2009 Fifth International Joint Conference on INC, IMS and IDC

Acquiring Mood Information from Songs in Large Music Database

Yi Liu
School of Information
Renmin University of China
Beijing, China
liuyi_lee@126.com

Yue Gao
School of Information
Renmin University of China
Beijing, China
gaoyue-yx@ruc.edu.cn

Abstract—**Automatic mood information acquiring from music data is an important topic of music retrieval area. In this paper, we try to find the strongest emotional expression of the song in large music databases. By analyzing hundreds of credible reviews from website, a 7 keywords mood model is constructed. 217 songs were collected in our dataset. Every song was divided into several 10s-long segments and our dataset containing 5929 music clips. We used Gaussian Mixture Model (GMM) and Support Vector Machine (SVM) as classifier and four feature selection algorithms to do mood classification experiments. A post-processing method was presented to find the strongest mood expression of each song. From the experiment result, we can see that SVM is the best classifier for mood classification, and Active Selection algorithm can remove weak features effectively. Using SVM classifier, the classification accuracy can achieves 83.33% with 40 features by using active selection algorithm, and 85.42% with 84 features which selected by ReliefF.**

Keywords-music mood classification; music mood taxonomy; music information retrieval; feature selection; mood detection.

I. INTRODUCTION

Nowadays, there are more and more music databases in our daily life. In music retrieval systems, mood is an important semantic attribute. It can express inherent emotional meaning of a song，and be used in semantic based music retrieval and recommendation. It is also very useful in music-related applications, such as medical treatment, background music of different scenes and personal appreciation. Therefore, automatic acquiring mood information from music data is an important topic of music information retrieval area.

Recent approaches about music mood automatic labeling following two stages: firstly, building model to describe mood taxonomy, and then obtaining mood information through classification. Now, some research using midi files, take keywords to represent their mood models. While, other research using audio files such as mp3, wav etc., take Thayer's model as their mood model. In this paper, we use acoustic music data and keywords model for acquiring mood labels in our study.

This paper is constructed as follows. Section 2 introduces related work. In section 3, an approach for building dataset is presented, and gives our mood model. Detailed mood information requiring method is presented in section 3, including feature extraction, feature selection, classification and post-processing. Section 4 shows our experiment and performance evaluation of the proposed algorithms and section 5 with conclusions and future directions

II. RELATED WORK

There has been a significant amount of research on mood classification. Mood taxonomy and classification methods are the main aspects.

A. *Mood taxonomy*

Establishing mood taxonomy is the first step in mood classification. As we known, there isn't a standard mood model. But we can summarize the mood models used before as two types: keywords model and acoustic feature model.

Hevner's adjective checklist is the most famous keywords model [1]. It is presented in 1930s and composed of 67 adjectives from eight clusters, which include Sober, Gloomy, Longing, Lyrical, Sprightly, Joyous, Restless and Robust. The other one is the model used in MIREX [3]. It has five classes and each of them using a group of similar words described. These models express rich affections which listeners can usually feel about. But it is difficult to connect the adjectives with acoustic features.

Thayer's model considers factors which affect emotion [2]. It proposed a two-dimensional mood model. By evaluate energy and stress, Music in this model can be divided into four clusters: Contentment, Depression, Exuberance and Anxious/Frantic. Because energy and stress can expressed by some acoustic features, many researches which use audio files take this model in their experiment. The Tellegen-Watson-Clark model [4] is derived from Thayer's model, it has 6 clusters.

B. *Classification Methodss*

Those work which used keywords models, usually took Midi files in their dataset. It's easy to get melody features such as pitch, duration etc. from Midi files, so those experiments show a good performance, the accuracy was about 96%. Some of these work also used audio files, but the performance was not very well. Tao Li [5] presented a mood model with 13 classes, also combined some of categories to form a model with 6 classes, each class depicted by some adjectives. The precision of classification was about 36% by using a dataset containing 499 music clips, each one was 30s long. Douglas Turnbull [6] [7] used 159 keywords to represent 159 classes, including 18 emotions. They created CAL500 data set. The accuracy can achieve 20% by using Gauss model and EM algorithm.

