

음성 데이터를 활용한 치매 징후 진단 프로그램 개발

송민지¹, 이민지², 김도은³, 최유진⁴,
¹²³⁴한양여자대학교 소프트웨어융합학과 학부생
withminjisong@gmail.com, lmj284550@gmail.com, doeun9170@gmail.com,
swells200483@gmail.com

Development of a Dementia Early Detection Program Using Voice Data

Min-Ji Song¹, Min-Ji Lee², Do-Eun Kim³, Yu-Jin Choi⁴
¹²³⁴Dept. of Software Convergence, Hanyang Women's University

요약

이 논문은 음성 데이터를 이용하여 치매 징후를 진단하는 프로그램을 개발하는 과정과 결과에 대해 소개한다. MFCC (Mel-frequency cepstral coefficients) 기술을 사용하여 음성 패턴을 추출하고 기계 학습 모델을 적용하여 치매 징후를 탐지하는 방법을 설명한다. 실험 결과는 치매 조기 진단 및 관리에 유용한 음성 기반 도구의 중요성을 강조한다.

1. 서론

치매는 노화가 진행됨에 따라 인구의 고령화로 인해 발생 빈도가 증가하고 있는 심각한 신경 질환 중 하나이다. 따라서 치매의 조기 진단과 관리는 급속한 인구 고령화에 대응하기 위한 주요 과제이다.

치매의 진단은 주로 인지 테스트, 뇌 영상 촬영 등을 통해 이루어지고 있다. 그러나 이러한 방법들은 비용이 많이 들고, 복잡하다. 반면, 음성 데이터를 활용한 치매 징후 진단은 비침투적이며, 비교적 저렴하게 적용할 수 있는 대안적인 방법이다.

이 논문은 Mel-frequency cepstral coefficients (MFCC) 기술을 사용하여 음성 패턴을 추출하고, 기계 학습 모델을 적용하여 치매 징후를 탐지하는 방법을 탐구하였다.

이 논문은 의료 분야와 기계 학습 분야의 융합을 통해 치매 징후 진단을 개선하고, 환자의 삶의 질을 향상시키는 데 기여할 것으로 기대된다.

2. 데이터 수집과 전처리

2.1 데이터 수집

치매 징후 진단을 위한 음성 데이터를 획득하기 위해 공개 데이터베이스인 TalkBank의 DementiaBank 섹션을 활용하였다. DementiaBank는 다양한 환자 그룹과 정상 대조 그룹의 음성 및 언어 샘플을 제공

하며, 환자들의 치매 진단 정보가 포함되어 있다. 데이터는 총 340개 수집했다.

2.2 데이터 구조

이진 분류 문제를 다루기 위해 "INV" (일반인)와 "PAR" (치매 환자)로 레이블을 매핑하였다. 각 샘플은 MFCC로 표현되며 각각 13개의 계수 및 길이 100의 시간 스텝으로 구성된다.

X shape: (340, 13, 100)

y shape: (340,)

2.3 데이터 전처리

MFCC (Mel-frequency cepstral coefficients)는 음성 및 오디오 신호 처리에서 주로 사용되는 특성 추출 기술 중 하나이다. 이 기술은 음성 신호를 주파수 영역에서 멜스케일로 변환하고, 이러한 변환된 주파수 영역에서 특징 벡터를 추출한다.

CNN 모델은 이미지 데이터처럼 2D 입력을 사용한다. 현재 input shape는 (None, 13, 100, 1)입니다.

"None"은 배치 크기를 나타낸다. 이 값은 훈련 중에 결정됩니다. 13은 MFCC의 계수 수, 100은 시간 스텝 수, 1은 색상 채널 수입니다.

3. 모델 설계 (Model Architecture):

이 연구에서는 CNN (Convolutional Neural Network) 모델을 사용하였다. 음성 데이터와 같은 2D 형식의 입력 데이터에 대해 잘 작동하는 아키텍처 중 하나이기 때문이다.

CNN 모델은 다음과 같은 구조로 설계되었다.

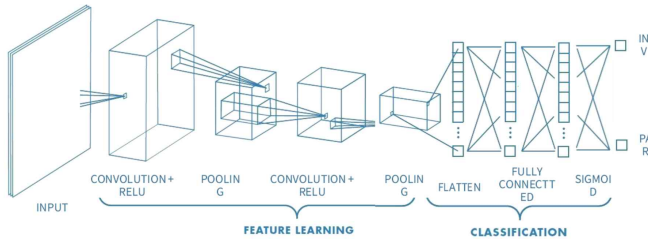


그림 1 사용한 CNN 모델의 구조 (참고한 이미지 출처: 참고문헌[2])

입력 데이터는 13x100 크기의 MFCC 특성으로 이루어진 2D 이미지이다.

