Musical Genre Classification using Gaussian Mixture Models

Su Myat Mon Oo, Khin San Aye
University of Computer Studies, Yangon
smmoucsy08@gmail.com

Abstract

Digital music is one of the most important data types, distributed by the Internet. Automatic musical genre classification is very useful for music indexing and retrieval. A method to recognize the genre of music audio is considered. In this paper, the input music is represented with DWT (Discrete Wavelet Transform) *coefficients and classifying the extracted features is performed using Gaussian Mixture Models* (*GMM*). *Using GMM the optimal class boundaries between four groups of genre namely, pop, classic, rock and jazz are obtained. The feature vector from feature extraction step uses wavelet coefficients by hierarchical decomposition as it is easy to implement as well as it can reduce the computation time and resources required. Given that GMM is a robust approach that could obtain very good performance and a solution based on it is powerful, the classification is mainly composed of GMM classifiers. The experimental results indicate that the proposed approach offer encouraging results.*

1. Introduction

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. Significant progress in network, data storage and retrieval technologies resulted in fact that there is a huge amount of musical recordings data available for users all over the world. These places are first of all commercial musical databases and popular commercial “mp3 download” sites in the World Wide Web. Nowadays, many music companies are putting music products on websites and customers can purchase them online. But from the customer point of view, they would prefer to listen to the highlights of the music and to know type of music before they make a decision on whether to purchase or not. Although genre classifiers are available on some websites, they are generated manually, which needs expensive manpower and is time-consuming.

Musical genre classification task falls into two major stages: feature extraction and classification. In this paper, a study of wavelet based feature extraction is presented in developing musical genre classification system. A feature vector based on discrete wavelet transform of input audio data is proposed. The features extracted are used to train and test through a classification scheme where GMM is employed as a base classifier. The iterative Expectation Maximization (EM) algorithm is used to estimate the parameters for each Gaussian component and the mixture weights.

This paper is organized as follows: Section 2 describes related work. Overviews of the DWT and the GMM are given in Section 3. Section 4 describes the processing step of the system. Architecture of the classification is also described in this section. Finally, results from music genre classification of songs are provided in Section 5.

2. Related Work

Initial study of wavelet based features extraction in the task of musical genre classification method used in [1]. Discrete wavelet transform method is used for audio classification method is in [2]. Jiang [3] used octave-based spectral contrast feature and GMM to classified music into five types. 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 is described in [4]. Hidden Markov Model is used in [6] for classification music genre. Methodology of automatic musical genre classification described by Tzanetakis in [7] represents an up-to-date system, based on advances feature extraction. Their proposed features are timbral texture features and rhythmic content features. There has also been some recent work on automatic music genre detection. In [8] standard features are combined with representations of rhythm and pitch content and classification performance are shown in the range of 60%. Xu [9] used Support Vector Machines for classification and MFCC, beat spectrum, LPC-derived cepstrum, zero-crossing rate as features to classify music into four genres: rock, pop, jazz and classic. All these methods tried to use in one classifier and several features to classify music into different genres at a time.

3. Background Theory

In this presented paper, the discrete wavelet transform is used as the main feature representation of music signal and Gaussian Mixture Model is used to differentiate among music genres.

3.1. Discrete Wavelet Transform

The discrete wavelet transform (dwt) was widely used for many applications such as watermarking, edge detection, image processing, speaker verification and … etc. The Discrete Wavelet Transform is a special case of the wavelet transform that provides a compact representation of a signal in time and frequency that can be computed efficiently. The DWT analysis can be performed using a fast pyramidal algorithm related to multirate filerbanks. As a multirate filter bank the DWT can be viewed as a constant Q filter bank with octave spacing between the centers of the filters. Each sub band contains half the samples of the neighboring higher frequency sub band. In wavelet analysis contains two parts: coarse approximation and details information. The approximation is the high-scale, low-frequency components of the signal. The detail is the low-scale, high frequency components. In the pyramidal algorithm the signal is analyzed at different frequency bands with different resolution by decomposing the signal into a coarse approximation and detail information. The coarse approximation is then further decomposed using the same wavelet decomposition step.

This is achieved by successive high pass [g] and low pass [h] filtering of the time domain signal (x[n]) and is defined by the following equations:

$$y_{high}[k] = \sum_n x[n]\, g\,[2k - n] \qquad (1)$$

$$y_{low}[k] = \sum_n x[n]\, h\,[2k - n] \qquad (2)$$

Features extraction involves information retrieval form the audio signal. A variety of different wavelet families have been proposed in the literature. In this paper, the 4 coefficient wavelet family (DAUB4) proposed by Daubechies [9] is used. Daubechies wavelets are the most popular wavelet. They represent the foundations of wavelet signal processing and are used in numerous applications. These are also called Maxflat wavelets as their frequency responses have maximum flatness at frequencies 0 and π.

The DWT is computed by successive low pass and high pass filtering of the discrete time domain signal as shown in Figure 1. This is called the Mallat algorithm or Mallat-Tree decomposition.

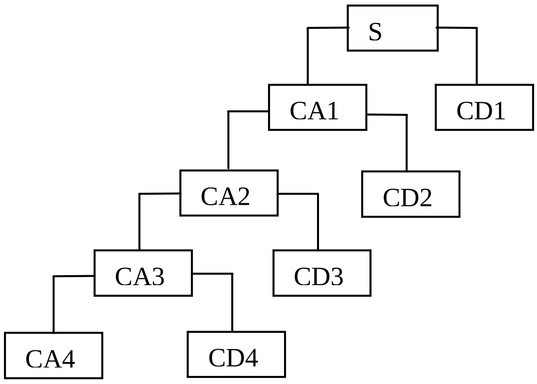

Figure 1: Decomposition Tree

3.2. Gaussian Mixture Models (GMMs)

The GMM method requires determining the parameters of the model based on the training set. The GMM classifier was implemented by first estimating the probability density functions of the features under the possible conditions, based on the training set. A new test set is then classified according to the likelihood ratio that is the ratio of the values of the pdfs of the classes at that point. The pdfs of the data sets were estimated by fitting a General Mixture Model.

The Gaussian means were first initialized by using the k-means clustering and then the model is refined using the Expectation Maximisation (EM) algorithm. For each class, the existence of a probability density functions are assumed as a mixture of a number of multidimensional Gaussian distributions. Equal prior likelihoods were assumed for each class and the decision rule was that points in the feature space for which one pdf was larger, were classified as belonging to that class. The well-known applications of GMM include image segmentation, edge detection, pattern recognition, motion tracking, voice recognition, and so on.

4. Proposed Method

This music genre classification framework is composed of two main parts: feature extraction followed by recognition or classification. Processing steps of the system is depicted in Figure 2 and the details are presented below.

4.1 Feature extraction

Feature extraction is the process of capturing the complex structure in a signal. Therefore, the discrete wavelet coefficients are used as the main feature of music signal. As these wavelet coefficients can be computed using pyramidal approach, feature extraction time is also reduced.

Preprocessing

- To have unique processing, the songs are converted to PCM encoded WAV files at sampling frequency of 22 kHz resolution 16 bits mono.

Framing

- Each song is partitioned into non-overlapping 23 ms (512 samples) frames so that there are 43 frames in every 1s interval of the song.
- For each frame, windowed by hamming

Wavelet Transform Feature Extraction

- The signal is decomposed into low pass (CA) and high pass (CD) for each decomposition step.
- Four level wavelet decomposition steps are performed and approximation coefficients at level 4 are used as features.
- Therefore, on every 1 s interval, a feature matrix of 43x32 wavelet coefficients is firstly extracted.
- Mean of level 4 approximation wavelet coefficients value is computed from every one second interval representing a feature vector of 32-dimension. The input parameters for GMM of a song is then n x 32 where n is the number of 1s segment contained in a song.

