Regular paper | Journal of the KIECS. pp. 571-578, vol. 13, no. 3, Jun. 30 2018, t. 89, pISSN 1975-8170 | eISSN 2288-2189
http://dx.doi.org/10.13067/JKIECS.2018.13.3.571

DNN을 이용한 오디오 이벤트 검출 성능 비교

정석환* · 정용주**

Comparison of Audio Event Detection Performance using DNN

Suk-Hwan Chung* · Yong-Joo Chung**

요 약

최근 딥러닝 기법이 다양한 종류의 패턴 인식에 있어서 우수한 성능을 보이고 있다. 하지만 소규모의 훈련 데이터를 이용한 분류 실험에 있어서 전통적으로 사용되던 머신러닝 기법에 비해서 DNN의 성능이 우수한지에 대해서는 다소 간의 논란이 있어 왔다. 본 연구에서는 오디오 검출에 있어서 전통적으로 사용되어 왔던 GMM, SVM의 성능과 DNN의 성능을 비교하였다. 동일한 데이터에 대해서 인식실험을 수행한 결과, 전반적인 성능은 DNN이 우수하였으나 세그먼트 기반의 F-score에서 SVM이 DNN에 비해 우수한 성능을 보임을 알 수 있었다.

ABSTRACT

Recently, deep learning techniques have shown superior performance in various kinds of pattern recognition. However, there have been some arguments whether the DNN performs better than the conventional machine learning techniques when classification experiments are done using a small amount of training data. In this study, we compared the performance of the conventional GMM and SVM with DNN, a kind of deep learning techniques, in audio event detection. When tested on the same data, DNN has shown superior overall performance but SVM was better than DNN in segment-based F-score.

키워드

Feedforward Neural Network, Gaussian Mixture Model, Machine Learning, Support Vector Machine
피드포워드 뉴럴 네트워크, 가우시안 믹스쳐 모델, 기계 학습, 서포트 벡터 머신

Ⅰ. 서 론

최근 사회 전반에 걸쳐 인공지능(Artificial Intelligence)에 대한 관심이 늘고 있으며 특히 딥러닝 기법은 그 중에서도 많은 주목을 받고 있다. 이는 딥러닝에 대한 새로운 학습 알고리즘의 개발과 더불어 이를 실행시킬 수 있는 부쩍 향상된 하드웨어 기술의 발전에 크게 기인하고 있다. 최근에 많은 연구가 이루어지고 있는 오디오 신호 분류에서도 딥러닝에 대한 많은 관심이 이어지고 있다. 대표적으로는 Tampere University에서 매년 주관하고 있는 DCASE(: Detection and Classification of Acoustic Scene and Events) 챌린지(challenge)를 들 수 있으며, 여기서는 여러 가지 오디오 분류 과제를 제안하고 연구자들이 자유롭게 자신들의 연구결과를 제출하고 평가 받으며 또한 워크샵을 통하여 그 결과에 대해서 발표하고 토

* 계명대학교 전기전자융합시스템공학과 (mester88@naver.com)
** 교신저자 : 계명대학교 전자공학과
· 접 수 일 : 2018. 05. 22
· 수정완료일 : 2018. 06. 03
· 게재확정일 : 2018. 06. 15
· Received : May. 22, 2018, Revised : Jun. 03, 2018, Accepted : Jun. 15, 2018
· Corresponding Author : Yong-Joo Chung
Dept. Electronic Engineering, Keimyung University,
Email : yjjung@kmu.ac.kr

론 할 수 있는 장을 제공하고 있다[1].

GMM(: Gaussian Mixture Model)과 SVM(: Support Vector Machine)은 전통적으로 오디오 분류 분야에서 많이 활용되었던 대표적 기법이다. 먼저 GMM은 사람의 비명이나 함성, 총소리의 검출에 있어서 우수한 결과를 나타내었으며[1], 사소한 소음에도 민감하게 반응하는 블랙박스의 오작동을 방지를 위해 사용되기도 하였다[2]. SVM은 오디오 검색에서 비교적 최근까지 많은 연구들에서 적용되어 왔다. 또한 SVM은 가정 내에서 발생하는 비명 소리를 검출하는데 있어서 좋은 성능을 보였다[3]. 이밖에도 총소리와 충격 소리, 폭발 소리, 비행기 소리 등의 분류를 위해서 SVM이 성공적으로 사용되기도 하였다[4].