Conv2D 레이어와 MaxPooling2D 레이어를 사용하여 특성 추출과 공간 차원 축소를 수행한다.

Flatten 레이어를 통해 2D 데이터를 1D로 변환한 후, Dense 레이어를 통해 신경망을 구성한다.

이진 분류 문제이므로 출력 레이어는 1개의 뉴런을 사용하며, sigmoid 활성화 함수를 적용한다.

하이퍼파라미터 조정

최적의 하이퍼파라미터 조합을 찾기 위해 Bayesian Optimization 기술을 활용하였다.

4. 모델 학습 (Model Training):

모델은 Adam 최적화 알고리즘을 사용하여 컴파일되었으며, 학습 중에는 조기 종료(Early Stopping) 기법을 사용하여 검증 데이터의 손실을 모니터링하고, 과적합을 방지하기 위해 최적의 에포크에서 학습이 종료된다.

```
training_history = train_model(X_train, y_train,
                                X_val, y_val, epochs=100, batch_size=32,
                                callbacks=[early_stopping])
```

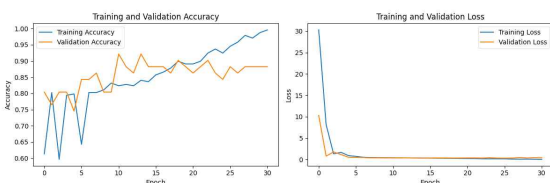


그림 2 학습 및 검증 데이터에 대한 정확도와 손실 그래프

5. 모델 평가 (Model Evaluation):

5.1 모델 정확도 (Test Accuracy)

치매 환자와 일반인을 구분하는 정도인 정확도는 88.5%로 나타났다.

5.2 Precision-Recall 곡선

정밀도는 66.7%로, 치매로 판별된 경우 중 66.7%가 실제 치매 환자인 비율을 나타낸다. 재현율은 88.9%로, 실제 치매 환자 중 88.9%가 정확하게 치매로 판별된 비율을 나타낸다.

5.3 F1 점수 (F1 Score)

모델의 정확성과 재현율 사이의 균형을 나타내는 지표인 F1 점수는 76.2%로 계산되었다.

5.4 모델 평가 그래프

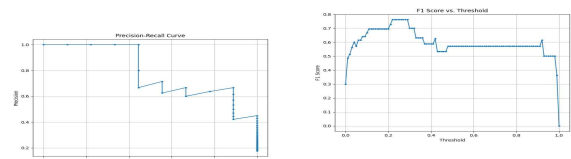


그림 3 정밀도-재현율 곡선

그림 4 임계값에 따른 F1 점수 변화 그래프

6. 결론 (Conclusion)

모델 성능: 개발한 CNN 모델의 성능은 만족스럽지만, 데이터 크기에 제한이 있어 데이터의 양 증가에 대한 잠재성을 가진다. 따라서 국내 데이터를 추가해 좀 더 좋은 결과를 만들 예정이다.

미래 연구 방향: 더 큰 데이터셋 확보, 다양한 음성 특성 고려, 음성 데이터 외의 다른 데이터와 통합한 ganzheitliche 접근법 개발

이러한 연구는 치매 징후 진단 및 치매 환자의 관리에 새로운 가능성을 제공할 것으로 기대된다. 치매 징후의 조기 발견은 환자의 삶의 질을 향상시키고 조기 개입을 가능케 하며, 의료 분야에서 중요한 역할을 할 것이다.

참고문헌

- [1] 문정현, “Deep learning-based speech recognition for Korean elderly speech data including dementia patients” KCI 한국통계학회
- [2] Nitin Panwar, 2018년 11월 16일, [https://nitin-panwar.github.io/Convolutional-neural-networks\(CNN\)-explanation-and-implementation-part-1/](https://nitin-panwar.github.io/Convolutional-neural-networks(CNN)-explanation-and-implementation-part-1/)

※ 본 논문은 과학기술정보통신부 정보통신창의인재양성사업의 지원을 통해 수행한 ICT멘토링 프로젝트 결과물입니다.