Reduction of Time for Speech Recognition System by Shrinking Adaptive Layers of ADAG for Real Time System using ADAT SVM

Rajkumar S. Bhosale, Narendra S. Chaudhari

Abstract*: Recently speech recognition becomes very major area for large vocabulary real time applications. In the existing work research was formulated for large number of words spoken by same speaker using Adaptive Directed Acyclic Graph (ADAG) with support vector machines. In today's era the emphasis is given to processing large vocabulary data process and recognition. In existing system, when spoken words are recognizing by number of adaptive layers causes increase in testing time for recognition of data occur. The aim of the proposed Reduced Adaptive Directed Acyclic technique (RADAT) is to develop a system to recognize test word in less amount of time than existing System [1]. In our current system it is possible to remove unnecessary finding distance of test word with number of times with training word. The proposed RADAT system handles this by applying reduction in number of adaptive layers to recognize any testing word and which result reduction in time. The experiment results of current system reduce time complexity without loss of recognition accuracy for any data of speaker dependent system. The results are obtained for various dataset like large vocabulary speech record as well for small size data to large vocabulary dataset. In our paper we use same feature extracted samples for training and testing. This method is work for various speech dataset like machine operation command in English or required language to operate computer system in real time application as well applied for offline application such as start, logoff, login, notepad, open, close, shutdown etc [5].

Keywords: Automatic Speech Recognition, ADAG, classification, DTA, Kernel, LPC, pre-processing, RADAT, SVM.

I. INTRODUCTION

Today's group of people interest is more towards hand free processing of acoustic signal with any hand held device like Smartphone, tablet etc.. So to perform dealing out recognition of speech commands or large vocabulary data is become crucial area of research. A support vector machine (SVM) is talented classifier among various classifiers for recognition of acoustic data. From many decades Researchers are using SVM because of its wonderful concert [6]. The SVM Classifier is both practically and theoretically considered as a binary classifier to solve multiclass problem. One-against-Rest SVM is one of the regular approach in recognizing speech dataset. It is used to separates binary SVM one from all others. The method is One-against-One (1-v-1) is also one of the trendy approach for classification in SVM. By the way here classification techniques it is conclude that, the max win algorithm One-against-One (1-v-1) is better for performance instead of using One-against-Rest (1-v-R) [1]. The 1-v-1 methods combined with ADAG SVM algorithm for garter evaluation of accuracy for training. Recognizing speech from large vocabulary data is one of the energetic areas of interest. In existing Systems, researchers mostly try to focus on how to increase prediction accuracy of recognition of speech instead of time complexity. In this paper we perform novel architecture, named as Reduced Adaptive Directed Acyclic technique (RADAT) SVM is proposed for reducing testing time by preserving the accuracy as compared to the ADAG SVM and Max Wins algorithm [2].

Accuracy and time are two important aspects in speech recognition. Given aspects will be significant when we are working in real time systems. An ADAG is a directed graph in which nodes (support vectors) are performed as per arrangement done in reversed triangular structure. To deal with C-class problem, the system encompasses C (C-1)/2 binary classifiers [3]. The ADAGSVM has C-1 internal nodes; each of these nodes is labeled with a Boolean function. The reversed triangle consist of C/2 nodes (rounded up) at the top, C/2 nodes in the second layer and so on until the lowest layer of a final node. From literature we recognized limitations of the ADAGSVM [4]. During testing, it is point out that ADAGSVM compares testing word with long evaluation path for predicting correct class. It is observed that increase in number of class increases testing time. Our contributed RADAT algorithm reduces time of testing in large vocabulary dataset. Due to reduction in time our proposed method outperforms for real time systems. So our RADAT algorithm is time efficient than ADAG. For training and testing, we used large vocabulary data set as well as small dataset like Marathi and other digits.

The next section II tries to brief the pre processing and SVM system. Then section III describes the summary of ADAG concept pointing out some discrepancies of same. Section IV focuses on modified architecture of ADAT, which reduces time complexity without hampering accuracy. The experimental results are illustrated in section V and finally conclusion of paper is summarized in section VI.

