오케스트라 연주에서 딥러닝 기반 악기 식별

정다현○, 박나은, 고재휘, 황혜수

서울시립대학교 컴퓨터과학부

dahyun21@uos.ac.kr, parkne0114@uos.ac.kr, kojaehwi0616@uos.ac.kr, hwang@uos.ac.kr

Deep Learning-Based Instrument Identification in Orchestra Performances

Dahyun Jeong○, Naeun Park, Jaehwi Ko, Heasoo Hwang

Department of Computer Science and Engineering, University of Seoul

요 약

본 연구는 오케스트라 음원에서 특정 시점에 동시에 연주되는 최대 16개 악기들을 식별하는 것을 목적으로 한다. 1개에서 16개 사이의 악기 소리가 포함된 4초 단위 음원을 입력 받는 CNN 기반 악기식별기 16개의 결과를 종합하여, 특정 시점에 연주 중인 악기를 식별한다. 제안한 방법은 연주 중인 악기들을 86.82%의 정확도로 예측했다. 또한 제안 방법의 실용성 평가를 위해 사용자 설문조사를 진행했다.

1. 서론

오케스트라 공연은 다양한 악기들이 어우러져 다채로운 음악을 창조한다. 이러한 음악적 복합성은 개별 악기의 소리를 식별하고 분석하는데 어려움을 준다. 최근 딥러닝 기술은 이미지 인식, 음성 인식 등 많은 분야에서 발전을 이끌어왔으며, 악기의 식별과 분류에서도 효과적이다. 악기의 식별은 소리를 듣고 특정 악기의 존재 여부를 판단하는 작업이며, 악기의 분류는 소리를 듣고 특정 악기로 구분하는 작업이다. 최근에는 합성곱 신경망(Convolution Neural Network, CNN)과 Mel-Frequency Cepstral Coefficient(MFCC)를 사용한 악기 식별 및 분류 연구가 활발히 진행 중이다[1-4].

[1]에서 제안한 방법은 악기의 소리를 듣고 해당하는 악기군으로 분류는 가능하나, 개별 악기의 식별은 진행하지 않는다. [2]에서 제안한 방법은 피아노, 기타, 드럼, 베이스 4 개의 악기에 대해서만 식별 가능한 모델을 구성했다. [3]에서 제안한 방법은 피아노를 포함한 5 개의 오케스트라 악기 소리를 특징에 따라 분류한 뒤, 악기를 식별하는 모델을 구성했다. [4]에서 제안한 모델은 다성음악에서는 악기 식별이 불가능한 모델로, 다성음악인 오케스트라에 적용이 불가하다. 공통적으로 이러한 선행 연구들은 주로 수치적 연구 결과만 제시했다. 또한 기술적 응용 및 산업적 활용을 위해서는 사용자 평가가 요구되나 실제 사용자들의 사용성 평가에 대한 구체적인 방법 및 결과를 제시하지 않았다. 본 연구는 이러한 한계를 확장된 악기 수와 사용자 평가 방법 제시 및 실행으로 극복하고자 한다.

본 연구의 목적은 CNN 기반 모델을 활용하여 오케스트라 음원에서 특정 시점에 어떤 악기가 존재하는지 식별하고 정확도 평가와 사용자 평가를 진행하는 것이다. 총 16 개의 서로 다른 악기의 존재 여부를 검출하는 악기식별기로 이루어져 있으며, 각 악기식별기의 결과를 종합하여 특정 시점에 연주 중인 모든 악기를 도출한다. 각 악기식별기는 정확도와 f1 score 지표를 활용해 평가하고, 10 명의 음악인과 20 명의 비음악인을 대상으로 설문조사를 진행해 사용자 평가를 진행했다. 제안한 방법은 오케스트라 음악에 대한 이해도를 높일 수 있으며 음악 인식, 음악 제작, 그리고 음악 정보 검색에 영향을 미칠 것으로 기대된다.

