

희소앙상블 코딩을 이용한 순차적 음악 회상 학습

한철호^o, 김병희, 장병탁
 서울대학교 컴퓨터공학부
 {chhan, bhkim, btzhang}@bi.snu.ac.kr

Sequential Music Learning for Recall with Sparse Ensemble Coding

Cheolho Han^o, Byoung-Hee Kim, Byoung-Tak Zhang
 School of Computer Science and Engineering, Seoul National University

요 약

이 논문은 음악 기억에 대한 계산 모델을 제안한다. 특히 회상에 대한 장기 기억에 초점을 두었다. 음악 심리학 연구를 통해 분절(segmentation)이 장기 기억 표현의 기초가 됨이 알려져 있다. 본 논문에서는 음악 회상이 순차적 정보 예측을 기반으로 이루어진다는 가정 하에, 기호적 음악에 대한 기억 및 회상 모델을 기존의 언어 모델 및 뉴럴 코딩의 특성을 취합한 희소 앙상블 코딩의 형태로 제안한다. 음악의 기억과 회상이 순차적인 정보 습득 및 예측으로 구현되는 과정에 대한 프레임워크와 알고리즘을 소개하고, 미디(MIDI) 형태의 팝송 멜로디 데이터에 모델을 적용하였다. 학습 과정 분석 및 멜로디 회상 실험 결과 제안한 모델의 가능성을 확인하였다.

1. 서 론

음악 회상에 대한 장기 기억을 표현하기 위해 분절(segmentation)은 핵심적으로 사용될 수 있다[1]. 언어 모델로서 활발히 응용되어온 가변차수 마코프(variable-order Markov) 모델[2]과 n-그램(gram) 모델[3]과 같은 분절 기반 모델들이 기호적 음악에 대한 회상 기억을 구축하기 위해 사용되었다.

순차적 예측은 과거의 사건이 주어졌을 때, 그 이후의 사건을 예측하는 것을 말한다. 순차적 예측은 음악 분야에서도 적용되어 왔다[2-3]. 이 논문에서는 순차적 예측에 대해 위 모델들처럼 분절에 기반을 둘 뿐만 아니라, 이에 더해 희소앙상블 코딩을 적용한다. 희소앙상블 코딩은 뉴럴 코딩에서 주로 나타나며, 여러 개체(뉴런, 기계)를 통해 어떤 작업(반응, 예측)을 하되, 개체 중 일부를 통하여 작업을 하는 방법을 일컫는다[4-7]. 희소앙상블 코딩은 실제 뇌에서 일어나는 현상이다. 또한 여러 개체로부터 시너지를 얻을 수 있고, 개체가 선택적으로 쓰이므로 전체 시스템이 효율적으로 운용될 수 있다. 따라서 희소앙상블 코딩은 음악 회상을 위해서도 반드시 적용되어야만 한다. 이 논문은 여러 가변 길이의 서열들이 동시에 예측에 참여하도록 하고, 또한 그 서열들 중 일부를 취함으로써 희소앙상블 코딩 형태의 알고리즘을 구현한다. 이를 통해 여러 서열로 얻는 시너지로 예측 성능을 높일 수 있다. 또한 서열의 일부를 취하지 않는다면 기억 용량의 제한으로 인해 구현이 불가능할 것이나, 희소앙상블 코딩을 적용함으로써 예측이 가능하게 된다.

이 논문은 희소앙상블 코딩에 기반을 둔 순차적 음악 회상 학습 알고리즘을 소개한다. 그리고 실험적으로, 팝송의 MIDI 파일을 순차적으로 학습하여, 제안하는 모델의 예측 성능을 보이고, 그 구조에 대해 분석한다.

2. 희소앙상블 코딩 기반 음악 회상 메모리

2.1 순차적 음표 예측

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

$$\hat{s}_i = \operatorname{argmax}_{s \in S} p(s|s_1^{i-1})$$

를 고를 수 있다.

하지만 과거의 모든 음표를 고려한 확률 $p(s|s_1^{i-1})$ 을 계산하기보다는 최근 $n-1$ 개의 음표가 주어졌을 때의 조건부 확률 $p(s|s_{i-n+1}^{i-1})$ 을 계산하고 이를 $p(s|s_1^{i-1})$ 의 근사값으로 사용하는 방법이 더욱 효율적이다. 이때 i 와 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_i 를 다음과 같이 예측할 수 있다.

