

Journal of the Korea Institute of Information and Communication Engineering

한국정보통신학회논문지 Vol. 24, No. 7: 842~848, Jul. 2020

생성적 적대 신경망(GAN)을 이용한 딥러닝 음악 장르 분류 시스템 모델 개선

배준

Deep Learning Music Genre Classification System Model Improvement Using Generative Adversarial Networks (GAN)

Jun Bae

*Associate Professor, Department of Computer Science, The University of Suwon, Hwaseong 18323, Korea

요 약

아이튠즈, 스포티파이, 멜론 등 음악시장은 바야흐로 스트리밍의 시대로 접어들었고. 음악 소비자의 취향에 맞는음악 선곡과 제안을 위해 음악장르 자동 구분 시스템에 대한 요구와 연구가 활발하다. 이전 논문에서 제안한 소프트 맥스를 이용한 딥러닝 음악장르 자동구분 투표 시스템을 더욱 발전시켜 생성적 적대 신경망(GAN)을 이용하여 이전 시스템의 미흡한 점이었던 장르 미분류 곡들에 대한 정확도를 높이는 방법을 제안한다. 이전 연구에서는 전체 곡을 작은 샘플 로 나누고 각각의 샘플을 CNN 분석하여 그 결과들의 총합으로 장르 구분을 하는 투표 시스템으로 곡 장르분류 정확도를 높일 수 있었다. 하지만 곡의 스펙트로그램이 곡의 장르를 파악하기에 모호한 곡의 경우에는 미분류 곡으로 남겨놓을 수밖에 없었다. 이 논문에서는 생성적 적대 신경망을 이용하여 미분류 곡의 스펙트로그램을 판독하기 쉬운 장르의 스펙트로그램으로 바꾸어 미분류 곡의 장르 구분 정확도를 높이는 시스템을 제안하고 그 실험결과 기존 방식에 비해 우수한 결과를 도출해낼 수 있었다.

ABSTRACT

Music markets have entered the era of streaming. In order to select and propose music that suits the taste of music consumers, there is an active demand and research on an automatic music genre classification system. We propose a method to improve the accuracy of genre unclassified songs, which was a lack of the previous system, by using a generative adversarial network (GAN) to further develop the automatic voting system for deep learning music genre using Softmax proposed in the previous paper. In the previous study, if the spectrogram of the song was ambiguous to grasp the genre of the song, it was forced to leave it as an unclassified song. In this paper, we proposed a system that increases the accuracy of genre classification of unclassified songs by converting the spectrogram of unclassified songs into an easy-to-read spectrogram using GAN. And the result of the experiment was able to derive an excellent result compared to the existing method.

키워드: 생성적 적대 신경망, 딥러닝, CNN, 음악장르 자동구분

Keywords: GAN, Deep Learning, CNN, Automatic Music Genre Classification

Received 23 March 2020, Revised 13 April 2020, Accepted 21 April 2020

* Corresponding Author Jun Bae(E-mail:thebigjun@gmail.com, Tel:+82-31-229-8391)

Associate Professor, Department of Computer Science, The University of Suwon, Hwaseong 18323, Korea

Open Access http://doi.org/10.6109/jkiice.2020.24.7.842

print ISSN: 2234-4772 online ISSN: 2288-4165

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(http://creativecommons.org/li-censes/by-nc/3.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © The Korea Institute of Information and Communication Engineering.

I. 서 론

디지털 미디어의 폭발적인 성장과 수많은 스트리밍 서비스 속에서 음악 장르의 자동분류에 대한 필요는 점 점 증가하고 있는 추세이다[1]. 이전 논문에서는 기존 CNN, RNN 등을 이용한 연구를 발전시켜 소프트맥스 레이어를 이용한 투표 시스템으로 곡 장르분류의 확신 도를 높이는 모델을 만들어 실험 결과 기존 방식에 비해 우수한 결과를 도출해내었다[2].

기존 방식은 곡들의 일정 부분만을 딥러닝 분석하여음악 장르를 구분하였기 때문에 곡의 구성이 복잡한음악의 장르 구분에는 약한 면을 보였으나 소프트맥스 레이어를 이용한 투표 시스템은 곡의 모든 부분을 일정하게 나누어 각각의 딥러닝 결과를 투표하여 곡의 장르를결정하는 시스템으로 장르구분의 확신도를 높였다.

하지만 이 방법에서도 이미지가 장르를 구분하기에 모호한 곡들은 미분류로 남겨놓을 수밖에 없었다.

이 논문에서는 생성적 적대 신경망(GAN)을 이용하여 판단이 모호한 음악의 스펙트로그램 이미지를 스타일과 도메인 전이(Style and Domain transfer)를 통해 좀더 판단이 용이한 장르의 이미지로 변화시켜 장르 구분의 정확도를 높이는 방법을 연구하고 기본 방식과 정확도 비교를 통해 우수성을 입증한다.

생성적 적대 신경망을 이용하여 스타일 전이를 하는 연구, 즉 재즈를 클래식으로 바꾸는 것 등에 대한 연구 는 많이 있었다. 이를 역으로 이용해 음악 장르 구분 능 력을 향상시킬 수 있을까? 판단이 모호한 음악의 스펙 트로그램 이미지를 스타일 전이를 통해 좀 더 판단이 용 이한 장르의 이미지로 변화시켜 장르 구분의 정확도를 높이는 방법에 대해 논의한다.

Ⅱ. 관련연구 및 기존연구 문제점

기존의 음악 장르 구분에 대한 연구는 신호 처리 이론을 바탕으로 한 음악의 특성을 추출해내는 방식이 주를 이루었다. 이후 이렇게 추출되어진 특성을 머신러닝을 이용해 장르를 구분하는 방법이 연구되었다. SVM(Support Vector Machine) 방식이 주를 이루었고 KNN(K-Nearest Neighbors), GMM(Gaussian Mixture Model) 등 여러 방식이 이용되었다.

CNN(Convolutional Neural Network)은 1990년대 후 반 필기인식 연구에서 시작되었다. 인간의 시각처리를 구현하여 이미지 처리에 알맞은 모델이다. 최근 연구로 모델의 레이어를 늘려 정확도를 늘렸고 특히 영상인식 분야에서 뛰어난 성능을 발휘한다.

RNN(recurrent neural network)은 순차 데이터 처리에 이용되는 모델로 Deep neural network의 중간 레이어 값을 재귀해 사용해 데이터 특성을 순차적으로 추출하는데 뛰어난 성능을 보여준다. 현재 LSTM(Long ,Short Term Memory)와 GRU(Gated Recurrent Unit) 등의 모델이 발전되어 음성인식 및 자연어 처리에 이용되고 있다[3].

기존 딥러닝을 이용한 음악 장르 분류 학습 모델은 음악의 스펙트로그램을 이용한 CNN 방식과 음악의 순차적 시그널 데이터를 입력으로 하는 RNN 방식, 그리고 이두 방식을 결합한 멀티모달 방식이 있다[4]. 이 방식들은음악 장르를 구분하는데 있어서 변화가 적은 짧은 음악을 분류하는데는 효과적이나 하나의 곡 안에서 변화가 많은 곡을 분류하는데는 적합하지 않을 수 있다. 왜냐하면 한곡 안에서 클래식처럼 조용한 부분과 록처럼 강한부분이 같이 있다면 그 곡의 시작부분 30초 정도를 분석하는 것만으로는 장르 분류에 오류가 있을 수 있다.

이전 연구에서는 이러한 오류를 줄이기 위하여 전체 곡을 작은 샘플로 나누고 각각의 샘플을 CNN 분석하여 그 결과들의 총합으로 장르 구분을 하는 투표 시스템으 로 곡 장르 분류 정확도를 높일 수 있었다. 하지만 곡의 스펙트로그램이 곡의 장르를 파악하기에 모호한 곡의 경우에는 미분류곡으로 남겨놓을 수밖에 없었다.

Ⅲ. 딥러닝 음악 장르 분류 투표 시스템 모델

3.1. 소프트맥스를 이용한 딥러닝 음악장르 자동구분 투표 시스템

이전 논문에서 논의한 딥러닝 음악 장르 분류 투표 시스템은 우선 음원의 데이터량을 줄이기 위해 스테레오 (stereo)파일을 모노(mono)파일로 만든 후 이를 스펙트로그램으로 변환하여 PNG 파일로 만든다. 이를 여러 개의 일정한 크기의 스펙트로그램으로 분할하여 각각을 CNN 분석을 하고 그 결과별 투표를 하여 그 곡의 장르를 결정하는 방식이다. (그림 1)

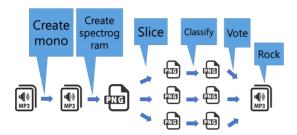


Fig. 1 Deep Learning Music genre automatic classification voting system flow chart (proposed model overview)

스펙트로그램을 일정한 길이로 잘라서 각 조각을 장르를 표시하는 개개의 샘플로 생각한다. 편의를 위해 128X128 픽셀의 정사각형으로 조각을 자르면 개개의 조각은 2.56초의 데이터를 가지게 된다.

모든 곡들을 정사각형의 스펙트로그램 이미지로 만든 후 각 음악 장르별로 수십만 개의 데이터셋을 가지게된다. 이제 Deep Convolutional Neural Network를 이용해 이 샘플들을 분류하도록 교육시키는데 Tensorflow's wrapper TF Learn을 사용한다[5].

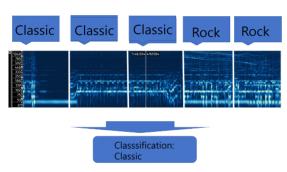


Fig. 2 Music Classification Voting System

6개의 장르 -하드코어, 록, 일렉트로, 클래식, 재즈 그리고 힙합-으로 나누어진 2,000개의 노래와 12,000개의 128X128 픽셀의 스펙트로그램 조각들을 이용한 결과이 모델은 90%의 정확도를 나타내었다. 이것은 노래의 작은 조각들을 사용한 것을 고려했을 때 상당히 좋은 결과라고 할 수 있다. 하지만 이 결과는 조각들의 분류에 대해 이야기하고 있는 것으로 전체 노래를 분류한 것이아니다. 딥러닝 음악장르 자동구분 투표 시스템그 곡의각의 샘플들이 장르에 대해 투표를 하고 가장 많은 표를 받은 장르를 그 곡의 장르로 선택하는 시스템이다.이로서 장르 예측의 정확도를 높일 수 있었다 (그림2).

이 시스템으로 새로운 곡의 장르 예측이 가능해졌지 만 여기서 투표 시스템을 더욱 발전시키기 위해 다음과 같은 수정을 가했다. 장르 구분 시스템의 마지막 레이어 에 소프트맥스 레이어를 추가해서 시스템이 장르를 지 정하기 보다는 그 가능성을 표시할 수 있도록 하였다. 이를 '분류 확신도'라고 명명한다. 이를 투표 시스템을 개선하는데 사용하였다. 예를 들면 낮은 확신도의 조각을 투표에서 제외하였다. 또한 투표에서 확실한 승자가 없으면 전체 투표 자체를 무효화하였다. 정확도 향상을 위하여 70% 미만의 투표를 받은 곡은 장르 구분에서 제 외시켜 장르 미분류곡으로 남겨두었다(그림 3).

이러한 미분류곡의 장르 분류 정확도를 개선하기 위해 생성적 적대 신경망(GAN)을 이용한 딥러닝 음악 장르 분류 시스템 모델을 제안한다.

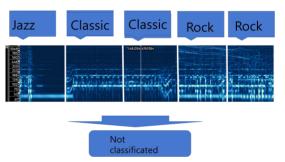


Fig. 3 Truncation of Music Classification

3.2. 생성적 적대 신경망(Generative Adversarial Networks)

생성적 적대 신경망(GAN)이란 두개의 네트워크로 구성된 심층 신경망 구조이다. 한 네트워크가 다른 네트워크와 겨루는 구조이기 때문에 적대적(Adversarial)이라는 의미가 이름에 포함되었다. 어떠한 종류의 데이터도 모방 학습이 가능하기 때문에 GAN의 잠재능력은 매우 크다. 음악, 이미지, 글 등 모든 분야에서 실제와 아주비슷한 새로운 결과물을 만들어 내도록 학습될 수 있다.

GAN을 이해하기 위해서는 생성(generative) 알고리 즘과 판별(discriminative) 알고리즘의 차이에 대해 알아야한다. 판별 알고리즘은 주로 입력데이터를 분류하는데 집중하고 클래스 간의 경계를 학습하는데 비해 생성알고리즘은 각 클래스의 분포를 모델링한다. 판별 알고리즘은 데이터 인스턴스의 피쳐가 주어지면 해당 데이터가 속한 레이블 또는 카테고리를 예측하게 된다. 즉 피쳐와 레이블 간의 상관관계에만 집중한다. 반면 생성

알고리즘은 이와 반대로 주어진 특정 피쳐로부터 레이블을 예측하는 대신 주어진 특정 레이블에서 피쳐를 예측하는 방식이다[6].

GAN은 생성기(generator)와 판별기(discriminator) 두 개의 신경망으로 구성되어 있다. 생성기 신경망은 새로운 데이터 인스턴스를 생성하고 판별기 신경망은 데이터의 참과 거짓을 평가한다. 즉 판별기는 검토한 각각의 데이터 인스턴스가 실제 트레이닝 데이터 세트인지아닌지를 판단한다(그림4). 예로서 MNIST 데이터 세트와 같은 손으로 쓴 숫자를 생성한다고 가정할 때 그것이 진짜인지 판단하는 것이 목표이다.

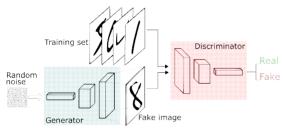


Fig. 4 Generative Adversarial Network framework [7]

한편 생성기는 판별기에게 입력할 새로운 이미지를 만들어낸다. 생성기는 만들어낸 가짜 이미지가 진짜 이 미지인 것처럼 보여지길 원한다. 여기에서 생성기의 목 적은 판별기가 진짜로 판단하게 만드는 손글씨 숫자를 만드는 것이다. 반대로 판별기의 목적은 생성기가 만든 가짜 이미지를 찾아내는 것이다.

GAN의 동작단계는 다음과 같다.

- 생성기가 입력된 임의의 수를 손글씨 숫자 이미지로 변환한다.
- 생성기가 만든 이미지를 실제 데이터 세트의 이미지들 과 함께 판별기에 전달한다.
- 판별기가 실제 이미지와 가짜 이미지를 구별하여 0과 1 사이의 확률값으로 변환한다. 1에 가까울수록 실제 이미지, 0은 가짜 이미지를 뜻한다.

MNIST의 경우 판별기망은 그에 공급되는 이미지를 분류할 수 있는 표준 컨볼루션망으로, 이미지를 실제 또 는 가짜로 표시하는 이항 분류기이다. 생성기는 어떤 의 미에서 역 컨볼루션망이다. 표준 컨볼루션 분류기가 이 미지를 가져와서 확률을 예측하기 위해 다운 샘플링을 하는 반면 생성기는 랜덤 노이즈 벡터를 가져와서 이미 지로 만드는 업 샘플링을 한다. 일반적인 컨볼루션 네트 워크는 최대 풀링(maxpooling)과 같은 다운 샘플링 기술을 통해 데이터를 처리하고 생성기와 같은 역 컨볼루션 네트워크는 새로운 데이터를 생성한다.

두 네트워크 모두 제로섬 게임처럼 서로 반대되는 목 적함수 또는 손실함수를 통해 최적화하려고 시도한다. 이것은 actor-critic model과 비슷하다. 판별기의 행동이 바뀌면 생성기의 행동도 변하고 그 반대의 경우도 마찬 가지이다. 각 에러는 서로에게 전달된다[8].

3.3. GAN을 이용한 스타일과 도메인 전이

(Style and Domain transfer)

신경망을 이용한 스타일과 도메인 전이는 머신러닝에서 매우 흥미진진한 분야이다. 대부분의 이전 작업은이미지 영역에 초점을 맞추었으며, 예를 들어 사진을 찍어 특정 화가의 스타일로 렌더링하거나 [9] 여름 동안촬영한 이미지를 겨울에 찍은 것처럼 변경할 수 있었다[10]. 도메인 전이가 흥미로운 건 그것들은 딥러닝 연구의 다른 분야로 옮겨갈 새로운 학습 기법의 개발을 요구하기 때문이다.

도메인 전이가 작동하기 위해서는 신경망 모델이 기본 도메인에 대한 깊은 이해를 가지고 있어야 한다. 이를 위해서는 이미지, 자연 언어 또는 음악과 같은 복잡한 데이터에서 중요한 기능을 추출해야 한다. 가변 자동코드[11](VAE) 및 생성적 적대 신경망과 같은 심층 생성 모델은 진정한 기초 데이터 분포를 학습하려고 하기때문에 이 작업에 매우 적합해 보인다. 따라서, 심층 생성 모델을 이용한 신경 스타일 전이는 심층 표현 학습연구[12]에서 매우 관련성이 높은 부분이다.

음악을 위한 도메인 전이는 실제로 가능한 많은 활용성을 가지고 있다. 예를 들어 음악가들은 종종 기존곡에 새로운 해석을 하여 리메이크곡을 만든다. 원곡과 완전히 다른 스타일의 리메이크곡을 만들기 위해서는 많은음악적 지식과 노력이 필요하다. 도메인 전이 시스템은이 과정을 상당히 가속화하거나 심지어 완전히 자동화할 수 있다.

신경망의 맥락에서 스타일 전이는 일반적으로 이미지의 명시적인 내용의 특징을 유보하고 다른 이미지의 명시적 스타일 특징을 적용하는 것을 말한다. 명시적인 스타일과 콘텐츠의 특징은 사전 훈련된 CNN에서 추출 된 것이다.

도메인 전이는 이미지 같은 전체 도메인 간의 매핑을

학습하는 것을 목표로 한다. 예를 들어, 도메인 전이는 도메인 A로부터 어떤 입력을 받아내어 그것이 마치 도 메인 B에 속하는 것처럼 보이도록 그것을 변경할 수 있 게 한다. 여기서 A와 B는 여름과 겨울, 또는 재즈와 클 래식이 될 수 있다.

이 논문에는 이러한 도메인 전이를 이용하여 미분류 곡의 스펙트로그램을 판독하기 쉬운 장르의 스펙트로 그램으로 바꾸어 미분류곡의 장르 구분 정확도를 높이 는 시스템을 제안한다.

3.4. 생성적 적대 신경망(GAN)을 이용한 딥러닝 음악 장르 분류 시스템

이 제안모델에서는 공유 생성기 G를 설정해 사용한다. 장르를 분류할 음악들에 모두 공통적으로 적용되고 점진적으로 학습된다. 점진적 네트워크(Progressive network)는 그림5의 첫 두 레이어로 시작한다. 두 개의 에포크마다 하나의 레이어가 모델에 추가되고 이전 반복의 변수가 이전 모델에서 넘어온 레이어에 입력된다. 새로운 계층에 대한 변수는 무작위로 초기화되고 정규화된다. 제1 계층은 제 2 계층이 제 1 계층으로부터 통과한 텐서를 재형성할 때 완전히 연결된 층이다. 생성기의 모든 후속레이어는 3D 컨볼루션 레이어로 바뀐다.

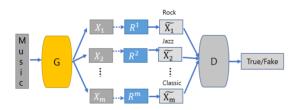


Fig. 5 Proposed Architecture

그림5는 제안된 아키텍처의 다이어그램으로서 공유 생성기 네트워크 (G) 및 공유 판별기 네트워크 (D)는 사이에 가변 텐서 (tensor) 크기를 갖는 리파이너 네트워크 (refiner network)로 점진적으로 훈련된다.



Fig. 6 Refiner Network

리파이너 네트워크(그림6)는 전체 네트워크와 결정 론적 이진법 뉴런(deterministic binary neurons)과 동일 한 크기의 텐서를 가진다. 리파이너 네트워크에서 사용 되는 잔여 단위 값은 두 컨볼루션 레이어의 커널 크기와 출력 채널 수를 나타낸다.

3.4.1. 리파이너 (Refiner)

리파이너 네트워크는 R 생성기 네트워크의 제 2 부분이며, 각각 Ri, $i=1,\ldots,M$ 으로 표시되는 M 개의 개별네트워크로 구성된다. 각 트랙마다 하나씩 (음악장르)리파이너 네트워크는 생성기에서 입력을 받아 트랙 축을 따라 분리한다. 리파이너의 작업은 생성기 $\hat{X}=G(z)$ 의 실제 출력을 \tilde{X} 로 표시되는 이진 값으로 세분화하는 것이다. 리파이너 출력은 트랙 축을 따라 모든 트랙에서 병합된 출력으로 구성된다.

3.4.2. 판별기(Discriminator)

판별기는 모든 훈련단계에 대한 생성기의 복제품이다. 판별기는 생성기 또는 원래 데이터 세트에서 생성된 출력을 입력으로 받는다. 두 번째 마지막 레이어와마지막 레이어에서 출력이 하나만 있는 3D 컨볼루션 레이어로 구성된다. 원본 데이터가 판별기에 공급될 때 실제데이터와가짜 데이터를 구별하도록 학습하고 생성기의 출력이 공급될 때 제공된 샘플이 실제인지가짜인지 (true 또는 false) 식별한다.

Ⅳ. 실험결과

4.1. 모델 적용 및 학습

이 실험에서는 6개의 각 장르별 -하드코어, 록, 일렉트로, 클래식, 재즈 그리고 힙합- 500개의 총 3,000개의 노래를 스펙트로그램으로 변환하고 128X128 픽셀로분할하여 34,200개의 스펙트로그램 조각들로 만들었다. 훈련의 마지막 단계(모든 계증이 추가될 때)에서 판별기의 입력 텐서와 생성기의 출력 텐서의 크기는 128×128 픽셀이다.

훈련의 마지막 단계에서 생성된 모델이 그림 5에 나와 있다. 입력 랜덤 벡터의 길이는 128이다. 사용하는 최적화는 Adam 최적화이며 생성기와 정제기를 정규화한다. 우리는 이진 뉴런을 가진 네트워크에 슬로프 어닐링 (slope annealing)[13]을 구현하는데, 여기서 시그모이드 조정된 직선 추정기(straight-through estimator)에서 시

그모이드 함수의 기울기에 1.1이 곱해진다.

각 에포크(epoch)는 32의 배치 크기를 사용하므로 에 포크 당 429회 반복된다. 모든 단계는 2개의 에포크에 대해 학습된다. 즉, 각 계층을 추가한 후 네트워크는 2개 의 에포크에 대해 학습된다.

4.2 실험결과

위와 같이 학습된 모델과 기존 모델들을 사용해 장르 분류를 실시하고 각각의 모델에서 장르 미분류곡의 비 율을 표와 같이 나타내었다(표1).

Table. 1 Classification Error Rate Comparison

| Model | Classification Error Rate (%) |
|-------------------------------------|-------------------------------|
| CNN | 16.3 |
| RNN | 19.2 |
| CNN+RNN | 13.4 |
| VOTING SYSTEM | 11.7 |
| VOTING SYSTEM+GAN (suggested model) | 8.1 |

장르 미분류 곡의 비율이 RNN 기법에 비해 CNN을 적용했을 때 적게 나왔으며 CNN과 RNN을 같이 적용 했을 때 더 적은 비율로 나타났다. 음악샘플을 나누어 각각의 CNN 결과를 투표하는 시스템이 이전 모델에 비 해 좋은 결과를 나타내었고 이 모델에 GAN을 적용한 것이 미분류 된 곡이 가장 적게 나오는 결과를 보였다.

Ⅴ. 결 론

아이튠즈, 스포티파이, 멜론 등 거대 스트리밍 사이트의 폭발적인 성장으로 음악 시장은 바야흐로 스트리밍의 시대로 접어들었다. 음악 소비자의 취향에 맞는 음악 선곡과 제안을 위해 음악장르 자동 분류 시스템에 대한 요구와 연구가 활발한 지금 이전 논문에서 제안한 소프트맥스를 이용한 딥러닝 음악 장르 자동구분 투표 시스템을 더욱 발전시켜 생성적 적대 신경망(GAN)을 이용하여 이전 시스템의 미흡한 점이었던 장르 미분류곡들에 대한 정확도를 높이는 방법을 제안하고 실험결과기존 방식에 비해 우수한 결과를 도출해낼 수 있었다. 이러한 방식을 이용하여 음악소비자가 좋아하고 원하

는 장르의 음악을 추천해주는데 더욱 정확한 큐레이션 이 가능해질 수 있을 것이다.

하지만 아직도 인간의 음악 장르 구분 능력에 비해 딥러닝 방식의 곡 장르 구분은 미흡한 점이 많은 것이 사실이다. 이 논문의 생성적 적대 신경망을 이용한 시스템은 미분류곡의 비율을 줄일 수 있었지만, 아직 유사한스펙트럼을 가진 곡의 장르분류는 완벽하지 않은 한계점이 있다. 향후 생성적 적대 신경망 시스템의 정확도를 개선하는 연구가 필요할 것으로 생각된다.

REFERENCES

- [1] S. Kim, D. Kim, and B. Suh, "Music Genre Classification using Multimodal Deep Learning," *International Journal of Information and Communication Engineering*, vol. 9, no. 4, pp.358-362, Aug. 2011.
- [2] J. Bae, J. Kim, "Deep Learnig Music Genre Automatic Classification voting System using Softmax" *Journal of the Korea Institute of formation and Communication Engineering*, vol. 23, no. 1, pp. 27-32, Jan.2019.
- [3] B.Macfee, "Learning Content Similarity for Music Recommendation," *Journal of latex class files*, vol. 6, no. 1, pp.1-2, Jan. 2017.
- [4] L. Maaten, G. Hinton, "Learning Content Similarity for Music Recommendation Visualizing Data using T-SNE," *Journal of Machine Learning Research*, vol. 9, no. 1, pp.2579-2605, Nov. 2008.
- [5] T.Yuan. "TF. Learn: TensorFlow's high-level module for distributed machine learning." arXiv preprint arXiv, pp.25-32, vol.1612, no.1.Jan.2016.
- [6] L.Yu, O.Tuzel. "Coupled generative adversarial networks." Advances in neural information processing systems. vol.3, no.2, pp.469-477, May. 2016.
- [7] B.David, T.Schumm, and L.Metz. "Began: Boundary equilibrium generative adversarial networks." *arXiv preprint arXiv*, vol.170, no.10717, pp 32-40, Nov. 2017.
- [8] T.Silva, Own work, CC BY-SA 4.0 [Internet]. Available: https://www.freecodecamp.org/news/an-intuitive-introducti on-to-generative-adversarial-networks-gans-7a2264a81394/
- [9] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *Computer Vision and Pattern Recognition (CVPR)*, IEEE Conference on. IEEE, vol.2, no.3, pp. 2414-2423, May. 2016.
- [10] J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired

- image-to-image translation using cycle-consistent adversarial networks," in *IEEE International Conference on Computer Vision, ICCV* vol. 35, no. 8, pp. 2242 2251, Jun. 2017.
- [11] D. P. Kingma, M. Welling, "Auto-encoding variational bayes," *CoRR*, vol 4, no.3, pp. 24 27, Sep. 2016.
- [12] Y. Bengio, A. C. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach.Intell.*, vol. 35, no. 8, pp. 1798 1828, Jan.2013.
- [13] J. Chung, S.Ahn, and Y. Bengio, "Hierarchical multiscale recurrent neural networks". *In Proc. ICLR*, vol. 24, no. 1, pp. 47-54, Mar. 2017.



배준(June Bae)

연세대학교 정치외교학과 졸업 상명대학교 컴퓨터음악대학원 졸업 수원대학교 음악테크놀로지과 교수 수원대학교 컴퓨터학부 박사과정

※관심분야: 전기차 사운드 디자인, AI 알고리즘 작곡, 머신러닝 플레이리스트 작성, 음성인식, DSP 설계