붙임2

2018 학년도 제 2 학기

연구논문/작품 중간보고서

제목	머신러닝을 통한 드럼 악보 추출	○ 논문( ) 작품(V)
GitHub URL	https://github.com/nelpsos/DrumMusicScore	
평가등급	지도교수 수정보완 사항	팀원 명단
A, B, F중 택1 (지도교수가 부여)	○ ○ ○	박동준 (학번:2013311872) 홍주경 (학번:2012311332)

2018 년 9월 18일

지도교수 : 윤희용 서명

Contents

0. 요약 ····· 1

1. 서론 ····· 2

1.1. 제안배경 및 필요성 ····· 2

1.2. 연구/논문 작품의 목표 ····· 3

1.3. 연구/논문 작품의 전체 Overview ····· 4

1.3.1. Sound Signal Extraction and Drum Signal Extraction ····· 4

1.3.2. Machine Learning ····· 4

1.3.3. Visualize with Musical Score ····· 4

1.4. 선행연구 및 기술현황 ····· 5

1.4.1. Hum On ····· 5

1.4.2. Machine Learning ····· 5

1.4.3. Artificial Neural Network ····· 5

1.4.4. Decision Forest ····· 6

1.4.5. MIR(Music Information retrieval) ····· 6

1.4.6. Visualization ····· 7

2. 관련연구 ····· 8

3. 제안 작품 소개 ····· 10

3.1. 시스템 구성 ····· 10

3.1.1 Ubuntu 16.04LTS ····· 10

3.1.2 Python v3 ····· 11

3.1.3 TensorFlow ····· 11

3.2. 이론적 배경 ····· 12

3.2.1 Datasets for Drum Sounds ····· 12

3.2.2 LibROSA ····· 12

3.2.3 Scikit-Learn ····· 13

4. 구현 및 결과분석 ····· 14

5. 결론 및 소감 ····· 16

6. 참고문헌 ····· 18

0. 요약

Recently, drum, the most popular and important percussion instrument in modern music, is an instrument that many people want to learn. However, many people are suffering from difficulties of making drum scores. Moreover many people feel that there are not enough materials about drum score, and also people found that making drum scores are labor-intensive work. a ppeal for lack of materials to find or make music.

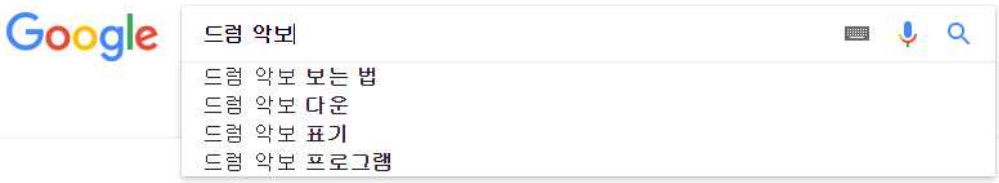
For those who are experiencing this difficulty, we want to create a product that extracts drum scores from voice files. The working principle of the work is as follows. First, from the voice file with drum sound, we classified each musical instruments' sound signal and extracted the data set of drum sound by using MIR(Music Information Retrieval) Technology After that, by applying one of machine learning algorithm and multiclass classification method, categorize classified drum sound by 7 kinds. Finally visualize the categorized data with the drum score.

1. 서론

1.1. 제안배경 및 필요성

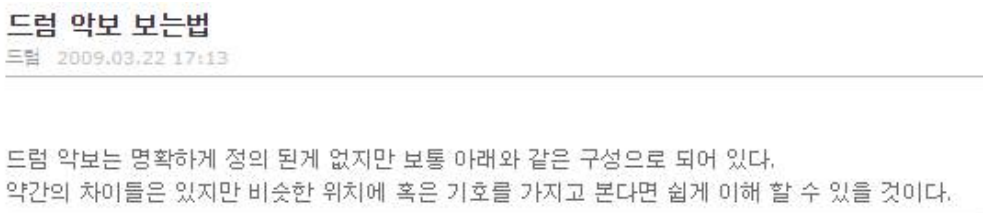
<Figure 1> Infographic of Survey Reportfrom JakPat:
People's Interest in Playing a Musical Instrument

Drums are the most popular and important percussion instrument in contemporary music. According to the <Figure 1>, a random survey of internet site JakPat, 49.18% of the survey participants replied that they have interest in percussion, including drums. As such, drums are not the only thing for specialized people, and many people are learning drums as a hobby.



