시끄러운 환경에서 기침 소리를 통한 코로나 19 모니터링 연구

이청안
주식회사 에이엘아이

cheonganlee.ali@gmail.com

Monitoring COVID-19 Through Cough Sounds in Noisy Environments

Cheongan Lee
ALI Co.,Ltd.

요 약

본 연구는 기침 소리를 통해 호흡기 감염병, 특히 코로나 19 를 모니터링할 수 있는 기술을 개발하는 것을 목표로 한다. 특히, 시끄러운 환경에서도 코로나 19 를 모니터링할 수 있는 기술을 개발하였다. 이를 위해 시끄러운 환경에서 기침소리만을 추출하고 인식하는 방법과 시끄러운 환경에서 녹음된 기침소리를 바로 분류하는 두 가지 방식을 실험하였다. 본 연구에서는 시끄러운 환경의 기침소리를 바로 분류하는 방식이 더 성능이 좋음을 확인하였다. 기존의 연구들은 조용한 환경에서만 효과적이었으나, 본 연구는 시끄러운 환경에서도 코로나 19 를 분류할 수 있는 모델을 개발하여 병원, 쇼핑몰, 시끄러운 가정 등 다양한 실제 환경에서 활용 가능성을 높였다.

Ⅰ. 서 론

코로나 19 의 조기 발견과 모니터링은 감염 확산을 줄이고 환자의 조기 치료를 가능하게 하여 공중 보건에 중요한 역할을 한다. 기침 소리를 분석하여 호흡기 감염병을 모니터링하는 기술은 비침습적이고 비용 효율적이지만, 기존 연구들은 주로 조용한 환경에서만 효과적이었다. 실제 생활에서는 병원, 쇼핑몰, 가정 등 다양한 시끄러운 환경에서 기침 소리를 정확히 인식하고 분석하는 것이 필요하다.

본 연구는 시끄러운 환경에서도 코로나 19 를 모니터링할 수 있는 기침 소리 분석 기술을 개발하는 것을 목표로 한다. 이를 위해 시끄러운 환경에서 기침 소리만을 추출한 후 분류하는 방법과 시끄러운 환경에서 녹음된 기침 소리를 바로 분류하는 두 가지 접근 방식을 실험하였다. 연구 결과, 후자의 방식이 더 높은 성능을 보였다.

Ⅱ. 본론

1. 데이터

본 연구를 위해 정상인의 기침 소리와 코로나 19 감염자의 기침 소리 데이터인 Coswara 데이터셋을 활용하였다.[1] 시끄러운 환경에서의 기침 소리 데이터는 공개된 것이 없어 이를 생성하기 위해 TAU Urban Acoustic Scenes 2019 데이터셋의 도시 환경 배경 소음과 FSDKaggle2018 데이터셋의 전경 소음을 사용하였다. Scaper 라이브러리[2]를 통해 배경 소음과 전경 소음을 혼합하여 6 초 길이의 소리를 생성하였으며, 학습 데이터, 검증 데이터, 테스트 데이터는 겹치지 않도록 구분하였다. 기침 소리 추출 모델을 학습하기 위해서는 Coswara 데이터셋에 추가로 AudioSet[3]의 기침 소리 데이터를 통합하여 소음이 혼합된 소리를 생성하였다.

2. 기침 소리 추출

표 1. 기침 소리 추출 성능

모델	SI-SNR
Waveformer	5.61
Waveformer-Cough	11.99

기침 소리 추출을 위해 Waveformer[4] 모델을 사용하였다. Waveformer 는 Dilated Causal Convolution 을 인코더로, 트랜스포머를 디코더로 사용하는 구조로, 다양한 소리를 선택적으로 추출할 수 있는 모델이다. 본 연구에서는 기침 소리 추출에 특화된 모델을 개발하기 위해 Waveformer 모델을 앞서 설명한 데이터를 사용하여 미세 조정하였다.

기침 소리 추출 모델의 성능을 평가하기 위해 Scale Invariant Signal to Noise Ratio(SI-SNR)를 사용하였다. SI-SNR 은 소리의 크기에 관계없이 노이즈 대비 시그널의 크기를 평가할 수 있는 지표로, 기침 소리 추출 모델의 성능을 객관적으로 평가하는 데 적합하다. 본 연구에서 제안하는 Waveformer-Cough 기침 소리 추출 모델은 SI-SNR 11.99 의 성능을 달성하여, 기존 모델(SI-SNR 5.61) 대비 두 배 이상의 성능 향상을 이루었다.