2. 제안 방법

2.1 데이터셋

단일 악기의 소리로 구성된 데이터들을 혼합하여, 여러 악기의 소리로 구성된 오케스트라 음원 데이터셋을 생성했다.

런던 필하모닉 관현악단에서 제공하는 오케스트라 악기들의 코드(Chord)별 사운드 샘플 데이터셋과 AIR Lab 에서 제공하는 오케스트라 악기들의 곡 별 연주 음원 URMP 데이터셋을 사용했다. 런던 필하모닉 데이터셋의 경우 URMP 데이터셋에 포함된 14 개의 모든 악기뿐만 아니라 이에 포함되지 않은 여러 타악기 데이터를 포함한다. 본 연구에서는 다양한 오케스트라 악기 식별을 위하여 두 가지 데이터셋에 모두 포함된 14 개의 악기와, 런던 필하모닉 데이터셋에 포함된 심벌즈, 탬버린을 포함하여 총 16 개의 악기를 선정했다.

표 1. 악기 별 데이터셋 수

악기	개수	악기	개수
바순	769	오보에	669
베이스 클라리넷	944	색소폰	837
첼로	1,038	탬버린	11
클라리넷	1,017	트롬본	948
크래쉬 심벌	7	트럼펫	741
콘트라 베이스	914	튜바	1,061
플루트	1,150	비올라	1,229
프렌치 호른	738	바이올린	1,904

단일 악기 데이터들을 모두 4 초 단위로 커팅(Cutting)했다. 4 초가 되지 않는 경우 패딩(Padding)을 통해 모든 데이터의 길이를 통일했다. 데이터를 4 초 단위로 맞춘 이유는 악기 특성을 반영하고 학습 효율을 위해서이다.

n 개(n=1~16)의 악기를 중복되지 않도록 선택했다. 선택된 악기 각각의 데이터셋에서 무작위로 하나의 데이터를 선택하고 그것들을 겹쳐 하나의 4 초짜리 오케스트라 음원 데이터로 만들었다. n 을 1 과 16 사이로 설정한 이유는 오케스트라의 특성상, 한 가지의 악기가 독주하는 상황과, 여러 악기가 동시에 연주되는 상황을 고려했기 때문이다.

이러한 방법을 통해 여러 악기들의 조합으로 4 초짜리 오케스트라 음원 데이터를 총 231,980 개(15GB) 만들었다. 또한 악기 개수 n 에 대해 훈련, 검증, 그리고 평가 데이터셋을 8:1:1 비율로 나누었다. 혼합된 악기의 개수에 따른 전체 데이터셋 분포는 표 2 에 나타나 있다.

만들어진 약 23 만 개의 데이터들을 MFCC 로 변환했다. MFCC 란 음성 데이터를 특징 벡터로 변환한 것이다. MFCC 를 통해 음성신호에서 유용한 정보를 추출할 수 있어 음성 및 오디오 신호처리 연구에서 사용되고 있다. 음원 데이터를 스펙트로그램으로 변환하고, 스펙트로그램의 주파수 척도를 Mel-Scale 로 변환해 Mel-스펙트로그램을 생성했다. 이후 Mel-스펙트로그램에서 주요 오디오 특징을 추출하여 MFCC 를 생성하고, 이를 정규화했다.

표 2. n 에 따른 오케스트라 음원 데이터셋 개수

악기수	개수	악기수	개수	악기수	개수	악기수	개수
1	13,980	5	17,000	9	18,000	13	10,000
2	14,000	6	18,000	10	18,000	14	10,000
3	15,000	7	19,000	11	18,000	15	10,000
4	16,000	8	20,000	12	10,000	16	5,000

