Naive Bayes Classifiers for Music Emotion Classification Based on Lyrics

Yunjing An[1,2,3], Shutao Sun[1], Shujuan Wang[1]
[1]Information Engineering School, Communication University of China
[2]Key Laboratory of Acoustic Visual Technology and Intelligent Control System, Ministry of Culture
[3]Beijing Key Laboratory of Modern Entertainment Technology
Beijing, 100024, China
an_yunjing@163.com

***Abstract*—There is a constantly growing interest in evaluating music information retrieval (MIR) systems that can provide effective management of the music resources. The crucial characteristic of music is its emotion, which reflect the human's perception. To do the automatic classification of Chinese music emotions more effective, we use the lyrics of music to analysis and classify music based on emotion. There are many algorithms to achieve text classification, and one of the most popular algorithms is Naive Bayes algorithm. Although it is simple, it can classify text effectively. In this paper, we crawl the music lyrics and their labels from a popular website named Baidu music and make our four different datasets. We also train four classifiers with different datasets and report their performance. We evaluate the classifiers trained by four different datasets, and the final accuracy we get is approximately 68%.**

Keywords—Naive Bayes algorithm; music emotion; lyrics.

I. INTRODUCTION

The essence of music emotion classification based on lyrics is text classification. There are many algorithms for text classification, such as SVM [1], Bayes, KNN, decision tree algorithm. There haven't form a unified test data and the evaluation method in the area of Chinese text categorization and the objectivity of the performance evaluation is difficult to achieve [2-3]. And there are some researchers have done lots of work in this field [4].

To get a better classifier, it's important to use an appropriate algorithm and get the correct dataset. Through investigate and survey, the algorithm we decide to use is Naive Bayesian Classification algorithm. Naive Bayes has very high learning efficiency and it can estimate all the probability just need a scan of the training data [5]. Naive Bayesian Classifier is a simple classifier based on applying Bayes theorem with independence assumptions.

The Naive Bayes classifier, which uses the Naive Bayesian formula to calculate the probability of each class A given the values B_i of all attributes for an instance to be classified, the conditional independence of the attributes given the class:

$$P(A \mid B_1..B_n) = P(A)\prod_i \frac{P(A \mid B_i)}{P(A)} \quad (1)$$

Naive Bayes classification is based on the assumption that the principle of maximum posteriori hypothesis to identify the object that most likely to be classified under the category. Bayes theorem shows the relation between one conditional probability and its inverse.

In this paper, we assume that the words of the lyrics are independent and of equal weight. Finally, using the above formula, we can get the conditional probability of each category, of which the biggest one is the category of the sample. The Naive Bayesian Classifier is a classification algorithm that is easy to implement. Owe to its simplicity of the model and the good classification performance, it's widely used in engineering application.

To carry out our work we need a great quantity of lyrics with the emotion label. Due to the complexity of the music emotion and the unpredictability of the human perception, the difficulty of the topic is obtaining a large of Manually-Tagged dataset. In this paper, our dataset which will be described in section II obtained from the internet. In section Ⅲ, four different datasets are used in our experiments, and we will get four different accuracy of four datasets. Section Ⅳ consists of some conclusion and our prospect of the music emotion classification.

II. DATASET

Now we are ready to describe the dataset we used. After investigation and study, considering the reliability and the difficulty of collecting, in this paper the music emotion labels we used are from Baidu music emotion labels. We get it by crawling the web of Baidu music. The emotion labels of the Baidu music contain sad, passionate, quiet, comfortable, sweet, inspirational, lonely, miss, romantic, yearning, joyful, soulful, happy, nostalgic, relaxed.

A. Lyrics Collection

We use the Scrapy-a framework of Python to achieve a simple crawler to crawl the related information of Baidu music which contain the singer, music name, lyrics and the category of the music. The music information we crawl from the page will be written to the database of MySQL. We get the labels and lyrics from database when we need.

978-1-5090-5507-4/17/$31.00 ©2017 IEEE
ICIS 2017, May 24-26, 2017, Wuhan, China

B. Emotion Model

It's necessary to choose the emotion model before we explore the music emotion. There are two emotion models which are defined by dimension, and they respectively are Russell emotion model and Thayer emotion model.

Russell model is established by two dimensions named arousal and valence [6], as is shown in Fig 1.

