Music Recommendation System Design Based on Gaussian Mixture Model

Yang Lu[1,a], Xuemei Bai[2,b,*], Feng Wang[3,c]

[1]Institute of Electric Information Engineering, Changchun University of Science and Technology, Changchun, 130022, China

[2]Institute of Electric Information Engineering, Changchun University of Science and Technology, Changchun, 130022, China

[3]Institute of Electric Information Engineering, Changchun University of Science and Technology, Changchun, 130022, China

[a]email: 15143019233@163.com, [b]email: xmbai@cust.edu.cn, [c]email: 770736667@qq.com

*: Communication author

Keywords: Gaussian Mixture Model; Music Classification; Music Recommendation

Abstract. The paper establishes a double-layer classifier based on Gaussian Mixture Model and the Thayer model to divide the music style into several categories. On the basis of effective verification of experiments, the music listening experience is added into the model to analyze and normalize two-dimensional data points and the tastes of music users or playing times also can be added to obtain new three-dimensional data points. Then the Gaussian Mixture Model is employed again for classifying the new three-dimensional data points. In this way, not only can the taste changing process of users for different music be analyzed, but also the similarity among different users can be calculated. Therefore, music can be recommended properly to music users.

Introduction

With the arrival of a series of emerging media, computer networks and music penetrate into every corner of our daily life gradually and the number of music users is growing. In the huge amounts of music information environment, we need to identify and manage the music signal as well as analyze the tastes of music users. Only in this way, can we make a progress in music recommendation work.

Previous music recommendation systems just analyzed the contents contained within music itself and the effect was too fuzzy. That is because the degree of emotion embodied in different music is not same even they are in the same style. These systems only analyzed the mood of the music and they did not tell the feelings contained in music to what degree and ignored the different music tastes of music users. Therefore, the method is lack of pertinence. The classification of music styles with normalizing results proposed in this paper contributes to a specific plane coordinates of two-dimensional data points and it can reflect the level of emotion in music. Then it will format three-dimensional data points combining with users' preference taste for music in the past. The music can be recommended to users to integrate these new data points again with the Gaussian mixture model.

The System Model and Algorithm

A. The Thayer Emotion Model

It is difficult to describe accurately which kind of style a piece of music exactly belongs to because of the rhythm variability and the music signal frequency complexity. The paper divides the music signals into four categories combining with the results of previous studies. Four classification model diagrams used in the article are shown in Figure 1, which is according to the converted Thayer model [1]. In this diagram, the vertical axis represents music emotional energy and strong indicator of the degree (shown as 'Arousal'), while the horizontal axis represents the positive or negative indicator of the music mood (shown as 'Valence'). Thus, the system coordinated with the four classifications is constructed. As is shown in Figure 1, it makes the music signals divide into four typical styles of Exciting, Peaceful, Sad and Annoying [2]. In this classification model, Gaussian mixture model is used for music classification.

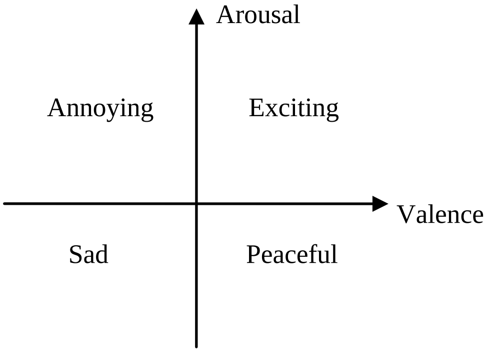

Fig.1. Four classifications model diagram for the music mood

B. Gaussian Mixture Model

Gaussian mixture model (GMM) is often used to describe mixture density distribution, in other words, it is a multiple Gaussian mixture distribution model [3]. It can easily and effectively express distribution characteristics of the same kind of data in feature space when dealing with classification problems, as the same kind of data in feature space are both in discreteness and concentration [4]. The Gaussian mixture model can simulate density distribution of arbitrary shape, no matter it is smooth or close [5]. That is why it has often been used in voice, image recognition and other aspects in recent years and got a great achievement.

C. EM Algorithm

Assume X {X_t, *t=1, 2, … K*} as a given feature vector sequence for training, while the likelihood of GMM can be express as

$$P(X|\lambda)=\prod_{t=1}^{T} p(X_t|\lambda) \tag{1}$$

The purpose of training is to find a group of λ to make maximum of P($X|\lambda$). EM algorithm is employed as a solution method because the calculation for the Gaussian mixture model is large and usually λ is a nonlinear function.

D. Feature Extraction of the Music Signal

There are some researches indicating that Mel Frequency Cepstrum Coefficient (MFCC) can be used as better classifications for audio features and it can improve the accuracy of the classification of audio [6]. According to the previous research experience, the paper will set the order of MFCC as 13. The analysis of short-term energy of the music provides an appropriate description method for rules of amplitude variation [7]. Short time average zero-crossing rate method refers to calculation of the average number of time that each frame signal goes through zero values, and its essence is the number of changing times of music signal sampling point. Short-term autocorrelation method is that we cut out some sample points of the music signal to get a signal by a short time window. Spectral centroid is the weighted average of the each component amplitude of the spectrum, which is equivalent to the spectral distribution of the ‘centroid’. Spectral flux is defined as the second-order distance of spectrum amplitude difference of the adjacent frames and reflects the size of spectral changes between adjacent frames [8]. Sub-band characteristics divided each frame short time spectrum into seven sub-bands in accordance with the rules of the octave and collected the maximum frequency component, the minimum frequency component and average value of each sub band into signature sequence. Thus, each frame has twenty-one dimensional characteristic values. It can effectively express spectral energy distribution of each sub-band.

The Experimental Process

A. The Pretreatment Process

Before experiment, it is necessary to preprocess the music. In order to calculate conveniently in this paper, every piece of music only extracts the music clips lasting 30 seconds that represents the most characteristics. It needs to do pre-emphasis, normalization and window frame after extracting fragments. The coefficient of the pre-emphasis of the paper is -0.9375 and it is added to the hamming window. The number of Gaussian mixture model class is 32. Too many classes will cause a burden of training and fitting and too few classes will not capture enough the music features on the contrary. Under the condition of the sampling frequency of 44100 by default, the frame length is 1024 and the frame is 512. There are 10 training music moods and 40 songs in total. Each emotion types have 5 test songs and 20 songs in total.

B. The Experimental Process

First, the models of $GMMV^+$ and $GMMV^-$ for the positive and negative shaft ends are established respectively, which represent the emotional intensity. The high intensity training samples are employed, which illustrate the Annoying and Exciting feelings to train $GMMV^+$ and the low intensity training samples of Sad and Peaceful mood are used to train $GMMV^-$. The two models represent the strong positive and the negative mood. In the same way, models of $GMMV^+$ and $GMMV^-$ on the both ends of the shaft of Valence are set, which express positive and negative emotions. In the process of the experiment, a two-layer classifier is formed. A test sample first determines the mood's intensity after the first layer classifier and it determines the positive and negative emotions after the second classifier. Usually when it comes to the question of multiple classifiers, it will cause the value allocation problem. For the distribution of the weights in the paper, seven tests are implemented to observe the effect of classifications under different weights and comparisons.

C. The Experimental Results

Several experiments were implemented for seven different weights. The weight distribution of arousal and the weight distribution of valence are shown in Table 1 and 2 as 'the weight of A/V'. 20 pieces of music are employed to test the accuracy for the weights of arousal and valence with 0.2/0.8, 0.3/0.7, 0.4/0.6, 0.5/0.5, 0.6/0.4, 0.7/0.3 and 0.8/0.2. It can be seen from Table.1 that the accuracy is the same when the weight distribution of arousal are 0.6 and 0.7, and the weight distribution of valence are 0.4 and 0.3 respectively.

Table.1. The influence of different weights

The weight of A/V	0.2/0.8	0.3/0.7	0.4/0.6	0.5/0.5	0.6/0.4	0.7/0.3	0.8/0.2
correct numbers	13	13	12	14	17	17	16
The error Numbers	7	7	8	6	3	3	4
accuracy	65%	65%	60%	70%	85%	85%	80%

10 more test samples are added to distinguish them further. Therefore, there are 30 samples for testing. The results are shown in Table.2. The results show that the weight distribution of arousal and valence is 0.7/0.3 achieved the best results in the 30 test samples.

Table.2. The comparison of two kinds of weights after increasing test samples

The weight of A/V	0.6/0.4	0.7/0.3
The correct numbers	25	26
The error numbers	5	4
Accuracy	83.3%	86%

Take the song of 'Best friend' as an example, a plane radar map is shown in Figure2.

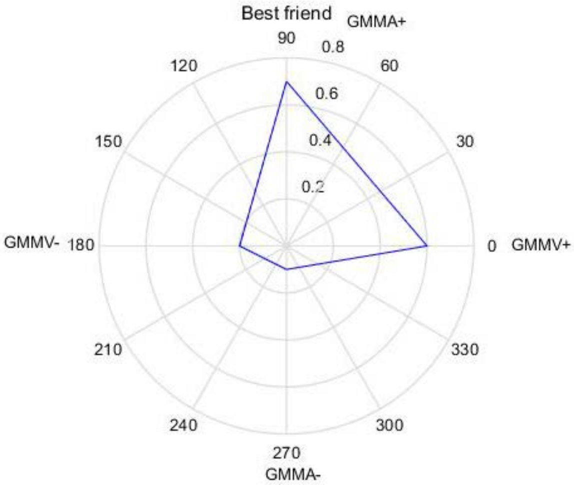

Fig.2. Radar map of ‘Best friend’

The value on GMMV$^-$ and GMMA$^+$ shafts is big for this piece of music. As we can see from Figure 2 obviously, it accounts for a larger area in the first quadrant, which represents the exciting mood and it meets the audience's perceptions of the song of ‘Best friend’.

The Similarity Detection for the User Model

A. The Data Normalization

Take GMMV$^+$ as an example to illustrate the process of data normalization to extract the feature parameters for the user played music. Set the characteristic vector for the V shaft as C_{xv}, for the GMMV$^+$ model, the posteriori probability provided the parameter of θ_v^+ is log $p(C_{xv} \mid \theta_v^+)$.

When the model parameter is θ_v^+, the maximum posteriori probability should be the value in θ_v^+ for the training feature vector C_v^+ of GMMV$^+$ and it is shown as follows.

$$\max \log p(C_v / \theta_v^+) = \log p(C_v^+ / \theta_v^+) \tag{2}$$

The minimum value is:

$$\min \log p(C_v / \theta_v^+) = \log p(C_v^- / \theta_v^+) \tag{2}$$

The data point of C, which is normalized with data points in axis of V^+ is shown as follows.

$$C(v^+) = \frac{\log p(C_{xv} / \theta_v^+) - \dfrac{\log p(C_v^- / \theta_v^+)}{K}}{\dfrac{\log p(C_v^+ / \theta_v^+)}{K} - \dfrac{\log p(C_v^- / \theta_v^+)}{K}} = \frac{K \log p(C_{xv} / \theta_v^+) - \log p(C_v^- / \theta_v^+)}{\log p(C_v^+ / \theta_v^+) - \log p(C_v^- / \theta_v^+)} \tag{3}$$

where, K presents the number of training model samples. In the same way, the normalized data points of C in the axis of V^- are:

$$C(v^-) = \frac{K \log p(C_{xv} / \theta_v^-) - \log p(C_v^+ / \theta_v^-)}{\log p(C_v^- / \theta_v^-) - \log p(C_v^+ / \theta_v^-)} \tag{4}$$

Therefore, the normalized data points of C in V-axis are:

$$C(v) = \begin{cases} C(v^+), if \ C(v^+) > C(v^-) \\ C(v^-), if \ C(v^-) > C(v^+) \end{cases} \tag{5}$$

Similarly, the normalized data point of C(a) of C can be obtained from A-axis. There are two-dimensional data points of C(v,a). A new music analysis data point of C(v,a,q) can be obtained when the users’ preference setting or the times of playing is added to the V-A plane, which can be expressed as Q.

B. The Recommended Process for User

a) This example shows the clustering result of a user's 30 pieces of music into 4 classes. It can be seen from Figure 3 that the playing times are few for the lower arousal and the playing times are more for the higher arousal and higher valence. The preference of a user for the different types of music can be analyzed from the clustering results. The clustering results and the time-stamping can find the preference variation process of the different types of music for a user and then music can be recommended to the user unpredictability.

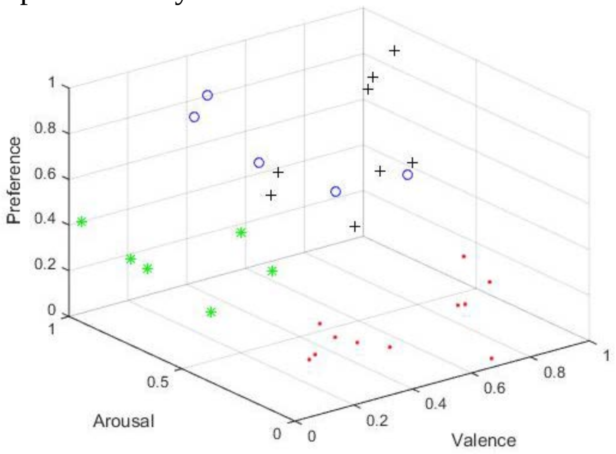

Fig.3. The clustering result of clustering a user's of 30 pieces of music into four classifications

b) The similarities among different users are calculated. Using each user's data to analyze $C(v,a,q)$ and train a GMM model. The following example is about two users.

Similarities between user A and user B is:

$$r(A,B) = \log p(C_A / \theta_A) + \log p(C_B / \theta_B) - \log p(C_A / \theta_B) - \log p(C_B / \theta_A) \tag{6}$$

Its maximum value is:

$$r_{\max}(A,B) = \log p(C_A / \theta_A) + \log p(C_B / \theta_B) \tag{7}$$

Therefore, the similarity between model A and model B is:

$$SIM(A,B) = \frac{r_{\max}(A,B) - r(A,B)}{r_{\max}(A,B)} = \frac{\log p(C_A / \theta_B) + \log p(C_B / \theta_A)}{\log p(C_A / \theta_A) + \log p(C_B / \theta_B)} \tag{8}$$

where, *SIM* (*A*, *B*) presents the similarities between user A and user B. With the information, we can calculate the similarity to judge the same preference points for difficult users, such as $GMM_{p1}, GMM_{p2}, \ldots GMM_{pn}$, and then popularize to the new music.

Conclusion

Based on previous music sentiment analysis, the paper normalized posterior probability of testing music with the known model parameters and added the users' preferences to the three-dimensional data. It can be seen from the simulation results that the radar map can show the main emotion contained in a song, which shows that the double classifier is effective. It can analyze the user's preference about different mood music using Gaussian mixture model in three-dimensional data, which can be targeted to achieve music recommendation.

Acknowledgement

In this paper, the research was sponsored by the Nature Science Foundation of Jilin Province (Project No.20150101013JC) and the Innovation Team Project of Jilin Provincial Science & Technology Department (Project No. 20130521015JH).

References

[1] K.C. Xu. Research and Implementation of Music Emotion Parameterized System. Guangzhou: South China university of Technology, 2013.

[2] X. Gong. Optimization and Implementation of Gaussian Mixture Model Parameters [J]. Journal of Southwest Jiaotong University, 2010.1, 2-3.

[3] M. Werner, C. Delucchi, H.J. Vogel, et al. The ATM-based Routing in LEO/MEO Satellite Networks with Intel Satellite Links [J]. IEEE Journal on Selected Areas in Communications, 1997(15), 69-82.

[4] Reynolds D A. Speaker Identification and Verification with Gaussian Mixture Speaker Models [J]. Speech communication, 1995, 17(1), 91-108.

[5] Y.G. Cheng, B.Y. Di. Wireless Local Area Network (LAN) Localization Algorithm Based on Gaussian Mixture Model [J]. Computer Engineering. 2009, 35(4), 25-27.

[6] D. Shi. Research on the Music Style Similarity Detection Algorithm. Dalian: Dalian university of Technology, 2013.

[7] Mesaros A, Astola J. The Mel-Frequency Cepstral Coefficients in the Context of Singer Identification[C]. ISMIR, 2005, 610-613.

[8] L. Wang, L.M. Du, J.L Wang. Music Emotion Classification Based on Adaboost. Journal of Electronics and Information, 2007, 29(9), 2067-2072.