iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** VI **Month of publication:** June 2020

DOI: http://doi.org/10.22214/ijraset.2020.6358

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue VI June 2020- Available at www.ijraset.com

Speech Recognition using Deep Learning Techniques

Reshma C R[1], Ratheesh S[2]
[1, 2]*Department of Computer Science, Kerala Technological University*

Abstract: Speech is the essential method of communication between human beings. Several researches are done on the use of machine learning for speech recognition. Speech recognition mechanisms of converting the recorded speech signals into the text are one of the challenging task. In this paper a framework for speech recognition is proposed.
Keywords: Speech, Speech Recognition, Speech Recognition Architecture, Machine Learning, Deep Learning

I. INTRODUCTION

Speech is the natural and effective method of human communication. Speech recognition is that the ability of a machine or program to spot words and phrases in speech and convert them to a machine-readable format.. Speech recognition works using algorithms through acoustic and language modelling. Acoustic modelling represents the connection between linguistic units of speech and audio signals; language modelling matches sounds with word sequences to help distinguish between words that sound similar.

Speech processing is one of the exciting research area of signal processing. Speech processing is the study of speech signals and the processing methods of these signals. The signals are processed in a digital representation, so speech processing is considered as a special case of digital signal processing, applied to speech signal. Some of the speech processing applications are Speech Coding, Text-to Speech Synthesis, Speech Recognition, Speaker Recognition and Verification, Speech Enhancement, Speech Segmentation and Labelling (Transcription), Language Identification, Prosody, Attitude and Emotion recognition, Audio-Visual Signal Processing and Spoken Dialog System [1].

The speech recognition software is easy to use and readily available. Speech recognition software is now frequently installed in computers and mobile devices, allowing quick access. Speech recognition offers the way to speak with the technology around us. The downside of speech recognition includes its inability to capture words because of variations of pronunciation, its lack support for several languages outside of English and its inability to sort through ground noise. These factors can cause inaccuracies. Speech recognition performance is measured by accuracy and speed. Accuracy is measured with word error rate. Speed is measured with the real-time factor. Other measures of accuracy include Single Word Error Rate(SWR) and Command Success Rate(CSR).

The conventional speech recognition systems are based on representing speech signals using Gaussian Mixture Models (GMMs) that are based on hidden Markov models (HMMs) [2]. The limitation of HMM is the requirement of large amount of training data. The GMM can successfully separate the noise from speech in noisy speech utterances but it increases the computational complexity.

Deep learning could also be a replacement area of machine learning research. Deep learning is one of the progressive and promising areas in machine learning for the future tasks involved in machine learning especially in the area of neural network. Deep learning is becoming a mainstream technology for speech recognition and has successfully replaced Gaussian mixtures for speech recognition. Deep learning consists of a multiple of machine learning algorithms fed with inputs in the form of multiple layered models. These models are usually neural networks consisting of varied levels of non-linear operations. The machine learning algorithms decide to learn from these deep neural networks by extracting specific features and knowledge.

II. LITERATURE REVIEW

Some surveys are conducted within the area of speech recognition. For instance, Morgan [3] conducted a review in the area of speech recognition assisted with discriminatively trained feed-forward networks. The main focus of the review was to shed the light on papers that employ multiple layers of processing prior to the hidden Markov model based decoding of word sequences. Some of the methods that incorporate multiple layers of computation for the purpose of either providing large gains for noisy speech in small vocabulary tasks or significant gains for high Signal-to-Noise Ratio (SNR) speech on large vocabulary tasks were described.

Hinton et al. [4], presents an overview on the use of deep neural networks that incorporate many number of hidden layers that are trained using some of the new techniques. The advantage of a feed-forward neural network that has quite a few frames of coefficients as an input and produces subsequent probabilities over HMM states as an output is revealed.

Kingbury et al. [5] presented the history of the development of the deep neural networks for acoustic models for speech recognition. The overview summary focused on the different ways that can be utilized to improve deep learning, which was classified into five different categories.

Deng et al. [6] conducted a summary on the work done by Microsoft since the year 2009 in the area of speech using deep learning. The paper focused on more recent advances which helped shed some light on the different capabilities as well as limitations of deep learning in the area of speech recognition.

