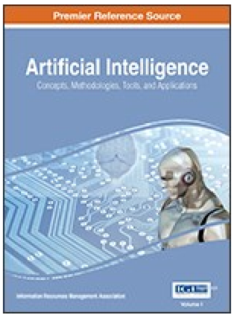

Artificial Intelligence: Concepts, Methodologies, Tools, and Applications, Volume I

by Information Resources Management Association (IRMA)
IGI Global. (c) 2017. Copying Prohibited.

Chapter 1: Definition of Artificial Neural Network

Sara Moein,
Washington University in Saint Louis, USA

ABSTRACT

In living creatures, the brain is the control unit and it can be divided in different anatomic and functional sub-units. An artificial neural network is a computational system for processing information as a response to external stimuli, which consists of a set of highly interconnected processing elements called neurons. It is very useful to have some knowledge of the way the biological nervous system is organized, since the artificial neural network is an inspiration of the biological neural networks. This chapter is an explanation of the Artificial Neural Network (ANN). The biological and mathematical definition of a neural network is provided and the activation functions effective for processing are listed. Some figures are collected for better understanding.

1 BIOLOGICAL NEURAL NETWORK

In the nervous system of the living creatures, there are fluid-filled sacs which bound by a lipid bilayer for separating the intracellular contents from the extracellular space and they are called Neurons, or brain cells. Inside the body, neurons are responsible to maintain a negative internal voltage, which is related to the extracellular space; ion channels and pumps maintain this potential difference. In most neurons of the central nervous system, spike is responsible to send the signals of neural activity, or rapid intracellular depolarization followed by repolarization; in order to adjust the neurons, it is necessary to communicate information about a neuron's activity. Some neurons communicate with simple resistive coupling, via channels that allow direction flow. However, for higher animals, most neurons in the central nervous system (CNS), communicate through chemical synapses: triggering the release of chemicals using the neural spike is called neurotransmitters into the extracellular space. These neurotransmitters bind to ion channels in adjacent neurons, causing a brief ionic current to flow into the neuron. The resulting current flow in the recipient neuron will be depolarizing, or hyperpolarizing and it depends on whether the neurotransmitter is excitatory or inhibitory, respectively.

It is very useful to have some knowledge of the way the biological nervous system is organized, since the artificial neural network is an inspiration of the biological neural networks.

There is difference among the nervous system of creatures. Most living creatures, which have the ability to adapt to a changing environment, need a controlling unit, which is able to learn. Higher developed animals and humans use very complex networks of highly specialized neurons to perform this task.

In the living creatures, the brain is the control unit and it can be divided in different anatomic and functional sub-units. Each unit is responsible to do certain tasks like vision, hearing, motor and sensor control. The brain is connected by nerves to the sensors and actors in the rest of the body.

There are a large number of neurons, about 10^{11} in average. This can be considered as the basic building bricks for the central nervous system (CNS). The connection points of the neurons are called synapses. The complexity of the brain is because of the massive number of highly interconnected simple units working in parallel, with an individual neuron receiving input from up to 10000 others (Bishop, 1995).

The structure and its processes in a simple cell are enormously complex. Even the `dendrites' that originating from the cell body are thin and widely branching fibers, reaching out in different directions and make connections to a larger number of cells within the cluster. Axons make the connections from the other cells to dendrites of one cell or directly to the body of the cell. There is only one axon per neuron. It is a single and long fiber, which transports the output signal of the cell as electrical impulses (action potential) along its length. The end of the axon may divide in many branches, which are then connected to other cells. The branches have the function to fan out the signal to many other inputs (Hassoun, 1995).

The structure of neuron has four main regions. There are two offshoots from the cell body, the dendrites, and the axon, which end in presynaptic terminals. The heart of the cell is called the cell body that contains the nucleus and maintaining protein synthesis. A neuron has many dendrites in a treelike structure, which branch out, and receive signals from other neurons. Each neuron has one axon that grows out from a part of the cell body called the axon hillock. The axon hillock generates the electrical signals to the axon down its length. These electric signals are called action potentials. The other end of the axon may split into several branches, which end in a presynaptic terminal. Action potentials are the electric signals that neurons use to convey information to the brain. All these signals are identical.

There are many different types of neuron cells found in the nervous system. The differences are due to their location and function. The basic performing function of the neuron is the following: all the inputs to the cell with various strength of the connection and various frequency of the incoming signal are summed up. The input sum is processed by a threshold function and produces an output signal.

The brain has the ability of processing in both a parallel and serial ways. The parallel and serial nature of the brain is readily apparent from the physical anatomy of the nervous system. The serial and parallel processing can be easily seen from the time needed to perform tasks. For example a human can recognize a picture of a person is about 100 ms. Given the processing time of 1 ms for an individual neuron this implies that a certain number of neurons, but less than 100, are involved in serial; whereas the complexity of the task is evidence for a parallel processing, because a difficult recognition task cannot be performed by such a small number of neurons. This phenomenon is known as the 100-step-rule (Schmidt, 2000)

Structurally the neuron can be divided in three major parts: the cell body (soma), the dendrites, and the axon, see Figure 1 and Figure 2.

Therefore, the brain determines what type of information is being received based on the path that the signal took. The brain analyzes the patterns of signals being sent and from that information it can interpret the type of information being received. The fatty tissue that surrounds and insulates the axon is called 'myelin'. Often short axons do not need this insulation.

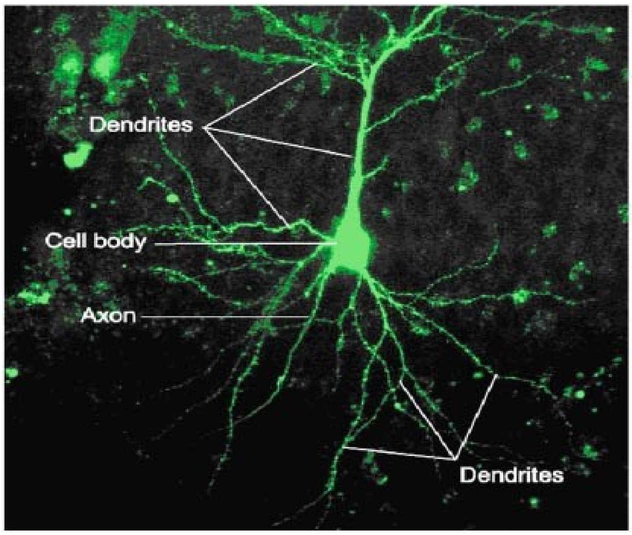

Figure 1: A biological neuron

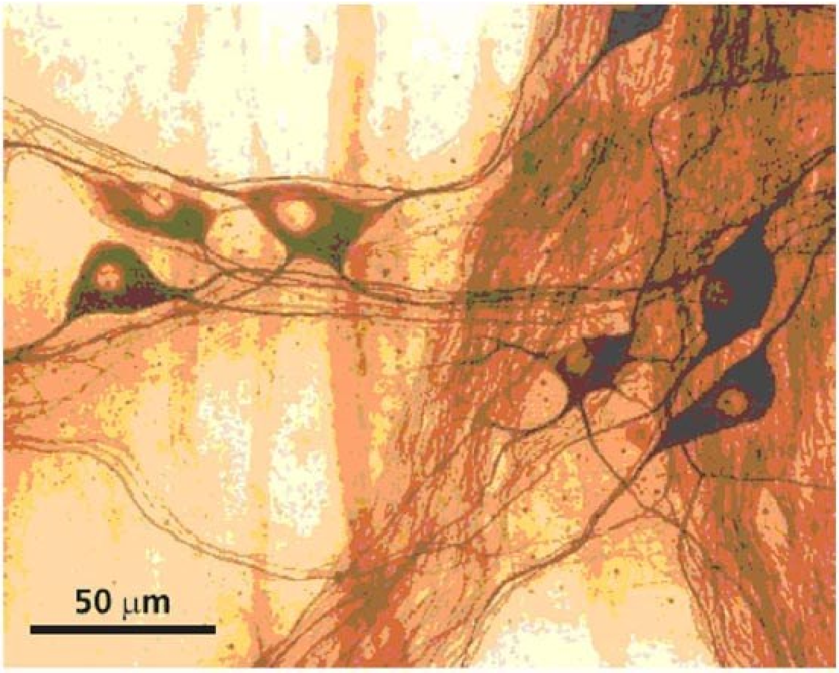

Figure 2: The main processes of the individual cells

The synapse is the area of contact between two neurons. The neurons do not actually physically touch. They are separated by the synaptic cleft, and electric signals are sent through chemical 13 interaction. The neuron sending the signal is called the presynaptic cell and the neuron receiving the signal is called the postsynaptic cell. The signals are generated by the membrane potential, which is based on the differences in concentration of sodium and potassium ions inside and outside the cell membrane. There are two categories to classify the neurons (Goyal & Harmsen, 2013). The first one is by their number of processes (or appendages), and the second one is by their function. If they are classified by the number of processes, they fall into three categories:

