강화 학습을 통한 자동 반주 생성

Automatic Generation of Music Accompaniment Using Reinforcement Learning

김나리 Nari Kim*, 권지용 Ji-Yong, Kwon**, 유민준 Min-Joon, Yoo***. 이인권 In-Kwon Lee****

요약

본 연구에서는 사용자가 입력한 멜로디에 따른 반주 음악을 자동으로 생성하는 방법을 제시한다. 시작되는 코드는 사용자의 멜로디에 의해서 생성이 되며, 그 다음 코드들은 코드들간의 전이확률이 정의되어있는 마르코프 체인(markov chain)의 확률 테이블을 이용하여 연속적으로 생성된다. 확률 테이블은 기존 음악의 샘플 데이터를 강화학습(reinforcement learning)을 이용하여 학습된다. 또한 실시간으로 재생되는 반주 코드는 매 상태 마다 주어지는 보상 값을 통해 더 나은 행동을 취할 수 있도록 학습해 나간다. 멜로디와 각 코드들간의 유사성은 피치 클래스 히스토그램을 이용하여 계산된다. 본 기술을 사용하여 주어진 사용자 입력에 조화로운 반주 코드의 자동 생성이 가능하다.

Abstract

In this paper, we introduce a method for automatically generating accompaniment music, according to user' s input melody. The initial accompaniment chord is generated by analyzing user's input melody. Then next chords are generated continuously based on markov chain probability table in which transition probabilities of each chord are defined. The probability table is learned according to reinforcement learning mechanism using sample data of existing music. Also during playing accompaniment, the probability table is learned and refined using reward values obtained in each status to improve the behavior of playing the chord in real-time. The similarity between user's input melody and each chord is calculated using pitch class histogram. Using our method, accompaniment chords harmonized with user's melody can be generated automatically in real-time.

핵심어: 자동 반주 생성, 강화 학습, 마르코프 체인, 전이 확률, 몬테카를로 학습.

Keyword : Automatic accompaniment generation, Reinforcement learning, Markov chain, transition probability, Monte Carlo learning.

본 논문은 문화관광부 및 한국문화콘텐츠진흥원의 문화콘텐츠기술연구소(CT) 육성사업의 연구결과로 수행되었음.

*주저자 : 연세대학교 컴퓨터과학과 석사과정 e-mail: wassupnari@cs.yonsei.ac.kr

**공동저자 : 연세대학교 컴퓨터과학과 박사과정 e-mail: mage@cs.yonsei.ac.kr

***공동저자 : 연세대학교 컴퓨터과학과 박사과정 e-mail: debussy@cs.yonsei.ac.kr

****교신저자 : 연세대학교 컴퓨터과학과 부교수 e-mail: iklee@yonsei.ac.kr

1. 서론

과학 기술의 발전과 함께, 컴퓨터를 이용한 음악의 창작 및 연주 분야에 많은 발전을 거듭해 왔다. 기존의 전통적인 악기만을 사용하던 데서 벗어나 최근에는 미디 장비 혹은 컴퓨터 음원을 이용한 작곡, 편곡 및 시퀀싱 등이 가능해 졌고, 더욱 적은 비용으로도 효율적으로 음악을 작곡하고 연주할 수 있는 시대가 되었다. 이에 더하여 수학적 논리 계산을 기반으로 한 음악의 작곡법 또한 제안되고 있다 [1, 2, 3].

본 논문에선 전통적 작곡법과 수학적 알고리즘을 토대로 한 새로운 반주 음악을 생성하는 방법을 제안하고자 한다. 음악 작곡 분야에 다양한 수학적 알고리즘이 활용되어왔다. 예를 들어 음악 작곡에 확률모델이나 마르코프 체인(markov chain), 형식 문법(formal grammar), 유한 상태 기계(finite state machine), 프랙탈(fractal)등의 기법이 사용될 수 있다 [1,3]. 하지만 이러한 수학적 모델만으로 작곡을 할 경우 대중들에게 친숙하지 않은, 음악적으로만 의미 있는 음악이 생성되는 경우가 많다. 본 논문에서는 일반적인 사람들을 위한 반주 음악을 생성하는 것을 목표로 한다. 따라서 우리는 전통적인 작곡법에 기반 하여 작곡된 기존의 음악들을 분석하여 이것을 수학적 모델로 구성하여 사용하는 방법을 사용한다.

