Speech Recognition by using Optimization Method

Siddhartha Mishra, Assistant Professor

Mr. Anil Kumar Pandey,Saarthak Srivastava, Simran Gupta ,sajal

Dept of Computer Science and Engineering B.Tech

Babu Banarasi Das National Institute of Technology and Management

Abstract— In this paper we discuss the importance of speech recognition as well as we show how speech recognition evolved in last few years. Speech Recognition is a challenging domain, which made it an important research topic. In its early stage or at the time of invention it was not an as effective as it now, therefore many researchers worked on this domain and made it an exceptional one. This paper also shows the technological perspective and the progress in the field of speech communication. We also discuss how the speech recognition changed the way the world as it is now with the help of ASR (Automatics speech recognition) system work. Speech Recognition is based on the voice of research object. It allows the machine to turn speech signal into text as command. The speech recognition technology is gradually becoming the key technology of the IT man machine interface. The paper describes the development of speech recognition technology and its basic principles, methods, reviewed the classification of speech recognition systems and voice recognition technology. This paper gives you the complete idea about speech recognition.

Introduction

Speech Recognition is a process in which the speech signals are converted into words or sequence of words with the help of algorithms which are implemented as the computer programs. The basic and natural form of human communication is called as speech where speech processing is one of the most exciting topics of processing of signals. Speech recognition made easy for computers to follow and understand the human voice commands and languages. There are certain systems of speech recognition that are to be trained to work effectively which means a human has to interact with the system and give some voice inputs with

which the system trains and understand the commands. As everyone knows that speech is the primary way in which humans communicate with each other.

Hence the speech recognition systems made the human life easier like these days we people have mobile assistants which help and notifies about our doubts, schedules that we have made on our phone and many more activities. The customer services use speech recognition to help the customers with problems and doubts. Other examples where speech recognition is used is banking, voice dictation, data entry, helps handicapped people, railway reservations etc.

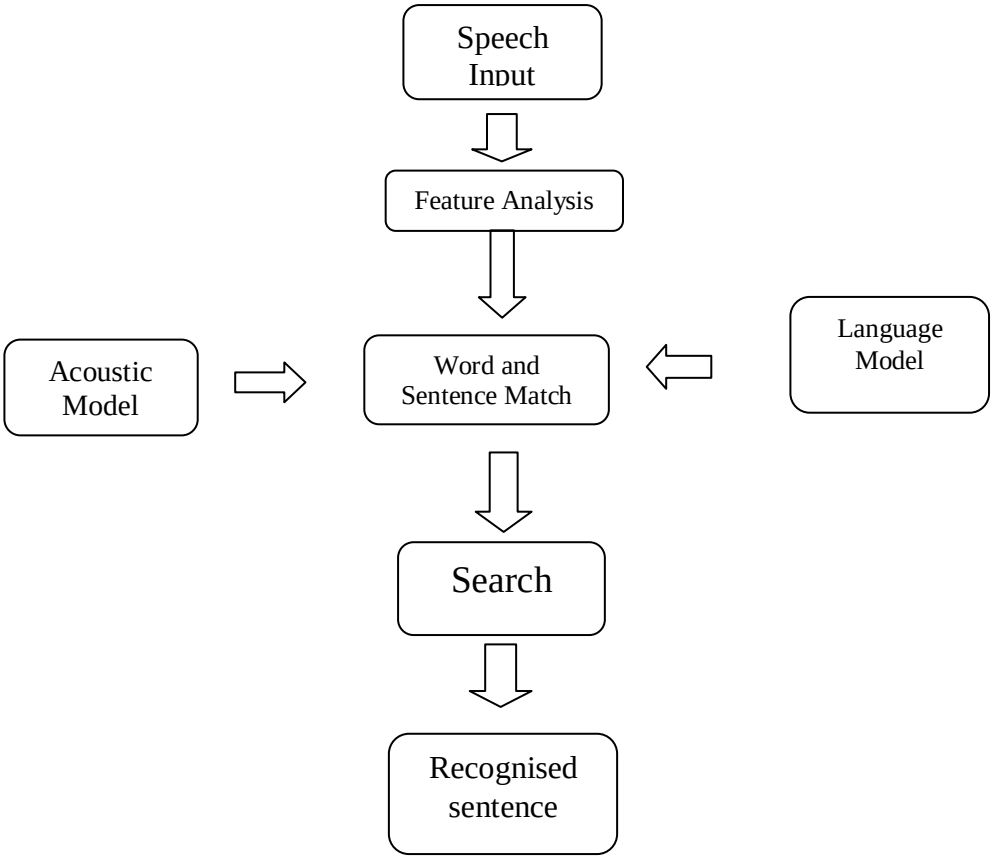

Figure 1: Simple Model of Speech Recognition

Speech recognition is divided into several types based on the utterances observer recognised. The classified types are:

