Multiclass Suport Class Support Vector Machine for Music Genre Classification

Nimesh Prabhu
Computer Engineering Department
Goa College of Engineering
Goa, India

Ashvek Asnodkar
Computer Engineering Department
Goa College of Engineering
Goa, India

Rohan Kenkre
Computer Engineering Department
Goa College of Engineering
Goa, India

ABSTRACT
Musical genres are defined as categorical labels that auditors use to characterize pieces of music sample. A musical genre can be characterized by a set of common perceptive parameters. An automatic genre classification would actually be very helpful to replace or complete human genre annotation, which is actually used. SVM have found overwhelming success in the area of pattern recognition. Finally we validate proposed algorithm with experimental results.

Keywords
Support vector machine, kernel function, optimal boundary, music genre classification..

1. INTRODUCTION
Browsing and searching by genre can be very effective tools for users of rapidly growing network music archives. The current lack of generally accepted automatic genre classification system necessitates manual classification, which is both time consuming and inconsistent.

Developments in Internet and broadcast technology enable users to enjoy large amounts of multimedia content. With this rapidly increasing amount of data, users require automatic methods to filter process and store incoming data. A major challenge in this field is the automatic classification of audio. During the last decade, several authors have proposed algorithms to classify incoming audio data based on different algorithms. Most of these proposed systems combine two processing stages. The first stage analyzes the incoming waveform and extracts certain parameters (features) from it. The feature extraction process usually involves a large information reduction. The second stage performs a classification based on the extracted features.

Support vector machine have made a significant impact in the area of pattern recognition. The support vector machine can be trained to discern the different criteria's used to classes, and it can do it so in a generalized manner allowing accurate classification of the inputs which are not used during classify into training.

The purpose of this paper is to do feasibility study of a music genre classification system based on music content using a multi-class support vector machine.

2. FRAMEWORK

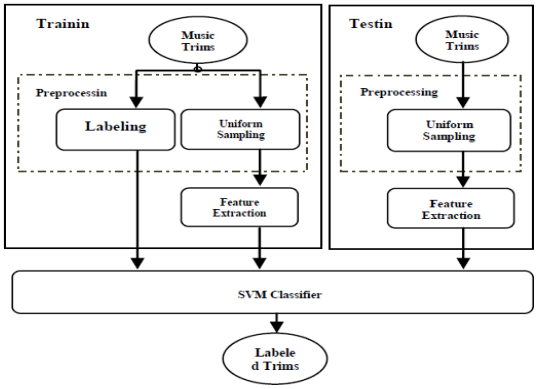


Figure 1:Design of overall Process

Figure 1 describes framework of whole process. First step is process of downloading dataset and installing matlab. Second step is feature extraction process. Third step is training and validation of neural network both standard and improved. Fourth step is testing of neural network.

3. PROPSED METHOLODOGY
Features

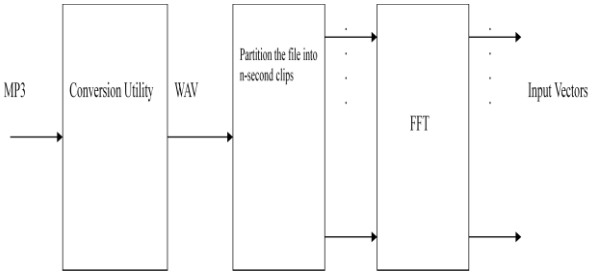


Figure 3: Feature Extraction Process

3.1 Dataset
First we need a dataset of music files to extract features. Marsyas (Music Analysis, Retrieval, and Synthesis for Audio Signals) is an open source software framework for audio processing with specific emphasis on Music Information Retrieval Applications. GTZAN Genre Collection, of 400 audio tracks each 30 seconds long. There are 4 genres represented, each containing 100 tracks. All the tracks are 22055Hz Mono 16-bit audio files in .wav format. We have chosen four of the most distinct genres for our research: classical, jazz, metal, and pop because multiple previous work has indicated that the success rate declines when the number of classifications is more.

