계층적 정보를 활용한 하이퍼네트워크를 통한 음악 예측

한철호O, 이상우, 장병탁
서울대학교 컴퓨터공학부
{chhan, slee, btzhang}@bi.snu.ac.kr

Music Prediction by Hypernetworks with Hierarchical Information

Cheolho HanO, Sang-Woo Lee, Byoung-Tak Zhang
School of Computer Science and Engineering, Seoul National University

요 약

이 논문은 계층적 정보를 활용한 하이퍼네트워크(hypernetwork with hierarchical information, HHI)를 통해 음악을 예측하는 기법을 제안한다. 음악은 구조적으로 계층을 가지고 있기에 계층적 정보는 음악 예측에서 필수적이다. 기존에 음악 예측을 위해 사용된 프레임워크 중 하나로 하이퍼네트워크가 있다. 이 프레임워크는 고수준의 개념들 간 상호작용을 통해 그 구조를 형성하는 개념적인 프레임워크로서, 음악의 자기조직적인 성질을 갖고 있어 음악 예측에 자주 활용되었다. 하지만 음악의 구조는 계층적으로 되어 있기 때문에, 이를 학습하는 구조 역시 계층적 정보를 필요로 한다. 본 논문에서는 음의 높이와 길이로 이루어진 음표를 HHI를 통해 학습 및 예측하였고, Beatles의 4/4 박자의 곡 중 160곡에 대하여 예측 정확도를 보였다.

1. 서 론

음악을 학습하고 기억하고 예측하는 데 있어서, 음악을 어떻게 표현할지 고려할 필요가 있다. 음악에 대한 기억을 표현하기 위해 음악을 사람이 이해할 수 있는 단위로 끊어 표현하는 분절(segmentation)은 핵심적으로 사용될 수 있다[1]. 예를 들어, 음악은 음 단위로 끊어서 나타낼 수 있다. 언어 모델로서도 활발히 응용되어온 가변차수 마코프(variable-order Markov) 모델[2]과 n-그램(gram) 모델[3]과 같은 분절 기반 모델들이 기호적(symbolic) 음악에 대한 회상 기억을 구축하기 위해 사용되었다. n-그램 모델은 음의 서열을 짧은 조각으로 나누고, 저장된 n-그램의 통계에 근거하여 주어진 문맥에 대해 예측한다. 이때 문맥에 일치하는 조각의 연장에 대한 확률 분포에 근거하여 예측하게 된다. 마코프 모델은 상태 간의 전이 확률을 나타낸다. 은닉 (hidden) 마코프 모델은 관찰되지 않는 은닉 변수를 가지며, 은닉 변수들이 관찰 변수에 대한 확률 분포를 결정하게 된다. n-그램 모델과 달리 은닉 마코프 모델은 이전의 모든 사건을 고려하여 예측을 하게 된다[3].

순차적 예측은 과거의 사건이 주어졌을 때, 그 이후의 사건을 예측하는 것이며, 음악 분야에도 적용되어 왔다[2-3]. 분절에 기반을 두고 순차적 예측을 하는 모델로 하이퍼네트워크(hypernetwork, HN)가 있다[4-11]. 특히 [11]에서는 음의 서열로부터 조합 자질(associative feature)을 실시간으로 학습하였다. 이에 더해, 이 논문은 시각적 인지 과정에서 활용된 공간적 계층 구조[12]에 착안하여, 시간상의 계층적 정보를 활용한 하이퍼네트워크(hypernetwork with hierarchical information, HHI)를 제안한다.

이 논문은 HHI의 구조와 학습 및 예측 알고리즘을 소개한다. 그리고 실험적으로, 팝송의 MIDI 파일을 순차적으로 학습하여, 제안하는 모델의 예측 성능을 보이고, 그 구조에 대해 분석한다.

