International Journal of Multidisciplinary Research and Development

Volume: 2, Issue: 6, 226-229
June 2015
www.allsubjectjournal.com
e-ISSN: 2349-4182
p-ISSN: 2349-5979
Impact Factor: 3.762

Dhavale Dhanashri
Affiliated to savitribai phule Pune University Department of Electronics, AISSMS-IOIT, Pune, Maharashtra, India.

S.B. Dhonde
Department of Electronics, AISSMS-IOIT, Pune, Maharashtra, India

Correspondence:
Dhavale Dhanashri
Affiliated to savitribai phule Pune University Department of Electronics, AISSMS-IOIT, Pune, Maharashtra, India.

Speech Recognition Using Neural Networks: A Review

Dhavale Dhanashri, S.B. Dhonde

Abstract
In this review paper firstly we will look after the types of neural networks and their introduction. Also the hybrid architecture of HMM and NN is also studied. Developments in the field of neural network will be discussed. Deep neural networks are mostly used for ASR systems. They give better performance as compare to traditional GMM. When used in hybrid architecture with HMM, deep neural networks give better performance as compared to HMM-GMM system.

Keywords: Speech recognition system, Neural network, Feedforward neural network, Recurrent neural network

1. Introduction
For communication purpose, speech is considered to be the most easiest way. For recognizing humans by analyzing their voices and understanding others, speech is mostly used. It is the useful interface to interact with machines. The process of converting speech signals into words with the help of an algorithm which is known as computer programme is called as speech recognition [1] To understand the human voice and languages for computer speech recognition technology proves to be very useful. There are three approaches to the speech recognition namely template based, knowledge based, stastical based approach [20].

Template-based approaches, in this approach speech is compared against number of templates. After that the best match is found out which is the required output. In this approach errors due to segmentation can be avoided, but as the recorded templates are fixed, variations in speech can be modeled by many templates which again becomes impractical. This is the main disadvantage of this approach.

Knowledge-based approaches, Knowledge based approach uses the information regarding linguistic, phonetic and spectrogram. In this approach we can model the speech variations but it is difficult to obtain. so this approach was judged to be impractical, and automatic learning procedures were sought instead.

Statistical-based approaches, in which variations in speech are modeled statistical using automatic learning procedures. Hidden markov model is used. The statistical models prepared make a priori modeling assumptions, which are liable to be inaccurate to the systems performance. This disadvantage of approach can be removed by using neural network [20, 1].
People understand the speech well. Depend on this, many speech and speaker recognition systems have been developed. Having long history in speech recognition, neural networks are mostly used in acoustic model [19]. The purpose of this review paper is to understand the proper usage of neural networks in the field of speech processing. In this review paper, different types of neural network methods that are used for speech recognition are explained [17]. Also it gives the basic idea about why they used and what are they. In recent years most of the work has been done in the field of neural networks. Results shows that neural networks are proved to be performance improving in case of speech recognition systems [16]. This paper basically deals with the different types of neural networks and hybrid approach of HMM and NN model for speech recognition.

2. Neural Network in the Field Of Speech Recognition
Neural network is mathematical model which is used to perform a particular function. It works like a human brain. Computers are trained by using machine learning algorithms to perform tasks by their own. There is need to focus on some properties of neural networks. It include

Trainability: Networks are trained to form association between input layer and output layer. This type of ability can be used to train the network how to classify speech patterns into phoneme categories**.**

Generalization. Networks don't just memorize the training data; rather, they learn the underlying patterns, so they can generalize from the training data to new examples. This is essential in speech recognition, because acoustical patterns are never exactly the same.

Nonlinearity. Networks can compute nonlinear, nonparametric functions of their input, enabling them to perform arbitrarily complex transformations of data. This is useful since speech is a highly nonlinear process.

Robustness. Networks are tolerant of both physical damage and noisy data; in fact noisy data can help the networks to form better generalizations. This is a valuable feature, because speech patterns are notoriously noisy.

Uniformity. Networks offer a uniform computational paradigm which can easily integrate constraints from different types of inputs. This makes it easy to use both basic and differential speech inputs, for example, or to combine acoustic and visual cues in a multimodal system.

Parallelism. Networks are highly parallel in nature, so they are well-suited to implementations on massively parallel computers. This will ultimately permit very fast processing of speech or other data [20]. Artificial neural network is the approach used for machine learning. These machine learning algorithms leads to many improvements in the field of speech recognition [10]. Artificial neural networks, these are the systems consisting of nodes (neurons)interconnected with each other. They works similar to the human brain [17]. Artificial neural network contains many processing elements connected which influence each others behavior via network of weight [20]. Each unit in the neural network computes a nonlinear weighted sum of its input and pass it over to the other units through the outgoing connections [20].

2.1 Different types of neural networks

There are different types of neural networks all over researchers searched. In order to map the complex inputs into simple outputs neural networks are used. They perform static pattern recognition for example such as an N-array classification of the input patterns [20].

