3D 특징 벡터를 이용한 영아 울음소리 분류

박정현, 김민서, 최혁순, 문남미
호서대학교 컴퓨터공학부
Parkjeonghyeon970609@gmail.com, kimminseo010620@gmail.com,
hyuksoon2001@gmail.com, nammee.moon@gmail.com

Classification of infant cries using 3D feature vectors

JeongHyeon Park, MinSeo Kim, HyukSoon Choi, Nammee Moon
Dept. of Computer Science and Engineering, Hoseo University

요 약

영아는 울음이라는 비언어적 의사 소통 방식을 사용하여 모든 욕구를 표현한다. 하지만 영아의 울음소리를 파악하는 것에는 어려움이 따른다. 영아의 울음소리를 해석하기 위해 많은 연구가 진행되었다. 이에 본 논문에서는 3D 특징 벡터를 이용한 영아의 울음소리 분류를 제안한다. Donate-a-corpus-cry 데이터 세트는 복통, 트림, 불편, 배고픔, 피곤으로 총 5 개의 클래스로 분류된 데이터를 사용한다. 데이터들은 원래 속도의 90%와 110%로 수정하는 방법인 템포조절을 통해 증강한다. Spectrogram, Mel-Spectrogram, MFCC 로 특징 벡터화를 시켜준 후, 각각의 2 차원 특징벡터를 묶어 3 차원 특징벡터로 구성한다. 이후 3 차원 특징 벡터를 ResNet 과 EfficientNet 모델로 학습을 진행한다. 그 결과 2 차원 특징 벡터는 0.89(F1) 3 차원 특징 벡터의 경우 0.98(F1)으로 0.09 의 성능 향상을 보여주었다.

1. 서론

영아는 일반적인 의사소통이 불가능하기 때문에 울음이라는 비언어적 의사 소통방식을 사용하여 본인의 모든 욕구를 표현하게 된다. 이때 욕구를 적절하게 해결해주지 못한다면 영아는 스트레스를 받게 된다.

하지만 양육자가 울음을 통해 영아의 욕구를 파악하는 것에는 어려움이 따르기 때문에 이 울음소리의 의미를 파악하기 위하여 오래전부터 다양한 연구가 이루어져왔다[1,2,3]. 연구에서 아이의 울음소리는 아이의 욕구에 따라 주파수와 패턴의 차이가 있음이 밝혀졌고 이를 인공지능 시스템을 통해 분석해본 결과 90%이상의 정확도로 분류가 가능하다는 것이 검증되었다[4].

본 연구에서는 영아 울음소리 분류의 정확도를 높이기 위하여 기존 연구에서 1 가지 특징벡터를 사용하는 것과는 달리 검증된 특징벡터 3 가지를 동시에 사용한다. spectrogram, Mel-Spectrogram, MFCC 특징 벡터를 추출해 3 차원 특징 벡터로 구성한 뒤 음성 분류에 뛰어난 성능을 보인 CNN 모델을 사용하는 방식을 제시한다.

2. 관련연구

2.1. 음성 분류 신경망

ResNet 은 ILSVRC 2015 에서 우승한 모델로, 신경망을 152 층의 깊이로 늘린 모델이다[5]. 기존 CNN 신경망의 기울기 소실/폭발 문제를 해결하기 위하여 Residual Block 을 사용한다. Residual Block 에는 skip/shortcut connection 이 있는데 이는 이전 층에서 넘어온 입력 x 를 출력 값 x 에 더하는 방식으로 기울기 소실 문제를 해결하였다. ResNet 기존 음성 분류 연구에서 가장 뛰어난 성능을 보여주었다[6].

<표 1> CNN 모델별 오디오 분류 성능표[6]

Architectures	AUC	mAP
Fully Connected	0.851	0.058
AlexNet	0.894	0.115
VGG	0.911	0.161
Inception V3	0.918	0.181
ResNet-50	0.916	0.182
ResNet-50	**0.926**	**0.212**

EfficientNet 은 다른 이미지넷 모델보다 적은 파라미터를 가지고 좋은 성능을 보여주는 모델이다[7].
이러한 이유로 본 연구에서는 ResNet 과 EfficientNet

을 사용한다.

2.2. 오디오 특징 벡터

오디오 신호를 학습에 사용하기 위해서는 오디오 신호의 특성을 수치적으로 나타내기 위하여 특징 벡터를 추출해야 한다[8]. 이러한 특징 벡터의 종류는 LPC (Linear Prediction reflection Coefficients), Mel-Spectrogram, MFCC(Mel-Frequency Cepstral Coefficient) 등이 있다. 본 연구에서는 Spectrogram, Mel-Spectrogram, MFCC 를 사용한다.

