MUSICAL GENRE CLASSIFICATION USING SUPPORT VECTOR MACHINES

Changsheng Xu, Namunu C. Maddage, Xi Shao, Fang Cao, Qi Tian

Laboratories for Information Technology
21 Heng Mui Keng Terrace, Singapore 119613

ABSTRACT

Automatic musical genre classification is very useful for music indexing and retrieval. In this paper, an efficient and effective automatic musical genre classification approach is presented. A set of features is extracted and used to characterize music content. A multi-layer classifier based on support vector machines is applied to musical genre classification. Support vector machines are used to obtain the optimal class boundaries between different genres of music by learning from training data. Experimental results of multi-layer support vector machines illustrate good performance in musical genre classification and are more advantageous than traditional Euclidean distance based method and other statistic learning methods.

1. INTRODUCTION

The rapid development of various affordable technologies for multimedia content capturing, data storage, high bandwidth/speed transmission and the multimedia compression standards such as JPEG and MPEG, have resulted in a rapid increase of the size of digital multimedia data collections and greatly increased the availability of multimedia contents to the general user. Digital music is one of the most important data types distributed by the Internet and the amount of digital music increases rapidly nowadays. How to effectively organize and process such large variety and quantity of musical data to allow efficient indexing, searching and retrieval is a big challenge.

Musical genre is an important description that can be used to classify and characterize music from different sources such as music shops, broadcasts and Internet. It is very useful for music indexing and content-based music retrieval. For human being, it is not difficult to classify music into different genres. Although to let computers understand and classify musical genre is a big challenge, there are still perceptual criteria related to the melody, tempo, texture, instrumentation and rhythmic structure that can be used to characterize and discriminate different musical genres.

In this paper, a novel automatic musical genre classification approach is presented. Music is divided into four major categories: classic, jazz, pop and rock. In order to discriminate different musical genres, a set of music features is developed to characterize music content of different genres and support vector machine (SVM) learning approach is applied to build a multi-layer classifier. For different layers, different features and support vectors are employed. In the first layer, music is classified into pop/classic and rock/jazz music using SVM to obtain the optimal class boundaries. In the second layer, pop/classic music is further classified into pop and classic music and rock/jazz music is classified into rock and jazz music. This multi-layer classification method can provide a better classification result than current existing methods.

2. RELATED WORK

A number of methods have been done to discriminate music, speech, silence, and environment sound [1]. It is extremely more difficult to discriminate musical genres than discriminate music, speech and other sounds. Several researches focus musical genre classification on MIDI files. Shan [2] investigated the classification of music style by melody from a collection of MIDI music. Chai [3] employed hidden Markov model to model and classify the melodies of Irish, German and Austrian folk song. Dannenberg [4] extracted 13 features from MIDI and used different classifiers to recognize music style. However, MIDI data is a structured format, so it is easy to extract features according to its structure. Actual sounds such as wav and mp3 files are different from MIDI, thus MIDI style classification is not practical in real applications. Matityaho [5] discriminated classic and pop music by using the average amplitude of Fourier transform coefficients and neural network. Soltau [6] classified music into rock, pop, techno and classic using HMM and ETM-NN to extract the temporal structure from the sequence of cepstral coefficients. Han [7] classified music into classic, jazz and pop using simple spectral features and the nearest mean classifier. Pye [8] used Mel-frequency cepstral coefficients (MFCC) and Gaussian mixture model (GMM) to classify music into six types of blues, easy listening, classic, opera, dance and rock. Jiang [9] used octave-based spectral contrast feature and GMM to classified music into five types. All these methods tried to use one classifier and several features to classify music into different genres at a time. However, according to music knowledge, it is easy to discriminate some genres of music (i.e., pop and classic) using some features, but it may be difficult to use same features to discriminate other genres of music (i.e., pop and rock). Therefore, to further improve the classification accuracy, we should consider using different features and classifiers to discriminate different genres of music.

3. FEAUTRE SELECTION

Feature selection is important for music content analysis. The selected features should reflect the significant characteristics of different kinds of music signals. In order to better discriminate different genres of music, we consider the features that are related to temporal, spectral and cepstral domains. The selected features include beat spectrum, LPC (linear predictive coding,

zero crossing rates, spectrum power and mel frequency cepstral coefficients.

3.1. Beat Spectrum

Beat spectrum is a measure to automatically characterize the rhythm and tempo of the music. Highly structured or repetitive music will have strong beat spectrum peaks at the repetition times. This reveals both tempo and the relative strength of particular beats, and therefore can distinguish between different kinds of rhythms at the same tempo.

The beat spectrum can be calculated from the music using three principal steps. First, the music is parameterized using a spectrum or other representation. This results in a sequence of feature vectors. Second, a distance measure is used to calculate the similarity between all pairwise combinations of feature vectors. The obtained similarity is embedded into a two-dimensional representation called similarity matrix. Finally, the beat spectrum can be obtained from finding periodicities in the similarity matrix, using diagonal sums or auto-correlation. Figure 1 illustrates the beat spectrum of pop, classic, rock and jazz music.

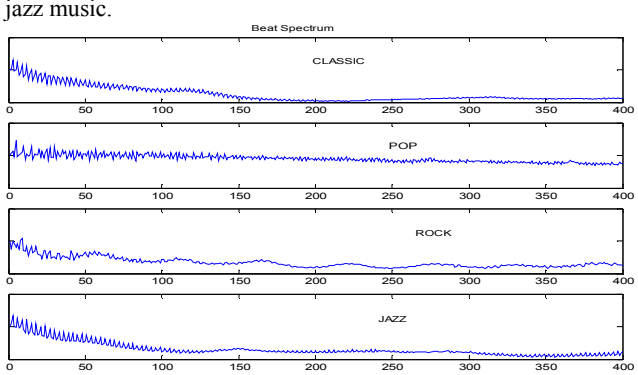

Figure 1Beat spectrum for pop, classic, rock and jazz

3.2. LPC-Derived Cepstrum

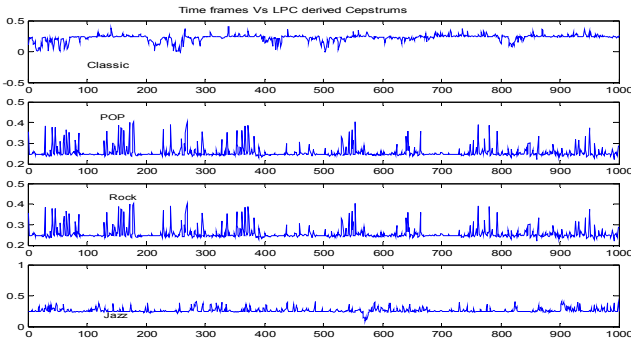

Figure 2 LPC-derived cepstrum for pop, classic, rock and jazz

The basic idea behind linear predictive analysis is that a music sample can be approximated as a linear combination of past music samples. By minimizing the sum of the squared differences (over finite interval) between the actual music samples and the linear predictive ones, a unique set of predictor coefficients can be determined. The importance of linear prediction lies in the accuracy with which the basic model applies to vocal signals in music. Figure 2 is an example of LPC (linear prediction coding) derived cepstrum for classic, pop, rock and jazz music.

3.3. Zero Crossing Rate

In the context of discrete-time signals, a zero crossing is said to occur if successive samples have different algebraic signs. The rate at which zero crossings occur is a simple measure of the frequency content of a signal. This average zero-crossing rate gives a reasonable way to estimate the frequency of sine wave. The number of zero crossing is also a useful feature in music analysis. Zero crossing rate is suitable for narrowband signals, but music signals include both narrowband and broadband components. Therefore, the short-time zero crossing rate can be used to characterize music signal. Figure 3 is an example of zero crossing rates for rock and jazz music.

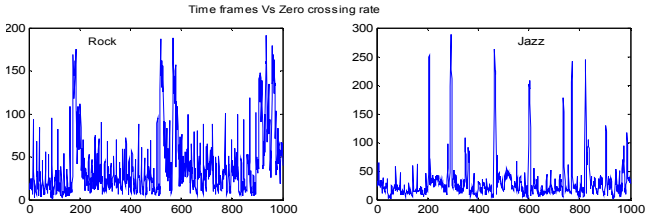

Figure 3 Zero crossing rates for rock and jazz

3.4. Spectrum Power

For a music signal $s(n)$, each frame is weighted with a Hanning window, $h(n)$:

$$h(n) = \frac{\sqrt{8/3}}{2}\left[1 - \cos(2\pi \frac{n}{N})\right] \quad (1)$$

where N is the number of samples of each frame.

The spectral power of the signal $s(n)$ is calculated as

$$S(k) = 10\log_{10}\left[\frac{1}{N}\left\|\sum_{n=0}^{N-1} s(n)h(n)\exp(-j2\pi \frac{nk}{N})\right\|^2\right] \quad (2)$$

The maximum is normalized to a reference sound pressure level of 96dB. Figure 4 is an example of short time energy for pop and classic music.

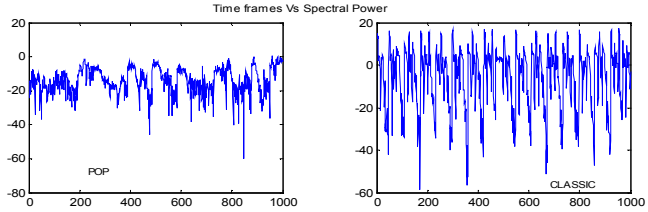

Figure 4 Spectrum power for pop and classic

3.5. Mel Frequency Cepstral Coefficient

The mel-frequency cepstrum has proven to be highly effective in automatic speech recognition and in modeling the subjective pitch and frequency content of audio signals. The mel-cepstral features can be illustrated by the Mel-Frequency Cepstral Coefficients (MFCCs), which are computed from the FFT power coefficients. The power coefficients are filtered by a triangular

band pass filter bank. The filter bank consists of K=19 triangular filters. They have a constant mel-frequency interval, and covers the frequency range of 0Hz – 20050Hz. Figure 5 is an example of MFCCs for pop and classic music.

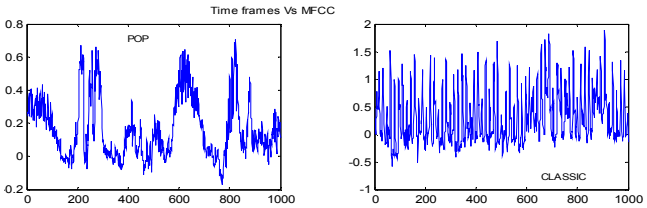

Figure 5 MFCCs for pop and classic

4. CLASSIFICATION

To achieve the best classification accuracy, a multi-layer classifier based on SVM is used to discriminate musical genres. In the first layer, music is classified into pop/classic and rock/jazz music according to the features of beat spectrum and LPC-derived cepstrum. In the second layer, pop/classic music is further classified into pop and classic music according to the features of LPC-derived cepstrum, spectrum power and MFCCs, and rock/jazz music is further classified into rock and jazz music according to the features of zero crossing rates and MFCCs. SVM is used in all layers and each layer has different parameters and support vectors. The system diagram of musical genre classification is illustrated in Figure 6.

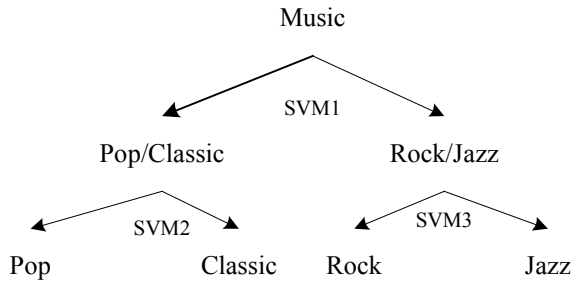

Figure 6 Musical genre classification diagram

4.1. Support Vector Machine

Support vector machine (SVM) learning is a useful statistic machine learning technique that has been successfully applied in the pattern recognition area [10].

Suppose we are given a set of training data $(x_1, x_2, ..., x_n)$ and their class labels $(y_1, y_2, ..., y_n)$, where $x_i \in R^n$ and $y_i \in \{-1, +1\}$, and we want to separate the training data into two classes. If the data are linearly non-separable but non-linearly separable, the non-linear SVM classifier will be applied. The basic idea is to transform input vectors into a high-dimensional feature space using non-linear transformation Φ, and then to do a linear separation in feature space.

To construct a non-linear SVM classifier, inner product $< x, y >$ is replaced by a kernel function $K(x, y)$.

$$f(x) = \mathrm{sgn}(\sum_{i=1}^{l} \alpha_i y_i K(x_i, x) + b) \quad (3)$$

The SV algorithm can construct a variety of learning machines by use of different kernel functions. Three kinds of kernel functions are constantly used. They are:

(1) Polynomial kernel of degree d

$$K(x, y) = (< x, y > +1)^d \quad (4)$$

(2) Radial basis function with Gaussian kernel of width $c > 0$

$$K(x, y) = \exp(-\|x - y\|^2 / c) \quad (5)$$

(3) Neural networks with tanh activation function

$$K(x, y) = \tanh(k < x, y > +\mu) \quad (6)$$

where the parameters k and μ are the gain and shift.

4.2. SVM Learning

We use non-linear support vector classifier to discriminate different musical genres. Therefore, classification parameters should be derived using support vector machine learning. Figure 7 illustrates a conceptual block diagram of the training process to produce classification parameters of classifier. The training process analyses musical training sample data to find an optimal way to classify musical frames into relevant genres. The derived classification parameters are used to discriminate different musical genres. Since we use three SVM classifiers and use different features to train these classifiers, the parameters corresponding to three classifiers are different.

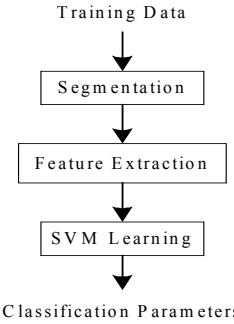

Figure 7 SVM learning process

5. EXPERIMENTAL RESULTS

To further illustrate and evaluate proposed musical genre classification approach, experiments are conducted for various genres of music samples.

5.1. Dataset Collection

The music dataset used in musical genre classification experiment contains 100 music samples. They are collected from music CDs and Internet and cover different genres such as classic, jazz, pop and rock. All data are 48.0kHz sample rate, stereo channels and 16 bits per sample. In order to make training results statistically significant, training data should be sufficient and cover various genres of music.

5.2. Classification Results

We select 60 music samples as training data including 15 pop music, 15 classic music, 15 rock music and 15 jazz music. Each sample is segmented into 2000 frames and the length of each frame is 882 sample points. Therefore, the total number of training data is 120,000 frames. For the SVM1 which is used to classify music into classic/jazz and pop/rock, 60,000 frames including 15,000 frames of each genre are used for training. For the SVM2 which is used to classify classic/jazz into classic and jazz, 40,000 frames are used for training. Among these training frames, 10,000 frames are from SVM1 training set with 5,000 frames of classic and jazz respectively; the other 30,000 frames are from new training frames with 15,000 frames of classic and jazz respectively. For SVM3 which is used for classify pop/rock into pop and rock, 40,000 frames are used for training. The training frames selected for SVM3 is similar to those for SVM2. 10,000 frames are from SVM1 training set and 30,000 frames from new training frame. The rest 40 samples are used as a test set.

Radial basis function with c=1 is used for SVM1 and c=2 for SVM2 and SVM3 as the kernel function in SVM training and classification. After training the SVMs, we use them as the classifiers to separate classic, jazz, pop and rock frames on the test set. The test set contains 10 classic music samples (20,000 frames), 10 jazz music samples (20,000 frames), 10 pop music samples (20,000 frames) and 10 rock music samples (20,000 frames). Table 1 shows the number of training and test data, support vectors obtained, and test error for SVM1, SVM2 and SVM3 respectively. It can be seen that our proposed approach can achieve an ideal result in musical genre classification.

Table 1 SVM training and test results

	SVM1	SVM2	SVM3
Training Set	60,000	30,000	30,000
Support Vectors	4325	8327	7684
Test Set	40,000	20,000	20,000
Error Rate	6.36%	7.42%	6.79%

5.3. Comparison

To further illustrate the advantage of proposed approach, we compare the performance of proposed method and other methods including nearest neighbor (NN), Gaussian mixture model (GMM) and hidden Markov model. The same training set and test set are used for these methods. Table 2 shows the comparison result among these methods. It can be seen that our proposed method achieves a significantly higher accuracy rate than other methods.

Table 2 Comparison result

	SVM	NN	GMM	HMM
Error Rate	6.86%	20.57%	12.31%	11.94%

6. CONCLUSIONS AND FUTURE WORK

We have presented and demonstrated an automatic classification approach for musical genres using multi-layer support vector machine learning. Beat spectrum, linear prediction coefficients, zero crossing rates, short time energy and mel-frequency cepstral coefficients are calculated as features to characterize music content. Three nonlinear support vector machine classifiers are developed to obtain the optimal class boundaries between classic/jazz and pop/rock, classic and jazz, and pop and rock by learning from training data. For each SVM learning and classification, different music features are used. Experiments show the multi-layer support vector machine learning method has good performance in musical genre classification and is more advantageous than traditional Euclidean distance based method and other statistic learning methods.

There are two directions that need to be investigated in the future. The first direction is to improve the computational efficiency for support vector machines. Support vector machines take a long time in the training process, especially with a large number of training samples. Therefore, how to select proper kernel function and determine the relevant parameters is extremely important.

The second direction is to make the classification result more accurate. To achieve this goal, we need to explore more music features that can be used to characterize the music content.

7. REFERENCES

[1] L. Lu, H. Jiang and H. J. Zhang, "A Robust Audio Classification and Segmentation Method", In *Proc. ACM Multimedia 2001*, Ottawa, Canada, 2001.

[2] M.K. Shan, F.F. Kuo and M.F. Chen, "Music Style Mining and Classification by Melody", in *Proc. of IEEE ICME02*, Lausanne, Switzerland, 2002.

[3] W. Chai and B. Vercoe, "Folk Music Classification Using Hidden Markov Models", In *Proc. of IC-AI01*, 2001.

[4] R.B. Dannenberg, B. Thom and D. Waston, "A Machine Learning Approach to Musical Style Recognition", In *Proc. of ICMC97*, 1997.

[5] B. Matityaho and M. Furst, "Neural Network Based Model for Classification of Music Type", in *Proc. of 18th Conv. Electrical and Electronic Engineers in Israel*, pp.1-5, 1995.

[6] H. Soltau, T. Schultz, M. Westphal and A. Waibel, "Recognition of Music Types", In *Proc. of IEEE ICASSP98*, pp.1137-1140, 1998.

[7] K.P. Han, Y.S. Pank, S.G. Jeon, G.C. Lee and Y.H. Ha, "Genre Classification System on TV Sound Signals Based on a Spectrogram Analysis", *IEEE Trans. on Consumer Electronics*, 55(1):33-42, 1998.

[8] D. Pye, "Content-Based Methods for the management of Digital Music", In *Proc. of IEEE ICASSP00*, pp.2437-2440, 2000.

[9] D.N. Jiang, L. Lu, H.J. Zhang, J.H. Tao and L.H. Cai, "Music Type Classification by Spectral Contrast Feature", In *Proc. of IEEE ICME02*, Lausanne, Switzerland, 2002.

[10] V. Vapnik, *Statistical Learning Theory*, Wiley, 1998.