인공지능을 활용한 드럼 연습 시스템

박현묵*, 박병용*, 최성규*, 김정민**
*단국대학교 소프트웨어학과
**KT

A Study on Drum Training Using Artificial Intelligence System

Byeong-yong Park*, Hyeon-Mook Park* ,Seong-Gyu Choi*, Jeong-Min Kim**
*Dept. of Software Engineering, Dan-kook University
**KT

요 약

이 논문에서는 인공지능 기반의 드럼 연습 시스템을 설명하고 있다. 먼저 사용자의 드럼 신호를 MIDI 파일로 변환하고, 이를 DB 에 저장되어있던 추천곡의 MIDI 파일과 비교하여 가장 유사한 것을 추천해준다. 또한 실시간으로 사용자의 드럼 연습을 도와주는 전문가 시스템역할을 함으로써 연주의 숙련도를 높여준다.

1. 서론

한국은 급격한 경제성장을 이룬 후 삶의 질이 급격히 향상되었다. 그로 인해 대한민국은 문화 강국이 되었고 대한민국 국민들의 엔터테인먼트에 대한 욕구는 점점 증가하고 있는 추세이다. 그중 많은 사람들은 취미로 악기를 연주하고 있거나 혹은 그러고 싶은 사람들이 많을 것이다.

또한, 2016 년 12 월 'AI 활용한 로봇 음악 산업이 뜬다'라는 기사가 발행되었다. 여러 가지 문제로 음악 산업이 맞이한 위기를 극복하는 해결책으로 인공지능 기술을 도입한 것이다. 음악 산업에서 인공지능의 역할은 이용자들 간의 데이터 비교, 이용자의 상황이나 성향에 맞는 새로운 컨텐츠 추천 등이다. 여기서 우리가 주목한 것은 데이터 비교와 새로운 컨텐츠 추천이다.

우리는 수 많은 악기 중에서 접근성, 경제성, 휴대성을 고려했을 때 효율적이라고 판단한 드럼 연습기를 선택했다. 물론 이미 드럼 연주에 대한 점수 표시 기능을 탑재한 제품이 시중에 있고, 태블릿을 사용한 드럼 연주 앱 또한 시중에 출시 되어있다. 그러나 우리는 실시간 채점기능과 사용자 연주에 맞는 곡 추천 기능을 인공지능을 통해 추가 구현함으로써 사용자가 드럼연습기와 의사소통을 하고있는 듯한 체험을 제공하도록 할 것이다.

2. 전문가 시스템

전문가 시스템(Experts System)은 생성 시스템(Production system)의 하나로서, 인공지능 기술의 응용분야 중에서 가장 활발하게 응용되고 있는 분야이다. 즉, 인간이 특정분야에 대하여 가지고 있는 전문적인 지식을 컴퓨터에 기억시킴으로써 전문가와 같은 판단, 추론을 컴퓨터가 수행하여 비전문가인 일반인도 전문지식을 활용할 수 있도록 하는 시스템이다.

생성 시스템의 하나인 전문가 시스템의 구조는 지식 베이스, 추론 기관, 사용자 인터페이스, 작업 메모리, 설명 부 시스템, 지식 습득 부시스템으로 구성되어 있다. 지식 베이스는 문제 해결에 대한 전문가의 지식이 포함된 부분이다. 추론 기관은 지식베이스에 있는 규칙들을 탐색하고, 추론과 통제를 위해 규칙을 선택하는 중요한 부분이다. 사용자 인터페이스는 사용자가 시스템을 사용할 때 접하게 될 부분이다. 작업 메모리는 사용자의 기록과 추론 과정에서 얻은 결과를 일시적으로 저장하는 부분이다. 설명 부 시스템은 사용자에게 전문가 시스템의 행동에 대해서 설명하는 부분이다. 지식 습득 부 시스템은 새로운 지식을 추가하거나 기존 지식을 수정할 수 있도록 지원하는 부분이다.

