ISSN: 2348 - 2117

International Journal of Engineering Technology and Computer Research (IJETCR)
Available Online at www.ijetcr.org
Volume 4; Issue 3; May-June-2016; Page No. 33-37

A Review: Speech Reorganization by Using Artificial Neural Network

Dr. Kavita[1], Dr. Akash Saxena[2], Jitendra Joshi[3]

[1]Research Supervisor, Jayoti Vidyapeeth Women's University, Jaipur, India.

[2]Research Co-Supervisor, *Compucom* Institute of Technology & Management, *Jaipur*, India.

[3]Research Scholar Jayoti Vidyapeeth Women's University, Jaipur, India.

E-mail: Lect.jitendra29@gmail.com

Abstract

This paper presents the basic idea of speech recognition, proposed types of speech recognition issues in speech recognition, different useful approaches for feature extraction of the speech signal with its advantage and disadvantage and various pattern matching approaches for recognizing the speech of the different speaker. Now day's research in speech recognition system is motivated for ASR system with a large vocabulary that supports speaker independent operations and continuous speech in different language.

Keywords:-Pattern Recognition, Automatic Speech Recognition (ASR), Acoustic Modeling, Language Modeling. Feature extraction, pattern matching, ANN, HMM, DTW, MFCC.

1. Introduction

Automatic recognition of speech by machine has been a goal of research for more than four decades. In the world of science, computer has always understood human mimics. The idea which generated for making speech recognition system is because it is convenient for humans to interact with a computer, robot or any machine through speech or vocalization rather than difficult instructions [1]. Human beings have long been inspired to create computer that can understand and talk like human. Since, 1960s computer scientists have been researching various ways and means to make computer record, interpret and understand human speech [2].

The fundamental aspect of speech recognition is the translation of sound into text and commands. Speech recognition is the process by which computer maps an acoustic speech signal to some form of abstract meaning of the speech. This process is highly difficult [3] since sound has to be matched with stored sound bites on which further analysis has to be done because sound bites do not match with pre-existing sound pieces. Various feature extraction methods and pattern matching techniques are used to make better quality speech recognition systems. Feature extraction technique and pattern matching techniques plays important role in speech recognition system to maximize the rate of speech recognition of various persons.

2. Classification of Speech Recognition System

A. Types of speech recognition system based on utterances

- *Isolated Words*

Isolated word recognition system which recognizes single utterances i.e. single word. Isolated word recognition is Suitable for situations where the user is required to give only one word response or commands, but it is very unnatural for multiple word inputs. It is simple and easiest for implementation because word boundaries are obvious and the words tend to be clearly pronounced which is the major advantage of this type. The drawback of this type is choosing different boundaries affects the results [4].

- *Connected Words*

A connected words system is similar to isolated words, but it allows separate utterances to be "run-together" with a minimal pause between them. Utterance is the vocalization of a word or words that represent a single meaning to the computer.

- *Continuous Speech*

Continuous speech recognition system allows users to speak almost naturally, while the computer determines its content. Basically, it is computer dictation. In this closest words run together without pause or any other division between words.

Corresponding author: Jitendra Joshi

Continuous speech recognition system is difficult to develop.

- *Spontaneous Speech*

Spontaneous speech recognition system recognizes the natural speech. Spontaneous speech is natural that comes suddenly through mouth. An ASR system with spontaneous speech is able to handle a variety of natural speech features such as words being run together. Spontaneous speech may include mispronunciation, false-starts and non words.

B. Types of speech recognition based on Speaker Model

Each speaker has special voice, due to his unique physical body and personality. Speech recognition system is classified into three main categories as follows:

- *Speaker Dependent Models*

Speaker dependent systems are developed for a particular type of speaker. They are generally more accurate for the particular speaker, but could be less accurate for other type of speakers. These systems are usually cheaper, easier to develop and more accurate .But these systems are not flexible as speaker independent systems.

- *Speaker Independent Models*

Speaker Independent system can recognize a variety of speakers without any prior training. . A speaker independent system is developed to operate for any particular type of speaker. It is used in Interactive Voice Response System (IVRS) that must accept input from a large number of different users. But drawback is that it limits the number of words in a vocabulary. Implementation of Speaker Independent system is the most difficult. Also it is expensive and its accuracy is lower than speaker dependent systems.