Ma et al. [7] presented the basics of the state of the art solutions for automatic spoken language recognition for both, computational and phonological perspectives. Huge progress was achieved in recent years in the area of spoken language recognition which was mostly directed by breakthroughs in relevant signal processing areas such as pattern recognition and cognitive science. Several main aspects relevant to language recognition was discussed.

Li et al. [8] provided an overview on modern noise robust techniques for automatic speech recognition developed over the past three decades. More emphasis was given on the techniques that have proven successful over the years and are likely to maintain and further expand in their applicability in the future. The examined techniques were categorized and evaluated using five different criteria.

Yogesh Kumar et al. [9] review the different aspects related to Automatic Speech recognition. They have elaborated the recent advancement in the speech recognition system, robust method for the development of an automatic speech recognition system and application of automatic speech recognition system in different fields.

Ram Paul et al. [10] provide a brief review of different DSP (Digital Signal Processing) based techniques applied for speech recognition. In this paper speaker recognition system are discussed.

Arul Valiyavalappil Haridasa et al. [11] a survey of speech recognition strategies suitable for human identification is discussed in this study. In this review, diverse issues included in speech recognition methodologies is distinguished and distinctive speech recognition procedures were studied to discover which qualities is tended to in a given system and which is disregarded.

FrankSeide et al. [12] proposed Context-Dependent Deep Neural-Network HMMs, or CD-DNN-HMMs, to speech-to-text transcription. CD-DNN-HMMs combine classic artificial-neural-network HMMs with traditional tied-state triphones and deep-belief network pre-training. Akshi Kumar et al. [13] a survey is provided on the application of three deep learning architectures in the field of speech recognition, namely, Deep Belief Networks, Convolutional Neural Networks and Recurrent Neural Networks.

Ossama Abdel-Hamid et al. [14] show that error rate reduction can be obtained by using convolutional neural networks (CNNs). It presents a concise description of the basic CNN and explain how it can be used for speech recognition. It proposes a limited-weight-sharing scheme that can better model speech features.

Akhilesh Halageri et al. [15] review the pattern matching abilities of neural networks on speech signal. Speech recognition involves capturing and digitizing the sound waves, converting them to basic language units or phonemes, constructing words from phonemes, and contextually analysing the words to ensure correct spelling for words that sound alike.

Rubi et al. [16] defined a three stage neural integrated model speech signal enhancement and use the decomposition integrated HMM model for speech feature transformation. For the feature extraction of speech Discrete wavelength transform (DWT) has been used which gives a set of feature vectors of speech waveform. The work has been done on MATLAB and experimental results show that system is able to recognize words at sufficiently high accuracy.

Kaisheng Yao et al. [17] evaluate the effectiveness of adaptation methods for context-dependent deep-neural-network hidden Markov models (CD-DNN-HMMs) for automatic speech recognition.

Alex Graves et al. [28] investigates deep recurrent neural networks, which combine the multiple levels of representation that have proved so effective in deep networks with the flexible use of long range context that empowers RNNs.

III.PROPOSED FRAMEWORK

The input is a speech file. The speech file is inputted to the speech recognition system. The speech recognition system converts the speech to text using different deep learning techniques. It produces the transcribed text as output. Then the performance metric which is the word error rate is computed for each technique. Based on the performance metric the best deep learning technique is chosen. The deep learning technique for which the performance metric is minimum is chosen as the best method for the given type of speech.