i. ISOLATED WORDS: This type lacks the audio signal. In this when there is a sample window then both sides should be quiet during the utterance. It accepts single word at a time. This type of system consists of listen and not listen states where the speaker or the person who is speaking has to be quiet after uttering

words. Hence whenever the speaker is in silent state or the system is in not listen state that time the system processes the word that is uttered by the speaker before. This type of system is also called as discrete speech recognition systems.

ii. CONNECTED WORDS: This type of system allows separate utterances to run together with minimum amounts of pauses. This kind of systems are also called as connected words system. This type of system is same as that of isolated words system.

iii. CONTINUOUS SPEECH: Speaker can almost speak naturally in this type of speech recognition as while the system recognizes the content that the speaker is speaking. This is like computer dictation. These systems are made with difficult utterance boundaries which allows the speaker to speak continuously.

iv. SPONTANEOUS SPEECH: These are used frequently for the speakers as there is no particular way to speak with it as the system can even recognize the sounds that are made by the speaker in between while thinking like 'UMM'. The automatic speech recognition systems with spontaneous speech are really difficult to code. This type of system allows the natural sounding and no rehearsed speeches from the speaker.

v. NATURAL LANGUAGE: In this type of system, they not only recognize the words of utterances by the speaker but also give the reply to the questions and the doubts from the speaker. Basically, the speech recognition systems are considered as two types which are speaker dependent systems and speaker independent systems.

vi. SPEAKER DEPENDENT SYSTEM: In this, the systems need to be trained in order to recognize or understand or process the utterances by the speaker.

vii. SPEAKER INDEPENDENT SYSTEM: In this, there is no need for the system to be trained as it recognizes and understand most user's voices.

Optimization

In optimization of a design, the design objective could be simply to minimize the cost of production or to maximize the efficiency of production. An optimization algorithm is a procedure which is executed iteratively by comparing various solutions till an optimum or a satisfactory solution is found. With the advent of computers, optimization has become a part of computer-aided design activities. There are two distinct types of optimization algorithms widely used today.

(a) Deterministic Algorithms.

They use specific rules for moving one solution to other. These algorithms are in use to suite some times and have been successfully applied for many engineering design problems.

(b) Stochastic Algorithms.

The stochastic algorithms are in nature with probabilistic translation rules. These are gaining popularity due to certain

properties which deterministic algorithms do not have.

Types of Optimization Techniques:

GENETIC ALGORITHM

A genetic algorithm (GA) is a method for solving both constrained and unconstrained optimization problems based on a natural selection process that mimics biological evolution. The algorithm repeatedly modifies a population of individual solutions. At each step, the genetic algorithm randomly selects individuals from the current population and uses them as parents to produce the children for the next generation. Over successive generations, the population "evolves" toward an optimal solution.

TRAVELING SALESMAN OPTIMIZATION

The traveling salesman problem is a classic problem in combinatorial optimization. This problem is to find the shortest path that a salesman should take to traverse through a list of cities and return to the origin city. The list of cities and the distance between each pair are provided. TSP is useful in various applications in real life such as planning or logistics. For example, a concert tour manager who wants to schedule a series of performances for the band must determine the shortest path for the tour to ensure reducing traveling costs and not making the band unnecessarily exhausted.

KNAPSACK OPTIMIZATION PROBLEM

The knapsack problem is a problem in combinatorial optimization: Given a set of items, each with a weight and a value, determine the number of each item to include in a collection so that the total weight is less than or equal to a given limit and the total value is as large as possible.

Basic Speech Features:

PROSODY: Prosodic phonology is a theory of the way in which the flow of speech is organized into a finite set of phonological units. It is also, however, a theory of interactions between phonology and the components of the grammar. Although many speech interfaces are already available, the need is for speech interfaces in local Indian languages. Application specific Indian language speech recognition systems are required to make computer aided teaching, a reality in rural schools. This paper presents the preliminary work done to demonstrate the relevance of an Oriya Continuous Speech Recognition System in primary education.

