



Detailed Survey of Speech Recognition using Machine Learning Algorithms

Dr. H S Prasantha

Professor

Department of Computer Science and Engineering
K.S. Institute of Technology, Bangalore, India

Abstract: Speech recognition is the process of converting human sound signals into words or instructions. The research of speech recognition involves many subject areas such as computer technology, artificial intelligence, digital signal processing, pattern recognition, acoustics, linguistics, and cognitive science. Our speech is made up of many frequencies at the same time. The actual signal is really a sum of all those frequencies stuck together. The conversation or speech that is captured by a microphone or a telephone is converted from acoustic signal to a set of words in speech recognition.

Keywords- speech, machine learning, recognition

1. Introduction

The structure of the speech recognition system includes a feature extraction algorithm, acoustic model, and language model and search algorithm. The major components and topics within the space of ASR are:

- feature extraction
- acoustic modelling
- pronunciation modelling
- language modelling
- Hypothesis search.

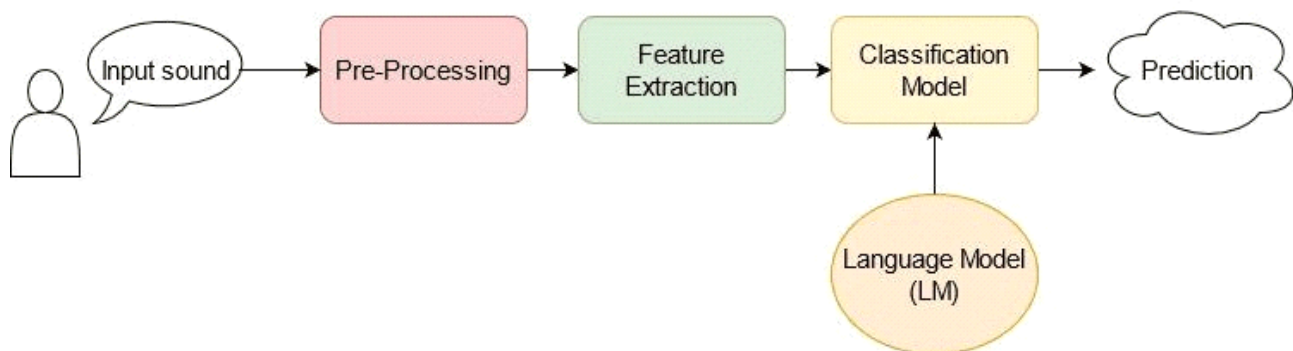


Figure 1: Speech Recognition Structure

Speech recognition has become a practical concept, which is now being implemented in different languages around the world and it is used in real-world human language applications, such as information recovery. Speech processing has a close relationship with computer linguistics, human-machine interaction, natural language processing, and psycholinguistics. Speech recognition can be extended to recognize speakers, exploiting the information present in the speech and various methods including exploiting from the excitation source.

2. SPEECH PROCESSING USING MACHINE LEARNING:

The human brain, like machine learning technology, is essential for speech recognition to interact with machines to humans. The machine learning methodology is used in a lot of assignments through the feature learning capability. The data modelling capability results attained supplementary than the performance of normal learning methodology. So, the speech signal recognition is based on a machine-learning algorithm to merge the speech features and attributes. As a result of voice as a bio-metric implication, the speech signal is converted into a significant element of speech improvement.

Machine learning consists of supervised and unsupervised learning among which supervised learning is used for the speech recognition objectives. Machine learning (ML) software can make measurements of spoken words through a set of numbers that represent the speech signal. There are several acoustic modelling machine learning techniques like Hidden Markov Model (HMM), Gaussian Mixture Model (GMM), Deep Neural Network (DNN), KNN algorithm, Naive Bayse algorithm.