2.2 오케스트라 악기 식별 방법

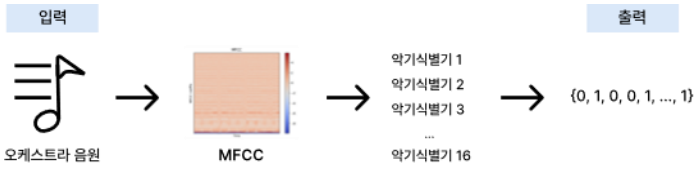

그림 1. 오케스트라 악기 식별 방법

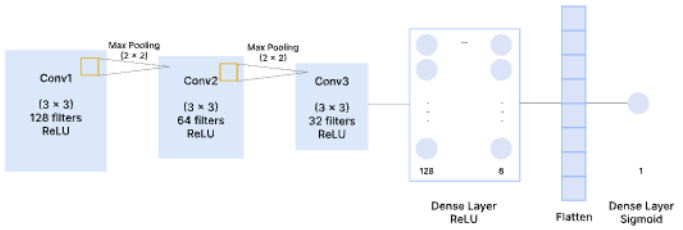

그림 2. 악기식별기 구조

그림 1 은 제안하는 오케스트라 악기 식별 방법을 보여준다. 16 개의 악기식별기를 통해 추출된 결과로 4 초짜리 오케스트라 음원에 연주 중인 모든 악기를 도출한다. 각 악기식별기는 입력된 데이터 내의 특정 악기의 존재 여부를 출력한다.

악기식별기의 입력은 음원의 MFCC 이고, 입력된 데이터 내에 악기가 존재하면 1 을, 그렇지 않으면 0 을 출력한다. 악기식별기는 그림 2 에 나와있는 구조를 갖는 CNN 모델이다. 텐서플로의 케라스 라이브러리를 사용하여 모델 생성 및 훈련을 진행했다[5].

CNN 모델의 기본적인 구조는 [2]를 참고했다. 각각 128, 64, 32 개의 필터를 가지는 3x3 커널의 컨볼루션 레이어로 이루어져 있으며 각 레이어는 맥스 풀링과 배치 정규화 레이어를 거친다. 컨볼루션 레이어 이후에는 3 개의 밀집층과 드롭아웃을 추가했다. 추가적으로 각 악기식별기 별로 모델 구조와 하이퍼 파라미터 수정을 통해 성능을 개선했다.

3. 결과

오케스트라 악기 식별 방법의 성능 평가를 위해 테스트 데이터셋을 기반으로 정확도와 F1-score 를 측정했다. 또한, 실용성 평가를 위해 사용자 평가를 진행했다.

3.1 성능 평가

표 3 악기식별기 별 정확도 및 F1-Score

	정확도	F1-Score
바순	0.88	0.89
베이스 클라리넷	0.92	0.81
첼로	0.86	0.85
클라리넷	0.83	0.84
크래쉬 심벌	1.00	1.00
콘트라 베이스	0.87	0.87
플루트	0.85	0.87
프렌치 호른	0.82	0.84
오보에	0.90	0.91
색소폰	0.82	0.84
탬버린	1.00	1.00
트롬본	0.84	0.86
트럼펫	0.86	0.88
튜바	0.90	0.90
비올라	0.84	0.85
바이올린	0.89	0.90

표 3 은 평가 데이터셋에 대한 각 악기식별기의 평균 정확도와 F1-score 이다. 탬버린과 크래쉬 심벌의 정확도가 가장 높았으며, 프렌치 호른과 색소폰의 정확도가 가장 낮았다. 프렌치 호른의 악기식별기는 0.82 로 가장 낮은 정확도를 보인다. 프렌치 호른의 소리는 다른 악기들에 비해 부드러운 소리를 낸다는 특징을 가지고 있다[6]. 이러한 특성으로 다른 악기들과 조화롭게 어우러져, 악기 식별에 어려움이 생겨 나온 결과로 보인다. 오보에, 튜바, 베이스 클라리넷의 정확도는 순서대로 타악기의 뒤를 따랐다. 악기군에 따른 평균 정확도는 타악기 (1.00), 목관악기(0.86), 현악기(0.86), 금관악기(0.86) 순으로 높았다. 타 분류에 비해 타악기의 평균 정확도가 높게 나타났다. 이러한 결과는 타악기의 특징적인 공명음, 진동 등의 음향 특성 때문으로 보인다.

