

멜로디 기반 화음 예측: 오토인코더와 FCNN을 이용한 통합적 접근

고서영, 이재호*

덕성여자대학교

20210808@duksung.ac.kr, *izeho@duksung.ac.kr

Melody-Based Chord Prediction: An Integrated Approach Using Autoencoders and FCNNs

Seo Young Koh, Jae Ho Lee*

DuksungWomen's Univ.

요약

본 논문은 인공지능영역을 활용하여 짧은 마디의 멜로디에서 화음을 예측하는 새로운 모델을 개발하고자 한다. 연구의 목적은 오토인코더를 사용하여 멜로디의 핵심적인 특징을 추출하고, 이를 바탕으로 FCNN을 통해 화음을 예측하는 것이다. 실험 과정에서는 각기 다른 12개의 화음에 특화된 오토인코더를 각각 독립적으로 학습시켰다. 이러한 방식은 각 화음에 최적화된 특징 추출을 가능하게 하며, 추출된 특징들은 FCNN에 입력되어 최종적으로 화음을 분류한다. 이는 멜로디와 화음 간의 관계를 이해하고 예측하는 데 있어서 음악 이론과 인공지능 기술의 융합을 통해 새로운 방향을 제시하며, 음악 분석과 작곡 지원 도구 개발에 유용한 기여할 수 있다.

I. 서론

멜로디와 화음 간의 상호작용은 음악 이론과 작곡의 중요한 부분을 이룬다. 멜로디는 연속적인 음의 높낮이 변화와 리듬의 조화로 이루어지며, 감정과 메시지를 전달하는 음악의 주된 수단이다. 반면, 화음(chord)은 높이가 다른 2개 이상의 음이 동시에 울려 합성되는 소리로, 음악의 조화와 구조를 형성하는 데 중요한 역할을 한다. 이 두 요소의 조화는 음악의 복잡성을 이해하는 데 있어 필수적이다.

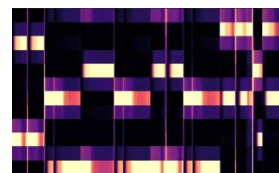
본 연구는 멜로디로부터 화음을 예측하는 모델을 개발하고자 한다. 연구의 목적은 오토인코더(Autoencoder)를 사용하여 멜로디로부터 동일 화음 영역의 핵심적인 각 화음마다 특징을 추출하고[1], 이를 바탕으로 완전 연결 신경망(Fully Connected Neural Network, FCNN)을 통해 해당 멜로디에 적합한 화음을 예측하는 것이다. 이러한 접근법은 기존 음악 이론과 머신러닝 알고리즘을 결합하여, 멜로디와 화음 간의 복잡한 관계를 이해하고자 한다. 음악 이론가, 작곡가, 그리고 음악 교육자들에게 멜로디와 화음 관계를 해석하고 창작하는데 있어 유용한 도구를 제공할 수 있다.

II. 본론

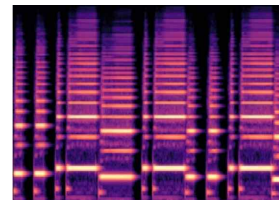
1. 데이터 전처리

본 연구에서 사용된 데이터는 2마디 길이의 멜로디로 구성된 wav 파일 형식으로, 각 파일은 전문 음악가에 의해 수작업으로 화음이 레이블링 되었다. 데이터의 전처리 과정은 다음과 같다. 각 wav 파일은 Mel Spectrogram으로 변환되었다. 이는 멜로디의 주파수 내용을 시간에 따라 시각화하는 데 사용되는 방법으로, 인간의 귀가 소리를 인식하는 방식을 모방하여 주파수를 처리한다. 동일한 wav 파일에서 Chromagram도 생성되었다. Chromagram은 음악의 12가지 다른 음(pitch)에 대한 강도를 시간에 따라 표시하여, 음악의 화성적 구조를 나타낸다. 생성된 Mel

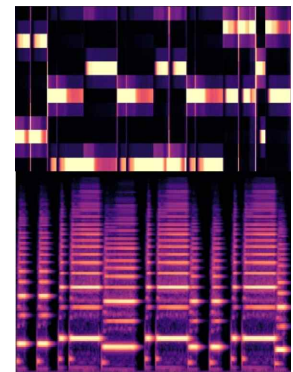
Spectrogram과 Chromagram은 세로 방향으로 결합되어 하나의 이미지를 형성한다. 이는 멜로디와 화음의 특징을 모두 포함하여, 모델이 두 가지 유형의 데이터를 동시에 고려할 수 있도록 한다.



(a) Chromagram



(b) Mel-spectrogram



(c) Combined Image

Figure 1. 음악 데이터의 이미지 변환 처리 결과 예시

: (a) 크로마그램, (b) 멜 스펙트로그램, (c) 합쳐진 이미지

2. 모델

2-1. 오토인코더

멜로디와 화음 데이터의 특징을 효과적으로 추출하고 복원하기 위해 Convolution 오토인코더를 구현하였다. 이 모델은 복잡한 오디오 데이터를 처리하고, 핵심 특징을 압축하는 데 중점을 두고 설계되었다.

인코더 부분은 입력 데이터의 특징을 효과적으로 압축하는 역할을 한다.

