

데이터 분포에 기반한 멜로디의 벡터 표현

Distribution-based Vectorization of Melody Lines

최길웅

Kiroong Choe

서울대학교

Seoul National University

krchoe@hcil.snu.ac.kr

서진욱

Jinwook Seo

서울대학교

Seoul National University

jseo@hcil.snu.ac.kr

요약문

본 연구에서는 멜로디를 음악적 사전 지식 없이 데이터의 분포만을 사용해서 벡터화하는 방법론을 분석하였다. 이를 위해 단선율 멜로디를 음표와 쉼표의 나열로 간주하고, 자연어 처리 기법을 사용하여 분절 단위를 추출하여 각 분절 단위마다 100 차원의 벡터 표현을 계산하는 파이프라인을 제안하였다. 7000 개의 합창곡 악보를 통해 분석한 결과, 본 연구에서 제안하는 벡터화 파이프라인을 통해 박자, 음고, 장르, 성부 정보를 담아낼 수 있음을 확인하였다. 이러한 결과는 특정 과업 또는 이론에 국한되지 않은 데이터의 일반적인 표현을 음악 정보검색의 다양한 응용 장면에 활용할 수 있다는 가능성을 시사한다.

주제어

멜로디, 벡터표현, 자연어 처리, 음악정보검색

1. 서론

멜로디는 다양한 음악 정보검색 과업에서 중요 단서로 활용될 수 있으나, 구체적인 문제 상황에 특정된 지식 표현은 일반화 가능성의 측면에서 극히 제한된다. 예컨대, 구체적인 과업을 위해 훈련된 특정 네트워크의 지식표상은 멜로디에 관한 정보를 간접적으로 담고 있지만, 이를 다른 과업에 전이시키기 까다롭다. 또한, 계산에 사용된 멜로디 특성(feature)이 음악 이론에 바탕을 둔 전처리에 기반하는 경우, 그 지식표상의 설명력은 해당 음악 이론에 한 번 더 국한된다. 본 연구에서는 주어진 코퍼스 내에서의 빈도와 분포에 기반하여 사전지식 없이, 다양한 과업에 활용할 수 있는 일반화된 지식 표상으로서 멜로디의 벡터 표현(혹은 임베딩)을 계산해내고자 하였다.

2. 멜로디 임베딩 파이프라인

이 장에서는 원 멜로디를 벡터 표현으로 바꾸기 위한 전체 과정을 기술한다.

2.1 멜로디의 정의

본 연구에서는 멜로디의 구성요소를 음표와 쉼표라고 정의하였다. 음표는 음정과 길이의 조합,

쉼표는 길이에 의해 유일하게 결정된다. 예를 들어, 음표의 예시는 E5 음정이 4 박 동안 지속되는 것이며, 쉼표의 예시는 2 박 동안 쉬는 2 분쉼표, 4 박 동안 쉬는 온쉼표 등이 있다. 또한 본 연구에서는 멜로디를 단선율에만 국한하여, 음표 혹은 쉼표가 일차원으로 나열된 것이 곧 멜로디라고 정의하였다.

2.2 멜로디 분절

멜로디를 의미 있는 구성 성분으로 쪼개기 위해서는 적절한 분절이 필요하다. 음악의 종류에 따라 2 마디 혹은 4 마디를 하나의 분절 단위로 보는 것이 일반적이거나, 본 연구에서는 음악적인 사전 지식 없이 분포만을 통해서 벡터 표현을 계산하기를 목표로 하므로 그러한 분절법을 사용하지 않았다. 또한, 멜로디 분절 단위의 시작과 끝이 마디의 시작과 끝과 항상 일치하지는 않는다. 따라서 주어진 코퍼스 내에서 빈도를 근거로 패턴을 만들어내기 위하여 SentencePiece[1] 모델을 사용하였다. 여기서 전체 멜로디는 문장, 쪼개진 멜로디 조각은 단어에 비유될 수 있다.

