www.koreaai.org

제1회 한국 인공지능 학술대회

1st Korea Artificial Intelligence Conference

Proceeding

일자 2020년 12월 16일(수) ~ 18일(금)

주최 한국통신학회

주관 한국통신학회 인공지능소사이어티

후원 국민대학교 에너지인터넷연구센터

경북대학교 ICT·자동차융합연구센터

준비위원

▣ 자문위원

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

▣ 조정위원장

장영민(국민대)

▣ 조정위원

강신각(한국전자통신연구원) 김대진(전남대) 김일규(한국전자통신연구원)
문영준(한국교통연구원) 박현철(한국과학기술원) 여 현(순천대)
정성호(한국외국어대) 정 송(한국과학기술원) 허 준(고려대)
홍인기(경희대)

▣ 운영위원장

한동석 (한국통신학회 인공지능소사이어티 회장)

▣ 운영위원

• 총무

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

• EDAS

김상철(국민대) 박경준(대구경북과학기술원) 최계원(성균관대)

• 등록

조성래(중앙대) 최준원(한양대)

• 섭외

김중헌(고려대) 이현우(한국전자통신연구원)

• 재무

김재일(경북대) 최윤호(부산대)

• 출판

김덕경(인하대) 김정곤(한국산업기술대) 석준희(고려대)

• 지역협조

김도현(제주대) 김선옥(한라대) 송왕철(제주대)
좌정우(제주대)

준비위원

• 홍보

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

• 특별세션

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

▣ 프로그램위원장

손인수(동국대) 한연희(한국기술교육대)

▣ 프로그램위원

권태수(서울과학기술대) 김광순(연세대) 김수민(한국산업기술대)
김원태(한국기술교육대) 김준수(한국산업기술대) 김평수(한국산업기술대)
박준구(경북대) 박현희(명지대) 박형곤(이화여대)
박혜영(경북대) 소재우(서강대) 신석주(조선대)
신수용(금오공과대) 신오순(숭실대) 윤주상(동의대)
이재호(덕성여자대) 이종혁(세종대) 정방철(충남대)

진행시간표

2020년 12월 16일(수요일)

시 간	주제 / 강사 / 내용	
13:00~13:30	행 사 준 비	
13:30~14:30 (60“)	**논문 발표 A-1** 좌장: 이동명 교수 (동명대) 인공지능 응용I	**논문 발표 A-2** 좌장: 박경준 교수 (DGIST) 딥러닝 & 최적화
14:30~14:40 (10“)	Break	
14:40~15:40 (60“)	**논문 발표 B-1** 좌장: 최계원 교수 (성균관대) 하드웨어 & 스마트팩토리	**논문 발표 B-2** 좌장: 김정곤 교수 (한국산업기술대) 지능형 통신 I
15:40~15:50 (10“)	Break	
15:50~16:20 (30“)	**Top Conference 초청 섹션 I** 좌장 : 한동석 교수 (경북대) (KAIST 이주호 교수)	
16:20~16:50 (30“)	**Top Conference 초청 섹션 II** 좌장 : 한동석 교수 (경북대) (KAIST 김범준 교수)	
16:50~17:00 (10“)	Break	
17:00~18:00 (60“)	**논문 발표 C-1** 좌장: 김정구 교수 (부산대) 영상처리	**논문 발표 C-2** 좌장: 최지웅 교수 (DGIST) 지능형 차량

진행시간표

2020년 12월 17일(목요일)

시 간	주제 / 강사 / 내용	
08:30~09:00	행 사 준 비	
09:00~10:00 (60“)	**논문 발표 D-1** 좌장: 정방철 교수 (충남대) 자연어 처리	**논문 발표 D-2** 좌장: 황인태 교수 (전남대) 강화학습
10:00~10:10 (10“)	Break	
10:10~11:10 (60“)	**튜토리얼 I** 좌장 : 조성래 교수 (중앙대) (국민대 김상철 교수)	
11:10~11:20 (10“)	개회식	
11:20 ~ 12:00 (40’’)	**초청강연** 좌장 : 손인수 교수 (동국대) (ETRI 지능화융합연구소 박종현 소장)	
12:00~13:30 (90“)	Lunch	
13:30~14:30 (60“)	**튜토리얼 II** 좌장 : 백명선 박사 (ETRI) (경북대 김재일 교수)	
14:30~15:00 (30“)	**Top Conference 초청 섹션 III** 좌장 : 한연희 교수 (한국기술교육대) (고려대 김중헌 교수)	
15:00~15:10 (10“)	Break	
15:10~16:10 (60“)	**논문 발표 E-1** 좌장: 강승택 교수 (인천대) 로봇 제어	**논문 발표 E-2** 좌장: 김동성 교수 (금오공대) 영어 논문 세션 I
16:10~16:20 (10“)	Break	
16:20~17:20 (60“)	**논문 발표 F-1** 좌장: 신수용 교수 (금오공대) 지능형 환경	**논문 발표 F-2** 좌장: 최윤호 교수 (부산대) 의료인공지능

진행시간표

2020년 12월 18일(금요일)

시 간	주제 / 강사 / 내용	
08:30~09:00	행 사 준 비	
09:00~10:00 (60“)	**논문 발표 G-1** 좌장: 강승택 교수 (인천대) 영어 논문 세션 II	**논문 발표 G-2** 좌장: 박혜영 교수 (경북대) 인공지능 응용 II
10:00~10:10 (10“)	Break	
10:10~11:10 (60“)	**논문 발표 H-1** 좌장: 최준원 교수 (한양대) 지능형 통신 II	**논문 발표 H-2** 좌장: 이규만 교수 (건양대) 지능형 보안 및 응용

목차

세션 A-1. 인공지능 응용 I

12월 16일(수요일) 13:30 ~ 14:30 (60분) | 좌장: 이동명 교수 (동명대)

A-1-1 개인 맞춤형 광고 제공을 위한 기계 학습 기반 한국인 성별 및 나이 인식 모델 생성 001
홍승준(경희대), 김명섭(경희대), 허의남(경희대)

A-1-2 인공지능 기반 지능형 범죄 위험 예측 및 대응 기술을 활용한 스마트 치안 기술 연구 003
백명선(한국전자통신연구원), 이용태(한국전자통신연구원), 박원주(한국전자통신연구원)

A-1-3 AWS 딥렌즈와 엣지 클라우드를 이용한 성별 및 나이 맞춤형 실시간 광고 스트리밍 서비스 005
김서현(경희대), 김명섭(경희대), 홍승준(경희대), 김명현(경희대), 허의남(경희대)

A-1-4 사이니지 사용자 맞춤형 광고를 제공하기 위한 엣지 클라우드 기반 얼굴 인식 모듈과 맞춤형 광고 서비스 개발 007
김명섭(경희대), 홍승준(경희대), 김서현(경희대), 김명현(경희대), 허의남(경희대)

A-1-5 국방 인공지능 동향과 국내 적용에 관한 연구 009
김상민(한화시스템)

A-1-6 K-스마트드라마 구현을 위한 인공지능 활용 방법 연구 011
최지애(칼빈대), 이일호(칼빈대), 권오병(경희대)

세션 A-2. 딥러닝 & 최적화

12월 16일(수요일) 13:30 ~ 14:30 (60분) | 좌장: 박경준 교수 (DGIST)

A-2-1 유전 알고리즘을 이용한 경량 인공지능 시스템에서의 하이퍼 파라미터 최적화 013
김병수(한국전자기술연구원), 전석훈(한국전자기술연구원), 황태호(한국전자기술연구원)

A-2-2 Single Memory를 활용한 뉴럴 네트워크 프로세서용 효율적 Processing Element Array 구조 제안 015
이재학(한국전자기술연구원), 김병수(한국전자기술연구원), 송보배(한국전자기술연구원),
황태호(한국전자기술연구원)

A-2-3 Meijer G-함수를 활용한 딥러닝 기반 설명 가능한 패스로스 모델 017
이현석(세종대)

A-2-4 자연계 최적법칙에 기반한 마이크로파 필터 설계법 019
이창형(인천대), 조정현(인천대), 서예준(인천대), 전문수(인천대), 이경민(인천대), 강승택(인천대)

A-2-5 딥 러닝에서의 딥 뉴럴 네트워크의 가중치 초기화 방법 020
홍정하(한국전자통신연구원), 여도엽(한국전자통신연구원)

목차

세션 B-1. 하드웨어 & 스마트팩토리

12월 16일(수요일) 14:40 ~ 15:40 (60분) | 좌장: 최계원 교수 (성균관대)

B-1-1 FPGA기반 저지연 데이터 이벤트 탐지 모듈 구현 022
윤기하(한국전자통신연구원), 김재인(한국전자통신연구원), 김성창(한국전자통신연구원)

B-1-2 AI(Deep Learning)을 이용한 S18650리튬이온배터리 SOC예측에 관한 연구 024
배정효(한국전기연구원), 진윤선(한국전기연구원), 백지국((주)아이이에스), 딘민차우(창원대),
김창순(창원대), 다오반권(창원대), 박민원(창원대)

B-1-3 지능형 엣지 컴퓨팅 시스템을 위한 소프트웨어 및 하드웨어 구현방안에 관한 연구 027
김재우(ICT융합특성화연구센터), 김동성(금오공과대)

B-1-4 하드웨어 성능에 따른 보행자 검출 성능평가 방법에 관한 연구 029
김희강(한국건설생활환경시험연구원), 손준우(한국건설생활환경시험연구원),
김창홍(한국건설생활환경시험연구원), 김지연(한국건설생활환경시험연구원),
조태식(한국건설생활환경시험연구원), 한동석(경북대)

B-1-5 BLDC 팬모터의 모터 DC 전압을 이용한 고장 진단 031
심준석(부산대), 조현진(부산대), 박정환(부산대), 김호원(부산대)

세션 B-2. 지능형 통신 I

12월 16일(수요일) 14:40 ~ 15:40 (60분) | 좌장: 김정곤 교수 (한국산업기술대)

B-2-1 연합학습을 통한 보안성 높은 무선통신 변조기술 탐지 모델에 관한 연구 033
서중하(한국과학기술원), 우태희(충남대), 박찬호(한국과학기술원), 강준혁(한국과학기술원)

B-2-2 인지 통신에서 자동 변조 분류를 위한 CNN 모델 설계 035
김승환(금오공과대), 김동성(금오공과대)

B-2-3 직렬 연결된 이미지 처리용 합성곱 신경망을 사용한 시간-주파수 채널 추정 037
김영찬(포항공과대), 장태준(포항공과대), 조준호(포항공과대)

B-2-4 IoT 기반의 전력 모니터링 시스템에서 군집화와 군집 대표 부하를 활용한 무손실 데이터 압축 방법 039
이지훈(광주과학기술원), 윤승욱(광주과학기술원), 황의석(광주과학기술원)

B-2-5 3GPP 실내환경에서 AI 기반 위치 추적 성능 개선 방안 041
오성현(한국산업기술대), 김정곤(한국산업기술대)

목차

세션 C-1. 영상처리

12월 16일(수요일) 17:00 ~ 18:00 (60분) | 좌장: 김정구 교수 (부산대)

C-1-1 Metric Learning 기반 Adversarial Example 탐지 가능성에 대한 연구 043
최석환(부산대), 신진명(부산대), 김정구(부산대), 최윤호*(부산대)

C-1-2 Data Augmentation & Augmenting Dataset for Facial Emotion Recognition 045
Jung Hwan Kim(경북대), Dong Seog Han(경북대)

C-1-3 특성기여도 분석 방법을 이용한 자동 증강의 영향에 관한 연구 047
김민기(경북대), 김재일(경북대)

C-1-4 데이터 증강과 전이학습을 활용한 주행환경에서의 감정 인식 모델의 성능 향상 실험 049
최준혁(포항공과대), 조현보(포항공과대)

C-1-5 공연 포스터의 이미지 특성을 활용한 딥러닝 기반 관객예측 051
조유정(경희대), 강경표(경희대), Yao hui(경희대), 권오병(경희대)

C-1-6 이미지 분류 네트워크에서의 효율적 훈련 기법에 대한 연구 053
배운지(동서울대), 이성진(동서울대)

세션 C-2. 지능형 차량

12월 16일(수요일) 17:00 ~ 18:00 (60분) | 좌장: 최지웅 교수 (DGIST)

C-2-1 심층 신경망 기반 차량 통신시스템의 신호 성상 분류 모델 055
김지훈(경북대), 한동석(경북대)

C-2-2 학습 기반 자율주행 제어 보조 모듈 설계 056
한경석(경북대)

C-2-3 차량 구성 요소 검출을 통한 오클루전 환경에서의 차량 검출에 관한 연구 058
배지환(국방과학연구소), 김태경(국방과학연구소)

C-2-4 ROS 기반 2륜 차량 시스템을 위한 딥러닝 기반 Monocular Visual Odometry 적용 060
최병찬(한양대), 남해운(한양대)

C-2-5 셀룰러 V2X 시스템을 이용한 자율 주행 시나리오 062
윤영진(경북대), 김지훈(경북대), 한동석(경북대)

C-2-6 V2X 기반 군집주행 차량과 주변 V2X 통신 차량간 통신 영향성 연구 064
구자후((주)웨이티즈), 한규동((주)웨이티즈), 정홍종((주)웨이티즈), 권순일((주)웨이티즈)

목차

세션 D-1. 자연어 처리

12월 17일(목요일) 09:00 ~ 10:00 (60분) | 좌장: 정방철 교수 (충남대)

D-1-1 텍스트 기반 지식요소 추출을 위한 온톨로지 활용 방안에 관한 연구 066
강유리(한화시스템)

D-1-2 한글, 영문, 숫자 및 특수기호가 혼합된 텍스트용 필기체 인식기의 구현 068
김홍숙(한국전자통신연구원), 김정시(한국전자통신연구원)

D-1-3 사전 훈련된 두 교차 연결을 통한 번역 성능 개선 070
오지은(한양대), 최용석(한양대)

D-1-4 심층학습을 이용한 한국어 음성변조에 관한 연구 072
김한(아주대), 이환용(아주대)

D-1-5 비디오 스크립트의 종결어미 태그를 이용한 비디오. 요약 방안 연구 074
신영주(헬스케어IT학과), 양진홍(인제대)

세션 D-2. 강화학습

12월 17일(목요일) 09:00 ~ 10:00 (60분) | 좌장: 황인태 교수 (전남대)

D-2-1 Max-Mean N-스텝 시간차 학습 076
황규영(한국기술교육대), 김주봉(한국기술교육대), 허주성(한국기술교육대), 한연희(한국기술교육대)

D-2-2 RF 충전 후방산란 CR 네트워크에서 효율적인 강화학습 기반 모드 최적화 078
오선애(숭실대), 신요안(숭실대)

D-2-3 MATLAB에서 회전형 도립 진자 제어를 위한 DDPG 기반 멀티에이전트 강화 학습 080
지창훈(한국기술교육대), 김주봉(한국기술교육대), 최호빈(한국기술교육대),
임현교(한국기술교육대), 한연희(한국기술교육대)

D-2-4 강화학습을 사용한 이미지 처리 기법 기반 적대적 사례 생성에 관한 연구 082
강효은(부산대), 김용수(부산대), 홍윤영(부산대), 이상현(부산대), 김호원(부산대)

D-2-5 수중 IoT 플러딩 영역 최적화를 위한 강화학습 프로세스에 대한 연구 084
강현우(한국폴리텍대학), 이성원(대구한의대), 서준호(경북대), 김동균(경북대)

D-2-6 수중 IoTCoAP에서 최적 메시지 타입 결정을 위한 강화 학습 기반 기계 학습 프로세스 086
이성원(대구한의대), 강현우(한국폴리텍대학), 서준호(경북대), 김동균(경북대)

목차

세션 E-1. 로봇 제어

12월 17일(목요일) 15:10 ~ 16:10 (60분) | 좌장: 강승택 교수 (인천대)

E-1-1 실내 환경 자율주행 로봇을 위한 객체 인지 모듈 개발 088
김명현(경희대), 김영인(경희대), 최인훈(경희대), 허의남(경희대)

E-1-2 인기 일체형 슬기 개인 로봇(PR-슬봇) 시스템 개발 연구 090
진용옥(경희대)

E-1-3 딥러닝 기반 음료 인식 및 로봇 제어 시스템 094
최인훈(경희대), 김명현(경희대), 김영인(경희대), 허의남(경희대)

E-1-4 딥러닝 기반 객체 인식을 통한 실내 위치 정보 탐색 자율 주행 로봇 개발 096
김영인(경희대), 최인훈(경희대), 김명현(경희대), 허의남(경희대)

E-1-5 심층 신경망 기반 추적기를 사용한 사용자 추종 로봇 098
손찬영(한국전자통신연구원), 이혜민(한국전자통신연구원), 이준구(한국전자통신연구원),
오지용(한국전자통신연구원)

세션 E-2. 영문 논문 세션 I

12월 17일(목요일) 15:10 ~ 16:10 (60분) | 좌장: 김동성 교수 (금오공대)

E-2-1 Reinforcement Learning Based Scheduling in Underwater NDN 100
Muhammad Toaha Raza Khan(경북대), Muhammad Saad Malik(경북대),
Muhammad Ashar Tariq(경북대), Md. Mahmudul Islam(경북대), Junho Seo(경북대),
Ru Yang(경북대), Dongkyun Kim(경북대)

E-2-2 Wind Speed Interval Forecasting Under Uncertainty Quantification Pattern Based on Deep Learning Method 102
Himawan Nurcahyanto(국민대), Aji Teguh Prihatno(국민대), Yeong Min Jang(국민대)

E-2-3 An Investigation on Feature Extraction and Feature Fusion Methods for Wearable Sensor-Based Human Activity Recognition 104
Nguyen Thi Hoai Thu(경북대), Dong Seog Han(경북대)

E-2-4 Artificial Intelligence based Internet of Things Security 106
Mitra Pooyandeh(동국대), Insoo Sohn(동국대)

E-2-5 Artificial Intelligence Platform Based for Smart Factory 108
Aji Teguh Prihatno(국민대), Himawan Nurcahyanto(국민대), Yeong Min Jang(국민대)

목차

세션 F-1. 지능형 환경

12월 17일(목요일) 16:20 ~ 17:20 (60분) | 좌장: 신수용 교수 (금오공대)

F-1-1 Prophet 모델을 사용한 기상데이터 예측 110
김준석(동의대), 김성희(동의대), 윤주상(동의대), 강재환(동의대)

F-1-2 전리층 총 전자량 데이터에 적용한 LSTM 기반의 지진 이상현상 탐지 111
조건우(광주과학기술원), 박동건(광주과학기술원), 김홍국(광주과학기술원)

F-1-3 미세먼지의 빅데이터/AI 분석 및 예측을 위한 IoT 측정 단말기 개발 113
우동식(대구가톨릭대), 백봉현((주)아르고스)

F-1-4 ICT 표준화전략맵 Ver.2021 기반 인공지능 적용 스마트헬스 분야 ICT 국제표준화 전략 연구 115
황유철(한국정보통신기술협회), 고준호(한국정보통신기술협회), 조수진(한국정보통신기술협회),
이영억(한국정보통신기술협회), 오구영(한국정보통신기술협회), 김대중(한국정보통신기술협회)

F-1-5 지능형 복합환경제어기 기반 토마토 병해 영상 분류시스템 설계 117
김태현(농촌진흥청 국립농업과학원), 이재수(농촌진흥청 국립농업과학원),
백정현(농촌진흥청 국립농업과학원), 최인찬(농촌진흥청 국립농업과학원),
곽강수(농촌진흥청 국립농업과학원), 김준용(서울대학교)

세션 F-2. 의료인공지능

12월 17일(목요일) 16:20 ~ 17:20 (60분) | 좌장: 최윤호 교수 (부산대)

F-2-1 뇌졸중 병변 분할을 위한 효율적인 U-Net 119
신현광(영남대), 최규상(영남대)

F-2-2 Octave U-net을 이용한 생체 의학 이미지 분할과 분류 방법 121
김화랑(경북대, 한국전자통신연구원), 김광주(한국전자통신연구원), 임길택(한국전자통신연구원),
최두현(경북대)

F-2-3 폐 영역 분할에서 적응형 활성화 함수의 유효성 검증 123
신호경(경북대), 김재일(경북대)

F-2-4 운전자의 얼굴을 검출하기 위한 딥러닝 얼굴 검출기와 객체 추적알고리즘 융합 시스템 125
유민우(경북대), 한동석(경북대)

F-2-5 YOLO 를 활용한 열화상 기반의 체온측정 인공지능 시스템 127
손진영(경북대), 김민영(경북대)

목차

세션 G-1. 영문 논문 세션 II

12월 18일(금요일) 09:00 ~ 10:00 (60분) | 좌장: 강승택 교수 (인천대)

G-1-1 Multi-agent clinical decision support systems: A survey 130
Sara Kaviani(동국대), Insoo Sohn(동국대)

G-1-2 Depthwise Separable Convolution for Facial Landmarks Detection 132
Savina Colaco(경북대), Dong Seog Han(경북대)

G-1-3 Machine Learning Approach to Detect and Classify Power Line Fault 134
Md. Habibur Rahman(국민대), Md. Morshed Alam(국민대), Yeong Min Jang(국민대)

G-1-4 On the Performance Gains of Federated Learning Edge Caching in Vehicular Internet of Things 136
Lilian C. Mutalemwa(조선대), Seokjoo Shin(조선대)

세션 G-2. 인공지능 응용 II

12월 18일(금요일) 09:00 ~ 10:00 (60분) | 좌장: 박혜영 교수 (경북대)

G-2-1 3D Depth 영상을 이용한 딥러닝 기반 이상 행위 인지 기술 139
김동칠(한국전자기술연구원), 박성주(한국전자기술연구원)

G-2-2 가속도 및 자이로 센서를 이용한 딥러닝 기반 행위인지 정확도 향상을 위한 심층 분석 처리 시스템 141
안영민(경희대), 이승진(경희대), 허의남(경희대)

G-2-3 자동 말투(Speech Style) 인식: 다자간 대화 상황에서의 화자인식 기술 개발 143
강가람(경희대), Jin Guangxun(경희대), 권오병(경희대)

G-2-4 PNCC와 합성곱 신경망을 이용한 능동 소나 표적 식별 145
이승우(국방과학연구소), 서익수(국방과학연구소), 한동석(경북대)

G-2-5 가사 Context 기반 음악간 유사도 산출에 관한 연구 146
박예은(포항공과대), 홍기석(포항공과대), 지봉준(포항공과대), 조현보(포항공과대)

G-2-6 효율적인 추천 시스템을 위한학습 컨텐츠의 상대적 특성 업데이트 방법 연구 148
윤봉영((주)태그하이브), 아가르왈 판카즈((주)태그하이브)

목차

세션 H-1. 지능형 통신 II

12월 18일(금요일) 10:10 ~ 11:10 (60분) | 좌장: 최준원 교수 (한양대)

H-1-1 Inherent Overestimation of DRL-Based Hybrid Beamforming for mmWave MIMO Systems: Behavioral Interpretation and Remedies 151
Dohyun Kim(University of Texas at Austi), Robert W. Heath Jr.(North Carolina State University)

H-1-2 Performance Analysis of Hybrid Deep Learning Model for Indoor localization 153
Alwin Poulose(경북대), Dong Seog Han(경북대)

H-1-3 Error Correction of Wearable Sensors using Sliding Window in Optical Camera Communication 156
Md. Faisal Ahmed(국민대), Israt Jahan(국민대), Yeong Min Jang(국민대)

H-1-4 Artificial Intelligence for Future 6G Cellular Networks: A Deep Learning Approach for Massive MIMO NOMA System 158
Muneeb Ahmad(금오공과대), Soo Young Shin(금오공과대)

H-1-5 수중 사물인터넷 환경에서 패킷 에러율 예측 및 네트워크 파라메터 최적화를 위한 기계학습 모델 160
김진홍(한국전자통신연구원), 이성원(대구한의대), 임길택(한국전자통신연구원), 김동균(경북대)

H-1-6 장소와 유형 분리 신경망 기반의 고정형 대상 인식 162
이형석(한국전자통신연구원), 이재호(한국전자통신연구원), 김도형(한국전자통신연구원), 류철(한국전자통신연구원)

세션 H-2. 지능형 보안 및 응용

12월 18일(금요일) 10:10 ~ 11:10 (60분) | 좌장: 이규만 교수 (건양대)

H-2-1 적대적 공격 방어를 위한 앙상블 기법에 관한 연구 164
김용수(스마트엠투엠), 강효은(부산대), 윤영여(부산대), 김민재(부산대), 김호원(부산대)

H-2-2 인공지능 프레임워크 신뢰성 표준 현황에 관한 연구 166
김성한(한국전자통신연구원), 최영환(한국전자통신연구원)

H-2-3 블록체인 기반 네트워크 관리 자동화 연구 168
최윤철(한국전자통신연구원), 박정수(한국전자통신연구원)

H-2-4 GEV 빔포밍을 위한 BiLSTM 기반 이진 마스크 추정 169
송일훈(광주과학기술원), 김홍국(광주과학기술원)

H-2-5 인공지능과 가상현실 간의 공진화 방안에 관한 연구 171
유건우(경희대), 황경화(경희대), 권오병(경희대)

H-2-6 Metric Learning 기반 Adversarial Example 탐지 가능성에 대한 연구 173
최석환(부산대), 신진명(부산대), 김정구(부산대), 최윤호(부산대)

제1회 한국 인공지능 학술대회

1st Korea Artificial Intelligence Conference

Proceeding

개인 맞춤형 광고 제공을 위한 기계 학습 기반 한국인 성별 및 나이 인식 모델 생성

홍승준, 김명섭, 허의남*
경희대학교

hongsj1022@khu.ac.kr, kms1205@khu.ac.kr, *johnhuh@khu.ac.kr

Machine learning based Korean gender and age recognition model generation to provide personalized advertisements

Seungjun Hong, Myeongseob Kim, Eui-Nam Huh*
Kyung hee Univ.

요 약

최근 언택트 시대를 직면함에 따라 얼굴인식을 통한 마스크 착용 정보, 출입 통제 시스템 등의 많은 기술들이 발명되었다. 본 논문에서는 얼굴인식을 이용하여 사용자에게 맞춤형 광고를 제공하는 서비스를 위한 한국인 성별 및 나이 인식 모델을 생성하는 과정과 결과를 제시하며, 향후 모델의 인식 정확도 개선 방안과 연구 방향을 제시한다.

Ⅰ. 서 론

최근 얼굴인식을 통한 여러가지 서비스들이 많이 등장하고 있다. 대표적인 예로 얼굴인식을 이용한 보안 시스템부터 관상 예측 어플리케이션, 가상 메이크업 어플리케이션까지 다양한 분야에서 활용되고 있다.

이러한 얼굴인식은 성별과 나이에 대한 학습을 통해 주어지는 개인의 얼굴에 따라 예측되는 결과를 활용한 맞춤형 광고 제공 서비스로 이어질 수 있다. 이 때 학습 모델의 정확도에 따라 제공되는 광고가 적합한지에 대한 판단이 이루어지기 때문에 적합한 데이터를 통한 학습 모델 생성이 중요하다.

본 논문에서는 이러한 서비스에 사용되는 국내 한국인을 대상으로 성별과 나이를 인식하는 모델을 개발하는 과정에서 데이터 셋의 구성이 학습 결과에 어떤 영향을 미치는지에 대해 논의한다. 같은 기계 학습 과정에서 여러 데이터 셋을 적용한 후 모델의 성능에 대한 비교 결과를 보여준다.

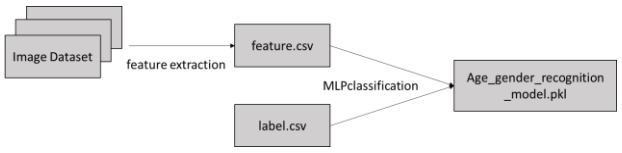
Ⅱ. 본 론

학습 모델 생성 과정

그림 1 은 기계 학습을 통한 학습 모델 생성 과정을 보여준다. 먼저 입력된 안면 이미지에서 얼굴을 인식한 뒤, 인식된 얼굴 내의 128 가지 특징 값을 추출하여 csv 파일로 저장한다. 각 이미지마다 성별과 나이를 라벨링한 csv 파일과 함께 추출한 특징 값을 scikit-learn 파이썬 모듈에서 제공되는 MLPclassifier 를 이용해 학습한다.

MLPclassifier 는 은닉층의 개수와 활성함수의 종류를 설정하여 입력 값을 다층 퍼셉트론을 통해 학습하는 함수이다. 본 논문에서는 128 개의 노드로 이루어진 128 개의 은닉층을 설정하였으며, 활성함수로는 relu 함수를 설정하여 학습하였다.

이미지를 학습함에 있어서 CNN(Convolutional neural network)을 이용한 학습이 적합하다는 것이 널리 알려져 있다[1]. 하지만 본 논문에서는 CNN 으로 이미지를 학습하기에 사양이 좋지 못한 개발 환경과 많은 시간이 소요된다는 점에 제한적이었기 때문에 계산량이 적은 MLP 를 사용하여 학습을 진행했다.



[그림 1] 학습 모델 생성 과정

데이터 셋의 구성과 라벨링

	LFW	AFAD	K-FACE
인종	전 세계	아시아	대한민국
데이터 양	약 13,000장	약 16만 장	약 432만 장
연령대 분류 라벨	Baby/ Child/ Youth/ Middle Age/ Senior	20대 / 30대 / 40대 / 50대	

[표 1] 학습에 활용한 데이터 셋

표 1 은 기계 학습에 사용한 데이터 셋의 종류와 구성을 보여준다. 학습 데이터로는 전 세계 유명인들로 구성된 LFW 데이터와 아시아인으로만 구성된 AFAD 데이터, 국내 aihub 에서 제공하는 한국인 안면 이미지(K-FACE) 데이터를 사용하였다[2-4].

데이터의 양은 각각 약 13,000 장, 약 16 만 장, 약 432 만 장으로 구성되어 있으며, 다양한 인종의 데이터로 이루어진 LFW 데이터 셋은 특정 연령에 대한 구분이 어려워 Baby, Child, Youth, Middle Age, Senior 로 라벨링을 하였다. 성별과 나이가 구분되어 제공된 AFAD 데이터 셋과 K-FACE 데이터 셋은 20 대, 30 대, 40 대, 50 대로 라벨링을 하였다.

400 명을 대상으로 30 단계의 조도, 20 개의 각도, 3 개의 표정, 6 종류의 가림 방법을 통해 구성한 K-FACE 데이터 셋은 표 2 와 같이 연령별 데이터가 편향되어 제공되는 한계점이 있었다. 이에 따라 20 대 데이터를 좌우 반전과 ±5°, ±10°씩 회전하는 작업을 거쳐 기존 데이터 양의 10 배로 늘렸고 다른 연령대의 데이터 양과 균일하게 학습을 진행하였다.

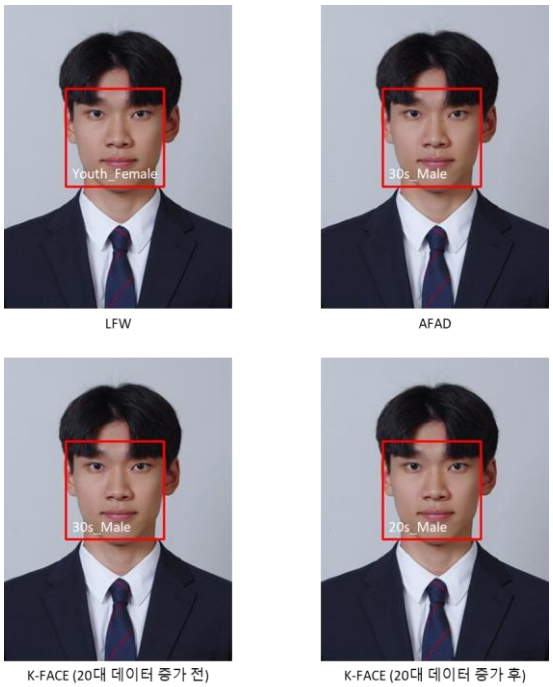
	남	여	계
20대	5	8	13
30대	67	76	143
40대	86	56	142
50대	56	46	102
계	214	186	400

[표 2] K-FACE 데이터의 연령별 분포(단위: 명)

Ⅲ. 결론

실험 결과

본 실험에서는 그림 2 와 같이 각 데이터 셋의 특징들을 학습한 모델들의 성능을 확인하기 위해 20 대 남성인 본 논문 저자의 사진을 사용하였다.



LFW AFAD

K-FACE (20대 데이터 증가 전) K-FACE (20대 데이터 증가 후)

[그림 2] 데이터 셋에 따른 학습 모델 별 성별 및 나이 인식 결과

LFW 데이터 셋을 이용한 학습 모델은 연령을 Youth 로 인식하였으나 성별을 여성으로 인식하였고, AFAD 데이터 셋을 이용한 학습 모델은 30 대 남성으로 인식하였다. 한국인의 성별과 나이 인식에 적합하지 않은 데이터를 사용하여 정확도가 낮음을 확인할 수 있다.

반면 K-FACE 데이터 셋을 이용한 학습 모델 중 20 대 데이터가 적게 학습된 모델은 30 대 남성으로 인식하였고, 20 대 데이터 양을 늘려 데이터의 편향 없이 학습된 모델은 20 대 남성으로 정확하게 인식하였다. 이는 학습에 적합한 데이터가 연령별 균일한 분포의 양으로 구성될 경우 정확도가 높음을 확인할 수 있다.

향후 학습 모델 개선 및 연구 방향

본 연구를 통해 기계학습을 통한 성별과 나이 인식 모델 생성 시 높은 정확도를 위해 인식 목표에 적합한 데이터 셋을 구성해야 하고, 분류할 클래스별 데이터의 양이 균일해야 함을 확인할 수 있다. 하지만 추가적인 실험으로 성별 인식에 비해 나이 인식의 정확도가 낮은 것을 확인할 수 있었다.

이를 개선하기 위한 향후 연구 방향으로는 안면 이미지 데이터의 매우 정밀한 부분을 특징으로 추출하여 학습할 수 있도록 하는 과정이 될 것이다. 또한 좀 더 좋은 환경에서 다층 퍼셉트론이 아닌 CNN 과 같은 다른 학습 알고리즘을 사용한 모델 생성과 수치화한 정확도와 함께 결과를 비교함으로써 성별과 나이를 이용한 맞춤형 광고 제공 서비스에 더욱 적합한 학습 모델이 사용될 것이다.

ACKNOWLEDGMENT

이 논문은 2020 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2019-0-01615, 온라인 동영상 광고를 제공하는 클라우드 기반의 무인 점포관리용 디지털사이니지 솔루션 개발)

참 고 문 헌

[1] A. Kriszhevky, I. Sutskever, and G. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, 2012

[2] “Labeled Faces in the Wild”, Computer vision lab at University of Massachusetts, http://vis-www.cs.umass.edu/lfw/”

[3] “The Asian Face Age Dataset(AFAD)”, GitHub Pages, https://afad-dataset.github.io/

[4] “한국인 안면 이미지 AI 데이터”, AIHub, 한국과학기술연구원, 한국정보화진흥원, http://kface.kist.re.kr/#/

인공지능 기반 지능형 범죄 위험 예측 및 대응 기술을 활용한 스마트 치안 기술 연구

백명선, 이용태, 박원주
한국전자통신연구원

sabman@etri.re.kr, ytlee@etri.re.kr, wjpark@etri.re.kr

A Study on the Smart Policing Technique using Intelligent Crime Risk Estimation and Response based on AI Technology

Myung-Sun Baek, Yong-Tae Lee, Wonjoo Park
ETRI

요 약

본 논문에서는 실제 치안현장에 효과적으로 적용할 수 있는 인공지능 기반의 지능형 치안기술을 설계하고 그 성능을 검증한다. 범죄의 위험상황을 초기에 감지하여 신속/정확한 상황 파악과 대응을 지원할 수 있도록, 경찰의 누적된 통계데이터를 기반으로 사전 접수 초기에 범죄의 유형을 인공지능 기술을 이용하여 예측할 수 있는 기술을 설계한다. 설계된 기술은 범죄 접수단계에서 범죄유형을 예측할 수 있으므로, 사건의 위험수준을 초기에 인지하여 신속한 초동대응방법을 도출하기 위한 참고 정보로 활용할 수 있다.

Ⅰ. 서 론

인공지능기술의 성숙과 더불어 다양한 인공지능 응용기술이 개발됨에 따라 치안분야에서도 인공지능을 활용한 지능형 스마트 치안 기술 개발에 대한 관심이 증가하고 있다 [1][2]. 국내 경찰청에서는 경찰대학 치안정책연구소내에 스마트치안지능센터, 치안과학기술연구실 등을 개설하여 치안기술의 지능화 및 첨단화를 위한 노력을 지속 수행하고 있다. 이러한 노력의 일환으로 국내 전체 범죄 발생건수는 2008 년 이후 점차 감소하고 있으며, 2015 년부터는 매년 크게 감소하고 있는 추세이다. 그러나 아직까지도 국민들은 가장 주된 사회불안 요인은 '범죄발생'이라 생각하고 있는 것으로 밝혀졌다 [4]. 따라서 범죄 발생에 대한 국민적 두려움을 해소하기 위해 추가적인 노력이 필요한 상황이다.

본 논문에서는 누적된 범죄/수사데이터와 실시간으로 수집되는 다양한 치안데이터를 효과적으로 관리 및 처리할 수 있는 빅데이터 처리기술 및 이러한 빅데이터를 효과적으로 활용할 수 있는 인공지능기술을 기반으로 하는 지능형 치안 기술을 설계 하고 설계된 기술의 성능을 검증한다. 설계된 기술은 인공지능 기반의 범죄 유형 추론 기술이다. 상기 기술을 설계/개발하기 위해 누적된 치안관련 통계데이터 및 가상의 치안데이터를 활용하였다. 치안관련 데이터는 피의자/피해자 등의 개인정보 및 민감정보를 포함하고 있으므로, 해당 정보를 개인정보 비식별화 등을 통해 대폭 수정하였으며, 공식 치안 데이터 양식을 준용하여 가상의 치안데이터를 대량으로 생성하여 활용하였다.

그림 1. 인공지능 기반 범죄 유형 추론 기술 활용방안

범죄 유형 추론 기술은 개발 단계에서는 상기 데이터들을 활용하여 인공지능 학습을 수행하고 이를 통해 신규 접수되는 사건의 내용 (텍스트 데이터)을 기반으로 범죄의 유형을 추론하는 기술이다. 해당 기술은 사건 내용을 포함하는 텍스트 데이터를 사용하여 총 21 종의 범죄 유형을 고려하여 해당 유형중 한가지의

범죄 유형으로 범죄유형을 추론하는 기술이다. 상기 개발된 기술을 활용하면 그림 1 에서와 같이 실제 치안현장에서 신규 사건 접수 시 사건유형/위험성 추론 결과를 참고하여 현장인력 파견/배치 등을 효과적으로 수행할 수 있다. 따라서 사건 발생시 신속한 대응으로 심각한 범죄로 확대되기 전 사전 대응이 가능하다.

Ⅱ. 범죄유형 추론 기술 설계

사건 접수 시 범죄 유형을 신속하게 판단하는 것은 범죄 현장 요원 파견 및 범죄 수사 방향 설정 등 초동 대응에 매우 중요한 요소이다. 본 논문에서는 범죄 접수 단계에서 텍스트기반의 사건데이터를 바탕으로 범죄 유형을 추론하는 기술을 설계한다. 기존에는 강력 범죄 7 종에 대해 간략하게 유형을 추론한 바 있다 [2]. 그러나 신종 범죄 출현 및 범죄의 유형이 다양화됨에 따라 기존 기술이 실제 현장을 반영하지 못하는 한계점과 더불어 특정 유형 분류 정확도가 현저히 낮다는 문제점도 있었다. 본 논문에서는 보다 확장된 21 종의 범죄 유형을 고려하였으며, 범죄유형 추론 방식을 고도화 하였다. 그림 1 은 범죄유형추론 시스템의 구조도를 보여준다. 그림에서와 같이 형사사법정보시스템(KICS: Korea Information System of Criminal Justice Services)양식에 따른 사건관련 텍스트데이터를 입력받는다. 입력받은 텍스트데이터로부터 유의미하다고 파악되는 단어들을 이용하여 키워드 사전을 생성한다. 생성된 키워드 사전 및 입력된 KICS 데이터를 이용하여 데이터 셋을 생성한다. 상기 데이터셋을 활용하여 딥러닝 기반의 인공지능 추론기를 생성한다. 최종적으로 GUI 가 탑재된 플랫폼형태의 시스템이 개발되어 실시간으로 신규 사건내용을 입력받아 해당 범죄유형을 추론하는 것이 가능하다. 추론기는 Deep Neural Network (DNN)형태의 추론기를 활용하였다 [4].

Ⅲ. 개발 기술의 성능 검증

본 절에서는 설계된 기술의 성능을 F1 score 를 기준으로 검증한다. 표 1 은 범죄유형 추론 기술의 성능을 검증한 결과이다. 본실험에서는 Ⅱ절에서 설명한 바와 같이 총 21 종의 범죄 유형을 포함하는 KICS 데이터를 활용하였다.

표 1 은 설계된 기술에 따른 범죄 유형 예측 성능을 보여준다. 표에서와 같이 각 유형별 추론 성능 및 전체 평균 추론 성능을 검증하였다. 표 1 에서와 같이 범죄유형 추론 성능은 0.89 점으로 매우 우수한 추론 성능을 제공하는 것을 확인할 수 있다.

표 1. 범죄유형 추론 기술 성능 검증 결과

	범죄유형	f1-score	samples
1	절도	0.98	192
2	손괴	0.91	80
3	공갈	0.98	62
4	약취유인	0.95	28
5	상해	0.83	82
6	폭행	0.85	107
7	강간	0.75	9
⋮	⋮	⋮	⋮
21	성풍속범죄	0.91	5
	전체 평균 성능	**0.89**	

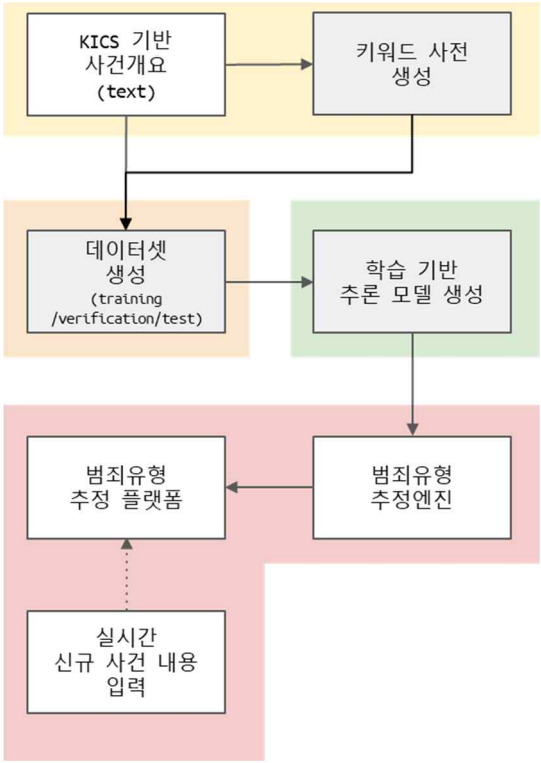


그림 2. 범죄유형 추론 시스템 구조도

Ⅴ. 결론

본 논문에서는 인공지능기반의 지능형 스마트 치안 기술을 설계하고 성능을 검증하였다. 개발된 기술은 딥러닝 기반의 기계학습 기술을 기반으로 텍스트 기반의 사건 접수 데이터를 바탕으로 해당 범죄의 유형을 추론하는 것이 가능하다. 해당 기술은 경찰의 초동대응에 도움을 주는 사건 유형 특정을 신속하게 수행할 수 있도록 지원한다.

ACKNOWLEDGMENT

이 논문은 2020 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] 장광호, “스마트치안”, 2020. 06.

[2] 백명선 외 4, “효과적 대응을 위한 기계학습 기반의 범죄 유형 및 범죄 위험스코어 예측 기술 연구”, 2020 한국통신학회 하계 종합학술대회, 2020. 08.

[3] M. Kim, D.-g. Lee, H. Shin, “Semi-supervised learning for hierarchically structured networks,” Pattern Recognition, vol 95, pp. 191-200, Nov. 2019.

[4] 통계청, 2018 년 사회조사(가족, 교육, 보건, 안전, 환경), 2018, (http//kostat.go.kr)

[4] M.-S. Baek, S. Kwak, J.-Y. Jung, H. M. Kim, D.-J. Choi, “Implementation Methodologies of Deep Learning-Based Signal Detection for Conventional MIMO Transmitters,” IEEE Trans. Broadcasting, vol. 65, no. 3, pp. 636-642, Sept. 2019.

AWS 딥렌즈와 엣지 클라우드를 이용한 성별 및 나이 맞춤형 실시간 광고 스트리밍 서비스

김서현, 김명섭, 홍승준, 김명현, 허의남*
경희대학교

seok02h@khu.ac.kr, kms1205@khu.ac.kr, hongsj1022@khu.ac.kr,
freckie@khu.ac.kr, *johnhuh@khu.ac.kr

Real-time advertisement streaming service customized to gender and age using AWS DeepLens and Edge Cloud

Seo-Hyun Kim, Myeongseob Kim, Seungjun Hong, Myung-Hyun Kim, Eui-Nam Huh*
Kyung Hee Univ.

요 약

AWS DeepLens는 딥러닝이 지원되는 비디오카메라이다. 본 논문에서는 사용자의 이용 시간이 짧은 사이니지에서 사용자 맞춤형 광고 서비스를 하기 위해 AWS DeepLens와 엣지 클라우드를 연계하여 별도의 제어 없이도 실시간으로 얼굴을 인식하여 맞춤형 광고를 제공하는 서비스를 구현한다.

Ⅰ. 서 론

AWS DeepLens는 세계 최초로 딥러닝이 지원되는 개발자용 비디오카메라이다[1]. AWS DeepLens를 사용하여 실시간으로 사람의 행동이나 얼굴, 객체 등을 딥러닝 모델을 통해 인식할 수 있다.

이는 실시간으로 인식한 객체의 맞춤형 서비스를 가능하게 한다. 그 예로 인식된 사람의 나이와 성별을 판단하여 맞춤형 광고를 제공할 수 있다. 또한 엣지 클라우드와 연동한다면 저지연 고성능 서비스를 구현할 수 있다.

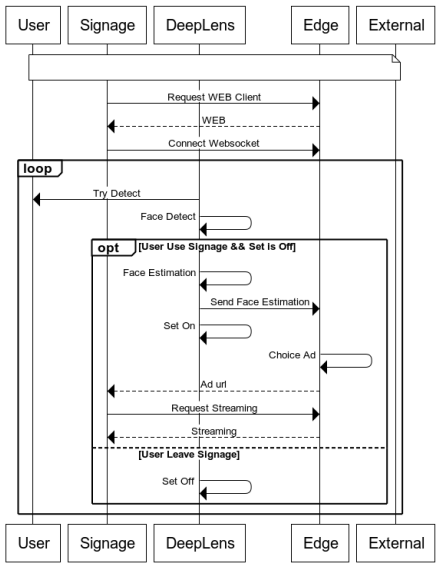
본 논문에서는 사용자 이용 시간이 짧아 실시간으로 처리해야 하는 사이니지에서도 맞춤형 광고를 제공하기 위한 방안을 제시한다. AWS DeepLens에 입력되는 얼굴에 따라 성별과 나이를 인식하고, 인식한 결과의 빠른처리를 위해 엣지 클라우드에서 광고 캐싱 및 선정하여 그에 따른 맞춤형 광고를 제공하는 실시간 광고 스트리밍 서비스 아키텍처를 설계하고 구현한다.

Ⅱ. 본 론

서비스 동작 과정

그림 1은 실시간 광고 스트리밍 서비스의 동작 과정을 나타내며 자세한 과정은 다음과 같다. 학습에 쓰인 모델은 K-FACE 데이터 셋을 이용해 학습시킨 모델을 사용하였다.

1. 엣지 클라우드와 연결되어 있는 사이니지에서 웹 서비스를 실행한다.
2. AWS DeepLens에서 실시간으로 촬영을 시작한다.
3. 촬영 중 얼굴이 인식되면 인식된 얼굴의 성별과 나이를 예측한다.
4. 성별과 나이는 학습시킨 모델을 활용하여 예측한다
5. 성별과 나이에 대한 예측 결과는 API를 이용하여 엣지 클라우드로 전송한다.
6. 엣지 클라우드에서는 API에 담긴 성별과 나이 정보에 맞는 광고가 선정된다.
7. 웹소켓을 통해 광고 url이 사이니지에 전달되어 광고 영상이 갱신된다.

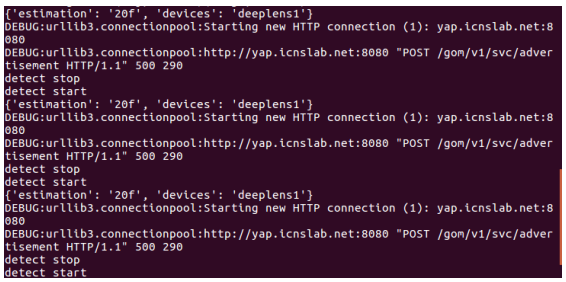


[그림 1] 광고 스트리밍 시퀀스 다이어그램

본 시나리오에서 스트리밍 서비스가 동작하는 중에 계속된 얼굴 인식이 있을 경우, 반복적인 API 호출과 광고 갱신이 이루어질 수 있다. 따라서 API를 한 번 호출한 후 인식한 얼굴이 사라질 때까지 플래그를 두어 API 호출을 멈추게 하였다. AWS DeepLens에서 인

식했던 얼굴이 사라진 후 새로운 얼굴이 인식되면 위 과정을 반복한다.

그림 2는 반복되는 얼굴 인식과 성별과 나이 예측 결과에 따른 API 호출 로그를 보여준다. 모델의 성별 예측에 관한 정확도는 높지만, 나이 예측에 관한 정확도는 개선이 필요하다.



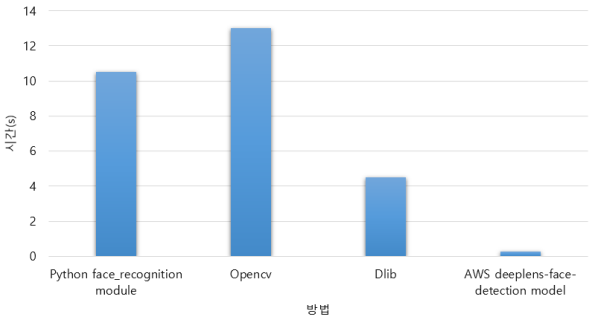
[그림 2] AWS DeepLens를 이용한
얼굴 인식 로그

서비스 성능 개선

사이니지는 사용자들의 이용 시간이 짧아 맞춤형 광고를 제공하기 위해서는 얼굴 인식 및 예측하는 과정에서 실시간으로 처리되어야 한다. 서비스 시간을 단축시키기 위해 얼굴을 인식하는 다양한 방법을 모색해보았다.

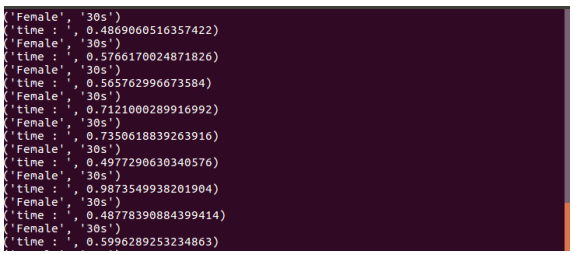
차트 1은 딥렌즈에서 얼굴을 인식하기 위해 사용한 방법에 따른 얼굴 인식 소요시간을 보여준다. Python의 face_recognition 모듈에서 제공하는 함수[2]를 이용한 경우에는 10~11초, Opencv 모듈에서 제공하는 함수[3]를 이용한 경우에는 13~14초, Dlib 모듈에서 제공하는 함수[4]를 이용한 경우에는 4~5초가 소요되었다.

반면, deeplens-face-detection model[5]을 이용한 얼굴 인식에는 0.2~0.3초가 소요되었다. 매 실행마다 소요시간에 대한 약간의 오차는 있었지만 deeplens-face-detection model을 사용한 얼굴 인식이 월등히 빨랐다.



[차트 1] 모듈에 따른 얼굴인식 소요시간 비교

그림 3은 AWS DeepLens에서 프레임을 가져오고 deeplens-face-detection model을 사용해 얼굴인식 후 인식된 얼굴을 학습 모델을 통해 나이와 성별을 판단하기까지 소요된 시간을 나타내는 로그이다. 약 0.4~1초가 소요됨을 알 수 있다.



[그림 3] AWS DeepLens내의
프레임 처리 결과 및 소요시간

Ⅲ. 결 론

본 논문에서는 사용자의 이용 시간이 작은 사이니지에서 사용자 맞춤형 광고 서비스를 하기 위해 AWS DeepLens와 엣지 클라우드를 활용하여 별도의 제어 없이도 얼굴을 인식하여 맞춤형 광고를 제공하는 서비스를 구현하고 성능 개선을 위해 다양한 얼굴 인식 모델을 비교해보았다.

이를 통해 별도의 제어 없이 맞춤형 광고를 제공할 수 있는 서비스를 개발하였으며, deeplens-face-detecion model을 이용하여 서비스 작동 시간을 단축하고 실시간 서비스를 구현할 수 있었다

그러나 나이를 추측함에 있어 모델의 정확도가 낮아 맞춤형 광고의 정확도가 낮다는 점을 확인하였고 이를 개선하기 위해 나이와 성별을 추측하는 모델에서 나이 추측에 대한 정확도를 높이는 방향으로 연구를 진행할 예정이다.

ACKNOWLEDGMENT

이 논문은 2020년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2019-0-01615, 온라인 동영상 광고를 제공하는 클라우드 기반의 무인 점포관리용 디지털사이니지 솔루션 개발)

참 고 문 헌

[1] “AWS DeepLens”, AWS,
https://aws.amazon.com/ko/deeplens/

[2] “face_recognition package”, Face Recognition,
https://face-recognition.readthedocs.io/en/latest/face_recognition.html

[3] “Face Detection in Video Capture”, Opencv,
https://docs.opencv.org/3.4/df/d6c/tutorial_js_face_detection_camera.html

[4] “face_detector”, dlib,
http://dlib.net/face_detector.py.html

[5], “AWS DeepLens Sample Projects Overview “, AWS DeepLens 개발자 안내서,
https://docs.aws.amazon.com/deeplens/latest/dg/deeplens-templated-projects-overview.html

사이니지 사용자 맞춤형 광고를 제공하기 위한 엣지 클라우드 기반 얼굴 인식 모듈과 맞춤형 광고 서비스 개발

김명섭, 홍승준, 김서현, 김명현, 허의남*
경희대학교

kms1205@khu.ac.kr, hongsj1022@khu.ac.kr, seok02h@khu.ac.kr,
freckie@khu.ac.kr, *johnhuh@khu.ac.kr

Development of edge cloud-based face recognition module and advertisement service to provide advanced signage advertisement

Myeongseob Kim, Seungjun Hong, Seo-Hyun Kim, myung-Hyun Kim, Eui-Nam Huh*
Kyung Hee University

요 약

본 논문에서는 사용자의 서비스 이용 시간이 짧은 사이니지에서 사용자 맞춤형 광고를 제공하기 위해 사이니지를 사용하는 사용자의 얼굴을 머신 러닝 모델을 가지고 인식하여 나이와 성별을 추측한 뒤 맞춤형 광고를 제공하는 엣지 클라우드 서비스를 제안하고 개발하여 평가한다.

Ⅰ. 서 론

무인화기기가 증가함에 따라 디지털 사이니지 사용자에 맞춤형 광고를 제공하고자 하는 맞춤형 광고 서비스를 필요로 하고 있다.

이를 지원하기 위한 방안으로 다양한 서비스와 데이터 집약적인 분석을 제공하는 클라우드 컴퓨팅을 고려할 수 있지만, 클라우드 컴퓨팅 작업이 중앙 집중적이고 원격으로 동작하기 때문에 사용자의 이용 시간이 짧은 사이니지와 같은 상황에서는 서비스를 실시간으로 지원하기 힘들다는 단점이 있다.

이를 해결하기 위해 사용자들과 가까운 위치에 엣지 클라우드 서버를 설치하고 분산된 리소스를 활용하여 실시간으로 사용자 맞춤형 광고를 제공하고자 한다.

본 논문에서는 사이니지 근처에 설치된 라즈베리파이 기반 엣지 클라우드에서 머신 러닝 모델을 사용한 얼굴 인식 및 인식된 얼굴의 성별과 나이에 맞는 맞춤형 광고를 제공하는 서비스를 설계 구현하고 평가하고자 한다.

Ⅱ. 본 론

얼굴 인식 및 맞춤형 광고 서비스 시나리오

라즈베리파이 클러스터와 Docker Swarm 을 사용하여 엣지 클라우드를 구축[1]하였고 사이니지에서 보내온 얼굴 이미지 사진을 클라우드에서 판단하지 않고 엣지 클라우드에서 판단하여 빠르게 맞춤형 광고를 송출 할 수 있도록 사이니지와 엣지 클라우드간 시나리오를 그림 1 과 같이 설계하였으며 설명은 아래와 같다.

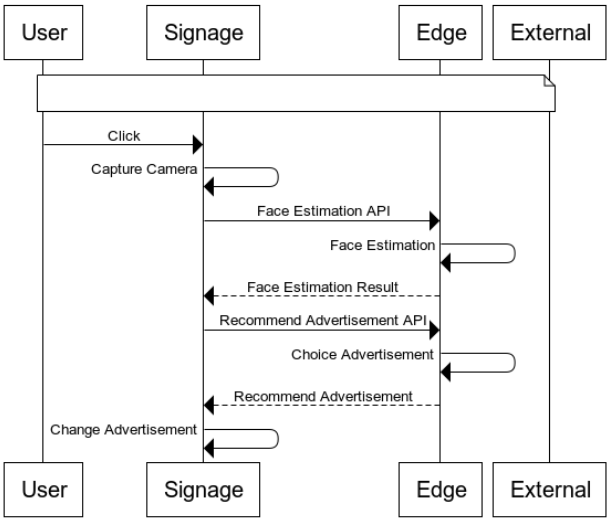


그림 1 시퀀스 다이어그램

1. 사이니지는 웹 기반의 페이지를 화면에 출력
2. 버튼 클릭시 사이니지에 연결된 카메라로 얼굴을 캡처
3. 캡처된 얼굴을 Restful API 를 사용하여 전송
4. 엣지 클라우드의 라우팅 기능을 활용하여 패킷을 엣지 클라우드에서 처리하도록 라우팅
5. 엣지 클라우드 내 얼굴 인식 API 서버가 머신 러닝 모델을 사용하여 얼굴의 성별과 나이를 추측하여 반환
6. 사이니지는 추측된 결과를 다시 Restful API 를 사용하여 맞춤형 광고 요청
7. 광고 API 서버는 엣지 클라우드 스토리지에 맞춤형 광고를 선정하여 사이니지로 광고 스트리밍 url 을 반환

8. 사이니지 웹 페이지는 광고 스트리밍 url 을 화면에 출력하여 사용자에게 맞춤형 광고를 제공

마이크로 서비스 아키텍처

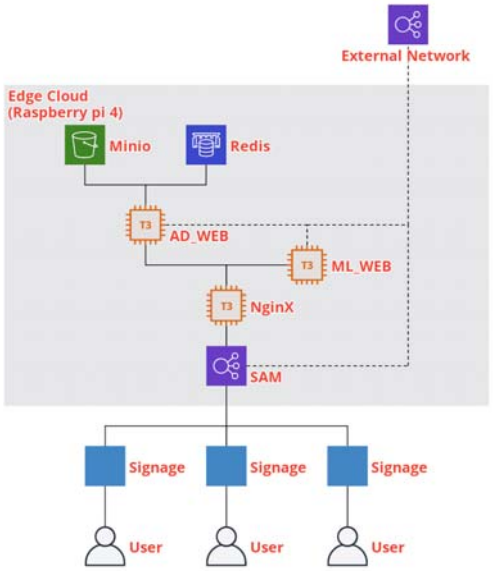


그림 2. 마이크로서비스 아키텍처

앞 절에서 설계한 시나리오에 따라 필요한 기능들을 마이크로 서비스로 나누어 개발 한 뒤 그림 2 와 같이 아키텍처를 구성하여 엣지 클라우드에 배치하였다. 각 서비스들의 설명은 다음과 같다.

NginX: 사용자의 서비스 요청시 필요한 마이크로 서비스로 요청을 라우팅 시키기 위한 리버스 프록시로 오픈 소스인 NginX 를 활용하였다.

ML_WEB: 사이니지가 전송하는 얼굴 이미지 데이터를 머신 러닝 모델을 활용하여 성별과 나이대를 추측하는 얼굴 인식 API 서버로 Python Flask 와 OpenCV, dlib 등을 사용하여 개발하였다. 얼굴 인식 모델은 오픈 소스[2]를 활용하였다.

AD_WEB: 광고 영상을 관리하고 클라우드에 광고 집행을 알리며 사이니지로 추측된 나이, 성별에 맞는 맞춤형 광고를 선정하여 광고 스트리밍 url 을 반환하는 광고 API 서버로 Python Flask 를 사용하여 개발하였다.

Minio: 엣지 클라우드에서 송출할 광고를 저장하기 위한 엣지 클라우드 스토리지 서버로 오브젝트 스토리지 오픈소스인 Minio 를 사용하였다.

Redis: 추측 결과와 광고 매핑을 위한 엣지 클라우드의 캐시 서버로 Key-Value 기반의 인메모리 DB 오픈 소스인 Redis 를 사용하였다.

Ⅲ. 결 론

본 논문에서는 사용자의 서비스 이용 시간이 짧은 사이니지에서 사용자 맞춤형 광고를 제공하기 위해 사이니지를 사용하는 사용자의 얼굴을 머신 러닝 모델을 가지고 인식하여 나이와 성별을 추측한 뒤 맞춤형 광고를 제공하는 엣지 클라우드 및 서비스를 제안하고 개발하였다.

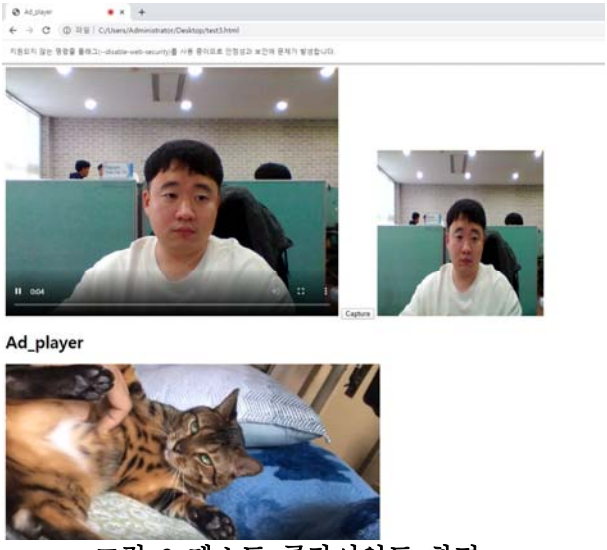


그림 3 테스트 클라이언트 화면

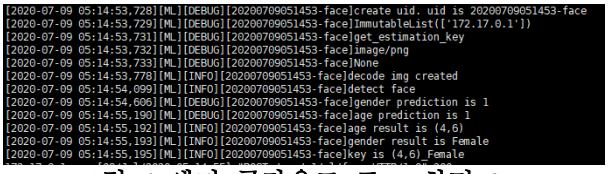


그림 4 엣지 클라우드 로그 화면 2

그림 2 는 서비스 테스트시 사용한 클라이언트 화면이고 그림 3 은 엣지 클라우드 로그 화면이다. 테스트 결과 시나리오대로 서비스가 동작함을 확인하였으나 사용한 얼굴 인식 모델의 정확도가 낮아 30 대 남성을 20 대 남성으로 예측하였고, 이로 인해 맞춤형 광고의 정확도가 떨어짐을 확인하였다. 또한 그림 3 에서 보여지듯이 엣지 클라우드에서 얼굴을 인식하는 과정에서 평균적으로 1.5 초가량 소요되는 것을 확인하였다.

향후 얼굴 인식 모델을 개선하여 정확도를 높임과 동시에 엣지 클라우드에서 이런 머신 러닝 모델의 개선 사항을 반영 할 수 있도록 기능을 추가가 필요해 보이며 현재 얼굴 인식 시간을 단축시키기 위하여 사용자 클릭을 기다리는 대신 엣지 클라우드가 자동으로 사람을 인식하는 시나리오로 개선하는 방법을 연구할 예정이다.

ACKNOWLEDGMENT

이 논문은 2020 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2019-0-01615, 온라인 동영상 광고를 제공하는 클라우드 기반의 무인점포관리용 디지털 사이니지 솔루션 개발).

교신저자: 허의남

참 고 문 헌

[1] 김윤곤, Tien-Dung Nguyen, 이가원, 허의남, "분산 클라우드 컴퓨팅을 위해 클라우드 연동을 고려한 에지 클라우드의 시스템 구조", 2018 한국컴퓨터종합학술대회 (KCC2018), 20180621, pp.1312-1314

[2] https://github.com/kairess/age_gender_estimation

국방 인공지능 동향과 국내 적용에 관한 연구

김상민
한화시스템

smkim0153@hanwha.com

A Study on the Trend of Defense Artificial Intelligence & Application of Defense Field

Sang Min Kim
Hanwha Systems.

요 약

최근 민간분야뿐만 아니라 국방분야에서도 인공지능 기술이 다양하게 활용되고 있다. 선진국들은 미래전을 대비하여 군사적 목적의 인공지능이 탑재된 무인체계의 전력화를 계획하고 있다. 하지만 현재 한국군이 보유한 인공지능 기술과 미래전을 위한 준비는 미흡한 상황이며 한국군에 최적화된 인공지능 기술 개발이 시급히 요구된다. 본 논문은 해외 사례를 통해 국내 수준을 파악하고 지휘통제 분야의 우선적 지능화 적용을 제안하며 그 방향에 대하여 고찰한다.

Ⅰ. 서 론

최근 인공지능 기반 서비스의 상용화가 성공적으로 이뤄짐에 따라 국내에서도 인공지능 기술의 비약적인 발전을 쉽게 체감할 수 있다. 음성인식 기술을 이용한 스마트 스피커, 자연어처리 기술을 이용한 챗봇 서비스, 영상인식기술을 이용한 자율주행 등은 일상생활에 적용된 대표적인 상용화 성공 사례라고 할 수 있다. 인공지능 기술은 일상생활뿐만 아니라 감염병 진단 보조, 화재 감지, 지진 예측 등과 같은 재난상황 대응을 위해서도 적극 사용되고 있으며 인공지능 기반의 미래전장을 대비하여 군사적 목적을 위한 국방 지능화 시도도 꾸준히 이어지고 있다.[1] 한편 민간분야의 상용화 기술을 국방분야에 단순 적용하는 방법으로는 만족스러운 성능을 기대하기 어렵다.[2] 인공지능 기술이 적용될 민간분야와 국방분야의 현실 세계가 매우 다르기 때문이다. 따라서 군사적 목적을 위한 별도의 인공지능 기술을 개발하거나 민군협력을 통해서 민군 사이의 간극을 조율하며 국방 지능화 개발을 이끌어야 한다. 미국은 미국방위고등연구계획국(DARPA)을 중심으로 1960 년대부터 인공지능 연구를 진행해왔으며 최근 20 억 달러를 투자하는 등 국방 인공지능을 위한 개발을 진행중이다. 중국은 민간분야에서 보유중인 정상급 인공지능 기술을 바탕으로 대미 군사력 열세를 극복하고자 민군융합 전략 등을 기반으로 2030 년 세계 최정상의 인공지능 기술력을 확보하려는 계획을 가지고 있다. 한편 우리나라는 2019 년 육군에 '인공지능연구발전처'를 창설하여 미래전에 대비하고 있으며 국방 지능화를 위한 동향 연구가 주로 진행되어왔다.[2]~[4] 하지만 선진국 대비 보유중인 인공지능 기술력과 군사적 목적의 인공지능 기술 개발을 위한 준비가 여전히 부족한 상황이다. 국방 지능화를 위한 출발이 늦은 만큼 선택과 집중 관점에서 효율적인 개발 방향이 필요하다.

본 논문은 국방 인공지능 적용 사례를 살펴보고 우리군에 우선적으로 도입되어야 할 국방 지능화 분야를 제시함으로써 신속한 국방 지능화에 기여하고자 한다.

Ⅱ. 국방 인공지능 동향

국방 인공지능 적용 사례는 매우 다양하나 본 논문에서는 크게 공격용과 방어용으로 나누어 소개한다.

먼저 공격용 사례는 미해군 로봇 잠수함에 탑재할 수 있는 인공지능 알고리즘 클로스 개발이다. 이 알고리즘이 탑재된 로봇 잠수함은 운용자 개입 없이 스스로 어뢰를 발사하는 살상 능력까지 갖추었다. 2022 년 실전배치를 목표로 하고 있으며 미 해군이 보잉에 의뢰해 제작한 로봇 잠수함(오르카)에 탑재될 계획이다.

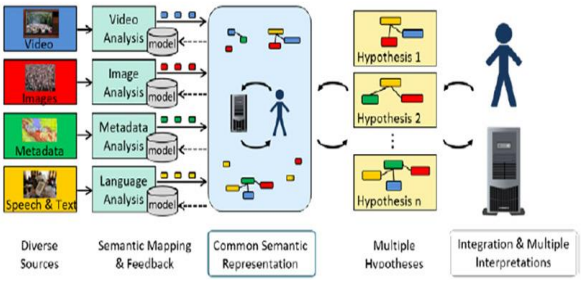


Figure 1 DARPA 의 지휘통제 지능화 개념도

방어용 사례로는 DARPA 의 인공지능 기반의 지휘통제 지능화를 위한 지능형 전장정보융합 학습모델과 비정형 전장정보자료 기반의 전장상황 분석 정보융합 학습엔진 개발이다.[5] 해당 프로젝트의 목적은 정보 분석가에 의존하여 전장상황을 분석하는 기존의 방법이 보유한 한계점을 벗어나 인공지능 기반의 분석을 통해 복수가설을 생성하고 지휘관의 결심을 지원하는데 있다.

Ⅲ. 국방 인공지능 개발 방향

본 논문은 기동/화력/지휘통제 등 전장기능 중 우선적으로 지능화가 적용되어야 할 분야를 제안한다.

국방분야에서 인공지능 기술의 활용은 크게 두 가지로 생각 할 수 있다.[2] 하나는 살상을 목적으로 하는 화력 등 공격 분야의 적용이며, 하나는 지휘통제와 같은 공격 지원 분야에서의 적용이다. 전자의 경우 살상 결과에 대한 책임과 살상의 기준 등 법적, 윤리적 문제를 우려하여 국제규범 창출을 위한 시도가 지속되고 있다.[6] 또한 알고리즘의 안정성 등 기술적인 문제 역시 가지고 있다. 현재 우리군이 보유한 군사용 인공지능 기술과 준비상태를 감안하여 공격 목적보다는 공격을 지원하는 분야의 지능화를 우선적으로 고려해야 한다. 본 논문은 특히 지휘통제 지능화의 우선적 적용을 다음 세 가지 측면에서 제안한다.

첫째, 민간분야에서 상용화된 기술의 접목이 용이하다. 인공지능 기반의 지휘통제체계 개발을 위하여 지능화가 필요한 대표적인 분야는 전장 정보융합 및 상황분석, 적 위협 및 방책 분석, 계획수립 및 방책추천 등이다. 이 분야의 지능화는 민간분야의 기술 적용이 가능하다. 예를 들어 음성인식, 통·번역, 정보검색, 기계학습 등이 이에 속한다. 민간분야 기술만으로는 개발의 어려움이 있는 제대 별 지휘결심지원을 위한 지식베이스 구축, 임무 목적형 학습 및 서비스화 등의 기술은 국방 핵심기술로 연구개발이 필요하며 민군협력으로 문제를 해결해나가야 한다.[7]

둘째, 한국군이 보유한 지휘통제체계의 한계점을 보완할 수 있다. 복잡하고 다양한 전장상황을 대비하여 지휘관의 전장인식을 지능화하는 전장관리체계(KJCCS, JFOS-K, ATCIS, MIMS 등)를 사용하고 있으나, 현 수준은 감시 정찰체계로부터 수집된 정보를 수작업으로 분석/정리하여 시스템에 저장 관리하는 정도이다. 육군전술지휘정보체계(ATCIS)를 국방온톨로지 기반으로 지능화하려는 연구가 진행되었으나[8] 지휘관의 야전 경험과 군사 전문성을 규칙기반에서 활용하는 수준으로 복잡한 상황과 불확실한 정보가 만연한 미래전장을 고려한다면 지능화를 위한 개발이 추가로 요구된다. 지휘통제체계가 지능화된다면 분석가의 경험에 의한 편견과 그로 인하여 간과될 수 있는 정보들을 좀 더 객관적으로 파악함으로써 수집된 정보들을 올바르게 해석할 확률을 높일 수 있다. 현재까지 보유한 지휘통체 지능화 경험과 관련 연구를 활용한다면 지휘통제분야의 지능화 수준은 빠르게 고도화 될 수 있을 것이다.

셋째, 지휘통제체계의 지능화에 따른 파급력이 크다. 기동/화력/지휘통제 등 전장기능 중 지휘통제의 범위를 벗어나는 전장기능은 없다. 즉 인공지능 기반의 지휘통제체계는 한국군 전력의 지능화를 직/간접적으로 가능케한다.

지금까지 지휘통제분야의 우선적인 지능화 적용에 대하여 논하였다. 더 나아가 성공적인 지휘통제 지능화 기술 개발 및 적용을 위해서 고려되어야 할 사항을 소개하고자 한다.

첫째, 클라우드 기반의 인공지능 서비스는 개발과 운영 측면에서 모둥 효율적이다. 하지만 현재 우리군은 제한적인 클라우드 서비스를 고려하고 있으며 시범사업을 준비하는 수준이다.[9] 효과적인 기술 적용을 위해서는 인공지능 알고리즘뿐만 아니라 군용 클라우딩 기술 등 지능화를 위한 인프라 개발이 함께 진행되어야 한다.

둘째, 현 인공지능 기술의 핵심인 딥러닝 알고리즘은 어떠한 정보로 학습을 하느냐에 따라 그 성능이 크게 좌우된다. 즉 성능이 보장된 군사적 목적의 인공지능 기술을 개발하기 위해서는 국방분야에서 실제 사용되는 충분한 양의 실자료 확보가 필수적이다. 하지만 보안상의 이유로 해당 자료가 군 외부에 공개되기는 쉽지 않을뿐더러 실자료는 학습을 위한 정보로 정제되어 있지 않기 때문에 정보의 전처리 과정 등이 요구된다. 따라서 인공지능의 기술적 배경을 고려한 실자료의 저장, 관리 방법 등의 충분한 논의가 필요하다.

Ⅳ. 결론

본 논문은 국방 인공지능 동향과 현재 국내 수준을 고려하여 여러 전장기능 중 지휘통제 분야에 우선적인 지능화 적용을 제안한다. 민간분야에서 보유한 인공지능 기술을 적극 활용하고 앞서 다룬 인공지능의 기술적 배경을 고려한 지능화 인프라 개발 등의 논의가 함께 이루어진다면 국내 국방 지능화 측면에서 유의미한 결과를 얻을 수 있을 것이다.

참 고 문 헌

[1] ETRI Insight 표준화 동향 2020-1.

[2] Jong-Kwan Lee, “Future Warfare and Military Artificial Intelligence Systems”, The Journal of Korean Institute of Communications and Information Sciences 44(4), 2019.4, 782-790.

[3] 국방 인공지능(AI) 활용방안 연구, 11-1290000-000628-01, 2017.3.

[4] Kyuyong Shin,"The Current Applications and Future Directions of Artificial Intelligence for Military Logistics",Journal of Digital Contents Society 20(12), 2019.12, 2433-2444(12 pages).

[5] Boyan Onyshkevych,"Active Interpretaion of Disparate Alternatives(AIDA)",DARPA,2017.

[6] 장기영. 국제.지역연구, 제 29 권 제 1 호 2020 201 - 226

[7] Sangheun Shim,” The Usage and Study of Intelligent Information Technologies for the Battlefield Awareness and Decision Support of Intelligent C2”,Journal of the KIMST,2019.

[8] Donghee Yoo,"Intelligent Army Tactical Command Information System based on National Defense Ontology",Journal of the Korea Society of Computer and Information 18(3), 2013.3, 79-89(11 pages).

[9] Jahoon Koo, “Design of Security Architecture for the Cloud-Based Korea Military Command and Control System”, The Journal of Korean Institute of Communications and information Sciences 45(2), 2020.2, 400-408

K-스마트드라마 구현을 위한 인공지능 활용 방법 연구

최지애, 이일호*, 권오병**
칼빈대학교, *칼빈대학교, **경희대학교

cja1321@gmail.com, *eli0410@hanmail.net, **obkwon@khu.ac.kr

Applying Artificial Intelligence to Realize K-Smart Drama

Ji Ae Choi, *Il Ho Lee, Ohbyung Kwon**
Calvin Univ., * Calvin Univ., **Kyung Hee Univ.

요 약

정부는 최근 인공지능(AI) 등 분야별 신기술 개발과 적용을 확대해 나갈 예정이며 콘텐츠와 더불어 온라인플랫폼이 해외로 진출할 수 있도록 중소 콘텐츠기업의 서버 구축과 마케팅 등 뉴미디어 콘텐츠에 대한 제작 지원을 확대하는 중이다. 이러한 상황에서 콘텐츠와 데이터, 인공지능 등의 결합을 통해 고부가가치 콘텐츠를 육성하며 영화의 경우 콘텐츠 배경 장소에서 활성화되는 위치기반 실감 콘텐츠, 인공지능 활용 콘텐츠 제작을 지원할 필요가 있다. 본 연구의 목적은 인공지능이 가미된 혁신적 드라마로서 K-스마트드라마를 제안한다.

Ⅰ. 서론

연극의 역사도 아날로그 시대에서 디지털 시대로의 대전환을 맞이하고 있는 것처럼 공연예술드라마의 여러 장르 중에서 이러한 변화와 혁신을 받아들인 상업적인 공연들 위주로-K시네마, K팝, K방송드라마 등-활성화되어 발전하며 존재감을 발휘해 가고 있다.

4차 산업혁명을 지향하는 산업을 스마트 산업이라 하듯이 공연예술분야에서는 4차 산업혁명의 정보, 기술을 도입하고 융합하는 연극을 'K-스마트드라마'라고 칭할 수 있다. 각종 문화예술의 공연과 관람이 원활하지 않은 상황에서 일명 K-드라마(연극포함) 콘텐츠의 제작은 한국의 글로벌 산업경쟁력에 중요한 수출 잠재력이자 자원인 상황에서, 인공지능 등 정보기술과 예술경영의 융합으로 활성화시켜야 하는 'K-스마트드라마'는 글로벌시장을 향한 한국의 새로운 먹거리로 등장하는 초기단계라 할 수 있다. 실제로 최근 정부는 디지털뉴딜과 그린뉴딜 사업을 수년간 발표, 추진하며 4차 산업혁명을 지원하는 스마트 산업화를 그 핵심으로 꼽았다. 이는 산업분야 별로 플랫폼 환경을 구축하는 것에 주안점을 두고 있기 때문이기도 하다. 그러나 아직 AI인공지능 기술이 스마트 드라마 제작에 활발하게 적용되고 있지는 않다.

이에 본 연구의 목적은 AI인공지능이 가미된 4차산업혁명기술을 수용하고 있는 광의에 드라마를 본 소고에서 K-스마트드라마라 칭하며 이러한 분야를 논하고자 한다.

Ⅱ. 스마트 연극

최근 초지능, 초연결, 초개성의 혁신 기술이 산업 전반적으로 확산되면서 연극분야에서도 실감현실, AI인공지능, 휴머노이드 로봇[1,2] 등을 활용하여 새로운 연극을 제작하려는 시도가 활발해지고 있다. 이러한 기술들은 연극공연 분야의 예술작품을 제작하는데 있어서 무대장치를 통한 효과와 출연하는 배우의 연기 방법, 연출의도에 맞는 일루젼 즉, 상상력의 표현이라든가 과거를 회상하는 것이라든가 등의 표현을 할 때 사용되고 있다[3].

오늘날 '스마트'라는 용어는 여러 영역에서 사용되고 있다. 스마트 연극을 정의하고 개념화하기 위하여 먼저 타영역에서 사용중인 '스마트'의 개념에 대해서 정리할 필요가 있다. 스마트폰, 스마트 제조, 스마트 팜, 스마트 시티, 스마트 헬스케어, 스마트 빌딩, 스마트

자동차, 스마트 교육, 스마트 미디어에 대한 정의를 살펴본 결과 '스마트'에 대한 다음과 같은 특성이 있는 것으로 보인다.

- 디지털 또는 ICT기술을 활용함: CPU, 메모리 등 컴퓨팅의 기본 요소 및 센서, 멀티미디어를 포함
- 디지털 데이터를 기반으로 수요자의 상황을 이해함: 센서 데이터 등을 활용하여 데이터를 기반으로 수요자의 특성이나 요구사항 등을 이해하는 기능을 제고
- 수요자의 경험을 증진함: 편리성(효율화, 자동화 등), 몰입감(흥미 등) 등이 제고
- 수요자 가치를 제고함: 안전, 편의, 행복, 학습 등 수요자의 가치를 제고하는 데 기여

스마트 연극이란 배우 혹은 관람객의 상황을 이해하고 경험을 증진하여 배우 혹은 수요자의 가치를 제고하게 해주는 ICT기반 연극이라고 정의할 수 있을 것이다. 따라서 연극에 비하여 스마트연극은 (작가와 희곡, 연출과 기획 및 제작, 배우와 인물캐릭터, 관객의 관람과 언텍트 선택) 편리성, 효율성, 지능성, 연결성, 순환성 등 공연작품의 제작과 완성도의 과학적인 각분야-석학자문단의 AI인공지능 수치화- 등이 더욱 제고된 연극이다.

스마트연극에는 다음과 같은 예가 있을 것이다.

- 연극, 공연 행위를 이미지 수치화(digitize an image)로 저장하고 추출 분석하도록 하는 것
- iPad나 스마트폰 등 스마트 전자장비를 활용매체로 하여 배우들이 공연을 수행하도록 하는 것
- 관객들과 배우들이 전자장비를 가지고 소통하는 것이며 또는 관객들이 전자장비로써 공연에 참여하는 것
- 전자장비(예: IoT)로 공연장을 컨트롤하거나 관객 각 개인이 자신에게 맞는 무대 연출을 선택한 스토리의 길을 찾아다니며(예: 이머시브연극) 공연 관람의 개인화를 체험하고 즐기는 것이 가능하도록 하는 것
- 몰입형 기술, 즉 VR 장비 등을 활용하여 가상적으로 그리고 에워싸는 듯(immersive)하게 공연을 보게 해주는 감정이입(empathy) 또는 매직이프(Magic if)에 집중되도록 하는 공연환경

Ⅲ. K-스마트드라마를 위한 인공지능 활용방안

K-스마트드라마 실현을 위한 인공지능의 공연에의 적용방안은 다음과 같다. 첫째, 공연 내용을 시각화하고 청각적인 음향과 앙상블을 이룬다. 둘째, 작품내용과의 연관성을 높이는 방법으로 온도, 습도, 후각적 자극을 통제한다. 셋째, 공연 데이터 분석을 통해 예지보전, 즉 관객의 불만족이나 관객수의 감소를 사전에 인지한다. 넷째, 기존의 공연 관련 빅데이터를 클라우드로 관리하여 누구나 관련 정보를 자신이 원하는 형태로 볼 수 있도록 한다. 다섯째, 인공지능 챗봇을 통해 대화를 통해 공연물을 추천하고 공연에 대한 보다 더 깊은 이해를 하도록 안내를 할 수 있다. 여섯째, 음성인식 기술을 통해 관람 중에서도 대사를 봄으로써 공연에 대한 이해를 높일 수 있다. 특히 청각 장애인에게는 대화 내용을 전달함으로써 만족감을 제공한다. 일곱째, GAN 기술을 사용하여 인공지능이 스스로 가상의 인물, 가상의 오브제, 가상의 배경을 제작하게 함으로써 무대장치를 효율적으로 제작한다. 마지막으로, 아카이브 형태로 존재하는 연기 동영상 학습을 통해 인공지능이 연기 동작을 예측하거나 새로운 연기 동작을 제안하게 한다.

이렇게 인공지능 기반 드라마를 위해서는 양질의 학습 데이터가 준비되어야 한다. 이를 위해 연극공연에 대한 아카이빙 작업이 필요하며, KOPIS 등 공개된 연극 DB 를 활용할 수 있다.

ACKNOWLEDGMENT

이 논문 또는 저서는 2020 년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2020S1A3A2A02093277)

참고문헌

[1] 박소영. (2019). 인공지능의 역사: 서사적 허구, 문화 상품, 그리고 과학적 사실로. 인간· 환경· 미래, (22), 91-118.

[2] 김영학. (2018). 로봇연극에 나타난 언캐니 연구-히라타 오리자의<일하는 나>,<사요나라>,<세 자매> 를 중심으로. 드라마 연구 (DR), 54, 5-31.

[3] 홍주영. (2013). 연극에 있어서의 일루젼(Illusion)과 관객의 역할. 연극교육연구, 23, 275-300.

유전 알고리즘을 이용한 경량 인공지능 시스템에서의 하이퍼 파라미터 최적화

김병수, 전석훈, 황태호

한국전자기술연구원

bskim4k@keti.re.kr, seokhun.jeon@keti.re.kr, taeo@keti.re.kr

Hyper Parameter Optimization in Lightweight Artificial Intelligence System Using Genetic Algorithm

Byung-Soo Kim, Seokhun Jeon, Tae-ho Hwang

Korea Electronics Technology Institute

요 약

현재 다양한 경량 인공지능 알고리즘이 개발되고 있으며 이러한 경량 인공지능 시스템의 학습 성능은 하이퍼 파라미터 값에 따라 성능이 크게 변동되거나 저하되는 문제점을 가지고 있다. 따라서 최근 최적의 하이퍼 파라미터를 찾으려는 방법에 대한 관심은 최근 높아지고 있으나 주로 휴리스틱한 방법이나 경험 법칙에 의해 하이퍼 파라미터가 결정되는 경우가 많다. 본 논문에서는 유전 알고리즘을 이용하여 경량 인공지능 시스템에서 사용되는 하이퍼 파라미터의 최적화 방법을 제안하였다. 제안하는 방법은 하이퍼 파라미터에 따라 변동하는 경량 인공지능 시스템의 성능 및 특성을 분석하여 고정되거나 제한된 범위 내에서 하이퍼 파라미터를 선정할 수 있어 제한적인 리소스 상에서의 높은 정확도와 효율적인 하드웨어 설계를 가능하게 한다. 제안한 방법은 MATLAB을 이용하여 시뮬레이션을 수행 각 응용 어플리케이션에 따라 하이퍼 파라미터를 최적화하여 경량 인공지능 시스템의 성능을 검증하였다.

Ⅰ. 서 론

최근 딥러닝(Deep Learning)기반 인공지능 기술이 다양한 분야에서 혁신적인 변화를 일으키고 있으나, 대부분 고사양의 서버 시스템, 대용량 스토리지, 클라우드에 의존적이다[1]. 이러한 한계를 극복하기 위해 경량 디바이스, 모바일 단말 등 엣지 디바이스에서 직접 학습과 추론이 가능한 경량 인공지능(Lightweight Artificial Intelligence)에 대한 연구가 활발히 이루어지고 있다[2,3]. 경량 인공지능은 사양이 낮고 저전력을 요구하는 임베디드 시스템에서 운용가능하며, 자체적으로 학습기능을 보유하고 있어, 응용 어플리케이션에 따라 특화된 학습 및 분류가 수행되어야 한다. 다양한 경량 인공지능 모델들이 개발되고 있지만, 최적의 성능을 낼 수 있는 하이퍼 파라미터(Hyper Parameter) 최적값에 대한 연구는 부족한 상황이며 주로 휴리스틱한 방법이나 경험 법칙에 의해서 하이퍼 파라미터의 최적값을 결정하고 있는 상황이다.

본 논문에서는 경량 인공지능 시스템에서 정확도를 높이기 위해 특정 개체가 아닌 전체적인 군집으로 탐색을 수행하여 전역 최소화를 구할 수 있는 특징을 가진 유전 알고리즘(Genetic Algorithm)을 사용하여 최적의 하이퍼 파라미터를 찾아내고, 이를 적용한 경량 인공지능 시스템이 제한적인 리소스 상에서 높은 정확도를 갖으며 효율적인 하드웨어 설계를 가능하게 하는 소프트웨어 모델을 제안한다.

본 논문은 2절에서 제안하는 유전 알고리즘을 이용한 경량 인공지능 시스템에서의 하이퍼 파라미터 최적화 방법을 설명하고, 3절은 다양한 응용에서의 실험결과를 통해 제안한 방법의 소프트웨어 모델 동작 검증결과를 설명하고, 4장에서는 본 연구의 결론과 함께 추가 연구 방향에 대해 소개하고자 한다..

Ⅱ. 제안하는 유전 알고리즘을 이용한 하이퍼 파라미터 최적화

유전 알고리즘은 1970년대 John Holland에 의해 개발된 자연 세계의 진화 현상에 기반 한 일종의 진화 연산 알고리즘이다[4]. 특정 개체가 아닌 제천적인 군집으로 탐색을 수행하여 복잡한 제약성을 가진 대규모의 최적화 문제들에 뛰어난 성능을 가지며, 국소 최소화가 아닌 전역 최소화를 구할 수 있는 특징을 가지고 있어 경량 인공지능 시스템의 하이퍼 파라미터 최적화에 적합한 방법이다.

본 논문에서 제안하는 유전 알고리즘을 이용한 경량 인공지능 시스템의 하이퍼 파라미터 최적화 방법은 아래의 표시된 그림 1과 같다.

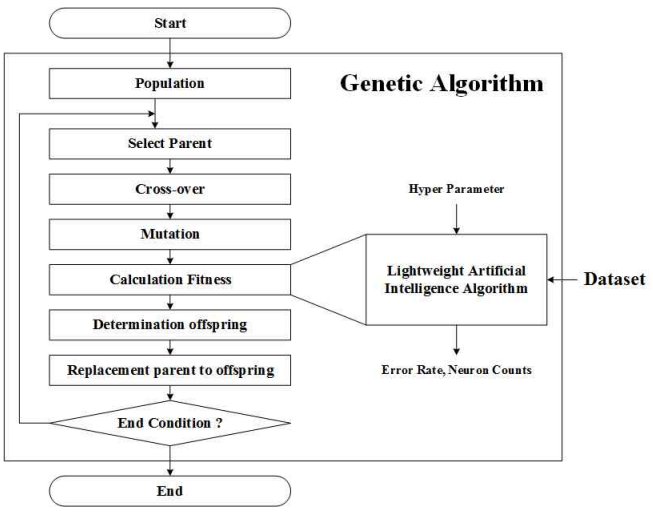


그림 1 제안하는 방법

먼저 초기 개체집단(Population)을 랜덤하게 생성하고, 각 개체에 대한 적합도 함수 값을 평가하고 그 값에 따라 정렬하면 초기 개체 집단 생성이 완료된다. 이후 전체 개체 집단 중 우수한 개체를 발굴하기 위한 Selection 과정을 수행하고, 선택되어진 개체를 조합하여 새로운 개체를 생성하는 Crossover 연산 과정을 수행한다. 이때, 유전 알고리즘에서 일반적으로 Single Point Crossover 연산을 사용한다. Crossover 연산 이후, 군집이 가지는 해공간의 제한성을 극복하고 보다 다양한 탐색 후보들을 얻기 위한 Mutation 연산을 수행한다. Mutation 연산 까지 수행한 개체들에 대해 경량 인공지능 알고리즘을 통해 적합도 함수 값을 계산하고, 이 개체의 적합도 함수 값이 처음 선택 되어진 개체보다 낮은 경우에 개체집단에서 선택된 개체(부모세대)를 최종 연산을 수행한 개체(자식세대)로 대체시킨다. 이러한 과정을 종료 조건이 만족될 때까지 계속 수행하여 최적의 하이퍼 파라미터를 탐색하게 된다. 제안하는 유전 알고리즘의 파라미터로는 반복 회수, 개체 집단의 크기, Selection 전략, Crossover 방법 및 확률, Mutation 확률 등이 있으며 제안하는 방법에서 사용한 파라미터는 아래의 표 1과 같다.

표 1 유전 알고리즘 파라미터

Iteration	100
Population	32
Selection 전략	Random
Crossover Type	1-Point Crossover
Crossover Probability	100%
Mutation Probability	5%

경량 인공지능 알고리즘은 그림 2와 같이 RCE 신경망(Restricted Coulomb Energy Neural Network)기반으로 구성되어 거리 기반으로 학습 및 인지 과정을 수행한다. RCE 신경망 내의 뉴런들은 입력 데이터와 뉴런 중심점 사이의 거리를 계산하고 반지름과 비교하여 입력 데이터가 해당 뉴런에 포함되는지를 판단한다. 이 때 뉴런의 최대 반지름 값과, 최소 반지름 값은 인공지능 알고리즘의 성능과 밀접하게 연관된 하이퍼 파라미터로 최적값을 찾기 위해 데이터 셋의 에러율과 활성화된 뉴런 유닛의 개수를 유전 알고리즘의 적합도 함수 값으로 사용한다.

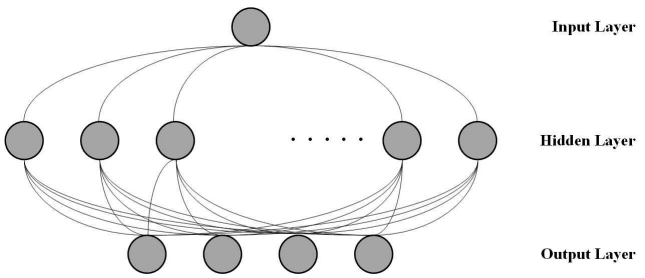


그림 2 RCE Neural Network architecture

Ⅲ. 시뮬레이션 결과

제안하는 유전 알고리즘을 이용한 경량 인공지능 시스템의 하이퍼 파라미터 최적화 방법의 검증을 위해 다양한 테스트 셋에 대한 시뮬에이션을 수행하였다. MATLAB 시뮬레이터를 통해 유전 알고리즘과, RCE 신경망을 모델링하였으며 UCI 데이터 셋을 사용하였다[5]. 2장에서 제안하는 하이퍼 파라미터 최적화 방법을 사용하여 각각의 데이터 셋에 대한 하이퍼 파라미터를 결정하고 정확도를 확인 하였으며 그 결과는 표 2와 같다.

표 2 Dataset에 따른 실험 결과

Dataset	Accuracy	Neuron Count
IRIS	100%	52
Parkinsons	100%	124
Wine	100%	122

Ⅳ. 결론

본 논문에서는 유전 알고리즘을 이용한 경량 인공지능 시스템의 하이퍼 파라미터 최적화 방법을 제안하였다. 제안하는 방법은 유전 알고리즘의 적합도 계산 함수를 경량 인공지능 알고리즘으로 대체하여 데이터 셋에 대한 경량 인공지능 알고리즘의 에러율과 활성화된 뉴런 유닛을 최적화 시킬 수 있는 하이퍼 파라미터 값을 결정하는 것이다.. UCI 데이터 셋을 통한 검증으로 유전 알고리즘을 이용한 경량 인공지능 알고리즘의 하이퍼 파라미터 최적화 가능성을 확인하였다. 추가적인 연구를 통해 제안하는 방법을 하드웨어로 구현하여 FPGA 플랫폼 또는 칩으로 제작하여 검증을 수행할 예정이며, Grid search, Random search 등의 다른 하이퍼 파라미터 최적화 알고리즘과의 비교 연구를 진행할 계획이다.

ACKNOWLEDGMENT

이 연구는 2020년도 산업통상자원부 및 산업기술평가관리원(KEIT) 연구비 지원에 의한 연구임('20009972').

참 고 문 헌

[1] Chen, Jiasi, and Xukan Ran. "Deep learning with edge computing: A review." Proceedings of the IEEE 107.8 (2019): 1655-1674.

[2] B. Kim, J. Lee, T. Hwang, and D. Kim, "Design of Lightweight Artificial Intelligence System for Multimodal Signal Processing," J. of the Korea Institute of Electronic Communication Sciences, vol. 13, no. 5, 2018, pp. 1037-1042.

[3] J. Cho, Y. Jung, S. Lee, and Y. Jung, "Vlsi implementation of restricted coulomb energy neural network with improved learning scheme." Electronics 8.5 (2019): 563.

[4] R. L. Haupt and S. E. Haupt, "Practical Genetic Algorithms," John Wiley & Sons, New York, 2004.

[5] Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

Single Memory를 활용한 뉴럴 네트워크 프로세서용 효율적 Processing Element Array 구조 제안

이재학, 김병수, 송보배, 황태호
한국전자기술연구원
jhk507@keti.re.kr, bskim4k@keti.re.kr, 3232semi@keti.re.kr, taeo@keti.re.kr

Efficient Processing Element Array for Neural Network Processor using Single Memory

Jeahack Lee, Byung-Soo Kim, BoBae Song, Tae-ho Hwang
Korea Electronics Technology Institute

요 약

본 논문은 단일 메모리를 활용하여 효율적인 구조를 가지는 PEA(Processing Element Array)를 제안하였다. 제안하는 PEA는 복수의 메모리 대신 단일 메모리를 사용하여 SoC 설계 시 placement 측면에서 면적을 줄일 수 있을 것으로 기대된다. 또한, latency를 줄여 제안하는 PEA를 활용한 NNP (Neaueal Network Processor)의 idle 시간을 줄여 효율을 높일 수 있다. 다양한 크기의 filter size를 지원 할 수 있도록 재구성 가능하도록 PEA를 설계하다. 시뮬레이션을 통해 동작 검증을 수행하였으며 matlab과 연동하여 결과를 비교하였다.

Ⅰ. 서 론

ImageNet 프로젝트에서 AlexNet [1], VGG [2]등 CNN(Convolutional Neural Network) 기반의 딥러닝이 뛰어난 성능을 보이며 학계에 뉴럴 네트워크에 관한 관심도가 높아졌다. 이후, 알파고[3]의 등장으로 전세계적으로 딥러닝에 대한 뜨거운 관심이 쏟아졌고, 다양한 연구 분야에서 뉴럴 네트워크를 적용하여 기존 알고리즘과 비교하여 향상된 성능에 대한 연구 결과들이 나오기 시작하였다.

딥러닝의 뛰어난 성능은 높은 연산량을 요구하여 실시간 처리에 부적합한 문제가 있다. 이러한 문제를 해결하기 위해 뉴럴 네트워크 연산을 가속할 수 있는 CUDA (Compute Unified Device Architecture) [4] 등 다양한 GPU 기반의 프레임워크가 연구되었다. 하지만, GPU 기반의 딥러닝 시스템은 많은 전력소모를 필요로 하기 때문에 서버에 적합하고 엣지 디바이스에 적용이 불가능하다. 저전력 구현을 위해 NNP (Neural Network Processor)에 대한 다양한 연구[5-7]가 수행되고 있다. 기존의 연구들은 메모리를 연산단위별로 나누어 사용하기 때문에 placement 측면에서 큰 면적을 사용하게 된다. 본 연구는 단일 메모리를 사용하면서 PE(Processing Element)의 효율을 높이는 구조를 제안하였다.

본 논문은 2장에서 제안하는 PEA(Processing Element Array)를 설명한다. 3장은 간단한 예시와 함께 시뮬레이션 결과를 통해 동작 검증을 수행하였다. 4장에서 본 연구의 결론과 함께 추가 연구 방향에 대해 논의하였다.

Ⅱ. 제안하는 NNP의 PEA

본 논문에서는 연속적인 convolution layer와 fully-connected layer 연산에 최적화된 PEA 구조를 제안한다. Convolution layer 연산은 수식 (1)과 같다.

$$y = \sum_i w_i x_i + b = \sum_i p_i + b, \quad (1)$$

y는 convolution layer 연산의 결과를 의미하고, w_i는 i번째 가중치(weight), x_i는 i번째 입력(ifmap), b는 편향(bias), p_i는 i번째 부분합(partial sum)으로 가중치와 입력의 곱을 나타낸다. Convolution layer의 filter size는 1x1, 3x3, 7x7 등 다양한 크기를 가질 수 있어 PEA를 재구성 가능하도록 설계가 되었다. Fully-connected layer는 convolution layer에서 filter size가 1x1의 연산과 수식적으로 동일하여 1x1 filter를 사용하는 연산으로 대체 가능하다. 따라서, 본 논문에서는 convolution layer 연산에 대해서만 다루었다.

Convolution layer 연산을 하드웨어 관점에서 보면 입력(x)을 SRAM에 저장 후 PEA에 입력하여 연산 결과를 얻게 된다. 행렬 연산을 위해 동시에 입력을 위해서는 filter size 만큼의 register를 사용하기 때문에 하드웨어 구조적으로 효율적이지 않고, 순차적으로 입력을 하여 하드웨어 효율적인 구조의 경우 단순한 연산기 배치로 convolution layer 연산이 불가능하다. 본 논문은 그림 1과 같이 PE를 chain 형식으로 연결하여 SRAM의 순차적인 입력에 대해 효율적 처리가 가능한 PEA 구조를 제안한다.

PEA는 N개의 PE와 하나의 adder tree로 구성된다. PE는 filter size에 따라 구성이 되어 다양한 filter size에 대해 처리가 가능하다. 예를 들어 3x3 filter size의 경우 9개씩 PE가 묶음으로 convolution layer 연산이 수행되며, $\lfloor N/9 \rfloor$ 개의 filter 연산이 동시에 수행 가능하다. PE는 그림 2(a)와 같이 입력(x_i)과 해당 가중치(w_i)의 곱연산을 통해 가중치를 계산

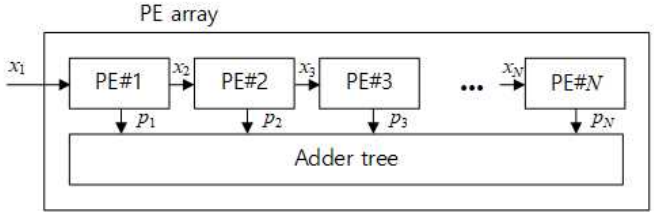


그림 1. 제안하는 PE array 구조.

하며, 다음 PE에 지연된 입력(x_{i-1})을 전달한다. Adder tree는 그림 2(b)와 같이 PE에서 연산한 부분합을 filter에 맞게 누적하는 연산을 수행한다. 새로운 filter 연산이 수행되는 경우 ps_i를 0으로 설정하여 누적 연산을 수행하고, 아닌 경우 이전에서 누적한 ps_i와 PE에서 연산한 p_i를 누적하여 다음 연산에 사용한다. 이와 같이 제안하는 PEA는 SRAM에서부터 출력되는 연속적인 입력에 대해 처리가 가능하여 latency를 줄일 수 있으며, 하드웨어 구현에도 적합한 구조이다.

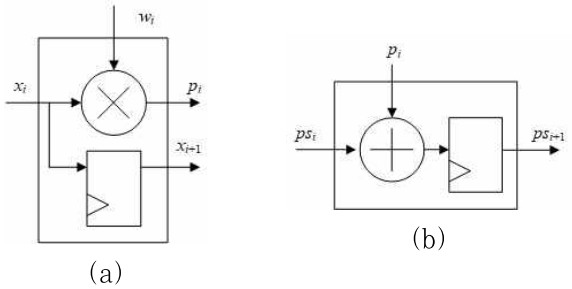


그림 2. (a) PE의 상세 구조. (b) Adder tree의 상세 구조.

Ⅲ. 시뮬레이션 결과

Verilog 언어를 사용하여 PEA를 설계하였으며 Cadence社의 irun을 사용하여 시뮬레이션을 수행하였다. 3x3 filter size에 대해 연산을 수행하였고, 가중치는 그림 3(a)와 같다. 입력은 그림 3(b)와 같으며 출력은 5bit fraction으로 shift 연산하여 그림 3(c)와 같다.

1	2	3
4	5	6
7	8	9

(a)

0	1	2	3	4
10	11	12	13	14
20	21	22	23	24
30	31	32	33	34
40	41	42	43	44

(b)

21	23	24
35	37	38
49	51	52

(c)

그림 3. 시뮬레이션 입력 (a) 가중치, (b) 입력, (c) 출력

설계한 PEA를 그림 3의 입력과 같이 수행한 결과 그림 4와 같이 동일한 출력 결과가 나오는 것을 확인 할 수 있다. 동일한 코드를 사용하여 1x1, 5x5, 7x7 filter size에 대해서도 시뮬레이션을 수행하였고 예측한 결과와 같은 출력을 얻어 동작을 확인하였다.

Ⅳ. 결론

본 논문에서는 하드웨어에 적합한 PEA 구조를 제안하였다. 제안하는 구조는 연속적인 입력에 대해 처리가 가능하며, 다양한 filter size를 지원하도록 재구성이 가능하도록 설계 되어 있다. 1x1, 3x3, 5x5, 7x7 filter size에 대해 시뮬레이션 검증을 수행하였으며 예측한 결과와 같은 출력을 얻을 수 있었다. 추가적인 연구를 통해 제안하는 PEA 구조를 사용하는 neural network processor 설계가 필요하며, FPGA 플랫폼 또는 SoC 설계를 통해 검증이 필요하다.

ACKNOWLEDGMENT

이 연구는 2020년도 산업통상자원부 및 산업기술평가관리원(KEIT) 연구비 지원에 의한 연구임('20009972').

참 고 문 헌

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Proc. Adv. Neural Inf. Process. Syst., 2012, pp. 1097-1105.

[2] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, arXiv:1409.1556. (https://arxiv.org/abs/1409.1556).

[3] https://deepmind.com/

[4] D. K. Panda, K. Tomko, K. Schulz, and A. Majumdar, "The MVAPICH Project: Evolution and Sustainability of an Open Source Production Quality MPI library for HPC," in WSSPE, 2013..

[5] D. Shin, J. Lee, J. Lee, J. Lee and H. Yoo, "DNPU: An Energy-Efficient Deep-Learning Processor with Heterogeneous Multi-Core Architecture," in IEEE Micro, vol. 38, no. 5, pp. 85-93, Oct. 2018.

[6] D. Shin, J. Lee, J. Lee and H. Yoo, "14.2 DNPU: An 8.1TOPS/W reconfigurable CNN-RNN processor for general-purpose deep neural networks," 2017 IEEE International Solid-State Circuits Conference (ISSCC), San Francisco, CA, 2017, pp. 240-241.

[7] J. Lee, C. Kim, S. Kang, D. Shin, S. Kim and H. Yoo, "UNPU: An Energy-Efficient Deep Neural Network Accelerator With Fully Variable Weight Bit Precision," in IEEE Journal of Solid-State Circuits.

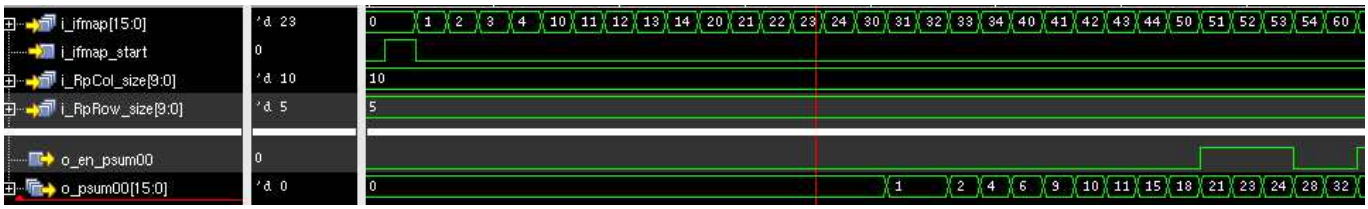


그림 4. 시뮬레이션 결과

Meijer G-함수를 활용한 딥러닝 기반 설명 가능한 패스로스 모델

이현석
세종대학교
hyunsuk@sejong.ac.kr

Explainable Deep Learning-Based Pathloss Model Using Meijer G-Function

Hyun-Suk Lee
Sejong University

요 약

딥러닝 기법을 활용하면 무선 측정 데이터로부터 효율적으로 정확한 심층신경망 패스로스 모델을 학습시킬 수 있으나, 심층신경망의 불투명성으로 인해 학습된 패스로스 모델의 특성을 분석하고 이해하기 어려운 문제가 있다. 본 논문에서는 학습된 심층신경망 패스로스 모델을 Meijer G-함수를 기반으로 한 수학적 표현으로 근사하여 설명 가능한 패스로스 모델을 구성하는 기법을 개발하고, 모의실험을 통해 제안 기법을 통해 찾은 모델이 심층신경망 패스로스 모델을 정확하게 근사하는 것을 보인다.

Ⅰ. 서 론

최근 통신 시나리오는 mmWave, 대규모 안테나 기술 등의 등장으로 인해 점점 더 복잡해지고 있다 [1]. 기존의 패스로스 모델링 방식은 무선 측정 데이터의 포스트 프로세싱을 통해, 패스로스의 특성을 조사하고, 이를 기반으로 패스로스 모델을 구성하는 형태로 진행되었다. 하지만 최근에 등장하는 복잡한 시나리오에서는 데이터의 복잡성과 다양성으로 인해 포스트 프로세싱이 어려워지고 높은 비용이 요구되므로 기존의 패스로스 모델링 방식이 사용되기 어려운 문제가 있다 [2]. 이를 해결하기 위해 최근 딥러닝 기술의 비약적인 발전과 함께 데이터로부터 패스로스 모델을 직접적으로 학습하여 전통적인 데이터의 포스트 프로세싱을 수행하지 않고도 정확한 패스로스 모델을 구하는 딥러닝 기법 기반의 효율적인 데이터-드리븐 패스로스 모델링이 연구되고 있다 [3].

이 같은 딥러닝 기반의 패스로스 모델링은 일반적으로 심층 신경망 (Deep Neural Network, DNN)을 이용한다. DNN을 이용한 패스로스 모델은 DNN이 갖는 높은 학습용량을 바탕으로 복잡한 패스로스 환경을 효과적으로 학습할 수 있게 한다. 하지만 DNN 모델의 경우 학습된 모델 내부의 동작 원리를 설명하기 어려운 한계가 있다 [4]. 이 같은 DNN 모델의 불투명성으로 인해 DNN 모델은 블랙박스 모델로 불린다. 따라서 DNN을 이용하여 학습된 채널 패스로스 모델의 경우 물리적인 송수신단 간 거리, 반송파 주파수 등과 같은 주어진 물리량으로부터 패스로스 출력값은 계산할 수 있으나, 주어진 물리량으로부터 패스로스가 어떻게 정해지는지 내부적인 원리를 설명할 수 없기 때문에 물리량과 패스로스 사이의 관계, 중요도와 같은 패스로스 모델 특성을 분석하고 이해하기 어렵다. 그러므로 이 같은 문제를 해결하기 위하여 DNN 패스로스 모델을 학습한 뒤, 더 나아가 사람이 이해할 수 있고 분석 가능한 형태로 표현할 필요가 있다.

본 논문에서는 위의 문제를 해결하기 위해 Meijer G-함수를 활용하여 DNN 패스로스 모델을 근사하는 이해 가능하고 분석 가능한 수학적 표현을 갖는 설명 가능한 패스로스 모델을 구하는 방법을 제안하고, 실험을 통해 Meijer G-함수 기반 설명 가능한 패스로스 모델이 학습된 DNN 패스로스 모델을 정확히 근사함을 보인다.

Ⅱ. 딥러닝 기반 패스로스 모델

본 논문에서는 딥러닝 기반 채널 패스로스 모델을 고려한다. 이 같은 패스로스 모델을 위해서는 순방향 신경망 (Feedforward Neural Network) 혹은 컨볼루셔널 신경망 (Convolutional Neural Network)과 같은 심층신경망 (DNN)이 사용되며, 신경망의 학습을 위해 무선 측정 데이터 혹은 시뮬레이션을 이용한 데이터셋이 활용된다. 일반적으로 채널 패스로스 모델은 송수신단 3차원 좌표, 송수신단 간 거리, 반송파 주파수 등의 물리적인 입력값들을 기반으로 송수신단 간 패스로스를 출력값으로 예측한다. 따라서 딥러닝 기반 패스로스 모델에서는 입력 feature로 갖고 송수신단 간 패스로스를 출력값으로 갖는 DNN을 구성하고 데이터를 이용하여 학습시켜 패스로스 모델을 구현한다 [2].

딥러닝 기반 패스로스 모델을 구성하는 DNN의 입력 feature와 출력값을 각각 $\mathbf{x}=(x_1, x_2, ..., x_K)^\top \in X$ 및 $y \in Y$라 정의한다. 여기서 K는 입력 feature의 개수를 의미하고 X 및 Y는 각각 입력 공간, 출력 공간을 의미하며, 일반성을 잃지 않고 입력 공간이 $X=[0,1]^K$와 같음을 가정할 수 있다. 위의 정의를 기반으로 물리적인 실제 무선 환경에서 입력 feature와 채널 패스로스와의 관계를 함수 $H: X \rightarrow Y$로 정의한다. 딥러닝 기반 패스로스 모델은 DNN $h: X \rightarrow Y$가 앞서 정의한 함수 $H(\mathbf{x})$를 근사하도록 (즉, $h(\mathbf{x}) \approx H(\mathbf{x}), \ \forall \mathbf{x} \in X$) 기계학습 기법을 이용하여 DNN을 학습시킨다.

Ⅲ. Meijer G-함수를 활용한 설명 가능한 패스로스 모델

딥러닝 기반 패스로스 모델으로부터는 DNN의 불투명성으로 인해 각 물리적 입력값들이 패스로스에 미치는 영향, 특성, 중요도 등을 분석하기 어려운 문제가 있다. 이를 해결하기 위해 DNN 모델의 불투명성을 해결할 수 있는 설명 가능한 형태의 패스로스 모델을 구성할 필요가 있고, 본 논문에서는 Meijer G-함수를 이용하여 DNN 패스로스 모델 $h(\mathbf{x})$를 이해 가능하고 분석 가능한 수학적 표현으로 나타내는 방법을 연구하였다.

이를 위해, 먼저 이해 가능하고 분석 가능한 수학적 표현들을 포함하는 class M을 정의하고, 설명 가능한 패스로스 모델을 구성하기 위해서

DNN 패스로스 모델 $h(\mathbf{x})$를 가장 정확히 근사하는 함수 $m \in M$을 찾을 필요가 있다. 함수 $m(\mathbf{x})$의 DNN 패스로스 모델 $h(\mathbf{x})$에 대한 근사 오차를 Mean Squared Error (MSE)과 입력 feature $\mathbf{x}$의 분포 $F(\mathbf{x})$를 이용하여 수학적으로 나타내면 아래 수식과 같다.

$$l(m,h) = \|m-h\|_2^2 = \int_X (m(\mathbf{x})-h(\mathbf{x}))^2 dF(\mathbf{x}) \quad (1)$$

앞서 정의한 근사오차를 이용하여 설명 가능한 패스로스 모델 $m(\boldsymbol{x})$를 찾는 최적화 문제를 아래와 같이 정의할 수 있다.

$$\arg\min_{m \in M} l(m,h) \quad (2)$$

Meijer G-함수는 파라미터 $a_i, b_j \in \mathbb{R},\ \forall i = 1,\ldots,p, j = 1,\ldots,q$ 및 $m,n,p,q \in \mathbb{N}$으로 특정되는 함수로써, 정의는 아래와 같다.

$$G_{p,q}^{m,n}\left(\begin{matrix} a_1,\ldots,a_p \\ b_1,\ldots,b_q \end{matrix} \middle| x \right) = \frac{1}{2\pi i} \int_{\mathcal{L}} \frac{\prod_{j=1}^{m} \Gamma(b_j - s) \prod_{j=1}^{n} \Gamma(1-a_j+s)}{\prod_{j=m+1}^{q} \Gamma(1-b_j+s) \prod_{j=n+1}^{p} \Gamma(a_j-s)} x^s ds$$

위의 정의에서 $\Gamma(\cdot)$은 Gamma 함수를 의미한다. Meijer G-함수는 주어진 파라미터에 따라 지수함수, 삼각함수, 초기하함수 등 알려진 많은 함수를 표현할 수 있다. 아래에서 Meijer G-함수를 이용하여 설명 가능한 패스로스 모델 $m(\mathbf{x})$를 파라미터 $\boldsymbol{\theta}$로 표현되는 함수 $M_{\boldsymbol{\theta}}(\mathbf{x})$로 정의한다. Kolmogorov superposition 정리에 따르면 어떤 다변수 함수도 표현할 수 있고, 이를 활용하여 다변수 함수인 패스로스 모델을 단변수 함수인 Meijer G-함수로 아래와 같이 표현한다.

$$M_{\boldsymbol{\theta}}(\mathbf{x}) = \sum_{i=0}^{r} G_{p,q}^{m,n}\left(\boldsymbol{\theta}_i^{\text{out}} \middle| \sum_{j=1}^{K} G_{p,q}^{m,n}(\boldsymbol{\theta}_{ij}^{\text{in}} | x_j) \right)$$

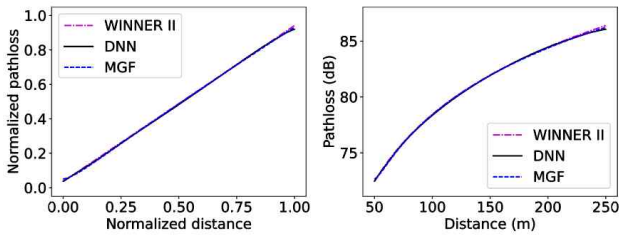
위의 식에서 파라미터 $\boldsymbol{\theta} = (\boldsymbol{\theta}^{\text{out}}, \boldsymbol{\theta}^{\text{in}})$, $\boldsymbol{\theta}^{\text{out}} = (\boldsymbol{\theta}_0^{\text{out}}, \ldots \boldsymbol{\theta}_r^{\text{out}})$, 그리고 $\boldsymbol{\theta}^{\text{in}} = \{(\boldsymbol{\theta}_{i1}^{\text{in}}, \ldots \boldsymbol{\theta}_{iK}^{\text{in}})\}_{\forall i \in \{0,\ldots,r\}}$를 의미하고 각 파라미터 $\boldsymbol{\theta}$들은 $a_1,\ldots,a_p$ 및 $b_1,\ldots,b_q$로 구성된다. 마지막으로 (m,n,p,q,r)은 $M_{\boldsymbol{\theta}}(\mathbf{x})$를 표현하는 초파라미터이다. 따라서 수식 (2)의 문제는 DNN 패스로스 모델을 근사하는 $M_{\boldsymbol{\theta}}$를 찾는 파라미터 최적화 문제로 변형된다. 파라미터 $\boldsymbol{\theta}$의 공간이 Θ 일 때,Meijer G-함수 기반 수학적 표현 class는 $M = \{M_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$로 정의되고, 주어진 공간 Θ에 따라 polynomial, algebraic, closed-form 등 폭넓은 수학적 표현 class를 표현한다. 따라서 분석 목적에 따라 적절한 파라미터 공간을 선택할 필요가 있다 [5]. Meijer G-함수 기반 설명 가능한 패스로스 모델을 구하기 위해 $M_{\boldsymbol{\theta}}$를 이용하여 수식 (2)의 문제를 아래와 같은 파라미터 최적화 문제로 다시 정의한다.

$$\arg\min_{\boldsymbol{\theta} \in \Theta} l(M_{\boldsymbol{\theta}}, h) \quad (3)$$

위 문제를 해결하기 위해서 근사오차를 최소화하는 최적 파라미터 $\boldsymbol{\theta}^*$를 구해야 한다. Meijer G-함수의 또 다른 장점은 파라미터에 대한 그래디언트를 수치적으로 구할 수 있다는 점이고, 이를 이용하면 일반적인 그래디언트 최적화 기법들을 활용하여 수식 (3)의 문제의 최적 파라미터를 효율적으로 구할 수 있다는 것이 알려져 있다 [5]. 구체적으로, 임의의 입력 feature를 생성하고 생성된 입력값들을 기반으로 경사하강법 (gradient descent algorithm)을 이용하여 근사오차를 최소화하는 파라미터 $\boldsymbol{\theta}^*$를 구하여 Meijer G-함수 기반 설명 가능한 패스로스 모델 $M_{\boldsymbol{\theta}}$를 얻는다. 패스로스 모델 $M_{\boldsymbol{\theta}}$는 수학적 표현으로 Taylor 급수 등을 이용하여 feature 중요도 및 상관도 등을 쉽게 분석할 수 있는 장점이 있다 [5].