A. Feedforward neural networks

It is the one way connection without back loop is used. It has only connections forward in time. Suppose a neuron is in layer a then it can only send data to neuron in layer b if b>a. The layers which are adjacent to each other can be connected together as in multilayer perceptrons Also there are shortcuts between the layers which are not adjacent [17]. In this one input is associated with one output so it is called as static. Feed forward connections are used by a time delay network which is to be used in classification of data with weighted delays [17]. There is a fixed weight mapping from inputs to outputs in feedforward networks as the weights of a feedforward neural network are fixed after training. It is confirmed that the state of any neuron is determined by its input and output pattern and not by its initial state or past state. It indicates that it is a static and no dynamics are involved. It is the most popular neural network used today.

B. Perceptrons and multi-layer perceptrons

It is one of the type of feed forward neural network. A perceptron is a simple neuron model that consists of set of inputs, weights regarded each input and the activation functions. Neuron performs the activated function to the weighted sum of inputs before sending the value to its output [17]. The perceptron model is shown in Figure 1, where *x* is an input vector, *w* is a weight vector and the activation function is a step function.

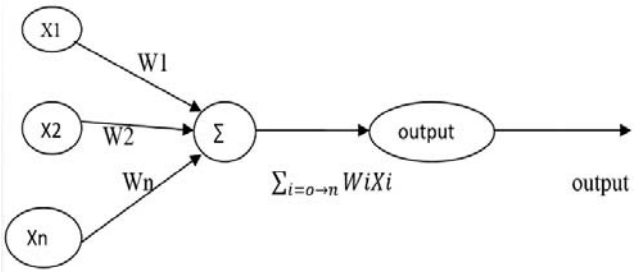

Fig 1: Model of perceptron

A multilayer perceptron has at least two layers of perceptron. It contains input layer, one or more hidden layer and output layer. Hidden layer works as a feature extractor. It uses nonlinear function like sigmoid or radial basis function to generate input [17]. The hidden layer consists of non-linear sigmoidal activation function neurons [19]. The number of neurons present in the hidden layer depends on the amount of input data, no of neurons in output layer, the needed generalization capacity of the network and size of the training set [19]. The output of all the neurons in the hidden layer act as a input to the next layer.

C. Recurrent Neural Network

In this type of neural network output of neuron is multiplied by a weight and fed bact to the neuron including delay [17]. Recurrent neural networks (RNNs) are a powerful model for sequential data [18]. Here the state of neuron is the input and the previous state of neuron itself. RNNs are inherently deep in time, since their hidden state is a function of all previous hidden states. In order to train RNNs, end to end training methods like connectionist temporal classification is used. It is for sequence labeling problems where the input-output alignment is unknown [18]. As compared to MLP, RNNs have achieved better performance in speech recognition. The training algorithm used is more complex and are sensitive which can cause problem [17].

D. Self-organizing maps

Self-organizing maps are the technique of converting high dimensional space to a smaller dimensional space. Because of this the input vectors which are close to each other corresponds to the neurons which are close to each other in map. There is a code vector associated with each neuron in the network which points to a corresponding neuron in the map. Competitive learning is used to train the network. These maps are used in speech recognition tasks mostly [17].

3. Learning

By training neural network, the network classifies data well and not overleam details of the training data. There are two ways to train a network either by supervised way or unsuperwised way. In superwised way the network is given a set of labeled data for learning and in case of unsuperwised way the task of the network is to find clusters of data that are

similar [17]. Previous instantiations of the neural network approach have used the backpropagation algorithm to train the neural networks discriminatively [6]. In the next sections we will see them in detail.

3.1 Supervised learning

In this learning, the training data is classified first using speech recognition system or manually. After that the network is trained by using this data to compute the data of classification. The network iteratively changes its weights to minimize a given cost function *E*. In error back propagation algorithm first the error is calculated at output and then it is send back to network and partial derivatives of error are calculated. Each neuron is updated and then new iteration starts. This algorithm is used for updates [17].

3.2 Unsupervised learning

In unsupervised learning there is no need to define the target output. It tries to figure out the underlying pattern or trend in the input data alone. Here no labeled data is given to the network. Instead of this there are some similarity measures like cosine distance that it uses to find the input vectors whose distance according to similarity measure is small.

