파일 내의 메타데이터 일관성을 이용한 메타데이터 없는 호흡음 분류에 관한 연구

박지현, 강승태, 정호영*
경북대학교

jihyunpia@knu.ac.kr, cdef3456@naver.com, *hoyjung@knu.ac.kr

Metadata-Free Respiratory Sound Classification Using Metadata Consistency within Files

Jihyun Park, Seungtae Kang, Ho-Young Jung*
Kyungpook National Univ.

요 약

본 논문은 각 파일의 호흡음 사이클들끼리 메타데이터가 같다는 점을 이용하여 메타 데이터 없이 호흡음 데이터의 고유한 이질성을 해결하는 학습방법을 제안한다. 호흡음 데이터는 같은 클래스의 데이터 사이에도 큰 다양성을 가지는데 이는 환자의 성별, 나이, 수집하는데 사용한 디바이스 등의 메타데이터의 영향이다. 최근의 연구들은 이 메타데이터를 함께 이용하여 이질성 문제를 해결하였다. 하지만 이러한 방법들은 데이터셋에 따른 제공되는 메타데이터 차이, 메타데이터 사용방법 등 고려할 부분이 많다. 우리는 각 파일의 호흡음 사이클들끼리 메타데이터가 같다는 점을 이용하여 각 파일을 하나의 도메인으로 정의하여 대조학습과 도메인 적응을 사용함으로써 추가적인 메타데이터 사용 없이 이질성 문제를 해결할 수 있는 방법을 제안한다. 이 방법을 사용할 때 같은 조건에서 ICBHI 2017 데이터셋에 대하여 메타데이터를 사용하지 않는 Audio-CLAP 대비 1.30%, 메타데이터를 사용하는 BTS 대비 0.63%의 성능 향상을 보였다.

Ⅰ. 서 론

본 논문에서는 호흡기 소리 분류 (respiratory sounds classification)를 메타데이터 없이 분류하는 접근방식을 제안한다. 호흡음 소리분류 문제는 호흡 소리데이터 샘플이 주어졌을때, 이 호흡음이 정상인지 비정상인지 만약 비정상이라면, 어떤 형식의 비정상인지 분류하는 문제이다. 호흡음 분류는 환자를 진단하고 경과를 관찰하는데 도움을 줄 수 있다. 최근 대두되고있는 호흡음 분류의 문제는 호흡음 데이터의 고유한 이질성이다. 이 문제를 해결하기 위한 최신의 연구들은 메타데이터를 함께 사용하는 접근법을 사용한다. 대표적인 예로 브리징 텍스트 및 사운드(BTS)는 대규모 언어 오디오 모델을 이용하여 호흡음과 메타데이터를 함께 사용한다. 하지만, 이러한 접근법은 실제 사용시에도 메타데이터가 필요하다는 문제가 있다. 우리는 메타데이터 없이도 호흡음 분류 문제를 해결 할 수 있는 모델을 제안한다. 각 파일을 각각의 도메인으로 정의하여 대조학습과 도메인 적응 훈련을 적용한 방법을 사용한다.

Ⅱ. 본론

가. Baseline

본 논문에서 사용한 baseline은 audio CLAP과 BTS 두가지 방법이다. 오디오 CLAP은 메타데이터 없이 호흡음만 사용한 모델이고, BTS는 호흡음과 메타데이터를 함께 사용한 모델이다.

우리의 실험에서는 ICBHI 2017 데이터셋을 사용한다. 이 데이터세트는 두 팀이 전자청진기를 사용하여 녹음한 126명의 피험자의 5.5시간 호흡 소리 녹음으로 구성되었다. 각각의 호흡 주기는 호흡 전문가들에 의해 균열, 쌕쌕거림, 둘다의 조합 또는 외래 호흡음이 포함되지 않는 것으로 표시된다.