Ⅳ. 모의실험

모의실험을 통해 Meijer G-함수 기반의 설명 가능한 패스로스 모델 (MGF)이 딥러닝 기법을 통해 학습된 DNN 패스로스 모델을 효과적으로 근사하는지 살펴본다. 이를 위해 임의의 거리를 생성하고 WINNER II 패스로스 모델에서 거리를 임의로 생성된 거리에 따른 패스로스에 평균 0 표준편차 0.1을 갖는 가우시안 잡음을 추가하여 1000개의 샘플을 갖는 패스로스 데이터셋을 생성하였고, 해당 데이터셋을 이용하여 100개의 unit으로 이루어진 2개의 hidden layer들로 구성된 DNN 패스로스 모델을 학습시켰다. 이후 학습된 DNN 패스로스 모델을 Meijer G-함수를 이용하여 설명 가능한 패스로스 모델로 근사하였다.



(a) 정규화된 패스로스 (b) 복원된 패스로스

그림 1. 패스로스 모델들의 비교

그림 1은 원래의 패스로스 모델 WINNER II와, DNN 패스로스 모델, Meijer G-함수 기반 패스로스 모델을 보여준다. WINNER II 모델로부터 생성된 데이터셋을 정규화시켜 DNN을 학습시켰으므로 그림 1(a)에서는 정규화된 패스로스 모델을 보여주고, 그림1(b)에서는 복원된 패스로스 모델을 보여준다. 두 그림 모두 DNN 모델이 효과적으로 WINNER II 모델을 학습한 것을 보여주며, Meijer G-함수 기반의 설명 가능한 모델 또한 DNN 모델을 효과적으로 근사하고 있는 것을 볼 수 있다. 위 결과를 통해 제안 기법이 DNN 패스로스 모델을 정확히 근사하는 Meijer G-함수를 기반으로하는 설명 가능한 패스로스 모델을 찾는다는 것을 알 수 있다.

Ⅴ. 결론

본 논문에서는 학습된 DNN 패스로스 모델을 Meijer G-함수를 기반으로 한 수학적 표현으로 근사하여 설명 가능한 패스로스 모델을 구성하는 기법을 개발했고, 모의실험을 통해 제안 기법을 통해 찾은 설명 가능한 모델이 DNN 패스로스 모델을 정확하게 근사하는 것을 보였다.

참 고 문 헌

[1] W. H. Chin, Z. Fan, and R. Haines, “Emerging technologies and research challenges for 5G wireless networks,” IEEE Wireless Commun., vol. 21, no. 2, pp. 106-112, Apr. 2014.

[2] J. Huang, C. Wang, L. Bai, J. Sun, Y. Yang, J. Li, O. Tirkkonen, and M. Zhou, “A big data enabled channel model for 5G wireless communication systems,” IEEE Trans. Big Data, vol. 6, no. 2, pp. 211-222, June 2020.

[3] J. Thrane, D. Zibar, and H. L. Christiansen, “Model-aided deep learning method for path loss prediction in mobile communication systems at 2.6 GHz,” IEEE Access, vol. 8, pp. 7925 - 7936, 2020.

[4] W. Guo, “Explainable artificial intelligence for 6G: Improving trust between human and machine,” IEEE Commun. Mag., vol. 58, no. 6, pp. 39-45, June 2020.

[5] A. M. Alaa and M. van der Schaar, “Demystifying black-box models with symbolic metamodels,” in Advances in Neural Information Processing Systems, 2019.

자연계 최적법칙에 기반한 마이크로파 필터 설계법

이창형, 조정현, 서예준, 전문수, 이경민, 강승택

인천대학교

S－kahng@inu.ac.kr

Nature－Inspired Optimization for Microwave Filter Designs

Changhyeong Lee, Junghyun Cho, Yejune Seo, Munsu Jeon, Gyungmin Lee, Sungtek Kahng

Incheon National University

Abstract

In this paper we present a stochastic optimization process based on genetic algorithm for filter synthesis. An L－band fourth order filter is designed and tested using the proposed algorithm.

Ⅰ. Introduction

Microwave filters play an important role in the wireless communication system. Chebyshev and Elliptic class of filtering function have found frequent application within the space of microwave communication system. The generic features of amplitude in band characteristics together with the sharp cutoff skirts gives an acceptable compromise between lowest signal degradation and highest interference rejection [1].

A filter can be design using transfer function by giving number of poles and zeros. The order of the filter depends on the number of poles. The number of zeros in the in the stop band region plays an important role in the performance of the filter. A filter can be synthesized by different methods. One of them which is being followed here is Genetic Algorithm (G.A.) optimization under the framework of Matlab. We have proposed the idea of synthesizing of filter by GA to obtain the location of poles and zeros.

Ⅱ. Proposed Synthesis Method

To verify the proposed filter synthesis method, it is applied to a symmetric 4th order L－band filter having four poles in the passband and two zeros in the lower and upper stop band regions.

$$s_{21}(s) = \frac{P(s)}{\varepsilon E(S)}$$

Where ε is the ripple factor and 's' is the complex frequency variable, E(s) is the polynomial of the poles and P(s) is the polynomial of the zeros.

$$cost\ function = \sum_{n_1=1}^{N_1} \left|S_{21spec.}(f_{n1}) - S_{21tried}(f_{n1})\right|^2 + \sum_{n_2=1}^{N_2} \left|S_{21spec.}(f_{n2}) - S_{21tried}(f_{n2})\right|^2$$

Figure 1 shows the normalized and real frequency response for the fourth order L－band filter over the frequency range of 1.67 GHz to 1.72 GHz. The return loss achieved is below －20 dB. The right end of Figure 1 shows the cost function error which is 5.4960 after 80 alterations.

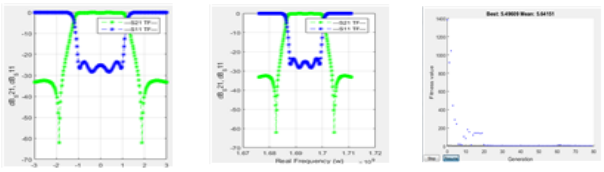


Figure 1. Frequency responses and cost function error

Ⅲ. Conclusion

In this paper a simple and efficient method for the synthesis of microwave filters has been presented. The synthesis procedure includes transfer function for the optimization of filters using genetic algorithm in the framework of Matlab.

ACKNOWLEDGMENT

This work was supported by the national project of Korean aerospace research institute.

References

[1] Richard J. Cameron, "General Coupling Matrix Synthesis Methods for Chebyshev Filtering Functions", IEEE Transaction on Microwave. Theory and Tech., Vol. 47, No. 4, April, 1999.

[2] Sungtek Kahng et al, "A Dual－Mode Narrow－Band Channel Filter and Group－Delay Equalizer for a Ka－Band Satellite Transponder", ETRI Journal, vol. 25, no. 5, Oct. 2003, pp.379－386.

딥 러닝에서의 딥 뉴럴 네트워크의 가중치 초기화 방법

홍정하, 여도엽
한국전자통신연구원

jhong@etri.re.kr, yeody@etri.re.kr

Weight initialization method for deep neural network in deep learning

Jungha Hong, Doyeob Yeo
Electronics and Telecommunications Research Institute

요 약

본 논문은 딥 러닝에서의 딥 뉴럴 네트워크의 효율적 학습을 위한 초기 가중치 설정에 관한 것으로, 비지도 학습을 이용하여 학습 데이터의 특징을 추출하는 딥 뉴럴 네트워크를 먼저 학습시킨 후, 학습된 딥 뉴럴 네트워크의 변수들을 지도 학습의 딥 뉴럴 네트워크 초기값으로 이용하는 방법을 제안한다.

Ⅰ. 서 론

딥 러닝에서는 손실함수(loss function 또는 cost function)가 가장 작은 값을 갖도록 딥 뉴럴 네트워크의 가중치(weight)를 학습시키는데, 주로 임의의 초기 가중치를 사용한다. 그런데 같은 구조의 딥 러닝을 학습시켜도, 가중치의 초기값에 따라 손실함수 값이 발산하는 경우도 발생하고, 수렴하더라도 수렴된 손실함수의 값이 다르게 나타날 수 있다. 또한 초기 가중치 설정에 따라 기울기 소실(gradient vanishing), 표현력(representation)의 한계 등 딥 러닝 학습에서 여러 문제를 야기할 수도 있다. 따라서, 딥 러닝 학습에서 초기 가중치 설정은 매우 중요한 역할을 한다.

딥 러닝 학습에서의 손실함수는 비볼록(non-convex)이고 많은 극소점(local minimum)을 갖기 때문에, 초기 가중치를 잘못 설정할 경우 최소점(global minimum)이 아닌 local minimum 으로 수렴할 가능성이 커지게 된다. 특히, 활성화 함수(activation function)가 시그모이드(sigmoid) 또는 정류 선형 유닛(ReLU)인 경우, 가중치 초기값의 절대값이 커지면 그래디언트(gradient) 소실(vanishing) 또는 폭주(exploding)가 일어나게 된다. 따라서, 딥 러닝에서 딥 뉴럴 네트워크가 안정적으로 수렴하기 위해서는 가중치 초기값을 작게 설정해야 하며 동일한 초기값을 갖지 않도록 랜덤하게 설정해야 한다. 가중치 초기값을 설정하는 방법은 다양하게 연구되어 왔으며, 주로 사용하는 방법으로는 Lecun initialization[1], Xavier initialization[2], He initialization[3] 등이 있다.

한편, 많은 데이터를 이용하여 비교적 층수가 얕은 네트워크를 학습시키는 경우에는 모델의 수용력(capacity)이 작기 때문에 과소적합(underfitting) 되는 경향이 있다.[4] 그러나, 네트워크의 층수가 깊어질수록 추론(inference) 시 계산량이 증가하여 계산 시간이 늘어나기 때문에 빠른 추론이 필요한 경우에는 네트워크의 층수를 줄이는 것이 중요하다. 즉, 많은 데이터를 이용하는 경우에도 층수가 얕은 네트워크를 이용하여 학습이 잘 되도록 할 필요가 있다.

따라서, 본 논문에서는 초기 가중치 값을 랜덤 하게 주지 않고, 입력 데이터의 특징이 잘 반영된 값으로 주는 방법을 제안한다. 제안한 가중치 초기화 방법을 사용하면 많은 데이터를 이용하여 학습시키는 경우에도 층수가 얕은 네트워크를 이용하여 학습이 잘 될 수 있도록 하는 것이 가능하게 된다.

Ⅱ. 본론

본 논문에서는 생산적 적대 신경망(Generative Adversarial Network, GAN)과 유사한 방식의 비지도 학습을 이용하여 지도 학습 기반의 딥 뉴럴 네트워크를 보다 안정적으로 학습시키기 위한 초기 가중치 설정 방법으로 학습 데이터의 특징을 추출하는 딥 뉴럴 네트워크를 먼저 학습시킨 다음, 학습된 딥 뉴럴 네트워크의 가중치 값들을 지도 학습 기반의 딥 뉴럴 네트워크의 초기 가중치 값으로 설정하는 방법을 제안한다.

일반적으로 지도 학습 기반의 딥 뉴럴 네트워크 기술에서 특징을 추출하는 네트워크는 그림 1 과 같이 층이 깊어질수록 네트워크의 폭이 점점 줄어드는 방향으로 설계가 된다. 한 예로, 가장 기본적인 딥 뉴럴 네트워크 구조인 다중 퍼셉트론(Multi-layer Perceptron, MLP)를 비롯하여 영상 데이터셋 학습을 위하여 많이 사용되는 합성곱 신경망(Convolutional Neural Network, CNN) 등에서의 네트워크는 대부분 폭이 점점 줄어드는 방향으로 설계가 되어 있다. 그림 1 에서 실선으로 나타낸 부분은 실제로 연산이 이루어지는 딥 뉴럴 네트워크이며, 점선으로 나타낸 부분은 입력 또는 결과 텐서들을 의미한다. 그림 2 는 일반적인 GAN 의 구조를 나타내고 있다.

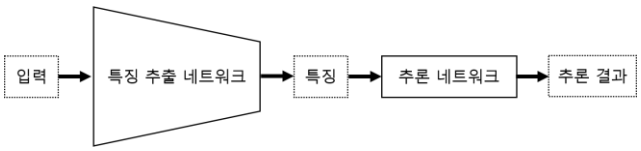


그림 1 지도 학습 기반의 딥 뉴럴 네트워크 구조

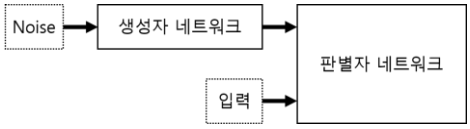


그림 2 Generative Adversarial Network (GAN) 구조

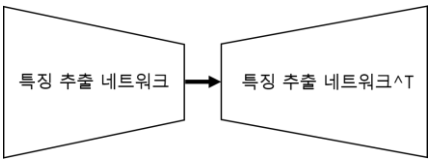


그림 3 생성자 네트워크 구조

본 논문에서는 주어진 데이터셋에 적합한 특징 추출 네트워크의 가중치 초기값을 얻기 위하여 그림 2 에서의 생성자 네트워크는 그림 3 과 같이 구성하고, 판별자 네트워크는 데이터셋에 따라 사용자가 자유롭게 구성하게 한다. 한 예로, 데이터셋이 영상 자료로 구성되어 있을 경우에는 판별자 네트워크를 CNN 으로 구성할 수 있으며, 일반적인 테이블 자료나 벡터로 구성되어 있을 경우에는 MLP 로 구성할 수 있다. 또한 그림 2 에서의 잡음(noise) 대신에 입력 값으로 주어진 데이터셋을 이용한다. 따라서 주어진 데이터셋에 적합한 특징 추출 네트워크의 가중치 초기값을 얻기 위한 장치는 그림 4 와 같이 구성된다. 그림 4 는 그림 2 의 GAN 구조와 유사하지만, GAN 은 생성자 네트워크의 입력으로 noise 를 주로 이용하는데 반해, 본 논문에서는 그림 4 와 같이 입력 데이터를 생성자 네트워크의 입력으로 이용한다는 차이가 있다.

그림 3 에서의 "특징 추출 네트워크"는 그림 1 에 나타낸 지도 학습 기반의 딥 뉴럴 네트워크에서 사용될 특징 추출 네트워크를 의미한다. "특징 추출 네트워크^T"는 특징 추출 네트워크에서의 계산 흐름과 반대로 이루어지는 네트워크이다. 생성자 네트워크를 통해 만들어지는 결과와 입력 데이터의 차원을 맞추기 위하여 그림 3 과 같이 생성자 네트워크를 구성하는 것이다.

그림 4 에서 판별자 네트워크는 입력 데이터로부터 생성자 네트워크를 통하여 나온 결과값인지, 아니면 원래의 입력 데이터인지를 판단하는 역할을 한다. 생성자 네트워크가 주어진 입력 데이터셋의 분포를 잘 학습하기 위하여, GAN 에서 제안된 손실 함수를 사용한다. 생성자 네트워크는 입력과 비슷한 데이터를 생성하도록 학습이 진행되고, 판별자 네트워크는 생성자 네트워크에서 만들어진 데이터인지, 아니면 원래의 입력 데이터인지를 잘 구별해내도록 학습이 진행된다. 생성자 네트워크와 판별자 네트워크 간의 적대적 관계를 통하여 결과적으로 생성자 네트워크는 원래의 입력 데이터들의 분포를 잘 학습할 수 있는 방향으로 학습이 진행된다. 또한, 생성자 네트워크를 구성하고 있는 특징 추출 네트워크에서는 입력 데이터의 특징을 효과적으로 추출해낼 수 있는 방향으로 학습이 진행된다.

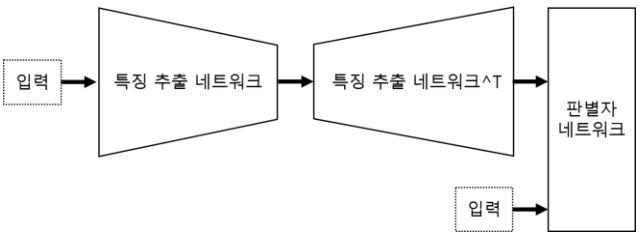


그림 4 딥 러닝에서의 딥 뉴럴 네트워크의 가중치 초기화 장치

따라서, 그림 4 의 장치를 통하여 가중치가 최적화된 특징 추출 네트워크를 그림 1 의 특징 추출 네트워크 초기값으로 설정하여 추론 모델을 학습시키면 층수가 얕은 네트워크를 이용하더라도 학습이 잘 될 수 있다.

Ⅲ. 결론

본 논문에서는 GAN 과 유사한 방식의 비지도 학습을 통하여 학습 데이터의 특징을 먼저 추출한 다음, 학습된 딥 뉴럴 네트워크의 가중치들을 초기값으로 설정하여 지도 학습에 이용하는 방법을 제안하였다.

제안한 가중치 초기화 방법을 사용하면 학습 데이터셋의 분포 및 특징을 먼저 학습한 다음 이를 가중치 초기값으로 설정하기 때문에, 층수가 얕은 네트워크를 이용하더라도 손실 함수가 발산하지 않고 수렴하는 효과가 있다.

딥 러닝에서 딥 뉴럴 네트워크의 층이 깊어지면 가중치들의 작은 변화가 출력 값의 큰 변화로 이어지는 불안정한 현상들이 생기기 때문에 처음부터 최적의 초기 가중치 값들을 사용한다면 출력 값들의 분포를 안정화시킬 수 있으며 학습 횟수가 많지 않더라도 성능이 좋은 모델을 만들 수 있는 효과가 있다.

ACKNOWLEDGMENT

이 논문은 2020 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2018-0-00841, IoT 디바이스를 위한 Lightweight 블록체인 표준개발).

참 고 문 헌

[1] LeCun, Y., et al. "Efficient BackProp," Neural networks: Tricks of the trade, pp. 9-50, Jan. 1998.

[2] Glorot, X. and Bengio, Y. "Understanding the difficulty of training deep feedforward neural networks," Proceedings of the thirteenth international conference on artificial intelligence and statistics, pp. 249-256, 2010.

[3] He, K., et al. "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," Proceedings of the IEEE international conference on computer vision, pp. 1026-1034, 2015.

[4] Goodfellow, I., et al. "Deep learning," The MIT press, Cambridge, 2016.

FPGA기반 저지연 데이터 이벤트 탐지 모듈 구현

윤기하, 김재인, 김성창
한국전자통신연구원
{gya, jaein, sungchang}@etri.re.kr

A Study on the FPGA based low-latency data event detector implementation

Yoon Giha, Kim Jaein, Kim Sung Chang
Electronics and Telecommunications Research Institute

요 약

이 논문에서는 저지연 IoT/IIoT 응용 서비스를 위하여 실시간으로 취득되는 센서 데이터로부터 사용자가 정의한 이벤트에 해당하는 데이터를 최소 지연으로 탐지할 수 있는 FPGA 기반의 하드웨어 모듈 구현에 대한 연구내용을 기술한다. Intel SoC FPGA 기반 개발보드를 활용하여 이벤트 탐지 모듈을 구현하였으며, RS485 통신기반으로 동작하는 전력품질 측정 센서를 활용한 실험환경에서 이벤트를 탐지 및 CPU로의 전송 시간은 평균 561.75μs 수준으로 확인되었다.

Ⅰ. 서 론

IoT(Internet of Things) 및 IIoT(Industrial IoT)에서는 각종 장치나 센서로부터 데이터를 수집하여 이를 가공 처리하는 데이터 수집 장치가 사용된다.[1, 2] 데이터 수집 장치는 단순히 데이터를 수집하여 타 인터페이스로 전달하는 일종의 게이트웨이 이며, 최근 수집한 데이터를 처리, 분석, 저장할 수 있는 엣지 컴퓨팅 기술이 게이트웨이에 적용되고 있다. 엣지 컴퓨팅기반의 게이트웨이는 각각의 응용서비스에 따라 범용적으로 적용될 수 있도록 다양한 수집용 인터페이스 및 복수의 프로토콜을 탑재한 프로세서 시스템으로 구성된다.[3] 이 논문에서는 프로세서 시스템이 탑재된 SoC FPGA를 활용하여 RS485기반의 전력센서로부터 획득되는 다양한 전력품질데이터의 사용자 정의 이벤트를 최소지연으로 획득하는 저지연 이벤트 탐지 모듈 설계 및 구현에 대하여 기술한다.

Ⅱ. FPGA기반 저지연 이벤트 탐지 모듈 구현

이 연구에서 제안하는 저지연 이벤트 탐지 기능은 데이터 획득 대상(센서)으로부터 데이터가 수신되어 수신버퍼에 쌓기 전 단계(Media Access Control Layer)에서 사용자가 정의한 이벤트 발생 여부를 판별하여 프로세서에 결과를 알릴 수 있도록 설계하였다. 하드웨어로 구현된 이벤트 탐지 모듈을 활용하면 프로세서로 처리하는 이벤트 판별에 관한 일련의 과정을 생략할 수 있어, OS(Operating System)기반 환경에서 복수 프로그램 동작에 의한 부하로부터 생길 수 있는 센서 데이터 취득 시점의 비주기성을 억제할 수 있다. 이 연구에서는 INTEL사의 SoC FPGA(System on a Chip with Field-Programmable Gate Array)인 Cyclone V SE 시리즈(5CSEBA6U23I7)를 탑재한 TERASIC사 개발보드(DE10-Nano Kit)를 활용하였다. 이 연구에 사용된 개발보드는 800MHz로 동작하는 Dual-core ARM Cortex-A9 프로세서 및 1GB DDR3 메모리를 포함한 컴퓨팅 시스템이 FPGA 내에 Hard IP(Intellectual Property) 형태로 집적되어, 프로세서와 FPGA간 내부 인터페이스를 통해 고속으로 데이터를 전달할 수 있다.[4] 아래 그림 1은 이 연구를 통해 설계한 SoC FPGA 기반 이벤트 탐지 모듈의 기능블록도를 나타낸다. 그림 1과 같이 FPGA 영역에서 하드웨어로 설계한 이벤트 탐지 기능블록은 UART(Universal Asynchronous Receiver/Transmitter) 송·수신 기능블록과 수신버퍼(RxBuff) 사이에서 1바이트 단위로 수신 프레임을 수집하여 실시간으로 이벤트를 탐지할 수 있도록 설계하였다. 사용자는 User Config. 기능블록을 통해 탐지대상 데이터의 위치, 크기, 탐지 조건 및 비교 데이터 정의할 수 있다.

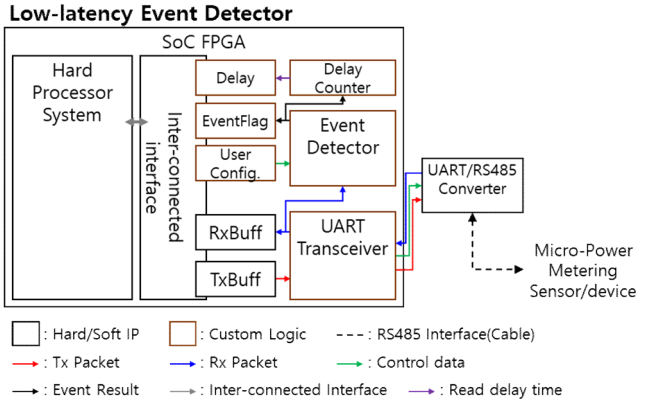


그림 1. SoC FPGA 기반 이벤트 탐지 모듈 블록도

수신 데이터를 실시간으로 감시하여 사용자 정의 이벤트 여부를 탐지한 결과는 SoC FPGA 내부 인터페이스를 통해 프로세서에서 읽을 수 있도록 설계하였다. 이벤트 탐지에 대한 결과를 EventFlag 레지스터에 갱신하며, 갱신과 함께 Delay Counter를 구동하여 프로세서가 이벤트 탐지 결과를 읽는데 소요되는 지연시간을 확인할 수 있도록 부가 기능블록을 추가하였다. 이벤트 탐지 결과 획득 지연시간은 Delay Counter를 통해 FPGA 영역에서 생성·계산되며, 이 기능을 바탕으로 구현한 저지연 이벤트 탐지 모듈의 이벤트 탐지 지연시간을 확인하였다. 이 연구에서 사용한 개발보드의 프로세서 시스템에는 리눅스 Ubuntu16.04 LTS(Long-Term Support)를 탑재하였다. 따라서, 프로세서가 내부 인터페이스를 통한 FPGA와 데이터 전달·획득에 소요되는 시간에 불규칙한 편차가 존재한다.[5] 프로세서에서

Delay 값을 읽은 뒤, 읽음 신호를 다시 한번 FPGA에 전달하는 절차까지 이벤트 탐지 지연시간으로 정의하였다. 이와 같은 환경에서 개발한 저지연 이벤트 탐지 모듈의 이벤트 탐지 지연시간은 최소 101.49μs에서 최대 972.26μs로 확인되었으며 평균 561.75μs 성능을 보이는 것으로 확인하였다. 그림 2는 본 연구에서 구현한 저지연 이벤트 탐지 모듈의 이벤트 탐지 지연시간을 데이터 획득시간 순으로 나열한 것이다.

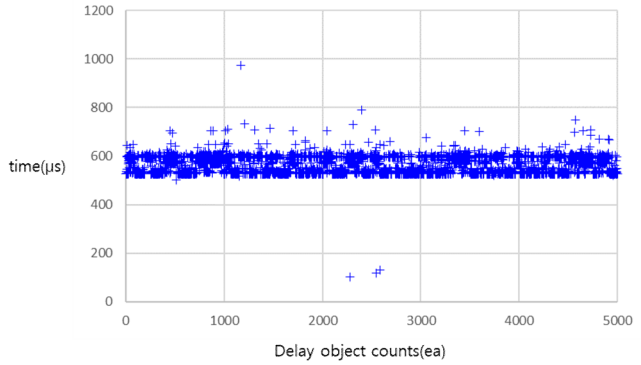


그림 2. 제안 모듈의 이벤트 탐지 지연시간 측정 결과(5,000개 샘플)

위 그림 2에서 보이는 바와 같이 FPGA 기반의 하드웨어 지원을 통한 이벤트 탐지 지연시간의 최대 편차가 870.77μs로 확인되었다. 아래 그림 3은 SoC FPGA 내에 집적된 Dual-core ARM Cortex-A9 프로세서에 의존한 소프트웨어 기반 이벤트 탐지 지연시간 결과 샘플을 나열한 것이다.

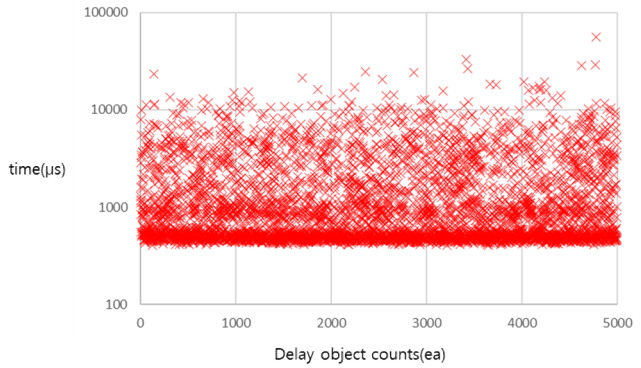


그림 3. S/W 기반 이벤트 탐지 지연시간 측정 결과(5,000개 샘플)

집적된 프로세서 기반의 소프트웨어 이벤트 탐지 지연시간은 최소 407.51us에서 최대 55.82ms으로 평균 1.70ms 수준으로 확인되었다.

위 그림 2, 3에서 보이는 것과 같이 제안한 FPGA기반 저지연 이벤트 탐지 모듈이 소프트웨어 처리기반의 이벤트 탐지 지연시간 편차 55.41ms 대비 약 1.57% 수준의 적은 편차로 동작하는 것을 확인할 수 있다. 아래 표 1은 제안한 저지연 이벤트 탐지 모듈과 소프트웨어 기반 이벤트 탐지 성능 비교표이다.

표 1. 소프트웨어 기반 이벤트 탐지 및 저지연 이벤트 탐지 모듈 성능 비교

Delay Type	*Event Detector	Software
Min.	101.49	407.51
Max.	972.26	55817.56
Deviation	870.77	55410.05
Avg.	561.75	1697.37

(Unit:μs)

표 1에서 보이는 것처럼 제안한 FPGA 기반의 저지연 이벤트 탐지 모듈은 소프트웨어 기반의 이벤트 탐지보다 약 4배에서 549.9배 우수한 성능을 갖는 것으로 확인되었다.

Ⅲ. 결 론

본 논문에서는 IoT/IIoT 분야에서 센서로부터 데이터를 수집함에 있어, 사용자 정의에 따른 이벤트를 최소지연으로 탐지하기 위한 SoC FPGA 기반의 저지연 이벤트 탐지 모듈 설계 및 구현에 대하여 기술하였다. 프로세서가 집적된 SoC FPGA는 제안한 이벤트 탐지 모듈을 프로세서에서 고속으로 접근하기 용이한 구조로 FPGA 영역에 구현된 하드웨어 기능으로부터 이벤트 탐지 결과를 획득하는데 평균 561.75μs 소요됨을 확인하였다. 이러한 기술은 복수의 센서로부터 데이터를 취득하며 이벤트를 탐지하는 응용의 하드웨어 지원부분, 인공지능 학습결과를 바탕으로 작성된 Look-Up Table 기반의 학습형 이벤트 탐지 등 IoT/IIoT 분야에서 특수목적 이벤트를 빠르게 탐지하여 대처하기 위한 기반 기술로 활용할 수 있다.[6, 7]

ACKNOWLEDGMENT

본 연구 논문은 한국전자통신연구원 연구운영지원사업의 일환으로 수행되었음.[20ZK1100, 호남권 지역산업 기반 ICT 융합기술 고도화 지원사업]

참 고 문 헌

[1] L. D. Xu, W. He and S. Li, "Internet of Things in Industries: A Survey," Nov. 2014. IEEE Transactions on Industrial Informatics, Vol. 10, no. 4, pp. 2233-2243.

[2] Young Seog Yoon, Hangjung Zo, Munkee Choi, Donghyun Lee, and Hyun-woo Lee. "Exploring the dynamic knowledge structure of studies on the Internet of things: Keyword analysis," Nov. 2018. ETRI Journal Vol. 40(6), pp. 745-758.

[3] Dell, "IoT Deployment Is Driving Analytics To The Edge", April. 2019. (https://www.delltechnologies.com/en-us/)

[4] terasIC, "DE10-Nano User Manual (rev. A/B Hardware) version 1.8", Feb. 2018. (https://www.terasic.com.tw)

[5] Intel, "Cyclone V Hard Processor System Technical Reference Manual," Sep. 2020. (https://www.intel.com)

[6] Giha Yoon, Jaein Kim, Geun-Yong Kim, Byunghee Son, and Hark Yoo. "Multiple RS-485 interface management FPGA design for Power micro-metering," ICPE 2019-ECCE Asia, pp. 2635-2640.

[7] Sung Il Na, Hyoung Joong Kim. "Design of Anomaly Detection System Based on Big Data in Internet of Things," Feb. 2018. Journal of Digital Contents Society, Vol. 19(2), pp. 377-383.

AI(Deep Learning)을 이용한 S18650리튬이온배터리 SOC예측에 관한 연구

배정효, 진윤선, 백지국*, 박민원**, 딘민차우**,김창순**,다오반권**

한국전기연구원, *(주)아이이에스, **창원대학교

jhbae@keri.re.kr, *kawabai@gmail.com,**capta.paper@gmail.com

A Study on the SoC estimation of 18650 Rechargeable Li-Ion battery by AI(DeepLearning)

JungHyo, Bae YunSeon,Jin, JiKook Baek*, MinWon Park**,MinhChau,Dinh**,

ChangSun,Kim**,Van Quan.Dao**

KERI, *IES Co. LTD., **Changwon Univ.

요 약

본 논문은 25℃, 60℃ 실험온℃ 조건에서 S18650 2차전지를 116싸이클 충/방전 후 수집한 데이터로 데이터 셋을 만들어 ANN(Artificial Neural Network) 신경망을 이용해 배터리충전상태 즉, State of Charge (SOC)를 예측하는 알고리즘을 개발하였다. 배터리 충/방전 데이터 셋은 전압, 전류, 온도(℃), SOC 데이터가 포함되어 있다. 이 중에 70% 데이터를 활용해서 AI 신경망 모델 학습용으로 사용하였으며, 30% 데이터는 AI모델 테스트 및 검증으로 사용하였다. AI모델은 입력, 출력, Hidden layers, Activation Functions 등으로 구성하였다. Hidden layers층은 총 2개이며 한 층에 32개 노드를 사용할 때에 예측한 결과가 25℃ 조건에서 SOC 평균 에러율은 0.32%, Max 에러율은 1.6% 이었으며, 60℃ 시에는 SOC 평균 에러율은 0.41%, Max 에러율은 2.2% 가 ℃출 되었다.

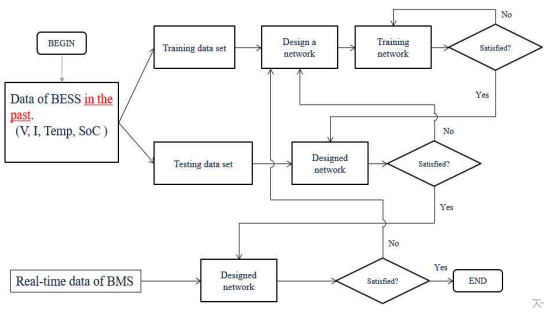
Ⅰ. 서 론

본 논문에서는 현재까지 여러 배터리 충전상태 예측 방법에 대해 조사를 했다. 그리고 각 방법에 대한 장단점을 분석하였다. 이 중에는 AI 딥러닝으로 배터리 충전상태 예측방법이 실시간, 비선형 및 다양한 유형의 배터리에 높은 적응성 등의 특징을 가지고 있다.[1]

그리고 요즘에 인공지능 및 AI시대가 핫한 이슈로 되는 바람에 세계 각 나라에는 많은 연구 및 관심을 갖고 있다. 또한 한국전기연구원에서 AI 인공지능으로 전기자동차 배터리 관리 시스템을 개발하는 과제를 수행하고 있는 내용 중에, 본 논문에서는 AI신경망을 이용해 S18650 리튬이온 2차전지 충전상태 (State of Charge 즉 SoC)를 예측하는 알고리즘을 개발 및 실험하였다.

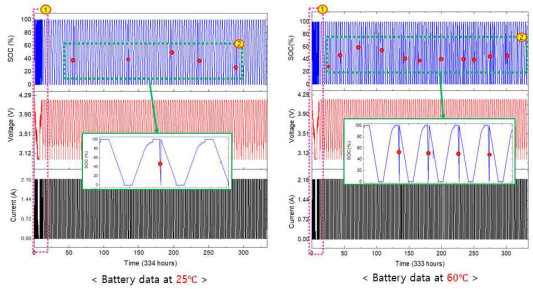
Ⅱ. 본론

먼저 배터리 SOC예측에 대해 AI 모델 개발 과정은 그림1 과 같다.



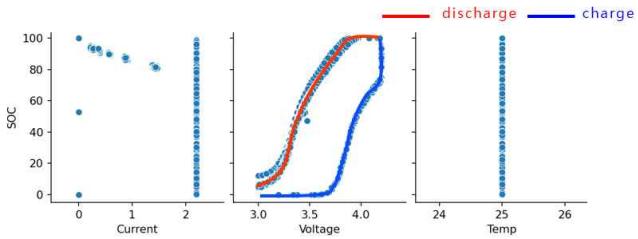
<그림 1. SOC예측에 대해 AI 모델 개발 과정>

이 과정에는 먼저 신경망을 설계하고 학습하는 것이다. 그리고 데이터 셋을 신경망 학습용, 테스트용, 실시간 검증용 등 3가지 용도로 사용하였다. 그리고 그림 2는 배터리데이터를 분석한 것이다.



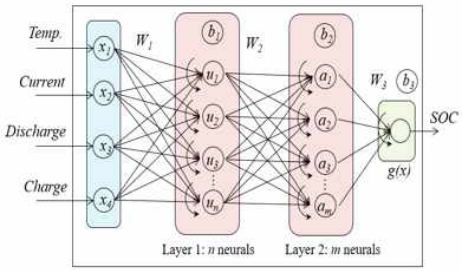
<그림 2 배터리 데이터 분석>

여기서 25℃/60℃ 각각에 대한 전압 전류 SOC값으로 분석하여 도시하였다. 1/2로 표시한 부분은 Trash data이며 정확한 AI모델을 개발하기 위해서 이런 Trash data를 제거하여 사용하였다.



<그림 3. 전압 전류 온도와 SOC 관계 분석 >

그림 3은 SOC와 전압과 강한 관계를 보이며, 지수함수로 표현할 수 있는 것으로 판단이 되며, 전압 값은 충전/방전으로 나눠서 input으로 사용해야 한다. ANN 신경망을 이용한 SOC예측 모델은 그림4와 같다[2].



<그림 4 ANN 신경망을 이용한 SOC예측 모델 >

AI 모델의 설계과정은 표1과 같다.

<표 1 AI 모델의 설계과정 >

STEP	설계 내용	비고
1	데이터분석 및 학습용 데이터 설정	-Training (80%) -Validation (10%) -Testing (10%)
2	신경망을 구축 및 학습	-Input layer: V(t), I(t), T(t) -Hidden layer: No. of layers -Output layer: SOC(t)
3	ANN의 active function 선정	-Sigmoid, Linear, Tanh, ReLu
4	ANN의 optimizer 선정	-Adam, AdaDelta, SGD, …
5	학습과정에 필요한 Loss function및 metrics선정	-Mean squared error (MSE) -Mean absolute error (MAE) -Metrics = [MSE, MAE]
6	최적 학습률 인자 α 확정	$0 < \alpha < 1$

설계한 AI모델로 학습한 후에 각각의 비교결과는 다음과 같다.

1. Activation function	No. of neural	Learning rate	Optimizer	Epoch	MAE (%)
Sigmoid (Hidden layer)	4 – 32 – 32 – 1	0.1	Adam	500	24.2
	4 – 32 – 32 – 1	0.01	Adam	500	3.12
	4 – 32 – 32 – 1	0.001	Adam	500	1.34
Tanh (Hidden layer)	4 – 32 – 32 – 1	0.1	Adam	500	22.1
	4 – 32 – 32 – 1	0.01	Adam	500	1.94
	4 – 32 – 32 – 1	0.001	Adam	500	0.32

*Active function & Learning rate 비교

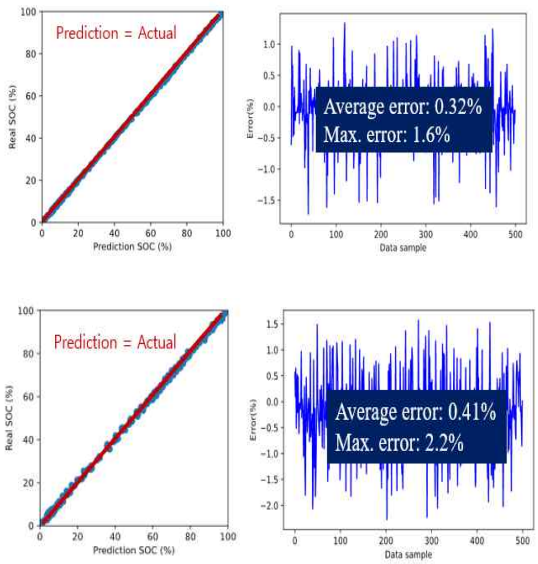
**Number of hidden layer 비교

2. Hidden layer	No. of neural	Activation	Epoch	MAE (%)
Single hidden layer (Basic ANN)	4 – 8 – 1	Tanh	500	21.32
	4 – 16 – 1	Tanh	500	8.36
	4 – 32 – 1	Tanh	500	6.27
	4 – 64 – 1	Tanh	500	6.11
Multiple hidden layers (Deep ANN)	4 – 32 – 32 – 1	Tanh	500	0.32
	4 – 32 – 32 –32 – 1	Tanh	500	0.33
	4 – 32 – 32 –32 – 32 –1	Tanh	500	0.31

<그림 5 각각 학습조건에서 SOC예측 에러율 >

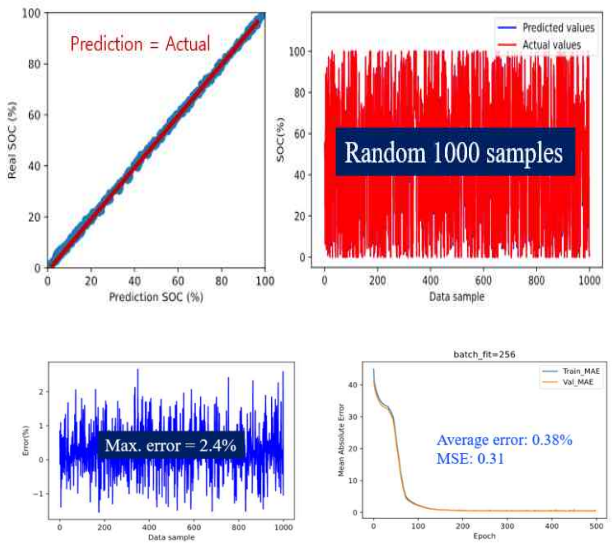
Activation function에서 32개의 노드로 설정(고정)한 이유는 노드 수는 8,16,32,64로 실험을 해 본 결과가 32개 노드를 가질 때에 에러 율이 가장 정확하고 정밀하였다. 그리고 Hidden layer에서 No. of neural은 4-32-32-1로 고정한 이유가 Hidden layer 1,3 층 일 때 보다 예측결과가 정밀하며 4층과 (즉 4-32-32-32-32-1) 비교하면 에어 율이 0.01%높지만 산력 소모량을 고려해 4-32-32-1이 가장 이상적인 설계였다.

그림6에서 25℃ Tanh Active function 및 32개 노드를 가진 Hidden Layer 층 2개를 이용할 경우, 평균 에러율은 0.32%가 되었다. 그리고 60℃ 평균 에러율은 0.41%가 되었다.



<그림 6 25℃ 및 60℃ 예측 에러율 >

또한 온도 상관없이 random으로 1,000개 데이터로 이용해서 예측할 때 나온 예측 에러율은 그림 7과 같다.



<그림 7 1000개 random 데이터로 예측 결과 >

1,000개의 샘플 충방전 데이트로 검증한 결과가 25℃나 60℃ 조건에서 평균 에러율은 0.38% 및 MAX 에러율이 2.4% 나타났다.

Ⅲ. 결론

본 논문에서는 AI 신경망 모델을 설계하고 S18650 배터리 충방전 데이터를 이용해 SOC예측하는 알고리즘을 개발 및 실험하였다. AI 신경망 모델설계과정은 6단계로 나눠서 설계했고, 많은 학습 후에 나온 결과를 분석하였으며, 결론적으로, ANN신경망에는 Active function은 Tanh로 선정하고, Hidden Layer은 2개로 정하고, 한 층에 32개 노드를 갖고 있을 때에 SoC예측 결과가 평균 에러율은 0.38% 및 MAX 에러율이 2.4% 로써, 매우 정밀하게 예측 되었다.

ACKNOWLEDGMENT

본 연구는 한국전기연구원 및 주식회사 아이이에스 수행 중인 'HKF와 Deep Learning을 이용한 화재예방용 고정밀 Smart-BMS 개발' 과제에게 지원을 받았다.

참 고 문 헌

[1] Yan Q. and Wang Y., "Predicting for power battery SOC based on neural network," 2017 36th Chinese Control Conference (CCC), 2017,4140-4143

[2] David J-B, Jesús F-A.,"Using Dynamic Neural Networks for Battery State of Charge Estimation in Electric Vehicles," Procedia Computer Science 130 2018, 533 - 540

지능형 엣지 컴퓨팅 시스템을 위한 소프트웨어 및 하드웨어 구현방안에 관한 연구

김재우, 김동성*
ICT융합특성화연구센터, *금오공과대학교
jaewookim@kumoh.ac.kr, *dskim@kumoh.ac.kr

A Study of Software and Hardware Implementation Schemes for Intelligent Edge Computing Systems

Kim Jae Woo, Kim Dong Seong*
ICT Convergence Center, *Kumoh National Institute of Technology

요 약

본 논문은 지능형 엣지 컴퓨팅 시스템을 위한 임베디드 디바이스 구현방식에 대해 비교분석 하였다. 엣지 컴퓨팅 시스템에서의 지능형 기술은 구현방식에 따라 소프트웨어 기술과 하드웨어 기술로 분류할 수 있다. 인공지능 기술에서 핵심 기술인 신경망을 소프트웨어 적인 방법으로 구현할 것인지 또는 하드웨어적인 방법으로 구현할 것인지에 따라 분류 될 수 있다. 본 논문에서는 각 구현 기술에 대하여 설명하고 몇 가지 관점에서 각 구현기술의 장단점을 도출하였다. 본 논문의 분석 결과는 지능형 임베디드 디바이스 개발에 참고할 수 있도록 가이드라인을 제시한다.

Ⅰ. 서 론

엣지 컴퓨팅은 중앙 네트워크의 부하분산기능과 빠른 대응시간을 위해 제안된 컴퓨팅 기술이다. 그러나 부하분산만의 목표가 아닌 정보의 획득과 처리 그리고 전달까지의 처리를 센서 디바이스 또는 모바일 디바이스와 같은 시스템의 엣지 단에서 수행하여 서비스의 빠른 실시간성을 목표로 하고 있다[1]. 한편 높은 신뢰성이 요구되는 산업용 시설 및 안전 설비에 관련된 기기들과 단말에서 단순히 정보를 습득하는 것만이 아닌 지능형 컴퓨팅이 요구되고 있다[2]. 따라서 모바일 기기에서 상황 및 객체 인식과 같은 지능형 서비스를 제공하기 위해 모바일 엣지 디바이스에서 지능형 컴퓨팅이 연구되고 있다[3].

엣지 디바이스와 같은 임베디드 디바이스에서 지능형 컴퓨팅을 하기 위해서 두 가지 인공지능 기술이 적용되고 있다. 첫째는 임베디드 보드에 고성능 CPU나 GPU를 하드웨어를 탑재하여 소형 PC나 소형 서버의 역할을 수행할 수 있을 뿐만 아니라, 지능형 컴퓨팅까지 수행 가능한 보드들이 개발되었다. 라즈베리 파이4, 라떼판다 등과 같은 보드는 모바일 CPU를 코어로 개발된 개방형 보드와, GPU를 탑재한 보드는 엔비디아의 제슨시리즈 보드들이 대표적이며 비맥스테크놀로시사의 Nuvo-5095GC와 같은 CPU와 GPU를 함께 탑재하고 있는 산업용 엣지 컴퓨팅 솔루션도 있다[4]. 이러한 GPU를 탑재하고 있는 보드위에 소프트웨어적인 딥러닝 알고리즘을 구현해야 한다.

두 번째는 인간의 신경 구조, 즉 뉴런을 모방한 하드웨어 뉴런을 병렬로 연결한 뉴로모픽 기술이다[5]. 즉 인간의 뇌의 하드웨어적인 뉴런구조를 반도체 칩으로 구현하여 병렬 뉴럴 네트워크를 구성한 기술이다. 일반적인 반도체 칩은 폰노이만 방식을 기본으로 데이터가 입력되면 이를 순차적으로 처리해 단순 작업을 빠르고 효율적으로 해내는 것에 최적화 되었지만 뉴로모픽 기술은 칩 안에 여러 개의 코어를 두고, 코어 내에 소자와 메모리가 뉴런 및 메모리 역할을 동시에 수행하는 구조로 설계되어있다.

본 논문에서는 첫 번째와 두 번째 경우를 각각 Software based Intelligent Edge Computing :SIEC) 시스템, Hardware based Intelligent Edge Computing :HIEC) 시스템이라 정의하였다. 뉴로모픽 칩을 이용한 학습을 위해서는 학습하고자 하는 데이터 셋트를 전처리하여 넣어주면 뉴로모픽 칩내의 뉴런소자들이 즉각적으로 반응하는 구조로 수행된다. 이런 기술은 실제 데이터에 대한 지속적인 학습과 빠른 적응이 필요한 로봇공학, 자율주행 등과 같은 인공지능 애플리케이션에 적용 될 수 있다. 그림1은 현재 SIEC/HIEC 시스템 디바이스 및 칩이다.

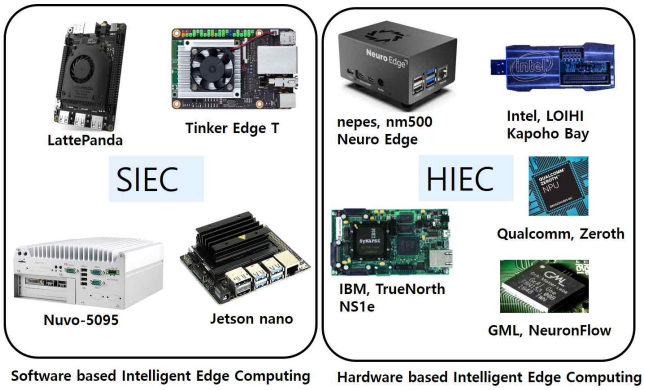


그림 1 지능형 엣지 컴퓨팅 디바이스(SW & HW)

본 논문에서는 SIEC 시스템과 HIEC 시스템을 비교 분석하고, 지능형 엣지 컴퓨팅 시스템 애플리케이션에 따라 어떤 시스템이 적합한지에 대해서 고찰하였다.

Ⅱ. 본론

본 장에서는 SIEC와 HIEC 시스템의 특성을 서술하고 각 관점에 따른 장단점을 비교 분석하였다.

항목	SIEC	HIEC
응답성	- 높은 정확도를 위한 학습 연산 및 추론을 위해 많은 시간이 걸림	- 학습, 추론, 저장이 동시에 처리하여 빠른 응답속도 - 지연시간 감소
정확도	- 하드웨어 성능과 인공지능 모델에 따라 높은 정확도를 기대할 수 있음	- 하드웨어로 이루어진 뉴런의 개수와 학습양에 비례하여 정확도가 높아짐
확장성	- 다양한 애플리케이션에서 적용가능 - 지능형 엣지서버로써 동작 가능	- 비정형 데이터를 처리하는 애플리케이션에 빠른 적용 가능
구현성	- 기존 인공지능 라이브러리 사용가능 - 데이터 인터페이스를 위한 구현 필요	- 뉴로모픽 칩에서 제공하는 학습 알고리즘 적용 - 데이터 인터페이스를 위한 구현 필요
효율성	- 전이학습을 위해 기존 학습모델을 이용하기에 용이 - 사용 성능에 따라 디바이스 선택의 폭이 다소 큼 - 학습 연산을 위한 딥러닝 가속기로 인해 소비전력이 큼	- 뉴로모픽 칩에 프로세스와 메모리가 통합되어 있어 별도의 메모리가 필요 없음 - 뉴런을 병렬로 구성해 저전력으로 구동

표 1 SIEC와 HIEC의 비교 테이블

1. SIEC(Software based Intelligent Edge Computing)

서론에서 전술하였듯이 SIEC 시스템은 종래의 인공지능 소프트웨어 기술을 사용한다. 종래의 인공지능 기술은 신경망을 복잡한 소프트웨어 코딩을 통해 구현하고, 구현된 인공 신경망은 다수의 레이어간 연결 및 일정한 데이터 전파를 통해 학습과 학습된 결과를 출력한다. 2012년 스스로 특징을 만들어내는 신경망인 딥러닝 기술이 등장한 이후 인공지능 알고리즘은 폭발적으로 성장하여 다양한 분야에 딥러닝 모델이 사용되고 있다. SIEC 시스템은 그간 발전에 딥러닝 기술을 라이브러리화 하여 제공되는 소프트웨어 개발환경을 쉽게 구축할 수 있으며 이를 이용하여 지능형 엣지 컴퓨팅 시스템을 구현할 수 있는 장점이 있다. 현재 출시되어 있는 대부분의 지능형 엣지 컴퓨팅 디바이스는 고성능의 CPU와 GPU가 탑재되어 있고 용도에 따라 특정 인터페이스를 내장하고 있는 임베디드 보드이다. 따라서 운영체제를 설치해야 하며 딥러닝 알고리즘을 사용하기 위한 라이브러리와 사용 환경을 설정하여 엣지 컴퓨팅을 구현할 수 있다.

2. HIEC(Hardware based Intelligent Edge Computing)

HIEC 시스템은 뉴로모픽 칩을 이용하여 엣지 컴퓨팅 시스템을 구현한다. 인간의 뇌의 구조를 모방한 하드웨어 반도체 칩을 이용하여 메모리 저장과 연산 그리고 연산결과 출력을 동시에 수행할 수 있는 구조이다. 따라서 정형화 되지 않는 데이터도 직관적으로 인식하고 인식부터 처리 까지 응답속도가 뛰어나다. HIEC 시스템은 뇌에 해당하는 뉴로모픽 칩과 눈과 입과 같은 입출력을 위한 주변 하드웨어 장치로 구성된다. HIEC 시스템을 구현하기 위해서는 SIEC 시스템과 같이 복잡한 연산을 위한 알고리즘 구현은 필요하지 않지만 뉴로모픽 칩이 데이터를 인식할 수 있도록 전처리 과정 그리고 뉴로모픽 칩을 구동하기 위한 드라이버를 주변 장치에 구현하여야 한다.

표1은 SIEC와 HIEC 시스템의 비교 테이블이다. 엣지 컴퓨팅 시스템을 대상으로 다섯 가지 관점으로 비교하였다. 엣지 컴퓨팅의 등장배경이라 할 수 있는 지연시간 및 응답속도와 같은 문제해결을 위해서는 SIEC 시스템에 비해 HIEC 시스템이 월등히 유용하다고 할 수 있다. 이는 뉴로모칩 하드웨어 반도체를 이용한 구조적인 처리방법 때문이다. 사례로 네패스사의 뉴로모픽 칩인 NM500과 CPU의 처리성능을 비교하였을 때 동일한 학습을 수행하기 위해 초당 수행 횟수가 NM500이 CPU보다 0.0005% 밖에 되지 않았다[6]. 또한 소비전력 면에서도 HIEC 방식이 더욱 뛰어나다. 반면 엣지 컴퓨팅의 애플리케이션이 점차 다양해지고 있음을 볼 때 좀 더 애플리케이션에 대한 확장성은 적용할 수 있는 기존 솔루션과 라이브러리를 이용하기에 용이한 SIEC 방식이 더 장점이 있다고 할 수 있다. 이는 구현성에서도 이점을 가질 수 있다. 기존 학습모델을 이용하는 전이학습의 경우 SIEC 방식은 기존 검증된 공개 학습 모델을 기초로 추가 학습을 수행하면 되지만, HiEC 의 경우 전이학습을 위한 기본 학습 모델또한 학습을 수행해야 한다.

Ⅲ. 결론

본 논문에서는 엣지 컴퓨팅을 위한 임베디드 디바이스에서의 인공지능 기술에 대해서 비교하였다. 엣지 컴퓨팅 시스템에서의 인공지능 기술을 하드웨어의 구조와 구현방식에 따라 소프트웨어 방식(SIEC)과 하드웨어 방식(HIEC)으로 정의하였다. 그리고 각 방식을 연구 및 개발을 위한 6가지 관점에서 비교 분석하였다. 분석 결과 통해 지능형 임베디드 장치 개발에 참고할 수 있다. 향후 연구로는 SIEC와 HIEC의 구체적 사례를 시험하여 사례를 통해 더욱 실제적인 분석을 수행할 것이다.

ACKNOWLEDGMENT

이 논문은 2020년도 정부의 재원으로 한국연구재단의 지원과 과학기술정보통신부 정보통신기획평가원의 지원을 받아 수행된 연구임
(2018R1A6A1A03024003,2019R1I1A1A01063895,IITP-2020-2020-0-01612)

참 고 문 헌

[1] Weisong S, Jie C, Quan Z, Youhuizi ,L Lanyu X," Edge Computing: Vision and Chanllenges,"IEEE Internet of Things Journal, vol. 3, No. 5, pp. 637-646, 2016

[2] Jae-Woo K, Dong-Seong K, “The System of Intelligent Edge Computing for Fault Tolerance of Industrial IoT Devices”, KICS Winter Conference 2020, pp281-282, Yong Pyong Korea, Fab 2020

[3] Guangxu Z, Dongzhu L, Yuqing D, Changsheng Y, Jun Z, Kaibin H, "Toward an Intelligent Edge: Wireless Communication Meets Machine Learning," IEEE Communications Magazine, vol 58, Issue 1, pp19-25, 2020

[4] "Datasheet : Nuvo-5095GC,"Neousys Technology 2001, (www.neousys-tech.com).

[5] Zhequ. Yu, Amir M. Abdulghani, Adnan Zahid, Hadi Heidari, Muhammad Ali. lmran, Qammer H. Abbasi, “An Overview of Neuromorphic Computing for Artificial Intelligence Enabled Hardware-Based Hopfield Neural Network”, IEEE Access, vol.8, pp.67085-67099, 2020.

[6] Nepes, NM500 User’s Manual Version 1.6.3, ser, Revised Edition. General Vision and Nepes Koea, Apr. 2019 [Online] Available: http://www.theneuromorphic.com/nm500/

하드웨어 성능에 따른 보행자 검출 성능평가 방법에 관한 연구

김희강, 손준우, 김창홍, 김지연, 조태식, 한동석*

한국건설생활환경시험연구원, *경북대학교

heekangkim@kcl.re.kr, joonsohn@kcl.re.kr, kch0420@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 Hardware Performance

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

Korea Conformity Laboratories, *Kyungpook National Univ.

요 약

실제 CCTV 카메라가 설치된 환경은 매우 열악한데도 불구하고 기존의 객체 검출같은 소프트웨어 알고리즘 성능 검증은 이를 반영하지 않아 실제 환경에서는 알고리즘 성능이 현저히 떨어지는 경우가 있다. 이를 보완하고 검증하기 위해서는 실제 환경기반에서 하드웨어 성능까지 고려된 환경에서 객체검출 성능을 검증하는 방법이 필요하다. 본 논문에서는 동일 카메라에서 설정 변경을 통해 다이나믹 레인지(DR) 값에 따른 객체 검출률을 비교한 결과 동일한 알고리즘의 객체 검출 소프트웨어 성능이 최대 225% 차이가 발생하였다.

Ⅰ. 서 론

CCTV(Closed-Circuit TeleVision)는 기존의 아날로그 신호로 송출되던 전송 방식이 IP 기반으로 발전하면서 지능형 CCTV로써 단순한 범죄 방지, 추적, 식별에 머무르지 않고, 영상 분석 기능을 통해 재난 감시, 교통관제, 불법 주정차 단속 등 다양한 분야에서 활용되고 있다.

지능형 CCTV의 발전으로 기존의 컴퓨터 비전 분야에서는 다양한 객체를 검출 및 인식과 관련된 연구를 수행해왔으며, 이러한 기술들이 점차적으로 CCTV에 접목되어 지능형 CCTV의 급진적인 발전을 이루고 있다.

CCTV의 발전으로 기존의 관제 역할에서 인공지능 지능형 CCTV로 실시간 객체를 검출, 인식, 추적 등 컴퓨터 비전 분야에서 발전된 기술들이 접목되고 있다. 객체 검출 등 인공지능 지능형 CCTV에 탑재되는 소프트웨어는 제한된 환경에서 구현되어 최상의 성능을 나타내는데 실제 CCTV가 설치된 실환경에서는 제대로 성능이 나오지 못하는 문제가 발생한다.

이러한 이유는 극한의 설치환경에서는 카메라 하드웨어 성능에 따라 제대로 영상이 표출되지 않기 때문이다. 이러한 하드웨어 성능을 고려하지 않고 단순 알고리즘 평가는 객관적인 지표로 설정할 수 없다.

본 연구에서는 조명이 없는 복도 입구에서 하드웨어 성능지표인 다이나믹레인지(Dynamic Range)의 성능에 따른 보행자 검출률을 비교하여 실환경 및 하드웨어 성능에 따른 객관적인 보행자 검출 성능 평가 방법을 제안한다.

Ⅱ. 본론

보행자 검출 장소는 한국건설생활환경시험실 시험실 복도에서 진행하였으며, 해당 복도는 그림 1과 같다. 복도의 길이는 약 30 m, 폭 약 2.5 m, 카메라 설치 천장 높이 약 3 m이며, 카메라는 실내에서 사용되는 돔(Dome) 카메라를 사용하여 보행자 검출을 진행하였다.

해당 시험 장소는 복도는 조명이 꺼진 상태에서 어둡고 복도 밖은 밝기 때문에 극한의 밝기 차가 발생하며, 다이나믹레인지(DR) 성능 값에 따른 영상의 품질 변화가 명확하므로 정량적으로 비교할 수 있다.

그림 1 보행자 검출 시험 환경

보행자는 흰색 가운을 입고 약 15 m 지점까지 걸어가서 돌아오는 방법으로 비디오 영상 클립의 시간은 약 24초로 녹화하여 700 개의 프레임을 추출하여 진행하였다.

동일한 카메라에서 설정을 변경하여 다이나믹레인지 성능을 3가지로 구분하였고, 다이나믹레인지는 국제표준 ISO 14524[1]를 기반하여 투과식 OECF 차트를 이용한다.

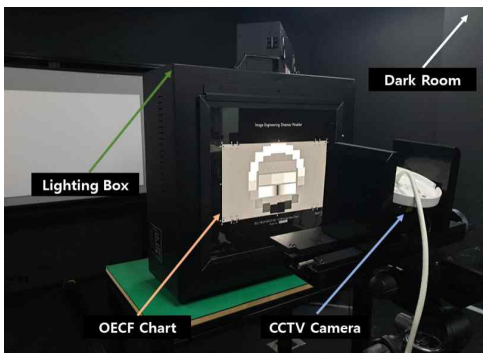


그림 2 다이나믹레인지 측정 방법

다이나믹레인지는 영상 품질 평가 소프트웨어인 IMATEST를 활용하여 Slope 기반의 다이나믹레인지를 비교한다. Lighting Box는 밝기 4,000 lx, 색온도 5,100 K의 조건에서 측정한다.

보행자 검출 알고리즘은 YOLO v3 모델을 활용했다. YOLO v3은 YOLO의 후속 버전인 YOLO v2의 아키텍처를 그대로 사용하였으나 아키텍처는 Darknet-19에서 Darknet-53으로 변경하고, FPN처럼 다양한 크기의 해상도 Feature Map을 이용해 Bounding box로 output을 나타내었다[2]. Darknet-53의 아키텍처는 아래 그림 3과 같다.

	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1	
	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2×	Convolutional	64	1 × 1	
	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
8×	Convolutional	128	1 × 1	
	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
8×	Convolutional	256	1 × 1	
	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
4×	Convolutional	512	1 × 1	
	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

그림 3 Darknet-53

보행자가 카메라로부터 약 15 m 거리를 왕복으로 걸었을 때의 클립 영상의 전체프레임 700 개를 추출하여 검출하였고, 영상의 차이를 확인하기 위해 보행자의 거리 약 15 m 위치에서 보행자 검출 Bounding Box의 위치를 화살표로 표시하였다.

그림 4 카메라 설정 별 검출 결과 영상

이러한 방법으로 1개의 클립당 700개의 단편영상으로 3가지 설정 영상 클립을 반복하여 보행자를 검출하여 전체 700개의 영상에서 검출한 영상의 개수를 구하여 검출률을 계산한다. 해당 시험에서는 1인 보행자 걸음으로 제한하는 방법으로 오검출의 요소를 제거하였기 때문에 오검출은 발생하지 않았다.

다이나믹레인지의 성능이 좋은 카메라는 객체 검출하기 좋은 환경의 카메라이며, 이는 난이도가 낮다고 할 수 있다. 동일 카메라의 설정을 변경하여 하드웨어 성능이 낮아지더라도 검출 성능이 유지되면 높은 성능이라고 할 수 있다.

〈표 1〉 다이나믹레인지에 따른 검출률 비교

설정 번호	Set #1	Set #2	Set #3
DR	29.5 dB	33.6 dB	48.4 dB
검출 영상 수	279 개	396 개	628 개
검출률	39.86%	56.57 %	89.71 %

위의 결과와 같이 카메라 하드웨어 성능지표인 다이나믹레인지에 따라 실환경에서의 보행자 검출률은 매우 차이가 났음을 확인할 수 있고, 객체 검출에 대해서 실환경 및 하드웨어 성능에 따라 다른 등급의 객체 검출 평가를 통해 객관적이고 정량적인 성능을 평가 할 수 있다.

Ⅲ. 결론

인공지능 지능형 CCTV의 발전에도 불구하고 인공지능 지능형 솔루션을 정량적으로 평가할 수 있는 방법이 부재하기 때문에 실환경에서 하드웨어 성능별 보행자 검출 결과를 도출한 결과 동일한 CCTV 카메라를 가지고 하드웨어 사양(설정값) 변경을 통해 동일한 객체 검출 알고리즘 성능이 최대 225% 차이가 발생하였다.

실 환경에서 카메라 하드웨어 성능에 따른 검출률 결과가 다르다는 연구결과를 제시하였으며, 인공지능 지능형 CCTV의 정량적인 성능 평가를 위해서는 하드웨어의 성능 별 솔루션을 평가해야하며, 본 논문에서는 다이나믹레인지를 하드웨어 성능지표로 설정하였고 다이나믹레인지의 성능별 검출률을 평가하여야 함을 제안하였다.

ACKNOWLEDGMENT

이 논문은 산업통상자원부 국가기술표준원에서 시행한 산업표준개발사업의 결과를 활용한 연구임

참 고 문 헌

[1] ISO 14524:2009 Photography – Electronic still-picture cameras – Methods for measuring opto-electronic conversion functions (OECFs).

[2] Redmon, Joseph, and Ali Farhadi. "Yolov3: An incremental improvement." arXiv preprint arXiv:1804.02767 (2018).

BLDC 팬모터의 모터 DC 전압을 이용한 고장 진단

심준석, 조현진, 박정환, 김호원*
부산대학교, 부산대학교, 부산대학교, *부산대학교
sugo312@pusan.ac.kr, wh77r77@pusan.ac.kr,
gg2059@pusan.ac.kr, *Howonkim@gmail.com

Fault Diagnosis Using DC Voltage Of BLDC Fan Motor

Jun Seok Shim, Hyun Jin Jo, Sang Hyun Lee, Ho Won Kim*
Pusan National Univ., Pusan National Univ., Pusan National Univ., *Pusan National Univ.

요 약

본 논문은 BLDC 팬모터의 DC 전압을 사용하여 오랜 사용으로 인해 결상된 BLDC 팬모터 또는 BLDC 팬모터 불량 제품의 고장을 진단하는 방법을 제안한다. 기존에는 모터의 진동 특성에 FFT 등 신호 분석 방법들을 모터의 진동 특성에 적용하여 모터의 고장을 진단하는 방법들을 제안하였다. 그러나 가전제품들 내부에 설치되는 공간적 특성상 광학 인코더 또는 가속도 측정 센서들을 부착하여 진동 특성을 추출하고, 추출된 진동 특성을 기반으로 하여 고장을 진단을 수행하기 어렵다. 따라서 본 논문에서 제안된 방법은 시계열 데이터를 처리하기 위한 모델인 BiLSTM을 BLDC 팬모터의 접지단에서 출력되는 모터의 DC 전압에 적용하여 고장을 진단하는 방법을 제안한다.

Ⅰ. 서 론

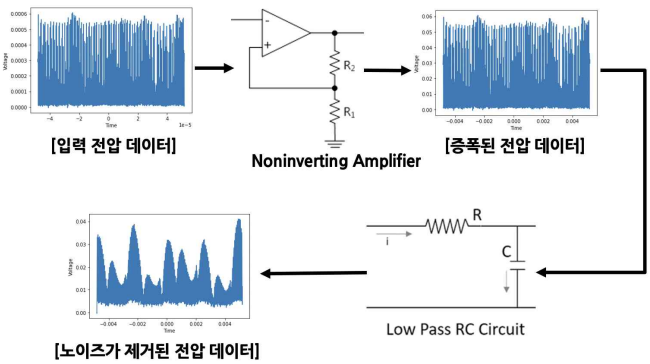
본 논문에서는 BLDC 팬모터의 DC 입력단 전압 정보만을 활용하여 모터의 고장 진단을 제안한다. 기존에는 모터에서 출력되는 펄스 기반의 고장 진단 로직 또는 모터의 진동 특성에 FFT와 같은 신호 분석 방법들[1]을 사용하여 모터의 고장 진단을 진행하였다. 그러나 기존의 펄스 기반의 모터 고장 진단의 경우 모터가 완전히 고장이 난 이후에야 모터의 고장을 진단할 수 있으며, 진동 특성 및 신호 분석 방법을 기반으로 한 모터의 고장 진단의 경우 가전제품 내에 광학 인코더 및 가속도 측정 센서를 부착하여야만 진동 특성을 추출할 수 있기 때문에 공간적인 제약이 있다.

따라서 본 논문은 BLDC 팬모터의 초기 고장 진단 및 가전제품 등에서 적용되는 공간적인 제약 등을 해결하기 위해 BLDC 팬모터의 접지단에서 출력되는 DC 전압을 사용하여 시계열 데이터를 처리하기 위한 모델인 BiLSTM 기반의 고장 진단 모델을 제안한다.

Ⅱ. 본론

Ⅱ-1. 데이터 전처리

본 논문에서 제안한 데이터 전처리 프로세스는 [그림 1]과 같다. BLDC 팬모터의 고장 진단을 위해 BLDC 팬모터의 접지단에서 출력되는 모터의 DC 전압을 증폭시키기 위해 비반전 증폭기가 사용된다. 비반전 증폭기는 입력 단자 간의 전위차(전압)보다 대개 백배에서 수 천배 큰 출력 전압을 생성하는 직류 연결형 고이득 전압 증폭기인 연산 증폭기(Op-amp)를 사용해 입력 전압의 부호와 일치하는 증폭된 출력 전압을 생성한다. 그 다음 데이터 전처리 단계로 생성된 전압의 노이즈를 제거하는 단계를 거친다. 본 논문에서 제안된 방식에는 노이즈를 제거하기 위해 저주파 통과 필터(Low Pass Filter, LPF)를 사용하였다.



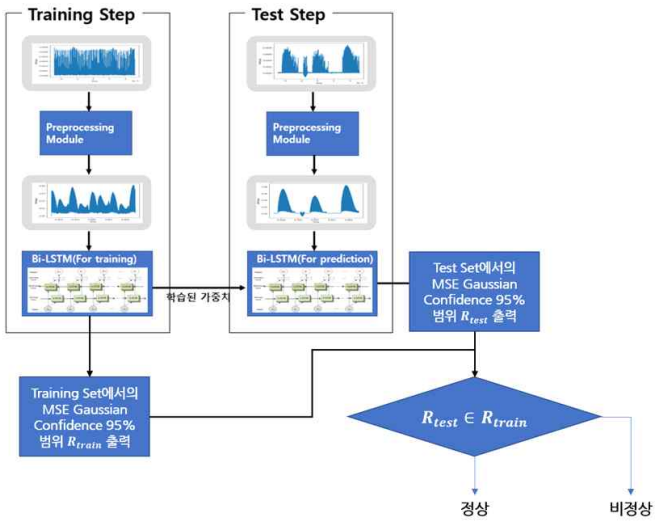
[그림 1] 데이터 전처리 프로세스

Ⅱ-2. 고장 진단 모델

BLDC 팬모터는 3상 인버터에 의해 구동되며 출력되는 DC 전압의 파형은 3상 전류가 합성된 파형이 그려진다. 따라서 본 논문은 BLDC 팬모터의 고장 진단을 위해 정현파의 고유한 특성인 주기성 및 시계열 데이터를 처리하기 위한 딥러닝 모델인 BiLSTM[3]을 활용하여 고장 진단을 진행하였다. BiLSTM은 시계열 데이터 처리를 위한 딥러닝 모델인 RNN의 문제점인 장기의존성을 개선한 모델이다. 해당 방법을 통하여 입력 시점인 (t) 시점에서 데이터의 기본 주기 (n)이후의 시점인 (t+n)시점에서의 예측값과 실제값 사이의 비교를 통해 고장 진단을 수행하였다.

[그림 2]는 데이터 전처리가 포함된 전체 고장 진단 모델을 나타낸다. 먼저 훈련부(Training Step)과 검증부(Test Step)으로 나뉘며 훈련부에서 정상 모터의 DC 접압에 대한 데이터만을 입력시켜 학습을 진행한다. 데이터 전처리부를 통해 노이즈 및 증폭 작업을 진행하고, 전처리된 데이터는 학습을 위해 BiLSTM에 입력된다. BiLSTM을 통해 정상 데이터에 대한

학습 가중치를 얻을 수 있으며, MSE 가우시안 분포 D_{train}의 정확도 95%의 신뢰구간인 R_{train}을 얻을 수 있다. D_{train}의 분포는 [그림 3]과 같다.



[그림 2] BiLSTM을 사용한 전체 이상탐지 모델

그 다음 검증부에서 훈련부와 같이 전처리를 진행하며 전처리된 데이터는 훈련부에서 학습된 가중치를 기반으로 예측값들을 출력한다. 출력된 예측값과 실제값을 비교하여 MSE 가우시안 분포 D_{test}의 정확도 95%의 신뢰구간인 R_{test}을 얻을 수 있다.

최종적으로 위에서 얻은 신뢰구간인 R_{train}과 R_{test}를 비교하여 R_{test}이 R_{train}에 포함되면 정상인 모터로, 포함되지 않는다면 고장난 모터로써 분류한다.

Ⅱ-3. 결과

본 논문에서는 검증을 위해 정상 데이터만 포함된 검증 데이터인 $Valid_{normal}$과 불량 모터의 데이터가 포함된 검증 데이터인 $Valid_{abnormal}$에서의 MSE 가우시안 분포인 D_{normal}, $D_{abnormal}$를 얻었다. 각 검증 데이터들의 분포인 D_{normal}, $D_{abnormal}$에서의 정확도 95% 신뢰 구간 R_{normal}, $R_{abnormal}$과 훈련부에서의 MSE 가우시안 분포 정확도 95% 신뢰구간인 R_{train}안에의 포함 여부를 통해 검증을 진행하였다.

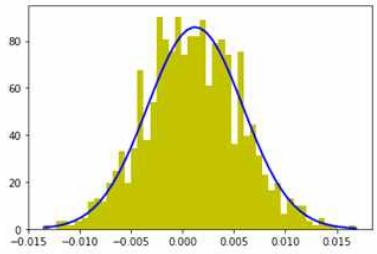
[그림 4]는 D_{train}과 D_{normal}를 비교한 그림이다. R_{normal}이 R_{train} 안에 포함되므로 $Valid_{normal}$은 정상적인 데이터임을 예측할 수 있으며, 실제 $Valid_{normal}$는 정상 모터의 데이터만이 포함된 데이터 셋이다.

[그림 5]는 D_{train}과 $D_{abnormal}$를 비교한 그림이다. $R_{abnormal}$이 R_{train} 안에 포함되지 않으므로 $Valid_{abnormal}$은 모터가 불량임을 예측할 수 있으며 실제 $Valid_{abnormal}$ 또한 불량 모터의 데이터가 포함된 데이터이므로 정확하게 정상, 불량의 모터를 분류해 낼 수 있었다.

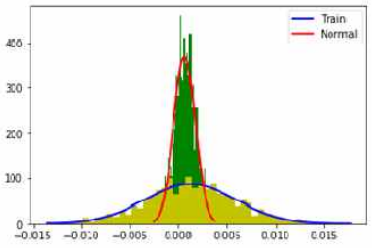
Ⅲ. 결론

본 논문에서는 BiLSTM을 사용해 BLDC 팬모터에서 나오는 DC 전압을 학습 및 테스트시킴으로써 시계열 데이터를 처리하기 위한 딥러닝 모델인 BiLSTM기반의 고장 진단 방식을 제안하였다.

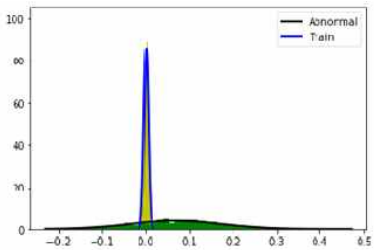
또한 기존의 펄스 기반 혹은 진동 특성 기반의 방식과 달리 BLDC 팬모터의 접지단에서 쉽게 추출 가능한 DC 전압과 MSE 가우시안 분포도를 통해 쉽게 고장 진단이 가능함을 실험을 통해 검증하였다.



[그림 3] 훈련부에서의 가우시안 분포 D_{train}



[그림 4] D_{train}와 D_{normal}을 비교한 그림



[그림 5] D_{train}와 $D_{abnormal}$을 비교한 그림

ACKNOWLEDGMENT

이 논문은 2020년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (2019-0-01343,융합보안핵심인재양성사업)

참 고 문 헌

[1] Zachar, Ryan, et al. "Utilizing spin-down transients for vibration-based diagnostics of resiliently mounted machines." IEEE Transactions on Instrumentation and Measurement 65.7 (2016): 1641-1650.

[2] M. Schuster, K. K. Paliwal, "Bidirectional recurrent Neural Networks," IEEE Transactions on Signal Processing, vol.45, no. 11, pp. 2673-2681, 1997.

[3] Flam, J. T., Chatterjee, S., Kansanen, K., & Ekman, T. (2011). Minimum mean square error estimation under gaussian mixture statistics. arXiv preprint arXiv:1108.3410.

연합학습을 통한 보안성 높은 무선통신 변조기술 탐지 모델에 관한 연구

서중하, *우태희, 박찬호, 강준혁
한국과학기술원, *충남대학교
junghaa.seo@kaist.ac.kr, seeles@o.cnu.ac.kr, kmapark@kaist.ac.kr, jhkang@kaist.edu

A Study on the Secured Radio Modulation Classification via Federated Learning

Seo Junghaa, *Woo Taehee, Park Chanho, Kang Joonhyuk
Korea Advanced Institute of Science and Technology, *Chungnam National University

요 약

본 논문은 각각의 수신기에서 수집된 데이터로 개별 학습된 모델의 가중치 파라미터를 중앙 서버로 보낸 후 글로벌 수신기 모델을 학습하는 과정의 반복을 통해 신호의 변조기술 탐지 성능을 향상시키는 연합학습 모델을 제안한다. 그리고 실험을 통해 개별 수신기가 악의적 사용자에 의해 오염되었을 때 글로벌 수신기 모델의 보안성을 확보할 수 있음을 확인하였다.

Ⅰ. 서 론

심층 신경망(DNN), 순환신경망(RNN) 등 다양한 기계학습 방법을 통해 인공지능은 여러 분야에서 인간의 인지능력을 뛰어넘는 성과를 보여주고 있다. 무선통신 분야에서도 인공지능을 통해 비선형적 채널 환경을 예측하고[1], 신호를 탐지하며[2], 기존 통신이론 기반의 분리된 소스코딩, 채널코딩을 하나의 시스템으로 연결하려는 End-to-end 오토인코더[3] 등이 심도있게 연구되고 있다. 기존의 기계학습 방법은 대용량의 데이터를 고성능 학습용 서버를 통해 처리/학습한다. 신호 변조기술 탐지 모델을 예로 들면, 수신기에서 수집된 신호 데이터는 학습용 서버로 수집, 학습된 탐지 모델을 다시 수신기로 보내게 된다. 이렇게 데이터가 이동하는 동안 네트워크 자원이 소모되며, 민감한 데이터의 경우 악의적 사용자가 이를 유출시키거나 학습 데이터를 오염시킬 수도 있다. 또한 서버에서 학습을 담당하기 때문에 학습 방법, 데이터의 양에 따라 많은 시간이 소요된다는 단점도 존재한다. 이를 해결하기 위해 연합학습(Federated Learning)[4] 개념이 등장하였다, 이 개념은 서버에 비해 성능이 다소 부족하지만 여러 대의 분산된 컴퓨팅 자원을 활용하여 각각 로컬 모델을 학습한 후 학습된 모델의 정보만을 중앙 서버에서 통합하여 글로벌 모델을 만드는 방법이다. 본 논문에서는 각각의 수신기에서 수집된 데이터로 개별 학습된 모델의 가중치 파라미터를 중앙 서버로 보낸 후 글로벌 수신기 모델을 학습하는 과정의 반복을 통해 신호의 변조기술 탐지 성능을 향상시키는 연합학습 모델을 제안한다, 그리고 실험을 통해 개별 수신기가 악의적 사용자에 의해 오염되었을 때 글로벌 수신기 모델의 보안성이 높음을 확인하였다.

Ⅱ. 본론

본 장에서는 연합학습 소개 및 연합학습을 기반으로 무선통신 변조기술을 탐지하는 수신기 모델을 제안하며, 글로벌 수신기 모델의 보안성 측정 실험을 통해 결과를 분석한다.

2-1 연합학습

일반적인 기계학습이 대용량의 학습용 데이터를 고성능 서버를 통해 학습하는 것과 달리 연합학습은 그림 1 과 같이 서버가 아닌 분산된 개별 컴퓨팅 자원에서 데이터를 처리하고 학습한다. 그리고 학습된 모델을 서버로 보내면(aggregation) 서버는 개별 컴퓨팅 자원에서 받은 모델을 글로벌 모델로 종합한 후 다시 개별 컴퓨팅 자원으로 보낸다(broadcast). 개별 컴퓨팅 자원에서는 서버와 연결되어 있지 않더라도 이전에 서버로부터 받은 글로벌 모델로 구동할 수 있으며, 학습 데이터가 아닌 모델만을 서버로 보내기 때문에 대용량의 네트워크를 유지할 필요가 없다. 또한 민감한 학습 데이터의 경우 내부에서만 처리되기 때문에 보안적 측면에서도 안정되어 있다. 서버에서는 회차별 개별 컴퓨팅 자원을 선정하고 이들로부터 받은 모델을 글로벌 모델로 종합을 하는 알고리즘을 통해 상대적으로 적은 데이터 처리만 하면 된다. 더욱 효율적인 네트워크 전송을 위해 가중치 파라미터를 압축하거나, 모델의 가중치 변화값을 전송, 손실함수의 경사도값을 전송하는 방법 등도 널리 연구되고 있다. 다음 장에서는 가중치 변화값 전송을 통해 효율적으로 네트워크 사용량을 줄인 연합학습 기반 무선통신 변조기술 탐지 모델을 제안한다.

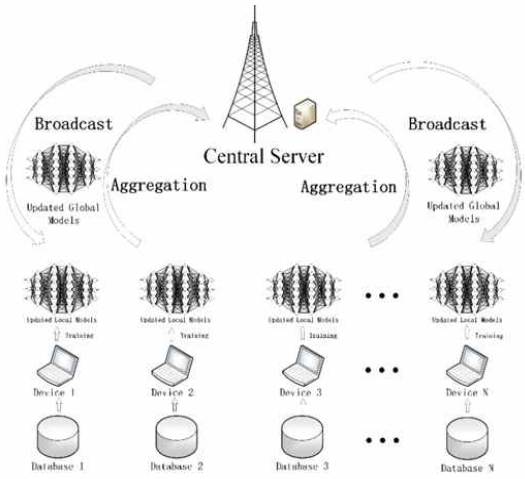


그림 1 연합학습 과정

2-2 연합학습 기반 무선통신 변조기술 탐지 모델

서버로부터 받은 $k-1$번째 글로벌 모델의 가중치 w_{k-1}를 기반으로 i번 수신기에서는 수신 신호 $r(t)$와 변조기술 y를 통해 수신기 모델 $f(w_k^i;)$ 를 학습하고, 가중치 변화값

$$\Delta w_k^i = w_k^i - w_{k-1},$$
$$\arg\min_{w_k^i} L(f(w_k^i; r(t)), y) \quad (1)$$

을 계산하여 서버로 보낸다. 이때, f는 DNN 기반 수신기 모델, L은 손실함수로, catrgorical cross-entropy를 사용한다. 서버는 개별 수신기로부터 받은 k번째 가중치 변화값들을 종합(Federated averaging)하여 글로벌 가중치 변화값

$$\Delta w_k = \frac{1}{N}\sum_{i=1}^{n} \Delta w_k^i \quad (2)$$

을 개별 수신기로 보내면 i번 개별 수신기는 다시 수신기 모델 $f(w_{k+1}^i;)$을 학습하고, 가중치 변화값

$$\Delta w_{k+1}^i = w_{k+1}^i - w_k^i + \Delta w_k \quad (3)$$

을 서버로 보내는 반복적인 과정을 진행한다.

실험에는 24개의 변조기술에 상응하는 무선신호[5]를 VGG[5] 구조로 학습되는 5개의 개별 수신기 모델을 사용하고 {1, 5}의 개별 수신기 반복학습 횟수(local epoch), {32, 64, 128}의 배치 사이즈로 구분하였다. 그림 2 는 학습 회차에 따라 글로벌 수신기 모델의 분류정확도를 측정한 것으로, 개별 수신기의 반복학습 횟수가 많을수록, 배치 사이즈가 작을수록 분류정확도가 높지만 학습시간도 길어진다.

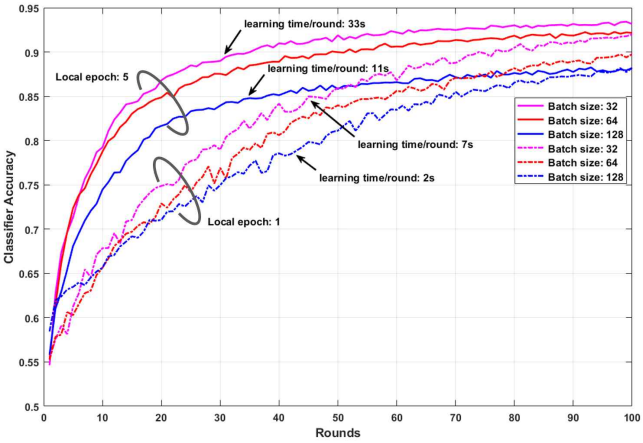


그림 2 연합 학습 회차별 글로벌 수신기 분류정확도 측정

2-3 위협 모델

연합학습은 서버가 개별 수신기의 학습 데이터를 볼 수 없고, 수신기를 통제하지 않기 때문에 악의적 사용자에 의해 오염되었을 때 학습 데이터의 무결성을 보장할 수 없다. 이 경우 오염된 데이터로 학습된 개별 수신기 모델이 글로벌 모델의 보안성에 영향을 미치는지를 확인하기 위해 개별 수신기의 학습데이터를 오염된 것(poisoning attack)으로 가정하였다.

실험에는 5개의 개별 수신기 모델이 1회의 반복학습 횟수와 32의 배치 사이즈로 학습되며, {11, 22, 33, 44}번째 회차에 하나의 수신기 모델의 학습 데이터를 무작위로 섞어 학습시킨 후 글로벌 수신기 모델의 분류정확도를 측정하였다. 그림 3 과 같이 연합학습 회차가 낮을 때는 개별 수신기의 오염이 글로벌 모델에 영향을 크게 미치지만, 연합학습 회차가 늘어날수록 개별 수신기의 오염도가 낮아지고, 이에 따라 글로벌 모델에 미치는 영향도 낮아진다.

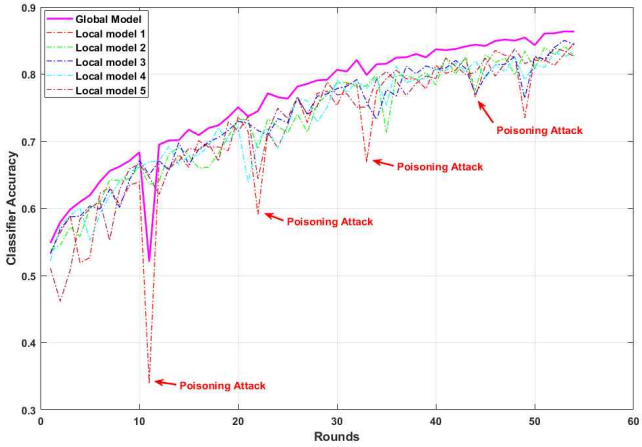


그림 3 개별 수신기의 학습데이터 오염시 분류정확도 측정

Ⅲ. 결론

본 논문은 기계학습의 방법 중 하나인 연합학습에 대해 소개하고 각각의 수신기에서 수집된 무선통신 신호로 개별 학습된 모델의 가중치 변화값을 서버로 보낸 후 글로벌 수신기 모델을 학습하는 과정의 반복을 통해 신호의 변조기술 탐지 성능을 향상시키는 연합학습 모델을 제안하고, 개별 수신기가 악의적 사용자에 의해 오염되었을 때 글로벌 수신기 모델의 보안성이 높음을 실험을 통해 확인하였다. 본 논문에서의 악의적 사용자에 의해 학습 신호가 오염된 것(포이즈닝 공격)을 가정하였으며, 추가적으로 악의적 사용자가 수신기 모델을 탈취하여 적대적 예제(adversarial example)나 백도어 공격을 하는 경우 글로벌 수신기 모델의 보안성에 대해서도 연구 중이다.

ACKNOWLEDGMENT

이 논문은 2020년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2018-0-00831, 이종 무선 네트워크를 위한 물리 계층 보안 기술 연구).

참 고 문 헌

[1] W. Lee, M. Kim, and D.-H. Cho, “Deep cooperative sensing: Cooperative spectrum sensing based on convolutional neural networks,” IEEE Transactions on Vehicular Technology, vol. 68, no. 3, pp. 3005 - 3009, 2019.

[2] H. Ye, G. Y. Li, and B.-H. Juang, “Power of deep learning for channel estimation and signal detection in ofdm systems,” IEEE Wireless Communications Letters, vol. 7, no. 1, pp. 114 - 117, 2017.

[3] T. O’Shea and J. Hoydis, “An introduction to deep learning for the physical layer,” IEEE Transactions on Cognitive communications and Networking, vol. 3, no. 4, pp. 563 - 575, 2017.

[4] Konečný, Jakub, et al. “Federated learning: Strategies for improving communication efficiency,” arXiv preprint arXiv:1610.05492, 2016.

[5] T. J. O’Shea, T. Roy, and T. C. Clancy, “Over-the-air deep learning based radio signal classification,” IEEE Journal of Selected Topics in Signal Processing, vol. 12, no. 1, pp. 168-179, 2018.

인지 통신에서 자동 변조 분류를 위한 CNN 모델 설계

김승환, 김동성*
금오공과대학교 ICT융합특성화연구센터
ksh001@kumoh.ac.kr, *dskim@kumoh.ac.kr

Design of Convolutional Neural Networks for Automatic Modulation Classification of Cognitive Radio

Seung-Hwan Kim, Dong-Seong Kim*
kit ICT Convergence Research Center.

요 약

본 논문은 인지 통신망에서 효과적인 자동 변조 분류를 위해 CNN 기술을 사용하여 변조 종류를 식별하는 모델을 제안하였다. 제안된 모델은 DEEPSIG DATASET: RADIOML 2018.01A 데이터셋을 사용하였으며 데이터셋은 24가지 종류의 변조 신호를 포함하고 있으며 수신된 신호는 클럭 오프셋, 채널 페이딩으로 인해 손상된 신호로 제안된 모델을 통해 분류 성능을 나타내었다. 제안된 모델은 최근 제안된 MCNet과 성능비교를 나타내었다.

Ⅰ. 서 론

무선통신 기술의 발달과 모바일 서비스의 증가로 제한적인 주파수 대역을 가지고 여러 가지 응용 분야에 대한 안정적인 서비스를 제공에 어려움을 가진다. 이러한 문제를 해결하고 주파수 대역의 사용효율을 높이기 위해 인지통신기술이 제안되었다[1]. 인지통신기술은 주 사용자의 채널사용에 간섭을 주지 않는 범위 내에서 2차 사용자가 채널을 사용하는 개념으로 이를 위해 채널 감지 및 채널 핸드오프 등의 기법에 연구가 진행되고 있다. 여기서 사용되는 채널은 다양한 사용자가 사용함으로 수신기에서 변조 방식에 대한 인지가 필요하다. 따라서, 무선 인지통신기술이 적용된 수신기는 변조기법을 자동적으로 선택하는 자동 변조 분류 기능을 포해야 한다. 자동 변조 분류의 개념은 전송되는 신호에 대한 사전지식 없이 자동적으로 신호의 변조 종류를 식별해내는 기능이다. 일반적으로 자동 변조 분류기능은 2가지 기법으로 구분할 수 있으며, 첫 번째로 최대 우도추정 기반의 분류 방식이 있고, 두 번째는 특징 기반의 패턴인식 방법이 있다. 최대 우도추정 방식은 베이지안 추론을 통해 최적의 값을 얻을 수 있는 장점을 가지만 높은 계산복잡도를 가지는 단점이 있다. 특징 기반의 패턴인식 방식은 입력 데이터 전처리, 특징추출, 마지막으로 분류 결정으로 진행된다. 본 논문에서는 DEEPSIG DATASET: RADIOML 2018.01A 데이터셋을 CNN (Convolutional Neural Network) 모델에 적용하여 변조 종류를 식별하는 방식을 제안하였다.

Ⅱ. 시스템 모델

본 논문에서 신호의 자동 변호 분류를 위해 사용된 DEEPSIG DATASET: RADIOML 2018.01A에서 사용된 프레임 수는 1,572,864이며 각 프레임은 1024 샘플 신호로 구성되며, 변조 타입마다 98,304의 프레임이 사용되었다. SNR 구간은 -10 dB에서 20 dB이며 2 dB의 간격을 가진다. 수집된 데이터는 클로오프셋과 Rayleigh 페이딩 채널을 통과한 손상된 신호이며 사용된 중심주파수는 900MHz의 ISM 대역을 사용하였다.

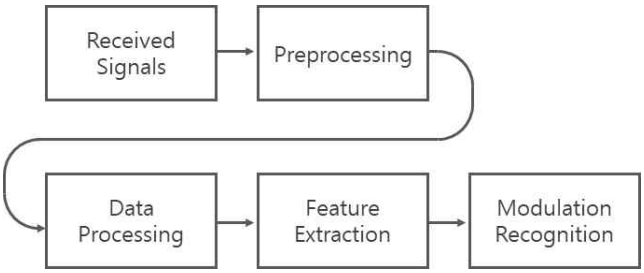


그림 1 Example of system model

수신기에 k번째 수신된 신호는 다음같이 표현할 수 있다. 채널 이득은 아래의 식으로 나타낼 수 있다.

$$y_k = \alpha e^{j(2\pi \triangle f t + \triangle \phi)} x_k + n. \quad (1)$$

여기서 α은 다중 경로의 진폭이며, $\triangle f$와 $\triangle \phi$는 캐리어 주파수와 위상의 오프셋, x_k는 변조신호 그리고 n는 백색가우시안잡음을 나타낸다. 수신된 신호는 데이터처리 부분에서 식2와 같이 RMS 정규화와 In-phase, Quadrature-phase 컴포넌트로 변환 뒤 딥러닝 기반 자동 변조 분류를 위해 전달된다.

$$\overline{s_i} = \frac{s_i}{\sqrt{\frac{1}{N}\sum_{i=1}^{N}|s_i|^2}}, \quad (2)$$

Ⅲ. 제안된 CNN 모델

일반적인 CNN 모델에서 하나의 합성곱 층에서 연산은 식3과 같이 표현할 수 있으며 스칼라 바이어스가 추가된다. 합성곱 층의 연산이 끝난 후 특징맵이 생성되며 이는 비선형 활성함수 층을 통과하게 된다.

$$conv_{x,y} = \sum_i w_i v_i + b. \quad (2)$$

여기서 w_i 합성곱 커널 가중치를 의미하고, v_i는 입력값을 의미한다. 제안하는 CNN은 SGDM (Stochastic Gradient Descent with Momentum) 최적화 기법을 사용하여 손실 값을 최적화하였으며, 여기서 사용되는 가중치(w)는 아래와 같이 나타낼 수 있다.

$$v_{t+1} := \alpha v_t - \eta \nabla Q(w_t), \quad (3)$$

$$where, Q(w_t) = \frac{1}{L}\sum_{t=1}^{L}(\hat{y_t} - y_t)^2$$

$$w_{t+1} := w_t + v_{t+1}, \quad (4)$$

여기서 v는 중간 가중치 값으로 각각의 반복 학습 때마다 사용되며 가중치(w)와 선형조합을 통해 갱신된다. α는 모멘트 계수이며, η는 학습률을 나타내고, $\hat{y}$는 추정값, 그리고 y는 참값을 나타낸다.

제안된 CNN 모델은 표1에서 요약하였으며, 각각의 컨볼루션층은 배치 정규화 층과 ReLU 활성화함수 층과 조합되어 있다. 입력층에서 입력 크기는 2x1024가지며 채널 수는 1이다. 제안된 모델의 ResNet 모델의 skip connection 기법을 적용하였으며 pointwise 합성곱 층을 적용하여 학습파라미터를 감소시켰다. 총 14개 컨볼루션층을 통해 특징맵을 추출하고 3개의 평균풀링층을 사용하여 학습되는 특징맵의 크기를 다운 샘플링 하였다. 여기서 FC (Fully Connection)층에 입력값이 전달되기 전 3x256 풀링 크기를 적용하여 128개의 채널만 연결되도록 하였다.

표 1 Configuration of CNN architecture

유형	출력크기	커널크기	파라미터 수
입력	2x1024	-	-
합성곱	6x1024x16	16x3x3	192
풀링	2x512x16	2x2	-
합성곱	6x512x32	2x2	-
풀링	3x256x32	2x2	-
블록1	3x256x32	32x1x1	1120
	3x256x8	8x3x1	792
	3x256x8	8x1x3	216
	3x256x32	32x1x1	352
블록2	3x256x64	64x1x1	2240
	3x256x16	16x3x1	3120
	3x256x16	16x1x3	816
	3x256x64	64x1x1	1216
블록3	3x256x128	128x1x1	8576
	3x256x32	32x3x1	12384
	3x256x128	128x1x1	4480
풀링	1x1x128	3x256	-
FC	24x128	-	3096
Softmax	24		

Ⅳ. 시뮬레이션 결과

본 논문에서 제안된 모델은 Matlab 2020b 프로그램을 통해 제안된 모델을 시뮬레이션 하였으며 결과는 그림 3, 4와 같이 비교결과를 Confusion Matrix 및 그래프로 나타내었다. 시뮬레이션을 위한 환경 설정은 최소 배치 크기 64, 최대 Epoch 45, 초기 학습비율 0.01 그리고 Epoch 40마다 0.01 학습 비율을 낮추도록 하였다. 제안된 모델에서 시험 데이터셋에 대한 정확도는 SNR 10 dB에서 91.48%로 최신 모델 MCNet[2]보다 5.42% 향상된 정학되를 보였다.

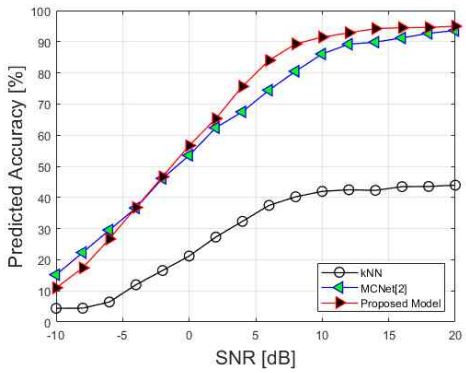


그림 2 The accuracy performance comparison

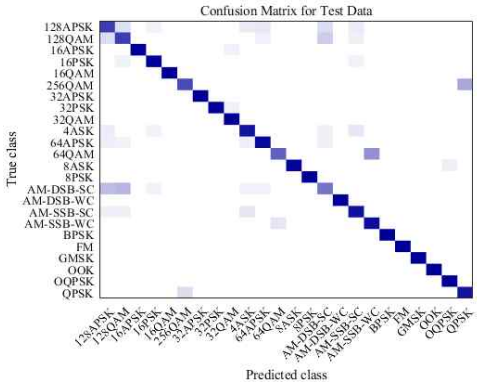


그림 3 Confusion matrix for 24 modulation types

Ⅴ. 결론

본 논문에서는 자동 변조 분류를 위해 CNN 기법을 사용하여 변조 종류를 식별하는 모델을 제안하였다. 제안된 모델의 성능은 24개의 변조 종류를 예측하는데 SNR (Signal-to-Noise Ratio) 10dB에서 91.48% 정확도를 가지며 최신 모델 MCNet보다 5.42% 높은 정확도 성능을 나타내었다.

ACKNOWLEDGMENT

이 논문은 정부(교육과학기술부)의 재원으로 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] T. Huynh-The, C. Hua, Q. Pham and D. Kim, "MCNet: An Efficient CNN Architecture for Robust Automatic Modulation Classification," IEEE Commun. Lett., vol. 4, no. 2, pp. 811-815, Api. 2020.

직렬 연결된 이미지 처리용 합성곱 신경망을 사용한 시간-주파수 채널 추정

김영찬, 장태준, 조준호
포항공과대학교

yckim321@postech.ac.kr, taejun8@postech.ac.kr, jcho@postech.ac.kr

Time-Frequency Channel Estimation Using Image Processing Convolutional Neural Networks

Youngchan Kim, Taejun Jang, Joon Ho Cho
POSTECH (Pohang University of Science and Technology)

요 약

최근 기계학습을 이용해 시간-주파수 채널 행렬을 추정하는 방법에 대한 연구가 다수 진행되었다. 본 논문은 이러한 채널 추정에 이미지 처리용 초해상도 합성곱 신경망과 잡음제거 합성곱 신경망을 연결하여 사용하는 기존 방법을 재검토하였다. 기존 연구에서는 초해상도 신경망과 잡음제거 신경망 모두를 적용 대상 채널 모델에 훈련하여 사용하는 방식을 제안하였다. 반면에 본 논문은 초해상도 신경망을 적용 대상 채널 모델에 훈련하는 과정을 생략하여도 기존 연구와 유사한 채널 추정 성능을 달성할 수 있음을 보였다. 즉, 잡음제거 신경망의 훈련만으로도 기존 성능에 근접함을 보임으로써 학습 시간을 크게 감소시킬 수 있음을 확인하였다.

Ⅰ. 서론

최근에 신경망을 이용한 기계학습이 크게 발달함에 따라 통신의 여러 분야에 기계학습을 적용하는 연구가 진행되었다. 채널 추정에도 기존의 통계적인 방법에서 탈피하여 기계학습을 이용하는 방법이 연구되었다 [1], [2]. 그 중에서 [1]에서는 이미지 처리용으로 개발된 기존의 신경망 2 개를 직렬 연결하여 채널 추정에 활용함으로써 인상적인 채널 추정 성능을 보여주었다.

[1]에서 사용하는 채널 추정 시스템은 파일럿 심볼을 이용하여 파일럿 위치에서의 채널을 추정한 후, 나머지 자원 요소(resource element)에서의 채널 추정값을 이미지 처리용 합성곱 신경망을 통해 구한다. 파일럿 위치만 추정된 시간-주파수 영역에서의 채널 행렬을 2 차원 이미지와 유사하게 취급하여 이미지 처리용으로 만들어진 합성곱 신경망을 사용한다. 이 과정에서 저해상도 이미지를 고해상도 이미지로 변환하는 초해상화(super resolution)를 위한 합성곱 신경망인 SRCNN 과 함께 이미지의 잡음 제거(denoising)를 위한 합성곱 신경망인 DnCNN 을 사용한다.

본 논문은 이 시스템에서 첫 번째 신경망인 SRCNN 을 일반적인 이미지에 대해서만 학습시켜도 목표 채널 모델의 수많은 채널에 대해 학습시킬 때와 유사한 성능을 보임을 확인하였다. 이를 통해서 학습 시간을 단축시킬 수 있다.

Ⅱ. 신호 및 시스템 모델

신호 모델은 다음과 같다. N_S 개의 부반송파와 N_D 개의 심볼이 있을 때, 시간-주파수 영역에서 채널을 통과한 수신 신호 행렬 $\mathbf{Y} \in \mathbb{C}^{N_S \times N_D}$의 i행 k열 원소는 다음과 같다.

$$Y_{i,k} = H_{i,k}X_{i,k} + W_{i,k}$$

여기에서 $\mathbf{H} \in \mathbb{C}^{N_S \times N_D}$는 채널 행렬, $\mathbf{X} \in \mathbb{C}^{N_S \times N_D}$는 송신 신호 행렬, $\mathbf{W} \in \mathbb{C}^{N_S \times N_D}$ 은 가산 백색 가우시안 잡음(additive white Gaussian noise) 신호를 뜻한다.

본 논문의 시스템에서는 파일럿 심볼을 이용한 채널 추정 기법을 사용한다. 파일럿 심볼의 위치가 정해져 있을 때, 시간 · 주파수 영역에서 파일럿 심볼 위치에 해당하는 채널의 값은 최소 제곱법(least squares)을 이용해 다음과 같이 구할 수 있다.

$$\hat{\mathbf{H}}_p^{LS} = \arg\min_{\mathbf{H}_p} \left\| \mathbf{Y}_p - \mathbf{H}_p \circ \mathbf{X}_p \right\|_F^2$$

여기에서 $\mathbf{Y}_p$, $\mathbf{H}_p$, $\mathbf{X}_p$는 각각 $\mathbf{Y}$, $\mathbf{H}$, $\mathbf{X}$에서 파일럿 심볼 위치를 제외한 원소의 값이 0 인 행렬을 나타내며 $\circ$는 아다마르 곱(Hadamard product) 연산자를 나타낸다. 이를 풀면 다음과 같은 식으로 각 파일럿 심볼의 추정값을 구할 수 있다.

$$\hat{H}_{i,k} = \frac{Y_{i,k}}{X_{i,k}}$$

[1]에서 제안한 채널 추정 시스템인 ChannelNet 은 파일럿 심볼을 이용해 최소 제곱법으로 파일럿 심볼 위치에서만 추정된 채널을 입력받아 모든 심볼에서 높은 정확도로 추정된 채널을 출력한다. 이러한 과정에서 그림 1 과 같이 2 개의 합성곱 신경망을 사용한다.

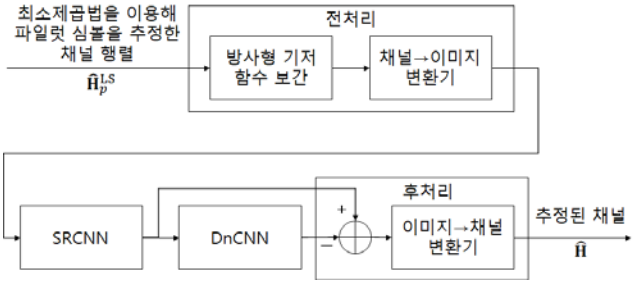


그림 1. 합성곱 신경망을 사용한 채널 추정 시스템 구조

첫 번째 합성곱 신경망인 SRCNN 은 본래 저해상도 이미지를 고해상도 이미지로 변환하기 위해 설계되었다[3]. 이 신경망은 보간된 저해상도 이미지를 입력받아 같은 크기의 고해상도 이미지를 출력한다.

그러나 우리는 이미지가 아닌 채널을 다루고 있으므로, 이를 이미지와 유사한 형식으로 변환하는 과정을 거쳐야 한다. 채널은 복소수 행렬로 표현할 수 있고 그레이스케일 이미지는 [0,1] 범위의 실수 행렬로 표현할 수 있기 때문에 전처리 과정으로 채널을 실수부와 허수부로 나누어 리스케일링하는 과정이 필요하다. 또한, [3]에서는 이미지를 보간하기 위해 쌍입방 보간법을 사용하였으나, 채널을 다룰 때에는 파일럿 심볼 위치가 격자점 모양으로 배치되지 않아도 처리할 수 있어야 하므로 보간 데이터 점 위치를 임의로 지정할 수 있도록 가우시안 커널을 사용한 방사형 기저 함수(radial basis function) 보간법을 사용한다.

두 번째 합성곱 신경망인 DnCNN 은 본래 잡음이 있는 이미지에서 잡음을 제거하기 위해 설계되었다[4]. 이 신경망은 잡음이 있는 이미지를 입력받아 최대한 잡음과 동일한 이미지를 출력한다. 따라서 신경망의 입력에서 출력을 빼면 잡음이 제거된 이미지를 구할 수 있다.

기존의 논문 [1]에서는 SRCNN 과 DnCNN 을 모두 적용하려는 채널 모델의 수많은 채널에 대해 학습시켰다. 그러나 본 논문에서는 SRCNN 은 일반적인 이미지에 대해 사전 학습된 것을 그대로 사용하고 DnCNN 만 채널에 대해 따로 학습하여도 시스템이 상당히 양호한 채널 추정 성능을 보임을 확인하였다.

Ⅲ. 시뮬레이션 결과

기본적인 시뮬레이션 매개 변수는 $N_S = 72$, $N_D = 14$를 사용하는 등 기존 논문[1]과 동일하게 설정하였다. 학습 및 성능 평가에 사용된 채널들은 채널 시뮬레이터를 이용해 생성되었으며, VehA 채널 모델과 SUI5 채널 모델의 두 가지 경우로 나뉘었다. 학습 데이터셋, 테스트 데이터셋은 각각 32000 개, 10000 개를 사용하였다. 또한 낮은 SNR 과 높은 SNR 의 경우로 나누어, 하나의 채널 모델 및 SRCNN 의 채널 추가 학습 여부에 따라 각각 DnCNN 을 2 개 학습하였다. 낮은 SNR 용 신경망은 SNR 12 dB 인 채널을 이용하여 학습하였고, 높은 SNR 용 신경망은 SNR 22 dB 에 맞추어 학습하였다.

그림 2 와 3 은 이렇게 학습시킨 시스템의 시뮬레이션 결과를 나타낸다.

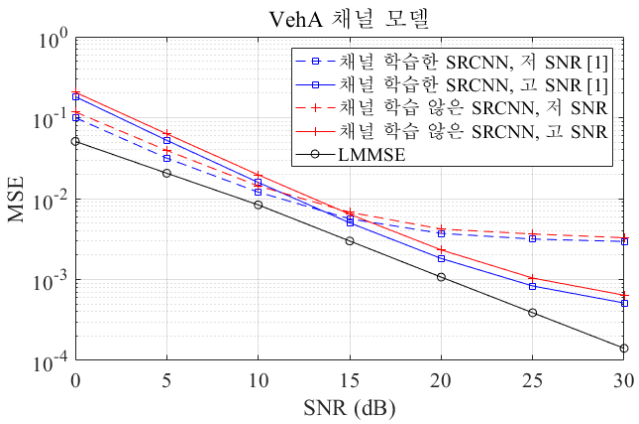


그림 2. VehA 채널 모델에서의 채널 추정값 MSE 비교

VehA 과 SUI5 채널 모두에서 낮은 SNR 을 목표로 학습한 신경망을 사용하는 경우, 채널에 대해 학습한 SRCNN 과 일반적인 이미지에 대해 학습한 SRCNN 의 MSE 차이가 1 dB 이내로 매우 작았다. 또한 높은 SNR 을 목표로 학습한 신경망을 사용하는 경우에도 마찬가지로 두 경우의 MSE 차이가 1.1 dB 이내로 매우 작은 것

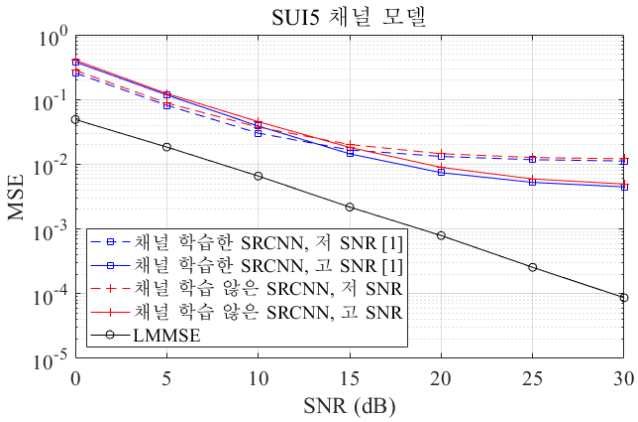


그림 3. SUI5 채널 모델에서의 채널 추정값 MSE 비교

을 확인할 수 있었다.

Ⅳ. 결론

본 논문에서는 채널 추정에 이미지 처리용 합성곱 신경망을 사용하는 기존 시스템을 재구현하고 분석하였다. 이를 통해 사용된 합성곱 신경망의 일부만을 채널에 대해 추가 학습시킴으로써 기존의 시스템과 유사한 채널 추정 성능을 보임으로써 학습에 필요한 시간을 크게 감소시킬 수 있음을 확인하였다.

ACKNOWLEDGMENT

이 논문은 부분적으로 2020 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구이며 [No. 2016-0-00123, 5G & Beyond 이동통신 시스템을 위한 정수 한정 MIMO 송/수신기 개발], 부분적으로 2020 년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 [No. 2018R1D1A1A02086169, QAM-FBMC: 6G & Beyond 이동통신 표준을 위한 새로운 파형 설계 연구].

참 고 문 헌

[1] M. Soltani, V. Pourahmadi, A. Mirzaei and H. Sheikhzadeh, "Deep learning-based channel estimation," in IEEE Communications Letters, vol. 23, no. 4, pp. 652-655, April 2019.

[2] H. He, C. Wen, S. Jin and G. Y. Li, "Deep learning-based channel estimation for beamspace mmWave massive MIMO systems," in IEEE Wireless Communications Letters, vol. 7, no. 5, pp. 852-855, Oct. 2018.

[3] C. Dong, C. C. Loy, K. He and X. Tang, "Image super-resolution using deep convolutional networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 2, pp. 295-307, 1 Feb. 2016.

[4] K. Zhang, W. Zuo, Y. Chen, D. Meng and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," in IEEE Transactions on Image Processing, vol. 26, no. 7, pp. 3142-3155, July 2017.

IoT 기반의 전력 모니터링 시스템에서 군집화와 군집 대표 부하를 활용한 무손실 데이터 압축 방법

이지훈, 윤승욱, 황의석*
광주과학기술원

{zoazoa61, ysw1207, *euiseokh}@gist.ac.kr

Lossless data compression based on clustering and cluster-representative load in IoT-enabled power monitoring system.

Lee Ji Hoon, Yoon Seung Wook, Hwang Eui Seok*
Gwangju Institute of Science and Technology (GIST)

요 약

본 논문은 사물 인터넷 (IoT, Internet of things) 환경에서 수집되는 전력 데이터의 압축을 목적으로 군집화 기술을 적용하여 군집 대표 부하를 활용한 무손실 압축 방법을 제안한다. 무손실 압축 방법은 원본 데이터의 손실 없이 복원할 수 있는 장점이 있어 중요한 정보의 보존을 위하여 활용되고 있다. 하지만, 개별 스마트 미터가 가지는 단일 데이터의 압축 방식은 압축률의 한계를 가지고 있어 이에 관한 효율적인 기술 개발이 중요하다. 제안된 압축 방식은 측정 장치로부터 수집된 전력 데이터의 프로파일 정보를 이용하기 위하여 군집화 기술을 적용하고 대표 및 비대표 부하로 나누어 유사 프로파일을 가진 데이터를 묶어 분류한다. 이후, 신호 간의 상관관계를 이용하여 선형 예측 부호화 기법을 적용함으로써 추정 신호를 생성한다. 이를 통해 원 신호와 추정 신호 간의 차이인 잔여 성분 신호를 생성하고 원 신호의 정보 엔트로피를 절감함으로써 단일 데이터의 압축 방식 대비 제안 방법을 적용하여 동일한 압축 알고리즘을 적용했을 경우 압축률이 향상됨을 확인하였다.

