

Performance Comparison of Multi-Class SVM Classification for Music Cultural Style Tagging

Sai Hseng Mao and Ei Ei Pe Myint

Abstract—In content centric music information retrieval (MIR); emotion, genre, similarity and style are well fashion in these days. Automatic classification of musical style is gaining more and more importance since it may serve as a way to structure and organize the increasingly large number of music files available on the Web. This paper presents cultural based music style classification task which is categorized three different cultural style; classical songs of Chinese, Indian and Myanmar. 1500 music pieces, 500 for each culture are employed for this work. Exclusively timbral texture feature set are extracted from audio files for training and testing. The Support Vector classifiers are applied for style classification based on cultural information. Performance of two different multiclass classification method, One-Against-One and One-Against-All (OAA) are compared as the main theme of this presentation. The experimental result show the 88.43% and 82.37% overall accuracy for three music cultural style classification. Moreover, the system achieved the highest accuracy of 89.13% in Myanmar culture using OAO support vector classifier and 87% in Myanmar culture using OAA support vector classifier respectively.

Index Terms—Music information retrieval, multi-class classification, OAO classification, OAA classification.

I. INTRODUCTION

The definition of style recognition is dissimilar. One of amongst defines musical style as “the identifiable characteristics of a composer’s music which are recognizable similar from one work to another”. Whatever the definition, music style and its recognition is something related to human nature: the average layperson can recognize the difference.

Automatic recognition of musical style is not an easy task. Even relatively simple stylistic manners of playing an instrument, such as playing energetically, playing lyrically or playing syncopation are difficult to detect reliably by automatic classification tools.

Culture and music are interrelated with each other because their style of playing music is differing as their culture differs. Therefore, music recognition using cultural style is important approach for automatic music recommending retrieving and playlist generation.

This paper employs a sort of timbral feature for comparing the performance of two One-Against-One (OAO) and One-Against-All (OAA) multiclass classification method.

The two multi-class classification method used by SVM is applied. Moreover, both of linear and Radial Basic Function

(RBF) kernel is also tested. It is compared the achievement of Precision and Recall for each classification method. What is more, the average training time of these two multi-class classifications are reported. To get the cultural influence from songs, we used classic songs for our experiment.

The rest of this paper is organized as follows. Section 2 discusses the related work of the system. The architecture of the system preprocessing, feature extraction and classification are discusses in Section 3 and 4. Evaluation and experimental results are present in Section 5.

II. RELATED WORK

There has already been some work toward automatic music style classification. In [1], self-organizing neural maps (SOM) is employed as music style classifiers of musical fragments. The monophonic melody track is extracted and cut into fragments of equal length. From these sequences, melodic, harmonic, and rhythmic numerical descriptors are computed from Musical Instrument Digital Interface (MIDI) and presented to the SOM. Their performance is analyzed in terms of separability in different music class of classical and jazz music. 430 jazz and 522 classical melodic samples of MIDI files, employing 8 bars are used to as train and test data. In experiment, 9 fold cross validation is applied and their accuracy is varied from 66% to 98%.

Early music style analysis usually assumed that each music piece of music has a unique style and they make use of music contents to construct a classifier for classifying each piece into unique style. Traditional music style classification approach is innovated by Tao Li et al [2]. Their basic idea is that an exact piece of music may match more than one, even several different styles. In their approach applied multi-label music style classification using Hypergraph integrated Support Vector Machine (HiSVM), which can integrate both music contents and music contents and tags for automatic music style classification. The 80 feature vector is extracted from each of audio music trimmed of different artists. Each of representative song obtained the style and tag description. It can be seen that their method cannot only take into account the music style correlations, but also the music tag correlations.

Liu *et al.* [3] presented cultural style based music classification using 6 different cultures Western, Chinese, Japanese, Indian, Arabic and African. Timbre feature, wavelet feature and musicology-based features sets are extracted. Decision tree, OAO and OAA SVM and k-NN classifiers are applied for classification purpose. The highest accuracy of 86.50% is achieved by SVM-OVA using all features.

In [4], cultural are treated as effective feature set for music

Manuscript received September 28, 2012; revised November 20, 2012.
S. H. Mao is with the Computer University (Lashio) Myanmar (e-mail: saih sengmao2012@gmail.com).

