Music Recommendation Based on Feature Similarity

Huihui Han
School of Computer Science & Technology
Donghua University
Shanghai, 201620, P.R. China
18302166037@163.com

Xin Luo
School of Computer Science & Technology
Donghua University
Shanghai, 201620, P.R. China
xluo@dhu.edu.cn

Tao Yang *
Network Education College
Donghua University
Shanghai, 201620, P.R. China
yangtao@dhu.edu.cn

Youqun Shi
School of Computer Science & Technology
Donghua University
Shanghai, 200051, P.R. China
yqshi@dhu.edu.cn

***Abstract*—As the information of online music resources continues to grow, it becomes more and more difficult for users to find their favorite music. Accurate and efficient music recommendation is very important. Music recommendation is a research hotspot in the field of speech processing. Calculate the mel frequency cepstral coefficient (MFCC) feature quantity by analyzing the characteristics of music content. Then, the feature quantities are clustered to compress the music feature values. Finally, the distance metric function is used to calculate the similarity between all music in the feature value database of the searched music. The closer the distance is, the higher the similarity is. According to the similarity, we can get the result of recommendation. The method recommended results have higher accuracy in experiments and provides an idea for music recommendations when user data is missing.**

Keywords—music recommendation, MFCC, K-means clustering, vector quantization, EMD

I. Introduction

Music is a necessary form of artistic expression in people's lives, and music-related industries are rapidly developing due to the convenience of the internet. Therefore, managing and searching for songs has become significant. Though music information retrieval techniques have been made many achievements in last ten years, the development of music recommender systems is still at a very early stage [1]. As early as 1995, the social information filtering system Ringo provided a personalized recommendation function for music. Users express their interest by rating music, and the system made them recommendation according to the interests of similar users. The recommended results were metadata such as music and artists [2]. There are three main methods for music recommendation, metadata-based recommendations, collaborative filtering, and content-based recommendations. The results of the metadata-based recommendation method are relatively simple and lack of pertinence; collaborative filtering requires a large amount of user data, and matrix sparsity problems occur when the amount of data is large; the content-based recommendation is based on audio and requires less user data. In this paper, we propose a music recommendation method based on content similarity, which is used to make accurate recommendation based on audio content similarity when user data is lacking.

978-1-5386-5514-6/18/$31.00 ©2018 IEEE

II. Literature Review

The content in content-based music recommendation refers to the rhythm, melody, and rotation, accompaniment and even timbre of music [3]. Reference [4] mentioned an audio retrieval method, where the audio fingerprint is obtained from the music spectrum. Reference [5] focused on optimization vector content feature for the music recommendation system, they created a database consisting of excerpts of music files, and then optimized using correlation analysis and Principal Component Analysis. In reference [6], they selected time energy, frequency energy, MFCC (mel frequency cepstral coefficient) and spectral envelope as the music features, these four features were integrated as the feature vectors that were calculated the fractal dimension by Hilbert transform. This method focuses on the certain degree of self-similarity between the whole and the local area. Reference [7] proposed a music retrieval method based on MFCC features, which is based on human hearing. There are also other ways to recommend music, such as collaborative filtering and machine learning. Reference [8] proposed a music recommendation mechanism to calculate the characteristics of songs accepted or rejected by the user through the Naive Bayes classifier, and continuously learn and update the classifier to obtain the recommendation results in accordance with the prescribed style. An audio fingerprint coding method based on signal power spectrum features is proposed as in [9], which has strong anti-noise ability.