Ⅰ. 서 론

최근, 스마트 가전과 사물 인터넷 (IoT, Internet of things)의 발달과 함께 관련 기술이 우리 일상생활에 적용됨에 따라 다양한 혜택이 제공되고 있다. 이에 따라 IoT 관련 기술의 기반으로 사용되는 전력 분석의 중요성 역시 증가하고 있다.[1] 예로서, 전력 데이터에 관한 많은 연구 중 비접촉식 개별 가전기기 부하 식별 (Non-intrusive load monitoring) 기술과 함께 측정 장비의 기술 수준이 높아지고 이에 따라 수집 데이터의 측정 간격을 줄이며 높은 샘플링을 가진 공개 데이터 세트의 크기가 점차 증가하였다. 이에 따라, 현재 상용화된 하드웨어의 사양은 제한적이므로 증가하는 데이터 세트의 효율적인 관리 문제는 중요해지고 있다. 따라서 데이터 세트의 크기를 효과적으로 압축하여 저장하고 압축 처리 속도를 줄이는 것이 필요하므로, 압축 알고리즘의 성능 향상에 관한 필요성이 제기되고 있다.[2]

압축 알고리즘은 크게 손실 압축과 무손실 압축 기술로 구분된다. 손실 압축 기술은 기존 정보에서 무의미한 정보에 관한 데이터를 제거한 후 압축하는 방법으로서, 높은 압축률을 얻는 것이 유리하여 이미지와 음원 데이터에 사용되고 있다. 반면, 무손실 압축 기술은 압축률 및 압축 속도와 비교하면 손실 압축보다 성능이 낮으나, 원본 데이터의 손실 없이 복원할 수 있는 장점이 있어 주로 텍스트 기반의 데이터와 같이 중요한 정보를 포함하는 데이터에 적용된다. 또한, 무손실 압축 알고리즘은 원본 데이터의 크기에 따라 계산 복잡도를 증가시키기 때문에 빅데이터의 처리 시, 압축 시간과 메모리 처리량이 비례하여 증가하므로 무손실 압축 알고리즘의 성능 개선은 중요하다. 그러나 기존의 개별 스마트 미터가 가지는 단일 데이터 기반의 압축 방식은 압축률의 한계치를 가지고 있어 데이터의 그룹화를 통한 추가적인 압축률 향상에 관한 연구와 분석이 진행되고 있다.[3]

이를 위해, 본 논문에서는 전력 IoT 환경에서 수집되는 데이터 세트에 관하여 세부 그룹의 정보를 대표해서 개체 간 연관 관계를 설정하는 구조 [4]를 적용하고, 선형 예측 방법을 적용한 무손실 압축 알고리즘 [5]을 제안한다. 제안하는 방식은 스마트 미터 장치를 통해 수집되는 다양한 전력 데이터의 프로파일 정보를 이용하여 비지도 학습 기반 군집화 기술을 활용하고 대표 및 비대표 부하로 나누어 적용한다. 그리고 군집 내 대표 부하와 비대표 부하의 상관관계를 이용하여 선형 예측 모델 기반의 추정 신호를 생성한다. 이후, 원 신호와 추정 신호의 차이 값인 잔여 성분 신호를 저장하여 정보 엔트로피를 절감하고 기존 압축 알고리즘의 압축률 향상에 기여한다.

Ⅱ. 군집 대표 부하를 활용한 무손실 데이터 압축 방법

본 논문에서 제안하는 압축 알고리즘은 그림 1 과 같이, 수집된 데이터에 군집화 기법을 적용한 후, 유사한 프로파일을 가진 데이터를 묶어 분류한다. 이후 군집 내 대표 부하와 비대표 부하를 선정하기 위해 유클리드 거리 (Euclidean distance)를 기준으로 군집의 중심 (Centroid)과 가장 가까운 데이터를 대표 부하로, 나머지 신호는 비대표 부하로 선정한다. 대표 부하를 기준으로

동일 군집 내 각 비대표 부하에 관하여 선형 예측 부호화 (LPC, Linear prediction coding) 기법을 적용하고 각각의 추정 신호를 생성한다. 이후, 원 신호와 추정 신호의 차이 값인 잔여 성분 신호를 생성하고 선형 계수 (Linear coefficient)와 함께 비대표 부하의 신호를 복원할 수 있는 정보로서 저장한다. 저장된 두 정보의 완벽한 복원을 위하여 무손실 압축 알고리즘을 적용함으로써 추가적인 압축률 향상을 기대한다.

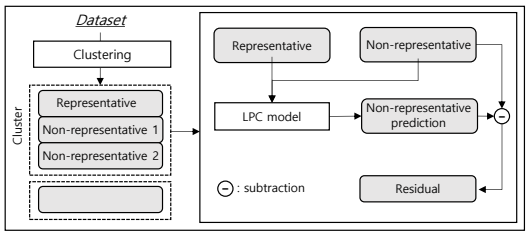


그림 1. 제안 방법의 구조도

Ⅲ. 실험 환경

제안 방법의 검증을 위한 실험 환경은 다음과 같다. 실험 데이터는 공공 데이터 세트 중 하나인 Pecan street [6]로 정하여 14 년도 9 개 주거형 데이터에 관한 1 분 단위 전력 데이터를 활용하여 실험을 진행하였다. 군집화 기법은 대표적으로 사용되는 K-평균 군집화 기법을 사용하였고 1 시간 단위 데이터의 중앙값을 기준으로 적용하였다. 최적 군집 수는 실루엣 기법 (Silhouette index)을 통해 4 개의 군집을 정하고, 군집의 중심 거리를 기준으로 가장 가까운 4526 번 집의 전력 정보를 대표 부하로, 나머지 4 개의 데이터를 비대표 부하로 선정하였다. 이후, 10 개의 선형 계수를 통해 LPC 기법을 적용하여 비대표 부하의 추정 신호를 생성하고 잔여 성분 신호를 구한 후 대표적인 무손실 압축 알고리즘인 LZMA, Bzip2, Gzip 등의 방법을 적용하여 압축을 진행하였다.[2]

Ⅵ. 실험 결과

제안 방법의 타당성 평가를 위해 먼저 (1)의 압축률을 적용하여 평가하고, 압축률과 정보 엔트로피 (2)의 상관관계를 확인하며 이론적으로 실험 결과를 분석한다.

$$Compressed\ ratio = \left(\frac{Compressed\ size}{Original\ size}\right) * 100 \quad (1)$$

$$H(x) = -\sum_x P(x)\log P(x) \quad (2)$$

제안 방법을 적용한 각 비대표 부하의 압축률 평가 결과는 그림2와 같이 전체적으로 향상되었으며, LZMA 알고리즘을 적용했을 경우 가장 향상된 결과를 확인하였다. 또한, 표1과 같이 압축률과 정보 엔트로피와의 상관성을 확인할 수 있어, 잔여 성분 기반 제안 방법의 압축 성능 향상 효과를 정보 이론적 지표로 확인할 수 있다.

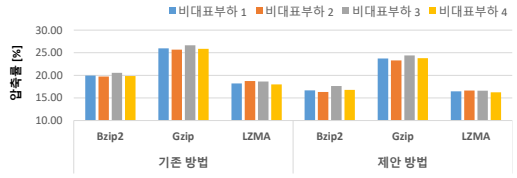


그림 2. 제안 방법에 관한 무손실 알고리즘 압축 평가

데이터	정보 엔트로피		압축률 [%]	
	원 신호	잔여 성분 신호	원 신호	잔여 성분 신호
비대표 부하 1	11.99	**11.07**	18.21	**16.45**
비대표 부하 2	12.94	**11.88**	18.74	**16.62**
비대표 부하 3	13.45	**13.43**	18.63	**16.59**
비대표 부하 4	11.07	**10.56**	17.98	**16.25**

표 1. 정보 엔트로피와 압축률의 시뮬레이션 결과

Ⅴ. 결론

본 논문에서는 IoT 환경에서 개별 스마트 미터가 가지는 단일 데이터 기반의 압축 한계를 극복하기 위하여 군집화 기법을 적용하고 군집 대표 부하를 활용하는 무손실 압축 방법을 제안한다. 제안된 방법은 동일 군집 내 유사 프로파일을 묶어 분류한 후, 대표 및 비대표 부하의 상관관계를 이용하여 생성한 잔여 성분 신호를 압축함으로써 신호가 보유한 정보 엔트로피를 절감한다. 이를 통해 기존 단일 데이터에 관한 무손실 압축 알고리즘을 적용했을 경우 대비 제안하는 방법 적용 시 압축률이 향상됨을 실험을 통해 확인하였다.

ACKNOWLEDGMENT

이 논문은 2020 년도 정부(과학기술정보통신부)의 재원으로 정보통신산업진흥원의 지원을 받아 수행된 에너지 AI 융합 연구개발 사업과 광주과학기술원의 GRI(GIST 연구원) 사업의 지원을 받아 수행된 연구임.

참 고 문 헌

[1] Miraz M. “A review on internet of things (IoT), internet of everything (IoE) and internet of nano things (IoNT),” IEEE ITA, pp. 219-224, 2015.

[2] Kriechbaumer T. “Waveform signal entropy and compression study of whole-building energy datasets,” ACM ICFES, pp. 1-9, 2019.

[3] Wang S. “A novel smart meter data compression method via stacked convolutional sparse auto-encoder,” Electrical Power & Energy Systems, p. 105761, 2020.

[4] Asimakopoulou G. “Leader-follower strategies for energy management of multi-microgrids,” IEEE Trans. Smart Grid, pp. 1909-1916, 2013.

[5] Wu J. “Lossless compression of hyperspectral imagery via clustered differential pulse code modulation with removal of local spectral outliers,” IEEE Signal Processing Letters, pp. 2194-2198, 2015.

[6] Pecan Street. “Dataport: the world's largest energy data resource,” Pecan Street Inc, 2015.

3GPP 실내환경에서 AI 기반 위치 추적 성능 개선 방안

오성현, 김정곤*
한국산업기술대학교 전자공학부
osh119@kpu.ac.kr, jgkim@kpu.ac.kr*

AI based Location Tracking in 3GPP Indoor Environment

Sung Hyun Oh, Jeong Gon Kim*
Dept. of Electronic Engineering, Korea Polytechnic University

요 약

이동통신 기술의 급속한 발달과 함께 GPS(Global Positioning System) 기술도 성장하고 있다. GPS 기술이 발달함에 따라 야외 환경에서 위치 측위 기술은 매우 높은 정확도를 갖게 되었다. 최근 실내 환경에서도 위치 측위 기술의 적용이 주목받고 있다. 하지만 GPS 기술은 복잡한 실내 환경에 존재하는 전파 손실 문제로 적용에 많은 한계가 존재한다. 따라서 본 논문에서는 WiFi(Wireless Fidelty) 통신을 사용하는 복잡한 실내 환경에서 사용자 위치 측위 기술에 관해 연구한다. 먼저 사용자 위치 측위를 위해 오프라인 단계에서 각 샘플 포인트를 배치하고 각 샘플 포인트에 대한 RSSI(Received Signal Strength Indicator) 값을 측정하여 핑거프린팅 데이터베이스를 구축한다. 그 후 실제 온라인 단계에서는 실제 사용자의 위치에 대한 RSSI 값을 측정하여 해당 값과 사전에 구축한 데이터베이스의 값으로 퍼지 매칭을 수행한다. 다음으로 PSO(Particle Swarm Optimization) 알고리즘 적용하여 실내 환경에서 위치 측위 정확도 향상 되었다. 시뮬레이션 통해 AI 기반 PSO 알고리즘 적용할 경우 예전대비 평균위치오차가 많이 감소해서 개선된 위치 측위 정확도를 얻게됨을 최종 확인하였다.

Ⅰ. 서론

오늘날 네트워크, 통신, GPS(Global Positioning System) 및 무선 센터 네트워크 등의 기술이 발달할수록 위치 기반 서비스의 중요성이 더욱 커진다. 이에 따라 위치 측위 기술이 출현하게 되고 실외 환경에서는 GPS를 기반으로 하여 높은 정확도의 위치 측위를 얻을 수 있게 되었다. 하지만 실내에서는 신호 손실로 인해 GPS를 적용하기에 한계가 존재한다. 이로 인해 실내 환경에서의 위치 측위 기술은 현재 중요한 연구 대상이 되었다.

위치 측위 기술의 적용 예로는 크고 복잡한 쇼핑몰에서 개인이 원하는 특정한 상점을 찾거나, 이러한 건물에서 화재가 발생하는 경우 소방관을 지원하는 것, 그리고 개인적인 목적이나 공공의 목적에 따라 다양한 실내 위치 측위 기술들이 연구되고 있다.[1] 현재 보편적으로 사용되어지는 실내 위치측위 기술로는 Bluetooth, UWB(Ultra Wide Band) 그리고 Wi-Fi(Wireless-Fidelity) 등이 있다. 그리고 기존의 센서 측위 기술에는 범위를 기반으로 하는 것과 범위를 사용하지 않는 방법이 사용되는데, 보편적으로는 범위를 기반으로 하는 기술을 사용한다. 그 중에서도 RSSI(Received Signal Strength Indicator)를 기반으로 하는 기술은 측위 정밀도가 가장 높으면서 비용도 저렴하기 때문에 범위 기술 중에서 가장 많이 사용된다.[2]

따라서 본 논문에서는 위치 측위 기술로는 Wi-Fi를 사용하고, 신호 세기 측정은 RSSI를 기반으로 한다. 또한 측위에 사용되는 기법은 핑거프린팅, 퍼지매칭, PSO(Particle Swarm Optimization) 알고리즘을 사용한다.

본 논문은 다음과 같이 구성되어 있다. 2절에서는 사용자 위치 측위를 위한 구체적인 방법을 제시하고 시뮬레이션을 수행한다. 3절에서는 본 논문의 결론을 맺는다.

Ⅱ. 본론

본 논문에서는 위치 측위를 위해 오프라인 단계에서 핑거프린팅 기법을 사용하여 데이터베이스를 구축하고, 퍼지매칭을 수행하여 얻은 추정된 사용자 위치와 가장 근접한 4개의 샘플 포인트를 선택한다. 그 후 선택된 4개의 샘플 포인트를 선으로 이어 제한된 영역을 설정하고, 제한된 영역 내에서 PSO 알고리즘을 사용하여 사용자 위치 측위를 완료한다.

먼저 핑거프린팅 기법은 총 B개의 Wi-Fi AP(Access Point)와 S개의 샘플 포인트를 사용한다고 가정한다. 모든 샘플 포인트는 각 AP에 대한 RSSI 값을 측정한다. 그 후 측정된 RSSI 값을 기반으로 핑거프린팅 데이터베이스를 구축한다. 구축이 완료된 핑거프린팅 데이터베이스 F_{DB}는 아래 식과 같이 RSSI 행렬로 표현할 수 있다.[3]

$$F_{DB}=\begin{Bmatrix} RSSI_1^1 & \cdots & RSSI_s^1 & \cdots & RSSI_S^1 \\ \vdots & & \vdots & & \vdots \\ RSSI_1^b & \cdots & RSSI_s^b & \cdots & RSSI_S^b \\ \vdots & & \vdots & & \vdots \\ RSSI_1^B & \cdots & RSSI_s^B & \cdots & RSSI_S^B \end{Bmatrix}_{B\times S} \quad (1)$$

여기서, $RSSI_s^b$는 b번째 AP와 s번째 샘플 포인트 사이의 RSSI 값을 나타낸다.

퍼지 매칭을 수행하기 위해 각 AP와 각 샘플 포인트 사이의 벡터 값을 계산한다. 계산된 벡터 값은 핑거프린팅 데이터베이스와 동일하게 행렬 형태로 나타낼 수 있으며, V라고 정의한다.

$$V=\begin{Bmatrix} d_1^1 & \cdots & d_s^1 & \cdots & d_S^1 \\ \vdots & & \vdots & & \vdots \\ d_1^b & \cdots & d_s^b & \cdots & d_S^b \\ \vdots & & \vdots & & \vdots \\ d_1^B & \cdots & d_s^B & \cdots & d_S^B \end{Bmatrix}_{B\times S} \quad (2)$$

여기서, d_s^b는 b번째 AP와 s번째 샘플 포인트 사이의 벡터 값을 의미한다. 그 후 온라인 단계에서는 실제 사용자 UE(User Equipment) u의 위치

* : 교신저자

에서 RSSI를 측정하며, 측정된 값은 아래와 같이 나타낼 수 있다.

$$H^{b}_{RSSI} = [RSSI^{1}_{u}, RSSI^{2}_{u}, ..., RSSI^{B}_{u}] \quad (3)$$

여기서, $RSSI^{b}_{u}$는 b번째 AP와 사용자 UE u 사이의 RSSI 값이다.

위에서 측정된 사용자 위치의 RSSI 값은 오프라인 단계에서 구축된 핑거프린팅 데이터베이스의 값과 퍼지 매칭을 수행한다. 이때 F_{DB}와 H^{b}_{RSSI}의 상관관계를 평가하면 유클리드 거리를 얻을 수 있다. b번째 AP에 대해 오프라인 단계에서 s번째 샘플 포인트의 핑거프린팅 데이터와 온라인 단계에서 UE u의 RSSI 값의 상관관계는 $\alpha^{b}_{u,s}$로 주어지며, 여기서 $0 \le \alpha^{b}_{u,s} \le 1$ 이다. 따라서 최소 거리는 아래와 같이 정의된다.

$$d_{u,s} = \| p_{u,b} - s \| = \sqrt{\sum_{b=1}^{B} (\alpha^{b}_{u,s} - 1)^2} \quad (4)$$

위 식의 결과 $d_u = [d_{u,1}, d_{u,2}, ..., d_{u,s}]$이며, 이는 유클리드 거리 벡터를 나타낸다. 유클리드 거리 벡터를 기반으로 UE u와 가장 인접한 샘플 포인트를 선택할 수 있다. 선택된 샘플 포인트를 연결하여 초기 PSO 알고리즘의 제한된 영역을 얻을 수 있다.

PSO는 지능형 진화 계산 알고리즘 중 하나이며, 다음과 같은 순서로 진행된다. 초기에 모든 입자의 초기화가 진행된다. 초기화된 입자는 탐색 공간 내에 랜덤하게 배치되며, 배치된 입자들은 최적의 해를 찾기 위해 탐색을 시작한다. 탐색을 진행하면서 각 입자들은 자신의 최적 위치인 $pbest$와 군집의 최적 위치인 $gbest$를 서로 공유할 수 있다. 입자들은 앞서 언급한 두 가지의 값을 기반으로 탐색을 진행한다. 알고리즘은 여러 번의 반복 중 최대 반복횟수를 달성한 경우와 목표 정확도를 달성한 경우에 종료된다. 아래는 입자의 다음 속도, 다음 위치 및 관성계수에 관한 수식이다.

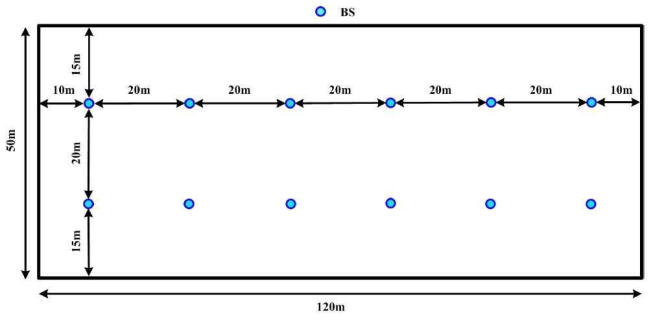
$$V_i(t+1) = w\, V_i(t) + \alpha\, \gamma\, [pbest_i(t) - X_i(t)] + \alpha\, \gamma\, [gbest(t) - X_i(t)] \quad (5)$$

$$X_i(t+1) = X_i(t) + V_i(t+1) \quad (6)$$

$$w = w_{\max} - \frac{t\,(w_{\max} - w_{min})}{T} \quad (7)$$

여기서, α는 가속계수, γ는 수축계수, ω는 관성계수, t는 현재 반복횟수, T는 최대 반복 횟수를 나타낸다.

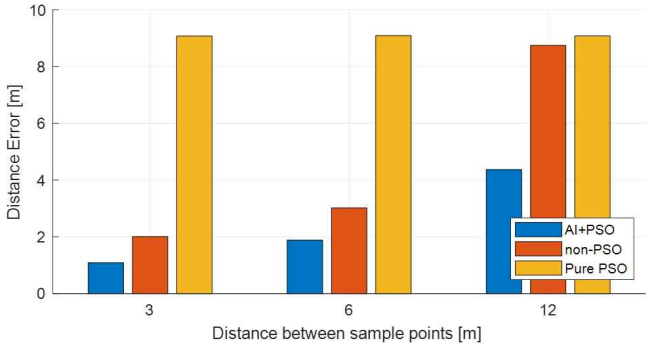
본 논문에서 시뮬레이션을 진행한 시스템 모델은 [그림 1]과 같다. 이는 3GPP에서 제안한 실내 오피스 환경이다.[4] 해당 그림에서 볼 수 있듯이 $120m \times 50m$의 비어있는 공간을 가정하였고, BS는 총 12개를 배치하였다. 시뮬레이션에 사용된 중요한 변수는 [표 1]에 정리하였다.



〔그림 1〕 3GPP 실내오피스 환경 정의

〔표 1〕 실험 파라미터

파라미터	값
방 크기	120mX50m
AP의 개수	12
샘플 포인트 간 거리	3,6,12〔m〕
입자의 개수	10
AP의 전력	20W
α, γ, $w_{\max}$, w_{min}, T	2, 0.3, 1, 0.4, 10



〔그림 2〕 샘플간격 따른 측위오차 실험결과

시뮬레이션 결과는 [그림 2]에 나타내었으며, 여기서 샘플 포인트 간 거리가 변화할 때에 따른 위치 정확도를 비교하였다. 결과 그래프에서 볼 수 있듯이 샘플 포인트 간 거리가 멀어질수록 위치 측위 오차는 더욱 커지는 것을 볼 수 있다. 하지만 제안하는 PSO 기반의 위치 추정 방식을 사용할 경우 기존의 방식보다 더 높은 측위 정확도를 달성할 수 있음을 최종적으로 확인할 수 있었다.

Ⅲ. 결론

본 논문에서는 3GPP에서 제안한 실내 오피스 환경에서 Wi-Fi를 기반으로 핑거프린팅, 퍼지매칭, PSO 알고리즘을 사용하여 사용자의 위치를 추적하는 방법을 제안하였다. 이는 오프라인 단계에서 구축하는 데이터베이스의 크기에 따라 정확도가 달라지며, 온라인 단계에서는 PSO 알고리즘에서 입자의 개수를 몇 개로 설정하느냐에 따라 정확도가 상이하게 된다. 추후 위와 같은 변수를 다르게 설정하여 최적의 샘플 포인트 개수 및 입자의 개수를 설정하는 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] Hu, J., and Wang, H. “WIFI indoor positioning algorithm based on improved Kalman filtering,” Proceedings of the 2016 International Conference on Intelligence Transportation, Big Data & Smart City(ICITBS), pp.349-352, December 2016.

[4] “Study on channel model for frequencies from 0.5 to 100 GHz (Release14)”, 3GPP TR 38.901

Metric Learning 기반 Adversarial Example 탐지 가능성에 대한 연구

최석환, 신진명, 김정구, 최윤호*
부산대학교, *부산대학교
daniailsh@pusan.ac.kr, sinryang@pusan.ac.kr, kimjg@pusan.ac.kr, *yhchoi@pusan.ac.kr

A Study on possibility of detection for adversarial examples based on metric learning

Seok-Hwan Choi, Jinmyeong Shin, Jeong Goo Kim, Yoon-Ho Choi*
Pusan National Univ., *Pusan National Univ.

요 약

딥러닝 모델은 다양한 분야에서 안정적인 성능을 보이지만 입력 이미지에 특정 노이즈를 추가하여 딥러닝 모델의 분류 정확도 감소를 유발하는 Adversarial Example에 매우 취약하다. 이러한 Adversarial Example을 방어하기 위한 기존 연구는 세 범주로 분류할 수 있다. (1) 모델 재학습 기반 방법; (2) 입력 변환 기반 방법; (3) Adversarial Example 탐지 방법. 하지만, Adversarial Example 생성 기법의 발전과 함께 발전하는 모델 재학습 기반 및 입력 변환 기반 방법과 달리 Adversarial Example 탐지 방법은 여전히 이진 분류 방법에 머물러 있다. 본 논문에서는 Metric Learning을 기반으로 한 다중 클래스 Adversarial Example 탐지 기법을 제안한다.

Ⅰ. 서 론

딥러닝 모델은 다양한 분야에서 활용되고 있으며, 자율 주행 자동차 및 맬웨어 분류와 같은 보안에 민감한 영역에서도 활용되고 있다. 하지만, 딥러닝 모델 활용의 증가와 함께 많은 보안 이슈도 등장하고 있으며, 그 중에서 입력에 사람이 인식 할 수 없는 노이즈를 추가하는 Adversarial Example은 딥러닝 모델의 분류 정확도 감소와 같은 심각한 문제를 야기할 수 있다[1]. 이러한 Adversarial Example을 방어하기 위해 많은 방어 방법이 제안되었으며 주로 세 가지 범주로 분류할 수 있다. (1) 모델 재학습 기반 방법[2]; (2) 입력 변환 기반 방법[3]; (3) Adversarial Example 탐지 방법[4]. 모델 재학습 기반 방법은 딥러닝 모델을 재학습하거나 새로운 모델로 학습하여 Adversarial Example을 방어할 수 있으며, 입력 변환 기반 방법은 Adversarial Example을 딥러닝 모델에 공급하기 전에 노이즈 제거 기법을 적용하여 Adversarial Example을 방어할 수 있다. 이러한 두 가지 방법은 Adversarial Example 생성 기법의 발전과 함께 발전하였다. 반면에, Adversarial Example 자체를 탐지하는 Adversarial Example 탐지 방법은 여전히 이진 분류 방법에 머물러 있다.

따라서 본 논문에서는 Adversarial Example 탐지 방법의 발전을 위해 Metric Learning을 기반으로 한 다중 클래스 Adversarial Example 탐지 기법을 제안한다.

Ⅱ. 본론

본 논문에서는 다중 클래스 Adversarial Example 탐지를 위해 Convolution Neural Network (CNN) 기반의 유클리드 임베딩 기술을 적용하였으며, 이를 그림 1에서 도시화 하였다. 구체적으로, 제안하는 방법은 정상 입력 및 Adversarial Example을 유클리드 공간에 맵핑하기 위해 학습 데이터셋 생성, Metric Learning 모델 학습, Advesrarial Example 탐지의 3 단계를 거쳐 동작한다.

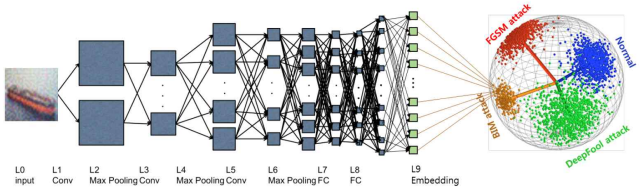


그림 1 제안하는 기법 개요도

2.1 학습 데이터셋 생성

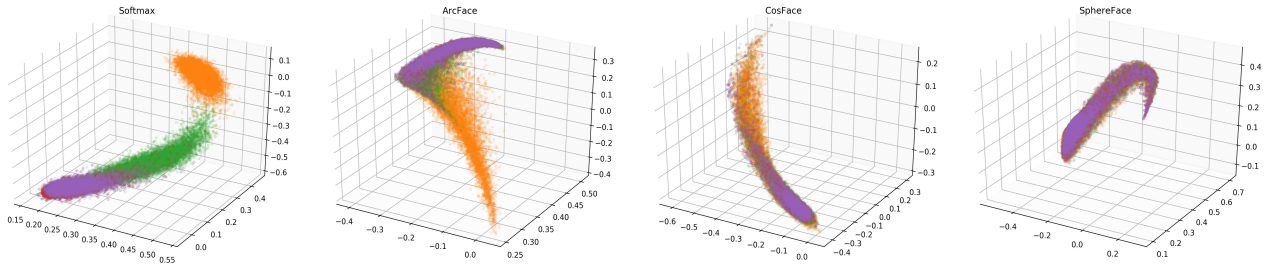
학습 데이터셋 생성 단계에서는 대상 모델에 대해 다양한 Adversarial Example 생성 기법을 적용하여 Metric Learning 모델 학습에 필요한 Adversarial Example을 생성한다. 본 논문에서는 대표적인 Adversarial Example 생성 방법인 Fast Gradient Sign Method (FGSM)[5], Basic Iterative Method (BIM)[6], DeepFool[7], C&W's Method[8]을 이용하여 Metric Learning을 위한 학습 데이터셋을 생성하였다.

2.2 Metric Learning 모델 학습

Metric Learning 모델 학습 단계에서는 원본 학습 데이터셋과 Adversarial Example 데이터셋을 사용하여 CNN 기반 Metric Learning 모델을 학습하였다. 본 논문에서는 대표적인 CNN 모델인 ResNet20을 사용하였다. 또한 본 논문에서는 Metric Learning 모델에서 주로 사용되는 Softmax, SphereFace[9], CosFace[10], ArcFace[11]의 손실함수를 이용하여 4개의 Metric Learning 모델을 학습하였다.

2.3 Adversarial Example 탐지

Adversarial Example 탐지 단계에서는 학습된 Metric Learning 모델을 이용하여 Adversarial Example을 탐지한다. 학습된 Metric Learning 모델은 입력 이미지를 유클리드 공간 상에 맵핑하므로 학습 데이터셋에



a Softmax 손실 함수 b ArcFace 손실 함수 c CosFace 손실 함수 d SphereFace 손실 함수

그림 2 다양한 손실 함수를 사용한 제안하는 기법의 유클리드 공간 상 맵핑 결과

포함된 Adversarial Example 뿐만 아니라 새로운 유형의 Adversarial Example도 탐지 할 수 있다. Metric Learning 모델 이 후 단계에서는 일반적인 분류 및 클러스터링 알고리즘을 적용할 수 있다. 따라서, 본 논문에서는 Metric Learning 모델을 통한 유클리드 공간 상에 맵핑된 결과만을 다룬다.

2.4 실험 및 검증

제안하는 방법의 성능을 평가하기 위해 본 논문에서는 CIFAR-10 이미지 분류 데이터셋에 대해 다양한 Metric Learning 손실 함수를 이용하여 실험하였다. 구체적으로, CIFAR-10 데이터셋의 전체 학습 데이터셋을 이용하여 대상 모델을 학습하였으며, 전체 테스트 데이터셋을 이용하여 Adversarial Example 생성 및 Metric Learning 모델을 학습하였다. 또한 대상 모델 및 Metric Learning 모델로는 ResNet20을 사용하였으며 Metric Learning의 출력 차원은 10차원으로 고정하였다.

그림 3은 다양한 손실 함수를 사용한 제안하는 방법의 결과를 보여준다. Softmax 손실 함수를 사용하여 Metric Learning 모델을 학습한 경우 정상 입력과 4개의 Adversarial Example 생성 기법을 유클리드 공간 상에 효율적으로 맵핑하는 것을 확인할 수 있다. 하지만, SphereFace, CosFace, ArcFace 손실 함수를 사용하여 Metric Learning 모델을 학습한 경우에는 정상 입력 및 4개의 Adversarial Example 생성 기법을 효율적으로 맵핑하지 못하였다. 이는 SphereFace, CosFace, ArcFace 손실 함수가 학습 시에 Adversarial Example 생성 기법의 특징을 반영하지 못한다는 것을 나타낸다.

Ⅲ. 결론

본 논문에서는 Metric Learning을 기반으로 한 다중 클래스 Adversarial Example 탐지 방법을 제안하였다. 다양한 손실함수를 이용한 실험을 통해 Adversarial Example의 유형을 분류 할 수 있는 가능성을 확인하였다. 하지만, 제안하는 방법은 DeepFool 및 C&W's Method과 같은 유사한 속성을 갖는 Adversarial Example에 대해 낮은 분류 성능을 보였다. 따라서, 향후 연구에서는 데이터 전처리 및 차원 감소 등의 방법을 적용하여 대부분의 Adversarial Example을 효율적인 분류할 수 있는 방안에 대한 연구가 수행되어야 할 것이다.

ACKNOWLEDGMENT

본 연구는 한국연구재단 논문연구과제 (NRF-2018R1D1A3B07043392) 지원 및 BK21플러스, IT기반 융합산업 창의인력양성사업단의 연구결과로 수행되었습니다

참 고 문 헌

[1] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, et al. “Intriguing properties of neural networks”. In: CoRR abs/1312.6199 (2013). arXiv: 1312 . 6199. URL: http://arxiv.org/abs/1312.6199.

[2] Ruitong Huang, Bing Xu, Dale Schuurmans, et al. “Learning with a Strong Adversary”. In: CoRR abs/1511.03034 (2015). arXiv: 1511.03034. URL: http: //arxiv.org/abs/1511.03034.

[3] Dongyu Meng and Hao Chen. “MagNet: a Two-Pronged Defense against Adversarial Examples”. In: CoRR abs/1705.09064 (2017). arXiv: 1705.09064. URL: http://arxiv.org/abs/1705.09064.

[4] Jiajun Lu, Theerasit Issaranon, and David Forsyth. “SafetyNet: Detecting and Rejecting Adversarial Examples Robustly”. In: The IEEE International Conference on Computer Vision (ICCV). Oct. 2017.

[5] Ian Goodfellow, Jonathon Shlens, and Christian Szegedy. “Explaining and Harnessing Adversarial Examples”. In: International Conference on Learning Representations. 2015. URL: http://arxiv.org/abs/1412.6572.

[6] Alexey Kurakin, Ian J. Goodfellow, and Samy Bengio.“Adversarial examples in the physical world”. In: CoRR abs/1607.02533 (2016). arXiv: 1607.02533. URL:http://arxiv.org/abs/1607.02533.

[7] Seyed-Mohsen Moosavi-Dezfooli, Alhussein Fawzi, and Pascal Frossard. “DeepFool: a simple and accurate method to fool deep neural networks”. In: CoRR abs/1511.04599 (2015). arXiv: 1511.04599. URL: http: //arxiv.org/abs/1511.04599.

[8] Nicholas Carlini and David A. Wagner. “Towards Evaluating the Robustness of Neural Networks”. In: CoRR abs/1608.04644 (2016). arXiv: 1608.04644. URL: http://arxiv.org/abs/1608.04644.

[9] Weiyang Liu, Yandong Wen, Zhiding Yu, et al. “Sphereface: Deep hypersphere embedding for face recognition”. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017, pp. 212 - 220.

[10] HaoWang, YitongWang, Zheng Zhou, et al. “Cosface:Large margin cosine loss for deep face recognition”.In: Proceedings of the IEEE Conference on Computer Visionand Pattern Recognition. 2018, pp. 5265 - 5274.

[11] Jiankang Deng, Jia Guo, Niannan Xue, et al. “Arcface:Additive angular margin loss for deep face recognition”. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019, pp. 4690 - 4699.

Data Augmentation & Merging Dataset for Facial Emotion Recognition

Jung Hwan Kim, Dong Seog Han
Graduate School of Electronics and Electrical Engineering, Kyungpook National University

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

Abstract

The facial emotion recognition (FER) system is desirable for many research fields such as game development, social workers, and the autonomous driving vehicles. The most popular research datasets, called FER 2013 and Extended Cohn-Kanade (CK+), were mainly tested by FER's researchers. However, testing only one dataset as FER 2013 or CK+ sometimes had limited further improvement of FER's performance even after data augmentation. Since the sampling FER datasets heavily affected to the FER's performance, we propose that merging datasets could be another method to improve the FER's performance other than using the data augmentation technique. By merging different datasets to magnify the number of training facial images, the FER performance improved 15.33% of validating accuracy.

I. Introduction

The facial emotion recognition (FER) is the future instrument for the game industry, social workers, and developing the autonomous driving vehicles. Improving FER system's performance is still in the primitive stage due to the scarcity of the FER datasets [5-6] for many FER researchers. The small number of facial images for training could lead the overfitting problem when the recent sophisticated neural networks and data augmentation were applied. Some FER researchers were decided to more facial images to improve the further FER's performance. But, others claimed that applying data augmentation on a small dataset could solve the data deficiency.

In essence, Kim *et al* [1] claimed that applying data augmentation onto the small number of facial images in FER dataset robustly solved a problem of the FER dataset's scarcity. They used the extended Cohn-Kanade (CK+) [6] dataset and applied with data augmentation to randomly manipulate facial images' transformation. Still, training with a small number of facial images could lead the bias result. We discovered that the given small number of face images to train performed well, but badly performed on newly detecting facial images even after the data augmentation.

In addition, Rosebroke [2] explained the fundamental concept of data augmentation. The data augmentation randomly added the jitteriness onto the original dataset's distribution, and magnify randomness of the dataset. The data augmentation generally solved an insufficient number of training images in a dataset, yet Sakai *et al* [3] displayed results which the large number of collected bio signals sometimes showed better performance than the small number of bio signals with data augmentation.

In this paper, we compared the performance of the Xception algorithms with and without applying the data augmentation on the FER 2013 [5] dataset. After the comparison, we collected the additional facial images and combined with different FER datasets to inspect the result of increasing the number of the dataset.

II. With and Without the Data Augmentation

FER 2013 had 48×48 pixels size of 35,813 facial images and 7 different categorical emotions: angry, disgust, fear, happy, neutral, sadness, and surprise. The first result from Fig. 1 without applying the data augmentation showed the overfitting problem during the training process. After applying the data augmentation of that data, the overfitting problem was resolved, and the performance was slightly improved from Fig. 2.

However, applying data augmentation on the small number of facial images could not improve the FER's performance further. Therefore, we were deterministic that increasing the number of facial images could potentially improve the FER's performance.

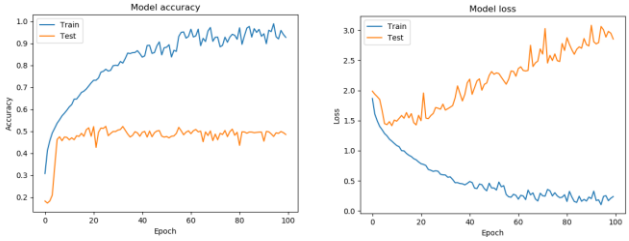


Fig. 1 Training the Xception model without data augmented FER 2013 Dataset.

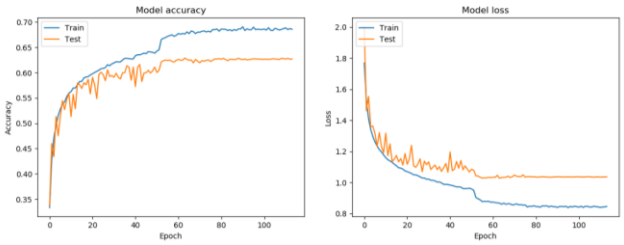


Fig. 2 Training Xception Model with data augmented FER 2013 Dataset.

III. Merging Datasets and Test Results

CK+ from Fig. 3 has 640×490 pixels size of 918 facial images within 8 categorical classifications: angry, contempt, disgust, fear, happy, neutral, sadness, and surprise. These facial images did not properly crop and a large portion of background pixels. Besides, training with the small number facial images even after applying data augmentation could lead the poor performance with the unseen dataset. The training result from CK+ dataset showed far superior than training with FER 2013 dataset, yet the model hardly detected the new emotional faces from unseen facial images. Training only CK+ dataset would not reach the robust FER performance.

We collected facial images from 60 video clips on YouTube and created intelligent signal processing lab (iSPL) dataset at Kyungpook National University from Fig. 3. The iSPL dataset contained 8,173 valid facials images and 7 categorical emotions as FER 2013. Although the performance showed better performance than the FER 2013 and CK+, the testing unseen dataset was still unable to detect the new facial emotions. Hence, we decided to merge all datasets together.

All facial images from different datasets such as FER 2013, CK+, iSPL have different size and different position of faces. To merge those different datasets without concerning of such the different sizes and position of facial images, we created the facial images threshing (FIT) machine. The FIT machine contained the multi-task cascade neural network (MTCNN) [7] and resizing program [2] that could symmetrically match to FER 2013 dataset. After all facial images became standardized to FER 2013's size and cropped faces by the FIT machine, we conducted a final experiment with the merged dataset.

The final result of the Xception algorithms and merged datasets from Fig.4 reached 76.32% of validating accuracy and increased 15.33% comparing with the 60.99% from Fig. 2. The experiment was applied with data augmentation in order to prevent from possible over-fitting problem. From Table I, applying confusion matrix's evaluation could confirm the robust improvement of the FER's performance. We used the unseen private facial images from the FER 2013 as simulating real-time testing.

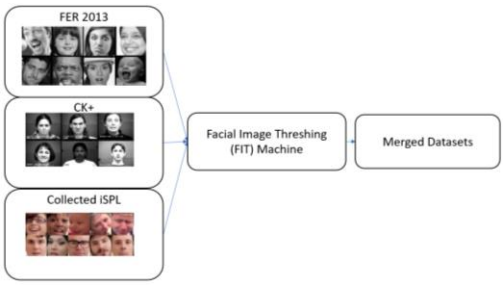


Fig. 3 Merging FER 2013, CK+, and iSPL Datasets by the FIT machine.

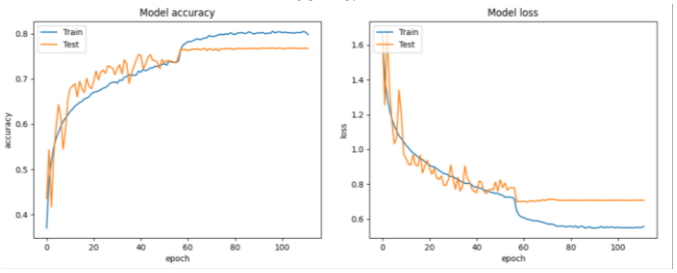


Fig.4 The final result from the merged datasets and the Xception algorithms.

Table I. The Result of Confusion Matrix from the unseen private test.

Datasets	Precision	Recall	F1 Score
FER 2013	61.6532%	58.7689%	59.4004%
Merged Dataset	**66.6236%**	**66.8845%**	**66.6779%**

IV. Conclusion

To conclude, data augmentation prevents from the over - fitting problem and also generalize the entire dataset. However, augmenting the small number of facial images by merging additional dataset proved to have further improvement of FER's performance but not with data augmentation.

ACKNOWLEDGMENT

This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (2016-0-00564, Development of Intelligent Interaction Technology Based on Context Awareness and Human Intention Understanding)

References

[1] J.-H. Kim, B.-G. Kim, P. P. Roy, and D.-M. Jeong, "Efficient facial expression recognition algorithm based on hierarchical deep neural network structure," IEEE Access, vol. 7, pp. 41273-41285, 2019.
[2] A. Rosebroke, Deep Learning for Computer Vision with Python, pp. 14-29, 2017.
[3] A. Sakai, Y. Minoda, and K. Morikawa, "Data augmentation methods for machine-learning-based classification of bio-signals," in 2017 10th Biomedical Engineering International Conference (BMEiCON), pp. 1-4.
[4] F. Chollet, "Xception: Deep learning with depthwise separable convolutions", in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1251-1258.
[5] P.-L. Carrier, A. Courville, I. J. Goodfellow, M. Mirza, and Y. Bengio, "FER-2013 face database," Universit de Montral, 2013.
[6] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, San Francisco, CA, 2010, pp. 94-101, doi:10.1109/CVPRW.2010.5543262.
[7] J. Brownlee, Deep Learning for Computer Vision: Image Classification, Object Detection, and Face Recognition in Python: Machine Learning Mastery, 2019.

특성기여도 분석 방법을 이용한 자동 증강의 영향에 관한 연구

김민기, 김재일
경북대학교
alsrlehd2@knu.ac.kr, threeyears@gmail.com

A Study on the Effect of Auto Augmentation using Feature Attribution Methods

Mingi Kim, Jaeil Kim
Kyungpook National University, Korea

요 약

데이터 증강은 데이터의 변하지 않는 특징을 학습하는 것을 목표로 한다. 본 논문에서는 특성기여도 방법을 바탕으로 자동 증강이 어떤 특징을 학습하고 있는지 확인하고, 특성기여도 방법을 사용했을 때의 한계점을 제시한다.

Ⅰ. 서 론

심층신경망(Deep neural nets)은 많은 양의 데이터에서 복잡한 패턴을 학습하며, 다양한 임무(Task)에서 높은 성능을 달성하고 있다. 하지만, 의료영상을 활용한 진단 보조를 비롯하여 실제 현장에서는 많은 양의 데이터를 확보하기 어렵기 때문에, 적은 데이터로부터도 일반화 성능이 높은 모델을 만드는 것은 중요한 문제이다. 이 문제를 해결하기 위해, 데이터 증강(Data Augmentation)이 많이 활용되고 있다. 데이터 증강에서는 이동(Translation), 좌우 뒤집기(Horizontal Flipping), 회전(Rotation)이 많이 사용되며, 이들은 데이터의 변하지 않는(Invariant) 특징을 학습하는 것을 목표로 한다. 데이터 증강은 데이터의 양과 다양성 모두를 증가 시킬 수 있는 방법이고, 선행 연구들로부터 분류문제에서 일반화 성능을 높일 수 있다는 것이 밝혀졌다.

데이터 증강을 적용했던 기존 연구에서 데이터 증강에 대한 강도(Magnitude)를 실험적으로 설계(Design)하는 경우가 많았고, 비용이 많이 드는 작업으로 여겨졌다. 최근 자동 증강(Auto Augmentation), Fast Auto Augmentation 논문에서 데이터의 미니배치(mini-batch)마다 데이터 증강을 다르게 적용하고, 이들의 오차율을 최소화 하는 방향으로 데이터 증강을 찾는 연구가 진행되었다 [1,2].

본 연구에서는 특성기여도 분석 방법(Feature Attribution Methods)을 사용하여 자동 증강이 어떤 특징을 학습하고 있는지 시각적으로 확인하고, 한계점을 제시한다. 여기서 특성기여도 분석 방법이란 입력이미지의 어느 부분이 모델 예측결과에 큰 영향을 미치는지 시각적으로 확인하는 방법을 일컫는다.

Ⅱ. 본론

1. 데이터셋과 모델 구성

CIFAR10[9] 데이터셋을 사용하여 실험을 진행하였다. 성능의 최종 테스트에 사용할 테스트 데이터셋(Test Dataset)을 제외한 50000개의 데이터에서 7500개의 데이터를 검증 데이터셋(Validation Dataset)으로 사용하여 초모수조정(Hyper-Parameter Tuning)을 수행한다. 학습에는 WideResNet(depth=4, widen_factor=2) 모델[8]을 사용했다.

2. 자동 증강

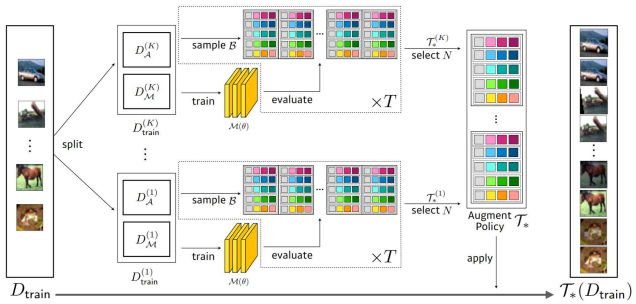


그림 1. 자동 증강 방법.

위 그림은 [2]에서 사용된 그림이다. 훈련 데이터셋(Train Dataset)을 K개의 구획(Fold)로 나누고, 각각의 구획에서 D_M과 D_A로 다시 나눈다. D_M을 이용하여 모델을 훈련 시킨 후, 초기 데이터 증강 T를 적용한 $T(D_A)$를 이용하여 오차율을 계산한다. 그리고 각 구획에서 가장 낮은 오차율을 가지는 N개의 데이터 증강 들을 모아 새로운 데이터 증강 T_*를 구성한다. 그리고 위의 과정을 반복하여 최종적인 데이터 증강 T_*를 확정한다.

3. 특성기여도 분석 방법의 구성

자동 증강이 어떤 특징을 학습하는지 시각적으로 확인하기 위하여, Grad-Cam[3], Smooth-grad[4], Integrated Gradient[5], Guided Backpropagation[6] 방법을 사용했다.

4. 실험결과

CIFAR10 데이터셋으로 Fast Auto Augmentation 방법을 사용하여 데이터 증강을 수행했다. 모델을 훈련시킨 결과 테스트 셋에서의 상위 1개 오차율(Top1 error)은 3.62%로 나타났고, Fast Auto Augmentation 방법을 사용하지 않았을 때의 상위 1개 오차율은 11.13%로 나타났다.

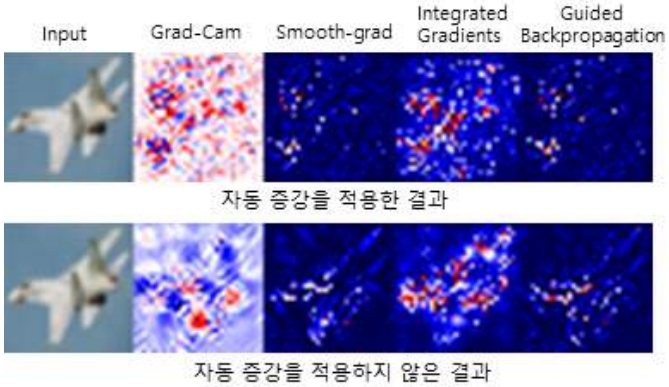


그림 2. 특성기여도 분석 방법을 이용한 자동 증강 효과 분석.

자동데이터 증강을 적용하지 않았을 때, 물체를 구별할 수 있는 특징을 강조하고, 자동데이터 증강을 적용했을 때, 물체의 특징 주변 배경을 강조하는 실험결과를 얻을 수 있다.

5. 실험결과 분석

자동 증강이 학습과정에서 배운 물체의 고유한 특징이 특성기여도 분석 방법을 통해 나타날 것이라 가정했었지만, 물체 주변을 강조하는 결과를 얻을 수 있었다. 위 결과를 얻게 된 원인을 세 가지로 추정할 수 있다. 첫 번째는 자동 증강을 통해 물체의 고유한 특징과 물체가 가지고 있는 맥락(Context)을 함께 학습했다는 것이다. 두 번째는 모델이 미니배치 내부의 검증 데이터셋의 오차율을 줄이는 방향으로 학습하는 특징과 인간이 중요하게 생각하는 특징이 다르다는 것이다. 세 번째는 특성기여도 분석 방법으로 데이터 증강이 학습할 때 배우는 특징을 정확하게 시각화 할 수 없다는 것이다. [7]에서 제시된 것처럼 특성기여도 분석방법은 적대적 공격(Adversarial Attack)에 취약하고, 모델의 입장에서 데이터 증강은 적대적 공격이라고 생각될 수 있다.

Ⅲ. 결론

자동 증강을 적용하여 모델을 학습시킬 때, 일반화 성능이 증가하는 것은 이미지를 구별할 수 있는 고유한 특징을 학습하기 때문이라고 생각된다. 이를 관찰하기 위해, 자동 증강을 적용한 모델에서 입력이미지의 기여도를 특성기여도 분석 방법을 통해 나타내어 보았다. 자동 증강을 적용하지 않은 모델은 물체의 구별할 수 있는 특징을 나타내는 반면, 자동 증강을 적용한 모델은 주변 배경을 강조하는 것이 관찰되었다. 하지만, 실험결과 분석에서 서술한 것처럼 다양한 해석이 있을 수 있고, 다른 실험조건(데이터셋과 모델의 변화)에서 실험을 수행하지 않았기 때문에 자동 증강이 추출하는 특징을 정확하게 관찰하기 위해서는 추가적인 연구가 필요할 것으로 사료된다.

ACKNOWLEDGMENT

이 논문은 2020년도 정부(교육부)의 재원으로 한국연구재단 기초연구사업의 지원을 받아 수행된 연구임(“2020R1I1A3074639”)

References

[1] E. D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, and Q. V. Le. Autoaugment: Learning augmentation strategies from data. In Proceedings of the IEEE conference on computer vision and pattern recognition, 2019.

[2] Sungbim Lim, Ildoo Kim, Taesup Kim, Chiheon Kim, and Sungwoong Kim. Fast autoaugment. arXiv:1905.00397 [cs.LG], 2019.

[3] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Gradcam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

[4] Daniel Smilkov, Nikhil Thorat, Been Kim, Fernanda Viégas, and Martin Wattenberg. Smoothgrad: removing noise by adding noise. arXiv preprint arXiv:1706.03825, 2017.

[5] Sundararajan, Mukund, Taly, Ankur, and Yan, Qiqi. Axiomatic attribution for deep networks. arXiv preprint arXiv:1703.01365, 2017.

[6] Springenberg JT, Dosovitskiy A, Brox T, Riedmiller M. Striving for simplicity: The all convolutional net. arXiv 2014

[7] Amirata Ghorbani, Abubakar Abid, and James Zou. Interpretation of neural networks is fragile. arXiv preprint arXiv:1710.10547, 2017.

[8] S. Zagoruyko and N. Komodakis. Wide residual networks. arXiv preprint arXiv:1605.07146, 2016.

[9] A. Krizhevsky. Learning multiple layers of features from tiny images. Tech Report, 2009.

데이터 증강과 전이학습을 활용한 주행환경에서의 감정 인식 모델의 성능 향상 실험

최준혁, 조현보
포항공과대학교 산업경영공학과

cjh0102@postech.ac.kr, hcho@postech.ac.kr

Ensemble Method using Data Augmentation and Transfer Learning for Face Recognition Model for driving environments.

Choi Jun Hyuk, Cho Hyunbo
Department of Industrial Engineering and Management, POSTECH

요 약

본 논문은 기존 감정 인식 모델이 주행 환경에서 성능이 하락하는 한계점을 극복하기 위해서, 데이터 증강과 전이학습을 동시에 활용한 프레임워크를 제안한다. 적용 환경에서 수집된 학습 데이터가 부족한 통해 해결하고자 한다. 데이터 증강 기술을 통해 감정 라벨링이 없이 주행환경에서 수집된 데이터, 감정이 라벨링된 데이터를 통해 주행환경을 위한 데이터를 생성하고, 해당 데이터에 기반하여 전이학습을 수행한다. 실험 결과 데이터 증강과 전이학습 기반 감정 인식 모델이 기존 감정 인식 모델의 결과에 비해 유의미한 성능 향상을 보였음을 확인하였다.

Ⅰ. 서 론

최근 자율주행차의 등장으로 운전자 맞춤형 콘텐츠 제공이 관심을 받고 있다. [1] 그 중에서 운전자의 얼굴 표정을 분석하여 감정을 도출하고 음악과 동영상 등의 맞춤형 콘텐츠를 제공하는 기술에 대한 연구가 활발히 이뤄지고 있다.

얼굴 표정 인식 기술을 위해 딥러닝 알고리즘이 좋은 성능을 보여주고 있지만, 높은 성능의 딥러닝 모델을 개발하기 위해서는 충분한 학습 데이터가 필요하다. 즉, 딥러닝 모델이 적용되는 환경의 다양한 상황이 포함된 학습 데이터가 필요하지만, 데이터 수집 비용 등의 한계로 현실적으로 어려운 상황이다.

해당 문제점을 해결하기 위해 대표적으로 데이터 증강과 전이학습 방법론이 각각 활용된다. [2] 본 논문에서는 데이터 증강과 전이학습을 함께 사용하여 주행환경에서의 감정 인식 모델의 성능을 향상하는 프레임워크를 제안한다. 나아가 프레임워크 적용 전후의 감정 인식 모델의 성능을 비교하여 데이터 증강과 전이학습 기술의 효용성을 검증하고자 한다.

Ⅱ. 본론

본 연구의 프레임워크는 다음 <그림 1>과 같이 두 단계로 구성된다.

2-1. 데이터 증강 Phase

데이터 증강이란 적은 양의 학습 데이터에 인위적인 변화를 가해 새로운 학습 데이터를 생성하는 기법이다. [4] 이 때, 적용 상황에 맞지 않게 증강된다면 오히려 모델의 성능이 하락한다.

본 Phase 에서는 데이터 증강 기술로 주행 환경의 특성을 충분히 담고 있는 데이터셋을 생성한다. 먼저 주행 환경에서 수집된 DrivFace 에서 주행 환경 내의 변수와 범위를 도출하고, KoreanFaceDB 에 해당 Policy 에 적용하여 데이터를 증강한다.

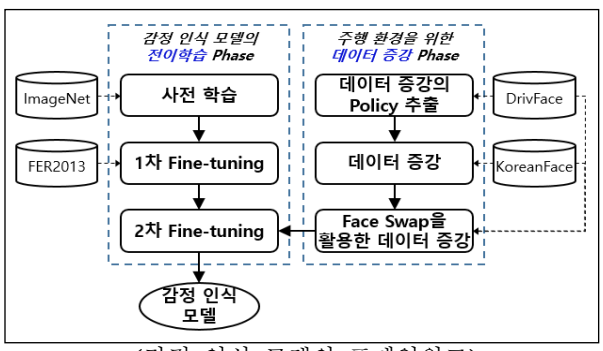
감정이 라벨링된 주행 환경의 얼굴을 생성하기 위해서, 주행 환경 속 얼굴 이미지와 다양한 감정이 라벨링된 Augmented KoreanFaceDB 에 Face Swap 기술을 적용하여 감정이 라벨링된 주행 환경 속 얼굴 이미지가 생성된다.

2-2. 전이학습 Phase

전이학습이란 기존 데이터로 사전 학습된 모델의 일부분 혹은 전체를 대상으로, 예측모델의 적용 상황에서 수집된 데이터로 재학습하여 적용 상황의 목적에 적합한 모델을 생성하는 기법이다. [3]

본 Phase 에서는 ImageNet, FER-2013 과 주행 환경을 위해 증강된 데이터를 활용하여 차례대로 기존 모델을 미세조정하여 주행 환경에 적합하게 모델을 재학습한다.

감정이 라벨링된 주행 환경의 얼굴을 생성하기 위해서, 주행 환경 속 얼굴 이미지와 다양한 감정이 라벨링된 Augmented KoreanFaceDB 에 Face Swap 기술을 적용하여 감정이 라벨링된 주행 환경 속 얼굴 이미지가 생성된다.



<감정 인식 모델의 프레임워크>

2-3. 실험 결과

검증 데이터는 “2-1 데이터 증강” 에서 생성된 데이터를 랜덤하게 추출하여 생성한다. 검증 데이터를 대상으로 증강된 데이터로 전이학습한 감정 인식 모델과 증강된 데이터를 사용하지 않은 감정 인식 모델의 성능을 비교하였다. 실험 결과 데이터 증강과 전이학습 기반 감정 인식 모델이 기존 감정 인식 모델의 결과에 비해 유의미한 성능 향상을 보였음을 확인하였다. 학습에 사용한 하드웨어는 CPU Intel i7-8700K, RAM 48GB, NVDIA GTX 1080Ti 이다.

Ⅲ. 결론

본 논문에서는 데이터 증강과 전이학습을 활용하여 주행환경에 적합한 감정 인식 모델 개발 방안에 대해 고찰하였다. 실험 결과, 제안 프레임워크를 적용한 결과 기존 대비 성능이 향상됨을 확인하였다. 나아가, 실제 주행환경에서 수집된 데이터를 추후에 확보하여 모델을 검증하고자 한다.

ACKNOWLEDGMENT

한국산업기술진흥원의 광역협력권 산업육성사업 내 “자율주행 자동차용 스마트 GLOVE BOX 편의장치 개발” 과제의 지원을 받아서 수행됨

참 고 문 헌

[1] Davies R. W." The Data Encryption standard in perspective,"Computer Security and the Data Encryption Standard, pp. 129-132.

[2] 이한수, et al. "전이학습 기반의 합성곱 신경망을 이용한 다중클래스 분류에 관한 연구." 한국지능시스템학회 논문지 28.6 (2018): 531-537.

[3] Sarkar, Dipanjan. "A Comprehensive Hands-on Guide to Transfer Learning with Real-World Applications in Deep Learning." (2018).

[4] Lim, Sungbin, et al. "Fast autoaugment." Advances in Neural Information Processing Systems. 2019.

공연 포스터의 이미지 특성을 활용한 딥러닝 기반 관객예측

조유정, 강경표*, Yao hui**, 권오병***
경희대학교, *경희대학교, **경희대학교, ***경희대학교

yujung251@khu.ac.kr, *kpkang0646@naver.com,
**zgyh97y@khu.ac.kr, ***obkwon@khu.ac.kr

Deep Learning-Based Audience Prediction Using Poster Image Characteristics

Yu Jung Cho, Kyung Pyo Kang *, Yao Hui**, Oh byung Kwon***
Kyung Hee Univ., * Kyung Hee Univ., ** Kyung Hee Univ., *** Kyung Hee Univ.

요 약

공연예술 기관에서의 공연에 대한 흥행 예측은 공연예술 산업 및 기관에서 매우 흥미롭고도 중요한 문제이다. 이를 위해 출연진, 공연장소, 가격 등 정형화된 데이터를 활용한 전통적인 예측방법론, 데이터 마이닝 방법론이 제시되어 왔다. 그런데 관객들은 공연안내 포스터에 의하여 관람 의도가 소구되는 경향도 있음이 분명함에도 불구하고, 포스터 이미지 분석에 의한 흥행 예측은 거의 시도되지 않았다. 최근에는 이미지를 통해 판별하는 CNN계열의 딥러닝 방법이 개발되면서 포스터 분석의 가능성이 열렸다. 이에 본 연구의 목적은 공연관련 포스터 이미지를 통해 흥행을 예측할 수 있는 딥러닝 방법을 제안하는 것이다. 이를 위해 KOPIS에 공개된 포스터 이미지를 학습데이터로 하여 Pure CNN, VGG-16, Inception-V3, ResNet50 알고리즘을 통해 예측을 수행하였다. 그 결과 흥행예측 정확도 80% 이상의 높은 성과를 보였다. 또한 공연 관련 정형데이터를 활용한 전통적 방법론과의 앙상블을 시도하였다. 본 연구는 공연예술 분야에서 이미지 정보를 활용하여 흥행을 예측하는 첫 시도이며, 본 연구에서 제안한 방법은 연극 외에 영화, 기관 홍보, 기업 제품 광고 등 포스터 기반의 광고를 하는 영역으로도 적용이 가능할 것이다.

Ⅰ. 서 론

공연예술 기관에서의 콘텐츠에 대한 흥행 예측은 공연예술 산업 활성화에 중요한 이슈이다. 과거 흥행 예측 방법으로는 전통적인 예측 기법인 다중회귀분석[1][2]과 Bass모형[2]외에 Random Forest, KNN, 인공신경망[3] 등 데이터 마이닝 알고리즘 등이 제안되어 왔다.

지금까지의 흥행 예측 연구는 대부분 출연진[4]이나, 공연장소[5][6], 일반인 평가[4], 전문가 평가[7]등 정형적인 공연 특성에 의한 것이었다. 그러나 최근 이미지 기반의 판별 문제가 CNN 등 딥러닝 알고리즘에 의하여 해결 가능해짐으로써 공연 포스터와 같은 이미지로도 흥행을 예측하는 시도가 가능 해졌다.

공연예술 분야의 흥행 관련 연구는 성과 예측 모형을 직접 제시하는 것보다는 연극의 장르(유희적, 교육적, 교훈적 연극), 원작의 유무, 연극의 속성[8]등 연극의 특성이 흥행에 미치는 연구가 대부분이다[9]. 예측 모형 연구가 활발하지 않은 이유는 영화만큼 예측 관련 데이터가 풍부하지 않고[8], 설문조사나 면접 방법 등을 통해 자료를 확보하기 때문에 성능이 좋은 예측 모형을 만들기가 용이하지 않기 때문이다.

이에 본 연구에서는 비정형 데이터인 공연관련 포스터 이미지를 통해 흥행을 예측해보고, 정형화된 공연 특성 데이터만으로 예측한 것과 성능을 비교해 봄으로

써 비정형 데이터인 연극 공연 포스터로 연극의 흥행 여부를 판단할 수 있는 지와 정형화된 공연 관련 데이터와 포스터 이미지 특성을 복합적으로 활용하여 흥행을 예측하였을 때 어느 것이 더 우수한 성능을 나타내는지 알아보고자 한다.

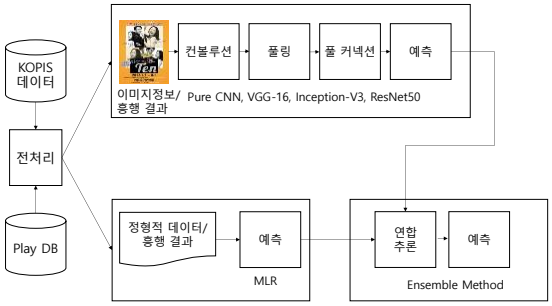
Ⅱ. 본론

본 연구에서 공연 포스터 이미지를 기반으로 흥행을 예측하기 위해 <그림1>과 같이 예측을 수행하고자 한다. 먼저 복수의 오픈 데이터를 수집하고 전처리를 한 후에 포스터 이미지 데이터에 대해서는 복수의 CNN 계열 알고리즘들로 흥행을 예측하고, 정형화된 데이터는 회귀분석을 통해 흥행을 예측한다. 그리고 두 가지의 흥행 예측 결과를 바탕으로 앙상블 기법을 수행한다. 연극공연에 대한 정보는 공연예술통합전산망인 KOPIS(http://www.kopis.or.kr)에서 API key를 발급받아 연극정보, 연극상세정보, 연극포스터 등을 확보하였다.

학습에 사용되는 모델은 CNN을 기반으로 한 모델이다. 수집된 데이터는 추정 관객수를 기준으로 관객수가 500명 미만일 경우 흥행실패, 500명 이상일 경우 흥행성공으로 분류하였다. 추정 관객수는 공연횟수와 1회 당 최대 유치 관객수, 즉 공연장 규모의 곱으로 구했다. 따라서 CNN 모델은 흥행 실패와 흥행성공 두 개의 클래스로 분류되도록 모델을 설계하였다. CNN 모델은 각각 3개의 Convolution layer와 maxpooling 층으로 이루어진 Pure CNN과 Python의 패키지 Keras의 내장 사전 훈련 모델로 있는 VGG16, Inception-v3, Resnet50모델을 사용하였다.

Ⅲ. 결론

공연예술 분야는 인공지능. 특히 딥러닝이 적용될 수 있는 유력한 분야임에도 불구하고, 그동안 많은 연구가 진행되지 못했다. 이에 본 연구는 연극 분야가 보유하고 있는 이미지를 포함한 오픈 빅데이터를 토대로 연극의 활성화를 위한 예측과 의사결정에 도움이 되는 딥러닝 방법을 제안했다는 데 의미가 있다.



<그림 1> 전체적인 프로세스

ACKNOWLEDGMENT

이 논문 또는 저서는 2020 년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2020S1A3A2A02093277)

참 고 문 헌

[1] 김상호, 한진만. (2014). “한국 영화의 흥행성과 결정요인 분석.” 사회과학연구, 53(1), pp.191-214.

[2]김보경, & 임창원. (2018). “GLS 와 Bass 모형을 결합한 하이브리드 모형을 이용한 영화 관객 수 예측.” 응용통계연구, 31(4), pp.447-461.

[3] 정찬미. (2020). “영화 관객 수 예측을 위한 머신러닝 기법의 성능 평가 연구.” 석사학위논문, 이화여자대학교

[4] 김연형, 홍정한 (2011). “영화 흥행 결정 요인과 흥행 성과 예측 연구.” 한국통계학회논문집, 18(6), pp.859-869.

[5] 권선주. (2014). “영화 흥행성과의 분석과 예측: 뉴스와 웹사이트 데이터 이용.” 문화경제연구, 17(1), pp.35-55.

[6] 우종필, & 이응환. (2018). “빅데이터 분석을 통한 천만 관객 영화 예측 모델.” 한국빅데이터학회지, 3(1), pp.63-71.

[7] 전범수, & 이준영. (2019). “한국 개봉 흥행 영화 평점이 성과에 미치는 영향: 일반 이용자와 전문가 평점의 비교.” 언론과학연구, 19(4), pp.227-253.

[8] 송은아. (2013). “장기공연 연극의 특성에 관한 연구.” 한국엔터테인먼트산업학회 학술대회 논문집, pp.132-136.

[9] 김태희, & 신형덕. (2013). “어린이공연의 내용, 장르, 원작유무가 공연 흥행에 미치는 영향: 등교기간의 조절효과를 중심으로.” 한국산학기술학회 논문지, 14(10), pp.4762-4768.

이미지 분류 네트워크에서의 효율적 훈련 기법에 대한 연구

배은지, 이성진
동서울대학교
ejbae25@du.ac.kr, sungjinlee@du.ac.kr

A Study on the efficient training methodology in the Image Classification Network

Eunjee Bae, Sungjin Lee
Dong Seoul University

요 약

최근 영상인식 기술의 성능적 발전은 수많은 데이터 축적과 딥러닝 네트워크의 심층화에 기인하여 왔다. 하지만, 이런 다양한 데이터들을 딥러닝 네트워크에 훈련시키는 것은 다양한 문제들을 유발시킨다. 적은 데이터 량에서 기인하는 오버피팅, 클래스 간의 데이터 양 차이에서 오는 클래스 불균형 (Imbalance), 멀티클래스 훈련 문제 등이 그것이며 본 논문은 Pascal VOC 데이터에서 이런 문제들을 발견하고 분석하였으며 해결책을 제시하고 실험으로 성능을 분석해 보았다.

Ⅰ. 서론

최근 들어, IT산업뿐 아니라 대부분의 산업 분야에서 딥러닝 기반 영상인식 기술이 주목받고 있으며, 이러한 관심은 ImageNet, Open Images Dataset, MS COCO 와 같은 수많은 데이터 셋 구축 [1][2][3] 과 딥러닝 네트워크 심층화로 이어졌다 [4][5][6].

하지만 실제로 이런 딥 뉴럴 네트워크를 커스텀 데이터 셋에서 훈련시켜 문헌에서 제공된 성능만큼을 얻어내기란 쉽지 않다. 보통 문헌에서 제공되는 성능은 매우 큰 데이터 셋에서 다양한 클래스를 분류해 내는 챌린지(예, ILSVRC (ImageNet Large Scale Visual Recognition Competition))에 최적화된 네트워크를 사용하지만, 실제 산업에서 사용되는 데이터 셋과 분류 클래스는 그에 비해 매우 작기 때문에 적합한 모델 선택에 실패할 시, 데이터 과적합 (Overfitting) 혹은 데이터 언더피팅 (Underfitting) 문제 등의 문제들이 발생하기 때문이다.

본 논문에서는 객체 분류 (Classification) 와 검출 (Detection) 성능 분석에 많이 사용된 PASCAL VOC 데이터 셋을 이용하여 VGG16 네트워크로 학습하였을 때의 결과를 통해 이 학습의 문제들에 대한 해결책들을 제시하고, 효과를 분석하였다. Pascal VOC 데이터 셋은 ImageNet 데이터셋에 비해 클래스 별 이미지 양이 적고 클래스 간 이미지 양의 차이가 크며 한 이미지 내에 여러 클래스가 포함되어 있는 경우가 많기 때문에 훈련시키기 까다로운 편이다.

본 논문은 이런 Pascal VOC 데이터 셋에 대해 분석하고 해당 데이터셋을 훈련시키는데 발생되는 어려운 점들, 오버피팅, 클래스 불균형, 멀티 클래스을 분석하였다. 그 후 각 문제점들을 위한 해결책들을 제시하고 이를 실험적으로 검증하였다.

Ⅱ. 데이터 분석과 훈련기법

실험을 위해 사람, 동물, 탈것, 실내와 관련된 20개의 클래스를 가지고 있는 PASCAL VOC 데이터 셋을 사용하였다. 그러나 이 데이터 셋을 그대로

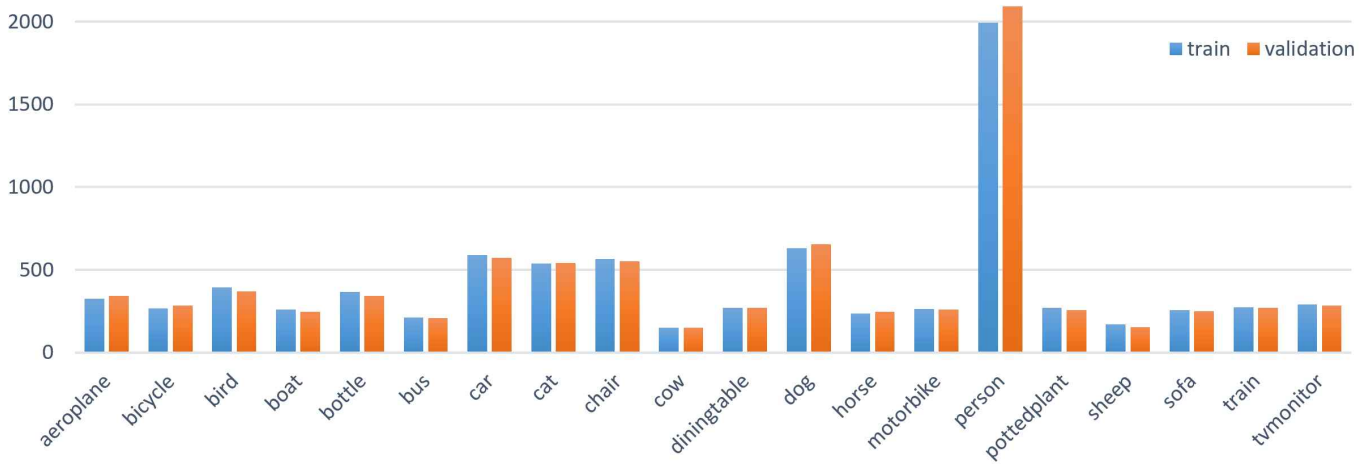


그림 1 Pascal VOC 데이터 셋의 클래스 구성

이용하여 학습하기에 여러 어려운 점들이 있다.
우리는 이러한 어려운 점들을 분석하고 해결하기 위해 데이터 전처리 과정에서 multi class 데이터를 제거해 보고, 부족한 훈련 데이터를 보충하여 실험해 보았으며, 클래스 간 데이터 비율을 평탄화하여 학습하였을 때의 정확도를 비교해보았고, 원 데이터 셋에서 가장 데이터가 많은 4개의 class를 선정해 학습하였을 때의 정확도도 비교해보았다.

1. Overfitting

그림 1에서 보듯이 Pascal VOC의 데이터 셋은 특정 클래스의 경우 이미지 개수가 매우 적은 것을 확인할 수 있다. 이런 데이터 부족은 Overfeatting 문제를 유발시켜 train 과 validation accuracy 의 차이를 발생 시킨다. 이 문제를 해결하기 위해 해당 클래스의 이미지가 최소 250여장이 되도록 추가적인 데이터 약 2500여장 보충을 수행하였다. 또한 해당 데이터 들은 기존 이미지와는 겹치지 않게 새로운 데이터로 크롤링하여 보충하였다.

2. Class Imbalance Problem

훈련 시 클래스 간 이미지 개수를 유사하게 맞추는 클래스 균등화 (Class balancing) 는 인식 시에 클래스 간 편향성을 줄여줄 수 있는 전처리 단계이다. 그림 1에서 보듯이, Pascal VOC 데이터 셋은 클래스 간 imbalance 가 심하다. 학습 데이터 (training data) 가 총 8,331장으로 이루어져 있으나 그 중 1,994장이 person class에 속해있었으며, 가장 data가 적은 class (151장) 와는 13배가량 차이가 난다. 본 논문에서는 20개 클래스 간 이미지 개수를 유사하게 맞추기 위해 이미지가 부족한 클래스는 위에서 설명한 것처럼 추가적인 데이터 보충 단계를 진행하였고 일부 데이터가 너무 많은 클래스는 250여장으로 균일하게 줄여서 전처리 작업을 수행하였고 이에 대한 영향도 분석하였다.

3. Multi-Class Training

원 데이터 셋인 Pascal VOC 에서는 한 이미지 내 여러 클래스 (multi-class) 가 다수 존재한다. 이런 다중 클래스를 포함하는 이미지는 딥러닝 네트워크 훈련 시 인식률에 악영향을 미칠 수 있게 된다. 그래서 이런 다중 클래스를 포함한 이미지를 제거하고 훈련함으로의 인식률 변화를 확인하였다.

Ⅲ. 실험

본 논문에서는 데이터 셋의 문제점들을 면밀히 관찰하기 위해 가장 보편적으로 사용되는 VGGNet16을 사용하였고 ImageNet에 Pretrain 된 가중치에 Header 만 추가하여 재학습 시켰다. 분석 데이터 셋의 조합은 다음과 같다.

- **ORG** : Pascal VOC 기본 데이터 셋
- **DS** : ORG 데이터 셋에 각 클래스 별로 300장 이상 이 되도록 데이터 보충
- **DSB** : DS 데이터셋에서 모든 클래스들이 균일하게 300장이 되도록 조정
- **NMC** : ORG 데이터 셋의 훈련 데이터에서 여러 클래스가 아닌 한 개의 클래스 만 포함하도록 조정한 데이터 셋
- **NMCB** : NMC 데이터셋에서 모든 클래스들이 균일하게 250장이 되도록 추가

그림 2에서 알 수 있듯이 ORG 데이터 셋에 비해 DS 데이터 셋에서 성능이 많이 향상된 것을 알 수 있다. 이로서, 데이터 양이 확보되는 것이 정확도 향상에 중요하다는 것을 알 수 있다.

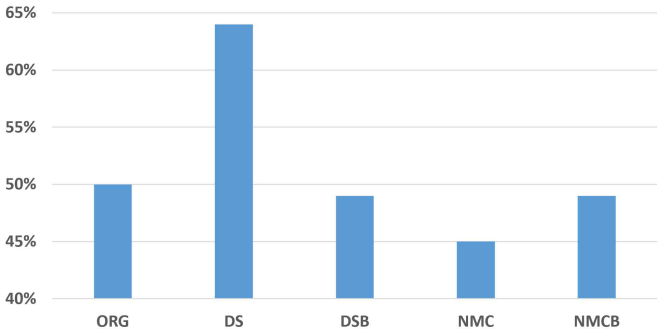


그림 2 Pascal VOC 데이터 셋 조합에 따른 VGGNet19 성능

두 번째로 클래스 별 불균형을 해결하기 위한 DSB 성능을 보면 오히려 조금 줄어든 것을 확인할 수 있다. 그 원인은 데이터 양이 많았던 일부 클래스의 데이터 양의 크기가 줄어듬으로서 해당 클래스의 정확도 감소가 일어났기 때문이다. 이로서 클래스 균형 보다는 데이터 양 확보가 정확도에서는 더 중요하다는 것을 알 수 있다.
세 번째로 ORG 훈련 이미지에서 멀티 클래스 이미지를 제거한 NMC 의 성능이 제일 줄어든 것을 알 수 있다. 그 이유는 NMC에서 데이터 양의 줄어들기 때문에 거기서 오는 성능 열화가 상당한 것으로 분석된다. 두 번째의 경우와 마찬가지지만 데이터 양이 적으면 클래스 균등화 혹은 멀티 클래스 제거와 같은 효과는 크지 않거나 오히려 열화의 원인이 된다.

Ⅳ. 결론

본 논문에서는 커스텀데이터 셋들에서 발생할 수 있는 데이터 부족, 클래스 불균형, 멀티 클래스 인식에 관한 어려운 점들을 분석해 보았고 실험결과를 통해 데이터 셋 양 확보가 성능에 제일 중요한 요인이라는 것을 보였다.

ACKNOWLEDGMENT

이 논문은 2020년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기본연구사업임(No. NRF-2019R1F1A1062878)

참 고 문 헌

[1] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li and L. Fei-Fei, ImageNet: A Large-Scale Hierarchical Image Database. IEEE Computer Vision and Pattern Recognition (CVPR), 2009

[2] https://opensource.google/projects/open-images-dataset

[3] Tsung-Yi Lin and etc, “Microsoft COCO: Common Objects in Context”

[4] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks”, NIPS 2015

[5] Karen Simonyan, Andrew Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition”

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, “Deep Residual Learning for Image Recognition”

심층 신경망 기반 차량 통신시스템의 신호 성상 분류 모델

김지훈, 한동석*
경북대학교 전기전자공학부
soji@knu.ac.kr, dshan@knu.ac.kr*

A Signal Constellation Classification Model of Deep Learning-based Vehicle Communication System

Jihun Kim, Dong Seog Han*
Kyungpook National Univ.

요 약

본 논문은 빠르게 변화하는 차량 통신 환경에서 인공 신경망을 이용한 수신 신호의 성상을 분류하는 모델을 제안한다. 차량 이동 환경의 무선 엑세스(wireless access in vehicular environment, WAVE) 시스템은 BPSK, QPSK, 16-QAM, 64-QAM과 같은 변조 방식을 사용한다. 이러한 변조 신호를 성상 이미지화하고 이를 학습한 인공 신경망 모델을 통해 이동 통신 채널의 영향을 받은 수신 신호의 성상을 분류한다. 변조 신호 성상 이미지의 분류를 위한 인공 신경망 모델은 합성곱 신경망 구조에 기반한 AlexNet을 적용한다. 분류 모델을 위한 변조 신호는 이동 통신 채널의 특성인 레일리 페이딩 채널에서 송수신된 신호를 이용한다.

Ⅰ. 서 론

차량 통신은 고속으로 이동하는 상황에서 차량과 차량, 차량과 노변 기지국 등과의 이동 통신 환경이다. 이동 통신 환경은 전파의 반사, 회절, 산란 현상을 모두 겪게 되어 수신 신호의 큰 변화를 일으킨다. 따라서 차량 통신 시스템의 수신부에서는 높은 수준의 수신 기법이 적용되어야 한다. 최근에는 복잡한 구조의 수신기를 심층 신경망 등으로 대체하는 접근법이 활발히 적용되고 있다. 심층 신경망을 적용한 방식은 기존의 통신 방식보다 효율적인 구조를 지원하고 실제로 대체 가능한 성능 수준을 보이고 있다. 본 논문에서는 차량 통신 시스템의 수신 신호 복조 과정을 심층 신경망을 적용하여 수신 신호를 분류하는 모델을 제안한다.

Ⅱ. 본론

심층 신경망은 이미지 및 비디오 인식, 자연어 처리 등 다양한 응용 분야에서 기존의 방식에 비해 효과적인 성능 및 결과를 도출하며 최근에는 많은 연구에 적용되고 있다. 본 논문에서는 빠르게 변화하는 채널의 특성을 극복하는 복조 성능에 도달하기 위해 심층 신경망을 이용한 효율적 복조 모델을 제안한다. 제안하는 알고리즘은 그림1과 같으며, 차량 통신 채널의 영향을 받은 수신 신호를 입력으로 하는 합성곱 신경망을 통해 복조 신호를 분류하는 구조이다. 심층 신경망은 많은 양의 데이터에 의존적이며 충분한 학습 데이터를 보유하고 있다면 최적의 분류 성능을 도출할 수 있다. 또한 심층 신경망 기반의 분류 모델은 변조 분류 작업의 복잡성을 크게 줄이는 효과가 있다. 이 논문에서는 합성곱 신경망의 대표적인 종류인 AlexNet[1]을 기본 구조로 하는 모델을 제안하였다. AlexNet은 이미지 데이터를 입력으로 합성곱 레이어와 ReLU 함수를 통해 네트워크를 학습한다. 제안하는 신호 성상 분류 모델은 이동 통신 채널 특성의 레일리 페이딩 채널에서 수신된 신호를 통해 차량 통신 환경을 반영한 신호를 학습하게 된다. 또한 차량 통신 환경의 신호 성상 분류를 위해 IEEE 802.11p(WAVE) 변조 기법으로 이진 위상 편이 변조(BPSK), 직교 위상 편이 변조(QPSK), 16 직교 진폭 변조(16-QAM), 64 직교 진폭 변조(64-QAM) 구조를 사용하였다. 제안하는 모델의 학습 및 테스트를 위해서 각 변조 방식 별 250개로 구성되며 각 데이터는 신호 대 잡음비 0dB에서 30dB까지의 범위에서 수집되었다. 수집 데이터는 AlexNet 입력을 위해 227 × 227로 변환하였다. 제안하는 심층 신경망 기반의 성상 분류 모델의 성능은 레일리 페이딩 채널의 신호에서 약 80%의 정확도를 보였다.

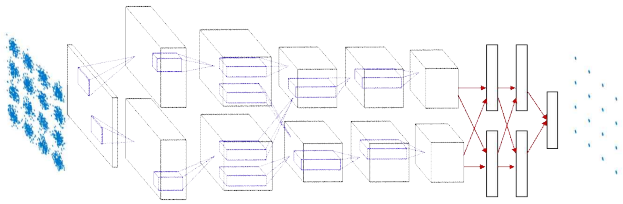


그림 1. 딥러닝 기반 신호 성상 분류 모델

Ⅲ. 결론

본 논문에서는 차량 통신 채널 환경에서 합성곱 신경망을 이용한 복조 신호 분류 모델을 제안하였다. 제안하는 모델은 차량 통신 표준의 변조 방식기반의 분류 성능을 실험하였으며 평균 80%의 분류 정확도 성능을 보였다.

ACKNOWLEDGMENT

본 연구는 산업통상자원부와 한국산업기술진흥원이 지원하는 5G기반 자율주행 융합기술 실증 플랫폼 과제(과제고유번호 : 1415169669)의 지원을 받아 수행하였습니다.

참 고 문 헌

[1] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Communications of the ACM 60.6 (2017): 84-90.

학습 기반 자율주행 제어 보조 모듈 설계

한경석*
*경북대학교

*kyoungsh@knu.ac.kr

Learning-based autonomous driving control auxiliary module design

Kyoungseok Han*
*Kyungpook National Univ.

요 약

본 논문은 자율주행 강인 제어 알고리즘의 보조를 위한 보조 모듈을 학습 기반으로 설계하였다. 강화학습 등 학습 기반의 제어 정책 (Control Policy) 설계에 관한 연구가 활발한데, 자율주행의 경우 학습되지 않은 주행환경을 마주했을 때 그 성능을 보장 할 수가 없다. 또한 기업마다 고유 기술을 기반으로 제어기가 이미 설계되어 있기 때문에 제어기 자체를 완전히 새롭게 설계하는 것 보다 기 설계된 제어기를 그대로 활용하면서 제어기로 부터 결정된 제어 출력을 필터링할 수 있는 보조 모듈을 제안하고자 한다. 본 논문에서 제안하는 추가모듈은 어떤 제어기에도 추가 될 수 있으며 충돌율 측면에서 전체 제어 성능을 향상 시킴을 확인하였다.

Ⅰ. 서 론

본 논문에서는 자율주행 강인 제어 알고리즘 개발에 있어서, 이를 보조하는 개념의 추가 모듈을 설계한다. 자율주행 연구는 산학연에서 활발하게 진행되고 있으나, 각 연구기관에서는 고유의 보유 기술을 기반으로 제어기를 설계하므로 같은 주행상황이라도 자율차는 다른 액션을 취할 가능성이 존재한다. 즉, 모든 주행 상황에 강건한 제어기 설계가 목표지만 이를 실현하는게 어려운게 사실이다.

또한 대부분의 자율주행 연구는 기업이 선도하므로 학계에서 개발된 제어기를 도입하는 것이 쉽지 않다. 따라서 본 논문에서는 디폴트 제어기는 그대로 사용하고, 이를 보조하는 추가 모듈을 학습기반으로 개발한다. 대부분의 주행상황에서는 디폴트 제어기가 올바른 선택을 할 것이지만, 자율차가 마주하지 않았던 상황을 마주했을 때는 추가된 모듈이 활성화되어서 사고를 방지한다.

추가된 모듈은 Action Governor 라고 명명되는데, 디폴트 제어기에서 선택된 액션을 필터링한다는 의미고 이는 학습기반으로 설계된다 [1]. 즉, 주행시뮬레이터를 활용하여 위험 주행상황을 충분히 학습하여 특정 주행상황에서 bad action 들을 action governor 버퍼에 저장 후 실제 주행시 action governor 버퍼의 bad action 을 필터링한다.

본 추가 모듈을 도입한다면, 어떤 디폴트 제어기와도 결합가능하여 제어 성능을 보조할 수 있으며 제어기를 설계한 기업의 입장에서도 부담없이 도입할 수 있을 것으로 예상된다.

Ⅱ. 본론

본 논문에서 제안되는 Action Governor 를 학습시키기 위해서는 적절한 주행 환경을 제공해주는 주행 시뮬레이터가 필요하다. 본 논문에서는 기 개발된 2 차선 혹은 3 차선 고속 주행상황에서의 주행 시뮬레이터를 활용한다[2].

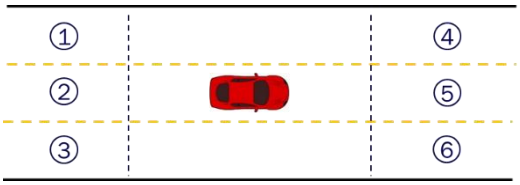


그림 1. 관측가능 공간 정의

그림 1 과 같이 자율주행차가 관측 가능 공간의 총 6 개의 공간으로 나누는데, 각 공간에 위치한 차량과의 상대 속도, 상대 위치 등을 고려하여 자율 주행차와 관계를 정의한다. 예를 들어 1 번위치의 차량이 자율차와 거리가 멀고 (상대거리), 자율주행차에 접근하는 중 (상대속도) 등과 같은 정의를 할 수 있고, 자율차는 6 개의 공간에 대한 상대거리/위치를 한번에 센싱하기 때문에 아래와 같은 메시지의 개념으로 현재 주행상황을 정의 할 수가 있다.

Distance: $d_{TV} \le d_{th}^1 \rightarrow$ close
$d_{th}^1 < d_{TV} \le d_{th}^2 \rightarrow$ nominal
$d_{th}^2 < d_{TV} \rightarrow$ far

Speed: $|v_{TV}| \le v_{th}^1 \rightarrow$ approaching
$v_{th}^1 < |v_{TV}| \le v_{th}^2 \rightarrow$ stable
$v_{th}^2 < |v_{TV}| \rightarrow$ moving away

여기서 d_{TV}, v_{TV} 는 상대 위치 및 속도, d_{th}, v_{th}는 위치 속도에 대한 threshold 를 나타낸다.

즉 한 공간에 상대위치/속도에 대한 3 × 3=9 가지의 조합이 가능하고, 총 6 개의 공간에 대해 정의가 필요하므로 메시지의 경우의 수는 9^6이 된다. 이와 같은 조합이 자율주행차가 센싱할 수 있는 메시지를 나타내는 최대의 수라고 할 수 있고, 모든 교통상황은 이와 같은 메시지로 표현 할 수있다고 가정한다.

다음으로는 임의의 Default Control Policy 를 사용하여 주행시뮬레이터 내에서 Action Governor 를 학습시킨다. 그림 2 가 본 연구에서 제안하는 전체 제어 블록다이어그램이고, Action Governor 가 Default Control Policy 뒤에 추가된 것을 확인 가능하다.

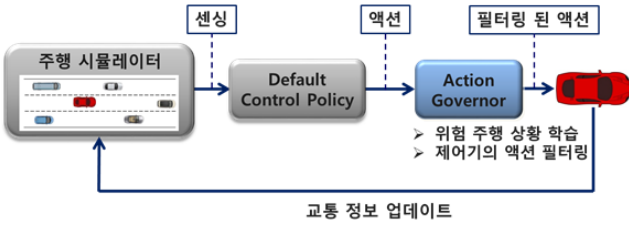


그림 2. Action Governor 가 추가된 제어 블록다이어그램

제안된 Action Governor 를 활용했을 때, 그림 3,4 로부터 어떤 디폴트 제어기에 추가하여도 일정 성능이 향상됨을 확인 할 수있다. 3 차선 고속 도로에서 제한조건을 위배하는 비율이 두 Case 모두에서 향상됨을 확인 가능하다. 두 그래프의 x 축은 주변 차량의 수를 나타내고, 교통량 많을 수록 상호작용이 많이 일어나기 때문에 제한조건 위배 비율이 높아지는 것을 확인 가능하다.

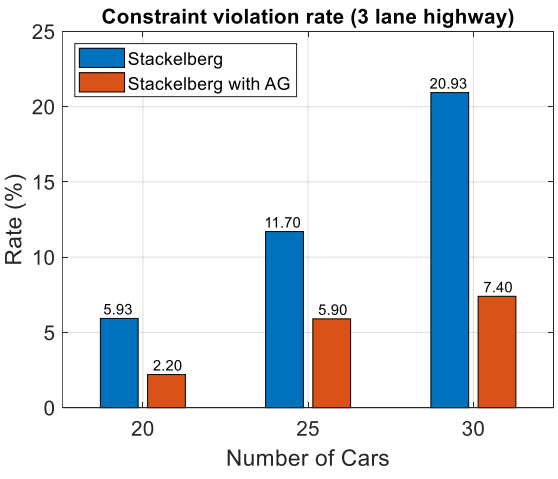


그림 3. Stackelberg 제어기와 Action Governor 결합시 성능 비교

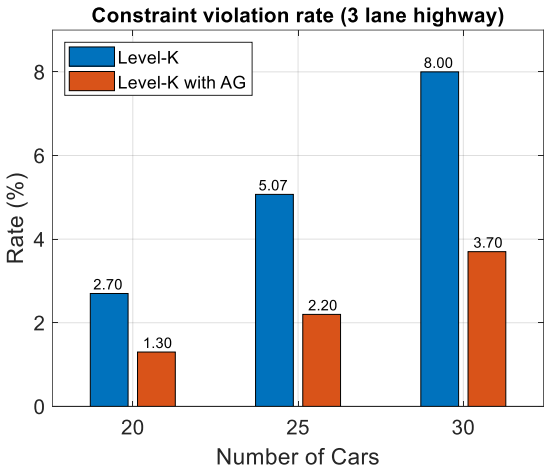


그림 4. Level-K 제어기와 Action Governor 결합시 성능 비교

Ⅲ. 결론

본 논문에서는 디폴트 제어기를 보조하는 개념의 학습기반 추가 모듈을 설계하는 기법에 대해 제안하였다. 기 개발된 주행시뮬레이터 상에서 Action Governor 를 설계하여, 디폴트 제어기 대비 그 우수성을 검증했다. 하지만 현재는 연속 제어 입력을 다루지 않았고 차량 동역학이 고려되지 않았기 때문에, 실제 주행환경과 동일한 학습환경에서 학습이 되어야 하는 점에 위배된다. 향후에는 상용 소프트웨어 등을 활용하여 차량동역학 및 연속제어 입력 등을 고려하여 보다 신뢰성 높은 Action Governor 를 수정/보완 하고자 한다.

ACKNOWLEDGMENT

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2020R1G1A1099830, 2020K1A3A1A39112277)

참 고 문 헌

[1] Li, N., Han, K., Girard, A., Tseng, H.E., Filev, D. and Kolmanovsky, I., 2021. Action Governor for Discrete-Time Linear Systems with Non-Convex Constraints. IEEE Control Systems.

[2] Li, N., Oyler, D.W., Zhang, M., Yildiz, Y., Kolmanovsky, I. and Girard, A.R., 2017. Game theoretic modeling of driver and vehicle interactions for verification and validation of autonomous vehicle control systems. IEEE Transactions on control systems technology, 26(5), pp.1782-1797.

차량 구성 요소 검출을 통한 오클루전 환경에서의 차량 검출에 관한 연구

배지환, 김태경
국방과학연구소

jihan1008@add.re.kr, ktk1501@add.re.kr

A Study on the Enhancement of Car Detection under Occlusion via Auxiliary Car Components Detection

Jihwan Bae, Taekyung Kim
Agency of Defense Development (ADD)

요 약

자율주행 차량의 주변 환경 인식에는 이미지를 이용한 물체 검출기(object detector)가 필수적으로 사용된다. 물체 검출기의 성능은 날이 갈수록 증가하고 있지만, 다양한 주행 상황에서 모든 물체를 강건하게(robust) 인식하지는 못하고 있다. 대표적으로 물체의 일부 또는 대부분이 다른 물체 혹은 기상 환경에 의한 외란으로 가려지는 경우가 있다. 물체의 전체적인 특징을 찾아내지 못 하면 인식률이 극명하게 떨어지는 현상을 PASCAL VOC2007[8] 데이터로 학습한 Faster-RCNN[1] 검출기에서 확인하였다. 본 논문에서는 이러한 오클루전 환경(occlusion environment)에서의 성능 저하를 완화하기 위해 추론 과정에서 차량 구성 요소 검출기를 사용해 추가적인 정보(supervision)를 제공하고, 이를 활용하는 방법론을 제안한다. 차량 구성 요소 검출 네트워크는 차량의 5 가지 구성 요소를 검출하도록 학습되었으며, 추론 과정에 보조 역할로 사용되었을 때 실험적으로 성능 향상을 보임을 관찰하였다.

Ⅰ. 서 론

물체 검출(object detection)은 이미지 상의 시멘틱(semantic) 물체들의 인스턴스(instance)를 검출하는 컴퓨터 비전 기술의 한 갈래이다. 물체 검출은 기본적인 분류 과제(vanilla classification task)의 일반화(generalization)라고 볼 수 있는데, 분류 과제에서처럼 물체 레이블(label)만을 출력하는 것이 아니라, 위치 정보 또한 출력 해 준다는 차이점이 존재한다. 딥러닝의 접목을 통해 물체 검출은 모바일에 탑재되어 있는 카메라의 얼굴 검출 기능부터 자율주행 차량의 도로 위의 상황 판단을 위한 차량 검출까지 다양하게 쓰이고 있다. 물체 검출기는 크게 두 가지 종류로 분류할 수 있는데, 2-스테이지(two stage) 검출기와 1-스테이지(one stage) 검출기이다. 가장 대표적인 2-스테이지 검출기로는 Faster-RCNN[1]이 있다. Faster-RCNN[1] 의 경우, 첫 번째 스테이지에서 지역 제안 네트워크(region proposal network)를 통해 물체 제안(object proposals)을 생성하고, 두 번째 스테이지에서 해당 제안들과 가공된 특징점(cropped feature)을 분류기(classification module)의 입력으로 주게 된다. 1-스테이지 검출기의 대표적 예라고 할 수 있는 YOLO[3]나 SSD[4]의 경우는 지역 제안 네트워크를 사용하지 않고 물체 검출을 위한 학습을 진행한다.

현재 이러한 물체 검출기들의 성능은 끊임없이 발전하여, 추론 속도와 정확도가 굉장히 높다. 때문에 일부 가려져있는 물체나, 주행 상황에서 가드레일이나 나무와 같은 장애물에 가려진 물체에 대한 검출[6]도 원활하게 이루어진다. 그러나 그 가려진 정도(occlusion)가 심할 경우, 물체 검출기의 성능은 확연하게 감소한다. 자율주행 차량이 코너를 돌 때의 시나리오를 상정해보면, 반대쪽에서 들어오는 차량이 주차되어 있는 차량에 가려 극히 일부분 밖에 이미지 상에 드러나지 않으면 물체 검출기에서는 검출이 불가능하다는 것이다.

본 논문에서는 검출하고자 하는 물체들의 오클루전이 심할 때 물체 검출기의 검출률 향상을 위한 보조 검출기와 이를 적용하는 방법론을 제안한다. 오클루전이 많이 일어난 상황이라도 차량의 앞 유리나 헤드라이트 등과 같은 차량 구성 요소에 대한 정보는 이미지 상에 나타날 수 있다[7]. 때문에 차량 구성 요소를 검출할 수 있는 검출기를 경계 상자(bounding box)를 출력하는 최종 추론 과정에 도입하면 물체 검출기가 검출하지 못 하는 물체에 대한 경계 상자 후보군을 확보할 수 있다.

Ⅱ. 본론

차량 구성 요소 검출기는 (1) 전면 라이트 / 후면 라이트, (2) 앞 유리 / 뒷 유리, (3) 측면 유리, (4) 타이어, (5) 차문, 총 5 개의 차량 구성 요소를 검출하도록 학습되었다. 검출기 네트워크로는 YOLOv3[2]를 사용하였고, 미리 학습된(pre-trained) DarkNet[3]의 가중치(weight)를 이용하여 학습을 진행했다. Microsoft Visual Annotation Tool(VoTT)로 100 개의 차량 데이터에 대해 어노테이션(annotation)을 하고, 90 개의 학습용 데이터셋(training dataset)과 10 개의 평가용 데이터셋(testing dataset)으로 분할하였다. 배치 크기(batch size)는 4 로 50 epoch 동안 학습을 시켰다.

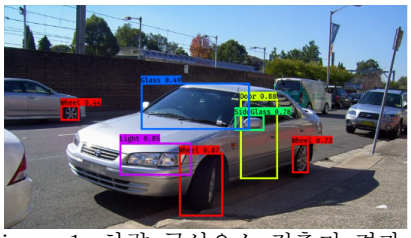


Figure 1. 차량 구성요소 검출기 결과

백본(backbone)으로 사용한 물체 검출기는 Faster-RCNN[1]으로 PASCAL VOC2007 데이터셋으로 학습되어 mAP(mean average precision)가 0.685 인 모델을 사용하였다.

오클루전이 심한 이미지에서 차량을 검출하기 위해 본 논문이 제시하는 방법론은 다음과 같다.

Algorithm 1:

Normal score threshold α
Given Faster-RCNN model $\boldsymbol{F}(\alpha)$
Given Part Detection model $\boldsymbol{P}$
Initialize β, that $\beta < \alpha$
begin:
  Let $\mathbb{R}$ be set of strong candidates: $\boldsymbol{F}(\alpha)$
  Let $\mathbb{R}'$ be set of weak candidates: $\boldsymbol{F}(\beta)$
  $\mathbb{S} \leftarrow \boldsymbol{P}$
  for $r = \{r_1, r_2, \ldots, r_N\} \in \mathbb{R}'$ **do**
    if$\{\exists\, s \subseteq r |$ for all $s \in \mathbb{S}\}$ **then**
      $\mathbb{W} \leftarrow \mathbb{W} + r$
    end if
  end for
  $\mathbb{R} \leftarrow$ merge $\mathbb{R}$ and NMS($\mathbb{W}$)
  return $\mathbb{R}$
end

Algorithm 1. 본 논문의 수도(pseudo)-알고리즘

백본 물체 검출기 F 에서 기존의 역치 값(threshold score) α와 α보다 작은 임의의 β로 얻은 경계 상자 집합을 각각 $\mathbb{R}$, $\mathbb{R}'$로 정의한다. 차량 구성 요소 검출기 P 가 이미지에서의 차량 구성 요소 경계 상자 집합 $\mathbb{S}$ 를 출력하면, $\mathbb{S}$의 원소 s가 하나라도 포함된 $\mathbb{R}'$의 원소(element) r_i를 새로운 경계 상자 집합 $\mathbb{W}$에 추가한다. $\mathbb{R}$ 은 기존 물체 검출기로부터 출력된 후보군(strong candidates) 이고, $\mathbb{W}$는 신뢰성이 떨어지는 후보군(weak candidates) 중 차량 구성 요소를 포함하고 있는 경계 상자 후보들의 집합이다. $\mathbb{W}$에서 NMS(non-maximum suppression)[5]를 적용하여 신뢰성이 떨어지는 원소들을 제거한다. 최종적으로 $\mathbb{W}$와 $\mathbb{R}$ 의 합집합이 본 논문이 제안하는 방법론의 결과물이 된다.

Figure 2 와 Figure 3 는 오클루전이 심하게 발생한 실제 환경 이미지에서 실험한 결과이다. 각각 좌측부터 백본 검출기의 검출 결과, 차량 구성 요소 검출기의 검출 결과, 본 논문의 방법론을 적용한 검출 결과이다. Figure 2 에서 백본 검출기로 검출되지 않던 가장 가까운 하얀색 차량이 본 논문의 방법론을 통해 검출됨을 확인할 수 있다. Figure 3 에서는 백본 검출기에서 물체가 단 하나도 검출되지 않았으나, 차량 구성요소 검출기에서 후면라이트가 검출되었고, 최종적으로 본 논문이 제시하는 방법론을 통해 검출할 수 있음을 확인하였다.

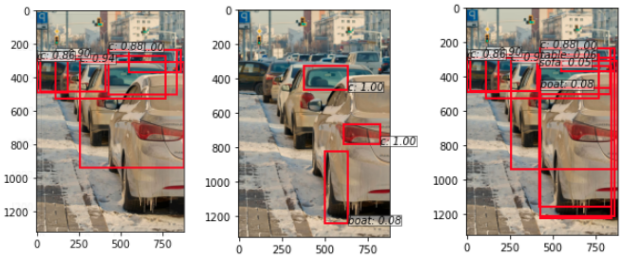


Figure 2. 실제 이미지에서의 실험 결과 - 1

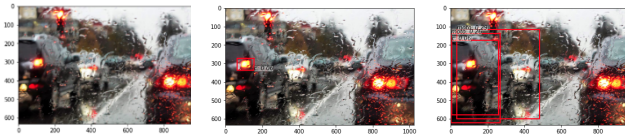


Figure 3. 실제 이미지에서의 실험 결과 - 2

Ⅲ. 결론

본 논문에서는 오클루전이 심하게 발생했을 때의 물체 검출기 정확도 향상을 위해 보조적으로 차량 구성 요소 검출기를 추론 과정에 적용하는 방법론을 제시하였다. 실제 환경에서 백본 물체 검출기가 실패했던 물체 검출을 해당 방법론을 통해 검출할 수 있음을 실험적으로 관찰하였다. 자율주행 차량이 도로 주행에서 겪을 수 있는 시나리오에 대한 연구로서, 주행 성능 향상에 도움이 될 수 있을 것이라 기대는 바이다.

ACKNOWLEDGMENT

본 연구는 국방과학연구소 사업의 일환으로 수행되었음.

참 고 문 헌

[1] Shaoqing Ren, Faster-RCNN: Towards Real-Time Object Detection with Region Proposal Networks, NIPS, 2015

[2] Joseph Redmon, YOLOv3: An Incremental Improvement, arXiv, 2018

[3] Joseph Redmon, You Only Look Once: Unified, Real-Time Object Detection, CVPR, 2016

[4] Wei Liu, SSD: Single Shot MultiBox Detector, ECCV, 2016

[5] Jan Hosang, Learning non-maximum suppression, CVPR, 2017

[6] Haochai Zhang, Towards Adversarially Robust Object Detection, ICCV, 2019

[7] MengMeng Xu, Missing labels in Object Detection, CVPR Workshop, 2019

[8] Mark Everingham, The Pascal Visual Object Classes(VOC) Challenge, IJCV, 2009

ROS 기반 2륜 차량 시스템을 위한 딥러닝 기반 Monocular Visual Odometry 적용

최병찬, 남해운

한양대학교

luwis93@hanyang.ac.kr, hnam@hanyang.ac.kr

Implementation of Deep Learning-based Monocular Visual Odometry on ROS-based 2WD system

Byung Chan Choi, Haewoon Nam

Hanyang University

요 약

본 논문은 Recurrent Neural Network (RNN) 기반 Monocular Visual Odometry 기법인 DeepVO를 ROS 기반 2륜 차량 시스템과 실내 주행 상황에 적용하기 위한 구현 방법을 제안한다. RNN 기반 딥러닝 네트워크인 DeepVO에 Monocular Visual Odometry를 학습시키기 위해 딥러닝용 GPU 서버 컴퓨터에서 KITTI 데이터셋을 사용하여 학습을 진행하였다. 학습된 네트워크를 별도의 ROS Host PC에 탑재한 후 ROS 기반 2WD 차량에서 전송되는 연속되는 실내 주행 이미지로부터 Odometry 연산을 수행하여 차량의 실내 위치 추정을 수행하였다. 본 논문은 주행 데이터셋에서 좋은 성능을 내는 딥러닝 기반 Visual Odometry 기법을 실제 주행 시스템에 적용한 결과와 구현 과정에서 발생하는 문제점을 파악하는 것을 목표로 하였다.

Ⅰ. 서 론

최근 딥러닝 기술의 급격한 발전으로 차량 및 로봇 위치 추정의 핵심 기술인 Visual SLAM과 Visual Odometry에 딥러닝을 적용하려는 연구가 활발히 진행되고 있다.[1][2][3] 이러한 딥러닝 기반 기법에 대한 활발한 연구는 딥러닝의 일반화 특성에 대한 높은 기대치에서 비롯된 것이라 할 수 있다.

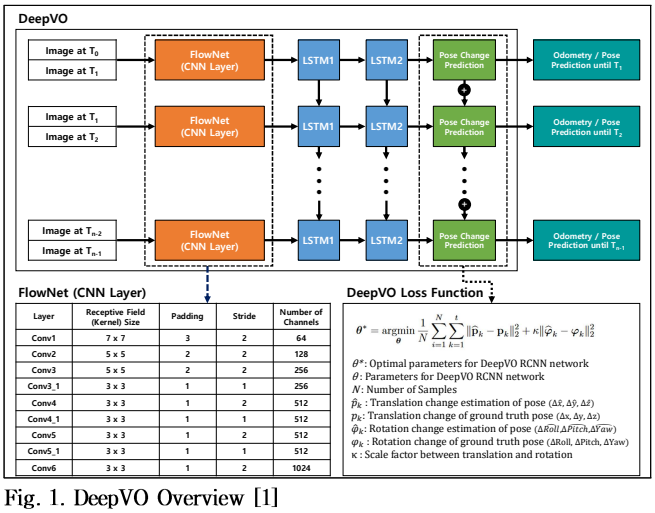
고전적인 Visual Odometry는 Feature Tracking, Point Cloud Tracking, Multi-view Geometry를 통해 카메라의 Pose 변화를 추정하고 이를 누적하여 이동 경로를 예측했다.[4][5][6] 하지만 기존의 방식은 Feature 추출 결과에 매우 의존적이다. 이로 인해 주행 환경 변화에 따른 Parameter Fine Tuning이 요구된다. Bundle Adjustment와 같은 최적화 기법을 사용하여 평균 Pose 추정 오차를 최소화시키지만 여러 주행 환경과 시나리오를 위한 일반화 특성을 확보하는 데에는 제한적이다.

딥러닝 기반 Visual Odometry는 기존의 방식과 달리 대용량의 주행 데이터셋에서 전방 이미지와 위치 정보 사이의 관계를 표현할 수 있는 모델을 Deep Neural Network 학습을 통해 얻어냄으로서 Visual Odometry와 관련 기능을 구현한다.[1][2][3] 다양한 광원 변화, 물체 이동, 회전 시나리오 등을 포함한 데이터셋에서 주행 이미지와 위치 정보 간의 관계를 표현할 수 있는 일반화된 모델을 도출하기 때문에 Robustness가 고전적인 방식보다 좋은 편이다.

대부분의 연구는 KITTI, nuScenes, CARLA와 같은 주행 데이터셋 내에서 학습 및 성능 평가에 집중하고 있다. 그러나 학습된 딥러닝 네트워크를 실제 로봇 또는 차량에 탑재하여 실 주행 상황에서의 성능 평가는 상대적으로 저조한 편이다. 본 논문에서는 [1]에 제시된 방법으로 RNN 기반의 Monocular Visual Odometry 네트워크 DeepVO를 구현하고 실제 2WD 소형 차량 Odometry에 탑재하기 위한 구현 과정, 실내 주행 성능 및 제한사항을 파악하는 것에 중점을 두었다.

Ⅱ. DeepVO 개요

DeepVO는 2017년 IEEE International Conference on Robotics and Automation (ICRA) 에서 발표된 딥러닝 기반 Visual Odometry 기법이다.[1] DeepVO는 연속적으로 입력되는 이미지 데이터를 시계열 데이터 (Sequential Data)로 취급하여, Monocular Visual Odometry 문제를 RNN 네트워크를 통해 해결하였다는 점이 특징이다. 연속적인 이미지로부터 Optical Flow를 추정할 수 있는 FlowNet과 시계열 데이터를 처리할 수 있는 RNN를 합친 Deep RCNN을 제시하여 KITTI 데이터셋에서 고전적인 Visual Odometry 기법보다 Robust하고 정확한 결과를 제시하였다.[1][2]



FlowNet (CNN Layer)

Layer	Receptive Field (Kernel) Size	Padding	Stride	Number of Channels
Conv1	7 x 7	3	2	64
Conv2	5 x 5	2	2	128
Conv3	5 x 5	2	2	256
Conv3_1	3 x 3	1	1	256
Conv4	3 x 3	1	2	512
Conv4_1	3 x 3	1	1	512
Conv5	3 x 3	1	2	512
Conv5_1	3 x 3	1	1	512
Conv6	3 x 3	1	2	1024

DeepVO Loss Function

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{t} \|\hat{p}_k - p_k\|_2^2 + \kappa \|\hat{\varphi}_k - \varphi_k\|_2^2$$

θ^*: Optimal parameters for DeepVO RCNN network
θ: Parameters for DeepVO RCNN network
N: Number of Samples
$\hat{p}_k$: Translation change estimation of pose ($\Delta\hat{x}, \Delta\hat{y}, \Delta\hat{z}$)
p_k: Translation change of ground truth pose ($\Delta x, \Delta y, \Delta z$)
$\hat{\varphi}_k$: Rotation change estimation of pose ($\Delta\hat{Roll}, \Delta\hat{Pitch}, \Delta\hat{Yaw}$)
φ_k : Rotation change of ground truth pose (ΔRoll, ΔPitch, ΔYaw)
κ : Scale factor between translation and rotation

Fig. 1. DeepVO Overview [1]

DeepVO는 연속되는 이미지 2개를 쌓아서 네트워크에 입력하고, 목표값으로 각 이미지에 해당되는 시간의 Pose 변화량을 학습하게 한다. 그리고 Long Short Term Memory (LSTM)를 통해 현재 시간대에서 추정한 정보를 다음 시간대에 전달하여 연속적인 Pose 변화량 추정을 할 수 있도록 한다. Pose에서 Translation 변화량인 Δx, Δy, Δz와 Rotation 변화량인 ΔRoll, ΔPitch, ΔYaw는 서로 단위와 범위가 다른 것을 감안하여 Loss Function을 Translation Error, Rotation Error를 분리하고 Weight를 적용한 Weighted Loss Function을 사용한다.

Ⅲ. 주행 시나리오 데이터셋을 이용한 RNN 학습

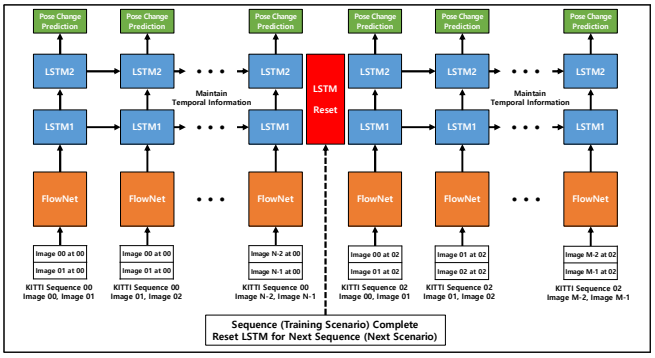


Fig. 2. Stateful RNN Training for DeepVO

DeepVO와 같은 RNN 구조의 딥러닝 네트워크를 학습시키기 위해서는 해당 RNN 문제가 Stateless case인지 Stateful case인지 확인해야한다. Stateless case의 경우 각 Batch가 독립적인 시나리오기이기 때문에 다음 Batch에 영향을 주지 않는다. Stateless RNN을 학습할 때에는 Batch마다 LSTM을 초기화하며 데이터셋을 Shuffling한다. [7] 그러나 Stateful case의 경우에는 각 Batch가 전체 시나리오의 일부이기에 다음 Batch에 영향을 준다. 이로 인해 각 Batch에 할당된 시나리오가 독립적이지 못하다. Stateful RNN을 학습할 때에는 이전 Batch에서 배운 정보의 흐름을 유지하기 위해 시나리오 기준으로 LSTM을 초기화하고 데이터셋을 Shuffling하지 않고 순차적으로 제공한다.[7]

KITTI와 같은 주행 데이터셋에서 학습할 경우 Batch 단위로 가져오는 데이터는 각 주행 시나리오의 일부이기 때문에 하나의 주행 시나리오가 끝나기까지 Stateful RNN case로 취급하여 DeepVO를 학습해야한다. 본 논문에서는 LSTM이 KITTI 데이터셋 시나리오 1개 완료마다 초기화되게 학습하였다.

Ⅳ. DeepVO ROS 기반 시스템 탑재

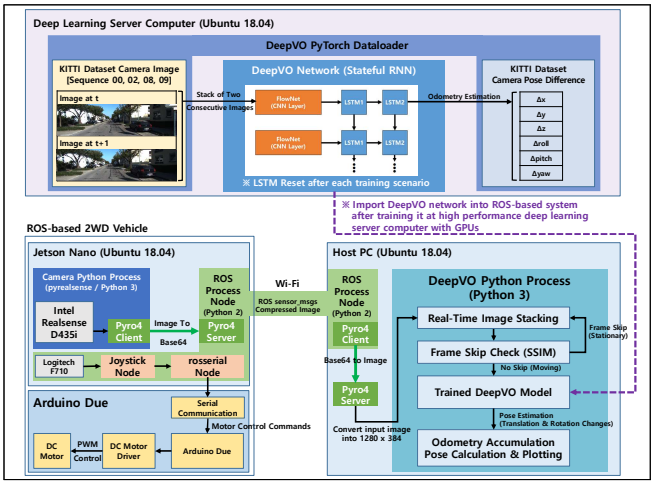


Fig. 3. System Architecture for DeepVO on ROS-based 2WD System

본 논문에서는 Nvidia 2080 Ti가 장착된 딥러닝용 GPU 서버 컴퓨터에서 KITTI 데이터셋과 PyTorch를 이용하여 DeepVO 네트워크를 학습시킨 후 ROS Host PC에 탑재하였다. ROS Host PC는 ROS 2WD 차량에서 연속적으로 전달하는 이미지를 입력받아 학습된 네트워크에 전달하여 Pose 변화량을 추정하고, 결과값을 누적해서 차량의 이동거리와 위치를 추정한다. Python3 기반 딥러닝 프로세스와 Python2 기반 ROS1 사이의 데이터 교환을 구현하기 위해 프로세스간 데이터 교환 라이브러리인 Pyro4를 사용하여 ROS 네트워크와 병행으로 작동하는 Python 프로세스 네트워크를 구성하였다.

