

딥러닝 기반의 음원검색 및 분류 시스템

이세훈*, 정의중^o

^o인하공업전문대학 컴퓨터시스템공학과

e-mail: seihoon@inhac.ac.kr*, panzz8307@naver.com^o

Deep Learning based Music Classification System

Sei-Hoon Lee*, Ui-Jung Jeong^o

^oDept. of Computer Engineering, Inha Technical College

● 요약 ●

본 논문에서는 음악을 듣고 어떤 음악인지 인식하고 판별하는 음원분류 시스템과 해당 기술 구현을 딥러닝을 통해 적용하도록 제안하였다. 제안한 시스템은 인공지능추진경망을 통해 음원파일을 여러 음원 특징 추출 모델에 따라 검출된 특징들을 학습하여 해당 음원의 고유한 보컬이나 반주의 특색 등을 찾아내어 이를 인식할 수 있도록 구현하였다. 이를 통해, 기존의 Fingerprint 방식의 데이터베이스 검색 시스템과는 다른 접근방식으로 보다 사람이 음악을 기억하는 방법에 가깝도록 구현하여 능동성과 유연성을 개선하고 다양한 응용분야로 활용할 수 있는 시스템을 제안하였다.

키워드: 음악 분류(Music classification), 딥러닝(Deep Learning), 소리특징인식(Sound feature recognition)

I. Introduction

최근 머신러닝 기술이 부각 받으며 음성 데이터를 인공지능이 활용하면서 사운드 처리 프로세스 분야에서도 기존까지 컴퓨터만으로는 하지 못했던 분야를 컴퓨터가 인간처럼 생각할 수 있도록 하는 연구들이 진행되고 있다[1]. 이런 인공지능 형태의 학습 방식은 기존의 Fingerprint 방식이 사운드를 주파수의 특징 점을 데이터베이스에서 검색 하는 구조에서 인간이 소리파형을 그보다는 복잡하고 다양한 연관 작용을 해내는 것과 같이 음원을 인식하거나 특징을 파악하는 분야로 발전할 수 있다. 따라서 음악을 듣고 어떤 노래인지를 넘어 가수나 반주의 특정 부분이 어떤 특징을 가지는지를 사람이 무의식적으로 인지하듯, 본 논문에서는 이를 찾아내는 방식을 인공지능경망이 음악을 듣고 구분하는 음원 분류 시스템을 제안한다.

II. System Design and Experiment

1. System Architecture

1.1 Sound source feature extraction algorithm

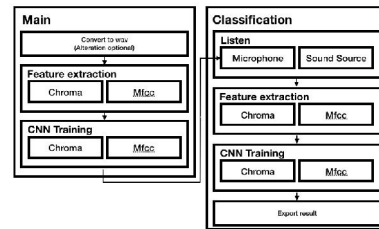


Fig. 1. System Sequence Flow

Fig. 1과 같이 본 시스템에서는 wav파일을 학습시키기 위해 특징 추출 알고리즘으로 특징을 추출한다. 먼저 각각 노래는 무작위로 선정한 노래 5개와 그 노래와 장르, 가수가 같거나 비슷한 노래 5개, 총 10곡을 선정하였다. 이후 두 가지 음원 특징 추출 알고리즘을 사용하여 두 개의 딥러닝 모델을 제작하며 화성정보를 표현하는 Chroma 특성 벡터와 음색을 표현하는 Mfcc[2]를 사용하여 각각 44100Khz 샘플링 데이터의 전체 음원파일에서 초당 100개의 데이터로 분할하여 사용한다.

1.2 CNN(Convolutional neural network)

본 시스템은 음악을 선율에 따라 검색하여 찾는 방법이 아닌 음원과 일의 고유 특징들과 음색을 학습하여 스스로 해당 음원의 특징을 찾아내는 데에 중점을 두었다. 이를 위해 CNN기반에 Relu, Hyperbolic tangent함수로 이루어진 3차 히든레이어를 각각 300, 300, 100개로 튜닝 하여 98%의 Accuracy를 보일 때 까지 학습된 10개의 음원 중 가장 그러한 특징이 많이 포함된 노래를 분류할 수 있도록 한다.

1.3 Music Classification

음악을 들려줄 때 학습한 음악을 인식하는 시나리오를 실험에 사용한 노트북의 내장 외부 마이크를 사용한 방법 외에도 실시간 스트리밍 미디어 소스에서 배경음악을 검출하는 use-case 모두 실험할 수 있도록 하였다. 녹음 변환한 음성 파일을 학습한 각각의 모델에 테스트해서 반환된 결과를 최적화하여[3] Fig.2와 같이 GUI화면상에 그래프와 결과를 출력한다.

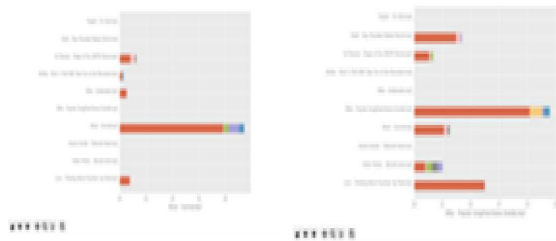


Fig. 2. Matplot based Classification GUI parts

2. Experimental results and evaluation

2.1 Experimental results

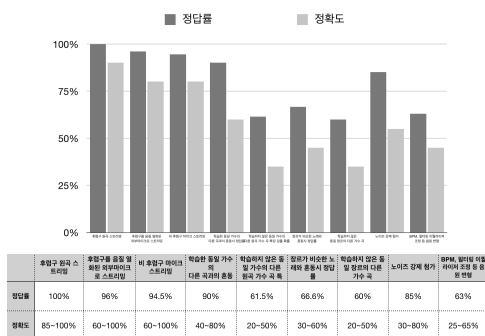


Fig. 3. Accuracy of sound source classification by scenario

기존의 Fingerprint 방식은 주파수의 특징을 데이터베이스에 대응 되는 패턴을 검색하는 방식이라 음원의 일치 여부를만 확인할 수 있었다. Fig.3의 결과는 본 시스템이 학습한 음악과 그 외에도 장르나 가수, 악기의 특색 등 음악의 특징들을 인식하는지 다양한 시나리오를 가정하여 청취 시에 연관성 있는 샘플을 분류한 정답률과 그 정확도를 나타낸 것이다.

2.2 Evaluation

원곡 청취 시 Fingerprint 방식과 비슷하게 정답률과 정확도는 월등하게 나타났으며, 본 시스템에서는 노래에 노이즈, 이펙트, 필터 또는 BPM을 조정하거나, 학습되지 않은 가수의 목소리 등 원곡에서 변형되어도 해당 음원의 특징을 판단하여 이를 비슷하다고 인식하거나 장르가 비슷한 음악들의 유사성을 인지하며 연관 있는 곡을 정답으로 표시하였다. 또한 여성 가수 10곡만을 학습한 데이터 셋의 실험 결과에서는 여성 목소리 특성들을 보다 중점적으로 특징하여 해당 특징을 더 잘 분류하였다. 즉, 학습 단계에서 초당 100개로 나뉜 샘플링 데이터들을 학습할 때 음원 고유의 특징들과 공통점을 찾아낸 것이다.

III. Conclusions

본 논문에서는 기존 내용기반의 음악 검색 어플리케이션보다 오차 범위가 유연하며 음악을 특정 구간을 기억하는 게 아닌 선율과 상관없이 그 속의 특징을 기억하기 때문에 짧게 인식하여도 정확도가 높으며 이는 기존 Fingerprint방식과는 다르게, 노래의 음색과 음계, 목소리 특색 등에 더 유연하게 대처하는 점을 확인하였다. 이러한 설계를 통해 향후 더 많은 곡들의 특징분류나 정확도를 개선하여 음원인식 분야와 응용 사운드 처리 분야 등에서 본 논문에서 제안한 인공지능 방식의 개선된 시스템의 방향성을 제시할 것으로 기대된다.

REFERENCES

[1] Google AI Blog. (2018). Looking to Listen: Audio-Visual Speech Separation. [online] Available at: <https://ai.googleblog.com/2018/04/looking-to-listen-audio-visual-speech.html?m=1> [Accessed 7 Jul. 2018].

[2] Kim, T. (2014). End-to-end learning for music audio. In: IEEE International Conference on Acoustics, Speech and Signal Processing.

[3] Hyun, H. Lee, K. (2016). Music segmentation using music source separation. In: The Korea Information and Communications Society.