978-0-7695-3769-6/09 $26.00 © 2009 IEEE
DOI 10.1109/NCM.2009.311

Thayer's model is used more frequently. According to this model, some related acoustic features can be extracted and used for classification. . L. Lu, D. Liu [8] [9] proposed an approach to mood detection using timbre, intensity and rhythm features, and took GMM as basic classifier. The experiment was done on 800 clips, and the accuracy can achieve 86%. Y. Feng [10] used two features: tempo and articulation, and neural network to do the experiment. By using 223 clips in their dataset, the precision is about 67%. In Yi-Hsuan Yang's [11] work, they chose 25s-long segments for each song in their dataset, which expressed the strongest emotion of the song, and then extracted 15 kinds of acoustic features for classification. By using Fuzzy Nearest-Mean classifier and stepwise backward selection method to remove weak feature, the accuracy can achieve 78.33%.

From all these work it can be seen that those who used keywords mood model usually get a lower performance. Using Thayer's model the performance seems to be well, but it only has four classes and few clips in their datasets, so it can not widely applied in our daily life. Besides, there is a few feature selection works in previous researches. Only in Yi-Hsuan Yang's work [11], stepwise backward selection method is used and the average accuracy is better than using all features.

III. Mood Taxonomy and Dataset

In this paper, a 7 keywords mood model is presented by analyzing hundreds of credible classical music reviews, and most of them are about Chinese music. According to these reviews, songs are collected to form our dataset.

A. *Mood Taxonomy*

Hevner's adjective checklist is the most influential keywords model. But for many classical songs, especially Chinese songs, Hevner's adjectives can not express their emotion accurately. So we analyze thousands of reviews from authority music websites, and do a statistical count of mood adjectives' term frequency. 7 adjective groups with high frequency were chosen at last, including sturdy, enthusiastic, lively, melodious, euphemistic, depressed, and anguished, as table Ⅰ shows. They can express the emotion of most classical music.

B. *Dataset*

In most researches, songs or clips in datasets are annotated by experts or some other persons. That would expend much time and the dataset should be small enough. If there are a few people take part in annotating work, the result may be not very credible. Because personal feelings is subjective, and could be easily affected by the annotator's emotion at that time. To avoid these problems, we choose well-known classical music in our dataset, especially Chinese music. Because Chinese people are very familiar with them and there are many credible reviews about them written by experts. Most songs and related reviews are downloaded from guqu.net [12]. We put part of them into dataset corresponding to our mood model and their reviews.

Finally, we got 217 songs in our dataset. In order to get the strongest emotion expressed by each song. Firstly, each song is converted to a uniform format: 16000HZ, 16 bits, mono channel, and then divided into 10s-long clips. Here we didn't choose longer duration for music clips because the emotion is likely to change in a long duration[13].The dataset contained 5292 clips, about 700-800 clips in each class, and 75% of them used in training set, 25% used in testing set. The clips came from the same song couldn't in two sets at the same time.

TABLE I. TABLE TYPE STYLES

Keyword	Synonym
sturdy	athletic, firm, powerful, robust
enthusiastic	energetic, impassioned, passionate, vehement
lively	vivid, vivacious, alive, brisk
melodious	pleasant
euphemistic	Periphrastic, roundabout
depressed	blue, black, dusky, mopish
anguished	sorrowful, heartbroken, sore, tearful

IV. Music Mood Classification

In this paper, we extracted 88 acoustic features, and then used four feature selection algorithms to decrease the dimension. SVM classifier and Gauss model are used for classification. In order to find the songs' strongest emotion, a post-processing method is applied.

