수를 의미하고, ω_{speed}는 목표지점에 대한 로봇 주행 속도를 나타내는 함수를 의미하고, $\omega_{obstacle}$는 로봇의 주행 방향이다.

현재 주행 속도와 로봇이 낼 수 있는 최고 주행 속도에서, 정지 시간까지의 소요 시간을 고려하여 장애물과의 충돌에 걸리는 시간을 예측 가능함을 나타내는 함수이다. $\gamma_{head}, \gamma_{speed}, \gamma_{obstacle}$ 는 각 함수의 속도변화에 따른 가중치를 의미한다.



Fig.1 자율주행플랫폼(CiLab X1)

Fig.1은, 본 논문의 실험을 위해 사용된 자율주행플랫폼 CiLab X1이다. 소형 카메라와 라이다로 구성이 되어 있으며, 본 논문에서는 카메라 모듈을 통해서 받은 영상만을 이용하였다. 카메라 영상을 통해 받은 영상 정보를 기반으로, 바이너리 DWA알고리즘을 통해 장애물 회피 주행 실험을 진행하였다. 장애물에 대한 정보는 자율주행플랫폼 CiLab X1의 전방에 부착된 카메라를 통해서 이미지 영상을 획득하였다.



Fig.2. 실험 환경

주행 실험은 위의 그림과 같이 회전 굴곡이 있는 환경으로 구성하였으며, 크기는 길이 3 m이며, 폭은 1.8 m크기 범위로 제안하였다. 양쪽의 흰색 차선을 두어 차선 유도가 가능한 구성으로 하였다.

자율주행플랫폼인, CiLab X1에 부착되어 있는 카메라를 통해서 받은 이미지 정보에서 주행 경로에 보이는 좌/우 흰색 차선에 대한 데이터를 식 (1)에 있는 $\omega_{obstacle}$ 에 적용 하였으며, 이 함수의 값의 변화에 따라서 경로를 유지하면서 주행이 가능하게 된다. 또한, 좌우에 있는 흰색 차선 값을 장애물로 설정하여, DAW알고리즘에 적용 하였으며, 자율주행플랫폼이 주행할 때, 경로를 유지하면서 주행하게 하였다.

Fig.3은 자율주행플랫폼인, CiLab X1이 실험 환경을 주행하던 중에 경로 이탈 방지를 위해서 적용한 바이너리DWA 알고리즘을 적용하는 과정에 대해서 나타내었다.

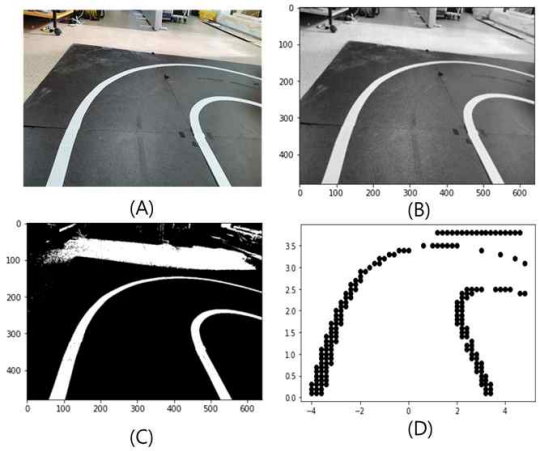


FIg.3. 차선(장애물) 검출

먼저, Fig.3(A)는 자율주행 차량에서 카메라를 통해서 받은 영상 데이터이다. Fig.3.(B)는 Fig.3(A)에서 받은 영상을, Gray scale로 변환하였다. 이는, 양쪽 차선을 인식하기 위해서, 다른 컬러 정보를 없애고, 차선의 특징점을 검출하기 위함이다. Fig3.(c)에서는, Gray scale 이미지에서 양쪽의 흰색 차선의 특징 점 및 Pixel 정보를 제거하였다.

```
[[0 0 0 ... 0 0 0]     array([[ 1.2,  3.8],
 [0 0 0 ... 0 0 0]            [ 1.4,  3.8],
 [0 0 0 ... 0 0 0]            [ 1.6,  3.8],
 ...                          [ 1.8,  3.8],
 [0 0 0 ... 0 0 0]            [ 2. ,  3.8],
 [0 0 0 ... 0 0 0]            [ 2.2,  3.8],
 [0 0 0 ... 0 0 0]]           [ 2.4,  3.8],
```

Fig.4. 차선(장애물) Binary 정보 추출

Fig.4 같이 Fig.3(C)에서 받은 영상 정보를 Pixel로 변환하고, Matrix 형태로 만들게 된다. 이렇게 만든 Matrix에 대해서, 바이너리로 변환하고, 이를 각각의 좌표 값을 설정하였다. 그렇게 되면, Fig.3.(D)와 같이 흰색 차선에 도트 정보만을 나타내게 된다. 이를 이용하여 임베디드 시스템에서 빠르게 연산이 가능하고, 연산한 정보를 DWA알고리즘을 통해서 경로 계획을 빠르게 추정이 가능하게 된다.

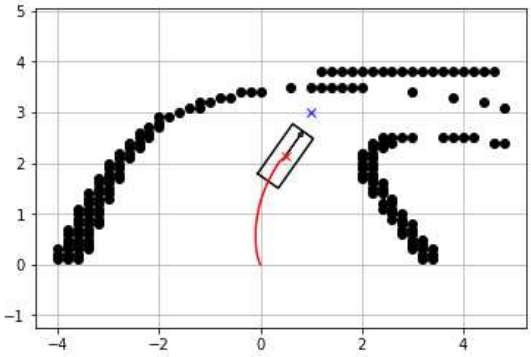


Fig.5. DAW 기법을 적용한 주행 실험 결과

위 그림은, 앞에서 설명한 Fig.3.을 통해 설정한 양쪽 흰색 차선을 장애물로 설정하고, 이를 통해서 목표 지점까지의 경로를 타겟을 성정한 모습을 시각화한 결과이다. 위 그림에서 검은색 점은 차선을 의미하며, 붉은색 선은 자율주행 플랫폼인 CiLab X1 주행 경로를 나타낸다.

위 실험 결과에서 알 수 있듯이 바이너리 DWA 기법을 적용하여 자율주행 플랫폼인 CiLab X1이 차선을 이탈하지 않고, 안정적으로 주행이 가능하게 하였다.

Ⅲ. 결론

본 연구는 라이다, 레이더, 카메라와 같이 복합적인 정보를 사용하지 않고, 카메라 영상만을 이용하여, 차선 인식 및 경로 회피를 하였다. 또한 임베디드 시스템에 최적하였으며, 소형 자율주행 차량에 적용이 가능함을 검증 하였다. 또한 DWA의 폭넓은 활용 방안을 제시 하였으며, 차선 기반으로 주행 및 차선 중앙 유지 하며, 주행이 가능하게 하였다. 기존 경로 계획보다 연산량이 적으며, 라이다와 같이 다른 센서들의 결합 없이 주행이 가능함을 검증하였다. 추후 장애물 회피 및 빠른 속도로 주행을 하면서 궤적을 추정하는 연구를 진행할 계획이다.

참 고 문 헌

[1] 윤태진, et al. "SLAM 을 이용한 자율주행 순찰 로봇 개발." 한국컴퓨터정보학회 학술발표논문집 27.2 (2019): 437-438.

[2] 홍인기, 이훈석, 심영대, 2015, “C-ITS와 스마트카 기술 동향 및 전망”, 정보통신기술진흥센터, 주간기술동향.

[3] 국토해양부, 국토교통과학기술진흥원, 2018, “첨단자동차 안전성 평가 기술개발 최종보고서, R&D 16PTSI1C054118-08", pp. 27~5

[4] Fox, Dieter, Wolfram Burgard, and Sebastian Thrun. "The dynamic window approach to collision avoidance." IEEE Robotics & Automation Magazine 4.1 (1997): 23-33.

[5] Jung, Euigon, Wonho Song, and Hyun Myung. "동적 및 정적 물체 회피를 위한 정밀 도로지도 기반지역 경로 계획." Journal of Korea Robotics Society 16.2 (2021): 112-121.

교통소외지역 사회문제 해결을 위한 자율주행 기반 서비스 어플리케이션에 관한 연구

조봉균, 이윤화, 김태형, 윤경수

지능형자동차부품진흥원

jbggg1@kiapi.or.kr, yhlee@kiapi.or.kr, thkim@kiapi.or.kr, kadbonow@kiapi.or.kr

A Study on Service Application of the Autonomous Driving for Solving Social Problems in the Transportation Marginalized Community

Jo Bong-Gyun, Lee Yun-Hwa, Kim Tae-Hyeong, Yun Kyung-Su

Korea Intelligent Automotive Parts Promotion Institute (KIAPI)

요 약

최근 다양한 모비리티 관련 기술에 대한 연구가 활발하게 이루어지고 있으며, 이를 기반한 서비스에 대한 관심도 높아지고 있다. 이러한 국내 대부분의 자율주행 기술 및 관련 서비스들은 공도를 기반으로 개발되고 있으며, 자율주행 차량을 호출하기 위한 휴대폰 어플리케이션도 공도 위주로 개발되고 있다. 그러나 자율주행 호출 서비스가 가장 필요한 곳은 시골 등과 같은 교통소외지역이며, 호출을 위한 휴대폰 어플리케이션 또한 노인들이 쉽게 사용할 수 있도록 개발되어야 한다. 본 논문에서는 이러한 교통소외지역의 노인들도 쉽게 사용할 수 있는 자율주행 호출 어플리케이션에 대하여 연구를 진행하고 향후 서비스를 위한 나아가야할 방향도 제시하고자 한다.

Ⅰ. 서 론

자율주행 기술은 다양한 IT 기술들과 기존 자동차 관련 기술이 결합되어 관련 산업 생태계를 파괴적으로 변화시키고 있다. 이러한 자율주행 기술 선점을 위해 기존 자동차 회사뿐만 아니라 대형 IT 기업들도 자율주행 기술 개발에 매진하고 있다. 국내에서는 이미 자율주행 자동차를 활용한 다양한 시범서비스가 진행되고 있으며, 이를 상업적으로 활용하기 위한 연구도 활발하게 진행 중이다.[1] 그러나 국내 자율주행 기술 및 관련 서비스들은 일반 공도 위주로 개발되고 있으며, 이러한 자율주행 기술 및 기반 서비스들이 절실히 필요한 시골 등과 같은 교통소외지역에서는 관련 기술 개발이 아직 미비한 실정이다. 또한 교통소외지역 인구의 대부분을 차지하는 노인들은 스마트폰 기반의 어플리케이션을 활용하기 어려우며, 이를 위한 호출 어플리케이션 개발 또한 상업적인 이유로 어려움을 겪고 있다. 본 논문에서는 이러한 교통소외지역의 노인들을 위한 자율주행 차량 호출 어플리케이션 개발 방향 및 시나리오에 대하여 기술하고, 실제 개발현황에 대하여 논의하고자 한다.

Ⅱ. 본론

다양한 정보를 작은 화면에 보여주며, 작은 어플리케이션 아이콘들이 배치되어 있는 스마트폰은 노인들이 활용하기에는 많은 어려움이 있다. 이러한 문제를 해결하기 위해서는 스마트폰에서 보여주는 정보들을 최대한 단순 및 간단하게 만들어야하며, 한 번의 손동작으로 어플리케이션 작동 절차가 단순 명료하게 순차적으로 눈에 익혀져야 한다. 이를 위해 교통소외지역에서 자율주행 차량을 호출하기 위한 어플리케이션을 최대한 단순 명료하게 만들었으며, 이를 지원하기 위해 태블릿을 활용한 어플리케이션을 개발하고 있다. 또한 자율주행차량 호출을 위해서 다음 그림과 같은 시나리오를 개발하였다.



그림 1. 자율주행 차량 호출 시나리오

Ⅲ. 결론

향후 개발 완료된 어플리케이션을 통해 자율주행 차량을 실제 호출하는 시범 서비스를 추진하고, 그 성능을 검증할 계획이다.

ACKNOWLEDGMENT

This work was supported by Institute of Information & communications Techology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT) (No. 2019-0-00399, Development of A.I. based recognition, judgement and control solution for autonomous vehicle corresponding to atypical driving environment)

참 고 문 헌

[1] 이재형, 김진희, 이설영(2021). “국내외 자율주행 대중교통 서비스 실증 사례 분석”. 교통기술과정책, 18(3),59-67.

머신러닝 기반 전기차 배터리 제어 관련 최신 기술 동향

정선호, 윤윤기, 강병도
지능형자동차부품진흥원
id4touch@kiapi.or.kr, ykyoon@kiapi.or.kr, headlamp@kiapi.or.kr

Recent Technical Trends in Machine Learning Based Control of EV Battery

Jung Sun-ho, Yun Yun-Ki, Kang Byung-Do
Korea Intelligent Automotive Parts Promotion Institute (KIAPI)

요 약

전기자동차는 최근 보급이 확대되고 있지만 빈번한 화재사고로 인해 안전 이슈가 대두되고 있고, 소비자는 배터리 충전 시간 단축 등 사용자 편의성을 요구하고 있다. 이러한 이슈들을 해결하기 위해 머신러닝 기반의 전기차 배터리 제어 기술이 주목을 끌고 있다. 본고에서는 머신러닝 기반으로 전기차 배터리의 수명을 조기에 예측하여 급속 충전 프로토콜을 최적화하는 기술과 전기차 배터리의 고장을 진단하는 기술에 대해 살펴본다.

Ⅰ. 서론

전기차 화재 사고로 인해 글로벌 전기차 제조사 중 GM 쉐보레 볼트가 6만 8천여대, BMW가 플러그인하이브리드 차량 2만 6900여대, 현대 코나 EV 7만7천여대 등 리콜이 발생하는 등 소비자의 불안감이 고조되고 있다. 국내에서는 '17~'20년 11월 기간 중 소방청에서 집계한 69건의 전기차 화재 사고로 인해 총 7.6억원 상당의 재산피해액이 발생했다. 국립과학수사연구원에 따르면 이러한 전기차 화재 사고의 절반 이상이 차량 하부에 설치된 배터리팩 어셈블리 내부의 전기적 요인에 의한 것으로 추정된다.

거듭되는 전기차 화재 사고 소식을 접하게 되면서, 배터리의 하드웨어 및 배터리 관리 시스템(BMS)에 대한 안전 기준의 중요성이 대두되고 있다. 유럽에서는 WP29의 REESS(Rechargeable Electric Energy Storage System)와 ELSA(Electric Safety) 전문가 입법회의를 통하여 유럽의 전기자동차 안전기준인 'ECE R100'이 제정되었다. 미국에서는 전기차 안전성 관련 세계기술기준인 GTR No. 20의 제정에 미국도로교통안전국(NHTSA)이 총괄로 참여하는 등 행보를 보이고 있다. 국내에서는 '21년 7월부터 개정된 국내 전기차 배터리 안전 기준인 '구동축전지 안전성 시험 기준'에서 전기차 배터리 안전성 시험을 기존 7항목에서 12항목으로 확대시켜 배터리의 하드웨어 뿐 아니라 BMS의 진단 기능에 대해서도 안전성 기준이 강화되고 있다.

전기차 배터리 관련 안전 기준 강화와 더불어 배터리의 안전성을 확보할 수 있도록 급속 충전 시 충전프로토콜 제어나, 배터리 고장 진단 등 머신러닝 기술을 필요로 하는 분야가 늘어가고 있다. 본고에서는 관련 연구에 대해 알아보기로 한다.

Ⅱ. 본론

1. 머신러닝을 이용한 전기차 배터리 급속 충전 프로토콜 최적화

전기차 충전시간 단축 수요로 인해 급속충전기 보급이 확대되고 있다. 급속 충전은 소비자의 편의를 향상시켜 주지만, 발열로 인해 배터리 내부가 손상을 입을 가능성이 있고 이는 배터리 사용 수명 단축을 초래할 수 있다. 따라서 배터리의 내부 열화를 방지하는 범위 내에서 최단 시간에 충전할 수 있도록 최적화된 충전 프로토콜의 설계가 필요하다고 볼 수 있다. 최근에는 머신러닝을 이용한 전기차 배터리 충전 프로토콜의 최적화가 시도되고 있다.

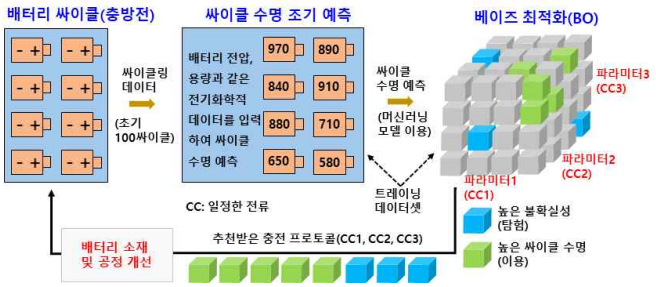


그림 1 클로즈 루프 최적화(CLO) 시스템 개요

연구팀은 배터리 수명을 최대화하기 위한 6단계의 충전 프로토콜 전류와 전압 프로필을 지정하여 파라미터 공간을 효율적으로 최적화하는 머신러닝 방법론을 제안했다[1]. 이 연구에서는 배터리의 수명을 최대화하고 10분 이내에 EV 배터리를 80%까지 충전하기 위한 최적의 프로토콜에 대해 머신러닝을 이용하여 찾는 것을 목표로 하고 있다. 충전 프로토콜 최적화를 위해 사용하는 클로즈 루프 최적화(CLO) 시스템은 그림 1과 같이 진행된다. 초기 100사이클의 배터리 전압, 용량과 같은 데이터는 사이클 수명의 조기 예측을 위한 입력치로 사용된다. 머신러닝 모델의 이러한 사이클 수명 예측은 이후에 베이즈 최적화(Bayesian Optimization, BO) 알고리즘으로 전달된다. 이는 예측된 사이클 수명이 높은 불확실성을 갖는 프로토콜을 테스트하는 '탐험(exploration)'과 높은 사이클 수명이 예측된 프로토콜을 테스트하는 '이용(exploitation)' 사이의 경쟁에서 균형을 맞추며

테스트할 다음 프로토콜을 추천한다. 이 과정은 테스트할 충전 프로토콜이 소진될 때까지 반복된다. 이러한 싸이클 수명의 조기 예측은 배터리 수명 예측을 위해 필요한 장시간의 테스트 사이클 수를 줄여 주어 효율을 개선시킬 수 있다. 고장나는 시점까지 충방전을 거친 배터리의 작은 트레이닝 데이터셋은 싸이클 수명 조기 예측 변수를 트레이닝하고 BO 하이퍼파라미터(학습 알고리즘의 샘플에 대한 일반화를 위해 조절되는 학습률, 에포크 수, 은닉층 수 등의 파라미터)를 설정하는 데 모두 사용된다. 향후 배터리 소재 및 공정 설계 또한 이 시스템을 적용할 수 있을 것으로 연구팀은 기대하고 있다.

기존 배터리의 충방전을 반복하여 배터리의 수명을 평가하는 방식은 시간과 비용이 소요되어 단시간에 합리적으로 배터리 수명을 예측하는 다양한 방법들이 제안되고 있다. 이런 수명 예측 방법들은 실제 충방전 환경을 적절히 반영하고 있는지, 그리고 실제 충방전 수명이 어느 정도 재현 가능한지를 명시하여야 하며, 재현성을 높여가기 위해 오차 보정 등의 노력이 필요할 것으로 생각된다.

2. 머신러닝 기반 전기차 배터리 고장 진단

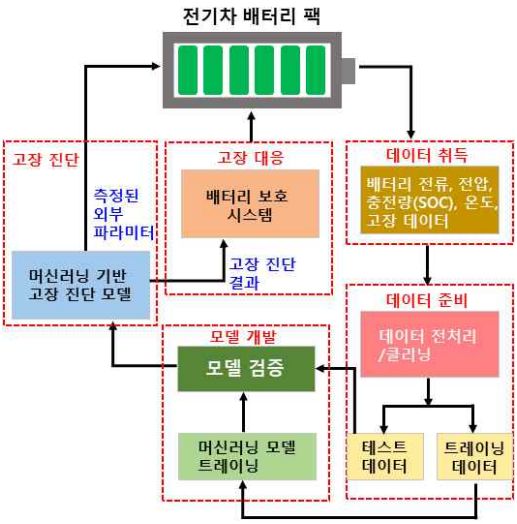


그림 2 머신러닝 기반 배터리 고장 진단

현재 전기차에 배터리 시스템에 탑재된 BMS는 일반적으로 스탠드 얼론으로 동작하여 일반 PC에 못미치는 처리 능력으로 배터리 감시를 수행하고 있다. 또한 배터리 충전량(SOC)은 배터리의 용량에 따라 달라지는데, 용량은 온도, 전류 크기 등의 환경에 따라 달라질 수 밖에 없음에도 BMS는 일반적으로 고정값의 용량을 입력하여 계산한다. 이러한 배터리 진단 기능에 관한 문제점를 극복하기 위해 머신러닝 기반의 배터리 고장 진단이 하나의 해결 방안으로 제시되고 있다[2].

그림 2에 머신러닝 기반 전기차 배터리 고장 진단 개요를 나타낸다. 전기차 배터리 고장 진단을 위해, 배터리 셀 전압, 전류 및 온도를 센서 및 데이터 수집 시스템을 이용하여 기록한다. 배터리 소모 패턴을 분석하고 실험을 통해 데이터를 생성한다. 배터리 고장 데이터는 배터리 작동 데이터와 실험실에서의 실험을 통해 수집되고, 데이터 클리닝 및 feature extraction*을 위해 원시 데이터의 전처리가 진행된다. 전처리를 거친 데이터는 트레이닝 데이터셋과 테스트 데이터셋으로 랜덤 분류된다. 머신러닝 기반 고장진단 전략을 수립하는 동안, 트레이닝 데이터셋은 알고리즘 러닝을 위해 사용되고, 테스트 데이터셋은 검증용으로 사용된다. 작동 전류, 단자 전압, 온도 등과 같은 배터리 관련 파라미터들은 검증된 머신러닝 기반 고장진단 방식을 사용하여 배터리의 고장을 검출하기 위해 사용된다. 고장 검출 신호는 배터리 보호 시스템에 대한 명령으로 사용된다.

이러한 고도화된 배터리 고장 진단은 스탠드 얼론으로는 현실적으로 어렵기 때문에, 클라우드 기반으로 배터리 상태 변화 모니터링 및 이상 신호 감지 시 충방전 차단 등을 실시간 대응할 수 있어야 전기차의 안전을 확보할 수 있을 것이다. 따라서, 제작 비용이나, 통신 지연 문제 등 클라우드 기반 BMS의 다양한 부가적인 이슈에도 관심을 가져야 할 것이다.

*feature extraction: feature는 학습 및 예측할 데이터의 특징, 항목으로, 여기서는 배터리의 전압, 전류, 온도 등의 데이터가 이에 해당되며, feature extraction은 feature가 유용한지 여부를 확인하는 과정

Ⅲ. 결론

전기차 배터리는 생산자의 생산 단가 인하나 성능 개선에 대한 압박을 꾸준히 받고 있어 배터리 소재 개발에 대한 지속적인 수요가 있을 것이며, 출시되는 배터리의 종류는 점차 다양해질 것으로 예측된다. 따라서 다양한 배터리 시스템에 대해 단 시간에 충전 프로토콜 최적화할 수 있는 머신러닝 기술 확보를 위해 배터리의 전기화학적 특성과 같은 트레이닝 데이터의 축적이 중요하다고 할 수 있겠다.

전기차 배터리의 고장 진단의 몇 가지 일반적인 과제가 있다. 예를 들면, 고장 임계값 정의 기준, 고장 시뮬레이션 툴 부재, 조기 고장 감지의 한계가 있겠다. 배터리 셀 노화, 배터리 셀 불일치(용량), 등에 대응하는 고장 진단 방법의 개발도 매우 중요한 과제라고 할 수 있겠다. 또한, 단일 고장 진단 방법의 내재적 한계로 인해, 정확성 향상을 목표로 하는 복합적인 고장 특성과 진단 방법을 결합하는 것이 유망한 추세이다.

전기차 배터리에 대한 머신러닝 기반 수명 예측이나 고장 진단이 아직 초기 단계로 해석되지만 갈수록 머신러닝 이론이 풍부해지고 배터리 데이터의 가용성이 높아져 고장 진단에 대한 신뢰는 높아져 갈 것으로 기대된다. 향후 HIL(Hardware-in-the-loop) simulation 등을 이용하여 배터리 관리시스템의 배터리 상태 추정 기능과 알고리즘의 계산 부하 등 적용 가능 여부를 검증해 나갈 예정이다.

ACKNOWLEDGMENT

본 연구는 국토교통부 국가연구개발사업의 연구비 지원을 받아 수행되었습니다.(과제번호: 20SDPT-C158092-01)

참 고 문 헌

[1] Attia, P. M., et al. "Closed-loop optimization of fast-charging protocols for batteries with machine learing," Nature '578, pp. 397-402, Feb. 2020.

[2] Samanta, A., et al. "Machine Learning-Based Data-Driven Fault Detection/Dignosis of Lithium-Ion battery: A Critical Review," Electronics '10, 1309, May 2021.

제2회 한국 인공지능 학술대회

The 2nd Korea Artificial Intelligence Conference

세션 A

물리적 정보 기반 강화학습을 이용한 로봇 팔 제어 연구

김성현, 장인국, 김현석, 박찬원, 박준희
한국전자통신연구원

{kim-sh, ingook, hertzkim, cwp, juni}@etri.re.kr

A Study on the Robot Arm Control with Reinforcement Learning based on Physical Information

Seonghyun Kim, Ingook Jang, Hyunseok Kim, Chan-Won Park, Jun Hee Park
Electronics and Telecommunications Research Institute

요 약

본 논문은 물리적 정보에 기반한 강화학습을 이용한 로봇 팔 제어 연구를 제시한다. 로봇 팔 제어 문제는 주어진 환경에서 특정 작업을 달성하는 것을 목표로 정의된다. 예를 들어, 로봇 팔을 이용하여 환경 내 놓여진 객체를 집거나, 밀거나 하는 형태로 객체를 특정 위치로 옮기는 작업을 정의할 수 있다. 로봇 팔 제어 문제는 관측과 행동을 통해서 상호작용이 이루어지는데, 이 때 관측 정보와 행동 정보가 연속된 값을 갖기 때문에 강화학습을 통해 로봇 팔 제어 문제를 해결하는 것은 매우 도전적인 문제이다. 본 논문에서는 로봇 환경을 구성하는 객체들에 대한 물리적 정보를 활용하여 효과적으로 로봇을 제어하는 방식을 제안한다.

Ⅰ. 서 론

본 논문에서는 물리적 정보에 기반한 강화학습을 이용한 로봇 팔 제어를 연구한다. 현재까지 강화학습은 심층학습 네트워크와 결합하여 불연속 관측 정보 및 행동 정보를 갖는 게임 환경에서 사람 수준의 성능을 달성하는 놀라운 성과를 이루었다 [1]. 이와 같은 심층학습과 결합한 강화학습, 즉 심층강화학습의 적용 분야는 게임 문제를 넘어서 실제 세계 문제를 풀고자 하는 분야로 확장되고 있다. 실제 세계의 문제 중 하나인 로봇 팔 제어 문제는 제조 및 물류 관리 등 다양한 분야에서 활용 가능한 문제이다. 로봇 팔 제어 문제는 기존 게임과 달리 연속된 관측 정보와 행동 정보를 다루기 때문에 난이도가 높은 문제에 속하고, 이를 해결하기 위한 많은 연구들이 진행되고 있다 [2]-[5]. 본 논문에서는 로봇 환경을 구성하는 객체들에 대한 물리적 정보를 활용하여 효과적으로 로봇을 제어하는 방식을 제안한다.

Ⅱ. 본론

그림 1 은 본 연구에서 고려하는 로봇 팔 작업 환경을 나타낸다. 그림 1 과 같이 로봇 팔 작업 환경은 6 축의 로봇 팔, 목적 객체(큐브), 센서 객체(원형) 등으로 구성되어 있다. 해당 로봇 팔 작업 환경에서 풀고자 하는 작업은 로봇 팔이 목적 객체를 들어서 센서 객체 위치로 옮기는 것으로 정의한다.



그림 1. 로봇 팔 작업 환경

1. 마르코프 결정 프로세스

로봇 제어 문제는 마르코프 결정 프로세스로 정의된다. 마르코프 결정 프로세스는 (상태, 행동, 전이확률, 보상)의 튜플 형태로 표현된다. 일반적으로 강화학습의 목적함수는 다음과 같이 감쇄보상의 합계에 대한 기대값으로 정의된다.

$$\boldsymbol{E}\left[\sum_{t=0}^{T}\gamma\cdot r(s_t,a_t)\right]$$

여기서, γ 는 감쇄 정도, T 는 전체 작업 시간, r 은 보상, s_t 는 상태, a_t 는 행동을 나타낸다. 상태는 로봇의 관절들의 회전각위치 및 끝단의 그리퍼에 대한 3 차원 위치정보, 목적 객체 및 센서 객체의 3 차원 위치정보를 포함한다. 행동은 관절들의 회전각속도와 그리퍼의 열림-닫힘의 이진신호를 포함한다. 보상은 그리퍼와 목적 객체 사이의 거리, 그리퍼가 목적 객체를 잡았는지 여부를 판단하는 이진값, 목적 객체와 센서 객체 사이의

거리, 그리고 목적 객체가 센서 객체에 위치하였을 때 추가적인 보상 등의 합으로 정의된다.

2. 확률적 정책 학습

로봇 제어를 위한 정책은 $\pi(a_t|s_t)$ 로 표현되고, 이는 상태가 주어졌을 때 행동의 확률분포로 정의된다. 로봇 제어와 같은 연속된 관측 및 행동에 대한 학습은 수많은 시행착오를 통한 경험으로부터 수행된다. 즉, 다양한 경험을 획득하는 탐색의 과정이 효과적으로 이루어져야 한다. 목적함수를 최적화하는 학습단계에서 탐색을 보다 효과적으로 하기위해, 정보이론의 엔트로피 개념을 목적함수와 결합할 수 있다 [6]. 즉, 목적함수에 추가적으로 정책에 대한 엔트로피 $H(\pi(a_t|s_t))$를 더하여, 행동의 확률분포를 넓게 가져가도록 최적화를 수행한다.

3. 물리적 정보 기반 확률적 정책 학습

상태 정보는 로봇 팔, 목적 객체, 센서 객체에 대한 각각의 고유 정보만을 포함하고 있다. 작업에서 요구하는 조건을 효율적으로 달성하기 위해서는 보다 직관적인 형태의 정보를 활용할 수 있다. 즉, 개별 객체에 대한 정보뿐만 아니라, 객체 사이의 관계를 정의하는 물리적 정보를 도출 활용함으로써, 보다 효율적으로 정책을 학습할 수 있다. 이와 같이 물리적 정보를 활용한 정책은 $\pi(a_t|s_t, z_t)$ 으로 표현되고, 여기서 z_t 는 물리적 정보를 나타낸다. 물리적 정보는 로봇 팔 작업 환경 내 객체 사이의 거리, 각도 등으로 정의할 수 있다.

물리적 환경을 활용하는 방식으로는 모델 기반 강화학습이 있다. 모델 기반 강화학습과 제안 방식을 비교하였을 때, 모델 기반 강화학습은 행동에 따른 상태전이 확률 분포에 대한 학습이 요구되는 반면, 제안하는 방식은 사람이 설계 가능한 물리적 정보를 직접 활용하기 때문에 학습에 대한 비용이 적고, 직접적인 해석이 가능하다는 장점이 있다.

Ⅲ. 실험 결과

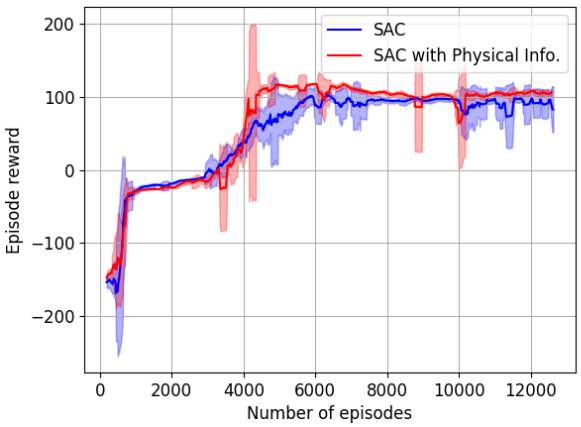


그림 2. 에피소드 보상 비교

제안하는 물리적 정보 기반 확률적 정책 학습 기법은 기존 확률적 정책 학습 중 하나인 Soft Actor-Critic(SAC)을 기반으로 물리적 정보를 결합하여 구현하였다. 그림 2 는 학습 진행에 따른 에피소드 보상에 대한 비교를 나타낸다. 그림 2 에서 확인할 수 있듯이, 물리적 정보를 결합함으로써 기존 SAC 기법보다 더 빠르게 학습이 수렴하는 것을 알 수 있다. 즉, 개별 객체에 대한 정보만 이용하는 것보다, 객체 사이의 관계를 입력으로 추가하였을 때 더 좋은 성능을 도출할 수 있다.

Ⅳ. 결론

본 논문에서는 로봇 환경을 구성하는 객체들에 대한 물리적 정보를 활용하여 효과적으로 로봇을 제어하는 방식을 제안한다. 기존 연구에서는 개별 객체에 대한 정보만을 활용하는 반면에, 본 논문에서는 객체 사이의 관계를 정의하는 물리적 정보를 함께 활용하는 방법을 제안하였다. 실험결과를 통해, 개별 객체에 대한 정보뿐만 아니라, 객체 사이의 관계를 정의하는 물리적 정보를 함께 활용함으로써 보다 효율적으로 확률적 정책을 학습할 수 있음을 확인할 수 있다.

ACKNOWLEDGMENT

본 연구 논문은 한국전자통신연구원 연구운영지원사업의 일환으로 수행되었음. [21ZR1100, 자율적으로 연결· 제어· 진화하는 초연결 지능화 기술 연구]

참 고 문 헌

[1] P. Henderson, R. Islam, P. Bachman, J. Pineau, D. Precup, and D. Meger. Mastering the game of go with deep neural networks and tree search. Nature, 529:484-489, Jan. 2016.

[2] Y.-J. Han, I.-S. Kim and Y.-D. Hong, “Optimization-based humanoid robot navigation using monocular camera within indoor environment,” ETRI Journal, vol. 40, no. 4, pp. 446-457, Aug., 2018.

[3] M. Zhang, J. Chen, X. Wei, and D. Zhang, “Work chain-based inverse kinematics of robot to imitate human motion with Kinect,” ETRI Journal, vol. 40, no. 4, pp. 511-521, Aug., 2018.

[4] M. Ilbeygi and M. R. Kangavari, “Comprehensive architecture for intelligent adaptive interface in the field of single-human multiple-robot interaction,” ETRI Journal, vol. 40, no. 4, pp. 483-498, Aug., 2018.

[5] L. Fan, Y. Zhu, J. Zhu, Z. Liu, O. Zeng, A. Gupta, J. Creus-Costa, S. Savarese, and L. Fei-Fei. Surreal: Open-source reinforcement learning framework and robot manipulation benchmark. Proceedings of The 2nd Conference on Robot Learning, 87:767-782, Oct. 2018.

[6] T.Haarnoja, A.Zhou, P.Abbeel, and S.Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. Proceedings of the 35th International Conference on Machine Learning, 80:1861-1870, Jul. 2018.

심층강화학습을 이용한 Cellular-V2X 요구사항 적응형 혼잡제어 알고리즘

양우열, 조한신
한밭대학교 전자공학과
30211173@edu.hanbat.ac.kr, hsjo@hanbat.ac.kr

Cellular-V2X Requirements Adaptive Congestion Control Algorithm Using Deep Reinforcement Learning

Wooyeol Yang, Han-Shin Jo,
Hanbat National University

요 약

Cellular-Vehicle to everything (C-V2X)는 Long Term Evolution 기반의 V2X통신을 지칭하며, 해당 통신 기술의 요구 사항으로는 메시지 수신 신뢰도와 통신 범위 등이 있다. 그러나 C-V2X는 고밀도 상황에서 높은 혼잡도에 의해 해당 요구 사항을 만족시키지 못하며, 이를 위해 V2X관련 표준 기관은 혼잡제어 알고리즘을 발표했다. 그러나 이와 같은 기존 혼잡제어 알고리즘은 C-V2X 요구 사항을 고려하지 않은 채 설계되었기 때문에 요구 사항을 만족시키기 어렵다. 따라서, 본 논문에서는 C-V2X 요구 사항을 만족시키기 위해 심층강화학습을 이용한 중앙 집중식 혼잡제어 알고리즘을 제안한다. 제안하는 알고리즘에서 에이전트는 기지국으로 네트워크의 혼잡도를 측정하여 모든 차량의 전송 주기를 일괄적으로 제어한다. 본 논문에서 이용한 심층강화학습 알고리즘은 Deep Q-Networks (DQN)와 Double DQN이며, 시뮬레이션 결과 두 알고리즘 모두 목표로 하는 C-V2X 요구 사항을 만족시킨다.

Ⅰ. Cellular-Vehicle to everything (C-V2X)

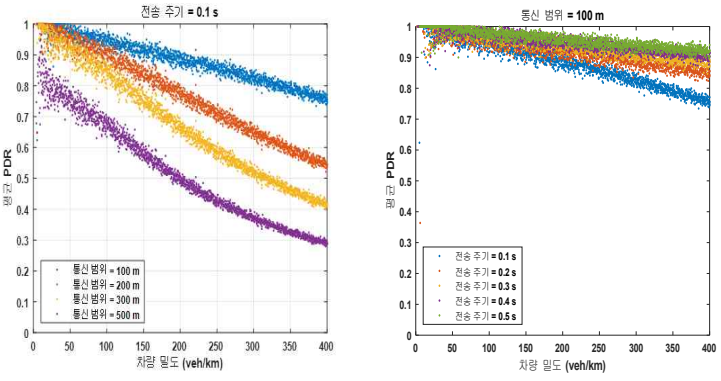
C-V2X는 Long Term Evolution 기반의 V2X통신을 지칭하며, 여기서 V2X란 차량이 유·무선망을 통해 다른 차량 및 도로 등 인프라가 구축된 사물과 정보를 교환하는 것 또는 그 기술을 말한다. 국제 표준화 단체 3rd Generation Partnership Project (3GPP)에서는 이러한 C-V2X 요구 사항을 기술 문서에서 제안하고 있으며, 그 예로 메시지 수신 신뢰도(Packet Delivery Ratio, PDR)는 통신 시나리오에 따라 80~95 %, 협력 인식 메시지의 통신 범위는 300~500 m로 제안하고 있다[1].

그러나 C-V2X는 고밀도 상황에서 통신 성능이 급격히 저하되기 때문에 PDR, 통신 범위와 같은 요구 사항을 만족시키기 어렵다. 이러한 문제점을 해결하기 각 V2X와 관련된 표준 기관들은 전송 전력제어, 전송 주기 제어와 같은 혼잡제어 알고리즘을 발표했지만, 이와 같은 기존 혼잡제어 알고리즘은 C-V2X의 요구 사항을 고려하지 않은 채 설계되었기 때문에 요구 사항을 효율적으로 만족시키지 못한다. 따라서 본 논문에서는 심층강화학습을 이용한 C-V2X 요구 사항 적응형 중앙 집중식 혼잡제어 알고리즘을 제안한다. 제안하는 알고리즘의 혼잡제어 기술은 전송 주기 제어를 이용하여 차량 간 패킷 충돌 현상을 조절한다. 이를 통해 목표로 하는 통신 요구사항을 만족시킬 수 있으며, 설정한 목표 통신성능에 따라 상승 시키거나 저하시킬 수 있다.

Ⅱ. 전송 주기 및 통신 범위설정에 따른 차량 네트워크 성능 변화

본 논문에서 활용한 C-V2X의 요구 사항은 PDR과 통신 범위이다. 다음 그림 1은 전송 주기, 통신 범위에 따른 차량 네트워크의 평균 PDR을 보여준다. 그림 1. (a)는 네트워크 내 차량 들의 전송 주기를 0.1 s로 고정시킨 결과이다. 결과에서 확인할 수 있듯이, 각 차량이 커버해야 할 통신 범위가 넓어질수록 평균 PDR이 낮아지는 것을 알 수 있다. 이러한 결과는 차량 간 송 수신 거리가 멀어짐에 따라 경로 손실에 의해 신호대 잡음 간섭비가 낮아지기 때문이다.

다음 그림 1. (b)는 통신 범위를 100 m로 설정하여, 차량 간 전송 주기 설정



(a) 통신 범위에 따른 평균 PDR (b) 전송 주기에 따른 평균 PDR

그림 1. 변수에 따른 차량 네트워크의 평균 PDR 변화

에 따른 네트워크의 평균 PDR을 보여준다. 결과에서와 같이 전송 주기를 늘릴수록 패킷 충돌 현상이 줄어들기 때문에, 평균 PDR결과가 높아지는 것을 확인할 수 있다. 그러나 네트워크가 혼잡하지 않은 상황에서 전송 주기를 늘리는 것은 메시지 지연을 야기하기 때문에, C-V2X 메시지 수신 신뢰도를 만족시킬 수 있는 적절한 전송 주기 선택이 필요하다.

최종적으로 목표로 하는 통신 범위 내 요구되는 평균 PDR 성능을 만족시키기 위해서는 통신 네트워크 상황에 따른 전송 주기 제어가 필요하다. 이는 네트워크 혼잡도에 따른 전송 주기의 조합 최적화 문제로 전환되며 높은 복잡성을 갖는다. 따라서, 본 논문에서는 이러한 조합 최적화 문제를 해결하기 위해 심층강화학습 알고리즘 중 Deep Q-Networks (DQN)와 Double DQN (DDQN)을 이용한다[2].

Ⅲ. 심층 강화학습을 이용한 C-V2X 요구 사항 적응형 중앙 집중식 혼잡제어 알고리즘

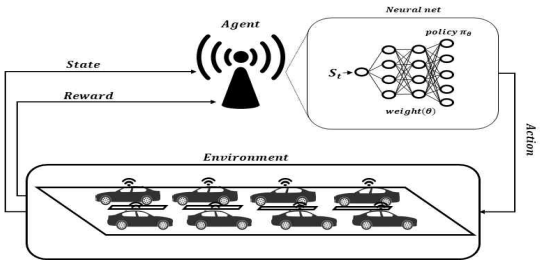


그림 2. 알고리즘의 강화학습 모델

제안하는 알고리즘에서 기지국은 심층강화학습을 통해 훈련되며, 각 차량은 도로의 혼잡도를 나타낼 수 있는 100 m 이내 노드 수를 기지국에게 전달한다. 이후 전달받은 정보를 토대로 기지국은 네트워크 내 모든 차량의 전송 주기를 변화 시키며, 이때 전송 주기는 모든 차량이 동일하게 부여받는다. 전송 주기는 기지국에 설정된 C-V2X 요구 사항 목표에 맞추도록 변화되며 이는 그림 2와 같은 강화학습 모델로 정의된다.

본 연구에서 적용된 심층강화학습 알고리즘인 DQN은 타겟 네트워크의 가중치 (θ')와 로컬 네트워크의 가중치 (θ)를 최적화하여, 에이전트의 정책 (π_θ)을 최적화시키는 알고리즘이다. 여기서 π_θ는 입력으로 t시점의 관측치 (State, S_t)를 입력으로 받아 행동 (Action, A_t)의 확률을 출력하는 함수이다. 최적화된 π_θ은 보상 (Reward, R_t)을 최대화한다. 제안하는 알고리즘에서 S_t는 100 m 이내 평균 노드 수로 설정했으며, A_t는 전송 주기로 0.1~0.5 s의 5개의 단계로 설정했다. 최종적으로 R_t는 C-V2X 요구 사항을 만족시키기 위해 다음과 같이 설정했다.

$$R_t = -|PDR_{Target} - OUTPUT| \quad (1)$$

여기서 PDR_{Target}은 목표로 설정한 네트워크의 통신 범위 내 평균 PDR이며, $OUTPUT$은 실제 측정된 평균 PDR값이다.

DQN알고리즘은 다음 수식과 같이 신경망의 손실 함수를 Mean Squared Error (MSE)로 정의한다.

$$L = \frac{1}{M}\sum_{i=1}^{M}(y_i + Q(S_i, A_i|\theta))^2 \quad (2)$$

위의 M은 experience replay memory에서 추출한 mini-batch 사이즈이며, $Q(S_i, A_i|\theta)$는 State에서의 Action의 가치를 나타내는 Q-value로 로컬 네트워크의 값이다. y_i는 목표 값으로 에피소드 종료 유/무에 따라 다음 식으로 정의된다.

$$\begin{cases} R_i & \text{for } S_t = S_T \\ R_i + \gamma \max_{A'} Q'(S_i, A'|\theta') & \text{for } S_t \neq S_T \end{cases} \quad (3)$$

식에서 γ는 감가율로 0과 1사이 값이며 T는 에피소드 종료 시점을 의미한다. $Q'(S_i, A'|\theta')$는 타겟 네트워크에서 추출한 Q-value이다. DQN과 달리 DDQN은 다음의 수식으로 y_i가 정의된다.

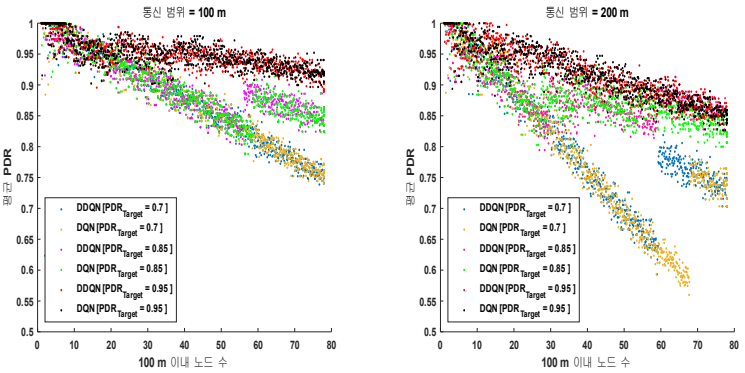
$$A_{\max} = \arg\max_{A'} Q(S_i', A'|\theta)$$
$$y_i = R_i + \gamma Q'(S_i', A_{\max}|\theta') \quad (4)$$

위와 같이 목표 값의 업데이트 수식을 다르게 하여 DQN과 달리 DDQN 은 학습의 편향을 줄여주는 효과를 낸다.

Ⅳ. 시뮬레이션 결과

표 1. 시뮬레이션 파라미터

파라미터	값
메시지 크기	300 byte
도로 길이	1 km
대역폭	10 MHz
안테나 이득 (G_t, G_r)	3 dB
목표 PDR (PDR_{Target})	0.95, 0.85, 0.7
통신 범위	100, 200 m
MCS index	3



(a) 통신 범위 = 100 m (b) 통신 범위 = 200 m

그림 2. 통신 범위와 목표 PDR에 따른 학습결과

시뮬레이션은 표 1의 파라미터를 이용했으며, DQN과 DDQN으로 각자 학습한 기지국의 성능을 평가한다. 시뮬레이션 평가는 차량 네트워크의 평균 PDR을 이용한다. 다음 그림 2는 통신 범위와 PDR_{Target}에 따른 두 가지 심층강화학습 알고리즘의 학습 결과를 보여준다. 그림 2. (a)는 통신 범위가 100 m로 설정했기 때문에 (b)보다 전반적으로 높은 수준의 평균 PDR을 보인다. 두 결과에서 모두 PDR_{Target}이 높을수록 기지국은 전송 주기를 0.5 s에 가깝게 이용했으며, PDR_{Target} 낮을수록 0.1s 가까운 전송 주기를 이용하였다. 또한 통신 범위가 100 m에서 200 m로 증가함에 따라 기지국은 같은 관측치에서 더욱 긴 전송 주기를 이용했다. 시뮬레이션 결과에서 DQN과 DDQN의 성능 차이가 크지 않은 이유는 시뮬레이션 환경에서 관측치에 대한 에이전트의 행동 양상이 뚜렷하기 때문으로 분석된다.

Ⅴ. 결론

본 논문에서는 심층강화학습을 이용하여 C-V2X 요구 사항에 적응적인 혼잡제어 알고리즘을 제안한다. 제안하는 알고리즘에서 에이전트는 기지국으로 네트워크 내 모든 차량의 전송 주기를 제어한다. 시뮬레이션 결과, 기지국은 네트워크 혼잡도에 따라 통신 범위 내 PDR_{Target}을 만족시키기 위해 유연한 전송 주기 제어를 보였으며, 이를 통해 PDR_{Target}을 만족시키는 것을 확인할 수 있었다. 추후 심층강화학습을 이용한 분산 혼잡제어 알고리즘을 연구할 계획이다.

참 고 문 헌

[1] Study on LTE support for Vehicle to Everything (V2X) services, 3GPP TR 22.885 V14.0.0, Dec. 2015.

[2] R. S. Sutton and A. G.Barto,Reinforcement learning: An introduction.MIT press, 2018

다관절 로봇팔을 위한 모방학습 기반 역기구학 해석 기술에 관한 연구

최진철, 박찬원, 박준희
한국전자통신연구원
{spiders22v, cwp, juni}@etri.re.kr

A Study on Imitation Learning-based Inverse Kinematics Analysis for multi-joint robot manipulators

Jinchul Choi, Chan-Won Park, Jun Hee Park
Electronics and Telecommunications Research Institute

요 약

본 논문은 다관절 로봇팔 제어를 위한 모방학습 기반의 역기구학 해석 기술을 제안한다. 제안기술은 로봇팔을 제어하는 시연 데이터셋을 모방하는 행동정책을 학습함으로써 로봇팔 말단부를 임의의 공간 좌표에 위치시키기 위한 다관절 제어값을 별도의 역기구학 모델 없이 도출할 수 있게 해 준다.

Ⅰ. 서 론

기구학(Forward Kinematics)은 로봇의 각 관절의 움직임에 따라 말단부(end-effector)의 공간좌표가 어떻게 변하는지를 분석하는 것이다. 대조적으로 역기구학(Inverse Kinematics)는 말단부가 특정 좌표에 위치하기 위한 관절 제어값을 도출하는 기법이다. 역기구학을 위해서는 기하학법, 대수행렬법, 자코비안 등이 활용되고 있으나, 계산이 복잡하고 로봇의 형태와 목표점에 따라 해답 도출에 실패하는 경우도 많다[1].

본 논문에서는 로봇팔 말단부를 특정 좌표에 도달시키는 다수의 시연 데이터를 모방학습함으로써 임의의 좌표에 말단부를 도달시키기 위한 관절값 결정 기법을 제안한다.

Ⅱ. 본론

본 논문에서는 마코프 결정 프로세스(Markov Decision Process)(MDP)라는 확률 모델로 문제를 표현한다. 학습을 위한 시연 데이터셋은 일련의 궤적 $T=\{\tau_1,\tau_2,...\}$으로 표현되며, 각 궤적 $\tau=\{(x,y,z),(s_1,a_1),(s_2,a_2),...\}$는 목표 공간좌표와 일련의 상태($s$)-행동($a$)쌍으로 구성된다. 여기서 상태 s는 각 관절의 상태값(joint position)이며, 행동 a는 변경해야 할 각 관절의 상태값이다. 본 논문의 목표는 로봇팔의 관절 상태가 s일때 말단부가 목표좌표 (x,y,z)에 도달하기 위해 취해야 하는 행동 a를 도출하는 정책 $\pi(a|s,(x,y,z))$을 학습하는 것이다.

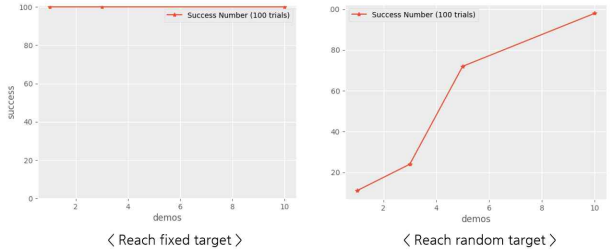
시연 데이터셋은 행동복제 기법을 통해 학습된다. 행동복제는 시연 데이터셋의 행동-상태 매핑 정보를 이용하여 시연자의 행동을 재연하는 일련의 규칙을 직접 도출한다. s_t^E와 a_t^E를 각각 t시간에서의 시연 데이터셋 상태와 행동이라고 하면 학습모델의 출력인 a_t를 결정하는 손실함수 L은 다음과 같다.

$$L=\Sigma_{(s_t^E,a_t^E,a_t)\in B}\parallel a_t^E-a_t\parallel_2^2 . \quad (1)$$

여기서 B는 에피소드 버퍼(Episodic Buffer)이다.

Ⅲ. 실험 결과

시연 데이터셋 획득과 학습모델의 테스트는 CoppeliaSim과 RLBench 프레임워크[2] 환경에서 6개의 회전관절 자유도(DoF)를 가진 UR3로봇을 이용하여 시뮬레이션되었다. 각 관절은 ±360° 범위의 상태공간(state space)과 행동공간(action space)을 가지며, 특정 공간 위치로 로봇암의 말단부를 위치시키는 작업(Reach fixed target)과 20×20×20㎤ 공간상에서 임의의 위치에 말단부를 위치시키는 작업(Reach random target) 등의 연속 제어 작업을 수행하였다. 그림 1은 시연 데이터량에 따른 모방학습 모델의 작업 성공률을 보여준다. 테스트에는 20k epoch을 학습한 모델이 사용되었으며, 100회의 에피소드를 반복하여 성공률을 측정하였다.



〈 Reach fixed target 〉 〈 Reach random target 〉

그림 1. 시연 데이터량에 따른 작업 성공률(%)

Ⅳ. 결론

본 논문에서는 모방학습을 통해 다관절 로봇팔의 역기구학을 해석하는 방법을 제안하였다. 실험 결과는 시연 데이터셋의 모방학습을 통해 임의의 공간좌표에 도달하는 로봇팔 관절제어가 가능함을 보여주었다.

ACKNOWLEDGMENT

본 연구 논문은 한국전자통신연구원 연구운영지원사업의 일환으로 수행되었음. [21ZR1100, 자율적으로 연결· 제어·진화하는 초연결 지능화 기술 연구]

참 고 문 헌

[1] A. Almusawi ,L. Dülger and S. Kapucu, " A New Artificial Neural Network Approach in Solving Inverse Kinematics of Robotic Arm," Comput. Intell. Neurosci. Article No. 5720163, 2016.

[2] S. James, Z. Ma, D. R. Arrojo and A. J. Davison, "RLBench: The Robot Learning Benchmark & Learning Environment," arXiv preprint Xi1909122712019.

딥러닝 기반 가짜뉴스 탐지 시스템에서 뉴스 카테고리의 영향 분석

한윤진*, 김성한**, 김근형*
*동의대학교, **한국전자통신연구원
qkr030@naver.com, sh-kim@etri.re.kr, geunkim@deu.ac.kr

Analysis of the effect of News Categories in Deep Learning-based Fake News Detection Systems

YoonJin Han*, Sunghwan Kim**, Geun-Hyung Kim*
*Dong-eui Univ., **ETRI

요 약

소셜미디어(SNS)가 보편화되어 개인 방송 채널이 증가하면서 거짓 정보를 언론 보도 형식으로 위장한 가짜뉴스가 꾸준하게 증가하고 있어 사회적 비용을 많이 발생시킬 것이다. 최근 딥러닝 기술을 활용하여 가짜뉴스를 자동으로 탐지하는 연구가 활발히 진행되고 있다. 본 논문에서는 가짜뉴스 탐지 모델 개발을 위해 설계한 메타 데이터 체계에 따른 가짜뉴스의 분류가 가짜뉴스 탐지 모델의 성능에 미치는 영향을 분석하였다.

Ⅰ. 서 론

한국언론진흥재단은 정치·경제적 이익을 얻으려 의도적으로 언론 보도 형식으로 유포된 거짓 정보를 가짜뉴스로 정의하고 가짜뉴스를 정보 형태에 따라 동영상, 오디오, 이미지, 텍스트로 구분하며 정보 성격에 따라 허위정보, 거짓 정보, 오인정보, 패러디, 루머로 분류한다[1]. 가짜뉴스 탐지 방법은 크게 비기술적 방법, 기술적 방법, 하이브리드 방법으로 분류한다[2].

본 논문에서는 기술적 가짜뉴스 탐지 방법인 딥러닝 기반 가짜뉴스 탐지 모델의 성능과 뉴스 데이터의 카테고리 분류의 관계 (뉴스 카테고리가 모델 성능에 미치는 영향) 를 LSTM(Long Short-Term Model) [3] 기반의 딥러닝 네트워크를 사용하여 분석하였다.

Ⅱ. 본론

본 논문에서는 다양한 미디어의 가짜뉴스를 콘텐츠 또는 맥락을 기반으로 검출하는 인공지능 모델을 개발할 때 활용되는 뉴스 데이터를 분류하기 위해 정립한 메타 데이터 체계[4]를 바탕으로 Kaggle[5]의 뉴스 데이터를 분류하여 뉴스 데이터의 분류가 딥러닝 기반 가짜뉴스 탐지 모델의 성능에 미치는 영향을 분석하였다. 정립한 메타 데이터 체계는 미디어에 공통이고 필수적인 메타 데이터 요소와 미디어 특화되고 선택적인 메타 데이터 요소로 구분하였고 공통적인 메타 데이터 요소는 카테고리, 제목, 본문 또는 콘텐츠에 해당하는 내용 또는 링크, 뉴스 게재 일자, 뉴스 게시자, 출처가 있다.

본 연구에서는 가짜뉴스 탐지 모델을 개발 및 활용할 때 뉴스 데이터를 메타 데이터 중 공통요소의 중요성을 강조하기 위해 실험을 진행하였다. 우리는 실험을 통해 메타 데이터의 공통요소 중 카테고리 요소가 가짜뉴스 탐지에 얼마나 영향을 미치는지 분석하였다. 실험에 사용한 인공지능 모델은 심층 신경망(DNN: Deep Neural Network)중 하나인 LSTM 네트워크로 구성하였고, 학습 및 테스트용 뉴스 데이터는 Kaggle의 뉴스 데이터 셋에서 진짜뉴스 21,417개와 가짜뉴스 23,481개를 추출하여 사용되었다. 검증용 뉴스 데이터는 학습용 뉴스 데이터의 일부를 활용한다. 사용된 Kaggle의 뉴스 데이터는 제목, 본문, 카테고리, 그리고 게시일자 정보로 이루어져 있다.

첫 번째 실험에서는 카테고리 상관없이 진짜뉴스 데이터 5,000개와 가짜뉴스 데이터 5,000개를 무작위로 추출하였다. 추출한 10,000개의 뉴스 데이터는 'news'로 분류된 데이터가 4,516개, 'worldnews'로 분류된 데이터가 2,410개, 'politics'로 분류된 데이터가 1,469개, 'left-news'로 분류된 데이터가 942개, 'government News'로 분류된 데이터가 330개, 'middle-east'로 분류된 데이터가 171개, 'us_news'로 분류된 데이터가 162개로 구성되었다. 첫 번째 실험에서는 학습 데이터 8,000개와 테스트 데이터 2,000개로 나누어 사용하였으며, 학습 과정에서 학습 데이터의 20%를 검증 데이터로 사용하였다.

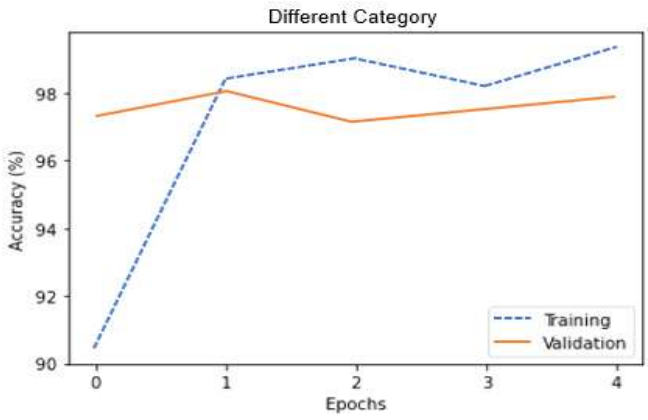


그림 1. 카테고리 분류가 없는 뉴스 활용 학습 결과.

두 번째 실험에서는 동일한 카테고리에 속한 진짜뉴스 데이터 5,000개와 가짜뉴스 데이터 5,000개를 무작위로 추출하였다. 10,000개의 뉴스 데이터는 'news'로 분류된 데이터로만 구성하였다. 첫 번째 실험과 같은 비율로 데이터를 분리하여 실험에 사용하였으며 두 가지 실험 모두 5번의 epoch를 반복하여 학습하였다.

그림 1은 카테고리를 분류하지 않은 실험의 결과를 나타낸다. 카테고리 종류에 상관없이 데이터를 추출하여 사용한 첫 번째 실험 결과(그림 1)를 보면 테스트 데이터는 epoch를 반복할수록 학습 정확도는 증가하나 검증 데이터는 정확도가 0.98을 넘지 않고 유지된다.

동일 카테고리에 속한 뉴스 데이터를 기반으로 실험한 결과는 그림 2와 같다. 여러 카테고리에 속한 뉴스 데이터를 기반으로 실험한 결과보다 동일 카테고리에 속한 뉴스 데이터를 기반으로 한 실험이 가짜뉴스 탐색 모델의 정확도가 높다.

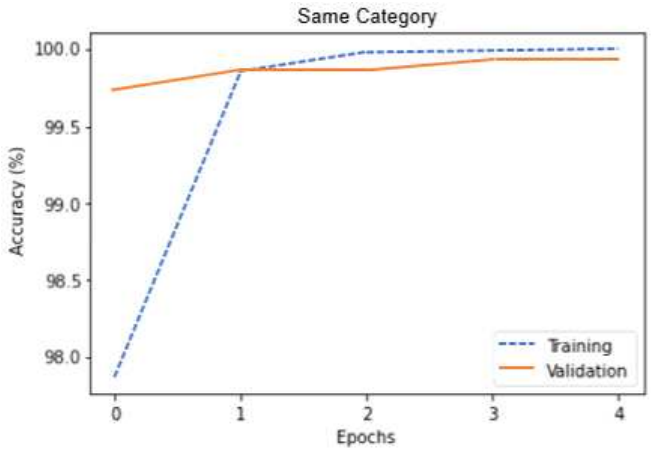


그림 2. 동일 카테고리로 분류된 뉴스 활용 학습 결과.

카테고리가 다른 뉴스 데이터를 이용한 실험 결과(그림 1)와 달리 동일한 카테고리로 분류된 뉴스 데이터를 이용한 두 번째 실험 결과(그림 2)를 보면 테스트 데이터는 epoch를 반복할수록 학습 정확도와 검증 정확도가 모두 증가한다.

학습된 두 모델에 테스트 데이터를 사용한 테스트를 하면 첫 번째 학습 모델의 가짜뉴스 탐지 정확도는 47.36%, 두 번째 학습 모델의 가짜뉴스 탐지 정확도는 98.68%이다. 테스트에 사용된 2,000개의 뉴스 데이터는 'news' 카테고리 데이터가 594개, 'worldnews' 카테고리 데이터가 469개, 'politics' 카테고리 데이터가 367개, 'left-news' 카테고리 데이터가 359개, 'government News' 카테고리 데이터가 150개, 'middle-east' 카테고리 데이터가 31개, 'us_news' 카테고리 데이터 30개로 이루어졌다. 그러나 학습에 사용된 10,000개의 뉴스 데이터 중 4,516개가 'news' 카테고리에 해당되어 테스트 시 가짜뉴스 탐지의 정확도가 낮았다.

실험 결과 학습 데이터를 세부 카테고리로 분류하지 않고 학습시킨 가짜뉴스 탐지 모델은 다양한 뉴스 데이터에 대해 가짜뉴스를 탐지할 때 정확도가 낮았다. 실험 결과로부터 유사한 특징(예: 카테고리)을 가지는 뉴스 데이터로 학습시킨 가짜뉴스 탐지 모델에 동일한 카테고리에 속한 뉴스에 대해서만 뉴스의 진위를 판단할 때 정확성이 높다고 추론할 수 있다.

인공지능 모델은 설계 시 설정한 조건 범위 및 지식의 범위 내에서만 잘 동작하는데 모델이 작동하도록 설계 또는 고려되지 않은 경우의 데이터에 대해서는 모델의 정확도가 낮아지는데 이와 같은 인공지능 모델의 지식 한계[6]를 식별하고 선언하는 것이 필요하다.

Ⅲ. 결론

본 논문에서는 가짜뉴스 탐지 모델 개발을 위해 정립한 메타 데이터 체계에 따른 가짜뉴스의 카테고리 분류가 가짜뉴스 탐지 모델의 성능에 미치는 영향을 분석하였다. 가짜뉴스가 다양한 미디어 기반으로 생성되고 있어 가짜뉴스를 분류할 수 있는 메타 데이터 체계가 요구되며 메타 데이터의 카테고리와 인공지능 모델의 지식 한계가 가짜뉴스 탐지 모델의 정확도에 미치는 영향을 분석하게 위해서 뉴스 데이터를 카테고리 분류 없이 학습 데이터로 활용한 가짜뉴스 탐지 모델과 하나의 카테고리에 속한 뉴스 데이터를 학습 데이터로 활용한 가짜뉴스 탐지 모델의 성능을 분석하였고 카테고리별 가짜뉴스 탐지 모델을 개발하는 것이 적합하다는 결론을 도출하였다.

가짜뉴스 탐지 모델을 활용하기 위해 뉴스를 메타 데이터 체계에 따른 올바른 분류가 가짜뉴스 탐지 모델의 성능에 영향을 미치기 때문에 뉴스 데이터의 자동 분류 기능이 요구되며 인공지능 모델의 지식 한계를 규정하고 지식 한계를 벗어나는 것에 대해서 결정을 유보할 수 있어야 한다.

감사의 글

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 한국 연구재단의 지원을 받아 수행된 연구임(No. NRF-2021R1F1A1047573).

참 고 문 헌

[1] 좌희정, 오동석, 임희석, "자동화기반 가짜 뉴스 탐지를 위한 연구 분석", 한국융합학회논문지, 제10권, 7호, pp15-21, 2019.

[2] 윤영석, 엄태원, 이재영, 이현우, 허재두, "페이크 뉴스 탐지 기술 동향과 시사점," 주간기술동향, 정보통신기술진흥센터, Oct. 2017.

[3] S. Hochreiter, J. Schmidhuber, "Long Short-Term Memory," Neural Computation, 9(8), pp. 1735-1780, 1997.

[4] 김근형, 김성한, 한윤진, "인공지능 기반 가짜뉴스 탐지 시스템을 위한 뉴스 메타 데이터 연구 분석," 한국멀티미디어학회 추계학술발표대회논문집, 제 23권 2호, 2020.

[5] Kaggle 홈페이지, https://www.kaggle.com.

[6] P. J. Phillips, C. A. Hahn, P. C. Fontana, D. A. Broniatowski, M. A. Przybocki, "Four Principles of Explainable Artificial Intelligence," NIST Report, Draft NISTIR 8312, August 2020.

길 안내 로봇의 신호등 없는 횡단보도 보행 안내를 위한 차량 인식 및 거리 측정 모듈 개발

최인훈, 홍승준, 김서현, 정홍주, 박수용, 허의남*

경희대학교

{inhun321, hongsj1022, seok02h, sub06038, clapd10, *johnhuh}@khu.ac.kr

An Implementation of Vehicle Recognition and Distance Estimation Module for Guiding Pedestrian Crossing without Traffic Lights of Road Guide Robot

Inhun Choi, Seungjun Hong, Seo-Hyun Kim,

Hong-Ju Jeong, Suyong Bahk, Eui-Nam Huh*

Kyung Hee Univ.

요 약

본 연구에서는 차량 인식 및 거리 정보를 제공하는 모듈 개발을 진행하였다. 이 모듈은 길 안내 로봇이 신호등 없는 횡단보도에서 주변 차량의 위치, 거리, 속도 정보를 이용하여 보행 가능 여부를 판단할 수 있도록 하였다. 이를 통해 보행약자들이 안정적인 횡단보도 보행 안내 서비스를 제공 받을 수 있다.

1. 서 론

기계 학습 기술이 발전함에 따라 최근 보행에 불편을 느끼는 시각장애인, 어린이, 노인 등의 보행약자들을 위한 길 안내 로봇에 활발히 개발되고 있다. 길 안내 로봇은 목적지까지의 길 안내뿐만 아니라 보행 시 마주칠 수 있는 위험 상황에 대해서도 안전하게 대처해야 한다. 특히 신호등 없는 횡단보도에서 보행약자들이 안전하게 횡단할 수 있도록 안내해야 한다.

따라서 본 논문은 딥러닝 기반 차량 인식 및 거리 측정 모듈을 사용하여 주변의 차량 정보들을 이용하여 사용자가 신호등 없는 횡단보도를 안전하게 보행할 수 있는 서비스를 제공하는 것을 목표로 한다. 차량 추적 알고리즘을 통해 움직이는 영상 속 동일한 차량을 판단하고 거리 데이터를 통해 차량들의 속도를 계산하여 길 안내 로봇이 보행약자들이 안전하게 횡단보도를 횡단할 수 있도록 돕는 모듈을 개발하였다.

2. 본론

2.1 객체 인식(Object Detection)

객체 인식은 객체의 클래스를 판단하는 Classification과 존재하는 위치를 판단하는 Localization으로 이루어 진다. 객체 인식 모델 크게 객체가 있을만한 후보 영역들을 만들고 이 후 분류기에 통과시키는 비교적 정확도가 높지만 연산량이 많은 2-Stage Detector와 한번의 CNN (Convolution Neural Network) 연산으로 클래스 정보와 위치 정보를 판단하는 비교적 정확도가 낮지만 연산량이 적은 1-Stage Detector로 나눌 수 있다.

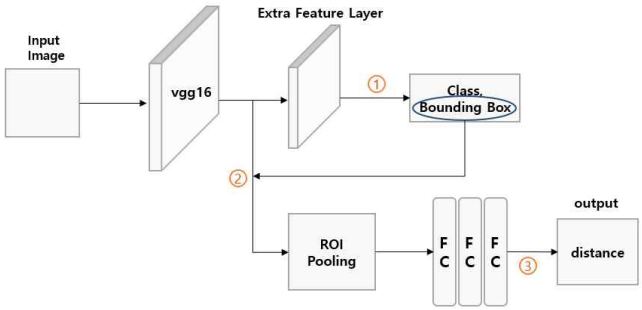
실제 도로 상황에서 사용하기 위해서 실시간성이 보장되어야 하므로 1-Stage-Detector 중 VGG16[1]을 특징추출 모델로 사용하는 SSD (Single Shot MultiBox Detector)[2] 모델을 선택하였다. 차량 인식이 목표이기 때문에 차량 인식 모델 학습을 위한 데이터셋으로 차량 이미지와 이미지에 존재하는 차량들의 Bounding Box가 포함된 KITTI Dataset[3]을 사용하였다.

2.2 단안 카메라 거리 측정(Monocular Camera based Distance Estimation)

카메라를 통한 거리측정 기술로는 크게 양안 카메라를 이용한 기술과 단안 카메라를 이용한 기술로 나눌 수 있다. 양안 카메라를 이용한 경우 정확도는 높지만 움직이는 물체에 카메라를 장착할 경우 충격이나 흔들림에 의해 교정된 값이 비틀릴 수 있고 비용이 비싸다는 단점이 있다[4]. 움직이는 로봇에 카메라를 장착해야하기 때문에 본 연구에서는 단안 카메라를 이용해서 거리를 측정[5]하기로 했다.

SSD를 통해 얻은 차량이 위치하는 Bounding Box정보를 이용하여 특징 맵에 투영시키고 Fully Connected Layer의 입력으로 사용하기 위해 ROI Pooling을 진행하여 모두 같은 크기의 특징 맵으로 만든다. 이 후 Fully Connected Layer들을 이용하여 카메라로부터 차량까지의 거리 값을 출력한다. 거리 측정 모델 학습을 위한 데이터셋도 차량 이미지와 카메라와 차량까지의 거리 데이터를 포함하고 있는 KITTI Dataset 을 사용하였다.

특징 맵을 얻기 위한 특징 추출모델로 Resnet, VGG16, Darknet을 이용하여 학습을 진행하였다. Darknet의 경우 차량이 아주 가까이에 있거나 아주 멀리 있을 경우 결과값이 실제 거리값과 차이가 많이 나는 결과를 얻었고 Resnet의 경우 RMSE 10.4325, VGG16의 경우 RMSE 6.7512의 오차를 얻었다. 따라서 최종적으로 VGG16을 특징 추출 모델로 선택하였다. 객체 인식 모델인 SSD모델에서도 VGG16을 특징 추출 모델로 사용하기 때문에 연산량을 줄일 수 있을 것으로 보인다. 이때의 ROI Pooling을 위한 피라미드 사이즈는 2*2이다.

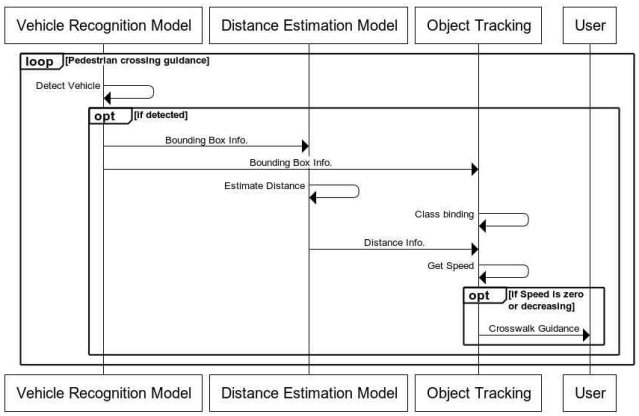


[그림 1] 차량 인식 및 거리 측정 모델 전체 구조

그림 1은 차량 인식 및 거리 측정 모델 전체 구조이다. 우선 SSD모델을 통해 차량의 위치 정보를 구한 후, 위치 정보를 통해 ROI Pooling을 이용해 크기가 일정한 특징 맵으로 만든 후 Fully Connected Layer를 통해 거리값을 출력한다.

2.3 객체 추적(Object Track)

차량의 속도값을 얻기 위해 움직이는 영상 속 동일한 차량을 Tracking하는 과정이 필요하다. 이전 frame에서 얻은 Bounding Box정보들과 현재 frame에서 얻은 Bounding Box정보들을 이용하여 교집합 영역 넓이를 합집합 영역 넓이로 나눈 값인 IoU (Intersection over Union)이 일정값보다 크면 같은 객체라고 판단한 후 같은 ID를 부여한 후 Tracking[6]한다. 해당 차량의 거리 데이터를 이용하여 접근하는 속도를 계산한다. 차량이 정지해 있거나 일정 속도 이하로 줄어들면 보행 가능하다고 판단한 후 보행자에게 보행 가능 여부를 알려준다. 그림 2는 횡단보도 보행 안내 시퀀스 다이어그램 이다.



[그림 2] 차량 인식 및 거리 측정 모델을 이용한 횡단보도 보행 안내 시퀀스 다이어그램

2.4 성능 평가

[그림 3]은 차량 인식 및 거리 측정 모델의 추론 결과이다. 차량 인식 모델은 정확도 약 93%, 거리 측정 모델은 약 RMSE: 6.7512의 오차를 보였다. 하지만 라즈베리 파이4 에서의 평균 처리 시간이 500ms 이상이 걸리는 것을 확인하였다. 실제 도로 환경에서 실시간으로 처리하기 어려운 것을 알 수 있다. 따라서 엣지 컴퓨팅 서버를 두어 실시간 이미지 전송을 통해 모델 처리 속도를 높여야 할 것이다.



[그림 3] 차량 인식 및 거리 측정 모델 추론 결과

3. 결론

본 연구는 횡단보도에서의 보행 가능 여부를 판단하기 위해 차량 인식 및 거리 측정을 통해 길안내 로봇에게 주변 차량의 정보를 제공해주는 모듈을 개발하였다. 약 93%의 정확도를 보이는 차량 인식 모델과 약 RMSE 6.7512의 오차를 보이는 거리 측정 모델을 이용하여 높은 성능의 네트워크와 엣지 서버를 이용한다면 안정적인 길안내 로봇의 성능을 기대할 수 있을 것으로 기대된다.

향후, 본 눈문에서 제시한 모듈을 길 안내를 위해 제작된 로봇에 적용하여 실제 도로 환경에서 처리 속도 및 정확도를 실험하고자 한다.

ACKNOWLEDGMENT

이 논문은 2021 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2017-0-00294, (대학 ICT 기초 연구실) 서비스 이동 지원을 위한 분산형 클라우드 핵심 원천기술 연구)

참 고 문 헌

[1] Karen Simonyan and Andrew Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition”, International Conference on Learning Representations, 2015

[2] Wei Liu, “SSD: Single Shot MultiBox Detector”, ECCV 2016

[3] Andreas Geiger, Piilip Lenz, and Raquel Urtasun. Are we ready for autonomous diving? thd kitti vision benchmark suite. Conference on Computer Vision and Pattern Recognition (CVPR), 2012

[4] [IITP] “딥러닝 기반 거리측정 기술 동향” - 김혜진 한국전자통신연구원 선임연구원

[5] Jing Zhu. “Learning Object-specific Distance from a Monocular Image”, ICCV 2019

[6] Himani S, "A Survey on Object Detection and Tracking Methods" International Journal of Innovative Research in Computer and Communication Engineering, pp. 2970-2978, 2014

Time-GAN을 활용한 트레이딩 알고리즘 선택

이재윤, 이주홍*, 최범기**, 송재원***

인하대학교, *인하대학교, **㈜큐헷지, ***㈜밸류파인더스

chris0258@naver.com, *juhong@inha.ac.kr, **bgchoi666@gmail.com, ***jwsong@valuefinders.co.kr

Trading Algorithm Selection Using Time-GAN

Jae Yoon Lee, Ju-Hong* Lee, Bum Gi Choi**, Jae Won Song***

Inha Univ., *Inha Univ., **QHedge Co. Ltd., ***ValueFinders Co. Ltd.

요 약

본 논문은 주식시장에서 Time-GAN 모델을 이용하여 여러 트레이딩 알고리즘들의 각각의 미래의 수익률 시계열을 생성한 후, 생성된 수익률 시계열의 분산과 평균을 이용하여 각 모델의 미래 수익률 예측 정확도와 수익률의 높고 낮음을 판단하고, Time-GAN에 의한 모델 예측도가 높고 수익률이 좋은 트레이딩 알고리즘을 선택한다. 선택된 트레이딩 알고리즘의 매수, 매도 시그널을 이용하여 안정적으로 최대의 수익률을 내는 투자 모델을 제안하였다.

Ⅰ. 서 론

현대 사회에서 금융시장을 분석하는 것은 그 사회의 경제를 분석하는 것 만큼이나 중요하기에 금융시장을 분석하는 것은 없이 논의할 수 없을 만큼 중요한 역할을 한다. 금융시장을 분석하는 방법은 크게 기본적 분석과 기술적 분석이 있다. 기본적 분석은 가치를 분석하는 방법이고, 기술적 방법은 데이터를 이용해 분석을 하는 방법이다. 기술적 분석을 이용하면 매수 또는 매도 시그널을 이용해 거래를 하는 방법이 있다.[1] 본 논문에서는 기술적 분석을 이용하는 트레이딩 알고리즘들을 이용하여 적은 변동성으로 수익을 내는 방법을 제안한다.

최근에 과거 데이터의 기술적 분석 방법을 기계학습, 신경망, 유전 알고리즘들을 이용하여 고도화한 트레이딩 알고리즘에 관한 연구들이 이루어졌다[2, 3, 4]. 다만 시간에 따라 통계적 특성이 변하는 비정상성 때문에 기술적 분석만으로는 완벽히 분석되지 않는다는 점이 있다. 즉 특정 구간에서는 잘 작동될 수 있지만, 모든 구간에서는 원하는 결과가 나오기 어렵다는 단점이 있다.[5] 이러한 문제점을 해결하기 위하여 Time-GAN (Time series Generative Adversarial Networks) 을 이용하여 각 트레이딩 알고리즘들이 어떤 성능을 보일지 예측하고 그 결과를 평가하여 여러 트레이딩 알고리즘 중에서 어떤 것이 해당 시점에서 가장 적합할지를 판단하여 투자에 사용할 트레이딩 알고리즘을 선택하는 투자 모델을 제안한다.[6] 이 모델은 2015년부터 2020년까지 5년간 삼성전자 데이터에 대하여 실험하여 좋은 성능을 보여주었다.

이후 사용되는 표기법은 다음과 같다.

$\bar{x}_t$: t 시점에서의 시가, 고가, 저가, 종가, 거래량을 가진 벡터

$(X_t)_k$: t-1 에서 t 시점까지 k 번째 알고리즘을 사용했을때의 수익률

T : 전체 데이터 크기

S : 주기 고정 길이

K : 트레이딩 알고리즘의 총 갯수

$(X_t)_k'$: k번째 Time-GAN 으로 생성한 t 시점 데이터

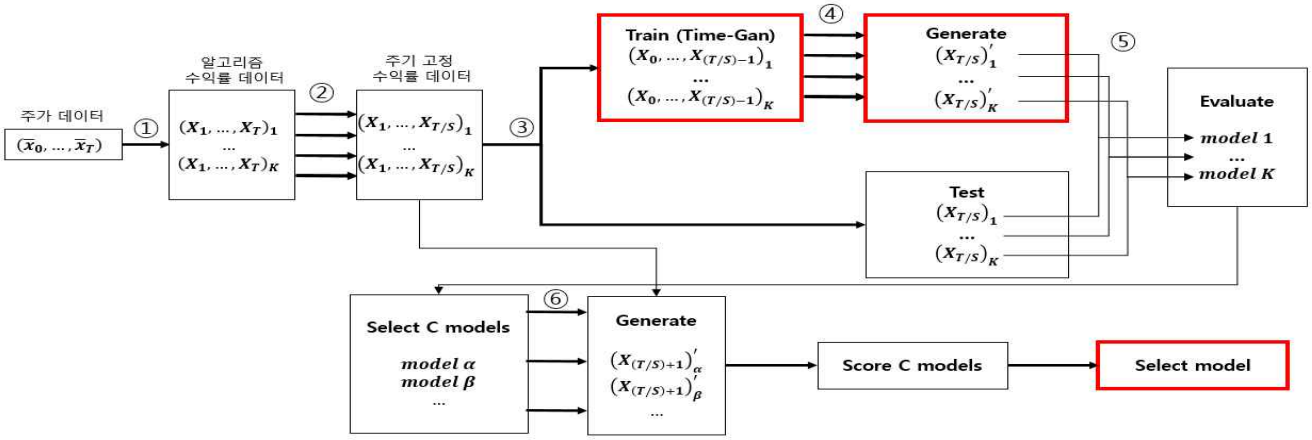


그림 1. 제안된 투자 모델 개요도

Ⅱ. 본론

본 논문에서 제안하는 투자 모델은 그림 1 과 같은 구성을 가진다. 전처리, 학습, 선택 3 단계로 분류되어 있다. 전처리 단계는 그림 1에서 ①과 ②로서 Time-GAN의 훈련 시계열 데이터를 준비하는 단계이다. ①은 주가 시계열 데이터를 K개의 트레이딩 알고리즘으로 각각 거래를 했을 때의 수익률 시계열들로 변환한다. ② 는 각 수익률 데이터를 압축해주는 과정이다. 예를 들면 압축주기가 5일이면 일별 데이터를 주별 데이터로 변환한다. 압축을 하면 학습할 데이터를 줄임으로써 학습시간을 줄일 수 있고, 압축하지 않았을 때 트레이딩 알고리즘들이 거래를 하지 않는 구간이 많이 발생하여 학습의 과적합이 되는 것을 방지하기 위함이다. 압축하게 되면 선택된 트레이딩 알고리즘이 교체되지 않고 압축한 주기 S 동안 동일한 트레이딩 알고리즘을 사용하게 된다. ②를 수행하면, K개의 트레이딩 알고리즘들의 수익률 시계열이 (T/S)의 압축된 수익률 시계열로 변환된다. 그림1에서는 훈련하기 위한 하나의 데이터만 표시하였으나 실제로는 과거의 여러 구간에 대해서 훈련 데이터를 생성한다. 학습 단계는 ③, ④, ⑤ 로 이루어져 있다. ③은 모든 K개의 압축된 수익률 시계열에서 마지막 수익률 데이터인 $X_{(T/S)}$를 제외하고 $X_{(T/S)}$ 이전의 시계열을 이용해 각 트레이딩 알고리즘별 수익률 시계열에 대해서 각각 다른 Time-GAN을 이용해 학습시킨다. ④는 학습된 Time-GAN을 이용해 $(X_{(T/S)})'$을 생성한다. ⑤는 생성된 결과의 평균과 분산을 이용하여 실제값을 비교하여 Time-GAN 예측 성능을 평가한다. ⑥은 선택단계로서 Time-GAN의 예측 성능이 좋은 상위 C개의 알고리즘을 선별한다. 그러면 선별된 알고리즘에 대한 Time-GAN의 예측 결과를 신뢰할 수 있다. 이후에 선별된 알고리즘에 대해서 다음 실제 거래구간에서 각 Time-GAN이 대응되는 알고리즘의 수익률을 예측하게 한다. 알고리즘들의 수익률 예측 결과 중에서 변동성이 작고 수익률이 높은 알고리즘을 최종적으로 선택한다. 만약 i번째 트레이딩 알고리즘이 선택되었다면 알고리즘은 다음 실제 거래 기간인 S일 동안은 i번째 트레이딩 알고리즘이 보내는 매수, 매도 시그널을 따라 투자를 하는 것이다.

본 논문에서는 제안한 투자 모델을 성능을 평가하기 위해 2015년 11월 11일부터 2020년 9월 29일까지 영업일 기준 1200일 삼성전자 주가에 대하여 sliding window 방식을 이용해 실험을 진행하였다. 초기 훈련기간은 2000년 1월 4일부터 2015년 11월 10부터 영업일 기준 3100일의 데이터를 사용하였다. 결과는 그림 2 와 같다. 1200일 동안의 트레이딩 알고리즘들의 수익률과 주가의 수익률을 나타낸 그래프로 빨간색 선은 주가의 종가이고 검은색은 투자 모델을 사용했을때의 결과, 나머지는 14개의 트레이딩 알고리즘들 이다.



그림 2. 삼성전자 2015년-2020년

	샤프 지수	최종 수익률	변동성
주가	0.73	117%	0.0169
투자 모델	1.13	116%	0.0096

표 1. 결과

표 1 에서 보는 바와 같이 최종 수익률은 주가가 117% 상승 하는 동안 투자 모델도 116%의 수익을 내었다. 투자 모델의 수익률은 비슷하지만 투자 모델의 변동성은 주가보다 약 2배 정도 작기 때문에 샤프 지수가 높게 나와서 안정적인 투자를 할 수 있는 모델임을 확인할 수 있다.

Ⅲ. 결론

본 논문은 트레이딩 알고리즘이 항상 좋은 결과를 내기는 힘들다는 한계점을 가진다는 단점을 보완하기 위해 여러 트레이딩 알고리즘을 사용해서 구간별로 적은 변동성으로 좋은 수익률이 예상되는 트레이딩 알고리즘을 찾는 투자 모델을 제안하였다. 제안 방법은 대상 종목의 주가와 비교해 더 높은 샤프지수를 보여주며, 제안된 투자 모델의 성과가 좋음을 보여주었다.

ACKNOWLEDGMENT

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (2019R1F1A1062094, 2017R1D1A1A02018319, NRF-2020R1F1A1069361)

참 고 문 헌

[1] Nuti, G., Mirghaemi, M., Treleaven, P., & Yingsaeree, C."Algorithmic trading." Computer, 44(11), 61-69. 2011

[2] Shao, X., Ma, D., Liu, Y., & Yin, Q. "Short-term forecast of stock price of multi-branch LSTM based on K-means." In 2017 4th International Conference on Systems and Informatics (ICSAI) (pp. 1546-1551). IEEE. 2017

[3] Selvin, S., Vinayakumar, R., Gopalakrishnan, E. A., Menon, V. K., & Soman, K. P. "Stock price prediction using LSTM, RNN and CNN-sliding window model." In 2017 international conference on advances in computing, communications and informatics (icacci) (pp. 1643-1647). IEEE. 2017

[4] Chung, H., & Shin, K. S. ."Genetic algorithm-optimized long short-term memory network for stock market prediction." Sustainability, 10(10), 3765. 2018

[5] Mills, T. C. . "Technical analysis and the London Stock Exchange: Testing trading rules using the FT30." International Journal of Finance & Economics, 2(4), 319-331. 1997

[6] Yoon, J., Jarrett, D., & Van der Schaar, M.. "Time-series generative adversarial networks." NeurIPS, 2019

학습자 정밀 평가를 위한 개인 맞춤형 테스트셋 생성 시스템

김연찬, 운봉영, 아가르왈 판카즈
㈜태그하이브

ychan@tag-hive.com, wbyoung@tag-hive.com, pankaj@tag-hive.com

Personalized Test Set Generating System for Advanced Student Assessment

Kim Yeoun Chan, Woon Bong Young, Agarwal Pankaj
TagHive, Inc.

요 약

본 논문은 코로나 시대에 주목을 받고 있는 에듀테크 시장에서 인공지능 기술의 주요 역할인 학습자 평가를 개별 문제 풀이가 아닌 다양한 난이도와 문제 배열 순서를 지닌 시험 상황에서 정밀 평가하기 위한 테스트셋 생성 시스템을 제안한다. 본 시스템은 Random Forest 분류 알고리즘을 사용한 정답률 예측 모델과 문제 배열 순서를 최적화하는 모델로 구성하였다. 본 연구에서 제안한 시스템을 적용할 경우 적용하지 않은 비교군에 비하여 예상 시험 점수에서 높은 성능을 확인하였다.

Ⅰ. 서론

최근 COVID-19의 영향으로 비대면 교육의 수요가 증가하면서 에듀테크(Edu-Tech) 시장에 대한 관심이 급격하게 높아지고 있다. 4차 산업혁명의 혁신적인 기술을 교육 분야에 접목하는 연구가 다방면으로 진행되고 있으며 특히 인공지능은 온라인 환경에서 학습 도중 발생하는 정답 유무, 문제 풀이 시간 등 다양한 데이터를 분석하여 개인 맞춤형 학습 및 고도화된 개인 평가를 제공하는 서비스에 주로 사용되고 있다[1]. 학습자의 평가에 있어 개별 문제 풀이 과정을 이용하는 연구[2,3]는 다양하게 진행되고 있지만 다양한 난이도와 문제 배열 순서의 변수를 가진 시험 상황에서의 연구는 부족한 편이다. 본 논문에서는 시험 상황에서 학습자를 정밀하게 평가하기 위한 개인 맞춤형 테스트셋 생성 방법을 제안한다.

Ⅱ. 개인 맞춤형 테스트셋 생성 시스템

본 논문에서는 그림 1과 같이 정답률 예측 모델과 문제 배열 순서 최적화 모델 기반의 학습자 정밀 평가가 목적인 개인 맞춤형 테스트셋을 생성하는 시스템을 제안한다.

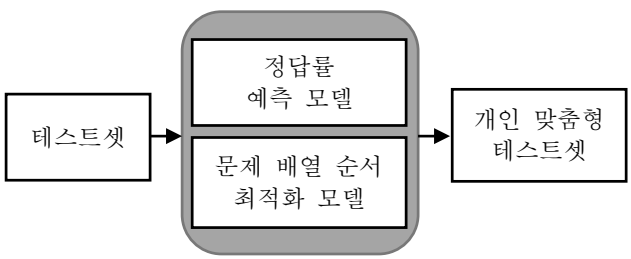


그림 1. 개인 맞춤형 테스트셋 생성 시스템 블록도

A. 정답률 예측 모델

본 모델은 분류 알고리즘을 유저들의 과거 학습 데이터를 이용하여 최적화한 후 정답 확률을 예측하도록 설계했다. 입력 데이터는 학습자의 개념 이해도, 문제의 난이도, 과거 테스트셋 내에서의 문제의 순번, 평균 문제 풀이 시간으로 이루어져 있으며 정답률을 예측하도록 학습한다. 개념 이해도는 학습자별 해당 문제에 요구되는 개념의 이해도를 0과 1 사이의 실숫값으로 수치화 하였으며 개념의 이해도가 높아질수록 1에 가까워진다. 문제의 난이도 역시 0과 1 사이의 실숫값으로 1에 가까울수록 높은 난이도를 의미한다. 평균 문제 풀이 시간은 각 문제의 전체 유저 학습 데이터를 기반으로 계산되었으며 온라인 환경에서 발생할 수 있는 접속 지연 문제, 디바이스 에러 등의 변수를 고려하여 상 하위 10%를 제외한 평균값을 사용한다. 과거 테스트셋에서 문제의 순번은 n개의 문제로 이루어진 과거의 테스트셋에서 해당 문제가 배열된 순번을 의미한다. 입력 데이터는 인도의 6-10학년 수학, 과학 과목이 대상인 본 연구팀의 학습 서비스 “Class Saathi”의 모의 시험 데이터를 사용하였으며 모의 시험은 학습자들이 시험 범위를 설정하면 해당 범위 내의 테스트셋을 제공한다. 제시된 입력 데이터를 이용하여 Random Forest, Support Vector Machine, K-Nearest Neighbors, Naïve Bayes 분류 알고리즘을 적용하여 성능을 비교하였다. 분류 알고리즘은 확률을 예측하여 50% 이상의 확률을 가지면 1, 반대의 경우 0을 출력하도록 초기 설계되어 있으나 분류 알고리즘의 신뢰도를 높이기 위하여 60% 이상의 확률을 가질 때 1, 반대의 경우 0을 출력하도록 설계한다. 성능 측정 결과, 표 1과 같이 Random Forest 분류 알고리즘을 적용하였을 때 가장 높은 성능을 보이며 정확도는 83.33%로 측정된다. 따라서 본 모델은 Random Forest 분류 알고리즘을 사용하여 정답률을 예측한다.

표 1. 분류 알고리즘의 성능 비교

Metric	Random Forest	SVM	KNN	Naïve Bayes
F1-score	**0.8895**	0.8543	0.8313	0.7103
AUC	**0.7707**	0.5008	0.5673	0.5619
ACC	**0.8333**	0.746	0.7288	0.7103

B. 문제 배열 순서 최적화 모델

본 모델은 학습자의 개념 이해도와 테스트 범위를 기준으로 무작위의 순서로 선별된 n개의 문제를 A에서 제시한 정답률 예측 모델을 사용하여 가장 고득점이 예상되는 문제 배열 순서의 개인 맞춤형 테스트셋을 생성한다. 해당 모델에서는 정답률 예측 모델이 1을 출력할 경우 정답 예상 문제, 0을 출력할 경우 오답 예상 문제로 정의한다. 본 모델은 무작위로 배치된 n개의 문제에 대하여 아래의 순서로 진행된다.

1) 초기 문제 배열 순서는 테스트셋에 임의의 배열 순서를 n/2번 부여하며 그 중 예측 정답률의 평균이 가장 높은 문제 배열 순서를 테스트셋에 부여한다
2) (1)의 과정 후 정답 예상 문제의 순번을 고정하고 나머지 순번을 오답 예상 문제에 무작위로 부여한 후 정답률을 예측한다
3) (2)의 과정을 오답 예상 문제의 수가 최소화될 때까지 반복한다
4) (3)의 과정 후 오답 예상 문제의 개수가 최소 두 개 이상이면 나머지 순번을 무작위로 부여하며 오답 예상 문제 집합의 예측 정답률 평균이 가장 높아지도록 순번을 부여한다

Ⅲ. 실험 및 결과

본 시스템의 성능을 검증하기 위하여 해당 시스템 외 난이도의 오름차순 및 내림차순을 활용한 문제 배열 순서에 관한 연구[4]를 참고하여 아래와 같이 비교군을 정의한다.

1) 무작위로 순번을 배열한 경우
2) 난이도의 오름차순으로 배열한 경우
3) 난이도의 내림차순으로 배열한 경우

총 100개의 테스트셋에 제안 방법을 포함한 네 가지 문제 배열 순서를 적용한 후 정답률 예측 모델을 사용하여 각 테스트셋 별 시험 점수를 예측한다.

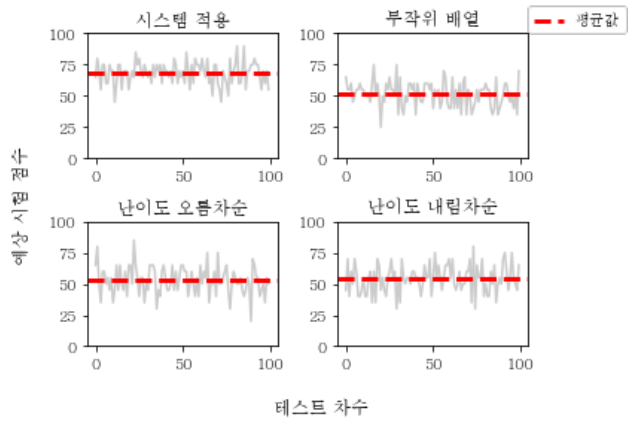


그림 2. 문제 배열 순서 별 예상 시험 점수

실험 결과, 그림 2와 같이 본 시스템을 적용하였을 때 예상 시험 점수의 평균값이 가장 높다는 점을 확인하였다. 문제 배열 순서 별 각 시험 점수의 평균값은 표 2와 같으며 본 시스템의 경우 시스템을 적용하지 않은 경우 대비 평균 29.42% 높은 성능을 확인하였다.

표 2. 문제 배열 순서에 따른 예상 시험 점수 평균

Metric	시스템 적용	무작위 배열	난이도 오름차순	난이도 내림차순
예상 시험 점수 평균	**68.3**	51.4	52.8	54.2

Ⅳ. 결론

본 논문에서는 학습자의 정밀 평가를 개별 문제 풀이 과정이 아닌 난이도가 다양하고 문제 배열 순서의 변수를 가진 시험 상황에서 적용할 수 있는 방법으로 과거 학습 데이터 기반의 정답률 예측 모델과 문제 배열 순서 최적화 모델을 사용한 개인 맞춤형 테스트셋 생성 시스템을 설계하였다. 본 시스템을 통한 개인 맞춤형 테스트셋의 경우 비교군 대비 29.42% 우수한 성능을 확인하였다. 추후 개인 맞춤형 테스트셋의 신뢰도 증가를 위해 정답률 예측 모델의 분류 알고리즘을 개선하여 성능을 향상하고 추가로 개인 맞춤형 테스트셋을 통한 정밀 평가 방식을 연구할 계획이다.

참 고 문 헌

[1] Luan, Hui, et al. "Challenges and future directions of big data and artificial intelligence in education." *Frontiers in psychology* 11 (2020).

[2] 김성훈, et al. "설명 가능한 AI 학습 지원 시스템 개발." *컴퓨터교육학회 논문지* 24.1 (2021): 107-115.

[3] Beck, Joseph E., and Beverly Park Woolf. "High-level student modeling with machine learning." *International Conference on Intelligent Tutoring Systems*. Springer, Berlin, Heidelberg, 2000.

[4] 김재철, 김성훈, and 김선희. "문항배열순서가 변별도와 난이도에 미치는 영향." *교육평가연구* 22 (2009): 409-427.

국민수요반응을 위한 기계학습 예측 모델 기반 고객기준부하 산정 방법

이현용, 고석갑, 이병탁
한국전자통신연구원

{hyunyonglee, softgear, bytelee}@etri.re.kr

A Method for Calculating Customer Baseline Load using Machine Learning Prediction Model for National Demand Response

HyunYong Lee, Seok-Kap Ko, and Byung-Tak Lee
Electronics Telecommunications Research Institute (ETRI)

요 약

본 논문은 국민수요반응을 위한 고객기준부하 산정 방법에 대해서 연구한다. 특히, 국민수요반응 프로그램의 참여자들이 보이는 다양하며 동적인 전력사용 패턴을 잘 반영할 것으로 기대되는 기계학습 예측 모델 기반 고객기준부하 계산 방법을 연구한다. 4 만여 세대의 아파트 전력사용량에 가볍고 빠른 기계학습 예측 모델들을 적용하여, 고객기준부하 계산 측면에서의 각 모델의 타당성을 검증한다.

Ⅰ. 서 론

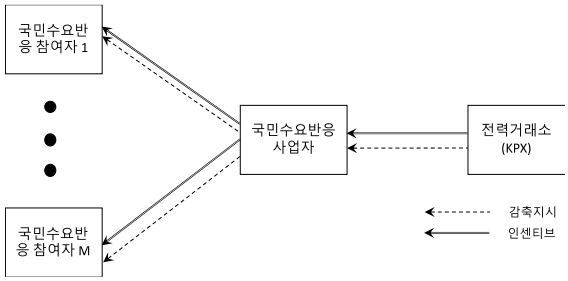
수요반응(Demand Response)이란 전력거래소가 전력 감축 지시를 발령하고 참여자들이 전력 사용량을 감소시키면 전력거래소에서 이에 상응하게 금전적으로 보상해주는 서비스이다 (그림 1). 초기 수요반응은 대규모 공장 중심으로 시작되었으며, 2018 년에는 개별 세대 및 상가 등을 포함하는 소규모 수요자원을 중심으로 하는 국민수요반응이 도입되었다[1][2].

수요반응 프로그램의 참여를 독려하는데 있어 중요한 것은 인센티브이다. 인센티브 계산에 있어 사용되는 것은 고객기준부하(Customer Baseline Load)로 전력거래소의 지시 등으로 전력 소비를 절감하지 않았다면 사용했을 평상시 전력사용량을 예측한 값이다. 따라서, 인센티브는 전력감축 지시가 적용된 시간 내에서의 고객기준부하와 실제 전력 사용량의 차이에 기반하여 계산된다.

초기 수요반응의 대상이었던 대규모 공장 등은 전력사용에 있어 뚜렷한 패턴이 있어서 전통적인 고객기준부하 계산 함수(예. MidXofY 등)를 적용할 수 있었다. 하지만, 전력사용 패턴이 다양한 요소에 의해 영향을 받을 뿐 아니라 매우 동적인 패턴을 보이는 국민수요반응 프로그램의 참여자의 경우에는 전통적인 계산 방법을 그대로 적용할 수 없다. 따라서, 국민수요반응 참여자의 전력사용 패턴을 고려한 고객기준부하 계산 방법이 필요하다.

본 논문에서는 국민수요반응을 위한 고객기준부하 계산 방법을 연구한다. 특히, 여러 기계학습 예측모델의 고객기준부하 계산 성능을 비교함으로써 각 방법의 적용 타당성을 검증하고자 한다.

Ⅱ. 본론

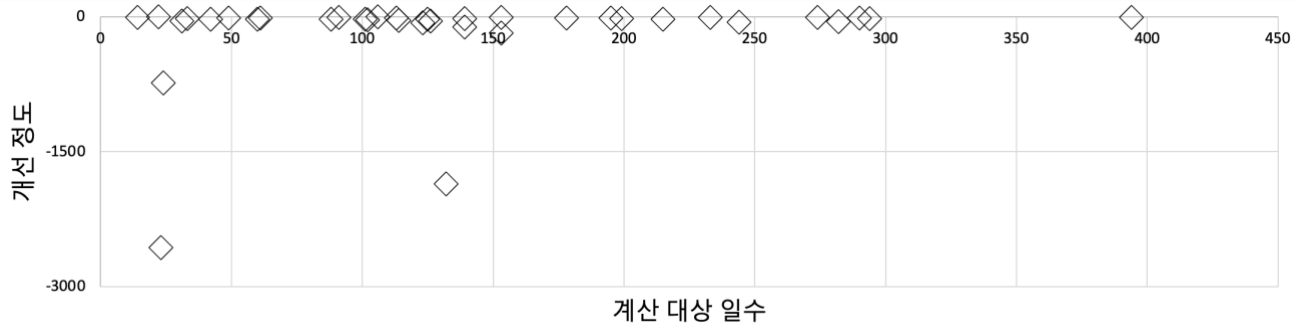


(그림 1) 국민수요반응 개념도.

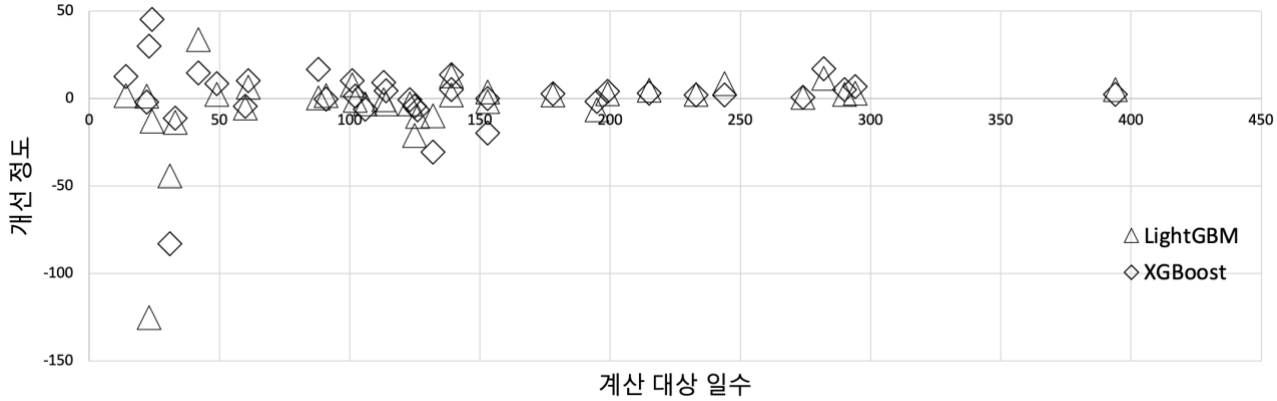
국민수요반응 고객기준부하. 국내에서 서비스 중인 수요반응 프로그램은 표준수요반응, 중소형수요반응, 국민수요반응이다. 국민수요반응을 제외한 두 프로그램에서는 인센티브를 참여자별로 계산한다. 반면, 국민수요반응에서는 일부 참여 세대들을 한 그룹으로 간주하고 그룹 단위로 인센티브를 계산한다. 이는 국민수요반응 참여자들의 전력 사용량이 적기 때문이다. 국민수요반응을 위한 그룹 단위의 고객기준부하 계산에는 Mid8of10 이 사용되고 있다. Mid8of10 은 고객기준부하 계산 대상이 되는 일을 기준으로 공휴일, 국민수요반응 적용일을 제외한 최근 10 일의 전력사용량 중에서 최소, 최대 전력사용량을 보이는 2 일을 제외한 나머지 8 일의 전력사용량을 평균하여 고객기준부하를 도출하는 방식이다.

기계학습 예측 모델. 본 논문에서 고려하는 기계학습 예측 모델은 online ARIMA[3], LightGBM[4], XGBoost 이다. 위 세 모델을 고려하는 이유는, 간단한 통계 기반의 전통적인 고객기준부하 계산 함수를 고려했을 때, 국민수요반응용 고객기준부하 계산 함수 또한 가벼워야 하기 때문이다. 위 세 모델은 GPU 를 필요로 하지 않으며, 빠른 장점이 있다.

데이터. 실험을 위해 36 개의 아파트 단지에 속한 4 만여 세대의 전력사용량 데이터를 사용하였다. 실험의 편의상 동일 아파트 단지에 속한 세대들을 국민수요반응을 위한 한 참여그룹으로 간주하였다. 그룹의 참여 세대 수는 13 부터 4,210 까지 다양하다.



(그림 2). Online ARIMA 에 의한 고객기준부하 계산 성능 개선 정도.



(그림 3). Boosting 알고리즘에 의한 고객기준부하 계산 성능 개선 정도.

데이터는 2016 년 11 월부터 2019 년 11 월까지 획득하였다. 하지만, 세대 별로 누락된 데이터가 있어 실험에 사용가능한 데이터 수는 이보다 적다. 고객기준부하 계산 대상의 시간은, 실제 국민수요반응이 가장 빈번하게 발생한 17 시부터 18 시로 하였다.

실험 결과. 세 가지 기계학습 예측 모델의 고객기준부하 계산 성능은 그림 2 와 그림 3 에 표현하였다. 그림 2 와 그림 3 은 현재 국민수요반응에서 사용되고 있는 Mid8of10 대비하여 실제 전력사용량 예측이 얼마나 더 개선되었는지를 보여준다. 양의 값이면 더 나은 예측 성능을, 음의 값이면 더 나쁜 예측 성능을 의미한다. 참고로, 본 실험에 사용한 데이터는 국민수요반응 이벤트가 미적용된 것으로, 고객기기준부하 계산 함수를 통해 계산된 값을 실제 전력사용량과 비교하여 각 계산 방법의 성능을 용이하게 파악할 수 있다. 그림 2 는 online ARIMA 의 성능을 보여주는데 거의 모든 경우에 저하된 예측 성능을 보인다. Online ARIMA 는 비록 가볍고 빠르지만, 과거 데이터의 유용한 패턴을 잘 활용하지 못하기 때문에, 예측 성능 측면에서는 안좋은 성능을 보인다. 그림 3 은 LightGBM 과 XGBoost 의 예측 성능을 보여준다. 총 36 개의 경우에 대해서, 26 개의 경우에 LightGBM(8 경우) 또는 XGBoost(18 경우)가 더 나은 예측 성능을 보인다. 특히, 계산 대상 일수가 클수록 LightGBM 이나 XGBoost 가 더 나은 성능을 보이는 경우가 더 많아진다. 이는, 활용가능한 학습 데이터가 많은수록 예측 모델의 성능이 증가함을 보여준다. 특히, 특정 계산 일수(199 일)이상의 경우에는 XGboost 나 LightGBM 이 모든 경우에 더 나은 성능을 보인다. 또한, XGBoost 가 LightGBM 보다 더 많은 경우에 성능 향상을 보이는 것을 확인할 수 있다.

Ⅲ. 결론

본논문에서는 국민수요반응을 위한 고객기준부하 계산 측면에서 3 가지 기계학습 예측모델의 타당성을 실제 데이터에 기반하여 검증하였다. 실험 결과 부스팅 알고리즘, 특히 XGBoost 가 기존 Mix8of10 뿐 아니라 다른 모델보다 더 나은 성능을 보이는 것을 확인하였다. 또한, 가용한 학습 데이터가 많을수록 그 성능이 더 증가하는 것을 확인하였다.

ACKNOWLEDGMENT

본 연구는 산업통상자원부(MOTIE)와 한국에너지기술 평가원(KETEP)의 지원을 받아 수행한 연구 과제입니다. (No. 20181210301570)

참 고 문 헌

[1] 정은아, 이경은, 김화영, 정소라, 이효섭, 서봉원, 이원종 “확장된 기술수용모델을 이용한 가정용 에너지 수요반응 프로그램 실증분석”, 한국 HCI 학회 논문지, vol. 12, no.4, pp.65-73, 2017.

[2] 오승훈, 이형옥, 박지원, 이병탁 "국민 DR 서비스를 위한 사물인터넷 표준화 연구", 2020, 대한전자공학회 하계학술대회 논문집

[3] Liu, C., et al. “Online ARIMA algorithms for time series prediction” in Proc. of AAAI, 2016.

[4] Guolin K., et al. "LightGBM: A highly efficient gradient boosting decision tree," in Proc. of NIPS, 2017.

[5] Tianqi X., et al. "XGBoost: A scalable tree boosting system," in Proc. of KDD, 2016.

인공지능 기초 의사결정나무학습을 위한 창의적 교육 프로그램 개발

최은선, 정유진, 현도익, 주연수, 김미진, 박남제*
제주대학교 창의교육거점센터, *제주대학교 교육대학 초등컴퓨터교육전공
{choi910624, yujinjung, pcc762cnd, mjkim, namjepark}@jejunu.ac.kr

Development of a creative education program for learning decision trees, the basis of artificial intelligence

Choi Eunsun, Jung Yujin, Hyun Doik, Joo Yeon-Soo, Kim Mijin, Park Namje*
Creative Education Center, Jeju National University
*Department of Computer Education, Teachers College, Jeju National University

요 약

인공지능은 4차 산업혁명 시대에 접어들면서 그 중요성이 더욱 커지고 있다. 각국에서는 인공지능의 중요성을 역설하며, 차세대 세계를 주도할 수 있는 기술로 인공지능을 주목하고 있다. 그러나, 국내에서는 많은 기업에서 인공지능 인재 부족을 외치고 있으며, 관련 인재 부족으로 기업에 인공지능 기술 도입을 포기하고 있다. 또한, 교육 내용으로서의 인공지능과 방법론에 대한 아직 초기 단계에 머물러 있다. 이에, 본 논문은 이러한 사회적 필요와 요구에 대응하여 학생들이 비교적 이해하기 용이한 의사 결정 나무를 교육 프로그램의 주제로 선정하였으며 다각화된 교수법을 통해 학생들의 창의력을 신장시킬 수 있도록 교육 프로그램을 개발하였다. 본 교육 프로그램은 초등학교 3~6학년을 대상으로 개발되었고, 2015 개정 교육과정의 도덕, 사회, 실과 수업과 연계시킬 수 있다. 더불어, 인문학적 요소를 포함하여 학생들의 문화적 리터러시 향상에 도움이 될 수 있으며, 온라인 워크시트 개발을 통해 원격 수업에서도 본 프로그램을 적용할 수 있다.

Ⅰ. 서 론

인공지능은 4차 산업혁명의 핵심 키워드로 떠오르면서 세상을 변화시킬 주요한 기술이며 핵심 가치로 평가받고 있다. 이에, 교육계에서도 인공지능은 중대한 화두로 떠오르고 있다. 인공지능은 세상과 소통하고 어려운 문제를 해결하기 위한 기술적 접근 방식이며, 오늘날의 학생들이 향후 직업에 관계없이 모든 생활을 하는 동안 접하게 될 기술이다.

미국에서는 인공지능발전협회와 컴퓨터과학교사협회가 주축이 되어 인공지능 교육을 위한 국가교육과정을 마련하고 있다. 또한, 다수의 대학에서는 인공지능 분야에서 연구 개발을 선도하면서 동시에 학생 대상 인공지능 교육 프로그램을 제공하기 위한 연구 프로젝트를 다수 수행하고 있다. 또 다른 인공지능 강대국인 중국에서도 유치원에서부터 인공지능을 접할 수 있도록 인공지능 로봇을 보조 교사로 도입하고 국가 차원에서 다양한 메이커 활동이 포함된 인공지능 교과서를 개발하여 시범적으로 적용하고 있다.

국내의 경우 대부분의 국민들이 인공지능 교육이 필요하다고 지각하고는 있으며, 국가 차원에서도 인공지능 융합 고교를 선정하거나 인공지능 관련 학과를 늘리는 등의 각고의 노력을 하고 있다[1]. 그러나, 인공지능 교육에 대한 구체적인 교수 내용이나 적절한 방법론에 대한 연구가 턱없이 부족한 실정이다. 따라서, 본 연구에서는 인공지능의 여러 기술적 분야 중 인공지능에 대한 기초 지식이 거의 없는 초등 학습자가 비교적 이해하기 쉬운 의사 결정 나무에 대해 학생들에게 흥미롭게 교수할 수 있는 방안을 소개하고자 한다.

Ⅱ. 이론적 배경

의사 결정 나무는 의사 결정 프로세스를 나타내는 표현이다. 인공지능의 의사 결정 나무는 과거의 의사 결정에서 얻을 수 있는 데이터를 기반으로 결론에 도달하는 데 사용된다. 또한 이러한 결론에는 값이 할당되어 향후 취해질 수 있는 조치를 예측할 수 있다. 의사 결정 나무는 다양한 문제와 가능한 결과에 대한 응답을 해석하고 학습하는 통계적, 알고리즘적 기계 학습 모델이다. 결과적으로 의사 결정 나무는 사용 가능한 데이터를 기반으로 특정 컨텍스트에서 의사 결정 규칙을 알고 있다. 학습 과정은 지속적이며 피드백을 기반으로 한다. 이것은 시간이 지남에 따라 학습의 결과를 향상시킨다. 이런 종류의 학습은 지도 학습(supervised learning)이라고 하며, 의사 결정 나무 모델은 지도 학습을 위한 지원 도구로 분류된다[2]. 의사 결정 나무는 사실과 가치에 기반한 과학적 의사결정 과정을 제공하기 때문에 학생들의 인공지능에 대한 기초적 지식 함양 뿐만 아니라, 학습자의 데이터 기반의 논리적 의사결정력을 향상하는데 기여한다.

한편, 창의 교육은 독창적인 사고와 행동의 발달과 관련된 학생의 성장을 촉진하는 새롭고 유용한 방법으로 가르치는 행위를 일컫는다. 창의적 교수는 교사가 학습을 전달하기 위해 사용하는 방법과 이러한 방법이 학생의 산출 결과에 미치는 전반적인 영향에 중점을 둔다. 창의적 교수 과정에서 교사는 학습자로 하여금 학습 자료에 대한 흥미를 불러일으킬 수 있도록 수업을 설계하고, 학습자가 스스로 문제를 창의적으로 찾도록 유도하거나, 특정 문제를 제시하고 학습자에게 모든 종류의 가용 자원을 적용하여 가장 만족스러운 해결책을 찾도록 유도한다. 이러한 과정들은 학습

자에게 새로운 도전을 마주했을 때 다양한 시각으로 문제를 해결할 수 있는 유연성을 제공하며, 논리적이며 분석적인 사고력이 증대된다는 인지적 특성을 가지고 있다[3]. 이에 따라, 본 연구에서는 의사 결정 나무를 인공지능의 교육적 콘텐츠로 창의적 인재를 양성할 수 있는 창의 교육 프로그램을 제안하고자 한다.

Ⅲ. 인공지능 기초 의사 결정 나무 교육 프로그램 개발

본 교육 프로그램은 인공지능 전문가와 교육 전문가들의 협력으로 초등학생의 인공지능 소양 함양을 목적으로 개발되었다. 이를 위하여 프로그램 개발 협의회를 실시하였으며, 문헌 연구를 통해 구체적인 교육 프로그램의 콘텐츠를 선정하고 내용 체계를 정립하였다.

본 논문에서 제안하는 인공지능 기초 의사 결정 나무 교육 프로그램은 초등학교 3학년부터 6학년을 대상으로, 2015 개정 교육과정에서 손쉽게 적용할 수 있도록 개발되었다. 연관된 학년(군)별 교과목 단원과 성취기준은 다음 <표 1>과 같다.

<표 1> 연관된 학년(군)별 교과목 단원과 성취기준

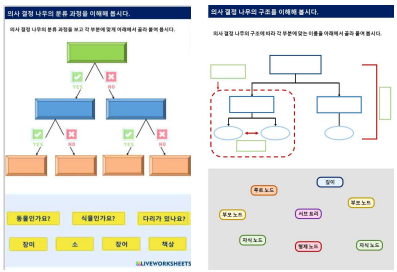
(군)	과목	단원	성취기준
3~4 학년	도덕	사회·공동체와의 관계	[4도03-02] 다문화 사회에서 다양성을 수용해야 하는 이유를 탐구하고, 올바른 의사 결정 과정을 통해 다른 사람과 문화를 공정하게 대하는 태도를 지닌다.
5~6 학년	사회	우리나라의 정치 발전	[6사05-04] 민주적 의사 결정 원리(다수결, 대화와 타협, 소수 의견 존중 등)의 의미와 필요성을 이해하고, 이를 실제 생활 속에서 실천하는 자세를 지닌다.
5~6 학년	실과	기술 시스템	[6실04-07] 소프트웨어가 적용된 사례를 찾아보고 우리 생활에 미치는 영향을 이해한다.

제안하는 교육 프로그램은 총 10차시로 계획되었으나 현장 교사의 필요에 따라 2-3차시로 구분하여 모듈형으로 수업에 적용할 수 있다. 교육 프로그램은 '의사 결정 나무 이해하기'와 '의사 결정 나무 적용하기' 로 크게 두 가지 영역으로 구분하였으며 교육 내용은 첫 번째 영역에서는 지도학습 알고리즘에 대한 전반적 이해를 기반으로 의사 결정 나무의 개념, 구조, 활용 분야, 랜덤 포레스트와의 비교, 설계 기준을 학습하며, 두 번째 영역에서는 의사 결정 나무와 랜덤 포레스트를 활용한 다양성 존중 방법과 설계 원리를 적용하여 타이타닉호의 탑승객을 구하는 수업을 실시한다. 자세한 수업 계획은 다음 <표 2>와 같다.

<표 2> 의사 결정 나무 수업 계획

차시	영역	내용	관련 활동
1	의사 결정 나무 이해 하기	지도학습 알고리즘	언플러그드 활동 온라인학습도구 (라이브워크시트) 활용
2		의사 결정 나무 개념	
3		의사 결정 나무의 구조	
4		의사 결정 나무 활용 분야	
5		의사 결정 나무와 랜덤 포레스트 비교	
6		의사 결정 나무 설계	
7	의사 결정	의사 결정 나무와 랜덤 포레스트를 활용한 다양성 존중 방법	학습 교구 활용
8	나무 적용 하기	의사 결정 나무와 랜덤 포레스트 설계 원리	
9		의사 결정 나무로 타이타닉호를 구하라 1	
10		의사 결정 나무로 타이타닉호를 구하라 2	

본 교육 프로그램은 인공지능 기술에 인문학적 요소를 포함하여 학생들의 문화적 리터러시 함양에도 도움이 될 수 있도록 설계하였는데 소크라테스의 문답법을 의사 결정 나무의 개념과 연계하여 학생들에게 소개하였다. 더불어, 코로나-19로 인한 블렌디드 수업에서도 용이하게 본 교육 프로그램을 적용할 수 있도록 온라인 인터렉티브 북 플랫폼인 라이브워크시트를 통해 온라인 학습지를 개발하였으며, 이는 다음 [그림 1]과 같다.



[그림 1] 의사 결정 나무 학습 온라인 워크시트

Ⅳ. 결론

본 논문에서는 초등학생들의 인공지능 소양 함양을 위하여 창의적인 교육 프로그램 개발한 결과를 나타낸다. 제안된 프로그램은 2015 개정 교육과정의 도덕, 사회, 실과 수업에서 적용할 수 있으며, 현장의 필요에 따라 10차시의 프로그램을 영역별로 쪼개어서 활용할 수 있다. 또한, 블렌디드 교육 현장에서 활용할 수 있도록 온라인 학습지를 개발하여 교사의 편의를 도모하였다. 제안하는 교육 프로그램은 향후 현장 시범 적용과 효과성 분석을 통해 수정과 보완을 거듭할 계획이다.

ACKNOWLEDGMENT

이 논문은 2021년~2023년도 교육부의 재원으로 한국과학창의재단의 지원을 받아 수행된 성과물임.

참 고 문 헌

[1] Ministry of Education, "Basic principles of artificial intelligence, now taught in general high schools," Ministry of Education, 2020, (https://www.moe.go.kr/boardCnts/view.do?boardID=294&boardSeq=79984&lev=0&searchType=null&statusYN=W&page=1&s=moe&m=020402&opType=N).

[2] Arundhati, N., Aamir. N. A., Siddharth. P., and Balwant A. S., "Overview of use of decision tree algorithms in machine learning," 2011 IEEE Control and System Graduate Research Colloquium, pp. 37-42, Jun. 2011.

[3] Eunsun C., Namje P., "Application and Development of Machine Learning Training Program based on Understanding K-NN Algorithm," Journal of the Korean Association of Information Education, pp.175-184, Feb. 2021.

디리클레 분포를 이용한 트레이딩 알고리즘의 앙상블에 관한 연구

정재용, 이주홍*, 최범기**, 송재원***
인하대학교, *인하대학교, **㈜큐햇지, ***㈜밸류파인더스
eoehd4108@naver.com, *juhong@inha.ac.kr, **bgchoi666@gmail.com, ***jwsong@valuefinders.co.kr

A study on ensemble of trading algorithms using Dirichlet Distribution

Jae Yong Jeong, Ju-Hong* Lee, Bum Gi Choi**, Jae Wong Song***
Inha Univ., *Inha Univ., **QHedge Co. Ltd., ***ValueFinders Co. Ltd.

요 약

금융시장에서 알고리즘 트레이딩은 중요한 역할을 한다. 시장의 변동성으로 인해서 알고리즘 트레이딩은 일관적인 결과가 나올 수 없고, 알고리즘마다 우수한 성능이 나오는 부분이 다르게 존재한다. 본 논문에서는 트레이딩 알고리즘의 결과에 대한 우수한 조합을 만들기 위한 앙상블 방법을 제안한다. 알고리즘 트레이딩에 적용하기 위한 다양한 앙상블 기법들이 존재하지만, 일부 앙상블 기법은 적절하지 않을 수 있다. 따라서 이론적으로 최적의 가중치 조합을 찾을 수 있는 BMA 기법을 변형한 앙상블 기법을 제안한다. 본 논문에서는 각 알고리즘의 가중치를 디리클레 분포로 가정하고, 분포로부터 샘플링을 수행하여 Transformer 모델로 알고리즘의 가중치를 학습한다. 실험 결과를 통해, 제안한 앙상블 기법이 알고리즘 트레이딩의 일반화 성능을 높였음을 보였다.

Ⅰ. 서 론

알고리즘 트레이딩이란 수학적 계산과 IT 시스템을 이용해 금융상품을 거래하는 것을 말한다. 2012년 미국에서 전체 거래의 85%를 차지할 정도로 알고리즘 트레이딩은 금융시장에서 중요한 역할을 한다[1]. 하지만, 시장의 변동성으로 인해서 알고리즘 트레이딩은 일관적인 결과가 나올 수 없고[2], 알고리즘마다 우수한 성능이 나오는 부분이 다르게 존재한다[3]. 따라서 알고리즘 트레이딩에 앙상블 기법을 사용하여 모델의 일반화 성능을 높이기 위한 다양한 시도들이 존재한다[2, 3, 4, 5].

[2]는 모델의 출력에 대한 동등한 가중치를 부여하여 조합한다. [3]는 가장 적절한 모델 1개를 선택하는 방식을 사용한다. [4]는 AND 연산자를 이용하여 불확실성에 대응하고자 한다. [5]는 앙상블 기법 중 대표적인 Bagging 기법을 적용한다. 이외에도 적용 가능한 앙상블 기법들을 [6, 7]에서 소개한다. 알고리즘 트레이딩에 적용하기 위한 다양한 앙상블 기법이 존재하지만, [8]는 기존에 알려진 앙상블 기법들이 주식 시장에 적용하기에 적절하지 않을 수 있음을 설명한다.

따라서 본 논문에서는 [9]에서 이론적으로 최적의 조합을 찾을 수 있다고 알려진 BMA(Bayesian Model Averaging)를 변형한 앙상블 기법을 제시한다. 기존의 BMA를 알고리즘 트레이딩에 적용하기 위해 2가지 접근을 사용한다. 첫째, 최근 정보에 집중하기 위해, 감소계수를 추가한다. 앙상블에서 적용될 모델들의 가중치가 시간에 따른 정보를 반영하기 위해서, 모델의 가중치를 디리클레분포로 가정하고, 훈련 구간에서 실제 모델의 성능을 기준으로 그 분포를 수정한다. 이때, 감소계수를 적용하여 가중치가 특정 모델에 지나치게 치우치는 것을 방지하고, 최근 정보에 가중치를 준다. 둘째, 훈련된 분포로부터 모델의 가중치를 샘플링하여 Transformer 모델을 훈련하여 다음에 나타나게 될 분포를 예측한다. 따라서, 특정 시점에 대한 과적합 된 분포를 그대로 사용하지 않을 수 있다. 제안한 앙상블 방법을 암호 화폐를 대상으로 수행하였다. 그 결과로서 앙상블 방법은 암호화폐의 가격 흐름보다 변동성이 매우 적었으며 수익률도 훨씬 높게 나타나서 성능이 매우 우수함을 볼 수 있었다. 따라서 제안한 앙상블 방법의 일반화 성능이 우수함을 볼 수 있다.

이후 사용되는 표기법은 다음과 같다.

K	알고리즘 트레이딩의 개수
w_i	i번째 알고리즘 트레이딩의 가중치 $(0 \leq w_i \leq 1, \sum_{i=1}^{K} w_i = 1)$
α_i	디리클레 분포의 i번째 파라미터
D_t^i	i번째 알고리즘 트레이딩의 t시점 결정 ($M_t^i \in \{-1, 1\}$, -1 매도, 1: 매수)
r_t^i	i번째 알고리즘 트레이딩 t시점 수익률

Ⅱ. 본론

본 논문에서 제안하는 앙상블 방법을 다음 과정을 통해서 진행한다. 앙상블을 수행하기 위한 각 알고리즘 트레이딩의 가중치가 디리클레 분포를 따른다고 가정한다.

$$W \sim Dir(\alpha_1, \alpha_2, \ldots, \alpha_K) \quad (1)$$

이때, 초기 $\alpha_i = 1 (i = 1, 2, \ldots, K)$로 초기값을 설정한다.

알고리즘 트레이딩의 앙상블 $D_t^{i,j,k}$를 D_t^i, D_t^j, D_t^k의 voting으로 얻은 결정이라고 하자. K개의 알고리즘 트레이딩의 n개의 앙상블은 ${}_K C_n$개의 조합이 나오게 된다.

${}_K C_n$개의 앙상블 조합으로부터 상위 m 개의 알고리즘 트레이딩을 찾는다. 이때, 알고리즘 트레이딩의 성능은 샤프지수(수식 2.)를 기준으로 평가한다.

$$Sharpe\ ratio = \frac{mean(r)}{stddev(r)} \quad (2)$$

상위 m 개의 i번째 알고리즘 트레이딩의 성능이 상위권이었던 횟수를 α'_i라고 하자. 가중치(W)를 알고리즘 트레이딩이 상위 m개의 모델에 포함되어 있을 확률의 의미로 해석하면, 다항분포의 모양으로 모델링할 수 있다.

$$P(\alpha' \mid W) = \frac{m!}{\alpha_1'!\alpha_2'!\ldots\alpha_K'!} w_1^{\alpha_1'} w_2^{\alpha_2'} \ldots w_K^{\alpha_K'} \quad (3)$$

수식 (1)은 (3)의 conjugate prior이기 때문에 베이지안 정리를 이용해

$$W \sim Dir(\alpha_1+\alpha_1', \alpha_2+\alpha_2', \ldots, \alpha_K+\alpha_K') \quad (4)$$

로 업데이트가 가능하다. 이때, 특정 모델에 가중치가 지나치게 집중되는 것을 막고 최근 정보에 대한 가중치를 주기 위하여 감소계수(c)를 적용한 업데이트를 한다.

$$W \sim Dir(\lfloor \alpha_1/c \rfloor + \alpha_1', \lfloor \alpha_2/c \rfloor + \alpha_2', \ldots, \lfloor \alpha_K/c \rfloor + \alpha_K') \quad (5)$$

따라서 특정 시점(t)의 가중치 분포 W_t를 찾아낼 수 있다. 하지만, 알고리즘 트레이딩의 앙상블을 수행하기 위해서는 미래 시점(t')의 가중치 분포 $W_{t'}$를 알아야 하는데, 이는 알 수 없는 분포이다. 이를 보완하기 위해서 [10]에서 제안한 Transformer 모델을 사용하여 미래 시점의 분포를 예측한다. Transformer 모델을 훈련하기 위해서 과거 시점들의 데이터 분포로부터 H번의 랜덤 샘플링을 아래와 같이 수행하고,

$$x_{(p)}^{(t)} \overset{iid}{\sim} Dir(W_t; \alpha) \; (p = 1, 2, \ldots, H)$$

훈련을 위한 데이터 셋을 아래와 같이 정의한다.

$$X_1 : (x_{(1)}^{(t-5)}, x_{(1)}^{(t-4)}, x_{(1)}^{(t-3)}, x_{(1)}^{(t-2)}, x_{(1)}^{(t-1)})$$
$$Y_1 : x_{(1)}^{(t)}$$
$$input\,data = (X_1, X_2, \ldots, X_H)$$
$$ouput\,data = (Y_1, Y_2, \ldots, Y_H)$$

이때, 목적함수는 크로스-엔트로피를 사용하여서 두 분포 사이의 차이를 줄여준다. 따라서, 미래 시점(t')의 데이터 분포는 이전 시점의 샘플링 $(x^{(t'-5)}, x^{(t'-4)}, x^{(t'-3)}, x^{(t'-2)}, x^{(t'-1)})$으로부터 추론 할 수 있다. 이때 추론된 값들을 $\widehat{x_p^{(t')}} \; (p = 1, 2, \ldots, H)$로 표현이 가능하다. 추론 된 값들의 평균($\bar{\hat{x}}$)을 미래 시점(t')의 알고리즘 트레이딩의 가중치로 사용할 수 있다. 즉, 아래와 같이 알고리즘 트레이딩의 앙상블을 수행한다.

$$D_{t'}^{ensemble} = \bar{\hat{x}}_i D_{t'}^i$$

위에서 제안한 앙상블 방법을 암호 화폐(비트코인)으로 실험한 결과는 (그림 1.)과 같이 나타난다. 실험 데이터로서 2020년 8월 23일부터 2021년 8월 4일까지의 비트코인의 30분 단위 시계열을 사용하였다. x축은 시간을 의미하고, y축은 시간에 따른 수익률을 의미한다. 파란색으로 표시된 비트코인의 경우의 변동성이 제안된 앙상블 방법의 변동성보다 훨씬 큰 것을 알 수 있고, 제안된 앙상블 방법의 최종 수익률이 더 높음을 볼 수 있다.



(그림 1.) 제안한 앙상블 기법을 적용한 결과

Ⅲ. 결론

본 논문에서는 시계열 데이터의 특성을 고려한 알고리즘 트레이딩의 앙상블 기법을 제안하였다. 각 알고리즘 트레이딩의 가중치를 디리클레 분포로 가정하고, 일정 시간 동안의 알고리즘 트레이딩 성능과 감소계수를 적용하여 분포를 업데이트한다. 또한 관측할 수 없는 미래 시점의 분포를 예측하기 위해서, 과거 시점에서의 디리클레 분포에서 데이터를 샘플링하여 Transformer 모델로 학습한다. 실제 트레이딩의 결과(그림1.)로서 제안된 앙상블의 성능이 모든 구간에서 꾸준히 유지되는 것을 볼 수 있다. 이를 통해 알고리즘 트레이딩의 일반화 성능이 우수함을 볼 수 있다.

ACKNOWLEDGMENT

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (2019R1F1A1062094, 2017R1D1A1A02018319, NRF-2020R1F1A1069361)

참 고 문 헌

[1] 안명호, 류미현 (2018) 머신러닝을 이용한 알고리즘 트레이딩 개발. 한빛미디어

[2] Ullah, A. K. M., et al. "Combining machine learning classifiers for stock trading with effective feature extraction." arXiv preprint arXiv:2107.13148 (2021).

[3] Yang, Hongyang, et al. "Deep reinforcement learning for automated stock trading: An ensemble strategy." Available at SSRN (2020).

[4] Giacomel, Felipe, Renata Galante, and Adriano Pereira. "An algorithmic trading agent based on a neural network ensemble: a case of study in North American and Brazilian stock markets." 2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT). Vol. 2. IEEE, 2015.

[5] Sun, Shaolong, Shouyang Wang, and Yunjie Wei. "A new ensemble deep learning approach for exchange rates forecasting and trading." Advanced Engineering Informatics 46 (2020): 101160.

[6] Sagi, Omer, and Lior Rokach. "Ensemble learning: A survey." Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 8.4 (2018): e1249.

[7] Huang, Faliang, Guoqing Xie, and Ruliang Xiao. "Research on ensemble learning." 2009 International Conference on Artificial Intelligence and Computational Intelligence. Vol. 3. IEEE, 2009.

[8] Jacobsen, Ben, Fuwei Jiang, and Hongwei Zhang. "Ensemble machine learning and stock return predictability." The 4th International Workshop on Financial Markets and Nonlinear Dynamics (Paris. 2019.

[9] Domingos, Pedro. "Bayesian averaging of classifiers and the overfitting problem." ICML. Vol. 747. 2000.

[10] Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems. 2017.

제한된 프런트홀의 셀탈피 MIMO 시스템을 위한 최적화 학습 기법

유대성, *이훈, ‡홍승은, 박석환
전북대학교, *부경대학교, ‡한국전자통신연구원
imcreative93@jbnu.ac.kr, hlee@pknu.ac.kr, iptvguru@etri.re.kr, seokhwan@jbnu.ac.kr

Learning-to-Optimize Approach for Fronthaul-Constrained Cell-Free MIMO Systems

Daesung Yu, *Hoon Lee, ‡Seung-Eun Hong, Seok-Hwan Park
Jeonbuk National Univ., *Pukyong National Univ.,
‡Electronics and Telecommunications Research Institute

요약

본 논문은 셀탈피 MIMO 시스템에서의 빔포밍 및 프런트홀 통신의 연동 최적화를 위한 딥러닝 적용 알고리즘을 연구한다. 기존 딥러닝 적용 기술을 검토한 후, 인공신경망의 입력으로 사용되는 여러 환경 변수 효과를 비교 분석한다.

Ⅰ. 서 론

셀탈피 (cell-free) MIMO 시스템은 기지국 안테나들의 커버리지 구분 없는 신호처리를 통해 효율적인 간섭 제어 및 스케줄링이 가능하다[1]. 이는 유사한 구조를 가진 기존 분산 안테나 시스템 및 클라우드 RAN 시스템에 관한 연구에서 충분히 검증된 바 있다 [2].

한편 최근 셀탈피 시스템의 빔포밍 및 프런트홀 통신 연동 최적화 문제의 해를 효율적으로 찾기 위해 딥러닝 기반의 최적화 학습 (learning-to-optimize) 기법이 [3]에서 제안되었다. 특히 랜덤한 환경 변수 (송신 전력, 프런트홀 용량)를 인공신경망 입력으로 활용하여 학습할 것을 제안했는데, 이 변수들의 입력 효과는 자세히 관찰된 바가 없다. 본 논문에서는 모의실험을 통해 인공신경망의 입력에 따른 효과를 관찰 및 분석한다.

Ⅱ. 본론

한 개의 central processor (CP), M 개의 access point (AP), K 개의 user equipment (UE)로 구성된 셀탈피 MIMO 시스템의 하향링크를 고려한다. 각 AP 및 UE 는 단일 안테나를 사용하며, 각 AP 와 CP 를 연결하는 디지털 프런트홀 링크의 용량은 C bps/Hz 이다. 각 AP 의 송신 전력은 P 로 제한된다. 해당 시스템의 빔포밍 및 프런트홀 통신 최적화를 위해 딥러닝 기반 최적화 학습 기술이 제안된 바 있다[4]. 환경 변수인 C 와 P 를 인공신경망의 추가 입력으로 활용하였다. 그러나 해당 변수들이 인공신경망의 입력으로 꼭 필요한지에 대한 논의는 생략됐다. 이에 본 논문은 그림 1 과 같이 입력 환경 변수에 따른 여러 경우에 대한 실험을 수행하여 그 성능들을 비교한다.

Ⅲ. 모의실험 결과

그림 2 에 도시한 실험 결과에 따르면, 환경 변수 C, P 를 모두 입력에서 제외할 경우 심각한 성능 손실이 발생한다. 그러나 둘 중 한 개 이상 입력으로 사용하면 의미 있는 성능 이득을 얻는다. 두 변수 중 P 의 입력 효과가 C 보다 더 큰 것을 확인할 수 있다. P 의 입력 범위가 더 넓기 때문인 것으로 추정된다.

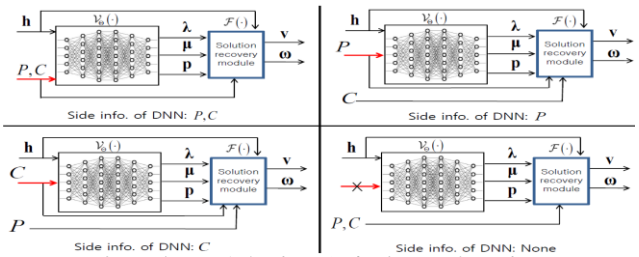


그림 1. 인공신경망 력 변수에 따른 4 개 조합

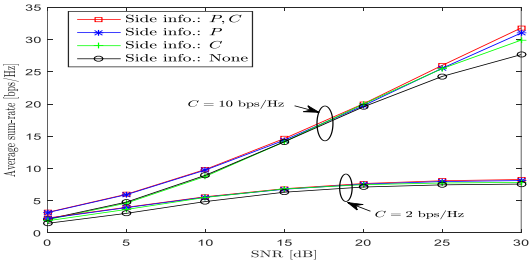


그림 2. 인공신경망 입력 변수에 따른 성능 관찰

ACKNOWLEDGMENT

This work was supported in part by the NRF of Korea Grants funded by the Ministry of Education under Grant 2021R1A6A3A13046157; in part by the Ministry of Science and ICT under Grant 2021R1C1C1006557 and Grant 2021R1I1A3054575; in part by the Institute of Information & communications Technology Planning & Evaluation (IITP) Grant funded by the Korea Government (MSIT) under Grant 2018-0-01659; and in part by the 5G Open Intelligence-Defined RAN (ID-RAN) Technique Based on 5G New Radio.

참 고 문 헌

[1] E. Nayebi, A. Ashikhmin, T. L. Marzetta, H. Yang, B. D. Rao, “Precoding and power optimization in cell-free massive MIMO systems,” *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, Jul. 2017.

[2] S.-H. Park, O. Simeone, O. Sahin, S. Shamai, “Joint precoding and multivariate backhaul compression for the downlink of cloud radio access networks,” *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5646-5658, Nov. 2013.

[3] D. Yu, H. Lee, S.-H. Park, S.-E. Hong, “Deep learning methods for joint optimization of beamforming and fronthaul quantization in cloud radio access networks,” arXiv:2107.02520, Jul. 2021.

톰슨 샘플링 기반 동적 프레임 슬롯 알로하

조영제, 황경호*
국립한밭대학교 컴퓨터공학과
yeongjejo@edu.hanbat.ac.kr, *gabriel@hanbat.ac.kr

Dynamic Framed Slotted ALOHA based on Thompson Sampling

Yeong-Je Jo, Gyung-Ho Hwang*
Hanbat National University, Dept. Computer Engineering

요 약

본 논문에서는 Q-learning 기반의 Slotted ALOHA 프로토콜인 ALOHA-Q 방식에서 노드의 수와 프레임 당 슬롯의 수가 차이가 있을 때 throughput 성능이 낮아지는 점을 보완하기 위하여 Slotted ALOHA에 톰슨 샘플링 강화학습 기법을 적용하여 샘플링 값에 따라 프레임 내의 슬롯 수를 조절하는 방안을 제안하였다. 슬롯에서의 패킷 충돌 확률과 빈 슬롯 발생 확률을 모두 감소시켜 Slotted ALOHA, ALOHA-Q 방식 대비 높은 throughput을 보여주는 것을 시뮬레이션을 통한 성능 분석으로 확인하였다.

Ⅰ. 서 론

Slotted ALOHA에 Q-learning 기반 강화학습을 적용한 ALOHA-Q 방식은 각각의 노드들이 프레임 내의 전송 성공률이 높은 슬롯에 패킷을 전송하게 하여 Slotted ALOHA 대비 성능을 높였지만 프레임 내의 슬롯의 개수가 노드 수보다 많으면 빈 슬롯이 생기고, 슬롯의 개수가 노드 수보다 적으면 슬롯에서 패킷 충돌 확률이 높아져 throughput 성능이 저하되는 문제가 있다.

본 논문에서는 Slotted ALOHA에 톰슨 샘플링 기반의 강화학습을 적용하여 노드들이 패킷 전송 성공 확률이 높은 슬롯에 패킷을 전송할 수 있도록 학습할 뿐만 아니라, 노드의 수에 따라 프레임 내의 슬롯 수를 조절하여 빈 슬롯을 줄이고 패킷 충돌 확률을 감소시켜 성능을 높이는 방식을 제안한다.

Ⅱ. 본론

2.1 Dynamic Framed Slotted ALOHA (DFSA)

Framed Slotted ALOHA에서는 프레임을 여러 개의 슬롯으로 나누어 슬롯의 시작점에서만 패킷을 전송할 수 있게 하여 Pure ALOHA 대비 성능을 높일 수 있지만, 프레임당 슬롯의 수를 고정으로 할당하여 슬롯 개수 대비 노드 개수가 많거나 작아질수록 성능이 떨어졌다. DFSA는 이러한 문제를 보완하여 프레임 내의 슬롯 수를 조절해가며 성능을 높이는 방식이다. Vogt 알고리즘은 프레임 내에서 충돌이 발생한 슬롯 수를 이용하여 다음 프레임 크기를 조절하는 알고리즘으로 슬롯에 충돌이 발생하면 해당 슬롯에는 최소한 2개 이상의 노드가 전송을 시도한 것을 의미하므로 충돌이 발생한 슬롯 수에 2를 곱한 만큼 다음 프레임 크기를 증가시킨다.[1]

DFSA 방식에서는 프레임의 크기 조절은 가능하지만, 노드가 랜덤하게 슬롯을 선택하여 성능의 한계가 존재한다.

2.2 ALOHA-Q

ALOHA-Q 방식은 Slotted ALOHA에 Q-learning 기반 강화학습을 적용하여 이전 프레임에서 선택한 슬롯에 대해 패킷 전송의 보상 값에 따라 노드들이 프레임 내의 슬롯 중 전송 성공률이 가장 높은 슬롯을 학습하는 방식이다. 이상적인 경우 다수의 노드가 서로 다른 슬롯에 패킷을 전송하여 패킷 충돌이나 빈 슬롯이 발생할 확률을 낮추어 성능을 높인다. 하지만 프레임 내의 슬롯 수가 고정이므로 노드 수와 차이가 날수록 성능이 낮아진다.[2][3]

2.3 톰슨 샘플링 강화학습 기반의 DFSA

톰슨 샘플링은 (1)과 같은 베타 분포를 이용한 알고리즘으로 α값과 β값을 이용하여 action에 대한 확률분포를 만들어 무작위로 값을 샘플링하여 그중 가장 큰 값을 가지는 action을 정한다. $\Gamma()$는 감마 함수를 의미하며 θ는 action에 대한 보상을 받을 확률을 나타내며 1-θ는 보상을 받지 못할 확률을 나타낸 값이다. Action에 대한 보상을 받게 되면 α값이 증가하여 분포가 1에 근접하게 되어 샘플링 값이 높을 확률이 높아지고 보상을 받지 못하면 β값이 증가하면서 분포가 0에 근접하게 되어 샘플링 값이 낮을 확률이 높아지게 된다.[4]

$$p(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1} \quad (1)$$

본 논문에서는 Slotted ALOHA에 톰슨 샘플링을 적용한 DFSA 방식을 제안한다. 노드들은 각자 프레임 내의 모든 슬롯에 대해 베타 분포를 가지고 α는 해당 슬롯에 패킷 전송을 성공한 횟수를 나타내고 β는 패킷 전송을 실패한 횟수를 나타낸다. <그림 1>은 두 슬롯에 대한 노드의 베타 분포를 표현한 예시다. 패킷 전송에 성공하면 α값이 증가하여 해당 슬롯에 대한 베타분포의 확률밀도함수의 값이 θ가 1에 가까워질 때 높아지고,

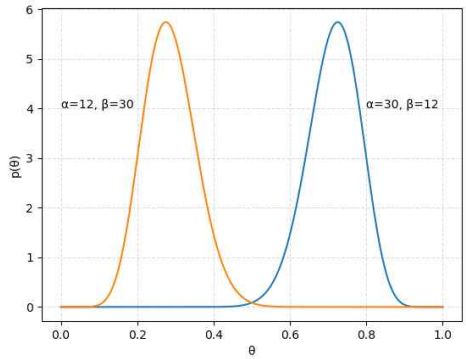


그림 1. α, β값에 따른 베타분포

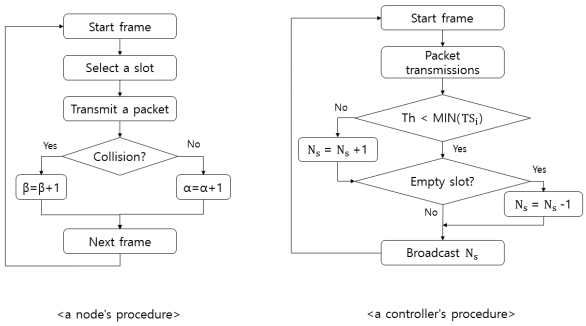


그림 2. 제안하는 톰슨 샘플링 강화학습 기반 Dynamic Framed Slotted ALOHA 동작

패킷 전송에 실패하면 β값이 증가하여 θ가 0에 가까울 때 확률밀도함수가 높아진다. 슬롯에 패킷을 전송하지 않으면 분포에는 변화가 생기지 않는다. 프레임이 진행될수록 베타 분포의 확률밀도함수의 폭이 좁아져 전송 성공확률이 높은 슬롯의 샘플링 값은 1에 가까워지고, 나머지 슬롯에 대한 샘플링 값은 0에 가까운 값이 되어 전송 성공 확률이 높은 슬롯을 선택할 가능성이 높아진다.

<그림 2>는 제안한 방안의 한 프레임 동안의 동작을 나타낸다. 모든 노드는 프레임 시작 시 톰슨 샘플링 알고리즘을 이용하여 각 슬롯의 샘플링 값을 구하고 그 중 가장 큰 샘플링 값의 슬롯을 선택하여 패킷을 전송하고 성공 여부에 따라 α, β값을 업데이트한다. 프레임의 크기를 조절하는 컨트롤러는 노드들이 선택한 슬롯에 대한 샘플링값(TS_i, 노드 i가 선택한 슬롯의 톰슨 샘플링 값)들의 최솟값이 미리 설정한 임계값 Th보다 낮으면 전송 성공률이 높은 슬롯을 선택할 수 없을 확률이 높으므로 프레임의 슬롯 수를 의미하는 N_S를 1 증가시킨다. 반대로 노드들의 샘플링 값이 모두 Th보다 높지만 빈 슬롯이 발생하면 불필요한 슬롯이 존재할 확률이 높으므로 N_S를 1 감소시킨다.

2.4. 시뮬레이션을 통한 성능 분석

본 논문에서 제안하는 방식의 성능을 분석하기 위해 Python을 이용하여 시뮬레이션을 진행하였고 임계값 Th는 0.5로 설정하였다.

<그림 3>은 노드 수에 따른 에피소드별 슬롯 수를 나타낸 것이다. 에피소드는 한 프레임 동안의 접속을 의미하며 첫 에피소드에서 프레임의 크기는 10개 슬롯으로 설정하였다. 에피소드가 증가함에 따라 노드 개수에 맞게 프레임의 슬롯 수가 조절되는 것을 볼 수 있다.

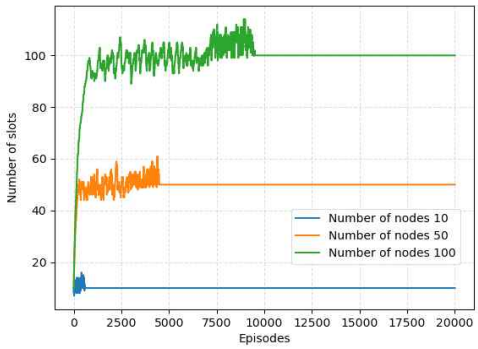


그림 3. 에피소드별 한 프레임의 슬롯 수

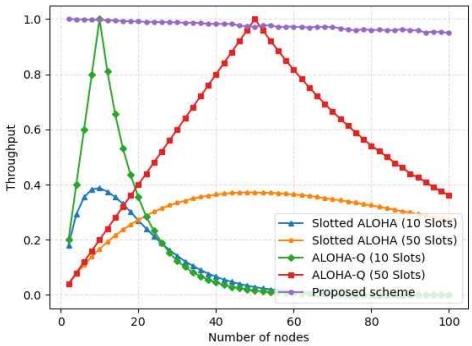


그림 4. 노드 수에 따른 throughput 비교

<그림 4>는 본 논문에서 제안하는 방식과 다른 방식의 성능을 비교한 것으로 100,000 에피소드 동안의 평균 throughput을 나타낸다. Throughput은 시뮬레이션 동안의 모든 슬롯 개수 대비 패킷 전송을 성공한 슬롯 개수의 비율이다. Slotted ALOHA, ALOHA-Q 방식 모두 프레임당 슬롯의 개수를 고정해 슬롯 수와 노드 수의 차이가 클수록 throughput이 감소하는 걸 확인하였고 본 논문에서 제안한 방식은 톰슨 샘플링 강화학습을 사용하여 프레임의 슬롯 개수를 동적으로 변화시키는 과정에서 노드의 수가 증가할수록 최적의 슬롯 개수를 찾는데 더 많은 에피소드가 필요하여 throughput이 점차 감소하지만 모든 경우에서 throughput이 1.0에 가까운 성능을 보여주었다.

Ⅲ. 결론

본 논문에서는 Slotted ALOHA에 톰슨 샘플링 강화학습을 적용한 DFSA 방식을 제안하였다. 톰슨 샘플링 값에 따른 프레임의 슬롯 개수를 동적으로 조절하고, 강화학습을 통해 전송 성공확률이 높은 슬롯을 결정하여 Slotted ALOHA와 ALOHA-Q 방식 대비 throughput 성능이 향상하였다.

참 고 문 헌

[1] H. Vogt, “Efficient object identification with passive RFID tags,” in Proc. Int. Conf. Pervasive Computing, pp. 98-113, 2002.

[2] R. S. Sutton and A. G. Barto, *Reinforcement Learning An Introduction*, MIT press, 2018.

[3] 이창규, 이승형, “강화학습 기반 Slotted ALOHA의 성능개선”, 한국통신학회논문지, vol.45, no.11, pp.1886-1892, 2020.

[4] D. J. Russo, B. V. Roy, A. Kazerouni, I. Osband and Z. Wen, “A Tutorial on Thompson Sampling,” Foundations and Trends in Machine Learning, vol.11, no.1, pp.1-96, 2018.

OFDM 시스템에서 딥 러닝 기반 변조 식별

*신정완, *송건호, *장민규, *윤동원†
*한양대학교 융합전자공학과
†dwyoon@hanyang.ac.kr

Deep Learning-based Modulation Identification in OFDM Systems

*Jeongwan Shin, *Geonho Song, *Mingyu Jang, and *Dongweon Yoon
*Department of Electronic Engineering, Hanyang University

요 약

자동 변조 분류는 무선 인지, 스펙트럼 감시 등의 비협력 통신 상황에서 가장 중요한 기술 중 하나이다. 본 논문에서는 직교 주파수 분할 다중화(OFDM, orthogonal frequency division multiplexing) 시스템에서 딥 러닝 기반으로 변조 방식을 자동으로 분류하고 식별 성능을 분석한다. OFDM 신호의 변조 분류를 위해 합성곱 신경망을 기반으로 변조 식별 신경망 구조를 설계하고 모의 실험을 통해 기존 논문 보다 향상된 변조 식별 정확도 성능을 가짐을 보인다.

Ⅰ. 서론

직교 주파수 분할 다중화(OFDM, orthogonal frequency division multiplexing)는 높은 스펙트럼 효율과 페이딩, 간섭 등의 완화 효과 때문에 무선 통신 시스템에서 중요한 기술 중 하나로 알려져 있다. 또한 무선 통신 기술이 빠르게 발전함에 따라 다양한 분야에서 비협력 통신이 보편화 되고 있다. 비협력 통신 상황에서는 송신 신호에 대한 변조 방식을 수신부에서 알지 못한다. 따라서 수신 신호로부터 정보를 얻기 위해 수신부는 수신 신호만을 이용하여 송신부에서 사용한 변조 방식을 추정해야 한다. 이에 따라 비협력 통신 상황에서는 수신 신호만을 이용하여 변조 방식을 분류하는 자동 변조 분류(AMC, automatic modulation classification)기술이 필요하다.

전형적인 AMC 기술은 우도 기반 AMC 와 특징 값 기반 AMC 가 있다[1], [2]. 우도 기반 AMC 는 최적의 성능을 제공하지만 높은 복잡도로 인하여 활용에는 한계가 있으며 특징 값 기반 AMC 는 변조 분류를 위한 특징 값을 선정해야 한다는 문제점이 있다. 이를 극복하기 위해 최근에는 LSTM, ResNet 과 같은 다양한 딥 러닝 기법에 기반한 AMC 연구가 활발히 진행되고 있다[3]-[8].

본 논문에서는 OFDM 시스템에서 합성곱 신경망(CNN, convolutional neural network) 기반의 AMC 방법을 제안하고 분류 성능을 분석한다. 이를 위하여 수신 신호 샘플을 이용해 CNN 모델을 훈련하고 훈련된 모델을 이용하여 수신 OFDM 신호의 변조 방식을 분류한다. 모의 실험을 통해 제안한 방법의 성능을 분석하고 제안한 방법이 기존 방법보다 우수한 변조 식별 정확도를 가짐을 보인다.

Ⅱ. CNN 기반 AMC

본 논문에서는 비협력 통신 상황에서 송신 OFDM 신호가 페이딩 채널을 통과하였다고 가정한다. 이 때 수신 신호는 다음과 같이 나타낼 수 있다.

$$r(k) = x(k) * h(k) + n(k) \quad (1)$$

여기서 $r(k), x(k), h(k), n(k)$는 각각 수신 신호, 송신 신호, 페이딩 채널의 임펄스 응답, 가산성 백색 가우스 잡음을 나타낸다.

수신 신호의 IQ (in-phase and quadrature phase) 성분을 분리하면 다음과 같은 수신 신호 IQ 배열을 얻을 수 있다.

$$R(k) = \begin{bmatrix} Re\{r(k)\} \\ Im\{r(k)\} \end{bmatrix} \quad (2)$$

여기서 $Re\{\}$와 $Im\{\}$은 각각 실수부와 허수부를 나타낸다. 본 논문에서는 (2)와 같은 형태의 배열을 CNN 구조의 입력으로 사용한다.

그림 1 은 본 논문에서 사용하는 CNN 구조를 나타낸다. 제안하는 CNN 구조는 입력층, 2 개의 합성곱층, 3 개의 완전 연결층, 그리고 출력층으로 구성되어 있다. 입력층에서는 (2)와 같은 형태의 배열을 이용하여 2 x 1,920 크기의 행렬을 한 샘플로 입력한다. 첫 번째 합성곱층에는 128 개의 합성곱 커널이 있고 합성곱 커널 행렬의 크기는 1 × 320 이다. 두 번째 합성곱층은 크기가 2 × 8 인 합성곱 커널 64 개가 있다. 완전 연결층의 뉴런의 수는 각각 256, 128, 4 이다. 출력층에서는 변조 분류 결과가 출력된다. 마지막 완전 연결층에서는 확률 분포 값을

얻기 위해 Softmax 함수를 활성화 함수로 사용하였으며 이를 제외한 나머지 완전 연결층과 합성곱 층에서는 활성화 함수로 PReLU 함수를 사용하였다. 또한 과적합을 방지하고 더 안정적인 학습을 위해 배치 정규화와 드롭아웃을 적용하였다.

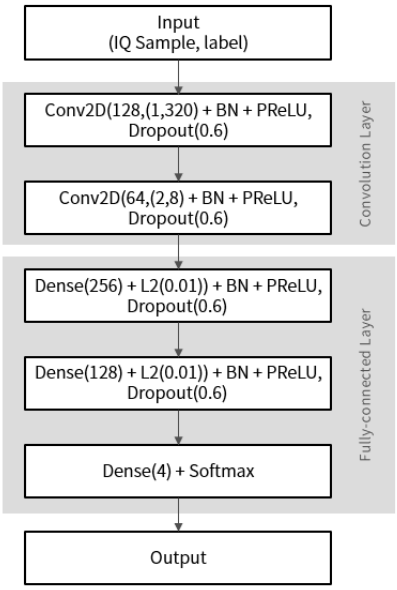


그림 1. AMC 를 위한 CNN 구조

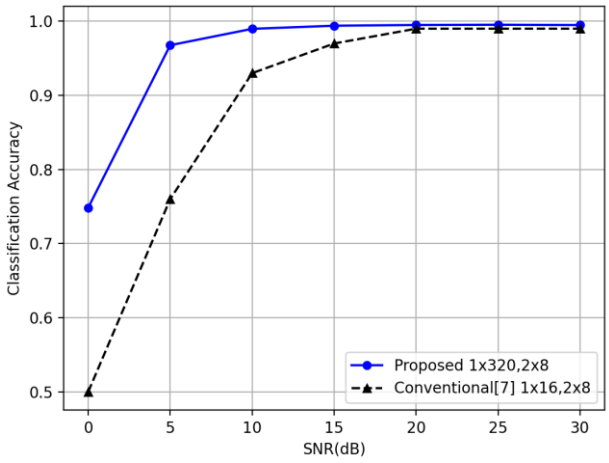


그림 2. 변조 식별 정확도

Ⅲ. 모의 실험 및 성능 분석

제안한 AMC 방법의 변조 식별 정확도 성능을 분석하기 위하여 다음과 같은 모의 실험을 진행하였다. 모의 실험에서는 CNN 구현을 위해 Keras, Tensorflow 2.4 와 Python 3.7 을 백엔드로 사용하였으며, 신경망을 최적화하기 위해 모멘텀을 포함한 확률적 경사 하강법을 채택하였다. 100 번의 epoch 에 대해 CPU Intel Xeon E5-2690 v4 6 개와 NVIDIA Tesla V100 1 GPU 가 포함된 플랫폼을 사용하였고 약 40 분의 훈련시간이 소요된다. 데이터셋은 Matlab 으로 생성하였고, 디지털 선형 변조 중 4 가지의 변조 방식(BPSK, QPSK, 8PSK, 16QAM)이 고려되었다. OFDM 신호는 기존 논문 [7]과의 비교를 위해 부 반송파 16 개, 푸리에 변환 크기 256, cyclic prefix 길이 64 로 가정하였다. 채널은 라이시안 페이딩 채널, 도플러의 편이 주파수 500hz, Rician K-factor 20 을 가정하였다. 신호대 잡음비 (SNR)는 -5~30dB 사이에서 무작위로 생성하였다. 각 변조 방식마다 17,000 개의 데이터 샘플이 있고, 각각의 데이터는 14,000/3,000 으로 트레이닝 및 검증 샘플로 나뉜다.

모의 실험 결과에 따른 변조 분류 정확도는 그림 2 와 같다. 그림 2 에서 볼 수 있듯이 제안한 AMC 방법은 변조 식별 정확도 90%를 기준으로 기존 논문[7] 대비 약 5dB 우수한 성능을 나타내는 것을 확인할 수 있다.

Ⅳ. 결론

본 논문에서는 OFDM 시스템에서 CNN 기반의 AMC 방법을 제안하고 분류 성능을 분석하였다. 모의 실험을 통해 제안한 AMC 방법이 기존 논문에 비해 우수한 변조 식별 성능을 나타내는 것을 확인하였다. 향후에는 다양한 채널 환경과 변조 방식을 고려하여 AMC 연구가 진행되어야 할 것이다.

참 고 문 헌

[1] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su, “A Survey of automatic modulation classification techniques: classical approaches and new trends,” *IET Commun.*, vol. 1, pp. 137-156, 2007.

[2] H.-C. Wu, M. Saquib, and Z. Yun, ‘‘Novel automatic modulation classification using cumulant features for communications via multipath channels,’’ *IEEE Trans. Wireless Commun.*, vol. 7, no. 8, pp. 3098-3105, Aug. 2008

[3] S. Kim and D. Yoon, “Automatic modulation classification in practical wireless channels,” in *Proc. ICTC*, Jeju, Korea, 2016, pp. 915-917.

[4] J. Lee, J. Kim, B. Kim, D. Yoon, and J. Choi, “Robust automatic modulation classification technique for fading channels via deep neural network,” *Entropy*, vol. 19, no. 9, pp. 454, Aug. 2017.

[5] J. Lee, B. Kim, J. Kim, D. Yoon, and J. Choi, “Deep neural network-based blind modulation classification for fading channels,” in *Proc. ICTC*, Jeju, Korea, 2017, pp. 552-555.

[6] Y. Wang, J. Yang, and G. Gui, “Data-driven deep learning for automatic modulation recognition in cognitive radios,” *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 4074-4077, Apr. 2019.

[7] S. Hong, Y. Zhang, Y. Wang, H. Gu, G. Gui, and H. Sari, “Deep learning-based signal modulation identification in OFDM systems,” *IEEE Access*, vol. 7, pp. 114631-114638, 2019.

[8] J. Shi, S. Hong, C. Cai, Y. Wang, H. Huang, and G. Gui, “Deep learning-based automatic modulation recognition method in the presence of phase offset,” *IEEE Access*, vol. 8, pp. 42841-42847, 2020.

RaPP 기반 네트워크 이상탐지 시스템

김미르, 계효선, 권민혜
숭실대학교

{48rlaalfm, ksh8301000}@soongsil.ac.kr, minhae@ssu.ac.kr

RaPP-based Network Anomaly Detection Systems

Miru Kim, Hyoseon Kye, Minhae Kwon
Soongsil University

요 약

네트워크 상의 트래픽이 증가함에 따라 네트워크 침입 탐지 시스템(Network Intrusion Detection System; NIDS)의 필요성이 대두되고 있다. 본 논문에서는 오토인코더(Autoencoder; AE)의 은닉층의 정보를 활용하여 기존의 오토인코더 방식보다 더 높은 성능으로 이상을 탐지할 수 있는 RaPP(Reconstruction along Projection Pathway)기반의 이상탐지 시스템을 제안한다. 오토인코더의 은닉층의 정보를 효과적으로 표현할 수 있도록 SAP (Simple Aggregation along Pathway)와 NAP(Normalize Aggregation along Pathway)를 활용하였고 기존의 오토인코더 방식과 성능을 비교하였다. 그 결과 RaPP 기반의 NAP 방식이 기존의 오토인코더 방식보다도 우수한 성능을 보임을 확인하였다.

Ⅰ. 서 론

네트워크 통신 기술의 발달로 네트워크 통신을 이용하는 기기의 종류와 양이 기하급수적으로 증가하면서 네트워크 트래픽 또한 급증하고 있다. 이에 따라 네트워크상에서의 악의적인 침입 시도의 방법과 빈도가 계속해서 늘어나고 있다. 따라서 공격을 사전에 탐지하는 네트워크 침입 탐지 시스템의 필요성이 대두되고 있다.

네트워크 침입 탐지 시스템은 기존의 알려진 공격 데이터의 패턴을 이용해서 침입을 탐지하는 오용탐지와 보편적인 정상 데이터의 패턴을 이용해서 침입을 탐지하는 이상탐지로 나누어진다[1,2,3]. 오용탐지는 사전에 알려진 공격 유형만을 탐지하는 방식으로 새로운 공격 유형에 즉각적인 대응을 하기 어려운 방식이다. 그러나 네트워크 트래픽 상에서 네트워크 공격 유형이 점차 지능화되고 다양화되고 있어 이에 대한 대비가 필수적이다. 이상탐지는 기존의 공격 유형뿐만 아니라 새로운 공격 유형을 탐지할 수 있어 현재의 문제점에 적합한 탐지 방식이기에 이에 대한 연구가 활발히 진행되고 있다.

머신러닝 기반의 대표적인 이상탐지 모델인 오토인코더[2,4]는 인코더(encoder)에서 입력데이터의 차원을 효과적으로 축소하고, 디코더(decoder)에서 축소된 데이터의 차원을 복원한다. 이렇게 복원된 데이터와 입력데이터 사이의 오차를 복원오차(reconstruction error)라고 정의 한다. 정상 데이터의 복원오차는 학습을 통해 작은 값을 가지게 되므로 오토인코더는 설정한 임계점보다 큰 복원오차를 가지는 데이터를 비정상 데이터로 탐지하여 이상탐지를 진행한다.

본 논문에서는 학습된 오토인코더의 성능을 효과적으로 향상시킬 수 있도록 은닉층의 정보를 활용하는 RaPP[5]를 사용함으로써 기존의 네트워크 침입 탐지 시스템의 성능을 향상시키는 새로운 시스템을 제안한다.

Ⅱ. RaPP 기반 이상탐지 시스템

오토인코더는 차원을 축소하는 인코더 부분과 축소된 차원을 다시 복원하는 디코더 부분으로 구성된다. 인코더는 g로 정의하며 l개의 층(layer)으로 구성되어 있다.

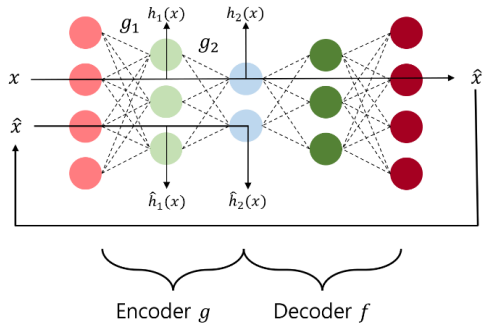


그림 1. RaPP 구조

또한 인코더의 $i\ (1 \le i \le l)$ 번째 은닉층의 축소과정은 g_i, 출력은 $h_i(x)$로 정의하고 디코더는 인코더와 동일하게 l개의 층으로 구성되어 있으며 f로 정의한다. 이렇게 정의된 오토인코더에 대해서 입력 데이터 x의 복원 데이터는 $\hat{x}$으로 표현한다.

$$h_i(x) = g_i \circ \cdots \circ g_1(x) \quad (1)$$

$$\hat{x} = f \circ h_l(x) \quad (2)$$

RaPP 기반 이상탐지 시스템은 복원 데이터 $\hat{x}$을 다시 인코더에 입력시켜 인코더의 i 번째 층의 출력 $\hat{h}_i(x)$ 을 확인한다.

$$\hat{h}_i(x) = h_i(\hat{x}) = h_i \circ f \circ h_l(x) \quad (3)$$

따라서, 입력 데이터 x와 복원 데이터 $\hat{x}$ 의 복원 오차만을 기반으로 이상탐지를 수행하는 기존의 오토인코더 방식과는 달리, RaPP 기반의 방식은 $h_i(x)$와 $\hat{h}_i(x)$의 차이를 비교한다. RaPP 구조는 그림 1 에 설명하였다. RaPP 기반 이상탐지 시스템에서는 은닉층의 복원오차를 확인하기 위해 SAP 와 NAP 적용한다. 입력데이터 x를 $h_0(x)$, 복원 데이터 $\hat{x}$를 $\hat{h}_0(x)$로 표현하면 기존의 오터인코더 방식(AE)과 SAP, NAP 는 (4), (5), (6)와 같은 수식으로 나타낼 수 있다.

표 1. Test set 데이터의 성능 측정 결과

	Pre-cision	Re-call	Fall out	Speci-ficity	Accu-arcy	F1-score	MCC
AE	0.91	0.99	0.09	0.90	0.94	0.95	0.90
SAP	0.83	0.99	0.20	0.79	0.89	0.90	0.82
NAP	0.97	0.99	0.02	0.97	0.98	0.99	0.97

$$s_{AE} = \left\| h_0(x) - \hat{h}_0(x) \right\|_2^2 \quad (4)$$

$$s_{SAP} = \sum_{i=0}^{l} \left\| h_i(x) - \hat{h}_i(x) \right\|_2^2 \quad (5)$$

$$s_{NAP} = \sum_{i=0}^{l} \left\| (h_i(x) - \hat{h}_i(x))^T V \Sigma^{-1} \right\|_2^2 \quad (6)$$

$$M = U \Sigma V^T \quad (7)$$

AE 와 SAP 방식은 복원오차 값에 L2 norm 을 적용하였다. SAP 방식에서 각 은닉층의 복원 오차 분포의 스케일이 상이할 경우, 분포의 스케일을 고려하지 않고 단순히 원점으로부터의 거리를 구하는 것은 모델의 정확도를 떨어뜨리게 되므로 이상탐지의 성능을 저하시킬 수 있다. 이에 SAP 방식의 문제점을 보완하기 위해 NAP 는 특이값 분해(Singular Value Decomposition; SVD)를 사용하여 은닉층의 복원오차 값을 정규화한 후에 L2 norm 을 적용하였다. SVD 는 $m \times n$ 행렬 M을 U, Σ, V의 3 개의 행렬로 분해할 수 있는 방식이며 (7)과 같은 수식으로 나타낼 수 있다. U는 좌특이벡터로 정의되며 V는 우특이벡터로 정의된다. U와 V는 기하학적으로 선형변환의 의미를 가지는 직교행렬(orthogonal matrix) 이고, Σ는 기하학적으로 스케일 변환을 의미하는 대각행렬(diagonal matrix)이다. NAP 는 은닉층의 복원오차 $(h_i(x) - \hat{h}_i(x))^T$를 SVD 하여 얻은 V로 선형변환 시킨 뒤에 Σ로 스케일을 변환하여 은닉층의 복원오차 분포 스케일을 하나의 스케일로 통일한다. 이에 따라 SAP 보다 높은 성능으로 이상탐지를 수행할 수 있다.

Ⅲ. 이상탐지 성능 평가 결과 비교

본 논문에서 정의한 SAP, NAP 의 성능평가를 위하여 시뮬레이션을 통해 AE 와 성능비교를 진행하였다. 네트워크 이상탐지 시스템의 성능을 평가하기 위해 Precision(정밀도), Recall(재현율), Fall-out, Specificity (특이도), Accuracy(정확도), F1-score, MCC(Matthews Correlation Coefficient)의 7 가지 성능지표를 사용하였다. 본 실험에서 사용한 데이터셋은 네트워크 이상데이터를 포함하고 있는 NSL-KDD 데이터셋[6]을 사용했으며, 10 개의 랜덤시드에 대한 성능결과의 평균값으로 성능 값을 구하였다. 훈련과 테스트에 모두 사용되는 정상데이터는 6:2:2 의 비율로 train set, validation set, test set 으로 구분하였다. Test set 은 학습과정에서는 사용하지 않고 성능평가를 위해서만 사용하였으며, test set 에 포함되는 비정상 데이터의 양은 정상 데이터와 동일하도록 설정하였다. 성능평가 결과는 표 1 에 정리하였다.

SAP 를 사용하였을 때는 기존 AE 방식에 비해 저하된 성능을 보였다. 이는 RaPP 의 은닉층의 복원오차가 다양한 스케일을 가지게 되기 때문이다. 반면, 은닉층의 복원오차에 정규화를 수행한 NAP 방식은 모든 성능 지표에서 다른 두 방식 대비 우수한 것을 확인하였다. 특히, F1-score 는 0.99 로 매우 높은 이상탐지 성능을 가짐을 확인하였다. 또한, Specificity 와 MCC 는 AE 대비약 8%의성능향상이 이루어짐을 확인하였다. 그림 2 는 test set 데이터의 NAP 기반 복원 오차 값 분포를 확인한 것이다. 이를 통해 정상과

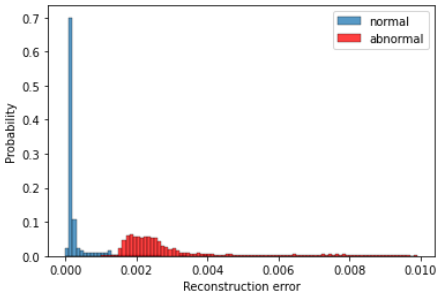


그림 2. NAP 방식의 복원오차 분석

비정상 데이터가 서로 다른 분포를 가지고 명확하게 구분되는 것을 확인할 수 있다. 정상데이터의 경우 성공적인 복원으로 복원오차 값이 0 에 가깝게 분포하였다. 반면, 비정상 데이터의 경우 복원오차 값이 정상데이터에 비해 큰 값인 0.002 를 중심으로 분포한다.

Ⅳ. 결론

본 논문에서는 은닉층의 정보를 활용하여 수행하는 RaPP 기반의 네트워크 이상탐지 시스템을 제안하였다. 은닉층의 정보를 활용하지 않는 기존 오토인코더 방식과, 복원오차의 SAP, NAP 의 성능 검증 결과, 기존 방식 대비 은닉층의 복원오차를 정규화 하지 않은 SAP 방식은 오히려 성능이 저하되는 반면, NAP 방식은 나머지 두 성능에 비해 우수한 성능임을 입증하였다. 따라서 NAP 방식을 적용하여 우수한 성능의 네트워크 이상탐지 시스템 구축을 기대할 수 있다.

ACKNOWLEDGMENT

이 논문은 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT 연구센터지원사업(IITP- 2021-2020-0-01602)과 방송통신산업기술개발사업(IITP-2021-0-00739)의 지원을 받아 수행된 연구임.

참 고 문 헌

[1] O. Depren, M. Topallar, E. Anarim, and M. K. Ciliz, “An Intelligent Intrusion Detection System for Anomaly and Misuse Detection in Computer Networks,” Expert Systems with Applications, vol. 29, no. 4, pp. 713-722, Nov. 2005.

[2] 김승주, 계효선, 권민혜, “오토인코더 기반 네트워크 침입탐지 시스템,” 한국통신학회 추계종합학술발표회, Nov. 2020.

[3] 계효선, 권민혜, “주성분 분석기반 저복잡 이상탐지 기술 연구,” 한국통신학회 논문지, vol. 46, no. 6, pp. 941-955, 2021 년 6 월.

[4] C. Zhou, and R.C. Paffenroth, “Anomaly Detection with Robust Deep Autoencoders,” ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Aug. 2017.

[5] K. H. Kim, S. Shim, Y. Lim, J. Jeon, J. Choi, B. Kim, and A. S. Yoon, "RaPP: Novelty Detection with Reconstruction along Projection Pathway," International Conference on Learning Representations, Sep. 2019.

[6] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, “A Detailed Analysis of the KDD CUP 99 Data Set,” IEEE Symposium on Computational Intelligence for Security and Defense Applications, Jul. 2009.

Nature-Inspired Microstrip Antennas For High-Directivity: Fractals and Genetics

Yejune Seo[1], Munsu Jeon[1], Junghyun Cho[1], Changhyeong Lee[2],
Jaume Anguera[3,4], Sungtek Kahng[1]
[1]Dept. of Info. & Telecomm. Engineering, Incheon National University, Incheon, South Korea
[2]Center for Advanced Meta-Materials, Daejeon, South Korea
[3]Technology Department, Fractus, Barcelona, Spain
[4]Electronics and Communications Department, Universitat Ramon LLull, Barcelona, Spain

Abstract

High directivity patch antennas with broadside directivity have practical interest since a narrow beam can be obtained without using an array of antennas. The solution becomes simpler because there is no need of using a feeding network. Two examples of high-directive microstrip antennas synthesized by fractals and genetic algorithm optimization are compared. Both simulated and measured results show the potential of fractals and genetic algorithms to design non-intuitive antennas having broadside and high-directivity performance with a single microstrip element. Said elements are useful for hot-spots.

Index Terms—fractals, genetic algorithm, optimization, high-directivity, microstrip antennas.

Ⅰ. INTRODUCTION

Achieving high-directivity by using a single element microstrip patch antenna is a challenging task, as conventional patches have moderate directivity of about 6-7 dBi [1]. A classical patch, such as a square patch, presents a broadside pattern at its fundamental mode (TM10 or TM01), with a low directivity. A broadside pattern can also be found at a higher order mode (TM30). However, the directivity is again limited, as the radiation pattern presents high secondary lobes. The classic method to increase directivity is to form an array of antennas. However, having as an objective to obtain a simple antenna structure, an array needs a feeding network which on one hand increases the mechanical complexity and on the other hand, the spurious radiation from said network may distort the desired radiation pattern. In order to avoid an array of antennas, various methods to improve the directivity of microstrip antennas have been reported in the literature, such as fractal-inspired antennas [3]-[5] or electromagnetic band gap, superstrates, zero-index metamaterial, modified Peano space-filling curve. In this sense, the present paper compares high-directivity microstrip antennas, where fractal geometry and GA is employed on radiating elements with same size.

The paper is divided as follows: in section II the basics on fractals-geometry the GA algorithm are presented. In section III, a high-directive microstrip antenna using fractal geometry and GA are compared. Finally, in section IV, conclusions are drawn.

Ⅱ. ON FRACTAL AND GENETIC ANTENNAS

Fractal-shapes have been useful to design small, multiband, and high directivity antennas. In particular, thanks to the space-filling properties, examples of small antennas have appeared in the literature. Also, thanks to the self-similarity property, multiband and high-directivity antennas have been designed.

In the present paper, the fractal Sierpinski triangle is used and compare with a GA design (Fig. 1). This element present at a higher order mode with a directivity larger than the directivity of the fundamental mode [3]. Considering a microstrip structure of h=1.52mm, printed on a ε_r=3.38 substrate, a bow-tie patch of 40 mm x 40 mm results in a 6.3dB at the fundamental mode resonating at 1.128GHz. This bow-tie patch is considered the zero iteration of the Sierpinski-patch. In contrast, for the Sierpinski-patch of 3 iterations (Fig. 1), the directivity obtained at a higher-order mode (3.86GHz) increases to 10.9dB (Fig. 2). The reason of this high-directivity is that the current distribution presents a self-similar behavior, i.e, repeated portions of in-phase radiation along the structure. On one hand, this in-phase portions results in a broadside-pattern. On the other hand, since the antenna is operating a high-order mode, the antenna is electrically larger than the antenna size when operating in the fundamental mode. Therefore, a larger directivity is obtained.

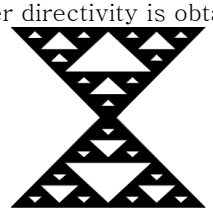


Figure 1 Sierpisnki bowtie of three iterations: it is based on the Sierpinski triangle applied to a bowtie

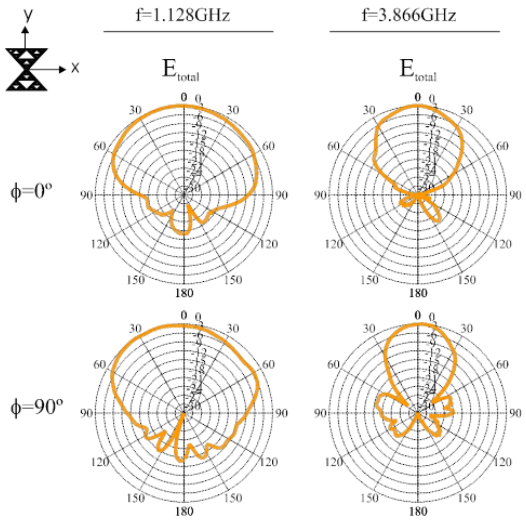


Figure 2 A comparison of the measured radiation pattern for the fundamental mode of the Sierpisnki-shaped antenna and for the higher-order mode

Regarding the genetic-inspired antenna, the proposed design is simulated in the High Frequency Structure Simulator (HFSS) environment in combination with a home-made GA code. The GA operation is written using Visual Basic Script (VBS) Writer and the .VBS file is called into HFSS environment to perform simulations. In order to compare with the fractal-shaped antenna, the same area of 40 mm x 40 mm is used as a reference as well as the same substrate.

The patch area is divided into small cells so as to overlap between adjacent cells. The conducting or non-conducting property of each cell is defined using binary encoding. If a cell is conducting, then the corresponding gene is assigned "1" and if a cell is non-conducting it is assigned "0". The purpose of the overlap between adjacent cells is to avoid having cells contacting by an infinitesimal point which may pose a connection problem when manufacturing the microstrip patch due to the tolerances of the chemical etching (Fig. 3).

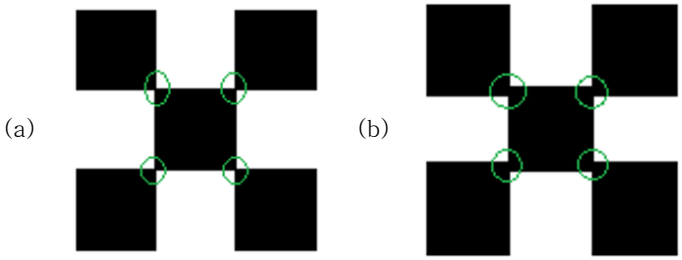


Figure 3 a) Traditional on/off building block with infinitesimal connections. b) Proposed scheme using overlapping based on a shifting of the cell following the vertical axis

The GA scheme is based on the cross-over and mutation operators (Fig. 4). The optimization is finished when the fitness presents convergence.

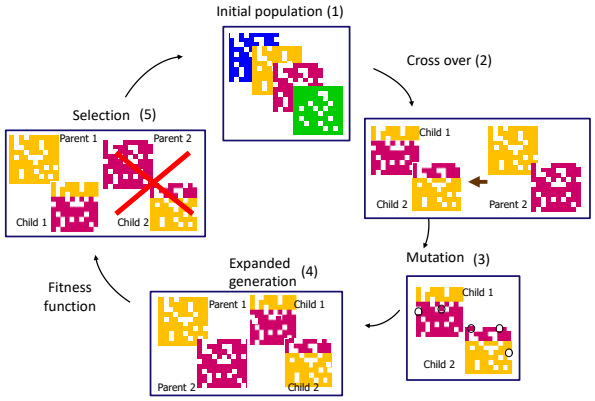


Figure 4Initial population consists on a random population; cross-over operator is used to create new individuals; elitism mutation is applied to ensure a globlal optimization; the fitness function selects the best individual to be on top of the population. Conductor is represented by colours whereas non-conductive portion is represented by white colour

The optimization process is described as above.

Ⅲ. CONCLUSIONS

Microstrip antennas with high-directivity characteristics designed using a single-fed point are interesting since there is no need to include a feeding network as is the case of microstrip arrays. In this sense, the fractal and GA designs presented in this paper, feature high-directivity including patterns with narrow beam at both planes. These designs are of engineering application for hot-spots.

ACKNOWLEDGMENT

This work was supported by the Priority Research Centers Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2020R1A6A1A03041954).

REFERENCES

[1] C. Balanis, Antenna theory and design, 2nd ed., Willey & Sons, 1997

[2] C. Borja et al, "High directivity fractal boundary microstrip patch antenna", IEE Electronics Letters, vol.36, nº9, pp.778-779, 2000.

[3] J. Angueraet et al, "Small and High Directivity Bowtie Patch Antenna Based on the Sierpinski Fractal", Microw. and Opt. Tech. Lett., Vol.31, nº.3, pp. 239-241, November 2001.

[4] J. Romeu, C. Borja, S. Blanch, "High Directivity Modes in the Koch Island Fractal Patch Antenna", IEEE Antennas and Propagation Symposium, Utah, USA, pp. 1696-1699, 2000

[5] J. Anguera et al, "Metallized Foams for Fractal-Shaped Microstrip Antennas", IEEE Antennas and Propagation Magazine, vol.50, nº6, Dec. 2008, pp.20-38.

제2회 한국 인공지능 학술대회

The 2nd Korea Artificial Intelligence Conference

세션 B

Sparse CNN을 이용한 위조 안저영상 포렌식 모델에 관한 연구

송호중, 한주혁, 김웅식*
건양대학교, *건양대학교
20806608@konyang.ac.kr, 21856503@konyang.ac.kr, *wskim@konyang.ac.kr

A Study on the Forensic Model of Forged Fundus Image Using Sparse CNN

Song Ho-Jung, Han Ju-Hyuck, Kim Woong-Sik*
Konyang Univ., *Konyang Univ.

요 약

본 논문은 최근 대두되는 의료이미지를 활용한 이미지 재생성 기술의 악용을 방지하기 위한 연구로 의료인들이 구분하기 어려울 정도의 재생성 이미지를 인공지능 모델로 구분해냈다. 우리는 의료이미지 중, 안저 이미지를 기반으로 연구했다. 데이터는 3개의 Kaggle 안저 이미지 데이터 셋을 사용했으며, 이를 전처리하여 위변조 데이터 셋을 구성하기 위한 위변조 모델에 입력으로 사용했다. 위변조 모델은 U-Net과 Cycle GAN모델을 사용하여 구성했으며, 생성된 데이터는 3명의 안과전문의와 2명의 일반의들로부터 검증을 받았다. 본 연구에서는 4개의 병변(정상, 녹내장, 당뇨성 망막병증, 황반변성)을 대상으로 원본 안저 이미지와 조작된 안저 이미지를 Sparse CNN에 기반으로 하는 포렌식 모델에 입력하여 위변조 유무를 판단한다. 본 논문의 모델 성능은 동일한 테스트 데이터를 5명의 의료진과 포렌식 모델이 그룹별 옵저버 테스트로 진행했다. 이는 포렌식 모델의 성능을 의료진과 비교하기 위한 방법이며, 그 결과 포렌식 모델의 성능이 전문의 그룹의 테스트 결과보다 전체적으로 약 10% 더 높은 검출률을 보였다. 따라서 포렌식 모델은 환자의 정보, 임상실험, 사회보험제도, 등 의료 데이터가 필요한 여러 영역에서 위변조 이미지로 인해 발생하는 문제를 사전에 검출 할 수 있다.

Ⅰ. 서 론

최근 인공지능 기술은 이미지 분야에서 많은 발전을 이루고 있다. 특히 이미지 재생성 기술은 실제와 구분하기 어려울정도의 유사한 이미지를 생성할 수 있다. 이는 뛰어난 기술적 발전이지만, 악용 시에 큰 피해가 예상된다. 특히 의료분야에서는 조작된 이미지가 환자의 생명에 직접적인 영향을 끼칠 수 있다. 또한 임상연구 조작, 사회보험 조작과 같이 사회적인 2차 피해를 발생시킬 수 있어, 그 피해가 치명적이다.[1] 따라서 본 논문에서는 여러 의료이미지 중, 안저 이미지를 대상으로 위조된 이미지를 검출해 낼 수 있는 인공지능 모델을 연구했다.

Ⅱ. 본론

본 연구에서는 인공지능 기반의 안저 이미지 포렌식 모델을 구성하기 위해, 실제 데이터와 이를 기반으로 조작된 위변조 데이터를 Kaggle에서 수집 및 위변조 하였다. 본 연구의 데이터 셋은 Kaggle의 Shanggong Medical Technology Co, Ltd.에서 수집한 Ocular Disease Recognition 데이터 셋을 사용했다.[2] 이는 5,000명의 환자를 대상으로 수집한 안저 이미지 데이터로 이미지 정보 및 진단 Label이 포함된다. 본 연구에서는 데이터 셋에서 수집된 약 6,000개(환자 1인당, 중복수집) 데이터에서 우측 안구 이미지만을 사용했으며, 이미지 품질을 고려하여 데이터를 분류하여 진행했다. 본 연구에 사용된 Shanggong 데이터 셋은 총 356개의 이미지이며, 정상(Normal) 114개, 녹내장(Glaucoma) 67개, 당뇨성 망막병증(Diabetic Retinopathy) 78개, 황반변성(Macular Degeneration) 97개로 구성된다. 또한 다양한 소스와 환경을 포괄하는 포렌식 모델을 구성하기 위하여, Kaggle의 MESSIDOR-2 DR Grades와 Glaucoma Detection 데이터 셋을 함께 사용했다.[3][4] 이 데이터 셋은 당뇨성 망막병증과 녹내장의 안저 이미지와 진단 Label을 포함한다.

본 연구에서는 데이터 위변조를 위하여, Cycle GAN과 U-Net을 사용하여 위변조 모델을 구성했다. 위변조 모델에 입력되는 데이터는 앞서 이야기한 데이터 셋에 이미지 크기 조정 및 전처리 과정을 진행하여 사용했다. 전처리 과정은 원본 데이터에 Sharp Filter를 적용했으며, 필터는 Gaussian Filter로 구성했다. 이는 원본 이미지의 크기를 축소하는 과정에서 발생하는 혈관 및 병변의 뭉개짐(Blur) 현상을 극복하기 위하여 사용했다. 또한 이미지의 크기 축소는 데이터 셋마다 상이한 크기를 맞추고 위변조 모델의 효율성을 증가시키는 역할을 한다.

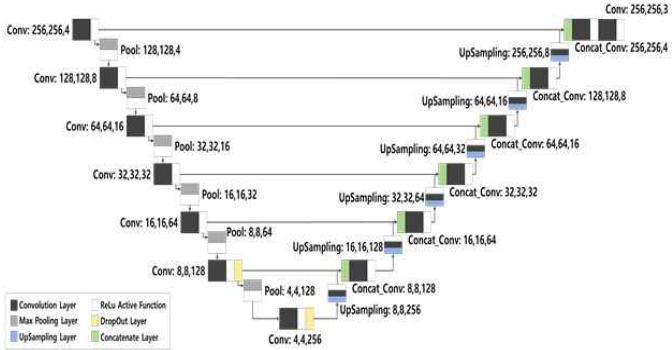


그림 1. 위변조 모델의 U-Net 구성

그림 1은 위변조 모델 중 U-Net 형태로 구성된 모델의 구성도이다. 본 모델은 원본 데이터를 위변조 시키기 위한 모델이며, CNN(Convolutional Neural Network)를 중심으로 구성되어 있다. 모델의 입력은 전처리된 데이터를 랜덤 샘플링하여 사용했으며, 입력의 채널은 4채널로 기본적인 3채널(RGB)과 혈관 Segmentation 채널을 흑백으로 구성했다.

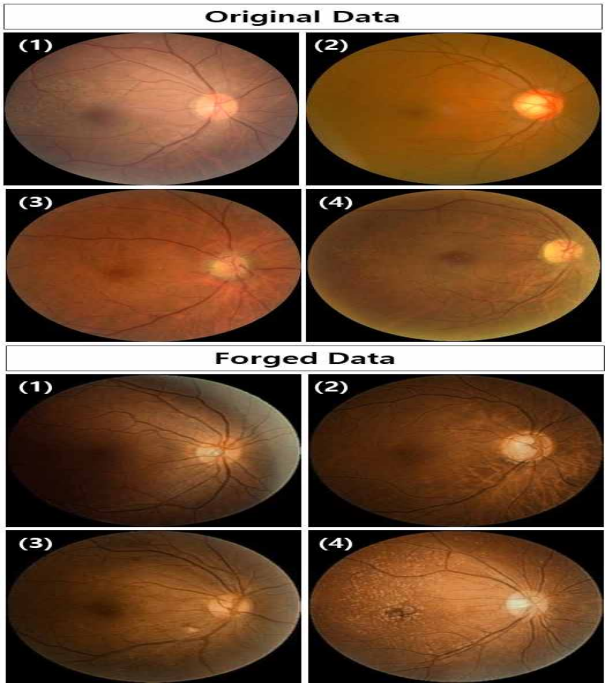


그림 2. 원본 및 위변조 이미지 비교

그림 2는 상단의 원본 이미지와 하단의 위변조 이미지를 비교한 그림이다. 위변조 이미지는 상단의 위조 이미지를 기반으로 만들어졌으며, 각 번호는 (1) 정상, (2) 녹내장, (3) 당뇨성 망막병증, (4) 황반변성이다. 본 연구에서는 위변조 된 데이터와 실제 데이터의 비교를 위하여 위변조 데이터를 생성했으며, 생성된 데이터를 검증하기 위해 안과전문의 3명(Specialist Group)과 일반의 2명(Doctor Group)에게 검증을 요청했다. 또한 포렌식 모델과의 성능비교를 진행했다.

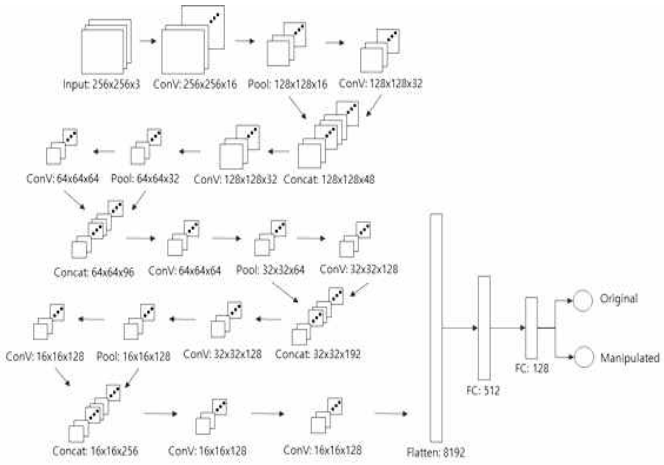


그림 3. 포렌식 모델 구성

그림 3은 실제 이미지와 위변조 이미지를 판단하기 위한 인공지능 모델이며, 모델 구성은 Sparse CNN의 형태로 구성했다.[5] 이는 최초 입력 데이터의 특징을 추출한 Feature Map이 CNN이 계산됨에 따라서 초기 값에 대한 가중치를 계산하지 못하는 문제를 해결하기 위해 사용했다. 또한 포렌식 모델은 2진 분류로 원본과 위변조를 판단하기 위해서, 병변별로 동일한 모델이 학습, 테스트 되었다.

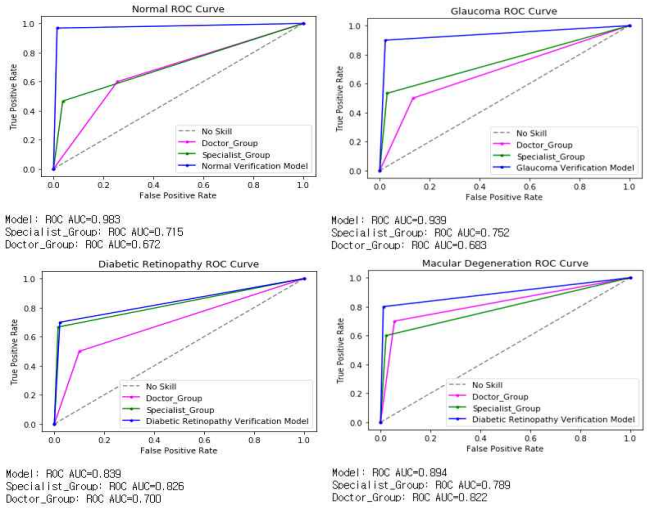


그림 4. 포렌식 모델 및 검증 그룹의 성능비교

본 연구의 결과는 그림 4와 같이 나타나며, 이는 본 포렌식 모델이 안과 전문의 및 일반의와 비교했을 때, 원본과 위변조된 이미지에 대한 판단 능력이 약 10.42% 더 높았다.

Ⅲ. 결론

본 논문에서는 안저 이미지를 대상으로 위변조된 이미지와 실제 이미지를 구분해낼 수 있는 포렌식 모델을 연구했다. 재생성 기술로 조작된 위변조 이미지는 의료현장의 전문의조차 구분하기 어려운 수준으로 재생성되어 환자의 정보, 임상실험, 사회보험제도, 등 여러 의료적, 사회적 문제를 발생시킬 수 있다. 따라서 본 논문의 Sparse CNN을 바탕으로한 포렌식 모델은 이러한 문제를 사전에 방지할 수 있는 기회를 제공한다. 향후연구로는 다양한 의료 이미지를 활용하여, 포렌식 모델을 연구할 예정이다.

ACKNOWLEDGMENT

이 논문은 2021년도 정부(산업통상자원부)의 재원으로 한국산업기술진흥원의 '바이오융복합기술 전문인력 양성사업'의 지원을 받아 수행된 연구임(No. P0017805).

참 고 문 헌

[1] Kim, H-J., Jung, D-C., and Choi, B-W. "Exploiting the vulnerability of deep learning-based artificial intelligence models in medical imaging: adversarial attacks.," Journal of the Korean Society of Radiology, pp. 259-273, 2019.

[2] Kaggle, "Ocular Disease Recognition: Right and left eye fundus photographs of 5000 patients.," (https://www.kaggle.com/andrewmvd/ocular-disease-recognition-odir5k).

[3] Kaggle, "MESSIDOR-2 DR Grades: Adjudicated DR Severity, DME, and Gradability for the MESSIDOR-2 fundus dataset.," (https://www.kaggle.com/google-brain/messidor2-dr-grades).

[4] Kaggle, "Glaucoma Detection: OCT Scans," (https://www.kaggle.com/sshikamaru/glaucoma-detection).

[5] Liu, B., Wang, M., Foroosh, H., Tappen, Marshall., and Penshy, M. "Sparse convolutional neural networks.," Proceedings of the IEEE conference on computer vision and pattern recognition, 2015.

인공지능 기반 COVID-19 진단 시스템 보안에 관한 연구

손인수
동국대학교 전자전기공학부
isohn@dongguk.edu

A Study on the Security of AI based COVID-19 Diagnosis Systems

Insoo Sohn
Dongguk University

1. 서론

의료 영상은 의학에 혁신적인 영향을 미치며 인체 장기, 세포 및 병리학 적 표본을 영상화하여 질병을 진단한다. 인공 지능에서 가장 효율적인 인공 신경망 기술 중 하나 인 심층 신경망 (DNN)을 응용하여 최근 의료 영상 분야에서 의료진단 개선에 효과를 보고 있다. COVID-19 진단 분야에서도 인공지능 기반 지능형 진단 기법에 대한 연구가 활발하게 연구가 되고 있으며 또한 AI 기반 COVID-19 시스템에 대한 사이버 공격 및 방어 대한 연구도 최근에 많은 관심을 받고 있다.

2. 본론

2.1 인공지능 기반 COVID-19 시스템 데이터

인공지능 기반 예측 및 분석 시스템에서 제일 중요한 요소 중 하나는 훈련 데이터이다. COVID-19 진단을 위한 인공지능 기반 시스템도 훈련 데이터가 중요한 역할을 하며 주요 데이터는 영상 기반의 흉부 엑스레이(CXR) 및 컴퓨터 단층촬영(CT-Scan) 이미지 데이터로 구성된다.

AI 기반 COVID-19 진단 시스템의 성능 평가 분석을 위해 사용되는 다양한 Dataset[1] 중에 제일 큰 규모의 샘플 수를 가지고 있는 Open Dataset 은 Wang et al.[2] 이 제안한 COVIDx 데이터 세트이며 현재까지 가장 큰 오픈 액세스 벤치마크 데이터 세트이다. 본 데이터 세트는 가장 많은 COVID-19 양성 환자의 사례를 기반으로 하며 13,8710 명의 환자의 13,975 개의 CXR 이미지로 구성되어 있다.

2.2 인공지능 기반 COVID-19 시스템 보안

인공지능 시스템에 가장 많이 적용하는 공격 방법은 적대적 훈련 세트를 증가시키는 Adversarial Attack 기법이며 인공지능 기반 COVID-19 시스템 공격을 위해 Qi et al.[3]은 SMIA(Stablilized Medical Image Attack) 공격 방법을 제안하였다.

인공지능 기반 COVID-19 시스템 공격에 대비한 방어기법에 대한 연구는 소수이며 최초로 Hirano et al.[4]은 인공지능 기반 COVID-19 시스템의 보안문제에 대한 분석을 하였다. 또한 CNN(Convolutional Neural Network) 기반 COVIDNet 모델을 COVID-19 공격 탐지를 위해 제안 하였으며 CXR 이미지를 사용하였다.

3. 결론

4 차 산업혁명의 중요한 인공지능 기술 기반의 의료영상 시스템 발전에서 환자의 개인정보 보호는 중요한 문제이며, 지능형 COVID-19 진단 시스템에 대한 사이버 공격에 대한 분석과 대응 기술 개발은 매우 중요하다. 본 논문에서는 최근에 연구되고 있는 AI 기반 COVID-19 시스템에 대한 사이버 공격 기술과 방어 기술을 검토하였다.

ACKNOWLEDGMENT

본 논문은 2018 년도 교육부의 재원으로 한국연구재단(2018R1D1A1B07041981)의 지원을 받아 수행된 기초연구사업입니다.

