International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 01 | Jan 2019 www.irjet.net p-ISSN: 2395-0072

Music Classification using Spectral Features and SVM

R. Thiruvengatanadhan

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

---

Abstract *- The huge growth of the digital music databases people begin to realize the importance of effectively managing music databases relying on music content analysis. Music classification serves as the fundamental step towards the rapid growth and useful in music indexing. Searching and organizing are the main characteristics of the music classification system these days. This paper describes a technique that uses support vector machines (SVM) to classify songs based on features using Spectral Features. Experimental results of multi-layer support vector machines shows good performance in musical classification and are more advantageous than traditional Euclidean distance based method and other statistic learning methods*

Key Words: *Feature Extraction, Zero Crossing Rate (ZCR), Short Time Energy (STE), Spectral Centroid, Spectral Flux and support vector machines (SVM).*

1. INTRODUCTION

Automatic music classification is a fundamental problem for music indexing, content-based music retrieval, music recommendation and online music distribution. Various large-scale datasets of Gigabytes of music information along with metadata and online music streaming services are available. Due to enormity of these datasets, scalable machine learning models that can categorize music information by different criteria such as artist, genre and music similarity are required. Numerous methods have been developed over the years to efficiently classify music information, but the hurdles remain [1].

Advanced music databases are continuously achieving reputation in relations to specialized archives and private sound collections. 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. Luckily, the progress in music data and music recovery has considerable growth in past years.

2. ACOUSTIC FEATURES FOR AUDIO CLASSIFICATION

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

2.1 Spectral Features

2.1.1 Zero Crossing Rate

The Zero Crossing Rate (ZCR) is a simple measure of the frequency content of a signal. For narrow band signals, the frequency content of the signal can be estimated using average ZCR. However, a broad band signal such as speech, it is much less accurate. The spectral properties can be roughly estimated using short time average zero crossing rate [4]. Each pair of samples is checked to determine where zero crossings occur and then the average is computed over N consecutive samples.

2.1.2 Short Time Energy

Short Time Energy (STE) is used in different audio classification problems. STE provides a basis for distinguishing voiced speech segments from unvoiced ones in speech signal. STE is a useful feature in distinguishing high quality speech from silence [5].

2.1.3 Spectral Centroid

A significant measure called spectral centroid is used in digital signal processing to characterize a spectrum. It indicates the “center of mass” of the spectrum and is perceptually connected with the brightness of sound. It is computed as the weighted mean of the frequencies present in the signal, which is calculated using a Fourier transform.

2.1.4 Spectral Flux

The average variation in value of spectrum between two adjacent frames in a given audio clip is called Spectral Flux (SF). Normally, speech signal consists of alternating voiced and unvoiced sounds in the syllable rate whereas this structure does not exist in music signals. Environmental sounds have the highest variation of spectrum flux than that of a speech and music [6]. Hence, SF is the significant acoustic feature for distinguishing environmental sounds

which exhibits strong periodicity. It also discriminates music, speech and environmental sounds effectively.

3. CLASSIFICATION MODEL

3.1 Support Vector Machine

A machine learning technique which is based on the principle of structure risk minimization is support vector machines. It has numerous applications in the area of pattern recognition [7]. SVM constructs linear model based upon support vectors in order to estimate decision function. If the training data are linearly separable, then SVM finds the optimal hyper plane that separates the data without error [8].

Fig. 2 shows an example of a non-linear mapping of SVM to construct an optimal hyper plane of separation. SVM maps the input patterns through a non-linear mapping into higher dimension feature space. For linearly separable data, a linear SVM is used to classify the data sets [9]. The patterns lying on the margins which are maximized are the support vectors.

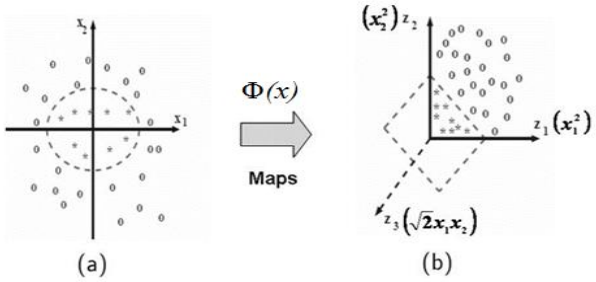

Fig -2: Example for SVM Kernel Function Φ(x) Maps 2-Dimensional Input Space to Higher 3-Dimensional Feature Space. (a) Nonlinear Problem. (b) Linear Problem.

