FEATURE SPACE MODIFICATION FOR CONTENT-BASED MUSIC RETRIEVAL BASED ON USER PREFERENCES

Keiichiro Hoashi, Kazunori Matsumoto, Fumiaki Sugaya

KDDI R&D Laboratories, Inc.
2-1-15 Ohara Fujimino-shi,
Saitama 356-8502, JAPAN

Hiromi Ishizaki, Jiro Katto

Dept of Science and Engineering,
Waseda University
3-4-1 Okubo Shinjuku-ku
Tokyo 169-8555, JAPAN

ABSTRACT

This paper proposes a feature space modification method for feature extraction of music, which is effective for the development of a content-based music information retrieval (MIR) system based on user preferences. The proposed method conducts clustering of all songs in the music collection, and utilizes the resulting cluster IDs as training data for feature space modification, and is capable to automatically generate a feature space which is suitable to the content of any music collection. Experiment results prove that the proposed method improves accuracy of user preference based MIR.

1. INTRODUCTION

Recent popularity of online music distribution services have enabled common users to personally accumulate a large scale music collection. Therefore, it has become difficult for users to efficiently find songs which they want to listen to. Development of an effective music information retrieval (MIR) system is, therefore, essential to realize satisfactory music distribution services, and improve usability of music applications.

This paper proposes a feature space modification (FSM) method, which is effective for feature extraction of music to be implemented effectively in a content-based MIR system which retrieves music that matches user preferences. The proposed FSM method is an expansion from the user preference based MIR algorithm presented by the authors in Reference [1]. Evaluation of the proposed method is conducted based on experiments on music data with subjective ratings, and the results prove that the method improves accuracy of user preference based MIR.

2. CONVENTIONAL MIR METHODS

In Reference [1], the authors have proposed an effective content-based MIR method, which retrieves songs that match music preferences based on a small set of example songs provided by the user. Our MIR method applies the tree-based vector quantization (TreeQ) algorithm developed by Foote[2]. The

TreeQ algorithm trains a vector quantizer based on training data, *i.e.*, a set of songs associated with category information. The quantization tree is grown so that the feature space is divided into regions that have maximally different class populations. The resulting tree can be used to vectorize any song, and automatically categorize songs by comparing the vectors of a song and all category vectors.

In the "genre-based" MIR method described in [1], the feature space is constructed based on a set of songs and their genre information. User profiles, *i.e.*, the vector expression of a user's musical preference, is then generated based on a small set of example songs, *i.e.*, songs which the user likes and/or dislikes. This approach, along with the implementation of relevance feedback, was proved to achieve accurate retrieval of songs which match users' preferences.

Several other reports have been presented regarding MIR based on user preferences. Chai *et al* have proposed an XML-based user model language for music[3], which enables retrieval of music based on user preferences. However, they have not conducted experiments to evaluate the effectiveness of their proposal. Grimaldi *et al* have developed a method to capture user preferences based on features effective for genre classification[4], which has achieved MIR accuracy competitive to our results in [1].

3. PROBLEMS

While our genre-based MIR method is capable of achieving MIR with sufficient accuracy, the diversity of music genre distribution of the songs in the music collection is assumed to apply a negative effect to the MIR results.

In order to generate a feature space by the TreeQ algorithm, a training data set of songs with appropriate genre information is necessary. In our previous experiments in [1], the "RWC Music Database: Music Genre"[5] was used as training data. The RWC Database consists of songs from a wide variety of genres, such as Pop, Rock, Classical, Traditional Japanese, etc. Therefore, the resulting feature space is expected to be a space where songs from various genres are
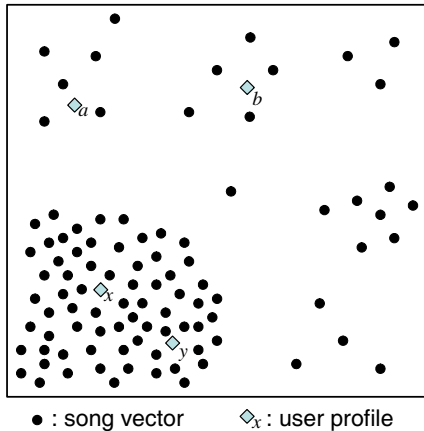
: song vector ◇$_x$: user profile

Fig. 1. Conceptual illustration of biased feature space and song vectors

evenly distributed.