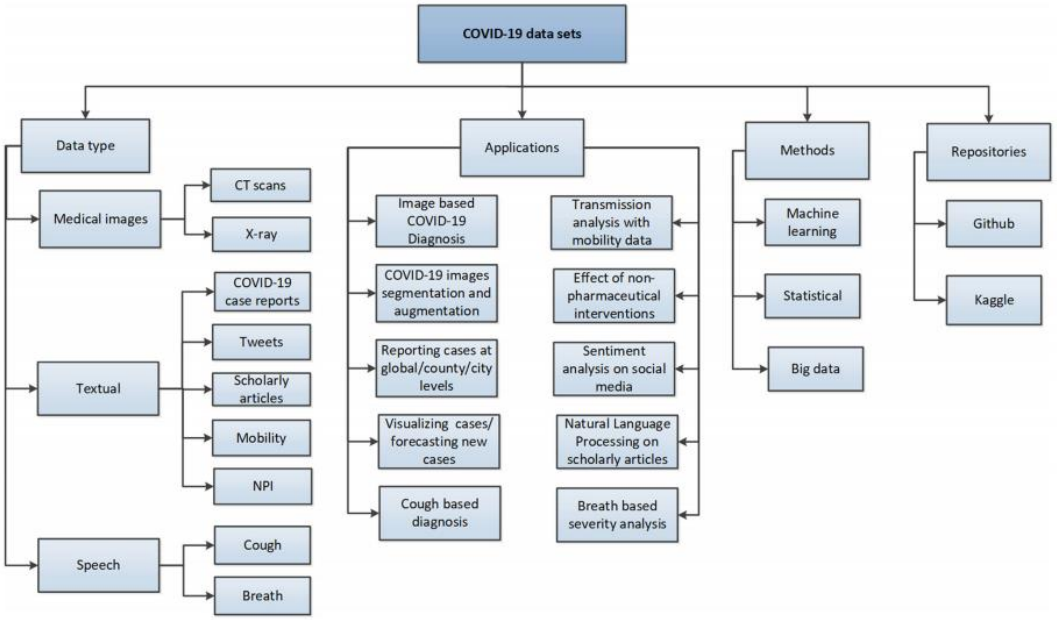


그림 1. COVID-19 Open Dataset [1]

참고 문헌

[1] J. Shuja, E. Alanazi, W. Alasmary, A. Alashaikh , "COVID-19 open source data sets: a comprehensive survey", Applied Intelligence, pp. 1296–1325, Sept. 2020.

[2] L. Wang, A. Wong, "COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest radiography images", arXiv preprint arXiv:2003.09871, 2020.

[3] G. Qi, L. Gong, Y. Song, K. Ma, Y. Zheng, "Stabilized medical image attacks", arXiv preprint arXiv:2013.05232v1, 2021.

[4] H. Hirano, K. Koga, K. Takemoto, "Vulnerability of deep neuural networks for detecting COVID-19 cases from chest X-ray images to universal adversarial attacks", Plos One vol. 15, no. 12, 2020.

멀티센서를 이용한 CNN기반 낙상감지 시스템

이영욱, 신수용*
금오공과대학교
thisis206@kumoh.ac.kr, wdragon@kumoh.ac.kr

Fall Detection System based on CNN using Multi-sensor

Yeong wook Lee, Soo Young Shin*
Kumoh National Institute of Technology

요 약

본 논문은 저비용으로 효율적이고 체계적으로 안전 취약 계층(어린이, 노인, 장애인 등)의 실내에서의 크고 작은 낙상사고를 방지하기 위해 본 논문에서 구현된 낙상감지 시스템을 이용하여 낙상사고를 미리 인지하고 예방한다. 제안된 시스템에서는 Thermal Array Sensor를 이용하여 데이터 샘플을 수집한 후 CNN기반 아키텍처인 yolo v3를 이용하여 낙상 알고리즘을 구현한다. 또한 UWB센서를 통하여 움직임 값을 측정하여 낙상 알고리즘의 정확도를 향상시켰다.

Ⅰ. 서 론

고령화는 전 세계적으로 많은 국가들이 직면한 세계적인 이슈이다. 고령화가 심화되면서 다양한 센서를 기반으로 한 스마트 웨어러블 제품이나 홈케어 시스템의 수요도 증가하였다[1]. 고령화 인구층을 위한 제품에는 주로 동작 감지 센서, 가스 누출 감지 센서, 블루투스 통신[2], 라이다 센서 등으로 다양하게 분포되어 있다. 하지만 이러한 스마트 제품은 비싼 재료비 때문에 제품의 가격 또한 비싸다. 또한 일반카메라를 사용하였을 경우 뛰어난 인식률을 가질 수 있지만 재실하고 있는 사람의 일상생활이 노출되기 때문에 사생활 보호 차원에서는 치명적이다. 본 논문에서 제안한 낙상감지 시스템은 낙상은 노인 뿐만 아니라 어린이, 장애인 등 안전 취약 계층에서 빈번하게 일어나는 사고이며 이러한 안전 취약 계층에게는 작은 낙상도 큰 위협이 될 수 있다. 본 논문에서는 Tharmal Array Sensor[3]와 UWB센서를 사용하여 낙상을 감지한다.

1) Thermal Array Sensor

Thermal Array Sensor는 Mexis사의 MLX90640 제품을 사용하였다.. 본 논문에서 사용된 Thermal Array Sensor는 Mexis사의 MLX90640 이다. 이 센서는 32x24의 해상도로 0.5Hz~32Hz까지 프레임 속도를 제어 할 수 있다. 이 센서는 I2C통신으로 제 일반카메라보다 성능은 저하되지만 사생활 보호 차원에서 상당히 높은 수준을 지니고 있다. 또한 가격면에서도 일반카메라보다 저렴하다는 특징으로 많은 분야에서 활용되고 있다.

논문에서 사용된 두 센서는 그림1에 나타나있다.

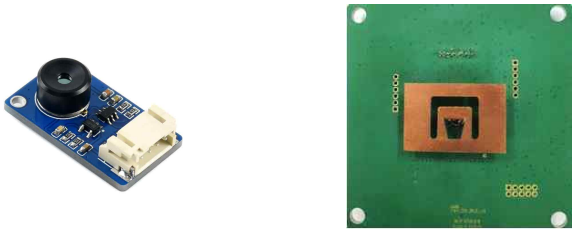


그림 1 MLX90640(좌), 천둥360(우)

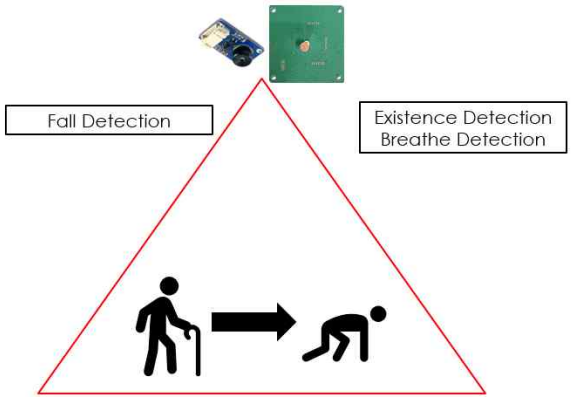


그림 2 제안된 시스템의 개요도

2) UWB Seonsor[4]

UWB Sensor는 하이밴드(7Ghz-9GHz)영역에서 울트라 와이드 밴드(Ultra Wide Band)기술을 기반으로 하는 임펄스 레이다로서 감지 범위 내에 있는 사람 또는 동물의 존재를 비접촉 원격 탐지하는데 사용 될 수 있다. 감지기는 사람의 호흡신호를 감지할 수 있으며 신뢰할 수 있는 포착 성능을 가지고 있다. 이 모델은 천장에 설치하여 최대 직경 10미터까지 미세움직임을 감지하고 최대 직경 7미터까지 호흡수를 감지할 수 있으며 UART통신을 지원한다.

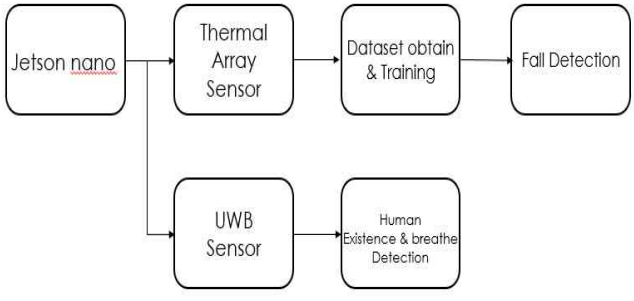


그림 3 시스템 블록도

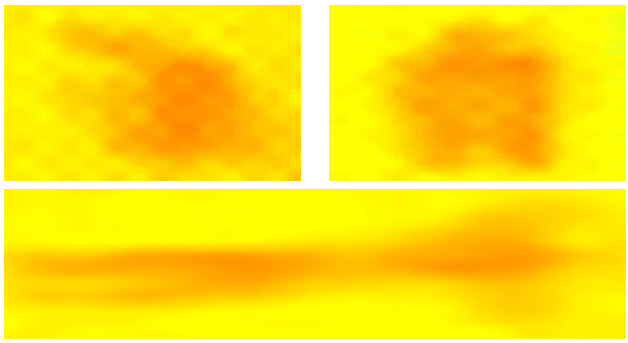


그림 4 Thermal Array Sensor에서 수집한 데이터 샘플 걷는모습(좌), 앉아있는 모습(우0, 쓰러진 모습(하)

Ⅱ. 본론

본 논문에서는 Thermal Array Sensor와 UWB 센서를 이용하여 낙상을 감지한다. 그림2는 제안된 시스템의 개요도이다. 두 개의 센서가 천장에 설치되어 일정범위 내의 사람을 탐지한다. UWB 센서는 사람이 아닌 물체(커피포트, 뜨거운 물체 등)이 있을 때 Thermal Array Sensor의 객체 인식을 방해 하는 요소를 식별해 줄 수 있다. 또한 재소자가 낙상이 발생하였을 때, 낙상사고는 넘어지는 과정이 정상적인 움직임보다 역동적이고 순식간에 일어나므로 UWB센서가 인식하는 움직임 수치가 급격히 상승하는 경향이 있다.그림3은 제안된 시스템의 블록도를 보여준다.

Thermal Array Sensor에서는 시각적 데이터를 수집하여 CNN 기반 아키텍처인 Yolo3를 이용하여 모델을 학습하고 테스트한다. Thermal Array Sensor에서 수집한 데이터를 이용하여 낙상감지 알고리즘을 구현한다. 데이터 샘플은 그림4와 같이 걷는 모습, 앉아있는 모습, 쓰러진 모습 세가지 클래스를 나누어 Thermal Array Sensor를 이용하여 2.7m의 높이의 천장에 설치하여 직접 수집하였다. 동시에 UWB센서에서는 영역 내의 사람의 재실여부를 판단하고 객체의 움직임을 파악함으로써 영역 내에서 사람으로 잘못 식별할 수 있는 물체를 구별 할 수 있다. UWB센서를 통하여 수집한 재실자의 움직임 감지 수준은 표 1에 나타나 있다. 정지해있거나 사람이 없는 공간에서는 센서값이 0이 나왔다. 천천히 걷는 것부터 일상활동은 100에서 1664정도의 값을 출력하였으며 낙상과 같은 역동적이고 빠른 동작에는 1795~2563정도의 값을 출력하였다.

Ⅲ. 결론

본 논문에서는 Thermal Array Sensor와 UWB센서를 이용하여 인간의 낙상을 감지할 수 있는 시스템을 제안하였다. 제안된 시스템을 구현하기 위해 CNN기반 아키텍처 yolo를 통하여 낙상 알고리즘을 구현하였다. yolo 학습 결과 그래프는 그림 5에 나타나 있다.또한 UWB센서를 통하여 잘못 탐지할 수 있는 요소에 대하여 오류를 줄임으로써 정확도를 좀 더 높일 수 있다. 향후 계획으로는 Thermal Array Sensor, UWB센서 뿐만 아니라 IMU센서 등의 다른 센서를 통하여 더욱 체계적인 시스템을 구현할 예정이며 무선 프로토콜을 구축하여 보호자가 피보호자의 안전 상태를 확인 할 수 있는 시스템을 구현할 계획이다. 또한 CNN을 이용하였을 때의 한계점을 극복하기 위해 LSTM, CNN기반 LSTM등의 적용 가능성을 검토할 예정이다.

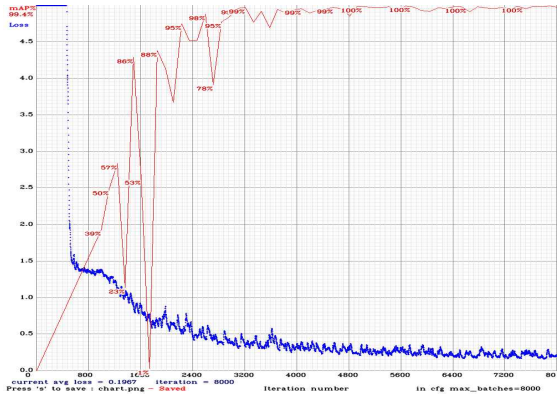


그림 5 학습 결과 그래프

	sensor 값
정지	0
걷기	100~497
일상활동	501~1664
격한활동	1795~2563

표 1 UWB센서에서 측정한 움직임 값

ACKNOWLEDGMENT

본 성과물은 중소벤처기업부에서 지원하는 2021년도 산학협력 거점형 플랫폼(R&D) (No. S3016551)의 연구수행으로 인한 결과물임을 밝힙니다.

This work was supported by project for Industry-Academic Cooperation Based Platform R&D funded Korea Ministry of SMEs and Startups in 2021.(Project No. S3016551)

참 고 문 헌

[1] Poncela, Alberto, et al. "Smart care home system: a platform for eAssistance." Journal of Ambient Intelligence and Humanized Computing 10.10 (2019): 3997-4021.

[2] Ha, Eun-Sil. "The implementation of smart care system for dementia patients." Journal of the Korea Academia-Industrial cooperation Society 15.6 (2014): 3832-3840.

and Home Healthcare, 2006. D2H2., April 2006, pp. 39–42.

[3] Trofimova, Anna A., et al. "Indoor human detection based on thermal array sensor data and adaptive background estimation." Journal of Computer and Communications 5.04 (2017): 16.

[4] Ruiz, Antonio Ramón Jiménez, and Fernando Seco Granja. "Comparing ubisense, bespoon, and decawave uwb location systems: Indoor performance analysis." IEEE Transactions on instrumentation and Measurement 66.8 (2017): 2106-2117.

독거노인 돌봄을 위한 전력데이터 기반 기계학습 이상진단 분석

김에덴, 고석갑, 이병탁
한국전자통신연구원

{kimed93, softgear, bytelee}@etri.re.kr

Analysis on machine learning anomaly detection using electricity data for care of elderly living alone

Kim E Den, Ko Seok Kap, Lee Byung Tak
Electronics and Telecommunications Research Institute (ETRI)

요 약

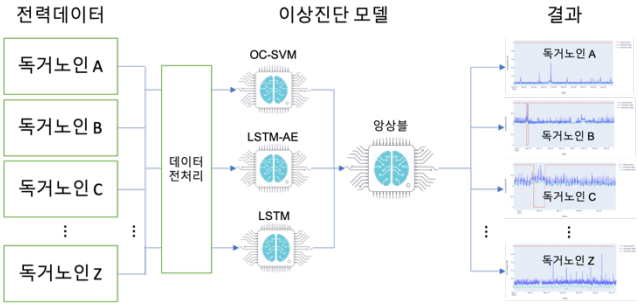
본 논문은 고령화로 인한 독거노인 인구의 증가로 돌봄 서비스에 관한 사회적 관심이 증가함에 따라 독거노인들의 전력데이터의 사용 패턴을 분석하여 딥러닝 및 머신러닝 모델을 통한 이상진단 기술을 소개하고자 한다. 실생활에서 전력 사용은 필수적이고 수집이 용이하기 때문에 전력 사용 패턴을 분석하여 독거노인의 이상상황을 판단한다. 본 논문에서는 세가지의 이상진단 모델을 제시함으로써 각 모델의 결과와 앙상블 기법을 통해 보다 신뢰성 있도록 이상진단 여부를 판별하고 독거노인의 이상 지수를 도출함으로 독거노인의 이상 정도에 대한 정보도 제공한다. 이와 같은 연구를 통하여 추후 사회복지사 또는 독거노인 보호자와의 연계를 통하여 사전 진단 혹은 이상 상황을 감지하여 효율적인 관리 및 서비스 등으로 활용될 수 있다.

Ⅰ. 서론

고령화 시대를 통하여 우리나라의 65 세 이상 인구는 현재 850 만명 정도로 늘어났고 홀로 사는 독거노인의 비율은 19.6%로 계속해서 늘어나고 있는 실정이다[1]. 이처럼 홀로 지내는 노인의 숫자와 비율은 늘어나고 있지만 이를 체계적으로 관리하고 보살 필 수 있는 인프라나 시설은 아직 부족하다. 이와 같은 문제를 해결하기 위해 독거노인 케어 서비스를 위한 기술들을 개발하고 있다. 그 중에 본 논문에서는 독거노인이 생활하는데 필수적인 요소인 전력 데이터를 수집하여 전력 수요 패턴 분석을 통한 AI 이상진단 연구에 대하여 소개한다.

기존의 독거노인 케어는 사람이 직접 확인하거나 전력 혹은 다른 수치들을 직접 모니터링하는 방법으로 대부분 구성되어 있다. 따라서 결국 인적 재원이 가장 중요하게 필요한 요소이다. 이러한 인적 재원을 보완하기 위해 최근 로봇, 스피커 등 많은 기술이 개발되고 있으나 이러한 서비스는 비용이 많이 요구된다는 점이 존재한다. 반면 본 논문의 연구는 전력데이터를 이용하기 때문에 보다 경제적으로 서비스가 가능하며 인공지능을 접목함으로 인적 재원의 손실을 줄일 수 있다는 장점이 있다. 비슷한 사례로 시행되는 서비스도 존재하는데, 본 논문에서는 전력 패턴을 통해 이상진단 결과 외에도 독거노인의 상태를 가늠할 수 있는 지표인 이상 지수를 추출하여 사전 돌봄 관리가 가능하다.

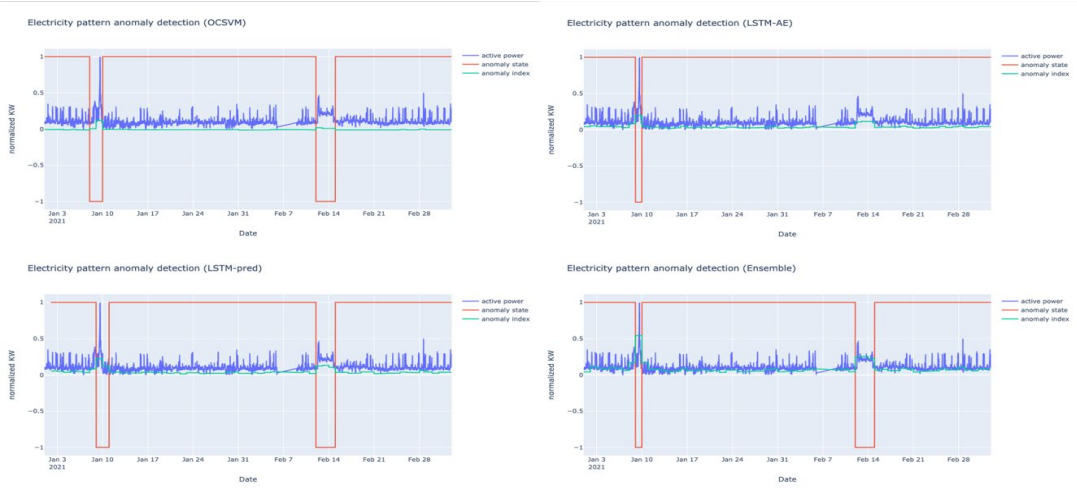
본 논문에서는 전력 패턴 이상진단을 위한 딥러닝 및 머신러닝 모델을 3 가지 사용하였다. 이는 경우에 따라 앙상블하여 단일 모델보다 신뢰성 있는 이상진단 결과를 도출할 수 있다. 이는 하루 단위로 학습되어 하루 전력 사용 패턴을 보고 이상상황을 진단하고 이상 지수를 도출하는데 사용된다.



(그림 1) 이상진단 모델 개념도

Ⅱ. 본론

데이터 및 전처리. 본 연구를 진행하기 위해서는 독거노인 가구에 설치된 AMI 기기를 통해 15 분 단위로 전력량 데이터를 수집한다. 15 분 단위로 수집된 데이터는 그 자체가 전력 소비 패턴을 나타내는 것이 아니기 때문에 하루 단위로 데이터를 재구성하여 분석하는 작업이 필요하다. 즉, 15 분마다 데이터를 수집하므로 하루의 전력 소비 패턴은 96 개의 데이터 포인트를 의미한다. 전력량 데이터를 AMI 기기로 수집할 때 정전 혹은 통신 등의 상황들로 인하여 데이터의 결측치가 발생하기도 한다. 이러한 경우 적절한 값을 채워 넣을 수 있으면 좋지만 장기간 결측치가 발생하는 경우도 있고 임의로 데이터의 값을 채워 넣었을 때, 이상진단의 측면에서는 에러가 높아질 수 있어 부적절하다. 따라서 이러한 경우에는 데이터를 하루단위로 재구성 할 때 하루 단위의 데이터 96 개중에 하나라도 결측치가 있으면 그 하루 전체의 데이터를 사용하지 않는다. 결측치 없이 수집된 정상 데이터만을 필터링하여 이상진단 모델 학습하는데 사용한다.



(그림 2) 데이터셋 2 에 대한 이상진단 및 이상지수 도출 결과

결측치가 제거된 데이터들은 딥러닝 및 머신러닝 모델에 적절히 사용되기 위하여 데이터 정규화 작업이 필요하다. 각 데이터마다 기저 전력과 피크 전력량이 달라 최대-최소 정규화를 통하여 모든 전력 데이터를 0 과 1 사이값으로 정규화를 진행하여 모델 학습 및 테스트 하기 위한 환경을 구축한다.

이상진단 모델. 본 논문에서는 이상진단을 위한 딥러닝 및 머신러닝 모델로 OCSVM 이상진단 방법, LSTM AutoEncoder 를 통한 이상진단 방법, LSTM 예측을 통한 이상진단 방법 이렇게 총 3 가지의 모델을 사용하였다. 먼저 OCSVM 방법은 하나의 클래스로 SVM 기법을 통하여 정상 범주를 정하고 그 범주를 벗어날 때에 이상상황이라고 판단하는 기법이다[2]. LSTM AutoEncoder 이상진단 방법은 LSTM 모델로 구성된 인코더, 디코더를 학습하고 새로운 데이터에 대한 복원 값의 차이를 계산한다. 그 차이의 범주에 따라 정상인지 비정상인지 판별한다. 마지막으로 LSTM 예측을 통한 이상진단 방법은 과거 데이터를 입력으로 LSTM 모델의 결과로 당일 전력패턴을 예측한다. 그 LSTM 예측값이 실제 전력 데이터와 얼마나 차이가 나는지를 계산하고 그 차이의 범주에 따라서 정상인지 비정상인지 판별한다[3]. OCSVM 은 이상진단에 가장 기본 모델로의 역할을 LSTM 은 시계열 특징을 통한 분석으로 오토인코더를 통해 현재 데이터로, 예측모델을 통해 과거 데이터로 이상판단을 함으로 이 세가지의 케이스를 고려하여 이상진단 모델들이 선택되었다.

실험 결과. 본 실험은 6 곳의 독거노인 가구의 전력데이터를 수집하여 진행하였다. 각 데이터 셋은 5~9 개월 동안 수집된 데이터이고 시간으로 트레이닝과 테스트 셋을 나누어서 실험을 진행하였다. 각 데이터의 구성은 다음 표 1 에서 나타낸다.

<표 1> 실험 데이터 정보

	트레이닝 셋	테스트 셋
데이터셋 1	2020.06.16 ~ 2020.09.30	2020.10.01 ~ 2020.10.30
데이터셋 2	2020.06.16 ~ 2020.12.31	2021.01.01 ~ 2021.03.05
데이터셋 3	2020.07.03 ~ 2020.12.31	2021.01.01 ~ 2021.02.06
데이터셋 4	2020.07.03 ~ 2020.12.31	2021.01.01 ~ 2021.03.27
데이터셋 5	2020.07.03 ~ 2020.12.31	2021.01.01 ~ 2021.03.05
데이터셋 6	2020.07.03 ~ 2020.10.31	2020.11.01 ~ 2020.12.15

그림 2 는 대표적으로 데이터셋 2 에 대한 이상진단 결과를 나타낸다. 이는 2020 년 6 월 16 일 부터 12 월 31 일까지의 데이터를 정상으로 학습시킨 후 2021 년 1 월 1 일 부터 3 월 5 일까지의 데이터를 통하여 실험한 결과이다. 각 모델의 결과를 보여주고 최종적으로는 세가지 결과를 통한 앙상블 결과를 나타낸다. 앙상블 모델은 각 모델로부터 추출된 이상지수 결과를 집계한 일종의 배깅 기법을 통하여 최종 결과를 도출하였다.

세부적으로 이상진단 결과를 살펴보면 육안으로도 전력량 패턴이 평소와 다르게 많이 사용된 날에 대부분의 모델은 이상상황이라고 감지하였다. 그림 2 를 통해 보면 1 월 9-10 일, 2 월 12-14 일이 이에 해당한다. LSTM-AE 모델의 경우 2 월 12-14 일을 정상으로 판단하였지만, 이상 지수는 다른 정상 상황에 비해 비교적 높은 값을 나타내고 있어 정상 범주에 있지만 이상상황에 대한 사전대비가 필요함을 알려준다. 최종 앙상블 모델에서는 세가지 모델의 이상지수를 집계하여 1 월 9 일과 2 월 12-14 일의 전력 패턴에 이상이 있다고 감지하였다. 그 중에서도 1 월 9 일의 이상지수가 높기에 이상 정도가 더 높음을 볼 수 있다.

Ⅲ. 결론

본 논문에서는 전력 소비 패턴 데이터를 통하여 독거노인 돌봄을 위한 이상진단 모델을 개발하고 실험하였다. 하루 단위로 데이터를 전처리 하고 3 가지 모델을 통하여 이상진단 및 이상 지수를 추출하여 사전 진단으로도 활용 가능성이 있어 독거노인 돌봄 서비스를 접목할 때 보다 유용한 서비스를 위한 연동이 가능하다. 또한 앙상블 기법을 통하여 단일 모델의 단점을 보완할 수 있어 보다 신뢰성 있는 결과를 도출할 수 있다.

ACKNOWLEDGMENT

본 연구는 산업통상자원부(MOTIE)와 한국에너지기술평가원(KITEP)의 지원을 받아 수행한 연구 과제임. (No. 20181210301570)

참 고 문 헌

[1] 통계청, 「장래가구추계 2017」, 「장래인구추계 2017」 2021.

[2] Li, K., et al, “Improving one-class SVM for anomaly detection.” In Proc. of ICMLC, 2003.

[3] Wang, R., et al, "Deep learning for anomaly detection." In Proc. of WSDM’20, 2020.

심전도 신호의 양자화 방법에 대한 심혈관 질환 분류시스템의 성능에 관한 연구

이희창, 윤태영, 여채윤, 오현영, 지예빈, 심성우, 강대성
인제대학교 헬스케어 IT 학과

gmlckd1254@oasis.inje.ac.kr, yty20173270@oasis.inje.ac.kr, ducodbs0516@oasis.inje.ac.kr, hy1013@oasis.inje.ac.kr, librenoblejyb@oasis.inje.ac.kr, mancity12@oasis.inje.ac.kr, mpkds@inje.ac.kr

A Study on the Performance of the Cardiovascular Disease Classification System for the Quantization of ECG Signals

Heechang Lee, Tae Young Yoon, Chaeyun Yeo, HyeonYeong Oh, Yebin Ji, Seongwoo Sim, Daesung Kang
Dept. of Healthcare IT, Inje Univ.

요 약

본 논문은 1 차원 신호의 형태학적 정보를 가지는 GLCM(gray level co-occurrence matrix), GLRLM(gray level run-length matrix) 행렬에서 특징값을 추출할 때 선행되는 양자화의 방법이 기계학습 기반의 심혈관 질환 분류 시스템에 어떠한 영향을 주는지에 관한 연구이다. 양자화 방법으로는 균일 양자화, 로이드-맥스 양자화, 히스토그램 평활화 양자화, 사분위수 양자화 방법 등을 적용하였고, 기계학습 방법은 결정 트리, 로지스틱 회귀, 랜덤 포레스트를 활용하였다. 전반적으로 로지스틱 회귀 방법이 다른 기계학습 알고리즘 보다 성능이 좋았고, 히스토그램 평활화 양자화 방법과 결합하였을 때 가장 좋은 성능을 보이는 것을 확인하였다.

Ⅰ. 서 론

기계학습을 통하여 심질환 관련 심전도 신호 분류에 사용되는 방법은 크게 두 가지가 있다. 심전도 신호에서 특징값을 직접 추출해서 기계학습 알고리즘에 적용하는 방법과 딥러닝을 활용하여 특징값을 자동으로 추출하는 End-to-End 방식으로 심질환을 분류하는 방법이다.[1] 특징값을 직접 추출할 때 자주 사용되는 방법으로는 심전도 신호의 피크 값 간의 거리, 퓨리에나 웨이블릿 변환에서 얻은 계수 값 등이 있다.[2] 본 연구에서는 최근 연구되는 새로운 방법으로, 심전도 신호의 형태학적 정보를 가지는 GLCM, GLRLM 행렬을 생성하여 특징값을 추출하였다. 일반적으로 GLCM 과 GLRLM 행렬을 생성할 때는 양자화가 선행되어야 한다. 이미지와 같은 2D 데이터에서 양자화의 영향에 관한 연구가 진행되었지만, 아직까지 심전도 신호와 같은 1D 신호에서는 관련 연구가 거의 없다.[3] 이에 관련하여 본 논문은 1D 신호에서 GLCM 과 GLRLM 행렬 생성시 선행되는 양자화 방법이 심질환 분류에 미치는 영향을 다룬다.

Ⅱ. 본론

본 연구에서 사용한 심전도 신호는 최근 Scientific Data 저널에 공개된 오픈데이터로 10,646 명의 심전도를 12 리드로 측정한 데이터셋이다.[4] 버터워스 저역 통과 필터, LOESS 평활기, NLM(non local means) 등과 같은 전처리를 하였고, 전처리 중에 손상된 58 명의 신호를 제외한 10,588 명의 심전도 신호를 실험에 사용하였다. 심전도 데이터는 11 개의 카테고리로 구성되어 있는데, 데이터 불균형을 고려하여 11 개의 심질환 카테고리를 4 개의 카테고리로 병합하여 사용하였다.

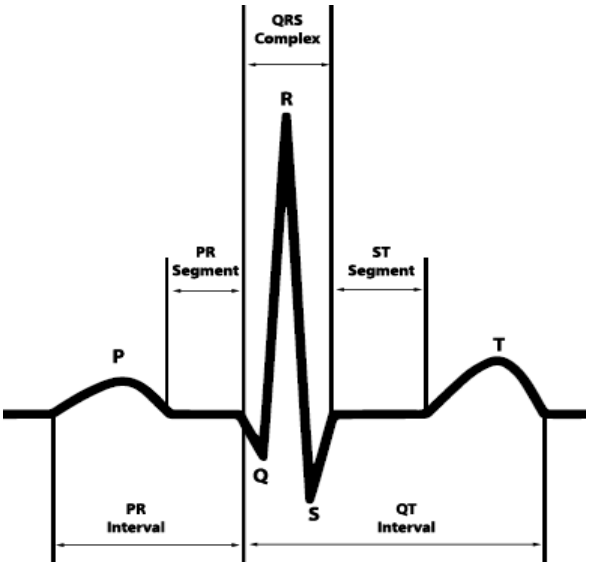


그림 1. 심전도 신호 [4]

데이터를 양자화하는 가장 간단한 방법은 균일 양자화(uniform quantization) 방법이다. 그림 1 과 같은 심전도 신호를 균일하게 양자화하면, R 피크에 비해 상대적으로 좁은 범위의 전압을 가지는 P, Q, S, T 피크 부근의 신호가 동일하거나 거의 같은 값으로 양자화 되는 경우가 생기는 반면 넓은 범위의 전압을 가지는 R 피크 값 주변의 신호가 다양한 값으로 양자화되기도 한다. 심혈관 질환을 분류할 때 R 피크 값도 중요하지만, P, Q, S, T 피크 값 역시 중요한 역할을 할 수 있다.[5] 본 연구에서는 P, Q, S, T 피크 주변에 존재하는 좁은 범위의 전압 신호를 보다 세밀하게 양자화하고, R 피크 부근의 넓은 범위의 전압 신호를 개략적으로 양자화할 수 있는 비균일 양자화 방법으로, 사분위수(quartile), 히스토그램 평활화(histogram equalization) 양자화 방법을 사용하였다. 사분위수 방법은 양자화 레벨의 75%를 사분위수 범위(IQR: 제 1 사분위와 제 3 사분위의 차)에 속하는 전압을 양자화하는데 사용하고, 나머지 25%의 양자화 레벨을 사분위수 범위 밖의 전압을 양자화하는데 사용하는 방법이다. 히스토그램 평활화 방법은 심전도 신호의 히스토그램을 먼저 생성한 후, P, Q, S, T 피크 주변에 밀집되어 있는 히스토그램을 평활화하여 많은 레벨의 양자화 값을 가지게 하고, R 피크 주변은 상대적으로 적은 레벨로 양자화하는 방법이다. 상기의 3 가지 양자화 방법과 대표적인 비균일 양자화 방법인 로이드-맥스(Lloyd-Max) 방법을 이용하여 심전도 신호를 양자화 하였다. 양자화 할 때 사용한 양자화 레벨은 32 단계이고, GLCM 행렬에서 기준점과 이웃하는 샘플의 오프셋 거리는 2 샘플 거리를 선택하였다.

양자화 된 심전도 신호에서 33 개의 특징값을 다음과 같이 추출하였다: 심전도 신호의 히스토그램에서 11 개, GLCM 행렬에서 9 개, GLRLM 행렬에서 13 개의 특징값을 각각 추출하였다. 본 연구에서 사용한 심전도 데이터는 12 리드 데이터이므로 한 명의 심전도 데이터에서 추출한 특징값은 모두 396 개이다.

양자화에 따른 심혈관 질환 분류를 위해 로지스틱 회귀(logistic regression)를 사용하였다. 분류 모델의 학습을 위해 전체 데이터의 80%를 학습데이터로 사용하였고, 나머지 20%는 테스트데이터로 사용하였다. 학습 모델의 유효성을 검증하기 위해 5 겹 교차검증(5-folded cross validation)을 사용하였다. 분석에 사용한 프로그램은 Python 3.7.3 과 Scikit-learn 0.24.2 이다.[6] 양자화 방법이 심혈관 질환 분류에 미치는 영향을 확인하기 위해서 테스트데이터의 정확도를 표 1 에 기술하였다.

	LR	DT	RF
균일 양자화 방법	**87.2**	76.7	86.6
로이드-맥스 방법	**87.6**	78.8	87.5
사분위수 방법	**88.1**	77.0	84.1
히스토그램 평활화 방법	**88.2**	76.0	83.4

표 1. 양자화 방법에 각 기계학습 알고리즘의 성능 비교. (LR: 로지스틱 회귀, DT: 결정트리, RF: 랜덤 포레스트)

Ⅲ. 결론

본 논문에서는 1 차원 심전도 신호의 형태학적 정보를 가지는 GLCM, GLRLM 행렬에서 특징값을 추출할 때 선행되는 양자화 방법이 심혈관 질환 분류시스템의 성능에 미치는 영향에 대해서 알아보았다. 표 1 의 결과에도 나와있듯이 로지스틱 회귀와 네 가지의 양자화 방법과 결합했을 때의 결과가 결정트리나 랜덤 포레스트보다 더 좋은 결과를 보였다. 전체적으로 보면, 로지스틱 회귀와 히스토그램 평활화 방법이 가장 높은 정확도를 보였다.

로지스틱 회귀와 네 가지 양자화 방법의 결과를 보면, P, Q, S, T 주변의 전압값을 세밀하게 양자화하는 사분위수와 히스토그램 평활화 방법이 균일 양자화 방법과 로이드-맥스 방법보다 더 좋은 결과를 보였다. 이는 P, Q, S, T 주변의 전압값이 심혈관 질환의 분류에 중요한 역할을 한다는 것을 간접적으로 보여준다고 볼 수 있다.

비록 본 실험에서 히스토그램 평활화 방법이 가장 좋은 성능을 보였지만, 이 방법이 항상 최선의 방법은 아니다. 본 실험에서는 심혈관 분류기로 로지스틱 회귀, 결정트리, 랜덤 포레스트만 고려하였고, 양자화의 레벨도 32 레벨, GLCM 에서 사용한 오프셋 거리도 2 샘플만 고려하였다. 보다 면밀한 결과를 위해서 SVM(support vector machine), kNN(k-nearest neighbor), XGBoost 등과 같은 다양한 분류기를 사용해야 하고, 양자화 레벨도 32 레벨 뿐만 아니라 16, 64, 128 레벨 등을 고려해야 한다. 그리고 GLCM 의 오프셋 거리도 1 샘플 거리에서부터 10 샘플거리까지 다양하게 바꾸어 가면서 실험이 행해져야 한다.

ACKNOWLEDGMENT

이 성과는 정부(과학기술정보통신부)의 재원으로

한국연구재단의 지원을 받아 수행된 연구임

(No.2020R1G1A1102881)

참 고 문 헌

[1] Parvaneh, S and Rubin, J. "Electrocardiogram Monitoring and Interpretation: From Traditional Machine Learning to Deep Learning, and Their Combination," Computing in Cardiology, 45, 2018.

[2] Arif M. et al. "Robust electrocardiogram(ECG) beat classification using discrete wavelet transform," Physiological Measurement, 29(5), 2008.

[3] Lee, G. N. et al. "Effect of quantization on co-occurrence matrix based texture features: An example study in mammography," Proc. SPIE 6144, Medical Imaging, 2006.

[4] Zheng, J. et al. "A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients," Scientific Data, 7(48), 2020.

[5] Magnani J.W. et al. "P wave duration is associated with cardiovascular and all-cause mortality outcomes: the National Health and Nutrition Examination Survey," Heart rhythm, 2011.

[6] https://scikit-learn.org/stable/index.html

자원이 제한된 자율주행 단말의 실시간 지능형 서비스를 위한 엣지 컴퓨팅 기반 이미지 처리 및 분석 시스템 개발

홍승준, 최인훈, 박수용, 정홍주, 김서현, 허의남*
경희대학교

hongsj1022@khu.ac.kr, inhun321@khu.ac.kr, clapd10@khu.ac.kr,
sub06038@khu.ac.kr, seok02h@khu.ac.kr, *johnhuh@khu.ac.kr

Development of edge computing-based image processing and analysis system for real-time intelligent service by autonomous vehicle with limited resources

Seungjun Hong, Inhun Choi, Suyong Bahk,
Hong-Ju Jeong, Seo-Hyun Kim, Eui-Nam Huh*
Kyung Hee Univ.

요 약

최근 자율주행을 이용한 지능형 서비스의 발전이 큰 주목을 받고 있다. 하지만 자율주행 단말의 제한된 자원으로 인해 이미지 분석을 이용한 다양하고 섬세한 서비스 제공에는 한계를 드러내고 있다. 본 논문은 이런 문제를 해결하고 실시간 지능형 서비스를 제공하기 위해 엣지 컴퓨팅을 활용한 실시간 이미지 처리 및 분석 시스템을 제안하고 평가한다.

Ⅰ. 서 론

최근 자율주행 분야는 센서와 컴퓨팅 기술의 발전으로 빠르게 진화하고 있다[1]. 그 중 합성곱 신경망(CNN, Convolutional Neural Network)을 이용한 이미지 분석 기술은 주변 상황에 대한 인지와 판단을 담당하며, 자율주행의 큰 축을 담당한다[2].

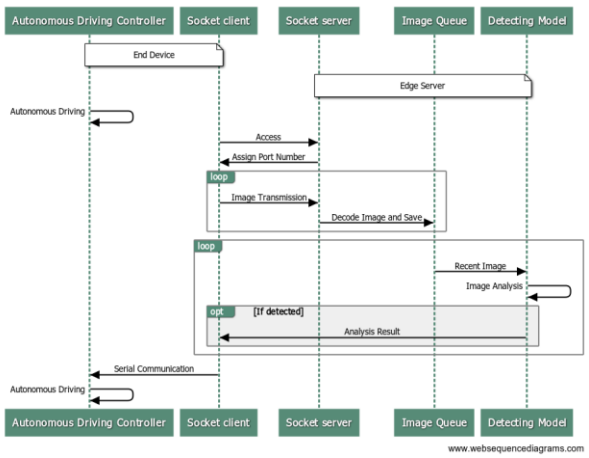
하지만 자율주행 단말의 제어를 담당하는 보드는 컴퓨팅 자원이 제한적이기 때문에 이미지 분석에 대한 온전한 성능을 낼 수 없다. 이로 인해 생기는 지연시간은 실시간으로 주행을 제어하고 지능형 서비스를 제공함에 있어서 큰 문제를 불러 일으킨다.

본 논문에서는 이러한 문제를 해결하기 위해 자율주행 단말과 엣지 서버간 통신을 통해 실시간으로 이미지를 처리하고 분석하는 시스템을 구현하고 평가한다.

Ⅱ. 본론

자율주행 단말과 같은 이동성 단말은 네트워크 제공과 서버와의 통신 방법에 대한 제한이 생긴다. 지속적인 이동으로 고정된 네트워크를 제공할 수 없으며, 주행 경로에 설치된 무선 네트워크를 변경하며 사용할 경우 단말의 IP 주소가 지속적으로 변경되어 서버에서 단말로의 데이터 전송이 불가능하다. 또한, 네트워크가 변경되는 시간에 서비스의 실시간성이 보장되지 못한다.

따라서 본 논문에서는 고정된 엣지 서버의 네트워크 주소를 이용하여 데이터를 주고받는 소켓 통신을 사용하였다. 단말의 네트워크는 스마트폰을 이용한 무선 네트워크 제공으로 고정되게 하였으며, 단말의 요청에 따라 할당된 서버의 소켓으로 데이터를 전송한다. 이는 단일 자율주행 단말이 아닌 다중 단말을 이용한 지능형 서비스 제공의 경우에도 통신이 가능한 확장성을 지닌다.



[그림 1] 이미지 분석 시퀀스 다이어그램

그림 1 은 자율주행 단말과 엣지 서버의 통신을 이용한 실시간 이미지 분석 시나리오를 나타내는 시퀀스 다이어그램이다. STM32 보드 제어를 통해 자율주행차가 주행을 시작하면 단말에서 서버로 소켓 연결 요청을 보낸다. 요청을 받은 서버는 소켓을 할당하고 단말로부터 주행 중 촬영되는 이미지를 해당 소켓으로 전송받는다. 전송받은 이미지는 디코딩 후에 큐에 저장되며 이는 주행이 끝날 때까지 반복된다. 이후 큐에 저장된 최근 이미지를 불러

와 학습된 모델을 사용하여 분석하고 인식 결과가 도출되면 분석을 멈춘다. 인식된 결과는 단말에 전송되고 STM32 보드와의 시리얼 통신을 통해 자율주행차를 제어한다.

이미지를 큐에 저장할 때는 엣지 서버의 디스크 용량을 고려하여 동일한 이름의 파일로 저장하고 학습 모델은 가장 최근 이미지를 분석하게 된다. 이 때, 이미지를 큐에 저장하는 프로세스와 큐에서 이미지를 가져와 분석하는 프로세스는 별도로 동작하기 때문에 두 동작이 겹치는 경우를 대비해 예외처리를 하였다.

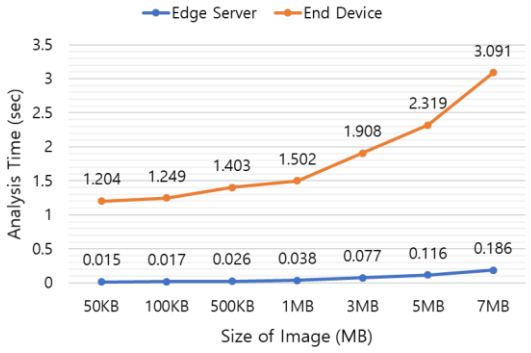
■ 이미지 분석시간 비교 및 평가

본 논문에서는 자율주행 단말로 Raspberry Pi 4B 4GB 모델을 사용하였으며, 엣지 서버는 Ubuntu 18.04.2 LTS Server 를 설치하였다. 그림 2 는 자율주행 단말과 엣지 서버의 상세 스펙을 나타낸다.

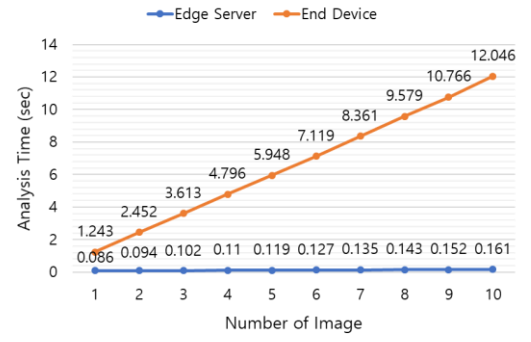
	End Device	Edge Server
CPU	ARM Cortex-A72 1.4GHz	Intel(R) Core(TM) i5-8400 CPU @ 2.80GHz
GPU	Broadcom VideoCore IV	NVIDIA GeForce GTX 1050Ti
RAM	4GB	16GB
DISK	128GB	120GB
OS	Raspbian GNU/Linux 10 (buster)	Ubuntu 18.04.2 LTS Server

[그림 2] 단말과 서버의 상세 스펙

이미지 분석에 사용된 학습 모델은 yolo-v3 tiny 모델을 사용하여 학습시킨 캔 음료 인식 모델이며[3], 단말에서 서버로 이미지를 전송하는 데 걸린 시간은 이미지 당 약 0.02 초가 소요되었다.



(A) 이미지 크기에 따른 분석 시간 비교



(B) 이미지 개수에 따른 누적 분석 시간 비교

[그림 3] 자율주행 단말과 엣지 서버의 이미지 분석시간 비교 그래프

그림 3 의 (A)는 픽셀을 제외한 모든 조건이 동일한 크기 별 이미지를 10 회 분석 후 평균시간을 측정한 그래프이다. 이미지의 크기가 증가함에 따라 평균 분석시간 또한 증가함을 관찰할 수 있으며, 엣지 서버에 비해 단말의 분석시간이 더 큰 증가율을 보인 결과, 7MB 를 기준으로 엣지 서버가 단말에 비해 약 94% 빠른 것을 알 수 있다. 이는 더 높은 정확도의 분석과 서비스 제공을 위해 이미지의 해상도를 높일 경우 더욱 긴 지연시간의 발생으로 이어진다.

그림 3 의 (B)는 50KB 크기의 이미지를 기준으로 누적 분석시간을 나타내며, 분석량과 비례하여 소요시간 또한 증가함을 알 수 있다. 이는 10 장의 이미지를 기준으로 단말이 약 98% 빠른 성능을 보이며, 결국 단말의 처리량이 많아질수록 지연시간으로 인한 지능형 서비스를 제공에 치명적일 것으로 평가된다.

Ⅲ. 결론 및 향후 연구 방향

본 논문에서는 자원이 제한된 단말에서 이미지를 직접 분석하지 않고 엣지 서버로 전송하여 처리 및 분석 후 결과를 수신하는 시스템을 구현 및 평가하였다.

이는 통신 시간과 이미지 분석시간을 모두 고려하여 비교하더라도 엣지 컴퓨팅을 이용한 분석이 실시간성 보장에 더 효과적이라는 것을 알 수 있다. 이는 본 논문에서 사용한 yolo-v3 tiny 모델과 같이 경량화된 학습 모델보다 신경망의 레이어가 추가될 경우에는 더 큰 효과를 기대할 수 있다.

이 시스템은 향후 시각 장애인을 위한 길 안내 서비스 개발에 사용될 것이며 길 안내 서비스에 적합한 모델을 생성하고 실시간으로 분석 후 자율주행차를 제어하는 시스템으로 개발할 예정이다. 또한 사용자의 증가로 다중 자율주행 단말에 지능형 서비스를 제공해야하는 경우를 고려하여 서비스를 확장해 나갈 예정이다.

ACKNOWLEDGMENT

이 논문은 2021 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2017-0-00294, (대학 ICT 기초연구실) 서비스 이동 지원을 위한 분산형 클라우드 핵심원천기술 연구)

참 고 문 헌

[1] Vasu Singh, Siva Kumar Sastry Hari, Timothy Tsai, Mandar Pitale; Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2021, pp. 122-128

[2] 이경민. "딥러닝을 이용한 영상 기반 자율 주행 자동차 시스템." 국내석사학위논문 세명대학교 대학원, 2018. 충청북도

[3] In-Hun Choi, “Robot Control System with Deep Learning based Beverage Image Recognition”, 1st Korea Artificial Intelligenc Conference, pp. 94-95, 2020.

LSTM 기반의 개선된 UAV 경로 알고리즘

김훈희, 이만희, 신수용*
금오공과대학교 IT융복합공학과
sdgo2363@kumoh.ac.kr, fordmore@kumoh.ac.kr, wdragon@kumoh.ac.kr

an Improved UAV path algorithm based on LSTM

Hoon Hee Kim, Man hee Lee, Soo Young Shin*
Department of IT Convergence Engineering Kumoh National Institute of Technology

요 약

본 논문은 LSTM(Long Short-Term Memory)의 특징을 활용하여 개선된 경로 알고리즘을 제안한다. 경로 알고리즘 DRRT*(Dynamic Rapidly Random Tree*)는 동적인 환경에서의 경로를 효율적으로 생성한다. 제안된 알고리즘은 보다 효율적인 경로 생성을 위해서 LSTM과 결합하고 그 장점을 이용한다. Lidar를 통해 감지된 주변 환경에 DRRT*가 가지로 뻗어 나아가며 생성될 때 LSTM의 시간적인 특성을 활용할 수 있다. 시스템 구성으로는 Bebop2를 기반으로 연산을 위한 Xavier NX와 지도 구성을 위해서 Rplidar_S1 모델과 비교적 연산이 빠른 Hector Slam을 활용한다.

Ⅰ. 서 론

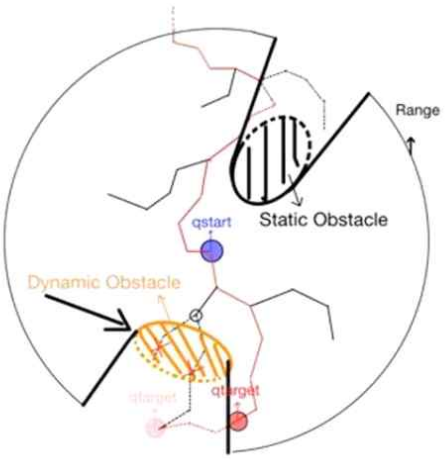
최근에 무인항공기(UAV)가 다양한 응용 분야로 활용되고 있다. UAV의 활용이 커지면서 제어하는 방식 또한 다양해지는데 컨트롤러, 음성, 제스처를 통한 사람이 조종하는 방식은 통신이나 명령에 대한 제약이 있다.[1] 따라서 UAV의 자율주행 기술은 다양한 분야에서 필요하다. 차량 자율주행의 경우 큰 인명 피해의 우려가 있어 자율주행에 단계를 나누고 그 단계에 따라 제한을 두기 때문에 완전자율주행을 실현하기에 시간이 더 필요하다. 그러나 드론 택시의 경우를 제외하면 UAV의 자율주행은 그 활용도가 더욱 크기 때문에 본 논문에서는 기존의 경로 알고리즘 DRRT*(Dynamic-Rapidly Random Tree*)에서 LSTM(Long Short-Term Memory)를 결합하여 시간에 따라 확산되는 RRT의 불필요한 경로들을 딥러닝 학습을 통해서 줄이고 효율적인 경로 생성하는 방법을 제안한다.

Ⅱ. 본 론

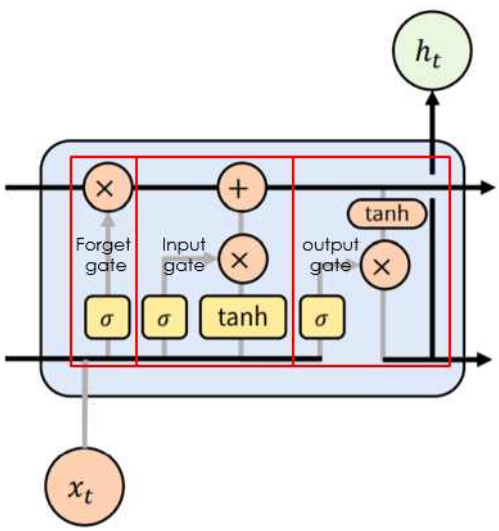
본론에서는 경로 생성 알고리즘인 DRRT*와 딥러닝의 LSTM에 대해 설명하고 두 기법을 결합하여 기존의 경로 알고리즘보다 효율적인 경로 생성 알고리즘을 제안한다. 시스템 구성으로는 Bebop2를 기반으로 연산을 위한 Xavier NX와 지도 구성을 위해서 Rplidar_S1 모델과 비교적 연산이 빠른 Hector Slam을 활용한다.

1. DRRT*

DRRT*(Dynamic Rapidly Radom Tree*)는 동적인 환경에서의 경로 생성 알고리즘이다. [그림 1]은 DRRT*의 생성과정을 보여준다. Lidar를 통해 주변의 장애물과의 거리를 파악하고 그 범위에서의 RRT(Rapidly Random Tree)를 생성한다. RRT는 구성 공간 내 무작위 노드를 선택하여 연결한 경로로 사방으로 퍼져 나아가며, 알려지지 않은 환경에서 경로 생성이 가능하다. DRRT*의 초기 생성 방식은 기존 RRT와 동일하다. 먼저 UAV의 중심부인 qstart에서 경로가 생성되고 장애물에 가려진 알려지지 않은 지역은 무효화 처리되며, 이는 동적 장애물에 대해서도 동일하다.



[그림 1] DRRT*의 생성



[그림 2] LSTM의 구조

또한 최근 영향을 받은 노드의 근처에서 재생성이 이루어진다. 위 과정을 반복하며 UAV가 움직일 때 Lidar를 통해 감지된 주변 환경도 변화하기 때문에 초기에 범위내 목표로 잡았던 qtarget에 도달하게 되면 경로를 다시 설정한다.[2]

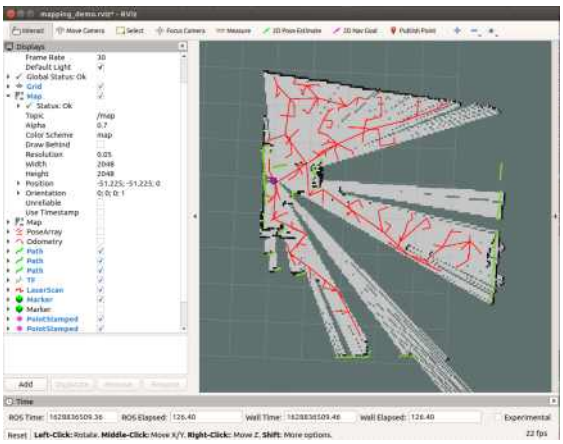
2. LSTM

LSTM(Long Short-Term Memory)는 RNN(Recurrent Neural Network)의 한 종류로서 긴 시간을 필요로 학습을 수행하는 문제에 대해서 즉, 시간적인 순서가 있는 문제를 해결하는데 보다 뛰어나다.[3] LSTM에는 [그림 2]와 같이 3개의 gate(forget gate, input gate, output gate) 그리고 Cell state로 구성되어있다. Cell state는 LSTM의 핵심으로 정보가 전달되는 길이고 각 gate에서 정보를 추가하거나 제거할 수 있다. 첫 번째로는 forget gate가 있다. 이전 값과 입력 x_t을 sigmoid layer를 통해 0과 1 사이의 값을 Cell state로 보내고 그 값이 0이면 모두 제거, 1이면 모두 전달의 의미를 갖는다. 두 번째로 input gate에서 새로운 정보 중 어떤 것을 저장할지 정하고 업데이트를 진행한다. 마지막으로 output gate에서는 sigmoid layer와 tanh 함수를 이용하여 출력으로 내보내고자 하는 부분만 내보낼 수 있다.[4]

이처럼 [그림 3]과 같은 경로 알고리즘의 시간적인 순서에 따른 확산을 장애물이나 환경 구조에 대해서 학습하고 UAV에 위험하거나 지나가기 힘든 환경에서의 불필요한 경로는 생성하지 않는 것으로 경로를 효율적으로 생성할 수 있다.

Ⅲ. 결 론

본 논문에서는 UAV의 자율주행을 위해서 기존의 경로 알고리즘 DRRT*에 LSTM 적용하는 방법을 제안하였으며, 경로 발생 단계에서 UAV가 진행하기 복잡한 경로나 장애물과 너무 가깝게 생성된 경로를 forget gate에서 제거하고 Cell state를 업데이트하여 목적지에 도달하는 최적의 경로를 생성한다. 이는 기존의 알고리즘보다 효율적인 경로를 생성할 수 있다. 향후 연구로는 UAV가 경로를 따라가기 수월하도록 하기 위해서 Pure Pursuit과 같은 Tracking 알고리즘으로 경로를 부드럽게 할 수 있다. 또한 2D 환경에 국한되지 않는 3D으로의 확장도 연구할 계획이다.



[그림 3] Lidar를 통한 구성 환경에서의 RRT 생성

ACKNOWLEDGMENT

"이 논문은 2021년도 정부(교육과학기술부)의 재원으로 한국연구재단의 대학중점연구소 지원사업으로 수행된 연구임"
(2018R1A6A1A03024003)

참 고 문 헌

[1] Sim, Seungbae, Hunyeong Kwon, and Hosang Jung. "A Study on Utilization of Drone for Public Sector by Analysis of Drone Industry." Journal of Information Technology Services 15.4 (2016): 25-39.

[2] Shome, Rahul, et al. "drrt*: Scalable and informed asymptotically-optimal multi-robot motion planning." Autonomous Robots 44.3 (2020): 443-467.

[3] Greff, Klaus, et al. "LSTM: A search space odyssey." IEEE transactions on neural networks and learning systems 28.10 (2016): 2222-2232.

[4] Sundermeyer, Martin, Ralf Schlüter, and Hermann Ney. "LSTM neural networks for language modeling." Thirteenth annual conference of the international speech communication association. 2012.

한국 밤시간대의 실시간 차선인식 기술 연구

고석현, 한지은, 임예원, 이성진
동서울대학교 전자공학과
kohjun3@gmail.com, sungjinlee@du.ac.kr

A Study on the Real-Time Lane Detection at Night in Korea

Seokhyun Koh, Jieun Han, Yewon Lim, Sungjin Lee
Department of Electronic Engineering, Dong Seoul University

요 약

최근 자율주행 기술에 대한 높은 관심으로 자율주행 기술은 운전 보조 기능을 넘어서 완전 자율주행 단계에 이르고 있다. 본 논문에서는 이런 과도기적 자율주행 기술 단계에서도 필수적인 영상인식 기술로 사용되는 차선인식 기술에 대해 다루었다. 특히, 한국의 도로 상황 환경에서 실시간 동작의 가능성과 밤 시간대의 성능에 대해 연구하였다. 이를 위해 기존 외국 지형 낮 시간대의 훈련 데이터에 더하여 한국 지형 낮 시간대의 데이터와 한국 지형 밤 시간대 데이터 셋을 구축하여 추가 훈련을 시켰으며 NVIDIA 2080ti 단일 GPU에서 그 정확도 및 동작 지연 성능을 측정하였다. 이런 추가 데이터셋 구축 및 추가 훈련의 결과 Ultra Fast Lane Detection 모델을 기준으로 한국 지형 밤 시간대에 76.9 % 의 정확도에서 92.4 % 의 정확도 향상을 이룰 수 있었다. 또한, 동작 지연 성능은 약 10ms에서 14ms 사이로 다른 모델과의 연동동작을 위해서는 모델/기기 최적화를 위한 노력이 요구된다는 것을 알 수 있었다.

Ⅰ. 서론

최근 딥러닝 기술의 발전으로 자율주행 기술은 운전 보조 기능 (ADAS: Advanced Driver Assistance System)을 넘어서 완전 자율주행 (FSD: Full Self Driving) 단계에 이르고 있다. 하지만, 완전 자율주행의 단계까지 수많은 기술들이 100%의 성능 안전성이 요구되고 있어 사실 이를 위해서는 아직 더 많은 단계가 남아있다고 볼 수 있다. 그럼에도 ADAS의 여러 요소 기술들은 우리의 자동차를 더욱 안전하고 편리하게 만들어 주고 있기 때문에 관련 요소 기술들은 더욱 관심을 받고 관련 기술 개발은 지속될 것으로 예상되고 있다.

본 논문에서는 이런 자율주행의 여러 요소 기술들 중 ADAS 의 한 부분인 LKA (Lane Keeping Assist)의 핵심기술이면서 완전 자율주행 (Full Self Driving)에서도 요구되는 기술인 Lane Detection 기술에 초점을 맞추어 연구하였다. 특히 Lane Detection 기술은 Lidar로는 사실상 구현되기 힘들고 Camera에 전적으로 의존해야 하기 때문에 빛에 대한 영향, 즉 밤에 그 성능이 보장이 될 수 있을지에 대해 중점을 두어 연구하였다.

이를 위해 여러 밤 데이터 셋이 요구되지만 기존의 데이터 셋에는 이런 데이터가 부족하여 해당 데이터 셋을 직접 확보하는 방법으로 확충하여 실험하였다. 데이터 확보를 통해 추가 훈련 시킨 결과 기존 76.9%의 정확도에서 92.4%의 정확도로 약 15% 정확도 향상을 이룰 수 있었다.

Ⅱ. Lane Detection Model

본 논문에서는 Lane Detection 모델 중 Fast Detection 성격에 있는 Ultra Fast Lane Detection 모델 [1] 에 기반하여 성능 분석을 수행하였다. 해당 모델은 그림 1과 같이 각 등 간격의 row 들을 anchor로 하여 각 column에 lane 성분이 있는지를 Cross-Entropy Loss를 이용하여 check하는 방식으로 detection을 수행한다.

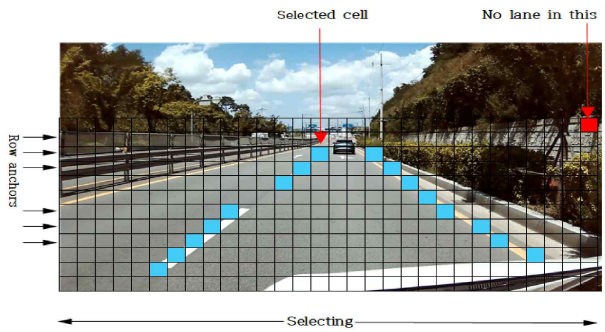


그림 1. Ultra Fast Lane Detection 모델의 동작 : 이미지 내에 여러 Row Anchor를 기준으로 각 Column에서 Lane 성분이 있는지를 Cross-Entropy 손실 함수를 이용하여 수치화하고 이를 학습

즉, 기존의 Segmentation 기반의 Lane Detection 들은 모든 Pixel들에 대해 Classification을 수행하는 반면 해당 모델은 Pixel들의 그룹이라 할 수 있는 Grid Cell 단위로 Classification을 수행하기에 여기서 가속화 성능이 나온다고 볼 수 있다.

Ⅲ. 데이터 셋 구축

Lane Detection을 위한 기본 훈련 데이터 셋으로 TuSimple [2] 를 사용하였다. 하지만 본 데이터 셋은 미국 지형의 낮 환경 도로 주행에서 만들어진 것이기에 한국 지형 밤 도로 환경에 적용하기에는 성능 열화가 발생할 수 있다. 그러므로 본 논문에서는 한국의 도로환경에서 밤 시간대의 주행 데이터 셋을 구축하였다. 또한 성능 비교를 위해 낮 시간대에서의 주행 데이터 셋 역시 구축하였다. 그림 2에서는 데이터 셋 구축을 위해 사용한 기

기들의 스펙을 나타내었다.



그림 2. 데이터 셋 구축을 위해 사용된 기기들

위 구축된 기기들 중 Lane Detection을 위해서 직접적으로 영향을 미치는 왼쪽 카메라 (운전자 기준)에서 수집되는 영상 이미지를 통해 데이터 셋을 구축하였다. 카메라는 FHD (1920x1080) 30 FPS 126° 화각을 지닌 WebCam을 사용하여 영상을 획득하였다. 경로는 동서울대학교를 중심으로 3가지 코스 즉, 교외 지형, 도심 지형, 고속도로 지형에 대해 낮 (12시30분~1시30분) 과 밤 (8시~9시) 시간대 각각에 대해 데이터를 획득하였다.

Ⅳ. 실험

실험을 위해 다음과 같은 Evaluation Set을 정하여 진행하였다.

- Basic Set : 기본 TuSimple 데이터 셋에서 훈련을 진행
- AugNight Set : 밤 데이터를 가공하기 위해 밝기 조절을 통해 Data Augmentation을 진행한 데이터를 통해 훈련을 진행
- RealDay Set : 실제 한국의 도로를 낮 시간대에 촬영하여 얻어진 데이터 셋에서 훈련을 진행
- RealNight Set : 실제 한국의 도로를 밤 시간대에 촬영하여 얻이진 데이터 셋에서 훈련을 진행
- RealDay and Night Set : 실제 한국의 도로 낮 시간대에 촬영하여 얻어진 데이터와 밤 시간대에 촬영하여 얻어진 데이터 셋을 합친 데이터셋에서 훈련을 진행

해당 데이터 셋에 대해 Epoch 50, Learning Rate 0.1, Batch Size 32 의 설정으로 훈련하여 얻어진 결과는 다음 Table 1과 같다.

Evaluation Set	밝기		속도
	실제 낮	실제 밤	(ms)
Basic Set Acc (%)	88.7	76.9	10.7
AugNight Set Acc (%)	78.3	71.9	10.7
RealDay Set Acc(%)	92.6	83.7	14.2
RealNight Set Acc (%)	89.2	92.4	14.2
RealDay and NightSet Acc(%)	91.7	91.7	14.2

Table 1. 5 가지 Evaluation Set에 대해 낮, 밤의 환경에서 정확도 실험

우선 Basic Set 의 정확도를 보면 낮 시간대에는 88.7% 의 정확도를 얻지만 밤 시간대에서는 76.9% 의 정확도를 얻는다. 이를 통해 TuSimple처럼 외국 도로 낮 시간 환경에서 얻어지는 데이터는 국내 자율주행 시 낮 시간대에는 어느 정도의 성능을 얻을 수 있으나 밤 시간대 자율주행 시 정확도에 한계가 있을 수 있다 라는 교훈을 얻을 수 있겠다.

반면 데이터가 부족하여 수행한 Data Augmentation 기반 AugNight 와 같은 경우 낮 데이터에 대한 정확도가 78.3%로 떨어지는 것은 물론 밤 데이터에 대한 정확도가 71.9%로 더 떨어지는 것을 확인할 수 있으며 이를 통해 일률적으로 밝기를 낮추는 것은 밤 환경의 특징을 반연하지 못하며 오히려 낮 환경에 가깝다는 것을 알 수 있다. 이것은 자동차 전조등이나 가로등같은 주변 광원에 의한 효과를 반영하지 못한 것으로 그 이유를 찾을 수 있겠다. 다음 그림 3, 4, 5는 Basic Set과 AugNIght Set, RealNight Set으로 훈련된 모델로 실제 한국 밤 영상에 대해 테스트 했을 때 차선 인식 정확도에 대한 시각화 자료이다.

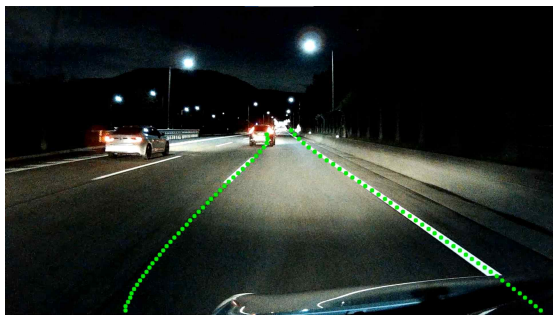


그림 3. Basic Set을 통해 훈련된 모델의 Lane Detection 결과

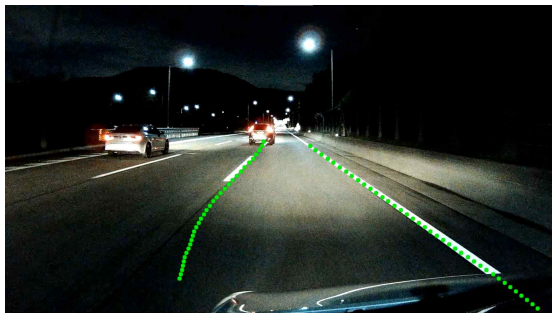


그림 4 AugNight Set을 통해 훈련된 모델의 Lane Detection 결과

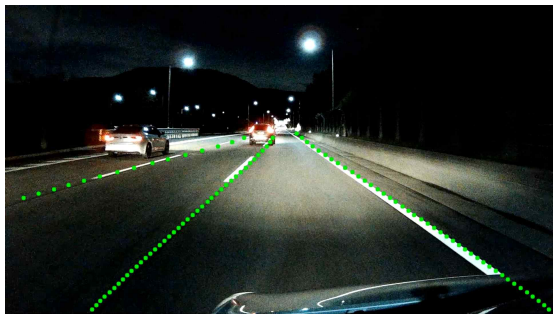


그림 5 RealNight Set을 통해 훈련된 모델의 Lane Detection 결과

마지막으로 RealDay Set을 통해 훈련된 기술의 정확도는 낮 시간대에 92.6 %, 밤 시간대에 83.7 % 정확도를 가지고 RealDay and NightSet을 통해 훈련된 기술의 정확도는 낮 시간대에 91.7%, 밤 시간대에 91.7%로 평균적으로 좋은 성능 향상을 이룰 수 있음을 알 수 있다. 이를 통해 자율주행 및 ADAS의 안정적 동작을 위해서는 낮 환경에서는 RealDay Set으로 훈련된 모델을 사용하고, 밤 환경에서는 RealNight Set으로 훈련된 모델을 사용하는 것이 가장 안정적인 성능을 확보할 수 있으며 메모리의 부족으로 해당 방법을 사용할 수 없다면 RealDay and NightSet으로 훈련한

모델을 사용하는것이 가장 좋은 방법임을 알 수 있었다.
이것을 통해 주행 시 겪을 수 있는 다양한 환경의 데이터 셋에 대해 훈련을 하는 것이 안정적 성능을 확보 할 수 있는 방안임을 알 수 있었다.

V. 결론

본 논문에서는 자율주행 및 ADAS의 핵심 기술 중 하나인 Lane Detection 기술을 다루었다. 특히 실시간 동작에 탁월한 Ultra Fast Lane Detection 모델을 통해 실 주행을 하는 환경 데이터 (지형, 시간) 에서 훈련을 하는 것이 좀 더 안정적 성능을 확보할 수 있는 방안 임을 알 수 있었다.

ACKNOWLEDGMENT

이 논문은 2021년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기본연구사업임(No. NRF-2019R1F1A1062878)

참 고 문 헌

[1] Zequn Qin, Huanyu Wang, Xi Li, “Ultra Fast Structure-aware Deep Lane Detection”

[2] CVPR 2017 Lane Detection Challenge, “https://github.com/TuSimple/tusimple-benchmark/tree /master/ doc/ lane_detection”

심층강화학습 기반 자율주행차량의 차선 변경 정책 안정성 평가

이동수, 권민혜*
숭실대학교
movementwater@soongsil.ac.kr, *minhae@ssu.ac.kr

Stability Evaluation on Lane-changing Policies for Deep Reinforcement Learning-based Autonomous Vehicles

Dongsu Lee, Minhae Kwon*
Soongsil University

요 약

자율주행기술과 같은 고신뢰 서비스(mission-critical service)는 상용화를 위해서는 학습 결과물의 안정성이 매우 중요하다. 본 연구에서는 자율주행차량의 안정적인 차선변경 정책을 위한 마르코프 의사결정(Markov Decision Process; MDP) 모델을 제안하고 대표적인 심층강화학습 알고리즘인 TD3와 PPO를 이용하여 자율주행차량을 학습시킨다. 학습된 자율주행차량의 속도분포의 엔트로피(entropy)를 확인하여 각 알고리즘별 학습 결과물의 안정성을 평가하였다. 그 결과 TD3를 기반으로 학습된 차량은 PPO 기반 학습차량에 비해 엔트로피(entropy) 값이 약 0.4 배 수준으로 낮아 더 안정적인 속도제어가 가능함을 확인하였다.

Ⅰ. 서 론

최근 인공지능 기술의 발전과 함께 자율주행기술의 연구가 활발히 진행되고 있다. 자율주행기술과 같은 고신뢰 서비스의 상용화를 위해서는 주행 효율성 뿐만 아니라 안정성 역시 중요한 척도가 된다. 본 연구에서는 차선 변경 학습을 위한 MDP 모델을 대표적인 심층강화학습 기반의 알고리즘인 Twin Delayed DDPG (TD3)[1] 와 Proximal Policy Optimization (PPO)[2]을 통해 자율주행차량을 학습시키고자 한다[3]. 그리고 불안정성 지표인 엔트로피를 이용하여 학습된 차량의 주행 정책 안정성을 비교한다.

Ⅱ. 차선 변경 학습을 위한 심층강화학습 문제 정의

모든 강화학습의 기본 동작은 마르코프 의사결정 과정(Markov Decision Process; MDP)으로 정의할 수 있다. MDP는 환경에 대한 완벽한 상태 정보를 획득할 수 있다는 가정이 존재하는데, 실제 환경에서는 부분적이거나 잡음(noise)이 섞인 불완벽한 상태 정보를 얻게 된다. 이러한 문제를 해결하기 위해 제안된 모델이 Partially Observable MDP (POMDP)이다. POMDP는 $< S, A, O, R, \gamma >$의 튜플로 표현할 수 있다. 튜플은 시간 t에서 유한한 상태 공간 집합 $s_t \in S$, 유한한 행동 공간 집합 $a_t \in A$, 유한한 관측 공간 집합 $o_t \in O$, 행동에 따른 보상 $R(s_t, a_t, s_{t+1},)$, 마지막으로 시간에 따른 감가율 γ을 포함한다.

Ⅱ.1. 도로 환경의 정의

본 논문에서는 차선변경이 필요한 2차선 원형 도로에서 한 대의 자율주행 차량과 여러 대의 비 자율주행차량이 혼재된 도로상황을 고려한다. 도로의 길이는 l이며 도로 내 차량의 집합 E는 자율주행차량 $[e_N]$과 비 자율주행차량 $[e_1, e_2 \cdots, e_{N-1}]$ 을 포함한다. 이와 같은 환경의 시간 t에서의 상태 정보는 $3N$차원으로 다음과 같이 정의 된다($s_t \in \mathbb{R}^{3N \times 1}$).

$$s_t = [v_1, p_1, k_1, v_2, p_2, k_2, \cdots, v_N, p_N, k_N]$$

여기서 v_n는 n 번째 차량의 속도, p_n은 n번째 차량의 위치, 마지막으로 k_n은 n 번째 차량이 위치한 차선 번호를 의미한다. 만약 특정 차량 e_n이 원형도로의 가장 바깥쪽 차선에 있다면 $k_n = 0$으로 정의하고 안쪽 차선으로 이동함에 따라 차선 번호가 1씩 증가한다.

Ⅱ.2. POMDP 정의

개체는 학습을 위해 시간 t에서의 환경 상태 정보 s_t를 관측하여 부분 관측 정보 o_t를 획득한다. 본 연구에서 자율주행차량 e_N이 관측 가능한 차량의 집합 $E_{obs} = [e_{l_0}, e_{l_1}, e_{f_0}, e_{f_1}]$은 전방 차량(leader) 두 대 e_{l_0}, e_{l_1}와 후방 차량(follower) 두 대 e_{f_0}, e_{f_1}, 총 네 대의 차량을 포함한다. 이때 관측 정보 o_t는 10차원으로 다음과 같이 정의 된다($o_t \in \mathbb{R}^{10 \times 1}$).

$$o_t = [v_{f0}, v_{f1}, v_{l0}, v_{l1}, v_N, P_{f0}, P_{f1}, P_{l0}, P_{l1}, k_N]$$

여기서 P는 자율주행차량의 위치 p_N과 관측 가능한 차량의 위치 p_{obs}의 상대거리를 의미한다.

개체가 매 시간 t마다 수행하는 행동 $a_t = \{a_{acc}, a_{lc}\}$는 가속도 조절 a_{acc}과 차선 변경 a_{lc}을 포함한다. 이때 a_{acc}는 유한하며 연속적인 실수 공간 $a_{acc} \in [acc_{\min}, acc_{\max}]$에서 정의 된다. $acc_{\min}$은 자율주행차량이 수행할 수 있는 최소 가속도를, $acc_{\max}$는 최대 가속도를 의미한다. a_{lc}는 유한하며 이산적인 실수 공간 $a_{lc} \in \{-1, 0, 1\}$ 에서 정의 된다. $a_{lc} = 0$인 경우 자율주행차량은 차선을 유지하며, $a_{lc} = -1$은 바깥쪽(우측)으로의 차선 변경을, 그리고 $a_{lc} = 1$는 안쪽(좌측)으로의 차선 변경을 의미한다.

개체는 주어진 상태를 관측한 뒤 행동을 선택하고 이를 바탕으로 보상 R_t을 얻게 된다. 차선 변경 및 가속도 조절 학습을 위한 보상함수는 다음과 같다.

$$R_t = \eta_1 \left(1 - \left| \frac{v_{t+1,N} - v^*}{v^*} \right| \right) + \eta_2 \left(\min \left[0, 1 - \left(\frac{s^*}{P_{t+1,f}} \right) \right] \right) + \eta_3 \times |a_{lc}| \min(0, P_{t+1,l} - P_{t,l}) \times P_{t,l} \quad (1)$$

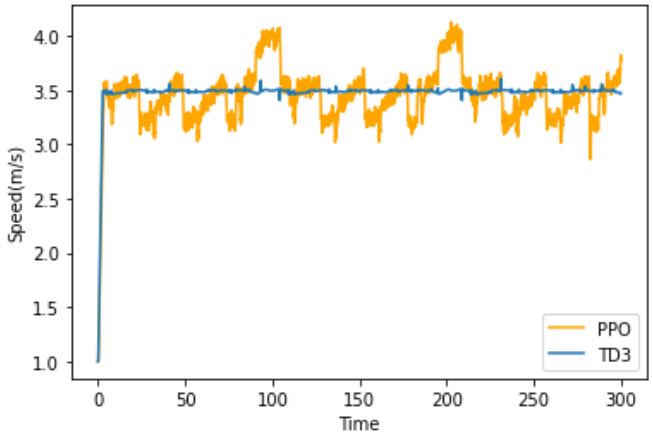


그림 1. 시간에 따른 자율주행차량의 속도 변화

보상함수는 세 가지 항으로 구분할 수 있다. $\left(1-\left|\frac{v_{t+1,N}-v^*}{v^*}\right|\right)$는 가속도 조절 후 자율주행차량의 속도$v_{t+1,N}$와 목표속도 v^*를 이용하여 원하는 속도로 주행할 수 있도록 한다. $\left(\min\left[0,1-\left(\frac{s^*}{P_{t+1,f}}\right)\right]\right)$는 차선 변경을 수행한 후 차량의 후방차량과의 상대거리 $P_{t+1,f}$ 와 안전거리 s^*를 고려한 안전거리 계수이다[4]. $|a_{lc}|\min(0,P_{t+1,l}-P_{t,l})\times P_{t,l}$ 는 차선 변경을 수행한 경우 t시간에서 선두차량과의 상대거리 $P_{t,l}$와 $t+1$시간에서 선두차량과의 상대거리 $P_{t+1,l}$ 비교하여 의미 없이 수행하는 차선 변경을 제어한다.

본 논문에서는 제시한 MDP 문제 해결을 위해 심층강화학습 알고리즘인 TD3와 PPO를 사용한다. 개체는 특정 상태에서 최적의 행동을 결정하는 정책을 학습한다. 각각의 정책 평가 및 업데이트에는 관측-행동 가치 함수 $Q(o,a)$및 관측 가치 함수 $V(o)$가 이용된다.

Ⅲ. 성능평가

본 연구에서는 앞서 제안한 POMDP문제를 풀기 위해 대표적인 심층강화학습 알고리즘인 TD3 및 PPO를 기반으로 자율주행차량을 학습 시킨다. 그 후, 알고리즘 별 학습 차량의 주행 정책의 안정성을 비교평가 하고자 한다.

Ⅲ.1 모의실험 설정

자율주행차량의 성능 평가를 위해 심층 강화학습 라이브러리와 교통 시뮬레이터를 통합한 프레임워크 FLOW[5]를 이용한다. 도로는 2차선 원형도로이며 $l=260m$이다. 전체 차량 $|E|=9$이며 비 자율주행차량 $|E_{non}|=8$이다. 이때 비 자율주행차량은 도로 내에서 일정한 간격으로 배치하였으며 Intelligence Driving Model (IDM) 컨트롤러[6]를 이용하여 주행한다. 모든 비 자율주행차량은 $1m/s$의 등속주행을 한다. 자율주행차량의 목표 속도 $v^*=3.5m/s$, 안전거리 s^*는 IDM 컨트롤러에서 제공하는 함수로 정의하며, 가속도 공간은 $acc_{\min}=-1m/s^2$, $acc_{\max}=1m/s^2$로 정의한다.

Ⅲ.2 자율주행차량의 주행 정책 안정성 평가

성능 평가에는 평균 속도 및 속도 분포에 대한 엔트로피를 이용하여 전체 에피소드에서 자율주행차량 정책의 안정성을 정량화하여 평가하였다. 엔트로피 H는 shannon entropy[7]를 이용하였다.

$$H=\sum_i(p_i)\ln(\frac{1}{p_i})$$

여기서 p는 개체가 경험한 주행 속도에 대한 확률을 의미한다.

표 1. 5개의 랜덤시드로 학습시킨 자율주행차량의 속도 및 엔트로피에 대한 평균과 분산

Algorithm	Speed $[m/s]$		Entropy (H)	
	mean	variance	mean	variance
PPO	$3.94m/s$	0.9×10^{-1}	4.87	0.13
TD3	$3.48m/s$	0.4×10^{-4}	2.05	0.13

주행 정책 안정성 평가는 그림 1의 학습 차량의 시간 변화에 따른 속도 변화 및 표1에서 5개의 다른 랜덤시드로 학습시킨 차량의 평균속도, 평균 엔트로피, 엔트로피의 분산을 통해 확인할 수 있다. TD3의 경우 에피소드 전체에서 큰 속도의 변화 없이 목표 속도인 $3.5m/s$를 유지한 채 주행하는 모습을 확인하였다. 반면 PPO의 경우 TD3와 비교하였을 때 속도의 변화 폭이 큰 것을 확인할 수 있다. 또한 PPO로 학습시킨 차량의 H는 TD3에 비해 2.4배 정도의 불안정성 척도를 보인다(표 1). 이는 PPO로 학습시킨 자율주행차량의 경우 일관된 정책이 아닌 불안정한 정책을 통해 주행한다는 것을 의미한다. 반면 TD3로 학습한 자율주행차량은 설정된 목표속도 $3.50m/s$에서 큰 변동이 없으며 낮은 엔트로피 값으로 안정적인 정책으로 주행하는 모습을 확인하였다.

Ⅳ. 결론

본 논문에서는 차선 변경을 위한 POMDP 모델을 제안하였다. 제안된 POMDP 모델로 디자인된 문제는 심층강화학습 알고리즘인 TD3와 PPO를 이용하여 성공적으로 해결하였다. 또한, 두 심층강화학습 알고리즘의 성능 차이를 비교하였다. TD3로 학습시킨 차량의 경우 PPO로 학습시킨 차량에 비해 목표속도 유지 능력이 뛰어남을 확인할 수 있었다. 결과적으로 논문에서 제안한 MDP 모델에 대해 TD3는 PPO보다 엔트로피가 0.4 배 낮은 수준으로 보다 안정적인 주행 정책을 수립함을 확인하였다.

사 사

이 논문은 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT 연구센터지원사업(IITP- 2021-2020-0-01602)과 한국 연구재단(NRF- 2020R1F1A1069182)의 지원을 받아 수행된 연구임.

참 고 문 헌

[1] S. Fujimoto, H. van Hoof, and D. Meger, "Addressing Function Approximation Error in Actor-Critic Methods," ICML, 2018.

[2] J. Schulman, F. Wolski, et al., "Proximal Policy Optimization Algorithms," arXiv preprint arXiv:1707.06347, 2017.

[3] 이동수, 권민혜, “심층강화학습기반 자율주행차량을 이용한 원형도로의 stop-and-go wave 현상 해결 전략 연구,” 한국통신학회 논문지, vol.46, no.10, 2021.

[4] 이동수, 권민혜, “PPO기반 자율주행차량의 효율적이고 안전한 차선 변경 정책 연구,” 한국통신학회 하계종합학술대회, 2021.

[5] C. Wu, A. Kreidieh, et al., "Flow: Architecture and Benchmarking for Reinforcement Learning inTraffic Control," arXiv preprint arXiv:1710.05465, 2017.

[6] M. Treiber, A. Hennecke, et al., "Congested Traffic States in Empirical Observations and Microscopic Simulations," Physical Review E, vol.62, no.2, pp. 1805-1824, 2000.

[7] C. E. Shannon, “A Mathematical Theory of Communication,” The Bell System Technical Journal, vol. 27, no. 3, pp. 379-423, 1948.

고정밀 지도를 이용한 모방 학습 기반 종단간 자율 주행

한승태, 조상재, 정명근, 김태선, 공승현*
한국과학기술원

seungtaeh@kaist.ac.kr, *skong@kaist.ac.kr

Imitation Learning-based End-to-end Autonomous Driving with High Definition Map

Seung-Tae Han, Sangjae Cho, Myung-Keun Jeong, Taeseon Kim, Seung-Hyun Kong*
Korea Advanced Institute of Science and Technology

요 약

참조 정책을 모방하도록 인공지능 네트워크를 학습시키는 모방 학습(Imitation Learning) 기반 종단간(End-to-end) 자율 주행 기술은 동일한 상태에서 취득한 관측치에 대해 사람에 따라 다른 행동을 취할 수 있다는 의사 결정 다양성 문제가 있다. 이러한 문제를 해결하기 위해, 본 논문에서는 차량 전방 이미지와 고정밀 지도 이미지로부터 차량의 조향각을 매핑하는 MIMSNet 을 제안한다. 그리고 직접 제작한 데이터셋을 사용하여 MIMSNet 과 기존 연구에서 제시된 네트워크를 학습 및 평가하고, 그 결과를 비교하여 MIMSNet 의 우월한 조향각 추정 성능을 입증한다.

Ⅰ. 서 론

자율 주행 기술의 대표적 패러다임으로는 모듈러 방식과 종단간(End-to-end) 방식이 있다. 모듈러 방식은 사람이 정의한 논리에 따라 인식, 판단, 그리고 제어 모듈 등으로 구성된다. 이러한 방식은 복잡한 주행 상황에서의 성능에 한계가 있으며, 높은 비용이 요구되어 효율적이지 않다.

반면, 종단간 방식은 인식에서부터 제어까지의 모든 과정을 단일 알고리즘으로 처리하기 때문에, 모듈러 방식의 단점을 극복할 수 있다. 최근 인공지능의 비약적 발전에 따라, 모방 학습(Imitation Learning) 기반 종단간 방식에 대한 연구가 활발히 진행되고 있다. 모방 학습은 사람이 직접 운전하는 상태에서 취득한 관측치와 그때 사람이 결정한 행동을 참조하여, 인공지능이 사람의 주행 정책을 스스로 모방하도록 학습하는 방식이다. 이러한 방식은 대표적으로 [1-4]에서 제시되었다.

[1]는 차량의 전방 이미지만을 조향각으로 매핑하는 기술을 제시하였고, 차선 유지 정도만을 실증하였다. 그러나, 해당 기술은 동일한 상태의 관측치에 대해 사람에 따라 서로 다른 행동을 취할 수 있다는 의사 결정 다양성 문제가 있다. 이를 해결하기 위해, [2, 3]는 방향 지시등 정보를 네트워크에 입력함으로써, 선회 방향 결정이 가능한 기술을 제시하였다. 하지만, 이 기술들은 데이터셋 구성 및 실증 시 방향 지시등을 수동으로 조작해야 한다는 한계가 있다. [4]는 단순화된 지도 이미지를 네트워크에 입력하는 기술들을 제시하였지만, 도로를 하나의 차선으로만 도시한 지도를 활용하기 때문에, 교차로 회전을 위한 차선 변경 등 다양한 환경에서의 적절한 행동을 출력하기에 한계가 있다. 이러한 한계는 네트워크에 도로의 지형 정보 및 전역 경로 정보를 반영하여 해결할 수 있다.

본 논문에서는 차량 전방 이미지와 고정밀 지도 이미지, 그리고 차량의 속도로부터 조향각을 출력하는 MIMSNet(Mapping from Image and Map to Steering Network)을 제안한다. 또한 MIMSNet 및 기존 연구에서 제시한 네트워크를 비교하여 MIMSNet 의 우월성을 입증한다.

Ⅱ. 본론

A. 데이터셋

본 논문에서는 한국과학기술원 내의 도로에서 사람이 직접 운전하며 측정한 차량 전방 이미지 i_f, 속도 v, GPS 정보, heading 과 그때 사람이 결정한 조향각 δ을 기록하여 데이터셋을 구성하였다. 이때, 데이터들은 맑은 날, 흐린 날, 비가 오는 날, 그리고 밤의 다양한 주행 환경에서 취득하였다.

i_f 는 해상도 640×480 의 원본 RGB 이미지를 관심 영역에 대해 크로핑(Cropping) 및 흑백 변환하여, 64×200 로 해상도를 조정하여 녹화하였다. 고정밀 지도 이미지 i_m 은 전역 경로 정보를 포함한 고정밀 지도를 차량의 heading 에 따라 회전시키고, GPS 측정치를 기준으로 전방 50m, 측방 각 25m 씩 관심 영역을 설정하여 크로핑한 후, 100×100 RGB 이미지로 변환하여 출력 및 취득하였다. 우리는 각 데이터를 15Hz 로 다운 샘플링 및 시간 동기화하여 데이터셋

$D = \left\{\langle i_{f.t}, i_{m.t}, v_t, \delta_{r.t}\rangle\right\}_{t=1}^{N}$ 를 구성하였고, 총 데이터의 개수 N은 100,000 이다. 학습 및 평가 데이터셋은 각각 다른 주행 경로에서 취득한 데이터로 구성하였고, 그 비율은 9:1 로 하였다.

B. 학습

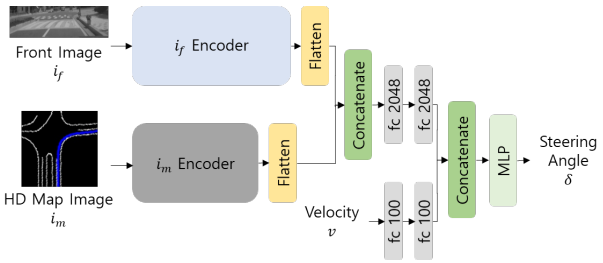


Fig. 1. 제안하는 MIMSNet 구조

본 논문에서 제시하는 MIMSNet 는 i_f, i_m, v로부터 $\hat{\delta} = f(i_f, i_m, v)$을 매핑한다. Fig. 1 과 같이 MIMSNet 은 i_f를 수정된(Modified) VGGNet[5]을 사용하여 $128 \times 4 \times 12$ 크기로 인코딩하고, 인코딩된 특징 맵(feature map)을 최종적으로 벡터화한다. i_m 은 i_f 비해 적은 특징들을 포함하기 때문에, 비교적 얕은 인공 신경망을 거쳐 $48 \times 4 \times 4$로 인코딩 된 뒤, 벡터화 된다. v는 두 개의 fully connected layer 를 거쳐 100 개의 뉴런으로 인코딩 된다. 출력되는 각 벡터들은 병합되어 Multilayer Perceptron(MLP)를 통해 조향각으로 회귀(Regression)된다. 이때, MIMSNet 의 마지막 Fully Connected Layer 을 제외한 모든 layer 이후에는 활성화 함수 Rectified Linear Unit(ReLU)를 적용한다.

MIMSNet 은 식 (1)과 같이 $\hat{\delta}$와 δ_r간의 Root Mean Square Error(RMSE)가 최소가 되도록 네트워크 파라미터 θ를 업데이트하며 학습한다.

$$\underset{\theta}{\text{minimize}}\ \text{RMSE}(f(i_f, i_m, v; \theta), \delta_r) \qquad (1)$$

Minibatch는 16 개이고, θ는 learning rate가 0.0001 인 조건에서 Adam 을 통해 최적화된다. 본 논문에서는 총 50 에포크의 과정을 NVIDIA GeForce GTX 1080 Ti GPU 를 사용하여 수행하였다.

C. 평가

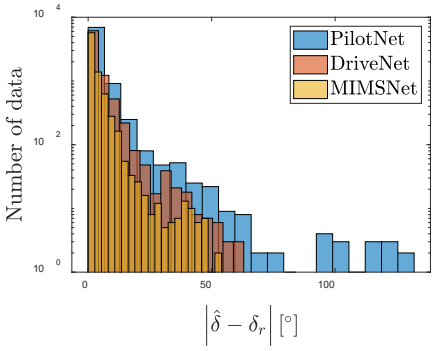


Fig. 2. PilotNet, DriveNet, 그리고 MIMSNet 의 평가 결과 오차 히스토그램

이 절에서는 학습된 MIMSNet, [1]에서 제시한 PilotNet, 그리고 [2]에서 제시한 DriveNet 을 평가 및 그 성능들을 비교한다. 우리는 공정한 비교를 위해 PilotNet 을 MIMSNet 의 i_f 인코더에 MLP 를 추가한 것으로 수정하였고, DriveNet 은 수정된 PilotNet 의 두번째 fully connected layer 출력 뉴런에 방향 지시등 정보를 병합하는 구조로 대체하였다. 방향 지시등 정보는 [2]에 제시된 바와 같이 100 배 스케일링하여 입력하였다. 우리는 MIMSNet 과 기존 2 가지 기술에 대해 참조 조향값과 평가 결과 조향값 간 차이의 절댓값을 계산하고, 각 값의 범위에 해당하는 데이터의 개수를 나타내는 히스토그램을 Fig. 2 에 도시하였다.

Fig. 2 의 성능 결과를 통해 MIMSNet 은 고정밀 이미지의 특징을 조향값 매핑에 반영함으로써, PilotNet 이 내포하고 있는 의사 결정 다양성 문제를 해결함을 알 수 있다. 또한, 히스토그램 상 MIMSNet 의 출력 오차가 DriveNet 에 비해 낮은 구간에 분포해 있다는 점에 대해, MIMSNet 이 고정밀 지도의 자세한 도로 지형 정보 및 전역 경로의 상대적인 곡률로부터 적절한 조향값을 추정할 수 있다는 것을 증명할 수 있다. Fig 2 에는 도시하지 않았지만, MIMSNet 은 교차로 회전을 위한 차선 변경과 같은 다양한 환경에서의 적절한 행동을 출력함을 실증을 통해 입증할 수 있었다.

Ⅲ. 결론

본 논문에서는 모방 학습 기반 종단간 자율 주행의 의사 결정 다양성 문제를 해결하기 위해 MIMSNet 을 제시하였다. 또한 직접 제작한 데이터셋을 사용하여 MIMSNet 과 기존 네트워크의 조향각 추정 성능을 비교하였고, 이를 통해 MIMSNet 의 우월성을 입증하였다. 본 연구에서는 한국과학기술원 내의 주행 데이터만을 사용하였지만, 미래에는 학습 데이터의 다양성을 제고하여 변화하는 환경에 강인한 자율 주행이 가능할 것으로 예상된다.

ACKNOWLEDGMENT

이 논문은 2021 년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2021R1A2C300837011, 예상 못한 상황에서 빠르게 적응하는 메타-강화학습기반 End-to-end 자율주행 기술 개발).

참 고 문 헌

[1] Bojarski, M. "End to end learning for self-driving cars," arXiv preprint arXiv:1604.07316, 2016.

[2] Hubschneider, C. “Adding navigation to the equation: Turning decisions for end-to-end vehicle control,” IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), 2017.

[3] Codevilla, F. “End-to-end driving via conditional imitation learning,” IEEE International Conference on Robotics and Automation (ICRA), 2018.

[4] Amini, A. “Variational end-to-end navigation and localization,” International Conference on Robotics and Automation (ICRA), 2019.

본 논문에서 실시한 실험 영상은 https://bit.ly/3hYGCSM 에서 확인할 수 있다.

[5] Simonyan, K. “Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv:1409.1556, 2014.

고급 챗봇 서비스를 위한 인공지능 마이크로서비스풀 공통 플랫폼

손승철, 고석갑, 이병탁
한국전자통신연구원

{sson, softgear, bytelee}@etri.re.kr

AI microservice pool common platform for advanced chatbot service

Seung-Chul Son, Seokkap Ko, Byung-Tak Lee
Electronics and Telecommunications Research Institute (ETRI)

요 약

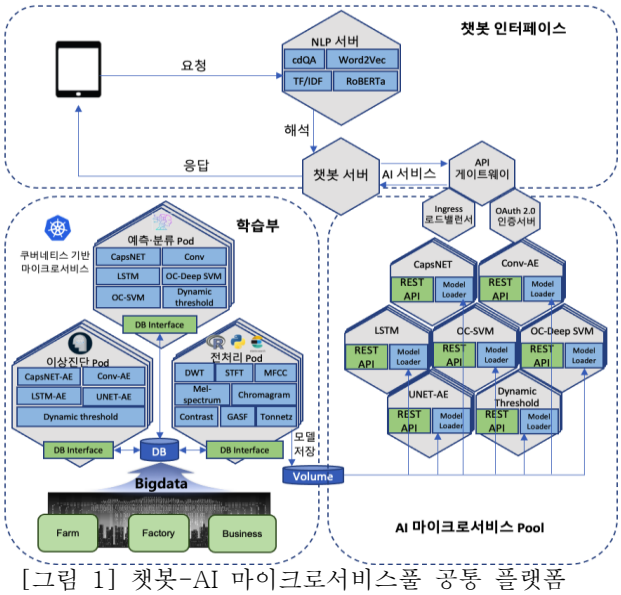
챗봇은 인간 및 컴퓨터간 상호작용을 위한 중요한 인터페이스 수단이다. 인공지능 기술이 성장하면서 챗봇을 이용한 서비스도 규칙 기반에서 벗어나 첨단 인공지능 서비스를 갖춰야 한다. 본 논문에서는 챗봇을 인공지능과 결합하여 산업 전반에 걸쳐 고급 챗봇 서비스를 제공할 수 인공지능 마이크로서비스풀 공통 플랫폼을 제안한다. 제안된 플랫폼이 마이크로서비스를 채택하기 때문에 플랫폼을 구성하는 다양한 기계학습과 목표서비스를 쉽게 배치하고 확장할 수 있으며, 새로운 인공지능 모델도 빠르게 적용할 수 있는 특징을 갖는다.

Ⅰ. 서 론

챗봇은 인터넷을 통해 인간과의 대화를 시뮬레이션 하도록 설계된 컴퓨터 프로그램으로 정의되며, AI(Artificial intelligence)의 급성장으로 사람과 컴퓨터간 상호작용을 위한 중요한 인터페이스 수단으로 사용되고 있다[1]. 최근에는 딥 러닝 및 NLP(Natural Language Processing) 기술보다 진보된 형태의 감성인지 기술 및 데이터 정형화 기술 등을 통해 챗봇 서비스는 사회 전 분야로 확대되고 있다. 챗봇 서비스는 지식 영역, 목표 서비스, 입력 처리 및 응답 생성 방법, 인적 지원 및 구축 방법과 같은 다양한 방식으로 분류할 수 있으며[2], 일반적으로 요청의 이해, 작업의 실행, 정보의 검색, 결과의 게시로 세분화되어 설계 및 구현될 수 있다[3]. 하지만 챗봇이 실제 인공지능 응용서비스와 결합되기 위해서는 데이터 과학자, 서비스 제공자, 서비스 수요자를 아우르는 공통의 통합 플랫폼이 요구된다. 본 논문에서는 숙박시설, 농장, 공장 등의 다양한 산업분야에 응용할 수 있는 이상진단, 예측, 분류 기능을 지원하는 공통 챗봇 마이크로서비스 플랫폼의 설계 및 구현에 초점을 둔다. 제안 플랫폼은 일반적인 챗봇 서비스의 설계 사상에 인공지능을 접목시킨 진보적인 형태의 아키텍처이다. 서비스 수요자 요청을 이해하기 위하여 NLP 를 활용하고, 작업의 실행과 정보의 검색을 위하여 이상진단, 예측, 분류를 위한 다양한 인공지능을 마이크로서비스 및 API(Application Programming Interface) 형태로 제공하여 목표 서비스에 쉽게 응용될 수 있도록 구성하였다.

Ⅱ. 본론

인공지능 기술은 빅데이터, 통신 및 컴퓨팅 기술의 비약적인 발전과 더불어 2012 년 이전에는 성능이 2 년마다 2 배씩 증가하였으나 이후에는 3~4 개월 마다 2 배씩 증가하여 무어의 법칙보다 약 7~8 배 이상 빠르게 발전하는 양상을 보인다.[4] 이러한 비약적 발전에는 인공지능 기술의 3 대 핵심요소인 데이터, 컴퓨팅 자원, 알고리즘이 모두 포함된다. 현대의 빅데이터 기술, 고성능 GPU(Graphics Processing Unit) 기술은 핵심요소 중 데이터 및 컴퓨팅 자원의 발전에 견인차 역할을 각각 해주었으며 마지막 핵심 요소인 알고리즘의 폭발적인 성장을 돕고 있다.



[그림 1] 챗봇-AI 마이크로서비스풀 공통 플랫폼

이번 장에서는 비약적으로 발전하고 있는 인공지능 알고리즘이 챗봇 서비스와 연동되어 산업 전반에 적용될 수 있는 플랫폼을 제안하고 그 플랫폼을 이용한 하나의 응용서비스로써 숙박시설의 이상여부를 진단하는 챗봇 서비스를 소개한다. 플랫폼은 서비스의 배치 용이성 확장성, 고가용성을 기본으로 제공해야 한다. 뿐만 아니라 좋은 성능의 인공지능 모델이 새롭게 개발되었을 때 신속한 적용도 가능해야 한다. 최근, 복잡한

애플리케이션을 논리적으로 분리하여 마이크로 서비스화 하고, 그것들을 느슨한 연결로 분산하는 서비스 지향적 MSA(Microservice Architecture)가 각광을 받고 있다[5]. 웹 서비스 제공 업체는 새로운 MSA 및 클라우드 컴퓨팅 패러다임을 수용하고 클라우드 데이터 센터로 애플리케이션을 빠르게 마이그레이션하고 있다. 마이크로서비스는 클라우드 기반 서비스를 위한 현대적 도구이자 시장 표준으로 자리매김 하였다. 이러한 마이이크로서비스는 제안 플랫폼에서 지향하는 배치 용이성, 확장성, 고가용성 및 모델 신속적용성을 제공하는 가장 적합한 기술이다. [그림 1]은 제안하는 고급 챗봇 서비스를 위한 인공지능 마이크로서비스풀 공통 플랫폼의 개요도 이다. 플랫폼은 크게 학습유닛, AI 마이크로서비스풀, 챗봇 인터페이스로 구성된다. 학습유닛은 수집된 다양한 산업 빅데이터를 통해 학습을 수행하기 위한 데이터 전처리 마이크로서비스풀을 제공한다. 전처리 마이크로서비스풀에는 DTW(Dynamic Time Warping), STFT(Short Time Fourier Transform), MFCC(Mel-Frequency Cepstral Coefficient) 등을 내장하였다. 이러한 전처리 기술은 수집된 데이터를 인공지능 학습에 적절하도록 데이터를 가공한다. 다음으로 제안 플랫폼은 예측 및 분류를 위한 인공지능 학습 마이크로서비스풀을 제공한다. 가공된 데이터의 레이블을 통한 지도학습 알고리즘이 주를 이룬다. 여기에는 CapsNET, conv, LSTM(Long short-term memory), OC-SVM(One-Class Support Vector Machine)등이 포함된다. 마지막으로 이상진단용 마이크로서비스풀은 오토 인코더 계열 및 dynamic threshold 로 구성되며 데이터에서 비정상 상황을 탐지하기 위한 인공지능 모델 학습 메커니즘을 제공한다. 학습유닛의 각 마이크로서비스들의 사용자는 데이터 과학자이며 그들이 전처리 및 인공지능 모델 학습을 용이하게 하기 위한 인터페이스를 별도로 제공하고 학습이 완료된 인공지능 모델은 공유 볼륨에 저장된다. AI 마이크로서비스풀은 학습유닛에 의해 학습된 모델을 실제 서비스에 적용하기 위한 마이크로서비스들로 구성된다. 각 마이크로서비스는 서비스 타겟에 맞춘 학습 모델 로더와 REST API 로 구성된다. 모든 서비스는 마이크로서비스로 구성되기 때문에 새로운 서비스의 배치, 기존 서비스의 갱신과 재배치, 서비스의 삭제 등을 유연하게 진행할 수 있다. 마지막으로 챗봇 인터페이스는 사용자 요청을 수신하고 해석하기 위한 NLP 마이크로서비스로 구성된다. cdQA(Closed Domain Question Answering), Word2Vec, TF/IDF(Term Frequency/Inverse Document Frequency), RoBERTa(Robustly Optimized BERT Pretraining Approach) 알고리즘을 탑재하여 사용자의 요청을 해석한다. 챗봇서버 마이크로서비스는 AI 마이크로서비스풀의 API와 통신하여 해석된 요청에 대한 실제 질의를 수행하고 사용자에게 응답한다. [그림 2]는 제안하는 플랫폼을 활용하여 구현한 스마트 스테이로써, 숙박시설 운영자를 목표로 구현한 서비스이다. 스마트 스테이가 제공하는 사용자 UI 페이지는 메인 홈, 방 목록, 챗봇, 로그기록 이다. 메인 홈은 숙박시설의 기본적인 정보를 비롯하여 전체 방의 이상 상태 및 대여/공실 상태들을 게시한다. 방 목록 페이지는 각 방의 세부 정보인 환경정보, 키 삽입/탈거 시간, 대실/숙박/공실 현황 등을 게시한다. 챗봇 페이지는 사용자의 요청에 대해 실시간 정보를 제공하거나 인공지능기반 이상진단 서비스를 제공한다. 그림 2 의 대화 예를 보면, 챗봇 서비스는 사용자의 현재 온도 요청에 대하여 유사 문장의 의미와 맥락을 해석하여 적절하게 응답하고 AI 마이크로서비스 풀과 결합된 고급 챗봇 서비스를 통해 이상진단도 제공한다.



[그림 2] 스마트스테이 챗봇 서비스

Ⅲ. 결론

본 논문에서는 숙박시설, 농장, 공장 등 다양한 산업분야에 이상진단, 예측, 분류 서비스를 제공할 수 있는 고급 챗봇 서비스를 위한 인공지능 마이크로서비스 풀 공통 플랫폼을 제안하였다. 플랫폼은 마이크로서비스 개발 방법론을 기반으로 구현되었기 때문에 개별 서비스의 쉬운 배치, 확장 및 새로운 NLP 기술이나 인공지능 모델의 신속한 갱신이 가능하다. 제안 플랫폼 기반의 스마트스테이 서비스 예제를 소개하였으며, 다양한 산업분야의 서비스 예를 들어, 스마트 팜, 스마트 공장, 스마트 상점 등에도 인공지능 기반의 고급 챗봇 서비스가 제공될 수 있다.

ACKNOWLEDGMENT

본 연구는 산업통상자원부(MOTIE)와 한국에너지기술평가원(KETEP)의 지원을 받아 수행한 연구 과제입니다. (No. 20181210301570)

참 고 문 헌

[1] Bansal, H., & Khan, R. (2018, April). A review paper on human computer interaction. International Journals of Advanced Research in Computer Science and Software Microservice Architecture: in Aligning Principles Practices and Culture Engineering, 8, pp. 53-56.

[2] Adamopoulou, E., & Moussiades, L. (2020, June). An overview of chatbot technology. In IFIP International Conference on Artificial Intelligence Applications and Innovations, Springer, Cham, pp. 373-383.

[3] Ahmad, N. A., Che, M. H., Zainal, A., Abd Rauf, M. F., & Adnan, Z. (2018, August). Review of chatbots design techniques. International Journal of Computer Applications, 181(8), pp. 7-10.

[4] Raymond Perrault et al., The AI Index 2019 Annual Report, AI Index Steering Committee, Human-Centered AI Institute (Stanford, CA: Stanford University, 2019), p. 66

[5] I. Nadareishvili, R. Mitra, M. McLarty, & M. Amundsen., "Microservice Architecture: in Aligning Principles Practices and Culture", O'Reilly Media, 2016.

SuMul 모듈을 이용한 퓨-샷 객체 검출 방법

임동주, 정순기
경북대학교 IT대학 컴퓨터학부
dddd1226@gmail.com, skjung@knu.ac.kr

Few-Shot Object detection with SuMul module

Dong Ju Im, Soon Ki Jung
Kyoungpook National University School of Computer Science and Engineering

요 약

본 논문은 퓨-샷 객체 검출 방법을 연구했다. AI의 궁극적인 목표인 인간처럼 사고하고 행동하는 것은 하위 단계인 딥러닝도 적용되어야 한다. 퓨-샷의 목표는 AI의 목표와 비슷하다. 퓨-샷 러닝은 적은 데이터를 이용하여 훈련하고 인간처럼 구분할 수 있도록 학습한다. 과적합 및 정확도를 해결하기 위해 샴 네트워크 기반으로 연구하며 학습하는 방법을 학습하기 위해 퓨-샷 학습 방식인 에피소딕 훈련 방법을 이용하여 훈련 및 테스트를 진행한다. FSOD 네트워크를 참조하고 본 논문에서 제안된 SuMul 모듈을 이용하여 정확도를 확인한다.

1. 서 론

현존하는 객체 검출 모델은 많고 다양하다. 하지만 모든 객체 검출 모델은 두 종류 분류가 가능하다. 객체의 경계 상자와 분류를 동시에 진행하는 1-Stage Detector와 객체의 경계 상자를 먼저 얻어내고 그 정보를 바탕으로 객체의 분류를 진행하는 2-Stage Detector가 있다. 두 종류의 대표적인 모델로 1-Stage Detector에는 YOLO가 있고, 2-Stage Detector에는 Faster RCNN이 있다. 모든 객체 검출 모델은 각자의 장점이 있고 단점이 존재하지만, 하나의 공통점을 가지고 있다. 모두 훈련에 필요한 데이터가 많다는 점이다.

AI의 궁극적인 목표는 인간처럼 사고하고 행동하는 것이다. 하위 단계인 딥러닝은 조금의 정보만으로 구분할 수 있는 인간처럼 사고 모델을 가지고 있다. 이러한 문제점을 해결하기 위해서 학습하는 방법을 학습하여야 한다. 이것이 퓨-샷 러닝이며 이를 이용한 객체 검출을 적은 데이터를 이용한 객체 검출 방법이라 한다.

2. 관련 연구

2.1 퓨-샷 러닝 (Few-Shot Learning)

기존 딥러닝 학습에는 수백 혹은 천개 이상의 데이터를 이용하여 학습하여야만 과 적합 문제나 정확도 그리고 딥러닝 모델의 신뢰성을 확립할 수 있다. 하지만, 이러한 방식은 인간처럼 사고하고 행동한다는 AI의 의의와는 다르다. 실제 인간은 어린아이도 사물을 구분하는 방법을 가르쳐 주지 않아도 구분해 내는 것을 알 수 있다. 퓨-샷 러닝은 이러한 문제점을 해결하고자 설계되었다.

퓨-샷 러닝은 훈련에 사용되는 데이터의 수가 한정되어있다. 1장만을 이용해 훈련하는 원-샷 러닝, 최대 100개의 클래스와 각각 25장의 데이터를 이용하는 퓨-샷 러닝, 데이터를 사용하지 않는 제로-샷 러닝이 있다.
각 훈련 마다 각기 다른 훈련 방법이 존재하며 그중 퓨-샷 러닝은 일반적으로 메타러닝 방법을 이용하여 훈련 및 테스트를 진행한다.

이전의 경험으로부터 새로운 태스크를 학습하는 방법을 학습하는 것이 메타러닝이다. 퓨-샷 러닝 모델이 완전히 새롭게 주어진 데이터에서도 잘 작동하도록 만들기 위해서는 메타 러닝이 필요하다. 퓨-샷 러닝 태스크에서는 에피소딕 훈련 (episodic training) 방식으로 메타 러닝을 시도한다. 에피소딕 훈련은 퓨-샷 태스크와 유사한 형태의 훈련 태스크를 통해 모델 스스로 학습 규칙을 도출할 수 있게 함으로써 일반화 성능을 높일 수 있게 한다.[1]

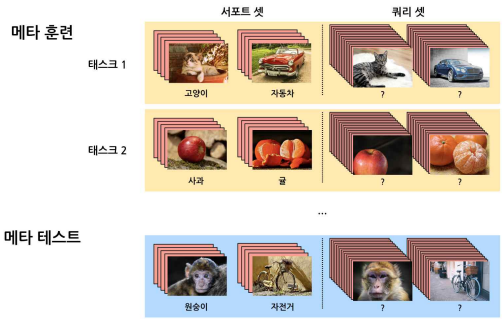


그림 1 에피소딕 훈련 방법[2]

2.2 샴 네트워크 (Siamese Neural Network)

샴쌍둥이라는 신체의 일부를 공유한다는 특징에서 착안된 네트워크로 동일한 두 네트워크의 구조가 서로 닮았으며 가중치를 공유한다.[3] 2005년 처음 구조가 발표되었으며 2015년에 Neural Network를 접목시킨 샴 네트워크가 발표되었다.

이 네트워크는 두 데이터의 특징을 추출하고 서로 비교하여 두 데이터가 같은 종류의 데이터인지 아닌지를 계산하고 결과를 도출해낸다. 적은양의 데이터를 이용하기 때문에 비교 군이 필요한 Few Shot을 이용한 네트워크에서 사용되고 있다. 본 논문에서도 해당 네트워크를 이용하여 연구를 진행한다.

3. 연구 내용

3.1 연구 방법

본 논문에서는 특징을 추출하기 위해 그림1과 같은 방법을 이용하였다.

기본적인 구조는 FSOD[4]에서 제안된 네트워크 구조를 사용한다. 비교군으로 사용될 서포트 데이터(Support data)와 객체를 찾으려는 쿼리 데이터(Query data)가 같은 가중치를 공유하여 특징을 추출한다. 그리고 Attention RPN모듈을 이용하여 추출된 쿼리 특징을 강화하여 얻어낼 수 있다. 그 다음 결과물을 이용하여 그림2의 SuMul 모듈에 적용하여 실험을 진행한다. 요소별 합과 곱을 이용하여 쿼리 특징에서 추출된 특징을 조금 더 강조하고 마지막에 쿼리 특징과 결과를 연결 및 FC작업과 활성화 함수 중 하나인 ReLu를 적용하여 찾고자 하는 특징만 추출해낸다.

실험은 총 3가지 항목을 진행한다. 첫 번째는 요소별 합과 쿼리 특징만을 이용하여 Concat, 두 번째는 요소별 곱과 쿼리 특징만을 이용하여 Concat, 마지막으로 SuMul를 이용하여 실험을 진행한다. 실험으로 확인할 것은 SuMul 모듈로 적용한 실험 결과가 정확도 향상에 기여를 하는지와 단일 결과에 비해 얼마나 정확도가 향상하는지가 이번 실험의 목표이다.

3.2 SuMul 모듈

퓨-샷 네트워크는 서포트를 데이터를 최대한 이용하여 쿼리 데이터에서 원하는 객체를 찾아내야 한다. 이를 위해 SuMul 모듈을 제안한다. SuMul 모듈은 Yang Xiao 논문의 Aggregator 모델을 참조하여 제작하였다.[5] 요소별 합성 합(Element Wise Sum)과 요소별 합성 곱(Element Wise Multiplicate)을 이용하여 객체의 특징을 극대화하기 위해 제안하였다. 쿼리 특징을 기준으로 서포트 특징을 요소별로 더하고, 곱한 결과를 원본 쿼리 특징과 Concat하여 결과를 도출해 낸다. 요소별 합은 세세한 부분의 특징을 검출할 수 있게 해주며, 요소별 곱은 강조될 부분을 더욱 강조하고 원본을 이용하여 기존에 추출된 특징에서 너무 벗어나지 않게 한다.

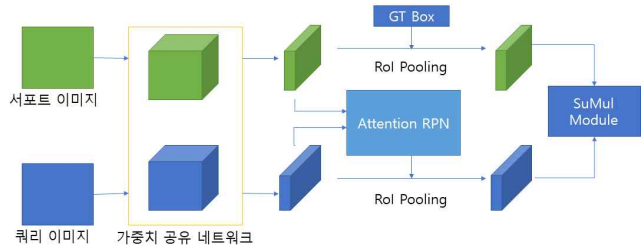


그림 2 FSOD 네트워크 구조에 SuMul 모듈을 합성

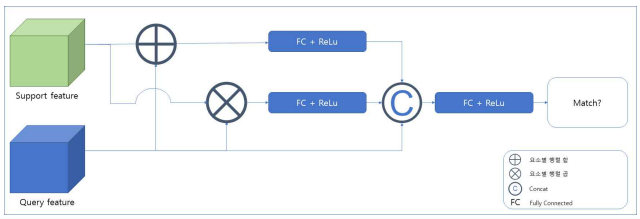


그림 3 제안된 SuMul 모듈

3.3 연구 결과

본 실험에서는 COCO 데이터셋에서 20개의 클래스와 그에 해당하는 이미지 5장씩 총 100장을 이용하여 훈련을 진행하고, 테스트 하려는 이미지에 해당하는 클래스를 가진 이미지를 이용하여 테스트를 진행했다. 테스트에 대한 정확도 척도는 AP(Average Precision)을 이용하였으며 결과는 표1과 같다.

3.4 연구 분석

테스트 결과 단순히 요소별 합성 합이나 곱을 따로 이용하는 것 보다 두 결과를 합성하여 이용하는 것이 결과가 좋게 나오는 것을 확인할 수 있다. 하지만, 기존 네트워크의 mAP는 11인 것을 생각하면 좋은 결과라 할 수 없다.

SuMul 모듈은 서포트와 쿼리의 요소별 합과 요소별 곱을 이용하기 위해 서포트 특징을 GAP(Global Average Pooling)을 이용하여 채널별 평균을 구한 뒤 해당 값을 쿼리에 적용하는 방식으로 진행하였다. 기존 네트워크의 경우 본 논문의 SuMul과는 다르게 GAP 없이 각 픽셀에 해당하는 결과들의 관계에 따라서 객체 검출을 진행하였기 때문에 더 나은 결과를 도출해 낸 것이다.

	mAP	AP50	AP75
Sum	5.93	13.92	4.12
Mul	6.80	15.76	5.01
Sum + Mul	7.59	16.41	6.37

표 1 테스트 결과

4. 결론 및 향후 연구

본 논문에서는 적은 데이터를 이용한 객체 검출 모델에서 특징을 더욱 극대화하기 위해 SuMul 모듈을 제안하였다. 제안된 모델은 추출된 서포트 특징을 최대한으로 이용하기 위해 요소별 합과 곱을 이용하였고 추출된 원본 특징도 이용하여 원본에서 최대한 벗어나지 않게 만들었다. 기존 네트워크의 mAP가 11인 것을 생각한다면, 결코 좋은 결과라고 할 수 없다. SuMul 모듈의 서포트의 이용 방식을 GAP로 채널 별 평균을 구하는 것이 아닌, 다양한 네트워크를 참조 및 연구하여 더 나은 방향으로 개선할 것이다.

향후 연구는 추출된 특징을 이용하여 객체 검출을 극대화하는 방법이 아닌, 특징 추출부터 다시 설계하여 서포트를 최대한으로 이용하는 방법을 연구할 예정이다.

ACKNOWLEDGMENT

이 연구는 한국연구재단(NRF-2019R1A2C1010786) 사업의 지원을 받아 수행된 연구임

참 고 문 헌

[1] Li, Da, et al. "Episodic training for domain generalization." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.

[2] 이수경, "1.퓨-샷 러닝(few-shot learning) 연구 동향을 소개합니다.," kakaobrain, 2019.11.06., (https://www.kakaobrain.com/blog/106).

[3] Koch, Gregory, Richard Zemel, and Ruslan Salakhutdinov. "Siamese neural networks for one-shot image recognition." ICML deep learning workshop. Vol. 2. 2015.

[4] Fan, Qi, et al. "Few-shot object detection with attention-RPN and multi-relation detector." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.

[5] Xiao, Yang, and Renaud Marlet. "Few-shot object detection and viewpoint estimation for objects in the wild." European Conference on Computer Vision. Springer, Cham, 2020.

LSTM 순환신경망을 이용한 비선형 시변 전력데이터 예측 및 분석기술에 관한 연구

최 성 수, 이 효 동
한국전기연구원

sschoi@keri.re.kr

A Study on the Prediction and Analysis of Nonlinear Time varying Power Data Using A LSTM Recurrent Neural Network

Sungsoo Choi, Hyo-Dong Lee
Korea Electrotechnology Research Institute (KERI)

요 약

본 논문은 인공지능(Ai) 알고리즘의 딥러닝 방법중 하나인 순환신경망(RNN)기법을 사용하여 전력시스템의 응답지연시간보상이 가능한 실시간 비선형 시변 전력데이터 예측과 분석에 관한 연구방향을 제시한다. 본 연구에서는 3개의 TSC와 2개의 TCR의 싸이리스터 스위칭 기능을 갖는 300MVar SVC (Static Var Compensator)가 설치된 6,000MVA 735kV 임의 계통을 대상으로 하고 있으며, 계통모선의 PCC(Point of Common Coupling)로 부터 전력데이터를 계측하여 딥러닝 학습, 예측과 분석이 가능하도록 하였다. 모의 계통에서 생성된 계측 전력데이터로 부터 장단기기억 (LSTM) 셀로 구성된 RNN딥러닝 구조를 설계하고 학습 데이터 전처리와 딥러닝 모델 도출을 위해 전력데이터의 최적 측정구간과 예측샘플 간격을 포함한 딥러닝 하이퍼 파라미터 최적화를 수행하였다. 또한 계측된 전력데이터와 예측데이터에 대한 이상진단을 돕기위해 IEEE519-2014 표준을 따르는 종합 고조파 왜형률(THD) 분석기능을 함께 구현하였다. 개발된 딥러닝 모델로 실험한 결과, 비선형 시변 전력데이터 예측성능은 약 96.45%로 매우 높은 정확성을 보여주었다.

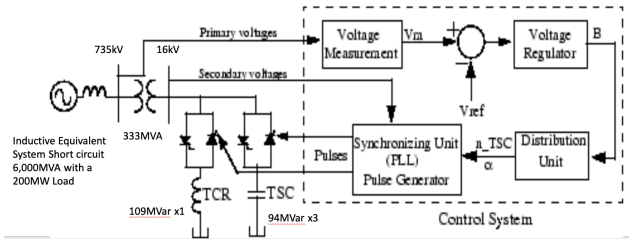
I. 서 론

최근 대규모 재생에너지 단지와 산업플랜트 확대에 따라 수변부하, 전력변환기기, 인버터 등 비선형 부하가 늘어나고 있는 추세이다. 비선형 부하에 전력을 공급하면, 부하에 흐르는 전류의 파형이 왜곡되고, 이에 의한 전압의 파형왜곡을 발생시키며 계통에 유입되는 고조파전류로 인한 전력계통에 연결된 콘덴서등에 과전류에 의한 과열로 전력설비관리의 어려움과 오동작을 일으키게 되는 원인이 되고 있다[1]. 동시에 교류계통은 송, 수신 양단의 전압 상차각에 따라서 전력을 수송하고 있지만 송전선이 장거리화하면 위상각이 증대하여 탈조, 난조에 이르게 되는데 중간점의 전압을 SVC에 의하여 유지하면 과도안정도가 대폭 향상된다. 대체로 공급변전소의 계통모선을 PCC 지점으로 하며 송전계통의 설비고장 등 계통무효전력 수급 불균형에 대비하고, 정상운전 및 상정고장시 안정도 향상을 목적으로 운영한다[2]. 그러나 기존 SVC PI 제어기는 전압안정화를 위한 처리지연으로 전력데이터가 계측되는 시점의 크기와 위상값이 현 제어시점과 틀어질 수 있는 문제를 가지고 있다. 제어하는 시점이 약간만 틀어져도 이에 따라 오히려 변형된 위상투입으로 인한 공칭값의 수십여배 큰 과전류가 발생할 수 있다. 이러한 문제점을 해결하기 위해 본 연구에서는 인공지능(Ai) 알고리즘의 딥러닝 방법중 하나인 순환신경망(RNN)기법을 사용하여 비선형 시변 전력데이터값을 미리 예측을 하여, 전력시스템 SVC 제어부의 응답지연시간을 획기적으로 줄이고자 한다. 또한 계측 및 예측된 전력데이터가 IEEE 519표준에서 정하는 고조파 전압 및 전류제한에 따르고 있는지 비교 분석해 보고자 한다.

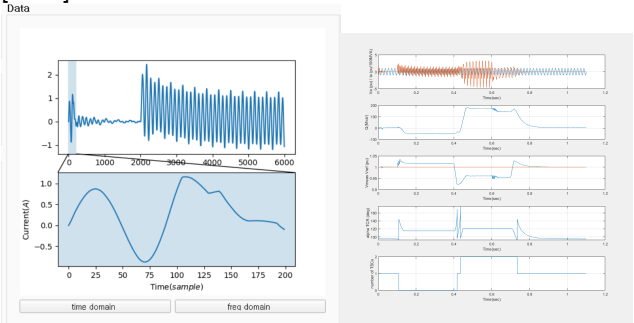
II. 본론

비선형 시변 전력데이터 예측 및 분석 대상

본 연구의 계통시험모델은 그림1에 보인 바와 같이 기본주파수 60Hz에서 200MW 인덕터 단락등가회로를 갖는 부하와 300MVar SVC (Static Var Compensator)가 설치된 6000MVA 735kV 계통을 대상으로 한다. SVC 설비는 109MVar TCR 한 개와 94MVar TSC 세 개로 구성된다. SVC 제어시스템은 계통모선 1차측으로 부터 전압값을 측정하여 싸이리스터 제어각 펄스들을 생성하고 싸이리스터 스위치 On/Off 신호를 발생시킨다. SVC 제어시스템을 5×10^{-5}샘플링시간으로 1.1초동안 연속적으로 운영하며 계통모선의 무효전력(Q)을 투입/방출시킴으로 계통전압을 안정화 시킨결과는 그림2에 보였다.



[그림1] 비선형 시변전력데이터 예측 및 분석을 위한 SVC 실험모델

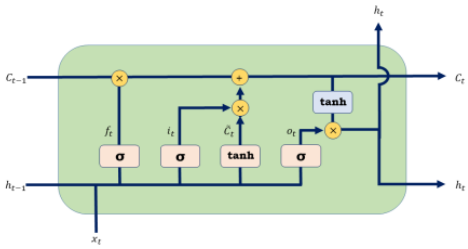


[그림2] 300MVar/-100MVar SVC 계통모의 실험결과: 비선형 시변 전력데이터 I_SVC 샘플일부(좌), 전력데이터 V_pri, I_svc계측값, 무효전력, 전압기준오차 및 TCR펄스,TSC 스위치 제어신호(우)

LSTM 순환신경망 딥러닝 알고리즘 모델 최적화 설계

RNN은 시계열 데이터를 처리하는데 적합한 딥러닝 알고리즘으로 잘 알려져 있다. 시간의 흐름에서 발생하는 데이터를 시계열 시퀀스 데이터라 한다. 이미지 추론으로 잘 알려진 CNN과 같은 신경망 구조에서는 모든 입력과 출력이 독립적으로 동작하기 때문에 시퀀스

데이터를 시간의 흐름에 반영 할 수 없다. 하지만 RNN의 출력 결과는 이전의 계산 결과에 영향을 받기 때문에 시퀀스의 모든 요소마다 적용이 가능하다는 장점이 있다. 그러나 RNN 딥러닝 알고리즘은 시퀀스 데이터 시간의 길이가 길어지고 층이 깊어질수록 학습성능이 크게 떨어지는 단점이 있다. 이를 해결하기 위해 나온 알고리즘이 LSTM셀을 적용한 RNN기법으로, 입력 시퀀스 데이터에 대해 cell state로부터 어떤 정보를 버릴 것인지 말 것인지 정한다. 즉 오래된 기억까지 잃어버리지 않고 전달하여 학습에 사용될 수 있도록 셀 내에 여러 게이트와 필터기능을 추가한 것으로 RNN의 장기기억 의존성문제 (Vanishing gradient problem)를 근원적으로 해결한다[3].



[그림3] 순환신경망을 구성하는 LSTM 셀 기본구조

본 논문에서는 순환신경망의 LSTM셀을 채택하여 변압기 1차측의 계통 모선 PCC에서 1.1초 동안의 i.i.d 통계분포를 갖는 Monte-Carlo 방법으로 비선형 시변 전력데이터를 생성한다. 추출된 전체 전력데이터들 중 80%는 학습데이터로 사용하고 이들 중 20%를 validation, 그리고 나머지 20%를 테스트데이터로 분할하여 딥러닝 엔진모델의 입력값으로 순차적으로 사용한다. 학습모델 구조와 최적의 파라미터 추출은 무효전력보상과 전압안정화를 위해 SVC제어 응답시간 내 처리와 전력데이터를 미리 확보할 수 있도록 설계하였다. LSTM을 이용하여 3개의 층으로 모델을 구성하고, 두 단의 FC 층을 두고 주어진 계층구간에서 부분학습을 진행하였으며 SVC 응답시간 내 처리가능한 시간마진에 해당하는 예측 샘플값을 갖도록 LSTM 모델의 구조는 many to one과 many to many 방식을 사용하였다. 하이퍼 파라미터 최적화를 위해서 Batch Size, Learning rate, 그리고 Optimizer 등의 파라미터 종류를 정한다. Batch Size는 40, 60, 128, 256, 512, 764, 1024로 가변 설정하고, Learning rate는 0.1, 0.01, 0.001, 0.0001, 0.00001로 변화를 줄 수 있도록 하였다. Optimizer 종류는 Adam, Adagrad, Adadelta, SGD, 그리고 RMSprop으로 가변하며 개별적으로 딥러닝 학습의 최적화를 수행한다. 반복되는 학습 결과, 최소의 Loss 값을 갖는 하이퍼 파라미터들을 최종 학습모델에 반영 하였다.

[표1] 최적화된 딥러닝 학습 최종 하이퍼 파라미터 결과

measured_ time	target_ti me	optimi zer	learning_ rate	batch_ num	epoch_ num
50	30	Adam	0.01	1024	20

비선형 시변 전력데이터 측정구간 및 예측 최적화

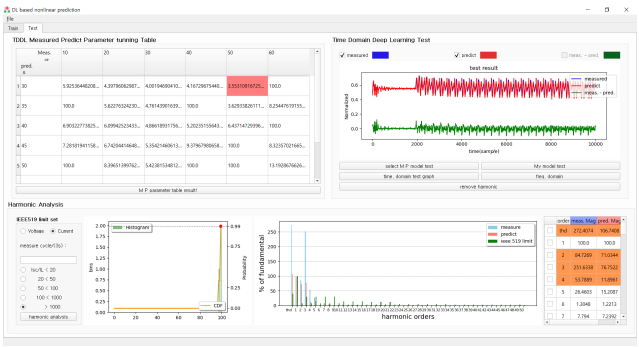
순환신경망 구조설계에 따른 비선형 시변 전력데이터 전처리와 측정구간 및 예측 최적화를 위한 파라미터 설정의 경우 측정구간 설정값을 10~60으로 하고, 예측 샘플간격을 10~60으로 번갈아 입력하며 최적화 실험을 수행하였고, 측정구간 50샘플과 예측 샘플간격 30의 최적 딥러닝 파라미터값을 최종적으로 추출하였으며, 이에 따른 예측 정확도성능은 약 96.45% 를 보였다.

전력데이터 분석을 위한 종합 고조파 왜형률 (THD) 측정

IEEE 519-2014 표준을 따르는 극단시간 고조파 측정을 통하여 THD를 분석한다. 전력데이터 분석을 위한 고조파 측정법으로는 극단시간고조파측정법(very short time harmonic measurements)과 단시간고조파측정법(Short time harmonic measurements)이 있다. 극단시간 고조파 측정법은 IEC 61000-4-30에서 제안한 3초 간격 기준 측정법이다. 60Hz 계통에서는 12cycle, 50Hz 계통에서는 10cycle이 되어야 한다. 극단시간 고조파 측정법은 아래 수식에 보인 바와 같이 측정시간이 24시간이며 12cycle 동안 측정된 값을 15회 연속하여 누적해 3초간의 값으로 고조파를 측정한다.

$$F_{n,vs} = \sqrt{\frac{1}{15}\sum_{i=1}^{15} F_{n,i}^2}\ ,$$

여기서 $F_{n,i}$는 전압(V) 또는 전류(I), n은 고조파 차수, i는 카운팅수, vs는 very short을 뜻한다. 고조파 측정으로 통계적 평가 방법이 사용된다. 통계적 평가 방법은 측정된 전압 및 전류 데이터를 통계적 누적분포 함수로 변환해 백분위 95% 또는 99%가 되는 값을 대표 THD로 사용한다. 대표 THD 값은 IEEE std 519-2014의 규제치 이내에 있어야 한다[4]. 이때 일간 측정 99% 백분위는 극단시간(3sec)값으로 IEEE std.519-2014 전압제한치의 1.5배보다 낮아야 한다. 그림4는 본 연구에서 개발된 비선형 시변 전력데이터 예측 최적화 파라미터 추출결과, 시뮬레이터 수행결과로 예측 전류파형 및 고조파 제한치에 따른 대표THD분석 비교결과를 보여준다.



[그림4] 비선형 시변 전력데이터 예측 및 분석 시뮬레이터 수행 결과 및 프로그램 GUI사진

III. 결론

본 논문에서는 SVC가 설치된 6000MVA 735kV계통모선 PCC에서 비선형 시변 전력데이터 기반 LSTM셀을 적용한 RNN 구조를 설계하였고, 50개의 계측 샘플 시퀀스데이터입력으로 30샘플 간격만큼 앞서서 정확히 예측하는 최적의 파라미터를 추출하였다. 설계된 학습모델을 통한 실험결과, 예측 전력데이터에 대해 96.45% 예측정확도 성능을 가지며 전력시스템에서 약 1.5ms 시간 마진을 확보할 수 있었다. 이는 기존 SVC 제어기에서 처리지연으로 인한 전력데이터가 계측되는 시점의 크기와 위상값이 현 제어시점과 틀어질 수 있는 문제를 주어진 시간내 즉각적으로 해결할 수 있음을 의미한다. 향후, 본 연구를 바탕으로 비선형 시변 전력데이터를 대상으로 하는 딥러닝 알고리즘에 기반한 다양한 온-타임 제어기 설계에 대한 활발한 연구가 기대되어 진다.

ACKNOWLEDGMENT

이 연구는 2021년도 정부(과학기술정보통신부)의 재원으로 국가과학기술위원회의 지원을 받아 수행된 한국전기연구원 기본사업임 (No. 21A01108)

참 고 문 헌

[1] 한국전력공사 계통계획처, "송전계통 고조파 관리기준", June 2018.

[2] 윤종수외, "SVC 설치 운전에 따른 계통의 영향성 평가", 전기학회논문지, 1187-1193, July 2010.

[3] Hochreiter, S. and Schmidhuber, "Long short-term memory". Neural Computation, 9(8), 1725-1780, J., 1997.

[4] IEEE Power and Energy Society, "IEEE Recommended Practice and Requirements for Harmonic Control in Electric Power Systems", IEEE std 519-2014.

연합학습에서 이상치 데이터가 미치는 영향 평가

정성우, 유준혁*
대구대학교, *대구대학교
learningstaedy0j0@gmail.com, *joonhyuk@daegu.ac.kr

Evaluating the Impact of Out of Distribution Data on Federated Learning

Seong-Woo Jeong, Joonhyuk Yoo*
Daegu University, *Daegu University

요 약

중앙 집중화된 전통적인 기계학습 방식은 개인정보 보호 문제와 컴퓨팅 자원이 제한된 에지 디바이스에서의 학습이 어려운 문제가 존재한다. 이를 해결하기 위해 새로운 분산 학습 패러다임인 연합학습에 대한 연구가 최근 활발하게 이루어지고 있다. 본 논문은 연합학습 프레임워크에서 사용자의 실수로 발생할 수 있는 이상치 데이터에 대한 영향력을 분석하기 위해 Non-IID CIFAR-10 데이터세트의 이미지 분류 작업을 위한 연합학습에서 SVHN 데이터 셋을 이상치 데이터로 하여 이상치 데이터가 포함되어 있는 클라이언트 비율과 이상치 데이터 비율에 따른 정확도를 비교한다. 실험 결과, 모든 클라이언트에서 10%의 이상치 데이터 비중을 가지고 있을 때 약 3% 정도의 정확도 감소가 나타나고, 이상치 데이터 비중이 증가하면서 정확도 감소폭도 함께 증가함을 보여준다,

Ⅰ. 서 론

최근 자율주행, 의료, 스마트 시티 등과 같이 개인정보 보호가 중요한 다양한 분야에서 연합학습이 적용되고 있다. 기존의 기계학습 방법은 에지 디바이스에서 생성된 데이터를 서버로 전송하여 중앙 집중화 방식으로 학습하기 때문에 데이터의 개인정보 보호에 대한 우려가 높다. 이와 더불어 에지 디바이스가 생성하는 방대한 양의 학습 데이터를 서버에서 처리하는데 필요한 컴퓨팅 자원이 기하급수적으로 증가하여 시간적/금전적 부담이 급증하였다. 이를 해결하기 위해 McMahan et al.이 연합학습(Federated Learning)이란 새로운 분산 학습 패러다임을 제시한 바 있다[1].

연합학습은 개인정보보호를 위해 클라이언트라 불리는 다수의 사용자 에지 디바이스가 저장하고 있는 원시 데이터를 중앙서버에 전송하지 않으면서 딥러닝 모델을 최적화하는 분산 학습 방법 중 하나이다. 연합학습의 최적화 방식은 그림 1과 같다. 먼저 각각의 클라이언트가 중앙 서버의 초기 모델을 수신하고 로컬 데이터를 통한 확률적 경사하강법(Stochastic Gradient Descent)을 수행한 뒤 생성된 매개변수를 서버로 전송한다. 그런 다음 서버는 모든 클라이언트의 매개변수를 수신한 후에 평균하여 초기 학습 모델을 업데이트 한다. 이러한 일련의 과정을 여러 번 반복 하여 최적화를 진행한다. 훈련에 사용되는 원시 데이터가 클라이언트에서만 관리되는 학습 방식 덕분에 개인정보 보호가 가능해지며 클라이언트의 로컬 데이터 전체를 보내는 것이 아닌 매개변수 값만 전송하기 때문에 상대적으로 서버에서 처리해야 하는 컴퓨팅 자원의 부담이 줄어든다.

하지만, 연합학습은 각 클라이언트에서 사용자 상호 작용에 의해 학습 데이터가 생성되기 때문에 사용자의 실수나 악의로 인한 이상치(OOD, Out of Distribution) 데이터가 발생하기 쉽다. 이상치 데이터란 일반적인 데이터 패턴과 다르게 매우 다른 패턴을 가지고 있는 데이터를 뜻한다. 예를 들면 '개' 이미지 데이터 집합에 '고양이' 이미지가 섞여 있는 경우 '고양이' 이미지를 이상치 데이터라고 볼 수 있다. 이러한 이상치 데이터는 다양한 실험을 통해 딥러닝 모델의 성능을 크게 좌우하는 것이 입증되었고, 이를 해결하기 위해 다양한 연구가 진행되고 있다[2,3]. 따라서 본 논문에서는

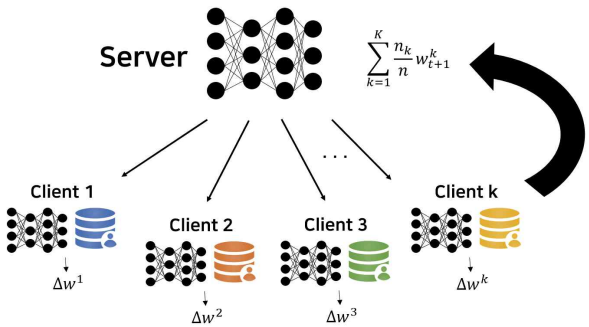


그림 1. 연합학습 최적화 방식

연합학습 프레임워크에서 이상치 데이터가 미치는 영향을 평가하여 문제점을 분석하고 초기 결과를 제시하고, 향후 연합학습에서 이상치 데이터의 적절한 처리방법의 필요성을 제기하고자 한다. 이와 더불어 이미지 분류 작업을 위한 연합학습시 이상치 데이터가 포함되어 있는 클라이언트의 비율과 이상치 데이터의 비율에 따른 영향력을 정확도 차이로 측정하였으며, 연합학습의 실제 적용 시 허용 가능한 목표 이상치 비율을 실험을 통해 제시한다.

본 논문은 다음과 같이 구성된다. Ⅱ장에서는 이미지 분류 작업에서 이상치 데이터가 미치는 영향에 대해 평가하기 위한 실험 방법을 설명한다. Ⅲ장에서는 실험 결과를 분석하고 Ⅳ장에서는 결론을 제공한다.

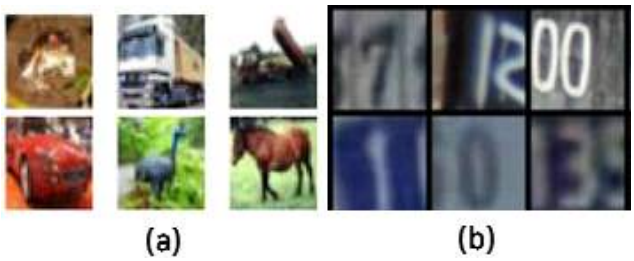
Ⅱ. 실험 평가 방법

이미지 분류를 위한 학습 데이터세트는 Non-IID CIFAR-10 데이터세트를 사용한다. 그림 2(a)에서 볼 수 있는 CIFAR-10 데이터세트는 컴퓨터 비전 연구에서 널리 사용되는 데이터세트 중 하나로 비행기, 자동차, 새, 고양이, 사슴, 개, 개구리, 말, 배 그리고 트럭 클래스에 대한 총 60,000개(학습 데이터 50,000개, 평가 데이터 10,000개)의 32x32 RGB 이미지 데이터이다[4]. Non-IID란 데이터가 독립적이지 않고 동일하지 않은 확률 분

포를 가지고 있다는 의미이다. 연합학습은 물리적으로 떨어진 곳에서 생성된 각 클라이언트의 로컬 데이터를 사용하기 때문에 학습 데이터를 Non-IID 데이터로 상정하고 성능 평가를 진행한다[5]. 따라서 CIFAR-10 학습 데이터세트 50,000개에서 10개의 클래스 중 5개의 클래스를 무작위로 선택하고, 이미지 데이터 수 또한 무작위로 하여 하나의 클라이언트에서 로컬 데이터세트를 생성한다. 이어서 모든 데이터가 포함되도록 총 10개의 로컬 데이터세트를 생성하였고, 이를 Non-IID CIFAR-10이라 정의하였다.

학습 데이터에 포함되지 않은 패턴을 가지는 이상치 데이터로는 SVHN(Street View House Numbers) 데이터세트를 사용한다. 그림 2(b)에서 볼 수 있는 SVHN 데이터세트는 Google이 Google 지도를 만드는 과정에서 촬영한 집들의 번호판 이미지이다. 이는 0부터 9까지 숫자 10개로 이루어져 있고 총 73,257개의 32x32 크기의 RGB 이미지 데이터이다[6].

이상치 데이터의 영향력 평가를 위해 본 논문에서는 SVHN을 Non-IID CIFAR-10 학습 데이터세트에 다양한 비율로 추가하여 실험을 진행하였고, 테스트 데이터세트로는 CIFAR-10의 테스트 데이터세트 10,000개를 그대로 사용하였다.



(a) (b)

그림 2. (a) CIFAR-10 데이터세트의 일부 이미지, (b) SVHN 데이터세트의 일부 이미지

실험 환경으로는, AMD RYZEN™ 7 1700 8-Core 3.0 GHz CPU와 GeForce RTX 2080 TI GPU 2개가 장착되어 있는 서버 컴퓨터를 사용하였고, Ubuntu 20.04 LTS 환경에서 PyTorch ver.1.7.1을 사용하여 실험을 진행하였다. 모델은 ResNet56을 사용하였고, 10개의 클라이언트를 가정하여 시뮬레이션을 진행하였다. 각 클라이언트에서는 10의 배치크기와 0.03의 학습률과 함께 확률적 경사하강법을 사용하여 5 epoch만큼 매 라운드 학습을 진행하였고, 학습된 각 클라이언트의 매개변수를 평균하여 총 100 라운드 동안 서버 모델의 업데이트를 진행하였다.

Ⅲ. 실험결과 분석

표 1은 클라이언트 및 이상치 데이터 비율에 따른 서버 모델의 정확도 값이다. 표에 나타나 있는 클라이언트 비율이란 이상치 데이터를 포함하는 클라이언트의 비율을 나타낸다. 다음으로 이상치 데이터 비율은 각 로컬 데이터세트에 추가된 이상치 데이터의 비율을 뜻한다.

Baseline은 이상치 데이터 없이 Non-IID CIFAR-10 데이터만으로 학습했을 때의 정확도 값이다. 클라이언트와 이상치 비율 모두 10%로 아주 작은 비율을 가지고 있을 때에는 0.1%의 아주 미미한 정확도 감소를 나타냈으며, 모든 클라이언트가 10%의 이상치 비율을 가지고 있을 때 약 3% 정도의 정확도 감소가 나타났다. 이 후 모든 클라이언트에서 이상치 비율의 증가에 따른 비교를 하였을 때 상당히 큰 폭으로 정확도 감소가 일어나는 것을 관측할 수 있었다.

이러한 결과를 바탕으로 연합학습의 실제 적용시 모든 클라이언트에서 이상치 데이터가 생성된다고 가정하였을 때 로컬 데이터세트에서 이상치 데이터가 10%의 비율까지는 허용할 수 있는 오차범위로 간주될 수 있다고 판단된다.

표 1. 클라이언트 및 이상치 데이터 비율에 따른 정확도 비교

Index	Client Ratio (%)	OOD Ration (%)	Acc (%)
Baseline	**0**	**0**	**69.5**
1	10	10	69.4
2	30	10	67.7
3	50	10	69.2
4	70	10	67.2
5	100	10	66.2
6	100	25	61.6
7	100	50	55.4
8	100	75	50.1
9	100	100	48.2

Ⅳ. 결론

본 논문에서는 연합학습시 이상치 데이터의 영향력을 평가하기 위해 클라이언트와 이상치 데이터 비중에 따른 비교실험을 진행하였다. 실험을 통해 모든 클라이언트에서 10%의 이상치 데이터 비중을 가지고 있을 때 약 3%의 허용 가능한 정확도 감소가 나타나는 것을 관측할 수 있었고, 연합학습의 실제 적용시 이를 허용 가능한 목표 비율로 제안하였다. 향후 본 연구를 확장하여 분산학습 프레임워크에서 이상치 데이터를 적절하게 보상할 수 있는 연합학습 알고리즘에 대한 연구를 진행할 계획이다.

ACKNOWLEDGMENT

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2020R1A2C1014768).

참 고 문 헌

[1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, “Communication-efficient learning of deep networks from decentralized data,” In Artificial intelligence and statistics, vol. 54, pp. 1273-1282, Apr. 2017.

[2] D. Hendrycks, M. Mazeika, and T. Dietterich, “Deep anomaly detection with outlier exposure.” arXiv preprint arXiv:1812.04606. Dec. 2018.

[3] D. Hendrycks, and K. Gimpel, “A baseline for detecting misclassified and out-of-distribution examples in neural networks,” arXiv preprint arXiv:1610.02136. Oct. 2016.

[4] A. Krizhevsky, and G. Hinton, "Learning multiple layers of features from tiny images," 2009.

[5] P. Kairouz, B. McMahan, ... & S. Zhao, “Advances and Open Problems in Federated Learning,” arXiv preprint arXiv:1912.04977. Dec. 2019.

[6] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, and A. Y. Ng, “Reading digits in natural images with unsupervised feature learning”, NIPS Workshop on Deep Learning and Unsupervised Feature Learning, Dec. 2011.

시각장애인 길 안내 서비스를 위한 딥러닝 기반 횡단보도 가이드 솔루션 개발

김서현, 최인훈, 박수용, 정홍주, 홍승준, 허의남*
경희대학교

{seok02h, inhun321, clapd10, sub06038, hongsj1022, *johnhuh}@khu.ac.kr

Development of deep learning based crosswalk guidance solution for pedestrian navigation service of the blind

Seo-Hyun Kim, Inhun Choi, Suyong Bahk, Hong-Ju Jeong, Seungjun Hong, Eui-Nam Huh *
Kyung Hee Univ

요 약

최근 시각장애인의 활동반경 확대 및 보행 안전을 위한 길 안내 서비스가 주목받고 있다. 본 논문에서는 그 중 사고 발생률이 높은 횡단보도 보행 상황을 위한 횡단보도 가이드 솔루션을 제안하며, 자원이 제한된 사용자 단말에서의 효율적인 자원 활용과 서비스 정확도 향상을 목표로 한다.

1. 서 론

시각장애인의 실외 보행 방법으로는 안내 보행, 흰 지팡이 보행, 안내견 보행, 전자기기 보행 등이 있다. 보행 방법 중 목적지까지 경로를 안내받을 방법은 전자기기 보행뿐이고, 휠체어를 타거나 손을 사용할 수 없는 시각장애인이라면 현재는 전자기기밖에 다른 단독 보행의 대안이 없다[1]. 이러한 이유로 시각장애인 길 안내 서비스의 필요성이 대두되고 있으며 안경, 앱, 로봇 등 다양한 단말 형태로 길 안내 서비스가 개발되고 있다[2]. 이러한 길 안내 서비스에 포함될 핵심 기능으로 횡단보도 가이드가 있다. 신호등 불빛을 인식하기 어려운 시각장애인을 위해 전방의 횡단보도 여부를 판단하고 신호등 불빛 상황과 횡단 가능 여부를 알려주어 사고 다발 지역인 횡단보도 상황에서 그들의 안전을 보장해주어야 한다.

횡단보도 가이드의 문제점은 대체로 신호등 인식에 있다. 횡단보도는 약 10m 에서 40m 에 육박하기까지, 그 길이가 매우 다양하다. 따라서 횡단보도도 길 건너편의 신호등을 인식할 때, 인식하는 단말로부터 신호등까지의 거리가 불규칙하다. 또한, 길안내 서비스를 주로 탑재하는 사용자 단말은 자원이 제한되어, 고성능을 기대하기 어렵다는 한계가 있다. 따라서 본 논문에서는 제한된 자원을 효율적으로 이용하며, 정확도 및 검출률이 99% 이상인 횡단보도 가이드 솔루션을 제안한다.

2. 본 론

2-1. 신호등 및 횡단보도 인식 모델

본 논문에서는 시각장애인을 위한 길 안내 서비스에 탑재할 수 있는 횡단보도 가이드 솔루션을 제안한다. 횡단보도는 길이가 매우 다양하여 전체 신호등 데이터셋을 이용하여 학습한 모델의 경우 전체적으로 검출률이 감소하고 특히 멀리 있는 신호등 검출률이 현저히 낮아지는 문제가 있다. 횡단보도 길이에 따라 데이터셋을 분리한 후 별도로 학습하여 위 문제를 해결할 수 있다. 횡단보도 길이를 그림 1 과 같이 15m 이하, 15m 초과, 30m 이하, 30m 초과 세 종류로 분류하였으며, 신호등 데이터 셋을 각 6,000 장씩, 총 18,000 장 수집하여, YOLOv5 모델을 통해 학습하였다.



① * 15m 이하 ② 15m 초과, 30m 이하 ③ 30m 초과

(* 횡단보도 길이)

[그림 1] 횡단보도 길이에 따른 신호등 데이터 예시

분리한 세 종류의 데이터셋을 각각 학습하였을 경우와 모든 데이터셋을 이용하여 한 번에 학습하였을 경우의 성능 결과는 표 1 과 같다.

Train Set / Test Set	① Test	② Test	③ Test
①+②+③ Train	Precision : 100 Recall :99	Precision : 99 Recall : 99	Precision : 99 Recall : 82
① Train	**Precision : 100 Recall : 100**	Precision : 99 Recall : 97	Precision : 98 Recall : 72
② Train	Precision : 100 Recall : 99	**Precision : 100 Recall : 99**	Precision : 99 Recall : 92
③ Train	Precision : 99 Recall : 81	Precision : 100 Recall : 98	**Precision : 100 Recall : 99**

Precision (%)
Recall (%)

[표 1] 모델 학습 데이터 및 성능 결과

표 1 에서 ①은 길이가 15m 이하인 횡단보도 건너편에서 촬영한 신호등 데이터셋, ②는 길이가 15m 초과, 30m 이하인 횡단보도의 신호등 데이터셋, ③은 길이가 30m 초과인 횡단보도의 신호등 데이터셋이다. 각 데이터셋을 학습과 테스트를 위해 분리하여 실험을 진행하였다. 모든 데이터셋을 이용하여 한 번에 학습하였을 경우 ③ 길이가 30m 초과인 데이터셋에 대하여 검출률이 낮은 것을 확인할 수 있다. 반면 데이터셋을 ①, ②, ③으로 분리하여 학습한 경우 다른 데이터셋에 대해서 인식률은 낮지만, 학습한 데이터셋과 같은 횡단보도 길이의 신호등 데이터에 대해서는 정확도 및 검출률이 99% 이상으로 높은 성능을 확인할 수 있다. 예를 들어 ③데이터셋을 이용하여 학습한 모델은 ①, ②데이터셋의 테스트셋에 대하여는 성능이 낮지만, ③데이터셋의 테스트셋에 대하여 높은 성능을 보임을 확인할 수 있다. 따라서 횡단보도 가이드에서 학습 모델 하나만을 이용해서 신호등을 인식하는 것보다는 횡단보도 길이에 따라 분리하여 상황에 맞는 인식 모델을 사용하는 것이 좋다는 결론을 도출할 수 있다.

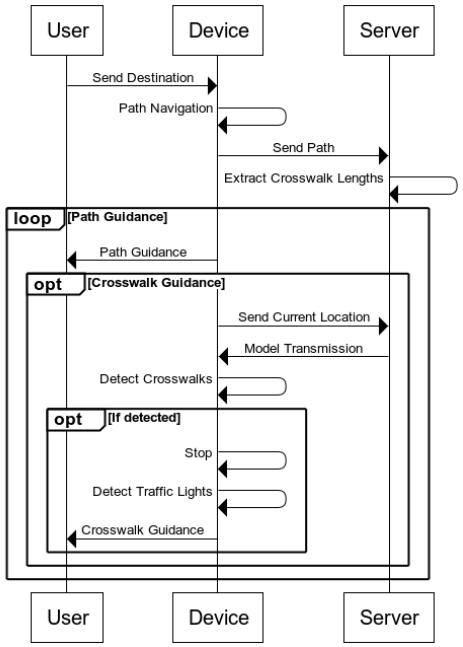
횡단보도 인식 모델의 경우, 20,000 장으로 구성된 횡단보도 데이터셋을 YOLOv5 모델로 학습하였으며, 횡단보도와 신호등 인식 결과는 그림 2 와 같다.



[그림 2] 횡단보도 및 신호등 인식 결과

2-2. 효율적 자원 활용을 위한 횡단보도 가이드 시나리오

횡단보도 가이드 전체 시나리오는 그림 3 과 같다.



[그림 3] 횡단보도 가이드 시나리오

횡단보도 상황에서 신호등 인식 시 앞서 생성한 3 가지 모델 중 하나만 이용하므로, 자원이 제한된 사용자 단말에 3 가지 모델을 전부 저장하는 것은 효율적이지 않다. 따라서 사용자 단말에는 모델을 한 가지만 저장하고, 상황에 따라 적절한 모델을 전송받아 교체하는 방식을 통해 한정된 자원을 효율적으로 활용할 수 있다.

시각장애인 길 안내 서비스에서 사용자가 출발지와 목적지를 입력하였을 때 단말에서 경로가 생성되며, 서버에서 경로 내 횡단보도의 길이를 모두 추출한다. 예를 들어 목적지 안내를 위해 주로 사용하는 TMAP 보행자 경로 안내 API 에서는 경로상 노드 정보를 통해 그림 4 와 같이 횡단보도 정보를 파악할 수 있다.



[그림 4] 경로 내 횡단보도 길이 추출 예시

경로 안내 중 단말에서는 현재 위치를 서버로 전송하고, 서버에서는 사용자가 다음으로 마주할 횡단보도 길이에 따라 적합한 모델을 단말로 전송한다. 단말에서 보행 중 전방의 횡단보도를 탐지하면 정지한 후 서버로부터 전송받은 모델을 이용하여 신호등을 인식한다. 신호등 불빛에 따라 횡단 가능 여부를 사용자에게 안내하고 길을 건넌다.

3. 결 론

전자기기를 통한 시각장애인 길 안내 서비스는 목적지까지의 경로 안내와 장애물회피, 횡단보도 가이드 기능을 모두 제공할 수 있기 때문에, 시각장애인의 활동반경을 확대하고 보행 안전을 보장하기 위해 필요한 서비스이다 [3]. 본 논문에서는 자원이 제한된 사용자 단말에서 정확도가 높으면서도 자원을 효율적으로 사용하는 횡단보도 가이드 솔루션을 개발 및 평가하였으며, 향후 시각장애인 길 안내 서비스에 탑재되어 횡단보도 상황에서 시각장애인의 안전을 보장할 수 있을 것으로 판단된다.

또한 어린이 및 노약자 등의 교통약자를 위한 서비스와 무인 배달 로봇의 서비스로 확장이 가능할 것으로 보인다.

ACKNOWLEDGMENT

이 논문은 2021 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2017- 0-00294, (대학 ICT 기초연구실) 서비스 이동 지원을 위한 분산형 클라우드 핵심원천기술 연구)

참 고 문 헌

[1] 시각장애인의 이해 - 보행, 한국 시각장애인 복지관 홈페이지,
http://hsb.or.kr/client/visually/visually3_3.asp

[2] 김창기, 이상도, 서정민. "스마트폰을 이용한 시각장애인 길안내 시스템." 한국컴퓨터정보학회 학술발표논문집, 22.2: 405-406 2014.

[3] 장영건, 차주현 “시각장애인용 길안내 서비스 시스템에 대한 연구”, 재활복지공학회논문지 v.11 no.4, 2017.

다양한 혼잡도에 적응하는 DQN 기반 Cellular V2X 분산혼잡제어

이은화, 정구선, *최주영, 문철
한국교통대학교, *연세대학교
20028606@ut.ac.kr, rntjs300@ut.ac.kr, cjy0103@yonsei.ac.kr, chmun@ut.ac.kr

Deep Q-Network Based Adaptive Distributed Congestion Control in Cellular V2X

Eun-Hwa Lee, Gu-Sun Joung, Joo-Young Choi*, Cheol Mun
Korea National University of Transportation, *Yonsei University

요 약

변동성이 있는 혼잡한 도로 상황에 따른 C-V2X(cellular vehicle-to-everything) 통신 시스템에서 패킷 충돌로 인한 패킷 정보 손실을 해결하기 위해 강화학습을 통해 차량마다 최적의 TTI(transmission time interval)를 조절하는 분산 혼잡제어 알고리즘을 제안한다. 제안된 기법이 기존 알고리즘보다 PDR(packet delivery ratio), IPG(inter-packet gap) 측면에서 높은 성능을 보여주는 것을 시뮬레이션 결과로 입증하고 다양한 혼잡도에서 학습한 결과와 단일 혼잡도에서 학습한 결과를 비교한다.

Ⅰ. 서 론

도로 환경의 안전과 관련하여 트래픽을 제어하기 위해 3GPP는 Release 14에서 LTE(long term evolution) C-V2X(cellular vehicle-to-everything) mode 4를 표준화하였다. mode 4는 셀룰러 인프라의 개입 없이 차량들이 SB-SPS(sensing based semi-persistent scheduling)를 통해 직접 무선자원을 할당하고 패킷을 전송한다. 하지만 혼잡한 도로 상황에서는 다수의 VUE(vehicular user equipment)가 동일한 무선자원을 예약하여 패킷이 충돌에 의해 손실되기 때문에, 이를 극복하기 위해 DCC(distributed congestion control)를 통해 전송 주기 TTI(transmission time interval)와 전송전력을 조절한다[1]. 아직 C-V2X에서 표준화된 DCC 알고리즘은 존재하지 않으나 IEEE 802.11p용 DCC표준인 SAE(society of automotive engineers)J2945/1[2]을 C-V2X에 적용하는 SAE J3161/1[3]을 개발하고 있다.

본 논문은 기존 SAE J2945/1의 전송 주기 제어 알고리즘을 강화학습 기법의 하나인 DQN(Deep Q-Networks)[5]을 이용하여 최적의 TTI를 결정하는 알고리즘을 제안한다. 또한 차량 밀도가 변하는 상황에서도 차량 밀도에 최적화된 TTI를 결정할 수 있는 다양한 차량 혼잡도에서 학습한 DQN 기반 분산 혼잡제어 알고리즘을 제안하고 이의 성능을 비교 분석하였다.

Ⅱ. 차량밀도에 적응하는 전송주기 제어 알고리즘

SAE J2945/1의 전송주기 제어 알고리즘은 시간 t에서 VUE i의 반경 100 m 내의 이웃 차량밀도 VD_t^i를 밀도계수 B와 비교하여 전송주기 TTI를 다음과 같이 결정한다[2].

$$TTI_t^i = \begin{cases} 100\ ms & VD_t^i \le B, \\ (100 \cdot VD_t^i)/B\ ms & B < VD_t^i < 6B, \\ 600\ ms & 6B \le VD_t^i \end{cases} \quad (1)$$

여기서, SAE J2945/1은 밀도계수 B를 25로 고정하므로, VD_t^i가 25 이상이 되면 혼잡도에 따라 TTI를 100ms에서 600ms까지 증가시키게 된다. 따라서, 밀도계수 B는 TTI를 조절해야 하는 혼잡구간과 최적 TTI 값을 결정하여 전송주기 제어 알고리즘의 성능을 좌우한다.

본 논문에서는 강화학습을 통해 다양한 차량 밀도에 따라 밀도계수 B를 최적화함으로써, 다양한 차량 밀도 환경에 적응하여 최적 TTI를 결정하는 분산 혼잡제어를 제안한다. 최적 B를 결정하기 위한 최적화 문제를 공식화하면 (2) 와 같다. 여기서, PDR(packet delivery ratio)을 최대화하되 CBR(channel busy ratio)을 target CBR에 근접하도록 한다. target CBR은 3GPP working document[4]와 같이 0.65로 설정하였다.

$$\mathbb{P}1: \max_{TTI_t^i} PDR_i[t,j] = \frac{\sum_{k\in\delta} C_{i,k}[t,j]}{\sum_{k\in\psi} F_{i,k}[t,j]} \quad s.t.\ \min_{TTI_t^i} \left| CBR^{target} - CBR_i[t] \right| \quad (2)$$

여기서, VUE i는 패킷을 송신하는 차량이며 ψ는 VUE i의 통신 범위에 있는 차량들의 집합이고 δ는 VUE i의 통신 범위에 있으며 패킷을 전송받은 차량들의 집합이다. VUE k는 δ와 ψ에 모두 포함되어 있어야 한다. $PDR_i[t,j]$은 서브프레임 t의 j번째 서브채널을 사용한 VUE i에서 오류없이 전송된 패킷 $C_{i,k}[t,j]$를 VUE i에서 송신한 총 패킷의 수인 $F_{i,k}[t,j]$로 나눈 것이며 VUE i의 패킷 성공률 이다. 식(2)는 $PDR_i[t,j]$을 최대화하며 CBR^{target}과 서브프레임 t에서 측정된 $CBR_i[t]$의 차이를 최소화한다.

Ⅲ. DQN 기반 혼잡도에 적응하는 전송주기 제어 알고리즘

DQN은 Q-learning을 deep neural networks로 근사화하여 Q-value를 구한다[5]. 네트워크 구조는 각 3개의 fully connected layer, ReLU activation layer와 action 개수만큼의 노드를 가진 output layer로 이루어져 있다. 그림 1과 같이 현재 time step t에서의 채널 이용률 $CBR_i[t]$을 state s_t^i로 네트워크의 입력단에 넣어서 출력으로 나온 Q-value들을 decaying ϵ-greedy를 이용해 exploration exploitation trade-off를 해주며 action a_t^i을 구한다. a_t^i는 B에 대한 이산값 {5,15,25,45,65}중에 하나이다.

에이전트의 리워드는 식 (2)를 기반으로 다음과 같이 결정된다.

$$R_j^t = PDR_i[t,j] + \left| CBR^{target} - CBR_i[t] \right| \quad (3)$$

리워드는 각 VUE가 DQN 에이전트가 되어 학습하는 과정에서 state에 따라 최적 action을 취할 수 있게 하는 지표로 사용된다. CTDE(centralized training decentralized execution)를 기반으로 하여 offline 학습을 하기 때문에 $PDR_i[t,j]$를 사용할 수 있다. 학습을 통하여 $(s_t^i, a_t^i, r_t^i, s_{t+1}^i)$ transition data가 완성되면 experience replay memory에 쌓이게 되고 이는 main network를 업데이트하는 minibatch data로 이용된다. batch size는 256이다. batch size만큼의 transition data를 통해 target network로 $\max_{a'}\widehat{Q}(s',a';\theta^-)$를 구하고 main network로 $Q(s,a;\theta)$를 구하여 아래의 식과 같이 Q-values의 MSE(mean square error) loss $L(\theta_t)$를 구한다.

$$L(\theta_t) = \mathbb{E}[(r(s,a)+\gamma \max_{a'}\widehat{Q}(s',a';\theta^-) - Q(s,a;\theta))^2] \quad (4)$$

위 식의 θ는 main network 파라미터로 (4)에 대한 gradient를 구하여 업데이트하고 θ^- target network 파라미터는 일정한 주기마다 main network와 동기화하여 안정적인 학습을 하도록 한다[5].

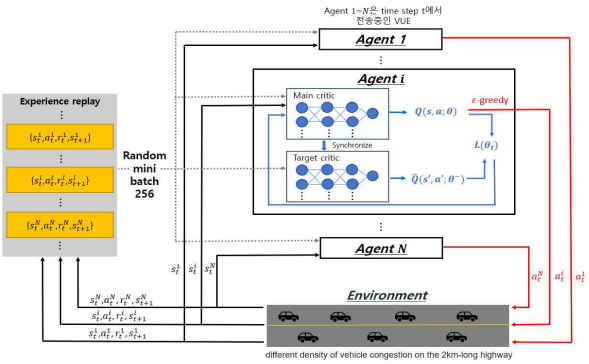


그림 1. 시뮬레이션에서 DQN 에이전트의 DCC 학습과정

Ⅳ. 시뮬레이션 결과

시뮬레이션은 LTEV2Vsim[6]을 기반으로 수행했다. LTEV2Vsim은 C-V2X 네트워크의 자원할당을 평가하기 위해 작성된 동적 시뮬레이터이다. 시뮬레이션의 환경 및 파라미터는 다음 표1과 같다.

표준 버전	C-V2X mode4	**MCS index**	6(=QPSK, 0.48)
시뮬레이션시간	20초	**패킷 크기**	300 bytes
중심 주파수	5.9 GHz	**SPS 임계값**	-94 dBm
대역폭	10 MHz	**전파모델**	WINNER Ⅱ, B1
도로 환경	2km 고속도로, 왕복 8차선		
차량 속도	60km/h, 표준편차 3km/h		
차량 밀도	2km 내 400, 600대		

표 1. 시뮬레이션 환경 및 시스템 파라미터

제안하는 알고리즘의 성능을 PDR과 평균 IPG(inter-packet gap) 관점에서 분석한다. 여기서 IPG는 동일한 송신 VUE와 수신 VUE 사이에서 연속적으로 오류 없이 수신된 2개의 패킷 사이 시간 간격이며, 평균 IPG는 송수신기 거리에 따른 시뮬레이션 환경 내 모든 VUE의 IPG에 대한 평균이다. 그림에서 fixed DQN은 고정된 하나의 차량 밀도에서 학습한 알고리즘이고 adaptive DQN은 다양한 차량 밀도에서 학습한 알고리즘이다.

그림 2는 2km 내 400대의 차량 밀도에서 학습한 fixed DQN과 400대와 600대의 차량 밀도에서 학습한 adaptive DQN 그리고 기존의 SAE J2945/1 알고리즘의 성능을 600대의 차량 밀도에서 테스트한 성능을 비교한 것이다. 제안하는 adaptive DQN은 기존 SAE J2945/1보다 PDR이 3% 개선되고 IPG는 10ms 정도의 미미한 증가에 그친다. 반면에 fixed DQN은 400대의 차량 밀도에 최적화된 과도하게 낮은 값의 TTI를 사용함으로써, 기존 알고리즘보다 PDR 성능이 최대 7% 감소하고 IPG가 과도하게 감소하는 결과를 보여준다. 그림 3은 2km내 600대의 차량 밀도에서 학습한 fixed DQN과 adaptive DQN 그리고 기존 알고리즘을 400대 차량 밀도에서 테스트한 성능을 비교한 것이다. adaptive DQN은 기존 SAE J2945/1에 비해 PDR이 개선되지만, 단일 혼잡도에서 학습을 수행한 fixed DQN은 600대의 차량 밀도에 최적화된 과도하게 큰 값의 TTI를 사용함으로써, PDR은 개선되나 IPG가 과도하게 증가하는 현상을 보여준다. 제안하는 다양한 차량 혼잡도에서 학습한 DQN 기반 분산 혼잡제어 알고리즘은 차량 밀도가 변하는 상황에서도 차량 밀도에 최적화된 TTI를 결정함으로써 기존 SAE J2945/1보다 우수한 PDR 성능을 보임을 확인할 수 있다.

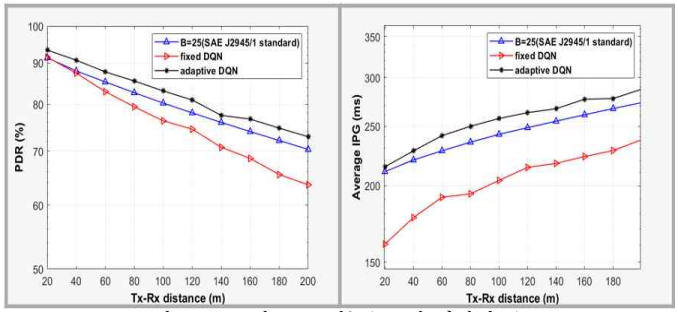


그림 2. PDR과 IPG 성능(600대 차량밀도)

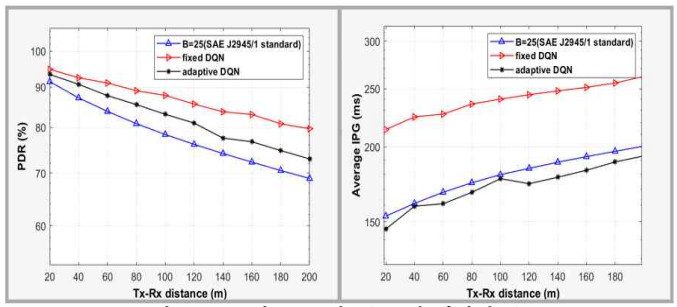


그림 3. PDR과 IPG 성능(400대 차량밀도)

ACKNOWLEDGMENT

본 연구는 국토교통부/국토교통과학기술진흥원의 지원(21AMDP-C160502-01)과 과학기술정보통신부 및 정보통신기획평가원의 Grand ICT 연구센터지원사업(IITP-2021-0-01462)의 연구결과로 수행되었음.

참 고 문 헌

[1] 정구선, 정창규, 문철, "C-V2X를 위한 지역 혼잡도 정보를 이용한 분산 혼잡 제어," 한국전자파학회논문지, vol. 32, no.2, pp.101-109, Feb. 2021.

[2] SAE International, on-board system requirements for V2V safety communications, Standard Doc. J2945/1, 2016.

[3] SAE International, on-board system requirements for LTE-V2V safety communications(under development), Standard Doc. J3161/1, Aug. 2019.

[4] Qualcomm, "R1-1611594. congestion control for V2V," 3GPP TSG RAN1-87, Tech. Rep., Nov. 2016.

[5] V. Mnih et al., "Human-level control through deep reinforcement learning," Nature, vol. 518, no. 7540, pp. 529-533, Feb. 2015.

[6] Web page of LTEV2Vsim. [Online]. Available: https://github.com/alessandrobazzi/LTEV2Vsim

k-NN 및 PSO 기반 실내 위치 측위 개선 방안

오성현, 김정곤*
한국산업기술대학교 전자공학부
osh119@kpu.ac.kr, jgkim@kpu.ac.kr*

Indoor Positioning with k-NN and PSO for an Improved Location Accuracy

Sung Hyun Oh, Jeong Gon Kim*
Dept. of Electronic Engineering, Korea Polytechnic University

요 약

4차 산업혁명이 시작되면서 LBS(Location Based Service) 산업이 함께 성장하고 있다. LBS 산업에서 가장 중요한 이슈는 다양한 위치 기반 서비스를 제공하기 위해서 사용자의 위치를 정확하게 추정해야 한다. 실외 환경의 경우 GPS(Global Positioning System) 등을 통해 비교적 높은 측위 정확도를 제공할 수 있다. 그러나 실내환경에서는 전파 복잡도로 인해 GPS 시스템을 적용하기에 제한적이다. 따라서 본 논문에서는 일반적인 실내환경에 적용되는 WiFi 통신을 사용하여 사용자의 위치를 측위하는 기술에 대해 연구한다. 먼저, 사용자 위치 측위 수행을 위해 RSSI(Received Signal Strength Indicator)와 결합하여 비교적 높은 정확도를 제공하는 핑거프린팅 기법을 적용한다. 이때 오프라인 단계에서 각 샘플 포인트를 배치한 후 RSSI 값을 측정하여 핑거프린팅 데이터베이스를 구축한다. 이후 온라인 단계에서는 핑거프린팅 데이터베이스와 실제 사용자의 위치에 대한 RSSI 값을 측정하여 지도학습의 한가지 기술인 k-NN(k-Nearest Neighbor) 알고리즘을 적용한다. 이때 실제 사용자로부터 가장 인접한 샘플 포인트를 도출하여 PSO(Particle Swarm Optimization) 알고리즘의 초기 탐색 영역을 제한한다. 이후 제한된 영역내에 입자들을 배포하여 최종적으로 사용자의 위치를 측위한다. 시뮬레이션을 통해 k-NN 및 PSO를 융합하여 사용하는 경우 기존 기법 대비하여 개선된 측위 정확도를 얻게됨을 확인할 수 있다.

Ⅰ. 서론

이동통신 기술이 발달함과 함께 LBS(Location Based Service) 제공을 위한 기술들이 고려되고 있다. 기존에는 GPS(Global Positioning System) 기술을 기반으로 실외환경에서 비교적 높은 측위 정확도를 제공할 수 있다. 그러나 실내환경의 복잡한 전파환경에 따른 신호손실 등의 문제로 GPS 기술을 적용하기에는 한계가 존재한다. 따라서, 이러한 문제를 해결하기 위해 WiFi, Bluetooth, UWB(Ultra Wide Band)등과 같이 실내환경에서 적용가능한 통신기술을 기반으로 높은 측위정확도를 제공하는 기술들이 중요한 연구 대상이 되었다[1].

실내환경에서 위치 측위 기술은 개인적인 목적이나 공공의 목적에 따라 요구되어지며, 주요한 예로는 크고 복잡한 쇼핑몰에서 고객 스스로 원하는 상점을 찾거나, 건물 내 화재 발생시 소방관을 지원하거나, 대형마트에서 할인정보와 같은 LBS 기술을 제공하기 위함이다. 현재 보편적으로 사용되고 있는 실내 위치 측위를 위한 이동통신 기술은 위에 언급되었듯이 WiFi, Bluetooth, UWB 등이 있다. 그리고 기존의 센서 측위 기술에서는 범위를 기반으로하는 기술과 범위를 사용하지 않는 방법이 존재한다. 이 중에서 일반적으로는 범위를 기반으로 하는 기술을 대부분 적용하고 있다. 그 중에서도 RSSI(Received Signal Strength Indicator)를 기반으로 하는 측위 기술은 핑거프린팅 알고리즘과 결합하여 사용할 경우 낮은 비용으로 높은 측위 정밀도를 얻을 수 있기 때문에 가장 보편적으로 사용되고 있다[2].

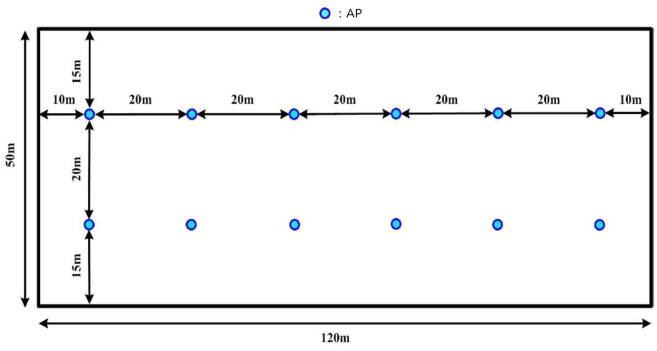
따라서 본 논문에서는 위치 측위를 위한 이동통신 기술로는 WiFi를 사용하고, 사용자 위치 측위를 위해 RSSI를 사용한다. 사용자의 위치를 측위하기 위해 초기에 핑거프린팅 맵을 구축하고, 핑거프린팅 데이터베이스와 실제 사용자의 RSSI 값 사이에 k-NN(k-Nearest Neighbor)알고리즘을 적용한다. k-NN 알고리즘은 PSO 알고리즘의 초기 탐색영역을 제한하기 위해 사용되며, 이는 측위 정확도 개선 및 빠른 측위 처리시간을 달성할 수 있도록 한다.

본 논문의 2절에서는 사용자 위치 측위를 위한 구체적인 과정에 대해 제시하고, 시뮬레이션 결과를 분석한다. 이후 3절에서는 본 논문의 결론을 맺는다.

Ⅱ. 본론

본 논문에서는 3GPP에서 제시한 실내 오피스 환경을 기반으로 한다[3]. 제시된 환경은 [그림 1]과 같으며, 크기는 $120m \times 50m$ 이다. 제시된 환경에서 총 12개의 AP(Access Point)가 배치되어 있으며, AP간 간격은 20m이다.

본 논문에서 제안하는 시스템의 전체적인 과정은 다음과 같다. 먼저, 핑거프린팅 기법을 적용하여 특정 위치에 대한 핑거프린팅 데이터베이스를 구축한다. 이후, 지도학습의 한가지 기술인 k-NN 알고리즘을 적용하여



〔그림 1〕 3GPP에서 제시한 실내 오피스 환경

* : 교신저자
이 논문은 2021년도 정부(산업통상자원부)의 재원으로 한국산업기술진흥원의 지원을 받아 수행된 연구임 (N0002429, 산업혁신인재성장지원사업)
이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2021R1F1A1063845)

사용자와 가장 인접한 샘플포인트를 도출한다. 도출된 샘플포인트를 기반으로 PSO 알고리즘의 초기 탐색영역을 제한하고, 영역 내에 입자들을 배포한다. 이후 PSO 알고리즘 과정을 수행하여 사용자의 위치를 얻는다. 적용된 기법들의 자세한 설명은 아래와 같다.

가장 처음에 핑거프린팅 기법이 다음과 같이 수행된다. 총 N개의 WiFi AP와 M개의 SP(Sample Point)가 실내환경에 배치되었다고 가정한다. 모든 SP는 각 AP로부터의 RSSI 값을 측정하고, 측정된 RSSI 값을 기반으로 핑거프린팅 데이터베이스를 구축한다. 구축된 핑거프린팅 데이터베이스는 아래와 같이 표현될 수 있다.

$$H^{RSSI}=\begin{Bmatrix} h_1^1 \cdots h_m^1 \cdots h_M^1 \\ \vdots \quad \vdots \quad \vdots \\ h_1^n \cdots h_m^n \cdots h_M^n \\ \vdots \quad \vdots \quad \vdots \\ h_1^N \cdots h_m^N \cdots h_M^N \end{Bmatrix}_{N\times M} \tag{1}$$

여기서, h_m^n는 n번째 AP와 m번째 샘플 포인트 사이의 RSSI 값을 나타낸다. 핑거프린팅 데이터베이스가 구축된 후 PSO 알고리즘의 초기영역을 제한하기 위해 지도학습의 한가지 기술인 k-NN 알고리즘을 적용한다.

k-NN은 지도학습 중 한가지 기술로 임의 입력이 특징 공간 내 k 개의 가장 가까운 데이터를 찾는 알고리즘이다. 다시 말해, 사용자의 위치로부터 가장 인접한 k개의 SP를 찾는 것이다. k-NN 알고리즘은 학습과정이 없는 방식이며, 새로운 데이터가 들어오는 시점에 기존 데이터 사이의 거리를 재서 이웃들을 선별한다. k-NN은 모델을 별도로 구축하지 않는 않기 때문에 Instance-based Learning이라고 불리기도 한다. k-NN은 거리 측정 방법으로 Euclidean Distance과 Manhattan Distance가 있다. 본 논문에서는 일반적으로 사용되는 Euclidean Distance를 기반으로 인접 SP를 도출하는 방법을 적용한다. 사용자 UE(User Equipment)로부터 가장 인접한 SP를 도출하는 방법은 아래와 같다.

먼저, 온라인 단계에서 사용자 위치에 대한 RSSI 값을 측정한다. 측정된 값은 아래와 같이 표현될 수 있다.

$$U_i^{RSSI}=[h_i^1, h_i^2, ..., h_i^n, ..., h_i^N] \tag{2}$$

여기서, h_i^n는 n번째 AP와 사용자 UE i 사이의 RSSI 값이다. 측정된 사용자의 RSSI 값을 새로운 데이터로 가정하고 핑거프린팅 데이터베이스 값과의 Euclidean Distance 계산을 통해 가장 인접한 SP를 도출한다.

$$d_{i,m} = \| U_i^{RSSI} - H_m^{RSSI} \| = \sqrt{\sum_{n=1}^{N} (h_i^n - h_m^n)^2} \tag{3}$$

여기서, $d_{i,m}$은 UE i와 SP m사이의 Euclidean Distance를 의미하며, 해당 값이 작을수록 UE i와 SP m이 인접한 것이다. 이후 k개의 가장 인접한 SP를 기반으로 PSO 알고리즘의 초기 탐색영역을 제한한다.

PSO 알고리즘은 지능형 진화계산 알고리즘이며, 탐색영역 내에 입자를 배포하여 최적의 해를 도출할 수 있다. 탐색을 진행하면서 각 입자들은 자신의 최적 위치인 $pbest$와 군집 내 최적위치인 $gbest$를 서로 공유한다. 입자들은 $pbest$와 $gbest$를 기반으로 탐색을 진행하여 최적의 해를 도출한다. 알고리즘은 최대 반복횟수에 도달하거나 목표 정확도를 달성하는 경우 종료된다. 아래는 반복에 따른 각 입자의 파라미터 변화를 나타낸다.

$$V_i(t+1) = w\,V_i(t) + a\,c\,[pbest_i(t) - X_i(t)] + a\,c\,[gbest(t) - X_i(t)] \tag{4}$$

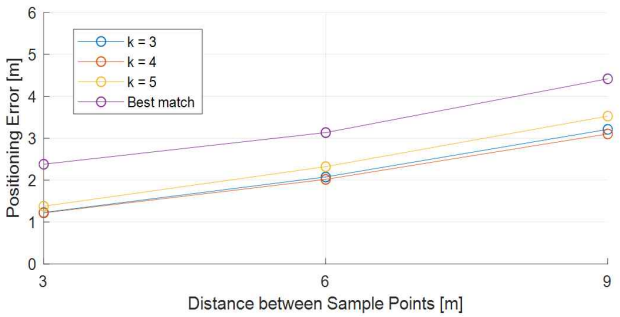
$$X_i(t+1) = X_i(t) + V_i(t+1) \tag{5}$$

$$w = w_{\max} - \frac{t\,(w_{\max} - w_{min})}{T} \tag{6}$$

여기서, $V_i(t+1)$은 각 입자의 다음 속도를 나타내고, $X_i(t+1)$는 각 입자의 다음 위치를 나타낸다. a는 가속계수, c는 수축계수, ω는 관성계수, t는 현재 반복횟수, T는 최대 반복 횟수를 나타낸다.

〔표 1〕 시뮬레이션 파라미터

파라미터	값
AP의 개수	12
샘플 포인트 간 거리	3,6,9〔m〕
k 값	3, 4, 5
입자의 개수	10
AP의 전력	20dBm
a, c, $w_{\max}$, w_{min}, T	2, 0.3, 1, 0.4, 10



〔그림 2〕 샘플포인트 간격에 따른 측위오차 성능 비교

제안 기법의 성능평가를 위해 MATLAB 2017b를 기반하여 시뮬레이션을 수행하였다. 시뮬레이션에 사용된 파라미터는 [표 1]에 정리하였다. 시뮬레이션 결과는 [그림 2]에 나타내었으며, 여기서 k의 값을 3, 4, 5로 설정하고, 샘플 포인트 간 거리가 변화할 때에 따른 위치 정확도를 비교하였다. Best match는 가장 인접도가 높은 SP의 좌표값을 UE의 좌표로 추정한 방식이다. 결과에서 확인할 수 있듯이, k의 값이 3과 4일 때 가장 높은 측위 정확도 성능을 달성함을 확인할 수 있다. 또한, SP간 간격이 좁을수록 높은 측위 정확도를 얻을 수 있는 것을 확인할 수 있는데, 이는 PSO 알고리즘의 초기 탐색영역이 작아질수록 높은 측위 정확도 성능을 달성함을 보여준다.

Ⅲ. 결론

본 논문에서는 3GPP에서 제시한 실내 오피스 환경에서 사용자의 위치 측위 정확도를 개선하는 연구를 수행했다. WiFi 통신을 기반으로 하며, 사용자의 위치를 측위하기 위해 핑거프린팅, k-NN, PSO 알고리즘을 융합하여 사용한다. 시뮬레이션 결과를 통해 오프라인 단계에서 구축되는 핑거프린팅 데이터베이스의 크기 및 k-NN 알고리즘에서 k 값에 따라 측위 정확도가 달라짐을 확인할 수 있었다. 본 논문에서는 SP간 간격이 3m이고, k 값은 4일 때 가장 높은 측위 정확도를 달성함을 확인하였다. 추후, PSO 알고리즘의 최적화를 통해 측위 정확도 개선을 위한 연구를 계획하고 있다.

참 고 문 헌

[1] Zhang, Y., Wang, H., and Wang, H. “Indoor Navigation System Design based on Particle Filter,” Proceedings of the 2016 International Conference on Intelligence Transportation, Big Data & Smart City(ICITBS), pp.105-108, December 2016.

[2] Zhao, C., and Wang, B. “A MLE-PSO Indoor Localization Algorithm Based On RSSI,” Proceedings of the 36th Chinese Control Conference (CCC), pp.6011-6015, July. 2017.

[3] “Study on channel model for frequencies from 0.5 to 100 GHz (Release14)”, 3GPP TR 38.901

무선 D2D 네트워크를 위한 자원 인지 및 위치 기반 클러스터링 연합학습

정준표, 고영배
아주대학교 AI 융합네트워크학과
{forjune, youngko}@ajou.ac.kr,

*임성화
*남서울대학교 멀티미디어학과
*sunghwa@nsu.ac.kr

Resource-aware, Location-based Clustered Federated Learning for Wireless D2D Networks

June-Pyo Jung, Young-Bae Ko
Dept. of AI Convergence Network Ajou University

*Sung-Hwa Lim
*Dept. of Multimedia Namseoul University

요 약

클라우드 서버와 클라이언트들의 직접적인 통신을 통한 기존 연합학습(Federated Learning)방식의 문제점으로 꼽히는 많은 통신 자원의 낭비, 중앙처리에 따른 부하, 실제환경에서 적용의 어려움 등의 문제를 개선하기 위해 거리 기반 클러스터링 연합학습의 효율적 방식을 제안한다. 클라우드 서버가 각각의 모든 클라이언트들과 직접적인 일대일 통신을 통해 모델 파라미터를 교류함으로써 발생하는 통신 자원의 소모를 줄이는 방식으로, D2D 통신에 기반하여 클라이언트들의 위치에 따른 제한된 거리(distortion)기반 K-mean 클러스터링을 통해 그룹과 그룹 리더를 선정한다. 선정된 그룹 리더는 그룹 내의 클라이언트들의 모델을 1 차적으로 취합하여 서버로 전송함으로써 클라우드 서버의 부하를 줄이는 동시에 클라우드 서버와 클라이언트들의 통신 자원 소모를 줄이는 연합학습 방식을 제안한다.

Ⅰ. 서 론

현재 네트워크 환경에서 기존의 머신 러닝은 디바이스에서 생성되는 데이터들을 클라우드 서버로 전송하여 데이터를 학습하는 방식을 사용한다. 하지만 클라우드 학습 처리 방식은 네트워크 대역폭, 개인정보 보호, 통신 자원 낭비 등의 문제점을 발생시키므로 모바일 환경에서는 적합 하지 않다[1].

이러한 문제점들을 해결하고자 최근 분산학습 방식의 하나인 연합학습(Federated Learning: FL)[1]이 주목받고 있다. 연합학습은 로컬 디바이스에서 생성되는 데이터를 클라우드로 전송하지 않고 로컬에서 학습함으로써 모델 파라미터 만을 클라우드 서버로 전송하여 업데이트 하는 방식이다.

하지만 기존 연합학습 방식을 모바일 환경에 효과적으로 적용하기 위해서는 학습 모델의 정확도의 수렴을 보장하는 동시에 제한된 통신과 컴퓨팅 자원에 대한 고려도 같이 이루어져야 한다. 따라서 연합학습의 다양한 프레임워크 중 하나로 클러스터링 기반 연합학습(Clustered Federated Learning: CFL)방식이 제안되었다[2]. 클라우드 서버와 모든 클라이언트들이 일대일 통신을 하는 기본의 방식을 개선하여, 클러스터링을 통해 그룹을 형성시켜 학습 모델을 취합한 후 그룹의 LN(Leader Node)만이 서버와 학습 모델을 교류하는 방식이다. 이를 통해 원거리에 위치한 클라우드 서버 와 클라이언트들 간의 통신 비용을 줄이면서 학습 모델의 수렴성을 보장한다[3].

기존의 CFL 관련 연구들은 학습 모델의 파라미터를 기준으로 클러스터링을 실시하였다. 그러나 모델 파라미터 기준의 클러스터링은 원거리에 위치한 클라이언트끼리 통신 가능성이 존재하므로 제한된 통신 자원의 모바일 환경에서 효율적이지 못하다. 따라서 본 논문에서는 D2D(Device to Device) 네트워크 환경을 가정하여 제한된 거리 내에서 클라이언트들의 위치를 기반으로 k-mean 방식의 클러스터링을 통해 그룹내의 모델들을 취합하여 LN 을 통해 원거리에 위치한 클라우드 서버와 클라이언트들의 직접적인 통신을 줄이는 자원 효율적 클러스터링 기반 연합학습 방식을 제안한다.

Ⅱ. 제안기법

그림 1.의 알고리즘은 본 논문에서 제안하는 L-CFL (Location-based Clustered FL) 이다. 클라이언트의 수 n 명, 설정한 전체 학습 라운드 T, *range* 범위에 존재하는 클라이언트들의 무작위 위치 배열 *location*, 모델 파라미터 w, 클라이언트간 최대 거리 임계 값 *threshold* 를 입력 값으로 설정하였다. *location* 값들을 입력으로 k-mean 클러스터링 함수를 통해 클라이언트들끼리 그룹화 하고 그룹의 개수에 따른 클러스터 내부 평균 거리 *distortion* 값을 구한다. 이때 설정한 제한된 거리 *threshold* 값을 넘지 않는 그룹의 최대 개수 m 을 구한다. 그림 2.는 *range* 가 2000*2000 인 환경에서 k-mean 클러스터링 방식을 활용하여 *threshold* 를 넘지 않는 그룹 m 개에 따른 그룹화된 산포도의 결과의 예를 나타낸다.

각 클러스터 내부에 위치한 클라이언트들 중 centroid 에 가장 가까운 클라이언트를 LN 으로 선정한다. 이후 기존 연합학습과 유사한 방식으로, 선정된 LN 은 클라우드 서버로부터 초기화된 학습 모델을 전송받은 후 클러스터 내부의 클라이언트들에게 학습 모델을 전송한다. 클러스터 내부의 클라이언트들은 전송받은 학습 모델을 각자의 로컬 데이터를 활용하여 훈련한다. 훈련이 끝나면 LN 으로 학습 모델을 전송하고 LN 은 학습 모델들을

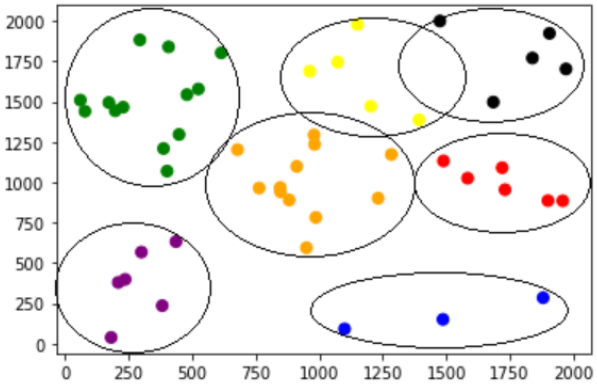
취합한 후 취합된 학습 모델을 클라우드 서버로 전송한다. 클라우드 서버는 LN 들로부터 받은 학습 모델을 취합한 후 학습 모델이 수렴할 때까지 T 라운드 동안 반복한다.

Algorithm 1: Location based Clustered Federated Learning

Input: Client set $C \leftarrow \{c_1, c_2, ..., c_n\}$,Leader node set LN$\leftarrow\{LN_1, LN_2, ..., LN_m\}$,
Initialized group model set$W_0^G \leftarrow \{w_0^{g_1}, w_0^{g_2}, ..., w_0^{g_m}\}$,
Total number of Communication T, Random location set *location*,
Threshold for distance between nodes *threshold*

kmmodels = **Kmeans**(*location*);
distortion = **Distortions**(*kmmodels*);
if ***Max***(*distortion*) $\leq$ *threshold* **then**
  m = ***Max***(*distortion*);
end
LN_m = ***findLN***($centroid_m$, *client.location*);
Server sends Initialized model W_0^G *to* LN_k
for *Communication Round t=1,2,...,T* **do**
  LN_m sends the model W_t^{gm} to all clients C belonging to the group;
  Clients belonging to each group train the model W_t^{gm} using local data;
  Clients belonging to each group update the model W_t^{gm} weights;
  LN_m aggregate the models W_t^{gm} updated in the group and sends W_t^{gm} to server;
  Server aggregates all LN_m models W_t^{gm}
end

<그림 1. Location -based Clustered Federated Learning Algorithm >



<그림 2. 클라이언트들의 위치에 따른 k-mean 클러스터링을 통한 그룹화된 산포도>

Ⅲ. 성능 평가

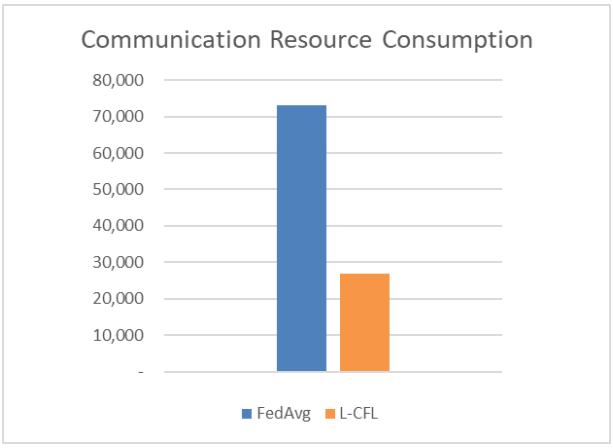
성능평가를 위해 D2D 네트워크 환경을 가정하였으며, 클라우드 서버의 위치는 2000m*2000m 의 중심점(1000,100)으로 설정하였다. 연합학습시 소모되는 통신 자원의 단위를 거리(미터)에 따라 1 unit 으로 설정하였다. 클러스터링을 통해 조직된 그룹 내부에 존재하는 클라이언트들 간의 D2D 통신 비용은 다음과 같은 이유로 서버와 클라이언트들의 직접적인 통신 비용의 1/10 수준인 0.1 unit 으로 설정하였다.

1) 클라우드 서버와 LN 간의 직접적인 통신 비용은 클러스터 내부 통신 비용의 10 배이다[4].[5].

2) 장거리 통신 거리와 잦은 backhaul 은 네트워크의 혼잡을 초래한다[4].

연합학습에 따른 통신 자원의 소모량을 비교하기 위한 벤치마크로 가장 보편적인 연합학습 방식인 FedAvg(Federated Averaging)[1]를 사용하였다.

그림 3. 은 본 논문에서 제안하는 L-CFL 기법의 통신 거리에 따른 자원 소모량을 FedAvg 와의 비교 그래프로 나타낸 것이다. FedAvg 방식의 거리에 따른 통신 비용은, 한 라운드동안 학습모델의 송신과 수신을 기준으로 36502 * 2 unit 이 소모되며, L-CFL 은 각 클러스터들의 내부 통신 비용 1106 * 2 unit, LN 과 서버와의 통신 5621 * 2 unit 을 합한 6727 * 2 unit 이 소모되었다. 이는 L-CFL 방식의 통신 비용이 FedAvg 방식의 통신 비용의 약 18%정도의 수준인 것을 확인할 수 있다.



<그림 3. FedAvg vs L-CFL 의 통신 거리에 따른 자원 소모량>

Ⅳ. 결론

본 논문은 기존 연합학습의 통신 자원 소모를 줄이기 위해 위치기반 k-mean 클러스터링을 통한 연합학습 기법을 제안하였다. 추후 본 연구를 확장하여 클라이언트들의 선택적 학습 참여를 통해 통신 자원 소모를 더 줄이면서 학습 모델의 유사도 기반의 클러스터링을 통한 연합학습 방식 연구를 통해 모바일 환경에서 자원 효율적이면서 높은 정확도의 학습 모델을 생성할 수 있는 연합학습에 대한 연구를 진행할 예정이다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 정보통신기획평가원의 대학 ICT 연구센터육성지원사업(IITP-2021-2018-0-01431)과 한국연구재단(No. 2021R1A2C1012776)의 지원을 받아 수행되었음.

참 고 문 헌

[1] B. McMahan, et. al., “Communication-Efficient Learning of Deep Networks from Decentralized Data,” in Artificial Intelligence and Statistics, pp. 1273- 1282, Apr. 2017.

[2] Sattler, F, et. al.,” Clustered federated learning: Model-agnostic distributed multitask optimization under privacy constraints”. IEEE transactions on neural networks and learning systems. 2020.

[3] Wang, Z., et. al. “Resource-Efficient Federated Learning with Hierarchical Aggregation in Edge Computing.” In IEEE INFOCOM 2021-IEEE Conference on Computer Communications pp. 1-10. May. 2021

[4] Y. C. Hu, M. Patel, D. Sabella, N. Sprecher, and V.young, “Mobile edge computing—a key technology towards 5g,” ETSI white paper, vol. 11, no. 11, pp. 1- 16, 2015.

[5] E. Cuervo, A. Balasubramanian, D. Cho, A. Wolman, S. Saroiu, R. Chandra, and P. Bahl, “Maui: making smart-phones last longer with code offload,” pp. 49- 62, 2010.

SDN(Software-Defined Network) Network상에서 인공지능을 활용한 네트워크장비의 장애예측 및 대응시스템구현에 관한 연구

함대훈*

*(주)에스비정보기술

*dhhaam@sbit.co.kr

An Implementation of SDN Controller that has the AI algorithm can both detect network device faults and automatically handle them.

Haam Dae Hoon*

*SBIT

요 약

본 논문은 SDN 네트워크상에서 발생하는 다양한 형태의 요인으로 말미암아 서비스에 지장을 주는 것을 장애라 정의 하고 이에 대한 해결책으로 인공지능의 여러 알고리즘을 적용하여 자동화를 통하여 인간의 개입 없이 판단하고 조치를 취하는 시스템 구현에 대해 논하였다. 장애에 대한 판단은 네트워크 장비들이 제공하는 명확한 정보를 바탕으로 네트워크 장비의 통계를 바탕으로 하는 것과 Traffic 정보의 Pattern을 분석하여 이를 근거로 장애를 판단하는 것이 있다. 장애로 판명될 경우, 이에 대한 처리는 경험적 바탕에 근거를 둔 여러 정책을 적용하여 서비스의 지장이 없도록 하는 시스템을 목적으로 개발하였으며 장애 종류별 대응 알고리즘은 부분적으로 개발 시험 중이다. 명확한 장애에 대한 자동화는 기대한 것 이상의 결과를 보여주었으며, 시스템 장애 시 관련된 장비를 서비스에서 제외시키고 서비스 중인 Traffic을 우회시키고 새로운 Traffic 요청에 에 대해 새로운 경로를 결정하기 위한 강화학습기반의 알고리즘을 개발하여 시험 망에서 시험 중이다. 본 논문은 이라한 시스템을 구성하는 요소, 알고리즘의 구현 단계, 그리고 SDN을 기반으로 했을 때의 장점에 대해 논하였다..

Ⅰ. 서 론

데이터 통신은 지난세기 발전을 가져와 제3의 산업혁명을 이끌었으며, 5G 네트워크로 진화하면서 인공지능, 자율주행 자동자등의 초고속, 초저지연에 대한 네트워크의 요구사항을 만족하도록 발전하고 있다. 네트워크는 데이터 전달에 대한 경로 설정을 결정하는 제어 평면 부분과 실제 데이터 전송을 담당하는 데이터 평면으로 나뉜다. 기존 라우터와 스위치들은 제어평면과 데이터 평면을 동시에 제어하고 하나의 물리적 공간을 통해서 제어를 위한 정보와 실제 데이터 전송을 하고 있으며, 제어정보는 제조사 고유의 소프트웨어를 사용하여 제조사간 호환은 불가능하였다. 이러한 기존네트워크의 한계를 극복하고자 SDN이 제안되었으며 SDN은 제어평면과 데이터 평면을 분리하고 제어평면에 대한 표준화를 진행하여 제조사간 호환을 보장한다. 이러한 SDN에서도 다양한 장애가 발생하고 있으며, 그 장애에는 라우터, 스위치 같은 장치에서 발생하는 장애 , 데이터의 이상흐름 그리고 알 수 없는 전체 네트워크 차원의 성능저하가 포함된다. 이러한 다양한 장애에 대해 이를 처리하기 위해 인공지능 기반의 자동화 알고리즘 및 네트워크의 성능 저하를 가져오는 다양한 형태의 장애에 대한 처리 연구[1]를 바탕으로 시스템의 이상 징후를 인지하고 선제적으로 대응할 수 있는 인공지능 알고리즘 및 이를 바탕으로 대응 시스템개발에 대한 연구를 하고자 한다.

Ⅱ. 본론

본 논문에서는 SDN의 Data 평면에서의 장애처리를 인공지능 알고리즘을 활용하여 자동적으로 감지하고 그에 대한 대응을 할 수 있는 시스템 개발 및 구현에 관해 논하고자한다. 시스템 구성은 그림1. 과 같다. 그림1과 같은 구성에서 SDN Controller에서 수집되는 SNMP (Simple Network Monitoring System)정보와 Traffic정보 그리고 통계정보를 활용하여 네트워크의 이상 징후를 파악한다..

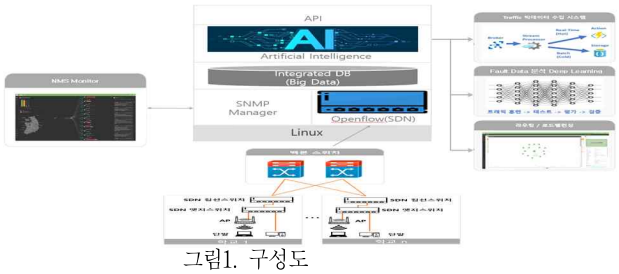


그림1. 구성도

네트워크 장비들이 전송하는 명확한 정보(Critical, Major, Minor 알람)과 장비의 통계정보를 조합하여 인공지능 알고리즘을 활용하여 Data 셑들과의 유사도 검사를 해서 유사도가 없다면 Single Failure로 간주하고 Traffic path를 다른 곳으로 Routing시킨 후 시나리오에 따라 장비를 복구 시도를 하고 그래도 복구가 안 될 때에는 Routing Path에서 제외시킨다.