최근에는 패턴 인식에서 우수한 성능을 보이는 DNN(: Deep Neural Network)을 영상처리와 오디오 검출에 적용한 사례들이 보고되고 있다. 8가지 종류의 오디오 소리를 가진 UrbanSound8K 데이터에 대해서 DNN을 적용한 결과 높은 인식률을 보였고, DNN의 파라메터를 변경해 가며 인식률 변화를 관찰하기도 하였다[5]. 또한 앞에서 언급한 DCASE 2016 챌린지의 3번째 과제에서 DNN을 적용한 많은 실험 결과들이 있었으며[6-7], 특히 많은 컨택스 정보를 DNN의 입력으로 넣은 결과에서 좋은 성능을 보였다[7]. 한편 CNN(: Convolutional Neural Network)를 사용하여 카메라 이미지로부터 화재 발생 여부를 알아내거나[8], 구조적으로 중요 정보를 놓칠 수 있는 CNN의 문제점을 보완하기 위해 Fully Convolutional Network와 Conditional Random Field를 적용한 사례들이 있다[9].

위에서 언급된 바와 같이, 오디오 검출에 대해 GMM과 SVM, DNN 각각의 방식들에 대한 다수의 연구 결과들이 도출되었다. 하지만 각각의 방식을 동일한 학습데이터를 바탕으로 비교 분석한 연구는 많지 않았다. 따라서 본 연구에서는 동일한 오디오 데이터를 활용하여 GMM, SVM과 DNN 인식기의 성능을 비교하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 오디오 검출을 위해 사용된 두 가지 특징 추출 방법을 소개하고 3장에서는 본 연구에서 수행된 인식기법들에 대한 소개와 오디오 분류 규칙을 정의하였다. 4장에서 실험결과를 도출하고 마지막으로 5장에서 결론 및 향후 과제를 논의한다.

II. 특징 추출

오디오 신호는 그 자체의 불규칙성으로 인하여 파형의 형태 그대로는 인식기의 입력으로 사용할 수 없다. 따라서 각 소리 신호의 특성을 잘 설명할 수 있는 특징 값이 필요한데, 본 연구에서는 오디오 신호의 특징을 위하여 로그-멜 필터뱅크(log-mel filterbank) 출력 값과 이를 바탕으로 하는 MFCC(: Mel-frequency Cepstral Coefficients) 값을 특징으로 사용하였다. 그림 1에는 이 2가지 특징 추출을 위한 과정을 나타내었다.

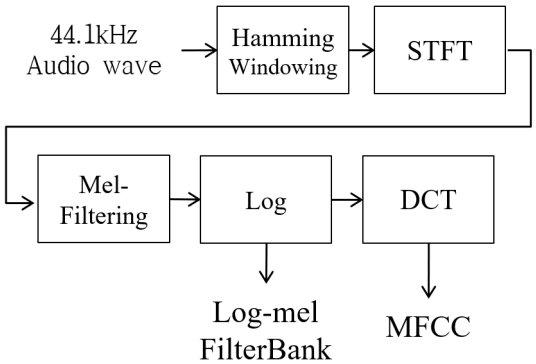

그림 1. 특징 추출 과정
Fig. 1 Feature extraction process

44.1kHz로 샘플링된 오디오 신호는 40ms 길이의 프레임(frame) 단위로 처리되며 프레임들은 20ms의 오버랩(overlap)을 허용한다. 각 프레임의 오디오 신호는 먼저 해밍(Hamming) 윈도우를 거쳐서 STFT(: Short Time Fourier Transform)을 통하여 주파수 영역으로 변환되게 된다. 이를 통해 얻어진 스펙트럼의 크기 값을 이용하여 40차원의 mel-scale의 필터뱅크가 얻어지며 로그를 취하여 로그-멜 필터뱅크 값이 구해진다. 이렇게 구해진 로그-멜 필터뱅크는 DNN의 특징으로 사용되었다.

