2013 IEEE International Symposium on Multimedia

Detecting Musical Genre Borders For Multi-label Genre Classification

Hiroki Nakamura, Hung-Hsuan Huang and Kyoji Kawagoe
Ritsumeikan University
Kusatsu, Shiga
{*is029088@ed, huang@fc,kawagoe@is*}*.ritsumei.ac.jp*

***Abstract*—In this paper, we propose a novel method to detect music genre borders for the music genre classification. The music genre classification is getting more important because music is influenced by an increasing amount of different musical styles. A general approach to classify music genres is a single genre labeling that usually gives the meaning of inherent stylistic elements to a musical piece. However this gives ambiguity in case of a music piece having multiple genres. To solve the problem, we consider separating the multi-label classification task into the single-label genre classification task. We propose a novel method to detect music genre borders for multi-label genre classification. The proposed method can find borderlines of different genres in music. Moreover, it is strongly expected to realize the multi-label genre classification to apply the single-label genre classification to each detected music segment.**

***Keywords*-genre border; music genre classification; music structure analysis**

I. Introduction

The multi-label genre classification is more important in the music genre classification because music is influenced by an increasing amount of different musical styles. Single genre labels usually assign the meaning of inherent stylistic elements to a music piece, however it increases ambiguity in case of musical pieces having multiple genres. Lukashevich [1] mentioned that it is obvious that the broad term ”World Music” is one of the most ill-defined tags when being used to lump all ”exotic genres” together.

To solve the problem, music having multiple genres is roughly segregated into two categories. The first is that some of music elements (such as melody, instruments, rhythm) are mixed with other style/performance, which is called ”Crossover”. However, it does not define a clear genre in the entire music by mixing other elements. The second is that music has the obvious change of genres on the time dimension. A typical example is ”Bohemian Rhapsody” by Queen, in which the genre change appears like a chapter in a movie. We focus on the second category and detect the genre borders in the music. It is more difficult to define a clear genre than in the first category.

In this paper, we propose an innovative method for detecting musical genre borders for the multi-label genre classification. A music piece is decomposed into genre segments by genre borders. Then multi-label genre classification is realized by applying single-label genre classification to each segments.

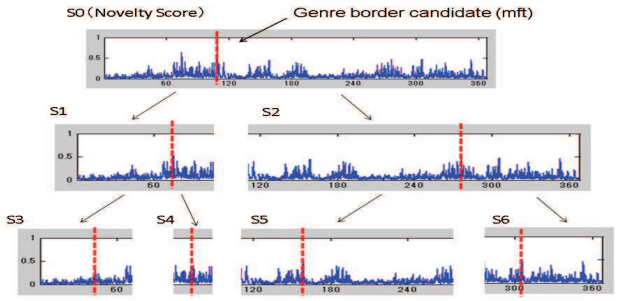

Figure 1. Detecting genre border candidates

II. Proposed Method

The proposed method is composed of two steps: genre border candidate generation, genre border determination.

A. Border Candidates Generation

Genre border candidates are generated from peaks in novelty score curves. Novelty score curves is calculated from MFCC, chroma features and rhythmograms by Foote’s method [2] for the purpose of automatically locating points where changes in acoustic features occur. In order to generate border candidates, we divide music segments using three thresholds about the genre change, which are $MAXSEG$, $SEGSIZE$, and τ. Throughout the paper, a segment is defined as a music sequence.

The first threshold, $MAXSEG$, is the maximum number of segments in one music piece. As the genre is usually used for arrangement in music, frequent genre changes tend to ruin musical impressions. Therefore, both the number of genre changes and the number of genres with a musical work have to be limited using $MAXSEG$.

The second threshold, $SEGSIZE$, is the size of a segment. When the length of a segment is extremely short, the segment can not contribute to the border detection process. This is because the music generally consists of a repetitive structure, where a base unit is mainly composed of four musical bars. As such, it is unlikely that a genre change could occur in fewer than four musical bars. Therefore, we define $SEGSIZE$ as $SEGSIZE = (60/TEMPO) * 16$, where $TEMPO$ is the speed of a music piece in beats per

978-0-7695-5140-1/13 $26.00 © 2013 IEEE
DOI 10.1109/ISM.2013.108

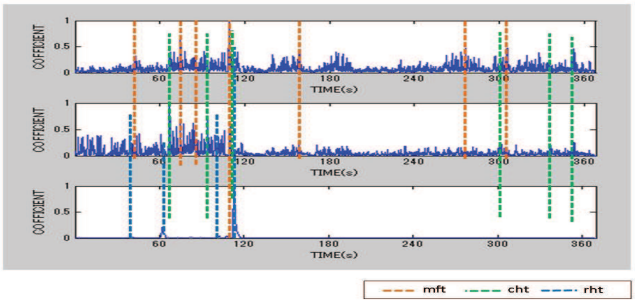

Figure 2. Determination of a genre border.

minutes and can be estimated by the method proposed by Lartillot [3].

The final threshold, τ, is specified for calculating a border from a large set of peaks. If the novelty scale value at a specified peak is greater than this threshold, then the peak will be classified as a border candidate.

An example of border candidates detection is shown in Fig. 1. Border candidates generated from MFCC, chroma, and rhythmogram are termed mft, cht, and rht, respectively.

B. Genre border Determination

In order to determine a genre border from a set of candidates, all the features are checked for whether or not they exist within the LAG. The LAG value is specified in one bar length units and is corresponds to $(60 = TEMPO) * 4(Insecond)$. This is because all the acoustic features should change at almost the same time. That is, if all the candidates obtained from the acoustic features exist within the LAG, then one of the border candidates is selected as a border. In this paper, we use the earliest border among the border candidates, as our preliminary evaluation results showed that genre changes are very likely to occur at the earliest border. An example of the genre border determination is shown in Fig. 2.

III. Evaluation

A. Dataset

We conducted evaluation using two datasets.

The first dataset was called the artificial dataset and was generated from the GTZAN dataset [4]. To create genre changes, musical pieces were generated from selecting two pieces that belonged to different genres and seamlessly combining them. The second dataset was called the real dataset included 30 songs, which we selected from commercially available pieces. Three knowledgeable music experts specified the correct genres and the genre borders for each work of music.

To compare the proposed method, we adopted one of the existing music structure analysis methods, which was developed by Jouni Paulus [5]. We compare the proposed method and the existing method by Paulus using four measures: the average F-measure, A_F, the average precision, A_P, the average recall , A_R, and the number of detected genre borders, A_{GB}.

TABLE I
Evaluation Result of Detecting Genre on Artificial Dataset. The Actual A_{GB} in Artificial Dataset is 1.0

	A_F	A_P	A_R	A_{GB}
Existing	0.31	0.20	0.71	3.24
Proposed	**0.50**	0.44	0.57	1.43

TABLE II
Evaluation Result of Detecting Genre on The Real Dataset. The Actual A_{GB} in The Real Dataset is 2.2

	A_F	A_P	A_R	A_{GB}
Existing	0.32	0.23	0.52	3.67
Proposed	**0.39**	0.21	0.53	3.10

B. Experiment Results

The evaluation results are shown in Table I and Table II. Both results indicate that the proposed method is more accurate than the existing method for the average F-measure, A_F. The proposed method can detect music borders more correctly than the existing method, mainly owing to the fact that our method innovates the thresholds concerned about the genre change.

IV. Conclusion

In this paper, we propose a method for detecting genre borders for multi-label genre classification. Foote's method is used as the basis of our detection method. From the experiment results, our method detected fewer genre borders incorrectly and was more accurate in detecting borders overall. Future work includes the introduction of the machine learning model approach to achieve a higher level of accuracy in genre border detection and further experiments on genre detection for music pieces related for multi-label genre classification.

Acknowledgment

This work is partially supported by KAKENHI #24300039.

References

[1] H. M. Lukashevich, J. Abeser, C. Dittmar and H. Grosmann, "From Multi-Labeling to Multi-Domain-Labeling: A Novel Two-Dimensional Approach to Music Genre Classification" in *Proceedings of ISMIR*, pp. 459-464, 2009.

[2] J. Foote. "Automatic audio segmentation using a measure of audio novelty". In *Proceedings of IEEE International Conference on Multimedia and Expo*, pp 452-455, 2000.

[3] O. Lartillot, "Mirtempo: tempo estimation through advanced frame-by-frame peaks tracking", MIREX, 2010.

[4] *GTZAN: http://marsyas.info/download/data_sets*

[5] J. Paulus and A. Klapuri. "Music structure analysis using a probabilistic fitness measure and a greedy search algorithm". In *Proceedings IEEE Transactions on Audio, Speech, and Language Processing*, pp. 1159-1170, 2009.