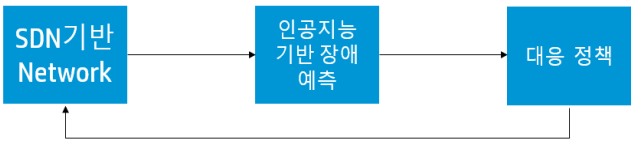


그림2. 장애 예측 및 정책 적용

Data Set	Value	Data Set	Value
CPU Load	30/50/70	Fan 상태	정상/이상
Memory	30/50/70	온도	50/60/70
Input Load	Bytes	Link Down	횟수
Output Load	Bytes	Rebooting	횟수

표 1. 장애에 대한 Data Set

한편, Traffic을 분석하여 이상 징후를 판단하는 것 중 대표적인 DDos(Distributed Denial of Service)공격에 대해서는 많은 연구가 되고 있으며[2] ,본 연구에서는 Clustering의 기법중 하나인 Random cut Forestfh 판별 한 후 STM알고리즘을 통한 DDoS를 판별한다. DDoS라 판별된 경우 경험에 의해 만들어진 정책을 통하여(Port 분리 /연결/ Port 분리)대응 한다.

장비의 장애. 또는 DDoS 공격을 받은 장비는 일시적으로 서비스에서 제외시키고 일정시간 후 서비스 복귀를 시키며, 서비스가 진행되던 또는 새로이 진행될 서비스는 다른 장비로 Routing할 수 있도록 강화학습 기반의 경로설정 알고리즘으로 최적의 경로를 선택하여 SDN스위치의 Flow Table을 제어한다.

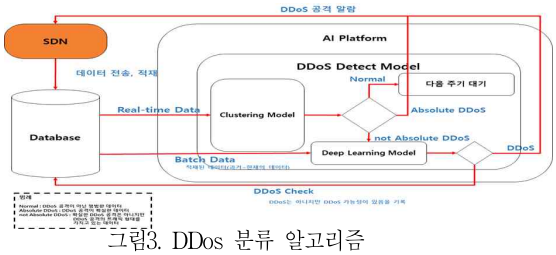


그림3. DDos 분류 알고리즘

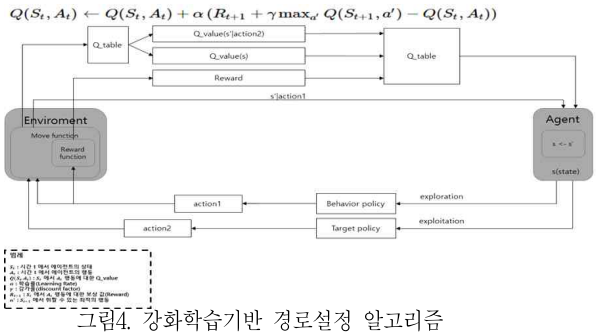


그림4. 강화학습기반 경로설정 알고리즘

Ⅲ. 결론

본 논문에서는 SDN상의 네트워크 환경 하에서 네트워크장비를 포함한 이상 징후 상태를 인공지능 알고리즘으로 종합적으로 파악 한 후 제어를 하기 위하여, 스위치 제조사가 다르더라도 SDN Openflow 규격을 준수하는 장비에 대해서 제어를 실시하여 전체 네트워크의 성능이 저하되지 않도록 하였으며 구체적인 및 시험은 아래그림과 같은 구성과 같이, SDN Controller와 빅 데이터 인공지능서버, L3 스위치3,대, 그리고 L2 스위치 3대로 망을 구성하여 시험 하였다. · · ·

군 물류 스마트화를 위한 IoT 적용 방안

하원용, 한충식
육군종합보급창
pengin0313@mnd.go.kr

Application of IoT for Smartening of Military Logistics

Won Yong Ha, Chung Sik Han
Army Consolidated Supply Depot

요 약

보급은 군 전체의 전투력에 큰 영향을 미칠 수 있으며 전시 상황에서 승리를 좌우할 수 있는 요소이기도 하다. 따라서 군은 안전 신속 정확한 보급을 위해 민간에서 빠르게 발전하고 있는 스마트 물류를 도입하고 있다. 본 논문은 군 물류 스마트화를 위한 IoT 적용 방안과 각 IoT 설비별 장단점을 서술하고 미래 군 물류의 발전 방향을 보여준다.

Ⅰ. 서 론

소비자 패러다임 변화와 정보통신 및 컴퓨터 기술의 발전으로 급변하는 물류 시장환경에 대응하고자 탄생한 개념이 “스마트 물류”이다. 특히, 스마트 물류는 인공지능(AI)과 정보통신기술(ICT)이 결합하여 물류현장의 자동화와 무인화 설비를 운영한다. 이러한 설비와 기술은 모두 데이터 분석을 기본으로 두고 있어 데이터의 양과 질이 전체적인 운영 최적화에 큰 영향을 미친다.

이러한 물류체계 변화는 군 물류에서도 적용된다. 현재 군 물류는 전체적인 물품 수가 증가하고 있으나 품목 수가 급증과 이러한 물류를 빠르고 정확하게 처리하기 위해 스마트화가 진행되고 있다. 과거 인력으로 해결했던 부분을 자동화나 무인 설비로 대체하려는 시도가 돋보이고 있으며 육군 종합보급창 예하 부대인 1보급단에서는 대규모 자동화 창고를 만들어 군 물류 스마트화에 앞장서고 있다. 또한, 2, 3보급단도 2025년까지 자동화 설비 구축이 계획되어 있으며 30%가 넘는 인력 감축 효과를 기대하고 있다.

그러나 군의 이러한 노력에도 군 물류는 보안과 예산 문제로 민간 물류 기업과 비교하여 스마트화가 느리게 진행되고 있다. 또한, 군 물류는 짧게는 수개월에서 길게는 수년 단위의 중단기 계획을 세우고 제조 업체로부터 물품을 납품받아 분배해야 해야하는 점, 이윤 추구가 아닌 안전·정확·신속한 보급을 최우선 목표로 하는 점, 타 업체나 부서와의 경쟁이 없다는 점, 언제 일어날지 모르는 전시상황을 고려하여 큰 혼란을 야기할 수 있는 모험적인 신기술이나 설비를 도입하기 어려운 점이 물류 스마트화에 지연을 일으키고 있다[3, 4].

위와 같이 군 물류는 많은 제약에도 불구하고 자동화 설비에 큰 투자를 하고 있지만, 전체적인 물류 데이터 축적에는 소극적이다. 스마트 물류는 자동화 설비 구축도 매우 중요하지만, 전체 체계 구축의 기초가 되는 데이터가 반드시 있어야 한다. 따라서 스마트 물류에서 가장 기초가 되는 데이터를 축적하는 일, 즉 IoT 설비를 구축하는 것을 자동화 설비 구축과 함께 고려해야 한다. 본 논문에서는 어떠한 IoT설비가 군 물류에 필요하며 적용했을 때 장단점이 무엇인지에 대하여 2장에서 기술하고 3창에서 차후 군 물류 스마트화의 미래에 관하여서 서술한다.

Ⅱ. IoT 적용 방안

IoT란 Internet of Things의 줄임말로 사물인터넷이라고도 한다. IoT는 센서와 동작 기계가 통합 플랫폼과의 정보교환을 통해 전체적인 작동체계 수성을 가능케 하는 기술로 스마트 물류에서 주로 데이터를 생성하는 역할을 한다. 따라서 IoT 설비 도입은 물류체계 내에서 어떠한 데이터가 필요하고 축적 및 분석할 것인지에 따라 달라진다[1].

현재 군 물류에서 필요한 데이터는 실시간 물품위치, 정보, 물량, 환경이 있다. 따라서 군 물류체계에서 필요한 IoT 설비는 빠른 정확한 입출고 현황을 파악할 수 있는 RFID(Radio Frequency Identification)와 창고 내에 물품의 위치를 실시간으로 정확히 알 수 있는 IPS(Indoor Positioning System), 물품의 배송 중 위치를 알 수 있는 GPS(Global Positioning System), 물품의 외부 환경을 측정하는 온·습도계가 있다. 본 논문은 이들의 군 물류 적용 방안을 서술하고 적용 시 장단점을 기술한다.

RFID는 원거리 태그(Tag) 리더 체계로 보통 10cm 이상의 거리에서 정보를 송수신 받을 수 있는 장치이다. RFID의 태그는 그 자체에서 RF 신호 송신이 가능하며 별도의 전원 공급 장치가 필요한 능동형과 리더기의 신호를 반사하여 전원 공급 없이 작동되는 수동형 태그가 있다. 또한, 인식 거리에 따라 구분되는 저주파와 고주파, 극초단파, 마이크로파가 있는데 주파수가 높을수록 인식 거리가 멀어진다. 이러한 특징을 가지고 있는 RFID는 두 가지 형태로 군 물류에 적용될 수 있다. 첫 번째로 물품 납품 단계에서 RFID 태그를 부착하여 입·출고 시간을 단축하고 정확한 물품 수량과 정보를 빠르게 습득하는 것이다. 대부분의 군 물품은 파레트에 실려 많은 수량이 배송되므로 1m 이상의 인식 거리를 가지는 고주파 리더기와 값이 낮고 극한 환경에서도 정보가 변형되지 않는 수동형 태그를 사용하는 것이 바람직 하다. 하지만, 각 물품의 태그장착을 제조공장에서부터 부착해야 하므로 각 제조사가 태그 제조 장비를 갖추고 있어야 하며 이에 따른 초기 구축 비용이 크다[2]. 두 번째 적용 방법은 극초단파를 사용한 능동형 태그를 파레트 하부에 부착하여 창고 내에서 움직임 제어와 위치 추적에 사용하는 것이다. 그러나 극초단파를 사용한 능동형 태그는 단가가 고주파 수

동형 태그와 비교하면 20배 이상 높으며 일정 기간이 지나면 배터리 교체 혹은 충전해야 한다. 또한, 능동형 태그는 크기가 상대적으로 크며 고장이 잦아서 이 장치를 관리 수리하는 엔지니어가 항시 상주해야 하는 단점이 있다.

IPS는 창고 내 물품 위치를 1% 미만의 오차 안의 범위에서 파악할 수 있는 장치로 창고 곳곳에 설치된 비컨(Beacon)을 이용해서 태그를 인식하여 위치를 파악한다. IPS는 빛, RF, 초음파 등을 사용하여 수신 방향과 각도를 계산하여 정확한 위치를 찾아낸다. 군 물류 창고에는 상대적으로 크고 무거운 물품이 많으며 재질 또한 금속이 많다. 따라서 빛이나 초음파 등을 이용한 IPS는 적절치 않아서 WiFi를 이용한 IPS를 구축해야 한다. IPS 설치는 물품의 위치를 파악할 수 있을 뿐만 아니라 실시간 추적도 가능하여 물품의 이동동선도 파악할 수 있다. 또한, 무기와 배터리등 위험성이 높은 물품에 붙착하여 보안을 높일 수 있다. IPS 태그는 직접 신호를 보내야 하는 발신기가 포함되어 있어야 해서 별도의 전원 공급장치가 필요한 단점이 있어 각 물품에 부착하기보다는 파레트 하부에 부착하는 것이 바람직하다.

GPS는 중궤도를 도는 20개 이상의 위성이 발신하는 마이크로파를 받아 수신기의 위치를 알아내는 체계이다. GPS의 주요 역할은 창고를 벗어나는 물품의 위치를 파악할 수 있고 배송되는 과정을 실시간으로 추적할 수 있다. GPS도입은 따로 배송에 이용되는 차량이나 기차, 항공기를 추적할 필요없이 통제실에서 물품의 위치를 파악 할 수 있으며 출도착 시간과 배송 동선을 파악하여 차후 배송노선 최적화와 도착시간 예측의 데이터로 활용되며 물품이 도난돼는 것을 방지하는 역할도 한다[5]. GPS는 IP와 같이 별도의 전원 공급장치가 필요하므로 파레트 하부에 부착하여 활용하는 것이 바람직 하며 해당 파레트는 반드시 회수되어야 한다.

온·습도계는 온도와 습도를 실시간으로 알 수 있는 장치로 현대 기술로 1cm이하의 크기로 제조할 수 있다. 온·습도계는 물품의 변질이 발생했을때 원인 분석과 이를 방지하기 위한 최적의 환경을 알아내는 데이터를 생성에 활용한다. 특히, 식자재나 배터리등 온도와 습도에 민감한 물품의 물류처리에 활용된다. 온·습도계는 실시간으로 치수를 측정하여 데이터를 저장 및 수신해야 하므로 물품에 각각 부착하기에는 비용적인 측면에서 바람직하지 않다. 따라서 파레트에 설치해야 하는데 파레트 바닥에 설치되면 온·습도에 왜곡이 일어날 가능성이 높다. 이런 왜곡을 방지하기 위하여 파레트 측면에 각각 설치하여 측정값에 신뢰를 높여야 한다.

Ⅲ. IoT와 군 물류의 미래

군 물류에 적용할 수 있는 IoT는 위에 설명한 RFID와 IPS, GPS, 온·습도계 이외에도 물류 취급 장비에 에너지 소비와 무게 측정 센서 등 무궁무진하다. 하지만 본 논문에서 위 네 가지만 언급한 이유는 이 IoT장비들이 만들어내는 데이터가 물류 최적화에 가장 중요한 창고 물품 위치 최적화와 스케줄링, 미래 주문량 예측 등을 가능케 하기 때문이다.

미래 대비 보급지원 역량 강화를 위해 한단계 더 발전시키기 위해서는 IoT 이외에 클라우드 및 모바일, BDA(Big Data Application), CPS(Cyber Physical System)등 소위 물류 4.0의 기술이 모두 도입되어야 한다. 이를 위해 첫번째 단계인 데이터 기반을 구축하는 IoT에 대하여 먼저 알아보았다[6]. 여기서 데이터는 단지 정확한 수불실적이나 물품의 정보만이 아닌 초단위의 물품 및 물류취급장비의 위치정보, 외부 환경정보, 실시간 재고량 등이 포함된다. 이런한 데이터는 <그림1>과 같이 신 물류 기술들의 기초가 되며 이런 기술들을 통한 상호 유기적인 데이터의 순환이 발생하며 새로운 데이터를 생산하여 스스로 발전하게 만드는 진정한 스마트 물류가 된다.

참 고 문 헌

[1] Acero, R., Torralba, M., Perez-Moya, R., Pozo, J. "Innovation in Logistics: Advanced Pooling and Robotization," Applied Science, Vol.10, No.106

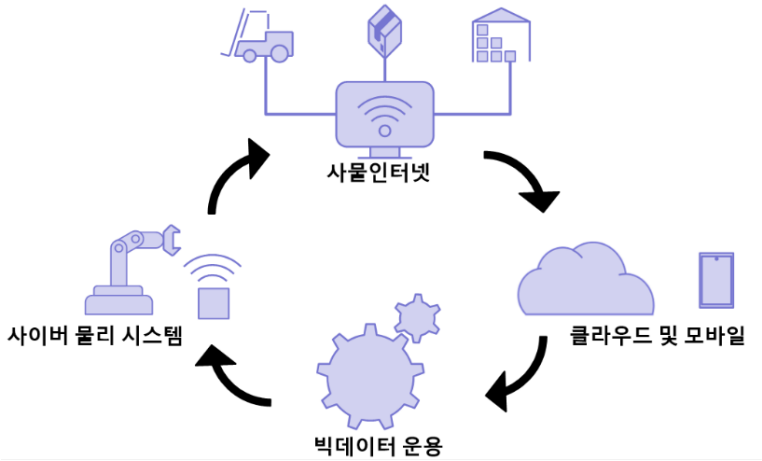
[2] Fraga-Lamas, P., Fernadez-Carames, T.M., Suarez-Albela, M., Castedo, L., Gonzalez-Lopez, M., "A Review on Internet of Things for Defense and Public Safety," Sensors, Vol.16, No.1644

[3] Hopkins, J., Hawking, P. "Big Data Analytics and IoT in Logistics: A Case Study," The International Journal of Logistics Management, Vol.29, No.2, 575-591, 2018

[4] Pawelcyk, M.M. "Contemporary Challenges in Military Logistics Support," Security and Defence Quarterly, Vol.3, 3-17, 2018

[5] Winkelhause, S., Grosse, E.H. "Logistics 4.0: a systematic review rowards a new logistics system," Inernational Journal of Production Research, Vol.58, No.2, 18-43, 2020

[6] 계중읍, 박판준, 김원태, 임채덕 "사이버물리시스템(CPS)과 사물인터넷(IoT)기술의 군사적 활용방안 및 추진전략," 전자통신동향분석, 30(4), 92-101, 2015



<그림1> 물류 4.0 기술의 데이터 순환 구조

제2회 한국 인공지능 학술대회

The 2nd Korea Artificial Intelligence Conference

세션 C

빅데이터 분석을 통한 개인정보자기결정권 영향 요인 모델 연구

이연수, 이욱한*
숙명여자대학교, *숙명여자대학교
pluto2002p99@sookmyung.ac.kr, *uhlee@sookmyung.ac.kr

A Study on Factors Affecting the Right to Control Personal Information through Big Data Analysis

Lee Yeon Su, *Lee Uk Han
Sookmyung Women's University, *Sookmyung Women's University

요 약

본 논문은 빅데이터의 시대 및 코로나 사태를 경험함으로써 개인정보의 활용 및 개인정보 보호의 필요성이 커짐에 따라 개인정보자기결정권 영향력이 커짐에 주목하였다. 이에 빅데이터 분석을 통하여 개인정보자기결정권에 영향을 미치는 요인 모델을 제시하는 데에 목적이 있다. 따라서 개인정보자기결정권에 대해 살펴본 후, 빅데이터 분석을 활용한 개인정보자기결정권 영향 요인 모델을 제안하고자 한다.

Ⅰ. 서 론

4차 산업혁명의 시대는 데이터의 시대이다. 사회에서 개인정보는 '자원'으로서 경제, 사회 등 다양한 분야에서 기능하게 된 한편, 한편 개인의 신분을 나타냄으로서 '인격'으로의 성격을 나타내기도 한다. 이렇듯 개인정보는 데이터 시대에 융합된 정체성을 갖고 사회 전반에서 활용된다.

사회가 고도로 발전하고 진화할수록 개인정보는 개인 데이터로서의 이용 가치가 커지게 될 것인 한편, 국가적 재난 상황에서 개인정보의 활용 가능성이 커짐에 따라 개인정보를 형성하는 개인의 권리 보호도 중요하게 될 것이다. 개인정보의 활용 및 보호가 중요시된 것으로 현재의 COVID-19를 들 수 있다. 우리나라는 코로나19 사태의 대응과정에서 신속한 정보공개를 통한 방역시스템의 구축은 세계적으로 주목받았다. 그러나 이와 같은 주목의 시선에는 부정적인 시각도 있었다. 정보공개 시스템 실행의 이면에는 국가인권위원회의 권고에도 불구하고 확진자의 성별, 나이, 이동경로 등 개인을 특정하거나 유추할 수 있는 정보가 공개되었고, 방역에 필요한 수준을 넘어선 정보내용 공개도 있었다.[1] 이처럼 국가에 의한 개인정보의 처리는 시민들의 더 많은 자유의 보장과 안전에 기여할 수 있지만 인권에 대한 위협이 될 수 있다. 이러한 점에서 민주주의의 이념과 법치국가의 원리에 입각하여 국가의 감시행위를 효율적으로 통제할 수 있는 법적 장치가 필요하다.[2] 개인정보자기결정권은 스스로에 대한 정보 노출을 억제하는 데만 중심이 있는 것이 아닌 스스로에 대한 정보의 흐름을 적극적으로 형성하기도 한다. 따라서 빅데이터 분석을 통한 개인정보자기결정권 영향 요인 모델을 제안함으로써 개인정보자기결정권의 실효적 보장에 기여하고자 한다.

Ⅱ. 본론

1. 개인정보자기결정권의 개념

개인정보자기결정권은 정보주체가 자신에 관한 정보에 대한 조사·수집·보관·처리·이용 등의 행위에 대해 동의 또는 거부할 수 있는 권리를 핵심적인 내용으로 하며, 경우에 따라 자신에 관한 정보에 관해 통제력을 행사하기 위하여 자신의 개인정보에 접근하여 열람하거나 그 정보의 정정·삭제·차단·처리정지 등을 요구하는 권리로 현출될 수 있다고 해석된다.[3] 따라서 이와 같은 개인정보자기결정권의 핵심은 정보주체인 개인이 자신의 의사에 따라 개인정보에 대해 조사·수집·보관·처리·이용하는 것에 대해 스스로가 동의 또는 거부할 수 있는 권리를 보장하는 것이다. 즉, 개인정보에 대한 정보주체의 통제권의 인정이 개인정보자기결정권의 핵심이라고 볼 수 있다.

2. 개인정보자기결정권 영향 요인 모델 제안

2.1 연구 설계

본 논문에서는 빅데이터 분석을 통해 개인정보자기결정권 영향 요인 모델을 제안하고자 한다. 이에 개인정보자기결정권에 관한 토픽에 따른 키워드를 분류 및 경향을 파악하기 위해 토픽모델링 분석과 개인정보자기결정권 상에 내재된 개인정보자기결정권에 대한 영향 요인을 파악하기 위해 회귀분석을 제안하고자 한다.

2.2 연구 방법

연구방법은 다음과 같다. 우선 연구의 데이터는 네이버(Naver), 다음(Daum), 구글(Google) 등의 대표적인 포털사이트에서 개인정보자기결정권을 키워드로 한 단어들을 수집한다. 데이터 수집 도구로써 텍스톰(Textom) 서비스를 이용하고자 하며, 분석 채널은 뉴스에 한정한다. 자료의 분석 기간은 3년으로 2018년 1월 1일부터 2020년 12월 31일까지 연구 대상의 기간으로 한정한다. 분석 데이터 정보는 다음(표 1.)과 같다.

표1. Research Data

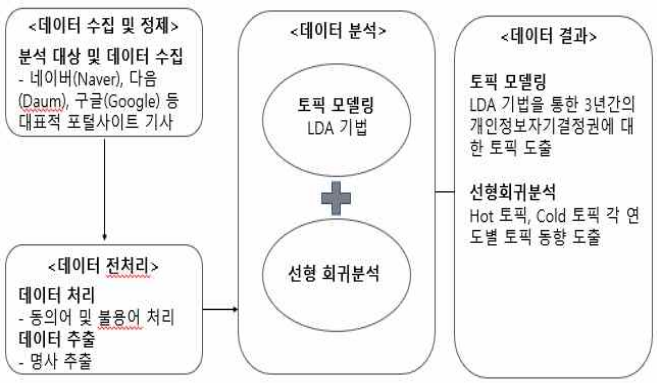
Target	Naver(news), Daum(news), Google(news)
Period	2018.1.1~2021.12.31
Tool	Textom, R Packages(3.6.1)
Key word	개인정보자기결정권

본 연구는 개인정보자기결정권 영향 요인 분석을 위해 토픽모델링 과 선형회귀분석을 제안한다. 먼저 토픽모델링은 문서가 가지는 잠재적 의미와 주제들을 찾아 추론해 내는 방법으로, 이를 통해 대량의 문서들에서 출현할 확률이 높은 키워드를 도출하여 해당 토픽의 의미와 주제를 찾는 확률 모형이다.[4] 단어들의 집합들로 이루어진 문서의 경우 토픽의 집합이 확률적으로 내포되어 있으며 단어들 간의 분포에 대한 분석을 시행함으로써 구조를 파악해 문서 내의 주제를 파악한다. 이를 통해서 문서에서 중요하게 다루어지는 주제를 파악할 수 있는 것이다. 문서 내 토픽을 분석하는 기법으로 최근 주제별 단어의 확률을 기반으로 하는 LDA(Latent Dirichlet Allocation) 기법이 있으며 이는 기존 토픽모델링의 단점을 보완하여 이루어지는 분석 기법이라는 점에서 의미 있다. LDA 알고리즘 모델 기법은 관찰 되어지는 변수(observed variable)와 문서 내 감춰져 있는 주제 즉, 보이지 않는 변수(hidden variable)를 찾고 추론해준다.[4][5] 따라서 LDA 분석을 통해 전체 문서들의 집합을 기반으로 주제 도출, 문서 내에서 주제의 비율 그리고 주제에 포함된 단어의 확률을 알아낸다. 이는 R Packages(3.6.1)를 이용하는 것을 제안한다.

다음으로 토픽모델링 분석 결과를 기반한 개인정보자기결정권에 영향을 미치는 요인을 분석하기 위해 선형회귀분석을 이용한다. 토픽에 대한 전체 분석 기간 동안의 비중에 대한 추세를 분석하여 비중의 증가 추세를 보이는 개인정보자기결정권 Hot 토픽과 감소 추세를 보이는 개인정보자기결정권 Cold 토픽에 대한 분석을 실시한다. 각 토픽의 추세를 증가와 감소로 나눠서 판단은 기준은 다음과 같다. 분석 데이터는 3년 간의 연도별 토픽의 비중 평균값을 활용하며, 연도는 독립변수, 개인정보자기결정권에 관한 토픽에 대한 비중값을 종속변수로 활용한다. 이를 통해 개인정보자기결정권 영향 요인에 대한 회귀계수 값을 산출하여 추이를 분석한다. 회귀 계수 값이 양수의 형태를 가질 경우 Hot, 음수의 형태를 가지면 Cold로 분류하여 토픽의 추이를 도출한다.

따라서 2018년부터 2020년인 3년간의 포털사이트 기사 데이터를 바탕으로 토픽 모델링 중 LDA 기법을 통해 개인정보자기결정권에 관한 주요 토픽을 추출 후 분류한다. 이후 각각의 토픽에 대한 3년 간의 추이 분석을 통하여 개인정보자기결정권의 기간별 증가(Hot) 주제와 빈도가 줄어드는 감소(Cold) 주제를 밝혀 선형회귀분석을 제안한다. 분석 설계 모델은 다음 (그림 1.)과 같다.

그림 1. 개인정보자기결정권 영향 요인 분석 설계 모델



Ⅲ. 결론

본 논문에서는 개인정보자기결정권에 대한 영향 요인 모델을 LDA 기법을 통한 토픽모델링 및 선형 회귀분석을 결합하여 제안하였다. 이와 같은 제안 모델을 통하여 개인정보자기결정권에 대한 영향 요인을 도출해냄으로써 개인정보자기결정권에 대한 영향 요인을 분석하고, 나아가 개인정보자기결정권에 대한 이해 및 향후 개인정보자기결정권의 개념 방향에 대해서 제시해 볼 수 있다는 점에서 의의가 있을 것이다.

다만 데이터에 있어 3년 간의 뉴스 데이터만으로 연구를 설계 했다는 점에서 한계가 있을 것이다. 이에 개인정보자기결정권에 대한 영향 요인을 파악하기 위한 키워드를 다양화하고 그 기간과 수집 범위를 확대한다면 보다 의미 있는 연구가 될 것으로 기대한다.

참 고 문 헌

[1] 조소영, “코로나19 시대의 개인정보자기결정권의 보호-감염병 예방법상의 개인위치정보 공개규정을 중심으로-”, 헌법학연구 제27권 제2호, pp. 179-205, 2021

[2] 권건보, “정보통신기술의 발달과 한국의 프라이버시권 관련 주요 쟁점”, 헌법재판연구, Vol.8, No.1, pp 121-153, 2021.

[3] 권건보, “개인정보의 헌법적 기초와 과제”, 저스티스 통권 제144호, pp 17, 2014.

[4] Bei, D. M. “Probabilistic topic models Communications of the ACM”, 55(4), pp. 77-84, 2012.

[5] 이원성, 손소영, “공간빅데이터 연구 동향 파악을 위한 토픽모형 분석”, 대한산업공학학회지, 제41권, 제1회, pp. 64-73, 2015.

[6] Bei, D. M, Ng, A Y., & Jordan, M. I., “Latent dirichlet allocation”, Journal of Machin Learning Research, 3, 993-1022, 2003.

전이학습 기반 사시 진단을 위한 안구운동영상 수집 및 소스 데이터 구성에 따른 성능 비교

주재한, 김동환, 최건호, 김석찬*

부산대학교, 부산대학교, 부산대학교, *부산대학교

jhjoo2018@pusan.ac.kr, dongh@pusan.ac.kr, 201624194@pusan.ac.kr, *sckim@pusan.ac.kr

A Study on Performance Comparison According to eye movement image Collection and Source Data Composition for Strabismus Diagnosis Based on Transfer Learning

Jaehan Joo, Donghwan Kim, Geonho Choi, Suk Chan Kim*

Pusan National Univ., Pusan National Univ., Pusan National Univ., *Pusan National Univ.

요 약

본 논문은 딥러닝 기반의 사시 판단 시스템을 설계하기에 앞서 비정형 데이터의 일종인 안구운동영상 수집 시 실제 의료진들의 진단 정확도, 선택률에 따른 합성곱 신경망의 성능을 비교하였다. 총 1700장의 사시환자 안구운동영상 데이터를 모델 훈련 및 테스트에 활용하였고, 데이터 클래스는 정상, 내사시, 외사시로 구분된다. 선택률에 따라 4개의 데이터 셋을 구성하였으며, 일반적으로 많이 이미지 구별 작업에 많이 사용되는 6개의 합성곱 신경망을 사용하여 각 데이터 셋에 대한 성능을 비교한다.

Ⅰ. 서 론

최근 딥러닝 기반의 의료 영상 분석 및 진단 기술의 연구 • 개발이 활발하게 진행 되고 있다[1]. 의료 분야에서의 딥러닝 활용 분야는 크게 병의 유무와 종류를 판단하는 구별 작업(classification task), 컴퓨터단층촬영(Computed Tomography, CT), 자기공명영상(Magnetic Resonance Imaging, MRI), 초음파 영상(Ultrasound Image)과 같은 정형화된 영상 내부의 환부나 신체 영역을 분리하는 객체 탐지(Object Detection), 영상 분할(Segmentation) 작업이 있다.

최근에는 수집 가능한 데이터 수 부족, 개인정보보호법 및 의학연구윤리심의위원회(Institutional Review Board, IRB) 통과, 임상적 확률에 의해서 방병률이 낮은 질병에 대한 데이터 수집 한계와 같은 문제점을 극복하기 위해서 생산적적대모델(Generative Adversarial Networks, GANs) 기반의 데이터 생성 및 융합 분야도 활발하게 연구 되고 있다[2 ~ 4].

현재 딥러닝 모델 학습을 위해 사용되고 있는 데이터 셋은 DICOM(Digital Imaging and Communications in medicine)과 같은 정형화된 형태로 제공되는 자기공명영상, 컴퓨터단층촬영, 초음파영상이 있고, 환자의 환부나 움직임을 직접 촬영한 영상이 있다. 정형화된 데이터 셋의 경우 실제 임상환경에서 사용하기 위한 목적으로 수집 되기 때문에 환자의 질환명, 정도, 처방, 진단 기록 뿐만 아니라 성별, 나이와 같은 메타데이터를 기반으로 구성되어 있다. 하지만 개인정보가 다량으로 포함되어 있기 때문에 의료 기관과 환자의 동의 없이 연구, 개발을 위한 데이터 수집이 어렵고, 해당 데이터를 수집하는 과정에서 고가의 의료 장비가 필요하기 때문에 일반인들이 데이터 수집에 참여하기 어렵다. 비정형 데이터, 즉 직접 촬영한 데이터는 개인정보보호 문제, 데이터 수집의 물리적 한계에 대해서 정형화된 데이터 대비 연구, 개발에 장점이 있다. 하지만 각 개인의 촬영 환경이 너무 다양하고, 질환에 대한 명확한 판단에 어려움이 있다. 따라서 비정형 데이터 기반의 딥러닝 모델 연구에는 높은 수준의 데이터 전처리 및 선별 과정이 필요하다.

본 연구에서는 딥러닝 기반의 사시 판단 시스템을 설계하기에 앞서 비정형 데이터의 일종인 안구운동영상 수집 시 실제 의료진들의 진단 정확도, 선택률에 따른 합성곱신경망(Convolutional Neural Networks, CNN)의 성능을 비교하고자한다. 안구운동영상은 사시 환자를 정면에서 촬영하며, 동공의 시선에 따라 총 9장의 영상으로 구성된다. 본 연구에서는 9개의 영상 중에서 정면을 응시하고 있는 영상을 사용한다.

Ⅱ. 본론

(1) 입력 데이터

본 논문에서는 합성곱신경망의 입력 데이터로 사시 환자의 안구운동영상을 사용한다. 안구운동영상은 사시 환자를 정면에서 촬영하며, 동공의 시선에 따라 총 9장의 영상으로 구성된다. 본 연구에서는 9개의 영상 중에서 정면을 응시하고 있는 영상을 사용한다.

(2) 데이터 수집

본 논문에서 사용하는 안구운동영상은 양산부산대학교병원 안과학교실로부터 수집하였다. 표 1에 보이는것과 같이 총 1,700명의 환자 데이터를 수집하였고, 데이터 클래스는 정상, 내사시(Esotropia), 외사시(Exotropia)로 구분된다.

(3) 데이터 선별

본 논문에서는 데이터 선별에 따른 합성곱 신경망의 성능 비교를 위해 양산부산대학교병원의 안과 전문의료진 10명이 데이터 라벨링과 선별에

참여하였다. 데이터 셋 구분을 위해 한 장의 안구운동영상에 대한 의료진들의 사시 종류 선택률에 따라 총 4개(모든 데이터 사용, 60%, 70%, 80%) 데이터 셋을 구성하였고, 구성된 데이터 셋 내부의 안구운동영상 수는 아래 표 2와 같다.

(4) 딥러닝 모델 훈련 및 테스트

본 논문에서는 데이터 선택률에 따른 일반적인 합성곱 신경망의 성능 비교를 위해 Xception, ResNet50V2, ResNet101V2, ResNet152V2, VGG16, VGG19모델을 사용하였다. 또한 구성한 데이터의 수가 부족하기 때문에 전이학습 기법 중 하나인 미세조정(fine tuning) 기법을 활용한다. 또한 4개의 데이터 셋 모두 훈련 데이터 셋, 검증 데이터 셋, 테스트 데이터 셋의 비율을 8:2:2로 구성하여 훈련 및 테스트를 진행하였다. 실험 훈련 파라미터는 표 3과 같다.

표 1. 데이터 수

Table. The number of Dataset

class	Normal	Esotroia	Exotropia
# Data	963	275	462

표 2. 각 데이터 셋의 안구운동영상 수

Table. The number of Data in each Dataset

Dataset	Normal	Esotroia	Exotropia
All	963	275	462
60%	925	249	435
70%	852	219	391
80%	753	192	337

표 3. 실험 훈련 파라미터

Table. Training Parameter

Dataset	Epoch	Learning Rate	batch size
Xception	100	0.0001	200
ResNet50V2	100	0.0001	200
ResNet101V2	100	0.0001	200
ResNet152V2	100	0.0001	200
VGG16	100	0.0001	200
VGG19	100	0.0001	200

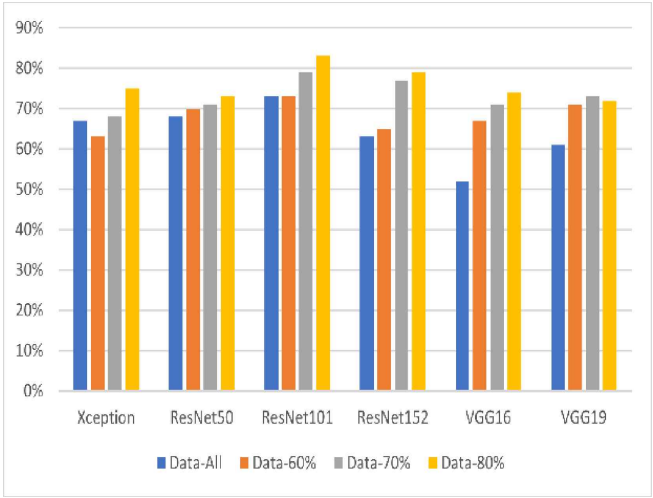


그림. 1. 데이터 셋, 딥러닝 모델 별 구별 테스트 정확도

Ⅲ. 결론

본 논문에서는 데이터 셋 구성 시 전문 의료진의 클래스 선별률에 따른 대표적인 합성곱 신경망 6개의 성능을 비교하였다. 6개의 합성곱 신경망을 4개의 데이터 셋으로 각각 훈련 및 테스트를 진행하였고, 그림 1은 4개의 데이터 셋에 대한 각 합성곱 신경망의 테스트 정확도를 나타낸다. 그림과 표에서 보이는 바와 같이 선택의 임계값을 크게 하였을 경우 훈련 데이터 셋의 수는 부족하지만, 훈련 및 테스트 정확도는 향상하는 것을 확인하였다. 또한 6개의 합성곱 신경망 중에서 ResNet101V2 모델을 의료진이 80% 이상 선택한 데이터 셋으로 학습하였을 때가 가장 성능이 좋은 것을 확인하였다. 하지만 현재 수집한 데이터 수가 적고, 클래스 간의 데이터 수 불균형으로 인하여 학습 과정중에 과적합 문제가 발생한 것을 확인하였고, 추후 연구에서는 생성적 적대 신경망과 같은 데이터 증강 기법을 활용하여 이러한 문제점을 해결하고자 한다.

ACKNOWLEDGMENT

본 연구는 2020년도 산업통상자원부의 재원으로 한국에너지기술평가원(KETEP)의 지원을 받아 수행한 연구 과제(No. 20204030200030)입니다.

이 논문은 국토교통부의 스마트시티 혁신인재육성사업으로 지원되었습니다.

참 고 문 헌

[1] Shen, D., Wu, G., and H. I."Deep Learning in medical image analysis,"Annual review of biomedical engineering 19, (2017), pp. 221-248.

[2] KER, Justin, et al. “Deep learning applications in medical image analysis,” Ieee Access, June 2017, pp. 9375-9389.

[3] Lundervol, Alexander S., and Arvid L.. "An overview of deep learning in medical imaging focusing on MRI." Zeitschrift für Medizinische Physik 29.2 (2019), pp 102-127.

[4] Cheng, Jie-Zhi, et al. "Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans," Scientific reports 6.1 (2016), pp. 1-13.

[5] Gulrajani, Ishaan, et al. "Improved training of wasserstein gans." arXiv preprint arXiv:1704.00028 (2017).

유방 촬영술을 이용한 딥러닝 기반 병리학적 완전관해율 (pCR) 예측 연구

신호경[1], 김원화[1,2], 김혜정[1,2], 김찬호[1], 김재일[1]

[1]경북대학교, [2]경북대학교병원

parkland106e@naver.com, greenoaktree9@gmail.com, ant637@naver.com,
knight970815@gmail.com, threeyears@gmail.com

A Deep Learning Study on Prediction of Pathological Complete Response (pCR) using Mammography

Ho Kyung Shin[1], Won Hwa Kim[1,2], Hye Jung Kim[1,2], Chanho Kim[1] and Jaeil Kim[1]

Kyungpook National Univ., Kyungpook National University Chilgok Hospitalniv.

요 약

병리학적 완전 관해율(Pathological Complete Response)은 유방암 선행화학요법(NeoAdjuvant Chemotherapy) 후 종양 조직의 유무를 판단해 환자의 생존율, 재발률 등 유방암 치료의 지표로 이용된다. 유방암 진단 과정에서의 pCR 예측은 향후 환자의 치료 방향 결정에 큰 영향을 미친다. 본 논문에서는 유방암 조기 진단에 이용되는 유방촬영술 영상을 통해 pCR 여부를 예측하는 멀티사이즈 패치 기반 모델을 제안한다. 종양과 주변 영역의 특징을 모델에 다양한 범위로 학습시키기 위해 전체 MLO, CC view 영상을 여러 번 리사이즈 한 후 동일한 크기의 패치를 뽑아 set로 묶어 모델의 입력으로 사용하였다. 제안 모델은 기존의 유방촬영술을 이용한 딥러닝 모델(AUC:0.5163)에 비해 우수한 예측 성능(AUC:0.8033)을 보였다.

Ⅰ. 서 론

유방암은 여성들을 대상으로 발병하는 암 중 발병률과 사망률이 높은 암종이다. 선행화학요법(NeoAdjuvant Chemotherapy)은 전이와 종양의 크기를 줄여 유방 보존술 성공을 확률을 높여주는 치료법이다. 병리학적 완전 관해율(Pathological Complete Response)은 NAC 후 결과를 평가하는 지표 중 하나로써 남아 있는 종양 조직 유무를 통해 환자의 생존률과 재발률 예측이 가능하다. NAC 이전에 환자의 pCR 달성 여부를 예측하는 것은 조기에 환자에게 더 효율적인 치료 방향 결정을 가능하게 한다.

유방암 진단에는 유방촬영술을 비롯해 MRI, 초음파 등 다양한 영상 기법이 사용되고 있다. 영상 기법들 중 MRI는 병리 영상의 종양 영역과 가장 유사하게 종양 영역을 나타내기 때문에[1, 2] MRI를 이용한 pCR 예측 연구가 많이 수행되었다[3, 4]. MRI가 pCR 예측 연구에서 가지는 장점에도 불구하고, 높은 비용, 긴 촬영 시간, 조영제의 필요성 등 제한점이 있다. 반면 유방촬영술은 유방 조직의 특징과 미세 석회화 등 주요 특징 확인이 가능하고, 유방암 조기 진단에 널리 사용된다는 장점이 있어[5], 유방촬영술을 통해 pCR 예측이 가능하다면 조기에 환자에 효과적인 치료법을 결정할 수 있다.

본 논문에서는 멀티사이즈 유방촬영술 패치를 이용한 pCR 예측 모델을 제안한다. 제안 모델의 성능 확인을 유방촬영술 영상을 이용한 유방암 진단에서 높은 성능을 보인 모델들과 비교하였다[6, 7, 8].

Ⅱ. 본론

본 연구를 위해 ** 병원에서 NAC 수행 전 유방촬영술 영상으로 총 288명의 영상을 수집하였다. 전체 데이터셋은 pCR 달성 144명과 미달성 144명 환자의 영상으로 구성되어 있고, pCR과 non-pCR 비율을 1 : 1로 유지한 상태로 학습용 데이터셋 228명, 검증용 데이터셋 60명으로 나누었다.

그림1은 제안한 모델의 전체 구조로 본 논문에서 제안한 멀티사이즈 유방촬영술 패치 모델(MSPatch-Net)은 patch extraction, feature

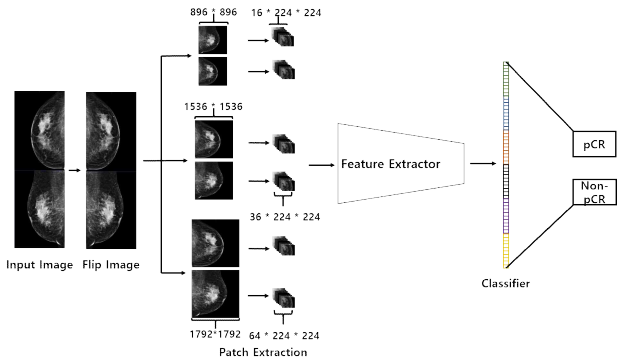


그림 1 멀티사이즈 유방촬영술 패치 모델

extractor, classifier로 구성되어 있다. 우선 입력 영상에서 동일한 위치의 패치를 뽑기 위해 patch extraction 전에 영상의 방향을 일치시켜준다. patch extraction에서 방향을 일치 시킨 영상을 896×896, 1536×1536, 1792×1792로 리사이즈한 후 동일한 크기의 패치(224×224)를 추출한다. 추출한 패치들은 채널 방향으로 합쳐져 총 6개의 patch set이 만들어지고 같은 크기로 변환된 MLO, CC view에서 추출한 patch set은 동일한 크기의 채널 수를 가진다. 그림 2는 patch extraction 과정이다. patch extraction을 통해 얻은 patch set들은 feature extractor의 입력으로 들어가 6개의 특징 벡터를 생성한다. feature extractor는 ResNet[9]을 백본 네트워크로 사용하는 6개의 서브 네트워크로 구성되어있고, 구조는 그림 3과 같다. 생성된 특징 벡터의 크기는 1×512이고, 각 특징 벡터들을 결합하여 classifier의 입력으로 사용한다.

classifier는 하나의 fully connected layer로 구성되어 1×3072 크기의 통합 벡터를 입력으로 환자의 pCR 달성 여부를 예측한다.

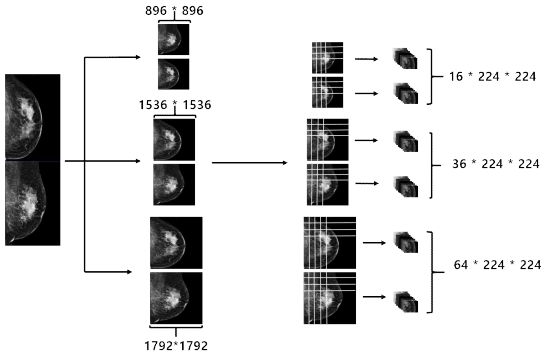


그림 2 Patch extraction 과정

제안한 모델의 pCR 예측 성능 검증을 위해 유방촬영술 영상을 이용한

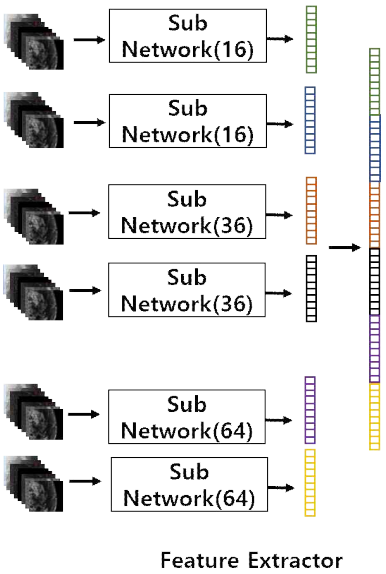


그림 3 Feature Extractor 구조

유방암 진단에서 높은 성능을 보인 세 모델[6, 7, 8]과 테스트셋으로 성능 비교하였다. 세 모델은 ROI 기반 모델로 영상에서 ROI를 추출해 학습, 성능 확인하였다. 모델 성능은 AUC(Area Under Curve), 정확도, 민감도, 특이도로 평가하였다.

비교 결과 세 비교 모델은 평균 AUC 0.5163, 정확도 0.5778, 민감도 0.3444, 특이도 0.7222를 보였다. 비교 모델 중 Arevalo et al[7]과 Rampun et al[8]은 민감도 0.9와 0.9333로 높은 민감도를 나타내는 것을 확인하였다. 제안한 모델은 AUC 0.8033, 정확도 0.75, 민감도 0.7333, 특이도 0.7667로 민감도를 제외한 다른 지표에서 세 모델 대비 우수한 성능을 보였다. 이 결과를 통해 ROI를 입력으로 사용하는 때에 비해 MLO, CC view 전체 영상을 모델의 입력으로 사용했을 때 높은 예측 성능을 보인다는 것을 알 수 있었다.

Ⅲ. 결론

병리학적 완전 관해율(pCR)은 유방암 치료법 중 하나인 선행 항암요법(NAC)의 결과를 평가하는 지표 중 하나이다. NAC 이전에 pCR 예측은 환자의 치료 방향 결정에 큰 영향을 끼친다. 본 논문에서는 기존 pCR 예측 연구에 사용되던 MRI 대신 유방암 진단 과정에서 널리 사용되는 유방촬영술을 통해 pCR을 예측해 환자에게 불필요한 과정을 생략하고, 효과적인 치료 방향 결정을 가능하게 하고자 했다. 제안한 멀티사이즈 유방촬영술 패치 모델은 리사이즈한 영상에서 동일한 크기로 추출한 패치셋을 모델의 입력으로 사용해 다양한 범위의 정보를 학습하게 하였다. 제안 모델은 AUC 0.8033으로 비교 모델들에 비해 우수한 예측 성능을 보였고, 이를 통해 유방촬영술의 MLO, CC view 전체 영상을 입력으로 사용하는 것이 pCR 예측 성능에 영향을 미치는 것을 알 수 있었다.

ACKNOWLEDGMENT

이 연구는 2021년도 산업통상자원부 및 산업기술평가관리원(KEIT) 연구비(20011875), 교육부 재원 한국연구재단 기초연구사업(No. 2020R1I1A3074639), 과학기술정보통신부 재원 한국연구재단 우수신진연구(No.2020R1C1C1006453)의 지원을 받아 수행된 연구임.

참 고 문 헌

[1] Marinovich, M.L., Houssami, N., Macaskill, P., Sardanelli, F., Irwig, L., Mamounas, E.P., Von Minckwitz, G., Brennan, M.E., Ciatto, S." Meta-analysis of magnetic resonance imaging in detecting residual breast cancer after neoadjuvant therapy." Journal of the National Cancer Institute 105, 2013.

[2] Marinovich, M.L., Macaskill, P., Irwig, L., Sardanelli, F., Mamounas, E., von Minckwitz, G., Guarneri, V., Partridge, S.C., Wright, F.C., Choi, J.H." Agreement between mri and pathologic breast tumor size after neoadjuvant chemotherapy, and comparison with alternative tests: individual patient data meta-analysis" BMC cancer 15, 2015.

[3] Duanmu, H., Huang, P.B., Brahmavar, S., Lin, S., Ren, T., Kong, J., Wang, F., Duong, T.Q.. "Prediction of pathological complete response to neoadjuvant chemotherapy in breast cancer using deep learning with integrative imaging, molecular and demographic data" International Conference on Medical Image Computing and Computer-Assisted Intervention, 2020.

[4] Qu, Y.H., Zhu, H.T., Cao, K., Li, X.T., Ye, M., Sun, Y.S. "Prediction of pathological complete response to neoadjuvant chemotherapy in breast cancer using a deep learning (dl) method"Thoracic cancer 11, 2020.

[5] Niell, B.L., Freer, P.E., Weinfurtner, R.J., Arleo, E.K., Drukteinis, J.S. "Screening for breast cancer," Radiologic Clinics 55, 2017

[6] L'evy, D., Jain, A. "Breast mass classification from mammograms using deep convolutional neural networks." arXiv:1612.00542, 2016

[7] Arevalo, J., Gonzalez, F.A., Ramos-Pollan, R., Oliveira, J.L., Lopez, M.A.G. "Representation learning for mammography mass lesion classication with convolutional neural networks." Computer methods and programs in biomedicine 127, 2016

[8]Rampun, A., Scotney, B.W., Morrow, P.J., Wang, H. "SBreast mass classification in mammograms using ensemble convolutional neural networks." 2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom)., 2018

[9] He, K., Zhang, X., Ren, S., Sun, J. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition., 2016

유방 초음파 진단 일반화 성능 향상을 위한 캡슐 네트워크 기반 분류 모델

김찬호 [1], 김원화 [1,2], 김혜정 [1,2], 김재일 [1]
[1] 경북대학교, [2] 경북대학교병원

knight970815@gmail.com, greenoaktree9@gmail.com, ant637@naver.com, threeyears@gmail.com

A capsule network-based model for improving diagnostic generalizability using breast ultrasonography

Chanho Kim[1], Won Hwa Kim[1,2], Hye Jung Kim[1,2], Jaeil Kim[1]
[1]Kyungpook National Univ., [2]Kyungpook National University Chilgok Hospital

요 약

최근 유방 초음파 영상을 활용한 기계학습 기반 컴퓨터 보조 진단 기술이 활발하게 발표되고 있으나, 유방 초음파 영상의 다양성과 기계학습 기술의 낮은 일반화 성능과 신뢰도 때문에 실제 진단 과정에 적용하기에 어려움이 있다. 본 논문은 초음파 영상에서 유방 종양 진단에 대한 일반화 성능을 높이기 위해 캡슐 네트워크 기반 유방 종양 감별 진단 모델을 제안한다. 또한, 다수의 초음파 영상 장비와 병원으로부터 획득한 영상 데이터를 내부 기관 데이터 셋과 외부 기관 데이터셋으로 나누어 제안한 모델의 성능을 평가하고, 일반화 성능 향상을 검증한 실험 결과를 소개한다.

Ⅰ. 서 론

유방 종양 진단에서 초음파 영상은 비침습적이고 인체 내부 조직에 대한 높은 시각화 성능과 안정성, 적은 비용을 장점으로 유방암 조기 진단에 일반적으로 사용되고 있다 [1]. 최근 초음파 영상 기기와 영상처리 및 기계학습 기술의 발전에 따라 종양의 양/악성 판별과 종양 검출을 수행하여 진단 정확도를 높이고자 하는 컴퓨터 보조 진단 기술이 활발하게 발표되고 있다 [2,3]. 하지만 유방 초음파 영상은 종양의 다양성과 잡음 등 여러 요인에 따라 영상 변이가 다양하게 나타나고 기계학습 기술의 낮은 일반화 성능과 신뢰도 때문에 실제 진단 과정에 적용하기에 어려움이 있다. 본 논문에서는 기계학습 기술의 일반화 성능을 높이기 위해 가우시안 피라미드와 캡슐 네트워크를 활용한 유방 초음파 분류 모델을 제안한다. 또한, 다수의 초음파 기기와 병원으로부터 수집한 유방 초음파 영상을 내부 기관 데이터 셋과 외부 기관 데이터셋으로 나누어 제안한 모델의 효용성을 평가한다.

Ⅱ. 본론

유방 초음파 영상 데이터

본 연구를 위하여 두 병원에서 내부 기관 데이터와 외부 기관 데이터를 획득하였다. 내부 기관 데이터는 총 1,568 장(양성 768 장, 악성 800 장)의 초음파 영상으로 모델 학습용 1,000 장(양성 500 장, 악성 500 장)과 학습 중 모델 튜닝용 200 장(양성 100 장, 악성 100 장), 모델 검증용 368 장(양성 168 장, 악성 200 장)으로 나누었다. 외부 기관 데이터는 257 장(양성 129 장, 악성 128 장)으로 모델 검증용으로 사용하였다. 수집된 영상 데이터는 익명화 및 비식별화를 거친 후 화소값을 0 과 1 사이로 정규화하였다. 유방 종양의 양/악성 분류는 임상 진단 결과를 따라 결정되었다.

유방 종양 감별 진단을 위한 모델 구조

본 논문에서는 가우시안 피라미드와 캡슐 네트워크를 활용한 유방 초음파 분류 모델을 제안한다. 가우시안 피라미드는 가우시안 필터를 사용하여 초음파 영상의 텍스쳐 정보를 제거하고 입력 영상의 크기를 줄여 다양한 크기의 종양과 그 형태를 모델에 학습시킨다. 가우시안 피라미드의 하위 계층에서 상위 계층으로 올라갈수록 영상의 텍스쳐 정보는 사라지고 종양의 모양에 대한 정보만 남게 된다. 이후 합성곱 신경망 모델을 통해 다중-크기 특징을 추출한다. 다양한 크기에 대한 정보를 반영하기 위해 합성곱 층을 병렬로 구성하였다. 이렇게 추출된 특징 맵으로 종양의 양/악성을 분류하기 위해 캡슐 네트워크를 활용하였다 [3]. 캡슐 네트워크는 캡슐이라는 벡터를 단위로 하여 연산을 수행한다. 스칼라값을 단위로 하는 기존의 신경망보다 고차원의 특징을 학습할 수 있다. 합성곱 모델을 거친 특징 맵을 캡슐 형태로 변형하여 다이나믹 라우팅 과정을 수행한다. 다이나믹 라우팅 과정은 아핀 변환을 통해 상위 캡슐을 예측하고 예측 벡터를 실제 아웃풋에 따라 라우팅하여 최종 출력 벡터를 계산한다. 이 과정을 반복하면서 실제 출력과 유사한 예측 벡터에 가중치가

추가되고 입력 영상에 없는 개체의 벡터는 짧아지고, 입력 영상에 존재하는 개체의 벡터는 길어지게 된다. 이후 비선형 함수인 스퀴시 함수를 통해 최종 출력 벡터의 길이는 양/악성에 대한 확률을 나타내게 된다.

진단 정확도 평가

분류 성능을 평가하기 위해 제안 모델과 비교 모델의 정확도와 민감도, 특이도를 평가하였다. 성능 비교를 위해 이미지 분류 문제에 좋은 성능을 보이는 GoogLeNet [5]와 Inception-v3 [6] 분류 모델을 초음파 영상으로 학습하였다. 또한, 제안 모델과 마찬가지로 종양의 텍스쳐 정보를 지우고 모양에 대한 정보를 학습하기 위해 입력 영상에 무작위 크기의 가우시안 필터를 적용하여 학습하였다. GoogLeNet 과 Inception-v3 의 경우 내부 기관 데이터 셋에서 각각 92.66%와 96.47%로 높은 성능을 달성하였지만, 외부 기관 데이터 셋에서는 70.04%와 77.43%로 성능이 하락하였다. 또한, 입력 영상에 가우시안 필터를 적용하여 블러된 영상을 학습한 GoogLeNet 과 Inception-v3 의 경우 내부 기관 데이터 셋에서 각각 93.75%와 95.11%의 정확도를 보여주었고, 외부 기관 데이터 셋에서 77.43%와 78.99%의 정확도를 달성하여 두 모델 모두 일반화 성능이 향상된 결과를 보여주었다. 제안 모델은 내부 기관과 외부기관 데이터 셋 모두 각각 97.01%와 88.72%로 가장 높은 성능을 달성하였다.

잡음에 대한 강건함 평가

잡음은 초음파의 내재한 특성으로 초음파 영상의 해상도를 훼손하여 영상을 해석하기 어렵게 만든다. GoogLeNet 과 Inception-v3 의 경우 잡음이 추가된 영상에 대한 정확도가 급격히 하락하였고, 두 모델 모두 블러된 영상을 통해 종양의 모양에 대한 정보를 학습하면 잡음에 대한 강건함이 일부 증가하였다. 이에 반해, 제안 모델은 잡음이 추가되어도 일정한 성능을 유지하였다. 이를 통해 다이나믹 라우팅 과정과 종양의 모양에 대한 학습 및 가우시안 피라미드를 통한 다중-크기 특징에 대한 학습이 잡음에 대한 분류 모델의 강건함을 향상할 수 있음을 알 수 있었다.

Ⅲ. 결론

본 논문에서는 캡슐 네트워크를 활용하여 초음파 영상에서 유방 종양을 감별 진단하기 위한 분류 모델을 제안한다. 제안한 방법은 가우시안 피라미드를 활용하여 종양 영상의 텍스쳐 정보뿐만 아니라 종양의 모양에 대한 정보와 다중-크기 특징에 대해 학습한다. 또한, 캡슐 벡터와 다이나믹 라우팅 과정을 통해 강건한 특징을 학습하여 기존의 합성곱 신경망보다 높은 일반화 성능을 보여준다. 이를 평가하기 위해, 다수의 초음파 장비와 병원으로부터 획득한 영상 데이터를 내부 기관 데이터 셋과 외부 기관 데이터 셋으로 나누어서 진단 정확도를 측정하고 일반화 성능을 검증하기 위한 실험을 진행하였다. 기존의 합성곱 신경망은 학습에 사용된 내부 기관 데이터 셋에서 높은 성능을 달성하였지만, 학습에 사용되지 않은 외부 기관 데이터 셋에서 오히려 성능이 감소하였다. 반면, 제안한 모델은 내부 기관과 외부 기관 데이터 셋 모두 높은 성능을 달성하였다. 또한, 입력 영상에 초음파 영상에서 흔히 발생하는 잡음을 추가하여 제안한 모델의 강건함을 측정하였다. 일반적으로 잡음의 세기를 높일수록 분류 모델의 성능이 떨어지지만, 제안한 모델은 잡음에 상관없이 일정한 성능을 보여주었다. 이러한 방법과 더불어 신뢰도 분석과 설명 가능한 인공지능을 통하여 컴퓨터 보조 진단 시스템의 범용성을 확보하고 실제 진단 과정에 활용할 수 있을 것으로 기대한다.

ACKNOWLEDGMENT

이 연구는 2021 년도 산업통상자원부 및 산업기술평가관리원(KEIT) 연구비(20011875), 교육부 재원 한국연구재단 기초연구사업(No. 2020R1I1A3074639), 과학기술정보통신부 재원 한국연구재단 우수신진연구(No.2020R1C1C1006453)의 지원을 받아 수행된 연구임.

참 고 문 헌

[1] Wang, Lulu. "Early diagnosis of breast cancer." Sensors 17.7 (2017): 1572.

[2] M. H. Yap, G. Pons, J. Martí, S. Ganau, M. Sentís, R. Zwiggelaar,A. K. Davison, and R. Martí, “Automated breast ultrasound lesionsdetection using convolutional neural networks,”IEEE journal ofbiomedical and health informatics, vol. 22, no. 4, pp. 1218- 1226, 2017

[3] S. Han, H.-K. Kang, J.-Y. Jeong, M.-H. Park, W. Kim, W.-C. Bang, andY.-K. Seong, “A deep learning framework for supporting the classificationof breast lesions in ultrasound images,”Physics in Medicine & Biology,vol. 62, no. 19, p. 7714, 2017

[4] Sabour, Sara, Nicholas Frosst, and Geoffrey E. Hinton. "Dynamic routing between capsules." arXiv preprint arXiv:1710.09829 (2017).

[5] Szegedy, Christian, et al. "Going deeper with convolutions." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

[6] Szegedy, Christian, et al. "Rethinking the inception architecture for computer vision." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

어린이 보호구역 내 사고 상황에서의 AEB 기능 효과성에 관한 실차 실험 연구

장우영, 최형수, 이학주, 박지수, 김홍익, 이태희*
지능형자동차부품진흥원
(wyjang, xu, lhj0601, hongik, jspark)@kiapi.or.kr, *thlee@kiapi.or.kr

A Study with real-vehicle test on the effectiveness of AEB function in school zone accidents

Jang Woo Young, Choi Hyung Soo, Lee Hak Ju, Park Ji Soo, Kim Hong Ik, Lee Tae Hee*
Korea Intelligent Automotive Parts Promotion Institute

요 약

본 논문은 2019년 제정된 특정범죄 가중처벌 등에 관한 법률 제5조의13(일명 '민식이 법') 관련하여 어린이 보호구역 내 사각지역에서 어린이가 달려 나오는 상황을 실차 시험 시나리오로 구성하고-이 상황에서 AEB(Autonomous Emergency Braking) 기능의 사고예방에 대한 효과성을 입증하기 위해 AEB Off 상태에서 운전자의 제동 능력과 AEB On 상태에서 차량 제동능력 간 비교 실차 시험을 진행하였다.

Ⅰ. 서 론

2019년 어린이 사망 교통사고 발생 후 특정범죄 가중처벌 등에 관한 법률 제5조의13(일명 '민식이 법')이 제정되었다. 해당 법률은 운전자의 부주의로 어린이가 사망할 경우 무기 또는 3년 이상의 징역에 처하고 피해자가 상해를 입으면 1년 이상 15년 이하의 징역이나 500만 원 이상 3000만 원 이하 벌금을 부과한다는 내용을 담고 있다.[1] TAAS(교통사고분석시스템)에 따르면 어린이 보호구역 내 사고 건수는 2017년부터 2020년 까지 차례대로 487건, 473건, 589건, 507건으로 감소되지 않고 있다.[2] 따라서 해당 법률이 제정된 목적으로는 어린이보호구역 내 어린이 교통사고 건수를 줄이겠다는 목적이 있는 것으로 파악된다. 하지만 일각에서는 운전자에 대한 과잉 처벌이라는 의견도 나타난다.[3] 민식이 법 시행 이후 어린이 보호구역 내 사고 상황을 가정하여 다양한 시뮬레이션 분석이 이루어지고 있다.[4]

본 논문에서는 어린이 보호구역 내 사각지역에서 어린이가 달려 나오는 상황을 실차 시험 시나리오로 구성하고 AEB(Autonomous Emergency Braking) Off 상태에서 운전자의 제동능력과 AEB On 상태에서 차량 제동능력 간 비교 실차 시험을 진행하고 충돌 여부, 정지거리, 충돌 속도에 대한 분석을 수행하였다.

Ⅱ. 본론

2.1 시나리오 구성

시험 시나리오는 실제 발생할 수 있는 상황으로 구성하기 위하여 어린이가 달리며 횡단보도를 지나간다고 가정하였으며 어린이 달리기 속도는 문화체육관광부의 국민체력측정 통계자료와 EuroNCAP 규격을 참고하였다.[5][6] 통계청에 등재되어 있는 자료에 따르면 2007년 초등학생 남자(1 ~ 4학년) 50 m 달리기 평균 속도는 16.2 km/h 이고 EuroNCAP 규격에서는 성인 보행자가 달리는 시나리오의 타겟 속도를 8 km/h로 규정하고 있다. 따라서 본 시험에서는 어린이 타겟 속도를 어린이가 통상적으로 달리는 속도로 구성하고자 50 m 달리기 속도와 EuroNCAP 규격 속도의 중간값인 12 km/h 로 선정하였다. 시험환경 구성은 도로교통공단 홈페이지에 따른 최소 차로 폭인 2.75 m 의 왕복 2차선으로 구성하였으며 차선의 폭은 0.15 m로 구성하였다.[7] 장애물 차량의 위치는 어린이 보호구역 내 등, 하굣길에 발생할 수 있는 상황을 구성하고자 주행 차로의 반대 차선에 끝부분에 위치하였으며, 대상 차량 속도는 어린이제한구역 속도인 30 km/h로 구성하고자 하였으나, 시험 진행 간 대상 차량 속도는 GPS 속도로 계측되기 때문에 계기판 속도보다 높은 점을 감안하여 20 km/h로 선정하였다. 시나리오 구성도는 그림 1과 같으며 실차 시험 현장 사진은 그림 2, 3과 같다.

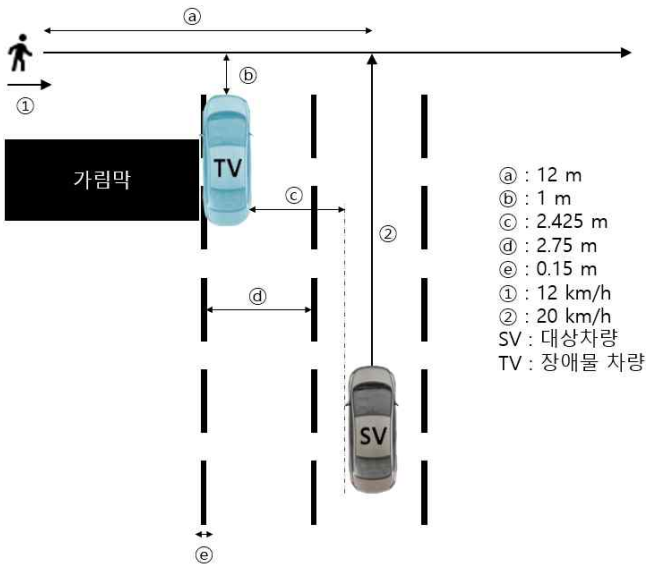


그림 1. 시나리오 구성도



그림 2. 실차실험 현장 사진 1



그림 3. 실차실험 현장 사진 2

2.2 대상차량 및 보행자 타겟 구성

그림 4와 같이 대상차량은 AEB 기능이 장착되어 있는 M사의 2017연식 E-Class로 선정하였으며 시험 장비는 ABDynamics사의 조향로봇(SR60), 브레이크 악셀로봇(CBAR), 그리고 차량의 거동과 속도 확인을 위한 OXTS 사의 3축 자이로와 DGPS(RT3002)로 구성하였다. 데이터 측정을 위한 Data acquisition(DAQ) 장비는 Dewesoft 사의 데이터 계측장비(SIRIUS)로 구성하였으며 측정 소프트웨어는 Dewesoft X3를 사용하였다. 장애물 차량으로는 T사의 승용차로 구성하였다. 보행자 타겟은 EuroNCAP 시험 장비인 Soft Pedestrian Target(SPT)를 사용하였으며 타겟 종류는 어린이 타겟으로 구성하였다. 그림 5와 같이 시험 장소는 지능형자동차부품진흥원 내 Proving ground에서 진행하였다.



그림 4. 대상차량 장비 장착 사진

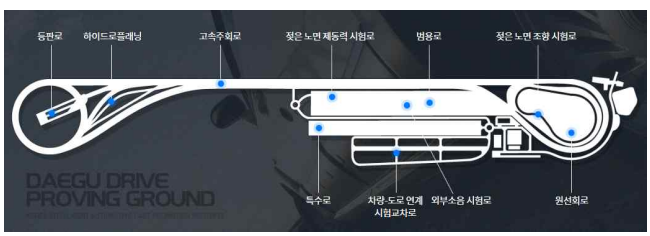


그림 5. Proving ground in KIAPI

2.3 AEB Off 상태의 운전자 선정

AEB Off 상태의 운전자는 시험의 공정성을 위해 실차시험 경력이 없고 위 시험을 처음 접하는 남자 4인, 여자 4인으로 선정하였다. 시험을 진행하기 전 어린이 타겟의 위치를 알려주지 않고 200 m 가량 직선주행 시험이라고 고지하였다. 해당 시험이 끝난 후 나이, 운전경력에 관한 정보와 타겟과 충돌을 방지하기 위한 제동 가능 여부에 대한 점수를 산정하는 감성평가를 진행하였다. 제동 가능 여부 점수는 5점(제동 가능), 4점(거의 제동 가능), 3점(보통), 2점(거의 불가능), 1점(제동 불가능)으로 구성하였다. 표 1과 같이 운전자 중 남성의 경우 나이 범위는 29세 ~ 34세, 운전 경력 범위는 1년 ~ 12년 이었다. 제동 가능 여부 평균 점수는 2점을 부여하였다. 표 2와 같이 여성의 경우 나이 범위는 24세 ~ 46세, 운전 경력 범위는 1년 ~ 26년 이었다. 제동 가능 여부 평균 점수는 1.25점을 부과하였다.

표 1. 남자 운전자 감성평가표

구분	남자1	남자2	남자3	남자4
나이 (세)	29	34	31	34
운전경력 (년)	1	10	10	12
제동 가능 여부	3	2	2	1

표 2. 여자 운전자 감성평가표

구분	여자1	여자2	여자3	여자4
나이 (세)	31	38	24	46
운전경력 (년)	10	10	1	26
제동 가능 여부	1	1	1	2

2.4 시험 결과

표 3. 시험 결과표

항목	구분	충돌/미충돌	충돌속도 (km/h)
AEB On	1회	미충돌	0
	2회	미충돌	0
	3회	미충돌	0
	4회	미충돌	0
AEB Off	남자 1	충돌	20
	남자 2	충돌	20
	남자 3	충돌	20
	남자 4	충돌	20
	여자 1	충돌	20
	여자 2	충돌	19
	여자 3	충돌	20
	여자 4	충돌	20

표 3의 시험 결과표를 보면 AEB Off 상태에서 사람이 운전했을 경우 모두 충돌이 발생하였으며 AEB On 상태에서는 충돌이 발생하지 않았다. 그림 6을 보면 AEB On 상태에서 TTC 0.42초에 FCW가 울렸으며 0.37초에 AEB 기능이 작동되었고 차량이 완전 정차 후 타겟과의 정지거리는 0.33 m 이다. 분석데이터는 DIAdem으로 분석하였다.

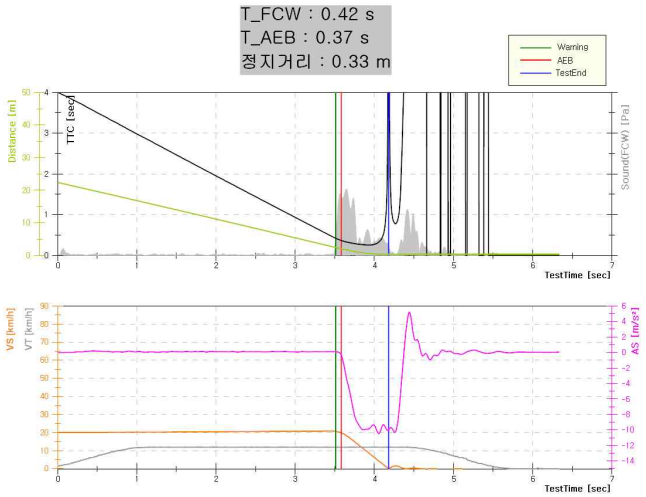


그림 6. AEB On 상황 분석 데이터

그림 7과 같이 운전자가 어린이 타겟을 인지한 시점은 타겟 전체가 차량 내 대시보드 카메라에 보이는 시점으로 판단하였다. 해당 시점의 TTC는 약 0.9초 이고 타겟과의 거리는 약 5 m 이다. 0.9초는 운전자의 인지반응 시간을 고려할 때 제동하기 충분하지 않은 시간으로 판단된다.[8] 따라서 운전자가 제동하기 힘든 상황이라는 것을 알 수 있었다. 그림 8은 AEB Off 상태에서 운전자가 제동을 한 여자 3의 분석데이터이며 충돌속도는 19 km/h이다. 그림 9의 분석데이터를 보면 나머지 인원 7명은 제동을 하지 못하였고 충돌속도는 20 km/h 이다. 측정 데이터는 Dewesoft X3 소프트웨어로 측정되었으며 분석 데이터는 DIAdem 소프트웨어로 분석하였다.



그림 7. 어린이 타겟 전체가 차량 내 대시보드 카메라에 보이는 시점

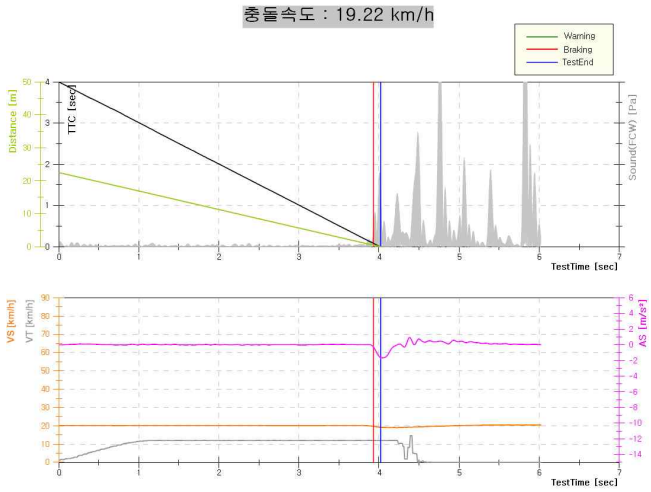


그림 8. AEB 기능 Off 상황 19 km/h 충돌 분석 데이터

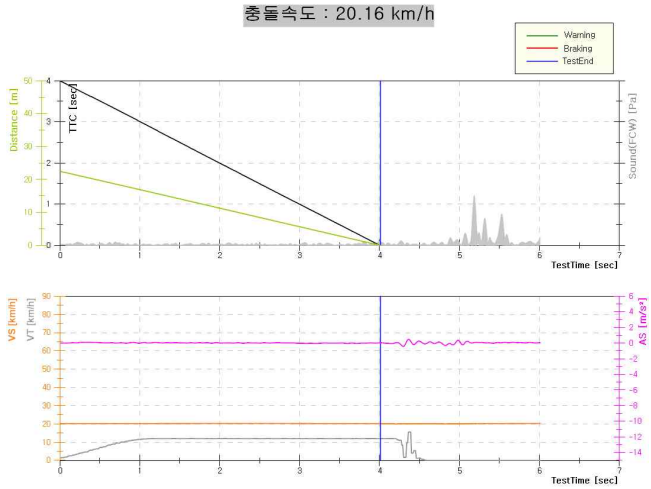


그림 9. AEB 기능 Off 상황 20 km/h 충돌 분석 데이터

Ⅲ. 결론

어린이보호구역 내 사각지역에서 어린이가 달려 나오는 상황을 실차 시험 시나리오로 구성하고 AEB Off 상태에서 운전자의 제동 능력과 AEB On 상황을 비교해본 결과 AEB Off 상황에서는 충돌이 발생하였고 AEB On 상황에서는 충돌을 피할 수 있었다. 해당 시험과 같이 AEB 기능이 없는 차량의 경우 규정 속도 이내로 운전하더라도 갑자기 달려 나오는 어린이와는 충돌을 피하기 어려우므로 일반인 운전자에게 민식이 법은 가혹하며, 따라서 AEB 기능의 의무화를 검토할 필요가 있다. 본 논문에서는 어린이 보호 구역에서 운전자가 피하기 힘든 사고를 방지하기 위한 AEB 기능의 사고 방지 효과성을 입증하였다. 향후에는 어린이보호구역 내 시나리오를 다양화 하고 차종별, 속도별 시험을 진행하여 발생할 수 있는 여러 가지 상황에서 사고방지를 위한 AEB 기능의 효과성을 입증하는 연구를 수행할 예정이다.

ACKNOWLEDGMENT

본 연구는 산업통상자원부 자동차 산업 기술 개발 사업 5G 기반 자율주행 융합 기술 실증 플랫폼(과제번호 : P0013840)에 의해 부분 수행되었습니다. 연구비 지원에 감사드립니다.

참 고 문 헌

[1] 국가법령정보센터, - 법령 특정범죄 가중 처벌 등에 관한 법률 (https://www.law.go.kr).

[2] TAAS(교통사고분석시스템) - 어린이 교통사고 관련 (https://www.law.go.kr).

[3] 박상민, "특정범죄 가중처벌 등에 관한 법률 제5조의13(일명 '민식이법')에 대한 형법 이론적 검토,", 한국법학회, pp. 151-172, Mar. 2021

[4] 정동훈, 이수범, "어린이 보호구역 어린이 사고 특성에 따른 AEBS 장착 차량의 사고감소 효과에 대한 연구,",대한교통학회, pp. 22-22, Feb. 2021

[5] KOSIS - 건강 체력 기준 : 초등학생 남자, (https://kosis.kr/statHtml/statHtml.do?orgId=113&tblId=DT_113_STBL_1012524).

[6] 권병헌, 김봉섭, 이선봉, "국내 도로환경과 AEB VRU Test Protocol v3.0.1을 고려한 AEB V2P 시험평가에 관한 연구.", 한국통신학회, pp. 174-175, Jun. 2019

[7] 도로교통공단 - 교통법규 QnA : 차선, 차로, (https://www.koroad.or.kr/kp_web/roadQnAList.do).

[8] 장명순, 이성관, 김진수, 홍성민, 이강진, "운전자 인지반응시간의 시대적 진화 연구,", 대한교통학회, pp. 55-61, Dec. 2011

신경망 탐색을 이용한 최적의 운전자 행위 분류 모델

성재호, 유민우, 김지훈, 한동석*
경북대학교

*dshan@knu.ac.kr

Optimal Driver Behavior Classification Model Using Neural Network Search

Jae Ho Seong, Min Woo Yoo, Ji Hun Kim, Dong Seog Han*
Kyungpook Univ.

요 약

본 논문에서는 최적의 운전자의 행위를 분류하기 위하여 ENAS(efficient neural architecture search)기반의 CNN(convolutional neural network) 모델을 제안한다. 기존 딥러닝 기반 DMS(driver monitoring system)는 CNN을 사용하여 운전자의 상태를 예측하는데 효과적이다. 그러나 CNN 기반의 시스템은 연산량이 많고 임베디드 기기와 같은 제한된 자원 환경에서 적용을 위한 모델 구조 최적화가 요구된다. 본 논문에서는 제안한 ENAS기반의 운전자 행위 분류 모델이 기존 경량화 모델의 분류 성능 보다 향상됨을 보였다.

Ⅰ. 서 론

운전자의 부주의한 행동은 교통사고의 주요한 원인이다[1]. 따라서 DMS(driver monitoring system)를 통해 운전자의 의도를 예측하고 긴급 상황에서 차량 제어에 개입하여 사고를 방지하는 것이 중요하다. 최근 DMS는 CNN(convolutional neural network)을 주로 사용한다. 예를 들어 CNN을 이용한 얼굴 검출, 헤드 포즈, 시선 추정 등의 알고리즘을 통해 운전자의 상태를 예측하고자 하는 연구가 활발히 진행중이다[2]. 기존의 CNN 기반의 딥러닝 모델은 운전자의 상태를 예측하는데 매우 효과적이다. 그러나 CNN 기반의 DMS는 계산량이 많고, 차량 내부에 적용을 하기 위한 임베디드 기기는 제한적인 자원 환경을 가진다. 따라서 DMS에 운전자 행위 분류 모델 적용을 위한 경량화 작업이 필요하다. 또한 딥러닝 기반의 분류 모델은 데이터의 종류와 네트워크의 구조에 따라 성능이 변화하므로 최적화 작업에 많은 노력과 시간이 소요된다. 이를 해결하기 위하여 목표 데이터에 모델을 자동으로 최적화할 수 있는 AutoML(automated machine learning) 기법들이 제안되었다. 그 중에서 NAS(neural architecture search) 기법을 사용하면 CNN 네트워크의 구조를 자동으로 설계하고 최적화할 수 있다 [3]. 그리고 NAS 기법 중 ENAS(efficient neural architecture search)는 모델의 가중치를 공유하여 단일 GPU로 짧은 시간안에 CNN 네트워크를 탐색할 수 있다[4].

본 논문에서는 제한된 자원 환경에서 운전자 행위 분류 모델을 적용하기 위하여 ENAS 기반의 CNN 모델을 제안한다. 실험을 통하여 수집한 운전자 행위 데이터셋에 대해 최적의 운전자 행위 분류 모델 탐색하고 제안한 모델의 분류 성능이 기존 딥러닝 기반 경량화 모델보다 성능이 향상됨을 보였다.

Ⅱ. 본론

2.1 ENAS 모델 생성 과정

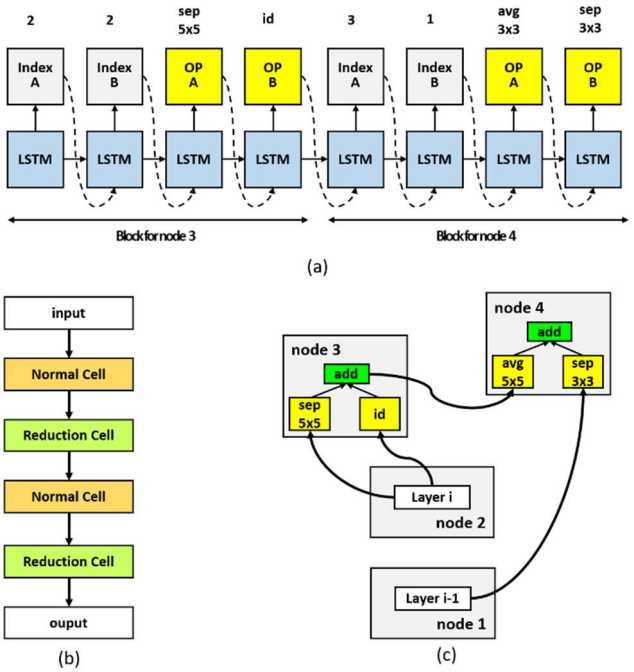


그림 1. ENAS 모델 생성 과정. (a) RNN 제어기. (b) CNN의 전체 네트워크 구조. (c) CNN 셀 생성 과정.

ENAS는 그림 1(b)와 같이 전체 CNN을 설계하는 것이 아닌, 작은 모듈 형태의 일반 셀과 축소 셀을 생성하고, 이를 연결하여 전체 CNN을 구성한다. 일반 셀은 같은 차원의 특징 맵을 생성하고, 축소 셀은 특징 맵의 높이와

너비를 반으로 줄인다. 그림 1(a)와 같이 RNN(Recurrent Neural Network) 제어기에서는 CNN 셀을 설계하기 위한 연산과 연산을 수행할 노드의 번호를 선택한다. 그림 1(c)는 CNN 셀 생성 과정을 나타낸다. 예를 들어 모듈이 B개의 노드를 가진다면 노드 1 과 2 는 이전 모듈의 출력 값을 그대로 입력으로 사용한다. 나머지 B-2 개의 노드는 현재 노드의 입력 값으로 제어기에서 선택된 연산과 노드의 번호를 조합한 후 각각의 연산이 적용된 특징 맵을 각 픽셀별로 값을 더한다. 사용된 연산으로는 depth-wise separable convolution 3x3 와 5x5, max pooling, average pooling, identity를 사용하였다. 생성된 CNN 모델은 학습을 시킨 후 검증 정확도를 계산하고 이를 강화학습의 보상으로 사용하여 제어기가 학습된다. 이 과정을 반복하여 제어기에서 최적의 모델을 설계할 수 있다.

2.2 ENAS 학습 과정

RNN 제어기 학습: 제어기에는 강화학습의 정책(Policy) $\pi(m;\theta)$과 보상 $r(m,w)$가 있다. m은 정책에서 선택된 모델이고, θ는 제어기의 매개변수, w는 생성된 모듈의 매개변수이다. Adam(Adaptive Moment Estimation)을 사용하여 보상의 기댓값인 $\mathbb{E}_{m\sim\pi}(m;\theta)\,[r(m,w)]$ 을 최대화하는 방향으로 제어기를 학습한다.

생성된 CNN 모델 학습: 이 단계에서는 CNN의 매개변수 w를 업데이트한다. w에대해 SGD(Stochastic Gradient Descent)를 사용하여 손실함수의 기댓값 $\mathbb{E}_{m\sim\pi}(m;\theta)\,[\mathcal{L}(m;w)]$를 최소화한다. $[\mathcal{L}(m;w)]$는 크로스 엔트로피 손실(Cross-entropy loss)를 사용하여 계산한다. 식(1)과 같이 몬테 카를로 추정(Monte Carlo estimate)을 사용하여 w에 대해 계산한다.

$$\nabla\mathbb{E}_{m\sim\pi}(m;\theta)\,[\mathcal{L}(m;w)] \approx \frac{1}{M}\sum_{i=1}^{M}\nabla_w\mathcal{L}(m_i,w) \quad (1)$$

m_i 는 $\pi(m;\theta)$ 에서 생성된 모델이다. $M=1$ 이 잘 작동하므로 단일 모델로 w 에 대한 생성된 모델의 기울기를 계산할 수 있다.

2.3 실험 환경 및 결과

본 장에서는 실험 환경 및 결과에 대해서 설명한다. 운전자 행위 데이터셋은 차량 내부 환경에서 카메라 사용해서 수집하였다. 표 1 과 같이 운전자의 행위는 음식 섭취, 평소 상태, 휴대폰 사용, 흡연으로 분류한다. CPU i7-6700 와 Nvidia Geforce GTX 1080 Ti 그래픽 카드를 사용하여 네트워크를 학습시켰고, 18 시간 이내에 CNN 모델을 탐색하였다. ENAS로 탐색된 일반 셀과 축소 셀 은 그림 2 와 같다. 제안한 ENAS 기반 CNN 모델은 운전자 행위 분류 데이터셋에 대해 약 11%의 테스트 오차를 보였다. 이는 표 2 와 같이 기존의 경량화 모델보다 비교적 낮은 파라미터 숫자를 가지며 약 5~8% 더 분류 성능이 높다.

표 1. 운전자 행위 데이터 구성

Class Name	Training data	Test data	Total
Eat	12,580	2,050	14,630
Normal	19,120	1,130	20,250
Phone	12,810	1,400	14,210
Smoke	17,650	1,030	18,680

표 2. 분류 모델 성능 비교

Model	ACC	Params
MobileNetV2	79.37	3.4M
MobileNetV3	83.59	5.4M
ENAS(Proposed Model)	88.48	3.6M

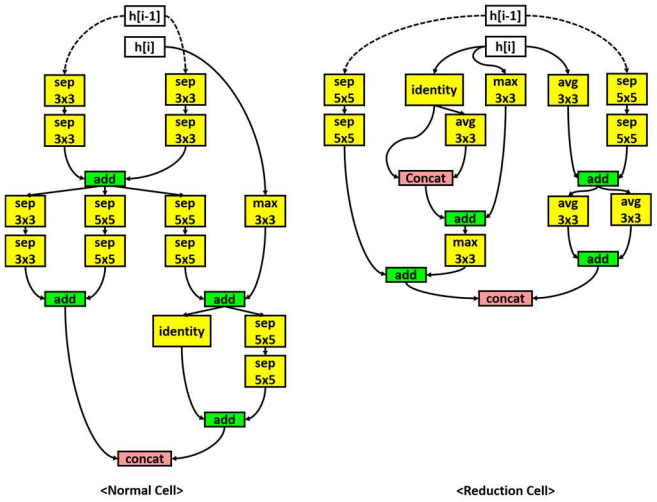


그림 2. ENAS 으로 생성된 일반 셀과 축소 셀

Ⅲ. 결론

본 논문에서는 최적화된 운전자 분류 모델을 설계하기 위하여 ENAS기반의 CNN 모델을 탐색하고 성능을 확인하였다. ENAS 기반 CNN 모델은 테스트 데이터에 대해서 약 11%의 오차의 성능을 보였다. 결론적으로 ENAS를 사용하여 신경망을 설계하면, 데이터 셋에 최적화된 고성능의 분류 모델을 탐색할 수 있다. 추후 연구에서는 ENAS의 탐색공간과 연산자를 수정하여 성능을 개선할 수 있을 것으로 기대한다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT 연구센터지원사업의 연구결과로 수행되었음(IITP-2021-2020-0-01808)

참 고 문 헌

[1] Fountas, Grigorios, et al. "Factors Affecting Perceived and Observed Aggressive Driving Behavior: An Empirical Analysis of Driver fatigue, and Distracted Driving." (2019).

[2] Vaca-Recalde, Myriam E., Joshué Pérez, and Javier Echanobe. "Driver Monitoring System Based on CNN Models: An Approach for Attention Level Detection." International Conference on Intelligent Data Engineering and Automated Learning. Springer, Cham, 2020.

[3] Ren, Pengzhen, et al. "A comprehensive survey of neural architecture search: Challenges and solutions." arXiv preprint arXiv:2006.02903 (2020).

[4] Pham, Hieu, et al. "Efficient neural architecture search via parameters sharing." International Conference on Machine Learning. PMLR, 2018.

시각장애인 길 안내를 위한 실외 자율주행 카트 개발

박수용, 홍승준, 최인훈, 정홍주, 김서현, 허의남*

경희대학교

{clapd10, hongsj1022, inhun321, sub06038, seok02h, *johnhuh}@khu.ac.kr

Outdoor Autonomous Driving System for Navigating Visually Impaired Pedestrians

Suyong Bahk, Seungjun Hong, Inhun Choi,

Hong-Ju Jeong, Seo-Hyun Kim, Eui-Nam Huh*
Kyung Hee Univ.

요 약

우리나라 안내견 사업의 공급 문제로 인공지능과 로봇 기술을 활용한 시각장애인 안내 서비스가 큰 주목을 받고 있다. 그러나 로봇의 높은 비용으로 실 수요층 보급에는 한계를 드러내고 있다. 본 논문은 이런 문제를 해결하고 시각장애인에게 실질적인 도움을 줄 수 있도록 보급형 싱글 보드 컴퓨터와 Open API, 오픈 소스 플랫폼을 활용한 실외 자율주행 카트를 제안한다.

Ⅰ. 서 론

우리나라의 안내견 사업은 1993 년도부터 시작되었으나 실 수요 인구에 비해 턱없이 부족한 공급만이 가능한 문제에 직면해 있다. 안내견 공급이 부족한 원인은 안내견을 육성하는데 약 1 년의 기간과 1 억원가량의 비용이 소요될 뿐만 아니라 그렇게 육성한 안내견 후보 중 일부만이 분양된다는 점에 있다.[1] 이에 따라 실제 안내견에 비해 생산되는 데 소요되는 시간 및 비용을 대폭 절감한 실내 및 실외 시각장애인 안내 로봇 서비스가 다수 제안 및 개발되고 있다.[2][3]

하지만 해당 서비스들의 경우 특수하게 제작된 고가의 하드웨어 및 소프트웨어를 사용하였기 때문에 실 수요층에게 보급되는 데 어려움이 있다. 또한 사용자가 이미 알고 있는 길에서 사용자를 보조하는 정도의 안내만을 해 줄 수 있다는 한계점을 지니고 있다.

본 논문에서는 라즈베리파이 및 보급형 센서, 마이크로컨트롤러를 활용한 실외 자율주행 카트를 설계하고, 사용자를 보조하는 정도를 넘어서서 사용자가 알지 못하는 길을 지나더라도 안전하게 목적지에 도달할 수 있도록 도보 안내 Open API[4] 및 오픈 소스 플랫폼 Robot Operating System(ROS)[5], Google TTS API[6]를 활용한 자율주행 및 경로 안내 서비스를 구현하였다.

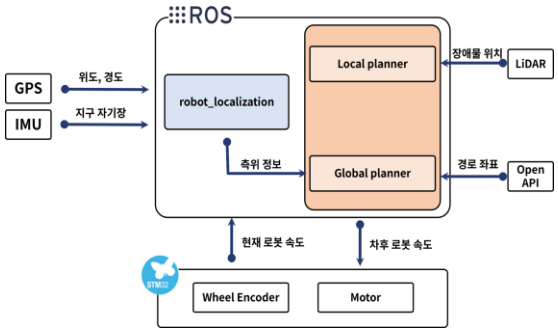
Ⅱ. 본 론



[그림 1. 구동 시나리오]

본 논문에서 제안하는 자율주행카트는 음성 인식을 통해 Open API 로부터 목적지까지의 좌표를 수신한 다음, 측위, 장애물 인식, 목적지까지의 로봇 경로 생성을 수행한다.

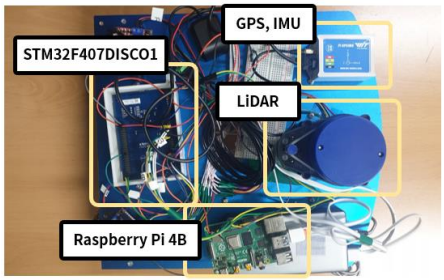
Ⅱ-1. 시스템 구성도



[그림 2. 시스템 구성도]

본 논문에서 제안하는 자율주행카트의 시스템 구성도는 그림 2 와 같다. 자율주행카트의 안내 명령과 안내 출력은 모두 음성으로 이루어진다. 목적지까지의 경로상에 있는 위도 및 경도 좌표가 Open API 를 통해 수신되면 텍스트 파일로 저장한다. 자율주행카트의 측위는 robot_localization 패키지[7]에서 GPS 와 IMU, Wheel Encoder 데이터에 확장 칼만필터[8]를 적용한 Sensor Fusion 으로 구현된다. 자율주행카트는 라이다 센서의 거리 정보를 바탕으로 Local planner 에서 Dynamic Window Approach[9]에 의한 장애물 회피를 먼저 수행한 다음, Global planner 에서 측위 정보와 수신한 좌표 정보를 바탕으로 Djikstra 알고리즘[10]을 이용하여 목표 지점까지의 최단 경로를 생성한다. 자율주행카트의 주행 중 모터 속도는 diff_drive_controller 패키지[11]에 의해 결정되고, 자율주행카트의 마이크로컨트롤러에 시리얼 통신으로 전달된다.

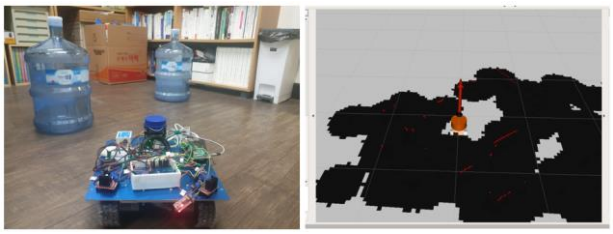
Ⅱ-2. 차량 프로토타입 제작



[그림 3. 실외 자율주행카트 프로토타입]

그림 3 은 제작된 자율주행카트의 프로토타입이다. GPS 및 IMU 센서는 자체적으로 칼만 필터가 적용된 wit 사의 WGAHRS1 모델을 사용하였다. 2D 라이다 센서는 EAI 사의 YDLiDAR X4 모델을 사용하였다. ROS 를 탑재한 싱글 보드 컴퓨터는 Ubuntu 18.04 가 설치된 Raspberry Pi 4B 4GB 모델을 사용하였다. 자율주행카트의 마이크로컨트롤러는 STM32F407DISCO1 보드를 사용하였다.

Ⅱ-3. 구현 결과



[그림 4. 장애물 회피 경로 생성 모습]

그림 4 는 장애물 회피 과정 예시를 보여준다. 라이다의 LaserScan 정보를 해석해 물체의 위치를 파악하고 로봇의 주행 경로를 계산한다.



[그림 5. 실외 경로 생성 모습]

그림 5 는 실외 경로 생성 모습을 보여준다. 카트가 순차적으로 이동해야할 GPS 좌표를 텍스트 파일로 저장하면 ROS 에서 경로 안내 노드를 실행하였을 때 해당 파일을 파싱하여 각 지점을 통과할 수 있는 경로를 생성한다.

Ⅲ. 결론 및 향후 연구 방향

본 논문에서는 싱글 보드 컴퓨터와 센서, 마이크로컨트롤러를 사용한 자율주행카트를 설계하였다. 제안하는 카트의 차별성은 다음과 같다.

1. 오픈 소스 플랫폼 및 Open API 를 적극적으로 활용하여 비용을 절감하고, 확장성을 증진하였다.
2. 음성 인식만으로 목적지까지 경로 안내 및 실외 자율주행이 동시에 가능하여 편의성이 증대된다.
3. 보급형 싱글 보드 컴퓨터 및 센서를 활용하여 시스템을 설계하여 상용화할 경우 보급이 용이하다.
4. 현재 안내견 사업을 통해 분양되는 안내견 개체보다 시간 및 비용이 적게 소요되며, 돌발 상황 등 다양한 상황에서도 다양한 기능을 수행할 수 있다.

본 연구는 안내견 공급 부족 문제를 해결함과 동시에 시각장애인의 생활 반경을 넓히는 데 큰 도움이 될 것이며, 향후 Open API 에서 간혹 보행자 도로가 아닌 차도 위에 경로를 설정하는 문제점을 해결하기 위해 보행자 도로를 감지하는 ROS 패키지를 제작하여 결합할 예정이다. 또한 AI 스피커 등 타 IoT 기기들과 결합하여 카트의 기능성을 증진시킬 계획이다.

ACKNOWLEDGMENT

이 논문은 2021 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2017-0-00294, (대학 ICT 기초연구실) 서비스 이동 지원을 위한 분산형 클라우드 핵심원천기술 연구)

참 고 문 헌

[1] 민동원. "자발참여형 공익 마케팅에 대한 이해." Korea Business Review 18.4 (2014): 1-25.

[2] Xiao, Anxing, et al. "Robotic Guide Dog: Leading a Human with Leash-Guided Hybrid Physical Interaction." arXiv preprint arXiv:2103.14300 (2021).

[3] Tan, Haobin, et al. "Flying Guide Dog: Walkable Path Discovery for the Visually Impaired Utilizing Drones and Transformer-based Semantic Segmentation." arXiv preprint arXiv:2108.07007 (2021).

[4] T MAP API. Available::http://tmapapi.sktelecom.com/main.html

[5] Quigley, Morgan, et al. "ROS: an open-source Robot Operating System." ICRA workshop on open source software. Vol. 3. No. 3.2. 2009.

[6] Google Cloud. Available::https://cloud.google.com/text-to-speech

[7] Moore, Thomas, and Daniel Stouch. "A generalized extended kalman filter implementation for the robot operating system." Intelligent autonomous systems 13. Springer, Cham, 2016. 335-348.

[8] Smith, Gerald L., Stanley F. Schmidt, and Leonard A. McGee. Application of statistical filter theory to the optimal estimation of position and velocity on board a circumlunar vehicle. National Aeronautics and Space Administration, 1962.

[9] Fox, Dieter, Wolfram Burgard, and Sebastian Thrun. "The dynamic window approach to collision avoidance." IEEE Robotics & Automation Magazine 4.1 (1997): 23-33.

[10] Dijkstra, Edsger W. "A note on two problems in connexion with graphs." Numerische mathematik 1.1 (1959): 269-271.

[11] Chitta, Sachin, et al. "ros_control: A generic and simple control framework for ROS." The Journal of Open Source Software 2.20 (2017): 456-456.

다중 에이전트 강화학습 기반 무인운반기 최적 경로 제어

최호빈, 김주봉*, 김찬명**, 한연희*
한국기술교육대학교 컴퓨터공학과, *한국기술교육대학교 미래융합공학전공,
**한국기술교육대학교 첨단기술연구소

{chb3350, *rlawnqhd, **cmdr, *yhhan}@koreatech.ac.kr

Multi-Agent Reinforcement Learning-Based Optimal Route Control for Automated Guided Vehicles

Ho-Bin Choi, Ju-Bong Kim*, Chan-Myung Kim**, Youn-Hee Han*
Department of Computer Science and Engineering, Korea University of Technology and Education,
*Future Convergence Engineering Major, Korea University of Technology and Education,
**Advanced Technology Research Center, Korea University of Technology and Education

요 약

유통 산업과 물류 산업의 발전에 맞춰 대규모 물류창고는 필수적인 요소가 되었고 그러한 물류창고들을 효율적으로 관리 및 운영하기 위해 여러 방법들이 연구되고 있다. 무인운반기는 고중량 물품을 안전하고 빠르게 운반할 수 있어 무인 물류창고에서 많이 사용되고 있다. 현재 사용되는 대부분의 무인운반기는 특정한 룰에 기반하여 제어되며 이 룰은 물류 창고의 레이아웃에 맞춰 구성되어야 한다. 본 논문에서는 특정한 룰없이 다중 무인운반기의 성능을 극대화하기 위해 다중 에이전트 강화 학습을 사용한다. 각 무인운반기는 다른 무인운반기와의 충돌을 피하며 자신의 임무를 가능한 한 빠르게 완료해야 한다. 검증을 위해 강화 학습 환경을 설계하고 대표적인 다중 에이전트 강화 학습 알고리즘인 QMIX 를 사용하여 실험을 진행해 무인운반기의 성공적인 최적 경로 제어가 가능함을 입증한다.

Ⅰ. 서 론

무인운반기(Automated Guided Vehicle, AGV)는 대규모 물류창고 무인화의 핵심이며 연구 및 적용이 가속화되고 있다 [1, 2]. 무인운반기로 물류창고의 모든 과정을 무인화 할 수는 없지만 운반 기능을 완벽하게 수행할 수 있어 인력 부족 문제를 해결할 수 있다 [3]. 즉, 사람이 피킹(Picking), 포장(Packing) 및 배송(Shipping) 등의 다른 업무에만 집중할 수 있도록 한다. 특히, 무거운 물건들을 빠르게 운반해야 하는 물류창고에서는 그 필요성이 부각된다.

강화 학습은 미리 준비된 학습 데이터 없이 에이전트가 환경을 탐험하며 학습 데이터를 만들어 학습하는 방식이다. 자세하게, 강화 학습 에이전트는 현재 State 에서 자신의 정책을 기반으로 한 Action 을 취한 후, 환경으로부터 Reward 를 받고 이 세 가지의 State, Action, Reward 정보를 활용해 기대 Reward 를 최대화 하는 방향으로 정책을 업데이트한다. 이 과정을 반복하며 에이전트의 정교한 정책이 수립된다. 한 개의 에이전트만이 존재하는 강화 학습은 단일 에이전트 강화 학습이라 부르며 비교적 어렵지 않게 학습을 진행할 수 있다. 하지만, 여러 개의 에이전트가 존재하는 강화 학습인 다중 에이전트 강화 학습은 단일 에이전트 강화 학습과 다르게 학습에 여러 어려움이 따른다. 대표적인 다중 에이전트 강화 학습 알고리즘으로 QMIX 가 널리 알려져 있으며 여러 변형이 발표되고 있다 [4]. 무인운반기 창고와 같은 환경은 시스템의 전체적인 성능을 위해 모든 에이전트들이 서로 협력하여 문제를 풀어야 하며 QMIX 는 그러한 목적에 적합하다. QMIX 는 각 에이전트의 정책과 더불어 모든 에이전트들의 네트워크 출력 값 및 전역 State 를 입력으로 받는 공동 정책이 존재한다. 이 공동 정책을 모든 에이전트의 정책 합이라 가정하고 학습하며 각 에이전트의 정책을 업데이트 시킨다.

Ⅱ. 본론

본 절에서는 강화 학습 환경의 시나리오와 State, Action, Reward 에 대해 자세히 설명한다. 각 무인운반기는 독립적인 에이전트가 되며 공유되는 하나의 정책을 통해 행동을 결정한다. 해당 정책과 더불어 공동 정책의 기대 보상을 동시에 최대화함으로써 시스템 전체의 성능 향상을 유도한다.

1. 시나리오

물류창고에는 무인운반기, 무인운반기 대기 장소, 선반, 선반 저장소, 피킹 스테이션 및 노동자가 존재한다 ([그림 1] 참조). 본 연구의 초점은 무인운반기의 제어에 맞춰져 있으므로 무인운반기의 제어를 제외한 다른 작업들은 무작위 방식으로 작동하도록 하였다. 즉, 필요한 선반 할당, 해당 선반을 운반할 무인운반기 할당, 피킹 작업이 완료된 선반을 위한 선반 저장소 선택

그리고 임무를 마친 무인운반기가 대기할 무인운반기 장소 선택은 가능한 범위에서 무작위로 결정된다. 나머지 모든 이동 또는 운반 작업들은 QMIX 를 사용한 다중 에이전트 강화 학습으로 제어되며 이동 또는 운반 작업마다 출발지와 목적지가 설정된다. 모든 무인운반기는 배터리가 충분하다고 가정하며 중앙시스템에 동기화되어 일정 시간 단위로 동기화되어 제어된다.

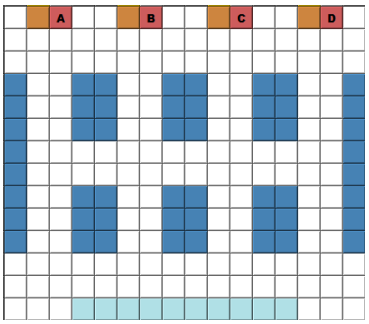


그림 1. 레이아웃

2. State

각 에이전트의 State 는 자신을 중심으로 한 5x5 크기의 지역 정보로 구성된다. 세 개의 채널로 이루어져 3x5x5 의 Shape 으로 표현되며 각 픽셀에는 해당 위치를 표현하는 정보가 들어있다. 해당 정보는 0.0 에서 1.0 사이의 실수로 표현된다. 첫 번째 채널에는 무인운반기를 제외하고 어떤 모듈이 어느 위치에 존재하는지의 정보가 들어있다. 이 정보를 통해 기본적인 길과 장애물의 위치를 판단할 수 있다. 두 번째 채널에는 에이전트인 무인운반기들이 어느 위치에 존재하는지의 정보가 들어있다. 이 정보를 통해 에이전트들 간의 충돌을 방지할 수 있다. 세 번째 채널에는 목적지까지 1.0 에 가까워지는 경사 정보가 들어있다. 이 정보를 통해 나아가야 하는 방향을 알 수 있다. 각 에이전트는 타임 스텝마다 이러한 세 채널의 함축된 정보를 사용하여 자신의 정책을 따르는 Action 을 선택한다.

3. Action

각 에이전트는 타임 스텝마다 자신의 정책을 따르는 행동을 수행한다. Action Space 는 {Stop, Move Forward, Move Right, Move Back, Move Left}이며 이 중 한 개를 선택하여 수행한다. 정지를 제외한 나머지 네 가지 Action 은 현재 에이전트가 바라보는 방향을 기준으로 한다. 이 바라보는 방향은 이전 타임 스텝의 Action 에 의해 결정되며 Stop 의 경우는 변경되지 않는다. 예를 들어, 에이전트가 직전 타임 스텝에 의해 바라보는 방향이 절대적 방향으로 오른쪽인 상태에서 Move Right Action 을 수행할 경우 움직이는 방향은 절대적 방향으로 아래쪽이 된다. 또, 에이전트는 자신의 바라보는 방향을 기준으로 회전된 State 를 사용한다. 즉, 에이전트는 바라보는 방향을 기준으로 State 를 인식하고 정책을 따르는 Action 을 선택한 후 바라보는 방향을 기준으로 Action 을 수행한다.

4. Reward

각 에이전트는 목적지에 도달하면 +1.0 Reward 를 받는다. 반대로 Action 이 장애물 또는 다른 에이전트의 위치나 Action 등으로 인해 충돌이 발생하는 경우 Action 이 Stop 으로 변경되고 -0.3 Reward 를 받는다. 그 외의 Reward 는 모두 0 이다.

이러한 설계를 통해 각 에이전트는 자신에게 임무가 할당되면 최단 타임 스텝 내에 목적지에 도달하는 것이 가장 높은 기대 Reward 를 갖는다. 따라서, 모든 에이전트들은 자신의 목적지에 빠르게 도달하기 위해서 장애물을 피해 가장 빠른 경로로 이동해야 한다. 하지만, 다른 에이전트와의 충돌을 고려하지 않는다면 목적지에 도달하는 타임 스텝이 더 길어질 수 있다. 즉, 타임 스텝마다 장애물과 다른 에이전트들을 고려하여 Action 을 결정해야 한다. 공동 정책은 모든 에이전트들의 정책 출력 값과 전역 State 를 종합하여 모든 에이전트의 기대 Reward 합을 최대화하는 방향으로 업데이트 된다. 그 과정에서 에이전트들의 정책에도 영향을 주어 에이전트가 공동의 목표를 위해 서로 협력하여 행동하도록 유도한다. 즉, 공동 정책은 환경의 기대 보상을 안정적으로 높일 수 있도록 개별 정책을 제어한다.

Ⅲ. 결론

본 연구의 실험에 사용된 레이아웃은 그림 1 과 같다. QMIX 알고리즘을 사용하였고 총 10 개의 에이전트가 존재한다. 에이전트는 녹색, 선반은 파란색, 피킹 스테이션의 입구는 빨간색, 피킹 스테이션의 출구는 황색으로 표현된다.

그림 2 는 학습 그래프이며 x 축, y 축은 각각 training steps, episode rewards 를 나타낸다. 약 8000 training step 이 진행되었으며 약 290 의 Reward 에 수렴하여 다수의 무인운반기를 성공적인 최적 경로로 제어할 수 있음을 확인할 수 있다.

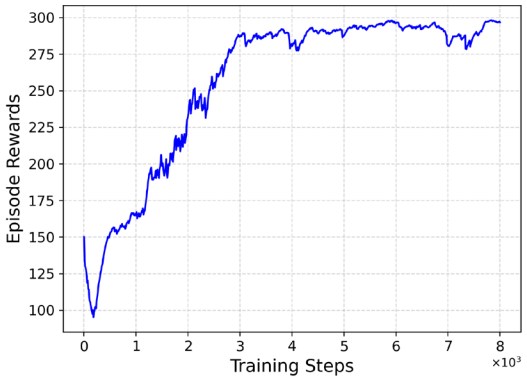


그림 2. 학습 그래프

ACKNOWLEDGMENT

이 논문은 2020 년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No. NRF-2020R1I1A3065610).

참 고 문 헌

[1] Han, Zengliang, et al. "Multi-AGV path planning with double-path constraints by using an improved genetic algorithm." PloS one, 2017.

[2] Kamoshida, Ryota, and Yoriko Kazama. "Acquisition of automated guided vehicle route planning policy using deep reinforcement learning." 2017 6th IEEE International Conference on Advanced Logistics and Transport (ICALT). IEEE, 2017.

[3] Yan, Xuejun, Canrong Zhang, and Mingyao Qi. "Multi-AGVs collision-avoidance and deadlock-control for item-to-human automated warehouse." 2017 International Conference on Industrial Engineering, Management Science and Application (ICIMSA). IEEE, 2017.

[4] Rashid, Tabish, et al. "Qmix: Monotonic value function factorisation for deep multi-agent reinforcement learning." International Conference on Machine Learning. PMLR, 2018.

자율주행차를 위한 실시간 거리예측 기술 연구

한지은, 고석현, 이성진
동서울대학교 전자공학과
haneul2069@du.ac.kr, sungjinlee@du.ac.kr

A Study on the Real-Time Depth Estimation for Autonomous Vehicle

Jieun Han, Sukhyun Ko, Sungjin Lee
Department of Electronic Engineering, Dong Seoul University

요 약

자율주행을 위해서는 자동차가 놓인 상황에 대한 3차원 인지 기술이 필수적이다. 이런 3차원 인지를 위해 주로 사용되는 Lidar 기반 3D Object Detection 기술 혹은 3D Reconstruction 기술은 높은 연산 복잡도로 인해 실시간 동작이 힘들다. 본 논문에서는 실시간 성능이 요구되는 자율주행 환경에서 2D Object Detection 기술에 후처리를 더해 3D Object Detection 에 준하는 결과를 낼 수 있는 2D-DEPTH (2D object detection based DEPTH estimation) 알고리즘을 제안하였다. 실험 결과 제안하는 2D-DEPTH 알고리즘은 1% 미만 의 정확도 손상으로 500% 이상의 가속화를 이룰 수 있었다.

Ⅰ. 서론

최근 딥러닝 기술의 발전으로 자율주행 기술은 운전 보조 기능을 넘어서 완전 자율주행 단계에 이르고 있다. 이런 자율주행의 핵심적인 기술 중에 하나 는 3 차원 주변 상황 인지 기술이다. 즉, 3D Object Detection 과 3D Reconstruction 과 같은 기술로서 이를 정확히 추론해 내기위해 Google, Intel 등의 진영에서는 Lidar를 사용하고 Tesla등의 진영에서는 Camera를 기반으로 해당 연산을 해낸다 [1].

하지만 이런 Lidar 기반 연산은 수많은 Point Cloud 데이터를 획득하여 3D Object Detection을 수행해야 하고 Camera역시 사용해서 Sensor Fusion 단계를 거치기 때문에 실시간 처리에 맞는지 Tesla 진영에서는 의문을 제기한다. 반면 Camera 기반 연산은 여러개의 Camera를 통해 부족한 2D 데이터들로 방대한 3D 데이터를 복원하기 위해 수많은 딥러닝 연산을 수행하므로 이 역시 실시간 처리에 의문이 제기되곤 한다.

본 논문에서는 이런 Lidar와 Multiple Camera에 의존하는 것이 아닌 Mono Camera 만으로 3D Object Detection 전체를 수행하는 것이 아닌 필수 3D 정보 만을 추론해 내는 2D-DEPTH 방식을 제안한다. 실험에 따르면 2D-DEPTH 알고리즘을 사용하여 기존 연산량을 20% 수준으로 감소시킬 수 있고 간단한 도로 상황에서는 정확도 손상이 기존 기술들에 비해 1% 미만이 될 수 있다.

Ⅱ. 2D-DEPTH 모델

본 연구에서는 2D Object Detection을 수행하고 그 결과에 기반하여 자율주행에 직접적으로 필요한 해당 객체의 Depth를 예측한다. 이를 위해 사용되는 2D Object Detection 모델로 RetinaNet [2] 모델을 사용하였다.

1. RetinaNet

RetinaNet 기술은 2D Object Detection 모델 중 1-Stage Detector로서 Focal Loss를 사용하여 Class Imbalance 문제, 즉 Easy Example (IoU가 0에 가까운 샘플, 훈련 용 데이터의 대부분을 차지하지만 훈련에 악영향을 미치는 영역) 에 낮은 가중치를 두고 Hard Example (IoU가 일정 값보다 큰 샘플, 훈련 용 데이터의 작은 부분을 차지하지만 훈련에 유의미한 Gradient를 가지는 부분) 에 높은 가중치를 두어 성능을 개선한 모델이다. 본 논문에서는 이 RetinaNet을 사용하여 관심 클래스로 KITTI dataset [3] 기준 Car, Pedestrian, Bicycle 만을 선정하였다.

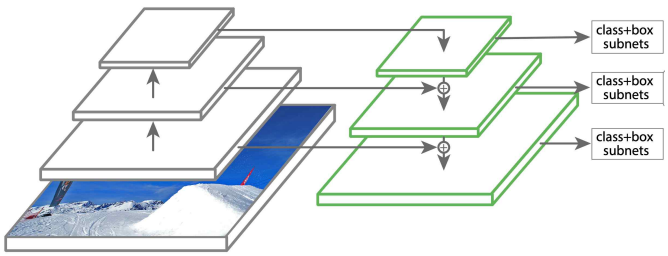


그림 1. RetinaNet 구조 및 동작

2. 2D Object Detection based DEPTH estimation

광학적으로 멀리 있는 물체는 작게 보이고 가까이 있는 물체는 크게 보인다. 이런 단순한 관찰은 Occlusion이 심한 환경에서는 다소 다를 수 있으나 처한 환경이 단순한 환경에서는 꽤 정확할 수 있다. 그래서 본 기술은 처한 환경이 교외지역에 집중해서 Lidar 나 3D Reconstruction 기술에 의존하지 않고 2D Object Detection의 결과 인 Bounding Box Size 만으로 Depth를 예측하여 3D Object Detection을 가속화 하려 한다. 또한 추론하고자 하는 Depth 역시 정확히 몇 m 가 아닌 실 주행에 의미가 있을 수 있도록 다음과 같이 4 단계 깊이로 구분하여 인식한다. 즉,

- **Depth-1**. 1.5m 이내에 있는 아주 가까운 초근접 거리에 위치함. 해당 위치에 관심 객체들이 검출되면 즉시 브레이크를 밟아 충돌을 방지하도록 함
- **Depth-2**. 1.5m 이상 4m 이내 거리에 관심 객체가 위치한 경우. 해당 위치에 관심 객체들이 검출되면 서서히 브레이크를 밟아 속도를 줄이거나 유

지하는 거리

- **Depth-3**. 4m 이상 10m 이내 거리에 관심 객체가 위치한 경우. 해당 위치에 관심 객체들이 검출되면 해당 객체를 주시하며 속도를 유지하는 거리
- **Depth-4**. 10m 이상 거리에 관심 객체가 위치한 경우. 해당 물체 때문에 속도를 줄이거나 유지하지 않으며 자율주행의 원 경로 만을 고려하여 주행

이런 4단계 깊이 구분으로 실제 주행에 도움이 되지 않고 불필요한 영역 (ex. 인도 영역, 하늘 영역) 의 depth 정보를 획득하려 하지 않고, 주행에 도움이 되는 영역 (ex. 내 차가 주행하는 차선 영역, 내 차와 근접한 차선 영역) 에 집중하여 4단계 depth 정보를 획득하는 방식이다. 또한 이런 4단계 깊이 구분을 위한 수치들은 자동차의 속도, 날씨 (비, 눈 등), 브레이크 상태 등 에 따라 달라 질수 있다.

그림 2는 해당 2D-DEPTH 알고리즘의 전체 동작과정을 나타내었다.

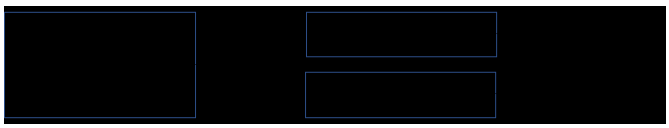


그림 2. 2D-DEPTH 알고리즘의 전체 동작과정

Ⅲ. 실험데이터 셋 구축

기존의 데이터 셋은 교통상황, 도로 상황에 대한 구분이 없어서 새로운 데이터 셋을 구축하였다. 즉, 제안하는 2D-DEPTH 알고리즘은 상대적으로 occlusion 이 적은 상황 즉, 도로교통 상황이 한적한 상황에 더 적합할 것으로 예상이 되고 도로교통 상황이 복잡한 상황에서는 상대적으로 부정확한 결과가 나올 것으로 예상되기 때문에 다양한 상황에 따른 정확도를 검증하기 위해 다음 그림 3과 같은 코스에 따라 그림 4의 기기들을 이용하여 실험 데이터 셋을 수집하였다.



(a) Urban Area (b) Downtown Area (c) Highway Area

그림 3. 도로 유형에 따른 3가지 주행 시나리오



그림 4. 데이터 셋 구축을 위해 사용된 기기들

그림 3에서 보듯이 다음의 3가지 상황에 따라 코스를 결정하였다.

- **UrbanArea**: 교외지역으로 교통이 한적하고 간단한 총 2차선 도로
- **DowntownArea**: 도심지역으로 교통이 복잡한 총 8차선 도로
- **HighwayArea**: 주행 속도가 빠른 고속도로

Ⅳ. 실험

본 논문에서는 2D-DEPTH 알고리즘의 성능 검증을 위해 다음의 Evaluation Set을 정하였다.

- **Lidar3D** : 라이다 기반의 3D Object Detection 기술
- **Cam3D** : 스테레오 카메라 기반의 3D Object Detection 기술
- **2D-DEPTH** : 제안하는 2D Object Detection 기반 Depth 예측 기술

이런 실험 셋에서 다음과 같은 정확도 및 동작 속도 성능을 얻어내었다.

Evaluation Set	주행 시나리오					
	Urban		Downtown		Highway	
	Acc	Lat	Acc	Lat	Acc	Lat
Lidar3D	100	310	100	350	100	330
Cam3D	94	220	90	240	91	230
2D-DEPTH	94	10	88	10	90	10

Table 1. 세 가지 Evaluation Set에 대해 Accuracy 및 Latency 비교

우선 Urban 지역의 성능를 보면 제안하는 2D-DEPTH 알고리즘은 Lidar3D 와 Cam3D 에 비해 정확도 성능은 거의 비슷하지만 (1% 미만) Latency 성능은 20% 수준으로 감소시킬 수 있음을 알 수 있다. 이는 Lidar3D 와 Cam3D 가 불필요한 영역의 3차원 복원에 지나치게 많은 자원을 할당하고 있음을 반증한다고 볼 수 있겠다. 하지만 Downtown, Highway 영역에서는 Latency 성능은 물론 20% 수준으로 감소시킬 수 있지만 정확도 성능이 다소 커서 (5~10%) 이를 위한 보정 혹은 추가 후처리 기술이 필요하다고 볼 수 있겠다.

Ⅳ. 결론

자율주행을 위한 필수 기술인 3D Object Detection의 연산 복잡도를 줄여 실시간 동작이 가능하도록 2D Object Detection 기반의 Depth estimation 알고리즘을 제안하였다. 해당 제안 알고리즘인 2D-DEPTH 알고리즘은 교외지역 Urban Area에서 기존 Lidar 기반 3D Object Detection 이나 Camera 기반 3D Object Detection 에 비해 정확도의 손상은 약 1% 미만으로 줄일 수 있었지만 동작 속도는 약 500% 가속화된 성능을 얻을 수 있었다.

ACKNOWLEDGMENT

이 논문은 2021년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기본연구사업임(No. NRF-2019R1F1A1062878)

참 고 문 헌

[1] Rui Qian, Xin Lai, Xirong Li, "3D Object Detection for Autonomous Driving: A Survey"

[2] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He and Piotr Dollár, "Focal Loss for Dense Object Detection"

[3] http://www.cvlibs.net/datasets/kitti/

인공지능 기반 스마트 수주/발주 처리를 위한 플랫폼 연구

금민경, 오암석*
동명대학교, *동명대학교

mkm6192@naver.com, *asoh@tu.ac.kr

A Study on Platform for Processing of Smart Obtain Order/Order Based on Artificial Intelligence

Min Kyung Keum, Am Suk Oh*
Tongmyong Univ., *Tongmyong Univ.

요 약

본논문에서는 인공지능, 사물 인터넷 등 4 차 산업혁명 기술을 기반으로 생산업체와 공급업체간의 공급 망의 가시성, 안전성, 효율성 향상을 위한 물류 표준을 준수하며 고도화 및 지능화된 스마트 SCM 솔루션에 대해 연구하고자 한다. 이를 위해, 자동 수주/발주 시스템, 자동 재고 관리 시스템, 적정 재고량 산출 시스템 등을 기술하고 각 시스템의 모듈 별 기능 명세와 플랫폼을 제시하였다.

Ⅰ. 서 론

본 논문에서는 인공지능, 사물 인터넷 등 4차 산업혁명 기술을 기반으로 생산업체와 공급업체간의 공급망의 가시성, 안전성, 효율성 향상을 위한 물류 표준을 준수하며 고도화 및 지능화된 스마트 SCM 솔루션을 제시하고자 한다. [1-2]

이와 같은 인공 지능 기반의 스마트한 SCM 솔루션 플랫폼 기술 지원을 위한 구체적인 관련 시스템들은 다음과 같다. [3]

○ 자동 수주/발주 시스템 : 협력업체와의 생산계획 정보, 발주 정보, 납품 정보, 품질판정정보, 재고 현황 등의 제품 데이터를 실시간 공유하는 웹 기반 시스템

○ 자동 재고 관리 시스템 : 사물 인터넷 센서 기반 자동 재고 관리(입고/출고) 시스템

○ 적정 재고량 산출 시스템 : 인공지능을 기반으로 결품에 대한 패턴 연구를 통해 안전 재고 발주 알고리즘

Ⅱ. 시스템 플랫폼

본 논문에서 제안하고자 하는 웹 기반 자동 수주/발주 시스템 구성도이며 고객사로 부터 발주 자료를 받아 협력사에 발주를 하고 업체의 출하부터 본사로 입고, 출하 준비, 검사, 품질관리, 고객 출하 까지를 관리하는 시스템이다. 각 모듈 별 기능 명세는 다음과 같다.

기능	내용	접목 기술
수주	• 고객사(원청사)로부터 납품 오더 수주하는 모듈 • 다양한 오더 인터페이스 방법 제공	① 직접입력(UI) ② 파일 업로드 ③ FTP interface ④ API:온라인 연계 ⑤ DB Link
발주	• 고객사로부터 받은 오더(발주번호)에 대해 협력사에서 공급해야 할 부품에 대해 발주 • 발주상태 모니터링(웹/모바일)	• 인공지능 모듈 활용 -다양한 변수를 고려한 오더 발주 -안전 재고, 안전 재고일, 납기,리드타임 등 -다양한 변수의 인공 지능적 산출

출하	• 협력사에서 납품을 위한 라벨 발행, 납품서발행 기능을 제공하여 납품 진척을 파악 가능	② 바코드 라벨 발행 ② IOT 장비-발주 번호 연계 등록
납입	• 협력사에서 납품한 생산 부품에 대해 입고 처리 • 재고조사(모바일 스캔)	① 라벨 스캔 ② IOT 자동 스캔
품질	• 입고 품에 대한 품질 검사 • 납기, 품질 등에 따른 업체 등급 설정	• 주기적으로 납기, 품질 등에 따른 오더 발주 변수 산출
출하	• 고객사에 부품 납품을 위한 출하 대기 등록	
기본	• 부품, 업체, 장비 등 기본정보 관리 • 업체 생산 물류 일정관리(달력)	• IOT 장비 등록 • AI를 위한 부품/협력사 등록

본 논문에서 제시하는 플랫폼은 자동 스마트 수주/발주 처리를 할 수 있는 플랫폼이며 사용자에게 어떤 OS나 어떤 웹 브라우저 에서든 수주/발주 정보를 제공 받을 수 있도록 구성하며 개발방법론에 따라 분석/설계되며 개발 플랫폼에서 통합 테스트, 사용자 수용 테스트(UAT)를 통해 운영 플랫폼에 적용된다. 각 아키텍처 별 상세 내용은 다음과 같다.

플랫폼	내용
운영 플랫폼	• 실제 운영 시스템이 탑재된 시스템 • 표준 웹(HTML5)기반 시스템 • JAVA, SRPRING, SqlServer(+ Oracle)
개발 플랫폼	• 개발 및 테스트를 위한 시스템 • 운영 환경과 동일한 구성 • 통합 테스트, 사용자 수용 테스트 환경 제공 • 분석/설계 환경
방법론	• 주관사의 개발방법론 • 개발 플랫폼에서 충분한 테스트를 거쳐 운영플랫폼에 적용 • 운영 적용 절차 방법론에 따른 적용

Ⅲ. 결론

본 논문에서는 인공지능, 사물 인터넷 등 4 차 산업혁명 기술을 기반으로 생산업체와 공급업체간의 공급 망의 가시성, 안전성, 효율성 향상을 위한 물류 표준을 준수하며 고도화 및 지능화된 스마트 SCM 솔루션을 제시하였다. 아울러, 자동 수주/발수 시스템, 자동 재고 관리 시스템, 적정 재고량 산출 시스템 등을 기술하고 각 시스템의 모듈 별 기능 명세와 플랫폼 그리고 각 시스템 모듈 별 기능 명세와 아키텍처 별 상세 내용을 제시하였다. .

.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기회평가원의 SW중심대학지원사업의 연구결과로 수행되었음 (1711102971).

참 고 문 헌

[1] KSA, "Global trend for smart factory and countermeasure strategy for standardization in Korea," KSA Policy Study 012 Issue Paper 2015-3, Jul. 2015.

[2] J. Davis, "NIST workshop on open cloud architectures for smart manufacturing," UCLA & SMLC (www.smartmanufacturingcoalition.org), May 2015.

[3] Joint Ministries, Policy to implement manufacturing innovation 3.0 strategy, 7th meeting to promote trade and investment in Korea, Mar. 2015.

딥러닝 기반 온습도 예측 모델을 통한 화재 조기 감지 시스템 구축

박치원, 홍기대, 허의남*
경희대학교
bamtorrr@khu.ac.kr, kidae92@khu.ac.kr, johnhuh@khu.ac.kr

Establishment of early fire detection system through deep learning-based temperature and humidity prediction model

Chi-won Park, Gi-Dae Hong, Eui-Nam Huh*
Kyung Hee Univ.

요 약

화재 조기 감지가 늦거나 감지 시스템의 잦은 오류로 생기는 안전 불감증으로 인해 화재의 골든타임을 놓치는 경우가 존재한다. 본 논문에서는 빠르고 정확한 화재 예측을 위해 슬라이딩 윈도우 프로세스 기반 온습도 예측 LSTM 모델과 화재 감지 시스템을 제안한다.

Ⅰ. 서 론

21년 6월 17일 대형 물류센터에서 화재 발생에 대한 대처가 늦어 큰 피해로 이어졌다. 화재 감지 시스템의 잦은 오류로 인한 안전 불감증이 원인이었다. 또한 화재는 발화되어 단 시간에 큰 피해를 일으키기 때문에 조기 감지가 중요하다. 따라서 화재로 인한 피해를 최소화하기 위해서 빠르고 정확한 예측이 요구된다.

본 논문에서는 발화요인 중 온도와 습도의 추이를 통해 그 값을 예측하는 화재 조기 감지 모델을 제안하며, 데이터 분석의 실시간성과 정확도 보장을 위해 이산적으로 데이터를 나누어 학습하는 방식이 아닌 연속적으로 이동하며 데이터를 학습하는 슬라이딩 윈도우 기법을 적용하고 평가한다. 이를 통해서 빠른 조기감지와 정확한 예측으로 골든타임을 놓치지 않고 오작동으로 인한 안전 불감증에 빠지지 않도록 한다. 또한 제안 모델에 따른 실시간 화재 감지 시스템을 구축한다.

Ⅱ. 본론



그림 1 데이터 수집 및 전송 저장

본 논문에서는 LSTM 예측 모델의 입력으로 사용하기 위한 데이터를 IoT 센서로부터 측정하여 RF 통신을 통해 라즈베리 파이 내 데이터베이스에 스트리밍 데이터를 수집한다. 실시간으로 축적되는 데이터 처리를 위해 IoT 기기에서 측정한 값을 RF 무선 모델인 ZIGBEE 모듈을 통하여 라즈베리 파이로 전송하고 전송된 데이터는 데이터베이스에 저장되어 사용된다.

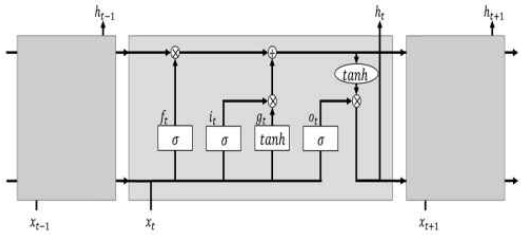


그림 2 LSTM 개념 구조

기존의 RNN 모델은 vanishing gradient라는 문제가 있어 학습하는데 한계가 있기에 이를 해결한 LSTM 모델을 이용하여 학습을 진행하였고 구현은 keras를 이용하여 모델의 신경망을 구축하였다. LSTM은 그림 2와 같이 기본적으로 Elman의 RNN과 같은 구조로 이루어져 있으나 각 state에 게이트(gate)들을 추가하여 전달할 정보들을 선택하고 셀 스테이트(cell-state)를 통해 결과 값을 갱신하여 전달한다.[1]

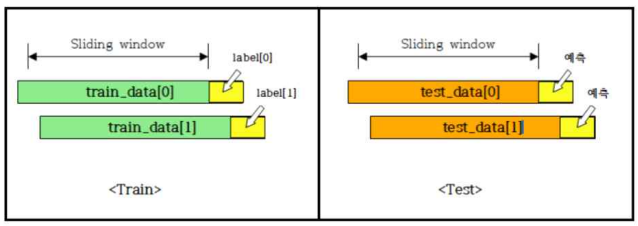


그림 3 Train, Test 데이터의 Sliding Window 구조

지정된 크기의 데이터 셋인 슬라이딩 윈도우는 그림3과 같이 일정한 간격으로 조금씩 이동하며 데이터를 분류한다. 분류된 데이터를 학습한 후 바로 다음에 위치한 데이터를 라벨로 지정하여 예측해보며 모델을 학습시킨다. 슬라이딩 윈도우가 일정한 간격으로 이동하며 겹치게 되는 구간

과 조금씩 변화되는 구간에 대해 학습한하는 방식은 기존의 이산적으로 데이터를 나누어 학습하는 모델보다 갑작스러운 변화에 더 빠르게 정확하게 예측할 수 있도록 해준다. 이는 단기간에 발화되어 빠르게 확산하는 화재를 예방하기에 더욱 적합하다.

예측 모델의 개발은 데이터 수집, 데이터 정규화, 데이터 정제, 모델학습 및 성능 검증, 예측의 단계를 거친다. 데이터는 IoT센서에서 수집하여 RF 모듈을 통해 라즈베리 파이에 전송하고 데이터 베이스에 저장한다. 데이터를 그대로 학습할 수 없기에 모든 데이터를 0~1 사이 값으로 변형해주는 MinMaxScaler 방법을 이용하여 데이터를 정규화한다. 또한 단기간 내 변화가 큰 경우에도 빠르고 정확하게 예측할 수 있도록 sliding window 기법을 모델에 적용시킨다. 모델 층을 구성하여 모델학습을 진행한다. 테스트 데이터를 이용하여 성능을 검증하며 그래프로 출력해본다.

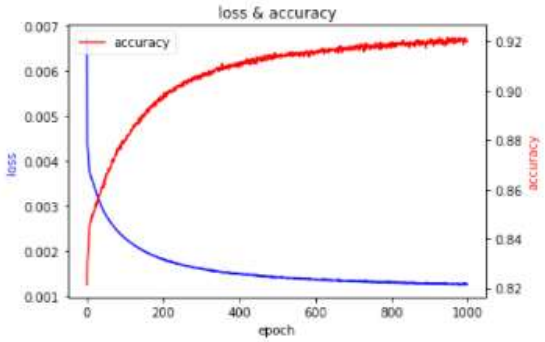


그림 4 학습 모델 오차 감소율과 정확도 증가율

데이터 셋의 테스트 정확도는 92.01가 나왔다. optimizer는 가장 높은 정확도를 보여주는 adam으로 설정하였고 batch 길이는 100, epoch는 1000으로 설정하고 실험을 진행하였다. 126141개의 데이터만으로 진행하였기 때문에 학습시간은 20분 정도 걸렸다. look_back의 길이가 34일 때 가장 높은 정확도가 나왔다.

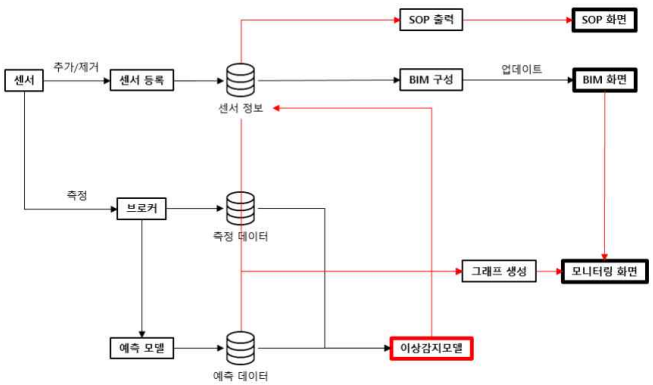


그림 5 화재 감지 시스템 구성도

실시간 화재 감지 시스템은 그림5과 같다. 센서를 추가하면 센서 정보 DB에 센서 고유 번호와 공간 정보 저장하고 연관된 센서 리스트에 업데이트된다. 센서를 통해 측정된 데이터를 측정 데이터 DB에 저장하고 예측 모델에 전달한다. 전달된 데이터를 통해 예측한 값을 예측 데이터 DB에 저장한다. 이상 감지 모델을 통해 측정 데이터와 예상 데이터를 분석하여 화재를 예측한다. 화재 예상 시 센서 정보 DB를 통해 연관된 센서 목록을 불러오고 해당되는 센서들에 대한 측정, 예측 데이터로 그래프를 생성하여 모니터링 화면에 출력해준다. 또한 해당 센서들의 공간 정보에 맞는 SOP를 불러와 SOP 화면에 출력해 화재 대처 방법을 제시한다.

Ⅲ. 결론

본 논문에서는 발화 후 단기간에 빠르게 확산되는 화재를 빠르고 정확하게 예측해 골든타임 내에 안전 불감증에 빠지지 않고 대처할 수 있도록 하는 딥러닝 기반의 예측 모델과 시스템 구조를 제안한다. 예측 모델이 제시하는 특징을 위해 이산적으로 데이터를 분류하던 기존 방식과 다르게 연속적으로 분류되어 학습할 수 있는 슬라이딩 윈도우 기법을 적용시켜 개발하고 성능을 제시한다.

IoT센서를 통해 수집되어 데이터베이스에 저장된 데이터로 학습 및 테스트를 진행하였고 그 결과 92.01% 사이의 결과 값을 얻을 수 있었다. LSTM 모델의 수와 look_back의 길이 및 batch를 바꿔가는 등 다양하게 코드를 수정하고 그 결과 값을 비교해가며 성능을 개선이 필요하다.

또한 슬라이딩 기법을 적용시킨 모델과 기존 모델의 성능 차이와 좀 더 다양한 데이터들로 비교 분석하여 제시한 모델의 취지에 맞도록 보완하고 이를 적용시킨 화재 감지 시스템을 구축할 예정이다.

ACKNOWLEDGMENT

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2021-0-00818, 스마트 플랜트 이상상태 조기감지를 위한 머신러닝 기반 저대역 영상/통신 Edge Computing 시스템)

참 고 문 헌

[1] 임장혁, “Detecting Disaster Information Sentences from Unstructured Data Using Recurrent Neural Network”, Seoul National University pp. 26.

화재 빅데이터 분석을 통한 화재 대응 위험도 모델링에 대한 연구

권은정, 변성원, 박현호, 손교훈
한국전자통신연구원

ejkwon@etri.re.kr, swbyon@etri.re.kr, hyunhopark@etri.re.kr, sonkh@etri.re.kr,

A Study on Fire Response Risk Modeling through Fire Big Data Analysis

Eun-Jung Kwon, Sungwon Byon, Hyunho Park, KyoHoon Son
Electronic and Telecommunications Research Institute (ETRI)

요 약

본 논문은 재난안전과 관련하여 국가적으로 활용 가치가 높은 화재 빅데이터에 대한 화재종류, 피해액, 연령별 환자증상, 환자발생지역 등과 같은 데이터의 특징을 탐색적으로 분석한다. 분석된 결과를 통해 인명피해와 재산 피해액에 따라 화재 발생 환경, 화제의 종류, 동원된 인력 및 자원의 규모를 산정함으로써 화재 대응에 대한 개선점을 도출하고자 한다. 이를 통해, 화재활동 조사서로 기록되어 생성된 화재 빅데이터로부터 화재 상황 별 위험등급을 산출하기 위한 방안을 모색한다.

Ⅰ. 서 론

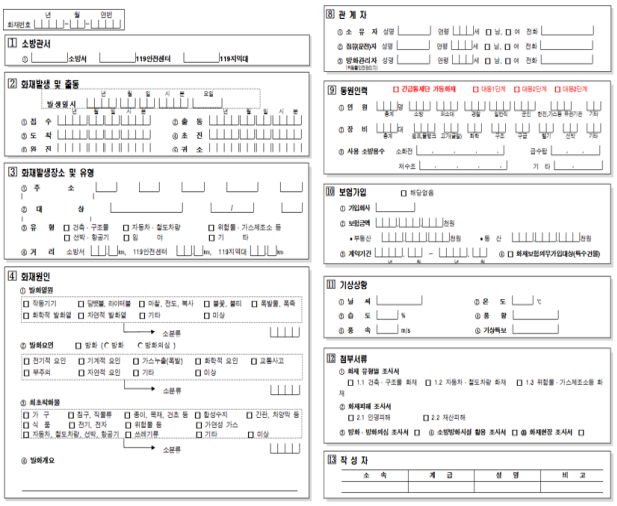
119 종합상황실에는 119 신고접수부터 재난대응 및 종결까지 전국 18 개 시도에서 표준화된 방식으로 긴급구조표준시스템를 운영되고 있다. 또한, 화재의 경우 국가화재정보시스템을 통해 화재 원인, 발화원인, 발생장소 등 화재 발생에서부터 과정에 대한 세부적인 조사내용을 기록, 관리하는 화재조사서와 화재기간, 장소, 피해액, 인명피해, 출동소요시간 등 수치적인 정보에 대한 화재통계정보에 대한 이용이 가능하다. 본 논문은 향후 화재 피해규모에 따라 화재대응 시 위험도를 산정하기 위해 화재신고로 접수되어 기록 보관되고 있는 화재현장 조사서에 기록된 데이터를 화재유형, 출동시간 및 안전센터와의 거리에 대한 관계를 고려하여 탐색적으로 분석한 내용을 정리한 것이다.

Ⅱ. 관련연구

2019 년 11,567,173 건의 119 신고로 출동관련 신고가 접수되었다[1]. 이중 화재는 40,103 건 발생하여 2,515 명의 인명피해를 발생하였다. 지역별 화재발생현황은 경기가 9,421 건으로 가장 많고 그 다음 서울이 5,881 건의 발생현황을 보이고 있다[1]. 이와 같이 국민의 생명에 직접적으로 영향을 미치는 화재를 포함한 구급, 구조와 같은 각종 위험 상황에 ICT 기술을 통해 대처하기 위해 소방안전 빅데이터를 이용한 대국민 안전을 위한 예방 대책이 개발되고 있다[2]. 재난상황을 가장 먼저 접하게 되는 119 상황실에서는 화재신고의 접수 시, 초동화재 진압을 위해 정확한 재난사고 발생지점, 재난사고 종류, 주변 중요 건물, 현장 진입을 위한 고속도로 및 터널의 진행방향 등과 같이 신속히 정확한 정보의 파악이 중요하다. 그 밖에 화재의 불꽃 유무, 내부에 사람이 보이는지 여부, 사람들의 대피여부, 출입문이 잠겨 있는 지 여부, 연소가 확대되는 지 여부, 소방차 진입에 어려움이 있는 지와 같은 '인명구조 상황 및 연소상황 파악' 이 요구된다. 신고접수와 동시에 재난유형, 재난위치, 신고자 전화번호 등을 확인하고 출동지령을 동시에 수행하게 된다. 이러한 화재 신고 접수 처리 과정에서 화재통계 빅데이터에서 인명, 재산 피해를 조합하여 화재위험도를 정의하고 예측하는 화재현장 위험도 예측 모델이 제시된 바 있다[3]. 소방기본법에 따라 화재조사결과서는 화재현장을 기록하여 화재예방지도나 소방행정 기초자료로 소방업무 전반에 또한 활용되고 있다[4].

Ⅲ. 화재 빅데이터 탐색적 분석

본 장에서는 화재현장조사로 기록되는 데이터 항목을 이해하고, 2020 년 대전시에서 발생한 화재신고에 대한 유형별 발생 건수와 화재발생지의 지리적 특징에 따른 화재대응에서의 의미를 파악 하고자 하였다. 이런 과정을 통해 화재유형별 피해규모의 결과값에 따라 앞으로 화재신고로 접수되는 사건을 과거 유사사건과 비교하여 대응정보에 활용할 수 있는 위험등급을 산정하고, 이를 활용하기 위한 방안을 제공하고자 한다.



(그림 1) 화재현장조사서

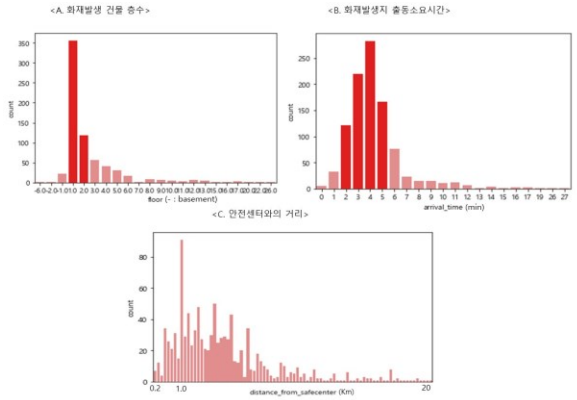
그림 1 화재조사서 양식을 통해 생성된 정형 및 비정형 형태로 생성되는 화재조사결과 데이터를 살펴보면 2020 년 약 800 건의 화재조사결과 데이터를 7 가지 화재유형으로 구분해 볼 수 있다. 유형별 분류된 화재발생 건수는 고층건물 223 건, 일반화재-주택 193 건, 일반화재-차량 108 건, 대형화재, 산불 49 건, 지하화재 2 건, 기타화재 415 건의 발생분포를 갖는다.



	화재구분	발생건수
0	기타화재	415
1	고층건물(3층이상,아파트)	223
2	일반화재(주택)	193
3	일반화재(차량)	108
4	대형화재(시장,공장)	49
5	산불	11
6	지하화재	2

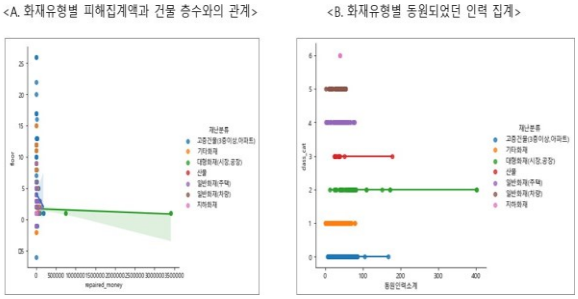
(그림 2) 화재유형별 발생건수

화재신고접수 시 과거발생 건수와 비교하여 위험등급을 정의하기 위해 문제정의를 수립하였다. 첫째, “화재발생 건물의 층수, 화재발생지 도착에 걸리는 소요되는 시간, 그리고 화재발생지와 안전센터와의 거리가 재산피해액과 관련성이 있다면 그 값은 비례하는가?”에 대한 데이터 탐색을 수행하였다.



(그림 3) 화재발생지 영향변수에 대한 특징 탐색

그림 3 과 같이 화재발생지의 발화건물의 대다수는 건물 1 층에서 발생했으며 고층건물인 26 층까지도 발생빈도가 존재함을 알 수 있다. 화재대응 골든타임과 관련하여 출동이후 화재발생지까지 소요시간은 평균 4.3 분으로 7 분이상 소요된 횟수로 약 50 건으로 확인된다. 또한 거리에 대한 영향변수로 소방차가 위치한 안전센터와의 는 평균 2.4km 에 존재하며 최고 20km 로 확인된다. 이러한 환경적 특성에 따라 화재유형별 피해액을 살펴본 결과 그림 4 와 같이 확인되었다.



(그림 4) 화재유형별 피해정도에 대한 데이터 탐색

화재유형별 피해집계액 및 동원인력에 대한 결과는 그림 4 에서와 같이 대형화재의 경우 평균 피해액 범위와 상당한 차이를 보이고 있으며, 다른 유형의 화재는 대부분 피해액이 비슷한 것으로 보인다. 대형화재가 저층에서 발생한 것을 알 수 있지만 고층 건물(3 층이상 아파트)물 화재의 경우도 피해액이 다른 유형에 비해 큰 것을 알 수 있다. 소방력으로 정의되는 동원인력의 소계에 대한 분포는 대형화재가 피해액과 비례하여 가장 많고, 산불 및 고층건물 화재가 그 뒤를 잇고 있다.

상기와 같은 결과를 통해 화재에 대한 위험은 ‘1 층 화재’의 경우 발생빈도는 낮지만 화재피해액의 규모가 크기 때문 ‘발화건물의 층(Floor)’와 일반화재(차량), 일반화재(주택) 및 기타화재의 경우 상대적으로 동원인력이 적은편이나 ‘화재 발생지로부터 안전센터와의 거리’는 멀수록 피해규모가 커지기 때문에 위험등급 정의에 고려되어야 할 것으로 보인다. 향후 각 화재유형별 발화개요 및 신고접수 시 접수자가 기록하는 신고내용의 주요 단어에 대한 특징을 살펴보고 화재피해액과의 관련성을 분석하여 위험등급에 적용 가능한 대표 단어를 추출하여 추가 연구를 진행하고자 한다.

IV. 결 론

본 논문에서는 화재조사서의 정형, 비정형 데이터를 통해 피해액 및 동원인력을 기준으로 화재유형 별 특징 분포를 파악하였다. 향후 발화개요 및 신고접수 시 신고내용에 대한 전체 문장의 비정형 데이터로부터 세부적인 키워드를 추출하여 위험상황을 세부적인 의미를 분석할 계획이다. 이를 통해 화재상황에 대한 위험도 모델링에 대한 대응 모델을 구현하고 성능을 평가함으로써 시스템 적용 가능성을 확인하고자 한다.

ACKNOWLEDGMENT

이 논문은 대한민국 정부(과학기술정보통신부)의 재원으로 한국연구재단 재난안전플랫폼기술개발사업의 지원을 받아 수행된 연구임 (과제번호 : NRF-2020M3D7A1103494

참 고 문 헌

[1] 2020 소방청 통계연보, 소방청

[2] 최갑용, " 소방안전 빅데이터 플랫폼 - 구현방안 및 발전방향, "정보공학회지, 2021, 48(1), pp. 37-46

[3] Do Hyoung Kim, Byung wan Jo, " A Study on the Development of a Fire Site Risk Prediction Model based on Initial Information using Big Data Analysis, " Journal of the Society of Disaster Information, Vol. 17, No. 2, pp. 245-253, June 2021

[4] Nikhila Raveendran, “Smart Firefighting Prediciton System Based on Artificial Intelligence and Big data,” 2020 Dec, DOI:10.13140/RG.2.2.18551.44963

K-Means 군집화 알고리즘을 이용한 주기 변동 랜덤 신호의 정렬 및 이벤트 추출

변성훈
선박해양플랜트연구소

byunsh@kriso.re.kr

Alignment and event extraction of period-varying random signal using K-Means clustering algorithm

Sung-Hoon Byun
Korea Research Institute of Ships and Ocean Engineering (KRISO)

요 약

본 연구는 주기적으로 반복되지만 시간에 따라 주기가 임의적으로 변화하는 주기 변동 신호에 대해 K-Means 알고리즘을 사용해 군집 신호를 추출하고, 추출된 군집 신호를 이용해 실제 주기를 추정하는 신호 정렬 및 추출 알고리즘에 대한 연구 결과를 소개한다. 대상 신호는 왕복 또는 회전 사이클 운동에 의해 주기성을 가지나, 기계적인 특성으로 사이클 길이(주기)가 변화하는 기계 진동 신호가 대표적이다. 제안된 알고리즘은 신호가 일정한 임계값 이하로 작아지는 경우 신호의 값을 0 으로 치환하게 되면, 남은 신호가 군집화되는 원리에 착안하여 K-Means 알고리즘을 이용해 군집화를 진행하였다. 이 때, 각 군집의 파형은 각 사이클마다 랜덤하지만 평균 전력은 유사하게 유지되는 점을 이용하여 기준이 되는 군집을 선정하고, 기준 군집 사이의 간격을 측정하여 주기를 추정하였다. 제안된 알고리즘을 실제 선박 발전기의 진동 신호에 적용하여 사이클 길이를 추정하였으며, 이를 발전기 사양의 표시 회전수에 의한 사이클 길이와 비교하여 알고리즘 수행 결과의 유효성을 확인하였다.

Ⅰ. 서 론

최근 다양한 기계 시스템의 진동 신호로부터 시스템 건전성을 모니터링하는 기술에 대한 연구가 활발하다. 건전성 모니터링을 위해 측정되는 진동 신호들은 많은 경우 주기적인 실린더 왕복 또는 프로펠러 회전으로 인하여 주기성을 가지지만, 기계의 작동 특성으로 주기가 일정하게 유지되지 않고 시간에 따라 조금씩 임의적으로 변화하는 특성을 보여준다. 따라서 타코미터를 이용해 왕복(회전)에 동기를 맞추고 신호를 수집한 경우가 아니면, 수집된 신호의 주기를 정확하게 파악하기 어렵고 다양한 이벤트 신호들의 시간에 따른 변화를 확인하는 데에 어려움이 따른다. 본 연구는 이와 같이 주기적으로 반복되지만 시간에 따라 주기가 임의적으로 조금씩 변화하는 주기 변동 신호의 주기를 추정하고, 한 주기 내의 이벤트 신호를 군집화하여 추출하는 방법에 대한 연구 결과를 소개하고자 한다.

Ⅱ. 문제 설명

본 연구에서 대상으로 하는 주기 변동 신호는 다음과 같은 특징을 가진다.

- 신호의 측정 시작 시간은 임의적이며, 측정 주기가 왕복(회전) 운동에 동기화되지 않음
- 신호는 주기적인 사이클이 반복되나, 각 사이클 길이, 즉 주기는 변화함
- 한 사이클은 기관의 행정 단계 특성을 반영하여 파악 가능한 개수의 군집 신호로 구성됨
- 각 군집 신호의 파형은 임의적이나, 각 군집 신호의 평균 전력은 유사하게 유지됨

Fig. 1은 선박에서 사용되는 발전기의 진동 신호[1]를 표시 회전수(nominal RPM) 정보를 이용해 사이클 단위로 분리하여 나타낸 것이다. 표시 회전수에 의한 사이클 길이는 실제 사이클 길이와 차이가 존재하므로, 각 신호 군집의 위치가 일정하지 않고 변화하는 것을 확인할 수 있다. 본 연구의 목적은 위와 같은 특징을 가지는 시계열 신호를 같은 기준 위치를 가지는 사이클 단위로 분리하고, 각 사이클의 주요 이벤트 신호를 추출하는 방법을 고안하는 것이다.

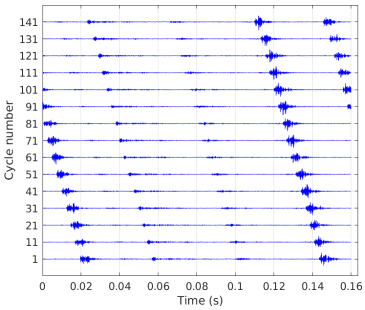


Fig. 1 Example of period-varying signal aligned using the nominal cycle length.

Ⅲ. K-Means 알고리즘을 이용한 신호 정렬 및 이벤트 추출

본 연구에서는 신호가 일정한 임계값 이하로 작아지는 경우 신호의 값을 0 으로 치환하게 되면, 남은 신호가 군집화되는 원리에 착안하여 K-Means 알고리즘[2]을 이용해 군집화를 진행하였다. 이 때, 각 군집의 파형은 시간에 따라 달라지지만 평균 전력은 유사하게 유지되는 점을 이용하여 기준이 되는 군집을 선정하고, 기준 군집 사이의 간격을 측정하여 한 사이클의 길이, 즉, 주기로 판단하였다.

Fig. 2는 본 연구에서 제안한 알고리즘을 정리하여 나타낸 것이다. 처음 초기화 단계 (Initialization)에서는 군집 중심 위치 정보가 없으므로 다수의 초기값을 입력하여 결과를 계산하고, 그 중에서 가장 작은 점-중심 거리합을 가지는 결과를 최종 결과로 사용한다. 이 때, 각 초기값은 K-Means++ 방법을 사용하여 생성한다[3]. 이후 반복 연산 단계(Iteration)에서는 이전 사이클 처리 과정에서 측정된 중심을 초기값으로 입력하여 K-Means 알고리즘을 수행하도록 하였다.

Input: signal sequence $\underline{x} = [x_0, x_1, ..., x_{L-1}]$ **, threshold** v **, number of clusters** N
Output: cycle starting index i_0 **, estimated cycle lengths** $\underline{M}$ **, index sets of clusters** $\underline{I}$

Initialization

1. Set initial starting index $i_0(=0)$ and initial cycle length M_0 (= nominal cycle length)
2. Set input signal sequence for clustering, $\underline{x} = [x_{i_0}, x_{i_0+1}, ..., x_{i_0+M_0-1}]$
3. Thresholding $\underline{x}$: make x_i zero for $i \in I_{<thd}$ *where* $x(i) < v$ and set $I_{>thd} = I_{<thd}{}^C$.
4. Cluster $I_{>thd}$ using K-Means algorithm with random initial centroid: get N clusters of index $\underline{I_0} = \{I_0{}^1, I_0{}^2, ..., I_0{}^N\}$ and centroid locations $\underline{c_1}$.
5. Select target cluster $I_0{}^t$ and update i_0 under the criterion:
   $i_0 = argmin_{i \in I^t} |x_i| \geq 0.5 * maximum\ of\ target\ cluster\ signal\ level$
6. Update the initial centroid $\underline{c_0} = \underline{c_1}$.

Iteration

7. for (k=1:MAX_ITER) do
   a. If $i_0 + M_0 > L$ do
      K=k-1
      break
      end if
   b. Set input signal sequence for clustering, $\underline{x} = [x_{i_0}, x_{i_0+1}, ..., x_{i_0+M_0-1}]$
   c. Thresholding $\underline{x}$: (same as 1-c)
   d. Cluster $I_{>thd}$ using K-Means algorithm with the initial centroid $\underline{c_0}$: get N clusters of index $\underline{I_k} = \{I_k{}^1, I_k{}^2, ..., I_k{}^N\}$ and centroid locations $\underline{c_1}$.
   e. Select the target cluster I^t and set i_1 under the criterion:
      $i_1 = argmin_{i \in I_k{}^t} |x_i| \geq 0.5 * maximum\ of\ target\ cluster\ signal\ level$
   f. Update the cycle length $M_0 = M_0 + i_1 - i_0$ and save $M_k = M_0$.
   g. Update the initial centroid $\underline{c_0} = \underline{c_1}$ and the starting index $i_0 = i_1$.
   h. k = k+1
8. end for
9. return i_0, $\underline{M} = [M_0, M_1, ..., M_K]$, $\underline{I} = [\underline{I_0}, \underline{I_1}, ..., \underline{I_K}]$

Fig. 2 Algorithm for cycle length estimation

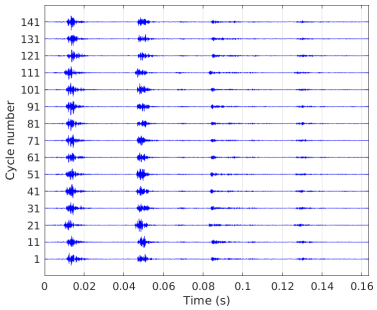


Fig. 3 Signal (in Fig. 1) aligned using the estimated cycle lengths.

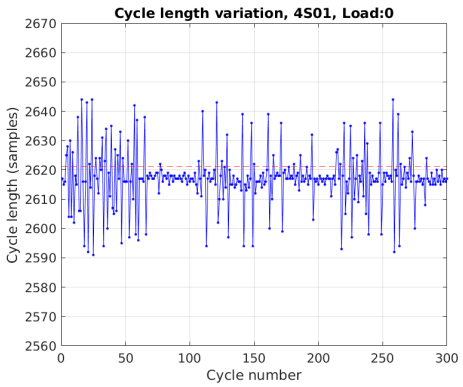


Fig. 4 Estimated cycle length (solid line with dots) and nominal cycle length (dashed) in samples.

Fig. 3은 제안한 알고리즘을 이용하여 Fig. 1의 신호를 사이클 길이 단위로 정렬한 결과이다. 유사한 전력 크기를 가지는 군집 신호들이 같은 시간 위치를 가지도록 정렬된 것을 확인할 수 있다. Fig. 4는 알고리즘으로부터 추정된 사이클 길이를 표시 회전수와 비교한 것으로, 시간에 따라 사이클 길이가 변화하며 평균적인 사이클 길이는 표시 회전수에 의한 사이클 길이 값과 유사하지만 약간 작게 나타났다.

Ⅳ. 결 론

본 연구는 주기적으로 반복되지만 시간에 따라 주기가 임의적으로 조금씩 변화하는 주기 변동 신호에 대해 K-Means 알고리즘을 사용해 사이클 내의 군집 신호를 추출하고, 추출된 군집 신호를 이용하여 실제 주기를 추정하는 알고리즘을 개발하였다. 제안된 알고리즘을 실제 선박 발전기의 진동 신호에 적용하여 사이클 길이를 추정하였으며, 이를 발전기 사양의 표시 회전수에 의한 사이클 길이와 비교하여 유사하게 사이클 길이가 추정되는 것을 확인하였다.

ACKNOWLEDGMENT

본 논문은 산업통상자원부 국가연구개발사업 "자율운항선박 핵심 기관시스템 성능 모니터링 및 고장예측 진단 기술 개발"으로 수행된 연구결과입니다.

참 고 문 헌

[1] 변성훈 외, "자율운항선박 기관시스템 상태 진단을 위한 진동 신호 분석 기법 고찰", 대한조선학회 2021 년 춘계학술대회 논문집, 2021.

[2] 오일석, 패턴 인식, 교보문고, 2008, pp.349-354.

[3] Arthur, David, and Sergei Vassilvitskii. k-means++: The advantages of careful seeding. Stanford, 2006.

LDAPS 예측 및 AWS 관측 데이터를 이용한 CNN-BLSTM 기반의 단기 기온 예측

전지민, 김홍국
광주과학기술원
jiminbot20@gm.gist.ac.kr, hongkook@gist.ac.kr

Short-term Temperature Prediction Based on CNN-BLSTM Using LDAPS-Predicted and AWS-Observed Data

Jimin Jun, Hong Kook Kim
Gwangju Institute of Science and Technology

요 약

본 논문에서는 단기 기온 예측 성능을 향상하기 위해 AWS (Automatic Weather Station) 관측 데이터와 LDAPS (Local Data Assimilation and Prediction System) 수치예보모델의 고해상도 데이터를 사용한 CNN-BLSTM (Convolutional Neural Network - Bi-directional Long Short-Term Memory) 모델 기반 모델의 성능을 평가한다. 특히, LDAPS 이미지의 크기에 따른 기온 예측 오차의 성능을 평가한다. 그 결과, 수치예보모델의 이미지 크기가 20x20일 때, 1.78 RMSE (Root-Mean-Squared Error), 1.39 MAE (Mean Absolute Error)로 가장 좋은 성능을 보였다.

Ⅰ. 서 론

최근 전 지구적 극단적인 기후 변화의 영향으로 국내외의 직간접적인 경제적 피해와 인명 피해가 증가하고 있다. 정밀한 기후 예측은 이러한 피해를 줄이고 예방하는데 필요한 가장 중요한 요소 중 하나이다. 기후변화의 가속화로 급변한 기상환경에 대한 정밀한 기후 예측의 필요성이 더욱 높아지고 있다[1]. 특히 극단적인 이상 기상 현상으로 급변하는 기상환경에 의해 증가하는 폭염, 국지성 집중호우, 태풍 등을 대비하기 위한 정확한 단기 기후예측은 매우 중요하다.

대표적인 기후 예측 방법으로는 수치예보모델이 있다. 수치예보모델은 해상도, 모델 지형, 동화 기법, 모수화 등에서 여러 가지 한계점을 가지고 있다[2]. 이를 보완하기 위해 딥러닝으로 접근하는 여러 연구가 진행되고 있다[3-5]. 특히 [3]에서는 관측 데이터와 수치예보모델 데이터를 각각의 신경망으로 특징을 추출하고, 각각의 특징을 주의집중 기법을 적용하여 기온을 예측하는 방법을 제안하였다. 여기서는, 중장기 기온 예측을 위해 수치예보모델로 RDAPS (Regional Data Assimilation and Prediction System) 예측 데이터를 사용하였다.

본 논문에서는 [3]의 신경망 모델을 활용하여 단기 기온 예측 성능을 높이고자 한다. 이를 위해 RDAPS 보다 해상도 높은 LDAPS (Local Data Assimilation and Prediction System) 예측 데이터를 활용한다. 특히, LDAPS 이미지 데이터의 크기에 따른 단기 기온 예측의 성능에 대해 논하기로 한다.

Ⅱ. 본론

2.1 데이터 구성

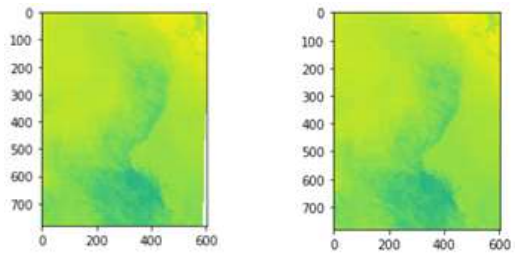
사용한 데이터는 기상청에서 제공한 관측 데이터와 수치예보모델 데이터로 구성되었다. 기간은 LDAPS 제공 시작일인 2012년 9월 10일부터 2015년 12월 31일까지 약 3년 4개월이다. 이 데이터를 6:2:2 비율로 training, validation, 그리고 test 데이터로 각각 분할하였다.

관측 데이터란 지역 날씨를 지상, 해상 혹은 고층에서 직접 관측한 데이터이다. 본 연구에는 AWS (Automatic Weather Station) 관측 데이터를 사용하였다. AWS는 관측 공백 해소 및 국지적인 기상 현상을 파악하기 위해 실시하는 지상관측을 말하고 전국 약 510여 지점에서 관측한다. 본 연구에서는 도봉구(지역 번호 406) 지점의 기온, 1시간 누적 강수량, 10분 평균 풍향 및 풍속, 습도 등 5개 변수 데이터를 사용하였다. 따라서 AWS는 변수 개수에 따라 5차원 벡터가 한 시간 간격으로 모델에 입력된다. 하지만 AWS는 점검이나 고장 등 각종 이유에 따라 결측될 수가 있어 모델 학습에 바로 적용할 경우 문제가 된다. 구체적으로는, 수집한 데이터 중 389시간에 문제가 있었고, 544개 데이터가 결측되었다. 따라서 본 연구에서는 [6]에서 제안한 방법에 따라 전처리를 수행하였다.

한편, 수치예보모델이란 역학과 물리학적 원리에 대한 방정식들을 슈퍼컴퓨터를 활용하여 대기 상태를 분석하고 미래의 대기 상태를 정량적으로 예측하는 일련의 과정이다. 기상청에서 제공하는 대표적인 수치예보모델은 영국기상청의 통합수치예보모델(UM)을 기반으로 구축된 GDAPS (Global Data Assimilation and Prediction System), RDAPS, LDAPS 등이 있다. 각각 모델들은 서로 다른 예측시간, 수평 격자의 크기와 간격을 가진다. 본 연구에서는 국지적인 지역의 단기 예측을 위해 예측 시간이 짧고 해상도가 높은 LDAPS를 선택하였다.

LDAPS의 공간 해상도는 1.5km이며, 연직으로 70층으로 구성되며, 3시간 간격으로 GDAPS로부터 경계장을 제공받아 1일 8회 예측을 수행하고. 등압면, 모델면, 단일면 자료 3종류가 제공된다. 본 연구에서는 등압면의 변수 중 10,000Pa의 기온을 입력 데이터로 사용하였다.

LDAPS 또한 전처리 과정이 필요하다. 예를 들면, 그림 1의 우측 하단에 결측으로 인한 공백이 생긴 것을 볼 수 있다. 이를 3x3 필터로 결측치를 필터의 평균값으로 채우는 방식으로 이미지를 보간하였다. 앞서 언급한 바와 같이 1일 기준 3시간 간격으로 8회만 제공되므로 AWS와 일대일 대응시키기 위해 좌표를 고정하고 시간 축으로 선형 보간을 진행하였다. 이때 LDAPS를 입력 크기에 따른 성능을 평가하기 위하여 관측 지점을 중



(a) (b)

그림 1. LDAPS의 전처리 전후 이미지 비교: 전처리 (a) 전 및 (b) 후

심으로 다양한 크기의 정사각형 모양으로 전처리를 수행하였다. 결과적으로, 각 이미지의 크기는 20x2, 40x40, 80x8, 320x320 등 4가지를 설정하였다. 이때, 320x320은 남한 전역으로 설정한 이미지의 크기에 해당한다.

2.2 단기 기온 예측 모델

2.2.1. 주의 집중 기법을 적용한 CNN-BLSTM 기반 모델

본 논문에서의 단기 기온 예측 모델[3]은 크게 특징 추출, 정보 융합, 예측 등 3단계로 구성되어 있다. 특징 추출에서는 AWS와 LDAPS의 특징을 서로 다른 구조로 추출한다. AWS의 특징 추출은 BLSTM 모델을 사용한다. 이 모델은 256개의 은닉 노드를 갖는 두 개의 BLSTM 층과 repeat vector 층, dense 층으로 구성되었다. 시간마다 5차원 벡터가 있고, 현재 시간으로부터 이전 시간까지 예측 시간만큼의 데이터를 입력으로 한다. 즉, 예측 시간이 6시간인 경우 입력 데이터의 크기는 6x5이다. 다음으로 첫 BLSTM이 양방향성에 의해 512차 벡터 하나를 출력한다. BLSTM의 512차원 출력이 예측 시간만큼 repeat vector 층으로 입력되어 이 층을 거친 벡터의 크기는 예측 시간x512이다. 이후 두 번째 BLSTM 층을 거쳐 5개 AWS 변수에 대한 특징이 추출된다. 동시에 앞서 설명한 동일한 BLSTM 모델 구조에 기온 변수만 입력으로 하여 특징을 추출한다.

수치예보모델 데이터의 특징을 추출을 위해서는 CNN-BLSTM기반 모델을 사용한다. 이 모델은 여러 Conv(Convolution) 층을 거쳐 나온 벡터를 flatten 층을 거쳐 1차원 벡터로 전환돼 AWS 특징 추출과 같은 구조의 BLSTM 층들을 거쳐 LDAPS 이미지 데이터에 대한 특징이 추출된다. 이후 정보 융합에서는 2번의 융합을 진행한다. 먼저 기온으로만 뽑은 특징과 LDAPS로부터 뽑힌 특징을 내적 연산을 통해 context vector를 만든다. 다음으로 AWS의 특징, LDAPS 이미지 데이터의 특징과 context vector를 연결해 concatenate 층을 만든다.

최종적으로 dense 층에 입력해 기온을 예측한다. 예측된 기온은 AWS의 기온 간의 평균 제곱 오차(MSE, Mean-Squared Error)를 손실함수로 설정하여 Adam Optimizer로 역전파된다. 본 논문의 모든 실험에서는 64 batch, learning rate 0.001로 설정하였으며, 80 epoch을 최대로 하여 early stopping 기법을 적용하였다.

2.2.2 LDAPS 이미지 크기에 따른 특징 추출

단기 예측을 할 때 관측지점을 중심으로 얼마만큼 넓을 영역을 학습에 사용해야 최적의 결과를 얻을 수 있을지에 대한 답을 얻기 위해 2.2.1절의 LDAPS 특징 추출 모델을 여러 가지로 변경하며 실험을 진행하였다.

특징 추출을 위한 CNN모델의 구조는 기본적으로 2차원 Conv 층, ReLU 활성함수, 2x2차원 max pooling 층이 한 블록으로 이뤄진 구조이다. 세부적인 파라미터와 블록의 수는 서로 다르다.

20x20 크기의 이미지 추출을 위해, 3x3 커널을 가진 Conv 층 2개, 5x5 커널을 가진 Conv 층과 한 개의 2x2차원 max pooling 층으로 구성되었다. 40x40 이미지에 대해서는 2개의 블록이 적용되었다. 첫 Conv 층은 5x5 커널과 8 채널, 두 번째는 7x7 커널과 32 채널을 가진다. 80x80은 3개의 블록이 적용되었다. 모든 Conv 층은 5x5 커널을 가지고 층이 깊어질수록 4, 8, 32 채널을 가진다. 마지막으로, 320x320 이미지를 위해서는 5개의 블록이 적용되었다. 모든 Conv 층은 5x5 커널을 가지고 층이 깊어질수록 4에서 64까지 2배씩 채널이 증가한다. 출력한 특징은 6x6x64 차원이다. 320x320 이미지 외의 나머지 이미지는 6x6x32 차원을 출력한다.

표 1. LDAPS 입력 이미지 크기에 따른 6시간 기온 예측 성능 비교

이미지 크기 / 평가 척도	20x20	40x40	80x80	320x320	기존모델[3]
RMSE(℃)	1.78	1.88	1.88	3.89	1.94
MAE(℃)	1.39	1.48	1.47	3.52	1.52

2.3 실험 및 결과

6시간 후 기온 예측에 대한 성능 평가를 위해 관측치와 예측치 간의 평균 제곱근 오차(RMSE, Root-Mean-Squared Error)와 평균 절대 오차(MAE, Mean Absolute Error)를 측정하였다. 표 1에서 보는 바와 같이, 이미지 크기가 320x320 일 때를 제외하면 [3]에서 제안된 기존 모델보다 좋은 성능을 보였다. 특히, 20x20 이미지를 입력으로 주었을 때 RMSE와 MAE가 각각 1.78과 1.39로 가장 작은 기온 예측 오차를 보였다. 단기 기온 예측에서 영상의 크기를 작게 볼수록 예측 오차가 적음을 알 수 있다.

Ⅲ. 결론

본 연구에서는 AWS 관측 데이터와 LDAPS 예측 수치 데이터를 활용한 단기 기온 예측 모델의 성능을 평가하였다. 기존의 RDAPS 수치 데이터에 비해 LDAPS 수치 데이터를 사용한 경우, 단기 기온 예측 오차를 줄일 수 있었다. 특히, LDAPS의 이미지 크기를 조절할 경우 기온 예측 오차를 더욱 감소시킬 수 있었다.

ACKNOWLEDGMENT

이 논문은 2021년도 광주과학기술원의 재원으로 GRI(GIST연구원) 사업의 지원을 받아 수행된 연구임.

참 고 문 헌

[1] 기상청, “2018 기상연감,” 2018.

[2] 기상청, “예보관 훈련용 기술서-수치예보.”

[3] S. Jeong, et al., “Temperature prediction based on bidirectional long short-term memory and convolutional neural network combining observed and numerical forecast data,” Sensors, 21(3), 941, 2021.

[4] Y.-G. Ham, et al., “Deep learning for multi-year ENSO forecasts,” Nature 573, pp. 568 - 572, 2019.

[5] C. K. Sønderby, et al., “MetNet: A neural weather model for precipitation forecasting,” 2020.

[6] I. Park, et al., “Temperature prediction using the missing data refinement model based on a long short-term memory neural network,” Atmosphere, vol. 10(11), 718, 2019.

On the Performance Gains of Decentralized Edge Caching Schemes in 5G and Beyond

Lilian C. Mutalemwa and Seokjoo Shin*
Department of Computer Engineering, Chosun University, Gwangju 61452, South Korea
Email: lilian.mutalemwa@gmail.com, *sjshin@chosun.ac.kr (corresponding author)

***Abstract*—Edge caching is one of the key technologies that enable low-latency communications in 5G and beyond 5G networks. Performance of an edge caching scheme may vary depending on the type of learning model. Thus, edge caching schemes present different performance when centralized or decentralized learning algorithms are used. In this study, we outline the performance gains of decentralized edge caching. We consider various edge caching frameworks based on deep learning (DL), deep reinforcement learning (DRL), and federated learning (FL) algorithms. It is shown that decentralized frameworks present viable mechanisms to outperform centralized frameworks, especially in dynamic and heterogeneous large-scale networks or complex IoT environments.**

Keywords—5G; ultra-reliable low-latency communications; edge caching; decentralized edge caching; federated learning.

I. INTRODUCTION

Ultra-reliable low-latency communications (URLLC) is probably the most talked-about 5G and beyond 5G (B5G) use case mainly because of the huge services it can support [1]. URLLC aims to deliver a vastly reliable mobile wireless network with extremely low latency requirements. Edge caching is one of the key technologies that facilitate low-latency communications in 5G and B5G networks. Particularly, mobile edge caching (MEC) has been regarded as a promising technique to provide low latency for content access [1]-[3]. In the MEC systems, popular contents can be cached in proximity to the edges of networks in edge devices (EDs), e.g. base stations (BSs) and user equipment (UE) (or mobile devices), which reduces massive duplicated traffic of content deliveries via backhaul networks and shortens the content delivery latency [3]-[26]. An example use case of a MEC system may be video caching on the network edge. At present, video delivery is the dominant traffic in the network. In the year 2022, more than 82% of network traffic will be video traffic, and this number will increase to 90% by the year 2025 [5]. In video delivery, reducing network delay is one of the key factors for improving users' quality of experience (e.g., significantly reduce the playback latency). Therefore, video caching is pushed from the cloud server to the edge network to provide end-users a low-delay video delivery [5].

Employing an appropriate caching algorithm is pivotal to increase the overall quality of experience in content distribution systems as 1% increase in hit rate can have a positive impact [1]. Therefore, in this study, we investigate the performance of centralized and decentralized edge caching solutions. We consider various edge caching frameworks based on deep learning (DL), deep reinforcement learning (DRL), and federated learning (FL) algorithms. Subsequently, we highlight the performance gains of the decentralized edge caching frameworks.

II. CENTRALIZED AND DECENTRALIZED EDGE CACHING SCHEMES

In general, edge caching-enabled systems may be identified as centralized or decentralized, based on their learning models [7], [27]. Due to bandwidth, storage, and privacy concerns, centralized edge caching systems are often impractical [26]. The centralized caching algorithms may result in overconsumed network resources during the training and data transmission processes [9]. The effects of centralized caching schemes can become severe in dynamic and heterogeneous large-scale systems. The centralized edge caching may be realized through DL or DRL algorithms. To address the challenges of centralized edge caching, decentralized caching schemes are considered. The decentralized edge caching frameworks may be realized through DL, DRL, or FL algorithms. Thus, all FL schemes are decentralized. FL is becoming increasingly popular in 5G and B5G networks because it is effective and privacy-preserving by design [28], [29].

Decentralized edge caching schemes are often considered because it is assumed that most 5G and B5G networks will be based on decentralized and infrastructureless communication to enable devices to cooperate directly over device-to-device spontaneous connections [30].

III. PERFORMANCE INVESTIGATIONS

In our investigations, we consider various edge caching solutions based on DL, DRL, and FL learning algorithms. Also, multi-agent DRL (MADRL) and distributed DL (DDL) are considered. The solutions are adopted from [7], [9], [17], [19]-[21], [31], and [32]. Then, we examine the performance of the solutions. Our observations are based on the analysis and evaluations presented in [7], [9], [17], [19]-[21], [31], [32]. A summary of the observations is presented in Table I.

The performance of the DL, DRL, and FL edge caching solutions is compared to the performance of two baseline solutions, the least recently used (LRU) and least frequently used (LFU) caching solutions which are commonly used by content providers [7], [21]. In the LRU algorithm, the system keeps track of the most recent requests for every cached data

TABLE I. PERFORMANCE FEATURES OF CENTRALIZED AND DECENTRALIZED EDGE CACHING SCHEMES.

Technique	Limitations	Example solution	Objective	Learning model	Performance
DL-based edge caching	• Traditional DL-based optimization and prediction schemes take a long running time of recursions for converging to the optimal model. • In centralized schemes, sending streams of raw training data to server can increase network traffic and energy consumption. • Most schemes cannot handle non-IID data or privacy preservation issues.	DLs [19]	• Reduce latency and backhaul network traffic for 5G mobile video streaming.	• Decentralized DL algorithm.	• Lower latency than LRU and LFU. • Higher cache hit rate than LRU and LFU.
		DLs2 [17]	• Reduce service delay for UEs and error of content demand prediction.	• Centralized DL algorithm.	• Lower latency than LRU and LFU. • Higher cache hit rate than LRU and LFU. • Higher latency than DDLs. • Lower cache hit rate than DDLs.
		DDLs [17]	• Reduce service delay for UEs and error of content request prediction while preserving privacy of UEs data.	• Decentralized DDL algorithm.	• Lower latency than LRU, LFU and DLs2. • Higher cache hit rate than LRU, LFU and DLs2.
DRL-based edge caching	• Requires intensive computation capacity for finding optimal model particularly in large-scale data. • Achieves reduced performance when UEs and network states are heterogeneous. • In large-scale data with massive UEs, centralized DRL schemes incur increased traffic on uplink wireless channels. • In large-scale data with massive UEs, it is challenging to perform decentralized DRL due to relatively weak computation capability of UEs. Also, it takes long time to train the DRL agent. • Decentralized DRL increases the energy cost at the UEs. • Most schemes cannot handle unbalanced and non-IID data or privacy preservation issues.	DRLs [32]	• Minimize communication cost and loss of data freshness in IoT applications.	• Centralized DRL algorithm.	• Significantly lower latency than LRU and LFU. • Significantly higher cache hit rate than LRU and LFU. • Higher latency than MADRLs. • Lower cache hit rate than MADRLs.
		MADRLs [21]	• Minimize content access latency and traffic cost in diversified 5G video streaming environment.	• Decentralized MADRL algorithm.	• Significantly lower latency than LRU and LFU. • Significantly higher cache hit rate than LRU and LFU. • Lower latency than DRLs. • Higher cache hit rate than DRLs.
		DRLs2 [7]	• Improve cache hit rate in media-enabled applications.	• Centralized DRL algorithm.	• Higher cache hit rate than LRU and LFU but lower than MADRLs2.
		MADRLs2 [7]	• Reduce latency and improve cache hit rate in media-enabled applications.	• Decentralized MADRL algorithm.	• Lower latency than LRU, LFU, and DRLs2. • Higher cache hit rate than LRU, LFU, and DRLs2.
FL-based edge caching	• When traditional algorithms such as FedAvg is used, FL suffers from a large number of communication rounds to convergence with non-IID datasets. Also, has high communication overhead.	FedDRLs [9]	• Reduce latency, performance loss, and backhaul traffic while improving hit rate in IoT systems.	• FL-based DRL algorithm.	• Significantly lower latency than LRU and LFU. • Significantly higher cache hit rate than LRU and LFU. • Latency and cache hit rate are comparable to a centralized DRL scheme.
		FedDRLs2 [20]	• Make mobile communication system cognitive and adaptive, reduce network traffic, and achieve near-optimal performance with low overhead of learning.	• FL-based DRL algorithm.	• Significantly higher cache hit rate than LRU and LFU. • Cache hit rate is comparable to a centralized DRL scheme.
		FedDRLs3 [31]	• Reduce transmission costs between IoT devices and EDs.	• FL-based DRL algorithm.	• Transmission time is comparable to a centralized DRL.

content. When the cache storage becomes full, the cached content which is requested least recently, is replaced by the new content [7], [9]. For the LFU algorithm, the system keeps track of the number of requests for every cached content. When the cache storage becomes full, the cached content which is least frequently requested, is replaced by the new content [7], [9]. We compare the performance of the solutions in terms of content delivery latency and cache hit rate. The cache hit rate is used to show how frequently the requested content is found in the local cache [7], [21]. Therefore, we assume that cache hit rate indicates the content acquisition reliability. That means, a high cache hit rate corresponds to high content acquisition reliability.

It is shown in Table I that the DL, DRL, and FL edge caching solutions are capable of achieving improved performance in terms of content delivery latency and cache hit rate to outperform the LRU and LFU solutions. The main reason for the poor performance in the LRU and LFU solutions is that, the algorithms in LRU and LFU do not consider the popularity of contents in the future. As a result, the solutions do not adapt well to the dynamically changing content popularity and they achieve low cache efficiency [24], [33]. For instance, the LFU framework is not able to reach a good performance in IoT environment because it does not consider the saltation and timeliness of the IoT data popularity [33]. It was also shown in [9], [19], [21] that the DL, DRL, and FL edge caching solutions are capable of achieving reduced backhaul network traffic to outperform the LRU and LFU solutions.

On the other hand, Table I shows that although the solutions with centralized DL, DRL, and FL algorithms perform better than the LRU and LFU solutions, the centralized solutions present reduced performance when compared to the solutions with decentralized algorithms. For example, it was shown in [17] that a DDL solution performs better than a centralized DL solution in terms of latency and cache hit rate. Furthermore, since the DDL framework only needs to collect the trained models from the EDs without considering any raw dataset transmission, the DDL framework was able to achieve reduced communication overhead. Also, the DDL was able to learn the dataset faster than the centralized DL as the number of the EDs was increased. It was also presented in [7], [21] that decentralized MADRL frameworks achieve better performance than centralized DRL frameworks. As an example, when a massively vibrant, diversified, and distributed video streaming environment was considered in [21], a decentralized MADRL framework presented better performance than a centralized DRL framework in terms of video access latency and the traffic cost. In [20], it was revealed that a FL scheme can consume significantly lower communication resources than a centralized DRL framework. In [20], [31], it was presented that the FL-based solutions achieve a lower number of dropped tasks, queuing delay, and transmission energy to outperform the DRL solutions.

Conversely, when the performance of the FL frameworks was investigated in [9], [20], [31], it was revealed that although FL schemes can address the challenges of DRL and DL frameworks, the FL schemes are less capable of achieving significantly improved performance in terms of content access latency and cache hit rate. For example, the performance of the FL schemes in [9], [20], [31] was comparable to the performance of the centralized DRL schemes once the model aggregation of FL was performed several times. That means, for the FL algorithms to achieve the performance level of the centralized DRL algorithms in terms of content access latency and cache hit rate, the FL algorithms must allow several rounds of model aggregation.

On the other hand, it was pointed out in [9], [31] that the performance level of the FL schemes is reasonable since the FL schemes assume more practical network conditions. As an example, the centralized DRL scheme considered in [9], [20], [31] assumed that the massive training data can be successfully uploaded to the ED without loss or delay. Considering the limitations of wireless channels, the assumption made by the DRL scheme may be impractical. Moreover, the work in [20], [34] considered the challenge that when not independent and identically distributed (non-IID) datasets are used, FL algorithms incur large number of communication rounds to converge to the global optimal. Consequently, [20], [25], [34] highlighted the technique of transfer learning as a potential solution for the challenge. Thus, [20], [25], [34] considered the use of transfer learning technique to improve the learning efficiency of the FL algorithms. It was pointed out in [34] that the transfer learning technique can ensure training is not initialized from scratch. In [25], it was demonstrated that personalized federated learning can significantly reduce the performance degradations caused by the non-IID data.

Several other benefits of using FL edge caching were highlighted in [15], [20], [22], [27], [34]. The following benefits were outlined: the system becomes more cognitive and robust, improved flexibility, reduced network traffic and energy consumption, privacy preservation, and improved stability despite loss of connectivity.

IV. CONCLUSION AND FUTURE WORK

This paper presents some investigations on the performance features of centralized and decentralized edge caching schemes. It is shown that decentralized frameworks present better performance than centralized frameworks in terms of content delivery latency and traffic cost. Moreover, the decentralized frameworks perform better in a massively vibrant, diversified, and distributed video streaming environment or in dynamic and heterogeneous large-scale networks where devices are resource-constrained, including in complex IoT environments. Furthermore, it is shown that FL edge caching presents viable mechanisms in 5G and B5G networks. However, traditional FL algorithms incur high communication overhead. Therefore, as part of our future work, we will explore the techniques to improve the communication efficiency of FL algorithms. In particular, we

will study the mechanisms of transfer learning and personalized FL.

ACKNOWLEDGMENT

This research is supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2018R1D1A1B07048338).

REFERENCES

[1] E. E. Ugwuanyi, M. Iqbal and T. Dagiuklas, "A novel predictive-collaborative-replacement (pcr) intelligent caching scheme for multi-access edge computing," *IEEE Access*, vol. 9, pp. 37103–37115, March 2021.

[2] J. Liang *et al*., "Multi-head attention based popularity prediction caching in social content-centric networking with mobile edge computing," *IEEE Communications Letters*, vol. 25, no. 2, pp. 508–512, Feb. 2021.

[3] X. Wang, R. Li, C. Wang, X. Li, T. Taleb and V. C. M. Leung, "Attention-weighted federated deep reinforcement learning for device-to-device assisted heterogeneous collaborative edge caching," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 1, pp. 154–169, Jan. 2021.

[4] X. Xia, F. Chen, Q. He, J. Grundy, M. Abdelrazek and H. Jin, "Online collaborative data caching in edge computing," *IEEE Transactions on Parallel and Distributed Systems*, vol. 32, no. 2, pp. 281–294, Feb. 2021.

[5] Y. Guan, X. Zhang and Z. Guo, "PrefCache: Edge cache admission with user preference learning for video content distribution," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 4, pp. 1618–1631, April 2021.

[6] S. K. Sharma, I. Woungang, A. Anpalagan, and S. Chatzinotas, "Toward tactile Internet in beyond 5G era: Recent advances, current issues, and future directions," *IEEE Access*, vol. 8, pp. 56948–56991, 2020.

[7] C. Zhong, M. C. Gursoy, and S. Velipasalar, "Deep reinforcement learning-based edge caching in wireless networks," *IEEE Trans. Cognit. Commun. Netw.*, vol. 6, no. 1, pp. 48–61, Mar. 2020.

[8] M. Mehrabi, D. You, V. Latzko, H. Salah, M. Reisslein, and F. H. P. Fitzek, "Device-enhanced MEC: Multi-access edge computing (MEC) aided by end device computation and caching: A survey," *IEEE Access*, vol. 7, pp. 166079–166108, Nov. 2019.

[9] X. Wang, C. Wang, X. Li, V. C. M. Leung, and T. Taleb, "Federated deep reinforcement learning for Internet of Things with decentralized cooperative edge caching," *IEEE Internet Things J.*, vol. 7, no. 10, pp. 9441–9455, Oct. 2020.

[10] S. He, J. Ren, J. Wang, Y. Huang, Y. Zhang, W. Zhuang, and S. Shen, "Cloud-edge coordinated processing: Low-latency multicasting transmission," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 5, pp. 1144–1158, May 2019.

[11] T. Zhang, X. Fang, Y. Liu, and A. Nallanathan, "Content-centric mobile edge caching," *IEEE Access*, vol. 8, pp. 11722–11731, Jan. 2020.

[12] U. Paul, J. Liu, S. Troia, O. Falowo, and G. Maier, "Traffic-profile and machine learning based regional data center design and operation for 5G network," *J. Commun. Netw.*, vol. 21, no. 6, pp. 569–583, Dec. 2019.

[13] S. Zhang, P. He, K. Suto, P. Yang, L. Zhao, and X. Shen, "Cooperative edge caching in user-centric clustered mobile networks," *IEEE Trans. Mobile Comput.*, vol. 17, no. 8, pp. 1791–1805, Aug. 2018.

[14] G. Zhu, Y. Wang, and K. Huang, "Broadband analog aggregation for lowlatency federated edge learning," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 491–506, Jan. 2020.

[15] S. Samarakoon, M. Bennis, W. Saad, and M. Debbah, "Federated learning for ultra-reliable low-latency V2 V communications," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2018, pp. 1–7.

[16] J. Konečný, H. B. McMahan, D. Ramage, and P. Richtárik, "Federated optimization: Distributed machine learning for on-device intelligence," *CoRR*, 2016. [Online]. Available: http://arxiv.org/abs/1610.02527

[17] Y. M. Saputra *et al.*, "Distributed deep learning at the edge: A novel proactive and cooperative caching framework for mobile edge networks," *IEEE Wireless Commun. Lett.*, vol. 8, no. 4, pp. 1220–1223, Aug. 2019.

[18] X. Fan, Y. Huang, X. Ma, J. Liu, and V. C. M. Leung, "Exploiting the edge power: An edge deep learning framework," *CCF Trans. Netw.*, vol. 2, no. 1, pp. 4–11, Dec. 2018.

[19] H. Pang, J. Liu, X. Fan, and L. Sun, "Toward smart and cooperative edge caching for 5G networks: A deep learning based approach," in *Proc. IWQoS*, Jun. 2018, pp. 1–6.

[20] X. Wang, Y. Han, C. Wang, Q. Zhao, X. Chen, and M. Chen, "In-edge AI: Intelligentizing mobile edge computing, caching and communication by federated learning," *IEEE Netw.*, vol. 33, no. 5, pp. 156–165, Sep. 2019.

[21] F. Wang, F. Wang, J. Liu, R. Shea, and L. Sun, "Intelligent video caching at network edge: A multi-agent deep reinforcement learning approach," in *Proc. IEEE INFOCOM*, Jul. 2020, pp. 2499–2508.

[22] S. Niknam, H. S. Dhillon, and J. H. Reed, "Federated learning for wireless communications: Motivation, opportunities and challenges," 2019, *arXiv:1908.06847*. [Online]. Available: http://arxiv.org/abs/1908.06847

[23] Y. Lu, X. Huang, Y. Dai, S. Maharjan, and Y. Zhang, "Differentially private asynchronous federated learning for mobile edge computing in urban informatics," *IEEE Trans. Ind. Informat.*, vol. 16, no. 3, pp. 2134–2143, Mar. 2020.

[24] Z. Yu, J. Hu, G. Min, H. Lu, Z. Zhao, H. Wang, and N. Georgalas, "Federated learning based proactive content caching in edge computing," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2018, pp. 1–6.

[25] Q. Wu, K. He, and X. Chen, "Personalized federated learning for intelligent IoT applications: A cloud-edge based framework," *IEEE Open J. Comput. Soc.*, vol. 1, pp. 35–44, May 2020.

[26] S. Wang *et al.*, "Adaptive federated learning in resource constrained edge computing systems," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 6, pp. 1205–1221, Jun. 2019.

[27] L. C. Mutalemwa and S. Shin, "A classification of the enabling techniques for low latency and reliable communications in 5g and beyond: Ai-enabled edge caching," *IEEE Access*, vol. 8, pp. 205502-205533, Nov. 2020.

[28] S. Savazzi, M. Nicoli, M. Bennis, S. Kianoush and L. Barbieri, "Opportunities of federated learning in connected, cooperative, and automated industrial systems," *IEEE Communications Magazine*, vol. 59, no. 2, pp. 16-21, Feb. 2021.

[29] J. Posner, L. Tseng, M. Aloqaily and Y. Jararweh, "Federated learning in vehicular networks: Opportunities and solutions," *IEEE Network*, vol. 35, no. 2, pp. 152-159, April 2021.

[30] S. Savazzi, M. Nicoli, and V. Rampa, "Federated learning with cooperating devices: A consensus approach for massive IoT networks," *IEEE Internet Things J.*, vol. 7, no. 5, pp. 4641–4654, May 2020.

[31] J. Ren, H. Wang, T. Hou, S. Zheng, and C. Tang, "Federated learning-based computation offloading optimization in edge computing-supported Internet of Things," *IEEE Access*, vol. 7, pp. 69194–69201, Jun. 2019.

[32] H. Zhu, Y. Cao, X. Wei, W. Wang, T. Jiang, and S. Jin, "Caching transient data for Internet of Things: A deep reinforcement learning approach," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2074–2083, Apr. 2019.

[33] B. Chen, L. Liu, M. Sun, and H. Ma, "IoTCache: Toward data-driven network caching for Internet of Things," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 10064–10076, Dec. 2019.

[34] W. Y. B. Lim *et al.*, "Federated learning in mobile edge networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 3, pp. 2031–2063, 3rd Quart., 2020.

사전 훈련된 딥 러닝 모델의 파인튜닝을 위한 최적의 배치 크기

아흐마드 이자즈, 신석주*
컴퓨터공학과, 조선대학교

ahmadijaz@chosun.kr, *sjshin@chosun.ac.kr (교신저자)

Optimal Batch Size for Fine-Tuning Pre-Trained Deep Learning Models

Ijaz Ahmad, Seokjoo Shin
Dept. of Computer Engineering, Chosun University.

요 약

Training a deep learning model from scratch significantly benefits from high resolution images. However, when using transfer learning, the higher resolution does not necessarily offer better performance and the model achieves better test accuracies on smaller images. The advantage of the smaller size is that it allows vast batch sizes during training, which can reduce model's training time. Therefore, in this study we have analyzed varying batch sizes when using a pre-trained MobileNetV2 model for classification task. We have performed our experiments on CIFAR 10 and CIFAR 100 datasets and the images are resized to 160x160. Our results have shown that the same or even improved accuracy can be achieved with less number of epochs when using larger batch size. For example, in the case of CIFAR10, there is no significant difference in the accuracy while in the case of CIFAR100 for batch size 512, there is a 2% improvement in the accuracy.

I. Introduction

Transfer learning (TL) is a machine learning method where what has been learned in one problem domain is repurposed to improve generalization in another related problem domain [1]. Usually, it is hard to get a dataset of sufficient size, e.g., medical image datasets, to train a network from scratch with random initialization. In such scenarios, TL improves the performance of a model through the transfer of knowledge and allows rapid progress. There are two main approaches of TL [2]. The first is to retrain only the classifier of the pre-trained model on the new dataset while its convolution network (ConvNet) is used as a fixed feature extractor. This approach is faster as only the classifier is trained. Alternatively, the second approach is to not only retrain the classifier but fine-tune the ConvNet on the new dataset as well. Fine-tuning a pre-trained CNN on the target data can significantly improve the performance of the model. However, when overfitting is of concern (e.g., when the new dataset is small), only a part of the ConvNet can be fine-tuned.

Transfer learning boosts the performance of a learning algorithm significantly; however, the pre-trained networks have certain constraints in terms of their architecture. For example, due to the wide availability of color images, the state-of-the-art pre-trained models are trained on color images. [3] has shown how to run pre-trained color image models on grayscale images. Their analysis has shown that comparable classification accuracy can be achieved when using grayscale images. The second challenge is that the available pre-trained models are trained for a specific input size. Therefore, require resizing the input images of different sizes.

The selection of optimal hyperparameter for training a deep learning model can avoid unnecessary computation [4]. The literature covers in detail the hyperparameter selection, e.g., image resolution, batch size, etc., when training neural networks from scratch. However, fewer studies have covered the selection of hyperparameter for transfer learning, e.g., [4] studied the momentum and [5] proposed Bayesian optimization to select learning rate for transfer learning. [6] studied the selection of optimal image resolution when fine-tuning a neural network. Their results have shown that the pre-trained models can achieve better accuracy on smaller size images. The smaller resolution of images can allow the use of larger batch sizes, which can reduce the training time. Therefore, in this work we have analyzed varying batch sizes when fine-tuning pre-trained MobileNetV2 model for classification task.

Ⅱ. Methods

MobileNetV2 [7] is a deep learning model for classification, detection and segmentation tasks. The model is suitable to run on low computational devices such as mobile devices, thanks to its smaller number of parameters. The model uses separable convolution as efficient building blocks as its predecessors, but with two new features: linear bottlenecks between the layers and shortcut connections between the bottlenecks. The pre-trained MobileNetV2 comes with weights for five input image sizes (96, 128, 160, 192 and 224). When training MobileNetV2 model on ImageNet datasets from scratch, larger image sizes offer better performance. However, [6] has shown the contrast that when fine-tuning the pre-trained model for the classification of the CIFAR datasets, the model

has better performance for image size of 160^2. Therefore, in our experiments we have used the pre-trained weights for input size of 160.

In our experiments, we have used the pre-trained MobileNetV2 model to classify CIFAR10 and CIFAR100 datasets [7]. In order to benefit from fine-tuning, we have initialized the ConvNet of the model with the pre-trained weights of MobileNetV2 and added a classifier on top of it as in [6]. The classifier has a global average-pooling layer followed by a drop out layer with ratio set to 0.5. The final layer is a dense layer and the number of nodes in the layer is equal to the number of classes in a dataset. We have perform our training in two stages: first, we have only retrain the classifier for 20 epochs. The initial learning rate was set to 10^{-2}. The learning was decremented by a factor of 10 when the validation accuracy stops improving. Fine-tuning a pre-trained CNN on the target data can significantly improve the performance of the model. Therefore, in the second stage of training, we have fine-tuned the whole network. The trained classifier from the previous stage and all the layers in the feature extractor were fine-tuned as well for 100 epochs. In the second stage, the initial learning rate was set to 10^{-4} be with a decrement factor of 10, in case the validation accuracy stops improving. In addition, to prevent the network from overfitting we have set early stopping criteria based on validation accuracy. During training, we have evaluated five batch sizes: 32, 64, 128, 256 and 512.

Ⅲ. Results and Discussion

In the simulations, we have used the MobileNetV2 pre-trained model as a classifier with the pre-trained weights for image resolution of 160x160. The experiments have been performed on the widely used CIFAR10 and CIFAR100 datasets [8]. The datasets consists of color images with resolution of 32x32. In our experiments, we have resized the images to 160x160 in order to be same as the network input size. Both datasets have 60k images with 50k training images and 10k test images. CIFAR10 has 10 classes and 6k images per class, while CIFAR100 has 100 classes and 600 images per class. In experiments, we have used 10% of training images as validation images.

To demonstrate the relationship between the batch size and classification accuracy, we have considered five batch sizes (32, 64, 128, 256 and 512) as shown in Fig. 1. We have plotted the validation accuracy as a function of the number of epochs for CIFAR10 in Fig. 1. (a) and CIFAR100 in Fig. 1. (b). It can be seen that varying batch sizes requires different number of epochs shown by the circle on the end of the line. The larger batch size requires fewer numbers of epochs than the smaller batch sizes. The abrupt change in the accuracy is the result of the learning rate decay. For CIFAR10, there is no significant difference in the accuracy for varying batch sizes. However, in the case of CIFAR100 for batch size 512, there is a 2% improvement in the accuracy. In addition, the benefits of the transfer learning can also be observed during training as the model starts with a better accuracy. We have summarized the achieved accuracy of MobileNetV2 on the CIFAR datasets for different batch sizes in Table I.

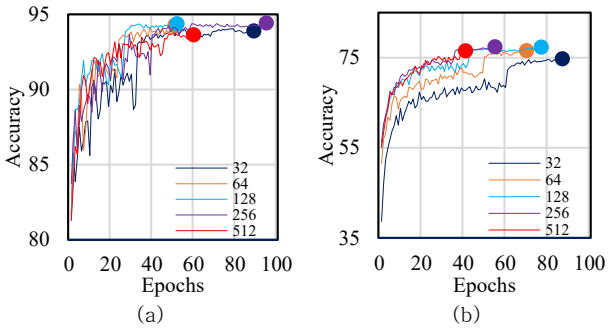


Fig. 1. Validation accuracy of pre-trained MobileNetV2 on (a) CIFAR10 and (b) CIFAR100 for different batch sizes.

TABLE Ⅰ. Classification accuracy of the pre-trained MobileNetV2 model on different batch sizes.

Batch Sizes	CIFAR10	CIFAR100
32	93.83	74.32
64	93.97	76.26
128	93.88	76.71
256	93.66	76.97
512	93.53	76.49

Ⅳ. Conclusion

The selection of optimal hyperparameter for training a deep learning model can avoid unnecessary computation. The literature extensively convers hyperparameter selection when training neural networks from scratch. However, there are few studies that cover the hyperparameter for transfer learning. Therefore, in this study we have analyzed various batch sizes for fine-tuning a pre-trained model. Our results have shown that the same accuracy can be achieved in fewer numbers of epochs with larger batch sizes.

Acknowledgment

This research is supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2018R1D1A1B07048338).

References

[1] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. Cambridge, Massachusetts: The MIT Pres, 2016.

[2] "CS231n Convolutional Neural Networks for Visual Recognition." https://cs231n.github.io/transfer-learning/ (Dec. 18, 2020).

[3] I. Ahmad and S. Shin, "An Approach to Run Pre-Trained Deep Learning Models on Grayscale Images," in *2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, Jeju, Korea, Apr. 2021, pp. 1-4.

[4] H. Li *et al.*, "Rethinking the Hyperparameters for Fine-tuning," *ArXiv*, vol. abs/2002.11770, 2020.

[5] R. J. Borgli, H. Kvale Stensland, M. A. Riegler, and P. Halvorsen, "Automatic Hyperparameter Optimization for Transfer Learning on Medical Image Datasets Using Bayesian Optimization," in *2019 13th ISMICT*, Oslo, Norway, May 2019, pp. 1-6. doi: 10.1109/ISMICT.2019.8743779.

[6] I. Ahmad and S. Shin, "Optimal resolution selection to run pre-trained deep learning models on tiny images," in *2021 한국차세대컴퓨팅학회 춘계학술대회*, Korea, May 21, pp. 172-174.

[7] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks,"

ArXiv180104381 Cs, Mar. 2019, Accessed: Mar. 08, 2021. [Online]. Available: http://arxiv.org/abs/1801.04381

[8] A. Krizhevsky, "Learning Multiple Layers of Features from Tiny Images," p. 60.

Evaluating Artificial Intelligence Mitigation Techniques for Countering Attack on Smart Factory SCADA Network

Love Allen Chijioke Ahakonye, Cosmas Ifeanyi Nwakanma, Jae Hyun Lee, Jae Min Lee, Dong-Seong Kim
IT Convergence Engineering, Kumoh National Institute of Technology.
(loveahakonye, cosmas.ifeanyi, antigen47, ljmpaul, dskim)@kumoh.ac.kr

Abstract

The new model of the Smart Factory (SF) compels all the industrial sectors to face an extensive digital evolution to be at the brim in a competitive scheme. As a result, Artificial Intelligence (AI) can be highlighted as a support to other technologies, such as cybersecurity. Following the trend of the vulnerability in the SF SCADA system, there is a serious need for cutting-edge mitigation solutions leveraging AI. This study aims at exploring AI techniques for the mitigation of attacks in the SCADA system. From the results of the evaluation, Weighted KNN, Fine KNN and Fine Tree are considered most ideal for SF SCADA attack mitigation.

Ⅰ. Introduction

Presently, industries have developed their business lines to offer the technology of critical industrial systems such as supervisory control and data acquisition (SCADA). This is to enable the inter-connectivity and remote accessibility of sub-stations. The smart factory (SF), SCADA system comprises the internet and industrial internet of things (IoT, and IIoT). This has facilitated the feasibility of a comprehensive implementation of remote administration which has bared the SCADA system to vulnerabilities and potential attacks. This increased the necessity for utmost safety for proficient administration and control, emerging in sustained industrial operations with premium achievement. Besides, the damage repair from attacks on SCADA and critical industrial systems can be overwhelming [1]-[4].

Mitigation techniques such as SCADA hardware security, network-based intrusion detection systems exist. Several studies have employed artificial intelligence (AI) techniques to resolve a variety of issues. AI is a technology that enables a machine simulation of human actions utilizing machine learning (ML). The automatic method of attack detection and classification in SF SCADA is based on AI techniques. However, the premise for determining the type of AI and its qualities establishes a research challenge.

Researchers have proffered ML prospects to tackle the advancement of intrusion. In [5], the authors utilized the Decision Tree, Random Forest, Naive Bayes, Logistic Regression and K-Nearest Neighbour (KNN) on a water storage facility testbed generated dataset. From their evaluation, the approach had an excellent performance in accuracy and false alarm rates with an unbalanced dataset. [6], had compared the performance of NB, KNN, ANN techniques. This technique performed poorly, hence, the ensemble technique was employed to enhance performance but this model requires minimal computational cost. Furthermore, deep learning (DL) techniques were proposed, and widely utilized for their ability to learn and predict meaningful attributes [7], [8].

Regardless of this number of studies, there is still a lack of convergence at establishing the most viable AI technique for attack detection in SCADA network. Hence, researchers face a dilemma at arriving at the AI technique of choice. Inspired by this challenge, this study seeks to achieve the following:

1) Utilization and appraisal of various AI techniques relevant to SCADA attack mitigation
2) Leveraging a current cyber-security dataset and choice of AI techniques, this work compared various ML models and their performances.
3) Preference in choosing the best proactive scheme based on computational complexity and accuracy in terms of model training time

Following section I is section II which is the overall system model. Section III is the evaluation of the AI techniques. The paper was concluded in section IV

Ⅱ. System Methodology

Modern smart factory's functionality are increasingly becoming complex. This is enabled by the possibility of massive connectivity of devices and sensors and thus, SCADA has become vulnerable to attack. Therefore, the need for efficient security is essential. To address SCADA network security vulnerability in the smart factory, this work presented a comparison of various AI algorithms developed to monitor communication traffic in encoded DNS and identify abnormal nature in the traffic. Fig. 1 depicts the system model of the study. The proposed architecture consists of 3 main phases; training, testing, and model determination phase. During the training and testing phase,the labelled cyber-security dataset which is extracted from normal web browsing activities and domain name service channelling mechanisms is split into training and testing sets and imported into the diverse ML algorithms after implementing a five (5) fold cross-validation. The test set was used to validate the performance of the training set. The choice and decision for the best model is regulated by the performance metrics (PM). The PM used in evaluating the models is the accuracy of the selected model, training time and the model's mis-classification error (MCE).

Fig 1. System Model of the Evaluation of AI Techniques for SF SCADA Vulnerability Mitigation

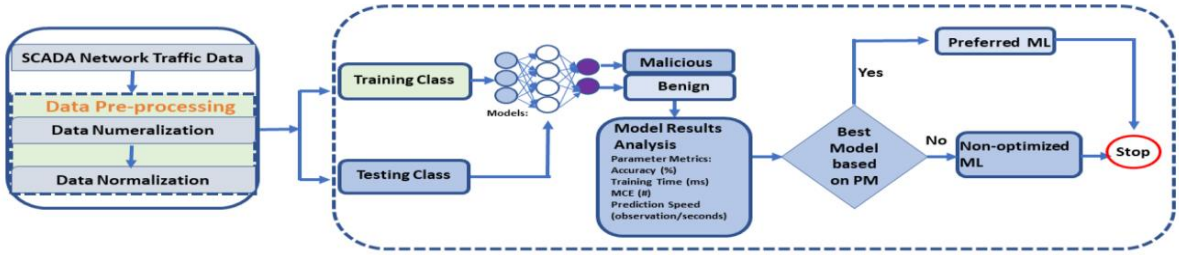


Table 1.
COMPARISON OF ACCURACY, TIME, MCE AND PREDICTION SPEED

Parameter Metrics	Weighted KNN	Fine KNN	Fine Tree
Accuracy (%)	99.2	99.2	98.1
Training Time (s)	8.8344	13.583	4.5778
MCE (#)	2134	2192	5096
Prediction Speed (obs/sec)	250000	440000	1100000

Table 2.
COMPARISON OF ACCURACY, TIME, MCE AND PREDICTION SPEED SHOWING 3 LEAST PERFORMED AI TECHNIQUES

Parameter Metrics	Bagged Tree	QDR	Coarse Tree
Accuracy (%)	99.4	92.9	95
Training Time (s)	85.488	3.8003	3.5817
MCE (#)	1512	19225	13357
Prediction Speed (obs/sec)	1600000	990000	1200000

Ⅲ. Performance Evaluation of various AI Techniques

A. Analysis and Comparison of various ML Algorithms for Intrusion Detection and Mitigation in SCADA

In this study, leveraging on a cyber-security dataset [9], MATLAB 2019Rb was used to examine all ML prospects. The algorithms include classes of Decision Trees, Discriminant Analysis, Logistic Regression, Naive Bayes, Support Vector Machines, K-Nearest Neighbours and Ensemble classifiers.

B. Summary of Performance Evaluation

A total of twenty-two (22) models were simulated, the result of the analysis for the top three best-performed algorithms is as shown in Table I and the least performed in Table II. The comparison is based on the parameter metrics in terms of model training time, accuracy and MCE. For validation, 30% of the dataset was reserved.

Ⅵ Conclusion

In this study, a relative analysis of various ML AI techniques was done. This was to determine the best ML prospect for attack mitigation in SF SCADA. From the evaluation, it is evident that several ML exhibit significant performance in terms of accuracy, MCE, training time and prediction speed. Nevertheless, considering the complexity of SF operations, the accuracy, detection time and number of misclassification errors are the basis for the choice of the preferred ML. Based on a trade-off of these parameter metrics, the Weighted KNN had the best performance occasioned by its versatility, interpretability, high accuracy and swift calculation time. However, it is not suitable for large dataset and sensitive to nosiy data.

ACKNOWLEDGMENT

This research work was supported by Priority Research Centers Program through NRF funded by MEST (2018R1A6A1A03024003) and the Grand Information Technology Research Center support program (IITP-2021-2020-0-01612) supervised by the IITP by MSIT, Korea.

References

[1] D.-S. Kim and H. Tran-Dang, "Industrial Sensors and Controls in Communication Networks," Computer Communications and Networks. Springer International Publishing, Cham, 2019.

[2] C. I. Nwakanma, F. B. Islam, M. P. Maharani, J.-M. Lee, and D.-S. Kim, "Detection and Classification of Human Activity for Emergency Response in Smart Factory Shop Floor," Applied Sciences, vol. 11, no. 8, 2021. [Online]. Available: https://www.mdpi.com/2076-3417/11/8/3662

[3] G. C. Amaizu, C. I. Nwakanma, S. Bhardwaj, J.- M. Lee, and D.-S. Kim, "Composite and Efficient DDoS Attack Detection Framework for B5G Networks," Computer Networks, vol. 188, p. 107871, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1389128621000438

[4] C. I. Nwakanma, M. S. Hossain, J.-M. Lee, and D.-S. Kim, "Towards Machine Learning Based Analysis of Quality of User Experience (QoUE)," International Journal of Machine Learning and Computing, vol. 10, no. 6, pp. 752–758, 2020. [Online]. Available: https://doi.org/10.18178/ijmlc.2020.10.6.1001

[5] M. Teixeira, T. Salman, M. Zolanvari, R. Jain, N. Meskin, and M. Samaka, "SCADA System Testbed for Cybersecurity Research Using Machine Learning Approach," Future Internet, vol. 10, no. 8, p. 76, Aug 2018. [Online]. Available: http://dx.doi.org/10.3390/fi10080076

[6] D. Upadhyay, J. Manero, M. Zaman, and S. Sampalli, "Intrusion Detection in SCADA based Power Grids: Recursive Feature Elimination Model with Majority Vote Ensemble Algorithm," IEEE Transactions on Network Science and Engineering, pp. 1–1, 2021.

[7] J. Li, Y. Qu, F. Chao, H. P. H. Shum, and L. Ho, Edmond S. L.and Yang, Machine Learning Algorithms for Network Intrusion Detection. Cham: Springer International Publishing, 2019, pp. 151–179. [Online]. Available: https://doi.org/10.1007/978-3-319-98842-96

[8] Z. Ahmad, A. Shahid Khan, C. Wai Shiang, J. Abdullah, and F. Ahmad, "Network Intrusion Detection system: A Systematic Study of Machine Learning and Deep Learning Approaches," Transactions on Emerging Telecommunications Technologies, vol. 32, no. 1, p. e4150, 2021

[9] "CIRA-CIC-DoHBrw-2020 Dataset," 2020. [Online]. Available: https://www.unb.ca/cic/datasets/dohbrw-2020.html

Depression Group Prediction with Passive Sensing: Prediction Accuracy and Window Sizes

Sabinakhon Akbarova, Youngtae Noh*

sabina@nsl.inha.ac.kr. ytnoh@nsl.inha.ac.kr*
Department of Electrical and Computer Engineering, Inha University, Incheon, Korea

Abstract

Depression diagnosis is an important issue, usually requiring mental health professionals' help, which is not affordable to everyone in terms of time and money. Passive sensing is an emerging technology allowing effortless depression detection with everyday mobile devices. One of the main challenges in eHealth applications is the improvement of prediction accuracy. In this work, we investigate the impact of feature extraction window sizes on depression group classification accuracy. Our findings illustrate the importance of window size hyperparameter and the positive tendency of accuracy with the increase of window size. The highest accuracy that we could achieve for multi-class classification is 77% for the window size of 14 days.

Ⅰ. Introduction

Depression is a common and dangerous illness accounting for almost 300 million people affected worldwide [1]. Although depression symptoms manifest in day-to-day life, they are not straightforward to diagnose and assess. Seeking help from mental health professionals is the most effective method for thorough diagnosis and individually tailored treatment. However, this approach is time-consuming, high-priced, and requires professional involvement.

Previous studies demonstrated the potential of smartphone passive sensing in depression detection since sensor data provides meaningful insights about behavioral patterns that, in turn, are highly correlated with depression [2-6]. It is common to extract features from raw sensor data within several hours epochs [2,4,6] or aggregate them into daily measures [3,4,5]. However, previous works lack investigating the effect of window sizes on depression group classification. Therefore, we conducted a study that makes the following contributions:

- We developed a data collection agent and passively collected sensor data from 60 people for two months to explore the feasibility of smartphone sensor data to detect depression severity.
- We predicted the depression severity with various window sizes for features extraction and showed the importance of the choice of the window size parameter.

II. Study design

Figure 1 illustrates the overview of the study procedure. The first step is participants recruitment and data collection. After data collection is finished, features are extracted from raw data and preprocessed for further analysis. As the last step of the experiment, we feed data to the machine learning classifiers and test the resulting models on unseen data.

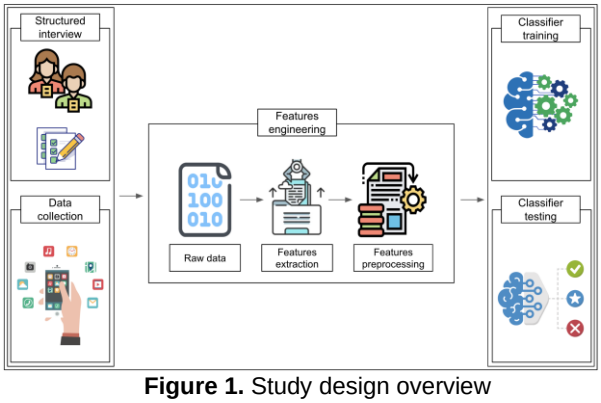


Figure 1. Study design overview

III. Data collection

The recruitment happened through online social communities. We recruited the participants on a first-come-first-serve basis. A total of four depression screening tools were conducted for depression severity assessment - Patient Health Questionnaire (PHQ-9) [7], Structured Clinical Interview for DSM-5 (SCID-5), Beck Depression Inventory-II (BDI-II) [8], and Korean Inventory of Depressive Symptomatology (K-IDS) [9]. Every applicant who wanted to participate in our study was required to, firstly, take a PHQ-9 online via Qualtrics [10]. We classified applicants into three groups (non-depressed, depressed group, severely depressed group) following the results of the PHQ-9 screening tool with cut-off scores of 10 and 20 (0-9 - non-depressed group, 10-19 - depressed group, 20-27 – severely depressed group). For the further pre-assessment, interviewers administered SCID-5 for depression severity confirmation because it is a well-established diagnostic assessment tool. The selected participants were required to fill in BDI-II and K-IDS tests online four times during the study with two weeks intervals.
As a result of the structured interview, we recruited 60 people (24 males, 36 females of average age = 25 years with a standard deviation of 6.1). Following the interview procedure described above, we identified three groups, namely non-depressed, depressed, and severely depressed, each comprising 20 participants. To verify the heterogeneity of the groups, we performed separate statistical tests for age and gender. We applied a chi-square test of independence to examine the relationship between gender and depression groups. The results ($X^2(2, N=60) = 0.41$, $p>0.05$) indicated no significant relationship between gender and depression group. The age characteristic is a continuous variable; and therefore, we used the Kruskal Wallis test, which resulted in $p>0.05$, indicating that the null hypothesis cannot be rejected and concluding that the group medians of age are equal for depression groups.
After being recruited, participants received the data collection agent "YouNoOne" and were instructed not to change the phone or uninstall the application. Data collection started only after a participant signed an informed consent form. The data collection duration was 60 days. We passively collected sensor data using the "YouNoOne" application. Table 1 summarizes sensor data collected with sampling rates and brief explanations. The collected data was securely transmitted to our cloud server via a TCP connection using SSL protocol [11]. The cloud server assigned random unique IDs for all participants to protect their identities.
The study procedure was approved by the Institutional Review Board of Yonsei University (No. 7001988-202103-HR-966-08)

Table 1. Collected data with sampling rates

Data source	Sampling rate	Details
Activity recognition	event based	timestamp of activity (still, walking, running, cycling, in a vehicle)
Applications usage	event based	timestamp and duration of used applications
Calendar	4 hours	number of calendar events
Calls	event based	timestamp and duration of missed, outgoing, incoming calls
GPS	5 minutes	latitude, longitude
Gravity sensor	15 minutes	x, y, z of phone position
Keystroke log	event based	timestamp and duration of typing, backspaces, autocorrection
Light sensor	15 minutes	ambient light value
Microphone	15 minutes	sound energy, pitch, MFCC
Notifications	event based	timestamp of notification, action (remove/click)
Pedometer	event based	timestamps of steps
Screen state	event based	timestamp of the screen on/off, lock/unlock
Significant motion sensor	event based	timestamp when sudden motion occurs
SMS	event based	timestamp and number of characters in an incoming message
Stored media	4 hours	number of images, videos, music files on a device
WiFi	30 minutes	SSID, BSSID

IV. Results

We performed data analysis after completing data collection. The data analysis pipeline starts from features engineering. We extracted more than 1500 features for different window sizes (around 300 features for each window size), namely a day, two days, three days, a week, two weeks. Our goal was to perform a comparative analysis of depression group classification accuracy for various window sizes. Each sensor has its own set of features, which we carefully selected based on related works. We calculated statistical characteristics such as mean, median, standard deviation, skewness, maximum, minimum, and contextual information (duration of homestay, phone usage duration at study places, etc.). The detailed list of features is out of the scope of this paper.

Features preprocessing pipeline consists of 3 steps: rare labels replacement, out-of-range values and outliers removal, missing data analysis. We treated labels that comprise less than 5% of the dataset as rare labels, which we combined into the “other” category. For example, we merged categories on a bicycle and in a vehicle into one. We treated duration features (call duration, moving duration, phone usage duration, etc.) that exceeded window size as out-of-range values and replaced them with missing data indicators. For outliers detection, we used 1% as a lower boundary and 99% as the upper limit, meaning that values falling out of borderline were considered outliers and regarded as missing data. Finally, we removed cases (rows of data) with more than 20% of missing values. We replaced the remaining missing values with the depression group median.

For the depression group classification, we decided to utilize tree-based ensemble algorithms, such as Random Forest (RF), AdaBoost, and XGBoost, because they do not require normal data distribution and are robust against overfitting. As hyperparameters, we used 100 decision trees in the RF classifier, 50 estimators for AdaBoost, and the maximum depth of 6 with a 0.3 learning rate for XGBoost. We split the dataset for train and test sets by 70% and 30%. We used GroupShuffleSplit from the scikit-learn library, where we treated each participant as a separate group. By doing so, we avoided having data from the same participant for training and testing sets. For model evaluation, we applied 10-fold cross-validation and calculated the accuracy as an average from the results. We repeated the experiment for all window sizes and summarized the results in figure 2. The highest achieved prediction accuracies are 68%, 73%, 77% for Random Forest, AdaBoost, and XGBoost, respectively. Figure 2 illustrates the generally positive prediction accuracy trend with the window size increase. Therefore, we conclude that by increasing the window size for features extraction, we can predict the depression group more accurately.

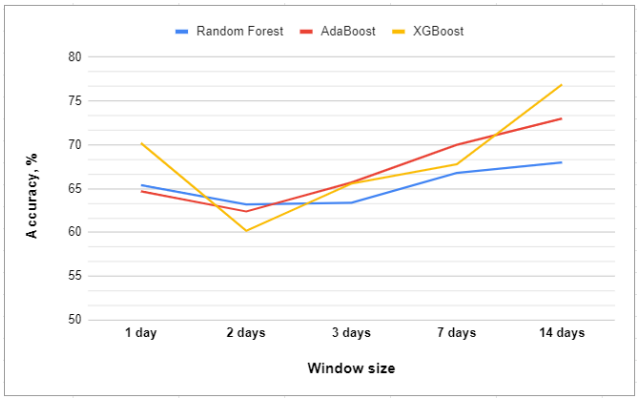


Figure 2. Depression group prediction accuracy for different window sizes

V. Conclusion

In this work, we investigated the influence of features extraction window size on depression group prediction. We collected sensor data from 60 participants from 3 depression groups (non-depressed, depressed, severely depressed) and extracted features from the data using different window sizes. Our results show that prediction accuracy generally increases with the increasing window size. In future work, we plan to experiment with more participants and more fine-grained window sizes.

ACKNOWLEDGMENT

This research was supported in part by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2019R1F1A1059898),

References

[1] World Health Organization. Depression 2020: https://www.who.int/news-room/fact sheets/detail/depression

[2] A. A. Farhan et al., "Behavior vs. introspection: refining prediction of clinical depression via smartphone sensing data," 2016 IEEE Wireless Health (WH), 2016, pp. 1-8, 10.1109/WH.2016.7764553.

[3] Rui Wang, Weichen Wang, Alex daSilva, Jeremy F. Huckins, William M. Kelley, Todd F. Heatherton, and Andrew T. Campbell. 2018. Tracking Depression Dynamics in College Students Using Mobile Phone and Wearable Sensing. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 2, 1, Article 43 (March 2018), 26 pages. https://doi.org/10.1145/3191775

[4] A. Ghandeharioun et al., "Objective assessment of depressive symptoms with machine learning and wearable sensors data," 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII), 2017, pp. 325-332, doi: 10.1109/ACII.2017.8273620.

[5] Saeb S, Zhang M, Karr CJ, Schueller SM, Corden ME, Kording KP, Mohr DC. Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. J Med Internet Res. 2015 Jul 15;17(7): e175. doi: 10.2196/jmir.4273. PMID: 26180009; PMCID: PMC4526997.

[6] Narziev, N.; Goh, H.; Toshnazarov, K.; Lee, S.A.; Chung, K.-M.; Noh, Y. STDD: Short-Term Depression Detection with Passive Sensing. Sensors 2020, 20, 1396. https://doi.org/10.3390/s20051396

[7] Kroenke, K.; Spitzer, R.L.; Williams, J.B. The PHQ-9: Validity of a brief depression severity measure. J. Gen. Intern. Med. 2001, 16, 606–613.

[8] Upton, J. Beck Depression Inventory (BDI). In Encyclopedia of Behavioral Medicine; Gellman, M.D., Turner, J.R., Eds.; Springer: New York, NY, USA, 2013; pp. 178–179.

[9] Journal of Korean Society for Depressive and Bipolar Disorders (우울・조울병), Vol.10(3): 131-151, 2012-10

[10] Qualtrics. Available online: https://www.qualtrics.com/

[11] Muhammad Salman, Touseef Javed Chaudhery, and Youngtae Noh. "Study on performance of AQM schemes over TCP variants in different network environments." IET Communications (2020).

데이터 상호 운용성 제공을 위한 딥 러닝 활용 시맨틱 정렬 적용 평가: 웹 오브젝트 기반 우울증 온토롤지 사례

정일영*, 강현국**
*DL 정보기술, **고려대학교

*iychong54@gmail.com, **kahng@korea.ac.kr

Deep Learning Utilized Semantic Alignment Evaluation for Data Interoperability Provisioning: Use Case of Web Objects based Depressive Disorder Ontologies

*Ilyoung Chong, **Hyun Kook Kahng
*DL Information Technology, **Korea Univ.

Abstract

This paper focuses the evaluation mechanism to support the semantic interoperability provisioning in healthcare applications with the experiments of emotion and depressive disorder symptom data. We develop the ontology model of depression states in WoO framework to enrich depressive disorder symptoms data with the semantics. The paper delivers a proof of concept to carry out the semantic interoperability evaluation of healthcare application of depressive disorder using hybrid deep learning for base ontology model and shared data representations.

I. Introduction

In today's healthcare environments, depressive disorder assistance is one of the important issue. To support depressive disorder (DD) healthcare service, its symptoms can be monitored with IoT enabled, wearable sensor devices and questionnaire activities. Web Objects based ontology model simplifies object representations and provides effective way to utilize virtualization environment shown in [3] and [4]. In the use case of Web of Objects(WoO) platform [4], a VO has been linked to an information resource, and a CVO provides the rules to represent a functional feature aggregated by multiple VOs.

Preparing the data for the analytics in the DD healthcare applications it has been semantically aligned through mediation procedure first. This involves aligning the different data models from heterogeneous sources to be mapped to the base ontology model following a shared vocabulary. The deep learning mechanism has been incorporated as a process to perform analytic procedures on the data aggregated by the ontologies composed by VOs and CVOs. In order to evaluate the semantic interoperability provisioning for DD healthcare applications has been designed. To realize the DD conditions, the ontologies of affective human health conditions have been implemented. To support the integration and semantic harmonization of the data models a base ontology, and shared data model and vocabulary havebeen applied.

This is to evaluate the ontology models in different settings considering the nature of time series data. In the developed model of CNN with LSTM, an extra layer has been cascaded with CNN model. In this setting the fully connected dense layer output is fedin the cell of LSTM. The models are trained on the shared data with integration of the data from three well known datasets.

II. Semantic Ontology Alignment Model on Deep Representation Learning Process

To deploy the automatic semantic ontology alignment model, a base ontology catalog has been setup to provide the repositories and data processing and management mechanism based on Web Objects framework [1][2][3]. The learning mechanism has been designed in microservice process. Depending the semantic matches entries are stored in the base ontology instances. These instances collectively form an interoperable shared data model to provide semantic coherence of multiple data models. The representation of the ontology graphs is developed using deep learning method to provide the semantic alignment. The alignment process is performed using the semantic association algorithm based on the ontology graph structure. In the last step, it combines the results of both alignments and perform semantic interoperability checking and then provide the final semantic interoperability provisioning as shown in Figure 1.

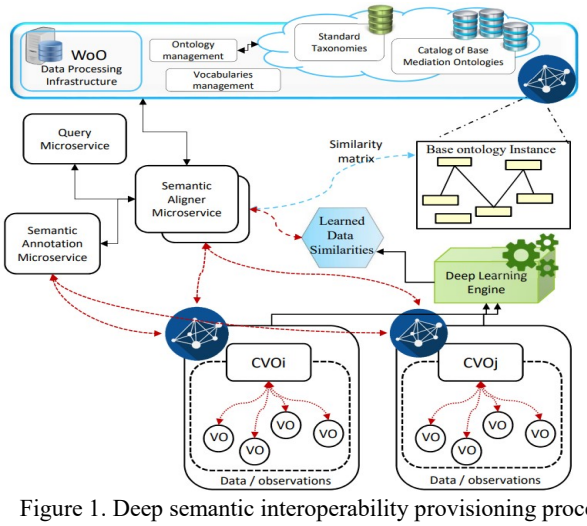


Figure 1. Deep semantic interoperability provisioning process

It is important to note that all the entity names and profile vectors are combined to form the integrated embedding. As the input to the three sub-networks, the embedding is executed based on the

mutual parameters for semantic similarity predictions. The networks are designed using three layers which comprise CNN along with the ReLU attachments. More importantly the resultant output of the sub-networks is combined together along with the features computed earlier. The network probabilities for source and target entities shown in Figure 2 have been defined as follows:

$$(1)\quad Z_{source} = Act_{ReLU}(\, Act_{ReLU}(Act_{ReLU}(V_{source}, theta_1); theta_2); theta_3)$$

$$(2)\quad Z_{target} = Act_{ReLU}(\, Act_{ReLU}(Act_{ReLU}(V_{source}, theta_1); theta_2); theta_3)$$

$$(3)\quad P(Z_{source}, Z_{target}) = \mathrm{SIGM}(\, Act_{ReLU}(Z_{source}, Z_{target}); theta_4); theta_5)$$

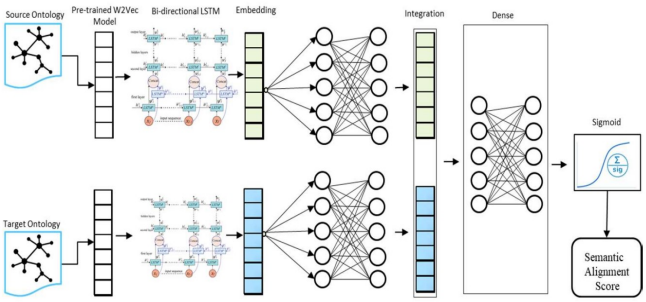


Figure 2. Deep semantic alignment architecture with enhanced embedding

In the developed model of CNN LSTM, an extra layer has been cascaded with CNN model. The model configuration has been kept same with addition of layer of the LSTM. In the Hybrid LSTM shown in Figure 3, cell parts learns the resultant output by updating their memory depending on their input and past history of the states. The model developed with 3 learning layers, has been trained on the data from the shared data representation model sharing the modalities of the different sensors.

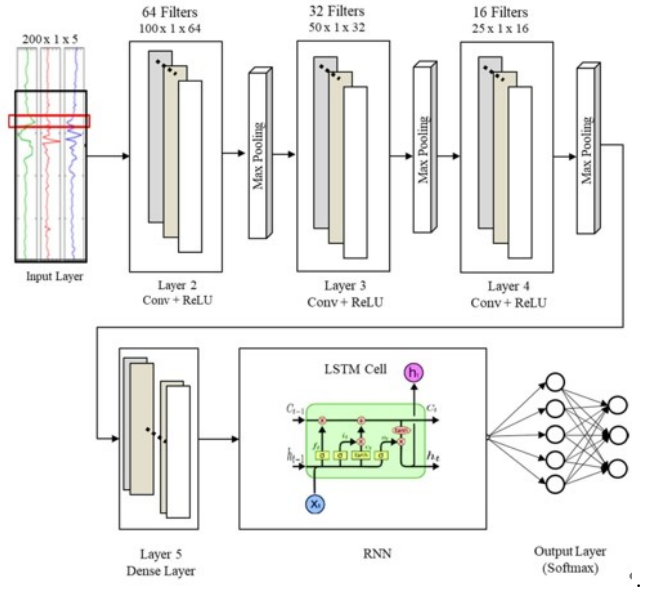


Figure 3. Hybrid LSTM based learning network

III. Experimental Results and Conclusion

The experiment on DEAP dataset is introduced to show the performance of DEAP dataset in terms of 2 learning based CNN model and 3 learning based Hybrid LSTM model. DEAP data sets based on sensor data as explained earlier. We have compared the accuracy achieved of our developed models (CNN and Hybrid model) with recent state of the art methods who utilize the AMIGOS, DEAP and ACGM dataset [8][9] shown in Table 1.

Moreover, in order to implement semantic interoperability features in DD healthcare application, Web Objects reference architecture [4] has been implemented, and the accuracy of the Hybrid LSTM deep learning model applied in this paper to provide semantic interoperability in DD healthcare applications was found to be relatively good as shown in Table 1 and figure 4. It has been also shown that the semantic data representation based on Web Objects contributes to achieve a better accuracy to support semantic interoperability provisioning. The Web Objects (VOs, CVOs and Microservice) to represent semantic ontology models should be well fitted with required performance enhancement of DD healthcare applications.

Based on this analysis, further development to support practical applications environment will be necessary to enhance semantic interoperability provisioning capability in real DD healthcare application environment.

Table 1. State of the art approaches comparison of accuracy measure for AMIGOS, DEAP and ACGM datasets

	AMIGOS	DEAP	ACGM
CNN	86.92 %	88.23 %	92.56 %
Hybrid(CNN+LSTM)	90.66 %	91.24 %	94.56 %



Figure 4. Achieved Accuracy for Training and Testing with respect to DEAP, AMIGOS Dataset

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MIST) (No. 2019R1F1A106372013)

참 고 문 헌

[1] M. G. Kibria, S. Ali, M. A. Jarwar, and Ilyoung Chong, “A framework to support data interoperability in web objects based IoT environments,” in 2017 International Conference on Information and Communication Technology Convergence (ICTC), Oct. 2017

[2] Sajjad Ali and Ilyoung Chong, “Semantic Mediation Model to Promote Improved Data Sharing Using Representation Learning in Heterogeneous Healthcare Service Environments,” MDPI applied science, pp. 1-30

[3] S. Ali, M. G. Kibria, and Ilyoung. Chong, “WoO enabled IoT service provisioning based on learning user preferences and situation,” in 2017 International Conference on Information Networking (ICOIN), Feb. 2017.

[4] ITU-T Y.4452: Functional Framework of Web of Objects. Available online: http://www.itu.int/rec/T-REC-Y.4452-201609-P

[5] Kim, A.Y.; Jang, E.H.; Kim, S.; Choi, K.W.; Jeon, H.J.; Yu, H.Y.; Byun, S. Automatic detection of major depressive disorder using electrodermal activity. *Sci. Rep.* **2018**, *8*, 17030

[6] Garcia-Ceja, E.; Riegler, M.; Nordgreen, T.; Jakobsen, P.; Oedegaard, K.J.; Tørresen, J. Mental health monitoring with multimodal sensing and machine learning: A survey. Pervasive Mob. Comput. 2018, 51, 1–26

[7] Noura, M.; Atiquzzaman, M.; Gaedke, M. Interoperability in Internet of Things: Taxonomies and Open Challenges. *Mob. Netw. Appl.* **2019**, *24*, 796–809.

[8] A. Miranda-Correa, M. K. Abadi, N. Sebe, and I. Patras, "AMIGOS: A Dataset for Affect, Personality and Mood Research on Individuals and Groups," Feb. 2017.

[9] S. Koelstra et al., "DEAP: A Database for Emotion Analysis using Physiological Signals." EEE Transactions on Affective Computing, vol. 3, no. 1, pp. 18-31, 2012

제2회 한국 인공지능 학술대회

The 2nd Korea Artificial Intelligence Conference

세션 D

그래프 신경망을 이용한 관망에서의 신호 소스 위치 예측 방법

고석갑, 김에덴, 이병탁
한국전자통신연구원
{softgear, kimed93, bytelee}@etri.re.kr

A Method for Signal Source Location Prediction in Pipe Networks Using Graph Neural Networks

SeokKap Ko, Eden Kim, and Byung-Tak Lee
Electronics Telecommunications Research Institute (ETRI)

요 약

본 논문은, 관망(pipe network)에 특이 신호가 발생하였을 때, 관망에 설치된 센서들로부터 수집된 데이터를 이용하여 그 신호 소스의 위치를 예측하는 방법을 제안한다. 센서의 설치 위치와 관망의 구조정보를 이용하여 구성한 그래프 정보와 센서에서 측정된 값을 입력으로 하고, 각 센서로부터 신호 소스와의 거리를 출력으로 하는, 그래프 신경망을 구성 및 학습한 뒤, 신호 소스의 위치를 예측하는 방법을 제시한다. 시뮬레이션을 통하여 제안된 그래프 신경망을 이용한 신호 소스 위치 예측 모델의 유용성을 검증하였다.

Ⅰ. 서 론

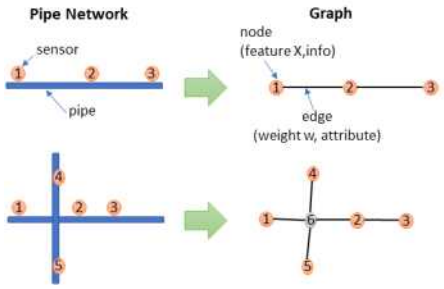
우리 주변에는 상수도 공급망, 하수도 망, 가스 공급망, 석유 공급망 등 복잡한 파이프로 구성된 다양한 관망이 구축되어 있다. 최근 사물인터넷 기술이 도입되면서 관망의 상태를 모니터링하기 위한 센서들을 관망의 곳곳에 설치하고 있다. 부식 및 노후, 공사, 사고, 지각변동, 기상변화 등으로 인해 파이프의 특정 지점이 손상될 수 있다. 이 경우 손상 지점에서 특이 신호가 발생하며 이 지점을 신호 소스라고 한다. 신호 소스를 찾는 기존의 연구는 물리적 특성과 이론에 근거한 모델 기반 접근 방법이었다[1][2]. 비교적 단순한 구조에서는 모델 기반 접근 방법이 좋은 성능을 내지만, 복잡한 실제 환경에서는 성능저하가 발생할 수 있다. 복잡한 관망 구조 및 신호 전파 특성, 저전력 센서의 데이터 수집능력 한계 등으로 인하여 관망에 설치된 센서들의 데이터로부터 신호 소스 위치를 예측하는 문제는 어렵다.

최근 기계학습 기반의 인공지능 기법이 다양한 분야에 적용되고 있다. 기계학습 모델에는, 일반적인 분류 및 회귀분석에 사용되는 다계층 신경망 및 이미지 분류 문제에 사용되는 콘볼루션 네트워크, 시계열 예측에 사용되는 LSTM(Long Short Term Memory) 등이 있다. 기계학습 모델 중 그래프 신경망은 노드의 특징값에 추가로 노드의 연결 관계를 함께 학습할 수 있다. 그래프 신경망 중 그래프 콘볼루션 네트워크는 그래프상 인접 노드와의 관계를 동일 가중치를 이용하여 학습하여 예측 성능을 향상 시킬 수 있다[3]. 본 논문에서는 관망의 구조와 관망 상에서의 센서의 위치를 그래프로 구성하여 그래프 콘볼루션 네트워크를 이용한다.

본 논문에서는 관망을 그래프로 변환하는 방법, 그래프 신경망을 이용하여 센서로부터 신호 소스 위치를 예측하는 방법에 대해 설명한다. 시뮬레이션을 통하여, 그래프 신경망을 이용한 신호 소스 예측 모델의 유용성을 검증하였다.

Ⅱ. 본론

(그래프구성방법) 그래프는 노드와 엣지로 구성된다. 노드는 특성 데이터를 가지고 있는 지점이다. 엣지는 노드와 노드 사이를 연결하는 선이며 연결 강도를 나타내는 가중치 및 속성을 가질 수 있다. 관망 구조 및 센서의 위치를 그래프로 구성하는 방법은 그림1과 같다. 한 파이프 상에 연속



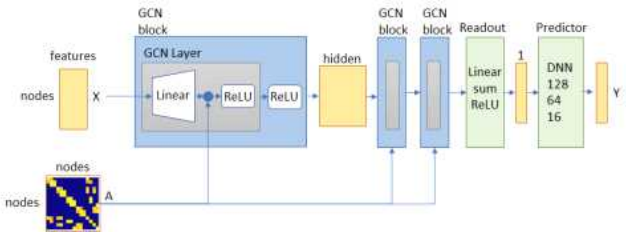
(그림1) 관망-그래프 구성 방법

적으로 센서가 배치된 경우, 센서를 노드로 하고 그 사이를 엣지로 연결한다. 엣지의 가중치는 노드 사이의 거리 및 파이프 속성(관종, 직경 등)을 고려하여 적용한다. 이 가중치는 나중에 그래프 신경망에 입력될 때 각 노드의 특성(feature) 데이터와 결합된다. 만약 파이프가 교차하는 경우, 교차 지점을 가상 노드로 만들고 교차지점 주변에 가까운 센서를 노드로 하고 엣지로 연결한다. 각 센서 노드에는 측정된 신호 데이터를 담게 된다. 교차점의 가상 노드는 측정 데이터가 없으므로, 관망에 따른 전파 함수(propagation function)을 이용하여 주변의 센서 노드로부터의 데이터를 취합하여야 한다. 본 논문에서는, 문제를 일반화 및 단순화 하기 위하여, 거리에 반비례하는 단순한 전파 함수를 이용하였다. 가상 노드의 특성 데이터는 다음 수식(1)과 같이 각 노드의 특성값(X)와 엣지의 가중치값(w)를 곱한 값의 가중치 평균으로 계산된다.

$$X_{virtual} = \frac{1}{\sum_{neighbors} w_i} \sum_{neighbors} w_i X_i \quad \text{....} \quad \text{수식(1)}$$

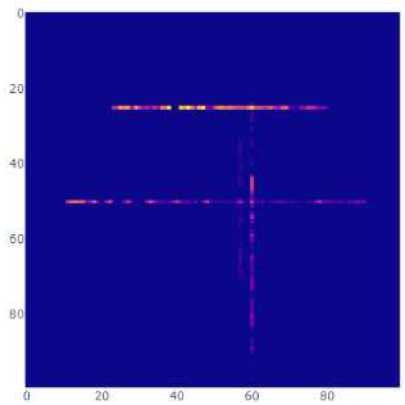
(그래프콘볼루션네트워크적용) 이미지 분류에 널리 사용되는 콘볼루션

네트워크는 인접된 점들의 값에 학습 가능한 가중치를 곱하여 새로운 값을 만들어 내고 이를 반복하여 분류 또는 회귀 문제를 해결한다. 마찬가지로 그래프 콘볼루션 네트워크는 엣지로 연결된 인접된 노드들의 값에 학습 가능한 가중치를 곱하는 과정을 통하여 분류 또는 회귀 문제를 해결한다. 그래프 콘볼루션의 입력은 각 노드의 특성값 X와 인접행렬 A를 취한다. 본 논문에서 그래프 콘볼루션의 출력으로는 각 센서로부터 신호 소스까지의 거리를 이용하였다. 그림2는 본 논문에서 사용한 그래프 콘볼루션 네트워크의 구조를 보인다. 콘볼루션 네트워크에서와 마찬가지로 그래프 콘볼루션 네트워크에서도 Gated Skip Connection 기법을 사용할 수 있다. 이는 그래프 콘볼루션 블록의 결과와 입력에 가중치를 주어 취합하는 방법이다[4].



(그림2) 그래프 콘볼루션 네트워크 구조

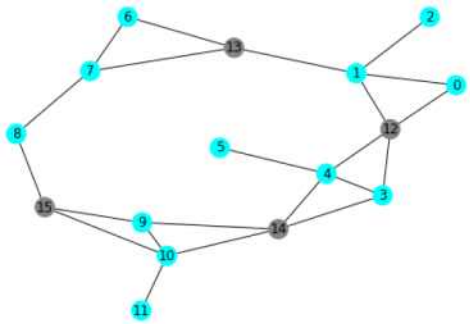
(신호소스위치예측) 그래프 콘볼루션 네트워크의 출력은 각 센서 노드로부터 신호 소스 위치까지의 거리이다. 정확한 신호 소스 위치를 예측하기 위해서는 각 센서 노드로부터의 거리가 중첩되는 부분을 찾아야 한다. 예측 모델에는 오차가 있기 때문에, 이를 고려하여 신호 소스 위치에 대한 확률을 누적하여 전체 관망에서 확률이 가장 높은 부분을 신호 소스 위치로 예측한다. 모델의 예측 결과 평가를 통해 오차의 평균값과 표준편차를 구한다. 이를 가우시안 분포 함수에 넣어 거리에 따른 확률밀도 함수를 구하였다. 그림3은 신호 소스 위치에 대한 확률 분포를 보인다. 여기서 확률이 가장 높은 지점의 위치를 신호 소스 위치로 판별한다.



(그림3) 신호 소스 위치에 대한 확률 분포

(관망신호소스위치시뮬레이션) 제안하는 그래프 신경망을 이용한 신호 소스 예측 모델를 평가하기 위하여, 시뮬레이션을 이용하였다. 시뮬레이션은 100mx100m 평면에서 4개의 직선관으로 구성된 관망을 임의로 생성한다. 파이프들은 서로 3지점 또는 4지점에서 서로 연결된다. 4 지점에서 연결된 경우, 루프를 형성하게 된다. 각 파이프 별로 3개의 센서가 임의의 위치에 배치된다. 그리고 관망상의 임의의 위치에 신호 소스가 위치한다. 신호 소스에서의 신호 출력 세기는 임의로 결정된다. 신호 소스로부터의 신호는 파이프를 이동하며 거리에 따라 감쇄한다. 분기점에서는 분기수에 맞게 감쇄한다. 각 센서들은 각 지점에서의 신호의 세기를 측정한다. 그림4는 한 시뮬레이션 시나리오에서의 관망 구조 정보를 토대로 그래프를 구성한 예시를 보인다. 파란색 노드는 센서 노드를 나타내며, 회색 노드는 파이프 연결 지점의 가상 노드를 나타낸다.

이와 같은 시뮬레이션 환경에서 10만 개의 시나리오 데이터셋을 구축하였다. 8만 개는 학습용으로, 1만 개는 검증용으로, 1만 개는 평가용으로 구성하였다. 다음 표는 실험결과를 나타낸다. 표의 값은 예측한 신호 소스 위치와 실제 신호 소스 위치와의 거리를 나타낸다. 신호 소스 위치 후보 상위 1개(first)의 평균 예측 오차는 13m이며, 오차 상위 3개 이내(within-third)의 평균 예측 오차는 9m였다.



(그림4) 관망 구조 정보를 그래프로 구성한 예시

prediction	mean	std	min	max
first	**13.9**	11.2	0.0	73.0
second	14.3	11.4	0.0	75.0
third	14.4	11,7	0.0	77.0
within-third	**9.7**	9.3	0.0	66.0

표 1. 시뮬레이션을 통한 평가결과(신호 소스 위치 오차 거리)

Ⅲ. 결론

본 논문에서는 그래프 신경망(그래프 콘볼루션 네트워크)를 이용한 관망에서의 신호 소스 위치 예측 방법을 제안하였다. 일반적인 기계학습과 달리 그래프 신경망은 기본적인 데이터 이외에 추가적인 형상(topology) 정보를 같이 입력할 수 있기 때문에, 관망에서의 신호 소스 위치 예측이 가능하다. 본 기술은, 관망 상의 누수 위치 탐지, 파손 위치 예측, 공사 지점 예측, 충격 지점 예측, 고장 위치 예측 등에 활용될 수 있을 것이다.

추가적으로, 다른 기계학습 모델과의 성능 비교 및 그래프 구성시 엣지의 가중치 설정 방법 개선, 그래프 콘볼루션 네트워크 내에서 인접행렬과 특성행렬의 단순곱이 아닌 다차원 함수 적용, 물리 모델과의 결합, 신호 소스 위치를 직접 예측하는 모델 개발 등의 연구를 통한 성능 개선이 필요하다.

ACKNOWLEDGMENT

본 연구는 산업통상자원부(MOTIE)와 한국에너지기술평가원(KETEP)의 지원을 받아 수행한 연구 과제입니다. (No. 20181210301570)

참 고 문 헌

[1] Yazdekhasti, S., Piratla, K.R., Matthews, J.C., Khan, A. and Atamturktur, S. (2018), "Optimal selection of acoustic leak detection techniques for water pipelines using multi-criteria decision analysis", Management of Environmental Quality, Vol. 29 No. 2, pp.

255-277. https://doi.org/10.1108/MEQ-05-2017-0043

[2] Ling Ling Ting, Jing Yuen Tey, Andy Chit Tan, Yeong Jin King, Faidz Abd Rahman, “Water leak location based on improved dual-tree complex wavelet transform with soft thresholding de-noising”, Applied Acoustics, Volume 174, 021, 107751, ISSN 0003-682X,https://doi.org/10.1016/j.apacoust.2020.107751.

[3] Zhang, S., Tong, H., Xu, J. et al. “Graph convolutional networks: a comprehensive review”. Comput Soc Netw 6, 11 (2019). https://doi.org/10.1186/s40649-019-0069-y

[4] Seongok Ryu, et al., “Deeply learning molecular structure-property relationships using attention- and gate-augmented graph convolutional network”, arXiv:1805.10988, (2018).

시각장애인 길 안내 서비스를 위한 Image Segmentation 기반 경로 보정 모듈 개발

정홍주, 최인훈, 홍승준, 김서현, 박수용, 허의남*
경희대학교
{sub06038, inhun321, hongsj1022, seok02h, clapd10, johnHuh}@khu.ac.kr

Image Segmentation-Based Route Correction Module for Road Guidance Service for Visually Impaired

Hong-Ju Jeong, Inhun Choi, Seungjun Hong,
Seo-Hyun Kim, Suyong Bahk, Eui-Nam Huh*
Kyung Hee Univ.

요 약

현재 상용되고 있는 보행자용 길 찾기 오픈 API는 때때로 차도로 보행자를 안내한다는 한계점이 있는데, 이는 시각장애인 길 안내 서비스를 제공하는 자율주행 로봇을 이용하는 사용자에게 위험하다. 따라서 본 논문은 이를 보완하고자 Depth Camera의 Image 정보와 Depth 정보로 보행자 도로 및 거리를 인식하고 GPS로 길을 찾아가는 Image Segmentation 기반 경로 보정 모듈을 제시하였다.

1. 서 론

여러 플랫폼에서 제공하는 보행자용 길 찾기 오픈 API를 사용하면 공통점으로 나타나는 문제가 있다. API를 제공하는 플랫폼이 지도에서 차도와 인도를 제대로 구분하지 않는 경우 보행자용 길 찾기 기능을 수행하였음에도 불구하고 보도가 아닌 차도로 GPS POINT(위도, 경도)가 설정된다. 이는 주로 아파트 단지나 학교 캠퍼스 등 집합건물이나 보도가 제대로 구성되어있지 않은 구역에서 나타난다. 이는 보행자 도로를 따라 자율주행을 하는 소·중형 로봇이 Map을 가지지 않고 목적지를 찾아갈 때 위험요소로 작용한다. 특히 시각장애인을 위한 자율주행 AI 길 안내 로봇 서비스를 제공할 때 치명적이다.

따라서 본 논문은 자율주행 로봇이 보행자용 길 찾기 오픈 API가 제공하는 GPS POINT를 기반으로 길 찾기를 수행하되, 보행자 도로를 인식하여 가능한 보도로 주행할 수 있게 하는 Image Segmentation 기반 자율주행 로봇 경로 보정 모듈을 제안한다.

2. 본론

본 논문에서는 Raspberry Pi 4B와 stm32f4discovery 보드를 탑재한 카트형 로봇을 활용한 모듈개발을 진행하였다. 뎁스카메라(Intel RealSense Depth Camera D435i)로 전방의 이미지 데이터와 거리 데이터를 수집하고 GPS-RTK2 보드(SparkFun GPS-RTK2-ZED-F9P)를 사용하여 GPS 데이터를 수집한다. 보행자용 길 찾기 Open API(TMAP API)로 목적지까지 경로를 설정한다. Raspberry Pi 4는 이미지 데이터와 거리 데이터를 분석하고 목적지까지 경로 데이터와 현재 GPS 정보를 종합하여 UART 통신으로 stm보드에 신호를 전달한다. stm보드는 모터 제어를 담당하여 디바이스를 보행자 도로로 움직이게 한다. AHRS/GPS 복합 네비게이션 센서 모듈(WTGAHRS1/UART)는 IMU 센서 역할을 하여 기기의 방향을 알 수 있다. 이를 통해 GPS POINT와의 각도를 알 수 있고 보행자 도로 유도용 POINT와 차량의 이동 경로를 결정할 수 있다.

2-1. 보행자도로 인식

본 논문에서는 AI Hub의 '인도(人道) 보행 영상'을 데이터셋으로 사용하였으며, 그 중 노면 데이터셋인 'Surface Masking 데이터셋'을 사용하였다. 이 데이터셋은 노면 특성(재질, 특수성, 파손여부 등)에 따라 20종으로 구분된 이미지가 51,318장이 포함되어있다[1].

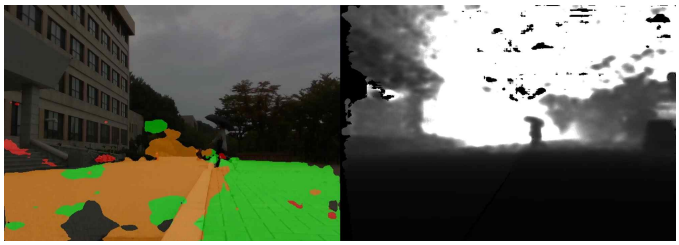
Surface Class	attribute
sidewalk	@blocks, @soil_stone, @cement, @damaged, @urethane, @asphalt, @other
braille_guiding_block	@normal, @damaged
roadway	@normal, @crosswalk
alley	@normal, @speed_bump, @crosswalk, @damaged
bike_lane	
caution_zone	@stairs, @grating, @manhole, @repair_zone, @tree_zone

New Class	Label	RGB Color
background	0	[0, 0, 0]
crosswalk	1	[255, 0, 255]
roadway	2	[255, 128, 0]
sidewalk	3	[0, 255, 0]

[그림 1] class별 RGB Color

[그림 1]은 본 논문이 보도와 도로 구분을 목적으로 하고 있다는 점에서 20종으로 구분된 데이터셋을 3종으로 통합했음을 나타내고 있다. 점자블록과 기타 보행자 도로 위 클래스들을 모두 sidewalk로 통합했으며

차도와 보도를 판단하기 위해 DeepLab v3+ 모델 기반의 segmentation-selectstar를 사용하였다[2]. 자율주행 로봇은 image segmentation으로 분류된 class의 정보를 통해 기기가 보도와 차도 중 어디에 있는지 알 수 있으며 결과는 [그림 2]와 같다.

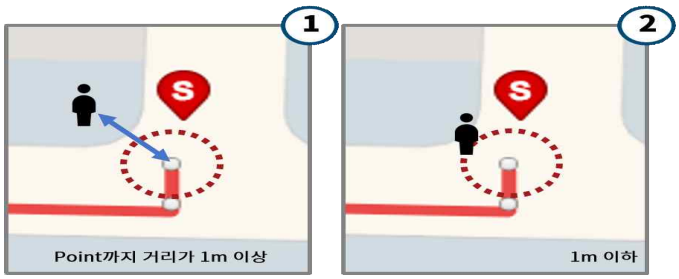


[그림 2] 이미지 데이터와 거리 데이터

[그림 2]의 왼쪽은 뎁스카메라로 촬영한 이미지에 학습한 모델로 예측한 결과이고 오른쪽은 이미지를 촬영할 당시 거리 데이터이다. 거리 데이터는 가정한 뎁스카메라 기준으로 약 10m까지 측정 가능하다.

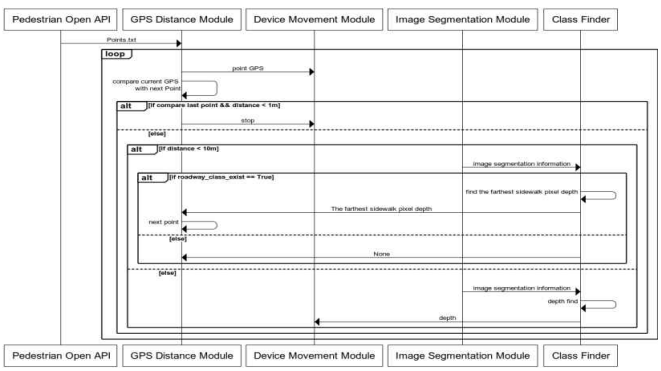
2-2. GPS POINT 설정

본 논문에서는 목적지까지 경로를 1차적으로 보행자용 길 찾기 Open API로 설정한다. 경로 상의 각 POINT들을 위도와 경도로 받아와 자율주행 로봇의 GPS 위치와 비교한다. 자율주행 로봇의 위치는 GPS 안테나를 통해 받아오고 GPS-RTK로 보정한다. GNSS 관측소의 실시간 데이터인 RTCM을 받아오기 위해 ROS(The Robot Operating System)에서 NTRIP 패키지를 사용하였다.



[그림 3] GPS POINT와 디바이스의 좌표 비교

자율주행 로봇의 10m 이내의 직선거리에 보행자 도로가 아닌 클래스가 포함되어 있다면 보행자 도로 끝 좌표에 POINT를 생성하고 향하게 하여 로봇이 차도를 거치지 않게 한다. [그림 3]에서는 버스정류장 근처 보도에서 도서관까지 보행자용 길 찾기 Open API를 실행하여 설정된 POINT들과 그에 따른 안내선을 볼 수 있다. 기존에는 POINT의 반경 1m 이내에 접근 했을 때 다음 POINT로 경로를 안내하는 방식으로 자율주행 로봇이 동작했다.



[그림 4] Image Segmentation 기반 경로 보정 모듈 시퀀스 다이어그램

[그림 4]는 Image Segmentation 기반 경로 보정 모듈의 시퀀스 다이어그램을 나타낸다.. 먼저 보행자용 길 찾기 Open API에서 제공한 첫 번째 POINT까지 디바이스가 자율주행을 해서 찾아간다. 그 후 GPS간 거리비교와 자율주행 로봇의 직선거리의 클래스 정보를 합하여 POINT가 보행자도로에 없음을 인지한다. 없다면 다음 포인트와 거리비교를 하여 10m 이상이면 IMU 센서를 이용하여 방향 및 거리 비교를 진행하여 두 번째 POINT로 이동하게 한다. 10m 이하이면 보행자 도로에 위치해 있는지 확인하고 없다면 다시 다음 POINT 방향으로 이동한다. 도착 지점이 차도로 확인되면 보행자도로 위에서 종료한다.

3. 결론 및 향후 연구방향

본 논문에서는 보행자용 길 안내 Open API에서 제공하는 목적지까지 경로에 따라 자율주행 로봇이 보도로 이동하도록 유도하는 기능을 제안하였다. 이 모듈로 집합건물을 포함한 보행자도로가 제대로 표기되지 않은 지역에서도 시각장애인을 위한 자율주행 로봇이 보행자 도로로 안전하게 길 안내를 할 수 있다. 본 논문에서 활용한 디바이스는 3 FPS의 성능을 나타내며 이를 포함하여 안정성을 향상시키고자 다음과 같은 향후 연구를 진행하고자 한다.

시각장애인의 안전을 더 보장하기 위해서는 보행자 도로가 존재하지 않아 인식이 불가능한 경우, 보행자 도로가 있는 곳으로 다른 경로를 지정해주는 기능을 추가해야 한다.

성능 개선을 위해서는 엣지 컴퓨팅으로 성능이 좋은 엣지 클라우드에서 예측단계를 진행할 수 있고, 주기적으로 지나온 경로를 엣지 클라우드에 업데이트할 수 있다.

엣지 클라우드에 학습 모델을 두고 디바이스에서 이미지와 거리 정보를 전송시켜 예측할 수 있다. 이 경우 본문에서 활용한 뎁스카메라(Intel RealSense Camera) 기준 거리 데이터만 640x360 해상도에 30FPS로 받아오기 위해서는 최소 네트워크 대역폭이 111Mbps 이상 필요하다[3]. 이 점을 고려했을 때 일반 네트워크는 부족하다고 판단한다. 그래서 5G 네트워크망이 활성화되면 디바이스의 성능이 부족해도 엣지 클라우드의 모델로 보행자 도로를 예측하여 반환하는 엣지 컴퓨팅이 가능할 것으로 예상한다. 또한 주기적으로 자율주행 로봇의 위치가 어떤 클래스에 속하는지 수집하여 한번에 엣지 클라우드에 전송하여 이전에 한번 이상 지났던 길이라면 엣지에서 데이터를 받아와 더 쉽게 보행자 도로로 유도가 가능할 것으로 예상한다.

ACKNOWLEDGMENT

이 논문은 2021 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2017-0-00294. (대학 ICT 기초연구실) 서비스 이동 지원을 위한 분산형 클라우드 핵심원천 기술 연구)

참 고 문 헌

[1] “인도(人道) 보행 영상”, AIhub, (주)테스트웍스, 한국지능정보사회진흥원, https://aihub.or.kr/aidata/136

[2] selectstarofficial. 2019. segmentationn-selectstar. https://github.com/selectstarofficial/segmentation-selectstar. (2019).

[3] “Intel® RealSense™ Depth Camera over Ethernet.” Intel Realsense Documentation. last modified 2020, accessed Aug 26, 2021, https://dev.intelrealsense.com/docs/depth-camera-over-ethernet-whitepaper

학습기반 측위 DB 를 이용한 미수집지점에서의 측위 인프라 유사도 분석

지명인, 조영수
한국전자통신연구원 지능화융합연구소 도시·공간 ICT 연구실
{myungin, choys}@etri.re.kr

Analysis of positioning infrastructure similarity at non-collection points using learning-based positioning DB

Myungin Ji, Youngsu Cho
Electronics and Telecommunications Research Institute (ETRI)

요 약

LTE 기반 측위는 위성신호의 수신이 어려운 실내 또는 지하 공간에서도 측위가 가능하므로 스마트폰에서 사용가능한 측위 자원 중 가장 가용성이 높다. 하지만, 넓은 실내, 사유지 등 차량수집 음영구간의 경우 측위 DB 생성이 어려운 문제가 있어 주변 수집정보를 이용한 추정이 필수적이다. 실데이터 학습을 통해 실제와 유사한 가상데이터를 생성하는 생성적 적대 신경망[1]은 일부 수집데이터를 이용한 전역 측위 DB 생성에 있어 좋은 해법이 될 수 있으며, 본 논문에서는 이를 활용하여 생성한 위치인식 용 DB 의 효용성을 미수집지점에서 측정한 정보와 가상으로 생성한 DB 정보 간의 유사도를 분석하여 제시한다.

Ⅰ. 서 론

최근 다양한 통신 인프라의 발전과 다량의 데이터 축적을 통해 언제 어디서나 정확한 위치정보를 제공하는 기술이 활발하게 개발되고 있다.

실외 환경에서는 주로 GPS 신호를 기반으로 하여 단말에 위치정보를 제공하지만 실내 환경에서는 LTE, Wi-Fi, BLE 등의 무선통신 신호를 활용하여 위치를 추정하며, 유동인구가 많은 큰 쇼핑몰 같은 경우에는 도보를 이용한 세부수집을[2], 이외 지역에는 차량을 이용한 건물주변 수집을 일반적으로 진행한다.

차량수집의 경우 차량에 GPS, LTE, Wi-Fi, BLE 등의 모듈을 설치한 후 차량이 이동할 수 있는 도로 상에서 수집한 무선통신 인프라 정보에 경위도 좌표 정보를 태깅하며 수집을 진행하는데, 넓은 지역에 대해 빠른 수집이 가능한 장점과 차량진입이 어려운 곳에는 수집이 이뤄지지 않는 단점이 존재한다.

이에 따라 [3]에서는 생성적 적대 신경망(Generative Adversarial Network, GAN) 방법을 활용하여 수집지역뿐만 아니라 미수집지역에 대해서도 LTE 측위 DB 를 생성하는 방법을 제안하였다. 본 논문에서는 생성한 측위 DB 가 실환경을 잘 모사하는 지 검증하기위해 여러 미수집지역에서 상용 스마트폰으로 수집한 측위 인프라 측정정보를 생성된 측위 DB 에 저장된 정보와 비교한다.

본 논문의 구성은 다음과 같다. 2 장에서는 생성된 측위 DB 를 수집데이터 만을 이용한 전파지도 DB 와 비교하고, 3 장에서는 미수집지점에서의 수집한 데이터에 대해 각 DB 와의 유사도를 평가하며 4 장에서는 결론을 맺는다.

Ⅱ. 생성적 적대 신경망 개념을 활용한 측위 DB[3]

LTE 측위 DB 생성에 생성적 적대 신경망 개념을 활용하기 위해, 먼저 차량수집 데이터를 판별자(Discriminator)가 학습하도록 하였다. 학습에 사용하는 데이터는 수집위치 정보와 해당 수집위치에서의 무선통신 인프라 정보 및 수신 신호세기 정보이다.

생성적 적대 신경망에서 생성자(Generator)는 가상의 무선통신 신호세기 정보를 생성하여 판별자가 갖는 정보와 유사한 전파지도를 만들어 내는데, 이를 위해 ① 각 기지국 위치를 임의로 설정, ② 기지국 위치로부터 각 수집지점과의 거리와 수신된 신호세기를 이용하여 전파모델 생성, ③ 생성한 전파모델을 이용하여 서비스 지역 전체에 수신 신호세기 추정 과정을 순차적으로 진행한다.

그림 1 은 특정 기지국(PCI: 289)에 대해 도심지역에서 차량수집 데이터를 활용하여 30m 간격으로 생성한 전파지도 DB(좌)와 생성적 적대 신경망 개념을 활용하여 생성한 측위 DB(우)를 비교한 것이다. 차량이 구석구석 진입하기 어려운 대규모 주거지역 및 상업지역, 공원 등의 구역에서는 미수집으로 인해 전파지도 DB 에 비어있는 공간이 측위 DB 에 비해 비교적 넓은 것을 볼 수 있다. 반면 측위 DB 는 수집되지 않은 지역에 대해서도 신호세기 정보를 추정하여 저장하므로 비어있는 공간없이 DB 가 생성되었다.

Ⅲ. 미수집지점에서의 측위 인프라 유사도 비교

심층학습 방법을 활용하여 생성한 측위 DB 가 차량수집이 진행되지 않은 지역에서 갖는 효용성을 분석하기 위하여 미수집지점에서 측정한 데이터를 전파지도 DB 와 측위 DB 를 이용하여 비교하였다. 시험지점은 테스트베드 내 차량으로 수집이 이루어지지 않은 구역의 중심부분

위주로 측정하였고, 일반적인 성능을 분석하기 위하여 총 14 개 지점에 대해 1,800 여 회의 LTE 시험데이터를 수집하였다.

그림 2 는 시험지점 중 한 곳인 모 중학교와 공영주차장 인접면에서 측정한 데이터에 대해 전파지도 DB 와 측위 DB 를 이용한 분석 결과를 도시화 한 것이다. 측정값과 DB 간 유사도는 두 집합의 유사도를 평가하는 대표적인 방법인 Jaccard 지수를 응용하여 측정값 중 DB 에 기록되어 있는 기지국의 수를 모든 측정된 기지국의 수로 나누어 계산하였고, 이는 다음의 식과 같다.

$$Similarity = \frac{\#\ of\ \ matched\ \ PCI}{\#\ of\ \ scanned\ \ PCI}$$

측정 가능 한 기지국의 수는 기지국 과의 통신환경뿐만 아니라 단말에 내장된 통신 칩셋에 따라 달라지게 되는데, 그림 2 의 시험에서는 총 8 개의 기지국이 검색되었다. 시험지점에서 반경 200m 이내에 있는 참조위치에 대해 전파지도 DB 를 사용한 경우 유사도는 평균 19.5%로 나타난 데에 비해 측위 DB 의 경우 평균 80.9%로 크게 향상되었다. 실제 위치추정 시 유의미한 지점으로 사용할 수 있는 측위 인프라 유사도 60% 이상을 보이는 지점의 유사도 평균은 63.8%에서 73.1%로 향상되어, 측위 DB 가 참위치 주변의 참조위치 유사도를 높이고 전체적인 유사도를 개선하는 효과를 확인하였다.

표 1 은 전체 14 개 지점의 시험데이터에 대해 참위치 반경 200m 이내 참조위치의 유사도 평균과 유사도 60% 이상 지점의 평균 유사도를 정리한 것이다. 그 결과, 심층학습 기반 측위 DB 를 이용한 위치추정 시 변별력을 크게 향상시킬 수 있을 것으로 기대되었다.

Ⅳ. 결 론

본 논문에서는 차량수집 데이터를 활용하여 생성한 전파지도 DB 와 학습을 통해 미수집지역에서도 신호세기를 추정한 측위 DB 를 이용하여 미수집지점에서의 측위 인프라 유사도를 분석하였다. 그 결과, 측위 DB 사용 시 참위치 주변의 측위 인프라 유사도가 크게 향상되었고, 유사도 60%이상 지점의 평균 유사도 또한 전파지도 DB 대비 높아져 무선통신 인프라 기반 측위 DB 생성 분야에서 심층학습 알고리즘의 활용 가능성을 확인하였다.

ACKNOWLEDGMENT

이 논문은 2021 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (2019-0-01401, 긴급구조용 측위 품질 제고를 위한 GPS 음영 지역 내 다중신호패턴의 학습 기반 3 차원 정밀측위 기술 개발).

참 고 문 헌

[1] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., et al. 2014, Generative Adversarial Nets, in Proc. of Advances in Neural Information Processing Systems, Montreal, Canada, 8-13 Dec 2014.

[2] Youngsu Cho, Myungin Ji, Jooyoung Kim, Yangkoo Lee, Sangjoon Park, Beomju Shin, and Taikjin Lee, “WARP-P: Wireless signal Acquisition with Reference Point by using simplified PDR-system concept and performance assessment,” in Proc. ION PNT, 2013.

[3] 지명인, 전주일, 조영수, "A Positioning DB Generation Algorithm Applying Generative Adversarial Learning Method of Wireless Communication Signals," JPNT 9(3), 2020.

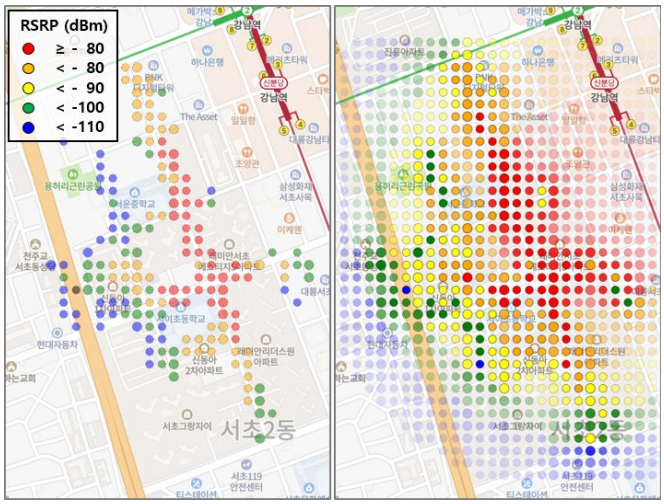


그림 1. 생성된 전파지도 DB 및 측위 DB 비교

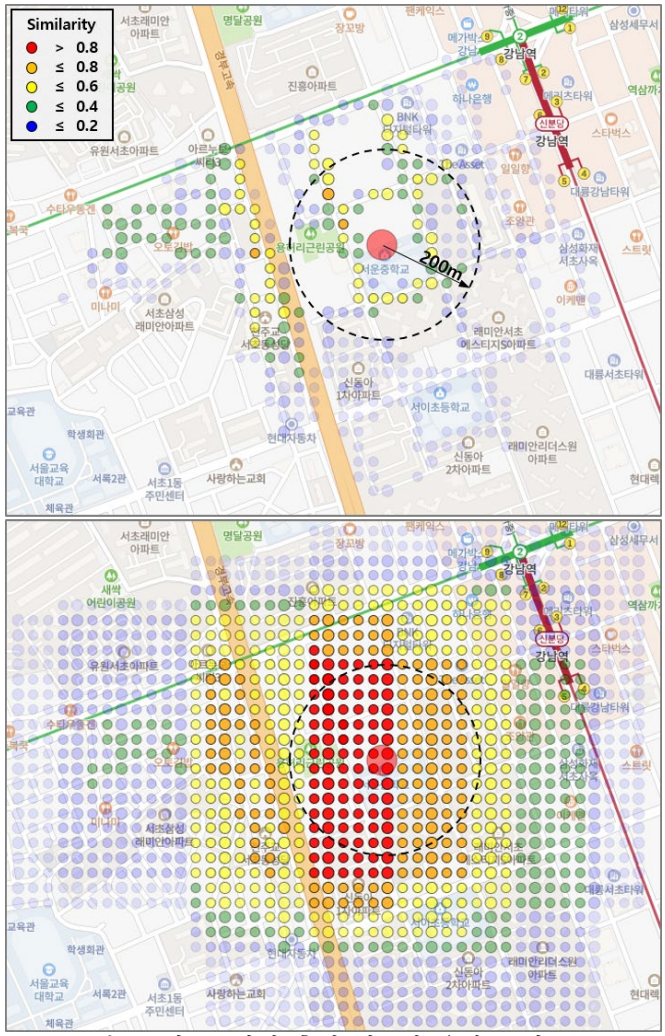


그림 2. 각 DB와의 측위 인프라 유사도 비교

표1. 전체 시험데이터의 유사도 평가 결과

	참위치 반경 200m 이내 참조위치의 유사도 평균	유사도 60% 이상 지점의 유사도 평균
전파지도DB	15.6%	67.7%
측위DB	69.7%	74.0%

딥러닝 기반 자동 변조 분류 위한 CNN 모델 성능 분석

김승환, 권기협, 김동성
금오공과대학교 ICT융합특성화연구센터
ksh001@kumoh.ac.kr, navkwon@kumoh.ac.kr, dskim@kumoh.ac.kr

Performance Analysis of Convolutional Neural Network Model for Deep Learning-Based Automatic Modulation Classification

Seung-Hwan Kim, Ki-Hyeop Kwon, Dong-Seong Kim
kit ICT Convergence Research Center.

요 약

본 논문은 무선 인지 네트워크에서 딥러닝 기반의 자동 변조를 위해 사용되는 CNN 모델에 대해 필터 크기에 따라 변화되는 예측성능에 대해서 분석하였다. 성능분석을 위해 사용된 데이터셋은 Deepsig의 RADIOML 2018.01a 그리고 RML2016.10b를 사용하였다. 기존의 모델에서 사용한 1×3의 필터를 1×5, 1×8 그리고 1×11 크기로 대체하여 나타나는 성능에 대해 나타내었다. 성능 분석에 따라서 1×8 필터가 최적의 필터 크기인 것을 보였다.

Ⅰ. 서 론

무선통신 기술의 발전을 통해 IoT 모바일 기기에 대한 서비스 증가로 주파수 자원에 대한 고갈 현상이 심화되면서 산업 분야[1], 군사 분야 등 다양한 응용 분야에 대한 서비스를 제공에 한계점을 가지게 된다. 이러한 문제를 해결하고자 인지통신기술이 보급되었고 이는 제한된 주파수 자원을 효율적으로 사용할 수 있도록 하게 한다. 인지통신기술은 면허 주파수 대역을 허가되지 않은 사용자(부 사용자)가 유휴 채널을 사용 가능하도록 하는 기술이며 부 사용자는 인가된 사용자(주 사용자)의 통신을 방해하지 않는 조건을 가지고 있다. 따라서 주 사용자의 통신에 간섭하지 않도록 하기 위해 인지통신을 가능하게 하고 부 사용자가 네트워크 채널을 사용하도록 하는 다양한 기법들이 구성되어 있으며, 그 중 자동변조분류 기법[2]은 부 사용자가 채널을 사용하기 위한 필수적 기법이다. 자동변조분류 기법은 수신된 신호를 사전지식 없이 변조 종류를 식별해내는 기법으로 기존의 2가지 방식을 통해 연구 되었다. 첫째는 최대 우도추정 기반의 분류 방식이며, 두 번째는 특징 기반의 패턴인식 방식이다. 최대 우도추정 방식을 적용 시 최적의 성능을 보장하지만 구현하기 위한 매우 높은 복잡도가 발생함으로 차선의 성능을 가지는 패턴인식 방식이 일반적으로 적용된다. 패턴인식 방식의 분류 진행과정은 입력 데이터 전처리, 특징추출, 마지막으로 분류결정으로 3단계로 진행되지만 딥러닝 기반 패턴인식 방식[3]은 입력 데이터 전처리와 특징추출/분류결정으로 2단계로 진행된다. 본 논문에서는 Deepsig Dataset: RML 2018.01a 데이터셋과 RML 2016.10b 데이터셋을 활용하여 각 필터 크기에 대한 성능을 분석하였다.

Ⅱ. 시스템 모델

송신부의 전송된 신호는 클럭 오프셋과 채널 페이딩에 의해 손상되어 수신부를 전송이 되며 k번째 수신된 신호는 식 1과 같이 나타낼 수 있다.

$$y_k = \alpha e^{j(2\pi \triangle ft + \triangle \phi)} x_k + n. \tag{1}$$

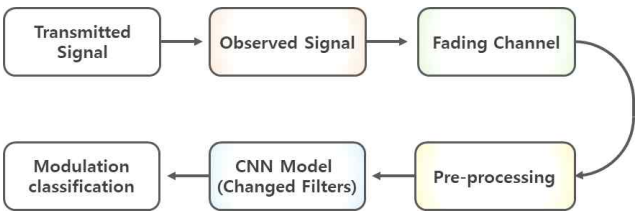


그림 1 Example of system model.

여기서 α은 다중 경로의 진폭이며, $\triangle f$와 $\triangle \phi$는 캐리어 주파수와 위상의 오프셋, x_k는 변조신호 그리고 n는 백색가우시안잡음을 나타낸다.

신호의 자동 변조 분류를 위해 사용된 데이터셋 RML 2018.01a은 1024 샘플을 1프레임으로 구성하여 총 884,736 프레임을 사용하였으며 각 변조 타입 당 4096 프레임이 사용하였다. RML 2016.10b의 경우 128 샘플을 1 프레임으로 구성하여 총 540,000 프레임이 사용되었으며 각 변조 타입 당 6,000 프레임이 사용하였다. RML 2018.01a의 경우 측정된 SNR 구간은 0 dB에서 16 dB이며 2 dB의 간격으로 예측성능을 나타내었고 RML 2016.10b의 경우 SNR 구간은 -10 dB에서 6 dB이며 2 dB의 간격으로 예측성능을 나타내었다. 수신된 신호 프레임으로 재구성하여 식 2를 통해 RMS 방식으로 정규화하였으며, 정규화된 신호는 In-phase, Quadrature-phase (IQ) 요소로 분리 후 딥러닝 네트워크에 학습 하였다.

$$\overline{s_i} = \frac{s_i}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}|s_i|^2}}, \tag{2}$$

Ⅲ. CNN 모델 성능분석

CNN 모델을 통한 특징을 추출하기 위해 컨볼루션 레이어를 통해 특징맵을 생성하고 액티베이션 레이어를 통해 비선형적으로 출력된 특징맵을 적층한다. 컨볼루션 레이어와 액티베이션 레이어 사이 배치 정규화 레이

어를 사용하여 일반화 성능을 높여 학습 속도와 정확도를 높이게 할 수 있다. 특징맵 생성 시 컨볼루션, 배치 정규화 그리고 액티베이션 레이어를 하나의 컨볼루션 블록의 기능으로 두었고 각각에 대한 식은 식 2-4를 통해 나타내었다.

$$z_\kappa = \sum_i w_i v_i + b, \tag{2}$$

여기서 v_i는 입력값, w_i는 컨볼루션 계수이고 그리고 b는 바이어스 값이다. 배치 정규화에 대한 부분은 식 3을 통해 나타내었다.

$$\acute{z}_\kappa = \frac{z_\kappa - \mu_B}{\sqrt{\sigma_B^2 + \rho}}, \tag{3}$$

여기서 μ_B는 최소 배치의 평균값이며 σ_B^2는 최소 배치의 분산값 그리고 ρ는 분모가 0이 되지 않도록 주어지는 작은 값이다. 마지막으로 액티베이션 레이어는 식 4를 통해 나타내었다.

$$f(\acute{z}_\kappa) = \begin{cases} 0, & \text{if } \acute{z}_\kappa < 0, \\ \acute{z}_\kappa, & otherwise, \end{cases} \tag{4}$$

제안하는 CNN은 SGDM (Stochastic Gradient Descent with Momentum) 최적화 기법을 사용하였다. 제안된 CNN 구조는 그림 2, 3을 통해 나타내었으며, 그림 3에서 ConvB는 컨볼루션 블록, MPool은 MaxPooling 레이어, APool은 AveragePooling 레이어, GAPool은 Groval APool 레이어, DO는 DropOut 레이어, FC는 Fully-Connected 레이어, SM은 Softmax 레이어 그리고 CO는 Classification Output을 각각 의미한다. 그림 2에서 CovnB의 (1×8), (3×1) 그리고 (1×1)은 컨볼루션 레이어 필터 크기를 나타내며 여기서 (1×8) 필터 크기 사용하는 각 레이어를 (1×3), (1×5) 그리고 (1×11)로 각각 크기를 적용하여 성능을 비교 분석하였다. 그림 2에서 제안된 CNN 모델은 총 3개의 블록과 출력으로 구성되며, 18개 컨볼루션 을 통해 특징맵을 추출하였다. 1024 또는 128 입력 길이는 ABlock에서 사용된 MPool를 통해 축소 하였고, CBlock의 APool을 통해 2-Dimension 특징맵을 1-D로 축소하도록 하여 계산복잡도를 낮추었다. 출력단에서는 FC 레이어에 전달된 입력 크기는 글로벌 평균풀링을 적용하여 1×1로 조정하여 학습 파라미터 수를 낮추었다.

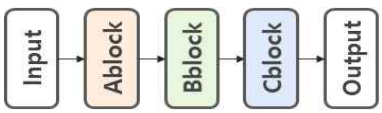


그림 2 신호 기반 CNN 아키텍처.

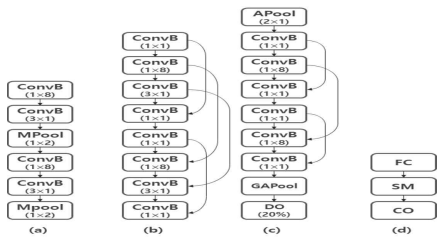


그림 3 CNN 모델 구조 (a)Ablock (b)Bblock (c)Cblock (d)Output.

Ⅳ. 시뮬레이션 결과

본 논문에서 제안된 모델의 성능 분석을 그림 3, 4를 통해 각 데이터셋에 대한 비교결과를 나타내었다. 시뮬레이션을 위한 환경 설정은 최소 배치 크기 64, 최대 Epoch 40, 초기 학습비율 0.1 그리고 Epoch 35마다 0.1 학습 비율을 낮추도록 하였다. 결과를 통해서 두 가지 데이터셋 모두 (1×8)의 필터를 사용 하였을 때, 최적의 성능을 가지는 것을 볼 수 있다. 각 필터의 성능 차이는 낮은 SNR에서 높은 SNR로 증가할 때, 따라서 최저 분류 정확도에서 최고 분류 정확도로 천이 될 때 각 필터에 대한 분류 성능이 뚜렷하게 나타나는 것을 볼 수 있다. 따라서 본 시험을 통해 필터 크기가 증가 할수록 최저 그리고 최대 성능을 높이는 것이 아니라 천이되는 과정에서 분류 성능을 높이는 것을 볼 수 있다.

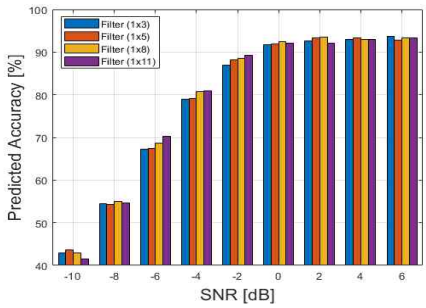


그림 4 Performance comparison with RML 2016.10b

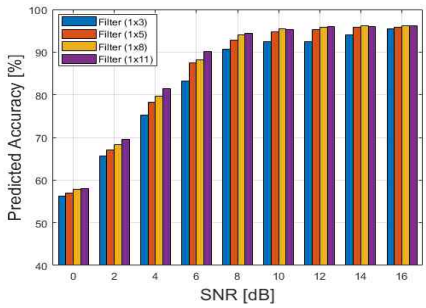


그림 5 Performance comparison with RML 2018.10a

Ⅴ. 결론

본 논문에서는 자동 변조 분류를 위해 CNN 모델을 제안하고 컨볼루션 필터 크기에 대한 성능을 분석하였다. 제안된 모델의 성능분석을 위해 RML 2018.01a, RML 2016.10b 데이터셋을 사용하였으며 (1×8)의 필터 크기를 사용할 때 최적의 성능을 보이는 것을 나타내었다.

ACKNOWLEDGMENT

이 논문은 정보통신기획평가원의 Grand ICT연구센터지원사업 (IITP-2020-2020-0-01612)과 한국연구재단의 대학중점연구소 지원사업으로 수행된 연구임(2018R1A6A1A03024003).

참 고 문 헌

[1] S.-H. Kim, J.-W. Kim and D.-S. Kim, "Energy Consumption Analysis of Beamforming and Cooperative Schemes for Aircraft Wireless Sensor Networks," *Appl. Sci.*, vol. 10, no. 12, pp. 4374-4391, Jun. 2020.

[2] S. -H. Kim, J. -W. Kim, W. -P. Nwadiugwu and D. -S. Kim, "Deep Learning-Based Robust Automatic Modulation Classification for Cognitive Radio Networks," *IEEE Access*, vol. 9, pp. 92386-92393, June 2021.

[3] S. -H. Kim, J. -W. Kim, V. -S. Doan and D. -S. Kim, "Lightweight Deep Learning Model for Automatic Modulation Classification in Cognitive Radio Networks," *IEEE Access*, vol. 8, pp. 197532-197541, Oct. 2020.

링크 이용률 향상을 위한 DQN 기반 TCP 혼잡제어 알고리즘

서상진*, 조유제

경북대학교 전자전기공학부

chil258@knu.ac.kr, yzcho@ee.knu.ac.kr

DQN-Based TCP Congestion Control Algorithm to Improve Link Utilization

Sang-Jin Seo* and You-Ze Cho

School of Electronic and Electrical Engineering, Kyungpook National University

요 약

기존 TCP 혼잡제어 알고리즘은 큰 대역폭과 짧은 Round Trip Time(RTT)으로 인해 가용 링크 대역폭을 모두 사용하지 못하거나, 혼잡이 발생할 때까지 Congestion Window(Cwnd)를 증가시켜 발생하는 처리량 손해 문제가 존재한다. 이를 해결하기 위해 기계 학습 기반 혼잡제어들이 제안되었는데, 기계 학습의 특성상 파라미터에 따라 처리량이나 링크 이용률과 같은 성능차이가 있음을 확인하였다. 본 논문에서는 NS3 시뮬레이터 기반의 실험 환경에서 Additive Increase/Multiplicative Decrease(AIMD)기반의 TCP 혼잡제어 알고리즘에 기계 학습의 일종인 Deep Q-Network(DQN)를 적용하여 성능 및 한계점을 실험 후 평가하였다.

Ⅰ. 서 론

기존 TCP 혼잡제어 알고리즘은 큰 대역폭과 짧은 Round Trip Time(RTT)으로 인해 가용 링크 대역폭을 모두 사용하지 못하거나, 혼잡이 발생할 때까지 Congestion Window(Cwnd)를 증가시켜 발생하는 처리량 손해 문제가 존재한다.

이를 해결하기 위해 기계 학습 기반의 혼잡제어 알고리즘인 RL-TCP[1], Q-TCP[2] 등이 제안되었지만, 기계 학습의 특성상 파라미터에 따라 처리량이나 링크 이용률과 같은 성능 차이가 있음을 확인하였다.

본 논문에서는 NS3-GYM toolkit 과 NS3 시뮬레이터 기반의 실험 환경에서 Additive Increase/Multiplicative Decrease(AIMD)기반의 TCP 혼잡제어 알고리즘에 기계 학습의 일종인 Deep Q-Network(DQN)를 적용해 단일 흐름에서의 처리량과 링크 이용률을 측정 및 분석하고, 한계점을 실험으로 확인하였다.

Ⅱ. 본론

A. State/Action space 및 Rewards function

본 논문에서는 학습을 위해 [Threshold(임계값), RTT, Cwnd, Throughput(처리량)]의 State space 와 식. (1)과 같은 Cwnd 증감에 관한 Action space 를 사용하였다.

$$Action = \begin{cases} Cwnd = Cwnd + SegmentSize \times (a-1), & a > 0 \\ Cwnd = Cwnd - SegmentSize, & a = 0 \end{cases}, \quad \text{식 (1)}$$

이때 a 는 DQN 학습 결과값으로, 현재 State 에서 네트워크의 상태를 다음과 같이 판단하도록 학습하였다.

$$a = \begin{cases} 0, \text{ 혼잡 위험 매우 높음 (감소)} \\ \begin{cases} 1, \text{ 혼잡 위험 높음} \\ 2, \text{ 혼잡 위험 보통} \\ 3, \text{ 혼잡 위험 낮음} \end{cases} \text{ (증가)} \end{cases}, \quad DQN \text{ 결과값}$$

수행된 Action 에 따른 Rewards 는 다음과 같은 수식으로 정의하였다.

$$Rewards = \frac{Throughput}{Limit\ Throughput}, \quad \text{식 (2)}$$

$$Throughput = \frac{Cwnd}{RTT}, \quad \text{식 (3)}$$

$$Limit\ Throughput = \frac{Cwnd\ before\ congestion}{RTT\ before\ congsetion}, \quad \text{식 (4)}$$

$$\begin{cases} \varepsilon, & Random\ Action \\ 1-\epsilon, & Best\ Rewards\ Action' \end{cases} \quad \varepsilon = 0.015, \quad \text{식 (5)}$$

식. (2)는 링크 이용률을 나타내며, Limit Throughput 을 기준으로 계산한다. 이때 Limit Throughput 은 현재 링크 용량에서 사용가능한 최대치에 가까운 처리량으로 식. (4)를 이용해 계산하며, Cwnd 와 RTT 는 마지막으로 혼잡이 발생하기 바로 전의 State 값을 사용한다.

Rewards 값이 1 을 초과할 경우 -1 의 값을 갖도록 페널티를 부여하였으며, 혼잡이 발생하지 않으면 Limit Throughput 을 다시 찾는 동작을 하게 된다. 또한 Rewards 에 따른 Action 의 선택은 식. (5)의 확률을 가지는 Epsilon-Greedy 정책을 사용하였다.

Time out 발생 시 표준 TCP 동작[3]에 따라 Slow Start 단계로 재 진입하게 되며, CUBIC 표준[4]에 따라 확장성에 초점을 두고 빠른 Cwnd 증가 및 최대 처리량 달성을 위해 Cwnd 감소 지수는 0.7 로 설정하였다.

B. 실험 환경



그림 1. NS3 시뮬레이터 실험 환경 구성도

그림 1 은 혼잡제어의 성능 및 링크 이용률 분석을 위한 NS3 시뮬레이터의 실험 환경을 나타낸다. 호스트는 실험 시나리오에 따라 NewReno, CUBIC, DQN 기반 TCP 혼잡제어 알고리즘을 사용해 데이터를 전송한다. 실험 환경의 RTT 는 100ms 로 설정하였으며, 300 초의 데이터 전송 시간을 가진다. 병목 링크 대역폭은 50Mpbs 고정과 100 초 주기로 50, 25Mpbs 를 왕복하는 2 개의 상황을 부여하여 실험하였다.

C. 단일 흐름에서 DQN TCP, NewReno, CUBIC 비교

그림 2 는 단일 흐름일때 각 혼잡제어의 Cwnd 를 비교 분석하기 위한 그래프이다. DQN 기반 TCP 혼잡제어 알고리즘은 최대 혼잡창의 크기를 학습하고 이를 유지하였다. 또한 Epsilon-Greedy 정책으로 랜덤 선택 도중 혼잡창을 증가시켜 중복 ACK 가 발생하더라도 빠르게 학습한 최대 처리량으로 복구하였다.

평균 처리량은 NewReno 가 43.93Mbps 로 가장 낮았고, DQN 기반 혼잡제어 알고리즘이 50.44Mbps 로 가장 높았다. 하지만 CUBIC 은 49.64Mbps 로, DQN 기반 혼잡제어 알고리즘보다 낮지만, 큰 차이가 없음을 확인하였다.

D. 병목 링크의 용량이 변화할 때의 실험

그림 3 은 단일 흐름에서 병목 링크의 용량이 변화할 때 각 혼잡제어의 링크 이용률을 비교하기 위한 Cwnd 그래프이다. NewReno 와 CUBIC 은 혼잡이 발생할 때까지 Cwnd 를 증가시키기 때문에 병목 링크의 용량이 변화하여도 지속적인 Cwnd 증가 및 혼잡 발생으로 인한 감소 동작을 반복하였다.

DQN 기반 TCP 혼잡제어 알고리즘은 처리량이 가장 높은 Cwnd 를 학습 후 유지하게 동작한다. 또한 Epsilon-Greedy 정책에 따른 확률로 Cwnd 를 증가시키며 새로운 최대 처리량을 탐색하는 과정에서 혼잡이 발생하면 Cwnd 를 감소시키고, 그렇지 않으면 다시 Cwnd 를 유지하게 된다.

링크 대역폭이 50, 25Mbps 일 때 각 혼잡제어에 따른 평균 처리량은 다음과 같다.

링크 대역폭	혼잡제어	평균 처리량 (Mbps)
50Mbps	DQN 기반	49.37
	NewReno	43.58
	CUBIC	48.67
25Mbps	DQN 기반	25.4
	NewReno	23.31
	CUBIC	25.34

실험에서 DQN 기반 TCP 혼잡제어 알고리즘의 평균 처리량은 모든 구간에서 NewReno 보다 최소 9%, 최대 13.3% 높았다. 하지만 CUBIC 과 비교했을 때는 최소 0.2%, 최대 1.4%의 미미한 평균 처리량 상승을 보였다.

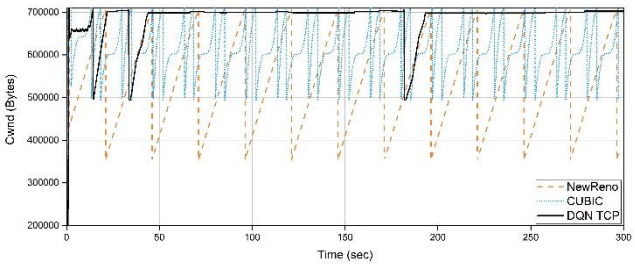


그림 2. 단일 흐름일때 각 혼잡제어의 Cwnd 비교

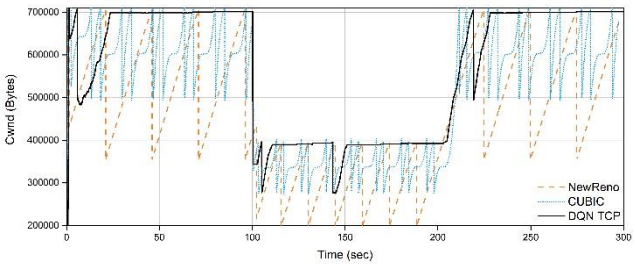


그림 3. 병목 링크 용량 변화 시 각 혼잡제어의 Cwnd 비교

Ⅲ. 결론

본 논문에서는 AIMD 알고리즘에 DQN 기반 혼잡제어 알고리즘과 기존 혼잡제어 알고리즘 사이의 단일 흐름 성능 차이와 링크 용량 변화 시 링크 이용률의 비교 및 분석을 위해 NS3 시뮬레이터로 실험환경을 구성하였다. 모든 실험에서 DQN 기반 TCP 혼잡제어 알고리즘은 학습한 Cwnd 를 최대한 유지하며 가장 높은 처리량을 보여주었지만, CUBIC 과의 처리량 차이가 약 1%로 미미하다는 한계점이 존재한다. 추후 AIMD 알고리즘이 아닌 비선형 Cwnd 증가 알고리즘을 적용해 성능 개선 연구를 진행할 계획이다.

ACKNOWLEDGMENT

This research was supported in part by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by Ministry of Education (No. NRF-2018R1A6A1A03025109) and by National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. NRF-2019R1A2C1006249).

참 고 문 헌

[1] Yiming Kong, Hui Zang, Xiaoli Ma, “Improving TCP Congestion Control with Machine Intelligence”, August. 2018, NetAI’18

[2] Wei Li, Fan Zhou, Kaushik Roy Chowdhury, “QTCP: Adaptive Congestion Control with Reinforcement Learning”, July-September. 2019, IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING, VOL.6, NO. 3

[3] “TCP Congestion Control”, rfc5681

[4] “CUBIC for Fast Long-Distance Networks”, rfc8312

동형암호를 이용한 프라이버시 보호 얼굴 검증 시스템

심준석, 조현진, 홍윤영, 김호원*

*부산대학교

sugo312@pusan.ac.kr, wh77r77@pusan.ac.kr, hyy0238@pusan.ac.kr, *howonkim@pusan.ac.kr

Privacy Protection Face Verification System using Homomorphic Encryption

Junseok Shim, Hyunjin Jo, Yoonyoung Hong, HowonKim*

*Pusan Univ.

요 약

최근 딥러닝 기술의 발전에 따라 컴퓨터 비전 분야에서 얼굴 검증 기술에 대한 다양한 연구가 제안되고 있다. 그러나 딥러닝 모델 훈련에 사용되는 데이터를 추출하는 공격인 모델 전도 공격에 의해 얼굴 검증 시스템에 등록된 이미지가 유출될 가능성이 있어 프라이버시 침해 문제가 제기되고 있다. 본 논문에서는 프라이버시 침해 문제를 해결하기 위해 Paillier 동형암호화와 샴 네트워크를 사용한 얼굴 검증 시스템을 제안한다. 이미지에 Paillier 동형암호화를 수행하고 샴 네트워크를 통해 암호화 이미지 사이의 얼굴 검증을 수행한다. ROC 커브를 통해 원본 이미지의 얼굴 검증 결과와 암호화 이미지의 얼굴 검증 결과를 비교하였다. 실험 결과, 원본 이미지와 암호화 이미지의 얼굴 검증 결과로 비슷한 AUC를 보였으며, 프라이버시 보호와 동시에 얼굴 검증이 가능함을 보였다.

Ⅰ. 서 론

얼굴 검증은 입력으로 들어오는 두 개의 얼굴 영상이 동일 인물인지 여부를 판단하는 1:1 검증 문제이다. 최근 딥러닝 기술의 발전으로 딥러닝 기술을 적용시킨 얼굴 검증(Face verification) 기술의 다양한 연구가 제시되고 있으며, FaceID 등 모바일 앱 로그인 및 결제 시스템에서 사용되고 있다. 그러나 딥러닝을 활용한 보안기술 연구가 활발히 이루어지는 가운데, 딥러닝을 악용한 연구사례들도 많이 나타나고 있다. 모델 전도 공격(Model inversion attack)은 딥러닝 모델의 학습에 사용된 데이터를 추출하는 공격기법이다. 주어진 입력의 출력 결과와 신뢰도를 분석하여 역으로 데이터를 추출해 낼 수 있다. 얼굴 검증 모델의 경우, 공격자가 모델의 학습에 사용된 이미지를 추출하여 프라이버시 침해 문제를 초래할 수 있다. 따라서 이미지가 유출되더라도 피해를 최소화하기 위해 데이터 암호화가 필요하다. Munirah 등[1]은 인공위성으로 촬영한 이미지를 보호하기 위해 동형암호를 적용한 모델 학습 기법을 제안했다. Ma 등[2]은 얼굴의 특징을 통해 등록 및 검증 시스템인 DeepID를 제안하였다. DeepID 시스템은 Convolutional Neural Network(CNN) 기반으로 추출된 얼굴 특징을 암호화 모듈을 통해 암호화한다. 시스템에 저장된 특징들과 암호화된 특징 간의 해밍 거리(Hamming distance) 값이 임계값 이하일 겨우 동일 인물이라고 판별하는 시스템이다.

본 논문에서 제안하는 시스템은 암호화를 통해 얼굴 이미지 데이터의 프라이버시를 보호하면서 얼굴 검증을 한다. 또한 [2]와 달리 기존의 CNN 기반의 얼굴 검증 시스템이 아닌 샴 네트워크(Siamese network)를 사용해 소량의 이미지만으로 얼굴 검증 결과에 높은 성능을 보인다.

Ⅱ. 본론

2.1 배경 지식

본 시스템에서는 프라이버시 보호 및 얼굴 검증을 위해 Paillier 암호 체계와 샴 네트워크를 사용했다. Paillier 암호 체계는 비밀키에 대한 정보가 없더라도 암호문만 이용하여 평문 간의 상수 곱과 덧셈 결과를 구할 수 있는 동형 연산이 가능한 특징을 가진다. 이미지의 동형암호화는 그림 2와 같이 먼저 이미지를 1차원 배열로 변형한 후 공개키를 통해 암호화한다. 암호화된 1차원 형태의 배열은 모델에 입력하기 위해 이미지의 원래 모양으로 변형하여 암호화를 완료한다.

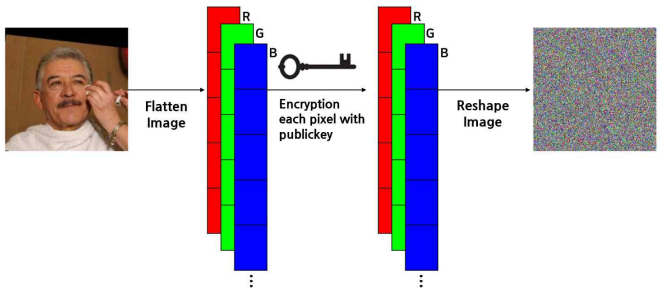


그림 1 이미지 암호화 과정

샴 네트워크는 그림 2와 같이 동일한 파라미터를 공유하는 트윈 네트워크(Twin network)를 포함하는 딥러닝 모델이다. 트윈 네트워크는 검증받고자 하는 이미지와 검증용 참조 이미지로 구성된 이미지 쌍을 입력받아 특징 벡터를 추출한다. Distance layer에서 추출된 특징 벡터들의 L_1거리를 계산한 후, Prediction layer에서 L_1거리를 이용해 유사도를 계산한다.

2.2 모델 구조

본 논문에서 제안하는 얼굴 검증 시스템의 샴 네트워크 모델은 $128 \times 128 \times 3$ 크기의 이미지 쌍을 입력받는다. 이미지 쌍은 트윈 네

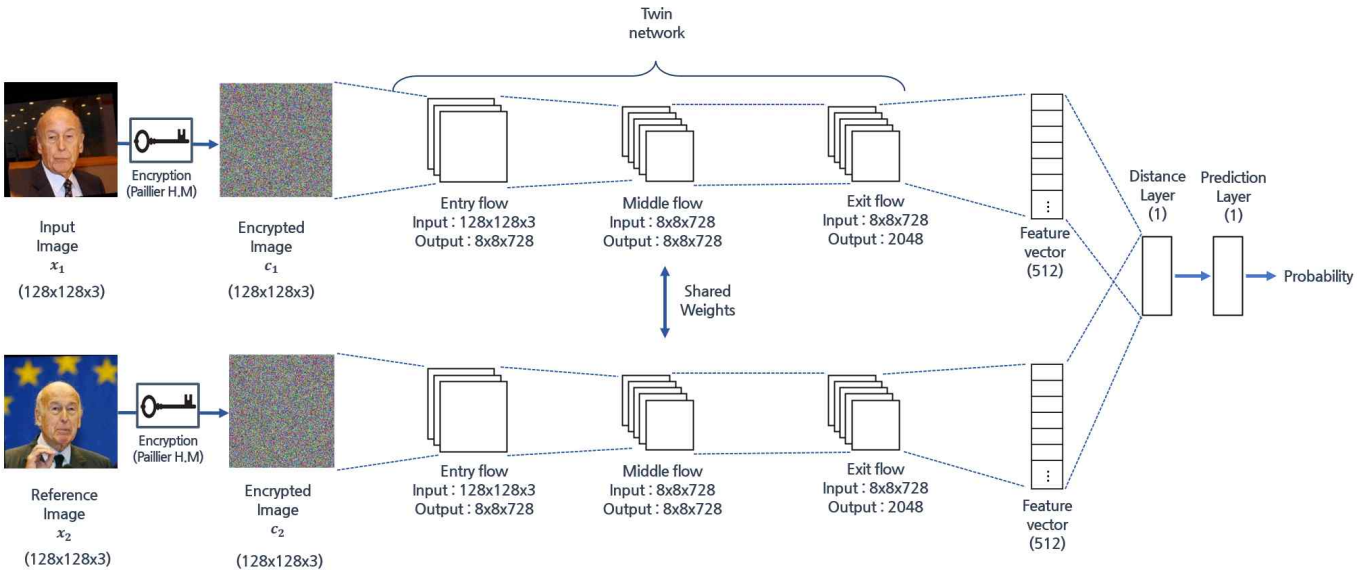


그림 2 시스템 구조

트워크에 입력되기 전 동형암호화를 거친다. 본 논문에서는 연산의 효율을 위해 ImageNet[3] 데이터 세트로 사전 학습된 XceptionNet[4]을 트윈 네트워크로 사용한다. 샴 네트워크의 Prediction Layer에서 출력되는 입력 이미지 쌍의 유사도는 입력 이미지의 인물들을 동일 인물로 판단하는 기준이 된다.

2.3 실험 및 실험 결과

본 연구에서는 얼굴 검증 모델의 학습을 위해 Labeled Faces in the Wild(LFW)[5] 데이터 세트를 사용한다. LFW 데이터 세트는 5,749명의 사람에 대해 13,233개의 이미지로 구성된다. 학습 및 검증을 위해 LFW 데이터 세트에 존재하는 1,680명에 대해 동일 인물로 구성된 이미지 쌍과 서로 다른 인물로 구성된 이미지 쌍으로 데이터를 구성했다. 동일 인물로 구성된 이미지 쌍은 1, 서로 다른 인물로 구성된 이미지 쌍에는 0의 레이블을 부여했다.

시스템의 성능 평가를 위해 원본 이미지와 암호화 이미지의 얼굴 검증 결과의 Reciever Operating Characteristic(ROC) 커브를 사용한다. 그림 3은 원본 이미지와 암호화 이미지의 ROC 커브 비교이다. 원본 이미지와 암호화 이미지의 AUC(Area Under the ROC curve)는 각각 0.99, 0.97이다. Park[8]에 따르면 AUC에 따른 정확도는 AUC가 0.5~0.7 일 때 낮은 정확도(less), 0.7~0.9 일 때 중등도(moderate), 0.9~1.0 일 때 높은 정확도(high)를 가진다. 따라서 암호화 이미지에 대한 얼굴 검증 결과와 원본 이미지의 얼굴 검증 결과가 큰 차이가 없이 높은 정확도의 결과를 가지는 것을 확인할 수 있다.

Ⅲ. 결론

본 논문에서는 얼굴 이미지에 대한 프라이버시를 보호하는 얼굴 검증 시스템을 제안하였다. 제안된 기술은 Paillier 동형암호를 사용해 암호화된 이미지를 학습 및 검증에 사용해 모델 전도 공격으로 인해 유출되는 이미지들에 대하여 프라이버시를 보호한다. 또한 샴 네트워크를 통해 적은 암호화 이미지 데이터에 대한 학습이 가능함을 보였다. 향후, 본 솔루션을 통해 얼굴 검증 이외 프라이버시 보호에 민감한 객체 탐지 분야 및 컴퓨터 비전 분야에서도 사용될 것으로 전망한다.

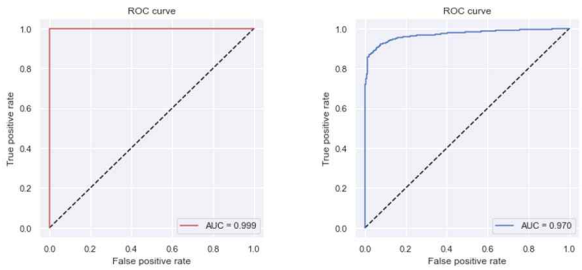


그림 3. ROC Curve 비교 (a) 원본 이미지의 ROC Curve, (b) 암호화 이미지의 ROC Curve

ACKNOWLEDGMENT

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (2019-0-01343,융합보안핵심인재양성사업)

참 고 문 헌

[1] Alkhelaiwi, Munirah, et al. "An Efficient Approach Based on Privacy-Preserving Deep Learning for Satellite Image Classification." Remote Sensing 13.11 (2021): 2221.

[2] Ma, Yukun, et al. "A secure face-verification scheme based on homomorphic encryption and deep neural networks." IEEE Access 5 (2017): 16532-16538.

[3] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems 25 (2012): 1097-1105.

[4] hollet, François. "Xception: Deep learning with depthwise separable convolutions." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.

[5] Park, S.-I., & O, T.-H. (2016). Application of Receiver Operating Characteristic (ROC) Curve for Evaluation of Diagnostic Test Performance. Journal of Veterinary Clinics, 33(2), 97. https://doi.org/10.17555/jvc.2016.04.33.2.97

완전 동형 암호화된 데이터를 위한 근사 합성곱 신경망의 근사 범위에 따른 성능에 관한 연구

이정현, 노종선*

서울대학교

jhlee@ccl.snu.ac.kr, *jsno@snu.ac.kr

A Study on the Performance of the Approximate Convolutional Neural Network According to Approximation Bound for Homomorphically Encrypted Data

Lee Junghyun, No Jong-Seon*

Seoul National Univ.

요 약

클라우딩 컴퓨팅의 보안을 유지하면서도 데이터에 대한 분석을 진행하기 위해 완전 동형 암호를 사용한 딥 러닝 모델이 최근에 활발하게 연구되고 있다. 딥 러닝 모델이 완전 동형 암호에 적용되기 위해서는 다항식으로 연산이 불가능한 함수들을 다항식으로 근사하는 것이 필요하다. 본 논문의 목표는 이러한 함수를 다항식으로 근사할 때 성능에 어떤 변화가 일어나는지 알아보는 것을 목표로 하였다. 특히, 본 논문의 시뮬레이션을 통하여 ReLU 함수의 다항식 근사 범위 설정이 기학습된 모델이 성능에 중요한 역할을 하고 있음을 확인할 수 있었다.

Ⅰ. 서 론

딥 러닝은 지속적인 기술 발전에 따라 데이터 분석을 위한 필수적인 방법이 되었고, 클라우딩 컴퓨팅을 활용하여 다양한 사람들에게 딥 러닝 모델을 활용한 데이터 처리 서비스를 진행하기도 한다. 근래에는 금융, CCTV 자료, 환자 의료정보와 같은 보안 유지가 필수적인 데이터의 분석에 딥 러닝을 활용기도 한다. 데이터를 가지고 있는 고객이 딥 러닝을 모델을 갖고 있는 다른 서버에게 데이터 분석을 의뢰한다고 할 때, 고객이 가지고 있는 데이터의 보안이 유지되어야 한다면 쉽게 데이터를 서버에게 전달할 수 없을 것이다. 이를 위해 고객은 데이터를 암호화하여 서버에게 전달해야 하는데, 암호화가 되어 있는 채로 데이터를 전달받은 서버는 복호화의 과정 없이 데이터 간의 원하는 연산을 진행하기가 힘들다. 이러한 문제를 해결할 수 있는 암호화 스킴이 동형 암호이다.

동형 암호(Homomorphic Encryption)는 암호화된 데이터 간의 연산, 덧셈과 곱셈을 지원하는 암호 스킴인데, 이 연산에는 특별한 복호화 과정이 필요없다는 것이 큰 장점이다. 따라서 동형 암호 스킴으로 고객의 데이터를 암호화할 수 있다면 고객의 데이터의 정보가 노출되지 않으면서도 서버가 원하는 연산을 수행하여 보안성이 유지되는 데이터 처리를 할 수 있게 되는 것이다. 동형 암호 스킴은 [1]에서 처음으로 격자 문제에 기반하여 제안되었고 그 이후로 동형 암호의 효율성 개선에 대한 연구가 활발히 진행되어 왔다. 최근 [2]에서 제안한 Homomorphic Encryption for Arithmetic of Approximate Numbers (HEAAN 또는 CKKS)인데, 이는 실수 연산에 적합하다는 특징이 있어 실제 데이터 분석에 활용 가능성이 높을 것으로 주목되고 있다.

동형 암호를 사용하여 데이터를 처리하는 방법에는 크게 두 가지가 있다. 첫째는 완전 동형 암호 (Fully Homomorphic Encryption) 기법으로서, 다른 도움 없이 서버가 제3자의 도움 없이 처음부터 끝까지 모든 연산을 할 수 있도록 하는 것이다. 이를 위해서 서버는 동형 암호에서 제한적으로 제공하는 덧셈과 곱셈만으로 데이터 처리를 진행해야 한다. 둘째는 부분 동형 암호 (Somewhat Homomorphic Encryption) 기법으로서 다자간 연산 (Multi-Party Computation) 방법을 사용하여 보다 넓은 범위의 연산을 가능하도록 하는 것이다. 부분 동형 암호의 큰 장점은 서버가 덧셈과 곱셈 이외의 연산을 비교적 자유롭게 할 수 있다는 것이지만, 이를 위해서는 특정 데이터를 고객과 주고받아야 한다. 이 과정에서 서버는 데이터 통신에 대한 비용뿐 아니라 자신이 갖고 있는 데이터 처리 기술(예를 들면 딥 러닝 모델의 구조)의 유출에 대한 위험을 감수해야 한다. 반면, 완전 동형 암호를 사용할 경우 데이터 처리에 대한 보안은 서버가 완벽하게 유지할 수 있다. 대신 아무리 간단한 연산이라도 덧셈과 곱셈이 아니라면 이를 구현하기가 까다롭다는 단점이 있다. 이렇듯 완전 동형 암호와 부분 동형 암호는 그 장단점이 선명하고 어느 것이 더 좋은 방법이라고 이야기하기가 어렵다. 본 연구에서는 완전 동형 암호에 대한 내용으로 논의를 진행하고자 한다.

비선형 활성 함수(non-linear activation function)는 딥 러닝의 발전에 가장 큰 기여를 한 기술 중 하나이고, 활발하게 사용되는 비선형 활성 함수의 예로 sigmoid, ReLU를 들 수 있다. 이러한 비선형 함수는 평문 상에서는 연산하기 상당히 쉬운 함수이나, 덧셈과 곱셈만 지원하는 암호문 상에서의 정확한 연산은 불가능하다. 딥 러닝에서 사용되는 주요 연산(합성곱 신경망 연산 등)은 대체로 선형 함수이기 때문에 동형 암호로 구현하는 것이 가능하다. 따라서 비선형 활성 함수를 어떤 방법으로 연산하느냐가 중요한 문제가 되고, 이에 대한 해결을 위해 크게 두 가지 방향의 연구가 진행되어 왔다. 첫째는 암호문 상에서 계산하기 쉬운 새로운 비선형 함수를 제안하는 것이다. 완전 동형 암호에서도 어렵지 않게 계산할 수 있는 $f(x)=x^2$과 같은 단순한 이차, 삼차 다항식을 사용하여 이에 맞는 새로운 딥 러닝 모델을 설계한 연구이다[3, 4, 5]. 둘째는 성능 개선에 유리한 ReLU 함수를 할 수 있는 만큼 최대한 정확하게 근사하여 기존의 딥 러닝 모델에 적용하는 연구이다[6]. ReLU 함수 $f(x)=\max\{x, 0\}$를 동형 암호에 적용하기 위해서, max 함수 또는 이를 위해 필요한 부호 연산(sign function)을 다항식으로 근사하는 연구가 진행되었고[7, 8], [6]은 [7]에서 제안한

minimax polynomial의 합성 기법을 사용하여 CIFAR-10와 ImageNet을 덧셈과 곱셈만으로 높은 정확도로 분류하는 데 성공하였다. 특히 [6]에서 제안한 방법은 ResNet과 같은 실제 성능이 검증된 기존의 딥 러닝의 모델의 구조와 기학습(pre-trained)된 파라미터를 그대로 활용할 수 있다는 장점이 있고, ReLU 함수를 정확하게 근사할수록 이미지 분류의 성능이 정확해진다. 본 연구에서는 ReLU 함수의 정확도에 대한 분석을 보다 자세히 진행하여 실제 암호화된 딥 러닝 모델에서 사용 가능한 다항식이 어떤 것인지, 특히 근사 범위가 어느 정도로 허용되는지에 대한 것을 알아보고자 한다.

Ⅱ. 본론

ReLU 함수를 계산하기 앞서 [7]에서는 아래와 같은 부호 함수

$$sgn(x)=\begin{cases}x/|x| & x\neq 0\\ 0 & x=0\end{cases}$$

를 다항식으로 근사하였다. 구체적으로, 주어진 precision parameter α에 대해, $-1\le x\le 1$의 구간에서

$$|sgn(x)-p_\alpha(x)|\le 2^{-\alpha}$$

를 만족하는 다항식 p_α가 제시되어 있다. 그리고 [6]에서는 이 다항식을 활용하여 $-B\le x\le B$의 구간에서 ReLU 함수 $f(x)=\max\{x,\ 0\}$를 RIGHT

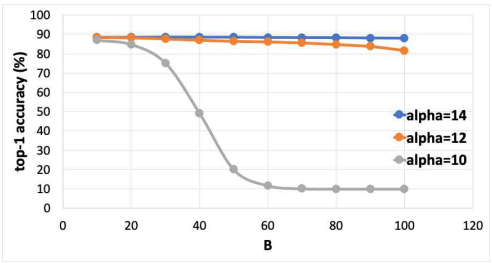
$$r_{\alpha,B}(x)=\frac{1}{2}(x+xp_\alpha(\frac{x}{B}))$$

와 같이 근사하였다. 이 다항식을 사용하면

$$|\ ReLU(x)-r_{\alpha,B}(x)\ |\ \le B(1/2)^{-\alpha}$$

의 오차 관계를 얻게 된다. Precision parameter α에 대해서는 모델의 성능 분석이 진행된 바 있지만[6], 근사 범위에 해당하는 값인 B 역시 ReLU 함수의 오차에 영향을 미치므로 여기에 대한 추가 분석이 필요하다.

본 연구에서는 CIFAR-10을 사용하여 기학습되어 88.36%의 top-1 정확도를 갖는 ResNet-20을 사용하였다. ResNet-20 안의 ReLU 함수를 위에서 언급한 다항식 $r_{\alpha,B}$로 바꾸었을 때 어떤 성능의 변화가 일어나는지에 대한 시뮬레이션을 진행하였다. Precision parameter α는 10, 12, 14를 사용하였고, 근사 범위에 해당하는 값인 B는 10부터 100까지 10을 간격으로 하여 실험하여 top-1 성능의 변화를 관찰하여 이를 [그림 1]에 정리하였다.



〔그림 1〕 B 값에 따른 ResNet-20의 top-1 정확도

[그림 1]에서 볼 수 있듯이 B가 증가할수록 top-1 정확도는 감소하며 이는 식 (1)에서도 유추할 수 있는 사실이다. α가 12와 14일 때는 B가 증가하여도 그렇게 큰 정확도 감소를 보이지 않는다. α가 14일 때는 87.94%까지 성능 저하가 일어났으며, α가 12일 때는 81.50%까지 일어났다. 그런데 α가 10일 때는 성능 저하가 B 값에 상당히 민감하게 일어남을 알 수 있다. B 값이 100까지 늘어나도 α가 12와 14일 때는 어느 정도 유의미한 성능을 얻지만, α가 10일 때는 B 값이 40 정도만 되어도 기학습된 모델의 정확도가 49.05%로 성능에 큰 손실이 있었다. 이는 특정 precision parameter α에 대해서는 B 값 역시 정밀하게 조절해야 올바른 근사 모델을 얻을 수 있다는 것을 의미한다.

B 값이 매우 작으면 정확도는 올라가겠지만, 기학습된 모델의 파라미터들이 어떤 값인가에 따라 B 값을 줄이는 데에 한계가 있다. 파라미터들의 값이 크면 모든 값들이 근사 범위 $-B\le x\le B$에 들어오기가 쉽지 않기 때문이다. 따라서, precision parameter α와 별개로 B에 대한 값 역시 정밀한 조절이 필요한 것으로 요구된다.

Ⅲ. 결론

본 연구에서는 완전 동형 암호 상에서 딥 러닝 모델을 적용하기 위해, ReLU 함수를 다항식으로 근사한 상황에서 근사 범위가 성능에 어떤 영향을 주는지 확인하는 시뮬레이션을 ResNet-20 모델 상에서 진행하였다. 그 결과, 특정 precision parameter α에 대해서는 근사 범위를 정하는 데 필요한 값인 B에 민감하게 성능이 변화하는 것을 확인하였다. 따라서 B 값 역시 기학습된 모델의 성능에 중요한 역할을 하고 있음을 알 수 있었다.

참 고 문 헌

[1] Gentry, C. (2009, May). Fully homomorphic encryption using ideal lattices. In Proceedings of the forty-first annual ACM symposium on Theory of computing (pp. 169-178).

[2] Cheon, J. H., Kim, A., Kim, M., & Song, Y. (2017, December). Homomorphic encryption for arithmetic of approximate numbers. In International Conference on the Theory and Application of Cryptology and Information Security (pp. 409-437). Springer, Cham.

[3] Gilad-Bachrach, R., Dowlin, N., Laine, K., Lauter, K., Naehrig, M., & Wernsing, J. (2016, June). Cryptonets: Applying neural networks to encrypted data with high throughput and accuracy. In International conference on machine learning (pp. 201-210). PMLR.

[4] Hesamifard, E., Takabi, H., & Ghasemi, M. (2017). Cryptodl: Deep neural networks over encrypted data. arXiv preprint arXiv:1711.05189.

[5] Chou, E., Beal, J., Levy, D., Yeung, S., Haque, A., & Fei-Fei, L. (2018). Faster cryptonets: Leveraging sparsity for real-world encrypted inference. arXiv preprint arXiv:1811.09953.

[6] Lee, J., Lee, E., Lee, J. W., Kim, Y., Kim, Y. S., & No, J. S. (2021). Precise Approximation of Convolutional NeuralNetworks for Homomorphically Encrypted Data. arXiv preprint arXiv:2105.10879.

[7] Lee, E., Lee, J. W., No, J. S., & Kim, Y. S. (2020). Minimax Approximation of Sign Function by Composite Polynomial for Homomorphic Comparison. IACR Cryptol. ePrint Arch., 2020, 834.

[8] Cheon, J. H., Kim, D., & Kim, D. (2020, December). Efficient homomorphic comparison methods with optimal complexity. In International Conference on the Theory and Application of

Cryptology and Information Security (pp. 221-256). Springer, Cham.

네트워크 플로우 데이터 기반 이상징후 탐지 인공지능 모델 성능 비교

박태정*, 최태정* , 서성관**, 손배훈**, 윤주범***

xowjdpk@gmail.com taeda593@gmail.com *
ths3847@gmail.com sbh960701@gmail.com** jbyun@sejong.ac.kr***

Performance Comparison of Anomaly Detection Using Artificial Intelligence Model based on Network Flow Data

Tae Jeong Park, Tae Jeong Choi* Seong Kwan Seo, Bae Hun Son **

Joo Beom Yun***

Sejong Univ(Undergraduate Student)* Sejong Univ(Graduate Student)**

Sejong Univ(Professor) ***

요 약

네트워크 기반 공격의 위험성은 서비스에 치명적이며 사회적, 경제적 손실을 야기한다. 따라서 네트워크를 위협하는 여러 공격에 대한 사전 관리와 예방이 손실을 줄일 수 있는 방법이다. 본 논문에서는 네트워크 플로우 기반 데이터로 이상징후를 탐지하는 인공지능 모델 SVM, AutoEncoder , LSTM AutoEncoder, Isolation Forest를 설계하고 그 성능을 비교하였다. 데이터는 플로우 기반 네트워크 데이터인 CICIDS 2017을 사용하여 인공지능 모델을 학습시키고 테스트했다. SVM , AutoEncoder , LSTM AutoEncoder , Isolation Forest 중 LSTM AutoEncoder 가 F1-Score 95.21 %로 가장 뛰어난 성능을 보였다.

Ⅰ. 서 론

네트워크 트래픽 이상탐지는 서비스 및 서버 보호를 위한 정책 중 가장 중요 한 구성 요소로서 정상 동작을 따르지 않는 플로우을 빠르고 정확하게 탐지하는 것이 무엇보다 중요하다. 불규칙한 패턴을 가진 플로우 데이터를 보다 잘 분류하기 위해서는 정보의 특성을 모델링할 수 있는 신경망 알고리즘이 필요하다.

이에 본 논문에서는 지도학습 모델인 SVM 과 비지도 학습 모델 Autoencoder (AE),LSTM AE(Autoencoder), Isolation Forest(I-Forest) 를 활용 한 네트워크 트래픽 이상탐지 모델을 설계하고 알고리즘 간 성능을 비교하였다.

Ⅱ. 본론

2.1 데이터 처리

인공지능 모델을 학습시키기 위해서는 많은 양의 데이터 샘플이 필요하다. 실험 한 네트워크 트래픽 이상탐지 신경망을 훈련하고 검증하기 위해 정상데이터와 DoS Golden Eye, DoS Hulk, DoS Slowloris, DoS SlowHTTP Test, Heart bleed로 진단 된 데이터 Intrusion Detection Evaluation Dataset (CIC-IDS2017) 을 사용하였다. 인공지능 모델에 입력으로 사용하기 위한 데이터 전처리 단계는 총 세단계로 구성 하였다.

첫번째 단계는 One-hot-encoding 으로 정상, 이상 데이터 라벨에 대응하는 값을 0 과 1로 구분 하였다. 두번째는 네트워크 플로우 기반 데이터를 Correlation 에 기반하여 Feature 를 추출하였다.

$$x_{i_}new = \frac{x_i - mean(x)}{stdev(x)} \qquad (1)$$

세번째로 추출 된 Feature 에서 데이터의 값을 식 (1)을 이용하여 데이터의 평균을 0, 분산을 1로 변환하여 표준화하는 과정을 거쳤다.

2.2 실험 인공지능 모델 구조

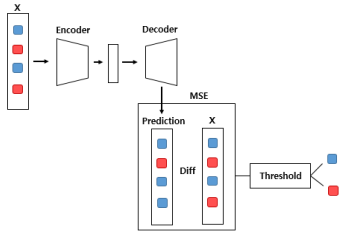


그림 1. AE 와 LSTM AE 모델의 신경망 구조

AE 와 LSTM AE 모델의 신경망은 인코더

계층과 디코더 계층으로 구성되었다. 정상 데이터가 신경망에 입력되면 인코더 층에서 압축되고 디코더 층에서 입력과 똑같이 복원되는 것을 목표로 복원 데이터의 손실을 최소화하는 과정을 통해 정상 데이터의 특징을 학습한다. 그림 1은 테스트 데이터 X가 AE와 LSTM AE 모델의 학습된 신경망에 입력되는 과정을 나타낸 것이다. X는 정상과 비정상 데이터로 구성되어 있다. X가 신경망을 거쳐 복원값(Prediction)이 반환되면 원본 X와 복원값을 비교하여 그 차이가 임계값(Threshold) 보다 클 경우 비정상 데이터로 분류하고, 작은 경우 정상 데이터로 분류하며 이상탐지를 수행한다.

SVM 모델은 정상과 비정상 데이터를 입력받아 결정 경계를 정의하고 테스트 데이터 입력시 결정 경계와 비교하여 정상과 비정상으로 분류하여 이상탐지를 수행한다.

Isolation Forest(I-Forest) 는 랜덤하게 차원을 선택하여 공간을 분할한다. 분할 된 공간에 군집 화 된 정상치는 공간내에서 많은 공간분할이 필요하지만 이상치는 정상 군집에서 멀리 떨어져 있으므로 적은 횟수의 공간 분할만으로 이상치를 고립 (isolation) 시킬 수 있다. 이 과정은 그림2의 Decision Tree 로 표현되는데 정상치 일수록 완전히 고립시키도록 Tree 의 깊이가 높아진다.

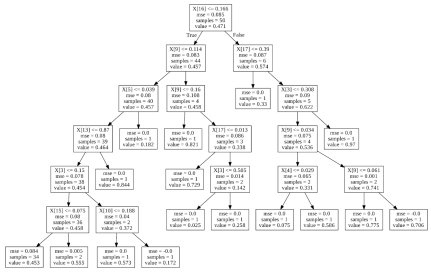


그림2. Isolation Forest Decision Tree

2.3 실험 및 성능비교

실험 데이터는 학습 60%, 검증 10%, 테스트 30%로 분할하였으며 검증의 타당성을 높이기 위해 모든 신경망 알고리즘에 대해 10겹 교차검증을 수행하였다. 표1 는 10겹 교차검증을 수행한 알고리즘 이상탐지 성능 평균을 나타낸다. 이상 탐지 신경망의 성능 평가 지표로 Accuracy, Precision, Recall, F1 Score를 사용하였다.

Model	Accuracy(%)	Precision(%)	Recall (%)	F1-score(%)
AE	92.83	100	89.71	94.53
I- Forest	63.52	70.86	82.40	76.19
SVM	84.57	75.91	96.46	84.96
LSTMAE	95.33	95.41	95.22	95.31

표 1. 알고리즘 이상탐지 성능

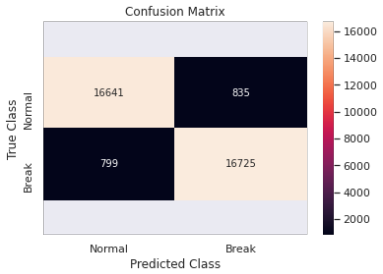


그림 3. LSTM-AE 이상 탐지 혼동행렬

LSTM AE을 적용한 이상탐지에서 95.33% Auccuracy를 기록하였고 이는 다른 알고리즘과 비교하여 2% 이상 높은 수치이다. LSTM AE는 Recall 과 F1-score 에서도 다른 알고리즘의 성능보다 높은 수치를 보여준다. 그림2는 LSTM-AE로 분류한 CICIDS 2017 데이터셋의 혼동행렬을 나타낸다. 약 5% 정도의 오분류 확인할 수 있다.

Ⅲ. 결론

본 논문에서는 네트워크 플로우 데이터를 사용하여 비정상 트래픽을 탐지하기 위한 네 가지 이상탐지 인공지능 모델을 비교했다. SVM, AE , LSTM AE , Isolation Forest 모델을 학습시켜 네트워크 플로우 데이터에 대한 이상탐지 성능을 측정하였다.

실험 결과 LSTM AE가 다른 인공지능 알고리즘보다 높은 성능을 보여준다. 이는 네트워크 이상탐지 알고리즘에 LSTM AE 가 적용 될 수 있음을 의미한다. 이후 LSTM AE 성능개선/최적화를 위한 연구를 진행 할 계획이다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터지원사업의 연구결과로 수행되었음“(IITP-2021-2018-0-01423)

참 고 문 헌

[1]K.J. Kim, "A survey of deep learning-based network anomaly detection," Cluster Computing, vol.27, pp. 949-961, Jan. 2019.

[2] T. N. Sainath, O. Vinyals, A. Senior and H. Sak, "Convolutional, long short-term memory, fully connected deep neural networks," Acoustics, Speech and Signal Processing, IEEE, pp. 4580-4584, 2015.

[3] Lombardi, M. Milano, and L. Benini, "Anomaly detection using autoencoders in high performance computing systems", In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, 2019.

이종데이터 연계 기반 계층적 범죄 네트워크 분석

지종호 [1], 김명준 [1], 박명건 [1], 신현정 [1*]
[1*]아주대학교

{baical77, junkim930, kack7090}@ajou.ac.kr, *shin@ajou.ac.kr

Hierarchical Crime Network Analysis with Heterogenous Data

Jong Ho Jhee[1], Myungjun Kim[1], Myungeon Park[1], Hyunjung Shin[1*]
[1*]Ajou Univ.

요 약

국내·외 범죄 수사와 분석을 위해서는 범죄자들과 범죄 사건에 대한 통합적 정보가 필수적이다. 하지만 이러한 정보의 대부분이 개인 정보 보호와 더불어 개인별, 그리고 사건별로 나뉘어져 있어, 필요한 상황에서는 분절되고 단편적인 정보만을 제공할 수 밖에 없다. 이러한 이종 정보 데이터를 통합한 범죄관련 네트워크 분석을 통해 인물 혹은 사건의 연계성, 중요성 등을 정확하게 인지함으로써 효율적·효과적인 범죄 수사에 도움을 줄 수 있다. 본 연구에서는 가상의 이종 데이터를 연계하여 통합된 범죄 네트워크를 개발한다. 범죄 네트워크 분석을 위하여 그래프기반 계층적 준지도 학습 알고리즘을 활용한다. 범죄사건과 인물 간의 서로 다른 네트워크를 연계한 계층적 네트워크를 분석하고, 이를 바탕으로 인물 네트워크에서 용의자 후보를 추론하는 실험을 진행하였다. 가상 데이터의 용의자와 추론된 용의자 후보의 비교를 통해 용의자 후보 추론의 정확도를 검증하였다.

Ⅰ. 서 론

국내·외 치안 분야에서 범죄 수사와 분석을 위해서는 범죄자들과 범죄 사건에 대한 정보의 통합이 필수적이다. 하지만 이러한 정보의 대부분이 개인 정보 보호로 인한 접근 제한과 더불어 개인별□사건별로 또 기관별로 나뉘어져 있어, 막상 필요한 상황에서는 분절되고 단편적인 정보만을 활용할 수밖에 없다. 이러한 이종 정보 데이터를 통합한 범죄 네트워크 분석을 통해 인물 혹은 사건의 연계성, 중요성, 위험성 등을 파악하여 범죄 수사에 도움을 줄 수 있다[1, 2].

범죄 수사관련 기관의 데이터는 대부분 개별 사건 중심으로 구성되어 있어, 연계·통합을 통한 관리와 분석이 어려운 상황이다. 데이터의 형식 또한 일괄적으로 통합되어 있지 않아 종합적 분석이 여의치 않다. 경찰기관의 수사나 적시 대응을 위해서는, 단편적 정보 즉, 범죄자, 용의자, 피해자, 단일 사건 등의 개별 정보만으로는 충분치 않다. 데이터 통합 분석 시스템의 개발을 통해 범죄자의 색출과 수사가 효과적으로 이루어질 수 있다. 구체적으로는 개별 단위의 수사에서 여력이 미치지 못해 미처 파악하지 못한 상황 혹은 단서들에 대한 정보를 전달할 수 있다[3].

본 연구에서는 먼저 사건 데이터 간의 유사성을 기반으로 하여 노드를 사건, 그 사건 간의 관계를 엣지로 정의하는 사건 네트워크를 구성한다. 또한 인물 데이터로부터 인물 간의 관계 정보를 추출하여 인물(노드), 그 사이를 잇는 선인 엣지(인물 간 관계)로 정의하는 네트워크를 구성한다. 각각의 네트워크는 이종의 데이터를 통합하여 설계한다. 서로 다른 형식의 데이터를 네트워크로 구성한 뒤, 네트워크를 통합하여 하나의 통합된 네트워크를 구성한다. 네트워크의 통합에는 그래프 인테그레이션(Graph integration)을 활용하였다. 이를 통해, 개별 네트워크 정보를 보존하면서 하나의 통합된 네트워크를 생성할 수 있다. 다음으로 생성된 사건 네트워크와 인물 네트워크를 연결하여 계층적 네트워크를 생성한다. 두 네트워크 간의 연결은 사건과 관련된 인물들을 사건에 연결하여 구성하였다. 결론적으로 사건과 인물 데이터를 각각 수평적으로 통합하고 인물/사건 네트워크를 수직적으로 연결한 계층적 구조의 네트워크를 구성한다.

계층적 네트워크에서의 준지도 학습 알고리즘을 통해 각각의 인물들에 대한 스코어(score)를 매겨 용의자 후보 가능성 순위를 도출하였다. 모델의 검증을 위하여 선정된 용의자 후보들과 실제 범죄 네트워크 상의 용의자들을 비교하여 정확도를 검증하였다. 그림 1 은 인물-사건 네트워크의 구조와 이를 이용한 용의자 후보 순위 도출 과정의 프레임워크를 보여준다.

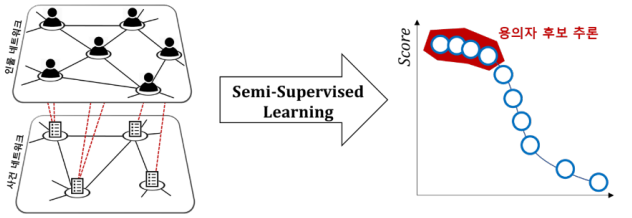


그림 1 인물-사건 네트워크에서 용의자 후보 순위를 스코어 산출을 통해 도출

Ⅱ. 범죄 네트워크 모델

공공기관과의 협업을 통해 획득한 범죄 관련 데이터를 이용하였다. 범죄의 발생 장소, 일시, 사건개요 등의 간략한 정보와 범죄와 관련된 인물들에 대한 기본적인 정보가 포함되어 있다. 각각의 사건 혹은 인물을 노드(node)로 정의하고, 노드와 노드의 관계를 엣지(edge)로 정의함. 노드와 노드의 관계는 유사도를 계산하여 적용함. 이를 통해, 사건이나 인물 데이터를 하나의 네트워크로 표현할 수 있다. 유사도는 similarity metric 을 이용하여 정의한다. 본 연구에서는 cosine similarity 와 Gaussian kernel 을 유사도로 사용하였다.

다수의 네트워크를 통합하기 위하여 그래프 인테그레이션 방법론을 적용한다. 그래프 인테그레이션은 각각의 네트워크들의 정보를 보존하면서 동시에 통합된 하나의 네트워크를 생성하는데 적합한 방법이다[5]. 임의의 네트워크 $G_i(V,E), i=1,\dots,K$가 K개 주어졌을 때, $V=\{v_1,\cdots,v_n\}$는 노드셋, $E\subseteq V\times V$는 엣지셋을 의미한다. 네트워크 통합은 다음의 식에 의해 이루어진다.

$$L(\beta)=\sum_{i=1}^{K}\beta_i L_i$$

여기에서 L_i는 네트워크 G_i의 그래프 라플라시안(graph Laplacian)을 의미하며, L은 통합된 네트워크의 그래프 라플라시안이다. 파라미터 β 값의 조정을 통해 각 그래프의 통합 그래프 내에서의 중요도를 반영하게 된다.

위와 같은 방법을 통해 두 개의 통합 범죄 네트워크(사건 /인물 네트워크)를 생성한다. 범죄사실(사건 텍스트) 포함한 여러 정보를 데이터로부터 추출하여 생성한 사건 네트워크는 전체 노드 500 개와 네트워크 밀도(network density) 약 2.9%로 구성되었다. 또한 인물의 여러 정보를 바탕으로 생성한 인물 네트워크는 전체 노드 1,402 개와 네트워크 밀도 약 5.9%로 구성된다.

Ⅲ. 범죄 네트워크에서의 용의자 후보 추론 실험

범죄(인물-사건) 네트워크에서 임의의 사건(노드) A 를 새로 발생한 사건으로 가정한다. 사건 A 로부터 유사한 사건과의 연결 엣지를 통해 인물 네트워크에 연결된 인물들이 용의자 후보로 선정된다. 주어진 정보 내에서 알려진 용의자들의 스코어(score)를 “1”로 부여하였을 때, 나머지 인물들에게 네트워크 상의 유사도에 따라 스코어가 0 에서 1 사이의 연속적인 값으로 전파되어 알려지지 않은 용의자 후보들에 대한 스코어가 도출된다. 인물들을 스코어 순위에 따라 정렬하여 용의자 후보들을 선정한다.

기존의 그래프기반 준지도학습 알고리즘에서는 계층적 네트워크의 분석에 어려움이 있다[4]. 따라서 용의자 후보를 추론하기 위하여 다계층 네트워크의 준지도 학습 알고리즘을 이용한다[6]. 이 알고리즘은 희소(sparse)한 행렬(네트워크)에서 효율적으로 인물(노드)들에 대한 스코어(score)를 산출해낼 수 있는 방법이다. 네트워크에서의 스코어 산출을 위한 목표함수는 아래와 같다.

$$\min_f (f-y)^T(f-y)+\mu f^T L(\beta) f$$

여기에서 f 는 스코어, y 는 노드의 레이블 값이며, μ 는 네트워크의 평활도(smoothness)를 조절하는 하이퍼 파라미터(hyperparameter)이다.

용의자 후보로 선정된 인물들 중 사건 A 와 관련된 인물들의 포함여부를 확인하였다. 모델의 평가 지표는 AUC 를 사용하였다. 전체 500 건의 사건에 대해서 500 번의 용의자 추론 실험을 반복 진행하는 LOO(Leave One Out) 방법을 수행하였다. AUC 는 평균 0.76 로 나타났다. 그림 2 는 대표적인 사건들의 평균 AUC 를 측정한 결과이다.

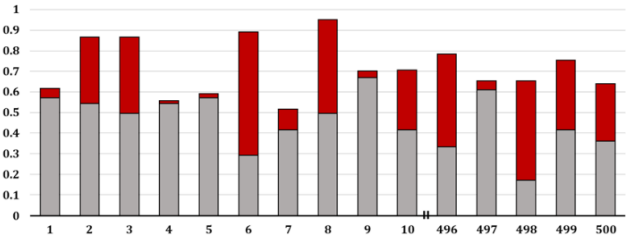


그림 2 모델 성능 측정 결과. 인물-사건 네트워크 간 연결 보강 전 AUC(회색), 인물-사건 네트워크 간 연결 보강 후 AUC(빨강색).

Ⅳ. 결론

본 연구에서는 범죄관련 사건/인물의 계층적 네트워크를 이종의 데이터로부터 설계하였다. 네트워크의 분석을 위하여 그래프 기반의 네트워크 통합 및 계층적 네트워크 준지도학습 알고리즘을 적용하였다. 설계한 계층적 네트워크를 활용하여 용의자 후보들을 스코어를 통해 추론하였다. 그리고 용의자 후보들의 정확도에 대한 실험을 수행하였다. 본 연구를 통해 범죄 프로파일링이 용이해짐으로써 기존 범죄와의 연계성에 대한 식견을 얻을 수 있을 것으로 기대하며. 사건 간 연결을 기반으로 관련된 용의자를 파악하거나, 연결된 인물들을 기반으로 사건 간의 관련성을 파악할 수 있을 것이다.

ACKNOWLEDGMENT

This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIT) (No. (No.2018-0-00440, ICT-based Crime Risk Prediction and Response Platform Development for Early Awareness of Risk Situation).

참 고 문 헌

[1] Chen, Hsinchun, et al. "Crime data mining: a general framework and some examples." computer 37.4 (2004): 50-56.

[2] Xu, Jennifer J., and Hsinchun Chen. "CrimeNet explorer: a framework for criminal network knowledge discovery." ACM Transactions on Information Systems (TOIS) 23.2 (2005): 201-226.

[3] Sparrow, Malcolm K. "The application of network analysis to criminal intelligence: An assessment of the prospects." Social networks 13.3 (1991): 251-274.

[4] Chapelle, Olivier, Bernhard Scholkopf, and Alexander Zien. "Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]." IEEE Transactions on Neural Networks 20.3 (2009): 542-542.

[5] H. Shin, A.M. Lisewski, and O. Lichtarge, “Graph Sharpening Plus Graph Integration: A Synergy That Improves Protein Functional Classification,” Bioinformatics, vol. 23, no. 23, pp. 3217-3224, 2007.

[6] Kim, Myungjun, Dong-gi Lee, and Hyunjung Shin. "Semi-Supervised Learning for Hierarchically Structured Networks." Pattern Recognition (2019).

탈중앙화 신원증명 기반 인증 시뮬레이션 모델에 관한 연구

김호윤, 신승수*
동명대학교
miask376@gmail.com, *shinss@tu.ac.kr

A Study on the Decentralized Identifier-Based Authentication Simulation Model

Ho Yoon Kim, Seung Soo Shin*
Tongmyong Univ.

요 약

본 논문은 탈중앙화 신원증명(DID: Decentralized Identifier)에 기반하여 인증 모델에 관해 연구한다. 온라인에서 사용자 인증 기술이 고도화되고 있지만 대부분 중앙집중형식으로 해킹 공격의 주요 요인이 된다. 분산원장기술을 기반으로 하는 탈중앙화 신원증명 방법을 이용하여 중앙집중형식에서 오는 문제점을 해결하고, 공개키 암호 알고리즘을 사용하여 보안성과 무결성을 해결한다. 특히 인증 모델은 실물 신분증을 대체할 수 있기 때문에 활용도가 높다.

Ⅰ. 서 론

ICT의 발달과 함께 온라인 사용자 인증 기술 또한 고도화되면서 오프라인과 마찬가지로 온라인에서 많은 일들을 처리한다. 코로나-19로 인해 비대면으로 신원을 증명하는 일들이 많은데 이는 실물 신분증이 가지는 한계점을 드러낸다. 온라인에서 신원확인 방법은 ID와 Password를 통해서 인증하거나, 연합 신원 모델(SSO: Single Sign On, 통합로그인)을 이용하여 인증한다[1]. 이는 중앙집중화 되어 사용자에게 편리함을 제공하지만 해킹 사고 시 많은 피해가 발생한다. 중앙집중형의 문제점을 해결하기 위해 분산원장기술의 발달과 이슈로 점차 탈중앙화 시스템으로 발전하고 있다. 탈중앙화 기술이 발전하면서 DID 또한 주목받고 있는데 DID는 자기주권 신원증명을 실현할 수 있다. 탈중앙화 구조를 바탕으로 하여 사용자가 직접 자신의 ID를 관리함으로써 데이터의 주권을 ID 주인에게 부여하여 서비스 제공자가 중앙집중식으로 관리하지 않고 사용자가 직접 관리하는 권한을 가진다.

본 논문에서는 자기주권 신원증명(SSI: Self-Sovereign Identity)을 실현시킬 수 있는 탈중앙화 신원증명 인증 모델을 제안하여 사용자가 직접 ID를 관리하고 데이터의 주권을 ID 주인에게 직접 부여한다[2,3].

Ⅱ. 관련연구

2.1 분산원장기술

분산원장기술은 P2P(Peer-to-Peer) 기술을 기반으로 신뢰할 수 있는 제3자 없이도 데이터를 분산 저장 및 공유하고, 동일한 데이터를 유지하여 무결성을 보장하는 기술이다. 분산원장은 블록체인을 구성하는 가장 중요한 요소 중 하나이며, 블록체인을 탈중앙화된 시스템으로 만들어주는 핵심 기술이다. 거래 기록 등의 데이터를 저장하는 데이터베이스를 중앙화된 서버가 소유하는 것이 아니라, 블록체인에 참여하는 모든 사람이 동일한 원장을 소유하고 관리하는 기술을 일컫는다. 분산원장의 특징은 모든 정보가 암호화되어 Append-only 방식으로만 원장에 저장되기 때문에 한 번 원장에 기록된 정보는 절대 수정할 수 없다. 불가변성의 특성은 블록체인 데이터에 대한 악의적인 변조를 불가능하게 만들어서 데이터에 대한 신뢰도를 향상시켜 준다[4].

2.2 DID

DID는 분산원장기술에 기반한 신원증명이며, 중앙 시스템으로부터 통제받지 않고 개인이 자신의 정보에 대해 통제권을 갖게 하는 디지털 신원관리 체계이다. DID는 DID 주소값과 1:1 관계를 가지는 DID document가 있고, DID document는 분산저장소에 저장한다. 분산저장소는 주로 블록체인을 사용하며 Bitcoin, Ethereum과 같은 비허가형 네트워크뿐만 아니라 Hyperledger와 같이 허가형 네트워크도 사용한다. DID document는 DID 소유 인증에 사용되는 PublicKey 값을 포함한다. DID는 DIF(Decentralized Identity Foundation)의 주도로 개념과 설계가 세워졌고, W3C(World Wide Web Consortium)가 주도하에 국제 표준(Decentralized Identifiers Standards)이 제정되고 있다[5].

Ⅲ. 인증 시뮬레이션 모델

3.1 모델 구성

모델의 주요 참여자는 발행기관(Issuer), 사용자(Holder), 검증인(Verifier), 그리고 DID 관련 정보를 저장하는 분산저장소로 Verifiable Data Registry가 있다. 발행기관은 신뢰기관으로써 사용자가 요청한 신원증명을 DID를 통하여 검증 후 발행한다. 사용자는 신원증명을 필요한 항목만으로 검증인에게 제출하기 위해 재가공한다. 검증인은 서비스 제공자이며 사용자로부터 제공 받은 신원증명의 진위 여부를 분산저장소를 통해 검증한 뒤 서비스를 제공한다.

3.2 구성 요소

사용자와 사용자 인증을 위해 이용되는 DID와 DID document가 있

고 사용자가 발행기관으로부터 발급받는 신원증명 ID 속성으로 VC(Verifiable Credential), 그리고 검증인에게 제출하기 위해 제공 ID 속성으로 VP(Verifiable Presentation)가 있다. VC 항목은 발행기관의 정보, 속성 값, 그리고 발행기관의 서명 데이터가 포함되며 그림 Fig. 1과 같다.

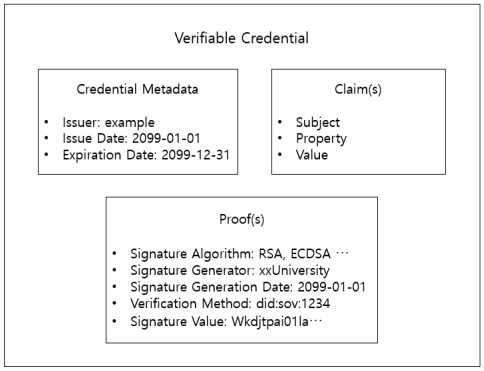


Fig. 1. VC Element

3.3 모델 시나리오

먼저 발행기관과 사용자는 DID Auth를 하기 위한 DID 생성과 공개키를 각각 생성한다. 사용자는 발행기관에 VC(Verifiable Credential) 발급을 요청하고 발행기관은 모바일 본인인증, I-PIN 등 1차 인증을 수행한다. 그리고 사용자의 DID를 통해 DID document 위치를 확인하여 획득한다. DID document를 획득한 뒤 Challenge를 생성하고 사용자에게 DID Auth를 요청한다. DID Auth통해 인증 후 사용자에게 VC를 발급해준다. 사용자는 발급받은 VC를 검증자에게 필요한 항목만으로 제출하기 위해 VP(Verifiable Presentation)로 가공하여 제출한다. 검증자는 제출 받은 VP를 검증하기 위해 Proof 항목으로 분산저장소를 통해 검증한다. 모델의 흐름도는 Fig. 2와 같다.

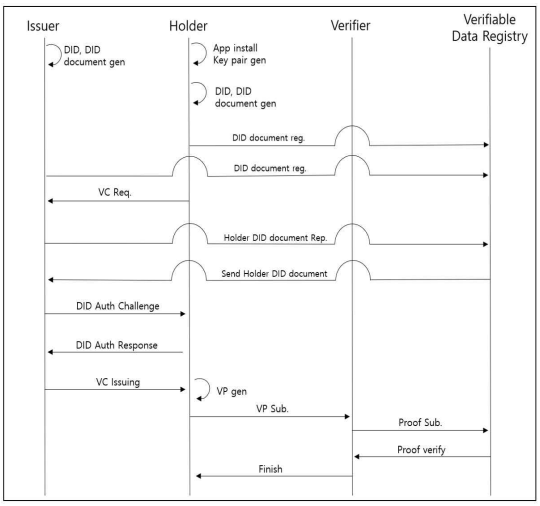


Fig. 2. Proposed System Flow Chart

DID Auth는 매번 다른 값의 Challenge로 진행되며 발행기관은 임의의 난수 값을 생성하여 사용자에게 보낸 뒤 약속된 암호 알고리즘으로 암호화하여 Response를 전달하고 결과 값을 비교해 인증을 수행한다. 이는 스니핑과 재사용 공격에 안전하고, PKI기반으로 중간자 공격에도 안전하다.

Ⅳ.결론

온라인에서 나를 증명하는 신분 확인 기술은 매우 중요하다. 본인인증은 다양한 기술을 이용하여 구현하지만 여전히 문제점이 있다. 기술이 발전함에 따라 편리하면서 보안 위협을 낮출 필요성이 있다. 본 논문에서는 탈중앙화 신원증명 기반 인증 모델을 제시하여 사용자의 데이터 주권을 강화하고자 한다. 제안 모델은 실물 신분증을 대체할 수 있으며 단순히 신분증을 모바일로 옮기는 것이 아닌 분산원장에 DID를 등록 후 필요 시 전자 신분증을 꺼내 이용하여 자기주권 신원증명을 실현하고자 한다. 향후 연구로는 플랫폼 간에 호환성 문제 해결과 표준화 작업에 대한 연구가 필요하다.

ACKNOWLEDGMENT

“This research was supported by the BB21plus funded by Busan Metropolitan City and Busan Institute for Talent & Lifelong Education(BIT).”

참 고 문 헌

[1] W. Li & C. J. Mitchell. (2020). User Access Privacy in OAuth 2.0 and OpenID Connect. IEEE EuroS&PW.

[2] Sovrin Protocol and Token White Paper. (2018). Sovrin Available online: https://sovrin.org/제-content/uploads/Sovrin-Protocol-and-Token-White-Paper.pdf

[3] What is self-sovereign identity. (2018). Sovrin Available online: https://sovrin.org/faq/what-is-self-sovereign-identity

[4] J. K. Lee. (2020). Hyperledger Fabric Configuration and Channel Development Case Study for Google Cloud-based Distributed Ledger Processing. Korean Association Of Computers And Accounting, 18(1), 19-39.

[5] Decentralized identity Foundation. (2019). DIF. https://identity.foundation

변동성 라벨링을 이용한 앙상블 머신러닝 트레이딩 시스템 개발

박명석, 유성주, 김재윤*
순천향대학교

pmsk980122@sch.ac.kr, sjyoo1572@naver.com, kimym38@sch.ac.kr*

Developing an ensemble trading system using the volatility labeling

Myeong Seok Park, Sungju Yoo, Jaeyun Kim*
Soonchunhyang Univ.

요 약

본 논문은 변동성 라벨링을 통한 앙상블 머신러닝 트레이딩 시스템을 제안한다. 변동성 라벨링은 수익률과 표준편차를 이용하여 학습데이터를 라벨링하는 방식이다. 또한 예측 성능 향상과 안정적인 트레이딩 시스템을 구축하기 위해 앙상블 기법을 적용한 결과 전통적으로 활용되고 있는 방법보다 우수한 결과를 도출하였다.

Ⅰ. 서 론

주식 시장 예측이 어려운 이유는 주식 데이터가 비선형성과 비정상성 특징을 가지고 있기 때문이다. 주가 데이터는 일정한 분산과 일정한 평균을 가지지 않으며, 자기상관관계에서 시차의 길이에 따라 예측이 가능하지 않아 통계적 모델만으로는 예측이 어렵고 복잡하다.[1] 이를 극복하기 위해 머신러닝을 이용한 주가 예측 연구가 활발해지고 있다. Rohit Choudhry [2]는 유전자 알고리즘과 SVM(Support Vector Machine)을 활용하여 주식 시장을 예측하기 위하여 다양한 기술적 지표를 입력 변수로 활용하였다. 다른 주가의 상관관계를 이용하여 기술적 지표의 유용성을 증명하고 주가 예측에 대한 성능을 확인했다. Dongdong Lv[3]는 미국 시장의 SPICS 424 개, 중국 시장의 CSICS 185 개를 실험 대상으로 적용하고 44 개의 기술적 지표를 머신러닝의 입력 변수로 활용했다. 머신러닝과 딥러닝 알고리즘의 성능을 비교했을 때, 일부 머신러닝 알고리즘이 주가 방향성 예측 성능이 더 좋은 것을 확인했다. Samuel Asante Gyamerah[4]은 두 개의 머신러닝 알고리즘(Adaptive Boosting 및 K-Nearest Neighbor)와 stacking 앙상블 분류기를 활용하여 주식의 움직임을 예측했다. 개별 머신러닝 알고리즘과 앙상블 분류기의 정확도를 비교했을 때 앙상블 분류기의 정확도가 모든 분류기 중에서 가장 높은 정확도를 보임을 확인했다. Shikha Metha[5] 또한 앙상블 머신러닝 접근 방식을 제안했다. 앙상블 학습 방식이 감소된 분산으로 최대 정확도를 달성할 수 있음을 확인했다.

그러나 위와 같은 연구들은 다음과 같은 한계점을 가진다. 첫째, 사용 데이터의 기간이 짧아 다양한 패턴 학습이 불가 했다는 점이다. 둘째, 주가의 상승과 하락의 기준을 단순히 가격의 차이를 바탕으로 정의했다는 점이다. 단순히 가격의 차이를 바탕으로 라벨을 정의하면 비선형성과 비정상성을 가지는 데이터를 사용하게 되는 문제가 발생한다. 셋째, 주가 예측을 활용한 트레이딩 시스템의 부재이다. 위와 같은 한계점을 바탕으로 본 논문에서는 주가 방향을 기존 방법과 다르게 정의하는 라벨링 방식(변동성 라벨링)과 가격의 차이를 바탕으로 상승과 하락을 정의하는 라벨링 방식(벤치마크 라벨링)을 비교하고 앙상블 머신러닝 트레이딩 시스템을 제안한다.

Ⅱ. 본론

본 논문의 제안 모델은 [Figure 1]과 같다. STEP 1 에서는 주식 데이터를 가져와 기술적 지표를 추출한 뒤 변동성 라벨링 방식을 적용한다. STEP 2 에서는 다양한 머신러닝 알고리즘을 이용해 학습 및 예측을 진행한다. STEP 3 에서는 도출된 예측 값을 이용해 트레이딩 시뮬레이션을 진행하고 트레이딩 시스템을 구축한다.

2.1 STEP 1

주식 데이터는 시가총액 100 위 종목을 가져왔으며 학습 데이터를 동일한 기간 동안 하기 위해 2008 년 주가 정보가 없는 종목은 제외하였다. 기술적 지표는 선행연구에서 소개된 19 개를 사용했다. 변동성 라벨링 공식은 아래와 같으며 α 는 표준편차의 계수를 의미하며 σ 는 표준편차를 의미하고 n_preiod 는 수익률의 기간을 나타낸다.

$$DOWN = -\alpha\sigma \leq n_preiod\ profit\%$$

$$UP = +\alpha\sigma \geq n_preiod\ profit\%$$

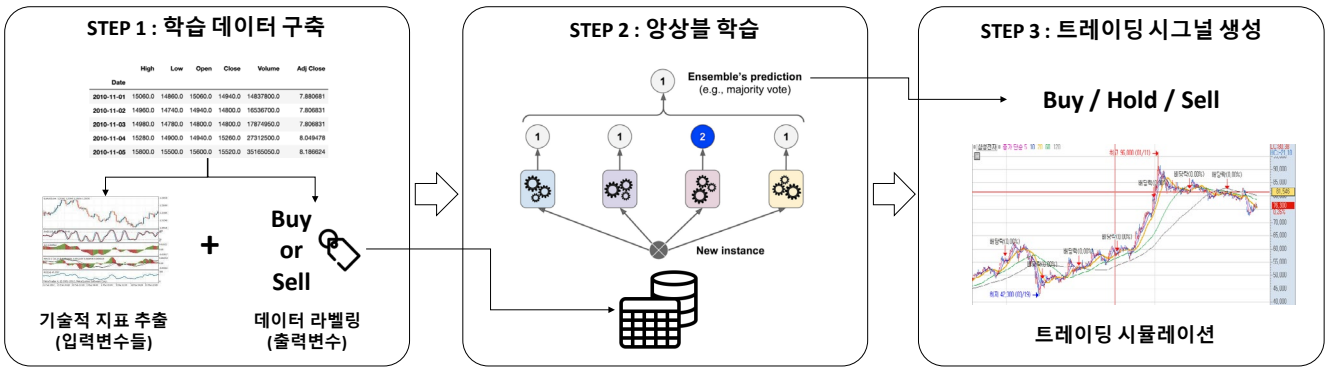


Figure 1 앙상블 트레이딩 시스템

2.2 STEP 2

앙상블 머신러닝 모델은 대표적인 방법인 보팅, 배깅, 부스팅 방법을 적용하였다. 보팅에 사용되는 알고리즘은 Decision tree, Logistic regression, Navie bayes, K-nn, MLP 사용하였으며 배깅은 Randomforest, 부스팅은 GBM 을 사용한다.

2.3 STEP 3

학습된 모델을 통해 예측 값을 생성하고 생성된 예측 값을 통해 트레이딩 시그널을 생성해 트레이딩 시뮬레이션을 진행한다. 매수 포지션이 발생하면 다음날 시가로 매수하며, 매도 포지션이 발생하면 다음날 시가로 매도한다. 매수 포지션이 발생하고 매도 포지션이 나오기 전까지는 보유하는 전략을 가지고 매도 포지션이 발생하고 매수 포지션이 나오기 전까지는 아무런 포지션을 갖지 않는다. 실거래와 유사한 결과를 도출하기 위해 거래 수수료는 0.015%를 부과하였다. 강한 매수 포지션과 매도 포지션을 잡기 위해 예측 임계 값을 down 은 0.25, up 은 0.75 로 조정한 뒤 진행했다. 임계 값을 변경한 이유는 강한 상승 신호와 하락 신호를 잡기 위함이다. 모델의 성능을 평가하기 위해 투자 성과 지표를 사용한다. 투자 성과 지표는 총 6 개로 거래 횟수(합계), 승률, 평균 수익, 평균 손실, Payoff ratio, Profit factor 가 있다.

Ⅲ. 결론

트레이딩 시뮬레이션을 실행한 결과는 Table 1, 2 과 같다. 벤치마크 라벨링과 변동성 라벨링의 결과를 비교해보면 평균 수익, 평균 손실, Payoff ratio, Profit factor 에서 벤치마크 라벨링 결과보다 변동성 라벨링의 성과가 높은 것을 확인할 수 있다. 변동성 라벨링에서 Payoff ratio 와 Proift factor 의 성능이 벤치마크 라벨링의 비해 개선이 된 것을 볼 수 있다. 변동성 라벨링에서 제안한 앙상블 모형인 보팅 모형이 다른 앙상블 기법인 배깅과 부스팅의 성과보다 높은 것을 확인할 수 있다. 이는 제안한 보팅 모형이 안정적인 수익을 가져간다는 것을 의미하며 다른 머신러닝 알고리즘보다 성능이 좋은 것을 확인할 수 있다.

Table 1 벤치마크 라벨링을 이용한 트레이딩 결과

모델	거래 횟수	승률	평균 수익	평균 손실	Payoff Ratio	Profit factor
Decision tree	266	0.48	3104.65	2878.16	1.06	0.99
K-nn	275	0.48	2948.85	2737.96	1.05	0.98
Naive Bayes	96	0.63	3338.36	4875.03	0.89	1.42
MLP	165	0.53	2962.44	2996.49	0.98	1.13
Logistic regression	130	0.57	3316.58	3187.07	1.03	1.28
Voting	164	0.53	2925.27	2992.88	1.00	1.15
GBM	206	0.51	2952.76	2918.26	1.03	1.06
Random forest	216	0.51	2932.76	2913.82	1.04	1.10

Table 2 변동성 라벨링을 이용한 트레이딩 결과

모델	거래 횟수	승률	평균 수익	평균 손실	Payoff Ratio	Profit factor
Decision tree	286	0.47	5444.10	4572.17	1.17	1.04
K-nn	195	0.45	6647.84	5197.36	1.30	1.09
Naive Bayes	212	0.45	6006.77	4852.66	1.26	1.07
MLP	160	0.44	7528.58	5616.73	1.38	1.14
Logistic regression	136	0.44	7882.94	5862.92	1.38	1.13
Voting	**185**	**0.44**	**8266.17**	**5740.48**	**1.42**	**1.29**
GBM	283	0.47	5517.42	4612.71	1.17	1.03
Random forest	153	0.45	7392.22	5625.25	1.36	1.13

ACKNOWLEDGMENT

본 연구는 2021 년 과학기술정보통신부 및 정보통신기획평가원의 SW 중심대학사업의 연구결과로 수행되었음(2021-0-01399)

참 고 문 헌

[1] 이형용 "한국 주가지수 등락 예측을 위한 유전자 알고리즘 기반 인공지능 예측기법 결합모형" Entrue Journal of Information Technology 7.2 pp.33-43 (2008) : 33.

[2] Choudhry, Rohit, and Kumkum Garg. "A hybrid machine learning system for stock market forecasting." World Academy of Science, Engineering and Technology 39.3 (2008): 315-318.

[3] Lv, Dongdong, et al. "An empirical study of machine learning algorithms for stock daily trading strategy." Mathematical Problems in Engineering 2019 (2019).

[4] Gyamerah, Samuel Asante, Philip Ngare, and Dennis Ikpe. "On stock market movement prediction via stacking ensemble learning method." 2019 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr). IEEE, 2019.

[5] Mehta, Shikha, et al. "Ensemble learning approach for enhanced stock prediction." 2019 Twelfth International Conference on Contemporary Computing (IC3). IEEE, 2019.

오토인코더를 이용한 불량 탐지에 관한 연구

김지섭, 김준한, 김상태, 심병효
서울대학교
{jskim, junhankim, stkim, bshim}@islab.snu.ac.kr

A Study on the Defect Detection using Autoencoder

Jiseob Kim, Junhan Kim, Sangtae Kim, and Byonghyo Shim
Seoul National University

요 약

많은 실제 현장에서는 불량 탐지 딥 러닝 모델의 학습을 위한 불량 데이터가 많지 않고 한정되어있다. 이런 경우 지도 학습 방식으로 훈련된 딥 러닝 모델은 특히 테스트 단계에서 불량을 탐색할 때 성능이 저하된다. 본 논문에서는 이러한 한계를 극복하기 위하여 오토인코더를 이용한 불량 탐지 기법을 제안한다. 시뮬레이션을 통하여 제안한 기법이 불량 탐지를 잘할 수 있음을 확인하였다.

Ⅰ. 서 론

최근 딥 러닝을 이용한 불량 탐지 기술에 대한 많은 연구가 이루어지고 있다. 방대한 양의 데이터가 있는 경우 불량 탐지를 위한 종단 간 솔루션을 제공하도록 딥 러닝 모델을 훈련할 수 있다 [1]. 하지만 실제 현장에서는 딥 러닝 모델의 학습을 위한 불량 데이터가 많지 않고 한정되어있다. 이런 경우 지도 학습 방식으로 훈련된 딥 러닝 모델은 특히 테스트 단계에서 불량을 탐색할 때 성능이 저하된다. 본 논문에서는 이러한 한계를 극복하기 위하여 오토인코더[2]를 이용한 불량 탐지 기법을 제안한다. 시뮬레이션을 통하여 제안한 기법이 불량 탐지를 잘할 수 있음을 확인하였다.

Ⅱ. 본론

원본 이미지는 클래스 내 변화, 조도 변화, 시점 변화로 인해 직접 비교하기 어렵기 때문에 이미지의 특징을 추출한 후 추출된 특징을 불량 탐지에 사용하고자 한다.

이미지의 특징을 추출하기 위해 특징 추출에 자주 사용되는 비지도 학습 기반 딥 러닝 모델인 오토인코더를 사용한다. 오토인코더는 인코더와 디코더의 두 개의 주요 부분으로 구성된다. 인코더 f_θ.는 입력 이미지를 더 작은 차원의 특징 벡터로 압축한다. 반면에 디코더 g_ϕ는 인코더의 출력인 특징 벡터를 이용하여 원래 입력 이미지를 복구한다. 오토인코더는 입력 이미지 x를 잘 재구성하는 방향으로 학습된다. 즉, 인코더와 디코더의 매개변수 θ및 ϕ를 학습하여 입력 이미지 x와 복구된 이미지 $\hat{x}= g_\phi(f_\theta(x))$사이의 평균 제곱 오차를 최소화한다.

$$(\theta^*,\phi^*) = \underset{(\theta,\phi)}{\operatorname{argmin}} \quad x-\hat{x} \parallel_2^2 = \underset{(\theta,\phi)}{\operatorname{argmin}} \parallel x-g_\phi(f_\theta(x)) \parallel_2^2$$

입력 이미지와 복구된 이미지가 충분히 비슷하다면 인코더의 특징 벡터는 원본 이미지를 복구하기 위한 충분한 정보를 가지고 있다.

오토인코더를 훈련시킨 후 재구성 오류를 최소화하기 위하여 모든 양품 이미지들의 특징 벡터를 계산한 후 평균을 취한다. 또한 모든 양품 이미지들의 특징 벡터와 평균 특징 벡터 사이의 유클리드 거리를 측정하고 최대값을 임계값 τ로 설정한다.

테스트 단계에서는 입력 이미지가 제공된다. 먼저 인코더를 통하여 해당 이미지의 특징 벡터를 추출한 다름 추출된 특징 벡터와 평균 특징 벡터 사이의 유클리드 거리를 측정한다. 만약 측정된 유클리드 거리가 임계값 τ보다 크다면 해당 이미지에 불량이 있다고 판단한다.

우리는 시뮬레이션을 통해 제안된 기법의 성능을 측정하였다. 훈련 단계에서는 양품 화면 이미지만 사용하여 훈련하였으며 테스트 단계에서는 네 가지 종류의 불량 이미지가 사용되었다.

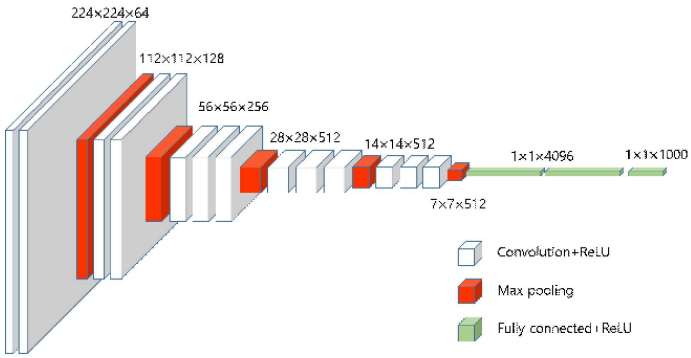


그림 1. 일반적인 VGG 네트워크의 구조

인코더로는 ImageNet 데이터셋에서 사전 훈련된 VGG 네트워크를 사용하였다 [3]. VGG 네트워크의 구조는 그림 1.과 같다. 인코더의 입력은 256x1024 사이즈의 RGB 이미지이며 출력은 131,072 차원으로 구성된 특징 벡터이다. 디코더의 구조는 인코더와 대칭으로 설계하였다. 우리는 3,000개의 양품 이미지를 이용하여 오토인코더를 훈련하였다.

표 1에는 시뮬레이션 결과가 요약되어 있다. 모든 불량 이미지들이 불량으로 성공적으로 분류되어 제안된 기법이 불량 탐지를 위해 효과적임을 확인할 수 있다.

표 1. 불량 탐지 시뮬레이션 결과

Class ID	Defect features	Total features
1	2,123	131,072
2	421	131,072
3	126	131,072
4	324	131,072

Ⅲ. 결론

본 논문에서는 오토인코더를 이용한 불량 탐지 기법을 제안하였다. 시뮬레이션 결과에 따르면 제안된 VGG 기반 오토인코더를 사용한 불량 탐지 기법은 적은 수의 학습 데이터 환경에서도 불량 탐지에 효과적임을 확인하였다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터육성지원사업의 연구결과로 수행되었음 (IITP-2021-2017-0-01637)

참 고 문 헌

[1] Chalapathy, R., & Chawla, S. (2019). Deep learning for anomaly detection: A survey. arXiv preprint arXiv:1901.03407.

[2] Sun, W., Shao, S., Zhao, R., Yan, R., Zhang, X., & Chen, X. (2016). A sparse auto-encoder-based deep neural network approach for induction motor faults classification. Measurement, 89, 171-178.

[3] Karen S., Andrew Z., "Very Deep Convolutional Networks for Large-Scale Image Recognition", Computer Vision and Pattern Recognition 2015.

해밍부호 기반 이진인공신경망 학습

이선용, 노종선
서울대학교 전기정보공학부,뉴미디어통신공동연구소

sonyongyi@snu.ac.kr, jsno@snu.ac.kr

Binary neural network learning strategy via hamming distance

SeonYong Lee, Jong-Seon No
INMC, Dept. of ECE, Seoul National Univ.

요 약

본 논문은 이진인공신경망 네트워크의 성능을 개선하기 위해서 새로운 학습 방법을 제시한다. 새로 제시된 방법을 통해 학습한 이진신공망 네트워크는 별다른 구조적 변화 없이 classification 층의 해밍부호 기반 classification 을 통해 좋은 학습성능을 보인다.

Ⅰ. 서 론

이진 인공 신경망은 (Binay Neural Network, 이하 BNN) 2016 년 Itay hubara, Yoshua Bengio 에 의해 제안된 메모리 사용을 극단적으로 제한한 인공신경망 구조이다[1]. 인공신경망 네트워크는 Hinton, Bengio, Lecun 에 의해 인공지능 분야의 혁신을 이루었고, 이미지 인식 및 기타 분야에서 사람을 뛰어넘는 성능을 보여주고 있다. GPU 의 발전으로 빠른 다중연산처리가 가능해졌기 때문에 머신러닝 기술이 대두되었으나, 그 계산량과, 전력소모등에 의해 실제 사용할 때 많은 단점이 따른다. 이에 인공신경망 네트워크를 경량화 하는 방법을 많은 사람들이 연구하고 있다. 그 중 이진인공신경망 네트워크는 기존에 제시된 모델들을 그대로 사용할 수 있어, 기존 모델들이 가진 특성을 반영할 수 있고, 학습 방법, 가중치등을 이용 할 수 있기 때문에 범용성이 높아 네트워크 경량화 분야에서 높은 가능성을 가지고 있다.

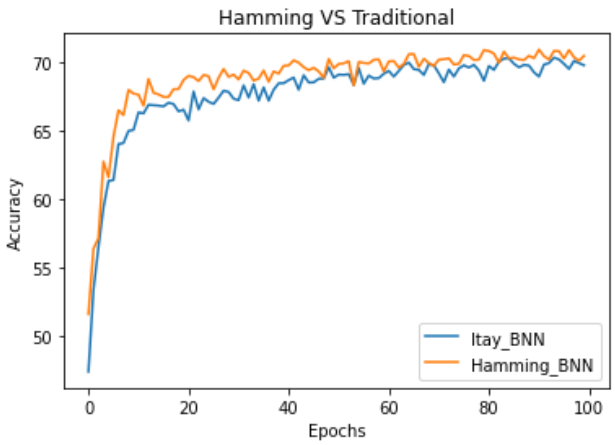
이진 인공 신경망은 모든 가중치를 이진화 하기 때문에 이에 알맞은 방식의 학습방법을 연구해야 한다는 의견이 존재한다. 그에 따라 Binary optimizer 라 불리우는 방식이 나왔고[2], 그 외에도 다양하게 연구되고 있다.

합성곱 인공신경망에서 분류층의 결과를 해밍부호로 나타내었을 때 별다른 조작 없이 adversarial attack 과 학습에서 좋은 결과를 나타냄이 알려져있다[3].

본 논문에서는 위의 결과들에서 착안해 이진 인공신경망의 학습단계에서 분류층을 해밍부호화하여 학습과정을 개선시킬 수 있음을 보였다.

Ⅱ. 본론

본 논문에서는 BNN 의 분류층을 해밍부호화 하여 성능을 개선하는 방법을 실험했다. 실험데이터셋은 CIFAR-10 데이터셋을 이용했고, BNN 구조는 VGG-small 네트워크를 사용했다. 데이터셋의 정규화 이외에는 다른 전처리를 사용하진 않았다. Classification layer 의 해밍 부호는 (7,4) 해밍부호를 이용했고, 10 개의 classification 을 (7,4) 해밍부호를 통해 나타내 주었다.



Ⅲ. 결론

본 논문에서는 해밍부호를 기반으로 분류층을 나타내는 방법을 BNN 에서 적용했고, 적용한 결과가 traditional BNN 보다 좋은 성능을 보여줌을 보였다. 추후의 연구로서 cifar-100, imagenet 등 다양한 분류를 요하는 데이터셋에 대해 RS 부호, LDPC 부호등 다양한 부호를 BNN 구조에 적용하는 연구가 수행될 예정이다

참 고 문 헌

[1] Hubara I, Courbariaux M, Soudry D, El-Yaniv R, and Bengio Y." Binarized Neural Networks," NeurIPS, pp.4107-4115, 2016.

[2] Helwegen K, Widdicombe J, Geiger L, Liu Z, Cheng K-T, and Nusselder, R." Latent Weights do not exist: Rethinking Binarized Neural Network Optimization," NeurIPS, 2019.

[3] Gunjan V, Ananthram S. "Error Correcting Output Codes Improve Probability Estimation and Adversarial Robustness of Deep Neural Networks, " NeurIPS, 2019

스마트 트렁크 오프너 시스템용 딥러닝과 레이더 시그너쳐를 이용한 발동작 인식 기술

송승언, 김봉석, 김상동, 이종훈*

첨단레이더연구소, 융합연구원, *대구경북과학기술원

*jhlee@dgist.ac.kr

Foot Movement Detection Scheme Utilizing a New Radar Signature and Deep Learning for Smart Trunk Opener Systems

Seungeon Song, Bong-seok Kim, Sangdong Kim, Jonghun Lee*

Advanced Radar Tech. Lab. Research Institute of Convergence, *DGIST.

요 약

본 논문은 스마트 트렁크 오프너 시스템을 동작시키는 데 필요한 새로운 레이더 시그너쳐와 딥러닝을 이용한 비접촉식 발동작 인식기술을 제안한다. 스마트 트렁크 오프너 시스템은 미래형 스마트 자동차 기술 중에서 양손을 자유롭게 사용할 수 없는 환경 또는 조건에서도 손대신 발을 사용함으로써 발동작을 통해서 사용자의 의도를 파악할 수 있는 기술이다. 본 논문에서는 사용자의 의도를 정확하게 파악하기 위해서 새로운 레이더 시그너쳐와 기존 딥러닝 모델을 활용한 transfer learning을 이용함으로써 다양한 발동작에 대한 인식 성능을 향상하고자 한다.

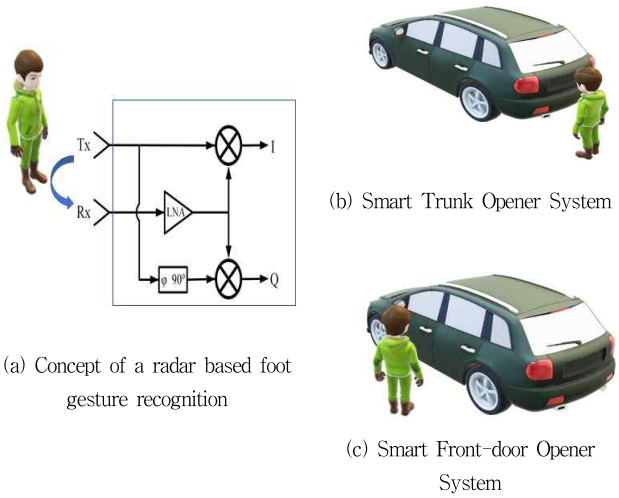
Ⅰ. 서 론

최근 레이더는 군수용뿐만 아니라 민수용으로 다양하고 폭넓게 응용이 되고 있다. 특히, 자율주행 자동차용 능동안전장치, 생체신호감지, 제스처 인식 등에 사용된다. 도플러 레이더는 연속적인 레이더신호를 이용하여 움직임이 있는 타겟에 대한 민감도가 높아서 다양한 제스처 인식용으로 크게 주목을 받고 있다.

특히, 다양한 제스처 인식중 발동작 인식기술은 양손의 사용에 제약이 있는 환경에 놓인 상태에서는 무엇보다도 편리하고 요긴한 기술이다. 기존 발동작 인식용 센서는 초음파, 라이다, 카메라, 레이더 등이 사용되고 있다. 이러한 센서 중에서 레이더는 조명 및 주변 환경 특히 습기 및 먼지 등에 대한 잡음에 강인하다는 강점이 있다. 이러한 장점에 의해서 STO(Smart Trunk Opener system), SDO(Smart Door Opener), 상지 장애인용 HMI(Human Machine Interface) 응용분야에 크게 주목받고 있다 [1-3]. 본 논문에서는 양손을 자유롭게 사용할 수 없는 상태에서 자동차의 트렁크를 자동으로 개폐하는 발동작 인식기술을 제안한다. 또한, 킥 동작 인식에 오류를 일으킬 수 있는 스윙 또는 굴러가는 물체와 구분할 수 있도록 레이더 센서를 이용한 딥러닝 기반 킥 동작인식 기술을 제안하고자 한다.

Ⅱ. 발동작 인식용 레이더 시그너쳐 및 인식 성능 평가

그림1은 레이더센서를 이용한 STO 시스템의 기본 개념을 나타내는 그림이다. 그림 1에서 알 수 있듯이, STO와 SDO 응용에는 다양한 인식용 발동작이 고려될 수 있다. 본 논문에서 고려중인 발동작은 킥 동작, 스윙 동작, 위아래로 빠르게 움직이는 태핑 동작, 그리고, 센서 밑으로 튀어서 지나가는 공 등이 고려된다. 특히, 센서 밑으로 튀어서 지나가는 공은 발동작과 상당히 유사한 동작으로 STO의 인식성능에 상당한 영향을 줄 것으로 여겨지기에 특별히 고려한 움직임이다. 또한, 움직이는 물체의 변화에 민감하게 반응하고 센서의 상용화를 위해서 저가형 센서를 고려해서 CW(Continuous Wave) 레이더 센서를 발동작 인식용으로 사용한다.



(a) Concept of a radar based foot gesture recognition

(b) Smart Trunk Opener System

(c) Smart Front-door Opener System

Fig. 1 레이더기반 발동작 인식 센서 개념 및 응용분야

그림2는 STO 시스템에 활용될 수 있는 다양한 발동작에 대한 레이더 시그너쳐를 나타내고 있다. 본 레이더 시그너쳐는 도플러 레이더를 통해서 발동작에 의해서 반사된 레이더 신호에 대한 시간-주파수 특성을 나타낸다. 본 논문에서는 발동작별 인식 성능을 향상시키기 위해서 기존 Alexnet 모델 기반으로 transfer learning을 적용한다. 획득된 레이더 시그너쳐는 Alexnet-transfer learning 모델의 입력으로 사용된다.

획득된 레이더 시그너쳐는 발동작별로 700개의 데이터를 확보함으로써 총 4,200개의 데이터를 확보하였다. 훈련용으로 발동작별 600개의 레이더 데이터를 사용하고 시험용으로 발동작별로 100개의 레이더 데이터를 사용하였다.

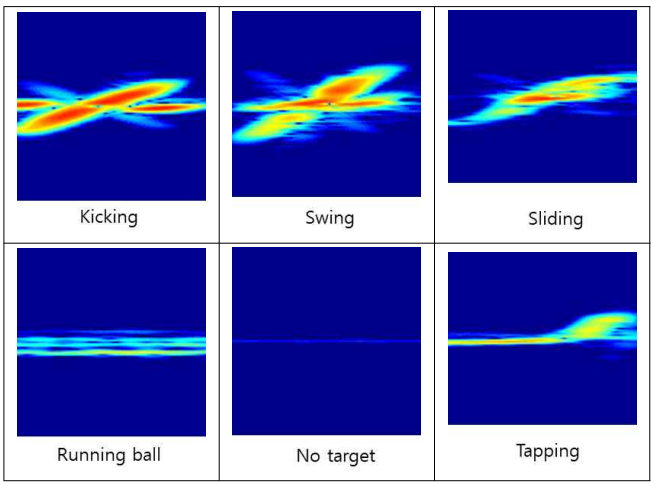


Fig. 2 STO용 다양한 발동작에 대한 레이더 시그너쳐 특성

그림 3은 Alexnet 딥러닝 모델을 기반으로 다양한 발동작에 대한 인식성능을 나타내는 confusion matrix를 나타낸다. 그림 3에서 알 수 있듯이, STO 시스템용 레이더 기반 발동작 인식 정확도는 대략 96%이상임을 알 수 있다.

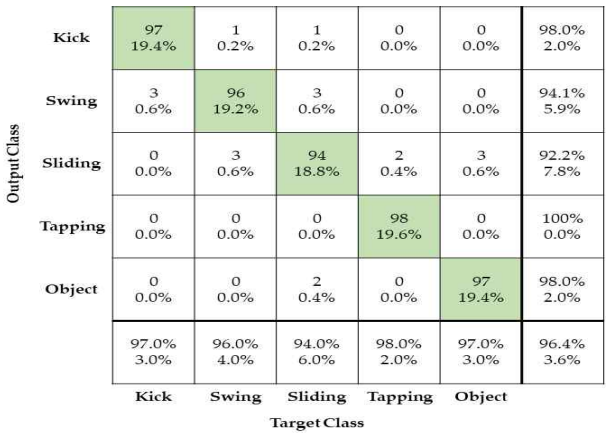


Fig. 3 레이더기반 발동작 인식성능용 confusion matrix

Ⅲ. 결론

본 논문에서는 스마트 트렁크 오프너 시스템용 딥러닝기반 레이더를 이용한 발동작 인식기술을 제안하였다. 본 연구를 통해서 STO용 다양한 발동작에 대한 시간 및 주파수 특성에 레이더 시그너쳐를 분석하고, SVD(Singular Value Decomposition)를 통해서 유의미한 정보를 추출하고, 그 특징정보를 포함하는 새로운 레이더 이미지를 재구성한 레이더 시그너쳐를 기존 딥러닝 모델인 Alexnet의 transfer learning을 통해서 우수한 발동작 인식성능을 확보하였다.

향후, 레이더를 이용하는 발동작 인식 센서는 레이더의 고유한 강점인 주변 환경과 잡음에 대한 강인한 특성 때문에 미래자동차의 다양한 비접촉식 제스쳐 인식 및 HMI 기술 등에 폭넓게 적용될 수 있을 것으로 기대된다.

ACKNOWLEDGMENT

※ 본 논문은 과학기술정보통신부에서 지원하는 대구경북과학기술원의 기관고유사업(21-IT-02), 5G기반 자율주행 융합기술 실증 플랫폼(P0013840)의 지원에 의해 연구되었음.

참 고 문 헌

[1] Shankar Y. and Santra A. “Valid Kick Recognition in Smart Trunks Based on Hidden Markov Model Using Doppler Radar,” In Proceedings of the 2019 International Radar Conference, RADAR 2019, 2019.

[2] Liu H., Yang R., Yang Y., Hou C., Hu Z. and Jiang T. “Human - Human Interaction Recognition Based on Ultra-Wideband Radar,” Signal, Image and Video Processing 2020.

[3] Chiang H.P., Lai C.F., Lai Y.H., Huang Y.M., “A Sensor-Based Feet Motion Recognition of Graphical User Interface Controls,” Multimedia Tools and Applications 2016.

기계학습에 사용되는 테스트 케이스 생성에 관한 고찰

정재호, 노종선
서울대학교 전기정보공학부 뉴미디어통신공동연구소
jaehoj@ccl.snu.ac.kr, jsno@snu.ac.kr

A Study on Test Case Generation for Machine Learning Systems

Jaeho Jeong and Jong-Seon No
INMC, Department of ECE, Seoul National University

요 약

본 논문은 기계학습에 관한 연구 중에서 데이터를 학습, 검증, 시험하기 위해 필요한 테스트 케이스를 잘 설계하고 적용하는 방법에 대한 최신 연구들을 살펴보았다. 현재 진행되고 있는 관련 연구 분야로는 테스트 케이스 생성, 테스트 케이스 감축, 테스트 케이스 우선 순위화 등이 있으며 최신 연구들이 컴퓨터 공학 분야에서 어떤 방법으로 테스트 케이스 생성과 관련된 성능을 발전시켰는지 알아본다.

Ⅰ. 서 론

일반적으로 기계학습(Machine Learning)에 관한 연구라고 하면 기계학습 그 자체의 성능을 높이기 위한 알고리즘 혹은 기계학습을 적용시킬 수 있는 새로운 분야에 관한 연구가 진행되기 마련이다. 하지만 이러한 연구를 진행하기 위해 데이터를 학습(training), 검증(validation), 시험(test)하는 데는 필요한 테스트 케이스(Test Case)를 적절하게 잘 만들고 적용하는 것이 매우 중요하다. 일반적으로는 기계학습에 MNIST, CIFAR-10, CIFAR-100, ImageNet 등 이미지 파일들로 구성된 데이터 세트를 많이 사용하지만, 다른 분야에 적용되는 기계학습의 경우에는 학습을 위한 데이터를 어떻게 설계하느냐에 따라 학습이 잘 안 될 수도 있고, 데이터 세트 전처리를 어떻게 하느냐에 따라 성능 차이가 크게 벌어질 수 있다. 그렇기 때문에 테스트 케이스 그 자체에 대한 연구도 활발히 진행되어야 하는데, 현재 진행되고 있는 관련 연구 분야로는 테스트 케이스 생성(Test Case Generation), 테스트 케이스 감축(Test Case Reduction), 테스트 케이스 우선 순위화(Test Case Prioritization) 등이 있다. 특히 이러한 기계학습용 테스트 케이스를 잘 선별해 내는 데에도 기계학습 그 자체가 다시 사용될 수 있다.

본 논문에서는 테스트 케이스 생성 기법이 적용될 수 있는 컴퓨터 공학 분야들에 대해 먼저 알아본다. 이후 이와 관련된 최신 연구들에서 각각의 분야들에 어떤 방법으로 테스트 케이스 생성과 관련된 성능을 발전시켰는지 분석한다.

Ⅱ. 본 론

총 3종류의 최신 연구를 바탕으로 각각의 연구가 진행된 분야에 대한 소개와 함께 이를 통해 테스트 케이스 생성, 감축, 우선 순위화 방식을 어떻게 적용시켰는지 자세히 알아보도록 한다.

- 소프트웨어 테스트(Software Testing)

소프트웨어 테스트란, 소프트웨어가 제대로 작동하는지, 버그가 발생할 수 있다면 어떤 부분에서 발생할 수 있는지 검증하는 작업이다. 개발이 완료된 소프트웨어를 출시하기 전에 이러한 테스트를 미리 많이 해봐야 소프트웨어의 보안이나 안정성이 보장되어 공격을 당해 악용되거나 제대로 작동을 못 하는 것을 방지할 수 있다. 이러한 소프트웨어 테스트를 진행할 때, 개발자들이 입력값을 직접 만들어서 테스트를 진행하게 된다면 속도도 느릴뿐더러 미처 생각해내지 못한 부분에 대한 테스트는 진행하지 못할 수도 있다. 따라서, 테스트 케이스 생성을 자동화시키는 방법에 대한 연구를 연구[1]에서 진행하였고, 그 과정에서 Focal Method라는 방식이 적용되었다. Focal Method란 그림 1에서 보이는 것처럼 전체 소스 코드들을 클래스(class), 메소드(method), 생성자(constructor) 등 여러 context로 적당히 쪼개는 방법을 의미한다.

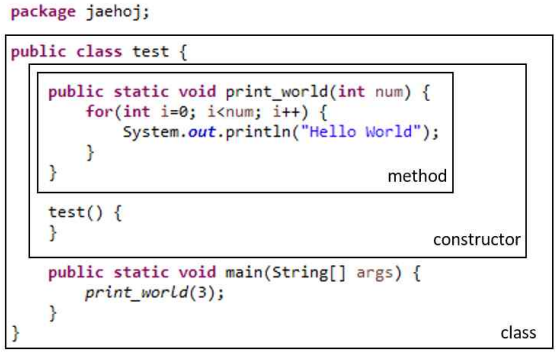


그림 1. Focal Method를 통해 자바 소스 코드를 구분한 예시

연구[1]에서는 오픈소스(open-source) 자바(Java) 소스 코드들을 긁어모아 코드 자체를 영어 문장들로 보거나 focal method를 이용하여 적당히 쪼개어 구분해 놓았다. 이후 이를 자연어 처리 기계학습 모델의 일종인 sequence-to-sequence transformer에 적용하였고, 이렇게 학습시켜 자동

으로 생성된 소스 코드들을 살펴본 결과, 기존 개발자들이 작성한 테스트 케이스들과 매우 유사하면서도 이해하기 쉽고 정확한 데이터 세트들을 작성할 수 있었고, 이를 온라인으로 공개하였다.

- CI (Continuous Integration)

CI란 동시에 여러 개발자들이 소스 코드를 수정하다 보면 발생할 수 있는 프로그램 버전 혹은 호환성 등과 같은 문제를 해결하기 위한 연구 분야이다. 전체 커다란 프로그램이 있을 때 일부 작은 부분의 코드만 수정이 되었다고 전체 프로그램을 다시 돌려보기에는 자원 낭비가 심할 수 있고, 코드가 일부 수정되어 샘플 입력을 돌려보는 과정에서 다른 한쪽의 코드가 수정된 것은 반영이 안 되어 있을 수도 있는 등, 여러 개발자들이 참여하는 프로젝트에서는 위와 같은 CI 이슈가 발생할 수 있다.

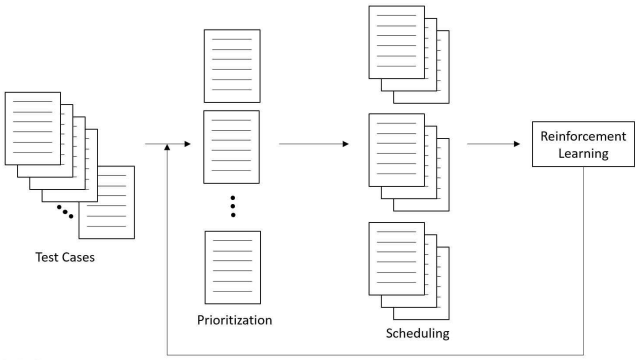


그림2. 테스트 케이스 우선 순위화 과정과 스케쥴링 과정을 학습하는 과정

연구[2]에서는 이러한 CI 과정에서 테스트 케이스를 자동으로 선택하고 우선 순위화하는 데 있어서 최초로 강화학습을 적용시켰고, 현재까지 나와있는 다른 테스트 케이스 자동생성 프로그램들과 비교했을 때 더 적은 cycle 수로 모든 테스트 케이스 검사를 진행할 수 있도록 우선 순위화하여 스케쥴링(scheduling) 방식으로 동시에 순차적으로 프로그램을 검증시키는 데에 성공하였다(그림 2). 이 과정에서 학습에 필요한 문제를 state, action, agent, policy, reward function 등의 notation으로 구분하여 실제 현업에서 사용되는 데이터도 강화학습에 적용될 수 있다는 점을 보일 수 있었다.

- 기호 실행(Symbolic Execution)

기호 실행이란 어떤 프로그램의 입력값으로 구체적인 값(concrete value)이 아닌 정해지지 않은 기호 값(symbolic value)이 들어가는 것을 의미한다. 변수에 들어갈 수 있는 모든 값이 입력으로 가능하기에 특정 분기 조건(branch condition)에서 어떤 방향으로 진행하는지 볼 수 있으며, 프로그램의 일정 부분으로 가기 위해서는 어떤 입력값을 넣어야 하는지 알 수 있다. 프로그램 진행 도중 분기 조건을 마주쳤을 때 포크(fork)를 통해 양쪽 경로를 모두 다 탐험할 수 있어 프로그램 내 모든 경로를 훑을 수 있기에 소프트웨어 테스트나 소프트웨어 보안 분야에 많이 사용된다.

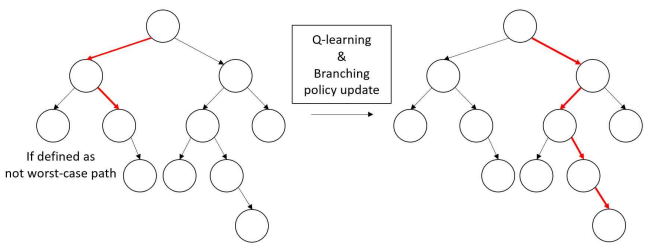


그림 3. 기호 실행 도중 다음 분기로 넘어가는 과정

연구[3]에서는 기호 실행을 통해 최악의 경로(worst-case path)값을 찾는 데 활용하였다. 그림 3에서와같이 특정 분기 정책(branching policy)을 통해 경로를 찾다가 Q-learning을 통해 최악의 경로값이 아니라고 판단이 되는 패턴이 등장한다면 해당 경로 탐험을 중단하고 이전 분기의 다른 경로로 탐험을 진행하게 된다. 이 과정에서 해당 분기 정책을 업데이트하고 이를 다시 학습하면서 기호 실행을 지속하게 된다. 일반적으로 분기 조건마다 탐험해야 하는 경로는 최소 2배씩 늘어나므로 복잡한 프로그램을 기호 실행하기 위해서는 경로가 기하급수적으로 증가하는 경로 폭발(path explosion) 문제가 발생하기에 작은 규모의 프로그램에서만 기호 실행을 진행할 수 있다는 단점이 존재한다. 그러나 연구[3]과 같은 경우에는 경로 폭발 문제가 발생하기 전에 특정 분기 아래의 모든 경로를 제거할 수 있으므로 더 큰 규모의 프로그램에 대해서도 최악의 경로를 제대로 탐색할 수 있었다.

Ⅲ. 결 론

본 논문에서는 테스트 케이스 생성 기법에 관한 최신 연구들을 살펴보면서 이들이 각기 다른 컴퓨터 공학의 분야에 어떤 식으로 적용될 수 있었는지 알아보았다. 주로 소프트웨어 테스트 분야에 사용된다는 점을 알 수 있었으며, 이들 연구에서의 주된 기여(contribution)는 테스트 케이스를 자동으로 만드는 데 성공하였고 비교 대상인 다른 소프트웨어들보다 이해하기 쉽고 성능이 좋은 테스트 케이스를 만들었으며 방대한 양의 고품질 데이터 세트 및 코드들을 만든 후 온라인으로 공개한 점들이었다. 비록 이번 논문에서는 테스트 케이스 생성, 감축, 우선 순위화에 관한 연구들을 모두 다 다루지는 못하였지만, 소프트웨어 엔지니어링 분야에서 관련 연구가 매우 많이 진행되고 있기에 이 분야들에 조금 더 관심 있게 본다면 기계학습 연구에 있어서 큰 도움이 될 수 있을 것이다.

ACKNOWLEDGMENT

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2021-0-00400, 저사양 디바이스 대상 고효율 PQC 안전성 및 성능 검증 기술 개발)

참 고 문 헌

[1] Tufano, Michele, et al. "Unit Test Case Generation with Transformers." *arXiv preprint arXiv:2009.05617* (2020).

[2] Spieker, Helge, et al. "Reinforcement learning for automatic test case prioritization and selection in continuous integration." *Proceedings of the 26th ACM SIGSOFT International Symposium on Software Testing and Analysis*. 2017.

[3] Koo, Jinkyu, et al. "Pyse: Automatic worst-case test generation by reinforcement learning." *2019 12th IEEE Conference on Software Testing, Validation and Verification (ICST)*. IEEE, 2019.

심전도에서 심장 박동의 특징을 추출하는 방법에 관한 연구

윤수진, 문소연, 권나영, 김동환
동국대학교 정보통신공학과
jjochocod@gmail.com

Methods of Extracting Features from ECG heartbeats

Yoon Soo Jin, Moon So Yeon, Kwon Na Young, Kim Donald
Dongguk Univ. Information and Communication Engineering

요 약

본 논문은 부정맥을 실시간으로 감지하는 IoT 디바이스를 개발하기 위하여, 심전도를 해석하는데 주로 사용되는 특징을 추출하는 방법에 대해 논술하였다. 그 유용성을 증명하기 위하여 MIT-BIH 부정맥 데이터베이스 Record 106 에서 추출된 특징들을 로지스틱 회귀에 적용하여 정상적인 맥박과 심방조기수축 맥박을 분류한 결과 95%의 정확도를 얻었다.

I. 서론

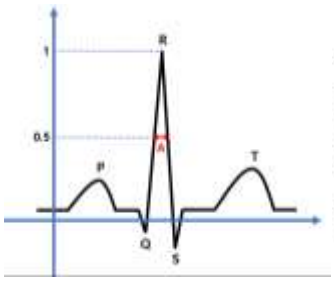
이 논문은 부정맥을 실시간으로 감지하는 IoT 디바이스를 개발하기 위한 사전 단계이다. 정상인 맥박과 비정상적인 맥박을 분류하기 위한 심장 박동의 특징은 매우 다양하다[1]. 이 논문에서는 심전도를 해석하는데 주로 사용되는 몇 가지 특징을 추출하는 방법에 대해 논술하였다. 그 유용성을 증명하기 위하여 MIT-BIH 부정맥 데이터베이스 Record 106 에서 추출된 특징들을 로지스틱 회귀에 적용하여 정상적인 맥박과 PAC(Premature Atrial Contraction) 맥박을 분류하였다. 그 결과 95%의 정확도를 얻었다.

II. MIT 데이터베이스

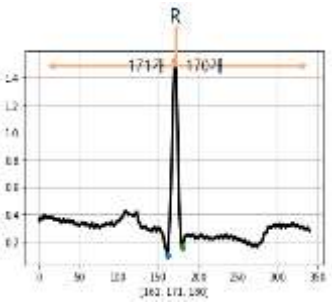
MIT-BIH 부정맥 데이터베이스는 수집한 4,000 개의 24 시간 보행 심전도 데이터 중 48 개를 PhysioNet 에 무료 공개하였다[2]. 이 논문은 정상적인 정상 신호인 Record 100 과 부정맥 신호인 Record 106 을 이용해 심장박동과 비정상적인 심장박동을 구별하였다.

Record 100 은 정상적인 신호로 총 2239 번의 정상 박동과 33 번의 조기 심방 수축(PAC), 1 번의 조기 심실 수축(PVC, Premature Ventricular Contraction)으로 구성되어 있다. Record 106 은 총 1507 번의 정상 박동, 520 번의 PVC 로 구성되어 있다. 최대 3 번 연속하여 심실 빈맥(VT, Ventricular Tachycardia)이 나타나기도 한다[3].

III. 심전도 신호의 특징



[그림 1]심전도 신호와 feature A



[그림 2]Heartbeat Window

1. 정상 심전도

심전도(ECG) 신호는 크게 P 파, QRS 군, T 파로 나뉜다. 정상 ECG 는 다음과 같은 조건을 만족한다[5]. P 파의 박동수가 60-100bpm 사이에 있어야 하며 P 파와 QRS 파가 1:1 로 대응해야 한다. PR interval 이 120-200ms 이어야 하며, QRS duration 이 100ms 이하가 되어야 한다. 또한 ST segment 가 0.5mm 보다 크고 T 파가 정상적이어야 하며, QT interval 이 남자의 경우 420ms 이하, 여자의 경우 430ms 이하여야 한다 [그림 1].

2. 부정맥 심전도

각 심장 질환을 검출하기 위해 확인해야 하는 ECG 의 특징은 다음과 같다.

[표 1] 심장 질환별 확인해야하는 ECG 특징

	Q wave	R wave	S wave	RR	P wave
PAC	✓	✓	✓	✓	
PVC	✓	✓	✓	✓	
Bradycardia				✓	
Tachycardia				✓	
AF	✓	✓	✓		✓
AFL	✓	✓	✓		✓

IV. 심전도 신호 특징 검출을 위한 알고리즘

1. Heartbeat Window

하나의 Heartbeat Window 는 1 개의 심장박동을 포함하고 있다. 그 크기는 R 의 인덱스를 중심으로 좌로 171 개, 우로 170 개의 샘플링 인덱스를 포함한다. 따라서, window 의 크기는 341 샘플이다. 이때 샘플의 주기는 2.8ms 이다 [그림 2].

2. QRS 인덱스

R 지점은 각각의 window 에서 전압이 최대가 되는 지점이다. find_peaks 이라는 Scipy 모듈을 이용하여 R 의 인덱스를 구하였다.

R 인덱스 직전의 최소 지점은 Q 인덱스이며 직후의 최소지점은 S 인덱스이다. Q 와 S 의 인덱스를 각각 R

인덱스로 초기화한 후, 그 지점으로부터 Q 의 경우에는 이전 값이, S 의 경우에는 이후의 값이 더 작은지를 판별해 최대한 작아지는 지점까지 반복문을 통해 구하였다.

이를 통해 Q, R, S 각각의 크기와 주기를 특징 set 으로 추출하였다.

3. RR

i 번째 Heartbeat Window 를 W[i]라고 하고, T_R[i]를 그 윈도우 안의 R 의 샘플링 인덱스라고 했을 때 RR0[i]부터 RR3[i]는 다음과 같이 계산된다.

RR0[i] = T_R[i-1] - T_R[i-2]
RR1[i] = T_R[i] - T_R[i-1]
RR2[i] = T_R[i+1] - T_R[i]
RR3[i] = average(RR[0][i], RR[1][i], RR[2][i])

T_R[-2], T_R[-1]은 그 영역이 존재하지 않으므로 RR0, RR1 를 구할 수 없다. 마찬가지로 주어진 데이터에서 R 의 개수가 n 개라고 하였을 때 T_R[n+1]이 존재하지 않으므로 데이터의 마지막 RR2 를 구할 수 없다. 이처럼 구할 수 없는 RR 의 데이터는 Null 처리하였다.

4. Ratio

W[i]에서 ratio0 은 RR1 에 대한 RR0 의 비율이다. ratio1 은 RR1 에 대한 RR2 의 비율이다. ratio2 는 RR1 에 대한 RR3 의 비율이다.

$$ratio0 = \frac{RR0}{RR1}, ratio1 = \frac{RR2}{RR1}, ratio2 = \frac{RR3}{RR1}$$

5. Cross Correlation

특정 심장 박동의 이상 여부를 파악하기 위해서는 해당 박동의 전후 박동과의 비교 또는 정상 박동과의 비교가 중요하다. 본 연구에서는 파이썬을 이용하여 피어슨 상관 계수(Pearson Correlation Coefficient)로 비교했다.

