International Journal of Scientific Research & Engineering Trends
Volume 5, Issue 6, Nov-Dec-2019, ISSN (Online): 2395-566X

Real Time Enhanced Speech Recognition Technique to Operate Computer System Using SVM:Review

Jadhav Sayali Kadam Madhuri Talreja Khushboo Khatake Sujata
Department of Information Technology
Vidya Prati Sthan's Kamalnayan Bajaj Institute of Engineering and Technology
Baramati, Maharashtra

Abstract **-Speech is unique human characteristic used as tool to communicate. Speech Recognition technology is one most of the fast growing engineering technologies. It has a number of applications in different areas and provides potential benefits. Nearly 20% people of the world are suffering from various disabilities, the speech recognition system in those particular cases provide a significant help to them, so that they can share information with people by operating computer through voice input. The main objective is to recognize the spoken word by same speaker using multi-class support vector machine. Because of their limitations and high cost, voice recognition systems have traditionally been used only in a few specialized situations. For example, such systems are useful in instances when the user is unable to use a keyboard to enter data because his or her hands are occupied or disabled. Instead of typing commands, the user can simply speak into a microphone. The performance of system gives better result with reducing the work of human in terms through input device or touch screen. This work is proposed to operate the computer system through the voice commands like open or close file or any computer operations.**

Keywords **-Speech, support vector machine, machine operation.**

I. INTRODUCTION

Speech recognition is a main core of spoken language systems. Speech recognition is a complex classification task and classified by different mathematical approaches[15]. The speech recognition is based on the speaker how speaking and the environment in which speaker giving the context message. In this work, we review the recent techniques related to Speech Recognition system based on the few factors which includes the feature-extraction phase, training phase and testing phase. The speech recognition system work with number of applications like speech to text and to recognize the spoken word etc. also it is possible to use voice communication with computers to operate the computer system. In this system the input data will be large voice data in .wave file to apply it to the pre-processing phase. Then the pre-processed samples of signal are taken for training and testing phase and these are machine command used to operate computer system to open particular application[14].The input signal data is in English language to check the efficiency of the proposed system. Input signal is captured using microphone. The input signals given to the system are totally isolated word. Features for each isolated word are extracted and those models were trained. SVM modeling technique is used to model each individual utterance. Thus, each isolated word segment from the test sentence is matched against these models for finding the semantic representation of the test input. The aim of the proposed system is to create intelligent voice operator system which can be used by handicap person or person having some motor disabilities. The provided system performs the computer operations without any input device[14]. The objective of the proposed system is to overcome the disadvantage of increasing cost and reducing speed while recognizing the spoken word from the dictionary which requires internet connection and hence cannot works offline, and develop a speech recognition system requiring minimum storage and standalone operation by devices, which is faster, requires less memory and gives better accuracy.

II. LITERATURE SURVEY

Boonserm Kijsirikul and Nitiwut Ussivakul, [1]worked on a method of extending Support Vector Machines (SVMs) for dealing with multiclass problems. A modified structure of the DDAG that has a lower number of decision levels and reduces the dependency on the sequence of nodes. Thus, the ADAG improves the accuracy of the DDAG while maintaining low computational requirement.

Xin Dong, Wu Zhaohui and Pan Yunhe, [2]They propose a novel approach that contains support vectors describing the hypersphere to separates the samples. We also generalize it in multi-class classification phrase.Support Vector Machines have been introduced in for solving pattern recognition and nonlinear function estimation problems.

Chen Junil, Jiao Licheng, [3] Described an overview of Structural Risk Minimization (SRM) principle, and describes the mechanism of how to construct SVMs. For a two-class pattern recognition problem.

K.P. Bennett, J.A. Blue, [4] This Shows result indicate that the method produces simple trees that generalize well with respect to other decision tree algorithms and single support vector machines.

C. J. C. Burges, [5] Provide an introductory yet extensive tutorial on the basic ideas behind Support Vector Machines (SVMs). Support Vector Machines that learn classification problem are specific to use hyper plane. They propose a novel approach that contains support vectors describing the hyper sphere to separates the samples.