Revised Manuscript Received on October 15, 2019
***** Correspondence Author
Rajkumar S. Bhosale *, Amrutvahini College of Engineering, Sangamner, Maharashtra, India. bhos_raj@rediffmail.com
Narendra S. Chaudhari, India Department of Computer Science, Indian Institute of Technology, Indore, MadhyaPradesh. nsc@iiti.ac.in

II. Preprocessing Technique And Support Vector Machine Classification

A. Speech Separations by End point Detection

The speech recognition first step starts from speech separation. The unwanted parts of signal are removed from start and end of signal. If continuous speech signals are applied for recognition then non-speech part of signal must remove from the given input signal [7].

In our experiments, the "Sound forge" tool is used for end point detection of input continuous wave signal as shown in fig.1. To isolate continuous speech signals this tool is used. 16-bit depth is used for recording of samples. The 44100 Hz rate of sample along with mono channel is used for recording of acoustic signal. Average noise is calculated by removing silent signals from recorded wave. We are using this methodology for online and offline large vocabulary dataset. Such system used to remove unwanted noise from given input signal.

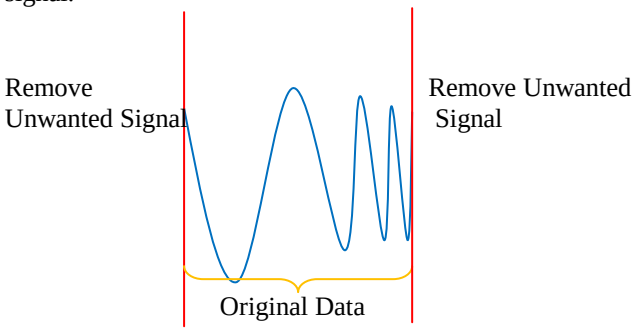

Fig.1. End Point Detection System.

B. Linear Predictive Coding Model for coefficient extraction

Linear Predictive Coding (LPC) is feature extraction method used mostly for audio signal processing. It is used to compute the coefficients of audio input signal. These signals are error free signals. LPC order along with 256 windowing size is used to find vector coefficients.

In proposed architecture of feature extraction, sample of speech at t time, $c(t)$ linear combination is calculated in this stage.

$$c(t) \approx b_1 c(t-1) + b_2 c(t-2) + \ldots\ldots + b_p c(t-l) \quad (1)$$

Pre-emphasis is the first step in analysis of LPC. After pre-emphasis frame blocking is performed. Frame blocking is followed by windowing. Autocorrelation and translation to cepstral coefficients is performed.

C. Dynamic Time Alignment (DTA) SVM

The length of Sample vector is get compared with applied input speech signal. Dynamic time alignment technique is used to equalize the size of variable length feature vectors. The Dynamic Time Alignment SVM (DTAKSVM) tool is used in kernel role [9].

Consider two vectors $\overline{L}$ and $\overline{M}$. Without loss of generality, short length vector is selected as $\overline{L}$. $\overline{L}$ chain first vector is 20. Over a window, Find 20 matching vectors in the $\overline{M}$ vector. The p size unit window is fixed. The window begins as of v1 vector of $\overline{M}$. Consider vk is finest match where $1 \le k \le p$. Repeat this step subsequently for v2 vector feature to k index.With above feature extraction, we get $\overline{L}$ and $\overline{M}^{new}$ as two analogous piece vectors.

D. Support Vector Machines

Any Support Vector machine is promising binary classifier for recognition of speech with considerable accuracy [11]. The basic idea of Support vector Machine is illustrated in next subsections.

Linear Separable Concept

The two-class generalized categorization problem is described below [3].

Consider D as a data input set of N samples. It consists of samples $\langle x_1, y_1 \rangle, \langle x_2, y_2 \rangle, \ldots \langle x_N, y_N \rangle$. For training x_i example sample with K length is used, by means of elements $x_i = \langle x_1, x_2, \ldots x_K \rangle$, Consider $y_i \in \{-1, +1\}$ will be a target value. The Classifier is created with $f(x)$ decision function.

$$f(x_i) = y_i, \forall \langle x\ , y_i \rangle \in D$$

.

