Automatic Music Recommendation System based on Music Characteristics

Sangho Kim*, Sungtak Kim*, Suk-bong Kwon*, Mikyong Ji*, Hoirin Kim*,
Jeong Hyun Yoon**, Han-kyu Lee**
* School of engineering, Information and Communications University (ICU)
** Electronics and Telecommunications Research Institute (ETRI)

Abstract In this paper, we present effective methods for automatic music recommendation system which automatically recommend music by signal processing technology. Conventional music recommendation system use users' music downloading pattern, but the method does not consider acoustic characteristics of music. Sometimes, similarities between music are used to find similar music for recommendation in some method. However, the feature used for calculating similarities is not highly related to music characteristics at the system. Thus, our proposed method use high-level music characteristics such as rhythm pattern, timbre characteristics, and the lyrics. In addition, our proposed method store features of music, which individuals queried, to recommend music based on individual taste. Experiments show the proposed method find similar music more effectively than a conventional method. The experimental results also show that the proposed method could be used for real-time application since the processing time for calculating similarities between music, and recommending music are fast enough to be applicable for commercial purpose.

Keyword: Music recommendation, Music characteristics, Rhythm feature, Timbre feature, the lyrics

1. Introduction

Recently, digital music is moving into the mainstream of consumer life. Sales of single track download in the US in 2004 rose to 142.6 million from 19.2 million in the second half of 2003 [1]. The digital music market is rapidly growing. As such, there has been great importance placed on efficient management of numerous digital music databases. In addition, people want personalized music services, nowadays. Automatic music summarization, music retrieval, and recommendation system are those kinds of services. In case of music recommendation system, the system usually use users' downloading pattern to recommend music to other customers. If some group has similar music downloading pattern, the music, to which someone in the group listen, can be recommended to other person in the same group because it is assumed that they have similar fondness if they have similar music downloading pattern. However, the method does not consider music characteristics. Thus, it is difficult to reflect user's taste on each song in the system. So, some system uses similarity between music to consider more song dependent recommendation [2]. However, the method use just fundamental timbre feature. So, it is difficult to calculate music-level similarity. Therefore, we use several things such as rhythm, timbre, and the lyrics to calculate music-level similarity in this work. In the aspect of music psychology, rhythm is the most fundamental factor in classifying the mood of music [3]. If we use this characteristic when we calculate similarity between music for recommendations, it will be more effective than just considering fundamental timbre feature. In addition, we consider similarity between words of songs because two songs could be regarded as dissimilar music although acoustic similarity between two music is very high because human could be dependent on the lyrics when they determine whether two music are similar or not. Thus, we also use the lyrics to calculate similarity between music by using simple natural language processing technique. And, the feature vector of each music, which user queried, are stored individually to recommend music based on individual fondness in the proposed system. Finally, we use an objective measure to evaluate the proposed methods. The results show that the

proposed method is relatively good in finding similar music, and the proposed method is applicable to real-time application.

2. Fundamentals

In this section, the methods how the features can be extracted are discussed. As it is explained, the rhythm pattern is important for human to determine the mood of music. Of course, melody or modality of music also affects the mood of music. It seems that the degree of influence of several factors for determining the mood of music is not well-generalized until now. However, it is generally believed that that the rhythm pattern is more important than other factors when human recognize the mood of music. By this reason, the rhythm pattern is used in this scheme. Of course, the timbre feature is also used for constructing a complete feature set. It is clear that much additional work will be required before a complete understanding of human perception how they determine the mood of music in the area of music psychology. Finally, words of songs are also important to determine the mood or style of music. Thus, similarity between the lyrics is also considered in this scheme by using simple natural language processing technology (NLP). The approach using NLP in this scheme might be a good trial to enhance performances although the technique is simple.

2.1 Rhythm feature extraction

To extract rhythm feature, Alonso's tempo estimation algorithm, which shows very good performance, is used and somewhat modified in this work [4]. The algorithm consists of several main modules which are onset detection and periodicity estimation. Firstly, the input audio signal is analyzed using the short-time Fourier transform (STFT). Then magnitudes on each frequency index along whole frames are compressed logarithmically. Secondly, magnitudes on each frequency index are filtered along time axis using low pass filter. After that, canny operator is used for detecting onset. The operator is famous for an edge detection tool in image processing area. Finally, the onset curve of each frequency index is summed and used to represent the rhythm information of a music clip. From the curve, onsets can be chosen from the peaks which are larger than some threshold which is set to 0.0 in our experiment. Then the final onset curve is used for median filtering. The final onset sequence is used for constructing rhythm feature. The first dimension is average of onset values, and the second dimension is the frequency of onset. After median filtering, the onset sequence can be used for the autocorrelation method to find periodicity of the sequence. Figure 1 shows the result of autocorrelation of onset sequence. As it can be seen, the example shows that rhythm pattern of the audio signal is not quite explicit because the second peak is not as big as other peaks. Thus, the music signal might be regarded as a kind of smooth music. However, we do not classify the mood of signal as a human language like 'exuberance', or 'contentment'. Just numerical values for the feature are extracted. So, the third dimension of the rhythm feature is average value of the largest peaks, and the fourth dimension of the feature is average value of the smallest valleys. To calculate each average value, four largest peaks and four smallest valleys are chosen.

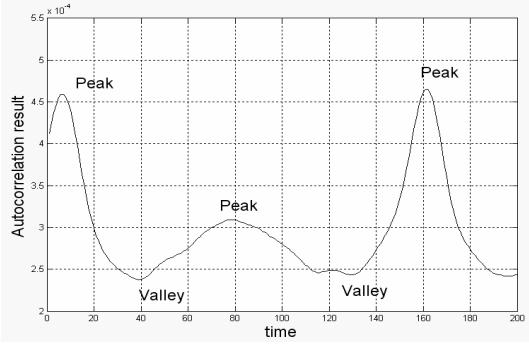

Figure 1 Examples of autocorrelation curve corresponding to the onset sequence

2.2 Timbre feature extraction

We extract features from acoustic music signals. The process of feature extraction is illustrated in Fig. 2. Mel-Frequency Cepstral Coefficient (MFCC) is well-known for speech and audio signal processing. Other features such as spectral contrast and shape features [5], [6] are also used for music signal processing. The spectral contrast feature may be more suitable for music signal processing than MFCC and the octave scale filter bank is also frequently used. But it depends on the application. In our experiments, MFCC was good enough to analyze the similarity and dissimilarity between signals. And, linear predictive cepstral coefficient (LPCC) was better in discriminating delicate differences of timbre. However, our focus is not on the differences of delicate timbre but on the explicit differences of music. Thus, MFCC is more reasonable. So, we use only MFCC features in this scheme. The number of bands is 40, but just 24 coefficients from the lowest order are taken to get smoother pattern of timber of music signal. Then mean and variance values of each dimension are calculated and stored as a feature vector.

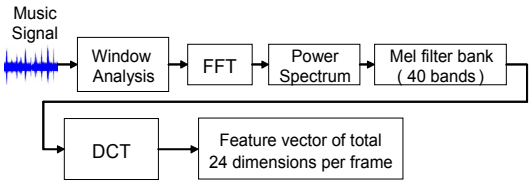

Figure 2 Basic process of the feature extraction

2.3 Similarity between the lyrics

In natural language processing area, so many researches were conducted to calculate the similarity between words [7]. To calculate the similarity, constructing word co-occurrence vectors is fundamental. After constructing the vector, it is possible to use the several distance measure such as cosine measure and Min/Max measure. For other application, distance between articles can be calculated mathematically [8]. The method also uses vector form to characterize the meaning of articles. If it is possible to construct feature vectors to represent the character of words or article, calculating distance between them is not difficult thing. To construct a word co-occurrence vector, it is needed to determine a format of the feature. The elements of the concrete format are called 'attribute'. For example, we could select three attributes such as <run, subject>, <run, object> and <eat, subject>, that is, att1=<run, subject>, att2=<run, object> and att3=<eat, subject>. Beside these, it is possible to select various combinations for attributes. Then, we could generate a vector of a word 'u' like <freq(u, att1), freq(u, att2), freq(u, att3)>. 'freq(u, att)' means that the number of frequency of co-occurrence of the word, u, with the attribute. On the other hand, the word co-occurrence vector could be constructed differently. The vector could have different form like <assoc(u,att1), assoc(u,att2), assoc(u,att3), …> The 'assoc(u,att)' means 'association', and shows how much the word 'u', and the attribute 'att' have strong or weak relation have. To simply calculate, assoc(u,att) could be freq(u,att). The different form of the association can be possible. By using this way, it is possible to construct a feature of special word, and then calculating distance between words is simple task. Likewise, to calculate distance between the words of songs, a similar approach is used, but the predefined attribute is not used in this scheme. Firstly, if song A has N different words, word occurrence list, which have information such as words and its occurrence time in the lyrics, is constructed like as shown in Table 1. Likewise, the word occurrence list of song B can be constructed. Then it is possible to calculate the distance between the lyrics of song A and B by using some equation. The equation can be set as

Table 1

Words in Song A	The number of word occurrence
Word 1(ex, love)	3
Word 2(ex, forever)	2
Word…	…
Word N(ex, peace)	8

$$sim(u,v)=\frac{1}{N_{SW}}\sum_{SW}\frac{\min[freq(u,SW),freq(v,SW)]}{\max[freq(u,SW),freq(v,SW)]}+N_{SW} \quad (1)$$

where, sim(u,v) is the similarity between the lyrics of song u and song v, SW means the same words occurred in both song u and v except reject words which are predefined words such as 'is', 'are', 'on' and something like that, N_{SW} is the number of same words occurred in both song u and v, freq(u, SW) is the number of occurrence of the word, SW in song u, the min[a,b] operation returns smaller value between a and b, and the max[a,b] returns greater value between a and b. The summation of the min/max operation is normalized by the N_{SW}. As a result, the left side term of the plus sign will be ranged from 0 to 1. If the lyrics of two songs are exactly same, the value will be 1. However, N_{SW} is added to the value because the number of same words in both songs is emphasized in this work. Thus, it is not difficult task to calculate distance or similarity between the lyrics of songs by this scheme. Of course, high-level NLP techniques such as semantic representation, machine translation and pragmatic translation of sentences were not used in this experiment. The approaches could be adopted in later works with cooperation of other NLP related laboratories.

3. Proposed method

3.1 Overall system

The proposed system could be organized as shown in Fig. 3. Firstly, client can push a button which is to represent user's fondness corresponding to the music user is listening in the present. Then information of the queried music such as song title, genre information, user ID can be transmitted to the service provider. The service provider finds the feature vector of the music, and select several test songs according to the information of the requested music in recent music database to calculate similarity between music. When calculating similarity between the feature vectors of music, both the feature vector of requested music, and the feature vector of user history are used to calculate the similarity. Thus, two sets of recommended lists are got from the results. Each category has two categories such as singer-based recommendation list, and genre-based recommendation list, which means that only reference songs in recent music database, which are same singer or genre, are selected to calculate the similarity. The singer or genre information can be known from the user query as previously explained. This overall system can be slightly modified by service providers. However, the basic process is not changed according to the system.

