iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** V **Month of publication:** May 2020

DOI: http://doi.org/10.22214/ijraset.2020.5328

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 V May 2020- Available at www.ijraset.com

Review on Speech Recognition using Deep Learning

Deepankar Dayal[1], Firoz Alam[2], Himanshu Varun[3], Neha Singh[4]
[1, 2, 3]*B. Tech Student,* [4]*Assistant Professor, Department Of Computer Science Engineering, RKGIT Ghaziabad, Uttar Pradesh*

Abstract: ***Speech recognition is the method of translating spoken words into text. The speech recognition process digitizes the sound waves into basic language units. Speech recognition is one of the most used technologies in today's life. This technology can be seen everywhere around a person, for example in phones, games, etc. The main purpose of the paper is to know the knowledge and the technology behind this superb invention.***
Keywords: Speech Recognition, Deep Learning, Machine Learning, Neural Networks, Hidden Markov Model, Synapses.

I. INTRODUCTION

What is Speech Recognition? The speech recognition process is the method to identify words and to convert them into machine-readable units. There are many speech recognition applications that are in use today, such as simple data entry, voice dialing, speech-to-text, etc.
The speed recognition is another method through which humans can interact with the machine, the example for the same is giving input to the computer using a keyboard.
The first attempt to build automatic speech recognition was made in the 1950s. Deep learning consists of a multiple machine learning algorithms fed with input in the form of multiple layered models. One of the early applications of deep learning was speech. The way words are spoken can be altered by accents and mannerisms depending upon the area where people live, but with the help of deep learning we will be able to remove this alteration of accent and mannerism. In this paper we will learn about how deep learning is being used worldwide to make more and more use of Speech Recognition.

II. REVIEW

A. *Hidden Markov Model*

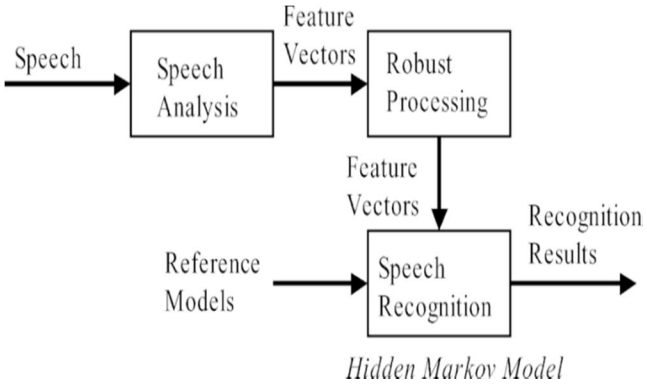


Fig.2.1 Hidden Markov Model

The existing systems for ASR use the complex statistical models which are the HIDDEN MARKOV MODEL. The HMM has been successful models in the field of deep learning.
HMM is called hidden because only the symbols emitted by the system are observable and not the underlying random walk between states. They output a sequence of symbols or quantities. Speech signals are stationary signals, hence, HMM is used in speech recognition. They are popular because they are simple and feasible when it comes about their use.
The disadvantages of these systems are:
They are statistically inefficient for modeling data that lie on or near nonlinear manifold in the data space.
1) It is expensive in terms of memory as well as compute time.
2) They cannot express dependencies between hidden states.
3) There are many possible HMMs for a given set of seed sequence and hence choosing one can be difficult

B. Suggested Method

The suggested system is to use learning algorithms without any assumptions. The algorithms using neural networks have been very successful for such type. For assumptions free algorithms, the neural networks are a great choice.

The advantages of neural networks are:

1) Can learn without any help.
2) In neural networks, the input is stored in their networks.
3) It can detect fault and still produce output.
4) Multi-tasker.

C. Viability

The only requirement is training data. In fact, many personal devices have transitioned from visual controlled interfaces to speech controlled interfaces in order to reduce cost as speech controlled interfaces have rapid growth.

Why to use deep learning in speech recognition?

1) Better representation with fewer parameters.
2) Deep learning made speech recognition possible to be use by people.
3) Deep learning can model high-dimensional features.