<Figure 2> Related keyword of "drum score" search in Google Korea

Despite of these popularization of drums, however, drummers including hobbyists faced large obstacles. According to <Figure 2>, when we search drum score in Google Korea, the first thing in the auto-complete related query is "How to view drum score". Even in the lecture about "how to view drum score", as in <Figure 3>, says "There are no clearly defined method to make drum score.". As such, reading and making drum score is a quite cumbersome work, and so if you are not a major in drums, writing your own drum score is a very difficult and demanding work. So when people are looking for a piece of drum score that they want, or when they want to write a new piece of drum score, they suffer from the lack of materials.



<Figure 3> Lecture about how to view drum score

1.2. 연구논문/작품의 목표

In order to make it easier for many people who are experiencing this difficulty, we want to create a program that can extract the drum music score by inputting the music file. And, with this work, we hope that many drummers will be able to find and make drum score easier, and hope to be able to lead a more convenient cultural life.

1.3 연구 논문/작품 전체 Overview

1.3.1. Sound Signal Extraction and Drum Signal Extraction

It is the process of getting the data sets of the voice file or the sound source prior to the machine learning. Refer to existing MIR-based technologies and understand how to extract sound signals from audio sources by using matlab program through internet research. In addition, the extracted signal undergoes secondary classification with 7 kinds for example, cymbals, snare drum, hi-hat, etc.. according to the drum sound. Classification will depend on the frequency of the drum part.

1.3.2. Machine Learning

Before letting the computer to learn the actual dataset, understand the various machine learning modeling techniques first and find the appropriate technique to use. Moreover by using Python and Matlab, we will study how to apply machine learning technique to the computer. In addition, when we input the real data set to let computer to learn, we will use an appropriate method and programming language. Finally, calculating the accuracy by applying various machine learning methods.

1.3.3. Visualize with Musical Score

According to the result from 3.3, visualize the data from result with open source software which help user to draw musical score. Prior to use that , we need to first study and understand the instructions of that software. Then we can finally draw a musical score by inputting data set with coding.

1.4 선행연구 및 기술현황

1.4.1. Hum On

There is an application called 'Hum On'. It is a simple concepted application that turns humming, or hum, into a musical note. However this simple concept contains really complex technique. The sound of humming is different for person by person. In addition, even the same sound of the same person is different from the sound of other situations. The 'Hum On' is the application that transfers this to the musical score.

Machine running is the basis for making this difficult technology available. 'Hum On' analyzed humming 'Big Data' by machine learning technology and applied the technique of reading user intention. With this technology, you can turn humming into musical score as well as other sounds.

1.4.2. Machine Learning

Machine learning means let machine to learn something. In other words, even if a person does not explicitly direct logic to a computer, it means that the computer can 'learn' through a large number of data sets and automatically solve the problem through the learned 'experience'. Like that, machine learning is a field of research that develops algorithms that allow machines to "learn" from data and to perform actions that are not explicitly specified in code.

1.4.3. Artificial neural network

Artificial neural network, which is a statistical learning algorithm of machine learning inspired by neural networks of biology. An artificial neural network refers to the entire model that has artificial neurons (nodes) that form a network due to defects of the synapses, which change the binding strength

of the synapses through learning and have problem solving ability.

Artificial neural networks include teacher learning that is optimized for problems by inputting teacher signals (answers), and companion learning that does not require teacher signals. In the case of this work, the actual music and its drum score can be used, so teacher learning will be used.

1.4.4. Decision Forest

Decision Forest is a forecasting model which connects observed value with desired value. It is a kind of forecasting modeling method using decision tree in machine learning. Classification tree is a tree model with finite target variable. In this tree structure, a leaf node is a class label and branch is a logical product of characteristics of class label.