3.2 Feature Extraction

A MP3 file is converted into WAV using wav converter software. A 30 seconds audio file stored in WAV format which is passed to a feature extraction process. The WAV format for audio is simply the right and left stereo signal samples. The feature extraction process calculates 16 numerical features that characterize the particular sample. One of the features is MFCC that again gives 12 values. Hence, in total 16 values are used to classify the music genres classification (MGC). Feature extraction process is carried out on many different WAV files to create a matrix of containing column's of feature vectors. Feature extraction matrix is used to train neural network.

3.3 Some Features that will be Extracted

3.3.1 Zero Crossing Rate:

The Zero crossing rates is the rate of sign-changes along a signal, i.e., the rate at which the signal changes from negative to positive or positive to negative. This feature has been used heavily in both speech recognition and music information retrieval, being a key feature to classify percussive sounds.

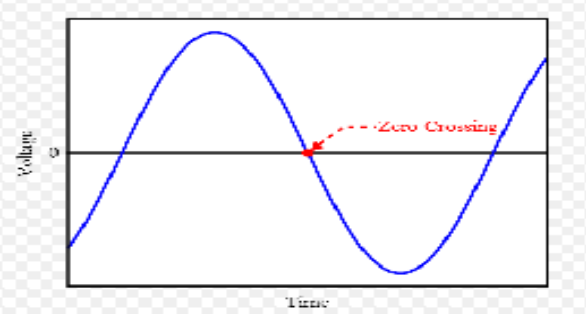


Figure 4: Zero Crossing Rate

$$\mathbf{ZCR} = \frac{\mathbf{1}}{\mathbf{T-1}}\sum_{\mathbf{t=1}}^{\mathbf{T-1}} \mathbf{II}\{\mathbf{s_t s_{t-1}} < 0\}$$

Where S is a signal of length T and the indicator function II{A} is 1 if its argument A is true and 0 otherwise.

3.3.2 Spectral Flux:

Spectral flux is a measure of how quickly the power spectrum of a signal is changing. It is calculated by comparing the power spectrum for the current frame against the power spectrum from the previous frame. It is usually calculated as the 2-norm between the two normalized spectra.

$$\textbf{Spectral flux} = (\boldsymbol{F(n)_t - F(n)_{t-1}})$$

Where $F(n)_t$ and $F(n)_{t-1}$ are normalised magnitudes of Fourier transform at current frame t and previous frame t-1.

3.3.3 Signal energy:

It is total energy of an audio file calculated by following formula:

$$\textbf{Signal Energy} = \sum_{n=1}^{N} |\boldsymbol{x(n)}|^2$$

where x(n) is feature vector

3.3.4 Mel Frequency Cepstral Coefficients:

In music genre classification, the Mel frequency Cepstrum is a representation of the short-term power spectrum of a audio. It is based on a linear cosine transform of a logarithmic power spectrum on a non-linear mel scale of frequency.

Mel-frequency cepstral coefficients (MFCCs) are coefficients that collectively make up an Mel frequency Cepstrum (MFC).

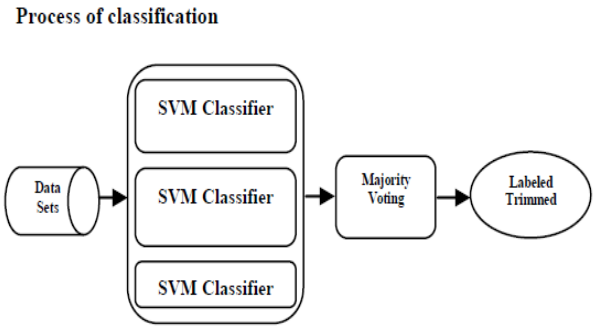
3.3.5 Root Mean Square Level (amplitude):

It is used to calculate root mean square level of amplitude of a audio signal for a continuously varying function or for the series of discrete values.

$$\textbf{RMS} = \sqrt{(\boldsymbol{x_1^2 + x_2^2 + \ldots . x_n^2})/\boldsymbol{n}}$$