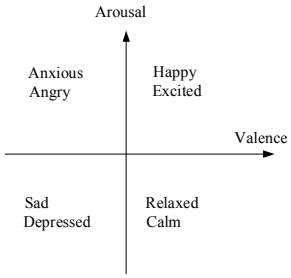

Fig 1. Emotion model of Russell

As Fig1 shows, the abscissa represents valence, and the ordinate stands for arousal. The four quadrants represent the four kinds of emotions. The first quadrant represents happy emotion and excited emotion, and the second quadrant represents anxious emotion and angry emotion, and the third quadrant represents sad emotion and depressed emotion ,and the fourth quadrant represents relaxed emotion and calm emotion.

Thayer model is based on two dimensions named energy and stress [7], as it's shown in Fig 2.

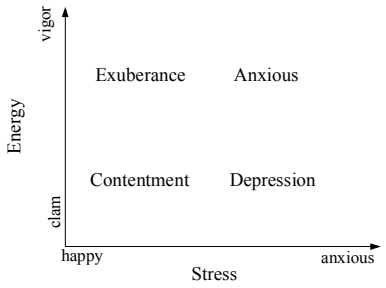

Fig 2. Emotion model of Thayer

As is shown above, energy-axis reflects the people's vitality in physiological, from clam to vigor. Stress-axis reflects the process of people from happy to anxious in psychology. It divides the music emotion into four categories: exuberance, anxious, contentment and depression.

On the one hand, this emotion classification model has some disadvantages. It compresses the complex emotion of human to a dimensional model, and thus it will lead to the loss of music information to some degree. On the other hand, there still exists some advantages about it. Firstly, this model is more suitable for emotion recognition because of its emotion characteristics. Secondly, the model uses the idea of dimension to describe the emotion of music, and it's easy to establish the connection with the music features. In addition, the two factors of energy and pressure in Thayer emotion model can be well matched with the audio features. Considering the emphasis of this paper, we decide to use this emotion classification model.

C. Maping of the Emotion Labels

Because the transform of Baidu music labels to the category of anxiety is inappropriate, so we transform the emotion labels of Baidu to three categories, namely contentment, depression and exuberance. We find the labels are complicated and various. It is difficult to classify the origin labels because they are so polysemous. We remove the categories of the ambiguous labels, such as miss, yearning, which we can transform to different emotion depend on the human emotion. If some music with mutually exclusive labels, they will not be used in experiment. We subjectively transform the other emotion labels to these three categories. As shown in TABLE I.

TABLE I. MAP OF THE EMOTION LABELS

Emotion labels of our work	Emotion labels of Baidu
Contentment	Quiet, Yearning, Romantic, Sweet, Healing
Depression	Waiting, Sad, Frustrated
Exuberance	Passionate, Inspirational, Happy, Joyful

As is shown above, we get 11 different labels of Baidu music to do experiments. According to the method above we complete the label transformation, finally we get 3552 songs that both have labels and lyrics. We get 346, 2175 and 1031 music with emotion of contentment, depression and exuberance respectively.

III. EXPERIMENTS

With the dataset defined in section II, in this section we will introduce the experiments. We will evaluate the performance of model by the accuracy which calculated by

$$\text{Accuracy}=\frac{\text{Musics have been correctly classified}}{\text{all music have been classified}}\text{x}100\% \quad (2)$$

From the dataset, we randomly pick up 2/3 of them for training and 1/3 for testing. Each lyric of song used for experiment will be segmented and picked up emotional words. For the word segmentation, we use a module of Python named Jieba, which have achieved good results in Chinese word segmentation. After segmentation and picking up emotion words from emotion dictionary, we get the dataset for our experiment. Considering the nature of the Bayes algorithm, we get four different datasets to do experiments and compare their results.

Finally, we get 2369 songs for training. In the training set, 231 songs of them belong to the category of contentment, and 1450 songs belong to the category of depression, and 688 songs belong to the category of exuberance. There are 1183 songs used for testing. In the testing set, 115 songs of them

belong to category of contentment, and the number of them belong to category of depression and category of exuberance are 725 and 343. We name this dataset as D-1. The experimental result of D-1 is shown in TABLE II.

TABLE II. EXPERIMENTAL RESULT OF D-1

Original Label / Test Result	Depression	Contentment	Exuberance
Depression	574	81	193
Contentment	66	14	28
Exuberance	85	20	122

For D-1, there are both Chinese lyrics and English lyrics, and most of them are Chinese. It's easy to find that the English lyrics are useless data for the experiment though the number of them is little. We expect that the dataset only contain Chinese lyrics can improve the classification effectively. So we design a different dataset which only contain Chinese lyrics. Finally, we get 3316 songs for experiment, and 271 songs of them belong to category of contentment, and 2091 songs belong to category of depression, and 954 songs belong to category of exuberance. We divide the dataset into training and testing by the same way, and we name this dataset as D-2. The experimental procedure is the same as before, and the experimental result of D-2 is shown in TABLE III.

