USING SUPPORT VECTOR MACHINE FOR MUSIC GENRE CLASSIFICATION

Lett Yi Kyaw, Renu
Computer University (Myeik), Myanmar
lettyilettyi@gmail.com, renushi@gmail.com

ABSTRACT

Musical genres are commonly used to structure the increasing amounts of music available in digital form on the Web and are important for music/audio information retrieval. Genre categorization for music/audio has traditionally been performed manually. Automatic music genre classification is very useful for music indexing and retrieval. In this paper, we present an efficient and effective automatic music genre classification approach. Music genre classification is processed in two parts, feature extraction and classification. A set of feature is extracted and used to characterize music content. A multilayer classifier based on support vector machine is applied to music genre classification. Support vector machines are used to obtain the optimal class boundaries between different genres of music by learning from training data .The classification results of the proposed feature set has 93% accuracy rate improvement in the multilayer SVM.

Keywords: Music genre classification, automatic music genre classification approach, Support Vector Machine.

1. INTRODUCTION

Browsing and searching by genre can be very effective tools for users of the rapidly growing networked music archives. The current lack of a generally accepted automatic genre classification system necessitates manual classification, which is both time-consuming and inconsistent. Most existing studies have focused on accomplishing the difficult task of features extraction from music/audio data. The nebulous and changing nature of genre definitions makes the task well suited to machine leaving systems such as Support Vector Machine [2].

The rapid increase in speed and capacity of computers has allowed the inclusion by music/audio as a type of data in many modern computer applications. Numerous music/audio recordings are dealing with in music/audio and multimedia applications. The effectiveness of their deployment in greatly dependent on the ability to classify and retrieve the music/audio files in terms of their sound properties or content. Users accustomed to searching, scanning and retrieving data can be frustrated by the inability to look inside the music/audio objects. Digital music is one of the most important data types distributed by the Internet and the amount of digital music increase rapidly nowadays. How to effectively organized and process such large variety and quantity of musical data to allow efficient indexing, searching and retrieval is a big challenge.

Music 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 music retrieval. To determined music genre of a music track by experienced managers is a labourious and time-consuming work. So, a number of music genre classification systems have been developed to deal with this problem.

Music genre classification, which is defined as the most restrict form (i.e., the computer classifies each music audio signals to one class), can be divided into two stages: feature extraction and classification method (classifier design). Feature extraction is important for music content analysis and the process of converting an audio signal into a sequence of feature vectors carrying characteristics information about the signal. These vectors are used by many different classification algorithms to classify the music/audio signal.

After the features are extracted from each music/audio file, classification method (classifier) is used. In this paper, music is divided into four categories: Country, Jazz, Classic and Pop. We experimented a state-of-the-art machine learning algorithm, Support Vector Machine (SVM), in the design of an automatic genre classifier over music information. A set of music feature is developed to characterize music content of different genres and Support Vector Machine (SVM) is applied to build a multilayer classifier. This multilayer classification method can improve a better classification result than current existing methods.

2. RELATED WORK

George Tzanetakis, et.al, proposed a set of features for representing texture and instrumentation. In addition a novel set of features for representing rhythmic structure and strength is proposed. The performance of those features sets has been evaluated by training statistical pattern recognition classifiers using real world audio collections. Based on the automatic hierarchical genre classification two graphical user interfaces for browsing and interacting with large audio collections have been developed. To evaluate the performance of the proposed feature set, statistical pattern recognition classifiers were trained and evaluated using data sets collected [5].

In this paper [3], an efficient and effective automatic musical genre classification approach was presented. A set of features is extracted and used to characterize music content. Their experimental results of multilayer support vector machines illustrated good performance in musical genre classification and are more advantages than other traditional Euclidean distance based method and other statistic learning methods.

Charles. and et.al, have chosen relevant feat-ures and appropriate classification. Features were extracted via spectral and time domain analysis, and then the LogitBoost algorithm is used to build and effective classifier for the data. Both Logitboost and Adaboost were well suited to this problem. They discussed the final feature set, why they chose those features, their final classification algorithm, and why they chose it [4].

3. FEATURE EXTRACTION

Feature extraction is a major stage in any classifying system in general, and in music/audio signal classification systems in particular. After the segmentation of the music/audio, several representative features are extracted from the music/audio files.

Feature extraction, using the correct feature representation, which usually varies in the several situations. Feature extraction or selection refers to the representation of the music waveform with a suitable set of characteristics, derived from its waveform. This is based on the fact that features of music change slowly within frames of a few milliseconds, in contrast with its waveform. Thus available signal information is compressed and represented in a more effective space. Audio applications usually do not operate on the original data, instead features that represent the content more efficiently, are computed.