However, it is unrealistic to expect any music collection to be composed of an equal amount of songs per music genre. If anything, the genre distribution of songs in a typical music collection is assumed to be focused on a small range of genres, which represent the taste of the collection's owner. Therefore, vectors of songs within the collection are expected to be biased towards whatever genre the collection is focused on.

A conceptual illustration of a biased music feature space is shown in Figure 1. If the user profile is located in a low-density area, as in user profiles a and b in Figure 1, it is fairly easy for the MIR system to retrieve songs which match user preferences, since the features of such songs are distinctive from most of the other songs in the music collection. However, if the user profile is located within the dense portion of the feature space, as in user profiles x and y, it becomes more difficult for the MIR system to achieve accurate MIR, since the songs are densely clustered within the vicinity of the user profiles, and are thus much more difficult to distinguish from each other.

4. CLUSTER-BASED FEATURE SPACE MODIFICATION

In order to overcome the problems described in the previous section, we have developed a feature space modification (FSM) method which improves accuracy of user preference based MIR, regardless of the genre distribution of songs in the music collection.

The general flow of the proposed FSM method (hereafter referred to as the "cluster-based FSM" method) is illustrated in Figure 2. First, an initial VQ tree is generated based on a music genre database, and all songs in the music collection are vectorized based on the resulting VQ tree. Next, all song vectors are clustered by the K-means clustering algorithm.

Table 1. Summary of user rating data

Category	Rating	Ratio(%)
C_g	5	15.2
	4	19.5
C_f	3	32.1
C_b	2	21.8
	1	11.5

The resulting cluster IDs, and song vectors which belong in each cluster, are then used as training data to generate a new, "cluster-based" VQ tree. Namely, the training data set is generated by extracting M songs from each cluster ($C_1, \cdots, C_n$ in Figure 2), which are closest to each cluster centroid. As a result, the feature space is modified according to the songs in the music collection. Finally, all songs in the music collection are re-vectorized based on the cluster-based VQ tree.

Cluster-based FSM solves the problems described in the previous section. Since the proposed method conducts clustering regardless of the genre distribution of the songs in the music collection, the method is robust to diverse genre distributions. For example, if a large majority of songs in the music collection belong to a single genre, such as "Pop," song vectors generated by the genre-based method will be biased towards the "Pop" vicinity of the feature space, as in the example shown in Figure 1. By applying cluster-based FSM, the VQ tree can automatically create sub-genres within the majority "Pop" genre. This enables extraction of distinctive features of songs, which cannot be sufficiently detected in the conventional feature space.

5. EXPERIMENTS

This section describes evaluation experiments conducted to compare the accuracy of the cluster-based FSM method with the conventional genre-based MIR algorithm.

5.1. Experiment data

The music collection for the MIR system is constructed by combining the experiment data set used in the experiments in Reference [1], with the CDs listed in the *uspop2002* music data set constructed by Ellis[6]. The total number of songs in the music collection is 6863.

Experiment data for our MIR experiments were collected by inviting 19 subjects to apply subjective ratings to the songs in the above music data set, ranging from 1 to 5 (Bad:1 $\sim$ Good:5). The rated data was then classified into three categories, according to the subjective ratings: "good songs" (C_g), "bad songs" (C_b), and "fair songs" (C_f). Categories C_g, C_b, and C_f consist of songs rated (4 or 5), (1 or 2), and 3, respectively. Table 1 shows the average ratio of songs per rating for all 19 subjects.
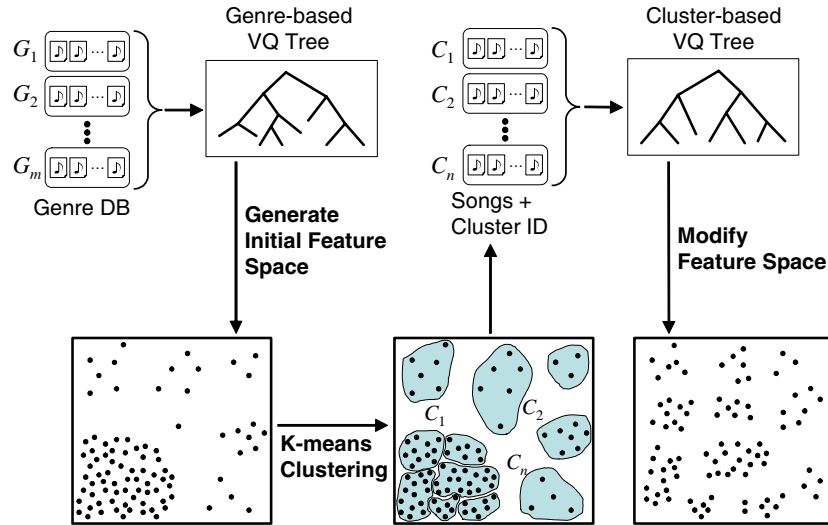
Fig. 2. Outline of cluster-based feature space modification method