Hiroshi Shimodaira, et al, [6] Support Vector Machine (SVM) is one of the latest and most successful statistical pattern classifiers that utilizes a kernel technique.Support Vector Machines have been introduced for solving pattern recognition problems and nonlinear function estimation problems.

Kartiki Gupta, Divya Gupta, [7]They had discussed some features extraction techniques and their performance. An analysis on LPC, RASTA and MFCC techniques in Automatic Speech Recognition System.

Pratik K. Kurzekar, Ratnadeep R. Deshmukh, Vishal B. Waghmare, Pukhraj P. Shrishrimal,[8] A Comparative Study of Feature Extraction Techniques for Speech Recognition System.The techniquesLPC, PLP and MFCC are the most frequently used features extraction techniques in the fields of speech recognition and speaker verification applications.

S. Gunn, [9] proposed the term SVM will refer to both classification and regression methods, and the terms Support Vector Classification (SVC) and Support Vector Regression (SVR) will be used for specification.

Singhal and Dubey, [10] had implemented automatic speech recognition for connected words using DTW/HMM for English /Hindi languages. In which author shows their work on isolated word recognizer for speaker dependent data works in both English as well as Hindi using DTW & HMM techniques, both techniques are compared in this paper.

Patil, Admuthe, &Zirmite, [11] had described isolated digit recognition using ear microphone data using MFCC, VQ & HMM ,here proposed model implement isolated digit into three stages, which are endpoint detection & speech segmentation, another one included feature extraction which is done by MFCC, & at last it conducted both VQ & HMM based classifier, works on isolated English digits .

Nandyala, [12] had described real time isolated word speech recognition system for human computer interaction. In which proposed model works on speaker dependent system, which is implemented using MFCC feature extraction technique & Dynamic programming algorithm. This work is used in tourism application. System obtained 88.0% accuracy.

MengdiYuet, Ling Chen, Jie Zhang, Hong Liu, [13] had implemented speaker age recognition based on isolated words by using SVM, here proposed model recognizes speaker's age using SVM with MFCC feature extraction technique of 4507 isolated words, which achieves 72.93% accuracy. Author also compare HMM & SVM both & based on that it is observed that SVM is more suitable.

III. METHODOLOGY

1.Architecture Diagram

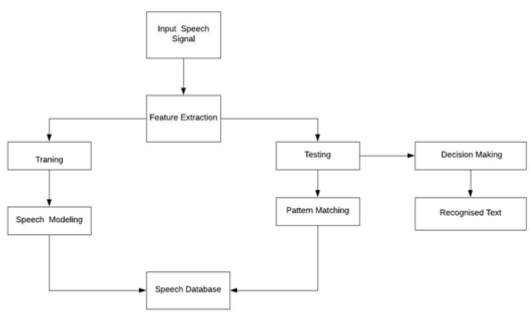

Fig.1. Architecture diagram[7].

2. Pre-Processing Phase- The input taken in the form of wave file is applied to pre-processing unit. In this phase using end point detection technique unwanted noise is removed from the input[14]. Hence, we get the clear speech signals and no non –speech signals. The next step is to segment the speech signals into the frames with frame size ranging from 10 to 25 milliseconds.
The architecture shows three important parts:
1. Feature Extraction phase
2. Training phase
3. Testing phase

3. Feature Extraction Phase-
Feature extraction is the process that captures speech properties such as fundamental frequency and the loudness of a signal the output that produced from feature extraction process are parametric numerical descriptions of signals.

Feature Extraction process aims to reduce data by extracting the most meaningful information form the signal which leads to reduces the dimensionality of the input vectors.Linear PredictiveCoding(LPC) is feature extraction technique used to calculate the features of sample speech signals.

4. Linear Predictive Coding

The input speech signals given by the speaker are compared to sample input features through time alignment algorithm. Then these input signals and sample features are calculated using Linear Predictive Coding (LPC) method. LPC is the most powerful, robust, accurate, reliable and popular speech analysis technique tool for speech recognition, compression and synthesis[7]. Also, it is one of the most useful methods for encoding good quality speech at a low bit rate and provides extremely accurate estimates of speech parameters.

