Speech Recognition Using Neural Networks

Manvendra Singh[1] Kamal Verma[2]

Digital Communication, RIET, Jaipur, Rajasthan, India
manvendra.hcst@gmail.com[1], kkverma_99@yahoo.com[2]

Abstract: **-Neural networks have seen an explosion of interest over the last few years, and are being successfully applied across an extraordinary range of problem domains, in areas as diverse as finance, medicine, engineering, geology and physics. They are used in areas ranging from robotics, speech, signal processing, vision, and character recognition to musical composition, detection of heart malfunction and epilepsy, and many more. In our paper, we have made an attempt towards illustrating the application of neural networks in Speech Recognition. Although, speech recognition products are already available in the market at present, their development is mainly based on statistical techniques which work under very specific assumptions. We elaborate the feasibility of an alternative approach for solving the problem more efficiently, in this paper.**

Keywords - Speech, Neural Networks

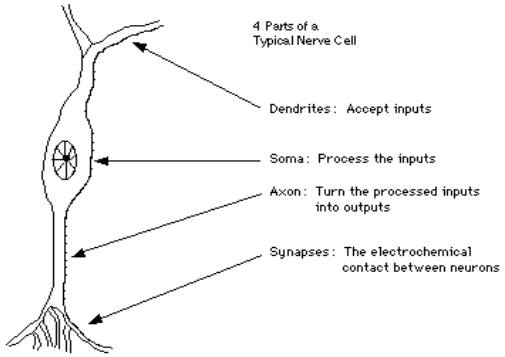
I. INTRODUCTION

A. Artificial Neural Network: -

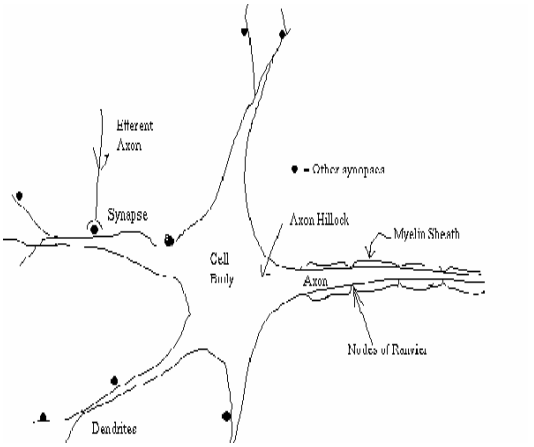
An Artificial Neural Network (ANN) is an information processing paradigm that is inspired by the way biological nervous systems, such as the brain, process information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems. ANNs, like people, learn by example. An ANN is configured for a specific application, such as pattern recognition or data classification, through a learning process. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons. This is true of ANNs as well [1].

B. Analogy to Brain: -

Much is still unknown about how the brain trains itself to process information, so theories abound. In the human brain is composed of a very large number (circa 10,000,000,000) of *neurons*, a typical neuron collects signals from others through a host of fine structures called *dendrites*. The neuron sends out spikes of electrical activity (electro chemical signal) through a long, thin stand known as an *axon*, which splits into thousands of branches.



At the end of each branch, a structure called a *synapse* converts the activity from the axon into electrical effects that inhibit or excite activity from the axon into electrical effects that inhibit or excite activity in the connected neurons. When a neuron receives excitatory input that is sufficiently large compared with its inhibitory input, it sends a spike of electrical activity down its axon. Learning occurs by changing the effectiveness of the synapses so that the influence of one neuron on another changes.



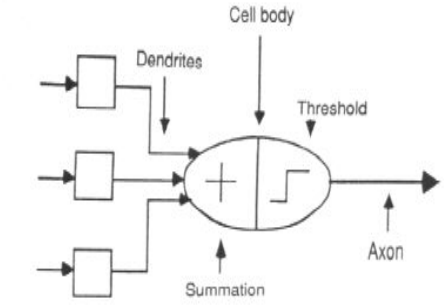
C. From Human Neurons to Artificial Neurons

To capture the essence of biological neural systems, an artificial neuron is defined as follows [6]:

It receives a number of inputs (either from original data, or from the output of other neurons in the neural network). Each input comes via a connection that has a strength (or *weight*); these weights correspond to synaptic efficacy in a

biological neuron. Each neuron also has a single threshold value. The weighted sum of the inputs is formed, and the threshold subtracted, to compose the *activation* of the neuron (also known as the post synaptic principle, or PSP, of the neuron).

