Review Of The Language Modeling Techniques Used In Automatic Malayalam Speech Recognition

Manju G[1], Usha K[2]
[1](Department Of Computer Science, Govt. College, Ambalapuzha, University Of Kerala, India)
[2](Department Of Statistics, Govt. College, Ambalapuzha, University Of Kerala, India)

Abstract:
Malayalam is a low resource language spoken by the people of Kerala, a state of southern India. There is a rapid advancement in the ASR since the last few years due to the emerging of advanced machine learning techniques and tools. Being a regional language Malayalam speech recognition still faces major challenges like lack of annotated speech corpus and language models. This paper discusses major research works and language modeling methods in the area of Automatic Speech Recognition of Malayalam language.

Background*: The purpose of the paper is to review recent advancements in Automatic Speech Recognition of Malayalam language, with a specific focus on acoustic modeling techniques used in various research works. The aim the paper is to explore the frameworks used in ASR and identify the scope for future research in this area.*

Materials and Methods: *In traditional ASR systems the likelihood of the phonemes is computed using either a generative model or discriminative model to decode the speech signal features into to text. The End to End speech models uses raw speech signal as input and generate conditional probability of phoneme class*

Conclusion: *It is identified that only a few works were done in this area continuous speech recognition and there is scope for using advanced tools of machine learning for building a large language model for Malayalam- a low resource language.*

Keywords*: Automatic Speech Recognition; Malayalam Language Model; Generative model; Discriminative model; End to End Model*

Date Of Submission: 29-04-2024 Date Of Acceptance: 09-05-2024

I. Introduction

Speech is the most natural form of communication so that even a non-expert can use a system without much training. Speech is very much faster than other modality which in turn reduces the response time. Speech itself contains many information such as age, gender, prosody, rhythm,, accent, emotion, health and speaker specific features etc [1] which are not contained in other mode of communication. Hand free aspect of speech helps to incorporate multiple modalities for communication and speech recognition can be applied in speech enabled services and applications. Automatic Speech Recognition is the process of converting speech signals into written text where the acoustic features of the signal is mapped into building units of speech such as phonemes, words, sentences etc.

ASR systems are classified as isolated word, connected word, continuous word and spontaneous word recognition based on the type of utterances Building ASR system for continuous word recognition is more difficult than isolated and connected word recognition [2]. ASR system for spontaneous speech recognition is the most complex in terms of processing and building the language model because of the unpredictable, inconsistent and non-domain specific nature of spontaneous speech. The development of machine learning techniques has led to a rapid advancement in the field of Automatic Speech Recognition (ASR) in the recent years. The high proportion of these achievements is constrained to languages having huge volume of annotated speech corpus. But low resource languages faces challenges in adopting the same methodology used for languages having massive dataset for training and testing the ASR model. Besides, the low resource languages face challenges like code switching, less fluent native speakers for data collection, and too many accents etc. [3]

II. Malayalam – A Low Resource Language

Malayalam is a morphologically complex language with 16 vowels and 36 consonants spoken by the people of Kerala, one of the southern states of India. It has seven nominal case forms, two nominal number forms and three gender forms. These forms are used as suffixes to the nouns for nominal inflection. Tense, mood, voice and aspect causes verb inflection [4]. Malayalam also has a canonical word order of Subject-

Object- Verb (SOV) agreement. Like any other regional languages in India, Malayalam is a low resource language in terms of availability of labeled speech-to-text corpora. Malayalam language has scarcity of experts having linguistic knowledge with technological know-how. It also lacks various language models. Like any other regional language accents of the speech varies in different areas of the state. It also faces challenges like code switching and non-native speakers

III. Speech Signal Processing

Speech is produced by the vibration of vocal folds, which exerts a pressure in the surrounding medium. When the propagated sound waves cause the vibration of the ear drum of the listener, the process of hearing is initiated. The speech has rich temporal and spectral variations. The smallest frequency produced is called fundamental frequency F0 which varies in different persons due to anatomical differences. All other frequencies are harmonics of the fundamental frequency. The concentration of acoustic energy around a particular frequency in the speech wave is called formants. There are several formants in a speech signal, each at a different frequency, numbered from lowest frequency as F1, F2, F3 etc. which is used to characterize phonemes. F1 and F2 itself can characterize all the vowels.

The speech processing begins by recording human voice which is then sampled to convert the speech signal into discrete form using a sampling frequency based on the Nyquist–Shannon sampling theorem. Fourier transform converts an audio signal from time domain to frequency domain. A discrete Fourier transform computes the frequency representation of digital audio and transforms a discrete sequence of time-domain speech samples into a discrete sequence of frequency-domain coefficients.

IV. Language Models Used In Malayalam Speech Recognition

Language model compute the probability of each phoneme. Conventional ASR system is an integration of acoustic model trained on annotated speech corpus, language model for recognizing text and pronunciation dictionary. In traditional ASR systems features like MFCC, PLP, wavelet coefficients are extracted from the transformed signals which in turn results in the dimensionality reduction of the input. The likelihood of the phonemes are computed using generative model or discriminative model and which is then decoded to text. Advanced machine learning techniques facilitate end-to-end speech models which take raw speech signal as input and generate conditional probability of phoneme class by combining all the components of traditional ASR system into a single model. The end-to-end architecture for speech recognition uses attention based method, CTC and CNN based raw speech model.

a) Generative Models:

A generative model is a statistical model of the joint probability distribution. Generative models can learn the underlying patterns in the given data set and can generate new data based on the probability distribution. In this statistical model, the prior probability P(Y) and likelihood probability P(X|Y) are estimated with the help of the training data and the posterior probability P(Y |X)is calculated using the Bayes Theorem;

$$\mathbf{P(Y/X) = (P(Y)\ x\ P(X/Y))/(P(X))}$$

[5] proposes the first Automated Malayalam speech recognition which uses Hidden Markov Models (HMM) for acoustic modeling using MFCC. It was developed as simple 8 command system which is then enhanced to 50 commands. Another work [6] developed a speaker independent Malayalam Isolated Digit Recognition is based on Perceptual Linear Predictive Cepstral Coefficient (PLP) and continuous density Hidden Markov Model (HMM) for ASR modeling. The data set used was a limited vocabulary of Malayalam digits.

[7] compares and evaluates the performance automatic speech recognition of continuous speech recognition of Malayalam language using HMM context dependent tied (CD tied) model, context dependent (CD) model and Context independent (CI) models. The ASR system modeled using Phoneme-based Hidden Markov Models (HMM) and MFCC features for constrained vocabulary of continuous digits. [8] proposes an ASR system which uses Perceptual Linear Predictive Cepstral coefficients and continuous density Hidden Markov Model for the recognition of connected digits for limited vocabulary.

[10] builds an acoustic Model using HMM with different Gaussian Mixtures for Isolated Malayalam Speech Recognition for medium vocabulary utilizing MFCC features. Another work developed [17] an ASR system using support vector machine (SVM) with Quandratic, Cubic, Fine Gaussian, Medium Gaussian Coarse Gaussian kernel functions for the recognition of vowels in the Malayalam language using MFCC features.

b) Discriminative Models:

Discriminative models, directly assume some functional form for posterior probability **P(Y|X)** and then estimate the parameters of **P(Y|X)** with the help of the training data to calculate the probability. The first work based discriminative model in Malayalam proposes a speaker independent Automatic Speech Recognition System for isolated Malayalam vocabulary using hybrid system consisting of wavelet packet decomposition and artificial neural network[9]. Another work proposes a Malayalam Speech to Text Conversion model using Deep Learning [11]. HMM for the classification and LSTM for the training is use for the recognition of isolated words with constrained vocabulary using MFCC features.

[12] developed Convolutional Neural Network (CNN) for the acoustic modeling of Malayalam speech data using spectrogram images. The Convolutional Neural Network is built with a set of Convolution and Fully Connected layers with Softmax layer for classification of speech data. The proposed system used a vocabulary of 4000 tokens. Another work on developing a conventional and syllable-based ASR system for Malayalam used DNN for speech modeling following syllable-based approach using MFCC features [13]. An ASR system for Malayalam language is designed to recognize around 5-10 isolated words by using deep learning and MFCC feature extraction technique [14].A related research work for Speech Emotion Recognition used Convolutional Neural Network-Based approach for classification of emotions in Malayalam speech[15]. Another speech recognition system developed for Malayalam ASR used a DNN-HMM (Deep Neural Network–Hidden Markov Model) based automatic speech recognition with MFCC features using sub word tokens for language modeling [16].

C) End-To End Models:

Modern neural architectures has enabled the development of end-to-end ASR systems that directly translate input raw speech signal into output sequence.[18] proposes an end-to-end model by fine tuning pre trained XLS-R model using publicly available dataset of Malayalam language. On top of the XLS-R transformer with 0.3 billion parameters CTC layer is attached for fine-tuning.

Table 1 summarizes the major works in Malayalam Automatic Speech Recognition.

Table 1

Author, Year	Classifier	Acoustic Model	DataSet features
[5], 2007	HMM	Generative	Limited dataset of isolated word
[6], 2011	HMM (Continuous density)	Generative	Limited dataset of Malayalam digit
[7], 2011	HMM CD tied model	Generative	Limited dataset of continuous Malayalam digit
[8], 2013	HMM	Generative	Connected Malayalam digit
[9], 2012	ANN	Discriminative	Limited dataset Isolated Malayalam Word
[10], 2018	HMM & GMM	Generative	Medium vocabulary, speaker dependent isolated speech corpus of 100 words
[11], 2021	Deep Learning model using HMM classification and LSTM	Discriminative	context independent, Isolated Malayalam Words of constrained vocabulary
[12], 2021	CNN	Discriminative	4000 words
[13], 2022	DNN	Discriminative	Syllable based with Limited vocabulary
[14],2021	HMM classification and LSTM	Discriminative	Isolated words with constrained vocabulary of 5-10 isolated words.
[15],2022	CNN	Discriminative	Limited Vocabulary
[16] , 2023	DNN- HMM	Discriminative	Limited Vocabulary
[17], 2023	SVM	Generative	Vowels in Malayalam, very small dataset
[18], 2023	Transformers	End-to-End (fine tuning of multilingual pre trained model)	18 hrs of Malayalam speech data