DeepVO만 사용해서 위치를 추정할 경우 정지 상황과 후진 상황에서 위치 추정 결과가 발산하거나 추정이 틀리는 것을 볼 수 있다. 왜냐하면 학습에 사용한 KITTI 데이터셋에는 후진과 정지 상황이 전진 주행 보다 적게 배정되어있기 때문이다.

정지 상황에서 위치 추정이 발산하는 것을 막기 위해서 Structural Similarity Index Measure (SSIM)을 이용하여 연속된 이미지가 동일한지 여부를 파악하고 동일할 경우 정지 상태인 것을 판단하는 Frame Skip을 도입해야한다. Frame Skip 기능을 도입함으로서 정지 상태에서 불필요한 Pose 추정 오차가 누적되는 것을 방지하여 전체 시스템을 안정화 시킬 수 있다. 후진 상황을 학습하기 위해서는 데이터셋을 반대로 실행하여 학습시키는 방법이 있으나 학습 소요시간이 증가한다는 단점을 내포하고 있다.

Ⅴ. 최종 구현 결과 및 실험 결과

KITTI 데이터셋에서 장거리 데이터셋 00, 02, 08, 09를 Training Set으로 설정하여 전진, 우회전, 좌회전 등 다양한 주행 시나리오를 딥러닝 네트워크에 학습시켰다. 그리고 Validation으로 비교적 단거리 데이터셋인 01, 03, 04, 05, 06, 07, 10을 선정하였다. 사용된 이미지 데이터의 비율은 Training 67.59%, Validation 32.41%이다. 네트워크 학습을 위해 Learning Rate는 0.001로 설정했으며, Adagrad Optimizer를 사용하였다. Pre-trained된 FlowNet 대신 학습되지 않은 초기 CNN 모델을 사용하여 학습의 전 과정을 관찰하였다. 학습을 통해 생성된 DeepVO 모델을 ROS 기반 2륜 차량의 Host PC에 탑재하였으며, 실내 주행하는 2륜 차량의 Odometry 추정을 수행하였다.

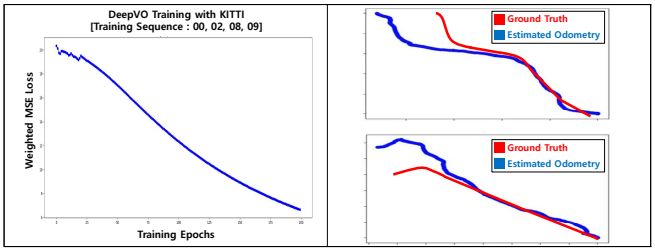


Fig. 4. Training Loss (Left) / Indoor Driving Test (Right)

Pre-trained된 FlowNet이 아닌 순정 상태의 초기 CNN 네트워크를 사용하기 때문에 Loss가 완만하게 감소하면서 Pose 변화량 추정이 학습되는 것을 볼 수 있다. 실내 주행 결과 Pose 추정 결과가 Ground Truth와 전반적으로 형태는 유사하나 4륜 Ackermann 구조의 차량 조향을 기준으로 학습된 모델이 2륜 차량의 Rotation Estimation과 Scale Estimation에 대해 부족한 면모를 보이고 있다. 그리고 4륜 Ackermann 구조 차량은 회전을 위해 전진이 병행되기에 데이터셋에서도 대부분 Pose 변화량에 전진이 반영되어있고 그에 맞춰서 네트워크가 학습되었다. 이로 인해 Pose 변화량을 4륜 차량 구조와 같이 전진이 병행된 형태로 추정하려는 경향이 있다.

Ⅵ. 결론

딥러닝은 여러 연구에서 다양한 주행 시나리오 학습을 통한 일반화된 Visual Odometry 모델을 도출하여 데이터셋에서 고전적인 Visual Odometry 기법보다 Robust한 결과를 보여주었다. 그러나 딥러닝 학습은 데이터셋의 구성에 의존적이기 때문에 데이터셋에 포함되지 않은 시나리오에 대해 취약한 점을 보인다. 그리고 딥러닝 기반 Visual Odometry는 장착한 차량 또는 로봇의 Motion Model, 사용하는 카메라의 Instrinsic Parameter 등을 암묵적으로 학습하기 때문에 사용 HW나 카메라 해상도가 변경될 경우 성능이 저하될 여지가 있다. 이를 보완하기 위해 SSIM와 같은 고전적인 영상처리 기법을 딥러닝 네트워크와 연계하여 시스템을 구현할 수 있다.

ACKNOWLEDGMENT

이 논문은 2020년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2019M3F6A1106108).

참 고 문 헌

[1] S. Wang, R. Clark, H. Wen and N. Trigoni, "DeepVO: Towards end-to-end visual odometry with deep Recurrent Convolutional Neural Networks," 2017 IEEE International Conference on Robotics and Automation (ICRA), Singapore, 2017, pp. 2043-2050

[2] A. Dosovitskiy et al., "FlowNet: Learning Optical Flow with Convolutional Networks," 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, 2015, pp. 2758-2766

[3] R. Wang, S. M. Pizer and J. Frahm, "Recurrent Neural Network for (Un-)Supervised Learning of Monocular Video Visual Odometry and Depth," 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, pp. 5550-5559

[4] D. Scaramuzza and F. Fraundorfer, "Visual Odometry [Tutorial]," in IEEE Robotics & Automation Magazine, vol. 18, no. 4, pp. 80-92, Dec. 2011

[5] R. Mur-Artal, J. M. M. Montiel and J. D. Tardós, "ORB-SLAM: A Versatile and Accurate Monocular SLAM System," in IEEE Transactions on Robotics, vol. 31, no. 5, pp. 1147-1163, Oct. 2015

[6] M. Labbé, F. Michaud. "RTAB Map as an Open Source Lidar and Visual Simultaneous Localization and Mapping Library for Large-Scale and Long-Term Online Operation," Journal of Field Robotics, vol. 36, no. 2, pp. 416 - 446, Mar. 2019

[7] M. A. Yilmaz and A. Murat Tekalp, "Effect of Architectures and Training Methods on the Performance of Learned Video Frame Prediction," 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 2019, pp. 4210-4214

셀룰러 V2X 시스템을 이용한 자율 주행 시나리오

윤영진, 김지훈, 한동석
경북대학교 대학원 전자전기공학부
dshan@knu.ac.kr

Autonomous Driving Scenario Using Cellular V2X System

Young Jin Yoon, Jihun Kim, Dong Seog Han
Kyungpook National University

본 논문은 3GPP(3rd Generation Partnership Project) LTE(Long Tem Evaluation)에서의 차량(vehicle) 통신에 대한 표준화 실험으로서, LTE Release 14 표준에서 논의되었거나 논의 중인 V2X(Vehicle to everything)에 대한 표준화 실험이다. 본 실험에서는 LTE와 GPS를 연동하여 차량의 위도, 경도를 실시간으로 받고 그 위도, 경도를 기준으로 서버와 통신하여 적절한 미션을 수행하였다. 미션 포함내용은 출발지의 위도, 경도 도착지의 위도, 경도를 통신을 통해 받음으로써 미션을 진행하는 데 문제가 없었다. 본 논문에서는 LTE C-V2X 시스템을 이용하여 메시지 규격을 사전에 정의함으로써 정확한 통신 성능을 보이는 것과 효율적인 설계를 통해서 GPS와 LTE를 병합하여 효과적인 통신 방식을 확인하였다.

Ⅰ. 서론

자율 주행에 필요한 3가지 핵심 기술로는 센서, 제어, V2X 통신을 들 수 있다. 그중 V2X(Vehicle-to-Everything) 통신에서는 운전 중 도로 인프라 및 다른 차량과 통신하면서 교통상황 등의 정보를 교환하거나 공유하는 통신을 의미한다. V2X 통신은 차량과 차량 간의 통신인 V2V (Vehicle-to-Vehicle), 차량과 보행자 단말 간의 통신인 V2P(Vehicle-to-Pedestrian), 차량과 도로변 유닛(Roadside Unit RSU) 간의 통신인 V2I(Vehicle-to-Infrastructure)를 포함하고 있다[2]. 본 논문에서는 사전에 만들어진 규격을 통해 LTE V2X를 실제 차 테스트를 하여 통신 성능을 분석한다.

Ⅱ. 본론

본 논문에서는 사전에 규격이 정해진 V2X 서버와 통신하는 과정을 보여준다. 실험에서는 실제 차량과 OBU, GPS를 설치하여 진행하였다. 그림 1. 에서는 실제 차량(기아 쏘울) 차량의 트렁크 상단에 다음과 같이 설치하였다. 그림 2. 에서는 사전에 규격 된 프로토콜을 이용하여 시스템을 구성하였다. 먼저 자율차 장치에서 TCP/IP 연결을 요청한다. 이것은 통신 방식을 맞추기 위해서 구성되어 진 것이다. 다음은 서버로 시스템 승인 요청을 진행한다. 그리고 서버에서는 사전 규격 된 프로토콜이 규격과 일치하면 자율차 장치에 승인이 완료되었다고 신호를 송신한다. 그리고 자율차에서는 차량의 위치를 보낸다. 차량의 위치는 GPS의 위도, 경도를 받아서 V2X 서버로 보낸다. 그러면 V2X 서버에서는 미션을 준다. 미션의 내용은 출발지의 위도, 경도 도착지의 위도, 경도를 줌으로써 자율차 장치에게 출발, 도착 위치를 제공함으로써 LTE 서버의 임무를 수행한다. 미션의 내용 및 미션 선택은 그림 3에서와같이 표현된다.
자율차가 서버로부터 미션을 받은 후 자동차 내부에서는 스스로 미션을 선택할 수 있는 알고리즘을 개발하여 직접 차량이 수행 가능한 알고리즘을 선택한 후 서버에게 미션 번호를 전송한다. 전송 후 서버에서는 다른 접속자가 사용하고 있다면 미션 수행 불가 메시지를 보낼 것이고 그렇지 않다면 미션 수행 가능 메시지를 보낸다. 그림 4에서는 차량 내부 서버와의 V2X 서버 관계도를 표현하였다.

그림 1 실제 차량에 설치된 WAVE 모듈(WAVE)

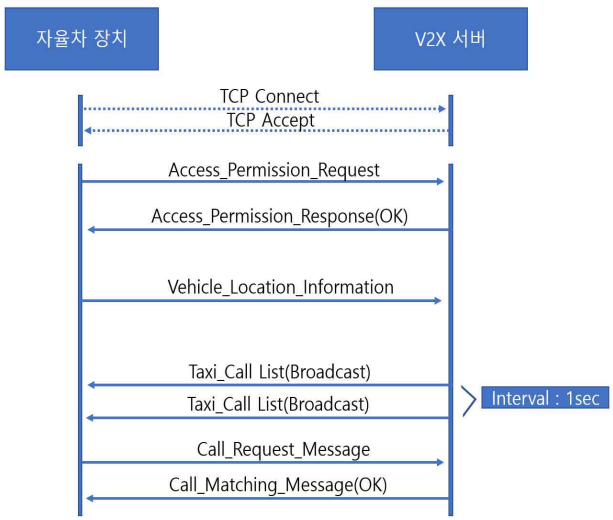


그림 2 V2X 전체 시스템 개략도

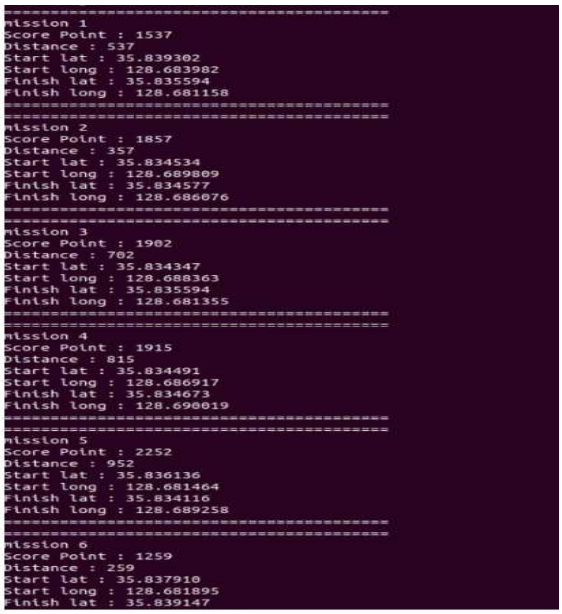


그림 3 LTE 미션 리스트

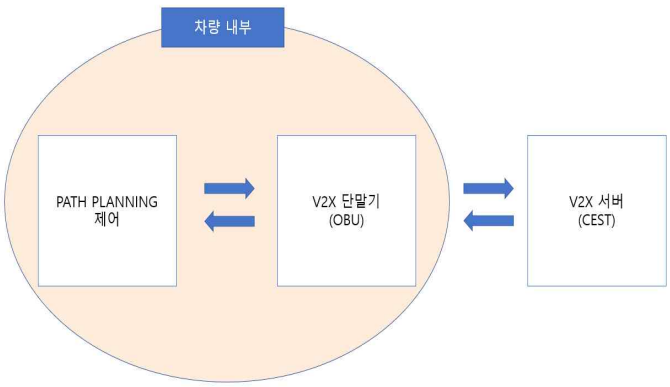


그림 4 차량 내부와 V2X 서버와의 통신

Ⅲ. 결론

본 논문에서는 자율 주행에 있어서 필요한 3가지 핵심 기술인 V2X 통신에서 셀룰러 데이터 기반의 LTE 통신을 테스트하여 성능을 테스트하였다. LTE 서버와 통신을 할 때 판단 알고리즘을 넣음으로써 서버에서 수행불가 메시지를 보내었을 때 차량 내부에서 예외처리를 함으로써 완성도를 높였다. 앞으로는 5G C-V2X를 이용하여 다양한 통신을 수행하여 더욱더 의미 있는 연구가 필요한 것으로 판단한다.

ACKNOWLEDGMENT

본 연구는 산업통상자원부와 한국산업기술진흥원이 지원하는 5G기반 자율주행 융합기술 실증 플랫폼 과제(과제고유번호 : 1415169669)의 지원을 받아 수행하였습니다."

참 고 문 헌

[1] Chen, Shanzhi, et al. "Vehicle-to-everything (V2X) services supported by LTE-based systems and 5G." IEEE Communications Standards Magazine 1.2 (2017): 70-76.

[2] Abboud, Khadige, Hassan Aboubakr Omar, and Weihua Zhuang. "Interworking of DSRC and cellular network technologies for V2X communications: A survey." IEEE transactions on vehicular technology 65.12 (2016): 9457-9470.

[3] Molina-Masegosa, Rafael, and Javier Gozalvez. "LTE-V for sidelink 5G V2X vehicular communications: A new 5G technology for short-range vehicle-to-everything communications." IEEE Vehicular Technology Magazine 12.4 (2017): 30-39.

[4] Chen, Shanzhi, et al. "LTE-V: A TD-LTE-based V2X solution for future vehicular network." IEEE Internet of Things journal 3.6 (2016): 997-1005.

[5] Schünemann, Björn. "V2X simulation runtime infrastructure VSimRTI: An assessment tool to design smart traffic management systems." Computer Networks 55.14 (2011): 3189-3198.

[6] Hobert, Laurens, et al. "Enhancements of V2X communication in support of cooperative autonomous driving." IEEE communications magazine 53.12 (2015): 64-70.

V2X 기반 군집주행 차량과 주변 V2X 통신 차량간 통신 영향성 연구

구자후, 한규동, 정홍종, 권순일
주식회사 웨이티즈

{jahu.ku, gyudong.han, hj, steve}@wayties.com

요 약

본 논문은 WAVE/DSRC 기반의 V2X 통신성능 시험이 개발 차종 내의 통신에만 치중되어 차종이 혼재되어 있는 실도로에서의 상호 영향성에 대한 부족한 연구를 해소하기 위해 V2X 통신을 수행중인 자율/협력주행 소형차와 군집주행중인 대형차(트레일러)가 혼재하는 자동차 전용도로를 가정하고, 시험 차량의 배치와 통신 혼잡도를 조절하여 다양한 통신 환경에서의 차량 그룹간 통신성능을 측정 후 분석 결과 및 시험 중 특이사항을 기술하고 있다.

Ⅰ. 서 론

최근 V2X 통신을 활용한 자율협력주행 기술은 소형 승용차만이 아닌 대형 화물차 및 버스에도 적용하기 위한 연구가 이루어지고 있다. 그러나 도로상의 승용차와, 버스, 화물차가 동시에 동일 도로상에서 자율/협력주행을 위해 V2X 통신을 수행 시의 상호영향성에 대한 조사가 부족한 실정이다. 본 연구에서는 근간의 연구를 바탕으로 고속도로에서 군집을 형성하고 주행중인 대형차 그룹과 V2X 서비스를 이용중인 소형차 그룹을 설정하고, 이 두 그룹이 실 도로 상에서 겪을 수 있는 다양한 상황을 재현, 통신성능 및 상호영향성을 분석하고자 한다.

Ⅱ. 시험 설계

대형 화물차의 운행차로는 도로의 최 외곽 차선이기에 일반 소형차는 항상 트럭의 좌측에 위치하며, 각 차선에 위치한 차량들의 평균 주행속도는 국내 고속도로 교통법에 정의된 최대속도에서 10km/h 씩 안전구간을 두고 다음과 같이 설정하였다. (승용차 100km/h, 대형차 80km/h). 군집을 형성하고 80km/h 의 속도로 이동중인 군집대열의 차간 안전거리(시간)는 1 초로 설정하였으며, 이 시간은 주행속도로 계산 시 22.2m 에 달한다. 소형차 그룹은 군집주행처럼 짧은 차간거리를 유지할 필요가 없으므로 평균 100km/h 로 주행 시 2 초의 차간거리로 설정하며 이는 55.6m 에 달한다. 본 시험에서는 한 대의 소형차량에 다수의 단말 및 안테나를 세트로 설치하고 단말의 출력 주기를 조정하여 방식으로 실제 소형차를 다수 배치하는 것과 유사한 통신 환경을 구현하였다.

표 1. 시험에서 사용된 차량 및 단말

	대형차 1 (T1)	대형차 2 (T2)	소형차 1 (V1)	소형차 2 (V2)
모델명 (전장)	엑시언트 (트레일러 포함 18.2m)		아이오닉 (4.47m)	소나타 (4.9m)
단말 수	1	1	6	6

그림 1. 시험에서 사용된 안테나 및 차량 별 배치
(좌: 대형차 1, 2 측면거울 내장, 우상: 소형차 1, 우하: 소형차 2)

일반적으로 차량 간 통신환경은 LOS(Line of Sight)로 표현할 수 있다. 통신환경에서의 LOS 여부는 확보된 프레넬 영역(Fresnel Zone)[3]의 정도로 판단되므로 이 시험에서는 소형차 V1, V2 사이의 무선 통신환경을 LOS, Near-LOS, Non-LOS 의 세 영역으로 구분하여 정의하였다. (대형차는 세 시험 모두 LOS 를 만족한다.)

표 2. 프레넬 영역에 따른 통신환경 및 시험 정의

시험 명	LOS	Near-LOS	Non-LOS
프레넬 영역	100%	0~60%	0%

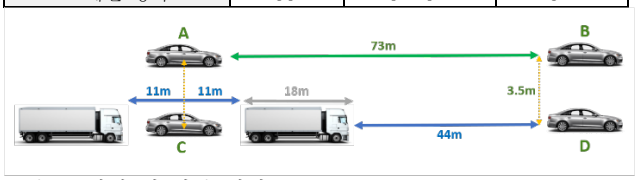


그림 2. 시험 별 차량 위치

표 3. 시험 별 배치된 소형차 위치(대형차 T1, T2 위치 고정)

	LOS	Near-LOS	Non-LOS
소형차 1 (V1)	B	D	D
소형차 2 (V2)	A	A	C

표 4. 시험 별 배치된 차량 그룹 간 LOS 환경

	LOS	Near-LOS	Non-LOS
T1 ↔ T2	LOS	LOS	LOS
T1 ↔ V1	LOS	LOS	LOS
T1 ↔ V2	LOS	LOS	Non
T2 ↔ V1	Near	Non	Non
T2 ↔ V2	LOS	LOS	LOS
V1 ↔ V2	LOS	Near	Non

대형차들이 주고받는 군집 상태 메시지(PSM: Platoon Status Message)와 그 외 차량들이 주고받는 일반 안전메시지(BSM: Basic Safety Message)는 아래와 같이 정의되었다. [표 5]

표 5. 시험 중 주고받는 V2V 메시지 설정 값

	채널	Tx 세기	패킷 길이	송신 주기
PSM	172	20 dB	400 Byte	50 Hz
BSM	172	20 dB	400 Byte	10, 20, 50 Hz

각 시험은 변인이 변경될 때 마다 반복적으로 최대 3 회 시행되었으며 각 시행마다 모든 단말이 N:N 으로 60 초간 주고받은 메시지의 개수를 기록, 총 예상 송신 메시지 수 대비 수신 메시지 수로 1 분간의 PER 을 계산하였다.

Ⅲ. 그룹간 V2V 통신 영향 분석

1. 군집 대형차 그룹 → 소형차 그룹

소형차의 주력 V2X 통신은 후방의 운전자가 확인 불가능한 전방 추돌 방지(FCW: Forward Collision Warning)나 비상제동경고(EEBL: Emergency Electric Brake Light)등의 대표적인 V2X 서비스를 제공하기 위해 해당 메시지들이 진행방향 전방에서부터 후방에 있는 차량들에게 안정적으로(PER 10% 이하) 전송되어야 한다.

LOS 환경의 소형차 그룹 간의 PER 은 최대 4.33%로 관측되었다. 이는 1 대의 차량 위에 여러 대의 안테나를

배치한 실험의 특성 상 일부 안테나의 프레넬 영역이 완벽히 100%를 확보하지 못한 것으로 보이며, 실제 측정 데이터에서도 특정 위치에 배치된 일부 안테나의 성능 저하만 관측되었다. Near-LOS 시험에서는 해당 환경을 구축하기 위해 V1 그룹이 군집 주행중인 T1 그룹의 전방으로 끼어들어 V1, V2 간의 프레넬 영역이 60% 이상 확보되지 못하는 상황을 구현하였으며, 이 때의 소형차 그룹 간 PER 은 최대 8.81% 로 LOS 환경에서의 시험에 비해 다소 성능 하락 경향을 보였다. V1 그룹이 군집 주행중인 T1 의 전방에 끼어들고, 후방의 V2 그룹이 군집 주행중인 T1, T2 사이에 들어가가 V1, V2 간의 프레넬 영역이 T1 에 의해 완벽히 가려지는 Non-LOS 시험의 PER 은 최대 80.91%로 FCW, EEBL 등의 V2X 서비스 요구 PER 10%를 넘는 결과를 보였다. 이는 일반적인 Non-LOS 상황에서의 통신성능 저하로 판단된다. 결론적으로 군집주행중인 대형차에 의한 소형차 V2V 통신 영향은 대형차에 의한 통신 혼잡보다는 장애물로서 소형차 간 LOS 환경에 간섭할 때 크게 영향을 받는다. [그림 3]

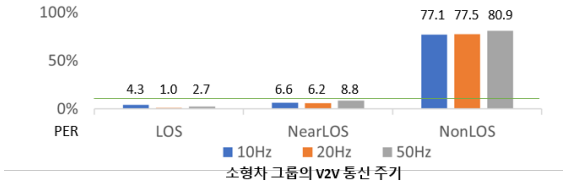


그림 3. 군집주행중인 대형차 그룹의 V2V 통신(50Hz)에 의한 소형차 그룹의 통신 주기 별 V2V 통신 영향 (V1 → V2)

2. 소형차 그룹 → 군집 대형차 그룹

군집주행중인 대형차 간의 군집통신성능은 소형차의 방해에 의해 PER 이 증가하는 경향을 보이긴 하였으나 T1 이 V1 과 V2 사이에 끼어 있는 Non-LOS 시험에서 각 소형차가 각각 6 개의 단말에서 50hz 로 T1 전후방 30 대씩의 방해 환경을 구축했음에도 불구하고 T1 과 T2 사이(V2→V1)의 최대 PER 이 8.14%로 관측되었다. 이는 대형차 그룹의 차량 높이에 의한 상호 LOS 환경 상시 확보로 인한 성능 우세로 보이나, 군집 그룹 간 전달되는 PSM 메시지의 군집 서비스 요구 PER 기준이 FCW, EEBL 등의 서비스 요구 PER 인 10%보다 엄격해질 경우 성능적 고려가 필요해 보인다. [그림 4]

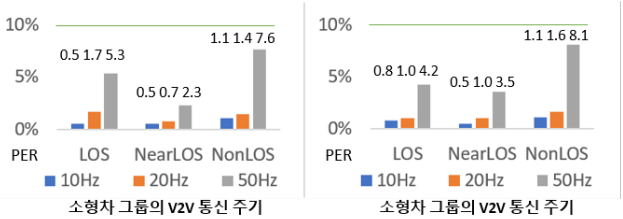


그림 4. 소형차 그룹의 V2V 통신 주기에 의한 군집주행중인 대형차 그룹의 V2V 통신(50Hz) 영향 (좌: V1 → V2, 우: V2 → V1)

그 외 특이할 만한 상황으로, 정의된 모든 시험에서 소형차 V1 은 대형차 T1 에 대해 항상 LOS 환경에서 통신을 주고받았지만 소형차 그룹의 단말 12 대에 의한 10Hz 방해 메시지 송출이라는 비교적 양호한 방해환경에서도 해당 그룹 간 PER 이 10%가 넘는 경우가 있었다. [그림 5] 이는 주행중인 T1 을 앞뒤로 둘러싼 V1, V2 가 중간에 위치한 T1 에 의해 서로를 숨겨진 노드로 판단하여 무선 충돌 회피를 시행하지 않고 메시지를 보낸 결과로, 대형차 T1 을 기준으로 채널 혼잡도가 높은 영역이 만들어져 해당 영역에 존재하는 모든 차량의 통신성능에 영향을 미치는 것으로 추측된다.

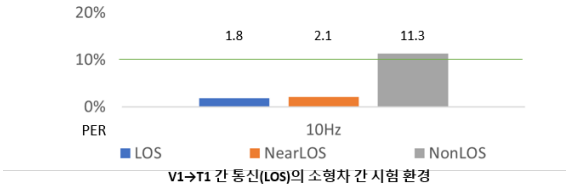


그림 5. 소형차간 Non-LOS 시험 시, V1→T1(LOS)의 성능저하

이는 V2X 단말이 충분히 보급된 후의 공도에서 V2X 통신을 하고 있는 소형차 그룹이 중간에 위치한 장애물(대형차)에 의해 양분되어 서로 Non-LOS 환경에 처할 경우, 해당 장애물을 중심으로 통신성능이 저하되는 채널 혼잡 영역이 발생할 수 있음을 의미한다. 이 경우 혼잡영역 중심의 대형차는 전방 수신 능력 저하로 인해 앞쪽에서 전달되는 FCW, EEBL 등의 후방 전달 서비스의 표준 PER 을 만족하지 못하는 경우가 발생하며, 이는 상호 영향에 의한 위험으로 판단할 수 있다. [그림 6]

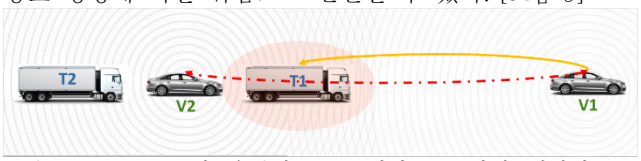


그림 6. V1, V2 간 숨겨진 노드 현상으로 인해 생성된 T1 주변의 통신혼잡영역과 영역에 방해받는 V1 → T1 통신성능

IV. 결론

본 논문에서는 자동차 전용도로 환경에서 WAVE/DSRC 기반의 V2X 서비스를 이용하는 소형/대형차가 혼재된 환경을 가정하고 상호간 V2X 통신 영향을 확인하기 위하여 각 그룹간 PER 성능을 측정/분석하였다. 군집주행중인 대형차 그룹은 주변의 소형차 그룹에 통신 장애물로써 통신 성능을 감소시키는 영향을 확인 할 수 있었으며, Near-LOS 와 Non-LOS 시나리오에서 각 최대 8.8%, 80.9% PER 을 나타내었다. 다수의 소형차 그룹이 만들어낸 V2V 통신 혼잡은 대형차 간의 통신 성능에 미치는 영향의 미미하였으나, 숨겨진 노드 상황이 발생하는 경우 소형차 대형차량 간 PER 성능이 최대 약 11%까지 높아지는 현상이 관측되었다. 이러한 결과는 통신 단말, 안테나 및 주변 환경 설정에 따라 달라질 수 있는 결과이며, 이에 관한 추가적인 시험 및 연구 수행을 통한 심층적인 통신 영향 분석이 필요하다.

ACKNOWLEDGMENT

본 자료는 국토교통부 국가교통과학기술진흥원의 「교통물류연구사업 (과제번호: 20TLRP-B147674-03) - V2X 기반 화물차 군집주행 운영기술 개발 연구과제」의 지원을 받아 작성되었습니다.

참 고 문 헌

[1] 조영, 권경주, 오철 - "고속도로 화물차 군집주행 적용구간 선정 연구" - Journal of Korean Society of Transportation (2018), 36(2), 98-111.

[2] Vukadinovic, V. et al., "3GPP C-V2X and IEEE 802.11p for Vehicle-to-Vehicle communications in highway platooning scenarios" - Ad Hoc Networks (March 2018) 74

[3] Joo, J., Jeong, H.-J., & Han, D. S., "Verification of Fresnel Zone Clearance for Line-of-sight Determination in 5.9 GHz Vehicle-to-Vehicle Communications" Wireless Pers Communication (2018) 101:239-249

텍스트 기반 지식요소 추출을 위한 온톨로지 활용 방안에 관한 연구

강유리
한화시스템

yuri.k54@hanwha.com

A Study on the utilization of ontology for extracting knowledge elements from text data

Yu Ri Kang
Hanwha Systems

요 약

다양한 종류의 정보가 실시간으로 발생하는 현대 전장환경에서는 적군의 거짓 정보, 아군의 자원 오류 요인 등으로 인해 혼재된 전장환경 정보 발생 가능성이 높다. 이러한 리스크가 있는 상황에서 지휘관이 올바른 상황판단을 하도록 지휘결심을 지원하는 인공지능 기술의 필요성이 높아지고 있다. 본 논문은 DARPA 의 AIDA 프로그램 중 GAIA 의 연구 분석을 통해 현 국방 환경에서 효과적으로 다종의 데이터로부터 추출한 지식요소 통합 및 정확한 상황 분석을 위한 지식요소 추출 단계들을 분석한다. 이 분석을 바탕으로 온톨로지를 활용한 텍스트 기반 지식요소 추출 방안의 국방 분야의 적용을 제안한다.

Ⅰ. 서 론

현대 전장환경에서는 적군의 거짓 정보와 아군 정찰 자원의 오류 등으로 인해 혼재된 전장환경 정보가 지휘관에게 전달된다. 또한 다종의 빅데이터가 초단위로 발생하여 축적되는 상황에서 이를 최대한 활용해 지휘관의 올바른 지휘결심을 지원할 인공지능 기술이 필요하다.

GAIA(GAIA: A Fine-grained Multimedia Knowledge Extraction System)는 DARPA AIDA 프로그램 중 멀티미디어 및 다국어 환경에서 지식요소를 추출하는 시스템이다. 추출한 지식요소들은 텍스트와 이미지에 표현된 상황을 분석하여 상황 분석 결과를 제시하는데 사용된다. GAIA 는 데이터 종류로는 텍스트, 이미지와 언어로는 영어, 러시아어, 우크라이나어를 다룬다. [1]

본 논문은 GAIA 의 AIDA 온톨로지와 GAIA 가 다루는 조건 중 영어 텍스트 데이터에서의 지식요소 추출 단계 분석을 통해, 온톨로지를 활용한 텍스트 지식요소 추출이 다종의 데이터 통합과 정확한 상황 분석에 효과적인 요소임을 제안한다.

Ⅱ. 본 론

본론 Ⅱ.1.에서는 2018 년부터 2020 년까지의 GAIA 온톨로지 구성을 보인다. Ⅱ.2.에서는 온톨로지를 세분화하여 다양화했을 때 지식요소 추출을 통해 상황 분석에서 얻을 수 있는 이점을 소개하며, 온톨로지를 활용한 영어 텍스트에서의 지식요소 추출 과정을 분석한다.

Ⅱ.1. AIDA 온톨로지

온톨로지는 지식요소 추출을 위해 추출할 정보를 정의하기 때문에 효과적으로 정보를 표현할 수 있는 구조가 필요하다. GAIA 는 Entity, Relation, Event 를 상황 표현을 위한 온톨로지 기본 구성요소로 두며, 특히 서로 다른 데이터 종류에 대해서도 상황을 표현할 수 있는 멀티모달 온톨로지를 구현 목표로 설정했다.

그러나 GAIA 는 기존의 온톨로지는 너무 대분화(coarse-grained)되어 있거나 너무 세분화(fine-grained)되어 있어 그대로 차용하지 않고, 여러 단계의 기준을 두고 YAGO 온톨로지를 정제하여 Entity 온톨로지를 구현하거나, 다수 온톨로지들에 대해 여러 단계의 기준을 두고 추가하여 Event 온톨로지를 구현함으로써 AIDA 온톨로지를 제작했다. [3] AIDA 온톨로지 예시는 Figure1 과 같다. [4]

```
/** Entity classes **/
IdcOnt:Person  a        owl:Class ;
      rdfs:subClassOf  aidaDomainCommon:CanHaveName ,
aidaDomainCommon:EntityType .

/** Event types **/
IdcOnt:Conflict.Attack
      a              owl:Class ;
      rdfs:subClassOf  aidaDomainCommon:EventType ;
      rdfs:subClassOf  [ a              owl:Restriction ;
                  owl:allValuesFrom  [ a          owl:Class ;
                              rdfs:label   "Attacker" ;
                             owl:unionOf  ( IdcOnt:GeopoliticalEntity
IdcOnt:Organization IdcOnt:Person )
                              ] ;
                  owl:onProperty     IdcOnt:Conflict.Attack_Attacker
                     ...
                ] .

/** Relation/Event arguments **/
IdcOnt:Conflict.Attack_Attacker
      a              owl:ObjectProperty ;
      rdfs:label        "Attacker" ;
      rdfs:subPropertyOf  owl:topObjectProperty .
```

Figure 1: AIDA 온톨로지 예시

GAIA 가 온톨로지를 세분화할 때는 각 Type 에 하위 레벨을 만드는 type.subtype.subsubtype 형태로, Type 을 세분화하는 방법을 이용했다. 그 형태는 Figure2 와 같다. [5]

1. top level type for the most coarse-grained level (e.g., PER)
2. type.subtype for the next level (e.g., PER.Politician)
3. type.subtype.subsubtype for the finest-grained level (e.g., PER.Politician.Governor)

Figure 2: AIDA 온톨로지 세분화 형태

서로 다른 종류의 데이터가 하나의 상황에 대해 서로 상충되는 정보를 포함할 때, GAIA 는 AIDA 온톨로지를 활용해 추출한 각 지식요소를 최신 개발한 Visual Grounding System 을 통해 단일의 일관된 정보를 표현하도록 멀티미디어 결과들을 통합한다. 이로써, 데이터 종류를 넘어서 Entity 상호참조해결을 통한 다종의 데이터로부터의 지식요소 추출 결과 통합이 가능하다.

GAIA 는 Entity 세분화 type 수를 2018 년 163 개에서 2019 년 187 개로, Event 세분화 type 수를 2018 년 114 개에서 2019 년 139 개, 2020 년 144 개까지 확장하며 AIDA 온톨로지를 정교하게 세분화했다. 가장 최신 발표된 2020 년 GAIA 의 AIDA 온톨로지 구성은 Figure3 과 같다. [2]

	Coarse-grained Types	Fine-grained Types
Entity	7	187
Relation	23	61
Event	47	144

Figure 3: 2020 년 AIDA 온톨로지 구성

Ⅱ.2. 온톨로지를 활용한 텍스트 기반 지식요소 추출

GAIA 는 Ⅱ.1.과 같이 온톨로지를 구성하여 지식요소 추출 시스템에 적용했다. GAIA 가 2018 년 연구부터 2020 년까지 온톨로지 세분화 변화를 기반한 지식요소 추출을 통해 상황 분석에서 얻을 수 있던 이점은 크게 2 가지이다.

1. 상황 이해도 향상
2. 이벤트 예측도 향상

Type 이 세분화 되면, 발생한 Event 에 어떠한 대상이 이어서 오는지에 따라 정확한 상황 이해를 지원하며, 또한 뒤에 어떠한 Event 가 등장할 가능성이 높을지 역으로 예측 확률도 높아진다. [2]

Figure4 는 GAIA 멀티미디어 지식요소 추출 전체 아키텍처이며, 영어 텍스트에서 지식요소 추출 단계와 각 방법은 아래와 같다. [1]

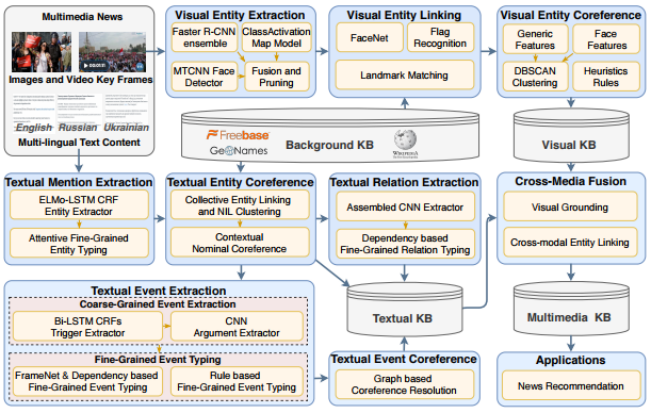


Figure 4: GAIA 멀티미디어 지식 추출 아키텍처

1. 텍스트 Entity 추출과 상호참조해결
   1.1. 대분화된 Mention 추출
      - ELMo LSTM-CRF 모델을 이용하여 Entity Mention 추출 및 "대분화 Type" 할당
   1.2. Entity Linking 과 상호참조해결
      - 대용량의 외부 데이터베이스를 포함한 지식베이스에 동일할 Entity 가 존재하는지 확인 및 상호참조해결
      - 존재하지 않을 시, 룰 이용해 NIL 클러스터링
   1.3. Entity Typing 세분화
      - 상호참조해결 정보, 지식베이스, 분류기를 이용하여 대분화 Type 을 바탕으로 "세분화 Type" 재할당
   1.4. Entity 중요도 랭킹
      - 가중치에 따라 Entity Mention 중요도 랭킹
2. 텍스트 Relation 추출
   - CNN 이용해 "대분화 Relation Type 할당" 후, Entity Type 제약 조건과 언어 의존 패턴 기반 혹은 룰 기반으로 "세분화 Relation Type" 재할당
3. 텍스트 Event 추출과 상호참조해결
   - Bi-LSTM CRF 와 CNN 사용해 "대분화 Event Type" 할당 후, 룰 비교를 통해 "세분화 Event Type" 재할당
   - 그래프 알고리즘을 통해 Event 상호참조해결

Ⅲ. 결 론

기존 인공지능 모델은 데이터 종류에 의존적이기 때문에 지식요소를 추출하여도 통합적으로 활용하여 분석하는데 어려움이 있었다. GAIA 는 AIDA 온톨로지를 구현해 지식요소 추출 및 데이터 통합에 활용함으로써 다종의 데이터로부터 추출한 지식요소를 통합하여 상황 분석을 위한 Event 정보를 보다 정확하고 풍부하게 표현하는 시스템을 만들었다.

현대 전장환경은 대량의 다종 데이터가 발생하며, Event 를 중심으로 정확한 상황 분석이 중요하다. 이 점은 혼재된 대량의 전장환경 정보를 통합하고 분석하여 지휘관의 지휘결심을 지원하는 지능형 전장인식 기술 개발의 필요성으로 이어진다.

본 논문은 GAIA 연구 분석을 통해, 현 국방 분야의 지능형 전장인식 기술 개발에 GAIA 의 온톨로지를 활용한 지식요소 추출 방법 적용이 적합한 것을 확인하였다. 향후 본 논문의 연구 결과를 기반으로 온톨로지를 활용한 텍스트 기반 지식요소 추출 시스템을 개발할 예정이다.

참 고 문 헌

[1] Li, Manling, et al. "Gaia at sm-kbp 2019-a multi-media multi-lingual knowledge extraction and hypothesis generation system." Proceedings of TAC KBP (2019).

[2] Li, Manling, et al. "Gaia: A fine-grained multimedia knowledge extraction system." Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations. 2020.

[3] Zhang, Tongtao, et al. "GAIA-A Multi-media Multi-lingual Knowledge Extraction and Hypothesis Generation System." TAC. 2018.

[4] NIST "Text Analysis Conference" NIST TAC, 30 August 2018, https://tac.nist.gov/tracks/SM-KBP/2018/ontologies/SeedlingOwlOntology, accessed 27 November 2020

[5] NIST "Text Analysis Conference" NIST TAC, 28 June 2019, https://tac.nist.gov/2019/SM-KBP/guidelines/SM-KBP_2019_Evaluation_Plan_V1.5.pdf, accessed 27 November 2020

한글, 영문, 숫자 및 특수기호가 혼합된 텍스트용 필기체 인식기의 구현

김홍숙*, 김정시
*한국전자통신연구원

*kimkk@etri.re.kr, sikim00@etri.re.kr

Implementation of Hand-written Letter Recognizer for Text of Korean Syllable, Alphabet, Digit and Special Symbol

Hongsoog Kim*, Jeong-Si Kim
Electronics and Telecommunications Research Institute

요 약

본 논문은 한글, 영문, 숫자 및 특수 기호를 포함한 텍스트를 대상으로 필기체 글씨 인식기 구현에 대한 연구로, 데스크탑 기반의 학습용 서버상에서 딥러닝 기반의 컨볼루션 네트워크를 이용하여 대상 글씨들을 학습하고, 학습된 모델을 임베디드 시스템상에 이식하여 실제 응용에 필요한 추론을 수행하였다. 실험 결과 학습용 데이터 셋과 분리된 검증용 데이타셋을 이용한 인식률 테스트에서 Top-1 accuracy 0.9777 의 높은 인식률을 확인하였다. 이를 기반으로 임베디드 보드상에서 필기체 인식기를 GUI 애플리케이션을 구현하여 실용화에 필요한 보완 사항들을 확인하고, 실무 활용 가능성을 확인하였다.

Ⅰ. 서 론

필기체 글자 인식과 같은 이미지 분류 문제에 있어서, 기존의 규칙 기반 시스템의 경우, 입력 데이터를 검토하여 규칙성을 파악하고, 규칙을 알고리즘화 하였으나, 이러한 방법은 제한된 범위내의 문제만을 해결할 수 있으며, 새로운 데이터에 대하여 유연한 대처가 불가능한 단점이 있었다[1,2]. 본 논문에서는 한글, 영문, 숫자 및 특수 기호를 포함한 2,448 개의 서로 다른 글자 클래스의 필기체 인식을 위한 딥러닝 기반 컨볼루션 네트워크의 학습 및 임베디드 시스템상에 이식하여 고정양식내에 포함된 글자들을 인식하는 애플리케이션화 과정까지의 학습용/검증용 데이터 셋의 준비과정, 모델 학습 과정, 입력 전처리, 및 추론 적용과정에서의 실무 경험을 소개한다.

Ⅱ. 본론

한글, 영문, 숫자 및 특수기호가 혼합된 필기체 인식을 위한 딥러닝 기반 모델 학습을 위한 학습용 이미지 데이터 셋 및 학습 과정의 검증을 위한 검증용 이미지 데이타셋을 먼저 준비하여야 한다.

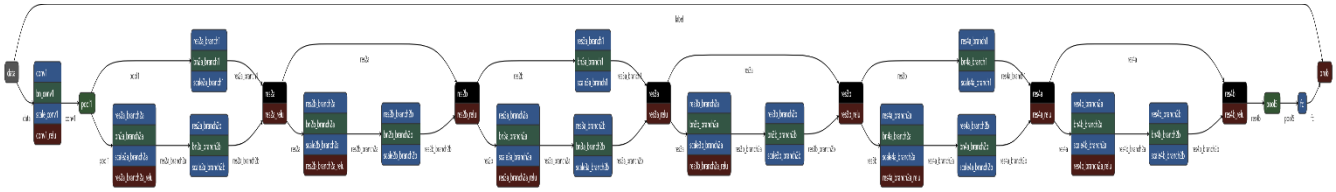
모델 학습 및 검증을 위한 필기체 이미지 데이터 셋들은 한국정보화진흥원에서 배포하는 한국어 글자체 이미지[3], SD19 숫자 영문 필기체 이미지[4], 자체 보유 글자 이미지셋 및 데이터 증강 (data augmentation) 기술에 기반한 자체 개발한 폰트 기반 필기체 글자 생성기를 통하여 확보하였다. 한국정보화진흥원에서 제공하는 한글 글자 이미지의 경우, 완성형 한글의 필기체 학습에 사용가능한 그레이 스케일 이미지들을 선별한 결과 전체 이미지의 약 35%정도인 271,858 장의 이미지들만을 사용하였다. 완성형 글자수 2,350 을 고려하면 글자당 115 장 정도를 활용할 수 있었기에, 부족한 이미지셋을 보충하기 위하여 자체 개발한 필기체 글자 이미지 생성기를 사용하여 공개된 인쇄체 및 필기체 폰트들로 생성한 글자 이미지에 Random Elastic Distortion 등의 방법을 적용하여 생성한 필기체 이미지들로 학습용/검증용 이미지 데이터 셋의 크기를 증강하였다.

인식 대상인 글자는 KS-X-1001 규격에서 정의하는 완성형 한글 2350 자, 알파벳 대소문자 52 자, 숫자 10 자, 주요 화폐 기호를 포함한 특수 기호 36 자를 포함한 총 2,448 개의 클래스로 구성되며, 확보된 33,239,371 장의 글자 이미지 파일에 대하여, 9:1 의 비율로 학습용 셋 29,557,881 장, 검증용 셋 3,281,490 장으로 분할하여 사용하였다. 학습에 사용된 이미지셋은 글자당 평균 12,074 장에 해당한다.

필기체 인식용 딥러닝 모델은 Github 에 공개된 ResNet-18 모델[5]을 대상으로 하는 2448 개의 글자 클래스 인식에 맞게 그림 1 과 같은 구조로 일부 레이어와 파라메터들을 수정하여 사용하였다. 학습에 사용된 딥러닝 프레임워크는 연구용으로 널리 사용되는 caffe[6]와 파이썬 기반 caffe 구동용 GUI 프로그램인 barista[7]를 사용하였다. 딥러닝 기반 필기체 인식 모델의 학습에 사용된 컴퓨팅 환경, 학습 모델, 학습 및 검증용 데이터 셋에 대한 내용을 표 1 에 정리하였다.

[표 1] 필기체 인식기 모델 학습 환경

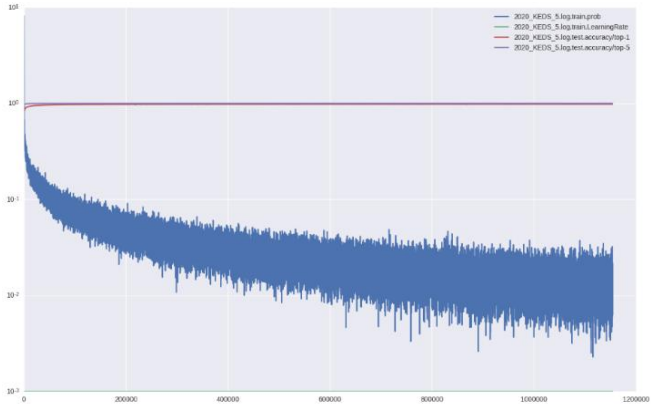
학습용 서버	CPU: Intel Core I7-6700K GPU: NVIDIA RTX2080 Ti Memory: 64GB (16GBx4) 딥러닝 프레임워크: BVLC Caffe
학습용 모델	Resnet-18 Modified 학습 파라메터 수: 3,404,864 개
학습용 검증용 데이터 셋	2,448 클래스에 대하여 학습용 글자 이미지: 29,557,881 장 검증용 글자 이미지: 3,281,490 장
하이퍼 파라메터	학습량: 50 epoch - train batch size: 1,280, -총 iteration 1,154,650 Optimizer: ADAM



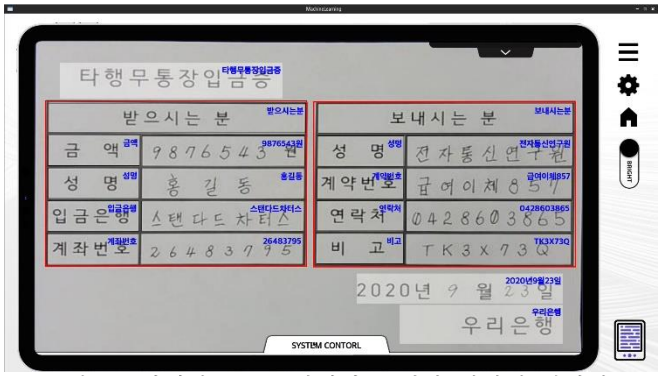
[그림 1] ResNet-18 을 수정한 학습용 모델 구조

학습 결과에 대하여 검증용 데이타셋을 통하여 정확도를 측정한 결과 50 epoch 의 학습 후 Top-1 accuracy 0.9777, Top-5 accuracy 0.9988 을 보였다. 그림 2 는 학습과정에서의 iteration 에 따른 accuracy, loss 를 측정한 그래프로, 보라색은 Top-5 accuracy 이며, 적색 커브가 Top-1 accuracy, 파란색 커브는 loss 를 나타낸다. 150,000 iteration 후에 Top-5 accuracy 는 0.95 이상을 보였으나, loss 의 감소는 이후에도 지속적으로 일어나서, Early stopping 없이, 설정한 전체 iteration 을 계속하여 진행하였다. 인식기용 딥러닝 모델의 전체 학습에 3 일 11 시간 30 분이 소요되었다.

학습용 서버에서 학습된 모델을 실제 서비스하기 위한 임베디드 보드상에 이식하여 은행에서 사용하는 타행 무통장 입금증 양식을 간략화 한 고정 양식용 글자 인식기를 구현하여 실용화 가능성을 평가하였다. 임베디드 보드상에서 글자 추론 기능은 자체 개발한 추론 가속 라이브러리리에 기반한 추론 엔진에서 학습 서버에서 학습된 모델 파라메터들을 로딩 하여 사용하였다. 그림 3 은 임베디드 보드상에서 Qt QML 을 사용하여 개발한 GUI 환경에서 고정 양식에서 글자 이미지에 대하여 추론한 결과 화면이다. 실제 이미지를 캡쳐하여 추론기의 입력으로 사용 시, 낱 글자 영역 추출에 있어서, 촬영시의 광원 등에 의한 왜곡 효과가 발생하여 이를 제거하기 위한 전처리 과정을 추가하였다.



[그림 2] 학습과정에서 accuracy, loss 변화 추이



[그림 3] 임베디드 보드상에서 구현한 필기체 인식기

Ⅲ. 결론

본 논문에서는 한글, 영문, 숫자, 특수기호가 혼합된 2,448 종의 서로 다른 글자에 대한 필기체 인식을 위한 딥러닝 기반 학습 모델에 필요한 데이터 셋의 준비 및 학습, 임베디드 보드로의 이식 및 응용 개발 과정에서의 작업들을 정리하였다.

개발된 학습 모델을 테스트하는 과정에서 잘 준비된 낱 글자들에 대한 검증용 데이타셋을 사용한 검증 결과는 98%이상의 정확도를 보였지만, 무통장 입금증을 카메라로 촬영하는 실무 상황에서 정확한 낱 글자영역 분리 과정이 있어야, 검증용 데이터 셋에서의 정확도와 같은 정확도를 확보할 수 있음을 확인하였다.

스캐너 이미지와 같이 선명한 입력 이미지를 확보하기 힘든 카메라 촬영 환경과 같은 실제적인 환경에서의 활용을 위하여 광원에 따른 글자 왜곡 제거, 정확한 낱 글자 영역 분리를 위한 이미지 전처리 기능들이 중요한 요소임을 확인하였으며, 이를 보완하기 위한 기술들을 현재 개발하고 있는 중이다.

ACKNOWLEDGEMENT

본 연구는 2020년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(No.2017-0-00142, 스마트기기를 위한 온디바이스 지능 형 정보처리 가속화 SW플랫폼 기술 개발)을 받아 수행된 연구임.

참 고 문 헌

[1] Y. LeCun, et al., “Gradient-Based Learning Applied to Document Recognition,” Proceedings of the IEEE, 86(11):2278-2324, 1998.

[2] Y. LeCun, et al. “Deep learning,” Nature, 521(7533):436-444, 2015.

[3] 한국정보화진흥원, “한국어 글자체 이미지,” 2020. (https://aihub.or.kr/aidata/133)

[4] Patrick J. Grother, “NIST Special Database 19 2nd Ed.,” 2016. (https://www.nist.gov/srd/nist-special-database-19)

[5] HolmesShuan, “ResNet-18-Caffemodel-on-ImageNet,” 2020. (https://github.com/HolmesShuan/ResNet-18-Caffemodel-on-ImageNet)

[6] Yangqing Jia, et al., “Caffe: Convolutional architecture for fast feature embedding,” Proc. of the 22nd ACM international conference on Multimedia, 2014.

[7] S. Klemm, et al., "Barista - A Graphical Tool for Designing and Training Deep Neural Networks," CoRR (Computing Research Repository) 2018.

사전 훈련된 두 Autoencoder 교차 연결을 통한 번역 성능 개선

오지은, 최용석*
한양대학교, *한양대학교
jiunoh@hanyang.ac.kr, *cys@hanyang.ac.kr

Improving Machine Translation via Cross-connecting Two Autoencoders

Oh Jiun, Choi Yong Suk*
Hanyang Univ., *Hanyang Univ.

요 약

본 연구에서는 기계번역 문제를 위하여 두 개의 오토인코더(autoencoder)를 각각 단일 언어에 대하여 사전 훈련한 후, 둘을 교차 연결하여 병렬 코퍼스로 번역을 학습하는 전이 학습 모델을 개발하였다. 두 오토인코더는 각자 입력 언어와 출력 언어에 대하여 denoising autoencoder 방식으로 사전 훈련되며, 입력 언어의 인코더와 출력 언어의 디코더를 연결한 후 둘 사이에 feature-mapping layer를 추가하여 병렬 코퍼스로 미세 조정된다. 이 모델을 이용하면, 단일 언어로 사전 훈련한 모델을 여러 언어 쌍을 위하여 반복적으로 재사용할 수 있으며, 번역을 학습시킬 수 있는 언어 쌍에 제한이 없다. 실험 결과, 본 논문의 방식이 기존 Transformer 모델보다 개선된 성능으로 교차 연결을 통한 재활용의 가능성을 보였다.

Ⅰ. 서 론

최근 자연어 처리 분야에서 BERT[1]가 전이학습(transfer learning) 기법으로 큰 성능 향상을 달성한 이래, 신경망 기계 번역(Neural Machine Translation)문제를 위하여서도 사전 훈련(pre-training)과 미세 조정(fine-tuning)을 통한 전이 학습이 보편적인 방법으로 자리 잡았다. 그런데 전이 학습을 위하여 모델을 단일 언어로 사전 훈련하는 데에는 많은 비용이 드는 반면에, 이미 사전 훈련된 모델을 다른 언어 쌍 번역에 대하여 재사용하는 방법은 많지 않다.

따라서 본 논문에서는 두 개의 오토인코더(autoencoder)를 각각 단일 언어에 대하여 따로 사전 훈련한 후, 둘을 교차 연결하여 병렬 코퍼스로 번역을 학습하는 전이 학습 모델을 제안한다. 이 모델을 사용하면 단일 언어로 사전 훈련한 모델을 언어 쌍에 구애받지 않고 계속 재사용할 수 있다. 예컨대 영어와 프랑스어 간의 번역을 위하여 학습한 영어 모델을 다시 영어와 독일어 간의 번역을 학습하는 데에 사용할 수 있다. 또한 이 방법은 처음부터 따로 학습된 모델을 사용하기 때문에 기존의 Transformer[2] 모델과 달리 단어 사전과 임베딩이 분리되어 있어 번역을 학습할 수 있는 언어 쌍에 제한이 없고 어떤 조합으로든 학습이 가능하다.

본 논문에서 사용된 오토인코더 모델은 Transformer이며, 데이터셋은 WMT 2014 english-french[3]이다. 성능 평가 지표는 BLEU 점수이고 측정을 위하여 tensor2tensor 스크립트[4]를 사용하였다. 실험 결과, 본 논문의 모델이 기존 Transformer 모델보다 우수한 성능을 보여 사전 학습의 효과와 사전 학습된 모델의 교차 재활용 가능성을 입증하였다.

Ⅱ. 본론

1. 사전 학습

두 모델은 단일 코퍼스를 이용하여 각각의 단일 언어에 대하여 denoising autoencoder[5]로 학습된다. 예컨대 영어 오토인코더는 오염된 영어 입력을 받아 오염 없는 영어 출력을 내고, 프랑스어 오토인코더는 오염된 프랑스어 입력을 받아 오염 없는 프랑스어 출력을 내도록 학습된다. 데이터에 noise를 주기 위하여 BERT의 Masked Language Model (MLM) 방법이 사용되었다. masking의 비율은 BERT와 동일하다. 전체 데이터 중 15% 토큰이 오염되는데, 그 15% 가운데 80% 토큰은 [MASK] 토큰으로, 10%는 임의의 토큰으로 대체하며, 10%는 원래의 토큰 그대로 둔다. 각 토큰은 sub-word 단위로 분리되며, 한 단어를 이루는 sub-word들 전체가 masking되도록 하였다. 예를 들어 한 단어 'student'가 'stud', '##ent'로 분리될 때, 둘 모두 [MASK] 또는 임의의 토큰으로 대치되도록 하였다. BERT와 달리, 디코더를 함께 학습하기 위하여 모델이 오염된 토큰뿐 아니라 원래의 입력 문장 전체를 재구성하도록 하였다. 사전 훈련 모델의 구조는 그림 1과 같다.

2. 미세 조정

미세 조정 단계에서 모델은 우선 사전 훈련된 가중치로 초기화된다. 이 때, 인코더는 입력 언어로 훈련된 모델의 인코더로 초기화되고, 디코더는 출력 언어로 훈련된 모델의 디코더로 초기화된다. 단 디코더 중 encoder-decoder attention 부분은 임의로 초기화되는데, 이 부분이 monolingual로만 학습되었던 모델이 cross-lingual 문제인 번역을 수행하기 위하여 새로 학습되어야 하는 부분이기 때문이다.

그런데 인코더와 디코더는 단일 언어 데이터로만 사전 훈련되었으므로, 단순히 교차 연결만으로는 번역을 제대로 수행하기 힘들 수 있다. 그렇기 때문에 다른 언어로 학습된 인코더와 디코더가 맞물리도록 둘 사이에 feature-mapping layer를 추가하였다. 이 레이어는 self-attention과 feed-forward network로 이루어져 있으며 dropout, residual connection, layer normalization이 적용되었다. 구조적으로 인코더 내부의 한 sublayer와 동일하며, 임의로 초기화된다. 이 상태에서 프랑스어 문장을 영어 문장으로 번역하도록 병렬 코퍼스를 이용하여 번역을 학습하였다.

Ⅲ. 실험 및 결과

실험에서 사용된 오토인코더 모델은 Transformer이다. 모델의 구현은 tensorflow tutorial[6]을 사용하였다. 컴퓨터 자원의 한계로 모델 크기를

Transformer-Base보다 축소하였다. 인코더와 디코더의 레이어 개수는 각 4개로 도합 8개이다. 임베딩 차원은 128, feed-forward filter size는 512이다. attention head는 8개가 사용되었다. dropout 비율은 0.1이다. 사용된 데이터는 WMT 2014 english-french 데이터셋으로 evaluation set은 newstest2013, test set은 newstest2014이다. 이때 사전 훈련에는 원 학습 데이터셋의 1/4, 미세 조정과 baseline에는 1/8을 사용하여 학습하였다. 단어 사전의 크기는 각 32K이며 sub-word encoding이 적용되었다. batch size는 64이고 토큰이 64개 이상인 문장은 제거하였다. 최적화 기법으로 Adam을 사용하였으며, learning rate schedule과 warmup step은 Transformer 원 논문의 수식을 그대로 적용하였다. β_1=0.9, β_2=0.999, ϵ=1e-6이다.

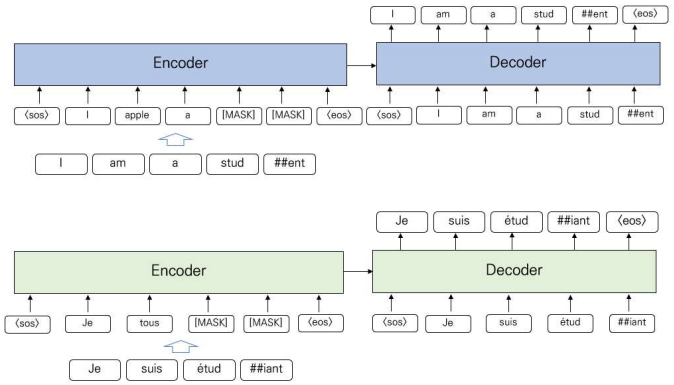


그림 1 사전 훈련 모델

위는 영어, 아래는 프랑스어 모델이다. <sos>는 문장의 시작을, <eos>는 문장의 끝을 나타낸다. 모델은 MLM 방식으로 noise가 있는 입력으로부터 원래의 입력을 복원하도록 학습된다.

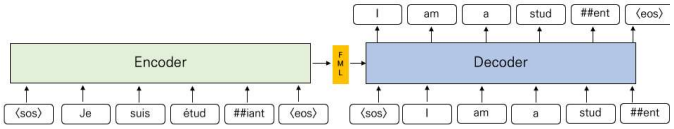


그림 2 미세 조정 모델

FML은 feature-mapping layer를 의미한다. 모델은 프랑스어 입력으로부터 번역된 영어 문장을 생성하도록 학습된다.

baseline은 사전 훈련을 하지 않고 임의로 초기화한 Transformer 모델이다. 미세 조정을 위한 1/8 데이터셋을 사용하였으며 모델의 크기, 최적화 기법, learning rate schedule은 위와 동일하다. Adam의 hyperparameter와 warmup step 모두 원 Transformer 논문과 같다. 또한 feature-mapping layer의 효과를 확인하기 위하여 feature-mapping layer 없이 미세 조정한 모델을 같이 비교하였다. feature-mapping layer 유무를 제외한 모든 데이터셋, 초기화 가중치, hyperparameter는 feature-mapping layer가 있는 모델과 같다.

실험 결과는 표 1에 나타나 있다. Transformer는 사전 훈련하지 않은 모델이며, ours (w/o FML)는 사전 훈련하고 feature-mapping layer가 없는 모델, ours (w FML)는 사전 훈련하고 feature-mapping layer가 추가된 모델이다. ours (w FML)의 결과는 임의로 초기화한 baseline보다 BLEU 점수가 확연히 높다. ours (w/o FML)의 결과는 baseline 결과보다는 다소 우세하지만, ours (w FML)보다 점수가 낮다. 이를 통하여 완전히 별개의 언어와 단어 사전을 이용하여 사전 훈련된 두 모델의 인코더와 디코더를 사후에 조립하여 미세 조정할 수 있으며, 이러한 재활용을 통하여 성능이 향상된다는 사실을 확인할 수 있다. 또한 재활용할 때에 추가적인 feature-mapping layer가 성능 향상에 유의미한 영향을 미친다는 사실 또한 확인할 수 있다.

모델	BLEU
baseline (random initialized)	24.18
ours (w/o FML)	24.57
ours (w FML)	**25.30**

표 1 실험 결과

baseline은 사전 훈련 없이 임의로 초기화된 Transformer 모델이다. ours (w/o FML)는 사전 훈련 후 교차 연결하고 feature-mapping 없이 미세 조정한 모델이며, ours (w FML)는 feature-mappng layer를 추가하여 미세 조정한 모델이다. 사용된 지표는 BLEU 점수이다.

Ⅳ. 결론

본 논문에서는 기계번역을 위하여 사전 훈련한 모델을 여러 언어 쌍을 위하여 효율적으로 재활용할 수 있는 새로운 전이학습 방법을 고안하였다. 이 방법은 두 개의 오토인코더를 각각 단일 언어 코퍼스로 학습한 후, 입력 언어의 인코더와 출력 언어의 디코더를 교차 연결하여 병렬 코퍼스로 학습한다. 이때 학습된 인코더와 디코더 사이에 feature-mapping layer를 삽입함으로써, 사전에 서로 다른 언어를 학습한 인코더와 디코더가 맞물려 번역을 학습할 수 있도록 하였다. 그 결과로 사전 훈련 없이 임의로 초기화한 Transformer보다 BLEU 점수가 상승하였으며, 사전 훈련된 모델의 재활용을 통하여 번역 성능을 향상시킬 수 있음을 보였다.

ACKNOWLEDGMENT

이 연구는 2019년도 산업통상자원부 및 산업기술평가관리원(KEIT) 연구비 지원(과제번호:10077553)과, 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(No. 2020R1A2C1014037), 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2020-0-01373, 인공지능대학원지원(한양대학교)).

참 고 문 헌

[1] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. "Bert: Pre-training of deep bidirectional transformers for language understanding." CoRR, 2018.

[2] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L. & Polosukhin, I. "Attention is all you need." In Advances in neural information processing systems, pp. 5998-6008, 2017

[3] https://www.tensorflow.org/datasets/catalog/wmt14_translate

[4] https://github.com/tensorflow/tensor2tensor/blob/master/tensor2tensor/bin/t2t_bleu.py

[5] Vincent, P., Larochelle, H., Bengio, Y., & Manzagol, P. A. "Extracting and composing robust features with denoising autoencoders." In Proceedings of the 25th international conference on Machine learning, pp. 1096-1103, 2008

[6] https://www.tensorflow.org/tutorials/text/transformer

심층학습을 이용한 한국어 음성변조에 관한 연구

김한, 이환용
아주대학교

hanish@ajou.ac.kr, hwan@ajou.ac.kr

A Study on the Korean Voice Conversion Systems Using Deep Learning

Kim Han, Lee Hwanyong
Ajou Univ.

요 약

본 논문에서는 영어와는 언어적 특징이 다른 한국어의 음성 변조 시스템을 딥러닝을 통해서 구현하는 방법을 제안한다. 먼저 실험을 위해서 KSS(Korean Single Speaker) TTS 데이터 세트과 추가적으로 음성인식에 주로 활용되는 Zeroth 데이터 세트를 학습 입력 데이터로 이용하였다. 학습 단계에서 필요한 hyper parameter 를 실험을 통해 제시하였다. 또한 hyper parameter 중 teacher forcing rate 를 학습 중간에 변경하여 줌으로써 학습 속도가 향상 되도록 하는 방법을 제안 하였다. 마지막으로 보다 향상된 변조 품질을 얻기 위해 기존 모델의 인디코더 학습 방식에 GAN 방식의 Discriminator 네트워크를 추가하였고 (1+1) 학습 전략을 통해 적대적 학습을 진행하였다. 이를 통해, 일주일 정도의 학습 결과로서 화자의 음성 특성 정도를 구분할 수 있는 수준의 음성 변조 결과를 얻을 수 있었다.

Ⅰ. 서론

음성 변조 VC(Voice Conversion)는 source 화자의 음성 특징을 기반으로 하여 target 화자의 특징을 자연스럽게 덧입혀 변환하는 기술을 의미한다. 음성 변조는 impaired speech 를 normal speech 로 변환하는데 사용하거나 음성 더빙과 같은 여러 응용 분야에 활용 되어왔다. 또한 이 분야는 언어의 음성학적 특성에 따라서 다른 접근이 필요하다. 따라서 영어에 기반한 음성 변조 연구 결과를 우리 말에 그대로 적용하였을 경우 좋은 결과를 얻기 힘들다. 따라서, 본 논문은 한국어 음성 변조 모델 구축하기 위해 다양한 실험을 수행하고 이 중 활용할 수 있는 학습 가이드라인을 제시하고자 한다. 본 연구에서는 Cotatron[1] 이란 오픈소스 프로젝트를 음성 변조 모델로 사용했다. Cotatron 을 활용한 음성 변조 품질을 향상 하기 위하여 학습 모델에 새로운 네트워크인 Discriminator 를 제안했다.

Ⅱ. 본론

Ⅱ-1. 학습 데이터 세트

#	전체화자 수	전체 학습 분량	화자당 데이터 양	음질	대사 중복 여부	언어
Libri	123	46 시간	평균 11 분	clean	x	영어
VCTK	109	34 시간	평균 22 분	clean	x	영어
KSS	1	12 시간	12 시간	clean	x	한국어
Zeroth	105	52 시간	평균 30 분	noisy	o	한국어

표 1. 본 논문에서 활용되는 데이터 세트 현황

본 논문에서 활용하는 Cotatron 음성 변조 모델은 한 쌍의 텍스트와 음성 파일로 이루어진 TTS 데이터 세트를 입력으로 요구한다. 위의 표 1 에서 알 수 있듯 활용 가능한 한국어 TTS 데이터 세트는 영어 데이터 세트와 달리 KSS(Korean Single Speaker) 데이터 세트만 유일하게 활용 가능하며, 오직 한 명의 화자만이 존재한다. 하지만 음성 변조를 위해서는 다양한 화자가 필요하다. 따라서 유사 도메인의 음성인식에서 105 명의 화자 기반 Zeroth 데이터 세트을 함께 활용하였다.

Ⅱ-2. 한국어 학습을 위한 실험

본 논문에서 활용하는 변조 모델의 경우 영어 모델을 학습하기 위해 123 명의 Libri 데이터 세트와 108 명의 VCTK 데이터 세트의 일부를 활용하여 풍부한 데이터를 가지고 학습을 진행하였다. 하지만 본 논문은 제한적인 한국어 데이터 세트에 대해 학습이 필요하다. 따라서, 먼저 12 시간 크기의 단일 화자의 KSS 데이터 세트만을 가지고 50k steps 만큼의 학습을 진행했다. 그 후, 다시 checkpoint 를 load 하여 105 명의 Zeroth 데이터 세트를 기반으로 100k steps 까지의 세밀한 학습을 진행하였다. weight decay, lr decay 등 각종 hyper parameter 및 학습 방식은 기존 모델과 동일하며 달라진 hyper parameter 는 목록은 표 2 와 같다.

실험을 통해 decoder 의 prenet 의 dropout 을 disable 시켰던 것이 학습 결과가 좋았기 때문에 활용하지 않았다. 또한 vocoder 모델이 mel-spectrogram 을 오디오로 변환하는 과정에서 이에 해당하는 MelGAN[4] 과의 호환을 위해 최저, 최대 주파수에 해당하는 f_{min} 과 f_{max} 를 각각 70, 8000 에서 0, 11025 로 바꿔주었다.

또한 영어 데이터 세트에 대해 학습하기 위해 Cotatron[1]은 초기 learning rate = $3 * 10^{-3}$ teacher forcing rate = 0.5 으로 설정하였다. 하지만 본 연구에서는 좀 더 빠르게 학습하기 위하여 learning rate 를 $6 * 10^{-3}$ 으로 바꿨고, 초기의 학습 방향을 정확하게 가이드 하기 위하여 teacher forcing rate 를 1.0 으로 설정하여 학습을 진행하다가 alignment 상태가 확인되면 다시 checkpoint 를 load 한 후 teacher forcing rate = 0.5 으로 설정하여 학습을 마무리 하였다. 위 방법의 적용을 통해 기존에는 150k-200k 까지 학습이 필요했던 이전의 여러 사례들에 비해 약 10 배 정도 빠르게 학습을 진행할 수 있었다. [5]

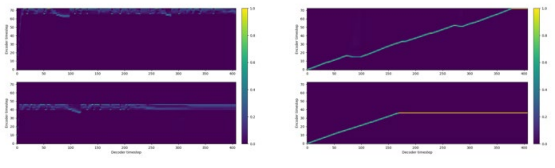


그림 3. 기존 방법(좌)과 본 논문의 방법(우) 비교 (step 2k 에서의 alignment 그래프)

#	learning rate	prenet dropout	TF_{rate}	f_{min}	f_{max}
기존 방식	$3 * 10^{-3}$	0.5	0.5	70	8000
본 논문	$6 * 10^{-3}$	0.0	1.0->0.5	0	11025

표 2. 개선 된 Hyper Parameter

Ⅱ-3. 성능 향상을 위한 적대적 학습

본 논문에서 활용하는 변조 모델의 네트워크 구조는 아래 그림 4 와 같다. voice conversion decoder 의 학습을 위해선 source 화자의 mel-spectrogram 과 target 화자의 ID 가 일치해야 입력 mel-spectrogram 에 관한 reconstruction loss 처벌이 가능하다. 즉, 모델은 화자가 일치하는 경우에만 한정적으로 reconstruction loss 에 의해 학습이 진행된다.

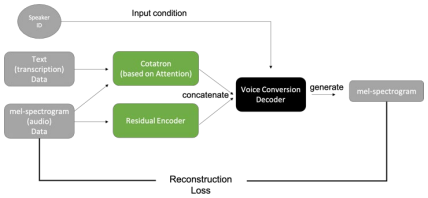


그림 4. 변조 모델의 training process

하지만 변조는 그림 5 의 inference 의 상황처럼 target 으로 하는 Speaker ID 와 source 라는 입력 mel-spectrogram 이 서로 다른 경우를 의미한다.

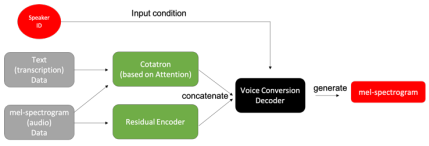


그림 5. 변조 모델의 inference process

따라서, 본 논문은 변조 라는 목적 자체에 어울리는 target 과 source 가 일치하지 않는 경우에 대한 학습까지 고려하여 Discriminator 네트워크를 그림 6 과 같이 구성하였다.

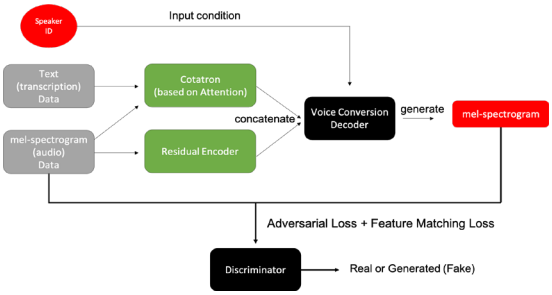


그림 6. 본 논문에서 사용한 적대적 신경망 구조

학습은 (1+1) 전략[3]을 활용하여 각 step 별로 번갈아 학습하였다. 학습 단계에서 한번은 기존의 학습 방식처럼 source 와 target 이 일치하는 paired data 를 학습하고 다른 한번은 source 와 target 이 일치하지 않는 unpaired data 를 학습하였다. 그렇기에, 각 단계마다 loss function 의 설정이 다르다. 전자는 모든 loss function 에 대해 활성화하였고, 후자는 feature matching loss 와 reconstruction loss 를 비활성화하여 학습 하였다.

Ⅲ. 결론

본 논문에서는 Cotatron 음성 변조 모델을 활용하여 한국어 음성 변조 시스템을 제안 하였다. 이를 위해 유사 도메인인 음성인식 한국어 데이터 세트을 활용하여 학습하였다. 또, 실험을 통해 한국어 Hyper Parameter 찾아냈고 보다 빠른 한국어 모델 학습 방법을 추가적으로 제시하였다. 또한 기존 모델의 변조 성능을 향상 시키기 위해 적대적 학습 방식을 추가 하였다. 이를 통해, 일주일 정도의 학습 결과로서 화자의 음성 특성 정도를 구분할 수 있는 수준의 음성 변조 결과를 얻을 수 있었다. 변조 결과는 source 화자와 target 화자의 특성이 뒤섞이는 문제가 여전히 존재한다. 추후 보다 향상 된 결과를 얻기 위해서는 target 화자의 감정과 억양과 같은 특성을 분리할 수 있는 네트워크 추가가 필요하다고 판단된다. 또한 source 화자와 target 화자가 일치 하지 않는 상황에서의 변조 성능 향상을 위한 방안들에 대한 추후 연구가 필요하다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW 중심대학사업의 연구결과로 수행되었음 "(2015-0-00908)"

참 고 문 헌

[1] Seung-won Park, Doo-young Kim, Myun-chul Joe, “Cotatron : Transcription-Guided Speech Encoder for Any-to-Many Voice Conversion without Parallel Data”, *Interspeech 2020*

[2] Yuxuan Wang, RJ Skerry-Ryan, Daisy Stanton, “Tacotron: Towards End-to-End Speech Synthesis”, *Interspeech 2017*

[3] Shuyang Gu, Jianmin Bao, Hao Yang, Dong Chen, Fang Wen, Lu Yuan, “Mask-Guided Portrait Editing with Conditional GANs”, *CVPR 2019*

[4] Kundan Kumar, Rithesh Kumar, Thibault de Boissiere, Lucas Gestrin, Wei Zhen Teoh, Jose Sotelo, Alexandre de Brebisson, Yoshua Bengio, Aaron Courville, “MelGAN: Generative Adversarial Networks for Conditional Waveform Synthesis”, *NeurIPS 2019*

[5] https://github.com/carpedm20/multi-speaker-tacotron-tensorflow/issues/4

비디오 스크립트의 종결어미 태그를 이용한 비디오 요약 방안 연구

신영주, 양진홍*
헬스케어IT학과, 인제대학교
20173262@oasis.inje.ac.kr, *jinhong@inje.ac.kr

A Study on the video summary method using the final ending tag of video script

Shin Yeong Ju, Yang Jinhong*
Inje Univ.

요 약

최근 온라인 동영상의 공급 증가에 따라 다양한 영상 플랫폼을 통해 콘텐츠 소비자들에게 소비되고 있다. 그 중, 모바일 중심의 소비를 중심으로 카드뉴스와 같은 새로운 스낵 미디어 형태의 영상 소비 방식 등이 등장 하였다. 특히 한정된 사용자의 영상 소비 시간에 맞춰 빠르게 콘텐츠를 이용할 수 있는 영상 요약의 중요성은 점차 증대되고 있다. 본 논문에서는 영상을 재생이 아닌 이미지로 소비하기 위해 스크립트를 기반으로 뉴스 영상의 주요 장면을 추출하였다. 이때 스크립트 기반의 영상 구간 정보 판독 시 발생하는 한계점을 해결하기 위해 한국어 형태소 분석기의 종결어미를 사용하였고, 추출 정확도에 대한 평가를 진행함으로써 해당 방식에 대한 근거로 활용하였다.

Ⅰ. 서 론

최근 다양한 영상 플랫폼의 등장 및 온라인 동영상의 공급 증가 그리고 소비하기 편리하도록 제작된 온라인 동영상의 보편화로 인해, 온라인 동영상의 이용률은 급증하고 있다.[1] 그 중, 온라인 동영상의 시청은 대부분 모바일을 중심으로 콘텐츠 소비자에게 소비되고 있는데, 이에 따라 card news, skimming 등의 형태로 영상 콘텐츠 소비를 위한 새로운 소비 방식이 등장하였다.

이와 같은 현상에 의해 뉴스, 시사 정보 이용과 같은 정보 전달의 매체로 신문이나 포털사이트가 아닌 유튜브와 같은 영상 플랫폼을 통해 소비하는 현상이 발생하고 있다.[2] 제한된 사용 시간에 의해 발생하는 새로운 영상 소비 방식으로 인해, 영상 요약 기술의 필요성이 증가하게 되었다.

본 논문에서는 영상 요약에 필요한 영상 주요 장면 추출을 위해 많은 양의 영상에 보다 가벼운 방법을 이용하기 위해 영상의 메타데이터인 스크립트를 기반으로 영상의 주요 장면을 추출하였고, 이때, 분석을 위해 한국어 형태소 분석기를 이용하였다. 그림 1은 실제 YouTube의 뉴스 장르의 영상 중 하나의 실제 스크립트인데 뉴스 영상의 스크립트의 특성에 따르면 영상의 프레임에 따라 자막이 나눠짐으로 자막이 문장 단위로 나눠 있지 않아, 정보를 전달함에 있어 문제가 발생하고 있다.

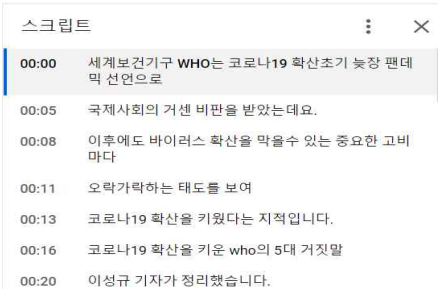


그림 1. 실제 YouTube 영상의 스크립트

따라서 스크립트 기반 영상 주요 장면 추출 시 발생하는 문제를 해결하고자 종결어미 태그를 이용하여 끊어져 있는 문장을 연결시켜 문장 단위로 스크립트를 추출하는 방식을 이용하였고 한국어 형태소 분석기의 종결어미 추출 정확도에 대한 평가 진행을 통해 해당 방식에 대한 근거로 활용하고자 한다.

Ⅱ. 관련 연구

기존 연구에 많이 사용된 대표적 한국어 형태소 분석기로는 KoNLPy가 있다. 이는 Tokenizing을 통해 명사, 동사를 구분하여 한국어 자연어 처리를 하는 파이썬 패키지로 분석기가 가지고 있는 미리 학습된 사전을 기반으로 단어를 추출하는 방식을 사용한다. 해당 분석기는 분석 속도가 빠르며 대중적으로 사용되지만 사전 내용을 기반으로 형태소 분석을 진행함으로 사전에 포함되어 있지 않은 단어는 추출할 수 없다는 한계점이 존재한다.[3]

따라서 본 논문에서는 머신 러닝 형태소 분석기 중 하나인 Kiwipiepy모듈을 사용한다. 파이썬 기반의 한국어 형태소 분석기로 학습한 데이터를 기반의 분석을 진행하므로 미등록 단어에 대한 추출 성능이 뛰어난 특징을 가진다.[4] 미리 학습된 단어 사전을 기반으로 단어를 추출하되, 미등록 단어는 사용자 사전을 추가하여 기존의 사전에 등록되어 있지 않은 단어도 추출할 수 있도록 제작된 Kiwipiepy를 이용해 KoNLPy를 사용함으로써 발생하는 한계점을 해결하고자 하였다

Ⅲ. 본론

3.1 한국어 형태소 분석기

본 연구에서 사용된 Kiwipiepy는 한국어 형태소 분석기 Kiwi(Korean Intelligen Word Identifier)의 파이썬 모듈로 또 다른 머신 러닝 형태소 분

석기인 Soynlp의 Word Extraction 기법을 바탕으로 만든 분석 방식을 이용한다. Kiwipiepy는 코퍼스로부터 미등록 된 단어를 추출하고 기존의 사전에 등록되지 않은 단어를 제대로 분석하기 위해 사용자 사전에 등록하여 사용한다.[5]

3.2 종결어미 태그를 이용한 뉴스 기사 분석

본 연구는 뉴스에서 종결어미 태그를 이용한 문장 추출을 진행하기 위해 뉴스가 많이 소비되는 플랫폼 중 하나인 포털사이트에 기재되는 100개의 뉴스 기사를 대상으로 실시하였다. 해당 뉴스 기사의 경우 영상의 나레이션에 대한 스크립트가 웹 페이지 상에 full-text로 제공되어 별도의 처리 없이도 정확한 영상의 스크립트를 획득할 수 있다. 그림 2는 본 논문에서 뉴스 기사를 분석하기 위한 시스템의 구조이다.

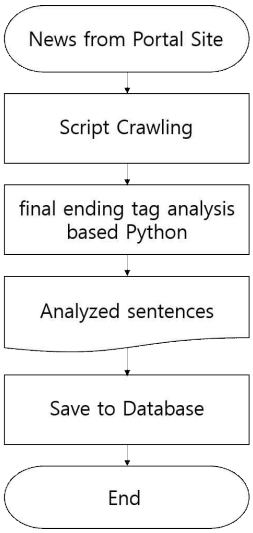


그림 2. 시스템 구조도

먼저, 뉴스 기사를 크롤링 후 이를 별도의 텍스트 파일로 생성하였다. 이후 Python의 Kiwipiepy 모듈을 사용하여 문장 단위로 파일을 읽으며 종결어미를 기준으로 문장을 분류하기 위해 품사 태그 중 종결어미에 해당하는 태그인 “EF”를 기준으로 문장의 개수를 측정하는 분석 방법을 사용하였다. 그 결과를 실제 사람이 평가한 Ground Truthset의 문장 개수와 Kiwipiepy를 통해 측정한 문장 개수를 비교하여 종결어미에 대한 추출 정확도를 측정하였다. 그림 3은 문장 개수를 측정하기 위한 알고리즘이다.

```
EndingTag = ["EF"] // 종결어미 태그
for result, count in result:
    for end in result:
        IF len(end[0]) > 1 and end[1] in EndingTag THEN
            return words.append(end[0])
// 문장과 종결어미 태그를 비교하여 문장 개수를 측정
```

그림 3. 종결어미 태그를 이용한 문장 개수 측정 알고리즘

표 1은 인터넷 뉴스 기사의 스크립트를 Kiwipiepy의 종결어미 태그 통해 분석하였을 때 도출할 수 있었던 결과로 실제 문장 개수 대비 형태소 분석기를 통해 측정한 문장 개수에 대한 비율을 나타내고 있다. 표 1은 10개의 그룹으로 만들어졌으며 하나의 그룹은 포털사이트의 언론사별 인터넷 기사 10개로 구성되어 있다.

표 1. 형태소 분석기의 종결어미 추출 정확도

그룹	정확도(%)
그룹 A	88.8%
그룹 B	92.3%
그룹 C	85.7%
그룹 D	93.3%
그룹 E	100%
그룹 F	86.6%
그룹 G	93.75%
그룹 H	100%
그룹 I	93.3
그룹 J	92.8%
평균	92.6%

Ⅳ. 결론

본 논문에서는 Kiwipiepy를 통해 미리 학습된 단어 사전을 기반으로 종결어미를 추출하였고 그 과정에서 미등록 단어에 대해서는 학습시켜 사용자 사전에 추가하는 과정을 진행하였다. 그 결과로 실제 문장 개수와 비교하여 종결어미로 추출한 문장의 개수에 대한 분석 정확도의 평균은 92.6%로 높은 정확도를 보여주고 있다. 이때, 정확도가 100%에 일치하지 않는 이유는 특정 문장에서 종결어미 태그가 아닌 연결어미 태그인 “EC” 태그로 인식되는 것이 이유였다. 이에 대해, 문장을 분석할 때 “EF”, “EC” 태그를 모두 사용하여 추출하는 방식을 이용한다면 분석 정확도를 높일 수 있을 것이라 기대한다. 스크립트를 전체로 합치고 문장 단위로 다시 나누는 과정은 뉴스 영상의 스크립트가 프레임 대비 문장 단위로 작성되어 있지 않아 발생하는 영상 정보 전달의 한계점에 대해 해결책으로 제시할 수 있을 것이다. 또한, 이를 통해 뉴스 영상의 스크립트에 대한 가독성을 높여줄 수 있을 것이라 기대한다.

결과적으로 본 논문에서는 뉴스 영상 요약 시, 발생하는 문제점을 해결할 수 있는 방안으로서 종결어미를 통한 뉴스 영상 스크립트를 재구성하는 방식을 제시하였다. 향후 이를 이용하여 뉴스 영상에 대한 요약 방식을 진행, 스크립트 기반 분석 및 주요 장면 추출을 진행할 수 있을 것이다. 또한, 장르별 테스트를 통해 뉴스 장르가 아닌 다른 장르에서의 영상 분석에도 이를 이용하여 스크립트 기반 영상 주요 장면 추출을 진행할 수 있다면 제한된 시간에 대한 새로운 콘텐츠 소비 방식을 제시할 수 있을 것이다.

ACKNOWLEDGMENT

"이 성과는 2020년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2019R1F1A1059357)"

참 고 문 헌

[1] NASMEDIA, “NPR, Netizen Profile Research“, pp. 17-19, 2020

[2] Korea Press Foundation, “Media Issue Volume.5 Number.3“, pp. 2-4, June, 2019

[3] D.W.Ko, J.J.Yang, “Korean Natural Language Processing and Analysis Using KoNLPy and Word2Vec“, Journal of the Korean Information Science Society, pp. 2140-2142. June, 2018

[4] lovit, soynlp, Retrieved Nov. 28, 2020, from https://github.com/lovit/soynlp

[5] ba2min, kiwipiepy, Retrieved Nov. 28, 2020, from https://github.com/bab2min/kiwipiepy

Max-Mean N-스텝 시간차 학습

황규영*, 김주봉*, 허주성*, 한연희*†
*한국기술교육대학교 미래융합공학전공

{to6289, rlawnqhd, chil1207, yhhan}@koreatech.ac.kr

Max-Mean N-step Temporal Difference Learning

Gyu-Young Hwang*, Ju-Bong Kim*, Joo-Seong Heo*, Youn-Hee Han*
*Future Convergence Engineering, Korea University of Technology and Education.

요 약

적절한 n을 선택할 경우 n-스텝 시간차 학습은 몬테카를로 방법과 1-스텝 시간차 학습보다 성능이 좋은 알고리즘으로 알려져 있지만 최적의 n값은 파라미터에 민감하며 편향-분산 트레이드오프 문제가 있기 때문에 n을 선택하는 것에 어려움이 있다. 기존 n-스텝 시간차 학습에서 n값 선택의 어려움을 해소하기 위해, 본 논문에서는 $1 \le k \le n$에 대한 모든 k-스텝 누적 보상의 최댓값과 평균으로 구성된 새로운 학습 타겟인 $\Omega - return$을 제안한다. 마지막으로 OpenAI Gym 의 Atari 게임 환경에서 기존 n-스텝 시간차 학습과의 성능 비교 평가를 진행하여 본 논문에서 제안하는 알고리즘이 기존 n-스텝 시간차 학습 알고리즘보다 성능이 우수하다는 것을 입증한다.

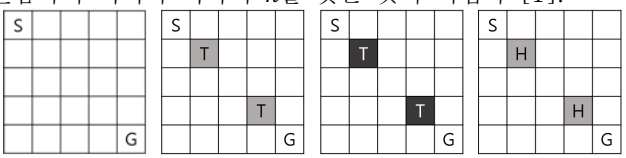
Ⅰ. 서 론

몬테카를로 방법은 완전 누적 보상을, 1-스텝 시간차 학습은 1-스텝 누적 보상을 학습의 업데이트 타겟으로 사용한다. n-스텝 시간차 학습은 몬테카를로 방법과 1-스텝 시간차 학습을 결합한 것으로, n-스텝까지 관찰된 누적 보상과 n 번째 스텝에서 기대되는 행동 가치가 결합된 값을 학습의 업데이트 타겟으로 사용한다. n 의 값을 적절하게 선택할 경우 n-스텝 시간차 학습은 강화 학습의 성능을 높일 수 있어 1-스텝 시간차 학습보다 좋을 수 있다. 하지만 파라미터 n은 강화 학습 환경의 변화에 대하여 민감하여 최적의 n을 찾는 것이 어렵고, n의 값이 커지면 n-스텝 누적 보상의 분산이 커지고 n 의 값이 작아지면 편향이 발생하는 편향-분산 트레이드오프 문제가 있다. 본 논문에서는 n -스텝 시간차 학습에서 최적의 n 을 선택해야 하는 문제를 해결하기 위해, $\Omega - return$ 이라는 새로운 n -스텝 업데이트 타겟을 제안한다.

Ⅱ. 배경

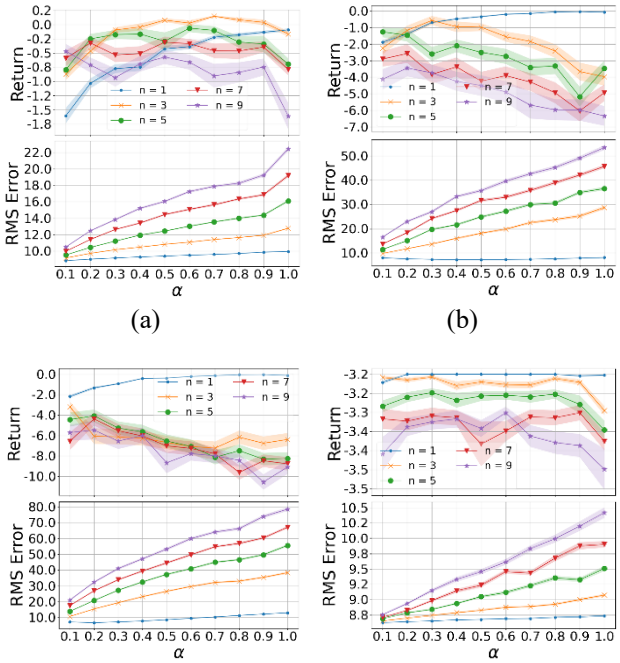
1. n-스텝 시간차 학습의 파라미터 민감성

(그림 1)은 네 가지의 5×5 그리드월드 환경이다. 매 스텝마다 -0.1 의 보상을 받고, 'S'에서 출발해 'G'에 도달하면 1.0 의 보상을 받으며 에피소드가 끝난다. (b)의 'T'에 도달하면 -3.0, (c)의 'T'에 도달하면 -6.0, (d)의 'H'에 도달하면 -3.0 의 보상을 받으며 에피소드가 종료된다. (그림 2)는 (그림 1) 환경에서의 실험 결과이다. 실험 결과와 같이 n은 실험 환경과 학습률 α의 변화에 민감하여 최적과 최악의 n을 찾는 것이 어렵다 [1].



(a) (b) (c) (d)

(그림 1) 네 가지의 5×5 그리드월드



(a) (b) (c) (d)

(그림 2) 네 가지의 5×5 그리드월드에서의 실험 결과

2. 편향-분산 트레이드오프

Q-learning n -스텝 누적 보상은 다음의 2 가지로 분리하여 볼 수 있다.

$$\sum_{k=1}^{n} \gamma^{k-1} R_{t+k} \quad (1)$$

$$\gamma^{n} \max_{a} Q_{t+n-1}(S_{t+n}, a) \quad (2)$$

수식(1)은 에이전트가 환경과 상호작용하며 받은 보상을 의미하며, 수식(2)는 행동 가치 함수 Q 를 사용하여 누적 보상을 추정한 값이다. n의 값이 커지면 수식(1)에 의해 확률적으로 행동하며 받은 보상의 비중이

† 교신 저자 한연희

커지므로 n-스텝 누적 보상의 분산이 커지게 된다. 반대로 n의 값이 작아지면, Q는 학습 중인 에이전트가 받을 누적 보상을 추정한 값이기 때문에 실제 값과는 차이가 있고 이로 인해 편향이 발생하게 된다 [2].

Ⅲ. 본론

본 논문은 $1 \le k \le n$에 대한 모든 k-스텝 누적 보상의 최댓값과 평균으로 구성된 $\Omega - return$이라는 새로운 n-스텝 업데이트 타겟을 제안한다.

제안하는 방법은 n-스텝 누적 보상에 기초한 가장 좋은 업데이트 타겟을 결정하는 것에서 시작된다.

Q-learning 알고리즘의 학습 타겟인 1-스텝 누적 보상을 보면 $\max_a Q_t(S_{t+1}, a)$가 사용된다. Q의 max를 취하는 것은 실제 값보다 과대평가되어 편향이 발생해 문제를 야기할 수 있지만, 만약 높은 보상을 받을 수 있는 상태-행동에 대한 Q 값이 과대평가되어 있다면 에이전트가 해당 영역을 탐험하도록 장려하여 학습에 이점을 줄 수 있다 [3]. 따라서 $1 \le k \le n$에 대한 모든 $G_{t:t+k}$ 값들의 최댓값을 G_{max}라 명명하고 다음과 같이 정의한다.

$$G_{max} = \max_k G_{t:t+k}, \quad 1 \le k \le n \tag{3}$$

만약 모든 상태에 대한 최적의 행동 가치 Q^*를 구했다면, $1 \le k \le n$에 대한 모든 $G_{t:t+k}$는 같은 값을 가질 것이다. 상태 S_t에서 행동 A_t가 최적 정책을 따른다면 다음과 같이 표기할 수 있다.

$$\frac{1}{n}\left[\sum_{k=1}^{n} G_{t:t+k}\right] - Q(S_t, A_t) \approx 0 \tag{4}$$

수식(4)로부터 $1 \le k \le n$에 대한 모든 $G_{t:t+k}$ 값들의 평균이 업데이트 타겟으로 좋다는 것을 알 수 있다. 따라서 이 평균을 G_{mean}이라 명명하고 다음과 같이 정의한다.

$$G_{mean} = \frac{1}{n}\left[\sum_{k=1}^{n} G_{t:t+k}\right] \tag{5}$$

G_{max}는 초기 단계에 학습을 가속화 시킬 수 있지만 Q값을 과대평가하는 경향이 있어 학습하는 동안 문제가 될 수 있다 [4]. 반면, G_{mean}은 학습 후반 업데이트의 타겟으로써 좋다. 그러므로 G_{max}와 G_{mean}을 결합한 업데이트 타겟인 $\Omega - return$을 제안한다.

$$\Omega - return = \beta G_{max} + (1-\beta) G_{mean}, \quad 0 \le \beta \le 1 \tag{9}$$

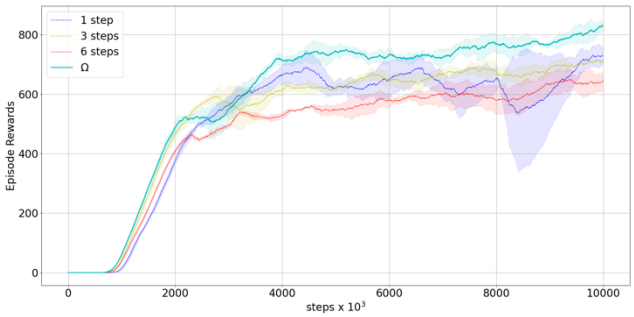
β는 G_{max}와 G_{mean}의 비율을 결정하는 파라미터이다. 본 논문은 파라미터 β를 다음의 수식으로 대체하였다.

$$\beta = \frac{\max_k |G_{t:t+k}| - \min_k |G_{t:t+k}|}{\max_k |G_{t:t+k}|} \tag{10}$$

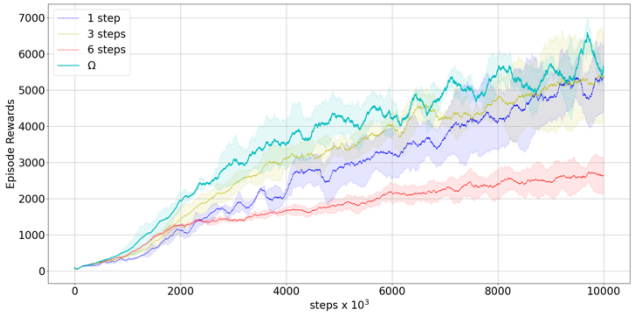
학습 초기에는 β 값이 높아 G_{max}의 비율이 높으며, 학습이 진행되면서 β 값은 줄어들어 G_{mean}의 비율이 높아진다.

Ⅳ. 실험

제안하는 $\Omega - return$의 성능을 평가하기 위해 Nature DQN, Double DQN, Dueling DQN, Prioritized replay buffer 알고리즘들을 통합한 기법 위에서 기존 n-스텝 시간차 학습과의 비교 실험을 진행한다. 실험 환경은 OpenAI Gym Atari 게임 중 Enduro 와 Seaquest 이다. 실험 결과는 (그림 3), (그림 4)와 같으며 3 번 측정한 것의 평균과 표준편차를 그래프에 나타내었다. 실험 결과를 통해 기존 n-스텝 시간차 학습 알고리즘의 성능이 1-스텝 시간차 학습보다 떨어지는 경우에도 본 논문에서 제안하는 알고리즘의 성능이 좋다는 것을 알 수 있다.



(그림 3) OpenAI Gym Enduro 환경에서의 실험 결과



(그림 4) OpenAI Gym Seaquest 환경에서의 실험 결과

Ⅴ. 결론

n-스텝 시간차 학습은 하이퍼-파라미터 n을 선택하기 어려운 문제가 있다. 본 논문에서는 $1 \le k \le n$에 대한 모든 n-스텝 누적 보상의 최댓값과 평균으로 구성된 $\Omega - return$이라는 새로운 n-스텝 업데이트 타겟을 제안한다. 실험은 DQN 알고리즘의 기본이 되는 Nature DQN 과 DQN 의 확장 알고리즘인 Double DQN, Dueling DQN, Prioritized replay buffer 알고리즘들을 통합한 기법 위에서 기존 n-스텝 시간차 학습과 제안하는 알고리즘의 성능 비교 평가를 진행하였다. OpenAI Gym Atari 게임 환경에서의 실험 결과 본 논문에서 제안하는 알고리즘이 기존 n-스텝 시간차 학습 알고리즘보다 성능이 우수하다는 것을 입증하였다. 따라서 $\Omega - return$이 기존 n-스텝 시간차 학습 알고리즘에서 적절한 n을 선택해야 하는 문제를 해결하는 새로운 학습 타겟으로 적절하다고 생각된다.

ACKNOWLEDGMENT

이 논문은 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No. 2018R1A6A1A03025526 및 No. NRF-2020R1I1A3065610).

참 고 문 헌

[1] Hessel, M., J. Modayil, H. van Hasselt, et al. Rainbow: Combining Improvements in Deep Reinforcement Learning. In *AAAI*, pages 3215-3222. AAAI Press, 2018.

[2] Hernandez-Garcia, J., Sutton, R. Understanding Multi-Step Deep Reinforcement Learning: A Systematic Study of the DQN Target. 2019. Cite arXiv:1901.07510 Comment: NIPS Deep Learning Workshop 2018.

[3] Lan, Q., Pan, Y., Fyshe, A., White, M. Maxmin Q-learning: Controlling the Estimation Bias of Q-learning. In *ICLR* 2020.

[4] Thrun, S., A. Schwartz. Issues in using function approximation for reinforcement learning. In *Proceedings of the Fourth Connectionist Models Summer School*. Erlbaum, 1993.

RF 충전 후방산란 CR 네트워크에서 효율적인 강화학습 기반 모드 최적화

오선애, 신요안*

숭실대학교 전자정보공학부

Email: {sunae0814@soongsil.ac.kr, yashin@ssu.ac.kr}

An Efficient Mode Optimization Based on Reinforcement Learning in RF-powered Backscatter CRNs

Shanai Wu, Yoan Shin*

School of Electronic Engineering, Soongsil University

(*Corresponding author)

요 약

본 논문은 RF 충전 후방산란 인지 무선 네트워크에서 2차 송신단말 (Secondary Transmitter; ST)이 1차 채널과 상호작용하면서 받는 보상을 통해 최적의 정책을 효율적으로 학습하는 방안을 제안한다. ST는 환경에서 변화되는 자신의 상태에 적합한 동작 모드를 수행하면서 주어진 시간 동안에 최대한 많은 데이터 패킷을 전송하는 것을 목표로 하며, 학습의 효율을 높이기 위해 Energy Outage가 발생하는 동작 모드를 선택한 경우에 Penalty를 부여하는 방안을 제안하였다. 랜덤하게 변화하고 예측이 어려운 환경에서 안정적인 학습이 가능한 Deep Q-network 알고리즘을 사용하였으며, ST가 학습된 인공신경망 모델로부터 모드를 선택하여 수행하는 모의실험을 통해 제안 기법의 학습 성능을 검증하였다.

Ⅰ. 서 론

주변의 무선 주파수 (Radio Frequency; RF) 신호로부터 에너지를 충전하는 RF 에너지 수집 기술이 센서 노드와 같은 저전력 단말의 자기 유지 가능한 (Self-sustainable) 에너지 공급 기술로 부상하고 있다. 주파수 이용 효율 극대화를 위한 인지 무선 (Cognitive Radio; CR) 기술과 결합한 RF 충전 CR 네트워크에서 2차 송신단말 (Secondary Transmitter; ST)은 주변의 1차 신호로부터 에너지를 수집하고, 1차 채널이 비어 있는 동안에 수집한 에너지를 사용하여 데이터를 전송하는 방식으로 동작할 수 있다. 따라서 ST의 전송 성능은 1차 시스템에 의해 결정되며, ST의 전송 성능을 개선하기 위해 주변 후방산란 통신 (Ambient Backscatter Communication; AmBC) 기술의 적용이 제안되었다[1]. AmBC는 주변 RF 신호를 반사하여 정보를 전송하는 통신 기술로서 전력소모가 적기 때문에 에너지 수집과 결합하여 효율적인 무선충전 통신 네트워크를 구성할 수 있다. 또한 RF 신호를 수신하는 단말로부터 7.2인치 이상 떨어져 있으면 후방산란 간섭을 발생하지 않는다는 연구결과가 보고된 바 있다[2].

RF 충전 후방산란 CR 네트워크에서 예측하기 어려운 1차 채널에 접근하여 점유 상태에 적합한 모드로 동작하면서 일정 시간 동안에 최대의 전송 성능을 얻기 위해, ST는 랜덤하게 변화하는 상태를 고려하여 순차적으로 동작 모드를 결정해야 한다. 이를 위해 본 논문에서는 강화학습의 활용 방안을 고려하였으며, 강화학습은 환경에서 시도한 행동과 그 결과로 나타나는 보상 사이의 상관관계를 시행착오를 통해 학습하는 방법이다.

Ⅱ. RF 충전 후방산란 CR 네트워크

본 논문에서는 1차 시스템과 2차 시스템이 각각 한 쌍의 송수신단으로 구성된 RF 충전 후방산란 CR 네트워크를 고려하였다. 1차 시스템에서 주사용자 (Primary User; PU)는 스펙트럼 대역에 접근할 수 있는 권한을 갖고 있는 사용자이며, 1차 채널의 상태는 PU의 전송 패턴에 따라 변화한다. 본 논문에서는 타임 슬롯 기반의 네트워크 모델을 고려하며, 따라서 임의 슬롯에서 1차 채널이 PU에 의해 사용될 확률이 p인 베르누이 분포에 따라 간단하게 모델링할 수 있다. 잔여 에너지가 충분한 경우에 ST는 1차 채널이 비어 있으면 Active 모드로 데이터를 전송하며, PU가 채널을 사용하고 있으면 RF 신호를 후방산란하여 정보를 전송하거나 에너지를 수집하여 배터리에 저장한다. ST는 타임 슬롯의 시작점에서 확률 λ로 발생하는 데이터를 저장장치에 보관하였다가 데이터가 저장된 순서에 따라 순차적으로 전송하며, 저장 공간이 부족한 경우에 가장 오래된 데이터부터 손실하게 된다. 전송 성능을 최대화하기 위해 ST는 1차 채널의 점유 상태에 적합한 모드로 동작해야 하며, 따라서 각 슬롯의 시작점에서 일정 시간 동안 스펙트럼을 센싱하는 방식으로 1차 채널의 점유 상태를 판단할 수 있다. 하지만 ST가 RF 신호로부터 수집 가능한 에너지가 제한적이기 때문에 부가적으로 소모되는 에너지를 최소화하기 위해, 본 논문에서는 에너지 수집 모드를 통해 1차 채널의 점유 상태를 판단하는 방안을 제안한다. 제안 기법은 슬롯을 두 개의 서브 슬롯으로 나누어 동작하며, 첫 번째 서브 슬롯에서 ST는 수집 모드를 통해 에너지 상태의 변화를 관찰하여 1차 채널의 점유 상태를 판단한다. 즉, 배터리의 잔여 에너지가 증가하면 PU가 1차 채널을 사용하고 있다고 판단하여 두 번째 서브 슬롯에서 주변 후방산란 모드로 동작하게 되고, 그렇지 않으면 1차 채널이 비어 있다고 판단하여 Active 전송 모드도 동작한다. ST는 에너지 수집을 통해 얻은 관찰 값으로 1차 채널의 점유 상태에 적합한 모드로 동작하면서 데이터를 성공적으로 전송하는 것도 중요하지만, 1차 채널과 상호작용 하면서 변화하게 될 자신의 상태에서 최적의 모드를 선택하면서 주어진 시간 동안에 최대한 많은 데이터를 전송하는 것도 중요한 문제이다. 따라서 ST는 순차적으로 동작 모드를 결정해야 하며, 본 논문에서는 강화학습을 통해 ST가 임의 상태에서 최적의 모드를 선택하는 정책을 학습하고자 한다.

Ⅲ. 제안하는 강화학습 기반 모드 최적화

ST가 순차적으로 동작 모드를 결정하는 문제에 접근할 수 있도록 하기 위해, 마르코프 결정 과정 (Markov Decision Process; MDP)을 통해 수학적으로 문제를 정의해야 한다. MDP를 구성하는 요소들로는 상태, 행동,

보상, 상태전이확률, 감가율 등이 있다. 따라서 본 논문에서는 첫 번째 서브 슬롯에서 에너지 수집 모드를 통해 얻은 관찰 값, 배터리의 잔여 에너지, 데이터 큐 (Queue) 등이 ST의 현재 상태 s가 되며, 이와 같은 상태들을 고려하여 ST가 두 번째 서브 슬롯에서 선택할 수 있는 행동 a들로는 Active 전송과 주변 후방산란뿐만 아니라 Idle 모드와 수집 모드가 있다. ST는 첫 번째 수집 모드를 통해 1차 채널이 비어 있다고 판단되어도 잔여 에너지가 부족하거나 저장된 데이터 패킷이 적은 경우에 Idle 모드로 동작하게 되며, 전송해야 하는 데이터가 적으면 1차 채널이 점유되었다고 판단하여도 Idle 모드로 동작하거나 배터리가 완전히 충전되지 않으면 에너지를 수집할 수 있다. ST는 Active 전송과 후방산란이라는 행동을 통해 데이터 전송에 따른 보상 (r)을 받게 되며, 상태에 적합한 행동을 선택하면 양의 보상을 받게 되고 상태에 적합하지 않은 행동을 선택하면 음의 보상을 받게 된다. 행동을 취한 후 다음 슬롯에 도달하게 될 상태 s'에는 확률적인 요인이 포함된다. ST는 에너지 상태와 데이터 큐의 변화를 스스로 관찰할 수 있는 반면에 환경에 해당하는 1차 채널의 상태 변화를 알 수 없다. 따라서 제안 기법은 에너지 수집 모드를 위한 서브 슬롯을 할당하여 1차 채널의 상태 변화를 관찰한다. 또한 ST가 행동을 결정하는 시점인 현재에 가까운 보상일수록 큰 가치를 갖도록 하기 위해 감가율을 사용하여 나중에 받게 될 보상의 가치를 감소할 수 있다.

본 논문에서는 Deep Q-network (DQN) 알고리즘을 사용하여 ST가 최적의 정책을 학습할 수 있도록 하였다. DQN은 Off Policy인 Q-learning 알고리즘이 Q값을 탐욕 (Greedy) 정책에 따라 업데이트하는 방식과 동일하게 경사하강법을 사용하여 오류함수를 최소화하도록 인공신경망의 가중치를 학습시킨다[3]. 반면에 환경에서 충분히 탐험하기 위해 행동 정책은 ϵ-탐욕 정책을 따른다. 또한 DQN은 환경에서 탐험하면서 얻은 (s, a, r, s') 샘플을 리플레이 메모리에 저장하였다가 학습에 사용하는 경험 리플레이 (Experience Replay)를 통해 샘플들의 상관관계가 학습에 주는 영향을 완화하였으며, 한 개의 샘플로 학습하는 것이 아니라 배치로 학습하기 때문에 학습이 안정적이다. 학습의 목표가 되는 정답이 타임 스텝마다 변하는 것을 방지하기 위해, DQN은 목표 인공신경망을 따로 구현하여 정답 역할을 하는 값을 제공하며 일정 시간 간격마다 학습하는 인공신경망의 가중치로 업데이트하여 준다. 리플레이 메모리는 사이즈가 정해져 있기 때문에 오래된 샘플들부터 삭제되며, 샘플들에 가중치를 할당할 수 없기 때문에 학습에 유리한 샘플들도 삭제된다. ST는 유한한 에피소드를 반복하면서 타임 스텝마다 인공신경망의 가중치를 업데이트하며, 본 논문에서는 Energy Outage (EO)가 발생하는 행동을 취하면 Penalty를 부여하는 동시에 에피소드를 강제로 종료시켜 나쁜 상황에서 신속하게 벗어나도록 하여 학습 효율을 향상시키는 방안을 제안하였다.

Ⅳ. 모의실험 결과 및 결론

제안 기법을 통해 ST가 달성 가능한 성능을 검증하기 위해 타임 스텝이 500인 에피소드를 반복하면서 학습을 수행하도록 하였으며, 학습 과정은 그림 1에서 도시한다. DQN 알고리즘의 하이퍼파라미터는 표 1에서 정리하였으며, 목표 신경망은 에피소드마다 업데이트하였다. 데이터 저장장치와 배터리의 용량은 모두 10으로 고려하였으며, 한 개의 슬롯을 기준으로 수집 가능한 평균 에너지와 Active 모드로 동작하면서 소모하는 에너지를 모두 1로 고려하였으며, Active 전송과 후방산란을 통해 각각 2개와 1개의 데이터 패킷을 전송할 수 있다고 가정하였다. 또한 데이터 발생 확률 λ와 1차 채널이 사용될 확률 p를 모두 0.5로 설정하였다. 그림 2는 EO Penalty를 적용하여 학습하는 과정을 도시하며, Penalty를 적용하지 않은 경우에 비해 학습 시간이 짧은 것을 확인할 수 있다. 그림 3은 ST가 두 가지 방식으로 학습된 DQN 모델로부터 탐욕 정책에 따라 행동하면서 얻을 수 있는 전송 성능을 도시하였다. 500개의 타임 스텝으로 구성된 유한한 에피소드를 500번 반복하면서 확인해 본 결과, 제안된 EO 패널티를 적용한 학습을 통해 DQN이 충분히 학습한 경우에 근접한 성능을 얻을 수 있음을 확인하였다.

표 1. DQN의 하이퍼파라미터

Hyper-parameter	Value
Number of hidden layers	2
Activation function	"ReLU" for hidden layers, "Linear" for output layer
Optimization	Adam Optimizer
Learning rate	0.0001
Discount factor	0.99
Epsilon	1 → 0.01
Batch size	64
Size of replay memory	2,000

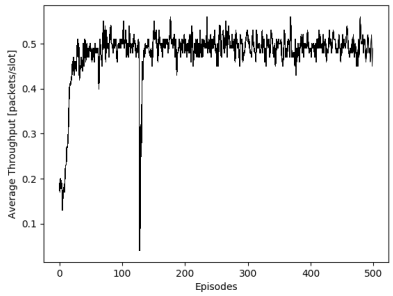


그림 1. DQN을 통한 제안 기법의 학습 과정 (Learning Steps: 250,000)

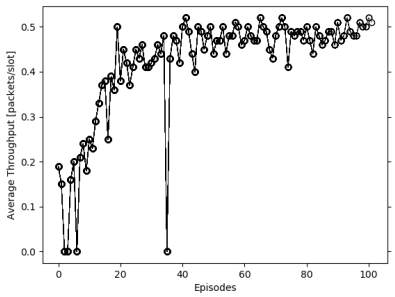


그림 2. 제안하는 EO Penalty를 적용한 학습하는 과정

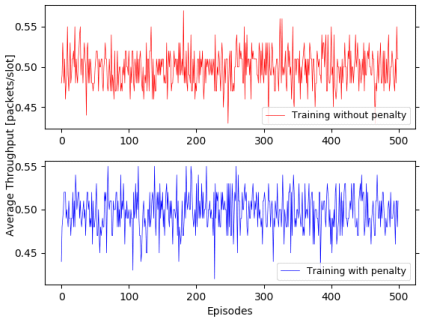


그림 3. 학습된 모델로부터 모드를 선택하여 수행한 성능 비교

ACKNOWLEDGMENT

본 논문은 2014년 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구결과임 (2014R1A5A1011478).

참 고 문 헌

[1] D. Hoang *et al.*, "Ambient backscatter: A new approach to improve network performance for RF-powered cognitive radio networks," *IEEE Trans. Commun.*, vol. 65, no. 9, pp. 3659-3674, Sept. 2017.

[2] V. Liu *et al.*, "Ambient backscatter: Wireless communication out of thin air," *Proc. ACM SIGCOMM 2013*, Hong Kong, China, Aug. 2013.

[3] V. Mnih *et al.*, "Playing Atari with deep reinforcement learning," *Proc. NIPS 2013*, Lake Tahoe, USA, Dec. 2013.