- *Speaker Adaptive Models*

Speaker adaptive speech recognition system uses the speaker dependent data and adapt to the best suited speaker to recognize the speech and decreases error rate by adaption [6]. They adapt operation according to characteristics of speakers.

C. Types of speech recognition based on Vocabulary

The size of vocabulary of a speech recognition system can affect the complexity, processing and the rate of recognition of ASR system. So that ASR systems are classified based on the vocabulary as following: • Small Vocabulary - 1 to 100 words or sentences • Medium Vocabulary - 101 to 1000 words or sentences • Large Vocabulary- 1001 to 10,000 words or sentences • Very-large vocabulary - More than 10,000 words or sentences.

3. Literature Review

The scientists ISO & Watanabe have suggested the NPM (Neural Prediction Model) a speech (voice) recognition neural network. The certain model uses the multilayer perceptions to predict the patterns (It is not for the recognition of pattern) [6]. And researcher Uchiyama has proposed a RNPM (Re-current Neural Prediction Model), and architecture of recurrent network for particular model.

The model which has proposed here used to be reducing the network size very significantly, with the high recognition rate such as original model, by keeping high learning efficiency, for the speaker autonomous isolated words [7].

The researcher Kuhn & Herzberg have described some of variations on the multiple layered recurrent networks training, which usually overcome the requirement for the externally-supplied target operation (function), avoid the back-propagation of an error derivatives in the time, minimize the training time, and improve the generalization. The applied problem of the speech recognition, such number of variations resulted minimum training iterations number. As well, they got that the recurrent networks haven't given the maximum improvement in performance over the challenging non recurrent network with the similar weights number [8]. The GRNNs (General Regression Neural Networks) have been practiced to a phoneme identification as well separated the word recognition in a clean speech. In the [9], researcher Amrouche et al. has extended particular approach to the word recognition in Arabic spoken following adverse conditions. The matter of fact that, the robustness of noise is used to be challenging issues in ASR (Automatic Speech Recognition) and much of the existing recognition techniques, that have shown to be extremely effective under the noise free conditions, as well it fail drastically in the noisy environments. The suggested system used to be tested for the recognition of Arabic digit at distinct SNR (Signal to-Noise Ratio) levels. The suggested scheme was compared successfully to a similar recognizers depended on the MLP (Multilayer Perceptions), the Elman RNN (Recurrent Neural

Network) and a discrete HMM (Hidden Markov Model). The result of experiment showed that a non-parametric regression use with suitable smoothing factor (spread) enhanced a generalization power of a neural network and global speech recognizer performance in the noisy environments.

The researcher Nakamura described [10] the phoneme PFN (Filter Neural Network) approach in order to recognize phoneme. The multilayer neural network PFN with the less hidden units than the input units equipped for every category of phoneme. Every network gets trained as the identity mapping by the data speech belonging to a category of phoneme. In a process of recognition, the similarity in between output data and input data is then computed for every network. The experimental results need to apply with the task of Japanese vowel recognition illustrated with the rates of PFN recognition at higher than certain conventional 3-layer PFN outputs and neural network usually represented the likelihoods of candidate.

The researcher Bottou [11] has focused on the speaker-independent, task of global word recognition with time-delay networks. Initially, we described the certain path of networks for the purpose of learning the feature extractors with the constrained back propagation. Such TDN (Time Delay Network) is described as the competent dealing with closely actualized problem: The reorganization of French digit. The researcher Ganapathy has presented [12] a technique of feature extraction depended on static & dynamic spectrum of modulation that derived from the long-term sub bands envelopes. The temporal sub-band estimation envelopes are used to be done by using the FDLP (Frequency Domain Linear Prediction). Certain envelopes of sub-band are compressed with the dynamic (adaptive loops) compression and static (logarithmic) compression. The sub bands of compressed envelopes are transformed to the spectral modulation components that are used for speech recognition features. The experiments get performed on the phoneme task of recognition by using hybrid HMM ANN system of phoneme recognition as well the ASR task by using a speech recognition system TANDEM. The suggested features give relative enhancements of 11.5% & 3.8% in the accuracies of phoneme recognition for TIMIT and CTS (Conversation Telephone Speech) correspondingly. And further, such improvements are usually being constant for the ASR tasks on database of OGI-Digits (relative 13.5% improvement).

