Automatic Music Genre Classification for Indian Music

S. Jothilakshmi [1+], N. Kathiresan [2]

[1] Department of Computer Science and Engineering, Annamalai University, Annamalai nagar, India

[2] Department of Information Technology, Bharathiyar College of Engineering and Technology, Karaikal, India

Abstract. Automatic genre classification from audio has been an area of active research due to its importance in music information retrieval systems. India is a multilingual and multi cultural country. The language and culture differ based on the geographical area, hence the music genre also varying based on the geographical area. So automating human capabilities for classifying different Indian music genre of songs is a difficult task. In this work, we explore the performance of various features extracted from the audio signal in terms of the separability of the five classes of Indian music using Gaussian mixture model (GMM) and k-nearest neighbour (kNN) classifier. Spectral shape features and perceptual features are used for the experiments. For this work we have selected more popular music genres of Indian music namely Hindustani, Carnatic, Ghazal, Folk and Indian western. The obtained results have clearly shown that GMM is best to classify the Indian music genres when compared to kNN. Moreover it shows that MFCC, Spectral centroid, Skewness, Kurtosis, Flatness, Entropy, Irregularity feature combination outperforms all other combinations with 91.25% of classification accuracy.

Keywords: Indian music, Music genres, Gaussian mixture model, k-nearest neighbour, Spectral shape features and perceptual features

1. Introduction

Music is one of the most popular types of online information and there are now hundreds of music streaming and download services operating on the World Wide Web. Some of the music collections available are approaching the scale of ten million tracks and this has posed a major challenge for searching, retrieving and organizing music content [1]. The vast quantity of music files on the web is making the manual classification of music libraries more and more infeasible. Hence automatic classification of music into different categories is an important task for retrieval and organization of music libraries [2].

Music genre is defined as classes or groups of songs that categorizes a collection of songs that have similar characteristics. It is a label created by music experts, so that these songs are easily described and recognized [3]. Music genres are categorical descriptions that are used to characterize music in music stores, radio stations and on the internet [4]. Automating human capabilities for classifying different genre of songs is a difficult task. Analyzing music contents is one of many ways to identify and group similar songs together. Automatic genre classification from audio has been an area of active research due to its importance in music information retrieval systems [5].

Most of the music genre classification techniques employ pattern recognition algorithms to classify feature vectors extracted from short time recording segments into genres. In general the features employed for music genre classification are related to the main dimensions of music including melody, harmony, rhythm, timbre and spatial location [6]. In the literature various supervised and unsupervised approaches are proposed for music genre classification. Most commonly used classifiers for supervised classification are

---

[+] Corresponding author. Tel.: + 91-9894693493; Fax: +91-04144-238080.

E-mail address: (jothi.sekar@gmail.com).

support vector machine, **k**-nearest neighbours, Gaussian mixture model, hidden Markov model, linear discriminant analysis and artificial neural networks [7], [8]. Unsupervised algorithms take the advantage of the similarity measure to organize the music collection with clusters of similar titles.

India is made up of several dozen ethnic groups, speaking their own languages and dialects, having distinct cultural traditions as well as music genres. The Indian music genres are based on geographical area of the country. The Indian music has multiple genres like classical, light classical, folk, hip hop, Indian western, and many others. For this work we have selected more popular music genres of Indian music namely Hindustani, Carnatic, Ghazal, Folk and Indian western. The two main traditions of classical music are Hindustani music and Carnatic music. Hindustani music, found in the northern and central regions and Carnatic music found predominantly in the peninsular regions. Ghazal is a light classical or semi-classical music. There are many subgenres under Indian folk music. Most of the Indian film music is based on Indian western genre [9].

In this work, we explore the performance of various features extracted from the audio signal in terms of the separability of the five classes of Indian music using Gaussian mixture model and k-nearest neighbour classifier. We organize the remainder of this paper as follows: Section 2 describes the commonly used features for music genre classification. Section 3 discusses the basic concepts of Gaussian mixture model and **k**-nearest neighbour classifier. The database used for the experiments is presented in Section 4. Section 5 then will be discussing the experiments and the classification results; and we conclude in Section 6 with conclusions and future directions in the field.

2. Feature Description

Extracting features is the first step of most pattern recognition systems. Indeed, once significant features are extracted, any classification scheme may be used. In the case of audio signals, features may be related to the main dimensions of music including melody, harmony, rhythm, timbre and spatial location [6], [10], [11]. We summarize here the main features used in genre characterization applications:

- **Temporal features**: features computed from the audio signal frame (zero-crossing rate, linear prediction coefficients, etc.).
- **Energy features:** features referring to the energy content of the signal (Root Mean Square energy of the signal frame, energy of the harmonic component of the power spectrum, energy of the noisy part of the power spectrum, etc.).
- **Spectral shape features:** features describing the shape of the power spectrum of a signal frame: centroid, spread, skewness, kurtosis, slope, roll-off frequency, variation, Mel-Frequency Cepstral Coefficients (MFCCs).
- **Perceptual features:** features computed using a model of the human earring process (relative specific loudness, sharpness, spread).