3. 코로나 19 분류

표 2. 코로나 19 분류 성능

기침 소리 추출 여부	학습 데이터 환경	ROC-AUC
미추출	조용한 환경	0.6451
추출	조용한 환경	0.6521
미추출	**시끄러운 환경**	**0.7830**

코로나 19 분류 모델은 AudioMAE[5] 모델을 기반으로 개발하였다. AudioMAE 는 스펙트로그램의 일부를 마스킹하고 이를 다시 예측하는 방식으로 학습된다. 본 연구에서는 원래 Coswara 데이터, 본 연구에서 생성한 소음 환경에서의 기침 소리 데이터 두 가지를 활용하여 모델을 미세 조정하여 두 모델을 만들었다. 분류를 위해 AudioMAE 의 인코더 출력을 MaxPooling 하고 단층의 Feed Forward Neural Network 를 사용하여 최종 출력을 내도록 모델을 구성하였다. 본 연구에서는 MaxPooling 이 기존 AudioMAE 가 사용한 MeanPooling 보다 약간 더 높은 성능을 보였다.

모델의 성능을 평가하기 위해 ROC-AUC 를 사용하였다. ROC-AUC 는 한계값 설정에 관계없이 모델의 분류 성능을 평가할 수 있는 지표로, 민감도와 특이도를 종합적으로 고려한다. 시끄러운 환경에서 녹음된 기침 소리를 조용한 환경에서의 기침 소리로 학습된 분류 모델로 테스트하면 ROC-AUC 0.6451 의 성능을 보였다. 시끄러운 환경에서 녹음된 기침 소리를 먼저 기침 소리 추출 모델로 기침 소리만 추출한 후 이를 조용한 환경에서의 기침 소리로 학습된 분류 모델로 테스트하면 ROC-AUC 0.6521 로 약간의 성능 향상이 있었다. 시끄러운 환경에서의 기침 소리로 학습한 분류 모델은 ROC-AUC 0.7830 의 성능으로 가장 좋은 성능을 달성하였다.

Ⅲ. 결론

본 연구는 기침 소리를 통해 호흡기 감염병, 특히 코로나 19 를 모니터링할 수 있는 기술을 개발하는 것을 목표로 했다. 조용한 환경과 시끄러운 환경에서 각각 기침 소리를 분류하는 모델을 개발하고, 두 가지 접근 방식을 비교했다. 첫 번째 접근 방식은 시끄러운 환경에서 기침 소리만을 추출한 후 분류하는 방법이었고, 두 번째 접근 방식은 시끄러운 환경에서 녹음된 기침 소리를 바로 분류하는 방법이었다. 연구 결과, 시끄러운 환경에서 기침 소리를 바로 분류하는 방식이 더 높은 성능을 보였다. 최근 많은 연구에서 인공지능 시스템을 여러 구분된 모듈로 만드는 대신 하나의 종단간 모델로 통합하는 방식이 좋은 성능을 거두고 있다. 이번 연구에서도 이런 연구들과 같은 결과를 얻을 수 있었다. 종단간 모델이 더 좋은 성능을 거둔 이유는 기침 소리 추출이 완벽하게 될 수 없고 이런 오류가 다음 모델에 영향을 주게 되는데 두 모델을 따로 구분해서 만들 경우 이런 오류를 해결할 수 없기 때문으로 보인다. 종단간 모델은 학습을 통해 이러한 문제를 보다 잘 해결하는 것으로 생각한다.

본 연구의 결과는 소리 기반 질환 모니터링 기술의 실용성과 정확성을 향상시켜, 공중 보건 및 의료 분야에 중요한 기여를 할 수 있는 잠재력을 보여준다. 앞으로의 연구에서는 데이터셋의 확장과 모델의 개선을 통해 더욱 다양한 환경에서의 적용 가능성을 높이고, 실제 환경에서의 테스트를 통해 실용성을 검증할 필요가 있다.

ACKNOWLEDGMENT

본 연구는 보건복지부의 재원으로 한국보건산업진흥원의 보건의료기술연구개발사업 지원에 의하여 이루어진 것임 (과제고유번호: RS-2023-00267328)

참 고 문 헌

[1] Bhattacharya D, Sharma NK, Dutta D, Chetupalli SR, Mote P, Ganapathy S, Chandrakiran C, Nori S, Suhail KK, Gonuguntla S, Alagesan M. "Coswara: A respiratory sounds and symptoms dataset for remote screening of SARS-CoV-2 infection." Scientific Data. 2023 Jun 22;10(1):397.

[2] Salamon J, MacConnell D, Cartwright M, Li P, Bello JP. "Scaper: A library for soundscape synthesis and augmentation." In2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA) 2017 Oct 15 (pp. 344-348). IEEE.

[3] Gemmeke JF, Ellis DP, Freedman D, Jansen A, Lawrence W, Moore RC, Plakal M, Ritter M. "Audio set: An ontology and human-labeled dataset for audio events." In2017 IEEE international conference on acoustics, speech and signal processing (ICASSP) 2017 Mar 5 (pp. 776-780). IEEE.

[4] Veluri B, Chan J, Itani M, Chen T, Yoshioka T, Gollakota S. "Real-time target sound extraction." InICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2023 Jun 4 (pp. 1-5). IEEE.

[5] Huang PY, Xu H, Li J, Baevski A, Auli M, Galuba W, Metze F, Feichtenhofer C. "Masked autoencoders that listen." Advances in Neural Information Processing Systems. 2022 Dec 6;35:28708-20.