본 연구에서 사용된 두 번째 특징인 MFCC는 앞에서

1) DCASE2018, http://dcase.community/

계산된 로그-멜 필터뱅크 값에 대하여 DCT(: Discrete Cosine Transformation)를 적용하여 얻어 지는데 각각 13차와 20차의 MFCC 특징을 구하였다. 본 논문에서는 선행연구에 기반하여 인식 모델별로 서로 다른 특징을 사용하였다. 먼저 SVM의 경우 c0를 포함한 13차의 계수를 특징으로 사용하였고, 20차의 계수에서 차분(derivative)과 차차분(acceleration) 계수를 추가하여 얻은 60차 MFCC에서 c0를 제거한 59차 계수를 GMM의 특징으로 사용하였다. 로그-멜 필터뱅크와 MFCC의 특징 추출을 위해서는 LibROSA를 사용하였다[2].

III. 인식기 구조

3.1 GMM

GMM은 M개의 가우시안 분포를 합하여 만들어진 통계적 확률모델로 전통적으로 오디오 이벤트 검출 분야에서 널리 사용되는 기법이다. GMM 확률분포는 식 (1)과 같이 표현된다.

$$b(\boldsymbol{O}_t) = \sum_{m=1}^{M} w_m N(\boldsymbol{O}_t|\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m) \quad (1)$$

$$N(\boldsymbol{O}_t|\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m) = \frac{1}{\sqrt{(2\pi)^N|\boldsymbol{\Sigma}_m|}} e^{-\frac{1}{2}(\boldsymbol{O}_t-\boldsymbol{\mu}_m)\boldsymbol{\Sigma}_m^{-1}(\boldsymbol{O}_t-\boldsymbol{\mu}_m)^T} \quad (2)$$

여기서 $N(\boldsymbol{O}_t|\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m)$은 N차 입력 벡터 $\boldsymbol{O}_t$에 대해 평균벡터 $\boldsymbol{\mu}_m$과 공분산행렬 $\boldsymbol{\Sigma}_m$을 가지는 단일 가우시안 확률 밀도 함수를 의미하며, $\boldsymbol{\omega}_m$은 각 단일가우시안 확률밀도 함수에 대한 가중치를 나타낸다.

본 논문에서 사용된 GMM 분류기는 DCASE 2016에서 제공하는 베이스라인 시스템을 기반하였으며 훈련 및 인식과정은 그림 2에 나타나 있다. 훈련과 인식을 위해 2절에서 언급한 59차 MFCC가 특징 벡터로 사용되었으며, 훈련 과정에서는 클레스에 해당하는 데이터와 해당하지 않는 데이터 각각에 대해서 GMM 모델을 생성하였다. 인식 과정에서는 테스트 데이터에 대해서 훈련과정에서 생성된 두 가지 GMM 모델 각각에 대해서 로그-우도(log- likelihood)를 계산하였다.

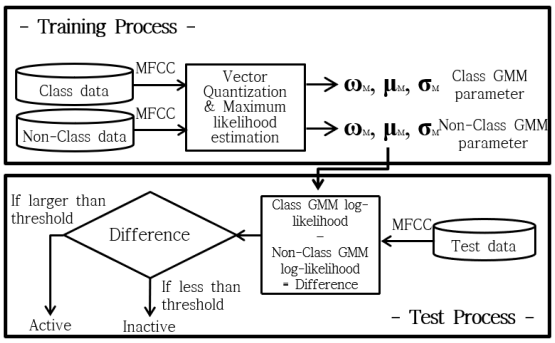

그림 2. GMM에 대한 훈련 및 인식 과정
Fig. 2 Training and recognition process of GMM

매 프레임마다 계산된 로그-우도는 1초 동안 누적되며 두 모델에서의 누적값의 차이값을 미리 설정된 임계값(threshold)과 비교해서 해당 클레스의 활성(존재)과 비활성을 결정한다. 이 1초 길이의 윈도우를 한 프레임씩 이동하여 테스트 데이터의 모든 프레임에 대해 활성과 비활성을 판단한 후 활성화된 프레임의 위치 정보를 사용하여 오디오 이벤트(특정 클레스)의 시작지점(onset)과 종료지점(offset)을 정한다[3]. 이와 같은 후처리(postprocessing) 과정은 아래 SVM과 DNN에 대해 동일하게 적용하였다.