우리는 머신러닝의 한 분야인 강화학습(reinforcement learning)[4,5] 방법을 사용한 새로운 반주 음악 작곡법을 제안하고자 한다. 강화학습은 어떤 환경을 탐색하는 에이전트가 있고 이 에이전트에게 앞으로 누적될 보상이 최대화 되는 일련의 행동으로 정의되는 정책을 찾아 나가게 하는 학습 방법이다. 강화학습을 기반으로 한 실시간 음악작곡을 위해 에이전트는 현재 음악의 진행되는 상태를 인식하고 그에 맞는 반주 코드를 생성하여야 한다. 그리고 생성된 반주 음악이 다음 상태에서 얼마나 적합할지를 판단하여 적절한 보상을 얻게 된다. 여기서 시스템은 주어진 보상 값을 통하여 각 상태에서 취한 행동이 적합 한지 아닌지를 학습하며 발전해 나간다.

마르코프 체인은 알고리즘 작곡 분야에서 자주 사용되어 온 방법 중에 하나이다. 하지만 대부분의 경우 각 음들의 전이확률을 정의하여 멜로디를 생성할 때 사용하고 있다[6]. 본 논문에서는 이와는 달리 코드들 간의 전이확률이 정의되어있는 마르코프 확률 테이블을 이용하게 된다. 최종적으로 학습된 모델에 기반을 두어 사용자 입력에 최적화된 반주음악의 생성이 가능하다.

2. 강화 학습 기반 모델

본 논문의 강화 학습 모델은 그림 1과 같이 구성된다. 크게 에이전트(agent)와 환경(environment), 정책(policy)이 있고 상태(state), 행동(action), 보상(reward)의 집합들로 이루어져 있다.

그림 1. 에이전트는 자신에게 주어질 보상이 최대화 되는 정책(π)을 찾아 가며 각 상태에서 특정 행동을 취하게 된다. 우리는 여기서 사용자로부터 입력에 맞게 생성 될 코드를 현재 상태(S)로 인식하고 높은 보상(R)이 기대되는 반주 코드를 행동(A)으로 정의한다.

2.1 상태-행동 모델

상태(S)와 행동(A) 모델은 그림 1과 같이 구성될 수 있다. 모델의 각 집합들을 간략히 정리하면 다음과 같다.

1. 환경에 속한 상태 집합 : $s_t \in S$
2. 행동 집합 : $a \in A(s_t)$
3. 보상 집합 : $R = r_0 + r_1 + \cdots + r_n$

하나의 마디를 위해 생성되는 한 마디의 반주 코드를 상태(S)로 분류하였다. 그리고 현재의 상태를 분석하여 다음 상태에서 나타날 수 있는 멜로디를 예측하고 그에 맞는 반주 코드를 생성하게 되는데, 이때 생성되는 한 마디의 반주 코드를 행동(A)으로 정의한다.

우리는 여기서 생성되는 반주 코드가 다음의 상태로 예상되는 멜로디와 잘 어울리는지를 판단하여 그에 합당한 보상(R)을 할당하였다. 반주 코드를 생성하는 부분에 있어서는 직접 음을 생성하는 방법도 있겠지만 여기서는 여러 샘플 데이터로부터 입력 받은 코드들을 학습하여 마르코프 체인을 이용한 확률 테이블로 저장한 후, 이를 이용하는 방법을 사용하였다. 멜로디와 반주 코드가 어울리는 정도를 평가하는 부분은 4장에서 설명된다.

2.2 보상-정책

강화 학습 모델은 궁극적으로 보상(R)을 최대화 하는 정책(π)을 찾아 가는 방향으로 학습해 나간다. 각 상태마다 취할 행동을 정의하는 정책과 그 행동에 대한 보상은 다음과 같은 식으로 정의된다.