E.-E. P. Myint is with the Department of Software Technology, Computer University (Lashio) Myanmar (e-mail: eieipemyint@gmail.com).

genre based style classification. Cultural features for music retrieval are presented in [5].

III. ARCHITECTURE OF THE SYSTEM

The overview of music style tagging based on cultural information is presented in Fig. 1.

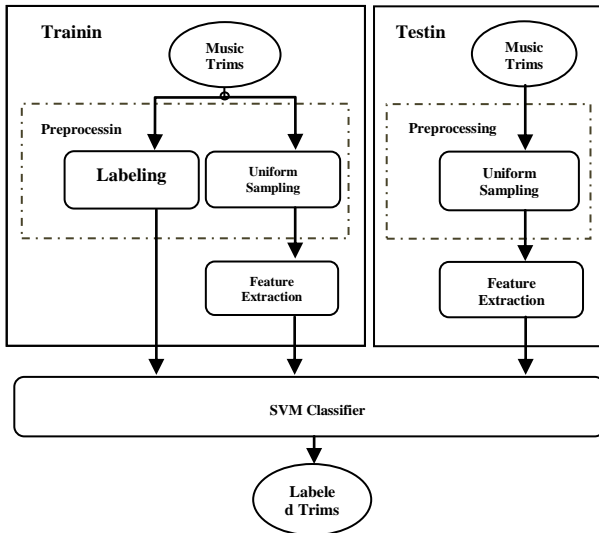


Fig. 1. Architecture of the system.

The system counted as three main categories; *pre-processing*, *feature extraction* and *tagging*. Before data is actually processes such as extracting features and classifying, data had to be recognized as the uniform format. Then, the necessary feature sets are extracted for tagging cultural label. In classification, the system employed OAO SVM and OAA SVM classifiers for performance comparison purpose. The detail tasks carried out by each component can be seen in the following.

A. Preprocessing

Altogether 204 songs, 82 for Chinese classical, 65 for India classical and 57 for Myanmar classical songs are collected. Before feature extraction, all the collected songs are down sampled to 22.05 kHz, 16 bit/sample, mono channel WAV format for uniform format. And songs are manually labeled related to their cultural style. 500 collections of music trims for each cultural group are utilized.

B. Feature Extraction

Every 30 second trims without initial 30 second are extracted from every song for feature extraction.

Some sort of timbre features in [6]; Mel frequency spectral coefficient, spectral centroid, spectral flux and spectral rolloff are extracted. Brief descriptions of extracted audio features are:

Mel-Frequency Cepstral Coefficients: Mel-frequency cepstral coefficients (MFCC) are perceptually motivated features that are also based on the STFT [6]. After taking the log-amplitude of the magnitude spectrum, the FFT bins are grouped and smoothed according to the perceptually motivated Mel-frequency scaling. Finally, in order to decorrelate the resulting feature vectors a discrete cosine transform is performed. In spite of the fact that 24 coefficients are used for high level song representation, the

system has found 13 coefficients provide the best music style tagging performance.

Spectral Centroid: amplitude weighted mean of the spectrum;

Spectral Flux: Euclidean distance between normalized spectrum distributions of two successive frames;

Spectral Rolloff: the frequency boundary. 95% of the magnitude distribution is concentrated experimentally.

Musical feature sets, in this paper, songs or music trims are divided into non-overlapping frames of 20-ms length, known as framing, for preliminary of feature extraction. The Mean, Standard Deviation and Variance of extracted feature, altogether 48 feature dimensions are applied for training and testing.

C. Tagging

Though SVM is binary classifier, the system adopted OAO and OAA multi-class classification method.

In the OAO approach, each class is compared to each other class [7]. A binary classifier is built to discriminate between each pair of classes, while discarding the rest of the classes. This requires building $K(K-1)/2$ binary classifiers. A voting is performed among the classifiers and the class with the maximum number of votes wins.

The simplest approach is to reduce the problem of classifying among K classes into K binary problems, where each problem discriminates a given class from the other $K-1$ classes [7]. For this approach, the system require $N = K$ binary classifiers. The classifier producing the maximum output is considered the winner, and this class label is assigned.

Because the system employ 3-class classification approaches, both the two multi-class classification method needed equal amount, 3 classifiers. For each binary SVM, the linear kernel and radius bias kernel function is used. The detail process of tagging is presented in Fig 2.