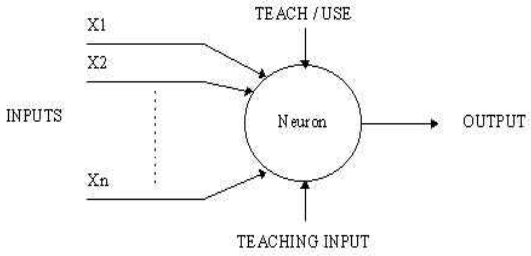
The activation signal is passed through an activation function (also known as a transfer function) to produce the output of the neuron [2].



II. WORKING

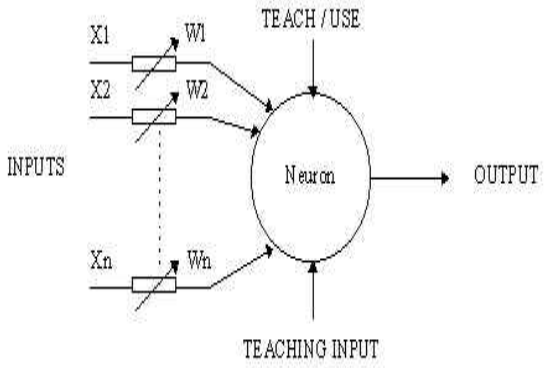
a. A simple neuron:-

An artificial neuron is a device with many inputs and one output. The neuron has two modes of operation; the training mode and the using mode. In the training mode, the neuron can be trained to fire (or not), for particular input patterns. In the using mode, when a taught input pattern is detected at the input, its associated output becomes the current output. If the input pattern does not belong in the taught list of input patterns, the firing rule is used to determine whether to fire or not [3].



b. An engineering approach:-

The previous neuron doesn't do anything that conventional computers don't do already. A more sophisticated neuron is the McCulloch and Pitts model (MCP). The difference from the previous model is that the inputs are 'weighted'; the effect that each input has at decision making is dependent on the weight of the particular input.[4] The weight of an input is a number which when multiplied with the input gives the weighted input. These weighted inputs are then added together and if they exceed a pre-set threshold value, the neuron fires. In any other case the neuron does not fire.



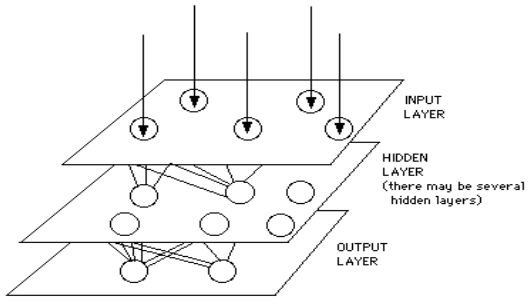
In mathematical terms, the neuron fires if and only if;

$$X1W1 + X2W2 + X3W3 + ... > T$$

The addition of input weights and of the threshold makes this neuron a very flexible and powerful one. The MCP neuron has the ability to adapt to a particular situation by changing its weights and/or threshold. Various algorithms exist that cause the neuron to 'adapt'; the most used ones are the Delta rule and the back error propagation. The former is used in feed-forward networks and the latter in feedback networks [5].

III. NETWORK LAYERS

The commonest type of artificial neural network consists of three groups, or layers, of units: a layer of "input" units is connected to a layer of "hidden" units, which is connected to a layer of "output" units. (see Figure below)



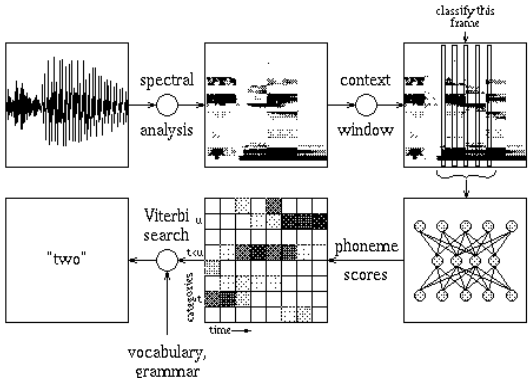
1. The activity of the input units represents the raw information that is fed into the network.
2. The activity of each hidden unit is determined by the activities of the input units and the weights on the connections between the input and the hidden units.
3. The behavior of the output units depends on the activity of the hidden units and the weights between the hidden and output units.