2.3 벡터 표현 계산

멜로디를 조각 단위로 분절하고 나면 주어진 데이터를 보편적인 언어 코퍼스와 동치로 생각할 수 있다. 따라서 Word2Vec[2] 등의 방법론으로 벡터 표현을 얻는 것이 가능한데, 여기서는 FastText[3]를 사용하였다. FastText 는 문장 내의 각 단어를 trigram 으로 쪼개어 subword 정보를 바탕으로 embedding 을 계산하므로, 멜로디 분절 과정에서 고려되지 않은 작은 음악적 단위를 포착할 수 있다.

3. 분석

3.1 데이터셋

분석을 위해 Choral Public Domain Library[4]에 수록된 7000 개의 합창곡 악보를 사용하였다. 해당 데이터셋은 하나의 곡을 소프라노, 베이스 등의 성부로 나누어 담고 있으며, 장르, 연도, 작곡가 등의 메타데이터를 함께 제공한다.

각 성부의 전체 멜로디를 하나의 독립된 멜로디로 취급하여 총 35,000 개의 멜로디가 분석에 사용되었다. 전술한 파이프라인을 통해 총 32,000 개의 멜로디 조각을 추출하고, 각 멜로디 조각에 대응되는 100 차원의 벡터 표현을 계산했다.

그 다음, 전체 멜로디의 벡터 표현은 그것을 구성하는 멜로디 조각들의 벡터 표현을 평균한 값으로 정의하였는데, 이는 투사되는 벡터 공간이 성기다(sparse)는 전체 하에 멜로디 조각들의 Bag of Words 로 생각할 수 있다.

3.2 단어 단위 분석

각 멜로디 조각(단어)이 하나의 실수 벡터에 대응되므로, 코사인 유사도[5]를 사용하여 특정 단어와 가장 유사한 단어가 무엇인지를 알아볼 수 있다. 그 결과, 그림 1 과 같이 질의에 해당하는 멜로디 조각과 음정이 같지만 리듬 패턴이 다른 멜로디 조각과 리듬 패턴이 다르지만 음정이 같은 멜로디 조각이 가장 가까운 것으로 나타났다. 코퍼스 분석 과정에서 음정과 박자 정보는 익명화되어 고려되지 않았음에도 유사한 음정과 박자를 지닌 멜로디 조각들이 서로 가깝게 인식된 것을 통해, 본 벡터 표현법이 코퍼스 내에서 유사한 분포를 가진 멜로디 조각들을 포착하고 있음을 알 수 있다.



그림 1 질의된 멜로디 조각과 가장 유사한 5 개의 멜로디 조각. 위에서부터 더 가까운 순으로 정렬되어 있으며, 질의 멜로디 조각과 음정이 같거나 박자 패턴이 유사한 멜로디 조각들로 구성되어 있음을 알 수 있다.

그림 2 는 계산된 벡터 표현이 장르 등의 의미 있는 정보를 잘 분별해내는지 파악하기 위하여 Canon 과 Carol 의 두 장르 중 하나에 해당하는 멜로디 조각들을 t-SNE 시각화 기법[6]을 통해 2 차원에 투영한 것이다. 각 장르별로 클러스터를 구성하고 있는 것을 확인할 수 있다. 그림 3 은 각

클러스터의 대표 멜로디 조각을 샘플링한 것인데, Carol 의 경우 반복적인 리듬 패턴이 주요한 반면, Canon 의 경우 화성 진행과 연관된 멜로디 패턴이 주가 됨을 확인할 수 있다.

