IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 3, March 2016

A contextual survey on Speech Recognition Systems for Natural Language

Priyanaka R[1], Manjunath A E[2]

Dept. of CSE, R V College of Engineering, Bangalore, India [1, 2]

Abstract: The paper presents an overview of the automatic speech recognition (ASR) systems with the goal of bringing forth how the advancements in field of ASR have been applied in attempt to develop a natural language ASR. Initial sections of this paper will introduce the basic building blocks of an ASR and interactions among them to implement a speech recognition system. The techniques used in each of these building blocks are briefly introduced. Later sections of this tutorial will focus on efforts at building these essential building blocks for implementing a Kannada ASR and techniques employed for this purpose.

Keywords: ASR, Acoustic, Vector Quantization, DTW.

I. INTRODUCTION

Speech is fundamental to communication; it allows communication of feelings and ideas using a language as a framework. This information is output as speech (sound) signals and encoded in complex form that humans are naturally programmed to decode it. ASR systems try to emulate this ability. Such a system has to deal with multitude of problems do decode information from speech signal such as identifying the speaker, detecting the language spoken, transcribing speech and understanding the speech. In an ASR system linguistic message is the information of interest. Extracting information from the speech is a complicated process. The variability in speech to linguistic, regional influences and environmental factors are key challenges in reliably extracting the relevant information from the speech signal.

Speech recognition is essentially described as a function that defines a mapping from acoustic evidence to single or sequence of words. Let X ={X1, X_2, X_3... X_t} represent the acoustic evidence that is generated in from a given speech signal and belong to the complete set of acoustic sequences, X. Let W={W_1,W_2,W_3,.....Wn}denote the sequence of n words, each belonging to a fixed and known set of possible words ω. There are two frameworks to describe the recognition function: template and statistic.

II. TEMPLATE FRAMEWORK

In the template framework recognition is performed by finding the possible sequence of words W that minimizes the distance function between the acoustic evidence X and sequence of word patterns (templates). So the problem is to find the optimum sequence of template patterns R* that best matches X as follows.

$$R^*=R_S * d\ (R_S, X)$$

Where Rs is a concatenated sequence of template patterns from some admissible sequence of words. Complexity of such a system grows exponentially with the length of sequence of words in W. This framework uses the methods such as dynamic time wrapping (DTW) and vector quantization (VQ) achieve the template matching.

III.STATISTICAL FRAMEWORK

In the statistical framework, recognizer will select the sequence of words that is more likely to produce given the observed acoustic evidence. Let P(W|X) denote the probability that words W were spoken given that evidence X was observed. The recognizer will select W' satisfying

$$W' = W_{\infty} * P\ (W)P(X/W)$$

Using the Bayes rule above can be re-written as

P(W|X) =(P(W)P(X|W)) (P(x))

Where P(W) is the probability that sequence of words W will be uttered. P(X/W) is the probability of observing the acoustic evidence X when speaker utters W, and P(X) is probability that acoustic evidence X will be observed. Since P(X) is constant under max operation the recognizer needs to select sequence of words W' that maximizes the product P(W)P(X|W).

IV.LITERATURE SURVEY

Kannada has been one of the languages of interest in attempts to build language models and speech databases among other Indian languages. There have been various studies in understanding acoustic properties of Kannada language, phonemic analysis, building isolated words recognizer and building language database with small to medium vocabulary ASRS. This section lists and briefly discusses such attempts. Background cannot be removed clearly even lost or the parameter is smaller, it peel away the original color and give a bad performance. At this paper we set the artificial parameter of k-means with 3 for one cloth contains on more than three colors.

The isolated words recognizer in Natural language speech [1] based on Discreet Wavelet Transform (DWT) and Principal Component Analysis (PCA). Process begins by computing the DWT of the speech and MFCC coefficients are calculated. For this purpose PCA process was applied to speech recognition. This effort created database of kannada isolated digits from 0 to 10.

The acoustic characteristics [2] of any language can by analysis of phonemics. Phone to phone relationship is studied using 15 million words text corpus. Four hundred and forty five samples and most frequently used Kannada words were selected to measure duration, intensity, frequency and formants of phonemes in different emotional status and at different levels. The spoken word perception [6, 7] with focus on how speech perception capacities are used in segmenting and recognizing words in fluent speech. A word spotting technique was used. Eighty Kannada and non kannada words with 5 words and 5 non kannada words appearing twice were given audio presentation. Results of the study indicated that Kannada words were spotted better than non- kannada words supporting lexical representation of the words.

The automated speech translation in Asian languages as part of A-STAR consortium. Six Indian languages Hindi, Marathi, Malayalam, Tamil, Telugu, Kannada and Indian spoken English as seventh language were part of the study. Three approaches were considered namely parallel phone recognition (PPR), SWRLM and Parallel-SWRLM. It used two types of classifiers in the study – Maximum Likelihood classifier and Gaussian classifier.

V. SYSTEM ARCHITECTURE

Many of the successful ASR systems are based on the statistical framework described in previous sub section. The Equation defines the components of speech recognizer. The prior probability P (W) is determined by the language model and the likelihood P (X|W) is determined by set of acoustic models and the process of searching over all possible sequences of words W that maximizes the product is performed by the decoder. Below schematic diagram shows the architecture of ASR.

The statistical framework for ASR focuses on four key issues need be addressed for above architecture to viable.

1. The acoustic processing problem – Deciding what acoustic data X is going to be estimated. Need is to use a representation that reduces the model complexity. In general the speech waveform is transformed in to sequence of acoustic feature vectors in a process called feature extraction. Most commonly used methods for feature extraction are Signal based analysis, Production based analysis, and linear predictive analysis.
2. Acoustic modeling problem, i.e to decide on how P (X|W) should be computed. Thus several acoustic models are necessary to characterize how speakers pronounce the words of W given the acoustic evidence X. Acoustic models are highly dependent on type of the ASR application (ex – dictation, voice commands etc..). In general constraints are imposed to make the Acoustic models computationally feasible. Most popular method for estimating the acoustic models is HMM (hidden Marko models).
3. The language modelling problem deals with decision on how to compute the a priori probability of P (W) for a sequence of words. The most popular method is based on Markovian assumption that a word in sentence is conditioned on only the previous N -1 words. Such a statistical method is called N-gram.
4. The search problem deals with how to find best word transcription W' for the acoustic evidence X, given acoustic and language models. Some of the popular methods used in search algorithm are Viterbi search, Stack decoding, N-Best and Multipass search.

ASR Classification

ASR systems can be classified based on the parameters that govern the operation of ASR.

1. Vocabulary size: Speech recognition is easy when vocabulary size is small. Small vocabulary is measured as tens of words, medium hundreds of words and large in thousands of words. Grammar constraints can also affect the complexity of the system.
2. Speaking style: Isolated word or continuous speech define this mode of operation. Isolated word is less complex that continuous speech as it does not need to with empty spaces around words sequence of words.
3. Speaker mode: ASR can be used by a specific user (speaker dependent) or by any speaker (speaker independent). Speaker dependent systems need to be trained for the specific speaker and are much easier to implement due to low variability of speech. Speaker independent systems are difficult implement find much broader use compared speaker dependent systems.

VI. CONCLUSION

Automatic speech recognition systems are still and evolving science. Much of the early work done in building frameworks for language and Acoustic models are being enhanced by implementation today. Across the world experimentation in local languages are on both funded by government and privately. The applications of a reasonably good ASR are manifold impacting wide cross section of the society. Hence it assumes great importance as an evolving branch of science that is evolving quickly.

REFERENCES

[1]. Cheong Soo Yee and abdul Manan ahmad, Malay Language Text Independent Speaker Verification using NN- MLP classifier with MFCC, 2014 international Conference on Electronic Design.
[2] http://crdo.up.univaix.fr/ExternalDisk0/preview/000836/node303.html
[3] Zahi N.Karam,William M.Campbell "A new Kernel for SVM MIIR based Speaker recognition "MIT Lincoln Laboratory, Lexington, MA, USA.
[4] GIN-DER WU AND YING LEI " A Register Array based Low power FFT Processor for speech recognition" Department of Electrical engineering national Chi Nan university Puli ,545 Taiwan.
[5] Nicolas Morales1, John H. L. Hansen2 and Doorstep T. Toledano1 "MFCC Compensation for improved recognition filtered and band limited speech" Center for Spoken Language Research, University of Colorado at Boulder, Boulder (CO), USA.
[6] M.A.Anusuya, S.K.Katti "Speech Recognition by Machine: A Review" International journal of computer science and Information Security 2009.
[7] Samudravijay K "Speech and Speaker recognition report" source: http://cs.jounsuu.fi/pages/tkinnu/reaserch/inde x.html Viewed on 23 Feb. 2010
[8] Sannella, M Speaker recognition Project Report report" From http://cs.joensuu.fi/pages/tkinnu/research/index.html Viewed 23 Feb. 2010
[9] C.S.Myers and L.R.Rabiner, A Level Building Dynamic Time Warping Algorithm for Connected Word Recognition, IEEE Trans. Acoustics, Speech Signal Proc.,ASSP-29:284- 297,April 1981.
[10] IBM (2010) online IBM Research Source:- http://www.research. ibm.com/Viewed 12 Jan 2010.