3. 3D 특징 벡터를 이용한 영아 울음소리 분류

3.1. 데이터 구성

학습과 평가에 사용되는 데이터 세트는 donate a cry-corpus 를 사용한다. 데이터의 형식은 'wav', 8000hz 샘플레이트, 6 초의 길이로 구성되어 있다. 데이터 세트의 클래스는 복통, 트림, 불편, 배고픔, 피곤 5 가지로 분류되어 있다. 하지만 클래스 불균형이 심해 기존 배고픔 데이터는 다운 샘플링을 진행해 30 개만 학습에 사용한다. 또한 모든 데이터는 청취하여 3 초 이상 영아의 울음 소리가 나오지 않으면 제외하였다. 그 결과 기존 총 457 개의 데이터 중 93 개를 학습용 데이터로 사용한다. 데이터의 구성표는 <표 2>와 같다

<표 2> donate a cry-corpus 구성표

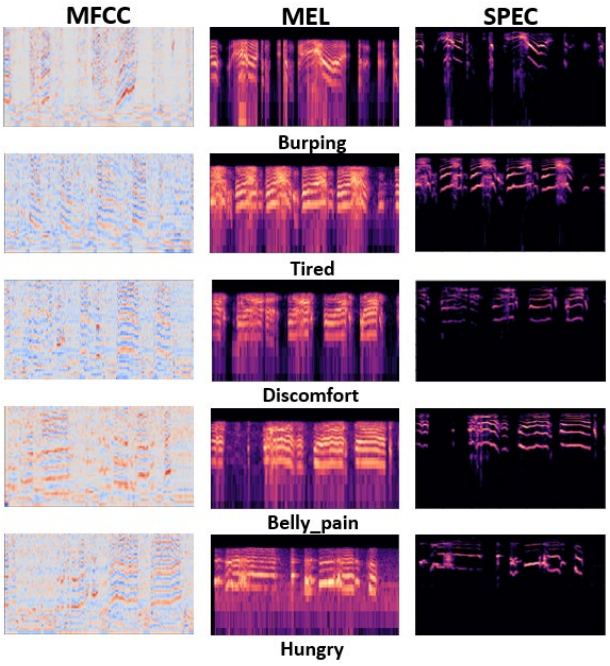
	기존	검사 후
배고픔	382	30
트림	8	7
복통	16	10
불편	27	24
피곤	24	22
총	**457**	**93**

제외된 데이터 364 개는 후에 테스트 세트로 사용한다.

3.2. 데이터 증강 및 전처리

학습용 데이터는 음성의 높낮이가 변하지 않게 오디오의 속도를 변경하는 템포 조절을 통하여 데이터 증강을 진행하였다. 데이터에서 템포 조절은 원래의 속도의 90%와 110%로 설정하는 경우가 가장 뛰어난 성능을 보인다는 연구 결과가 있다[9]. 이에 기존 93 개의 데이터를 279 개로 증강하였다.

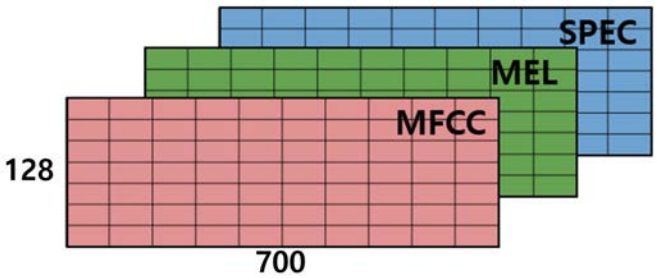
증강된 데이터들은 파이썬 라이브러리인 Librosa 를 사용하여 Spectrogram, Mel-Spectrogram, MFCC 특징 벡터화 한다. 각 클래스별 특징 벡터는 (그림 1)과 같다 특징 벡터들은 Min-Max 스케일링을 통하여 데이터의 피처가 0 과 1 사이에 위치하도록 전처리를 진행한다



(그림 1) 각 클래스의 특징벡터

4. 실험

본 연구에서는 데이터의 특징 벡터의 개수를 128 로 설정하였다. 하나의 2 차원 특징 벡터는 128 * 700 의 크기를 가지고 있다. 이러한 각각의 2 차원 특징 벡터들을 한 채널씩 담당하게 하여 3 차원 특징 벡터로 구성하여 학습에 사용하였다. 이렇게 만들어진 INPUT 의 모양은 (3, 128, 700)으로 (그림 2)와 같다



(그림 2) 3차원 특징 벡터 구성

제안된 실험의 구현은 Pytorch 를 사용하여 수행하였다. 실험은 279 개의 데이터를 학습용 데이터 223 개 검증용 데이터 56 개로 8:2 비율로 데이터를 구성하였다. <표 3>은 실험에 사용된 하이퍼파라미터의 설정표이다.