Avg = $(1/n)\sum_k$ (W[k]) where W[k] is a normal heartbeat
Corr1 = numpy.correlation(W[i], W[i+1])
Corr2 = numpy.correlation(W[i-1], W[i])
Cxy = numpy.correlation(W[i], Avg)

6. A

R 점의 크기(amplitude)값이 1 이 되도록 정규화 된 각각의 window 에서 Q 점과 R 점 사이의 크기가 0.5 가 되는 지점의 시간축 값과 R 점과 S 점 사이의 크기가 0.5 가 되는 지점의 시간축 값의 차이를 구해 A 배열에 넣는다 [그림 1].

V. 유용성 실험 결과

Section 4 에서 도출한 특징들의 유용성을 측정하기 위하여 Record 100 과 Record 106 에 로지스틱 회귀를 적용한 실험 결과값을 소개한다.

1. 모델 생성 및 평가

본 논문에서 구한 Record 100 에서 뽑은 16 가지의 심장 박동의 특징들을 input features X 로 사용하였다. X 는 Scikit Learn 툴의 fit_transform 함수를 이용하여 정규화 해주었다. Section 4 에서 구한 데이터 (X, Y)의 70%는 훈련 데이터셋으로, 30%는 테스트 데이터셋으로 사용하였다. Record 100 을 통해 학습하고 Record 106 의 부정맥 여부를 판단하여 도출한 결과값을 소개한다.

분류 모델이 어떠한 유형의 예측 오류를 일으키는지 알아보기 위하여, Scikit Learn 의 confusion_matrix 함수를 사용하여 오차 행렬을 구하였다.

$$\text{Precision} = \frac{TP}{TP+FP}, \text{Recall} = \frac{TP}{TP+FN}$$
$$\text{F1-Score} = 2 * \frac{Precision * Recall}{Precision + Recall}$$

정확도(Precision)는 모델이 True 라고 예측한 것 중에서 실제로 True 인 것의 비율이고, 재현율(Recall)은 실제 True 인 것 중에서 모델이 True 라고 예측한 것의 비율이다. F1 스코어는 정확도와 재현율의 조화 평균으로, 데이터 label 이 불균형 구조일 때 모델의 성능을 정확하게 평가할 수 있어 모델의 성능 평가에 사용된다.

[표 2]는 Record 100 과 Record 106 으로 각각 모델을 만들어, 교차하여 예측한 모델의 성능에 관한 표이다.

[표 2] 16 개 특징으로 훈련된 로지스틱 모델의 성능

생성 레코드	Record 100	Record 106
정확도	0.7409	0.9674
정밀도	0.7569	0.9941
재현율	0.9595	0.9727
F1 스코어	0.8463	0.9833

VI. 결론

본 논문은 부정맥을 실시간으로 감지하는 IoT 디바이스를 개발하기 위한 사전 단계이다. 논문은 해당되는 특징들을 추출하는 방법에 대해 논술하였다.

본 논문을 기초로 다음과 같은 후속연구를 진행할 예정이다: 1) 비정상적인 리듬의 다른 여러 특징들을 추출하는 연구가 필요하다, 2) 여러 ML 알고리즘을 적용하여 다양한 비정상적인 리듬을 예측한다, 3) 실시간으로 비정상적인 리듬을 탐지하는 IoT 디바이스를 개발하고, Azure 클라우드와 연동하여 환자의 생명을 구하는 시스템을 개발한다.

이 연구를 통해 심질환을 앓고 있는 환자들을 실시간으로 모니터링하여 생명을 구할 수 있는 시스템이 개발되기를 기원한다.

ACKNOWLEDGMENT

이 논문은 2020 년도 동국대학교의 개별연구 재원으로 지원을 받아 수행된 연구임.

참 고 문 헌

[1] Alarsan, F.I., Younes, M. Analysis and classification of heart diseases using heartbeat features and machine learning algorithms. J Big Data6,81(2019)

[2] Moody GB, Mark RG. The impact of the MIT-BIH Arrhythmia Database. IEEE Eng in Med and Biol 20(3):45-50 (May-June 2001).

[3] Moody GB. MIT-BIH Arrhythmia Database Directory. https://physionet.org/physiobank/database/html/mitdbdir/mitdbdir.htm. Accessed 2021

[4] 김성환. 신경계 중환자실에서 흔한 부정맥. J Neurocrit Care 2018;11[1]:7-12(2018).

[5] 이명용. (2013). EKG I. 대한내과학회 춘계학술발표논문집, 2013(0), 6-7.

Gradient boosting Classifier와 Feature importance를 이용한 EEG 신호처리에 따른 Hyponogram예측 모델에 관한 연구

류진수, 김규형, 조위덕*
아주대학교, 아주대학교, *아주대학교
s8620494@ajou.ac.kr, kviw0220@gmail.com, *chowd@ajou.ac.kr

A Study on the Hyponogram Prediction Model according to EEG Signal Processing Using Gradient Boosting Classifier and Feature importance

Ryu Jin Su, Kim Kyu Hyeong, Cho We Duke*
Ajou Univ., Ajou, Univ, *Ajou Univ.

요 약

본 논문은 기존의 뇌파, 호흡 수, 심박 수 등의 여러 데이터를 사용하여 수면의 깊이를 예측하는 모델과 다르게 2-channel의 EEG 데이터를 통해 수면의 깊이를 의미하는 Hyponogram을 예측하는 모델에 관한 연구로 뇌파의 주파수 대역에 따라 나눠 총 8개의 feature를 생성하여 머신러닝을 진행하였으며, 여러 방법으로 신호처리를 진행하였다. 신호처리를 진행한 데이터를 Gradient boosting Classifier를 사용하여 생성된 모델의 각 Feature importance를 확인하여 모델별 예측 결과에 대해 비교, 분석을 진행하여 신호처리 방법에 따른 예측률 차이를 확인하였다.

Ⅰ. 서 론

본 논문에서는 수면 시 2-channel의 EEG 뇌파 데이터[1][2]의 여러 신호처리 방법을 이용하여, 수면 상태를 나타내는 Hyponogram의 Wake(W), 렘수면 상태(R), 비 렘수면 상태(NR1, NR2, NR3, NR4)를 머신러닝을 이용해 예측 모델을 만드는 연구에 관한 내용이다. 머신러닝 알고리즘은 Gradient Boosting Classifier 알고리즘을 이용하였다. 각 수면 상태별로 신호처리 방법에 따른 예측률 차이를 확인하기 위해 신호처리 방법을 다르게 하여 모델을 생성하고, 예측률에 따른 이유를 feature importance를 비교하여 분석하는 연구를 진행하였다.

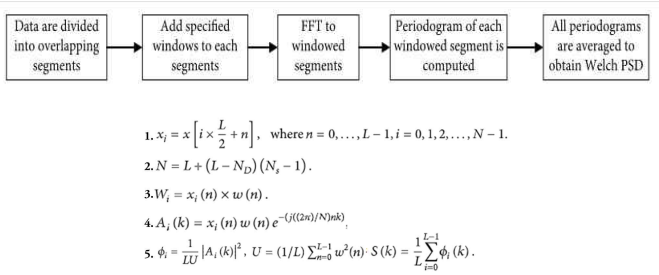
Ⅱ. 본론

본 논문에서는 뇌파(EEG)의 신호처리를 진행하였다. EEG 신호처리의 sequence는 다음과 같다. EEG signal data를 FFT로 변환한 데이터, welch-PSD으로 변환한 데이터 셋으로 나눈다. welch-PSD[3]으로 변환 시 windows 함수를 hann, hamming, blackman 세 가지 함수를 사용하여 데이터 셋을 나눈다. 각 windows 함수의 수식은 아래와 같다.

$0.5-0.5\cdot cos\left(\frac{2\cdot\pi\cdot k}{N-1}\right)$	$0.54-0.46\cdot cos\left(\frac{2\cdot\pi\cdot k}{N-1}\right)$	$0.42-0.5\cdot cos\left(\frac{2.0\cdot\pi}{N-1}\cdot k\right)+0.08\cdot cos\left(\frac{4.0\cdot\pi}{N-1}\cdot k\right)$
1. Hann Window	2. Hamming Window	3. Blackman Window

Table 1. 사용한 window 함수 수식

EEG signal의 데이터를 python을 이용하여 FFT, Welch PSD 두 가지 방법으로 주파수 영역으로 변환한다. 그 후 EEG의 각 wave 별로 주파수 영역에서 나누어 평균을 구한다. Welch PSD 같은 경우 변환 시 Window 함수를 사용하는데, 이 Window 함수에 따라 그래프의 특성이 변한다. Welch 함수를 구하는 sequence와 알고리즘은 아래와 같다.[4]



1. $x_i = x\left[i\times\frac{L}{2}+n\right]$, where $n=0,\ldots,L-1, i=0,1,2,\ldots,N-1$.
2. $N = L+(L-N_D)(N_s-1)$.
3. $W_i = x_i(n)\times w(n)$.
4. $A_i(k) = x_i(n)\,w(n)\,e^{-(j((2\pi)/N)nk)}$.
5. $\phi_i = \frac{1}{LU}|A_i(k)|^2$, $U = (1/L)\sum_{n=0}^{L-1} w^2(n)$, $S(k) = \frac{1}{L}\sum_{i=0}^{L-1}\phi_i(k)$.

Figure 1. Welch sequence와 알고리즘

위 알고리즘에 대한 파라미터는 다음과 같다. n은 EEG signal의 개수, 는 EEG signal을 segment로 나눈 것이다. 이 segment의 길이는 L이며 각 segment는 겹치는 부분이 존재한다. 는 겹치는 점의 수 는 segment의 수이다.

w[n]은 segment에 대한 window 함수이다. 사용하는 window 함수의 종류에 따라 수식을 변경할 수 있다.

는 각 segment의 Periodogram, U는 window 함수의 평균전력, S(k)는 PSD이다.

위 알고리즘을 따라 welch PSD를 구한 후 EEG의 wave 별 구간으로 나누어야 한다. EEG wave의 구간은 Delta wave : 1~4Hz, Theta wave : 4~7Hz, Alpha wave : 8~12Hz, Beta wave : 12~30Hz이다. welch PSD를 구한 후 각 wave의 구간별로 나누어 평균을 구한 값을 machine Learning의 parameter로 사용하였다. 한 channel당 4개씩 총 8개의 feature를 통해 machine Learning을 진행하였다.

본 연구에서 사용한 machine Learning 알고리즘은 Gradient boosting classifier로 여러 개의 학습기를 연결하여 강한 학습기를 만드는 앙상블 방법인 부스팅 방법 중 하나로 이전 예측기가 만든 잔여 오차에 새로운

예측기를 학습시키는데 Negative gradient(residual)를 최소화시키면서 학습시키는 알고리즘이다.

$$F_0(\mathbf{x}) = \arg\min_\rho \sum_{i=1}^{N} L(y_i, \rho)$$
For m = 1 to M do:
$$\tilde{y}_i = -\left[\frac{\partial L(y_i, F(\mathbf{x}_i))}{\partial F(\mathbf{x}_i)}\right]_{F(\mathbf{x})=F_{m-1}(\mathbf{x})}, \quad i = 1, N$$
$$\mathbf{a}_m = \arg\min_{\mathbf{a},\beta} \sum_{i=1}^{N} [\tilde{y}_i - \beta h(\mathbf{x}_i; \mathbf{a})]^2$$
$$\rho_m = \arg\min_\rho \sum_{i=1}^{N} L(y_i, F_{m-1}(\mathbf{x}_i) + \rho h(\mathbf{x}_i; \mathbf{a}_m))$$
$$F_m(\mathbf{x}) = F_{m-1}(\mathbf{x}) + \rho_m h(\mathbf{x}; \mathbf{a}_m)$$
endFor
end Algorithm

Figure 2. Gradient bossting classifer algorithm

Hyponogram 상태별 Gradient boosting classifier를 통한 Machine Learning 결과 Test score는 아래와 같다.

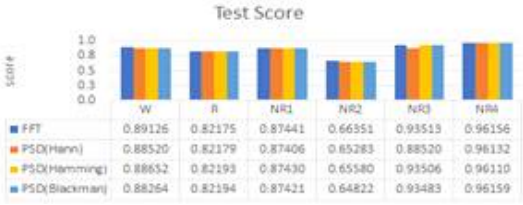


Figure 3. Hyponogram 상태별 Test score

신호처리별로 같은 Hyponogram 일 때의 test score는 크게 차이가 나지 않았으며 Non-rem 수면상태 중 NR2 상태 모델의 test score는 0.65 정도로 다른 상태에 비해 낮은 예측률이 나왔으며 NR4일 때 0.96 정도로 높은 예측률을 보여주었다. 예측률이 낮게 나오는 이유를 확인하기 위해 FFT로 신호처리한 경우 각 Hyponogram 상태별 모델의 Feature importance[6]를 확인하였으며 Feature importance는 아래의 과정을 통해 구해.

- Inputs: fitted predictive model m, tabular dataset (training or validation) D
- Compute the reference score s of the model m on data D (for instance the accuracy for a classifier or the R^2 for a regressor).
- For each feature j (column of D):
  - For each repetition k in $1, \ldots, K$:
    - Randomly shuffle column j of dataset D to generate a corrupted version of the data named $\tilde{D}_{k,j}$
    - Compute the score $s_{k,j}$ of model m on corrupted data $\tilde{D}_{k,j}$
  - Compute importance i_j for feature f_j defined as:

$$i_j = s - \frac{1}{K}\sum_{k=1}^{K} s_{k,j}$$

Figure 4. Scikit Learn Gradient boosting의 Classifier Feature importance



Figure 5, FFT처리 데이터 model의 Feature importance

상대적으로 높은 예측률이 나타나는 Non-rem 수면 상태와 다르게 NR2의 Feature importance는 FFT - r_delta가 FFT -f_delta에 비해 높게 나타나지만 다른 상태에서는 반대로 FFT -f_delta의 feature importance가 높게 나타난다.

하지만 이것이 model의 score가 낮게 나오는 직접적인 원인이 아닐 수 있기 때문에 Non-rem 수면 상태 중 신호 처리 방법에 따른 모델별 예측값의 차이가 다소 발생한 NR3의 feature importance를 비교해 본 결과 아래와 같다.

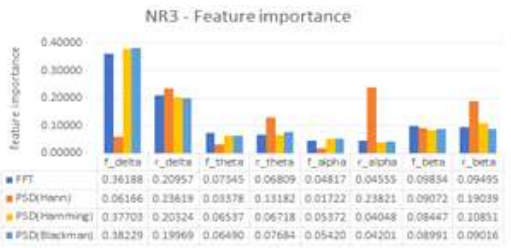


Figure 6. NR3상태 신호처리 방법에 따른 Feature importance

NR3에서 PSD(Hann)의 score가 다른 model 들에 비해 낮게 score가 나왔는데 feature 별 importance를 보면 PSD(Hann)의 f_delta에 대한 importance가 낮고 활동 중에 나타나는 머리 뒤쪽(r) alpha 파와 beta 파의 importance가 높게 설정되어 결과적으로 다른 model 들에 비해 score가 낮게 나타났다. 이를 통해 Non-rem 수면 예측 모델에서 f_delta의 importance가 높아야 예측률이 높게 나올 것이라는 것을 알 수 있다.

Ⅲ. 결론

본 논문에서는 2-channel의 수면 시 EEG 데이터를 통해 Gradient boosting classifier로 machine learning 학습을 진행하고 신호처리 방법에 따른 score의 차이와 score가 낮은 경우 낮게 나온 원인을 모델별 feature importance를 통해 확인하였다. 대부분의 뇌파 연구 데이터는 많은 channel의 데이터를 사용하기 때문에 일상적으로 사용 가능한 제품에 적용하는 데에는 한계가 있다. 하지만 신호처리를 통해 한 개의 channel의 데이터를 여러 개의 feature로 나누면 적은 수의 channel의 데이터로도 활용이 가능하다는것을 확인하였다. 또한 신호처리 방법이 머신러닝 결과 차이를 통해 뇌파를 이용한 머신러닝 진행시 머신러닝 알고리즘 뿐만아니라 뇌파 신호를 어떻게 처리하는지도 중요한 요소임을 확인할 수 있다.

ACKNOWLEDGMENT

이 논문은 한국연구재단의 이공분야기초연구 (기본연구)의 지원을 받아 수행된 연구임(NRF-2019R1F1A1063128).

참 고 문 헌

[1] B Kemp, AH Zwinderman, B Tuk, HAC Kamphuisen, JJL Oberyé. Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG. IEEE-BME 47(9):1185-1194 (2000).

[2] Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., ... & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation [Online]. 101 (23), pp. e215 - e220..

[3] Anis Ameera, A. Saidatul, Z Ibrahim. Analysis of EEG Spectrum Bands Using Power Spectral Density for Pleasure and Displeasure State. Anis Ameera et al 2019 IOP Conf. Ser.: Mater. Sci. Eng. 557 012030

[4] Qi Xiong, Xinman Zhang, Wen-Feng Wang, Yuhong G. A Parallel Algorithm Framework for Feature Extraction of EEG Signals on MPI. Computational and Mathmatical Methods in Medicine, vol.2020, 10 pages, 2020.

[5] Friedman, J. H. (February 1999). "Greedy Function Approximation: A Gradient Boosting Machine".

[6] scikit-learn.org, https://scikit-learn.org/stable/modules/permutation_importance.html

알츠하이머형 치매 진행단계 구분을 위한 딥러닝 성능 비교 연구

이사랑, 송은채, 강재환, 김준석*
동의대학교 산업ICT공학과, 인공지능그랜드ICT 연구센터
dltk812@gmail.com, aue020@naver.com, [jh.kang, *junsuk.kim]@deu.ac.kr

Classification of Alzheimer's Disease Progression using Deep Learning Algorithms

Sa Rang Lee, Eun Chae Song , Jae-Hwan Kang, Junsuk Kim*
Dong-Eui Univ.

요 약

이 논문에서는 자기공명영상 (MRI) 을 이용하여 알츠하이머형 치매 진행의 5단계를 예측하는 딥러닝 모델들을 직접 설계하고 학습시켜 모델의 성능을 비교하는 실험을 하였다. 다양한 딥러닝 알고리즘 가운데 컨볼루션 신경망 (convolutional neural network)을 사용하고 추가로 전이학습 (transfer learning)을 활용하였으며 각 모델의 정확도와 오차율을 비교하였다. 결과적으로 전이학습을 사용한 딥러닝 모델이 단순 컨볼루션 신경망 보다 치매 진행단계 분류에서 뛰어난 성능을 보인다는 것을 확인하였다. 이 연구 결과는 딥러닝 기술을 사용한 MR 영상 기반 치매단계 구분 연구에 기초 자료가 될 것으로 생각된다.

Ⅰ. 서론

알츠하이머는 치매를 일으키는 가장 흔한 퇴행성 질환으로 서서히 발병하며 점진적으로 진행되는 경과가 특징적이다 [1]. 치매의 진행 단계를 예측하는 다양한 방법들 중 자기공명영상이미지 (MRI)를 활용한 진단은 뇌 피질의 수축 및 뇌 심실의 확장을 시각적으로 확인할 수 있다는 장점 때문에 의학적으로 가장 많이 사용되는 방법이다. 하지만 MR 이미지의 시각적인 정보만을 사용하여 치매의 진행 단계를 예측하기 위해서는 전문적인 임상 지식과 축적된 경험이 필수적이다. 최근 인간의 시각으로 추출해내기 힘든 바이오마커 (bio-marker) 추출, 의료진 부족으로 인한 물리적인 시간 부족 문제를 해결하기 위하여 딥러닝 알고리즘을 활용하는 다양한 연구가 진행중이며, 특히 컨볼루션 신경망 (CNN) 방법은 매우 우수한 성능을 보이는 것으로 보고되고 있다 [2,3].

이 논문에서는 알츠하이머형 치매 진행과 관련된 5단계 (Early Mild Cognitive Impairment: EMCI, Late Mild Cognitive Impairment: LMCI, Mild Cognitive Impairment: MCI, Alzheimer's Disease: AD, Cognitively Normal: CN)의 MR 이미지를 이용하며, 진행단계를 예측하기 위하여 CNN 기반 딥러닝 모델들을 직접 설계하고 학습시켜 모델의 정확도를 비교 및 분석하려고 한다.

Ⅱ. CNN (convolutional neural network)

알츠하이머형 치매 진행의 5단계를 분류하기 위해 이미지 분석에 널리 활용되고 있는 CNN을 사용하였다. CNN은 이미지 및 비디오 인식, 추천 시스템, 이미지 분류, 의료 이미지 분석 및 자연어 처리에 다양하게 응용된다 [4]. CNN는 수십 또는 수 백개의 계층을 가질 수 있으며, 각 계층은 영상의 서로 다른 특징을 검출한다. CNN의 계층 중 Convolutional Layer는 Convolutional 필터와 활성화 함수 그리고 Pooling layer를 반복적으로 조합하여 특징을 추출한다. Convolution layer는 합성곱 연산을 통해 적절한 특징맵 (Feature Map)을 구성한다. 상위 Convolution layer는 하위 Convolution layer의 출력을 이용하여 좀 더 높은 수준의 특징들을 추출하게 된다. 또한 각 Convolution layer는 필터를 이용하여 특징을 추출하게 되는데, 이 필터는 고정된 것이 아니라 학습에 의해 학습 데이터에 맞게 결정이 된다 [5]. 특징맵을 추출 후 특징맵에 활성화 함수를 적용하게 되는데 활성화 함수는 주로 비선형 함수로 되어 있으며 예를 들어 시그모이드 (Sigmoid), ReLu (Rectified Linear Unit) 등이 있다. 시그모이드 함수는 입력값이 아무리 커도 미분 값의 범위가 제한됨으로써 층이 많을수록 gradient 값이 0에 수렴하는 문제를 가지고 있다. 이로 인해 그래디언트 소실 문제를 야기한다. 이러한 문제를 해결 한 것이 ReLu 함수이다. ReLu 함수는 입력에 대한 출력을 특정 범위에 한정하지 않는 비선형 함수로 입력으로 양수 값이 들어오면 값을 그대로 출력하고 음수 값이 들어오면 0을 출력한다 [6]. 이 논문에서는 활성화 함수로 ReLu 함수를 사용하였다. Pooling layer는 입력 정보를 최댓값, 최솟값, 평균값 등으로 압축하여 데이터의 연산량을 줄여주는 절차다. 요약하면 CNN은 입력 데이터에 대해서 필터를 통해 이미지의 특징을 추출하고 (Convolution layer) , 특징을 강화하고 이미지의 크기를 축소(Pooling layer)한다. 이 과정을 통해 입력된 2차원 데이터의 적절한 특징을 추출한다. 이후 Fully connected layer에서 1차원으로 변환된 feature map을 이용하여 최종 분류한다.

Ⅲ. 실험 결과 및 분석

° 데이터 구성 및 전처리

실험에 사용되는 데이터는 kaggle 에서 알츠하이머의 진행단계에 따라 수집되어있는 오픈 데이터 셋을 사용하였다 (https://www.kaggle.com/madhucharan/alzheimersdisease5classdatasetadni) [7]. 분석에 사용된 MR 이미지의 피험자 정보는 모두 익명화 되어있으며, 뇌 이미지는 horizontal plane 이며 256x256 사이즈를 가지고 있다. 학습 이미지의 개수는 총 1,101개로 EMCI 204개, LMCI 61개, MCI 198개, AD 145개, CN 493개이며, 테스트 이미지의 개수는 총 195개로 EMCI 36개, LMCI 11개, MCI 35개, AD 26개, CN 87개이다. CNN 모델의 학습 데이터 개수를 증가시키기 위하여 Data augmentation을 적용하여 이미지를 임의로 변형 및 개수를 증가시켰다 [8]. 이 연구에서 사용한 Data augmentation 방식은 Horizontal flip, Width shift, Height shift이다. Horizontal flip은 50% 확률로 이미지를 수평으로 뒤집으며 뇌 영상과 같이 원본 이미지에 수평

비대칭성이 없을 때 효과적이다. Width shift, Height shift는 이미지를 수평 또는 수직으로 랜덤하게 평행 이동시키는 방법이다. 학습용의 배치 사이즈는 22, 테스트용의 배치 사이즈를 13으로 하였다.

° CNN 모델

Convolution layer에서는 층을 3개로 하였고 각 계층에서 필터의 개수는 32, 32, 64개로 진행하였다. 활성화 함수는 모두 ReLu 함수를 사용하였고, Pooling layer에서는 Max Pooling을 사용하였다. 마지막으로 Fully Connected Layer를 거쳐 softmax 함수를 통해 확률 값으로 변환하여 최종 결과를 출력한다 [6]. 본 논문에서는 모델을 실행할 epoch를 30으로 하였으며 학습 결과를 정확도로 평가하였다 (그림 1). 최종 정확도는 44.55% (baseline accuracy = 20%) 이고 오차율은 1.40 였다.

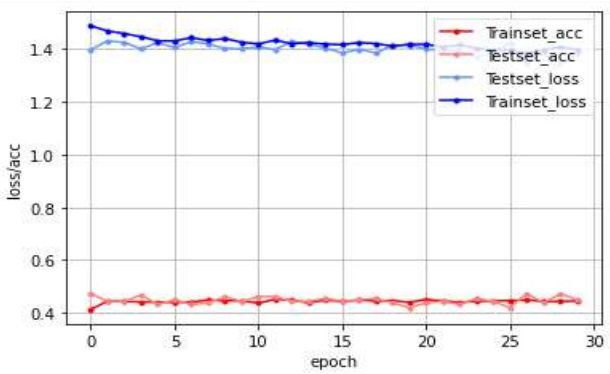


그림 1. CNN 모델 학습 결과 그래프

° 전이학습 (transfer learning)

CNN 모델에서 깊은 신경망을 훈련하기 위해서는 충분한 양의 데이터가 필요하기 때문에 많은 양의 데이터가 확보되지 않는 상황에서는 과적합 문제가 발생할 수 있다. 전이학습은 위와 같은 문제를 해결하기 위한 방법으로, 학습 데이터가 부족한 분야의 모델 구축을 위해 기존의 풍부한 데이터를 바탕으로 일반성을 학습한 모델의 일부 계층을 재사용함으로써 모델의 학습 속도를 높이고 분류 성능을 높인다고 알려져있다 [9].

기존 CNN 모델에 전이학습 방식을 적용해보았다. 학습 이미지의 배치 사이즈는 22, 테스트 이미지의 배치 사이즈는 13으로 지정하였고 이외의 데이터를 증가시키고 불러오는 작업은 기존 CNN 모델과 동일하게 진행하였다. 특히 VGG16의 학습된 가중치를 가져와서 적용하였다.

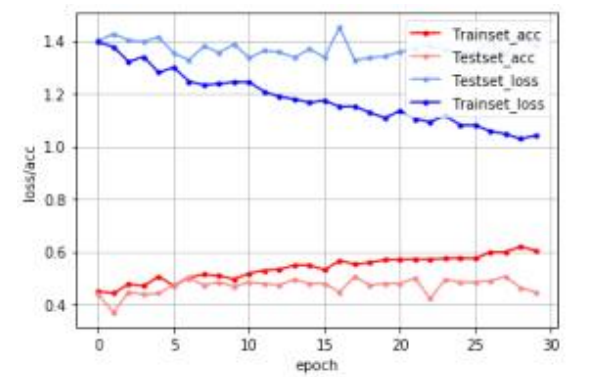


그림 2. 전이학습 모델 학습 결과 그래프

전이학습의 최종 정확도는 60.4% 오차율은 1.04로 나타났다. CNN의 최종 정확도인 44.55%와 비교하여 볼 때, 전이학습을 활용할 경우의 분류성능이 우수하다고 볼 수 있다. 또한 학습 속도 및 추세 측면에서도 CNN 모델에 전이학습을 활용하는 것이 분류 성능 향상에 큰 도움을 주고 있음을 알 수 있었다. 그림 2를 보면, 에포크의 횟수가 증가할수록 빠른 속도로 loss가 감소하고 accuracy가 증가하는 것을 확인할 수 있다.

Ⅳ. 결론

본 논문에서는 알츠하이머형 치매 진행단계 구분을 위하여 CNN과 전이학습 알고리즘을 구현하였다. 이를 통해, 치매 진행단계를 5단계로 구분하였으며 오차율과 정확도를 가지고 각각의 성능을 평가하였다. 결과적으로 단순 CNN보다 전이학습을 활용했을 때 정확도가 약 15%만큼 높게 나오며 오차율은 약 25% 가량 낮은 것을 확인하였다. 이 연구 결과를 통하여 의료 인공지능 분야에서 전이학습 활용의 중요성을 확인하였으며, 향후 유사 연구의 기초자료로 사용될 것으로 기대된다.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (2021R1F1A1055814) and the Grand Information Technology Research Center support program (IITP-2020-0-01791) grant funded by the Korea government (MSIT).

참 고 문 헌

[1] Brookmeyer, Ron, et al. "Forecasting the global burden of Alzheimer's disease." Alzheimer's & dementia 3.3 (2007): 186-191.

[2] Son, Joo Hyung, Kyeong Tae Kim, and Jae Young Choi. "Alzheimer's Disease Classification with Automated MRI Biomarker Detection Using Faster R-CNN for Alzheimer's Disease Diagnosis." Journal of Korea Multimedia Society 22.10 (2019): 1168-1177.

[3] Kim, Ji-Yul, and Soo-Young Ye. "Diagnostic Classification of Chest X-ray Pneumonia using Inception V3 Modeling." Journal of the Korean Society of Radiology 14.6 (2020): 773-780.

[4] Bae, Chang-Hui, et al. "An Experimental Comparison of CNN-based Deep Learning Algorithms for Recognition of Beauty-related Skin Disease." Journal of the Korea Society of Computer and Information 25.12 (2020): 25-34.

[5] Lee, JeongHwan, Byeong Man Kim, and Yoon Sik Shin. "Effects of preprocessing and feature extraction on CNN-based fire detection performance." Journal of the Korea Industrial Information Systems Research 23.4 (2018): 41-53.

[6] Choi, Gwang-Mi, and Yu-Jeong Jeong. "Efficient Iris Recognition using Deep-Learning Convolution Neural Network." The Journal of the Korea institute of electronic communication sciences 15.3 (2020): 521-526.

[7] https://www.kaggle.com/datasets

[8] Kim, Chisung, et al. "Development of Automatic Crack Detection Technology in Welded Surface using Laser Active Thermography and CNN Deep Learning." Journal of the Korean Society for Nondestructive Testing 40.3 (2020): 163-173.

[9] Kim, Yeonho, and Namgyu Kim. "Deep Learning-based Pes Planus Classification Model Using Transfer Learning." Journal of the Korea Society of Computer and Information 26.4 (2021): 21-28.

심리상담 내담자의 심리변화 시각화시스템

손우정, 손동현, 김주성, 강수환*

부산대학교, *라인플러스

handfriendship1998@gmail.com, europeanboy@naver.com, kjs_0322@naver.com, lionwithnoname@gmail.com

Visualization system of the flow of emotion change of client in psychotherapy

WooJung-Son, DongHyun-Son, JuSung-Kim, SuHwan-Kang*

Pusan National Univ., *LINE Plus

요 약

심리상담이란 일상에서 마주할 수 있는 다양한 문제 상황으로부터 야기되는 정서적, 정신적 고통을 상담자와 내담자가 함께 공유하며 해결해나가는 과정을 말한다. 심리상담을 행하는 과정에서 상담자가 내담자의 고민과 심리를 정확하게 파악하는 것이 만족스러운 결과를 내는 것에 중요하다. 본 논문에서는 심리상담시 상담자가 AI 기술을 활용하여 심리상담을 받는 내담자의 심리변화를 상담자가 쉽게 파악할 수 있는 방법을 제안한다.

Ⅰ. 서 론

심리상담이란 일상에서 마주할 수 있는 다양한 문제 상황으로부터 야기되는 정서적, 정신적 고통을 상담자와 내담자가 함께 공유하며 해결해나가는 과정을 말한다. 현재 상담업계에서는 심리치료, 상담 등의 분야에 전문 자격을 가진 상담사가 내담자의 말과 행동, 심리 등을 파악하고 직접 기록하는 형식으로 상담이 진행되는데, 자연어처리 기술의 발전으로 인공지능 기술을 심리상담 업계에 도입하기 위한 연구들이 활발히 진행되고 있다.

인공지능 기술을 상담에 활용하기 위해 행해진 대부분의 연구들은 인공지능 상담사를 도입하여 전문 상담사의 역할을 대신하는 것에 초점이 맞춰져 있다. 하지만 심리치료와 상담 분야의 인공지능 기술 적용이 아직 시작단계에 있어, 적용에 따른 효과를 철저하게 검증하기 보다는 기술 자체를 소개하고 이를 활용하는 방안에 주안을 두는 것이 현재까지 진행된 대부분의 연구들이다. 심리적 개입을 통한 정신건강 상의 변화를 검증하는 것은 통상 오랜 관찰을 요하는데, 그것을 장기적으로 추적하고 그 효과성을 검증한 연구 사례가 부족하다.[1] 이러한 이유로 인공지능 상담 서비스의 도입과 상용화는 아직 시기상조이며, 현재 대부분의 상담센터에서는 전문상담사가 직접 상담을 수행하고 있다. 따라서 본 논문은 인공지능 기술을 상담사의 보조도구로써 활용함으로써 실제 심리상담 업계에 보다 직접적인 활용과 도움이 가능하도록 한 방법을 소개한다.

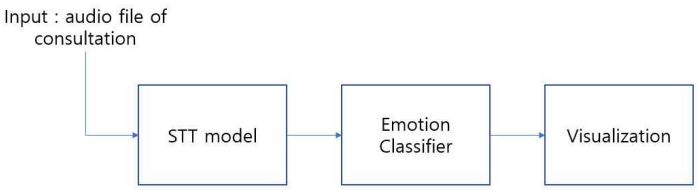
심리상담이 이루어지는 과정에서 상담사가 내담자의 말과 심리변화를 알아차리는 것은 올바른 결과를 이끌어내는 데 있어 중요한 요소이다. 하지만 상담사마다 다른 상담경력 및 인지능력, 해당 상담시의 컨디션 등은 같은 내담자의 동일 문제 상황을 마주하더라도 다른 결과를 도출할 수 있게 한다. 만약 내담자의 말을 듣고 그것의 감정 변화를 동일한 기준으로 관찰하며, 실제 상황에서 빠르게 적용할 수 있도록 시각화하는 시스템을 보조도구로써 사용한다면, 위에서 언급한 여러 외부적 변수로부터 강건한 상담을 행할 수 있다. 이는 내담자의 상담사 교체 요청, 상담사의 수퍼바이저(실습, 수련을 감독하는 경력 상담사)로의 도움 요청 및 인수인계 등 실제 빈번히 발생하는 다양한 상황에서 활용될 수 있다는 장점이 있다.

앞서 설명한 시스템을 도입하기 위해 본 연구에서는 인공지능 기술을 사용한 방법론을 제시한다. 우선 STT(speech to text) 기술을 이용해 심리상담 내용을 텍스트로 기록한다. 그 다음 감정분류모델을 이용하여 기록된 상담내용을 28개의 감정군으로 분류한다. 마지막으로 상담사가 쉽게 결과물의 의미를 파악할 수 있도록 시각화하여 타임라인 그래프로 표현한다. 이러한 프로세스를 통해 제한된 상담시간의 효용 가치를 극대화하고, 다양한 유형의 내담자에 대한 이해와 관찰의 '객관성'을 높이는 것에 기여하는 것이 본 시스템의 연구 목표이다.

본 논문의 2장에서는 제안된 시스템의 설계와 구조에 대해 소개한다. 3장에서는 시스템을 실험한 과정과 결과에 대해 설명한다. 마지막으로 결론과 의의, 한계와 필요한 후속 연구를 서술한다.

Ⅱ. 본론

본 논문에서 제안하는 방법론의 시스템 설계는 [그림1]과 같다. 본 장에서는 STT model, Emotion Classifier, Visualization 순으로 시스템에 대해 기술하였다.



[그림1] 시스템 설계

2.1 음성텍스트변환(STT model)

음성 파일을 감정분류모델이 처리하기 쉽도록 텍스트 파일 형태로 바꾼다. 음성 파일을 텍스트 파일로 바꾸는 과정에는 mozilla 재단의 오픈소스 음성인식 엔진인 DeepSpeech의 사전훈련(pre-trained)된 모델을 이용하였다. DeepSpeech는 Google의 Tensorflow를 기반으로 하는데, MFCC(Mel Frequency Cepstral Coefficient)를 이용하여 음성의 특징벡터를 추출하며, RNN을 핵심 엔진으로 5개의 은닉노드 층을 두어 학습하는 음성인식 모델이다. 변환오류율은 6.5%로, KALDI 등의 타 음성인식 모델에 비해 상대적으로 간단하면서 높은 성능을 보여 선택하였다.[2]

본 연구에서는 DeepSpeech에서 제공하는 사전훈련 모델 중 'pbmm','scorer'모델로 STT-model을 구성하였다. pbmm은 음성신호를 text로 전사(transcribe)하는 음향모델(acoustic model)이고, scorer은 pbmm과 별도로 전사된 text 후보들에 대해 점수(score)를 매기는 외부 언어모델(language model)이다.

사전훈련된 언어모델에 적용한 설정(configuration)은 다음과 같다.

beam_width는 1000으로 두었다. 이는 음성신호의 전사가 일어나는 언어모델의 디코딩(decoding)과정에서 수행되는 beam search의 tree width이다. 언어모델의 가중치 상수인 language_model_weight은 0.93으로 두었다. 언어모델이 산출하는 점수는 말뭉치의 어휘 수가 많아질수록 작아지는데, 이를 보정해주는 상수 값인 WIP(word insertion penalty)를 1.18로 설정했다.

위와 같이 구성한 STT의 수행과정은 다음과 같다. 먼저 모델은 음성파일을 입력으로 받아들이는데, 입력의 sampling rate는 학습에 사용된 데이터의 sampling rate와 동일하게 16KHz로 샘플링하였다. 전사과정에서 중요한 요소 중 하나는 문장을 구분하는 것인데, [3]에서 제시한 0.8초를 기준으로 어휘 사이의 공백기간이 0.8초를 넘어가면 문장이 바뀌는 것으로 인식하게 하였다. 타임라인 그래프를 출력하여 시각화하는 것이 본 시스템의 목적이므로, 출력된 각각의 문장마다 타임스탬프를 함께 기록하였다. 마지막으로 전사된 문장과 타임스탬프를 csv형식으로 저장하였다.

2.2 감정분류모델(Emotion Classifier)

2.1절의 과정을 거쳐 텍스트로 전사된 음성내용을 입력으로 받아 해당하는 세그먼트에 적절한 감정을 라벨링해주는 분류모델을 설계한다. 학습에 사용된 데이터셋, 분류모델의 설계 순으로 기술하였다.

2.2.1. 데이터셋

해외 소셜 뉴스 웹사이트 'Reddit'의 영어 댓글을 수집하여 그와 적절한 감정을 라벨링하는 형식으로 구축한 구글의 공개 데이터셋 'GoEmotions'을 사용한다.[4] 총 58,009개의 세그먼트가 샘플로 사용되며, 각각의 샘플은 27가지의 감정으로 라벨링되어있고, 샘플의 sequence length의 최대길이는 30이다. 전체 샘플 중 훈련 데이터셋으로는 43,410개, 테스트 데이터셋으로는 5,427개, 검증 데이터셋으로는 5,426개로 분배하여 사용하였다. 심리상담의 경우, 내담자의 발언에 항상 감정이 실려있지 않을 수 있으므로, 27개의 감정군에 감정적 중립상태인 'Neutral'을 추가하여 총 28가지의 감정군을 사용하였다.

2.2.2. 모델 설계

감정분류모델은 구글의 Transformer기반 사전훈련 모델인 BERT(Bidirectional Encoder Representations from Transformers)의 소형화 버전인 BERT-base를 미세조정(fine tuning)하여 설계하였다.[5] [4]에서 소개한 방법에 따라 선행학습된 모델에 밀집 출력층을 추가하였고, 심리상담의 경우 내담자의 발언에 여러 가지 감정이 내포되어있을 수 있으므로 다중레이블이 가능하도록 시그모이드 크로스 엔트로피 손실함수를 사용하였다. 또한 발언의 문맥을 파악하는 것에 중점을 두어 토큰단위가 아닌 세그먼트단위로 분류를 수행하였다. 이외의 모델 설정(configuration)을 위한 파라미터는 hidden_size 768, hidden_act gelu, vocab_size 30522, hidden_dropout_prob 0.1, num_attention_heads 12, max_position_embeddings 512, num_hidden_layers 12, attention_probs_dropout_prob 0.1 와 같이 구성되었다. 미세조정을 위한 하이퍼파라미터는 batch_size 4, learning_rate 0.00005, epochs 4 가 사용되었다.

2.3 시각화(Visualization)

시각화에는 구글에서 제공하는 Google Charts 중 timeline API를 이용하였다. 2.2절의 수행결과를 tsv형식으로 저장하였다. tsv파일에는 text로 전사된 문장, 발화 시작시각, 발화 종료시각, 감성 레이블, 감성 score(0~1 사이의 확률)이 기록되어 있다. 이 파일을 읽어들이는 자바스크립트 코드를 작성하여 x축은 시각, y축은 감성 레이블이 표기된 타임라인 그래프를 브라우저에 렌더링하였다. 결과 그래프는 Ⅲ장에서 확인할 수 있다.

Ⅲ. 실험 방식 및 시스템 실험 결과

본 시스템은 심리상담 중 내담자의 발언을 마이크 등으로 녹취하여 그것을 저장한 음성파일을 모델에 입력하면, 감정흐름의 시각화 결과를 컴퓨터의 브라우저에서 확인하는 상황에서 이용하는 것으로 한다. 심리상담에서는 주로 내담자의 사적인 고민 혹은 정신적, 정서적 문제가 드러날 수 있으므로 프라이버시 보호를 위해 내용을 발설하지 않는 것을 원칙으로 한다. 따라서 전문 상담사가 행하는 실제 상황 대신, 심리상담 중 과거경험을 이야기하는 상황을 가정하여 저자가 직접 해당 상황을 묘사하여 음성 파일에 녹취하였다.

제안한 시스템을 batch형식으로 수행하는 쉘스크립트 파일인 run.sh를 수행하면, 먼저 음성파일을 텍스트로 전사해주는 speech2text.py 파일이 실행되어 전사된 텍스트가 기록된 임시파일을 얻는다. speech2text.py의 설명은 Ⅱ장의 2.1절에 기술되어 있다. 이후 2.2절의 감성분류모델이 구현된 mybert_classifier.py파일이 내부적으로 수행되어 전사된 텍스트파일에 감성레이블이 표기된 tsv파일을 얻는다. 마지막으로 타임라인 그래프를 그려주는 자바스크립트 코드를 이용해 결과를 시각화한다. 음성파일과 시스템을 시연한 결과 동영상은 아래 링크에서 확인할 수 있다.
https://drive.google.com/drive/folders/1U3S5Q2zM26UPd-VvLpK6LQLfnfDurOTD?usp=sharing

Ⅲ. 결론

본 연구에서는 STT, 감정분류모델을 이용해 심리상담시 내담자의 감정흐름을 관찰하고 시각화하는 방법을 제시하고 실험하였다. 결과를 타임라인으로 보여줌으로써 누구나 쉽게 감정의 흐름을 이해할 수 있고, 상담사에 관계없이 같은 결과를 제공하므로 본 연구의 목표를 달성했다는 것에 의의가 있다.

한계점으로는 음성과 감정라벨이 주어진 테스트 데이터셋 확보의 어려움으로 시스템 성능평가의 어려움이 있었다. 화자식별 기능의 부재로 복수의 내담자가 존재하는 집단상담시 화자별 녹취가 필요하다는 단점이 있다. 본 연구에서는 내담자의 말을 녹취하는 방법을 택했지만 컴퓨터 혹은 기타 주변기기를 이용해 실시간 음성인식을 구현하여 시각화하는 방법도 생각해 볼 수 있다. 이에 관해서는 후속 연구를 기약한다.

참 고 문 헌

[1] 김도연, 조민기, 신희천, 상담 및 심리치료에서 인공지능 기술의 활용: 국외사례를 중심으로, May, 2020.

[2] Awni Hannun, Carl Case, Jared Casper et al., Deep Speech: Scaling up end-to-end speech recognition, Dec, 2014.

[3] Mattias Heldner, Jens Edlund, Pauses, gaps and overlaps in conversations, Oct, 2020.

[4] Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko et al., GoEmotions: A Dataset of Fine-Grained Emotions, Jun, 2020.

[5] Jacob Devlin, Ming-Wei Chang, Kenton Lee et al., BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, Oct, 2018.

기상 IoT 센서 데이터를 활용한 서리 예측 연구

서재하[1], 안은지[2], 문애경[3], 박주영[3]

[1]경희대학교, [2]안동대학교, [3]ETRI

jaeha.5eo@khu.ac.kr, {akmoon, jypark}@etri.re.kr

Frost Prediction Research Using IoT Sensor Weather Data

Jaeha Seo[1], Eunji An[2], Aekyeong Moon[3], Juyoung Park[3]

[1]Kyunghee Univ., [2]Andong Univ., [3]ETRI

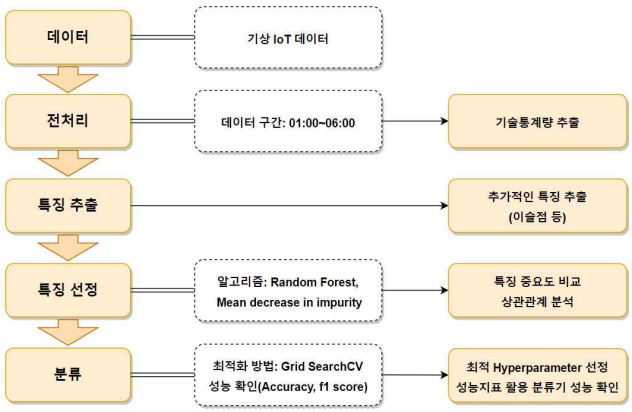
요 약

서리 피해에 미리 대응하지 못해서 발생하는 농가의 손실을 최소화하기 위해서는 정확한 서리 예보가 필수적이다. 본 논문에서는 서리 발생에 영향을 주는 기상인자에 머신러닝을 적용하여 서리 발생을 예측하는 방법을 제안한다. 농업 기상 IoT 데이터를 전처리한 후 랜덤 포레스트를 기반으로 한 특징 선정을 통해 서리 발생에 주요한 영향을 주는 변수들을 도출하고, 교차 검증과 적합한 성능지표를 통해 모델을 최적화하였다. 실제 서리 관측 데이터를 통해 검증을 수행한 결과 우수한 성능을 가지는 랜덤 포레스트 기반의 서리 예측 모델을 구현할 수 있었다.

Ⅰ. 서 론

최근 지구온난화 현상으로 인해 작물의 개화시기가 앞당겨지고 있고 이때 발생하는 서리는 작물 생육에 큰 영향을 미친다. 서리가 발생하는 시기를 정확히 예측할 수 있다면 농민들은 서리를 대비할 수 있고, 서리 발생으로부터 오는 손실을 막을 수 있다[1]. 따라서, 봄, 가을철 서리로 인한 농가 피해를 최소화하기 위해 정확한 서리 발생 예측이 필요하다. [2][3]에서는 서리 예측의 정확성을 높이기 위한 서리 발생일의 기온 위주 기상 특성을 제시하였다. 서리는 일반적으로 날씨가 맑으면서 춥고 바람이 잘 불지 않을 때 발생한다. [4]의 연구에서 서리 발생 예측을 위해 도출된 기상인자는 최저기온, 평균풍속, 상대습도, 이슬점 온도이다. 서리가 발생한 날이 서리가 발생하지 않은 날에 비해 최저기온, 이슬점 온도, 평균풍속은 낮게 나타나고 상대습도는 높게 나타난다.

본 논문에서는 농업 기상 IoT 데이터를 전제로 머신러닝 알고리즘을 사용하여 서리 발생을 예측한다. 그림 1은 본 논문에서 제시한 서리 예측을 위한 전체적인 데이터 분석 단계를 나타낸다. 특징 선정 단계에서는 랜덤 포레스트 기반의 특징 중요도를 활용하여 서리 발생에 주요한 영향을 주는 변수들을 결정한다. 특징 중요도를 계산하기 위하여 랜덤 포레스트 분류기 기반의 MDI(Mean Decrease in Impurity) 모델을 활용한다. 분류 단계에서는 서리 예측 모델의 구축을 위해 분류기 알고리즘을 구현하며 최적화 및 두 가지 성능지표를 활용하여 분류기 성능을 확인한다. 이 연구를 통해 서리 예측 정확도를 높임으로써 서리 피해를 사전 예방하고자 한다.



[그림 1] 데이터 분석 단계

Ⅱ. 본론

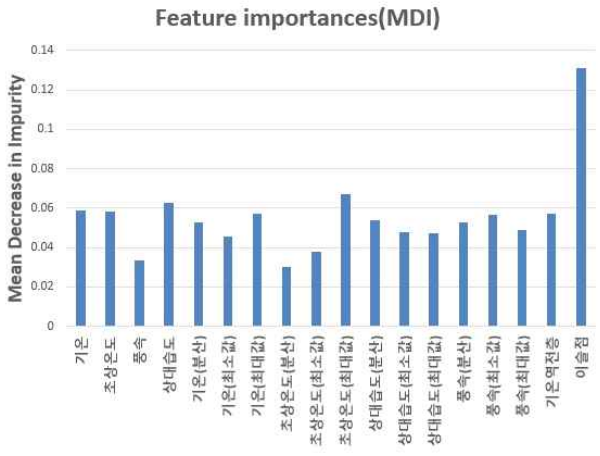
2.1 데이터 전처리 및 특징 선정

본 연구에서는 2020년 1월 1일부터 5월 31일까지 경상북도 영천시 화북면의 IoT 기상 관측 장비에서 수집한 기온, 상대습도, 강수량, 초상 온도 등의 센서 데이터를 활용한다. 관측 기간 152일 중 서리가 관측된 날은 29일이고, 서리 현상이 발생하는 새벽 시간대(01:00 ~ 06:00)의 데이터를 분류 모델 학습에 사용한다.

머신러닝 모델을 학습시키기 위해서 설정한 시간대의 데이터를 요약하여 대표할 수 있는 새로운 데이터 세트를 산

술 평균, 최대/최솟값, 분산 등의 기술 통계량을 활용하여 생성한다. 또한, 기존 연구를 바탕으로 서리 발생에 영향을 주는 변수로서 기온역전층[1]과 이슬점을 추가 정의한다.

본 연구에서는 MDI 알고리즘을 이용하여 각 특징의 중요도를 계산한 뒤, 높은 중요도의 특징들을 활용하여 모델을 학습한다. MDI 중요도는 트리형 알고리즘에서 각 데이터가 split될 때 불순도의 감소분에 따라 각 특징의 중요도를 계산한다.



[그림 2] 특징 중요도(MDI)

그림 2는 각 특징의 MDI 중요도를 나타낸다. 추가한 기온역전층과 이슬점이 가장 높은 중요도를 가졌고, 상대습도의 최솟값과 기온의 분산 값 또한 높은 중요도를 가진다. 이 결과에 따라 기온역전층과 이슬점, 상대습도의 최솟값, 그리고 기온의 분산 값을 활용하여 모델을 학습한다.

2.2 실험 및 성능 평가

본 장에서는 랜덤 포레스트의 기본 설정 모델과 초매개변수(Hyper Parameter) 최적화 모델의 성능 지표 비교를 통하여 머신러닝 모델의 발전 가능성과 이를 사용한 서리 예측 가능 여부를 확인한다.

랜덤 포레스트 분류기 모델을 생성할 때 설정하는 트리 개수, 트리의 최대 깊이 등의 여러 초매개변수에 따라 분류기의 성능 차이가 날 수 있다. 본 연구에서는 Scikit-learn에서 제공하는 GridSearchCV를 통해 초매개변수 변화에 따른 성능을 측정하여 초매개변수를 설정한다. 표 1은 기본 설정 모델과 최적화 모델의 초매개변수를 비교한 표이다.

	Parameter Name	Value
Default	Number of estimator	100
	Max depth of estimator	None
	Class weight	None
Optimized	Number of estimator	1000
	Max depth of estimator	3
	Class weight	'balanced'

[표 1] 초매개변수 비교

두 모델 성능 평가는 아래의 식으로 계산되는 두 가지 성능지표(Accuracy, F1-score)를 통해 진행한다.

$$(Accuracy) = \frac{TP+TN}{TP+FN+FP+TN} \quad (1)$$

$$(F1-score) = 2 \times \frac{\mathrm{Precision} \times recall}{\mathrm{Precision} + recall} \quad (2)$$

서리가 관측된 날의 데이터 세트보다 그렇지 않은 날보다 4배 정도 많은 편향된 데이터 세트를 모델의 학습에 활용했기 때문에, 이러한 편향된 데이터 세트의 성능을 잘 표현해 줄 수 있는 F1-score 값이 중요하다. 기본 설정 모델의 Accuracy와 F1-score는 각각 0.8, 0.25이고, 최적화 모델은 0.93, 0.83으로 두 지표 모두 최적화 이후 상승하였다.

Ⅲ. 결론

본 연구에서는 서리 발생 예측에 필요한 중요 인자들을 도출해보고, 랜덤 포레스트 모델 최적화를 통해 서리의 발생을 예측해보았다. 전체 데이터 세트의 숫자가 크지 않았고 편향적인 데이터 세트를 활용하였지만 적절한 특징 데이터를 도출하고, 최적의 초매개변수를 설정함으로써 서리의 발생을 정확하게 예측할 수 있는 가능성을 실험을 통해 확인하였다. 보다 많고 다양한 기상 데이터를 확보한다면 본 연구에 기초하여 정확한 서리 예측으로 농작물 피해를 예방할 수 있을 것으로 기대된다.

ACKNOWLEDGMENT

본 결과물은 농림축산식품부의 재원으로 농림식품기술기획평가원 첨단생산기술개발사업의 지원을 받아 연구되었음 (317015-6)

참 고 문 헌

[1] Yosuke Tamura, Liya Ding, Kosuke Noborio, Kazuki Shibuya, "Frost Prediction for Vineyard Using Machine Learning", 2020 Joint 11th International Conference on Soft Computing and Intelligent Systems and 21st International Symposium on Advanced Intelligent Systems (SCIS-ISIS), Dec 2020

[2] 권영아, 이효신, 권원태, 부경온, "서리 피해 방지를 위한 서리 발생일의 기상 특성에 대한 연구", 대한지리학회지 제43권 제6호, 2008.12

[3] 노일석, 도해원, 김수옥, 이승재, "2018년 봄철 서리 발생일 기상 분석", 한국기상학회 학술대회 논문집, 2019.10

[4] Y.S. Kim, K.M. Shim, M.P. Jung, I.T. Choi, "Study on the Estimation of Frost Occurrence Classification Using Machine Learning Methods", Korean Journal of Agricultural and Forest Meteorology, Vol. 19, No. 3, (2017), pp. 86~92, 2017

제2회 한국 인공지능 학술대회

The 2nd Korea Artificial Intelligence Conference

세션 E

치안현장 의사결정 지원을 위한 상황 대응 분류 모델

권은정, 변성원, 박현호, 장동만, 정의석
한국전자통신연구원

ejkwon@etri.re.kr, swbyon@etri.re.kr, hyunhopark@etri.re.kr, dmjang@etri.re.kr, esjung@etri.re.kr

A Situational Response Classification Model to Support Decision-Making in Police Field

Eun-Jung Kwon, Sungwon Byon, Hyunho Park, Dongman Jang, Euisuk Jung
Electronic and Telecommunications Research Institute (ETRI)

요 약

본 논문은 수시로 변화하는 치안현장에서 발생하는 데이터를 분석하여 현장경찰관들이 접하는 각종 사건 유형에 따라 대응정보를 제공하는 분류모델을 제안한다. 이를 위하여 112 신고를 받고 출동하는 현장경찰관에게 제공되는 출동지령 데이터, 사건현장에서 현장경찰이 발화하는 질의 데이터의 특징을 분석하여 치안업무를 지원하기 위한 상황대응 모델을 제시한다. 현장 상황 유형 기준에 따라 특징이 되는 주요 변수 및 데이터의 주요 키워드를 가중치 단어로 선정하여 모델 학습 수행 및 서비스 가능성을 제공한다.

Ⅰ. 서 론

범죄현장에 출동하는 현장 경찰관은 지령정보를 통해 사건발생현장의 상황을 인지하고 현장으로 출동하게 된다. 이과정에서 현장 상황에 따라 경찰은 현장 상황 대응에 필요한 법령 및 경찰 업무준수 지침 등을 숙지할 필요가 있다. 또한, 현장의 위험 상황에 따라 신속하게 대응하기 위한 방안을 강구해야 한다. 이러한 판단은 현장 경찰관 개인의 능력 및 경험에 따라 대응 조치가 달라질 수 있으므로 효과적인 현장 대응이 가능한 지원 시스템 필요하다. 통계에 따르면 2019 년 112 신고에 대한 출동지령은 연 총 12 백 54 만 5 천건으로, 중요범죄 40 만건의 출동지령이 포함된다. 이러한 이유로 본 눈문에서는 현장경찰이 각종 사건에 대응 시 업무지원을 효과적으로 제공받을 수 있도록 치안현장 데이터기반 상황 대응 분류 모델을 제안한다.

Ⅱ. 관련연구

112 신고가 접수되면 출동지령을 통해 사건 발생 위치로 신고 종류와 긴급성에 따라 출동 경찰에게 신고 내용과 출동 단계가 전달된다. 현장 경찰은 전달받은 신고 내용을 이해하고 순찰차로 이동하는 과정에서 현장의 상황을 추측, 도착 후 상황 업무를 수행하게 된다. 현장 상황에 대한 대응 방안은 업무범위 지침내에서 현장경찰의 판단에 의해 이루어진다. 현장경찰에게 경찰 순찰차 내부 탑재 단말기기 또는 경찰 보유 개인 휴대기기를 통해서 현장상황에 대응하기 위한 적합한 안내정보를 적시에 제공받게 된다면 치안현장의 안전은 더욱 강화될 것이다.

범죄는 강력범죄, 절도, 폭력범죄, 지능범죄, 풍속범죄, 교통범죄 등 약 15 개의 죄종으로 분류되고, 각 죄종은 다시 세부 항목으로 정의되어 범죄 발생 시간, 발생장소 기준으로 세부적인 통계분석 결과 확인이 가능하다[1]. 정형적인 데이터 분석 결과 외에 112 신고내용으로부터 텍스트 마이닝 분석을 통해 행위 주체와 대상, 행위 유형, 장소에 관한 키워드 분류를 통해 112 신고 분석시스템 및 범죄발생 예측시스템의 구축에 대한 시사점을 연구한 바 있다[3]. 또한, 최근 언어모델이 단순히 순차적으로 특정 단어 기준으로 이후 단어를 예측하지 않기 때문에 신경망의 모든 레이어에서 문장 내 단어 순서, 단어 간의 관계, 문맥 정보를 확인할 수 있는 딥러닝 기반 언어모델을 활용하여 기존 전통적인 방법의 비정형 데이터 분석의 한계를 극복한 사례들이 제시되고 있다. 이러한 연구성과는 기존 데이터 분석 방법 대비 대용량의 학습데이터가 필요하고 이로 인해 학습 시간도 증가하지만, 벡터의 저장공간에 대한 희소성 문제점, 동음 이의어 구별 등 기존 언어모델의 한계점을 극복할 수 있기때문에, 사전 학습된 언어모델을 통해 도메인 말뭉치를 이용하여 다운스트림 태스크 성능의 개선으로 활발히 이용되고 있다[3]

Ⅲ. 치안 현장 상황대응 분류 모델링

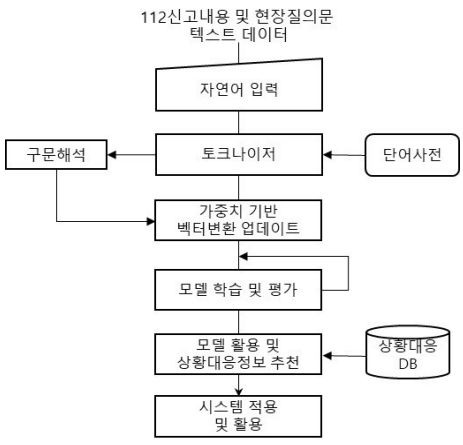
본 장에서는 112 상황실에 접수되어 현장경찰의 이동 기기로 현장상황 대응을 위한 정보를 자동 제공하기 위한 모델 구조와 학습용 데이터로 사용되는 데이터 구조 및 속성에 대해 설명한다. 모델 학습을 위한 데이터 중 개인정보가 포함된 부분은 정규 표현식, 고유 개체명 인식기를 이용하여 데이터 전처리 과정을 수행한다. 또한, 상황 별 문맥상 의미가 통하지 않는 데이터도 학습 데이터에서 제외한다.

그림 1 은 본 논문에서 제안하는 모델학습을 위한 전체 워크플로우 처리에 대한 개념을 표현한 것이다.

그림 1 은 경찰 출동지령시스템을 통해 신고접수 종별 분류코드 기준에 따라 텍스트 형태로 현장 경찰의 이동 단말로 대응정보를 전달하기 위한 모델 구조이다. 상황 대응에 대한 세부기준은 현장상황의 위험도를 고려해야 하기 때문에 출동지령 데이터를 구성하는 주요 단어 및 문맥의 의미에 가중치를 부여하는 구조를 갖는다. 이에 따라 본 논문에서는 치안 도메인에서 상황 별 단어가 가지는 빈도와 위험 유형에 따라 해당 단어에 대한 가중치를 설정, 출동지령 텍스트 문장 구분을 위한 1 단계 작업을 수행한다. 예를 들어, '폭력'으로 분류된 출동지령 텍스트는 '때리고', '맞음', '과격' 과 같은 단어에 대한 가중치 값이 "방화"로 분류된 출동지령 텍스트의 주요 키워드 값보다 높은 값을 갖도록 한다. 가중치 기법을 기준으로는 출동지령 텍스트 문장의 전체 문맥에 대한 이해가 어렵기 때문에 문맥에 대한 의미까지 파악하기 위해 학습 알고리즘이 필요하다. 본 논문에서는 상황 별 가중치가 높은 단어는 치안 도메인 용어사전을 정의하여 선정한다. 출동지령 텍스트 데이터는 토크나이저를 이용하여 문장을 공백 단위로 분절, 분절된 단어를 다시 서브단어로 구분하여 벡터 값으로 변환된 값을 통해 모델 학습이 이루어진다. 이 과정에서 사전 학습된 언어모델을 통해 다운스트림 태스크 성능이 개선될 수 있도록 모델을 학습하고 그 결과에 대한 성능을 검증한다.

모델 학습에 사용되는 데이터세트의 일부 항목에 대한 예는 표 1 과 같다. 112 신고 접수 시 기록되는 사건종별과 현장경찰이 사건 현장에서 종결처리과정에서 분류하는 값이 다름을 알 수 있다. 다시 말해, 신고를 받고 출동하는 현장경찰은 출동지령 시 제공되는 일부 정보만으로는 현장의 다양한 상황을 대응하기 어렵다고 이해할 수 있다. 실제 실험에 적용된 출동지령 데이터가 딥러닝 알고리즘을 수행하기에 충분하지 않다면 특정 범죄 죄종으로 편향(overfitting)될 수 있기 때문에 성능 개선을 위해 데이터 증강 및 랜덤 오버 샘플링 기법을 통한 학습방식이 요구된다.

본 논문에서 데이터를 증강하는 방식은 치안현장 도메인 용어사전을 정의하고 활용한다. 학습용 데이터로 문장을 구성하는 주요 단어 중심으로 자주 등장하는 단어와 관련된 단어를 관계어로 정의, 추가 생성 하여 학습용 데이터로 보강한다. 이방법은 표 1 의 '신고내용' 텍스트 문장에 대해 예측되는 '사건종별' 값을 획득하여 상황대응 DB 로부터 관련된 대응정보 추천 받을 수 있도록 시스템에 적용한다.



(그림 1) 현장 상황 대응 분류 모델 적용을 위한 워크플로우 흐름도

(표 1) 상황대응분류를 위한 가정폭력 데이터 샘플 예시

코드	사건종별	종결분류	종결시사건코드	신고내용
C1	폭력	현장종결	가정폭력	위층에서 엄청싸우는소리가 들린다고/ 신고자201호에 살고있고/
C1	소음	현장종결	가정폭력	가동 엘리베이터에서 조등 말을 심하게 때리서 우는소리가 크게 들린다며
C1	가정폭력	검거	가정폭력	남편이 베트민턴 채로 아이들을 위협 다급
C1	가정폭력	현장종결	가정폭력	가족이 싸운다며
C1	가정폭력	현장종결	가정폭력	아버지가 치매이신데, 어머니를 때리려고 한다고, 어머니한테 연락 왔다며,어머니,
C1	가정폭력	현장종결	가정폭력	부부싸움하고 있다/ 빨리오라며
C1	가정폭력	검거	가정폭력	동생이 술마시고 정신없이 날뛴다며경찰요청
C1	가정폭력	허위오인	가정폭력	술취한 아들이 행패를 부린다며/그동안 신고를 많이 했는데 조치가 잘 되지 않아서 불안하다며조치가 필요하다고./
C1	가정폭력	현장종결	가정폭력	옆집에서 싸우고 있다고가정폭력 같다고 502호에서
C0	가정폭력	현장종결	가정폭력	[CODE0 선지령]생명의 위험을 느낀다고//아버지가 몽둥이 들고 쫓아온다고//신고자 핸드폰 놓고와서 연락처 없다고//
C1	가정폭력	검거	가정폭력	아들이 폭행을 하고 있다고
C2	가정폭력	현장종결	가정폭력	아침에 아들과 좋지 않는 일이 있어서 경찰서 갔다와다/지구대 들렀더니 집 근처에 오면 112에 신고하라고 했다며/
C1	가정폭력	비출동 종결	가정폭력	정신이 이상한것 같다는/둘째 언니를 때린다고/빨리와달라고
C1	가정폭력	현장종결	가정폭력	형이 아버지하고 싸움이 있다며//형이 물었다며//누나도 다쳤다며/

현장에 출동한 경찰이 실제 출동 당시 전달받은 내용과 현장 상황이 다른 양상으로 전개가 된 경우, 경찰 업무 지원 휴대기기에 설치된 현장지원 어플리케이션을 통해 질의문을 받고, 변경된 현장 상황에 따른 적합한 대응정보를 수신 받을 수 있도록 한다. 모델의 성능 및 정성적인 평가를 위해서 사건 종별 레이블링 된 값이 정확히 매칭되었는지 확인하기 위해서는 고차원 단위의 '신고내용' 텍스트 문장을 구성하는 단어들의 벡터값을 2 차원으로 축소 변환 후 '종결시사건코드'와 관계를 t-SNE 를 통해 시작적으로 표현하여 확인한다. 또한 분류모델 평가지표인 F1 Score 에 따라 합성 데이터를 추가 생성하거나 딥러닝 모델의 하이퍼파라미터를 조정하여 모델 성능 개선을 위한 파인튜닝 작업을 수행한다.

IV. 결 론

빅데이터 기반 지식 구축 시장은 꾸준히 성장하고 있고 공공분야에서도 새로운 서비스를 통한 경쟁력 제고 필요성이 대두되고 있다. 본 논문은 현장경찰의 업무를 지원하기 위해 치안현장의 각종 상황들에 대응하기 위해 112 출동지령 및 현장 경찰의 주요 질의내용을 통해 적합한 대응정보 제공을 위한 전체 워크플로우를 제안하였다. 성능이 우수한 모델은 향후 테스트 케이스를 설계하여 실제 시스템에 적용 가능성을 확인할 계획이다.

ACKNOWLEDGMENT

이 논문은 2021 년도 정부(경찰청)의 재원으로 지원받아 수행된 연구결과임 [내역사업명: 112 긴급출동 의사결정 지원 시스템/연구개발과제번호: PR08-03-000-21]

참 고 문 헌

[1] 경찰범죄통계-범죄 발생 상황 관련 특성, 경찰청, https://stat.kosis.kr/nsieu/view/tree.do?task=branchView&id=132_13204*MT_OTITLE&hOrg=132

[1] 박진수, 박현호, 조준택, " 112 신고내용에 대한 텍스트 마이닝 분석과 시사점, "한국치안행정논집, 제 3 회(2018), pp. 159-186

[3] J. Devlin, M.W. Chang, K. Lee, K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," arXiv:1810.04805

IoT센서와 영상 학습을 통한 조기 이상 탐지 시스템 설계

홍기대, 박치원 허의남*
경희대학교
{kidae92, bamtorrr, *johnhuh}@khu.ac.kr

Early Anomaly Dectection System Through IoT Sensor and Image Learning

Gi-Dae Hong, Chi-Won Park, Eui_Nam Huh*
Kyung hee University.

본 논문은 딥러닝 기반의 영상 인식 모델과 IoT 센서를 이용하여 조기 이상 탐지 시스템을 제안하였다. CCTV 영상을 통해 딥러닝 모델의 영상 인식 결과와 센서값에 LSTM 기술을 접목시켜 이를 기반으로 조기 이상을 판단한다. 이는 LSTM을 통한 이상 데이터 변화 추세를 감지하고 화재 감지 센서 위치가 화재 발생 위치와 거리가 떨어져 있을 경우, 탐지하는 시간이 늦을 수 있는 점을 Object Detecting 기술을 더해, 빠르고 정확한 이상 탐지를 돕는다.

Ⅰ. 서 론

화재감지 및 경보시스템의 설계 목표는 화재발생 초기 단계에서 화재징재를 발견 및 피난의 개시를 신속하게 통수하는 것이다. 하지만 화재감지기는 다양한 건물 및 환경에 설치되기 때문에 주위 온도를 포함한 다양한 환경적 측면을 심층적으로 고려하지 않으면 적절하게 감지하지 못하는 경우가 발생할 수 있다.[1]

최근 대형 물류센터 사고 화재의 경우에도 화재 감지기의 잦은 오작동으로 인해 근무자들의 안전불감증이 사고 규모를 키운 것으로 분석됐다.[2] 더 나아가 연기 또는 온도 감지로만 작동하는 화재 감지 시스템의 구식화 문제가 오작동으로 인한 안전불감증 상승 요인으로 볼 수도 있다. 또한 CCTV 화면에 연기와 화재가 감지 됐음에도 불구하고, 아무도 인지를 못한 것 또한 사고를 키웠다고 볼 수 있다.[3] 이에 따라 화재 경보기는 다양한 데이터를 바탕으로 실화재와 비화재를 정확하게 구분하고 이에 정확한 경보를 제공하는 것이 가장 중요한 관건이다.[4]

본 논문에서는 온도와 같은 단일 인자로 화재 감지를 하는 방법이 아닌, 온도 감지기에 LSTM(Long Short-Term Memory)의 접목과 딥러닝 기반의 영상 학습 기술을 이용하여, 신속하고 정확한 비상 상황 판단을 하는 조기 이상 탐지 시스템을 제안한다.

Ⅱ. 본론

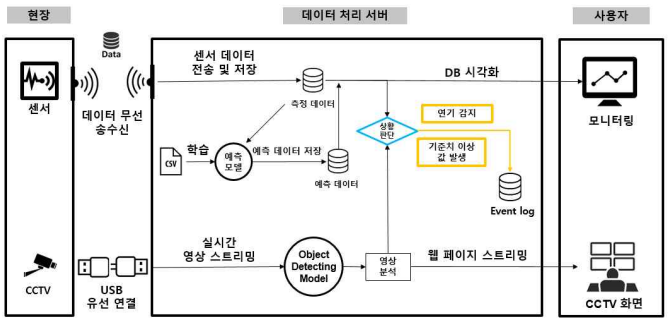
Ⅱ-1. 시스템 구조

그림 1은 조기 이상 탐지 시스템의 전체 구성도이다. 시리얼 통신을 지원하는 온, 습도 센서가 ZigBee RF 무선 송수신기를 통해 1초 간격으로 데이터를 전송하고, DB에 저장이 된다. 저장된 데이터는 LSTM 기반으로 학습된 모델을 통해 20개 단위로 10초 뒤 예측값을 생성 및 저장하는데 사용된다. 학습모델은 화재 발생과 같은 갑작스런 온도 변화를 예측 할 수 있도록 급격한 온도변화가 있는 데이터셋 10만개를 구축하여 학습시켰다.

측정된 데이터와 예측 데이터는 웹페이지를 통해 실시간으로 모니터링이 가능하다. 값의 기준은 정상, 관심, 주의, 경고, 심각 총 5단계로 구분하며 관심 값 이상부터는 Event log에 남는다.

설치된 CCTV는 실시간으로 해당 구역을 스트리밍한다. 스트리밍 영상은 Object Detecting Model을 통해 실시간으로 분석된다. 해당 모델은 연기가 찍힌 데이터셋[5]을 기반으로 학습 시켰다.

LSTM 예측 모델과 Object Detecting Model이 감지 및 분석한 결과를 토대로 조기 이상 탐지 시스템을 구현하였다.



〔그림 1〕 조기 이상 탐지 시스템 구조도

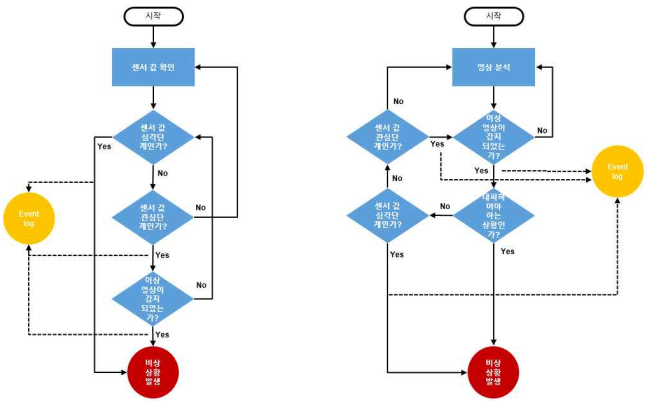
Ⅱ-2. 시스템 시나리오

그림 2는 본 논문에서 제안하는 시스템의 전체 흐름도(workflow)이며, 센서 데이터와 영상 분석 결과에 따른 두 개의 흐름도로 나뉜다.

먼저 측정 및 예측값은 정상, 관심, 주의, 경고, 심각 총 5단계로 분류된다. 그림 2 (좌)는 측정 및 예측값에서 먼저 이상 값이 탐지되었을 때의 시스템 흐름도이다. 데이터값이 위험단계로 측정될 경우 즉시 비상 상황 발생으로 판단한다. 관심 단계라면 이상 영상 감지 여부에 따라 상황 판단을 하며, 데이터 값이 관심 단계 이상에다가 연기가 감지되었다면 비상 상황 발생으로 판단한다. 이상 영상이 감지되지 않았다면 데이터 확인, 영상 분석 결과를 지속적으로 확인한다. 관심 값 이상값과 이상 영상 인식 결과들은 모두 Event log에 저장된다.

그림 2 (우)는 센서 위치에 따라 이상 영상이 먼저 감지되는 경우에 대한 흐름도이다. 이상 영상이 감지 된 경우, 연기의 크기에 따라 비상 상황으로 판단한다. 그렇기 않은 경우 센서 데이터를 참고하게 된다. 측정 및 예측 값이 심각 단계이면, 또한 비상 상황으로 판단하며 관심 값 이상 단계

면 이상 영상감지 여부 및 센서 값 단계를 주시하며 판단하게 된다.



〔그림 2〕 (좌) 센서 데이터 시스템 흐름도
(우) 영상 분석 시스템 흐름도

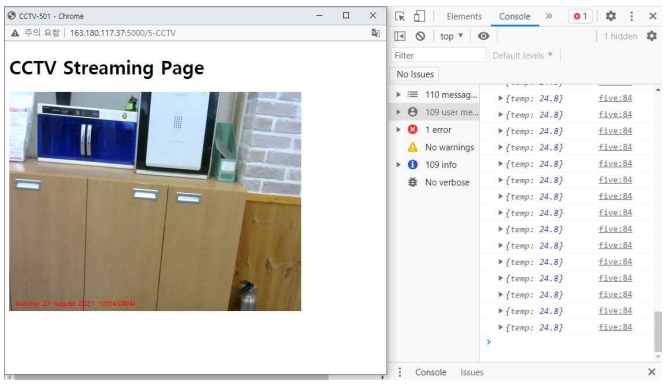
Ⅱ-3. 화재감지 시스템 프로토타입

그림 3은 온/습도 센서 값과 LSTM의 예측 결과를 시각화 한 모습이다. 이는 실시간 값과 예측값이 저장되는 DB를 연동시켜 Grafana로 출력했으며, 이를 통해 해당 구역의 상태 및 변화를 쉽게 관찰할 수 있다.



〔그림 3〕 실시간 센서값과 예측값의 시각화

그림 4는 해당 구역 화면을 웹페이지로 띄워 실시간으로 모니터링을 할 수 있도록 구성했다. Logitech C270 웹캠을 사용했으며, Grafana와 별개로 CCTV 페이지에서 센서 데이터를 console log로 확인 할 수 있다. 이 웹캠을 통해 Object Detecting Model이 화염 및 연기를 감지하며, 영상 감지 결과와 센서 데이터를 기반으로 상황 판단을 한다.



〔그림 4〕 CCTV Streaming Page

Ⅲ. 결론

본 논문에서는 기존의 단일 센서에 의존한 화재 감지 방식을 보완하기 위해, 센서값을 활용한 LSTM 예측값과 딥러닝 기반의 영상 인식 결과를 기반으로 신속, 정확한 조기 이상 탐지 시스템을 소개하였다.

본 연구에서 가상 연기를 발생시켜 웹캠을 통해 실험을 해보았으나 작은 연기는 인식 정확도가 낮게 측정되었다. Object Detection Model 학습 데이터 셋이 대형 화재로 인해 발생한 연기를 바탕으로 진행하여, 정확도가 낮게 나온 것으로 예측된다. 연기의 크기, 색깔 등을 고려하여 다양한 데이터 셋을 바탕으로 인식 정확도가 높은 학습모델을 만들어 향후 연구를 추가 진행할 계획이다.

또한 화재 발생 인지를 판단 할 수 있는 요소는 온도를 포함한 CO2, CO 등 여러 요소가 존재한다. 다양한 센서를 활용하여, 화재 발생 시 복합적인 요소들의 연관성을 파악해 LSTM 학습을 한다면, 보다 정확한 조기 이상 탐지 시스템이 될 것으로 기대된다.

ACKNOWLEDGMENT

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2021-0-00818, 스마트 플랜트 이상상태 조기감지를 위한 머신러닝 기반 저대역 영상/통신 Edge Computing 시스템).

참 고 문 헌

[1] 최문수, 홍성호, 이상호, 박상태, 유송현 “가연물에 따른 화재감지기 응답특성에 관한 실험연구” 2011년도 추계학술논문발표회 논문집 p-32

[2] 한겨레 https://www.hani.co.kr/arti/area/capital/1004118.html ,2021-07-19

[3] 매일경제 https://www.mk.co.kr/news/society/view/2021/06/588449/ ,2021-06-17

[4] “신개념 화재 감지용 CO & CO2 멀티가스 감지기 개발” ㈜인바이런먼트 리딩 테크놀리지(이엘티) 한국소방산업기술원 kfi-2009-산지-008

[5] Wildfire Smoke Dataset, roboflow, 2020.10
https://public.roboflow.com/object-detection/wildfire-smoke/fork

로지스틱회귀 모델을 이용한 익명처리 데이터의 유용성 측정

성민경, 이진규, 정성룡, 김윤중*, 김동례*

한국정보통신기술협회, ㈜이지서티*

[mksung, jklee, srjeong]@tta.or.kr, [yjkim, drkim]@easycerti.com

Data Usefulness Measurement on an Anonymized Dataset Using Logistic Regression Model

Sung Min Kyoung, Lee Jin Kyu, Jeong Seong Ryong,

Kim Yun Joong*, Kim Dong Rye*

Telecommunications Technology Association, EASYCERTI Co., Ltd.*

요 약

데이터는 4차산업혁명의 원동력이 되는 중요한 자원이지만 개인정보 처리 문제로 인해 활용에 큰 제약을 받고 있다. 이를 해결하기 위하여 정부는 가명·익명처리를 통한 비식별 처리 데이터 활용을 법·제도 개편을 통해 가능하게 하고 있다. 본 논문에서는 대표적인 익명화 기법인 K-익명성 모형에 의한 익명처리 데이터의 유용성을 로지스틱회귀 모델의 예측 정확도 측정을 통해 알아본다. 또한, 예측에 사용되는 데이터를 변경하면서 측정을 수행하여 데이터 유용성 측정에 대한 모델을 제안한다.

Ⅰ. 서 론

빅데이터는 제4차 산업혁명의 원동력이라 할 수 있을 만큼 필수적인 요소이다. 빅데이터는 기존산업 발달, 신산업 도출, 새로운 지식 축적 등 다양한 분야에 활용될 수 있으나 개인정보 보호의 문제로 유통·활용이 매우 제한되고 있다. 이에 대한 대안으로 데이터 비식별 처리를 통해 데이터를 가명화 또는 익명화하여 활용하는 방안이 각국의 정부로부터 제시되고 있다. 비식별 데이터의 활용을 통해 개인정보 보호와 빅데이터 활용 이라는 두 가지 상충되는 목적을 동시에 달성할 수 있다.

데이터 비식별화는 개인을 구분/구별할 수 있는 정보의 일부 또는 전체를 삭제·수정하여 개인의 식별성을 낮추는 기술이다. 비식별 수준이 클수록 개인정보 보호는 강해지고 데이터 유용성은 낮아지며, 비식별 수준이 낮을수록 개인정보 보호는 약해지고 데이터 유용성은 높아진다. 즉, 데이터 활용 상황, 환경, 목적에 따라 비식별 수준이 결정되어야 하며, 유용성 측정은 해당 과정에서 필수적으로 수행되어야 하는 작업이다.

본 논문에서는 빅데이터의 안전한 활용을 위한 비식별 처리 데이터의 유용성을 로지스틱 회귀 모델의 예측 정확도를 통해 측정한다. 또한, 학습 및 추론(예측)에 활용되는 데이터의 속성(컬럼) 변경에 따른 분석을 통해 데이터 유용성 측정에 대한 모델을 제안한다.

Ⅱ. 배경지식

1. 기존연구

비식별 처리 데이터의 유용성 측정을 위한 기존 연구는 크게 두 가지 방식으로 나뉜다. 첫 번째 방식은 비식별된 데이터 자체의 변경 정도를 원본데이터와 비교하는 방법이다[1,2]. 이 방식은 공식화된 수식을 통해 정량적인 수치를 얻을 수 있지만 데이터를 활용하는 상황을 반영하지 못하는 단점이 있다. 두 번째 방식은 비식별 데이터를 로지스틱 회귀 등 학습/추론 모델에 적용하여 원본데이터의 예측 정확도와 비교하는 방식이다[3,4]. [3]에서는 로지스틱 회귀 모델을 이용한 통계모형 정확도에 기반한 비식별 데이터의 품질을 측정하였다. [4]는 본 연구의 선행 연구로 다양한 비식별화 수준과 모델에 따른 로지스틱회귀 모델 정확도를 측정하였다.

2. 데이터 비식별 처리 모델

데이터 비식별 처리는 가명처리와 익명처리로 나뉘며, 본 논문에서는 익명처리 모델을 활용하므로 이와 관련된 KLT 및 차분프라이버시 모델에 대해서 소개한다.

K-익명성은 데이터에 포함된 개인을 적어도 데이터 내의 다른 K-1명과 구분이 되지 않게 데이터를 수정·삭제 하는 모델이다[5]. K-익명성을 만족하는 데이터집합 내에서 개인 식별 확률은 1/K 이하로 보장된다.

L-다양성은 같은 데이터 값을 가진 그룹에 속한 개인들의 민감한 개인정보 속성의 값이 서로 다른 L개 이상 존재하게 데이터를 수정·삭제 하는 모델이다[6]. L-다양성 모델을 만족하는 데이터집합은 최소 L-익명성을 만족하여 일반적으로 K-익명성 모델에 비해 더 높은 개인정보 보호 수준을 가진다.

T-근접성은 같은 데이터 값을 가진 그룹에 속한 개인들의 민감한 개인정보 속성분포와 전체 데이터의 민감한 개인정보 속성 분포를 비교하여 두 분포의 차이가 T이하가 되게 데이터를 수정·삭제·삽입 하는 모델이다[7].

차분프라이버시는 개별 질의에 대한 응답에 잡음을 삽입하여 응답하며, 통계 데이터를 배포할 때는 전체 데이터에 일부 잡음을 삽입하여 배포함으로써 개인정보를 보호하는 모델이다[8]. 각각의 상황에서 데이터 사용자는 정해진 수학적 모델을 준수하는 수준만큼의 잡음이 포함된 결과를 받게 된다.

Ⅲ. 본 문

1. 실험목적

본 실험의 목적은 익명 처리된 데이터의 유용성을 로지스틱회귀 모델의 예측 정확도를 통해 측정하는 것이다. 원본데이터의 예측 정확도를 기반으로 하여 낮은 수준의 익명화로부터 높은 수준의 익명화 데이터의 예측 정확도 변화를 살펴본다. 또한, 전체 데이터 중 학습에 사용되는 컬럼을 변경시키면서 예측 정확도를 측정하여 그 결과를 살펴봄으로써 새로운 데이터 유용성 측정 모델을 제안한다.

2. 실험설계

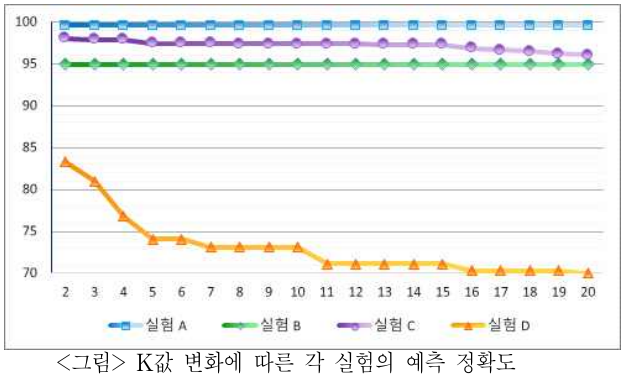
본 실험에서는 UCI Machine Learning의 Adult Dataset을 이용한다[9]. Adult Dataset은 나이, 교육정도, 성별, 결혼여부 등에 따른 급여의 상관관계를 나타낸 데이터이다. 데이터는 15개의 컬럼으로 구성되어 있으며, 'Salary'를 로지스틱회귀 모델의 종속변수로, 'Salary'를 제외한 14개 컬럼을 로지스틱회귀 모델의 독립변수로 설정한다. 14개 컬럼 중 9개 컬럼('Age', 'Workclass', 'Education', 'Relationship', 'Martial-status', 'Occupation', 'Race', Sex', 'Native-country')을 '식별가능정보'로 설정하여 비식별 처리 하였다. 종속변수 'Salary'컬럼은 '>50K'와 '<=50K' 두 가지 값을 가진다(원본데이터의 '>50K'와 '<=50K' 비율은 3:1). 익명화 모델은 K-익명성을 적용하였으며(K값은 2~20), ARX툴[10,11]을 이용하여 익명 데이터를 생성하였다.

텐서플로우 라이브러리를 통한 로지스틱회귀 모델을 이용하여 학습데이터와 테스트데이터를 8:2 비율로 나누어 'Salary'를 예측하였다. 비식별 처리된 데이터의 유용성을 측정하기 위해 다음과 같은 4가지 실험을 수행하였다.

실험 A	원본데이터 14개 컬럼 전체로 학습
실험 B	비식별대상 9개만 컬럼 원본 상태로 학습 (나머지 5개 컬럼 학습에 이용하지 않음)
실험 C	비식별 처리된 9개 컬럼+원본 5개 컬럼 학습
실험 D	비식별 처리된 9개 컬럼만 학습 (나머지 5개 컬럼 학습에 이용하지 않음)

3. 실험결과

<그림>은 K값 변화에 따른 실험 A ,B ,C ,D의 예측 정확도를 나타내고 있다. 실험 A와 B는 비식별 처리되지 않았으므로 K값 변화에 따른 예측 정확도 차이가 없다. 실험 C의 결과를 통해 2~20 사이의 K 값에서 95% 이상의 예측 정확도가 나왔다. 이는 비식별 대상 9개 컬럼만 원본 상태로 학습한 실험 B 보다 높은 수치로 비식별 처리 대상 컬럼이 있더라도 원본데이터가 보존된 컬럼이 일부 있다면 예측 정확도가 더 상승함을 알 수 있다. 또한 실험 D의 결과를 통해 비식별 처리된 데이터만으로 학습할 경우 예측 정확도가 매우 떨어짐을 알 수 있다. 이를 통해, 비식별 수준(여기서는 K값) 보다는 학습에 사용된 원본데이터 포함 여부가 더 큰 영향을 미침을 볼 수 있다.



<그림> K값 변화에 따른 각 실험의 예측 정확도

4. 유용성 측정모델 제안

본 실험 결과를 통해 로지스틱회귀 모델을 이용한 비식별 처리 데이터 유용성 측정모델을 제안한다. 우선 실험 A와 B를 통해 원본데이터 전체컬럼으로 학습한 모델과 비식별 처리 대상 컬럼들의 원본으로 학습한 모델의 예측 정확도를 측정하여 비교를 위한 기준점으로 삼는다. 그 후 실험 C를 통해 비식별 처리된 컬럼들과 비식별 처리되지 않은 컬럼들을 이용하여 예측 정확도를 측정하고, 실험 D를 통해 비식별 처리된 컬럼들만 학습에 이용하여 예측 정확도를 측정한다. 실험 A와 C의 비교를 통해 비식별 처리된 데이터의 유용성을 간접 측정할 수 있으며, 실험 B와 D의 비교를 통해 비식별 처리된 컬럼들의 유용성 정도를 측정할 수 있다.

Ⅳ. 결 론

본 연구에서는 로지스틱회귀 모델의 예측 정확도를 이용하여 비식별 처리 데이터의 유용성을 측정해 보았으며, 실험을 통한 유용성 측정 모델을 제안하였다. 학습에 사용되는 데이터가 변화함에 따른 예측 정확도 결과 분석을 바탕으로 유용성을 측정하는 한 방법을 보여주었다.

본 연구는 그동안 비식별 처리된 데이터 자체의 변경정도 측정에 치우쳤던 유용성 측정기법과는 다르게 데이터 활용 측면에서 유용성을 측정하는 방법을 사용하였다는 것에 의미가 있다. 다만, 다양한 타입의 데이터와 비식별 기법에 대한 내용을 다루지 못한 한계가 있다.

향후연구는 다음과 같다. 다양한 데이터 분포를 가진 데이터에 대한 결과 분석으로 데이터 분포에 따른 적절한 비식별 기법 적용방식을 선정하는 방법론을 연구해 볼 수 있다. 또한, 각 컬럼별 특성을 고려하여 각 컬럼에 대한 비식별 처리 수준을 결정하는 방법을 연구할 수 있다.

ACKNOWLEDGMENT

본 연구는 대용량 정형 데이터 대상 개인정보 가명/익명화를 위한 자동처리 기술 연구과제로 수행되었습니다.(2021-0-00634)

참 고 문 헌

[1] D. Robllo-Monedero, J. Forne, M. Soriano, and J. P. Allepuz, "k-Anonymous microaggregation with preservation of statistical dependence," Information Science, Vol.342, 2016.

[2] D. Sanchez, J. Domingo-Ferrer, S. Martinez, and J. Soria-Comas, "Utility-preserving differentially private data releases via individual ranking microaggregation," Information Fusion, Vol.30, 2016.

[3] 전희주, 이현지, 연규필, 김동례, "통계모형의 정확도에 기반한 비식별화 데이터의 품질 측정", 한국콘텐츠학회논문지 19(5), 2019.

[4] 성민경, 강지성, 유정승, "비식별 조치 데이터가 기계학습에 미치는 영향 분석", 한국정보과학회 학술발표논문집, 2019.

[5] L. Sweeney, k-anonymity: A model for protecting privacy, International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, Vol. 10, No. 5, pp. 557-570, 2002

[6] A. Machanavajjhala, J. Gehrke, D. Kifer and M. Venkitasubramanian, ℓ-diversity: Privacy beyond k-anonymity, in Proceedings of the IEEE International Conference on Data Engineering, 2006

[7] N. Li, T. Li and S. Venkatasubramanian, t-Closeness: Privacy beyond k-anonymity and ℓ-diversity, in Proceedings of the IEEE International Conference on Data Engineering, 2007

[8] C. Dwork, Differential Privacy. In: van Tilborg H.C.A., Jajodia S. Encyclopedia of Cryptography and Security. Springer, Boston, MA, 2011.

[9] UCI Machine Learning Repository, www.archive.ics.uci.edu

[10] ARX Data Anonymization Tool, arx.deidentifier.org

[11] Fabian Prasser, Florian Kohlmayer, Ronald Lautenschlaeger, Klaus A. Kuhn., ARX - A comprehensive tool for anonymizing biomedical data., in Proceedings of the AMIA 2014 Annual Symposium, November 2014, Washington D.C., USA.

딥러닝 및 이미지화를 통한 악성코드 분류 시스템

조현진*, 홍윤영*, 김호원*
*부산대학교
hyunjin@islab.re.kr, hyy0238@pusan.ac.kr, howonkim@gmail.com

Malware Classification System Using Deep Learning and Imaging

Jo Hyun Jin*, Hong Yoon Yung*, Kim Ho Won*
*Pusan National Univ.

요 약

최근 분산서비스거부(DDoS) 공격, 랜섬웨어, 개인정보 유출, 해킹 등 악성코드로 인한 피해가 급증하고 있다. 특히 기존 악성코드를 분석하던 기법으로는 은닉기술, 자가변형 등 발달한 분석 방해 기법으로 지능화된 악성코드들의 분석이 어려운 실정이며 악성코드의 종류가 다양해지고 있어 새로운 형태의 악성코드가 많이 등장하고 있다. 본 논문에서는 딥러닝 모델과 이미지화된 악성코드를 이용하여 악성코드를 분류하는 시스템을 제안한다. 시스템에서 사용된 딥러닝 모델인 샴 네트워크는 유사도를 기반으로 분류를 하여 새로운 형태의 악성코드에 대해서도 비슷한 악성코드 패밀리로 분류가 가능하다. 또한 적은 데이터로도 학습이 가능한 One-shot learning이 가능하여 적은 양의 데이터를 가진 악성코드 패밀리에 대해서도 분류가 가능하다. 제안하는 시스템은 평균 97.5%의 정확도로 악성코드를 분류하며 데이터가 적은 악성코드 패밀리에 대해서 약 90%의 정확도로 분류하며 일반적인 CNN(Convolutional Neural Network)과 비교하여 좋은 성능을 보인다.

Ⅰ. 서 론

최근 연구에서는 이미지 인식에 뛰어난 성능을 보이는 딥러닝을 이용하여 이미지화된 악성코드를 분석하는 기법들이 제안되고 있다 [1] [2]. 이미지화된 악성코드들은 같은 패밀리 간 유사성이 높게 나오기 때문에 이미지 인식을 하는 딥러닝으로도 높은 분류 성능을 확인할 수 있다. 하지만 기존 이미지 인식 모델들은 소량의 데이터로 훈련된 악성코드에 대해서 유사한 악성코드 패밀리가 있더라도 제대로 분류하지 못 하는 결과를 보인다. 이에 본 논문에서는 유사도 기반의 분류를 하는 샴 네트워크를 악성코드 분류에 적용하여 기존 패밀리와 유사한 형태의 소량의 악성코드 패밀리에 대해서도 악성코드로 분류되도록 하는 기법을 제안한다.

Ⅱ. 본 론

2.1 샴 네트워크

Gregory 등[3]은 하나의 데이터로만 학습하여도 정확한 추론이 가능한 One-shot learning 기반의 샴 네트워크를 제안하였다. 샴 네트워크는 동일한 파라미터를 공유하는 트윈 네트워크(Twin-network)를 포함하는 딥러닝 모델이다. 그림 1과 같이 트윈 네트워크 중 상위 네트워크는 입력으로 검증받고자 하는 이미지가 주어지고, 하위 네트워크는 입력으로 검증용 참조 이미지가 주어진다. 트윈 네트워크는 두 이미지로부터 특징 벡터(Feature vector)를 추출하고 특징 벡터간의 L_1거리를 계산하여 벡터 간의 유사도를 계산하는 Prediction layer로 전달한다. Prediction layer에서 특징 벡터 간의 유사도를 구하는 식은 식 1과 같다.

$$\sum_j \left(\alpha_j \left| h_1^{(j)} - h_2^{(j)} \right|\right) \quad (1)$$

h_1과 h_2는 상위 네트워크와 하위 네트워크에서 출력된 각각의 특징 벡터, α_j는 특징 벡터의 j번째 특징 간 L_1거리에 곱해지는 Prediction layer의 파라미터이다. Prediction layer는 특징 벡터 간의 유사도에 시그모이드 함수를 적용하여 0과 1사이의 확률을 출력한다. 그림 1과 같은 구조로 입력 이미지가 참조 이미지와 동일한 클래스이면 특징 벡터 간의 유사도가 커지도록, 다른 클래스면 유사도가 작아지도록 학습을 하여 새로운 데이터에 대해서 재학습 없이 높은 성능의 분류가 가능하다.

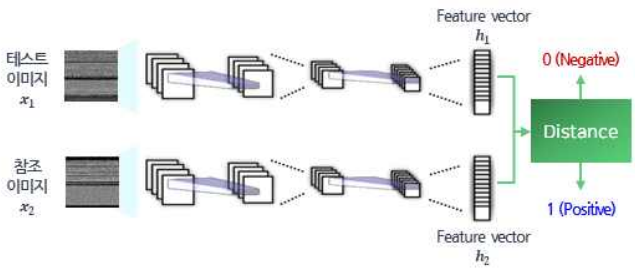


그림 1. 시스템 구조

2.2 시스템 구조

본 논문에서 제안하는 시스템의 구조는 그림 1과 같다. 데이터세트는 Microsoft Malware Dataset [4]를 사용하였다. 해당 데이터세트는 총 9가지의 악성코드 패밀리로 구성이 되어있다. 각 데이터는 Hex값으로 구성된 바이너리 덤프 파일을 가지고 있으며 이를 0은 검정색, 255는 흰색으로 변경하는 Gray scale 이미지화를 적용한다. 샴 네트워크의 트윈 네트워크는 MobileNet V2 [5]로 구성하였다.

2.3 실험 결과

본 장에서는 샴 네트워크의 성능을 기존 이미지 분류 딥러닝 모델 [6]과 비교하여 성능을 검증하였다. 동일한 데이터세트로 학습한 2개의 모델

의 정확도를 비교하였으며 각 클래스 별 Confusion Matrix로 성능을 비교하였다. 사용된 데이터세트의 문제점은 특정 클래스의 데이터가 다른 클래스에 비해 부족한 점이다. 해당 클래스는 Simda 패밀리 클래스이며 기존 이미지 분류 모델에서는 해당 클래스의 분류 성능이 낮게 나온다.

```
Class Name : [1], a number of Class : 1541,
Class Name : [2], a number of Class : 2148,
Class Name : [3], a number of Class : 2937,
Class Name : [4], a number of Class : 455,
Class Name : [5], a number of Class : 42,
Class Name : [6], a number of Class : 751,
Class Name : [7], a number of Class : 389,
Class Name : [8], a number of Class : 1222,
Class Name : [9], a number of Class : 1013,
```

그림 2 . 데이터세트의 각 클래스 분포

샴 네트워크는 One-shot learning이 가능하여 작은 클래스의 데이터에 대해서도 좋은 성능의 분류가 가능하다. 그림 3과 그림 4는 기존 이미지 분류 모델과 샴 네트워크의 Confusion matrix이다. 5번째 패밀리에 대해서 일반 분류 모델은 58%의 정확도를 보이는 반면 샴 네트워크는 90%를 보이는 점을 확인할 수 있다.

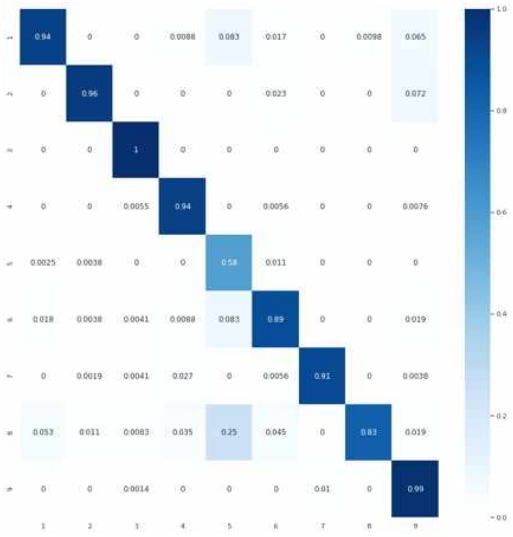


그림 3 . 기존 이미지 분류 모델의 Confusion matrix

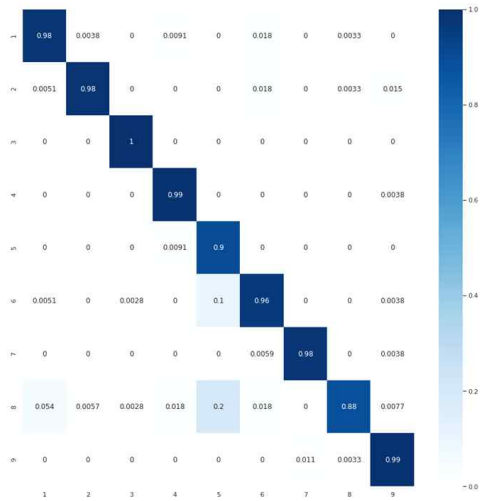


그림 4 . 샴 네트워크의 Confusion matrix

Ⅲ. 결론

본 논문에서는 샴 네트워크를 사용하여 소량의 악성코드 패밀리에 대해서도 높은 성능으로 분류가 가능한 기법을 제안하였다. 새로운 형태의 악성코드들이 많이 생기고 있으며 이에 소량의 데이터만 확보되어도 정확한 분류가 가능한 샴 네트워크를 악성코드 분류에 적용하여 높은 성능을 보이는 점을 검증하였다. 추후 악성코드의 이미지화 방법을 다양하게 하여 각 성능을 검증할 예정이다.

ACKNOWLEDGMENT

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2019-0 00097, 스마트공장 네트워크 인프라용 보안 칩 및 실시간 제어프로토콜 보안 기술 개발)

참 고 문 헌

[1] 조영복. "딥러닝 기반의 R-CNN 을 이용한 악성코드 탐지 기법." 한국디지털콘텐츠학회 논문지 19.6 (2018): 1177-1183.

[2] 석선희, and 김호원. "Convolutional Neural Network 기반의 악성코드 이미지화를 통한 패밀리 분류." 정보보호학회논문지 26.1 (2016): 197-208.

[3] Koch, Gregory, Richard Zemel, and Ruslan Salakhutdinov. "Siamese neural networks for one-shot image recognition." ICML deep learning workshop. Vol. 2. 2015.

[4] Ronen, Royi, et al. "Microsoft malware classification challenge." arXiv preprint arXiv:1802.10135 (2018).

[5] Sandler, Mark, et al. "Mobilenetv2: Inverted residuals and linear bottlenecks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

[6] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems 25 (2012): 1097-1105.

연합학습 기반 오토인코더 모델을 이용한 네트워크 이상탐지 시스템에 관한 연구

계효선, 권민혜
숭실대학교

ksh8301000@soongsil.ac.kr, minhae@ssu.ac.kr

Federated Learning Based Stacked Autoencoder for Anomaly Detection on Networks

Hyoseon Kye, Minhae Kwon
Soongsil University

요 약

네트워크 공격을 사전에 탐지하는 동시에 고성능 컴퓨팅의 비용 및 개인정보 보호의 문제를 해결하여 네트워크 상에서 안정성을 확보하는 것의 중요성이 높아지고 있다. 본 논문에서는 연합학습을 기반으로 오토인코더를 사용한 이상탐지 시스템을 제안한다. 제안하는 시스템의 성능을 평가한 결과, 본 제안 시스템이 기존의 중앙학습 방식과 유사하게 우수한 성능을 보이는 것을 확인하였다.

Ⅰ. 서 론

네트워크 보안 강화를 위해 보편적인 정상 트래픽의 패턴을 기준으로 악성 트래픽을 탐지하는 이상탐지에 대한 연구가 활발히 이루어지고 있다. 기존에는 중앙학습(centralized learning; CL) 방식으로 머신러닝 모델을 구축하고 있으나, 개인정보 보호와 서버 운영 비용에 대한 한계점을 극복하고자 연합학습이 제안되었다[1, 2]. 본 연구에서는 연합학습을 기반으로 stacked autoencoder를 활용한 네트워크 이상탐지 시스템을 제안한다.

Ⅱ. 연합학습 및 오토인코더 기반 이상탐지 연구

Ⅱ.1. 연합학습의 정의

연합학습은 중앙 서버의 글로벌 모델을 클라이언트의 로컬 학습을 통해 학습하는 방식이다. 클라이언트에서는 전송받은 글로벌 모델의 파라미터 정보를 클라이언트의 로컬 데이터로 학습하여 중앙 서버에 업데이트하고, 중앙 서버에서는 업데이트된 정보들을 취합하여 글로벌 모델을 갱신하는 방식으로 학습이 이루어진다. 이때, 서버와 클라이언트는 모델의 파라미터 정보를 communication round 마다 주고받으며 학습을 진행한다.

연합학습의 클라이언트의 수는 K로, 클라이언트의 로컬 데이터는 $X_k(1 \le k \le K)$로 정의한다. 로컬 데이터 X_k는 N 개의 데이터의 샘플 수와 M 개의 데이터의 차원(feature)의 수로 구성되어 $N \times M$ 차원 $(X_k \in \mathbb{R}^{N \times M})$으로 이루어진다. 데이터 샘플 $x_{k,n}$의 인덱스 집합은 $P_k=[x_{k,1}, x_{k,2}, \cdots, x_{k,N}]$로 정의하며 $N = |P_k|$이다. 중앙 서버는 매 communication round 마다 학습에 참여할 클라이언트를 uniform random으로 선택한다. 이때, 선택되는 클라이언트의 비율을 $C(0 < C \le 1)$로 정의하고, 매 communication round 마다 참여하는 클라이언트의 개수는 $K \cdot C$로 표현한다. 중앙 서버는 연합학습을 시작하기 위해 초기화된 글로벌 모델을 생성하고, 학습에 참여할 클라이언트를 선택하여 글로벌 모델의 파라미터를 전송한다. 글로벌 모델의 파라미터는 d차원을 가지는 $w(w \in \mathbb{R}^d)$로 정의한다. 선택된 클라이언트들은 각자 보유하고 있는 로컬 데이터 X_k를 이용하여 글로벌 모델의 파라미터 w를 기반으로 학습을 수행한다. 모델의 w는 입력값인 데이터 샘플 $x_{k,n}$과 예측값의 차이인 loss를 기반으로 역전파를 통해 업데이트 된다. 로컬 모델의 loss function은 $F_k(w)$로 정의되며, $x_{k,n}$의 loss function은 $f_{k,n}(w) = \ell(x_{k,n}; w)$로 정의한다. 로컬 학습에서 모델을 최적화하기 위해 $F_k(w)$를 최소화할 수 있는 w를 찾는 것을 목표로 하며 이는 (1)과 같이 표현할 수 있다.

$$\min_{w \in \mathbb{R}^d} F_k(w) \qquad where \qquad F_k(w) = \frac{1}{N}\sum_{n \in P_k} f_{k,n}(w) \quad (1)$$

t번째 communication round에서 클라이언트로부터 업데이트된 모델 파라미터 w_t^k를 중앙 서버로 전송한다. 중앙 서버는 업데이트 된 파라미터 w_t^k를 (2)와 같이 평균값으로 취합하고 글로벌 모델 파라미터 w_{t+1}을 업데이트하여 글로벌 모델을 갱신한다.

$$w_{t+1} = \frac{1}{K}\sum_{k=1}^{K} w_t^k \quad (2)$$

연합학습은 글로벌 모델이 수렴할 때까지 이 과정을 반복하면서 학습을 진행한다.

Ⅱ.2. Stacked Autoencoder기반 이상탐지

본 논문에서는 연합학습 기반의 중앙 서버 및 클라이언트 모델로 stacked autoencoder를 사용하여 이상탐지 시스템을 구성하였다. Stacked autoencoder는 여러 개의 은닉층(hidden layer)으로 구성된 심층 신경망(deep neural network)형태의 오토인코더이다[3]. 오토인코더는 비지도 학습 방식으로 데이터를 학습하고 재구성하기 위한 모델로, 인코더(encoder)와 디코더(decoder)로 구성된다. 인코더와 디코더의 층은 $L(1 \le i \le L)$로 동일하며 인코더는 $g = g_L \circ \cdots \circ g_1$, 디코더는 $f = f_1 \circ \cdots \circ f_L$로 정의한다. 오토인코더 A를 사용한 클라이언트의 데이터 샘플 $x_{k,n}$에 대한 출력값은 $A(x_{k,n}) = f \circ g(x_{k,n})$으로 정의할 수 있다. 인코더는 데이터 차원의 수를 입력 노드의 수로 설정하여 데이터를 입력층(input layer) g_1으로 입력받아 학습을 진행한다. 인코더 g는 입력 데이터의 차원을 축소시켜 의미 있는 정보만을 가지고 있는 잠재 공간(latent vector)을 형성한다. 디코더는 잠재 공간을 이용

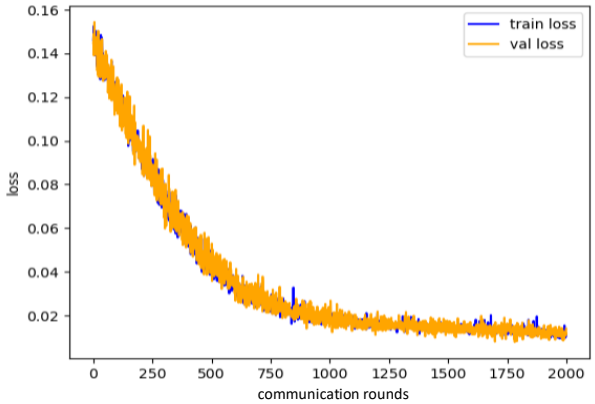


그림 1. Communication rounds 에 따른 loss 변화

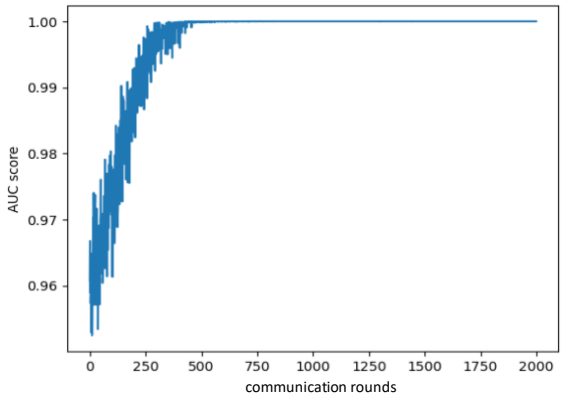


그림 2. Communication rounds 에 따른 AUC 변화

하여 축소한 데이터의 차원을 복원시킨 복원 데이터(reconstruction data) $\hat{x}_{k,n} = A(x_{k,n})$를 생성한다. 학습은 원형 데이터와 복원 데이터의 차이를 최소화하는 방향으로 이루어진다. 이를 위해 평균 제곱 오차(Mean Squared Error; MSE)를 loss function으로 사용하여 역전파를 진행하며, 이를 통해 수식(1)은 (3)과 같이 표현된다.

$$F_k(w) = \frac{1}{N}\sum_{n\in P_k} \ell(x_{k,n}; w)$$
$$= \frac{1}{N}\sum_{n=1}^{N}(\hat{x}_{k,n} - x_{k,n})^2 \quad (3)$$

오토인코더 기반의 이상탐지를 위해 정상 데이터를 오토인코더로 학습시킨다. 복원 데이터와 원형 데이터의 복원 오차(reconstruction error)는 ϵ로 정의한다. 정상 데이터의 복원 오차를 기반으로 설정한 특정 임계값(threshold)을 h로 표현한다. 임계값 h를 설정하기 위한 복원 오차는 (4)와 같이 평균 절대 오차(Mean Absolute Error; MAE)를 이용하여 계산한다.

$$\epsilon = \frac{1}{N}\sum_{n=1}^{N} |\hat{x}_{k,n} - x_{k,n}| \quad (4)$$

이상탐지를 수행하는 과정에서 입력 데이터의 복원 오차가 $\epsilon \geq h$이면 이를 비정상으로 간주하여 이상을 탐지한다.

Ⅲ. 제안하는 방식의 실험 설정 및 성능평가

Ⅲ.1. 성능평가 실험 설정

본 논문에서 제안하는 네트워크 이상탐지 시스템은 연합학습을 기반으로 stacked autoencoder 모델을 적용하였다. 실험에서 중앙 서버 및 모든 클라이언트들이 동일한 stacked autoencoder 구조를 가지는 것을 전제로 한다. 연합학습을 위해 500개의 고정된 데이터 샘플을 가진 100개의 클라이언트를 생성하였다. 실험의 학습 및 성능 검증을 위해 NSL-KDD 데이터셋을 사용하였다[4]. NSL-KDD train 데이터셋을 6:4의 비율로 train set과 validation set으로 분리하여 제안하는 시스템의 학습과 검증을 진행하였다. 모델의 평가를 위해 NSL-KDD test 데이터셋을 사용하였다. 제안하는 시스템의 정량적인 평가를 위해 정상 및 비정상 데이터를 탐지한 비율을 토대로 수치를 나타내는 성능지표인 F1-score와 MCC를 사용하였고, 기존 중앙학습 방식과 성능을 비교하였다.

Ⅲ.2 성능평가 결과

본 제안 시스템은 communication rounds에 따른 loss curve를 나타내는 그림1을 통해 overfitting 없이 loss가 수렴하면서 모델이 최적화되었음을 확인하였다. Loss가 수렴하기 시작하는 500번째 communication round 이후부터 AUC 또한 안정적으로 1에 가깝게 수렴하는 것을 그림 2를 통해 확인하였다. 표 1은 연합학습 FL과 중앙학습 CL의 성능을 나타낸 결과이다. CL은 F1-score와 MCC 모두 1에 가까운 높은 성능을 보였다. 이에 반해 FL은 100개의 클라이언트가 분산적으로 학습을 진행하였음에도 CL과 유사하게 높은 성능을 보이는 것을 확인하였다.

표 1. 연합학습(FL), 중앙학습(CL) NSL-KDD 성능평가

	F1-score	MCC
FL	**0.9917**	**0.9747**
CL	0.9999	0.9997

Ⅳ. 결론

본 논문은 연합학습을 기반으로 stacked autoencoder 모델을 적용한 네트워크 이상탐지 시스템을 제안한다. 제안한 시스템은 기존의 중앙학습 방식과 유사하게 우수한 성능을 보였다. 이를 통해 본 제안 시스템은 중앙학습 방식의 한계점인 고성능 컴퓨팅 파워의 비용 및 개인정보 보호 문제를 해결하면서 성공적으로 최적화된 글로벌 모델을 구축하여 이상탐지를 진행할 수 있음을 확인하였다. 본 연구 결과를 바탕으로 클라이언트에서 이상탐지 시스템을 직접 적용하여 서버 운영 비용 및 프라이버시 면에서 더 안정적인 네트워크 서비스를 제공할 수 있음을 확인하였다.

ACKNOWLEDGMENT

이 논문은 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT 연구센터지원사업(IITP- 2021-2020-0-01602)과 방송통신산업기술개발사업(IITP-2021-0-00739)의 지원을 받아 수행된 연구임.

참 고 문 헌

[1] 계효선, 권민혜, “모델 파라미터의 평균값을 이용한 연합지도학습 기반 네트워크 이상탐지 시스템,” 통신정보 합동학술대회, Apr. 2021.

[2] B. McMahan, E. Moore, D. Ramage, S. Hampson, B. A. y. Arcas, “Communication-Efficient Learning of Deep Networks from Decentralized Data,” Artificial intelligence and statistics. PMLR, Apr. 2017.

[3] A. M. Vartouni, S. S. Kashi, M. Teshnehlab, “An Anomaly Detection Method to Detect Web Attacks Using Stacked Auto-Encoder,” Iranian Joint Congress on Fuzzy and Intelligent Systems, pp. 131-134, Feb. 2018.

[4] https://www.unb.ca/cic/datasets/nsl.html

메타버스 플랫폼에서 인터랙션을 위한 단일 RGB 카메라 기반 경량화 자세 추정 모델에 관한 연구

남성현, 송하영, 김영원*

한국전자기술연구원, *한국전자기술연구원

sadguest@keti.re.kr, shy1230@keti.re.kr, *kimforever920@keti.re.kr

A Study on the Lightweight Pose Estimation Model Based on Single RGB camera for Interaction in Metaverse Platform

Nam Seong Hyeon, Song Ha Young, Kim Young Won*

Korea Electronics Technology Institute., *Korea Electronics Technology Institute.

요 약

본 논문에서는 MobileNet을 기반으로 광학식 모션캡처 장비, 깊이 카메라 등을 사용하지 않고 단일 RGB 카메라만 활용하여 실시간으로 사용자의 자세를 추정하는 경량화 모델을 구현하였다. Refinement 스테이지와 OneEuroFilter를 적용하여 자세 추정 오류를 정제하였으며 이를 통해 메타버스 플랫폼과 현실세계간 유연한 인터랙션 및 아바타 동기화를 가능하게 한다.

Ⅰ. 서 론

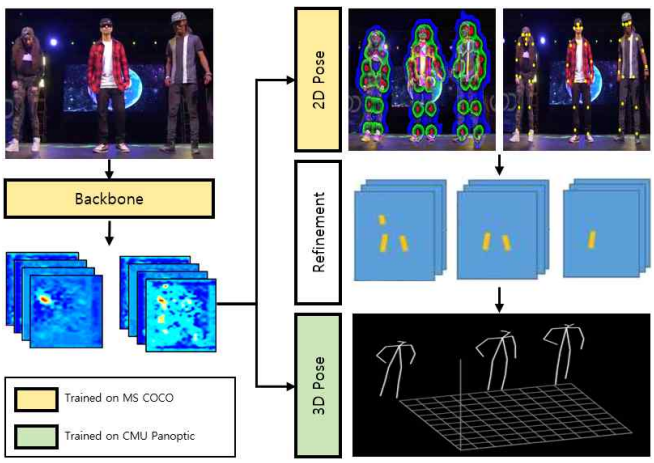
최근 코로나 팬데믹 이후 메타버스(Metaverse)에 대한 키워드가 떠오르고 있다. 메타버스는 XR, AI, Big data, 5G 네트워크, 블록체인 등 최신 기술의 복합체이다. 메타버스에서는 현실에서 나를 표현할 수 있는 아바타를 통해 3D 가상세계를 체험할 수 있으며 메타버스 내 활동은 현실과 분리되지 않고 현실의 연장선상에서 발생하는 활동을 포함한다. 현실 세계가 가상공간과 결합돼 현실 세계가 가상공간으로 확장된다는 의미다[1]. 몰입감 있는 메타버스를 즐기기 위해서는 아바타에 현실 세계의 움직임이 투영되어야 하며 이를 위해 사용자의 움직임을 추적하는 기술이 요구된다.

움직임 추적 기술은 OptiTrack과 같은 광학식 모션 캡처 장비를 사용하거나 Azure Kinect와 같은 ToF(Time-of-Flight) 방식의 깊이 카메라를 사용하는 경우가 일반적이다. 광학식 모션 캡처의 경우에는 측정 대상에 마커를 부착하여 측정 정확도를 높여 모션 측정을 가능하게 한다. 하지만 몸에 마커를 부착해야 하기 때문에 몰입도가 떨어질 수 있으며 초기 구축 비용 또한 높다. 깊이 카메라의 경우에는 depth 해상도에서 고화질을 지원하는 장비를 찾기 어렵고 취득한 데이터의 노이즈가 높아 정확한 자세 추적에 어려움을 가진다. 이러한 문제점을 해결하기 위해서는 부착식 센서, 깊이 카메라 등 별도의 장비를 사용하지 않고, 고해상도 RGB 카메라만으로 관절 위치를 추적하여 자세 및 움직임을 추정하는 기술이 필요하다. 컴퓨터 비전에서 사람의 신체 관절을 나타내는 키포인트가 어떻게 구성되어 있는지 위치를 측정하고 추정하는 문제인 Pose Estimation의 최신 연구 결과[2, 3]에 따르면 RGB 카메라만으로 기존의 방식보다 뛰어난 자세 및 움직임 추적에 대한 정확도를 보였다.

본 논문에서는 단일 RGB 카메라 기반 자세 추정 기술을 통하여 메타버스 플랫폼과 현실 세계 간 유연한 인터랙션 및 메타버스 아바타와의 움직임 동기화를 위한 딥러닝 기반 경량화 자세 추정 모델을 구현한다.

Ⅱ. 본론

딥러닝 기반 자세 추정 모델의 백본은 CNN 기반으로 설계된 MobileNet을 사용하였다. 기존의 다른 CNN 기반 모델과 비교하여 MobileNet은 Depthwise Separable Convolution 구조를 사용한 경량화 모델이다. 백본의 weight 파라미터들은 고정하고 2D, 3D 자세 추정을 위해 추가한 layer들을 학습시켰다.



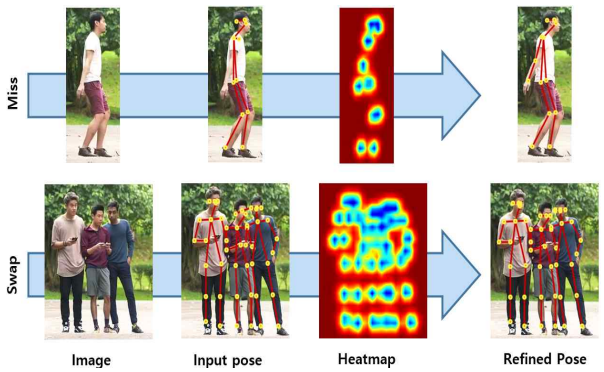
[그림 1] 경량화 자세 추정 모델 네트워크

[그림 1]은 논문에서 구현한 경량화 자세 추정 모델 네트워크를 보여준다. Top-Down 방식은 우수한 정확도를 갖추었지만, 다중 사용자 및 실시간 자세 추정에서는 실행 시간이 길어 적합하지 않다. 따라서, 이미지 내에서 별도의 Person Detector를 사용하지 않고 키포인트와 연결 정보를 추출할 수 있는 Bottom-up 방식을 사용하였다. 자세 추정 모델은 추출된 특징맵을 입력 값으로 하여 2D Pose 스테이지에서는 2D 키포인트 히트맵과 PAFs(Part Affinity Fields)를 추출하고 이를 통해 2D 자세의 초기 상

태를 추정한다. 추출되는 키포인트는 COCO 키포인트 포맷을 따라서 18개로 지정하였으며 각 키포인트는 목 키포인트를 루트로 하여 좌우 팔다리, 좌우 귀와 루트, 좌우 눈과 루트로 엣지를 구성하였다. 키포인트 추정 좌표와 ground-truth 좌표의 유사성 평균인 OKS(Object Keypoint Similarity)을 계산하기 위하여 다음 식을 사용했다.

$$OKS = \sum_i [\exp(\frac{-d_i^2}{2s^2k_i^2})\delta(v_i > 0)] / \sum_i [\delta(v_i > 0)]$$ d_i는 추정 좌표와 ground-truth 좌표 사이의 Euclidean distance를 나타내며, s는 인물(객체)의 크기, v_i는 groun-truth의 visibility flag이며 추정 좌표의 flag는 평가 과정에서는 사용되지 않는다. 신체에 있는 각 키포인트는 서로 다른 크기를 가지고 있어 falloff를 제어하기 위한 k_i 값을 가지고 있다.

Refinement 스테이지에서는 2D Pose 스테이지에서의 출력을 입력으로 받아 ground-truth와 비교하여 jitter, inversion, miss, swap 등의 오류를 정제한다. [그림 2]는 Refinement 스테이지를 3번 반복하여 miss와 swap 오류를 정제한 결과이다.



[그림 2] Refinement 스테이지에서의 오류 정제 과정

영상의 경우 프레임 단위로 자세를 추정하기 때문에 정적인 이미지에 비해서 jitter 오류의 발생 확률이 높다. 프레임 단위에서 이전 프레임의 추출된 키포인트와 현재 프레임에서 추출된 키포인트 간 OKS를 비교하여 유사성을 검사하고 가장 잘 매칭된 키포인트에 OneEuroFilter[4]를 적용하여 jitter 오류를 정제하였다.

3D Pose 스테이지에서는 특징맵과 키포인트 히트맵, 정제 과정을 거친 2D 자세를 기반으로 2차원 자세를 3차원 위치 정보로 변환하여 3D 자세로 보여준다.

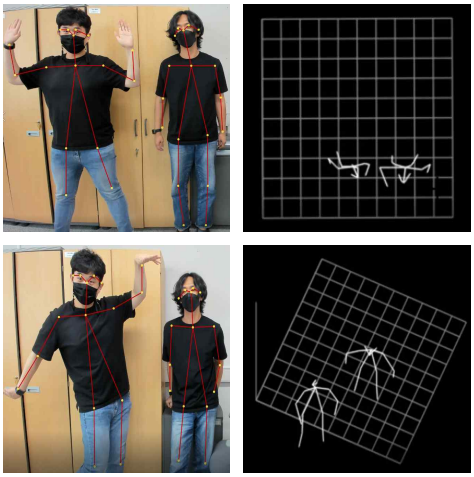
자세 추정을 위하여 2D Pose 스테이지는 MS COCO 데이터셋, 3D Pose는 CMU Panoptic 데이터셋을 학습시켰다. 학습은 CPU는 Intel i9-108050K, Nvidia GeForce RTX 3080 환경에서 진행되었으며 2D Pose에 대하여 COCO validation set에서 검증 결과는 <표 1>과 같다.

<표 1> COCO validation 결과

Method	AP	AP^{50}	AP^{75}	AP^{M}	AP^{L}	AR
MobileNetV1 (368x368)	36.1	61.3	35.5	30.5	44.9	43.1
MobileNetV1 (256x256)	38.0	64.0	38.0	32.5	46.9	44.7
MobileNetV1 (128x128)	39.0	65.3	40.1	32.9	48.4	45.5

검증 결과, 128x128에서 AP 39%를 달성하였고 3D Pose에서는 MPJPE(Mean Per Joint Position Error) 100mm의 결과를 보였다.

[그림 3]은 정제 과정을 모두 마친 2D, 3D 자세의 최종 추정 결과이며 2명의 사용자 대상으로 평균 37fps 환경에서 동작하였다.



[그림 3] 실시간 자세 추정 결과

Ⅲ. 결론

본 논문에서는 MobileNet을 백본으로 사용하여 고성능의 디바이스가 아닌 환경에서도 연산량 대비 정밀도 높은 자세를 추정하고 단일 RGB 카메라만으로 사용자의 2D, 3D 자세를 추정하는 모델을 구현 및 검증하였다.

이를 통해 별도의 장비 없이 RGB 카메라만으로 실시간으로 사용자의 자세 추정을 가능하게 하여, 추출되는 사용자 자세 3D 데이터를 통해 메타버스 플랫폼 내 아바타와 현실 세계 사용자 간 동기화 및 인터랙션을 가능하게 한다. 추후 연구에서는 Occlusion robust 알고리즘 및 3D 자세 추정 알고리즘 고도화 진행한다.

ACKNOWLEDGMENT

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2021년도 문화기술 연구개발 사업으로 수행되었음(과제명:메타버스 AI융합 기반 확장형 공원 플랫폼 기술개발 및 실증, 과제번호: R2021040269, 기여율: 100%)

참 고 문 헌

[1] 고선영, 정한균, 김종인, 신용태. "메타버스의 개념과 발전 방향," Korea Information Processing Society Review, pp. 7-16, 2021.

[2] Mehta, D. et al. "Single-Shot Multi-person 3D Pose Estimation from Monocular RGB," '18 International Conference on 3D Vision (3DV), pp. 120-130, Oct. 2018.

[3] Mehta, D. et al. "XNect: Real-time Multi-Person 3D Motion Capture with a Single RGB Camera," ACM Transactions on Graphics, pp. 1-17, Jul. 2020.

[4] Casiez, G., Roussel, N., & Vogel, D. "1€ filter: A simple Speed-based Low-Pass Filter for Noisy Input in Interactive Systems," In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 2527-2530, May. 2012.

로블록스 플랫폼을 활용한 메타버스·AI 융합기반 인터랙티브 콘텐츠 제안

송하영, 남성현, 김영원*
한국전자기술연구원, *한국전자기술연구원
shy1230@keti.re.kr, sadguest@keti.re.kr, *kimforever920@keti.re.kr

Proposal of interactive content based on Metaverse and AI convergence using the Roblox platform

Song Ha Young, Nam Seong Hyeon, Kim Young Won*
Korea Electronics Technology Institute, *Korea Electronics Technology Institute

요 약

최근 비대면 서비스에 대한 수요 증가와 가상 생태계 구성을 위한 ICT 기술산업의 성장으로 차세대 가상공간 플랫폼인 메타버스가 부상하고 있다. 메타버스는 현실세계와 같은 사회·경제·문화 활동이 이뤄지는 3차원의 가상세계를 말하며, 아바타를 활용해 단순히 게임이나 가상현실을 체험하는 것이 아닌 사용자들 간에 사회적 활동이 가능하다. 본 논문에서는 부산의 용두산 공원을 세계적으로 점유율이 가장 높은 메타버스 플랫폼인 로블록스와 XR, AI, 5G 등의 기술을 바탕으로 메타버스·AI 융합기반 인터랙티브 콘텐츠 개발 방안을 제안하고자 한다.

Ⅰ. 서 론

최근 온라인상의 만남이 보편화되는 비대면 사회가 지속되면서 가상의 세계에서 다른 사람들과의 상호작용이 가능한 메타버스(Metaverse)가 주목받고 있다. 메타버스는 범용기술(General Purpose Technology)의 복합체로, XR(eXtended Reality)기술을 비롯한 Data, Network, AI(Artificial Intelligence)와 같은 다양한 범용기술이 복합 적용되어 구현되어 현실과 가상의 경계가 사라지고 있다. 메타버스 플랫폼은 현실의 세계를 가상세계에 투영하여 구현할 수 있다는 점을 바탕으로 국내외로 문화·예술, 교육, 홍보, 엔터테인먼트 등 다양한 분야에서의 적용으로 일상생활에서 쉽게 접할 수 있게 되어 활용도가 높아지고 있다.[1]~[3] 대표적인 메타버스 플랫폼으로는 로블록스(Roblox), 포트나이트(Fortnite), 마인크래프트(Minecraft), 제페토(Zepeto)가 있으며 국내외로 가장 점유율이 높은 플랫폼은 로블록스이다. 로블록스는 사용자들에게 자신의 아바타와 자신만의 가상 공간을 창조하는 서비스를 제공하고, 게임내에서 사용되는 가상화폐인 로벅스(Robux)를 이용하여 다른 사람들과의 사회적 활동이 가능하다는 점과 로블록스 스튜디오(Roblox Studio)를 통해 직접 게임을 제작하고 배포 할 수 있다는 특징이 있다.[4]

본 논문에서는 부산의 근현대사 제 1호 도심공원인 용두산 공원을 로블록스 플랫폼으로 구현하여 인공지능 기술과 실감형 기술을 기반으로 확장 가능한 인터랙티브 콘텐츠 개발 방안에 대해 제안하고자 한다.

Ⅱ. 메타버스·AI융합 기반 확장형 인터랙티브 콘텐츠 구현

메타버스의 핵심은 현실세계가 가상공간과 결합되어 현실세계가 가상공간으로 확장되는 것에 있다. 가상세계에서 다중 사용자들이 콘텐츠를 공유하고 함께 상호교류하며 원활한 사회·문화·경제적 활동이 가능해야 한다. 본 논문에서는 다수의 사용자들이 실시간으로 콘텐츠에 접속하고 서로 상호작용 할 수 있는 메타버스 플랫폼인 로블록스를 활용하여 메타버스·AI 융합기반 확장형 인터랙티브 콘텐츠 개발 방안을 제시한다. 로블록스는 자체적으로 제공하는 로블록스 스튜디오를 이용하여 사용자가 가상세계에서 서로 간의 의사소통 및 상호작용이 가능하도록 콘텐츠를 직접 제작할 수 있다. 최근 로블록스 이용자는 월간활성이용자수(MAU)는 1억5천만명, 일평균 접속자수는 4천만명에 달한다. 국내외로 점유율이 가장 높은 로블록스는 PC, Mobile, VR기기 등 다양한 접근방법을 제공하여 사용자가 원하는 방식으로 콘텐츠 체험이 가능하여 사용자의 접근성을 높여 전 세계 로블록스 사용자들에게 AI 융합형 메타버스 콘텐츠를 알리는데 가장 적절한 플랫폼이라 할 수 있다.

[그림 1]은 메타버스·AI융합 기반 인터랙티브 콘텐츠의 전체적인 구상도를 보여준다. 부산의 용두산 공원의 주요 상징물을 <표 1>과 같이 로블록스 플랫폼 내에서 가상세계를 현실과 유사하게 제작함으로써 사용자들의 몰입도를 높일 수 있으며, 용두산 공원의 주요 상징물인 꽃시계, 종각, 충무공 이순신 동상, 용탑, 부산 타워, 팔각정 등의 요소들을 소재로 사용자의 이동 동선을 고려한 실감형 콘텐츠를 구현하여 메타버스 속 가상세계에 있는 사람들이 콘텐츠를 체험하고, 경험을 공유하면서 사용자들이 상호작용 할 수 있다. 또한 AI 융합 기술을 활용하여 VR 콘텐츠 내에서 가상의 3D 아바타와 현실적이고 효과적인 인터랙션이 가능하도록 콘텐츠를 구성하였다. 컴퓨터 비전에서 사람의 신체 관절 위치를 측정하고 추정하는 Pose Estimation 기술을 바탕으로 딥러닝 기반 경량화 자세 추정 모델을 활용하여 단일 RGB 카메라만으로 메타버스 플랫폼과 현실 세계 간 유연한 인터랙션 및 메타버스 아바타와의 움직임 동기화가 가능하도록 하였다.

[그림2]는 인터랙티브 콘텐츠 구상도 중 가상과 현실이 연결되는 윈도우 포탈 콘텐츠를 나타낸다. 5G 기술을 기반한 실세계와 가상세계 간 연결형 VR 윈도우 포탈 콘텐츠로 VR 콘텐츠 내에서 용두산의 실시간 상황을 뷰잉할 수 있다. 용두산 공원의 대형 미디어월에서는 가상의 용두산과 현실의 용두산이 공존하는 장관이 펼쳐지도록 개발하고, 대형 미디어월을 통해 실시간 용두산 VR을 볼 수 있으며 접속한 유저들과 실제로 만날 수 있다. 또한 VR내 윈도우 포탈 개발을 통해 곳곳의 스크

린을 통해 같은 시간에 용두산 공원에 실재하는 사람을 만나는 것도 가능 할 것으로 보인다.

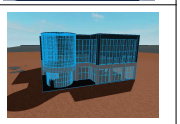
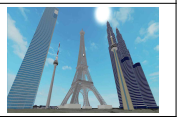
메타버스로 구현된 용두산 공원을 체험하는 VR 콘텐츠 구현 예시를 나타낸 [그림 3]에서는 3D 가상 객체로 제작된 환경에 자신의 아바타가 투영된 모습을 볼 수 있다. 사용자의 3D 아바타가 용두산 공원 곳곳을 누비며 종각, 이순신 장군 동상 등 주요 장소들을 모티브로 한 콘텐츠 체험을 할 수 있으며, 딥러닝 기반 동작분석 기술을 바탕으로 사용자의 움직임과 위치를 추적함으로써 실감형 인터랙티브 콘텐츠 구현이 가능하다. 사용자들은 로블록스로 구현된 메타버스·AI 융합 기반 확장형 인터랙티브 콘텐츠 내에서 사용자들끼리 자유롭게 소통하고, 로벅스와 같은 가상화폐를 활용하여 경제적 활동 등의 사회적 활동을 경험할 수 있다.

이처럼 메타버스 플랫폼을 활용하면 현실세계와 가상세계를 연결하는 콘텐츠를 통해 온라인상에서도 실제와 유사한 사회적 경험이 가능하며 시공간적 제약없이 사용자에게 다양한 상호작용을 통해 새로운 경험을 제공하는 데 유용하게 활용될 것이다.



[그림 1] 메타버스·AI융합 기반 인터랙티브 콘텐츠 구상도

<표 1> 로블록스 플랫폼 내 용두산 공원 주요상징물 제작 예시

No	제작 개체	제작 예시	No	제작 개체	제작 예시
1	꽃시계, 종각		4	중앙광장	
2	충무공 이순신 동상		5	전시관	
3	용탑		6	부산타워, 팔각정	



[그림 2] 실세계-가상세계간 연결형 VR 윈도우 포탈 콘텐츠



[그림 3] 로블록스를 활용한 인터랙티브 콘텐츠 구현 예시

Ⅲ. 결론

최근 대면 서비스에 대한 수요 증가와 가상 생태계 구성을 위한 ICT 기술산업의 성장으로 메타버스를 활용한 콘텐츠들이 교육, 홍보, 엔터테인먼트 등 다양한 분야에서 다채롭게 적용되어 개발되고 있다.

본 논문에서는 메타버스·AI 융합 기반 확장형 인터랙티브 콘텐츠 개발 방안을 제시하였다. 전 세계적으로 활용도와 점유율이 가장 높은 로블록스 플랫폼을 활용하여 사용자들의 접근성을 높이고, 용두산 공원을 구현하여 공원 내 주요 장소들에 관한 콘텐츠를 사용자가 인터랙션 할 수 있는 가능성도 제시하였다. 앞으로 로블록스와 같은 다양한 메타버스 플랫폼을 통해 시공간적 제약 없이 주요 관광지들을 여행할 수 있는 콘텐츠들이 다양하게 개발될 것으로 예상되며 국내 관광지의 경우 현실과 가상세계를 연결하는 메타버스 플랫폼을 활용함으로써 국내외 사용자들의 접근성을 높일 수 있을 것이다.

ACKNOWLEDGMENT

본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2021년도 문화기술 연구개발 사업으로 수행되었음(과제명: 메타버스 AI융합 기반 확장형 공원 플랫폼 기술개발 및 실증, 과제번호: R2021040269, 기여율: 100%)

참 고 문 헌

[1] Nevelsteen, Kim JL. "Virtual world, defined from a technological perspective and applied to video games, mixed reality, and the Metaverse." Computer Animation and Virtual Worlds 29.1 : e1752, 2018

[2] Dionisio, John David N., William G. Burns III, and Richard Gilbert. "3D virtual worlds and the metaverse: Current status and future possibilities." ACM Computing Surveys (CSUR) 45.3: 1-38, 2013

[3] Sun, Yuan, and Masoud Gheisari. "Potentials of Virtual Social Spaces for Construction Education." EPiC Series in Built Environment 2 : 469-477, 2021

[4] Lee, Byong-Kwon. "The Metaverse World and Our Future." Review of Korea Contents Association 19.1 : 13-17, 2021

Group Attention U-Net을 이용한 단일 영상 연무제거

홍찬의[1], 박진효[2], 최현덕*

전남대학교

217643@jnu.ac.kr, 206070@jnu.ac.kr, *ducky.choi@jnu.ac.kr

Single Image Dehazing based on Group Attention U-Net

Hong Chan Eui[1], Park Jin Hyo[2], Choi Hyun Duck*

Cheonnam National University

요 약

본 논문에서는 CA(Channel Attention), SA(Spartial Attention), ResNet[3]로 구성된 BottleNeck을 U-Net[1]구조로 접목시켜서 활용한 영상 연무 제거 모델을 제안한다. 영상에서 연무는 균형적으로 형성되지 않고, 특정 부분에 더 많이 일어나고 있음을 알 수 있다. 기존의 연무제거 모델로 주로 사용하는 모델구조인 ResNet[3]은 CNN(Convolution Neural Network)과 skip connection 방식을 사용하여 진행하였다. 이 방식에서는 지역적인 부분에만 작동된다는 점에서 큰 약점이 있다. CA과 SA을 추가하여 글로벌한 정보에서도 손실이 없도록 개선하였다. 제안하는 방법은 이전에 연구된 image dehazing 방법에 비해 PSNR(Peak Signal-to-Noise Ratio)과 SSIM(Structural Similarity Index)에서 우수한 결과를 나타냈다.

Ⅰ. 서 론

최근 인공지능 기술들이 발전하면서 영상을 통해 정보를 추출하고 이를 활용하는 컴퓨터 비전 연구들이 활발하게 진행되고 있다. 대기 상황이 안 좋은 경우 여러 입자들의 영향을 받아 색상 왜곡, 낮은 콘트라스트 등으로 영상이 손상되어 컴퓨터 비전을 토대로 진행하는 방법의 성능을 저하시킨다. 특히 객체 감지가 필요한 자율주행 자동차, 방범용 카메라 등에는 인명피해가 발생하는 치명적인 문제로 작용할 수 있다. 이런 문제를 해결하기 위해 연무로 인해 손상된 영상에서 연무가 제거된 영상을 복구하는 영상 연무 제거 기술을 목표로 한다.

Ⅱ. 본론

단일 영상 연무 제거 기술은 가시성이 낮아진 영상으로부터 연무가 제거된 영상을 복원하는 것을 목표로 한다. 대기 산란 모델은 연무의 간단한 근사치를 제공하며 다음과 같이 공식화된다.

$$I(z) = J(z)t(z) + A(1 - t(z)) \quad (1)$$

I(z)는 관측된 연무 영상, A는 지구 산란광, t(z)는 대기투과율, J(x)는 연무가 제거된 영상이다. 대기 산란 모델은 영상의 dehazing이 A,t(z)에 대한 지식 없이 결정되지 않는 문제임을 보여준다. 다음과 같이 공식화된다.

$$J(z) = \frac{(I(z) - A)}{t(z)} + A \quad (2)$$

공식 1과 2를 통해 연무가 낀 영상에 대해 지구 산란광과 대기 투과율을 적절하게 추정하면 연무가 제거된 영상을 복원할 수 있음을 알 수 있다.

초기 대기 산란 모델은 시각 정보를 추출하여 단일 영상의 연무를 제거하는 연구가 많이 진행되고 있다. DCP(Dark Channel Prior)[5]는 성능이 뛰어난 사전 지식 기반 방법들 중 하나이며 연무가 없는 영상은 R,G,B 채널 중에서 적어도 하나의 채널은 낮은 명도 값을 갖는다라는 가정을 기반으로 제안한다. 그러나 사전 지식 방법은 특정 사례에서는 잘 작동하지 않는다. 최근 딥러닝 기법이 떠오르면서 AOD-Net[6], DehazeNet[7]과 같이 신경망을 사용하는 연무제거 기술이 많이 발달하고 있다. 기존 방법과 비교하여 딥러닝 방법은 end-to-end 방식을 사용하여 처음부터 끝까지 기계가 알고리즘을 찾는 방법으로 사람이 개입함으로써 생기는 오류를 줄인다. 이를 빅데이터에 적용하면 우수한 품질의 영상을 얻을 수 있다.

본 논문은 CA, SA, ResNet[3]을 bottleneck구조로 구성한 GAB(Group Attention Block)을 기반으로 U-Net[1] 구조를 활용한 새로운 end-to-end 방식의 신경망을 제안한다. 그림 1은 제안하는 신경망 구조이다.

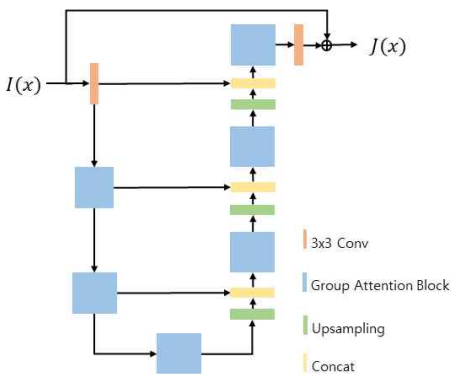


그림 1. 제안하는 GAB 기반 연무제거 모델

U-Net[1]은 영상의 크기를 줄이면서 이미지 내의 중요한 정보를 직접 전달(skip-connection)하여 디코더에서 선명한 이미지를 얻게 해 정확하게 연무 제거를 하는 데에 효과적인 구조를 가진다. 신경망의 구조에서 GAB 구조는 아래의 그림 2처럼 구성되어 있다.

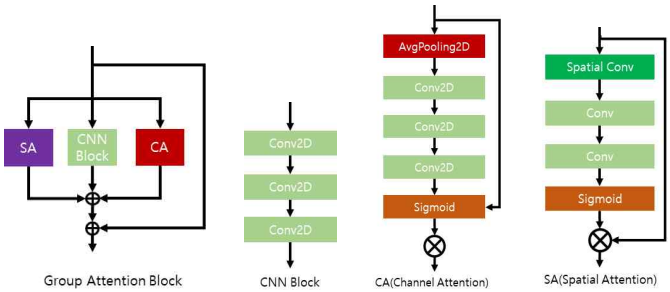


그림 2. GAB 구조

GAB은 CA, SA, CNN을 bottleneck 구조로 구성하고 skip connection을 통해 residual 구조를 사용한다.

CNN Block은 CNN을 통해 image feature을 생성한다. CA는 averagepooling을 이용해 $C\times W\times H$에서 $C\times 1\times 1$로 만들고 CNN을 통해 attention해서 channel feature를 추출한다. SA는 CNN을 이용해 $C\times W\times H$에서 $1\times W\times H$로 만들어 CNN를 통해 attention해서 spatial feature를 추출한다. CA, SA, CNN을 bottleneck 구조로 이용함으로써 채널단위, 픽셀단위, 영상단위로 구성된 다양한 관점으로 영상에서 연무의 특징을 분석하여 효과적으로 복원한다.

GAB의 bottleneck 구조는 ResNext[4]에서 사용되는 Group Convolution과 유사한 구조로 형성이 되어 있으며, Group Convolution에서는 CNN을 3개 이상 구성해야 좋은 성능이 나타난다. 기존에 CA, SA은 2개의 CNN으로 구성했지만, 각각 CNN이 3개씩 구성하여 보다 좋은 성능을 이끌어냈다.

제안하는 방법을 학습할 때 학습용 데이터셋으로 RESIDE (Realistic Single Image Dehazing)이라는 합성 및 실제 흐릿한 이미지로 구성된 대규모 벤치마크를 사용하였다. ITS (Indoor Training Set) 110,000여개, OTS (Outdoor Training Set) 313,959여개 등등으로 구성되어 있다. train data와 ground truth간의 MSE(Mean Square Error)를 비용함수(Loss Function)로 설정하고, 초기 learning rate를 0.0001로 기준으로 10epoch마다 0.2배씩 늘려가면서 image dehazing 실험을 진행하였다. 실험 결과 PNSR 지표는 30.96, SSIM 지표는 0.954로 보여주고 있다. 표 1에서 제안하는 방법이 기존의 연무제거 모델보다 연무제거에 효과적이며, 우수한 성능이 나타나는 것을 확인할 수 있다.

	SSIM	PSNR
DCP[5]	16.62	0.82
AOD-Net[6]	19.06	0.85
DehazeNET[7]	21.14	0.86
GFN[8]	22.30	0.880
Ours	31.07	0.954

표 1. 지표분석

Ⅲ. 결론

본 논문에서는 단일 영상의 연무를 효과적으로 제거하기 위해 CA, SA, ResNet[3]을 bottleneck방식으로 구성된 block으로 기반으로 U-Net[1]구조를 활용한 새로운 end-to-end 방식의 신경망을 제안했다. 제안하는 네트워크는 구조는 다양한 attention을 통해 다각도로 영상의 특징을 분석하여, 영상의 세부사항과 색상 복원에 강력한 이점을 갖고 있다. 제안하는 새로운 신경망은 RESIDE Dataset를 토대로 진행된 image dehazing 실험을 통해 효과적으로 검증되었으며, super-resolution, denosing, deraining과 같은 작업에도 탁월할 것으로 예상된다.

ACKNOWLEDGMENT

본 논문(저서)은 2020년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(No. 2020R1G1A1101070)을 받아 수행된 연구임.

참 고 문 헌

[1] Ronneberger O., Fishcer P. and Brox T., “U-Net[1]: Convolutional Networks for Biomedical Image Segmentation,” Medicial Image Computing and Computer-Assisted Intervention-MICCAI 2015, pp 234-241, Nov. 2015.

[2] Park J., Woo S., Lee J.Y. and Kweon I.S., “BAM: Bottleneck Attention Module,” BMVC, Oct. 2018.

[3] He K., Zhang X., Ren S., and Sun J., “Deep Residual Learning for Image Recognition,” 2016 IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778, June. 2016.

[4] Xie S., Girshick R., Dollar P., Tu Z., and He K., “Aggregated Residual Transformations for Deep Neural Networks,” 2017 IEEE Conference on Computer Vision and Pattern Recognition, pp. 5987-5995, July. 2017.

[5] He K., Sun J. and Tang X., “Single Image Haze Removal Using Dark Channel Prior,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 12, pp. 2341-2353, Dec. 2011.

[6] Li B., Peng X., Wang Z., Xu J. and Feng D., "AOD-Net: All-in-One Dehazing Network," 2017 IEEE International Conference on Computer Vision, pp. 4780-4788, Dec. 2017.

[7] Cai B., Xu X., Jia K., Qing C. and Tao D., “DehazeNet: An End-to-End System for Single Image Haze Removal,” IEEE Transactions on Image Processing, vol. 25, no. 11, pp. 5187-5198, Nov. 2016.

[8] Ren W., Ma L., Zhang J., Pan J., Cao X., Liu W. and Yang M., “Gated Fusion Network for Single Image Dehazing," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3253-3261, June. 2018.

[9] Li, B., Ren, W., Fu, D., Tao, D., Feng, D., Zeng, W., and Wang, Z., “Benchmarking single-image dehazing and beyond,” IEEE Transactions on Image Processing, vol. 28, no. 1, pp.492 - 505, Jan. 2018.

딥 러닝 기반 지게차 충돌 방지 시스템

박재한, 신수용
금오공과대학교
qkrwogks7094@kumoh.ac.kr, wdragon@kumoh.ac.kr

Deeplearning-basedforkliftcollisionpreventionsystem

Jae Han Park, Soo Young Shin
Kumoh National Institute of Technology

요 약

본 시스템은 LED와 카메라를 결합하여 객체를 탐지한다. 딥러닝을 사용한 객체 탐지는 사람과 지게차를 탐지하고 실시간으로 객체들의 거리를 측정하여 충돌 위험이 있을 때 경고한다. 또한 공장의 조명과 함께 연동되어 충돌위험 거리에 따라서 다른 색상의 조명으로 근로자들에게 알려준다. 제안되는 시스템은 광각 카메라를 사용하여 적은 수의 SMART LED로 넓은 구역의 충돌 사고를 방지한다.

Ⅰ. 서 론

최근 공장과 같이 많은 물류가 이동하는 장소에서는 지게차가 많이 사용되고 있다. 하지만 지게차의 경우 많은 양의 물류를 적재할 시 운전자가 전면의 시야를 확보하기 어렵고 이 때문에 사람과 충돌하는 사고가 증가하고 있다.[1] 또한, 기존에는 이런 사고를 방지하기 위해서 신호수가 지게차를 따라다니며 지게차의 사각을 보조했다. 하지만 이런 방법은 현재 잘 지켜지고 있지 않으며 좁은 공간에서 신호수와 지게차가 모두 공간을 차지하기 힘든 경우도 존재한다. 본 논문에서는 공장의 조명시스템과 객체 충돌 방지 시스템을 결합하여 이런 문제를 방지하고자 한다. 조명과 함께 카메라를 설치하면 항상 일정한 밝기를 보장받을 수 있고 그로 인하여 객체를 조금 더 수월하게 탐지할 수 있다. 또한 광각 카메라를 사용하여 적은 수의 SMART LED 모듈로도 큰 공간을 탐지할 수 있다. 충돌 방지 시스템에는 딥 러닝을 사용한 객체 인식 시스템을 채택하였으며 다양한 각도와 상황에서 지게차와 사람을 탐지하고 객체 간의 거리를 인식한다.

Ⅱ. 본론

본 시스템은 크게 2가지 부분으로 구성된다. 첫 번째는 객체 탐지이며 이 부분에서 우리는 지게차와 사람을 탐지한다. 두 번째는 사람과 지게차의 충돌을 감지하는 충돌위험 감지 알고리즘 부분이다.

1. 객체 탐지

객체 탐지는 Yolo v3-tiny를 사용하여 진행하였다. 클래스의 경우 3가지로 분류되었다. 클래스는 Forklift, Handforklift, Person 이며, 지게차와 사람이 같이 존재하는 사진 약 450장과 사람만 존재하는 사진 650장이 사용되었다. 하지만, 확보된 데이터세트의 숫자가 적어 학습에 무리가 있다고 판단하여 image augmentation 작업을 진행하여 주었다. 작업 이후 총 이미지의 숫자는7000장이 되었으며 그림 2와 같은 회전, 반전을 이용한 augmentation이 진행되었다.

라벨링 작업에는 그림 3의 LabelImg 툴을 사용하였으며[2], 트레이닝의 경우 이후 젯슨 nano에서 사용할 것을 고려하여 Yolov3-tiny를 사용하였다.[3] 학습 결과는 그림4과 같으며, 추후 데이터세트 추가를 통하여 다른 클래스 학습도 생각해 볼 수 있다.

2. 객체 충돌 감지 알고리즘

그림 4는 훈련된 결과물로 객체를 탐지하는 사진이다. 우리는 객체를 인식하고 지게차와 사람의 거리가 가까울 때 충돌 경고를 하도록 설정하였다. 현재는 객체 간의 위험 거리는 15픽셀로 설정되어있으며, 객체의 바운딩박스 가장자리끼리의 거리가 15픽셀 이하가 되면 위험하다고 판단하여 경고하게 된다. 또한 연동된 LED의 색상을 적색으로 바꾸어 조명 아래 작업 중인 작업자들과 지게차의 운전자가 현재 충돌위험이 있음을 인지할 수 있도록 한다. 추후 거리별 혹은 충돌 대상의 위치에 따라서 LED의 색상을 다르게 표기하여 운전자가 어디에 충돌위험 대상이 있는지 인지할 수 있도록 하고자 한다.

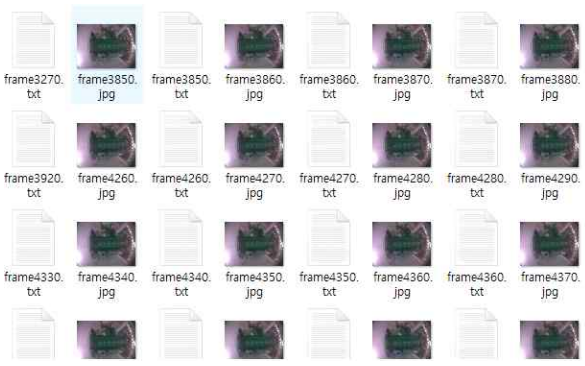


그림 1 라벨링을 위해 준비된 사진과 주석 텍스트 파일

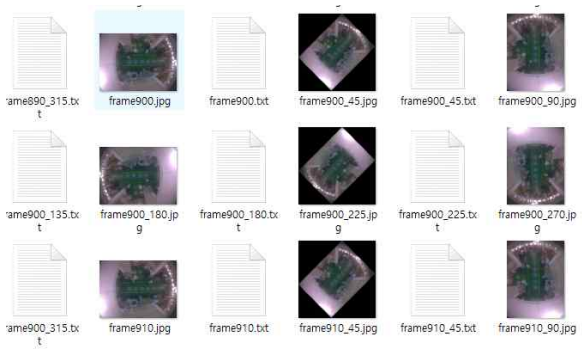


그림 2 augmentation을 사용한 결과물

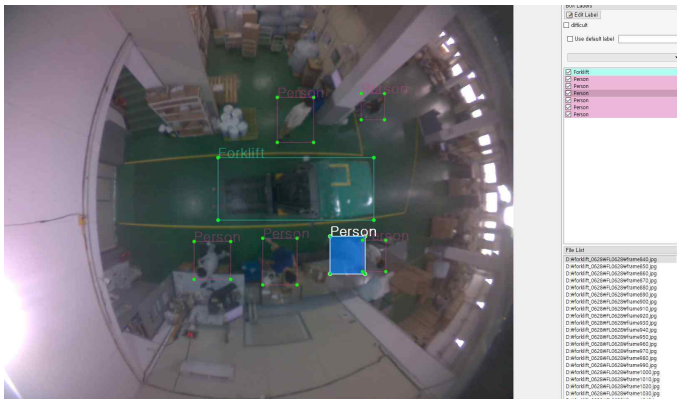


그림 3 LabelImg를 사용한 라벨링 작업

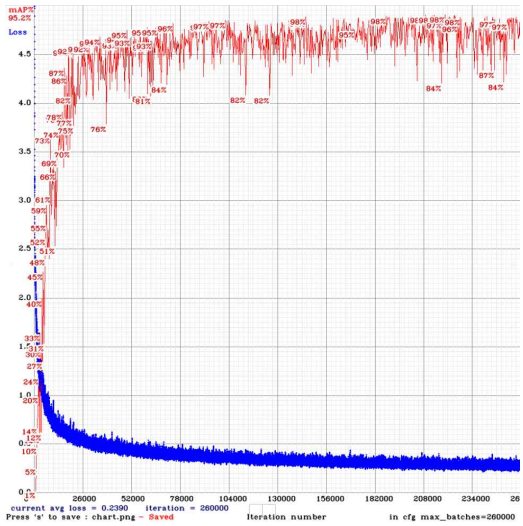


그림 4 학습 결과 그래프

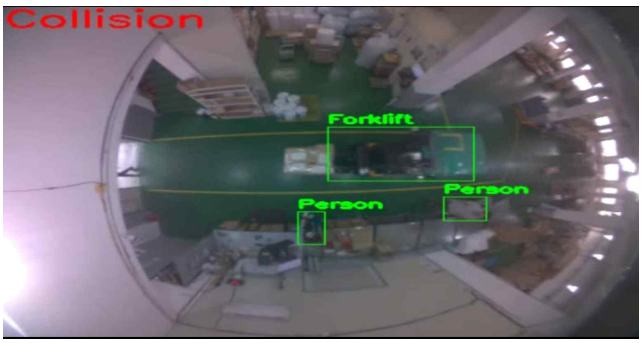


그림 5 객체 인식 기반의 충돌 알고리즘

Ⅲ. 결론

본 시스템은 딥 러닝을 사용하여 사람과 지게차를 검출하고, 검출된 객체들을 이용하여 충돌을 방지한다. 하지만, 광각렌즈의 경우 왜곡이 심하여 객체 간의 거리가 정확하게 측정되기 힘든 문제가 있다. 이후 카메라 왜곡 보정 코드를 통하여 객체 간의 정확한 거리를 측정하고자 한다. 현재는 지게차와 사람만을 탐지하여 충돌 경고만 알려주지만, 이후 추가적인 클래스와 데이터세트를 사용하여 낙상을 감지하고, 감지된 사고를 관제실에 알리는 방향으로 연구를 진행하고자 한다. 또한, 온도 센서 등을 장착하여 공장의 위험구역을 설정하고 위험구역에 사람이 출입하면 관제실로 경고를 보내는 시스템도 개발하고 있다.

ACKNOWLEDGMENT

"이 논문은 2021년도 정부(교육과학기술부)의 재원으로 한국연구재단의 대학중점연구소 지원사업으로 수행된 연구임 (2018R1A6A1A03024003)

참 고 문 헌

[1] http://news.tvchosun.com/site/data/html_dir/2021/06/05/2021060590040.htmlDavies

[2] https://github.com/tzutalin/labelImg

[3] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," 2018.

다중 에이전트 강화학습을 이용한 UAV-자원 할당 최적화에 대한 기술 동향

조현희, 김중헌
성균관대학교, 고려대학교
gt7746d@g.skku.edu, joongheon@korea.ac.kr

Survey on Reinforcement Learning Technologies for Resource Allocation in Unmanned Aerial Vehicles (UAVs)

Hyunhee Cho and Joongheon Kim
Sungkyunkwan University, Korea University

요 약

예상하기 어려운 다양한 환경(Environment) 변화나 UAV(Unmanned Aerial Vehicle) 개체의 에너지 또는 컴퓨팅 자원의 한계를 극복하려는 목적으로 강화학습(Reinforcement Learning, RL) 알고리즘을 UAV 시스템에 도입한 연구는 활발하게 진행되고 있다. UAV 군집 주행과 같이 다수의 개체가 존재하는 환경에서 다중 에이전트 강화학습을 기반으로 영상 처리 및 정보 보안을 위해 데이터 할당을 최적화하는 목적의 UAV 시스템에 도입되는 연구를 살펴보았다. 본 논문은 UAV 기반 시스템에 도입한 다중 에이전트 강화학습을 각 에이전트가 할당받는 데이터의 양질 및 서비스 최적화 목적과 데이터 손실과 같은 비용(Cost) 최소화 목적으로 분류하여 최신 연구 동향을 분석하였다.

Ⅰ. 서 론

UAV는 높은 기동성과 넓은 적용 범위, 사용자와 Line-of-Sight (LOS) 링크를 맺을 수 있는 장치로서 주목받고 있다[1]. 현재 크기, 비행 반경 및 고도, 이용 목적에 따라 다양한 종류의 UAV가 [2]와 같이 분류되고 있어 실제 연구 및 산업 분야에 활용되고 있다. UAV는 정찰이나 전투와 같은 군사적 용도와 통신 중계, 무인 배송, 조난 구조 시스템과 같이 민간용으로도 사용되는데 그 중 드론 촬영이나 데이터 처리 분야에 관심이 높아지면서 UAV의 데이터 품질을 고려한 자원 할당 시스템이 주목받고 있다.

본 논문에서는 최근 UAV의 자원 할당 최적화 목적으로 적용되는 강화학습(Reinforcement Learning, RL) 기술 동향을 분석한다. UAV 시스템에서 고려해야 할 자원을 1) 각 에이전트가 소비하는 시간을 적절히 분배하는 시간 할당(Time allocation), 2) 수집하는 데이터의 양과 그 데이터의 품질을 최적화하는 데이터 할당(Data Allocation), 그리고 3) 개별 에이전트가 차지할 수 있는 공간의 양을 지정하는 공간 할당(Space Allocation)으로 크게 분류한다. 이 중 데이터 할당은 UAV들이 개별적인 데이터 수집을 할 때, 데이터의 품질 및 사용자가 받는 서비스의 최적화와 통신량으로 발생하는 잡음 문제 최소화 등으로 세분화할 수 있다.

Ⅱ. 본론

인공지능 기술을 통한 학습보다는 UAV 성능 자체의 개선에 대하여 진행되었던 기존 연구에서와 달리, 최근의 UAV 기반 연구들은 환경(Environment) 변화에 대한 따른 대응과 배터리 용량 등의 한계를 극복하려는 목적[3]으로 강화학습 알고리즘을 도입하여 UAV 시스템의 개선을 연구하는 경향을 보인다. 본 장에서는 UAV 환경에서 자주 응용되는 다중 에이전트 강화학습에 대한 개념을 서술하고, UAV 기반 시스템에 도입한 다중 에이전트 강화학습이 양질의 데이터 수집 및 서비스 최적화 그리고 통신 과정에서 발생할 수 있는 데이터 손실 최소화를 목적으로 세운 최신 연구 동향을 논한다.

2-1 다중 에이전트 강화학습

특정 UAV 시스템에 도입되는 강화학습 알고리즘은 사용하는 UAV의 개수에 따라 달라지기도 한다. UAV가 군집 주행을 하는 경우, 일반적으로 MARL(Multi Agent Reinforcement Learning) 알고리즘이 사용되고 있다. MARL 알고리즘[4]은 단일 에이전트 기반 강화학습만으로 해결될 수 없는 트래픽 제어 또는 작업 분배 시스템 등에 사용되어 강화학습의 열풍과 더불어 재조명받고 있다. 그림 1에서 볼 수 있듯이, MARL은 단일 에이전트 강화학습과 마찬가지로 MDP(Markov Decision Process) 기반의 시스템 최적화 문제를 해결하는데, 상태(State)와 각 에이전트가 받는 보상(Result)은 모든 에이전트의 공동 행동(Joint action)에 영향을 받는다[5].

2-2 데이터 양질 및 서비스 최적화

UAV는 무선 통신(Wireless Communication)을 통해 데이터를 전송 또는 수집할 수 있다. 무선 통신 센서로부터 데이터를 수집하고 중앙 기지국에 데이터를 전송하는 ADCP(Aerial Data Collection Problem) 문제[6]가 제안되었다. 해당 문제를 해결하기 위해 DQN(Deep Q-Network) 알고리

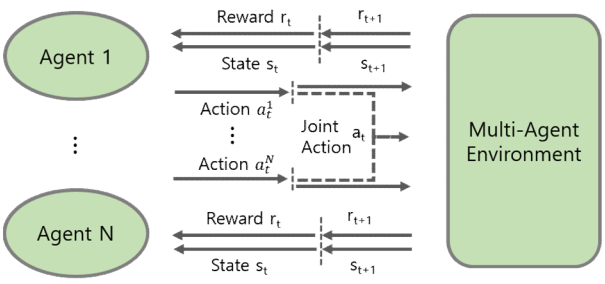


그림 1 다중 에이전트 강화학습 구조

즘이 UAV의 서비스를 받는 사용자들의 QoS(Quality of Service)와 데이터 수집 정도를 파악하는 것을 목적으로 사용되었다.

첫 번째 연구에서는 UAV를 통해 수집된 데이터가 사용자의 개별 휴대용 장치로 전송될 때, 전송되는 데이터의 QoS를 휴대용 장치의 이동통신망(Cellular Network)에서 보장받기 위하여 다중 에이전트 DQN 알고리즘을 고려한 모델[7]이 제시되었다. 또 다른 연구[8]에서는 기존 DQN 알고리즘을 뛰어넘는, PER(Prioritised Experience Replay)와 고정된 Q 타겟(Q-targets)을 기반으로 DDDQN(Double Dueling Deep Q-Network) 모델을 설계하여 오버피팅(Over-fitting) 문제의 해결 및 성능 향상을 이끌어내었다. 해당 연구는 빅데이터 기반의 사례 연구를 이용하여 양적인 QoE(Quality-of-Experience)와 질적인 QoS(Quality-of-Service)를 모두 고려하였다. 마지막으로, MARL 알고리즘 기반으로 분산형 사물인터넷(IoT) 장치로부터 다중 UAV가 수집하는 데이터양을 최대화하는 연구가 있다. 해당 연구에서는 시스템에 존재하는 UAV의 방향을 조정함으로써 수집되는 데이터양을 최적화하는 목적을 달성하였다. 제안된 네트워크 구조[9]는 DRL(Deep Reinforcement Learning)을 통한 Dec-POMDP(Decentralized-Partially Observable Markov Decision Process) 기법으로 변환되어 데이터 수집 작업을 UAV 간 분배하여 최적화하였다. 이처럼 여러 제약 조건을 동시에 고려하면서도 최적화 목적을 수행하는 UAV 시스템이 연구되는 추세이다.

2-3 데이터 손실 최소화

UAV 기반 환경에서 발생하는 데이터 손실은 네트워크 연결 방해(Interference), 통신량의 급증, 또는 직접적인 위협 등을 받을 때 발생한다. 데이터 손실은 개인 정보 유출, 저장되지 않은 정보의 삭제 그리고 촬영 결과물 퀄리티 저하의 이유로 문제가 되고 있다. 문제를 해결하기 위하여 다중 에이전트들끼리의 정보 교환 없이 사용자와 UAV가 상호작용하는 상황에서 자원 할당에 따른 보상을 최대화하려는 목적의 연구가 진행되었다. 첫 번째 연구에서는 모든 에이전트가 독립적으로 의사결정 알고리즘을 실행하지만, Q-러닝 기반의 공통적인 다중 강화학습 구조[10]를 공유하는 에이전트 독립 방법이 제안되었다. 두 번째 연구[11]에서는 사물인터넷 노드에서의 버퍼 오버플로우(Overflow)와 채널 손실로 인한 데이터 전송 오류를 줄이기 위하여 비행경로와 데이터 수집 스케줄링을 POMDP(Partially Observable Markov Decision Process)로 공식화하여 UAV가 수집하는 데이터양을 최대화하였다. 이 연구를 통해 대규모 환경 감지 기술 발전을 위한 사물인터넷 네트워크에서의 데이터를 수집 과정에서 데이터 손실을 줄이는 것이 가능해졌다. 마지막으로 DQN 알고리즘를 개선한 DDPG(Deep Deterministic Policy Gradient) 알고리즘을 기반으로 DeFRAS(Deep Reinforcement Learning based Flight Resource Allocation Scheme)을 제안하여 UAV의 방향과 순간 속도를 조절하여 데이터 손실을 최소화한 연구[12]도 있다.

Ⅲ. 결론

강화학습 알고리즘을 도입하면서 UAV 시스템은 더 높아진 기동성과 넓어진 활용 범위를 가지게 되었고, 자체적인 물리적 한계를 극복해 나가며 급속도로 발전해오고 있다. 특히, 최근 정보 보안 및 위험 방지 분야에 주목도가 높아지면서 자원 할당 중 데이터 할당 최적화 연구의 발전이 전망되고 있다. 본 논문에서는 자원 할당 최적화 목적의 UAV 기반 시스템에 도입한 다중 에이전트 강화학습을 데이터의 양질 및 서비스 최적화 목적과 데이터 손실의 최소화 목적으로 분류하여 최신 연구 동향을 살펴보았다. 다양한 데이터 종류와 환경 특성에 따라 상황별 고려 요소를 추가하여 확장하는 것이 앞으로의 연구 과제이다.

ACKNOWLEDGMENT

This research is funded by Ministry of Health and Welfare (HI19C0842).

참 고 문 헌

[1] B. Li, Z. Fei, and Y. Zhang, “UAV Communications for 5G and Beyond: Recent Advances and Future Trends,” *IEEE Internet Things Journal*, vol. 6, no. 2, pp. 2241 - 2263, Apr. 2019.

[2] A. Fotouhi et al., "Survey on UAV Cellular Communications: Practical Aspects, Standardization Advancements, Regulation, and Security Challenges," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3417-3442, 2019.

[3] A. Carrio, et al., “A Review of Deep Learning Methods and Applications for Unmanned Aerial Vehicles,” *Journal of Sensors*, vol. 2017, no. 3296874, Aug. 2017.

[4] L. Canese, et al., “Multi-Agent Reinforcement Learning: A Review of Challenges and Applications,” *Applied Science,* 2021.

[5] K. Zhang, Z. Yang, and T. Basar, “Multi-Agent Reinforcement Learning: A Selective Overview of Theories and Algorithms,” *arXiv preprint arXiv:1911.10635*, 2021.

[6] C. Caillouet, F. Giroire, and T. Razafindralambo, “Efficient data collection and tracking with flying drones,” *Ad Hoc Networks*, vol. 89, pp. 35-46, 2019.

[7] F. Wu, et al., "Cellular UAV-to-Device Communications: Trajectory Design and Mode Selection by Multi-Agent Deep Reinforcement Learning," *IEEE Transactions on Communications*, vol. 68, no. 7, pp. 4175-4189, 2020.

[8] W. Guo, "Partially Explainable Big Data Driven Deep Reinforcement Learning for Green 5G UAV," *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, pp. 1-7, 2020.

[9] H. Bayerlein, et al., "Multi-UAV Path Planning for Wireless Data Harvesting With Deep Reinforcement Learning," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 1171-1187, 2021.

[10] J. Cui, Y. Liu and A. Nallanathan, "Multi-Agent Reinforcement Learning-Based Resource Allocation for UAV Networks," *IEEE Transactions on Wireless Communications*, vol. 19, no. 2, pp. 729-743, Feb. 2020.

[11] K. Li, et al., "Joint Flight Cruise Control and Data Collection in UAV-Aided Internet of Things: An Onboard Deep Reinforcement Learning Approach," *IEEE Internet of Things Journal*, vol. 8, no. 12, pp. 9787-9799, 2021.

[12] K. Li, W. Ni and F. Dressler, "LSTM-characterized Deep Reinforcement Learning for Continuous Flight Control and Resource Allocation in UAV-assisted Sensor Network," *to appear in IEEE Internet of Things Journal*, 2021.

무선 이미지 전송을 위한 딥러닝 기반의 효율적인 결합 소스-채널 부호화

임수빈°, 김정현°, 송홍엽*
순천향대학교°, 연세대학교*
qlstn9150@gmail.com, kimjh@sch.ac.kr, hysong@yonsei.ac.kr

An Efficient Joint Source-Channel Coding Based on Deep Learning for Wireless Image Transmission

Subeen Leem°, Junghyun Kim°, Hong-Yeop Song*
Soonchunhyang Univ.°, Yonsei Univ.*

요 약

본 논문은 무선 통신 시스템에서 이미지 전송을 위한 결합 소스-채널 부호화 기법에 대해 다룬다. 기존의 딥러닝 기반 결합 소스-채널 부호화 기법인 DJSCC의 신경망 구조를 개선하여 향상된 성능을 갖는 새로운 딥러닝 기반의 결합 소스-채널 부호화 기법을 제안한다. 실험을 통해 제안한 기법이 기존의 DJSCC 기법보다 뛰어난 성능을 보임을 확인하였다.

Ⅰ. 서 론

결합 소스-채널 부호화 기법은 소스 압축률과 채널 부호율을 가변적으로 적용하여 서비스 품질을 만족시키면서 대역폭 사용 효율을 높이는 부호화 기법이다. 전통적인 소스 부호화는 JPEG나 JPEG2000과 같은 압축 알고리즘을 통해 이미지 픽셀들 간의 상관관계를 줄이고 전송되는 정보량을 줄인다. 또한 채널 부호화는 LDPC, Turbo, Polar 부호와 같은 오류정정부호를 사용하여 무선 채널 환경에서 발생하는 오류를 극복하기 위한 탄력적인 전송을 가능하게 한다.

최근 인공 신경망을 이용하여 대부분의 채널 환경에서 기존의 소스-채널 부호화의 성능을 능가하는 딥러닝 기반의 결합 소스-채널 부호화 기법인 DJSCC(Deep Joint Source-Channel Coding)[1]이 제안되었다. 해당 기법은 채널의 변화에도 급격한 성능 저하를 겪지 않는 장점이 있다. 그러나 일부 채널 환경에서는 여전히 기존의 소스-채널 부호화의 성능보다 낮은 성능을 보이므로 여전히 개선의 여지가 남아있다.

본 논문에서는 DJSCC의 구조적 한계를 극복하여 더 나은 성능을 갖는 결합 소스-채널 부호화 기법을 제안한다. 특히 DJSCC에서 사용한 단순한 구조의 CNN에 층 간 연결을 추가한 DenseNet[2] 구조를 송수신기에 활용함으로써 이미지 특징 추출 기능을 향상시켰다.

Ⅱ. 본론

DJSCC에서 사용한 신경망 구조는 송신기와 수신기 모두 기본 컨볼루션 층이 5개 쌓여있는 CNN 구조로 되어있다. DJSCC의 성능을 개선하기 위해 기존 송신기 구조에서 두 번째 층의 출력을 세 번째와 네 번째 층의 출력에 연접하고, 세 번째 층의 출력을 네 번째 층의 출력에 연접하였다. 또한 기존 수신기 구조에서 첫 번째 층의 출력을 두 번째와 세 번째 층의 출력에 연접하고, 두 번째 층의 출력을 세 번째 층의 출력에 연접하였다. 이러한 밀집한 신경망 구조를 통해 파라미터 수를 크게 증가시키지 않고도 이미지 특징 추출 기능을 향상시킬 수 있다.

제안하는 DJSCC-DN 기법의 구조가 그림 1에 표현되어 있다. 송신기에서는 0~255의 픽셀 값을 갖는 입력 이미지를 0~1로 정규화 한 후, 다섯 개의 컨볼루션 층으로 전달한다. 각 컨볼루션 층의 활성함수로 비선형 활성화함수인 PReLU[3]를 사용한다. 그림 1의 각 블록 위에 쓰여진 표기는 (필터 가로 × 필터 세로 × 필터 수 / 스트라이드 간격)을 의미하는데, 마지막 컨볼루션 층의 필터 수 c는 학습 시에 정의하는 대역폭 압축률에 따라 조절된다. 이어서 평균 전력 제약 조건을 만족하게 하는 정규화[4]를 진행한 후, 잡음 채널을 통해 전송된다. 수신기는 다섯 개의 전치 컨볼루션 층과 역정규화 층으로 구성되어 송신기와 대칭되는 구조로 설계하였다. 마

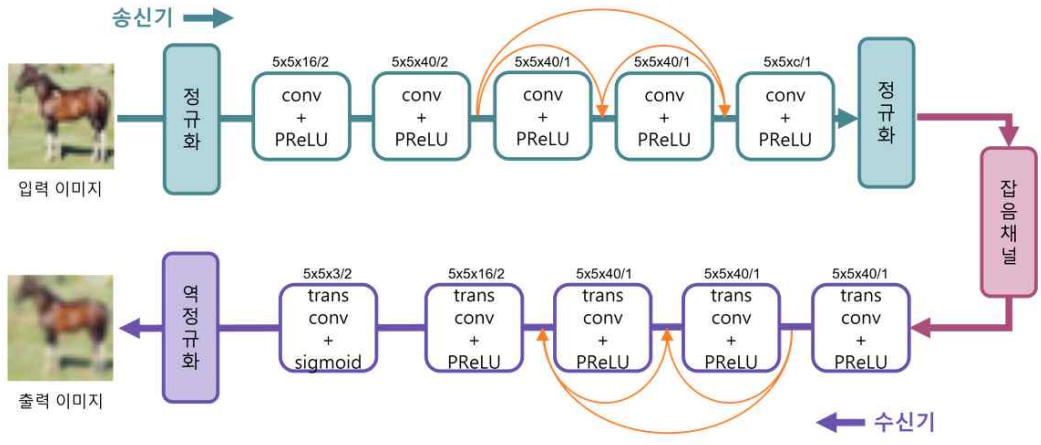


그림 1. DJSCC-DN의 구조

지막 역정규화 층은 전치 컨볼루션 층의 출력 값을 원본 이미지의 픽셀 값의 범위인 0~255로 맞춰주는 역할을 한다.

학습 및 테스트에 사용한 데이터셋은 32x32크기의 컬러 이미지들로 구성된 CIFAR-10이다. 이 중 50,000장의 이미지는 학습 데이터로, 10,000장의 이미지는 테스트 데이터로 분류하여 사용했다. 아래 표는 DJSCC-DN의 학습에 사용된 하이퍼 파라미터를 나타낸 것이다.

파라미터	설정값
손실 함수	MSE(Mean Squared Error)
배치 크기	100
최적화 방법	Adam, 학습률=0.001
반복 학습 횟수	20
학습 방법	SNR=0,10,20dB마다 학습

송신된 이미지와 수신된 이미지 간의 손실 정보를 평가하기 위해 최대 신호 대 잡음비(Peak Signal-to-Noise Ratio, PSNR)[5]을 성능 지표로 사용하였다. 이는 다음의 식 (1)과 같이 정의된다.

$$PSNR = 10\log_{10}\frac{MAX^2}{MSE}\ (dB) \quad (1)$$

여기서 MAX는 이미지의 한 픽셀이 가질 수 있는 최댓값이고, MSE는 송신 이미지와 수신 이미지의 각 픽셀 오차를 제곱하여 평균 낸 값이다. 즉, PSNR값이 클수록 손실이 적은 이미지라고 할 수 있다.

그림 2는 기존 DJSCC 기법과 제안하는 DJSCC-DN 기법의 대역폭 압축률에 따른 PSNR 성능을 비교한 것이다. 여기서 대역폭 압축률은 입력 이미지가 n차원 실수 벡터로 표현되고 채널 입력이 k차원 복소수 벡터로 표현될 때 이들 간의 비율인 k/n로 정의된다. 학습 및 테스트를 위한 채널 환경으로 신호 대 잡음비(Signal-to-Noise Ratio, SNR)를 0dB, 10dB, 20dB로 고려하였다. 또한 대역폭 압축률은 0.06, 0.26, 0.49인 경우를 고려하였다. 고려한 각 채널 환경에서 모든 대역폭 압축률에 대해 제안하는 DJSCC-DN 기법이 기존 DJSCC 기법 대비 같거나 더 나은 성능을 보이는 것을 확인할 수 있다.

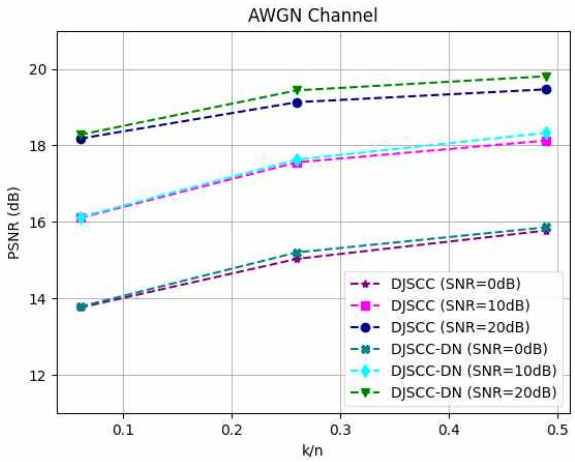


그림 2. 대역폭 압축률(k/n)에 따른 PSNR 성능 비교

그림 3은 대역폭 압축률을 0.49로 고정하고 다양한 테스트 채널 환경에서 기존 DJSCC 기법과 제안하는 DJSCC-DN 기법의 PSNR 성능을 비교한 것이다. 신호 대 잡음비가 0dB, 10dB, 20dB인 환경을 타켓으로 학습된 각 모델에 대해 타겟 채널 환경과 유사한 테스트 환경에서 제안하는 DJSCC-DN 기법이 기존 DJSCC 기법 대비 같거나 더 나은 성능을 보이는 것을 확인할 수 있다. 특히 채널 환경이 좋을수록 이 격차는 크게 증가하였다.

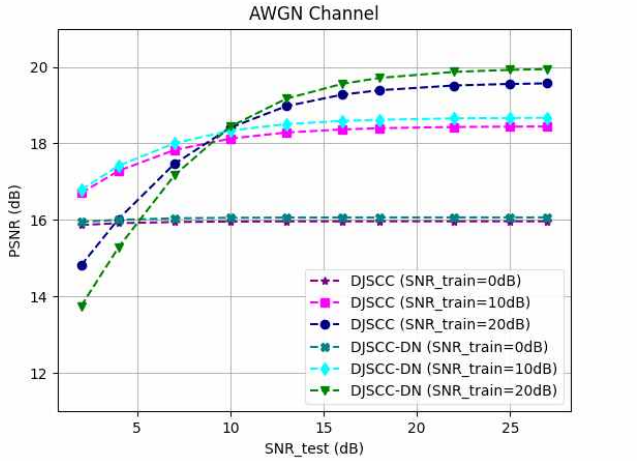


그림 3. 테스트 환경에서의 신호 대 잡음비에 따른 PSNR 성능 비교

Ⅲ. 결론

본 논문에서는 무선 통신 시스템에서 이미지 전송을 위한 딥러닝 기반 결합 소스-채널 부호화 기법인 DJSCC의 구조를 개선하여 DJSCC-DN을 제안하였다. 다층으로 구성된 CNN 기반 송수신기 구조에 층 간 연결을 추가하여 특징 추출 기능을 향상시켰다. 향후 추가 성능 개선을 위한 연구 방법으로 새로운 신경망 구조 활용과 PSNR 뿐만 아니라 SSIM[5]을 성능 지표로 활용한 학습 방법과 다양한 채널 환경에서의 성능 비교 및 분석 등을 고려할 수 있다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No.2020R1A2C201196911).

참 고 문 헌

[1] E. Bourtsoulatze, D. Burth Kurka, and D. Gündüz, “Deep Joint Source-Channel Coding for Wireless Image Transmission,” *IEEE Transactions on Cognitive Communications and Networking*, vol. 5, no. 3, pp. 567-579, Sep. 2019.

[2] G. Huang, Z. Liu, L. Van Der Maaten, and K. Weinberger, “Densely Connected Convolutional Networks,” in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017 pp. 2261-2269.

[3] T. Jiang and J. Cheng, “Target Recognition Based on CNN with LeakyReLU and PReLU Activation Functions,” in *Proc. International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC)*, 2019, pp. 718-722.

[4] I. Kozintsev and K. Ramchandran, “Robust Image Transmission over Energy-Constrained Time-Varying Channels using Multiresolution Joint Source-Channel Coding,” *IEEE Trans. Signal Process.*, vol. 46, no. 4, pp. 1012-1026, Apr. 1998.

[5] D. Ziou and A. Horé, “Image Quality Metrics: PSNR vs. SSIM,” in *Proc. International Conference on Pattern Recognition (ICPR)*, Istanbul, Turkey, 2010 pp. 2366-2369.

독거노인 고독사 방지를 위한 다양한 실시간 행동인식 방법

정우진
경희대학교 전자공학과

woojinchung@khu.ac.kr

Various Real-Time Behavior Detection to Prevent Dying Alone of Elderly People Living Alone

Woo Jin Chung
KyungHee Univ.

요 약

본 논문에서는 독거노인 고독사 예방을 위해 opensource yolo 의 실시간 객체 인식을 이용한 다양한 독거노인 행동 인식 시스템을 제안한다. 기존에 사용하던 움직임 감지 센서는 단순한 움직임만을 감지하기 때문에 독거노인의 현재 상황을 제대로 판단하지 못하는 한계가 존재한다. 본 논문에서는 이러한 문제점을 극복하기 위하여 객체인식을 통해 검출된 사람과 침대의 위치정보를 토대로 독거노인의 행동상태를 취침, 외출, 일반 생활, 이상 상태 등 4 가지로 구분하였다. 객체의 인식률과 행동의 인식률을 각 mAP 와 recall 로 나타내어 제안한 시스템의 결과가 신뢰성이 있음을 보였다.

Ⅰ. 서론

독거노인 고독사 관련된 문제가 제기되어 왔다 [1]. 이에 움직임 감지 PIR 센서를 기반으로 단순한 움직임을 인지하는 방식이 사용되고 있다. 움직임이 일정시간 이상 없는 경우 문제가 생긴 것으로 판단되어 경보가 울린다 [2]. 그러나 잠들어 있거나 외출을 하는 경우에도 움직임이 없기에 false 알람 확률이 커지고 외출 후 문제가 발생한 경우 감지를 못하는 문제가 존재하는 등 다양한 생활 패턴을 반영한 동작을 감지 하지 못한다.

본 논문은 이러한 문제들을 해결하기 위해 단순한 움직임만 감지하는 것이 아니라 사람의 행동을 보다 정확하게 판단하여 독거 노인의 고독사를 미리 방지하는데 도움이 되도록 한다. 실내에 있는 사람의 다양한 동작 검출을 위해서 기계 학습을 이용한 행동 인식을 하고 방 안에서 사람의 위치정보를 기반으로 현재 행동 상태를 '취침 상태', '일반 생활 상태', '외출 상태', '이상 상태'로 구분하여 행동을 인식하였다. 본 논문에서 분류한 4 가지 행동에 대해 Yolo 알고리즘을 이용하여 실험하고 검출 확률을 보였다.

Ⅱ. 본론

본 논문에서는 Yolo 를 이용하여 고독사 방지를 위한 4 가지 행동 유형을 분류하고 각각에 대한 행동 검출을 분류하고 이를 판단하는 방법에 대해서 설명을 한다.

1. 독거 노인 행동 분류

사람의 움직임이 장시간 없는 경우가 위험한 경우이므로 움직임이 없는 상태를 상세 분류한다. 본 논문에서는 방 안 사람의 행동을 '일반 생활 상태', '취침 상태', '외출 상태', '이상 상태' 총 4 가지로 나눈다.

첫 번째로 '일반 생활 상태'는 움직임이 비주기적으로 일어나는 상황을 간주한다

두 번째로 '취침 상태'는 사람이 침대에 있는 경우 침대안에 있는 사람과 사람을 각각 검출하여 사람이 침대 위에 있음을 판단한다. 침대 객체의 크기가 사람 객체의 크기보다 크기 때문에 인식된 사람의 객체가 침대보다 크지 않으면 안에 사람이 있다고 간주할 수 있으나 사람의 수면 상태에서 움직임이 있을 경우 사람 객체의 크기가 침대보다 커지는 경우가 발생한다. 그러므로 사람 객체의 중심점이 침대 객체 내부 일정 범위 안에 위치하는지 검출함으로써 움직임이 있는 '취침 상태'를 검출한다. 수면 중 사망을 검출하기 위해 주기적으로 침대 안 사람 객체의 움직임을 검출한다.

세 번째로 '외출 상태'는 사람이 문으로 나가서 존재하지 않은 상태이다. 이에 따라 사람 객체가 일정시간 이상 검출되지 않는 경우 '외출 상태'로 구분하였다.

네 번째 이상 상태의 경우 독거노인의 건강에 문제가 발생한 상태이다. 먼저 '취침 상태'에서 움직임이 일정시간 이상 없는 경우 이상이 생긴 경우라고 간주한다. 외출 후 이상 상태가 발생하였을 상황을 고려하여 '외출 상태'에서 일정시간 시간 이내에 사람의 움직임이 검출이 안된 경우를 외출 후 이상 상태로 구분한다. 그리고 '일반 생활 상태'에서 사람이 쓰러져 움직이지 못하고 있을 상황을 고려하여 '일반 생활 상태'에서 사람의 중심점이 일정 범위 이상 움직이지 않은 상태로 일정시간이 경과한 경우도 이상 상태로 구분하였다.

2. Yolo

Yolo(you only look once)는 객체 인식을 위해 CNN(convolutional neural network)으로 만들어진 open source 알고리즘이다. 기존의 R-CNN(recurrent-CNN)과 같은 객체 인식 모델은 입력 이미지 안에서 bounding box 생성, classifier 적용, 분류와 후처리 등을 수행한다[3]. 그러나 yolo 는 이 모든 절차를 하나의 회귀 문제로 간주하여 하나의 파이프라인으로 구성하기 때문에 처리속도가 다른 알고리즘에 비해 연산이 빠른 장점(프레임당 약 0.029 초)이 있고 새로운 이미지에서도 검출 정확도가 높다[3].

본 논문에서는 Yolo 를 이용해 일정 시간마다 방 안 영상 속 객체를 일정 시간 간격에서 인식된 객체들의 위치 변화를 판단하여 다양한 독거노인 행동을 인식할 수 있는 프로그램을 제시하였다.

3. YOLO 를 통한 객체 인식 시스템

Yolo 의 객체 인식은 다음과 같은 순서로 진행이 된다. 먼저 입력된 영상의 이미지를 resize 한다. 본 논문에서는 yolov2 608x608 을 사용하였기에 608x608 로 resize 한다.

그림 1 은 yolo 의 architecture 를 나타낸 그림이다. 그림 1 의 단일 CNN 은 각 grid cell 마다 5 개의 anchor box 를 설정하고 그 box 가 객체일 확률을 예측하고 클래스를 분류한다[3][4].

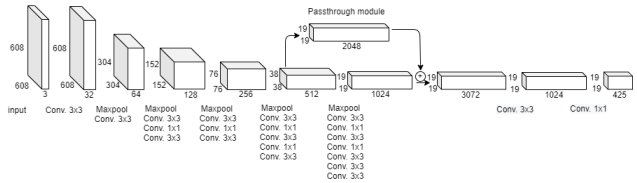


Fig. 1. Yolov2 architecture

Convolution network 의 결과로써 5 개 anchor box 의 정보를 가진 19x19 feature map 을 얻을 수 있다. 이 feature map 에서 영상 속 객체에 대한 bounding box 의 좌표점들을 얻을 수 있고 이를 객체의 위치를 사용한다[3][4].

4. 실험결과

본 논문에서 제안한 독거 노인 다양한 행동 상태를 검출하는 실험은 Darknet 에서 80 개의 클래스를 가진 coco dataset 2014 를 yolov2 로 학습시킨 가중치를 사용하여 진행하였다. 위에서 나눈 4 가지 행동 상태를 구분하기 위해서 coco data set 의 기존 80 개 클래스 중 '사람'과 '침대' 클래스만 인식하도록 하였다. 인식된 2 가지 객체의 bounding box 정보를 이용하였다. 본 논문에서 고려하는 행동 상태 판단 예는 그림 2 와 같다.

실험은 촬영 영상을 프레임 단위로 추출한 7467 장의 사진으로 진행되었다. '사람', '침대' 2 가지 객체 클래스에 대한 객체인식 성능을 mAP 로 나타낸 값은 표 1 과 같다. '이상 상태'를 제외한 3 가지 행동 클래스에 대한 행동인식 recall(sensitivity)은 표 2 와 같다. 이상 상태의 경우 '취침 상태', '외출 상태', '일반 생활 상태' 판단을 기반으로 시간적인 조건이 붙는 것이기 때문에 각각의 검출율을 따른다. 본 시스템의 행동 상태의 평균 인식률(이상 상태 검출율)은 94.1%의 정확도를 보였다.

'취침 상태'의 경우 yolo 의 특성상 bounding box 가 물체의 방향과 상관없이 직사각형으로 검출되기 때문에 사람이 침대의 구석에서 취침하는 경우 검출 오류 확률이 증가하게 된다. 그러나 카메라를 침대의 각 변이 화면의 축들과 평행하도록 설치한다면 검출율이 대폭 상승될 것이다.

Table 1. Object detection performance of the system

침대	사람	mAP
99.8%	90.3%	95.1%

Table 2. Behavior detection performance of the system

취침 상태 recall	일반 생활 상태 recall	외출 상태 recall	Mean recall
83.6%	99.6%	99.3%	94.1%

Table 3. Behavior detection performance of the system

이상 상태 missing rate	5.9%



Fig. 2. Behavior detection examples of the classes

Ⅲ. 결론

본 논문에서는 독거노인의 다양한 행동 인식을 통해 이상 상태를 판단 후 경보를 울려 독거노인의 고독사를 방지할 수 있는 시스템을 제안한다. Open source 객체인식 모델인 yolo 를 사용하여 객체인식을 하였고 이를 바탕으로 독거노인의 행동을 인식을 하였다. 행동 상태의 인식 결과가 높은 정확도를 보였다. 추후 연구에서는 일반 생활 상태를 더욱 세분화하여 더욱 다양한 이상 상태를 판단하고 검출 오류를 더 줄일 예정이다.

참 고 문 헌

[1] 송영신, "여성 노인 1 인 가구의 실태 및 정책적 개선방안," 이화젠더법학, 7(2), pp. 33-72, 2015.

[2] 장한솔, 김수정, 박영호, "SilverLinker: IoT 센서 기반 독거노인 케어 플랫폼," 한국디지털콘텐츠학회 논문지, 19(11), 2235-2245, 2018.

[3] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, " You Only Look Once: Unified, Real-Time Object Detection," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779-788, 2016

[4] Joseph Redmon, Ali Farhadi, "YOLO9000: Better, Faster, Stronger," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 7263-7271, 2017.

전동킥보드 불법행위 판별을 위한 딥러닝 시스템

전하용
*최 예원, 김 찬욱, 이 현성
경남대학교 MSC교육 융합대학
경남대학교 컴퓨터공학부
* wony9802@naver.com , gcw9438@naver.com , dlgustjd3513@gmail.com , hayongj@kyungnam.ac.kr

Deep learning system that determination illegal activities on scooter

*Yewon Choi , Chanwook Kim, Heonsung Lee
저자 소속 *Kyungnam University Department of Computer Science

요 약

전동킥보드의 사용량이 증가함에 따라 이슈가 되고 있는 사회적 문제점들을 해결하고자 연구를 진행한다. 이 문제점들을 해결하기 위해 Deep learning 기반인 Deep sort를 알고리즘을 활용한다. 전동킥보드의 도로교통법 규정 위반인 헬멧 미착용, 2인 이상 탑승, 규정 속도위반 등의 행위를 하는 탑승자들을 Detection하고 Tracking을 하여 불법행위를 인식하고자 한다.

키워드: YOLO, deep sort, AI, detection, tracking

Ⅰ. 서 론

최근 공유 PM(Personal Mobility, 개인형 이동장치) 서비스인 전동킥보드가 무섭게 성장하며 개인의 편리함과 이용량이 증가하며 관련된 사고도 함께 증가하였다. 이러한 사고의 주된 원인은 탑승자의 음주, 미숙한 운전 실력 등 이지만 중상을 입는 환자 대부분이 헬멧 미착용, 2인 이상 탑승, 과속으로 도로교통법을 위반했기 때문이다. 실제 국회 행전안전위원회 소속 박완수 의원이 공개한 자료를 보면 올해 5~6월 간 헬멧 미착용으로 단속된 사례는 전국 5400건 정도에 달하며 사망사수도 급격히 증가하고 있다.1) 이러한 이유로 전동킥보드의 안전 문제는 많은 주목을 받고 있지만 현실적으로 도로 경찰이 모든 불법행위를 단속하는 것에는 한계가 있다. 그래서 Deep learning 기반의 YOLO V5(You only look once)의 deep sort를 사용하여 킥보드 불법행위를 단속하고 범칙자를 tracking 할 수 있게 구현하였다. 단속 대상은 헬멧 미착용자, 속도가 규정된 속도보다 빠를 경우 혹은 2인 이상 탑승한 불법 행위자이다. 나아가 본 연구를 활용하여 다른 이동수단에도 활용할 수 있다.

Ⅱ. 관련 연구

2.1 YOLO v5

YOLO는 you only look once 뜻으로 이미지를 한번 보는 것으로 object의 종류와 위치를 추측, 단일 네트워크를 통해 여러 개의 테두리 상자에 대한 객체 확률을 계산하여 보여 주는 알고리즘이다. 2)
YOLO의 첫 개발 이후부터 계속해서 발전되어가고 있다. YOLO는 빠르게 이미지에서 객체를 탐지하는 모델로 큰 3가지 특징이 있다. 첫 번째는 이미지 전체를 단 한 번만 본다. 두 번째는 통합된 모델을 사용한다. 다른 객체 탐지 model들은 다양한 전처리 모델과 인공 신경망을 결합하여 사용한다. 하지만 YOLO는 단 하나의 인공 신경망에서 전부 처리한다. 마지막으로 실시간 객체 탐지가 가능하다. 이 특징이 YOLO를 유명하게 만든 것으로 YOLO가 높은 성능으로 객체를 탐지하는 model은 아니지만 여러 장의 이미지를 탐지할 수 있다.

2.2 Deep sort

Deep sort는 MOT(Multiple Object Tracking)의 종류 중 하나인 Detection-Based Tracking에 포함되어있다. Deep sort 알고리즘의 개념은 복잡하지 않지만 Kalman Filter, SORT 기술, NMS(Non Maxmimum Suppressions) 기술, re-id 기술 등 다양한 기술을 사용한다. 요약하자면 Deep Sort는 Kalman Filter를 기본으로 Deep feature (Re-Id)를 추가로 반영하여 헝가리안 알고리즘을 수행한다. 3)

Ⅲ. 연구

3.1 알고리즘

본 연구에 적용된 전체적인 순서도는 그림 1과 같다. 헬멧, 킥보드, 사람을 Detecting하고 각각의 Object중심점 거리를 계산하는 알고리즘을 기반으로 먼저 사람과 킥보드 중심점 사이 거리로 탑승자인지를 결정한다. 그리고 킥보드 탑승자와 헬멧 중심점 사이 거리로 착용유무를 판단 후 탑승자와 사람의 중심점 사이 거리로 수가 탑승자가 2인 이상인지 확인한다. 마지막으로 그림2 공식을 적용하여 속도가 25km/h인지 확인한다. 모든 규정을 지킨 각각의 object는 하나의 object로 결합하며 “GOOD”로 표기한다. 하나라도 적합하지 않을 경우 object를 “BAD”로 표기한다.

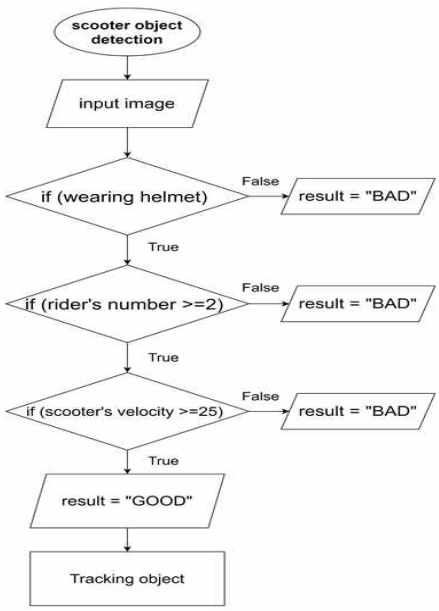


그림 1. flow chart

$$velocity = \frac{distance}{time}$$

그림 2. 속도계산 공식

3.2 평가 및 검증

동영상과 이미지 등을 크롤링을 활용하여 데이터를 수집하고, 이미지 라벨링을 통해 YOLO v5s로 필요한 모델을 재학습 진행시켜 모델을 완성한다. 학습에 사용한 이미지의 크기는 640x640으로 총 22530장이다.

표 1. Evaluation(V100)

batch/ epohcs	person	scooter	helmet	all classes
8/100	0.977	0.982	0.984	0.981 mAP@0.5
16/100	0.981	0.982	0.985	0.983 mAP@0.5
32/100	0.981	0.985	0.988	0.984 mAP@0.5
64/100	0.979	0.982	0.987	0.983 mAP@0.5

연구 환경은 Ubuntu20.04, CUDA11.0, NVDIA-tesla-v100, pytorch1.7.1, python3.8로 진행하였을 때 표 1에서처럼 다양한 batchsize와 epoch를 사용한 결과 가장 좋은 정확도를 보인 것은 batchsize는 32, epochs는 100이다. 높은 정확도를 보여줄 것이라 예상한 batchsize는 64, epochs는 100은 overfitting 되어 실제 상황에선 오히려 낮은 정확도를 보였다. 결과적으로 해당 object들은 평균 95% 이상의 정확도로 성공적으로 detection을 하였다.



그림 2. object detection



그림 3. 모든 규정 준수

그림2는 각각의 object들의 detiction 정확도로 높은 정확도를 보여주는 것을 확인 할 수 있다. 그림 3은 그림1의 모든 규정을 준수하였을 때 결과로이다.

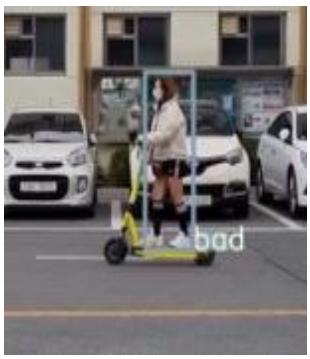


그림 4. 헬멧 미착용



그림 5. 2인 이상 탑승

그림 4와 그림 5는 그림 1의 모든 규정을 준수 하지 않아 “BAD”로 감지한 결과이다. 그림 4는 탑승자가 헬멧을 미착용으로 그림 5는 전동킥보드에 탑승한 탑승자가 2명 이상으로 규정을 준수 하지 않았다.

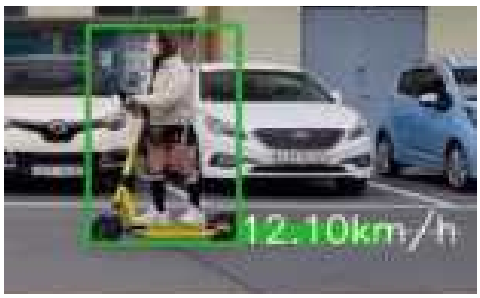


그림 6. 킥보드 속도 측정

그림 6은 그림2의 공식을 적용하여 전동킥보드의 속도를 측정한 한 예시이다.