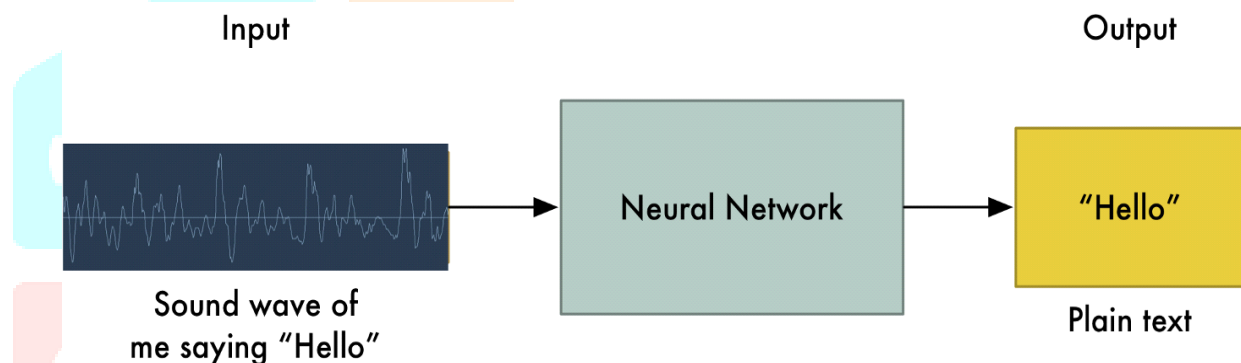


Figure 2: Sound Wave detection using a Deep Neural Network

3. LITERATURE SURVEY

TITLE	AUTHORS	ML METHODOLOGY	REMARKS
Machine Learning Paradigms for Speech Recognition: An Overview	Li Deng and Xiao Li, IEEE	Maximum likelihood linear regression(MLLR), Hidden Markov model (HMM), Bayesian-sensing HMM	To develop automatic systems capable of generalizing from previously observed examples, by constructing or learning functional dependencies between arbitrary input and output domains.
Machine Learning in Automatic Speech Recognition: A Survey	JayashreePadmanabhan& Melvin Jose Johnson Premkumar	Gaussian mixture models, Hidden Markov models, Machine learning, Support vector machines, SVM/HMM hybrid systems, MLP/HMM systems,	Markov models used in prediction problems. determining the probability of the model in generating the sequence. Determining the most likely state sequence that generates the outcome is decoding, using the Viterbi algorithm; updating the model parameters for

			maximizing the likelihood of occurrence is training, adopted by Baum-welch algorithm.
Applications of Support Vector Machines to Speech Recognition	AravindGanapathiraju, Jonathan E. Hamaker, Joseph Picone	Support vector machines(SVM), aussian mixture models(GMM), Hidden Markov models(HMM), Expectation–Maximization (EM) algorithm, Empirical risk minimization (ERM)	SVM automatically controls generalization and parameterization as part of the overall optimization process. The Switchboard (SWB) task, which is based on two-way telephone recordings of conversational speech, is very different from the AD task in terms of acoustic confusability and classifier complexity
Speech Recognition for COVID-19 Keywords Using Machine Learning	Wael Ben Amara, Amani Touihri, Salma Hamza.	Artificial neural network (ANN),Support Vector Machine (SVM).	The proposed method in this research is a way to detect individuals carrying the virus based on the list of symptoms they provide, but in practice this should be done very precisely in order to avoid false outputs and especially false negatives.
An Evaluation on Speech Recognition Technology based on Machine Learning	Tribhuwan Kumar, Rajesh K, KalyanChakravarthi, SumanRajest	Speech Recognition with RNN.	The speech signal recognition is based on a machine-learning algorithm to merge the speech features and attributes. As a result of voice as a bio-metric implication, the speech signal is converted into a significant element of speech improvement. The experimental result shows the improved proposed speech recognition algorithms accuracy with another state of the art method.
Machine Learning Based Speech Emotions Recognition System	Dr. Yogesh Kumar, Dr. Manish Mahajan	Support Vector Machine (SVM) classifier. Convolution Neural network (CNN) classifier. K-Nearest Neighbours (KNN).	Performance analysis based on different machine learning techniques for different languages. In the paper, the detailed review on KNN, SVM, CNN classifier for speech emotion recognition system is used. The overall aim is to work on the emotions generated by humans using the above mentioned algorithms.

Speech Recognition Using Machine Learning	VineetVashisht, Satya Prakash yadav, Aditya Kumar pandey	Speech computer or speech synthesizer, text-to-speech(TTS) system,Neural Machine Translation(NMT)	Voice detection with real-time predictive voice translation device optimization using multimodal vector sources of information and functionality was presented. The key production and commitment of this work is the manner in which external information input is used to increase the system's accuracy, thereby allowing a notable improvement, compared to the processes of nature. In addition, a new initiative has been launched from an analytical standpoint, while remaining a realistic one, and was discussed.
Convolutional Neural Networks for Speech Recognition	Ossama Abdel-Hamid, Abdel-rahman Mohamed, Hui Jiang, Li Deng, Gerald Penn, and Dong Yu	Hybrid ANN-HMM framework, Convolution Neural network (CNN) classifier.	This paper describes how to apply CNNs to speech recognition in a novel way, such that the CNN's structure directly accommodates some types of speech variability. This Hybrid CNN-HMM approach delegates temporal variability to HMM, while convolving along the frequency axis creates a degree of invariance to small frequency shifts, which normally occur in actual speech signals due to speaker differences.
“Controlling Email System Using Audio with Speech Recognition and Text to Speech” 2021	Harsh D Shah, Amit Sundas, Shabnam Sharma.	A hidden Markov model (HMM)	The proposed system is intended to build a framework for control email over voice and reduce the effort of manual typing the mail along with saving time. The powerful advantage of the HMM approach combined with MFCC features is better suited to these criteria and provides strong recognition results
“Sign Languages to Speech Conversion Prototype	Malli Mahesh Chandra,RajkumarS,LakshmiSutha Kumar.	Support vector machines (SVMs)	In this paper the user can speak English or some Indian languages through the gestures using proposed prototype. Around 22 gestures in ASL and

using the SVM Classifier”			11 gestures in ISL are trained and tested successfully. An accuracy of 100% is achieved for ISL database with 25% test data and 75% training data. And an accuracy of 98.91% is achieved for ASL database with 25% test data and 75% training data
---------------------------	--	--	--

4. TOOLS

- Python programming higher installed in windows 10 or any distribution of Linux.
- Arduino IDE software
- Hardware:MPU6050 sensor
- HC05 Bluetooth module.
- Training of an ANN model using Coding Source in MATLAB.
- Python open source library called TensorFlow1 is used in order to train the neural network.

5. APPLICATIONS

- The first successful use of DNN-based acoustic models for large vocabulary continuous speech recognition (LVCSR) employed Bing voice search data. The system obtained a sentence accuracy of 69.6% on the test set, compared to 63.8% obtained by a GMM/HMM baseline.
- Used in Speech emotion recognition (SER) which is a type of speech recognition whose purpose is to establish a speaker’s underlying emotional state by analyzing their voice.
- Popular apps such as Amazon’s Alexa, Apple’s Siri and Google Maps employ speech recognition.
- Voice search: a digital assistant to help surf the web and search through to help accomplish different tasks.
- Smart home devices leverage speech recognition technology to carry out household tasks, such as, turning on the lights, boiling water, adjusting the thermostat, and more.
- Speech recognition enables hands free computing. Its use cases include, but are not limited to: Writing emails, Composing a document on Google Docs, Automatic closed captioning with speech recognition (i.e. YouTube), Automatic translation, And sending texts.
- Healthcare: Doctors and nurses leverage dictation applications to capture and log patient diagnoses and treatment notes.
- Speech recognition technology has a couple of applications in sales. It can help a call center transcribe thousands of phone calls between customers and agents to identify common call patterns and issues.
- Security: As technology integrates into our daily lives, security protocols are an increasing priority. Voice-based authentication adds a viable level of security.