The most important aspect of LPC is linear predictive filter which allows the value of the next sample to be determined by a linear combination of previous samples. In order to implement LPC and generate the features the input speech signals needs to pass through pre-emphasize, the output of pre-emphasize acts as the input to frame blocking where the signal is blocked into frames of N samples[7]. The next step after frame blocking is windowing where each frame is windowed in order to reduce signal discontinuity at the beginning and end of every frame. Hamming frame is an example of typical frame. After windowing each windowed frame is auto correlated and the highest autocorrelation value gives the order of LPC analysis and finally LPC coefficients are derived[7].

Common applications of LPC : Speech Vocoders, Spectral Analysis, Pitch Estimation(voice changers), analysis/synthesis of instrument sounds(voice box),Compression.

5. Training Phase:

5.1 Support vector classification

The classification problem can be restricted to consideration of the two-class problem without loss of generality. In this problem the goal is to separate the two classes by a function which is induced from available examples[9]. The goal is to produce a classifier that will work well on unseen examples, i.e. it generalizes well. Consider the example in Figure. Here there are many possible linear classifiers that can separate the data, but there is only one that maximizes the margin (maximizes the distance between it and the nearest data point of each class). This linear separable case is termed the optimal separating hyperplane[9].

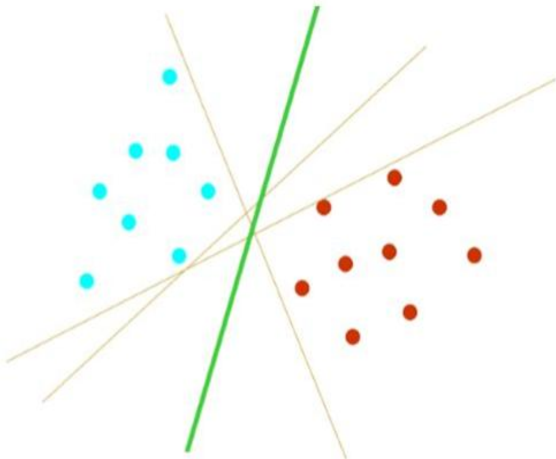

Fig.1.Classification(Linear Separable Case)[9].

5.2 Linear Separable Case

Suppose some data points, each belonging to one of two sets, are given and we wish to create a model that will decide which set a new data point will be in. In the case of support vector machines, a data point is viewed as a p-dimensional vector (a list of p numbers), and we want to know whether we can separate such points with a (p − 1)-dimensional hyper plane This is called a linear separable case. There are many hyper planes that might classify (separate) the data. One reasonable choice as the best hyper plane is the one that represents the largest separation, or margin, between the two sets. So, we choose the hyper plane so that the distance from it to the nearest data point on each side is maximized. If such a hyper plane exists, it is known as the maximum-margin hyper plane and the linear classifier it defines is known as a maximum margin classifier. More formally, given some training data D, a set of n points of the form

$$\mathcal{D} = \{(\mathbf{x}_i, y_i) \mid \mathbf{x}_i \in \mathbb{R}^p,\ y_i \in \{-1, 1\}\}_{i=1}^{n}$$

where the yi is either 1 or −1, indicating the set to which the point xi belongs. Each xi is a p-dimensional real vector. We want to find the maximum-margin hyperplane that divides the points having yi=1 from those having yi=-1. Any hyperplane can be written as the set of points x satisfying-b=0,where, denotes the dot product and w the (not necessarily normalized) normal vector to the hyperplane.

6 Non-Linear Separable Case:

In the linearly separable case, SVM is trying to find the hyperplane that maximizes the margin, with the condition that both classes are classified correctly. But in reality, datasets are probably never linearly separable, so the condition of 100% correctly classified by a hyperplane will never be met.