PHONEMES: The sounds of language are classified into what are called phonemes. A phoneme is minimal unit of sound that has semantic content. e.g., the phoneme AE versus the phoneme EH captures the difference between the words “bat” and “bet”. Not all acoustic changes change meaning. For instance, singing words at different notes doesn’t change meaning in English. Thus changes in pitch does not lead to phenemic distinctions

MEL FREQUENCY CEPSTRUM COEFFICIENT (MFCC):

Mel frequency cepstrum coefficient (MFCC) technique MFCC represents the power spectrum for speech signal on basis of transformation of the speech signal. MFCC generate mimics of the human auditory system. In Mel frequency scale, linear frequency spacing is less than process. In feature extraction process, continuous speech is entered as input for windowing. Before transformation stage windowing reduces the disruption process. After that, a speech signal which is in a continuous form converted in frames of the window. Then these frames are passed on to Fourier transformation process which transforms frames of the window into a spectrum. Review of Techniques of existing speech recognition

In 1877 Thomas Edison is the first person who invented the very first device, phonograph that can record and reproduce the sound. It is very fragile and considered as prone to damage. Later in 1879 Thomas Edison invented the first dictation machine which is considered as the improved version of his phonograph. In 1936 At bell labs a team of engineers led by Homer Dudley, invented the first speech electronic synthesizer called as Voder (Voice Demonstrator). In 1939 The patent was confirmed for Dudley for his invention Voder. In 1952 At bell labs a team designed a machine which is capable of understanding spoken digits. In 1962 IBM demonstrated shoebox that can understand 16 spoken words from the speaker at fair. In 1971 A device is invented by IBM named as Automatic call identification system using which a person can talk and receive the spoken answers from another person from the device. Early 80’s The technique named Hidden Markov Model is being put in use in machines for the first time since then. Mid 80’s IBM started working on a machine that can understand nearly 20,000 spoken words and it was named as Tangora. In 1987 The invention of World of Wonders a Julie doll which is toy is done and it is trained to respond to the speaker’s voice and brought the speech recognition home. In 1990 machine Dragon Dictate was invented and launched by Dragon company which is considered as the first speech recognition machine for customers. In 1993 The first built-in speech recognition and voice enabled control software was introduced to the apple computers which is known as Speakable items. In 1993 first large vocabulary speech recognition system names Sphinx II was invented by Xuedong Huang. Then in 1996 first commercial product named Med Speak came into light which is capable of recognizing continuous speech, it was invented by IBM. In year 2007 GOOG-411 was launched by the Google company that served as the foundation for the future voice search product. It is a

telephone-based directory service. 2011 was the year when Apple company launched the digital personal assistant named as Siri. It can not only understand the speech by the user but also it does the appropriate actions based on the speech. Then in year 2014 The voice-controlled speaker called Echo which is powered by Alexa is invented by Amazon. This Echo is kind of similar to Cortana and Siri but is different in many aspects.

Classic Optimization Problem:

PRINCIPLE COMPONENT ANALYSIS (PCA) -: Principal Component Analysis (PCA) is a statistical procedure that uses an orthogonal transformation that converts a set of correlated variables to a set of uncorrelated variables. PCA is the most widely used tool in exploratory data analysis and in machine learning for predictive models. It is a traditional eigen vector base method called as karhuneu-loeve expansion. It is good for guassian data, Non-linear feature extraction,`Fast,Eigen vector based,Linear map.

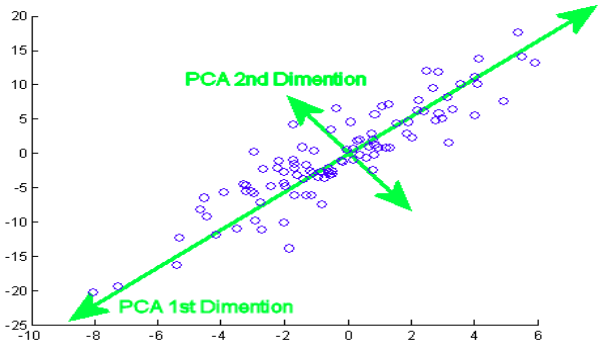

Figure 2: Principle component analysis (PCA)

LINEAR DISCRIMINATE ANALYSIS (LDA) -: Linear Discriminant Analysis or LDA is a dimensionality reduction technique. It is used as a pre-processing step in Machine Learning and applications of pattern classification. This method is better than PCA. Non-linear feature extraction, Fast, Eigen vector based, Supervised linear map.