Ⅳ. 결론

본 연구는 전동킥보드의 도로교통법 규정 위반행위를 감지하기 위해 킥보드, 헬멧 탑승자를 각각 구분하여 Detecting하고 Tracking 하는 것을 제시한다. 이를 통해 탑승자들에게 언제 어디에서든 교통법을 준수해야 한다는 경각심을 줄 수 있다. 이 기술을 더 발전시켜 킥보드 크기와 카메라 사이의 거리, 주행 거리를 측정하여 인도나 킥보드 전용도로(정해진 도로)가 아닌 장소에서 주행 여부 확인이 가능하다. 그리고 하나의 카메라가 아닌 여러 개의 카메라를 이용하여 동일객체 인식하고, 주행 경로 및 사고 발생 시 초기 사고 장소의 유추까지 기대할 수 있다. 킥보드뿐만이 아니라 앞으로 안전규제를 새롭게 적용받는 교통수단들이 재분류 될 수 있어 언제든 적용할 수 있어 앞으로 사람이 직접 단속하고 만약 범칙자가 도주할 경우 Tracking을 통해 빠르게 추적할 수 있다.

ACKNOWLEDGMENT

본 논문은 과학기술정보통신부 지원을 받아 경남 SW융합클러스터 2.0 특화산업 강화사업으로 지원받아 진행하였음 수행한 연구 결과임.

참 고 문 헌

[1]https://m.khan.co.kr/national/national-general/article/202107251640001#c2b

[2] 박준우, 김지훈, 이진서, 김영덕. "딥러닝 기반 YOLO를 이용한 도로내 CCTV 화재 상황 판별 및 전파." 정보 및 제어 논문집 . (2020): 530-531.

[3] N. Wojke, A. Bewley and D. Paulus, "Simple online and realtime tracking with a deep association metric," 2017 IEEE International Conference on Image Processing (ICIP), 2017, pp. 3645-3649, doi: 10.1109/ICIP.2017.8296962.

고정지역 감시 엣지 AI 단말용 자동 DB 생성 도구

허재윤, 이준구*, 최윤원*, 백장운*
영남대학교, *한국전자통신연구원
dustnia11@yu.ac.kr, *{leejg01679, yunwon.choi, jwbaek98}@etri.re.kr

Automatic DB Generation Tool for Edge AI Device Monitoring Particular Area

Jaeyoon Heo, Joon-Goo Lee*, Yun Won Choi*, Jang Woon Baek*
Yeungnam Univ., *ETRI

요 약

본 논문은 생활방범 혹은 도로방범용 지능형 CCTV와 같이 특정 지역을 감시하는 고정형 엣지 AI 단말의 객체검출기 학습을 보다 효율적으로 수행할 수 있는 지역 최적화 DB 자동 생성 도구를 제안한다. 기존의 엣지 AI 단말용 객체검출기는 경량 모델을 사용하기 때문에 오검출률 및 미검출률이 높기때문에 카메라의 실제 감시영역에 대한 학습 DB를 생성하여 재학습하는 지역 최적화 과정이 필요하다. 하지만, 사람이 직접 태깅을 진행하는 방식은 시간이 오래 걸리고 카메라가 감시하는 지역마다 지역최적화를 위한 개별 학습을 진행해야 하는 문제가 있다. 이러한 문제를 해결하기 위해 본 논문에서는 고정지역 감시 엣지 AI 단말용 자동 학습 DB 생성 도구를 개발하였고, 자동 생성된 학습 DB를 이용하여 객체검출기를 재학습하였다. 실험을 통해 자동생성된 학습 DB와 사람이 직접 생성한 학습 DB에서 재학습한 객체검출기의 성능을 비교하였다. 실험 결과, 제안한 도구에서 자동생성된 학습 DB로 재학습한 객체 검출기가 사람이 직접 생성한 학습 DB로 재학습한 객체검출기와 유사한 성능을 보이는 것을 확인할 수 있었다.

Ⅰ. 서 론

최근 컴퓨터 비전과 딥러닝을 활용한 객체 검출 기술이 확대되면서 다양한 분야에서의 응용 서비스가 활용 사례가 증가하고 있다. 이러한 변화에 따라 딥러닝 기술도 고성능 GPU 환경을 가정한 구조만이 아니라 엣지 단말과 같이 경량화된 딥러닝 구조에 대한 연구도 진행되고 있다. 경량화를 목적으로 두는 엣지 단말을 위한 딥러닝 기술은 기존의 고성능 GPU 환경을 가정한 구조에 비해 빠른 속도를 보이지만, 낮은 정확도 성능을 가지고 있어 지역 최적화 학습으로 보완한다.

횡단보도와 같은 특정 지역을 고정 감시하는 카메라에서 수행하는 객체 검출은 객체의 형태가 유사하고, 대부분 고정된 영역을 이동하기 때문에 다른 데이터보다 높은 객체 검출 정확도를 얻을 수 있지만, 학습에 필요한 데이터는 해당 지역의 영상으로만 이루어지기 때문에 DB의 총 개수에 제한이 있을 수 있고, 각 카메라가 서로 다른 학습 결과물을 도출해야 하므로 시간과 인력이 과도하게 필요한 단점이 발생한다. 이와 같은 과정에서 소모되는 자원을 절약하기 위해 지역 최적화 DB 구축 과정의 자동화가 필요하다.

본 논문에서는 이전 연구[1]를 바탕으로 영상에서 배경을 추출하여 기존 딥러닝 모델을 통해 검출된 객체를 합성해서 학습 DB를 자동으로 생성하는 도구를 개발하였고, 자동 생성된 학습 DB를 이용하여 생성된 검출모델의 성능 평가를 통해 DB 생성 도구의 유용성을 확인하였다.

Ⅱ. 본론

2.1 지역 최적화 DB 자동 생성 도구

본 논문에서 개발한 지역 최적화 DB 자동 생성 도구의 GUI 구조는 <그림 1>과 같다. 동영상 플레이어와 같은 재생에 필요한 버튼과 현재 프레임 위치를 나타내는 프로그레시브 바가 배치되어있으며, 배경생성, 배경선택, 합성이미지생성의 3가지 모드 변경 버튼이 있고, 각 기능에 필요한 옵션이 중앙에 위치한다. 객체 오검출과 미검출 상황을 보완하기 위한 수동 annotation 기능 또한 지원한다. 배경생성 모드에서는 입력영상에 대해 배경을 생성한다. 배경선택 모드에는 생성된 배경 중에서 합성 이미지 생성에 사용된 배경을 선택한다. 합성이미지생성 모드에서는 입력영상에서 검출한 객체와 생성된 배경을 합성하여 이미지를 생성하고, 검출 객체에 대한 annotation 파일도 함께 생성한다. 이때, 검출 객체는 confidence 값이 특정 임계값 이상으로 검출 신뢰도가 높은 값만 합성이미지 생성에 사용된다.

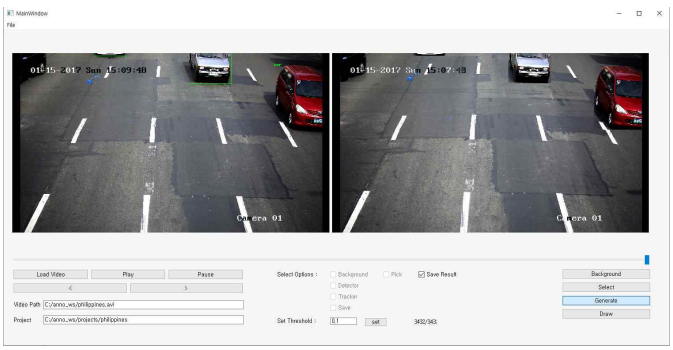


그림 1. 지역 최적화 DB 자동 생성 도구

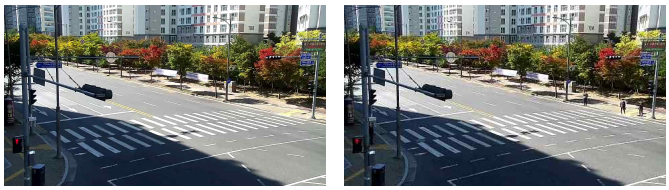
2.2 자동 생성 개요

지역 최적화 DB 자동 생성 도구는 움직이는 동영상에서 정적인 객체만

추출한 배경과 yolo 알고리즘[2]을 통해 검출한 객체들을 합성하는 과정을 거쳐 학습용 이미지가 생성된다.

동영상의 일정 구간 내 프레임들의 평균을 구하면 움직이는 객체들이 제거되고 정적인 객체만 영상에 남기 때문에 배경을 추출할 수 있으며, 정확하게 배경만 추출하기 위해서는 더 넓은 구간의 프레임 구간을 입력해야 하지만 이는 많은 메모리를 소모하므로 추출한 배경 중에서 사용자가 선택한 프레임만 합성용으로 사용된다. 이와 같은 방법으로 추출된 배경은 <그림 2>의 (가)와 같이 나타난다.

합성에 사용될 객체는 최초 생성에 한해 yolov4-tiny[2]의 기본 가중치 모델을 사용하며, 검출된 객체를 배경에 합성하면 <그림 2>의 (나) 같은 결과를 얻을 수 있다. 최초 생성에는 미검출 및 오검출 객체가 다수 존재하지만, 지역 최적화 DB를 구축하고 해당 DB를 기반으로 다시 학습시킨 가중치 모델을 얻는 과정을 반복하면 점차 정확도가 향상되어 비교적 짧은 시간 내에 준수한 성능의 객체 검출기가 만들어진다.



(가) 동영상으로부터 추출한 배경 (나) 검출된 객체를 합성한 결과물

그림 2. 지역 최적화 DB 생성 과정

2.3 성능 평가

자동 DB 생성 도구를 이용하여 생성된 학습 DB와 사람이 직접 생성한 학습 DB의 객체 검출기의 성능을 비교하였다. 횡단보도를 바라보는 고정 지역 감시 카메라로부터 얻은 1596프레임의 영상에 대해 실험을 수행하였다. 사람이 직접 생성한 DB를 통해 학습한 모델과 지역 최적화 DB를 이용해 반복 학습한 모델 중 성능 지표가 높게 나온 것을 비교하였다. 직접 생성한 DB를 학습시키는데 사용된 시간은 약 13시간이며, DB 자동 생성은 3회 실시, 학습 시간은 약 3시간이 소모되었다. 성능 지표는 평균 정밀도를 이용하여 비교했다. 실험 결과 사람이 직접 생성한 DB를 통해 학습한 모델은 *person* 객체에 대해 79.65% 평균 정밀도를 기록하였으며, 자동 생성된 DB를 통해 학습한 모델은 2회차에서 78.88%, 3회차에서 78.55%로 사람이 직접 생성한 DB를 이용한 학습과 유사한 성능을 보였다.

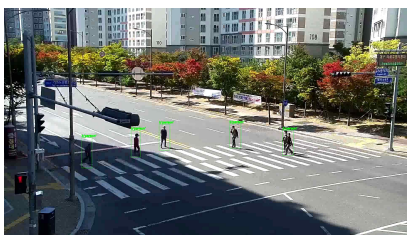
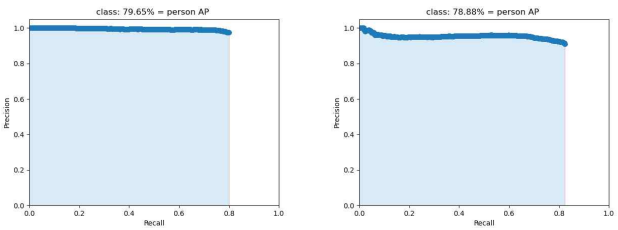


그림 3. 보행자 검출 결과



(가) 사람이 생성한 학습 DB (나) 자동 생성된 학습 DB

그림 4. 지역 최적화 객체 검출기 precision, recall 그래프 비교 (보행자)

두 번째 실험은 차량이 이동하는 사거리를 고정 감시하는 카메라로부터 얻은 2880프레임으로 구성된 영상을 이용했다. 실험 환경은 이전과 동일하며, 직접 생성한 DB를 학습시키는데 사용한 시간은 약 20시간, 자동 생성 DB를 이용한 학습은 약 3시간이 소모되었다. 실험 결과 직접 생성한 DB로 학습한 모델은 *car* 객체에 대해 98.44% 평균 정밀도를 기록하였으며, 자동 생성된 DB를 통해 학습한 모델은 2회차에서 97.98%로 사람이 직접 생성한 DB를 이용한 학습과 유사한 성능을 보였다.

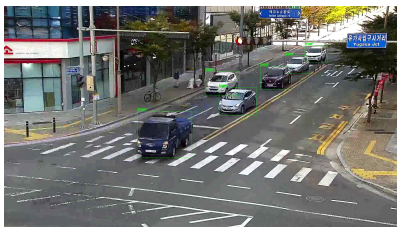
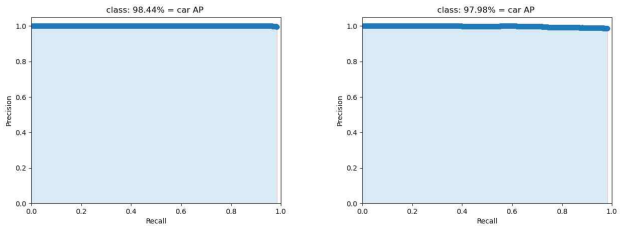


그림 5. 차량 검출 결과



(가) 사람이 생성한 학습 DB (나) 자동 생성된 학습 DB

그림 6. 지역 최적화 객체 검출기 precision, recall 그래프 비교 (차량)

Ⅲ. 결론

본 논문에서는 고정지역 감시 엣지 AI 단말을 효율적으로 학습시키는 방법으로 영상에서 추출한 배경에 검출한 동적 객체를 합성하는 자동 DB 생성 도구를 구현하였다. 자동 DB 생성 도구를 이용하면 기존에 학습된 객체 검출기를 이용하여 새로운 DB를 생성, 재학습하는 방식으로 사람이 직접 DB를 구축하는 것보다 수 배 빠르게 작업을 완료할 수 있으며, 그 결과물의 성능 지표도 큰 차이를 보이지 않음을 확인하였다.

향후 연구는 자동 생성된 DB로 사용자가 직접 yolov4-tiny 네트워크를 학습시키는 단계를 개선하여 프로그램에서 자동으로 학습에 필요한 데이터를 입력받고 지정 횟수만큼 반복 학습시키는 기능을 도입하여 지역 최적화 학습의 완전 자동화를 목표로 한다.

ACKNOWLEDGMENT

본 연구 논문은 한국전자통신연구원 연구운영지원사업의 일환으로 수행되었음. [21ZD1100, 대경권 지역산업 기반 ICT 융합기술 고도화 지원사업]]

참 고 문 헌

[1] J. Lee and J. Baek, “An Automatic Database Generation Algorithm for Local Optimization of CNN Object Detector for Edge Devices,” Proc. of IEEE International Conference on Consumer Electronics Asia 2020 (ICCE-Asia 2020), 2020.

[2] A. Bochkovskiy, C. Wang, H. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” arXiv:2004.10934[cs.CV].

이기종 공작기계 상태진단을 위한 제조 및 센서 데이터 통합 관리에 관한 연구

이정환, 이문영, 윤성재, 나중찬, 이성희

한국전자통신연구원

{jeonghwan.lee, munyounglee, sj.yoon, njc, slee0003}@etri.re.kr

A Study on Integrated Management of Manufacturing and Sensor Data for Heterogeneous Machine Tool Status Diagnosis

Jeong-Hwan Lee, Munyoung Lee, Sungjae Yoon, Jungchan Na and Seong-hee Lee

Electronic and Telecommunications Research Institute

요 약

현재 공작기계 상태진단 기술은 제조사에서 제공하는 제조 데이터로만 공정/장비의 이상을 진단하고 있으며, 제조 데이터는 제조사 자체 프로토콜을 사용하는 비개방형 구조로 되어 있다. 이에, 본 논문에서는 인공지능을 활용하여 이기종 공작기계 핵심부품 상태진단을 수행하기 위하여, 공작기계 제조사에서 제공하는 제조 데이터와 공작기계의 환경과 고장 데이터를 수집하기 위한 목적으로 설치된 센서에서 제공하는 센서 데이터를 수집하기 위한 OPC-UA 개방형 네트워크 구조와 데이터 수집 방법과 통합 데이터 구조에 대해 제안하고자 한다.

Ⅰ. 서 론

정보통신기술(Information & Communications Technology, ICT)의 발달과 함께 제조 분야의 스마트화는 가속화되고 있다. 장비에 연결된 각종 센서와 통신장비를 통해 실시간으로 현장 데이터를 수집하여 장비 상태 및 공정 과정을 모니터링할 뿐만 아니라, 인공지능 기술을 접목하여 장비 상태 진단과 예지 보전도 수행하고 있으며 이러한 체계적인 분석과 관리를 통해 생산성을 높이고 있다. CNC(Computerized Numerical Control) 공작기계의 핵심 기술인 CNC 컨트롤러 또한 사물인터넷과 인공지능 기술 등이 결합된 첨단기술 복합체로 성장하고 있다. CNC 컨트롤러 시장의 대부분을 화낙, 미쓰비시, 지멘스 등과 같은 외산 업체가 선점하고 있으며, 이들 업체들은 자신들만의 특정 CNC 컨트롤러를 사용함으로써 자신들만의 시장을 만들어가고 있다. 이러한 상황에서 공작기계는 제조사에서 제공하는 인터페이스와 자체 프로토콜을 채용하여 사용하고 있어, 상태 진단을 위한 데이터 수집도 제조사의 도움없이는 쉽지 않다. OPC-UA(OPC Unified Architecture)는 산업용 자동화 애플리케이션을 위한 통신 프로토콜로, 유연하고 완벽하게 독립적으로 동작한다. 인더스트리얼 4.0의 구현을 위한 이상적인 통신 프로토콜로 간주되고 있다.

본 논문에서는 공작기계 화낙과 미쓰비시 공작기계 장비에서 제공하는 제조 데이터를 OPC-UA 개방형 산업 제어 프로토콜을 이용하여 수집하고, 공작기계의 환경과 고장 데이터를 수집하기 위한 목적으로 추가적으로 설치된 센서에서 제공하는 센서 데이터를 수집하기 위해 데이터 수집 장치(Data AcQuisition, 이하 DAQ) 와의 연동 방법과 통합 데이터 구조에 대해 제안하고자 한다.

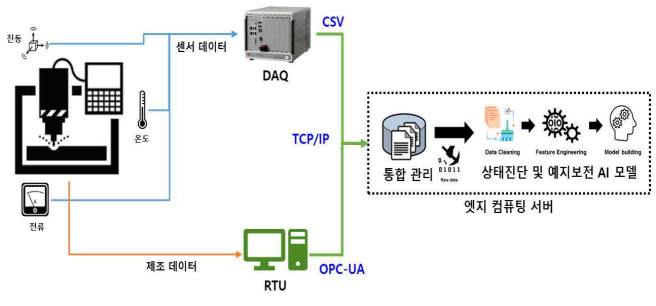


그림 1. 공작기계 개방형 데이터 수집 네트워크 개념도

Ⅱ. 제조 및 센서 데이터 수집

본 논문에서는 화낙 F0i-MD 와 미쓰비시 FCA70P-2AVU 컨트롤러가 내장된 CNC 공작기계를 테스트벤치로 구축하였다. 공작기계 상태진단을 목적으로 사용하기 위한 공작기계에서 제공하는 제조 데이터와 스핀들, 이송계와 같은 공작기계 핵심부품의 상태를 진단하기 위해 사용되는 센서들에 대해서 살펴보고자 한다.

가.제조데이터

화낙 F0i-MD 컨트롤러에서 수집하는 제조 데이터 개수는 34개이며, 미쓰비시 FCA70P-2AVU 컨트롤러에서 수집하는 개수는 340개이다. [그림 1] 과 같이 원격 단말 장치(Remote Terminal Unit, 이하 RTU)는 각 컨트롤러 인터페이스로부터 수신되는 제조 데이터를 수집하여 이를 OPC-UA 개방형 산업 제어 프로토콜로 변환하는 역할을 수행한다. 또한, RTU에서 수집된 제조데이터를 통합 관리하기 위한 엣지 컴퓨팅 서버는 50ms마다 (1초에 20번) RTU에게 OPC-UA 프로토콜을 이용하여 [표 1]과 같이 상태진단에 사용될 제조 데이터를 요청한다.

표 1. 상태 진단용 화낙 컨트롤러 제조데이터 예시

Idx	Data Type	DisplayName	Description
0	Int32	axisdata_3-0-0_load	부하 (실제 사용전류/정격전류 * 100)
1	Int32	axisdata_3-1-0_speed	주축 이송 속도
2	Int32	axisdata_current_x	부하 전류량
3	Int32	axisdata_load_x	서보모터 부하 (사용 전류 / 정격전류)
4	Int32	diagnosis_closed_loop_err_x	Closed loop error
5	Int32	diagnosis_cumul_feedback_x	피드백 적산(SV)
6	Int32	diagnosis_cumul_n_pulse_x	지령 펄스 적산(NC)
7	Int32	diagnosis_cumul_sv_pulse_x	피드백 적산(SV)
8	Int16	diagnosis_eff_current_x	유효 전류량
9	Int16	diagnosis_exclt_phase_curr_x	Reference current value
10	Float	diagnosis_pos_error_x	서브보트 위치 편차
11	Short	diagnosis_pulse_temp_x	펄스규격 전주
12	Int16	diagnosis_r_phase_curr_x	
13	Int32	diagnosis_ref_count_x	레퍼런스 카운트
14	Int8	diagnosis_servo_temp_x	서보모터 온도
15	Int8	diagnosis_temp	주축모터 온도
16	Int16	diagnosis_vcll_info_x	서보모터 전압
17	Int32	dynamic2_feed_x	이송 속도
18	Int32	loop gain_x	위치 루프 게인

공작 기계의 위치정보, 고장이력 등도 HMI(Human Machine Interface)에 연결되어 상태를 분석할 수 있다.[1]

나.센서 데이터

스마트 공장 시스템에서는 기계, 장비, 설비, 부품에 센서를 설치하여 제어하기 위한 유연한 통신 환경이 구축되어야 한다.[2] 이에, 공작기계 스핀들, 이송계 등의 핵심 부품 상태 진단을 위해 추가적으로 진동, 전류, 온도, 소음 센서들을 부착하였다. 진동 센서는 스핀들의 주축및 베어링 헤드와 이송계의 X,Y,Z 축에 5개를 부착하여 진동 특성을 측정하였으며, 온도 센서는 스핀들과 공작기계 내부 및 외부에 스핀들, 이송계, 구조물의 열변위 측정을 목적으로 3개를 설치하였다. 소음 센서는 공작기계 내부에 1개만 설치하여 가공시 발생하는 소음을 감지하기 위한 목적으로 측정하였고, 전류 센서는 스핀들과 이송계의 모터 소비 전류 측정을 통해 가공 부하를 확인하기 위한 목적으로 6개를 설치하였다.

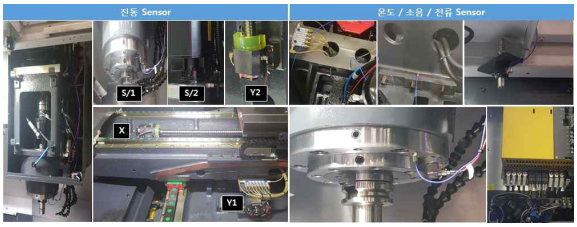


그림 2. 핵심 부품 상태 진단을 위한 센서 부착 예시

데이터 수집 장치(DAQ)와 센서들과는 BNC 케이블로 상호 연결되어 있으며, DAQ 와 엣지 컴퓨팅 서버는 이더넷으로 연결되어 있다. DAQ는 초당 20KHz 단위로 진동, 소음, 전류 데이터를 수집하고, 온도 데이터는 1Hz 단위로 데이터를 수집한다. DAQ는 수집된 센서 데이터를 내부 스토리지에 저장하고 이를 엣지 컴퓨팅 서버로 전송한다. 엣지 컴퓨팅 서버로 전송하는 방법은 센서 데이터가 수집되는대로 TCP/IP 기반 스트림으로 전송 방식과 수집된 센서 데이터를 1초 단위로 CSV(comma-separated values) 형식 파일로 전송하는 방식으로 설계가 가능하다. 본 연구에서는 DAQ 가 윈도우 기반의 운영 체제로 되어있기 때문에 TCP/IP 기반의 데이터 스트림으로 센서 데이터를 전송하기에 다소 성능에 무리가 있어, CSV 형식의 파일 형태로 전송하는 방식을 채용하였다.

다.통합데이터구조

공작기계에서 전송하는 제조 데이터와 추가적으로 부착된 센서에서 전송하는 센서 데이터는 RTU와 DAQ를 통해 엣지 컴퓨팅 서버로 전달된다. 엣지 컴퓨팅 서버내 데이터 수집 모듈은 수집된 제조 데이터와 센서 데이터를 1초 단위로 In-memory 방식의 통합 데이터 형태로 생성 및 관리한다. 수집된 센서 데이터와 제조 데이터는 공작기계 상태진단 인공지능 모델의 원시 데이터로 사용된다.

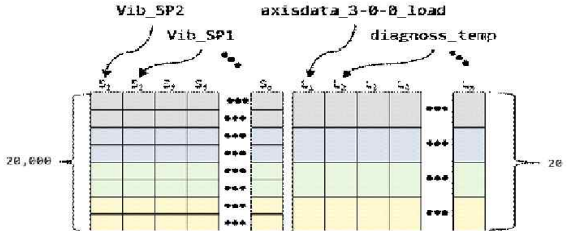


그림 3. 인메모리 통합데이터 구조

[그림 3] 의 통합데이터 구조를 살펴보면, 센서의 개수는 S_1, S_2,...S_n 의 n 개를 가지므로 본 연구에서는 6개의 센서가 부착되어 n값은 6이 된다. 샘플링 레이트가 가장 큰 수집 센서를 기준으로 하여 초당 수집되는 데이터의 개수가 결정된다. 본 논문에서는 센서 데이터 수집의 최대 샘플링은 20KHz 이므로 센서 데이터의 초당 수집 개수는 20000개이다. 상태 진단을 위해 사용하는 화낙 컨트롤러의 제조 데이터의 개수는 L_1, L_2,...L_n 의 m개를 가지므로 [표 1]과 같이 m값은 19개이며, 제조 데이터의 샘플링 레이트는 20Hz이므로 제조 데이터의 초당 데이터 수집 개수는 20개이다. 공작기계의 상태 진단 및 예지 보전을 마이크로 세컨드(us), 밀리 세컨드(ms) 내에 실시간으로 처리해야 하는 사항이 아니므로, 제조 데이터와 센서 데이터간의 데이터 동기화는 1초 단위로 생성된 데이터를 기반 동기화 하였다. 이렇게 생성된 통합 데이터는 상태 진단 인공지능 모델이 사용하기 위하여 데이터 전처리 모듈로 전달된다. 데이터 전처리 과정은 인공지능 모델 성능에 중요한 역할을 담당한다. 다양한 종류의 센서에서 수집한 원시 데이터에서 불필요한 부분을 제거하거나 인공지능 모델에 적합한 형태로 데이터의 형태를 변환하는 등의 과정을 의미한다.[3] 센서 데이터 1개(50us 분량)에 동일한 로그 데이터 1000개로 변환하여 단일 데이터셋으로 구성한다.

Ⅲ. 결론

본 논문에서는 화낙, 미쯔비시와 같은 이기종의 외산 공작기계 컨트롤러에서 제공하는 제조 데이터를 OPC-UA 프로토콜을 이용하여 수집하는 방법과 공작기계 핵심부품의 상태 정보와 환경을 모니터링하기 위해 부착된 다양한 종류의 센서 데이터를 TCP/IP 기반 CSV 파일 전송 방식으로 데이터를 수집하는 방안을 제안하였다. 또한, 수집된 제조 데이터와 센서 데이터를 상태 진단 인공지능 모델의 원시 데이터로 사용하기 위한 통합 데이터 구조에 대해서도 살펴보았다.

차년도 연구에서는 본 연구를 통해 개발된 연구 결과물을 개방형 임베디드 하드웨어 플랫폼에 적용하여, 이기종 CNC 공작기계를 운영하는 현장 환경에서 제조 데이터와 센서 데이터를 수집하고 통합 데이터를 생성 및 관리하는 기능 안정화를 실증할 계획이다. 또한, 공작기계 상태 진단 인공지능 모델과 연동하여 성능 개선 관련 연구도 추가 병행할 예정이다.

ACKNOWLEDGMENT

이 연구는 2021년도 산업통상자원부 및 산업기술평가관리원(KEIT) 연구비 지원에 의한 연구임(No. 20007891).

참 고 문 헌

[1] D.B. Choi, S.H. Kim, S.I. Jeong, "System Configuration of the Machine Tools Monitoring and Data Processing," The Korean Society of Manufacturing Technology Engineers, pp. 95, May. 2018.

[2] Young-woon Lee, Young-ju Choi, ong-Hyeok Lee, Byung-Gyu Kim, Seung-Woo Lee, Jong-Kweon Park, " Development of Distributed Smart Data Monitoring System for Heterogeneous Manufacturing Machines Operation," Journal of Digital Contents Society, pp. 1175-1182, Oct. 2017.

[3] Sungjae Yoon, Munyoung Lee, Jeonghwan Lee, Seong-hee Lee and Jungchan Na, " Fault Diagnosis Using Artificial Intelligence for the Spindle of Machine Tools," The Korean Society of Mechanical Engineers, pp. 401-408, February 2021.

멀티태스크 러닝을 이용한 심층신경망 기반 기계 소리 이상 진단

최지훈, 최정우*
한국과학기술원, *한국과학기술원
wlgns2533@kaist.ac.kr, *khepera@kaist.ac.kr

Deep Learning based Machine Sound Anomaly Detection usingMulti-TaskLearning

Choi Ji Hoon, Choi Jung Woo*
KAIST, *KAIST

요 약

소리에 기반하여 기계류의 이상 탐지를 수행하는 딥러닝 기법은, 기계류 작동 시에 발생할 수 있는 다양한 비정상 데이터를 사전에 획득하기 어려운 문제를 안고 있다. 때문에 정상 기계의 소리 데이터만을 가지고 사전 정의된 태스크를 수행하여 정상 데이터의 특징을 학습하는 자기 지도학습 방법이 널리 사용되고 있다. 본 논문은 기존 모델들이 특정한 단일 태스크로 자기 지도학습을 수행하면서 발생하는 한계를 극복하기 위하여 과거의 소리 데이터로 미래를 예측하는 복원 태스크와 서로 다른 기계 종류를 분류하는 분류 태스크를 동시에 수행하는 새로운 멀티태스크 모델을 제안한다. 제안하는 멀티태스크 모델은 WaveNet 기반의 특징 추출부를 소리 재현부와 분류부가 공유하는 구조를 가지고 있으며, 기계류 이상 탐지에 있어 기존의 단일 태스크 모델에 비해 더 높은 성능을 보여준다.

Ⅰ. 서 론

소리에 기반하여 기계류의 이상 유무를 탐지하는 딥러닝 기법은 산업의 여러 분야에서 사용되고 있다[1]. 이 과정에서 가장 큰 문제점은 기계의 정상 소리는 상시 측정이 가능해 데이터 확보가 쉽지만, 비정상 소리는 드물게 발생하고 그 발생 양상을 예측하기 어렵기 때문에 훈련 데이터를 수집하기 어렵다는 것이다. 이 때문에, 정상 소리 데이터만을 이용한 자기 지도학습 방법이 연구되고 있다.

기계 소리 이상 탐지를 위한 여러 모델 학습 방법 중에는, 과거의 정상 소리로 미래의 정상 소리를 예측 혹은 복원하는 자기 회귀 과정을 통해 정상 소리의 특징을 학습하는 복원 학습 방법과 기계 종류를 분류해 기계 간 소리의 특징을 학습하는 분류 학습 방법이 있다. 자기 회귀 태스크를 통해 훈련된 모델은 정상 기계 소리에 대해서는 학습한 기계의 특징을 이용해 잘 복원하게 되어 복원 오차가 작지만, 비정상 기계 소리에 대해서는 정상 기계 소리의 특징이 없으므로 복원을 잘하지 못해 큰 복원 오차를 나타내게 된다. 반면, 기계 종류 분류 태스크는 서로 다른 모델의 기계 데이터를 분류하도록 학습하며, 학습 과정에서 기계 종류별 정상 소리의 특징을 학습하게 된다. 학습된 모델은 정상 기계 소리에 대해서는 기계 종류를 예측할 확률이 높게 나오지만, 비정상 소리에 대해서는 분류 예측 확률이 낮게 나오게 된다.

본 논문에서는 기존에 존재하는 기계 소리 이상 탐지 모델의 성능을 향상하기 위해, 복원 학습 방법과 기계 id 분류 학습 방법을 동시에 적용한 새로운 멀티태스크 모델을 제안한다.

Ⅱ. 본론

본 논문에서 제안하는 새로운 멀티태스크 모델은 입력 신호의 공통적인 특징을 추출하는 부분인 특징 추출 모델, 추출한 특징을 이용해 신호를 복원하게 되는 복원 모델, 특징을 이용해 기계 id를 분류하는 분류 모델 세 부분으로 나눌 수 있다. 모델의 전체 구조는 그림 1에 나타냈다.

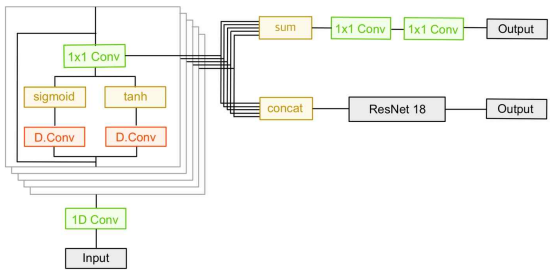


그림 1. 멀티태스크 모델의 구조

먼저 모델의 특징 추출부는 WaveNet[2]의 구조를 사용했다. WaveNet은 End-to-End로 음성 데이터를 생성하는 것을 목적으로 사용되는 생성 모델이지만, 이를 변형한 스펙트로그램 미래 예측 모델이 제안된 바 있다[3]. WaveNet의 특징인 causal convolution을 이용해 일정 시간의 스펙트로그램 데이터를 입력으로 받으면 다음 시간 프레임의 주파수 bin 값을 예측하게 된다. 본 연구에서는 WaveNet의 residual block 부분을 분리하여, 스펙트로그램의 복원 과정과 기계 소리 분류 과정에 사용될 공통 특징을 추출하는 추출부로 사용한다. 특징으로 뽑아낸 residual block의 출력은 각각 복원 모델과 분류 모델의 입력으로 사용되어 스펙트로그램 복원과 기계 소리 분류를 동시에 수행한다. 복원 모델은 1x1 convolution을 두 번 사용했고, 분류 모델은 18-layer ResNet[4]을 사용했다. 모델의 손실함수는 복원 모델의 손실함수와 분류 모델의 손실함수 합으로 나타냈다.

$$L_{\text{total}} = L_{\text{recon}} + L_{\text{class}} \qquad (1)$$

식 (1)에서의 L_{recon} 은 입력과 출력 간 스펙트로그램의 L_2 값을 사용했고, L_{class} 은 categorical cross entropy를 사용했다.

모델 학습을 위한 데이터 세트는 DCASE2020 Challenge Task2에서 제공한 기계 소리를 사용했다[5]. ToyCar, ToyConveyor, fan, slider, pump, valve 총 6개의 기계 종류가 있고, 각 기계 종류별로 다른 기계 번호를 나타내는 id로 세분되어있다. 10초의 기계 소리 음원 파일을 STFT를 통

해 스펙트로그램을 구했다. 이때 음원의 sampling rate는 16000, window 크기는 2048, hop length는 512, 주파수 bin 개수는 128개로 설정했다. 전처리 과정으로 기계 종류마다 정규화를 진행한 후 모델의 입력으로 사용했다. 데이터의 다양성을 위해 데이터 증강 기법도 사용했다. 서로 다른 id를 갖는 같은 종류의 두 기계의 스펙트로그램을 일정 비율로 섞어 새로운 이미지를 생성한다. 이때 복원 모델은 새로운 이미지 자체를 복원하고, 분류 모델은 새로 생성된 이미지가 각 기계 id에서 얼만큼의 비율로 섞였는지를 예측하게 된다. 데이터의 배치 크기는 64, 학습 방법은 Adam, learning rate는 10^{-4}을 사용했고, 100 epoch까지 학습한 결과를 이용했다.

데이터의 평가 방법은 기계 종류에 따라 다른 방법을 사용한다. 전 주파수 영역에 대해 기계 소리가 관측되는 valve와 slider는 학습할 때처럼 입력과 출력 사이의 L_2값을 사용한다. 나머지 기계들은 특정 주파수 대역에서 기계 소리가 관측되므로, 입력과 출력 이미지 차이의 시간 평균을 취해 주파수 벡터를 구한 뒤, Mahalanobis distance를 계산한다[3].

$$L_{\text{recon}}^{ij} = \sqrt{(x_{ij}-m_j)^T \Sigma^{-1}{}_j (x_{ij}-m_j)} \qquad (2)$$

식 (2)에서의 x_{ij}는 기계 종류 j의 i번째 데이터로부터 얻은 주파수 bin 벡터, m_j는 학습시킨 기계 종류 j의 데이터로부터 얻은 주파수 벡터들의 평균, Σ_j는 기계 종류 j의 데이터 세트로 구한 주파수 bin 간의 공분산 행렬이다. 분류 성능에 대한 평가는 학습할 때처럼 categorical cross entropy를 사용한다. 이상 탐지 분류를 위한 이상 점수는 복원에 대한 점수와 분류에 대한 점수의 Mahalanobis distance를 계산한다.

$$\text{Score} = \sqrt{(l_{ij}-d_j)^T S^{-1}{}_j (l_{ij}-d_j)} \qquad (3)$$

식 (3)에서의 l_{ij}는 기계 종류 j의 i번째 데이터로부터 얻은 손실함수 벡터, d_j는 학습시킨 기계 종류 j로부터 얻은 손실함수 벡터들의 평균, S_j는 기계 종류 j의 데이터 세트로 구한 손실함수 간의 공분산 행렬이다.

모델의 이상 탐지 성능 평가를 위한 지표로는 ROC-AUC와 pAUC를 사용했다[5]. 성능 비교를 위해, 자기 회귀 task만 수행하는 WaveNet 모델, 분류 task만 수행하는 ResNet 모델, 그리고 새로 제안하는 멀티태스크 모델과 DCASE 2020 Challenge에서 가장 좋은 성능을 보인 모델[6]의 ROC-AUC와 pAUC를 비교하였다.

	WaveNet	ResNet	Ours	Giri[6]
ToyCar	0.82/0.65	0.88/0.87	0.90/0.86	**0.95/0.90**
Toy-Conveyor	0.67/0.54	0.55/0.55	0.70/0.57	**0.81/0.66**
fan	0.91/0.78	0.70/0.70	**0.92/0.86**	0.82/0.79
pump	**0.90**/0.78	0.77/0.77	0.89/**0.83**	0.87/0.80
slider	0.94/0.82	0.73/0.72	**0.99/0.96**	0.97/0.90
valve	0.78/0.57	0.87/0.87	**0.97/0.92**	0.97/0.91

표 1. 기계 종류별 이상 진단 결과 (ROC-AUC/pAUC)

표 1에서 볼 수 있듯이 6종의 기계 중 pump를 제외한 모든 기계에 대해 이상 진단 성능이 향상됨을 알 수 있다. 특히 valve에 대해서는 ROC-AUC 값이 0.1이나 증가한다. 표 2에서의 기계 id 별 결과를 비교해보면 자기 회귀 모델이나 분류 모델 중 한쪽의 성능이 좋지 않더라도, 멀티태스크 모델은 기계 이상 진단 성능이 높은 것을 알 수 있다.

Ⅲ. 결론

본 논문에서는 기계 소리 이상 진단을 위해 활용되고 있는 여러 딥러닝 학습 방법 중 자기 회귀 방법과 분류 방법을 동시에 학습하는 새로운 모델을 제안했다. 이상 탐지 성능 평가 지표인 ROC-AUC를 이용해 모델의 성능을 평가했을 때 하나의 작업을 수행하는 단일 모델에 비해 멀티태스크 모델이 좋은 성능을 보이는 것을 확인하였다.

기계 종류 (개수)		WaveNet	ResNet	Ours
ToyCar (7000 / 2459)	id_01	**0.798**	0.656	0.774
	id_02	0.857	**0.881**	0.857
	id_03	0.704	0.969	**0.980**
	id_04	0.923	**0.999**	0.995
ToyConveyor (6000 / 3509)	id_01	0.696	0.607	**0.732**
	id_02	0.642	0.510	**0.674**
	id_03	0.673	0.532	**0.701**
fan (6251 / 1875)	id_00	**0.800**	0.502	0.788
	id_02	0.924	0.801	**0.967**
	id_04	0.938	0.566	**0.940**
	id_06	0.962	0.919	**0.998**
pump (5766 / 856)	id_00	0.903	0.783	**0.935**
	id_02	**0.881**	0.570	0.707
	id_04	0.913	**0.996**	0.991
	id_06	0.906	0.735	**0.923**
slider (5174 / 1290)	id_00	**0.999**	0.945	0.997
	id_02	0.862	0.987	**0.991**
	id_04	0.982	0.500	**0.994**
	id_06	0.916	0.500	**0.986**
valve (5822 / 879)	id_00	0.635	0.950	**1.00**
	id_02	0.878	0.529	**0.898**
	id_04	0.885	**1.00**	0.983
	id_06	0.713	1.00	**1.00**

표 2. 기계별 데이터 개수(train / test)와 기계 id 별 ROC-AUC 결과

ACKNOWLEDGMENT

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. NRF-2020M2C9A1062710)

참 고 문 헌

[1] Yumi, O., Yoshifumi, O., Takafumi, K., Soichiro, T., and Osamu, H. "Anomaly detection of motors with feature emphasis using only normal sounds," Proc. IEEE ICASSP pp. 2800 - 2804, 2013.

[2] Oord, A. V. D., Dieleman, S., Zen H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., and Kavukcuoglu, K. "Wavenet: A generative model for raw audio," 2016, (https://arxiv.org/abs/1609.03499).

[3] Daniluk, P., Gozdziewski, M., Kapka, S., and Kosmider, M. "Ensemble of auto-encoder based and wavenet like systems for unsupervised anomaly detection," Challenge on Detection and Classification of Acoustic Scenes and Events, 2020.

[4] He, K., Zhang, X., Ren, S., and Sun, J. "Deep residual learning for image recognition." Proc. IEEE conference on computer vision and pattern recognition, 2016.

[5] Koizumi, Y., Kawaguchi, Y., Imoto, K., Nakamura, T., Nikaido, Y., Tanabe, R., ... and Harada, N. "Description and discussion on DCASE2020 challenge task2: Unsupervised anomalous sound detection for machine condition monitoring," 2020, (https://arxiv.org/pdf/2006.05822.pdf)

[6] Giri, R., Tenneti, S. V., Cheng, F., Helwani, K., Isik, U., Krishnaswamy, A. "Unsupervised Anomalous Sound Detection Using Self-Supervised Classification and Group Masked Autoencoder for Density Estimation" Challenge on Detection and Classification of Acoustic Scenes and Events, 2020.

Bi-LSTM 을 활용한 제주도 전력 수요 중요 변수 선정에 관한 연구

김진실, 전유정, 양여진, 정병진, 조완섭
충북대학교

kim94203069@gmail.com, ddbb0147@gmail.com, yangyeojin79@g.cbnu.ac.kr,
jeongb1030@gmail.com, chowansub@g.cbnu.ac.kr

A Study on the Selection of Variables for Power Demand Management in Jeju Island Using Bi-LSTM

Kim Jin Sil, Janu Yu Jung, Yang Yeo Jin, Jeong Byeong Jin, Cho Wan Sub
Chung-Buk National University

요 약

본 논문은 안정적인 전력 수요 관리에 도움이 되고자 기상정보, SNS 의 계절성 단어 언급량과 긍·부정 감성어 그리고 유동인구를 설명변수로 사용하여 예측 성능을 향상하는 변수 조합을 연구하였다. 그 결과 기상정보가 가장 유용한 변수로 확인되었으며, 유동인구 데이터를 함께 사용했을 때 예측 성능을 높이는 것으로 나타났다. SNS 의 계절성 단어 언급량 및 긍·부정 감성 변수는 예측 성능을 오히려 떨어뜨리는 요소로 작용하는 것을 확인하였다.

Ⅰ. 서 론

한국의 전력 소비량은 꾸준하게 증가하고 있으며, 계절 등 다양한 요인에 따라 그 사용량이 큰 폭으로 증감한다. 공급 예비력 유지가 정책상 이슈로 언급되면서 안정적인 전력 수급을 위한 연구가 중요해지고 있다. 현재까지 정확한 전력 수요 예측에 대한 다양한 시도가 있었으며, 기상 데이터는 대부분의 연구에서 활용되고 있다.

정희원(2016), 조세원(2019)은 기상정보를 활용한 전력수요 예측 연구를 진행하였고, 기상 정보 이외의 요인에 대한 추가 연구가 필요함을 언급하였다[1][2].

공성배(2015)는 기상 데이터와 함께 전력을 많이 소모하는 계절과 관련된 단어의 SNS 언급량을 전력 수요 예측에 대한 보정 지표로 활용할 것을 제안하였다[3].

본 논문에서는 기상정보와 SNS 에서의 계절성 단어 언급량, 게시글의 긍·부정 감성어 그리고 유동인구를 변수로 사용하여 제주도의 전력 사용량을 예측한다. 세 그룹의 변수를 조합해 전력수요를 예측하고, 가장 효과적인 방안을 탐색하였다.

Ⅱ. 본 론

2.1 데이터 셋

본 연구에서 사용된 데이터는 총 6600 행의 제주도 전력 수요 데이터이다. 자세한 설명은 다음과 같다.

(1) 한국전력거래소의 2019 년 4 월~12 월 제주도 시간별 전력 수요 데이터
(2) 기상청의 방재 기상정보
(3) 일자 별 제주도 유동인구
(4) SNS 계절 키워드 언급량 및 긍·부정 감성어 데이터

(1)의 제주도 시간별 전력 수요 데이터는 목표 변수이고, (2)~(4)는 예측에 사용할 설명변수이다. 기상정보는 기상자료개방 포털에서 강수량, 기온, 습도, 풍속 데이터를 수집하였으며, 제주도가 관광도시라는 특성을 고려하여 유동인구 데이터를 설명변수로 추가하였다.

또한 기상 정보와 전력 수요가 밀접한 관계를 보이는 것에 착안하여 계절성 및 온도 관련 키워드가 전력 수요에 영향을 미치는지 확인하기 위해 인스타그램, 트위터, 카페, 뉴스, 블로그 등 여러 SNS 에서 '덥다', '춥다' 등의 키워드에 대한 언급량과 감성 분석 데이터를 변수로 사용하였다. 해당 데이터는 관련 서비스를 제공하는 웹사이트를 통해 수집하였다.

본 연구는 어떤 설명변수들을 조합하여 가장 좋은 성능을 보이는지 확인하고, 이후 실시간 단기 전력 수요 예측에서 가장 합리적이고 신뢰도 높은 예측 결과를 제공하는 방법을 탐색하였다.

2.2. 연구방법

기상, SNS 계절성 단어 언급량 및 긍·부정 감성어, 유동인구로 구성된 설명변수를 기상, SNS, 유동인구, 기상+유동인구, 기상+SNS, SNS+유동인구, 기상+SNS+유동인구 7 가지로 조합하여 각 변수별 Bi-LSTM 모델을 만들어 비교하고 어떤 설명변수들을 조합하면 가장 좋은 성능을 보이는지 확인하였다.

Bi-LSTM 은 BRNN 과 LSTM 을 조합한 모델로, 양방향성을 갖는 LSTM 으로 정의할 수 있다. 포워드 계층, 백워드 계층 안에 각각 망각 게이트, 입력 게이트, 출력 게이트가 있어 오래된 과거의 데이터를 망각한다는 BRNN 의 단점과 결과 모델이 직전 패턴을 기반으로 수렴하는 경향을 가지는 LSTM 의 한계를 모두 해결한다[5].

고상준 외(2018)는 LSTM 과 양방향 LSTM 을 비교하며 Bi-LSTM 의 예측 성능이 더 우수함을 밝혔다[4].

본 논문에서 사용한 데이터는 1 년 이내의 기간에 대한 데이터로, 여러 계절을 한 번만 포함하는 전력 데이터이기 때문에 예측에 있어 단기적인 과거 패턴에 과도하게 종속되는 것을 방지하고자 Bi-LSTM 알고리즘을 선택하

였다.

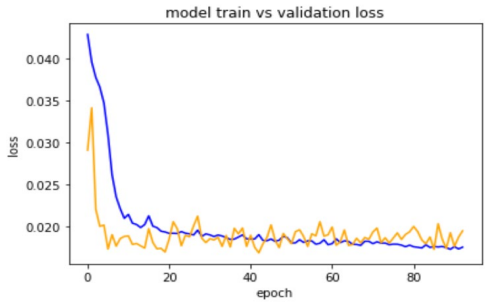
2.3 실험 결과

사용 변수에 따른 Bi-LSTM 예측 성능은 <표 1>과 같다.

변수	RMSE	MAPE
기상	0.342	46.387
SNS	0.382	70.947
유동인구	0.382	70.812
기상 + 유동인구	0.335	42.746
기상 + SNS	0.465	91.035
SNS + 유동인구	0.407	74.063
기상 + SNS + 유동인구	0.436	72.624

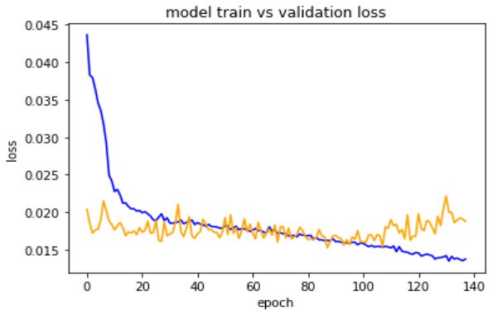
<표 1> 실험 결과 예측 성능

단독으로 사용했을 때 가장 성능이 높은 변수 집합은 기상 데이터로 RMSE 는 0.342, MAPE 값은 46.387 로 나타났다. 결과는 아래 <그림 1>과 같다.



<그림 1> 기상변수만 적용한 훈련-검증 오차 history

가장 높은 성능을 보인 변수 집합은 기상과 유동인구의 집합으로 RMSE 는 0.335, MAPE 는 42.746 으로 기상 정보를 단독으로 사용했을 때보다 예측 성능이 개선된 것을 확인할 수 있다.



<그림 2> 기상, 유동인구 변수를 적용한 훈련-검증 오차 history

하지만 SNS 데이터는 예측 성능을 떨어뜨리는 요인으로 작용하는 것이 관찰되었다. 이는 데이터 수집에서 있었던 한계로 인한 것으로 추측된다.

본 논문의 목적이었던 SNS 데이터를 포함하는 경우에 모든 결과에서 설명력이 떨어지는 현상을 볼 수 있었다. SNS 변수를 포함한 모델의 성능이 떨어진 이유는 다음과 같은 한계점으로 인한 것으로 분석된다.

첫째, 데이터의 시점이 다른 채널을 함께 사용한 점이다. 트위터가 비교적 실시간성을 띠고 있다고 하나 과거 경험에 대한 후기 및 기록 특성 이 강한 다른 채널과의 혼용으로 SNS 유저의 현재 상태에 대한 실시간 반영이 불가하여 전력 소비 예측에 활용이 어렵다. 일별 데이터 및 합계로 추출되어 수집 단어의 의미에 대한 정확성이 반영되기 어렵다.

둘째, 전력 수요 데이터와 기상 및 유동인구 데이터는 제주도 데이터인 반면, SNS 를 통해 수집한 데이터는 사용자의 현재 위치 특성을 반영할 수 없다. 따라서 SNS 데이터를 이용하여 전력 수요의 보정 지표로 활용하기 위해서는 위치기반 서비스를 이용하여 위치 특성을 반영할 수 있는 데이터를 수집해야 할 것으로 보인다.

Ⅲ. 결론

본 논문에서는 제주 전력수요 예측의 정확성을 높이기 위하여 SNS 계절성 단어 언급량 및 긍·부정 감성어 데이터와 유동인구 데이터를 이용하여 전력 수요를 예측하고 성능을 살펴보았다. 실험 결과 기상정보와 유동인구 변수를 활용했을 때 가장 성능이 좋은 것으로 나타났다. 유동인구는 단독으로는 유용하지 않았으나 기상과 함께 사용했을 때 모델의 설명력을 높였다. SNS 데이터는 오히려 성능 하락의 요인으로 작용하여 부적절한 설명변수로 판단되었다.

연구 결과 제주의 전력 수요를 예측하기 위해서는 제주의 기상 정보와 함께 유동인구 예측 보정 지표로 활용하는 것이 전력 수요 예측의 정확성을 개선하는데 도움을 주는 것으로 나타났다. 전력 수요의 예측에 있어 도시의 특성을 반영하는 데이터를 보정지표로 사용하는 것은 도시의 전력 수요 예측 모델의 예측력 향상에 도움이 될 것으로 보인다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 대학 ICT 연구센터지원사업의 연구결과로 수행되었음(IITP-2018-0-01396).

참 고 문 헌

[1] 정희원, 구본희, 차준민. 2016. 제주계통의 기온변화 민감도를 반영한 주말 전력수요예측. 전기학회논문지 65: 718-723.

[2] 조세원, 박래준, 송경빈. 2019. 기상에 대한 동네예보를 활용한 제주도의 특수일 전력수요예측알고리즘. 조명·전기설비학회논문지 33: 1-8.

[3] 공성배, 장민석, 주성관. 2015. 빅데이터 분석을 이용한 단기전력 수요예측. 대한전기학회 학술대회 논문집: 144-145.

[4] 고상준, 윤호영 and 신동명. (2018). 양방향 LSTM을 활용한 전력수요 데이터 예측 기법 연구. 한국소프트웨어감정평가학회 논문지, 14(1), 33-40.

[5] 정호철, 선영규, 이동구, 김수현, 황유민, 심이삭, 오상근, 송승호, 김진영. "에너지인터넷에서 1D-CNN 과 양방향 LSTM 을 이용한 에너지 수요예측." 전기전자학회논문지 23.1 (2019): 134-142.

CCTV 영상 기반 신경망을 이용한 유기물 탐지

박민성*, 손창대. 최성진, 이현지

*(주)마크애니

*parkms@markany.com, cdson@markany.com, choisj@markany.com, hjlee@markany.com

Abandoned Object Detection using Neural Network in Video-Surveillance

Park Min Seong*, Son Chang Dae, Choi Sung Jin, Lee Hyun Ji

*MarkAny.

요 약

본 논문은 합성곱 신경망을 이용하여 CCTV 영상 환경에서의 유기물을 탐지하는 알고리즘을 제안한다. 이 알고리즘은 입력 영상에서 합성곱 신경망을 통해 물체를 탐지하고, 탐지된 물체의 정보를 이용하여 유기물을 판별하는 방법으로 설계된다. 물체 탐지를 위한 합성곱 신경망은 AI Hub 데이터셋을 통해 학습되었다. 제안하는 알고리즘은 AVSS 데이터셋을 통해 기존 유기 알고리즘에 비해 높은 성능을 보인다.

Ⅰ. 서 론

유기물이란 일정 시간 동안 움직임이 없는 물체 혹은 물체의 소유자와의 거리가 가까이 있지 않은 물체로 정의된다[1]. 유기물 탐지는 매장에서 고객이 물건을 두고 갔을 때 이를 발견하여 되찾아주거나 쓰레기 무단투기 금지 장소에서 무단 투기자를 찾아내는 등 일상생활에서 유용하게 사용될 수 있다. 뿐만 아니라 테러가 빈번하게 일어나는 도시에서 유기물은 폭탄, 화학물 등 위험 물질의 후보가 될 수 있기 때문에 유기물 탐지는 매우 중요하다. 과거에는 관제 요원이 수많은 CCTV 영상을 직접 감시했지만, 최근 딥러닝의 발전으로 딥러닝을 이용한 유기물 관제 시스템이 유기물을 탐지하는데 도움을 줄 수 있다. 따라서 본 논문은 CNN을 이용하여 CCTV 영상 내 유기물을 탐지하는 알고리즘을 제안한다.

Ⅱ. 본 론

유기물을 탐지하기 위해 먼저 사람과 유기물 후보가 될 수 있는 물체를 탐지해야 한다. 이를 위해 본 논문은 합성곱 신경망을 활용한 방법을 제안한다. 합성곱 신경망을 학습하기 위한 데이터는 한국지능정보사회진흥원이 운영하는 AI Hub[2]에서 제공하는 유기 관련 영상을 사용한다.

합성곱 신경망 학습을 위해 AI Hub 유기 데이터셋[3]에서 초당 2초의 이미지를 추출하여 학습데이터로 사용한다.. 본 논문에서 탐지하고자 하는 유기물 후보는 캐리어와 가방이므로, 각 이미지는 사람, 캐리어, 가방 클래스와 바운딩 박스로 레이블링 된다. 추가적으로 물체와 사람 사이의 관계에 대한 정보를 얻기 위해 물체를 들고 있는 사람 클래스를 추가한다. 여러 이미지 내 정지된 물체가 중복되는 경우는 1개의 이미지를 제외하고 마스킹 된다. 라벨링 된 이미지와 구축한 학습 데이터셋 정보는 각각 그림 1과 표 1과 같다.

본 논문에서는 실시간 CCTV 환경에서 제안하는 알고리즘을 적용하기 위해 YOLO[4]를 객체 탐지 모델로 사용한다. YOLO[4]의 입력은 1920x1080 해상도의 CCTV 이미지를 608x608로 다운 샘플링된 이미지이고, 출력은 객체에 해당하는 클래스와 바운딩 박스의 좌표 (x,y,w,h)이다.



그림 1. 라벨링 된 데이터

클래스			총 이미지 수
번호	종류	개수	
0	사람	6,011	6,113
1	물체를 들고 있는 사람	2,307	
2	캐리어	1,897	
3	가방	1,566	

표 1. 학습 데이터셋

학습을 위한 하이퍼파라미터로 학습률(learning rate)은 0.0003, 배치 사이즈는 2, 이터레이션(iteration)은 25,000을 사용한다. 스케줄러 기법은 StepLR을 사용하고, 10,000, 13,000, 15,000, 18,000, 21,000 이터레이션 마다 0.1의 gamma 값을 설정한다. 학습 데이터 셋의 클래스 불균형을 고려하기 위해 각 클래스에서 얻어지는 loss마다 0.094, 0.245, 0.300, 0.361의 비율로 곱해서 최종적인 loss를 계산한다.

YOLO[4] 성능을 확인하기 위해 545장의 이미지를 평가 데이터셋으로 이용된다. 그림 2는 YOLO의 학습/평가 loss의 그래프이다. YOLO의 성능은 평가 데이터셋에 대하여 각 클래스마다 AP(Average Precision) 값

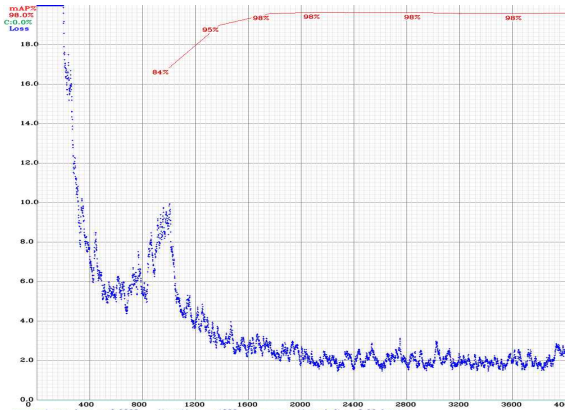


그림 2. 학습/평가 loss 그래프

이 99%, 99%,100.00%, 98%을 보이고, 99%의 mAP(Mean Average Precision) 값을 얻었다.

본 논문은 클래스 0, 2, 3을 이용한 알고리즘 Ⅰ과 클래스 0, 1, 2, 3을 이용한 알고리즘 Ⅱ를 제안한다. 알고리즘 Ⅰ은 입력 이미지에 대해 객체 탐지 모델이 객체를 찾았다면 그 객체가 유기 후보물인지 판별하고, 유기 후보물인 경우 이전 이미지에서의 유기물 후보와 현재 이미지에서의 유기물 후보의 IOU를 계산하여 물체가 멈춰 있는지 판별한다. 또한 사람과 유기 후보물 사이의 거리를 계산하여 주인이 있는 물체인지 아닌지 판별한다. 거리를 계산하기 위해 사람의 중심 좌표 또는 두 발의 중점 좌표와 물체의 중심 좌표가 사용된다. 최종적으로 알고리즘 Ⅰ은 물체가 일정 프레임 이상 연속적으로 멈춰있고, 주인이 없는 경우일 때 유기로 판별한다.

알고리즘 Ⅱ은 입력 이미지에 대해 객체 탐지 모델이 객체를 찾고, 찾은 객체와 이전 프레임의 유기 상태에 따라서 나뉘어서 동작한다. 알고리즘 Ⅱ에서 물체가 멈춰있는지 판별하는 방법은 알고리즘 Ⅰ과 같다. 주인이 있는 물체인지 아닌지 판별하기 위해서는 유기물 후보의 바운딩 박스가 물체를 들고 있는 사람의 바운딩 박스 내에 있는지를 확인한다. 이때 물체를 들고 있는 사람의 옆모습과 같이 유기물 후보가 가리워짐(occlusion)을 고려하기 위해 가리워짐 발생 시간을 계산하여 일정 시간 이상인 경우 유기로 판별한다.

제안하는 알고리즘의 성능은 AVSS[5] 데이터셋의 유기 영상을 이용하여 평가된다. 알고리즘 Ⅱ가 [6]에서 제안하는 알고리즘에 비해 높은 성능을 보인다. 제안하는 알고리즘 Ⅰ, Ⅱ의 자세한 성능은 표 2와 같다.

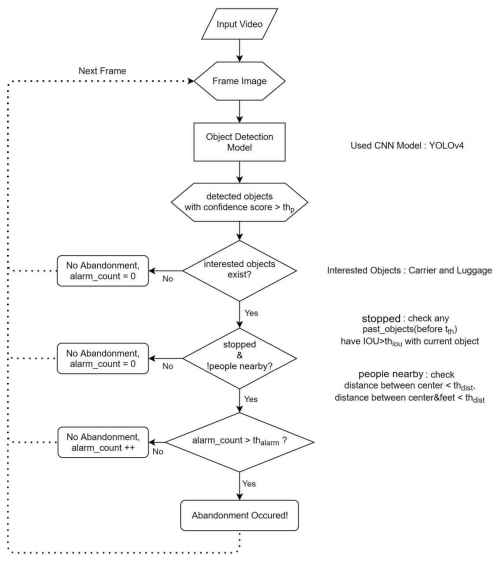


그림 3. 알고리즘 Ⅰ

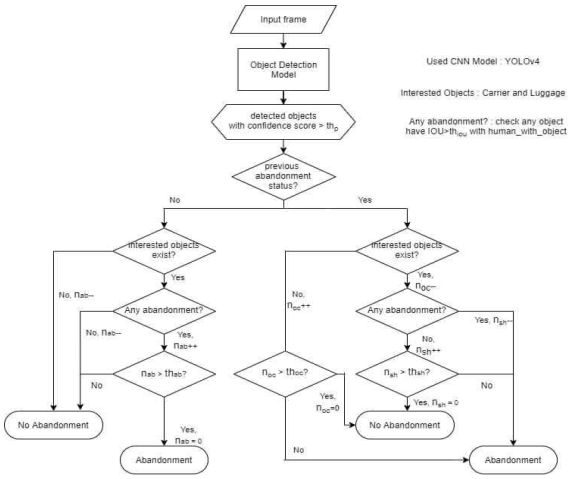


그림 4. 알고리즘 Ⅱ

	precision	recall	f1 score
SOD+CCNN [6]	97.77%	51.89%	67.80%
SOD+CCNN [6] (Generated Sample)	97.48%	66.59%	79.13%
알고리즘 Ⅰ(ours)	86.64%	80.66%	83.50%
알고리즘 Ⅱ(ours)	**98.49%**	**85.23%**	**91.38%**

표 2. AVSS[5] 데이터셋에 대한 성능 비교

Ⅲ. 결 론

본 논문에서는 CCTV 영상 기반 합성곱 신경망을 이용한 유기물 탐지 알고리즘을 제안한다. 제안한 알고리즘은 실시간 CCTV 환경의 유기물 발생 지역에서 사용될 수 있다.

ACKNOWLEDGMENT

본 연구는 2021년도 과학기술정보통신부(IITP)의 재원으로 지원을 받아 수행되었다. (No. 2021-0-00348, 2D/3D 영상 통합 분석을 이용한 클라우드 기반 무인점포 환경 대응형 영상보안시스템 개발)

참 고 문 헌

[1] Elena L., Juan C. S., Diego. O., and Jose M. M. "Abandoned Object Detection in Video-Surveillance: Survey and Comparison", MDPI, 2018.

[2] 한국지능정보사회진흥원, AI Hub (http://aihub.or.kr/).

[3] AI Hub 이상행동 CCTV 영상. (http://aihub.or.kr/aidata/139).

[4] Bochkovskiy. A., Wang, C. Y., and Liao H. Y. M., "YOLOv4: Optimal speed and accuracy of object detection." [Online]. Available: http://arxiv.org/abs/2004.10934.

[5] Bayona A., SanMiguel J. C., and Martínez J. M., "Comparative evaluation of stationary foreground object detection algorithms based on background subtraction techniques". In Proceedings of the IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 25 - 30, September 2009.

[6] Sorina S., and Radu T. I., "Real-Time Deep Learning Method for Abandoned Luggage Detection in Video". [Online] available: http://arxiv.org/abs/1803.01160.

제2회 한국 인공지능 학술대회

The 2nd Korea Artificial Intelligence Conference

세션 F

Reverse Low-High 라벨링 방법을 이용한 XGBoost 트레이딩 시스템 개발

한예찬, 박태신, 김재윤*
순천향대학교

ychan.sch@gmail.com, ptaesin@gmail.com, kimym38@sch.ac.kr*

Developing XGBoost Trading System using Reverse Low-High Labeling

Yechan Han, Taeshin Park, Jaeyun Kim*
Soonchunhyang Univ.

요 약

본 논문에서는 트레이딩 시스템 개발 시 데이터 라벨링의 중요성에 대해 논의한다. 기존의 Up-Down 라벨링 방법은 주가의 등락 정도를 파악하지 못한다는 한계점이 있다. 본 논문에서는 데이터 라벨링 시 Reverse Low-High 방법을 사용하며, KOSPI 200 과 S&P 500 주가 지수로 실증 분석을 진행한다. 실증 분석 결과 제안된 방법은 기존의 Up Down 방법보다 우수한 성과를 나타내는 것을 확인하였다.

1. 서 론

주식 투자에 대한 관심이 높아지고, 주가 데이터에 대한 접근이 쉬워지면서 머신러닝 기술을 활용한 다양한 투자 방법이 제안되고 있다. [1] 머신러닝 기술을 활용한 투자 방법은 주로 과거의 주가 데이터를 이용하여 매매 규칙을 발견하고 매매 규칙에 따라 백테스팅을 수행하여 검증하는 절차를 거친다. 주가의 상승과 하락을 예측하는 분류 문제의 경우, 이 과정에서 주가 데이터를 라벨링 하는 작업이 수행된다. 많은 연구에서는 주가 데이터 라벨링 시 Up-Down 라벨링 방법을 사용한다. [2][3] Up-Down 라벨링 방법은 t+1 시점의 주가가 t 시점의 주가보다 오르면 Up, 내리면 Down 으로 라벨링을 수행한다. 이러한 라벨링 방법은 단순히 주가의 등락만을 구분하며, 등락 정도를 파악하지 못하기 때문에 이후 머신러닝 모델을 사용한 예측 과정에서 문제가 발생할 수 있다. 본 논문에서는 데이터 라벨링 작업 시 Sezer 가 제안한 라벨링 방법(본 논문에서는 해당 라벨링 방법을 Low High 라벨링 방법이라 정의)을 변환한 Reverse Low High 라벨링 방법을 사용하며, 기술적 지표와 XGBoost 모형을 활용하여 트레이딩 시스템을 구축한다. [4]

2. 본 론

2.1. 실험 데이터

본논문에서는 KOSPI 200 과 S&P 500 일별 주가 지수 데이터를 사용하였으며, 2012 년 1 월 1 일 ~ 2017 년 12 월 31 일 기간 (6 년)의 데이터를 학습 기간으로, 2018 년 1 월 1 일 ~ 2020 년 12 월 31 일 기간 (3 년)의 데이터를 테스트 기간으로 설정하였다.

2.2. 실험 방법

- 데이터에서 기술적 지표를 추출하고 데이터 라벨링 작업을 수행하여 학습 데이터를 구축한다.
- XGBoost 모형을 사용하여 학습 데이터로 모형을 학습한다. (학습 데이터에서는 Up 과 Down 신호만 사용하며 Hold 신호는 사용하지 않는다)
- 학습된 XGBoost 모형을 사용하여 테스트 데이터를 예측하며, 예측 값으로 매매 신호를 생성하여 거래를 수행한다.

2.3. Reverse Low High 라벨링

본논문에서는 기존의 Low High 라벨링 방법을 변환한 Reverse Low High 라벨링 방법을 제안한다. Reverse Low High 라벨링 방법은 아래와 같으며, 샘플 데이터로 라벨링을 수행한 결과는 Figure 1 과 같다.

- window size 를 설정한다. (본 연구에서는 window size 를 31 로 정하여 실험한다)
- 한 칸씩 이동하면서 window size 기간의 ∨ 모양 최저점과 ∧ 모양 최고점을 찾는다. 예를 들어 window size 의 크기가 5 라면 3 번째 날(가운데 위치한 기간)의 주가가 5 일의 기간 중 최저가인지 최고가인지를 확인한다.
- 최저점은 Down 으로 최고점은 Up 으로 설정하며, 나머지 부분은 모두 Hold 로 설정한다.

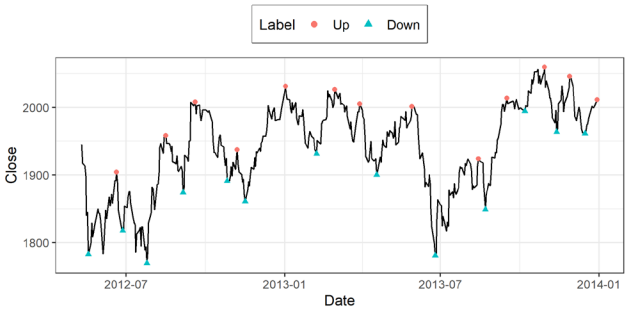


Figure 1 Reverse Low High 라벨링 결과

2.4. 매매 신호 생성

XGBoost 모형으로 예측한 값을 임계값 0.5 기준으로 기준 값 보다 작으면 Up, 기준 값 보다 크면 Down 으로 변환한다. 다음으로 변환한 값에서 처음 발생하는 Up 과 Down 을 찾아 각각 Buy 와 Sell 신호로 변경한 다음, 나머지 연속되는 신호는 모두 Hold 로 처리한다. 거래는 모두 1 주 단위로 진행되며, Buy 신호가 발생한 경우 다음날 시가로 매수하고 Sell 신호가 발생한 경우 다음날 시가로 매도한다. Figure 2, 3 은 샘플 데이터의 매매 신호 생성 전, 후를 보여준다.

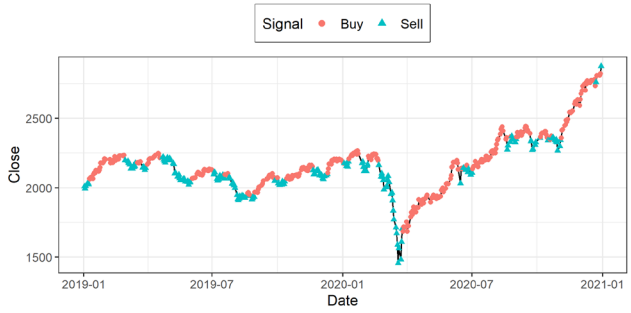


Figure 2 매매 신호 생성 전

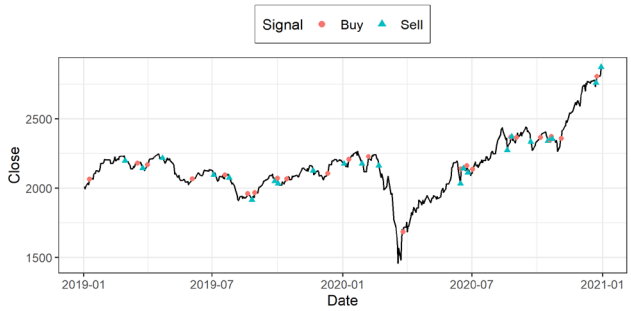


Figure 3 매매 신호 생성 후

2.5. 실험 결과

Table 1 은 KOSPI 200 주가 지수로 실험한 결과로 Up-Down 방법과 기존의 Low High 방법으로 라벨링을 수행하는 경우 누적 수익 값이 음수가 되어 포인트 손실이 발생하는 것을 확인 할 수 있다. 반면 Reverse Low-High 라벨링의 경우 778 포인트의 수익이 발생하는 것을 볼 수 있다. KOSPI 200 주가 지수의 경우 테스트 기간에 매수 후 보유(Buy&Hold) 전략을 취하는 경우 345 포인트의 수익이 발생하는데 이에 비해 2 배 이상의 포인트 수익이 발생한 것을 확인할 수 있다. Table 2 는 S&P 500 주가 지수로 실험한 결과로 KOSPI 경우와 마찬가지로 Up-Down 방법과 Low-High 방법으로 라벨링하는 경우 포인트 손실이 발생한다. S&P 500 주가 지수는 테스트 기간에 매수 후 보유 전략을 취하는 경우 1049 포인트의 수익이 발생하는데, Reverse Low-High 라벨링의 경우 KOSPI 경우와 마찬가지로 Buy & Hold 전략에 비해 2 배 이상의 포인트 수익을 내는 것을 확인할 수 있다. Up-Down 라벨링 방법은 KOSPI 200 과 S&P 500 두 경우 모두 트레이딩 성과가 매우 좋지 않은 것을 확인할 수 있다.

Table 1 KOSPI 200 트레이딩 결과

Method	N_trade	Win %	AVG gain	AVG loss	Cum profit
Up Down	141	0.46	24.60	33.92	-978
Low High	28	0.57	40.85	90.55	-432
Reverse Low High	28	0.46	105.26	39.33	778

Table 2 S&P 500 트레이딩 결과

Method	N_trade	Win %	AVG gain	AVG loss	Cum profit
Up Down	108	0.47	32.88	56.99	-1571
Low High	37	0.43	22.73	91.68	-1561
Reverse Low High	38	0.71	102.47	14.14	2611

3. 결 론

본 논문에서는 트레이딩 시스템 개발 시 데이터 라벨링의 중요성에 대해 논의하였다. 기존의 Low High 라벨링 방법을 변환한 Reverse Low-High 라벨링 방법의 우수성을 확인하기 위해 KOSPI 200 과 S&P 500 주가 지수 데이터를 바탕으로 실증 분석을 진행하였다. 분석 결과 Up-Down, Low-High 방법에 비해 Reverse Low-High 방법의 거래 성과가 더 우수한 것으로 나타났다. 그러나 본 연구는 KOSPI 200, S&P 500 주가 지수 만을 대상으로 실증 분석이 진행되었다는 한계가 있다. 따라서 채권, 원자재 등 다양한 상품에 제안 방법을 적용하여 유용성을 확인할 필요가 있다.

ACKNOWLEDGMENT

본 연구는 2021 년 과학기술정보통신부 및 정보통신기획평가원의 SW 중심대학사업의 연구결과로 수행되었음(2021-0-01399)

참 고 문 헌

[1] Dash R. and Dash PK. “A hybrid stock trading framework integrating technical analysis with machine learning techniques” *The Journal of Finance and Data Science*, 2016.

[2] Zhong X. and Enke D. “Predicting the daily return direction of the stock market using hybrid machine learning algorithms” *Financial Innovation*, 2019.

[3] Nabipour M., Nayyeri P., Jabani H., Shahab S. and Mosavi A. “Predicting stock market trends using machine learning and deep learning algorithms via continuous and binary data: A comparative analysis” *IEEE Access*, 2020.

[4] Sezer OB. and Ozbayoglu AM. “Algorithmic financial trading with deep convolutional neural networks: time series to image conversion approach” *Applied Soft Computing*, 2018.

디노이징 셀프 어텐션 네트워크 기반 정형 데이터 결측값 보간 기법

이도훈, 김한준*
서울시립대학교

yesdhlee95@uos.ac.kr, *khj@uos.ac.kr

Structured Data Missing Value Imputation Based Denoising Self-Attention Network

Do-hoon Lee, Han-joon Kim*
University of Seoul

요 약

기존 머신러닝 기반 정형 데이터 결측값 보간 기법은 수치형 변수에 한해 적용되거나, 변수별로 예측 모형을 만들기 때문에 비효율적이다. 이에 본 논문은 수치형 변수와 범주형 변수가 혼합되어 있는 정형 데이터에 직접적으로 적용 가능한 결측값 보간 모델 DSAN(Denoising Self-Attention Network)를 제안한다. DSAN 은 디노이징 기법과 셀프 어텐션을 결합하여 견고한 특징 표현 벡터를 학습하고, 멀티 태스크 러닝 기법을 통해 결측된 다중 변수에 대한 대체값을 병렬적으로 예측한다. UCI 테이블 데이터 세트를 이용하여 결측값 보간 실험을 수행하여 제안 모델의 유효성을 입증한다.

Ⅰ. 서 론

대표적인 데이터 기반 의사결정 방법론의 기반 기술인 머신러닝은 기본적으로 완전한 고품질 데이터를 요구하지만, 실제 주어지는 입력데이터는 다양한 이유에 의해 결측값을 포함하게 된다. 결측값은 정보 손실을 야기하여 분석에 대한 신뢰성 하락과 모델 성능 하락의 주요 원인이 된다. 일반적으로 데이터 결측 문제를 해결하기 위해, 데이터 보간(Data Imputation) 기법을 활용하여 결측값을 적절한 대체값으로 채워 넣는다.

본 논문은 정형 데이터에 대해 직접적으로 활용 가능한 결측값 보간 모델 DSAN(Denoising Self-Attention Network)을 제안한다. DSAN 은 디노이징 기법과 셀프 어텐션을 결합하여 결측 입력 데이터에 대해 견고한 특징을 학습한다. 이후 멀티 태스크 러닝 기법을 통해 각 변수 별 대체값을 병렬적으로 예측한다.

Ⅱ. 관련 연구

2.1 결측값 보간 기법(Missing Value Imputation)

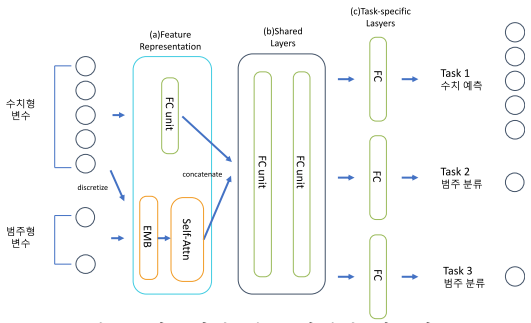
결측값 보간 기법은 적절한 대체값을 추정하여 결측값을 대체하는 기법을 말한다. 평균값이나 최빈값과 같은 기초 통계량 기반의 보간은 데이터가 크고 복잡해질수록 유효성이 떨어지며, 이를 해결하기 위해 머신러닝 기반의 결측값 보간 기법이 연구되고 있다. 머신러닝 기반 기법은 예측 모델 기반의 방법론과 생성 모델 기반의 방법론을 주로 사용한다.

2.2 정형 데이터를 위한 어텐션 메커니즘

최근 어텐션 메커니즘을 정형 데이터 도메인에 적용하기 위한 연구가 진행되고 있다. TabNet[1]은 순차적 어텐션 메커니즘을 활용하여 각 의사결정 단계에서 중요한 특징을 선택함으로써 해석 가능하고 효율적인 학습을 가능하도록 하였다. TabTransformer [2]는 셀프 어텐션을 기반으로 한 Transformer 의 인코더를 사용하였다. 셀프 어텐션을 통해 범주형 변수에 대한 문맥적 임베딩(Contextual Embedding)을 학습하여, 높은 예측 정확도를 달성했다.

Ⅲ. 디노이징 셀프 어텐션 네트워크

본 논문이 제안하는 디노이징 셀프 어텐션 네트워크(DSAN)의 구조는 [그림 1]과 같으며, 멀티 태스크 러닝 기법을 통해 각 변수 별 대체값을 예측한다. DSAN 은 역할에 따라 크게 3 가지 모듈인 특징 표현 모듈, 공유 레이어 모듈, 개별 태스크 레이어 모듈로 구성되어진다.



[그림 1] 디노이징 셀프 어텐션 네트워크 구조

특징 표현 모듈은 데이터를 입력 받아 대체값 예측을 위해 유용한 특징을 학습한다. 범주형 변수의 범주 값과 이산화한 수치형 변수 값에 대해 임베딩 후 셀프 어텐션을 통해 변수 간 연관성 정보를 포함하는 문맥적 임베딩(Contextual Embedding)을 학습한다. 또한 이산화하지 않은 수치형 변수 값을 FC(Fully Connected) 유닛을 통해 병렬적으로 특징을 학습하여 정보 손실을 보완한다. 병렬적으로 학습한 FC 유닛은 문맥적

임베딩과 결합(concatenate)하여 공유 레이어 모듈의 입력으로 활용된다. 이후 공유 레이어 모듈에선 각 변수 예측에 필요한 공유 파라미터를 학습하고, 개별 태스크 레이어 모듈에서 각 변수의 대체값 예측에 필요한 독립적인 파라미터를 학습한다. 수치형 변수의 경우 원래 값을 복원하도록 학습하고, 범주형 변수의 경우 원래 범주 값으로 분류하도록 학습하게 된다.

DSAN 은 학습과정에서 디노이징 기법[3]을 적용하여 일부 값을 제거한 입력 데이터를 받은 후 원래 값을 정답으로 맞추도록 학습한다. 디노이징 기법으로 결측 패턴에 대한 견고한 특징 표현을 학습하기 위해, DSAN 은 특정 변수의 결측 여부에 대한 임베딩 벡터를 학습한다. 이는 특정 변수의 결측 여부와 다른 변수의 관측값 간 연관성을 학습하게 하여, 모델의 예측 성능을 높일 수 있었다.

Ⅳ. 실험 및 결과

4.1 실험 데이터 및 방법

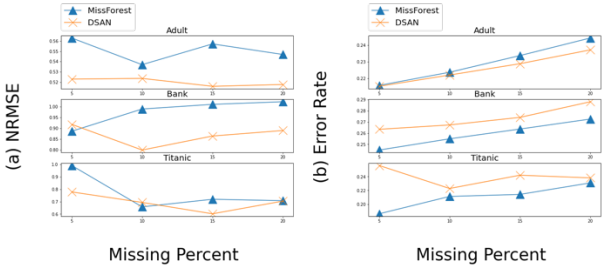
DSAN 의 유효성을 보이기 위해 UCI 정형 데이터 세트인 'Adult', 'Bank', 'Titanic' 데이터로 결측값 보간 실험을 수행한다. 실험은 세 단계로 이루어지며, 첫째 단계에선 완전한 데이터에 대해 MCAR(Missing Completely At Random)[4] 방식으로 5~20%의 결측값을 생성한다. 둘째 단계에선 결측값 보간을 수행하여 원래 값과의 오차를 측정한다. 이 때 수치형 변수의 경우 NRMSE 를, 범주형 변수의 경우 Error Rate 을 평가 척도로 삼는다. 셋째 단계에선 정제한 데이터를 이용해 이진 분류 모델을 학습하여 성능을 분석한다. 이 때 분류 모델의 성능 평가 지표로 AUC-ROC 를 활용한다.

[표 1] 실험 데이터 셋

데이터 셋	범주형 변수 개수	수치형 변수 개수	레코드 개수
Adult	9	6	30162
Bank	10	7	11162
Titanic	4	5	712

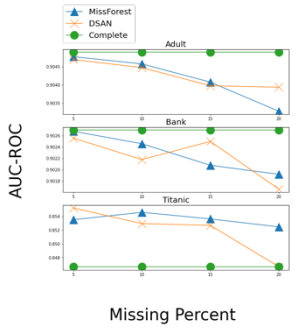
4.2 실험 결과

결측값 보간 모델 MissForest[5]를 비교 모델로 하여 실험 결과를 분석한다. [그림 2]의 (a)는 수치형 변수에 대한 오차를 의미하며 낮을수록 원래 값에 근사한 값으로 보간했음을 의미한다. DSAN 이 비교 모델 대비 평균 7% 향상된 성능을 보이는 것을 확인할 수 있다. [그림 2]의 (b)는 범주형 변수에 대한 오차를 의미하며, 제안 모델인 DSAN 이 비교 모델 대비 약 1% 낮은 성능을 보였다. 하지만 레코드 개수가 가장 큰 'Adult' 데이터 셋의 경우 DSAN 의 성능이 더 좋았으며, 이는 레코드 수가 클 수록 제안 모델이 우수한 성능을 보이는 것으로 해석할 수 있다.



[그림 2] 실험 결과(NRMSE, Error Rate)

[그림 3]의 경우 정제한 데이터를 학습한 이진 분류 모델의 성능을 나타낸다. 원래 데이터를 학습한 이진 분류 모델 성능 대비 평균 0.1% 미만의 차이를 갖으며, 이는 DSAN 이 정제한 데이터가 이 후 수행하는 작업에서 유의미한 결과를 낼 수 있음을 의미한다.



[그림 3] 실험 결과(AUC-ROC)

Ⅴ. 결론

본 논문은 DSAN 기반의 정형 데이터 결측값 보간 기법을 제안하였다. 수치형 변수 보간의 경우 비교 모델 대비 7% 향상된 성능을 보였으며, 범주형 변수의 경우 데이터 레코드 수가 커질수록 높은 성능을 보이는 것을 확인하였다. 또한 보간한 데이터를 학습한 이진 분류 모델이 원래 데이터 학습 대비 0.1% 미만의 차이를 보이며 유효성을 입증하였다.

ACKNOWLEDGMENT

* 사사의 글: 이 논문은 2020 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(No.2020-0-00121, 데이터 품질 평가기반 데이터 고도화 및 데이터셋 보정 기술 개발)과 과학기술정보통신부 및 정보통신기술진흥센터의 대학 ICT 연구센터지원사업으로 수행된 연구(IITP-2020-2018-0-01417)이며, 또한 2018 년도 정부(교육부)의 재원으로 한국연구재단의 지원(No. NRF-2018R1D1A1A02086148, 시멘틱 텍스트 큐보이드 기반 자가증식 기술 및 딥러닝 학습 기술)을 받아 수행된 연구임

참 고 문 헌

[1] S. O. Arik and T. Pfister, "TabNet: Attentive Interpretable Tabular Learning", arXiv preprint arXiv:1908.07442, 2019.

[2] X. Huang, A. Khetan, M. Cvitkovic and Z. Karnin, "TabTransformer: Tabular Data Modeling Using Contextual Embeddings", arXiv preprint arXiv:2012.06678, 2020.

[3] P. Vincent, H. Larochelle, Y. Bengio and P.A. Manzagol, "Extracting and Composing Robust Features with Denoising Autoencoders", International Conference on Machine Learning, 2008.

[4] D. B. RUBIN, "Inference and missing data", Biometrika, Vol. 63, pp 581-592, 1976.

[5] DJ. Stekhoven and P. Buhlmann, "MissForest-non-parametric missing value imputation for mixed-type data", Journal of Machine Learning Research, Vol.28, No.1, pp 41-75, 1997

감염병 전파 과정을 위한 마르코프 체인 기반 SEIVR 모델 연구

안수진, 권민혜
숭실대학교

asujin331@soongsil.ac.kr minhae@ssu.ac.kr

Markov Chain based SEIVR Model for Infection Transmission

Sujin Ahn Minhae Kwon
Soongsil University

요 약

코로나 바이러스(SARS-CoV2; COVID-19)의 감염 양상을 파악하기 위해서는 다양한 시나리오를 포함할 수 있는 수학적 모델링이 중요하다. 본 논문에서는 백신 접종 상태를 포함한 5 가지의 상태(state)와 상태 전이 확률(state transition probability)을 정의하여 마르코프 체인 기반 감염병 전파 모델을 제안한다. 그 후, 대한민국 COVID-19 데이터를 이용하여 감염병의 확산 과정을 본 모델로 설명하고 시뮬레이션 결과를 확인하였다.

Ⅰ. 서 론

COVID-19 은 사람들의 전방위적 행동 뿐만 아니라 사회적, 경제적인 면까지 영향을 미치고 있다. 감염병이 지속되면서 정부는 장기적이고 효율적인 대비책을 세우는 것이 필수적이다. 실질적인 감염병 대응 전략을 수립하기 위해서는 감염병에 대한 이해가 선행되어야 한다. 감염병의 유행은 현실에서의 실험적 검증이 불가능하므로 수학적 모델링이 감염병에 대한 이해와 유행에 대응하는 계획 수립에 중요한 역할을 하고 있다. 본 연구에서는 감염병 전파 모델을 이산적인 확률 과정(stochastic process)인 마르코프 체인(Markov chain)에 기반하여 제시한다. 각 상태의 값과 일일 확진자 수를 비교하고, 감염병 전파 과정에 대해 설명한다.

Ⅱ. 이산시간 마르코프 체인 기반의 SEIVR 모델 제안

본 논문에서는 이산시간 마르코프 체인 (discrete-time Markov chain) 모델을 기반으로 상태와 상태 전이 확률을 다음과 같이 정의하였다.

Ⅱ.1 상태 정의

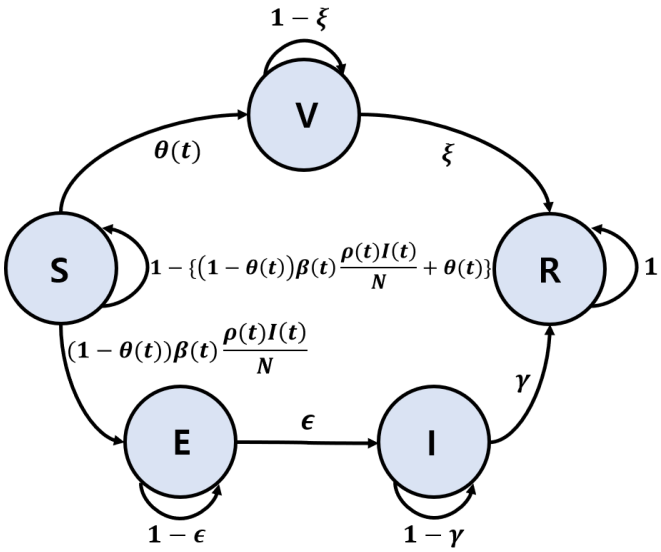
한 명의 인구는 특정 시간 t 에 복수의 상태에 속할 수 없으며, 각 상태의 인구수는 음이 아닌 정수이다. 본 모델에서는 인구의 출생과 자연 사망을 고려하지 않았기 때문에 모든 상태의 인구수의 합은 전체 인구수 N 으로 고정된다. 본 논문에서는 S, E, I, V, R 5 가지 상태로 정의하였다.

$\boldsymbol{S}$ (***Susceptible***) 아직 감염이 되지 않은 감수성자 상태를 의미한다. 한번 S 상태를 벗어나면 다시 돌아올 수 없다.

$\boldsymbol{E}$ (***Exposed***) 감염되었지만 바이러스 배출을 하지 않는 잠복 상태를 의미한다. S 상태에서 I 상태의 개체와의 접촉을 통해 E 상태로 전이된다.

$\boldsymbol{I}$ (***Infectious***) 감염병의 잠복기가 지나 전염 능력이 존재하는 상태로, 감염 상태라고 한다.

$\boldsymbol{R}$ (***Recovered***) 면역 체계를 갖춘 회복 상태이거나 사망한 상태이다. I상태에서 벗어나 회복기를 지나거나 백신 접종 이후에는 면역 체계를 갖추게 된다.



[그림 1] SEIVR Model 의 State Transition Diagram

$\boldsymbol{V}$ (***Vaccinated***) 백신 접종을 1 번이라도 실시한 백신 접종 상태이다. 면역력 생성을 위한 일정 기간이 지나면 R 상태로 전이된다.

S 상태에서 R 상태로 가기까지의 역방향은 고려하지 않는다. 즉, S 상태에서 다음 상태로 전이된 이후엔 다시 S 상태로 복귀할 수 없다. R 상태는 흡수 상태(absorbing state)로, R 상태에 진입하게 되면 다른 상태로 이동할 수 없게 된다. t 시점에서 K 상태의 값 $\mathrm{K}(t)$ 는 K 상태에 머무르는 인구수이다. t 시점의 감수성 상태 값을 $S(t)$, 잠복기 상태 값을 $E(t)$, 감염 상태의 값을 $I(t)$, 회복 상태의 값을 $R(t)$, 백신 접종 상태의 값을 $V(t)$로 표현할 수 있다.

Ⅱ.2 상태 전이 확률

마르코프 체인은 $t-1$시점까지의 과거 관측 값과 t시점의 현재의 관측 값을 알고 있는 상황에서 $t+1$시점의 미래의 예측 값의 조건부 확률 분포가 현재의 관측 값에 의존한다고 가정한다[1]. $p_{\mathrm{K}J}(t)$는 t 시점에서 K상태라는 정보가 주어진 상황에서, $t+1$시점에서 J 상태에 있을

조건부 확률로 상태 전이 확률이다($t \geq 0$). $p_{KK}(t)$는 t시점에서 여전히 K 상태에 머무를 확률이다. 역방향의 상태 전이 확률은 0이다. [그림 1]은 상태 간의 이동 과정과 확률을 나타내었다.

$$p_{SS}(t) = 1 - \left\{(1-\theta(t))\beta(t)\frac{\rho(t)I(t)}{N} + \theta(t)\right\}$$

$$p_{SE}(t) = (1-\theta(t))\beta(t)\frac{\rho(t)I(t)}{N}$$

$$p_{SV}(t) = \theta(t)$$

$$p_{EE}(t) = 1-\epsilon\,, p_{EI}(t) = \epsilon$$

$$p_{II}(t) = 1-\gamma\,, p_{IR}(t) = \gamma$$

$$p_{VV}(t) = 1-\xi, p_{VR}(t) = \xi$$

$$p_{ES}(t), p_{IE}(t), p_{RI}(t), p_{RV}(t), p_{VS}(t) = 0$$

$\theta(t)$는 t시점에서 S 상태 값 중 일일 1 차 백신 접종자 수의 비율이다. $\rho(t)$는 t시점에서 I 상태 값 중에 격리되지 않은 비율이다. $\beta(t)$는 t시점의 감염 전파율 (transmission rate)이다. $\beta(t)$는 감염병 대응 정책과 병의 특성에 종합적으로 영향을 받는다. $1/\epsilon$는 감염된 이후 전염력이 생기기까지의 기간인 평균 잠복 기간(mean latent period), $1/\gamma$는 감염 전파 가능 기간(mean infectious period), $1/\xi$ 는 한번이라도 백신 접종을 한 이후 완전한 면역력이 생기기까지 걸리는 평균 기간으로 정의한다. $t+1$ 시점에서 각 상태의 값은 다음과 같다[2,3,4].

$$S(t+1) = S(t) - X_{SE}(t+1) - X_{SV}(t+1)$$

$$E(t+1) = E(t) + X_{SE}(t+1) - X_{EI}(t+1)$$

$$I(t+1) = I(t) + X_{EI}(t+1) - X_{IR}(t+1)$$

$$V(t+1) = V(t) + X_{SV}(t+1) - X_{VR}(t+1)$$

$$R(t+1) = R(t) + X_{IR}(t+1) + X_{VR}(t+1)$$

여기서, $X_{KJ}(t+1)$은 $t+1$ 시점에서의 K 상태에서 J 상태로 전이될 때의 인구 수를 나타내는 확률변수로 $p_{KJ}(t)$의 확률을 갖는 이항 분포(binomial distribution)를 따른다. 이항 분포 성질에 의해 K상태에서 J 상태로 전이될 때의 기대되는 상태 값은 다음과 같다.

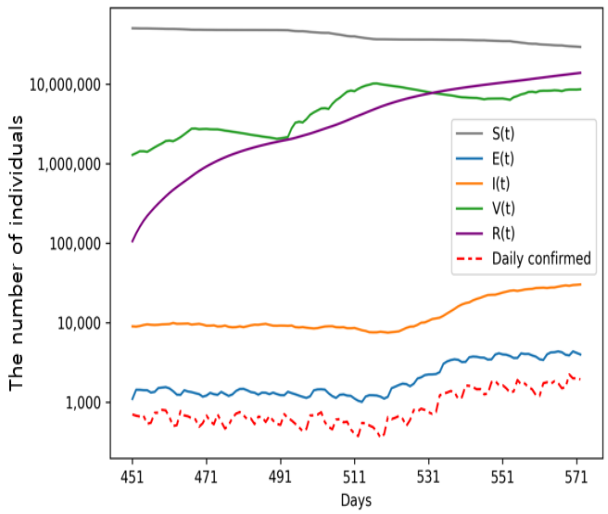
$$X_{\mathrm{KJ}}(t+1) \sim Binom\big(\mathrm{K}(t), p_{\mathrm{KJ}}(t)\big)$$

$$\mathbb{E}[X_{\mathrm{KJ}}(t+1)] = \mathrm{K}(t)\, p_{\mathrm{KJ}}(t)$$

Ⅲ. 대한민국 COVID-19 데이터 기반 SEIVR 모델 결과 분석

[그림 2] 에서는 COVID-19 발병 시작일 (2020 년 1 월 20 일)을 기준으로 451 일 (2021 년 4 월 15 일)~ 572 일 (2021 년 8 월 14 일)간의 각 상태에 속하는 인구수와 일일 확진자 수를 비교해보았다. ϵ, γ, ξ는 각각 $\frac{1}{2}$, $\frac{1}{9}$, $\frac{1}{49}$ 의 값을 갖는다고 가정하였다[5]. 각 상태 값의 합은 전체 인구수를 따르도록 하였다. S 상태 값은 감염병이 지속됨에 따라 감소하고 있다. 이는 S 상태에 머무르는 인구가 E 상태나 V 상태로 전이되기 때문이다. S 상태에서 E, I 상태를 거치거나 V 상태를 거치면 R 상태에 도달하게 된다. R상태에 머무르는 인구는 더 이상 다른 상태로 이동할 수 없기 때문에 지속적으로 증가하게 된다.

V 상태 값은 백신 접종 현황에 따라, E, I 상태 값은 감염병의 전파 양상에 따라 증감한다. 일일 확진자 수와



[그림 2] 각 상태 값과 일일 확진자 수 변화 양상

비교해 본 결과, E, I 상태 값의 변화 양상은 일일 확진자의 변화 양상과 비슷함을 확인할 수 있다. 또한, I상태 값은 E 상태 값보다 큼을 확인할 수 있다. 이는 $\frac{1}{\epsilon} < \frac{1}{\gamma}$ 이므로 한 인구가 E 상태에서 I 상태로 변화하는 속도가 I 상태에서 R상태로 변화하는 속도보다 빠르기 때문이다.

Ⅳ. 결론

본 논문에서는 백신 접종을 고려하여 상태에 따른 감염병의 확산 과정을 살펴보았다. 본 모델에서는 감염 전파율 $\beta(t)$, 접종 비율 $\theta(t)$, 격리되지 않은 비율 $\rho(t)$ 등의 변수들을 포함하여 기존의 단순화된 결정론적 감염병 모델보다 현실적인 부분들을 고려하여 반영하였다. 또한, 마르코프 체인을 기반으로 SEIVR 모델의 각 상태 값을 추정하고, 각 상태의 변화를 확인하였다.

ACKNOWLEDGEMENT

이 논문은 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT 연구센터지원사업(IITP- 2021-2020-0-01602)과 한국 연구재단(NRF- 2020R1F1A1069182)의 지원을 받아 수행된 연구임.

참 고 문 헌

[1] H. Taylor, S. Karlin, “An Introduction to Stochastic Modeling,” Academic Press, 1998.

[2] C. Twumasi, L. Asiedu, E. N. Nortey, “Markov chain modeling of HIV, tuberculosis, and Hepatitis B Transmission in Ghana,” Interdisciplinary perspectives on infectious diseases, 2019.

[3] Z. Xu, B. Wu, U. Topcu, “Control strategies for COVID-19 epidemic with vaccination, shield immunity and quarantine: A metric temporal logic approach,” PloS one, vol. 16, no. 2, 2021.

[4] 안수진, 권민혜, “순간 재생산지수에 기반한 COVID-19 확진자 수 변화 예측 시스템 연구,” 한국통신학회 하계종합학술대회, June 2021.

[5] J. T. Wu, K. Leung, G. M. Leung, “Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan China: a modelling study,” The Lancet, vol. 395, no. 10225, pp. 689-697, 2020.

다중안테나 통신시스템 성능 분석을 위한 회귀 기반의 확률 밀도 함수 근사 방법론

이영석, 이기훈, 정방철
충남대학교
yslee@o.cnu.ac.kr, kihun.h.lee@cnu.ac.kr, bcjung@cnu.ac.kr

A Novel Regression Method to Approximate Probability Density Function for Performance Analysis of MIMO Communication Systems

Young-Seok Lee, Ki-Hun Lee, Bang Chul Jung
Chungnam National University

요 약

본 논문에서는 다중안테나(multiple-input multiple-output: MIMO) 통신시스템의 성능 분석을 위한 회귀 기반의 확률 밀도 함수(probability density function: PDF) 근사 방법론을 제안한다. 이를 위해 실제로 다루기 어려운 형태를 갖는 MIMO 채널의 최대 고유값 분포 PDF를 근사하고, 모의실험을 통해 제안한 방법론의 효용성을 검증했다. 구체적으로, 회귀를 기반으로 근사한 PDF와 실제 PDF를 비교했으며, 이를 기반으로 수행한 신호 대 잡음비(signal-to-noise ratio: SNR) 대비 비트 당 오류율(bit-per-error rate: BER) 성능 분석 결과 또한 거의 일치하는 것을 확인했다.

Ⅰ. 서 론

인공지능(artificial intelligence: AI) 분야의 발전과 동시에 다양한 분야에서 AI를 응용한 기술이 높은 효용성을 입증하고 있다. 통신 분야에서도 AI를 적용한 다양한 기술이 개발되고 있으며, 성능 향상을 입증한 AI 기반의 통신 기술은 초고신뢰, 초저지연 등 차세대 통신시스템의 높은 성능 지표를 달성하는 데 활용할 수 있을 것으로 전망된다 [1], [2]. 몇 가지 예로, [3]에서는 다중안테나(multiple-input multiple-output: MIMO) 기반 통신시스템에서 채널 상태 정보(channel state information: CSI) 추정으로 인한 부하를 완화하기 위해 기계학습 기술이 적용되었고, [4]에서는 딥러닝 기반의 하이브리드 빔포밍 기술이 제안되었다.

본 논문에서는 이와 다른 관점으로 통신시스템의 수학적 성능 분석을 위한 회귀(regression) 기반의 방법론을 제안한다. 통상적으로 무선통신 기술의 성능 예측을 위해서는 채널의 통계적 특성(randomness)으로부터 평균 성능을 도출하기 위해 많은 횟수의 실험을 수행하고 평균 내는, 이른바 몬테카를로(Monte-Carlo) 방법의 모의실험을 수행한다. 하지만, 시스템에 따라 이 방법은 매우 긴 시간 소요가 필요하며, 이에 따라 통신시스템의 성능을 수학적으로 분석하고 이로부터 통찰을 도출하는 것은 통신 분야의 전통적인 연구 흐름이다.

한 예로 MIMO 채널의 최대 고유값에 대응하는 고유벡터를 기반으로 빔포밍하는 최대 고유값 빔포밍(혹은 SVD 빔포밍) 기법에서 최대 고유값의 확률 밀도 함수(probability density function: PDF)는 그 형태가 수학적 분석을 위해 다루기 쉽지 않다 [5], [6].

Ⅱ. 최대 고유값 빔포밍 기법의 시스템 모델

본 논문에서는 각각 N 개의 안테나를 갖는 송신단과 M 개의 안테나를 갖는 수신단을 고려한다. 이때, 송수신단 모두 각 안테나 사이 채널 상태 정보(CSI)를 알고 있다고 가정하며, 앞서 언급한 대로 이를 활용한 최대 고유값 빔포밍 기술을 고려한다. 이에 따라 최대 고유값에 대응하는 고유벡터를 기반으로 송수신 빔포밍을 적용하면, 최종 수신 신호 y는 등가적으로 다음과 같이 표현할 수 있다 [5]:

$$y = \mathbf{u}_s^H \mathbf{y} = \mathbf{u}_s^H \mathbf{H} \mathbf{v}_s^H x + \mathbf{u}_s^H \mathbf{n} = \lambda_s x + \tilde{n}, \qquad (1)$$

여기서 $\mathbf{H} \in \mathbb{C}^{M \times N}$는 송수신단 각 안테나 사이의 무선 채널로 구성된 행렬이며, 모든 채널은 서로 독립이고 $\mathcal{CN}(0, 1)$의 동일한 분포를 따른다고 가정한다. 또한, $\mathbf{v}_s \in \mathbb{C}^N$와 $\mathbf{u}_s \in \mathbb{C}^M$ $(s = \min(N, M))$는 최대 고유값 $\phi_s = \lambda_s^2$에 대응되는 우특이 벡터와 좌특이 벡터를 각각 나타내며, $\mathbf{n} \in \mathbb{C}^M$은 신호 송수신 과정에서 발생하는 열잡음을 의미한다. 모든 잡음은 $\mathcal{CN}(0, N_0 \cdot \mathbf{I}_M)$ 분포를 따른다고 가정한다. 본 논문에서는 추가적으로 송신단에서 유효 무선 채널 이득(ϕ_s)이 특정 임계치인 η 이상일 때만 변조 신호 x를 송신하는 기회적 전송을 고려한다.

Ⅲ. 제안하는 확률 밀도 함수 근사 방법론

기회적 전송을 고려함에 따라 최대 고유값 빔포밍 기법에서 최대 고유값(ϕ_s)의 PDF $\tilde{f}_{\phi_s}(\phi_s)$는 [6]의 (17, $f_{\phi_s}(\phi_s)$)로부터 다음과 같이 잘린 분포(truncated distribution)로 유도된다:

$$\tilde{f}_{\phi_s}(\phi_s) = f_{\phi_s}(\phi_s \mid \phi_s \geq \eta) = \frac{\left|\boldsymbol{\Psi}_{\mathrm{c}}(\phi_{\mathrm{s}})\right| \mathrm{tr}\left(\boldsymbol{\Psi}_{\mathrm{c}}^{-1}(\phi_{\mathrm{s}}) \boldsymbol{\Phi}_{\mathrm{c}}(\phi_{\mathrm{s}})\right)}{\prod_{k=1}^{s} \Gamma(t-k+1)\Gamma(s-k+1) - \left|\boldsymbol{\Psi}_{\mathrm{c}}(\phi_{\mathrm{s}})\right|} U(\eta), \qquad (2)$$

여기서 $\left\{\boldsymbol{\Psi}_{\mathrm{c}}(x)\right\}_{i,j} = \gamma(t-s+i+j-1, x)$, $\left\{\boldsymbol{\Phi}_{\mathrm{c}}(x)\right\}_{i,j} = \phi_s^{t-s+i+j-2} e^{-\phi_s}$, $i, j = \{1, 2, \cdots, s\}$이고 $t = \max(N, M)$이며, $\Gamma(\cdot)$와 $\gamma(\cdot)$는 각각 감마 함수 및 하부 불완전 감마 함수를 의미한다. 또한, $|\cdot|$는 행렬식(determinant), $\mathrm{tr}(\cdot)$은 행렬의 대각합(trace), 그리고 $U(\cdot)$는 단위계단 함수(unit step function)를 나타낸다.

한편, 식 (2)는 최대 고유값 빔포밍 기반 통신시스템의 수학적 성능 분석에 있어 다루기 어렵다는 한계를 갖는다. 본 연구팀은 모의실험과 (2)로부터 최대 고유값 빔포밍 기반 기회적 전송 기법에서 최대 고유값의 PDF가 지수함수의 형태를 보이는 것을 도출했으며, 이에 따라 실제 PDF를 근사 지수 V에 기반하여 다음과 같이 근사한다:

$$\tilde{f}_{\phi_s}(\phi_s) \approx \bar{f}_{\phi_s}(\phi_s) = \sum_{v=1}^{V} \alpha_v \phi_s^{v-1} e^{-\beta_v \phi_s} + \delta, \qquad (3)$$

여기서 각 파라미터 $\{\alpha_1, \cdots, \alpha_V, \beta_1, \cdots, \beta_V, \delta\} \in \mathbb{R}^{2V+1}$는 몬테카를로 방법을 통해 획득한 ϕ_s의 PDF 또는 (2)로부터 계산되는 입출력 쌍 $\left\{\phi_s, \tilde{f}_{\phi_s}(\phi_s)\right\}$을 데이터(ground truth, $\{X_{\mathrm{data}}, Y_{\mathrm{data}}\}$)로 활용한 회귀 문제를 통해 도출한다. 한편, (3)에서 $V = 1$ 또는 2인 경우 각각 통신시스템의 수학적 성능 분석에서 자주 활용되는 지수분포 형태 또는 지수분포와 카이 제곱(chi-squre) 분포의 합 형태로 도출되는 것을 확인할 수 있다.

(3)의 각 파라미터를 도출하기 위해 본 논문에서 제안하는 회귀 최적화 문제는 다음과 같이 공식화(formulation)할 수 있다:

$$\text{(P1)} \quad \min \sum_{\phi_s \in X_{\mathrm{data}}} \left\{ \left| \bar{f}_{\phi_s}(\phi_s) - \tilde{f}_{\phi_s}(\phi_s) \right|^2 \right\}$$
$$\text{s. t.} \quad \bar{f}_{\phi_s}(\phi_s) \leq \tilde{f}_{\phi_s}(\phi_s), \forall \phi_s \geq \eta.$$

마지막으로, (3)은 음수 값을 가질 수 있으며, 수학적 정의에 따라 PDF는 음의 값을 가질 수 없으므로 다음과 같이 0과 max 함수를 통해 최종 근사 PDF를 정의한다:

$$\hat{f}_{\phi_s}(\phi_s) = \left[\bar{f}_{\phi_s}(\phi_s)\right]^+ = \left[\sum_{v=1}^{V} \alpha_v \phi_s^{v-1} e^{-\beta_v \phi_s} + \delta\right]^+, \qquad (4)$$

여기서 $[\cdot]^+$는 $[x]^+ := \max(0, x)$을 나타낸다.

표 1. 송수신 안테나 수에 따른 근사 PDF 파라미터 ($V=2$)

$N\times M$	2×2	3×3	4×4
α_1	-6.0414	0.9026	-151.3761
β_1	0.5776	-0.0014	0.3928
α_2	5.0299	1.0138	35.6170
β_2	0.7587	0.4510	0.4694
δ	0	-0.9412	-0.0012
η	3.18	6.21	9.47
ζ	13.26	14.67	19.88

Ⅳ. 모의실험 결과 및 결론

본 절에서는 앞서 회귀를 통해 근사한 PDF (4)의 정확도를 검정하기 위해 몬테카를로 방법 또는 (2)를 통해 도출되는 최대 고유값의 실제 PDF와 비교·분석한다. 또한, 통신시스템의 수학적 성능 분석 활용 예로 전송 신호 x가 이진 위상 편이 변조(binary phase shift keying: BPSK) 신호일 때, 최대 고유값 빔포밍 기반 기회적 전송 기법의 신호 대 잡음비(signal-to-noise ratio: SNR) 대비 비트 당 오류율(bit-per-error rate: BER) 성능을 분석하고 모의실험 결과와 비교함으로써 (4)를 검정한다.

그림 1은 기회적 전송을 고려한 최대 고유값 빔포밍 기법에서 몬테카를로 방법을 통해 획득한 최대 고유값의 정규화된 히스토그램과 수식 (2) 및 본 논문에서 근사한 PDF (4)를 도시한 결과이다. 본 논문에서는 송수신 안테나 개수가 같은 상황($N=M$)을 고려했으며, 송신단의 전송확률은 0.5로 가정했다. 또한, 그림 1은 $N=\{2,3,4\}$일 때 최대 고유값의 PDF를 나타낸다. 근사한 PDF는 근사 지수 2 ($V=2$)로부터 도출했으며, 파라미터 값은 표 1과 같다. 표 1에서 η는 전송확률이 0.5일 때 최대 고유값의 임계값을 의미하며, ζ는 근사한 PDF의 함수값이 0일 때의 최대 고유값을 나타낸다. 근사한 PDF는 몬테카를로 방법 및 기존 분석 결과(2)와 거의 일치하는 것을 보여준다.

한편, 이론적으로 변조 신호가 BPSK일 때, 무선 채널 이득을 고려한 SNR 대비 BER 성능은 다음과 같이 유도된다:

$$\mathbb{E}[\mathrm{P}_e]=\int_{\eta}^{\zeta}\mathrm{Q}\left(\sqrt{\frac{2\phi_s}{N_0}}\right)\hat{f}_{\phi_s}(\phi_s)\,d\phi_s$$
$$\approx\left[-\frac{\alpha_1}{2\beta_1}\left(\frac{\mathrm{erf}(\mathrm{A}(\beta_1)\mathrm{B}(\phi_s))}{\mathrm{A}(\beta_1)}+e^{-\beta_1\phi_s}\mathrm{erfc}(\mathrm{B}(\phi_s))\right)\right.$$
$$-\frac{\alpha_2}{4\beta_2^2}\left(\frac{(3\mathrm{A}^2(\beta_2)-1)\mathrm{erf}(\mathrm{A}(\beta_2)\mathrm{B}(\phi_s))}{\mathrm{A}^3(\beta_2)}\right.$$
$$+2e^{-\beta_2\phi_s}(\beta_2\phi_s+1)\mathrm{erfc}(\mathrm{B}(\phi_s))$$
$$\left.\left.-\frac{2\beta_2N_0\mathrm{B}(\phi_s)}{\sqrt{\pi}\,\mathrm{A}(\beta_2)}e^{-\mathrm{A}^2(\beta_2)\mathrm{B}^2(\phi_s)}\right)\right)$$
$$\left.+\frac{\delta N_0}{2}\mathrm{erf}(\mathrm{B}(\phi_s))+\phi_s\mathrm{erfc}(\mathrm{B}(\phi_s))-\frac{N_0e^{-\mathrm{B}^2(\phi_s)}\mathrm{B}(\phi_s)}{\sqrt{\pi}}\right]_{\eta}^{\zeta}\quad(5)$$

여기서 $\mathrm{A}(x)=\sqrt{xN_0+1}$, $\mathrm{B}(x)=\sqrt{x/N_0}$, $\zeta=\hat{f}_{\phi_s}^{-1}(0)$ 이다.

그림 2는 BPSK를 고려한 최대 고유값 빔포밍 기반 기회적 전송 기법의 BER 성능 모의실험 결과와 본 논문에서 근사한 PDF를 통해 분석한 결과 (5)를 나타낸다. 결과적으로, 근사 PDF에 기반한 성능 분석 결과가 모의실험과 거의 일치하는 것을 확인할 수 있다. 따라서, 본 논문에서 제안하는 PDF 회귀 방법이 수학적으로 다루기 어려운 PDF를 근사함으로써, 다양한 통신시스템의 통신 성능을 쉽게 분석할 수 있음을 검정하였다.

본 논문에서는 다중안테나 기반 통신시스템의 수학적 분석을 위한 회귀 기반의 PDF 근사 방법론을 제시했다. 이는 통신시스템 성능 분석에 있어, 수학적으로 다루기 어려운 PDF를 분석에 용이한 형태로 근사하는 것이며, 한 예로 최대 고유값 빔포밍 기반 기회적 전송 기법에서 최대 고유값의 PDF를 근사했다. 모의실험을 통해 근사한 PDF가 실제 PDF와 거의 일치하는 것을 확인했으며, 실제 BPSK 변조를 고려한 SNR 대비 BER 성능의 수학적 분석 결과를 모의실험 결과와 비교함으로써 본 논문에서 제안한 방법론의 효용성을 입증했다.

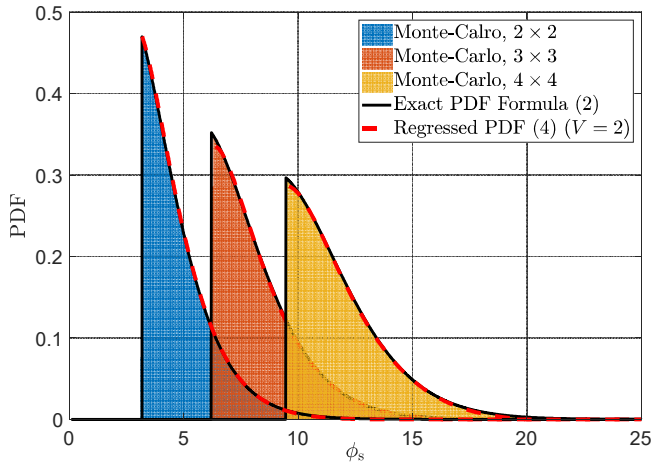


그림 1. 기회적 최대 고유값 빔포밍의 최대 고유값 PDF 비교 결과

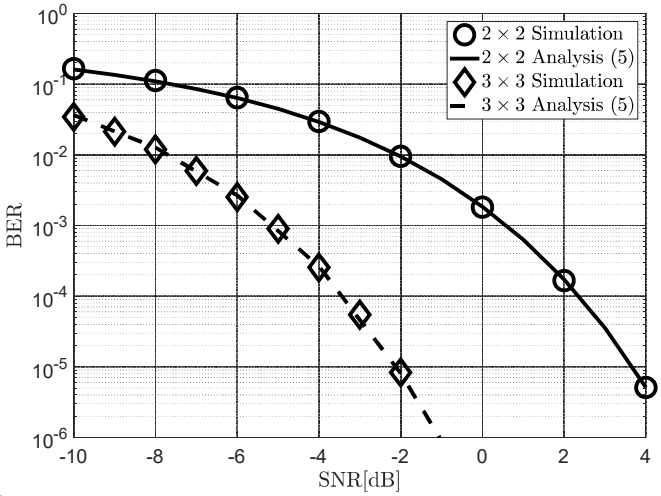


그림 2. 근사 PDF에 기반한 기회적 최대 고유값 빔포밍의 BER 성능

ACKNOWLEDGMENT

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (2021-0-00486, ABC-MIMO: 증강 빔 라우팅 기반 차세대 다중 입출력 통신 시스템).

참 고 문 헌

[1] T. O'shea and J. Hoydis, "An introduction to deep learning for the physical Layer," *IEEE Trans. Cogn. Commun. Netw.*, vol. 3, no. 4, pp. 563-575, Oct. 2017.

[2] S. Ali, W. Saad, D. Steinbach, I. Ahmad, and J. Huusko, "White paper on machine learning in wireless communication networks," *University of Oulu 6G Research Visions*, no. 7, Jun. 2020.

[3] J. Joung and B. C. Jung, "Machine learning based blind decoding for space-time line code (STLC) systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 5154-5158, May 2019.

[4] T. Peken, S. Adiga, R. Tandon, and T. Bose, "Deep learning for SVD and hybrid beamforming," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6621-6642, Oct. 2020.

[5] M. Raja and P. Muthuchidambaranathan, "BER performance of SVD-based transmit beamforming with various modulation techniques," in *Proc. 5th IEEE Int. Conf. Ind. Inf. Syst (INDIN).*, pp. 155-160, Jul.-Aug. 2010.

[6] M. Kang and M. -S. Alouini, "Largest eigenvalue of complex Wishart matrices and performance analysis of MIMO MRC systems," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 3, pp. 418-426, Apr. 2003.

양산라인 실시간 품질 검사를 위한 딥러닝 기반 엣지컴퓨팅 비전 검사 장치

김량수, 유학, 김근용, 김성창
한국전자통신연구원

rskim@etri.re.kr

Deep Learning-based Edge Computing Vision Inspection Device for Real-time Quality Inspection on Mass Production Line

Ryangsoo Kim, Hark Yoo, Geun-Yong Kim, Sung chang Kim
Electronics and Telecommunications Research Institute (ETRI)

요 약

본 논문에서는 공장의 양산라인에서 생산되는 인덕터 제품의 코어 크랙 여부를 실시간으로 검출하기 위한 딥러닝 기반 엣지컴퓨팅 응용솔루션을 소개한다. 인덕터 제품의 비전 데이터 취득을 위해 4 채널 조명이 장착된 비전 검사 장치를 자체 제작하여 인덕터 코어 크랙 데이터셋을 구축하였고, Pytorch 프레임워크를 활용하여 컨벌루션 신경망으로 구성된 딥러닝 모델 학습을 수행하였다. 학습이 완료된 딥러닝 모델의 실험결과 코어 크랙 판별 정확도는 약 99%임을 확인하였으며, 학습이 완료된 모델을 Nvidia Jetson Nano 엣지컴퓨팅 플랫폼에서 수행한 결과 약 9 FPS 로 추론이 가능함을 확인하였다.

Ⅰ. 서 론

전자제품에 주요부품으로 사용되는 인덕터는 부품 단가가 몇백원 수준의 비교적 저가의 부품으로 인식된다. 그러나, 인덕터의 불량으로 인해 발생하는 In-Rush Current 가 TV 와 같은 고가의 전자제품 회로에 유입되는 경우 SMPS 와 같은 주요 부품에 지속적인 타격을 입혀 제품 수명을 단축시키는 원인이 된다. 이와 같은 문제를 해결하기 위해서는 인덕터를 생산하는 양산라인에서 정밀한 제품 검사를 수행하여 인덕터 패키징 크랙 유무를 판별하는 검사가 필수적이다. 하지만 자동화된 양산라인에서 생산 일정에 차질없이 수많은 인덕터의 크랙 유무를 판별하기 위해서는 실시간으로 제품의 이미지 데이터를 취득하고 판별해내는 검사 장치가 필요하다.

최근 딥러닝 기술의 급격한 발전으로 인해 다양한 실생활 이미지와 영상 데이터 분석이 가능한 딥러닝 기반 비전 시스템이 다양한 응용서비스 분야에 적용되고 있다[1]. 특히, 딥러닝 기술이 기존의 기계학습 알고리즘 수준을 뛰어넘는 성능을 선보이면서 영상 기반 결함 검사 기술이 다양한 분야에 적용되고 있다[2]. 본 논문에서는 비전 기반 양산라인 실시간 품질 검사를 위해 딥러닝 기술을 적용한 엣지컴퓨팅 비전 검사 장치를 소개한다. 엣지컴퓨팅 기반 비전 검사 장치는 인덕터 코어의 크랙 유무를 판별하기 위해 4 채널 조명을 제어하여 인덕터 코어 이미지 데이터를 취득하고 딥러닝 모델 기반 코어 크랙 유무 판별을 통해 인덕터의 불량 여부를 판별한다. 인덕터 코어 크랙 유무 검출을 위한 딥러닝 모델의 학습을 위한 데이터셋은 자체 제작한 비전 검사 장치를 통해 양품과 불량품 인덕터 샘플에 대한 이미지 데이터를 획득하였다. 딥러닝 기반 불량 유무 판별을 위하여 컨벌루션 신경망으로 이진 분류(Binary Classification) 모델을 구성한 뒤 자체적으로 확보한 데이터셋을 활용하여 학습을 수행하였다. 학습이 완료된 딥러닝 모델은 엣지컴퓨팅 플랫폼인 Nvidia Jetson Nano 보드에서 실시간 추론을 수행하여 실시간으로 불량 유무 판별이 가능한지 여부를 실험을 통해 확인하였다.

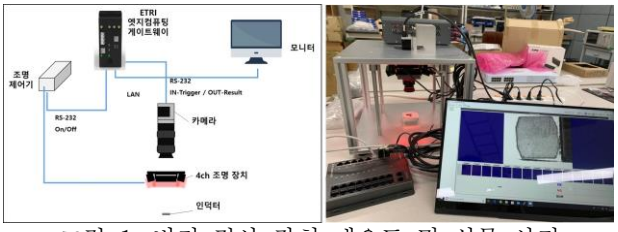


그림 1. 비전 검사 장치 개요도 및 실물 사진

Ⅱ. 비전 검사 장치

기존의 인덕터 제품 불량 여부를 판별하는 작업은 작업자가 단일 조명 아래에서 인덕터를 상하좌우로 움직여가며 코어 상단 패키지 부분의 크랙 또는 찍힘 여부를 확인하는 방식으로 수행된다. 이는 코어 상단의 크랙 또는 미세한 찍힘 여부가 조명의 각도에 따라서 보여지는 형태가 다르기 때문이다. 따라서, 인덕터 코어의 미세한 찍힘까지 판별할 수 있는 딥러닝 모델을 학습하기 위해서는 다양한 각도의 조명을 제어하여 다수개의 이미지 데이터 확보가 가능한 비전 검사 장치가 필요하다. 이를 위해, 우리는 4 채널의 조명 제어가 가능한 비전 검사 장치를 그림 1 과 같이 자체 제작하였다.

그림 1 과 같이 자체 제작한 비전 검사 장치를 이용하여 인덕터 코어 크랙 불량 검출을 위한 데이터셋을 확보하였다. 인덕터 코어 크랙 불량 검출 데이터셋은 총 1,563 개의 인덕터 샘플을 통해 수집하였으며, 각 샘플은 조명 각도에 따라 4 개의 이미지 데이터를 생성하여 총 6,252 개의 이미지 데이터를 생성하였다. 6,252 개의 이미지 중 정상 샘플

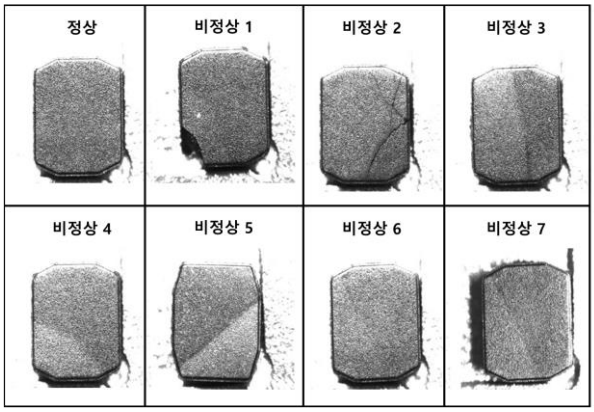


그림 2. 인덕터 코어 크랙 불량 검출 데이터셋 샘플

이미지는 5,444 개이고 비정상 샘플 이미지는 808 개이며, 딥러닝 학습용 데이터와 학습이 완료된 모델의 성능 검증을 위한 테스트 데이터의 비율은 1:1 로 설정하였다. 딥러닝 모델의 학습을 수행함에 있어서 딥러닝 모델의 과적합(overfitting) 여부를 확인하기 위한 검증용 데이터는 학습용 데이터의 20%를 따로 분리하여 사용하였다.

그림 2 는 비전 검사 장치를 통해 획득한 인덕터 코어 크랙 이미지 데이터셋 샘플이다. 그림 2 에서 정상 샘플은 코어 표면이 매끄러운 반면에 비정상 1 과 같이 코어의 상단부가 깨지거나 비정상 2 와 같이 크랙이 발생하거나 비정상 3~7 에서와 같이 찍힘이 발생한 샘플이 비정상 샘플 데이터로 확보되어 있다. 위의 그림에서 비정상 6 과 비정상 7 은 동일한 샘플에 대하여 조명 각도 변경에 따라 촬영된 이미지 데이터이다. 동일한 미세 불량에 대하여 비정상 6 이미지에서는 육안으로 확인이 어려운 반면에 다른 각도의 조명 환경에서 촬영된 비정상 7 이미지는 미세 불량이 좀 더 명확하게 드러나는 것이 확인 가능하다.

위와 같이 자체적으로 확보한 데이터셋을 활용하여 인덕터 코어 크랙 불량 검출 여부를 판별하는 딥러닝 모델 학습을 수행하였다. 딥러닝 학습은 Pytorch 프레임워크를 이용해 수행되었으며, 이미지 기반 데이터셋으로 불량 검출을 수행하기 때문에 컨벌루션 신경망으로 구성된 딥러닝 모델을 생성하였다. 그림 3 은 해당 딥러닝 모델에 대한 구조도를 나타낸다.

Ⅲ. 실험 결과

비전 검사 장치를 통해 수집한 인덕터 코어 크랙 불량 검출 데이터셋을 활용하여 그림 3 의 구조로 구성되어 있는 딥러닝 모델 학습을 수행하였다. 모델의 학습은 TITAN X 12GB GPU 가 장착되어 있는 딥러닝 서버(Intel i7 8core)에서 20 epochs 동안 수행되었으며, 단일 epoch 마다 테스트 데이터셋을 활용해 모델의 성능평가를 수행하였다. 표 1 은 20 epochs 중에서 가장 성능이 좋았던 19 epoch 학습 완료된 딥러닝 모델의 성능 지표 결과값을 나타낸다. 인덕터 코어 크랙 검출 데이터셋의 경우 정상 샘플과 비정상 샘플의 데이터 밸런스가 맞지 않기 때문에 F1-score 를 모델 성능의 지표로 사용하였다.

표 1 의 결과에서 알 수 있듯이 학습이 완료된 딥러닝 모델은 정상 및 비정상 샘플 판별에 대한 정밀도(Precision)는 100%를 달성하였으나 비정상 샘플에 대한 민감도(Recall)의 경우 98%를 달성함을 확인하였다. 비정상 샘플에 대한 민감도 98%가 의미하는 것은 비정상 샘플 중에서 2% 정도가 딥러닝 모델 추론

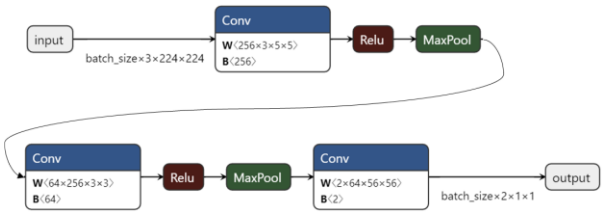


그림 3. 컨벌루션 신경망 기반 딥러닝 모델 구조

표 1. 인덕터 코어 크랙 판별 딥러닝 모델 추론 결과

	Precision	Recall	F1-score	Support
비정상	1.0	0.98	0.99	448
정상	1.0	1.0	1.00	2688

결과 정상으로 판별되고 있음을 의미한다. 인덕터 코어의 결함으로 인한 피해를 최소화하기 위해서는 정상 샘플에 대한 민감도를 손해보더라도 비정상 샘플에 대한 민감도를 100% 수준으로 향상시키는 것이 바람직하다. SVM 과 같은 기존의 기계학습 방법의 경우 매개 변수를 조정하는 것이 가능하지만 딥러닝 모델의 경우 loss 함수에서 비정상 샘플을 정상으로 인식하는 케이스에 대한 loss 값에 가중치를 줘서 조절이 가능할 것이라 예상된다.

학습이 완료된 딥러닝 모델을 엣지컴퓨팅 환경에서 실험하기 위하여 Nvidia Jetson Nano 보드에서 추론 성능을 측정하였다. Jetson Nano 보드에서의 딥러닝 모델의 추론 실험을 위해 오픈소스 프로젝트인 TVM 컴파일러를 이용하여 Pytorch 모델을 TVM 런타임에서 실행가능하도록 컴파일을 수행했으며, 실험에 사용된 Jetson Nano 는 cuda 10.2 버전을 기반으로 한 Jetpack 4.4 이미지를 사용하였다. 딥러닝 서버에서 추론 성능은 약 83FPS 정도가 나오지만 Nvidia Jetson nano 보드에서는 약 9FPS 로 처리가 가능함을 확인하였다.

Ⅳ. 결론

본 논문에서는 공장 양산라인에서 생산되는 인덕터의 크랙 여부를 실시간으로 판별하기 위한 딥러닝 기반 엣지컴퓨팅 비전 검사 장치를 소개하였다. 비전 검사 장치를 개발함에 있어서 자체 제작한 4 채널 조명 제어 가능 비전 검사 장치를 이용해 인덕터 코어 크랙 검출을 위한 데이터셋을 확보하였고, 컨벌루션 신경망으로 이뤄진 이진 분류 딥러닝 모델을 학습시킨 뒤 엣지컴퓨팅 플랫폼에 적용하여 실시간으로 인덕터 코어 크랙 유무 판별이 가능함을 확인하였다.

ACKNOWLEDGMENT

본 연구 논문은 한국전자통신연구원 연구운영지원 사업의 일환으로 수행되었음. [호남권 지역산업 기반 ICT 융합기술 고도화 지원사업 (21ZK1100)]

참 고 문 헌

[1] K. Yun *et al.*, "Vision-based garbage dumping action detection for real-world surveillance platform," in *ETRI Journal, vol. 41, no. 4, pp. 494-505, Aug. 2019.*

[2] 이건영, 정선재, 서기성, "결함 검출을 위한 임베디드 딥러닝 시스템," 전기학회논문지, 69(2), pp. 325-330, 2020.02.

한국지형에서의 카메라 기반 실시간 3차원 객체인식 기술 연구

한지은, 고석현, 임예원, 이성진
동서울대학교 전자공학과
haneul2069@du.ac.kr, sungjinlee@du.ac.kr

A Study on the Camera based 3D Object Detection in Korea

Jieun Han, Sukhyun Ko, Yewon Lim, Sungjin Lee
Department of Electronic Engineering, Dong Seoul University

요 약

자율주행 기술의 실현을 위해 필수적인 3차원 객체 인식 기술에는 카메라 기반 방식과 Lidar 기반 방식의 두 가지 접근법이 있다. 자율주행기술은 안전성이 중요시되기 때문에 빛 반사, 어두움, 착시 현상 등의 영향으로 상황에 따라 불안정할 수 있는 카메라 기반의 3D Object Detection 기술보다는 Lidar 기반 방식이나 그 둘의 Multi-Sensor Fusion 기술이 주로 다루어지고 있다. 본 논문에서는 다양한 환경에서 카메라 기반의 3D Object Detection 기술의 성능을 측정해보고 그 가능성을 분석해 보았다. 특히, 한국 지형과 여러 주행 환경 (도심, 교외, 고속도로) 에 따라서도 안정적으로 잘 동작할 수 있는지를 평가하기 위해 자체 데이터 셋을 구축하여 여러 상황에 따른 실험을 해 보고 각각의 시나리오에서 좀 더 성능을 높이기 위한 전략들을 도출하였다.

Ⅰ. 서론

최근 딥러닝 기술의 발전으로 자율주행 기술은 운전 보조 기능을 넘어서 완전 자율주행 단계에 이르고 있다. 그런 완전 자율주행을 위해서는 주변 상황에 대한 3차원 인식과 이을 통한 주행가능 경로 파악이 필수적이여서, 관련된 3D Object Detection 기술에 산업과 학계가 주목하고 있다 [1][2].

특히 이런 3D Object Detection 기술에는 카메라 기반 방식과 Lidar기반 방식의 두 가지 접근법이 있다. 우선 카메라 기반 방식은 상대적으로 데이터 양이 적고 해당 객체들의 특징에 대한 표현력이 우수하여 좋은 환경에서는 높은 정확도에 실시간 처리가 가능하다. 하지만 다양한 광학적 어려움, 즉 빛 반사, 어두움, 착시 현상 등의 영향으로 상황에 따라 불안정한 성능을 낼 수 있는 게 사실이다. 반면 Lidar기반 방식은 해당 객체들의 3D 위치를 Point Cloud Data 로 상대적으로 정확하게 측정할 수 있지만 해당 데이터의 양이 많고 표현력 (Intensity) 이 상대적으로 부족하여 Lidar만으로는 객체의 class를 정확하게 분별해 내기 힘들어서 이를 보강하기 위해 카메라와 함께 사용하는 Multi-Modal Sensor Fusion 방식을 주로 사용한다 [1][2]. 하지만 실시간으로 처리되어야 하는 자율주행 상황에서는 두 입력신호 각각에 대한 3D Object Detection 적용 및 Multi-Modal 신호처리를 위한 정합과정등을 위한 느린 동작 속도에 대한 위험성이 존재한다.

그러므로 실시간 처리 및 정확도 모두를 확보할 수 있도록 본 논문에서는 카메라 기반 3D Object Detection 기술에 대해 집중하여 해당 기술의 불리한 상황에서도 정확도를 높일 수 있는 방법에 대해 연구하였다. 특히 지역적으로 훈련 데이터와 다른 상황들, 즉 한국 지형과 여러 주행 환경 (도심, 교외, 고속도로), 시간대 (낮, 밤) 에서 어느 정도의 성능 열화가 발생하는지 분석하였고 이런 환경 분석을 위해 자체 데이터 셋을 구축하여 시나리오 별 성능 향상을 위한 기법들을 제시하였다.

Ⅱ. 3D Object Detection Model

본 논문에서는 3D Object Detection 모델 중 Fast Detection 성격에 있는 RTM3D 모델 [3]에 기반하여 성능 분석을 수행하였다. 해당 모델은 Fast Detection을 위해 입력이미지로 1장만을 사용한다. 그림 1과 같이 1장의

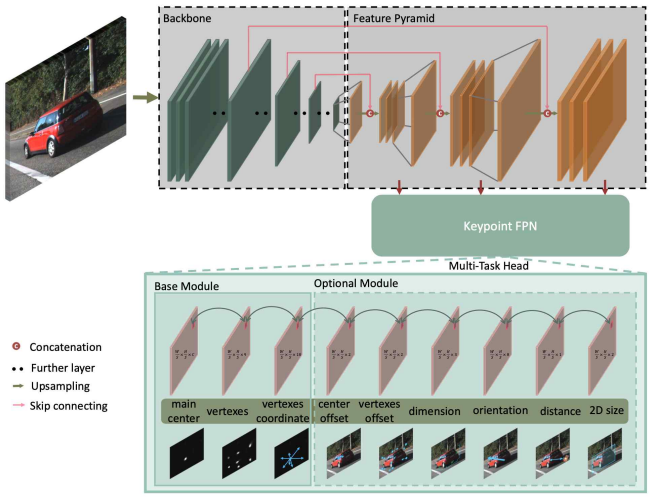


그림 1. RTM3D 의 동작 구조도 [3]

이미지를 입력받아 FPN (Feature Pyramid Network) 과 Multi-Task Head을 통해 9개의 Keypoint (main center, vertexes, vertexes coordinate, center offset, vertexes offset, dimension, orientation, distance, 2D size) 에 대한 Heatmap을 얻어낸다. 결과적으로 3D Bounding Box B를 위한 Rotation R, Position T, Dimension D를 얻어낸다. 즉, 기존 2D Object Detection에서 많이 사용하는 FPN 방식에 2D Keypoint Feature들을 얻어내고 이를 3D Bounding Box에 Projection 함으로서 성능 도출 시간을 단축시켰다고 볼 수 있겠다.

Ⅲ. 데이터 셋 구축

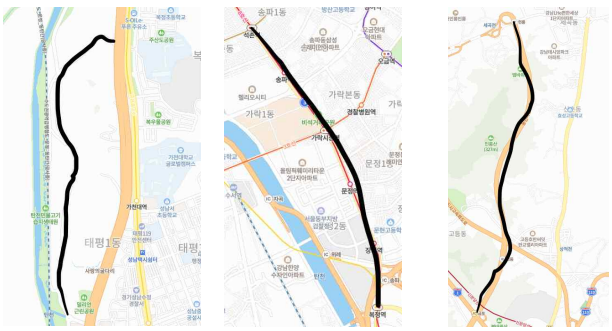
3D Object Detection을 위한 기본 훈련 데이터 셋으로 KITTI [4] 를 사용하였다. 하지만 본 데이터 셋은 미국 지형의 낮 환경 도로 주행에서 만들어진 것이기에 한국 지형과 여러 주행환경 및 밤 도로 환경에 적용하기에는 성능 열화가 발생할 수 있다. 그러므로 본 논문에서는 한국의 도로환경에서 다양한 주행 시나리오와 낮, 밤 시간대를 고려하여 주행 데이터 셋을 구축하였다. 그림 2에서는 데이터 셋 구축을 위해 사용한 기기들의 스펙을 나타내었다.



그림 2. 데이터 셋 구축을 위해 사용된 기기들

위 구축된 기기들 중 주행을 위해서 직접적으로 영향을 미치는 왼쪽 카메라 (운전자 기준)에서 수집되는 영상 이미지를 통해 데이터 셋을 구축하였다. 카메라는 FHD (1920x1080) 30 FPS 126° 화각을 지닌 WebCam을 사용하여 영상을 획득하였다. 경로는 동서울대학교를 중심으로 3가지 코스 즉, 교외 지형, 도심 지형, 고속도로 지형에 대해 낮 (12시30분~1시30분) 과 밤 (8시~9시) 시간대 각각에 대해 데이터를 획득하였다.

또한 기존의 데이터 셋은 교통상황, 도로 상황에 대한 구분이 없어서 새로운 데이터 셋을 구축하였다. 다양한 상황에 따른 정확도를 검증하기 위해 다음 그림 3과 같은 코스에 따라 그림 2의 기기들을 이용하여 실험 데이터 셋을 수집하였다.



(a) Urban Area (b) Downtown Area (c) Highway Area

그림 3. 도로 유형에 따른 3가지 주행 시나리오

Ⅳ. 실험

실험을 위해 다음과 같은 Evaluation Set을 정하여 진행하였다. 우선 기본 KITTI 데이터 셋에 대해

- KITTI dataset : 기본 KITTI 데이터 셋에서 훈련을 진행

을 정하였다. 또한 그림 3에서 보듯이 다음의 3가지 상황에 따라 코스를 결정하였다.

- Urban Area Driving dataset: 교외지역으로 교통이 한적하고 간단한 총 2차선 도로 주행 데이터 셋
- Downtown Area driving dataset: 도심지역으로 교통이 복잡한 총 8차선 도로 주행 데이터 셋
- Highway Area driving dataset: 주행 속도가 빠른 고속도로 주행 데이터 셋

해당 데이터 셋에 대해 IoU 0.5 이상을 Positive의 기준으로 삼았다.

Train	Test			
	KITTI	UA	DA	HA
KITTI	54	51	45	50
UA	51	54	52	53
DA	53	54	54	54
HA	52	53	51	53

Table 1. 4가지 주행시나리오에 대해 정확도 (Average Precision) 실험

위 실험결과에서 보듯이 차량의 밀도가 높은 한국 도심지역 (Downtown)에서의 성능이 일반적으로 열화되어 있는 것을 확인할 수 있다. 하지만, train 셋에 관련 한국 도심지역 데이터를 더 넣어서 훈련 시키면 그 정확도 손실은 다소 복구할 수 있는 것을 알 수 있다. 그 이유로 미국에서 주로 주행되는 차량의 모양과 한국에서 주로 주행되는 차량의 모양이 꽤 달라서 (예, 한국의 시내 버스) 한국 고유의 차량들은 잘 인식이 안되기 때문이다. 하지만, 이런 데이터를 추가로 훈련 데이터에 넣어서 훈련시키면 그 정확도 손실이 복구된다는 것을 알 수 있다.

반면, 교외지역 (Urban) 이나 고속도로 지역 (Highway) 과 같은 곳에는 차량의 밀도가 다소 적고 Kitti 와 비슷한 승용차 위주의 차량들이 많아서 원본 데이터 (Kitti)로 훈련 시킨 것과 거의 비슷한 정확도를 나타내는 것을 알 수 있다.

Ⅳ. 결론

본 논문에서는 미국지형에서 훈련된 KITTI 데이터 셋으로 한국지형에서 3D Object Detection 적용을 위한 시나리오 별 성능평가를 수행하였다. 실험결과 주행되는 차량의 종류, 밀도가 성능평가에 영향을 준다는 것을 알 수 있었고 이를 극복하기 위해서는 추가적인 데이터 셋 확보와 훈련이 중요한 전략이라는 것을 보였다.

ACKNOWLEDGMENT

이 논문은 2021년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기본연구사업임(No. NRF-2019R1F1A1062878)

참 고 문 헌

[1] Rui Qian, Xin Lai, Xirong Li, “3D Object Detection for Autonomous Driving: A Survey”

[2] Edardo Arnold, etc, “A Survey on 3D Object Detection Methods for Autonomous Driving Applications”

[3] Peixuan Li, Huaici Zhao, Pengfei Liu, Feidao Cao, “RTM3D: Real-time Monocular 3D Detection from Object Keypoints for Autonomous Driving”

[4] http://www.cvlibs.net/datasets/kitti/

생성적 적대 신경망 기반 Image-to-Image translation 기술 연구 동향

황도연, 김재석, 최윤호
부산대학교
h1d2y3@pusan.ac.kr, tjr3117@pusan.ac.kr, yhchoi@pusan.ac.kr

Research trend of Image-to-Image translation technology based on generative adversarial neural network

Hwang Do-Yeon, Kim Jae-Seok, Choi Yoon-Ho
Pusan National Univ.

요 약

인공지능 기술이 발달하고 다양한 분야에 적용됨에 따라 특정 도메인(domain)을 가진 이미지를 다른 도메인을 가진 이미지로 변환하는 Image-to-Image translation 기술이 발달하였다. 특히 생성적 적대 신경망(generative adversarial network, GAN)이 적용된 Image-to-Image translation 모델들은 성능 지표에서 뛰어난 향상을 보여주었으며 이를 통해 사용자들에게 편의를 제공하였다. 본 논문은 GAN을 적용한 Image-to-Image translation 모델들을 조사하였으며 해당 모델들의 목표와 작동 방식을 분석하였다.

Ⅰ. 서 론

Image-to-Image translation은 특정 도메인(domain)을 가진 입력 이미지를 목표 도메인을 가진 출력 이미지로 변환하는 것을 목표로 하는 분야이다. 최근 인공지능 기술이 발전하고 인공지능 모델들이 Image-to-Image translation 분야에 적용됨에 따라 해당 분야에서도 뚜렷한 성능 개선이 이루어지고 있다. 특히 인공지능 모델 중 하나인 생성적 적대 신경망(generative adversarial network, GAN)은 경쟁적 학습을 통해 양질의 이미지를 생성할 수 있어 이미지를 출력해야 하는 Image-to-Image translation 분야에서 적극적으로 사용되고 있다. 본 논문은 현재 뛰어난 성능을 보여주고 있는 GAN 기반 Image-to-Image translation 모델들의 목표 및 작동방식을 분석하였다.

본 논문의 구성은 다음과 같다. 2장에서는 GAN 기반 Image-to-Image translation 모델들을 간략하게 소개하고 해당 모델의 목표 및 작동 방식을 설명한다. 3장에서는 해당 내용을 바탕으로 본 논문의 결론을 맺는다.

Ⅱ. 본론

GAN이 적용된 Image-to-Image translation 모델들은 generator와 discriminator의 경쟁적 학습을 통한 입력 이미지와 출력 이미지 간 매핑 함수 학습을 목표로 한다. 본 장에서는 GAN의 학습과정을 바탕으로 각 모델들의 목표 및 작동 방식에 대해 분석하였다.

2.1 Pix2Pix

Convolutional neural network(CNN)를 사용한 기존 Image-to-Image translation 모델들은 출력 이미지의 평균을 구하는 방식으로 학습되어 흐릿한 출력 이미지를 생성하는 한계를 가지고 있다. 이러한 문제를 해결하기 위해 Isola 등은 GAN을 통해 Image-to-Image translation을 수행하는 pix2pix를 제안하였다[1]. Pix2pix는 generator와 discriminator로 구성되어 있으며 generator는 입력 이미지를 변환하고 discriminator는 이미지에 대해 실제/거짓 여부를 판단한다. Generator는 discriminator를 속이는 것을 목표로 하며 discriminator는 이미지를 정확히 감별하는 것을 목표로 한다. 이러한 두 모듈의 경쟁적 학습은 generator가 실제 이미지와 유사하게 이미지를 변환할 수 있도록 하며 이를 통해 출력 이미지가 흐릿하게 생성되는 문제를 해결한다. 또한 Pix2pix는 conditional GAN을 사용하여 여러 개로 세분화된 Image-to-Image translation의 다양한 문제들을 하나의 모델로 해결할 수 있다는 장점을 가지고 있다.

2.2 CycleGAN

Pix2pix는 비교적 선명한 출력 이미지를 생성할 수 있지만 학습을 위해서 paired data가 필요하다는 단점이 있다. 이러한 문제를 해결하기 위해 Zhu 등은 unpaired data로도 학습이 가능한 CycleGAN을 제안하였다[2]. Unpaired data는 ground truth가 없으므로 경쟁적 학습만을 사용하여 이미지 변환을 학습할 경우, discriminator는 속일 수 있지만 입력 이미지와 무관한 이미지로 변환되는 mode collapse가 발생할 수 있다. 이러한 문제를 해결하기 위해 CycleGAN은 두 개의 generator-discriminator 쌍으로 구성된 네트워크에서 cycle consistency loss를 적용하여 학습을 수행한다. 먼저, generator G는 도메인 X에 해당하는 실제 이미지 x를 도메인 Y의 분포를 따르는 $G(x)$로 변환한다. 반대로, generator F는 도메인 Y에 해당하는 이미지 y를 도메인 X의 분포를 따르는 $F(y)$로 변환한다. 이때, G와 F는 서로 역함수이므로 $G(x)$는 F를 통해 x로 복원될 수 있다. Cycle consistency loss는 복원된 $F(G(x))$와 x의 차이를 통해 계산되며 이를 통해 CycleGAN은 두 이미지가 최대한 유사하도록 학습된다. 두 이미지가 유사하기 위해서는 F로부터 x가 복원될 수 있도록 $G(x)$가 x의 특성을 보존하고 있어야 한다. 따라서 cycle consistency loss를 통해 학습이 수행될 경우, G는 Y의 분포를 따르는 동시에 x의 특성을 보존하는 변환을 수행할 수 있도록 학습된다.

2.3 UNIT

확률적 모델링 관점에서 unpaired data를 이용한 Image-to-Image translation 학습은 각 주변분포(marginal distribution)를 따르는 두 개의 이미지 집합을 사용하여 결합분포(joint distribution)를 추론하는 것으로 볼 수 있다. Liu 등은 해당 문제를 해결하기 위해서는 결합분포에 대한 추가적인 가정이 필요함을 지적하며 특수한 가정을 바탕으로 이미지를 생성하는 UNIT을 제안하였다[3]. 구체적으로, UNIT은 대응되는 한 쌍의 이미지가 shared-latent space에서 동일한 latent representation으로 사상될 수 있다는 가정을 사용한다. 해당 가정이 적용된 이미지 변환을 수행하기 위해 UNIT은 두 개의 generator-discriminator 쌍에 두 개의 encoder E_1, E_2를 추가하여 latent representation을 찾을 수 있도록 하였다. 이러한 구조에서 입력 이미지 x는 E_1을 통해 latent code인 z로 변환되며 generator G_1은 z를 입력으로 출력 이미지를 생성한다. 여기서 E_1-E_2, G_1-G_2는 각각 대칭되는 모듈과 일부 레이어의 가중치(weight)를 공유하며, 이를 통해 앞선 가정을 충족하는 이미지 변환이 수행될 수 있도록 한다.

2.4 AGGAN

CycleGAN은 unpaired data를 이용한 Image-to-Image translation을 가능하게 했으나 이미지 변환을 수행할 경우, 변환의 대상이 되는 전경(foreground) 뿐만 아니라 배경(background) 또한 변환되는 문제가 있다. 구체적으로 CycleGAN은 이미지 전체에 대한 기존(source) 도메인 이미지의 분포와 목표(target) 도메인 이미지의 분포 간의 차이를 줄이는 데만 집중하여 이미지 내부에서 변환의 목표가 되는 지역을 파악하기 힘들다는 한계를 가진다. 이러한 문제를 해결하기 위해 Mejatti 등은 attention 모듈을 적용한 Image-to-Image translation 모델인 AGGAN을 제안하였다[4]. AGGAN은 CycleGAN과 같은 두 개의 generator-discriminator 쌍에 두 개의 attention 모듈인 A_s, A_t를 추가하였으며 G를 통해 x를 $G(x)$로 변환하고 A_s를 통해 x에 대한 attention map인 $A_s(x)$를 생성한다. 최종 이미지 x'는 다음과 같은 방법을 통해 생성된다.

$$x' = A_s(x) \odot G(x) + (1 - A_s(x)) \odot x \quad (1)$$

(1)에서 $\odot$는 성분별 곱(element-wise product)을 의미하며 $A_s(x) \odot G(x)$는 이미지의 전경을 의미하고 $(1 - A_s(x)) \odot x$는 이미지의 배경을 의미한다. AGGAN은 attention 모듈을 통한 전경과 배경을 구분하는 이미지 변환을 수행함으로써 우수한 품질의 출력 이미지를 생성할 수 있도록 하였다.

2.6 StarGAN

언급된 Image-to-Image translation 모델들은 두 개의 도메인에 대해서만 이미지 변환이 가능하다는 점에서 확장성이 부족하였다. Choi 등은 이러한 문제를 해결하기 위해 하나의 모델로도 다양한 도메인에 대한 이미지 변환이 가능한 StarGAN을 제안하였다[5]. StarGAN은 기존 모델과는 다르게 generator에 입력 이미지뿐만 아니라 목표 도메인 레이블도 함께 입력으로 사용된다. Discriminator는 주어진 이미지에 대해 실제/거짓 여부뿐만 아니라 해당 이미지의 도메인 또한 판단한다. 새롭게 추가된 generator와 discriminator의 역할을 학습하기 위해 StarGAN은 다음과 같은 domain classification loss를 손실 함수로서 추가하였다.

$$L_{cls}^{r} = E_{x,c'}[-\log D_{cls}(c'|x)] \quad (2)$$

(2)는 domain classification loss에서 discriminator를 학습시키기 위한 부분이다. StarGAN의 discriminator는 실제 입력 이미지 x가 주어질 경우 해당 이미지에 대응하는 도메인인 c'을 정확히 예측하는 것을 목표로 학습된다.

$$L_{cls}^{f} = E_{x,c}[-\log D_{cls}(c|G(x,c))] \quad (3)$$

(3)은 domain classification loss에서 generator를 학습시키기 위한 부분이다. x를 입력으로 generator를 통해 c 도메인으로 변환된 이미지 $G(x,c)$가 주어질 경우, 해당 이미지는 discriminator에 의해 c로 분류되는 것을 목표로 학습된다. StarGAN은 이러한 학습과정을 통하여 단일 모델이 입력 이미지를 다양한 도메인의 이미지로 변환할 수 있도록 하였다.

Ⅲ. 결론

본 논문은 GAN을 사용한 Image-to-Image translation 모델들인 pix2pix, CycleGAN, UNIT, AGGAN, StarGAN을 분석하였다. 각 모델들은 발전과정에 따라 새로운 모듈과 학습방식을 제안하였으며 이를 통해 기존 모델들의 문제를 해결하고 성능을 개선하였다. 추후 이루어질 Image-to-Image translation 연구에서도 기존 모델의 한계를 진단하고 해당 문제를 해결하기 위한 합리적인 방법론이 제안되어야 할 것이다.

ACKNOWLEDGMENT

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원(2019-0-01343,융합보안핵심인재양성사업) 및 정보통신기획평가원의 대학ICT연구센터육성지원사업(IITP-2021-2020-0-01797)에 의하여 지원되었음.

참 고 문 헌

[1] ISOLA, P., Zhu, J. Y., Zhou, T., and Efros, A. A. “Image-to-image translation with conditional adversarial network,”. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1125-1134, 2017.

[2] Zhu, J. Y., Park, T., Isola, P., and Efros, A. A. “Unpaired image-to-image translation using cycle-consistent adversarial networks,” In: Proceedings of the IEEE international conference on computer vision. pp. 2223-2232, 2017.

[3] Liu, M. Y., Breuel, T., and Kautz, J. “Unsupervised image-to-image translation networks. In: Advances in neural information processing systems,” pp. 700-708. 2017.

[4] Alami Mejjati, Y., Richardt, C., Tompkin, J., Cosker, D., and Kim, K. I. “Unsupervised Attention-guided Image-to-Image Translation,” Advances in Neural Information Processing Systems, pp. 3693-3703, 2018.

[5] Choi, Y., Choi, M., Kim, M., Ha, J. W., Kim, S., and Choo, J. “Stargan: Unified generative adversarial networks for multi-domain image-to-image translation,” In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 8789-8797, 2018.

이미지 노이즈 제거를 위한 다중 오토인코더 기반 점진적 복원에 관한 연구

김형선, 허청환, 이한음, 황현택, 강상길*
인하대학교

hskim@inha.edu, princehur@naver.com, vmfksxp89@naver.com, hwanght@inha.edu, *sgkang@inha.ac.kr

A Study on the Progressive Restoration Based on Multi-Autocoder for Image Noise Removal

Kim Hyeong Seon, Hur Cheong Hwan, Lee Han Eum, Hwang Hyeon Taek, Kang Sang Gil*
Inha Univ.

요 약

본 논문은 딥 러닝 접근 방식을 통해 저하된 이미지 복원 과정을 크게 세 단계로 나누어 점진적으로 복원하는 방법을 제안한다. 먼저 U-Net 구조를 활용하여 저하된 이미지의 전체적인 노이즈 제거 작업을 한 다음 이미지 내의 핵심 요소를 찾아 두 개의 VAE 를 활용하여 부분적인 노이즈를 제거하고 선명도를 향상시킨다. 전체와 핵심부분에 대해 각각 노이즈 제거 작업을 완료한 이후 전체 이미지에 대해 로컬 정보를 유지한 고화질의 피처맵과 결합하여 해상도 복원 작업을 한다. 이 방법론을 통해 노이즈가 포함된 인물 이미지 복원 작업에서 좋은 성능을 달성했다.

Ⅰ. 서 론

이미지 복원 연구[1, 2, 3]는 다양한 데이터셋에서 다양하게 발전해왔다. 이미지 복원 연구 중에서도 노이즈 제거를 위한 데이터에는 빗자국이 포함된 이미지 데이터셋[4]과 인위적으로 Gaussian 노이즈를 입힌 데이터셋[5, 6] 등이 있지만 인물 사진에 반복적인 패턴의 노이즈가 강하게 포함된 데이터는 없었다. 윤곽이 파악되면서 SSIM(Structural Similarity Index) 수치가 0.5 이하로 측정되는 데이터셋을 구축하였고, 해당 데이터셋에서 기존의 연구들에 비해 훨씬 좋은 성능을 보이는 모델을 제안한다.

Ⅱ. 본론

그림 1 은 본 연구에서 제안하는 다중 오토 인코더 기반 점진적 이미지 복원 네트워크의 시스템 구조도이다. 시스템은 크게 U-Net 구조[7]를 활용하여 전체적인 노이즈 제거 작업을 진행하는 단계(Step 1)와 이미지 내의 핵심 요소인 얼굴 부분을 찾고 두 개의 VAE 를 활용하여 얼굴의 세부 요소들을 복원하는 단계(Step 2), 전체 이미지를 대상으로 해상도 복원 작업을 진행하는 단계(Step 3)로 구성되어 있다.

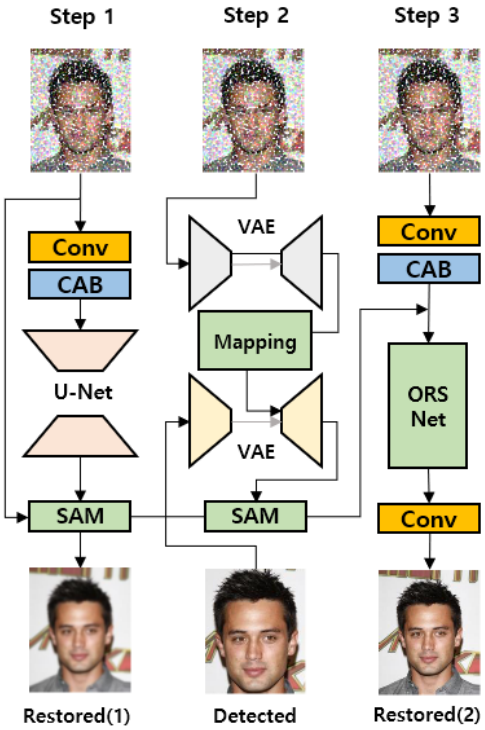


그림 1. 시스템 구조도

Step 1 은 U-Net[7]과 SAM[8, 9]으로 구성되어 있다. 먼저, Encoder-Decoder 기반의 네트워크를 기반으로 U-Net 구조[6]를 활용하여 Encoder 에서 Context 를

학습하고, Skip Connection 을 통해 Decoder 에서 Feature Map 과 Context 를 결합하여 정확한 Localization 을 한다. 그리고 SAM(Supervised Attention Module)[4, 5]을 통해 이후 단계로 중요 정보를 전달한다. 여기서 SAM 은 이미지 복원에서 중요한 부분들을 표현한 Attention Guided feature 를 생성하고 모델이 집중해야 하는 노이즈 부분을 강조하여 Step 3 에서 활용한다.

Step 2 는 먼저 이미지의 얼굴 부분을 Detection 하고 고해상도의 이미지와 함께 두 개의 VAE(Variational Auto Encoder)[3]를 훈련시켜 각각의 이미지를 변환하고, 두 VAE 로부터 나온 두 개의 이미지를 합성쌍 데이터로써 학습한다. 여기서는 Pascal VOC dataset 을 랜덤하게 둘로 나누어 하나는 도메인 A, 나머지는 도메인 B 로 지정하여 각각의 VAE 학습에 활용한다. LSGAN[2]의 loss function Least-square-loss 를 이용하여 두 VAE 로부터 나온 이미지의 얼굴 부분을 합성함으로써 최적의 결과값으로 이미지를 복구한다.

Step 3 는 Step 1, 2 에서 노이즈를 제거한 이미지에 Super Resolution 작업을 하는 단계이다. 여기서 ORSnet[9]은 3 개의 ORB(Original Resolution Block)으로 구성되어 있고, 한 개의 ORB 는 8 개의 CAB(Channel Attention Block)으로 구성되어 어떤 채널이 이미지 내의 노이즈를 제거하는 데에 중요한지 파악하고 High resolution feature 들을 생성하여 이미지의 해상도를 복원한다.

기존의 Denoising 기술로 이미지 복원이 어려운 정도의 이미지 데이터를 만들기 위해 SSIM 수치가 0.5 이하로 측정되도록 하여 약 2,000 장의 이미지 데이터를 수집하였다. 이 중 1,550 장은 훈련용으로, 나머지 450 장은 테스트용으로 사용하였다.

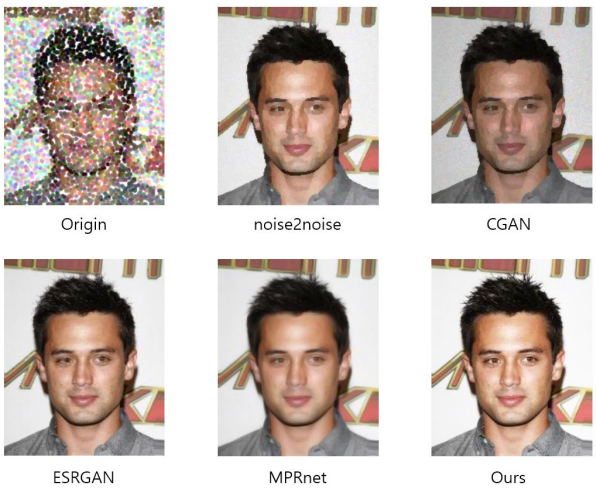


그림 2. 다른 Denoising 기술과 비교

결과적으로 이미지 유사도를 알아볼 수 있는 SSIM 수치를 측정하여 테스트 데이터 전체 평균 0.8804 의 수치를 달성하였고, 본 연구 이외의 다른 Denoising 기술 중 가장 높은 수치인 MPRnet 의 0.8189 보다 훨씬 더 향상된 성능을 보였다.

Ⅲ. 결론

본 논문에서는 기존에 존재하는 다단계 이미지 복원 기술의 과정에 이미지 내에서 핵심적인 요소를 찾고, 그 부분에 더 집중하는 복원 과정을 포함시켜 전체적인 이미지의 복원도를 향상시키는 시도를 하였다. 그 시도를 통하여 기존의 이미지 복원 연구에서 다루지 않았던 강한 노이즈가 포함된 이미지 복원을 성공적으로 마쳤다.

다양한 종류의 강한 노이즈가 포함된 이미지 데이터 중 밝기의 심각한 저하로 인해 첫 번째 단계의 복원 작업으로 인물이나 물체의 윤곽선이 뚜렷하게 복원되기 힘든 경우에 다른 이미지에 비해 저조한 결과를 보였다. 향후 연구에서 우리는 노이즈 제거 작업과 핵심 요소 복원 작업 사이에 이미지의 밝기를 적절히 조절하여 물체의 윤곽선을 드러내기 위한 단계를 추가하여 형체를 알아보기 힘들 정도로 어두운 이미지에서도 좋은 복원 성능을 보이는 모델 개발을 위한 연구를 다루고자 한다.

ACKNOWLEDGMENT

“이 논문은 2019 년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업이다. (NRF-2019R1A2C1008048)”

참 고 문 헌

[1] Lehtinen, Jaakko, et al. "Noise2noise: Learning image restoration without clean data." arXiv preprint arXiv:1803.04189 (2018).

[2] Mao, Xudong, et al. "Least squares generative adversarial networks." Proceedings of the IEEE international conference on computer vision. 2017.

[3] D. P. Kingma and M. Welling, “Auto-encoding variational bayes,” arXiv preprint arXiv:1312.6114, 2013.

[4] Wang, Tianyu, et al. "Spatial attentive single-image deraining with a high quality real rain dataset." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.

[5] Zhang, Kai, et al. "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising." IEEE transactions on image processing 26.7 (2017): 3142-3155.

[6] Slepian, David. "The one-sided barrier problem for Gaussian noise." Bell System Technical Journal 41.2 (1962): 463-501.

[7] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015.

[8] Zhang, Yulun, et al. "Image super-resolution using very deep residual channel attention networks." Proceedings of the European conference on computer vision (ECCV). 2018.

[9] Waqas Zamir, Syed, et al. "Multi-Stage Progressive Image Restoration." arXiv e-prints (2021): arXiv-2102.

하이브리드 트랜스포머 기반 비지도 단안 카메라 영상 깊이 추정 방법

황승준, 박성준, 김규민, 백중환
한국항공대학교
s-eungju-n@hanmail.net, jdwns1011@naver.com, gyumin93@gmail.com, jhbaek@kau.ac.kr

Unsupervised Monocular Depth Estimation Using Hybrid Transformer

Seung-Jun Hwang, Sung-Jun Park, Gyu-Min Kim, Joong-Hwan Baek
Korea Aerospace Univ.

요 약

본 논문은 비지도 학습 기반 단안 카메라 영상 깊이 추정에 CNN과 비전 트랜스포머 네트워크를 혼합한 하이브리드 네트워크를 적용하는 방법에 대해 기술한다. 최근의 연구는 CNN은 컨볼루션 연산을 통해 지역적 특징을 추출에 적합하며, 비전 트랜스포머는 다중 자기주의 모듈에 기반한 광역적 특징추출에 적합하다고 알려져 있다. 본 연구에서는 일반적인 비지도 깊이 추정 네트워크의 인코더-디코더 구조에서 지역적 특징을 추출하는 단계에서는 CNN을 사용하고, 광역적 특징을 추출하는 단계에서는 비전 트랜스포머를 사용하는 하이브리드 모델 기반 깊이 추정 네트워크를 제안한다. 제안하는 네트워크는 실험을 통해 기존의 네트워크보다 적은 파라미터와 연산량으로 기존 네트워크와 유사한 성능을 보임을 확인하였다.

Ⅰ. 서 론

영상에서 3D 깊이 정보를 예측하는 기술은 자율주행, AR, VR 및 로봇 비전 등 3D 공간 지도 생성이 요구되는 분야에 활발히 적용되고 있다. 기존의 센서 방식의 깊이 추정 방식은 정확도는 높지만, 높은 가격과 낮은 해상도의 문제점이 존재한다. 최근의 카메라 기반의 깊이 추정 연구는 상당히 활발히 진행되었으며, 성능 또한 드론과 같은 상용 제품에 적용될 정도로 높은 수준을 보인다.

카메라 기반의 깊이 추정 연구는 센서로부터 취득된 데이터를 직접 학습하는 지도 학습 모델과 인접한 프레임 두 프레임의 이미지에서 다른 한쪽 이미지를 재구성하는 시점 합성을 통해 깊이를 추정하는 비지도 학습 모델로 나누어진다[1]. 지도 학습 모델이 데이터를 직접 학습하기 때문에 더 높은 성능을 보이고 있으나, 센서로부터 3D 데이터를 획득할 수 없는 경우에는 지도 학습 모델 적용의 어려움이 있다. 최근의 연구는 비지도 학습 모델 방식도 지도학습 방식과 유사한 수준의 성능을 보인다.

또한, 최근의 비전 트랜스포머 모델은 이미지 분류, 분할, 검출 분야에서 최고성능을 달성하였다[2, 3]. 비전 트랜스포머는 패치단위의 임베딩 벡터를 쿼리, 키, 값으로 나누어 연산하는 다중 자기주의 메커니즘과 MLP(Multi Layer Perceptron)로 구성되어 있다. 이를 지도학습에 적용한 연구에서는 CNN 기반의 Resnet과 비전 트랜스포머 모델을 혼합한 하이브리드 모델이 가장 좋은 성능을 보임을 실험으로 증명하였다[4].

본 논문에서는 비지도 깊이 추정 모델에서 효율적인 깊이 네트워크 추정을 위한 하이브리드 트랜스포머 기반 깊이 추정 네트워크를 제안한다. 기존 깊이 추정 네트워크의 인코더-디코더 구조에서 지역적 특징 추출을 위한 CNN기반의 Resnet 모델과 광역적 특징 추출을 위한 트랜스포머의 혼합 모델을 제안한다.

Ⅱ. 본 론

본 논문에서는 최근의 연구에 따라 비지도 학습 기반 깊이 추정을 위해 깊이 및 포즈 네트워크를 동시에 학습한다[5]. 네트워크는 타겟 이미지 I_t와 소스 이미지 I_s로부터 타겟 이미지 시점으로 재구성된 이미지 $\hat{I}_{s\ t}$사이의 재구성 에러 L_p를 최소화하는 뷰합성 과정을 거치면서 학습이 이루어진다. 재구성 이미지는 예측된 타겟 깊이와 포즈를 투영하고 소스이미지로부터 샘플링하여 얻을 수 있다. 이때 예측된 깊이와 포즈는 각각의 네트워크에 의해 추정되어진다. 재구성 이미지 생성을 위한 뷰합성 과정에 대한 식 (1)을 아래에 보인다.

$$\hat{I}_{s\to t} = I_s \ \ proj(D_t, T_{t\to s}, K)\rangle \quad (1)$$

기존의 비지도 모델의 손실 함수는 닮음 유사도 정도인 SSIM(Structural Similarity Index Measure)과 $L1$ 거리를 조합한 이미지 재구성 손실과, 깊이가 이미지 경계 불연속에 따라 달라지도록 제약하는 깊이 유연 손실을 사용한다. 본 연구에서도 기존의 연구를 따라 위의 손실함수를 사용한다.

최근의 연구는 이미지의 뷰합성이 아닌 특징 표현의 뷰합성을 수행하였다. 또한, 재구성 소스 특징과 타겟 특징간의 $L1$거리를 다채널로 누적하는 비용 용량 메커니즘을 적용하여 성능을 향상 시켰다[6]. 본 연구에서는 Resnet을 이용한 기존 모델의 인코더-디코더 구조에서 광역적 특징 추출자에 해당하는 깊은 레이어에 트랜스포머를 적용한다.

CNN기반의 모델은 컨볼루션 연산을 사용하기 때문에 인접 영역에 대한 특징 표현 특성이 있으며, 트랜스포머는 자기주의 메커니즘으로 전체 영역을 표현하는 특성을 갖는다. 따라서, 모든 레이어를 동일한 구조의 연속으로 설계하는 방법이 아닌, 레이어에 따라 다른 모델 블록을 적용하도록 지역적 블록과 광역적 블록을 제안한다.

지역적 블록은 기존 Resnet의 처음 두 레이어를 그대로 사용한다. 광역

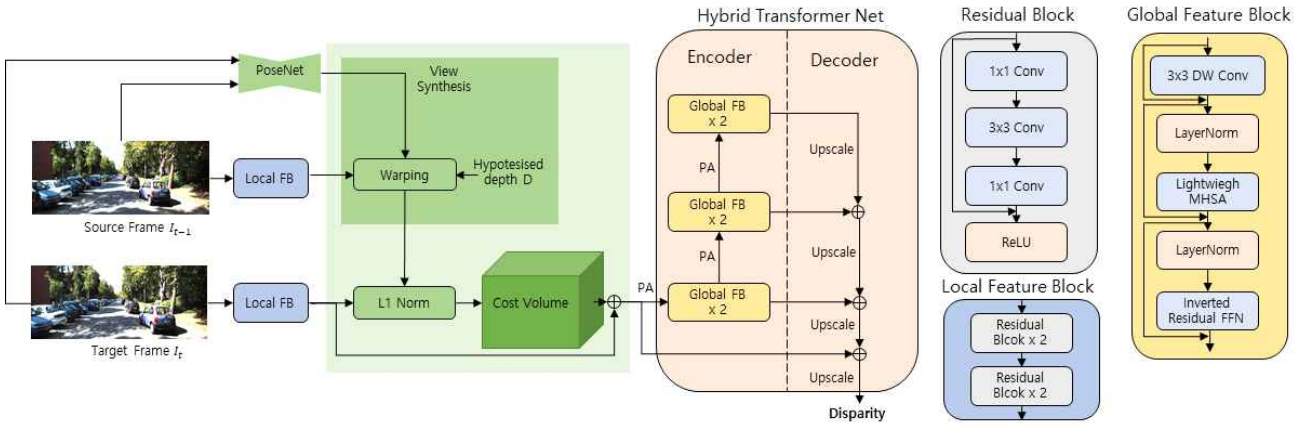


그림 1. 하이브리드 트랜스포머 모델 기반 단안 카메라 깊이 추정 모델

적 블록은 키와 값을 절반으로 줄인 가벼운 다중 자기주의 모듈과 역잔차 피드포워드 네트워크로 구성한다[7]. 광역적 블록은 각 레이어에 2개의 연속된 층으로 구성하였으며, 제안하는 전체 구조도를 그림 1 에 보인다.

Ⅲ. 실험 결과

본 논문에서는 제안하는 모델의 성능 평가를 위해 KITTI 데이터셋 기반 실험평가를 진행한다. 트랜스포머 모델은 다중 자기주의 연산에 따라 컨벌루션보다 더 적은 파라미터와 연산 수로 광역적 특징추출이 가능하다. 기존 Resnet18 기반 인코더-디코더 모델과 하이브리드 인코더-디코더 모델의 파라미터 차이를 표 1에 보인다.

표 1. 인코더-디코더 파라미터

	Params(M)	MACs(G)
Resnet18 [6]	14.277 M	4.461
Hybrid	**11.097 M**	**3.406**

기존 Resnet18 기반과 제안한 하이브리드기반 모델의 정량적 성능 평가를 표 2에 보인다. 평가 지표에서 AbsRel, SqRel, RMSE는 에러 평가지표로 낮을수록 좋으며, $\delta < 1.25$는 임계치 정확도 평가지표로 높을수록 좋은 지표이다. 본 논문에서 제안하는 하이브리드 모델이 기존 모델에 비해 보다 적은 파라미터와 연산량으로 유사한 성능을 나타냄을 알 수 있다.

표 2. 정량적 평가표

	AbsRel	SqRel	RMSE	$\delta < 1.25$	$\delta < 1.25^2$	$\delta < 1.25^3$
Resnet18 [6]	**0.098**	0.770	**4.459**	**0.900**	**0.965**	**0.983**
Hybrid	0.099	**0.757**	4.478	0.898	0.964	**0.983**

정성적 평가를 위한 결과 이미지를 그림 2에 보인다. 기존 모델보다 적은 파라미터와 연산량을 필요로 하지만 유사한 깊이 추정 결과를 보임을 확인한다.

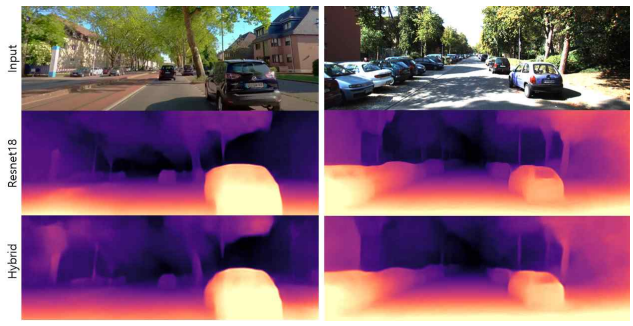


그림 2. 기존 모델과 제안하는 하이브리드 모델의 깊이 추정 결과

Ⅳ. 결 론

본 논문에서는 비지도 학습기반 단안 카메라 영상 깊이 추정에 CNN기반 Resnet18 모델과 비전 트랜스포머 모델을 혼합한 하이브리드 모델을 제안하였다. 제안하는 하이브리드 모델은 기존 모델에 비해 보다 적은 파라미터와 연산량을 필요로 하지만, 기존 모델과 유사한 성능을 확인하였다. 추후 연구로 기존 자세 추정 모델 경량화를 진행할 계획이다. 깊이 추정 네트워크의 자세 추정 모델 또한 Renet18기반의 인코더-디코더 모델을 사용한다. 자세 추정 모델은 두 개의 입력이미지에서 자세를 추정하기 때문에, 두 이미지간의 상관관계를 분석하는 상호 작용 트랜스포머 모델을 사용하여 보다 효율적인 네트워크를 구성할 수 있을 것으로 예상된다.

ACKNOWLEDGMENT

본 연구는 경기도 지역협력 연구센터 사업의 일환으로 수행하였음. [GRRC항공2017-B04, 지능형 인터랙티브 미디어 및 공간 융합 응용 서비스 개발].

참 고 문 헌

[1] Ming, Yue, et al. "Deep Learning for Monocular Depth Estimation: A Review." Neurocomputing (2021).

[2] Liu, Ze, et al. "Swin transformer: Hierarchical vision transformer using shifted windows." arXiv preprint arXiv:2103.14030 (2021).

[3] Dong, Xiaoyi, et al. "CSWin Transformer: A General Vision Transformer Backbone with Cross-Shaped Windows." arXiv preprint arXiv:2107.00652 (2021).

[4] Ranftl, René, Alexey Bochkovskiy, and Vladlen Koltun. "Vision transformers for dense prediction." arXiv preprint arXiv:2103.13413 (2021).

[5] Guizilini, Vitor, et al. "3d packing for self-supervised monocular depth estimation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.

[6] Watson, Jamie, et al. "The Temporal Opportunist: Self-Supervised Multi-Frame Monocular Depth." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.

[7] Guo, Jianyuan, et al. "CMT: Convolutional Neural Networks Meet Vision Transformers." arXiv preprint arXiv:2107.06263 (2021).

STFT 를 활용한 비가시영역 형상 복원에 관한 실험적 연구

신의현, 장승우, 김광수*
성균관대학교

shineh96@skku.edu, sewo@skku.edu, *kim.kwangsu@skku.edu

Experimental Study on Non-Line-of-Sight Image Reconstruction using STFT

Ui Hyeon Shin, Seungwoo Jang, Kwangsu Kim*
Sungkyunkwan Univ.

요 약

비가시영역(Non-Line-of-Sight) 형상 복원은 사람이 직접 관찰할 수 없는 영역의 사물을 인식하고 시각화하는 기술로 광학(Optics), 무선(Radio frequency), 음향(Acoustic) 장비 등을 사용한다. 기존 음향 장비를 사용한 연구는 소리의 시간과 진폭(Amplitude)을 사용하기 때문에, 노이즈가 있는 환경에서는 신호가 중첩되어 물체로부터 반사된 신호를 구분하기 어렵다. 따라서, 본 논문에서는 가청 주파수(Acoustic frequency)를 시간에 따른 주파수 대역별 신호로 변환하는 STFT(Short-Time Fourier Transform)을 사용하여 노이즈가 있는 환경에서 비가시영역 형상 복원에 필요한 정보를 획득하는 것을 실험을 통해 증명한다.

1. 서론

비가시영역 형상 복원은 벽, 장애물 등에 의해 가려 사람이 직접 관찰할 수 없는 영역의 사물을 인식하고 시각화하는 기술이다. 자율주행, 원격감지 등의 분야에 활용되며, 광학(Optics), 무선(Radio frequency), 음향(Acoustic) 장비 등을 사용한다. 그 중, 음향 장비는 광학, 무선 장비에 비해 저렴하고, 가청 주파수는 넓은 주파수 대역(20Hz - 20,000Hz)을 사용하기 때문에 노이즈에 강건하며, 소리의 회절성으로 한 번에 넓은 범위의 정보를 취득할 수 있어 비가시영역 연구에 적합하다.

가청 주파수(Acoustic frequency)를 활용한 기존 연구[1]는 반사벽을 통해서만 소리가 반사되고 외부 소음으로부터 격리된, 노이즈가 없는 환경에서 신호의 진폭을 측정한다. 직접음, 반사벽에 반사된 1 차 반사음, 물체에 반사되어 돌아온 2 차 반사음을 구분하여, 신호가 물체에 반사된 후 돌아오는데 걸리는 시간을 통해 형상을 복원한다. 노이즈가 없는 환경에서는 <그림 1 (좌)>와 같이 반사음 간의 중첩이 발생하지 않기에 신호의 진폭을 통해 반사음을 구분할 수 있다. 하지만, 모든 벽에서 소리가 반사되어 노이즈가 있는 환경에서 신호의 진폭을 사용하면, <그림 1 (우)>와 같이 여러 반사음과 소음이 중첩되어 직접음과 반사음을 구분할 수 없다.

이를 해결하기 위해, 가청 주파수 신호를 시간, 주파수, 신호강도 3 개의 축으로 변환하는 STFT(Short-Time Fourier Transform)[2]를 사용하면 중첩된 신호를 구분하고 주파수 대역별 정보를 얻을 수 있다. 얻은 정보를 통해서 직접음과 반사음을 구분할 수 있고(그림 3), 잔향을 통해서 비가시영역의 물체를 확인할 수 있다(그림 4). 본 논문에서는 노이즈가 있는 상황에서 STFT 를 활용하여 구분된 직접음, 반사음, 잔향의 정보를 통해서 비가시영역의 물체를 탐지하는 것을 실험적으로 증명한다.

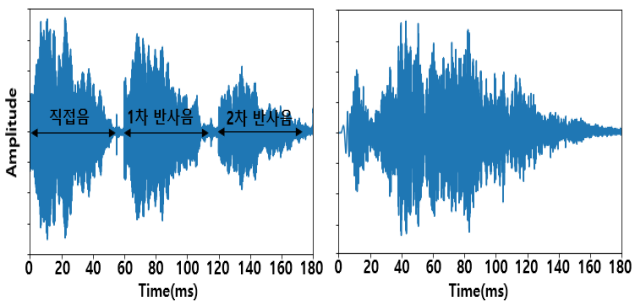


그림 1. 신호의 진폭 사용 시 발생하는 문제점. 노이즈가 없는 환경(좌)에서는 직접음과 반사음이 구분되지만, 노이즈가 있는 환경(우)에서는 신호가 중첩되어 직접음과 반사음을 구분할 수 없다.

2. 본론

2.1 실험 환경

본 논문에서는 노이즈가 있는 비가시 상황을 설정하기 위해 <그림 2>와 같이 실험실을 구성한다. 가로 6.52m,

세로 8.51m 의 공간에 나무 재질의 가림벽으로 공간 내의 물체를 직접 관찰할 수 없도록 하고, 스피커에서 발생한 신호는 반사벽을 거쳐 플라스틱 재질의 마네킹에 도달한 뒤, 다시 반사벽을 거쳐 마이크를 통해 수집된다.

실험환경은 가청 주파수를 활용한 비가시영역 복원 기초 연구[3]와 같이 8 개의 스피커를 사용하여 20Hz-20,000Hz 의 선형 Chirp 신호[4]를 0.1s 동안 발생시키고, 1 개의 측정용 마이크를 사용하여 반사된 신호를 녹음한다. 이때, 직접음과 반사음 이후에 남아있는 잔향까지 녹음하기 위해 0.4s 를 녹음한다.

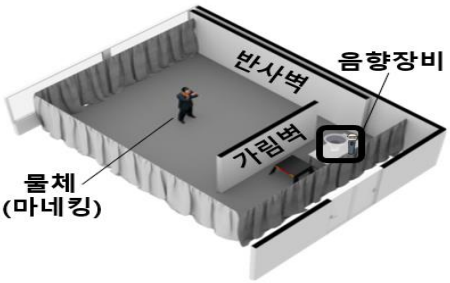


그림 2. 실험실 구조도

2.2 빈 공간과 마네킹이 있는 공간에서의 STFT 실험 결과

신호를 STFT 로 변환하면 신호가 중첩되더라도 주파수 대역별로 나누어 확인 가능하기 때문에 회환 된 신호의 주파수 영역대를 구분하여 소음을 제거할 수 있고, <그림 3>과 <그림 4> 같이 직접음과 반사음을 구분할 수 있다.

또한, 회환 된 신호를 STFT 로 변환하면 잔향을 통해 비가시영역의 물체를 탐지할 수 있다. <그림 3>과 <그림 4>는 각각 빈 공간과 마네킹이 존재하는 공간에서 회환 된 신호의 STFT 값을 보여준다. <그림 3>에서는 잔향에 특정 주파수 값이 관측되지 않지만, <그림 4>에서는 잔향에서 특정 주파수 값이 관측되는 것을 확인할 수 있다. 즉, 마네킹이 존재하는 공간의 잔향에서 선형 Chirp 신호가 더 오래 남아 있다.

이처럼 물체의 존재 여부에 따라 STFT 가 구분 가능한 차이를 보이므로 STFT 를 통해 비가시영역의 형상 복원에 필요한 정보를 획득할 수 있다. STFT 는 신호의 진폭에 비해 더 많은 정보를 포함함으로 추후 딥러닝 모델을 적용하여 비가시영역 형상 복원에 활용 가능성이 크다.

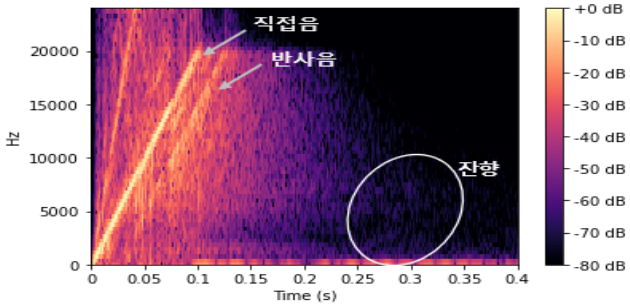


그림 3. 빈 공간에서의 STFT

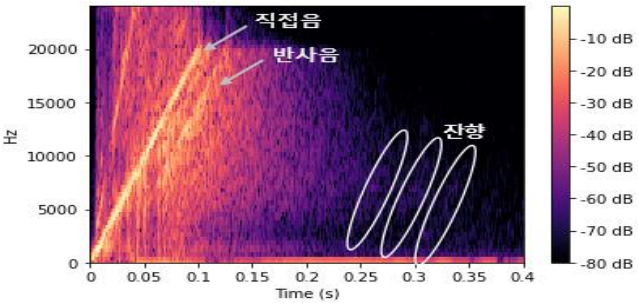


그림 4. 마네킹이 존재하는 공간에서의 STFT

3. 결론

본 논문은 가청 주파수 신호를 STFT 로 변환하여 노이즈가 있는 환경에서 발생 가능한 문제점을 해결해, 비가시영역 형상 복원에 필요한 정보를 획득하는 것을 실험을 통해 증명하였다. STFT 는 진폭만 사용했을 때와 달리 반사음과 소음 등 중첩된 신호를 구분하고, 물체의 존재 여부에 따라 잔향에서 뚜렷한 차이를 보인다. 이를 통해, 추후 이를 딥러닝 모델에 적용하여 비가시영역의 형상 복원에 활용할 수 있을 것으로 기대한다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 지역지능화혁신인재양성(Grand ICT 연구센터) 사업의 연구결과로 수행되었음 (IITP-2021-2015-0-00742)

참 고 문 헌

[1] Lindell, David B., Gordon Wetzstein, and Vladlen Koltun. "Acoustic non-line-of-sight imaging." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019.

[2] Biondi, Biondo. "Migration using short-time Fourier transforms." (1950).

[3] 장승우, 김광수. "가청 주파수를 활용한 비가시 영역 복원 기초 연구." 한국통신학회 학술대회논문집 (2021): 1088-1089.

[4] Mann, Steve, and Simon Haykin. "The chirplet transform: A generalization of Gabor's logon transform." Vision interface. Vol. 91. 1991.

등간격 평면배열 안테나에서의 Deep Neural Network 기반 방위각 및 고각 탐지

지오근, 정현진, 서형욱, 김선우
한양대학교 융합전자공학부
{wldhrms, hyeonjingo, rlrprl, remero}@hanyang.ac.kr

Azimuth and Elevation Estimation on Uniform Planar Array using Deep Neural Network

Ohgeun Jee, Hyeonjin Chung, Hyeongwook Seo, and Sunwoo Kim
Department of Electronic Engineering, Hanyang University

요 약

본 논문에서는 등간격 평면배열 안테나에서 수신 신호 방향탐지 시 보다 낮은 계산 복잡도를 요구하는 DNN (Deep Neural Network) 기반 방위각 및 고각 탐지 알고리즘을 제안한다. 제안하는 DNN 구조는 공분산 행렬의 상삼각행렬을 입력하여 추정 도래각을 출력하는 회귀 구조를 가진다. 시뮬레이션을 통해 활성화 함수에 따른 방향탐지 성능을 비교하였으며, 제안 알고리즘과 기존 방향탐지 알고리즘인 MUSIC (multiple signal classification)과의 RMSE (root mean squared error)와 실행시간을 비교한다.

Ⅰ. 서 론

방향탐지는 주로 레이다, 군수 통신 등에 사용되며 최근 이동통신에서의 배열 안테나 신호 처리기술이 일반화됨에 따라 민간 분야에도 확대 적용되고 있다 [1]. 하지만 하드웨어 발전에 따라 배열을 구성하는 안테나 수가 증가하는 추세이며, 이때 MUSIC (multiple signal classification)과 같은 기존 방향탐지 알고리즘 사용 시 높은 계산 복잡도가 문제가 될 수 있다 [2, 3]. 본 논문에서는 방위각 및 고각 추정 시 요구되는 계산 복잡도를 줄이고자 DNN (Deep Neural Network) 기반 방향탐지 알고리즘을 제안한다.

Ⅱ. 시스템 모델

그림 1과 같이 등간격 평면배열 안테나에서 x 축 상 m 번째, y 축 상 n번째 안테나의 좌표 $\boldsymbol{d}_{m,n}$ 은 $\boldsymbol{d}_{m,n}=[(m-1)d,(n-1)d]$를 만족하며 여기서 d 는 인접한 두 안테나 간의 거리를 의미한다. 방위각이 ϕ, 고각이 θ 일 때 배열의 x 축과 y 축에 따른 조향벡터 $\boldsymbol{a}(u), \boldsymbol{a}(v)$는 다음과 같이 나타낼 수 있다.

$$\boldsymbol{a}(u)=[1, e^{j2\pi(\frac{d\cos\phi\sin\theta}{\lambda})}, \cdots, e^{j2\pi(\frac{d(M-1)\cos\phi\sin\theta}{\lambda})}]^T$$
$$\boldsymbol{a}(v)=[1, e^{j2\pi(\frac{d\sin\phi\sin\theta}{\lambda})}, \cdots, e^{j2\pi(\frac{d(N-1)\sin\phi\sin\theta}{\lambda})}]^T, \quad (1)$$

여기서 u, v 는 다음과 같이 나타낼 수 있다.

$$u=\frac{2\pi d}{\lambda}\cos\phi sin\theta,\ v=\frac{2\pi d}{\lambda}\cos\phi sin\theta\ . \quad (2)$$

등간격 평면배열 안테나에서의 D 개의 snapshot으로 구성한 최종 수신 신호 X는 다음과 같이 표현할 수 있다.

$$\boldsymbol{X}=\boldsymbol{AS}+\boldsymbol{W}\in C^{MN\times D}\ , \quad (3)$$

여기서 $\boldsymbol{W}\in C^{MN\times D}$ 는 noise 행렬을 의미하고 수신 신호 행렬 $\boldsymbol{S}$ 는 $\boldsymbol{S}=[\boldsymbol{s}_1, \boldsymbol{s}_2, \cdots, \boldsymbol{s}_P]^T\in C^{P\times D}$ 이며, 벡터 $\boldsymbol{s_i}$ 는 길이가 D 인 i 번째 신호 벡터를 의미한다. $\otimes$는 Kronecker product를 의미한다. $\boldsymbol{A}$ 와 $\boldsymbol{a_i}$는 다음과 같이 나타낼 수 있다.

$$\boldsymbol{A}=[\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_P]\in C^{MN\times P}\ , \quad (4)$$

$$\boldsymbol{a}_p=\boldsymbol{a}(u_p)\otimes\boldsymbol{a}(v_p)\in C^{MN\times 1}\ . \quad (5)$$

수신 신호 X의 공분산 행렬 R_X는 다음과 같이 나타낼 수 있다.

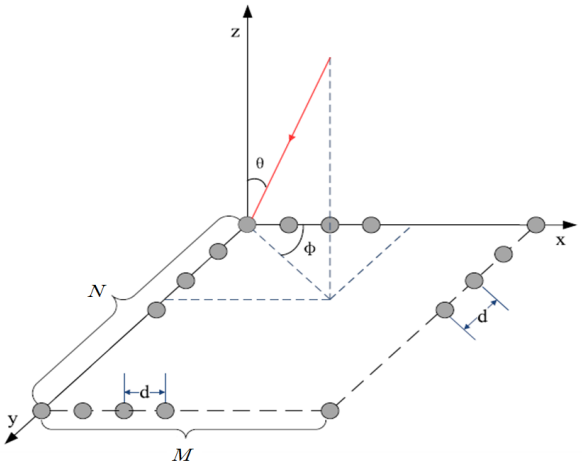


그림 1. 등간격 평면배열 안테나 구조

$$\boldsymbol{R_X}=E[\boldsymbol{XX^H}]=\boldsymbol{AR_SA^H}+\sigma^2\boldsymbol{I_{MN}}$$
$$\approx\frac{\boldsymbol{XX^H}}{D}\in C^{MN\times MN}, \quad (6)$$

여기서 $\boldsymbol{R}_S$ 는 신호원 $\boldsymbol{S}$ 의 공분산 행렬, σ 는 noise의 power, $\boldsymbol{I_{MN}}$ 은 $MN\times MN$ 인 단위행렬을 나타낸다. $\boldsymbol{R_X}$ 는 에르미트 행렬이기에, $\boldsymbol{R_X}$ 의 상삼각행렬을 벡터화한 $\boldsymbol{b}$ 는 다음과 같이 나타낼 수 있다.

$$\boldsymbol{b}=[R_{11}, R_{12}, R_{22}, \cdots, R_{1L}, \cdots, R_{LL}]\in C^{(MN+1)MN/2}. \quad (7)$$

Ⅲ. DNN 기반 방향탐지 알고리즘

DNN의 입력은 복소수가 될 수 없기에 (7)에서의 $\boldsymbol{b}$ 의 실수부와 허수부를 분리한 $\bar{\boldsymbol{b}}$ 와 이를 정규화한 $\boldsymbol{z}$ 를 입력으로 한다. $\boldsymbol{z}$ 는 다음과 같이 나타낼 수 있다.

$$\bar{\boldsymbol{b}}=[R_{11}, Re(R_{12}), Im(R_{12}), \cdots, Re(R_{1L}), Im(R_{1L}), \cdots, R_{LL}]\in R^{(MN)^2}, \quad (8)$$

$$\boldsymbol{z}=\frac{\bar{\boldsymbol{b}}}{\|\bar{\boldsymbol{b}}\|}\in R^{(MN)^2}\ . \quad (9)$$

DNN의 출력 학습 데이터 $\boldsymbol{y}$ 는 $\boldsymbol{y}=[\alpha_1, \beta_1, \cdots, \alpha_P, \beta_P]\in R^{2P\times 1}$ 로

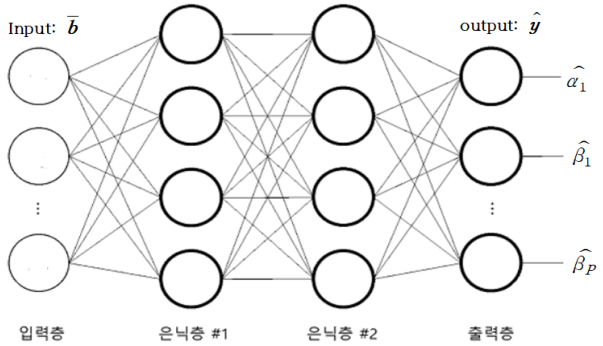


그림 2. 공분산 행렬의 상삼각행렬을 입력으로, 방위각 및 고각을 출력으로 갖는 DNN 구조

설정하였다. 이때, α_i, β_i 는 각각 i 번째 신호의 실제 도래각 (방위각 및 고각)이다. 이와 같이 DNN은 회귀형식으로 설정하였으며, 신호원의 개수 P는 사전에 주어진다고 가정한다. DNN의 구조는 그림 2와 같이 도식적으로 나타낼 수 있다.

손실 함수는 MSE (mean squared error)로 설정하였으며 $\|\boldsymbol{y}-\hat{\boldsymbol{y}}\|^2$와 같이 나타낼 수 있다. 여기서 $\hat{\boldsymbol{y}}$ 는 신호원의 추정 도래각으로 DNN의 출력값이다. DNN의 하이퍼 파라미터는 표 1과 같다.

Ⅳ. 시뮬레이션 환경 및 결과

본 논문에서는 P 를 2로 고정하였다. 100,000개의 train data set의 SNR (signal-to-noise ratio)은 1dB부터 10dB까지 무작위로 설정하였고, RMSE 산출을 위해 총 500 test data를 사용하였다. 시뮬레이션은 Intel i7-8750H CPU, NVDIA GeForce GTX 1050, RAM 20GB을 사용해서 진행하였다. 이외 시뮬레이션 환경 파라미터는 표 2와 같다.

표 1. DNN의 하이퍼 파라미터

하이퍼 파라미터	설정값
Activation function	ReLU, PReLU
Number of layers, nodes	4, 300
Batch size	64
Number of epochs	150

표 2. 시뮬레이션 환경 파라미터

파라미터	설정값
안테나 개수 (M,N)	5, 5
안테나 간격 (d)	0.5λ (λ: 신호의 파장)
Snapshot 개수 (D)	2048
방위각, 고각 범위	$30^\circ \le \theta_i \le 150^\circ, 60^\circ \le \phi_i \le 80^\circ$

그림 3에서 초기 epoch에서는 PReLU로 구성된 네트워크가 ReLU로 구성된 네트워크에 비해 전반적으로 적은 MSE 값을 보였으나, epoch가 거듭됨에 따라 두 DNN의 MSE가 유사해져 MSE 차이가 0.2 정도로 근소해짐을 확인하였다.

그림 4는 PReLU 사용 시 SNR에 따른 Monte Carlo 수가 500인 RMSE (root mean squared error) 결과이다. 제시한 알고리즘이 MUSIC 알고리즘에 비해 낮은 SNR에서 보다 낮은 RMSE를 가지는 것을 확인하였으며, 높은 SNR에서는 성능 차이가 좁혀지는 것을 확인하였다.

또한, 본 논문에서 제안한 DNN 기반 알고리즘이 MUSIC 알고리즘에 비해 추정시간 측면에서 장점을 가진다. MUSIC 알고리즘의 경우 각도 스펙트럼 분할 수가 많아질수록 소요 시간이 길어지는데, MUSIC의 각도

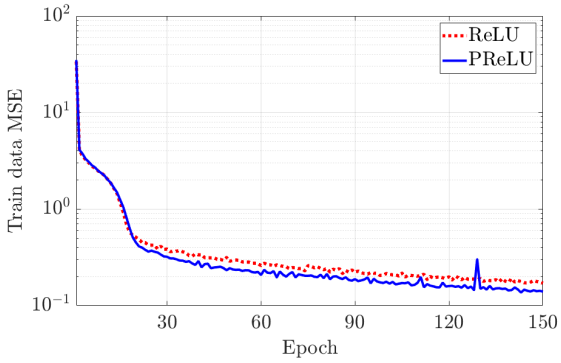


그림 3. 각 활성화 함수의 Epoch에 따른 train data MSE

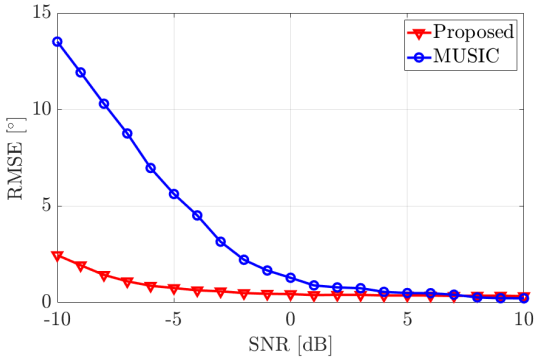


그림 4. test data의 SNR에 따른 DNN과 MUSIC 알고리즘의 RMSE 결과

스펙트럼을 0.5° 단위로 분해한 경우 도래각 추정에 평균 0.2초가 소요되었다. 반면, 본 논문에 제안하는 DNN 기반 알고리즘은 평균 0.05초가 소요됨을 확인하였다.

Ⅴ. 결론

본 논문에서는 DNN 기반 신호 방위각 및 고각 추정 알고리즘을 제시한다. 본 논문에서 제안하는 알고리즘은 신호의 공분산 성분을 바로 입력하여 방위각 및 고각을 출력하는 회귀 방식이다. 시뮬레이션 결과를 통해 SNR이 낮을 때 제안 알고리즘의 RMSE가 MUSIC 알고리즘보다 낮음을 확인하였으며, 실행 시간 또한 MUSIC 알고리즘에 비해 평균적으로 4배 가량 빠름을 확인하였다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터육성지원사업의 연구결과로 수행되었음 (IITP-2021-2017-0-01637)

참 고 문 헌

[1] 장병준, "방향탐지기술: 튜토리얼, 최신동향 및 연구전망," *한국전자파학회논문지*, 30(8), pp. 607-618, 2019년 8월.

[2] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas, Propag.*, vol. 34, no. 3, pp. 276-280, Mar. 1986.

[3] S. Yang and L. Hanzo, "Fifty Years of MIMO Detection: The Road to Large-Scale MIMOs," IEEE Commun. Surveys, Tut., vol. 17, no. 4, pp. 1941-1988, Fourthquarter, 2015.

뉴스 기반의 주가 변동폭 예측 시스템에 관한 연구

류재학, 조성혁, Md Habibur Rahman, 장영민
국민대학교
jhak@kookmin.ac.kr, hang8709@kookmin.ac.kr, habib182@kookmin.ac.kr, yjang@kookmin.ac.kr

A Study on the News-based stock price fluctuationprediction system

Ryu Jae Hak, Cho Seong Hyeok, Md Habibur Rahman, Jang Yeong Min
Kookmin Univ.

요 약

본 논문은 회사에 대한 긍정적/부정적 뉴스가 주가에 미치는 영향을 학습시킨 뒤 새로 입력되는 뉴스 데이터가 회사에 긍정적인지 부정적인지를 판단하고, 과거의 뉴스 감정과 주가 등락 데이터를 기반으로 새로운 주가 등락을 예측한다. LSTM 기반 주가 예측이 기존의 주가 데이터만으로 미래의 주가를 올바르게 예측하기 어렵다는 점을 개선하고자 하였다.

