SVM and GMM Based Speech Recognition using PNCC

R. Thiruvengatanadhan
Assistant Professor, Department of Computer Science and Engineering,
Annamalai University, Annamalainagar, Tamil Nadu, India.
thiruvengatanadhan01@gmail.com

Abstract:*Automatic recognition of speech to make fast between human and machine communication. This paper describes a technique that uses Support Vector Machine (SVM) and Gaussian Mixture Model (GMM) to recognized speech based on features using Power Normalized Cepstral Coefficients (PNCC). Displaying methods, for example, SVM and GMM were utilized to demonstrate every individual word which is prepared to the framework. Each segregated word Segment utilizing Voice Activity Detection (VAD) from the test sentence is coordinated against these models for finding the semantic portrayal of the test input discourse. Experimental results of GMM shows good performance in recognized rate.*

Keywords: Feature Extraction; Voice Activity Detection (VAD); Power Normalized Cepstral Coefficients (PNCC); support vector machines (SVM) and Gaussian Mixture Model (GMM).

1. Introduction

Research in Speech Recognition by machines had been done for almost four decades. Applications Discourse acknowledgment is an intricate grouping errand and characterized by various numerical methodologies: acoustic-phonetic methodology, design acknowledgment approach, man-made reasoning methodology, dynamic time traveling, connectionist approaches and backing vector machine. There have additionally been applications to discourse acknowledgment issues, to be specific phonetic order and post-arrangement of discourse acknowledgment speculations. Huge discourse information bases such a Television program, radio stations, CDs and DVDs are accessible on the web and the need to sort out such immense data sets becomes basic these days [1]. As Large Vocabulary Continuous Speech Recognition (LVCSR) is flawed, programmed discourse records contain blunders. Because of capacity limitations, research identified with discourse ordering and recovery has gotten a lot of consideration. As capacity has gotten less expensive, enormous assortment of spoken reports is accessible on the web, yet there is an absence of satisfactory innovation to clarify them. Manual record of discourse is exorbitant and furthermore has security requirements [2].

2. Voice Activity Detection

Voice Activity Detection (VAD) is a strategy for finding voiced portions in discourse and assumes a significant function in discourse mining applications. VAD disregards the extra sign data around the word viable. It tends to be likewise seen as a speaker free word acknowledgment issue. The essential standard of a VAD calculation is that it removes acoustic highlights from the info sign and afterward contrasts these qualities and limits as a rule separated from quietness. Voice movement is announced if the deliberate qualities surpass the limit. Something else, no discourse action is available [3].

VAD discovers its use in an assortment of discourse correspondence frameworks like coding of discourse, perceiving discourse, hands free communication, sound conferencing, discourse upgrade and wiping out of sound. It distinguishes where the discourse is voiced, unvoiced or continued and gains smooth ground of the discourse cycle [4]. An edge size of 20 ms, with a cover of half, is considered for VAD. RMS is separated for each casing. Figure 1 shows the disconnected word partition.

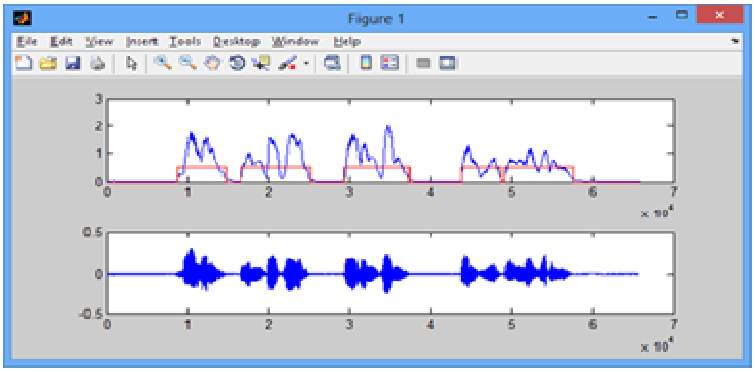

Figure 1 Isolated Word Separations.