전문가 시스템의 적용 사례로 미국의 통신회사 GTE 사의 전화교환기의 오류 발견 및 유지보수를 지원하는 COMPASS, 벌링턴 노던 철도 회사의 철도 고장의 해결사 SMP, 제너럴 모터스 자동차회사의 기술자를 능가하는 찰리 시스템, 맥도널 더글러스 항공기 회사의 항공기 유지 보수 시스템 FLEAT 등이 있다.

3. MIDI

3.1 MIDI 와 mp3 의 특징

음악을 재생할 수 있는 파일의 형식 중 MP3(MPEG Audio Layer-3)와 MIDI(Musical Instrument Digital Interface)가 있다. MP3 는 MPEG-1 의 오디오 규격으로 개발된 손실압축 포맷이며, 소리나 음악이 녹음된 파일의 형식이다. MIDI

는 전자 악기 간 디지털 신호를 주고 받기 위해 각 신호를 규칙화한 일종의 규약이며, 소리에 대한 정보(음의 크기, 악기 종류 등)가 기록된 파일의 형식이다. 간단히 표현하면 MP3 는 파형이고, MIDI 는 신호라고 할 수 있다.

본 시스템에서는 녹음이 아닌 진동에 의해 전달되는 신호를 처리하여 응용하기 때문에 MIDI 파일 형식을 사용하기로 한다.

3.2. MIDI text 형식

미디 파일은 파일 정보를 기록한 헤더와 파일 내용을 기록한 트랙으로 구성되어 있다. 트랙은 MIDI, SysEx, Meta 의 3 가지 event 로 구성되어 있다. MIDI 는 재생 위치에 해당하는 정수 값, 재생 출력 상태, 악기에 대한 채널, 음표 종류인 note, 음의 세기 등을 포함한다. SysEx 는 임의의 내용을 넣을 수 있는 부분이다. Meta 는 저작자 등의 추가 정보를 포함한다.

다음은 이해를 돕기 위해 미디 텍스트 파일의 예시를 분석한 내용이다.

```
MFile 0 1 120
```

위 부분은 헤더 부분이다. MFile 은 미디 파일이라는 것을 표시하는 부분이며, 순서대로 트랙 개수와 관련된 포맷, 트랙 개수, 사용할 시간 간격(단위 시간)과 관련된 값을 나타내는 부분이다.

```
MTrk
0 Meta TrkName "rudi2"
0 Tempo 500000
0 TimeSig 4/4 24 8
0 Meta TrkEnd
TrkEnd
MTrk
0 Meta TrkName "Addictive Drums 2 01"
```

위 부분은 트랙 부분이다. MTrk 는 트랙 부분의 시작을 표시하는 부분이며, 위 표시된 부분은 Meta event 에 해당한다.

```
0 On ch=1 n=38 v=100
120 Off ch=1 n=38 v=64
480 On ch=1 n=40 v=100
600 Off ch=1 n=40 v=64
…
```

위 부분은 트랙 부분이며, MIDI event 에 해당한다. 각 줄마다 순서대로 재생 위치, 사용 여부, 악기 채널, 음표, 세기에 대한 값을 나타낸다. 드럼 연습기를 사용하므로 악기 채널과 음표의 값은 고정한다.

```
3480 Meta TrkEnd
TrkEnd
```

위 부분은 트랙 부분의 끝을 표시하는 부분이며, Meta event 에 해당한다.

4. 추천 시스템

4.1 특징값

기존의 MIDI 컨트롤러로부터 받은 디지털 수치를 표준 미디 파일 포맷(SMF)에 맞게 생성한다. 그 파일에서 이제 특징값이란 것을 추출해야 하는데, 여기서 특징값의 의미는 실제 연주자가 친 하나의 MIDI 파일과 시스템이 가지고 있는 MIDI 파일을 비교하려고 할 때 무엇을 두고 비교하는 가에 대한 어떤 특정한 값을 대조해서 비교하는데 그곳에 필요한 게 이를 특징값이라 한다.