There are some music recommendation methods in addition to content-based recommendations. In reference [10], they proposed a content-based music recommender system for the cold start problem. This system is based on a set of attributes derived from psycho-logical studies of music preference. The results demonstrate the effectiveness of these attributes in music preference estimation. Music classification is useful for music recommendations. Reference [11] described a technique

that uses support vector machines to classify songs based on features using Mel Frequency Cepstral Coefficients. Experimental results of multi-layer support vector machines shows good performance in musical classification. Reference [12] proposed a hybrid method to seam lessly integrate the automatically learnt features and collaborative filtering. The results show that the method significantly improved the performance of collaborative filtering. Since the process of listening to music is closely related to emotions, reference [13] proposed a music recommendation method that incorporating the user's emotional factors, which added the recommendation results of emotional classification and user emotional matching. The audio fingerprint extraction method based on time-frequency analysis only takes signal energy as a parameter, which can't fully characterize the complexity and irregularity of the signal. Therefore, an audio fingerprint extraction method based on wavelet packet transform is proposed as in [14].

In this paper, music recommendation is based on content similarity method, and the content is represented by MFCC feature value. The recommendation process is roughly divided into three steps: first, the music feature vector is calculated, then the distance between the vectors is calculated to represent the similarity of the song, and finally the recommended result is obtained.

III. Recommendation Algorithm Based on MFCC Feature

The recommendation algorithm based on MFCC feature similarity belongs to content-based recommendation, and the content is quantized by MFCC feature values. The recommendation process is shown in Fig.1. First, we collected six types of music including: happy, quiet, romantic, sad, inspirational and excited, which constitute the basic database. In this paper, we extract the MFCC feature quantity, cluster it, and obtain the feature value of the song. Then, we put the metadata such as the singer and the music name into the database, together with their feature value. When the user inputs the audio to be recommended, the same feature value is extracted, and the distance is compared with the existing music in the database. If the absolute value is relatively small, it means that the two musics are similar in content, that is, the recommendation according to the input music.

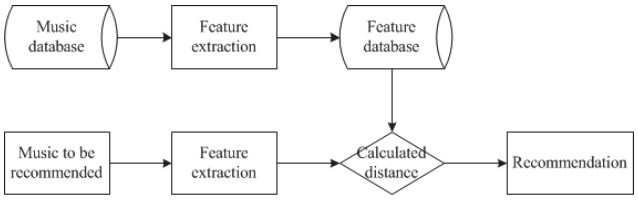

Fig. 1. Content-based music recommendation flow chart

A. Extract MFCC Feature

MFCC is a speech feature extracting technology. The extracting process is shown in Fig.2. The steps are as follows:

1) Intercept fragment: The duration of the song is generally 3-5 minutes, and there is often some duplicates of content. The calculation of the feature value of the whole song will be very large, which is unnecessary. Pampalk et al. [15] proposed to select only 30 seconds of the song center as the processing object. In this paper, only 30 seconds of the song is selected for analysis.

2) Pre-emphasis: In order to compensate the high-frequency portion where the speech signal is suppressed, the high-frequency resonance peak is highlighted, and the pre-emphasis filter differential speech signal is used to amplify the high-frequency portion. The pre-emphasis filter is defined as shown in (1), and p is the pre-emphasis coefficient.

$$S_n^{'} = S_m - p \times S_{n-1} \tag{1}$$

3) Framing and windowing: The audio signal usually changes constantly. To simplify the calculation, it is assumed that the audio signal does not change much in a short time, that is, within one frame. In order to reduce the spectral energy leakage, a Hamming window is added for each frame of the signal to perform discrete Fourier transform.

4) Mel filter bank: The Meyer scale reflects the auditory characteristics of the human ear. The lower the frequency, the narrower the interval, the higher the frequency and the wider the interval. Usually 20-40 triangle filters are applied to the power spectrum. The amplitude of each frame is separately multiplied by each filter, and the obtained value is the energy value of the frame data in the corresponding frequency band of the filter.

*5) Discrete cosine transform(DCT): T*ake the logarithm of the obtained energy value, use the discrete cosine transform to directly obtain the low frequency information, remove the correlation between the various dimensional signals, and map the signal to the low dimensional space. What obtained is the MFCC feature.

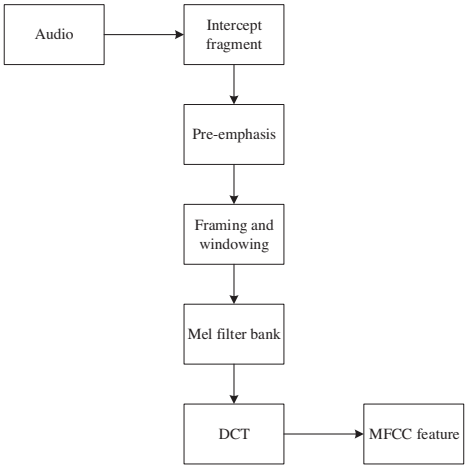

Fig. 2. MFCC extraction flow chart

B. Vector quantization

The Vector Quantization technique is a data compression and coding technique that combines several scalar data into a vector and then quantizes it in the vector space, which can

compress the data while ensuring that the information is not lost. Linde, Buzo and Gray proposed a training sequence based VQ design algorithm, called LBG-VQ algorithm. The algorithm generates m code vectors as follows:

1) Randomly select m vectors from the feature quantity, and set them as the initial code vector;

2) For each vector in the feature quantity set, find the nearest code vector according to the Euclidean distance and classify it into the cluster of the code vector;

3) Calculate the center vector of each cluster and use it as the new code vector of the next iteration, repeating the second step;

4) If the total average distortion of the two iterations is less than the threshold, or reaches the upper limit of the number of iterations, the iteration is stopped output the codevector.

After vector quantization, the feature vectors are divided into m classes, and the audio feature values are calculated. The eigenvalue *P* consists of three parts: the average vector of each classification μ_{p_i}, the covariance of each type $\sum p_{m_i}$, and the number of vectors in each category as weight w_{p_i}, in the form of:

$$P = \left\{ \left(\mu_{p_1}, \sum p_1, w_{p_1} \right), \ldots, \left(\mu_{p_m}, \sum p_m, w_{p_m} \right) \right\}$$

Through the K-means clustering, the distance between vectors is K-L distance, and the final music feature value *M* is obtained, as shown in (2), where *m* is the number of classifications, A_i is the mean vector of each category of the eigenvalue *P*, σ_i Is the covariance matrix of *P*.

$$M = mA_i\sigma_i \tag{2}$$

C. Vector quantization

After the feature vector library is established, the distance between the songs is quantized using the EMD (Earth Mover's Distance) algorithm. The smaller the distance, the closer the similarity is. The EMD distance is the minimum cost of distributing from one distribution to another under some constraints and can be used to represent similarities between two multidimensional distributions.

The feature quantity of the song P is $P = \left\{ \left(M_{p_1}, w_{p_1} \right), \ldots, \left(M_{p_m}, w_{p_m} \right) \right\}$, where M_{p_i} is the characteristic of the classification, w_{p_i} is the weight of the class, and similarly, $Q = \left\{ \left(M_{q_1}, w_{q_1} \right), \ldots, \left(M_{q_m}, w_{q_m} \right) \right\}$ is the feature quantity of the song *Q*. The ground distance matrix $D = d_{ij}$ represents the ground distance between M_{p_i} and M_{q_j}, and the minimum cost function of global consumption is as shown in (3):

$$W = \sum_{i=1}^{m} \sum_{j=1}^{m} d_{ij} f_{ij} \tag{3}$$

Which should meet the following constraints:

$$f_{ij} \geq 0 \tag{4}$$

$$\sum_{i=0}^{m} f_{ij} \leq w_{p_i} \tag{5}$$

$$\sum_{i=1}^{m} f_{ij} \leq w_{q_j} \tag{6}$$

$$\sum_{i=1}^{m} \sum_{j=1}^{m} f_{ij} = \min \left(\sum_{i=1}^{m} w_{p_i}, \sum_{i=1}^{m} w_{q_j} \right) \tag{7}$$

