APPLICABILITY OF HIDDEN MARKOV MODEL APPROACH FOR SINHALA SPEECH RECOGNITION – A SYSTEMATIC REVIEW

L.A.D.S.A.Molligoda[1] and P.G. Wijayarathna[2]

Department of Industrial Management, Faculty of Science, University of Kelaniya.

E-mail: [1]supun95@gmail.com, [2]gamini@kln.ac.lk

ABSTRACT

Language is the humans’ most important media of communication and speech is its primary medium. Every person is familiar to his/her mother-tongue from his/her childhood. Thus it gives the ease of communication between human beings. Generally, transfer of information between human and machine is accomplished via keyboard, mouse etc., but human can speak more quickly than typing. Speech input offers high bandwidth information and relative ease of use. Here we deal with the interaction between human and machine via speech recognition. In the current context, Speech Recognition is realized by different approaches. Pattern recognition approach with hidden markov model is the most used and popular approach by researches as of now. This study was focused on Sinhala speech recognition and objective was to review on the applicability of hidden markov models as pattern recognition algorithms for Sinhala speech recognition. Applicability is measured by using inclusion and exclusion criteria and search strategy. This will include three main components; Sinhala speech recognition, Hidden Markov Model for speech recognition and Sinhala Language compatibility for HMM. Reviews based on research papers published on above domains have been taken into consideration in search of advances of the Hidden Markov Model approach. Outcomes of the research draw a curtain of various techniques on extended hidden markov models and show the need of further researches/experiments on hidden markov model variants for Sinhala speech recognition.

***Key words*:** ASR, Hidden Markov Model, Pattern Recognition

1. INTRODUCTION

From the early ages to present speech was the most powerful and common way of communication between human beings. Verbal communication may differ from the language they use. But still it is the most convenient and direct way of communication among humans. In contrast, communication between human and machine still requires an intermediate help. Most commonly we use interfaces like keyboard, mouse & screen etc., operated with the help of software. Developing a simple software interface i.e. an ASR (Automatic Speech Recognition) gives the alternative to those hardware interfaces thus enhance the human computer interaction. ASR is the process of taking utterance of speech signal as inputs and convert them into a text sequences as close as possible to the spoken data. [1] There exist many difficulties in developing ASR; different speaking styles in different languages [5] and environmental disturbances are major barriers. Hence those will lead the researchers to experiment using various algorithms and techniques.

Basically, there are three approaches to speech recognition, namely the acoustic-phonetic approach, the pattern recognition approach and the artificial intelligence approach. The acoustic-phonetic approach has not been widely used in most commercial applications. [2] A limited success has been obtained because of the lack of good knowledge of the acoustics phonetics and the related area. Pattern recognition approach has been used successfully to design a number of commercial recognition systems. This approach is a popular choice for most ASR system nowadays because it is simple and is computationally feasible to use when compared to artificial intelligence approach [3].

Recognition algorithm plays a major part in pattern recognition approach. Hidden markov model is a powerful statistical algorithm for pattern recognition. Many speech recognition studies were done by using this algorithm, but the applicability of such an algorithm for different languages has not been evaluated yet.

Sinhala is the native language of Sri Lanka and implementation of ASR for Sinhala language is an emerging research area. Therefore, the main objective of this article review and peer reviewed literature is investigating to identify, the applicability of hidden markov models as pattern recognition algorithms for Sinhala speech recognition and identify the state of the art of HMMs for speech recognition.

2. METHODOLOGY

2.1 Inclusion and exclusion criteria

The inclusion criteria for the studies included in this review were as follows: (1) article domains were related to ASR using HMMs; (2) primary focus was to understand the type of HMMs used in ASR and their consequences; (3) articles were published in a peer-reviewed journals; (4) articles were available in English.