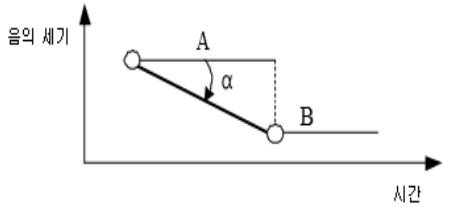
4.2 특징값 추출

MIDI 파일의 특징값이라 하면 고유의 가지고 있는 연주 정보, 즉 박자 / 키 / 음의 높이 / 길이 / 세기 등이 있는데, 여기서 아두이노를 통한 신호를 드럼 신호로 바꾸게 되면 가장 두드러지는 값이 바로 음의 세기(Velocity)이다. 이 세기에 대한 값을 바로 비교하는 것이 아니라 음에 대한 상대적인 평균의 차를 통해서 비교한다. 굳이 음에 대한 평균의 차를 이용하는 이유는 같은 연주였다 할 지라도 세기의 절대값은 주변환경에 따라서 변할 수 있는 소지가 있기 때문이다. 따라서 평가 기준으로써 각 음의 상대적인 차이를 이용함으로써 주변환경에서의 방해요소를 제거하기로 한다.

$$X = V - V_{avg}$$
(여기서 V_{avg}는 음의 $Velocity$의 평균)

(수식 1) 음의 세기값 정규화

외부환경을 받지 않도록 [수식 1]을 이용하여 각 음의세기에 대하여 정규화 처리를 해주도록 한다.



(그림 1) 음의 세기값 정규화

그 후에, 1 차적인 처리를 거친 두개의 음 A 와 B 사이에서 특징값이 추출되는데, 이 값은 다음 식을 통해 도출된다.

$$Feature = \tan(\alpha)$$

(수식 2) 특징값 추출 식

[수식 2]를 통해 연주의 평가 기준으로써 박자와 음의세기를 고려한 평가가 가능해진다.

4.3 특징값 비교

사용자의 연주로 부터 얻어낸 특징 값과 그에 대응하는 악보 연주 특징 값을 비교한다.

$$i = \tan(\alpha') \qquad Similarity(i,j) = \frac{i \cdot j}{\|i\| \, \|j\|}$$
$$j = \tan(\alpha)$$

(수식 3) 유사도 측정

코사인 유사도는 두 벡터 값 사이의 각의 코사인 값을 이용하여 두 벡터 i, j 의 유사도를 측정하는 방법인데 주로 데이터 마이닝이나 검색에 사용된다. 코사인 유사도를 통해서 나오는 값은 - 1 ~ 1 사이의 수이며 그 값이 1 에 가깝다는 것은 두 벡터가 유사하는 뜻이고 반대로 -1 에 가까울 수록 서로간 상관성이 떨어진다는 뜻이다.

이를 이용해서 악보연주로 부터 추출된 특징 값들을 Training dataset 하고 연습자의 연주로부터 추출 된 특징 값들을 Test dataset 으로 한다. 훈련의 결과로 도출된 기준선과 사용자의 연주를 [수식 3]을 통해 비교함으로써 사용자의 드럼 연주를 평가할 수있게 된다.

4.4 추천

연주를 분석하고 판단할 수 있게됨으로써 사용자의 연주에 맞는 추천곡 기능이 가능해지게 된다.
예를 들어, 데이터베이스에 연주법에 따른 루디먼트, 연주의 장르, 연주에 맞는 대중음악가요들에 대한 특징 값 정보가 들어있다고 하자. 결과적으로 우리는 사용자 연주로 부터 얻은 특징 값들을 기준으로 하여 데이터베이스에 저장되어 있는 것들과 비교가능하게 되며, 따라서 사용자의 연주와 유사한 것들을 추출하여 알려줄 수 있게된다.

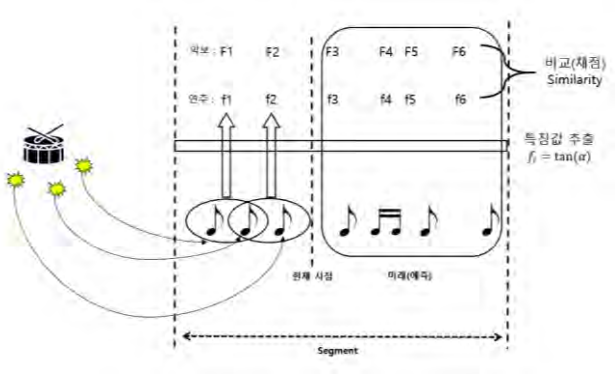
5. 실시간 채점 시스템

인공지능 드럼연습기는 사용자의 연주에 따른 실시간 악보 출력 및 점수를 보여준다. 대략적인 방법은 다음과 같다. 만약 사용자가 드럼을 2 번 이상 치게 될 경우에 연속한 두 음 간의 관계에 대한 특징값이 도출 된다. 매 충격마다 특징값은 생성되고 이렇게 생성된 추출값의 집합은 일정 구간 단위로 점수 채점에 사용된다.