2. HHI를 통한 음표 예측

2.1 순차적 음표 예측

순차적 예측은 음악 등에서 널리 사용되어왔다[2-3, 8-9]. 음표의 집합을 S라고 하자. 순차적으로 i번째 음표 $s_i \in S$를 예측하기 위해, 이전까지의 음표의 서열 s_1^{i-1}이 주어졌을 때 새로 s가 나타날 조건부 확률 $p(s|s_1^{i-1})$을 계산한 후 이 확률을 가장 크게 하는 음표

$$\hat{s}_i = \arg\max_{s \in S} p(s|s_1^{i-1})$$

를 고를 수 있다.

하지만 과거의 모든 음표를 고려한 확률 $p(s|s_1^{i-1})$을 계산하기보다는 최근 $n-1$ 개의 음표의 서열 s_{i-n+1}^{i-1}이 주어졌을 때의 조건부 확률 $p(s|s_{i-n+1}^{i-1})$을 계산하고 이를 $p(s|s_1^{i-1})$의 근삿값으로 사용하는 방법이 더욱 효율적이다. 이때 i와는 독립적으로 관찰되는 n 길이의 서열 $g_1^n = s_{i-n+1}^{i}$에 대한 분포를 통해, g_1^{n-1}에 s_{i-n+1}^{i-1}을 대입한 후 s_i를 다음과 같이 예측할 수 있다.

$$\hat{s}_i = \arg\max_{g_n \in S} p(g_n | g_1^{n-1} = s_{i-n+1}^{i-1}) \quad (1)$$

이러한 설정은 가변차수 마코프 모델 [2], n-그램 모델 [3], 하이퍼네트워크 모델 [8-11] 등에 이용되어 왔고, 이 논문에서도 동일한 설정 하에서 시간상의 계층적 정보를 반영한 알고리즘을 제안한다.

2.2 계층적 정보를 활용한 하이퍼네트워크(HHI)

[8-11]에서는 하이퍼네트워크(hypernetwork, HN)를 통해 순차적으로 음표를 예측하였다. 하이퍼네트워크는 노드(node)와 여러 개의 노드를 묶는 하이퍼에지(hyperedge, HE)로 구성된다. 이때, 차수(order)가 2인 하이퍼에지로만 구성된 하이퍼네트워크일 경우, 일반적인 노드와 간선(edge)을 갖는 그래프가 된다. 하이퍼네트워크를 구성할

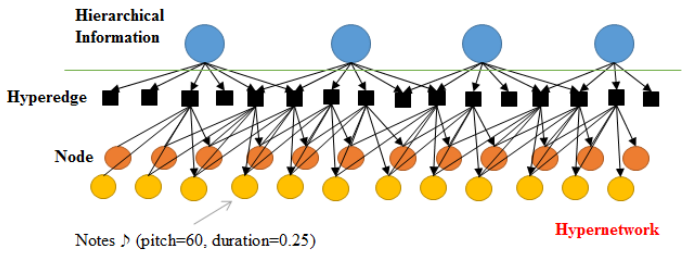

그림 1 계층적 정보를 활용한 하이퍼네트워크의 구조

때, 보통 최대 차수를 정해두고 2에서 최대 차수까지의 차수를 갖는 하이퍼에지들을 포함하도록 한다. 이때, 각 노드는 음표 등의 한 가지 개념을 나타내며, 하이퍼에지는 개념들의 집합을, 하이퍼에지의 가중치는 그 개념들의 상관도를 나타낸다.

하이퍼네트워크의 학습 및 예측 방법은 [8-11]에 기술되어 있다. 관측된 데이터와 하이퍼에지들을 비교하여 다양한 차수의 유사한 하이퍼에지들을 선택하여 이들의 가중합을 통해 예측을 진행한다. 학습 시에는 예측 결과를 원곡과 비교하여 가중치를 조정한다. 또한 가중치가 낮은 하이퍼에지들을 새로운 하이퍼에지들로 대체하는 작업을 수행한다.