Where n = number of samples

3.4 Support Vector Machine Algorithms



These are supervised learning models with associated learning algorithms that analyze data and recognise models used for classification and regression analysis. The basic SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the output. Given the set of training examples each marked as belonging to two categories and SVM training algorithm builds some model that assigns the new example into one category or another.

$$g(x_i) = w^T x_i + b = 0$$
$$y_i = \{1 \mid g(x_i) \geq 1\}$$
$$y_i = \{-1 \mid g(x_i) \leq -1\}$$

We first plotted all the nodes in the hyperplane, all the nodes are not linearly separable hence we use kernel function to separate the nodes. Once we apply kernel function, all the nodes are separable in the hyperplane. Then we find an optimal boundary to classify the genres.

The optimal boundary is also known as Margin & is given by:

$$m = \frac{2}{\sqrt{w \bullet w}} = \frac{2}{|w|}$$

In order to compare several MKL algorithms, we perform 10 different experiments on four data sets that are composed of different feature representations. We use both the linear kernel and the

Gaussian kernel in our experiments; we will give our results with the linear kernel first and then compare them with the results of the Gaussian kernel. The kernel matrices are normalized to unit diagonal before training.

4. EXPERIMENTAL RESULTS

The accuracy was calculated for the music genre dataset. The highest accuracy recorded was 63%.

Table 2: Confusion matrix for MGC

	Pop	Jazz	classical	metal
Pop	**24**	**1**	**0**	**0**
Jazz	**10**	**10**	**0**	**0**
Classical	**1**	**6**	**18**	**0**
Metal:	**0**	**14**	**0**	**11**

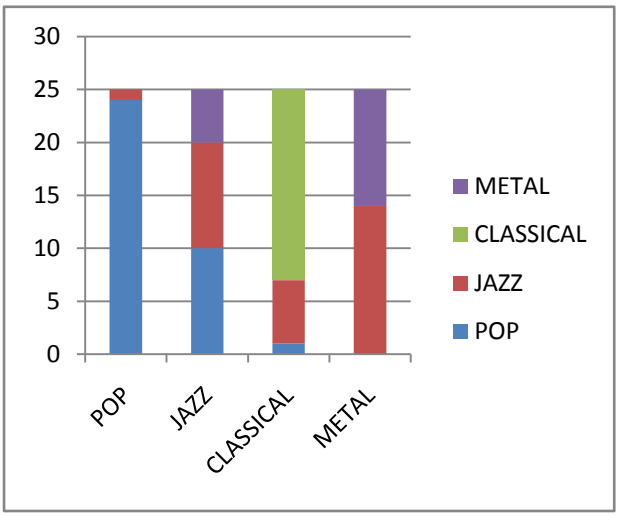


Figure 5: Graphical representation of Confusion matrix

5. CONCLUSION AND FUTURE RESEARCH

From the above result we can see that Jazz and Metal are not classified accurately due to overlapping features in them. Though good results are obtained for our GTZAN datasets, we can try it for more datasets.

There is a significant amount of work on multiple kernel learning methods. This is because in many applications, one can come up with many possible kernel functions and instead of choosing one among them, we are interested in an algorithm that can automatically determine which ones are useful, which ones are not and therefore can be pruned, and combine the useful ones. Or, in some applications, we may have different sources of information coming from different modalities or corresponding to results from different experimental methodologies and each has its own (possibly multiple) kernel(s). In such a case, a good procedure for kernel combination implies a good combination of inputs from those multiple sources.

6. REFERENCES

[1] G. Tzanetakis and P. Cook, “Musical Genre Classification of Audio Signals” In IEEE Trans. Acoust. Speech, Signal Processing , vol.10, ,N°5, July 2002.

[2] G. Tzanetakis and P. Cook, “Audio analysis using the discrete wavelet transform” in Proc. Conf. Acoustics and Music Theory Applications, Sept.2001.

[3] T. Heitolla, "Automatic Classification of music signals ", Master of Science Thesis, February 2003.

[4] R. Duda, P. Hart and D. Stork, “Pattern Classification” , John Wiley & Son, New York, 2000.