2.2 Search strategy

An extensive initial search was done by using IEEEexplore, Research Gate and Science direct online journal databases. Search process was consisted of following phases: (1) Speech recognition using hidden markov models (2) Sinhala speech recognition (3) hidden markov models for pattern recognition

3. RESULTS AND DISCUSSION

3.1 Evolution of HMM

Hidden Markov model is a key statistical approach built around 1980s and widely used in speech recognition domain. It is a doubly stochastic process which has an underlying stochastic process that is hidden or not observable, but can be observed by another stochastic process that produces a sequence of observations. There exist two types of HMMs used in speech recognition; Discrete Hidden markov Model-DHMM and Continuous Density Hidden markov Model-CDHMM. CDHMM is more capable in modeling inter-speaker acoustic variability compared to DHMM. [4]

Most commonly used extensions to standard HMMs is to model the state-output distribution as a mixture model. In the architecture of an HMM-Based Recognizer, a single Gaussian distribution was used to model the state–output distribution. Therefore model assumes that the observed feature vectors are symmetric and unimodal. In practice this is a rare case. For example, speaker, accent, language and gender differences tend to create multiple modes in the data. To address this problem, researchers have introduced the mixture of Gaussians (Gaussian Mixture Model-GMM) to replace single Gaussian state-output distribution which is a highly flexible distribution able to model, for example, asymmetric and multi-modal distributed data. R.K Aggarwal [1] state that depending on the language, accuracy of the speech recognizer will vary on selecting different range of Gaussian mixtures. Xiaodong [6] present a novel approach which extends the conventional GMHMM by modeling state emission (mean and variance) as a polynomial function of a continuous environment dependent variable. This has been used to improve the recognition performance in noisy environments by using multi condition training.

In the training phase of speech recognition using HMM, modern practices use discriminative training techniques to optimize some classification related measures of the training data. [2] Maximum Mutual Information (MMI), Minimum Classification Error (MCE), Minimum Bayes Risk (MBR) and minimum phone error (MPE) are some examples for such techniques.

According to the Rabinar [2] there are three canonical problems to solve with HMMs:

1. Given the model parameters, compute the probability of a particular output sequence. This problem is solved by the Forward and Backward algorithm.

2. Given the model parameters, find the most likely sequence of (hidden) states which could have generated a given output sequence, solved by the Viterbi algorithm and posterior decoding.

3. Given an output sequence, find the most likely set of state transition and output probabilities. Solved by the Baum-Welch algorithm.

From the review it could be stated that HMMs are the most commonly used and recognized as the state of the art of speech recognition. Variants of HMMs can be applied to answer issues like robustness, speaker dependency and language barrier.

3.2 Sinhala Speech Recognition

In the context of Sinhala language there exist only a few research on speech recognition. Which have used the conventional HMM with the support of HTK software package. They have mainly used the Bi-gram language model with no use of extended HMM variants such as GMHMM, CDHMM etc.

3.3 Language applicability for HMM

Sinhala is a language which belongs to the Indo-Aryan branch of the Indo-European languages. According to the R.K.Agrawal [1] using high range of Gaussian Mixture models will be more applicable for Sinhala language. Also N-gram language modelling can be applied for Sinhala speech recognition.

The comparison of various speech recognition research related to Indo-Aryan languages based on recognition type, recognition technique and language models are presented below in Table 1.

4. CONCLUSION

The statistical pattern recognition approaches via HMM, has become dominant in development of ASR system. This happens due to the strength of HMM which includes the availability of its mathematical framework, simple architecture, feasibility to use and the high accuracy for its performances. Particularly for Indo-Aryan languages including Sinhala, usage of extended HMM variants are absent. CDHMM can be applied for LVCSR combined with GMMs to enhance the accuracy levels of speech recognition with robustness. The work review in this paper is a step towards the development of such type of systems.

