Music Genre Classification using GMM

R. Thiruvengatanadhan[1]

[1]Assistant Professor/Lecturer (on Deputation), Department of Computer Science and Engineering, Annamalai University, Annamalainagar, Tamil Nadu, India

---***---

Abstract - Automatic music genre classification is very useful in music indexing. Tempogram is one of the feature extraction method uses in classification of musical genre that is based temporal structure of music signals. Searching and organizing are the main characteristics of the music genre classification system these days. This paper describes a new technique that uses support vector machines to classify songs. Gaussian mixture model classify music audio into their respective classes by learning from training data. The proposed feature extraction and classification models results in better accuracy in music genre classification.

Key Words: Feature Extraction, Tempogram and Gaussian mixture model (GMM).

1. INTRODUCTION

Musical genres have no strict definitions and boundaries as they arise through a complex interaction between the public, marketing, historical, and cultural factors. This observation has led some researchers to suggest the definition of a new genre classification scheme purely for the purposes of music information retrieval [1]. In addition to this, the advancement in digital signal processing and data mining techniques has led to intensive study on music signal analysis like , content-based music retrieval, music genre classification, duet analysis, Musical transcription , Musical Information retrieval and musical instrument detection and classification. Musical Instrument detection techniques have many potential applications such as detecting and analyzing solo passages, audio and video retrieval, music transcription, playlist generation, acoustic environment classification, video scene analysis and annotation etc.

Automatically extracting music information is gaining importance as a way to structure and organize the increasingly large numbers of music files available digitally on the Web. It is very likely that in the near future all recorded music in human history will be available on the Web. Automatic music analysis will be one of the services that music content distribution vendors will use to attract customers. Due to improvements in internet services and network bandwidth there is also an increase in number of people involving with the audio libraries. But with large music database the warehouses require an exhausting and time consuming work, particularly when categorizing audio genre manually. Music has also been divided into Genres and sub genres not only on the basis on music but also on the lyrics as well [2]. This makes classification harder. To make things more complicate the definition of music genre may have very well changed over time [3]. For instance, rock songs that were made fifty years ago are different from the rock songs we have today.

2. ACOUSTIC FEATURES FOR AUDIO CLASSIFICATION

An important objective of extracting the features is to compress the speech signal to a vector that is representative of the meaningful information it is trying to characterize. In these works, music features namely Tempogram features are extracted.

2.1 Tempogram

An element which gives shape to the music in temporal dimension is the rhythm. Rhythmic feature arranges sounds and silences in time. A predominated pulse called beat which serves as basis for temporal structure of music is induced [242]. Tempogram captures the local tempo and beat characteristics of music signals. The Fourier tempograms are used in the research work.

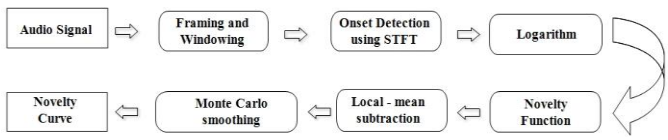

Fig -1: Novelty Curve Computations.

Human perceives rhythm as a regular pattern of pulses as a result of moments of musical stress. Abrupt changes in loudness, timbre and harmonic causes the occurrences of musical accents [4]. In instruments like piano, percussion instruments and guitar, occurs a sudden change in signal energy accompanied by very sharp attacks. A novelty curve is based on this observation and is computed for extracting meaningful information regarding note onset e.g. pieces of songs which are dominated by instruments [5]. In the pre-processing, stage short segmented frames have been extracted and windowed. The novelty curve computed, as described above, indicates peaks which represent note onset values [6]. A hamming window function is applied to avoid boundary problems as smoothing [7]. Novelty curve computation is shown in Fig. 1. The Fourier tempogram is calculated. Tempo related to musical context is a measure of beats per minute. Finally, the histogram is computed for each frame resulting in 12 dimensional feature vectors.