① 매 연주 마다 특징값 추출

② 추출된 특징 값들은 일정길이 단위로 채점

여기서 주의할 점은 아래 그림에서도 알 수 있듯이 실시간 채점 시스템을 하기 위해서는 미래에 있을 음에 대한 예측이 필요하다. 왜냐하면 채점의 단위는 하나의 특징 값이 아니라 개발자가 지정한 일정한 길이의 Segment 단위로 이루어 질 것이기 때문이다.



(그림 2) 실시간 채점

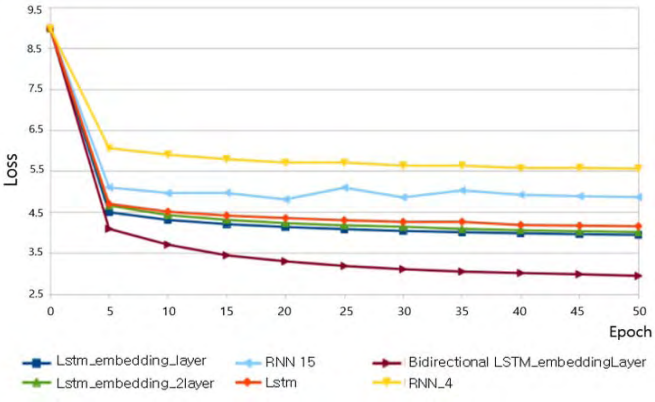
5.1 Segment 단위의 채점

Segment 단위의 실시간 채점을 할 경우에는 현시점에서부터 Segment 의 끝단 까지 매번 예측을 해야한다. 따라서 이러한 채점 시스템을 속도면에서 효율성을 떨어뜨린다. 그럼에도 불구하고 Segment 단위의 채점을 선택한 이유는 평가의 정확성 때문이다. 소리와 같은 연속적인 데이터를 판단하는데 있어서는 과거의 정보를 고려하는 것이 중요하다. 이와 비슷한 경우로 예를 들자면, 자연어 처리가 그러하다. 한 문장의 의미를 이해하기 위해서는 각각의 단어가 아닌 전반적인 문맥의 파악이 중요하다. 낱개의 짧은 조각만 가지고는 전체적인 흐름파악이 어렵기 때문이다. 따라서 처리 지연에 큰 영향을 주지 않는 범위 내에서 알맞은 Segment 길이를 설정하는 것이 핵심이라고 볼 수 있다.

5.2 LSTM 을 통한 단위 segment 예측

LSTM 은 최근 음성 인식, 필기 인식, 그리고, 기계 번역을 비롯한 자연어 처리 에서 탁월한 성능을 보이며 기계 학습의 핵심 방법론으로 자리 잡았다. 뿐만 아니라, 로봇 컨트롤, 동작인식, 시간적 예측 등에도 널리 활용된다. LSTM 을 통해서 오래된 과거의 결과까지 고려하여 앞으로의 연주를 예측해 볼 수 있다.

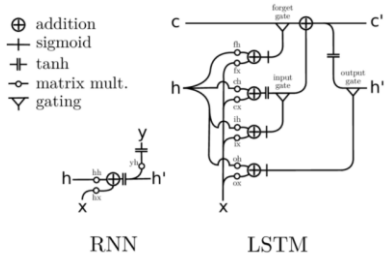
그러나 일반적으로 음성인식 오픈소스의 대표주자는 RNN 이다. 하지만 RNN 은 오랜 과거의 정보를 기억하는데 있어서 LSTM 보다 성능이 떨어지는 단점이 있다. 두 오픈소스의 성능을 비교하면 다음과 같다.