These features can be categorized into short time features and long time features. The short time features are computed at regular time intervals, over short windows of typical length between 10 and 60ms. These are mainly based on spectrum-derived quantity within a short segment. In contrast, the long time features are computed over larger windows commonly called as texture windows. The typical length of texture window is 1 second. These features characterize the variation of spectral shape or beat information over a long segment [6], [10].

3. Model Description

The major interest of supervised classification is that one does not need to explicitly describe musical genres: the classifier attempts to form automatically relationships between the features of the training set and the related categories. For our work, Gaussian mixture model and *k*-nearest neighbour classifier are used and they are described as follows:

k-Nearest Neighbours (*k*NN): It is a non-parametric classifier based on the idea that a small number of neighbours influence the decision on a point. More precisely, for a given feature vector in the target set, the *k*

closest vectors in the training set are selected (according to some distance measures) and the target feature vector is assigned the label of the most represented class in the k neighbour (there is actually no other training than storing the features of the training set). kNNs are evaluated in the context of genre classification in [6].

Gaussian Mixture Models (GMM): GMMs model the distribution of feature vectors. For each class, we assume the existence of a probability density function expressible as a mixture of a number of multidimensional Gaussian distributions. The iterative Expectation Maximization (EM) algorithm is usually used to estimate the parameters for each Gaussian component and the mixture weights. GMMs have been widely used as classifiers by using a maximum likelihood criterion to find the model best suited to a particular song. They have been used to directly model musical genres [6].

4. Database

The dataset consists of following five genres: Hindustani, Carnatic, Ghazal, Folk and Indian western. Each genre class contains 100 audio recordings of 30 seconds long. The selected 30-second excerpt was obtained from approximately starting 30 seconds after the beginning of the piece because which may not be representative of the whole piece. A total of 500 30-second duration excerpts equally distributed between five genres were extracted from the commercially available CDs. The selected audio clips were filtered to limit the bandwidth to 11.025 kHz.

5. Experiments and Results

The selected genres are highly similar with each other; hence the features are selected from spectral shape features and perceptual features. These features are based on spectral components and human perception model of the music. The extracted music features are MFCC, spectral centroid, skewness, kurtosis, flatness, entropy, spectral roll-off, irregularity, spread and brightness. These low level features are extracted from each analysis window of duration 20ms with 50% overlap between windows. With 11 dimensional MFCC, the dimensionality of the complete feature vector is 20. Classification is performed by kNN rule and GMM classifier.

For each track, the distribution of feature vectors is estimated with a GMM with 1, 3, 8 and 10 Gaussians. Initially performance of individual group of features was studied with 1, 3, 8, and 10 Gaussians and the results are shown in Table 1. The results show that while increasing the number of Gaussians the performance also increased, but after eight numbers of Gaussians the performance is decreased. Hence for the rest of the experiments we used only eight numbers of Gaussians.

Table 1: Performance of Individual Group of Features using GMM (in %)

Number of Gaussians	1	3	8	10
MFCC	28.13	31.25	37.50	37.50
Spectral centroid	25.00	18.75	25.00	21.88
Spectral entropy	34.38	25.00	25.00	12.50
Spectral irregularity	25.00	25.00	25.00	18.35
Brightness group (brightness, roll off and spread)	28.13	28.13	28.13	28.13
Skewness group (skewness, kurtosis, and flatness)	15.63	21.88	25.00	25.00

Likewise we conducted experiments to tune parameters of k nearest neighbour classifier by selecting the number of neighbours (k) as 3, 5, and 10 with individual group of features and the results are shown in Table 2. From the experimental results we observed that the performance with 3 neighbours is better than others, hence it is used for the rest of the experiments.

Table 2: Performance of Individual Group of Features using k Nearest Neighbours (in %)

Number of neighbours	3	5	10
MFCC	28.12	37.50	31.25
Spectral centroid	18.75	18.75	21.88
Spectral entropy	43.75	34.38	31.25
Spectral irregularity	28.13	25.00	25.00
Brightness group (brightness, roll off and spread)	34.38	32.38	31.25
Skewness group (skewness, kurtosis, and flatness)	28.13	21.88	12.50

After selecting the parameters for the classifiers, we conducted experiments to study the performance for various combinations of features. The selected basic feature to form the combination of features is MFCC. The MFCC vectors are commonly used in audio analysis and are described as timbral features because they model the short-time spectral characteristics of the signal onto a psychoacoustic frequency scale. With this features one or more spectral shape features and/or perceptual features are combined to analyze the best feature combination to identify the genre of Indian music. The classification accuracy of various feature combination is shown in Table 3. The performance is drastically increased while combining entropy with MFCC feature. But by adding the features like spread and brightness with other features the performance is poor. We also observed that the performance of kNN is better than GMM for feature combinations given in rows 2, 4 and 5 in Table 3. But it is not increasing beyond 65% for any combination of features. The best performance obtained from GMM is 91.25% which is better than the overall performance of kNN.