MATLAB에서 회전형 도립 진자 제어를 위한 DDPG 기반 멀티에이전트 강화 학습

지창훈[1], 김주봉[1], 최호빈[1], 임현교[2], 한연희[1+]
한국기술교육대학교 미래융합공학전공[1], 한국기술교육대학교 창의융합공학협동과정[2]
{koir5660, rlawnqhd, chb3350, glenn89, yhhan}@koreatech.ac.kr

Multi-Agent Reinforcement Learning for Rotary Inverted Pendulum in MATLAB

Chang-Hun Ji[1], Ju-Bong Kim[1], Ho-Bin Choi[1], Hyun-Kyo Lim[2], Youn-Hee Han[1]
Future Convergence Engineering, Korea University of Technology and Education[1]
Department of Interdisciplinary Program in Creative Engineering, Korea University of Technology and Education[2]

요 약

강화 학습은 최적의 행동을 찾을 때까지 반복 학습을 수행한다. 이런 강화 학습의 특징은 많은 장점에도 불구하고 현실 세계에서 강화 학습 적용을 어렵게 만든다. 본 논문은 제어분야에서 제어 시스템을 설명하기 위해 많이 사용되는 Rotary Inverted Pendulum을 3D 모델링 하여 가상 환경을 구축하고 구축된 가상 환경의 시뮬레이션을 통해 강화 학습의 반복적인 학습으로 인한 실제 환경의 제약을 해결할 수 있다. 따라서, 본 논문에서는 구축된 가상 Rotary Inverted Pendulum 시뮬레이션을 이용하여 Multi-Agent 강화 학습을 수행함으로써 이를 검증한다.

Ⅰ. 서론

강화 학습은 주어진 환경에서 에이전트가 임의의 행동을 선택하여 문제를 해결하는 기계학습의 한 종류이다. 강화 학습은 게임, 로봇 제어, 네트워크 통신과 같은 복잡한 환경에 적용할 수 있고 관련 분야의 전문 지식이 없어도 최적의 행동을 찾아낼 수 있다는 장점이 있어 주목받고 있다 [1]. 그러나 강화 학습은 최적의 행동을 찾기 위한 반복 학습으로 인해 환경의 제약이 있다. 환경의 제약은 현실 세계에서의 강화 학습 적용을 어렵게 한다. 반복 학습으로 인한 강화 학습 환경의 제약을 해결하기 위해 본 논문에서는 Rotary Inverted Pendulum(RIP) 가상 환경에서 Multi-Agent 강화 학습을 수행하여 환경의 제약을 극복한다.

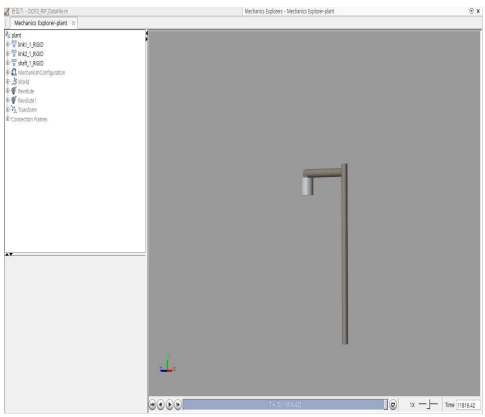
Ⅱ. 본론

1. Rotary Inverted Pendulum 환경

RIP는 비선형적이고 불안정한 동적 시스템으로 실험이 간단하여 제어 시스템 분야에서 검증을 위한 환경으로 사용되어왔다. RIP는 Pendulum, Motor와 Pendulum을 연결하는 Arm으로 구성되어 있으며, Pendulum을 도립 시키고 유지하는 환경이다. 제어 시스템 분야에서는 RIP를 제어하기 위해 복합적으로 여러 controller를 동시에 사용한다 [2].

RIP 실제 환경에 강화 학습을 적용하여 훈련 시키기에는 부품의 소모, 에너지 소비와 오랜 시간의 학습이 필요로 하는 제약 사항들이 존재한다. 따라서, 실제 환경의 제약 사항들을 해결하기 위해 시뮬레이션 환경에 강화 학습을 적용하여 훈련을 진행한다.

본 논문은 대표적인 3D 모델링 프로그램 중 하나인 Solid Works로 RIP를 3D 모델링하고 공학용 프로그램인 MATLAB의 시뮬레이션을 이용해 강화 학습을 위한 환경을 구성한다. MATLAB 시뮬레이션으로 구성한 RIP를 통해 Pendulum 각도 θ_P, Pendulum 속도 w_P, Arm 각도 θ_A와 Arm 속도 w_A를 얻는다. MATLAB 시뮬레이션은 초기 상태일 때 θ_P, θ_A = 0이다 (<표 2> 참조). (그림 1)은 MATLAB 시뮬레이션의 초기 상태를 보여준다.



(그림 1) Rotary Inverted Pendulum MATLAB 시뮬레이션 초기 상태

2. 강화 학습 에이전트

강화 학습 에이전트는 환경으로부터 상태(State)와 보상(Reward)을 받아 행동(Action)을 수행하고 다시 환경으로부터 상태와 보상을 받는 상호작용을 한다. 이러한 상호작용을 바탕으로 보상을 최대화하는 것이 에이전트의 목표이다.

RIP를 비롯한 로봇 제어는 연속적인 제어 값을 추출해야 한다. 따라서 본 논문은 심층 네트워크를 이용해 지정한 행동 범위 내에서 연속적인 행동을 뽑아내는 Deep Deterministic Policy Gradient (DDPG) 알고리즘을 적용한 에이전트를 사용한다 [3].

본 논문에서는 RIP를 역할에 따라 swing-up, balancing 2단계로 나눈다. swing-up은 Pendulum을 도립 시키는 역할이고, balancing은 도립 시킨 Pendulum을 유지 시키는 역할이다. swing-up과 balancing의 행동 범위는 상이하기 때문에 행동 범위가 다른 2개의 에이전트를 이용한다. swing-up과 balancing을 나누는 기준과 행동 범위는 <표 1>과 같다.

+ : 교신저자 한연희(한국기술교육대학교)

<표 1> 강화학습의 에이전트 기준, 행동 범위, 해당 status

에이전트 종류	swing-up	balance
기준	$\theta_P < 177^\circ$, $\theta_P > 183^\circ$	$177^\circ < \theta_P < 183^\circ$
행동 범위	-0.035 ~ 0.035(torque)	-0.01 ~ 0.01(torque)
해당 status	swing-up, swing-up to balance	balance, balancing to swing-up

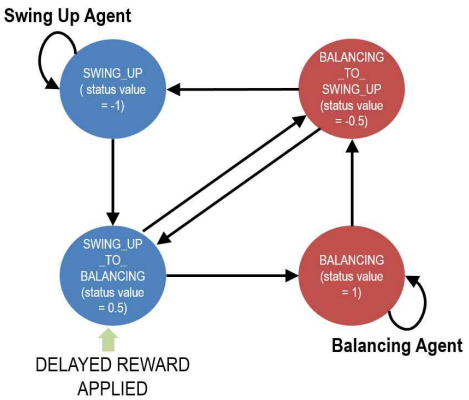
3. 학습 환경 및 구조

본 논문에서는 MATLAB의 python API를 이용해 MATLAB 시뮬레이션을 수행할 때 python으로 구성된 강화 학습 환경에 정보를 실시간으로 전달한다. 우리는 MATLAB 시뮬레이션이 주는 정보를 에이전트가 이해할 수 있도록 재구성하여 상태와 보상을 정의한다.

<표 2> 전달받는 정보, 에이전트의 상태, 보상

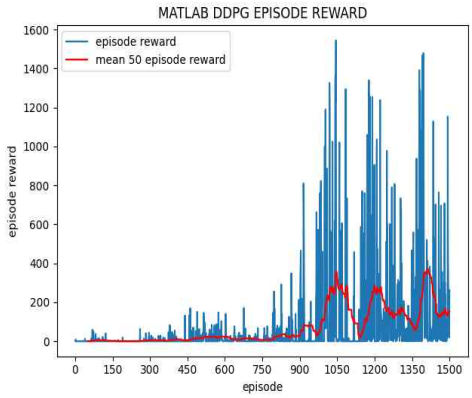
환경이 시뮬레이션에서 전달받는 정보					
Pendulum 각도 (radian)		Pendulum 속도 (radian/s)	Arm 각도 (radian)		Arm 속도 (radian/s)
$0 < \theta_P < 2\pi$		w_P	$0 < \theta_A < 2\pi$		w_A
상태					
$\cos(\theta_P)$	$\sin(\theta_P)$	w_P	$\cos(\theta_A)$	$\sin(\theta_A)$	current status value = (e.g. -1, -0.5, 0.5, 1)
보상					
$\{(2*\pi-\theta')-0.001 * w_P^2 - 50 * \lvert action \rvert * \theta_P\}/1000$ $\theta' = \begin{cases} 2\pi - \theta_P & \text{if } \theta_P > \pi \\ \theta_P & else \end{cases}$					

<표 2>는 MATLAB 시뮬레이션에서 받은 정보로 상태와 보상을 재구성하는 것을 보여준다. 상태는 ($\cos(\theta_P)$, $\sin(\theta_P)$, w_P, $\cos(\theta_A)$, $\sin(\theta_A)$, current status value)으로 정의된다. current status value는 4개로 구성된 status 중 현재 status를 나타내는 value 값을 말한다 ((그림 2) 참조). current status value를 상태에 넣음으로 에이전트는 현재 어떤 status인지 알 수 있다. 보상 식의 θ'는 완전히 도립한 Pendulum의 각도인 π를 가장 높은 값으로 갖도록 θ_P를 변환시킨 것이다. 에이전트는 θ_P가 π에 가깝고 w_P^2와 |action|이 작을수록 큰 보상을 받는다.



(그림 2) status 구성 및 status value

(그림 2)와 같이 RIP 강화 학습 환경은 4개의 status로 나뉜다. swing-up to balancing status와 balancing to swing-up status는 에이전트가 바뀌는 step에만 할당된다. swing-up status와 balancing status는 에이전트가 유지되면 할당된다. 각 status의 에이전트는 <표 1>에서 알 수 있다. 강화 학습 환경이 swing-up to balancing status를 할당 받게 되면 에이전트에게 보상을 넘겨주지 않고 보류시킨다. 그 후 강화 학습 환경 에이전트가 바뀌어 balancing to swing_up status를 할당받게 되면 그동안 balancing status에서 누적된 보상들로 보류 시켰던 swing_up to balancing status의 보상을 대체한다. 이로 인해 swing-up 에이전트는 balancing이 오래 유지되면 더 높은 보상을 얻기 때문에 도립 시킨 후 유지되도록 쉽게 행동을 뽑는다.



(그림 3) 학습 결과 그래프

(그림 3)의 파랑 선은 매 episode의 보상이고 빨간 선은 50 episode의 평균 보상이다. (그림 3)을 보면 DDPG를 이용한 학습의 episode가 증가함에 따라 보상이 증가하는 것을 볼 수 있다. 따라서, 가상 환경에서 RIP의 제어 문제를 성공적으로 해결하고 있음을 알 수 있다.

Ⅲ. 결론

본 논문에서는 강화 학습 환경의 제약을 극복하기 위해 RIP 가상 환경에서 학습을 수행하였다. RIP 시스템은 서로 상이한 행동 범위를 가진 2단계로 구분되었기 때문에 2개의 강화 학습 에이전트를 이용해 학습을 진행하였고 학습이 성공한 것을 확인하였다.

가상 환경의 학습 성공은 실제 환경의 학습에도 좋은 영향을 미칠 수 있다. 성공한 가상 환경을 가진 강화 학습의 좋은 경험들을 실제 환경에서 강화 학습을 수행할 때 에이전트에게 미리 줄 수 있다면, 에이전트의 학습이 빠르게 수렴하는 것을 기대할 수 있다 [4]. 현재 가상 환경의 강화 학습을 끝내고 실제 환경과 연동하기 위한 환경을 구축 중이고 계속적으로 피드백하고 있다.

ACKNOWLEDGMENT

이 논문은 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행 기초연구사업임(No. 2018R1A6A1A03025526 및 NRF-2020R1I1A3065610)

참 고 문 헌

[1] Mnih et al. "Human-level control through deep reinforcement learning." Nature 518 , no. 7540 (2015): 529--533.

[2] Potsaid el al. "Optimal mechanical design of a rotary inverted pendulum.." Paper presented at the meeting of the IROS, 2002.

[3] Lillicrap el al. "Continuous control with deep reinforcement learning.." Paper presented at the meeting of the ICLR, 2016.

[4] Parisotto el al. "Actor-Mimic: Deep Multitask and Transfer Reinforcement Learning.." Paper presented at the meeting of the ICLR (Poster), 2016.

강화학습을 사용한 이미지 처리 기법 기반 적대적 사례 생성에 관한 연구

강효은, 김용수, 홍윤영, 이상현, 김호원*

*부산대학교

{hyoeun0915, dkgoggog0329, hyy0238, jdsd2233, howonkim}@pusan.ac.kr

A Study on the Adversarial Example with Reinforcement Learning for Image Processing

Kang Hyo Eun, Kim Yong Su, Hong Yoon Young, Lee Sang Hyun, Kim Ho Won*

*Pusan National Univ

요 약

딥러닝 모델은 영상, 음성 및 자연어 처리 등 다양한 분야의 산업에서 사용되고 있다. 특히, 최근 활발히 연구가 진행되고 있는 자율 주행 분야에서도 도로 교통표지판 검출 및 판단에 영상인식 딥러닝 모델을 채택하는 추세이다. 이에 따라 딥러닝 모델의 취약점을 악용한 공격에 관한 연구도 활발히 이루어지고 있다. 본 논문에서는 공격 대상 교통표지판의 마스크(Mask)를 추출하고 원본 이미지에 합성하여 적대적 사례(Adversarial Example)를 생성하는 기법을 제안한다. 적대적 사례를 생성할 때 강화학습 기반 이미지 처리 기법을 사용하여 원본과 유사하면서 타겟 클래스에 대한 공격을 수행할 수 있도록 구현하였다. 제안하는 공격 기법은 VGG16, ResNet32, MobileNetV2에 적용하여 이미지 분류 모델의 오작동이 생길 수 있음을 검증하였다.

Ⅰ. 서 론

최근 인공지능의 이미지 인식 결함을 드러내는 적대적 사례(Adversarial examples) 공격이 큰 이슈로 떠오르면서 이를 위한 공격 및 방어 모델에 대한 연구가 화두가 되고 있다. 적대적 사례 공격은 목표로 하는 딥러닝 모델의 성능 저하를 위해 입력 데이터에 눈에 보이지 않는 오류(perturbation)을 더하는 공격이다. 정교하게 생성된 오류는 사람이 인식하기 어렵지만, 딥러닝 모델의 성능 저하로 연결될 수 있다. 따라서 딥러닝 모델의 예측 정확도 하락, 미탐율 증가 등을 불러일으킨다.

특히, 자율주행 영역에서 도로 교통표지판은 모양과 색상이 규격화되어 있어 공격 대상으로 사용될 위험이 있다. 도로 교통표지판을 대상으로 공격할 경우 돌발 상황이 발생하여 차량 혹은 차량 탑승자에게 치명적인 결과를 불러일으킬 수도 있다. 본 논문에서는 공격 대상 교통표지판의 특징 마스크(Mask)를 추출하여 교통표지판에 합성함으로써 교통표지판 인식 모델이 오작동하도록 공격한다. 마스크를 합성할 때 이미지 밝기 변화, 픽셀 값 변화 등 강화학습 기반 이미지 처리 기법을 사용하여 눈에 보이지 않는 적대적 사례를 생성하는 알고리즘을 제시한다.

Ⅱ. 본론

본 연구에서는 A3C(Asynchronous Advantage Actor-Critic) [1] 기반 적대적 사례 생성 방안에 대하여 소개한다. A3C는 여러 에이전트가 독립적으로 개별 환경에서 행동(action), 보상(reward)을 발생시켜 나가며 경험 데이터를 만든 후 loss로부터 계산한 gradient를 글로벌 네트워크로 업데이트한다. 글로벌 모델은 각 로컬 모델에 의해 비동기적으로 업데이트되며, 각 로컬 모델은 업데이트된 글로벌 모델의 파라미터를 복사한다.

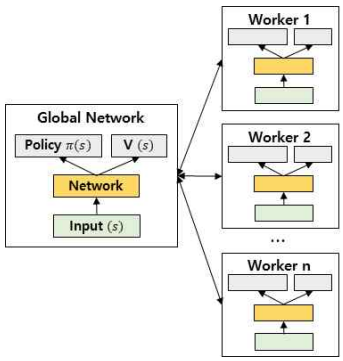


그림 1. A3C 기본 구조

본 연구에서 제안하는 강화학습 기반 적대적 사례 생성 프로세스는 아래 그림2와 같다. 환경(Environment)에서 타겟 교통표지판 이미지가 주어지며, 최초 입력 이미지에 객체 검출 알고리즘인 HOG[2] 알고리즘을 적용하여 표지판의 특징 마스크를 추출한다. 마스크를 합성한 이미지로부터 A3C 학습이 진행된다.

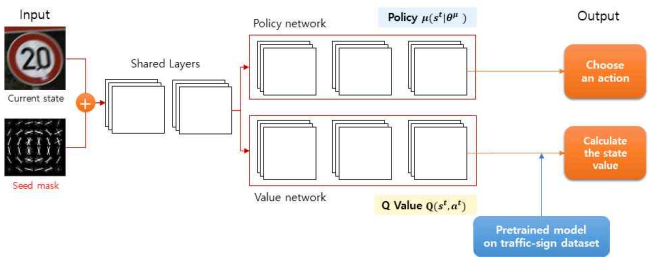


그림 2. A3C 기반 적대적 사례 생성 프로세스

액터(Actor)는 현재 이미지에 대하여 액션을 수행하며 이미지를 변경한다. 평가자(Critic)는 수정된 이미지에 대하여 원본 이미지와의 차이를 계

산하고 현재 state의 가치를 평가한다. 이 과정을 여러 로컬 모델이 독립적으로 수행하여 최종적으로 주어진 표지판에 대하여 최적의 교란 신호(perturbation)를 생성한다.

표 1. 수행 가능한 Action 목록

Seq	Action	Range
1	Pixel value+1	0~255
2	Pixel value -1	0~255
3	Blur	3×3
4	Do nothing	

본 연구에서는 GTSRB(German Traffic Sign Recognition Benchmark) [3] 데이터세트를 사용하여 실험을 진행하였다. GTSRB 데이터세트는 도로교통표지판 43개의 클래스에 대하여 총 50,000장의 이미지로 구성되어 있다. 공격 모델은 이미지 분류 문제에 높은 성능을 보이는 VGG16[4], ResNet32[5], MobileNetV2[6]으로 공격 성능을 실험하였다. 실험결과는 표2와 같다. 정상 모델의 정확도와 적대적 공격을 적용한 모델의 정확도를 기록하였다.

표 2. 각 공격 대상 모델의 분류 정확도

Target Model	Original Acc.	After Attack Acc.
VGG16	91.9%	16.3%
ResNet32	92.1%	12.7%
MobileNetV2	92.3%	15.8%

Ⅲ. 결론

본 연구에서는 타겟 이미지에 대한 특징 마스크와 강화학습 기반 이미지 처리 기법을 사용하여 적대적 사례를 생성하는 방법을 제안하였다. 제안하는 기법은 주어진 마스크 합성 이미지에 대하여 최적의 적대적 사례를 생성하기 위한 이미지 처리 action을 학습한다. 학습이 완료되면 입력 이미지에 대한 공격 과정을 시각화하여 확인할 수 있다. 향후에는 물리적 대상에 대한 공격 실험을 추가하여 실제 환경에서도 강인한 공격 모델과 이에 대한 방어 모델을 연구할 계획이다.

ACKNOWLEDGMENT

This work is financially supported by Korea Ministry of Land, Infrastructure and Transport(MOLIT) as 「Innovative Talent Education Program for Smart City」

참 고 문 헌

[1] Mnih, Volodymyr, et al. "Asynchronous methods for deep reinforcement learning." International conference on machine learning. 2016.

[2] Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection." 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05). Vol. 1. IEEE, 2005.

[3] J. Stallkamp, et al., "Man vs. Computer: Benchmarking machine learning algorithms for traffic sign recognition," Neural Networks : the official journal of the International Neural Network Society, 32, 2012

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

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

[6] Sandler, Mark, et al. "Mobilenetv2: Inverted residuals and linear bottlenecks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

수중 IoT 플러딩 영역 최적화를 위한 강화학습 프로세스에 대한 연구

강현우 [1], 이성원 [2], 서준호 [3], 김동균 [3]
[1] 한국폴리텍대학 대구캠퍼스, [2] 대구한의대학교 스마트 IT 융합학부,
[3] 경북대학교 IT 대학 컴퓨터학부
hwkang@kopo.ac.kr, lsw5359@dhu.ac.kr, junhoseo@knu.ac.kr, dongkyun@knu.ac.kr

A Study on the Reinforcement Learning Process for Flooding Area Optimization in Underwater IoT

Hyunwoo Kang[1], Sungwon Lee[2], Junho Seo[3], Dongkyun Kim[3]
[1]Korea Polytechnics, [2] Deagu Hanny Univ., [3]Kyungpook Nat. Univ.

요 약

본 논문은 수중 센서네트워크에서의 대표적인 라우팅 기법인 DFR (Directed Flooding based Routing)에서 플러딩(Flooding) 영역을 효율적으로 설정하기 위하여 강화 학습을 사용하는 방법을 제안한 논문이다. 실측된 데이터를 기반으로 임의의 링크 품질을 설정한 이후 패킷의 수신율과 전송 횟수에 의해서 노드들이 최적의 플러딩 영역을 학습하는 방법을 제안하였다. 향후 라우팅 프로토콜뿐 아니라 다양한 네트워크 파라메터를 최적화하는데 있어 제안한 기법을 응용할 수 있을 것으로 예상된다.

Ⅰ. 서 론

최근 수중 센서 네트워크는 오염도 측정, 전략 감시, 항만 트래픽 관리 등 다양한 응용이 수행되고 있으며, 이에 따라 센서 노드뿐만 아니라 수중 드론, 경량형 잠수함, 고래 등 수중동물에 부착한 생체 센서 등 다양한 통신장비들이 사용되고 있는 수중 IoT (Internet of Underwater Things) 환경을 구축하고 있다[1]. 또한 인공지능 등 다양한 최신 기술을 수중 IoT 환경에 적용하여 새로운 응용의 신뢰성 있는 동작을 지원하려는 연구가 진행되고 있다[2]. 이러한 연구들은 주로 라우팅 계층과 데이터 링크 계층의 프로토콜을 중심으로 진행되고 있는 추세이다[3].

기존 수중 센서 네트워크에서는 수중 통신의 낮은 전송 신뢰성을 극복하기 위해 라우팅 프로토콜로 플러딩 기반 라우팅 프로토콜을 주로 사용한다. 대표적인 라우팅 프로토콜인 DFR(Directed Flooding based Routing) 프로토콜에서는 각 노드가 주변 노드들과의 평균 에러율을 측정하고, 측정된 에러율에 따라 다중 경로의 숫자를 조절하는 방법을 사용하고 있다[4]. 다시 말하여, 링크 품질이 낮을수록 다수의 경로를 사용하여 특정 경로의 패킷이 손실되더라도 다른 경로로 패킷이 전달될 수 있도록 하고, 반대로 링크 품질이 높다면 다중 경로의 숫자를 줄여 네트워크 자원을 절약한다. 이러한 에러율을 고려한 플러딩 기반 라우팅 프로토콜은 노드들이 균일하게 분포하고 노드들 간 성능이 거의 유사한 기존 센서 네트워크에서는 높은 신뢰성을 제공하면서도 네트워크 자원을 절약할 수 있는 효율적인 라우팅 프로토콜이다.

그러나, 수중 IoT 환경에서는 다양한 성능의 노드들이 응용에 따라 배치되어 동작하기 때문에 노드들간 통신 성능이 다르고, 노드간의 배치도 불균일한 경우가 많다. 이러한 환경에서 DFR 이 동작할 경우에는 다음과 같은 다양한 문제가 발생한다고 알려져 있다. 아래 그림 1 은 전송 노드 T 가 싱크 노드로 패킷을 전송하는 상황에서 발생할 수 있는 이러한 문제를 도식화한 것이다.

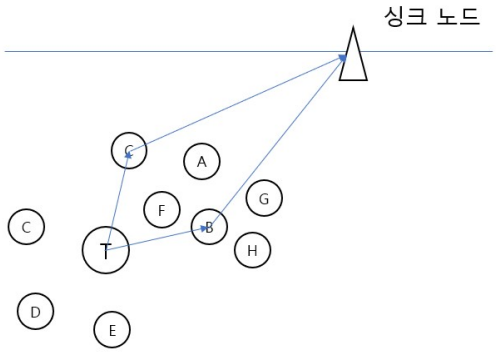


그림 1. DFR 의 플러딩 영역

1) 후방 이웃 노드의 ETX 반영 문제: 위 그림 1 에서 노드 C, D 와 E 는 노드 T 의 후방에 있으므로, base angle 을 아무리 작게 설정하더라도 전송 영역에 포함되지 못한다. 하지만 기존 DFR 에서는 모든 이웃 노드의 링크 품질 ETX 을 평균에 반영하기 때문에, 만약 노드 C, D, E 의 링크 품질에 좋거나 나쁘다면 전송 영역이 불필요하게 작아지거나 커질 수 있다.
2) 전송 영역 설정 패러독스(paradox): 위와 같은 문제를 해결하기 위해 전송 영역에 포함되는 노드의 링크 품질만을 고려하여 플러딩 영역을 설정하고자 하는 연구가 진행되었다. 이러한 연구에서는 명백한 후방 노드인 노드 C, D, E 의 링크 품질을 평균 링크 품질에 반영하지 않을 수 있으나, base angle 에 따라 플러딩 영역에 포함될 수도, 그렇지 않을 수도 있는 지역에 위치한 노드 B, C 의 링크 품질을 적절히 반영하는 방법은 아직까지 제안되지 못했다.

3) 비대칭 노드 분포 문제: 위 그림 1 에서 노드 C 와는 달리 노드 B 에는 많은 이웃 노드가 존재한다. 이 경우 노드 C 가 전송 영역에 포함되도록 전송 영역을 설정하지 않는다면, 노드 B 쪽으로만 패킷 전송이 발생하여 전송에 참여한 노드의 숫자는 충분하더라도 양질의 다중경로가 형성되지 못하게 된다.

이러한 문제들을 해결하기 위해서는 이웃 노드들의 위치, 배치 상태, 이동성 및 대기 큐의 길이 등 다양한 메트릭을 동시에 고려하여 최적의 플러딩 영역을 설정하기 위한 연구가 필요하다. 이러한 연구는 기존 수중 센서 네트워크에서도 진행되었으나, 기존 연구에서는 둘 이상의 메트릭을 한번에 고려하기에는 역학론적인 방법론이 부족하였기 때문에 최적의 플러딩 영역에 영향을 미칠 수 있는 요소를 제안하는 방향으로 진행되었다. 하지만 최근에는 기계학습을 활용하여 다양한 메트릭들을 함께 고려하여 최적의 값을 찾아내는 연구가 성과를 보이고 있다. 따라서 본 논문에서는 다양한 요소들을 활용하여 노드들을 미리 기계 학습시킨 후, 수중에 배치함으로써 최적의 플러딩 영역을 설정하도록 하는 과정과 그 학습 방법론을 제안한다.

Ⅱ. 본론

본 논문에서 제안하는 기계학습 기반 플러딩 영역 설정 프로세스는 다음과 같이 동작한다.

1) 기계학습 환경 생성 과정: 전송 노드 T 의 주변에 랜덤하게 생성된 N 개의 노드를 랜덤한 위치에 배치시킨다. 이후 랜덤하게 생성한 각 센서 노드들에게 34%~87% 사이의 랜덤한 링크 품질을 설정한다. 위 수치는 대한민국에서 측정한 실측 데이터를 기반으로 하였으며, 향후 전송 기법의 발전에 따라 수정될 수 있다. 또한 전송 노드 T 로부터 2 Hop 이상, 3 Hop 이하의 통신 거리에 노드들을 균일하게 배치한다. 본 논문에서는 이러한 노드들은 보상 노드라 칭한다.
2) 초기학습 과정: 현재 설정된 이웃 노드들의 분포 map 과 각 이웃 노드들의 링크 품질을 State 로 지정하고, 전송자 노드는 랜덤한 base angle 을 설정하여 패킷을 전송한다. 이 때 이 패킷을 수신한 모든 노드는 항상 자신이 수신한 패킷을 브로드캐스트한다.
3) 강화학습의 보상 과정: 보상 노드들은 자신의 패킷 수신 유무를 로그로 남겨 싱크 노드에게 전송한다. 또한 보상 노드들이 아닌 노드들은 자신이 패킷을 전송에 참여했는지를 로그로 남겨 이를 싱크 노드에게 전송한다. 싱크 노드는 패킷을 수신했다는 로그를 남긴 보상 노드들에 따라 Reward 값을 증가시킨다. 이후 전송에 참여한 이웃 노드 수에 따라 Reward 값을 감소시킨다.
4) 보상 피드백 과정과 Action: 싱크 노드는 측정된 Reward 값을 소스 노드에게 전송한다. 소스 노드는 수신된 Reward 값이 base angle 값이 최대화될 수 있도록 새로운 base angle 을 선택하는 Action 과정을 수행한다. 하지만 일반적인 전송자 노드가 기계학습을 수행할 정도의 연산능력을 보유하지는 못하므로, 이 프로세스는 전송자 노드와 하드웨어적으로 연결된 외부 병렬처리 장비를 사용하여 진행한다. 해당 테스트 케이스에서 base angle 이 더 이상 변화하지 않는다면 (플러딩 영역의 변경으로는 더 이상 Reward 값을 증가시키지 못한다면) 노드의 배치와 링크 품질을 변경하여 새로운 학습을 수행한다.
5) 노드의 배치와 동작: 기계학습이 완료된 장비는 외부 병렬처리 장비와 분리하여 실제 수중 환경에 설치되어 동작한다. 이 과정에서 주기적인 hello 메시지를 교환하되, 자신의 해당 hello 메시지에는 자신의 위치 정보를 포함시켜 전송한다. 이웃 노드들의 모든 위치 정보와 평균 에러율을 수집한 전송자 노드는 자신의 학습 데이터를 활용하여 최적의 base angle 을 선택하여 전송함으로써 전송 신뢰성을 높이고 네트워크 자원을 절약할 수 있다.

Ⅲ. 결론

본 논문에서는 수중 IoT 환경에서 사용되는 플러딩 기반 라우팅 프로토콜에서 최적의 플러딩 영역을 설정하기 위해 강화학습법을 적용할 수 있는 학습 프로세스를 제안하였다. 특히 제안된 학습 프로세스는 랜덤한 노드들의 배치와 링크 에러율을 State 로, 보상 노드들의 수신률과 전송 횟수를 Reward 로, 플러딩 영역을 설정하는 것을 Action 으로 하여 강화학습을 수행하였다. 이 학습법을 통해 강화학습된 노드들이 기존 방식에 비해 효율적으로 플러딩 영역을 설정한다는 사실이 검증된다면, 이러한 학습 프로세스를 라우팅 파라메터 외에도 더욱 다양한 네트워크 파라메터의 최적화에 적용할 수 있을 것으로 예상된다.

ACKNOWLEDGMENT

이 논문은 2020 년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (NRF-2016R1D1A3B01015510)

참 고 문 헌

[1] 서준호, 김민석, 이성원, 김동균. (2019). 적은 네트워크 오버헤드를 소모하여 중복 주소 감지를 수행하는 IoUT 를 위한 클러스터 헤더 선정 기법. 한국통신학회 학술대회논문집, pp. 1582-1583.

[2] J. Yan, Y. Gong, C. Chen, X. Luo and X. Guan, "AUV-Aided Localization for Internet of Underwater Things: A Reinforcement-Learning-Based Method," in IEEE Internet of Things Journal, vol. 7, no. 10, pp. 9728-9746, Oct. 2020.

[3] E. Liou, C. Kao, C. Chang, Y. Lin and C. Huang, "Internet of underwater things: Challenges and routing protocols," 2018 IEEE International Conference on Applied System Invention (ICASI), Chiba, 2018, pp. 1171-1174.

[4] Daeyoup Hwang and Dongkyun Kim, "DFR: Directional flooding-based routing protocol for underwater sensor networks," OCEANS 2008, Quebec City, QC, 2008, pp. 1-7

수중 IoT CoAP 에서 최적 메시지 타입 결정을 위한 강화 학습 기반 기계 학습 프로세스

이성원 [1], 강현우 [2], 서준호 [3], 김동균 [3]
[1] 한국폴리텍대학교 (강현우), [2] 대구한의대학교 스마트 IT 융합학부,
[3] 경북대학교 IT 대학 컴퓨터학부
lsw5359@dhu.ac.kr, hwkang@kopo.ac.kr, junhoseo@knu.ac.kr, dongkyun@knu.ac.kr

A Reinforcement Learning based Machine Learning Process for Optimized Message Type Determination in CoAP for Underwater IoT

Sungwon Lee[1], Hyunwoo Kang[2], Junho Seo[3], Dongkyun Kim[3]
[1]Deagu Hanny Univ., [2]Korea Polytechnics, [3]Kyungpook Nat. Univ.

요 약

본 논문은 수중 IoT 네트워크에서의 대표적인 전송 프로토콜인 CoAP(Constrained Application Protocol)에서 강화학습을 사용하여 메시지 타입을 효율적으로 설정하기 위한 학습 프로세스 및 기법을 제안한 논문이다. 제안된 학습 프로세스에서는 우선 강화학습법을 사용하여 응용의 정상적인 동작을 지원하는 최소한의 Confirmable 메시지를 찾아낸다. 이후 해당 강화학습 데이터를 기반으로 하여 첫 번째 데이터가 전달되었을 때 N 번째 패킷이 응용의 정상적인 동작에 유의미한 영향을 미칠 가능성을 확률 함수로 계산한다. 소스 노드는 해당 확률함수를 기반으로 하여 전송해야 할 메시지의 타입을 결정하고 전송한다.

Ⅰ. 서 론

최근 수중 센서 네트워크는 오염도 측정, 전략 감시, 항만 트래픽 관리 등 다양한 응용이 수행되고 있다. 이에 따라 수중 드론, 경량형 잠수함, 수중동물에 부착한 생체 센서 등의 통신장비들이 IoUT (Internet of Underwater Things) 환경을 구축한다[1]. 이러한 IoUT 환경에 인공지능 등 다양한 인공지능 관련 기술을 프로토콜에 적용하여 새로운 응용의 신뢰성 있는 동작을 지원하려는 연구가 진행되고 있다[2][3].

그 중에서도 CoAP(Constrained Application Protocol)는 응용의 요구에 따라 메시지 전송방식을 변경할 수 있는 대표적인 응용/전송 계층 프로토콜이다. CoAP 에서는 응용이 전송 신뢰성이 높은 데이터 패킷을 생성했을 경우에는 패킷 손실 시 재전송을 수행하는 Confirmable 메시지를 전송하여 신뢰성을 높인다. 반대의 경우에서는 손실을 감수하는 Non-Confirmable 메시지 타입으로 전송하여 효율성을 높인다. 이러한 메시지 타입은 응용이 어떤 과정으로 데이터를 생성했는가에 의해 결정된다. 예를 들어, 네트워크 제어에 관련된 SYN 메시지들은 Confirmable 메시지로 전송하고, 일반 페이로드를 포함한 데이터 패킷은 Non-confirmable 메시지로 전송할 수 있다.

하지만 CoAP 에서는 데이터의 생성과정에 의해서만 메시지 타입을 결정하므로, 같은 종류의 데이터에는 같은 전송 방식이 사용된다. 이로 인해, 감시 정찰 응용 등에서 연속적으로 찍은 스냅샷들이 모두 Confirmable 메시지로 전송되거나 또는 모두 Non-confirmable 메시지로 전송되게 된다. 전자의 경우에는 비슷한 형태의 정보를 과도하게 재전송하면서 불필요한 네트워크 리소스를 소모하게 되고, 후자의 경우에는 낮은 신뢰성에 의해 중요한 내용이 손실될 수 있다. 간단한 해결책으로는 연속적으로 촬영된 내용중 일부만을 랜덤하게 전송하는 방법을 생각할 수 있으나, 이 경우 중요한 내용이 촬영된 부분이 전송에서 제외되어 응용이 비정상적으로 동작할 수 있다.

이러한 문제를 극복하기 위해서는 정보를 수집한 센서 노드가 정보의 중요도를 스스로 판단하여 메시지 타입을 결정할 수 있어야 한다. 최근 기계학습 연구에서는 지도학습 또는 비지도학습 등을 사용하여 데이터에 특정 내용의 포함 유무를 판단하는 기술이 발전하였으나, 연속적인 스냅샷 등에서는 대부분의 경우 동일한 레이블이 포함되기 때문에, 이러한 기술들로는 정확한 데이터 타입을 결정할 수 없다.

이러한 문제를 해결할 가능성이 있는 기술로는 퍼지 이론(Fuzzy Theory)이 언급되고 있다. 퍼지 이론은 인공지능의 한 분야로, 특정 요소가 집합에 소속될 가능성을 확률 함수로 나타내어 불명확한 상황에 대한 수학적인 접근을 정의한 이론이다. 이러한 퍼지 이론에서는 확률 함수를 정확하게 설정하는 것이 중요하며, 최근에는 기계학습을 사용하여 확률함수를 설정하는 연구가 시도되고 있다.

따라서 본 논문에서는 동일한 형태의 레이블을 가진 데이터들이 중복 생성되었을 경우 기계 학습된 머신으로부터 생성된 퍼지함수를 활용하여 CoAP 메시지 전송 방식을 선택하는 과정과 그 학습 방법론을 제안한다. 또한, 본 논문에서는 연속 촬영된 스냅샷(Snapshot) 사진들을 대상으로 하지만, 동일한 레이블을 가지는 데이터가 연속적으로 생성되는 응용에 동일하게 사용될 수 있다.

Ⅱ. 본론

본 논문에서 제안하는 기계학습 기반 확률함수 설정 프로세스는 다음과 같이 동작한다.

1) 기계학습 환경 생성 과정: 본 논문에서 제안하는 기계학습 프로세스에서는 센서 노드가 수중에 배치되기 전, 소스 노드와 싱크 노드를 직접 연결하여 다음과 같은 학습 과정을 거친다. 싱크 노드가 N 번째 데이터를 수신하였을 경우 N+1 메시지는 34%~87% 사이의 확률로 전송되도록 설정한다. 이 수치는 대한민국에서 측정한 실측 데이터를 기반으로 하였으며, 향후 전송 기법의 발전에 따라 수정될 수 있다.
2) 강화학습 프로세스 과정: 수신된 스냅샷을 사용하여 설정된 응용 (잠수함 또는 어군 정보 추출)을 동작시킨다. 이 때 응용이 정상적으로 동작하였을 때는 Reward 를 부여하고, 그렇지 못했을 경우에는 Reward 를 감소시킨다. 이후 동일한 상황에서 손실된 메시지 중 일부를 Confirmable 메시지로 전송되도록 설정되었다고 가정하여 손실없이 전송시켜 다시 Reward 를 부여한다.
3) 재학습 과정: 강화학습 프로세스를 수행하여 Reward 가 최고가 되는 전송 방식을 결정한 이후에는, 이번에는 다시 Confirmable 중 일부를 랜덤하게 손실시키면서 응용의 성공적인 동작을 확인한다. 만약 손실이 발생했을 때에도 응용이 정상적으로 동작한다면 Reward 를 증가시키고, 응용이 동작하지 못한다면 Reward 를 감소시키는 재학습 과정을 설계한다.
4) 확률함수 계산: 재학습과정을 수행한 노드는 첫 번째 패킷이 정상적으로 전달되었을 때 N 번째 패킷이 Non-confirmable 메시지 타입으로 전송되어 랜덤한 손실을 겪였을 때 Reward 가 일정 임계치 이상으로 증가할 확률을 계산한다. 이후 이 확률의 오차를 최소화하는 2 차원 확률함수를 계산한다. 다시 말하여, 첫 번째 패킷이 전달되었을 때(첫 패킷은 항상 Confirmable 메시지로 전달됨을 가정), N 번째 패킷이 응용의 정상적인 전달에 유의미한 영향을 미칠 것이라고 생각될 확률을 확률함수로 계산하는 것이다. 해당 확률함수를 계산한 후, 센서 노드를 수중에 설치한다.
5) 센서 노드의 메시지 타입 결정: 위 과정을 통해 학습된 센서 노드는 다음과 같이 동작한다. 동일한 레이블을 가지는 데이터가 연속적으로 생성되었을 경우 첫 데이터는 항상 Confirmable 메시지로 전송한다. 이후 첫 번째 데이터가 정상적으로 전달되었을 때 N 번째 데이터가 유의미할 확률 K 을 학습된 확률함수를 사용하여 계산한다. 만약 확률 K 가 미리 설정된 임계치 이상일 경우에는 센서 노드가 해당 메시지를 Confirmable 메시지로 전송하고, 임계치 이하일 경우에는 Non-confirmable 메시지로 전송한다.
6) 전송률 피드백 과정: 위 과정에 따라 계산된 확률함수에 의해 첫 번째 패킷 이후의 패킷이 Confirmable 메시지로 전송될 수 있다. 또한, CoAP 표준에서는 Non-confirmable 메시지가 계속 선택되었을 경우에는 혼잡을 감지할 수 없으므로 16 개의 패킷 중 최소 2 개의 패킷은 Confirmable 메시지로 전송되도록 정의하고 있다. 첫 패킷 이후의 Confirmable 메시지를 수신한 싱크 노드는 두 Confirmable 메시지 사이에 수신한 Non-confirmable 메시지를 ACK 메시지에 피기백하여 전송한다. 해당 ACK 메시지를 수신한 소스 노드는 종단간 수신율을 확률함수에 반영하여 새로운 데이터가 유의미할 확률 K 를 계산하여 메시지 타입을 결정한다.

Ⅲ. 결론

본 논문에서는 수중 IoT 환경에서 응용이 동일한 레이블 정보를 가지는 데이터를 연속적으로 생성시켰을 때 패킷의 일부만을 Confirmable 메시지로 전송하여 네트워크 자원을 절약하면서도 응용의 정상적인 동작을 지원하는 CoAP 메시지 선정 기법과 그 프로세스를 제시하였다. 제시된 프로세스에서는 최소한의 Confirmable 메시지를 사용하여 응용이 정상적으로 동작할 때 최고점의 보상이 되도록 강화학습을 설계하였다. 이후 강화학습된 데이터를 기반으로 하여 패킷을 보내는 방식에 따라 보상이 증가할 확률을 확률함수로 계산하는 할 수 있도록 설계하였다. 해당 확률함수를 사용하여 첫 번째 이후에 생성된 패킷이 응용의 정상적인 동작에 유의미한 영향을 미칠 확률이 일정 이상일 경우에 한해 해당 메시지를 Confirmable 메시지로 전송하도록 설계하였다.

ACKNOWLEDGMENT

이 논문은 2020 년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (NRF-2016R1D1A3B01015510)

참 고 문 헌

[1] 서준호, 김민석, 이성원, 김동균. (2019). 적은 네트워크 오버헤드를 소모하여 중복 주소 감지를 수행하는 IoUT 를 위한 클러스터 헤더 선정 기법. 한국통신학회 학술대회논문집, pp. 1582-1583.

[2] J. Yan, Y. Gong, C. Chen, X. Luo and X. Guan, "AUV-Aided Localization for Internet of Underwater Things: A Reinforcement-Learning-Based Method," in IEEE Internet of Things Journal, vol. 7, no. 10, pp. 9728-9746, Oct. 2020.

[3] E. Liou, C. Kao, C. Chang, Y. Lin and C. Huang, "Internet of underwater things: Challenges and routing protocols," 2018 IEEE International Conference on Applied System Invention (ICASI), Chiba, 2018, pp. 1171-1174.

[4] Daeyoup Hwang and Dongkyun Kim, "DFR: Directional flooding-based routing protocol for underwater sensor networks," OCEANS 2008, Quebec City, QC, 2008, pp. 1-7

[5] Junho Seo, Sungwon Lee, Muhammad Toaha Raza Khan, and Dongkyun Kim. 2020. A new CoAP congestion control scheme considering strong and weak RTT for IoUT. In Proceedings of the 35th Annual ACM Symposium on Applied Computing (SAC '20). Association for Computing Machinery, New York, NY, USA, 2158-2162.

실내 환경 자율주행 로봇을 위한 객체 인지 모듈 개발

김명현, 김영인, 최인훈, 허의남*
경희대학교

{freckie, rladuddls3, inhun321, *johnhuh}@khu.ac.kr

An Implementation of Object Recognition Module for Indoor Self-Driving Robot

Myung-Hyun Kim, Yeong-In Kim, In-Hun Choi, Eui-Nam Huh*
Department of Computer Science and Engineering, Kyung Hee University

요 약

최근 비약적으로 발전한 기계 학습 기술은 자율주행 자동차를 실현하는데 큰 축을 담당하고 있다. 다만 실내 주행의 경우, 여러 환경적 제약 조건에 따라 자동차의 현재 위치를 파악하기에 어려움이 있다. 이러한 문제점에 주목하여 본 연구에서는 실시간 영상에서 실내 복도의 문패를 인식해 실내 위치 정보를 제공하는 모듈 개발을 목표로 한다.

1. 서 론

미국의 아마존(Amazon)에서는 물류센터에서 물건을 옮기는 자율주행 로봇이 이미 운영 중이다. 다만 이는 실외나 다름없는 거대한 환경에서의 운행으로, 건물 복도와 같은 협소한 실내 환경과는 거리가 많다. 따라서 본 논문에서는 좁은 통로가 존재하고 사람들의 왕래가 잦으며, 차선 등 가이드라인이 없는 실내 환경에서의 개발을 진행했다.

또한 본 논문에서 개발을 목표로 하는 자율주행 로봇은 실내 운행용으로, GPS 등의 도움을 받기 어려워 정확한 현재 위치를 획득할 수 없다. 대신, 복도를 주행하며 로봇에 부착된 카메라로 실시간 영상을 분석, 복도의 문에 붙어 있는 문패(Door Plate)를 인식해 위치 정보를 획득하는 것을 목표로 한다.

2. 본 론

2.1. 객체 인식 (Object Detection)

객체 인식은 입력 영상에서 여러 객체에 대한 Classification 과 Localization 을 진행한다. 객체 인식을 수행하는 딥 러닝 아키텍처인 R-CNN(Regions with Convolutional Neural Networks)와 YOLO 를 먼저 살펴보고 본 연구에 반영했다.

R-CNN 은 CNN 과 같은 딥 러닝 기반 이미지 분류기를 특징 추출기(Feature Extractor)로 사용하여 객체 인식에 높은 성능을 보인 예시이다. R-CNN 은 입력 영상에서 객체가 있을 것이라고 예측되는 위치를 찾고, 각 객체를 CNN 모델에 통과시켜 특징 벡터를 얻은 후 이를 SVM Classifier 에 통과시키는 2-Stage Object Detection 방식을 수행한다.[1]

다만 이 아키텍처는 비교적 정확하지만 매우 느리다는 단점이 있어, 이를 실시간에서 사용할 수 있게 개선한 모델인 YOLO 를 선택하였다.[2]

본 연구는 문패 속 숫자를 판별하는 것이 목표이므로, YOLO 모델을 숫자 인식 용도로 제한하여 학습했다. 데이터셋으로는 각 숫자의 경계 박스(Bounding Box) 좌표 데이터와 수 데이터를 포함하고 있는 SVHN (Street View House Numbers) 데이터셋[1]을 사용하였다. YOLO 는 자전거, 강아지 등 80 종류의 실생활 객체를 포함한 coco 데이터셋[2]을 사용하였지만, 본 연구에서는 0~9 한 자리의 숫자를 10 종류의 객체로 사용하였다.

[그림 1] SVHN 데이터셋에 경계 박스를 표시한 그림

2.2. 문패 이미지 획득과 숫자 추출 과정의 분리

YOLO 의 실시간 영상 분석은 NVIDIA Titan X GPU 를 사용하여 초당 30 프레임의 처리 속도[3]를 보여준다. 앞서 제시된 R-CNN 보다 처리 속도가 빠르지만, 데스크탑 용 GPU 를 공간 제약 상 사용하기 힘든 자율주행 로봇의 환경에서 모든 프레임을 YOLO 모델로 처리하기에는 큰 어려움이 있다. 따라서 최대한 경량화한 모델 기반으로 구현하며, 문패 이미지를 먼저 획득한 후 YOLO 모델로 처리하도록 알고리즘을 구현하였다.

또한, 라즈베리 파이[3]에서 YOLO 의 처리 속도가 매우 느리므로 사실상 모든 프레임을 YOLO 모델로 처리가 불가능

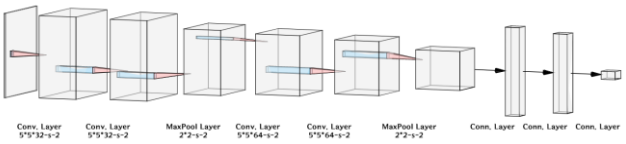
[1] http://ufldl.stanford.edu/housenumbers (format 2)
[2] https://cocodataset.org/#home
[3] Raspberry Pi 4 Model B 4GB 기준

하다. 따라서 본 연구에서는 문패 이미지 획득 과정과 숫자 추출 과정을 분리하여 별개의 연구 목표로 설정하였다.

문패 이미지 획득 과정은 최대한 실시간에 가까운 처리 속도가 요구되며, 숫자 추출 과정은 최대한 높은 정확도가 요구된다. 따라서 문패 이미지 획득 과정은 초당 15 프레임 이상 처리 속도와 80% 이상의 정확도를, 숫자 추출 과정은 500ms 이하의 처리 시간과 95% 이상의 정확도를 목표로 설정하였다.

2.3. 문패 이미지 획득 과정

문패 이미지 획득 과정은 후보 이미지 추출 단계와 문패 이미지 여부 판별 단계, 총 두 단계로 구성하였다. 후보 이미지 추출 과정은 OpenCV 라이브러리를 이용하여 Canny Edge 검출, Contour 넓이 비교, Contour 비율 비교 등의 과정을 거쳐 후보 이미지를 생성한다. 인식된 Contour 은 가로와 세로 비율이 1.0 ~ 3.0 인 이미지로 제한한다. 이렇게 생성된 이미지는 48*48 Grayscale 이미지로 리사이즈되어, 문패 이미지 판별 모델로 전달된다.



[그림 2] 문패 이미지 판별 모델 아키텍처

문패 이미지 판별 모델은 48*48*1 Grayscale 이미지를 입력으로 받아 문패 이미지인지 여부를 출력하는 이진 분류 모델(Binary Classification)이다. 모델 출력이 1 이 되어 문패 이미지로 판별되면 해당 이미지가 크롭된 프레임 이미지 전체를 YOLO 기반 숫자 추출 모델로 전달한다.

2.4. 숫자 추출 과정

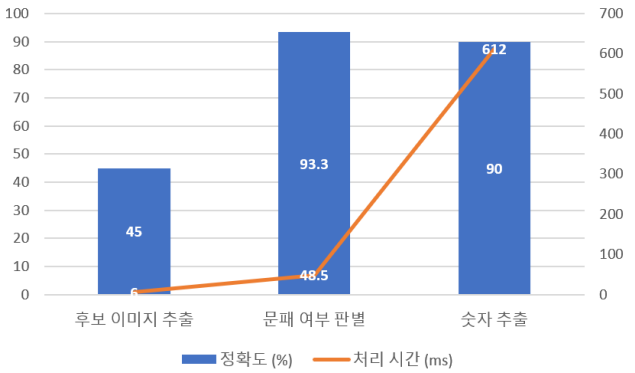
숫자 추출 과정은 프레임 이미지를 YOLO 기반 숫자 추출 모델에 통과하는 단계와 모델의 결과를 처리하여 현재 위치를 문자열로 추출하는 단계로 구성하였다. 프레임 이미지가 숫자 추출 모델에 입력되면, 모델은 인식된 각 숫자들을 박스 좌표와 함께 예측 값을 반환한다. 숫자 추출 모델은 경량화를 위해 YOLOv5s 모델을 SVHN 데이터셋으로 학습하였다.

모델을 통해 예측된 숫자 박스들을 좌표 기준으로 좌상단부터 정렬하여 예측 값을 조합해 숫자 문자열을 최종적으로 반환한다.

그러나 노이즈가 100%로 걸러질 수 없기 때문에 정확한 예측이 실패하는 경우도 있다. 이를 해결하기 위해 최근 20 프레임의 예측 값 중 가장 빈도가 높은 값을 선택해 출력하는 알고리즘을 추가로 구현하였다.

2.6. 성능평가

[그림 3]은 각 과정 별 평균 처리 시간과 정확도를 나타낸 그래프이다. 비교를 위해 문패 이미지 획득 과정의 두 단계를 모두 나타냈다. 또한 숫자 추출 과정의 정확도는 실제 문패의 정보와 예측한 문자열이 정확히 같은 경우 성공하였다고 설정했다.



[그림 3] 각 과정 별 평균 처리 시간 및 정확도

한 프레임의 처리에서 후보 이미지 추출 과정은 6ms 이하의 평균 처리 시간을 보여 전체 처리 과정에 유의미한 영향을 주지 않았다. 그러나 정확도가 45%로 크게 낮아 문패 여부 판별 모델로 한 번 더 필터링을 수행할 필요가 있다.

문패 여부 판별 모델은 48.5ms 의 평균 처리 시간을 보여주어 초당 약 20 프레임을 처리할 수 있다. 정확도 또한 93.3%로 목표치인 80%을 초과하는 성능을 보였다.

숫자 추출 모델은 약 90%의 정확도를 보였지만 612ms 의 평균 처리 시간을 보여 사실상 실시간으로 처리가 어려운 것을 알 수 있다. 따라서 앞의 문패 이미지 획득 과정의 정확도를 최대한 높여 정확한 문패 이미지만이 숫자 추출 모델에 전달되도록 해야 할 것이다.

3. 결론 및 향후 연구

본 연구는 상황에 따라 유연한 대응을 하는 실내 자율주행을 위해, 실내 복도의 문패를 인식해 자율주행 로봇의 위치 정보를 획득하는 모듈을 개발하였다. 약 93.3%의 정확도를 보이는 문패 여부 판별 모델, 90%의 정확도를 보이는 숫자 추출 모델을 이용해 우발적으로 변화하는 실내 환경에도 안정적인 주행을 기대할 수 있다.

향후 연구로, 문패 이미지 획득 과정의 전체적인 정확도를 98% 이상으로 올려 노이즈를 최대한 배제하여 보다 더 최적화된 모듈을 구현할 계획이다. 또한 모듈을 계속 구동하는 방법보다, 문이 인식되었을 때만 모듈을 가동하는 방법으로 라즈베리 파이의 컴퓨팅 자원을 절약할 계획이다.

ACKNOWLEDGMENT

이 논문은 2020 년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No. 2017-0-00294, (대학 ICT 기초연구실) 서비스 이동 지원을 위한 분산형 클라우드 핵심원천기술 연구)

참 고 문 헌

[1] Shanqing Ren, ed., “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks”, 2015.

[2] Joseph Redmon, ed., “You Only Look Once: Unified, Real-Time Object Detection”, pp. 5-6, 2015.

[3] *ibid.* p. 1.

인기 일체형 개인 슬깨비 [슬봇] 시스템 구조 연구

A Study on the Development of man machine integration for the personal intelligence Robot [PR-Inbot] System.

진 용옥[경희대] Y.O..Chin [Kyunghee UNV]

00. 요약[summary]

PC[슬기 틀]나 PT[슬기 전화]를 포함하여 기계와 인간이 일체가 되는 슬기로봇[PR-슬기깨비] 시스템 구성에 대하여 기술한다 PC와 PT를 통합 연동하고 중복기능을 조정하며 슬깨비는 중앙 관제와 조절로 전체를 제어한다 만물 감성체[SoE]와 5G를 연동하여 원방의료 보조기, 개인건강 돌보미, 개인 통신기지국, 개인비서, 개인 레이다 등 휴대 착의형 개인 슬깨비를 설계하고 제작에 대하여 논의한다

Describes the system configuration of Seulgibot [PR-Seulgikebi] in which machines and humans are integrated including PC[Seulgiframe] and PT[SeulgiPhone]. PC and PT are integrated and redundant functions are adjusted.

The whole is controlled by central control By linking all sensing[SoE] and 5G, we design and manufacture portable personal sequins such as medical aids, personal health caregivers, personal communication base stations, personal assistants, and personal radar

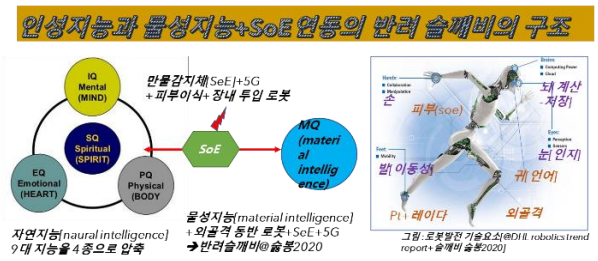
iscuss

열쇠말 [keyword]

열쇠말[KEY WORD]	영문	약어
인간지능	Human intelligence	HI
물성지능	Material intelligence	MI
인공지능	Artificial intelligence	AI
만물감지체	Senser of everything	SoE
슬깨비	Sapience robot	SR
착의컴퓨터	Wearing computer	WC
슬깨비 운영체계	Robot operating system	ROS

10. 머리 말 [서론]

슬깨비[로봇]sms 물성지능[MQ]과 자연지능[NQ] 그리고 감촉체가 통합 연동되는 구조로 정의한다[그림1] 물성지능은 기계나 물질에 부여되는 지능으로 주로 AI를 지칭하지만 여기서는 만물 촉지체 [SoE]에 의한 자연 지능과 연동지능으로 정의한다 만물 감지체란 피부 접촉이나 피부이식으로 장내로 삽입되는 나노 수준의 생체 사이틀[bio interface]이다 모든 입출력은 사이틀로 연결되며 고글[안경]에 통합 실장하고 AR[증간현실]이나 사방 입체 투영[홀로그램]으로 표출되거나 표시된다



[그림1; 지능통합 연동 개념]

[Figure 1; Intelligence Integration Interlink Concept]

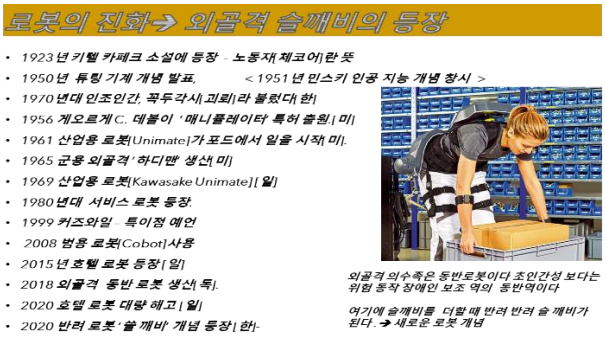


그림2. 슬깨비의 진화와 반려 슬깨비 개념

Figure 2. Evolution of robots and concept of companion robots.

20. 몸말[본론]

왜 인기 일체형 슬깨비 인가?

인공지능(人工知能, artificial Intelligence, AI)은

학습, 추론, 지각, 논증, 언어 이해 등을 인공적으로 구현한 컴퓨터 시스템이다. 인간과 생물이 갖고 있는 지능 자연지능 natural intelligence과 다르며 기계 등에 인공적으로 시연(구현)한 것이다. 1951년 민스키의 인공지능 개념이 등장하고 50년에는 튜링 테스트가 제시되고 반도체 기술이 폭증하면서 인공 지능 기계가 자연지능을 능가할 것이라는 전망이 나왔다[@커즈와일 특이점 1999] 2045년에 역전되고 인류는 정복당하고 세로운 기계 인류가 등장하리라는 전망이다

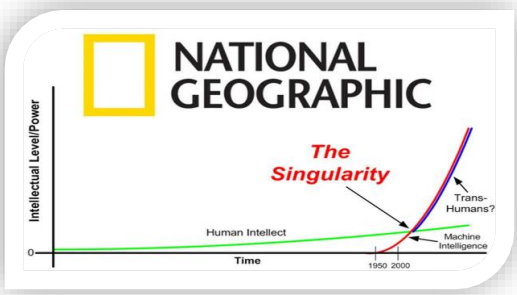


그림 3. 특이점 개념
Figure 3. singularity concept

이를 방지하는 방법으로 인공도덕 장치[AMA]가 제시되고 있지만 지역마다 각가 다른 도덕율 때문에 실효성이 의문시되고 있다 이 논문에서는 인간과 로봇이 공존하는 방안에 대하여 다룬 것이다

인물성 동론과 슬깨비의 인격체

18세기 조선 성리학자들은 인물성 동이론에 대하여 100년간 논변을 벌였다 금수와 이적[夷狄-청나라]도 깨달으면 사람으로 볼 수 있겠는가 하는 화두였다 동론이 우세하여 송시열의 북벌 정책을 청산하고 만청 제국에 대하여 교린에서 사대정책으로 전환되었다 정조 때는 북학파가 형성되어 19세기에는 개화파로 연결되었으며 호론의 이론의 자세는 후일 위정척사로 이어졌다 똑같은 시각이 로봇에서도 적용될 수 있다 로봇을 무자비하고 도덕성이 결여된 기계로 볼 것인가 아니면 인간에 준하는 인격체로 볼 것인가? 하는 점이다 특이점을 주장하는 사람들은 대학을 설립하면서 슬깨비에 대해 경계를 가지고있다 그러나 슬깨비를 인간과 대응한 관계로 대결이 아니라 공유와 동반이나 반려의 시각에서 발전 시켜 나가야 한다고 생각할 수 있다 인간과 슬깨비가 적절한 역할을 분담을 추구해야 하는 이유가 여기에 있다

그림4 인물성 동이론 학맥과 발전계통도
Figure4. Theory and Its Development System

로봇 운영체계[ROS]

쓸깨비 운영체계는 통신 7계층의 맨 윗 단계 응용계층의 운영체계이다 윈도, 니눅스 등 기계어로 기술되지만 본체[platform]의 구조는 j로 다른 별개의 운영체계이다. 기저[인프라]층에서 입력은 문자정보가 주력이지만 개인 슬기깨비에서는 문자보다 음성과 넋골 소리말[뇌언어]이 보다 중요하며 음성언어에서는 정음 한글이 가장 유력하다 음절간 분절이 명확하고 양자역학적 자질 문자 특징을 가지고 있기 때문이다[그림2] 자질문자란 파동적 발화신호와 입자성 본 뜻이 서로 일치하는 알고리즘[@05, 09]이다.

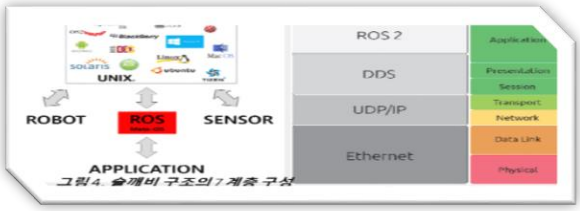


그림 5. Ros 계층 준위도
Figure 5. Ros Layer Levels

이를 뇌파 언어와 발성언어에서 동시에 입력 가능하다 뇌파 언어와 발성언어를 변환장치를 이용하면 발성할 수 없는 인간[농아자]과 슬깨비 사이에 뇌파로 서로 소통할 수 있다 귓속 장착 뇌파 언어를 별도로 수신하는 장치를 사용하면 슬깨비와 인간이 언어 소통장치로 활용할 수 있다

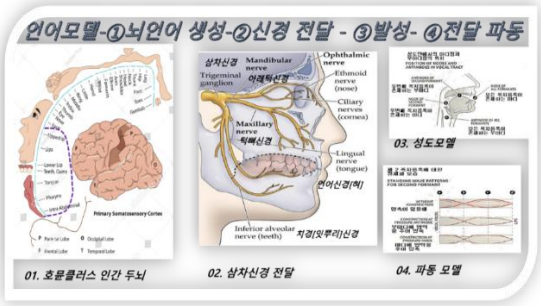


그림 6 – 뇌파언어와 발성언어의 변환과정
Figure 6 – Transformation of brain wave language and vocal language

인성지능과 언어정보의 상호관계
인간의 지능은 9종이 있지만 그 중에서 언어지능이 56%이상 차지하며 이해와 기억 추론은 물론 상호 교신에 의한 정보의 교류와 후성 유전에도 결정적으로 작용하다 최근 발성언어와 뇌파언어가 상호 교신한다 사실이 밝혀졌다
PC와 PT 그리고 PR은 기능과 감각기관이 서로 달라 기능이 분산되어 있다 그러나 이를 통합하고 정음 한글[@00] OS를 공유하면서 이들 3개의 기기는 슬깨비에 의하여 통합 제어된다 [그림2] 검진이나 건강 돌 보미가 첨가되며 개인 대표 기지국 역할도 수행한다

3종슬기 융복합 시스템
슬깨비는 슬기 틀과 슬기 전화의 단순 물리적 통합을 넘어 중복 기능을 재 조정하고 재 조합하
새로운 기능을 정의하고 이를 추가하는 것이다 예를 들면 나노 IoE 소자를 교직 의류[스마클로]에 실장하여 착의형 기기[wearing computer]를 지향할 수 있다 나노 마이크로 로봇을 장내에 투입하여 질병을 예방하고 치료 및 건강 유지의 보조 기구로 활용한다 또 양말에다 레이다 기능을 장착해 자율 운전 보조 기구로 활용하면서 운전자와 함께 운전할 때는 졸음 방지 등에 활용할 수 있으며 보행시의 장애물을 탐지하여 충돌을 사전에 피할 수 있다
행위 관찰용 AR[증강현실] 콘텍트 렌즈, 사방투영입체[홀로그램] 표시창 무잡음 골도 음성입력장치 등은 모두 3종 슬기 융합기구에서 공유할 수 있다

그림7 3종 슬기도구의 융복합 구도
Figure 7 Convergence structure of three types of intelligence device

30.맺음 [결론]
개인용 슬깨비는 사무실이나 가정보다는 이동 정보생활에 보다 더 유용할 것으로 판단된다 이전의 산업혁명과 4차 산업혁명의 핵심적 차이는 인공지능(AI) 활용여부에 있다 유럽의회는 20127년 1월 전자인간(Electronic

Personhood)에 법인격을 부여하고 슬깨빈의 지위, 개발, 활용에 대한 기술적, 윤리적 가이드라인을 제시한 바 있다. "기계의 자유가 커질수록 도덕적 기준이 더욱 필요하다는 취치로 AMA[@11김동욱] 등이 강력하게 요구되지만 지역에 따른 윤리가치는 서로 달라[@12-MIT 2018] 기준 설정이 힘들다

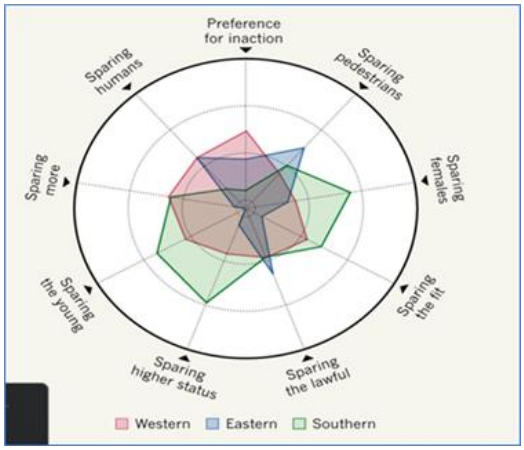


그림 8, 지역바다 다른 윤리 기준

Figure 8, different ethical standards in rezone

그러므로 동반 쓸깨비를 정의하고 개인이나 단체와 기관에서 협업체계를 구축하여 공동으로 대응하는 체계를 구축하는 것이 필요하다

다음은 적용할 내용물 개발이다 한국은 동방예의지국의 전통이 있다 홍익인간[전통유학]주의와 대동주의[공자유학]가 아직 도 명백히 살아있다 이는 세계인이 공유할 수 있는 윤리적 가치를 제공할 수 있으며 인류가 공유해야 할 공통 자산이다 윤리와 도덕 학습 내용을 디지로그 형태로 재개발하고 개발하고[@신윤식 스마트논어 2020+온사람136] 동반 쓸깨비에 장착하여 세계로 도전할 준비가 필요하다

40. 참고문헌[용어해설]

01. 튜링 – 1950년 사람과 기계가 분간할 수 없으면 지능을 가진다

02. 이경구- 조선 철학의 왕국 -호락논쟁 이야기- 2018- 033)248-2900~2902

03. 민스크-1951년 첫번째 신경 네트워크 기계 SNARC 발표[최초인공지능알고림즘] .

04. 카즈와일- 1950년 튜링 기계가 나오고 2045년에는 특이점 도달

05. 정음한글 : ICCKL '01 ISO 유니코드 콘소시엄 특별 권고안- 북정음 조선글.

06.; 뇌파 언어 변환 교신; 스페인 바로셀로나 대학 2017

07. KAIST; 반도체 사방 투영 입체[홀로그램] 표시창 2018

08. 호뮨쿨루스; 대뇌 컽질에서 신체기관 담당 비율에 따라 그려 놓은 인형 모형

09. 샘슨(Geoffrey Sampson) ; 자질문자; 영국 서섹스 대학의 세계 문자시스템 1985

10. 이 옥근; 스마클로 협회 2017

11. 김종욱 로봇도 윤리(AI 로봇의 인공윤리 개발 연구) 교수(동아대 kjwook at dau.ac.kr 2020

12. MORAL ROBOT [도덕보봇]에 대한 MIT 미디어 랩의 연구 2018 -지역마다 다름

13. 신윤식; 스마트 논어 2020 [

14. 황 병수; 온사람 136

저자

술봉 진 용옥[Chin-YongOk-陳庸玉. 麗陽人]

경희 대학교 명예교수 [양자전파공학], 공학박사, 정보통신기술사,

심곡서원 장의(斯文後學)[사]

한미컨 학술연합공동의장,[사]정암 학회장

주 관심 : ①양자전파통신 ②미디아 발달사 ③언어정보학 ④양자악학

+82 010 8923-3402,

(p3soolbong@naver,com) 경기도 용인시 성복2로 86- LG 117동1701호

딥러닝 기반 음료 인식 및 로봇 제어 시스템

최인훈, 김명현, 김영인, 허의남*
경희대학교

{inhun321, freckie, rladuddls3, *johnhuh}@khu.ac.kr

Robot Control System with Deep Learning based Beverage Image Recognition

In-Hun Choi, Myung-Hyun Kim,Yeong-In Kim, Eui-Nam Huh*
Kyung Hee University

요 약

본 논문은 딥러닝 기반의 음료 인식 모듈을 통해 로봇과 객체와의 상대적인 위치를 조정하여 로봇이 원하는 음료를 집을 수 있는 로봇 제어 시스템을 제안하였다. 제안한 시스템은 딥러닝 인식 결과에 따라 매카넘 휠을 이용하여 객체와 로봇간의 위치를 조정하여 중심을 맞춘 후 초음파 센서 값에 따라 로봇 팔을 제어하여 객체를 집도록 구현하였다. 이를 통해 로봇의 앞에 놓여진 여러 객체들 사이에서 원하는 객체를 로봇이 피킹할 수 있도록 하여 무인 카페 시스템 등 다양한 분야에 활용 가능한 시스템을 제안하였다.

1. 서 론

산업용 로봇과 자율주행 기술이 발전함에 따라 최근 자율주행 배달 로봇이 활발히 개발, 도입되고 있다. 자율주행 배달 로봇은 주문자가 있는 곳까지 스스로 이동하여 주문한 물건을 전달한다. 하지만 로봇에게 배달하고자 하는 물건을 실어주는 것은 여전히 사람의 몫이다. 따라서 본 논문은 딥러닝 기반의 객체 인식 모듈과 로봇 팔을 사용하여 편의점과 같이 수 많은 상품이 존재하는 매장에서 로봇이 스스로 물건을 찾아 싣는 시스템을 개발하고자 했다.

2. 본론

2-1. 시스템 구조

Fig. 1 은 시스템의 전체 구성도이다. 시스템은 주문을 접수받는 웹서버, 객체 인식을 위한 카메라와 라즈베리파이, 로봇의 바퀴를 제어하기 위한 stm32f4disc 보드와 모터드라이버(l298n), 6 축 로봇 팔을 제어하기 위한 stm32f4disc 보드로 이루어져있다. 라즈베리파이는 웹서버에 폴링 방식을 통해 1 초에 한 번씩 정보를 요청하며, 라즈베리파이와 2 개의 stm 보드는 시리얼 통신을 통해 데이터를 주고받도록 하였다.

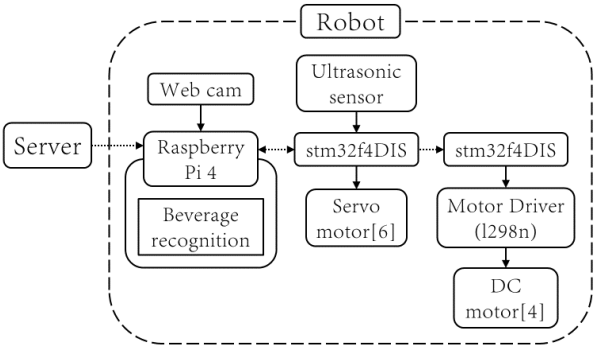


Fig. 1 시스템 구조도

2-2. 시스템 시나리오

Fig. 2 는 시스템의 전체 흐름도(workflow)이다. 사용자가 웹을 통해 주문을 완료하면 로봇은 좌우 이동을 하며 음료를 찾기 시작한다. 객체 인식 모델이 주문 받은 음료를 발견하면 음료와 로봇의 상대적인 위치를 고려하여 음료와 로봇의 중심이 일치할 때까지 객체 인식 결과에 따라 위치를 미세 조정한다.

음료와 로봇의 중심이 완전히 일치하면 로봇의 6 축 팔을 제어하여 음료를 피킹하기 시작한다. 물체를 잡기에 최적의 자세에서 점점 팔을 앞으로 뻗어 집게 사이에 물체가 들어오도록 한다. 로봇팔에 장착된 초음파 센서 값에 따라 집게 사이에 물체가 들어왔는지 판별하여 집게 사이에 물체가 완전히 들어오면 물체를 집어 카트에 싣는다.

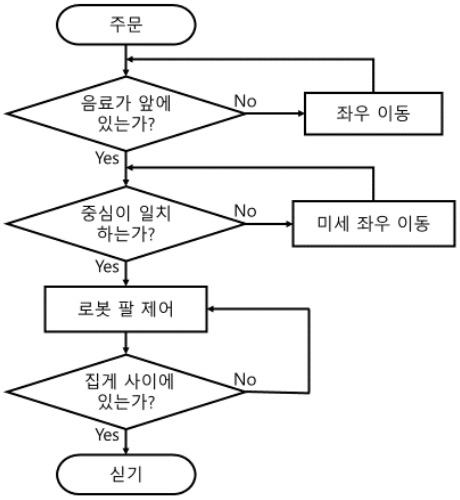


Fig. 2 시스템 흐름도

2-3. 객체 인식

객체 인식 모델은 객체의 종류를 판단하는 classification 과 객체의 위치를 판단하는 localization 으로 이루어 진다. 이는 두 stage 를 동시에 수행하는 1-stage detector 와 순차적으로 수행하는 2-stage detector 로 나눌 수 있다.

2-stage detector 모델들은 입력 이미지에서 객체가 있을만하다고 판단되는 수많은 후보 영역들을 만들고 각각의 영역들을 convolutional neural network(CNN)의 입력으로 사용한다. 이런 2-stage detector 모델들은 비교적 정확도가 높지만 연산량이 매우 많아서 느리다는 단점을 가지고 있다. 실내 자율주행 로봇의 특성상 한정된 자원에서 실시간으로 구동되어야 하므로 한번의 CNN 모델로 객체의 종류와 위치를 판단하는 1-stage detector 모델중 yolo-v3 모델을 경량화한 yolo-v3 tiny 모델을 선택하였다.

모델 훈련을 위해 사용한 데이터는 직접 찍은 1120 장의 이미지를 사용하였다. 5 종류의 음료를 모두 같은 비율로 위치, 각도등을 다르게 하면서 이미지를 찍었다. 이후 각 음료의 경계 박스(bounding box)좌표와 음료의 종류(class)를 포함시킨후 밝기를 5 단계로 조정하는 data augmentation 을 진행한 후 사용하였다.

Fig. 3 음료의 종류와 위치를 인식한 모습

2-4. 로봇 제어

본 논문의 로봇은 4 륜 구동 메카넘휠을 장착하여 전방향으로 이동이 가능하다. 메카넘휠의 특성에 따라 전방향 이동이 가능하려면 이동하고자 하는 방향에 따라 모터의 회전방향을 달리 해주어야 하는데 이를 위해서 l298n 모터드라이버를 사용하여 모터의 회전방향을 설정해주었다. 로봇의 휠 회전 속도와 로봇 팔의 자세는 pulse width modulation(PWM) 방식으로 제어했다.

로봇이 딥러닝 기반 객체 인식 결과에 따라 음료와 로봇의 중심을 완전히 일치시킨 후 로봇 팔을 제어하여 음료를 피킹할 때 사용되는 초음파 센서 값은 실시간 제어에 유리하도록 인터럽트 방식으로 센서값을 수신하였다. 로봇 제어 알고리즘은 c 언어를 활용하여 trueSTUDIO 개발환경에서 구현하였다.

3. 결론

본 논문에서는 음료 인식 모듈을 통해 로봇이 원하는 음료를 집을 수 있는 로봇 제어 시스템을 제안하였다. 이러한 시스템에는 객체 인식 모델로 yolo-v3 tiny 모델을 도입하여 객체를 0.5 초 이하의 시간으로 음료의 위치와 종류를 판단하도록 설계하였다. 이러한 로봇이 자동으로 객체의 위치와 종류를 판단하고 피킹하는 시스템을 통해 무인 배달 서비스를 설계한다면 배달 서비스 부분에 큰 변화를 가져올 수 있을 것으로 기대된다.

ACKNOWLEDGMENT

이 논문은 2020 년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No. 2017-0-00294, (대학 ICT 기초연구실) 서비스 이동 지원을 위한 분산형 클라우드 핵심원천기술 연구)

참 고 문 헌

[1] Joseph Redmon, Ali Farhadi. “YOLOv3: An Incremental Improvement”

(https://pjreddie.com/yolo)

[2] Ross Girshick Jeff Donahue Trevor Darrell Jitendra Malik. “Rich feature hierarchies for accurate object detection and semantic segmentation Tech report (v5)” 2014

딥러닝 기반 객체 인식을 통한 실내 위치 정보 탐색 자율 주행 로봇 개발

김영인, 최인훈, 김명현, 김승직, 허의남*

경희대학교

{rladuddls3, inhun321, freckie, ksj961323, *johnhuh}@khu.ac.kr

An Implementation of Self-driving Robot to Search for Indoor Location Information through Object Recognition based on Deep Learning

Yeong-In Kim, In-Hun Choi, Myung-Hyun Kim, Seung-Jik Kim, Eui-Nam Huh*
Kyung Hee University

요 약

본 논문은 실내 환경에서 딥러닝 기반의 객체 인식 결과를 기반으로 출발지점에서 목표지점까지 자율 주행하는 로봇 제어 시스템을 제안하였다. 제안한 시스템은 360 도 라이다 센서를 통한 실시간 주변 환경 인식 결과를 이용하여 장애물 회피 및 중앙 주행을 수행하고 딥러닝 기반의 객체 인식을 통해 획득한 실시간 위치 정보를 이용하여 목표지점 도착여부를 파악한다. 이를 통해 로봇은 실시간으로 변하는 주변 환경의 제약없이 스스로 목표지점까지 도달 가능하며, 실내 무인 배달 서비스 등 다양한 분야에 활용이 가능할 것이다.

1. 서 론

본 논문은 실내 환경에서 자율 주행 로봇의 위치 정보 탐색에 관한 내용으로, 딥러닝 기반 객체 인식을 통해 로봇의 현재 위치 정보를 추정하는 시스템을 제안하고 있다. GPS 등과 같은 위성항법 장치의 도움을 받기 힘든 실내 환경에서는 로봇의 정확한 위치 추적이 매우 어려운 상황이다. 또한 건물 복도와 같은 실내 환경은 길이 매우 좁으며 보행자가 수시로 나타나는 등 변화가 매우 많은 환경이므로, 관성항법을 통한 위치 추정 또한 로봇의 제어에 어려움이 생긴다.

따라서 본 논문은 360 도 라이다 센서와 딥 러닝 기반 객체 인식을 통해 주변 환경을 파악하며 스스로 목표 지점에 도달하는 자율 주행 로봇 제어 시스템을 제안하고자 한다.

2. 본론

2-1. 시스템 구조

Fig. 1 은 시스템의 전체 구성도이다. 시스템은 목적지를 입력으로 받는 웹서버, 객체 인식을 위한 카메라와 라즈베리파이, 로봇의 바퀴를 제어하기 위한 stm32f4discovery 보드와 모터드라이버(l298n), 실시간 주변 환경을 인식하기 위한 360 도 라이다 센서로 이루어져있다. 라이다 센서와 stm 보드, stm 보드와 라즈베리파이 간에는 시리얼 기반의 UART 통신을 통해 데이터를 주고 받는다.

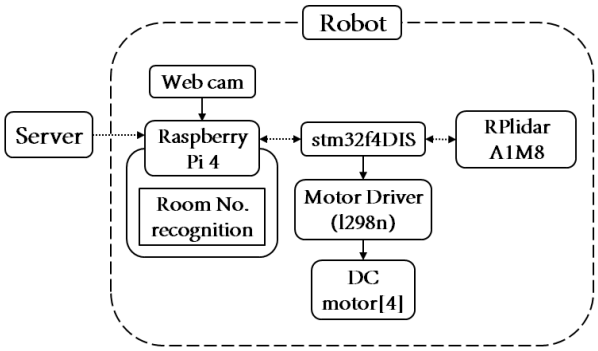


Fig. 1 시스템 구조도

2-2. 로봇 주행

본 논문의 로봇 휠은 4 륜 구동 메카넘휠로써 전방향으로 이동이 가능하다. 4 개의 휠에서 발생하는 벡터의 합의 방향이 로봇이 향하는 방향이 되며 각 휠의 벡터 방향은 모터의 회전 방향에 따라 결정된다. 모터의 회전 방향은 정방향 혹은 역방향으로 설정이 가능한데 이는 모터 드라이버를 사용하여 설정해주었다. 모터의 회전 속도는 pulse width modulation(PWM) 방식으로 제어하였다. 기본 주행 속도는 범위가 0 에서 9999 사이인 pulse width 가 4000 일 때로 설정하였다.

또한 PID 제어기를 사용하여 로봇이 효율적으로 목표속도에 도달하도록 하였다. PID 제어는 피드백 제어기의 형태로 목표치와 실제 측정치를 비교하여 오차를 계산하고 이를 이용하여 제어값을 계산하는 구조이다.

모터의 실제 회전 속도는 엔코더를 사용하여 측정한다. 본 논문에서는 각 PWM 값에 따른 평균적인 엔코더 출력 값을 표로 만들어 전달함수 형태로 관계식을 구하였다.

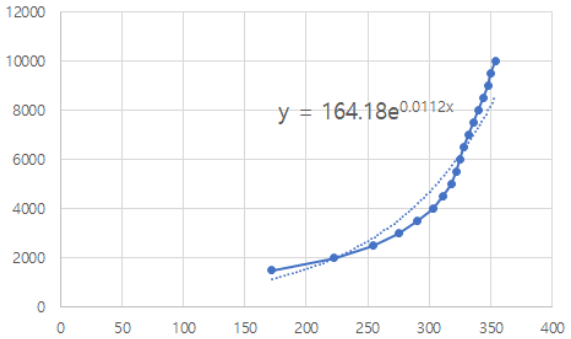


Fig. 2 엔코더 출력 변환 함수

2-3. 실시간 주변 환경 인식

본 논문은 360 도 라이다 센서를 사용하여 자율주행 알고리즘을 구현하였다. 로봇 전방의 180 도에 대한 거리 데이터만 주행에 사용하였으며 180 도를 좌우로 나누어 거리의 평균을 계산하고 좌우 평균의 차이를 PWM 제어값에 활용하였다. 거리 측정에 실패한 경우 좌우 평균 값 계산에 영향을 미치지 않도록 하기 위해 측정에 성공한 횟수를 세어 평균 값 계산에 사용하였다. 이에 따라 로봇이 수시로 나타나는 장애물을 빠르게 우회하고 좁은 건물 복도를 안정적으로 중앙주행하도록 하였다.

2-4. 딥러닝 기반 위치 탐색

본 연구에서는 실내 위치 정보 획득을 위해 딥 러닝 기반 문패(Door Plate) 인식 모델을 구현하였다. 문패 인식 모델은 Fig. 1 의 Web cam 의 영상을 입력으로 하여 실시간으로 분석, 문패 이미지를 추출하고 포함된 숫자를 인식해 출력하는 모델이다.
입력 프레임에서 문패를 인식하는 과정은 OpenCV 라이브러리를 통한 Canny Edge 검출, Contour 분석 등을 거치고, 디자인한 CNN 모델을 통과하게 된다. 이렇게 인식된 문패 이미지는 YOLO 모델 기반의 숫자 추출 모델에 전달된다.
문패 이미지 판별 모델은 직접 디자인한 CNN 모델로, 48*48 Grayscale 이미지를 입력으로 하여 0, 1 의 결과를 도출하는 이진 분류 (Binary Classification)을 수행한다. Table 1.은 문패 이미지 판별 모델의 아키텍처를 표현한 테이블이다.

TABLE 1. 문패 이미지 판별 모델 아키텍처

CONV. 2D	filter=32	(5, 5)
CONV. 2D	filter=32	(5, 5)
DROPOUT	0.5	
MAXPOOL. 2D	stride=2	(2, 2)
CONV. 2D	filter=64	(5, 5)
CONV. 2D	filter=64	(5, 5)
DROPOUT	0.5	
MAXPOOL. 2D	stride=2	(2, 2)
FULLY CONNECTED	1024	
FULLY CONNECTED	2	

문패 이미지 판별 모델은 SVHN Validation Set 의 경우 99.8%의 정확도를 보여주었으며, 실제 주행에서 수집한 데이터의 경우 93.3%의 정확도를 보여주었다.
이렇게 인식된 문패 이미지는 YOLO 모델 기반의 숫자 추출 모델에 전달된다.
숫자 추출 모델은 YOLOv5s[1] 기반으로 SVHN 데이터셋 [2]을 학습하였다. 기존 YOLO 모델은 80 종류의 실생활 객체를 학습했으나 본 논문의 숫자 추출 모델은 10 종류의 숫자 객체를 인식 대상으로 설정하였다.

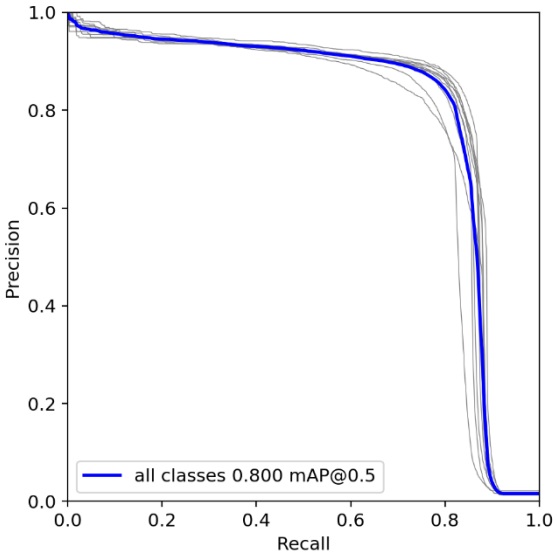


Fig. 3 숫자 추출 모델의 Precision-Recall Curve

Fig. 3 은 숫자 추출 모델의 성능을 평가하기 위해 Precision 과 Recall 을 나타낸 그래프이다. IoU(Intersection over Union)가 0.5 인 경우 0.8 의 mAP(mean Average Precision) 값을 보여주었다.

3. 결론

본 논문에서는 방 호수 인식 모듈을 통해 로봇이 사용자가 원하는 곳까지 자율주행하는 로봇 제어 시스템을 제안하였다. 로봇은 라이다 센서를 통해 실시간 주변 환경 변화에 신속하게 반응하며 주행할 수 있었고, PID 제어기를 사용하여 보다 안정적인 주행을 할 수 있었다. 문패 인식 모델은 약 93.3%의 정확도를 보였고, 숫자 추출 모델은 mAP@0.5 의 값이 80%로 두 모델 모두 높은 성능을 보였다.
본 논문의 시스템은 향후, 로봇에 서스펜션을 장착해 충격을 완화하여 영상의 흔들림을 감소시키고, 사람이 붐비는 환경에서는 일시적으로 저속으로 주행하는 저속 주행 모드를 추가하는 등 기능을 향상시킬 예정이다. 또한 문패 이미지 판별 모델의 정확도를 높여 노이즈를 줄이고, 이를 통해 숫자 추출 모델의 호출 횟수를 줄여 전체적인 시스템의 효율을 증대시킬 계획이다.

ACKNOWLEDGMENT

이 논문은 2020 년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No. 2017-0-00294, (대학 ICT 기초연구실) 서비스 이동 지원을 위한 분산형 클라우드 핵심원천기술 연구)

참 고 문 헌

[1] Joseph Redmon, ed., “You Only Look Once: Unified, Real-Time Object Detection”, p. 6, 2015.

[2] http://ufldl.stanford.edu/housenumbers (format 2)

심층 신경망 기반 추적기를 사용한 사용자 추종 로봇

손찬영, 이혜민, 이준구, 오지용
한국전자통신연구원 대경권연구센터
{cysohn, leehaemin90, leejg01679, jiyongoh}@etri.re.kr

A User Following Robot Using Deep Neural Network Based Tracker

Chanyoung Sohn, Hea-Min Lee, Joon-Goo Lee, and Jiyong Oh
Daegu-Gyeongbuk Research Center
Electronics and Telecommunications Research Institute

요 약

본 논문은 이동 로봇이 사용자를 따라다니며 다양한 협업이 가능한 사용자 추종 기능을 다룬다. 본 연구를 통해 개발된 사용자 추종 기술은 RGBD 카메라 신호로부터 사용자의 위치를 실시간으로 계산하는 추적 모듈과 로봇이 사용자를 적절하게 추종하기 위해 사용자의 위치로부터 주행 명령을 계산하는 주행 모듈로 구성된다. 정적 환경에서 반복 수행된 사용자 추종 실험은 본 논문의 사용자 추종 기술이 효과적이다는 사실을 보여준다.

Ⅰ. 서론

2016년 3월 알파고의 등장 이후 인공지능 기술은 4차 산업혁명의 핵심기술로 자리 잡아 현재 다양한 분야에서 활용되고 있으며 로봇 분야에서도 인공지능 기술이 적극적으로 활용되고 있다. 로봇은 전통적으로 인간의 노동력을 대체하거나 노동환경을 개선하기 위해 개발되었다. 특히 농업, 수산업을 비롯한 1차 산업에서부터 물류창고 및 배송, 제조공장, 병원 등 다양한 산업에서 요구되는 중량물 운반은 로봇이 담당하기에 적합한 작업 중 하나이다. 그러나 농작물 수확, 택배 배송 등은 중량물을 단순히 목적지까지 이송하는 것 이외에 적재물을 운반하는 과정에서 사용자의 작업이 빈번하게 요구된다. 사용자 추종[1]은 이러한 요구에 맞추어 개발되고 있는 대표적인 인간-로봇 협업 기술이다. 본 연구에서는 인공지능 기반의 추적 기술을 활용하여 사용자 추종 로봇을 개발한다. 특히 개발된 사용자 추종 로봇은 VOT-RGBD2019[2] Challenge에서 우승한 SiamDW-D 구조[3]를 활용하여 사용자가 카메라의 FOV 밖으로 사라지거나 다른 사람에 의해 가려진 뒤에 다시 나타나더라도 추적 중이던 사용자를 강인하게 추종할 수 있다.

Ⅱ. 본론

1. 로봇 및 프로그램의 구조

그림 1. 로봇 플랫폼 (좌), 사용자 추종 프로그램 흐름도 (우)

그림 1은 본 연구에서 사용된 2개의 구동 휠과 4개의 캐스터를 갖는 차동 휠 (differential wheel) 타입의 로봇 플랫폼과 사용자 추종을 위한 프로그램의 흐름을 보여주고 있다. ROS 기반의 사용자 추종을 위한 프로그램은 추적 모듈과 주행 모듈로 구성된다. 추적 모듈은 지면으로부터 120cm 높이에 장착된 Intel Realsense D435 카메라로부터 RGBD 데이터를 입력받아 20 fps이상의 속도로 사용자의 상대 위치를 주행 모듈로 전달한다. 주행 모듈은 현재 사용자의 상대 위치로부터 사용자를 추종하기 위한 로봇의 선속도와 각속도를 계산하여 모터 드라이버로 전달한다.

2. 추적 모듈

2.1. 심층 신경망 기반 추적 모듈

추적기는 target(추적 대상)과 search(검색될 전체) 이미지를 각각의 신경망에 통과시키는 Siamese Tracker 구조를 갖는다. 특히 추적 속도가 빠른 RPN (region proposal network) 구조[4]를 지니며 추적 정확도 및 강인성을 향상시키기 위해 설계된 CIResNet[3]을 backbone으로 사용한다. 신경망의 출력은 추적 대상의 영상 내 2차원 중심점과 경계 상자의 너비와 높이, 신뢰도 값이다.

2.2. 추적 대상의 3차원 좌표 추정

로봇이 추종 대상을 따라다니도록 만들기 위해서는 실제 세계의 3차원 좌표가 필요하다. 이때 좌표계는 카메라 좌표계를 사용하며 로봇에 부착된 RGBD 센서의 중심을 원점으로 하는 좌표계이다. 카메라 좌표계에서의 추적 대상 좌표를 구하기 위해 먼저 추적 대상의 깊이 정보를 먼저 획득해야 한다. 일반적으로 깊이 정보는 depth 이미지의 화상 강도로 표현되며, 객체 영역의 평균 화소 값을 이용한다. 하지만 객체 영역은 수시로 변하는 배경을 포함하며, depth 센서의 특징으로 인한 결측이 발생한다. 이를 해결하기 위해 다음과 같이 경계 상자의 너비와 높이의 일정 비율만큼을 포함하는 중심 상자의 좌표를 구한다.

$$l = m - w \times r/2, \quad r = m + w \times r/2 + 1,$$
$$t = n - h \times r/2, \quad b = n + h \times r/2 + 1$$

여기서 l, r, t, b는 각각 중심 상자의 좌, 우, 상, 하의 좌표이며, (m, n)은

중심 상자의 중심점 좌표이다. 또한 r은 경계 상자와 중심 상자의 비율을 의미하는데 본 연구에서는 0.1로 설정하였다(그림 2 참조). 이 중심 상자 내에서도 깊이 정보 결측치가 발생하게 되므로 해당 값들을 최대한 제거해야 한다. 중심 상자를 기준으로 평균과 표준 편찻값을 구하고, 중심 상자 내 임의의 값의 편차가 표준 편찻값에 비례한 기준값보다 작은 값들에 대해서만 평균을 구한다. 이 값을 추적대상의 깊이 값으로 사용한다.

다음으로 영상 좌표계의 경계 상자의 중심 좌표 (m,n)과 추적 대상의 깊이 값 d를 이용하여 추적대상의 카메라 좌표계에서의 좌표 (x,y,z)를 구한다. 이를 위해 센서 카메라의 내부 파라미터 행렬을 사용한다.

$$x = d \times (m - p_x)/f_x, \quad y = d \times (n - p_y)/f_y,$$

$$z = \sqrt{x^2 + y^2 + d^2},$$

여기서 f는 초점 거리이며, p_x는 주점의 x좌표, p_y는 주점의 y좌표이다.

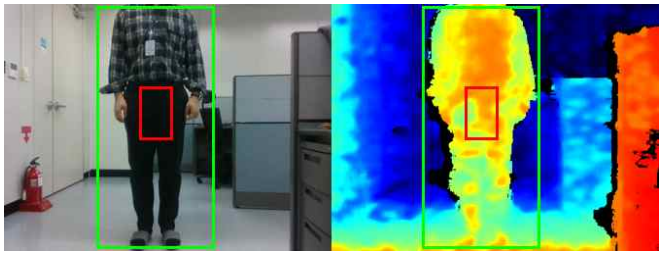


그림 2. 경계 상자와 중심 상자를 표시한 RGB 이미지(좌)와 색상화된 깊이 정보(우)

2.3. 제어기 입력 좌표계 변환

로봇의 모터를 구동하기 위한 제어기는 로봇과 추종 대상과의 상대 거리를 사용한다. 이때 θ는 카메라 방향에 수직인 영상 중심점을 지나는 평면으로부터 카메라 방향과 추종 대상 간의 각도를 의미한다.

$$\phi = \sin^{-1}\left(\frac{x}{z}\right), \quad x_r = z\cos(\phi), \quad y_r = z\sin(-\phi)$$

앞서 구한 x_r, y_r, z, ϕ를 제어기에 전달하게 되는데, 낮은 신뢰도를 갖는 추적 결과에 대한 로봇의 오작동을 방지하기 위해 신경망이 추론한 신뢰도 값이 일정 값 이하일 때 제어기에 해당 값을 전달하지 않는다.

3. 주행 모듈

구현한 모바일 로봇은 두 가지 센서 정보를 이용해 주행한다. 첫 번째, 영상처리 모듈에서 로봇이 따라가야 할 추종 대상의 상대 위치와 각도 차이, 거리 차이, 신뢰도 값이 계산되어 입력된다. 두 번째, 엔코더에서 글로벌 좌표 기준 실시간 좌표와 각도 정보가 입력된다. 이 두 데이터를 이용한 경로 계획을 통해 로봇이 추종 대상을 따라 주행하게 만든다. 그림 3의 왼쪽과 같은 이륜 구동 로봇을 제어하는 시스템을 구현하기 위해 다음과 같은 모델을 이용한다.

$$\rho = \sqrt{(x_r - x_i)^2 + (y_r - y_i)^2}, \quad \theta = \phi - \tan^{-1}\left(\frac{y_r - y_i}{x_r - x_i}\right)$$

$$v = K_1\rho\cos\theta, \quad w = -K_1\sin\theta\cos\theta - K_2\theta,$$

여기에서 (x_i, y_i)와 (x_r, y_r)은 각각 로봇의 현재 위치와 추종 대상의 위치 정보이고, ρ는 추종 대상과의 거리, θ는 추종 대상과의 각도 차이다. 또한 K_1과 K_2는 주행 제어를 위한 제어기의 이득에 해당한다. 주행이 시작되면 로봇은 이미지 센서로부터 도출된 정보로 추종 대상의 위치 정보를 만든다. 추적 모듈로부터 목표 지점의 좌표를 전달받으면 그림 3의 왼쪽과 같은 모바일 로봇의 기구학 모델을 기반으로 로봇의 경로를 계획한다. 위 식을 바탕으로 로봇에 선속도 값(v)과 각속도 값(w)을 제어 입력으로 주어 로봇을 이동시킨다.

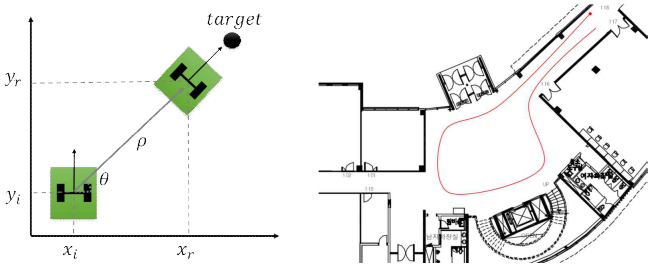


그림 3. 이륜 구동 모바일 로봇(좌), 사용자 추종 실험 환경 및 시험 경로(우)

4. 실험

개발된 사용자 추종 로봇의 성능을 검증하기 위해 그림 3의 오른쪽과 같이 비교적 넓고 지나가는 사람들이 많지 않은 실내 공간에서 미리 정의된 경로를 따라 움직이며 추종의 성공률을 계산하였다. 사용자 추종은 로봇이 출발지부터 도착지까지 지정된 사용자를 추종하며 이동하는 동안 관찰자의 개입이 없는 경우 성공이라고 간주하였다. 추종 실험에는 다른 사람이 사용자와 로봇 사이를 지나가는 의도적인 방해 활동을 포함하였다. 그럼에도 불구하고 개발된 사용자 추종 로봇은 20번의 반복 실험에서 20번 모두 사용자 추종에 성공하였다. 이러한 사실은 본 연구를 통해 개발된 사용자 추종 기능이 효과적이며 강인하다는 것을 의미한다.

Ⅲ. 결론

본 연구에서는 로봇이 사용자를 따라다니며 사용자와 협업할 수 있게 하는 사용자 추종 기능을 개발하였다. 제안된 사용자 추종 프로그램은 SiamDW 구조의 심층 신경망을 활용하여 RGBD 카메라의 입력을 받아 실시간으로 사용자의 위치를 추적하는 추적 모듈과 현재 사용자의 위치를 입력받아 해당 사용자를 추종하기 위한 로봇의 이동 선속도와 각속도를 계산하는 주행 모듈로 구성된다. 그리고 정적 환경에서 수행한 반복 실험을 통해 개발된 프로그램이 사용자 추종에 효과적이며 다른 사람이 로봇과 사용자 사이를 이동하는 방해에도 강인하다는 것을 입증하였다.

ACKNOWLEDGMENT

본 논문의 연구는 한국전자통신연구원 연구운영지원사업의 일환으로 수행되었음. [20ZD1130, 지능제어기반 스마트 기계 및 로봇 기술 개발].

참 고 문 헌

[1] Islam, M. J., Hong, J., and Sattar, J., "Person-Following by Autonomous Robots: A Categorical Overview," The International Journal of Robotics Research 38, no. 14, pp. 1581-1618, Dec. 2019.

[2] Kristan, M. et al., "The Seventh Visual Object Tracking VOT2019 Challenge Results," In Proceedings of 2019 IEEE/CVF International Conference on Computer Vision Workshop, pp. 2206-2241, 2019.

[3] Zhang, Z. and Peng, H., "Deeper and Wider Siamese Networks for Real-Time Visual Tracking," In Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4591-4600, 2019.

[4] Li, B., Yan, J., Wu, W., Zhu, Z., and Hu, X., "High Performance Visual Tracking with Siamese Region Proposal Network," In Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8971-8980, 2018.

Reinforcement Learning Based Scheduling in Underwater NDN

Muhammad Toaha Raza Khan, Muhammad Saad Malik, Muhammad Ashar Tariq*, Md. Mahmudul Islam, Junho Seo, Ru Yang, Dongkyun Kim
School of Computer Science and Engineering, Kyungpook National University, South Korea
**Department of Artificial Intelligence, Kyungpook National University, South Korea*
{toaha,maliksaad,tariqashar,mislam,junhoseo,yr0818,dongkyun}@knu.ac.kr

***Abstract*—Day by day the demand for monitoring the marine environment and exploring the ocean is increasing. However, the limited bandwidth and long propagation delays and transmission delays make it challenging. In this paper, we propose the Reinforcement Learning (RL) based efficient scheduling for underwater nodes which communicates with the sink node using Named Data Networking (NDN) protocol. The RL based approach is applied at the sink node which provides assistance to the sensor nodes in scheduling for transmitting data. RL based approach assist in avoiding collision of packets that engenders efficient use of an acoustic channel.**

***Index Terms*—Reinforcement Learning (RL), Named Data Network (NDN), Underwater, Surface Sink, Sensor Nodes**

I. Introduction

Internet of underwater things (IoUT) is defined as the network of various smart underwater objects which can communicate to each other [1]. Different types of underwater entities including autonomous underwater vehicles (AUVs), underwater sensor nodes, surface sinks and so on for the network in IoUT. The underwater nodes in IoUT, however, have a limited processing capability and battery. Therefore, these nodes need to communicate using some lighter protocol for minimum resource consumption. Constrained application protocol (CoAP) is a lightweight internet of things (IoT) protocol developed for the constrained IoT devices. Since the underwater nodes are constrained, CoAP is a highly suitable candidate for communication among nodes, AUVs and surface sinks or ships. The CoAP has a minimum header size of 4 bytes and supports reliable communication with a built-in congestion control mechanism. A detailed overview of the CoAP can be found in [2].

The communication of multiple underwater objects like AUVs, nodes, etc. with the surface sink requires scheduling to communicate over same channel. Numerous scheduling schemes are available for communication between surface sink and underwater objects. Two scheduling schemes based on time domain multiple access (TDMA) are discussed in [3] for communication between underwater sensor nodes and surface sink, named as Transmit Delay Allocation MAC (TDA-MAC) and Accelerated TDA-MAC. The TDA-MAC uses ping messages to calculate the propagation delay between each node and surface sink and then sends a transmit delay instruction (TDI) packet to each node, informing the amount of time it has to wait to start transmission after receiving the request (REQ) packet from the sink.The accelerated TDA-MAC caters the channel underutilization issues in TDA-MAC.

The limited bandwidth and slow propagation speed of acoustic signals leads to low data throughput for underwater networks. Machine learning techniques such as RL approach is applied to medium access control that engenders efficient use of an acoustic channel. In [4] RL based mechanism is applied for distributed scheduling in underwater networks to avoid collision. However, if two nodes sense the channel simultaneously before taking action, there is also a probability of collision. If the two nodes share the neighboring information between each other such as for neighbor discovery, the exchange of beacon messages can reduce the probability of collision. However exchange of beacon messages can increase the network overhead. In this paper we proposed the semi-centralized scheduling for transmission of data by the nodes using RL based mechanism.

The remainder of this paper is as follow. Section II gives the overview of ALOHA in underwater networks. Section III discuss about the proposed approach for scheduling in underwater NDN. Finally the conclusion is drawn in Section IV.

II. ALOHA in Underwater Networks

Underwater sensor networks (UWSNs) is quite similar to terrestrial wireless networks, both shares a common channel for message propagation. Having a common channel for transmission and reception leads to collision when multiple devices access the shared medium. To reduce the data packet collision there should be a mechanism which will control the allocation of the common channel to various users. MAC protocols do the job for efficient channel access mechanism. The primary task of a MAC protocol is to avoid collision when allocating channel access to different nodes. Pure ALOHA (P-ALOHA) was the earliest contention-based MAC protocol invented in the 1970s. Slotted ALOHA (S-ALOHA) was proposed to enhance P-ALOHA. S-ALOHA divides the transmission time into multiple slots [5]. But still there is a probability of collision if two nodes sense the channel simultaneously.

III. Reinforcement Learning based Scheduling in Underwater NDN

In underwater networks there is long propagation delays and transmission delays. In underwater NDN based network, the channel for transmission is divided into slotted ALOHA. Each node when have data to send, sense the transmission channel and sends the data on the respective free slot. The channel is divided into limited number of larger time span slots because of long propagation delay. When nodes sense the channel for data transmission there is a probability of collision when both nodes sense the channel simultaneously for transmission. In order to overcome this problem we proposed the centralized scheduling of data transmission for nodes using RL approach. This scheme still allows the nodes to behave in distributed manner rather than fully controlled by the surface sink as in centralized scheduling scheme. Wherein distributed scheduling scheme which is applied at the underwater nodes increase the network overhead, if nodes share the beacon message with the neighboring nodes which is not desired for underwater networks.

Machine learning based RL mechanism is applied at the surface sink. The surface sink gives the opinion regarding the scheduling of data transmission and broadcast the message for other nodes not to transmit at the respective time. The RL based machine learning is applied at the surface sink which is considered as the agent as shown in Fig. 1.

RL based mechanism consists of state and action taken by the agent and in return gets the reward. The surface sink observes the current state and based on optimal policy takes the action. The surface sink gets the message packet from nodes. From message packet it extracts the control information and the data information. The data information comprises of 3 dimensional hierarchical scheme such as \\location\time \type. The receive data message gives information regarding the location of the sensing region, at what time data was sensed and the type of sensed data such as salinity and turbidity. The control information collected from the message packet gives information regarding the received interference strength in channel (I_t), channel gain between sending node and the surface sink (H_t), selected slot indices (N_t) for transmission and (L_t) represents the proportion of bits remaining to transmit. So the state comprises of number of observations which includes as given by equation (1).

$$S_t = [I_t, H_t, N_t, L_t] \tag{1}$$

The agent based on the policy takes the action gives reward to the decision made. The surface sink after receiving the message packet sends the scheduling intervals to the nodes and also sends the acknowledgement to the sender. The agent after taking the action rewards +1 if no collision happens, -1 if collision happens and -0.5 if the sink node receives the duplicate packet. The collision is detected if the sink does not receive the message packet at the given slot assigned to the sending node. The objective of the RL is to find the policy to maximize the expected cumulative reward as given equation (2).

$$R_t = E[\sum_{n=0}^{\infty} \beta^n r_t + n] \tag{2}$$

where $\beta \epsilon$ [0, 1] is the discount factor.

The state transition and reward are stochastic and modelled as a Markov decision process (MDP), where the state transition probabilities and rewards depend only on the state of the environment and the action taken by the agent. The transition from s_t to s_{t+1} with reward r_t when action a_t is taken can be characterized by the conditional transition probability, $p(s_{t+1}, r_t | s_t, a_t)$.

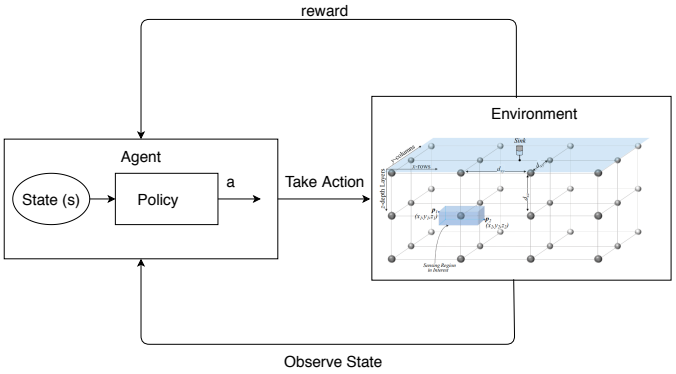


Fig. 1. Reinforcement Learning for Scheduling in Underwater NDN Network

IV. Conclusions

In this paper, we propose the RL based scheduling scheme for the transmission of data. This RL based approach is applied at the surface sink which performs the scheduling on the behalf sensor nodes to avoid the collision of data packets. RL based potentially offers opportunities for underwater NDN network design, due to its adaptive capability and its responsiveness to environmental changes.

V. Acknowledgement

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2016R1D1A3B01015510).

References

[1] C.-C. Kao, Y.-S. Lin, G.-D. Wu, and C.-J. Huang, "A comprehensive study on the internet of underwater things: applications, challenges, and channel models," *Sensors*, vol. 17, no. 7, p. 1477, 2017.

[2] M. A. Tariq, M. Khan, M. T. Raza Khan, and D. Kim, "Enhancements and challenges in coap—a survey," *Sensors*, vol. 20, no. 21, p. 6391, 2020.

[3] N. Morozs, P. Mitchell, and Y. V. Zakharov, "Tda-mac: Tdma without clock synchronization in underwater acoustic networks," *IEEE Access*, vol. 6, pp. 1091–1108, 2017.

[4] S. H. Park, P. D. Mitchell, and D. Grace, "Reinforcement learning based mac protocol (uw-aloha-q) for underwater acoustic sensor networks," *IEEE Access*, vol. 7, pp. 165531–165542, 2019.

[5] K. Chen, M. Ma, E. Cheng, F. Yuan, and W. Su, "A survey on mac protocols for underwater wireless sensor networks," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 3, pp. 1433–1447, 2014.

Wind Speed Interval Forecasting Under Uncertainty Quantification Pattern Based on Deep Learning Method

Himawan Nurcahyanto
Electrical Electronics Engineering
Kookmin University
Seoul, South Korea
himawanurcahyanto@gmail.com

Aji Teguh Prihatno
Electrical Electronics Engineering
Kookmin University
Seoul, South Korea
aji.teguh@gmail.com

Yeong Min Jang
Electrical Electronics Engineering
Kookmin University
Seoul, South Korea
yjang@kookmin.ac.kr

***Abstract*— Global demand for energy is on the rise. The incorporation of renewable energy sources into the grid presents an engineering and economic challenge. Wind and solar power are considered to be the next generation of electricity. However, the wind is often difficult to forecast, as wind speed is typically stochastic and non-stationary. The prediction depends on the site of the users and the feasibility of the forecasting should be observed under technological and regulatory conditions. In this paper, we propose a wind speed interval forecasting under the uncertainty quantification pattern. It is well accepted that wind differs in patterns and weather conditions. Deep learning called Recurrent Neural Network (RNN) algorithm is applied to seeking for the optimal prediction error weights. The implementation considering the two-season situation that carried out by observation of the three-month wind speed pattern. The result shows that the trained and tested model can achieve higher quality forecasting value.**

Keywords— Deep learning, optimal prediction, artificial intelligence, recurrent neural network, wind speed forecasting.

I. INTRODUCTION

Wind speed interval estimation plays an important role in renewable energy generation. Especially in solar power generation and wind power generation. The output of wind energy and solar energy is outstanding among many new energy sources. The International Energy Agency (IEA) forecasts that wind capacity is projected to increase by 60% (325 million KW) by 2019 [1]. As wind speed can change drastically in just a few hours, the issue of wind power generation and solar power generation lies in its reliance on wind fluctuations. In the wind power generation, it further hinders incorporation of getting wind power. In the solar power generation case, the wind speed determines the cloud motion condition that affects the output of the power system. Therefore, prediction of reliable wind speed is an important prerequisite for large-scale production and utilization.

II. MODELLING TECHNIQUES AND STRUCTURES

In several power systems, the stability and sustainability of power generation and the reduction of greenhouse gas emissions are key concerns to consider. Wind power prediction, which is commonly considered to be a highly variable time sequence, plays a key role in overcoming such challenges [2]. Generally, the studies divided into four categories: physical model, conventional statistical model, spatial correlation model and artificial intelligence model [1]. The proposed wind speed forecast methods offer point estimates of future values. In practice, the precision of point estimates can be influenced by variability in model parameters and input data. For practical applications, information on the uncertainty of forecasts is important to properly manage the energy system. [**3].**

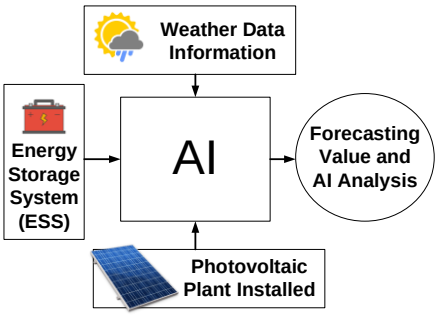


Figure 1 Modelling techniques and structure of wind speed forecasting

Artificial intelligence construction prediction model is proposed to minimize the error estimation of the wind speed condition. Variable wind speed data is used to decompose complex wind speed time series. The Recurrent Neural Network (RNN) model is designed to predict wind speed and the error. The prediction error for this model is given by weight and accumulated to obtain the width of the forecast interval.

A. Data Collection

The condition of wind speed will vary under changes in the weather situation. Voltage and current output power will change because it is influenced by wind speed. In the case of solar power generation, wind speed will determine the cloud size that affects the total solar radiation to the photovoltaic system. Besides, in the wind generation system, it will determine the rotation of the wind turbine system. The inherent variability and uncertainty affecting renewable energy sources which have a major impact on the power supply, and accuracy and reliability forecasts of the power production from renewable energy sources are required at different time scales. For this purpose, forecasting the performance of renewable energy sources is crucial for their efficient incorporation into the grid and for dealing with their uncertain and intermittent existence [3].

B. *Wind Speed Forecasting*

Wind prediction is complicated due to the high degree of uncertainty and variance of the wind. Windspeed is a time series that can be defined as a collection of observations of a parameter or a set of parameters taken at several time intervals. Time scale needs to be established. Wind speed forecasting in this paper is described every one hour ahead of prediction. These intervals are typically of normal duration. If the time step between data points is not compatible or data is incomplete, this should be corrected to a standard time step if the data is to be used for forecasting purposes. Real-world time series are very diverse. Some time-series data shifts slowly and reasonably smoothly. Monthly energy demand can be a sequence of periods like this. Other time series can exhibit relatively chaotic behavior, making them difficult to predict [4].

III. Implementation Result of Deep Learning Model For Wind Speed

Comprehensive studies should be included in a systematic assessment of wind speed forecasting process, as this will improve confidence in the results. We use wind velocity datasets for solar panels located in South Korea. The data used is from January 2020 until April 2020.

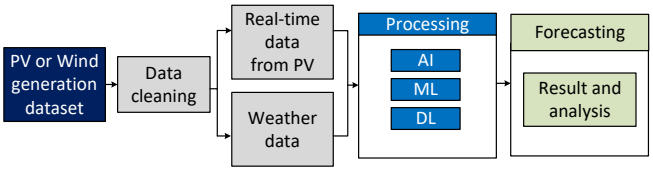


Figure 2. Wind speed prediction architecture.

This paper uses Jupyter software, which gathers the data from the renewable energy company. The algorithm obtains the precision of analytics. We can predict the value of their data in the future using the RNN algorithm by using wind speed data.

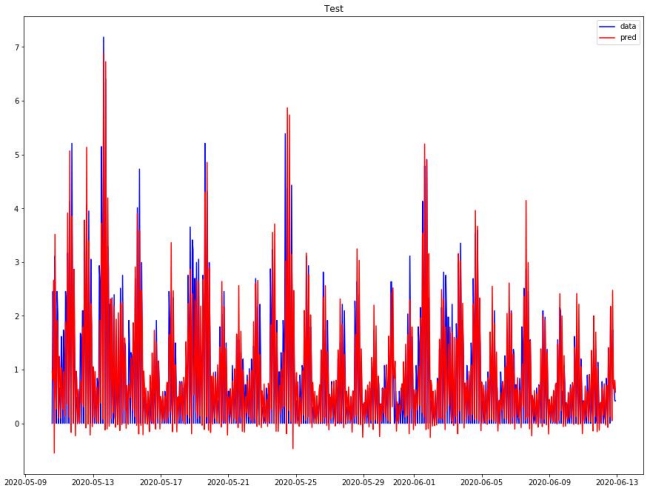


Figure 3. Wind speed predicted

Figure 4. Wind speed forecasted

IV. Conclusion

Implementation of the AI algorithm is presented to provide data stability, reliability, and data pattern-based interoperability. The algorithm will find the pattern of the data by using deep learning and will forecast using RNN in the next stage. This implementation can be used as the Energy Factories approach technique. Hence, in terms of electricity efficiency, we should have a reasonable maintenance plan. To give the outcome of the forecast, more reliably and effectively, structural changes were essential for this implementation. In order to offer the outcome of prediction, more reliably and accurately, sustainable improvements necessary for this study.