Before classification, the trained data are trained using SVM. After that, each tested song is injected to every three binary classifiers for classifying Chinese/India/Myanmar classical According to the results by three binary classifiers, the system picks up the candidate label which have majority vote count among 3 classification consequences.

Sets of 48 feature dimensions are extracted from each music clip and prepared as training and testing data. The system used and external 10-fold cross-validation on the training and testing. Each of one fold dataset is leaved out for testing for each of one fold classification. The rest datasets are injected into the classifiers to be trained. After that, omitted one fold is injected into the classifiers for testing purpose.

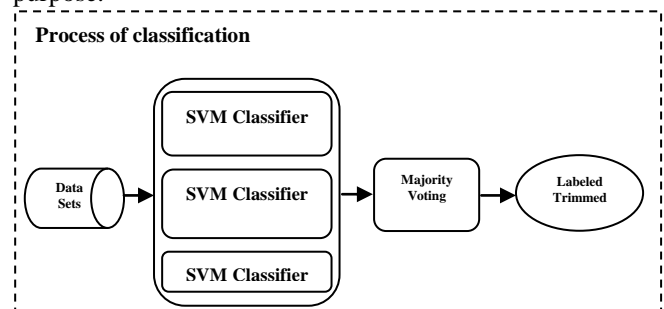


Fig. 2. Classification process.

IV. EVALUATION

The system applied OAO and OAA SVM classifier sets of liner and RBF kernel. To decide the suited sigma value for this music style classification system various sigma values, 1 to 10 is tested alternatively. Finally, RBF sigma value is empirically chosen as 3 for this experiment.

The performance measure using SVM Linear kernel and SVM RBF kernel for OAO approach and SVM Linear kernel and SVM RBF kernel for OAA approach is presented in Table I.

TABLE I: CLASSIFICATION RESULTS

	Chinese	Indian	Myanmar	Missed	Unclassified
SVM Linear (OAO)	420	436	447	161	36
SVM RBF (OAO)	441	451	458	144	6
SVM Linear (OAA)	390	414	435	94	167
SVM RBF (OAA)	373	415	444	46	222

According to Table I, both of OAO SVM and OAA SVM introduced unclassified region. The resulted classification presented that OAA SVM devoted much un-classifiable region in both using linear and RBF kernel. Much error rate in OAO SVM introduced than error rate in OAA SVM in both kernels, linear and RBF.

To clearly show the discrimination power of feature set in style recognition, the corresponding performance on each style cluster using SVM OAO with linear kernel, SVM OAO with RBF kernel, SVM OAA with linear kernel, SVM OAA with RBF kernel, is listed in Table II, Table III, Table IV and Table V respectively. Where each row corresponds to the actual cultural style cluster and each column corresponds to the predicted cluster.

TABLE II: CONFUSION MATRIX USING OAO SVM LINEAR KERNEL

	Chinese Songs	Indian Songs	Myanmar Songs
Chinese Songs	420	46	19
Indian Songs	39	436	14
Myanmar Songs	32	11	447

TABLE III: CONFUSION MATRIX USING OAO SVM RBF KERNEL

	Chinese Songs	Indian Songs	Myanmar Songs
Chinese Songs	441	48	8
Indian Songs	42	451	6
Myanmar Songs	37	3	458

TABLE IV: CONFUSION MATRIX USING OAA SVM LINEAR KERNEL

	Chinese Songs	Indian Songs	Myanmar Songs
Chinese Songs	390	33	11
Indian Songs	25	414	4
Myanmar Songs	18	3	435

TABLE V: CONFUSION MATRIX USING OAA SVM RBF KERNEL

	Chinese Songs	Indian Songs	Myanmar Songs
Chinese Songs	373	27	3
Indian Songs	9	415	4
Myanmar Songs	3	0	444

According to the confusion matrix, Chinese classical songs are mostly missed classified as Indian classical songs, it brings in 7.7% and Indian classical songs are much missed classified as Chinese classical songs brings in 5.75% conversely. Myanmar classical songs are mostly missed classified as Chinese classical songs distinctly, 4.5%.

Table VI lists the recall and precision of cultural style music classification categorized 3 different culture, Chinese, Indian and Myanmar classic songs.