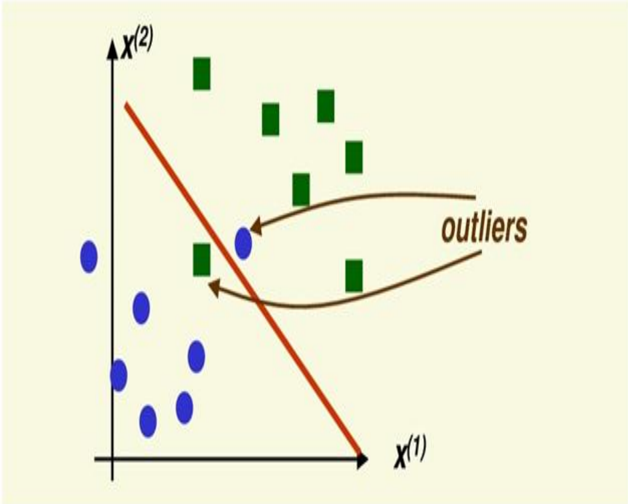

Fig.2. SVM Classification(Non-Linear Separable Case).

Therefore, since the training data is now non-linearly separable, we must admit that the hyperplane found will misclassify some of the samples. This misclassification is a new variable in the optimization that must be taken into account. The new model has to include both the old requirement of finding the hyperplane that gives the biggest margin and the new one of generalizing the training data correctly by not allowing too many classification errors.
Non-linear case can use any linear separable case after lifting data into a higher dimensional space.

7. Testing Phase

7.1 Adaptive Directed Acyclic Graph SVM:

An Adaptive Directed Acyclic Graph (ADAG) is a Directed Acyclic Graph with a reversed triangular structure. In an N-class problem, the system comprises N(N-I)/2 binary classifiers[1]. The ADAG has N-1 internal nodes, each of which is labelled with an element of a Boolean function. The nodes are arranged in a reversed triangle with Ni2 nodes (rounded up) at the top, N12 nodes in the second layer and so on until the lowest layer of a final node.

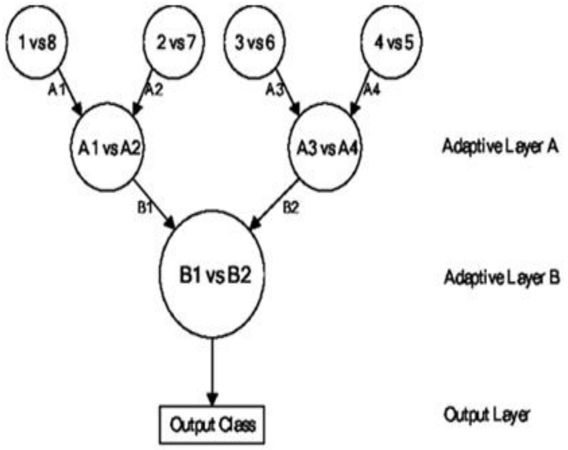

Fig.3. Adaptive Directed Acyclic Graph[1].

To classify using the ADAG, starting at the top level, the binary function at each node is evaluated. The node is then exited via the outgoing edge with a message of the preferred class. In each round, the number of candidate classes is reduced by half[1]. Based on the preferred classes from its parent, the binary function of the next-level node is chosen. The reduction process continues until reaching the final node at the lowest level. The value of the decision function is the value associated with the message from the final leaf node nodes. The ADAG requires only N-l decision nodes to be evaluated in order to derive an answer[1].

Table 2 Comparatively Study of LPC,SVM,RastaAnd MFCC:

Techniques	Features	Merits	Limititions
1.Linear Predictive Coding(LPC)	Very useful and practical approach for speech signal at high frequency.	It is readable ,robust and accurate technique for proving parameters that describe linear system representing vocal tract.[4]	Cannot distinguish words with same vowel sounds.[5]
2.SVM(Support Vector Machine)	For classification of input command SVM is most likely method.[9]	It supports the support vector in a pattern recognition problem of two classes.[9]	Only contributed for binary classification .[9]
2 a.Linearly Separable	It is used for generalised classification problem of two class.[14]	It removes the negative points from the problem.[14]	It produces same solution like separable case. [14]

2 b.Nonlinearly Separable	It is used for preventing from vector loss.[14]	For measuring input "Higher Dimensional Feature Space "is used. [14]	Some constraints must be satisfied by adjustable kernel like the implementation of exponential radial basis function. [14]
3.Relatively Spectral Filtering(RASTA)	Widely used popular technique for input signals that have environmental noise or speech with noisy disturbance .[8]	Highly reliable and robust useful for capturing frequency with low modulation. [8]	It needs to compare with PLP for better performance.[1]
4.Mel Frequency Cepstral coefficient(MFCC)	Cepstral domain based analysis technique for speech recognition that matches the human auditory system. [8]	Very efficient and accurate with low complexity. [8]	Background noise can affect and hamper the quality of MFCC result.

IV. CONCLUSION

We conclude Speech recognition system using Support Vector Machine. Feature are extracted by using Linear Predictive Coding and those models trained successfully by using Support Vector Machine. SVM learning algorithm has been used for the recognized speech by learning from training data. Training of the system requires creating a pattern representative for the features of class using one or more patterns that correspond to speech sounds of the same class. Many models are available out of them Support Vector Machine (SVM) is widely used and accepted as it is efficient algorithm for training and recognition.

REFERENCES

[1]. Boon serm Kijsirikul and NitiwutUssivakul, "Multiclass Support Vector Machine Using Adaptive Directed Acyclic Graph",In Proceedings of International Joint Conference on Neural Networks, IEEE 2002, pp 980-985.
[2]. Xin Dong, Wu Zhaohui and Pan Yunhe, "A New Multi-Class Support Vector Machines", IEEE 2001, pp 1673-1676.
[3]. Chen Junil, Jiao Licheng, "Classification Mechanism of Support Vector Machines", IEEE 2000.
[4]. K.P. Bennett, J.A. Blue, "A Support Vector Machine Approach to Decision Trees", IEEE 1998, pp 2396-2401.
[5]. C. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition",Knowledge Discovery and Data Mining, 2(2), 1998.
[6]. Hiroshi Shimodaira, et al, "Support Vector Machine with Dynamic Time-Alignment Kernel for Speech Recognition", Eurospeech 2001, Scandinavia.
[7]. Kartiki Gupta1 ,Divya Gupta2" An analysis on LPC, RASTA and MFCC techniques in Automatic Speech Recognition System"2016 IEEE
[8]. Pratik K. Kurzekar1 ,Ratnadeep R. Deshmukh 2, Vishal B. Waghmare 3 , Pukhraj P. Shrishrimal 4 "A Comparative Study of Feature Extraction Techniques for Speech Recognition System" M.Tech Student, Department of Computer Science and Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad (MS), India.
[9]. S. Gunn, "Support vector machines for classification and regression". Technical Report, University of Southampton Image Speech and Intelligent Systems Group, 1997.
[10]. Singhal Shweta, Dubey Rajesh Kumar, "Automatic Speech Recognition For Connected Words Using DTW/HMM For English/Hindi Languages ", IEEE 2015 .
[11]. Patil Mahesh K, Admuthe, Prof L S, Zirmite, Prashant P, "Isolated Digit Recognition Using Ear Microphone Data Using MFCC, VQ & HMM", IJETT, ,Jun-2014, VOL.12, ISSUE 7, pp (322-325).
[12]. Nandyala, Siva Prasad, "Real Time Isolated Word Speech Recognition System For Human Computer Interaction" International Journal of Computer Applications, NOV-2010, VOL.12, ISSUE2.
[13]. MengdiYuet, Ling Chen, Jie Zhang, Hong Liu, "Speaker Age Recognition Based On Isolated Words 282-286).By Using SVM", IEEE, 2014, pp(282-286) .
[14]. Rajkumar Bhosale, Dr. Narendra Chaudhari. Real Time Enhanced Speech Recognition Technique to Operate Computer System Using SVM 2017.