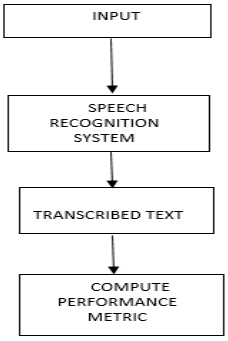


Fig. 1 Proposed Framework

IV. CONCLUSION

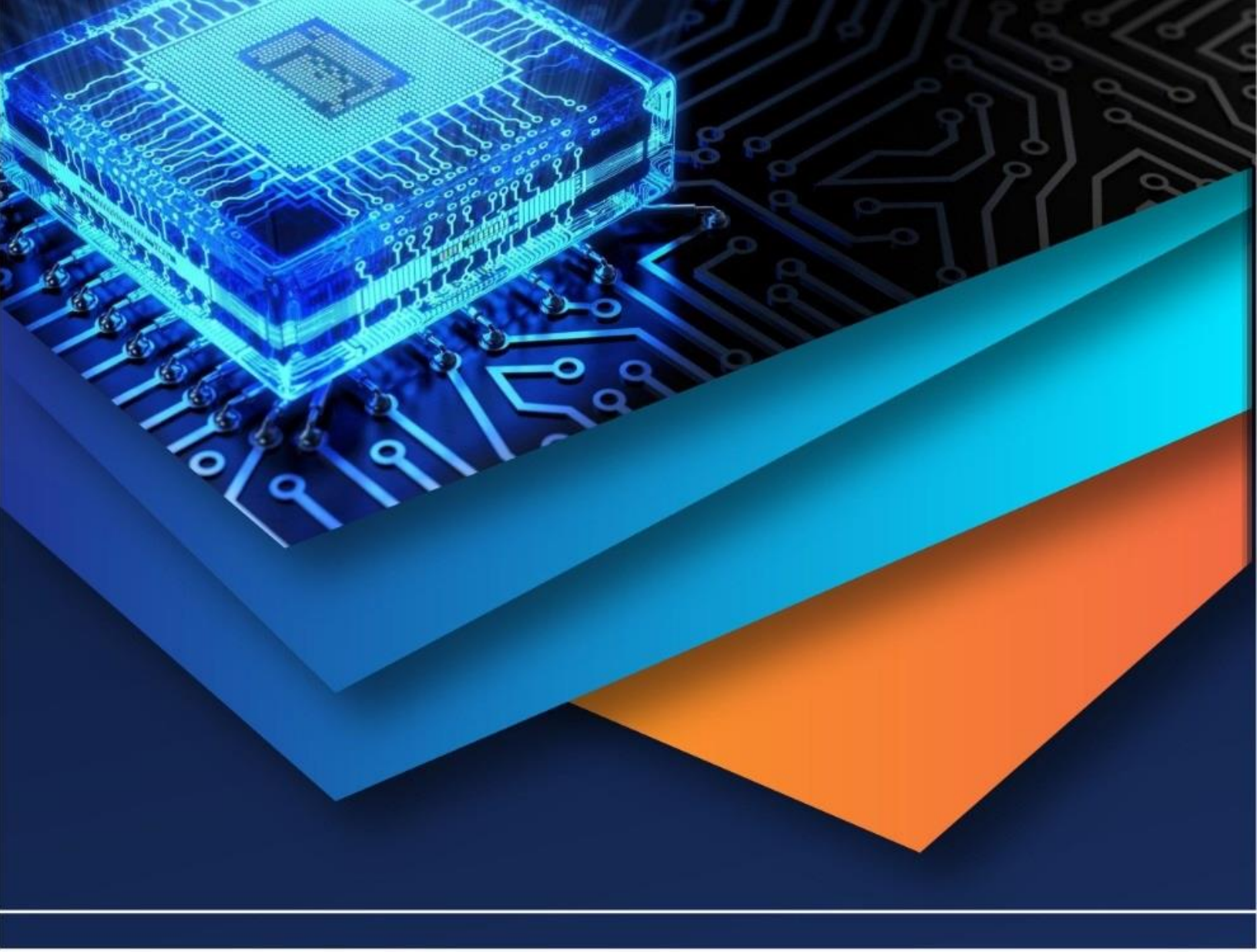
The most important way for humans to communicate with each other and acquire information is with the help of speech. Among various speech processing problems, automatic speech recognition (ASR) for converting recorded speech automatically to text is one of the most challenging tasks. In the future we can work on any of the language using different improved approaches that gives better results than existing work done on it.

V. ACKNOWLEDGMENT

We are grateful to the anonymous reviewers for their valuable suggestions.

