iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 **Issue:** V **Month of publication:** May 2017

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

A Review on Speech Segmentation Technique

Vikas Malik[1], Rekha[2]
[1,2] Assistant Professor, [2]M. Tech Student
BPS Mahila Vishwavidyalaya Khanpur Kalan (Sonipat)

Abstract: ***As information technology has an impact on more and more aspects of our daily lives, the problem of communication between human beings and information-processing devices become increasingly important. Up to now such communication has been run almost entirely by means of keyboards and screens, but speech is by far the most widely used, natural and fast means of communication for people. Unfortunately, machine capabilities for interpreting speech are still poor in comparison to what a human can achieve. Speech segmentation is the process of identifying the boundaries between words, syllables, or phonemes in spoken natural languages. The term applies both to the mental processes used by humans, and to artificial processes of natural language processing. Speech segmentation is a subfield of general speech perception and an important sub problem of the technologically focused field of speech recognition, and cannot be adequately solved in isolation. In this paper we study about the existing techniques used for speech classification.***

Keywords: - Speech, SVM, Feature extraction, Type Classification

I. INTRODUCTION

Data mining, an interdisciplinary subfield of computer science, is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Aside from the raw analysis step, it involves database and data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating. The term is a misnomer, because the goal is the extraction of patterns and knowledge from large amount of data, not the extraction of data itself. It also is a buzzword and is frequently applied to any form of large-scale data or information processing (collection, extraction, warehousing, analysis, and statistics) as well as any application of computer decision support system, including artificial intelligence, machine learning, and business intelligence. The popular book "Data mining: Practical machine learning tools and techniques with Java" (which covers mostly machine learning material) was originally to be named just "Practical machine learning", and the term "data mining" was only added for marketing reasons. Often the more general terms "(large scale) data analysis", or "analytics" or when referring to actual methods, artificial intelligence and machine learning are more appropriate.

II. LITRETURE SURVEY

In this section we study about the research work done by researcher and scientists. The study of existing work moves us towards some issues.

In the year 2001 they provided a tutorial on learning and inference in hidden Markov models in the context of the recent literature on Bayesian networks. The perspective makes it possible to consider novel generalizations of hidden Markov models with multiple hidden state variables, multi-scale representations, and mixed discrete and continuous variables. Although exact inference in these generalizations is usually intractable, one can use approximate inference algorithms such as Markov chain sampling and variation methods. They conclude this review with a discussion of Bayesian methods for model selection in generalized HMMs. [1]

In the year 2007 he described a problem in speech recognition and also in automatic phonetic transcription from read speech is accurate segmentation of the incoming speech signal into syllable-sized segments. One common and also quite simple algorithm is to use the intensity from the original signal and the intensity from one or more band pass filtered versions of the signal. These are compared using different criteria to determine the syllabic boundaries in the speech signal [2].

In the year 2008, they presented a discriminative training algorithm that uses support vector machines (SVMs), to improve the classification of discrete and continuous output probability hidden Markov models (HMMs). The algorithm uses a set of maximum-likelihood (ML) trained HMM models as a baseline system, and SVM training scheme to restore the results of the baseline HMMs. It turns out that the rescoring model can be represented as an un-normalized HMM. They described two algorithms for training the

©IJRASET: All Rights are Reserved

un-normalized HMM models for both the discrete and continuous cases. One of the algorithms results in a single set of un-normalized HMMs that can be used in the standard recognition procedure (the Viterbi recognizer), as if they were plain HMMs [3]. They used a toy problem and an isolated noisy digit recognition task to compare their new method to standard ML training. Their experiments show that SVM rescoring of hidden Markov models typically reduces the error rate significantly compared to standard ML training.

In the year 2009 they proposed a more natural approach is to segment the speech signals on the basis of time-frequency analysis. Boundaries are assigned in places where some energy of a frequency band rapidly changes. In this research they apply the discrete wavelet transform (DWT) to analyses speech signals, the resulting power spectrum and its derivatives. This information allows us to locate the boundaries of phonemes. It is the first stage of speech recognition process. Additionally they presented an evaluation by comparing their method with hand segmentation. The segmentation method proves effective for finding most phoneme boundaries [4].

In the year 2011they proposed a novel face recognition algorithm based on Gabor texture information is proposed in this paper. Two strategies to capture it are Gabor Magnitude-based Texture Representation (GMTR) which is characterized by using the Gamma density to model the Gabor magnitude distribution and Gabor Phase-based Texture Representation (GPTR), characterized by using the Generalized Gaussian Density (GGD) to model the Gabor phase distribution. The estimated model parameters serve as texture representation. Experiments are performed on Yale, ORL and FERET databases to validate the feasibility of the method. The results show that GMTR based and GPTR-based SVM classifier both significantly outperform the widely used Gabor energy based systems and other existing subspace methods. [5]

In the year 2012 he described handwritten Numeral recognition plays a vital role in postal automation services especially in countries like India where multiple languages and scripts are used Discrete Hidden Markov Model (HMM) and hybrid of Neural Network (NN) and HMM are popular methods in handwritten word recognition system. The hybrid system gives better recognition result due to better discrimination capability of the NN. A major problem in handwriting recognition is the huge variability and distortions of patterns. Elastic models based on local observations and dynamic programming such HMM are not efficient to absorb this variability. But their vision is local. But they cannot face to length variability and they are very sensitive to distortions. Then the SVM is used to estimate global correlations and classify the pattern. Support Vector Machine (SVM) is an alternative to NN [6]. In Handwritten recognition, SVM gives a better recognition result. The aim of this research was to develop an approach which improves the efficiency of handwritten recognition using artificial neural network.

In the year 2013 they presented [7] strong research and development interests in multimedia database in order to effectively and efficiently use the information stored in these media types. The research effort of the past few years has been mainly focused on indexing and retrieval of digital images and video. In video retrieval, the most common use of audio information is for automatic speech recognition and the subsequent use of the generated transcript for text retrieval. However, the audio information can also be used, more directly, to provide additional information such as the gender of the speaker, music and speech separation and audio textures such as fast speaking sports announcers. So that the applications describing the content and the applications using the corresponding descriptions can interoperate, it is necessary to define a standard that specifies the syntax and semantics of these multimedia descriptions.

In the year 2014 they described about emotion detection in speech processing is one of the burning arenas in data mining field. Detecting the motion of the speech is not that easy as it seems to be. Many different researchers have tried their approach in this field but accuracy is the major factor of the processing. Their basic problem is to detect the kind of emotion gets detected from a pitch file. This would be done with the help of the HMM algorithm which would identify the frequency parameters. Then after finding the exact length of the file, they will have to get into the predefined clusters. Mugging into the predefined clusters would be achieved by the SVM algorithm and each cluster will roll back to a result value. The exact cluster which would give us the maximum probilitical analysis of the file would be their target cluster [8]. This work was done previously with the help of ANN algorithm and they have provided an accuracy of about 92.1%.Their problem would be increasing this accuracy ratio, in comparison to the ANN module.

III. SEGMENT FEATURES

Segment length: The length of the segment.

Words: The number of words recognized inside the segment.

Boundary length: The length of the boundary token (i.e. the silence or non-speech region) recognized at the end of the segment.

Boundary confidence: The confidence score of the boundary token recognized.

©IJRASET: All Rights are Reserved

Warping factor variance: The variance of the VTLN warping factors classified inside the segment. This feature should give an estimate of the speaker homogeneity in a segment.

IV. PROBLEM STATEMENT

As in most natural language processing problems, one must take into account context, grammar, and semantics, and even so the result is often a probabilistic division (statistically based on likelihood) rather than a categorical one. Though it seems that co-articulation - a phenomenon which may happen between adjacent words just as easily as within a single word - presents the main challenge in speech segmentation across languages.

V. MOTIVATION

The research is motivated by the observation that almost all native speakers perceive, relatively easily, the acoustic characteristics of their own language when it is produced by speakers of the language. Small variations within a phoneme category, sometimes different for various phonemes, do not change significantly the perception of the language's own sounds. Several methods are introduced based on similarity measures of the Euclidean space spanned by the acoustic representations of the speech signal and the Euclidean space spanned by an auditory model output, to identify the problematic phonemes for a given speaker. The methods are tested for groups of speakers from different languages and evaluated according to a theoretical linguistic study showing that they can capture many of the problematic phonemes that speakers from each language mispronounce.

VI. OBJECTIVE

Step1: Study about voice segmentation and classification methods.
Step2: Take seven samples of voices like apple, banana, kiwi, lime, orange, peach, and pineapple on which improved and existing techniques applied.
Step3: Modify SVM according to working of this in multi-dimensional. New technique is known Multi-SVM.
Step4: Data goes through the training operation then this trained data is tested. Now data is ready for classification.
Step5: Apply Multi-SVM technique and get result.
Step6: Apply HMM (Hidden Markov Model) on the same sample of voices.
Step6: Find expected results.

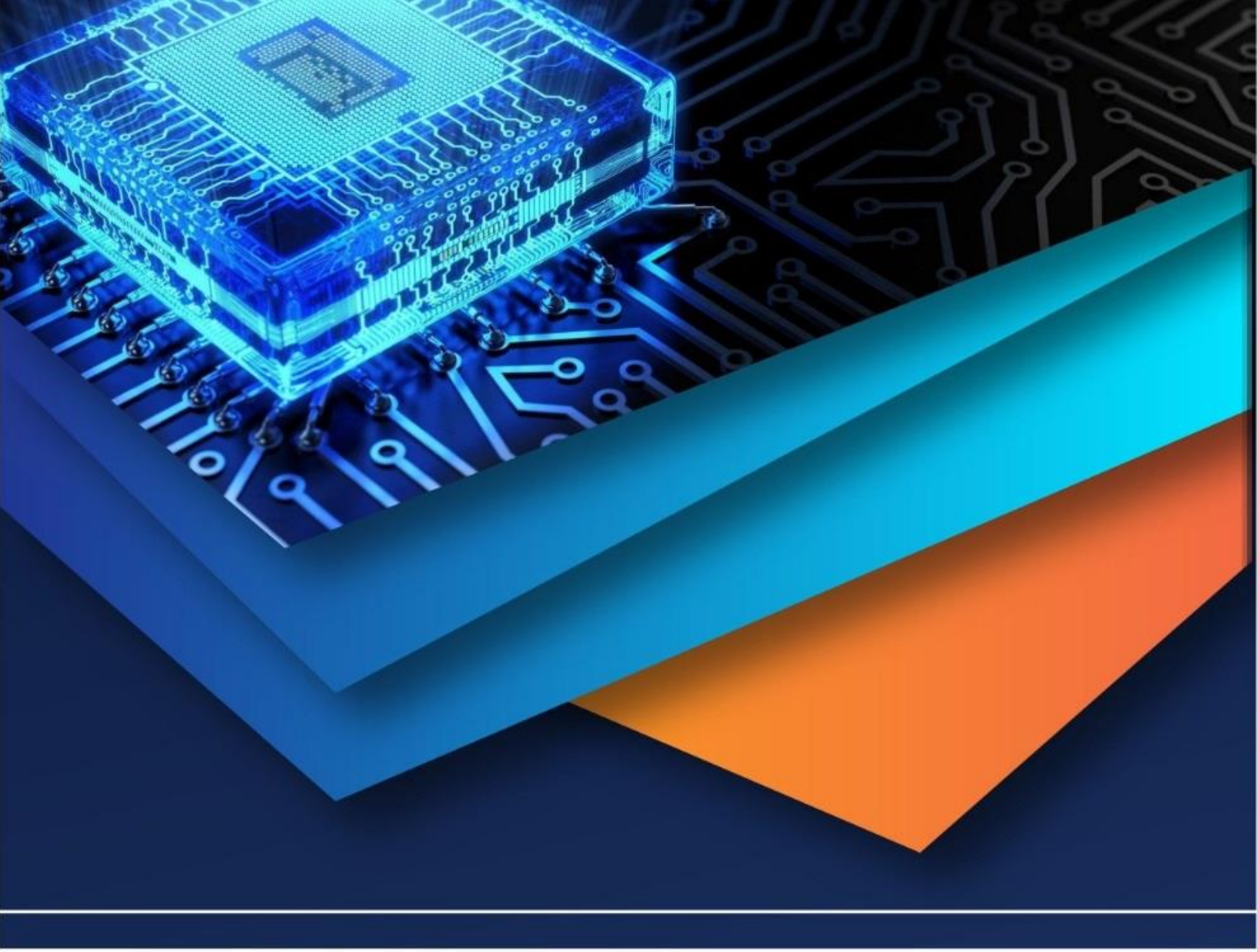
VII. CONCLUSION

Successful completion of the voice classification of the given voice files for both algorithms. Along with this we have the completion of classification for the two algorithms of modified SVM & HMM classification techniques. Comparison is performed by these algorithms on voice file. Only some samples of voice are taken and classify voice according words. In future the proposed algorithm can be applied on video files. The variety and quantity of data is constant in this work so in future we can vary these issues also.

REFERENCES

[1] Ghahramani, Z., "An Introduction to Hidden Markov Models and Bayesian Networks", International Journal of Pattern Recognition and Artificial Intelligence, 2001
[2] Lars Eriksson, "Algorithms for Automatic Segmentation of Speech" Lund University, Dept. of Linguistics Working Papers 35 (2007), Page: 53-61
[3] Alba Sloin and David Burshtein, "Support Vector Machine Training for Improved Hidden Markov Modeling" IEEE Transactions On Signal Processing, Vol. 56, No. 1, January 2008.
[4] Bartosz Zi´ołko, Suresh Manandhar, Richard C. Wilson and Mariusz Zi´ołko, "Wavelet Method of Speech Segmentation" (http://www-users.cs.york.ac.uk/bziolko/), 2012.
[5] Latha Parthiban, "A Novel Face Recognition Algorithm with Support Vector Machine Classifier", International Journal of Mathematics and Scientific Computing, Vol. 1, No. 1, 2011
[6] Anshuman Sharma, "Handwritten digit Recognition using Support Vector Machine" http://arxiv.org/ftp/arxiv/papers/1203/1203.3847.pdf, 2012
[7] Khin Myo Chit, K Zin Lin, "Audio-Based Action Scene Classification Using HMM-SVM Algorithm", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 2, Issue 4, April 2013.
[8] Vijaya lakshmi B, Dr.Sugumar Rajendran, "Survey on Speech Emotion Detection Using SVM and HMM", IJIRCCE (An ISO 3297: 2007 Certified Organization) Vol. 2, Issue 8, August 2014.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)