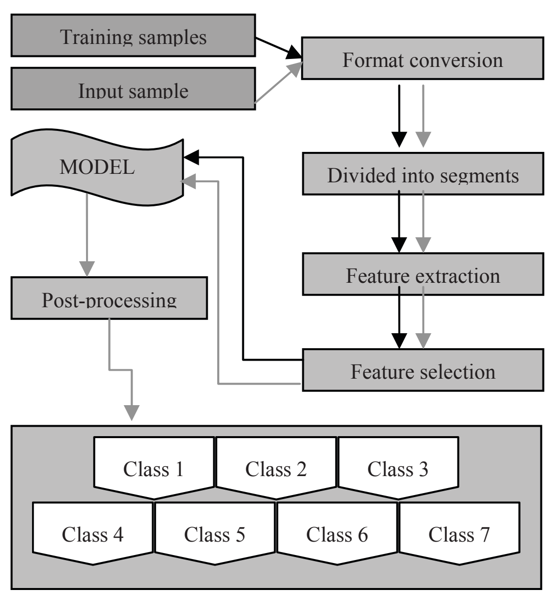

Figure 1. model generator and mood classifier

The process of classification can be described in two aspects: model generator and mood classifier. Using features extracted from training samples, a classification model can be generated. In figure 1, the dark arrow directs this process. And then, the model is applied to classify input samples, as the light arrow directs.

A. *Feature Extraction*

Each clip is divided into 0.5 overlapping 32ms-long frames. Jaudio [14] [15] was used for the feature extraction process. The extracted features fall into four categories: timbre, intensity, rhythm and others. The first three sets can express mood information to some degree and very important for mood detection. The last feature set also affects music emotion but it's hard to explain how it works.

1) Timbre features

Happy songs usually sound bright and clear, while grief ones sound dark and melancholy. Timbre features can be used to judge whether the emotion is negative or positive. The timbre features we used are listed as follows: Centroid, Rolloff Point, Flux, Zero Crossing, Strongest Frequency Via Zero Crossing, Strongest Frequency Via Spectral Centroid, Strongest Frequency Via FFT Maximum, Compactness, MFCC, LPC, Peak based Spectral Smoothness. We calculated the mean and standard deviation over all frames. This led to a total of 64 timbre features.

2) Intensity features

Intensity features can used to judge whether the emotion is very strong or not. For example, if songs express a positive emotion, then using intensity features we can get whether it is enthusiastic or lively. In this paper, the intensity features are RMS and Fraction of Low Energy Windows. By calculating the mean and standard, we got 4 intensity features.

3) Rhythm feature

Through rhythm features, we also can get some information about whether the music emotion is positive or negative. Fast songs tend to be happier than slow ones. We extracted rhythm features including Beat Sum, Strongest Beat, and Strength of Strongest Beat. Also by calculating the mean and standard, led to 6 rhythm features.

4) Other features

Other acoustic features, although it's difficult to say how they affect music mood, but they also play an active role. This can be seen from the result of feature selection. These features include Spectral Variability, Method of Moments, and Relative Difference Function. We got 14 features by calculating mean and standard.

B. *Feature Selection*

Feature selection is an important technology in pattern recognition. If the feature vector contained enough information of different classes, then input samples can be classified correctly. But it's difficult to learn whether the information is enough or not and which feature can play an active role. If features extracted as much as possible, the dimension would increase sharply, and many features maybe very weak or related to each other. So choosing suitable features to describe a certain classification model is very important, it can not only save time spend on model generator and spaces used for store samples, but also improve classification performance.

In this paper, we used four feature selection algorithms, Stepwise Forward Selection method (SFS), ReliefF, Fisher Rule and Active Selection. Among them, ReliefF and Fisher Rule are Filter feature selection methods. SFS is a type of wrapper approach. We proposed Active Selection method based on both filter and wrapper approaches. The testing samples used in feature selection are chosen from training samples randomly. We made a comparison on these algorithms in our experiment.

1) SFS method

SFS is a simple approach for feature selection. Firstly, each feature is used in classification, and the best one would be put into final feature set. And then greedily adds the best feature sequentially until no more accuracy improvement can be obtained. Generally, this method can decrease the dimension effectively. Because it's based on classifier, so for a certain test set, the accuracy could be higher than using all features. But if the test set change, the result may become worse. So it's not stable enough.

