A Comparative study of HMM and SVM in Malayalam Digit Recognition

Harsha K M[#1], Facila Chinchu O[#1], Cini Kurian[#1], Kannan Balakrishnan[#1]

[#1]*Department of Computer Applications, Cochin University of Science and Technology*
Kochi, Kerala, India

[1]916harsha@gmail.com
[2]facilabani@gmail.com
[3]cinikurian@gmail.com
[4]bkannan@cusat.ac.in

***Abstract*— A primary medium for the human beings to communicate through language is Speech. Automatic Speech Recognition is wide spread today. Recognizing single digits is vital to a number of applications such as voice dialling of telephone numbers, automatic data entry, credit card entry, PIN (personal identification number) entry, entry of access codes for transactions, etc. In this paper we present a comparative study of SVM (Support Vector Machine) and HMM (Hidden Markov Model) to recognize and identify the digits used in Malayalam speech.**

***Keywords*— SVM (Support Vector Machine), HMM (Hidden Markov Model), MFCC (Mel Frequency CepstralCoefficients), Speech Recognition.**

I. INTRODUCTION

Speech is the physical production of sound using respiratory and phonatory Systems. Speech Recognition converts spoken words to text by using computers without targeting a single speaker. That is Speech Recognition is program which identifies the words and phrases in spoken language and convert them to machine understandable form. The applications of speech recognition include speech to text, voice dialling, call routing etc.

Speech is a very natural phenomena and very difficult to control because it can vary in terms of gender, accent, pronunciations, pitch, volume and speed. Some times during transmission it is distorted by noise and echoes. So speech Recognition is a very complex problem. Another problem with the speech recognition [1][2][3]system is Out of Vocabulary that is the word spoken during speech is not in the dictionary.

Malayalam is one among the 22 languages spoken in India with about 38 million speakers. Malayalam belongs to the Dravidian family of languages and is one of the four major languages of this family with a rich literary tradition. The majority of Malayalam speakers live in the Kerala, one of the southern states of India and in the union territory of Lakshadweep. There are 37 consonants and 16 vowels in the language. It is a syllable based language and written with syllabic alphabet in which all consonants have an inherent vowel /a/. There are different spoken forms in Malayalam even though the literary dialect throughout Kerala is almost uniform.

Speech Recognition of Malayalam Digit is a major research topic and not much work is done in this area based on HMM and SVM except “Isolated Malayalam Digit Recognition Using Support Vector Machines”[14],’’Speech Recognition of Malayalam Numbers’’ [13], and ”Perceptual Linear Predictive Cepstral Coefficient for Malayalam Isolated Digit Recognition”[15].

“Speech Recognition of Malayalam Numbers” system employs Mel frequency cepstrum coefficient (MFCC) as feature for signal processing and Hidden Markov model (HMM) for recognition. In this work, a public domain speech recognition development toolkit (CMU sphinx [16] is used for training and decoding. “Isolated Malayalam Digit Recognition Using Support

Vector Machines" is created by using Mel Frequency Cepstral Coefficients (MFCC) and Support Vector Machines (SVM).

II. METHODOLOGY

The basic issue in speech recognition is dealing with two kinds of variability: acoustic and temporal [4]. Acoustic variability covers different accents, pronunciation, pitches, volume, and so on, while temporal variability covers different speaking rates. Development of a better acoustic modelling is the main task in Speech recognition research.

Consequently the need for a discriminative classifier with good generalization and convergence property has necessitated the evolution of new machine learning paradigm called SVM classifier. Support Vector machine which is kernel based machine learning tool which has shown its powerful performance in classification problems ranging from particle identification, face identification, text categorization, engine knock detection, to bioinformatics.SVM is fundamentally a linear classifier, however the powerful utilization of kernel functions allows SVM to function as a non linear classifier. Hence it can work on data having high dimensionality.

support vector machines (SVMs) are learning models that analyse data and recognize patterns, used for classification and regression analysis. The underlying concept behind an SVM is structural risk minimization [8]. A learning machine is chosen that minimizes the upper bound on the risk (or test error), which is a good measure of the generalizability of the machine. This is estimated as the ratio of misclassified vectors over the total number of training vectors when using a "leave-one-out" method [9]. It can be shown that this is equal to the ratio of expected number of support vectors to the total number o training vectors.

Support vector machine (SVM) is the state-of-the-art classifiers derived from statistical learning theory. It is supervised learning algorithm with better generalized properties and with limited number of training patterns [10].SVM is introduced for classifying linearly separable classes of objects. SVM resolves the classification problems by separating the data into two categories by using an n dimensional hyper plane. SVM determines the hyper plane that maximizes the margin between classes. For any particular set of two classes of objects, an SVM finds the unique hyper plane having the maximum margin. SVM represents the classified outputs as support vectors that determine the maximum margin hyper plane. This maximum margin solution enable SVM to outperform compared to other nonlinear classifiers, particularly in noisy environments. Moreover , SVM can also be used to classify classes that cannot be classified with a linear classifier. A unique property of SVMs is that they simultaneously minimize the empirical classification error and maximize the geometric margin; hence it is also known as maximum margin classifier [11]. A support vector machine forpattern classification is built by mapping the input pattern xinto a high-dimensional feature vector v using a non lineartransformation f(x), and by constructing an optimal hyperplane in the feature space. Non linear transformation f(x) should be such that the pattern classes are linearly separable in feature space [12]. A function called 'kernel' is used to map the data from input space to feature space .

The main draw backs are : i) SVMs, being a static classifiers, adaptation of the variability of duration of speech utterances is very difficult ; ii) ASR faces multiclass issues while SVMs are originally formulated as a binary classifier and iii) SVM training algorithms are very weak in managing huge databases typically used in ASR.

HMM is a statistical model in which it is assumed to be in a Markov process with unknown parameters, the challenge is to find all the appropriate hidden parameters from the observable states. Hence it can be considered as the simplest dynamic Bayesian network. In a regular Markov model, the state is directly visible to the observer, and therefore the state transition probabilities are the only parameters. However, in a Hidden Markov model, the state is not directly visible (so-called hidden), while the variables influenced by the states are visible. Each state has a probability distribution over the output.

Acoustic modelling uses some probability measures to realize the sounds using statistical models. Generally most of the current speech recognition systems, the acoustic modelling components of the recognizer are almost exclusively based on HMM [5][6][7]. HMM provides a statistical framework for modeling speech patterns using a Markov process [6] that can be represented as a state machine. The temporal evolution of speech is modelled by the Markov process in which each state is connected by transitions, arranged into a strict hierarchy of phones, words and sentences. The probability distribution associated with each state in an HMM, models the variability which occurs in speech across speakers or even different speech contexts.

III. Feature Extraction

For recognition of speech, the signals have to be represented with some specific features. MFCC is the well known popular method of feature extraction. To capture the phonetically important characteristics of speech, signal is expressed in Mel-Frequency Scale [12]. This scale has a linearly frequency spacing below 1000Hz and a logarithmic spacing above 1000Hz. MFCCs are less susceptible to the physical conditions of the speakers' vocal cord [17], compared to the speech wave forms. The block diagram of the feature extraction process is shown in figure 2.

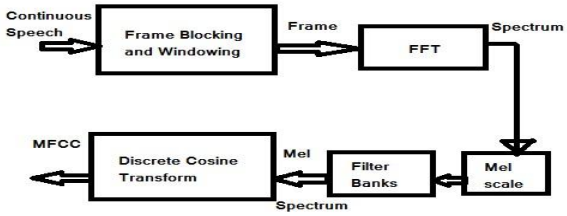

Figure2 : Steps involved in the computation of MFCC

IV. Database

In "Isolated Malayalam Digit Recognition Using support Vector Machines"[14], and in ''Speech Recognition of Malayalam Numbers'' [13], , the database used contains 100 isolated spoken words from 10 speakers. Each speaker uttered numbers from zero to nine separately. The recording was done in normal office environment with a high quality microphone with 16 kHz sampling frequency and quantized at 16 bit.

.

V. Design and Development of System

The structure of a standard Speech Recognition system[13] is depicted in the figure1

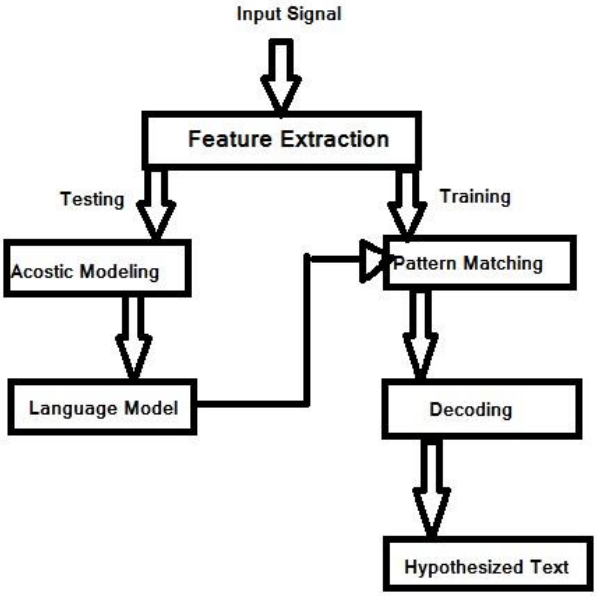

Figure 1: Speech Recognition system

The goal of acoustic modelling is to characterize the statistical variability of the feature set determined above for each of the basic sounds (or words) of the language. The purpose of the language model, or grammar, is to provide a task syntax that defines acceptable spoken input sentences and enables the computation of the probability of the word string, given the language model . The job of the pattern matching module is to combine information from the acoustic model, the language model to find the optimal word sequence.

VI. Results And Discussions

Digit recognition system for Malayalam language is implemented by HMM [13] and SVM[14]. In both the system MFCC is used for feature extraction. Recognition system for Malayalam Digits using Hidden Markov Models gave an accuracy of 98.5% while that of SVM gave and accuracy of 97.6%. Both methods show very good accuracy. In SVM based application we use only first 20 features but in HMM based application we uses 39 dimensional vector from each frame of 25 milliseconds. The higher accuracy of HMM based system may be due to the large number of features. It can be concluded that both methods shows reasonable well accuracy in the task of digit recognition.

VII. Conclusions

Spoken number recognition system provides a user-friendly interface for feeding numeric data into computers. This paper provide a comparative study of SVM and HMM for speech recognition especially on Malayalam digits. From this study we identified that HMM and SVM provide reasonable good accuracy for Malayalam digit recognition. The futures prospects of this work are to incorporate more number of feature vectors for SVM application.

VIII. References

[1]B. Plannerer," An Introduction to Speech Recognition", c 2001,2002,2003 Bernd Plannerer, Munich, Germany plannerer@ieee.org

[2]Lawrence Rabiner and Biing-Hwang Juang," Fundamentals of Speech Recognition", Prentice Hall, New Jersey, 1993 (or later editions)

[3] Xuedong Huang, Alex Acero, and Hsiao-Wuen Hon," Spoken Language Processing: A Guide to Theory, Algorithm, and System Development", Prentice Hall, New Jersey, 2001 (or later editions)

[4] Saumudravijaya K, "Hindi Speech Recognition" (2001), J. Acoustic Society India, 29(1), pp 385-395.

[5] Dimov, D., and Azamanov, I. (2005)." Experimental specifics of using HMM in isolated word Speech recognition" .International Conference on Computer Systems and Technologies – CompSysTech '2005'.

[6] F.Felinek, "Statistical Methods for Speech recognition" MIT Press, Cambridge, Massachusetts, USA,1997.

[7] Rabiner L R, "A Tutorial on Hidden Markov Models and selected Application s in Speech Recognition" Proc. IEEE, vol. 77, 1989, pp. 257 – 286.

[8]. V. Vapnik, The Nature of Statistical Learning Theory, Springer-Verlag, New York, NY,USA,1995.

[9]. B. Schölkopf, Support Vector Learning., Ph.D Thesis, R. Oldenbourg Verlag Publications, Munich, Germany, 1997.

[10] A. Ganapathiraju, J. Hamaker, and J. Picone, "Hybrid SVM/HMM architectures for speech recognition," in Proc. ICSLP, Beijing, 2000.

[11] C. Chandra Sekhar, W.F.Lee, K. Takeda and F. Itakura Acoustic Modeling of Sub word Units Using Support Vector Machines, Workshop on Spoken Language Processing , TIFR, Mumbai, India.

[12] S.S Stevens and J. Volkman (1940), " The relation of pitch to Frequency", American Journal of Psychology, vol. 53(3), pp 329-353..

[13] Cini Kurian, Kannan Balakrishnan," Speech Recognition of Malayalam Numbers", 978-1-4244-5612-3/09/$26.00 @2009 IEEE.

[14] Cini Kurian , Firoz Shah.A, Kannan Balakrishnan" Isolated Malayalam Digit Recognition Using Support Vector Machines". International Conference on Communication Control and Computing Technologies (ICCCCT), 2010 IEEE ,2010.

[15] Cini Kurian, Kannan Balakrishnan, "Perceptual Linear Predictive Cepstral Coefficient for Malayalam Isolated Digit Recognition". Trends in Computer Science, Engineering and Information Technology Communications in Computer and Information Science Volume 204, 2011, pp 534-541.

[16] http://cmusphinx.sourceforge.net.

[17] Rabiner L R, "A Tutorial on Hidden Markov Models and selected Application s in Speech Recognition" Proc. IEEE, vol. 77, 1989, pp. 257 – 286.