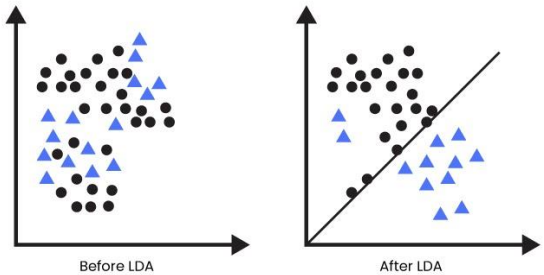

Figure 3: *Linear discriminate analysis (LDA)*

INDEPENDENT COMPONENT ANALYSIS (ICA) -: Independent Component Analysis (ICA) is a machine learning technique to separate independent sources from a mixed signal. Unlike principal component analysis which focuses on maximizing the variance of the data points, the independent component analysis focuses on independence, i.e. independent components. It is used for demixing of nonguassian fistributed features. Used for blind course separation. Non- linear feature extraction,Linear map,Iterative nonguassian.

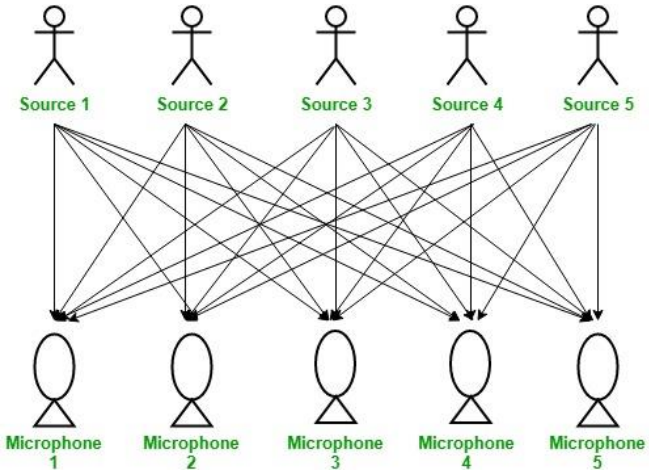

Figure 4: *Independent component analysis (ICA)*

Linear predictive coding -: Linear predictive coding (LPC) is a method used mostly in audio signal processing and speech processing for representing the spectral envelope of a digital signal of speech in compressed form, using the information of a linear predictive model. Mainly used for lower order feature extraction.Static feature extraction,10-16 lower order coefficient.

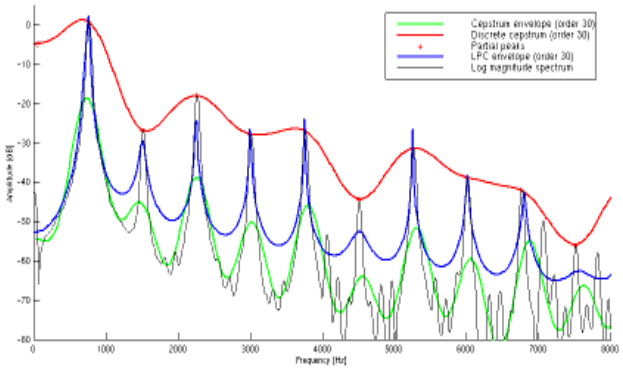

Figure 5: Linear predictive coding

Cepstral analysis -: Cepstrum Analysis is a tool for the detection of periodicity in a frequency spectrum, and seems so far to have been used mainly in speech analysis for voice pitch determination and related questions. In that case the periodicity in the spectrum is given by the many harmonics of the fundamental voice

frequency, but another form of periodicity which can also be detected by cepstrum analysis is the presence of sidebands spaced at equal intervals around one or a number of carrier frequencies. It is used for the representation of spectral envelope. Static feature extraction, Power spectrum.