5.2. Experiment method

First, in order to generate user profiles for each subject, N songs are randomly selected from the good-rated songs (*i.e.*, songs which belong to category C_g) of each subject. Next, two user profiles are generated based on the selected N songs. One profile is generated by inputting the N songs to the genre-based VQ tree, which is generated by using the RWC Database as training data. The other profile is generated by inputting the N songs to the cluster-based VQ tree. Two sets of song vectors are also generated for the songs in the music collection, by using the genre-based and cluster-based VQ trees.

Next, scores of all songs in the music collection are calculated based on the vector similarity between the song vectors and user profiles, for both methods. All songs are sorted according to the calculated scores, and the top 50 songs are extracted as MIR results. The ratio of songs which belong to categories C_g, C_f, and C_b within the extracted MIR results are then calculated, and used as the evaluation measure of MIR accuracy.

5.3. Results

First, in order to examine whether or not songs in a music collection are biasedly distributed within the genre-based feature space, we conducted a simple music categorization experiment to categorize all songs in the experiment music collection to the genres in the RWC Database, and examine the distribution of the genre categorization results.

Categorization of all songs in the music collection are conducted according to the method described in Reference [2], where vector similarity between the song vector and all category vectors are calculated, and each song is classified to the category with the highest similarity to the song. Results of this genre categorization experiment showed that 70.3% of
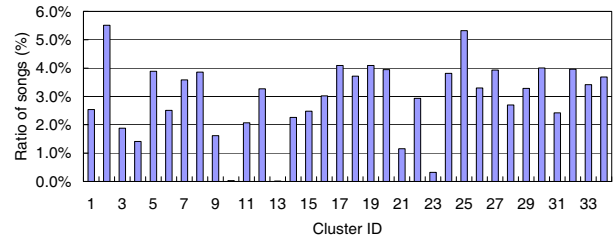

Fig. 3. Ratio of songs categorized to each cluster ($K = 34$)

the songs in the experiment data set were categorized to the "Pops" genre of the RWC Database. This result proves that the distribution of vectors of songs in a music collection can be heavily biased within the feature space generated based on genre information.

The same categorization experiment was conducted for the cluster-based FSM method, by generating a vector for each cluster, and categorizing all songs to a specific cluster, following the categorization method mentioned above. For fair comparison to the genre categorization experiment, the number of clusters for this experiment is set to 34, which is equal to the number of categories in the previous experiment.

Figure 3 illustrates the ratio of songs categorized to each cluster. Results in this Figure show that no more than 6% of all songs are categorized to a specific cluster. Therefore, the distribution of song vectors generated by the cluster-based FSM method is clearly more significantly dispersed compared to the conventional feature space. This result proves that, as hypothesized in Section 4, the proposed FSM method is capable of creating a feature space where song vectors are evenly distributed, regardless of the genre distribution of the songs in the music collection.

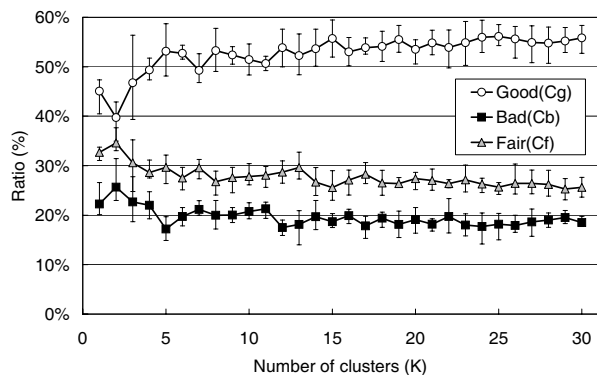Next, we compare MIR accuracy of the two methods.

Fig. 4. Average, maximum and minimum ratio of songs of categories C_g, C_f, and C_b in MIR results for various number of clusters

Experiments were conducted for the cluster-based method with the number of clusters (K) ranging from 1 to 30. In order to compensate the randomness of the data extracted for user profile generation, we conducted five sets of experiments for each parameter value of K. For these experiments, the number of songs selected from each cluster (M) as training data for cluster-based VQ tree generation is fixed to 3.

Figure 4 illustrates the average, maximum, and minimum ratio of songs in categories C_g, C_f, and C_b, for all experiments conducted per K. Note that the results of the conventional genre-based method are illustrated at the leftmost point of the graph in Figure 4.

As clear from Figure 4, the ratio of songs in category C_g has significantly improved by applying the cluster-based FSM method, with the exception of the results when $K = 2$. Namely, the average ratio of songs in C_g has improved from 45.1% for the genre-based method to 56.0%, which was achieved for the cluster-based FSM method when $K = 15$. Furthermore, the ratio of songs in C_b has also decreased for the proposed method, compared to the results of the genre-based method. This indicates that the proposed cluster-based FSM method has improved MIR accuracy.

Further analysis on the results in Figure 4 shows that the ratio of good songs in the MIR results gradually increases in proportion to the increase of K. However, since the minimum cluster size (*i.e.*, the number of songs in a specific cluster) is extremely small when K is set at a large value (for example, the minimum cluster size when $K = 25$ is 15, *i.e.*, 0.2% of all songs in the music collection), the constant improvement of C_g ratio is assumed to be an effect of overfitting to the experiment music collection.

Overall, the above experiment results show that the proposed cluster-based FSM method is capable of achieving highly accurate MIR based on user preferences, and is superior to the conventional genre-based method, proving the effectiveness of the proposed method.

6. CONCLUSION

In this paper, we have proposed a feature space modification (FSM) method for content-based music information retrieval (MIR) based on user preferences. The proposed cluster-based FSM method utilizes clustering results of songs in the music collection to reconstruct the feature space. The proposed method is superior to the conventional genre-based method, since the feature space dynamically adapts to genre distribution of any music collection. Results of evaluation experiments conducted on a music collection with user ratings have proved that the cluster-based FSM method contributes for improvement of user preference based MIR accuracy.

This research contributes to the development of an effective MIR system, which encourages discovery of unknown songs from a large scale music database, and also enables users to efficiently find preferable songs from their personal music collections. We expect that development of such systems will further promote the prosperity of current music distribution services and applications.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] K. Hoashi *et al*: "Personalization of user profiles for content-based music retrieval based on relevance feedback", Proc. of ACM Multimedia 2003, pp 110-119, 2003.

[2] J. Foote: "Content-based retrieval of music and audio", Proceedings of SPIE, Vol 3229, pp 138-147, 1997.

[3] W. Chai, B. Vercoe: "Using user models in music information retrieval systems", Proc. Int'l Symposium on Music Information Retrieval, 2000.

[4] M. Grimaldi, P. Cunningham: "Experimenting with music taste prediction by user profiling," Proc. of 6th ACM SIGMM Int'l Workshop on Multimedia Information Retrieval, pp 173-180, 2004.

[5] M. Goto *et al*: "RWC Music Database: Music Genre Database and Musical Instrument Sound Database," Proc. of the 4th Int'l Conference on Music Information Retrieval (ISMIR 2003), pp.229-230, 2003.

[6] D. Ellis: "The *uspop2002* Pop Music data set", List available at http://www.ee.columbia.edu/%7Edpwe/research/musicsim/uspop.html, 2003.

[7] J. Foote: "TreeQ software," http://treeq.sourceforge.net/