Ⅰ. 서 론

현재 넓게 분포되어있지만 눈에 띄지 않는 알고리즘 우위 영역은 주식 시장이다. 새로운 거래 알고리즘이 등장해 월스트리트의 거래 방식을 바꾸기도 하며, 알고리즘 거래는 미국 전체 거래의 약 70%, 인도와 같은 신흥 경제국에서는 약 40%의 거래량을 차지하는 것으로 추정된다.[1] 더 나아가 단순 거래뿐만이 아닌 주가를 예측하는 알고리즘 또한 활발하게 연구되었고, 그 중 하나가 LSTM을 이용한 주가 예측이다. 그러나 주가는 실수 체계의 값이고 무작위성이 강하기 때문에 LSTM 모델에서의 학습은 불확실한 중간값이나 의미 없는 값을 반환하는 경향이 강해진다.[2]

그러나 주가는 완전히 무작위적으로 결정되는 값은 아니다. 주가에 영향을 미치는 요소가 존재하고, 그 중 하나는 뉴스에 포함된 기업의 정보이다. 일일 웹 직, 간접 뉴스의 상승/하락 추세는 일일 평균 수익률에 영향을 미친다. 주가에 영향을 미칠 수 있는 기업에 대한 특정 사건이 발생하였을 때 웹 뉴스를 통해 투자자들의 심리가 좌우되고, 이는 곧 주가의 변화를 야기할 수 있음을 뜻한다.[3] 뉴스의 데이터가 주가의 등락을 온전히 결정하지는 않지만, 등락의 방향성을 잡기 위한 유의미한 데이터로서 가치를 갖게 된다.

본 논문에서는 크롤링한 뉴스 데이터를 기반으로 뉴스의 감성(긍정적/부정적)의 단계와 과거 뉴스에서 보여주었던 주가 변동폭의 경향성을 학습시켜 주가 자체의 완벽한 예측이 아닌 주가의 변화 추세를 예측하는 것을 목적으로 하고 있다.

Ⅱ. 본론

2.1 주가 예측 모델 구현 과정

Kaggle 데이터(https://www.kaggle.com/aaron7sun/stocknews)를 기반으로[5] 주가 정보를 모은다. 이후 뉴스 제목을 크롤링하여 정규화를 거치고, VADER(Valence Aware Dictionary and sEntiment Reasoner) 감정 분석을 이용해 텍스트 분석 과정을 진행했다. 결과는 다음과 같다.

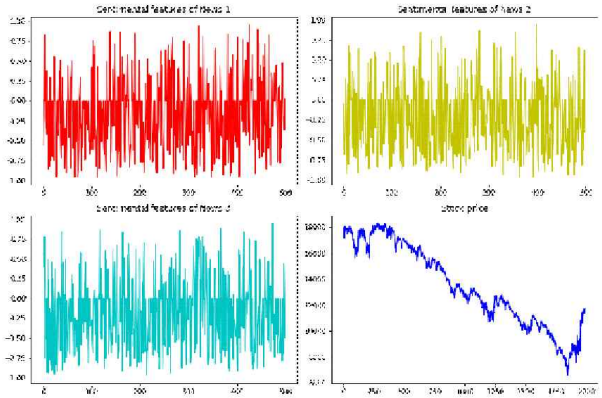


그림 1. 텍스트 특징 추출 데이터 및 주가 데이터의 그림

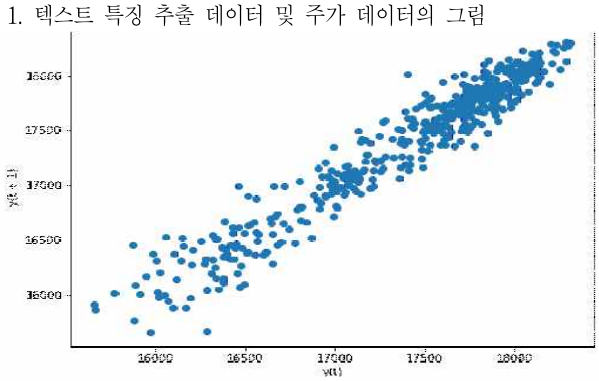


그림 2. 현재 단계(t)와 다음 단계($t+1$)데이터의 상관관계

또한 현재 단계의 $y(t)$와 다음 단계의 상관관계$y(t+1)$를 확인하여, 양의 상관관계를 가짐을 결론하였다.

2.2 VADER 감성 분석

VADER(Valence Aware Dictionary and sEntiment Reasoner)는 규칙 기반 감정 분석 라이브러리 및 어휘를 뜻한다. Amazon의 Mechanical Turk를 통해 얻은 데이터를 기반으로 감정 샘플이 없더라도 감정 분류를

빠르게 실행할 수 있다. 또한 VADER 알고리즘은 다른 머신러닝 및 딥러닝 접근 방식과 비교했을 때 계산 면에서 효율적이며, 내장된 학습 라이브러리로 구성되어 있어 모델을 별도로 학습하지 않는다.

VADER 알고리즘은 주관성과 극성 개념을 사용한다. 극성은 문장에 표현된 긍정적 혹은 부정적 감정을 -1과 +1 사이로 나타낸 부동 소수점 값이며, 주관성은 문장의 주관적인 의견이나 감정 또는 판단을 나타내는 0에서 +1 사이의 부동 소수점 값이다. 감정 관련 단어의 사전을 기반으로 하는 감정 분석 유형에 속하며. 이 접근 방식에서 사전의 각 단어는 해당 단어가 얼마나 긍정적인지 부정적인지에 따라 등급이 매겨진다.

VADER 분석의 결과는 데이터의 텍스트가 복잡한 경우 많은 이점을 갖는다. 기존의 수동적 단어 평가의 단점이였던 고비용과 시간 소모 문제를 해결하고, 정확도를 높여 많은 양의 텍스트 데이터에서도 결과를 빠르고 정확하게 반환한다.

VADER는 텍스트 묶음을 분석할 때 텍스트의 단어가 사전에 있는지 확인 후, 3개의 감정 항목과 종합 결과값을 생성한다. 3개의 감정 항목은 각각 부정적, 중립적, 긍정적 텍스트의 비율을 나타내며, 종합 결과값은 -1에서 1 사이의 범위로 표준화된 감정 항목의 합계이다. [4] 이번 실험에서, 데이터셋의 문장 중 “U.N. Chief Admits He Removed Saudi Arabia From Child-Killer List Due to Extortion”에 VADER 알고리즘을 적용했을 때, 다음과 같은 결과가 출력되었다.

감정	결과값
긍정	0.155
부정	0.0
중립	0.845
종합	0.296

표 1. VADER 알고리즘 적용 결과

2.3 실험 과정

실험은 다음과 같은 과정으로 진행되었다. 먼저 인터넷에서 하루 단위로 3개의 주요 뉴스 헤드라인을 크롤링 후 VADER 감성 분석을 거친다. 이러한 시계열 데이터를 종합하여 입출력 행렬의 형태로 변환한다. 그리고 이 데이터 중 70%는 학습 데이터, 10%는 검증 데이터, 20%는 훈련 데이터로 분배하였다. 이후 LSTM 모델에 데이터를 훈련시켰다. 모델 매개변수는 다음과 같다.

매개변수	값
레이어	LSTM,Dropout
뉴런 개수	100,50
손실 함수	Mean Square Error
옵티마이저	Adaptive Moment Estimation
에포크	50
배치 크기	25

표 2. 훈련 모델의 매개변수

Ⅲ. 결론

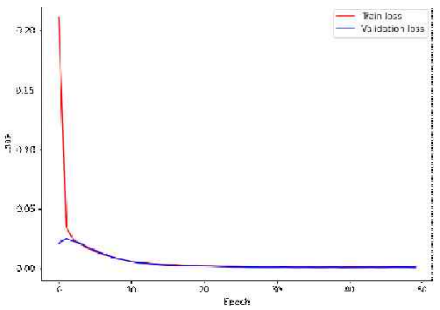


그림 3. 훈련 횟수에 따른 훈련 세트 손실과 검증 세트 손실

실험 결과 훈련을 반복함에 따라 훈련 세트 손실과 검증 세트 손실이 매우 비슷해 과적합을 걱정할 필요 없이 최적점에 도달했음을 확인했다.

성능 측정 지표인 mse, mae, mape를 측정한 결과 mse=0.000847, mae=0.023270, mape=0.108286의 결과가 도출되었다.



그림 4. 예측값과 실제값의 비교 그래프

훈련을 거듭함에 따라 미래 주가의 정확한 수치를 예측하는 능력은 다소 부족하지만, 주가 변동의 방향성과 유사한 추세를 보임을 확인하였다.

본 논문에서는 기존의 LSTM 주가 예측 모델의 단점을 보완하기 위해, 뉴스 데이터에 기반한 기업의 감정 평가를 이용하였다. 주가의 등락을 예상하는 모델을 제안하고, 이를 실제로 실험하여 주가 변동의 방향성을 성공적으로 예측함을 확인했다. 더욱 정확한 예측을 위해 뉴스의 감정 외에 추가적인 종속변수가 필요함이 예상된다.

ACKNOWLEDGMENT

참 고 문 헌

[1] Cameron Turner, "The Future of Algorithmic Trading", https://www.experfy.com/blog/the-future-of-algorithmic-trading, May 2021

[2] TBACKING, "LSTM 기반 주가 예측의 실상", https://tbacking.com/2020/08/16/lstm-%EA%B8%B0%EB%B0%98-%EC%A3%BC%EA%B0%80-%EC%98%88%EC%B8%A1%EC%9D%98-%EC%8B%A4%EC%83%81 , Aug 2020

[3] Sang Soo Kim,"A Study on the Relation of Web News and Stock Price", 2012, p. 201

[4] Monali Yadav, "Sentimental Analysis on Audio and Video using Vader Algorithm", 2019, p.192

[5] Sun, J. "Daily News for Stock Market Prediction", https://www.kaggle.com/aaron7sun/stocknews, 2016

Deep CNN을 통한 손 제스처 인식 및 ASL-파벳 변환

이건, 윤리나, Ida Bagus Krishna Yoga Utama, 장영민
국민대학교
kmuleegun@naver.com, flsk0102@naver.com, krishna@kookmin.ac.kr, yjang@kookmin.ac.kr

Hand gesture recognition and ASL-phabet conversion through deep convolutional neural network

Gun Lee, Ri Na Yoon, Ida Bagus Krishna Yoga Utama, Yeong Min Jang
Kookmin Univ

요 약

수화는 청각장애와 언어장애를 가진 사람들에게 없어서는 안 되는 중요한 대화수단이다. 본 논문에서는 영상인식 기술을 통해 알파벳을 인식하고 CNN 딥러닝을 통해 영상을 통해 구별되는 손 제스처를 문자로 변경해주는 알고리즘에 대해 연구하였다. 우선 CNN을 통해 손 제스처를 인식하여 학습시키고 이후에 새로운 손 제스처를 인식시키고 ASL-파벳을 예측하는 것이다. 제안된 CNN 아키텍처를 사용한 ASL-파벳 예측 정확도는 테스트 데이터에서 96.7%, 훈련 과정에서 98.3% 정확도를 보여주었다. 제안한 CNN 아키텍처는 다른 방법과 비교하여 가장 높은 정확도를 달성한다.

Ⅰ. 서 론

소리를 듣지 못하는 청각장애인들은 말을 제대로 하지 못하는 언어장애까지 수반하는 경우가 많다. 듣지 못하기 때문에 자신의 말을 제대로 교정할 수 없기 때문이다. 그들은 타인과의 대화를 위해선 수화를 사용한다. 하지만 이러한 수화는 많은 사람들에게 이야기 할 때는 불편할 수 밖에 없다. 언어를 시각으로 전달해야하기 때문이다. 또한 장애인들은 항상 도움을 구해야 하는 상황에 처하거나 수화를 알지 못하는 사람과의 대화에 처할 수 있다. 수화를 사용하는 사람들을 돕기 위해 수화를 영상인식을 통해 그들이 표현하는 언어, 수화를 자동적으로 문자로 변환해주는 아키텍처를 개발하였다. 여기서는 수화의 손짓을 감지하고 분류하는 방법을 제안한다. 데이터베이스는 미국 수화의 개방형 데이터베이스를 사용한다. 이 데이터베이스는 온라인에서 사용할 수 있으며 무료로 다운로드할 수 있다. 미국 수화 데이터베이스는 손 제스처를 분류하고 손 제스처를 알파벳으로 번역하기 위해 제안된 수정된 CNN 아키텍처를 훈련하는 데 활용된다.

Ⅱ. 데이터 세트

청각 장애와 언어 장애를 가지고 있는 사람들이 사용하는 수화에도 다양한 종류가 존재한다. 무려 100개 이상의 다른 수화가 사용되며 인도 수화, 미국 수화, 이탈리아어 수화 및 많은 수화가 존재한다. 그 중 미국 수화(ASL)는 북미에서 4번째로 많이 사용될 정도로 많이 사용되는 언어이다. ASL은 미국뿐만 아니라 영어가 주요 의사소통으로 사용되는 30개 이상의 다른 국가에서도 사용된다. ASL은 American Manual Alphabet으로 널리 알려진 26개의 제스처 기호 집합으로, 영어 사전에 있는 다양한 단어를 표현하는데 사용할 수 있다. 우리가 사용하는 데이터 세트는 미국 수화용 MNIST(Modified National Institute of Standards and Technology database) 수화이다. 미국 수화는 원래 26개의 알파벳으로 구성되어 있다. 그러나 알파벳 J와 Z는 모션 기반 알파벳이기 때문에 이 두 알파벳은 이 데이터 세트에서 제외된다. 따라서 이 데이터 세트에서는 24개의 알파벳만 사용한다. 데이터 세트는 총 34,267개의 이미지로 구성되어 있으며, 학습 데이터용 이미지 27,455개와 테스트용 이미지 7,172개로 나뉜다. 데이터 세트의 각 이미지는 28 x 28 픽셀의 비슷한 크기를 가지며 모두 회색조 형식을 사용한다.

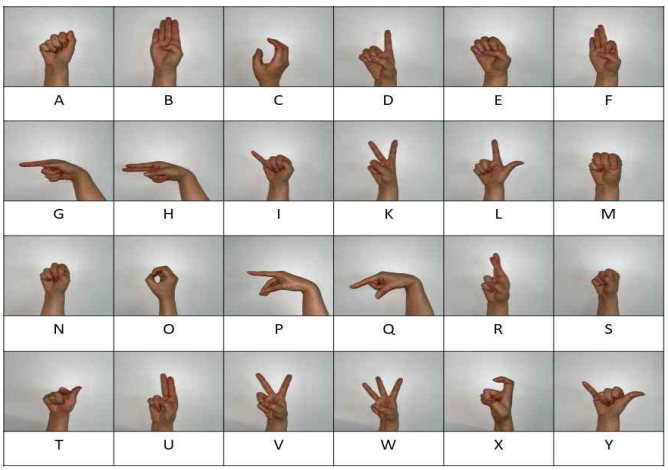


그림 1. 미국식 수화 (ASL)

Ⅲ. 시스템 아키텍처

Convolutional neural network(CNN 또는 ConvNet)은 수동으로 특징을 추출할 필요 없이 데이터로부터 직접 학습하는 딥 러닝을 위한 신경망 아키텍처이다. CNN은 영상에서 객체, 얼굴, 장면 인식을 위한 패턴을 찾을 때 특히 유용하다. 자율주행 차량 및 얼굴 인식 응용 분야와 같이 객체 인식과 컴퓨터 비전이 필요한 응용 분야에서는 CNN을 많이 사용한다. 이 CNN을 사용하여 영상에서 손 제스처를 인식하여 표현하고자 하는 문자를 가져올 수 있다. 이 연구에서 우리는 6개의 컨볼루션 레이어, 6개의 배

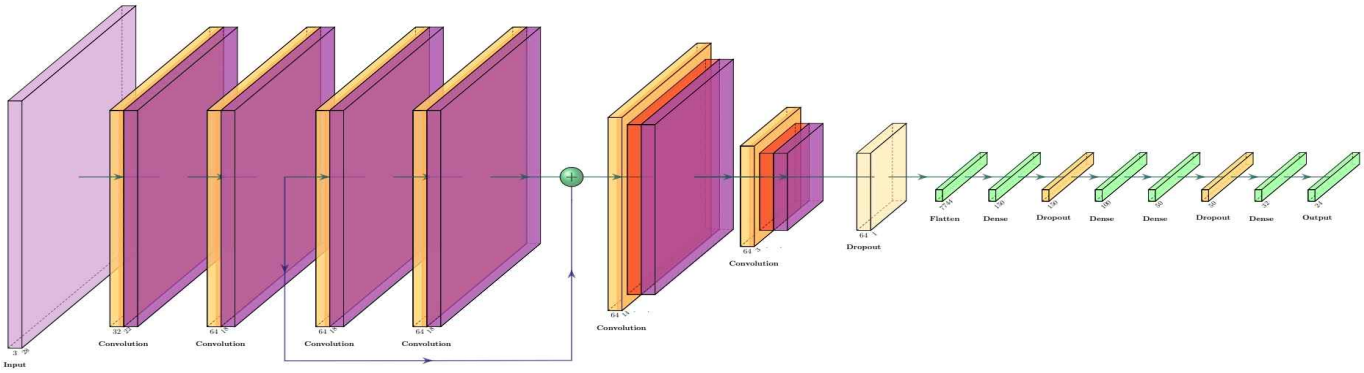


그림2. 제안된 CNN 아키텍처

치 정규화 레이어, 2개의 풀링 레이어, 3개의 드롭아웃 레이어, 5개의 고밀도 레이어로 구성된 CNN 아키텍처를 제안한다. 우리는 또한 두 번째 컨볼루션 레이어와 네 번째 컨볼루션 레이어 사이에 잔여 연결을 구현한다. 각 컨볼루션 레이어에서 서로 다른 크기의 커널을 적용하고 ReLU 활성화 함수를 사용하여 컨볼루션 레이어를 활성화한다. 또한 모든 컨볼루션에 대해 배치 정규화를 적용한다. 여러 드롭아웃 레이어가 CNN 아키텍처 전체에 구현되어 모델을 정규화하여 네트워크의 과적합을 줄이는 데 도움이 된다. 컨볼루션 레이어는 이미지에 포함된 특별한 정보를 추출하는 것을 목적으로 한다. 그런 다음 이미지를 분류하기 위해 고밀도 레이어를 사용한다. 조밀한 레이어는 각 뉴런이 ReLU에 의해 활성화되는 서로 다른 수의 뉴런으로 구성된다. 데이터 세트의 각 클래스를 나타내는 24개의 뉴런으로 구성된 출력 레이어이기 때문에 Softmax는 마지막 밀집 레이어만 활성화한다. Softmax 활성화 함수는 출력 레이어의 각 뉴런에 대한 확률 값을 생성한다.

Ⅳ. 결과 및 토론

Ⅳ-Ⅰ. 데이터 전처리

데이터 세트의 이미지 크기는 모두 28 x 28픽셀이며 이미 회색조 형식이다. CNN에 이미지를 삽입하기 전에 이미지의 각 픽셀 값을 다시 조정한다. 원래 픽셀 값의 범위는 0-255이다. 여기에서 0-1 범위의 값을 다시 조정한다. 모든 이미지를 동일한 방식으로 처리하고 출력 범위가 0-1인 Softmax 기능을 사용하기 때문에 이 범위 값을 선택한다.

Ⅳ-Ⅱ. 교육 및 테스트

훈련 과정을 위해 훈련 데이터를 80%-20%로 나눈다. 80% 데이터는 훈련에 사용되고 20% 데이터는 훈련 프로세스의 모든 시대에 대한 네트워크 성능 검증에 사용된다. 이 연구에서는 100개의 epoch 훈련을 실행한다. 훈련 과정에서 우리는 항상 검증 과정에서 나온 정확도 값을 모니터링 한다. 훈련에서는 항상 최고의 정확도를 달성하는 최고의 모델을 저장한다. 훈련 과정이 끝난 후, 우리는 이전에 모델에서 볼 수 없었던 이미지 데이터 세트를 사용하여 테스트 과정을 수행한다. 훈련 과정에서 저장한 최상의 모델을 로드하고 해당 모델을 사용하여 테스트 데이터 세트를 기반으로 예측한다. 교육 및 테스트 과정에서 Crosse entropy loss 함수를 사용하여 손실을 측정한다. 손실을 측정하는 것은 모델 무게를 수정하여 모델 성능을 향상시키는 데 중요하다.

Ⅳ-Ⅲ. 결과 및 분석

훈련 과정에서 우리는 0.005의 손실 값과 98.3%의 정확도를 달성했다. 한편 테스트 과정에서 손실 값 0.09와 정확도 96.7%를 달성했다. 훈련과 테스트 과정의 정확도가 크게 다르지 않기 때문에 만족할만한 결과를 얻은 것 같다. 이는 모델이 과적합 또는 과소적합 되지 않고 모델이 이전에 해당 데이터에 대해 학습한 적이 없음에도 불구하고 올바르게 분류할 수 있도록 모델이 매우 잘 정규화 되었음을 의미한다. 이 결과는 네트워크의 잔여 연결 때문이기도 하다. 비선형 활성화 함수가 네트워크를 따라 그라디언트가 사라지거나 폭발할 수 있기 때문에 잔차 연결은 그라디언트가 직접 흐르도록 하고 비선형 활성화 함수의 영향을 받지 않는다. 우리는 본 논문의 결과를 다른 논문의 결과와 비교한다. 표 1로 다른 방법과의 비교한 결과를 제시한다.

표 1. 다른 방법과의 정확도 비교 결과

방법	정확도 (%)
LeNet [1]	82.19
CapsNet [1]	88.93
CapsNet augmented [1]	95.08%
CNN [2]	95.3%
Ours	96.7%

Ⅴ. 결론

본 논문에서는 MNIST 수화 데이터 세트를 사용하여 훈련되는 CNN 아키텍처를 사용하여 손 제스처를 감지하고 분류하는 작업을 발표했다. 제안된 방법의 성능은 훈련 과정에서 98.3%의 정확도와 테스트 과정에서 96.7%의 정확도를 생성할 수 있음을 약속한다. 본 결과를 통해 deep CNN을 사용한 손 제스처 인지 방식을 더 보완하여 수화를 사용하는 사람들에게 보다 더 편한 삶을 제공할 수 있기를 바란다.

참 고 문 헌

[1] Bilgin, M., & Mutludogan, K. (2019). American sign language character recognition with capsule networks. 2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT). IEEE.

[2] Chakraborty, D., Garg, D., Ghosh, A., & Chan, J. H. (2018). Trigger detection system for American sign language using deep convolutional neural networks. Proceedings of the 10th International Conference on Advances in Information Technology - IAIT 2018. New York, New York, USA: ACM Press.

[3] MathWorks “https://kr.mathworks.com/discovery/convolutional-neural-network-matlab.html”

행동감지를 이용한 운동자세 교정 시스템 설계 및 구현

이철민*, 이동명*
*동명대학교 컴퓨터공학과
captaink7428@gmail.com, *dmlee@tu.ac.kr

Design and Implementation of Exercise Posture Correction System using Behavior Detection

Cheol Min Lee*, Dong Myung Lee*
*Dept. of Computer Engineering, Tongmyong Univ.

요 약

본 논문은 헬스 트레이너 없이도 사용자가 바른 자세로 운동할 수 있게 도움을 주는 행동감지를 이용한 운동자세 교정 시스템을 설계하고 구현하였다. 제안 시스템은 관절 좌표 추정 단계, 자세 판단 단계, 운동자세 교정 단계로 구성된다. 실험공간은 카메라의 영상에 머리부터 무릎까지 나오는 2.5m × 2.0m으로 설정하였다. 푸시업은 총 120회 실시되었고 푸시업 인식 정확도는 틀린 자세의 경우는 모두 100%, 바른 자세의 경우는 평균 95%로 측정되었다. 차후 제안한 시스템을 보완한 후 사용 범위를 확장해서 다른 운동의 자세 교정 시스템도 개발해 볼 예정이다.

Ⅰ. 서 론

최근 코로나의 단계적인 변이 바이러스 확산으로 사회적 거리 두기 상황이 계속 악화되고 있는 실정이다. 이러한 영향으로, 헬스장을 다니던 사람들도 집에서 운동하는 홈트레이닝 문화가 매우 빨리 정착되고 있다. 보통 개인적인 취향에 따라 운동을 하면 건강에 도움은 되겠지만 효과적인 운동에 의한 신체단련은 어렵다. 왜냐하면 바른 자세로 운동을 하지 않으면 근육이나 관절을 손상시킬 수 있기 때문에 헬스 트레이너의 도움이 필요하다.

본 논문은 최근 이러한 코로나 상황에서 특히 실내에서 헬스 트레이너 없이도 사용자가 바른 자세로 운동할 수 있도록 도움을 줄 수 있는 행동감지를 이용한 운동자세 교정 시스템을 설계하고 구현하였다.

Ⅱ. 관련연구

본 논문에서 제안하는 시스템에 관련된 유사연구로 두가지를 소개하면 다음과 같다.

1. 실시간 공간매핑을 이용한 객체 이동경로 추적

[1]은 오픈포즈 기반에서 실시간 공간 맵핑을 이용하여 객체의 이동경로를 추적하는 연구이다. 이 연구에서 제안하는 알고리즘은 사용자 추적 단계, 실시간 공간 맵핑 단계, 이동경로 추정단계로 구성된다. 사용자 추적 단계는 Human Pose Estimation 기술의 오픈포즈를 이용하여 사용자의 관절 좌표 정보를 획득하고 양쪽의 발목 좌표의 중심점을 사용자의 위치 정보로 지정한다. 실시간 공간 맵핑 단계에서는 공간의 추정 영역을 지정하고 Matrix 변환을 적용한 후 사용자 객제 정보와 매칭한다. 그리고 마지막 단계인 이동경로 추적 단계에서는 사용자의 위치 정보를 추출하고 저장하고 사용자의 이동경로를 획득한다.

2. 기계 학습을 사용한 실시간 자국어 수화 인식

[2]는 MediaPipe 및 기계 학습을 사용한 실시간 자국어 수화 인식 연구이다. 이 연구에서 제안하는 알고리즘은 MediaPipe의 Hand Pose Estimation 라이브러리와 다양한 머신러닝 모델을 이용하여 수화를 학습하고 인식한다. 사용한 머신러닝 모델은 DeepCNN, CNN (Convolutional Neural Networks), KNN (k-Nearest Neighbors), Random Forest, ANN (Artificial Neural Network), SVM 등인데, SVM (Support Vector Machines)에서 99%대, 평균적으로 CNN에서 97%, KNN과 Neural Network을 결합한 모델은 96%, DeepCNN에서가 94%의 정확도를 가짐을 확인하였다.

Ⅲ. 시스템 설계

1. 설계 고려사항

사용자의 신체를 표현하고 키포인트 (Key Point) 좌표를 추출하기 위해서 MediaPipe라이브러리를 설치한다. 그리고 좌푯값을 계산하기 위해 numpy를 설치한다. 카메라에 가려진 신체 키포인트 좌표는 오류에 원인이 될 수 있으므로 촬영에 방해요소가 적은 왼팔을 기준으로 시스템을 설계하였다.

2. 운동 자세교정 시스템의 구성

제안 시스템은 그림 1과 같이 3단계로 구성된다. 1단계는 사람의 자세를 시각적으로 확인할 수 있는 관절 좌표 추정 단계, 2단계는 푸시업 자세를 판단하는 자세 판단 단계. 그리고 3단계는 푸시업 자세가 정확한지를 판단하고 운동자세를 교정하는 운동자세 교정 단계이다.

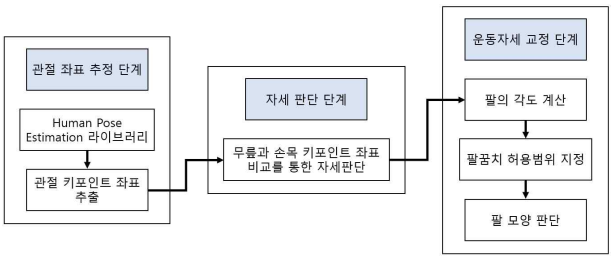


그림 1. 운동자세 교정 시스템의 단계 구성

3. 운동 자세교정 시스템 설계

관절 좌표 추정 단계에서는 HPE 라이브러리를 이용하여 사용자의 관절 키포인트 좌표를 추출한 후 이를 영상의 픽셀 좌표로 시각적 나타낸다. 푸

시업 자세 판단 단계에서는 추출한 무릎 좌표의 y축 좌표의 값이 팔꿈치의 y축 좌표의 값 보다 작으면 엎드림 또는 푸시업 자세로 판단한다.

운동자세 교정 단계에서 팔의 각도는 어깨, 팔꿈치 그리고 손목의 좌표를 이용하여 계산한다. 팔꿈치의 오차 허용범위 (L)는 손목의 x좌표에 팔꿈치 오차범위 가중치 (W)를 통해 계산한다. W는 식 (1)과 같이 팔꿈치 좌표(x_1, y_1)와 손목 좌표(x_2, y_2)간의 거리를 구한다. L은 손목의 x축을 기준으로 팔꿈치 오차범위 가중치 W를 더한 값인 최대 허용범위와 x축을 기준으로 W를 뺀 값인 최소 허용범위를 합산한 값이다. L은 사용자와 카메라의 거리에 따라 조정될 수 있다. 팔 모양 판단에서 팔의 각도가 일정 각도 이상으로 줄어들지 않으면 운동 횟수로 카운트하지 않는다.

육안으로 팔꿈치의 좌표 오차의 허용 범위의 설정은 팔꿈치 좌표와 손목의 좌표를 3으로 나눈 값으로 사용해야 하지만, 식 (2)와 같이 x_2에서 ±연산을 하기 위해 3에서 2를 나눈 값인 6을 적용하였다.

$$W = \frac{\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}}{6} \quad (1)$$

$$L = (x_2 + W) + (x_2 - W) \quad (2)$$

Ⅳ. 실험 및 결과

1. 실험 환경

환경구축에 사용된 언어는 python 3.8버전이고 라이브러리는 anaconda를 이용해서 정리하였다. Human Pose Estimation 기술을 사용하기 위한 라이브러리는 MediaPipe를 사용했다. 실험에는 엔비디아 그래픽카드 지포스 지티엑스 1050ti가 탑재된 노트북을 사용하였고 Windows 10 환경에서 실험을 진행하였다. 실험공간은 카메라에 실험자의 신체가 머리부터 무릎까지 나오는 공간인 2.5m × 2.0m으로 설정하였다.

2. 실험 및 결과

실험은 사람이 카메라 앞에서 푸시업을 하는 방식으로 진행되었다. 푸시업은 총 120회 실시되었고, 바른자세 푸시업을 카메라의 왼쪽, 중앙, 그리고 오른쪽 위치에서 각각 20회를, 틀린 푸시업 또한 동일한 위치에서 각각 20회 진행하였다. 사람과 카메라와의 거리 설정은 사용 카메라의 해상도에 따라 달라지지만 이번 실험에서 사람과 카메라간의 거리는 실험에 사용된 카메라의 해상도를 고려해서 대략 2.5m 거리에서 설정하였다.

그림 2의 왼쪽은 올바르지 못한 푸시업 자세를 그리고 오른쪽은 바른 푸시업 자세 나타내는 것이다. 올바르지 못한 푸시업 자세의 경우 회당 횟수의 변화가 전혀 없었지만 바른 푸시업 자세의 경우는 회당 횟수가 증가함을 확인할 수 있었다.

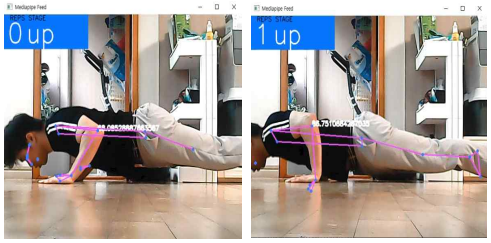
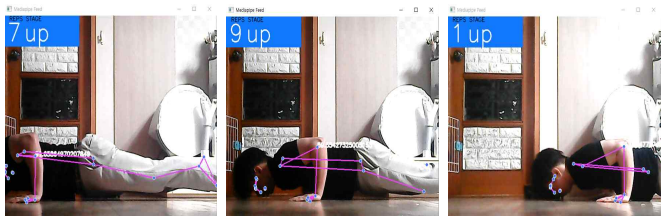


그림 2. 실험장면

실험은 그림 3과 같이 카메라의 왼쪽 (A), 중앙 (B), 오른쪽 (C) 위치에서 진행하였고, 바른 자세와 틀린 자세에 따른 푸시업 인식 정확도를 표 1에서 보는 바와 같이 측정 및 분석하였다. 바른 자세의 푸시업 인식 정확도는 A위치에서 90%, B위치에서 100%, 그리고 C위치에서 95%로 각각 인식되었다. 그리고 틀린 자세에서의 카운터 횟수는 A, B, C의 위치에서 모두 0개로 인식하였기 때문에 푸시업 인식 정확도는 100%로 인식 되었다. 그러나 바른 자세 푸시업에서 실험결과로 얻어진 인식률은 카메라에 붙은 볼록렌즈가 이미지의 왜곡을 유발하여 푸시업을 하는 팔의 각도나 무릎의 좌표를 부정확하게 인식하게 하는 요인으로 추측하고 있다.



왼측에서 촬영 (A) 중앙에서 촬영 (B) 오른쪽에서 촬영 (C)

그림 3. 사용자의 푸시업 위치

표 1. 푸시업 인식 정확도

자세 유형	사용자 위치	카운트횟수	푸시업 인식 정확도
바른 자세	A	18/20	90%
	B	20/20	100%
	C	19/20	95%
틀린 자세	A	0/20	100%
	B	0/20	100%
	C	0/20	100%

Ⅴ. 결론

본 논문에서는 HPE의 MediaPipe 라이브러리를 이용하여 사용자의 푸시업 자세를 교정하는 시스템을 설계 및 구현하였다. 푸시업 자세판단 단계에서 팔의 모양을 판단하기 위한 각도 계산 알고리즘, 푸시업이 바른 자세인지 아닌지를 판단하는 팔꿈치 오차 허용범위 지정 알고리즘을 제안하였다. 그리고 운동자세 교정 단계에서 팔이 접혔다 펴졌음을 확인할 수 있는 팔 모양 판단 알고리즘을 제안하였다.

실험결과, 틀린 자세는 카운터가 되지 않았고, 바른 자세는 횟수마다 카운터가 되었음을 확인하였다. 푸시업을 하는 위치에 따라 이러한 현상이 발생한 이유는 카메라의 볼록렌즈로 인한 이미지 왜곡 현상이 푸시업을 하는 팔의 각도나 무릎의 좌표를 부정확하게 인식하게 하는 것으로 추측하고 있다. 이 문제점은 규명중에 있으며, 차후 제안한 시스템을 보완한 후 사용 범위를 확장해서 다른 운동의 자세 교정 시스템도 개발해 볼 예정이다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학 지원 사업의 연구결과로 수행되었음(2018001874004).

본 논문은 부산광역시 및 (재)부산인재평생교육진흥원의 BB21플러스 사업으로 지원된 연구임.

참 고 문 헌

[1] 신승현, 김태완, 이동명, “오픈포즈 기반 실시간 공간 맵핑을 이용한 객체 이동경로 추적 알고리즘에 관한 연구,” 2020년도 한국통신학회 하계종합학술발표회 논문집 (ISSN: 2383-8302), vol. 72, 용평리조트 타워콘도, pp.1363-1364, 2020. 8. 12-14.

[2] Arpita Halder and Akshit Tayade, “Real-time Vernacular Sign Language Recognition using MediaPipe and Machine Learning,“ International Journal of Research Publication and Reviews (ISSN 2582-7421), Vol (2), Issue (5), pp.9-17, 2021.

합성곱 신경망을 이용한 자동 변조 분류 기법 설계와 성능 분석

오영우, 최우열*
조선대학교
snlyoungwoo@gmail.com, *wyc@chosun.ac.kr

Design and Performance Analysis of Automatic Modulation Classification Using Convolutional Neural Network

Youngwoo Oh, Wooyeol Choi*
Chosun University

요 약

본 논문은 합성곱 신경망(Convolutional neural network, CNN)을 이용하여, 자동 변조 분류 기법을 설계 및 구현한다. 입력데이터는 Inphase-quadrature(I/Q) 데이터를 이용하며 3개의 아날로그 변조와 8개의 디지털 변조를 포함한 11종에 대하여, 5-25dB까지 다양한 신호 대 잡음 비율(Signal to noise ratio, SNR) 환경에서의 변조 인식 성능을 분석하였다. 실험 결과를 통해, CNN 기반 자동 변조 분류 기법이 타 기법들과 비교해 월등히 우수한 성능을 보이는 것을 확인하였으나, 신호 대 잡음 비율에 따른 분류 정확도의 편차와 M-PSK 계열 변조 간 오인식 문제를 보인다. 따라서, 이러한 식별 정확도의 개선을 위해 상황에 맞는 적절한 학습데이터 선정 및 신경망 구조의 최적화가 요구된다.

Ⅰ. 서론

자동 변조 분류(Automatic modulation classification, AMC)는 무선 통신 시스템에서 전송 신호에 대한 사전 정보 없이 수신 신호의 변조 방식을 예측하는 기술이다. 특히, 군사 통신에서의 적 신호 탐지와 인지 통신 분야에서의 사용자 신호를 블라인드 검출하기 위해 사용된다.

전통적인 자동 변조 분류 기법으로는 우도 함수 기반(Likeliood-based, LS)의 기법[1]과 특징값 기반(Feature-based, FB) 기법[2]이 존재한다. 우도 함수 기반의 기법은 최적의 성능을 제공하지만, 수신 신호에 대한 정확한 사전 정보가 요구된다. 반면 특징값 기반의 분류 기법에서는 수신 신호의 주기적 특성을 추출하는 방법에 따라 성능의 편차가 크게 나타난다.

이러한 한계점들을 극복하고자 최근 딥러닝(Deep learning, DL) 기반의 자동 변조 분류 기법에 관한 연구가 활발히 진행되고 있으나, 대다수의 기법이 가우시안 채널에 대해서 설계되어, 페이딩 환경에서의 불안정한 예측 성능을 보인다[3]. 따라서, 본 논문에서는 페이딩 환경을 고려한 합성곱 신경망 기반의 자동 변조 분류를 설계하고, 신호 대 잡음 비율의 변화에 따른 인식 정확도 성능의 평가를 수행한다.

Ⅱ. 본론

2.1 페이딩을 고려한 데이터셋 구성

본 연구에 활용된 데이터셋은 총 105,243개의 프레임으로 구성된다. 각 프레임의 길이는 1,024개의 샘플로 구성되며, 샘플 당 8개의 심볼을 적용하였다. 총 변조 종류는 11개로 표 1과 같이 디지털 변조 8개와 아날로그 변조 방식 3개로 도출하였다. 이때, 전체 11개 클래스를 가지며, 클래스별로 최소 10,000개의 프레임이 사용되었다. 신호 대 잡음 비율은 5dB에서 25dB이며 5dB의 간격으로 각 프레임을 생성하며, 채널에 의해 손상되는 합성 파형을 생성하기 위해, Rayleigh fading, AWGN, 주파수·클록 오프셋 등을 통해 손상된 I/Q 신호를 수집하였다. 해당 디지털 변조 유형의 중심주파수는 902MHz, 아날로그 변조 유형의 중심주파수는 100MHz로 지정하였다.

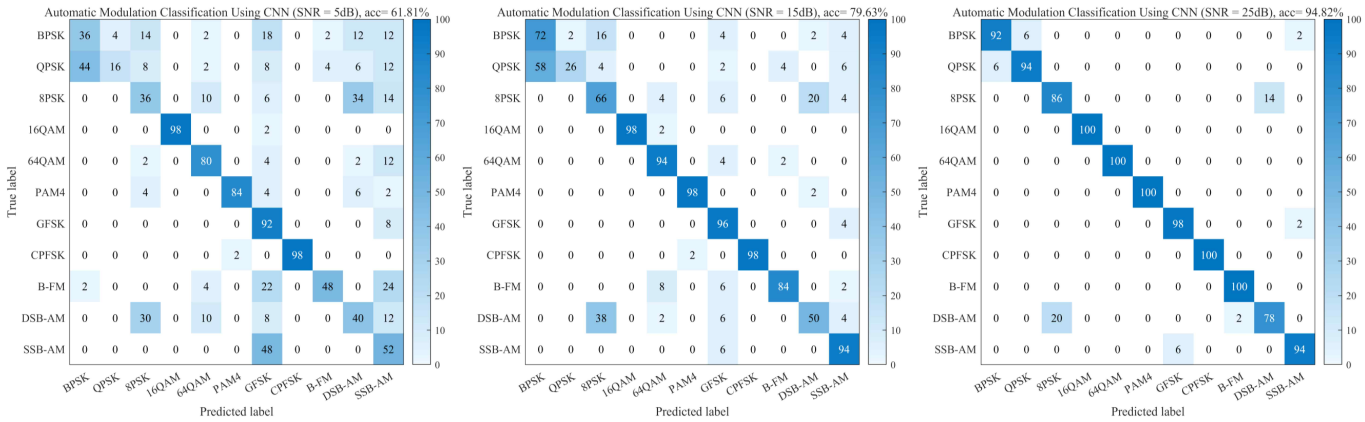
Number of training frame	78,540 (75%)
Number of validation frame	15,708 (15%)
Number of test frame	10,995 (10%)
Sample length	1024
SNR (dB)	5dB~25dB
Modulation type (11)	BPSK, QPSK, 8-PSK, 16-QAM, 64-QAM, PAM4, GFSK, CPFSK, B-FM, DSB-AM, SSB-AM

표 1. 자동 변조 분류를 위한 데이터셋

2.2 CNN 기반의 자동 변조 분류

본 논문에서는 실제 무선 채널에서 나타날 수 있는 현상들을 반영하기 위해, 변조 신호에 대한 Rayleigh fading, 위상 회전 및 AWGN 특성이 반영된 다양한 입력데이터를 생성하였다. 이때, 전체 데이터 중 학습데이터는 75%, 검증데이터 15%, 실험에 사용되는 데이터를 10%로 적용하였으며, 최대 학습 회수(epoch)는 15회, 배치사이즈는 512로 설정하였다.

사용된 CNN 모델의 경우, 6개의 합성곱 계층과 1개의 완전 연결 계층으로 구성된다. 합성곱 계층은 배치 정규화 계층과 ReLU 활성화 함수 계층, 최댓값 폴링 계층으로 구성되었으며, 신호 대 잡음 비율 변화에 따른 정확도에 대한 성능 분석을 수행한다.



(a) SNR = 5dB (b) SNR = 15dB (c) SNR = 25dB

그림 1. 신호 대 잡음 비율 변화에 따른 CNN 기반 자동 변조 분류의 혼동행렬

2.3 시뮬레이션 결과 분석

그림 1은 SNR 변화에 따른 CNN 기반 자동 변조 분류의 정확도를 혼동행렬(Confusion matrix, CM)로 표현한 결과를 보여준다. 비교적 신호 대 잡음 비율이 낮은 그림 1(a)의 경우, 페이딩 및 주파수·위상 오프셋 등으로 인해 전체적으로 변조 분류의 정확도가 떨어지는 것을 확인할 수 있다.

그림 1(b)는 비교적 신호 품질이 개선된 환경에서의 변조 분류 성능을 나타낸다. (a)에 비해, BPSK, 8-PSK, 64-QAM, PAM4, SSB-AM 변조에 대한 분류 성능이 크게 개선되었으며, 오인확률 또한 일정 수준 감소하였다. 그러나, BPSK, QPSK, DSM-AM과 같은 일부 변조 분류에서는 여전히 낮은 정확도를 보인다.

신호 대 잡음 비율이 높은 그림 1(c)에서는 기존에 발생하는 오인확률이 크게 개선되었음을 확인할 수 있다. 그러나, M-PSK 계열의 디지털 변조 신호 간 오인식 문제가 존재하는데, 이는 I/Q 데이터를 생성하는 과정에서 추가되는 주파수·위상 오프셋과 Rayleigh fading 채널 통과에 따른 constellation의 위상 회전으로 인해 발생한다. 또한, 이러한 위상 변화는 변조 신호 간의 유사성을 증가시키며, 결과적으로 신경망의 학습 성능에 영향을 준다.

그림 2는 자동 변조 분류 기법에 따른 변조 예측 성능을 나타낸다. 이때, CNN 기반의 자동 변조 분류 기법은 기존의 방식들과 비교해 29-47%가량의 개선된 정확도를 보이는 것을 확인할 수 있다. 그러나, 신호 대 잡음 비율 정도에 따라 정확도의 편차가 커지는 것 역시 확인할 수 있다. 따라서, 안정적인 성능을 보장하기 위해 상황에 맞는 적절한 학습데이터를 선정하거나, 신경망 구조의 최적화를 통하여 변조 분류 성능의 신뢰성을 회복할 수 있을 것으로 예상된다.

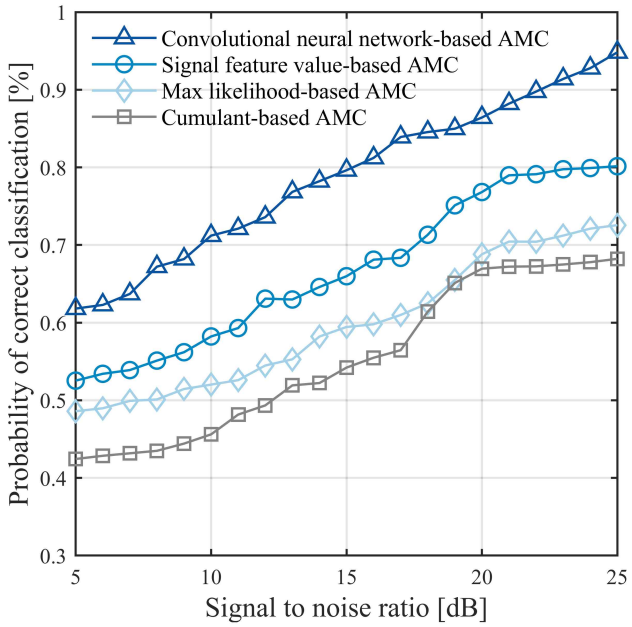


그림 2. 자동 변조 분류 기법에 따른 예측 정확도 비교

Ⅲ. 결론

본 논문에서는 CNN 기반의 자동 변조 분류 기법을 설계 및 구현하였으며, 주파수 및 위상 오프셋과 Rayleigh fading 채널 특성을 고려한 상황에서의 SNR 변화에 따른 변조 정확도 성능을 평가하였다. 실험 결과를 통해, CNN 기반의 변조 분류 기법이 타 기법들과 비교해 월등히 우수한 성능을 보이는 것을 확인하였으나, 신호 대 잡음 비율에 따른 분류 성능의 편차와 M-PSK 계열 간 오인식 문제를 보인다. 따라서, 향후 연구에서는 이러한 문제를 해결하기 위해 최소 평균 제곱(Least mean squares, LMS) 기반의 채널 등화를 입력데이터에 적용하여, 개선된 CNN 기반 자동 변조 분류 기법을 설계 및 구현할 예정이다.

ACKNOWLEDGMENT

본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(No. NRF-2019R1F1A1063345)과 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학지원사업의 연구결과로 수행되었음 (2017-0-00137).

참 고 문 헌

[1] F. Hameed, O. A. Dobre and D. C. Popescu, "On the likelihood-based approach to modulation classification." *in IEEE Transac. Wireless Commun.*, vol. 8, no. 12, pp. 5884-5892, Dec. 2009.

[2] V. D. Orlic and M. L. Dukic, "Automatic modulation classification algorithm using higher-order cumulants under real-world channel conditions." *in IEEE Communications Letters*, vol. 13, no. 12, pp. 917-919, Dec. 2009.

[3] 김정현, 김아경, 유소영. "자동 변조 분류를 위한 딥러닝 기법들의 성능 분석." *한국통신학회 학술대회논문집 (2020).* pp. 26-27.

다수의 병렬 가중 모델 기반 스마트팩토리 에너지 사용량 예측 알고리즘에 관한 연구

여현영+, 심서영+, 홍수민+, 차채연*, 박형곤+*

이화여자대학교 전자전기공학전공+, 이화여자대학교 스마트팩토리융합전공*

{eyhy1016, yeni2000, hsum916, chaeyeon.cha}@ewhain.net, hyunggon.park@ewha.ac.kr

Energy Consumption Prediction Algorithm for Smart Factory Based on Multiple Parallel Weighted Models

Hyeon-yeong Yeo+, Seoyoung Sim+, Sumin Hong+, Chaeyeon Cha* and Hyunggon Park+*

Department of Electronic and Electrical Engineering, Ewha Womans University+

Graduate Program in Smart Factory, Ewha Womans University*

요 약

본 논문은 머신러닝을 통한 스마트팩토리의 에너지 소비량 예측의 정확성을 높이고자 다수의 병렬 가중 모델을 활용하는 방법을 제안한다. 제안하는 방법은 여러 일 동안 수집된 하루 에너지 사용량 데이터를 하루씩 나누어 다수의 모델에 병렬로 학습시킨 후 각각의 모델의 성능을 평가한 가중치를 예측값에 곱하여 합하는 방식으로 최종 예측치를 얻는다. 실제 스마트팩토리 데이터를 활용한 실험을 통하여 제안한 방법이 하나의 머신러닝 모델에 직렬로 학습시키는 방법보다 평균 제곱근 편차가 작고 예측 성능이 우수함을 확인하였다.

Ⅰ. 서 론

스마트팩토리는 4차 산업 혁명의 핵심 요소 중 하나로 다양한 센서 등을 활용하여 상황을 인지하고 실시간으로 반응할 수 있는 지능형 공장을 의미한다[1]. 이러한 스마트팩토리의 빅데이터 수집, 생산 라인 모델링, 예측 및 생산 지능화에 대한 연구가 전 세계적으로 활발히 이루어지고 있다[2]. 그 중 스마트팩토리의 핵심 기술인 IoT(Internet of Things)와 관련하여 주목받고 있는 스마트미터는 실시간으로 에너지 사용량 데이터를 수집할 수 있고 더 나아가 본 논문의 목적인 에너지 사용량 예측을 통한 에너지 효율성 향상 및 생산 관리 결정에 중요한 역할을 한다[3]. 그러나 생산 활동 전반이 자동화, 지능화 및 자율화된 스마트팩토리의 특성상 작동 시간 및 에너지 사용량에 대한 변동 폭이 크고 데이터 수집 및 측정 과정에서 잡음이 포함될 가능성이 있다. 이러한 변동 폭은 머신러닝을 통한 예측 성능을 저하시켜 예측 성능 향상에 근본적인 한계치를 만든다[4], [5]. 본 논문에서는 이러한 머신러닝의 성능을 향상시키기 위하여 활용 데이터를 병렬로 학습시킨 다수의 개별적인 머신러닝 모델을 만든 후 각 모델의 예측 결과에 가중치를 주어 최종 예측값을 얻는 방법을 제안한다.

Ⅱ. 본론

A. 문제 정의

본 논문에서는 n일 동안 수집된 하루 에너지 사용량 데이터를 통해 $n+1$일째 하루 에너지 사용량을 예측하고자 한다. $i(i=1, \cdots, n)$일 하루를 균등한 시간 간격으로 나누어 총 m번 에너지 사용량을 측정하였을 때, $t(t=1,\cdots,m)$번째 시간 간격에서 수집된 에너지 사용량 데이터를 $y_i(t)$라 하자. i일 하루 동안 수집된 에너지 사용량 벡터를 $\boldsymbol{x_i}$라 하면, 벡터 $\boldsymbol{x_i}$는 다음과 같이 표현된다.

$$\boldsymbol{x_i}=\left(y_i(1), y_i(2), \cdots, y_i(m)\right). \quad (1)$$

n일 동안 수집된 하루 에너지 사용량 벡터의 집합을 $\boldsymbol{X_n}$이라 하자. 집합 $\boldsymbol{X_n}$는 다음과 같이 표현된다.

$$\boldsymbol{X_n}=\{\boldsymbol{x_i} | i=1,\cdots,n\} \quad (2)$$

n일 동안 수집된 에너지 사용량 벡터의 집합 $\boldsymbol{X_n}$을 머신러닝의 훈련 집합으로 사용하여 예측한 $n+1$일의 에너지 사용량 벡터를 $\hat{\boldsymbol{x}}_{\boldsymbol{n+1}}$라 하면, 벡터 $\hat{\boldsymbol{x}}_{\boldsymbol{n+1}}$는 다음과 같이 표현된다.

$$\hat{\boldsymbol{x}}_{\boldsymbol{n+1}}=\left(\hat{y}_{n+1}(1), \hat{y}_{n+1}(2), \cdots, \hat{y}_{n+1}(m)\right) \quad (3)$$

본 논문에서는 예측의 정확도를 평가하기 위해 평균 제곱근 편차(RMSE)를 사용하였으며 각 시간 간격 t에서 예측한 에너지 사용량과 실제 에너지 사용량을 통해 구한다.

$$RMSE\left(\boldsymbol{x_{n+1}}, \hat{\boldsymbol{x}}_{\boldsymbol{n+1}}\right)=\frac{1}{m}\sqrt{\sum_{t=1}^{m}\left|y_{n+1}(t)-\hat{y}_{n+1}(t)\right|^2} \quad (4)$$

변동 폭이 큰 데이터를 기반으로 일반적인 머신러닝을 통해 $\hat{\boldsymbol{x}}_{\boldsymbol{n+1}}$을 예측할 때, 예측의 정확도가 떨어지는 경향이 있다[4], [5]. 본 논문에서는 이를 해결하기 위해 데이터 전처리와 제안 방법을 통해 파동이 큰 데이터에서도 예측의 성능을 높이고자 한다.

B. 제안 방법

본 논문에서는 지수 이동 평균으로 잡음을 제거한, 하루 동안 1분 간격으로 총 1,440번($m=1{,}440$) 측정된 하루 동안의 에너지 사용량 데이터 $\boldsymbol{x_i}$를 사용하여 다수의 머신러닝 모델을 병렬로 학습시킨 뒤, 학습된 모델의 예측 성능을 기반으로 한 가중 모델을 제안한다. 본 논문에서 제안하는 모델은 그림 1과 같다. 하나의 머신러닝 모델을 M^i라 하면, M^i는 $\{\boldsymbol{x_i}, \boldsymbol{x_{i+1}}\}$를 기반으로 i일을 입력하였을 때 $i+1$일의 하루 에너지 사용량을 예측하도록 학습한다. $n+1$일을 예측하기 위해 활용되는 다수의 병렬 머신러닝 모델들은 $n-1$일 동안 수집된 데이터 집합 $\boldsymbol{X_{n-1}}$을 $\{\boldsymbol{x_i}, \boldsymbol{x_{i+1}}\}$ $(i=1,\cdots n-2)$로 나누어 병렬로 학습하는 과정을 통해 총 $n-2$개 생성된다.

제안하는 모델은 학습된 M^i에 $\boldsymbol{x_{n-1}}$을 입력하였을 때 n일에 대해 예측한 성능을 기반으로 가중치를 정한다. 머신러닝 모델 M^i가 n일에 대하여 예측한 결과를 에너지 사용량 벡터 $\hat{\boldsymbol{x}}_{\boldsymbol{n}}^{\boldsymbol{i}}$라 하면, $\hat{\boldsymbol{x}}_{\boldsymbol{n}}^{\boldsymbol{i}}$와 실제 n일에 대한 데이터 $\boldsymbol{x_n}$을 통해 각각의 모델 M^i에 대한 $RMSE(\boldsymbol{x_n}, \hat{\boldsymbol{x}}_{\boldsymbol{n}}^{\boldsymbol{i}})$인 e^i 을 계산한다. 편차와 예측 성능은 반비례하므로 이 편차의 역수로 가중치 w^i를 얻게

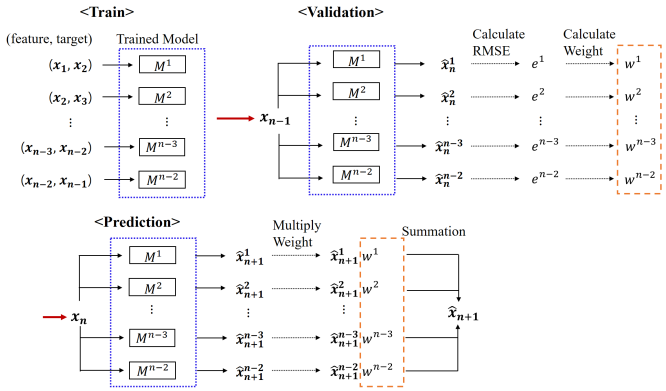


그림 1. 병렬 생성 모델을 활용한 제안 방법.

되며 가중치 w^i $(i=1, \cdots, n-2)$는 다음과 같이 표현된다.

$$w^i = \frac{1/e^i}{\sum_{i=1}^{n-2} 1/e^i} \tag{5}$$

제안 모델은 병렬로 학습된 각각의 머신러닝 모델 M^i에 $\boldsymbol{x_n}$ 데이터를 입력해 $n+1$일을 예측한 $\hat{\boldsymbol{x}}_{n+1}^{i}$을 추출하고 M^i마다 계산된 가중치 w^i를 예측값 $\hat{\boldsymbol{x}}_{n+1}^{i}$에 각각 곱한 후 더하여 $n+1$일에 대한 예측값 $\hat{\boldsymbol{x}}_{n+1}$을 얻는다.

$$\hat{\boldsymbol{x}}_{n+1} = \sum_{i=1}^{n-2} w^i \hat{\boldsymbol{x}}_{n+1}^{i} \tag{6}$$

Ⅲ. 구현 및 실험 결과

본 논문에서 제안한 모델의 성능을 평가하기 위하여 독일 쾰른-아헨 지역의 FINESCE 시범 스마트팩토리에서 1분 간격으로 수집한 4가지 기계 레이저 절단기(laser cutting), 벤딩기(bending), 3D 로봇 웰터(3D robot welter), 레이저 형삭기(laser shaping)의 하루 동안 에너지 사용량 데이터를 사용하였다[6]. 결측값이 없는 데이터를 활용하기 위하여 2015년 5월 18일부터 6월 15일까지 4주 동안 수집된 데이터를 사용하여 에너지 사용량을 분 단위로 모두 더한 총 에너지 사용량 데이터를 활용하였다. 실험에서 사용한 머신러닝 모델은 Decision Tree, Adaboost 및 kNN(k-Nearest Neighbors)이며 본 논문에서는 각각을 DT, AB, kNN으로 나타낸다.

제안 방법의 병렬 모델 개수에 따른 예측 성능을 평가를 위해 n $(n=4,5 \cdots, 17)$을 지정하여 데이터 길이에 맞춰 머신러닝 모델 M^i를 각각 2,3,⋯,15 개 생성하여 $n+1$일의 에너지 사용량을 예측하였다. 그림 2에서 실험에 사용한 3개의 머신러닝 모델에 대해 병렬 모델을 13개 사용하였을 때 예측 성능이 가장 우수하고 모델을 2개 사용한 경우보다 성능이 평균 14.41% 향상된 것을 확인할 수 있다. 이는 경향성이 없는 데이터를 학습할 때, 데이터를 많이 사용할수록 모델이 정확하게 학습하기 때문이다. 또한, 본 논문에서는 일반적인 머신러닝 방법과 제안 방법의 성능을 비교하기 위해 활용 데이터 길이 $n=15$로 하였으며, 제안 방법에서는 병렬 모델 개수를 13개로 지정하여 15일 동안 수집한 데이터로 16번째 날의 하루 에너지 사용량을 예측하였다. 표1에서 임의의 4일에 대해 제안 방법이 일반적인 머신러닝 방법에 비해 AB 최대 2.2%, DT 최대 17.2%, kNN 최대 8.2%로 예측 성능이 향상된 것을 확인할 수 있다.

Ⅳ. 결론

본 논문에서는 변동폭이 큰 스마트팩토리 데이터의 예측 성능을 향상하기 위해 병렬로 생성된 머신러닝 모델의 예측 성능에 따른 가중치를 부여하여

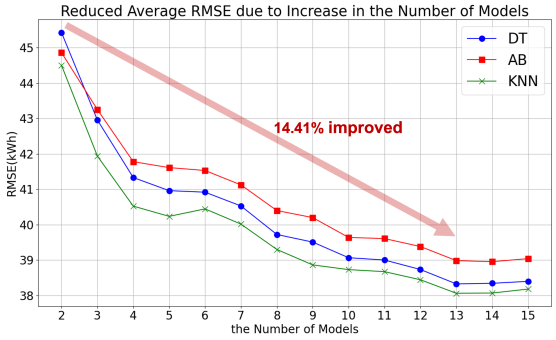


그림 2. 병렬 모델 개수에 따른 제안 방식의 예측 성능.

표 1. 직렬 학습 모델 1개 , 병렬 모델 13개 예측 성능 비교

RMSE (kWh)	Day 1			Day 2			Day 3			Day 4		
	DT	AB	kNN	DT	AB	kNN	DT	AB	kNN	DT	AB	kNN
Series	46.46	41.74	41.90	47.12	44.18	43.74	48.75	41.22	43.85	37.63	31.78	32.77
Proposed	40.10	40.84	39.90	42.81	43.26	42.50	40.37	40.92	40.27	30.03	30.93	29.58

최종 예측을 하는 다수의 병렬 가중 모델을 제시하였다. 실험을 통해 기존 방법보다 우수한 예측 성능을 확인하였으며 병렬 모델의 수가 증가할수록 학습에 사용된 데이터가 많아 성능이 더욱 향상되는 것을 확인할 수 있었다. 잡음이 있고 경향성이 없는 데이터에 대하여 제안 방법을 통한 예측을 통해 높은 예측 정확성을 보일 것으로 기대된다.

ACKNOWLEDGMENT

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(No.2021-0-00739, 분산/협력AI 기반 5G+ 네트워크 데이터 분석 기능 및 제어 기술 개발, No. 2019-0-00024, 네트워크 자동화를 위한 개방형 네트워크 데이터 분석 기반 지도형 애자일 머신러닝 기술 개발)과 2021년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. NRF-2020R1A2B5B01002528).

참 고 문 헌

[1] 정병도, “스마트 팩토리 연구동향,” TTA(한국정보통신기술협회) journal, 181호, pp. 46-51, Jan. 2019.

[2] X. Xu and Q. Hua, "Industrial Big Data Analysis in Smart Factory: Current Status and Research Strategies," *IEEE Access*, Vol. 5, pp. 17543-17551, 2017.

[3] F. Shrouf, J. Ordieres and G. Miragliotta, "Smart factories in Industry 4.0: A review of the concept and of energy management approached in production based on the Internet of Things paradigm," *IEEE International Conf. on Industrial Engineering and Engineering Management*, pp. 697-701, 2014.

[4] S. Gupta and A. Gupta, “Dealing with Noise Problem in Machine Learning Data-sets: A Systematic Review,” *Procedia Computer Science*, Vol. 161, pp. 466-474, 2019.

[5] G. Li, J. Zrimec, B. Ji, J. Geng, J. Larsbrink, A. Zelezniak, J. Nielsen, and M. K. Engqvist, “Performance of Regression Models as a Function of Experiment Noise,” *Bioinformatics and Biology Insights*, Vol. 15, pp. 1-10, Jan. 2021.

[6] *Smart Energy Data: Aachen / Cologne Smart Factory*, Aachen / Cologne, Germany, Aug. 25, 2015. [Online] Available: https://data.lab.fiware.org/dataset/smart_energy_data-_aachen__cologne_smart_factory

최대 풀링 레이어를 위한 저면적 재구성 가능한 하드웨어 구조

김은서, 신동엽, 임용석
한국전자기술연구원 스마트네트워크센터

golroom97@keti.re.kr, dongyeob.shin@keti.re.kr, busytom@keti.re.kr

Low Area Reconfigurable Hardware Architecture for Max Pooling Layer

Eun-Seo Kim, Dongyeob Shin, Yong-Seok Lim

Smart Network Center
Korea Electronics Technology Institute

요 약

합성곱 신경망(Convolutional Neural Network)의 적용 범위가 확장됨에 따라 풀링 레이어(Pooling Layer)의 필터 크기 및 스트라이드 크기 또한 다양해지고 있다. 이에 본 논문에서는 풀링 레이어의 필터 크기와 스트라이드 크기를 가변적으로 사용할 수 있는 하드웨어 구조를 제안한다. 결과적으로 재구성 가능한 최대 풀링 하드웨어 구조는 최대 풀링 연산을 처리하는 면적을 감소시킬 수 있다.

Ⅰ. 서 론

합성곱 신경망(Convolutional Neural Network)은 이미지 분류, 객체 인식 분야에서 가장 뛰어난 성능을 보인다. 합성곱 신경망은 특징 추출(Feature Extraction)과 분류(Classfication) 기능으로 나누어지며, 이 중 특징 추출은 일반적으로 특징맵(Feature Map) 연산을 위한 합성곱 레이어(Convolutional Layer)와 특징맵의 크기를 줄여 과적합(Overfitting)을 방지하는 풀링 레이어(Pooling Layer)로 구성된다[1]. 풀링의 종류에는 최대 풀링(Max Pooling), 평균 풀링(Average Pooling) 등이 있으며 특히 최대 풀링은 대부분의 합성곱 신경망에서 사용된다. 최대 풀링은 신경망 구조에 따라 적절한 필터(Filter) 크기와 스트라이드(Stride)를 가질 수 있다. 최근 NASNet 과 같이 NAS(Neural Architecture Search)를 통해 자동 생성하는 최적의 신경망 구조들은 2x2, 3x3, 5x5, 7x7 등 다양한 필터 크기 중 최적의 필터 크기를 갖도록 설계되고 있다[2].

본 논문에서는 다양한 필터 크기의 최대 풀링을 지원할 수 있는 저면적 재구성 가능한 하드웨어 구조를 제안한다. 기존의 최대 풀링 하드웨어는 고정된 필터 크기에 적합하게 설계되어 있어, 이를 가변하기 위해서는 큰 면적의 비효율적인 구현이 불가피하다[3]. 제안하는 하드웨어는 폴딩(Folding) 구조를 통해 면적 오버헤드를 최소화하며 필터 크기를 가변할 수 있도록 설계되었다.

Ⅱ. 본론

2.1. 최대 풀링 레이어

최대 풀링 레이어는 합성곱 레이어의 연산 결과인 특징 맵을 입력으로 하여 필터 크기와 스트라이드에 따라 특징 맵의 크기를 감소시킨다. 그림 1 은 2x2 필터, 스트라이드 2 인 경우와 3x3 필터, 스트라이드 2 인 경우의 최대 풀링 동작을 각각 나타낸다. 그림 1 과 같이 최대 풀링은 필터 윈도 영역의 픽셀 중 가장 큰 값만 추출하여 다음 레이어의 입력이 될 축소된 특징 맵을 생성한다. 하드웨어 구현 시, 필터 크기에 따라 최대 풀링 하드웨어에 입력되는 픽셀 개수 및 비교 연산 횟수가 달라지며, 스트라이드 크기에 따라 처리해야 하는 픽셀들이 달라지는 것을 고려해야 한다.

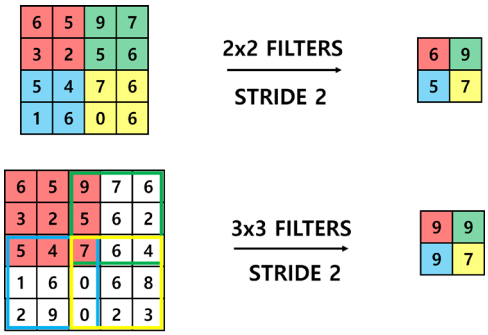


그림 1. 최대 풀링 연산

2.2 제안 기법

먼저 기존 기법과 제안 기법에서 공통적인 최대 풀링 하드웨어의 입력 부분은 다음과 같다. 그림 2 는 32x32 입력 특징 맵의 최대 풀링 하드웨어 입력 예시이다. 그림에서는 입력 특징 맵을 열 단위(32x1)로 처리(행 단위도 동일)하는 것을 가정한다. 그림 2 처럼 스트라이드의 크기가 2 일 때 첫 번째 열인 column[0]에 존재하는 32 개의 데이터 중 짝수 번째 행의 데이터가 최대 풀링 연산의 시작점이다. 시작점이 되는 데이터를 기준으로 연속적인 7 개의 데이터가 하나의 풀링 연산의 입력이 된다. 그림 2 의 최대 풀링 하드웨어 입력 부분은 MUX 와 FIFO 로 구성되어 최대 풀링 오퍼레이터(Max Pooling Operator) 하드웨어 앞단에 위치한다.

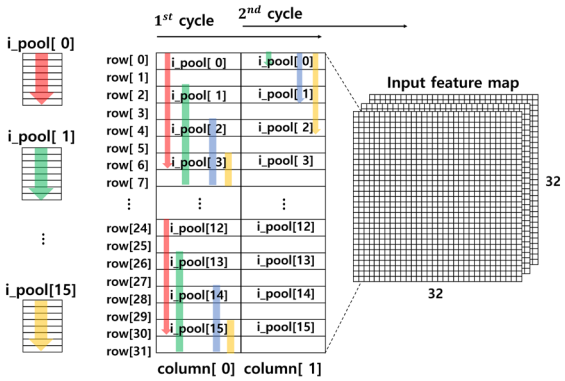


그림 2. 최대 풀링 하드웨어 입력 데이터 예시

그림 3 은 기존 필터 크기를 가변하는 최대 풀링의 하드웨어[3] 구조를 나타낸다. 그림 3 과 같이 기존 최대 풀링 오퍼레이터 하드웨어는 입력 픽셀에 대한 최댓값 비교 연산을 동시에 수행하기 때문에 가변하는 필터 크기 중 가장 큰 필터 크기에 따라 면적이 결정되고, 이는 그보다 작은 필터 크기로 동작 시에 자원을 낭비하는 요소가 된다. 예를 들어, 그림 3 에서와같이 2x2, 3x3 필터를 가변할 수 있도록 구현하는 경우, 2x2 필터 크기로 최대 풀링 시 3x3 최대 풀링에 맞춰 설계된 최댓값 비교 연산기의 하나의 입력은 사용하지 않게 된다. 즉, 기존 최대 풀링 하드웨어는 가변하는 필터의 크기 차이가 벌어질수록 효율이 떨어질 수 있다.

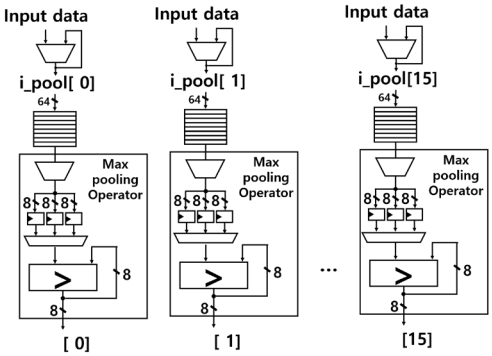


그림 3. 필터 크기를 가변하기 위한 기존 최대 풀링 하드웨어 구조

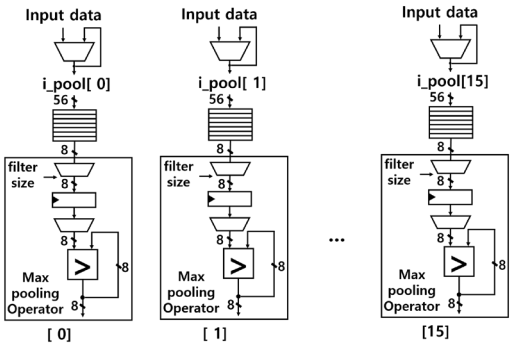


그림 4. 필터 크기를 가변하기 위한 제안 최대 풀링 하드웨어 구조

본 논문에서 제안하는 하드웨어 구조는 그림 4 에 나타나 있다. 그림 4 에서 제안하는 최대 풀링 하드웨어는 폴딩 구조를 활용하여 최댓값 비교 연산을 두 개의 피연산자에 대해서만 수행하기 때문에 연산을 처리하는 면적이 필터의 크기에 의해 제한되지 않는다. 또한 1) 필터의 크기에 따라 입력 데이터 순서를 결정하는 MUX 를 통해 비교 연산자의 입력이 될 수 있는 데이터의 개수를 가변하고 2) 비교 연산의 반복 횟수는 카운터를 통해 제어하여, 단일 하드웨어로 필터의 크기를 2x2 부터 7x7 까지 다양하게 사용하는 것이 가능하다. 동시에 제안하는 하드웨어 구조는 비교 연산기의 모든 입력을 사용하여 높은 하드웨어 활용도를 가질 수 있다.

2.3. 구현 결과

표 2 는 기존 최대 풀링 하드웨어 구조와 제안하는 하드웨어 구조 결과를 비교한 것이다. 하드웨어 구현 결과는 합성 결과를 사용하였으며, 합성을 위해서는 시놉시스(Synopsys) 디자인 컴파일러(Design Compiler) 툴과 130nm CMOS 공정 라이브러리를 사용하였다. 10MHz 클록 주파수에서 합성하였고, 기존 하드웨어도 공정한 비교를 위해 동일한 조건에서 합성하여 결과를 비교하였다(redesign). 제안하는 최대 풀링 하드웨어는 9,265um^2 의 면적을 가져, 기존 하드웨어의 48,504um^2 와 비교하여 80.9%의 면적을 줄일 수 있었다.

	기존[3]	제안
공정 [nm]	130	130
동작 주파수 [MHz]	10	10
면적 [um^2]	48,504	9,265

표 2. 최대 풀링 하드웨어 구현 결과 비교

Ⅲ. 결론

본 논문에서는 다양한 필터 크기를 갖도록 재구성 가능한 최대 풀링 하드웨어를 저면적으로 구현하고자 하였다. 제안하는 최대 풀링 하드웨어는 기존의 병렬 처리 구조가 아닌 폴딩을 통한 직렬 처리 방식을 통해 하드웨어 면적을 약 80.9% 감소시킬 수 있었다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부의 재원으로 정보통신기획평가원의 정보통신방송기술개발지원사업의 연구결과로 수행되었습니다 (IITP- 2020-0-01077, IoT 다중 인터페이스 기반의 데이터센싱, 엣지컴퓨팅 분석 및 데이터공유 지능형 반도체 기술 개발)

참 고 문 헌

[1] H. Gholamalinezhad, and H. Khosravi, “Pooling methods in deep neural networks, a review,” arXiv, no. September, 2020.

[2] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le., “Learning transferable architectures for scalable image recognition,” arXiv: 1707.07012, 2017. 2, 6

[3] L. Du et al., “A reconfigurable streaming deep convolutional neural network accelerator for Internet of Things,” IEEE Trans. Circuits Syst.I, Reg. Papers, vol. 65, no. 1, pp. 198- 208, Jan. 2018.

심층 신경망 가속을 위한 유연한 곱셈 누산 연산기 기반 가변 필터 처리 시스톨릭 어레이 하드웨어

임현서, 신동엽, 권대길, 임용석
한국전자기술연구원 스마트네트워크연구센터

hihi981025@keti.re.kr, dongyeob.shin@keti.re.kr, tgkwon@keti.re.kr, busytom@keti.re.kr

Reconfigurable Multiply Accumulate Unit Based Variable Filter Processing Systolic Array Hardware for Deep Neural Network Acceleration

Hyunseo Lim, Dongyeob Shin, Tai-Gil Kwon, Yong-Seok Lim

Smart Network Research Center
Korea Electronics Technology Institute

요 약

본 논문은 비트-정밀도와 곱셈-누산(Multiply-Accumulate) 연산의 출력 개수를 조절할 수 있는 유연한 곱셈-누산기 및 이를 적용하여 한 번에 처리하는 심층 신경망 필터 개수를 가변 가능한 시스톨릭 어레이(Systolic Array) 하드웨어 구조를 제안한다. 제안하는 구조는 빠르게 많은 필터 처리를 요구하는 레이어에서 기존 구조보다 효율적이다.

Ⅰ. 서 론

심층 신경망(Deep Neural Networks)은 수백만 개 이상의 가중치(Weight) 파라미터 수와 그에 대응되는 입력 데이터와 가중치 값 사이의 반복적인 곱셈-누산(Multiply-Accumulate) 연산을 통해 컴퓨터 비전, 자연어 처리 등 다양한 분야에서 높은 정확도를 보여주고 있다. 하지만 수많은 곱셈-누산 연산으로 인해 하드웨어 구현 시 많은 전력 및 에너지를 소모하는 문제점이 있다. 따라서 곱셈-누산 연산의 하드웨어 에너지 소모를 줄이기 위해, 곱셈-누산 연산의 피연산자(Operand)인 입력 데이터와 가중치의 비트-정밀도를 감소시키되 심층 신경망 정확도는 유지하는 양자화(Quantization) 기법이 많이 사용되고 있다[1]. 예를 들어, 양자화를 통해 입력 데이터와 가중치의 비트-정밀도를 절반씩 줄이면 하나의 단위 곱셈-누산 연산에 소모하는 하드웨어의 전력/에너지/면적 소모를 1/4 로 줄일 수 있다. 정확도를 유지하기 위해서, 양자화를 통한 비트-정밀도는 심층신경망 모델별로 혹은 같은 모델 내에서도 레이어(Layer)별로 최적의 값이 달라질 수 있으며, 최근 관련 연구들이 급증하고 있다[2]. 따라서, 심층 신경망 모델 및 레이어에 따라 비트-정밀도를 가변할 수 있는 가변 비트-정밀도 곱셈-누산기[3] 및 이를 적용한 심층신경망 연산 하드웨어 설계가 중요해지고 있다.

본 논문에서는 비트-정밀도와 출력의 개수를 조절 가능한 곱셈-누산기 및 이를 적용한 심층 신경망 연산용 시스톨릭 어레이(Systolic Array) 하드웨어를 제안한다. 제안하는 하드웨어는 비트-정밀도 가변 시 모드(Mode) 변경을 통해 곱셈-누산 연산의 출력 개수를 가변하여, 처리하는 심층신경망 필터 개수를 조절할 수 있다.

Ⅱ. 본론

2.1. 기존 가변 비트-정밀도 곱셈-누산기 구조

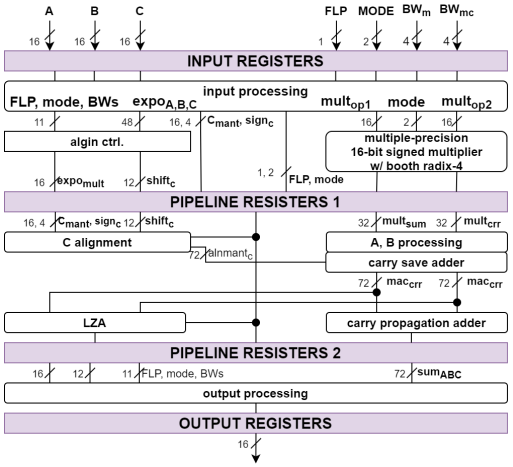


그림 1. 기존 가변 비트-정밀도 곱셈-누산기 구조

그림 1 은 기존의 가변 비트-정밀도 곱셈-누산기 구조이다. A, B, C 를 입력 데이터로 하여 A*B+C 의 곱셈-누산 연산을 수행한다. 이때, A 와 B 의 비트-정밀도에 따라 여러 개의 곱셈을 동시에 처리하는 것이 가능하다. 예를 들어 A, B 가 16bit 일 때 1 개의 곱셈

연산을 처리할 수 있지만, A, B 가 8bit 일 때는 2 개의 곱셈 연산을 동시에 처리하여 A0*B0 + A1*B1 + C 의 연산을 수행할 수 있다. 그림 1 에서 볼 수 있듯이, 기존의 가변 비트-정밀도 곱셈-누산기는 항상 모든 곱셈기 결과를 전부 더해서 하나의 결괏값을 출력한다.

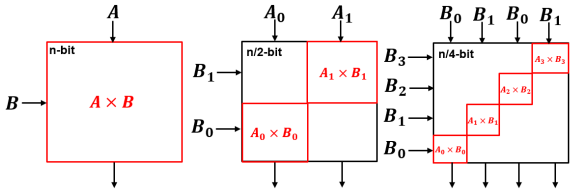


그림 2. 비트-정밀도 변화에 따른 곱셈기 동작

그림 2는 비트-정밀도 변화에 따른 곱셈기의 동작 변화를 나타낸다. 입력 A, B가 n-bit, n/2-bit, n/4-bit인 경우 1, 2, 4개의 곱셈 연산이 동시에 가능하다. 하지만 기존 가변 비트-정밀도 곱셈-누산기는 입력으로 받는 A, B는 가변이 가능하지만 누산 연산의 출력은 고정되어 있다. 대부분의 심층신경망 연산 하드웨어는 곱셈-누산기 여러 개를 시스톨릭 어레이와 같은 배열로 구성하여 처리한다. 곱셈-누산기가 여러 개의 곱셈-누산 연산을 처리하더라도 전부 누산하여 하나의 곱셈-누산 연산 결과만 출력한다면, 전체 연산 측면에서 좀 더 다양한 배열 처리를 할 수 있는 가능성을 제한하게 된다.

2.2. 제안하는 유연한 곱셈-누산기 구조 및 적용

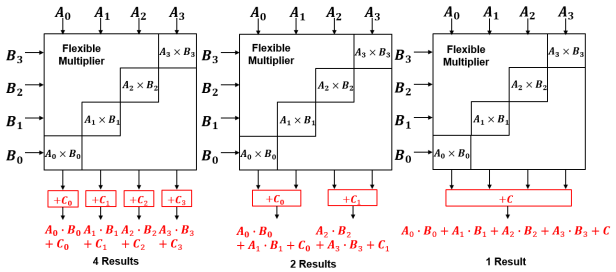


그림 3. 곱셈- 누산 연산 출력 수의 가변이 가능한 누산기 구조

그림 3은 n/4-bit의 곱셈 연산 시, 출력의 가변 형태를 보여준다. 기존에 고정된 개수 만을 출력했던 구조와 달리, 다양한 형태의 누산 결과를 출력할 수 있다. 이러한 구조는 곱셈 결과 누산에 사용되는 누산기의 구조에 간단한 MUX 만을 추가하여 사용 가능하므로, 기존과 하드웨어 면적의 차이가 거의 없다.

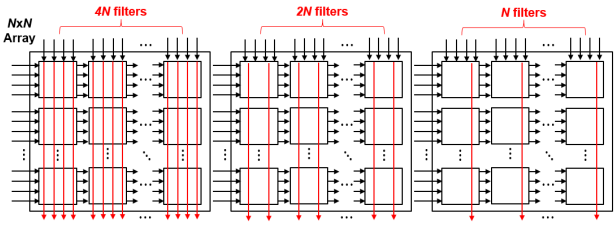


그림 4. 연산 출력의 가변에 따른 필터 처리 수 변화

그림 4 는 본 논문에서 제안한 가변 비트-정밀도 곱셈-누산기를 기본 연산 단위로 하는 시스톨릭 어레이 형태의 하드웨어 가속기이다. 그림 4 와 같이 NxN 배열의 가속기에서 제안하는 곱셈-누산기 1 개가 2 개 혹은 4 개의 곱셈-누산 연산을 처리하는 경우 2N, 4N 개의 필터 연산도 동시에 처리가 가능해진다. 심층 신경망의 레이어 특성에 따라 동시에 처리하는 필터 수가 많을수록 좋은 경우에는 고정된 개수의 필터 연산만을 수행하는 구조보다 더 많은 필터 연산을 동시에 처리할 수 있으므로 효율적이다.

2.3. 구현 결과

표 1 은 본 논문에서 제안하는 곱셈-누산기(MAC)와 제안하는 MAC 에 기반한 8x8 시스톨릭 어레이의 하드웨어 구현 결과이다. 제안하는 하드웨어의 구현을 위해서 시놉시스(Synopsys) 디자인 컴파일러(Design Compiler) 툴과 130nm CMOS 공정 라이브러리를 사용하여 10MHz 동작 주파수에서 합성을 진행하였다. 제안하는 MAC 은 148,305 um^2 의 면적을 가지며, 제안하는 8x8 시스톨릭 어레이는 9,492,205 um^2 의 면적을 갖는다.

	1 MAC	8x8 MAC Array
공정 [nm]	130	130
동작 주파수 [MHz]	10	10
면적 [um^2]	148,305	9,492,205

표 1. 제안한 하드웨어 구조의 합성 결과

Ⅲ. 결론

본 논문에서는 비트-정밀도의 가변과 곱셈-누산 출력의 개수 가변이 가능한 새로운 곱셈-누산기 하드웨어 구조를 제안하였다. 비트-정밀도의 가변을 통해 여러 곱셈 연산을 동시에 처리할 수 있고, 출력의 개수를 조절하여 빠르게 많은 필터 처리를 요구하는 레이어에서 기존 고정된 출력 개수를 처리하는 구조보다 효율적이다. 기존 대비 하드웨어 면적 오버헤드는 5% 미만이며, 130nm 공정, 10MHz 동작 주파수에서 8x8 시스톨릭 어레이로 구현 시 면적은 9,492,205 um^2 이다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부의 재원으로 정보통신기획평가원의 정보통신방송기술개발지원사업의 연구결과로 수행되었습니다 (IITP- 2020-0-01077, IoT 다중 인터페이스 기반의 데이터센싱, 엣지컴퓨팅 분석 및 데이터공유 지능형 반도체 기술 개발)

참 고 문 헌

[1] B. Jacob, S. Kligys, B. Chen, M. Zhu, M. Tang, A. Howard, H. Adam, and D. Kalenichenko, “Quantization and training of neural networks for efficient integer-arithmetic-only inference,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jul. 2017, pp. 2704-2713.

[2] K. Wang, Z. Liu, Y. Lin, J. Lin, and S. Han, “HAQ: hardware-aware automated quantization,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2019, pp. 8612-8620.

[3] H. Zhang, D. Chen, and S.-B. Ko, ‘‘New flexible multiple-precision multiply-accumulate unit for deep neural network training and inference,’’ IEEE Trans. Comput., vol. 69, no. 1, pp. 26- 38, Jan. 2020.

반도체 식각 공정에서 커널 밀도 추정 기반의 식각 종점 탐지

최준규, 김보배, 심영준, 임성빈*

숭실대학교

cjg@soongsil.ac.kr, kbb@nate.com, simyoungjun@gmail.com, *sbi@ssu.ac.kr

Detection of Etching End Points Based on Kernel Density Estimation in Semiconductor Etching Process

Jungyu Choi, Bobae Kim, Youngjun Sim, Sungbin Im*

Soongsil University

요 약

빠른 속도로 향상되는 컴퓨터의 계산 능력으로 인해 기계학습은 빠르게 발전하여 광범위한 분야에서 응용하여 사용되고 있다. 현재는 반도체 제조 공정에서도 기계학습에 많은 이목이 집중되고 있다. 따라서 본 논문은 반도체 제조 공정에서 식각 종점 탐지을 위한 Kernel Density Estimation 기법 적용을 제안한다. 무작위로 선택한 Data로 100번의 실험을 반복한 결과, 커널 밀도 추정 기법은 train data를 사용해 확률 함수를 구하고 test data를 실시간으로 식각 구간을 탐지할 수 있어 반도체 식각 공정 적용에 적합한다고 판단된다.

Ⅰ. 서론

빠른 속도로 향상되는 컴퓨터의 계산 능력으로 인해 기계학습은 빠르게 발전하여 광범위한 분야에서 응용하여 사용되고 있다. 그 중에서도 기계학습을 이용한 시계열 이상 탐지 기술은 금융 및 항공우주를 포함하여 IT, 보안 및 의료에 이르기까지 다양한 산업에서 직면한 치명적이고 예기치 못한 상황에 대한 정보를 제공한다 [1]. 현재는 반도체 제조 공정에서도 마찬가지로 기계학습에 많은 이목이 집중되고 있다.

Ⅱ. 본론

커널 밀도 추정은 수많은 추정법 중 커널 함수를 이용한 밀도 추정 방식이다. 밀도 추정이란 어떤 변수가 가질 수 있는 값 혹은 그 값을 가질 수 있는 가능성의 정도를 추정하는 것을 의미한다. 기계학습, 확률, 통계 등에서 말하는 밀도는 확률 밀도(Probability Density)를 의미한다. 즉, 어떤 변수(variable) x의 밀도를 추정하는 것은 x의 확률 밀도 함수(PDF, Probability Density Function)를 추정하는 것과 같은 의미이다. 어떤 변수 x의 확률 밀도 함수 f(x)는 식 (1)과 같이 표현할 수 있다.

$$P(a \le x \le b) = \int_a^b f(x)dx \qquad (1)$$

f(x)가 그림 1과 같다고 하면, f(a), f(b)는 각각 x=a, x=b에서의 확률 밀도 즉, 변수 x가 a라는 값을 가질 수 있는 상대적인 가능성(Relative Likelihood)을 의미한다.

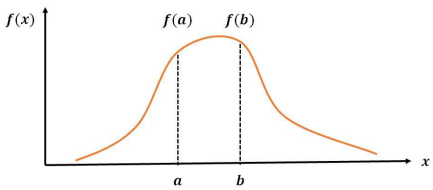


그림 1 확률 밀도 함수 예

밀도 추정은 크게 Parametric 방식과 Non-Parametric 방식으로 나뉜다. Parametric 밀도 추정은 사전에 확률밀도함수에 대한 모델을 정해놓고 데이터들로부터 모델의 파라미터만 추정하는 방식이고, Non-Parametric 밀도 추정은 관측된 데이터로부터 히스토그램을 구하여 확률밀도함수로 사용하는 방식이다. Parametric 밀도 추정은 관측된 데이터가 정규분포를 따른다고 가정하면 평균과 분산만 구하면 되기 때문에 밀도 추정 문제가 간단하게 된다는 문제가 발생하고, Non-Parametric 밀도 추정은 히스토그램 방법을 사용할 경우 그림 2와 같이 bin의 경계에서 불연속성이 나타나고, bin의 크기와 시작 위치에 따라서 히스토그램이 달라지는 문제점이 있다.

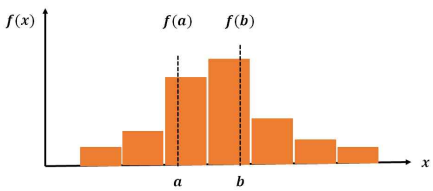


그림 2 히스토그램 예

커널 밀도 추정은 평탄한 곡선이므로 확률밀도함수의 세부사항을 더

잘 나타내며 경우에 따라서는 비일관성을 시사하기도 한다. 그리고 모든 표본점의 위치를 사용하기 때문에 표본에 포함된 정보를 더 잘 드러내는 등 히스토그램에 비해 장점이 훨씬 크다 [2]. 그림 3은 데이터 분포에 따른 커널 밀도 추정을 통해 구한 확률 함수의 예이다. 커널 밀도 추정은 히스토그램의 단점을 보완할 수 있기에 본 연구팀은 커널 밀도 추정 기반의 반도체 플라즈마 식각 공정에서 식각 종점을 탐지하는 알고리즘을 제안한다.

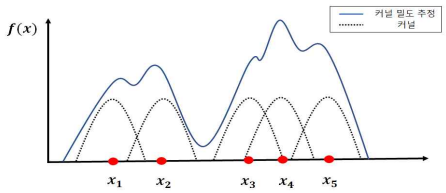


그림 3 커널 밀도 추정을 통해 구한 확률 함수 $f(x)$ 예

Ⅲ. 실험

본 논문에서는 가우시안 커널 밀도 추정을 통해 식각 종점을 탐지하였다. 커널 밀도 추정 모델에 사용된 데이터는 실제 반도체 플라즈마 식각 공정에서 OES를 사용하여 시간에 따른 파장별 Intensity 값을 측정한 것이다. 여기에 사용된 데이터는 총 1,911장의 Wafer에서 측정하였다. 수집한 데이터에는 총 2,048개의 파장이 있다. OES 장비를 통해 수집한 데이터는 방대한 양의 정보를 가지고 있기 때문에 실제 실험에서 사용하는 데이터는 식각 종점 탐지에 영향이 크다고 판단되는 파장 387nm, 520nm, 700nm, 777nm만 추출하여 사용한다. 4개의 파장 Intensity 값을 커널 밀도 추정에 적용하여 비식각 구간과 식각 구간에 대한 확률 함수를 구하고 이를 통해 식각 종점을 탐지한다. 그림 4는 커널 밀도 추정 기반의 식각 종점 탐지 흐름도이다.

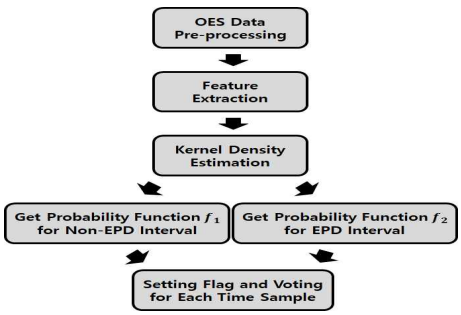


그림 4 커널 밀도 추정 기반의 식각 종점 탐지 흐름도

우리는 가우시안 커널 밀도 추정 실험을 통해 그림 3과 같은 결과를 얻을 수 있었다. 그림 5는 Test Wafer 711장 중에서 1장의 가우시안 커널 밀도 추정을 통해 얻은 확률 함수이다. 여기서 f_1은 비식각 구간에 대한 확률 함수이고 f_2는 식각 구간에 대한 확률 함수를 의미한다.

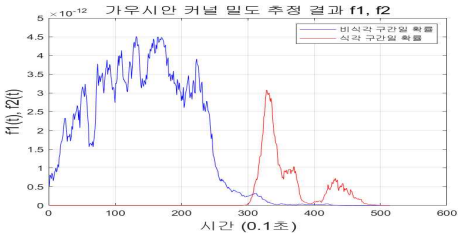


그림 5 가우시안 커널 밀도 추정 결과 f_1, f_2 샘플

가우시안 커널 밀도 추정을 통해 구한 확률 함수 f_1, f_2를 비교하여 f_1이 크면 0, f_2가 크면 1로 flag를 설정하였고 이 flag의 반복을 확인하여 최적의 식각 종점 탐지를 계산하였다. 그 결과 flag가 12번 반복되었을 때 98.49 % 라는 가장 높은 정확도를 보였다. Flag 반복 수를 확인한 이유는 f_2가 f_1보다 커졌을 때는 식각 구간이 시작되었음을 의미하는 것이지 그 순간이 식각 종점임을 의미하지 않기 때문이다. 그리고 데이터의 잡음으로 인해 flag 값은 0과 1이 반복되기 때문에 가장 먼저 flag가 1인 시점은 식각 종점과 오차가 클 수 있다. 그림 6은 flag 반복 수에 따른 가우시안 커널 밀도 추정 모델의 정확도 그래프이다.

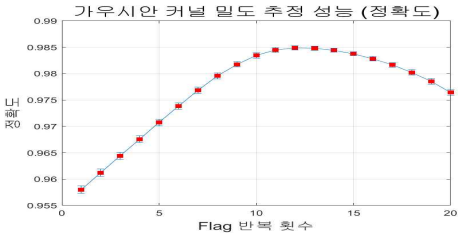


그림 6 가우시안 커널 밀도 추정 성능 (정확도)

Ⅳ. 결론

본 논문에서는 반도체 제조 공정에서 플라즈마 식각 종점 탐지를 위해 커널 밀도 추정 기법을 사용했다. 랜덤한 711개의 Wafer 데이터에 대해 100번의 정확도 평가를 수행했고 98.49 %라는 결과를 얻었다. 커널 밀도 추정은 train data를 사용하여 확률 함수를 구하여 test data를 실시간으로 식각 구간을 탐지할 수 있어 반도체 식각 공정 적용에 적합하다고 판단한다.

ACKNOWLEDGMENT

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2020-0-02102, (2단계 : 기술개발) 반도체 식각장비 실시간 공정진단을 위한 AI 기반 다채널 광 스펙트럼 분석플랫폼 개발).

참 고 문 헌

[1] Chandola, Varun, Arindam Banerjee, and Vipin Kumar. "Anomaly detection: A survey." ACM computing surveys (CSUR) 41.3 (2009): 1-58.

[2] Węglarczyk, Stanisław. "Kernel density estimation and its application." ITM Web of Conferences. Vol. 23. EDP Sciences, 2018.

플라즈마 식각 공정에서 CNN과 KDE의 반도체 식각 종점 탐지 비교

김보배, 최준규, 심영준, 임성빈
숭실대학교
kbb@nate.com, cjg@soongsil.ac.kr, simyoungjun@gmail.com, sbi@ssu.ac.kr

Comparison of Semiconductor Etch Endpoint Detection Performance of CNN and KDE in Plasma Etching Process.

BoBae Kim, JunGyu Choi, YoungJun Sim, Sungbin Im
Soongsil Univ.

요 약

미세화(Scailng)되고 있는 반도체 플라즈마 식각공정에서 수율 향상을 위해 최적의 식각 종점을 찾는 것이 중요하다. 따라서 본 논문에서는 최적의 식각 종점을 탐지하기 위해 반도체 플라즈마 식각 공정에서 생성되는 데이터를 활용하여 두가지 방법을 적용하여 비교했다. 첫 번째 방법은 인공지능이며 두 번째는 커널 밀도 추정이다. 인공지능 모델은 Convolutional Neural Network (CNN)을 적용했고 확률 분포 방법은 Kernel Density Estimate (KDE)를 적용하였다. 연구에 사용된 데이터는 플라즈마 식각 과정에서 얻은 Optical Emission Spectroscopy (OES) 데이터로 웨이퍼의 왼쪽, 중앙, 오른쪽, 전체 구간에서 측정한 4채널 데이터이다. 실험결과 학습 시간 및 정확도 측면에서 CNN의 성능이 더 우수했다.

Ⅰ. 서 론

최근 산업의 고도화 및 컴퓨팅 성능의 향상으로 반도체 공정이 미세화되고 있다. 이에 따라 공정과정상의 편차와 성능에 대한 불확실성이 증가하고 있다.[1] 반도체 공정에서 사용되는 플라즈마 식각의 미세한 제어는 어려우며 수율 향상을 위한 최적의 식각 종점을 예측하는 것은 중요한 일이다.[2] 따라서 본 논문에서는 플라즈마 식각 공정에서 나오는 다채널 OES 데이터를 활용하여 인공지능과 커널 밀도 추정 방법을 적용해 EndPoint Detection (EPD)의 성능을 비교한다.

본 논문의 구성은 다음과 같다. 본론에서는 본 논문에서 사용한 두 가지 방법과 실험에 사용된 데이터에 대해 설명한 후 결과를 보여준 후 결론에서는 결론과 향후 연구 방향에 대해 서술한다.

Ⅱ. 본론

연구에 사용된 CNN 모델은 딥러닝에서 가장 많이 사용되는 알고리즘으로 데이터에서 특징을 직접 학습하고 패턴을 활용하여 데이터를 분류하는 알고리즘이다.[3] 본 논문에서 사용한 CNN 모델의 구조는 다음과 같다.

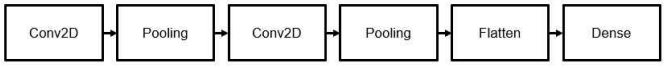


그림 1 CNN 모델 구조

KDE는 커널 함수를 활용하여 변수에 대한 Probability Density Function (pdf)을 추정하는 방법이다.[4] KDE로 얻은 확률밀도함수는 데이터의 히스토그램을 스무딩한 결과로 볼수있으며 bandwidth 값에 따라 그 정도가 달라진다.

본 논문에 사용된 데이터는 플라즈마 식각 과정에서 추출한 4채널 OES 데이터로 웨이퍼의 왼쪽, 중앙, 오른쪽, 전체 구간에서 측정되었으며 아래 그림 2에 측정 지점을 그림으로 나타냈다.

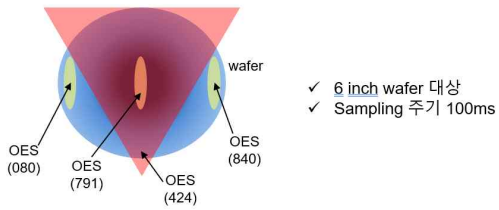


그림 2 OES 데이터 측정 지점

총 36장의 웨이퍼에 대한 4채널 데이터인 144 (36 x 4)개의 데이터로 학습데이터와 검증데이터의 비율은 6:4로 설정하였다. 원본 데이터는 식각 시간에 따른 2,048개의 파장으로 sampling 주기는 100ms 이다. 그 중 데이터에서 특징이 잘 나타나는 파장 4가지 (240nm, 450nm, 703nm, 777nm)를 선택하여 학습했으며 각 파장의 추세는 아래 그림 3과 같다.

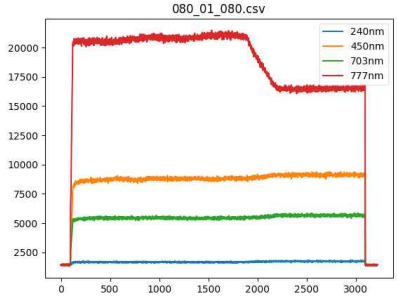


그림 3 학습에 사용된 파장의 추세 그래프

그림 3에서 777nm 파장 외의 다른 파장들은 그래프의 변화가 적어보이지만 777nm 파장의 Intensity 세기가 크기 때문이다.

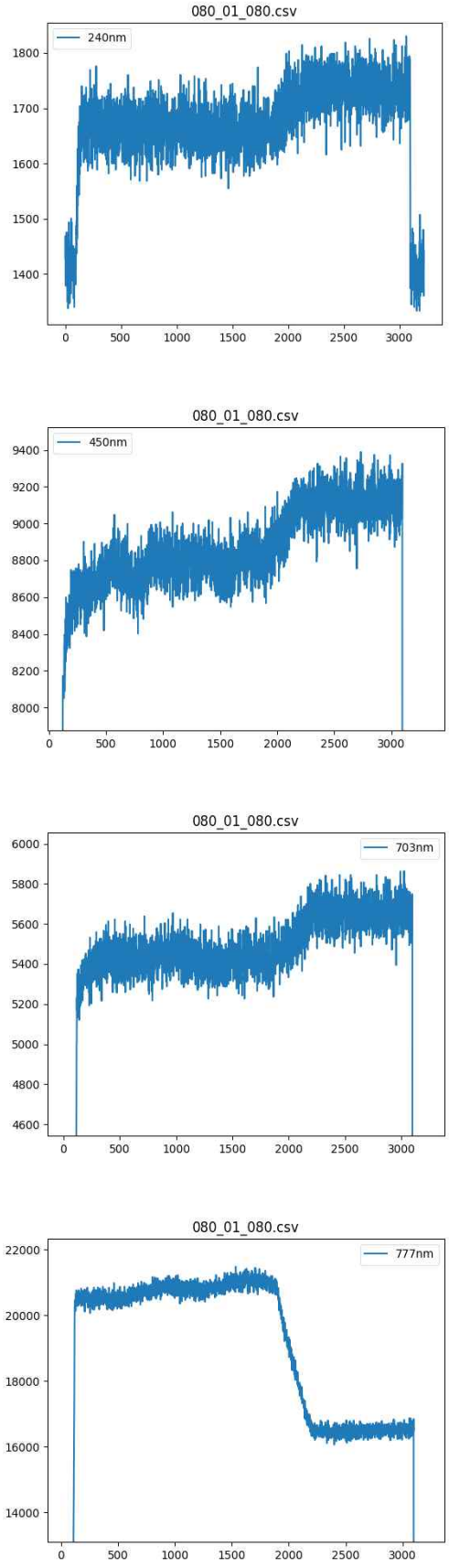


그림 4 각 파장의 개별 추세 그래프

위 그림 4는 파장들을 개별로 그린 그래프이다. 기울기가 급격하게 변화하는 구간을 EPD라고 설정하였다.

앞서 서술한 데이터를 사용하여 진행된 실험 결과를 아래 표 1에 정리하였다.

표 1 모델의 정확도 및 학습 시간

모델	정확도(%)	학습 시간(s)
CNN	95.6616	7,542
KDE	90.4718	449,346

Ⅲ. 결론

본 논문에서는 플라즈마 식각에서 방출되는 4채널 OES 데이터를 두 가지 방법의 알고리즘에 적용하여 비교했다. 그 결과 학습 시간에서는 압도적으로 CNN이 빨랐고 정확도 측면에서도 마찬가지로 CNN의 성능이 더 좋았다. 적은 데이터로도 우수한 성능을 보였으므로 데이터의 양이 더 많아진다면 정확도는 더욱 올라갈 것으로 판단되며 더 많은 데이터를 확보하여 추가 연구를 진행해야 할 필요성이 있다.

ACKNOWLEDGMENT

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2020-0-02102, (2단계 : 기술개발) 반도체 식각장비 실시간 공정진단을 위한 AI기반 4채널 광 스펙트럼 분석플랫폼 개발)

참 고 문 헌

[1] Kim, H., Lim, D. E., & Lee, S. “Deep learning-based dynamic scheduling for semiconductor manufacturing with high uncertainty of automated material handling system capability”, IEEE Transactions on Semiconductor Manufacturing, 33(1), 13-22, 2020

[2] Shane A. Lynn, John Ringwood, Senior Member, IEEE, and Niall MacGearailt, Global and Local Virtual Metrology Models for a Plasma Etch Process, IEEE TRANSACTIONS ON SEMICONDUCTOR MANUFACTURING, VOL. 25, NO. 1, FEBRUARY 2012.

[3] Albawi, S., Mohammed, T. A., & Al-Zawi, S. “Understanding of a convolutional neural network”, In 2017 International Conference on Engineering and Technology (ICET), pp. 1-6, Ieee, 2017.

[4] Lemenkova, P. “Processing oceanographic data by Python libraries NumPy, SciPy and Pandas”, Aquatic Research, 2(2), 73-91, 2019.

제2회 한국 인공지능 학술대회

The 2nd Korea Artificial Intelligence Conference

세션 G

Partial Isolation Type Saddle-FinFET (Pi-FinFET)s 에서 Single Charge Trap 에 의한 누설전류 분석

박진효,김건,김동영,김수연,박제원

이명진*

전남대학교, *전남대학교

206070@jnu.ac.kr, rjslchd@naver.com, dy3351@gmail.com, bollic4504@gmail.com,
you05281@naver.com, *mjlee@jnu.ac.kr

A Study on the S-TAT current in Partial Isolation Type Saddle-FinFET (Pi-FinFET)s

Jin Hyo Park, Geon Kim, Dong Young Kim, Su Yeon Kim, Je Won Park
Myoung Jin Lee*
Chonnam Univ.,*Chonnam Univ.
Corresponding author: Myoung Jin Lee

요 약

본 논문은 single charge trap 에 의한 누설전류를 분석하는 방법으로 기존의 saddle type FinFET (SFinFET)과 SFinFET 의 드레인(drain) 영역 하단에 절연체(insulator)를 고립시킨 partial isolation type saddle FinEFT(Pi-FinFET) 구조를 비교하였다. 3D TCAD 를 이용하여 드레인 영역 하단에 single charge trap 의 위치를 조정해가며 single charge trap 이 누설전류 메커니즘에 미치는 영향을 분석하였다. 그 결과 Pi-FinFET 이 single charge trap 의 위치와 상관없이 SFinFET 보다 개선된 성능을 나타내는 것을 확인했다. 본 논문에서는 single charge trap 에 의한 누설전류를 예측하는 방법을 통해 DRAM refresh 특성 향상을 위한 소자 디자인 가이드라인을 제시한다.

Ⅰ. 서 론

반도체 동작에 있어 결함(defect)은 소자 특성에 치명적인 영향을 미친다. Defect 는 외부 환경에 의해 생길 수도 있고 과도한 stress 에 의해 발생하기도 한다. 이는 문턱전압과 transconductance 그리고 carrier mobility 에 영향을 주며 Trap-Assisted Tunneling (TAT)에 의해 주된 leakage path 로 작용될 수 있다 [1]. 이러한 문제는 반도체 소자의 신뢰성 및 성능과 직접적인 연관이 있다 [2]. 그렇기 때문에 소자 특성을 분석하는 데 있어 defect 에 대한 정확한 예측이 필요하다. 그러기 위해서 본 논문은 defect 가 DRAM cell transistor 의 누설전류에 미치는 영향에 대해서 분석했다. DRAM 동작에서 defect 의 정확한 위치를 파악하는 것이 불가능하기 때문에 누설전류에 영향을 미치는 드레인 하단의 다양한 위치에서 single charge trap 이 누설전류에 미치는 영향에 대해 분석했고 그 평균값을 이용해 소자를 비교하는 방법을 제시한다 [3]. 이 방법에 따르면 Pi-RFinFET 은 RFinFET 보다 2 배 작은 누설전류 변동 값을 보였다.

Ⅱ. 본론

본 논문에서는 modified saddle type FinFET (RFinFET) 과 RFinFET 의 드레인 하단 insulator 를 삽입한 Pi-RFinFET 구조를 사용하였고[4], 이 소자들은 3D 소자 시뮬레이터(sentaurus TCAD)로 제작하였다 [5]. 그림 1-(a) 는 Pi-RFinFET 의 3 차원 구조를 나타낸다. RFinFET 은 modified saddle type FinFET 으로 side gate 가 channel 영역에만 존재하는 구조로, 게이트와 드레인의 overlap region 을 줄여 GIDL 을 개선한 소자이다. Pi-FinFET 의 buried insulator 는 등방성 식각을 통해 형성할 수 있으며, 에칭 시간을 조절하여 측면 깊이를 조절할 수 있다. 그림 1-(b) 는 RFinFET 의 단면도이며 초록색 점선으로 된 사각형은 Pi-RFinFET 에서의 buried insulator 의 영역을 나타낸다. 검은색 점은 본 논문에서 사용한 single charge trap 의 위치이다.

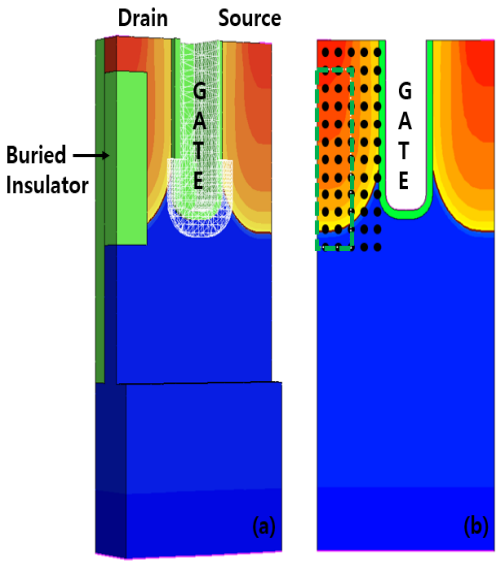


그림 1. (a) Pi-RFinFET 의 3D 구조와 (b) 단면도 및 single charge trap 의 위치.

그림 2-(a) 는 Pi-RFinFET 소자에서 single charge trap 위치에 따른 누설전류 그래프이다. 검은색 그래프는

single charge trap 을 고려하지 않은 결과이고 검은색 선위에 검은색 원이 존재하는 그래프는 파란색 그래프값의 평균값이다. 그림 2-(b) 는 single charge trap 에 따른 그래프의 평균값과 single charge trap 을 고려하지 않은 그래프를 나타낸다. RFinFET 의 경우 single charge trap 을 고려하였을 경우 누설전류가 20% 증가하였고, Pi-RFinFET 의 경우 25%가 증가하였다. 이를 통해 정확한 누설전류 예측을 위해 single charge trap 에 대한 고려가 필요함을 확인할 수 있다. 그림 2-(b) 를 보면 Pi-RFinFET 이 RFinFET 보다 2 배 작은 변동 폭을 가지고 있음을 확인할 수 있고, 이를 통해 Pi-RFinFET 이 RFinFET 보다 single charge trap 에 영향을 덜 받는다는 것을 확인할 수 있다.

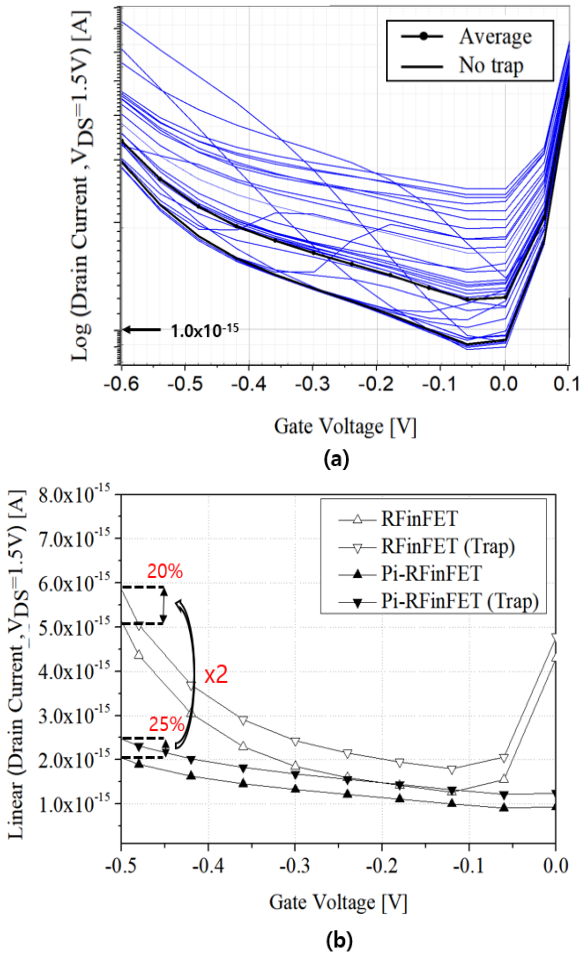


그림 2. (a) Pi-RFinFET 에서 single charge trap 에 따른 leakage current (b) Pi-RFinFET 과 RFinFET 에서 single charge trap 유무에 따른 그래프 비교. Single charge trap 은 acceptor like trap 에 E_t=E_i

그림 3 은 RFinFET 과 Pi-RFinFET 에서의 electric field 분포를 나타낸다. Pi-RFinFET 의 경우 절연체의 영향으로 강한 electric field 가 상단에서만 형성되는 것을 확인할 수 있다. 이는 절연체 내부에서도 potential drop 이 발생하게 되고 드레인 하단 영역에서 발생하는 potential modulation 현상이 게이트와 드레인 사이의 전계가 감소하는 결과로 이어지기 때문이다. 반면 RFinFET 의 경우 강한 전계가 드레인 하단에 더 넓게 분포하고 있다. 전계의 감소는 누설전류의 메커니즘인 tunneling 현상이 발생할 확률이 줄어듦을 의미한다. 따라서 같은 위치에 single charge trap 이 존재하더라도 전계의 분포가 한정적이고 작은 전계를 가지고 있는 Pi-FinFET 이 RFinFET 에 비해 trap 의 영향을 덜 받는다는 것을 확인할 수 있다.

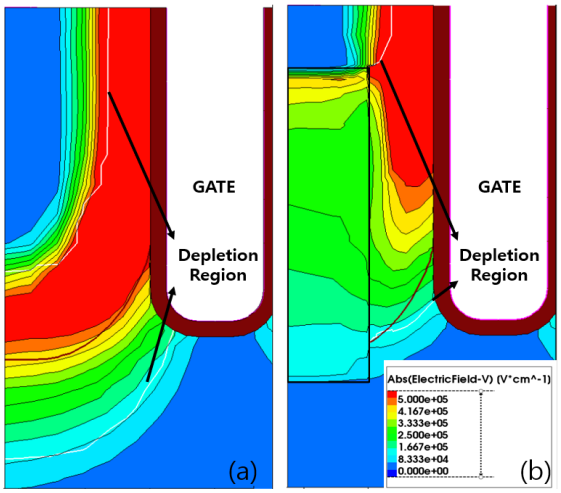


그림 3. (a) RFinFET 과 (b) Pi-RFinFET 에서의 electric field

Ⅲ. 결론

본 논문에서는 RFinFET 과 Pi-RFinFET 에서 single charge trap 에 의한 누설전류를 분석하는 방법을 제시한다. Single charge trap 은 특정 위치에서 소자에 치명적으로 작용하며, 이 분석 방법을 통해 그 위치를 확인하여 공정 시 그 위치에 대한 피드백을 통해 소자의 신뢰성을 높일 수 있다. 또한 single charge trap 에 의한 누설전류 변동의 평균값을 통해 소자 간 trap 에 대한 취약성을 비교할 수 있으며 이를 통해 DRAM refresh 특성 향상을 위한 소자 디자인 가이드라인을 제시할 수 있다.

ACKNOWLEDGMENT

This work was supported in part by the BK21 FOUR Program (Fostering Outstanding Universities for Research, 5199991714138) and in part by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT; grant number 2018RIA2B600821613). The EDA tool was supported by the IC Design Education Center (IDEC), Korea.

참 고 문 헌

[1] S.R. Stiffler." Oxidation-induced defect generation in advanced DRAM structures," IEEE Transactions on Electron Devices, vol.37, pp. 1253-1287, May.1990.

[2] K.N. Kim. "A New Investigation of Data Retention Time in Truly Nanoscaled DRAMs," IEEE Electron Device Letters, vol.30,pp. 846-848, Aug.2009

[3] J.H. Park. "S-TAT Leakage Current in Partial Isolation Type Saddle-FinFET(Pi-FinFET)s," IEEE Access, vol.9, pp. 111567-111575, Aug. 2021.

[4] Y.K. Kim. "Partial Isolation Type Saddle FinFET(Pi-FinFET) for sub-30nm DRAM Cell Transistors," Electronics, vol.8, pp. 8.1-8.6,Oct.2019.

[5] Synopsis Inc. TCAD Sentaurus Manual, Synopsis®, version D-2013.03; Synopsis Inc.: Mountain View, CA, USA, 2013

유도전동기에서 다변량 커널 밀도 추정 기반의 이상치 탐지

심영준, 최준규, 김보배, 임성빈*

숭실대학교

simyoungjun@gmail.com, cjg@soongsil.ac.kr, kbb@nate.com, *sbi@ssu.ac.kr

Outlier Detection Based on Multivariate Kernel Density Estimation in Induction Motors

Youngjun Sim, Jungyu Choi, Bobae Kim, Sungbin Im*

Soongsil University

요 약

4차 산업 혁명 시대에는 사물 인터넷 기술 보급과 함께 점차 더 다양한 단말들이 사물 인터넷에 연결되고 있고, 그에 따라 단말들이 생성하는 사물 데이터도 급증하는 추세이다. 스마트 기술, 데이터, 자동화, 상호 연결성, 인공지능 등의 기술은 4차 산업 혁명의 핵심 기술이며, 이러한 기술은 공장과 산업 조직의 운영 방식을 혁신적으로 변화시키고 있다. 그 중 예측 유지 보수 기술(Predictive Maintenance)은 현재 가장 눈에 띄는 IIoT(IIoT, Industrial Internet of Things) 적용 사례이자 스마트 팩토리의 확대 및 고도화를 위한 시작점이라 볼 수 있다. 이를 위한 장애 감지 및 프로세스 모니터링은 프로세스 제어의 핵심 기술 중 하나이다. 본 논문에서는 이상 탐지의 기법 중 하나로 유도 전동기의 이상 탐지를 위한 다변량 커널 밀도 추정(MKDE, Multivariate Kernel Density Estimation)을 제안한다. 3상 유도 전동기에서 획득한 전류 데이터를 이용하여 실험한 결과, 특정한 확률분포함수와 상관관계에 대한 정의 없이 오직 데이터만을 사용하여 결합 확률 밀도 함수를 추정하고, 라벨이 없는 데이터에 대해 이상치를 탐지할 수 있다는 점에서 기계 시설물의 이상치 탐지에 적합하다고 판단된다.

Ⅰ. 서론

4차 산업혁명시대의 스마트 기술, 데이터, 자동화, 상호 연결성, 인공지능 등의 기술은 공장과 산업 조직의 운영 방식을 혁신적으로 변화시키고 있다. 이상 탐지는 스마트 팩토리의 확대 및 고도화를 위한 핵심 기술 중 하나이며 제조업뿐만 기계 센서 모니터링, 휴먼 헬스 모니터링, 금융, 기후 변화, 통신 데이터 등 광범위한 분야에 응용되고 있다. 그 중 유도 전동기는 산업계의 중요 전력설비로서 각종 생산설비 및 유틸리티 설비의 구동을 위한 핵심역할을 수행하고 있으며, 현재 유도 전동기의 고장 진단 방법에 대하여 많은 연구가 이루어지고 있는 추세이다.

Ⅱ. 본론

푸리에 변환(FFT, Fast Fourier Transfrom)이란 시간에 대한 함수 신호를 주파수 성분으로 분해하는 작업이다. 고속 푸리에 변환(Fast Fourier Transfrom, FFT)는 이산적인 함수 신호에 대한 푸리에 변환과 그 역변환을 빠르게 수행하는 효율적인 알고리즘이다.

함수 신호를 푸리에 변환하여, 주파수 스펙트럼으로 주파수 성분을 분석하였을 때 기본파와 고조파(harmonics) 성분으로 구분할 수 있다. 고조파가 상대적으로 기본파에 비해 클수록 왜곡이 커지며 즉, 이는 순수한 정현파의 파형에서 멀어지게 된다는 의미이다. 전 고조파 왜율(THD, Total Harmonic Distortion)이란 고조파로 인한 파형의 왜곡 정도를 나타내는 지표이다. 고조파 성분이 많을수록 THD 값이 높아지게 되면 정현파로부터의 왜곡이 심하다는 것을 의미한다. THD 계수에 대한 일반 방정식은 다음과 같다.

$$THD(w) = \frac{\sqrt{\sum_{n>1} Y_n^2(w)}}{Y_1} \quad (1)$$

Y_1은 기본파의 크기이며 Y_n은 고조파의 크기이다.

밀도 추정이란 데이터로부터 변수가 가질 수 있는 모든 값의 확률, 즉 확률 밀도 함수(PDF, Probability Density Function)를 추정하는 행위이다. 확률 밀도 함수는 확률 변수의 분포를 나타낸 함수인데, 구간 [a, b]에서 확률 변수 X가 해당 구간에 있을 확률은 다음과 같이 표현이 가능하다..

$$P(a < X < b) = \int_a^b f(x)dx \quad (2)$$

이러한 밀도추정 방법은 다양한 종류가 존재하는데, 크게 Parametric estimation 과, Non-parametric estimation으로 나뉜다. Non-parametric estimation 은 순수 관측 데이터만으로 확률 밀도 함수를 추정하는 기법인데, 가장 간단한 형태는 히스토그램(Histogram)형태이다. 하지만 히스토그램을 이용한 밀도 추정은 불연속성이 나타나고, 고차원 데이터에는 메모리 등의 문제가 있다. 이를 보완한 대표적인 추정법에는 본 논문의 주제인 Kernel Density Estimation이 있다. KDE는 커널(Kernel) 함수를 이용한 밀도 추정방법이다. 커널 함수는 원점을 중심으로 대칭이며 적분 값이 1인 양의 함수로 정의할 수 있다.

KDE는 이들을 사용하여 주어진 데이터의 분포를 예측하는데 다음과

같은 수식을 사용한다 [1].

$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{x-x_i}{h}) \quad (3)$$

K는 커널 함수를 뜻하고, h는 커널함수의 Bandwidth 값으로 커널의 분산을 조절하는 파라미터이다. Bandwidth 값이 작을수록 커널이 뾰족한 형태를 띄고, 클수록 커널이 완만한 형태가 된다. 데이터의 확률 밀도 함수를 히스토그램으로 추정했을 때와 KDE를 이용했을 때의 그림 1과 같다. 초록색은 값이 작은 Bandwidth를 사용한 것이고 붉은색은 값이 큰 Bandwidth를 사용한 것이다.

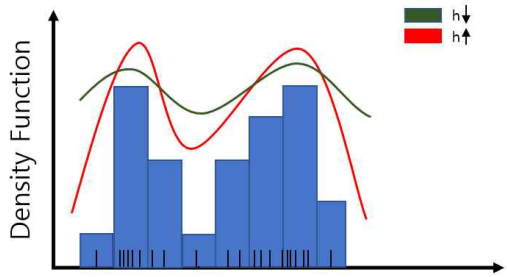


그림 1 히스토그램, 커널 함수를 이용한 확률 밀도 함수

다변량 커널 밀도 추정(MKDE)은 특정한 확률 분포 함수(CDF, Cumulative Distribution Function)와 상관 관계에 대한 정의 없이 오직 데이터만을 사용하여 결합 확률 밀도 함수 (JPDF, Joint Probability Density Function)를 추정하는 비 모수적 통계 모델링 방법이다. MKDE는 분포함수의 형태가 고정되지 않고 데이터에 따라서 유연하게 표현 할 수 있기 때문에 분포 함수의 표현의 자유도가 매우 높은 기법이다. MKDE는 각 데이터에 동일한 커널 함수 (kernel function)를 정의하고 모든 커널 함수를 최적의 대역폭에 따라서 합하면 최종적으로 결합 확률 밀도 함수가 추정되게 된다 [2]. Silverman's rule of thumb에 의하면 최적의 대역폭을 구하는 식은 다음과 같다 [3].

$$h_i = \sigma_i \left\{ \frac{4}{n(d+2)} \right\}^{\frac{1}{(d+4)}}, \quad i = 1,2,...,d, \quad (4)$$

σ_i는 i번째 feature에 대한 표준편차이며 d는 차원의 수이다.

Ⅲ. 실험

본 논문에서는 가우시안 함수를 커널 함수로 사용한 다변량 커널 밀도 추정을 이용해 유도 전동기의 이상을 탐지하였다. 이 연구에 사용된 데이터 세트는 AI Hub에서 제공한 "기계 진동 및 전류 데이터"로 기계 시설물의 상태 모니터링 및 고장 진단을 위해 활용하는 진동 및 전류 데이터이다 [4]. 총 11948개의 전류 신호 데이터 중 정상 데이터와 비 정상 데이터의 수는 각각 10000개, 1948개 이다. Train set와 Test set을 각각 5974개로 나누어 진행 하였다. FFT 변환 후, 신호를 분석하여 기본 주파수와 전고조파 왜율 (THD) 을 특징 추출에 이용하였다. R, T, S의 3상의 신호에 대한 3개의 120Hz 기본 주파수와, THD를 계산하여 총 6개의 feature를 생성하였다. Bandwidth는 Silverman's rule of thumb를 이용하여 계산하였다. 그림 2는 전류 신호 데이터를 FFT하여 나타낸 주파수 스펙트럼, 그림 3은 다변량 커널 밀도 추정 기반의 이상치 탐지 흐름도이다.

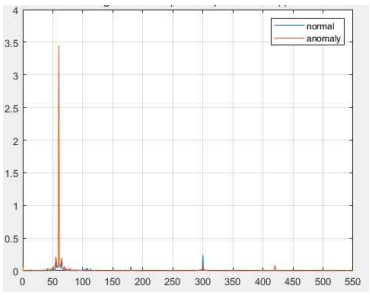


그림 2 전류 신호 FFT 분석

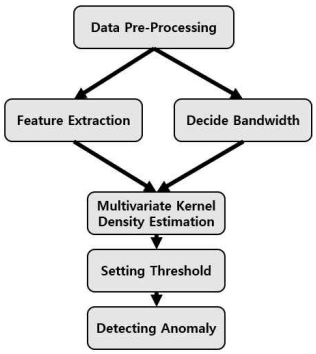


그림 3 다변량 커널 밀도 추정에 기반한 이상치 탐지 흐름도

다변량 커널 밀도 추정을 이용한 이상치 확률과, Silvermans's rule of thumb를 이용한 대역폭 값을 조정하면서 최적의 Threshold를 계산하였다. 그 결과 98.93% 정확도를 얻을 수 있다.

Ⅳ. 결론

본 논문에서는 유도 전동기의 이상치 탐지를 위해 다변량 커널 밀도 추정 기법을 사용했다. 총 5974개의 test 데이터에 대해 정확도 평가를 수행하였고, 98.9%의 결과를 얻었다. 다변량 커널 밀도 추정은 라벨이 없는 데이터로 모델의 훈련이 가능하다. 또한 상대적으로 적은 수의 이상 데이터를 사용하여 데이터 불균형이 있는 상황에서도 좋은 성능을 내는 것으로 보인다. 따라서 기계시설물에 대한 이상치 탐지에 적합하다고 판단된다.

ACKNOWLEDGMENT

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2021R1F1A1061403, 반도체 식각 공정에서 TadGAN을 이용한 Predictive Maintenance 기반의 EPD 추정)

참 고 문 헌

[1] Parzen, Emanuel. "On estimation of a probability density function and mode." The annals of mathematical statistics 33.3 (1962): 1065-1076.

[2] Y.-J. Kang, Y. Noh, and O.-K. Lim, "Development of MKDE-ebd for Estimation of Multivariate Probabilistic Distribution Functions," Journal of the Computational Structural Engineering Institute of Korea, vol. 32, no. 1, pp. 55 - 63, Feb. 2019.

[3] Scott, David W. Multivariate density estimation: theory, practice, and visualization. John Wiley & Sons, 2015.

[4] "AI 허브.", 기계시설물 고장 예지 센서. 2021년 6월 18일 수정, 2021년 8월 30일 접속, *https://aihub.or.kr.*

그래프 신경망을 이용한 전력 거래 파트너 매칭 서비스

강효은, 김용수, 홍윤영, 김호원*

*부산대학교

{hyoeun0915, yongsu, hyy0238, howonkim}@pusan.ac.kr

Partner Matching Service on Electric Power Transaction using Graph Neural Network

Hyo Eun Kang, Yong Su Kim, Yoon Young Hong, Ho Won Kim*

Pusan National University

요 약

최근 국내 전력 거래 시장은 K-RE100 캠페인 참여 확산 및 직접PPA 제도의 등장에 따라 신재생에너지 중심으로 조성되고 있다. 본 논문은 P2P 전력거래 플랫폼에서 전력 구매 희망자와 가까운 지역의 적합한 분산자원 공급자와 매칭시켜주는 그래프 신경망 기반 지역 맞춤형 전력 거래 서비스를 제안한다. 전력 구매 요청 발생 시 구매 희망자와 인근 공급자들의 정보를 조합하여 전력 최적의 거래 파트너 후보군을 추천한다. GraphSAGE를 사용하여 정확도 95%를 달성하였으며, 웹 기반 전력 거래 플랫폼을 구현하여 서비스 실현 가능성을 검증하였다.

Ⅰ. 서 론

전세계적으로 소규모 분산전원 보급이 확대됨에 따라 에너지 프로슈머의 전력산업 참여 기회도 증가하는 추세이다. 특히, 에너지 프로슈머 등장으로 전력시장에 변화가 발생하면서 이해관계자의 역할 변화가 발생하고 있다. 그에 따라 전력회사는 계통신뢰도, 전력거래, 부가서비스 측면에서 에너지 프로슈머 확산에 대비하여야 한다.

한편, 에너지 프로슈머의 성장을 저해하는 요소는 잉여전력이며, 잉여전력을 해소할 수 있는 방안이 필요하다. 이를 위해 P2P 전력거래 시장이 점차 커지고 있는 추세이다. 국내에서는 프로슈머 사업을 통한 전력거래 허용, 상계거래 범위 확대 및 소규모 전력 중개사업 허용 등을 통해 P2P 시장의 활성화 방안을 모색하고 있다[1]. P2P 프로슈머 시장에서는 수요와 공급이 지속적으로 변화하며 거래의 단위가 작고 참여자가 많게 되므로 AI, 블록체인 등 신기술을 활용한 전력 거래 플랫폼의 지능화가 필수적이다.

본 논문에서는 신재생에너지 지역생산 및 소비 체계 활성화를 위해 최적으로 전력 구매 희망자와 전력공급자를 매칭하는 지역 맞춤형 전력 거래 파트너 매칭 플랫폼을 제안한다.

Ⅱ. 본론

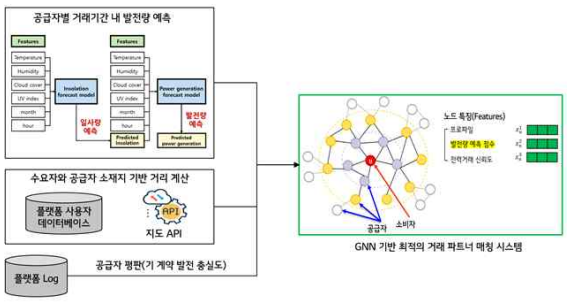
본 논문에서는 GraphSAGE[2]을 이용하여 인근 지역 내 최적의 전력 거래 파트너 후보군 추천 모델을 구현하였다. GraphSAGE는 Node간의 연결성에 대해 표현력이 우수한 그래프 신경망 모델이다. 전력 거래 파트너 매칭 플랫폼은 GraphSAGE 모델을 기반으로 전력 구매 희망자와 인근 공급자들의 정보를 조합하여 최적의 거래 파트너를 추천한다.

전력 거래 파트너 매칭 알고리즘은 크게 Graph를 구현하기 위한 데이터 전처리, 에너지 프로슈머와 전력 수요자간의 연결 여부를 예측하는 Link Prediction Task[3] 과정으로 나눌 수 있다.

Graph는 각 Node는 플랫폼에 참여한 에너지 프로슈머와 전력 수요자를 타나낸다. 따라서 플랫폼 사용자에 대한 프로파일을 Node의 Feature로 사용한다. 전력 수요자는 소재지 위치 정보(위도, 경도), 전력 공급량 및 단가(원/kWh) 등을 설정한다. 에너지 프로슈머는 발전소 위치 정보(위도, 경도), 희망 계약 전력량 및 단가(원/kWh), 계약 기간 등을 설정한다. 또한, Node feature에는 공급자의 발전량 예측 정보를 포함한다. 전력 수요자는 발전 설비가 없으므로 발전량 예측 값은 -1로 설정하였다. Node feature를 생성함으로써 사용자의 정보와 사용자간의 연결 정보를 동시에 학습할 수 있어 모델의 성능을 높일 수 있다.

Link Prediction Task는 에너지 프로슈머와 전력 수요자간 거래 가능성에 대해 예측하는 것이다. 전력 거래가 가능한 다수의 에너지 프로슈머 후보군을 추천하는 문제이므로 Multi-Label Classification[4]으로 정의할 수 있다.

Link Prediction Task는 기존 Node 간 연결 정보를 학습하여 미래 연결 가능성이 있는 Node를 예측하므로 과거 거래 기록이 필요하다. 본 연구에서는 가상의 거래 기록을 생성하기 위해 에너지 프로슈머 정보는 공공데이터포털의 남부발전소 태양광발전소 정보를 사용하였으며, 전력 공급량과 희망 단가와 등은 임의로 설정하였다. 전력 소비자의 정보는 임의로 구성하였다. 에너지 프로슈머와 전력 소비자 간 연결 정보는 사용자 프로파일을 기준으로 우선순위를 구성 및 매칭하여 학습에 활용하였다.



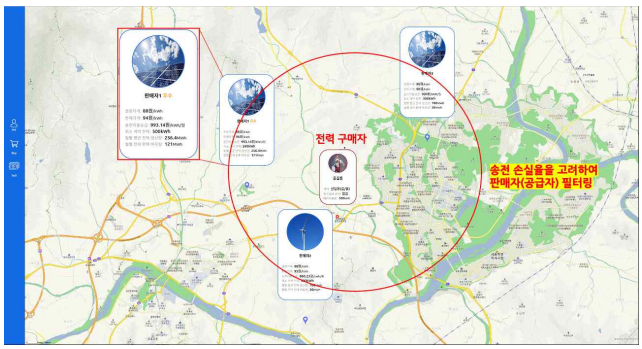
〔그림 1〕 전력 거래 파트너 매칭 서비스

본 논문에서 실험은 Python 3.7에서 TensorFlow 2.4.1과 Deep Graph Library[5]를 사용하여 구현하였다. 성능평가는 GraphSAGE, SimpleGNN(Simple Graph Neural Network), CNN(Convolutional Neural Network)을 사용하여 진행하였다.

	Train	Test
GraphSAGE	97.3%	95.4%
SimpleGNN	90.3%	93.1%
CNN	80%	78.5%

〔표 1〕 성능평가 결과

본 연구에서 제안하는 모델을 활용한 전력 거래 파트너 매칭 플랫폼의 기술적 구현 가능성을 검증하기 위해 POC(Proof of Concept)를 구현하였다. 아래 [그림 2]와 같이 전력 거래 파트너 매칭 플랫폼은 제공하여 전력 구매 희망자와 소재지 인근 공급자를 조합하여 공급자와 중개계약을 진행할 수 있다.



〔그림 2〕 전력 구매 희망자 파트너 추천 서비스 예시

Ⅲ. 결론

본 논문에서는 그래프 신경망을 이용하여 전력 거래 플랫폼의 전력 거래 파트너 매칭 서비스를 위한 모델을 제안하였다. 본 연구는 에너지 프로슈머의 사업 다각화를 통한 마이크로그리드 참여를 촉진시켜 신·재생에너지 지역생산 및 소비 체계 활성화 요인을 발견하는 데 의의가 있다. 전세계적으로 전력거래 회원 급증 및 RE100 참여 확산에 따라, 본 논문에서 제안하는 솔루션이 상용화로 확대가 가능해질 것으로 기대된다.

ACKNOWLEDGMENT

* 본 연구는 국토교통부/과학기술정보통신부/국토교통과학기술진흥원의 스마트시티 혁신성장동력 프로젝트 지원으로 수행되었음(과제번호 18NSPS-B149388-01)

참 고 문 헌

[1] 박찬국, 김양수, “우리나라 P2P 전력거래 가능성 연구”, 에너지경제연구원 수시연구보고서, 2016.

[2] Hamilton, William L., Rex Ying, and Jure Leskovec. "Inductive representation learning on large graphs." , Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017.

[3] Zhang, Muhan, and Yixin Chen. "Link prediction based on graph neural networks." Advances in Neural Information Processing Systems 31 (2018): 5165-5175.

[4] de Carvalho, André CPLF, and Alex A. Freitas. "A tutorial on multi-label classification techniques." Foundations of computational intelligence volume 5 (2009): 177-195.

[5] Wang, Minjie, et al. "Deep graph library: A graph-centric, highly-performant package for graph neural networks." arXiv preprint arXiv:1909.01315 (2019).

스마트글라스 및 경량 OCR 기반 축산동물 귀표 식별 시스템

전광명, 류인철, 김누리, 임채준, 오영우*, 전찬준*, 최우열*
인트플로우 주식회사, 조선대학교*
{kmjeon, inchul_ryu, somi-u, cjleem}@intflow.ai, snlyoungwoo@gmail.com, {cjchun, wyc}@chosun.ac.kr

Livestock eartag recognition system based on smart glasses and lightweight OCR

Kwang Myung Jeon, Inchul Ryu, Nuri Kim, Leem Chaejun, Youngwoo Oh*, Chanjun Chun*, Wooyeol Choi*
Intflow Inc., Chosun University*

요 약

본 논문은 스마트글라스와 경량화된 광학문자인식 (Optical Character Recognition, OCR) 기술을 활용한 축산동물 귀표 식별 시스템을 제안한다. 제안하는 시스템은 축산개체관리에 필수적인 동물식별용 귀표 (Ear Tag)를 축산 현장에서 신속하게 식별하기 위해 디스플레이와 카메라가 포함된 스마트글라스를 활용한다. 스마트글라스에 부착된 카메라는 화각 내 포함된 귀표의 영역을 식별하고, 해당 영역 내의 식별번호를 인식하기 위해 OCR을 활용한다. 이때 OCR은 스마트폰 혹은 그에 준하는 저전력 웨어러블 디바이스에서 실시간으로 구동되어야 하므로 일정 수준의 정확성을 유지하면서 낮은 복잡도를 지녀야 한다. 이에 제안하는 시스템은 귀표 OCR을 위한 문자영역 검출 및 문자식별 신경망에 대한 채널 가지치기 및 경량화를 수행하여 모바일 디바이스에서의 실시간 동작을 구현한다. 제안하는 방법의 성능을 검증하기 위해 스마트글라스의 카메라로 촬영한 한우의 귀표사진 1,000장에 대한 문장식별 정확성 및 반응 속도를 평가하였다. 평가결과 제안된 시스템은 평균 99.8%의 정확성을 지니면서 평균 62.3ms의 프레임당 처리시간을 보이며 실시간 구동이 가능함을 보였다.

Ⅰ. 서 론

축산물 이력관리는 가축의 출생에서부터 도축, 포장처리, 판매에 이르기까지의 정보를 기록, 관리하여 위생안전에 문제가 발생할 경우 그 이력을 추적하여 신속하게 대처를 하기 위한 제도이다 [1]. 국내에서는 현재 한우와 젖소의 사육과정에서 이력관리를 수행하고 있다. 이때 소와 젖소 개체의 고유 식별 정보를 담는 것이 귀표 (Ear Tag)이다. 귀표는 이력추적을 위한 개체식별번호를 표시하기 위해 문자와 숫자 및 바코드를 기재하여 귀에 부착할 수 있도록 제작한 표로, 인쇄형, 단추형, 재부착용 귀표로 그 종류를 나눌 수 있다 [2]. 12자리의 고유 숫자열로 구성된 개체식별번호는 가축 개체별 생에 주기를 담는 고유 식별자이다. 따라서 출생, 사양, 질병, 번식, 출하성적 등 가축 별 생애 전과정을 관리하기 위해서는 귀표의 개체식별번호의 인식이 필수적으로 선행되어야 한다. 하지만 현재의 귀표 인식은 사람의 육안으로 수행하는 경우가 대부분이며, 이를 자동화하기 위한 RFID 귀표 역시, RFID의 짧은 통신거리와 수신기 설치 및 운영에 따른 비용 부담의 이유로 축산 내 제한적인 영역에서만 활용어 왔다 [3]. 최근에는 스마트폰 기반 광학문자인식 (Optical Character Recognition, OCR)기술이 널리 보급되어 산업의 다방면에 널리 쓰이고 있다. 현재의 스마트폰 기반 OCR 기술 수준은 축산업 현장에서 귀표를 인식하는데 충분한 성능이다.하지만 양손이 자유롭지 못한 축산현장의 작업자들이 스마트폰을 조작하여 활용하기에는 그 이용성이 떨어지는 단점이 있다. 또한 충분한 정확성을 지니는 스마트폰 기반 OCR 기술들은 대부분 인공지능 서비스 제공자를 통한 클라우드 API 기반으로 동작하므로 축산현장에 사용하기 위해서는 API 활용에 따른 지속적 비용 발생 및 농장 내 인터넷 연결성이 보장되어야 한다는 제약이 존재 한다. 이러한 단점들을 극복하기 위해서는 축산현장에서 OCR 기술을 실용적으로 활용하기 위한 요소기술간 융합이 요구된다.

본 논문에서는 스마트글라스와 경량화된 OCR 기술을 활용한 축산동물 귀표 식별 시스템을 제안한다. 제안하는 시스템은 축산현장에서 작업자가 손쉽게 귀표를 식별하기 위한 목적으로 디스플레이와 카메라가 포함된 스마트글라스를 활용한다. 스마트글라스에 부착된 카메라는 화각 내 포함된 귀표의 영역을 자동으로 검출한 다음, 해당 영역 내의 식별번호를 인식한다. 기존의 클라우드 API 기반 OCR을 대신하여, 제안하는 시스템은 스마트폰 혹은 그 역할을 담당하는 저전력 웨어러블 디바이스에서 실시간으로 귀표인식을 수행하기 위한 경량화된 객체검출 및 문자열식별 신경망을 탑재한다. 스마트글라스와 영상인식 기술을 실용적으로 조합하여, 제안하는 시스템은 대부분 수기로 기록하는 한우, 젖소, 모돈 등 가축 개체별 사양관리 현황정보를 디지털 전환하는데 적합한 사용자 인터페이스를 제공한다.

서론에 이어 Ⅱ장에서는 제안하는 시스템의 HW 및 SW 구성, 그리고 각 신경망의 구조와 경량화 전략을 상세히 설명한다. 이어서 Ⅲ장에서는 제안 시스템의 실험 및 그 결과를 논의하고, 마지막으로 Ⅳ장에서 결론을 맺는다.

Ⅱ. 제안하는 축산동물 귀표 식별 시스템

제안하는 축산동물 귀표 식별 시스템은 카메라와 증강현실 디스플레이를 포함하는 스마트글래스를 주요 사용자 인터페이스 장치로 활용한다. 스마트글라스는 휴대가 가능한 저전력 모바일 디바이스와 연결되며 해당 디바이스에는 디스플레이 표출과 카메라 제어, 그리고 영상 기반 귀표 식별 연산을 수행하는 귀표 식별 SW가 포함된다. 귀표 식별 SW의 핵심 기능은 카메라로부터 입력된 동영상 프레임으로부터 귀표가 존재하는 영역을 탐지하고, 탐지영역에 표출된 귀표 번호를 식별하여 개체식별번호를 확인하는 것이다. 따라서 귀표식별 SW에는 딥러닝 기반 귀표영역 검출 모델과 귀표 내 개체식별번호 식별 모델이 구현된다. 각 모델은 개별 학습을 통해 주어진 데이터에 대한 최적 성능을 달성 후 경량화 과정을 추가적으로 거쳐 정확성 대비 연산 효율성을 극대화 하는 과정을 거친다. 각 모델 개발에 대한 세부 사항은 다음과 같다.

1. 귀표영역 검출 모델

제안하는 시스템의 귀표영역 검출 모델은 매우 낮은 계산량을 지니면서 비교적 높은 정확성을 제시하는 것으로 알려진 YOLOX [4] 구조를 기반으로 개발하였다. 다양한 YOLOX 구조 중에서 모바일 장치내 실시간 구현을 목적으로 설계된 YOLOX-Nano를 채택하였는데, 이 구조는 객체검출 분야에서 성능평가의 척도로 쓰이는 COCO (Common Objects in Context) [5] 평가 데이터셋 기준으로 25.3%의 평균 AP를 보여 종래의 경량 객체검출 모델 대비 우수한 정확성을 보인다 [5]. 또한 모델을 구성하는 파라메터 수는 0.91M으로 종래에 다양한 분야에 활용된 YOLOv3 모델의 파라메터 수인 63.00M의 1.4% 수준으로 대단히 낮다고 할 수 있다. 제안하는 시스템의 귀표영역 검출 모델은 COCO 학습용 데이터로 학습된 80-class YOLOX-Nano 모델 파라메터를 초기값으로 설정 후, 자체 수집한 귀표 이미지와 영역 레이블링 데이터셋을 활용하여 전이학습을 수행한다. 전이학습 과정에서 평가 데이터셋에 대한 AP (Average Precision)이 수렴하면 학습을 중단하고, 채널 가지치기 및 양자화를 수행하여 추가적인 경량화를 수행한다.

2. 개체번호 인식 모델

자동차 번호판 인식과 같은 영상기반 숫자열 인식 분야는 숫자가 일렬로 나열되어 있기 때문에 주로 객체검출이 아닌 문자열 인식 모델을 주로 사용한다 [6]. 하지만 그림 1에 나타난 것과 같이 귀표의 경우 숫자의 배치가 일렬로 되어있지 않고 상하 좌우에 나누어 배치되어 있으므로, 각 위치별 숫자를 식별하고 재정렬 할 필요가 있다. 이에 제안하는 시스템의 개체번호 인식 모델 역시 귀표영역 검출과 마찬가지로 YOLOX-Nano를 기본적으로 선택 하였다. 개체번호 인식의 경우 앞서 검출된 귀표영역 내의 이미지에서만 동작을 수행하며, 숫자 영역을 검출하고 그 영역 내 숫자를 식별한다. 식별된 12자리 숫자는 최상단 좌측의 숫자를 기준으로 3행으로 구분한 뒤 각 행별 숫자를 순서대로 나열하여 12자리의 개체식별번호로 정렬한다.

3. 신경망 모델의 경량화

상기 소개한 귀표영역 검출 및 개체번호 인식 모델들 모두 YOLOX-Nano 구조를 기반하여 이미 매우 경량화 된 모델들이지만, 추가적인 전력 효율의 이득을 얻기 위해 필터 가지치기 (Pruning)와 양자화를 수행한다. 가지치기 기법은 신경망을 구성하는 층 (Layer)에 따라 적합한 전략을 채택해야 한다. 제안하는 시스템의 신경망들은 대부분 합성곱 신경망 (CNN, Convolutional Neural Network)로 구성된 관계로, 합성곱 신경망의 가지치기에 큰 효과를 보이는 필터 가지치기 [7] 방법을 활용하였다. 구체적으로, 각 합성곱 층별 가지치기 전략으로 L1 거리 비교 전략 [7]을 사용하였으며, 가지치기 강도는 출력층에서 입력층으로 갈수록 강도가 낮아지도록 하여 입력 직후 얻는 저차원의 특징추출 정보의 손실에 따른 오류 전파의 영향을 낮추도록 하였다. 가지치기 된 모델들은 가지치기 전과 5% 이내의 성능 저하를 만족하도록 합성곱 층별 가지치기 강도 파라메터를 그리드 서치 방식으로 조정하며 훈련을 반복 하였다. 이렇게 얻어진 각 모델 파라메터들은 모바일 포팅을 위한 변환 툴을 활용하여 양자화 한다. 본 논문에서는 안드로이드 모바일 디바이스에서의 구동을 위해 Tencent社의 NCNN library [8]와 OpenCV Mobile [9]을 활용하였다.

4. 안드로이드 기반 귀표식별 SW

귀표식별 SW는 안드로이드 디바이스에서 구동하도록 앱 형태로 개발하였다. 앱은 카메라 API를 통해 영상을 입력받아 경량화된 귀표영역 검출 모델 및 개체번호 인식 모델에 이미지 프레임을 전달하여 사용자의 시야 안에 있는 축산동물 귀표의 개체식별 번호를 인식한다. 인식된 개체식별번호는 별도의 축산 개체관리 플랫폼과 연계하여 현장의 가축정보를 갱신하는데 활용한다.



그림 1. 한우 귀표 일반형에 대한 영역 검출 및 개체번호 인식 동작 예

표 1. 이표 인식 신경망 모델의 가지치기 전/후 성능 비교

	가지치기 전		가지치기 후	
	AP50	No. Parameter	AP50 (감소율)	No. Parameter (감소율)
이표영역 검출 모델	94.7%	0.891M	91.4% (3.4%)	0.596M (33.0%)
개체번호 인식 모델	92.2%	0.901M	88.0% (4.5%)	0.632M (28.1%)

Ⅲ. 성능 평가

제안하는 축산동물 귀표 식별 시스템의 성능은 다음과 같이 평가하였다. 먼저 귀표영역 검출 모델과 개체번호 인식 모델을 각각 1만장의 합성 이미지로부터 전이학습 기법으로 학습 시켰다. 다음으로 각 모델의 검증 데이터에 대한 정확성 척도인 AP50 (Average Precision, IoU=0.5)이 5% 이내로 낮아지는 조건을 만족하도록 필터 가지치기를 진행하였다. 표 1은 두 모델에 대한 가지치기 전 후 정확성과 파라메터 수에 대한 비교를 나타낸다. 실험 결과, 귀표영역 검출 모델은 경량화 후 33.0%의 파라메터 감소 효과와 3.4%의 정확성 감소를 보였으며, 개체번호 인식 모델은 경량화 후 28.1%의 파라메터 감소 효과와 4.5%의 정확성 감소 효과를 보였다. 이는 각 모델들이 이미 대단히 경량화된 구조를 지녔지만 필터 가지치기를 통해 미미한 성능감소를 대가로 추가적인 파라메터 감소가 가능함을 보인다.

다음으로 경량화가 진행된 각 모델들은 16비트 부동소수점 (fp16) 정밀도로 양자화 후 NCNN을 통해 안드로이드 앱 형태로 구현하였다. 이 안드로이드 앱을 상용 안드로이드 폰에 설치한 후 상용 AR 글래스와 연결하여 전체 시스템을 구성 하였다. 스마트글래스상에서의 구동 화면 및 UI 구성은 각각 그림2와 그림3에서 확인할 수 있다.

전체 시스템의 귀표인식 동작 성능을 검증하기 위해 스마트글라스의

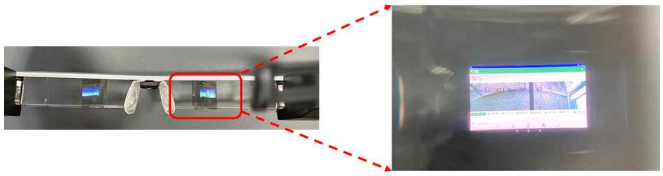


그림 2. 스마트글래스 스크린에서의 제안시스템 구동 화면

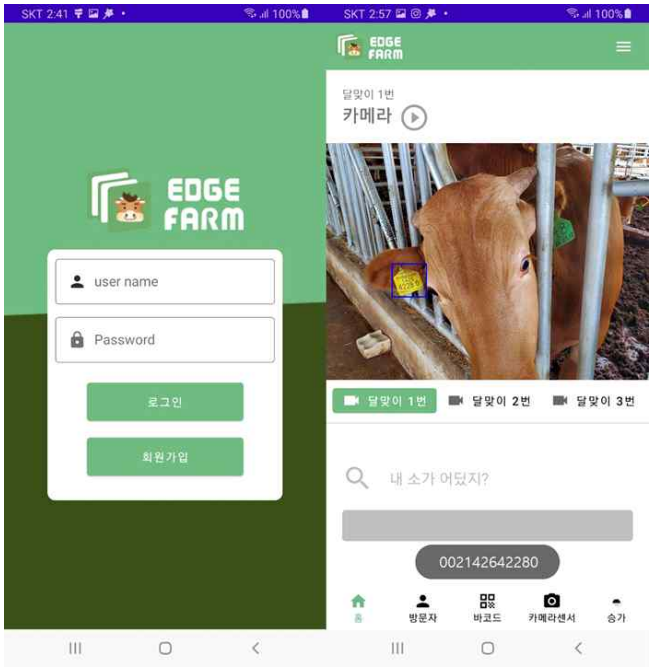


그림 3 제안시스템의 앱 UI 구성 화면

카메라로 촬영한 농장 내 한우의 귀표사진 1,000장에 대한 문장식별 정확성과 반응 속도를 평가하였다. 평가 결과 제안하는 시스템은 평균 99.8%의 정확성을 지니면서 평균 62.3ms의 프레임당 처리시간을 보이며 실시간 구동이 가능함을 보였다.

Ⅳ. 결론

본 논문에서는 스마트폰에서 구동 가능한 수준의 딥러닝 기반 객체검출 기술과 스마트글래스 UI를 활용한 축산동물 귀표 식별 시스템을 제안하였다. 본 시스템은 지난 20여년간 활용되어 온 축산동물 귀표의 표준을 준수하면서도 작업자가 스마트폰과 스마트글래스를 통해 현장의 개체별 정보를 실시간으로 인식할 수 있다는데 의의가 있다. 이후 축산분야에서 수기로 관리되어 온 축산개체별 관리대장을 실시간으로 읽어내는 경량화된 OCR 모델의 설계 및 구현방법에 대한 후속연구가 진행된다면 축산업 분야의 효율성 재고를 위한 디지털 전환의 속도를 높이는데 일조할 것으로 기대된다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학지원사업의 연구결과로 수행되었음. (2017-0-00137)

참 고 문 헌

[1]Lee, S. G., Song, W. I., Hong, Y. S., Park, M. H., Im, H. J., So, H. G., ... & Jeon, G. J. (2005). 국내 축산물 원산지 추적관리를 위한 생산이력관리시스템. In Proceedings of the Korean Society for Food Science of Animal Resources Conference (pp. 274-277). Korean Society for Food Science of Animal Resources.

[2] 한우자조금 관리위원회 소식지, 7월호, 2009.7

[3] Ng, M. L., Leong, K. S., Hall, D. M., & Cole, P. H. (2005, August). A small passive UHF RFID tag for livestock identification. In 2005 IEEE International Symposium on Microwave, Antenna, Propagation and EMC Technologies for Wireless Communications (Vol. 1, pp. 67-70). IEEE.

[4] Ge, Z., Liu, S., Wang, F., Li, Z., & Sun, J. (2021). Yolox: Exceeding yolo series in 2021. arXiv preprint arXiv:2107.08430.

[5] Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ... & Zitnick, C. L. (2014, September). Microsoft coco: Common objects in context. In European conference on computer vision (pp. 740-755). Springer, Cham.

[6] Wang H, Li Y, Dang L-M, Moon H. Robust Korean License Plate Recognition Based on Deep Neural Networks. Sensors. 2021; 21(12):4140. https://doi.org/10.3390/s21124140

[7] Li, H., Kadav, A., Durdanovic, I., Samet, H., & Graf, H. P. (2016). Pruning filters for efficient convnets. arXiv preprint arXiv:1608.08710.

[8] nihui et al. ncnn. https://github.com/Tencent/ncnn, 2017.

[9] nihui et al. opencv-mobile, https://github.com/nihui/opencv-mobile, 2021.

딥러닝을 이용한 풍력 블레이드의 상태 구분

김용빈, 김용민, 류상우, 김영민, 최인식*
한남대학교 전기전자공학과
*recog@hnu.kr (교신저자)

State Classification of Wind turbine blade using Deep Learning

Yong-Bin Kim, Yong-Min Kim, Sang-Woo Ryu, Young-Min Kim, In-Sik Choi
Dept. of Electrical and Electronic Eng., Hannam Univ.

요 약

본 논문은 풍력 블레이드의 동작 상태를 구분하기 위해 CNN(convolutional neural network)을 설계하고 구분 결과를 제시한다. 구분하고자 하는 블레이드의 상태는 외관상으로 '정상 블레이드(normal blade)' 및 '파손된 블레이드(broken blade)'를 가정하고, 동작 상태로는 '정상 회전(normal rotation)' 및 '불균형 회전(imbalance rotation)' 상태를 가정하였다. 풍력 블레이드의 상태별로 각각 다른 CNN 구분기를 설계하였다. 첫 번째 구분기는 블레이드의 외관 상태를 구분하는 구분기이며, 두 번째 구분기는 블레이드의 동작 상태를 구분하는 구분기를 설계하였다. 실험 결과 첫 번째 구분기는 5 dB의 SNR에서 99.83%, 두 번째 구분기는 같은 조건에서 98.68%의 구분성능을 보여 주었다.

Ⅰ. 서 론

풍력 블레이드는 한번 고장이 나면 정비가 어렵고 많은 정비 비용이 발생하므로 고장을 미연에 방지하기 위한 모티터링 시스템이 매우 중요하다. 참고문헌 [1]에서는 풍력 블레이드를 레이더로 모니터링할 때 블레이드의 동작 상태를 예측하기 위한 특성 벡터 추출 방법과 추출된 특성벡터를 이용하여 상태를 인식하는 매우 간단한 구분기 설계 기법을 제시하였다. 본 논문에서는 구분 성능을 보다 향상시키기 위해 CNN을 이용한 딥러닝 구분기를 이용하였으며, CNN의 입력 영상으로는 스펙드로그램(spectrogram)을 이용하였다.

Ⅱ. 본론

본 논문에서는 레이더를 이용한 풍력 블레이드의 모니터링에 있어서 블레이드의 상태를 자동으로 인식하기 위해 CNN을 이용하였으며, 입력 영상으로는 그림 1과 그림 2에서 제시한 각 상태별로 획득한 스펙트로그램 영상(350×263 pixels)을 이용하였다. 구분 실험을 위해 외관 상태를 구분하는 구분기와 동작 상태를 구분하는 구분기를 따로 설계하였다.

-5 dB부터 5 dB까지 SNR을 변화시켜가며 구분 실험을 진행하였다. 그 결과 SNR이 증가하면 구분 성능도 증가함을 알 수 있었으며, 구분기 1은 5 dB에서 99.83%의 구분 성능을 보였으며, 구분기 2는 98.68%의 구분 성능을 보였다. 이러한 결과는 참고문헌 [1]에서 제시한 86.56%와 96.62%보다 구분 성능이 향상됨을 알 수 있었다.

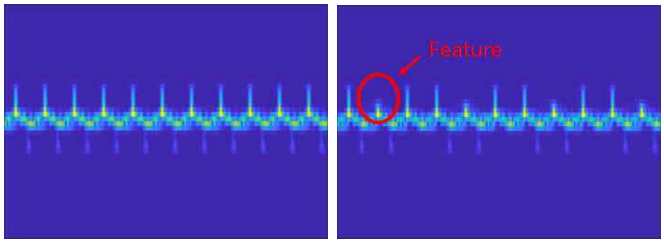


그림 1. 정상 블레이드와 파손된 블레이드의 스펙트로그램 (구분기 1)

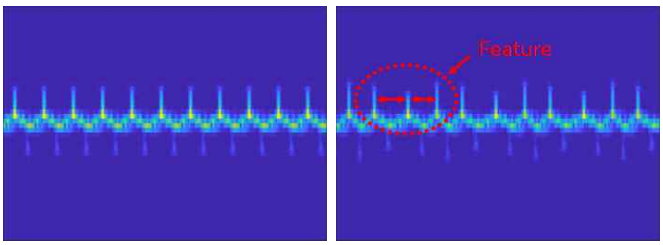


그림 2. 정상 회전과 불균형 회전의 스펙트로그램 (구분기 2)

블레이드 상태 구분을 위해 사용된 CNN의 구조는 350×263의 입력 영상으로부터 특성을 추출하기 위해 3개의 컨벌루션 및 풀링 계층을 가지며, 최종 구분 단계에서는 2개의 완전 연결 계층(fully connected layer)을 가지고 있다.

Ⅲ. 결론

본 논문에서는 24 GHz 도플러 레이더에 기반한 풍력 블레이드 모니터링 시스템에서 수신된 레이더 신호로부터 풍력 블레이드의 동작 상태를 자동으로 인식하기 위해 딥러닝 구분기를 이용하였다. 또한 CNN 구분기 입력 영상으로는 각 상태별로 획득한 스펙트로그램 영상을 이용하였다. 실험 결과 선행 연구인 참고문헌 [1]에서 사용한 구분기에 비해 CNN 구분기를 이용했을 때 구분성능이 향상됨을 알 수 있었다.

ACKNOWLEDGMENT

본 연구는 한국전력공사의 사외공모 기초연구(개별)에 의해 지원되었음(과제번호: R19XO01-50). 이 논문은 2021년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2021R1I1A3043120).

참 고 문 헌

[1] 김용빈, 최인식, "풍력 블레이드 상태 구분을 위한 특성벡터 추출 및 구분기 설계", 한국정보기술학회 논문지, Vol. 19, No. 6, pp. 88-97, Jun. 2021.

CaRFE: Candidates Recursive Feature Elimination for Black-Box Model

Song Youngjae, Kim Kwangsu*

Sungkyunkwan University, *Sungkyunkwan University

thddudwo1313@gmail.com, *kim.kwangsu@skku.edu

Abstract

RFE(Recursive Feature Elimination) is a widely-used greedy algorithm for a feature selection. However, the greedy feature selection does not guarantee obtaining an optimal feature subset as there is no method to predict perfectly precise feature importance for a black-box model such as a deep neural network. Thus, this paper proposes a novel variation of RFE for the black-box model, namely CaRFE(Candidates RFE). Through two experiments in the solar power forecasting dataset, we empirically show that CaRFE outperforms RFE in model performance and can reach a nearly optimal feature subset.

Ⅰ. Introduction

RFE[1] is a greedy feature selection technique eliminating unimportant features through evaluating the feature importance. Although there have been proposed several variation of the RFE[4, 5] so far, they usually focus on the traditional machine learning model rather than a black-box model such as deep neural network. As there is no method to predict perfectly precise feature importance for deep neural network, the greedy manner of the RFE misleads it to reach an optimal feature subset for deep neural network. Thus, we need an improved version of the RFE which is applicable to a black box model such as deep neural network.

Motivated by this, we propose a practical feature elimination framework, CaRFE(Candidates Recursive Feature Elimination). The CaRFE employs a candidates search to maintain the algorithm much less greedy and exploits both the feature importance and model performance metric. This paper computes a feature importance and model performance with KernelSHAP[2] and customized NMAE(Normalized Mean Absolute Error), respectively.

Through two experiments in the solar power forecasting dataset, we show that the CaRFE outperforms the RFE in model performance. Additionally, we identify that the CaRFE can reach a nearly optimal feature subset by comparing it with an exhaustive search. The contribution of this work is that we introduce an improved and practical version of the RFE appliable to a black-box model.

Ⅱ. CaRFE: Candidates Recursive Feature Elimination

CaRFE consists of 2 stages per one iteration to eliminate one unimportant feature. The first stage identifies candidate feature subsets C_i, $i \in 1,2,\ldots,p$, that the most unimportant feature is eliminated from the initial feature set F. Given the $C_1, C_2, \ldots, C_p$, the second stage stores a feature subset with the best model performance among them. Then the feature subset is assigned for the following initial feature set. The procedure above is recursively iterated until the number of remained features becomes k. Finally, among the stored candidates, CaRFE select a feature subset with the best performance. Algorithm 1 presents the details about the CaRFE algorithm. The CaRFE becomes RFE if we assign p to 1 and do not care about the model performance. The CaRFE is much less greedy than RFE due to the candidate search and able to reach a nearly optimal feature set as the user searches for the appropriate hyper-parameter p.

Algorithm 1: CaRFE

Input: F = Initial feature set
$M(F)$ = trained model with feature set F
$S(M)$ = list of feature importances for M
p = number of candidates
k = number of final features

Initialize the empty list *result*
while $|F| > k$ **do**
  Sort $S(M(F))$ in ascending order
  /* stage 1 starts */
  for $i = 1, 2, \ldots, p$ **do**
    $C_i \leftarrow F$
    $s_i \in S(M(C_i))$
    Remove the feature with s_i from C_i
  end
  /* stage 2 starts */
  Let C_{best} a feature subset with the best performance among $C_1, C_2, \ldots, C_p$
  $F \leftarrow C_{best}$
  Append C_{best} to *result*
end
Let *final* a feature subset with the best performance in *result*
return *final*

Ⅲ. Experiment

We apply RFE and exhaustive search to compare with the CaRFE. We work with the solar power forecasting dataset of the Ulsan solar plant in South Korea. The initial feature set F consists of 10 features: 'season',

'cos(time)', 'GHI', 'DNI', 'Temperature', 'Humidity', 'Wind_x', 'Wind_y', 'clearness' and 'Visibility'. In addition, we set 3 for both k and p. We implement an MLP architecture constructed by 1 dense layer with 256 neurons, 3 dense layers with 128 neurons, 1 dense layer with 64 neurons and 1 dense layer with 32 neurons sequentially. Dropout layers are placed with a 0.2 dropout rate between each layer. The activation function is ReLU for all dense layers. The optimizer is the Adam[3] with a 0.001 learning rate and the epochs are 100. The model performance metric used here is defined below.

$$NMAE = \frac{100}{c} \times \frac{1}{N} \sum_{i=1}^{N} \mid y_i - f(x_i) \mid$$
$$i \in \{ i \mid y_i \geq c \times 0.1 \}$$

The y_i and $f(x_i)$ indicate an actual power generation and the predicted one, respectively. And the c stands for the capacity of the solar panel in the Ulsan solar plant: 500. N is the number of all instances for the dataset.

The experiments are consist of 2 parts: comparison to the RFE and comparison to the exhaustive search. We evaluate the performance of the CaRFE in both cases with the NMAE metric.

3.1 Comparison to the RFE

We examine how low NMAE the CaRFE and the RFE can reach. To guarantee the reliance for the result, we repeat the experiment 10 times with different random seeds and take the average of NMAE. Figure 1 presents the average NMAE plot by 10 repetitions of the CaRFE and the RFE. It shows that the CaRFE can reach way lower NMAE, 4.294, whereas RFE can only get 4.608 for the lowest NMAE.

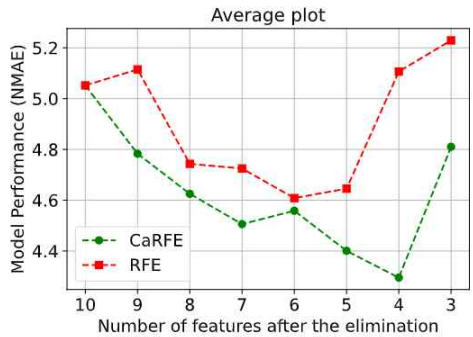


Fig 1. Average behaviors of NMAE during CaRFE and RFE

Additionally, since CaRFE is much less greedy in nature, the NMAE is lower than RFE over the whole interval. This indicates that CaRFE can select important features in practice regardless of the number of features a user wants to use.

3.2 Comparison to the exhaustive search

This experiment aims to ascertain whether the feature set obtained by CaRFE resembles the optimal feature set. We evaluate NMAE for all possible feature combinations with the exhaustive search for the given features: 968 cases in total, whereas CaRFE tries 21 cases for the given setting. If there is no significant difference between NMAE from the best feature set by CaRFE and exhaustive search, we can conclude that the CaRFE can find an almost optimal feature set in the given setting.

Table 1. CaRFE and exhaustive search comparison

	CaRFE	Exhaustive search
Best feature subset	season cos(time) GHI Wind_y	season cos(time) GHI Visibility
NMAE	4.212	4.191

Table 1 presents the best feature subset by the CaRFE and the exhaustive search. There is no significant gap between them with respect to the NMAE and the qualitative comparison. This implies that the CaRFE can find a nearly optimal feature set, unlike the RFE.

Ⅳ. Conclusion

In this work, we present the CaRFE, the improved version of RFE. The CaRFE exploits both the model performance metric(NMAE) and feature importance computed by KernelSHAP with candidates search. The CaRFE is more reliable than the RFE and applicable to a black box model. The two quantitative experiments show that the CaRFE outperforms the RFE and can find nearly optimal feature set in the solar power forecasting dataset.

ACKNOWLEDGMENT

This research was partly supported by the MSIT(Ministry of Science and ICT), Korea, under the Grand Information Technology Research Center support program(IITP-2021-2015-0-00742) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation) and Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.2020-0-00990, Platform Development and Proof of High Trust & Low Latency Processing for Heterogeneous·Atypical·Large Scaled Data in 5G-IoT Environment)

Reference

[1] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines" Machining Learning, vol. 46, no. 1-3, pp. 389-422, 2002.

[2] Scott M Lundberg and Su-In Lee. "A unified approach to interpreting model predictions" In Advances in Neural Information Processing Systems, pp. 4768 - 4777, 2017.

[3] Diederik Kingma and Jimmy Ba. Adam, "A method for stochastic optimization" In ICLR, 2015.

[4] Chen and Jeong "Enhanced recursive feature elimination" In IEEE sixth international conference on machine learning and application. pp. 429-435, 2007

[5] H. Jeon and S. Oh "Hybrid-recursive feature elimination for efficient feature selection" Appl. Sci, 10. pp. 3211, 2020

Performance Analysis of Clipping based PAPR Reduction Technique in UFMC Systems

Esmot Ara Tuli, Jae Hyun Lee, Dong-Seong Kim, Jae Min Lee
IT Convergence Engineering, Kumoh National Institute of Technology.
(esmot, antigen47, dskim, ljmpaul)@kumoh.ac.kr

Abstract

Universal filtered multi-carrier (UFMC) is one of the strongest candidates waveform to overcome the shortness of orthogonal frequency division multiplexing (OFDM) and for the future mobile communication system. UFMC system provides robust, high spectral efficiency and minimal out-of-band(OOB) emission communication system. However, UFMC system has high peak-to-average power ratio (PAPR) problem like other multi-carrier transmission systems. This paper study performance of Clipping method in different threshold level for PAPR reduction in UFMC system.

Ⅰ. Introduction

Wireless communication systems are changing rapidly to fulfill the demand for Internet-of-Everything (IoE) and real-time communication. It reflects on economic development and flexibility [1]. UFMC system is the combined form of orthogonal frequency division multiplexing (OFDM) and Filter Bank Multicarrier (FBMC). UFMC carries the basic feature of OFDM system. Instead of filtering the whole frequency band, it uses sub-band filter like FBMC system. In UFMC system, each sub-carriers are divided into sub-bands. and filter is applied to each sub-band. For filtering UFMC system used Dolph-Chebyshev filter.

Like other multicarrier transmission systems UFMC system also have PAPR problem which reduces system performance. In general PAPR reduction technique can be classified into three groups as distortion, encoding, and probabilistic approaches. Each of the systems has its own limitations. Among all PAPR reduction techniques, clipping is the simplest method. It requires low computational complexity and minimum hardware complexity unlike selected mapping (SLM), partial transmit sequence (PTS), Tone Reservation [2] etc. The main contribution of this paper are as follow:

- Reduce PAPR in UFMC system using the Classic Clipping method.
- As a performance analysis of PAPR reduction technique reduces complementary commutative-distribution function (CCDF) compare to original UFMC system [3].
- BER graph also almost same as original signal in Additive White Gaussian Noise(AWGN)

The rest of the paper is organized as follows. system model and PAPR introduces in section II, In section III, describes Clipping and CCDF method. Section IV, presents the simulation work. Finally, the paper is conclude in section V and provides some future work direction

Ⅱ. System model and PAPR

Figure 1, shows UFMC transmitter block diagram. As from the figure input signal first mapped as QAM symbol and divided into sub-bands with fixed size. Then the signal goes through the N-point inverse fast Fourier transform (IFFT). The signal can be defined as

$$\mathrm{x_b} = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} x_b(c) e^{\frac{j2\pi cn}{N}} \quad (1)$$

Where x_b are the sub-bands, x_b is the sub-carrier. Dolph-Chebyshev filter is applied and combined together and make the signal as

$$\mathrm{s(n)} = \sum_{b=1}^{B} x_b * f_b \quad (2)$$

Where f_b is the filter impulse on each sub-band. PAPR of the candidate signal $\mathrm{s(n)}$ can be expressed as-

$$PAPR = 10\log_{10}\frac{P_{peak}}{P_{average}} = 10\log_{10}\frac{\max_n[|s(n)|^2]}{E[|s(n)|^2]} \quad (3)$$

where $\mathrm{E[.]}$ is the expected value. P_{peak} is the highest instantaneous power and $P_{average}$is the average power of the signal.

Ⅲ. Clipping and CCDF

PAPR can be reduced by clipping the signal at a fixed level. Consider the signal clipped at level and phase remain same. Then clipping described as

$$c(s) = \begin{cases} s, & s \le C \\ C, & s > C \end{cases} \quad (4)$$

where s is previous step of clipping, $\mathrm{c(s)}$ is the result signal after clipping and C is the clipping level

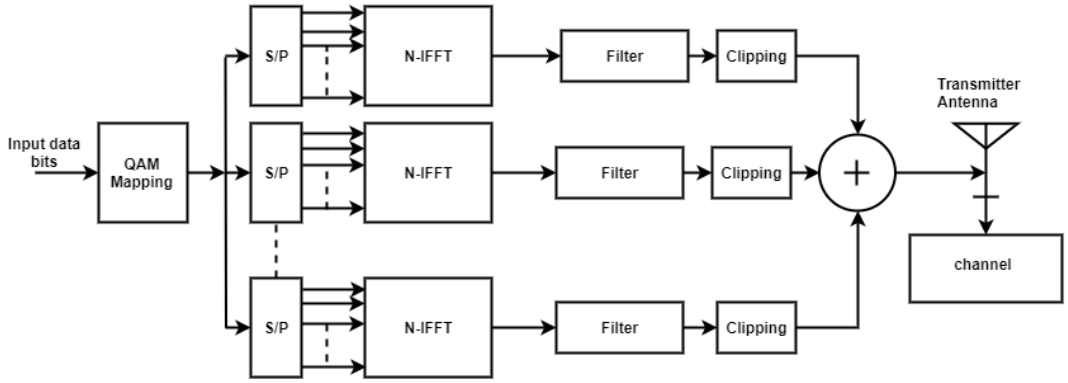


Fig. 1: Block diagram of typical UFMC transmitter.

PAPR performance of a signal can be measured by Complementary Cumulative Distribution Function (CCDF), which can be defined as-

$$\mathrm{P}(\mathrm{PAPR} > PARP_0) = 1 - (1 - e^{-PARP_0})^N \quad (5)$$

where $PARP_0$ defined as clipping level. This equation measures the probability of a signal PAPR exceed the clipping level $PARP_0$

Ⅳ. Simulation and performance evaluation

Clipping based PAPR reduction of UFMC signal simulation was performed in MATLAB software. 4 QAM signal divided into 10 sub-band of size 20 and goes through the system with size of fast fourier transform(FFT) is 512. After inverse FFT, Chebyshev filter is applied in each sub-band. Clipping is applied in each sub-band. Finally all the signal are grouped together and formed UFMC signal which next pass through the channel.

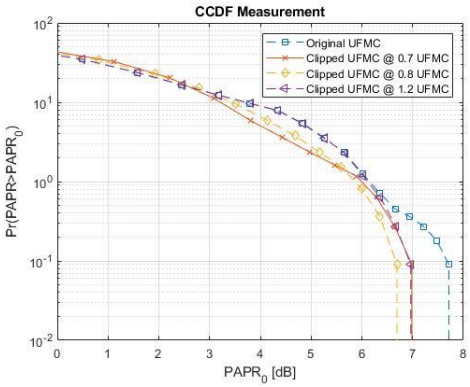


Fig. 2: PAPR value of UFMC system at different clipping level

Ⅴ. Conclusion

Among all PAPR reduction method, clipping is simple to implement and computational complexity low. In this study we investigate the PAPR performance and bit error rate at different clipping level. However the optimal clipping factor is 1.2 where PAPR reduced with low BER rate.

As we can see in the figure 2 and 3, clipping level 1.2 is optimal. At this level the PAPR is reduced and also it does not effect in bit error rate. As bit error rate is same as original signal. In clipping level .8 PAPR is the lowest but clipping at this level create BER.

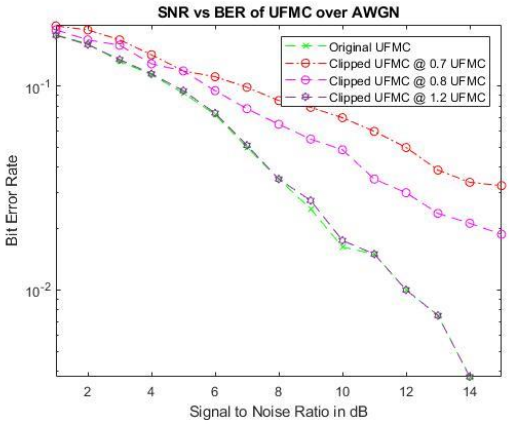


Fig. 3: BER of the UFMC system at different clipping level

ACKNOWLEDGMENT

This research work was supported by Priority Research Centers Program through NRF funded by MEST (2018R1A6A1A03024003) and NRF grant funded by Korea government(MSIT)(2019R1F1A1064055) and the Grand Information Technology Research Center support program (IITP-2021-2020- 0-01612) supervised by the IITP by MSIT, Korea.

참 고 문 헌

[1] K. P. Dirgantoro, W.-P. Nwadiugwu, J. M. Lee, and D.-S. Kim, "Dual fieldbus industrial iot networks using edge server architecture," Manufacturing Letters, vol. 24, pp. 108-112, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2213846319301518

[2] M. I. Al-Rayif, H. E. Seleem, A. M. Ragheb, and S. A. Alshebeili, "Papr reduction in ufmc for 5g cellular systems," Electronics, vol. 9, no. 9, 2020. [Online]. Available: https://www.mdpi.com/2079-9292/9/9/1404.

[3] D. Nugroho and D.-S. Kim, "Precoded dct and low complexity slm for papr reduction in ofdm systems," International Journal of Engineering and Industries, vol. 4, no. 4, pp. 43-49, 2013.

보행자 시점의 멀티 도메인 교차로 분류 문제에서의 도메인 제거 손실 함수 비교

Marcella Astrid[1,2], Muhammad Zaigham Zaheer[1,2], 이승익 [1,2]
[1] 과학기술연합대학원대학교, [2] 한국전자통신연구원

{marcella.astrid, mzz}@ust.ac.kr, the_silee@etri.re.kr

Comparisons of Domain-Removal Losses in Multi-Domain Pedestrian-View Intersection Classification

Marcella Astrid[1,2], Muhammad Zaigham Zaheer[1,2], Seung-Ik Lee[1,2]
[1]University of Science and Technology
[2]Electronics and Telecommunications Research Institute

Abstract

For pedestrian-view intersection classification model that can run on multi-domain input, i.e., indoor and outdoor, we adopt method from Kim *et al.*, [1]. The training objective is to not extract domain related features by using negative entropy loss on domain prediction. However, with similar goal, Alvi *et al.* [2] utilize confusion loss, i.e., cross entropy to uniform distribution. In this work, we compare the two losses on our multi-domain pedestrian-view intersection classification model.

Ⅰ. Introduction

To assist navigation of small robots, several researchers [3, 4] propose pedestrian-view intersection classification methods. However, these models are implemented only on one domain, i.e., outdoor domain. To incorporate both indoor and outdoor domains, we adopt method from Kim *et al.* [1] in order to remove the domain information in the extracted features. In this paper, we focus on comparing two losses, i.e., negative entropy and confusion loss, utilized in multi-domain training to reduce domain information in the extracted features.

Ⅱ. Methodology

We adopt model from Kim *et al.* [1] to train our multi-domain intersection classification model. As seen in Fig. 1, the architecture consists of two branches. The first branch is utilized to predict intersection categories $\hat{\boldsymbol{y}} \in \mathbb{R}^2$ and the second branch is utilized for domain prediction $\hat{\boldsymbol{z}} \in \mathbb{R}^2$. The training objective is to prevent the base network $\mathcal{F}$ to extract domain information while still extracting the features related to the intersection classification. To correctly classify the intersection, we train the model to minimize cross entropy loss between prediction $\hat{\boldsymbol{y}}$ and its ground truth $\boldsymbol{y}$. In addition to the loss, to remove the domain from the features extracted by the $\mathcal{F}$, Kim *et al.* [1] incorporate negative entropy loss which has to be minimized by $\mathcal{F}$ as:

$$\min_{\mathcal{F}} \alpha \left(\sum_{i=1}^{2} \hat{z}_i \log \hat{z}_i \right), \tag{1}$$

where α is weighting hyperparameter. In order to minimize this loss, $\mathcal{F}$ should not extract any domain information so that the entropy of $\hat{\boldsymbol{z}}$ is maximized, i.e., unconfident $\hat{\boldsymbol{z}}$ prediction. The loss value and its gradient are visualized in Fig. 2(a).

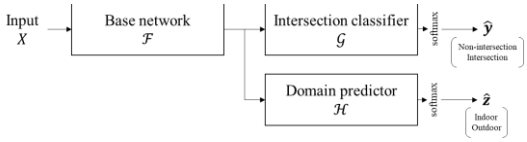


Figure 1. Architecture setup for our multi-domain pedestrian-view intersection classification during training. At test time, only the base network and the intersection classifier are utilized.

Instead of negative entropy, with similar goal to the negative entropy loss, some researchers [2] utilize confusion loss, i.e., cross entropy loss with respect to the uniform distribution as:

$$\min_{\mathcal{F}} \alpha \left(-\sum_{i=1}^{2} 0.5 \log \hat{z}_i \right). \tag{2}$$

This loss is minimized if the domain predictor predicts 0.5 confidence for each indoor and outdoor class, i.e.,

it is unable to predict the domain from the extracted features. The loss value and its gradient can be seen in Fig. 2(b). In this work, we focus on comparing Eq. (1) and (2) in our pedestrian-view intersection classification model.

Additionally, following Kim *et al.* [1], we also add an adversarial training for the domain prediction. $\mathcal{F}$ maximizes cross entropy between domain prediction $\hat{\boldsymbol{z}}$ and its ground truth $\boldsymbol{z}$. On the other hand, $\mathcal{H}$ minimizes the loss. During test, only $\hat{\boldsymbol{y}}$ is considered for accuracy measurements.

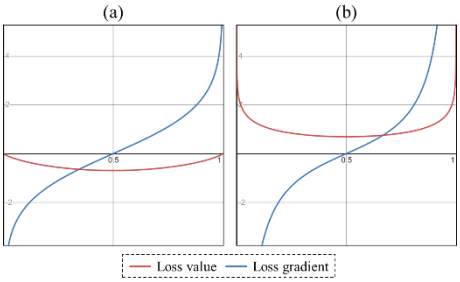


Figure 2. Graphs representing loss values and their gradients with respect to the prediction: (a) $y = x\log x + (1-x)\log(1-x)$ similar to negative entropy in Eq. (1), (b) $y = -0.5\log x - 0.5\log(1-x)$ similar to confusion loss in Eq. (2). In this graph, x-axis is confidence of indoor class $\hat{z}_1$ and since the final output is softmax, confidence of outdoor class is $\hat{z}_2 = 1 - \hat{z}_1 = 1 - x$.

Ⅲ. Experiments

For indoor dataset, we collected 2,020 and 791 images for non-intersection and intersection categories, respectively. The dataset is sampled from various YouTube videos recorded in multiple scenes. For outdoor dataset, we collected 2,092 and 2,713 images for non-intersection and intersection categories, respectively. The dataset is recorded in urban areas and parks. Each indoor and outdoor dataset is first divided into train, validation, and test sets with ratio of 5:2:3. Then, we combine the train and validation set of both domains as the overall training and validation sets. However, we test our model in each domain separately.

For our model, we utilize ImageNet-pretrained ResNet-18 [5]. The $\mathcal{G}$ and $\mathcal{H}$ branches start from the third residual block. Fine-tuning using our dataset is conducted using Adam, learning rate 10^{-5}, weight decay 10^{-6}, and batch size 64. The highest validation accuracy out of 500 training epochs are selected for testing. We run each experiment five times and report the average.

For training, the input image is first cropped into random size between range of 0.8 to 1.0 from the original size then resized to 224×224. Additionally, we add random horizontal flip, contrast, brightness, sharpness, and color balance augmentation. Finally, we normalize the input values to range 0 to 1. For preprocessing during validation and test, we only resize and normalize the input.

Table 1 shows the accuracy comparisons between models trained using negative entropy loss and models trained using confusion loss on various α settings. With higher values of α, model trained using negative entropy tends to perform better than using confusion loss. This phenomenon can be caused by the gradients. As seen in Fig. 2, confusion loss tends to have steeper gradient than the negative entropy loss which makes it more susceptible with higher α.

Table 1. Test accuracy comparisons between models trained on negative entropy and confusion loss on different weighting α.

	Negative Entropy (Eq. (1))		Confusion Loss (Eq. (2))	
α	Indoor	Outdoor	Indoor	Outdoor
1	**80.07**	**68.62**	69.61	57.16
0.1	**80.67**	**68.90**	73.99	63.73
0.01	**80.92**	**68.08**	80.75	67.64
0.001	**81.03**	67.85	80.82	**68.25**
0.0001	79.93	67.60	**80.78**	**68.37**

Ⅲ. Conclusion

In this paper, we compare two losses, i.e., negative entropy and confusion loss, utilized in multi-domain setting of pedestrian-view intersection classification. The results show that using a higher weighting factor, negative cross entropy is more stable compared to the confusion loss.

ACKNOWLEDGMENT

This work was supported by the ICT R&D program of MSIT/IITP. [2019-0-01309, Development of AI Technology for Guidance of a Mobile Robot to its Goal with Uncertain Maps in Indoor/Outdoor Environments].

참 고 문 헌

[1] B. Kim, H. Kim, K. Kim, S. Kim, and J. Kim, "Learning not to learn: Training deep neural networks with biased data," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 9012-9020.

[2] M. Alvi, A. Zisserman, and C. Nellåker, "Turning a blind eye: Explicit removal of biases and variation from deep neural network embeddings," in *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, 2018, pp. 0-0.

[3] M. Astrid, M. Z. Zaheer, J.-H. Lee, J.-Y. Lee, and S.-I. Lee, "What do pedestrians see?: Visualizing pedestrian-view intersection classification," in *2020 20th International Conference on Control, Automation and Systems (ICCAS)*. IEEE, 2020, pp. 769-773.

[4] M. Astrid, J.-H. Lee, M. Z. Zaheer, J.-Y. Lee, and S.-I. Lee, "For safer navigation: Pedestrian-view intersection classification," in *2020 International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, 2020, pp. 7-10.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.

Deep Learning for Corrupted Image-Text Matching

Sunwoo Kim, Seongwook Shin, and Byonghyo Shim
Seoul National University

{swkim, swshin, bshim}@islab.snu.ac.kr

Abstract

Image-text matching plays a vital role in connecting clean images and texts. In many real-world applications, since a large number of corrupted images exists, it is very difficult to match images and texts accurately. In this paper, we propose a novel corrupted image-text matching scheme, referred to as robust corrupted image-text matching (RCIT), suitable for various real-world scenarios.

Ⅰ. Introduction

In recent years, deep learning (DL) has achieved great success in a variety of real-world problems such as network localization [1]. The application of DL techniques in the field of image-text matching has also emerged as an active topic in the last three years [2]. Image-text matching is a task to search for images that match with their corresponding texts (i.e., sentences) and vice versa. Although great progress has been made, image-text matching remains challenging in many real-world applications such as google image search due to a large number of corrupted images. When image quality degrades, it is highly likely that the conventional image-text matching schemes cannot perform image-text matching effectively.

In this paper, we propose a novel image-text matching scheme suitable for various real-world scenarios. The main idea of the proposed scheme, henceforth referred to as *robust corrupted image-text matching* (RCIT), is to gather the matching triplets (clean and corrupted images and text) in the common embedding space. It is now well-known that many image-text matching schemes rely on common embedding space, where the distance between paired image and text instances is reduced, to infer the image-text similarity [3]. By gathering matching triplets together in the embedding space, we can achieve accurate image-text matching even with corrupted images.

Ⅱ. Visual Semantic Embedding

In this section, we briefly review the basics of visual semantic embedding (VSE) approach used in image-text matching. First, we use the image feature extractor (e.g., CNN) and text feature extractor (e.g., RNN) to obtain the set of visual and text features from image and text (i.e., sentence), respectively. Then, the extracted features are aggregated by feature aggregators into the common embedding space:

$$\mathbf{v} = f_{VISUAL}(\{\phi_n\}_{n=1}^{N})$$
$$\mathbf{u} = f_{TEXT}(\{\psi_n\}_{n=1}^{N})$$

where $\mathbf{v}$ and $\mathbf{u}$ are visual and text instance vectors of same dimension in the common embedding space.

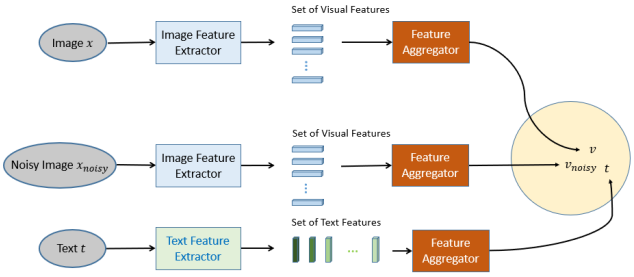


Figure 1. Overall Training Process of RCIT

Finally, the image-text similarity score is defined as the dot product between $\mathbf{v}$ and $\mathbf{u}$. In the test phase, the similarity scores are used to rank a query text against all possible images and vice versa.

Ⅲ. Corrupted Image-Text Matching

Fig. 1 depicts the overall training process of RCIT scheme. The key operation of RCIT is to jointly minimize the distance between clean image-text pair and corrupted image-text pair. In doing so, we can learn the proper embeddings representation of corrupted images in the common embedding space Distinctive feature of RCIT over the conventional image-text matching scheme is that matching clean and corrupted images and texts are used simultaneously in the training process. To jointly train the three feature extractors, we combine the two bidirectional ranking losses used for matching clean image-text and corrupted image-text pairs with the loss used for matching corrupted image-clean image pair [2]. That is,

$$L = L_{CI-T} + L_{I-T} + L_{CI-I}$$

Since three modules are trained together, the similarity representations are shared, allowing greater capacity in characterizing the complex alignments between corrupted images, clean images, and sentences.

IV. Simulation Result

In this section, we present the simulation results to evaluate the image-text matching performance of the proposed RCIT. For dataset, we use Flickr30k, which

	Text Retrieval	Text Retrieval	Text Retrieval	Image Retrieval	Image Retrieval	Image Retrieval
Techniques	R@1	R@5	R@10	R@1	R@5	R@10
RCIT	50.1	76.8	84.6	35.3	61.5	72.1
VSE++	39.7	56.6	61.9	30.6	52.3	59

Table 1. Corrupted Image-Text Matching performance

consists of 29,783 training images, 1,000 images for validation and testing [2]. To generate the corrupted image from the given clean image, we contaminate with the clean image with either speckle or salt-pepper noise with noise rate of 0.6. Note that noise rate is defined as the fraction of number of unclean pixels over the total number of pixels of an image. For fair comparison, we train VSE++ with corrupted images. As performance measure, Recall at K (R@K), which is defined as the proportion of queries whose ground-truth exists within the top K chosen candidates, is considered. In Table 1, we evaluate the image-text matching performance of RCIT and compare it with that of VSE++, a state-of-the-art VSE-based image-text matching technique [3]. From the result, we observe that R@1, R@5, and R@10 of RCIT are much higher than those of VSE++.

V. Conclusion

In this paper, we proposed a novel image-text matching scheme suitable for many real-world applications where a large number of images with pool quality exists. We demonstrated from the numerical evaluations that the proposed scheme is very effective in terms of matching image and text accurately regardless of image quality.

ACKNOWLEDGMENT

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program (IITP-2021-2017-0-01637) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation).

REFERENCE

[1] S. Kim and B. Shim, "Localization of internet of things network via deep neural network based matrix completion," in Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC), pp. 1766-1770, Oct. 2020.

[2] H. Diao, Y. Zhang, L. Ma, and H. Lu, "Similarity reasoning and filtration for image-text matching," arXiv:2101.01388, 2021.

[3] F. Faghri, D. J. Fleet, R. Kiros, and S. Fidler, "VSE++: Improving visual-semantic embeddings with hard negatives," in Proc. Brit. Mach. Vis. Conf. (BMVC), pp. 1-13, 2018.

Anti-Drone Systems Design: Safeguarding Airspace through Real-Time Trustworthy AI Paradigm

Simeon Okechukwu Ajakwe, Rubina Akter, Da Hye Kim, Golam Mohatsin, Dong Seong Kim, Jae Min Lee
Department of IT Convergence Engineering, Networked System Laboratory
Kumoh National Institute of Technology, Gumi, South Korea
(simeon.ajakwe, dahem91,golam248, dskim, ljmpaul)@kumoh.ac.kr, rubinaakter2836@gmail.com

Abstract

This paper provides a noble direction for guaranteeing safety of airspace for the use of drones in various transportation and logistics services in a smart city. Yolov5 deep learning model was used for visual drone detection and payload identification. The model showed 99.9% drone detection accuracy in less time but misidentified some of the attached objects due to its concealment. For robust discerning and countering of drones' usage for illegal, devious, and deadly purposes, the model's improvement is necessary to maintain sanity in airspace through responsible AI products for next-generation scenarios.

Ⅰ. Introduction

The polarization and proliferation of use of drones and its' derivatives for transportation and logistics without adequate check of legitimacy and privacy protection has sparked global concerns as it portends great security risk [1]. The drones' inexpensiveness, ease of operation, and dynamic evolution of its underlying technologies has created a gruesome challenge for developing an encompassing system for countering usage of drones for illegal and obnoxious purposes.

Though, Drone Transportation System (DTS) guarantees speedy service delivery, the influx of drones into the airspace is creating social panic as incidences of drone accidents, drone hijacking, terrorist attacks, privacy invasion, across border smuggling, etc., are on the increase [2]. These concerns question the verity of Artificial Intelligence (AI) responsiveness in maintaining technological friendly environment in its quest to mimic humans.

Anti-drone systems help to defend against drone accidents, and counter terrorist and illegal airspace activities perpetrated using unmanned aerial vehicles (UAV) [3]. However, the narrow mindedness of existing anti-drone systems in terms of militarization, lack of adaptive and simultaneous multi-drone detection, identification, interception, and neutralization method, poor response procedures, absence of drone flight regulation, amongst others, demands an environmentally friendly, robust, reliable, and ethical based approach to keep sanity in the airspace and restore societal confidence in AI technologies and its products.

This paper provides a conceptual framework on trustworthy anti-drone system design for maintaining sanity in the airspace. It explores and emphasizes the need for convergence technologies in the design of responsible AI products.

Ⅱ. Main subject

An Anti-drone system is an event-triggered critical-mission based hard real-time system deployed to detect a drone's location, determine its legality/illegality, harmfulness or otherwise, and decide the best method of neutralizing its' operation. These multi-dimensional task cuts across the convergence of computer vision, cognitive reasoning, wireless sensor, network, edge computing, blockchain, etc. [4]. The design of this event-triggered critical-mission real time system must therefore incorporate demanding response time, predictable peak load performance, environmentally friendly pace control and autonomous safety measure features as seen in figure1.

AI especially Deep Convolutional Neural Network (DCNN) coupled with detection techniques (visual, acoustic, radio frequency, thermal, and radar) has proven a veritable tool for solving complex dynamic event-driven problems as drone detection and identification [5]. DCNN Object detection models such as Faster RCNN, Yolo, SSD, MobileNet, etc. provides accurate and timely drone detection. Yolov5 model is used in this paper for object detection and identification respectively.

Besides identifying drone's source and its intended action in real-time, accurate and quick identification and interception of the payload (explosives, gun, recording device, etc.) attached to a drone is of utmost critical importance as the drone is just a vehicle conveying the object.

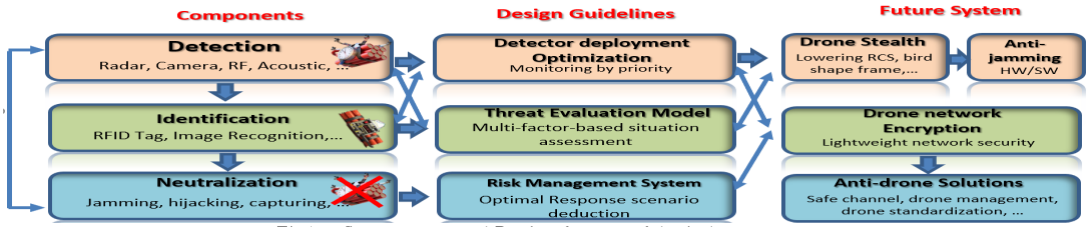


Fig1: Components and Design factors of Anti-drone system

To achieve this, a super-positioning strategy is formulated with features of drone model, characteristics and sensitivity of deployment area, and assessing potential threat level (high/low).
The perceived damage likely to occur R is a function of the payload attached, flight path and response time expressed as:

$$R = (R_{object} + R_{path})^{Rtime}, \quad \cdots (1)$$

where R_{object} is physical threat level, R_{path} is flight path and R^{time} is response time with maximum value set at, β, and 1 respectively. Physical threat level (R_{object}) is a function of the drone's weight/mass and speed derived from its kinetic energy given as:

$$K = \tfrac{1}{2}mv^2, \quad \cdots (2)$$

where m and v are drone's mass and speed respectively. A variation from the default drone weight is a signal that a potential payload is attached. Timely visual identification of the attached object is of vital importance in deciding the neutralization strategy to adopt. Hence, the overall damage of a drone in airspace is a function the drone size, noise, payload, and scanability as shown in equation 3.

$$R_{total} = \max [\alpha (R_{object,\,kinetic} + R_{object,noise} + R_{object,\,payload} + R_{object,\,scannable}) \times N_{drone}], \quad \cdots (3)$$

with N_{drone} being the number of swarming drones.

To achieve this requires transiting from weak AI (Artificial Narrow Intelligence) to General AI-based Anti-drone system which must have specialized environment-sensitive multi-drone detection capability, intuitive payload identification ability, adaptive and situation-based defensibility, co-operative interconnectivity and security alertness, scalability, and above all, efficient non-militarized neutralization strategy in handling envisaged UAVs.

Experimental simulation was carried out using 2000 datasets generated through testbed setup for drone detection and payload identification. Table 1 shows the dataset description and result of drone detection signifying high accuracy within minimal time.

Table1: Dataset Description and Detection Result

Type of drone	No. of items	Data Split (%)	Precision	Detection Time
Anafi-Ext	500	Training, Test, Valid (70, 20,10)	99.9%	0.022s
DJIFPV	500		100%	0.022s
Other drones	1000		99.9%	0.021s

For payload object identification, seven objects; bomb, guns, missile, etc. were attached to several drones and positioned at different distances as shown in the results in figure 2 and figure 3 respectively.

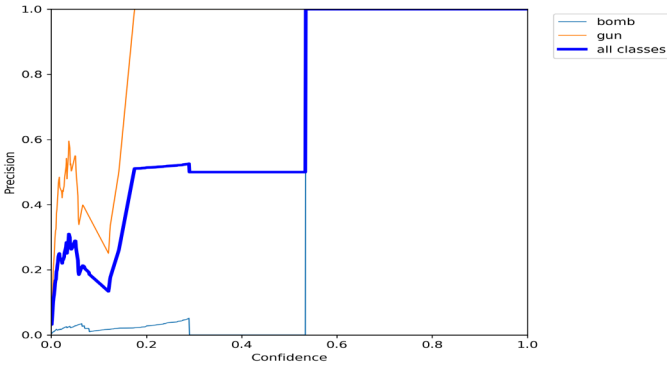


Figure 2: Model Precision of Payload

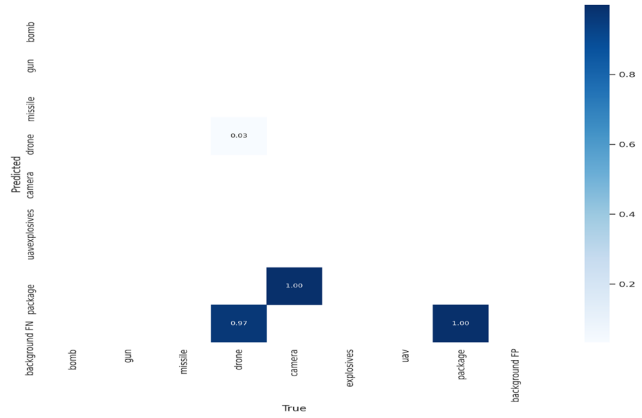


Figure 3: Confusion Matrix of Payload

The results in figure 2 shows that with Yolov5 model, payload attached to drones at various distances can be visually identified which determines the anti-drone's adaptive response strategy. However, as seen in figure 3, there are still misidentification of payloads due to object concealment which could result in wrong response. Hence, model improvement is necessary.

Ⅲ. Conclusion

Maintaining sanity and security in airspace because of drone usage influx remains a daunting challenge. This paper emphasized the need for responsive AI-driven real-time identification and interception of illegal drones to guarantee environmental safety. Unravelling payload content is crucial for accurate visual payload identification. More research efforts should be made in this direction.

ACKNOWLEDGMENT

This research work was supported by Priority Research Centers Program through NRF funded by MEST (2018R1A6A1A03024003) and the Grand Information Technology Research Center support program (IITP-2021-2020-0-01612) supervised by the IITP by MSIT, Korea.

References

[1] Chen H., Wang Z., and ZhangL," Collaborative Spectrum Sensing for Illegal Drone detection: A Deep-Learning-Based Image classification Perspective," China Communications 17(2), (2020) 81-92

[2] Haviv E. Elbit, "Drone Threat and CUAS Technology: WhitePaper, Elbit Systems 1 (01), 2019, 1-19

[3] Seongjoon P, Hyeong T.K, Sangmin L. Hyeontae Joo, and Hwangnam K, "Survey on Anti-Drone Systems: Components, designs, and Challenges", IEEE Access 9(2021), 42635-42659

[4] Seung-Hwan Kim, Jae-Woo Kim, Dong-Seong Kim, "Energy Consumption Analysis of Beamforming and Cooperative Schemes for Aircraft Wireless Sensor Networks" Applied Sciences, vol.10, no.12, pp.4374-4391, June 2020

[5] R. Akter, V. -S. Doan, G. B. Tunze, J. -M. Lee and D. -S. Kim, "RF-Based UAV Surveillance System: A Sequential Convolution Neural Networks Approach," 2020 International Conference on Information and Communication Technology Convergence (ICTC), 2020, pp. 555-558

Transformer 를 이용한 Image Captioning

김동훈, Xu Jiajie, 심병효*
서울대학교

dhkim@islab.snu.ac.kr, jiajiexu@islab.snu.ac.kr, *bshim@islab.snu.ac.kr

A Study on the image captioning using transformer

Donghoon Kim, Jiajie Xu, Byonghyo Shim*
Seoul National University

요 약

본 논문은 vision 과 language domain 의 정보를 동시에 학습하는 image captioning task 에 대해서 Transformer 를 적용하는 기법에 대해서 다룬다. Vision 과 language domain 을 동시에 학습하는 multi-modality 학습 기법은 image captioning task 뿐만 아니라 text-to-image retrieval, video question and answering, video moment retrieval 등 다양한 영역에서 사용된다. 그런데 수많은 task 들 중에서도 image captioning task 는 이러한 수많은 multi-modality 중에서도 두가지 도메인에 신경망의 이해도를 측정하는 데에 있어 가장 많은 주목을 받고 있다. 이는 현 시점의 많은 연구들이 image captioning task 를 pretrain task 를 이용하고 down stream task 로 다양한 task 를 가능하게 하고 있다는 점에서 확인 가능하다. 본 논문은 이러한 image captioning task 에 대해서 transformer 구조를 활용하여 coco dataset 에 대해서 scratch 로부터 end-to-end 로 학습시켜 image captioning task 를 수행 해 보았다.