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

이러한 설정은 가변차수 마코프 모델 [2], n-그램 모델 [3], 하이퍼네트워크(hypernetwork) 모델 [8-9] 등에 이용되어 왔고, 이 논문에서도 동일한 설정 하에서 기존과 알고리즘을 보완한 희소앙상블 코딩 기반의 예측 알고리즘을 제안한다.

2.2 희소앙상블 코딩 기반 알고리즘

이 논문은 특히 [6]에서 소개된 희소앙상블 코딩을 이용하여, 고정된 n 대신 가변 길이의 음표 서열을 통해 s_i 를 예측한다. 모든 n 길이의 서열의 집합은 음표 집합 S 의 데카르트 n -제곱(n -th Cartesian power)

$$S^n = \{(g_1, \dots, g_n) | g_i \in S, i = 1, \dots, n\}$$

과 같다. 이때 길이가 2인 서열부터 최대 길이가 N 인 서열까지 고려하여 집합 $\Sigma = \bigcup_{i=2}^N S^i$ 를 정의할 수 있다.

$\sigma \in \Sigma$ 의 길이가 n 일 때, 앞의 $n-1$ 개의 성분을 $\sigma_{\text{hist}} = \sigma_1^{n-1}$, 뒤의 1 개의 성분을 $\sigma_{\text{pred}} = \sigma_n$ 이라 하고, σ 를 σ_{hist} 와 σ_{pred} 의 연결 $\sigma_{\text{hist}} : \sigma_{\text{pred}}$ 로 나타낼 수 있다. 관찰된 음표 서열 s_1^{i-1} 의 문맥 집합

$$C(s_1^{i-1}) = \{s_{i-n+1}^{i-1} | s_i \in S, n = 2, \dots, N\}$$

에 하나의 음표 s 를 더 연결한 집합을

$$D(s_1^{i-1}, s) = \{c : s | c \in C(s_1^{i-1})\} \subset \Sigma$$

라 할 때, s_i 를 다음과 같이 예측할 수 있다.

$$\hat{s}_i = \operatorname{argmax}_{s \in S} \sum_{\sigma \in D(s_1^{i-1}, s)} w_\sigma p(\sigma_{\text{pred}} | \sigma_{\text{hist}}), \quad (2)$$

여기에서 w_σ 는 σ 에 대한 가중치이다. (1)에서는 하나의 고정된 길이의 서열로부터 확률을 계산하여 예측을 하는 반면, (2)에서는 여러 길이의 서열로부터 개별적으로 확률을 계산하고, 이들의 가중합을 통해 예측을 하게 된다. 즉, 앙상블 코딩을 적용하는 것이다.

이때 집합 $D(s_1^{i-1}, s) \subset \Sigma$ 에 속하는 서열 σ 에 대한 가중치 w_σ 를 모두 이용하기 위해서는 Σ 의 크기

$$|\Sigma| = |S|^2 + |S|^3 + \dots + |S|^N = \frac{(|S|^{N+1} - |S|^2)}{|S| - 1}$$

만큼의 저장 공간을 필요로 한다. 이는 대략 $|S|^N$ 정도로 매우 크다. 하지만 희소앙상블 코딩을 적용하여 이보다 적은 정보로도 s_i 를 예측할 수 있다. 즉, 고정된 크기의 집합 $T \subset \Sigma$ 에 서열을 저장하고, $D(s_1^{i-1}, s)$ 대신

$$D'(s_1^{i-1}, s) = T \cap D(s_1^{i-1}, s)$$

을 이용하여 s_i 를 다음과 같이 예측할 수 있다.

$$\hat{s}_i = \operatorname{argmax}_{s \in S} \sum_{\sigma \in D'(s_1^{i-1}, s)} w_\sigma p(\sigma_{\text{pred}} | \sigma_{\text{hist}}) \quad (3)$$

이때 관찰된 적이 없거나 적은 서열 σ 를 다루기 위해 PPM-C([2]) 방식으로 계산한다.