이는 다음과 같은 계층들로 구성되어 있다. 첫 번째 Convolution 계층은 32개의 필터를 사용하며, 'relu' 활성화 함수와 'same' 패딩을 적용한 3*3의 convolution 계층이다. 공간적 차원을 축소하면서 중요한 특징을 유지하기 위한 2*2 크기의 Max Pooling 계층을 사용한다. 두 번째 Convolution 계층은 16개의 필터를 사용하며, 다시 'same' 패딩과 'relu' 활성화 함수를 적용한다. 최종 인코딩 계층은 또 다른 Max Pooling 계층을 통해 인코딩 과정을 마무리한다. 디코더 부분은 인코더에서 압축된 데이터를 원래의 차원과 형태로 복원한다. 이 과정은 다음 계층들을 통해 이루어진다. 인코딩된 데이터는 64개의 필터를 사용하는 3*3 Convolution 계층을 거치고, 이어서 2*2 Up Sampling 계층을 통해 차원이 확대된다. Cropping 계층을 사용하여 이미지의 크기를 조정하고, 이를 통해 출력 데이터의 크기를 원래 입력 데이터의 크기와 일치시킨다. 마지막으로, 입력 데이터의 채널 수와 동일한 크기의 필터를 사용하는 convolution 계층을 통해 최종 출력 이미지를 생성한다. 여기서는 'Sigmoid' 활성화 함수를 사용한다.

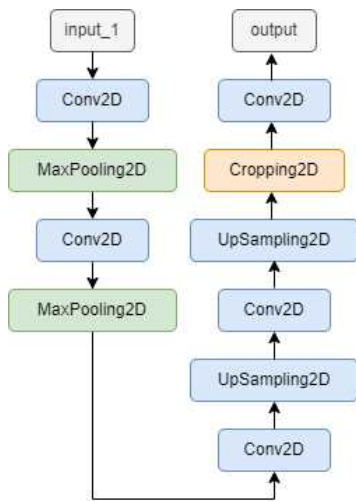


Figure2. 오토인코더 모델 구조도

2-2. FCNN

화음 분류를 위한 완전 연결 신경망이다. 오토인코더의 인코더를 통해 추출된 데이터를 입력으로 받아, 화음을 예측하는데 사용된다. FCNN의 구조는 다음과 같다. 오토인코더로부터 출력된 feature 맵을 1차원 벡터로 평탄화한다. 이후 3개의 Dense층으로 완전 연결 계층을 이룬다. 각각 128, 64, 128개의 유닛을 가지고 있다. 최종 Dense 계층은 12개의 유닛을 가지며 최종적으로 화음을 분류한다. 각 유닛은 특정 화음을 나타낸다.

3. 학습

화음 예측을 위해 특화된 12개의 오토인코더를 구축하고 학습시키는 방법을 채택하였다. 각 오토인코더는 특정 화음에 해당하는 멜로디 데이터로만 훈련되었다. 이 접근 방식의 목적은 각 화음에 특화된 특징 추출을 가능하게 하는 것이다. 학습 과정은 다음 단계로 구성된다.

화음 별로 레이블링 된 멜로디 데이터를 각각의 화음에 해당하는 12개의 그룹으로 분할한다. 각 그룹의 데이터를 사용하여 해당 화음에 특화된 오토인코더를 별도로 학습시킨다. 각 오토인코더는 특정 화음에 최적화된 특징을 추출할 수 있게 된다. 각 오토인코더의 인코더만을 사용하여 학습된 특징을 추출한다. 추출된 특징을 입력으로 사용하여, FCNN을 통해 최종 화음을 분류한다. 이 과정에서 FCNN은 12개의 다른 특징 세트 각각에

대해 화음을 예측하도록 훈련된다.

III. 실험 결과 분석

테스트 데이터를 12개의 학습된 오토인코더 각각에 통과시켜, 각 화음에 대한 특징 값을 추출한다. 추출된 특징들을 각각 FCNN에 입력하여 화음을 예측한다. 이때, FCNN의 출력은 softmax 함수를 통해 화음의 확률 분포를 나타낸다. 각 FCNN의 출력 중 가장 높은 softmax 값을 가진 화음을 최종 예측 결과로 선택한다. 이는 모델이 가장 확신하는 화음 예측을 나타낸다.

IV. 결론

본 연구는 멜로디로부터 화음을 예측하는 새로운 접근 방식을 제시하고 구현하였다. 이 연구의 주요 기여는 12개의 화음별 특화 오토인코더와 이를 활용한 FCNN 기반의 화음 분류 방법의 개발이다. 멜로디의 특징을 화음별로 효과적으로 추출하고, 이를 기반으로 정확한 화음 예측을 가능하게 한다. 실험 결과는 제안된 모델이 멜로디 데이터를 바탕으로 화음을 예측하는 데 있어 정확도 향상을 보여준다.

이 연구는 음악 이론과 인공지능의 접목을 통해, 음악 분석과 작곡 지원 도구 개발 분야에 도움을 줄 수 있다.

ACKNOWLEDGMENT

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원 받아 수행된 연구임(과제번호- 2022R1A2C1009951).

참 고 문 헌

- [1] ZHANG, Yifei. A better autoencoder for image: Convolutional autoencoder. In: ICONIP17-DCEC. Available online: http://users.cecs.anu.edu.au/Tom.Gedeon/conf/ABCs2018/paper/ABCs2018_paper_58.pdf (accessed on 23 March 2017). 2018.