- **Unipolar Neurons:** That have a single process (dendrites and axon are located on the same stem), and are most common in invertebrates.
- **Bipolar Neurons:** The dendrite and axon are the neuron's two separate processes. Bipolar neurons have a subclass called pseudo-bipolar neurons, which are used to send sensory information to the spinal cord.
- **Multipolar Neurons:** Are most common in mammals. Examples of these neurons are spinal motor neurons, pyramidal cells and Purkinje cells (in the cerebellum).

If classified by function, neurons again fall into three separate categories:

- **Sensory, or Afferent Neurons:** Which provide information for perception and motor coordination.
- **Group of Providing Information (or Instructions):** To muscles and glands and is therefore called motor neurons.
- **Interneuronal:** Contains all other neurons and has two subclasses. One group called relay or projection interneurons have long axons and connect different parts of the brain. The other group called local interneurons are only used in local circuits.

One of the characteristics of the biological neural systems is that usually have a very high fault tolerance. Various experiments with people with brain injuries have shown that damage of neurons up to a certain level does not necessarily influence the performance of the system, though tasks such as writing or speaking may have to be learned again. This can be regarded as re-training the network.

2 DEFINITION OF ARTIFICIAL NEURAL NETWORK

An artificial neural network is a computational system consisting of a set of highly interconnected processing elements, which are called neurons for processing information as a response to external stimuli. The simplistic representation for emulating the signal integration and threshold firing behavior of biological neurons is called an artificial neuron. An artificial neuron uses the mathematical equations to emulate the behavior of a biological neuron. Like their biological counterpart, artificial neurons are bound together by connections that determine the flow of information between peer neurons. Stimuli are transmitted from one processing element to another via synapses or interconnections, which can be excitatory or inhibitory. If the input to a neuron is excitatory, it is more likely that this neuron will transmit an excitatory signal to the other neurons connected to it. Whereas an inhibitory input will most likely be propagated as inhibitory.

Artificial neural networks are the mathematical model for computational paradigms with a structure and operation that resembles that of the mammal brain. Other names of artificial neural networks or in short neural networks are connectionist systems, parallel distributed systems or adaptive systems, because they are composed by a series of interconnected processing elements that operate in parallel. There is lack of centralized control in neural networks in the classical sense, since all the interconnected processing elements change or "adapt" simultaneously with the flow of information and adaptive rules. One of the important aims of neural networks (ANN) is to shape the functional characteristics and computational properties of the brain in performing the cognitive processing tasks such as sensorial perception, concept categorization, concept association and learning. However, recent studies focused on the development of neural networks for applications such as pattern recognition and classification, data compression and optimization.

Complexification in artificial neural networks can prove to be as important, as it is in the development of natural neural systems. It is important in artificial development to unleash fitness potential otherwise left untouched and constrained by a fixed neural topology. Complexification in neural networks is a vital process in the development of the brain in any natural system. Complexification in human brains happens in several different ways, by growth, by pruning and by reorganization. The first form of complexification happens from before birth and goes on up to adulthood, as the brain is formed. During this period neurons and interconnections grow and hence complexifies the network. The second form of complexification happens through continuous pruning.

Connections between neurons have to be used for them not to fade away and eventually possibly disappear. This concept is called neural Darwinism, as it is similar to normal evolution; where the fittest, in this case connections, survive. The third form of complexification happens through reorganization. In some cases, for yet unknown reasons, connections detach themselves from neuron and reconnects to another. Mostly, reorganization in natural systems have a detrimental effect, but some might have unexpected positive effects (Hassoun, 1995). In the literature, a wide variety of definitions and explanations for the terms Artificial Neural Network (ANN) and Neural Computing can be found. The following definitions are balanced towards computing but are nevertheless very comprehensive in my opinion and they offer a wide range of views of what an ANN is.

The definition by Igor Aleksander includes a very wide range of methods and applications in the field of neural computing:

> *Neural computing is the study of networks of adaptable nodes, which through a process of learning from task examples, store experimental knowledge and make it available for use. (Aleksander & Morton, 1995)*

There are two important features to simulate the intelligent behavior the abilities of memorization and generalization are essential. These are basic properties of artificial neural networks. Table 1 shows the definitions according to the Collins English Dictionary.