Audio-CLAP(contrastive learning audio pretraining)는 CLAP의 오디오 브랜치에 분류기를 부착하여 지도 오디오 분류에 사용되는 모델이다. BTS 논문에서 기준선으로 좋은 성능을 입증했으며, 메타데이터를 활용하는 BTS의 오디오 부분과 일치하기 때문에 기준선으로 선택되었다. 분류기는 하나의 조밀한 레이어를 사용하여 다른 모든 설정은 BTS 논문의 설정과 일치한다. BTS 는 메타데이터를 언어로 변환하여 호흡음과 함께 CLAP 에 입력한 다음 분류기를 부착하는 모델이다. 현재 ICBHI 2017 데이터 세트에서 메타데이터를 사용하는 모델 중 최첨단(SOTA) 성능을 보여준다. 분류기는 하나의 조밀한 레이어를 사용하고 다른 모든 설정은 BTS 논문의 설정과 일치한다.

나. 실험 설정

각 사이클은 CLAP 의 사전 훈련 데이터와 일치하도록 48kHZ로 리샘플링 된 다음 8초의 길이로 조정된다. BTS 실험에서 메타데이터는 논문에 설명된 대로 “ 이 소리는 성인 남성 환자의 왼쪽 전방 흉부에서 메디트론 청진기를 사용.” 라는 형태로 변환된다. 모든 실험에서 우리는 LAION-Audio-630K 에서 사전 훈련된 CLAP 을 사용하며, 분류기는 하나의 조밀한 레이어로 구성되고, 프로젝터는 두 개의 조밀한 레이어로 구성된다. 우리는 학습률이 5E-5 인 Adam 옵티마이저를 사용하여 미세조정하고 코사인 스케줄링을 적용한다. 미세 조정은 배치 크기가 8일 50 개의 에포크에 걸쳐 수행된다. 분류 손실은 범주형 교차 엔트로피 손실을 사용한다. 대조 학습 손실은 infoNCE 손실을 사용한다. 대조학습을 사용 할 때 각 표본에는 8개의 음성 표본이 포함된다.

N의 값은 1,3,5,7,9 의 그리드 검색을 사용하여 결정된다.

다. 결과

Method	Specificity	Sensitivity	score
Audio-clap	**70.38**	48.94	59.66
BTS*	<u>72.6</u>	48.06	60.33
Method1	72.22	<u>49.35</u>	<u>60.76</u>
Method2	71.75	**50.16**	**60.96**

표 1. ICBHI 2017 데이터셋에서 RSC의 성능 *은 메타데이터의 사용을 나타낸다. 최고의 성능과 두 번째로 우수한 성능은 각각 굵은 글씨와 밑줄로 강조 표시된다.

표 1은 ICBHI 2017 데이터 세트 공식 분할에 대한 호흡음 분류 성능을 보여준다. 메트릭은 데이터 세트 논문의 정의를 따른다. 모든 방법은 동일한 사전 학습된 CLAP을 사용하며 5번의 반복 실험의 평균 결과를 나타낸다. 저희가 제안한 방법 1과 2는 기준 오디오-CLAP보다 1.30점 높은 점수를 달성할 뿐만 아니라 메타데이터를 사용하는 BTS논문의 점수를 0.63점 초과한다.

Method 1

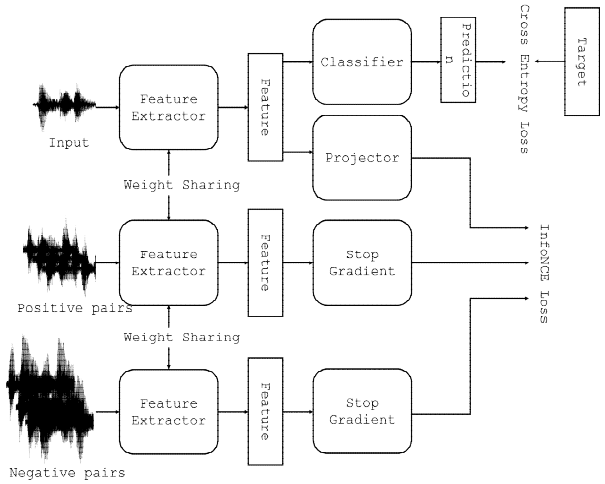

그림 1 method1 structure

분류 대상이 동일한 데이터 쌍은 서로 가까워야 하며, 동일한 파일(메타데이터가 동일함) 내의 주기로 만들어진 쌍은 가능한 한 멀리 떨어져 있어야 한다.