The researcher Jamil Arous (2012) has investigated that the use of bothe dynamic and static c neural networks in the recognition of phoneme. Besides that, the research paper also suggests the cooperative model of static & dynamic neural networks. And the co-operative model integrates with the decision system for the recognition of phoneme. The coding of Mel cepstrum has applied to describe the frames in speech signal. Selected frames features are utilized to train the model based on neural networks. The relative study illustrate that the suggested co-operative model gives more accurate rates of recognition both in the generalization and auto-coherence test.
In particular research paper, we studied the cooperative dynamic and static neural networks model. The NN models case study is the recognition phoneme. The results are given below:

- The Elman NN and PMC give nearly the similar rates of recognition both in generalization and auto-coherence test.
- The decision strategy based on probability gives more accurate rates of recognition.
- The co-operative system gives the best rates of recognition.
- Entire NN-based models can not to separate the “near” phonetically vowels.

They have given the hidden Markov and hybridizing NN models for ameliorate recognition rates and to produce frontier in between the phonetically phonemes of “near”. With the similar context, we have given other ways learning to combine the NN models by utilizing boosting and bagging techniques. So as a future work, they have suggested combining neural networks with related areas like soft-computing and fuzzy logic in general to contribute the complex decision-making in particular fields.
The researcher Ravindra A. Athale in year 1992 suggested that the function of input-output transfer of the neuron is vitally in behavior detaction of the greatly internally connected network. They described the exclusive move toward the realization of the function of neuron transfer that described by

a Weber-Fechner Law by exploring a dynamical behavior of ET (Electron Trapping) materials. As well the interaction in between the blue light (close to infrared light intensity) and ETD (Electron Trap Density) closely resembles which needed by the action f mass law. The suggested scheme may realize a neuron's complex dynamical behavior in the sole ET thin-film particularly in the spatially continuous manner. Its features may result in the higher density of neurons contrasted with an approach that combines the discrete sub-cells functioning a constituent functions. The paper re-presents analytical description of equations of Neural Net STM. Where the equations are illustrating the ET dynamics as well as the data of initial experiments.

The suggested optical architecture design uses different optical system properties to estimate an excitation level I of net input and to transmit it to entire neurons to give an adaptive threshold and in order to control lateral inhibitory signals strength. The ET material dynamics is particularly used to execute a multiplicative inhibition. As well the computationally easy operation of the thresholding and level shifting is functioned by the custom circuits of CMOS. The entire system complexity is implementing to fix the point dynamics which minimized by resorting to an inherent ET material behavior. The whole optical system may be employed in the compact manner as well as with the compatibility of input-output with the storage of analog optical by encoding the long-term storage of memory. Which in turn, it may lead to extreme sophisticated system of optical multi-layer neural net depended on the models of ART (Adaptive Resonance Theory).

The researcher Jean-Philippe and S. Draye,et.al in year 1996 have represent a research paper in that they explored a dynamical features of the model of neural network that presents the two adaptive parameters types: The classical weights in between time constants and units connected with every artificial neuron. The aim of particular study is to give the strong theoretical basis for simulating and modeling the dynamic re-current neural networks. And to gain particular, they studied the statistical distribution effects of time constants and weights on dynamics of network and it will make the statistical analysis for the neural transformation.

4. Conclusion

In this review paper the basics of speech recognition system and different approaches available for feature extraction and pattern matching has been discussed. Using these various techniques rate of speech recognition can be improved and better quality speech recognition can be developed. In future there will be focus on development of large vocabulary speech recognition system and speaker independent continuous speech recognition system.