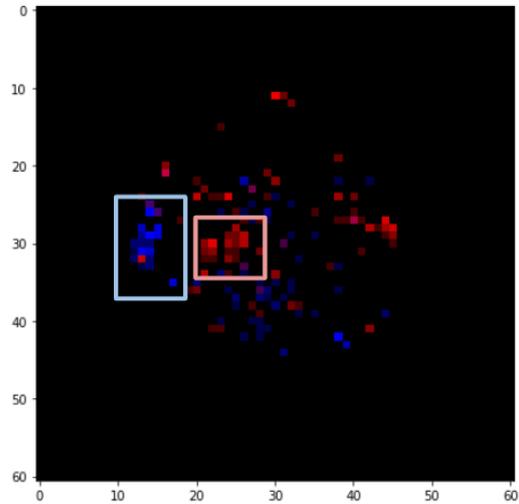


그림 2 Carol(빨강)과 Canon(파랑)에 해당하는 멜로디 조각의 t-SNE 시각화



그림 3 Carol(왼쪽)과 Canon(오른쪽)을 대표하는 멜로디 조각

3.3 Sentence-Level Observations

단어 벡터를 단순 합산하여 구해진 멜로디 벡터가 해당 멜로디의 의미 있는 성분을 포착하고 있음을 확인하기 위해 추가적인 분석을 수행하였다. 그림 4는 멜로디 조각(단어) 벡터 중 소프라노 혹은 베이스 성부에 속한 것만을 시각화한 것이다. 그림과 같이 명백하게 두 그룹으로 쪼개어진 것을 확인할 수 있으며, 이는 소프라노와 베이스를 구성하는 음정이 현저하게 차이 나기 때문으로 해석할 수 있다.

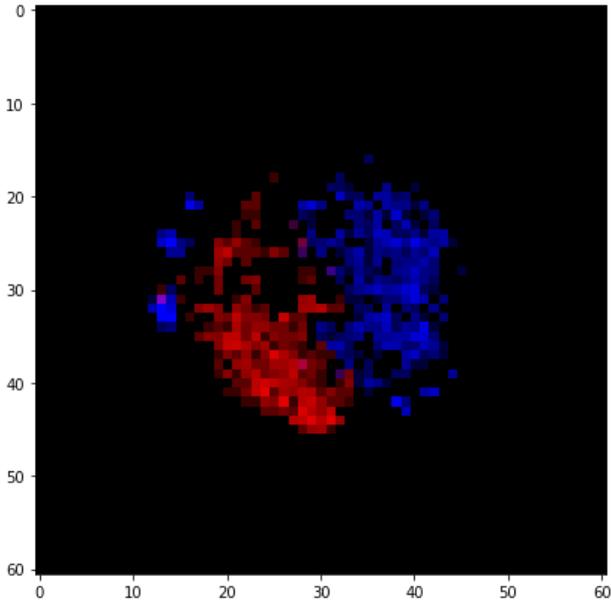


그림 4 소프라노(빨강)와 베이스(파랑)를 구성하는 멜로디 조각의 t-SNE 시각화

한편, 그림 5 는 소프라노와 베이스를 구성하는, 멜로디 조각이 아닌 전체 멜로디의 벡터를 시각화한 것인데, 그림 4 와 달리 소프라노와 베이스가 분별되지 않고 오히려 작은 단위로 응집되어 있는 것을 확인할 수 있다. 이는 음정의 차이를 상쇄할 수 있는 멜로디적 유사성이 반영된 것이라 추측할 수 있다. 해당 추측을 검증하기 위해, 그림 6 에서는 서로 다른 성부의 멜로디가 같은 곡에 속해 있을 때와 다른 곡에 속해 있을 때의 코사인 유사도를 비교하였고, 같은 곡에 속한 서로 다른 두 성부가 더 높은 유사도를 보이는 경향이 존재함을 확인하였다.