- positive pair : 레이블이 동일한 데이터 쌍
- negative pair : 동일한 파일의 데이터 쌍 r

방법 1 과 같이 쌍을 정의할 때 양성 쌍은 전체 데이터 세트에서 동일한 클래스 레이블을 가진 모든 데이터를 사용할 수 있지만 음성 쌍은 동일한 파일의 데이터만 사용할 수 있다.(동시에 기록됨) 그 결과 양의 쌍에 비해 음의 쌍이 훨씬 적으며, 이는 대조학습 나리오에서 일반적이지 않다. 이를 해결하기위해 방법 2 는 양의 쌍과 음의 쌍을 교환하고, 기울기 반전 레이어를 사용하여 메타데이터에 불변하는 표현을 얻는다. 호흡음 분류 데이터셋은 대부분 하나의 녹음 파일에 여러 호흡 사이클이 포함된다. 한번에 녹음된 사이클이기 때문에 이 사이클들은 모두 같은 메타데이터를 갖는다. 주의할 점은 파일이 다르다고 메타데이터가 꼭 다르지는 않다는 점이다. 이 점에 착안하여 우리는 각각의 파일을 각각의 도메인으로 가정하여 대조학습과 도메인 적응을 적용하는 방법을 제안한다. 방법 1 은 기존의 분류 손실함수와 함께 대조학습의 손실함수를 추가한 것이다. 레이블이 동일한 데이터 쌍을 양의 쌍으로, 같은 파일에 있는 데이터상을 음의 쌍으로 정의하여 지도 대조학습을 진행한다.

Method 2

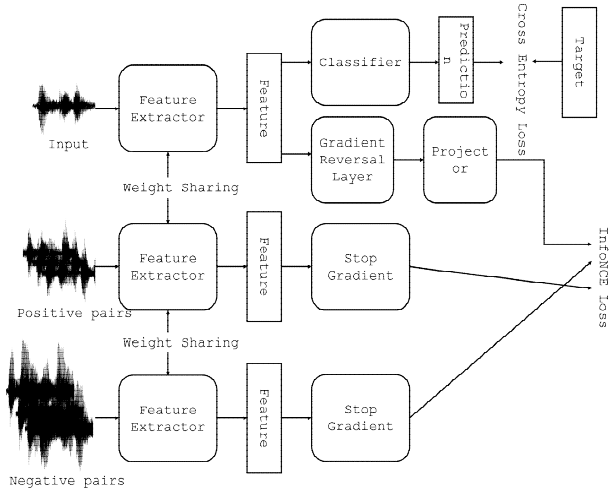

그림 2 method2 structure

방법 2 는 대조학습에 더하여 기울기 반전 계층을 이용한 도메인 적대적 학습을 추가로 사용한 방법이다. 방법 1 과 반대로 레이블이 동일한 데이터 쌍을 음의 쌍으로, 같은 파일에 있는 데이터상을 양의 쌍으로 정의하여 지도 대조학습을 진행한다. 두 방법 모두 지도 대조학습의 학습이 불안정하여 분류 손실함수가 무시되는 현상이 나타나 분류 손실함수를 N 회 반복하여 학습한 후 대조학습 Loss 를 학습하였다. N 은 [1, 9]범위를 grid search 하여 탐색하였다.

Ⅲ. 결론

Determining N

우리의 학습 알고리즘에는 새로운 하이퍼파라미터인 N이 도입된다. N은 대조적 손실 훈련이 얼마나 많은 분류 손실 훈련을 반복하는지를 결정한다. 본논문에서는 대조학습과 기울기 반전 레이어를 이용한 두가지 호흡음 분류 훈련방법을 제안한다. 이러한 방법들은 실제 메타데이터를 사용하지 않고도 메타데이터를 사용하는 효과를 달성하여, 스코어 메트릭을 기준으로 메타데이터를 사용하지 않는 Audio-CLAP에 비해 1.30, 메타데이터를 사용하는 BTS에 비해 0.63의 성능 향상을 보여준다.

Performance variations according to specific meta-data

호흡음 분류에 메타데이터를 사용하는 이유는 호흡음 데이터의 고유한 이질성 때문이다. 이는 호흡음이 환자와 이를 포착하는 데 사용된 장치에 따라 달라진다는 것을 의미한다. 따라서 메타데이터를 사용하여 원하는 효과는 메타데이터에 불변하는 표현과 점수를 얻는 것이다. 다양한 특정 메타데이터에 대한 호흡음 분류 성능을 보여준다. 메타데이터를 사용하는 BTS와 그렇지 않은 Audio-CLAP을 비교하면, 이전에 좋은 성능을 보였던 메타데이터 범주(예: 소아, 메디트론)가 크게 감소한 반면, 성능이 좋지 않은 범주(예: 트라체아)는 개선된 것으로 관찰된다. 유사한 관찰 결과는 원래 BTS 논문에서도 확인할 수 있다. 방법 1과 방법 2도 비슷한 경향을 보인다. 소아과와 메디트론의 경우 세 가지 방법의 성능이 모두 감소한 반면, 기관의 경우 세 가지 방법의 성능이 모두 증가했다. 이는 우리의 방법이 메타데이터를 포함함으로써 얻은 것과 유사한 효과를 달성한다는 것을 나타낸다.

Time and Space complexity

훈련 시간과 관련하여 엔비디아 RTX 3090의 경우 N=1일 때 방법 1과 방법 2 모두 Audio-CLAP보다 약 5배, N=5일 때 약 3배 더 오래 걸렸다. GPU 메모리 사용량도 약 1.5배 증가했다. 하지만 BTS 논문과 달리 모델 구조와 입력이 동일하게 유지되어, 시간과 공간의 복잡성을 모두 유지했기 때문에 우리 방법의 추론 시간은 변하지 않았다.

Conclusion

본 논문에서는 대조 학습과 기울기 반전 레이어를 이용한 두 가지 호흡음 분류 훈련 방법을 제안한다.
이 방법들은 효과를 달성한다. 실제 메타데이터를 사용하지 않고 메타데이터를 사용하는 경우, 스코어 메트릭을 기준으로 메타데이터를 사용하지 않는 Audio-CLAP에 비해 1.30, 메타데이터를 사용하는 BTS에 비해 0.63의 성능 향상을 보여준다.

ACKNOWLEDGMENT

참 고 문 헌

[1] Sangmin Bae, June-Woo Kim, Won-Yang Cho, Hyerim Baek, Soyoun Son, Byung Hoon Lee, Chang Woo Ha, Kyongpil Tae, Sungnyun Kim, and Se-Young Yun. Patch-mix contrastive learning with audio spectrogram transformer on respiratory sound classification. In Interspeech, 2023.

[2] Ke Chen, Xingjian Du, Bilei Zhu, Zejun Ma, Taylor BergKirkpatrick, and Shlomo Dubnov. Hts-at: A hierarchical token-semantic audio transformer for sound classification and detection. In IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, 2022.

[3] June-Woo Kim, Sangmin Bae, Won-Yang Cho, Byungjo Lee, and Ho-Young Jung. Stethoscope-guided supervised contrastive learning for cross-domain adaptation on respiratory sound classification. ArXiv, abs/2312.09603, 2023.

[4] June-Woo Kim, Miika Toikkanen, Yera Choi, Seoung-Eun Moon, and Ho-Young Jung. Bts: Bridging text and sound modalities for metadata-aided respiratory sound classification, 2024.