3.1. Music Feature Extraction

Music feature extraction involves processing a recording with the aim of generating numerical representations of what are, hopefully traits of the recording that are characteristics of the category or categories that it should be classified as belonging to. These features can be grouped together into feature vectors that serve as the input to the classification systems [1].

3.1.1 Zero Crossing Rate

In the context of discrete-time signals, a zero crossing is said to occur if successive sample have different algebraic sign. The rate at which zero crossing occurs 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 size 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.

$$Z_m = \sum_{n} \left| sign[x(n)] - sign[x(n-1)] \right| w(m-n) \qquad (1)$$

where the sign function is $sign[x(m)] = \begin{cases} 1 & x(m) \geq 0 \\ -1 & x(m) \angle 0 \end{cases}$,

n = time index of the zero crossing rate

ZCR essentially reflects the frequency content of the signal. Higher ZCR implies high frequency and lower ZCR means low frequency. While calculating ZCR, it is necessary to first remove the DC content to get the accurate reflection of frequency content.

3.1.2 Pitch

Pitch is the perceptual counterpart of the physical frequency. It is the perceived frequency of a sound. Pitch cannot be measured physically, since it is an auditory sensation. Two sounds with measurably different frequencies do not need to have two different pitches but difference in the perceived pitch implies different frequencies. Mean and variance of the time dependent pitch are used as features.

3.1.3 Mel Frequency Cepstral Coefficient

Mel Frequency Cepstral Coefficients (MFCC) is introduced in 1980 to the field of speech processing. They have become the most widely used short-time features in speech classification problems like speech recognition and speaker verification as they provide a compact (as a result of DCT) and accurate representation of the speech magnitude spectrum and also because they have a Gaussian distribution, which allow for excellent matching with currently referred back-ends such as Gaussian mixture models and Hidden Markov models.

$$Mel(f) = 2595 \log_{10} \left(1 + \frac{f}{700}\right) \qquad (2)$$

Where, f = physical frequency in Hz

Mel = perceived frequency in mels

The idea of a perceptual frequency scale has led to the investigation of the benefits of using a frequency axis that is warped to correspond to the mel scale. One of the techniques used to obtain the new frequency axis is to use a filter bank. Since the filter bank is applied in the frequency domain, the modified spectrum of the signal thus consists of the output power of these filters when the input is the signal obtained at the previous step. It has been found that the perceived loudness of audio signals to be approximately logarithmic and hence the logarithm of the power spectrum is taken. And then the log mel spectrum is converted back to the time domain and the result is Mel Frequency Cepstral Coefficients. Discrete Cosine Transform is applied in this step [6].

4. CLASSIFICATION

4.1. Manual Music Genre Classification

Musical Genre is widely used to classify and describe titles, both by the music industry and the consumers. There are therefore many different genre taxonomies available. Manual input is clearly not sufficient to describe precisely millions of titles. Manual input is therefore mostly useful as bootstrap to test research ideas, or as a comparison base to evaluate automatic algorithms. The bottom taxons were very difficult to describe objectively. Only the taxonomy designers would be able to distinguish between slightly different taxons. The taxonomy was very sensitive to music evolution. Music evolution has notable and well-known effects in genre taxonomies [7].

4.2. Automatic Musical Genre Classification

There have been numerous attempts at extracting genre information automatically from the audio signal, using signal processing techniques and machine learning schemes. They all proceed in two steps:

- **Frame-based Feature extraction**: the music signal is cut into frames, and a feature vector of low-level descriptors of timbre, rhythm, etc. is computed for each frame.

- **Machine Learning/Classification**: a classi-fication algorithm is then applied on the set of feature vectors to label each frame with its most probable class: its “genre”. The class models used in this phase is trained beforehand, in a supervised way.

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 Country/Jazz and Classic/Pop music according to the features. In the layer Country/Jazz music is further classified into Country and Jazz music according to the featured and Classic/Pop music is further classified into Classic and Pop music according to the features of SVM is used in all layers and each layer has different parameters and support vectors. The system diagram of music genre classification is illustrated in Figure 1.

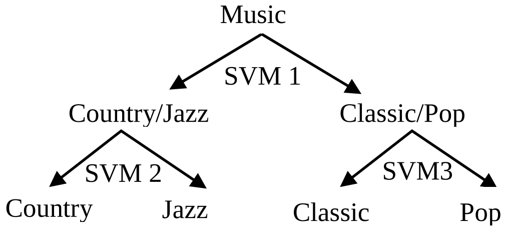


Figure 1. Musical Genre Classification Diagram

4.3. Support Vector Machine

Universal feed-forward networks are known as Support Vector Machines (SVM). SVM can be used for pattern classification and nonlinear regression. Basically, the support vector machine is a linear machine with some very nice properties. To explain how it works, it is perhaps easiest to start with the case of separable patterns that could arise in the context of pattern classifications. In the context, the main idea of a support vector machine is to construct a hyperplane as the decision surface in such a way that the margin of separation between positive and negative examples is maximized [8].