HHI는 좀 더 긴 시간 범위를 반영하고 시간상의 계층적 정보를 고려하여 음표를 예측하기 위해 설계되었다 (그림 1). HHI는 계층적 정보를 활용해 예측하는 하이퍼네트워크로서 두 층으로 이루어져 있다. 하이퍼네트워크 층 위에 하이퍼네트워크의 하이퍼에지(hyperedge)들에 대한 계층적인 정보를 담은 노드들로 구성된 층이 있다.

2.2.1 계층적인 정보

계층적인 정보를 나타내는 상위 층의 각 노드는 하위 층의 하이퍼에지 집합의 부분집합이다. 부분 집합을 정하는 한 가지 방법은 하이퍼에지를 클러스터링(clustering)하는 것이다. 다른 방법은 하이퍼에지의 어떤 속성을 클러스터링하는 것이다. 곡의 한 마디를 실수 혹은 정수 값을 갖는 벡터로 나타낼 때, 각 마디 별로 k-평균 클러스터링(k-means clustering)을 수행할 수 있다. 하이퍼에지와 한 마디가 공통되는 음표의 서열을 갖고 있을 때, 그 하이퍼에지를 어떤 마디가 속하는 클러스터와 연결시킬 수 있다. 따라서 마디의 서열이 하위 층의 하이퍼에지의 부분집합의 서열이 되어, 상위 층의 각 노드를 이루게 된다.

2.2.2 계층적인 정보를 활용한 예측

하이퍼네트워크 층에서의 예측은 기존의 하이퍼네트워크와 비슷한 방식으로 진행하는데, 계층적인 정보를 어떻게 활용할 것인가 하는 문제가 있다. 먼저, 상위 층에서는 각 마디 별로 계층적인 정보를 제공한다. 다음으로 하위 층은 상위 층에서 제공된 계층적은 정보를 바탕으로 하여, 해당되는 하이퍼에지들을 통해 다음 노드인 음표를 예측한다.

3. 실험 및 결과

제안된 알고리즘을 통해 순차적으로 음표를 예측하는

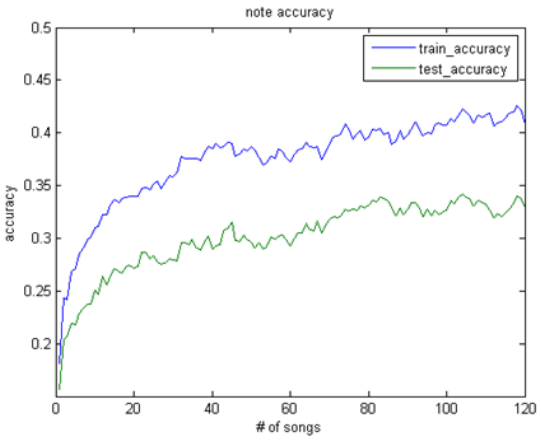

그림 2 학습 곡의 증가에 따른 음의 높이 예측 정확도 변화.

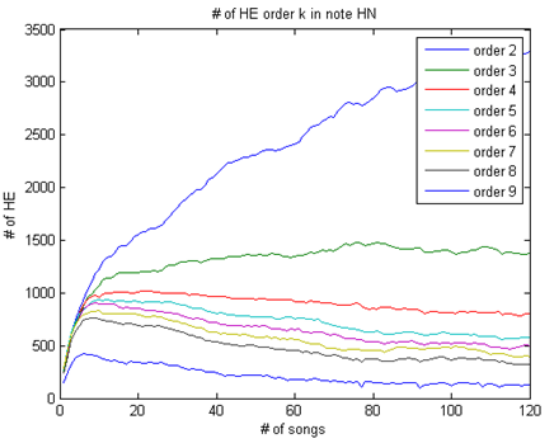

그림 3 음의 높이를 예측하는 하이퍼네트워크(HN)의 구조. 학습 초기에는 높은 차수(order)의 하이퍼에지(HE)들이 높은 비율을 차지하지만, 학습이 진행됨에 따라 낮은 차수의 하이퍼에지들이 늘어난다.