III. ARCHITECTURE OF ASR

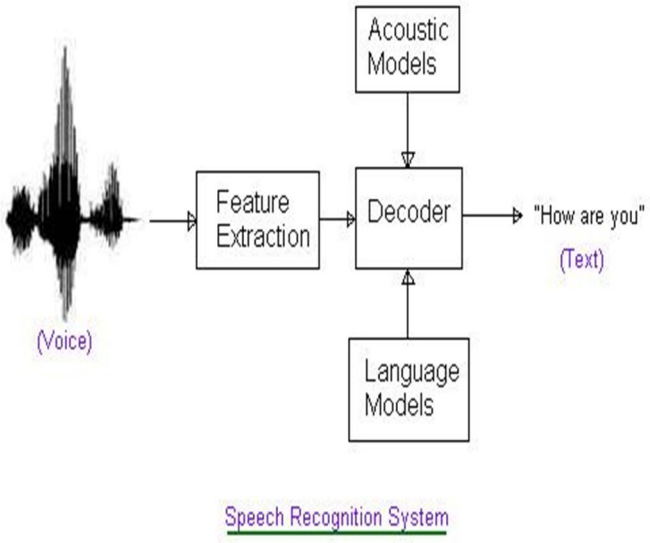


Fig.3.1 Architecture Of ASR

IV. IMPLEMENTATION

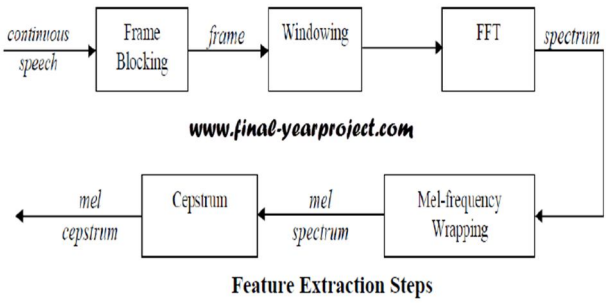


Fig.4.1 Feature Extraction

A. Signal Preprocessing

We all know that the sound produced by a human being depends upon his/her vocal tract. It is the shape of the vocal tract that determines what sound has to come out. If the shape of the vocal tract determined correctly, this can give us an accurate phenome. The job of the MFCC is to accurately represent the envelope of the short-time power spectrum.

Feature extraction is the process of obtaining different features such as power, pitch, etc. from the speech signal.

The features are extracted as follows:

1) Signals are framed into short frames.
2) Applying a hamming window.
3) Apply filter bank to the power spectra.
4) Mel frequency wrapping.
5) Creating a window of adjacent frames to capture phenome.

B. Neural Networks

1) Architecture of Neural Network

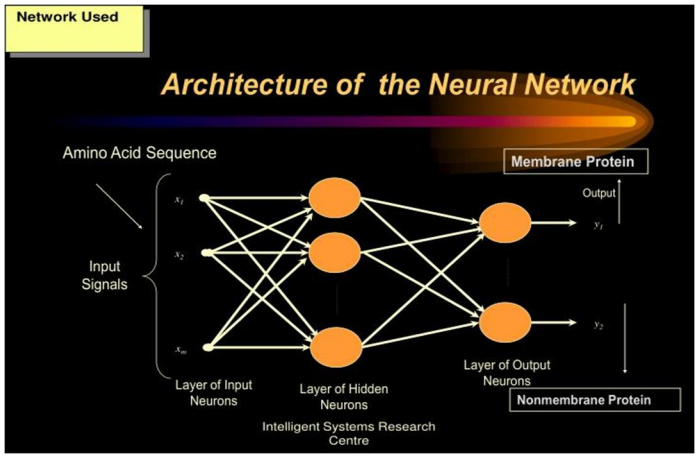


Fig.4.2 Architecture Of Neural Network

2) What are "Neural Networks"?