평가 데이터셋에 대해 16 개의 악기식별기를 통해 특정 시점에 연주 중인 모든 악기 목록을 예측했다. 이를 실제 연주 중인 악기 목록과 비교하여 정확도를 측정했다. 오케스트라 악기 식별 정확도는 86.82%로 평균 16 개 중 약 14 개의 악기에 대한 정답을 올바르게 예측하는 것으로 나타났다.

3.2. 사용자 설문조사

제안한 시스템의 실용성 평가를 위해 클래식 악기에 익숙한 음악인과, 비음악인을 대상으로 구글 폼을 이용해 설문조사를 진행했다. 음악 전공 여부, 오케스트라 관련 활동 참여 여부, 그리고 1 년간 오케스트라 연주 관람 및 참여 횟수 등의 사전 질문을 통해 음악인과 비음악인을 구별했다. 음악인 10 명, 비음악인 20 명을 대상으로 설문조사를 진행했다. 4 초 단위의 오케스트라 음원을 듣고 음원에서 들리는 악기를 사용자가 모두 선택하도록 했다. 설문조사에서는 최소 2 개, 최대 5 개의 악기가 연주된 오케스트라 음원을 사용했다. 다음 단계에서는 악기식별기가 존재한다고 예측한 악기 목록을 함께 제공하며 같은 음원에 대해 다시 한번 답을 선택하도록 했다.

설문조사 결과, 비음악인의 경우, 예측된 악기 목록 제공 전의 정답률은 75.7%이었으며 목록 제공 후 84.6%로 약 9% 증가했다. 음악인의 경우, 목록 제공 전 정답률은 88%이었으며 제공 후 91%로 3% 증가했다. 사후 질문으로 악기식별기가 예측한 악기 목록가 음악감상에 어떤 영향을 미치는지에 대한 답변으로 "헷갈리는 음에 대한 가이드라인을 제시한다", "놓쳤던 선율을 확인할 수 있다", "악기마다의 표현에서 음악적 감상을 느낄 수 있었다" 등의 긍정적인 답변이 주를 이뤘고, "선입견을 준다", "생각한 것과 예측값이 다른 경우 혼란을 준다" 등의 답변도 일부 있었다. 마지막으로 "향후 제안하는 모델을 이용할 의향이 있는가?" 질문에 비음악인의 70%와 음악인의 60%가 "그렇다"라고 답했다.

4. 결론

본 연구는 딥러닝 기술을 활용해 오케스트라 음원에서 동시에 연주되는 다수의 악기를 식별하는 시스템을 제안하고 사용자 평가를 진행했다. 향후 연구에서는 n 에 따른 정확도의 경향성을 파악하고, 악기의 연주 빈도와 가중치를 고려해 성능을 개선할 계획이다.

5. 참고문헌

[1] Philippe Hamel et al., "Automatic Identification of Instrument Classes in Poly-Instrument Audio." ISMIR, 2009.
[2] Blaszke M. and Kostek B., "Musical Instrument Identification Using Deep Learning Approach." Sensors (Basel), 2022.
[3] Nagawade M. S. and Ratnaparkhe V. R., "Musical instrument identification using MFCC." IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), 2017.
[4] Mahanta, S.K. et al., "Exploiting cepstral coefficients and CNN for efficient musical instrument classification." Evolving Systems, 2023.
[5] "Keras | Tensorflow Core", Tensorflow, 2020 년 06 월 05 일 수정, https://www.tensorflow.org/guide/keras?hl=ko
[6] Philip Farkas.,『The Art of French Horn Playing』. Alfred Music Publishing Co., Inc. (1995)