잡음 환경에서 경고음 인지력 향상을 위한 다중 합성곱 신경망 기반의 소리 분류 알고리즘 연구

강하림[1], 김도연[1], 고형일[1], 나승대[2], 김명남[3*]

경북대학교 대학원 의용생체공학과[1], 경북대학교병원 의공학과[2],

경북대학교 의과대학 의공학교실[3]

A Study on Sound Classification Algorithm based on Multiple Convolutional Neural Networks to Improve Alert Sounds Recognition in a Noisy Environment

H.L. Kang[1], D.Y. Kim[1], H.I. Koh[1], S.D. Na[2], and M.N. Kim[3*]

1 Department of Medical & Biological Engineering, Graduate School, Kyungpook National University

2 Department of Biomedical Engineering, Kyungpook National University Hospital

3 Department of Biomedical Engineering, School of Medicine, Kyungpook National University

*kimmn@knu.ac.kr

Abstract

Digital hearing aids include signal processing technology that removes noise present in amplified sound due to the development of various technologies. However, noise removed by signal processing technology has a warning sound notifying an operator of danger, and there is a problem in that a wearer who does not recognize the removed warning sound is exposed to danger. In this study, we proposed an MCNNs model that improved the performance of the PCNNs model, which can analyze sound data locally with two domain layers. The proposed method is a model in which a general convolutional layer is added to the PCNNs model, and experiments were conducted to learn and classify it in various ways than before. As a result of the experiment, it was expected that the performance of the model would increase as the amount of data computation increased. However, it was confirmed that the loss value increased due to the convolutional layer data added to improve the performance, and this decreased the model performance.

1. 연구 배경

디지털 보청기는 대표적인 청력 재활 도구이며 착용자의 난청 정도에 맞게 소리를 증폭하고 전달해 주는 기기이다. 하지만, 디지털 보청기는 주변 잡음에 따라 성능이 저하될 수 있기 때문에 잡음을 제거하는 디지털 신호처리 기술을 포함한다[1]. 기존의 신호처리 기술로 제거되는 잡음에는 난청자에게 위험을 알리는 경고음이 있다. 그리고 기존의 신호처리 기술로 제거되는 경고음을 인지하지 못한 디지털 보청기 착용자는 위험에 노출되는 문제가 생긴다. 이러한 문제를 보완하고자 최근에는 최근에는 보청기 관련 딥 뉴럴 네트워크 (deep neural network)를 기반으로 하는 음성 향상 알고리즘 연구가 이루어지고 있다[2].

병렬 합성곱 신경망(PCNNs, parallel convolutional neural networks) 모델은 딥 뉴럴 네트워크 기반으로 하는 모델이다. PCNNs 모델은 합성곱 필터를 기반으로하며 잡음에 존재하는 경고음을 분류한다. 그리고 단일 도메인층을 가지는 합성곱 신경망(CNN, convolutional neural network) 모델만으로 분석하는데 한계가 있는 복잡한 잡음 환경을 분석하기 위하여 시간 축의 특징을 분석하는 합성곱 층과 주파수 축의 특징을 분석하는 합성곱 층을 동시에 사용하여 분석한 특징을 학습하고 경고음의 유무를 분류한다.

본 연구에서는 두 개의 도메인 층으로 소리데이터를 국부적으로 분석할 수 있는 PCNNs 모델의 성능을 개선한 다중 합성곱 신경망(MCNNs, multiple convolutional neural networks) 모델을 제안하였다. 제안한 방법은 소리데이터를 국부적으로 분석할 수 있는 PCNNs 모델에 소리 데이터를 포괄적으로 분석하는 합성곱 층을 추가한 모델이며 소리 신호의 특징을 기존의 PCNNs 모델 방법보다 다양한 방면으로 학습하고 경고음의 유무를 분류하도록 하였다.