V. Conclusion

This paper explores advancements in techniques and tools used for Automatic Speech Recognition of Malayalam language. Only a few end-to-end speech recognitions systems were developed for the modelling of continuous speech recognition of Malayalam language. The main constraint in building large language model for Malayalam is the lack of availability of annotated speech corpus. By building and training a model using a large dataset the accuracy of Malayalam ASR systems can be improved more. No work is reported in spontaneous speech recognition. Additionally, modelling of spontaneous Malayalam speech involves challenges like mass variety of vocabulary and sentence formations caused by mispronunciations, disruption in the flow of speech etc. Researches in the field of Automatic Speech Recognition of Malayalam language also demand the development of variety of datasets for training and testing various language models.

References

[1] Rashmi C R ,Review Of Algorithms And Applications In Speech Recognition System, International Journal Of Computer Science And Information Technologies, Vol. 5 , No.4 , Pp. 5258-5262, 2014.

[2] A Review On Speech Recognition Technique, International Journal Of Computer Applications, Vol 10, No.3, Pp. 16-24. 2010.

[3] Mohamed Hashim Changrampadi, A. Shahina, M. Badri Narayanan, A. Nayeemulla Khan, End-To-End Speech Recognition Of Tamil Language, Intelligent Automation & Soft Computing,Vol. 32, No.2, Pp. 1309-1323,2022.

[4] Manohar K, Jayan A R And Rajan R,Mlphon: A Multifunctional Grapheme-Phoneme Conversion Tool Using Finite State Transducers. Ieee Access, Vol.10, Pp. 97555–97575,2022.

[5] “Sharika - Malayalam Speech Recognition System”, Shyam.K, Icist, 2007.

[6] Cini Kurian, Kannan Balakrishnan, Malayalam Isolated Digit Recognition Using Hmm And Plp Cepstral Coefficient”, International Journal Of Advanced Information Technology, Vol. 1, No.5, Pp.31-38,2011.

[7] Cini Kurian, Kannan Balakrishnan, Development& Evaluation Of Different Acoustic Model For Malayalam Continuous Speech Recognition, International Conference On Communication Technology And System Design,2011.

[8] Cini Kurian, Kannan Balakrishnan, Connected Digit Speech Recognition System For Language”, Indian Academy Of Sciences,Sadhana, Part 6, Vol.38, Pp. 1339–1346, 2013.

[9] Sonia Sunny, David Peter S, K Poulose Jacob, Development Of A Speech Recognition System For Speaker Independent Isolated Malayalam Words, International Journal Of Computer Science & Engineering Technology, 2012

[10] Lekshmi.K.R, Elizabeth Sherly,An Acoustic Model For Isolated Malayalam Speech Recognition With Different Gaussian Mixtures(7), National Conference On Indian Language, Cochin University Of Science And Technology, 2018

[11] Lekshmi K R And Sherly E, An Asr System For Malayalam Short Stories Using Deep Neural Network In Kaldi. International Conference Of Artificial. Intelligence And. Smart Systems, Pp. 972–979, 2021.

[12] K R Lekshmi And Elizabeth Sherly, An Acoustic Model And Linguistic Analysis For Malayalam Disyllabic Words: A Low Resource Language, International Journal Of Speech Technology, Vol24(2), No.10, Pp. 483-495, 2021.

[13] Jasminsashish ,Abraham Samuel, Rajeev Rajan, Astudy On Conventional And Syllable-Based Approaches For Automatic Speech Recognition In Malayalam, Sadhana, Vol, 47, No,284,Pp.1-5,2022.

[14] Arun H P, Jithin Kunjumon, Sambhunath, Ancy S Ansalem, Malayalam Speech To Text Conversion Using Deep Learning, Iosr Journal Of Engineering, Vol. 11, No. 7, Pp. 2278-8719, 2021

[15] V K Muneer, K P Mohamed Basheer, Rizwana Kallooravi Thandil, Convolutional Neural Network-Based Automatic Speech Emotion Recognition System For Malayalam, Indian Journal Of Science And Technology, Vol. 16, No.46, Pp.4410-4420, 2022.

[16] Kavya Manohar, Jayan A R And Rajeev Rajan, Improving Speech Recognition Systems For The Morphologically Complex Malayalam Language Using Subword Tokens For Language Modeling, Eurasip Journal On Audio, Speech And Music Processing, Vol.2023, No.47, Pp. 2023:47, 2023.

[17] Leena G Pillai, D Muhammad Noorul Mubarak, Malayalam Language Vowel Classification Using Support Vector Machine For Children, Sadhana, The Indian Academy Of Sciences, Volume 48, No. 41, Pp. 1-10,2023

[18] Kavya Manohar,Gokul G. Menon,Ashish Abraham,Rajeev Rajan,A. R. Jayan ,3automatic Recognition Of Continuous Malayalam Speech Using Pretrained Multilingual Transformers, International Conference On Intelligent Systems For Communication, Iot And Security (Iciscois),2023.