2) ReliefF method

K. Kira [16] present Relief algorithm in 1992 and it is acknowledged as a good algorithm on feature evaluation. And then in 1994, Kononenko I. [17] improved this method and presented ReliefF algorithm. ReliefF algorithm is able to detect conditional dependencies between attributes. It is very efficient for feature evaluation and can quickly find some weak features in the feature set. But for those which have negative effect to some classes but play an active role in others, would be hardly found by ReliefF. So this algorithm has limitations in feature selection. The greatest advantage of ReliefF is that it can remove apparent weak features effectively.

3) Fisher rule

Fisher rule computed inner-class and inter-class distance. Usually strong features' inner-class distance is much smaller than inter-class distance. Then according to the difference of two distances, we can arrange those features, and choose strong ones for classification. In some classification experiments, fisher rule performs well in feature selection. It can remove some weak features and slightly increase the accuracy. But our purpose is to find the strongest emotion expressed by a song. So effective for clips' classification doesn't mean effective for songs'.

4) Active selection

Active selection is very useful in the classification of Chinese folk songs [18]. It pays a great attention on error samples in classification, and tries to add features that help correct them. In order to distinguish from the training set and testing set used in classification. Here we use Strain and Stest to represent training samples and testing samples used in feature selection. They all come from the training set. F(set1, set2) means extracting features in set1 by using samples in set2 to compose a feature set. The algorithm is presented as follows:

a) Input F(all, Strain) and F(all, Stest).

b) The F(all, Strain) is used as input to relief, the output feature set is remarked as Init.

c) The first N features in Init is used to compose F(Init, Strain) and F(Init, Stest).

d) Input F(Init, Strain), F(Init, Stest) into SVM classifier, and get an accuracy and error samples (marked as

Serror). If this is first time to d) or the accuracy is higher than last one, then go to e), otherwise return to f).

e) Input F(all, Serror) into reliefF, the best feature in output which not in Init would be added into Init.

f) If d) returned, remove the feature in the output that just added in Init and return to e). Until return M times. otherwise Go to d).

The difficulties in using active selection method are deciding the initiate number of features (N) and the return times (M). If the N is very small, then the algorithm may be stop at local optima. Otherwise, some weak features can not be removed. In our experiment, we assign 30 to N, because on our own experience, the dimension of best feature set is from 30 to 80, so the features would neither too few to express mood information or too much to express much redundant information. If M is not assigned, the algorithm would stop when there is no feature can be added. This may take time, but can ensure there is no strong feature dropped away.

C. Classifier

In previous work, many researches used Gaussian Mixture Model (GMM) to model each feature set regarding each mood cluster, and Expectation Maximization (EM) algorithm to estimate the parameters of the Gaussian component and mixture weight [8][9].

Support Vector Machine (SVM) is a machine learning method. It delivers state-of-the-art performance in real-world applications. LibSVM is a tool made by Chih-Jen Lin [19] [20]. It is an integrate software that support vector classification and multi-class classification. In MIREX audio mood classification task [3], using LibSVM as classifier performed very well. In this paper, we'll use two classifiers and made a comparison in our experiment.

D. Post-processing

Our purpose is to find the strongest emotional expression of a song, not a clip. So after classification, according to predict classes of the clips in each song, we calculate the probability of each song in each class. That is , for each song S, S={c_1,c_2,...,c_n}, c_i (i=1,2,...,n) is a clip which comes from S, and n is the total number of clips from S. our goal it to express S as S={C_1,C_2,...,C_m}, C_i means i-class, and m is the class number. In our model, m is 7. After classification, c_i would be classified into C_k, then the set {c_1,c_2,...,c_n} can be grouped into {C_1,C_2,...,C_m}. As the result, we can get the number of clips in each class ($\sum C_i$). So the result of classification for each song can be expressed by a probability vector:

$$P(S)=\frac{1}{n}\{\sum C_1,\sum C_2,...,\sum C_m\},$$

The biggest one in P(S) will be chosen as the final result. But there is a situation that the vector has more than one maximum value. That is sensible because the song may have multi-emotion, and some of them expressed strongly so we can't figure out which is stronger. In our experiment, if one of the biggest probabilities is corresponds to the label assigned before, then we think it found the strongest emotion.