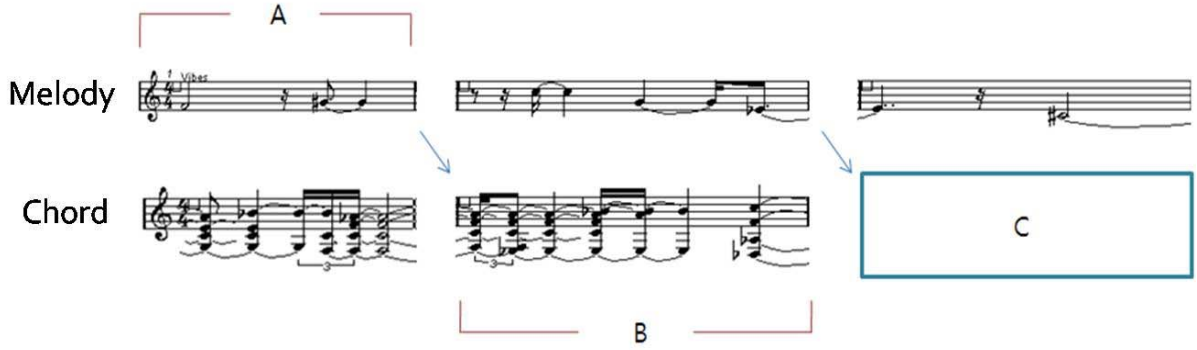

그림 2 A와 같은 초기 상태에서 멜로디 코드를 분석하여 다음 상태에서 나타날 만한 반주 코드를 B와 같이 미리 예상한다. 새로운 멜로디의 상태로 전이가 되면 새롭게 생성된 멜로디와 B에서 생성된 반주코드가 얼마나 잘 어울리는지를 분석하여 그에 합당한 보상을 해 준다. 그리고 다시 현재 상태를 분석하여 다음의 코드 C를 생성하여 그때의 멜로디와의 조화를 판단한다.

$$R = \sum_{t} \gamma^{t} r_{t} \quad (0 \le \gamma \le 1), \tag{1}$$

$$\pi : S \rightarrow A, \tag{2}$$

여기서 R은 보상 집합 원소들의 합을 나타내는데 미래의 보상 값을 세기위한 요소로 γ를 사용한다. π는 현재 상태에서 취하는 특정 행동으로의 맵핑을 나타낸다.

정책(π)은 현재의 상태에서 다음 상태로 전이 될 때에 높은 확률 값을 갖는 행동을 취하는 것으로 정의하였다. 마르코프 체인을 이용한 look-up 테이블의 확률 값을 고려하여 다음의 행동을 선택하게 되는데 이때 취한 행동이 다음의 상태와 어울리지 않을 때에는 보상되는 값을 이용하여 확률 값을 조절한다.

3. 상태 전이

본 연구에서는 마르코프 확률 테이블을 이용하여 현재의 상태를 분석하고 다음 마디에 생성될 코드를 결정한다. 초기 상태에서는 사용자의 입력으로 부터 들어온 멜로디 마디를 분석하여 그 멜로디와 어울리는 코드를 생성한다. 즉 사용자의 멜로디를 바탕으로 하여 첫 마디만을 고려한 한 마디의 코드를 생성한다. 생성된 첫 번째 코드를 바탕으로 다음 코드를 연속적으로 생성하며 이때 다음으로 생성될 코드는 멜로디에 의존하지 않고 확률테이블을 통해 생성된다(그림 2). 멜로디를 고려하지 않고 코드만 고려해서 다음 상태를 결정하게 되므로 상태의 전이를 나타내는 테이블의 확률 값을 정의하기 위한 몇 가지 음악적 규칙이 필요하다.

먼저, 코드를 토닉, 도미넌트, 서브도미넌트의 세 가지 코드 군으로 나누었을 때 그림 3에서와 같이 코드의 진행 형태를 정의할 수 있다[7]. 즉 일반적으로 토닉은 도미넌트나 서브도미넌트로 진행하고, 서브 도미넌트는 토닉이나 도미넌트로 진행하며, 도미넌트는 토닉으로 진행한다. 반드시 다른 군으로 진행해야 하는 것은 아니며 자신의 군으로도 진행이 가능하고 도미넌트에서는 서브도미넌트로의 진행이 그리 흔하지 않다.

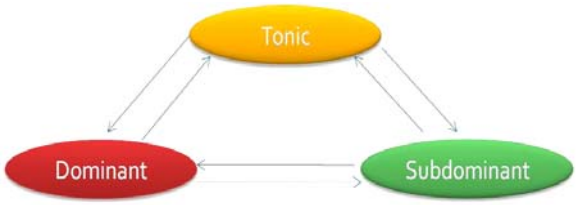

그림 3. 토닉, 서브도미넌트 그리고 도미넌트 간의 일반적인 관계. 토닉-서브도미넌트, 토닉-도미넌트 서로 간의 이동은 자연스럽지만, 도미넌트에서 서브도미넌트은 일반적으로는 사용되지 않는다. 서브도미넌트에서 도미넌트로의 이동은 매우 일반적이다.