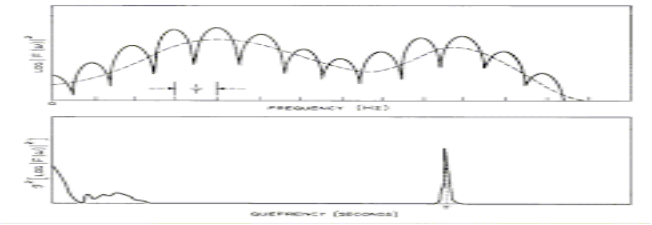

Figure 6: Cepstral analysis

Speech recognition vs Face Recognition

speech recognition software provides a simple way to get words into a document without having to be delayed in the process ,Since most people speak faster than they write. This speed is what makes many people seek out its use. Typing, on the other hand, can slow down the communication process. When working with a client or completing a task, the use of speech recognition tools facilitates easy note taking, use of other materials, and professional eye contact. Each of these activities is limited when someone has to type information into a computer behind a screen Although it should always be proofread, speech recognition software can result in a document more or less free of errors. In addition, newer programs tend to be well designed and can offer reliable results for some applications. It can help to increase productivity in many businesses, such as in healthcare industries. It can capture speech much faster than you can type You can use text-to-speech in real-time. The software can spell the same ability as any other writing tool.

Face recognition threat to individual privacy. Facial recognition carries its own security risk. Unlike a password, biometric data cannot be changed. If my fingerprint data leaks, this is not something I can “reset” like a password. To make things worse, your face can be scanned anytime and anywhere, without your consent. This means that your biometric data might actually be stored in a range of databases, whose security measures might be inadequate. This technology is also known to work relatively well on white and male faces, while having a high rate of inaccuracy on people of colour, especially if they are female. This means that people from ethnic minority groups, who already tend to have less easy access to services and amenities, will encounter an extra barrier when authorities use facial recognition in these contexts. Lawbreakers can use facial recognition technology to perpetrate crimes against innocent victims too. They can collect individuals’ personal information, including imagery and video collected from facial scans and stored in databases, to commit identity fraud.

Future Scope: This paper comprises the information on speech recognition that is known or which is invented or designed or understood or explained or spoken of or discussed at length since the time it was known to have existed. What is known about speech recognition is very limited and the gap between what is known and what is yet to be explored is magnificent. This paper can help the researchers to understand the speech recognition properly also it can help to know what had happened in the field of speech recognition till now. The future inventions should be more exciting in this field for example consider an environment where large number of speakers are there and the machine's work is to understand who is speaking from those large number of speakers and do the appropriate action according to the command or the speech. This paper can help the researcher to grasp the knowledge about the technology.

Conclusion: Speech is one of the most effective and natural ways of communications. Due to the interest in this field, many machines were invented in the past decades that could recognize, understand and respond to the speech. As we can see there is really a tremendous growth in this area and also many applications, software and machines were invented. The main difference between genetic algorithm and traditional algorithm is that genetic algorithm is a type of algorithm that is based on the principle of genetics and natural selection to solve optimization problems while traditional algorithm is a step by step procedure to follow, in order to solve a given problem. There are also practical limitations which hinder the use of services and applications.

Reference:

[1] Anasuya, M. A., & Katti, S. K. (2009). Speech recognition by machine: A review. International Journal of Computer Science and Information Security, 6, 181-205.

[2] Walker, D. (1975). The SRI speech understanding system. *IEEE transactions on acoustics, speech, and signal processing*, *23*(5), 397-416.

[3] Bhagath, P., & Das, P. K. (2004). Acoustic Phonetic Approach for Speech Recognition: A Review. *Language*, *77*, 93.

[4] Diller, T. (1979, April). Phonetic word verification. In *ICASSP'79. IEEE International Conference on Acoustics, Speech, and Signal Processing* (Vol. 4, pp. 256-261). IEEE.

[5] Green, P., & Wood, A. (1986, April). A representational approach to knowledge-based acoustic-phonetic processing in speech recognition. In *ICASSP'86. IEEE International Conference on Acoustics, Speech, and Signal Processing* (Vol. 11, pp. 1205-1208). IEEE.

[6] Pisoni, D. B., & Luce, P. A. (1987). Acoustic-phonetic representations in word recognition. *Cognition*, *25*(1-2), 21-52.

[7] Reddy, D. R. (1966). Approach to computer speech recognition by direct analysis of the speech wave. *The Journal of the Acoustical Society of America*, *40*(5), 1273-1273.

[8] Myers, C., & Rabiner, L. (1981). A level building dynamic time warping algorithm for connected word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *29*(2), 284-297.

[9] XU, Yi (2011) SPEECH PROSODY: A METHODOLOGICAL REVIEW University College London

[10] Padmalaya Pattnaik & Shreela Dash (2012) Raman College of Engineering, Bhubaneswar