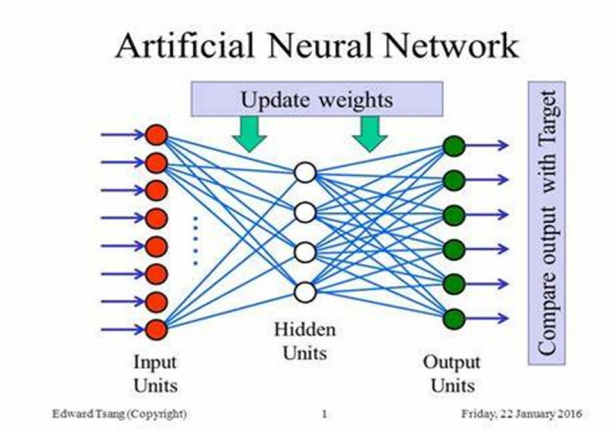


Fig.4.3 Layers Of Neural Network

"Neural networks" are computational algorithm that stimulates the behavior of biological systems composed of neurons. It is an algorithm that is based on the model of a human's neuron. The human brain consists of millions of neurons that help in processing signals in the form of electrical and chemical signals. The neurons are connected with a special structure known as synapses. Synapses allow neurons to pass the signals. Neural network can recognize a pattern. It corresponds to dendrites and synapses. Each arc in the network associated with a weight at each node.

A neural network is a technique that works the same way human brains processes information. They have large number of processing units that work parallel to each other and create meaningful outputs.

Neural network consists of three layers:

1) *Input Layer:* It helps in representing the raw information that is fed into the network.

2) *Hidden Layer*

a) It determines the activity of every hidden unit.

b) The activities of the input units and the weights depend upon the connections between input and hidden units**.**

c) There may be multiple hidden layers.

3) *Output Layer:* The conduction of the output units depend upon the connection between the input and hidden units as well as hidden and output units.

The Basic characteristics of neural networks are:

a) Processing unit

b) Connections

c) Computing procedure

d) Training

i) *Processing Unit:* A neural network contains a large number of processing units similar to the neurons in the human brain. All the units in the network operate at the same time. At every moment each unit computes a scalar function of its input and produces outputs.

ii) *Connections:* The units in a network are organized into a topology by a set of connections shown as lines (weights) in the diagram. Every weight has a real value, although sometimes the value is limited. The value of the weight describes the influence of a unit on its neighbor. The positive causes one unit to excite another while negative to inhibit another. Weight can change due to training but they change very slowly. The weights can be represented as lines (can be shown in diagram). The lines (weights) are connected from input units to hidden units and hidden units to output units.

A neural network can be connected with any kind of topology which is best suited for a particular type of application. For example:

- Unstructured networks
- Layered networks
- Recurrent networks
- Modular networks

iii) *Training:* It is the most general sense that means adapting its connections so that the network exhibits desired computational behavior to its all inputs. The process involves modifying the weights and sometimes modifying the actual topology. The topological change helps in improving the speed of the learning of network, by constraining the class of the function and this can be controlled by adjusting the learning rate and momentum.

iv) *Hidden Layer:* The hidden layer is the layer between the input and output algorithm. In the hidden layer, the function applies weight to the input and produces the output of the desired input. The hidden layer performs a nonlinear transformation of the inputs which enter the network. They are called hidden because they are not possible to be seen from the inputs or the outputs.