Acknowledgement

This research was financially supported by the Ministry of Trade, Industry and Energy (MOTIE) and Korea Institute for Advancement of Technology (KIAT) through the International Cooperative R&D program (Project ID:P0011880)

References

[1] Y. Zhang, S. Gao, J. Han and M. Ban, "Wind Speed Prediction Research Considering Wind Speed Ramp and Residual Distribution," in IEEE Access, vol. 7, pp. 131873-131887, 2019, doi: 10.1109/ACCESS.2019.2940897.

[2] M. Khodayar, J. Wang and M. Manthouri, "Interval Deep Generative Neural Network for Wind Speed Forecasting," in IEEE Transactions on Smart Grid, vol. 10, no. 4, pp. 3974-3989, July 2019, doi: 10.1109/TSG.2018.2847223.

[3] R. Ak, O. Fink and E. Zio, "Two Machine Learning Approaches for Short-Term Wind Speed Time-Series Prediction," in IEEE Transactions on Neural Networks and Learning Systems, vol. 27, no. 8, pp. 1734-1747, Aug. 2016, doi: 10.1109/TNNLS.2015.2418739.

[4] C. W. Potter and M. Negnevitsky, "Very short-term wind forecasting for Tasmanian power generation," in IEEE Transactions on Power Systems, vol. 21, no. 2, pp. 965-972, May 2006, doi: 10.1109/TPWRS.2006.873421.

An Investigation on Feature Extraction and Feature Fusion Methods for Wearable Sensor-Based Human Activity Recognition

Nguyen Thi Hoai Thu, Dong Seog Han
School of Electronics and Electrical Engineering, Kyungpook National University

thunguyen@knu.ac.kr, dshan@knu.ac.kr

Abstract

Feature extraction is usually considered as one of the most essential parts in wearable sensor-based human activity recognition (HAR) and classification tasks in general. In this paper, we carry out an investigation into feature extraction methods of both conventional machine learning and deep learning as well as feature fusion of these two approaches for wearable sensor-based human activity recognition. The hand-crafted features and automatically learning features are combined in order to provide the most useful information for the classification task. The experimental results on a benchmark dataset indicate that using hand-crafted features with deep learning models can give a better performance compare to other feature extraction methods.

Ⅰ. Introduction

Recently, human activity recognition (HAR) has gained great attention as its contribution to the domain of healthcare and human-computer interaction. With the ubiquity of smart wearable devices which contain powerful sensors, human activities, and abnormal behaviors can be automatically detected using the sensor data.

Research on human activity recognition can be generally grouped into two main approaches: conventional machine learning (ML) approach and deep learning (DL) approach. Conventional ML methods have been widely applied to HAR for the last two decades, in which the system contains two main parts: feature extraction and activity classification. Essential features are extracted from the sensor data by using several feature extraction methods in both time domain (e.g., mean, standard deviation) and frequency domain (e.g., Fourier Transform, Wavelet Transform) before being fed into some conventional classification models such as *k*-nearest neighbors (kNN) and support vector machine (SVM) [1]. Although this approach has succeeded in gaining significant achievements, it still has some limitations as the feature extraction often requires domain knowledge.

In the last few years, in the rapid growth of deep learning algorithms and powerful computational resources, several studies have delved into applying DL to human activity recognition [2, 3]. With extraordinary architectures such as convolutional neural networks (CNN) and long short-term memory (LSTM), deep learning has opened a new approach for human activity recognition where the features can be automatically extracted without expert knowledge. In addition, these deep features also help improve the performance of HAR, especially in complex activity recognition. However, there is an opening question which is whether DL automatic feature extraction methods always outperform the conventional methods.

Ⅱ. Method

In order to answer the question, we carry out several experiments on different feature extraction methods: The Wavelet transform, CNN, and LSTM on the public HAPT dataset [4]. The dataset contains data collected from accelerator and gyroscope embedded in a smartphone mounted at the waist of the users. Thirty participants carried out 12 activities: 6 basic activities and 6 postural transitions.

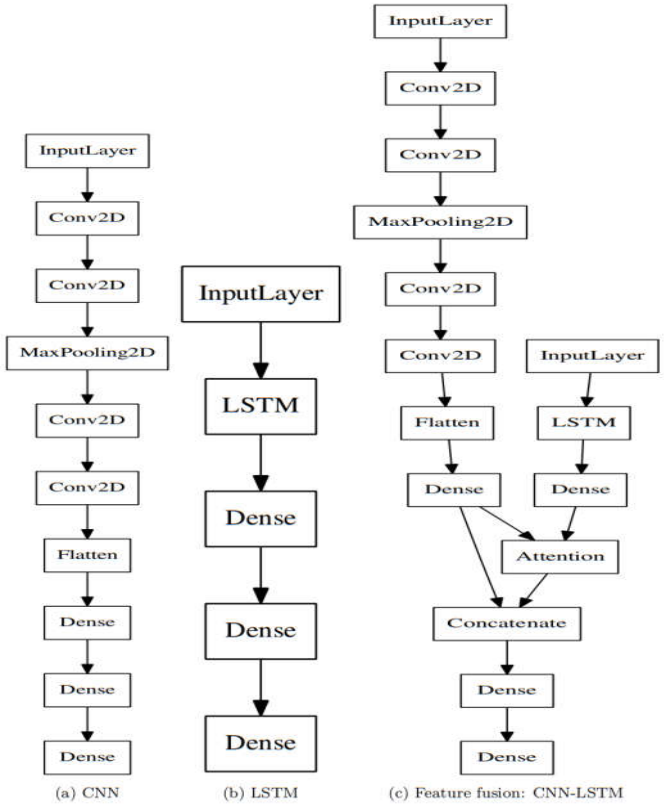


Figure 1. Three considered deep learning based HAR architectures

First, the data is split into fix-sized windows with an overlap of 50%. The Haar mother wavelet is used to extract the discrete wavelet transform (DWT) coefficients from the sensor data. Six HAR models have been made from the combination of 3 main architectures (CNN, LSTM, CNN-LSTM) and 2 types of input data (raw data, DWT coefficients). The detailed structures of the 3 main architectures are shown in Fig. 1. The CNN and LSTM models are implemented with standard sequential connections in which the CNN model contains several convolutional layers and maxpooling layer, followed by fully connected and softmax layers. In the feature fusion CNN-LSTM model, instead of connecting CNN and LSTM sequentially, the two sub-models are parallelly operated. An attention mechanism proposed by Luong *et al.* [5] is exploited to combine two outputs from the two sub-models.

The dataset is randomly split into 80% for training and 20% for validation. A $L1$-regularizer is used in all six models in order to avoid overfitting. In order to make a comprehensive investigation, two conventional ML methods: SVM and KNN ($k = 7$) are implemented. Each model is run for 10 experiments and the average accuracy is used as a performance metric. The results from the models are shown in Fig. 2. It can be clearly seen that, in most of the cases, the hand-crafted features give higher accuracy than the raw data except for the CNN model where the raw data achieve only 0.5% higher than the discrete wavelet transform features. In the feature fusion CNN-LSTM model, although both types of input data are used, it is not the one that gets the highest accuracy. The LSTM model which uses DWT data as input gets the highest accuracy and 3% higher than the raw input data.

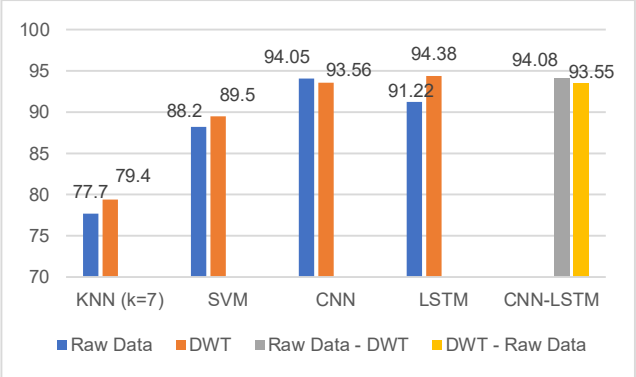


Figure 2. Performance comparison of different models with different input types

Ⅲ. Conclusion

In this paper, several feature extraction methods in both conventional and deep learning approaches have been implemented and applied to HAR. The experimental results indicate that although deep learning approach can automatically extract features from the raw data, in some cased, by exploiting the strength of domain knowledge in hand-crafted features, we can improve the performance of the system.

ACKNOWLEDGMENT

This work was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (2016-0-00564, Development of Intelligent Interaction Technology Based on Context Awareness and Human Intention Understanding).

References

[1] O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," IEEE Communications Surveys & Tutorials, vol. 15, no. 3, pp. 1192-1209, 2012.

[2] O. Steven Eyobu and D. S. Han, “Feature representation and data augmentation for human activity classification based on wearable IMU sensor data using a deep LSTM neural network," Sensors, vol. 18, no. 9, p. 2892, 2018.

[3] N. T. H. Thu and D. S. Han, “Utilization of postural transitions in sensor-based human activity recognition," in 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), 2020, pp. 177-181.

[4] J.-L. Reyes-Ortiz, L. Oneto, A. Sam`a, X. Parra, and D. Anguita, “Transition-aware human activity recognition using smartphones," Neurocomputing, vol. 171, pp. 754-767, 2016.

[5] Luong, M. T., Pham, H., & Manning, C. D. Effective approaches to attention-based neural machine translation. arXiv preprint arXiv:1508.04025, 2015.

Artificial Intelligence based Internet of Things Security

Mitra Pooyandeh Insoo Sohn

Division of Electronics & Electrical Engineering

Dongguk University

mitra.p@dongguk.edu isohn@dongguk.edu

Abstract

Artificial Intelligence (AI) and Internet of Things (IoT) is a soft smart revolution in industry 4.0. IoT is a world of sensors that connecting the physical objects such as computers, vehicles, appliances and other devices together and collect data over a wired or wireless network. On the other hand, these IoT systems are exposed to various types of cyber and physical attacks. As the modern threats continues to enlarge on, AI techniques, as intelligent methods which can learn and decide without the human intervention, have been widely used to enhance the IoT security in different ways. In this paper we survey the existing methods in which one of the AI methods has been used to defend against malicious attacks to IoT.

1. Introduction

The number of things connected to the Internet and using IoT technology is estimated at 14.2 billion as of 2019, and it reaches 25 billion by 2021 [1], and by 2025 more than 75 billion devices will be connected to the Internet [2]. Wireless Sensor Networks that monitor and control electric transmission tower, traffic lights, industrial machineries, and healthcare systems are kinds of IoT devices that attacks against them has bad influence on critical systems [3, 4]. Therefore, security is the greatest challenge for IoT. On the other hand, due to latency, transferring data to the cloud for processing is a time-consuming method. Hence, edge computing is an appropriate solution for transferring data processing to the edges. This causes to expose data to more attacks. One of the most recent approaches to enhance the IoT security is to utilize artificial intelligence (AI) methods. AI investigation continues to advance and it has gradually been applied to many fields of IoT security [2]. In this paper we review the most recent application of AI methods to increase the IoT security.

2. Materials

To know more about the probable security threats to IoT systems and existing AI methods to defend against different types of attacks, in this section, we explain briefly about the IoT systems, their security threats, and recent investigations about the application of AI for IoT security.

2.1. Internet of Things (IoT)

Internet of Things (IoT) now refers to billions of physical devices around the world that are connected to the internet, and which are collecting and sharing data with each other. Physical devices can refer to connected medical devices, a biochip transponder (think livestock), a solar panel, a connected automobile with sensors that alert the driver to a myriad of possible issues (fuel, tire pressure, needed maintenance, and more) or any object, outfitted with sensors, that has the ability to gather and transfer data over a network. Therefore, in the IoT the type of communication is machine-machine (M2M). IoT systems have three layers: *application layer* that provides service to users, *network layer* including GSM, WiFi, 3-5G, etc. ,and *perception layer* which consists of physical and MAC layer [5]. For successful implementation of Internet of Things (IoT), the important prerequisites including real time needs, availability of applications, data protection and user privacy, execution of the applications near to end users, and access to an open and interoperable cloud system.

2.2. IoT security threats

IoT systems exposed to different types of attacks including active and passive cyber attacks and physical attacks.
In the active attacks, malicious acts are carried out against data confidentiality as well as data integrity. They can also aim for unauthorized access and usage of the resources or the disturbance of an opponent's communications. There are variants of active attacks such as sybil, jamming, spoofing, Dos, data tampering, and malicious input attacks.
Passive attacks are performed in a way that it cannot be detected easily. This is due to the fact that the adversaries do not make any radio emissions. In passive attacks, attackers are typically hidden, and tries to collect data from communication lines. These types of attacks can be divided into different groups including eavesdropping, node malfunctioning, node tampering/destruction, and traffic analysis types [6]. Physical attacks refer to the attacks that physically damage IoT devices. The attackers do not need any network to attack the system. Therefore, this kind of attacks are subjected to physical IoT devices such as mobile, camera, sensors, routers, etc. by which the attackers interrupt the service.

2.3 AI used for IoT Security

As the modern threat landscape continues to enlarge on, adding artificial intelligence (AI) to a security strategy result in maintaining an effective security position. Considering the speed and complexity of modern cyber threats, network

security teams need the support of machine learning and other AI-based capabilities to detect, secure, and mitigate the attacks. Artificial Intelligence (AI) is classified to four major techniques: *machine learning* (ML), *fuzzy model*, *probabilistic models*, and *Metaheuristics* [7]. Recently, AI has enhanced the security of IoT in device authentication, DoS attack's defense, intrusion detection, and malware. AI techniques has unique solutions for these threats. The common processes of AI solutions are data collection, data pre-processing, model selection, data transformation, train and test, and model deployment [8]. Machine learning techniques provide the security for IoT by affecting the three layers of IoT with various methods such as Supervised, Unsupervised, and Reinforcement learning [9, 10].

In supervised learning the output is classified based on the input with a learning algorithm as in classification problems. In unsupervised learning there is not output for input data and the data is classified as what happens in clustering. In reinforcement learning the machine learns from interactions with human. Machine learning uses several techniques for IoT security such as classifying security attacks, Active learning for intrusion detection, security analytics learning-based malware detection system, learning-based authentication system, and hybrid intrusion detection system [11, 14].

Another AI method that is called Metaheuristic is a procedure which is designed to find a good solution to a difficult optimization problem. This algorithm is generally used for feature selection and tries to mimic biological, physical, and natural phenomena [12]. To increase the IoT security this method is used for intrusion detection and attack recovery.

There is also a fuzzy model technique that works based on fuzzy logic. Fuzzy logic is a method of reasoning which is similar to human reasoning. The approach of fuzzy logic includes all intermediate possibilities between digital values yes and no. This model is used for privacy and identify management, malware and attack detection by applying various methods such as clustering, classification, and ranking. The efficiency of security risk management depends on the speed and quality of clustering and classification of security threats [13].

Finally, the probabilistic model is a way to prove the existence of a structure with certain properties in combinations. The behavior of probabilistic systems modeled as discrete-time Markov chains (DTMCs), MDPs, or CTMCs. Indeed this model used for complex system attacks such as anomaly learning and detection, and security analytics.

3. Conclusion

In this paper we reviewed existing methods for enhancing the security of IoT devices and detection and mitigation techniques against attacks to IoT with AI techniques.

It turns out that various AI methods such as machine learning, fuzzy model, probabilistic model, and metaheuristics model have been used to make IoT secure against cyber and physical attacks. It has been deduced that AI-based methods have shown outstanding improvement in IoT security against cyber and physical attacks specifically in intrusion, anolamy, and malware detection. Certainly, in future due to artificial intelligence constant breakthroughs the IoT security will have a clear vision and it may soon offers the means to successfully secure the IoT.

Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRFK) funded by the Ministry of Education (2018R1D1A1B07041981).

References

[1] Seoyeon Kim, Jisu Park, Jaehyeok Jeong, Survay of IoT Platforms Supporting Artificial Intelligence, 2019.

[2] Massimo Merenda, Carlo Porcaro, Demetrio Iero, Edge Machine Learning for AI-Enabled IoT Devices: A Review, 2020.

[3] Kaviani, Sara, and Insoo Sohn. "Defense Against Neural Trojan Attacks: A survey." Neurocomputing (2020).

[4] Ioannis Stellios, Panayiotis Kotzanikolaou, Mihalis Psarakis, Cristina Alcaraz, Javier Lopez, A Survey of IoT-Enabled Cyberattacks: Assessing Attack Paths to Critical Infrastructures and Services, 2018.

[5] Syeda Manjia Tahsien, Hadis Karimipour, Petros Spachos, Machine learning based solutions for security of Internet of Things (IoT): A survey, 2020.

[6] Martins O. Osifeko, Gerhard P. Hancke, and Adnan M. Abu-Mahfouz, Artificial Intelligence Techniques for Cognitive Sensing in Future IoT: State-of-the-Art, Potentials, and Challenges, 2020.

[7] He Fang, Xianbin Wang, Fast Authentication and Progressive Authorization in Large-Scale IoT: How to Leverage AI for Security Enhancement?, 2019.

[8] Hui Wu, Haiting Han, Xiao Wangof, Sengli Sun, Research on Artificial Intelligence Enhancing Internet of Things Security: A Survey, 2020.

[9] Kotenko, I., Saenko, I., & Ageev, S. (2015, August). Countermeasure security risks management in the internet of things based on fuzzy logic inference. In *2015 IEEE Trustcom/BigDataSE/ISPA* (Vol. 1, pp. 654-659). IEEE.

[10] Abebe Abeshu Diro, Naveen Chilamkurti, Distributed attack detection scheme using deep learning approah for Internet of Things, 2018.

[11] Holzinger, Andreas. "Machine learning for health informatics." Machine Learning for Health Informatics. Springer, Cham, 2016. 1-24.

[12] Nahla Shatnawi, Qutaibah Althebyan, Wail Mardini, Detection of Insiders Misuse in Database Systems, 2011.

[13] Jiawei Su, Danilo Vasconcellos Vargas, Sanjiva Prasad, Lightweight Classification of IoT Malware based on Image Recognition, 2018.

[14] Kaviani, S., & Sohn, I., Influence of random topology in artificial neural networks: A survey. ICT Express, 6(2), 145-150 (2020).

Artificial Intelligence Platform Based for Smart Factory

Aji Teguh Prihatno
Electrical Electronics Engineering
Kookmin University
Seoul, South Korea
aji.teguh@gmail.com

Himawan Nurcahyanto
Electrical Electronics Engineering
Kookmin University
Seoul, South Korea
himawanurcahyanto@gmail.com

Yeong Min Jang
Electrical Electronics Engineering
Kookmin University
Seoul, South Korea
yjang@kookmin.ac.kr

***Abstract*— The effect of Artificial Intelligence (AI) on smart manufacturing is increasing rapidly. The AI-based open-source web application enables users to build and distribute live code, calculations, visualizations, and explanatory text documents, and also supports the data cleaning and conversion, numerical simulation, mathematical modelling, machine learning, and much more are used by the industries. This paper describes the processing data particle $PM_{2.5}$ and creates prediction using an AI algorithm from those data which the result can be implemented in the smart factory. To support the accuracy of the prediction of the AI method in the smart factory environment, the author uses Jupyter Notebook based on the source web application. This deployment will lead to performance improvement, cost reduction, process management, shortened product cycle production times, and increased productivity for the manufacturing sector.**

Keywords*— *Artificial Intelligence (AI), AI Platform, $PM_{2.5}$, Smart Factory, Manufacturing, Jupyter Notebook

I. INTRODUCTION

The advancement of Artificial Intelligence (AI) technology has continued to grow rapidly especially in the field of a smart factory. AI is one of the main key roles in smart factories that will lead the market size to expand. From statistics [1]. This shows us the increase in the revenue of the AI global market from 2018 to 2024 which is estimated to be around 126 billion USD as shown in Figure 1 below.

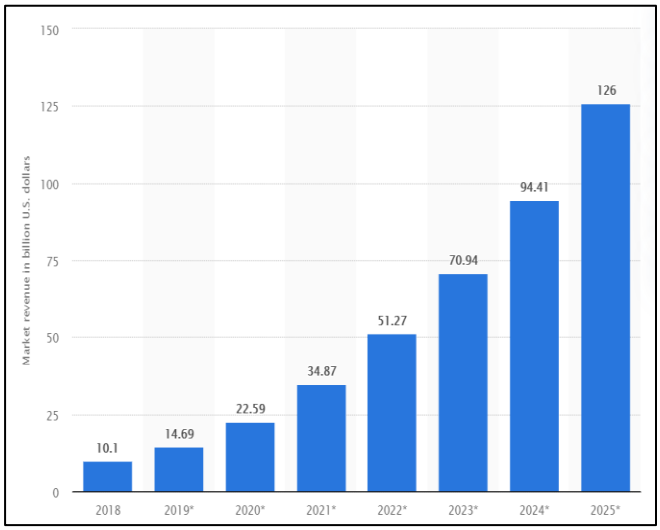


Figure 1. Revenues from the AI software market worldwide from 2018 to 2025 *(in billion U.S dollars)*

The increment of AI global market as Figure 1 shows, driven by three factors. The first factor, the growth of demand for the application of predictive maintenance and machinery inspection, which are widely spread usage of computer vision cameras in machinery inspection. The second factor, the implementation of the Industrial Internet of Things (IIoT), and the third factor is the use of big data in the manufacturing industry. This growing of AI global market needs to reduce operating costs and machine downtime also complements the growth of the application of predictive maintenance and machinery inspection in industries [2].

As a big element of this digital transformation, AI is being touted. Even with new innovative Manufacturing 4.0 innovations, the majority of connected devices in manufacturing, including initiation, management, tracking, and feedback, are still unable to make decisions without human intervention. Infusing knowledge into these physically linked things will increase the value that can be produced from them exponentially. The purpose of a smart factory is endorsed by AI; one that will function with minimal human contact [3].

In addition to improving the environment and the quality of indoor air quality in smart factories, AI-based smart factory can also optimize efficiency, quality, cost, and resource management processes at the global production level. It is also directly proportional to the size of the AI market that drives the growth of the size of the smart factory market.

This paper explains a smart factory that runs the AI Platform, related to the measurement of environmental values in the plant, such as Particulate Matter ($PM_{2.5}$).

II. AI PLATFORM BASED ON OPEN SOURCE

To process and train the data of $PM_{2.5}$ in the Smart Factory, the authors have implemented the AI algorithm using Jupyter Notebook. The Jupyter notebook is based on a set of collaborative computing open standards. For interactive computing on the internet think HTML and CSS. Third-party developers will utilize these open standards to create customized applications with embedded interactive computing.

With its modular framework, Jupyter Notebook expands the notebook to visualization, multimedia,

collaboration, and more beyond code. It stores code and output, along with markdown notes, in an editable document called a notebook, in addition to running the code. This is sent from browser's users to their notebook server when they save it which stores it as a JSON file with a.ipynb extension on the disk. The design architecture of Jupyter Notebook and its interface can be seen in Figure 2.

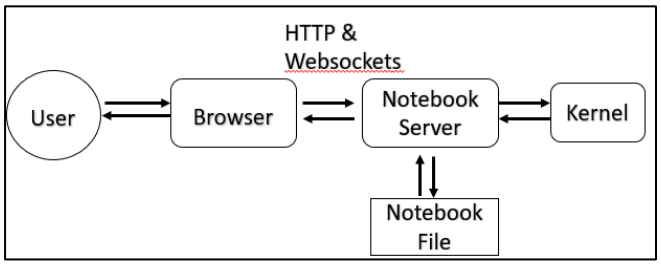


Figure 2. The Jupyter Notebook Interface

The notebook server is responsible for saving and loading notebooks, so even if the user doesn't have the kernel for that language, the user still can edit notebooks. There's nothing the kernel knows about a notebook document: it simply gets code cells sent to run when the user runs them [4]. The kernel of a notebook is a "computer engine" that executes the code found in the Notebook file. A python code is executed by the ipython kernel which kernel exists for many other languages (official kernels). The associated kernel is automatically launched when the user opens a Notebook File. The kernel performs the computation and generates the results when the notebook is executed (either cell-by-cell or with the menu Cell -> Run All). The kernel can consume important CPU and RAM depending on the type of computation. Remember that the RAM is not released until the user has shut down the kernel.

III. Implementation and Result

To predict $PM_{2.5}$ concentration in a factory environment, the authors were using RNN (Recurrent Neural Network) that we can shape several sets of sequences, from the data using time from a continuous data set, and also from certain sets of data sequences, we can observe the correlation between sequences. To form a network, Simple RNN has several neuron-nodes. Per node (neuron) has a real-valued activation that varies in time. Each relation has a real-value weight, which can be modified in each case. [5].

The human presence in the cleanroom is the most easily observed associations in $PM_{2.5}$ data, according to the aim of this research is to get the smart factory environment. The $PM_{2.5}$ concentration will also increase as more individuals come into the cleanroom.

After processing the data using Jupyter Notebook, we get the experimental result to predict one day ahead of the concentrate of $PM_{2.5,}$ in respectively, the accuracy of Mean Absolute Error (MAE) 0.08, Root Mean Square Error (RMSE) 0.11, and Mean Square Error (MSE) 0.01.

As we can see from Figure 3, the graph of prediction in the blue line is near to the real data which the line is red.

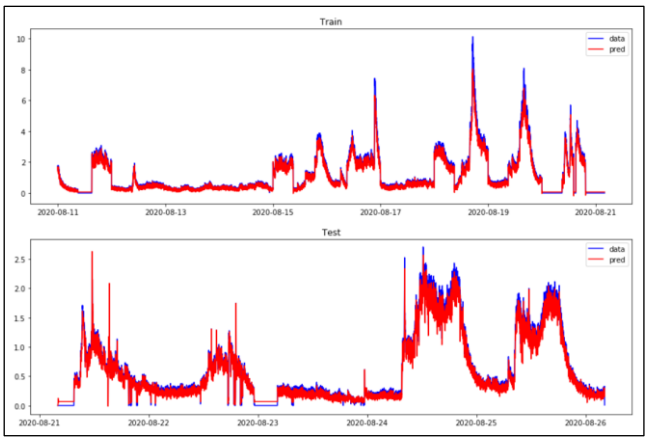


Figure 3. Result Implementation. Prediction $PM_{2.5}$ using RNN Method.

IV. Conclusion

The purpose of the work presented was to implement AI by building prediction machines as the basis for the smart factory with time-series data. The RNN model reveals substantial results in the long-term $PM_{2.5}$ concentration based on historical data. Nevertheless, to improve the accuracy of the prediction machine and extend the reach of the smart factory operated by the AI network, the model needs to be further and more varied in the future. Finally, the prediction of $PM_{2.5}$ status will assist operators in the operational policy and allocation of smart factory capital.

Acknowledgement

This work was supported by the Ministry of Trade, Industry, and Energy (MOTIE) and Korea Institute for Advancement of Technology (KIAT) through the International Cooperative R&D program (Project ID:P0011880)

References

[1] Statista, "'Global size of the smart factory market in 2019 and 2024,'" *Online*. https://www.statista.com/statistics/872289/worldwide-smart-factory-market-size/.

[2] M. and Markets, "Artifical Intelligence in Manufacturing Market," 2020. https://www.marketsandmarkets.com/Market-Reports/artificial-intelligence-manufacturing-market-72679105.html.

[3] J. Lee, J. Singh, and M. Azamfar, "Industrial artificial intelligence," *arXiv*, 2019.

[4] Jupyter, "Jupyter Notebook Architecture." https://jupyter.readthedocs.io/en/latest/projects/architecture/content-architecture.html.

[5] Y. T. Tsai, Y. R. Zeng, and Y. S. Chang, "Air pollution forecasting using rnn with lstm," *Proc. - IEEE 16th Int. Conf. Dependable, Auton. Secur. Comput. IEEE 16th Int. Conf. Pervasive Intell. Comput. IEEE 4th Int. Conf. Big Data Intell. Comput. IEEE 3rd Cyber Sci. Technol. Congr. DASC-PICom-DataCom-CyberSciTec 2018*, pp. 1068–1073, 2018, doi:10.1109/DASC/PiCom/DataCom/CyberSciTec.2018.00178.

Prophet 모델을 사용한 기상데이터 예측

김준석, 김성희, 윤주상, 강재환*
동의대학교 산업 ICT 공학과, *동의대학교 인공지능그랜드 ICT 연구센터

[junsuk.kim; sh.kim; jsyoun; *jh.kang]@deu.ac.kr

Meteorological Data Forecasting using Facebook Prophet Library

Junsuk Kim, Sung-Hee Kim, Joosang Youn, Jae-Hwan Kang*
Dong-Eui Univ., * Dong-Eui Univ. Grand ICT Research Center

요 약

본 논문은 시계열 데이터 예측에서의 훌륭한 성능으로 주목받고 있는 Facebook 의 Prophet 모델을 사용하여 기상 데이터 (온도, 습도, 풍속)를 학습시키고 시계열 데이터 변화양상을 예측하는 연구를 진행하였다. Prophet 의 예측 성능은 다양한 모델 성능 평가 기법 (Index of Agreement; IOA, Mean Error, Normalized Root Mean Squared Error; NRMSE)을 사용하여 수치화하였다. 그 결과, Prophet 모델은 온도 데이터 예측에는 탁월한 성능을 보여주는 반면 습도와 풍속 데이터 예측에는 좋은 성능을 보여주지 못하였다. 이 연구결과는 향후 인공지능 알고리즘 또는 다양한 시계열 예측 기술을 사용한 기상 데이터 예측 연구에 기초 자료가 될 것으로 생각된다.

I. 서 론

최근 회귀분석, 심층 신경망 등 다양한 기술을 활용한 시계열 예측 연구가 활발히 진행되고 있다 [1,2]. 이 연구에서는 시계열 예측 분석 대회에서 특별한 전략을 사용하지 않고 데이터만을 활용하여 훌륭한 예측 성능을 보여주었던 Facebook 의 시계열 분석 라이브러리 Prophet 을 사용하여 기상 변화 (온도, 습도, 풍속) 를 예측해 보았다 [3]. 이 모델을 통해 예측된 기상데이터 값을 실제 데이터와 비교하여 예측 정확도를 평가하였다.

II. 본론

본 논문에서는 2018년 Taylor S.J. 와 Letham B. 이 발표한 Facebook 의 시계열 분석 라이브러리 Prophet 을 사용하여 기상 인자들의 시계열 예측 모델을 구현하였다 [3].

분석에 사용된 기상데이터는 기상청 기상자료 개방포털 (https://data.kma.go.kr/) 에서 제공하는 온도, 습도, 풍속 데이터를 사용하였다. 특히, 강원도 고성군 지역의 3 년간 (2016.01.01. ~ 2018.12.31.) 의 관측치를 트레이닝에 사용하고 그 이후의 4 개월간 (2019.01.01. ~ 2019.04.30.) 의 관측치를 성능평가를 위한 테스트에 사용하였다. 입력 데이터로는 각 시간별로 제공되는 온도, 습도, 풍속 관측치를 사용하였으며, 각각 기상데이터 종류에 따라 시계열 예측 모델 입력으로 트레이닝 데이터와 테스트 데이터를 가진 두 개의 데이터 시퀀스가 사용되었다.

Prophet 모델의 예측 성능을 정량적으로 평가하기 위해, 즉, 모델의 예측 결과를 실제 관측 값과 비교하기 위하여 각각의 기상데이터 조건에 대하여 IOA (Index of Agreement), ME (bias), NRMSE (Normalized Root Mean Squared Error) 를 계산하였다. 이 평가 방법들은 인공지능 알고리즘을 활용하여 시계열 예측력을 평가하는 용도로 많은 연구에서 사용되고 있다 [2,4].

III. 결론

온도 데이터 예측의 경우, IOA는 0.78, ME는 0.50, NRMSE는 14.35% 의 결과를 얻었다. 또한, 습도 데이터 예측의 경우 IOA는 0.38, ME는 -5.02, NRMSE는 35.95% 의 결과를, 풍속 데이터 예측의 경우 IOA는 0.36, ME는 0.12, NRMSE는 35.41% 의 결과를 얻었다. 결과적으로 세 종류의 입력데이터 중에서 Prophet 모델은 온도 데이터의 예측 능력이 습도, 풍속 데이터 예측력 보다 우수한 것으로 나타났다. 추가로 모델의 예측 결과와 실제 관측 값 간의 상관관계를 분석하였으며, 그 결과 온도의 경우 $r = 0.64$, $p < .01$, 습도의 경우 $r = 0.29$, $p < .01$ 로서 통계적으로 유의미한 유사성을 확인하였으며, 풍속의 경우 $r = 0.12$, $p = .19$ 로 통계적으로 유사성을 찾지 못했다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 Grand ICT 연구지원센터 지원사업의 연구결과로 수행되었음 (IITP-2020-0-01791).

참 고 문 헌

[1] Yu, P-S., Chen, S-T., Chang, I-F. (2006) Support vector regression for real-time flood stage forecasting, Journal of Hydrology, vol. 328, no. 3-4, pp. 704-716.

[2] Kim, D., Kim, J., Kwak, J., Necesito, I. V., Kim, J., Kim, H. S. (2020) Development of Water Level Prediction Models Using Deep Neural Network in Mountain Wetlands, Journal of Wetlands Research, vol. 22, no. 2, pp. 106-112.

[3] Taylor, S. J., Letham, B. (2018) Forecasting at scale, The American Statistician, vol. 72, no. 1, pp. 37-45.

[4] Cho, K., Lee, B-Y., Kwon, M., Kim, S. (2019) Air quality prediction using a deep neural network model, Journal of Korean Society for Atmospheric Environment, vol. 35, no. 2, pp. 214-225.

전리층 총 전자량 데이터에 적용한 LSTM 기반의 지진 이상현상 탐지

조건우, 박동건, 김홍국
광주과학기술원

joungju257@gist.ac.kr, dongkeon@gist.ac.kr, hongkook@gist.ac.kr

LSTM-based Earthquake Anomaly Detection Applied to Total Electron Current Data

Geon Woo Cho, Dong Keon Park, Hong Kook Kim
Gwangju Institute of Science and Technology

요 약

본 논문에서는 전리층 총 전자량(TEC, Total Electron Current) 데이터에서의 이상현상을 long short-term memory (LSTM) 기반 모델로 탐지하는 기법을 제안한다. 또한 제안된 방식은 Gaussian mixture model, K-means clustering, 그리고 support vector machine 의 기계학습 기반의 기법과 비교한다. 실제 지진 데이터와 비교하여 이상현상 탐지 성능을 비교한 결과, LSTM 기반의 이상현상 탐지 성능이 기존의 기계학습 기반의 성능과 비교하여 F1-score 로 약 20% 향상됨을 확인하였다.

Ⅰ. 서 론

전리층 총 전자량(TEC, Total Electron Current) 데이터는 지구 전리층에 있는 전자의 밀도를 나타내는 지표로, 지진과의 상관 관계는 꾸준히 논의되어 왔다[1]. 하지만, 규모 6.0 이상의 지진 중 특정한 사례에 대해서만 분석된 것이 대부분이었다[2, 3]. 본 논문에서는 시계열 데이터 학습에 사용되는 지도 학습 모델 중 하나인 long short-term memory (LSTM)[4] 기반의 모델을 이용한 TEC 데이터에서의 이상현상을 탐지하는 기법을 제안한다. 제안된 기법의 성능은 2016 년 1 년간 미국에서 지진이 가장 많이 발생한 경도 -117°~-120°, 위도 35°~40° 사이의 TEC 데이터를 이용하여 실제 지진 데이터와 비교하여 평가된다. 또한, 머신러닝 기법 중에 Gaussian mixture model (GMM), K-means clustering, 그리고 support vector machine (SVM) 기반의 기법과도 성능을 비교한다.

Ⅱ. 본론

2.1 TEC 데이터

TEC 데이터는 NOAA 에서 2016 년 1 월 1 일 0 시부터 12 월 31 일 24 시까지 15 분 간격 (총 35040 개), 위도와 경도를 1 도 간격으로 관측한 것을 사용하였으며[5], 위도, 경도 별로 관측된 TEC 데이터들에 평균을 취한 후 이용하였다. 단위는 $10\ \mathrm{TECU} = 10^{17}\ electrons/m^2$ 이다. TEC 데이터는 태양 활동에 영향을 크게 받아 계절별 분포 차이가 크기 때문에, 학습 정확도 향상을 위해 1~3 월, 4~6 월, 7~9 월, 10~12 월 4 분기로 나누어 사용하였고, 전체 데이터를 학습시킨 결과와 비교하였다.

전체 TEC 데이터의 70%를 학습 데이터, 나머지 30%를 평가 데이터로 분리하였다. [3]의 연구에서는 지진 발생 이전 3 일 안에 TEC 데이터에서 두드러지는 변화가 나타남을 보인 반면, 본 연구에서는 규모 4.5 이상의 지진에 대한 정보를 이용하였기에 본진이 끝난 후에도 여진이 있을 것을 감안하여 지진 전후로 3 일을 이상현상으로 레이블링하였다.

2.2 방법론

2.2.1 K-means Clustering

전체 데이터에 대해 군집의 수를 1 개부터 늘려가며 elbow point 로 군집의 개수를 6 개로 정하였다. K-means clustering 기법을 통해 6 개의 군집을 형성한 후, 군집들의 중심으로부터 가장 먼 거리에 위치한 군집에 속한 데이터들을 이상현상으로 분류하였다.

2.2.2 One-class Support Vector Machine

One-class SVM 에 기반한 학습 기법을 적용해 정상 학습 데이터에 대한 경계를 학습한 후, 전체 데이터를 넣어 학습된 경계 밖에 있는 데이터들을 이상현상으로 판별하였다.

2.2.3 Gaussian Mixture Model

전체 혼합 개수를 3 개로 설정한 후, expectation-maximization (EM) 알고리즘을 이용하여 정상 데이터에 대한 확률 분포를 추정하였고, 문턱치보다 높은 확률 값을 가지는 데이터들을 이상현상이라고 탐지하였다.

2.2.4 Long Short-Term Memory Model

70%의 학습 데이터 중 정상 데이터만 선택하여 LSTM 이 정상일 때의 분포를 예측할 수 있도록 학습시켰다. 학습된 파라미터를 이용해 10 단계 예측값을 도출하였고, 예측된 결과를 바탕으로 다음과 같은 Mahalanobis distance 기반의 anomaly score 를 정의하였다.

$$\text{Anomaly Score} = (e^{(i)} - \mu)\Sigma^{-1}(e^{(i)} - \mu)^T \quad (1)$$

여기서, μ와 Σ 는 평균 및 공분산을, e^i 는 i(범위: 0~35039)번째 index 에서의 오차를 의미한다. 식 (1)의 anomaly score 가 문턱치 값보다 높은 값을 가지고 있으면 이상현상이라고 판별하였다.

2.3 실험 및 성능 평가

데이터들을 학습시킬 때, 잡음이 약간 섞인 데이터에 대해서도 비슷한 결과를 얻을 수 있도록 아래와 같은 데이터 증강 기법을 적용하였다.

표 1. 잡음 인가 기반의 데이터 증강 기법

데이터 증강 기법
1: Total data = {Original data}
2: Noise ~ $N(0,1)$, $i = 0$
3: **While** $i < 0.05$
4: Total data ← Original data + $\sigma \times$ Noise $\times i$ (σ 는 Original data 의 표준 편차)
5: $i = i + 0.0005$

LSTM 의 학습 결과에 대한 예시는 그림 1 과 같다. 초록색 그래프는 1 단계 예측에 대한 결과이고, 파란색 그래프는 이전 단계에 대한 값을 주지 않고 재귀적으로 예측한 결과로, 정상 데이터에 대한 모델의 학습 정도를 확인할 수 있다.

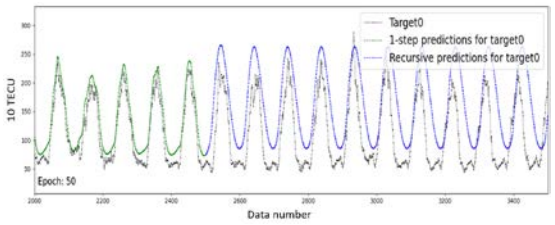


그림 1. LSTM 기반의 TEC 데이터 예측 (target0: 봄 데이터, 전체 데이터 번호의 범위: 0~35039)

LSTM, K-means clustering, SVM, GMM 을 이용해 이상 현상을 탐지한 결과에 대한 예시는 그림 2 와 같다.

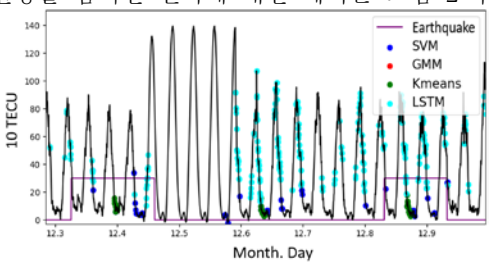


그림 2. LSTM, K-means clustering, SVM, GMM 기반의 이상현상 탐지 성능 비교

전체 데이터 35,040 개 중 이상현상이라고 표기된 값은 총 5,159 개로, 이상현상의 개수에 비해 정상일 때의 개수가 많은 불균형 데이터이기에, 정확도로는 예측 결과에 대한 분석을 정확하게 할 수 없었다. 따라서 precision 과 recall 의 조화 평균에 해당하는 F1-score 를 이용해 예측 결과를 분석하였다. LSTM 에서의 F1-score 는 anomaly score 에 대해 임의의 문턱치 값을 잡았을 때 얻어지는 F1-score 중 가장 큰 값으로 하였다. 예측 결과는 표 2 와 같다.

표에서 보는 바와 같이, F1-score 측정 결과, LSTM 기반의 기법의 F1-score 가 기존의 기계학습 기반의 기법보다 높고, 계절별 학습에 대해서는 LSTM 기반 기법이 이상현상을 F1-score 로 약 20%만큼 잘 예측하고, 전체 데이터에 대한 학습에서는 이상현상을 약 10%만큼 잘 예측함을 확인할 수 있었다. 또한, LSTM 을 이용해 학습시킬 시, 전체 데이터를 모두 학습시킨 기법에 비해 데이터를 계절별로 따로 학습시킨 기법이 이상현상을 약 15%만큼 잘 예측함을 확인할 수 있었다.

표 2. 각 기법별 단일모델과 계절별 모델의 F1-score 비교

Method	단일 모델	계절별 모델			
		spring	summer	autumn	winter
K-means	0.000	0.016	0.005	0.017	0.011
GMM	0.033	0.031	0.001	0.011	0.013
SVM	0.031	0.009	0.000	0.004	0.032
LSTM	0.121	0.197	0.218	0.223	0.378

Ⅲ. 결론

본 논문에서는 TEC 데이터에서의 이상현상 탐지를 위해 제안된 LSTM 기반의 기법의 F1-score 가 GMM, K-means clustering, 그리고 SVM 에 비해 좋음을 확인하였다. 또한, TEC 데이터의 계절별 분포 차이가 있음을 각 계절별 F1-score 를 통해 간접적으로 확인할 수 있었다. 하지만, LSTM 기반 기법의 F1-score 가 상대적으로는 높지만 절대적인 수치를 봤을 때 이상현상 탐지를 효율적으로 하지는 못함을 확인할 수 있었다. 이는 disturbance storm time (Dst) 지수, K(George) 지수에 의해 발생하는 잡음이 지진에 의해 발생하는 TEC 데이터의 변화보다 더 크거나 불규칙하게 발생할 때가 많아 학습이 제대로 이루어지지 않아 그런 것으로 판단된다. 추후 규모가 큰 지진이라는 제약을 두어 잡음의 영향을 상대적으로 줄이고, Dst, K 지수 등을 활용한 잡음제거 기법을 도입하는 등의 학습 방법을 도입해 LSTM 기반의 이상현상 탐지 기법의 성능을 높이고자 한다.

ACKNOWLEDGMENT

본 연구는 광주과학기술원 전기전자컴퓨터공학부 오디오지능연구실 인턴십의 결과이며, 2020 년도 광주과학기술원 GRI(GIST 연구원)의 지원을 받아 수행된 연구임.

참 고 문 헌

[1] M. Hayakawa, “Earthquake prediction with electromagnetic phenomena,” in *Proc. AIP Conference,* vol. 1709. no. 1. p. 020002, 2016.

[2] W. Liu and L. Xu, “Statistical analysis of ionospheric TEC anomalies before global M w ≥ 7.0 earthquakes using data of CODE GIM,” *Journal of Seismology*, vol. 21, no. 4, pp. 759-775, 2016.

[3] J. Y. Liu, Y. I. Chen, Y. J. Chuo, and H. F. Tsai, “Variations of ionospheric total electron content during the Chi-Chi earthquake,” *Geophysical Research Letters*, vol. 28, no. 7, pp. 1383-1386, 2001.

[4] P. Malhotra, A. Ramakrishnan, G. Anand, L. Vig, P. Agarwal, and G. Shroff, “LSTM-based encoder-decoder for multi-sensor anomaly detection,” arXiv:1607.00148, 2016.

[5] N. G. D. Center, “Real-time US-Total Electron Content: Vertical and Slant” *NOAA National Centers for Environmental Information* (NCEI), 11-Jul-2006. (https://www.ngdc.noaa.gov/stp/iono/ustec/products/)

미세먼지의 빅데이터/AI 분석 및 예측을 위한 IoT 측정 단말기 개발

우동식, 백봉현*
대구가톨릭대학교, *(주)아르고스
dswoo@cu.ac.kr, *wefbbh@argosinc.co.kr

Development of a IoT measurement terminal for big data/AI analysis and prediction of fine dust

Woo Dong Sik, Back Bong-Hyun*
Daegu Catholic Univ., *Argos, Inc.

요 약

본 논문에서는 지역별로 상이한 환경조건(산업, 산림, 인구밀도, 자동차 등)과 다양한 산업 분야에서 발생하는 환경변수 데이터와 지역(구군별)별 발생되는 실제 대기 정보(미세먼지)를 빅데이터 및 인공지능 기법을 이용하여 미세먼지의 원인 분석과 예측이 가능한 빅데이터 플랫폼에 사용가능한 태양광을 이용한 IoT 측정 단말기를 개발하였다. 이를 통하여 최근 이슈화 되고 있는 미세먼지의 보다 정확한 세부 지역별 예측에 있어 빅데이터와 인공지능 기술을 활용한 플랫폼 서비스에 활용 가능함을 보였다.

Ⅰ. 서 론

최근 대기환경오염의 위험에 대한 환경 문제가 이슈화되고 있으며, 특히 잦은 초미세 먼지의발생으로 인체 유해성 문제가 점차 대두되고 있다. 2016년부터 정부에서는 경유차규제와 석탄화력 발전소 규제를 통한 저감 정책을 발표하였으나, 정확한 미세먼지의 발생원인과 해결책을 제시하지 못하고 있다. 대기환경에 영향을 미치는 지역별 환경변수의 차이로 정확한 미세먼지 등의 예측정보 제공의 어려움이 있다[1-2]. 따라서 지역적 환경조건에 기반하여 공공데이터기반 환경변수를 위한 빅데이터 처리 플랫폼과 상세 지역별 대기질 정보 및 데이터를 제공하기 위한 통계분석 및 인공지능 기술기반 예측시스템을 제안하고 이러한 환경데이터 수집을 위한 IoT 측정 단말기를 제안하였다.

Ⅱ. 본론

기존의 기상청 데이터를 기반으로 하는 전국단위의 미세먼지 측정예보 방식의 신뢰성 부족을 감안하여 친환경 대기질 측정 장비로부터의 상세 지역별 대기질 데이터 및 공공데이터 측정하도록 하였다. 지역별 환경변수 특성과 기상청 데이터를 결합한 빅데이터 수집·저장·분석이 가능한 플랫폼을 제안하였다. 지역 또는 소규모 단위의 보다 정확한 미세먼지의 예측을 위해 그래디언트 부스팅기법의 인공지능기법을 적용하여 상세 지역별 대기질정보 및 각 지역별 정확한 미세먼지 예측을 통한 대국민 생활안전 서비스를 제공할 수 있다. 이를 위해 로라(LoRa)망을 통한 태양광발전 소규모 미세먼지 측정기로부터 수집된 기상데이터(미세먼지, 초미세먼지, 풍향, 이산화탄소 등)를 구간별/지역별로 수집하여 빅데이터분석 처리를 통해 구간별/지역별 상세한 대기질맵(지도)를 구축하도록 하고, 외부데이터(기상청, 국도지리정보, 시도별 자동차 보유량 및 지역별 공장, 농업 등의 생산환경)등의 각 기관별로 제공되는 다양한 데이터포맷의 지역 환경변수를 연계하여 지역의 미세먼지 발생원인과 예측할 수 있는 빅데이터 수집·저장·분석시스템과 시각화 서비스를 개발하였다. 제안된 플랫폼의 구성도는 그림 1과 같다. 웹 UI를 통한 데이터의 수집·관리, DataBase화와 통계분석도구 R을 활용하여 빅데이터 통계분석 및 시각화 및 자동 레포팅을 위한 자동화 통계가 가능하도록 하였다.

1) 데이터 수집

미세먼지 예측을 위해 대기질 측정 Device로부터 대기환경을 측정하여 대기 및 고도·위치 등의 정보를 소프트리얼타임으로 로라망을 통해 통합 수집모듈로 송신하는 네트워크 모듈을 개발하였다. 공공데이터 수집 모듈은 지역의 정확한 대기질정보와 대기질 예측 및 발생원인을 분석하기 위해 기상청, 환경관리공단 및 통계청 등에서 측정된 데이터를 수집할 수 있도록 하였다. 통합데이터 수집 모듈은 Device 및 공공데이터에서 수집될 데이터의 포맷이 상이함에 따라 분석에 필요한 데이터 포맷 및 수집경로를 관리할 수 있도록 한다. 그림2는 제안된 태양광을 이용한 LTE IoT 측정 단말기의 사진이다. LTE Cat. M1 통신 모듈을 적재하고, 아두이노보드와 미세먼지 측정 모듈 연동을 하였다. 태양광발전패널과 내부배터리를 연동하여 사용하고 TCP/IP 네트워크를 통한 데이터 수신이 가능하도록 하였다.

2) 데이터 저장

수집된 대량의 데이터의 안정적, 신뢰적 저장을 위해 HDFS(Hadoop Distributed File System)시스템을 구축하고, 분석된 결과에 대한 빠른 저장과 조회를 위해 NoSQL(MongoDB)를 활용하도록 한다.

3) 데이터 분석

수집된 데이터로부터 대기오염에 대한 고차원 통계분석을 위해 오픈소스 기반의 통계R을 웹버전으로 개발 한다. 또한 스케줄링에 의해 지속적으로 수집된 데이터의 통계분석 결과를 레포팅할 수 있도록 하였다. 대기오염 예측분석은 각 지역별 Device로부터 수집된 대기정보와 공공데이터로부터 수집된 공장 등의 지역현황 정보 및 기상청으로 부터의 전국 기상데이터 등을 기반으로 각 지역별(시군구동)의 상세한 미세먼지 현황과 예

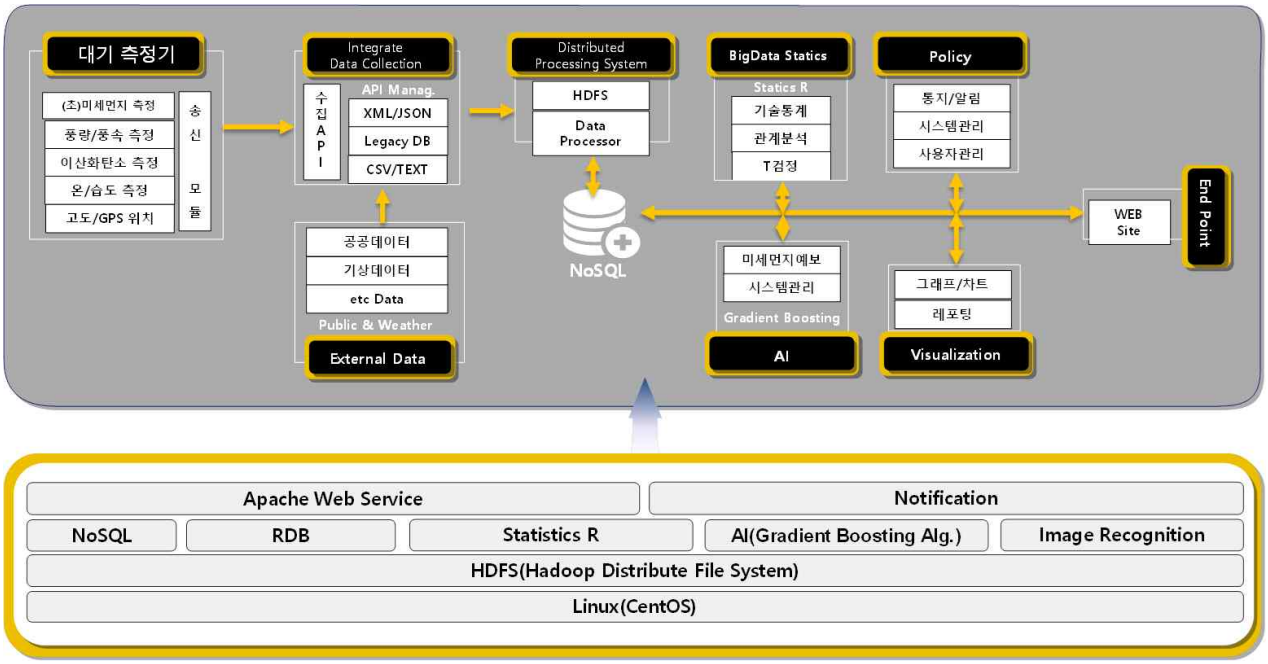


그림1. 제안된 플랫폼 구성도

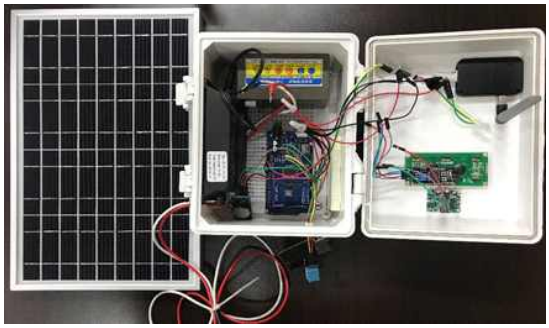


그림2. 제작된 LTE기반 IoT 측정 단말기

측을 통한 우리동네 미세먼지 정보를 산출하며, 수집된 데이터를 AI기법 중 그래디언트 부스팅 알고리즘(Gradient Boosting Algorithm)을 적용하여 대기질 분석 및 예측을 하였다.

4) 시각화 및 데이터제공

분석된 결과물의 시각화 전달을 위해 웹에 미세먼지의 현황을 실시간으로 조회할 수 있는 웹사이트를 활용하며 회원관리, 외부데이터 연계 및 정책 등을 설정할 수 있도록 백오피스를 활용할 수 있다. 원본데이터 및 빅데이터 분석정보를 CSV, XML, JSON 형식으로 데이터를 제공할 수 있도록 API를 제공할 수 있다. 그림 3은 에어코리아 API와 연동하여 홈/미세먼지 농도를 PM10, PM2.5로 구분하여 시간대별로 노출하여 최종 수집 시간의 선택 지역의 단일 미세정보 노출 사례를 보여주고 있다.

Ⅲ. 결론

본 논문에서는 기존의 기상청 데이터를 기반으로 하는 전국단위의 미세먼지 측정예보 방식에서의 고려되지 못한 지역별환경변수 특성과 기상청 데이터를 결합한 빅데이터 수집·저장·분석 플랫폼을 위한 IoT 측정 단말기를 개발하였다. 개발된 측정 단말기로 수집된 데이터를 활용하여 지역단위의 보다 정확한 미세먼지의 예측을 위해 그래디언트 부스팅 기법의 인공지능기법을 적용하여 정확한 미세먼지 예측이 가능하도록 하였다.

그림3. 홈/미세먼지 AI Air API 사례

제안된 시스템은 세분화된 대기질 정보분석과 예측 시스템 등 다양하게 응용될 수 있을 것이다.

ACKNOWLEDGMENT

본 논문은 ㈜아르고스와 산학공동연구에 의하여 수행되었음.

참 고 문 헌

[1] 강성원외, “환경 빅데이터 분석 및 서비스 개발,” 한국환경정책·평가연구원, 2017.

[2] 임준묵, “기상환경데이터와 머신러닝을 활용한 미세먼지농도 예측 모델,” *한국IT서비스학회지* 제18권, 제1호, pp. 173-186, 2019.

ICT 표준화전략맵 Ver.2021 기반 인공지능 적용 스마트헬스 분야 ICT 국제표준화 전략 연구

황유철, 고준호, 조수진, 이영억, 오구영, 김대중*

한국정보통신기술협회

hyc@tta.or.kr, jhko@tta.or.kr, sjcho@tta.or.kr, 2duddjr@tta.or.kr, ohky@tta.or.kr, kdj@tta.or.kr*

A Study on the International Standardization Items and Strategies in Intelligence Smart Health based on the Ver.2021 ICT Standardization Strategy Map

Hwang YouChul, Ko Jun Ho, Cho Su Jin, Lee Young Eok, Oh Ku Yeong, Kim Daejung*

Telecommunications Technology Association

요 약

본 논문의 표준화전략맵 Ver.2021은 스마트헬스 기술의 인공지능 기술을 적용한 영상데이터에 관한 IPR 현황 분석, 국제표준화기구의 동향과, 국내표준화기구의 현황을 살펴보았다. 이러한 동향 및 현황 분석을 통해 국제표준화 대응 전략을 도출했다. 본 논문은 스마트헬스의 인공지능 관련 영상데이터에 대한 국제 표준화 추진전략을 살펴본다.

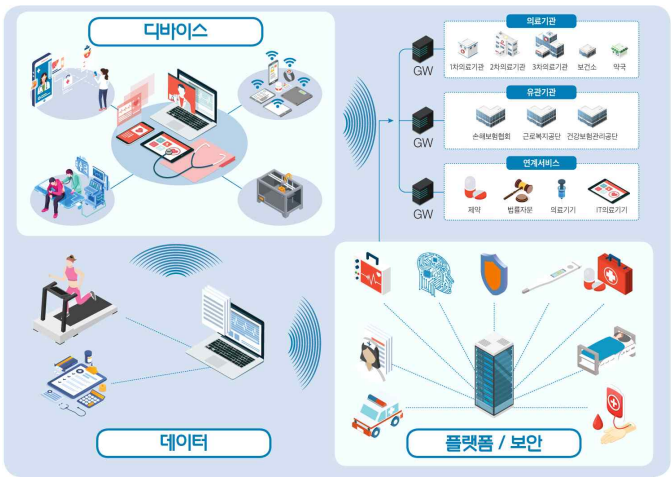
Ⅰ. 서 론

2019년 12월 중국 우한을 시작으로 전 세계적으로 코로나바이러스 감염증19(COVID-19, 이하 코로나19) 확진자 및 사망자가 지속적으로 증가함에 따라 인공지능을 적용한 스마트헬스 분야가 더욱 주목받고 있다. 또한, 스마트헬스 분야 비대면 서비스·산업 육성을 통해 경기회복 및 세계적 흐름인 디지털 전환(Digital Transformation)을 선도할 수 있을 것으로 기대되고 있다.

정보통신방송 분야의 기술은 시장 변화가 급속히 진행되어 적시에 대응하는 것이 최우선으로 강조되고 있으나 이에 따르는 표준은 미비한 실정이다. 표준은 기술개발 결과를 시장에 연결하는 다리(Bridge) 역할을 하고 있어, 관련 시장 선점을 위한 글로벌 표준 경쟁은 갈수록 치열해지고 있다. 이에 한국정보통신기술협회(이하 TTA)에서는 매년 롤링 플랜(Rolling-Plan)으로 ICT 전략 기반 신산업의 시장 선점을 위해 국내외 기술정책 및 동향을 반영한 유망 중점기술별 표준화 전략을 수립하여 ICT 표준화전략맵을 발간하고 있다.

TTA ICT 표준화전략맵은 사전조사분석을 통한 중점기술 선정, 전문가 회의를 통한 기술별 현황분석 및 전략수립과 중점 표준화 항목을 도출한다. 사전조사분석은 ITU, JTC1, ETSI, IEEE 등 주요 국제표준화기구의 핵심 표준화 항목, 미국, 유럽, 일본, 중국 등 국내·외 주요 ICT 정책, 가트너, IDC, ETRI, KISA 등 국내외 주요기관 보고서 및 매체 선정 유망기술 분석을 통해 이루어진다. 또한 약 160여회 회의를 통해 기술별 현황분석 및 전략수립, 약 340개의 표준화 항목을 도출을 한다.

ICT 표준화전략맵 Ver.2021에서는 스마트헬스 기술의 정의로 블록체인, IoT, 5G 등 ICT 기술을 기반으로 개인의 일상 건강정보 및 의료정보를 연결하여 질병의 예방, 상태파악, 진단, 건강관리 등 맞춤형 보건의료서비스를 제공하기 위한 기술들로 정의되며, 기존의 오프라인 및 병원중심의 헬스케어 서비스를 뛰어넘어 보건, 복지를 한 단계 업그레이드 시켜 건강한 삶을 보장할 수 있는 IT기반 헬스케어 서비스를 위한 기술로 정의하였다. 이러한 스마트 헬스 분야의 국제 공식/사실 표준화 기구에 대응하기 위한 국내 포럼, 기관들의 역할을 명시했다. 또한, 국제 표준화 대응방안, 국내 표준화 추진계획, IPR 확보 가능분야 및 확보 방안, 기술개발-표준화-IPR연계 방안등의 전략을 제시했다. 본 논문에서는 인공지능을 적용한 의료영상 데이터에 대한 국제표준화 전략에 대해 살펴보고자 한다.



(그림 1 ICT 표준화전략맵 Ver.2021 스마트헬스 기술 개요)

Ⅱ. 본론

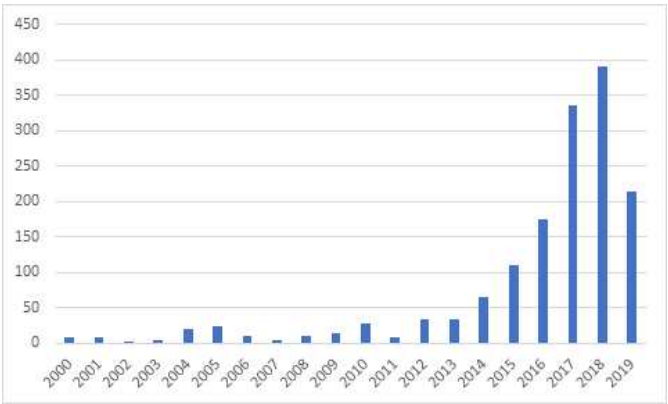
Ver.2021 ICT 표준화전략맵 스마트헬스에서는 인공지능 기술 적용을 위한 의료영상 데이터 표준을 중점표준화항목으로 선정하여 이에 대한 전략을 수립하였다. 인공지능 기술 적용을 위한 의료영상 데이

터 표준은 의료정보와 관련하여 ISO TC215, HL7, DICOM, LOINIC, SNOMED, IoMT 등 다양한 표준화기구에서 오랫동안 표준화 작업을 진행해하고 있다. 이와 관련하여 표준화기구로 ISO, IEC JTC1 SC42에서 용어 정의, 서비스 유즈케이스 개발, 인공지능 기술 특성과 특징 연구 등에 대해 국제표준화 추진을 진행하고 있으며, 한국을 포함한 22개 국가가 활발히 참여 중이다.

의료 인공지능 분야 표준화의 경우 자동진단보조 시스템 및 측정 장비 등과 관련된 표준이 이루어졌으며, 인공지능 기술 적용을 위한 의료데이터 구축 관련한 표준은 미비한 실정이다. 의료데이터 구축 관련된 대표적인 개발 사업으로, 서울아산병원, 분당서울대병원, 아주대병원 등 국내 병원은 산자부 지원의 'CDM 기반 데이터 통합 구축' 사업을 진행하며 데이터 표준화 사업을 진행 중이다.

의료정보와 관련해서 ISO, TC215 등 다양한 표준화기구에서 오랫동안 표준화 작업을 진행하고 있지만, 의료 인공지능 데이터 분야 관련된 국제 표준 역시 미비하며 표준을 만들기 위한 초기 작업 중 이다.

인공지능 기술 적용을 위한 의료영상 데이터 표준 IPR 동향 분석을 위해 의료영상, 인공지능, 딥러닝, DICOM, 의료데이터 구축과 같은 핵심키워드를 사용하여 42개의 검색식을 도출하여 분석을 진행하였다. 세계적으로 2010년부터 2020년까지 총 1,501건의 특허가 출원되었으며, 2014년까지 두자리 수의 특허 출원이 지속되었으나, 2015년 이후 특허출원이 급속이 증가하여 2018년에는 390건까지 특허가 출원되었다. 2010년부터 3년 단위 구간으로 나누어 볼 때 2010년~2012년 총 68건, 2013년~2015년 총 209건 그리고 2016년~2018년 총 901건이다. 우리나라는 총 179건의 특허를 출원하였고, 전체의 12%를 차지하고 있다. 스마트헬스 분야 중 인공지능을 기술을 적용한 스마트헬스 분야에서의 활용이 많음에 따라 특허 출원 양이 최근 활발함을 알 수 있다.



인공지능기반기술 관련한 국내 표준화 연구가 추진되고 있으며, 인공지능 시스템의 프레임워크, 연산 인터페이스, 활용 기술, 서비스 유즈케이스, 의료영상 데이터 공개 가이드라인 등의 표준 개발을 진행중으로 점차 인공지능 기술을 적용하기 위한 데이터 방법 및 시스템 구축 설계 등의 표준화가 이루어질 것으로 전망된다. TTA 인공지능기반기술 PG(PG1005)에서는 의료영상 DICOM 데이터에서의 인공지능 기술을 활용하기 위한 가이드라인으로, DICOM 영상 학습데이터의 명세 및 검수 등의 내용으로 2019년 9월에 표준이 제안 진행 중에 있다.

또한, 국제표준화기구에서는 의료 인공지능과 관련하여 2017년 JTC1 SC42가 신설되었고, 2018년 5월에 첫 총회를 중국에서 개최하였다. 일반적인 인공지능 시스템과 관련된 전반적인 분야의 표준 개발 목적으로 신설되었으며, 아직 의료 인공지능과 관련된 국제 표준은 전무하며 표준을 만들기 위한 초기 준비 작업을 진행 중이다.

Ⅲ. 결론

본 논문에서는 스마트헬스 분야의 인공지능 기술 적용을 위한 의료영상 데이터 표준화 항목에 대해 살펴보았다. 인공지능 기술의 발달로 인해 국제표준화기구인 JTC1 SC42가 신규로 설립되어 인공지능 기반 표준들을 개발하고 있으며, ITU-T FH-AI4H에서는 의료 인공지능의 활용사례들에 대해서 수집하면서 논의를 진행 중. 의료 인공지능 자체에 대한 표준은 개발 중이며, 특히 인공지능 기술 적용을 위한 데이터 구축 기술에 대한 표준은 미비한 실정임. 이에 과제 승인이 되어 개발을 준비, 진행하는 단계이다. 국내는 TTA 스마트헬스 PG(PG1005)를 통해 의료영상 딥러닝 학습 데이터 구축 표준에 대해 과제 승인 후 개발 중에 있다.

이처럼, 스마트헬스 분야의 시장규모가 계속 커지는 만큼 국내·외 기업 간 특허 경쟁력을 가지기 위해 국내 기업들의 해외특허권 확보가 필요하며, 스마트헬스 표준화기구인 TTA 스마트헬스 PG(PG1005)를 통해, 인공지능 기술 육성 및 활성화를 위한 의료영상 공개 가이드라인 표준에 대해 아이템을 발굴하고 표준화를 진행할 필요가 있다. TTA 표준화전략맵의 국제 표준화 전략을 통해 이러한 시장 요구에 맞춘 국제표준 대응과 개발이 필요하며 향후 중점 표준화 항목 전략에 대한 추적/조사에 대한 연구도 필요할 것으로 생각된다.

ACKNOWLEDGMENT

본 논문은 2020년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2017-0-00059, ICT 표준화 체계 분석 및 전략 연구)

참 고 문 헌

[1] 식약처, "빅데이터 및 인공지능기술이 적용된 의료기기의 허가·심사 가이드라인", 2017.12

[2] 식품의약안전처, "2019년 식품의약품 안전백서", 2019.06.

[3] 지식산업정보원, "인공지능 클라우드 딥러닝 기술 동향 및 ICT 융합 헬스케어 산업 실태 분석", 2017.11.

[4] TTA, Ver2021. ICT 표준화전략맵, 2020.12

[5] 보건복지부, " K-방역 3T (Test-Trace-Treat) 국제표준화 추진전략", 보도자료, 2020.06.

지능형 복합환경제어기 기반 토마토 병해 영상 분류시스템 설계

김태현, 이재수, 백정현, 최인찬, 곽강수, 김준용*
농촌진흥청 국립농업과학원, 서울대학교*
thkim8205@korea.kr, butiman@korea.kr, butterfy@korea.kr, inchchoi@korea.kr, kskwak@korea.kr, tombraid@snu.ac.kr*

A Design of Tomato Disease Classification Based on Artificial Intelligence Multiplex Environmental Control System

Taehyun Kim, Jaesu Lee, Jeonghyun Baek, Inchan Choi, Kangsu Kwak, Junyong Kim*
National Institute of Agricultural Sciences, *Seoul National Univ.

요 약

본 논문은 표준 기반의 인공지능 복합환경제어 시스템을 이용해 작물 병해영상을 자동 취득하고 환경 정보에 맞게 병해 영상 정보를 증강하여 딥러닝 모델을 활용해 분류 및 피드백하는 시스템을 설계하였다. 제안하는 방법은 온실 내 산란광을 측정해 광량에 따라 영상에 발생할 수 있는 손실을 보정한 데이터를 증강 시키고 모델을 통해 분류한 뒤 정확도 결과 값을 비교함으로써 정확도가 높은 쪽으로 손실 보정이 가능하도록 피드백 하는 시스템을 설계하였다.

Ⅰ. 서론

최근 인공지능 모델의 발전에 힘입어 온실 내의 작물의 생체정보와 환경 정보를 융합하여 병해 진단과 작물 관리에 적용하여 온실 운영 및 작물 생산의 효율성 증대와 관련한 연구가 많이 진행되고 있다. 이 중 온실 내의 작물 생산 관련 농업 분야의 경우 질병 검출 및 분류[1], 최적 환경 조건을 찾아내는 작물의 표현체 분석[2], 재배 현황 분석을 위한 환경 정보 메타데이터 생성[3] 등을 머신러닝 기법을 이용하여 수행함으로써 실시간으로 환경과 작물의 상호작용을 통한 피드백과 보상 관계 등을 통해 생산력과 이윤의 증대를 추구하고 있다. 병해에 의한 농작물 피해를 최소화하기 위한 작업의 자동화를 위해서는 작물 영상 및 환경 정보를 자동으로 취득하는 시스템과 분류 작업을 수행할 수 있는 진단 모델이 있어야 한다. 현재 영상 분류 작업에는 CNN(Convolution Neural Network)[4] 모델이 좋은 성능을 나타내어 널리 사용되고 있다. 작물의 질병을 분류하는 문제를 사람이 파악하여 해결한다고 가정했을 때, 질병의 병징이 나타난 위치 정보와 더불어 해당 병징이 나타난 환경 정보를 활용하는 것이 좀 더 명확하게 분류하는데 도움이 될 수 있다. 예를 들어 영상의 경우 CNN 분석을 할 경우 GAP(Global Average Pooling)을 수행함으로써 특징 맵의 평균값을 구하는데 이 경우 공간 정보의 평균값을 사용하기 때문에 원본(RAW) 영상의 병징의 경계가 얼마나 명확하게 구분 되냐에 따라 병징의 검출 성능을 많이 좌우하게 된다[5]. 이를 위해 애초에 학습시킬 때 되도록 해당 병징의 많은 원본(RAW) 영상을 확보하려고 하고 이를 학습하는 모델이 다양한 학습을 할 수 있도록 데이터 증강을 통해 모델의 검출 및 분류 정확도를 높이려고 하지만 실증을 해보면 모델을 학습시킨 사진의 온실 환경과 그 외의 온실의 환경에 따라 같은 병해라도 검출 정확도에는 많은 차이가 나타나는 것이 사실이다. 본 논문에서는 이러한 부분에 착안하여 환경에 따른 영상 차이를 극복하기 위한 방안으로 이동식 영상 장치를 통해 일정시간마다 실시간으로 취득한 영상데이터와 연동되는 인공지능 기반의 복합환경제어시스템을 이용해 작물의 영상을 실시간으로 변형, 증강 및 결과 피드백을 통해 작물의 병해 영상 분류 정확도를 높이는 시스템을 설계하였다.

Ⅱ. 본론

본 논문에서는 병해 데이터를 수집할 때 온실 정보를 함께 수집하여 온실 정보를 토대로 영상의 병해 데이터를 변형, 증강시켜 분류하는 시스템을 고안하였다. 제안 된 설계안은 병해 영상 자동 취득 장치와 연동되는 인공지능 모델이 탑재 된 표준 기반의 복합환경제어 시스템을 활용하여 실시간으로 영상과 환경 데이터를 취득하여 분석하는 시스템이며, 이를 위해 Faster RCNN 기반 병해 영상 분류기와 작물 영상 취득 장치, 아두이노 기반의 레퍼런스 보드를 활용하여 센서노드와 인공지능 기반 복합환경제어시스템을 올린 시범장치를 구성하였다. 영상장치는 농촌진흥청 스마트팜개발과에서 개발 된 작물 영상 취득 장치를 활용하였다.

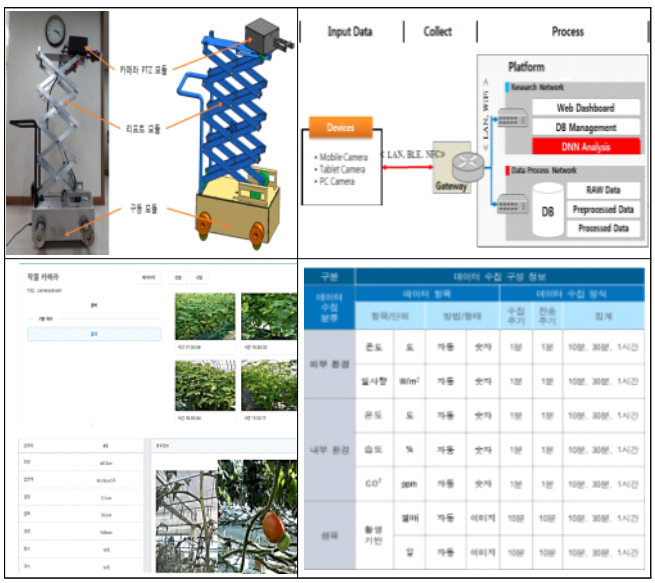


그림 1. 영상 취득 장치 및 데이터 수집 설계

본 연구에서 사용 된 모델의 기본 딥러닝 구조는 VGG-16 feature extractor의 Faster R-CNN 구조를 사용한다. 이 Faster R-CNN은 CNN

backbone, ROI pooling layer 및 fully connected layer로 구성되며, 분류 및 바운딩 박스 regression을 위한 두 개의 브랜치가 있다. Region Proposal Network(RPN)는 backbone convolution neural network에 이미지가 입력되어 실행되며, CNN backbone에서 출력되는 feature map의 모든 지점에 대해 네트워크는 해당 위치의 입력 이미지에 개체가 있는지를 학습하고 크기를 추정해야 한다. RPN의 바운딩 박스 제안은 backbone feature map에서 ROI 풀링 계층에 의해 feature를 pooling하는 데 사용되며, ROI 풀링 레이어는 기본적으로 a) backbone feature map의 제안에 해당하는 영역을 선택하고, b) 이 해당 영역을 고정된 수의 하위 window로 나누기, c) 고정된 크기 출력을 제공하기 위해 하위 window을 통해 max pooling을 수행한다. 현재 구현 된 모델은 토마토 병해에 대해 궤양병, 잎곰팡이, 잿빛곰팡이, 흰가루병, 황화잎말림바이러스에 대해 검출이 가능하다[6].

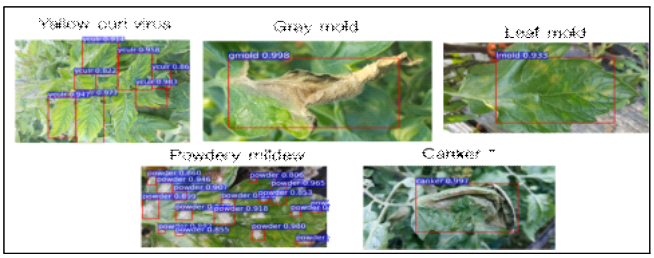


그림 2. 토마토 병해 진단 모델

해당 연구에서는 병해 영상 장치를 인식 하기 위해 표준 기반 복합환경제어시스템에 현행 인공지능 모델을 탑재해서 온실 환경 정보와 병해영상 데이터를 처리하여 데이터 변형, 증강 및 피드백을 통해 미검출 데이터 검출 및 병해 진단 분류 정확도를 향상시킬 수 있는 시스템을 고안하였다. 본 연구 설계에서 고려 된 복합환경제어시스템은 장치 간 호환을 위한 KS X 3267과 TTAK.KO-10.1172를 준수하는 표준 기반 인터페이스를 채택해 영상 장치를 PnP 방식으로 인식하고 python 코드로 구현 된 CNN 모델을 탑재하여 환경 수집 및 영상 분석이 가능한 오픈 소스 기반 시스템으로 아두이노 환경에 4채널 릴레이 모듈과 센서노드를 탑재하였다.

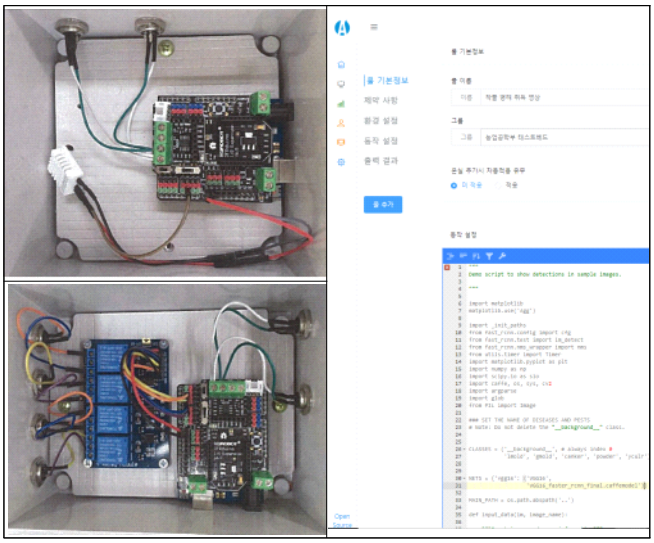


그림 3. 영상분석장치

상기 장치를 이용해 취득한 영상을 보정하기 위한 광 데이터는 일사량(W/㎡) 기준으로 측정하였으며, 날씨에 따른 광환경 분석을 위해 일사량 추정식[6]을 활용하였다. 영상 데이터는 기본적으로 RAW를 기준으로 분석 후 10초 이내로 검출이 안될 경우 또는 검출 결과 피드백에서 유사도가 50% 이하로 나타날 경우 광 환경에 따른 일사량 추정식을 토대로 3단계(청천공, 부분 담천공, 담천공) 영상 밝기 조절 및 90, 180, 270도 회전, 상하, 좌우 반전 등 영상 당 최대 8개의 증강 데이터를 활용해 검출 또는 피드백 데이터를 기준으로 10% 이상 유사도 향상이 있는지를 검증할 수 있도록 설계하였다.

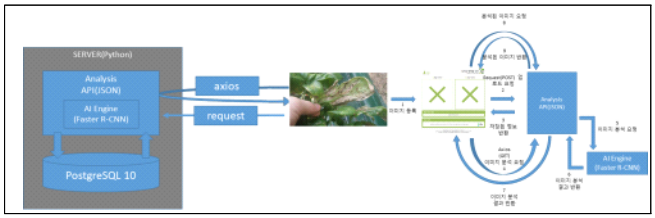


그림 4. 병해진단 검증 설계

Ⅲ. 결론

본 논문에서는 이동식 영상 장치를 이용하여 영상을 취득하고 영상과 환경정보를 복합적으로 처리 가능한 레퍼런스 센서노드를 포함한 복합환경 제어시스템을 구성하고 제어시스템 내에 인공지능 분류 모델을 통해 실시간으로 환경 변화에 맞게 증강 된 영상 데이터를 분류하고 피드백 함으로써 광 환경에 따른 영상 손실로 인해 병해가 미분류 되는 것을 방지하기 위한 시스템을 설계했다. 제안 된 설계안은 병해 영상 자동 취득 장치와 연동되는 인공지능 모델이 탑재 된 표준 기반의 복합환경제어 시스템을 활용하여 실시간으로 영상과 환경 데이터를 취득하여 분석하는 시스템을 구성하였다. 향후 연구에서는 병해 분류 및 예찰까지 수행할 수 있도록 온습도 등의 환경 정보와 병의 징후에 대한 문헌 정보까지 포함하여 분석할 수 있는 시스템을 구축하고자 한다.

ACKNOWLEDGMENT

본 연구는 농림축산식품부의 재원으로 농림식품기술기획평가원의 1세대 스마트팜 산업화 기술개발사업의 지원을 받아 연구되었음 (과제번호320085-01)

참 고 문 헌

[1] A. F. Fuentes, S. Yoon, J. Lee, and D. S. Park, “High-performance deep neural network based tomato plant diseases and pests diagnosis system with refinement filter bank,” Frontiers in Plant Sci. , vol. 9, Aug. 2018.

[2] F. Fiorani and U. Schurr, “Future scenarios for plant phenotyping,” Annu. Rev. Plant Biology , vol. 64, pp. 267-291, Apr. 2013.

[3] P. L. Suarez, A. D. Sappa, and B. X. Vintimilla, “Learning image vegetation index through a conditional generative adversarial network,” 2017 IEEE ETCM, pp. 1-6, Oct. 2017.

[4] C. Szegedy, et al., “Inception-v4, inceptionresnet and the impact of residual connections on learning,” Thirty-first AAAI Conf. Artificial Intell., pp. 4278-4284, San Francisco, USA, Feb. 2017.

[5] M. Lin, Q. Chen, and S. Yan, “Network in network,” arXiv preprint arXiv:1312.4400 , 2013.

[6] 이준환, 김태현, 박종한, 이준용, “토마토 병해진단 웹 UI 고도화 및 전문가활용시스템 구축에 관한 연구”, 농업기술기회평가원 1세대 스마트 플랜트팜 고도화 및 실증 과제 완결보고서, 2020.07.

[7] 장성택, 장성주, “자연광 다층 작물재배를 위한 광선반 시스템에 관한 연구”, Journal of the Korea Institute of Ecological Architecture and Environment v.13 no.2, pp.61-66, 2013.

뇌졸중 병변 분할을 위한 효율적인 U-Net

신현광, 최규상
영남대학교
shg3786@ynu.ac.kr, castchoi@ynu.ac.kr

e-UNet: Efficient U-Net for Brain Stroke Lesion Segmentation

Hyunkwang Shin, Gyu Sang Choi
Yeungnam Univ.

요 약

뇌졸중은 전 세계적으로 가장 흔한 신경학적 질환이며, 일반적으로 방사선 전문의가 직접 뇌졸중 병변을 분할한다. 이는 많은 시간이 소요되며, 전문의의 주관적 인식에 의존하게 되기 때문에 뇌졸중 병변을 자동으로 분할하는 연구가 중요하다. 최근 딥 러닝 발전과 함께 컨볼루션 네트워크 기반의 이미지 분할 연구가 많이 진행되고 있지만, 많은 파라미터 수와 장기 의존성 문제가 존재한다. 이 문제를 해결하기 위해 기존의 U-Net 구조에 e-block과 Non-local block을 적용한 e-UNet(Efficient UNet)을 제안한다. e-UNet의 성능을 평가하기 위해 ATLAS(Anatomical Tracings of Lesions After Stroke) 데이터를 활용했으며, 기존의 SegNet, U-Net, 2D Dense-UNet 보다 더 나은 성능을 달성했다.

Ⅰ. 서 론

뇌졸중은 뇌의 혈관이 터지거나 막힘으로써 해당 부위가 손상되어 발병되며, 전 세계적으로 두 번째 사망원인이다[1]. 일반적으로 뇌졸중 병변은 방사선 전문의가 직접 MR 이미지에 포함된 병변을 분할한다. 이는 많은 시간이 소요되며, 전문의의 주관적 인식에 의존하게 되기 때문에 뇌 병변 분할을 위한 자동화 연구가 필요하다[2]. 뇌졸중 병변은 그림 1과 같이 모양과 경계가 명확하지 않으며, 다양한 위치에서 발생 된다.

최근 몇 년간 딥 러닝 발전과 함께 컨볼루션 네트워크 기반의 이미지 분할 연구가 진행되고 있으며, 대표적인 모델로 SegNet[3], U-Net[4], 2D Dense-UNet[5] 등이 있다. U-Net은 의료 영상 분할의 대표 모델로 인코더와 디코더 구조에 skip connection을 결합한 구조로 이루어져 있다. 하지만, 고정된 컨볼루션 크기와 다운 샘플링에 의해 로컬 수용 필드 및 특징 정보 재사용이 제한됨으로 크기 및 경계가 명확하지 않아 뇌 병변 분할 결과에 영향을 미친다. 또한, 기존의 이미지 분할 모델들은 많은 학습 파라미터들이 요구된다.

본 논문에서는 위에서 언급한 문제를 해결하기 위해 e-UNet 모델을 제안한다. e-UNet은 기존의 U-Net에 사용되는 컨볼루션을 깊이별 분리 컨볼루션(Depthwise Separable Convolution; DSC)[6] 기반의 e-block으로 대체함으로써, 학습 파라미터를 줄인다. 또한, Non-local block[7]을 통해 장기 의존성 문제 해결하며, 기존의 SegNet, U-Net, 2D Dense-UNet 모델과 성능을 비교한다.

Ⅱ. e-UNet

본 논문에서는 뇌졸중 병변 분할을 위한 e-UNet을 제안하며, 구조는 그림 2와 같다. e-UNet은 기존 U-Net 구조에서 컨볼루션 연산과 브릿지(brige) 위치에 e-block 및 Non-local block이 적용된다. 기존 U-Net은 각 계층에서 두 번의 3×3 컨볼루션 연산을 수행하며, $2(K^2CM)$ 파라미터 수가 요구된다. K는 필터의 크기, C와M은 입력 채널 수, 필터의 개수를 의미한다.

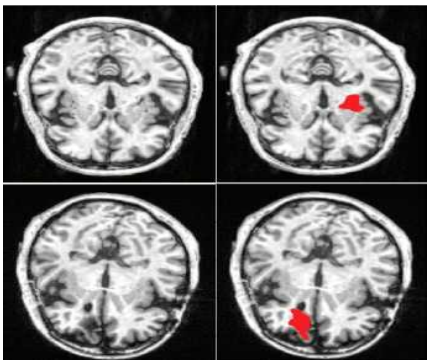


그림 1. 뇌졸중 병변 분할의 예

제안하는 e-block은 1×1 컨볼루션을 사용하여 필터 수를 결정하고, 두 번의 깊이별 컨볼루션(Depth-wise Convolution)과 1×1 컨볼루션을 통해 나온 결과와 합치게 된다. 첫 번째 연산은 3×3 깊이별 컨볼루션을 통해 공간적 특징을 추출하고, 1×1 컨볼루션을 통해 출력 채널의 수를 2배 늘린다. 그 후, 다시 3×3 깊이별 컨볼루션과 1×1 컨볼루션을 통해 입력 채널 수와 동일하게 출력 채널을 맞춘다. 제안하는 e-block의 파라미터 수는 $CM+C(K^2+2M)+2C(K^2+M)$이며, 각 계층은 $C(3K^2+5M)$ 만큼의 파라미터 수를 가지게 된다. 따라서, 각 계층의 두 파라미터 수 차이는 $e-block$ 파라미터 수/기존 컨볼루션 파라미터 수 $= 3/2M+5/2K^2$ 배 차이난다.

Ⅲ. 실험 및 결과

3.1 데이터셋

뇌졸중 병변 분할에 대한 e-UNet의 성능을 평가하기 위해 ATLAS(Anatomical Tracings of Lesions After Stroke) 데이터[2]를 사

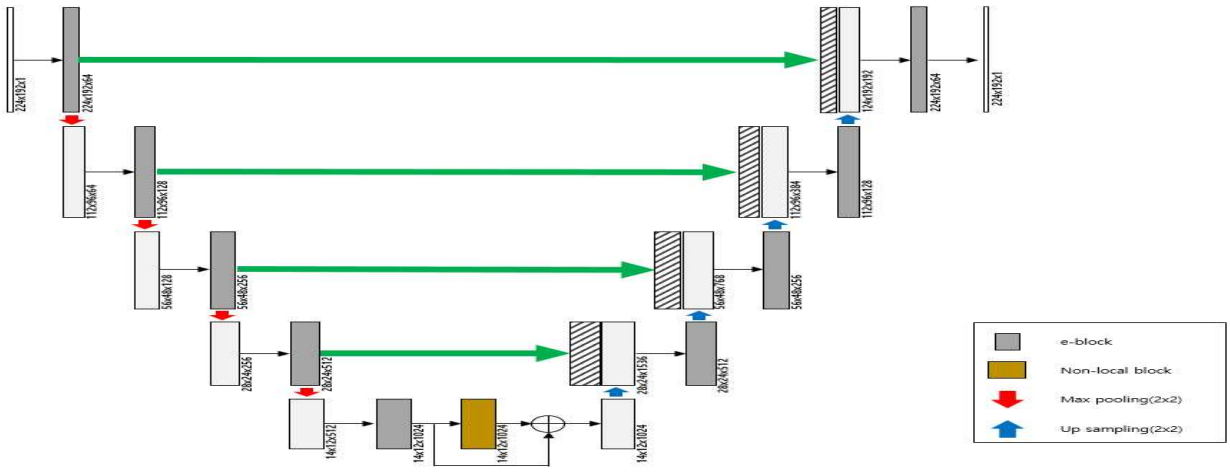


그림 2. e-UNet 구조

용했다. ATLAS는 다양한 병변이 포함된 229개의 정규화된 T1-가중치 (T1-weighted) 3D MRI로 구성된다. 각 3D 이미지는 189개의 슬라이스로 구성되며, 233×197 이미지에서 224×192영역만 네트워크의 입력으로 사용한다.

3.2 실험 결과

본 논문에는 5-교차 검증(5-fold cross validation)을 통해 제안하는 모델을 검증하며, 평가지표로 Dice score, 정밀도, 재현율을 사용한다. 표 1은 2D Dens-Unet, U-Net, SegNet 모델과 제안하는 e-UNet 간의 성능을 비교한 결과이며, 기존 모델에 비해 Dice score와 정밀도가 향상된 것을 확인할 수 있다. 또한, e-UNet의 파라미터 수는 11.5M로 기존 모델보다 적은 파라미터 수와 연산량으로 모델을 학습할 수 있다.

표 1. 성능 결과

	Dice	정밀도	재현율	파라미터 수
2D Dense-UNet	0.4741	0.5613	**0.4875**	41.3M
SegNet	0.277	0.394	0.2532	29.5M
U-Net	0.461	0.599	0.445	34.5M
e-UNet	**0.480**	**0.624**	0.451	**11.5M**

Ⅳ.결론

본 논문에서는 뇌졸중 분할을 위해 e-UNet 모델을 제안했으며, e-block을 통해 학습 가능한 파라미터의 수를 효율적으로 줄였다. 또한, Non-local 블록을 통해 장기 의존성 문제를 해결했으며, ATLAS 데이터를 통해 제안 모델이 기존 모델보다 적은 파라미터 수와 뇌졸중 분할에 우수함을 검증했다.

ACKNOWLEDGMENT

이 논문은 2019년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구(2019R1A2C1006159)이며, 미래창조과학부 및 정보통신기술진흥센터의 대학ICT연구센터육성 지원사업의 연구결과로 수행되었으며 (IITP-2020-2016-0-00313), 2016년도 산업통상자원부 및 산업기술평가관리원(KEIT) 연구비 지원에 의한 연구임(No.10063130, 익스트림 트랜잭션 및 분산 확장성을 제공하는 대용량 비휘발성 메모리(SCM) 기반의 차세대 인메모리 빅 데이터베이스 시스템 상용 기술 개발)

참 고 문 헌

[1] Johnson W., Onuma O., Owolabi M., et al.: Stroke a global response is needed. Bulletin of the World Health Organization 94(9), 634 (2016)

[2] Liew, S. L., Anglin, J. M., Banks, N. W., et al.: A large, open-source dataset of stroke anatomical brain images and manual lesion segmentations. Scientific data (2018)

[3] Badrinarayanan V., Kendall A., Cipolla R.: Segnet: A deep convolutional encoderdecoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence 39(12), pp. 2481 - 2495 (2017)

[4] Ronneberger O., Fischer P., Brox T.: U-net: Con- volutional networks for biomedical image segme- ntation. In: International Conference on Medical Image Computing and Computer-Assisted Inter- vention. pp. 234 - 241 (2015)

[5] Li X., Chen H., Qi X., et al.: H-denseunet: Hybrid densely connected unet for liver and tumor segme- ntation from ct volumes. IEEE Transactions on Me- dical Imaging 37(12), 2663 - 2674 (2018)

[6] Chollet F.: Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1251 - 1258 (2017)

[7] Cao, Y., Xu, J., Lin, S., Wei, F., & Hu, H. Gcnet: Non-local networks meet squeeze-excitation networks and beyond. In Proceedings of the IEEE International Conference on Computer Vision Workshops. pp. 0-0 (2019)

Octave U-net 을 이용한 생체 의학 이미지 분할과 분류 방법

김화랑 [1,2], 김광주 [2], 임길택 [2], 최두현 [1]
[1] 경북대학교, [2] 한국전자통신연구원

Khr1393@knu.ac.kr, kwangju@etri.re.kr, ktl@etri.re.kr, dhc@ee.knu.ac.kr

Biomedical Image Segmentation and Classification Using Octave U-net

Hwa-Rang Kim[1,2], Kwang-Ju Kim[2], Kil-Taek Lim[2], Doo-Hyun Choi[1]
[1] Kyungpook National University, [2] Electronics and Telecommunications Research Institute

요 약

생체의학 이미지를 이용한 환자의 질병 분석이나 예측은 병리학자에 의한 분석과 함께 추가적인 검사를 필요로 하는 경우가 자주 발생한다. 특히 암과 같은 심각한 질병에 대해서는 다양한 추가적인 절차의 검사가 필요한데, 이는 많은 인력과 비용을 요구한다. 본 논문에서는 인공지능을 이용한 질병 검사 분야 중 생체 의학 이미지 분석 분야에 U-net 기반의 대장암 이미지 분할 및 분석 방법을 제안한다. 기존 U-net 을 사용한 모델의 성능향상을 위해 합성곱 신경망을 옥타브 합성곱 신경망으로 대체하였다. 옥타브 합성곱 신경망은 고주파와 저주파 성분에 대한 연산에 차별을 두는 모듈로써, 고주파에 대해서는 큰 특성맵으로, 저주파에 대해서는 고주파 특성맵의 반의 크기를 가지는 특성맵으로 연산한다. 본 논문에서 제안한 Octave U-net 을 이용한 생체 의학 이미지 분할 및 분류 성능은 0.775 F1 score 와 0.803 Jaccard score 로 기존의 U-net, U-net++, DeepLabV3+ 보다 F1 score 는 각각 0.172, 0.143, 0.070 만큼, Jaccard score 는 각각 0.126, 0.078, 0.040 만큼 성능 향상이 있었다.

Ⅰ. 서 론

대장암은 암으로 인한 사망률에 크게 관여하는 치명적인 병이다. 대장암에 대한 면역 치료에 대해 환자가 잘 반응하는지에 대한 지표로 MSI (Microsatellite instability)가 사용되는데, MSI 는 MSI-High(MSI-H)와 MSI-Low(MSI-L), MSI-Stable(MSI-S)로 분류된다. MSI-H 는 좋은 예후 증상을 나타냄을 의미하며, MSI 상태에 대한 병리학적인 보고는 수술로 절제된 모든 대장암 사례에 대해 강력히 권장된다.

하지만 실제로는 추가적인 유전자 검사와 면역 조직화학 검사를 받아야 하기 때문에 모든 환자들이 MSI 테스트를 받기는 어렵다. 이에 대한 해결책으로 MSI 예측 알고리즘을 개발하여 추가적인 비용이나 시간 지불 없이 슬라이드 이미지만 이용하여 예후 증상을 예측할 수 있도록 할 수 있다. 예측 알고리즘을 위하여 통계적, 혹은 수학적 접근법을 적용할 수 있겠지만, 많은 영역에서 인간의 수학적, 과학적 지식을 압도하는 딥러닝의 발전 이후, 많은 연구자들은 딥러닝을 이용한 알고리즘 개발에 몰두해 왔다.

본 논문에서도 딥러닝을 이용한 생체의학 이미지 분석을 한다. 많은 딥러닝 알고리즘 중에서도 생체의학 이미지에서 뛰어난 성능을 보였던 U-net [1]은 간단한 U 자 모양의 합성곱 신경망 구조를 이용하여 많은 연구자들의 관심을 받았다. 실제로 U-net 보다 몇몇 데이터셋에 대해서 뛰어난 성능을 보인 다른 모델들(DeepLabV3+ [2], U-net++ [3] 등)이 있지만 본 실험에서는 비교적 간단한 구조인 U-net 을 변형하여 몇몇 모델들보다 우수한 성능을 낼 수 있음을 확인하였다. U-net 의 변형은 Octave 합성곱 신경망[4]를 이용하였으며, 옥타브 합성곱 신경망은 기존 합성곱 신경망보다 가벼우며, 효율적인 연산을 통해 성능 향상을 할 수 있는 모듈이다. 본 실험에서는 대장암 이미지 데이터셋에 대해 U-net, U-net++, DeepLabV3+ 와 제안하는 Octave U-net 의 성능을 비교하였다.

Ⅱ. 본론

본 실험에서는 대장암 이미지 데이터셋에 대해 종양 부위를 분할하고, 분할된 종양 부위에 대해 MSI 상태를 예측하는 분류 알고리즘 개발을 목표로 한다.

그림 1 은 본 실험에서 사용한 옥타브 합성곱 신경망의 특징맵을 나타낸다. 기존의 합성곱 신경망과는 달리 옥타브 합성곱 신경망은 특징맵의 반은 크기를 반으로 하여, 저주파에 대해 글로벌한 특징을 해석하며, 고주파에 대해 미세한 부분에 대한 해석을 한다. 그림 2 는 본 실험에서 사용한 U-net 구조이며, 기존의 합성곱 신경망이 옥타브 합성곱 신경망으로 대체된 모델이다.

테이블 1 에서 각 모델 별 성능 비교를 하였다. 본 실험에서는 분류 평가 기준으로 F1 score 를 사용하였고, 분할 평가 기준으로 Jaccard score 를 사용하였다. 테이블 1 에서 볼 수 있듯이 본래의 U-net 이나 U-

net++구조, DeepLabV3+ 구조보다 뛰어난 성능을 보였다.

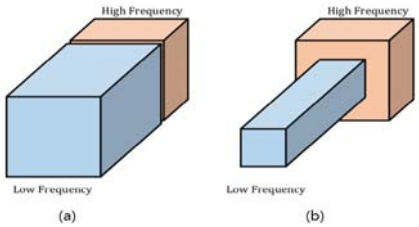


그림 1. (a) 기존의 합성곱 신경망 특징맵 (b) 옥타브 합성곱 신경망의 특징맵

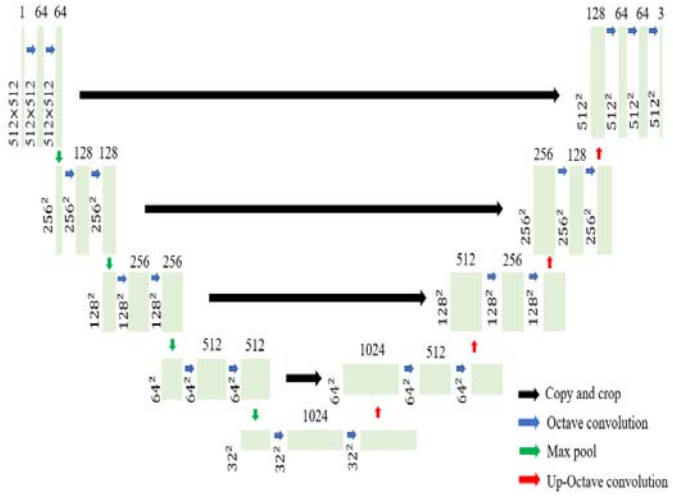


그림 2. 본 실험에서 사용된 Octave U-net 구조

테이블 1. 각 모델 별 성능 비교

모델	F1 Score	Jaccard Score
U-net	0.603	0.677
U-net++	0.632	0.725
DeepLabV3+	0.705	0.763
Octave U-net	0.775	0.803

그림 3 은 대장암 이미지 데이터에 대한 Octave U-net 의 분할 결과를 보여준다. 하얀색 영역이 종양 부위를 나타내며, 종양 부위는 다시 MSI 상태로 분류된다. 한 이미지에 대해 여러 MSI 상태는 나타나지 않으며, MSI-H 는 1, MSI-S 나 MSI-L 는 0 으로 구분된다. 본 실험에서 사용한 대장암 이미지 데이터는 총 6,746 장이며, 5,000 장이 훈련 데이터셋, 1,746 장이 테스트 데이터셋으로 쓰였다.

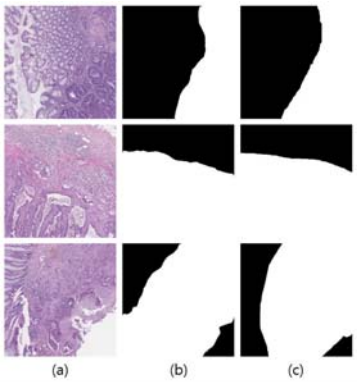


그림 3. 대장암 이미지 분할 결과 (a) 대장 이미지 (b) 예측 결과 (c) 정답 마스크

Ⅲ. 결론

본 논문에서는 대장암 이미지의 분석을 위해 U-net 과 옥타브 합성곱 신경망을 조합하여 모델을 구축하고 실험하였다. 그 결과, 병리학자가 아닌 일반인의 눈으로는 구분하기 어려운 종양 부위(그림 3 참조)에 대해 상당히 비슷하게 예측해냈으며, 기존의 성능이 증명된 다른 모델보다 우수한 성능을 보였다.

ACKNOWLEDGMENT

This work was supported by Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korean government. [20ZD1200, Development of ICT Convergence Technology for Daegu-GyeongBuk Regional Industry].

De-identified pathology images and annotations used in this research were prepared and provided by the Seoul National University Hospital by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number: HI18C0316).

참 고 문 헌

[1] Ronneberger O., Fischer P., Brox T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N., Hornegger J., Wells W., Frangi A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science, vol 9351. Springer, Cham.

[2] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, Hartwig Adam, "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation," Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 801-818

[3] Zhou Z., Rahman Siddiquee M.M., Tajbakhsh N., Liang J. (2018) U-net++: A Nested U-Net Architecture for Medical Image Segmentation. In: Stoyanov D. et al. (eds) Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. DLMIA 2018, ML-CDS 2018. Lecture Notes in Computer Science, vol 11045. Springer, Cham.

[4] Chen, Yunpeng, et al. "Drop an octave: Reducing spatial redundancy in convolutional neural networks with octave convolution." Proceedings of the IEEE International Conference on Computer Vision. 2019.

폐 영역 분할에서 적응형 활성화 함수의 유효성 검증

신호경, 김재일
경북대학교
parkland106e@naver.com, threeyears@gmail.com

Validation of adaptive activation function in lung segmentation

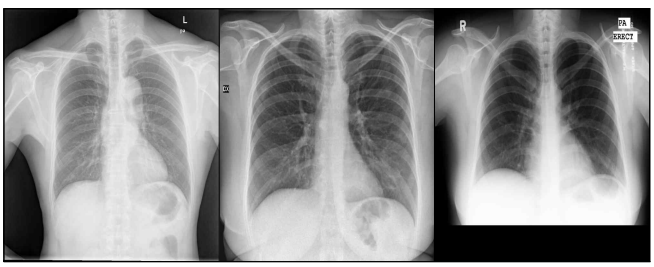
Ho Kyung Shin, Jaeil Kim
Kyungpook National Univ.

요 약

본 논문은 공개 데이터셋으로 학습한 폐 영역 분할 모델에서 적응형 활성화 함수에 따른 성능 비교를 통해 성능 향상의 유효성을 검증하였다. 폐 영역 분할을 위한 베이스라인 모델로 U-Net++[1]를 사용하였고, ReLU, ELU[2], MPELU[3], EPReLU[4], EELU[5] 다섯 가지 활성화 함수의 성능을 비교하였다. 모델 간 성능 비교를 위해 Shenzhen, Covid, Montgomery 3가지 데이터셋을 교차로 학습, 검증을 진행하였다. 성능 비교 결과 MPELU를 적용한 모델들에서 성능이 제일 높았고, MPELU와 같은 적응형 활성화 함수가 고정형 활성화 함수보다 성능이 더 높은 것을 확인하였다. 이를 통해 폐 영역 분할에서 적응형 활성화 함수 사용 시 성능 향상이 유효하다는 것을 확인하였다.

Ⅰ. 서 론

폐 영역 분할은 흉부 영상에서 신체 외곽 배경과 폐를 제외한 다른 신체 부분을 제거함으로 폐 영역을 찾는 것으로 폐 내 병변 검출 정확도 향상에 영향을 미치며 최근 차원 축소[6], U-Net모델[7] 등 딥러닝을 이용한 폐 영역 분할 기법이 활발히 연구되고 있다. 하지만 데이터에 의존적인 딥러닝의 특성에 따라 학습데이터 수집 기관마다 딥러닝 모델의 성능차이가 나타날 수 있다. 외부 데이터에 대해 성능이 떨어지는 이유는 그림 1과 같이 같은 흉부 X-Ray 영상임에도 촬영기기, 조작 방법으로 인해 기관마다 다른 영상 특징을 가지기 때문이다. 데이터셋 간 차이로 인한 모델의 성능 저하를 해결하기 위해 이미지 단위의 전처리, 데이터 어그멘테이션(Data augmentation) 기법, 모델 구조개선 등 관련 연구가 활발히 진행되고 있다.



(a) Shenzhen (b) Covid (c) Montgomery

그림1 Shenzhen, Covid, Montgomery 데이터셋 속 흉부 X-Ray 영상 간 차이

본 논문에서는 세가지 공개 데이터셋을 이용하여 폐 영역 모델에서 ReLU, ELU, MPELU, EPReLU, EELU 다섯 가지 활성화 함수의 일반화 성능을 비교하고, 적응형 활성화 함수를 통한 성능 향상의 유효성을 검증하고자 한다.

Ⅱ. 본론

딥러닝에서 활성화 함수는 입력으로부터 다음 층으로 전달할 출력를 결정하는 역할을 하고, 출력에 비선형성을 추가해 모델이 비선형적 패턴을 학습할 수 있게 해준다. 모델에 비선형성을 추가해주는 대표적인 활성화 함수인 Rectified Linear Unit(ReLU) 함수는 가중치들의 합이 음수일 때 활성화되지 않는 Dying ReLU, 음수 값에서 기울기가 소실되는 문제가 있다. 문제를 해결하기 위해 ReLU를 기반으로 한 활성화 함수들이 제안되었고, MPELU, EPReLU, EELU와 같은 적응형 활성화 함수는 학습 파라미터를 통한 활성화 함수의 학습으로 데이터에 맞추어 모델의 성능을 높이고 있다.

적응형 활성화 함수를 통한 폐 영역 분할 모델의 성능 향상의 유효성을 검증하기 위해 고정형 활성화 함수 ReLU, ELU와 적응형 활성화 함수 MPELU, EPReLU, EELU의 성능을 비교하는 실험을 하였다. 베이스라인 모델로 U-Net++를 사용해 흉부 X-Ray 영상에서 폐 영역 분할을 진행하였다. U-Net++ 모델은 Encoder와 Decoder 두 서브 네트워크와 두 서브 네트워크를 연결하는 Nested dense convolution block으로 구성된다. 활성화 함수에 따른 성능 비교를 위해 두 서브 네트워크 속 Convolutional block과 Nested dense convolution block을 구성하는 활성화 함수들을 바꿔가며 실험을 진행하였다. 실험에는 Shenzhen, Covid, Montgomery 세 가지 데이터셋에서 각 572, 226, 148장을 사용했고, 전처리로 영상에 CLAHE(Contrast Limited Adaptive Histogram Equalization) 알고리즘을 적용해 대비를 증가시켰다. 모델 간 성능 비교를 위해 Dice coefficient (수식 1)를 사용하였다.

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|} \quad (1)$$

그림 2는 활성화 함수에 따른 모델 성능 비교를 위한 실험 프로세스이다. 실험은 데이터셋을 학습, 튜닝, 테스트 데이터로 나누어 학습 데이터와 튜닝 데이터으로 모델을 학습시킨 후 테스트 데이터와 나머지 두 데이터셋

의 성능을 확인하였다.

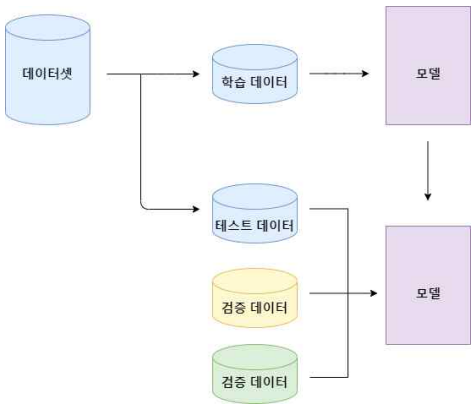


그림2 활성화 함수에 따른 일반화 성능 비교를 위한 실험 프로세스

적응형 활성화 함수들의 학습 파라미터 초기화를 바꿔가며 실험을 진행하였고, 각 데이터셋으로 학습시키고 테스트한 결과는 그림 3과 같다.

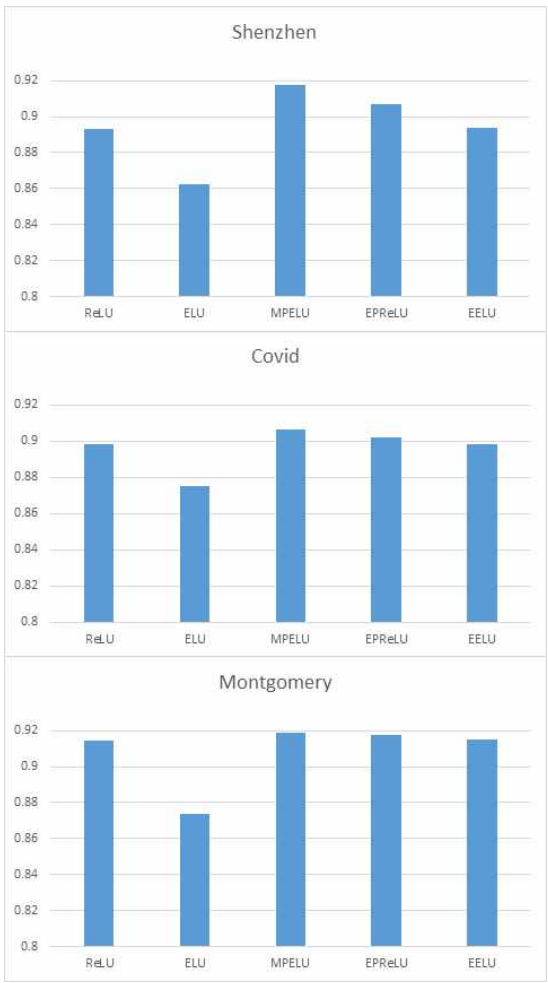


그림3 각 데이터셋으로 학습시킨 모델의 활성화 함수별 성능

실험 결과 MPELU가 각 데이터셋을 학습시킨 모델에서 평균 성능이 0.9173, 0.9067, 0.9190으로 가장 높은 성능을 보였다. 대부분의 테스트에서 적응형 활성화 함수를 사용한 모델들이 고정형 활성화 함수를 사용한 모델보다 성능이 높은 것을 확인하였다.

Ⅲ. 결론

본 논문에서는 폐 영역 분할 모델에서 적응형 활성화 함수를 통한 모델 성능 향상의 유효성 검증을 위해 다섯 가지 활성화 함수(ReLU, MPELU, EPReLU, EELU, ELU)를 비교하였다. 활성화 함수 간 성능 비교를 위해 모델에서 활성화 함수를 바꿔가며 Shenzhen, Covid, Montgomery 데이터셋을 교차로 학습, 테스트하였다. 그 결과 Montgomery 데이터셋으로 학습하고 테스트한 경우를 제외하면 나머지 모두에서 MPELU가 가장 높은 성능을 보였고, 적응형 활성화 함수가 고정형 활성화 함수보다 성능이 높은 것을 확인하였다. 이를 통해 데이터 수가 작은 폐 영역 분할 모델에서 적응형 활성화 함수 사용 시 성능 향상의 유효성을 검증하였다.

ACKNOWLEDGMENT

"This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(2020R1I1A3074639)"

참 고 문 헌

[1] Z.W. Zhou, M.M.R. Siddiquee, N. Tajbakhsh and J.M. Liang, "UNet++: A Nested U-Net Architecture for Medical Image Segmentation," Deep Learning in Medical Image Anylysis and Multimodal Learning for Clinical Decision Support, pp: 3-11, 2018

[2] D.-A. Clevert, T. Unterthiner, S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (elus)", arXiv preprint arXiv:1511.07289 (2015)

[3] Y. Li, C. Fan, Y. Li, Q. Wu, Y. Ming "Improving deep neural network with multiple parametric exponential linear units", Neurocomputing, 301 (2018), pp. 11-24.

[4] X. Jiang, Y. Pang, X. Li, J. Pan, Y. Xie "Deep neural networks with elastic rectified linear units for object recognition" Neurocomputing, 275 (2018), pp. 1132-1139

[5] D. Kim, J. Kim and J. Kim, "Elastic exponential linear units for convolutional neural networks", Neurocomputing, vol. 406, pp. 253-266, Sep. 2020

[6] Gordienko, Y., Kochura, Y., Alienin, O., Rokovyi, O., Stirenko, S., ang, P., Hui, J., Zeng, W. "Dimensionality reduction in deep learning for chest x-ray analysis of lung cancer." 10th International Conference on Advanced Computational Intelligence, Xiamen, China. arXiv preprint arXiv:1801.06495 (2018)

운전자의 얼굴을 검출하기 위한 딥러닝 얼굴 검출기와 객체 추적알고리즘 융합 시스템

유민우, 한동석*

경북대학교

dshan@knu.ac.kr

Deep learning face detector and object tracking algorithm convergence system to detect driver's face

Min Woo Yoo, Dong Seog Han*

Kyungpook National Univ.

요 약

본 논문은 운전자 모니터링 시스템에서 가장 많은 연산시간이 소비되는 운전자 얼굴 검출 알고리즘을 최적화하기 위하여 운전자 얼굴 검출 알고리즘과 객체 추적을 융합하여 고속으로 운전자의 얼굴을 검출하는 시스템에 관한 것이다. 기존의 운전자 모니터링 시스템의 얼굴 검출 방법은 딥러닝 객체 검출 알고리즘을 사용하여 얼굴을 검출한다. 하지만 딥러닝 객체 검출기는 매우 많은 연산을 소비하는 문제가 있다. 본 논문에서는 연산시간을 줄이기 위해 운전자 얼굴 검출 알고리즘과 객체 추적알고리즘을 융합하여 연산시간을 줄인다. 이러한 과정을 통해 운전자 모니터링 시스템의 얼굴 검출 연산시간을 줄여 안정적으로 운전자의 상태를 모니터링 할수 있을 것이다.

Ⅰ. 서 론

현재 정확한 검출을 위해 딥러닝 기술을 사용하여 운전자 모니터링 시스템 개발 및 실제 적용을 위해 테스트를 진행하고 있다. 운전자 모니터링 시스템이란 운전자의 얼굴을 검출한 뒤 얼굴의 특징을 분석하여 전방주시 태만이나 졸음 등을 판별하는 기술이다. 이를 구현하기 위해서는 얼굴의 특징 중 얼굴의 특징점, 시선 얼굴 방향 등이 검출되어야 한다. 얼굴 특징 중에서도 눈동자 검출은 졸음과 시선 검출에 동시에 사용되므로 정확한 검출을 해야 한다. 하지만 이러한 시스템은 실시간으로 동작하여야 하지만 딥러닝의 많은 연산자원 소비로 인해 어려움이 있다. 그중 운전자 모니터링 시스템에서 가장 많은 연산자원을 소비하는 알고리즘은 운전자의 얼굴 검출하는 알고리즘이다. 따라서 본 논문에서는 가장 많은 연산자원을 소비하는 얼굴 검출모델을 객체 추적 모델과 병합하여 최적화하는 방법을 제시한다.