5. PROPOSED SYSTEM DESIGN

In particular, we may use the support vector learning algorithm to construct the following three types of learning machines (among others):

1. Polynomial learning machines
2. Radial-basis function networks
3. Two-layer perceptrons (i.e., with a single hidden layers)

That is, for each of these feed-forward networks we may use the support vector learning algorithm to implement the learning process using a given set of training data, automatically determining the required number of hidden units.

Stated in another way: Whereas the back-propagation algorithm is devised specifically to train a multilayer perceptron, the support vector learning algorithm is of a more generic nature because it is wider applicability. The system overview of support vector machine is described in Figure 2.

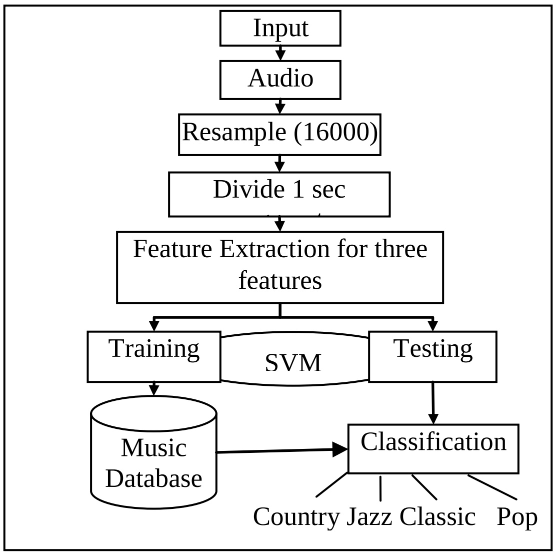


Figure 2. The Proposed System

Figure 3 shows data flow diagram of the proposed system. The steps of the music genre classifications are as follows:

(i) Resampling
(ii) Feature Extraction
(iii) Classification
- Training
- Testing
(iv) Classification stage

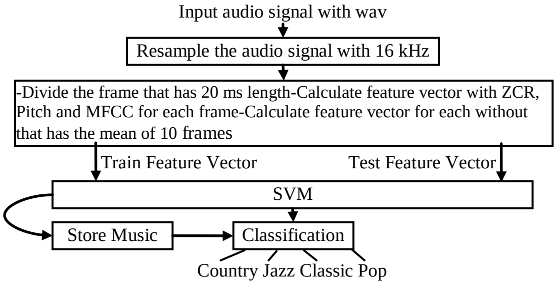

Input audio signal with wav
Resample the audio signal with 16 kHz
-Divide the frame that has 20 ms length-Calculate feature vector with ZCR, Pitch and MFCC for each frame-Calculate feature vector for each without that has the mean of 10 frames
Train Feature Vector
Test Feature Vector
SVM
Store Music
Classification
Country Jazz Classic Pop

Figure 3. Data Flow Diagram of the Proposed System

6. EXPERIMENTAL RESULTS

The MP3 files were used in the experiment are collected from the internet and Yinnmar studio. The MP3 music is selected from various categories (country, jazz, classic and pop). The music files are partitioned into two parts: the training and test. The training music is 20 minutes. We evaluated the system using four labeled data sets, each 5 minutes long.

We select 80 music samples as training data including 20 pop music, 20 jazz music, 20 classic music and 20 pop music. Each sample is segmented into 2000 frames and the length of each frame is 320 sample points. Therefore, the total number of training data is 160,000 frames. For the SVM1 which is used to classify music into country/jazz and classic/pop, 80,000 frames including 20,000 frames of each genre are used for training. For the SVM2 which is used to classify country/jazz into country and jazz, 40,000 frames, are used for training. Among these training frames, 20,000 frames are from SVM1 training set with 10,000 frames of jazz and pop respectively; the other 40,000 frames are from new training frames with 20,000 frames of jazz and pop respectively. For SVM3 which is used for classify country/classic into country and classic, 40,000 frames are used for training. The training frames selected for SVM3 is similar to those for SVM2. 20,000 frames are from SVM1 training set and 40,000 frames from new training frame. The rest 40 samples are used as a test set.