<표 3> 하이퍼 파마리터 설정표

Parameter	value
optimizer	adam
epochs	100
batch	16
learning rate	0.001

첫번째 실험은 기존 음성 분류 연구에서 가장 뛰어난 성능을 보였던 모델인 ResNet-50 을 사용하여 각 특징 벡터들을 사용해 학습을 진행하였다. 학습된 모델들은 테스트 세트를 사용하여 F1-score 를 측정하였다.

<표 4> 특징 벡터 별 성능표

Feature Vector	MODEL	F1 (macro)
Spectrogram	Resnet - 50	0.82279
Mel-Spectrogram	Resnet - 50	0.92440
MFCC	Resnet - 50	0.93111
ALL (3D Vector)	Resnet - 50	**0.98710**

실험 결과 2 차원 특징 벡터에서는 MFCC 의 성능이 가장 좋은 것으로 나타났다. 각각의 spectrogram, Mel-Spectrogram, MFCC 의 2 차원 특징 벡터를 묶어 구성한 3 차원 특징 벡터를 사용한 모델이 가장 높은 성능을 보였다.

두번째 실험은 3 차원 특징 벡터를 사용하여 ResNet 과 EfficientNet 모델을 적용해 학습을 진행하여 각 모델 별 F1-score 를 측정하였다.

<표 5> 모델별 성능표

Feature Vector	MODEL	F1 (macro)
ALL (3D Vector)	Resnet - 18	0.95996
ALL (3D Vector)	**Resnet - 50**	**0.98710**
ALL (3D Vector)	Resnet - 101	0.98665
ALL (3D Vector)	EfficientNet – B0	0.94635
ALL (3D Vector)	EfficientNet – B3	0.96113

모델은 3 차원 특징 벡터를 사용한 경우 기존의 연구의 결과처럼 Resnet - 50 에서 가장 높은 성능을 보였다.

5. 결론

본 논문에서는 오디오 신호에서 수치화 시킨 Spectrogram, Mel-Spectrogram, MFCC 의 2 차원 특징 벡터를 3 차원 특징 벡터로 묶어 학습을 진행하는 방식을 제안하였습니다. 제안한 방법을 donate a cry-corpus 데이터 세트를 사용하여 ResNet 모델과 EfficientNet 모델을 사용하여 성능을 비교하였습니다. 각각의 2 차원 특징 벡터 경우 평균 0.89(F1-score)으로 나타났지만 3 차원 특징 벡터의 경우 0.98(F1-score)으로 0.09 의 성능 향상을 보여주었다. 또한 영아 울음소리 분류의 모델로는 Resnet – 50 이 0.98(F1-score)으로 가장 높은 성능을 보여주었다.

ACKNOWLEDGEMENT

본 연구는 과학기술정보통신부와 정보통신기획평가원의 SW 중심대학사업의 연구결과로 수행되었음 (2019-0-01834)

참고문헌

[1] P. S. Zeskind and B.M. Lester, "Acoustic features and auditory perceptions of the cries of newborns with prenatal and perinatal complications", Child Dev., Vol. 49, No. 3, pp. 580-589, Sep. 1978.

[2] T. Murry and P. Amundson, "Acoustical characteristic of infant cries: fundamental frequency", Child Lang., Vol. 4, No. 3, pp. 321- 328, Oct. 1977.

[3] WhyCry Technology, http://www.why-cry.com, Apr. 05, 2019

[4] I.K Hwang*, H. B. Song "AI-based Infant State Recognition Using Crying Sound" Journal of KIIT. Vol. 17, No. 7, pp. 13-21, Jul. 31, 2019

[5] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEEconference on computer vision and pattern recognition. 2016.

[6] Shawn Hershey, Sourish Chaudhuri, “CNN Architectures for Large-Scale Audio Classification” International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE (2017)

[7] Tan, Mingxing and Quoc V. Le. “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. ” ArXiv preprint arXiv:1905.11946, 2019.

[8] J.D. Lim, S.W. Han, B.C. Choi, B.H. Chung "The Technology of the Audio Feature Extraction for Classifying Contents"

[9] S.G Lee, S.M Lee, “Data Augmentation for DNN-based Speech Enhancement” Journal of Korea Multimedia Society Vol. 22, No. 7, July 2019(pp. 749-758)