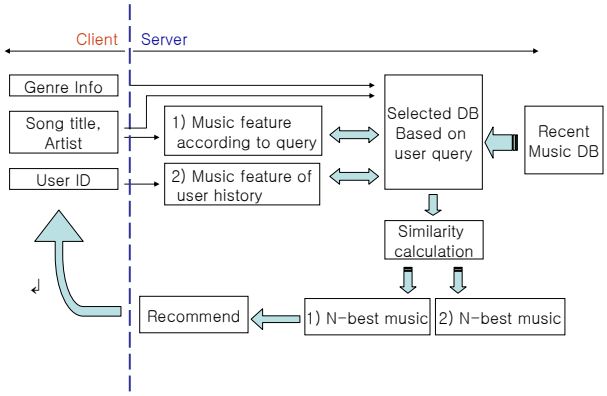

Figure 3 Graphical explanation of recommendation system

3.2 Similarity calculation

Firstly, the rhythm feature, and the timbre feature of all music are extracted as explained in previous chapter. The dimension of the rhythm feature is 4, and the dimension of the timbre feature is 48 because mean and variance values exist on each dimension. Thus, total 52 dimension vector are constructed at each song. Then it is possible to use several distance measures. In this scheme, we used the Euclidean distance measure. It is defined as follows. The measure is very simple but it is easy to emphasize the differences between each numerical values of each dimension. As a result, it is possible to get several distance values from the calculation, and then the values could be sorted by ascending order. We could select just a couple of songs from the result. In this experiment, just 4 songs are selected.

$$S_e(i,j) = \| v_i - v_j \|^2 \quad (2)$$

After finding songs from the acoustic similarity, the lyrics similarity is calculated as previously explained if the lyrics of the user queried music exists. Then it is possible to find the most similar music to the user requested music.

3.3 User history

The feature vectors of user requested music are stored individually because the proposed system aims to analyze and store the pattern of individual fondness. By using the history, we could recommend music based on his or her own style. To store the history, the feature vectors are stored at each genre category. For example, if user id is HCI2007, the feature vectors of the rock music requested are stored at a text file named HCI2007_rock. Likewise, the history for other genre could be stored. The text files include the number of queried music, and accumulated feature vectors. Firstly, the number of queried music is needed to accumulate the feature vectors. The calculation for newly history is as follows.

$$F_{NEW}(i) = (F_{OLD}(i) \times N_{QUERY} + F_{RECENT}(i)) / (N_{QUERY} + 1) \quad (3)$$

where, $F_{NEW}(i)$ is i-th feature value of the newly stored history, $F_{OLD}(i)$ is the i-th feature value of the previously stored history, N_{QUERY} is the number of queried music, $F_{RECENT}(i)$ is the i-th feature value of recently queried music. Suppose that the number of queried music is 3 in 'HCI2007_rock.txt" file, and feature values are 11, 12, and 31. And, suppose that the feature values of recently queried music are 10, 21, and 20. Then feature values of the newly stored history are (11*3+ 10)/4, (12*3+ 21)/4, and (31*3+ 20)/4. Of course, the number of dimension of the feature could be different, but the calculation process is exactly same as explained in the example.

4. Experimental results and discussion

To evaluate the proposed method, objective, and subjective test are conducted. For objective test, we categorized the groups of similar music, and the ground-truth data are used for evaluation. For subjective test, the degree of similarity with the reference music is assigned ranged from 3 to 1. The higher value means that there is more similarity between music. Actually, recommendation is very difficult task to evaluate the performance. Human being is not stable existence in the aspect of emotion although it depends largely on individuals. In addition, environmental situation can affect human feelings, and then it affects the recommendation performances. These kinds of problems could be interesting research topic with cooperation with other academic fields such as music psychology, cognitive science, and music therapy. However, we just adopted conventional evaluation methods in this experiment. The evaluation methods might not be the best solution for evaluation, but we could recognize the proposed method is relatively good at finding similar music compared to the method using just a kind of the feature like timbre. Firstly, 10 test songs in each genre group are selected as a query, and 100 songs are used for the similarity calculation. Then 4 songs which have the largest similarity values are selected. After that, each song is assigned by a value ranged from 3 to 1. Then the values are averaged. Table 2 shows the results. In this experiment, the similarity between the lyrics of songs is not considered. Only acoustic similarity by rhythm pattern feature and timbre feature is used to know the effect of using two different features. The results show that using two different features is better than using just timbre feature. Of course, it is clear that

using several features might affect the performance positively, but it is important to know what kinds of feature will be suitable for finding similar music. Thus, it could be possible to know that using rhythm pattern feature give positive results for the performance as the feature is already verified as an important factor for human when they recognize music.