Decision tree is visual and explicit. It is used to express decision making process and decision made or data itself. In this project, artificial neural network will be replaced if decision forest is more precise than artificial neural network.

1.4.5. MIR(Music Information Retrieval)

MIR is a technology which analyzes and extracts rhythmic information of audio contents automatically and visualize it. MIR is a kind of Digital signal process(DSP), and it is studied more these days. MIR analyzes specific vectors made by parameterizing audio signal, with techniques such as pattern matching. In plain language, MIR changes music sound to 'sound signal'. Music is just a gathering of some sounds. However, if the music is changed to electric signal, it becomes information. It can be sole signal or splited signal, to be sent to someone. In now, especially in web or mobile, MIR is applied to common applicaion.

1.4.6. Visualization

There are not many programs which can draw music score with computer. Moreover, few programs provide circumstance where we can use data from machine learning to draw music score.

In this project, we use a software named 'Lilypond'. Lilypond is complete open source software. It has all functions required to typing music score. Also, many users are improving it animately and can use it for free, because it is open source software.

2. 관련연구

First, [1] "Machine Learning for Audio, Image and Video Analysis : theory and applications" is about machine learning for audio, image and video analysis. In this paper, writers used Bayesian theory of decision, Gaussian density, expactation and maximization algorithm, K-Means, etc. for machine learning. They applied the methods to speech and handwriting recognition, automatic face recognition and video segmentation and keyframe extraction.

Second, [2] "Machine Learning for Audio, Image and Video Analysis" is similar to [1], but more theoretical than it.

Third, [3] "Application of Symbolic Machine Learning to Audio Signal Segmentation" is about machine learning to audio signal segmentation. In this paper, witers address a data-driven approach to the problem of automatic segmentation of speech. They put music into phones and notes respectively that makes use of symbolic mahine learning techniques. This paper divides whole segmentation process into four steps:

1. Series of non-linear transformations are used for building first-order features that allow easy detection of segmentation candidates.
2. Fetures that describe sound properties in the neiborhood of a segmentation candidates are developed.
3. The set of segmentation candidates is transformed into machine learning data set by labeling candidates in accordance to the annotated speech corpus.
4. Supervised symbolic machine learning methods are applied resulting in segmentation rules.

Fourth, [4] "DNN을 이용한 오디오 이벤트 성능 검출 비교" and [5] "CNN과 CRNN을 이용한 오디오 이벤트 검출 비교". In this project, we have to choose which neural network is better. There are three candidates: Deep Neural Network(DNN), Convolutional Neural Networ(CNN), Convolutional Recurrent Neural Networ(CRNN). We read this papers considering selecting one of these three candidates and reached the following conclusion: The simple is the best. We choose DNN to solve this problem.

Finally, [6] "Automatic Classification of Musical Audio Signals Employing Machine Learning Approach" is about classification of musical audio signals. In this project, classifying the voice datasets is a important issue. We have 7,476 sounds of datasets by 200 drums, so classification is an unavoidable and indispensable task. This paper can classify audio signals, and we read to classify our datasets.

3. 제안 작품 소개

3.1 시스템 구성

In this part we will going to introduce our development environment. What OS, Program for coding and deep learning method did we chose.

3.1.1 Ubuntu 16.04 LTS

<Figure 4> Simple Screenshot for Ubuntu 16.04 LTS

For OS, we selected Linux Ubuntu version 16.04 LTS. Figure 4 shows the simple screenshot of Ubuntu. To briefly introduce this environment, it's code name is Xenial Xerus and is Debian GNU/Linux based. Ubuntu has several important characteristics which are first, it is optimized for personal desktop environment. It provides comfortable user interface. Second, it is free software based. So we can use this OS for free. Also, there are several improvements from previous version which was Ubuntu 14.04 LTS. First, basic setting is now

python v3. In previous version, python v2 was used. We will explain about python 3 at next stage. Furthermore, lots of applications are improved for user convenience.

3.1.2 Python v3

For programming language, we used Python 3 which has been growing in popularity over the last few years. This is an interpreted language so is passed straight to an interpreter that runs the code directly. This makes for a quicker development cycle because we just type in our code and run it, without the intermediate compilation step.

Recently Python is taking center stage as a machine learning language. This is because of its simplicity, it has machine learning library and it manages memory for user.

3.1.3 TensorFlow

TensorFlow is an open source software library for numerical computation using data-flow graphs. It was originally developed by the Google Brain Team within Google's Machine Intelligence research organization for machine learning and deep neural networks research, but the system is general enough to be applicable in a wide variety of other domains as well.

TensorFlow is cross-platform. It runs on nearly everything: GPUs and CPUs—including mobile and embedded platforms etc.

We installed tensorflow to our desktop by following commands.

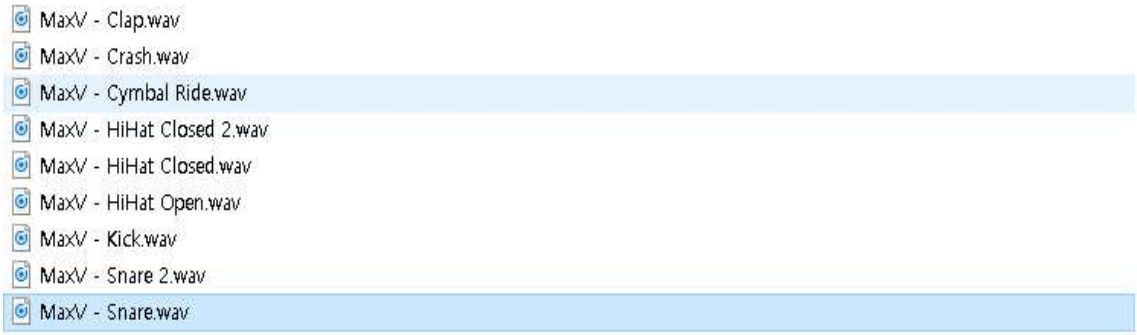
```
$ pip install --upgrade tensorflow
$ pip install --upgrade tensorflow-gpu
```

```
$ pip3 install tensorflow
```

<Figure 5> Installing Tensorflow

3.2 이론적 배경

3.2.1 DataSet for Drum sounds



<Figure 6> How the data are arranged

For the first thing we do for this project, is to find various Drum audio datasets for deep learning method. The more audio source file we have, the more accurate result will be given.

However it was impossible for us to get tons of drum audio files by self recording. So we found a opensource datasets in google. This dataset for drum machine contains 200 type of drum sounds which are from various drum producing company. It contains electrical, classic and vintage etc. Those drum sound data are agains divided in to numerous types which are snare, clap, cymbals, hihat and toms. Figure 5 shows how the datasets are divided and arranged. Since there were some unneeded sounds and data are not arranged, we unified the file names and left only files we needed. By those various kinds of drum sounds we can cover various type of audios.

3.2.2 LibROSA

LibROSA is a python package for music and audio analysis. It provides the building blocks necessary to create music information retrieval systems.

3.2.3 Scikit-learn

Scikit-learn is an open source tool that can be used to data mining and data analysis, built on NumPy, SciPy and Matplotlib.

4. 구현 및 결과분석

We used Tensorflow in machine-learning. The program learned a learning dataset and used testing dataset. Datasets have datas with 20 dimension vectors. In result, the program assumed the answer with very low error rate. For example, in <figure 6>, Its error rate was below 0.001% in 200 times of machine learning with simple 2 dimension vector.

Then, we searched datasets to use in machine learning. We found 7,476 sounds by 200 drums. We used LibROSA and Scikit-learn to analysis these datasets.

First, LibROSA is a python package for music and audio analysis. It provides the building blocks necessary to create music information retrieval systems. So we used LibROSA to load drum sound datasets to the program and

```
0 [[ 1.1097053  -0.69301444]] [0.2927837]
20 [[ 0.4946692  -0.15650275]] [0.24358404]
40 [[0.26195386 0.06302895]] [0.2725705]
60 [[0.16616987 0.1469284 ]] [0.2874659]
80 [[0.1269457  0.17928873]] [0.29448244]
100 [[0.11094496 0.19186881]] [0.29762965]
120 [[0.10443697 0.1967919 ]] [0.2989986]
140 [[0.10179596 0.19872913]] [0.29958194]
160 [[0.10072613 0.19949488]] [0.29982698]
180 [[0.10029329 0.19979872]] [0.29992884]
200 [[0.10011838 0.19991964]] [0.29997087]
```

<Figure 6> A simple Tensorflow example

change them MFCC. MFCC is the most popular feature value and can

be gained by MFC. The mel-frequency cepstrum (MFC) is a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear mel scale of frequency. MFC can be splited to this six steps:

1. Cut audio signal of input time domain down to small frames.
2. Compute Periodogram estimate(Periodogram Spectral Estimate) of Power Spectrum for each frames.
3. Apply Mel Filter Bank to Power Spectrum of 2, then plus energy to each filter.
4. Take log to every filter bank energy of 3.
5. Take DCT to value of 4.
6. Get Coefficients 2~13 of result of DCT.

After this 6 steps, the Coefficients is called MFCC. We assume 20 MFCCs, and if it lacks, we will increase that up to 50.

Scikit-learn is an open source tool that can be used to data mining and data analysis, built on NumPy, SciPy and Matplotlib. We will use this classifier using MFCC vectors as input by LibROSA and 4 drum sound classes as output.

In this step, we must prevent Overfitting. We don't have to use all of 7,476, but have to use 4,476 to learn and 3,000 to test.

We will analysis, use to learn, test and use based on actual data. However, other loud sounds which is similar to drum sounds can be recognized as drum sounds. To prevent this, we are considering adding the dMFCC and ddMFCC to our learning data or make to learn the sounds of other instruments as well.

5. 결론 및 소감

5.1. Works for future

For the future we will use TensorFlow to let computer learn sampled data (the drum signals) from audio file. Then we will make a code to classify 4 different drum sounds. For this work, we might increase the amount of drum sound types. At the initial perspective, we decided to divide drum sounds for 4 different types cymbals, snare, toms, and kicks. However, now we realized that there are more sounds that we need to focus. At last we will make a code to match drum score and specific drum signal. this is also called visualization step. The major thing in this step is to show the specific drum signal to the musical score.

5.2. Conclusion and Perspective

By advancing this graduation work, we have experienced many things that we cannot experience in general undergraduate courses.

First of all, with my team member, we got together in a place and made a lot of technical discussion. This made us to increase major knowledge and learn how to solve various problems.

Moreover, since this was the first time for us to do this big project, we systematically divided the project into parts and solved them step by step. This made us to solve problem fluently.

Finally, we used pair programming method to code our project. By using this method, we shared our major knowledge and anguished specific problem together. This made us to more concentrate on our work. Plus, we can also immediately proceed the code review and help to find each other's errors.

6. 참고문헌

[1] Francesco Camastra, Alessandro Vinciarelli. "Machine learning for audio, image and video analysis : theory and applications", London : Springer, 2015.

[2] Yu, J. "Machine Learning for Audio, Image and Video Analysis Francesco Camastra and Alessandro Vinciarelli", INTERNATIONAL SOCIETY FOR OPTICAL ENGINEERING (SPIE) 2009. Berlin; Great Britain; Springer, 2005.

[3] Raskinis, A., Raskinis, G. "Application of Symbolic Machine Learning to Audio Signal Segmentation" Berlin; Great Britain; Springer, 2005.

[4] 정석환, 정용주, "DNN을 이용한 오디오 이벤트 검출 성능 비교" 한국전자통신학회, 2018.

[5] 정석환, 정용주, "CNN과 CRNN을 이용한 오디오 이벤트 검출 성능 비교", 한국정보기술학회, 2018.

[6] Kostek, B. , Kupryjanow, A. , Zwan, P. "Automatic Classification of Musical Audio Signals Employing Machine Learning Approach" New York; Audio Engineering Society, 2011.

[7] Davis, S. Mermelstein, P. (1980) Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences. In IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 28 No. 4, pp. 357-366

[8] X. Huang, A. Acero, and H. Hon. Spoken Language Processing: A guide to theory, algorithm, and system development. Prentice Hall, 2001.