2.연구 방법

본 연구에서는 TIMIT 음성 데이터 베이스와 UrbanSound8k 잡음 데이터 베이스를 사용하였다. 그리고 잡음이 혼합된 음성 데이터와 잡음과 경고음이 혼합된 음성 데이터를 구현하였다. 구현한 데이터는 시각화하여 시간 축과 주파수 축의 변화에 따라 달라지는 진폭의 특징을 분석할 수 있는 스펙트로그램으로 전처리하였다. 전처리한 스펙트로그램의 시간 영역은 10개의 프레임(W)으로 분할하였고 주파수 영역은 40개의 멜 밴드(H)로 분할하여 40x10 크기를 가지도록 하였다.

기존의 PCNNs 모델은 합성곱 필터의 높이 (h)와 너비(w)를 수정하여 데이터의 특징을 시간 축과 주파수 축으로 각각 학습하는 방법으로 수식 (1)과 (2)와 같다.

$$FC_{ij} = \sum_{a=0}^{H-1}\sum_{b=0}^{w-1} S_f(p_i + a,\ p_j + b)\, F_h(a,b) \qquad (1)$$

$$MC_{ij} = \sum_{a=0}^{h-1}\sum_{b=0}^{W-1} S_m(p_i + a,\ p_j + b)\, F_w(a,b) \qquad (2)$$

여기서, S는 입력데이터, p_i와 p_j는 입력데이터 높이와 너비 위치, F는 합성곱 필터, a와 b는 합성곱 필터의 이동 방향을 나타낸다.

수식 (1)은 프레임 축을 기준으로 초당 2.5프레임씩

획득하는 입력 데이터를 세분화한다. 그리고 세분화한 프레임 축에 따른 주파수의 특징을 국부적으로 추출하는 합성곱 필터를 나타낸다. 수식 (2)는 40개의 멜 밴드 축을 기준으로 획득하는 데이터를 세분화한다. 그리고 세분화한 멜 밴드 축에 따른 시간 특징을 국부적으로 추출하는 합성곱 필터를 나타낸다.

본 연구에서 제안한 모델은 PCNNs 모델에 합성곱 필터의 높이와 너비가 동일한 일반적인 합성곱 층을 추가한 방법으로 수식 (3)과 같다.

$$NC_{ij} = \sum_{a=0}^{h-1} \sum_{b=0}^{w-1} S(p_i + a,\ p_j + b)\, F(a,b) \qquad (3)$$

수식 (3)은 프레임 축과 멜 밴드 축의 높이와 너비가 동일한 데이터의 특징을 포괄적으로 분석하는 합성곱 필터를 나타낸다.

그림 1은 다중 합성곱 신경망 모델의 구조를 나타낸다. 제안한 모델은 2개의 합성곱 필터와 최대 풀링 필터로 이루어진 특징 추출 영역과 추출한 특징을 분류하는 분류 영역으로 구현하였다.

3. 연구 결과

본 연구에서는 제안한 MCNNs 모델을 사용하여 다양한 잡음환경에 나타나는 경고음 유무를 분류하였다. 다양한 강도의 잡음 환경을 구현하기 위하여 ±10 dB 범위에서 5dB 간격으로 SNR을 적용하여 데이터를 구현하였다. 그리고 잡음이 혼합된 음성 데이터와 잡음과 경고음이 혼합된 음성 데이터에 SNR 적용하여 각 dB 별로 5,542개를 가지는 데이터 셋을 구축하였다. 구축한 데이터 셋의 70%는 훈련 데이터로 사용하였고, 30%는 테스트 데이터로 사용하였다.