4. Hybrid HMM/NN Models

Combination of hidden Markov Model and Neural Network works as an alternative paradigm for ASR started between 1980s.In hybrid NN-HMM model each output unit of NN is trained to estimate the posterior probability of a continuous density HMM's state given the acoustic observations [7]. Use of combination of HMM and NN for speech recognition gives better results than GMM.As comparing to GMM, neural networks gives the same performance but require smaller amount of parameters [17]. Most of the work on the hybrid approach used context-independent phone states as labels for NN training and considered small vocabulary tasks. ANN-HMMs were later extended to model context-dependent phones and were applied to mid-vocabulary and some large-vocabulary ASR tasks [7]. There are some limitations for this hybrid approach. By using only backpropagation to train the network makes it challenging to exploit more than two hidden layers well. In the language processing field and speech recognition neural networks are used widely. There have been numerous applications of neural networks in these fields. Neural networks particularly deep networks with many hidden layers are capable of modeling complex structures [9]

There are three main reasons which were responsible for the use of neural networks as high-quality acoustic models: (1) making the networks deeper makes them more powerful, hence deep neural networks (DNN); 2) initializing the weights sensibly and using much faster hardware makes it possible to train deep neural networks effectively, and 3) using a larger number of output units greatly improves their performance [12]. As compared to other networks, deep neural networks have higher modeling capacity with the same number of parameters. But deep neural networks are harder to train, both as stochastic top-down generative models and as deterministic bottom-up discriminative models [9]. The DNN architecture can be used for multi-task learning in several different ways and DNNs are far more effective than GMMs at leveraging data from one task to improve performance on related tasks [12].

5. Conclusion

This paper is showing that the neural networks are the most important in the field of speech recognition. They came with a new solution for large database. It is an interesting research area. Neural networks act like a human brain. In this paper we gave overview of the types of the neural networks and hybrid architecture of HMM and NN model. Hybrid architecture of HMM and NN works well for the acoustic model of speech recognition. Later in that deep neural networks are involved. They work well in noise also. Research is going on in this field to find out some more facts about neural network.

6. References

1. 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,pp.181-205, 2009
2. Bo Li and Khe Chai Sim, "A Spectral Masking Approach to Noise-Robust Speech Recognition Using Deep Neural Networks", IEEE/ACM Transactions On Audio, Speech, And Language Processing, Vol. 22, No. 8, pp. 1296-1305, AUGUST 2014
3. Y. Bengio, "Learning deep architectures for AI," Foundat. and Trends Mach. Learn., vol. 2, no. 1, pp. 1–127, 2009
4. Xiaohui Zhang, Jan Trmal, Daniel Povey, Sanjeev Khudanpur, "Improving Deep Neural Network Acoustic Models Using Generalized Maxout Networks",IEEE Interrnational Conference On Acoustic, Speech and Signal, pp 214-219, 2014
5. Xicai Yue, Datian Ye, Chongxun Zheng, Xiaoyu Wu, "Neural networks for improved text independent speaker identification", IEEE Engineering In Medicine And Biology, pp 53-58, April 2002
6. Abdel-rahman Mohamed, George E. Dahl, and Geoffrey Hinton, "Acoustic Modeling Using Deep Belief Networks", IEEE Transactions On Audio, Speech, And Language Processing, Vol. 20, No. 1, pp. 14-22, JANUARY 2012
7. George E. Dahl, Dong Yu,Li Deng,and Alex Acero, "Context-Dependent Pre-Trained Deep Neural Networks for Large-Vocabulary Speech Recognition", IEEE Transactions On Audio, Speech, And Language Processing, Vol. 20, No. 1, pp, 30-42, JANUARY 2012
8. Ke Chen, Ahmad Salman, "Learning Speaker-Specific Characteristics with a Deep Neural Architecture", IEEE Transactions On Neural Networks, Vol. 22, No. 11, pp 1744-1756, November 2011
9. Ruhi Sarikaya, Geoffrey E. Hinton, and Anoop Deoras, "Application of Deep Belief Networks for Natural Language Understanding", IEEE/ACM Transactions On Audio, Speech, And Language Processing, Vol. 22, No. 4, pp 778-784, April 2014
10. Geoffrey Hinton, Li Deng, Dong Yu, George E. Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior, Vincent Vanhoucke, Patrick Nguyen, Tara N. Sainath, and Brian Kingsbury, "Deep neural networks for acoustic modeling in speech recognition", IEEE Signal Processing Magazine, pp 82-97, November 2012
11. Li Deng, Geoffrey Hinton, and Brian Kingsbury, "NEW TYPES OF DEEP NEURAL NETWORK LEARNING FOR SPEECH RECOGNITION AND RELATED APPLICATIONS: AN OVERVIEW", IEEE Publication, pp 8599-8603, 2013
12. Jonas Gehring, Wonkyum Lee, Kevin Kilgour, Ian Lane, Yaije Miao, Alex Waibel, "Modular Combination of Deep Neural Networks for Acoustic Modeling", INTERSPEECH 2013

13. Veera Ala-Keturi, “Speech Recognition Based on Artificial Neural Networks”, Helsinki hnology Institute of Technology, 2004
14. Alex Graves, Abdel-rahman Mohamed and Geoffrey Hinton, “SPEECH RECOGNITION WITH DEEP RECURRENT NEURAL NETWORKS”, ICASSP, pp 6645-6649, 2013
15. Wouter Gevaert, Georgi Tsenov, Valeri Mladenov, “ Neural Networks used for Speech Recognition”, Journal Of Automatic Control, University Of Belgrade, Vol. 20, pp 1-7, 2010
16. Joe Tebelskis, “Speech Recognition using Neural Networks”, May 1995