Table 3: Performance of Combined Features (in %)

Feature Combinations	Classification Accuracy (%)	
	kNN	GMM
MFCC, Entropy	47.50	53.75
MFCC, Spectral centroid	41.25	38.125
MFCC, Spectral centroid, Skewness	44.38	44.38
MFCC, Spectral centroid, Skewness, Kurtosis	51.25	48.13
MFCC, Spectral centroid, Skewness, Kurtosis, Flatness	55.00	51.25
MFCC, Spectral centroid, Skewness, Kurtosis, Flatness, Entropy	61.25	83.75
MFCC, Spectral centroid, Skewness, Kurtosis, Flatness, Entropy, Irregularity	64.38	91.25
MFCC, Spectral centroid, Skewness, Kurtosis, Flatness, Entropy, Irregularity, Brightness	64.38	86.88
MFCC, Spectral centroid, Skewness, Kurtosis, Flatness, Entropy, Irregularity, Rolloff	64.38	87.88
MFCC, Spectral centroid, Skewness, Kurtosis, Flatness, Entropy, Irregularity, Rolloff, Spread	61.25	80.63

6. Conclusion and Future Work

India is a multilingual and multi cultural country. The language and culture differ based on the geographical area, hence the music genre also varying based on the geographical area. In this paper, the problem of automatic music genre classification of popular and highly similar music genres of Indian music is examined. The experiments are conducted by selecting the features and classifiers commonly used by

other music genre classification systems. The obtained results have clearly shown that GMM is best to classify the Indian music genres when compared to kNN. Moreover it shows that MFCC, spectral centroid, skewness, kurtosis, flatness, entropy, and irregularity feature combination outperforms all other combinations with 91.25% of classification accuracy.

Indian music is having more number of sub genres under each category; especially under light classical and folk. Future work will include an investigation by adding more number of sub genres in all categories. Only analysis window features are considered for the experiments, the performance of the system should be evaluated by including texture window features also.

References

[1]   Michael A. Casey, Remco Veltkamp, Masataka Goto, Marc Leman, Christophe Rhodes, and Malcolm Slaney, "Content-Based Music Information Retrieval: Current Directions and Future Challenges", *Proceedings of the IEEE*, Vol. 96, No.4, April 2008, pp. 668-696.

[2]   T. Lidy, R. Mayer, A. Rauber, P. J. Ponce de Leon, A. Pertusa, and J. M. Inesta, "A Cartesian Ensemble of Feature Subspace Classifiers for Music Categorization", *11th International Society for Music Information Retrieval Conference (ISMIR 2010)*, 2010, pp. 279-284.

[3]   Noor Azilah Draman, Campbell Wilson, and Sea Ling, "Modified AIS based Classifier for Music Genre Classification", *11th International Society for Music Information Retrieval Conference (ISMIR 2010)*, 2010, pp. 369-374.

[4]   George Tzanetakis, Georg Essl, and Perry Cook, "Automatic Musical Genre Classification of Audio Signals", *IEEE Transactions on Speech and Audio Processing*, 10(5), July 2002, pp. 293-302.

[5]   Sujeet Kini, Sankalp Gulati, and Preeti Rao, "Automatic Genre Classification of North Indian Devotional Music", *National Conference on Communications*, 2011, pp. 16-20.

[6]   N. Scaringella, G. Zoia, and D. Mlynek, "Automatic Genre Classification of Music Content: A Survey", *IEEE Signal Processing Magazine*, 2006, 23(2), pp. 133–141.

[7]   Ioannis Panagakis, Emmanouil Benetos, and Constantine Kotropoulos, "Music Genre Classification: A Multilinear Approach", *9th International Society for Music Information Retrieval Conference (ISMIR 2008)*, 2008, pp 583-588.

[8]   C. Marques, I. R. Guilherme, R. Y. M. Nakamura, J. P. Papa, "New Trends in Musical Genre Classification using Optimum-Path Forest, "*12th International Society for Music Information Retrieval Conference (ISMIR 2011)*, 2011, pp. 699-704.

[9]   *http://en.wikipedia.org/wiki/Music_of_India*.

[10] Kaichun K. Chang, Jyh-Shing Roger Jang, and Costas S. Iliopoulos, "Music Genre Classification via Compressive Sampling", *11th International Society for Music Information Retrieval Conference (ISMIR 2010)*, 2010, pp. 387-392.

[11] Thibault Langlois, and Gonc¸alo Marques, "A Music Classification Method Based on Timbral Features", *10th International Society for Music Information Retrieval Conference (ISMIR 2009)*, 2009, pp. 81-86.