3.2 SVM

SVM은 두 개의 클래스를 구분할 수 있는 경계의 마진(margin)을 최대화하는 학습 기법으로 미지의 패턴을 인식하는 비확률적 분류 모델이다. SVM 인식기는 식 (3)과 같이 표현된다.

$$d(\boldsymbol{x}) = \sum_{k=1}^{K} \boldsymbol{w}_k G(\boldsymbol{x}_k, \boldsymbol{x}) + b \quad (3)$$

여기서 $\boldsymbol{x}$는 테스트 특징 벡터 이고, $\boldsymbol{w}_k$와 $\boldsymbol{x}_k$, b는 각각 훈련과정에서 얻은 가중치와 서포트 벡터(support vector), 바이어스(bias)이며, K는 서포트 벡터의 수를 의미한다. 이러한 SVM 파라메터를 훈련과정에서 구하기 위해 2절에서 언급된 13차 MFCC를 사용하였다. 또한 프레임 단위의 입력 특징을 직접 사용하기 보다는 인접한 여러 프레임의 특징 평균값을

2) LibROSA, https://librosa.github.io/librosa/
3) DCASE 2016 Baseline system, https://github.com/TUT-ARG/DCASE2016-baseline-system-matlab

SVM의 입력으로 사용함으로써 인식 성능이 향상됨을 확인 할 수 있었다. 본 논문에서는 4개 프레임(80ms)의 평균값을 SVM의 입력으로 사용하였다. 그림 3에서는 SVM의 학습과 인식과정이 나타나 있다.

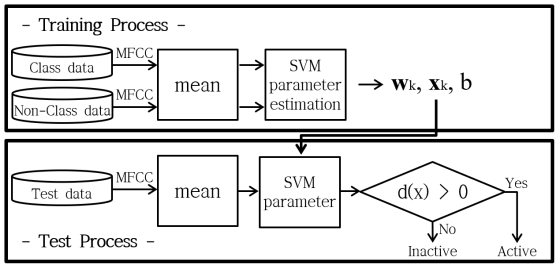

그림 3. SVM에 대한 훈련 및 인식 과정
Fig, 3 Training and recognition process of SVM

3.3 DNN

본 논문에서는 오디오 이벤트 검출에 있어서 이전에 연구된 두 가지의 DNN 구조에 대해 실험하였으며 그 구조는 표 1과 같다[10-11]. 본 논문에 사용된 DNN은 하나의 입력층과 여러 개의 은닉층(hidden layer) 및 하나의 출력층으로 구성되어 있다. 본 연구에서는 은닉층의 활성함수로는 ReLU(: Rectified Linear Unit)를 사용하였고, 훈련 시 과적합(overfitting)을 방지하기 위해 각 은닉층마다 dropout 기법을 사용하였으며 dropout rate는 두 개의 구조 각각에 대해서 0.25와 0.20을 적용하였다. 출력층은 다중 클래스 분류(multi-class classification)를 위해 분류하고자 하는 클래스 개수만큼의 유닛을 사용하였고, 각 유닛에 적용되는 활성 함수로는 sigmoid 함수를 사용하였다.

표 1. 실험에 사용된 두가지 DNN 구조
Table 1. The two DNN structures used in the experiment

	DNN1	DNN2
input	5 frames	5 frames
hidden1	1600 units(relu)	50 units(relu)
hidden2	1600 units(relu)	50 units(relu)
output	sigmoid	sigmoid
optimizer	Adam, SGD	Adam, SGD

DNN의 입력으로는 해당프레임을 포함한 주변의 5개의 프레임을 연결해 사용하였으며, 전체적으로 200차의 로그-멜 필터뱅크 특징으로 구성되었다. 학습과정에서 과적합을 방지하기 위해서 앞에서 언급된 dropout 기법 외에도 본 연구에서는 조기 종료(early stopping)를 적용하였다. DNN1에서는 최대 200회 반복(epoch)을 수행하며 매 회마다 검증(validation) 데이터를 대상으로 비용 함수의 값이 개선되지 않으면 조기 종료한다. DNN2에서도 DNN1과 마찬가지로 최대 200회 반복을를 수행하지만 100회 이후에 비용함수 값이 개선되지 않으면 조기 종료가 수행된다. DNN의 파라메터를 개선시키는데 사용되는 최적화 기법으로는 Adam(: Adaptive Moment estimation)과 SGD(: Stochastic Gradient Descent)를 적용하였으며 서로간의 성능을 비교하였다.