3. Power Normalised Cepstral Coefficients (PNCC)

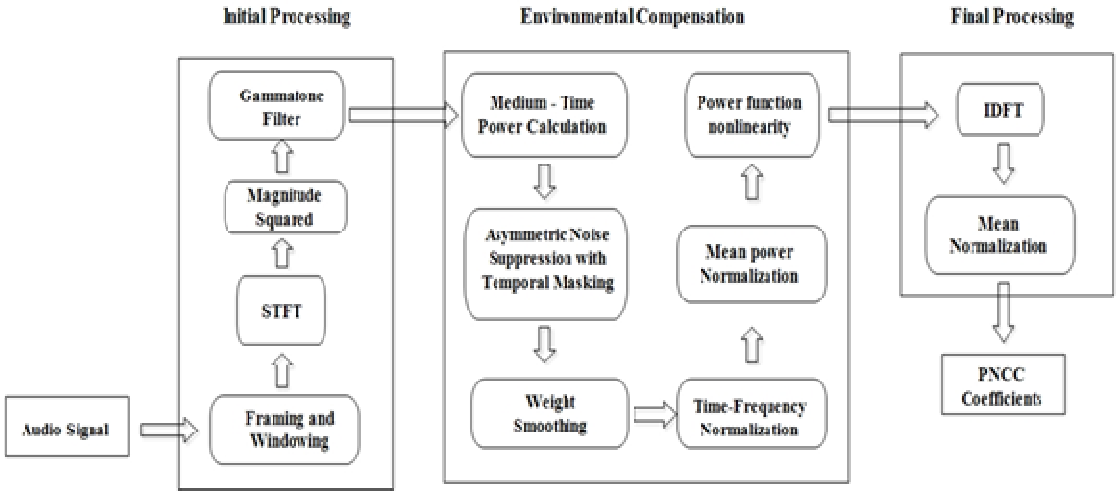

Figure 2 PNCC Feature Extractions

PNCC is notable for the high exactness of programmed discourse acknowledgment frameworks even in high-commotion conditions [5]. PNCC is an acoustic component which plays out the calculation utilizing on the web calculations continuously and gives high precision even in uproarious conditions [6]. (PNCC) is notable for the precision of

programmed discourse acknowledgment frameworks, even in high-clamor conditions. In Figure 2 Shows the square chart for the extraction of PNCC.

4. Support Vector Machine (SVM)

An AI procedure which depends on the standard of structure hazard minimization is uphold vector machines. It has various applications in the zone of example acknowledgment. SVM develops direct model dependent on help vectors so as to assess choice capacity. In the event that the preparation information are straightly detachable, at that point SVM finds the ideal hyper plane that isolates the information without mistake [7]. Fig. 3 shows a case of a non-direct planning of SVM to develop an ideal hyper plane of partition. SVM maps the information designs through a non-straight planning into higher measurement highlight space. For directly detachable information, a straight SVM is utilized to group the informational collections [8]. The examples lying on the edges which are expanded are the help vectors.

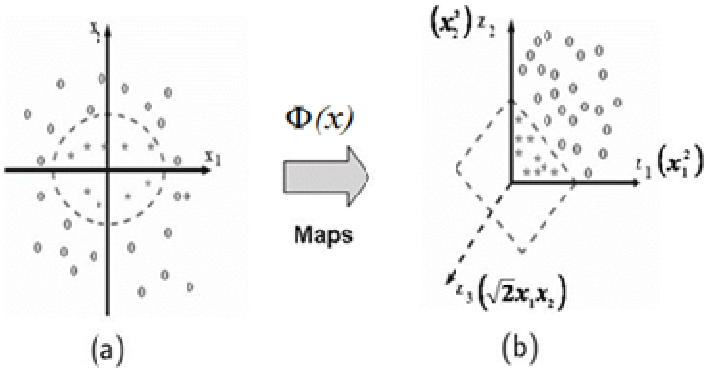

Figure 3 Example for SVM Kernel Function

5. Gaussian Mixture Models (GMM)

The likelihood conveyance of highlight vectors is demonstrated by parametric or non-parametric techniques. Models which accept the state of likelihood thickness work are named parametric. In non-parametric demonstrating, negligible or no presumptions are made with respect to the likelihood circulation of highlight vectors. In this part, we quickly survey GMM, for sound characterization. The reason for utilizing GMM is that the dissemination of highlight vectors extricated from a class can be demonstrated by a combination of Gaussian densities. The iterative Expectation Maximization (EM) calculation is utilized to assess the boundaries of GMM. EM calculation is one of the most mainstream bunching calculations used to assess the probabilistic models for each Gaussian part. The Expectation step (E-step) and Maximization step (M-step) are iterated till the union of the boundary [9]. EM calculation discovers most extreme probability assessment of boundaries. The E-step figures Expectation of probability accepting boundaries and M-step processes greatest probability assessments of boundaries by expanding the normal probability found in E-step.

6. Experimental Results

6.1.Dataset Collection

A complete dataset of 100 data, going from 5 to 10 seconds term, inspected at 16 kHz and encoded by 16-digit is recorded. Voice movement recognition is performed to segregate the words in every discourse record utilizing RMS energy envelope.

6.2.Acoustic feature extraction

In this work the pre-accentuated signal containing the nonstop discourse is taken for testing. Through VAD the detached words are removed from the sentences. Accordingly outlines which are unvoiced excitations are eliminated by thresholding the fragment size. Highlight PNCC are separated from each edge of size 320 window with a cover of 120 examples. Along these lines it prompts 13 PNCCs separately which are utilized independently to speak to the disconnected word fragment. During preparing measure each segregated word isisolated into 20ms covering windows for separating 13 PNCCs highlights. Utilizing VAD secluded words in a discourse is isolated.

6.3.Classification

N-SVMs are made for each word. The preparation cycle examines discourse preparing information to locate an ideal method to order discourse outlines into their individual classes. The inferred help vectors are utilized to order discourse information. For testing 13 dimensional PNCC highlight vectors were given as contribution to SVM model and the separation between every one of the element vectors and the SVM hyperplane is acquired. The normal separation is determined for each model. The content relating to the question discourse is chosen dependent on the greatest separation. A similar cycle is rehashed for various inquiry discourse, and the exhibition is considered. The exhibitions of discourse acknowledgment for various SVM parts are thought about for PNCC acoustic highlights are appeared in Figure 4.

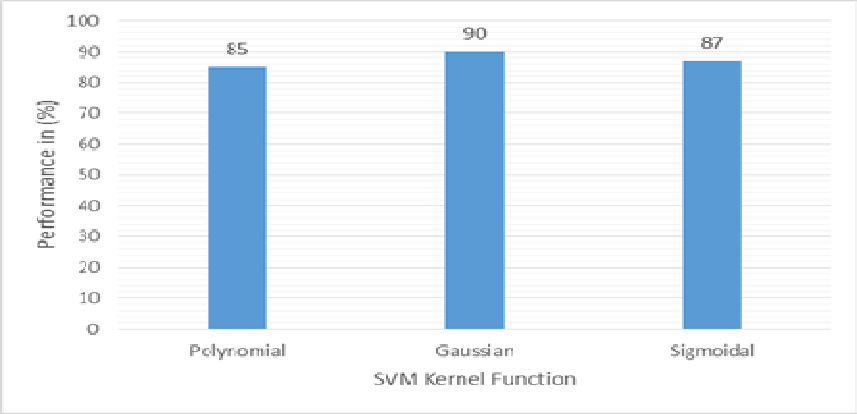

Figure 4. Performance of speech recognition rate in different SVM kernel function.

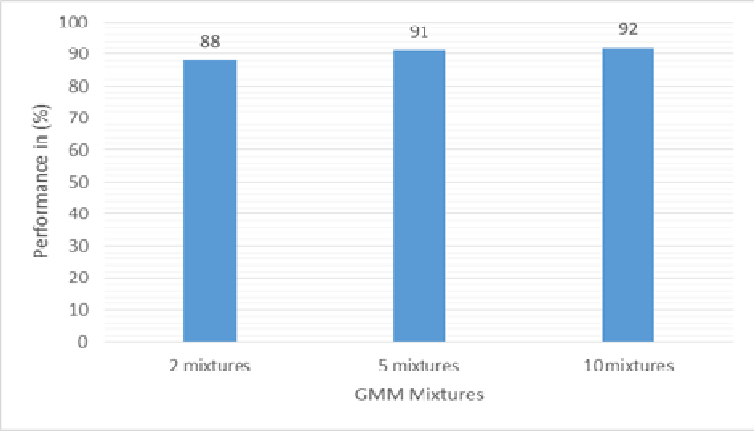

Figure 5 Performance of GMM for different mixtures.

In this paper, we have proposed an automatic music genre classification system using GMM. PNCC is calculated as features to characterize music content. 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. GMM method has good performance of accuracy rate is 92% shown in Figure 5.

7. Conclusion

In this paper, we have proposed discourse acknowledgment framework utilizing SVM and GMM. Voice Activity Detection (VAD) is utilized for isolating individual words out of the constant discourses. Highlights for each detached word are separated and those models were prepared effectively. SVM is utilized to demonstrate every individual expression. MFCC is determined as highlights to describe sound substance. SVM and GMM learning calculation has been utilized for the perceived discourse by gaining from preparing information. Test results show that the proposed sound help vector technique has great execution in GMM is 92% compare with SVM classifier.

REFERENCES

[1] *YaliZheng, Chisaki, Y. and Usagawa T., "Speech/Music Indexing for Audio Life Logs from Portable Device Record," IEEE International Conference on Advanced Computer Science and Information Systems, 2013, pp. 173 -178.*

[2] *Iswarya, P. and Radha, V, "Speech and Text Query Based Tamil - English Cross Language Information Retrieval system," International Conference on Computer Communication and Informatics, Coimbatore, 2014, pp. 1 -4,*

[3] Khoubrouy, S. A. and Panahi, I.M.S., "Voice Activation Detection using Teager-Kaiser Energy Measure," International Symposium on Image and Signal Processing and Analysis, pp. 388-392, 2013.

[4] Tayseer M F Taha and Amir Hussain. "A Survey on Techniques for Enhancing Speech", International Journal of Computer Applications, February 2018.

[5] Xin Yan and Ying Li, "Anti-noise Power Normalized Cepstral Coefficients for Robust Environmental Sounds Recognition in Real Noisy Conditions," Fourth International Conference on Computational Intelligence and Communication Networks, 2012, pp. 263-267

[6] Chanwookim, Stern, R.M. "Power-Normalized Cepstral Coefficients (PNCC) for robust speech recognition" IEEE International Conference on Acoustics, Speech and Signal Processing, 25-30 March 2012, pp:4101 –4104.

[7] Shweta V Raut and Madhu M Nashipudimath. Review: Sentiment Analysis using SVM Classification Approach. International Journal of Computer Applications 181(37):1-8, January 2019.

[8] ManishaPrajapati and ArchitYajnik and. POS Tagging of Gujarati Text using VITERBI and SVM. International Journal of Computer Applications 181(43):32-35, March 2019.

[9] MdMahadiHasanNahid, MdAshraful Islam and MdSaiful Islam. Comparison of VQ and GMM for Text Independent Speaker Identification System for The Bengali Language. International Journal of Computer Applications 178(47):18-21, September 2019.