본 논문에서 생성되는 반주 코드는 C key의 7개의 다이아토닉 코드(C, Dm, Em, F, G, Am, Bm^{-5})이며 입력되는 멜로디도 C Key라고 가정한다. 일반적으로 C key의 다이아토닉 코드들은 다음과 같이 3가지 코드 군으로 분류된다[8].

- Tonic family : C, Em, and Am
- Subdominant family : F and Dm
- Dominant family : G and Bm^{-5}

위와 같은 음악적 규칙들을 이용하여 코드의 진행 상태를 마르코프 확률 테이블로 나타내었다. 초기 확률 테이블은 표 1과 같은 형태이다.

	Am	Bm^{-5}	C	Dm	Em	F	G
Am	0.142	0.142	0.142	0.142	0.142	0.142	0.142
Bm^{-5}	0.18	0.18	0.18	0.05	0.18	0.18	0.05
C	0.142	0.142	0.142	0.142	0.142	0.142	0.142
Dm	0.142	0.142	0.142	0.142	0.142	0.142	0.142
Em	0.142	0.142	0.142	0.142	0.142	0.142	0.142
F	0.142	0.142	0.142	0.142	0.142	0.142	0.142
G	0.18	0.18	0.18	0.05	0.18	0.18	0.05

표 1 코드 군의 진행 규칙에 따라 정의된 초기에 주어진 확률 테이블

현재의 상태에서 전이 될 다음 상태는 코드의 진행 규칙에 따른 각각의 확률 값으로 나타내며, 몬테카를로법(Monte Carlo method)[9]을 사용하여 상태 변환이 학습된다. 즉 매 상태에서 생성되는 새로운 코드가 멜로디와 얼마나 잘 어울리느냐를 분석하여 어울리는 정도에 따라 적당한 보상 값을 할당한다. 할당 된 보상 값은 다시 확률 테이블의 값에 반영되며, 각 행의 확률 값의 합은 1이 되어야 하므로 해당되는 행의 모든 확률 값을 정규화한다. 표 2는 한곡의 입력 멜로디를 통해 3번 정도 학습시킨 후 변화된 확률 테이블의 예이다.

	Am	Bm^{-5}	C	Dm	Em	F	G
Am	0.133	0.200	0.133	0.133	0.133	0.133	0.133
Bm^{-5}	0.166	0.242	0.166	0.046	0.166	0.166	0.046
C	0.141	0.151	0.141	0.141	0.141	0.141	0.141
Dm	0.108	0.148	0.148	0.148	0.148	0.148	0.148
Em	0.117	0.163	0.143	0.143	0.143	0.143	0.143
F	0.131	0.144	0.144	0.144	0.144	0.144	0.144
G	0.111	0.203	0.192	0.053	0.192	0.192	0.053

표 2 강화 학습을 통해 변화된 확률 테이블의 예

4. 멜로디와 반주 코드의 조화도 분석

사용자로부터 입력으로 들어온 멜로디와 강화 학습을 통해 얻어진 반주 코드와의 어울림의 정도를 음악적으로 정확히 정의하는 데에는 어베일러블 노트(available note)등의 모드(mode)이론 등 여러 가지 음악적 이해가 필요하다. 하지만 본 논문에서는 몇 가지 제한을 두고 기본적인 반주 코드를 생성하는 것에 초점을 둔다. 예를 들어 코드 네임이 C일 경우 음악의 장르나 연주자에 따라서 종류의 코드들(예를 들어, 재즈와 같은 느낌으로 연주하고 싶다면 C_6나 CM_7등)이 이용되는데 이와 같은 장르를 고려하는 부분은 본 논문에서 고려하지 않고 간략화 하였다.

