International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 09 | Sep 2018 www.irjet.net p-ISSN: 2395-0072

Music Classification using MFCC 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 Mel Frequency Cepstral Coefficients (MFCC). 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, *Mel Frequency Cepstral Coefficients* (*MFCC)* 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 MFCC features are extracted.

2.1 Mel Frequency Cepstral Coefficients

Mel Frequency Cepstral Coefficients (MFCCs) are short-term spectral based and dominant features and are widely used in the area of audio and speech processing. The mel frequency cepstrum has proven to be highly effective in recognizing the structure of music signals and in modeling the subjective pitch and frequency content of audio signals [4].

The MFCCs have been applied in a range of audio mining tasks, and have shown good performance compared to other features. MFCCs are computed by various authors in different methods. It computes the cepstral coefficients along with delta cepstral energy and power spectrum deviation which results in 26 dimensional features. The low order MFCCs contains information of the slowly changing spectral envelope while the higher order MFCCs explains the fast variations of the envelope [5].

MFCCs are based on the known variation of the human ears critical bandwidths with frequency. The filters are spaced linearly at low frequencies and logarithmically at high frequencies to capture the phonetically important characteristics of speech and audio. To obtain MFCCs, the audio signals are segmented and windowed into short frames of 20 ms. Magnitude spectrum is computed for each of these frames using Fast Fourier Transform (FFT) and converted into a set of mel scale filter bank outputs.

The human ear resolves frequencies non-linearly across the audio spectrum and empirical evidence suggests that designing a front-end to operate in a similar non-linear manner improves the performance. A popular solution is therefore filter bank analysis since this provides a much more straightforward route to obtain the desired non-linear frequency resolution. However, filter bank amplitudes are highly correlated and hence, the use of a cepstral transformation in this case is virtually mandatory. Fig. 1 describes the procedure for extracting the MFCC features.

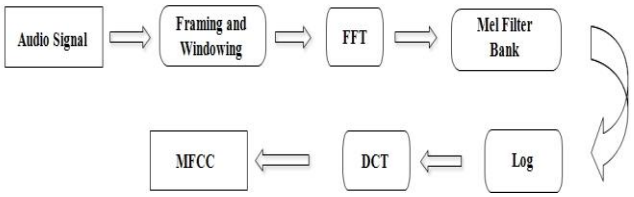

Fig -1: Extraction of MFCC from Audio Signal.

Mel frequency to implement this filter bank, the window of audio data is transformed using a Fourier transform and the magnitude is taken. The magnitude coefficients are then binned by correlating them with each triangular filter. Here, binning means that each FFT magnitude coefficient is multiplied by the corresponding filter gain and the results are accumulated. Thus, each bin holds a weighted sum representing the spectral magnitude in that filter bank channel.

Logarithm is then applied to the filter bank outputs. Discrete Cosine Transformation (DCT) is applied to obtain the MFCCs. Since the mel spectrum coefficients are real numbers, they are converted to the time domain using the DCT. In practice, the last step of taking inverse Discrete Fourier Transform (DFT) is replaced by taking DCT for computational efficiency. The cepstral representation of the speech spectrum provides a good representation of the local spectral properties of the signal for the given frame analysis. Typically, the first 13 MFCCs are used as features.

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 [6]. 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 [7].

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 [8]. The patterns lying on the margins which are maximized are the support vectors.

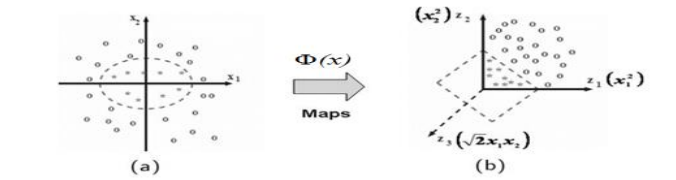

Fig -2: Example for SVM Kernel Function Φ(x) Maps 2-Dimensional Input Space to Higher 3-Dimensional Feature Space. (a) Nonlinear Problem. (b) Linear Problem.

The support vectors are the (transformed) training patterns and are equally close to hyper plane of separation. The support vectors are the training samples that define the optimal hyper plane and are the most difficult patterns to classify [9]. 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 [10].

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. MFCC 13 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 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	88%
Gaussian	91%
Sigmoidal	87%

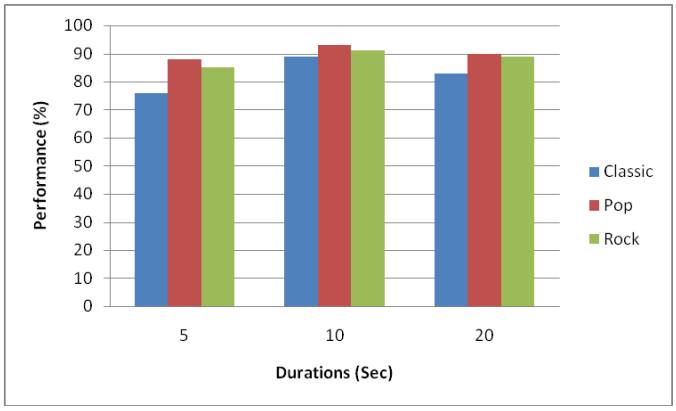

Chart -1: Performance of music classification for different duration of music clips

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. CONCLUSIONS

In this paper, we have proposed an automatic music classification system using SVM. MFCC is 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 93%.

REFERENCES

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

[2] Serwach, M., & Stasiak, B. (2016). 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. (2014). Radboud Universiteit Nijmegen Bachelor thesis Information Science Finding musical genre similarity using machine learning techniques, 1–25.

[4] O.M. Mubarak, E. Ambikai rajah and J. Epps, "Novel Features for Effective Speech and Music Discrimination," IEEE Engineering on Intelligent Systems, pp. 342-346, 2006.

[5] A. Meng and J. Shawe-Taylor, "An Investigation of Feature Models for Music Genre Classification using the Support Vector Classifier," International Conference on Music Information Retrieval, Queen Mary, University of London, UK, pp. 604-609, 2005.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.