IV. EVALUATION FROM OUTSIDE

The synergism of web and phone technologies has led to a new innovative voice web network. The

voice web requires a voice recognition and authentication system incorporating a reliable speech recognition technique for secure information access across the Internet [7].

Today speech recognizers deal with more than 50000 words, fluently spoken [8]. They are adaptable to the speaker if not speaker-independent. There are commercially available "dictate systems" (Dragon Systems, Kurzweil) which need powerful PCs to accomplish their tasks (Pentium processor, Windows or 6MB RAM). In this paper, we use an alternative approach which is based on neural networks bound to be simpler and economic.



The target-word pronunciations are then expanded into strings of phonetic-based categories, and a Viterbi search is used to find the best path through the matrix of probabilities for each legal string. The output of recognition is the word string that corresponds to this best path [9].

V. APPLICATIONS

Basically, most applications of neural networks fall into the following five categories:

a. *Prediction:* -Uses input values to predict some output. e.g. pick the best stocks in the market, predict weather, identify people with cancer risk.

b. *Classification:* -Use input values to determine the classification. e.g. is the input the letter A, is the blob of the video data a plane and what kind of plane is it.

c. *Data association:* -Like classification but it also recognizes data that contains errors. e.g. not only identify the characters that were scanned but identify when the scanner is not working properly.

d. *Data Conceptualization:* -Analyze the inputs so that grouping relationships can be inferred. e.g. extract from a database the names of those most likely to by a particular product.

e. *Data-Filtering:* -Smooth an input signal. e.g. take the noise out of a telephone signal.

6. CONCLUSION

Neural networks, with their remarkable ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques.

Our innovative technique for speech recognition is good enough to prove that quantizing

speech into frames suffices up to the mark. The human speech is an inherently dynamical process that can be properly described as a trajectory in a certain feature space.

The new approach developed for training the neural network's architecture proved to be simple and very efficient. It reduced considerably the amount of calculations needed for finding the correct set of parameters. If the traditional approach had been used instead, the amount of calculations would have been higher.

ACKNOWLEDGEMENT

We would like to thank RIET Faculty Members & Management Team besides it we would like to specially thank to Mr. Ashok Sirohi Head of ECE Dept. at RIET Jaipur.

REFERENCES

[1] Fausett L., *Fundamentals of Neural Networks*, Prentice-Hall, 1994. ISBN 0 13 042250 9 or

[2] Gurney K., *An Introduction to Neural Networks*, UCL Press, 1997, ISBN 1 85728 503 4

[3] Haykin S., *Neural Networks* , 2nd Edition, Prentice Hall, 1999, ISBN 0 13 273350 1 is a more detailed book, with excellent coverage of the whole subject.

[4] W. Hess, "Pitch and voicing determination," Adv. Speech Signal Process., pp. 3–48, 1992.

[5] W. Hess, Pitch Detemination of Speech Signals. Berlin,Germany: Springer-Verlag, 1983.

[6] V. Emiya, B. David, and R. Badeau, "A parametric method for pitch show that F-HMUSIC is more robust against colored noise than estimation of piano tones," in Proc. IEEE Int. Conf. Acoust., Speech, HMUSIC and WLS. The price that is paid for computational Signal Process., 2007, pp. 249–252.

[7] J. Luo, P.-C. Zai, and Y.-N. Jian, "Fault diagnosis of power transformer based on ellipsoidal basis functional neural network," in Proc. IEEE Wavelet Anal. Pattern Recognit., 2007, pp. 695–698.

[8] A.Asuncion andD.Newman,UCIMachine Learning Repository, Schl. Inf. Comput. Sci., Univ. California, Irvine, CA, 2007.

[9] C. M. Bishop, Neural Networks for Pattern Recognition. Oxford, U.K.: Oxford Univ. Press, 1995.