Memorizing is an obvious task in learning. It can be done by identifying the concept behind the input data, or by storing the input samples explicitly, and memorizing their general rules. The ability to identify the rules, to generalize, allows the system to make predictions on unknown data. Despite the strictly logical invalidity of this approach, the process of reasoning from specific samples to the general case can be observed in human learning.

The following description by Laurene Fausett of artificial neural networks includes only the connectionist research approach (Fausett, 1994).

"An artificial neural network is an information-processing system that has certain performance characteristics in common with biological neural networks. Artificial neural networks have been developed as generalizations of mathematical models of human cognition or neural biology, based on the assumption that:

- Information processing occurs at many simple elements called neurons.
- Signals are passed between neurons over connection links.
- Each connection link has an associated weight, which, in a typical neural net, multiplies the signal transmitted.

- Each neuron applies an activation function (usually nonlinear) to its net input (sum of weighted input signals) to determine its output signal."

Robert L. Harvey focuses very much on the biological model. His definition excludes most parts of logical neural networks from the field of neural networks:

> *An artificial neural network is a dynamical system with one-way interconnections. It carries out the processing by its response to inputs. The processing elements are nodes; the interconnections are directed links. Each processing element has a single output signal from which copies fan out*.

Table 1: Definitions for two basic features in intelligent behavior

To memorize	To commit to memory; learn so as to remember.
To generalize	To form general principles or conclusions from detailed facts, experience, etc.

It can be thought that the field of neural networks is related to artificial intelligence, parallel processing, machine learning, statistics, and other fields. The attraction of neural networks is that they are best suited to solving the problems that are the most difficult to solve by traditional computational methods.

Partridge et al. (1996) listed several potential of neural network over conventional computation and manual analysis:

- Implementation using data instead of possibly ill defined rules.
- Noise and novel situations are handled automatically via data generalization.
- Predictability of future indicator values based on past data and trend recognition.
- Automated real-time analysis and diagnosis.
- Enables rapid identification and classification of input data.
- Eliminates error associated with human fatigue and habituation.

One advantages of generalization is to remove the need to store a large number of input samples. Features common to the whole class need not to be repeated for each sample - instead the system needs only to remember which features are part of the sample. This can significantly reduce the amount of memory needed, and produce a very efficient method of memorization.

Neural Network (NN) is an AI technique with the capability of learning from a set of training data and constructs weight matrixes to represent the learning patterns. NN is a network of many simple processors or units (Sarle, 1999). It simulates the function of human brain to perform tasks as human does. As an example, a study on approximation and classification in medicine with incremental neural network shows superior generalization performance compared with other classification models (Jankowski, 1999). NN has been employed in various medical applications such as coronary artery (Lippmann, 1995), Myocardial Infarction (Heden, Ohlsson, Rittner, Pahlm, Haisty, Peterson, & Edenbrandt, 1996), cancer (Street, Mangasarian & Wolberg, 1996; Karkanis, Magoulas, Grigoriadou, & Schurr, 1999) and brain disorders (Pranckeviciene, 1999). In Karkanis et al. (1999) NN was implemented as a hybrid with textual description method to detect abnormalities within the same images with high accuracy.

3 MATHEMATICAL MODEL OF ARTIFICIAL NEURAL NETWORK

There are three basic components of importance for creating a functional model of the biological neuron.

- The synapses of the neuron are modeled as weights. The strength of the connection between an input and a neuron is noted by the value of the weight. Negative weight values reflect inhibitory connections, while positive values designate excitatory connections.
- The second component model the actual activity within the neuron cell. An adder sums up all the inputs modified by their respective weights. This activity is referred to as linear combination.
- An activation function controls the amplitude of the output of the neuron. An acceptable range of output is usually between 0 and 1, or -1 and 1 (Hassoun, 1995).

In the artificial neural network, each unit performs a relatively simple job: receive input from neighbors or external sources and use this to compute an output signal, which is propagated to other units. Apart from this processing, a second task is the adjustment of the weights. The system is inherently parallel in the sense that many units can carry out their computations at the same time. Within neural systems, it is useful to distinguish three types of units: input units, which receive data from outside the neural network, output units, which send data out of the neural network, and hidden units whose input and output signals remain within the neural network. During operation, units can be updated either synchronously or asynchronously. With synchronous updating, all units update their activation simultaneously; with asynchronous updating, each unit has a (usually fixed) probability of updating its activation at a time *t*, and usually only one unit will be able to do this at a time. In some cases, the latter model has some advantages.