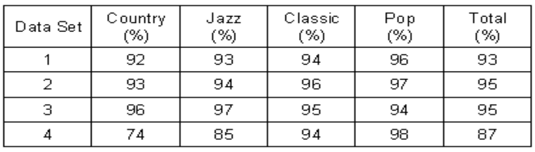
After training the SVMs, we use them as the classifiers to separate country, jazz, classic and pop frames on the test set. The test set contains 10 country music samples (20,000 frames), 10 jazz music samples (20,000 frames), 10 classic music samples (20,000 frames) and 10 pop music samples (20,000 frames). 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	**80,000**	**40,000**	**40,000**
Support Vectors	**5767**	**11103**	**10245**
Test Set	**40,000**	**20,000**	**20,000**
Error Rate	**7.26%**	**8.42%**	**7.79%**

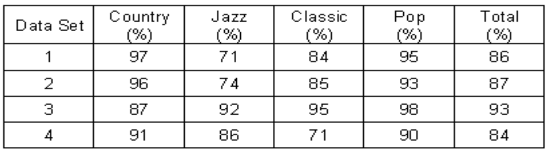
Table 2, 3, 4 and 5 show the results on four data sets by using the training data. When MFCC in conjunction with ZCR features were used, a correct classification rate of 90% was obtained. When the inputs was a combination of MFCC and pitch the classification rate was 91% which is almost the same as when pure MFCC features where used. When the MFCC features were used conjunction with ZCR and Pitch, a correct classification rate of around 93% was obtained. This result was the best result among the classification results.

Table 2. Classification Accuracy Results Using MFCC



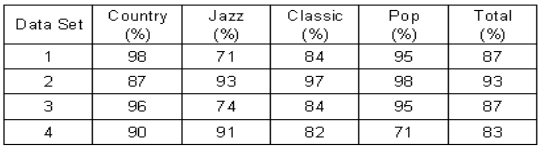
Data Set	Country (%)	Jazz (%)	Classic (%)	Pop (%)	Total (%)
1	92	93	94	96	93
2	93	94	96	97	95
3	96	97	95	94	95
4	74	85	94	98	87

Table 3. Classification Accuracy Results Using MFCC+ZCR



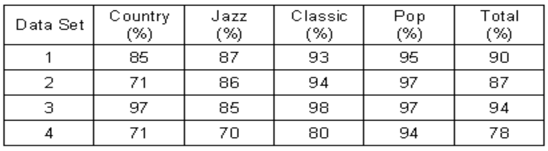
Data Set	Country (%)	Jazz (%)	Classic (%)	Pop (%)	Total (%)
1	97	71	84	95	86
2	96	74	85	93	87
3	87	92	95	98	93
4	91	86	71	90	84

Table 4. Classification Accuracy Results Using MFCC+Pitch



Data Set	Country (%)	Jazz (%)	Classic (%)	Pop (%)	Total (%)
1	98	71	84	95	87
2	87	93	97	98	93
3	96	74	84	95	87
4	90	91	82	71	83

Table 5. Classification Accuracy Results Using MFCC+ZCR and Pitch



Data Set	Country (%)	Jazz (%)	Classic (%)	Pop (%)	Total (%)
1	85	87	93	95	90
2	71	86	94	97	87
3	97	85	98	97	94
4	71	70	80	94	78

7. CONCLUSION AND FUTURE WORK

7.1. Conclusion

We have presented and demonstrated an automatic classification approach for musical genres using multi-layer support vector machine learning. Zero crossing rates, pitch and mel-frequency cepstral coefficients are calculated as features to characterize music content. There support vector machine classifiers are developed to obtain the optimal class boundaries between country, jazz, classic and pop 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 other statistic learning methods.

7.2. Future Work

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 feature that can be used to characterize the music content.

REFERENCES

[1] C. McKay, "*Automatic Genre Classification of MIDI Recordings*", MA-Thesis, Music Technology Area, Department of Theory, Faulty of Music, McGill University, Montreal, June 2004.

[2] C. McKay, "*Using Neural Networks for Musical Genre Classification*", Faulty of Music, McGill University, 555 Sherbrooke Street West, Montreal, Quebec, Canada H3A 1E3.

[3] C. Xu, N.C Maddage, X. Shao, F. Cao and Q. Tian, *"Musical Genre Classification Using Support Vector Machines"* , Laboratories for Information Technology 21 Heng Mui Keng Terrace, Singapore 119613.

[4] C. Tripp, H. Hung and M. Pontikakis, *"Waveform-Based Musical Genre Classification"*.

[5] G. Tzanetakis, G. Essl and P. Cook, *"Automatic Musical Genre Classification of Audio Signals"*,
Computer Science Department, 35 Olden Street Princeton NJ 08544 +1 609 258 5030.

[6] J.C. Wang, J.F Wang, C.B. Lin, K.T. Jian and W.H. Kuok, *"Content-Based Audio Classification Using Support Vector Machine"*, Department of Electronical Engineering, National Cheng Kaung University.

[7] J.J Aucouturier and F., Pachet, *"Representing Music Genre: A State of the Art"*, SONY Computer Science Laboratory, Paris, France.

[8] S. Haykin, "*A Comprehensive*", Neural Network, 2nd Edition, McMaster University Hamilton.