Ⅱ. 본론

본 논문에서 사용한 얼굴 검출기는 SSD(single shot multibox detector)를 사용한다[1]. 얼굴 검출기를 학습하기 위해서 약 40만 개의 얼굴 데이터가 있는 Wider face 데이터를 사용했다[2]. 얼굴 추적을 하기 위해 median flow 추적기를 사용한다[3]. median flow의 객체 추적을 위해서 특징점이 필요하다. 객체 추적에서 특징점을 검출하는 방법은 랜덤한 특징점을 잡거나 shi-tomasi알고리즘을 사용하여 특징점을 검출한다. 하지만 운전자 모니터링 시스템에서는 얼굴을 검출한 뒤 얼굴의 특징점을 검출하는 것이 일반적이다. 그러므로 얼굴 추적의 특징점은 얼굴의 특징점 검출기를 사용한다. 얼굴의 특징점 검출 방법은 Dlib의 64개 얼굴 특징점 검출기를 사용했다. 그림 1은 각 알고리즘의 검출 결과를 비교한 것이다. 얼굴 검출 알고리즘보다 추적 모델을 혼합한 모델이 약 3배 빨랐다.

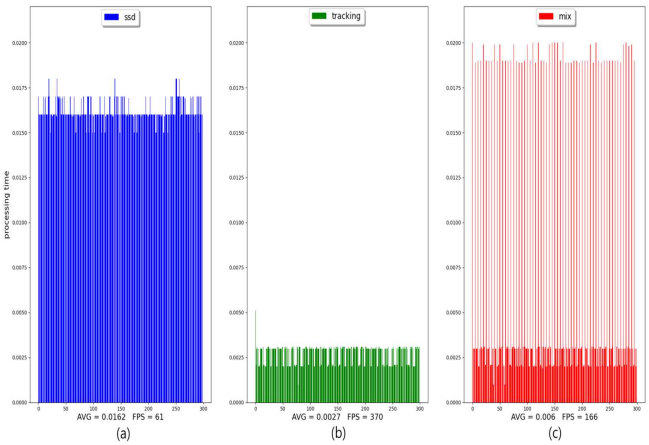


그림 1. 알고리즘별 연산시간 비교: (a) 딥러닝 얼굴 검출, (b) 추적 모델, (c) 딥러닝 얼굴 검출과 추적 모델을 혼합한 모델

Ⅲ. 결론

본 논문에서는 운전자 모니터링 시스템에서 가장 많은 연산량을 소비하는 운전자 얼굴 검출모델의 연산 속도를 줄이기 위해서 얼굴 검출모델과 객체 추적 모델을 병합하는 방법을 제시했다.

ACKNOWLEDGMENT

이 논문은 2020년도 정부(산업통상자원부)의 재원으로 “5G기반 자율주행 융합기술 실증 플랫폼” 사업의 지원을 받아 수행된 연구임(P0013840)

참 고 문 헌

[1] Liu, Wei, et al. "Ssd: Single shot multibox detector." European conference on computer vision. Springer, Cham, 2016.

[2] Yang, Shuo, et al. "Wider face: A face detection benchmark." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

[3] Kalal, Zdenek, Krystian Mikolajczyk, and Jiri Matas. "Forward-backward error: Automatic detection of tracking failures." 2010 20th International Conference on Pattern Recognition. IEEE, 2010.

YOLO 를 활용한 열화상 기반의 체온측정 인공지능 시스템

손진영, 김민영*
경북대학교, *경북대학교

sony0416@knu.ac.kr, *minykim@knu.ac.kr

Thermal image-based body temperature Measurement AI System using YOLO

Son Jin Yeong, Kim Min *
Kyungpook National Univ., * Kyungpook National Univ.

요 약

본 논문은 기존 열화상 카메라를 이용한 체온측정 방법의 한계인 방해요소에 의한 측정 문제 개선과 체온 측정 정확도 향상을 위한 온도 피팅 방법에 대하여 작성하였다. 위 문제에 대한 해결을 위하여 얼굴 및 방해요소 감지는 YOLOv4, 체온 측정은 얼굴 영역의 온도를 히스토그램을 통해 체온 측정 영역을 분리하였다. 분리된 영역에서 이루어지는 체온 측정은 온도 피팅 작업을 통하여 측정 결과를 보완하였다.

Ⅰ. 서 론

본논문에서는 기존 체온 측정 방법이 사람의 얼굴, 이마, 미간 등 위치에서 한 부분의 온도만을 측정하거나 체온 측정 영역을 설정하여 영역안의 온도를 측정하는 방법으로 이루어져 있다. [2, 3, 5, 6, 7] 위 방법들은 측정 위치가 가려지면 온도 편차가 크게 발생하고, 얼굴 감지가 안되는 문제를 발생하게 된다.

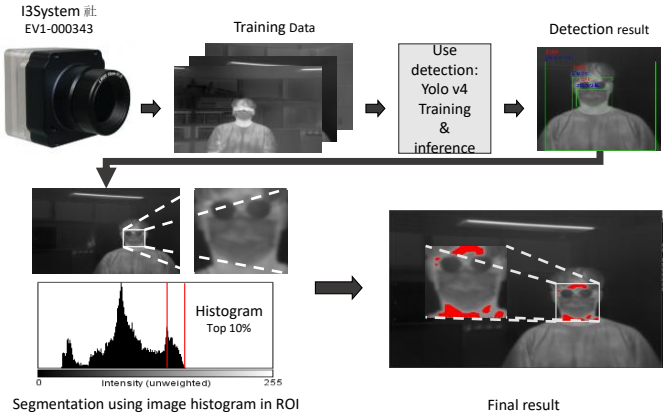


그림 1. YOLO를 활용한 열화상 기반의 체온측정 인공지능 시스템 개념도

이러한 문제를 해결하기 위하여 특정 영역에 제한되지 않는 감지와 영역 단위의 체온 측정 방법이 필요하다. 그리하여 딥러닝 방법 중 YOLOv4 를 이용한 사람 및 방해요소 감지로 특정 영역이나 위치에 제한되던 문제를 해결하며 마스크의 경우 착용 여부를 확인할 수 있다. 또한, 히스토그램을 통해 체온 측정 영역 분리 방법을 제시한다. [1, 3]

Ⅱ. 본론

본논문에서는 얼굴과 방해요소를 실시간 감지하기 위하여 YOLOv4 를 사용하고 있다. [1] 감지된 얼굴 영역에서 히스토그램을 통한 밝기 데이터 기반으로 체온 측정 영역 분리가 된다. 히스토그램은 영상처리에 있어 흔히 쓰이는 방법 중 하나이며 이미지 전체의 밝기 분포와 같은 이미지 특성을 알 수 있다. [3]

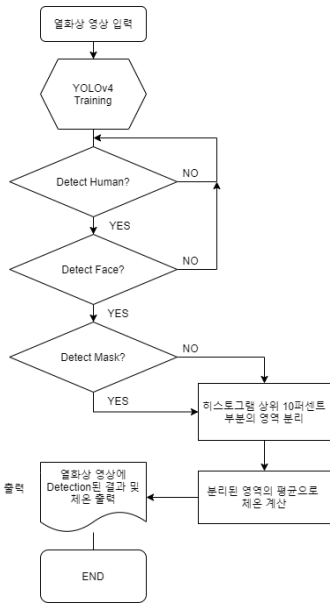


그림 2. 알고리즘 Flow-chart

이를 바탕으로 열화상 카메라 영상의 밝기 데이터를 사용하여 얼굴 영역안의 체온 측정 영역 분리에 사용된다.

본논문에서 제안하는 체온 측정 영역 분리 알고리즘은 YOLOv4 를 통해 얼굴과 방해요소를 감지하고 얼굴 영역의 밝기 분포를 히스토그램으로 데이터를 얻은 후 밝기 상위 n %에 해당하는 부분의 픽셀만큼 체온 측정 영역을 분리한다. n 은 상위 밝기 값의 분포 비율이다.

분리된 영역의 데이터를 사용하여 정확한 체온 측정을 위하여 온도 피팅이 필요하다. 온도 조절기를 사용하여 수조에 담은 물의 온도를 조절하며 수은 온도계와 열화상 카메라(i3system 社의 EV1-000343)를 사용하여 34 ~ 44℃ 범위를 0.5℃ 단위로 증가하며 온도 변화 측정 실험을 하였다. [7] 열화상 카메라와 온도 측정 포인트 사이의 거리는 0.6m 로 설정하였으며 수은 온도계는 수조 중앙에 위치하도록 설치하여 온도를 측정하였다.

열화상 카메라에서 얻은 데이터는 밝기 값을 기준으로 하였으며 그 외 온도 조절기와 수은 온도계는 표시되는 온도를 기준으로 하였다.

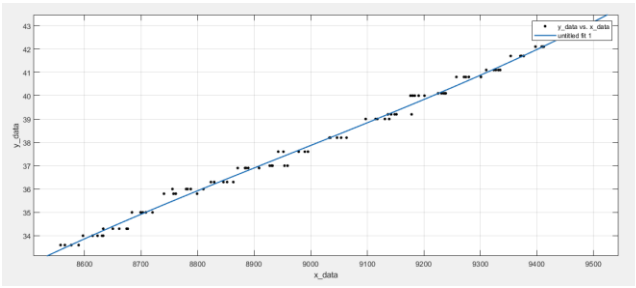


그림 3. 온도 피팅 실험 결과 그래프

$$T = \frac{(x - 5000)}{100} \qquad (1)$$

$$T = 3.045e^{-9} * x^3 - 8.186e^{-5} * x^2 + 0.7433 * x - 2241 \qquad (2)$$

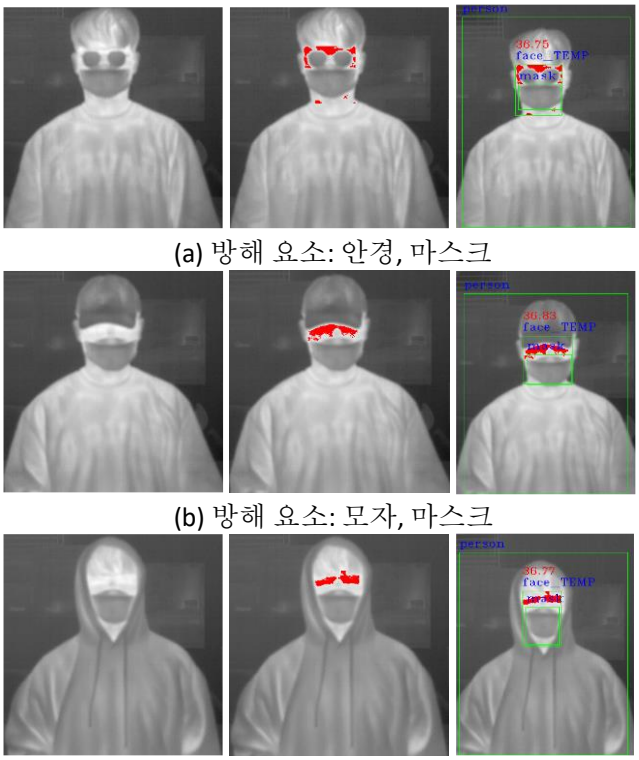
(1), (2) 식에서 T는 측정되는 체온을 의미하고, x 는 밝기 데이터를 의미한다.

식 (1) 은 실험에서 사용한 열화상 카메라의 체온 측정 식이며, 식 (2) 는 그림 3 에 보이는 온도 피팅 후 체온 측정 식이다.

Ⅲ. 결론

본논문에서는 YOLOv4 를 이용하여 사람, 얼굴, 방해요소(마스크, 안경 등)을 감지하고, 감지한 얼굴 영역의 히스토그램을 바탕으로 체온 측정 영역을 분리한다. [1, 3] 분리하여 얻은 데이터 값의 평균으로 체온을 측정하는 방법을 제안한다.

기존 방법들에 비해 특정 영역과 방해요소에 제한을 받지 않고 히스토그램을 통해 분리된 영역 평균과 온도 피팅 작업으로 인한 체온 측정 정확도 상승과 방해요소에 의해 받는 영향이 감소된다는 장점이 있다. 그림 4 에서 볼 수 있듯이 얼굴 영역안에서 체온 측정 방해요소인 마스크를 제외하여 분리하고 체온 측정하는 결과를 볼 수 있다.



(a) 방해 요소: 안경, 마스크

(b) 방해 요소: 모자, 마스크

(c) 방해 요소: 후드 티, 마스크

그림 4. YOLOv4를 통한 얼굴 및 방해요소 감지와 히스토그램 기반으로 측정 영역 분리 결과:
Blue Font – Object Name,
Red Font – Temperature Value

향후 본논문에서 제안하는 알고리즘을 임베디드 시스템화와 방해요소에 따라 분리되는 영역 비율의 차등화 알고리즘 업그레이드를 목표로 하고 있다. 또한, 블랙 바디를 사용한 온도 피팅 작업을 통해 체온 측정 결과 값의 정확도를 향상 및 거리에 따라 체온 측정 보조 작업과 다양한 종류의 방해요소와 마스크의 추가 학습을 진행할 예정이다.

ACKNOWLEDGMENT

이 논문은 2020 년도 정부 (산업통상자원부)의 재원으로 한국산업기술진흥원의 지원을 받아 수행된 연구임 (N0002428, 2020 년 미래형자동차 R&D 전문인력양성사업)..

참 고 문 헌

[1] Alexey Bochkovskiy, YOLOv4: Optimal Speed and Accuracy of Object Detection, arXiv:2004.10934, Apr 2020, https://arxiv.org/abs/2004.10934/

[2] Jia-Wei Lin, A Thermal Camera based Continuous Body Temperature Measurement System, ICCV Workshop, 2019.

[3] 정병조, 장성환, 열적외선 이미지를 이용한 영상 처리, 한국산학기술학회논문지 Vol.10 No.7, pp. 1503-1508, 2009.

[4] https://www.dahuasecurity.com

[5] https://www.dvs.co.uk

[6] http://www.meeso.cn

[7] http://i3system.com/

Multi-agent clinical decision support systems: A survey

Sara Kaviani Insoo Sohn

Division of Electronics & Electrical Engineering

Dongguk University

saraka1365@gmail.com isohn@dongguk.edu

Abstract

Agent-based clinical decision support systems are important to the medical industry and they can be used to improve the healthcare quality in various ways. These systems assist doctors and nurses to make diagnosis, prognosis and treatment, gathering medical data, and remote monitoring the elderly or patient people at home. This paper reviews some existing applications of multi-agent clinical decision support systems in medical and clinical workflows and problems.

1. Introduction

Decision making in medical areas is vital most of the time and have always been accompanied by mistakes due to big data management problems, incomplete or low amount of data or lack of knowledge and experience in nurses and primary care physicians who are in charge of initial physical examinations [1]. Clinical decision making follows a complicated multi-stage process, that includes data collection, diagnosis, prognosis and treatment suggestion and planning [2]. Clinical decision support systems (CDSS) are typically designed to integrate a medical knowledge base, patient data and an inference engine to generate specific advice for decision making [3]. Recently, agent-based systems has been hailed as a new method for intelligent software systems. Multi-agent CDSSs allow doctors and nurses to gather sufficient information from distributed data sources in a short time and process the data in different ways in order to help with medical and clinical decisions and enhance the decision's accuracy. The areas these systems can help in is diverse from storing and retrieval of medical records and data management to patients history and X-Ray analysis for the purpose of diagnosis and gathering and examining real-time data from monitors. In this paper we survey existing research in multi-agent CDSSs to review current trends in this area and investigate the capability of multi-agent models to provide support for the clinical needs.

2. Basics of multi-agent CDSS

In this section we briefly explain agents, multi-agent models and how they are used in medical sector. Then we discuss clinical decision support systems and their general role in assisting clinicians.

2.1 Agents

The word agent has many definitions in intelligent systems but the more common definition is that an agent is a system that can autonomously interact with, perceive, and act upon the information it perceives from the environment [4]. Therefore, some commonly used characteristics of an intelligent agent are reactivity, autonomy, learning, cooperation, reasoning, communication and mobility.

2.2 Multi-agent systems

Multi-agent system (MAS) is a loosely coupled network of agents that cooperate with each other to solve the problems that are beyond the individual capabilities or knowledge when there is no global control system. Therefore, to solve a specific problem, different agents have to focus on a different area or can only solve a specific part of a problem. In these systems, typically there is a core agent which is in contact with all other agents and coordinate agent activities.

2.3 CDSS

To improve health care, clinical decision support systems (CDSSs) support clinicians, staff, patients, or other individuals with knowledge and specific information or intelligence at appropriate times. The CDSS possibly will provide suggestions, but the clinician have to filter the information, review the suggestions, and decide whether to take action or what action to take.

3. Applications

Multi-agent systems has been widely integrated with DSSs to support medical and clinical decision making. Multi-agent CDSS (MA-CDSSs) can help in various areas which we are going to explain in this section.

3.1 Data repository management and data mining

One of the important applications of MA-CDSSs is to benefit non-specialist users providing easy access to clinical data. Having access to more data make the decision making process easier and more accurate. Therefore, to gather useful information, there are several steps that must be taken such as securely connecting a network of clinical centers and gathering distributed data [5,6], sharing data, data synthesis and pre-processing,

data mining [7,8] ,and feature extraction [9] which can be done by MA-CDSSs.

3.2 Diagnosis, prognosis and treatment

A specific type of CDSSs are diagnostic decision support systems (DDSS) which are developed to provide potential diagnosis, prognosis and treatment corresponding to given signs and symptoms. Applying multi-agent models in CDSSs, common decision making mistakes made by less experienced nurses in history and physical examination (H & P) processes [1], incomplete inputs, and primary care Physicians [10] can be eliminated. Moreover, in diagnosis of some sort of disease, that requires the combination of many different types of data including family and patient histories, laboratory results, imaging results and physical findings, such as Cardiac disorder [3,11] and variants of brain tumors [12,13], multi-agent CDSSs have shown to be successful. Moreover, machine learning has been integrated with MA-CDSSs in innovative ways. As an example in IMASC systems, a distributed CDSS has been proposed to assist physicians in diagnosing a certain disease by using supervised learning techniques [3].

3.3 Patient remote monitoring

A cornerstone technology in remote monitoring elderly and patient people is the decision support systems. These systems can generate risk detection decisions by gathering (e.x. via sensors which are attached to the patient), and analyzing data and finally sending the decisions to a remote monitoring center to take action when there is a real threat. There are various ways to use DSSs to improve the remote health care procedure such as encapsulating specific DSSs in intelligent agents that can exchange knowledge and intelligence, ALISA project ,QuoVAD project [14], and AGALZ as an autonomous agent to guarantee that the Alzheimer patients assigned to the nurses are given the right care [15].

4. Conclusion

This paper has presented a survey of recent research on the multi-agent clinical decision support systems to support decision making in clinical and medical problems. Multi-agent CDSSs can assist in different ways such as data gathering, management and information retrieval, diagnosis and treatment, and patient remote monitoring. Researches on using these systems for diagnosis and treatment is more extensive and most of CDSSs are merely used to retrieve information and assist physicians and nurses to discover new patterns. In none of the previous researches attempts were made to reduce the decision making time but mostly attempt to increase the decision quality. In most of the papers, the researchers tries to give innovative ideas rather than investigating and assessing their efficiency which makes it difficult to compare different methods.

Acknowledgments

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRFK) funded by the Ministry of Education (2018R1D1A1B07041981).

References

[1] Akbari, Z. and Unland, R., International Conference on Practical Applications of Agents and Multi-Agent Systems (pp. 29-41). Springer, Cham (2018).

[2] Perreault, L.E. and Metzger, J.B., Journal of Healthcare Information Management, 13, pp.5-22 (1999).

[3] Czibula, G., Czibula, I.G., Cojocar, G.S. and Guran, A.M., International Conference on Complexity and Intelligence of the Artificial and Natural Complex Systems (pp. 185-190). IEEE (2008).

[4] Ahmad, H.F., International Symposium on High Assurance Systems Engineering, Proceedings. (pp. 101-107). IEEE (2002).

[5] Canfield, K., Ramesh, V. and Quirologico, S., Internati onal Conference on System Sciences (Vol. 4, pp. 523-532). IEEE (1998).

[6] González-Vélez, H., Mier, M., Julià-Sapé, M., Arvanitis, T.N., García-Gómez, J.M., Robles, M., Lewis, P.H., Dasmahapatra, S., Dupplaw, D., Peet, A. and Arús, C., Applied intelligence, *30*(3), pp.191-202 (2009).

[7] Zaidi, S.Z.H., Abidi, S.S.R. and Manickam, S., Proceedings IEEE Symposium on Computer-Based Medical Systems (CBMS 2002) (pp. 339-342), 2002.

[8] Siddiqa, A., Niazi, M., Mustafa, F., Bokhari, H., Hussain, A., Akram, N., Shaheen, S., Ahmed, F. and Iqbal, S., In International Conference on Information and Communication Technologies (pp. 134-139). IEEE (2009).

[9] Gatta, R., Vallati, M., Dinapoli, N., Masciocchi, C., Lenkowicz, J., Cusumano, D., Casá, C., Farchione, A., Damiani, A., Van Soest, J. and Dekker, A., Artificial intelligence in medicine, 96, pp.145-153 (2019).

[10] Guo, X., Yu, H., Miao, C. and Chen, Y., In IJCAI (pp. 6521-6523), (2019).

[11] Hudson, D.L. and Cohen, M.E., In Computers in Cardiology (pp. 633-636). IEEE (2002).

[12] Arús, C., Celda, B., Dasmahaptra, S., Dupplaw, D., Gonzalez-Velez, H., Van Huffel, S., Lewis, P., Ariet, M.L.I., Mier, M., Peet, A. and Robles, M., In IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology Workshops (pp. 208-211). IEEE (2006).

[13] Gonzalez-Velz, H., Mier, M., Arus, C., Celda, B., Van Huffel, S., Lewis, P., Peet, A. and Robles, M., (2006).

[14] Dhouib, M.A., Bougueroua, L. and Węgrzyn-Wolska, K., In Computer Information Systems–Analysis and Technologies (pp. 74-84). Springer, Berlin (2011).

[15] Corchado, J.M., Bajo, J., De Paz, Y. and Tapia, D.I., Decision Support Systems, 44(2), pp.382-396 (2008).

Depthwise Separable Convolution for Facial Landmarks Detection

Savina Colaco and Dong Seog Han*
School of Electronics and Electrical Engineering, Kyungpook National University,

Daegu, Republic of Korea

savinacolaco@knu.ac.kr, dshan@knu.ac.kr*

Abstract

Facial landmarks detection has been used as important information for problems such as head pose estimation, facial emotion expression, face recognition, and face modelling. The facial keypoints are point information on the face such as eye corners, eye centre, mouth corners, nose, jawline, eyebrow points and so on. In this paper, depthwise separable convolution is used to predict vital keypoints on the face which is trained with public datasets with additional data. The face is detected with a widely used face detector. The predicted keypoints are mapped on a face detected to detect keypoints in real-time. The model is evaluated with wing loss and adaptive wing loss.

Ⅰ. Introduction

Facial landmark detection is also known as a facial alignment problem, is one of the challenging problems in the field of computer vision [1]. The landmarks can be used in applications such as face recognition, facial emotion recognition, self-driving cars and so on. With improved landmark detection, facial information can solve various facial alignment problems. Building a system with Convolutional neural networks (CNN) has been widely popular since it outperforms the traditional approach in speed and accuracy. Using CNN with deep structure, landmarks can be predicted and detected simultaneously. It can extract a high level of features needed for the prediction. In this paper, the deep learning approach depthwise separable convolution is used to predict the facial keypoints and mapped with the detected face in real-time.

Ⅱ. Experiment

The model trained with 300W [2] dataset which consists of XM2VTS, AFW, HELEN, LFPW and IBUG with additional data adding up to 112K grayscale images with 68 (x, y) coordinates. The input size scaled to 112x112 resolution. The model is implemented with Keras framework with epoch at 300 and batch size of 100. It uses Adam optimizer with learning rate fixed to 10^{-3} throughout the training. The model is evaluated with wing [4] and adaptive wing loss [5] as described by equations 1 and 2 respectively.

$$wing(x) = \begin{cases} \omega \ln\left(1 + \frac{|x|}{\varepsilon}\right) & \text{if } |x| < \omega \\ |x| - C & \text{otherwise} \end{cases} \quad (1)$$

where ω is a non-negative constant which sets the range of the nonlinear part to $(-\omega, \omega)$, ε limits the curvature of the nonlinear region and $C = \omega - \omega \ln(1 + \omega / \varepsilon)$ is a constant that smoothly links linear and nonlinear parts. The parameters are set to $\omega = 10$ and $\varepsilon = 2$.

$$Awing(y, \hat{y}) = \begin{cases} \omega \ln\left(1 + \left|\frac{y - \hat{y}}{\varepsilon}\right|^{\alpha - y}\right) & \text{if } |(y - \hat{y})| < \theta \\ A|y - \hat{y}| - C & \text{otherwise} \end{cases} \quad (2)$$

where y and $\hat{y}$ are ground truth and predicted values, respectively. Unlike wing loss ω as the threshold, Θ is the new variable threshold to switch between linear and non-linear part. The ω, Θ, ε, and α are positive values. $A = \omega(1/(1+(\Theta/\varepsilon)^{(\alpha - y)}))(\alpha - y)((\Theta/\varepsilon)^{(\alpha - y - 1)})(1/\varepsilon)$ and $C = (\Theta A - \omega \ln(1+(\Theta/\varepsilon)^{(\alpha - y)}))$ and are used to make smooth and continuous loss function at $|y - \hat{y}| = \Theta$. Similar settings from the paper (11) are used such as $\omega = 14$, $\Theta = 0.5$, $\varepsilon = 1$, and $\alpha = 2.1$.

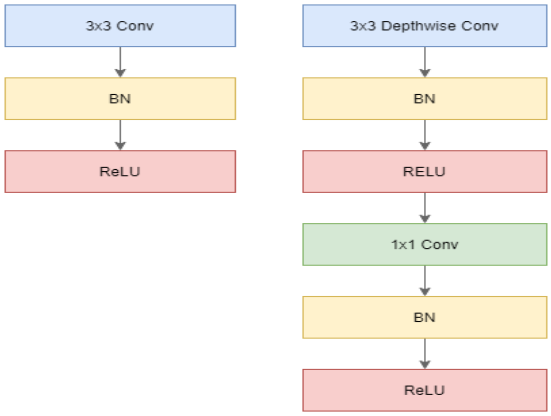


Figure 1: (Left) standard convolution layer and (right) MobileNet

The facial landmarks are predicted using mobileNet [6] model which is based on depthwise separable convolutions. The depthwise separable convolution is depthwise convolutions followed by pointwise convolution. In Fig. 1, the standard convolution layer is followed by batch normalization (BN) and rectified Linear units (ReLU). In MobileNet, the depthwise separable convolutions with depthwise and pointwise layers followed by batch normalization and ReLU. The depthwise separable convolution into two layers, where one layer for filtering and another for combining inputs whereas the standard convolution both filters and combines inputs into a new set of outputs in one step.

Single-shot detector (SSD) with ResNet as the backbone model is used to detect the user's face in the input image. In Figs. 2 and 4, the model accuracy with wing and adaptive wing loss functions are depicted in the plot with width multiplier 1 and 0.5, respectively.

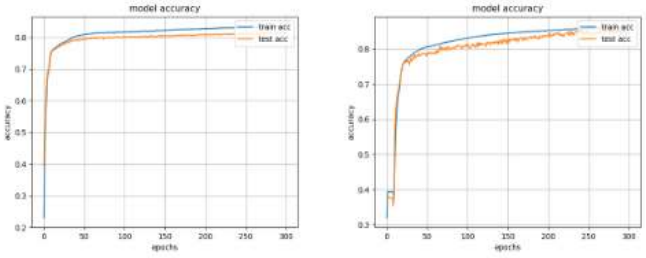


Figure 2: Model accuracy Loss functions with width multiplier=1

Figure 3: Facial keypoint detection with width multiplier = 1

In Figs. 3 and 5, the facial keypoints detection in real-time is demonstrated with width multiplier 1 and 0.5 respectively for wing and adaptive wing loss functions.

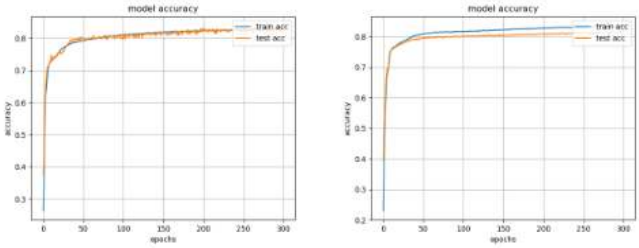


Figure 4: Model accuracy Loss functions with width multiplier=0.5

Figure 5: Facial keypoint detection with width multiplier = 0.5

Table 1: Model Accuracy of MobileNetV1 with wing loss and Adaptive wing loss

Model	Loss function	Width multiplier	Accuracy
MobileNet V1	Wing loss	α =1	84.4%
		α =0.5	82.5%
	Adaptive wing loss	α =1	84.9%
		α =0.5	81.5%

The model trained is 3.3M and 0.9M parameters in total for width multiplier 1 and 0.5, respectively. The model trained with adaptive wing loss has higher accuracy compared to wing loss. The model loss of adaptive wing loss is considerably lower than wing loss. The model sensitive to extreme head poses orientation and occlusion.

Ⅲ. Conclusion

In the paper, MobileNetV1 is used to predict the 68 facial keypoints. It is mapped with the detected face using an SSD detector with ResNet. The facial landmarks can be lost if the initial face detector failed to detect faces. Facial landmark detection is sensitive to extreme facial poses, occlusion and illumination conditions which can be improved.

ACKNOWLEDGMENT

This research was supported by the Ministry of Trade, Industry & Energy (MOTIE), Korea Institute for Advancement of Technology (KIAT) through 5G-based autonomous driving convergence technology demonstration platform task (task number: 1415169669).

References

[1] S. Shi, “Facial Keypoints Detection,” arXiv preprint arXiv:1710.05279, 2017.

[2] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, “300 faces in-the-Wild challenge: The first facial landmark localization challenge,” in Proc. IEEE Int. Conf. Comput. Vis. Workshops, Dec. 2013, pp. 397– 403

[3] Z.-H. Feng, J. Kittler, M. Awais, P. Huber, and X.-J. Wu, “Wing loss for robust facial landmark localisation with convolutional neural networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Salt Lake City, UT, USA, 2018, pp. 2235– 2245.

[4] X. Wang, L. Bo, and L. Fuxin, “Adaptive Wing Loss for Robust Face Alignment via Heatmap Regression,” in Proc. IEEE/CVF Int. Conf. Comput. Vis, Seoul, Korea, 2019, pp. 6971-6981.

[5] A. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, *et al.*, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," *ArXiv,* vol. abs/1704.04861, 2017.

Machine Learning Approach to Detect and Classify Power Line Fault

Md. Habibur Rahman, *Student Member, IEEE,* Md. Morshed Alam, and Yeong Min Jang, *Member, IEEE*

***Abstract*—In recent years, precise fault detection and localization of the power system are focused as the major research interest to ensure system protection and enhance the power quality. To protect power transmission line, identification of faulty phases is necessary and fault should be cleared accurately and reliably as fast as possible. Digital signal processing methods has made it easier to analyze faulty signal and efficiently detect precise fault. In this paper, we present machine learning approaches to detect and classify power line fault from the voltage and current signal.**

***Index Terms*—Empirical mode decomposition (EMD), Hilbert huang transform (HHT).**

I. Introduction

MODERN power suppliers aim at providing reliable and high quality power to the consumer. Fault occurrence in power systems creates impediment to this regard. Since faults can-not be avoided, it is evitable to recognize the faults and restore the supply of power by clearing the fault [1]. The major fault scenarios that are observed in power transmission line are Single line-to-ground faults, Line-to-line faults, Double line-to ground faults and three phase faults. These fault occurs mainly due to the lightening, falling trees on line and equipment male function. For the protection of power system, a protection scheme based on the operation of relays and circuit breakers has been developed. Various computational tools based on signal processing techniques has been employed for the proper functioning of protective devices. But, recently with the development of digital signal processing techniques and evolution of machine learning algorithm, every system is turning into an intelligent system [2]. In this paper, we have presented how a power system can be intelligent to detect and classify fault by itself. The rest of the paper is organized as follows: In section II, we have demonstrated machine learning approaches to detect and classify faults. We have studied feature extraction method in brief there. We have concluded our article in section III.

II. Proposed Approach

A. Workflow for Detecting and Classifying Fault

Recent advancements in the sensors for collecting various data popularize machine learning algorithms for decision making purpose. It has achieved paramount importance in complex, large and heterogeneous data processing where manual investigation may arise ambiguity. Over the years, various machine learning algorithms have been developed in order to perform classification, regression, and in general pattern recognition. Among these the most common algorithms are the support vector machine (SVM), the neural networks, the k-nearest neighbor (KNN), the logistic regression, the decision tree, the random forests, the linear discriminant analysis (LDA), the case-based reasoning, the naive Bayes, and the fuzzy logic [3]. In the following figure, methodology sing machine learning algorithm to classify the fault is given below:

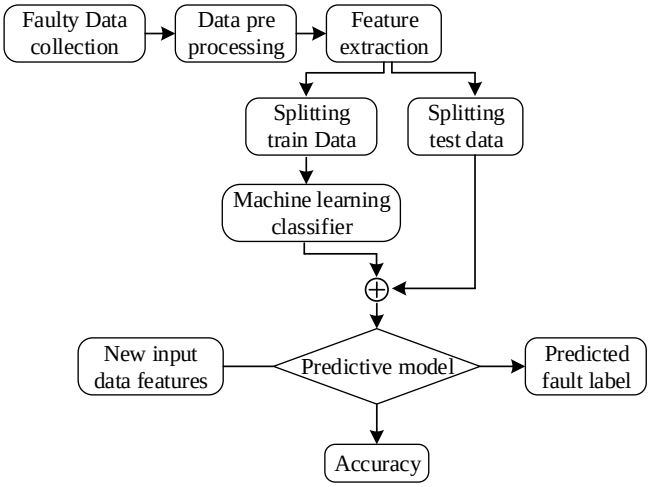


Fig. 1. Proposed methodology.

B. Feature Selection and Predictive Model Generation

Over the years, many soft computing techniques including Fourier Transform, Wavelet Transform and S-Transform are emerged as the popular techniques to detect the transient of power system network [4], [5]. Necessary features are extracted from these techniques. These extracted features are employed to train the machine learning classifier namely ANN and Fuzzy logic as well to detect and classify precise fault. Power system fault detection based on Wavelets and artificial neural networks (ANN) have been proposed in [6]. But the methods are limited due to having low accuracy during high impedance faults and when the fault inception angle is near zero [7]. Wavelet based combined fuzzy logic classifier is also introduced in the literature [8] for classification of faults of power system. Heavy load on power system also limits the accuracy of Fuzzy logic classifier [9]. Recently, EMD and HHT has been proposed for feature extraction instead of

Md. Habibur Rahman, Md Morshed Alam, and Yeong Min Jang are with the Department of Electronics Engineering, Kookmin University, Seoul 02707, South Korea (e-mail: rahman.habibur@ieee.org; yjang@kookmin.ac.kr).

TABLE I
FAULT CLASSIFICATION LOGIC TABLE.

A phase	B phase	C phase	Fault type
0	1	1	AG
1	0	1	BG
1	1	0	CG
0	0	1	AB
1	0	0	BC
1	1	0	CA
1	1	1	ABC

wavelets of voltage and current signal. The best features for the intended purpose are:

1) Energy distribution of instantaneous amplitude.
2) Standard deviation of amplitude.
3) Standard deviation of phase.

80% features out of the whole data set are splitted for training and the rest are kept for testing in general. Then, the trained data set are feeded into any of the machine learning classifier to train and generate the predictive model for detecting and classifying fault. Based on the Following logic table faults are classified.

C. Performance Measures for Classification

To properly evaluate the validity of the model we need to use proper performance measurement. In case of classification task we use precision, recall and F-score for each class defined by following equations. All these metrics may be obtained from confusion matrix. Precision represents as the number of examples correctly classified as class divided by the number of all the examples labeled by the classifier. Recall is the number of examples correctly classified as class divided by the number of all the examples of class in the data. F-score is a harmonic mean of the above.

$$Precision = \frac{True\ Positives}{Total\ Predicted\ Class} \tag{1}$$

$$Recall = \frac{True\ Positives}{Total\ Example\ of\ Class} \tag{2}$$

$$Fscore = \frac{Precision*Recall}{Precision+Recall} \tag{3}$$

III. CONCLUSION

In this paper we have studied in details how EMD and HHT assists to extracts features from voltage and current signal. Employing those features how machine learning algorithm can be applied detect and classify the fault. We have also demonstrated performance parameter to check our predictive model which will ensure the validity of our predictive model.

ACKNOWLEDGMENT

This research was financially supported by the Ministry of Trade, Industry and Energy (MOTIE) and Korea Institute for Advancement of Technology (KIAT) through the International Cooperative RD program (Project ID:P0011880).

REFERENCES

[1] J. A. Jiang, C. S. Chen, and C. W. Liu, "A new protection scheme for fault detection, direction, discrimination, classification, and location in transmission lines," *IEEE Trans. Power Del.,* vol. 18, no. 1, pp. 34–42, Jan. 2003.
[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
[3] Binkhonain, M., Zhao, L. (2019). A review of machine learning algorithms for identification and classification of non-functional requirements, *Expert Systems with Applications: X,* 1, 100001. doi:10.1016/j.eswax.2019.100001
[4] Aravena JL, Chowdhury FN. A new approach to fast fault detection in power systems, *inInt Conf Intell Syst Appl Power Syst,* Florida, USA; 1996. p. 328–32.
[5] T. Dalstein, and B. Kulicke, "Neural Network Approach to Fault Classification for High Speed Protective Relaying," *IEEE Trans. Power Del.,* vol. 10, no. 4, pp. 1002-1011, Apr 1995.
[6] Youssef OAS. Combined fuzzy-logic wavelet-based fault classification technique for power system relaying. *IEEE Trans Power Delivery*, 2004;19(2):582–9.
[7] B. Das, "Fuzzy Logic-Based Fault-Type Identification in Unbalanced Radial Power Distribution System," *IEEE Trans. Power Del.,* vol. 21, no. 1, pp. 278-285, Jan 2006.
[8] Shukla S, Mishra S, Singh B. Empirical-mode decomposition with hilbert transform for power-quality assessment. *IEEE Trans Power Delivery* 2009;24:2159–65.
[9] Huang NE, Wu MLC, Long SR, Shen SSP, Qu W, Gloersen P, Fan KL. A confidence limit for the empirical mode decomposition and Hilbert spectral 18 analysis. Proc R Soc A: Math Phys Eng Sci 2003;459:2317–45.

On the Performance Gains of Federated Learning Edge Caching in Vehicular Internet of Things

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—**In fifth generation (5G) wireless networks, one main challenge of edge artificial intelligence is to train machine learning models by aggregating a large amount of data that are distributed at different edge devices. Sending training data to a centralized cloud server introduces prohibitive communication overhead. Moreover, many types of data contain privacy-sensitive personal data. Federated learning (FL) is a recently proposed machine learning paradigm that allows to collaboratively train a shared model for many users without direct access to the raw data. In this study, we highlight the performance gains of FL edge caching. Subsequently, we outline the advantages and challenges of FL edge caching in vehicular Internet of things. It is shown that FL edge caching provides robust, stable, and flexible systems while ensuring a privacy-preserving solution for the data owners.**

Keywords— Edge caching; federated learning; 5G; vehicular internet of things.

I. INTRODUCTION

With the emergence of the fifth generation (5G) wireless networks, various devices in communication networks are able to connect and exchange information more efficiently than ever before [1]. Connected devices in the Internet of things (IoT) continuously generate enormous amount of data which would be requested by IoT application users. Transmitting requested IoT data from cloud servers would lead to increased network traffic and long delays. Therefore, edge caching mechanisms are used to ensure reduced latency and network traffic. Edge caching utilizes storage resources of edge nodes (ENs) by caching popular data files at the ENs. Then, user equipments (UEs) such as client vehicles in vehicular IoT can obtain the cached files from the ENs such as road side units without explicitly communicating with the cloud servers [2].

Federated learning (FL) edge caching is a decentralized machine learning (ML) edge caching technique that allows to collaboratively train a shared model for many UEs without direct access to the raw data. The technique allows distributed UEs to train local machine learning models on their local datasets and upload it to a server for a global model aggregation. FL edge caching is radically different from other more established techniques which upload raw data samples to servers.

In this study, we present an overview of the operational features of FL edge caching. Then, we highlight the performance gains of FL edge caching in 5G application scenarios. Moreover, we outline the advantages and challenges of FL edge caching in vehicular IoT.

The remainder of this paper is organized as follows. Section II presents an overview of the FL process and the unique features of the FL mechanisms. Section III outlines the advantages and challenges of FL edge caching in vehicular IoT. In section IV, the paper is concluded.

II. FEATURES OF FEDERATED LEARNING

A. Federated Learning Process

In general, FL presents an efficient training pattern which involves three main phases as demonstrated in Fig. 1. In phase 1, the global model on the server initializes model parameters and then all the UEs download the shared global model. In phase 2, each UE trains the model on its local data independently. Thus, each UE computes an individual update based on its local dataset. In phase 3, all local trained models are uploaded to the server via a secure protocol tunnel and are aggregated to learn a new global model. Then, the new global model is sent back to the UEs [3]-[5]. The learning process is iterated for many communication rounds until the global model is able to converge to the global optimal, a threshold level is achieved, or a desirable training accuracy is achieved. To ensure reduced communication overhead, FL frameworks employ selective communication such that only the important or relevant updates are transmitted in each communication round [3]. Only the model parameters or gradients are transmitted instead of the raw data.

B. Unique Features of FL

FL has the following unique properties as compared to other decentralized ML algorithms. (1) Ability to handle not independent and identically distributed (non-IID) data; the training data on a given UE is typically based on the usage of the device by a particular user, the wireless environment the UE experiences, the computation capability of the UE, and the energy consumption of the UE. Hence, any UE local dataset will not be representative of the training data of all UEs [6], [7]. In FL, the challenge of non-IID data can be met by merging the updates of the models by using the FederatedAveraging (FedAvg) algorithm [6]. (2) Unbalanced data; some users will make intensive use of a service or app than others. This can result in varying amounts of local training data. Furthermore, some UEs may have more computation tasks to be handled and some may experience more states of mobile networks. This can result in unbalancing training data among the UEs [6], [7]. Also, this challenge can be addressed by the FedAvg algorithm [6]. (3) Limited communication; UEs

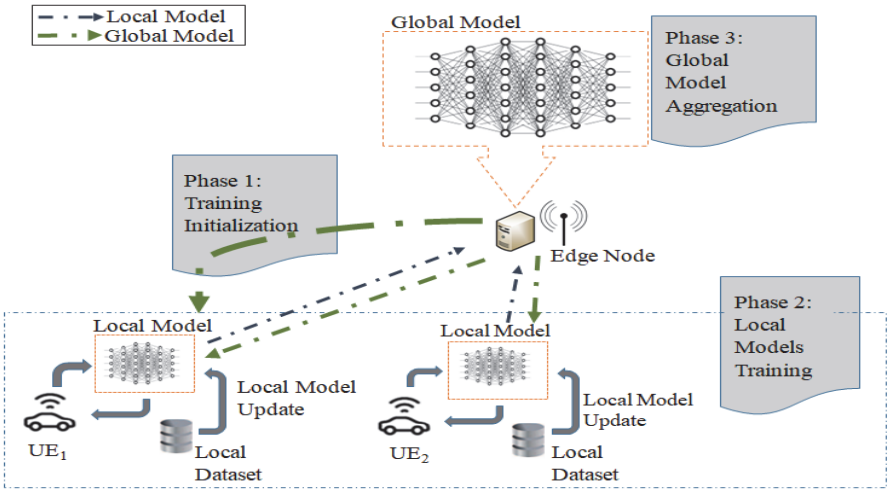


Fig. 1. An illustration of FL process.

are frequently and unpredictably offline, on slow or expensive connections, or they are allocated with poor communication resources [6], [7]. However, in FL, additional computation could decrease the consumption of communication rounds needed to train a model. Moreover, FL only asks a part of UEs, in one round, to upload their updates, which handles the situations where UEs are unpredictably offline [6]. (4) Privacy preservation; FL has the ability to collaboratively train a learning model on their individually gathered data, without revealing their privacy-sensitive data to a centralized server [5], [8]. Also, the information that needs to be uploaded to the server is the minimal update necessary. This feature is particularly useful in applications where datasets from the UEs are privacy-sensitive. Furthermore, the techniques of secure aggregation and differential privacy can be employed to ensure privacy preservation of data in the local updates [3], [6].

C. *Performance Gains of FL Edge Caching*

Several benefits of using FL in edge caching systems were demonstrated in [3], [6], [9], [10]. We summarize the benefits as follows. (1) System becomes more cognitive; in systems with a large number of UEs, the UEs can acquire various, abundant and personalized data for updating the global learning model. The data could include the quality of the wireless channel, the remaining battery life and the energy consumption, and the immediate computation capability. On the ENs, the cognitive data could include the computation load, the storage occupation, the number of wireless communication links, and the task queue states waiting for handling. As a result, the use of abundant and personalized distributed data instead of centralized training data ensures the system is more cognitive. (2) System becomes more robust; since FL can address the key issues such as the availability of the UEs, unbalanced data, and non-IID data, the performance of the systems is not easily affected by the unbalanced data or poor communication environment. Furthermore, its ability to handle non-IID data allows massive UEs in dynamic environments to train their own models without considering the overall negative effects. (3) Improved flexibility; in FL, additional computation could be used to decrease the number of communication rounds to train a model. The additional computation can be achieved by increasing the computation per UE. Therefore, UEs can decide to vary the number of mini-batches in training to adjust the communication cost. (4) Reduced network traffic and energy consumption; the decentralized training can significantly reduce the network traffic and energy consumption by sending only the features of interest rather than raw data. (5) Stability despite loss of connectivity; FL does not rely on synchronization among learners. Hence, even during a loss of connectivity between the ENs and UEs, the UEs can still build their local models. The feature is particularly important for highly dynamic and mission-critical applications such as vehicular IoT [9].

III. Federated Learning Edge Caching in Vehicular Internet of Things

A. *Suitability of FL Edge Caching in Vehicular IoT*

FL attracts a lot of interests from a large number of industries due to growing privacy concerns. Future vehicular IoT systems, such as cooperative autonomous driving and intelligent transport systems (ITS), feature a large number of devices and privacy-sensitive data where the communication, computing, and storage resources must be efficiently utilized. Therefore, FL edge caching is a promising approach to solve the existing challenges [11], [12].

The use of FL edge caching in vehicular IoT was considered in [3], [9], [11], [12], [13]. It was presented that the novel services in vehicular IoT systems present problems which can be addressed through the use of FL edge caching. The services demand unprecedented high reliability, high accuracy, and quick response. Furthermore, some services experience an extreme variance in their resource demands with respect to time, location, context, as well as individual users. Also, the vehicles are equipped with different types of sensor devices that generate and handle privacy-sensitive data, and the environments vary with time and road types. In such scenarios, FL edge caching helps to realize intelligent

vehicular IoT systems and privacy-preserving collaboration among different vehicles and road side units.

B. Challenges of FL Edge Caching in Vehicular IoT

In general, traditional FL algorithms suffer from the challenges of massive communication overhead and non-IID data [14], [15]. The accuracy of FL algorithms is significantly reduced when highly non-IID data is used as compared to when IID data is used. The communication overhead of FL mainly comes from the global model aggregation and update processes [3], [5], [8]-[12], [14], [15]. For instance, there exists FL algorithms that require each UE to communicate its full gradient update which may be in the size of gigabytes based on the learning architecture, and its millions of parameters. The size can increase to reach petabytes when the training is conducted on large-scale datasets [5]. The FedAvg is an example of FL algorithms that suffer from massive communication overhead. In [18], it was shown that the communication overhead in FL can impact other parameters such as the model accuracy and training time.

In vehicular IoT environments, the deployment of FL incurs several key challenges as highlighted in [3], [12], [13]. (1) The selection of UEs (client vehicles) for FL should address the mobility, communication bandwidth, and the specific scenarios that the UEs could represent. The mobility of vehicles makes it difficult to maintain continuous synchronized communication between the server and the vehicles. The consideration of mobility and communication bandwidth ensures a successful dissemination of the learning model and an accurate aggregation of local updates from the UEs. The consideration of vehicle scenarios guarantees that selected data for training includes a wide range of samples, avoiding the over fitting for a non-representative scenario. (2) Most vehicular IoT applications have stringent latency and reliability constraints. Therefore, it is important to devise an enhanced FL architecture which is more suitable for vehicular environments. (3) The dynamicity of vehicular environments makes the communication and computational resource allocation particularly difficult. Consequently, it becomes important to design efficient resource allocation algorithms that could satisfy the need of FL. (4) Centralized curator for aggregation is vulnerable to security threats which can lead to failure of the whole learning process. An asynchronous FL architecture was presented in [13] to mitigate the challenge.

IV. CONCLUSION

This paper presents an overview of the operational features of FL edge caching. It highlights the performance gains of FL edge caching. Furthermore, it outlines the advantages and challenges of FL edge caching in vehicular IoT. It is shown that FL frameworks exchange only model parameters learned locally at client vehicles. As a result, FL edge caching provides a better privacy-preserving solution for the data owners while ensuring robust, stable, and flexible systems.

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] Z. Ning, et al, “Joint computing and caching in 5g-envisioned internet of vehicles: A deep reinforcement learning-based traffic control system,” *IEEE Transactions on Intelligent Transportation Systems*, Early Access Article, 2020.

[2] 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 Of Things Journal*, vol. 6, no. 2, pp. 2074–2083, April 2019.

[3] W. Y.B. Lim, et al., “Federated learning in mobile edge networks: A comprehensive survey,” *IEEE Communications Surveys & Tutorials*, Early Access Article, 2020.

[4] 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 of Things Journal*, Early Access Article, 2020.

[5] M. Asad, A. Moustafa, and T. Ito, “FedOpt: Towards communication efficiency and privacy preservation in federated learning,” *Applied Sciences*, vol. 10, 2864, 2020.

[6] 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 Network*, vol. 33, no. 5, pp. 156–165, October 2019.

[7] H. B. McMahan, et al, “Communication-efficient learning of deep networks from decentralized data,” [Online]. Available: https://arxiv.org/abs/1602.05629, 2017.

[8] L. Wang, W. Wang, B. Li, “CMFL: Mitigating communication overhead for federated learning,” in *Proc. IEEE 39th Int. Conf. Distrib.Comput. Syst. (ICDCS)*, Jul. 2019, pp. 954–964.

[9] S. Samarakoon, M. Bennis, W. Saad, and M. Debbah, “Federated learning for ultra-reliable low-latency v2v communications,” in *Proc. 2018 IEEE Global Communications Conference (GLOBECOM)*, December 2018.

[10] S. Niknam, H. S. Dhillon, and J. H. Reed, “Federated learning for wireless communications: Motivation, opportunities and challenges,” 2020. [Online]. Available: arxiv preprint arxiv:1908.06847, 2020.

[11] J. Zhang, and K. B. Letaief, “Mobile edge intelligence and computing for the internet of vehicles,” *Proceedings of the IEEE*, vol. 108, no. 2, pp. 246–261, February 2020.

[12] Z. Du, et al, “Federated learning for vehicular internet of things: Recent advances and open issues,” *IEEE Open Journal of the Computer Society*, vol. 1, pp. 45–61, May 2020.

[13] Y. Lu, X. Huang, Y. Dai, S. Maharjan, and Y. Zhang, “Differentially private asynchronous federated learning for mobile edge computing in urban informatics,” *IEEE Transactions On Industrial Informatics*, vol. 16, no. 3, , pp. 2134–2143, March 2020.

[14] Q. Wu, K. He, and X. Chen, “Personalized federated learning for intelligent iot applications: A cloud-edge based framework,” *IEEE Open Journal of the Computer Society*, vol. 1, pp. 35–44, May 2020.

[15] Y. Sun, S. Zhou, and D. Gündüz, “Energy-aware analog aggregation for federated learning with redundant data,” in *Proc. ICC 2020*, 7-11 June 2020.

[16] L. Liu, J. Zhang, S. Song, and K. B. Letaief, “Edge-assisted hierarchical federated learning with non-iid data,” [Online]. Available: https://arxiv.org/abs/1905.06641, 2019.

[17] J. Mills, J. Hu, and G. Min, “Communication-efficient federated learning for wireless edge intelligence in iot,” *IEEE Internet Of Things Journal*, vol. 7, no. 7, pp. 5986–5994, July 2020.

[18] H. Sun, S. Li, F. R. Yu, Q. Qi, J. Wang, and J. Liao, “Towards communication-efficient federated learning in the internet of things with edge computing,” *IEEE Internet of Things Journal*, Early Access Article, May 2020.

3D Depth 영상을 이용한 딥러닝 기반 이상 행위 인지 기술

김동칠, 박성주
한국전자기술연구원
dckim@keti.re.kr, bpark@keti.re.kr

Deep Learning-based Anomaly Action Recognition Scheme using 3D Depth Images

Dongchil Kim, Sungjoo Park
Korea Electronics Technology Institute

요 약

본 논문은 딥러닝 기반 이상 행위 인지 기술을 제안한다. 제안된 기법은 3D Depth 영상을 이상 행위 인지에 활용함에 따라 기존의 2D 영상 기반 영상 보안 분석 기술의 단점을 보완함으로써, 저조도 환경에서의 객체 및 이벤트 탐지가 가능하다. 이를 위해, R(2+1)D 모델을 수정하고 실환경 영상 데이터셋을 기반으로 학습하여 이상 행위 모델을 생성하였다. 실험 결과를 통해, 제안한 딥러닝 기반 이상 행위 인지 기술의 성능을 확인하였다.

Ⅰ. 서 론

CCTV 환경에서 영상 분석 기술은 카메라로부터 취득한 영상 정보를 분석하여 객체 및 이상 행위를 탐지하는 기술이다. 최근, 딥러닝을 이용한 영상 분석 기술 성능이 급격하게 향상됨에 따라, 딥러닝 기술을 활용하여 객체 탐지 및 이상 행위를 인지하는 기술들이 증가하고 있다. 그러나 기존 2D 카메라 기반의 영상 분석은 저조도 및 정확한 행위 인지가 필요한 환경에서 객체 탐지 및 이상 행위 인지가 어렵다 [1-2].

이를 해결하기 위해, 본 논문에서는 3D 카메라 기술을 활용하여 취득한 3D Depth 영상을 기반으로 이상 행위 인지 기술을 제안하였다. 제안된 기술은 저조도 환경에서 실시간으로 이상 행위를 인지하기 위해, R(2+1)D 모델[3]을 수정하여 이상 행위 인지 모델을 생성하였다. 이상 행위 모델을 학습하기 위해 3D 카메라를 이용하여 다양한 뷰에서 취득한 학습 데이터를 이용하였다.

Ⅱ. 본론

이상 행위 인지 모델의 학습을 위해, 실환경 기반 위험 행위 영상 데이터셋(KETI-RGBD)과 오픈 영상 데이터 셋을 취득하였다. 그림 1, 2와 같이, 자체적으로 취득한 KETI-RGBD 데이터 셋은 RGB, Depth, 저조도 이미지 정보를 취득하였으며, 총 13종 행위의 비디오 샘플로 구성된다. 비디오 샘플은 10명 이상의 사람을 이용하여 3개 이상의 다른 뷰에서 촬영하였다. 데이터 셋 취득에 사용된 3D 카메라는 마이크로소프트의 Kinect V2와 Azure Kinect DK를 활용하였다. 13종의 행위 종류는 버림, 펀치, 발차기, 쓰러짐, 나타남, 사라짐, 배회, 가상경로통과, 구타, 군집, 위장, 기물파손, 비디오신호아웃이다. 오픈 영상 데이터 셋으로 중국 북경대학교에서 제작한 PKU-MMD를 이용하였다. PKU-MMD는 Kinect V2를 이용하여 RGB, Detph, 관절, IR 이미지 정보를 취득하며, 총 54종 행위, 1,076개의 비디오 샘플로 구성된다. 비디오 샘플내에는 20,000개의 행위가 포함되어 있으며, 66명의 사람을 이용하여 3개의 다른 뷰에서 촬영하였다 [1].

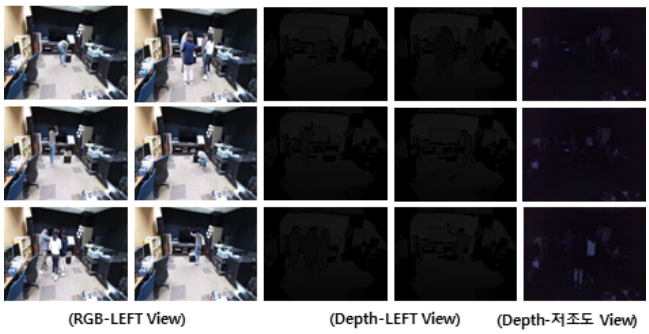


그림 1 Azure Kinect DK 기반 영상 데이터 셋 취득(측면-좌)

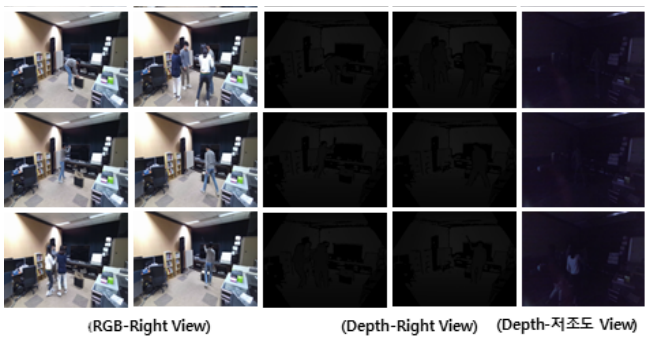


그림 2 Azure Kinect DK 기반 영상 데이터 셋 취득(측면-우)

3D Depth 영상을 이용한 딥러닝 기반 이상 행위 인지 기술의 구조는 그림 3과 같다. 취득된 이상 행위 영상 데이터 셋을 행위 인지 모델과 객체 탐지 모델에 각각 입력하여 학습한다. 3D Depth 이미지 시퀀스를 그림 4와 같이, R(2+1)D 모델에 입력하여 학습을 진행하며, 정답 행위와 추론한 행위의 오차가 최소가 되도록 네트워크의 가중치를 업데이트한다. 또한, 실환경에서 이상 행위 인지의 성능 향상을 위해, 표 1과 같이 하이퍼파라

미터를 수정하였다. 또한, 제안된 기술은 R(2+1)D 모델에서 3D 영상 입력을 통해 행위 인지가 가능하도록 수정하였으며, 연속적인 영상 입력에 따른 실시간 행위 인지 처리를 위해 슬라이딩 윈도우 기술을 적용하였다.

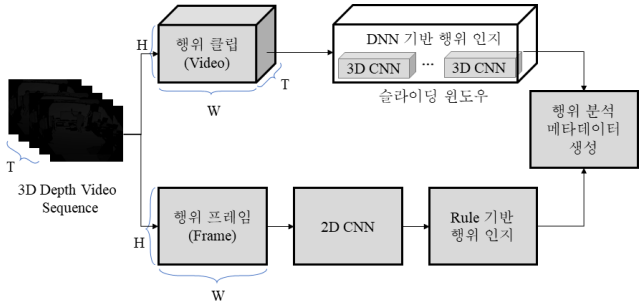


그림 3 이상 행위 인지 구조

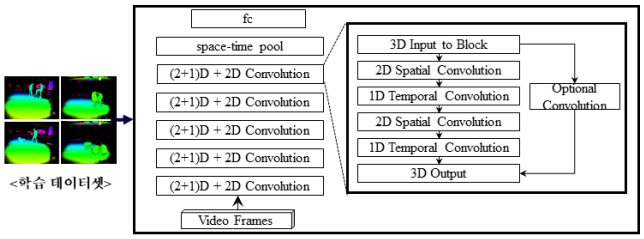


그림 4 R(2+1)D 학습 모델

Ⅲ. 결론

본 논문은 3D Depth 영상을 활용한 딥러닝 기반 이상 행위 인지 기술을 제안하였다. 제안된 기술은 실환경에서 3D Depth 영상을 이용한 이상 행위를 인지하기 위해, 3D 카메라를 통해 이상 행위 영상 데이터 셋을 취득하였다. 또한, 3D Depth 영상 기반 이상 행위를 인지하기 위한 R(2+1)D 모델을 수정하였다. 이를 통해, 그림 5와 같이, 밝은 환경뿐만 아니라, 저조도 환경에서도 위험 행위 인지가 가능하며, 인지 성능이 76.72%임을 확인하였다.

표 1 하이퍼파라미터 설정

Hyper-parameters	Valesexplored
Optimizer	SGD
Loss function	Cross entrophy Loss
Activation function	ReLU(Hidden layer)
	Linear(Output layer)
Learning rate	0.01
Batch size	10
Epochs	45

ACKNOWLEDGMENT

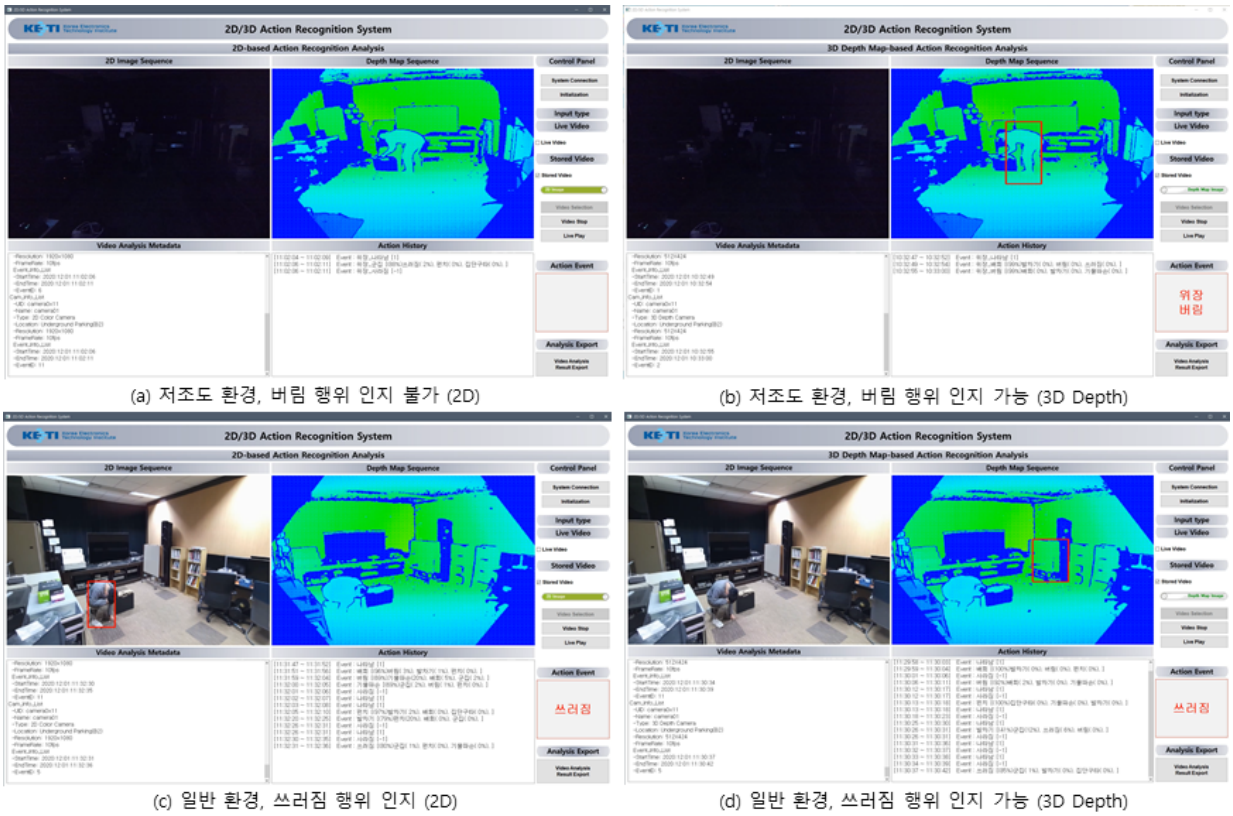
본 논문은 2020년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No. 2014-3-00077, (딥뷰-2세부) 대규모 실시간 비디오 분석에 의한 전역적 다중 관심객체 추적 및 상황 예측 기술 개발).

참 고 문 헌

[1] 김동칠, 박성주, “3D 영상 기반 위험 행위 인지 기술,” 2020년 대한임베디드공학회 추계학술대회, 2020. 11..

[2] S. Park and D. Kim, “Video Surveillance System Based on 3D Action Recognition”, International Conference on Ubiquitous and Future Networks (ICUFN), pp. 868-870 , Jul. 2018.

[3] D. Tran, H. Wang, L. Torresani, J. Ray, Y. LeCun, and M. Paluri, “A Closer Look at Spatiotemporal Convolutions for Action Recognition”, IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), pp. 6450-6459, Jun. 2018.



(a) 저조도 환경, 버림 행위 인지 불가 (2D)

(b) 저조도 환경, 버림 행위 인지 가능 (3D Depth)

(c) 일반 환경, 쓰러짐 행위 인지 (2D)

(d) 일반 환경, 쓰러짐 행위 인지 가능 (3D Depth)

그림 5 3D Depth 영상을 활용한 딥러닝 기반 이상 행위 인지 기술 결과

가속도 및 자이로 센서를 이용한 딥러닝 기반 행위인지 정확도 향상을 위한 심층 분석 처리 시스템

안영민, 이승진, 허의남

경희대학교

aem9704@khu.ac.kr, seungjin@khu.ac.kr, johnhuh@khu.ac.kr

Deep-Level Analysis Processing System for Improving Accuracy of Deep Learning Neural Newtwork-based Activity Recognition Using Acceleration and Gyro Sensors

Youngmin An, Seung-Jin Lee, Eui-Nam Huh

Kyung Hee University

요 약

최근 높은 스마트폰 보급률은 다양한 스마트폰 내장 센서 활용에 대한 기회를 제공한다. 가속도 및 자이로 센서 역시 대표적인 스마트폰 내장 센서 중 하나로 운동량 측정, 행동인지 등의 서비스 제공에 중요한 역할을 하고 있다. 본 논문은 가속도 및 자이로 센서에서 발생하는 시계열 데이터를 바탕으로 행위인지 정확도 향상을 위한 심층 분석 처리 시스템을 제안한다. 본 분석 처리 시스템은 사용자의 신체활동을 판단하기 위해 사용자가 등에 가속도 및 자이로 센서를 착용한 채 수집된 앉는 중/서는 중/걷는 중/눕는 중/누워있는 중/누웠다 서는 중/가만히 있는 중(앉아있거나 서있는 중) 7가지 행동 유형에 대한 충분한 시계열 데이터가 제공됨을 가정한다. 본 논문에서 제안하는 행위인지 정확도 향상을 위한 심층 분석 처리 시스템으로 행위인지를 수행하면 85% 이상의 정확도로 앉는 중/서는 중/걷는 중/눕는 중 4가지의 행동 유형을 예측할 수 있다.

Ⅰ. 서 론

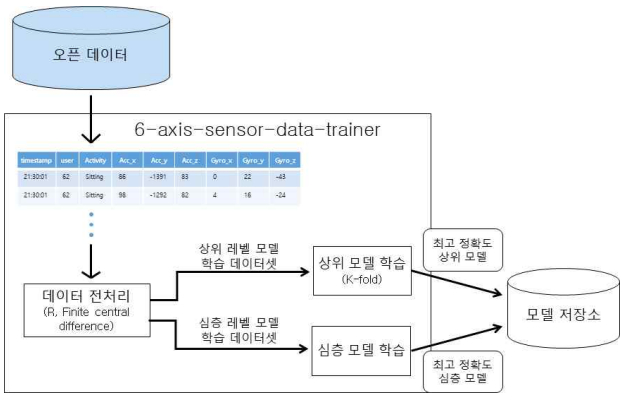
사용자 행동 유형은 '요리하는 중', 'TV보는 중', '조깅', '윗몸 일으키기' 등 분류 기준과 응용 목적에 따라 매우 다양하게 존재한다. '요리하는 중', 'TV보는 중' 행동 유형들은 헬스케어 서비스에서 모니터링의 대상으로서 활용될 수 있다. '조깅', '윗몸 일으키기' 행동 유형들은 스마트 홈트레이닝 산업에 활용될 수 있다[1-2]. 이렇게 다양한 행동 유형들은. 앉기/서기/걷기/눕기 4가지 행동 유형으로부터 파생된다. 예를 들어 '요리하는 중'은 주로 '서기' 행동을 취한 채 행해지며 'TV보는 중'은 '앉기' 행동을 취한 채 행해진다. 즉 행위인지를 필요로 하는 다양한 목적의 응용 서비스들은 앉기/서기/걷기/눕기 4가지 행동 유형에 대한 행위 인지를 기반으로 세분화할 수 있다.

본 논문에서는 가속도 3축 및 자이로 3축에서 발생하는 시계열 데이터를 바탕으로 4가지 행동 유형 앉기/서기/걷기/눕기 행위인지를 위한 딥러닝 기반 심층 분석 처리 시스템을 제공하여 4가지 행동에 기존보다 높은 정확도를 가지도록 한다.

Ⅱ. 본론

● 학습 컨테이너의 행동 예측 모델 학습 과정

[그림 1]은 학습 컨테이너의 학습 과정이다. 학습에 필요한 데이터는 가속도 x, y, z 3축, 자이로 x, y, z 3축, 타임스탬프, 행동 유형, 사용자 ID로 이루어진 시계열 오픈 데이터를 데이터 저장소에 보관하였다. 학습하고자 하는 사용자 ID와 함께 실행되면 컨테이너는 우선 데이터 저장소에서 해당 사용자에 대한 시계열 데이터를 타임스탬프 순으로 오름차순 정렬하여 모두 불러온다. 그 후 개인별 두 종류의 행동 예측 모델 생성 과정이 이루어지는데 이때 학습에 사용되는 모델은 시계열 데이터를 학습시키기에 적절한 Recurrent Neural Network(RNN) 의 한 종류인 Long Short-Term Memory(LSTM) 이다[3]. 개인별 두 종류의 모델을 생성하는 이유는 상위 행동 예측 모델에서 7가지 행동 유형 중 앉는 중/서는 중/가만히 있는 중(앉아있거나 서있는 중) 3가지 행동 유형에 대해 정확도가 떨어지는 경향이 있기에 이를 보정하기 위해서 해당 3가지 행동 유형에 대해서만 심층 레벨 행동 예측 모델을 생성한다.



〔그림 1〕 학습 컨테이너의 학습 과정

▶ 상위 레벨 행동 예측 모델 생성 과정

우선 7가지 행동 유형에 대해 데이터 셋들을 생성하는 전처리 과정을 거친다. 가장 오래된 레코드를 시작 레코드로 하여 오름차순으로 총 50개의 레코드를 추출한다. 추출된 50개의 레코드가 단일 행동 유형을 나타낸다면 데이터 셋 후보로 선정한다. 이어서 이전 시작 레코드로부터 오름차순으로 5개 떨어진 레코드를 다음 시작 레코드로 선정하고 새로운 시작 레코드로부터 오름차순으로 50개의 레코드를 추출하여 단일 행동 유형을 나타내면 데이터 셋 후보로 선정한다. 이 과정을 시계열 데이터 셋 전체에 수행한다. 그 다음 7가지 행동 유형의 데이터 셋들이 각각 균일한 개수를 가지도록 랜덤 추출하여 개수를 균일하게 맞춘다.

전처리 과정을 거쳐 생성된 데이터 셋들은 학습 정확도의 신뢰성을 보장하기 위해 5개의 군집으로 랜덤하게 분할하여 K=5의 K-fold 교차 검증을 진행한다. 학습 데이터 셋을 충분히 보장하기 위해 Validation Set은 따로 확보하지 않는다. 실험결과로부터 RNN의 hidden size는 150으로 선정하였다. 100부터 1,000까지 100단위의 epochs마다 K-fold 교차검증으로 도출된 정확도들의 평균치를 구하고 가장 높은 정확도를 보이는 epochs만큼 전체 데이터 셋을 다시 학습하여 상위 레벨 행동 예측 모델로 모델 저장소에 생성한다.

hidden size	100	150
epochs	acc.	acc.
100	89.20	91.53
200	87.26	93.47
300	89.33	92.64
400	91.98	93.35
500	91.46	92.61
600	89.96	90.21
700	86.97	89.42
800	85.92	89.89
900	82.91	85.76
1000	87.09	88.25

〔표 1〕 hidden size에 따른 정확도

〔그림 2〕 hidden size에 따른 정확도 비교

▶ 심층 레벨 행동 예측 모델 생성 과정

심층 레벨 행동 예측 모델은 앉는 중/서는 중/가만히 있는 중(앉아있거나 서 있는 중) 3가지 행동 유형에 대해 정확도를 보정하기 위해서 생성한다. 상위 레벨 모델 생성 과정과 다르게 처음 레코드로부터 50개의 레코드들을 추출한 후 동일한 행동 유형을 가질 경우 바로 데이터 셋 후보로 선정하지 않고 각 가속도 3축, 자이로 3축들에 대해 미분을 적용한다. 각 축의 한 값을 $f(x)$라고 할 때 $f(x+h)$와 $f(x-h)$는 [수식1]과 같이 Taylor serier 전개식으로 표현할 수 있으며 이들의 차이에 $2h$를 나눠주어 [수식2]처럼 'Finite central difference'를 구할 수 있으며 이를 근사된 미분값으로 활용할 수 있다.

데이터 셋 후보를 구성하기 위해 추출된 50레코드들에서 세 번째 레코드 값들과 첫 번째 레코드 값들의 차이로 이루어진 하나의 레코드, 네 번째 레코드 값들과 두 번째 레코드 값들의 차이로 이루어진 하나의 레코드처럼 총 48개의 미분 근사된 레코드를 재구성하여 데이터 셋 후보로 선정한 뒤 3가지 행동 유형들이 균일한 개수의 데이터 셋을 가지도록 랜덤 추출하여 최종 데이터 셋을 선정하고 20%를 Test Set으로 교차 검증 없이 학습하여 100부터 1,000까지 100단위의 epochs마다 정확도를 도출하고 가장 높은 정확도를 보이는 epochs로 심층 레벨 행동 예측 모델을 선정하여 모델 저장소에 저장한다.

$$f(x+h)=f(x)+hf'(x)+\frac{1}{2}h^2f''(x)+\frac{1}{3!}h^3f'''(x)+O(h^4)\ldots$$
$$f(x-h)=f(x)-hf'(x)+\frac{1}{2}h^2f''(x)-\frac{1}{3!}h^3f'''(x)+O(h^4)\ldots$$

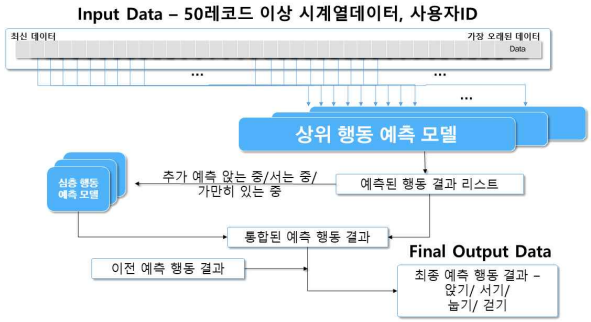
〔수식 1〕 Taylor series 전개식

$$f(x+h)-f(x-h)=2hf'(x)+\frac{2}{3!}h^3f'''(x)+O(h^4)$$
$$\frac{f(x+h)-f(x-h)}{2h}=f'(x)+O(h^2)$$

〔수식 2〕 Taylor series 전개 차이를 활용한 Finite central difference 식

● 행동 예측 컨테이너의 행동 예측 동작 과정

[그림 3]는 행동 예측 컨테이너의 행동 예측 동작 과정을 나타낸다. 50레코드 이상의 시계열 데이터가 사용자 ID와 함께 요청이 들어오면 해당 사용자의 두 종류 모델을 모델 저장소로부터 불러온다. 그 후 가장 최신 50레코드를 예측 데이터 셋으로 상위 레벨 행동 예측 모델에 투여하여 예측된 행동 결과를 반환받는다. 만약 예측된 행동 유형이 앉는 중/서는 중/가만히 있는 중에 속할 경우 예측 데이터 셋을 48개의 미분된 레코드들로 변환하고 심층 레벨 행동 예측 모델에 투여하여 보정된 행동 결과를 반환받는다. 동일한 방식으로 가능할 때까지 시계열 데이터의 마지막 레코드로부터 5개씩 오래된 레코드로 이동하면서 예측 결과를 반환받아 순서대로 리스트에 저장한다. 그 후 '앉았다 서는 중'과 '누웠다 서는 중'은 '서는 중'으로, '누워있는 중'은 '눕는 중'으로 변형하여 통합시킨다. 그리고 weighted voting을 실시하는데 리스트를 타임스탬프의 오름차순으로 정렬한 후 가장 최신의 행동에 더 높은 weight를 주기 위해 [1, 2, 2, 3, 3, 3, 4, 4 ...] 패턴의 수열의 각 자리의 대응되는 행동 유형에 voting한다. 다음으로 정렬된 리스트에서 바로 이전 행동 결과와 동일한 행동 결과가 연속해서 나온 행동 유형에 weight 2점씩 voting 한다. 그 후 가장 높은 voting 점수를 나타내는 행동 유형을 통합된 행동 예측 결과로 채택한다. 마지막으로 이전 최종 행동 예측 결과를 바탕으로 현재 통합된 행동 예측 결과를 비교하여 안지/서기/걷기/눕기 4가지 행동 유형 내의 현재 최종 행동 예측 결과를 도출하여 응답한다.



〔그림 3〕 행동 예측 컨테이너의 행동 예측 동작 과정

Ⅲ. 결론

본 논문에서 제시한 시스템으로 4가지 행동에 대해 행위인지를 실시할 경우 85% 이상의 정확도를 보인다. 이는 IoT 웨어러블 센서 기반의 헬스케어 시스템이나 스마트 홈트레이닝 서비스 등에서 유용하게 활용될 것으로 기대된다. 또한 4가지 행동은 다른 구체적 행동의 근간이 되므로 계층적인 행위인지 접근방식에서 이 시스템이 기초적인 역할을 할 수 있을 것이다. 향후, 본 논문에서 제시한 시스템을 웨어러블 센서를 이용한 요양시설 환경에서의 실시간 신체활동 케어 시스템에 적용하여 실제 구현 시스템에서의 정확도를 실험하고자 한다.

ACKNOWLEDGMENT

이 논문은 2020년도 정부(산업통상자원부)의 재원으로 한국산업기술평가관리원의 지원을 받아 수행된 연구임 (No. 20002610, 스마트 의류 플랫폼 기반 신체활동 취약자 실시간 케어용 웨어러블 상황인식 센서 디바이스 기술 개발)

참 고 문 헌

[1] 이승진, 안영민, 허의남. (2020). 웨어러블 센서를 이용한 요양시설 환경에서의 AI 기반 실시간 신체활동 케어 클라우드 시스템 설계. 한국통신학회 추계종합학술발표회 논문집, 80-81

[2] 이원주, 허태호, 이승룡. (2014). 손목착용형 웨어러블 가속도센서를 이용한 운동 행위인지 시스템. 한국정보과학회 학술발표논문집, 399-401.

[3] 신수연, 차주헌. (2018). 다중모드 센서와 LSTM 기반의 딥 러닝을 이용한 인간의 행동인식 시스템. 대한기계학회 논문집 A권, 42(2), 111-121.

자동 말투(Speech Style) 인식: 다자간 대화 상황에서의 화자인식 기술 개발

강가람, Jin Guangxun*, 권오병**
경희대학교, *경희대학교, **경희대학교

1st9aram@khu.ac.kr, *jinguangxun0407@khu.ac.kr, **obkwon@khu.ac.kr

Automatic Speech Style Recognition: Development of Speaker Recognition Technology in Multilateral Conversation

Kang Ga Ram, Jin Guangxun*, Kwon Oh Byung**
Kyung Hee Univ., * Kyung Hee Univ., ** Kyung Hee Univ.

요 약

종결어미 중에는 존어체와 경어체가 있어, 화자의 높임법을 추측하게 하여 화자와 청자 사이의 우열을 파악하는데 유용하기도 하다. 또한 종결어미는 시간의 흐름에 따라 새로운 종결어미가 등장하기도 하고 사라지기도 하며, 그 의미가 변천하기도 하는 매우 역동적인 양상을 보인다. 이에 본 연구의 목적은 화자인식의 정확도를 개선하기 위해 화자가 발화한 문장에 등장하는 종결어미의 등장 특성을 바탕으로 화자인식 하는 방법을 제안하는 것이다. 이를 위해 '응답하라 1994'라는 K-Drama 자막 데이터를 학습데이터로 하여 종결어미에 대한 문장 시퀀싱에 의한 w-vector 를 기반으로 화자인식을 수행하였다. 그 결과 음성 정보에만 의존하여 i-vector, x-vector, 딥러닝 등의 방법을 혼합하여 화자인식 하는 방법을 보완하려고 했다. 본 연구는 종결어미에 기반한 문장 시퀀싱이라는 방법을 제안한 최초의 연구이며, 향후 실존하는 음성인식 시스템과 함께 활용되어 화자인식에 의한 지능형 대화 시스템과 이를 기반한 전자거래 및 각종 음성 기반 서비스에 활용될 것을 기대한다.

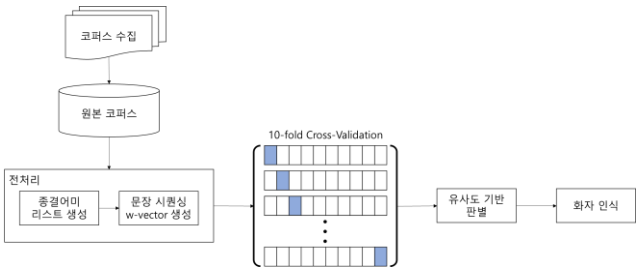
Ⅰ. 서 론

화자인식(Speaker Recognition)이란 특정인의 음성 샘플로부터 그가 누구인지를 자동으로 구별하는 기술이다[1]. 이를 위해 화자인식은 음성 샘플을 주요 입력 데이터로 필요로 한다[2]. 휴대용 기기의 발전과 음성 기술, 오디오 콘텐츠 분야 등이 계속해서 확장됨에 따라, 화자인식 기술의 중요성은 더욱 부각되고 있으며, 일반적인 거래나, 법의학 분야에서도 중요한 수단으로 사용되고 있다[3]. 그동안 음성 파일을 기반으로 하여 그 음성의 화자가 누구인지를 자동으로 판정하려는 화자인식 연구는 화자 판정의 정확도를 올리는 목표를 가지고 진행되어왔다. 음원에서 개선된 특징을 추출하는 접근법이 많이 제안되는데[4], 예를 들어 음성 신호를 네트워크의 입력으로 넣어 주기 전에, VAD(Voice activity detection) 및 feature extraction 과정을 거쳐 잡음을 제거하려는 접근법이 있다[5]. 최근에는 화자인식 성능을 제고하기 위해 CNN 과 같은 딥러닝 모델을 활용하기도 한다[6]. 학습된 CNN 의 feature vector 와 화자와의 유클리드 거리(Euclidean distance)를 기반으로 화자인식 하는 접근은 그 한 예이다[7]. 또한 음성 정보에는 감정 정보가 포함되어 있어, 이를 활용하여 화자인식의 정확도를 높이는 접근도 가능하다[8]. 또한 음성 정보를 낮은 차원의 정보로 변경한 i-vector[9]나 x-vector[10]를 기반으로 일련의 판별자(예: Probabilistic Linear Discriminant Analysis(PLDA))[11]를 활용하여 i-vector 또는 x-vector 에 등장하는 화자가 누구인지를 인식하는 방법을 제안하기도 한다[12]. 또한 GAN 기술을 활용하여 화자인식의 성능을 제고하는 노력도 최근 소개되고 있다[13].

Ⅱ. 본론

본 연구의 목적은 화자인식의 정확도를 개선하기 위해 화자가 발화한 문장에 등장하는 종결어미의 등장 특성을 바탕으로 화자인식 하는 방법을 제안하는 것으로 이를 위해 <그림 1>과 같이 진행하고자 한다. 또한 종결어미의 등장 특성을 바탕으로 화자인식 하는 방법을 제안하는 것이므로 음성 데이터를 텍스트 데이터로 변환하는 STT(Speech-to-Text) 단계는 생략한다. 화자인식을 위한 학습 데이터는 청각장애를 앓고 있는 분들을 위한 한글 자막 제작 모임인 '공이자막(https://blog.naver.com/dr_kisabi)'에서 제공하는 '응답하라 1994'라는 K-Drama 자막 데이터를 다운로드 받아서 사용하였다. 이후 전처리를 위해 다운로드 받은 '.smi' 파일을 '.txt' 파일 형태로 변환하고 먼저 이진 분류(Binary Classification)를 통한 화자인식을 실험하기 위해 남자 주연 1 명, 여자 주연 1 명으로 데이터셋을 구성한다. 생성된 데이터셋을 통해 종결어미 리스트를 생성하기 위해 형태소 분석기 RHINO 3.7 을 사용하였다. 해당 형태소 분석기를 통해 종결어미를 추출하고 중복 값이 존재하지 않게 종결어미 리스트를 생성한다. 문장 시퀀싱(sentence sequencing)이란 학습할 세션(session)에 등장하는 종결어미 빈도와 감성분석 결과를 통해 w-vector 를

생성하는 과정이다. 앞서 생성된 종결어미 리스트를 활용하여 각 화자별로 구성된 세션의 w-vector 를 계산하여 생성한다. 이후 10-Fold Cross-Validation 을 통해 각 세션에 생성된 w-vector 값을 활용하여 Cosine Similarity 계산 후 유사도 기반 판별을 통해 화자를 인식한다.



<그림 1> 전체적인 프로세스

Ⅲ. 결론

본 연구에서 우리는 종결어미에 대한 문장 시퀀싱에 의한 w-vector 를 기반으로 화자 인식하는 방법을 제안했다. 이를 통해 기존의 전통적인 음성인식 후 잡음 요소 제거 등을 수행하고 난 후에 음성 정보에만 의존하여 i-vector, x-vector, 딥러닝 등의 방법을 혼합하여 화자인식 하는 방법을 보완하려고 했다. 이를 위해 실제 존재하는 K-Drama 대본을 가지고 코퍼스를 구축하여 실험하였으며, 그 결과로 문장 시퀀싱 방법은 기존의 화자인식의 정확도를 더욱 개선한바 유용한 딥러닝 전처리 방법이 될 수 있음을 보였다. 본 연구는 종결어미에 기반한 문장 시퀀싱이라는 방법을 제안한 최초의 연구이며 화자인식의 정확도 제고에 유용함을 보인 것이다. 향후 실존하는 음성인식 시스템과 함께 활용되어 화자인식에 의한 지능형 대화 시스템과 이를 기반한 전자거래 및 각종 음성 기반 서비스에 활용될 것을 기대한다.

ACKNOWLEDGMENT

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea(NRF-2018S1A5A2A03036394).

참 고 문 헌

[1] Wang, N., Ching, P. C., Zheng, N., & Lee, T. (2010). Robust speaker recognition using denoised vocal source and vocal tract features. IEEE transactions on audio, speech, and language processing, 19(1), 196-205.

[2] Ramachandran, R. P., Farrell, K. R., Ramachandran, R., & Mammone, R. J. (2002). Speaker recognition—general classifier approaches and data fusion methods. Pattern recognition, 35(12), 2801-2821.

[3] 소순원. (2019). 자유 발화 데이터를 사용한 심층 인공 신경망 기반 화자 정보 분류 모델 개발 (Doctoral dissertation, 한양대학교).

[4] 강지훈, 김보람, 김규영, & 이상훈. (2020). MCE 기반의 다중 특징 파라미터 스코어의 결합을 통한 화자인식 성능 향상. 한국산학기술학회 논문지, 21(6), 679-686.

[5] 채석완. (2019). 교사-학생 학습 방법을 활용한 잡음에 강인한 화자 인식 (Doctoral dissertation, 서울대학교 대학원).

[6] Bhattacharya, G., Alam, M. J., & Kenny, P. (2019). Deep speaker recognition: Modular or monolithic?. In INTERSPEECH (pp. 1143-1147).

[7] 정희승, 윤상혁, & 박능수. (2020). 합성 샴 신경망을 이용한 화자 인식. 전기학회논문지, 69(1), 164-169.

[8] Huanjun, B., X. Mingxing, and F. Z. Thomas, “Emotion Attribute Projection for Speaker Recognition on Emotional Speech”. EUROSPEECH 2007, Antwerp, pp. 758-761, 2007.

[9] Dehak, N., P. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, “Front-end factor analysis for speaker verification,” IEEE Transactions on Audio, Speech, and Language Processing, vol. 19, no. 4, pp. 788-798, 2011.

[10] Garcia-Romero, D., Snyder, D., Sell, G., McCree, A., Povey, D., & Khudanpur, S. (2019). x-Vector DNN Refinement with Full-Length Recordings for Speaker Recognition. In INTERSPEECH (pp. 1493-1496).

[11] Ioffe, S., “Probabilistic linear discriminant analysis,” Computer Vision-ECCV 2006, pp. 531-542, 2006.

[12] Snyder, D., Garcia-Romero, D., Sell, G., McCree, A., Povey, D., & Khudanpur, S. (2019, May). Speaker recognition for multi-speaker conversations using x-vectors. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 5796-5800). IEEE.

[13] Chen, G., Chen, S., Fan, L., Du, X., Zhao, Z., Song, F., & Liu, Y. (2019). Who is real bob? adversarial attacks on speaker recognition systems. arXiv preprint arXiv:1911.01840.

PNCC와 합성곱 신경망을 이용한 능동 소나 표적 식별

이승우, 서익수, 한동석*

국방과학연구소, *경북대학교

swlee06@add.re.kr, seois@add.re.kr, *dshan@knu.ac.kr

Active Sonar Target Discrimination with Power-Normalized Cepstral Coefficients and Convolutional Neural Network

Lee Seung Woo, Seo Ik Su, Han Dong Seog*

Agency for Defense Development, *Kyungpook National Univ.

요 약

본 논문은 능동 소나 해상 실험 데이터에 대하여 특징 정보를 추출하고 Convolutional Neural Network(CNN)를 이용하여 수중 표적을 식별한 결과에 대해 분석하였다. Mel-Frequency Cepstral Coefficients(MFCC)보다 Power-Normalized Cepstral Coefficients(PNCC)를 이용하여 특징 정보를 추출하고 그 결과를 CNN을 이용하여 식별을 하였을 때 더 높은 식별률이 나타남을 확인할 수 있었다.

Ⅰ. 서 론

능동소나를 이용해서 수중 표적을 탐지할 때 수중 표적이 아닌 다른 원인에 의해 반사되어 오는 신호를 클러터 라고 하는데 능동 소나를 이용해서 수중 표적을 탐지할 때는 표적에 비해 클러터가 훨씬 많이 존재하기 때문에 표적에서 반향된 신호와 그렇지 않은 클러터 신호를 식별해서 구분해 주는 것은 매우 어려운 상황이다 [1]-[2]. 본 논문에서는 이러한 문제점을 극복하기 위해 능동 소나 해상실험 데이터에 대하여 PNCC를 이용하여 특징을 추출하고 그 결과에 대해 CNN을 이용하여 식별하는 기법을 제안하였다. 해상 실험 데이터 중 표적 데이터가 부족함에 따라 표적 데이터를 확장하였으며, 표적 식별을 위해 PNCC를 이용하여 능동 소나 반향음으로부터 특징 정보를 추출하고 이를 CNN을 이용하여 표적 식별을 수행하였다. PNCC를 이용한 식별 결과와 비교를 위해 MFCC를 이용하여 특징 정보를 추출하고 CNN을 이용하여 표적 식별을 수행하였다.

Ⅱ. 본론

음성인식 분야에서 널리 사용되고 있는 MFCC에 비해 최근에 개발된 PNCC가 음성인식의 경우 우수하다고 알려져 있으며 PNCC는 MFCC에 비해 잡음에 강인한 특성을 가진다 [3]. 능동 소나를 운용하는 해양 환경은 배경잡음이 큰 환경임에 따라 PNCC를 이용하여 특징을 추출할 경우 MFCC로 특징 추출을 할 경우 보다 표적 식별 성능이 우수하다. 본 논문에서는 능동 소나 빔형성 결과로부터 PNCC를 이용하여 특징을 추출하고, 식별기로 CNN을 이용하는 기법을 제안한다. 인간의 청각 특성과 능동 소나 운용 환경을 고려하여 특징을 추출하고 그 결과를 이미지화 하여 이미지 식별 분야에서 최근 널리 사용되고 있는 CNN을 식별기로 이용하였다. 실험에 사용된 데이터는 동해 해상에서 능동 소나를 이용한 해상실험 데이터이며 이를 이용한 식별 결과는 표 1과 같다. 표 1로부터 표적과 클러터에 대한 식별 결과 MFCC 보다 PNCC를 이용해서 특징을 추출하고, 이를 CNN을 이용하여 식별을 하였을 경우 식별률이 더 높음을 알 수 있다.

표 1. MFCC와 PNCC에 대한 식별 결과.

특징 추출	구분	식별 결과	
		표적	클러터
MFCC	표적	97.234 %	2.766 %
	클러터	1.493 %	98.507 %
PNCC	표적	98.617 %	1.383 %
	클러터	0.896 %	99.104 %

Ⅲ. 결론

본 논문에서는 능동 소나 해상실험 데이터를 이용하여 수중 표적을 탐지할 경우 PNCC와 MFCC를 이용하여 특징추출을 수행하고 그 결과에 대해 CNN을 이용해서 식별하였을 때 각각의 식별률에 대해 비교하였다. 해상실험 데이터 중 표적 데이터 확보가 제한적임에 따라 데이터 확장기법을 이용하여 데이터의 양을 증가시켰으며, 이를 이용하여 식별할 때 MFCC 보다 PNCC를 이용할 경우 표적과 클러터에 대한 식별률이 높음을 확인할 수 있었다. 향후 더 다양한 해양환경에서의 해상실험 자료를 획득하여 각각의 경우에 대한 식별률을 비교해 볼 예정이다.

참 고 문 헌

[1] R. Harrison, C. Yang, C.-F. Lin, T. Politopoulos, and E. Chang, “Classification of Underwater Targets with Active Sonar,” Proceedings of IEEE AEROCS, pp.534-538, Westlake Village, 1993.

[2] J. G. Kelly, R. N. Carpenter, J. A. Tague, and N. K. Haddad, “Optimum Classification with Active Sonar: New Theoretical Results,” Proceedings of ICASSP 1991, vol. 2, pp.1445-1448, Toronto, 1991.

[3] C. Kim and R. M. Stern, “Power-Normalized Cepstral Coefficients (PNCC) for Robust Speech Recognition,” IEEE Trans. Audio, Speech, Lang. Process., VOL. 24, NO. 7, July 2016, pp.1315-1329.

가사 Context 기반 음악간 유사도 산출에 관한 연구

박예은, 홍기석, 지봉준, 조현보*
포항공과대학교, 포항공과대학교, 포항공과대학교, *포항공과대학교

yeunpark@postech.ac.kr, kisuk13577@postech.ac.kr, bongari82@postech.ac.kr,
*hcho@postech.ac.kr

Similarity Evaluation of Music based on Context of Lyrics

Ye Un Park, Ki Suk Hong, Bong Jun Ji, Hyun Bo Cho *
POSTECH, POSTECH, POSTECH, *POSTECH

요 약

음악은 가장 인기 있는 콘텐츠 중 하나로 시장의 규모가 꾸준히 증가하고 있으며, 최근에는 다양한 산업에서도 활용되고 있다. 이에 따라, 음악의 특징, 사용자의 기호 등에 따라 음악을 추천하는 서비스에 관한 관심이 증가하고 있다. 하지만 시장의 규모가 커짐에 따라 추천할 수 있는 음악의 선택지는 많아졌으나 수많은 선택지 중에 어떤 음악을 어떤 방법으로 선정하여 추천해야 하는지에 관한 어려움이 있다. 음악 추천 기법 중 아이템 기반 필터링은 사용자가 선호하는 음악과 유사한 음악을 추천하는 방식으로, 음악간 유사도를 계산하는 것이 주요 과제이다. 이를 위해 음악의 특성을 통해 유사도를 산출하는 방법에 관한 연구가 필요하며, 본 논문에서는 음악의 가사를 기반으로 음악간 유사도를 산출하는 모델을 제안하고자 한다.

Ⅰ. 서 론

최근 음악 콘텐츠 규모 증가와 정보통신 기술로 인한 접근성 향상으로 음악 콘텐츠에 접근하기는 쉬워졌으나, 사용자가 음악에 대한 모든 정보를 사전에 취득하고 선별하여 사용하기 어려워졌다. 따라서, 사용자들이 편리하게 음악 콘텐츠를 사용하기 위해 음악을 추천해주는 서비스가 증가하고 있다. 음악을 추천하는 다양한 방법이 있으나, 사용자가 선호하는 음악과 유사한 음악을 추천하기 위해서는 음악간 유사도를 계산하는 것이 주요 과제 중 하나이다. 본 논문에서는 가사의 Context 를 활용하여 음악간 유사도를 산출하는 모델의 프레임워크를 제안하고, 실험을 통해 제안한 프레임워크의 Feasibility 를 검증하고자 한다.

Ⅱ. 본론

가사의 Context 를 활용하여 음악간 유사도를 산출하는 프레임 워크는 Figure1 과 같이 모델 학습 과정과 모델 실행과정으로 구성된다.

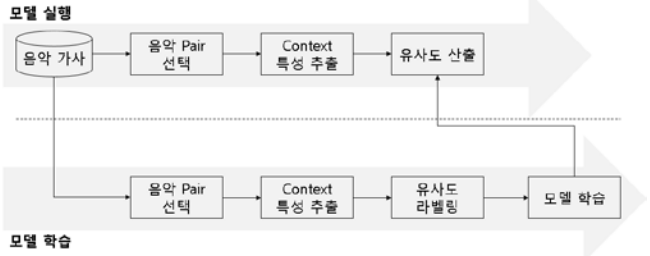


Figure 1

모델의 학습과정에서는 음악 가사 데이터에서 무작위로 음악 Pair 를 선택한 후의 음악 간 Context 를 표현하는 특성을 추출한다. 선택한 음악 Pair 가 서로 유사하다고 판단되는 경우는 1, 유사하지 않다고 판단되는 경우는 0 으로 라벨링을 한다. 이후에는 음악 pair 의 특성과 라벨링 된 유사도 간의 상관관계를 학습하여 유사도 산출 모델을 개발한다. 모델의 실행과정에서는 동일하게 음악 Pair 를 선택한 후, 특성을 추출한다. 추출된 특성을 미리 학습된 모델에 입력하여 음악 간 유사도를 산출한다.

Ⅲ. 결론

본 논문에서는 가사의 Context 를 활용하여 음악간 유사도를 산출하는 프레임워크를 제안하였다. 가사를 통해 음악의 Context 를 표현하는 특성을 추출하여 라벨링 된 유사도와 특성 간의 상관관계를 학습하여 이를 통해 음악 간 유사도를 산출하는 유사도 산출 모델을 개발하였다. 본 논문에서 제안한 프레임워크는 사용자에게 음악을 추천할 때에 유사한 음악을 탐색하고 선별하는 데에 사용할 수 있을 것으로 기대된다.

ACKNOWLEDGMENT

이 연구는 한국산업기술진흥원의 광역협력권 산업육성사업 (P0002315) 내 "자율주행 자동차용 스마트 GLOVE BOX 편의장치 개발" 과제의 지원을 받아서 수행되었다.

참 고 문 헌

[1] 성보경, 정명범, and 고일주. "음악 특징점간의 유사도 측정을 이용한 동일음원 인식 방법." 한국컴퓨터정보학회논문지 13.3 (2008): 99-106

[2] 임성수, and 조성배. "사용자 감정 및 환경을 고려한 퍼지추론 기반 음악추천 시스템." 한국정보과학회 학술발표논문집 31.2Ⅱ (2004): 541-543.

[3] 이승준, 서봉군, and 박도형. "소비자 감성 분석 기반의 음악 추천 알고리즘 개발." 지능정보연구 24.4 (2018): 197-217

[4] 이재환. 가사의 감정 분석과 구조 분석을 이용한 노래간 유사도 측정. Diss. 서울대학교 대학원, 2016

[5] 이세환, 이주환. "음악청취 상황에 따른 감성 기반 음악 추천 모델 평가 연구: 스포티파이 음악추천 알고리즘 중심으로." 한국디지털콘텐츠학회 논문지 21.7 (2020): 1301-1309

[7] Patra, Braja Gopal, Dipankar Das, and Sivaji Bandyopadhyay. "Retrieving similar lyrics for music recommendation system." Proceedings of the 14th International Conference on Natural Language Processing (ICON-2017). 2017.

효율적인 추천 시스템을 위한 학습 컨텐츠의 상대적 특성 업데이트 방법 연구

운봉영, 아가르왈 판카즈
㈜태그하이브

wbyoung@tag-hive.com, pankaj@tag-hive.com

A Method to Update the Relative Characteristics of Learning Contents For Efficient Recommendation System

Woon BongYoung, Agarwal Pankaj
TagHive, Inc.

요 약

본 논문은 추천 시스템의 핵심 요소인 컨텐츠 관리 방법에 대하여 연구한다. 학습 컨텐츠의 경우 난이도, 스킬, 챕터 등의 특성을 부여 받는데 사용자 군집의 실력에 따라 상대적인 특성이 발생하게 된다. 이를 해결하기 위하여 사용자 군집의 문제 풀이 데이터에 기반하여 사전에 부여된 난이도를 업데이트하고, 로지스틱 회귀모델을 활용하여 문제 정답 확률 예측을 위한 학습의 결과로 생성된 스킬 분포 값을 통해 스킬 적합도를 판단하는 방법을 제안한다.

Ⅰ. 서론

전 세계적으로 코로나 시대가 도래하면서 다양한 사회의 모습이 변화하고 있다. 교육 시장의 경우 대면 학습이 어려워지면서 온라인 교육 시장이 급격히 성장하였고 이와 더불어 에듀테크(Edu-Tech)에 관심도 높아지고 있다. 이전의 온라인 교육(이하 비대면 학습)의 경우 소비자가 사전에 만들어진 컨텐츠를 온라인으로 시간의 자유롭게 소비할 수 있다는 장점이 주요했지만 최근에는 각 사용자의 실력 및 특성 등을 정확하게 분석하고 이를 기반으로 적절한 컨텐츠 및 학습 방향을 추천하며 학습의 효율성을 극대화하는 다양한 개인화 추천 시스템이 개발되고 있다[1].

개인화 추천 시스템의 성능을 높이기 위해서는 다양한 유저에 대한 정확한 특성 분석과 양질의 많은 컨텐츠가 요구된다. 학습 컨텐츠의 경우 난이도, 스킬 등의 특성들이 정확하게 부여되어 있어야 하지만 컨텐츠 개발자의 주관 또는 컨텐츠 사용 분야, 시기에 따라 상대적인 차이가 발생한다. 예를 들어 현재 가장 높은 난이도를 부여 받은 컨텐츠는 사용자 군집의 기본 실력 수준에 따라 상대적으로 낮은 레벨로 분류 될 수 있고, 시간이 지나면서 전체적인 교육 수준의 향상으로 인하여 낮은 레벨로 분류 될 수 있다. 또한, 새로운 교육 정책에 따라 챕터 구성이 달라지고 스킬 이름이 달라지는 경우 기존 컨텐츠에 부여된 특성을 수정해야만 한다. 이처럼 추천 시스템을 높은 성능 확보를 위해서는 환경에 따라 발생하는 상대적인 특성을 반영하는 것이 중요하다.

이에 따라 본 논문에서는 사용자 데이터 통계를 활용하여 사용자 군집 실력에 따라 난이도를 업데이트 하고 스킬의 적합도를 판단하여 컨텐츠의 상대적 특성을 효율적으로 반영할 수 있는 방법을 제안한다.

Ⅱ. 본론

본 논문에서는 데이터에 기반하여 사용자 군집의 실력의 변화에 따라 컨텐츠 난이도를 업데이트하고 정답 확률 예측 모델을 활용하여 컨텐츠의 특성 적합도를 판단하는 방법을 제안한다.

1) 난이도 업데이트 시스템

난이도 업데이트는 Fig 1 의 순서로 이루어진다. 먼저 현재 사용자 군집의 문제 풀이 데이터를 수집하고 각 문제의 정답자, 오답자의 평균 실력을 계산한다. 사용자의 실력은 문제에 대한 지식 수준 정도를 구분하고 문제의 정답 유무에 따라 레벨을 부여한다[2]. 이후 새로운 난이도는 식 1 과 같이 생성한다. 계산식에서 해당 문제의 정답 확률과 오답자의 평균 실력을 곱하는 이유는 정답률이 높은 경우 현재 사용자 군집에게는 해당 문제가 쉬운 문제로 받아들여지기 때문에 난이도를 낮춰야 하고, 반대로 정답률이 낮은 경우 현재 사용자 군집에게 해당 문제가 어렵게 받아들여져 난이도를 높이는 방향으로 업데이트 되기 때문이다.

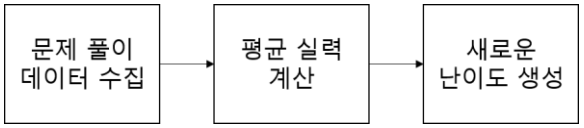


Fig 1. 난이도 업데이트 시스템 블록도

$$\Delta = (\beta * \mu) + ((1-\beta) * \nu) \quad \cdots\cdots\cdots\cdots\cdots\cdots \text{식 1}$$

Δ = 새로운 난이도
β = 해당 문제의 정답률
μ = 오답자 평균실력
ν = 정답자 평균 실력

최종 생성된 새로운 난이도는 현재 난이도와 λ 매개변수를 통해 평균을 계산한다. 단일 횟수로 난이도를 업데이트 하는 경우 현재 사용자 군집의 실력 분포에 의존도가 높아지는 문제가 발생하기 때문에 문제 풀이 데이터가 충분히 쌓였을 경우 난이도를 복수 횟수로 재 업데이트하여 정규분포를 고려한다.

2) 스킬 적합도 판단 시스템

본 논문에서의 스킬은 어떠한 과목 챕터의 하위 개념이며 문제를 풀기 위해 필요한 능력을 일컫는다. 스킬의 적합도를 판단은 Fig 2 와 같은 과정으로 진행한다. 현재 사용자 군집의 문제 풀이 데이터를 수집하고 정답 확률 예측 모델을 학습한다. 학습 결과로 얻은 데이터 기반의 스킬 분포 값과 현재 컨텐츠에 부여된 스킬을 비교하여 스킬의 적합도를 분석한다.

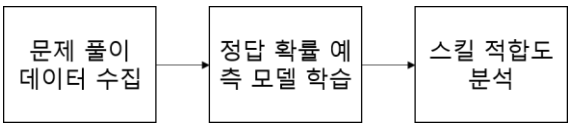


Fig 2. 스킬 적합도 판단 시스템 블록도

정답 확률 예측 모델의 경우 특정 사용자가 아직 풀지 않은 어떠한 문제에 대하여 정답을 맞출 수 있는 확률을 계산하기 위하여 사용되었으며 대표적으로 아래와 같은 방법들이 있다.

① Collaborative filtering: 어떠한 사용자가 새로운 문제를 접할 때 기존 사용자 군집에서 현재 사용자와 비슷한 특성을 지닌 사용자의 사전 정보를 활용하여 예측
② Logistic regression: 사용자 군집의 정보를 학습하여 어떠한 문제를 맞출 확률에 대한 예측

본 연구에서 활용하는 정답 확률 예측 단계는 학습 결과로 얻은 스킬 분포 값을 활용하는 것이 주된 목적이다.

따라서 현재 사용자 군집의 문제 풀이 데이터를 기반으로 컨텐츠의 특성을 분석하기 위한 방법이 요구하고 사전 유사 군집의 정보 없이 독립적인 학습을 진행하기 때문에 로지스틱 회귀 모델을 이용하여 사용자의 문제 정답 확률을 예측하였으며, 현재 사용자 군집(U)의 스킬별 실력(S) 데이터(U-S matrix)와 각 문제(Q)의 스킬 분포(S') 초기 랜덤값(Q-S' matrix)을 입력 데이터로 사용하고 문제 풀이 데이터 Y-matrix 를 학습 목표 데이터로 활용한다. 식 2 에 따라 생성한 정답 확률 예측 값을 Y-matrix 와 비교하여 학습하게 되고 스킬 적합도를 판단하기 위해서 Q-S' matrix 를 학습 매개변수로 사용한다. U-S matrix 에는 각 사용자의 스킬의 현재 실력 값이 들어가고, Q-S' matrix 의 경우 각 문제에 대한 랜덤 초기 값을 갖는다. 또한 Y-matrix 의 경우 사용자가 해당문제를 맞춘 경우 1, 틀린경우 0 으로 값을 지정한다.

본 연구는 정답 확률 예측이 아닌 스킬 분포 값을 계산하는 것이 목적이기 때문에 Fig 3 과 같이 학습을 수행하고 나면 매개변수인 Q-S' matrix 는 각 문제에 대한 스킬 분포 값을 얻게 되고 기존의 문제에 부여된 스킬의 개수에 따라(본 연구팀 컨텐츠의 경우 보통 1~2 개) 각 문제의 가장 높은 peak 값 1 개 혹은 2 개 값을 갖는 스킬을 추출한다.

U-S matrix * Q-S' matrix = 정답 확률 예측……식 2

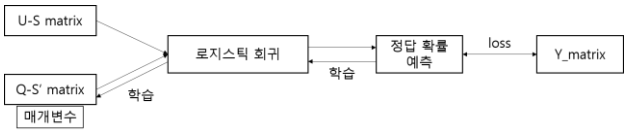


Fig 3. 정답 확률 예측 시스템 블록도

본 과정은 사용자 군집의 실력 값과 학습 결과로 생성된 문제의 스킬 분포 값을 통해 사용자의 문제 정답 확률을 정확하게 예측 한다면 매개변수인 Q-S' matrix 는 해당 문제에 대해 적절한 스킬 분포 값이 부여되었다고 판단할 수 있다. 따라서 학습된 Q-S' matrix 에서 peak 값을 갖는 스킬과 실제 각 문제에 부여된 스킬을 비교함으로써 스킬 적합도를 판단하여 스킬 특성을 업데이트 한다.

Ⅲ. 실험 결과

본 챕터에서는 앞에서 제안한 1) 난이도 업데이트 시스템, 2) 스킬 적합도 판단 시스템의 실험 및 결과에 대하여 설명한다. "Class saathi"는 인도의 6-10 학년의 수학, 과학 문제를 제공하며 초반 유저의 실력을 분석한 뒤 Fig4 와 같이 실력에 맞는 문제를 추천하여 단기간 실력 향상에 도움을 주는 앱으로 본 실험은 "Class Saathi" 앱의 데이터에 기반한다.

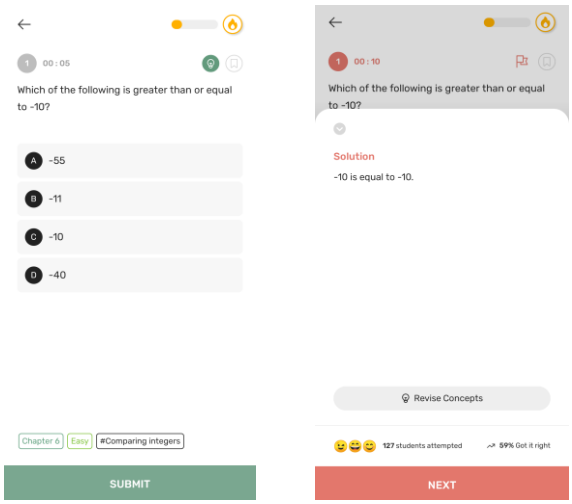


Fig 4. Class Saathi 퀴즈 화면 예시

1) 난이도 업데이트 실험

Fig4 의 왼쪽 그림에서와 같이 각 문제는 각기 다른 난이도를 가지고 있고, 앱의 화면에서는 Easy, Medium, Hard 로 표기되는 것과 달리 DB 에는 0~1 의 값으로 상세하게 분류 되어 있다. 각 문제를 접하는 사용자 그룹군의 실력에 적합한 난이도를 재부여 하기 위해서 최소 100 번 이상 풀어진 문제에 대하여 난이도 업데이트를 실시한다.

문제 ID (5 개 예시)	기존 난이도	새로운 난이도
06AeF110e0sBT8SOwgXp	0.66	0.527856

7THy1ZZsWAXJmcsnvk6B	0.33	0.516037
02XoWLZtk26RCt5SEtV1	0.33	0.612946
hg32mmDvbQkbgWR5EHtL	0.66	0.48694
01K10LcUw6TVU4trmKXb	0.2	0.319829

Table 1. 문제 난이도 업데이트 결과

식 1 에 따라 문제의 새로운 난이도를 생성한 결과는 Table 1 과 같다. 이를 통해 기존 문제의 난이도는 새로운 난이도로 대체하여 부여되며 유저의 문제 풀이 기록이 충분히 쌓인 뒤 난이도가 업데이트 된 문제의 데이터 통계를 분석한다. 분석 기준은 각 문제의 정답율이 약 70%가 되는 것이 목표이다.

문제 ID (5 개 예시)	정답율 (%)	유저에 의해 풀어진 횟수
06AeF110e0sBT8SOwgXp	74.8	127
7THy1ZZsWAXJmcsnvk6B	78.23	124
02XoWLZtk26RCt5SEtV1	76.42	106
hg32mmDvbQkbgWR5EHtL	64.71	102
01K10LcUw6TVU4trmKXb	64.58	96

Table 2. 난이도 업데이트 문제의 정답율

각 문제는 난이도가 업데이트 된 이후의 유저에 의해 풀어진 횟수가 많은 순으로 나열되어 있으며, 정답율은 최저 64%에서 최고 74%로 목표 정답율인 70%에 근접하였다. 각 문제들은 분기 혹은 풀어진 횟수에 따라 1 년 기준으로 약 4-6 회 업데이트 되어 사용자 그룹군의 실력의 정규 분포에 맞는 난이도를 부여 받게 된다.

2) 스킬 적합도 판단 실험

Fig 4 와 같이 각 문제에는 스킬이 1 또는 2 개 부여되어 있다. 업데이트 되어 사용자 그룹군의 실력의 정규 분포에 맞는 난이도를 부여 받게 된다. 해당 스킬들은 문제 생성자에 의해 사전에 부여 되며, 본 실험에서는 유저의 문제 풀이 데이터를 활용하여 이에 대한 적합도를 판단한다.

Logistic model 을 활용하여 Q-S' matrix 를 학습하며 이후 결과는 Fig5 와 같다. 왼쪽은 실제 스킬이 부여된 인덱스 번호이고 오른쪽은 학습 뒤 얻어진 Q-S'의 결과이다. 이처럼 각 문제에 대하여 얻어진 Q-S'의 결과 값을 통하여 실제 문제의 스킬 인덱스와 비교하며 스킬 적합도를 판단 할 수 있다. 하지만 Fig6 과 같이 스킬 적합도의 성능의 경우 해당 문제가 유저에게 풀어진 데이터 셋의 크기에 따라 성능의 차이가 발생할 수 있다.

Fig6 은 해당 문제의 Q-S' matrix 의 학습이 완료된 전체 값을 보여준다. 일반적으로 학습된 Q-S' matrix 의 max 값을 취해 비교하고 문제에 스킬이 복수로 부여되어 있는 경우 max 2 개 값까지 추출 한 뒤 비교한다.

Fig 5. 문제(v1iNuUfvpd8NJqnhqUni)에 부여된 스킬 인덱스와 Q-S' matrix 값

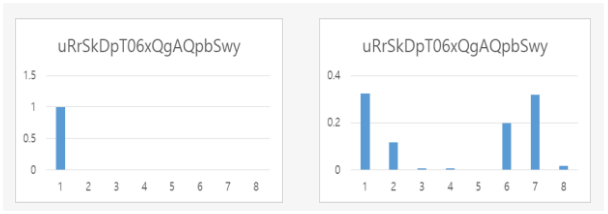


Fig 6. 문제(uRrSkDpT06xQgAQpbSwy)에 부여된 스킬 인덱스와 Q-S' matrix 값

Ⅳ. 결론

본 논문에서는 추천시스템의 높은 성능을 확보하는데 기반이 될 수 있도록 컨텐츠의 상대적 특성을 업데이트 하는 방법에 대하여 소개하였다. 본 연구는 현재 사용자 군집의 데이터에 기반하여 새로운 난이도를 부여하고 컨텐츠의 스킬 적합도를 판단함으로써 컨텐츠가 갖는 사용자 군집의 상대적 특성 차이 의존도를 낮추고 활용도를 높였다. 각 사용자의 군집의 데이터를 충분히 수집하면 제안 방법의 효과를 극대화할 수 있으며 난이도 업데이트 시스템, 스킬 적합도 판단 시스템 모두 단일 횟수의 적용 보다는 복수의 횟수를 반복함으로써 사용자 군집의 특성이 정규 분포 형태로 반영될 수 있도록 하는 것이 중요하다. 추후 연구에서는 단일 횟수의 업데이트 만으로 상대적 특성을 업데이트 하고 대량의 사용자 군집 데이터를 실시간으로 활용할 수 있는 방법에 대하여 논의한다.

ACKNOWLEDGEMENT

본 연구는 2020 년도 중소벤처기업부의 기술개발사업 지원에 의한 연구임. [S2814142]

참 고 문 헌

[1] 이석준, and 이희춘. "협업 필터링 추천에서 대응평균 알고리즘의 예측 성능에 관한 연구." *Information Systems Review* 9.1 (2007): 85-103.

[2] Jia, Bing, Yongjian Yang, and Jun Zhang. "Study on Learner Modeling in Adaptive Learning System." *JCP* 7, no. 10 (2012): 2585-2592.