(그림 3) LSTM vs RNN

RNN 의 loss 는 데이터 셋의 수가 많아 질수록 감소하는데 한계를 보이는 반면에 LSTM 은 보다 효과적으로 loss 를 떨어뜨리는 모습을 볼 수 있다.

RNN 과 LSTM 모두 동일한 연산을 하는 뉴런들이 연산 결과를 다음 뉴런으로 넘겨주는 구조를 취하고 있다. 그렇다면 두 기법 간의 성능 차이는 어디서 나오는 것일까? 그것은 바로 각 뉴런에 위치한 연산 게이트의 차이로 부터 나온다.



(그림4) RNN노드 vs LSTM

RNN 은 tan h 연산 하나만을 가지는 반면에 LSTM 은 4 개의 게이트를 통해 보다 정교한 정보전달을 하게 된다. 특히 맨 상단에 위치해있는 forget gate 라는 요소를 활용하여 필요한 정보를 더하거나 제거하는 기능을 구현하게 된다.

이와 같은 이유로 우리는 segment 단위의 연주를 위한 예측 오픈소스로써 LSTM 을 선정하였으며 이를 이용하여 보다 정확한 미래 연주를 예측한다.

6. 결론

위 설계 방법 통해서 우리는 데이터 분석 및 인공지능 기술을 이용한 똑똑한 드럼 연습기를 제작할 수 있게 된다. 신기술을 이용한 새로운 기능은 기존의 시스템이 지니고 있던 한계를 뛰어넘게 해줌으로써 사용자에게 보다 풍부한 경험을 제공한다. 앞으로 인공지능 기술은 대중화가 될 것이다. IoT 분야가 생활 속 가정, 가전 제품으로 확장되고 있기 때문이다.모든 사물과 통신할 수 있는 인프라가 구축된 현시점에서 인공지능 분야가 개입할 수 있는 여지는 무궁무진 할 것이다. 비록 현재 우리가 서술한 지금의 드럼연습기의 형태는 1인의 드럼 연습의 형태에 초점을 맞추고 기술 하였으나, 후에는 개발 사항에 다양한 악기를 고려하고 이에 덧붙여 IoT통신을 이용한 사용자간 커뮤니티를 형성할 수 있도록 한다면 자신의 집안에서 2인 이상의 연습자들이 모여 합주가 가능할 수 있을 시대도 머지않았을 것이라 생각한다.

7. 감사의 말

우선 논문을 작성하기 위해 물심양면 지원을 해주었던 한이음 감사를 표합니다. 끝으로 논문 작성을 위해 고군분투한 팀원들 각자에게 서로 수고했다는 말을 전하고 싶습니다.

참고문헌

[1] 毛鍾植 “선율 정보를 이용한 음악의 유사도 계산 알고리즘” 인하대학교 컴퓨터정보공학 석 박사 학위논문, 2000

[2] 유석인 “전문가 시스템의 성공 사례” http://dl.dongascience.com/magazine/view/S199006N029, 과학동아 1990 년 6 월호

[3] 김재희 “인공지능의 기법과 응용” 교학사, 1988

[4] 유석인 “인공지능 원론” 교학사, 1988

[5] 인공지능 - [9] 생성 시스템 http://booolean.tistory.com/659, 2017

[6] MIDI 파일 분석하기 #1. Header, Track 개요 http://frontjang.info/entry/MIDI-%ED%8C%8C%EC%9D%BC-%EB%B6%84%EC%84%9D%ED%95%98%EA%B8%B0-1-Header-Track-%EA%B0%9C%EC%9A%94

[7] MIDI 시스템 이란 http://rennflav.blogspot.kr/2012/06/midi-system-exclusive.html

[8] MIDI 파일 구조 http://www.somascape.org/midi/tech/mfile.html#structure

[9] 동요 MIDI 파일 http://kwin.net/mid/dong/adong.htm

[10] MIDI 구조의 이해 - PrCh https://discussions.apple.com/thread/7180057?start=0&tstart=0

[11] 박상준 “기계 학습을 이용한 내용 기반의 음악 장르 분류” 서울대학교 공학석사학위논문, 2002