3. CLASSIFICATION MODEL

3.1 Gaussian Mixture Models

Parametric or non-parametric methods are used to model the distribution of feature vectors. Parametric models are based on the shape of probability density function [8]. In non-parametric modeling only minimal or no assumption regarding the probability density function of feature vector is made [9]. The Gaussian mixture model (GMM) is used in classifying different audio classes. The Gaussian classifier is an example of a parametric classifier. It is an intuitive approach when the model consists of several Gaussian components, which can be seen to model acoustic features. In classification, each class is represented by a GMM and refers to its model. Once the GMM is trained, it can be used to predict which class a new sample probably belongs to [10].

The probability distribution of feature vectors is modeled by parametric or non-parametric methods. Models which assume the shape of probability density function are termed parametric. In non-parametric modeling, minimal or no assumptions are made regarding the probability distribution of feature vectors. The potential of Gaussian mixture models to represent an underlying set of acoustic classes by individual Gaussian components, in which the spectral shape of the acoustic class is parameterized by the mean vector and the covariance matrix, is significant.

Also, these models have the ability to form a smooth approximation to the arbitrarily-shaped observation densities in the absence of other information [11]. With Gaussian mixture models, each sound is modeled as a mixture of several Gaussian clusters in the feature space. The basis for using GMM is that the distribution of feature vectors extracted from a class can be modeled by a mixture of Gaussian densities.

The motivation for using Gaussian densities as the representation of audio features is the potential of GMMs to represent an underlying set of acoustic classes by individual Gaussian components in which the spectral shape of the acoustic class is parameterized by the mean vector and the covariance matrix[12]. Also, GMMs have the ability to form a smooth approximation to the arbitrarily shaped observation densities in the absence of other information. With GMMs, each sound is modeled as a mixture of several Gaussian clusters in the feature space [13].

A variety of approaches to the problem of mixture decomposition have been proposed, many of which focus on maximum likelihood methods such as expectation maximization (EM) or maximum a posteriori estimation (MAP). Generally these methods consider separately the question of parameter estimation and system identification, that is to say a distinction is made between the determination of the number and functional form of components within a mixture and the estimation of the corresponding parameter values.

4. IMPLEMENTATION

4.1 Dataset Collection

The music data is collected from music channels using a TV tuner card. A total dataset of 100 different songs is recorded, which is sampled at 22 kHz and encoded by 16-bit. In order to make training results statistically significant, training data should be sufficient and cover various genres of music.

4.2 Feature Extraction

In this work fixed length frames with duration of 20 ms and 50 percentages overlap (i.e., 10 ms) are used. An input wav file is given to the feature extraction techniques. Tempogram 12 dimensional feature values will be calculated for the given wav file. The above process is continued for 100 number of wav files.

4.3 Classification

When the feature extraction process is done the audio should be classified genre music. The extracted feature vector is used to classify whether the audio is speech or music. A mean vector is calculated for the whole audio and it is compared either to results from training data or to predefined thresholds. We select 75 music samples as training data including 25 classic music, 25 pop music and 25 rock music. The rest 25 samples are used as a test set.

Gaussian mixtures for the three classes are modeled for the features extracted. For classification the feature vectors are extracted and each of the feature vectors is given as input to the GMM model. The distribution of the acoustic features is captured using GMM. We have chosen a mixture of 2, 5, 10 mixture models. The class to which the audio sample belongs is decided based on the highest output.

The performance of the system for 2, 5 and 10 Gaussian mixtures is shown in Table.1. The distribution of the acoustic features is captured using GMM. The class to which the speech and music sample belongs is decided based on the highest output. Table.1 shows the performance of GMM for speech and music classification based on the number of mixtures.

Table -1: Performance of GMM for different mixtures.