REFERENCES

[1] Karpagavalli S and Chandra E," A Review on Automatic Speech Recognition Architecture and Approaches", International Journal of Signal Processing, Image Processing and Pattern Recognition Vol.9, No.4, 2016, pp.393-404, 2016.
[2] Ali Bou Nassif, Ismail Shahin, Imtinan Attili, Mohammad Azzeh, Khaled Shaalan, "Speech Recognition Using Deep Neural Networks: A Systematic Review", IEEE, vol. 7, 2019.
[3] N. Morgan," Deep and wide: Multiple layers in automatic speech recognition", IEEE Trans. Audio, Speech, Lang. Process., vol. 20, no. 1, pp. 7–13, 2012.
[4] G. Hinton et al.," Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups", IEEE Signal Process. Mag., vol. 29, no. 6, pp. 82–97, 2012.
[5] L. Deng, G. Hinton, and B. Kingsbury," New types of deep neural network learning for speech recognition and related applications: An overview", in Proc. IEEE Int. Conf. Acoust., Speech Signal Process., pp. 8599–8603, 2013.
[6] L. Deng et al.," Recent advances in deep learning for speech research at Microsoft", in Proc. IEEE Int. Conf. Acoust., Speech Signal Process., pp. 8604–8608, 2013.
[7] H. Li, B. Ma, and K. A. Lee," Spoken language recognition: From fundamentals to practice", Proc. IEEE, vol. 101, no. 5, pp. 1136–1159, 2013.
[8] J. Li, L. Deng, Y. Gong, and R. Haeb-Umbach," An overview of noise- robust automatic speech recognition", IEEE/ACM Trans. Audio, Speech,Language Process., vol. 22, no. 4, pp. 745–777, 2014.
[9] Yogesh Kumar, Dr. Navdeep Singh," A Comprehensive View of Automatic Speech Recognition System - A Systematic Literature Review", International Conference on Automation, Computational and Technology Management (ICACTM), IEEE, 2019.
[10] Ram Paul, Rajender Kr. Beniwal, Rinku Kumar, Rohit Saini," A Review on Speech Recognition Methods",International Journal on Future Revolution in Computer Science Communication Engineering Volume: 4 Issue: 2, 2018.
[11] Arul Valiyavalappil Haridas, Ramalatha Marimuthu, Vaazi Gangadharan Sivakumar," A critical review and analysis on techniques of speech recognition: The road ahead", International Journal of Knowledge-based and Intelligent Engineering Systems 22, 39–57, 2018.
[12] FrankSeide, GangLi, DongYu, "Conversational Speech Transcription Using Context-Dependent Deep Neural Networks", ISCA, 2011.
[13] Akshi Kumar, Sukriti Verma, Himanshu Mangla, "A Survey of Deep Learning Techniques in Speech Recognition", International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), IEEE, 2018.
[14] Ossama Abdel-Hamid, Abdel-rahman Mohamed, Hui Jiang, Li Deng, Gerald Penn, Dong Yu, "Convolutional Neural Networks for Speech Recognition", IEEE/ACM Transactions On Audio, Speech, and Language Processing, VOL.22, NO.10, 2014.
[15] Akhilesh Halageri, Amrita Bidappa, Arjun C, Madan Mukund Sarathy, Shabana Sultana, "Speech Recognition using Deep Learning", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (3), 3206-3209, 2015.
[16] Rubi, Chhavi Rana, "Review on Speech Recognition with Deep Learning Methods", International Journal of Computer Science and Mobile Computing, Vol.4 Issue.8, pg. 301-307, 2015.
[17] Kaisheng Yao, Dong Yu, Frank Seide, Hang Su, Li Deng, Yifan Gong, "Adaptation of Context-Dependent Deep Neural Networks for Automatic Speech Recognition", IEEE, 2012.
[18] Alex Graves, Abdel-Rahman Mohamed, Geoffrey Hinton, "Speech Recognition with Deep Recurrent Neural Networks", IEEE, ICASSP, 2013.
[19] R. Lawrence and B.-H. Juang," Fundamentals of Speech Recognition", Prentice-Hall, Inc., (Engelwood, NJ), 1993.
[20] Lim Sin Chee, Ooi Chia Ai, Sazali Yaacob," Overview of Automatic Speech Recognition System", Proceedings of the International Conference on Man-Machine Systems (ICoMMS) 11 – 13, 2009.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY