Music Data Representation and Information Retrieval Using Vector-based Similarity Scores

Hanchao Li

The Faculty of EEC, Coventry University, Coventry, UK, CV1 5FB

lih30@uni.coventry.ac.uk

ABSTRACT

Large repository of music can be accessed or downloaded over the internet. The Music Information Retrieval (MIR) system, music storage and music representation are based on either audio-based or symbolic-based method. Audio-based method can store continuous sound waves well, however, it has limited in illustrating the content flow of the music. Symbolic-based method has well in represent the content of the music including recognize similar patterns but not suitable for electronic music (EM). In this project, we proposed a data model which describe the music information using both Music Definition Language (MDL) and Music Manipulation Language (MML). We developed a hybrid method that combines both audio, contour and rhythm features, and employed a Deep Learning based classification mechanism that maps variations of music pieces to their corresponding originals, based on similarity scores. The newest experimental results obtained accuracy of stable 84%, and the potential has evaluated to be 96%. Moreover, if the music were stored by MDL and MML, the overall performance lies between symbolic-coded files and audio files. For the sample music track of 18 seconds, the file size is 14.3kB and the soundwave generated is better than MIDI file and closer to the audio file.

1. INTRODUCTION

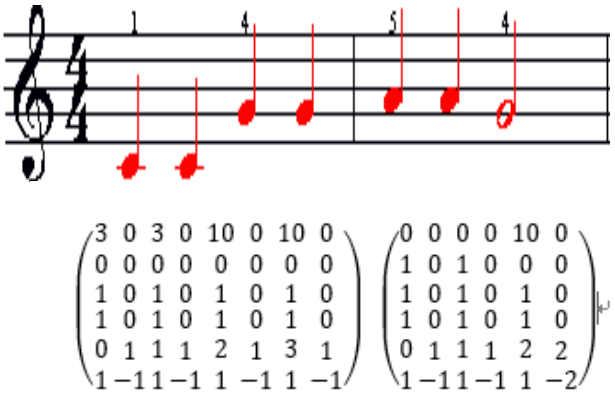
Music Data Storage, Music Data Representation and Music Information Retrieval (MIR) systems, such as music plagiarism detection tools and music search engines, or even music recommendation systems, are getting more important and popular these days. My project is to propose a data model naming MDL and MML, and apply techniques from Deep Learning, this can achieve a better performance. The outline of the project has shown in Figure 1.

Figure 1. Overall System Outline.

2. DEEP LEARNING FROM STAGE ONE [1]

We've designed MDL and MML at this stage. MDL represents what notes should be played and MML models how the notes are played. Essentially, MDL is similar to the existing MIDI files. To address the limitations of MIDI files, MML is designed to shape audio waves to a stream file (e.g. MP3) since the information on the continuous change of frequency, loudness etc. is coded inside MML in order to achieve a better performance. The examples of MDL and MML has shown in Figure 2.

$$Key: \begin{pmatrix} Mod\ 12\ for\ Notes \\ Set\ Number \\ Amplitude \\ Bar\ Number \\ Time\ for\ the\ notes \\ Duration \end{pmatrix} \begin{pmatrix} \Delta M \\ \Delta S \\ \Delta A \\ \Delta B \\ \Delta TN \\ \Delta D \end{pmatrix}$$



$$\begin{pmatrix} 3 & 0 & 3 & 0 & 10 & 0 & 10 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 2 & 1 & 3 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 0 & 10 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 2 & 2 \\ 1 & -1 & 1 & -1 & 1 & -2 \end{pmatrix}$$

Figure 2. Examples of MDL and MML.

With the proposed data format, we developed of a hybrid method on the proposed musical data model that combines contour, rhythm and audio fingerprints. The method features a modified deep learning based classification mechanism that maps variations of music pieces to their corresponding originals. The Deep Reinforcement Learning involves Strategy Tree passed through a SOM-based Neural Network, different routes can gain, a maximum overall accuracy as a reward. From the 50 selected sample music tracks, the newest experiment provides an accuracy of stable 84%, and the potential has evaluated to be 96%. This is caused by that, one is more unsupervised learning, while the other is more semi-supervied learning.

Furthermore, it will be better for Origins and Rhythm Vari-

ations, to use Audio Fingerprinting method; for Key Variations and Reduction, to use Symbolic-based method; while for Expansion, to use Symbolic-based method but with different equation. For the details of the equations used can refer to [1]'s equation number 2 and 3 respectively.

3. INVESTIGATION FROM STAGE TWO[2]

From the case investigation and related literature review carried out in [2], we can conclude some of the following conclusions:

- Lack information on music flows for audio-based MIR.
- MIR systems cannot find variations to its origin for audio-based MIR.
- Slide notes from modern pop songs/electronic music cannot be represented in this way for symbolic-based MIR.
- Symbolic-based MIR systems had very low accuracy on rhythm based search.
- Current symbolic files, such as MIDI or equivalent music sheet, can be extracted from audio files.
- Accuracy for rhythm variation can be improved after applying SOM.
- MDL-based system can obtain a reasonable accuracy for 10 bonus tracks, which does not lie inside the training set.

Some of those conclusion points have been dealt by our new approach.

4. REPRESENTATION FROM STAGE THREE[3]

The sample track we've used to generate the sound waveform is the first 18 seconds of the song title: "Zero For Conduct" by Block B – Bastarz. With the sample music track, at this stage, we have successfully generated a MDL and MML data file, of the file size of 14.3kB in Excel format, and a soundwave from MDL and MML based file. The comparison graph of the soundwaves obtained in [3] has shown in Figure 3.

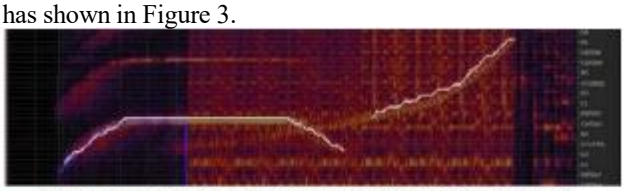


Figure 11a. Audio Channel: Pitch-Time Comparison.

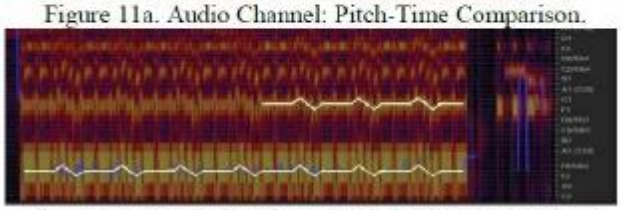


Figure 11b. Symbolic Channel: Pitch-Time Comparison.

Figure 3. Pitch-Time Comparison from Paper [3].

The purple line and the red-yellow graph was generated by the Adobe Audition spectral pitch display using the original audio file, the white line is generated by our demo-system using MatLab.

Based on this, we've also designed a media framework for the future media player, if using MDL and MML as storage and representing data type. The framework has shown in Figure 4.

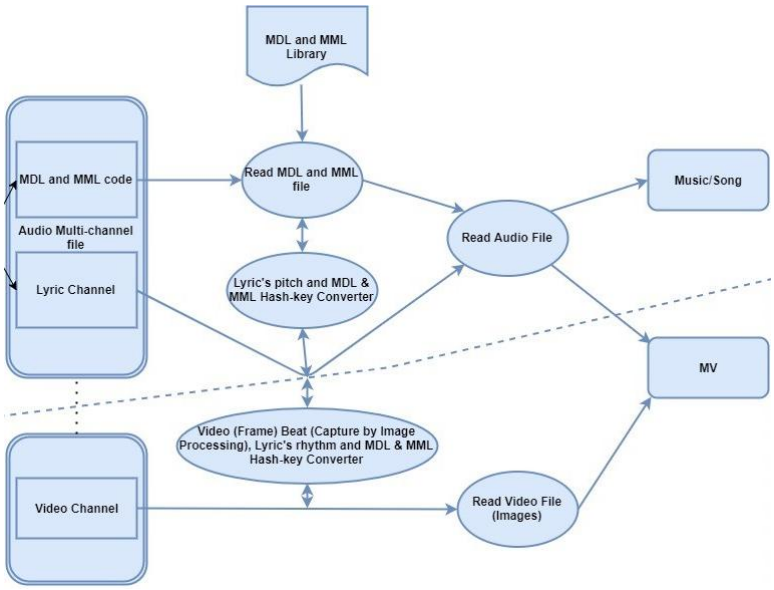


Figure 4. Preliminary Media Player Framework.

5. CONCLUSION AND FUTURE WORK

We have finished implement a demonstration for stage 1 and stage 3, and evaluate existing systems at stage 2. As mentioned in Section 3, MDL and MML can store music information as well as classify variations to its origin. On the other hand, MDL and MML can generating sliding notes for Electronic Music as well as classify with rhythm feature. The storage size lies between the audio and the symbolic file. The performance (classification, representation) is acceptable.

In the future, we need to cover the final stage (4), which is to convert music file into MDL-and-MML-based file using signal processing. We may also compare the system which does cover song detection / variation detection with an existing system to compare the performance of the algorithm.

6. REFERENCES

[1] H. Li, X. Fei, K-M. Chao, M. Yang, and C. He: "Towards A Hybrid Deep-Learning method for Music Classification and Similarity Measurement" *Proc. of e-Business Engineering (ICEBE), 2016 IEEE 13th International Conference*, pp. 9-16, 2016 (To be extend to a Journal Paper with the newest result).

[2] H. Li, Z. Tang, X. Fei, K-M. Chao, M. Yang, and C. He: "A survey of Audio MIR systems, Symbolic MIR Systems and a MDL Demo-System" *Proc. of e-Business Engineering (ICEBE), 2017 IEEE 14th International Conference*, 2017.

[3] H. Li, D.Y.F. Zuo, X. Fei, K-M. Chao, M. Yang, and C. He: "MDL & MML: A Coding Scheme for Music Representation and Storage" *Proc. of e-Business Engineering (ICEBE), 2017 IEEE 14th International Conference*, 2017.