V. ADVANTAGES OF NEURAL NETWORKS

A. Neural networks have the ability to learn complex structures.

B. Discriminative naturally.

C. Modular in design

D. Can be mixed into larger systems/structures.

E. Can be used easily with techniques like HMMs.

VI. APPLICATIONS OF DEEP LEARNING

A. Speech recognition

B. Character recognition

C. Signature verification application

D. Human face recognition

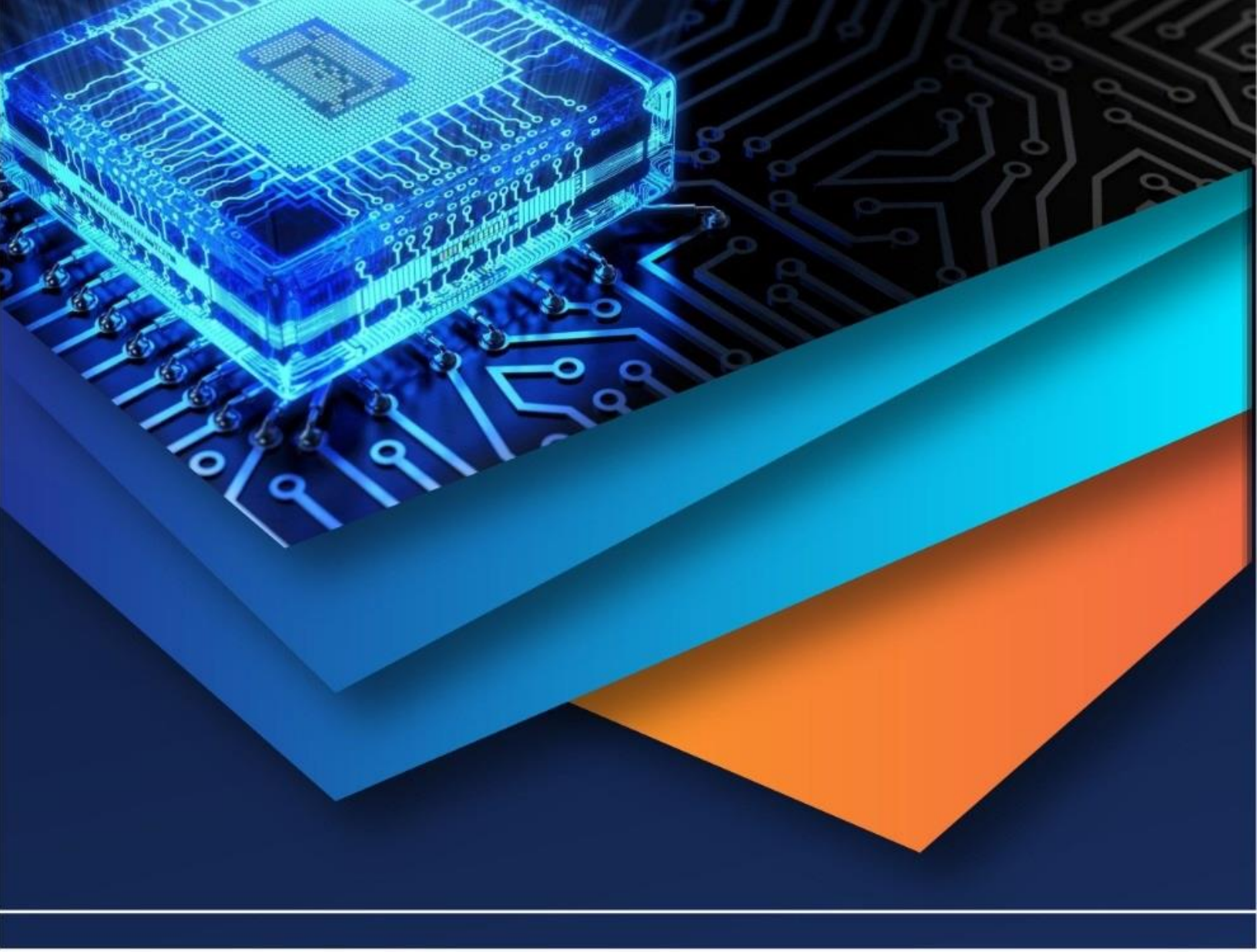
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com

VII. CONCLUSION

This paper tells the use and importance of using neural networks in place of models like HMM for speech recognition. Neural networks are suitable for predicting series because of learning only from examples, without any additional information. A combination of probabilistic neural network and recurrent neural network recognizes 98% of phonemes correctly, followed by HMM. HMM still dominate the field, but many researchers have begun to explore ways in which neural networks can enhance the accuracy of HMM-based systems. There is no second thought in the fact that the speech recognition is one of the most used and useful inventions in today's time.

REFERENCES

[1] Speech recognition using neural networks, Joe Tebelskis, May 1995, CMU-CS-95-142
[2] Speech recognition – A Deep Learning Approach, Dong Yu, Li Deng, Microsoft Research, ISBN 978-1-4471-5779-3
[3] Wikipedia
[4] Speech recognition using deep learning algorithms, Yan Zhang,
[5] S. Young, Large Vocabulary Continuous Speech Recognition: A Review, IEEE Signal Processing Magazine, volume 13, no 5, pp. 45-57,1996
[6] Google



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY