

CNN, 코사인 유사도를 이용한 장르 예측 기반 음악추천시스템

2022.07.08. KUBIG 컨퍼런스

뮤직~ 큐! (Music KU) 팀
박상준, 박지우, 이승은, 김지호

목차

- 1 주제 설명
- 2 Dataset
- 3 Model Description
- 4 Result
- 5 결론 및 제언

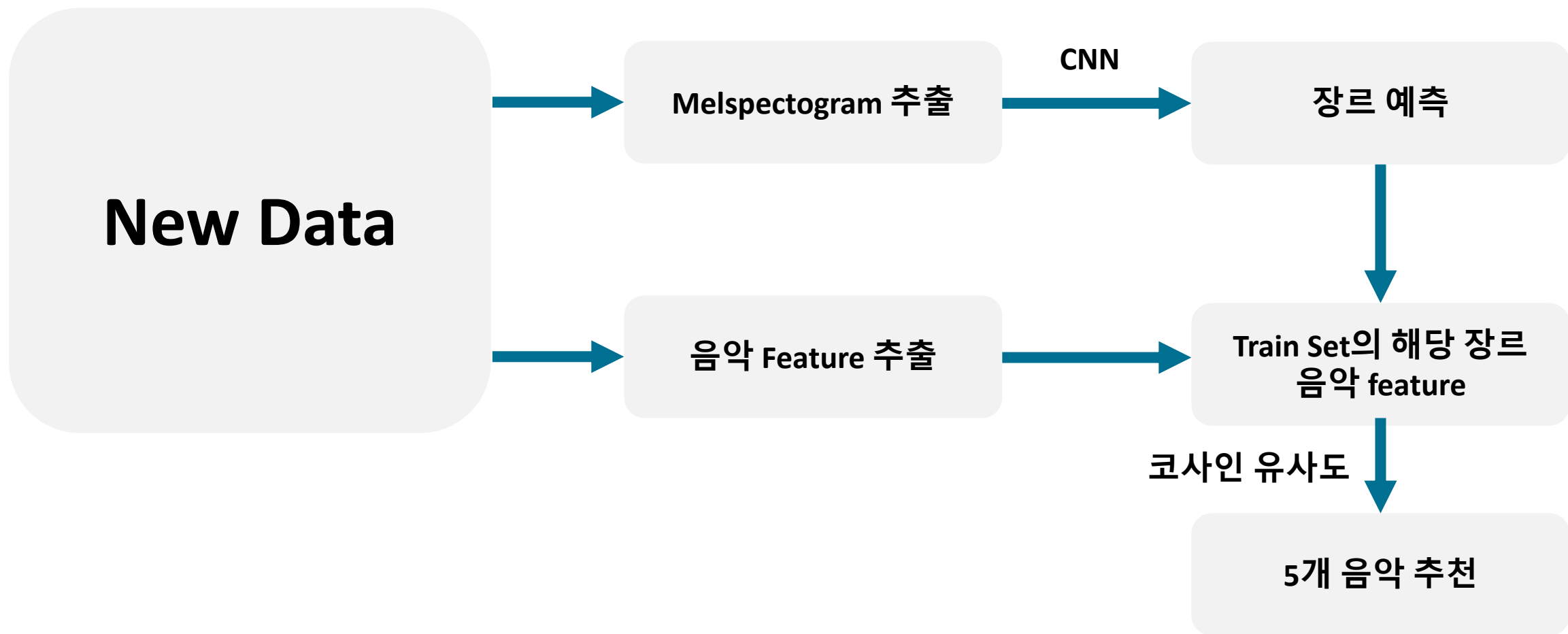




Part 1

주제 설명

1 주제 설명



장르 분류 후, 분류한 장르 내의 음악 feature 값으로
cosine similarity 계산 후 5개의 음악 추천



Part 2

Dataset





Dataset

수집 방식

: 유튜브 음악 추출 (배포하지 않도록, 모델 train/test를 위해서만 사용)

Hold-Out

: fixed test dataset for all models (앞서 소개한 dataset 800개는 train, validation dataset으로 사용 / test dataset은 각 장르별 30개, 총 120개를 더 수집하여 사용)

Stratified K-Fold

: 5-fold 사용



Part 3

Model Description



Step 0

Mel-spectrogram Extraction

def. Mel-spectrogram

감상자가 인지할 수 있는 피치의 크기로 주파수 단위를 변환한 스펙토그램

```
S = librosa.feature.melspectrogram(y, sr=sr)
S_DB = librosa.amplitude_to_db(S, ref=np.max)
```

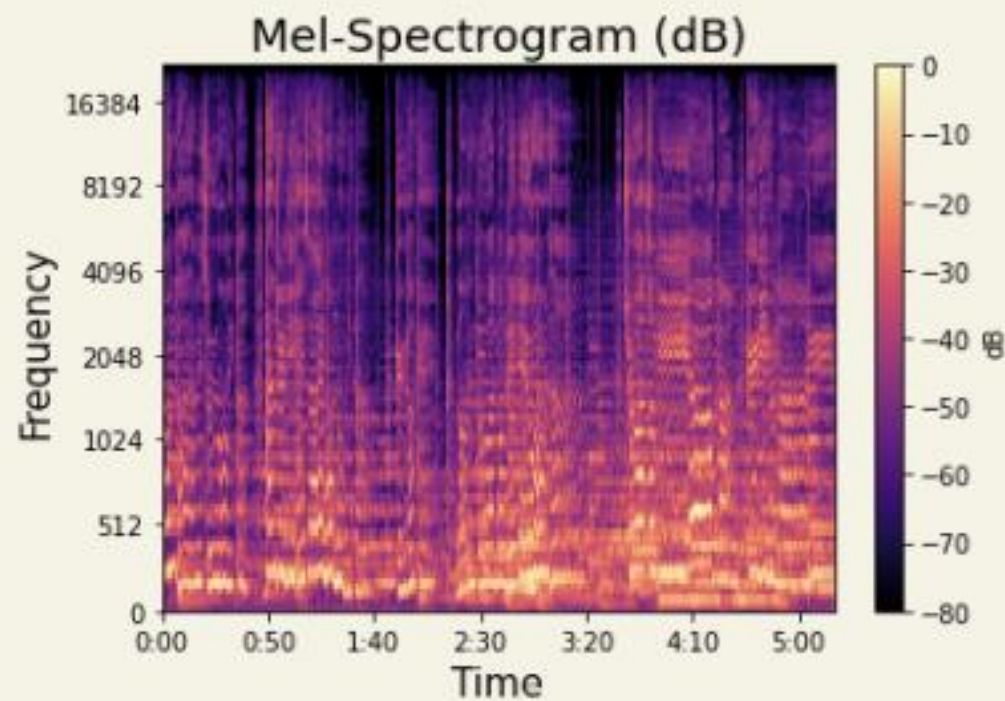
- y: 파형의 진폭 [amplitude] 값
- sr [sampling rate]: 초당 몇 개의 샘플을 가지고 있는지

(16000Hz로 설정, 사람의 목소리 모두 포함 가능)

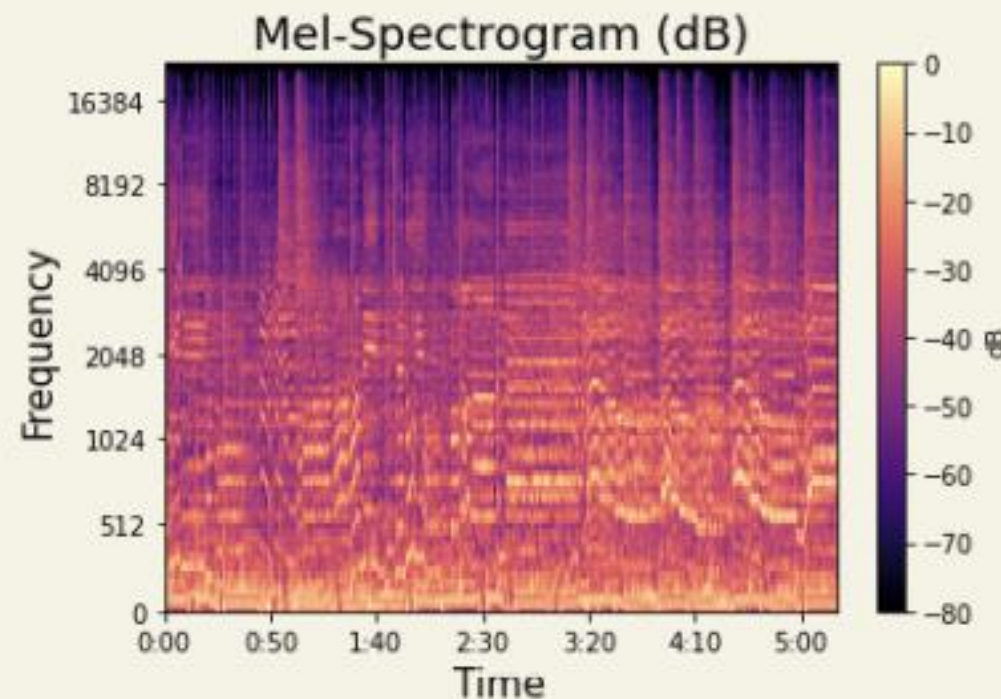
3

Step 0

Mel-spectrogram Visualization



〈박효신-야생화의 Melspectrogram〉



〈혁오-TOMBOY의 Melspectrogram〉

Step 1

Genre Classification based on Mel-spectrogram

1) CNN

Resnet50, Xception, Mobilenet (V1)

2) Attention-based computer vision models

Vision Transformer (Base/16 (384), Large/32 (384))

3) RNN (Sequence)

RNN, LSTM, GRU

Step 2

Feature Extraction

- 1) 음악 (.wav) 파일 60 – 90 sec (killing part 구간 자르기)
- 2) 자른 음악 파일 (.npy)로 feature extraction

Step 2

Feature View

Zero crossing rate	rate of sign changes along a signal
Zero crossing	positive에서 negative or negative에서 positive 변화 지점의 합
Spectrogram	시간축과 주파수 축의 변화에 따라 진폭의 차이 반영
Harmonics	고조파로 기본 주파수의 한 파동
Shock Wave	액체, 기체, 등의 유체 속을 음속보다 빠르게 지나갈 때의 파동
Chromagram	Pitch Class와 관련됨. 음악 작품 간의 유사성 측정을 설명하는 강력한 방법 제공
Spectral Centroid	주파수의 스펙트럼의 가중 평균
Spectral Rolloff	신호의 모양을 측정. 고주파수가 0으로 감소하는 주 파수
Spectral Bandwidth	주파수의 범위/폭 (최고 주파수와 최저 주파수의 차 이)
BPM	분당 빠르기 (Beats Per Music)

3

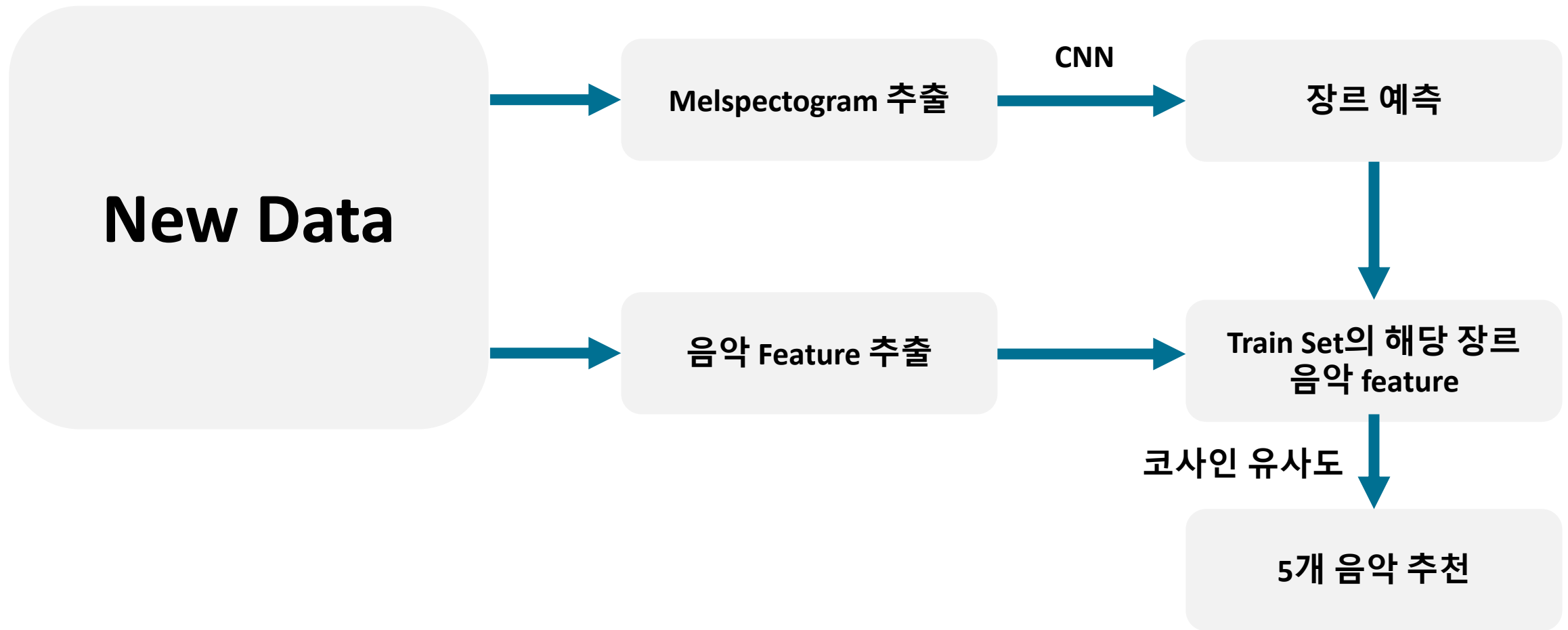
Step 3

Cosine Similarity

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

```
def cos_sim(a, b):  
    return dot(a, b)/(norm(a)*norm(b))
```


3 Step 4- Wrap Up



장르 분류 후, 분류한 장르 내의 음악 feature 값으로
cosine similarity 계산 후 5개의 음악 추천



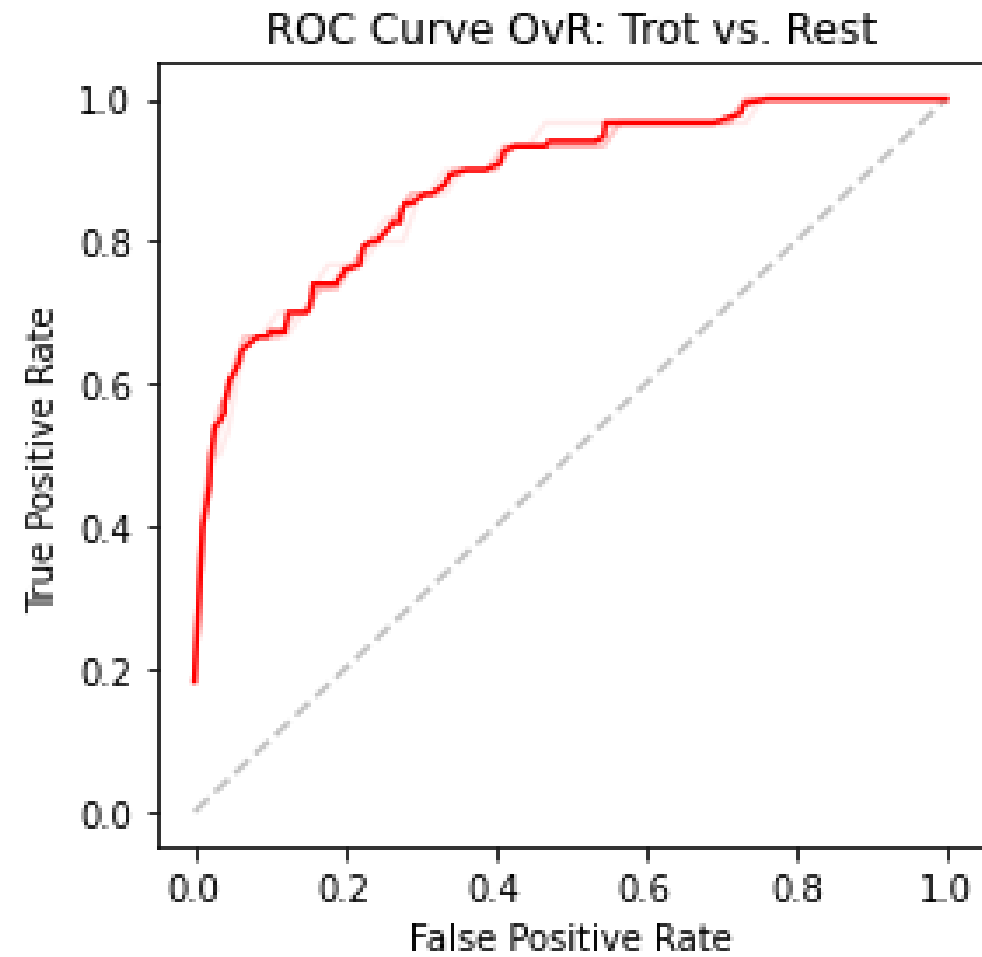
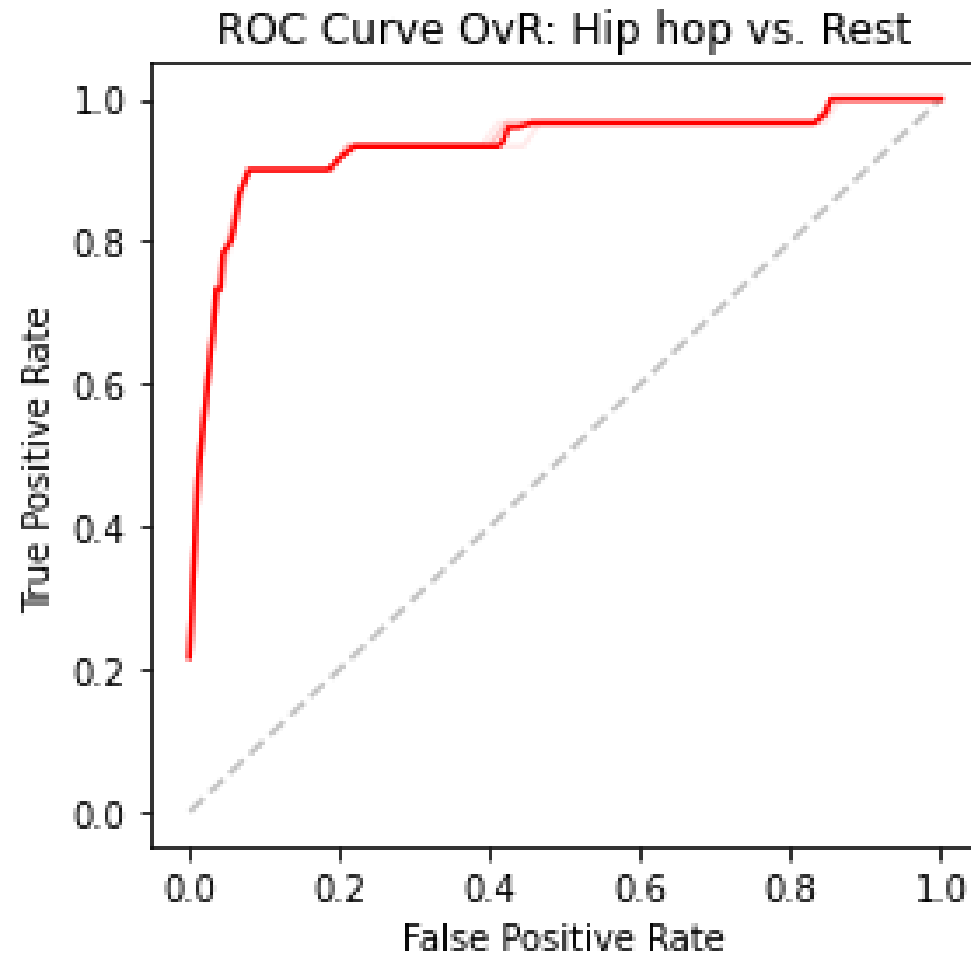
Part 4

Result

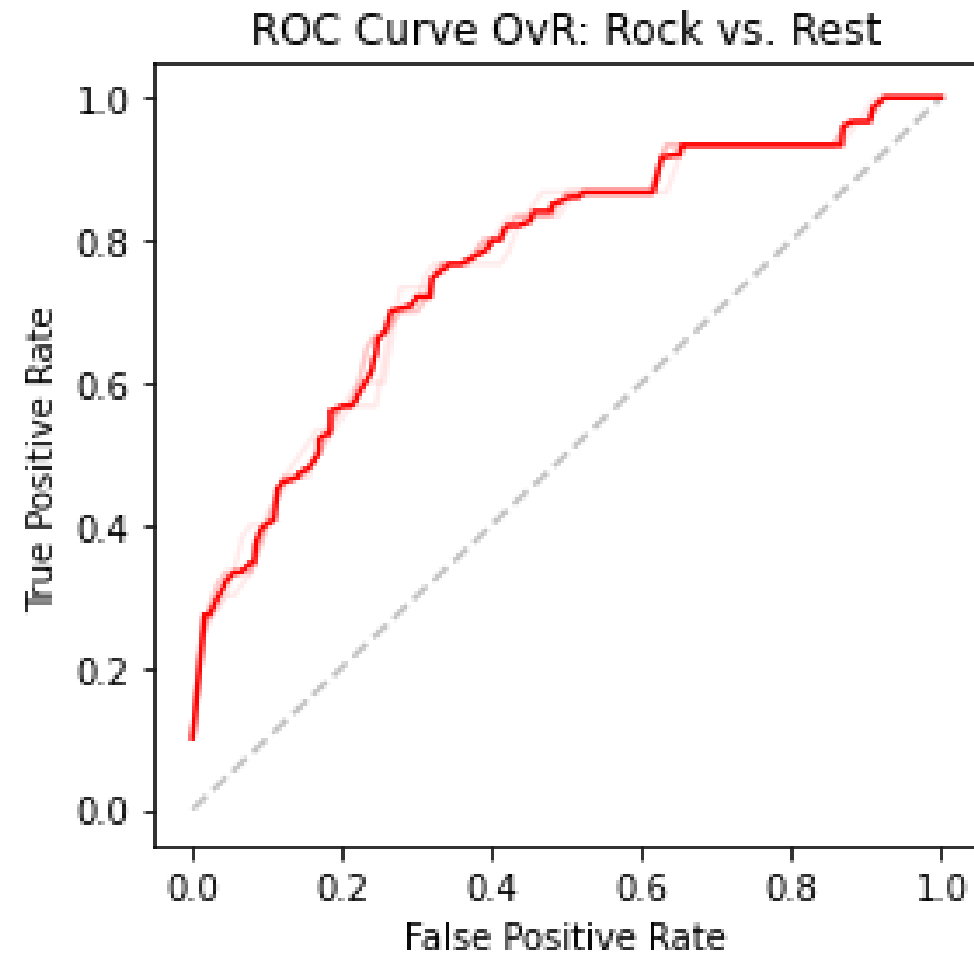
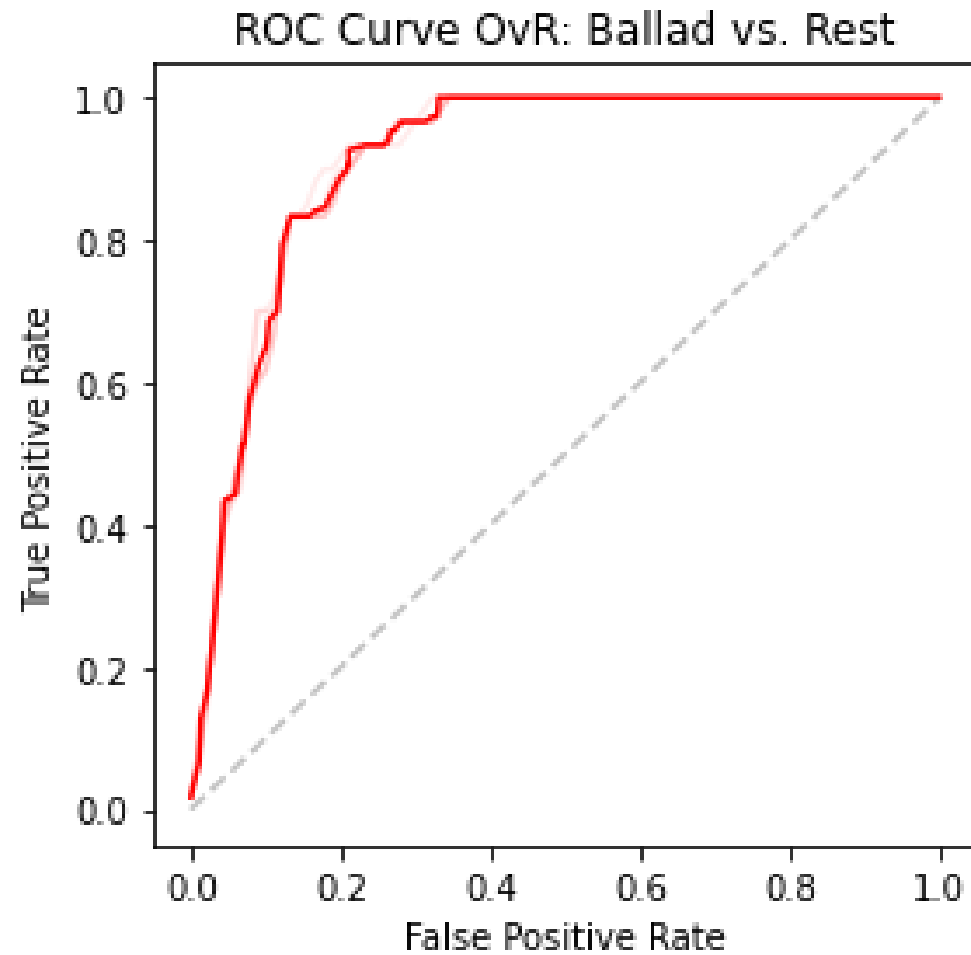


CNN Models	Classification Accuracy (5-fold, mean)	standard deviation (5-Fold)
Resnet50	69.17	0.008
Xception	70.83	0.003
Mobilenet (V1)	69.17	0.011

4 ROC Curve – Xception



4 ROC Curve – Xception



4 ROC Curve – Xception

AUC	0.8748
Sensitivity	0.9000
Specificity	0.7100

4 for research

1) Attention-based computer vision models

Vision Transformer (Base/16 (384), Large/32 (384))

-> classification accuracy: 25.50%

* (ViT) mel-spectrogram을 patch 단위로 잘라서 분석하는 것이 효과적이지 않음.

4 for research

2) RNN, LSTM, GRU

-> classification accuracy: 18.00%

pre-trained weight가 없고 random weight initialization을 해서 성능이 나빴던 것은 아닌지 생각

-> curriculum learning 시도 (mel-spectrogram을 patch 단위로 잘라서 학습 후 weight 생성 (3배))

-> 원본 mel-spectrogram에 transfer learning 적용) -> 성능 개선 X

(3배 이상으로 진행해보고 싶었으나 Google Colaboratory 환경에서 RAM 부족으로 실행 불가)

4 트로트 추천 예시

```
Reccomendation_system('414')  
#414: 임영웅 그대 나만 믿어요
```

```
Converting 414 .mp3 to wav...  
Trimming 414 ...  
Extracting Features of 414 ...  
Reccomending Top 5
```

	title	artist	release_date	genre
468	참 좋은 날	이찬원	2021.12.09	트로트
524	코로나19 이겨냅시다	태진아	2020.03.24	트로트
488	살아생전에	홍자	2019.06.07	트로트
460	물망초	정동원	2021.11.17	트로트
404	오라버니	금잔디	2013.03.01	트로트

그대 나만 믿어요 (임영웅)

트로트 장르 분류 후

(5-Fold / 최빈값으로 장르 분류)

트로트 장르 내의 나머지 곡들 중

현재 곡과 코사인 유사도가 가장 높은

5개를 선정하여 추천



Part 5

결론 및 제언

“CNN, 코사인 유사도를 이용한 장르 예측 기반 음악추천시스템”

- 한계
 - 최종적인 평가 기준은 음악을 추천 받은 개인의 주관적 선호도 장르 예측, 분류로 “평가”될 수 있는 모델을 만들어보았지만 한계 존재
 - 음악 추출 개수 한계. 보유한 데이터셋 내에서만 음악 추천 가능

4 결론 및 제언

- 후속 연구

Supervised Learning: class 세분화

(락 분류 어려움 – 발라드락, 메탈락, 헤비락, ...)

Unsupervised Learning:

장르 "분류" 보다는 음악 feature로 유사한 음악들을 clustering

*

Reference

- <https://kaen2891.tistory.com/m/39> [Mel-spectrogram]
- <https://gist.github.com/gvyshnya/ee61c83111ae043f7a49b8829ba994bc> [Music Feature]



you are ♪
What you listen to

♪ 감사합니다 ♪