TABLE VI: RECALL AND PRECISION

		Chinese	Indian	Myanmar	Missed	Unclassified
SVM Linear kernel (OAO)	Recall	84.00%	87.2%	89.40%	10.73%	2.4%
	Precision	85.54%	88.4%	93.13%	10.73%	2.4%
SVM RBF kernel (OAO)	Recall	88.20%	90.1%	93.13%	9.6%	0.4%
	Precision	84.81%	89.8%	97.03%	9.6%	0.4%
SVM Linear kernel (OAA)	Recall	78.00%	82.8%	87.00%	6.26%	11.1%
	Precision	90.00%	92.0%	96.70%	6.26%	11.1%
SVM RBF kernel (OAA)	Recall	74.60%	83.0%	88.80%	3.07%	14.8%
	Precision	96.80%	93.9%	98.45%	3.07%	14.8%

The system achieves maximum recall of 93.13% using OAO SVM RBF kernel. Maximum achieved precision is 98.45% OAA RBF kernel.

Required training time: While OAO SVM linear kernel inserts an hour, 23 minutes and 6 seconds, though OAA SVM linear kernel introduced 5 hours, 53 minutes and 47 seconds. Meanwhile, OAO SVM RBF kernel introduced 48 minutes and 26 seconds, OAA SVM RBF kernel put in 3 hours, 45 minutes and 24 seconds. All the training time discussed in this paper is calculated upon Intel® Core TM i5-2410M CPU

@ 2.30GHz with 2 GB running over Microsoft Window 7 Home Premium.

V. CONCLUSION

Because each piece of music has a unique style, automatic classical/traditional music style tagging is interested for music information retrieval. This paper presented cultural music style recognition using 3 different Music style categories; Chinese classical songs, Indian Classical songs and Myanmar classical songs. Unique timbral texture feature set is extracted for testing and training. Entirely 1500 music trims is used for applying music style classifying. The powerful machine learning tools SVM is employed in this experiment. This paper emphasized performance comparison of 2 multi-class classification methods, OAO and OAA SVM classification. According to the experimental results, this testing gives appropriate classification accuracy, 93.17% in OAO SVM classification although it is only timbral texture feature are used. Consequent results by the system, we can conclude the OAO SVM endorses much classification than OAA SVM classifier. What is more, OAO SVM required much less training time than OAA SVM distinctly.

ACKNOWLEDGEMENT

Author would like to thank to his supervisor Dr. Ei Ei Pe Myint and who is directed or indirect support during his researches. Author would not be standing here before you all without the love, care and continuous support from my family: mother, elder sister and younger sister. Nothing can sufficiently speak my word of thanks to them all.

REFERENCES

- [1] P. J. Ponce de Le'ón and J.M. G'nesta, "Feature-driven recognition of music styles," in *Proc. of the 1st Iberian Conference on Pattern Recognition and Image Analysis*. LNCS, 2652. pp. 773–781, 2003.
- [2] F. Wang *et al.*, "Tag Integrated Multi-label Music Style Classification with Hypergraphs," in *Proc. of the 10th International Society for Music Information Retrieval Conference (ISMIR)*. Kobe, Japan, 2009.
- [3] Y. Liu *et al.*, "Cultural Style Based Music Classification of Audio Signals," in *Proc. IEEE international conference on Acoustic Speed and Signal Processing*, pp. 57-60, 2009.
- [4] B. Whitman and P. Smaragdis, "Combining Musical and Cultural Features for Intelligent Style Detection," in *Proc. of the 3rd International Conference on Music Information Retrieval (ISMIR '02)* pp. 47-52, October 2002.
- [5] W.W. Cohen and W. Fan, "Webcollaborative filtering: recommending music by crawling the web," *WWW9/Computer Networks*, vol. 33, no. 1-6, pp 685-698, 2000.
- [6] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 5, pp. 293–298, Jul 2002.
- [7] M. Aly, "Survey on Multiclass Classification Method," Technical report, Caltech, USA, 2005.



Sai Hseng Mao received the B. C. Tech. (Hons:) and M.C.Tech degree from Computer University Lashio, Myanmar. His research interests are in the areas of signal processing, machine learning for music content analysis with emphasis on music information retrieval.



Ei Ei Pe Myint received Master degree from the University of Computer Study Mandalay (UCSM), Myanmar and awarded Ph.D degree in University of Computer Studies, Yangon (UCSY) at 2010. She is Assistant Lecture, Department of computer software technology in Computer University Lashio, Myanmar. She is interested in the areas of Digital Audio Processing especially in Music Recognition.