Ⅰ. 서 론

Vision domain 과 language domain 은 기계 학습 분야에 있어서 가장 큰 축들 중에 하나이다. Vision 도메인의 경우 CNN 모델을 통한 image classification, detection 그리고 segmentation 과 같은 다양한 ask 에서의 괄목할 만한 성과를 이루어 왔다. Language domain 역시 최근 Transformer 를 이용한 BERT[1], GPT[2], 그리고 다양한 파생 모델들이 탄생하고 translation, text generation, text sentimental classification, grammatical error classification, 그리고 question-answering 등의 task 들에서 human level 을 뛰어 넘는 놀라운 성과를 보이고 있다. 특히 OpenAI 를 통해서 발표된 GPT3[3]와 같은 초대형 모델이 앞에서 언급된 수많은 task 들에 대해서 zero-shot 성능이 State-of-the-art(SOTA)를 달성하는 것을 보여주며 natural language domain 에서 기계학습의 무한한 가능성을 보여주고 있다.

Vision 과 language domain 각각에서의 놀라운 성과에도 불구하고 두 domain 을 통합시키는 multi-modality learning 은 그만한 관심을 받지 못했다. Multi-modality learning 이란 인간의 다인지 학습 능력을 모방하여 다양한 서로 다른 domain 의 신호를 통합하여 처리하는 학습 방법을 의미한다. video question answering 이나 image captioning 과 같이 두가지 서로 다른 domain 의 data 를 모두 사용하는 task 들의 경우 두 domain 의 통합이 반드시 필요하다. 뿐만 아니라 Multi-modality learning 은 각 domain 의 단점을 보완해 주고 장점을 부각시키면서 성능을 향상시킨다. 예를 들어 image classification 에서 language representation 을 이용하여 학습을 시킴으로써 zero-shot capability 를 향상시키는[4] 등의 domain specific 한 task 에서의 다양한 문헌이 발표되고 있다.

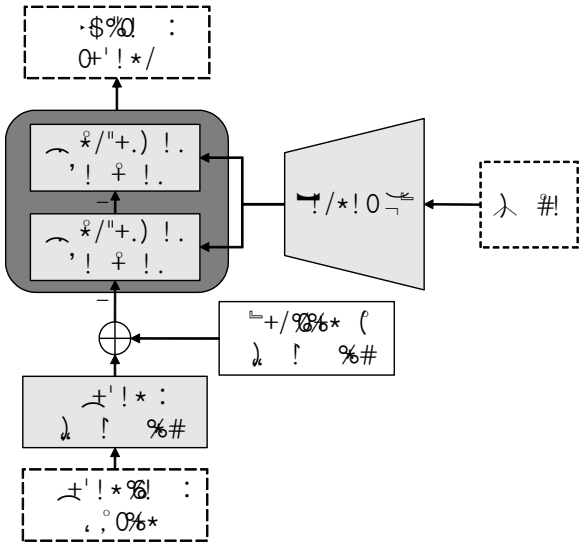


그림 1 모델 구조

이렇게 multi-modality 의 중요성이 부각되고 있는 가운데 해당 분야에서 가장 활발히 연구되고 있는 분야는 vision 과 language 의 통합 학습에 대한 분야이다. Vision 과 language 통합의 대표적인 task 들에는 vision 과 language domain 을 통합하여 수행하는 image captioning, text-to-image retrieval, video question and answering, video moment retrieval 등이 있다. 이러한 수많은 task 들 중 최근 발표되는

연구에서는 language model 에서 영감을 받아 image caption generation[6]과 masked token prediction [6, 8]을 pretrain 으로 활용하고, 나머지 task 들을 downstream task 로 fine-tuning 한다.

본 연구에서는 vision-and-language learning 에서 널리 pretrain task 로 이용되고 있는 image captioning 을 위해 transformer 와 resnet-50 을 scratch 부터 end-to-end 로 학습시켜 image captioning task 를 수행했다.

Ⅱ. 본론

본 연구에서 사용한 모델의 구조는 그림 1 과 같다. Transformer decoder 의 경우 original transformer[9]의 decoder 를 그대로 사용하였다. Resnet-50 도 마찬가지로 기존 model 의 구조를 그대로 사용했다. Tokenizer 의 경우 Byte Pair Encoding (BPE)를 통해 COCO caption 에서 추출한 10K 사이즈의 corpus 를 생성하여 이용하였다.

우선 caption 과 image 입력을 받으면, caption 은 tokenizer 로 token 들로 변환을 한다. 이때 캡션 token 맨 앞에 [sos] 토큰을 붙인다. 그리고 hidden size 가 1024 인 token embedding 으로 변환하고 같은 사이즈의 learnable parameter 인 positional embedding 을 더해주어 Transformer Decoder 에 Query 로 입력한다. Image 의 경우 높이와 너비를 모두 224 로 Crop 한다. 그리고 Resnet-50 을 통과해서 생성된 7*7*2048 (H*W*C)의 feature 를 channel 크기가 1024 가 되도록 Linear projection 시켜준 뒤, 49 개의 feature pixel 데이터들을 Transformer Decoder 의 Key 및 Value 로 입력한다. Output 은 shifted token 을 label 로 하여 Cross-entropy 를 이용하여 학습한다.

batch size 의 경우 64 로 하고 270 epoch 동안 학습시켰다. 학습에 사용된 optimizer 는 Stochastic Gradient Decent (SGD) 이다. momentum 은 0.9 를 weight decay 는 0.0001 이다. Learning rates 은 Resnet 의 경우 0.2 를 Transformer 의 경우에는 0.001 을 사용했다. 둘다 10 epoch 동안을 warm up step 으로 하는 cosine learning rate decay 를 이용했다.

이렇게 학습된 모델의 metric 은 CIDER score 와 SPICE score 를 통해 확인했는데, 결과는 표 1 과 같다.

Metric name	Score
CIDER	93.7
SPICE	18.0

표 1 Image captioning metric

Ⅲ. 결론

본논문에서는 Transformer 를 이용하여 scratch 부터 end-to-end 로 학습시킨 image captioning 을 확인 해 보았다. 입력 이미지에 대한 출력 caption 은 그림 2 와 같이 확인 가능하며 표 1 에서 나온 성능을 통해 사람이 labeling 한 caption 과 구분이 거의 불가능 함을 알 수 있다. 추후에는 model 구조를 변경하고 image 단에서의 loss 를 추가하여 학습을 시켜 성능을 향상시켜 볼 예정이다.

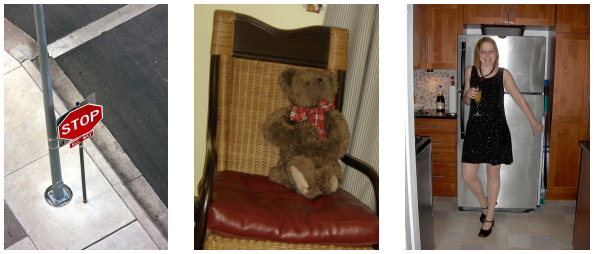


그림 2 생성된 캡션 예시

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT 연구센터육성지원사업의 연구결과로 수행되었음. (IITP-2021-2017-0-01637)

참 고 문 헌

[1] Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).

[2] Radford, Alec, et al. "Improving language understanding by generative pre-training." (2018).

[3] Brown, Tom B., et al. "Language models are few-shot learners." arXiv preprint arXiv:2005.14165 (2020).

[4] Radford, Alec, et al. "Learning transferable visual models from natural language supervision." arXiv preprint arXiv:2103.00020 (2021).

[5] Liu, Yang, et al. "Use what you have: Video retrieval using representations from collaborative experts." arXiv preprint arXiv:1907.13487 (2019).

[6] Lei, Jie, et al. "Less is more: Clipbert for video-and-language learning via sparse sampling." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.

[7] Desai, Karan, and Justin Johnson. "Virtex: Learning visual representations from textual annotations." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.

[8] Li, Xiujun, et al. "Oscar: Object-semantics aligned pre-training for vision-language tasks." European Conference on Computer Vision. Springer, Cham, 2020.

[9] Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems. 2017.

HSV 색공간을 기반한 사람 및 차량 객체의 실시간 색상 추출

장시예
마크애니

syjang@markany.com

Real-time Color Extraction of People and Vehicles Based on HSV Color Space

Jang Si Ye
MarkAny

요 약

본 논문에서는 CCTV 영상에서 검출된 사람 및 차량 객체의 색상을 실시간으로 추출하는 알고리즘을 제안한다. 기존 방식은 객체 전면부 이미지를 대상으로 색상 추출을 하기 때문에 CCTV 화각 이미지에는 적용하기 어렵다. 그리고 다수의 CCTV 영상에서 색상을 추출하기 위해서는 실시간처리가 가능해야한다. 이러한 문제를 해결하기 위해 본 논문에서는 객체 크기 정보를 이용하여 색상 추출 영역과 대표 픽셀을 선정하는 연구를 진행하였다. 또한 조도 및 휘도 변화에 강인한 HSV 색공간을 사용하여 정확도를 향상시켰다. 제안하는 알고리즘을 적용한 결과 주야간, 악천후 등의 다양한 환경의 CCTV 이미지 3,080 개에서 각각 차량 93.42%, 사람 77.32%의 정확도를 달성하였다.

Ⅰ. 서 론

최근 범죄예방 및 사회질서 유지를 목적으로 공공기관의 CCTV 가 점차 증가하고 있고, 그림 1 과 같이 2020 년의 총 CCTV 수는 1,336,653 대로 2019 년 대비 16.4% 늘어난 수준이다[1].

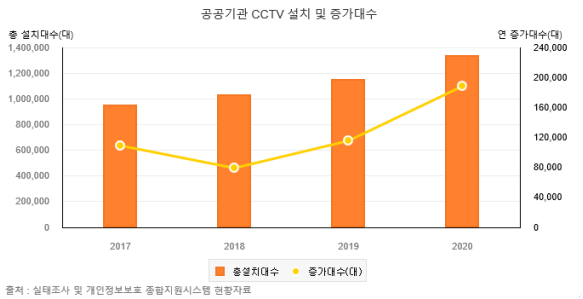


그림 1. 공공기관 CCTV 설치 및 증가대수 [1]

통합관제센터에서는 관제요원 1 명당 272 대의 CCTV 를 관제해야하는 실정으로[2], 늘어나는 CCTV 관제에 어려움이 있다. 따라서 이와 같은 문제를 해결하기 위해 실시간 모니터링이 가능한 지능형 관제 시스템을 위한 연구 및 개발이 진행될 필요가 있다.

기존 연구는 객체의 색상 추출을 위하여 차량의 전면부 이미지를 사용한다[3]. 해당 방법의 경우 추출 관심 영역을 찾아 정확도를 높인다는 장점이 있지만, 객체의 특정 각도에서 색상을 추출하기에 다양한 각도에서 관찰되는 CCTV 의 객체에 적용시키는 것은 어렵다. 다른 방법으로는 해리스 코너 검출을 이용해 색상 확률 지도를 만들어 대표 색상을 추출하는 방법이 있다[4]. 이 방법은 96.1%의 높은 정확도를 갖지만 멀티스케일 탐지 방식을 사용하므로 초당 9-10 개의 이미지를 처리하기에 실시간성이 필요한 관제 시스템에 적합하지 않다.

따라서 본 연구는 CCTV 관제 환경에서 필요한 실시간성을 보장하고 각도 및 거리에 강인한 색상 추출 알고리즘을 제안하고자 한다.

Ⅱ. 본 론

CCTV 에서 탐지되는 객체는 30° ~90° 의 화각으로 관찰되기 때문에 카메라에 다가올수록 객체 하단의 비율이 커지게 된다. 사람은 팔, 다리 등의 관절에 의해 다양한 포즈를 가지게 되므로 몸통 부분에서 색상을 추출하는 것이 오탐의 가능성이 적다. 차량은 창문 영역이 많은 비중을 차지하는 특성이 있기 때문에 창문 영역을 제외할 필요가 있다. 본 알고리즘에서는 이러한 화각과 객체별 특성을 고려하여 색상 추출 영역 및 대표 픽셀을 선정하고자 한다.

사람 객체의 경우, 색상 추출 영역은 객체의 너비와 높이를 고려하여 몸체를 가장 정확하게 잡을 수는 범위로 설정한다. 그림 2 와 같이 상의는 객체 높이의 1/4~1/2 지점과 너비의 1/4~3/4 지점을 목 아래 몸통 영역으로 지정하고, 하체는 객체 높이의 3/5~4/5 지점과 너비의 1/4~3/4 지점을 사타구니 및 허벅지 영역으로 지정한다.



그림 2. 사람 색상 추출 영역 예시

차량 객체의 경우, 객체가 박스 영역 안에 꽉 차기 때문에 색상 추출 영역을 따로 설정하지 않는다. 차량은 창문이 오탐의 주 원인인데, 그 모양이 객체의 각도에 따라 일정치 않기 때문에 영역 설정이 어렵다. 차선책으로 차량 객체에서 추출된 검은색의 일정 부분은 창문이라고 간주하고 검은색의 최종 점수에서 일정 수치를 제거하는 방식을 사용한다.

영역을 설정한 후에는 속도 향상을 위해, 지정 간격을 건너뛰며 대표 픽셀을 추출한다. 객체의 크기에 따라 건너뛸 픽셀의 수를 설정하며, 이것을 픽셀 선정 간격이라고 칭한다. 픽셀 선정 간격은 아래와 같은 공식(수식 1)을 따르며, x 는 객체의 너비를 의미한다.

$$f(x)=\begin{cases}3 \ if \ \ 0\le x\le 30\\ 5 \ if \ 30<x\le 50\\ 7 \ if \ 50<x\end{cases}$$

수식 1. 대표 픽셀 선정 간격

제안하는 알고리즘에서는 HSV 색상 모델을 사용한다. HSV 모델은 색을 표현하는 방법으로 색상, 채도, 밝기로 구분되는 3 차원 색상 공간이다. 해당 모델은 색상과 명암 요소가 분리됨으로써 흔히 사용되는 RGB 모델보다 조도 및 휘도에 강인하며, 인간이 인지할 수 있는 색상 스펙트럼을 가장 잘 표현하고 있기에 채택한다 [5].

$$V=\max(B,G,R)$$
$$S=\begin{cases}\dfrac{V-\min(B,G,R)}{V} & (V\neq 0)\\ 0 & (V=0)\end{cases}$$
$$H=\begin{cases}60\left(\dfrac{G-B}{V-\min(B,G,R)}\right) & if \ V=R\\ 60\left(\dfrac{B-R}{V-\min(B,G,R)}+2\right) & if \ V=G\\ 60\left(\dfrac{R-G}{V-\min(B,G,R)}+4\right) & if \ V=B\\ if \ H<0 \ then \ H=H+360\end{cases}$$

수식 2. RGB to HSV color converting

선정된 대표 픽셀은 변환 공식(수식 2)을 사용하여 색상 공간을 변환시킨다. 전체 픽셀을 변환시키지 않고, 대표 픽셀을 선정하여 필요한 부분만 변환시키는 것은 계산량을 줄여 색상 추출 속도를 향상시킬 수 있는 방안이다. 각 요소의 범위는 H[0-180], S[0-255], V[0-255]로 설정하여 사용한다.

이렇게 추출된 HSV 값을 통해 색상의 범위를 나눈다. 색상은 Axalta 의 아시아 선호 색상 통계[6]를 참고하여 유채색은 초록색, 빨간색, 파란색, 노란색으로, 무채색은 검정색, 흰색으로 나눈다. 무채색은 육안으로 구분이 명확하지 않은 회색을 제외하고 검정색과 흰색으로만 분류한다.

수식 3 과 같이 빨간색, 파란색, 초록색, 노란색은 색상(H)의 범위를 서로 겹치지 않게 분할하였고, 조도차와 빛반사를 고려하여 채도(S)와 명도(V)의 범위를 지정해주었다. 흰색과 검은색은 전범위의 색상(H)값을 가지되, 채도(S)와 명도(V)를 양극단의 범위로 지정하였다.

blue	92	≤	H	≤	130,	90	≤	S	≤	255,	90	≤	V	≤	255
green	34	≤	H	≤	91,	50	≤	S	≤	255,	60	≤	V	≤	255
yellow	16	≤	H	≤	33,	50	≤	S	≤	255,	50	≤	V	≤	255
red(1)	0	≤	H	≤	15,	80	≤	S	≤	255,	60	≤	V	≤	255
red(2)	131	≤	H	≤	180,	80	≤	S	≤	255,	60	≤	V	≤	255
white	0	≤	H	≤	180,	0	≤	S	≤	30,	190	≤	V	≤	255
black	0	≤	H	≤	180,	0	≤	S	≤	255,	0	≤	V	≤	50

수식 3. 색깔 별 HSV 값의 범위

수식 3 의 범위에 따라 대표 픽셀들을 대상으로 색상 히스토그램을 생성한다. 이후 가장 빈도수가 높은 색상이 객체의 최종 색상으로 판정된다.

Ⅲ. 성능평가

자체 제작한 그림 3 와 같은 CCTV 데이터 셋 Ground Truth(이하 GT)로 평가한다. 다양한 크기, 각도, 색상, 종류 등의 객체 외형적 요소와 다양한 배경 및 조도, 날씨 등의 환경적인 요소를 고려하여 차량 2,174 장, 사람 906 장(상하의 1,812 개)의 이미지를 선정하였다. 컴퓨터의 사양은 Window10, RAM 16G, AMD Ryzen 7 1700 Eight-Core Processor 이며 GeForce GTX 1060 1 개를 사용하였다.



그림 3. 성능평가에 사용한 이미지와 그 결과

사람 객체의 정확도는 77.32%이며, 초 당 10,394 객체를 처리할 수 있고, 차량 객체의 정확도는 93.42%이며, 초 당 3,485 객체를 처리할 수 있다. 차량의 평균 속도가 더 느린 이유는 GT 로 사용된 차량 객체의 평균 해상도(238x187)가 사람 객체(74x184)보다 크기 때문이다.

Ⅳ. 결 론

본 논문에서는 CCTV 방범 뷰에서 사람과 차량 객체의 HSV 색상 판별 알고리즘에 대해 탐구했다. 사람과 차량의 각도, 거리, 움직임 등의 특성을 파악하여 최적의 색상 추출 영역을 지정하였으며, 객체의 크기에 따라 대표 픽셀을 선정하였다. 또한 색 지정을 위해서 조도와 휘도에 강인한 HSV 색상 공간을 사용하였으며, 선정된 픽셀을 대상으로 픽셀 단위의 색공간 변환을 통해 알고리즘의 속도 향상을 도모했다.

제안한 알고리즘을 적용했을 때 평균 약 85.37%의 색상 추출 정확도와 1 초당 약 6,940 개의 처리 속도를 달성했다. 실시간성 및 각도와 거리에 강인하여 실시간 지능형 관제 시스템에서 색상 추출을 수행 할 수 있다.

차후 연구 주제로는 사람 상하의의 정확도를 높이기 위한 방법으로 이중 색상, 장애물, 배경색 등의 노이즈를 정밀하게 제거하는 것이 있다.

참 고 문 헌

[1] e-나라지표 : 공공기관 CCTV 설치 및 운영, [Online], Available:https://www.index.go.kr/potal/main/EachDtlPageDetail.do?idx_cd=2855

[2] 국회입법조사처, “CCTV 통합관제센터 운영실태 및 개선방안”, NARS 입법정책보고서 제 29 호, pp. 32, 2019

[3] 곽민호, 이효종, “차량의 전면부를 이용한 색상 인식 알고리즘 개발”, 대한전자공학회 학술대회, pp. 856-859, July, 2013

[4] 김광주, “Vehicle Detection and Color Recognition using Deep Learning for Intelligent Traffic Surveillance Systems”, 국내박사학위논문 경북대학교 대학원, 2020

[5] Kim, Tae-Hun, Gi-Seung Kim, Young-Chul Song, Gang-Soo Ryu, Byung-Jae Choi, and Kil-Houm Park. “A Color-Based Medicine Bottle Classification Method Robust to Illumination Variations”, Journal of Korean Institute of Intelligent Systems 23, no. 1, pp. 57- 64, Feb, 2013

[6] Axalta Coating Systems, “Global Automotive 2020 Color Popularity Report”, 2020

특징 피라미드 네트워크의 특징 압축을 위한 레벨 간 특징 예측 방법

김민섭, 이종석, 심동규*
광운대학교
{minsub20, suk2080, *dgsim}@kw.ac.kr

Inter-level Feature Prediction Network for Feature Pyramid Network Feature Compression

Kim Min Sub, Lee Jong Seok, Sim Dong Gyu*
Kwangwoon Univ.

요 약

본 논문은 FPN(Feature Pyramid Network)기반 네트워크에서 추출된 특징맵을 압축하는 구조를 제안한다. FPN구조의 네트워크는 해상도별 특징맵이 추출되기 때문에 일반적으로 영상에 비해 데이터가 크기 때문에, 영상에 대한 압축에 비해 더 높은 압축율을 요구한다. 이에 본 논문은 모든 해상도의 특징맵 전체를 압축하지 않고, 심층 신경망을 통해, 낮은 해상도의 특징맵으로 높은 해상도 특징맵을 예측한 뒤 압축을 진행하는 방법을 제안한다. 제안한 방법의 결과는 압축해야 할 대상의 데이터양은 75% 감소함과 동시에, 객체 검출 성능(mAP)은 모든 특징맵을 보내는 방법 대비, 0.2% 감소했다.

Ⅰ. 서 론

최근 이미지, 비디오에서 객체 검출, 객체 추적, 객체 영역 분할을 수행하기 위한 심층 신경망에는 FPN[1]구조가 널리 사용되고 있다. FPN의 기본 구조는 그림 1 과 같이 영상에서 해상도별 특징맵을 추출하는 backbone을 포함한 FPN부분과 특징맵을 입력으로 넣어 객체 검출을 수행하는 detector head 부분으로 이루어져 있고, FPN의 기본 구조를 기반으로 한 객체 검출 네트워크들은 추가적인 변화를 주어 성능 향상을 위한 연구들 진행되고 있다.[2][3][4]

다른 한편으로는 심층 신경망들을 모바일 또는 에지 디바이스에서 사용하려는 시도가 있다.[5] 하지만 모바일이나 에지 디바이스에서 사용하기에는 객체 검출 신경망이 너무 많은 계산량을 필요로 한다는 문제점이 있다. 이를 해결하기 위해서 FPN 과정 까지는 높은 컴퓨팅 파워를 가지는 서버를 사용하고, FPN에서 출력된 특징맵을 에지 디바이스에 전달하여 객체 위치나 크기와 같은 객체 탐지 결과를 에지 디바이스에서 추론하는 방법에 대한 연구가 진행되고 있다. [6]

하지만, FPN에서 나온 결과물인 특징맵은 일반적으로 입력 이미지보다 더 많은 채널 수를 가지며, FPN구조상 다수개의 특징맵을 출력하기 때문에 입력대비 많은 데이터양을 가지고 있다. 큰 데이터양는 통신 네트워크 환경 부하가 생겨, 지연과 같은 문제점이 발생할 수있다. 따라서, 본 논문에서는 FPN에서 출력 된 다수개의 특징맵의 데이터양을 감소시키기 위하여 낮은 해상도의 특징맵을 이용하여 높은 해상도의 특징맵을 예측하는, FPN의 특징 압축을 위한 레벨 간 특징 예측 방법을 제안한다. 제안하는 방법은 높은 해상도의 특징맵을 학습된 신경망을 통해 예측하여 생성함으로써 해당 해상도의 특징맵에 대한 전송을 생략하여 전체적으로 전송되는 데이터양을 75% 감소시킬 수 있으며, 객체 검출 성능은 모든 해상도의 특징맵을 보냈을 경우 대비 0.2% 감소했다. 본 논문의 구성은 다음과 같다. 2장에서는 본 논문에서 제안한 방법에 대해서 자세히 설명하고, 3장에서는 제안하는 방법의 성능을 평가하고 결론을 맺는다.

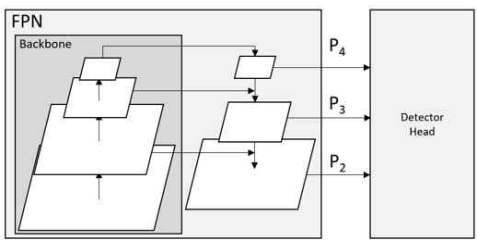


그림 1. FPN기반 객체 검출 네트워크의 기본구조

Ⅱ. 본론

본 논문에서는 FPN구조의 네트워크에서 출력 된 특징맵 중, 하위 해상도의 특징맵을 입력으로 상위 해상도의 특징맵을 예측한 뒤 압축을 진행하는 방법을 제안한다. 본 논문에서 사용한 FPN의 구조의 네트워크에서 출력된 특징맵은 P2, P3, P4, P5와 P5를 max pooling으로 down sampling한 P6로 이루어져있다. 그림 2는 본 논문에서 제안하는 구조로, Encoder와 Decoder로 구성되어있다. Encoder는 FPN에서 출력 된 특징맵에서 가장 높은 해상도의 특징맵인 P2를 제외한 P3, P4, P5, P6를 압축해서 Decoder로 전송한다. Decoder는 전송받은 압축된 특징맵을 복원 후, 가장 낮은 해상도 특징맵은 P′6를 제외한 P′3, P′4, P′5를 본 논문에서 제안한 네트워크에 입력하여 예측된 P2인 P′2를 생성한다. 마지막으로, P2′, P3′, P′4 ,P′5, P′6를 detector head에 입력하여 객체 검출을 수행한다. 제안한 네트워크가 P2를 예측하는 과정은 다음과 같다. 입력받은 특징맵들을 P2의 해상도(H,W,C)에 맞춰 보간 한 후, 채널 축 방향으로 이어붙인다. 이때 보간방법은 최단입점 보간법(Nearest neighbor interpolation)을 사용한다. 이후, 합성곱 층 과 GDN(Generalized Divisive Normalization)[7] 층을 통해 예측한 P2′가 출력된다. GDN의 하이퍼 파라미터로 beta를 1e-6, gamma를 0.1, offset을 2의 -18승을 사용했다.

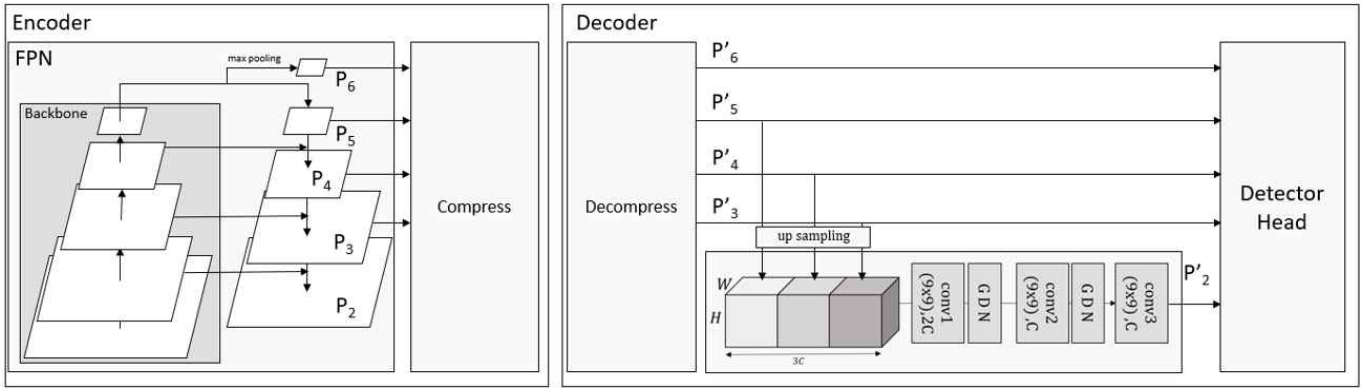


그림 2. 제안하는 특징맵 압축 구조

III. 실험환경 및 실험결과

본 논문에서 제안하는 방법의 성능을 평가하기 위하여 제안하는 특징 피라미드 네트워크의 특징 압축을 위한 레벨 간 특징 예측 방법을 python 3.7.9기반의 pytorch 1.7.1 딥러닝 프레임 워크를 이용하여 구현했다. 제안하는 방법의 객관적 압축 성능 평가를 위해 FPN 에서 나온 특징맵을 선형 양자화 및 SBMP 기반 CABAC[8]을 사용하여 압축한 뒤, 압축된 데이터의 크기 및 복원된 특징맵을 사용한 객체 검출 정확도를 모든 특징맵을 보내는 경우(*Anchor*), 그리고 P3를 P2의 해상도로 최단입점 보간법을 적용한 것을 P2로 사용하는 경우(*Nearest neighbor interpolation*)와 비교했다. 객체 검출 정확도는 *AP-S, AP-M, AP-L,mAP*에 대해서 측정하며 이는 각각 소형, 중형, 대형에 대한 객체 검출 성능과 평균적인 객체 검출 성능을 의미한다. 객체 검출 정확도 비교를 위한 네트워크는 detectron2[9] 를 사용하여 구현했다. 구현한 네트워크는 ResNeXt-101(32x8d)[10]을 FPN의 backbone으로 사용했으며, 특징맵을 입력받아 객체 검출을 수행하는 detector head는 Faster R-CNN[11]의 detector head 부분을 사용했다.

제안하는 네트워크를 학습시키기 위하여 COCO data set[12]의 train2017에서 10000장을 training set으로 사용하고, val2017에서 5000장을 validation set으로 사용했다. 성능 비교를 위한 Test set으로는 train2017에서 training set을 제외한 5000장을 사용했다.

표 1. 제안한 방법의 성능 비교

	data size(bit)	*mAP(%)*	*AP-S(%)*	*AP-M(%)*	*AP-L(%)*
Anchor	3773 k	57.595	37.755	64.244	73.022
Nearest neighbor interpolation	848 k	52.073	25.242	63.272	73.018
Proposed	848 k	57.387	36.429	63.961	73.023

표1에서 제안한 방법은 모든 특징맵을 보내는 방법(*Anchor*) 대비 75% 정도의 bit 감소를 보임과 동시에, 객체 검출 성능(*mAP*(mean Average Precision))은 0.2% 하락했다. 특히, 소형 물체 탐지 성능을 나타내는 *AP-S*에서의 성능 하락 정도가 크게 줄었음을 확인 할 수 있다.

IV. 결론

본 논문에서는 FPN 구조의 네트워크에서 하위 해상도의 특징맵을 입력으로 상위 해상도의 특징맵을 예측한 뒤 압축을 진행하는 방법을 제안했다. 본 논문에서 제안한 방법은 압축 할 대상의 총 데이터양을 75% 감소시킴과 동시에, 객체 검출 성능(mAP)은 모든 특징맵을 보내는 방법 대비, 0.2% 감소했다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT 연구센터지업사원(IITP-2021-2016-0-00288)의 연구결과로 수행되었음

참 고 문 헌

[1] Lin, T. Y. "Feature Pyramid Networks for Object Detection," Proc.CVPR, pp. 2117-2125, Jul. 2017.

[2] Ghiasi G. "Nas-fpn: Learning scalable feature pyramid architecture for object detection," Proc.CVPR, pp. 7036-7045, Jun. 2019.

[3] Li, Zeming "Detnet: A backbone network for object detection," arXiv preprint arXiv:1804.06215, 2018.

[4] Gong, Y. "Effective fusion factor in FPN for tiny object detection," Proc.CVPR, pp. 1160-1168, Jun. 2021.

[5] Robert J. Wang "PELEE: A Real-Time Object Detection System on Mobile Devices," arXiv preprint arXiv:1804.06882, 2018.

[6] Ren, J. "Distributed and efficient object detection in edge computing: Challenges and solutions," IEEE Network 32.6, pp. 137-143, 2018.

[7] Balle J. "End-to-End Optimized Image Compression," Proc.ICLR, Apr. 2017.

[8] Haase P. "State-based multi-parameter probability estimation for context-based adaptive binary arithmetic coding," Proc.DCC, pp. 163-172, Mar. 2020.

[9] Wu Y. "Detectron2," 2019, (https://github.com/facebookresearch/detectron2).

[10] Saining, X. "Aggregated Residual Transformations for Deep Neural Networks," Proc.CVPR, pp. 1492-1500, Jul. 2017.

[11] Shaoqing, R. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," NIPS, pp. 91-99, Dec. 2015.

[12] Lin, T. Y. "Microsoft COCO: Common Objects in Context," ECCV, pp. 740-755, 2014.

초분광 데이터 분석용 인공지능 플랫폼 연구

김거식, 김계은, 김정은, 강현서
한국전자통신연구원 광ICT융합연구실
keosikis@etri.re.kr, kyeeun_kim@etri.re.kr, j.kim@etri.re.kr, hskang87@etri.re.kr

A Study on artificial intelligence platform for analyzing hyperspectral data

Keo Sik Kim, Kyeeun Kim, Jeong Eun Kim, Hyun Seo Kang
Electronics and Telecommunications Research Institute

요 약

본 논문은 초분광 데이터 분석을 위한 인공지능 플랫폼에 대한 것으로, 보다 자세하게는 머신러닝/딥러닝 알고리즘을 이용하여 초분광 데이터를 분석하는데 필요한 전처리, 특징 추출, 분류기 학습, 학습데이터 구축 등의 기능이 통합 구현된 소프트웨어 플랫폼에 관한 연구이다. 인공지능 플랫폼은 Python 언어를 기반으로 구현되었으며, 각 기능은 레고블록 방식의 모듈 결합을 통해 동작하게 함으로써 사용자의 기능 확장이 용이하도록 설계하였다.

Ⅰ. 서 론

초분광 이미징은 대상체의 분광 정보와 공간 정보를 동시에 획득하는 기술로, 대상체의 광 특성을 이용하여 상태, 구성, 변이 등에 대한 정보를 이미징한다. 이런 장점을 활용하여 최근에는 식품, 의료/바이오, 환경, 제조 등 다양한 분야에 적용이 시도되고 있다[1]. 초분광 데이터 분석 방법으로는 신경망(neural network), 서포트 벡터 머신(Support Vector Machine, SVM), 군집화(clustering)와 같은 전통적인 머신러닝 기반 알고리즘이 주로 사용되었으나, 최근에는 컴퓨팅 성능 향상과 분석 기술 발전으로 인해 CNN(Convolutional Neural Network), 오토인코더(Auto Encoder, AE), 적대적 생성 네트워크(Generative Adversarial network, GAN)와 같은 딥러닝 기반 학습 알고리즘도 연구되고 있다[2]. 본 논문에서는 현재 개발 중인 인공지능 알고리즘 적용 초분광 데이터 분석 플랫폼의 소스코드 구조 설계와 실제 적용 사례에 대해 서술한다.

Ⅱ. 초분광 데이터 분석 플랫폼 설계

초분광 데이터 분석 플랫폼은 Python 언어(ver. 3.8)를 기반으로 TensorFlow(ver. 2.4.1), opencv-python(ver. 4.5.1), scikit-learn(ver. 0.24.1), spectral python(ver. 0.22.1) 등의 오픈 소스 라이브러리를 이용하여 구현되었다. 초분광 데이터는 ENVI(ENvironment for Visual Images) 파일 포맷 중 BSQ(Band Sequential) 방식에 따라 저장되며, 데이터 구조는 [columns × rows × channels]이다. 초분광 데이터는 헤더 파일(*.hdr)과 데이터 파일(*.raw)로 각각 저장된다. 데이터 파일에는 3차원 초분광 원본 데이터(cube)가 저장되고, 헤더 파일에는 다음과 같은 메타정보가 저장된다.

- acquisition time: 초분광 데이터 측정 시각
- bands: 분광 밴드수(분광해상도)
- pixels: 라인 당 픽셀수(라인축 공간해상도)
- lines: 초분광 획득 시 라인스캔 횟수(스캔축 공간해상도)
- integration time: 라인스캔 당 노출시간
- DMD_lines/scan: 라인스캔 당 노출된 DMD 라인수
- wavelength: 분광 파장

그림 1은 플랫폼의 프레임워크이다. 플랫폼은 크게 전처리, 분류, 분할/검출 기능이 있으며, 초분광 데이터의 픽셀 단위 분석 알고리즘을 중심으로 구성된다. 전처리 기능은 초분광 데이터 특징 추출 및 차원 감소(decomposition)를 위한 PCA(Principle Component Analysis), LDA(Linear Discriminant Analysis), NCA(Neighborhood Components Analysis)가 구현되어 있으며, 샘플 데이터 정규화를 위한 MSC(Multiplicative Scatter Correction), SNV(Standard Normal Variate)도 구현되어 있으며, 향후 t-SNE(t-Distributed Stochastic Neighbor Embedding)가 추가될 예정이다. 분류 알고리즘은 지도 학습(supervised learning)인 심층 신경망(Deep Neural Network, DNN)과 SVM 분류기가 구현되어 있으며, 향후 K-means clustering, GAN과 같은 unsupervised learning 알고리즘과 CNN과 같은 딥러닝 알고리즘도 추가 구현할 예정이다. 분할/검출 알고리즘은 향후 Faster R-CNN, YOLO, SSD, RFCN과 같은 다양한 딥러닝 알고리즘을 탑재할 예정이다.

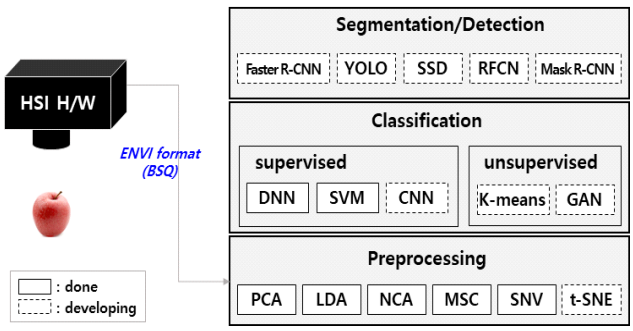


그림 1. 초분광 데이터 분석 플랫폼의 프레임워크

그림 2는 초분광 데이터 분석 플랫폼의 Pseudo code이다. 플랫폼의 주요 기능은 초분광 학습데이터를 이용한 인공지능 모델 학습('train')/재학습('retrain') 및 추론 결과의 시각화('inference')이다. 각 기능은 세부 동작들이 구현된 하부 모듈들을 레고블록 방식의 조합에 의해 최종 구현된다. 학습 과정에서는 학습데이터셋 생성(make_dataset), 모델 학습(train_model), 시각화(show_result) 블록의 조합으로 구현된다. 재학습 과정에서는 기 학습모델 및 데이터셋 갱신(prepare_for_retrain)

블록이 추가된다. 추론 결과 시각화 과정에서는 학습데이터셋 생성(make_dataset)과 시각화(show_result) 블록의 조합으로 구현된다.

```
Input: cube, data_info, mode, cla_mode, prepg_mode

CASE mode OF
          'train'     : sequence_1
          'retrain'   : sequence_2
          'inference': sequence_3
ENDCASE


def sequence_1:
          make_dataset
          train_model
          show_result

def sequence_2:
          make_dataset
          prepare_for_retrain
          train_model
          show_result

def sequence_3:
          make_dataset
          show_result


def make_dataset:
          load cube
          get data from selected area using data_info
          split them to train/valid/test dataset
return dataset

def train_model:
          preprocess dataset by prep_mode
          train a model by cla_mode
return model, prep_model

def prepare_for_retrain:
          load pre-stored dataset, model, prep_model
          update dataset (create/append/replace)
return dataset, model, prep_model

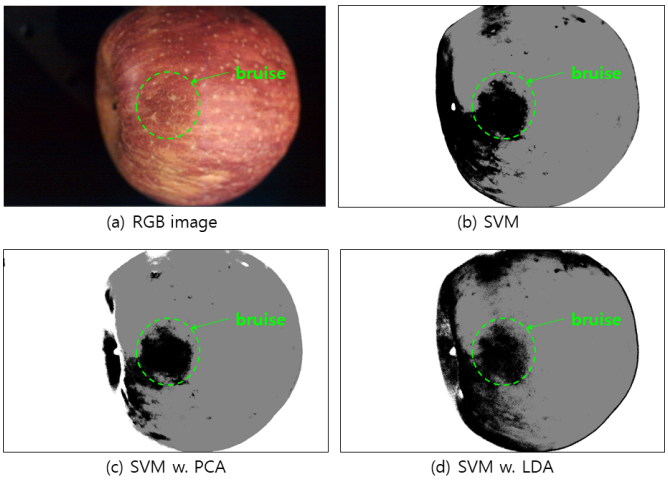
def show_result:
          show result of inference
          save model, prep_model
          update dataset
```

그림 2. 초분광 데이터 분석 플랫폼의 Pseudo code

Ⅲ. 적용 및 향후 연구 계획

그림 3은 상용 초분광 이미징 장치인 B150U (IMEC Int., Belgium)를 이용하여 사과 샘플로부터 수집한 초분광 이미지 데이터를 개발된 플랫폼을 이용하여 분석한 결과이다. 그림 2(a)는 샘플의 RGB 이미지이며, 중앙 부위에 멍(bruise)이 있으나 실화상에서는 잘 드러나지 않았다. 그림 2(b)는 사과의 정상 부위와 손상 부위의 초분광 데이터를 학습데이터셋(train/validation/test dataset)으로 구성한 후, SVM 모델 학습 및 추론을 통해 정상(gray color)/손상(black)/배경(white)으로 각각 분류한 결과이다. 그림 2(c)와 (d)는 각각 PCA와 LDA로 초분광 데이터 전처리 후 SVM으로 추론한 결과이다. 실험 결과 RGB 이미지에서는 잘 드러나지 않는 멍 부위가 초분광 데이터 전처리 및 신경망 추론을 통해 시각적으로 잘 구분될 수 있음을 확인하였다. 각 조건에서 Test dataset의 정확도는 SVM(99.66%), SVM w. PCA(99.79%), SVM w. LDA(98.59%)이었다.

현재 개발된 플랫폼은 픽셀 단위 머신러닝 알고리즘을 중심으로 기능 구현되어 있으나, 분석 성능 향상을 위해 향후 2D-CNN과 같은 패치 단위 알고리즘으로 기능을 확장할 예정이다. 그리고 학습데이터셋이 충분히 확보된다면 이상부위 진단을 위한 End-to-End 방식의 3D-CNN이나 Mask R-CNN(Region-based CNN)과 같은 딥러닝 알고리즘으로 기능을 지속 확장할 계획이다.



(a) RGB image (b) SVM

(c) SVM w. PCA (d) SVM w. LDA

그림 2. 초분광 분석 플랫폼 적용 사례

ACKNOWLEDGMENT

본 연구는 한국전자통신연구원 연구운영지원사업의 일환으로 수행되었음. [21ZK1100, 호남권 지역산업 기반 ICT 융합기술 고도화 지원사업]

참 고 문 헌

[1] Khan M. J. et al., "Modern Trends in Hyperspectral Image Analysis: A Review," IEEE Access, 6, pp. 14118-14129, 2018.

[2] Khan U. et. al, "Trends in deep learning for medical hyperspectral image analysis," IEEE Access, 9, pp. 79534-79548, 2021.

A Design of Traffic Accident Analysis Algorithm based on Scene Recognition

Min Htet Thar*, Cheol Min Lee**, Dong Myung Lee**
*Dept. of Computer & Media Engineering, Graduate School of Tongmyong University
**Dept. of Computer Engineering, Tongmyong University

*htatmin332@gmail.com, **captaink7428@gmail.com, **dmlee@tu.ac.kr

Abstract

In this paper, a scene recognition-based traffic accident analysis algorithm was proposed to automate the vehicle insurance claim process, and the performance was confirmed by establishing a vehicle damage dataset. As a result, it was confirmed that the average detection accuracy of the damaged vehicle of the proposed algorithm is 95%. The proposed system is processed based on the RCNN (mask region of convolutional network) model. Usage models include Convolutional Neural Network (CNN), Region Proposal Network (RPN), and Region of Interest (RoI). The results of this study are expected to be very useful in the field of automated traffic accident type identification and vehicle damage estimation.

I. INTRODUCTION

Recently, among computer vision technologies for extracting image information, research on scene recognition and object detection is being actively conducted. Most of the research uses deep learning algorithms such as Convolutional Neural Network (CNN), Residual Network (ResNet), and Region of Convolutional Network (RCNN) to extract image data. Computer vision covers the core technology of automated image analysis used in many fields such as medicine, machine vision, military and autonomous vehicles.

Applications of computer vision technology today are improving the insurance industry using automobile damage recognition and traffic analysis systems. Using these systems in the field makes it easy to analyze accident types, recognize vehicle damage, predict what types of repairs will be required, and estimate costs. Therefore, it has many advantages, such as reducing the time and effort required for human inspection and providing comfortable service.

Therefore, in this paper, we design a traffic accident type analysis algorithm using the vehicle damage location and vehicle damage volume dataset. In designing the algorithm, the type of traffic accident was analyzed using the vehicle damage location recognition result, and the degree of damage was considered using the vehicle damage recognition result. The dataset is a trained mask RCNN model and analyzes its recognition performance.

II. RELATED STUDIES

A vehicle damage detection segmentation algorithm based on the improved Mask RCNN was proposed by [1]. This algorithm is proposed to quickly solve the traffic accident compensation problem and uses an improved Mask RCNN to train the dataset and improve the detection accuracy. The results show that the improved Mask RCNN has better detection accuracy and AP value.

Vehicle damage detection and classification using deep learning-based algorithms has been proposed using VGG16 and VGG19 [2]. This algorithm detects damage in vehicle advertisements and evaluates their location and severity. In detecting damage, VGG19 is 95.22% accurate and VGG16 is 92.56%. At the location of the damage, VGG19 has an accuracy of 76.48% and VGG16 has an accuracy of 74.39%. Comparing these two results, VGG19 outperforms VGG16 in vehicle damage detection and classification.

III. PROPOSED ALGORITHM

3.1 Design Philosophies

The direction of the algorithm design is to automate the vehicle insurance claim process by analyzing the types of vehicle accidents [3]. Above all, in order to analyze the accident type, the classification process of vehicle damage location and vehicle damage size is important. This paper presents the design philosophies of the dataset construction and recognition process as follows.

- To build datasets of vehicle damage location and vehicle damage volume, each dataset collected 20 images for training and 5 images for validation and 3 images for testing.
- In image annotation process, each image marked using the making tool labeling and save as json format.
- In training process, it was trained by 100 epochs with prep-trained coco dataset and the learning time was about 30 minutes.
- After training the dataset, recognition process is run using trained dataset and Mask RCNN framework [4].

3.2 System Overview

In this paper, we proposed a scene recognition-based traffic accident analysis algorithm to automate the auto insurance claim process. In order to analyze the type of traffic accident, it is necessary to detect the vehicle's path and detect the type of traffic accident. This algorithm can detect and use parts of the damage location to help consider what type of accident has occurred. To figure out what kind of repair is needed and how much it will cost, this algorithm can also detect standard damage volume

levels such as A, B, S, N levels based on the ABI category information.

3.3 Algorithm Design

The proposed algorithm is designed based on an object detection system including a dataset and a deep learning model as shown in Fig. 1. The main core of algorithm design is **traffic accident types analysis**. Estimating the location of vehicle damage is very important for analyzing the type of traffic accident.

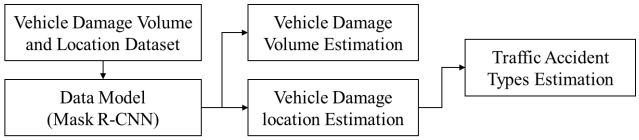


Fig. 1. Algorithm concept.

For the analysis process, a dataset of vehicle damage and location images is collected and trained by the Mask RCNN data model. We also use the vehicle damage data set and the Mask RCNN framework to perform the vehicle damage amount and damage location estimation process. After obtaining the results, we use the location of the vehicle damage to perform an accident type analysis.

3.3.1 Vehicle Damage Volume Estimation

Building a vehicle damage image dataset is an important part of both **vehicle damage volume estimation** and **vehicle damage location estimation**. Therefore, the proposed algorithm collects enough vehicle damage images and performs image annotation on each image. The image dataset is then trained as a data model. After training the data set, we perform the vehicle damage estimation process using the trained dataset.

In the estimation process as shown in Fig. 2, a vehicle accident image is input and the image is extracted as a feature map. Using this feature map, the damage part is extracted with the bounding box, and the extracted damage part is classified at what level of damage volume.

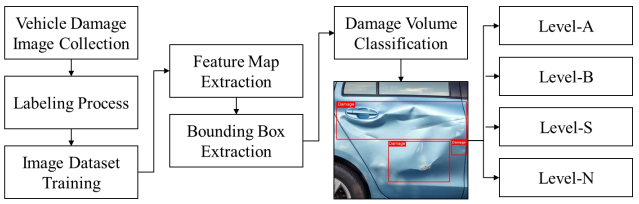


Fig. 2. Architecture of vehicle damage volume estimation.

In this paper, the standard of vehicle damage volume is defined according to Association of British Insurers (ABI) category information such as level A, B, S and N as shown in Table I.

3.3.2 Vehicle Damage Location Estimation

It is also important to build another dataset for estimating the location of vehicle. Building the dataset proceeds in the order of collecting enough vehicle damage images, performing image annotation processing on each image, and training the image dataset with the data model shown in Fig. 3.

TABLE I. ABI CATEGORY INFORMATION OF VEHICLE INSURANCE

ABI category information of vehicle insurance	
Level-A	Scrap only
Level-B	Break and crush body shell or chassis
Level-S	Structurally damaged but repairable
Level-N	Non-structurally damaged but repairable

The vehicle damage location estimation function defines vehicle damage locations as front, left, right, and rear. The vehicle damage location estimation part plays an important role in the **traffic accident types analysis**.

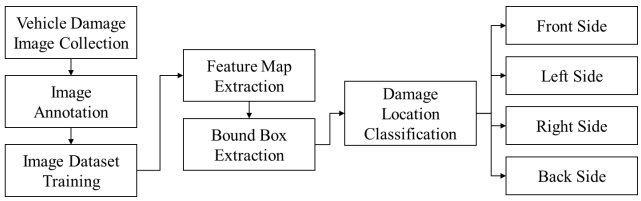


Fig. 3. Architecture of vehicle damage location estimation.

3.3.3 Vehicle Accident Types Analysis

A vehicle accident analysis algorithm was designed for only two vehicle accident situations. First, the crash vehicle is set to VEHICLE -A and VEHICLE -B, both crash vehicles detect the damage location, and then analyze the vehicle accident type through the damage location. According to Fig. 4, the proposed algorithm can recognize the types of traffic accidents from front to right, front to left and right, and front to rear.

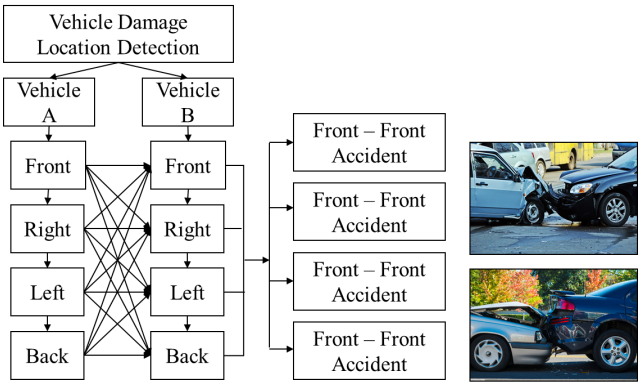


Fig. 4. Architecture of vehicle accident types analysis.

Considering the type of accident, if the damage location of VEHICLE-A is in the front and the location of damage in VEHICLE-B is in the front, it is recognized as a front accident. If the breakage location of VEHICLE-A is at the front and the breakage location of VEHICLE-B is at the rear, the system recognizes this as a front-to-rear accident.

IV. EXPERIMENTS AND RESULTS ANALYSIS

4.1 Experimental Environment

4.1.1 Labeling

The labels are predetermined by the machine learning engineer and are chosen to provide information about the computer vision model shown in the image as shown in Fig. 5.

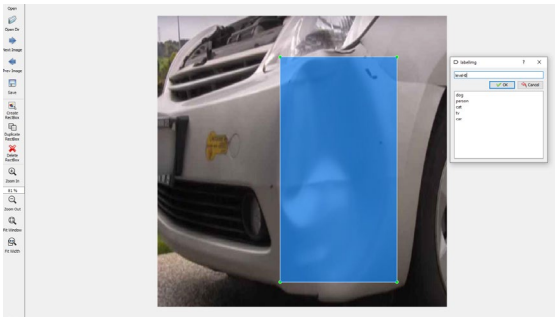


Fig. 5. Labeling in vehicle damage image.

4.1.2 Experimental environment information

As shown in Table 2, the hardware used for the experiment is a GeForce RTX 2080Ti GPU, and the software is Python 3.7, Tensorflow 1.3.0, Keras 2.0.8, Mask RCNN framework, Window10 64bit.

TABLE II. SPECIFICATION OF HARDWARE AND SOFTWARE FOR EXPERIMENTS

	Specification	
HW	GPU	GeForce RTX 2080Ti
SW	Python version	3.7
	TensorFlow version	1.3.0
	Keras version	2.0.8
	Framework	Mask RCNN
	Operation system	Window 10 64bit

4.2 Experiment Results on Vehicle Damage Detection

To validate the accuracy of vehicle damage volume, an experiment was performed using 20 images of damaged vehicles and 20 images of undamaged vehicles. According to Fig. 6, vehicle damage detection results are tested using a confusion matrix table. The experimental results confirmed that in the damaged vehicle image test, the system correctly detected 19 out of 20 images and correctly detected all images of the undamaged vehicle image.

	Damaged	Undamaged
Damaged	19	1
Undamaged	0	20

Fig. 6. Confusion matrices for vehicle damage detection.

According to Fig.7, the average detection accuracy of the damaged vehicle image test is 95%, and the average detection accuracy of the undamaged vehicle image is 100%. Since the performance of detection accuracy can vary depending on the dataset and data model, improvements in the dataset and data model can improve the detection accuracy performance.

V. FURTHER STUDIES AND CONCLUSION

In this paper, we propose the traffic accident analysis algorithm based on scene recognition to automate the vehicle insurance claim process. As a result of the experiment, it was confirmed that the vehicle damage detection accuracy of the proposed algorithm was 95%.

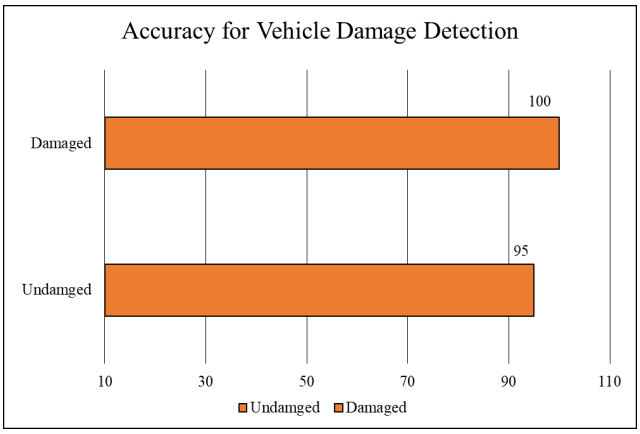


Fig. 7. Vehicle damage detection accuracy.

The purpose of this paper is to detect the location of damage to the vehicle and the amount of damage to the vehicle, ultimately to analyze the accident type, and to estimate the repair type and cost. However, it is very difficult to accurately detect the location and volume of vehicle damage, so better datasets and data models need to be built to improve performance. In practice, we tried to solve the problem using the popular models Mask RCNN, Yolov5, and detectron2, but there was still difficulty in detecting damage.

However, in the future, we plan to use the new data model to improve the algorithm and also detect vehicle routes to better analyze types of traffic accidents.

ACKNOWLEDGMENTS

This research was supported by the MISP (Ministry of Science, ICT & Future Planning), Korea, under the National Program for Excellence in SW supervised by the IITP (Institute for Information & communications Technology Promotion) (2018001874004).

This research was supported by the BB21plus funded by Busan Metropolitan City and Busan Institute for Talent & Lifelong Education (BIT).

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2019R1F1A1062670).

REFERENCES

[1] Qinghui Zhang, Xianing Chang, and Shanfeng Bian Bian, “Vehicle-Damage-Detection Segmentation Algorithm based on Improved Mask RCNN,” IEEE Access vol. 8, pp.6997-7004, Jan. 2020.

[2] Phyu Mar KYyu and Kuntpong Woraratpanya, “Car Damage Detection and Classification,” Proceedings of the 11th International Conference on Advances in Information Technology, pp.1-6, July 2020.

[3] Kalpesh Patil, Mandar Kulkarni, Anand Sriraman, Shirish Karande, “Deep learning-based car damage classification,” 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), pp.50-54, 18-21 Dec. 2017.

[4] Kaiming He, Georgia Gkioxari, Piotr Dollár, Ross Girshick, “Mask R-CNN,” Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, pp.2961-2969, 22-29 Oct. 2017.

Motor imagery 디코딩을 위한 Residual Dense Network

Permana Deny, 최계원
성균관대학교

denypermana@g.skku.edu, kaewonchoi@skku.edu

Residual Dense Network for motor imagery decoding

Permana Deny, Kae Won Choi
Sungkyunkwan University

요 약

이 논문은 residual dense network 을 사용한 Motor Imagery 의 디코딩 프로세스에 대한 연구이다. 본 연구에서는 뇌파(EEG)를 사용하여 Motor Imagery 중 뇌파의 신호를 수집하고 처리한 다음 신호를 세 가지 다른 Motor Imagery 신호로 디코딩하는 모델을 학습시켰다. 학습된 모델에서 피험자의 의도를 예측하기 위한 코헨(Cohen)의 카파 상관계수(Kappa)는 0.82 로 계산되었다.

Abstract

The paper studied a decoding process of motor imagery using adopted residual dense network. We acquired and processed the brain signal during motor imagery using electroencephalogram (EEG), then train the model to decode the signal into three different imagery intention. The proposed method achieved Cohen's value in predicting the subject the intention by 0.82.

Ⅰ. Background

Hand movement phases during motor imagery provide insight into the underlying neural mechanism in controlling human limbs. Using an EEG to record brain activity, we investigated the feasibility of brain signals during motor imagery (MI) from different imagery actions to establish a control path between human and computer.

This paper studied a brain-computer interface (BCI) to decode brain activity during motor imagery. The challenge of the decoding method lay on the brain signal that has non-stationary characteristics along the time. Signal representation method through short-time Fourier transform (STFT) and decoding method such as deep learning are proposed in the previous studies to address these challenges.

The decoding method aims to identify the signal distinction of motor imagery performed by the subject through transforming the signal to a smaller matrix representing the subject intention.

We propose a residual dense network architecture [1] consisting deep layer to decode the subtle characteristic from the different intention of movement phases during motor imagery. The adopted model consists of the convolutional network that initially extracts local features of image by accessing the previous input from concatenation and estimating the error through residual connection. These paths lift vanishing gradient in designing deeper layer towards the training process. Therefore, the model is able to generalize the feature to an acceptable degree.

In all, we evaluate the proposed method on the EEG signal during motor imagery to verify the effectiveness of decoding the motor imagery in predicting subject intention.

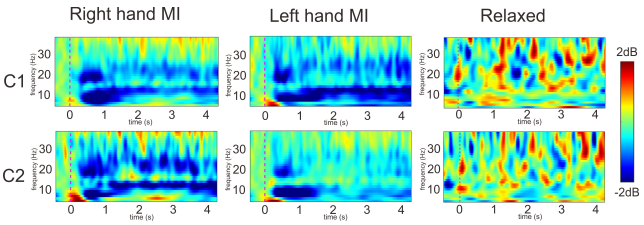


Figure 1. The average event related changes in spectral power of hand movement phases during motor imagery for channel C1 and C2. The STFT heighten the spectral change throughout the time, signified by deep blue color that indicates event related desynchronization (ERD).

Ⅱ. Discussion

We utilized an openBCI device which is consisted of 16 EEG channels, with 10-10 standard channel mapping. In order to acquire the motor imagery data,

we asked a subject to perform imagery movement of the right hand, left hand, and relaxed.

This experiment has a total 3000 trials (1000 trials for each type of imagery intention) in the dataset. The EEG was then, bandpass-filtered between 1 and 50Hz followed by a notch-filter at 60Hz. Finally, we extracted the signal from time period one-second pre-instruction and 4 seconds during motor imagery and did an STFT.

We completed the signal analysis to deliver a dataset, to then train a residual dense model. In the model framework, deep residual connection x aims to find underlying map h(x) from F(x) which is defined by:

$$h(x) := F(x) + x \quad (1)$$

The connection makes it possible to estimate the error of x, by estimating the error from residual function. Thus, when x is optimal, then $F(x) \rightarrow 0$ is achieved during training. Therefore, the approach becomes:

$$F(x) := h(x) - x \quad (2)$$

Furthermore, the adopted DenseNet [2] in the weight layer is to fit the residual value from x. The presence of DenseNet architecture is to design deep layer without severe vanishing gradient, by employing skip connection of previous layer to the end of the output.

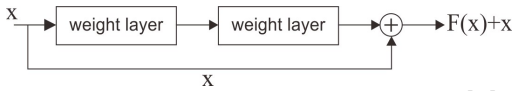


Figure 2. A building block of residual learning [3]. The tensor x represents layer input from previous connection. In this research, the residual layer consists of layer connection from DenseNet architecture.

The model consists of certain depth of residual and concatenation layers that is preserved on figure 3. It consisted of $\theta \approx 36M$ trainable parameters accounting each layer. As the result of feature representation, we impose an input size x:=[16,256,2] to the model, while the output y:=[1,3]. In the training phase, we occupy a callback function to update the learning rate by monitoring loss value. The model is trained for 200 epochs with batch size 32, and implemented Adam optimizer with initial learning rate 0.0001, together with categorical cross entropy as the loss function.

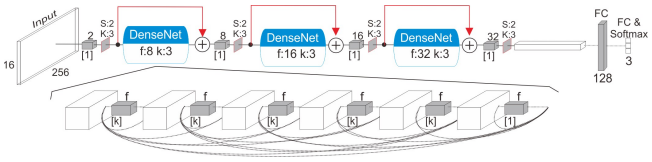


Figure 3. Abstract architecture of the adopted Residual Dense Network. The grey box indicates the Conv2D coupled with BatchNormalization layers. The above diagram is network with residual connection. The below diagram is DenseNet architecture with concatenation. The f represents filter number, while k represents [k,k] kernel number

The model is then evaluated and validated using a test dataset, while the evaluation that is calculated from confusion matrix is reported on the table 1. The significant result from this decoding method is achieved on the offline decoding, with the relaxed intention is highly distinguishable than the other intention.

Additionally, we reported different model properties using the same architecture to optimize the network and hyperparameter. Nonetheless, the deeper model does not achieve comparable result. It indicates that by adding or reducing the trainable parameter or network does not simply solve the decoding problem. Therefore, we further propose a deeper decoding method, different way than directly feeding the feature representation into neural network to explore feature connection over the EEG channels.

Table 1. Confusion matrix evaluation, PPV stand for positive predictive value, TPR for true positive rate, and k for Cohen's k-value. The results are respected to testing test for 900 trials from three different intention.

N=300	Right hand	Relaxed	Left hand	PPV
Right hand	267	8	23	0.89
Relaxed	6	272	21	0.90
Left hand	39	9	251	0.83
TPR	0.85	0.94	0.85	k=0.82

Ⅲ. Conclusion

In this paper, we studied a decoding method for motor imagery from three different intention. By comparing the true positive rate from three different intention, we observed the significant achievement on the relaxed state, hence the two other intentions have the same evaluation result. We achieved overall result performed by Cohen's k-value by 0.82.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 ICT 명품인재양성 사업의 연구결과로 수행되었음 (IITP-2021-2020-0-01821).

Reference

[1] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, Yun Fu. "Residual Dense Network for Image Super-Resolution". IEEE Converence on computer vision and pattern recognition, pp. 2472-2481. 2018.

[2] Huang Gao, Zhuang Liu, Laurens Van Der Maaten, Kilian Q.Weinberger. "Densely Connected Convolutional Networks". IEEE Converence on computer vision and pattern recognition, pp. 4700-4708. 2017.

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. "Deep Residual Learning for Image Recognition". IEEE Converence on computer vision and pattern recognition, pp. 770-778. 2016.

YOLOv5 Inference Performance Analysis on Edge with different Machine Learning Frameworks

James Rigor Camacho, Angela Caliwag, Wansu Lim
Kumoh National Institute of Technology

Abstract

In recent years, one of the key innovations in manufacturing automation is the improvement of object detection and classification through surveillance. To have a more reliable system that performs monitoring in a manufacturing facility, machines are integrated with object detection algorithms to help ensure the quality and sorting of products accurately and efficiently. One of the recent and widely used deep learning-based object detection algorithms is the YOLOv5, also known as "You Only Look Once" version 5. YOLOv5 is supported by machine learning frameworks in deep learning, providing different execution performance. The goal of this paper is to determine the best suited framework for YOLOv5 to achieve the optimal hardware performance on edge based on the inference time and inference speed of the model. The performance comparative results revealed that among all the frameworks that were tested on the YOLOv5 model, TensorRT gave the best overall performance. Its average inference time to process the prepared test images is 133ms which has a percent difference of 22.11%, 41.70%, and 65.87% from the other frameworks such as ONNX, TensorFlow, and PyTorch, respectively. In the TensorRT framework, YOLOv5 was able to increase its hardware performance in terms of inference speed, yielding 8fps and 12fps, respectively, based on its specified precision of FP32 and FP16. Hence, YOLOv5 in TensorRT framework can be deployed in a resource-constrained device such as edge AI with improved inference performance for applications that involve object detection and classification.

Ⅰ. Introduction

Manufacturing industries have allowed innovation in their manufacturing processes to increase their economic productivity that could keep up with the increasing demand of the public. The production and consumption of commodities have increased since the beginning of the industrial revolution, and it is that innovation in the manufacturing industries have helped to transform and to boost their productivity in the services provided. These services have aimed at manufacturing flexibility and rapid product innovation to gain competitive advantages in the manufacturing industry [1]. Recent innovative ideas in the manufacturing industry include hybridizing the tracking system with a Convolution Neural Network (CNN)-based object detection [2]; and integrating an improved Singe Short Detector (SSD) network to a defect detection system in manufacturing [3].

These methods, such as Faster R-CNN and SSD, for object detection applications are popular for their efficient feature extraction ability [3]. The increase in demand for manufacturing automation has been realized with use of these deep-learning methods. It has become the interest in automated factories to develop and improve the production line performing complex and time-efficient tasks. However, object detection methods that have been recently proposed are not all well-suited to cover complex environments and to perform time-efficient tasks. Scenes that have imperfect background can affect the detection performance of a surveillance system [4]. A surveillance system in manufacturing would either need to purchase a better system or use a compact system integrated with the most efficient and accurate object detection model to improve the efficiency of object detection in a complicated background. This brings the need for a more reliable system that performs monitoring in a manufacturing facility and integrated with an object detection algorithm to help ensure the quality and sorting of products accurately and efficiently.

One of the widely used deep learning-based object detection algorithms is known as "You Only Look Once" or YOLO. It is a regression-based object detection system that delivers class probabilities for the discovered images omitting the region proposal process [5]. This process was performed differently and was compared to other state-of-the-art models, the summary of YOLOv5 comparison to state-of-the-art models is presented in Table I. From the table, the methods are analyzed based on the complexity of their model, flexibility, and inference speed. These are the factors that greatly affect the hardware performance of a model when deployed on an edge. In [6], Fast R-CNN was the inspiration of the framework proposed in the

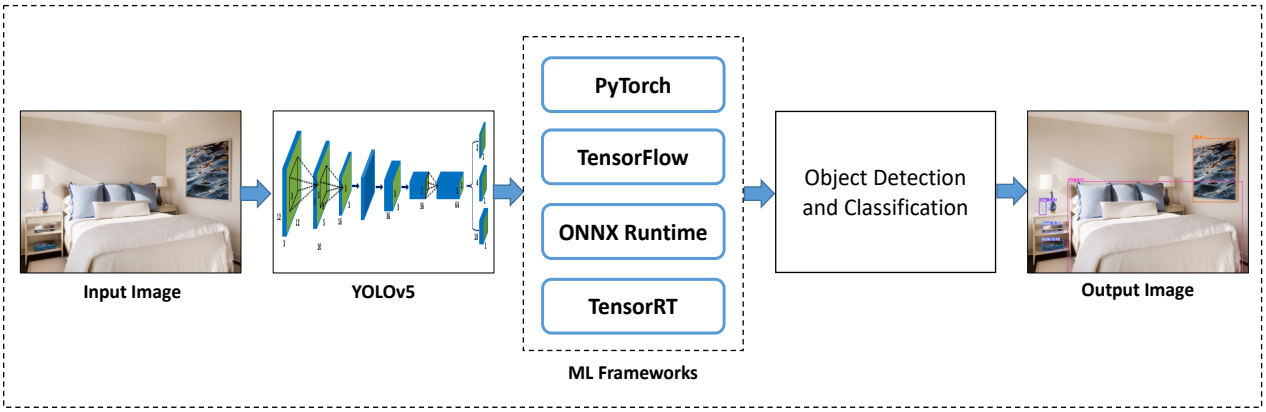


Fig. 1. Overview of the architecture for object detection and classification using YOLOv5 in different ML frameworks.

Table I
Summary of YOLOv5 comparison with state-of-the-art models for Object detection application

Method	Less Complexity	Flexibility	Inference Speed
Fast R-CNN [6]	✔	✘	7 FPS
SSD [7]	✔	✔	85 FPS
YOLOv5 [8]	✔	✔	155 FPS

paper. It is an object detection model that can achieve superior performance because of its CNN features. It is composed of two processes that involves regional proposal and object recognition [7]. Both parts of fast R-CNN can determine the object that appears on the image and specify the region proposal of that image. Which shows that the structure of this model has redundant functions and can be further improved to speedup the computation or processing of object detection. Morever, the region-based network of the model is not optimized which brought false detection results when applied on a complex environment. There are some studies that have tried to improve the outpuHence, it would require a lot of processing power when deployed on an embedded system. To resolve this constraint in an object detection system, the SSD method was introduced. In [8], SSD can handle complex environments by introducing default boxes in different aspect ratios to a larger input image. The object detected by the default boxes will be categorized based on the scores generated by the network. This predicts the object category and adjusts the box to better fit the shape of the object. Moreover, the model is flexible enough to be deployed in an embedded system and even applied for real-time applications. However, the inference performance of this model was inefficient for high quality real-time detection which brought to the idea of developing a YOLO model. Although the SSD has an inference speed of 85FPS, which is faster than the 7FPS of a Fast R-CNN model, the latest version of the YOLO series which is YOLOv5 [9] is capable of achieving 155FPS in real-time applications. Hence, among other models, YOLOv5 has been recorded to achieve faster speed, but lower accuracy. Considering its potential in the improvement of object detection algorithms, YOLO has become the hot topic by researchers to improve the surveillance system in manufacturing automation. The progression from the original version of YOLO up until its 5th version, has outperformed several models especially in terms of real-time inferencing. It is the latest version released for the YOLO series that is reliable, efficient, and simple to use.

In summary, although there are state-of-the-art models suitable for object detection applications, these models have some serious concerns. These concerns are as follows: 1) object detection algorithms produce false detection findings when applied in a complicated environment which indicates the lack of suitability and right information in the study, 2) the analysis in the improvement of the inference performance of a model was inefficient to determine a factor that contributes to the high quality real-time detection, and 3) deep-learning models when deployed on an embedded device would demand a lot of processing power to perform difficult and time-efficient tasks.

To address the aforementioned concerns, our study has focused on benchmarking of YOLOv5 using known frameworks like TensorRT, ONNX Runtime, Pytorch, and Tensorflow on edge AI devices. The performance of the model is analyzed based on the inference performance of the device. These parameters would determine which framework has the highest efficiency of the YOLOv5 on edge AI device. It is, thus, aimed to attain a high-efficiency YOLOv5 model with the lowest possible hardware requirements on deep learning applications. The experimental analysis results will decide and explain which framework is best suitable for YOLOv5 on edge to conduct an accurate and efficient object detection task in a manufacturing environment.

Ⅱ. Methodology

The overview for the architecture of object detection and classification using YOLOv5 on different machine learning frameworks is shown in Fig.1. This pipeline is composed of input images, YOLOv5 model, machine learning frameworks (i.e., PyTorch, TensorFlow, ONNX, and TensorRT), object detection



Fig. 2. NVIDIA Jetson Nano (Edge AI device).

and classification algorithm, and output images. A set of input images is fed into the pretrained YOLOv5 model where different machine learning frameworks are used in performing the object detection and classification algorithm. The output images are expected to have bounding boxes on the objects detected with their classifications. The inference time and speed of detecting objects from each image will also be outputted which varies depending on the ML framework that was used.

A. The Network Architecture

YOLO is a family of deep learning-based object detection architectures and models pretrained on the COCO dataset. Before YOLO, the process of classifying objects in an input image was difficult, slow-moving, and not very energy efficient. Instead of going through the same process, the YOLO algorithm simplifies object detection and classification tasks by treating it as a regression problem and employing a one-stage algorithm to predict object classes and positions immediately. Specifically, it divides the input images into a grid system and makes predictions on all the objects detected in the images by computing all the features at the same time. It can precisely predict and locate the object using bounding box coordinates.

Dataset: To evaluate the performance of the YOLOv5 model based on the inference speed at different frameworks, a benchmark data set is utilized. Fig.3 contains the test images that were picked from the pool of available images online aimed to best demonstrate the object detection and classification capabilities of the model. These images show the possible installation and applications of this study namely, pedestrian, streets, school facilities, backyard/garden, urban park, bus stop, and different parts of a house. Moreover, these images can be found in [10-19]. Hence, the dataset is important when re-training a model for a specific application. The more dataset is used to train the model, the higher the improvement in its accuracy and performance.

YOLOv5: It is the latest version released for the YOLO object detection series which is 3x faster as compared with YOLOv4. By just looking at the image once, it can detect the objects with an expected speed of 140 FPS (Frames Per Second). In addition to that, the size of YOLOv5 is 9x smaller making it suitable to efficiently deploy on edge AI devices having low computational capability. YOLOv5 is supported by machine learning frameworks in deep learning, providing different execution performance.

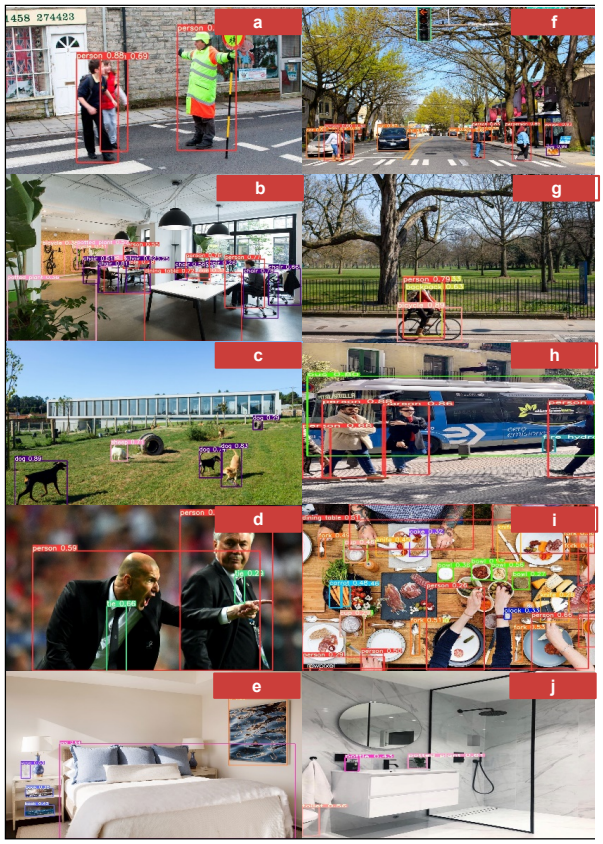


Fig. 3. Test Images. (a)Pedestrian, (b)Office, (c)Park, (d)Zidane, (e)Bedroom, (f)Road, (g)Bike, (h)Bus Stop, (i)Dinning, (j)Bathroom.

Frameworks: YOLOv5 was introduced with a PyTorch framework. The aspiration of researchers to continuously improve the architecture of the YOLO model has brought to the idea of converting a model defined in PyTorch into a different framework format (i.e., ONNX, TensorFlow, and TensorRT). These frameworks have a corresponding effect on the performance of the YOLOv5 model in terms of inference speed and accuracy. A benchmarking analysis on the results is undertaken and compared statistically and qualitatively with state-of-the-art models to evaluate the effectiveness of each framework and investigate the performance of the YOLOv5 model in an edge AI device.

B. Hardware Platform

The hardware platform used as an edge AI device in the experimental setup is the NVIDIA Jetson Nano device. Fig.2 shows the actual device where the YOLOv5 model is deployed and performs the object detection and classification tasks. It is a resource-constrained device, which means that it has limitations on resources such as computational power and memory

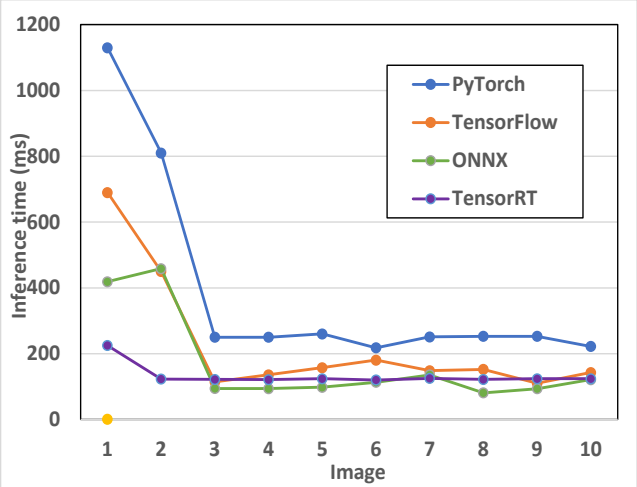


Fig. 4. Inference time of processing test images in different machine learning frameworks.

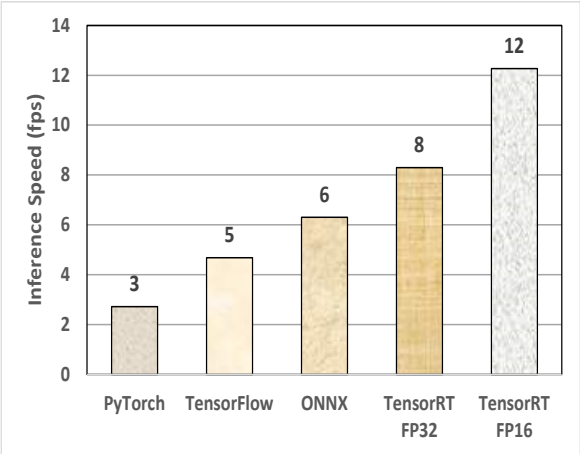


Fig. 5. Inference speed of different ML framework With NVIDIA Jetson Nano.

demand. For the YOLOv5 model, this is the perfect hardware platform to analyze the hardware performance of the model with different frameworks.

It is important to understand the behavior of these models in actual implementation since most applications would require installation setup. It is compact and compatible with deep learning-based models, specifically, YOLO models since the structure have less complexity and less energy requirements.

Ⅲ. Results and Discussion

In this section, the experimental setup and results are discussed to provide a comprehensive benchmarking analysis of the YOLOv5 model with different machine learning frameworks.

Experimental Setup: NVIDIA Jetson Nano was used as the edge AI device to deploy the deep learning network, YOLOv5 model, in different frameworks for object detection and classification tasks. This edge AI device is popular for research that is related to building autonomous machines and complex AI systems because of its compactness and capabilities. Despite being small, it is a powerful edge AI device that is well-suited for real-world application and experimental setup. Deploying the model in this device provides the inference speed and inference time of processing object detection and classification tasks in every input image. The hardware performance of the YOLOv5 model in different frameworks is evaluated with the following parameters: inference speed and inference time of the model design.

A. Deep Learning Inference Performance

The YOLOv5 model was implemented with different frameworks and ran inference on the test images. The results for the inference time of processing the prepared test images in different ML frameworks is shown in Fig.4. Given the trend on the inference time of these ML frameworks, it gives the intuition that Pytorch took the longest time to identify and classify the objects in the test images. On the other hand, TensorRT is the framework that has the fastest average inference time of 133ms to process the test images. Despite that, all the frameworks for YOLOv5 were able to maintain their average inference time for the succeeding test images starting the 3rd image. Moreover, among all the test images, the image (i) for dinning in Fig.3 took the longest for all the frameworks to detect and classify the objects in the image. The resulting percent difference of the different frameworks from the overall average inference time are 7.71%, 37.23%, 37.46%, and 39.22% for TensorRT, ONNX, TensorFlow, and PyTorch, respectively.

For the inference speed of the different frameworks, the YOLOv5 is deployed on edge AI devices and the results are recorded during the performance test. The graph in Fig.5 for inference speed of different ML frameworks with NVIDIA Jetson Nano shows that TensorRT FP16 and TensorRT FP32 got the fastest inference speed among the other frameworks with an inference speed of 12fps and 8fps, respectively. The result also shows that the YOLOv5 model at TensorRT FP16 format is 2x faster than the ONNX framework and 4x faster than the PyTorch framework. In reference to the average inference speed of the YOLOv5 model in different frameworks, the percent difference of their individual results is a 9.95%, 18.46%, 33.20%, 57.14%, and 75.29% for ONNX, TensorRT FP32, TensorFlow, PyTorch, and TensorRT FP16, respectively.

From the overall performance of the YOLOv5 model on edge for object detection and classification tasks, TensorRT framework got the best performance based on the inference time and inference speed results. TensorRT managed to process all the test images within an average of 133ms (inference time) with an inference speed of 8fps for FP32 and 12fps for FP16. To achieve an optimal performance for the YOLOv5 model, the best-suited framework to use is the TensorRT FP32. It might not have the fastest inference speed compared to FP16, but it has the most stable and high output results, which is 9.95% percent difference from the overall average inference speed, compared to all other frameworks.

Ⅳ. Conclusion

In this paper, a performance comparative study is performed in the YOLOv5 model with different ML frameworks to determine which model architecture is best suited for the application of object detection and classification tasks. This paper utilized an edge AI device to observe the hardware performance of the YOLOv5 model through its inference time and inference speed. Its goal is to develop a YOLOv5 model that can run on low-resource devices and give real-time performance. In terms of overall performance for the YOLOv5 model, the acquired average inference speed is 6fps which is 2x slower on PyTorch but 2x faster on the TensorRT FP16 framework. For inference time, TensorRT got the fastest average time of 133ms which has a percent difference of about 22.11%, 41.70%, and 65.87% for ONNX, TensorFlow, and PyTorch, respectively. Therefore, based on the results, TensorRT is the best suited framework for YOLOv5 with an average inference time of 133ms and an inference speed of 8fps and 12fps for TensorRT FP32 and TensorRT FP16, respectively. It was proven to provide the best inference performance for the YOLOv5 model on the application of object detection and classification algorithms. For future works, TensorRT provides INT8 optimization as well which can be used to other Jetson series like NVIDIA Jetson TX2 and NVIDIA Jetson Xavier. This could further optimize the inference performance of the YOLOv5 model by enabling the FP16 in the TensorRT framework.

ACKNOWLEDGMENT

This work was supported by the Ministry of SMEs and Start-ups, S. Korea (S2829065, S3010704), and by the National Research Foundation of Korea (2020R1A4A101777511, 2021R1I1A3056900).

References

[1] Bi Xinhua, Chen Taibo, Yu Baojun and Yu Cuiling, "Manufacturing flexibility and rapid product innovation: Two key manufacturing decision issues in turbulent business environment," 2009 Chinese Control and Decision Conference, 2009, pp. 4963-4967, doi: 10.1109/CCDC.2009.5194921.

[2] M. Jiang, K. Shimasaki, S. Hu, T. Senoo and I. Ishii, "A 500-Fps Pan-Tilt Tracking System With Deep-Learning-Based Object Detection," in IEEE Robotics and Automation Letters, vol. 6, no. 2, pp. 691-698, April 2021, doi: 10.1109/LRA.2020.3048653.

[3] J. Yang, S. Li, Z. Wang and G. Yang, "Real-Time Tiny Part Defect Detection System in Manufacturing Using Deep Learning," in IEEE Access, vol. 7, pp. 89278-89291, 2019, doi: 10.1109/ACCESS.2019.2925561.

[4] Y. Tian, R. S. Feris, H. Liu, A. Hampapur and M. Sun, "Robust Detection of Abandoned and Removed Objects in Complex Surveillance Videos," in IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 41, no. 5, pp. 565-576, Sept. 2011, doi: 10.1109/TSMCC.2010.2065803.

[5] W. Liu et al. (Dec. 2015). "SSD: Single shot multibox detector."[Online].Available:https://arxiv.org/abs/1512.02325

[6] Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. In: CVPR. (2016)

[7] S. Hsu, C. Huang and C. Chuang, "Vehicle detection using simplified fast R-CNN," 2018 International Workshop on Advanced Image Technology (IWAIT), 2018, pp. 1-3, doi: 10.1109/IWAIT.2018.8369767.

[8] H. Li, L. Deng, C. Yang, J. Liu and Z. Gu, "Enhanced YOLO v3 Tiny Network for Real-Time Ship Detection From Visual Image," in IEEE Access, vol. 9, pp. 16692-16706, 2021, doi: 10.1109/ACCESS.2021.3053956.

[9] X. Wang, H. Ma, X. Chen and S. You, "Edge Preserving and Multi-Scale Contextual Neural Network for Salient Object Detection," in IEEE Transactions on Image Processing, vol. 27, no. 1, pp. 121-134, Jan. 2018, doi: 10.1109/TIP.2017.2756825.

[10] M. Waghorn and (image: Getty), "Children can't safely cross the road until they are over 14 years old, scientists warn," Mirror.co.uk, 20-Apr-2017. [Online]. Available: https://www.mirror.co.uk/news/uk-news/children-cant-safely-cross-road-10259133.

[11] "3d visualization architecture," Service4money.com. [Online]. Available: https://www.service4money.com/3d-visualization-architecture/.

[12] "Dogs-and-cats-hotel-by-raulino-Silva-arquitecto-00," Aasarchitecture.com. [Online]. Available: https://aasarchitecture.com/2020/10/dogs-and-cats-hotel-by-raulino-silva-arquitecto.html/dogs-and-cats-hotel-by-raulino-silva-arquitecto-00/.

[13] "Ancelotti looks forward to Madrid-Bayern showdown with 'icon' Zidane," Loop News, 02-Apr-2017. [Online]. Available: https://tt.loopnews.com/content/ancelotti-looks-forward-madrid-bayern-showdown-icon-zidane.

[14] H. Mendelsohn, "If you've always wanted to sleep on a cloud, copy these white bedroom ideas," Housebeautiful.com, 01-Apr-2012. [Online]. Available: https://www.housebeautiful.com/room-decorating/colors/g1215/white-bedrooms/.

[15] S. Davis, "Seattle takes new steps to fine-tune traffic signals for people walking and rolling during COVID-19 health crisis," Seattle.gov, 07-May-2020. [Online]. Available: https://sdotblog.seattle.gov/2020/05/07/seattle-takes-new-steps-to-fine-tune-traffic-signals-for-people-walking-and-rolling-during-covid-19-health-crisis/.

[16] "Bloomberg," Bloomberg News.

[17] "YOLOv5 environment construction and target detection - Programmer Sought," Programmersought.com. [Online]. Available: https://programmersought.com/article/52126720263/.

[18] "Group of people dining concept," 123Rf.com. [Online]. Available: https://www.123rf.com/photo_67127031_group-of-people-dining-concept.html.

[19] Thetrendspotter.net. [Online]. Available: https://www.thetrendspotter.net/small-bathroom-design-ideas/

Reinforcement Learning based resource scheduling for network slicing in vehicular networks

Muhammad Ashar Tariq, Malik Muhammad Saad, Mahmudul Islam, Muhammad Toaha Raza Khan, Junho Seo, Dongkyun Kim
School of Computer Science and Engineering, Kyungpook National University, Republic of Korea
{tariqashar,maliksaad,mislam,toaha,jhaco,dongkyun}@knu.ac.kr

***Abstract*—The number of mobile devices is increasing at an unprecedented rate which also increases the traffic demands. Given the limited network resources, it is natural to slice the physical network into several logical virtual networks to fulfill the service demands. The network slices providing different services on demand have certain QoS requirements which need to be fulfilled in order to provide smooth service to users in a mobile network. Specifically, in a vehicular environment, the high dynamicity calls for real-time management and scheduling of resources. In order to utilize the limited network resources efficiently while honoring the QoS requirement of each slice, it is needed to schedule the resources in real-time based on the service demands. The proposed scheme provides a reinforcement learning-based solution to schedule the resources of a base station for different slices in the vehicular environment for various on-demand services. On the basis of required and available resources, as well as the priority of service, the base station schedules its resources among different slices in real-time.**

***Index Terms*—Reinforcement Learning (RL), Network Slicing, Resource Scheduling, Vehicular Networks**

I. Introduction

With the proliferation of mobile devices, the demand for mobile data traffic is also increasing unprecedentedly. Mobile network operators are expected to meet 1000 times more data traffic demand in the coming decade to support various services such as Augmented Reality (AR), Virtual Reality (VR) for several vertical industries including the automobile industry. The requirements of these vertical industries will be much higher than that of traditional broadband users. The structure of current mobile networks does not allow it to satisfy the diverse requirement of the multifaceted services of the vertical industries as the networks are not scalable and flexible. To satisfy the various requirements in terms of performance, network availability, security, and monetary cost, network slicing has been proposed by various researchers from both academia and industry. The network slicing paradigm has emerged from the well-established virtualization techniques such as Network Function Virtualization (NFV) and Software-Defined Networking (SDN) [1], [2]. The primary concept of network slicing is to divide a physical network into several logical ones which can be assigned for different vertical-specific services. Each sliced part is an isolated virtual network containing several Virtual Network Functions (VNFs) sharing a common physical network infrastructure.

While the core networks can be easily virtualized by utilizing the network slicing technology, slicing of the Radio Access Network (RAN) is quite challenging due to the limited resources [3]. To satisfy the diverse quality of service (QoS) requirements demanded by the individual users, the realization of the RAN slicing needs to be done properly. In traditional RAN, the resources are assigned to several users in a way so that the efficiency of the spectrum sharing is maximized. So, in RAN slicing the primary objective is to allocate the resources among the slice owners in a more flexible and controllable manner so that the resource utilization efficiency is maximized [4]. Slice owners should be given some flexibility so that they can maintain the isolated slices which will prevent the service interruption and improve the utilization of the limited resources. Another challenging issue of RAN slicing is that the resources must be allocated efficiently among the users in real-time [5]. In mobile environments such as vehicular networks, the users move rapidly from one place to another which makes resource management more critical.

Vehicular networks are known for their dynamic characteristics of the node. As the vehicles on the road travel from one place to another with high speed, the users come across various base stations within a short time. Intelligent transportation systems require a high data rate with QoS for the realization of high definition (HD) video streaming, on-demand traffic information, smart navigation, autonomous driving, and several other applications [6]. For emergency applications such as accident reporting, ITS require low latency connectivity between vehicles and the data center. Depending on the purpose and requirements of the application the physical network can be divided into multiple logical units and assigned for user-specific applications. Due to scarcity of resources, while slicing the network, resource management has to be done efficiently so that every user can satisfy their diverse requirements. This paper aims to provide real-time resource scheduling among multiple slices. The limited resources of a base station are scheduled among slices based on the required on-demand services in order to utilize and distribute the resources efficiently while fulfilling the QoS requirements of the service.

The remainder of this paper is as follow. Section II gives the system architecture. Section III discuss about the proposed scheme for scheduling of resources in network slicing. Finally

the conclusion is drawn in Section V.

II. System Architecture

In this paper, we consider multiple slices deployed on base stations (BS). The BSs are enabled with software-defined Network Function Virtualization (NFV). The physical network is logically abstracted into the sliced layers as shown in Figure . We consider $S_M = \{1, 2, 3, \ldots, M\}$ sliced networks. Each sliced network is responsible to meet user requirements. To meet diverse service requirements in terms of bandwidth and latency, the network is sliced specifically considering the constraints. The constraints such as some users require low latency and high reliability, some may require high bandwidth and are not delay-sensitive, and so on. Each sliced or logical network can serve multiple users, likewise, the set $U = \{U_1, U_2, U_3, \ldots, U_M\}$ indicates the set of users where $1, 2, 3, \ldots, M$ indicates the sliced network. For example, a user watching HD 4K video requires high bandwidth and low latency where an IoT monitoring device may require low bandwidth and high latency can be affordable. Similarly, in [7] experimental evaluation has been done using multiple configurations of virtual radios considering orthogonal frequency division multiplexing (OFDM) and filter bank multicarrier (FBMC) waveform.

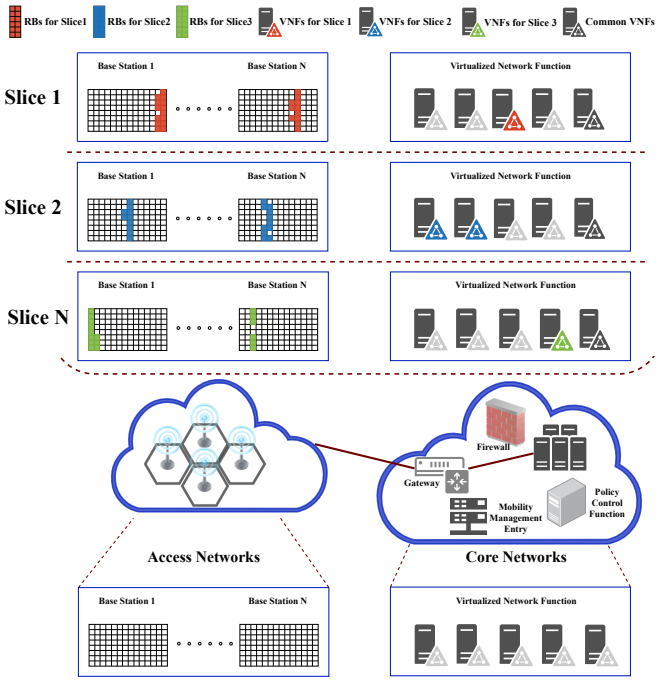


Fig. 1. Resource scheduling in network slicing

III. Reinforcement Learning based Resource Scheduling in Network Slicing

As the wireless networks evolve, a permanent topic under discussion is resource management that involves, resource allocation and scheduling. In the proposed architecture, different slices are composed of various virtual network functions (VNFs) and are allocated the required resources. Each slice may cover different BSs and the corresponding resources of the BS are distributed among slices. The slice instances are created based on the diverse service requirements for instance, a slice may require higher bandwidth and lower latency for HD video streaming services whereas for emergency services, high reliability and ultra-low latency may be required. Depending upon the service requirements resources are scheduled among different slices. In the proposed scheme, we use reinforcement learning for real-time scheduling of resources among different slices based on the service requirements. Each BS acts as an agent and performs the actions of scheduling the resources like bandwidth, VNFs, computation resources, etc. to different slices in order to fulfil the quality-of-service (QoS) requirements of that slice. The proposed reinforcement learning model is explained in the following subsection where the states, actions and reward policy are clearly described.

A. Reinforcement Learning

RL based mechanism consists of state and action taken by the agent and in return gets the reward. The BS observes the current state and based on optimal policy takes the action. In order to utilize the limited resources of the BS efficiently, the resources are optimally scheduled among different slice instances on real-time service requirements while maintaining the QoS. For example, if at a particular time instance, only video streaming services are required by the vehicle(s), the BS will allocate most of the bandwidth resources to fulfill the QoS requirements of the video streaming slice. However, in the case when there is some road hazard and emergency message dissemination is to be broadcasted with the images of the hazardous region, priority will be given to ultra-reliable low latency communication (URLLC) slice and the resources will be scheduled to that slice.

The BS acts as an agent and performs the actions of scheduling resources based on the observations of the current state which include the required resources, available resources, and service priority. The service requirements are defined in the service level agreement (SLA) at the time of slice creation. This tells us the required resources by a service. It is represented by R_r. The available resources at the BS are represented as R_a. Finally, the third observation parameter of the state vector, service priority is represented as P. Therefore, the state vector is given by equation 1.

$$S_t = [R_r, R_a, P] \tag{1}$$

Based on the policy, the agent takes an action and gets a reward for the decision made. The BS employs a scheduling algorithm to move the limited available resources among different slices to efficiently utilize the resources. The specific actions are defined based on the scheduling algorithm used. The goal of employing reinforcement learning is to find an optimal policy and maximize the cumulative reward. The reward function is represented by equation 2.

$$R_t = E[\sum_{n=0}^{\infty} \beta^n r_t + n] \quad (2)$$

where $\beta \epsilon [0, 1]$ is the discount factor.

The stochastic transitions of state and the reward are modeled using the Markov decision process (MDP). The probabilities of rewards and state transitions are dependant on the environment and the decision taken by the agent i.e. action. The conditional transition probability is represented as $P(s_{t+1}, r_t | s_t, a_t)$.

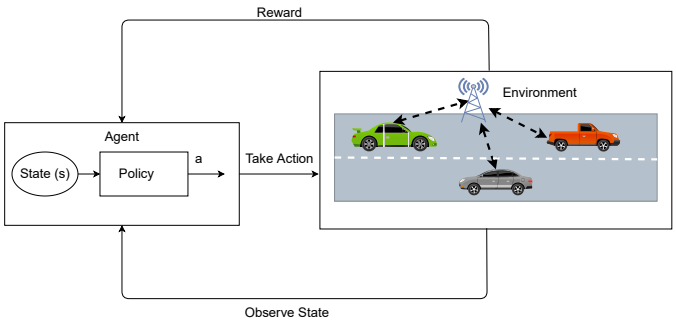


Fig. 2. Reinforcement Learning for resource scheduling in network slicing

IV. Acknowledgement

This study was supported by the BK21 FOUR project (AI-driven Convergence Software Education Research Program)funded by the Ministry of Education, School of Computer Science and Engineering, Kyungpook National University, Korea (4199990214394).

V. Conclusions

In this paper, we propose a reinforcement learning based scheduling of resources for network slicing in the vehicular network environment. The resources like bandwidth, computation, VNFs, and others are scheduled among the slices providing different services such as video streaming, emergency message dissemination, voice-over-IP, and so on. Since each service and hence each slice has different resource requirements that may change in real-time depending on the demand, the limited resources of the base station are to be managed and scheduled among different network slices. The proposed reinforcement learning based solution provides real-time scheduling of resources among slices for efficient resource utilization while fulfilling the QoS requirements.

References

[1] L. U. Khan, I. Yaqoob, N. H. Tran, Z. Han, and C. S. Hong, "Network slicing: Recent advances, taxonomy, requirements, and open research challenges," *IEEE Access*, vol. 8, pp. 36009–36028, 2020.

[2] S. Zhang, "An overview of network slicing for 5g," *IEEE Wireless Communications*, vol. 26, no. 3, pp. 111–117, 2019.

[3] H. Wang, Y. Wu, G. Min, J. Xu, and P. Tang, "Data-driven dynamic resource scheduling for network slicing: A deep reinforcement learning approach," *Information Sciences*, vol. 498, pp. 106–116, 2019.

[4] M. Yan, G. Feng, J. Zhou, Y. Sun, and Y.-C. Liang, "Intelligent resource scheduling for 5g radio access network slicing," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 8, pp. 7691–7703, 2019.

[5] R. Li, Z. Zhao, Q. Sun, I. Chih-Lin, C. Yang, X. Chen, M. Zhao, and H. Zhang, "Deep reinforcement learning for resource management in network slicing," *IEEE Access*, vol. 6, pp. 74429–74441, 2018.

[6] M. M. Islam, M. T. R. Khan, M. M. Saad, and D. Kim, "Software-defined vehicular network (sdvn): A survey on architecture and routing," *Journal of Systems Architecture*, p. 101961, 2020.

[7] M. M. Saad, F. A. Bhatti, A. Zafar, S. Jangsher, D. Kim, and M. Maqsood, "Air-interface virtualization using filter bank multicarrier and orthogonal frequency division multiplexing configurations," *Transactions on Emerging Telecommunications Technologies*, vol. 32, no. 2, p. e4154, 2021.

A Lightweight Facial Emotion Recognition System Using Network Optimization Methods

Erick C. Valverde, Angela C. Caliwag, Wansu Lim
Kumoh National Institute of Technology

Abstract

Facial emotion recognition system has been recently developed for more advanced applications of face detection. The FER system classifies the human expression in various categories such as angry, disgust, fear, happy, sad, surprise, and neutral. Conventional FER systems have issues with low accuracy and high resource requirements. In order to increase the accuracy, an extreme version of Inception V3, known as Xception, is leveraged in this paper. To enable low resource requirements, a 68-landmark face detector from Dlib is used. Moreover, to develop a lightweight FER model, different network optimization methods are applied to Xception. The optimization methods used are pruning and quantization to support lower computational costs and reduce memory usage, respectively. Furthermore, to increase the inference speed of the FER system, a deep learning (DL) compiler is used to implement advanced optimization techniques to the model. The objectives of these optimization methods are experimentally demonstrated by the proposed lightweight FER system as compared to VGG-Net and ResNet50. Hence, the proposed system can be used to realize an efficient FER system in real-time inference.

Ⅰ. Introduction

Emotion recognition based on artificial intelligence (AI) has been recently developed to realize a better human-computer interaction. It is the process of identifying human emotions by analyzing facial expressions [1], decoding voice patterns [2], monitoring eye movements [3], or examining brain signals [4]. Since humans have cultural differences or distinct ways of expressing emotions, AI-based emotion recognition has long been a broad area of study. That is, the challenges of human diversity make it harder for computers to draw accurate conclusions. Nonetheless, researchers have been exerting efforts to alleviate the issues with emotion recognition by utilizing different supervised learning task algorithms. This could increase the potential of creating a more accurate emotion recognition algorithm based on human interactions with computers through visual, sounds, or neurological signals.

One of the most prominent fields of study nowadays regarding emotion recognition is facial emotion recognition (FER). FER system refers to an emotion recognition process that analyzes human face expressions to identify a specific emotion. As shown in Fig. 1, this system involves two tasks: 1) face detection task and 2) facial emotion classification task. First, the human face is detected from the image acquired. Second, the detected face is analyzed by an FER algorithm to classify which emotion it displays. The human emotions are commonly categorized as follows: 1) angry, 2) disgust, 3) fear, 4) happy, 5) sad, 6) surprise, and 7) neutral. These seven emotions are accountable for the complex implementation of FER which typically provide a various range of accuracy. For instance, [5] obtained the highest accuracy in happiness, while the lowest accuracy is observed in fear, which results to a low overall accuracy of 76%. [6] explored different

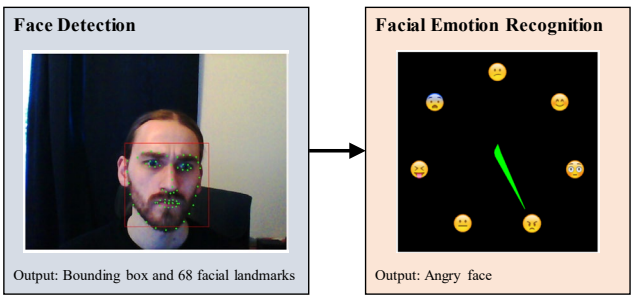


Fig. 1. Demonstration of an FER system [9].

supervised classification task algorithms for FER using K-nearest neighbor (KNN) and artificial neural network (ANN), separately. But both KNN-based and ANN-based FER methods lead also to low accuracy results (i.e., 54.16% and 66.66%, respectively) due to the corresponding low accuracies from different emotion categories. To resolve this issue with the accuracy, deep learning (DL)-based methods are proposed to improve the overall accuracy of FER algorithms up to 90% [7]. By this, researchers focus more on developing deeper networks or known as convolutional neural networks (CNNs) to find more complex features and increase the accuracy of FER system. The complexity of a CNN model refers to millions of connections within an architecture. It aims to search for significant facial patterns that can bring even up to more than 95% accuracy to FER system [8]. However, the inference speed and computational costs of FER system are often disregarded in exchange for higher accuracy results.

To cope with the issues of CNN models, some researchers

TABLE I
Comparison of Face Detection Methods [15]

Metrics / Methods	Haar	HOG	LBP
Detected Frames	653484	772954	503516
Detected Faces	652451	772954	503350
TPR (%)	78.23	**92.68**	60.37
FNR (%)	21.76	7.31	39.64

proposed to develop lightweight FER models. A lightweight FER model refers to an FER model with reduced resource requirements and faster inference speed. Several studies create transfer learning-based lightweight FER models by using state-of-the-art pre-trained DL models like VGG-Face [9], VGG16 [10], and Deep Face Convolutional Neural Network (CNN) [10]. However, their studies benchmark only which pre-trained models can provide both high accuracy and less complex architecture as compared to other existing models. That is, the original structure of the pre-trained models is not reduced, and so, no relative improvement can be observed in the results. To realize a novel lightweight FER model, several optimization methods can be used. An optimization method removes unimportant and redundant connections within a CNN model's architecture, albeit at the risk of a significant reduction in accuracy. The commonly used types of optimization methods are: 1) pruning, 2) quantization, and 3) DL compiler. Pruning refers to effective search and elimination of unimportant and redundant connections within the CNN model to reduce hardware requirements [11]. On the other hand, quantization refers to the conversion of the model from higher to lower precision format to support faster computation [12]. Meanwhile, DL compiler targeted the hardware architecture of the model to maximize the performance of the available hardware, at less computational costs, during DL inference [13]. These three optimization methods can be used to build a lightweight FER model that can provide better inference performance over existing pre-trained models.

In this paper, we proposed a lightweight FER model by network optimizations. The proposed method aims to have an efficient FER system by reducing the complexity of the DL-based FER model. The application of several optimization methods can optimize both the software and hardware components of the FER system. The optimization method will be carefully implemented to prevent any potential model accuracy loss. The contributions in this paper are summarized as follows:

1) To reduce computational costs, pruning is applied to remove unimportant and redundant connections within the architecture of the FER model.
2) To reduce the memory usage and increase inference speed of FER system, quantization is used to save the lightweight FER model at lower precision format.
3) To further increase inference speed while maintaining the reduced computational costs, DL compiler will be used to redesign the FER model to leverage with the available hardware of the device.

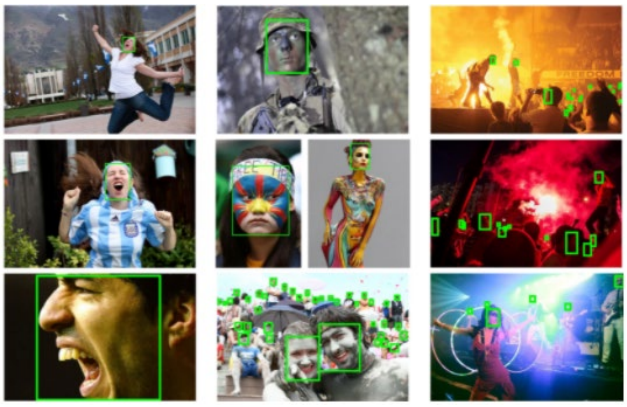


Fig. 2. Samples of WIDER FACE dataset.

The remainder of this paper presents in detail the development of FER system, different network optimization methods, the proposed methodology, and conclusion.

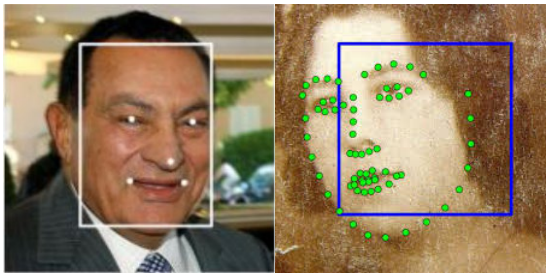
Ⅱ. Related Works

In this section, the development of facial emotion recognition (FER) system is discussed in detail. This is followed by the discussion of different network optimization methods that can be used to realize better inference performance in the FER system.

A. Development of Facial Emotion Recognition System

The current FER systems are divided into two tasks: 1) face detection task and 2) facial emotion classification task. In the FER system, the human face is detected and fed to an algorithm to analyze patterns and classify facial expressions. The emergence of state-of-the-art face detection algorithms led to the development of FER systems.

The traditional face detection algorithms like Viola-Jones detector [14] achieves real-time face detection of human using different feature descriptors [15], such as Haar-like features, histogram of oriented gradients (HOG), and linear binary pattern (LBP), to extract features and output the coordinates of the bounding box of a face. On one hand, Haar-like features method uses different rectangular filters to check the presence of a face. It has an integral image method and cascading classifiers to quickly calculate the sum of pixel values, and ignore background and non-face objects, respectively. On the other hand, HOG method calculates the histograms of gradients to check edges and corners from the images for extracting useful features. Lastly, LBP method checks textures by comparing pixels and generating binary patterns. It is used for various computer vision tasks due to its ability to check illumination invariance from an image. As shown in Table I, HOG obtained the highest true positive rate (TPR) while the highest false negative rate (FNR) is observed in LBP. Although the traditional methods are widely used because of its low computational costs, it is limited in practical applications due to its relatively low accuracy results. To address this issue, DL is used to create more advanced methods and extract meaningful features that can be used in various applications like face and object recognition. One of the most popular DL-based face detection models is multi-



(a) (b)

Fig. 3. Comparison of (a) MTCNN and (b) DAN.

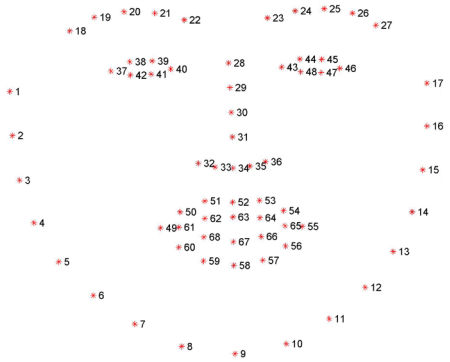


Fig. 4. Representation of 68 facial landmarks based on iBUG 300-W dataset annotations.

task cascaded convolutional neural network (MTCNN) [16]. MTCNN is composed of three cascaded CNNs, namely P-Net, R-Net, and O-Net, to provide the coordinates of bounding boxes and facial landmarks (i.e., eyes, nose, and two corners of the mouth) of the human face. As shown in Fig. 2, it solves issues with multiple posture variations, various human expressions, and lighting conditions, as tested on WIDER FACE dataset [17]. Another multi-stage CNNs for face alignment and detection is deep alignment network (DAN) [18]. DAN is a robust method that uses landmark heatmaps and feature images that transfers the landmark information between stages. As shown in Fig. 3, DAN generates a total of 68 facial landmarks, while MTCNN only generates 5. Thus, DAN can be used more effectively in complex application like in FER systems. However, since both MTCNN and DAN are built with multiple stages of CNNs, it requires high computational costs, which yield to low inference speed during real-time implementation. In this paper, a pre-trained 68-facial landmark detector trained in iBUG 300-W dataset (see Fig. 4) by Dlib [19] is used. The 68-facial landmark detector provides similar output like DAN but with reduced computational costs. It can be used to enable faster inference speed of FER system.

In line with the advancement of DL-based models, there are several facial expression or FER models that use CNN architecture to realize higher accuracy results. The traditional FER models [6] use K-nearest neighbor (KNN) and shallower network like artificial neural network (ANN). Both are used to classify human expression or emotion based on seven categories: 1) angry, 2) disgust, 3) fear, 4) happy, 5) sad, 6) surprise, and 7) neutral. These categories contribute to the complexity of building an FER model in addition to human diversity. Mostly, each emotion has huge similarity to other categories, which make it difficult to distinguish from other types of emotions. As a result, a non-parametric algorithm, such as a KNN-based model, cannot be utilized to effectively categorize the wide range of human emotions because it simply calculates and analyzes the distance between a sample and seven different emotional categories. Similarly, a shallower network, like an ANN-based model, is not efficient to extract significant features of several human emotions. The accuracy of ANN-based model reached up to 60% only, which is 10% higher than KNN-based model. For this reason, researchers focus more on developing FER models with deeper networks to learn more useful features. Deeper networks are generally leveraged using CNN architectures. It consists of millions of connections to extract and learn useful features that can be used to classify certain emotion based on the human face. [7] proves that increasing the depth of a network can improve the accuracy results of FER models up to 90%. The number of parameters within a CNN architecture is increased by applying more layers and using smaller kernels, which can learn more complex patterns. However, DL-based FER models suffer from low inference speed issue due to high computational cost requirement. To solve this issue, several researchers exerted efforts on developing lightweight FER models. A lightweight FER model is an optimized algorithm to support low computational cost and high inference speed. Some studies benchmark commonly used pre-trained models for FER task by transfer learning. They compare VGG-Face [9], VGG-16 [10], and Deep Face CNN [10] in terms of computational and memory costs. Although the existing pre-trained models can provide high generalizations, their requirements remained high. In addition, deeper networks can easily have problem of overfitting due to considerable number of parameters. Hence, key to lightweight FER models is to significantly reduce the parameters without affecting the accuracy. Most of the parameters are observed in the last fully connected layers of the CNN models. For instance, in the VGG-Net [20], the last fully connected layers include 90% of the total parameters. To reduce the parameters, Inception V3 [21] replaces the last fully connected layers with Global Average Pooling operation, which takes the average of the elements in feature image. To further reduce the parameters, an extreme version of Inception V3, known as Xception [22], is developed. Xception utilizes deep residual learning and depth-wise separable convolutions to separate the feature extraction and composition processes. Most of the efficient FER models developed nowadays are based on Xception architecture [23]. Although the computational cost of FER system is reduced by leveraging with Xception architecture, further optimizing this is still a challenge. In this paper, we proposed a lightweight FER model by applying different network optimization methods.

B. Network Optimization Methods

There are several network optimization methods that can effectively reduce CNN models, realize faster inference speed, and lower computational costs. In this paper, we used different optimization methods, such as pruning, quantization, and DL compiler, to reduce the complexity and requirements of Xception model for the FER task.

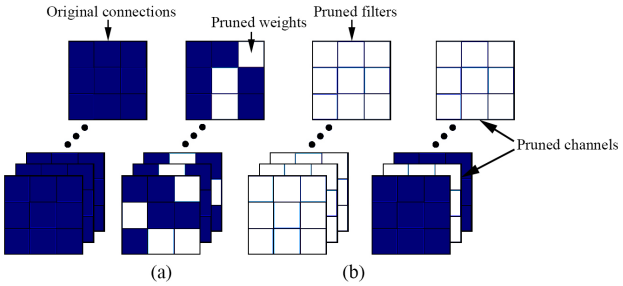


Fig. 5. Comparison of (a) unstructured pruning and (b) structured pruning. Note that unstructured pruning prunes weights while structured pruning prunes channels and/or filters.

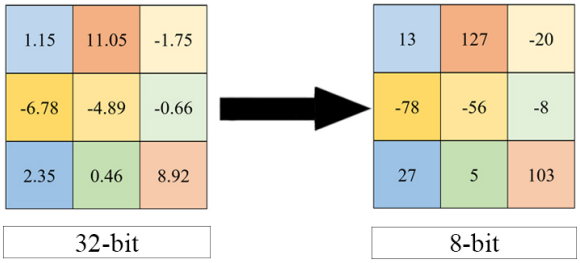


Fig. 6. General process of quantization. Sample of converting 32-bit floating-point number to 8-bit integer.

Pruning refers to an effective search and elimination of unimportant and redundant connections within the CNN model [11]. The weights, filters, and channels are the connections in the CNN model that build up its enormous complexity. By applying pruning method, the complexity of the model can be reduced. As a result, the model's memory and computational costs will be reduced during implementation. Pruning can be categorized as unstructured and structured. Unstructured pruning prunes redundant weights of a CNN model in fine-grained nature, as shown in Fig. 5(a). It effectively reduces the computational costs of the model. On the other hand, structured pruning refers to pruning of insignificant channels and/or filters of a CNN model in coarse-grained nature, as shown in Fig. 5(b). It can leverage the hardware parallelism of the device in expense of the substantial reduction in the storage footprint of the model. For comparison, unstructured pruning is more flexible than structured pruning in searching for optimized pruning structure. So, the former is easier to implement and typically achieves higher compression rate without affecting the accuracy results. In this paper, we used unstructured weight pruning to reduce the computational costs of Xception model while securing zero or negligible accuracy loss.

Meanwhile, **quantization** refers to the conversion of the bit representation of each weight from the CNN architecture into lower precision format [12]. Since CNN models are stored in 32-bit floating-point (FP32), it can be converted to either FP16 or even lower integer (INT) like INT8 and INT4 by using quantization. As a result, the memory usage is reduced, and the inference speed is increased. Quantization can be categorized as uniform and non-uniform. A uniform quantization has an equal step size or quantization level, which provides significant acceleration to the model. On the other hand, a non-uniform quantization does not have an equal

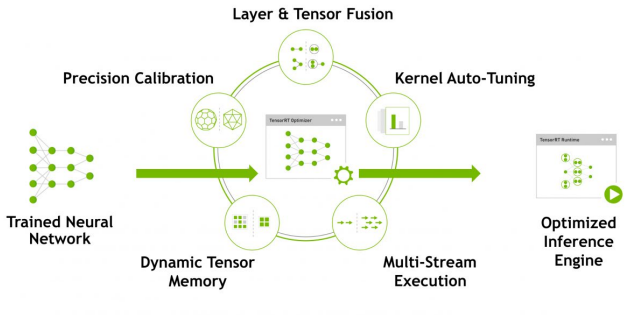


Fig. 7. Advanced performance optimizations using NVIDIA TensorRT.

quantization level, but usually achieves higher compression rate. The general process of quantization is shown in Fig. 6. In this paper, we used a uniform quantization to enable high inference speed to the FER system.

Lastly, a **DL compiler** is used to leverage with the hardware architecture of the model [13]. It can optimize models by redesigning and compiling its architecture to easily access hardware optimizations for faster inference. It includes advanced performance optimization, varied computation graph optimization, tensor optimization, and half precision format support. There are various DL compilers that can be used to assist inference acceleration, such as TensorFlow-Lite (TFLite), Alibaba Mobile Neural Network (MNN), Open Neural Network Exchange (ONNX), and NVIDIA TensorRT. These DL compilers have similar processes of optimizing CNN models. An example is shown in Fig. 7, which represents the advanced performance optimizations using NVIDIA TensorRT. In this paper, a DL compiler is utilized to compile the Xception model into an efficient end-to-end framework in order to improve the inference performance of the FER system.

Ⅲ. Proposed Lightweight FER System

The schematic diagram of the proposed lightweight FER system is shown in Fig. 8. As shown in the figure, there are three main processes in the proposed lightweight FER system: 1) face detection of the input image, 2) application of network optimization to the FER model, and 3) facial emotion classification.

First, for the face detection of the input image, we used a pre-trained landmark detector from Dlib library to identify 68 key points or facial landmarks marked at certain x and y coordinates in the human face. The key points localize the region around the human face, such as the eyebrows, eyes, nose, mouth, chin, and jaw. The Dlib 68-landmark face detector also provide the coordinates of a bounding box enclosing the human face. It is trained on the iBUG 300-W dataset, which is built by the Intelligent Behavior Understanding Group (iBUG) at Imperial College London. As shown in Fig. 9, the iBUG 300-W dataset contains "in-the-wild" images collected from the internet and corresponding 68 facial landmarks and bounding box annotations. It has more than 4000 static images, each having single face of various poses, expressions, and illuminations. The trained model is tested on 300-W test dataset collected also "in-the-wild" images, 300 indoor and 300 outdoor, each image has multiple

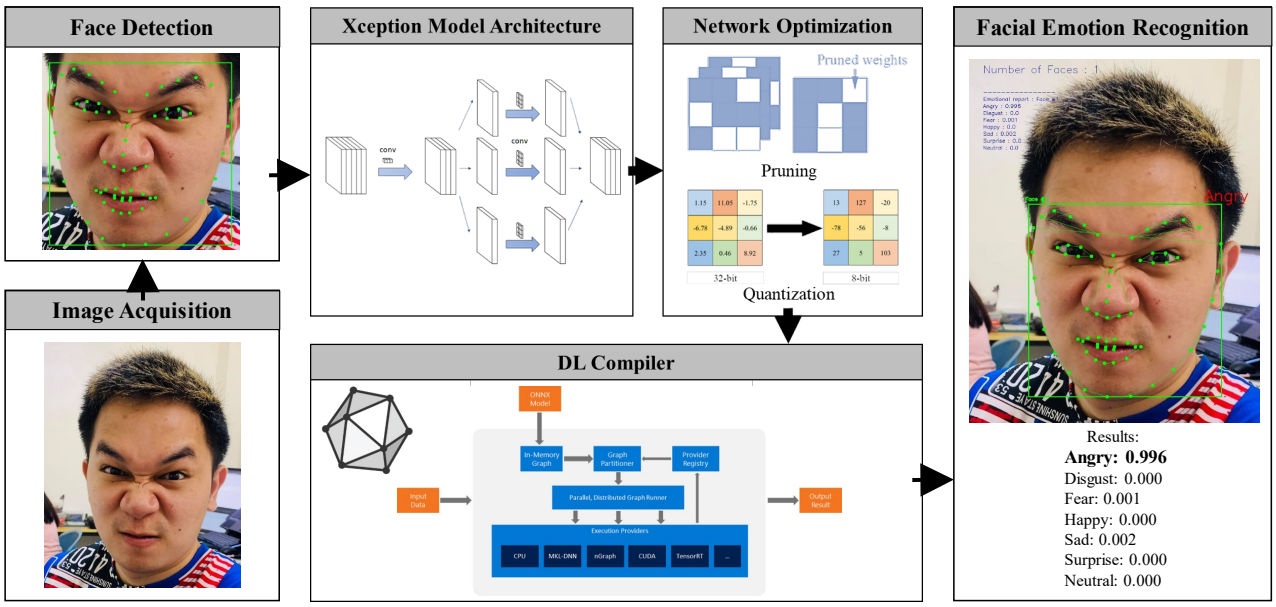


Fig. 8. Schematic diagram of the proposed lightweight FER system.

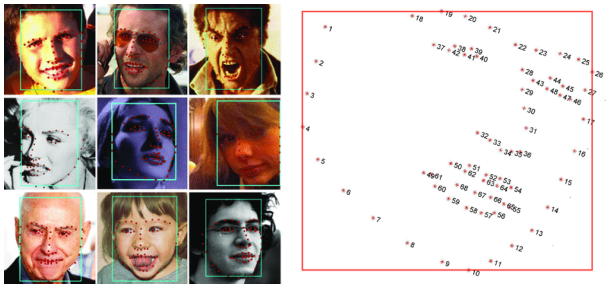


Fig. 9. Samples of iBUG 300-W dataset with annotations.



Fig. 10. Samples of 300-W test dataset (outdoor – left, indoor – right).

faces and huge facial variations, see samples in Fig. 10. The pre-trained face detection model has an efficient inference performance in detecting faces based on 300-W test dataset. For this reason, the Dlib 68-landmark face detector is used in this paper to extract useful patterns from several types of emotions in human faces. It can provide an efficient calculation to distinguish one specific emotion to another for an accurate FER system.

Second, for the application of network optimization to the FER model, we used pruning, quantization, and DL compiler to the Xception model to develop lightweight FER system. An Xception model has better accuracy and relatively smaller parameters than the existing CNN models (e.g., VGG-Net and ResNe50) as trained and tested on FER2013 dataset (see

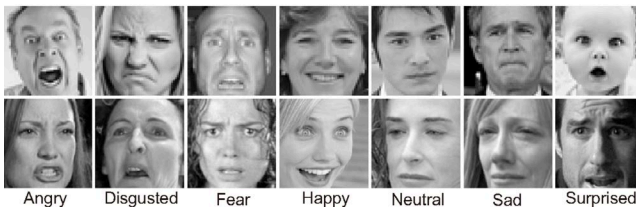


Fig. 11. Samples of FER2013 dataset.

samples in Fig. 11). However, it still has issues with high computational costs, high memory usage, and low inference speed during FER implementation. In this paper, we utilized several optimization methods that targeted both software and hardware components of the FER system to solve the issues.

Initially, we used unstructured pruning method to reduce the total number of non-zero parameters of Xception model by removing redundant and insignificant parameters. This optimization method will significantly reduce the computational costs of the model. Typically, the compression ratio of the pruned CNN architecture relative to the uncompressed is given by

$$R = \frac{\sum_{i=1}^{L} U_{w,f,c}{}^{(i)}}{\sum_{i=1}^{L} P_{w,f,c}{}^{(i)}}, \tag{1}$$

where R is called the compression ratio between uncompressed (U) and pruned models (P); $U_{w,f,c}{}^{(i)}$ and $P_{w,f,c}{}^{(i)}$ represents the total number of weights (w), filters (f), and/or channels (c) up to layer L of the uncompressed and pruned CNN architecture, respectively.

Then, we used a uniform quantization method to reduce the memory usage and increase the inference speed of the Xception model by converting its high precision format from FP32 to FP16 or INT8 low precision format. General weights quantization is computed by

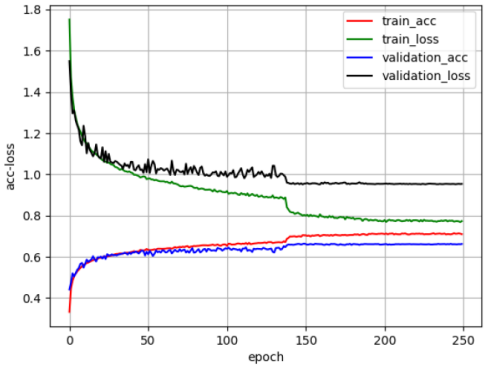


Fig. 12. Accuracy and loss curve of [23] based on FER2013 dataset.

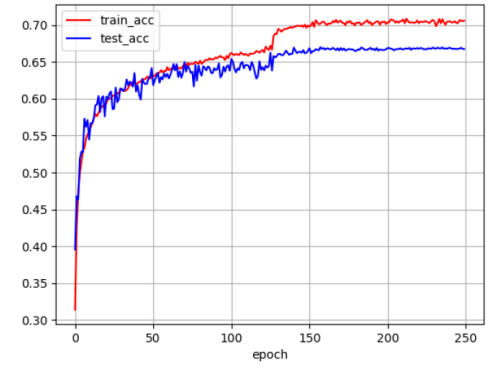


Fig. 13. Train and test accuracy of [23] based on actual test set.

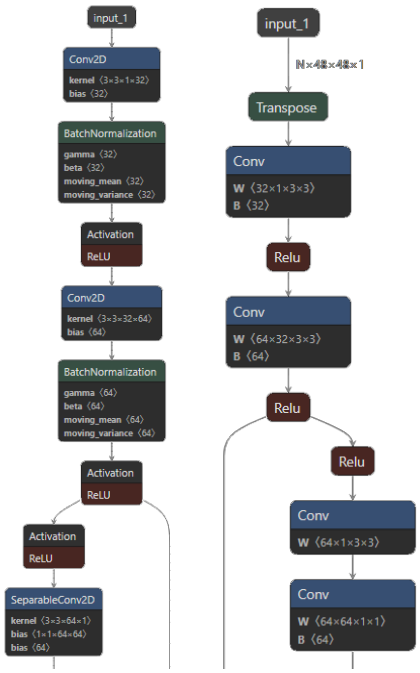


Fig. 14. Representation of the proposed lightweight FER model in (a) Keras format and (b) ONNX format.

$$W_b = round(S_b W_{fp}) \in \left[-\left(\frac{2^b-1}{2}\right), \left(\frac{2^b-1}{2}\right)\right] \quad (2)$$

$$s_b = \frac{2^b-1}{2M_W} \quad (3)$$

$$M_W = max\left(abs(W_{fp})\right), \quad (4)$$

where W_b represents the quantized weight tensor, b is the desired lower-precision format, s_b is the quantization scale factor, W_{fp} is the original weight tensor in higher-precision floating-point format, and M_W is the absolute maximum weight from W_{fp}. The quantized weight tensor is obtained by the product of the quantization scale factor and each weight from the original weight tensor rounded to the nearest integer. The quantized weights are bounded by the symmetrical dynamic range of the desired lower-precision format. This will reduce the memory usage of Xception model by more than 50% and slightly increase its inference speed.

The general problem of applying pruning and/or quantization is due to the difficult search of the compression hyperparameters in each layer of the CNN model. There are several approaches that can be used to effectively optimize the hyperparameters. One of these approaches is known as a heuristic method or manual tuning of the hyperparameter per layer of the CNN architecture. Another one is a black box optimization or Bayesian optimization method, which is an automatic hyperparameter-search approach. However, these approaches are inefficient because they rely only on repeated trial-and-error procedures to determine the ideal hyperparameters in a layerwise manner, which frequently results in a significant accuracy loss. To solve this problem, we implemented a constrained approach of pruning and quantization methods to minimize any possible accuracy degradation after optimization. This approach will not consider any compression hyperparameters to eliminate the risk of having low accuracy results.

Finally, we employed a DL compiler to further optimize and increase the inference speed of Xception model. Each framework (e.g., PyTorch, Keras, and TensorFlow) has its own distinct representation of a computation graph which often leads to a restriction in using a model from one framework to another. We solved this issue by using an Open Neural Network Exchange Format (ONNX), which is designed to standardize layer definitions of a network and support most of deep learning model formats. In this paper, a Keras framework is employed to train and optimize Xception model. Then an existing Keras to ONNX converter is utilized to reconstruct the network graph with the equivalent operators on ONNX format. This will allow a performance-focused inference optimizer, called ONNX Runtime, to automatically utilize the available hardware accelerators and runtime on the host device. Hence, improving the performance of the model. This inference engine partitions the execution graph into subgraphs and run each subgraph on the most efficient execution provider such as CUDA and TensorRT. By applying pruning, quantization, and DL compiler, a lightweight FER model is developed, as shown in Fig. 8.

Lastly, for the facial emotion classification, the proposed lightweight FER model shows better performance when implemented in real-time inference, as shown in the last diagram in Fig. 8.

TABLE II
Actual Test Results of The Proposed Lightweight FER Model

Lightweight FER Model	Angry	Disgust	Fear	Happy	Sad	Surprise	Neutral
Angry	**0.996**	0.000	0.001	0.000	0.002	0.000	0.000
Disgust	0.002	**0.997**	0.000	0.000	0.000	0.000	0.000
Fear	0.003	0.000	**0.839**	0.000	0.128	0.027	0.003
Happy	0.001	0.000	0.001	**0.992**	0.000	0.005	0.002
Sad	0.160	0.001	0.029	0.003	**0.414**	0.001	0.391
Surprise	0.004	0.000	0.035	0.003	0.000	**0.957**	0.000
Neutral	0.156	0.000	0.050	0.008	0.271	0.004	**0.511**

TABLE III
Comparison of Different FER Models

Model	Train set accuracy (%)	Test set accuracy (%)
VGG-Net	98.98	59.32
ResNet-50	98.87	57.48
CNN	99.7	58.9
HOG+ CNN [41]	-	61.86
Improved Inception [42]	-	66.41
Network from [43]	-	66.4
CNN-based+ Softmax [45]	-	65.03
ShallowNet [47]	-	63.49
Lightweight FER Model	**71.00**	**67.00**

TABLE IV
Hardware and Inference Performance of Different FER Models

Model	Parameters	CPU (%)	MEM (%)	FPS
VGG-Net	87566680	21.50	3.2	-
ResNet-50	25500000	49.78	2.9	-
CNN	95263	10.83	12.0	-
Lightweight FER Model	**58423**	**17.50**	**3.1**	**21**

IV. Experiment

In this section, we evaluated the performance of the proposed lightweight FER model. First, the loss and accuracy results of training the lightweight FER model is demonstrated. Then, the performance of the lightweight FER model is compared to other existing FER systems in terms of the computational costs, memory usage, and inference speed.

The Xception model is trained on FER2013 dataset for a total of 250 epochs. The model is being optimized, using the pruning and quantization methods, and trained simultaneously. Fig. 12 shows the potential change curves of the loss and accuracy values after 250 epochs of training and optimizing the model. As can be seen in the figure, the train accuracy reached 71% while the validation accuracy achieved 67% as tested on FER2013 dataset. Fig. 13 shows the actual test accuracy of the model which reached up to 67% as validated on "in-the-wild" images from the internet. The results of the proposed lightweight FER model are comparable with these figures.

Furthermore, the model trained in Keras framework is compiled into ONNX format. Fig. 14 shows the difference in architectures of the proposed model. The ONNX format of the Xception model has fewer blocks as compared to Keras. This means that some of the computation graph within the architecture is efficiently optimized. Table II shows the actual test result of the proposed lightweight FER model across different emotion categories. The proposed model accurately classified the test images based on their specific emotion. The highest accuracy is observed to angry, disgust, and happy, while the lowest accuracy is observed to sad and neutral. The comparison of the accuracy of the proposed lightweight FER model to other existing FER models is shown in Table III. As can be seen in the table, the proposed model has the highest test accuracy result reaching 67%. Other models, such as VGG-Net, ResNet-50, and CNN, have overfitting problems since their accuracy on train set is close to 100% but their resulting test set accuracy only reached about 60%. Moreover, Table IV shows the hardware and inference performance of the proposed lightweight FER model. As shown in the table, the proposed model has the least number of parameters. In addition to that, the computational cost of the model based on CPU is second only to CNN, however, the significant difference in their memory usage makes the proposed model (3.1%) superior to CNN (12%). The inference speed measured in frame per second (fps) of the proposed model is relatively high (21 fps). Hence, the proposed lightweight FER model outperforms other existing models.

V. Conclusion

In this paper, we proposed a lightweight FER system that realizes an efficient inference performance during real-time implementation. We experimentally demonstrated that the application of network optimization methods, like pruning and quantization, can effectively reduce the computational costs and memory usage of the Xception model without affecting its accuracy results. Also, employing DL compiler, like ONNX, can further increase the inference speed of the model. This is by utilizing the half precision support and TensorRT execution of the compiler. Our obtained experimental results show that the proposed lightweight FER system, with the used of Dlib 68-landmark face detector, outperforms other existing FER models specially in terms of real-time inference. Specifically, our proposed system achieved smaller number of parameters (58423 only), low CPU and memory usage (17.50% and 3.1%), and relatively high inference speed reaching 21 fps, as compared to other existing FER systems.

ACKNOWLEDGMENT

This work was supported by the Ministry of SMEs and Start-ups, S. Korea (S2829065, S3010704), and by the National Research Foundation of Korea (2020R1A4A101777511, 2021R1I1A3056900).

References

[1] H. Liu, J. Zeng and S. Shan, "Facial Expression Recognition for In-the-wild Videos," 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020), 2020, pp. 615-618, doi: 10.1109/FG47880.2020.00102.

[2] K. Tarunika, R. B. Pradeeba and P. Aruna, "Applying Machine Learning Techniques for Speech Emotion Recognition," 2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2018, pp. 1-5, doi: 10.1109/ICCCNT.2018.8494104.

[3] R. S. Soundariya and R. Renuga, "Eye movement-based emotion recognition using electrooculography," 2017 Innovations in Power and Advanced Computing Technologies (i-PACT), 2017, pp. 1-5, doi: 10.1109/IPACT.2017.8245212.

[4] B. Xue, Z. Lv and J. Xue, "Feature Transfer Learning in EEG-based Emotion Recognition," 2020 Chinese Automation Congress (CAC), 2020, pp. 3608-3611, doi: 10.1109/CAC51589.2020.9327161.

[5] Chung, K.-M., Kim, S., Jung, W. H., & Kim, Y. (2019). Development and validation of the yonsei face database (yface db). Frontiers in Psychology, 10. https://doi.org/10.3389/fpsyg.2019.02626

[6] R. Ranjan and B. C. Sahana, "An Efficient Facial Feature Extraction Method Based Supervised Classification Model for Human Facial Emotion Identification," 2019 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT), 2019, pp. 1-6, doi: 10.1109/ISSPIT47144.2019.9001839.

[7] E. Pranav, S. Kamal, C. Satheesh Chandran and M. H. Supriya, "Facial Emotion Recognition Using Deep Convolutional Neural Network," 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), 2020, pp. 317-320, doi: 10.1109/ICACCS48705.2020.9074302.

[8] M. A. Ozdemir, B. Elagoz, A. Alaybeyoglu Soy and A. Akan, "Deep Learning Based Facial Emotion Recognition System," 2020 Medical Technologies Congress (TIPTEKNO), 2020, pp. 1-4, doi: 10.1109/TIPTEKNO50054.2020.9299256.

[9] J. Schwan, E. Ghaleb, E. Hortal and S. Asteriadis, "High-performance and lightweight real-time deep face emotion recognition," 2017 12th International Workshop on Semantic and Social Media Adaptation and Personalization (SMAP), 2017, pp. 76-79, doi: 10.1109/SMAP.2017.8022671.

[10] A. Atanassov and D. Pilev, "Pre-trained Deep Learning Models for Facial Emotions Recognition," 2020 International Conference Automatics and Informatics (ICAI), 2020, pp. 1-6, doi: 10.1109/ICAI50593.2020.9311334.

[11] H. Yang, S. Gui, Y. Zhu and J. Liu, "Automatic Neural Network Compression by Sparsity-Quantization Joint Learning: A Constrained Optimization-Based Approach," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020, pp. 2175-2185, doi: 10.1109/CVPR42600.2020.00225.

[12] F. Tung and G. Mori, "Deep Neural Network Compression by In-Parallel Pruning-Quantization," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 3, pp. 568-579, 1 March 2020, doi: 10.1109/TPAMI.2018.2886192.

[13] Y. Cai, H. Li, G. Yuan, W. Niu, X. Tang, B. Ren, and Y. Wang, "YOLObile: Real-Time Object Detection on Mobile Devices via Compression-Compilation Co-Design," 2020 arXiv:2009.05697. [Online]. Available: https://arxiv.org/abs/2009.05697

[14] Y. Zhou, D. Liu and T. Huang, "Survey of Face Detection on Low-Quality Images," 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), 2018, pp. 769-773, doi: 10.1109/FG.2018.00121.

[15] A. Adouani, W. M. Ben Henia and Z. Lachiri, "Comparison of Haar-like, HOG and LBP approaches for face detection in video sequences," 2019 16th International Multi-Conference on Systems, Signals & Devices (SSD), 2019, pp. 266-271, doi: 10.1109/SSD.2019.8893214.

[16] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," IEEE Signal Processing Letters, vol. 23, no. 10, pp. 1499-1503, 2016.

[17] S. Yang, P. Luo, C. C. Loy and X. Tang, "WIDER FACE: A Face Detection Benchmark," In Proc. of IEEE Conf. CVPR, 2016.

[18] M. Kowalski, J. Naruniec and T. Trzcinski, "Deep Alignment Network: A Convolutional Neural Network for Robust Face Alignment," 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2017, pp. 2034-2043, doi: 10.1109/CVPRW.2017.254.

[19] Y. Liu, Z. Xu, L. Ding, J. Jia and X. Wu, "Automatic Assessment of Facial Paralysis Based on Facial Landmarks," 2021 IEEE 2nd International Conference on Pattern Recognition and Machine Learning (PRML), 2021, pp. 162-167, doi: 10.1109/PRML52754.2021.9520746.

[20] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, arXiv:1409.1556. [Online]. Available: https://arxiv.org/abs/1409.1556

[21] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2818-2826, doi: 10.1109/CVPR.2016.308.

[22] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 1800-1807, doi: 10.1109/CVPR.2017.195.

[23] N. Zhou, R. Liang and W. Shi, "A Lightweight Convolutional Neural Network for Real-Time Facial Expression Detection," in IEEE Access, vol. 9, pp. 5573-5584, 2021, doi: 10.1109/ACCESS.2020.3046715.

시계열 패턴 군집분석을 기반으로 LSTM을 이용한 단기 전력 수요 예측

오민지, 김주하, 엄지성, 최은선, 김재성, 조완섭*
충북대학교 빅데이터학과, *충북대학교 경영정보학과
minji.oh1020@gmail.com, wldus1907@gmail.com, ejs0414@naver.com, tmxk147@gmail.com,
comkjsb@gmail.com, *wscho63@gmail.com

Short-Term Load Forecasting using Time Series Clustering and LSTM

Min Ji Oh, Joo Ha Kim, Ji Seong Eom, Eun Seon Choi, Jae Sung Kim, Wan Sup Cho*
Chungbuk National Univ. Department of Bigdata Science,
*Chungbuk National Univ. Department of Management Information

요 약

고도화된 산업으로 인해 우리나라의 전력사용량은 증가하고 있어 전력사용량 예측에 대한 요구가 있으며, 정부의 스마트팩토리 추진 목표에 맞추어 설비가동률을 높이는 기업이 증가하고 있다. 스마트팩토리 운용에서 전기요금 감축과 에너지의 효율적 운용을 위한 전력 소비 패턴 관리 및 전력사용량 예측에 대한 중요성이 대두됨에 따라 본 연구에서는 건물의 특성과 요일에 따른 변화 및 시간대에 따른 특성을 시계열 분석에 반영하는 방안을 제안하였다. 형태 기반 거리척도인 DTW를 사용하여 요일과 시간대에 따라 시계열 데이터를 군집화한 뒤 LSTM 예측모델을 적용하여 군집화에 따른 전력 수요량 예측성능을 비교한 뒤 군집에 따른 전력 패턴의 차이를 분석하였다. K-means 군집화에 따라 군집의 개수를 2개로 설정한 뒤 군집의 특성을 반영한 LSTM모델의 예측력이 가장 뛰어났음을 확인하였다.

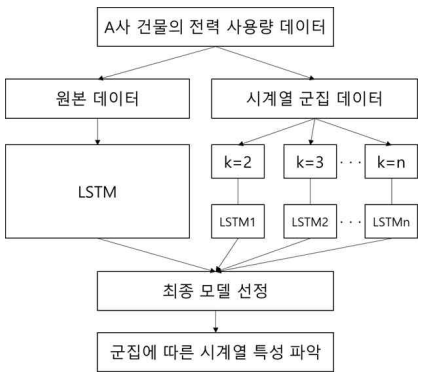
Ⅰ. 서 론

고도화된 산업으로 인해 우리나라의 전력사용량은 매년 증가하고 있다. 또한 정부의 스마트팩토리 추진 목표에 따라 국내 중소·중견기업에 스마트팩토리가 보급되면서 설비가동률을 높이는 기업이 증가하고 있다[1]. 스마트팩토리 운영에서 에너지 소비 감축은 주요 이슈로 떠오르고 있으며 건물의 효율적인 운용을 위한 에너지 소비 패턴 관리 및 예측에 대한 요구가 있다[2]. 건물의 에너지를 효율적으로 운용하려면 건물의 전력사용 패턴을 파악하고 이를 적절하게 분산시킬 필요가 있다. 특히 산업시설에 부과하는 전기요금은 전기 사용이 발생한 시간대와 요일을 구분하여 중부하, 최대부하, 경부하대로 나누어 요금을 산정하기 때문에 전력 사용 계획에 따른 전력 소비 분배는 생산비용을 낮추는 데 기여할 수 있다. 동일한 양의 전력을 사용하더라도 사용시간 및 사용패턴에 따라 전력요금이 달라지기 때문에 에너지의 효율적인 운용을 위해 시간적 패턴을 고려하여 전력 소비량을 예측해야하며 최대부하 시간대에 전력량을 줄이는 노력이 요구된다. 정확한 전력사용량을 예측하기 위해서는 건물의 개별 특성과 요일에 따른 변화, 시간대에 따른 특성을 반영해야한다. 하지만 이러한 요소들은 변화가 크기 때문에 정상성을 유지하기 어려운 특징이 있어 일반적인 통계적 시계열 분석에 반영하기 어렵다고 알려져 있다[3]. 본 연구에서는 요일와 시간대에 따른 시계열 특성을 반영하기 위해 형태 기반 거리척도인 DTW를 사용하여 데이터를 군집화한 뒤 LSTM 예측 모델을 적용하여 군집화에 따른 전력 수요량 예측 성능을 비교한 뒤 군집에 따른 전력 패턴의 차이를 분석한다.

Ⅱ. 연구 방법

1.분석프로세스

본 논문에서는 A사 건물의 전력 사용량 데이터를 활용하여 전력 사용량의 시계열 패턴을 기반으로 전력 수요 예측 분석을 수행하고 군집에 따른 전력 사용량 시계열 특성을 파악하는 연구를 제안한다. 분석 구조도는 <그림1>과 같다. 원본데이터를 LSTM 모델에 적용한 전력 수요 예측과 시계열 데이터를 군집 개수(k)에 따라 나누어 LSTM 모델을 적용한 전력 수요 예측을 수행한 뒤 각 예측모델의 성능을 비교하여 최종 모델로 선정한다. 마지막으로 군집에 따른 시계열 데이터의 차이를 비교 및 분석한다.



<그림1> 분석 구조도

2.DTW

DTW는 Dynamic Time Warping 의 줄임말로 형태를 기반으로 하는 탄력적 거리 측정 방식이다. 시계열 그래프 패턴 사이의 거리를 측정할 때 시간 축을 확장하거나 수축하는 방식으로 이를 왜곡하여 시계열 패턴 사이의 거리를 최적정렬 방법으로 찾는 알고리즘이다[4]. 주로 음성인식 분야에서 사용되며, 두 개의 시퀀스의 길이가 달라도 유사도를 측정할 수 있다는 장점이 있다. 시계열 군집분석에서는 거리 측정 방식으로 유클리디안(Euclidean)을 사용하지 않고 ETW를 사용한다. 개별 관측치 간의 동질성에 따라 거리를 측정하는 유클리디안 방식과 달리 DTW는 각 지점과 가장 유사한 위상의 점을 찾아 시계열 데이터의 시퀀스를 비교적 정확하

게 측정할 수 있다[5].

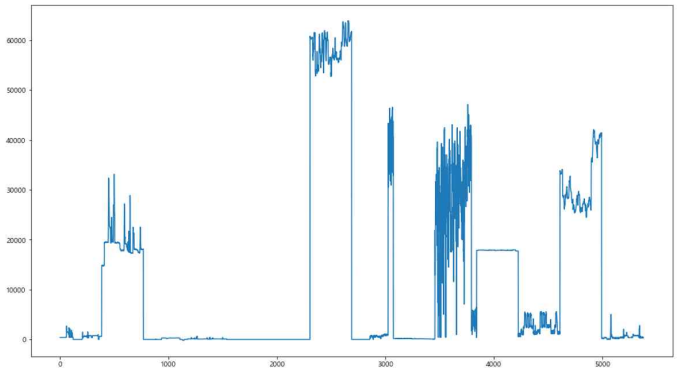
3.LSTM

LSTM은 Long Short Term Memory의 줄임말로 순환신경망(RNN)의 일종이다. 주기성을 가진 시계열 패턴을 예측하는데 사용되며 과거의 시계열 데이터와 현재의 데이터 조합으로 미래 데이터에 영향을 줄 수 있는 순환구조 형태를 지니고 있다[6]. 순환구조를 가진 특성으로 인해 음성인식, 문자, 시계열 예측 등의 분야에서 사용되고 있다.

Ⅲ. 분석 결과

1.데이터설명

본 논문에서 활용한 데이터는 A사 건물의 2015년 5월 3일부터 2015년 7월 28일까지 25일간 기록된 전력 사용 데이터이다. 센서에 의해 수집된 5분 단위 데이터를 1시간 단위로 합산하여 시간대별 시계열 데이터로 구성하였다. 전력 사용 데이터의 평일과 주말의 비율은 5:5로 구성되었다. 주7일에서 평일과 주말이 차지하는 비율이 5:2인 것을 고려하였을 때 연구에 사용된 데이터는 주말의 비율이 높은 것이라 판단할 수 있다. 또한 <그림2>를 통해 해당 기간의 부하유효전력(W) 사용량의 시계열 분포를 확인하였다. 탐색적 데이터 분석을 통해 전력 사용량이 10,000 W 이상인 구간과 10,000 W 미만인 구간으로 나누어진다는 것을 알 수 있었다. 특정 기간에 따라 데이터의 시계열 분포가 달라진다는 점을 통해 A사 건물의 전력 사용량 데이터는 시계열 군집화 기법으로 전력 수요 예측을 하는 것이 적합할 것이라고 판단하였다.



<그림2> 전력 사용량의 시계열 분포

전력 수요 예측을 위한 시계열 군집화를 수행하기 위해 패턴 분류에 사용한 입력 변수는 <표1>과 같다. 연구에 사용한 데이터는 1년이 채 되지 않는 기간 동안 수집된 데이터이며 수집 기간이 5월에서 7월 사이이므로 '월' 변수와 '온도' 변수 데이터 포인트마다 차이가 두드러지지 않았기 때문에 패턴 분류의 입력 변수로 사용하지 않았다. <그림2>에서처럼 부하유효 전력은 시간의 흐름에 따라 패턴을 보이므로 패턴 분류의 입력 변수로 사용하면서 하루를 기준으로 부하 유효 전력을 누적 계산한 부하누적 전력량(Wh)을 입력 변수에 추가하여 군집분류의 타당성을 높이고자 하였다. 또한 요일에 따른 패턴을 추적하기 위해 평일과 주말 여부를 확인할 수 있는 패턴 정보(Label)을 추가하였다.

입력 변수	설명
부하유효전력 (W)	전원으로부터 공급되어 부하에서 유효하게 이용되는 전력
부하누적전력량 (Wh)	하루를 기준으로 한 누적 전력량 값 (매일 오전 0으로 리셋됨)
주말 여부	평일:0, 주말:1

<표1> 군집화에 사용한 패턴 입력 변수

전력 사용량 예측모델에 사용된 변수는 <표2>와 같다. 전력 사용량 예측에 있어 군집화 수행 유무가 예측 성능에 영향을 주는지 확인하기 위해 부하유효전력과 군집화를 통해 부여한 군집번호를 변수로 활용하였다. 데이터가 수집된 기간이 짧아 데이터의 총 레코드가 8,064개 행인 점을 감안하여 과거 3시간 동안의 전력 사용량 데이터와 군집번호를 바탕으로 1시간 이후의 전력 사용량을 예측하는 방법을 사용하였다.

변수	설명	비고
전력사용량	1시간 이후의 전력 사용량	종속변수
전력사용량 (t, t-1, t-2, t-3)	현재 및 과거 3시간의 전력 사용량	독립변수
군집번호	군집화 결과 번호	

<표2> 전력 사용량 예측 모델의 변수

2.성능평가지표

본 논문에서 예측모델의 성능 평가지표로 사용한 것은 MAE, RMSE, R2이다. MAE는 식<1>과 같이 실제 값과 예측 값의 차이를 절대값으로 변환해 평균화한 것으로서 오차의 크기가 그대로 반영된다는 장점이 있다. RMSE는 식<2>와 같이 실제값과 예측값의 차이를 제곱하여 평균을 계산한 MSE의 제곱근을 계산한 것이다. MSE는 오류의 제곱을 구하므로 실제 오류 평균보다 더 커지는 문제가 있어 이를 해결하기 위해 RMSE를 사용하였다. R2는 결정계수라고 불리며 추정한 모형이 주어진 자료에 적합한 정도를 계산하는 척도이다. 식<3>과 같이 R2는 종속변수 y의 분산 중에서 적합된 모형이 설명가능한 부분의 비율을 의미하므로 적합된 모형의 설명력을 측정할 수 있다.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - y_i'| \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^{T}(y_t - \hat{y}_t)^2}{T}} \quad (2)$$

$$R^2 = \frac{\sum_i(\hat{y}_i - \bar{y})^2}{\sum_i(y_i - \bar{y})^2} \quad (3)$$

3.실험결과

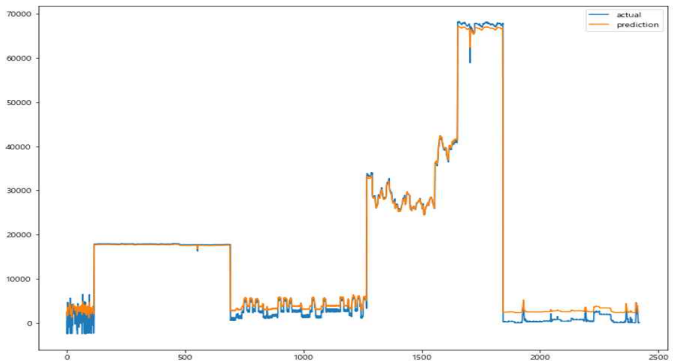
군집화의 개수에 따른 예측 성능을 비교하기 위해 군집화를 진행하지 않은 데이터와 k=2, 3, 4, 5 로 설정하여 군집화를 수행한 데이터의 성능을 비교하였다. 시계열 군집분석은 DTW를 이용한 Time Series K-Means 기법을 사용하였으며, 전력 수요량 예측 알고리즘은 LSTM을 사용하였다. 군집의 개수를 2에서 5 사이로 설정하여 군집 개수에 따른 예측 성능을 비교하였다. 예측 성능의 개선 여부를 확인하기 위해 군집화를 사용하지 않은 데이터 학습 결과를 기준점으로 사용하였다. <표3>과 같이 군집분석을 수행한 예측모델은 모든 k에 대해 RMSE는 기준점보다 낮고, R2는 기준점보다 높다는 것을 확인하여 군집분석을 통한 시계열 예측이 효

과적임을 입증하였다. 가장 예측성능이 뛰어난 모델은 군집을 2개로 나눈 뒤 군집번호를 독립변수로 반영하여 LSTM에 적용한 것으로 드러났다.

	기준	k=2	k=3	k=4	k=5
MAE	1187.13	1050.24	1479.08	1694.90	1658.92
RMSE	3005.30	1560.39	2379.09	2792.62	2976.78
R2	0.94	0.99	0.98	0.97	0.97

<표3> 군집화에 따른 전력 사용량 예측 성능 비교

<그림3>은 군집의 개수(k)를 2로 설정한 뒤 평가 데이터로 실제값과 예측값의 일치여부를 시각적으로 확인한 것이다. 시간의 흐름에 따른 전력 사용량의 변화를 정확하게 잡아내었다는 것을 알 수 있다.



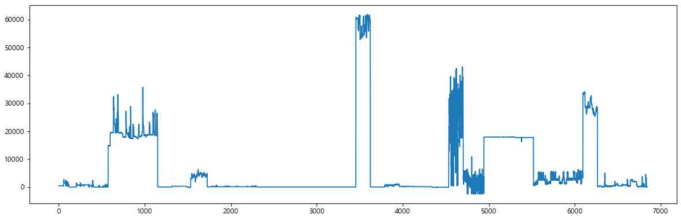
<그림3> 전력 수요량 실제값과 예측값 비교

두 개의 군집으로 나눈 어떤 차이를 보이고 있는지 알아보기 위해 군집1과 군집2의 전력 사용량에 대한 기술통계량을 <표3>과 같이 확인하였다. 두 군집은 평균 전력사용량과 전력사용량 중앙값에서 큰 차이를 보였다. 그러므로 시간에 따른 전력 사용량 패턴에서 구분되는 차이점이 발견되었을 때 적절한 패턴 군집화 방법을 적용하는 것이 전력 수요량을 정확히 예측하는 데 기여할 수 있다고 판단할 수 있다.

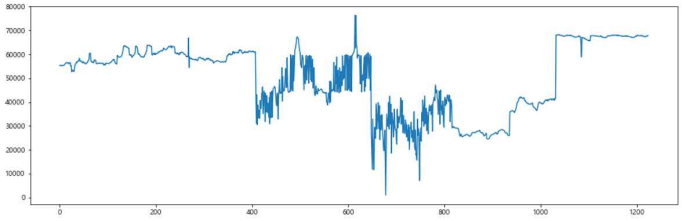
	군집1	군집2
비율	전체의 84%	전체의 16%
평균	6469 W	50047 W
중앙값	327 W	56218 W
표준편차	11929	14395

<표3> 군집에 따른 기술통계량 비교

군집1에 속하는 데이터의 시계열 그래프 <그림4>와 군집2에 속하는 데이터의 시계열 그래프 <그림5>를 비교하였다. <그림4>는 전력 사용량이 10,000 W를 넘지 않는 데이터가 다수인 반면 <그림5>는 전력 사용량이 10,000 W 이상인 데이터가 분포하고 있음을 확인할 수 있다.



<그림4> 군집1의 전력 사용량 시계열 그래프



<그림5> 군집2의 전력 사용량 시계열 그래프

Ⅳ. 결론

본 논문에서는 요일와 시간대에 따른 시계열 특성을 반영하기 위해 형태 기반 거리척도인 DTW를 사용하여 데이터를 군집화한 뒤 LSTM 예측 모델을 적용하여 군집화에 따른 전력 수요량 예측 성능을 비교한 뒤 군집에 따른 전력 패턴의 차이를 분석하였다. 원본데이터를 LSTM 모델에 적용한 전력 수요 예측 모델과 군집 개수(k)에 따라 시계열 데이터를 나누어 LSTM 모델을 적용한 전력 수요 예측 모델의 성능을 비교한 결과, 군집을 2개로 나누어 LSTM을 수행한 모델의 예측 성능이 가장 우수했다. 두 개 군집의 기술통계량을 비교한 결과 평균 전력 사용량과 전력사용량의 중앙값에서 유의한 차이가 있음을 확인하였다. 이처럼 시계열 패턴 사이에 차이점이 발견되었을 때 적절한 패턴 분류 방법을 적용하여 학습 모델을 생성한다면 개별 건물 특성과 시간적 특성을 반영할 수 있어 좀더 정확한 전력 수요량 예측을 가능하게 할 수 있다고 판단할 수 있다. 다만 연구에 사용된 데이터의 수집 기간이 1년이 되지 않는다는 점에서 다양한 계절적 특성을 반영할 수 없었다는 것이 한계점으로 남아 있다. 향후 데이터 수집 기간을 늘려 계절적 특성을 반영할 수 있는 시계열 패턴 분석을 수행하여 연구의 확장성을 이어나갈 것이다.

ACKNOWLEDGMENT

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2020-2018-0-01396) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation)

참 고 문 헌

[1] 정종필, 신광섭(2020). 융합연구리뷰. 융합연구정책센터. Vol6.

[2] 박재연, 허태욱, 고정길(2021). 한국통신학회 학술대회논문집, p132-133.

[3] Yang, L., & Yang, H(2019). Analysis of different neural networks and a new architecture for short-term load forecasting. Energies, 12(8), 1433.

[4] 이예담, 임재현, 최동구(2020). 형태 기반 군집화와 랜덤포레스트를 이용한 개별 건물의 단기 전력 수요 예측. 경영과학 37(4), 21-31.

[5] Paparrizos, J., & Gravano, L.(2015). k-shape: Efficient and accurate clustering of time series. In Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data (pp. 1855-1870).

[6] 김진석, 김기일(2019). 전력 수요관리를 위한 전력사용량 패턴예측에 관한 연구. 한국통신학회 학술대회논문집.

모바일 애드혹 네트워크 구축을 위한 Q-learning 기반 노드 위치 결정 전략 연구

김나영°, 권민혜*, 박형곤°

이화여자대학교 전자전기공학과 스마트팩토리융합전공°, 숭실대학교 전자정보공학부*

nayoung.kim@ewhain.net, minhae@ssu.ac.kr, hyunggon.park@ewha.ac.kr

Q-learning Based Node Positioning Strategy for Mobile Ad-hoc Network Formation

Nayoung Kim°, Minhae Kwon*, Hyunggon Park°

Graduate Program in Smart Factory, Dept. of Electronic and Electrical Engineering, Ewha Womans University°

School of Electronic Engineering, Soongsil University*

요 약

드론을 비롯해 이동 가능한 노드로 구성된 모바일 네트워크가 제시되며 이동성 제어의 중요성이 커지고 있다. 본 논문은 Q-learning을 기반으로 모바일 무선 노드의 이동성을 제어하여 높은 네트워크 처리율을 유지하는 애드혹(ad-hoc) 네트워크를 형성하는 방안을 제안한다. 제안한 방안은 한정된 주변 노드 정보를 바탕으로 릴레이 노드가 초기 고정 위치에서 시작하여 소스 노드-목적 노드 간 네트워크를 구축하는 상황을 가정한다. 시뮬레이션 결과를 통해 모바일 무선 노드가 분산적으로 이동성을 제어하여 애드혹 네트워크 구축을 위한 각자의 역할을 학습함을 확인하였다.

Ⅰ. 서 론

최근 무인 항공기와 드론을 사용한 애드혹 네트워크(Flying Ad-Hoc Network, FANET)와 같은 모바일 애드혹 네트워크(Mobile Ad-Hoc Network, MANET)의 활용 가능성이 높아지며, 재난 상황에서의 비상 네트워크[1]나 산불 예측 및 감시[2] 등에 적용하는 연구가 제안되고 있다.

효율적인 MANET은 적은 수의 모바일 기기를 사용해 넓은 통신 영역을 포함할 수 있어야 한다[3]. 모바일 기기의 송신 에너지와 이에 따른 송신 범위는 한정되어 있으므로 데이터를 소스 노드에서 목적 노드까지 릴레이 노드로 형성된 애드혹 네트워크를 통해 전송하기 위해서는 릴레이 노드들의 위치 결정이 중요하다. 일반적으로는 중앙 제어를 통해 최적의 노드 위치를 결정해왔으나, 이 경우 최적 네트워크를 결정하기 위한 모든 데이터가 가용 가능한 형태로 존재해야 한다. 또한, 노드 수가 증가함에 따라 필요로 하는 데이터의 양과 계산 복잡도가 크게 증가하므로 중앙 제어 없이 릴레이 노드의 이동성을 분산적으로 제어하는 방안이 제시되었다[4], [5].

본 논문에서는 분산적으로 노드 위치를 결정하기 위하여 강화학습(reinforcement learning) 기반으로 릴레이 노드의 이동성을 제어하는 알고리즘을 제안한다. 이를 통해 릴레이 노드들은 스스로 네트워크 상황을 판단하여 적절한 역할을 학습해 애드혹 네트워크를 형성할 수 있다.

Ⅱ. 이동성을 갖는 무선 노드의 Q-learning 기반 이동성 제어

Q-learning은 마르코프 결정과정(Markov Decision Process, MDP)으로 모델링하는 강화학습 알고리즘이다. MDP는 $< S, A, R, \gamma >$로 구성되며, S는 유한한 상태 공간(state space), A는 유한한 행동 공간(action space), R은 행동에 따른 보상(reward), γ는 감가율(discount factor)을 나타낸다. 학습 주체인 에이전트(agent)는 환경과 상호 작용해 누적 보상을 최대화하는 것을 목표로 학습한다.

2-1. 모바일 애드혹 네트워크 구성

본 논문은 고정된 소스 노드와 목적 노드, 이동성을 갖는 무선 릴레이 노드로 구성된 애드혹 네트워크를 고려한다. 이때 릴레이 노드는 학습 주체인 에이전트이다. 소스 노드와 목적 노드는 충분히 멀리 떨어져 있어 데이터를 전송하기 위해서는 릴레이 노드를 통해 소스 노드와 목적 노드가 연결되어야 한다. 모든 릴레이 노드의 송신 에너지는 같으며, 이에 따른 송신 범위 δ도 동일하다. 릴레이 노드는 이동 후 네트워크를 형성하며, 데이터를 동시에 전송한다. 릴레이 노드의 수는 N이며 소스 노드와 목적 노드를 포함한 전체 노드의 수는 M이다. i번째 릴레이 노드는 $n_i\,(0 \le i < N-1)$로 나타낸다.

2-2. Q-learning 기반 노드 이동성 제어

State space 상태 $s_{i,t} \in S$는 시간 t에서 릴레이 노드 n_i와 주변 노드 집합 $H_i = \{n_j \mid 0 \le j < M-1, i \ne j\}$와의 유클리디안 거리(Euclidean distance) $d_{i,t}(j)$와 각도 $a_{i,t}(j)$로 정의한다. 송신 범위 δ가 주변 노드와의 거리 $d_{i,t}(j)$보다 작은 경우 주변 노드 n_j에 대한 상태 정보는 존재하지 않는다.

Action space 행동 $a_{i,t} \in A$은 시간 t에서 릴레이 노드 n_i가 이동한 거리 $l_{i,t}(0 \le l_{i,t} \le l_{\max})$와 회전한 각도 $\theta_{i,t}(0 \le \theta_{i,t} < 2\pi)$의 조합이다. 시간 t에서의 n_i의 위치를 $(x_{i,t}, y_{i,t})$라고 할 때, 시간 $t+1$에서의 위치는 다음과 같이 나타낼 수 있다.

$$(x_{i,t+1}, y_{i,t+1}) = (l_{i,t} \times \cos\theta_{i,t} + x_{i,t}, l_{i,t} \times \sin\theta_{i,t} + y_{i,t})$$

Reward function 보상 $r_{i,t} \in R$은 행동에 따른 보상 항과 처벌 항으로 구성한다. 본 논문에서 에이전트의 학습 목표는 목적 노드와 같은 위치에서부터 이동하여 소스 노드-목적 노드 간 네트워크 연결을 유지하는 것이다. 따라서 보상 항은 네트워크 처리율(throughput) $\phi_t(s_{i,t})$과 단위 전송 범위 당 주변 노드와의 유클리디안 거리와 각도의 함수 $F(s_{i,t})$의 합으로 표현한다. 네트워크 처리율 $\phi_t(s_{i,t})$은 단위 시간당 전송되는 데이터의 양으로, $\phi_t(s_{i,t}) \ge 0$이면 소스 노드-목적 노드 간 연결이 이루어졌음을 의미한다. $F(s_{i,t})$는 (1)과 같이 정의하며

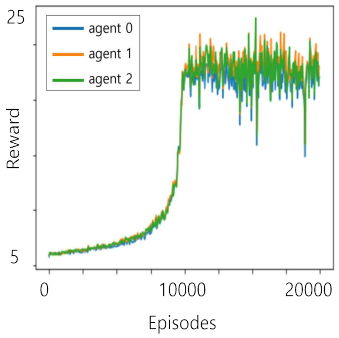


그림 1 에이전트의 학습 곡선

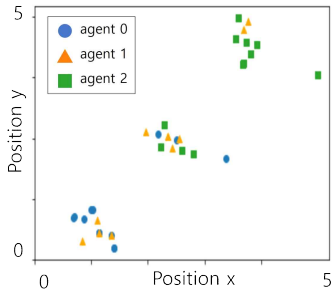


그림 2 에이전트의 탐색 초기 15% 기간 동안 위치

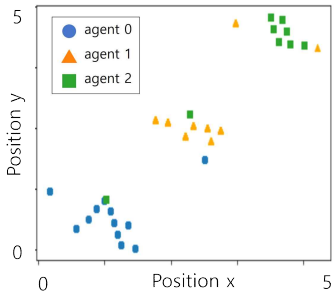


그림 3 에이전트의 최종 위치

$$F(s_{i,t}) = \frac{\sum_{j=0}^{|H_i|} d_{i,t}(j)}{|H_i| \times \delta} + \frac{\sum_{j=0}^{|H_i|} \alpha_{i,t}(j)}{|H_i| \times 180^o} \quad (1)$$

주변 노드의 거리 $d_{i,t}(j)$가 송신 범위 δ에 가까울수록, 주변 노드와의 각도 $\alpha_{i,t}(j)$가 180도에 가까울수록 1에 가까운 값을 갖는데 이는 노드 위치를 전송 범위에 맞추어 최적화했음을 의미한다. $F(s_{i,t})$는 릴레이 노드의 이동을 장려하는 역할을 한다. 이때 $|H_i|$는 주변 노드의 수를 나타낸다. 처벌 항은 행동에 따른 에너지 소모량 $G(a_{i,t}) = G(l_{i,t}, \theta_{i,t})$이다. 따라서 에이전트는 에너지 소모를 감소시키며 네트워크 처리율을 향상시키는 방향으로 학습된다.

본 논문의 보상 함수는 식 (2)와 같이 보상 항과 처벌 항의 합으로 표현한다. 계수 ω_t와 상수 k는 각각 처벌 항과 네트워크 처리율의 가중치를 나타낸다. ω_t는 학습이 진행됨에 따라 선형적으로 증가하며 릴레이 노드 위치의 수렴을 가속화 한다.

$$r_{i,t}(s_{i,t}, a_{i,t}) = k \times \phi_t(s_{i,t}) + (1-\omega_t) \times F(s_{i,t}) - \omega_t \times G(a_{i,t}) \quad (2)$$

2-3. 이동성 갖는 무선 노드의 Q-learning 기반 애드혹 네트워크 구축

네트워크 구축 시뮬레이션은 5×5 영역 내에 5개의 노드로 구성한다. 소스 노드와 목적 노드는 각각 1개이며 중간 노드는 3개이다. 소스 노드는 목적 노드에 직접 데이터를 전송할 수 없으며, 모든 릴레이 노드는 송신 범위 $\delta = 2$를 유지한다. 릴레이 노드 n_j는 시간 t에서 $\theta_{i,t} \in \left\{0, \frac{\pi}{2}, \pi, \frac{3}{2}\pi\right\}$만큼 회전한 뒤, $l_{i,t} \in \{0,1,2\}$만큼 이동한다. 릴레이 노드의 초기 위치는 목적 노드와 $(5,5)$로 같으며 소스 노드의 위치는 $(0,0)$이다.

그림 1은 에이전트의 학습 곡선(learning curve)를 나타낸다. 탐색(exploration)이 끝난 후 일정한 범위의 보상으로 수렴하는 학습 곡선을 통해 학습이 성공적으로 이루어졌음을 알 수 있다.

그림 2와 3은 에이전트인 릴레이 노드들의 위치를 나타낸다. 그림 2는 탐색 기간 초기 15% 기간 동안 1000개 에피소드마다 릴레이 노드 위치의 최빈값을 보여준다. 탐색 초기 15% 기간에서 노드 n_0의 평균 위치는 $(1.5, 0.9)$, n_1은 $(2.4, 2.0)$, n_2는 $(3.5, 3.0)$이다. 그림 3은 에이전트인 릴레이 노드의 최종 위치를 나타낸다. 그림 3을 통해 각 릴레이 노드는 학습 결과 특정 위치에 수렴함을 알 수 있는데, 평균적으로 노드 n_0은 $(1.1, 0.4)$, n_1은 $(2.3, 2.0)$, n_2는 $(3.8, 3.4)$에 위치한다. 이러한 최종적인 자리 선정은 노드 간 팀 내 역할 분배가 이루어졌다는 것을 의미한다. 즉, 애드혹 네트워크 구성에 있어 릴레이 노드는 소스 노드와의 연결, 릴레이 노드 간의 연결, 그리고 목적 노드와의 연결을 나누어 담당하게 된다. 이때, 그림 2와 3을 통해 네트워크 형성을 위한 역할분배는 탐색 기간 중 초기의 경험을 바탕으로 결정된다는 것을 알 수 있다.

Ⅲ. 결론

본 논문에서는 이동성을 갖는 무선 노드들이 Q-learning 기반으로 분산적으로 이동 거리와 회전 각도를 결정하여 애드혹 네트워크를 구축하는 방안을 제안하였다. 제안한 방안에 의해 릴레이 노드들은 이동성을 제어하며 연결 방식에 대한 역할을 스스로 학습해 소스 노드-목적 노드 간 네트워크를 형성한다. 시뮬레이션 결과를 통해 각 릴레이 노드는 사전에 역할을 정의하지 않아도 학습 과정에서 애드혹 네트워크 내에서 각자의 역할을 학습하여 특정 위치에 수렴함을 확인하였으며, 그 위치는 탐색 과정 초반에 결정됨을 확인하였다.

ACKNOWLEDGMENT

본 연구는 2020년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(NRF-2020R1A2B5B01002528, NRF-2020R1F1A1069182) 및 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(2021-0-00739)을 받아 수행되었습니다.

참 고 문 헌

[1] M. Deruyck, J. Wyckmans, L. Martens and W. Joseph, "Emergency ad-hoc networks by using drone mounted base stations for a disaster scenario," *IEEE WiMob*, pp. 1-7, 2016

[2] M. Y. Arafat and S. Moh, "Bio-inspired approaches for energy-efficient localization and clustering in UAV networks for monitoring wildfires in remote areas," *IEEE Access*, vol. 9, no. 1, pp. 18649-18669, 2021.

[3] 김나영, 권민혜, 박형곤, "랜덤 이동 모델을 따르는 무선 노드의 Q-learning 기반 애드혹 네트워크 구축 전략," 한국통신학회 논문지, Nov. 2021.

[4] M. Y. Arafat and S. Moh, "A Q-Learning-based topology-aware routing protocol for flying ad hoc networks," *IEEE IoT Journal*, pp.1-1, Jun. 2021.

[5] N. I. Mowla, N. H. Tran, I. Doh and K. Chae, "AFRL: Adaptive federated reinforcement learning for intelligent jamming defense in FANET," *Journal of Communications and Networks*, vol. 22, no. 3, pp. 244-258, June 2020.

레벨링과 분류기를 통한 자율주행에서의 차량인식

안재오, 이준연*
동명대학교, *동명대학교

dkswodh0518@naver.com, *jylee@tu.ac.kr

On-Road Vehicle Detection using Labeling and Classification in Autonomous Driving

Jaeoh An , Junyeon Lee*
Tongmyong Univ., *Tongmyong Univ.

요 약

자율주행차량의 상용화와 함께 차량 인식에 대한 기술이 많이 연구되어 왔다. 차량인식기술은 전통적인 이미지처리 기법을 사용하는 방법과 딥러닝 기법을 사용하는 방법이 있는데, 두 종류 모두 주간에는 훌륭한 인식성능을 보여주지만, 야간에는 후미등의 모양과 밝기, 위치 등 여러 요인으로 인하여 인식률이 현저히 떨어진다. 따라서 본 논문에서는 라벨링을 거쳐 선행차량을 검출하는 알고리즘을 제안한다.

Ⅰ. 서 론

자동차의 향상된 운전자 보조 시스템 (ADAS : Advanced Driver Assistance System)뿐만 아니라 자율 자동차 주행을 위한 강화학습과 같이 자율 주행 자동차에 필요한 기술에 대한 연구가 활발히 이루어지고 있다. 그 중에서도 차량 인식에 대한 개발은 자율 주행 자동차의 상용화에 있어서 중요한 연구요소중의 하나이다[1, 2]. 자율 주행에 있어서 신호등이나 교통표지판의 인식, 차선 인식과 같이 다른 교통 요소들의 인식 역시 중요하다. 그러나 차량 인식의 경우 사람의 생명과 직결되는 문제이기 때문에, 자율주행에 있어서 가장 주요한 과제 중 하나이다[3].

이러한 자율 주행 차량에서 물체를 인식 방법에는 카메라, 라이다(LiDAR), 레이더(Radar) 센서 등을 사용하는 기법들이 존재한다. 이 중 카메라를 통한 차량 인식을 위한 알고리즘은 크게 이미지 프로세싱을 이용한 방법 [4]과 딥러닝을 이용한 방법 [5]으로 나뉜다. 두 방법을 이용한 이전 연구들의 경우 주간 차량인식에 있어서 훌륭한 결과를 보여주지만, 야간에서의 인식률은 좋지 못하다. 마치 사람이 밤에 물체를 인식하는데 있어서 어려움을 겪는 것처럼 자율주행 시스템의 인식도 마찬가지이다.

따라서 본 논문에서는 레벨링과 분류기를 거쳐 선행차량을 인식할 수 있는 알고리즘을 제안한다.

Ⅱ. 관련연구

2.1 전통적인 이미지 처리를 이용한 차량인식

이미지 프로세싱을 이용한 차량인식 방법 중 가장 기본적인 방법은 바로 캐니를 이용한 에지 검출기이다 [6]. 캐니 에지 처리 이전에, 알고리즘은 가우시안 필더링 기법을 이용해 영상을 소프트닝시킨 후, gradient vector의 크기를 계산한다. 마지막으로, hysteresis thresholding을 사용하여 연결된 엣지들을 얻을 수 있다.

2.2 딥러닝을 이용한 차량인식

최근 들어 계산량을 빠르게 처리할 수 있는 하드웨어

들의 보급과 함께, 차량 인식에 대한 연구들은 기존의 전통적인 이미지 처리 기법을 이용한 방법에서 딥러닝을 이용한 기법들로 연구의 추이가 변화하고 있다[7].

2.3 영상분할(vehicle segmentation)

영상분할 기법은 이미지 세그멘테이션 기법에 대해서 다룬다[8]. CNN을 이용한 연구가 활발해지면서, 이미지 세그멘테이션 기법 역시 CNN을 이용하여 해결할 수 있게 되었다.

Ⅲ. 선행차량 검출구조

본 논문에서 제안하는 야간 차량 검출 시스템은 각 화소의 밝기 값을 기준으로 분류기를 구축하고, 학습을 통해 구축된 분류기로 영상에서 차량 후보를 탐색한다. 분류기 구축은 Haar 유사특징을 이용하기 때문에 차량 샘플의 경우 일관성 유지가 요구된다.

레벨링 (Lebelling) 과정을 거쳐 차량의 전조등 및 후미등 후보 영역을 추출하고, 추출된 후보 영역을 합성곱 신경망기반 분류기 (classification)에 입력하여 해당 후보영역 이 전조등 (또는 후미등)인지 비차량 인지를 판별하는 과정으로 구성된다. 학습을 통해 구축된 분류기로 영상에서 차량 후보를 탐색한다. 탐색에 사용하는 템플릿의 크기는 고정되어 있기 때문에 다양한 크기로 보이는 차량을 모두 검출할 수는 없다. 이를 해결하기 위해서는 입력 양상의 크기를 몇 단계로 줄여 피라미드 영상을 구축한다.

제안 방법에서는 빠른 후보영역 추출을 위해 이진화와 라벨링을 사용하는데, 이는 야간 환경에서 빠르고 효과적으로 차량 후보 영역의 추출이 가능하기 때문이다. 전조등이나 후미등이 아닌 후보 영 역이 생성되지만 합성곱 신경망의 성능에 따라 효과적으로 분류 가능하다

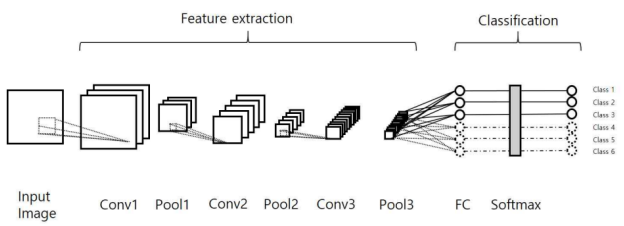


그림 1 검출 분류기 구조

Ⅳ. 결론

본 논문에서는 성능의 신뢰성이 필수적이며, 빠른 속도로 검출되어야 하는 주행가능 영역의 차량을 검색하여 검출 및 분할할 수 있는 알고리즘을 제시하였다. 야간 도로 주행 영상에 적용한 실험을 통해 제안된 신경망 기법이 기존의 다중 클래스 기반 SVM 방법과 유사한 검출성능을 보임으로써 합성곱 신경망 학습방법이 야간 차량 검출에 적용이 가능함을 확인하였다.

.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학지원사업의 연구결과로 수행되었음 (1711102971).

참 고 문 헌

[1] T. Hwu, J. Isbell, N. Oros, and J. Krichmar, “A self-driving robot using deep convolutional neural networks on neuromorphic hardware,” in Proc. IEEE International Joint Conference on Neural Networks (IJCNN), 2017.

[2] A. Forechi, T. Oliveira-Santos, C. Badue, and A. F. D. Souza, “Visual global localization with a hybrid WNN-CNN approach,” in Proc. IEEE International Joint Conference on Neural Networks (IJCNN), 2018.

[3] Z. Sun, G. Bebis, and R. Miller, “On-road vehicle detection: A review,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 5, pp. 694- 711, May 2006

[4] A. Jazayeri, H. Cai, J. Y. Zheng, and M. Tuceryan, “Vehicle detection and tracking in car video based on motion model,” IEEE Transactions on Intelligent Transportation Systems, vol. 12, no. 2, pp. 583- 595, June 2011.

[5] B. Huval, T. Wang, S. Tandon, and et. al., “An empirical evaluation of deep learning on highway driving,” arXiv:1504.01716, 2015.

[6] J. Canny, “A computational approach to edge detection,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 8, no. 6, pp. 679- 698, November 1986.

[7] S. Sivaraman and M. M. Trivedi, “Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis,” IEEE Transactions on Intelligent Transportation Systems, vol. 14, no. 4, pp. 1773- 1795, December 2013.

[8] F. Yu and V. Koltun. “Multi-scale context aggregation by dilated convolutions,” in Proc. International Conference on Learning Representations (ICLR), 2017.

사용자의 선호도를 고려한 콘텐츠 캐싱을 위해 추천시스템을 활용하여 연속적인 공간에서 콘텐츠를 캐싱하는 심층강화학습 모델

이우빈, 임민중

동국대학교 컴퓨터정보통신공학과

woobin0357@dongguk.edu

A Deep Reinforcement Learning Model that Caches Contents in a Continuous Space using Recommendation System for Contents Caching considering User Preferences

Woobin Lee, Minjoong Rim

Department of Information and Communication Engineering

Dongguk University

요 약

본 논문에서는 네트워크 엣지에서 사전 콘텐츠 캐싱 문제를 연구한다. 연구된 모델에서 네트워크 엣지의 캐시는 추천시스템을 활용하여 사용자 개개인의 선호도를 예측하고 사용자에게 추천이 될 만한 적절한 콘텐츠를 캐싱하는 결정을 할 수 있다. 이 문제를 해결하기 위해 강화학습 기반의 프레임워크를 제안하며, 무한히 증가하는 콘텐츠 수와 한정된 캐시 저장 공간에 따른 캐싱 결정의 복잡도 문제에 대처하기 위해 콘텐츠를 연속적인 공간에 배치시키는 것을 제안한다. 실험에서는 제안된 접근 방식이 사용자 개개인의 특성을 고려하지 않는 보편적인 선호도에 따른 캐싱 알고리즘과 비교하여 사용자들의 캐싱된 콘텐츠 선호도 측면에서 더 많은 이득을 산출한다는 것을 보여준다.

Ⅰ. 서론

오늘날 유튜브, 넷플릭스 등의 비디오 콘텐츠를 스트리밍하는 서비스는 심화를 넘어 포화상태로 보인다. 영상의 품질 자체가 올라감에 따라 콘텐츠 자체의 용량 또한 증가하고 있다. 다양한 기기에서 오는 사용자의 요청과 콘텐츠 용량의 증가로 인해 트래픽이 급증하는 피크 시간에는 네트워크의 백홀과 프론트홀에서 병목현상이 발생할 수 있다. 따라서 이를 해결하기 위한 방법 중 하나로 인공지능을 활용하여 네트워크 엣지에서 콘텐츠를 캐싱하는 연구가 활발히 진행되고 있다. 이러한 콘텐츠 캐싱에서는 추천시스템을 같이 고려하기도 한다. 추천시스템은 사용자에게 추천될 만한 콘텐츠를 추천하고, 이러한 추천될 만한 콘텐츠를 캐시에 저장한다면 캐시의 효율이 훨씬 높아질 것이다. [1]에서는 추천시스템을 활용하여 각 사용자의 기기에서 최적의 콘텐츠를 캐싱하고 기기간 통신을 통해 캐싱된 콘텐츠를 공유하여 기지국의 네트워크 부담을 줄인다. [1]과는 달리 본논문에서는 캐싱을 할 콘텐츠를 스스로 학습하고 결정하기 위해 강화학습 기반의 모델을 제안한다. 이전 연구에서는 사용자 개개인의 특성을 고려하지 않고 보편적인 선호도에 따라 콘텐츠 캐싱을 하지만 본논문에서는 추천시스템을 활용하여 사용자 개개인의 선호도를 예측함으로써 사용자에 따라 적절한 콘텐츠를 사전 캐싱한다. 추가적으로 무한히 증가하는 콘텐츠의 수와 한정된 캐시 저장 공간 문제를 해결하기 위해 콘텐츠를 연속적인 공간에 배치시키는 것을 제안한다. [2]에서는 강화학습으로 사용자의 실시간 피드백을 학습하여 사용자가 선호하는 상품을 추천하는 웹 페이지를 업데이트 한다. 상품들은 특성에 따라 연속적인 공간에 배치가 되어 있기 때문에 강화학습 모델은 연속적인 공간에서 비슷한 상품을 찾는다.

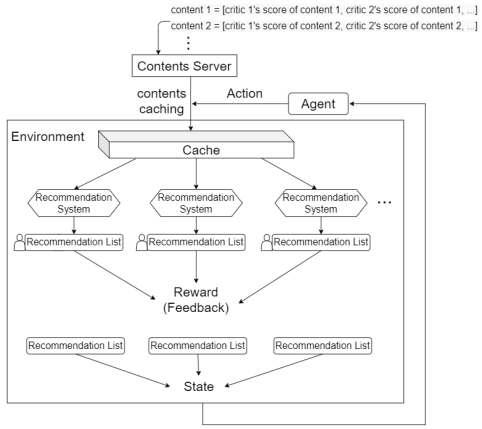
본논문에서는 콘텐츠에 점수를 매기는 비평가 요소를 도입하여 [2]에서처럼 콘텐츠들을 연속적인 공간에 배치한다. 제안한 방법은 콘텐츠의 수가 무한히 늘어나는 실제 환경에서 캐싱 결정의 복잡도가 높아지는 것과는 무관하게 적절한 캐싱 결정을 내릴 수 있도록 한다.

Ⅱ. 본론

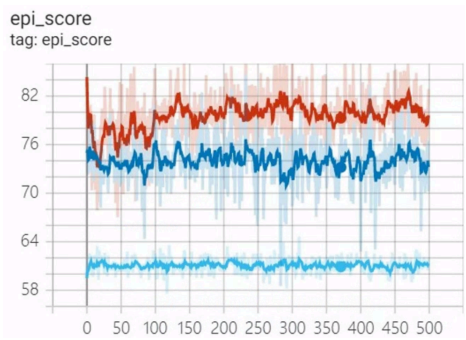
본논문에서 두 가지 시스템을 가정한다. 첫 번째로, 잠재 요인 협업 필터링(Latent Factor Collaborative Filtering) 기법처럼 사용자들의 콘텐츠에 대한 잠재 요인이 고려되는 추천시스템이 존재한다고 가정한다. 이러한 추천시스템에서는 사용자들은 각 잠재 요인들에 따라(e.g. 영화 장르라는 잠재 요인) 잠재 요인 선호도를 가질 수 있고 콘텐츠들을 잠재 요인들로 표현할 수 있게 된다. 본논문에서는 이러한 잠재 요인을 나타내는 비평가라는 개념을 도입하여 각각의 특성을 갖는(특정한 잠재 요인을 갖는) 비평가들이 콘텐츠에 대해 평점을 매긴다. 두 번째로, 각 사용자들의 선호도가 어떤 비평가들의 어떤 조합과 비슷한지 찾아내도록 학습된 시스템이 있다고 가정한다. 이를 통해 사용자가 아직 평점을 매기지 못한 새로운 콘텐츠에 대해서도 잠재 요인을 나타내는 비평가들의 점수를 통해 콘텐츠 예상 선호도 점수를 매길 수 있다. 본논문에서는 가정한 시스템들과 결합하여 강화학습을 통해 사용자 개개인의 선호도가 어떤 비평가와 비슷한지를 예측하여 추천될 만한 콘텐츠를 사전 캐싱하는 프레임워크를 제안한다. [그림 1]은 제안하는 시스템 모델의 강화학습 알고리즘으로 에이전트와 상호작용을 하는 환경을 보면, 캐시가 지원하는 사용자는 여러 명이 있고 사용자마다 추천시스템이 존재하며 잠재 요인을 나타내는 비평가가 여려 명 있고 추천시스템은 이 비평가의 콘텐츠 평점을 통해 구한 사용자의 예상 콘텐츠 평점을 이용하여 콘텐츠를 추천한다. 에이전트는 캐시 내에서 각 사용자들에게 만들어지는 추천리스트들을 상태로 받아 어떤 콘텐츠로 캐시를 구성할지 사전 결정하는 행동을 한다. 에이전트의 행동으로 캐시가 구성되면 캐시 내 콘텐츠들이 바뀌면서 추천리스트가 바뀌게 되고, 캐시에서 사용자들에게 추천된 콘텐츠의 예상 평점을 통해 보상을 받게 된다. 추천된 콘텐츠의 사용자 예상 평점이 높다면 높은 보상을 얻게 된다. 에이전트는 이 과정을 반복하면서 현재 캐시를 사용하는 사용자의 특성에 따라 어떤 콘텐츠가 추천이 될 만한 콘텐츠인지를 학습하게 된다. 하지만 실제 환경에서는 새로 만들어지는 콘텐츠의 개수가 무한히 증가할 것이고 이를 한정된 저장 공간을 가진 캐시에 저장해야 한다. 콘텐츠 개수가 증가할수록 콘텐츠를 캐싱하는 결정의 복잡도가 크게 증가할 것이다. 따라서 이러한 문제를 해결하기 위해 콘텐츠를 연속적인 공간에 배치시키는 것을 제안한다. 비평가들이 콘텐츠에 매긴 점수에 따라 콘텐츠 특성을 구성하게 되고 콘텐츠는 비평가 수만큼의 차원의 크기를 갖는 연속적인 공간에 배치된다. 아무리 콘텐츠 개수가 늘어나더라도 최적의 콘텐츠는 연속적인 공간에 배치된 콘텐츠들 중에서 가장 가까운 콘텐츠를 선택하면 된다. 이를 통해 콘텐츠 개수가 늘어나면서 캐싱 결정의 복잡도가 높아지는 것과는 무관하게 적절한 캐싱 결정을 내리는 것이 가능해진다. 실험은 500 개의 콘텐츠와 5 명의 비평가가 존재하고 4 명의 사용자가 에피소드마다 선호하는 비평가가 계속 바뀌는(사용자가 계속 바뀌는) 환경에서 진행된다. [그림 2]는 500 번의 각 에피소드에서 사용자들의 캐시 내 콘텐츠 예상 평점의 합으로 성능을 나타낸 그래프이다. 사용자 개개인의 선호도를 고려하지 않은 보편적인 선호도에 따른 콘텐츠 캐싱(파란색)과 랜덤 콘텐츠 캐싱(하늘색)과 비교하여 제안하는 모델(빨간색)이 사용자의 선호도를 가장 만족시키는 것을 볼 수 있다.

Ⅲ. 결론

본논문에서는 추천시스템을 활용하여 사용자 개개인의 선호도를 고려하여 콘텐츠를 사전 캐싱하는 모델을 제안한다. 이전의 많은 연구가 콘텐츠 각각을 고려함으로써 고려하는 콘텐츠의 수가 제한되는데 반해서 본논문은 콘텐츠를 연속적인 공간에서 고려함으로써 무한히 많은 콘텐츠를 고려할 수 있다.



[그림 1] 제안하는 강화학습 알고리즘



[그림 2] 전반적인 성능 비교

ACKNOWLEDGMENT

본 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No2019R1F1A1056963). 본 논문은 삼성전자 미래기술육성센터의 지원으로 이루어진 결과임 (SRFC-IT1702-13).

참 고 문 헌

[1] Y. Wang, M. Ding, Z. Chen and L. Luo, "Caching Placement with Recommendation Systems for Cache-Enabled Mobile Social Networks", IEEE Communications Letters, vol. 21, pp. 2266-2269, 2017

[2] Xiangyu Zhao, Long Xia, Liang Zhang, Zhuoye Ding, Dawei Yin, Jiliang Tang, "Deep Reinforcement Learning for Page-wise Recommendations", RecSys '18: Proceedings of the 12th ACM Conference on Recommender Systems, pp. 95-103, 2018

궤적에서 객체의 중심점과 변위 추정을 통한 영상 객체 탐지 연구

손하영, 이유진, 최계원*

성균관대학교

shyds628@g.skku.edu, leeyoujin225@g.skku.edu,*kaewonchoi@skku.edu

A Study on Video Obejct Detection by Estimation of Center and Movement of the object on the trajectory

Ha Young Son, You Jin Lee, Kae Won Choi*

Sungkyunkwan Univ.

요 약

본 논문에서는 영상 속에서 원거리에 위치하거나 해상도에 비해 매우 작아 형상이 명확하지 않아 탐지가 잘 되지 않는 객체를 탐지하기 위해, Anchor-Free 방식으로 Object Detection을 수행하는 CenterNet을 적용하고, 동시에 객체의 움직임을 추정하는 네트워크인 STTNet(Short Term Tracklet Network)을 만들었다. 이 네트워크에 VisDrone-MOT 데이터셋을 훈련시켜 영상의 일정 시간 간격의 프레임 사이에서 객체의 위치와 변위를 추정하여 영상 객체 탐지를 수행하였다.

Ⅰ. 서 론

최근 다양한 교통 상황과 물류 환경에서, Object Detection을 통한 Visual Surveillance의 적용이 연구되고 있다. 이 중 항만 환경은 안전 사고 방지와 보안을 위해 객체마다 탐지가 필요하지만, 크기가 제각각인 다양한 장애물이 존재하거나 움직이며 사람의 눈으로 파악하기 어려운 넓은 구역을 한꺼번에 감시해야 하는 환경이라는 문제점이 있다.

특히 이러한 환경의 경우 한 번에 넓은 환경을 감시하기 위해 낮은 지점이 아니라 높은 지점에 폐쇄회로 카메라가 설치된다는 특수성이 있다. 따라서 영상 속의 객체는 카메라로부터 상당히 원거리에 위치하며, 이로 인해 객체의 형상이 명확하지 않고 해상도에 비해 매우 작다는 문제점이 발생한다. 이러한 객체를 잘 식별하기 위해서는 전후 프레임을 참고하여 추가적인 시간적 정보를 학습에 활용할 필요가 있다.

이렇게 추가적인 시간적 정보를 학습에 활용하는 방법에는 정지된 단일 이미지 한 장이 아니라 영상을 입력으로 Object Detection을 수행하는 방법이 있다. Video Object Detection과 Object Tracking이 그 예이다.

영상을 입력으로 사용할 경우 단일 정지 이미지에서는 발생하지 않고 영상에서만 발생하는 문제가 있는데, 객체의 움직임에 따라 모션 블러가 발생하거나 포커스를 놓치기도 한다는 점이다. 이를 해결하기 위해 CNN을 거친 Feature 외에도 다양한 Feature을 사용하기도 한다. Optical Flow와 유사하게, 객체의 움직임을 나타내는 Vector를 Flow Network를 통해 Feature로 만든 Flow Field[3]를 인접한 프레임에서 모아 Video Object Detection에 사용한 FGFA[4]가 그 예이다.

영상의 각 프레임마다 Object Detection을 수행할 때 대부분 Faster-RCNN[5], YOLO-v3[6]와 같은 Anchor-Based 방식을 사용하지만, 특수한 상황에 맞추어 Anchor-Free 방식을 사용하기도 한다. 대표적인 Anchor-Free 방식인 CenterNet[1]은 Anchor 없이 각 객체들의 중심점을 추정하여 객체들을 하나의 점으로 표현하는 네트워크이다. 이 CenterNet을 Object Tracking에 적용한 사례는 FairMOT[2]가 있다.

본 논문에서는 영상의 작은 크기의 원거리 객체를 탐지하는 것이 목표로, 시간적 정보를 얻기 위해 여러 장의 이미지를 쌓아 훈련에 사용했다. 또한, 바운딩 박스를 찾는 것이 중점이 아니며 주로 작은 객체를 탐지해야 하기 때문에 높은 해상도를 가진 단일 레이어를 사용하는 것이 유리하다는 점에서 Anchor-Free 방식인 CenterNet을 기반으로 객체를 탐지했다. 또한, 갑자기 방향을 바꾸기도 하는 객체의 움직임을 좀 더 잘 파악하기 위해 객체의 변위 정보도 함께 예측했다.

Ⅱ. 본론



그림 1. VisDrone 데이터셋의 이미지와 그 라벨링

훈련 데이터셋은 VisDrone-MOT2019[7]을 사용했으며, 이 데이터셋은 Multi Object Tracking을 위해 만들어진 데이터셋으로, 드론이 고점에서 천천히 비행하면서 지상을 촬영한 영상에 객체의 클래스, 위치, ID를 프레임마다 라벨링하였다. 그림 1처럼 차량이 많은 교차로, 사람이 많이 지나다니는 공원과 학교 등 여러 환경을 높은 지점의 원거리에서 다양한 시간대에 촬영한 영상들이 우리의 목표와 잘 부합하여 훈련에 사용하였다. 훈련에는 56개의 시퀀스(총 24,201 프레임)를 사용했고, 테스트에는 17개의 시퀀스(총 6,618 프레임)을 사용했다.

입력으로는 그림 2와 같이 시퀀스 상의 연속된 이미지를 8장 단위로 쌓아 하나의 샘플을 만든 뒤, 이 샘플들을 모아 배치를 만들어 훈련에 사용했다. 샘플을 쌓는 단위가 작아질수록 샘플 내부의 프레임끼리의 시간적 인접성이 커지고, 커질수록 그 반대가 된다. 각 샘플의 처음 프레임과 마지막 프레임에서 객체의 중심점과 변위를 목표로 훈련했다.

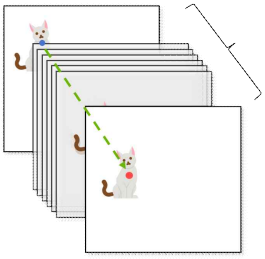


그림 2. 입력 샘플의 구조. 파란색 점은 시작 이미지 물체의 Center, 빨간색 점은 마지막 이미지 물체의 Center이다.



그림 3. 출력 이미지 샘플. 빨간색은 시작 이미지의 확률 분포 히트맵, 파란색은 마지막 이미지의 확률 분포, 초록색은 움직임을 나타내는 화살표.

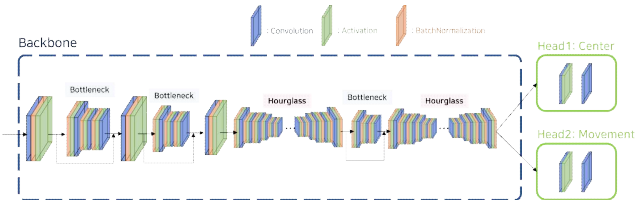


그림 4. STTNet의 훈련 네트워크 구조. 여러 개의 Bottleneck 레이어와 Hourglass 구조 레이어로 이루어진 백본 네트워크와 두 개의 탐지 Head로 이루어져 있다.

훈련한 백본 네트워크는 그림 4처럼 고해상도의 정보를 유지하기 위해 Skip Connection이 존재하는 Bottleneck 구조 레이어와 Hourglass 구조 레이어 여러 층으로 이루어진 깊은 네트워크로, 이를 거친 후 두 개의 Detection Head를 통해 출력을 만들었다. Hourglass 레이어는 Bottleneck 레이어를 재귀적으로 쌓아 만들었다. Head 중 하나는 각 클래스마다 객체의 중심점이 있을 확률의, 즉 Class score의 분포도를 찾는 Head이며, 다른 하나는 각 클래스마다 객체의 변위를 찾는 Head이다.

각 Head마다 훈련에 사용하는 Loss는 다음과 같다. 중심점을 찾는 Loss인 L_c는 CenterNet에서 사용된 Loss와 같이 Focal Loss[8]를 사용하였다.

$$L_c = -\frac{1}{N}\sum \begin{cases} (1-\hat{Y})^{\alpha}\log(\hat{Y}) & \text{if } Y=1 \\ (1-Y)^{\beta}(\hat{Y})^{\alpha}\log(1-\hat{Y}) & otherwise \end{cases} \quad (1)$$

$\hat{Y}$는 Prediction, Y는 Target, N은 이미지의 중심점 개수, α와 β는 Focal Loss의 하이퍼 파라미터로 각각 2와 4를 사용했다. 객체의 변위를 찾는 Loss인 L_m은 smooth L1 loss를 사용하여 훈련했다.

그림 3은 이 훈련을 통해 얻을 수 있는 출력을 보기 쉽도록 샘플의 첫 번째 시작 이미지와 마지막 이미지를 겹쳐 만든 이미지 샘플이다. 출력은 클래스마다 입력 샘플 첫 번째 시작 이미지의 객체 중심점의 확률 분포, 마지막 이미지의 객체 중심점의 확률 분포, 객체의 X축 변위, 객체의 Y축 변위로 이루어져 있다.

Ⅲ. 결론

클래스는 크게 사람, 이륜차(자전거, 오토바이), 사륜차(일반 차량, 트럭, 버스 등)의 총 세 가지로 나누어 학습하였다.

학습 결과, 8장의 이미지를 쌓아 만든 샘플 내에서 객체의 움직임을 따라 변위의 방향이 달라짐을 알 수 있었고, 객체가 위치한 곳에 객체가 위치할 확률이 문턱값보다 높음을 알려주는 키포인트를 볼 수 있었다. 이를 통해 영상 내의 일정 기간 프레임 내에서 객체의 움직임과 위치를 추정할 수 있음을 알 수 있었다.

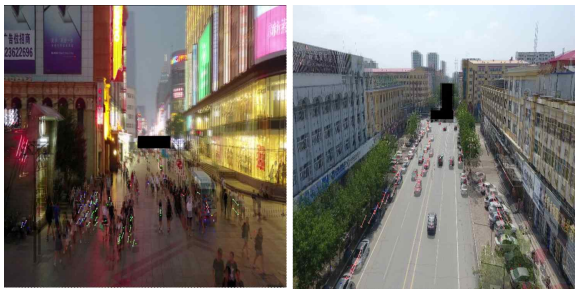


그림 5. 각 클래스 별로 출력을 보기 쉽도록 만든 이미지. 빨간색은 사륜차를, 파란색은 이륜차를, 초록색은 사람의 위치와 변위를 나타낸다.

그러나 샘플을 쌓는 프레임 수가 늘어나 샘플의 시간적 거리가 증가할수록, 카메라가 급격히 흔들리거나 크기가 작은 물체를 탐지하지 못하는 모습을 보였으며 다른 클래스로 추정하는 문제가 발생했다. 따라서 다른 프레임의 시간적 정보를 융합하고 보정하는 Feature Aggregation 방법을 통해 영상 객체 탐지를 개선할 필요가 있다.

ACKNOWLEDGMENT

이 논문은 2021년도 정부(해양수산부)의 재원으로 해양수산과학기술진흥원의 지원("IoT기반 지능형 항만 물류 기술 개발"사업의 "스마트 항만 IoT 융합•운영 기술 개발")을 받아 수행된 연구임.

참 고 문 헌

[1] Zhou, Xingyi, Dequan Wang, and Philipp Krähenbühl. "Objects as points." arXiv preprint arXiv:1904.07850 (2019).

[2] Zhang, Yifu, et al. "Fairmot: On the fairness of detection and re-identification in multiple object tracking." arXiv preprint arXiv:2004.01888 (2020).

[3] Zhu, Xizhou, et al. "Deep feature flow for video recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. (2017).

[4] Zhu, Xizhou, et al. "Flow-guided feature aggregation for video object detection." Proceedings of the IEEE International Conference on Computer Vision. (2017).

[5] Ren, Shaoqing, et al. "Faster R-CNN: towards real-time object detection with region proposal networks." IEEE transactions on pattern analysis and machine intelligence 39.6 (2016): 1137-1149.

[6] Redmon, Joseph, and Ali Farhadi. "Yolov3: An incremental improvement." arXiv preprint arXiv:1804.02767 (2018).

[7] Wen, Longyin, et al. "Visdrone-mot2019: The vision meets drone multiple object tracking challenge results." Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. (2019).

[8] Lin, Tsung-Yi, et al. "Focal loss for dense object detection." Proceedings of the IEEE international conference on computer vision. (2017).