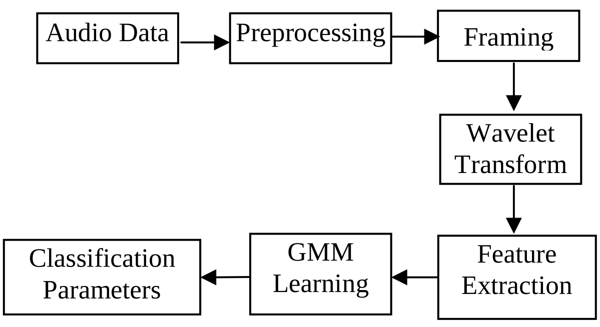

Figure 2: Processing Steps of the System

4.2 Classification

- Learn the classification parameters of GMM through DWT coefficients using EM algorithm.
- Expectation Step: computes an expectation of the likelihood by including the variables.
- Maximization Step: computes the maximum likelihood estimates of the parameters by maximizing the expected likelihood found in the E-step.
- To achieve the best classification accuracy, a multi-layer classifier based on GMM is used.

4.3 Architecture of Classification

Four music genres are considered to classify in this paper. A hierarchical classification architecture is designed so that number of classifier to be used is only three GMMs. To obtain the classification parameters, each GMM is learned through Expectation Maximization algorithm after selected wavelet coefficients have input to it.

In the first layer, music is classified into pop/classic group and rock/jazz group by classifier GMM1. In the second layer, pop/classic music group is further classified into either pop music or classic music using GMM2. Rock/jazz music group is further classified into either rock music or jazz music using GMM3. The diagram of musical genre classification is illustrated in Figure 3.

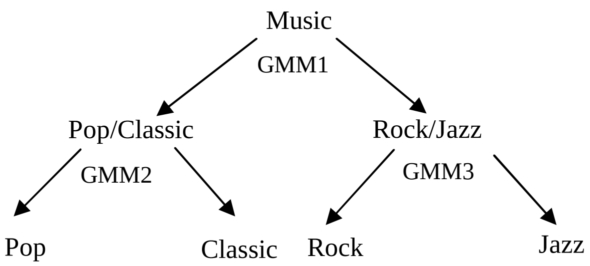

Figure 3: Architecture of the Classification

5. Experiments

To evaluate musical genre classification experiments have been conducted using a music dataset containing 4 genres: pop, classic, rock and jazz. In this dataset 60 songs have been used for training and 120 songs for testing. They are collected from music CDs and Internet. Most the songs collected data items are in .mp3 format with 44 kHz. There is no overlap between training and testing songs.

Table 1 shows the training songs database employed for learning the parameters of GMMs. For each group of genre, 15 songs from various artists have been applied.

Table 1: Training songs of the system

No	Genre	Artist	No of songs
1	Pop	Michael Jackson, Automic kitten, Jennifer Lopez	15
2	Classic	Mozart, Beethoven, Saint-seans Tchaikovsky, Charles-Canille	15
3	Rock	Bonjovi, Linkin Park, Paramore	15
4	Jazz	B.B.King	15

Similarly, in Table 2, the information for test songs has been given. The presented approach has been experimented with 40 songs for each pop and rock genres and 20 music pieces each classical and jazz music genres.

Table 2: Testing songs of the system

No	Genre	Artist	No of songs
1	Pop	Michael Jackson, Celine Dion, Richard Marx, Shakira	40
2	Classic	Beethoven, Johann Sebastian, Frederic Chopin, Suppe	20
3	Rock	Bonjovi, Paramore, Pearl Jam	40
4	Jazz	Callie Cardamon, B.B.king, Trichotomy, Grove Washington	20

In this test data, songs of various artists have been selected. The system performance is measured with a confusion matrix given in Table 3.

According to this table, it is found that jazz music is often categorized as rock music. Some rock songs are misclassified as pop songs and vice versa as the line between these genres is blurred at sometimes. Table 4 summarizes the classification accuracy and error rate for all test genres. A graphical representation of this table is also shown in Figure 4.

Table 3: Confusion Matrix of Classification Results

Type	Pop	Classic	Rock	Jazz
Pop	36	0	3	1
Classic	0	19	0	1
Rock	4	0	36	0
Jazz	2	0	6	12

Table 4 : Classification Accuracy and Error Rate

Type	Accuracy	Error Rate
Pop	90%	10
Classic	95%	5%
Rock	90%	10%
Jazz	60%	40%
Average	86%	14%

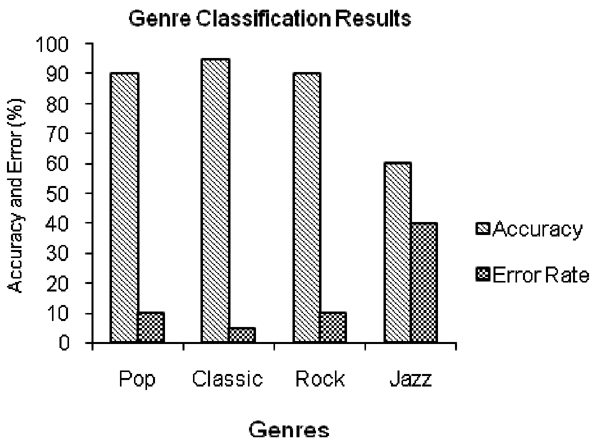

Figure 4: Genres Classification Results

6. Discussions

In early works of music genre classification, various feature vectors are used such as beat spectrum, LPC derived cepstrum, zero crossing rate etc. These feature extraction methods take long time in the training process. In this system, discrete wavelet transform is only used for feature extraction. Although only one feature is extracted to represent the music styles, the accuracy of this method is satisfactory. Furthermore, it is easy to implement and can reduce the feature extraction time. With hierarchical structure of genre classification, modeling time of classification structure is also minimized.

7. Conclusion

An automatic classification approach for musical genres using binary tree Gaussian mixture model learning is presented. Discrete Wavelet Transform is calculated as features to characterize music content. Three Gaussian mixture model classifiers are developed to obtain the optimal class boundaries between pop/classic and rock/jazz, pop and classic, rock and jazz by learning from training data. Experiments show the multi-layer Gaussian mixture model learning method has good performance in musical genre classification. Experimental results report that the overall accuracy of the method is about 95% on the pop, rock and classic songs. Among the test genres, accuracy on jazz music is as low as to 60%.

8. References

[1] L. Chen, E. Bovbel, M. Dashouk "Wavelet-based feature extraction for musical genre classification using SVM," *Proc. Of IASTE07, pp: 260-265, 2007.*

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

[3] J. D. ning, L. Lie, Z.H. Jiang, C.L. Hong, T. J. Hua, "Music Type classification by spectral contrast features", *Proc. Of IEEE International Conference on Multimedia and Expo, August 2002.*

[4] D. Pye "Content-based methods for the management of digital music". *Proceedings of IEEE International Conference on Acoustic, Speech and signal processing 2000, Volume 4, 2000.*

[5] L. R. Pabiner and R. W. Schafer, "Digital Processsing of Speech Signals", Prentice-Hall, Englewood Cliffs, NJ, USA, 1988.

[6] X.Shao, C.Xu and M. Hankarhalli, "Unsupervised Classification of musical genre using HMM" in IEEE International Conference of Multimedia Explore (ICME) , Taibei, Taiwan, Chan ,2004.

[7] G. Tzanetakis, P.Cook, "Musical genre classification of Audio Signals", *IEEE Trans on speech and Audio Processing, vo-10, No.5, July2002.*

[8] G. Tzanetakis, G. Essl, P.Cook, "Automatic Musical Genre Classification of Audio signals" *ISMIR01, 2001.*

[9] G. Xu, N. C. Maddage , X. Shao , F. Cao and Q. Tian, " Musical Genre classification using Support Vector Machines" *IEEE International conference on Acoustic, Speech and Signal Processing2003.*