전체적인 예측 알고리즘은 다음과 같다. 먼저, MIDI 파일의 다양한 길이의 연속된 음표들의 서열 σ 을 통해 가변 길이의 마이크로코드를 구성한다. 문맥 $C(s_1^{i-1})$ 이란 최근의 관찰된 음표 서열을 의미하는데, 마이크로코드 중 바로 다음의 예측하려는 음표를 제외한 나머지 음표들이 문맥과 일치하는 마이크로코드 $D(s_1^{i-1}, s)$ 들이 예측에 참여한다. 이때 모든 마이크로코드들이 예측에 참여하는 것이 아니라 선택된 일부의 마이크로코드 $D'(s_1^{i-1}, s)$ 을 통해 예측을 하게 된다.

$l(\sigma)$ 는 σ 의 길이, $I(p)$ 는 명제 p 가 참이면 1, 거짓이면 0이라고 하자. 학습 알고리즘의 의사코드는 다음과 같다.

초기 모델 생성:

- 서열로부터 유도되는 가변 길이의 음표 서열 집합 Σ 중에서 일부 T 를 임의로 선택
- T 내의 서열 σ 에 초기 가중치 w_σ 할당

국부적 학습과 전역적 학습 반복:

국부적 학습:

- 문맥 $C(s_1^{i-1})$ 초기화
- 문맥에 맞는 서열을 T 에서 찾아 식 (3)을 통해 예측
- 예측에 활용된 서열 $\sigma \in D'(s_1^{i-1}, \hat{s}_i)$ 에 대해 원곡과 비교하여 가중치 $w_\sigma := w_\sigma + (-1)^{I(s_i \neq \hat{s}_i)} l(\sigma)$ 갱신

전역적 학습:

- 가중치 w_σ ($\sigma \in T$) 정규화
- 가중치 w_σ 가 낮은 서열 $\sigma \in T$ 들을 기존에 선택되지 않았던 $\Sigma \setminus T$ 내의 서열들로 대체

3. 실험 및 결과

제안된 알고리즘을 통해 순차적으로 음악을 학습하고 회상하는 실험과 그 결과를 제시한다. 실험에 앞서, 단순한 곡에 비해 분석이 어려운 팝 그룹 ABBA의 4/4 박자의 곡 중 앨범 별로 골고루 30곡을 선별하여, C장조(A단조)로 변환하였다. 이 곡들은 모두 MIDI 파일로 저장되어 있는데, 이를 통해 음표를 추출해낼 수 있다. 이 곡들의 음표를 포함한 음표의 개수는 12,647 개이고, 음표의 종류의 수는 513 개이다. 각 곡을 학습할 때, 곡의 전반부는 훈련 집합으로, 후반부는 시험 집합으로 사용되었다. 이 곡들을 한꺼번에 학습하지 않고, 순차적으로 학습하였다.

적중률은 훈련 집합에 대해 0.95, 시험 집합에 대해 0.6 정도로 나타난다(그림 1). 과적합(overfitting)이 나타나면서 서서히 T 내에 길이가 긴 서열이 많이 저장됨을 알 수 있다(그림 2). 마이크로코드의 최대 차수는 9 이하로도 충분한 적중률을 갖는 것으로 보인다(그림 3).

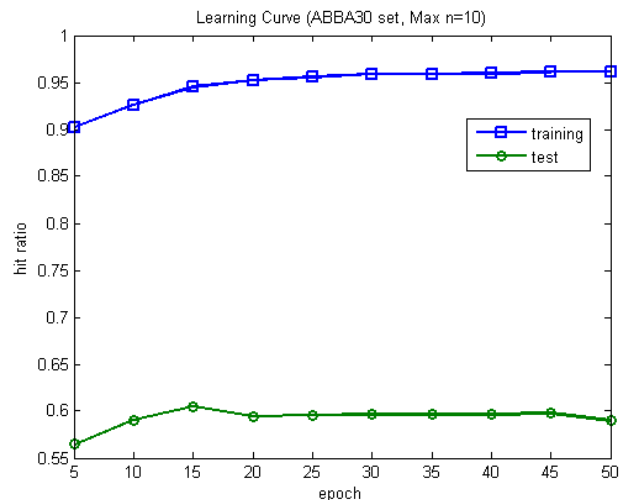


그림 1 학습데이터와 테스트데이터에 대한 학습 곡선(Hit ratio vs. epoch.). epoch=15 이후 과적합(overfitting)이 발생함을 볼 수 있다.