Table 2

	Rock	Dance	Ballad
Proposed	2.05	1.875	2.475
Timbre	1.625	1.5	2.075

Secondly, we considered the lyrics similarity with the acoustic similarity. 10 songs are randomly selected as query songs, and there are randomly selected 30 reference songs in database for the similarity calculation. All songs have text files of the lyrics. And, similar music is categorized in same group for evaluation. After finding 10-best similar music using the acoustic similarity, the lyrics similarity is used for finding 4-best similar music. Then the number of songs which are included in same group with a query song is counted. Table 3 shows the results.

Table 3

Test song index	Timbre	Proposed
1	2	3
2	1	3
3	1	3
4	2	2
5	1	1
6	0	1
7	2	2
8	1	3
9	1	2
10	1	1
Total (ea)	12	21

As it can be seen, the proposed method is relatively better than using just timbre feature. Actually, in this experiment, it is not easy to know the effect of using the lyrics similarity. However, other experiments also show that the effect of using text for similarity calculation do not have much impact on the performance. There are several reasons. Firstly, the lack of various songs could be the first reason. We tried to get various songs which have diverse contents in the aspect of the lyrics, but it was not easy task because the content of the lyrics of song is often abstract and ambiguous, and difficult to understand. So, semantic or pragmatic understanding of the lyrics was not easy task. Thus, grouping songs was difficult, but we simply categorized songs by the occurrence frequency of words. If the word, 'love' is frequently occurred in the music, the music is clustered in the love song category. Secondly, the method for similarity calculation between the lyrics used just the word occurrence. Machine translation or semantic or pragmatic translation was not used. Thus, the performance was not good as expected. However, those approaches such as machine translation are needed to use high-level knowledge and know-how related to the natural language processing, but the researches are out of scope in this scheme. Of course, we can try those approaches in later works. Despite that, the similarity between the lyrics of music could be effective to avoid bad recommendation results although it does not use high-level NLP technologies because if a song has the word 'love' several times, the probability which recommended songs could have words such as 'power', 'hate' and 'kill' will be low.

5. Conclusion and future works

We proposed automatic music recommendation system. The system or method use rhythm pattern feature and timbre feature for acoustic similarity between music. In addition, the methods use the lyrics similarity to consider the similarity between words of songs. The results show that using two acoustic features is better than using one kind of feature like timbre. And, considering the lyrics similarity could be helpful to avoid bad recommendations. As previously explained, it is needed to study how human can get good feelings from music recommendation results, and what kind of recommendation is effective. In addition, we need to devise different feature extraction and similarity calculation methods. Nowadays, music recommendation on the web is very active field [9], [10]. It is hoped that this proposed system could stimulate further investigation in this field.

References

[1] International Federation of the Phonographic Industry (IFPI), 2005 Digital Music Report.

[2] Beth Logan, “Music recommendation from song sets," *ISMIR 2004*, Universitat Pompeu Fabra, Barcelona, Spain, October 10-14, 2004.

[3] Rudolf E. Radocy and J. David Boyle, *Psychological foundations of musical behavior.* Charles C. Thomas Publisher, Ltd.

[4] Miguel Alonso, Bertrand David and Gael Richard, "Tempo and beat estimation of musical signals," In Proceedings of ISMIR 2004, Barcelona, Spain, October 2004.

[5] Dan-Ning. Jiang, Lie. Lu, Hong-Jiang Zhang, Jian-Hua Tao and Lian-Hong Cai. “Music type classification by spectral contrast feature,” In Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), Lausanne (Switzerland), August 2002.

[6] Lie Lu, Dan Liu and Hong-Jiang Zhang, “Automatic mood detection and tracking of music audio signals,” IEEE transactions on audio, speech, and language processing, Vol. 14, No.1, January 2006.

[7] Robert Dale, Hermann Moisl, and Harold Somers, *Handbook of Natural Language Processing*, Marcel Dekker, Inc.

[8] Michael Fleischman, and Eduard Hovy, "Recommendations Without User Preferences: A Natural Language Processing Approach," IUI'03, January 12-15, 2003, Miami, Florida, USA.

[9] http://www.owlmm.com/

[10] http://foafing-the-music.iua.upf.edu/index.html