DNN 학습과 인식을 위해 TensorFlow를 백엔드(back-end)로 한 keras를 사용하였고, 이를 통해 다양하고 신뢰도 높은 DNN 모델을 구성할 수 있었다[12].

Ⅳ. 실험결과

본 논문의 실험에서 사용된 오디오 데이터는 DCASE 2016 Task3에서 제공된 데이터를 사용하였다. 해당 데이터는 오디오 파일과 주석 파일(annotation data)을 가지고 있다. 주석 파일에는 오디오 파일 내에 존재하는 오디오 이벤트의 시작지점과 종료지점을 초단위로 기록한 정보가 있으며 이를 바탕으로 인식기를 훈련하고, 테스트 후 평가를 수행한다. 모든 DNN에 인식 결과는 10회 반복 실험하여 평균값과 최대값, 최소값을 기록하였다.

4.1 인식 방법

본 연구에서 사용된 성능 지표는 세그먼트 기반의 방식과 이벤트 기반의 방식으로 나눠진다.

세그먼트 기반의 방식은 테스트 데이터를 1초단위로 나누고 각 1초마다의 성능을 나타내는 것이고, 이벤트 기반 방식은 특정 클레스의 이벤트가 발생할 때마다의 성능을 나타내는 것을 의미한다. 이를 위해 사용되는 성능지표로는 ErrorRate와 F-score가 사용되고 그들은 TP(: True Positive)와 FP(: False

Positive), FN(: False Negative)에 의해 계산되며, 각 항목은 아래와 같이 정의된다.

① TP: 세그먼트 혹은 이벤트 내에 최소 하나의 프레임이 활성화되어 있고, 주석 데이터 또한 동일한 지점에서 동일한 이벤트가 존재할 때.

② FN: 세그먼트 혹은 이벤트 내에 검출이 이루어지지 않았으나, 주석 데이터에는 이벤트가 존재할 때.

③ FP: 세그먼트 혹은 이벤트 내에 최소 하나의 프레임이 활성화되어 있으나, 주석 데이터에서는 존재하지 않을 때.

이렇게 정의된 지표는 전체 테스트 데이터에 걸쳐 누적되어 인식 성능의 요소로 사용된다. 정확도(precision)와 민감도(recall)는 식(3)과 같이 계산된다.

$$P=\frac{TP}{TP+FP} \quad R=\frac{TP}{TP+FN} \tag{3}$$

또한 두 지표의 조화 평균을 통해 최종적으로 (4)와 같이 F-score가 구해진다.

$$F=\frac{2 \cdot P \cdot R}{P+R} \tag{4}$$

두 번째 지표인 ErrorRate는 FN와 FP를 사용해 대체(substitution), 삭제(deletion), 삽입(insertion)을 결정하고 식(5)과 같이 계산된다.

$$ErrorRate=\frac{\sum S(k)+\sum D(k)+\sum I(k)}{\sum N(k)} \tag{5}$$

여기서 k는 각각 세그먼트의 수와 이벤트의 수를 의미한다[13].

4.2 실험 결과

표 2와 표 3에서 세그먼트 기반의 인식성능과 이벤트 기반의 인식성능을 나타내었다. 표 2에서 DNN은 GMM보다 우수한 성능을 보였다. DNN1과 DNN2는 거의 동일한 인식 성능을 보여 주었는데, 이는 주어진 학습 데이터에서 은닉층의 유닛의 개수가 인식성능에 거의 영향을 미치지 않음을 알 수 있었다. SVM은 4가지 인식기 중 가장 뛰어난 F-score성능을 보여주었다. 반면에 ErrorRate에서는 SGD 최적화를 사용한 DNN1이 가장 우수하였다.

표 2. 세그먼트 기반의 성능지표
Table 2. Segment based performance Metrics

	F-score	ErrorRate
GMM	24.30%	0.970
SVM	**32.50%**	1.020
DNN1(Adam)	29.7±1.3%	0.98±0.02
DNN1(SGD)	28.8±0.6%	**0.87±0.10**
DNN2(Adam)	29.7±0.7%	0.92±0.01
DNN2(SGD)	27.6±0.4	0.87±0.01

표 3. 이벤트 기반의 성능지표
Table 3. Event based performance Metrics

	F-score	ErrorRate
GMM	1.20%	2.020
SVM	1.50%	4.300
DNN1(Adam)	1.9±0.35%	3.07±0.05
DNN1(SGD)	**2.2±0.40%**	**1.90±0.05**
DNN2(Adam)	1.3±0.5%	2.68±0.07
DNN2(SGD)	1.5±0.3%	2.12±0.03

표 3에서는 세그먼트 기반의 성능과 달리 F-score와 ErrorRate 모두 SGD 최적화를 사용한 DNN1이 우수했다. 따라서 표 2와 표 3으로 판단한 결과 GMM, SVM에 비해 DNN이 가장 우수함을 알 수 있었다.