Inherent Overestimation of DRL-Based Hybrid Beamforming for mmWave MIMO Systems: Behavioral Interpretation and Remedies

Dohyun Kim
Department of Electrical and Computer Engineering
University of Texas at Austin
Austin, TX 78712, United States
Email: dohyun.kim@utexas.edu

Robert W. Heath Jr.
Department of Electrical and Computer Engineering
North Carolina State University
Raleigh, NC 27695, United States
Email: rwheathjr@ncsu.edu

***Abstract*—Recently, many machine learning-based hybrid beamforming algorithms have been studied to implement the practical mmWave dense MIMO systems with high spectral efficiency. Hybrid beamforming algorithm based on deep reinforcement learning (DRL), is claimed to be the state-of-the-art technique regarding the computation time to achieve high spectral efficiency. Nonetheless, DRL is known to suffer from *overestimation*, which reinforces the algorithm to converge to a suboptimal behavior. Herein, we investigate overestimation in DRL-based hybrid beamforming using the angle representation of analog precoder. We discuss possible directions, based on the behavioral interpretation, to handle the overestimation.**

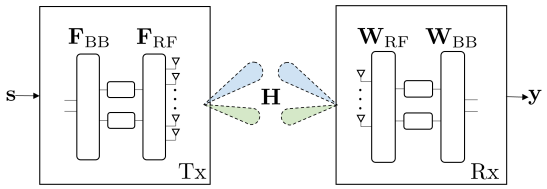
I. Introduction

Hybrid beamforming (HBF) enables the practical implementation of mmWave dense MIMO systems with high spectral efficiency [1]. Its effectiveness comes from the separation of the analog/digital domain, reducing the number of costly components such as converters between the analog/digital domain. Among a myriad of work applying the modern machine learning tools to HBF, in this paper, we focus on algorithms based on *deep reinforcement learning* (DRL). The benefit of DRL-based HBF is the short online computation time and robustness to channel estimation error [2].

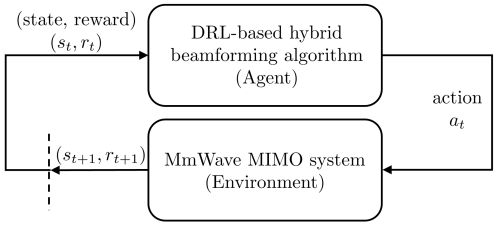
Overestimation in DRL is a well-known issue, that can make the algorithm converge to a suboptimal behavior. Moreover, overestimation is inherent and exists whenever the function estimator is imprecise [3]. Double Q-learning [4] is known as a ubiquitous solution for DRL with discrete states, but it is not suitable for DRL with continuous states which correspond to dense MIMO systems. Meanwhile, clipped double Q-learning [5] can handle DRL with continuous states.

To the best of the authors' knowledge, overestimation in DRL-based HBF has not been studied. Herein, we observe the overestimation behaviors of inherent states and discuss possible remedies.

The contribution of this paper are the following:



(a) Configuration of HBF for dense MIMO system



(b) Algorithm flow of DRL for HBF

Fig. 1: DRL-based HBF concepts: (a) Hardware, (b) Algorithm

- We properly observe and interpret the overestimation behavior of the inherent state in DRL-based HBF algorithms.
- We discuss remedies to the overestimation, providing experimental results on toy examples. We sketch to provide intuition towards an extension to practical problems.

II. Overestimation problem in DRL-based HBF

We implement an exemplary DRL-based HBF, similar to [2], with only one learning parameter $\mathbf{F}_{\mathrm{RF}}^{(t)}$. We consider a 128 by 16 MIMO system with 2 radio frequency chain and data stream in Figure 1a. We apply the learning model in Figure 1b with state $s_t = \{\mathbf{F}_{\mathrm{BB}}, \mathbf{F}_{\mathrm{RF}}^{(t-1)}, \mathbf{W}_{\mathrm{RF}}, \mathbf{W}_{\mathrm{BB}}\}$, action $a_t = \{\mathbf{F}_{\mathrm{BB}}, \mathbf{F}_{\mathrm{RF}}^{(t)}, \mathbf{W}_{\mathrm{RF}}, \mathbf{W}_{\mathrm{BB}}\}$, and reward r_t that corresponds to s_{t+1} and channel $\mathbf{H}$. We consider a narrowband channel model with $N_p = 2$ path clusters with angle of arrival vector $(0, \frac{\pi}{16})$ and angle of departure vec-

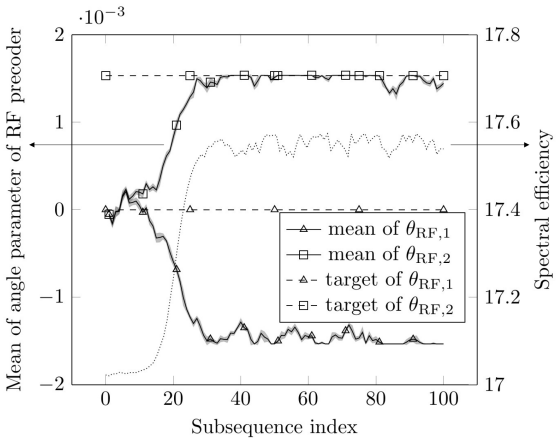


Fig. 2: Statistical interpretation of angle parameter $\boldsymbol{\theta}_{\mathrm{RF}}$ of RF precoder (per subsequence for selected SNR of 5dB and $\tau = 0.5 \cdot 10^{-3}$)

tor $(0, \frac{\pi}{128*16})$. We represent $\mathbf{F}_{\mathrm{RF}}^{(t)}$ with angle parameters $\theta_{\mathrm{RF},1}$ and $\theta_{\mathrm{RF},2}$ following the beam steering fashion [6].

Figure 2 depicts the overestimation issue in DRL-based HBF. The spectral efficiency staggers up to Subsequence 10, reaching almost 17. After Subsequence 15, we observe an increase in spectral efficiency, linearly up to Subsequence 30. Again, after Subsequence 40, we observe negligible oscillation. The means of angle parameter $\theta_{\mathrm{RF},1}$ and $\theta_{\mathrm{RF},2}$ are both zero at Subsequence 1. The mean of angle parameter $\theta_{\mathrm{RF},2}$ tends to increase up to Subsequence 30, where it lies near its target. However, the mean of angle parameter $\theta_{\mathrm{RF},1}$ does not converge to its target. It tends to decrease up to Subsequence 30, where it lies near $-1.5 \cdot 10^{-3}$ not converging it to its target of zero. We interpret that the overestimation of the state $\theta_{\mathrm{RF},1} = -1.5 \cdot 10^{-3}$ is the main source leading to suboptimal behavior. To be specific, the baseline starting with $\theta_{\mathrm{RF},1} = 0$, explores $\theta_{\mathrm{RF},1}$ around 0. Due to the imprecision of function estimator, the value of a negative $\theta_{\mathrm{RF},1}$ becomes higher than that of $\theta_{\mathrm{RF},1} = 0$. The overestimation of value induces a poor policy to select negative $\theta_{\mathrm{RF},1}$. The poor policy then results in a bad estimation of value. Overall, the overestimation accumulates throughout the recursive update of value [3]. We observe the accumulated error as a "drift" in $\theta_{\mathrm{RF},1}$, causing its tendency of decreasing.

III. Control of overestimation in DRL

Overestimation in DRL with discrete states can be improved by the use of separate networks, respectively for selecting and evaluating an action in the max operator of value updates [4]. Similarly, for continuous states, deep deterministic policy gradient (DDPG) separately trains target networks and online networks in an actor-critic learning fashion [7]. The target networks are delayed copies of online networks, where a parameter τ controls the delay. For stable learning, DDPG requires a small τ. The small τ, however, slows the change of target actor network, eventually making the target networks and online networks similar. Therefore, the practical use of DDPG needs further solution of overestimation.

On one hand, using the *minimum value estimate* of two separate networks is a quick remedy of overestimation in DDPG, at the cost of additional computation from the extra networks [5]. On the other hand, *multi-step bootstrapping* method [8] in the value estimate without additional networks introduces underestimation that needs further investigation in DRL-based HBF. Overall, the behavioral interpretation of DRL-based HBF using angle representation is interesting to observe the effect of remedies introduced by [5], [8].

IV. Conclusive Remark

Interpretation in DRL-based HBF is important, in the sense that it allows us to observe the behavioral details more than just its explicit performance. We illustrated the overestimation behavior in DRL-based HBF, with an exemplary implementation using angles, which is not explicit based on observed spectral efficiency. As a specific result, the poor "beam steering angle" behaviors accumulate overestimation errors, eventually lead to a suboptimal value of spectral efficiency throughout the learning process.

The behavioral interpretation, and remedies of the overestimation, of DRL-based HBF using angle representation make more margin of the tradeoff between computation time and spectral efficiency. Using the minimum value estimate and multi-step bootstrapping method can be further combined to control the overestimation.

Acknowledgment

This work was partially supported by the U.S. Army Research Labs under grant W911NF-19-1-0221.

References

[1] R. W. Heath *et al.*, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE journal of selected topics in signal processing*, vol. 10, no. 3, pp. 436–453, 2016.

[2] Q. Wang *et al.*, "Precodernet: Hybrid beamforming for millimeter wave systems with deep reinforcement learning," *IEEE Wireless Communi. Letters*, vol. 9, no. 10, pp. 1677–1681, 2020.

[3] S. Thrun and A. Schwartz, "Issues in using function approximation for reinforcement learning," in *Proc. Connectionist Models Summer School Hillsdale, NJ. Lawrence Erlbaum*, 1993.

[4] H. Van Hasselt *et al.*, "Deep reinforcement learning with double Q-learning," *arXiv preprint arXiv:1509.06461*, 2015.

[5] S. Fujimoto *et al.*, "Addressing function approximation error in actor-critic methods," *arXiv preprint arXiv:1802.09477*, 2018.

[6] O. El Ayach *et al.*, "The capacity optimality of beam steering in large millimeter wave mimo systems," in *IEEE Int. Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2012, pp. 100–104.

[7] T. P. Lillicrap *et al.*, "Continuous control with deep reinforcement learning," *arXiv:1509.02971*, 2015.

[8] L. Meng *et al.*, "The effect of multi-step methods on overestimation in deep reinforcement learning," *arXiv:2006.12692*, 2020.

Performance Analysis of Hybrid Deep Learning Model for Indoor localization

Alwin Poulose and Dong Seog Han*
School of Electronics Engineering, Kyungpook National University,
Daegu, Republic of Korea
alwinpoulosepalatty@knu.ac.kr, dshan@knu.ac.kr*

Abstract

Localization using Wi-Fi received signal strength (RSSI) signals gives accurate results for indoor localization. However, the signal interference, multipath effects and non-line of sight conditions (NLOS) from the indoor experiment area degrade the localization performance. To compensate for these localization challenges that exist in Wi-Fi RSSI based localization systems, we propose a hybrid deep learning model (HDML) based localization system which uses RSSI heat maps instead of raw RSSI signals. The HDLM in the proposed system utilizes the combined form of convolutional neural network and long short-term memory network (CNN-LSTM) architecture and improves the system's localization performance. The experiment results and analysis show that the proposed HDML based localization system reduces the localization error with the help of RSSI heat maps and gives better localization performance than CNN and LSTM models. The proposed architecture archives 88% model accuracy for localization than other deep learning models.

Ⅰ. Introduction

Localization using Wi-Fi received signal strength (RSSI) [1] is an effective localization approach when the inertial measurement unit (IMU) sensor based [2] or camera-based [3] localization systems show high margins of localization errors. In Wi-Fi RSSI based localization approach, the system utilizes access points (APs) in the experiment area and estimates the user distance from APs. The localization accuracy of the Wi-Fi RSSI based localization systems depends on the accurate user distance estimation from APs. To estimate the distance from APs, a free space path loss model (FSPLM) [4] is used which utilize the RSSI values from APs. The RSSI signal values from APs are easily fluctuate with indoor channel conditions such as multipath effects, non-line of sight (NLOS) conditions and signal interferences. To stabilize the RSSI data from APs and enhance the indoor localization performance, we propose a localization system which uses the RSSI heat maps instead of raw RSSI values and estimates the user position. The proposed system feed the heat maps into a hybrid deep learning model (HDLM) and predicts the user x and y position values. The HDLM is a combined form of convolutional neural network and long short-term memory network (CNN-LSTM) [5] and estimate the user position accurately.

In this paper, we analyze the performance of the proposed HDLM for indoor localization. The proposed HDLM gives accurate localization results for RSSI heat maps and improves the localization performance. To validate our proposed system performance, we compare the proposed HDLM performance with CNN and LSTM models and analyses the model accuracy for indoor localization. Through extensive experiments and result analysis, we demonstrate the superior performance of the proposed HDLM with CNN and LSTM models. The rest of the paper is organized as follows; Section II presents the proposed indoor localization system using Wi-Fi RSSI heat maps. In Section III, we discussed the experiment results and analysis of the proposed localization system and Section IV concludes the paper.

Ⅱ. Proposed Indoor Localization System Using Wi-Fi RSSI Heat Maps

The proposed indoor localization system effectively utilizes the advantage of RSSI heat maps and reduced the localization error. The RSSI heat map-based localization approach reduced the localization challenges faced by Wi-Fi RSSI signals and gives better localization performances. Fig. 1 shows the proposed indoor localization system using Wi-Fi RSSI heat maps [6].

In Fig. 1, the proposed system collects the RSSI data from APs and generate RSSI heat maps for each location. The unique characteristics of the each RSSI heat map are a useful information for the localization and each heat map pattern represents the x and y location values of a particular location from the experiment area. The generated heat maps are used as input for the HDLM and the model predicts the user x and y positions. The HDLM in the proposed system uses a CNN-LSTM architecture and train the CNN-LSTM with heat maps. The model after training is ready for location prediction and gives the user's x

and y positions for test data. As compared to CNN and LSTM models, the proposed HDLM uses the combined features of CNN and LSTM models and gives the best results for indoor localization. The CNN model in the HDLM gives the classification results with spatial features. The output of the CNN model is given to the input of the LSTM model and which predicts the location results with sequential data. The HDLM model uses the image as the input and gives user's location information.

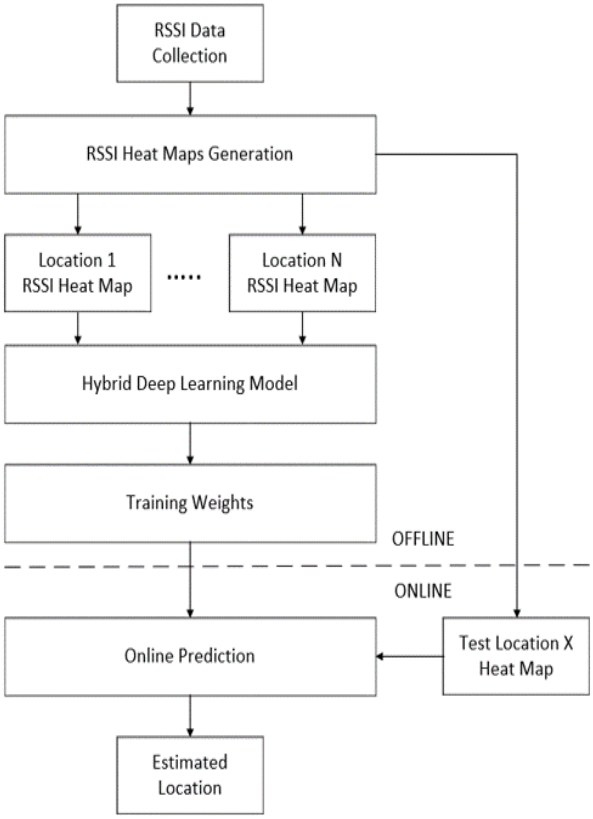


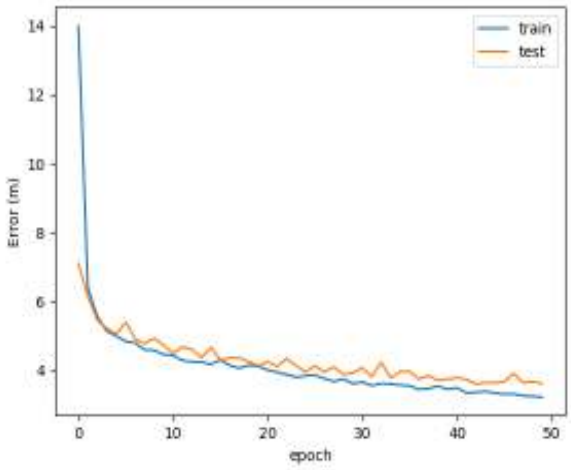
Fig. 1 Proposed indoor localization system using Wi-Fi RSSI heat maps.

Ⅲ. Experiment and Result Analysis

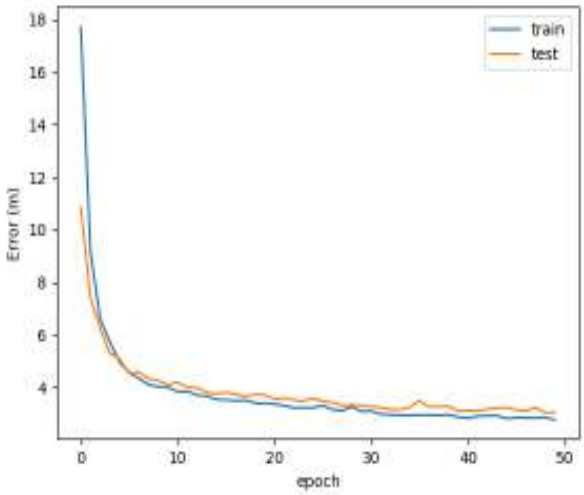
To evaluate the localization performance of the proposed HDML based indoor localization system, we did an experiment with an Android smartphone. The user holds the smartphone in his hand and walked in the experiment area. The Wi-Fi RSSI receiver module in the smartphone collects the RSSI values from APs through an Android application. The collected RSSI data used for heat map generation and generated 3000 heat maps for model training and 1500 heat maps for testing. Fig. 2 shows the localization performance of the different models.

From Fig. 2, the proposed HDLM model achieves a minimum localization error than other models and converges the localization error quickly. As compared to other models, the HDLM gives accurate localization results with RSSI heat maps. The model accuracy from each model is summarized in table 1.

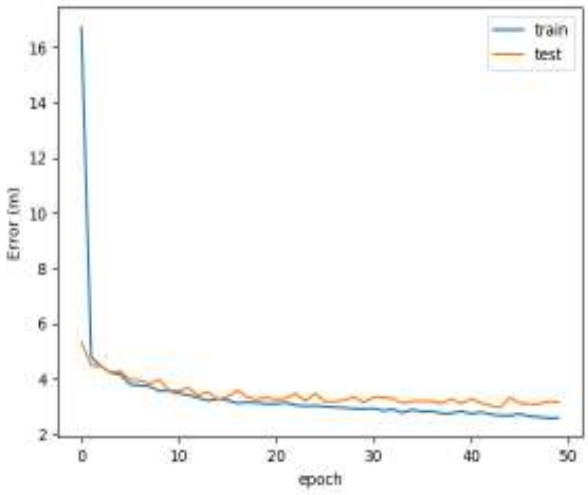
From Table 1, the proposed HDLM gives the best model accuracy than other models and reduces the localization error for Wi-Fi RSSI based localization systems. From the experiments and result analysis, the proposed HDLM outperforms existing approaches and gives accurate user position results for indoor localization.



(a) CNN



(b) LSTM



(c) Proposed HDLM

Fig. 2 Localization performance of the different models.

Table 1: Accuracy comparison of models.

Model	Accuracy (%)
CNN	82.30 %
LSTM	85 %
Proposed HDLM	88 %

IV. Conclusion

In this paper, we analyze the performance of the HDLM based indoor localization system using Wi-Fi RSSI heat maps. The experiment and result analysis show that the proposed HDLM based indoor localization system provides accurate localization results for Wi-Fi RSSI based localization systems. The Wi-Fi RSSI heat map-based localization system is an alternative solution for Wi-Fi RSSI signal interferences. The heat map-based localization system is also reducing the multipath effects for indoor environments and gives better results for indoor localization. In the future work, we intend to add the advanced deep learning architectures such as generative adversarial networks for Wi-Fi RSSI heat maps-based localization systems.

Acknowledgment

This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (2016-0-00564, Development of Intelligent Interaction Technology Based on Context Awareness and Human Intention Understanding).

References

[1] A. Poulose, J. Kim, and D. S. Han, "A sensor fusion framework for indoor localization using smartphone sensors and Wi-Fi RSSI measurements," Applied Sciences, vol. 9, p. 4379, 2019.

[2] A. Poulose, J. Kim, and D. S. Han, "Indoor localization with smartphones: Magnetometer calibration," in 2019 IEEE International Conference on Consumer Electronics (ICCE), 2019, pp. 1-3.

[3] A. Poulose and D. S. Han, "Hybrid Indoor Localization Using IMU Sensors and Smartphone Camera," Sensors, vol. 19, p. 5084, 2019.

[4] M. Shchekotov, "Indoor localization method based on Wi-Fi trilateration technique," in Proceeding of the 16th conference of fruct association, 2014, pp. 177-179.

[5] Y. Li and W. Dai, "Bitcoin price forecasting method based on CNN-LSTM hybrid neural network model," The Journal of Engineering, vol. 2020, pp. 344-347, 2020.

[6] M. T. Hoang, B. Yuen, K. Ren, X. Dong, T. Lu, R. Westendorp, et al., "A CNN-LSTM Quantifier for Single Access Point CSI Indoor Localization," arXiv preprint arXiv:2005.06394, 2020.

Error Correction of Wearable Sensors using Sliding Window in Optical Camera Communication

Md. Faisal Ahmed
Department of Electronics Engineering
Kookmin University
Seoul, South Korea
faisal.ahmed@ieee.org

Israt Jahan
Department of Electronics Engineering
Kookmin University
Seoul, South Korea
israt.eee2k11@gmail.com

Yeong Min Jang
Department of Electronics Engineering
Kookmin University
Seoul, South Korea
yjang@kookmin.ac.kr

***Abstract*— In this article, we have implemented an optical camera communication-based system to collect data from the human body using a pulse oximeter sensor. The system is mainly designed for convenient operation, long-term use, warning status, etc. So, the data collection and process in the output can show some unusual data in the receiver, this unusual data is eliminated by using a sliding window technique. After the inclusion of the sliding window, we can eliminate the noisy data as a result the BER will improve as well. We have also shown the improvement of performance by using a sliding window at the receiver in Python 3.7.**

Keywords—optical camera communication, pulse oximeter sensor, sliding window

I. INTRODUCTION

Optical camera communication (OCC) is an interesting topic in the field of healthcare application because of its radiofrequency free properties. The healthcare data are collected using a wearable sensor connected to a communication device to provide healthcare services across the users. The transmission of health information can be done wired and wireless. Wired-connected sensors are odd, costly, and consume high power [1]. On the other hand, currently, radio-frequency (RF)-based devices, such as Bluetooth, ZigBee, and 6LowPAN are mostly used for wireless communication objectives [2-5]. However, it can cause serious damage to human health and negative biological effects in the human body when long-term involvement with electromagnetic radiation (EMR) originating for RF. Some literature has given the overall implementation of the OCC system for eHealth application, but the error rate is quite high. By motivating this we have designed a sliding window algorithm at the receiver to remove the unwanted signal for lower bit-error-rate. We have also designed a system that collects data of heart rate (HR) and blood oxygen saturation (SpO_2) from a pulse oximeter sensor and send the data using LED array as transmitter [6], [7]. The camera is used as a receiver and processed in python environment with neural network for LED detection and data collection. After collecting the data at the receiver, the sliding window technique is used to reduce the unwanted noise at the receiver. In that case, we can improve the error performance at the receiver. Actually the error is happened due to the motion interference of the input sensor or due to the channel noise at the receiver [8].

II. SYSTEM OVERVIEW

The system is designed considering indoor scenario, that monitored the patient's health simultaneously. The sensor is attached with a finger of a patient and collected the HR and SpO_2 data. The sensor is connected with a patch circuit attached as an arm in the patient's body. The patch is composed of an LED array and microcontroller, here LED array is used as a source of optical signal and the microcontroller as a processor for that optical signal. The signal is modulated with the data using color intensity modulation using the microcontroller unit. Based on the symbol characteristics the color of the RGB LED is changed inside the LED array. A camera is used for surveillance and receiving data from the LED array using OCC simultaneously. Neural networks are developed to detect and recognize each LED and its color in the LED array using the camera, which is also used for the surveillance objective, simultaneously. However, for detecting the individual LED of a 4×4 LED array we use Darkflow in python environment with Open CV for training huge amount of image datasets. The weight of the trained image is used to test and label of each LED. In that case each LED shows the variation of color in different intensity grayscale level of R, G, and B code. After, recognizing the proper signal based on that three code the program extracts the signal and store it for correction purpose. So before processing data and recognizing color the binary scale, grayscale in different threshold is performed. Afterward, deletion and erosion is performed based on the image stripe. So after reorganization the data is demodulated based on the color intensity in different level of gray scale. Afterward, the data are processed using sliding window technique in different window size. We take the average of every window sets. If the set of window containing any error, then the average value and threshold value shows large margin of variation. After assuming the error value of a window we can eliminate the value and store a clean data sets in separate file. Finally, the data are transmitted to a cloud server, which can be further accessed by any authorized person using a private user ID and password. Before sending to the cloud server the data is processed using sliding window to improve the performance. The entire system overview is shown in Figure 1, including the sliding window technique.

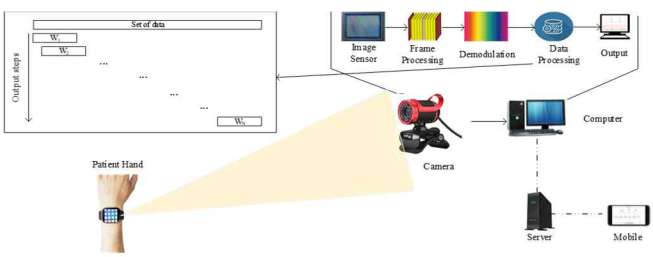


Figure 1: The overall OCC system architecture with sliding window algorithm in data processing at the receiver.

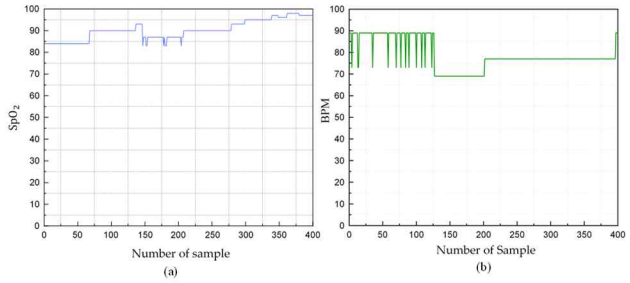


Figure 2: (a) Blood oxygen saturation and (b) heart rate, data collection after employing sliding window algorithm at the receiver.

III. Performance analysis

In the receiver, the SpO_2 and BPM data is collected using camera. Afterward, the data is processed using sliding window algorithm with different window size. The window size mainly depend on the number of sample in an interval data is collected. After applying sliding window scheme the SpO_2 and HR data is shown in Figure 2 (a) and (b) with respect to the number of sample. In the receiver, we took 400 samples of data in one interval for employing sliding window scheme to generate the graphs. As shown in Figure 2, several unstable spikes are generated within the initial samples. The reason is the initial stirs of the patch while attaching the sensor. The exact BPM and SpO2 values can be observed after 125 samples and 200 samples, respectively. The experiments were performed on a male volunteer whose basic information are: 62 kg weight, 167 cm height, and 26 years old. We also calculate the mean absolute error (MAE) of the data of the volunteer in different lighting conditions and considering four different indoor conditions. Firstly, in nighttime, the data is taken using multiple room light sources and a single light source with low intensity, separately. The same procedure is applied in daytime using slight sunlight entering the room. It can be seen that the MAE increases when the interferences are high. Finally, the data are sent to an IoT cloud server where an authorized person can access the data using a login ID and password.

IV. Data decoding

The retrieval of the received signal significantly depends on the orientation of the transmitter. It will not be sophisticated to maintain a fixed position of the patch as the placement of the hand may be altered based on the patients' requirement. However, different orientation of the LED array will generate substantial errors in the data. Thankfully, we have avoided the challenge by measuring the amount of inflection of the LED array. As mentioned in transmitter section, three LEDs are reserved in OFF position to indicate the starting and ending points of each IR, BPM, and SpO2 data. The three positions will be unchanged in the LED matrix disregarding any inflection. Thus, in the original LED matrix, the amount of orientation is calculated locating the positions of the OFF LEDs. The data may vary significantly, therefore, it is possible for any of the other LEDs to progress in the OFF state if the data is too small. However, as we defined the OFF LEDs at the end of each dataset, the newly OFF LEDs will appear sequentially just before the specific OFF LED, therefore, the positions of the OFF LEDs will be determined easily. After measuring the inflection angle, the starting point of the dataset is defined. Then, the symbols are decoded using the color code sequence in the LED matrix. Finally, the data for IR, BPM, and SpO2 are stored into three separate CSV files.

V. Conclusion

A real-time health monitoring system based on OCC with the improvement of system error performance is proposed in this paper. A MAX30102 sensor is used to collect the IR, SpO2, and BPM data, and connected to a patch mounted on the patient's hand. The patch is composed of an LED array that is used to transmit the data to a webcam. As the OCC system is itself a higher security system so we don't need any additional security protocol to improve the performance of the system. In the receiver, each LED in the array is detected using a NN. Also, another feature-extraction-based NN is used to recognize the colors precisely. In addition, a mechanism based on sliding window technique is developed to assuage the challenge error free data collection at the receiver. The whole data decoding and ECG signal generation procedures are performed in Python 3.7. Finally, the data is processed and accessed by a remote monitor and stored in a cloud server.

Acknowledgment

This work was supported by the Ministry of Trade, Industry, and Energy (MOTIE) and Korea Institute for Advancement of Technology (KIAT) through the International Cooperative R&D program (No. P007800004).

References

[1] M. F. Ahmed, M. K. Hasan, M. Shahjalal, M. M. Alam, and Y. M. Jang, "Design and Implementation of an OCC-Based Real-Time Heart Rate and Pulse-Oxygen Saturation Monitoring System," *IEEE Access*, vol. 8, pp. 198740-198747, 2020.

[2] B. Jiao, "Anti-motion interference wearable device for monitoring blood oxygen saturation based on sliding window algorithm", *IEEE Access*, vol. 8, pp. 124675-124687, 2020.

[3] W. Huang, P. Tian and Z. Xu, "Design and implementation of a real-time CIM-MIMO optical camera communication system", *Opt. Express*, vol. 24, pp. 24567-24579, Oct. 2016.

[4] O. S. Alwan and K. Prahald Rao, "Dedicated real-time monitoring system for health care using ZigBee", *Healthcare Technol. Lett.*, vol. 4, no. 4, pp. 142-144, Aug. 2017.

[5] V. P. Tran and A. A. Al-Jumaily, "A novel oxygen-hemoglobin model for non-contact sleep monitoring of oxygen saturation", *IEEE Sensors J.*, vol. 19, no. 24, pp. 12325-12332, Dec. 2019.

[6] M. Hasan, M. Shahjalal, M. Chowdhury and Y. Jang, "Real-time healthcare data transmission for remote patient monitoring in patch-based hybrid OCC/BLE networks", *Sensors*, vol. 19, no. 5, pp. 1208, Mar. 2019.

[7] M. Z. Chowdhury, M. T. Hossan, M. Shahjalal, M. K. Hasan and Y. M. Jang, "A new 5G eHealth architecture based on optical camera communication: An overview prospects and applications", *IEEE Consum. Electron. Mag.*, vol. 9, no. 6, pp. 23-33, Nov. 2020.

[8] M. F. Ahmed, M. K. Hasan, M. Shahjalal, M. M. Alam and Y. M. Jang, "Experimental demonstration of continuous sensor data monitoring using neural network-based optical camera communications", *IEEE Photon. J.*, vol. 12, no. 5, pp. 1-11, Oct. 2020

6G 셀룰러 네트워크를 위한 인공지능: massive MIMO-NOMA 딥러닝 해석

무니브 아흐매드, 신수용*
금오공과대학교

muneeb.ahmad@kumoh.ac.kr, *wdragon@kumoh.ac.kr

Artificial Intelligence for Future 6G Cellular Networks: A Deep Learning approach for Massive MIMO NOMA System

Muneeb Ahmad, Soo Young Shin*
Kumoh National Institute of Technology, South Korea.

요 약

The upcoming sixth-generation (6G) wireless networks are expected to lay the foundation for intelligent networks powered by isolated artificial intelligence (AI). For this, we assume that 6G wireless networks will operate with automatic on-demand configuration to improve network performance and service types. Non-orthogonal multiple access (NOMA) has received much attention as a major candidate for fifth-generation (5G) mobile communications systems. In this article, authors are exploiting deep learning (DL) approach that suggests that DL is essential to find the optimal sequence of channel gain and signal detection for 6G communication systems.

Ⅰ. 서 론

Data-driven research towards an adaptive and intelligent method has gained the attention of the researchers after the echoes of 6th generation (6G) future wireless networks began to chime. The advents in computing methods from Machine Learning (ML) to Artificial Intelligence (AI) bridged by Deep Learning (DL) approaches have led to the consideration of autonomous management and service classification to reconfigure the demands of future wireless networks. These kinds of data-driven methods show strong potential to realize the ambition of fully developed intelligent 6G wireless communication. In the recent future, the amenity of mass connections, intelligent human-machine interface and huge data traffic, AI seems to be a magic wand that can make the system learn, perform and enhance the operations by exploiting the operational knowledge in the form of data.

A subfield of AI is commonly known as DL and is implemented widely in various engineering streams such as image and video processing, data processing and wireless communication for channel estimation and signal detection [1]. Towards the implementation and fulfil the requirements of 6G network, several kinds of research are ongoing [2]. Among them, Massive Multiple-Input Multiple-Output (mMIMO) and Non-orthogonal Multiple Access (NOMA) systems are believed to have the potential to address the capacity demands and the spectral efficiency of the network respectively. NOMA and mMIMO integrated system if aided with DL methods; can deliver astonishing results due to the ability of deep neural networks to process high dimensional data to enhance the detection performance of the above-mentioned integrated system [3].

This paper examines the ability of downlink mMIMO-NOMA system based on the DL methods for channel estimation for future 6G cellular networks. In the following section, the system model and the ability of AI-driven networks is briefly discussed. Furthermore, the results are provided to give insight into the possible improvements in conventional networks to fulfil the requirements of future communication networks as shown in Figure 1.

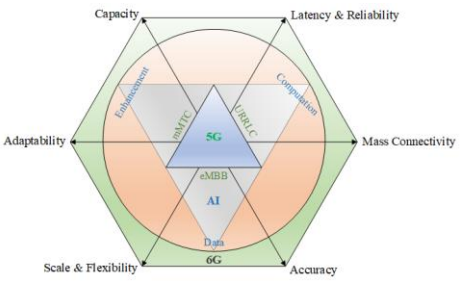


Figure 1: Key requirements and characteristics of 6G

Ⅱ. 본론

Unlike the traditional orthogonal multiple access (OMA) systems, NOMA utilizes power in a non-orthogonal fashion to enhance the spectrum efficiency of the network. Figure 2 shows the basic architecture

of the NOMA system. This paper explains the implementation of DL methods in wireless cellular systems to estimate the radio channel quality that is essential for the design of transmitting data.

In the mMIMO-NOMA system, the users are accumulated in a cluster and are served using multiple antennas at the base station (BS) via zero-Forcing (ZF) beamforming technique [4]. Let's assume the total number of users in the system are K, for the i^{th} user the signal can be expressed as y(t):

$$y(t) = \sum_{i=1}^{K} \sqrt{p_i} s_i(t) \quad (1)$$

where p_i is the power coefficient and s_i is the signal of the i^{th} user (i = 1,2,⋯,K). Deep Neural Network (DNN) is a deeper version of a neural network that generally consists of three types of layers: input, hidden, and output. For n layer network, the output of the n^{th} layer y_n can be expressed as:

$$y_n = f(w_n * y_{n-1} + b_n) \quad (2)$$

where w_n is the weight matrix and b_n is the bias vector. For a classic DNN, the activation function is the sigmoid and is limited to [0, 1]. This article adopts a DNN in a mMIMO-NOMA system, the system divides the process of detection into channel estimation, MMSE detection and signal decision. The deep learning method can perform all these procedures as a single process. The matrix for the mMIMO-NOMA signal at the transmitter side can be expressed as:

$$M = (M_1, M_2, \ldots, M_L) \quad (3)$$

where M_l is the m^{th} transmission antenna and it can be expressed as:

$$M_l = \sum_{k=1}^{K} \sqrt{P_k M_l^k} \quad (4)$$

NOMA technique strongly depends on the Successive Interference Cancellation (SIC) detection and it needs to be continuous process of decoding, reconstructing, and signal canceling. Allocated power to the users according to the channel gain is (P_1>P_2>⋯>P_K). Figure 2 also presents a DL model for MIMO-NOMA-DL signal detection. Keeping the online block inactive, the offline training will continue in the training state. The input of the DNN system will be the received mMIMO-NOMA signal and the DNN will optimize parameters by considering labelled data as supervised data. After the completion of training process the testing phase will commence. The results are generated when online block access the DNN and offline block is suspended.

Ⅲ. 결론

Figure 3 shows the gain of deep learning method over the conventional channel equalization schemes such as the least square (LS) and minimum mean square error (MMSE).

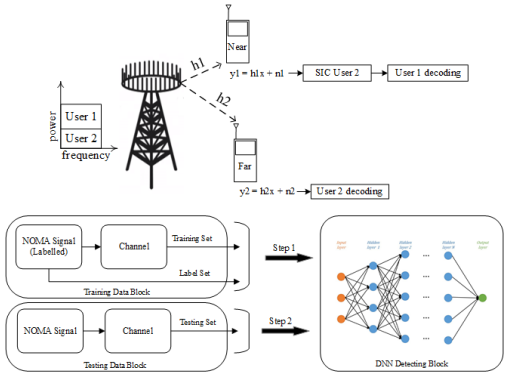


Figure 2: mMIMO-NOMA system model embedded with DL structure

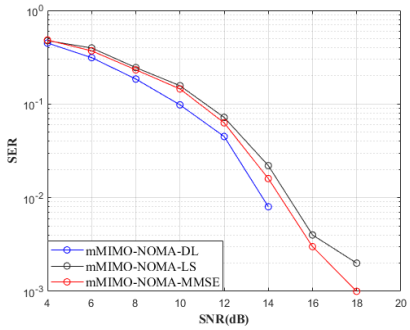


Figure 3: Performance comparison of mMIMO-NOM-DL with different channel estimation schemes

The MMSE inherently leads in the performance comparison with LS, where both the techniques are conventional. Compared to the DL method it is clear from the curves in Figure 3 that DL leads in the performance over LS and MMSE. For the increased signal-to-noise ratio (SNR), the symbol-error-rate (SER) improves systematically. Ths gives intuition that the system's throughput increases if the SER is optimized and hence the spectral efficiency also increases leading to the mass connectivity and high data rate.

In future works, online learning and testing can be considered, also other advanced DL approach can be developed for better signal detection or estimation.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government. (MSIT) (No. 2019R1A2C1089542).

참 고 문 헌

[1]Ma, Siwei, et al. "Image and video compression with neural networks: A review." IEEE Transactions on Circuits and Systems for Video Technology (2019).

[2]Wang, Ning, et al. "Pilot Contamination Attack Detection for NOMA in 5G mm-Wave Massive MIMO Networks." IEEE Transactions on Information Forensics and Security 15 (2019): 1363-1378.

[3]You, Xiaohu, et al. "AI for 5G: research directions and paradigms." Science China Information Sciences 62.2 (2019): 21301.

[4]Le, Quang Nhat, et al. "Learning-Assisted User Clustering in Cell-Free Massive MIMO-NOMA Networks." arXiv preprint arXiv:2011.07549 (2020).

수중 사물인터넷 환경에서 패킷 에러율 예측 및 네트워크 파라메터 최적화를 위한 기계학습 모델

김진홍 [1], 이성원 [2], 임길택 [1] 김동균 [3]
[1] 한국전자통신연구원, [2] 대구한의대학교 스마트 IT 융합학부,
[3] 경북대학교 IT 대학 컴퓨터학부
jinhong@etri.re.kr, lsw5359@dhu.ac.kr, ktl@etri.re.kr, dongkyun@knu.ac.kr

A Machine Learning Model for Prediction of Packet Error Rate and Network Parameter Optimization in Internet of Underwater Things

Jinhong Kim[1], Sungwon Lee[2], Kil-Taek Lim[1] and Dongkyun Kim[3]
[1]Electronics and Telecommunications Research Institute, [2] Deagu Hanny Univ.,
[3]Kyungpook Nat. Univ.

요 약

본 논문은 수중 사물인터넷(Internet of Underwater Things)환경을 구축함에 있어 수중환경의 특성으로 인한 낮은 전송 신뢰성을 극복하기 위해 주변환경의 변화에 맞춰 네트워크 프로토콜의 동작에 영향을 미치는 파라메터를 최적화하는 기계학습 프로세스 모델을 제시한다. 제안하는 프로세스 모델에서는 실측된 패킷의 전송로그 데이터를 바탕으로 전송 에러를 예측하는 함수모델을 기계학습을 통해 생성하며, 생성된 함수모델을 검증하는 과정을 통해 발생하는 오차를 일정한 임계값 이하로 유지할 수 있도록 인자들을 제어한다. 그리고 함수모델을 통해 예측된 에러율을 기반으로 최적의 네트워크 파라메터를 선정하고 프로토콜에 적용함으로써 패킷의 전송 신뢰성을 향상시킨다.

Ⅰ. 서 론

최근 수중 센서 네트워크는 오염도 측정, 전략 감시, 항만 트래픽 관리 등 다양한 응용이 수행되고 있으며, 이에 따라 센서 노드뿐만 아니라 수중 드론, 경량형 잠수함, 고래 등 수중동물에 부착한 생체 센서 등 다양한 통신장비들이 사용되고 있는 수중 사물인터넷(Internet of Underwater Things, 이하 수중 IoT) 환경을 구축하고 있다. 또한 인공지능 등 다양한 최신 기술을 수중 IoT 환경에 적용하여 새로운 응용의 신뢰성 있는 동작을 지원하려는 연구가 진행되고 있다. 이러한 연구들은 주로 라우팅 계층과 데이터 링크 계층의 프로토콜을 중심으로 진행되고 있는 추세이다[1][2].

특히, 라우팅과 데이터 링크 계층에서는 수중 통신의 낮은 전송 신뢰성을 극복하기 위해 주변 노드들과의 링크 품질을 측정하고, 측정된 링크 품질에 따라 프로토콜의 동작을 변화시키는 기법이 많이 적용되고 있다. 예를 들어, 대표적인 라우팅 프로토콜인 DFR(Directed Flooding based Routing) 프로토콜에서는 각 노드가 주변 노드들과의 평균 에러율을 측정하고, 측정된 에러율에 따라 다중 경로의 숫자를 조절하는 방법을 사용하고 있다[3]. 이러한 프로토콜들에서는 주변 노드들과의 링크 에러율을 측정하는 방식으로 ETX(Expected Transmission Count)를 사용하고 있다. 현재 사용되고 있는 ETX 는 각 노드들이 주기적인 hello 메시지를 교환하고, 이 메시지들의 수신 확률을 통해 ETX 를 계산한다.

하지만, 이러한 주기적인 hello 메시지를 활용한 ETX 측정 방식은 네트워크의 에러율이 급격하게 변화하는 환경에서는 현재 에러율을 즉시 반영할 수 없는 단점이 있다. 특히, 대역폭이 충분한 육상 네트워크와는 달리, 대역폭이 부족한 수중 네트워크에서는 hello 메시지의 전송 주기가 몇 십초 수준으로 설정된다. 이에 링크 품질의 변화에 따라 ETX 가 유의미하게 변화하기 위해서는 분단위 이상의 시간이 소모된다.

이러한 문제를 해결하기 위해, 육상 센서 네트워크에서는 잡음 대비 신호의 세기(SNR, Signal-Noise Ratio)를 사용한다. 그러나, 자연적인 신호가 거의 발생하지 않는 전파 대역과는 달리, 음파 대역을 사용하는 수중 통신은 SNR 만으로 링크 에러율을 예측할 수 없다. 또한 현재까지의 여러 실험결과에 따라 수중 통신의 링크 에러율은 수심, 염분도, 조류 등 다양한 요소들이 복합적으로 관여한다는 사실이 알려져 있다.

뿐만 아니라, 링크 품질에 따라 프로토콜의 동작 또는 네트워크 파라메터를 변화시키는 프로토콜이 다수 사용되면서, 다른 계층의 동작에 의해 자신의 파라메터가 영향을 받는 경우도 발생한다. 예를 들어, 링크 품질이 낮을 때 전송 성공률을 향상시킬 수 있는 데이터 링크 계층 프로토콜이 적용되더라도, DFR 프로토콜은 측정된 에러율만을 사용하여 플러딩 영역을 설정하여 불필요한 네트워크 리소스를 소모하게 된다. 따라서 다수의 외부 요소와 다른 계층의 네트워크 파라메터 등 복합적 요소를 한꺼번에 고려하여 전송 성공률을 예측하고 최적의 파라메터를 결정할 수 있는 통합 수식을 만들어 내는 연구는 매우 어려운 일로 알려져 있었다.

그러나 최근 인공지능 기술의 발전에 따라 다양한 제어 불가능 인자들을 고려하여 최적의 제어가능 인자를 결정하는 새로운 방법론들이 개발되고 있다. 본 논문에서는 다양한 요소들을 활용하여 모델링/기계학습을 수행하는 알고리즘과 기계학습 모델을 제안한다.

Ⅱ. 본론

본 논문에서 제안하는 기계학습 프로세스는 다수의 제어가능 인자와 제어불가능 인자를 기계학습 모델의 입력 인자로 사용한다. 본 논문에서 사용하는 제어 가능 인자와 제어불가능 인자는 다음과 같다.

- 제어가능 인자: 노드의 이동방향, 노드의 속도, 전송 신호의 크기, 신호간 간격(interval), 해당 계층의 네트워크 파라메터
- 제어불가능 인자: 수심, 염분도, 조류방향, 타 계층의 네트워크 파라메터

제어 불가능 인자는 센서 노드에 부착된 센서 장비를 통해 수집할 수 있음을 가정한다. 또한, 다수의 6LoWPAN (IPv6 over Low-Power Wireless Personal Area Network)에서 사용되는 서비스 계층(Service Layer)의 개념을 도입하여 각 계층간 네트워크 파라메터를 공유하도록 설계한다.

1) 에러율 측정 및 네트워크 파라메터 설정 통합 프로세스 모델: 다음 그림 1 과 같이, 각 노드들은 이전 시점에서 측정된 패킷 전송 성공률과 함께 제어 불가능 인자와 제어가능 인자들을 입력값으로 하여 에러율 예측 프로세스를 실행한다. 이후 예측된 에러율을 사용하여 현재 에러율에 맞는 프로토콜의 동작을 설정한 후, 해당 동작에 알맞게 네트워크 파라메터를 설정하는 과정을 거친다.

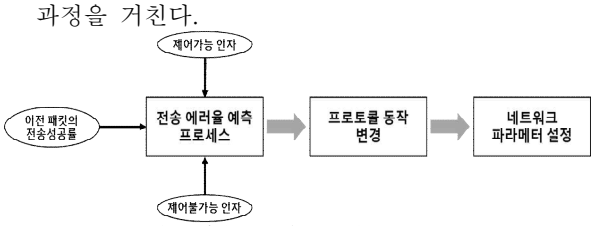


그림 1 제안된 통합 프로세스 모델

2) 에러율 예측 프로세스: 위 통합 프로세스 모델의 성능을 향상시키기 위해서는 에러율 예측 프로세스의 정확도의 향상이 중요하다. 본 논문에서는 기계학습을 사용하여 그림 2 의 과정을 거쳐 전송 에러율 예측 프로세스의 성능을 향상시킨다.

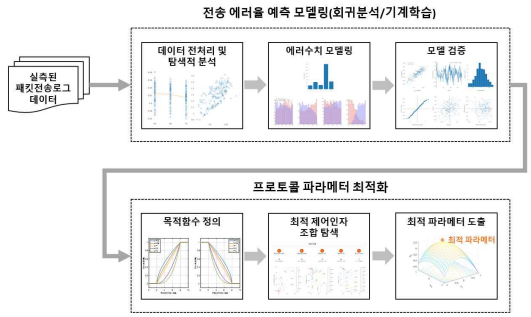


그림 2 전송에러율 예측 및 파라메터 최적화 모델

위 모델에서는 각 노드가 예측한 에러 수치와 함께 패킷 전송과정에서 실측된 로그 데이터를 활용하여 탐색적 데이터 분석을 수행한다. 이후 예측된 수치와 실측 수치 간의 오차를 최소화할 수 있는 2 차원 함수를 기계학습함으로써 제어가능 S 인자와 제어 불가능 인자간 반영 비율의 모델을 완성한다. 이후 기계학습 과정에서 해당 모델을 검증하되, 이 때 발생하는 일정 임계점 이상의 오차가 발생하는 경우 해당오차가 임계점 이하가 될 수 있도록 노드가 제어가능 인자의 값을 수정하도록 함으로써 에러율의 예측 성공률을 일정 임계값 이상으로 유지하도록 한다.

3) 프로토콜 파라메터 최적화 프로세스: 위 1,2 에서 제안된 프로세스를 통해 기계학습된 모델을 적용할 경우 각 노드는 일정 수준 이하의 오차율로 전송 에러율을 예측할 수 있다. 이후 노드는 예측된 에러율을 기반으로 하여 최적의 네트워크 파라메터를 선정하는 프로세스를 수행한다. 이 과정에서 각 노드들은 기계학습을 통해 각 파라메터에 따른 전송 성공률 향상을 위한 목적함수를 계산한다. 이후 새로운 에러율이 예측될 때 마다 해당 목적함수에 따른 최적의 네트워크 파라메터를 조합하여 각 프로토콜에 적용한다.

Ⅲ. 결론

본 논문에서는 수중 센서 네트워크에서 발생하는 에러율 반영 딜레이 문제와 에러율을 고려하여 설정되는 네트워크 파라메터간 상호 의존성 이슈를 해결할 수 있는 기계학습 프로세스 모델을 제시하였다. 제시된 모델에서는 패킷 전송률에 영향을 미치는 요소들을 제어가능인자와 제어불가능 인자로 구분하고, 해당 인자들에 따라 기계학습된 2 차원 수식 모델을 생성한다. 이후 모델을 통해 예측된 결과와 실측된 패킷 전송 로그데이터를 사용하여 파라메터 최적화 모델 함수를 디자인하고, 현재 예측된 에러율에 따라 각 계층별 최적의 파라메터를 설정할 수 있는 조합을 탐색하여 사용한다. 이 과정에서 에러율이 일정 임계값 이상일 경우 노드가 제어 가능 인자를 수정함으로써 예측 에러율과 실측 에러율의 오차를 일정 수준 이하로 유지할 수 있도록 한다.

ACKNOWLEDGMENT

본 연구는 한국전자통신연구원 주요사업의 일환으로 수행되었음[20ZD1120, 대경권 지역산업 기반 ICT 융합기술고도화 지원사업]

참 고 문 헌

[1] En-Cheng Liou, Chien-Chi Kao, Ching-Hao Chang, Yi-Shan Lin and Chun-Ju Huang. " Internet of underwater things: Challenges and routing protocols," in *2018 IEEE International Conference on Applied System Invention (ICASI), April* 2018, pp. 129-132.

[2] Mohamed Ammar, Khalil Ibrahimi, Mohammed Jouhari and Jalel Ben-Othman." MAC Protocol-Based Depth Adjustment and Splitting Mechanism for UnderWater Sensor Network (UWSN)," in *2018 IEEE Global Communications Conference (GLOBECOM),* Dec. 2018, pp. 1-6.

[3] Daeyoup Hwang and Dongkyun Kim." "DFR: Directional flooding-based routing protocol for underwater sensor networks," *in 2008 IEEE OCEANS*, Sept. 2008, pp. 1-7.

장소와 유형 분리 신경망 기반의 고정형 대상 인식

이형석, 이재호*, 김도형***, 류철***
한국전자통신연구원

hyslee@etri.re.kr, *bigleap@etri.re.kr, **dhkim@etri.re.kr, ***ryuch@etri.re.kr

Fixed object recognition based on place and type separation neural network

Hyung-Seok Lee, Jae-Ho Lee*, Do-Hyung Kim**, Cheol Ryu***
Electronics and Telecommunications Research Institute

요 약

본 논문에서는 해결하고자 하는 문제는 위치 고정형 객체들이 서로 다른 장소에 존재할 때 유형이 비슷하다 하더라도 장소와 대상을 정확히 인식하는 것이다. 이미지 기반으로 고정형의 대상 객체를 인식하는 문제를 풀기위한 한 방법으로 장소 인식과 객체 유형 인식을 분리하여 수행한 다음 결과를 통합하여 최종 인식을 하는 방법을 제시한다.

Ⅰ. 서 론

포스트 스마트폰으로 전망되어지는 스마트 글래스[1]에서 정면에 보이는 물체를 인식하는 것이 중요한 핵심 기능이다. 서로 다른 장소에서 같은 유형이나 비슷한 대상들이 다수가 있을 경우에 있어서 단순한 객체의 유형 분별을 한 딥러닝 신경망[2]을 단순히 사용하는 것은 적절치 않다. 객체가 존재하는 배경과 함께 촬영된 이미지들을 학습한 단일 신경망을 통하여 문제를 해결하는 방법을 먼저 생각해 볼 수 있다. 즉, 다른 장소에 있는 같은 유형의 대상 물체가 신경망에서 서로 다른 카테고리 이 방법은 대상을 특정 장소에 설치한 다음 학습할 이미지를 확보해야 하므로, 만일 장소를 이동한다면 새로운 학습 데이터를 확보하고 학습도 새로 수행해야 하는 문제가 있다. 그리고, 다른 장소에 같은 유형의 물체가 있는 경우 신경망에서 서로 다른 카테고리로 분별하기 어렵게 된다.

본 논문에서는 이미지 기반으로 고정형의 대상 객체를 인식하는 문제를 풀기위해서 장소 인식과 객체 유형 인식을 분리하여 수행한 다음 결과를 통합하여 최종 인식을 하는 방법을 제시한다.

Ⅱ. 본론

해결하고자 하는 문제는 위치 고정형 객체들이 서로 다른 장소에 존재할 때 유형이 비슷하다 하더라도 장소와 대상을 정확히 인식하는 것이다. 서로 다른 공간에서의 같은 유형을 분별할 수 있어야 한다. 이러한 문제를 해결하기 위해서 다중 구조 신경망을 활용하여 객체 종류 분류 뿐 만이 아니라 객체가 설치된 공간 정보도 동시에 사용할 필요가 있다. 이미지에 포함된 객체의 유형을 알고 그 객체가 존재하는 공간의 위치 정보를 알 수 있으면 대상을 더욱더 정확하게 식별할 수 있기 때문이다. 공간과 객체 이미지를 분리하여 각각을 인식하고 결과를 종합하여 대상을 추론함으로써. 공간에 대한 학습과정만 거치면, 객체의 유형에 대한 기존의 딥러닝 신경망을 그대로 활용할 수 있는 장점이 있다.

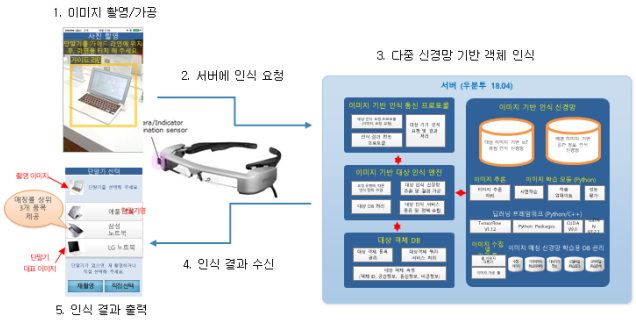


그림 1 이중 신경망 기반 객체 인식 구성

그림 3 에서 공간과 객체를 각각 분별하는 이중 신경망 기반 객체인식 시스템의 구성을 보여준다. 단말은 안드로이드 기반으로 서버는 우분투 리눅스를 기반으로 구현하였다. 단말과 서버와의 통신은 웹서비스를 통한 REST 인터페이스를 사용하였다. 장소 및 기기를 분별하기 위한 기본 모델을 Inception v3[3]을 사용하였다. 기존 학습된 모델로부터 출력단만 본 시스템에 맞게 재 구성하고 재 학습하는 작업을 수행하였다.

안경형 단말(Moverio BT-350)에서 인식하고자 하는 물체를 카메라 중앙의 사각 박스에 두고 인식 버튼을 누르면, 인식 이미지가 서버로 전송되고 서버에서 인식 결과를 수신하는 구조이다. 배경을 포함한 전체 촬영 이미지를 사용하여 장소를 분별한다.

실험을 위하여 3 개 장소와 8 개의 종류의 기기를 대상으로 단말에서 촬영된 장소와 기기 이미지를 서버에

전송한 후 인식 결과를 단말이 수신하여 인식률과 처리 시간 정보를 기반으로 성능 분석을 하였다. 이를 위하여 장소와 기기들 각각 100 개 이상의 이미지를 학습하였다.

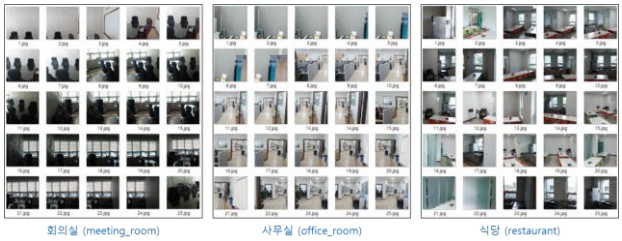


그림 2 장소 학습 데이터

그림 3 장소 이미지 수집 및 가공 예

장소의 경우 배경 이미지를 활용하는 장소의 경우 신경망의 입력은 종횡비 1:1 을 요구하므로 그림 3 와 같이 가공하여 사용하였다.

그림 4 학습에 사용된 8 개의 대상 기기 이미지 예

실험은 3 개 장소에서 10 개 기기를 대상으로 기기당 10 회 인식 실험을 수행하였다. 10 개 중 중복 유형의 기기는 2 종이어서 실제 기기 종류는 총 8 종이다. 대상 인식 시 장소 및 기기의 매칭 값이 모두 주어진 임계 값 이상일 경우 인식 성공으로 간주하였다. 3 개 장소에서 각각 등록되지 않은 기기 혹은 배경 등을 대상도 포함하여 전체 실험 수는 130 회이다.

그림 5 인식률 성능 평가

그림 5 는 실험 결과를 보여준다. 인식률은 장소와 기기 모두 인식에 성공한 수치, 오인식률은 장소와 기기 중 하나라도 실패, 미인식률은 장소와 기기 중 하나라도 임계값을 만족시키지 못하여 인식에 실패한 비율이다.

실험 결과 인식률은 장소 매칭값 0.6 이상, 기기 매칭값 0.7 이상일 경우, 83.85%로 측정되었고, 장소 매칭값 0.6 이상, 기기 매칭값 0.75 이상일 경우, 80.77%로 측정되었다. 매칭값에 따라 인식률에 차이가 발생하며, 매칭값이 너무 낮을 경우 오인식률이 증가하여 인식률이 저하되었다.

수행 시간에 대한 성능 측정은 안경형 단말(Moverio BT-350) 뿐 만이 아니라 스마트폰에서 수행하여 시간 성능을 측정하였다. Moverio BT-350 에서는 약 6 초 정도의 인식 시간이 소요되었으며, 이는 최신 스마트폰에 비해 약 2 배 느린 속도를 보였다.

	이미지 촬영	인식 요청 및 결과	전체 인식 시간
안경형 단말(Moverio BT-350)	1.309초	4.903초	**6.212초**
스마트폰(Galaxy Note 9)	0.961초	3.031초	3.992초
스마트폰(Galaxy S10)	0.793초	2.317초	3.110초

Ⅲ. 결론

본 논문에서는 스마트 글래스 같은 웨어러블 장치에서 사용자가 보이는 대상 객체에 대한 서비스를 제공하는 비주얼 서비스를 지원하기 위하여 이미지 기반으로 대상 객체를 인식하는 한 방법으로, 장소와 기기 유형을 각각 다른 신경망을 통하여 인식하고 결과를 종합하는 방식을 제안하고 실험을 통하여 성능을 분석하였다.

본 연구에서는 인식 대상을 포함하는 한 이미지로부터 장소를 인식 분류하는 기법을 사용하였는데 이 이미지의 배경만으로는 장소 인식에 다소 오류가 있음을 확인하였다. 이를 보완하는 차기 연구로 여러 개의 이미지로부터 인식을 한 후에 이 결과를 종합하여 장소를 판단하는 것을 생각해 볼 수 있다. 또한 본 논문에서는 대상 객체가 특정 장소에 고정되어 있는 것만 고려하였는데, 이동형 객체에 대한 문제에 대해서도 차기 연구과제가 될 수 있을 것이다.

ACKNOWLEDGMENT

이 논문은 2020 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (2018-0-00226, 포스트 스마트폰 시대를 대비한 Trusted Reality 핵심기술 개발)

참 고 문 헌

[1] K. Song, G. Kim, T. Kim, C. Ryu, S. Park, J.H. Lee, J.K. Lee, and S. Hwang, "Trusted Reality Technology, from a Post-Smartphone Perspective," Electronics and Telecommunications Trend, 2018.

[2] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," Proceedings of the 25th International Conference on Neural Information Processing Systems(NIPS) Volume 1 pp. 1097-1105, Dec. 20

[2] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, Zbigniew Wojna, "Rethinking the Inception Architecture for Computer Vision," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)

적대적 공격 방어를 위한 앙상블 기법에 관한 연구

김용수, 강효은*, 윤영여*, 김민재*, 김호원*
스마트엠투엠, *부산대학교
yongsu@smartm2m.co.kr, *{hyoeun0915, yeo8006, rlaalswo9, howonkim}@pusan.ac.kr

A Study on Ensemble Methods for Defense against Adversarial Attacks

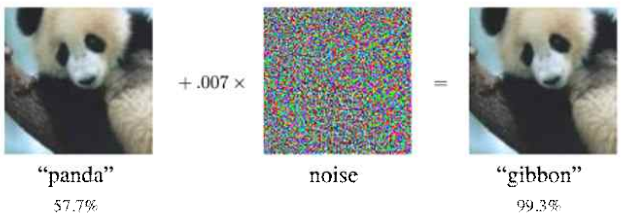
Yongsu Kim, Hyoeun Kang*, Youngyeo Yun*, Minjae Kim*, Howon Kim*
SmartM2M, *Pusan National Univ.

요 약

딥러닝을 비롯한 인공기술은 다양한 분야에서 뛰어난 성능을 보여주고 있지만, 의도적인 잡음을 추가하여 인공지능을 오작동시키는 적대적 공격(Adversarial Attack)에 취약한 것으로 밝혀졌다. 이는 인공지능 기술 기반 차세대 자율주행 시스템 및 군사용 장비 등에 악용될 경우 큰 문제를 야기할 수도 있다. 본 논문에서는 특정 인공지능 모델의 취약점을 공략하여 이루어지는 적대적 공격을 완화하기 위해, 유사한 성능을 가지는 다양한 모델 구조 간 앙상블 기법을 적용하는 방법에 관한 연구를 제안한다. 제안하는 기법의 효과를 증명하기 위해 CIFAR-10 데이터세트 및 FGSM, PGD, C&W 공격 기법을 사용하여 실험을 진행하였으며, 단일 모델을 대상으로 생성된 적대적 사례(Adversarial Example)에 대해 모두 분류 정확도가 향상되는 효과를 보였다. 또한, PGD 기반 적대적 훈련 기법에 비해 적대적 사례에 대한 분류 정확도가 약 13.45% 향상함으로써 높은 수준의 적대적 공격 방어 성능을 제시함을 보였다.

Ⅰ. 서 론

DNN(Deep Neural Network)를 비롯한 딥러닝 기술은 얼굴 인식, 질병 진단, 자율 주행 등 다양한 시스템에 사용되어 뛰어난 성능을 보여주고 있다. 하지만 최근 연구에 따르면 딥러닝 기술은 의도적으로 미세한 잡음을 추가하여 오작동하게 만드는 적대적 공격(Adversarial Attack)에 취약한 것으로 밝혀졌다.



[그림 1] 적대적 공격 예시

그림 1은 적대적 공격의 대표적인 예시이다. 팬더 사진을 팬더라고 정상적으로 인식하는 딥러닝 모델이 있을 때, 이미지에 미세한 잡음을 추가하여 긴팔원숭이로 잘못 인식하도록 할 수 있다. 적대적 공격 관련 연구는 그림 1의 예시와 같이 주로 이미지 및 영상 분석 관련 딥러닝 모델을 대상으로 진행되었다. 입력 이미지에 미세한 교란 신호(perturbation)를 추가하여 생성한 악의적인 이미지를 적대적 사례(Adversarial Example)라고 한다. 이러한 적대적 사례를 생성하는 대표적인 공격 기법은 FGSM(Fast Gradient Sigh Method)[1], PGD(Projected Gradient Descent)[2], C&W(Carlini and Wagner) Attack[3] 등이 있다.

적대적 공격에 대한 방어 연구로는 먼저 적대적 사례를 학습 데이터에 포함해 정상적으로 인식하도록 모델을 재학습시키는 적대적 훈련(Adversarial Training)이 있다. Goodfellow 등[1]은 많은 적대적 사례들에 대해 충분한 적대적 훈련이 이루어진다면 적대적 공격에 대한 강인성(Robustness)을 갖는다고 주장하였다. 현재 주로 사용되는 강력한 공격 기법인 PGD를 기반으로 수행한 적대적 훈련은 적대적 공격 방어 연구의 성능 평가를 위한 벤치마크로 많이 사용되고 있다.

또한 노이즈 제거와 같은 이미지 변형 기반의 적대적 공격 방어 연구가 많이 제안되었다. 대표적으로 Feature Squeezing[4], MagNet[5], Feature Denoising[6]과 같은 방법들이 있다. 하지만 Athalye 등[7]은 이미지 변형 기반의 적대적 공격 방어 기법은 단순히 기울기 정보에 혼동을 주어 공격을 막는 방법이라고 주장하며, 변형에 대한 함수를 미분 가능한 다른 함수로 근사하여 공격하는 BPDA(Backward Pass Differentiable Approximation) 기법을 제안하였다. 이를 통해 대부분의 이미지 변형 기반 적대적 공격 방어 기법을 우회할 수 있음을 보여주었다.

본 논문에서는 위와 같은 기존 방어 연구의 한계점을 개선하기 위해, 이미지 변형 기반이 아닌 특정 모델과 유사한 역할과 성능을 가지는 서로 다른 모델 구조 간의 앙상블 기법을 적용하여 적대적 공격을 방어하는 연구를 제안한다.

Ⅱ. 본 론

본 논문의 실험은 다음과 같은 과정으로 진행된다. 먼저, CIFAR-10[8] 데이터세트에 대해 서로 다른 분류 모델을 각각 학습한다. 실험에서 사용한 분류 모델은 대표적인 CNN(Convolutional Neural Network) 모델인 VGGNet[9], ResNet[10], MobileNet[11], EfficientNet[12]을 사용하였다. 다음은 FGSM, PGD, C&W 공격 기법을 적용하여 각 분류 모델에 대해 적대적 사례를 생성하고 분류 정확도를 확인한다. 마지막으로, 각 분류 모델 간 앙상블 기법을 적용하여 각 적대적 사례에 대한 분류 결과를 확인하여 제안하는 기법의 적대적 공격 방어 성능을 평가한다.

표 1은 CIFAR-10 데이터세트에서 100장의 테스트 이미지를 추출하여, 공격 기법을 적용하지 않은 정상 이미지와 각 공격 기법을 적용하여 생성한 적대적

[표 1] CIFAR-10 정상 이미지 및 적대적 사례에 대한 각 분류 모델의 분류 정확도 비교

구 분	None	FGSM	C&W	PGD	평 균
VGGNet	91%	38%	63%	33%	56.25%
ResNet	79%	14%	37%	13%	35.75%
MobileNet	82%	28%	47%	28%	46.25%
EfficientNet	79%	15%	77%	6%	44.25%

사례에 대한 각 분류 모델의 분류 정확도를 비교한 것이다. 적대적 공격 기법을 적용한 결과 모든 분류 모델에 대해 분류 정확도가 낮아지는 것을 확인할 수 있다.

다음으로, 본 논문에서는 적대적 공격을 완화하기 위해 각 분류 모델의 softmax 형식의 출력값을 합산한 후 평균을 취하여 가장 높은 확률을 가지는 class를 선택하는 Soft Voting 기반의 앙상블 기법을 적용하였다. 또한, 표 1을 보면 분류 모델 별로 각 공격 기법에 대해 분류 정확도가 낮아지는 정도에 대한 차이가 있음을 알 수 있다. 즉, 분류 모델 별로 적대적 공격에 대한 강인성의 차이가 존재한다. 본 논문에서는 이러한 특징을 이용하여 강인성이 높은 모델에 높은 가중치를 부여하여 앙상블 기법을 적용하는 Weighted Voting 방식을 추가적으로 사용하였다.

[표 2] 정상 이미지 및 적대적 사례에 대한 제안하는 앙상블 기법의 분류 정확도 비교

구 분	None	FGSM	C&W	PGD	평 균
Soft	81%	58.25%	76.75%	55.75%	67.94%
Weighted	83%	61.5%	78%	59.25%	75.44%

표 2에서 제안하는 앙상블 기법의 적대적 사례에 대한 분류 정확도는 기존의 각각의 분류 모델에서의 분류 정확도에 비해 모두 향상된 것을 확인할 수 있다. 여기서 각 공격 기법에 대한 분류 정확도는 각 분류 모델에 대해 생성한 적대적 사례를 모두 입력하여 평균값을 계산한 것이다. 또한, 강인성이 높은 모델에 높은 가중치를 부여한 Weighted Voting 방식이 Soft Voting 방식보다 모든 경우에 있어 방어 효과가 높은 것을 확인할 수 있다. PGD 기반 적대적 훈련을 거친 모델은 CIFAR-10의 적대적 사례에 대한 평균 분류 정확도가 약 45.8%[2] 이며, 제안하는 기법의 PGD 기반 적대적 사례에 대한 평균 분류 정확도는 약 59.25%로써 적대적 공격에 대한 방어 성능이 뛰어남을 알 수 있다.

Ⅲ. 결 론

최근 딥러닝 기술은 지속적인 발전으로 다양한 분야에서 뛰어난 성능을 보여주고 있지만, 악의적인 데이터로 인해 오작동을 하는 것과 같이 적대적 공격에 취약한 것으로 밝혀져 많은 이슈가 되고 있다. 이러한 적대적 공격을 막기 위한 많은 연구들이 진행되었지만, 이를 우회하는 공격 연구가 나옴으로써 여전히 적대적 공격에 대한 이슈를 해결하지 못하고 있는 상태이다. 본 논문에서는 유사한 역할과 성능을 가지는 다양한 모델 구조 간 앙상블 기법을 적용하여 적대적 공격을 완화하는 방법을 제안하였다. 제안하는 기법은 FGSM, C&W, PGD 공격 기법을 통해 생성한 적대적 사례에 대한 분류 정확도를 향상함으로써 적대적 공격 방어 성능을 입증하였다. 또한, 현재 적대적 공격 방어 연구의 성능 평가를 위한 벤치마크로 주로 사용되는 PGD 기반 적대적 훈련 기법에 비해서도 높은 수준의 방어 성능을 제공함을 보였다.

향후 연구 계획으로는 기존의 노이즈 제거와 같은 이미지 변형 기반의 방어 기법과 제안하는 기법의 결합 구조와, 방어율을 더욱 향상시킬 수 있는 새로운 방어 구조를 제안할 계획이다.

ACKNOWLEDGMENT

이 논문은 2020년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2019-0-00706,정보보호핵심원천기술개발)

*이 논문은 2020년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (2019-0-01343,융합보안핵심인재양성사업)

참 고 문 헌

[1] I. J. Goodfellow, J. Shlens and C. Szegedy, “Explaining and Harnessing Adversarial Examples,” *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*, 2015.

[2] A. Madry et al., “Towards Deep Learning Models Resistant to Adversarial Attacks,” *Proceedings of the 6th International Conference on Learning Representations (ICLR)*, 2018.

[3] N. Carlini and D. Wagner, “Towards Evaluating the Robustness of Neural Networks,” *IEEE Symposium on Security and Privacy*, pp. 39-75, 2017.

[4] W. Xu, D. Evans and Y. Qi, “Feature Squeezing: Detecting Adversarial Examples in Deep Neural Networks,” *25th Annual Network and Distributed System Security Symposium (NDSS)*, 2018.

[5] D. Meng and H. Chen, “MagNet: A Two-Pronged Defense against Adversarial Examples,” *Proceedings of the ACM SIGSAC Conference on Computer and Communications Security*, pp. 135-147, 2017.

[6] C. Xie et al., “Feature Denoising for Improving Adversarial Robustness,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 501-509, 2019.

[7] A. Athalye, N. Carlini and D. Wagner, “Obfuscated Gradients Give a False Sense of Security: Circumventing Defenses to Adversarial Examples,” *Proceedings of the 35th International Conference on Machine Learning (ICML)*, 2018.

[8] A. Krizhevsky, V. Nair and G. Hinton, “CIFAR-10 (Canadian Institute for Advanced Research),” from https://www.cs.toronto.edu/~kriz/cifar.html.

[9] K. Simonyan and A. Zisserman, “Very Deep Convolutaional Networks for Large-Scale Image Recognition,” *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*, 2015.

[10] K. He et al., “Deep Residual Learning for Image Recognition,” *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770-778, 2016.

[11] A. G. Howard et al., “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications,” *arXiv preprint arXiv:1704.04861*, 2017.

[12] M. Tan and Q. V. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” *Proceedings of the 36th International Conference on Machine Learning (ICML)*, pp. 6105-6114, 2019.

인공지능 프레임워크 신뢰성 표준 현황에 관한 연구

김성한, 최영환
한국전자통신연구원

sh-kim@etri.re.kr, yhc@etri.re.kr

A Study on the Artificial Intelligence Trust Standardization

Kim Sung Han, Choi Young Hwan
ETRI

요 약

본 논문은 국내외 인공지능 프레임워크 신뢰성 관련 표준 현황을 분석하고, 특히 공적 표준화 기구인 ITU-T 및 JTC 1/SC 42 에서의 표준 개발 주요 사항에 대해 기술하고 있다. 이를 통해 향후 인공지능 신뢰성 관련 국제표준화 활동을 위한 참고로 활용 가능하다.

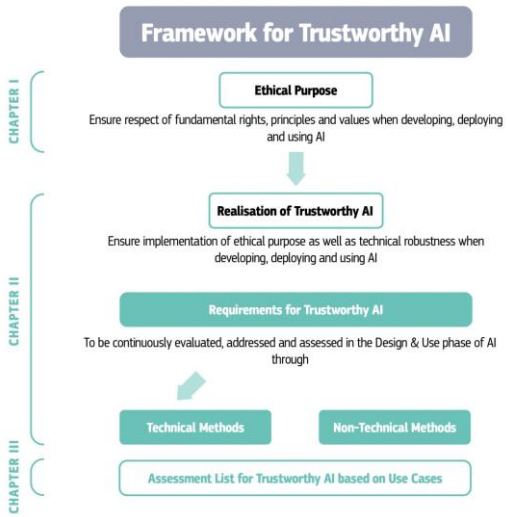
Ⅰ. 서 론

일반적인 트러스트(trust)는 사전적 의미로 타인의 미래 행동이 자신에게 호의적이거나 또는 최소한 악의적이지는 않을 가능성에 대한 기대와 믿음을 말한다. 즉, 사람을 믿고 그 사람의 미래 행동이 좋은 결과로 이어질 것이라는 믿음에 따른 행동에 대한 약속이다. 기술적인 측면에서 기존에 보안 및 프라이버시와 함께 사용자와 시스템 간의 자신감이나 확신, 신뢰하거나 의지, 믿음, 강도, 무결성, 의존성, 기대, 보증 등을 판단하고 이를 바탕으로 의사결정이 이루어질 수 있도록 제공한다.

신뢰할 수 있는 AI 를 위한 프레임워크에 대한 연구는 국제기구 및 단체에서 진행되고 있다. 이중에서 유럽에서 만든 AI 윤리 가이드 라인은 세 개의 주제로 구성되며, 각각은 추가 추상화 수준에 대한 지침을 제공하며 신뢰할 수 있는 AI 를 달성하기위한 프레임워크를 구성한다.

- 윤리적 목적: 본 주제는 AI를 다루는 모든 사람들이 준수해야하는 핵심 가치와 원칙에 중점을 둔다. 이는 EU 수준에서 EU 조약과 유럽 연합 기본권 헌장에 규정된 가치와 권리에 명시 되어있는 국제 인권법을 기반으로 한다.
- 신뢰할 수 있는 AI의 실현: 좋은 의도만으로는 충분하지 않다. AI 개발자, 배포자 및 사용자도 이러한 원칙과 가치를 기술 및 사용에 실제로 구현하기 위해 조치와 책임을 취하는 것이 중요하다. 또한 기술적 인 관점에서 시스템이 가능한 한 견고하다는 예방 조치를 취하여 윤리적 목적이 존중되더라도 AI가 의도하지 않은 해를 입히지 않도록 해야 한다. 따라서 신뢰할 수 있는 AI에 대한 요구 사항을 식별하고 이를 실현하는 데 사용할 수 있는 잠재적인 방법 (기술적 및 비 기술적)에 대한 지침을 제공한다.
- 평가 목록 및 사용 사례: 앞서 제시된 윤리적 목적과 구현 방법을 기반으로 AI 개발자, 배포자 및 사용자가 신뢰할 수 있는 AI를 운영 할 수 있는 예비 및 비포괄적 평가 목록을 설정한다.

이 지침의 구조는 아래 (그림 1) 신뢰 AI 프레임워크 가이드라인과 같다[1].



(그림 1) 신뢰 AI 프레임워크 가이드라인

Ⅱ. 표준화 현황

본 절에서는 ITU-T 및 JTC 1/SC 42 표준화 기구에서 AI 트러스트 관련 표준화 현황에 대해 기술한다.

ㅇ ITU-T Q16/13

본 Question은 네트워크의 기능을 고도화하기 위하여 기존의 상황인지 기술 등을 확장하여 지식기반으로 동적 네트워크 제어 및 관리가 가능토록 하는 표준개발을 추진하고, 미래 ICT 인프라에서 신뢰성 있는 통신

네트워크 및 서비스 제공을 위한 중점 기술의 표준화를 담당하고 있다.

ㅇ 미래 트러스트 ICT 인프라 표준화를 위한 CG

ICT융합, 사물인터넷 및 Connect 2020 결의가 채택되었으며, "미래 트러스트 및 지식 인프라" 표준의 필요성을 논의하기 위한 워크숍 및 트러스트 위한 CG-Trust를 만들었고 트러스트 정의, 유즈 케이스, 주요 기술 항목 발굴 및 향후 표준화 전략 등을 담은 기술 보고서 개발 작업 중이다.

ITU-T CG-Trust는 정량적 및 정성적 지표를 바탕으로 트러스트 관리가 이루어질 수 있도록 트러스트 레벨, 트러스트 품질 및 트러스트 인덱스 등과 같은 개념을 정립하고, 주요 기술 이슈를 도출하고, 핵심 유스 케이스 등에 대한 분석을 진행하는 등 향후 트러스트 표준화에 대한 큰 방향을 제시하고 있다[2].

번호	작업 아이템	권고 초안 명	추진일정
1	Y.3051 (ex Y.trusted-env)	The basic principles of trusted environment in ICT infrastructure	2017-02
2	Y.3052 (ex Y.trust-provision)	Overview of trust provisioning for ICT infrastructures and services	2017-02
3	Y.3053 (ex Y.trustnet-fw)	Framework of trustworthy networking with trust-centric network domains	2018-01
4	Y.3054 (ex Y.trustworthy-media)	Framework for trust-based media services	2018-05
5	Y.3053.Amendment	Trustworthy networking deployment architecture and procedures	2018-12
6	Y.trust-index	Trust index for ICT infrastructures and services	2021-07
7	Y.trust-arch	Functiaonl architecture for trust enabled service provisioning	2020-07
8	Y.SNS-trust	Framework for evaluation of trust and Quality of Media in Social Networking Services	2021-07
9	Y.trust-pdm	Framework for trust-based personal data management platform	2020-07
10	Y.PII-Did	Prioritization based de-identification methods for personally identifiable information	2020.07
11	Y.OBF_trust	Open bootstrap framework enabling trustworthy networking and services for distributed diverse ecosystem	2020.12

JTC 1/SC 42 WG3는 신뢰성 관련 표준 문서를 다수 개발하고 있으며 현재 진행중인 문서는 아래와 같다.

번호	작업 아이템	권고 초안 명
1	ISO/IEC CD 23894	Artificial Intelligence - Risk Management
2	ISO/IEC AWI TR 24027	Artificial Intelligence (AI) - Bias in AI systems and AI aided decision making
3	ISO/IEC TR 24028:2020	Artificial Intelligence (AI - Overview of trustworthiness in Artificial Intelligence
4	ISO/IEC DTR 24029-1	AI-Assessment of the robustness of neural networks - Part 1: Overview
5	ISO/IEC AWI 24029-2	AI-Assessment of the robustness of neural networks - Part 2: Formal methods methodology
6	ISO/IEC AWI TR 24368	Artificial intelligence (AI) – Overview of ethical and societal concerns
7	ISO/IEC WD 5059	Software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – Quality Model for AI-based systems
8	ISO/IEC AWI TR 5469	Artificial intelligence – Functional safety and AI systems

여 지식기반으로 동적 네트

Ⅲ. 결론

본 논문에서는 인공지능 신뢰성 표준 이슈에 대해 공적 표준 기구인 ITU-T 와 JTC 1/SC 42 에서 진행중인 표준 현황에 대해 기본적인 소개를 하였다. 인공지능 신뢰성 이슈는 사회적, 제도적인 측면부터 기술적인 측면까지 다양한 현안 사항이 있지만 본문에서는 표준화 측면에서 진행되는 부분에 대해 일부 언급하였다. 본 논문은 향후 인공지능 신뢰성 이슈에 대해 구체적이고 체계적인 접근을 위한 가이드로 활용 가능리라 사료된다.

.

참 고 문 헌

[1] "A Layered Model for AI Governance," Urs Gasser and Virgilio A.F. Almeida, Harvard University, IEEE Internet Computing, 2017.

[2] ITU-T Y.3052, "Overview of trust provisioning in information and communication technology infrastructures and services", 2017.

블록체인 기반 네트워크 관리 자동화 연구

최윤철, 박정수
한국전자통신연구원

cyc79@etri.re.kr, pjs@etri.re.kr

A Study on the Blockchain based Network Management Automation

Choi Yunchul, Park JungSoo
Electronics and Telecommunications Research Institute

요 약

네트워크에 SDN/NFV 기술을 적용하여, 다양한 서비스 환경에 적합한 리소스를 동적으로 할당한다. 또한 네트워크 관리에도 적용하여 효율성을 높이고 유연하고 민첩한 서비스를 제공한다. 그러나 관리자의 실수나 예기치 않은 장애가 발생하거나 보안 설정의 누락과 같은 문제가 발생하면 해결에 어려움이 있다. 이러한 문제를 해결하기 위하여 인텐트 기반 네트워크 기술이 나오게 되었다. 이러한 인텐트 네트워킹 기술을 기반으로 네트워크 관리 자동화를 구성하는 구조에 대해서 기술하였다.

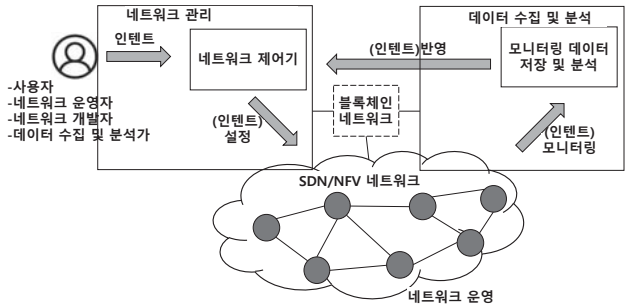
Ⅰ. 서 론

네트워크에 SDN/NFV[1,2] 기술을 적용하여, 다양한 서비스 환경에 적합한 가상화된 네트워크 리소스를 동적으로 할당할 수 있는 소프트웨어 기반의 개방화된 네트워킹 기술이 부각되고 있다. SDN 기술은 스위치와 같은 네트워크 장비의 제어 부분을 데이터 전송 부분과 분리하고, 네트워크 장비의 기능을 정의할 수 있는 오픈 API 를 외부에 제공한다. 그리고 소프트웨어 기술을 활용하여 다양한 네트워크 설정 및 제어등을 지원하는 응용을 구성하고, 이를 기반으로 중앙관리할 수 있게 도와준다. 이러한 SDN 환경에서도 네트워크 관리자가 장비 관리를 자동화하는 대에는 한계가 있다. 관리자의 실수로 예기지 않은 장애가 발생하기도 하고, 인지하지 못하는 보안 설정 누락으로 막대한 피해가 발생하기도 한다.

이러한 문제를 해결하기 위한 인텐트 기반 네트워킹 (Intent Based Networking, IBN) 기술이 이 나오게 되었다. IBN 기술은 관리자가 원하는대로 오류를 최소화하여 자동으로 네트워크 및 보안 설정을 하는 것을 목표로 하고 있다. 본 논문에서는 인텐트 기술을 활용하는 네트워크 환경에서 블록체인 기술을 적용하여 관리 자동화를 구성하는 방법에 대해서 기술하고자 한다.

Ⅱ. 본론

인텐트를 사용하는 네트워크 환경에서는 사용자, 네트워크 운영자, 개발자 또는 데이터 수집 및 분석가에 의해서 인텐트 설정을 내린다. 그리고 이렇게 설정한 인텐트는 네트워크 제어기를 통하여 SDN/NFV 네트워크 환경에 적용이 된다. 그러나 이러한 과정에 최초 입력한 사용자의 의도대로 네트워크에 반영이 되었는지 확인하는 것이 필요하고, 이를 위해서 모니터링 인터페이스가 필요하다. 그리고 이러한 모니터링 정보를 저장 및 분석하는 모듈이 필요하고, 분석 결과를 다시 네트워크 제어기에 전달하여 인텐트를 수정하는 작업이 필요하게 된다. 이러한 인텐트 사이클을 네트워크 관리 자동화로 정의하였다. 이를 위한 시스템 구성은 그림 1 과 같다.



[그림 1] 블록체인 기반 네트워크 관리 자동화 프레임워크

그림 1 에 나와 있는 블록체인 네트워크 각각의 네트워크 제어기, 데이터 수집 및 분석, SDN/NFV 네트워크 에서 주고 받는 인텐트 정보를 저장하여, 사고 발생시에 원인 규명에 활용 할수 있다. 또한 블록체인이 가지고 있는 스마트 계약 기술을 활용하여 인텐트 사이클 자동화를 수행하게 하는데 도움을 줄 수 있다.

Ⅲ. 결론

본 논문에서는 인텐트 네트워크 환경에서 네트워크 관리 자동화를 위해서 블록체인을 사용하는 구조를 살펴보았다. 그리고 모니터링 데이터에 대한 분석이 필요함을 확인하였다.

ACKNOWLEDGMENT

이 논문은 2020 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2018-0-00841, IoT 디바이스를 위한 Lightweight 블록체인 표준개발).

참 고 문 헌

[1]A. Manzalini et al., "Software-Defined Networks for Future Networks and Services," White Paper of IEEE SDN4FNS Workshop, 2014.

[2]N. McKeown et al., "OpenFlow: Enabling Innovation in Campus Networks," ACM, vol. 38, no. 2, pp. 69-74, Apr. 2008.

GEV 빔포밍을 위한 BiLSTM 기반 이진 마스크 추정

송일훈, 김홍국
광주과학기술원

ilhoon1204@gm.gist.ac.kr, hongkook@gist.ac.kr

BiLSTM-Based Binary Mask Estimation for Generalized Eigenvalue Beamformer

Ilhoon Song, Hong Kook Kim
Gwangju Institute of Science and Technology

요 약

본 논문에서는 bidirectional long short-term memory (BiLSTM) 신경망을 학습하여 generalized eigenvalue (GEV) 빔포밍을 위한 이진 마스크를 추정한다. 그리고 이를 GEV 빔포머의 가중치를 구하는 과정과 후처리 과정에 각각 적용하여 기존 BiLSTM 기반 GEV 빔포머와 비교해 본 결과, perceptual evaluation of speech quality 와 speech-to-distortion ratio 수치에서 각각 12.63%와 15.06%가 향상됨을 확인하였다.

Ⅰ. 서 론

빔포밍(Beamforming)은 잡음 환경에서 입력된 다채널 신호 각각에 특정 가중치를 주어 가중합을 구하는 방식으로 목표 음원 방향의 음질을 향상시키는 기법이다. 빔포밍 기법 중에서는 minimum variance distortionless response (MVDR)[1]과 generalized eigenvalue (GEV) 빔포밍[2]이 주로 연구되고 있다. MVDR 빔포머는 목표 음원 방향의 이득은 1 로 유지하면서 출력 잡음의 크기를 최소화한다. GEV 빔포머는 여러 음원 방향에 대한 신호 대 잡음 비를 구하여 이중 가장 큰 값을 찾으며 이를 토대로 목표 음원을 향상시킨다. 최근에는 딥러닝 기술이 발전함에 따라 빔포머 가중치를 구하는 과정에서 인공 신경망 구조를 사용하고 있으며 현재까지 bidirectional long short-term memory (BiLSTM) 신경망 구조를 이용한 GEV 빔포머가 높은 성능을 보이고 있다[3].

본 논문에서는 BiLSTM 신경망을 학습하여 이진 마스크(ideal binary mask)를 추정하고, 이를 통해 GEV 빔포머 가중치를 구한다. 또한, 가중치를 구한 이후에 후처리 과정으로 음성에 대한 이진 마스크를 적용하여 perceptual evaluation of speech quality (PESQ), speech-to-distortion ratio (SDR) 수치로 성능을 평가한다.

Ⅱ. 본론

2.1 GEV 빔포머 모델

음성과 잡음이 혼합되어 있는 다채널 입력 신호는 다음 식과 같이 정의할 수 있다.

$$\mathbf{Y}_{f,t} = \mathbf{X}_{f,t} + \mathbf{N}_{f,t} \tag{1}$$

여기서, $\mathbf{Y}_{f,t}$는 short-time Fourier transform (STFT)를 적용한 다채널 입력 신호, $\mathbf{X}_{f,t}$는 음성 신호, $\mathbf{N}_{f,t}$는 잡음 신호이다. f는 주파수이며 t는 음성 프레임을 의미한다. GEV 빔포머 가중치 $\mathbf{w}_f$ 는 [2]에 의해 다음 식으로 표현된다.

$$\mathbf{w}_f = \underset{\mathbf{w}_f}{\operatorname{argmax}} \frac{\mathbf{w}_f^H \mathbf{\Phi}_f^{(X)} \mathbf{w}_f}{\mathbf{w}_f^H \mathbf{\Phi}_f^{(N)} \mathbf{w}_f} \tag{2}$$

여기서, $\mathbf{\Phi}_f^{(X)}$와 $\mathbf{\Phi}_f^{(N)}$은 각각 음성과 잡음에 대한 공분산 행렬이며 고유값은 아래 식으로 구할 수 있다.

$$\left\{\mathbf{\Phi}_f^{-(N)} \mathbf{\Phi}_f^{(X)}\right\} \mathbf{w}_f = \boldsymbol{\lambda} \mathbf{w}_f \tag{3}$$

고유값 분해식에 의해 여러 고유값 중 가장 큰 고유값에 해당하는 빔포밍 가중치가 결정되며 이는 여러 음성 발화 방향에 대해 신호대잡음비를 구하여 최종 목표 음원 방향을 추정하는 것이다. 또한, 음성과 잡음에 대한 공분산 행렬은 다채널 입력 신호와 음성 성분의 마스크 $M_{f,t}^{(X)}$ 및 잡음 성분의 마스크 $M_{f,t}^{(N)}$로부터 얻어질 수 있다.

$$\mathbf{\Phi}_f^{(X)} = \sum_{t=1}^{T} M_{f,t}^{(X)} \mathbf{Y}_{f,t} \mathbf{Y}_{f,t}^H , \tag{4}$$

$$\mathbf{\Phi}_f^{(N)} = \sum_{t=1}^{T} M_{f,t}^{(N)} \mathbf{Y}_{f,t} \mathbf{Y}_{f,t}^H . \tag{5}$$

본 논문에서는 음성 성분의 마스크를 1, 잡음 성분의 마스크를 0 으로 판단하는 이진 마스크를 사용하며 이러한 마스크는 BiLSTM 신경망 기반으로 훈련된다. 이진 마스크 훈련 시, 실제 정답 값과 가장 가까운 값이 예측되도록 비용 함수로써 binary cross entropy (BCE)를 사용한다. 훈련 목표 값으로 정답 음성에 대한 이진 마스크와 신경망의 출력에 대한 이진 마스크를 비교하여 손실을 최소화 하도록 한다.

그림 1 은 BiLSTM 을 이용하여 이진 마스크가 훈련되는 전체적인 과정을 보여 준다. 그림 1(a)는 특징 추출 과정으로써 다채널 입력 신호를 STFT 적용 하여 513 주파수 빈으로 만들고 이를 BiLSTM 입력으로 준다. 그림 1(b)는 신경망 훈련 과정으로써 BiLSTM 층 2 개와 전결합 층(fully-connected layer) 2 개, 이후 sigmoid

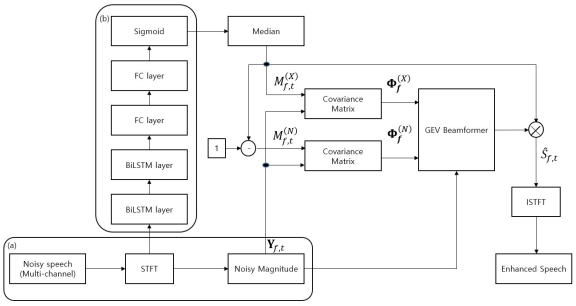


그림 1. BiLSTM 을 이용한 이진 마스크 추정 모델; (a) 특징 추출 과정, (b) 신경망 훈련 과정

함수를 거쳐 음성에 대한 다채널 이진 마스크를 생성한다. 다채널 이진 마스크는 median 연산을 통해 단일 채널 이진 마스크로 생성되며 잡음에 대한 이진 마스크는 $(1 - M_{f,t}^{X})$ 연산하여 생성한다. GEV 가중치를 구한 이후에는 다채널 입력신호에 대한 가중합으로 나타낼 수 있다. 여기서 후처리 과정으로 음성에 대한 이진 마스크를 적용하여 잡음을 최소화할 수 있으며 이렇게 향상된 음성에 대한 추정치 $\hat{S}_{f,t}$ 는 다음 식과 같다.

$$\hat{S}_{f,t} = \mathbf{w}_f^H \mathbf{Y}_{f,t} \cdot M_{f,t}^{(X)} \quad (6)$$

이후, inverse STFT 을 통해 신호를 주파수 축에서 시간 축으로 변환하며 최종적으로 향상된 단일 채널의 음성 신호를 얻을 수 있다.

2.2 실험 및 성능 평가

실험을 위해 CHiME-3 7,138 개 발화 문장으로 구성된 Bus, Cafe, Pedestrian area, Street 잡음 환경에서의 훈련 데이터로 학습을 진행하였다[4]. 테스트 과정에서는 1,320 개 발화 문장으로 구성된 테스트 데이터로 평가하였다. 성능 평가 결과, 기존 BiLSTM 기반 GEV 빔포머(BiLSTM-GEV)에 후처리 과정으로 음성에 대한 이진 마스크를 적용한 모델(BiLSTM-GEV+IBM)이 잡음을 최소화하여 각각 모든 잡음 환경에 대해 PESQ, SDR 수치를 평균적으로 각각 12.63%, 15.06% 향상시킬 수 있었다.

표 1. GEV 빔포머 모델 별 PESQ, SDR score 비교

Noise Type	Model	PESQ	SDR (dB)
Bus	Noisy data (6-ch)	1.71	0.28
	BiLSTM-GEV	2.88	5.70
	BiLSTM-GEV+IBM	3.14	8.14
Cafe	Noisy data (6-ch)	1.51	1.15
	BiLSTM-GEV	2.60	7.01
	BiLSTM-GEV+IBM	3.01	7.20
Pedestrian	Noisy data (6-ch)	1.50	1.26
	BiLSTM-GEV	2.68	7.03
	BiLSTM-GEV+IBM	3.04	7.14
Street	Noisy data (6-ch)	1.51	0.63
	BiLSTM-GEV	2.69	6.92
	BiLSTM-GEV+IBM	3.02	7.83

Ⅲ. 결론

본 논문에서는 GEV 빔포밍을 위한 BiLSTM 기반 이진 마스크 추정 기법을 소개하고, GEV 빔포머의 가중치를 구하는 과정과 후처리 과정에서 각각 이진 마스킹을 적용하였다. 그리고 이를 CHiME-3 데이터셋을 사용하여 성능 평가한 결과, PESQ, SDR 수치에서 보다 높은 성능을 보이는 것을 확인할 수 있었다.

ACKNOWLEDGMENT

본 연구는 2020 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구이며(No. 2019-0-01767, 드론을 활용한 재난 대응을 위한 기계학습 기반 음향지능 기술 개발), 그리고 2020 년도 광주과학기술원 GRI (GIST 연구원)의 지원을 받아 수행된 연구임.

참 고 문 헌

[1] C.-Y. Chen and P. P. Vaidyanathan, “Quadratically constrained beamforming robust against direction-of-arrival mismatch,” *IEEE Trans. Signal Process.*, vol. 55, no. 8, pp. 4139-4150, Aug. 2007.

[2] E. Warsitz and R. Haeb-Umbach, “Blind acoustic beamforming based on generalized eigenvalue decomposition,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 5, pp. 1529-1539, July 2007.

[3] J. Heymann, L. Drude, A. Chinaev, and R. Haeb-Umbach, “BLSTM supported GEV beamformer front-end for the 3rd CHiME challenge,” *IEEE 2015 Automatic Speech Recognition and Understanding Workshop (ASRU)*, pp. 444-451, 2015.

[4] J. Barker, R. Marxer, E. Vincent, and S. Watanabe, “The third ‘CHiME’ speech separation and recognition challenge: dataset, task and baselines,” in *Proc. of IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pp. 504-511, 2015.

인공지능과 가상현실 간의 공진화 방안에 관한 연구

유건우, 황경화*, 권오병**
경희대학교, *경희대학교, **경희대학교
yukw@khu.ac.kr, *you7i@khu.ac.kr,
**obkwon@khu.ac.kr

Study on Coevolution Method between Artificial Intelligence and Virtual Reality

Kun Woo Yoo, Kyungwha Hwang*, Ohbyung Kwon**
Kyung Hee Univ., * Kyung Hee Univ., ** Kyung Hee Univ.

요약

마샬 맥루한에 의하면 기본적으로 모든 미디어의 변이는 유기체적, 지속적 진화의 특성을 가지고 있는 것으로 파악하고 있다. 이를 4차산업혁명 기술로 확장하면 인공지능의 변이도 유기체적, 지속적 진화의 특징을 가지며, 이 유기체에는 다른 4차산업혁명기술 뿐 아니라 사회 요소도 포함된다고 볼 수 있다. 그러므로 인공지능은 다른 정보기술들과 상호작용을 통해 상대방을 진화시킬뿐더러 자신도 진화하게 된다. 본 연구의 목적은 인공지능과 가상현실 기술이 공진화하는 현상을 고찰하는 것이다.

I. 서론

4차 산업혁명 시대를 이끄는 핵심기술로 인공지능이 주목받고 있으며, 정교화된 인공지능을 개발하기 위해 학습의 질을 개선하기 위한 빅데이터 기술, 알고리즘의 효율적 구동을 위한 클라우드 기술, 센싱데이터 확보를 위한 사물인터넷 기술 등 다양한 기술들이 기여하며 같이 발전하고 있다.

이에 비해 사용자에게 가상의 환경을 마치 현실처럼 느끼게 하여 실재감을 지각시킨다는 점에서 인간이 특정 상황이나 환경에서 어떤 판단과 행동을 하는지 이해하는데 도움이 되는 가상현실 기술도[1,2,3] 인공지능과 공진화할 가능성을 가지고 있다. 하지만 아직 본격적으로 인공지능과 가상현실의 공진화 방안을 체계적으로 정리한 연구가 거의 없다. 특히 선행연구들은 주로 인공지능이 가상현실의 기술이나 콘텐츠 정교화에 도움을 줄 수 있다는 일방향적 관점에서 이루어지고 있음을 밝혀주고 있지만[4,5], 가상현실이 인공지능 발전에 어떠한 영향을 줄 수 있는지에 대한 연구는 거의 없다.

따라서 본 연구의 목적은 인공지능과 가상현실이 공진화하는 방안을 제안하는 것이다. 이를 위해 먼저 인공지능이 가상현실의 기술적 발전을 가능케 하는 현상과 동이세 가상현실 기술이 인공지능의 기술적 발전을 가져오는 현상을 정리하고자 한다.

II. 선행연구

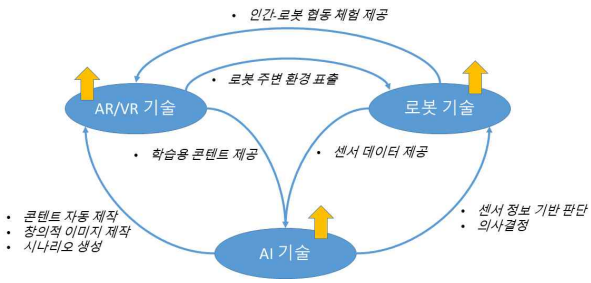
자폐 스펙트럼 장애를 지닌 아동에게 5주 동안 가상현실을 통한 사회인지훈련을 실시한 결과 자폐증상이 개선되었다는 연구[6]나 가상현실을 통해 경도인지장애를 개선시킬 수 있다고 전망한 연구[7]는 가상현실이 단순히 인간의 현재의 모습만을 반영하는 것이 아닌 미래의 어떤 변화를 이끌어내는 데에도 영향을 미칠 수 있음을 시사하고 있다. 이에 가상현실 기술은 게임, 관광, 쇼핑, 교육, 의료 등 다양한 분야에서 활용되고 있으며[8], 많은 기업들은 가상현실을 통한 사용자의 경험을 데이터화하기 위해 노력하고 있다.

III. 공진화 방안

첫째, 인공지능은 가상현실의 기술이나 콘텐츠 정교화에 도움을 줄 수 있다[4,5]. 가령, Kress et al.[9]은 AI의 딥러닝 알고리즘이 디스플레이와 광학 아키텍처를 최적화하여 인간의 시각 시스템에 맞는 가상현실 공간을 이해하고

구축하는데 도움을 줄 수 있다고 밝히고 있다.

또한 가상현실 기술은 단순히 사용자에게 이미 만들어진 콘텐츠를 제공하는 소극적 역할에서 더 나아가 사용자와 함께 사용자에게 적합한 맞춤형 콘텐츠를 제작해나가는 수준으로까지 발전해 나갈 것으로 전망할 수 있다. 이때, 인공지능이 제공하는 맞춤형 콘텐츠는 가상현실 기술에 대한 사용자 수용을 가속화하고, 인공지능은 가상현실 속에서 사용자들이 어떤 판단과 행동을 하는지 학습함으로써 인간에 대한 이해가 가속화될 것으로 판단된다. 즉, 인공지능과 가상현실 기술은 인간을 더 잘 이해하는 공진화 방향으로 발전할 것을 기대해볼 수 있다.



<그림1> 공진화 방안

ACKNOWLEDGMENT

이 논문 또는 저서는 2020년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2020S1A3A2A02093277)

참고문헌

[1] Cowan, K., & Ketron, S. (2019). A dual model of product involvement for effective virtual reality: The roles of imagination, co-creation, telepresence, and interactivity. *Journal of Business Research*, 100, 483-492.

[2] Deng, X., Unnava, H. R., & Lee, H. (2019). "Too true to be good?" when virtual reality decreases interest in actual reality. *Journal of Business Research*, 100, 561-570.

[3] Tussyadiah, I. P., Wang, D., Jung, T. H., & tom Dieck, M. C. (2018). Virtual reality, presence, and attitude change: Empirical evidence from tourism. *Tourism Management*, 66, 140-154.

[4] Li, M., Li, L., Jiao, R., & Xiao, H. (2017, October). Virtrul reality and artificial intelligence support future training development. In 2017 Chinese Automation Congress (CAC) (pp. 416-419). IEEE.

[5] Luck, M., & Aylett, R. (2000). Applying artificial intelligence to virtual reality: Intelligent virtual environments. *Applied artificial intelligence*, 14(1), 3-32.

[6] Didehbani, N., Allen, T., Kandalaft, M., Krawczyk, D., & Chapman, S. (2016). Virtual reality social cognition training for children with high functioning autism. *Computers in human behavior*, 62, 703-711.

[7] Cavedoni, S., Chirico, A., Pedroli, E., Cipresso, P., & Riva, G. (2020). Digital Biomarkers for the Early Detection of Mild Cognitive Impairment: Artificial Intelligence Meets Virtual Reality. *Frontiers in Human Neuroscience*, 14.

[8] Bricken, M., & Byrne, C. M. (1993). Summer students in virtual reality: A pilot study on educational applications of virtual reality technology. In *Virtual reality* (pp. 199-217). Academic Press.

[9] Kress, B., Pace, M., & Chatterjee, I. (2020, August). Artificial intelligence (AI) as a key enabling technology for next generation mixed reality (MR) experiences leading to mass adoption in enterprise and consumer spaces. In *Emerging Topics in Artificial Intelligence 2020* (Vol. 11469, p. 1146906). International Society for Optics and Photonics.

Metric Learning 기반 Adversarial Example 탐지 가능성에 대한 연구

최석환, 신진명, 김정구, 최윤호*
부산대학교, *부산대학교
daniailsh@pusan.ac.kr, sinryang@pusan.ac.kr, kimjg@pusan.ac.kr, *yhchoi@pusan.ac.kr

A Study on possibility of detection for adversarial examples based on metric learning

Seok-Hwan Choi, Jinmyeong Shin, Jeong Goo Kim, Yoon-Ho Choi*
Pusan National Univ., *Pusan National Univ.

요 약

딥러닝 모델은 다양한 분야에서 안정적인 성능을 보이지만 입력 이미지에 특정 노이즈를 추가하여 딥러닝 모델의 분류 정확도 감소를 유발하는 Adversarial Example에 매우 취약하다. 이러한 Adversarial Example을 방어하기 위한 기존 연구는 세 범주로 분류할 수 있다. (1) 모델 재학습 기반 방법; (2) 입력 변환 기반 방법; (3) Adversarial Example 탐지 방법. 하지만, Adversarial Example 생성 기법의 발전과 함께 발전하는 모델 재학습 기반 및 입력 변환 기반 방법과 달리 Adversarial Example 탐지 방법은 여전히 이진 분류 방법에 머물러 있다. 본 논문에서는 Metric Learning을 기반으로 한 다중 클래스 Adversarial Example 탐지 기법을 제안한다.

Ⅰ. 서 론

딥러닝 모델은 다양한 분야에서 활용되고 있으며, 자율 주행 자동차 및 맬웨어 분류와 같은 보안에 민감한 영역에서도 활용되고 있다. 하지만, 딥러닝 모델 활용의 증가와 함께 많은 보안 이슈도 등장하고 있으며, 그 중에서 입력에 사람이 인식 할 수 없는 노이즈를 추가하는 Adversarial Example은 딥러닝 모델의 분류 정확도 감소와 같은 심각한 문제를 야기할 수 있다[1]. 이러한 Adversarial Example을 방어하기 위해 많은 방어 방법이 제안되었으며 주로 세 가지 범주로 분류할 수 있다. (1) 모델 재학습 기반 방법[2]; (2) 입력 변환 기반 방법[3]; (3) Adversarial Example 탐지 방법[4]. 모델 재학습 기반 방법은 딥러닝 모델을 재학습하거나 새로운 모델로 학습하여 Adversarial Example을 방어할 수 있으며, 입력 변환 기반 방법은 Adversarial Example을 딥러닝 모델에 공급하기 전에 노이즈 제거 기법을 적용하여 Adversarial Example을 방어할 수 있다. 이러한 두 가지 방법은 Adversarial Example 생성 기법의 발전과 함께 발전하였다. 반면에, Adversarial Example 자체를 탐지하는 Adversarial Example 탐지 방법은 여전히 이진 분류 방법에 머물러 있다.

따라서 본 논문에서는 Adversarial Example 탐지 방법의 발전을 위해 Metric Learning을 기반으로 한 다중 클래스 Adversarial Example 탐지 기법을 제안한다.

Ⅱ. 본론

본 논문에서는 다중 클래스 Adversarial Example 탐지를 위해 Convolution Neural Network (CNN) 기반의 유클리드 임베딩 기술을 적용하였으며, 이를 그림 1에서 도시화 하였다. 구체적으로, 제안하는 방법은 정상 입력 및 Adversarial Example을 유클리드 공간에 맵핑하기 위해 학습 데이터셋 생성, Metric Learning 모델 학습, Advesrarial Example 탐지의 3 단계를 거쳐 동작한다.

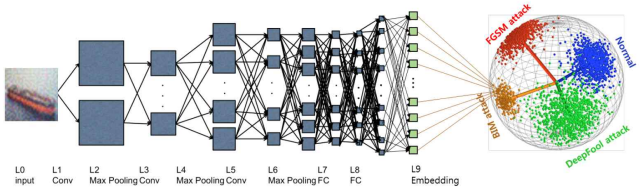


그림 1 제안하는 기법 개요도

2.1 학습 데이터셋 생성

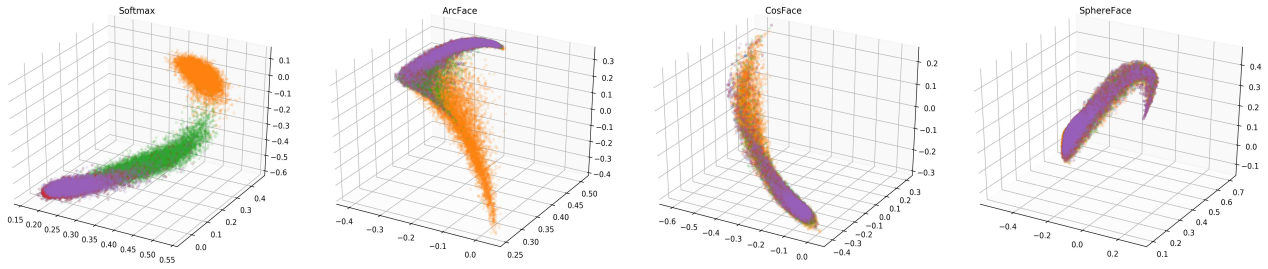
학습 데이터셋 생성 단계에서는 대상 모델에 대해 다양한 Adversarial Example 생성 기법을 적용하여 Metric Learning 모델 학습에 필요한 Adversarial Example을 생성한다. 본 논문에서는 대표적인 Adversarial Example 생성 방법인 Fast Gradient Sign Method (FGSM)[5], Basic Iterative Method (BIM)[6], DeepFool[7], C&W's Method[8]을 이용하여 Metric Learning을 위한 학습 데이터셋을 생성하였다.

2.2 Metric Learning 모델 학습

Metric Learning 모델 학습 단계에서는 원본 학습 데이터셋과 Adversarial Example 데이터셋을 사용하여 CNN 기반 Metric Learning 모델을 학습하였다. 본 논문에서는 대표적인 CNN 모델인 ResNet20을 사용하였다. 또한 본 논문에서는 Metric Learning 모델에서 주로 사용되는 Softmax, SphereFace[9], CosFace[10], ArcFace[11]의 손실함수를 이용하여 4개의 Metric Learning 모델을 학습하였다.

2.3 Adversarial Example 탐지

Adversarial Example 탐지 단계에서는 학습된 Metric Learning 모델을 이용하여 Adversarial Example을 탐지한다. 학습된 Metric Learning 모델은 입력 이미지를 유클리드 공간 상에 맵핑하므로 학습 데이터셋에



a Softmax 손실 함수 b ArcFace 손실 함수 c CosFace 손실 함수 d SphereFace 손실 함수

그림 2 다양한 손실 함수를 사용한 제안하는 기법의 유클리드 공간 상 맵핑 결과

포함된 Adversarial Example 뿐만 아니라 새로운 유형의 Adversarial Example도 탐지 할 수 있다. Metric Learning 모델 이 후 단계에서는 일반적인 분류 및 클러스터링 알고리즘을 적용할 수 있다. 따라서, 본 논문에서는 Metric Learning 모델을 통한 유클리드 공간 상에 맵핑된 결과만을 다룬다.

2.4 실험 및 검증

제안하는 방법의 성능을 평가하기 위해 본 논문에서는 CIFAR-10 이미지 분류 데이터셋에 대해 다양한 Metric Learning 손실 함수를 이용하여 실험하였다. 구체적으로, CIFAR-10 데이터셋의 전체 학습 데이터셋을 이용하여 대상 모델을 학습하였으며, 전체 테스트 데이터셋을 이용하여 Adversarial Example 생성 및 Metric Learning 모델을 학습하였다. 또한 대상 모델 및 Metric Learning 모델로는 ResNet20을 사용하였으며 Metric Learning의 출력 차원은 10차원으로 고정하였다.

그림 3은 다양한 손실 함수를 사용한 제안하는 방법의 결과를 보여준다. Softmax 손실 함수를 사용하여 Metric Learning 모델을 학습한 경우 정상 입력과 4개의 Adversarial Example 생성 기법을 유클리드 공간 상에 효율적으로 맵핑하는 것을 확인할 수 있다. 하지만, SphereFace, CosFace, ArcFace 손실 함수를 사용하여 Metric Learning 모델을 학습한 경우에는 정상 입력 및 4개의 Adversarial Example 생성 기법을 효율적으로 맵핑하지 못하였다. 이는 SphereFace, CosFace, ArcFace 손실 함수가 학습 시에 Adversarial Example 생성 기법의 특징을 반영하지 못한다는 것을 나타낸다.

Ⅲ. 결론

본 논문에서는 Metric Learning을 기반으로 한 다중 클래스 Adversarial Example 탐지 방법을 제안하였다. 다양한 손실함수를 이용한 실험을 통해 Adversarial Example의 유형을 분류 할 수 있는 가능성을 확인하였다. 하지만, 제안하는 방법은 DeepFool 및 C&W's Method과 같은 유사한 속성을 갖는 Adversarial Example에 대해 낮은 분류 성능을 보였다. 따라서, 향후 연구에서는 데이터 전처리 및 차원 감소 등의 방법을 적용하여 대부분의 Adversarial Example을 효율적인 분류할 수 있는 방안에 대한 연구가 수행되어야 할 것이다.

ACKNOWLEDGMENT

본 연구는 한국연구재단 논문연구과제 (NRF-2018R1D1A3B07043392) 지원 및 BK21플러스, IT기반 융합산업 창의인력양성사업단의 연구결과로 수행되었습니다

참 고 문 헌

[1] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, et al. “Intriguing properties of neural networks”. In: CoRR abs/1312.6199 (2013). arXiv: 1312 . 6199. URL: http://arxiv.org/abs/1312.6199.

[2] Ruitong Huang, Bing Xu, Dale Schuurmans, et al. “Learning with a Strong Adversary”. In: CoRR abs/1511.03034 (2015). arXiv: 1511.03034. URL: http: //arxiv.org/abs/1511.03034.

[3] Dongyu Meng and Hao Chen. “MagNet: a Two-Pronged Defense against Adversarial Examples”. In: CoRR abs/1705.09064 (2017). arXiv: 1705.09064. URL: http://arxiv.org/abs/1705.09064.

[4] Jiajun Lu, Theerasit Issaranon, and David Forsyth. “SafetyNet: Detecting and Rejecting Adversarial Examples Robustly”. In: The IEEE International Conference on Computer Vision (ICCV). Oct. 2017.

[5] Ian Goodfellow, Jonathon Shlens, and Christian Szegedy. “Explaining and Harnessing Adversarial Examples”. In: International Conference on Learning Representations. 2015. URL: http://arxiv.org/abs/1412.6572.

[6] Alexey Kurakin, Ian J. Goodfellow, and Samy Bengio.“Adversarial examples in the physical world”. In: CoRR abs/1607.02533 (2016). arXiv: 1607.02533. URL:http://arxiv.org/abs/1607.02533.

[7] Seyed-Mohsen Moosavi-Dezfooli, Alhussein Fawzi, and Pascal Frossard. “DeepFool: a simple and accurate method to fool deep neural networks”. In: CoRR abs/1511.04599 (2015). arXiv: 1511.04599. URL: http: //arxiv.org/abs/1511.04599.

[8] Nicholas Carlini and David A. Wagner. “Towards Evaluating the Robustness of Neural Networks”. In: CoRR abs/1608.04644 (2016). arXiv: 1608.04644. URL: http://arxiv.org/abs/1608.04644.

[9] Weiyang Liu, Yandong Wen, Zhiding Yu, et al. “Sphereface: Deep hypersphere embedding for face recognition”. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017, pp. 212 - 220.

[10] HaoWang, YitongWang, Zheng Zhou, et al. “Cosface:Large margin cosine loss for deep face recognition”.In: Proceedings of the IEEE Conference on Computer Visionand Pattern Recognition. 2018, pp. 5265 - 5274.

[11] Jiankang Deng, Jia Guo, Niannan Xue, et al. “Arcface:Additive angular margin loss for deep face recognition”. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019, pp. 4690 - 4699.