Table 1: The comparison of various speech recognition research related to Indo-Aryan languages

Author	Year	Research Work	Recognition Type	Recognition Technique (HMM variant & Language model)	Language	Accuracy/ Findings
Thilini Nadungodage, Ruwan Weerasinghe	2011	Continuous Sinhala Speech Recognizer	Continuous/ Speaker Independent Small vocabulary	HMM + HTK Bi-gram language model	Sinhala	75% Sentence, 96% word
WGTN Amarasinghe, DDA Gamini	2012	Speaker Independent Sinhala SR for voice dialing	Isolated digits/ Speaker Independent	HMM + HTK No use of N-gram models	Sinhala	87.37% quite room, 82.19% noisy
P.G.N. Priyadarshani et, al,	2012	DTW based Speech Recognition for Isolated Sinhala Words	Isolated words/ Speaker Dependent Small vocabulary	DTW	Sinhala	89%-99% Varies
Mohit Dua, R.K.Aggraval, Virender Kadyan, Shelza Dua	2012	Punjabi ASR using HTK	Isolated/ Speaker Independent	HMM + HTK Gaussian Mixture Model	Panjabi	94.08%
Ting Chee Ming	2007	Malay continuous speech recognition using CDHMM	Large Vocabulary continuous SR - LVCSR/ Speaker Independent	CDHMM Baum-Welch and Viterbi/Segmental K-mean uni-gram and bi-gram models	Malay	78.75%
Zhao Lishuang , Han Zhiyan	2010	Speech Recognition System Based on Integrating feature and HMM	Large vocabulary Speaker independent Vowels	Genetic Algorithm + HMM	Chinese	High Accuracy
R. Thangarajan, A.M. Natarajan and M. Selvam	2008	Phoneme Based Approach in Medium Vocabulary Continuous Speech Recognition in Tamil language	Medium vocabulary Speaker independent phonemes	HMM Baum-Welch training CMU Sphinx-4	Tamil	93.07% triphone model is best suited for LVCSR

Javed Ashraf , Dr Naveed Iqbal, Naveed Sarfraz Khattak, Ather Mohsin Zaidi	2010	Speaker Independent Urdu Speech Recognition Using HMM	Small Size Vocabulary Speaker Independent, Isolated words	HMM	Urdu	Little variation in WER for new speakers
R.K. Aggarwal and M. Dave	2011	Using Gaussian Mixtures for Hindi Speech Recognition System	Small size Speaker independent Continues	HMM +Gaussian Mixture	Hindi	Low range of Gaussian mixtures for Indo-Aryan languages
M. Kalamani, S. valarmathi, M. Krishnamoorthi	2014	Hybrid modelling algorithm for continuous Tamil Speech Recognition	Medium Size, Continues, Speaker Independent	HMM, Expectation Maximization GMM	Tamil	4% Word error rate

5. REFERENCES

[1] R. K. Aggarwal and M. Dave *“Acoustic modelling problem for automatic speech recognition system: conventional methods (Part 1)”* International journal Speech Technology, Springer, Vol.14, issue 2, 2011.

[2] L. R. Rabiner, *“A tutorial on hidden Markov models and selected applications in speech recognition”* Proc.IEEE77 (2):257-286.February 1989.

[3] M. A. Anusuya, S. K. Katti, *“Speech Recognition by Machine: A Review”*, International Journal of Computer Science and Information Security (IJCSIS), Vol. 6, No. 3, 189-196, 2009

[4] J. F. Mark J. F. Gales, Katherine M. Knill, et.al., *“State-Based Gaussian Selection in Large Vocabulary Continuous Speech Recognition Using HMMs”*, IEEE Transactions on Speech and Audio Processing, Vol. 7,No. 2, March 1999.

[5] M. Weintraub et al., *“Linguistic constraints in hidden markov model based speech recognition”*, Proc.ICASSP, pp.699-702, 1989.

[6] X. Lui, et.al. *“A study of variable parameter Gaussian mixture HMM modeling from Noisy speech recognition”*, IEEE Transactions on Audio, Speech and Language processing, Vol.15, No.1, Jan. 2007.