Equation (4) means that only one direction can be moved from P to Q; Equation (5) and (6) indicate that the mobile unit must be smaller than its weight value, and the receiving unit cannot be greater than its weight value; Equation (7) indicates that the total number of mobile units cannot exceed the sum of the weights. After calculating the workload F, normalize it as in (8):

$$EMD(P,Q) = \frac{W(P,Q)}{\sum_{i=1}^{m} \sum_{j=1}^{m} f_{ij}} \tag{8}$$

IV. Case Study

TABLE I. Music Data Set

Style	Number	Size
happy	174	601MB
quiet	166	616MB
romantic	177	656MB
Sad	186	755MB
encourage	160	628MB
excited	181	714MB
total	1044	3.87GB

We collected six types of music as shown in Table I, which constitutes the recommended basic database, intercepting 30 seconds of each song as the processing object. The pre-

emphasis coefficient is 0.97, the frame length is 30ms, the adjacent frame interval is 10ms, and the number of Meyer filter Group is 20.

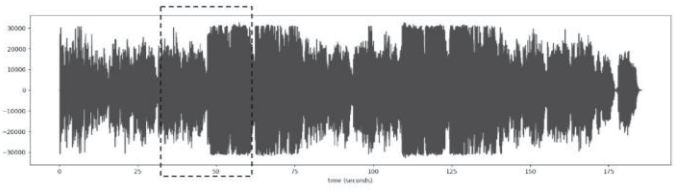

Fig. 3. Waveform of song A

As in Figure 3 below is a waveform of a song A. It can be seen that some of the waveforms are similar. When the interception is 30-60s, the segment in the dotted line in the figure can cover the repeated part. Therefore, the 30 to 60 second fragment of the song is taken as the research object.

A song C is randomly taken from the music database, and the recommended results obtained by the recommendation algorithm described in the text are as shown in Table II below. If the song of the recommendation result belongs to the same category as the song C, it is proved that the recommendation result is valid, that is, the recommendation result that the song C is similar in content is obtained. In this experiment, the accuracy of the results was 80%.

TABLE II. ONE RECOMMENDATION RESULT

recommended results	Distance	Belongs to Same Category
1	20.00	Y
2	70.67	Y
3	74.23	Y
4	79.94	Y
5	82.76	Y
6	83.97	N
7	84.37	Y
8	87.59	N
9	88.16	Y
10	89.49	Y

In addition, we took a total of 10 pieces of music in each type of music for a small range of experiments. The results are shown in Fig.4 below. The average accuracy of 12 recommendations is 89%. In this paper, the music recommendation method based on collaborative filtering is selected for comparison. According to Word2Vec, the music item is converted into a dense vector, and the similarity between items (Euclidean distance), is calculated, and the recommended result is obtained. The method compares the accuracy of the single song by the recommendation, and then compares it with the similarity-based algorithm proposed in this paper. The experimental results are shown in Fig.5. The average accuracy of multiple recommendations is only 57%. The experimental results demonstrate the effectiveness of the MFCC feature-like recommendation algorithm mentioned in this paper.