표 1. 합성곱 신경망 모델의 학습 정확도와 테스트 결과

Model	Filter Size	Test Accuracy
CNN	3x3	0.889
	4x4	0.898
	5x5	0.899
PCNNs	40x2, 2x40	0.974
MCNNs	40x2, 2x40, 3x3	0.965

표 1은 제작한 데이터의 특징을 학습하고 분류한 모델의 테스트 결과를 나타낸다. 표 1의 결과를 보면 합성곱 필터의 크기가 커지고 분석되어지는 데이터 연산량이 많아질수록 테스트 정확도가 증가하는 것을 확인하였다. PCNNs 모델의 정확도는 97.4%로 필터의 크기가 일정한 단일 도메인 층을 가지는 CNN 모델보다 성능이 약 8% 개선되는 것으로 확인할 수 있었다. 그리고 MCNNs 모델은 97%로 CNN 모델보다 성능이 약 7% 높은 반면 PCNN 모델 보다 성능이 약 0.9% 낮은 것을 확인할 수 있었다.

4. 결론

본 연구에서는 데이터의 특징을 편향적으로 학습하고 분류하는 PCNNs 모델의 한계를 보완하기 위하여 MCNNs 모델을 제안하였다. 제안한 방법은 소리데이터를 국부적으로 분석할 수 있는 PCNNs 모델에 소리 데이터를 포괄적으로 분석하는 합성곱 층을 추가한 모델로 기존보다 다양한 방면으로 학습하고 분류하는 실험을 진행하였다. 그 결과 기존의 PCNNs 모델보다 제안한 MCNNs 모델의 성능이 떨어지는 것을 확인할 수 있었다. 데이터의 연산량이 증가할수록 모델의 성능이 높아질 것으로 예상하였지만, 성능 향상을 위해 추가한 합성곱 층 데이터로 인하여 손실값이 증가하게 되고 이로 인하여 모델 성능이 감소되는 것을 확인할 수 있었다. 향후 본 연구에서 얻어낸 결과들은 다양한 환경에서 손실되는 소리 정보들을 보다 효율적으로 분석하는데 도움이 될 것으로 판단된다.

5. Acknowledgements

This study was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. NRF-2017M3A9E2065284, 2022R1A2C2009716).

6. 참고 문헌

[1] J.Y. Jeong, J.H. Bahng, and J.H. Lee, "Efficacy of a Closed-set Auditory Training Protocol on Speech Recognition of Adult Hearing Aid Users," *Journal of Otorhinolaryngology-Head Neck Surgery*, Vol. 64, No. 2, pp. 70-76, 2020.

[2] S.W. Seo, H.W. Suh, H.J. Yu, W.Y. Seon, and S.J. Park, "Hazardous Sound Classification for the Hearing-impaired Using Deep Neural Networks," *Proceeding of Korean Institute of Information Scientists and Engineers 2017 Korea Software Congress Conference*, pp. 799-801, 2017.

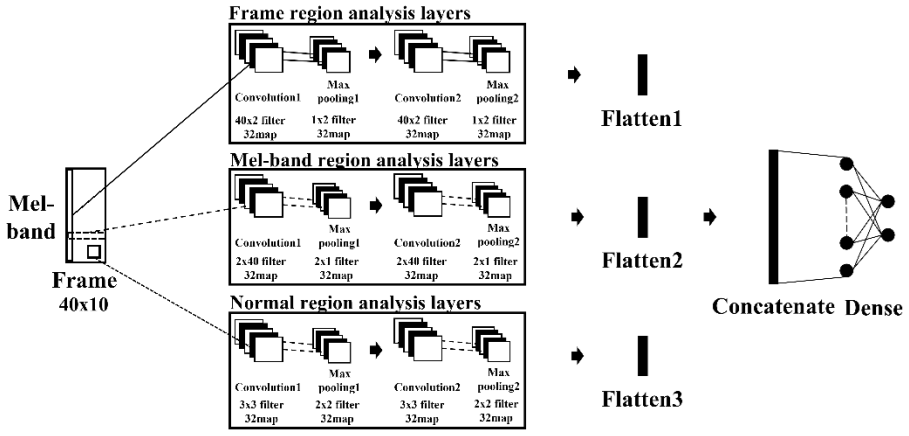

그림 1 제안한 다중 합성곱 신경망 모델 구조