아래 그림 5와 6은 DNN 훈련 과정에서 Adam 최적화와 SGD 최적화의 반복 횟수별 비용함수 값의 변화를 나타낸 것이다. train loss는 훈련 데이터에 대한 비용함수 값이며, SGD 최적화보다 Adam 최적화에서 빠르게 0에 수렴함을 알 수 있었다. 한편 val loss는 검증 데이터에 대한 비용함수 값이고, Adam 최적화에서는 지속적으로 증가한 반면 SGD 최적화에서는 크게 변화하지 않았다. 이를 통해 Adam 최적화가 SGD 최적화에 비해 상대적으로 과적합에 취약함을 알 수 있었다. 또한 DNN 학습 데이터로 계산한 train loss의 값과 달리, val loss는 훈련에 사용된 데이터가 아니므로 비용함수의 차이가 크게 나타남을 알 수 있었다.

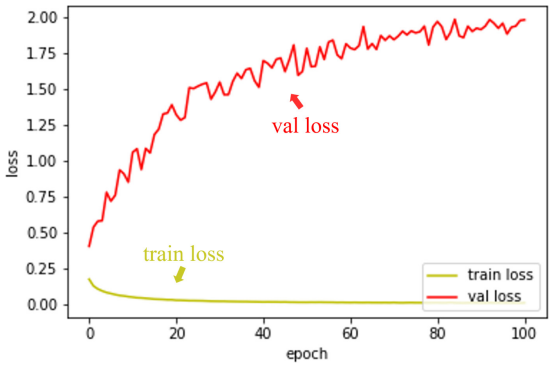

그림 5. DNN을 Adam으로 최적화했을 때, 반복 횟수별 훈련 데이터와 검증 데이터에 대한 비용함수의 변화

Fig. 5 When DNN is optimized as Adam, cost function changes for epoch-specific training data and verification data

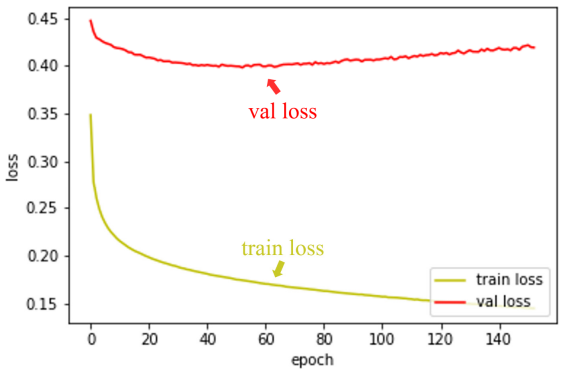

그림 6. DNN을 SGD으로 최적화했을 때, 반복 횟수별 훈련 데이터와 검증 데이터에 대한 비용함수의 변화

Fig. 6 When DNN is optimized as SGD, cost function changes for epoch-specific training data and verification data

V. 결 론

DNN기법은 기존의 머신러닝 기법들과 비교하여 좋은 인식성능을 보임이 알려져 있다. 그러나 본 논문의 연구 결과 훈련 데이터가 충분하지 않는 경우 DNN은 SVM과 GMM에 비해 월등히 나은 성능을 보이지는 않았다. 또한 DNN에 사용된 Adam 최적화 기법은 SGD 최적화 기법에 비해서 수렴속도가 빠름을 알 수가 있었다. 그러나 훈련 데이터에 대해서 Adam 최적화 기법은 sgd 최적화 기법에 비해 과적합 문제가 발생함을 알 수가 있었다. 이는 본 연구에 사용된 훈련데이터의 양이 충분하지 않은 것에 기인한다고 판단된다.

향후 과제로 다양한 데이터에 대한 반복 실험과 더불어 최근 널리 사용되는 CRNN 기법을 오디오 이벤트 검출에 적용하고자 한다.