GMM	2	5	10
Classic	94%	93%	94%
Pop	89%	87%	87%
Rock	90%	91%	93%

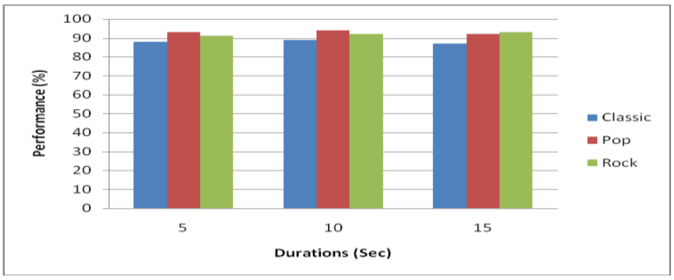

Chart -1: Performance of audio classification for different duration of speech and music clips

Audio classification using GMM gives an accuracy of 94.9%. The performance of GMM for different duration as shown in Chart 1 shows that when the mixtures were increased from 5 to 10 there was no considerable increase in the performance. With GMM, the best performance was achieved with 10 Gaussian mixtures.

5. CONCLUSION

In this paper, we have proposed an automatic music genre classification system using GMM. Tempogram is calculated as features to characterize audio content. GMM learning algorithm has been used for the classification of genre classes of music by learning from training data. The proposed classification method is implemented using EM algorithm approach to fit the GMM parameters for classification between classic, pop and rock by learning from training data. Experimental results show that the proposed audio GMM method has good performance in musical genre classification scheme is very effective and the accuracy rate is 94%.

REFERENCES

[1] F. Pachet and D. Cazaly, “A classification of musical genre,”inProc.RIAO Content-Based Multimedia Information Access Conf., Paris,France, Mar. 2000.

[2] Serwach, M., & Stasiak, B., GA-based parameterization and feature selection for automatic music genre recognition. In Proceedings of 2016 17th International Conference Computational Problems of Electrical Engineering, CPEE 2016.

[3] Dijk, L. Van., Radboud Universiteit Nijmegen Bachelorthesis Information Science Finding musical genre similarity using machine learning techniques, 1–25, 2014.

[4] Mi Tian, Gy ¨ orgy Fazekas, Dawn A. A. Black, Mark Sandler, On the use of the tempogram to describe audio content and its Application to music structural segmentation, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2015.

[5] Venkatesh Kulkarni, Towards Automatic Audio Segmentation of Indian Carnatic Music, Master Thesis, Friedrich Alexander University, 2014.

[6] Eronen, A. and Klapuri, A., “Music Tempo Estimation with k-NN regression,” IEEE Transactions on Audio, Speech and Language Processing, vol. 18, no. 1, pp. 50-57, 2010.

[7] Gui, Wenming & Sun, Yao & Tao, Yuting & Li, Yanping & Meng, Lun & Zhang, Jinglan., A Novel Tempogram Generating Algorithm Based on Matching Pursuit. Applied Sciences, 2018.

[8] Tang, H., Chu, S. M., Hasegawa-Johnson, M. and Huang, T. S., “Partially Supervised Speaker Clustering”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. 5, pp. 959-971, 2012.

[9] Chunhui Wang, Qianqian Zhu, Zhenyu Shan, Yingjie Xia and Yuncai Liu, “Fusing Heterogeneous Traffic Data by Kalman Filters and Gaussian Mixture Models,” IEEE International Conference on Intelligent Transportation Systems, pp. 276-281, 2014.

[10] Sourabh Ravindran, Kristopher Schlemmer, and David V. Anderson, “A physiologi-cally inspired method for audio classification,” Journal on Applied Signal Processing, vol. 9, pp. 1374–1381, 2005.

[11] Menaka Rajapakse and Lonce Wyse, “Generic audio classification using a hybrid model based on GMMs and HMMs,” in IEEE Int’l Conf. Multimedia Modeling, February 2005, pp. 1550–1555.

[12] Poonam Sharma and Anjali Garg. Feature Extraction and Recognition of Hindi Spoken Words using Neural Networks. International Journal of Computer Applications 142(7):12-17, May 2016.

[13] Sujay G Kakodkar and Samarth Borkar. Speech Emotion Recognition of Sanskrit Language using Machine Learning. International Journal of Computer Applications 179(51):23-28, June 2018