Mathematically, this process is described in Figure 3.

From this model, the interval activity of the neuron can be shown to be:

(1) $$V_k = \sum_{j=1}^{p} w_{kj} xj$$

3.1 Activation Function

As mentioned previously, the activation function acts as a squashing function, such that the output of a neuron in a neural network is between certain values (usually 0 and 1, or -1 and 1). In general, there are three types of activation functions, denoted by Φ(.). First, there is the Threshold Function, which takes on a value of 0 if the summed input is less than a certain threshold value (v), and the value 1 if the summed input is greater than or equal to the threshold value.

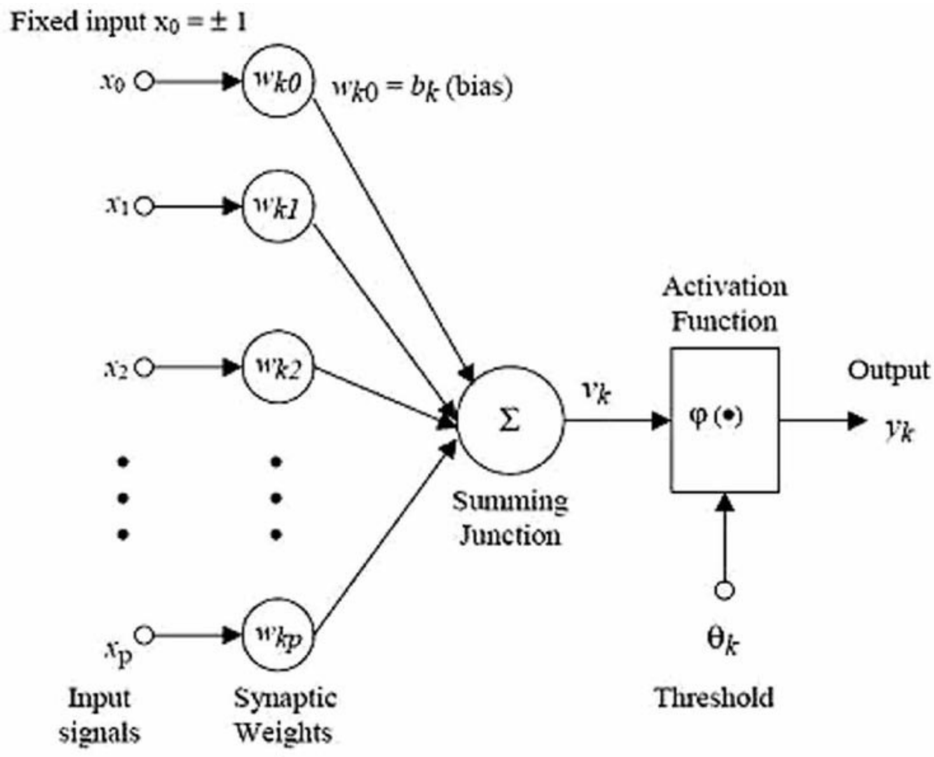

Figure 3: The mathematical process of artificial neural network

The inputs received by a single processing element (depicted in Figure 3) can be represented as an input vector $A= (a_1, a_2, a_n)$, where a_i is the signal from the i^{th} input. A weight is associated with each connected pair of neurons. Hence weights connected to the j^{th} neuron can be represented as a weight vector of the form $W_j=(w_{1j}, w_{2j}, , w_{nj})$, where w_{ij} represents the weight associated to the connection between the processing element a_i, and the processing element a_j. A neuron contains a threshold value that regulates its action potential. While action potential of a neuron is determined by the weights associated with the neuron's inputs (Equation 1), a threshold v modulates the response of a neuron to a particular stimulus confining such response to a pre-defined range of values. Equation 2 defines the output y of a neuron as an activation function *f(x)* of the weighted sum of *n+1* inputs. These *n+1* correspond to the *n* incoming signals. The threshold is incorporated into the equation as

(2) $$y = f(v_k)$$

(3) $$f(x) = \begin{cases} 1 & if \quad x \succ 0 \\ 0 & if \quad x \leq 0 \end{cases}$$