멜로디와 반주 코드를 잘 조화시키는 가장 기본적이면서 효율적인 방법은 멜로디에 존재하는 음을 이용하여 코드를 생성하는 것이다. 예를 들어 '도', '미', '솔'로 구성된 멜로디가 있다면 C코드가 가장 적합할 것이다. 구체적으로 멜로디를 구성하는 음들의 분석은 히스토그램을 이용한다. 한 마디의 멜로디를 구성하는 음들이 있을 때, 이 음들을 각 반음계의 12개의 음(정확히 정의하면 12개의 피치 클래스(picth class))에 속하는 빈도수를 계산하여 생성되는 히스토그램을 정의한 후, 이를 각 코드의 히스토그램과 비교하여 얼마나 잘 조화되는지 분석해 낸다. 수치적으로 나타내면 식 1과 같다.

$$\min_{k} \arg\left(\sum_{i=1}^{12} (M(i) - C_k(i))^2 \right) \qquad \text{(식 1)}$$

$M(i)$는 멜로디의 히스토그램에서 i번째 피치 클래스의 빈도수. $C_k(i)$는 k번째 코드의 히스토그램에서 i번째 피치 클래스의 빈도수. 본 논문에서 C_k는 C코드의 7가지 다이어토닉 코드를 의미한다.

본 논문에서 제안한 방법으로 반주 코드를 생성할 때, 식 1을 이용하여 구해진 최적의 코드만을 이용하면 같은 코드가 자주 반복되어 지루해 질 수가 있으므로, 생성된 코드의 코드 군을 고려하여 이 코드 군안에서 교환이 이루어질 수 있도록 할 수도 있다. 또한 이들의 대리코드를 사용하는 것도 멜로디와 잘 조화되면서 다양한 느낌의 반주를 생성할 수 있는 방법 중 하나이다.

5. 결과

본 논문에서 제안한 방법으로 실험을 해본 결과 일반적으로 입력 멜로디가 반복적이고 비슷한 느낌으로 흘러갈 때 학습 효과가 높은 것을 확인 할 수 있었다. 하지만 멜로디가 계속적으로 변화하거나 음악의 분위기가 크게 달라지는 부분에 있어서는 좋은 결과를 얻지 못하였다.

우리의 방법으로 얻은 코드 음악 파일과 원본 음악의 코드를 비교해 본 결과 약 50%정도 원본과 같아지는 것을 확인하였다. 향후 연구 과제로 제안한 여러 사항들을 보완하면 더 좋은 결과를 얻을 수 있을 것이라 기대된다.

6. 결론 및 향후 연구 과제

본 연구에서는 강화 학습을 기반으로 한 자동 반주를 생성하는 방법을 소개하였다. 본 기술을 이용하면 실시간으로 멜로디와 잘 어울리는 코드를 자동으로 생성 할 수 있다.

입력 멜로디를 기반으로 자동으로 생성되는 코드는 상태의 변화에 있어서 이전의 코드만을 고려했기 때문에 많은 제약 사항들이 있었다. 또한 상태의 전환과 멜로디와의 조화를 분석하는 데에 한 마디의 단위를 사용하였는데 실제로 코드는 한 마디 안에서 여러 번 변화할 수 있기 때문에 반

마디 정도의 단위로 멜로디와 코드를 고려한다면 더 좋은 결과를 기대 할 수 있을 것이다.

본 연구에서는 멜로디를 기반으로 그와 어울리는 코드를 생성하였다. 반대로 코드를 기반으로 다양한 멜로디를 생성하는 시스템도 가능할 수 있다. 이처럼 코드와 멜로디를 각각 자동으로 생성하고, 이들을 잘 조화롭게 엮을 수 있다면 새로운 곡을 작곡하는 시스템도 기대 할 수 있을 것이다.

참고문헌

[1] Iannis Xenakis, Formalized Music – Thought and Mathematics in Music, Pendragon Press 1970.

[2] Min-Joon, Yoo., In-Kwon, Lee., and Jung-Ju, Choi., Background Music Generation Using Music Texture Synthesis. In Proceedings of International Conference on Entertainment Computing, 2004. p. 565-570.

[3] Karlheinz Essl: Algorithmic Composition. in: Cambridge Companion to Electronic Music, ed. by N. Collins and J. d'Escrivan, Cambridge University Press, 2007.

[4] L.P.Kaelbling, M.L.Littman, and A.W.Moore. Reinforcement learning: A survey. Journal of Artificial Intelligence Research, 1996.

[5] R.S.Sutton and A.G.Barto. Reinforcement Learning: An Introduction. MIT Press, 1998.

[6] Curtis Roads. The Computer Music Tutorial. MIT Press, 1996.

[7] 길옥윤 : 알기쉬운 경음악 편곡법. 세광음악출판사, 1993.

[8] Levine, K. : The Jazz Theory Book. Sher Music Co. 1996.

[9] R.Y. Rubinstein and D.P. Kroese. Simulation and the Monte Carlo Method (second edition). New York: John Wiley & Sons, 2007,