감사의 글

이 논문은 2017년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임.(NRF-2015R1D1A1A01059925)

References

[1] L. Gerosa, G. Valenzise, M. Tagliasacchi, F. Antonacci, and A. Sarti, “Scream and Gunshot Detection in Noisy Environments,” In *Proc. the IEEE Conf. on Signal Processing,* Poznan, Poland, Sept. 2007.

[2] J. Park, J. Lim, J. Yang, J. Kyung, and M. Hahn, “False Positive Movie Clip Decision in Black-box Using Car Door-Closing Sound Classification,” In *Proc. the Institute of Electronics Engineers of Korea,* vol. 2014, no. 6, 2014, pp. 761-763.

[3] W. Huang, T. Chiew, H. Li, T. Kok, and J. Biswas, “Scream detection for home applications,” In *Proc. the IEEE Conf. on Industrial Electronics and Applications,* Taichung, Taiwan, June 2010.

[4] S. Oh, J. Uee, H. Lee, Y. Chung, and D. Park, "Abnormal Sound Detection and Identification in Surveillance System,” *J. of Korean Institute of Information Scientists and Engineers,* vol. 39, no. 2, 2012, pp. 144-152.

[5] M. Lim, D. Kim, K. Kim, and J. Kim, “Audio Event Classification Using Deep Neural Networks,” *J. of the Korean Society of Speech Sciences,* vol. 7, no. 4, 2015, pp. 27-33.

[6] D. Wei, J. Li, P. Pham, S. Das, and Shuhui Qu, Florian Metze, “Sound Event Detection for Real

Life Audio DCASE Challenge," In *Proc. European Signal Processing Conf. on Detection and Classification of Acoustic Scenes and Events,* Budapest, Hungary, Sept. 2016.

[7] Q. Kong and I. Sobieraj, W. Wang and M. Plumbley, "Deep Neural Network Baseline for DCASE Challenge 2016," In *Proc. European Signal Processing Conf. on Detection and Classification of Acoustic Scenes and Events,* Budapest, Hungary, Sept. 2016.

[8] S. Bang, "Implementation of Image based Fire Detection System Using Convolution Neural Network," *J. of the Korea Institute of Electronic Communication Sciences,* vol. 12, no. 2, 2017, pp. 331-336.

[9] S. Lim and D. Kim, "Semantic Segmentation using Convolutional Neural Network with Conditional Random Field," *J. of the Korea Institute of Electronic Communication Sciences,* vol. 12, no. 3, 2017, pp. 451-456.

[10] E. Çakır, G. Parascandolo, T. Heittola, H. Huttunen, and T. Virtanen, "Convolutional Recurrent Neural Networks for Polyphonic Sound Event Detection," *EEE/ACM Trans. Audio, Speech, and Language Processing,* vol. 25, no. 6, 2017, pp. 1291-1303.

[11] A. Mesaros, T. Heittola, A. Diment, B. Elizalde, A. Shah, E. Vincent, B. Raj, and T. Virtanen, "DCASE 2017 Challenge setup: Tasks, datasets and baseline system" In P*roc. DCASE 2017 - Workshop on Detection and Classification of Acoustic Scenes and Events,* Munich, Germany, Nov. 2017.

[12] Y. Lee and P. Moon, "A Comparison and Analysis of Deep Learning Framework," *J. of the Korea Institute of Electronic Communication Sciences,* vol. 12, no. 1, 2017, pp. 115-122.

[13] A. Mesaros, T. Heittola, and T. Virtanen, "Metrics for polyphonic sound event detection," *Applied Sciences,* vol. 6, no. 6, 2016, pp. 321-337

저자 소개

정석환(Suk-Hwan Chung)

2017년 계명대학교 전자공학과 졸업 (공학사)
2017년~현재 계명대학교 일반대학원 전기전자융합시스템공학과 석사과정
※ 관심분야 : 인공지능, 오디오 검출

정용주(Yong-Joo Chung)

1988년 서울대학교 전자공학과 졸업 (공학사)
1990년 한국과학기술원 전기및전자공학과 졸업(공학석사)
1995년 한국과학기술원 전기및전자공학과 졸업(공학박사)
1999년 ~ 계명대학교 전자공학과 교수
※ 관심분야 : 음성인식, 멀티미디어신호처리