6. REFERENCES

1. Deng, L., & Li, X. (2013). “*Machine Learning Paradigms for Speech Recognition: An Overview.*” IEEE Transactions on Audio, Speech, and Language Processing, 21(5), 1060–1089.
2. Padmanabhan, J., & Johnson Premkumar, M. J. (2015). “*Machine Learning in Automatic Speech Recognition: A Survey.*” IETE Technical Review, 32(4), 240–251.
3. Ganapathiraju, A., Hamaker, J., & Picone, J. (2004). “*Applications of Support Vector Machines to Speech Recognition.*” IEEE Transactions on Signal Processing.
4. Wael Ben Amara, Amani Touihri2, Salma Hamza "Speech Recognition for COVID-19 Keywords Using Machine Learning" Vol.8, Issue.4, pp.51-57, August (2020).

5. Tribhuwan Kumar, Rajesh K, Kalyan Chakravarthi, Suman Rajest "An Evaluation on Speech Recognition Technology based on Machine Learning" Webology, Volume 19, Number 1, January, 2022.
6. Dr. Yogesh Kumar, Dr. Manish Mahajan "Machine Learning Based Speech Emotions Recognition System" VOLUME 8, ISSUE 07, JULY 2019.
7. Vineet Vashisht, Satya Prakash yadav, Aditya Kumar pandey. "Speech Recognition Using Machine Learning". Article in IEIE Transactions on Smart Processing and Computing · June 2021
8. Ossama Abdel-Hamid, Abdel-rahman Mohamed, Hui Jiang, Li Deng, Gerald Penn, and Dong Yu , "Convolutional Neural Networks for Speech Recognition", IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 22, NO. 10, OCTOBER 2014
9. Harsh D Shah , Amit Sundas, Shabnam Sharma," Controlling Email System Using Audio with Speech Recognition and Text to Speech". 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO) Amity University, Noida, India. Sep 3-4, 2021
10. Chandra, M. M., Rajkumar, S., & Kumar, L. S. (2019). "Sign Languages to Speech Conversion Prototype using the SVM Classifier." TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON).
11. Dr. H S Prasantha, "NOVEL APPROACH FOR IMAGE COMPRESSION USING MODIFIED SVD", International Journal of Creative Research Thoughts (IJCRT), Volume 8, Issue 8, Page 2234-2243, Aug 2020
12. Dr. H S Prasantha, "IMPLEMENTATION OF IMAGE COMPRESSION USING FAST COMPUTATION OF SVD ON DM642", International Journal of Creative Research Thoughts (IJCRT), Volume 8, Issue 8, Page 2364-2368, Aug 2020
13. Prasantha, H, H Shashidhara, K N B Murthy, and M Venkatesh. "Performance Evaluation of H.264 Decoder on Different Processors." International Journal on Computer Science & Engineering. 1.5 (2010): 1768. Web. 7 Apr. 2013.
14. H. S. Prasantha, H. L. Shashidhara, and K. N. Balasubramanya Murthy. Image compression using SVD. In Proceedings of the International Conference on Computational Intelligence and Multimedia Applications, pages 143–145. IEEE Computer Society, 2007.
15. Gunasheela K S, H S Prasantha, "Compressive sensing for image compression: survey of algorithms", Proceedings of Emerging Research in Computing, Information, Communication and Applications, ERCICA, Springer publication, Bengaluru, 2018
16. K N Shruthi, B M Shashank, Y. Sai Krishna Saketh, H.S Prasantha and S. Sandya, "Comparison Analysis Of A Biomedical Image For Compression Using Various Transform Coding Techniques", IEEE, pp. 297-303, 2016