TABLE III. EXPERIMENTAL RESULT OF D-2

Original Label / Test Result	Depression	Contentment	Exuberance
Depression	584	57	160
Contentment	64	13	34
Exuberance	85	20	124

In this paper, we also investigate the accuracy of emotion classification only under valance dimension, in which the music are classified to two classes, positive and negative. So we transform the Baidu music labels under the rules as follows:

Positive: Quiet, Yearning, Romantic, Sweet, Healing, Passionate, Inspirational, Happy, Joyful.

Negative: Waiting, Sad, Frustrated.

After the transformation, we get 1377 songs belong to positive category, and 2175 songs belong to negative category. As usual, 2/3 of them are used as training dataset, and 1/3 of them are used as testing dataset. After dividing, there are 918 songs of positive category and 1450 songs of negative category in the training dataset. There are 459 songs of positive category and 725 songs of negative category in testing dataset. We name this dataset as D-3. Using the same way to divide the music, and the experimental procedure is the same as before. The result of D-3 is shown in TABLE IV.

TABLE IV. EXPERIMENTAL RESULT OF D-3

Original Label / Test Result	Positive	Negative
Positive	78	21
Negative	381	704

The D-3 is same to D-1 which have lyrics both of Chinese and English. Then we define another dataset named D-4 which is same to D-3 that only contain Chinese lyrics. We expect the result of D-4 will higher than D-3. Finally we get 3316 songs for experiment, and 1225 songs of them belong to positive category, and 2091 songs belong to negative category. The training and testing datasets are divided according to the proportion of 2:1. Using the same way to do experiment. The result of D-4 is shown in TABLE V.

TABLE V. EXPERIMENTAL RESULT OF D-4

Original Label / Test Result	Positive	Negative
Positive	180	122
Negative	228	575

To verify the performance based on different datasets, we train datasets D-1, D-2, D-3 and D-4, then make a comparative experiment and give the four results are shown in TABLE VI.

TABLE VI. ACCURACY OF THE DIFFERENT DATASET

Dataset	Accuracy	Accuracy (%)
D-1	(574+14+122)/1183.0	60
D-2	(584+13+124)/1105.0	65
D-3	(78+704)/1184.0	66
D-4	(180+575)/1105.0	68

As is shown above, the accuracy is calculated by the formula (2), and the result of D-4 is the highest of our experiments. Comparing the experimental result of D-1 with D-2, we find that the English lyrics are noise data, which influent the accuracy seriously. Comparing the experimental result of D-2 with D-3, we find that the result of two categories is better than three categories, which depend on the formula of Bayes and it also tell us that the number of datasets influence the accuracy.

IV. CONCLUSION

This paper describe the principle of Bayesian algorithm, and the reason why we use the Bayesian algorithm. We focus on designing an effective classifier to classify music by music emotion based on lyrics. First, we collect the dataset by crawling the web of Baidu music. Then, we choose one

emotion model from the two usual models. In order to validate the influence of different dataset on the classification performances, we use four different datasets for training and compare the results. Because of the polysemous of the Chinese words, we cannot sum up it just use the thousands of songs. We expect to classify the music using the audio of it based on emotion, and contrast the performance with this paper.

ACKNOWLEDGMENT

This paper is partly supported by “Science and Technology Innovation Project of Ministry of Culture (WHB201520)” and “The Project of National Science and Technology Support Plan (2015BAK22B01)”.

REFERENCES

[1] J. Rocchio, G. Salton, "Relevance Feedback in Information Retrieval", *The SMART System*, pp. 67-88.

[2] G. Salton, A. Wang, C. S. Yang, "A vector space model for automatic indexing [J]", *Communication of ACM*, vol. 18, no. 11, pp. 613-620, 1975.

[3] T. Zhou, The study based on WWW information technology [D], Nanjing:university, 1999.

[4] C. R. Luo, Automatic Chinese text classification based on the statistical method, 2004.

[5] G. Zheng, Y. Tian. "Chinese web text classification system model based on Naive Bayes." E-Product E-Service and E-Entertainment (ICEEE), 2010 International Conference on. IEEE, 2010.

[6] J. Russell. A circumplex model of affect[J]. Journal of Personality and Social Psychology, 1980,39(6):1161-1178.

[7] R. E. Thayer. The bionsvchology of mood and arousal[M]. America: Oxford University Press,1989.