References

1. Malay Kumar, R K Aggarwal, Gaurav Leekha and Yogesh Kumar "Ensemble Feature Extraction Modules for Improved Hindi Speech Recognition System", International Journal of Computer Science Issues, Vol. 9, Issue 3, No 1, May 2012. Pukhraj P. Shrishrimal, Vishal B. Waghmare, Ratnadeep Deshmukh, "Indian Language Speech Database: A Review", I international Journal of Computer Application, Vol 47– No.5, June 2012.
2. Hemakumar, Punitha, "Speech Recognition Technology: A Survey on Indian languages", International Journal of Information Science and Intelligent System, Vol. 2, No.4, 2013.
3. Sanjivani S. Bhabad, Gajanan K. Kharate, "An Overview of Technical Progress in Speech Recognition" International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 3, March 2013.
4. M. A. Anusuya, S.K. Katti, "Speech Recognition by Machine: A Review", (IJCSIS) International Journal of Computer Science and Information Security, Vol. 6, No. 3, 2009.
5. Pratik K. Kurzekar, Ratndeep R. Deshmukh, Vishal B. Waghmare, Pukhraj P. Shrishrimal, "Continuous Speech Recognition System: A Review", Asian Journal of Computer Science and Information Technology, 2014.
6. R.K. Aggarwal, M. Dave "Integration of Multiple acoustic and language models for improved Hindi speech Recognition system", Springer Science Business Media, LLC 2012.
7. Bishnu Prasad Das1, Ranjan Parekh, "Recognition of Isolated Words using Features based on LPC, MFCC, ZCR and STE, with Neural Network Classifiers", International Journal of Modern Engineering Research ,Vol.2, Issue.3, May-June 2012.

8. S.B. Magre, R.R. Deshmukh, "A Review on Feature Extraction and Noise Reduction Technique", International Journal of Advanced Research in Computer Science and Software Engineering Vol 4, Issue 2, February 2014.

9. Rajesh Kumar Aggarwal, Ph.D_Thesis, "Improving Hindi Speech Recognition Using Filter Bank Optimization and Acoustic Model Refinement", December 2012.

10. Ankit Kumar, Mohit Dua, "Continuous Hindi Speech Recognition using Monophone based Acoustic Modeling", International Journal of Computer Applications® (0975 – 8887), 2014.

11. Gaurav, Devanesamoni Shakina Deiv, Gopal Krishna Sharma, Mahua Bhattacharya, "Development of Application Specific Continuous Speech Recognition System in Hindi", Journal of Signal and Information Processing, 2012, 3, 394-401

12. Preeti Saini, Parneet Kaur, "Automatic Speech Recognition: A Review", International Journal of Engineering Trends and Technology- Volume4Issue2-2013.

13. Hanmin Jung, Sung-Pil Choi, Seungwoo Lee, "Survey on Kernel-Based Relation Extraction".Chapter 1.

14. Sanjivani S. Bhabad, Gajanan K. Kharate, " An Overview of Technical Progress in Speech Recognition", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 3, March 2013.

15. Vimala.C, Dr.V.Radha, "A Review on Speech Recognition Challenges and Approaches", World of Computer Science and Information Technology Journal, Vol. 2, No. 1, 1- 7, 2012.

16. Ranu Dixit, Navdeep Kaur, "Speech Recognition Using Stochastic Approach: A Review", International Journal of Innovative Research in Science, Engineering and Technology, Vol. 2, Issue 2, February 2013.

17. Joe Tebelskis, "Speech Recognition using Neural Networks", PHD thesis, School of Computer Science Carnegie Mellon University Pittsburgh, Pennsylvania May 1995.

18. Abhishek Thakur, Naveen Kumar, "Automatic Speech Recognition System for Hindi Utterance with Regional Indian Accents: A Review", International Journal of Electronics & Communication Technology, Vol. 4, April – June 2013.

19. Simone Marinai, Marco Gori, "Artificial Neural Networks for Document Analysis and Recognition", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, No. 1, January 2000.

20. Santosh K.Gaikwad, Bharti W.Gawali, Pravin Yannawar, "A Review on Speech Recognition Technique" International Journal of Computer Applications, Vol 10– No.3, November 2010.

21. M. Chandrasekar, M. Ponnavaikko, "Tamil speech Recognition: a complete model", Electronic Journal «Technical Acoustics» 2008.

22. Rashmi C R, "Review of Algorithms and Applications in Spe ech Recognition System", International Journal of Computer Science and Information Technologies, Vol. 5 (4), 2014.

23. R. Vinay Chand, M.Veda Chary, "Wireless Home Automation System with Acoustic Controlling", International Journal of Engineering Trends and Technology (IJETT) – Volume 4 Issue 9- Sep 2013.