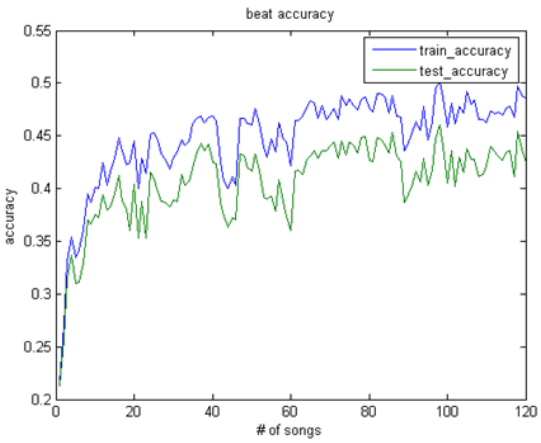

그림 4 학습 곡의 증가에 따른 음의 길이 예측 정확도 변화.

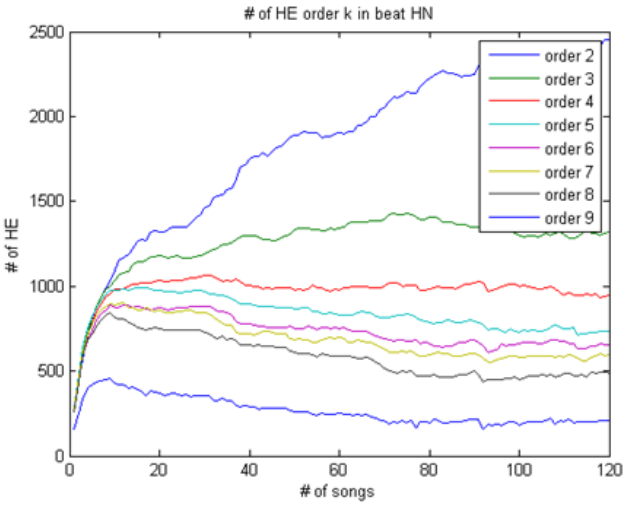

그림 5 음의 길이를 예측하는 하이퍼네트워크(HN)의 구조. 학습 초기에는 높은 차수(order)의 하이퍼에지(HE)들이 높은 비율을 차지하지만, 학습이 진행됨에 따라 낮은 차수의 하이퍼에지들이 늘어난다.

실험과 그 결과를 제시한다. 실험에 앞서, 단순한 곡에 비해 분석이 어려운 록 밴드 The Beatles의 4/4 박자의 곡 중 160곡을 선정하였고, 이 중 120곡을 학습 데이터로, 40곡을 테스트 데이터로 사용하였다. 이 곡들은 모두 MIDI 파일로 저장되어 있는데, 이를 통해 음표를 추출해 낼 수 있다. 학습 데이터가 지속적으로 새로 추가되는 상황에서 하이퍼네트워크를 학습하였다.

음의 높이(그림 1, 2)와 음의 길이(그림 3, 4)를 예측하는 하이퍼네트워크를 따로 구성하였고, 하이퍼에지의 최대 차수는 9로 하였다. 학습 곡이 늘어날수록 학습데이터로 사용된 곡과 테스트데이터로 사용된 곡에 대한 정확도가 점차 증가하였다. 이때 하이퍼네트워크를 구성하는 하이퍼에지들의 차수의 분포를 보면, 차수가 낮은 하이퍼에지들이 주를 이루었다.

4. 논의

기본적인 하이퍼네트워크는 n-그램과 비슷한 마이크로코드[6-11]를 통해 예측을 하되, 희소앙상블 코딩의 철학이 반영되어있다. 이 논문은 한 층의 하이퍼네트워크를 사용하는 대신, 하위 층의 하이퍼에지를 이용해 계층적 정보를 구성하여, 계층적 정보를 활용한 하이퍼네트워크를 구성하였다. 시간상의 계층적 정보를 고려함으로써 예측 시 좀 더 넓은 시간 범위를 반영하였다.