TABLE II. RESULTS OF GMM (BEFORE POST-PROCESSING)

	C1	C2	C3	C4	C5	C6	C7
C1	52.63	6.43	7.60	5.26	1.17	22.81	4.09
C2	0.95	62.09	24.17	5.69	1.42	4.27	1.42
C3	6.35	8.99	57.14	25.40	1.06	1.06	0.00
C4	2.71	40.27	11.77	29.86	0.45	13.58	1.36
C5	8.85	0.00	3.65	0.00	46.35	35.42	5.73
C6	12.15	9.35	2.34	0.47	7.48	63.08	5.14
C7	22.10	26.22	1.87	5.24	8.24	11.99	24.35

TABLE III. RESULTS OF SVM (BEFORE POST-PROCESSING)

	C1	C2	C3	C4	C5	C6	C7
C1	57.31	6.43	4.68	6.43	7.02	15.21	2.92
C2	4.27	77.25	3.32	9.48	0.00	2.84	2.84
C3	11.64	5.29	65.08	15.87	1.06	0.00	1.06
C4	5.88	11.77	11.77	54.75	1.36	7.69	6.79
C5	1.56	0.52	0.00	17.19	64.06	8.85	7.81
C6	7.48	5.14	1.40	6.07	16.82	56.54	6.54
C7	11.24	6.37	5.99	3.37	9.36	17.98	45.69

V. EXPERIMENTAL RESULTS

In this section, the experimental results of two classifiers and four feature selection methods are giving. The classification accuracies of GMM and SVM are compared for deciding which one is used for the further experiments. Moreover, the results of different feature selection methods reveal which algorithm is more suitable for mood classification. All these work is done by using the dataset we have clarified above.

A. Classifer

TableⅡ and TableⅢ are the confusion matrixes before post-processing by using GMM and SVM. TableⅣ and Ⅴ are the ones after post-processing. Firstly, we should notice that the clips' label is the same as the song which contained them. For most songs, their emotion maybe change many times. That is, more than one mood existed in a song. Therefore, many clips may not have correct label in our dataset. But this would not affect us to find the strongest emotion in a song, because most clips in a song would express the strongest emotion. The accuracies of clips classification can be seen in first two tables. But these accuracies do not compatible with the post-processing results. They just told us that how many clips classified in accordance with their song. The average accuracies are 49.49% and 59.45% respectively. SVM seems performs better.

TABLE IV. RESULTS OF GMM (AFTER POST-PROCESSING)

	C1	C2	C3	C4	C5	C6	C7
C1	80	0	0	0	0	20	0
C2	0	88.9	0	11.1	0	0	0
C3	0	14.3	71.4	14.3	0	0	0
C4	0	20	0	60	0	20	0
C5	0	0	16.5	0	67	0	16.5
C6	0	0	0	0	0	100	0
C7	16.7	33.3	0	0	0	0	50

TABLE V. RESULTS OF SVM (AFTER POST-PROCESSING)

	C1	C2	C3	C4	C5	C6	C7
C1	60	0	0	0	0	40	0
C2	0	100	0	0	0	0	0
C3	0	14.3	85.7	0	0	0	0
C4	0	0	10	80	0	10	0
C5	0	0	16.5	0	67	16.5	0
C6	0	0	0	0	0	100	0
C7	0	0	0	0	0	0	100

Table IV and Table V are the confusion matrixes of post-processing by using GMM and SVM. These results show whether the songs' strongest emotion is found correctly. Figure 2 shows the performance of two classifiers in each mood-class. Except the first class, SVM apparently performed better, especially in the 7-class. The average accuracies are 72.92% and 85.42%. We can see using SVM is better than using GMM. So the following experiments will use SVM as basic classifier.