(4) $$f(x) = \frac{1}{1 + e^{-x}}$$

(5) $$f(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$$

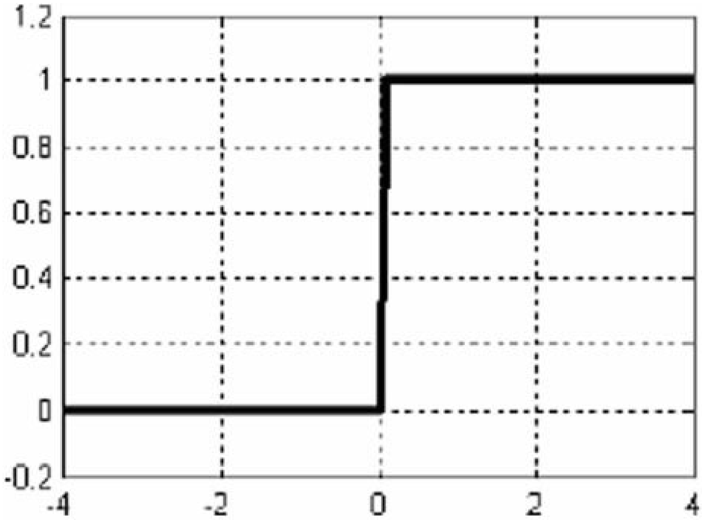

Figure 4: Step function

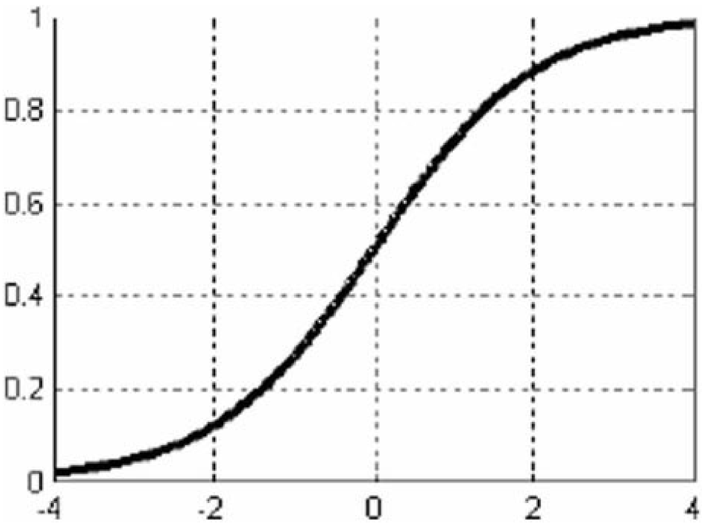

Figure 5: Sigmoid function

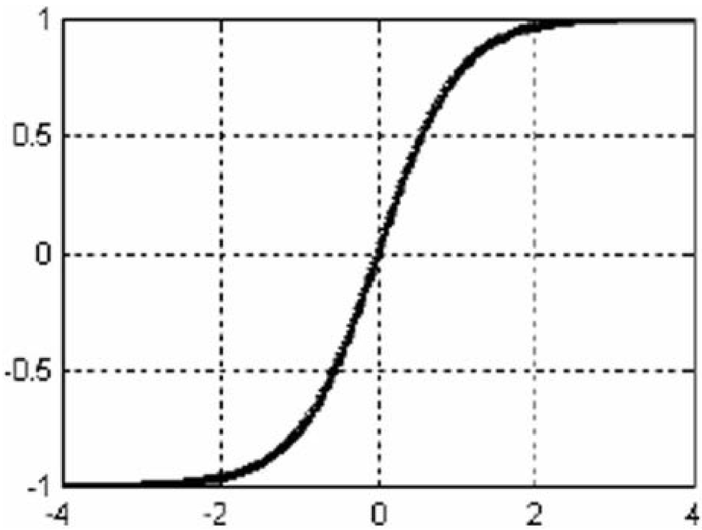

Figure 6: Hyperbolic tangent function

3.2 Training

Since the output(s) may not be what is expected, the weights may need to be altered. Some rule then needs to be used to determine how to alter the weights. There should also be a criterion to specify when the process of successive modification of weights ceases. This process of changing the weights, or rather, updating the weights, is called training. A network in which learning is employed is said to be subjected to training. Training is an external process or regimen. Learning is the desired process that takes place internal to the network. In the next chapter various types of neural network are explained (Hassoun, 1995).