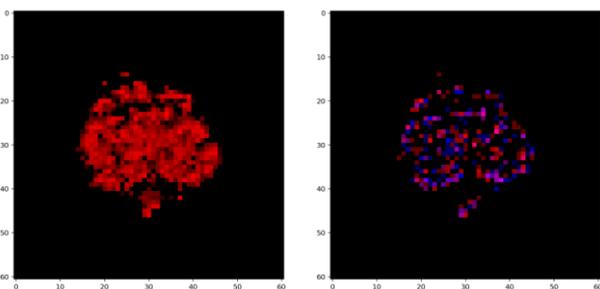


그림 5 모든 멜로디 벡터(왼쪽)와 소프라노(오른쪽; 빨강) 혹은 베이스(오른쪽; 파랑)의 멜로디 벡터의 t-SNE 시각화

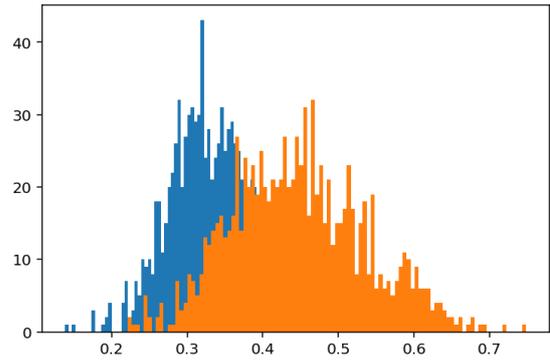


그림 6: 같은 곡에 속한 다른 성부의 멜로디의 거리(파랑), 다른 곡에 속한 두 성부 멜로디의 거리(주황)의 히스토그램. x 축은 코사인 유사도를 나타내며, y 축은 빈도를 나타냄.

4. 결론

본 연구에서는 음악의 멜로디를 구성하는 분절 단위를 데이터의 분포 기반으로 찾아내고 벡터화하는 파이프라인을 제안하였다. 이를 통해 음악이론 등의 사전지식 없이도 박자 유사성, 음악의 장르, 음악의 성부 등의 개별 특성이 벡터 표현상의 차이로 나타나는 것을 알 수 있었다. 특히, 같은 악보의 다른 성부가 구성음의 차이에도 불구하고 벡터 공간에서 더 근접하게 나타났는데, 이는 개별 특성을 넘어선 멜로디 자체의 표상이 벡터 표현에 포착됨을 보였다는 의의가 있다.

그러나 분석 방법이 관찰에 주로 기반하여 다각적인 측면에서 검증하지 못한 것은 아쉬움으로 남는다. 특히 데이터셋에 존재하지 않는 임의의 멜로디로 질의하거나, 음악이론적으로 유사한 멜로디 간의 관계를 파악하는 등 데이터 외적 정보와의 비교분석을 통해 본 연구의 외적 타당성을 강화할 필요가 있다. 예를 들어, 푸가(Fugue)에서 자주 사용되는 뒤집어진 멜로디 관계가 벡터 공간에서 어떻게 나타나는지를 분석해볼 수 있다.

또한 두 마디가 서로 인접한 관계인지 아닌지를 판단하는 등, 실제 분류과업(classification task)에 해당 벡터 표현을 사용할 때 어느 정도의 성능을 얻을 수 있는지도 알아볼 필요가 있다. 이와 같은 후속 연구가 이어진다면, 본 연구에서 제안하는 분포 기반의 멜로디의 벡터 표현을 노래 유사도 판별, 노래 추천, 커버곡 탐지와 같은 실제 음악 정보검색 과업에 활용할 수 있을 것이다.

참고 문헌

1. Kudo, T., Richardson, J. SentencePiece: A Simple and Language Independent Subword Tokenizer and Detokenizer for Neural Text

Processing. arXiv preprint
arXiv:1808.06226 (2018).

2. Mikolov, T., Chen, K., Corrado, G., & Dean, J. Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781. (2013).
3. Choral Public Domain Library.
<https://www.cpd.org/>
4. Christopher D. M., Prabhakar R., and Hinrich S. Introduction to Information Retrieval. Cambridge University Press, USA. (2008).
5. van der Maaten, L. & Hinton, G. Visualizing Data using t-SNE . Journal of Machine Learning Research, 9. (2008).