The support vectors are the (transformed) training patterns and are equally close to hyperplane of separation. The support vectors are the training samples that define the optimal hyperplane and are the most difficult patterns to classify [10]. Informally speaking, they are the patterns most informative of the classification task. The kernel function generates the inner products to construct machines with different types of non-linear decision surfaces in the input space [11].

4. EXPERIMENTAL RESULTS

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. The objective of overlapping neighboring frames is to consider the temporal characteristic of audio content. An input wav file is given to the feature extraction techniques. Spectral 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 music should be classified. 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. For the SVM-1 which is used to classify music into pop and classic used for training. For the SVM-2 which is used to classify classic and rock are used for training. Table 1 shows Performance of music classification in different SVM kernel function.

Table -1: Performance of music classification in different SVM kernel function.

SVM Kernels	Performance
Polynomial	86%
Gaussian	89%
Sigmoidal	88%

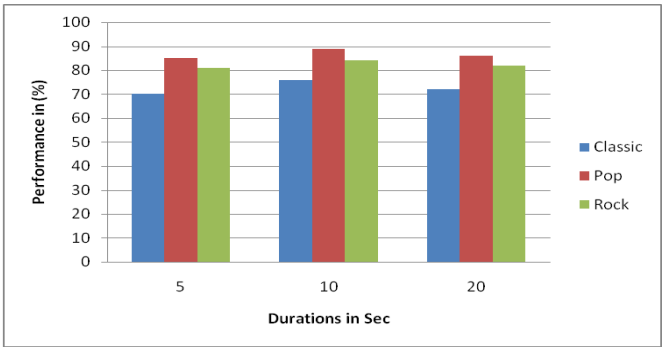

Chart -1: Performance of music classification for different duration of music clips

The performance of SVM for different duration as shown in Chart 1 shows that when the duration were increased from 10 to 20 there was no considerable increase in the performance.

5. CONCLUSION

In this paper, we have proposed an automatic music classification system using SVM. Spectral features are calculated as features to characterize audio content. SVM learning algorithm has been used for the classification of genre classes of music by learning from training data. Two

nonlinear support vector machine classifiers are developed to obtain the optimal class boundaries between classic and pop, pop and rock by learning from training data. Experimental results show that the proposed audio support vector machine learning method has good performance in musical genre classification scheme is very effective and the accuracy rate is 89%.

REFERENCES

[1] Z. Fu, G. Lu, K. M. Ting, and D. Zhang. A survey of audio-basedmusic classification and annotation. Multimedia, IEEE Transactionson, 13(2):303–319, 2011.

[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] J. Saunders, "Real-time Discrimination of Broadcast Speech/Music," International Conference on Acoustics, Speech and Signal Processing, Vol. 2, pp. 993–996, Atlanta, May, 1996.

[5] G. Peeters, "A Large Set of Audio Features for Sound Description," Technical representation, IRCAM, 2004.

[6] Breebaart J and McKinney M, "Features for Audio Classification," International Conference on Music Information Retrieval, 2003.

[7] Chungsoo Lim Mokpo, Yeon-Woo Lee, and Joon-Hyuk Chang, "New Techniques for Improving the practicality of a SVM-Based Speech/Music Classifier," IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 1657-1660, 2012.

[8] Hongchen Jiang, JunmeiBai, Shuwu Zhang, and Bo Xu, "SVM-Based Audio Scene Classification," IEEE International Conference Natural Language Processing and Knowledge Engineering, Wuhan, China, pp. 131-136, October 2005.

[9] Lim and Chang, "Enhancing Support Vector Machine-Based Speech/Music Classification using Conditional Maximum a Posteriori Criterion," Signal Processing, IET, vol. 6, no. 4, pp. 335-340, 2012.

[10] Md. Al Mehedi Hasan and Shamim Ahmad. predSucc-Site: Lysine Succinylation Sites Prediction in Proteins by using Support Vector Machine and Resolving Data Imbalance Issue. International Journal of Computer Applications 182(15):8-13, September 2018.

[11] Hend Ab. ELLaban, A A Ewees and Elsaeed E AbdElrazek. A Real-Time System for Facial Expression Recognition using Support Vector Machines and k-Nearest Neighbor Classifier. International Journal of Computer Applications 159(8):23-29, February 2017.