5. 결론 및 향후 연구

이 논문은 이전의 음표들이 관찰되었을 때, 그 다음의 음표를 차례로 예측하는 순차적 예측 문제에 접근하였다. 데이터가 계속 순차적으로 주어지는 것을 고려하여, 학습 곡이 점차 증가하는 상황에서 학습하도록 하였다. 또한 예측 시 좀 더 넓은 시간 범위를 고려하고, 음악이 가지는 시간상의 계층 구조를 반영하기 위해, 시간상의 계층을 고려한 HHI를 통해 순차적으로 음표를 예측하였다.

이후 상위 층을 구성하는 다양한 방법이 개발될 것으로 예상된다. 새로운 클러스터링 방법이나, 상위 층에서의 예측 등이 개발되면, 앞으로 여러 층을 갖는 하이퍼네트워크 또한 고려해 볼 수 있다.

감사의글

이 논문은 2015 년도 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원(R0126-15-1072-SW 스타랩, 10044009-HRI.MESSI)을 받아 수행된 연구임.

참고문헌

[1] B. Snyder (Sept. 18, 2012). Memory for music. *Oxford Handbooks Online*.

[2] R. Begleiter, R. El-Yaniv, and G. Yona, “On Prediction Using Variable Order Markov Model,” *Journal of Artificial Intelligence Research*, vol. 22, pp. 285-421, 2004.

[3] M. A. Rohrmeier and S. Koelsch, “Predictive information processing in music cognition. A critical review,” *International Journal of Psychophysiology*, vol. 83, pp. 164-175, 2012.

[4] S. Wu, S. Amari, and H. Nakahara, “Population coding and decoding in a neural field: a computational study,” *Neural Computation*, vol. 14, pp. 999-1026, 2002.

[5] H. Lee, A. Battle, R. Raina, and A. Y. Ng, “Efficient sparse coding algorithms,” *Advanced in Neural Information Processing Systems*, 2006.

[6] B.-T. Zhang, J.-W. Ha, and M. Kang, “Sparse population code models of word learning in concept drift,” in *Proc. of Annual Meeting of the Cognitive Science Society (CogSci 2012)*, pp. 1221-1226, 2012.

[7] B.-J. Lee, J.-W. Ha, K.-M. Kim, and B.-T. Zhang, “Evolutionary concept learning from cartoon videos by multimodal hypernetworks,” *IEEE Congress on Evolutionary Computation*, pp. 1186-1192, 2013.

[8] H.-W. Kim, B.-H. Kim, and B.-T. Zhang, “Evolutionary Hypernetworks for Learning to Generate Music from Examples,” *FUZZ-IEEE 2009*, pp. 47-52, Korea, Aug. 20-24, 2009.

[9] 구조학습 기반의 서열 데이터 재현 기법, 김병희, 장병탁, *한국정보과학회 가을학술발표 논문집*, 제39권 2(B), pp. 201-203, 2012.

[10] 희소앙상블 코딩을 이용한 순차적 음악 회상 학습, 한철호, 김병희, 장병탁, 한국정보과학회 동계학술발표회 논문집, pp. 631-633, 2014.12.

[11] 엔트로피를 최대화 하는 실시간 조합 자질 구축 방법, 이상우, 허민오, 장병탁, 2013 한국컴퓨터종합학술대회(KCC2013)논문집, pp. 1405-1407, 2013.06.

[12] Fritz, M., Andriluka, M., Fidler, S., Stark, M., Leonardis, A., & Schiele, B. (2010). Categorical Perception. In H. I. Christensen, G.-J. M. Kruijff, & J. L. Wyatt (Eds.), *Cognitive Systems*. Berlin: Springer Berlin Heidelberg.