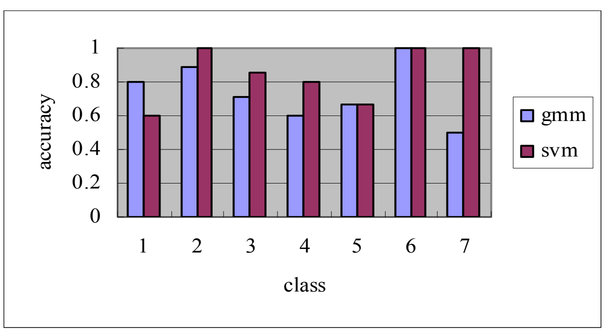

Figure 2. Result of using two classifiers for 7-mood classification

B. Feature selection algorithms

1) SFS

Stepwise forward selection algorithm selects 41 features for our data set. The classification result is showed in Table VI and Table VII. On the one hand, although the dimensions are decreased from 88 to 41, but it performed not well in some classes, especially in 7-class, the accuracy in this class decreased a lot, both on clips and songs. The average accuracy after post-processing is decreased either, from 85.42% to 79.12%.

2) ReliefF

Table VIII and Table IX are the results of using ReliefF method, the dimension is 84. Accuracy of clips is a little high than using all features. And after post-processing, the average accuracy is the same as in Table V. Although the dimensions don't decreased a lot by using this method, but it can remove redundant features effectively.

TABLE VI. RESULT OF SFS (BEFORE POST-PROCESSING)

	C1	C2	C3	C4	C5	C6	C7
C1	47.37	4.68	3.51	7.60	11.70	19.88	5.26
C2	3.79	67.77	9.95	10.43	0.47	1.90	5.69
C3	16.93	5.82	58.20	17.46	0.53	0.53	0.53
C4	2.26	15.84	11.31	59.73	1.81	6.33	2.71
C5	1.56	0.52	0.00	17.71	67.71	11.98	0.52
C6	9.35	8.41	7.01	6.07	9.35	44.86	14.95
C7	9.36	5.99	7.12	7.87	18.35	31.84	19.48

TABLE VII. RESULT OF SFS (AFTER POST-PROCESSING)

	C1	C2	C3	C4	C5	C6	C7
C1	60	0	0	0	0	40	0
C2	0	89	0	11	0	0	0
C3	0	0	100	0	0	0	0
C4	0	10	0	90	0	0	0
C5	0	0	0	17	83	0	0
C6	0	0	0	0	0	80	20
C7	0	0	0	16.7	0	50.3	33

TABLE VIII. RESULT OF RELIEFF (BEFORE POST-PROCESSING)

	C1	C2	C3	C4	C5	C6	C7
C1	57.31	7.02	4.09	5.85	7.02	15.79	2.92
C2	4.27	77.73	3.32	8.53	0.00	3.32	2.84
C3	11.11	5.29	64.55	16.40	1.06	0.00	1.59
C4	4.98	11.77	13.12	54.30	1.36	7.69	6.79
C5	1.56	0.52	0.00	17.19	64.06	8.85	7.81
C6	7.94	5.61	0.93	6.07	16.36	57.01	6.07
C7	10.86	7.12	5.62	3.37	10.86	16.11	46.07

TABLE IX. RESULT OF RELIEFF (AFTER POST-PROCESSING)

	C1	C2	C3	C4	C5	C6	C7
C1	60	0	0	0	0	40	0
C2	0	100	0	0	0	0	0
C3	0	14.3	85.7	0	0	0	0
C4	0	0	10	80	0	10	0
C5	0	0	0	16.5	67	0	16.5
C6	0	0	0	0	0	100	0
C7	0	0	0	0	0	0	100

Figure 3. the classificatin accuricies for the 7-mood taxonomy by using different feature selection algorithms

3) Fisher Rule

The clips' average accuracy is 59.59%, high than using all feature, SFS and ReliefF methods. But as we said above, some clips may not be correct labeled, so even though they are classified in accordance with their labels, that is the clips' accuracy is high, it doesn't mean the strongest emotion can be found correctly. Here from Table ⅩⅣ we can see that after post-processing, the accuracy is 81.52%, it's lower than using all feature and ReliefF's result.

4) Active Selection

We choose the first 30 features as initial feature set and use SVM as basic classifier. The dimension is decreased from 88 to 40 and average accuracy of clips and songs is 57.41% and 83.33%, a little lower than using all features. We can see that Active Selection method can remove redundant features effectively, it's better than SFS method on the whole.