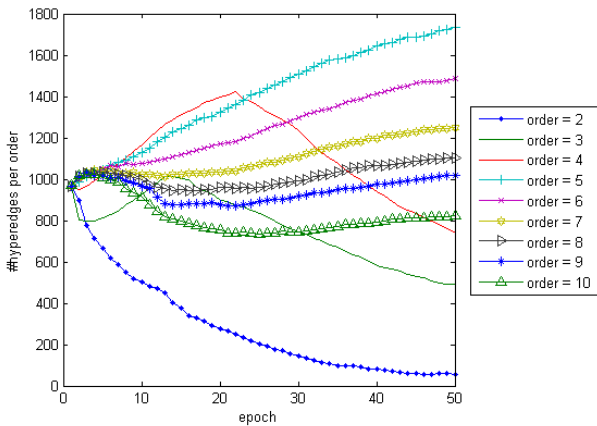


그림 2 반복학습 과정에서 희소양상블 코딩 기반 음악 회상 학습 모델의 구조의 변화(number of hyperedges vs. epoch). 과적합이 발생하는 시점 이후 차수가 큰 마이크로코드 비중이 급격히 증가함을 확인할 수 있다.

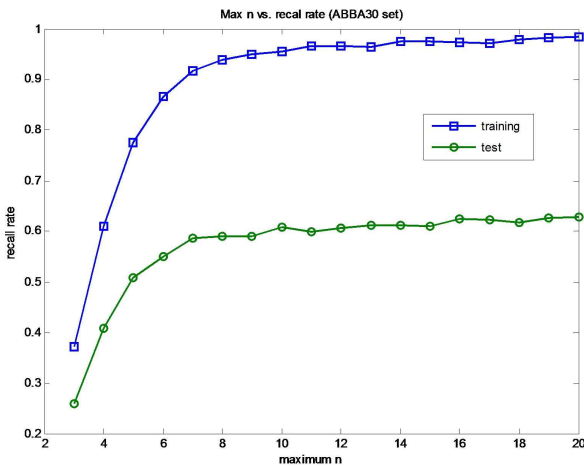


그림 3 회상 적응률(recall rate)과 최대 차수(maximum order) 간의 관계. 최대 차수 10 이후로는 큰 성능의 변화가 없다.

4. 논의

n-그램 모델은 서열을 짧은 조각으로 나누고, 저장된 n-그램의 통계에 근거하여 주어진 문맥에 대해 예측한다. 이때 문맥에 일치하는 조각의 연장에 대한 확률 분포에 근거하여 예측하게 된다. n-그램 모델에 대해 자주 쓰이는 기법으로서 관찰된 적이 없거나 적은 문맥을 다루는 영탈출(zero-escape) 기법, 다양한 길이의 문맥으로부터의 예측을 결합하는 평탄화(smoothing) 기법이 있다[3].

마코프 모델은 상태 간의 전이 확률을 나타낸다. 마코프 모델에서 상태가 전이되는 확률은 이전의 입력에 대해 독립이고, 현재의 입력과 상태에 대해 종속이다. 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].

이 논문은 n-그램과 비슷한 마이크로코드[6-7]를 통해 예측을 하되, 희소양상블 코딩을 적용하여 예측 성능을

높이고 효율성을 높이게 되었다.

5. 결론 및 향후 연구

이 논문은 단서 멜로디가 주어졌을 때 이어서 나오는 음표를 차례로 예측하는 멜로디 회상 문제에 접근하였다. 순차적 음악 회상 알고리즘을 구현하기 위해 일부의 가변 서열만이 저장되고 여러 서열을 통해 예측하는 희소양상블 코딩을 사용하였다. 팝송 멜로디에 대해 학습된 정보는 고정된 크기의 집합 T 에 대해 저장된다.

이후 가중치 갱신 규칙은 지금보다 더 개선될 여지가 있을 것으로 예상된다. 장기적으로는 음악을 자동으로 계속 학습하면서, 회상 또는 생성할 수 있는 인공지능 머신을 구현하는 것을 목표로 한다.

감사의글

이 논문은 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구이며(NRF-2010-0017734-Videome), 정부(미래창조과학부 및 정보통신기술진흥센터)의 정보통신·방송 연구개발사업 지원(10035348-mLife, 14-824-09-014, 10044009-HRI.MESSI)을 일부 받았음.

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.