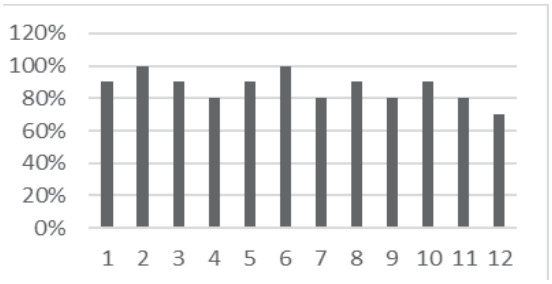

Fig. 4. Result of content-based recommendations

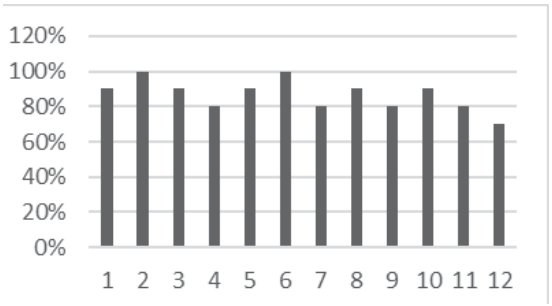

Fig. 5. Result of collaborative filtering recommendation

V. CONCLUSION

In this paper, a music recommendation algorithm based on MFCC eigenvalue similarity is proposed. The experimental results prove the validity of the inter-music distance algorithm. The average accuracy of multiple recommendation results reaches 87%. In addition, because of the limited amount of data, it is unrealistic to cover all the songs. Therefore, when the same song belongs to multiple categories, and the collected music is only in one category, it will affect the accuracy of the recommendation results. In addition, when the amount of data becomes larger, how to improve the recommendation speed is also a remaining work.

ACKNOWLEDGMENT

The authors would like to thank Guangdong Province collaborative innovation and platform Environmental Science build of special funds (2014B090908004); Dongguan City professional town innovation service platform construction project “Dongguan City Humen garment Collaborative Innovation Center”, whose constructive comments and suggestions helps us to improve the quality of this paper.

REFERENCES

[1] Y. Song, S. Dixon, and M. Pearce, "A survey of music recommendation systems and future perspectives," The International Symposium on Computer Music Modeling and Retrieval. pp. 395-410, 2012.

[2] U. Shardanand, and P. Maes, "Social information filtering: algorithms for automating “word of mouth”," Sigchi Conference on Human Factors in Computing Systems. ACM Press. Addison-Wesley Publishing Co. pp. 210-217, 1995.

[3] J. Foote, "An overview of audio information retrieval," Multimedia Systems. vol. 1, pp. 2-10, 1999.

[4] A. Wang, "An industrial strength audio search algorithm," Interior Symposium on Music Information Retrieval. pp. 7-13, 2003.

[5] P. Hoffmann, A. Kaczmarek, P. Spaleniak, and B. Kostek, "Music recommendation system," Journal of Telecommunications and Information Technology, pp. 59-69, 2014.

[6] B. Li, Q. Tao, and X. Li, "Music feature extraction based on fractal dimension theory for music recommendation system," 5th International Conference on Measurement, Instrumentation and Automation, pp. 538-542, 2016.

[7] X. Luo, X. Liu, R. Tao, and Y. Shi, "Content-based retrieval of music using mel frequency cepstral coefficient," Computer Modelling and New Technology. vol. 20, No. 1, pp. 1356-1361, 2016.

[8] K. Tada, R. Yamanishi, and S. Kato, "Interactive music recommendation system for adapting personal affection: IMRAPA," International Conference on Entertainment Computing, pp. 417-420, 2012.

[9] M. Lu, H. Zhang, and Q. Shen, "Realization of audio fingerprint based on power spectrum feature," Electronic Measurement Technology, vol. 39, No. 9, pp. 69-72, 2016. (*in Chinese*)

[10] M. Soleymani, A. Aljanaki, F. Wiering, and R. C. Veltkamp, "Content-based music recommendation using underlying music preference structure," IEEE International Conference on Multimedia and Expo, pp. 1-6, 2015.

[11] R. Thiruvengatanadhan, "Music Classification using MFCC and SVM," International Research Journal of Engineering and Technology, vol. 05, pp. 922-924, September 2018.

[12] X. Wang, and Y. Wang, "Improving Content-based and Hybrid Music Recommendation using Deep Learning," The 22nd ACM International Conference on Multimedia, pp. 627-636, November 2014.

[13] C. Ju, and S. Wang, "User-mood incorporated hubrid music-recommendation method," Journal of the China Society for Scientific and Technical Information, vol. 36, No. 6, pp. 578-589, 2017. (*in Chinese*)

[14] J. Zhu, and K. Deng, "An approach to audio fingerprinting extraction based on improved wavelet packet," Electronic Science and Technology, vol. 29, No. 3, pp. 30-34, 2016. (*in Chinese*)

[15] E. Pampalk, "Computational models of music similarity and their application in music information retrieval," Phd Thesis Technischen Universitat Wien, pp.1-140, 2006.