TABLE X. RESULT OF FISHER RULE (BEFORE POST-PROCESSING)

	C1	C2	C3	C4	C5	C6	C7
C1	57.3	5.8	4.1	5.3	7.0	18.1	2.3
C2	4.7	76.3	2.8	10.0	0.0	3.8	2.4
C3	11.1	3.7	66.7	15.9	0.5	0.5	1.6
C4	4.5	14.5	15.4	51.6	0.9	6.3	6.8
C5	3.1	0.5	0.0	17.2	63.5	7.3	8.3
C6	7.0	3.7	1.9	5.1	16.4	57.9	7.9
C7	10.9	5.6	4.9	3.0	7.1	20.6	47.9

TABLE XI. RESULT OF FISHER RULE (AFTER POST-PROCESSING)

	C1	C2	C3	C4	C5	C6	C7
C1	60	0	0	0	0	40	0
C2	0	100	0	0	0	0	0
C3	0	0	100	0	0	0	0
C4	0	0	20	70	0	10	0
C5	0	0	0	16.5	67	0	16.5
C6	0	0	0	0	20	80	0
C7	0	0	0	0	0	17	83

Different feature selection algorithms are compared in figure3. Among them, ReliefF and Fisher Rule are Filter feature selection methods, and they did not make the dimensions decrease a lot, this can be seen from Table ⅩⅣ. ReliefF can remove some apparent weak features, but that is not enough, some redundant features also exist in the feature set. Fish Rule is better than ReliefF in removing features to some extent, but it remove some useful features so the final accuracy decreased. This maybe partly because our clips are not labeled accurately, otherwise the accuracy of clips should compatible with the final result. SFS and Active Selection algorithms are both used classifier. Training samples also divided randomly into 75% and 25% when using these methods to do feature selection. From Table ⅩⅣ we can see that the dimension decreased to 41 and 40, and the accuracy is not bad. So these two methods can remove weak features better than ReliefF and Fisher Rule. But some useful features maybe removed either, because they may play an active role for some classes and have side effect for others. In these two algorithms, Active Selection method is a combination of filter and wrapper approaches, it's better than SFS as a whole.

TABLE XII. RESULT OF ACTIVE SELECTION (BEFORE POST-PROCESSING)

	C1	C2	C3	C4	C5	C6	C7
C1	52.05	1.75	5.85	9.94	10.53	17.54	2.34
C2	4.74	71.56	6.64	13.27	1.42	0.47	1.90
C3	7.41	5.29	63.49	21.69	1.06	0.53	0.53
C4	5.88	16.29	19.01	50.23	0.45	2.26	5.88
C5	0.52	2.08	0.00	17.19	65.10	6.25	8.85
C6	5.14	4.21	2.34	4.21	17.29	57.01	9.81
C7	10.11	11.61	3.37	5.62	8.99	14.23	46.07

TABLE XIII. RESULT OF ACTIVE SELECTION (AFTER POST-PROCESSING)

	C1	C2	C3	C4	C5	C6	C7
C1	80	0	0	0	0	20	0
C2	0	89	0	11	0	0	0
C3	0	0	100	0	0	0	0
C4	0	10	20	70	0	0	0
C5	0	0	0	16.5	67	0	16.5
C6	0	0	0	0	0	100	0
C7	0	0	0	17	0	0	83

TABLE XIV. AVERAGE CLASSIFICATION ACCURACIES BY USING DIFFERENT ALGORITHMS

	SFS	ReliefF	Fisher Rule	Active Selection
Dimension	41	84	78	40
Accuracy (%)	79.12	85.42	81.52	83.33

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we created 7-class mood model and presented different feature selection algorithms and classifiers for mood classification. SVM is chosen as basic classifier because it has better performance. Four feature selection methods are compared. Although the classification accuracy is well by using ReliefF and Fisher Rule, the dimension can not be decreased a lot. We can see that active selection have a relatively better performance in removing weak features. It can make the dimension decreased a lot without accuracy dropped much.

In the future, we will find an approach to label training clips more accurately, then the result of clips classification can compatible with the songs' result. For active selection algorithm, how many features should be chosen into initial feature set is still a problem. We will try other way to choose initial features. And from the experimental result, we can see different classifiers and feature selection algorithms usually perform better in some classes and not well in others. Then these classifiers or feature sets maybe can be combined and perform even better.

VII. ACKNOWLEDGMENTS

We gratefully acknowledge the support of the National Natural Science Foundation of China under grant 60873098.

REFERENCES

[1] Hevner, K., Expression in music: a discussion of experimental studies and theories. Psychological Review, 1935.42:P.246-248

[2] Thayer, R. E., The Biopsychology of Mood and Arousal, Oxford University Press, 1989.

[3] Hu X., J.Stephen Downie, The 2007 Mirex Audio Mood Classification Task: Lessons Learned, Proceedings of the 9th International Conference on Music Information Retrieval, ISMIR'08, Philadelphia, USA, 2008

[4] K. Trohidis, G. Tsoumakas, G. Kalliris, and I. Vlahavas, Multi-label Classification of Music into Emotions, Proceedings of the 9th International Conference on Music Information Retrieval, ISMIR'08, Philadelphia, USA, 2008

[5] T. Li, M. Ogiharai, Detecting Emotion in Music, Johns Hopkins University, 2003

[6] D. Turnbull., L. Barrington, D. Torres, G. Lanckriet, Towards Musical Query-By-Semantic-Description using the CAL500 Data Set, Music Information Retrieval (ISMIR'07), July 2007: 439-446

[7] D. Turnbull, L. Barrington, D. Torres, G. Lanckriet, Semantic Annotation and Retrieval of Music and Sound Effects, IEEE Transactions on Audio, Speech, and Language Processing, 2008: 467-476

[8] Liu D., Lu L., Zhang H.J., Automatic Mood Detection from Acoustic Music Data, Proc. of the 4th Int. Conf, Music Information Retrieval (ISMIR'03), Washington, DC, USA, October 2003: 81-87

[9] Lu L., Liu D., Zhang H.J., Automatic Mood Detection and Tracking of Music Audio Signals, IEEE Transaction on Audio, Speech, and Language Processing, 2006: 5-18

[10] Feng Y.Z., Zhuang Y.T., Pan Y.H., Popular Music Retrieval by Detecting Mood, Music Information Retrieval (ISMIR'03), July 2003: 375-376

[11] Yang Y.H., Liu C.C., Chen H.H., Music Emotion Classification: A Fuzzy Approach, Association for Computing Machinery (ACM'06), October 2006:81~84

[12] www.guqu.net

[13] Z. Xiao, E. Dellandrea, W. Dou, and L. Chen, What is the Best Segment Duration for Music Mood Analysis, Content-Based Multimedia Indexing, 2008: 17-24

[14] D. McEnnis, C. McKay, I. Fujinaga, and P. Depalle, Jaudio: A Feature Extraction Library, Music Information Retrieval (ISMIR'05), September 2005: 600-603

[15] D. McEnnis, C. McKay, and I. Fujinaga, jAudio: Additions and Improvements, Music Information Retrieval (ISMIR'06), October 2005.

[16] K. Kira, L A. Rendell, The Feature Selection Problem: Traditional Methods and a New Algorithm, Proc of Ninth National Conference on Artificial Intelligence, 1992:129-134.

[17] Kononenko, I., Estimating attributes: analysis and extensions of Relief. European Conference on Machine Learning ,1994: 171–182

[18] Y. Liu, J. Xv, L. Wei, T. Yun, The Study of the Classification of Chinese Folk Songs by Regional Style, Proceedings of the First IEEE International Conference on Semantic Computing (ICSC 2007), September 2007: 657-662

[19] C.-W. Hsu, C.-C. Chang, C.-J. Lin. A practical guide to support vector classification. Technical report, Department of Computer Science, National Taiwan University. July, 2003.

[20] R.-E. Fan, P.-H. Chen, and C.-J. Lin. Working set selection using the second order information for training SVM. Journal of Machine Learning Research 6, , 2005: 1889-1918