4 SUMMARY

A generic artificial neural network can be defined as a computational system consisting of a set of highly interconnected processing elements, called neurons, which process information as a response to external stimuli. Artificial neural networks are computational paradigms based on mathematical models that unlike traditional computing have a structure and operation that resembles that of the mammal brain. In the artificial neural network, each unit performs a relatively simple job: receive input from neighbors or external sources and use this to compute an output signal, which is propagated to other units.

REFERENCES

Alexopoulos, E., Dounias, G. D., & Vemmos, K. (1999). *Medical diagnosis of stroke using inductive machine learning*. In *Proceedings of Machine Learning and Applications* (pp. 20–23). Chania, Greece: Machine Learning in Medical Applications.

Coiera, E. (2003). *Guide to medical informatics, the Internet and telemedicine*. London: Arnold.

Fausett, L. (1994). *Fundamentals of neural networks*. In *Architectures, algorithms, and applications*. Upper Saddle River, NJ: Prentice-Hall.

Goyal, M. R., & Harmsen, E. W. (Eds.). (2013). *Evapotranspiration: Principles and applications for water management*. Waretown, NJ: Apple Academic Press. doi:10.1201/b15779-7

Hansen, L. K., & Salamon, P. (1990). *Neural network ensembles*. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12 (10), 993–1001. doi:10.1109/34.58871

Hassoun, M. H. (1995). *Fundamentals of artificial neural network*. Cambridge, MA: MIT.

Heden, B., Ohlsson, M., Rittner, R., Pahlm, O., Haisty, W. K., Peterson, C., & Edenbrandt, L. (1996). *Agreement between artificial neural networks and human expert for the electrocardiographic diagnosis of healed myocardial infarction*. *Journal of the American College of Cardiology*, 28, 1012–1016. doi:10.1016/S0735-1097(96)00269-0 PMID:8837583

Jankowski, N. (1999). *Approximation and classification in medicine with IncNet neural networks*. In *Proceedings of Machine Learning and Applications* (pp. 53–58). Chania, Greece: Machine Learning in Medical Applications.

Jones, M. T. (2005). *AI application programming* (2nd ed.). Hingham, MA: Charles River Media Inc.

Jorand, M. I., & Bishop, C. M. (1996). *Neural networks* (Technical Report No. A. I. Memo No. 1562). Artificial Intelligence Laboratory, Massachusetts.

Jorgensen, T. D., & Haynes, B. (2008). *Complexifing artificial neural network through reorganization*. In *Proceedings of the World Congress on Engineering, IWCE 2008*, London.

Karkanis, S. A., Magoulas, G. D., Grigoriadou, M., & Schurr, M. (1999). *Detecting abnormalities in colonoscopic images by textual description and neural networks*. In *Proceedings of Machine Learning and Applications* (pp. 59–62). Chania, Greece: Machine Learning in Medical Applications.

Lippmann, R. P., Kulkolich, L., & Shahian, D. (1995). *Predicting the risk of complications in coronary artery bypass operations using neural networks*. *Advances in Neural Information Processing Systems*, 7, 1053–1062.

Partridge, D., Abidi, S. S. R., & Goh, A. (1996). *Neural network applications in medicine*. In *Proceedings of National Conference on Research and Development in Computer Science and Its Applications (REDECS'96)* (pp. 20 – 23). Universiti Pertanian Malaysia: Kuala Lumpur, Malaysia.

Pranckeviciene, E. (1999). *Finding similarities between an activity of the different EEG's by means of a single layer perception*. In *Proceedings of Machine Learning and Applications* (pp. 49–52). Chania, Greece: Machine Learning in Medical Applications.

Rao, V. B., & Rao, H. V. (1995). *C++ Neural Networks & Fuzzy Logic*. New York: MIS Press.

Sarle, W. S. (1999). *Neural network FAQ, part 1 of 7: Introduction. [Periodic posting to the Usenet Newsgroup comp.ai.neural-nets]*. Retrieved from *ftp://ftp.sas.com/pub/neurl/FAQ.html*

Schmidt, A. (2000). *Biological neural network*. Retrieved June 2012 from *http://www.teco.edu/~albrecht/neuro/html/node7.html*

Street, W. N., Mangasarian, O. L., & Wolberg, W. H. (1996). *Individual and collective prognostic prediction*. Paper presented at the Thirteenth International Conference on Machine Learning.