The equation 2 and 3 are used to take apart the vectors with the use of hyper plane and removes pessimistic points from vectors.

$$x_i \cdot w + c \ge +1 \quad \text{for } y_i = +1 \quad (2)$$

$$x_i \cdot w + c \le -1 \quad \text{for } y_i = -1 \quad (3)$$

***Non-Linear Separable* Concepts**

One of the disadvantages of linear separable case is that loss of data points while separating the vectors. So a novel method is implemented for data point disjointing in high dimensional space [8].In non-linear separable data, input data is mapped using higher dimensional attribute space. Various kernel functions are used for matching support vector data given as follows [10].

Thus, input space attribute mapping can be completed by the inner product substitute as shown in equation:

$$x_i \cdot x_j \longrightarrow \Phi(x_i) \cdot \Phi(x_j) \quad (4)$$

So for separation of feature vector and mapping is done by using non-linear concept.

Calculation of each Φ is not needed explicitly. So, K (xi, xj) function is used, which compute feature space the inner product as follows.

$$K(x_i \cdot x_j) = \Phi(x_i) \cdot \Phi(x_j) \quad (5)$$

In subsequent step, inner product calculation between two distinct vectors is done. Optimization of the kernel problem is performed using the constant variable. Equation 5 is for optimization problem. The equation 6 is the functional representation of kernel and is used to solve the problem of high dimensional space. In equation 6, constant α is used.

$$\max_{\alpha} W(\alpha) = \max_{\alpha} -\frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\alpha_i\alpha_j b_i b_j K(x_i \cdot x_j) + \sum_{i=1}^{l}\alpha_i \quad (6)$$

Exponential radial basis function is implemented to satisfy few constrains as shown in equation 7.

$$K(a_i, a_j) = e^{-|a_i - a_j|/2\sigma^2} \quad (7)$$

In this work, σ is replacement of c..

The different user defined values are experimented for σ to give best performance of speech recognition.

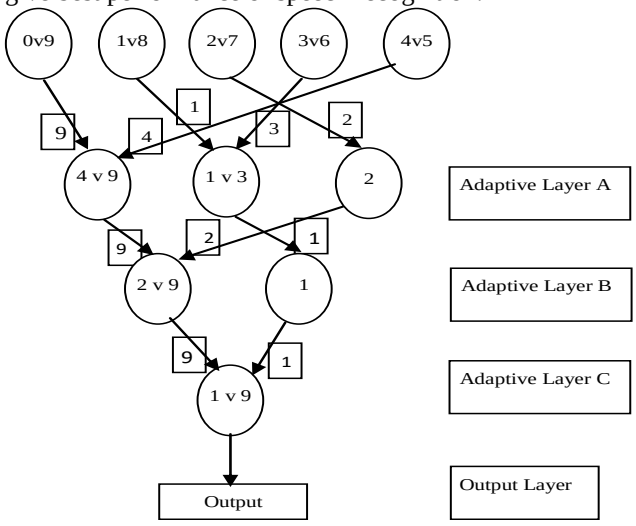

Fig. 2.The structure of an ADAG classifier for 10 class problem

III. ADAPTIVE DIRECTED ACYCLIC GRAPH SVM

The reverse triangular flow is adopted in ADAG SVM which is directed acyclic algorithm. As demonstrated in fig.2.

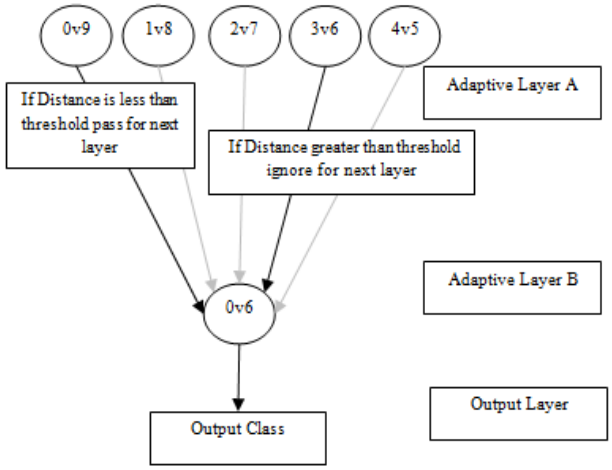

Fig.3. Proposed structure of an RADATSVM classifier for 10 class problem

C-Class problem use $(C-1)/2$ classifiers with $C-1$ internal nodes.

The low computation time is one of the characteristics of Multi-class SVM.Proposed RADAT is better time reduction than Multiclass SVM. It has more accuracy than max Wins method. For training phase, it encompasses one vs. one classifier. One binary SVM for each couple of classes is formed. For C support vectors, $(C-1)/2$ binary SVMs are created. The reverse triangle is formed by arranging nodes like $C/2$ at top, second layer C/22 and continued to only one node in last level [1].

Binary decision function from top to bottom is evaluated with the use of reverse triangle concept. At every level, classes are reduced by half number. Parent nodes are evaluated to select class in next layer. This reduction process in each level is done using binary function. It continues to the lowermost stage.

IV. REDUCED ADAPTIVE DIRECTED ACYCLIC TECHNIQUES SVM (RADATSVM)

In this paper we are going to propose novel approach of speech recognition which reduces time for testing. The plan of ADAG and RADATSVM is support for large vocabulary speech signal. In this paper, time complexity decreases for small dataset but same system can be applicable for large vocabulary dataset and vowel dataset. In Adaptive Directed Acyclic Graph, every plane is checked repeatedly for testing the speech data. As shown in Fig 2, every layer is having fixed number of internal nodes. In experiment, we have observed that though the test data is away from both planes, we are considering one minimum distance plane. This limitation of ADAG is overcome in our proposed System. In this system instead of forcefully matching testing word, we are omitting the plane which is long distance from test word. If the test word is having less than threshold distance from any of the two planes then only minimum distance plane is considered in next adaptive layer. This novel method not only reduces the number of nodes but also reduces the number of adaptive layers. In this system, we use threshold value for the distance between test word and the plane. This threshold value depends on the support vectors of training data and support vectors in test word. If the distance of applied testing data and trained samples are more than threshold then that particular node will be ignore for next adaptive layer for comparisons. So this will reduce the number of nodes in the subsequent layers. Decreasing number of nodes reduces time complexity of the speech recognition.

We are illustrating our proposed method of RADAT with simple example of Marathi digit recognition. As shown in the Fig 3, when we are checking distance between test word and node 0v9, if the distance is less than threshold value then output class is migrated in the next layer. This is shown by dark line. We are obtaining same result with node 3v8.So the output class of node 3v8 is carried in next adaptive layer. But nodes 1v8, 2v7, and 4v5 are having more distance than the threshold value with testing data. Therefore their output

class has been ignored in next adaptive layer and it shown by faint lines. This leads to reduction of nodes in subsequent layers. Reduction in nodes in subsequent layers largely affects to reduction in testing time.

If we look to the proposed system architecture it reduces the time for testing dataset by preserving accuracy of the existing system. This system maintains the same result as per previous system with slightly reducing time complexity.

V. RESULTS

Before The performance of our modified architecture RADAT SVM is tested for large vocabulary English word dataset. Here RADAT system gives better performance in minimum number of adaptive layers. So instead of comparing test word with number of planes with training samples, the proposed system will consider minimum number of planes in successive layers.. If the large vocabulary dataset is present for training and testing, then system will take less time for testing than ADAG SVM.

The experiments are carried out for large vocabulary dataset as well as Thai vowel dataset. 10 samples are taken of each word. First of all the training is given to these 10 samples of each word. The features are extracted and support vectors are calculated. SVM binary classifier is used for classification of test data. The training and testing data is of same speaker. This RADAT system uses speaker dependent data. In our experiment, the test data gives best results for threshold value of 7.0000. The experimental results show that without reducing accuracy, RADAT reduces the time for testing. The final result has been summarised and are shown in table 1 as follows.

Table- II: Results of Proposed RADATSVM algorithm with existing ADAGSVM for large vocabulary dataset and standard Thai Vowel dataset

System	Average	Testing Time(seconds)
ADAG	97	15.2
Proposed RADAT	**97**	**14.6**

In this experiment standard Thai vowel dataset and large vocabulary considered for training and testing and then average accuracy is calculated using existing and proposed system. The RADAT technique shows better result in term of testing time for the input speech signal and ADAG SVM takes more time for testing the input speech signal. We carried out the experiments using MATLAB software.

When experiments are carried out for small dataset, the testing time reduction is less. For large vocabulary dataset and Thai Vowel dataset, reduce more testing time. Real time system requires minimum time complexity. By looking our new approach of reduction in time of testing speech data, we can use this system for real time application without any problem.

VI. CONCLUSION

ADAG SVM is one of the promising binary classifier for speech recognition. In speech recognition accuracy and time are important parameters of efficiency. In this paper, we proposed variant of ADAGSVM by reducing number of planes in recognition. It reduces the testing time of speech data without compromising accuracy of recognition. As the existing system unnecessary transfers the output class of node in successive layers, the proposed system omits it by comparing distance of test data with given nodes. This new approach not only reduces the number of nodes in adaptive layers but also causes reduction of adaptive layers. The experiment shows that there is considerable reduction in time required for testing as compared to the previous systems. Now we can say that RADAT become one of the fastest algorithms for testing of words or any other dataset. Such a system is applicable for different real time application.

ACKNOWLEDGMENT

The moral support and advice are given by my guide Respected Dr. Narendra Chaudhari throughout my research work. Due to his helpful nature, I am always thankful to him.

Also I would like to thanks' my wife Archana Bhosale without her it was not possible for me to complete my work in a given time.

REFERENCES

1. Boonserm Kijsirikul and Nitiwut Ussivakul, "Multiclass Support Vector Machine Using Adaptive Directed Acyclic Graph", 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", International Conference on Signal Processing Proceedings. 16th World Computer Congress 2000 IEEE 2000.
4. Dişken, Gökay, et al. "A review on feature extraction for speaker recognition under degraded conditions." IETE Technical Review 34.3 ,2017,pp 321-332.
5. Rajkumar Bhosale, Narendra Chaudhari "Real times enhance speech recognition technique to operate computer system using SVM" 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT).
6. Sarma, Biswajit Dev, and SR Mahadeva Prasanna. "Acoustic–phonetic analysis for speech recognition: A review." IETE Technical Review 35.3 (2018): pp 305-327.
7. K.P. Bennett, J.A. Blue. "A Support Vector Machine Approach to Decision Trees", IEEE 1998, pp 2396-2401.
8. C. J. C. Burges. "A Tutorial on Support Vector Machines for Pattern Recognition". Knowledge Discovery and Data Mining, 2(2), 1998.
9. Hiroshi Shimodaira, et al, "Support Vector Machine with Dynamic Time-Alignment Kernel for Speech Recognition", Eurospeech 2001, Scandinavia.
10. Nello Cristinini and Jhon Shawe-Taylor. "An Introduction to Support Vector Machine and other kernel-based learning methods", Cambridge University Press 2000.
11. S. Gunn. "Support vector machines for classification and regression". Technical Report, University of Southampton Image Speech and Intelligent Systems Group, 1997.

AUTHORS PROFILE

Bhosale R.S. received the B.E. Degree in the Computer science and Engineering from Dr. Babasaheb Marathwada University, Aurangabad and completed his post graduation in Computer science and Engineering from Swami Ramananda Teerth Marathwada University, Nanded. He is currently pursuing his Ph.D. in Computer science and Engineering from Sant Gadge Baba Amravati University, Amravati Maharashtra, India.

Narendra S. Chaudhari is currently with Indian Institute of Technology, Indore since Aug 2009 as a Professor of Computer Science and Engineering. He was Director of Visvesvaraya National Institute of Technology (VNIT) Nagpur (M.S.) since June 2013 till June 2018 (five year term). Since Sept 2010 to June 2013, he was Dean - Research and Development (Dean - R&D), IIT Indore. Since 2001 to July 2009, he was with the School of Computer Engineering, Nanyang Technological University (NTU), Singapore. From 1990 to 2001, he was Professor of Computer Science Science in M.Sc. DRDO Program, Ministry of Defense (Govt. of India) at Devi Ahilya University, Indore and from 1988 to 1990.