SVM and GMM based Speech/music Classification using SBC

Dr. R. Thiruvengatanadhan
Assistant Professor
Department of Computer Science and Engineering
Annamlai University, Annamalainagar, Tamilnadu, India
Email Id: thiruvengatanadhan01@gmail.com

Abstract: *The audio refers to speech, music as well as any sound signal and their combination. The accuracy of the classification relies on the strength of the features and classification scheme. In this work, Subband Coding (SBC) features are extracted from the input signal. After feature extraction, classification is carried out, using Support vector machine (SVM) and Gaussian mixture model (GMM) model.*

Keywords: **Feature Extraction, Subband coding (SBC), Support vector machine (SVM) and Gaussian mixture model (GMM)**

1. INTRODUCTION

Audio refers to speech, music as well as any sound signal and their mixture. Audio consists of the fields namely file name, file format, sampling rate, etc. The need to naturally arrange, to which class a sound has a place, makes sound order and classification an arising and significant exploration region [1]. During the recent years, there have been numerous investigations on programmed sound grouping utilizing a few highlights and strategies. An information descriptor is frequently called an element vector and the cycle for removing such component vectors from sound is called sound element extraction. Generally an assortment of pretty much complex depictions can be extricated to highlight one bit of sound information. The effectiveness of a specific component utilized for examination and arrangement relies extraordinarily upon the application, the extraction cycle and the wealth of the depiction itself. Digital analysis may differentiate whether an acoustic file contains speech, music or other audio entities [2].

2. SUBBAND CODING (SBC)

Acoustic element extraction assumes a significant part in developing a sound arrangement framework. The point is to choose highlights which have huge among class and little inside class discriminative force. Discriminative intensity of highlights or capabilities tells how well they can separate various classes. Subband Coding (SBC) fuses the excitation in the discourse signal while mel-scale examination fuses properties of human hear-able framework [3]. In this work a bunch of highlights are separated dependent on the multi-rate subband investigation or wavelet examination of focused on discourse. The Discrete Cosine Transform (DCT) of subband energy for each casing in the discourse signal is separated utilizing perceptual wavelet parcel change.

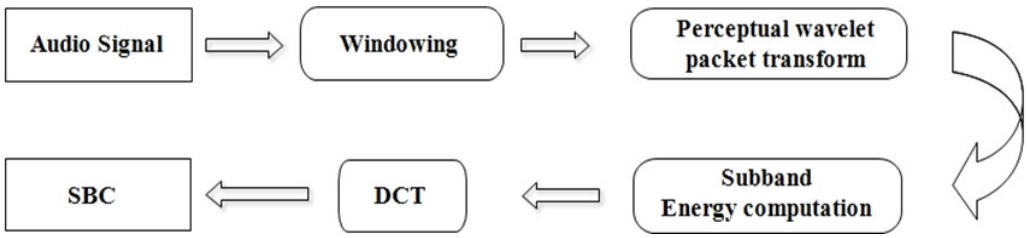

Figure 1. SBC Feature Extractions

This wavelet parcel change can be refined by two channel banks: low pass channel and high pass channel separately [4]. The current work is engaged to acquire the high energy data in the fell channel save with its wavelet bundle tree [5]. Figure 1 shows the block diagram of the extraction procedure of SBC feature.

3. SUPPORT VECTOR MACHINE (SVM)

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]. Figure 2 shows an example of a non-linear mapping of SVM to construct an optimal hyper plane of separation.

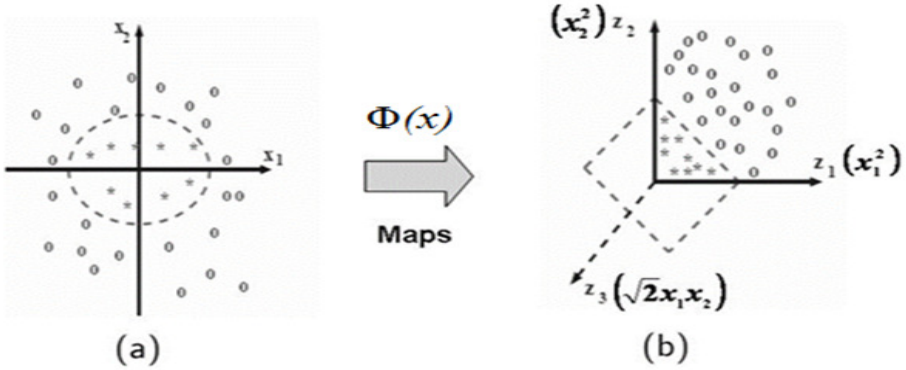

Figure 2 Example for SVM Kernel Function Φ(x) Maps 2-Dimensional Input Space to Higher 3-Dimensional Feature Space. (a) Nonlinear Problem. (b) Linear Problem.

The help vectors are the (changed) preparing designs and are similarly near hyperplane of division. The help vectors are the preparation tests that characterize the ideal hyperplane and are the most troublesome examples to order [8]. Casually, they are the examples generally instructive of the order task. The portion work creates the inward items to develop machines with various kinds of non-straight choice surfaces in the information space [9].

4. GAUSSIAN MIXTURE MODEL (GMM)

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 [10]. In non-parametric displaying just insignificant or no supposition with respect to the likelihood thickness capacity of highlight vector is made [11]. The Gaussian blend model (GMM) is utilized in grouping distinctive sound classes. The Gaussian classifier is an illustration of a parametric classifier. It is a natural methodology when the model comprises of a few Gaussian parts, which can be believed to display acoustic highlights. In arrangement, each class is spoken to by a GMM and alludes to its model. When the GMM is prepared, it very well may be utilized to anticipate which class another example presumably has a place [12]. Also, these models have the ability to form a smooth

approximation to the arbitrarily-shaped observation densities in the absence of other information [13].

5. EXPERIMENT AND RESULTS

5.1 The database

Execution of the proposed sound change point discovery framework is assessed utilizing the Television broadcast sound information gathered from Tamil stations, containing various terms of sound to be specific discourse and music from 5 seconds to 60 minutes. The sound comprises of shifting spans of the classifications, for example music followed by discourse and discourse in the middle of music and so forth, Audio is examined at 8 kHz and encoded by 16-bit.

5.2 Acoustic feature extraction

The element is extricated from each casing of the sound by utilizing the element extraction methods. Here the SBC highlights are taken. An information wav record is given to the component extraction procedures. The component esteems will be determined for the given wav record.

5.3 Classification

When the feature extraction process is done for the speech and music be classified. For the SVM which is used to classify speech and music used for training. Table 1 shows Performance of Speech/music classification in different SVM kernel function.

Table 1. Performance of Speech/music classification in different SVM kernel function.

SVM Kernels	Performance
Polynomial	87%
Gaussian	93%
Sigmoidal	85%

Gaussian mixtures for the two classes are modeled for the features extracted. We have chosen a mixture of 2, 5, 10 mixture models. Table.1 shows the performance of GMM for speech and music classification based on the number of mixtures.

Table 2. Performance of GMM for different mixtures.

GMM	2	5	10
Speech	92%	93%	91%
Music	93%	94%	90%

6. CONCLUSION

In this paper, we have proposed speech/music classification system using SVM and GMM. SBC is calculated as features to characterize audio content. The proposed classification method is implemented using EM algorithm approach to fit the GMM parameters for classification between speech/music by learning from training data. Experimental results show that the proposed audio GMM method has good performance

in Speech and Music classification scheme is very effective and the accuracy rate is 94% compared with SVM.

REFERENCES

[1] *H Watanabe SM, Kikuchi H (2010) Interval calculation of em algorithmfor gmm parameter estimation. Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium pp 2686–2689*

[2] *C. Panagiotakis and G. Tziritas. A speech/music discriminator based on rms and zero-crossings,.IEEE Trans. Multimedia, 7(5):155–156, February 2005.*

[3] *Zhu Leqing, Zhang Zhen "Insect Sound Recognition Based on SBC and HMM," International Conference on Intelligent Computation Technology and Automation, IEEE, pp. 544-548, 2010.*

[4] *Chaya. S, Ramjan Khatik, Siraj Patha and Banda Nawaz, "Subband Coding of Speech Signal Using Scilab", IPASJ International Journal of Electronics & Communication (IIJEC), vol. 2, Issue 5, 2014.*

[5] *Mahdi Hatam and Mohammad Ali Masnadi-Shirazi, "Optimum Nonnegative Integer Bit Allocation for Wavelet Based Signal Compression and Coding," Information Sciences Elsevier, pp. 332-344, 2015.*

[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] *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.*

[9] *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.*

[10] *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.*

[11] *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.*

[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*