International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868
Foundation of Computer Science FCS, New York, USA
Volume 12– No.21, June 2019– www.ijais.org

Deep Learning an Overview

Fahad Masood Reda
Department of Information Systems, Faculty of Computing and Information Technology
King Abdul-Aziz University, Jeddah, Kingdom of Saudi Arabia

ABSTRACT
Deep learning is an essential element in technological advancements globally in modern times. Understanding the different constituents that succeed deep learning through advanced AI is therefore paramount in dissecting the impact of deep learning in the society today. This text focuses on the working mechanisms of neural networks in detail. The paper explores the close relationship between the brain's neurons and the artificial neural networks. The paper has been subdivided into subsections that cover important aspects of deep learning with respect to its application in resolving real-world problems that may be too challenging to solve using conventional means.

General Terms
FeedForward Neural Network, Artificial Neurons, Activation Function,Deep Neural Network, Deep Learning

Keywords
FeedForward Neural Network, Artificial Neurons, Activation Function, Deep Neural Network, Deep Learning

1. INTRODUCTION
This section provides an insight into several crucial subtopics that were applied in the data acquisition to satisfy the methodology chapter. The subtopics in the section have been discussed in detail to ensure that important points have not been left hanging. For the purposes of simplification, the subsections may contain other subtopics as well.

2. DEEP NEURAL NETWORKS
A deep neural network can be defined as an artificial neuron network with multiple layers separating between the input and output. From a different perspective, the massive set of rules that have been used to influence machine learning and bring about great success. In the same line, one should note that the neural network models are derived from biological studies of human neurology structure. The latest artificial neural networks are known the deep neural networks. The deep neural networks have proved to be efficient and accurate in the different functionalities that they are set to perform. Deep neural networks work on the concept of approximation. This important aspect has proven priceless as it enables the application of the neural networks in resolving any major challenges in deep machine learning more in the challenges involving the input and output spaces. In addition to providing a solution in understanding the complex mechanisms in which the human brain works, deep neural networks provide a noble set of algorithmic solutions to daily challenges that are dictated by the working concepts of the human brain. The human brain uses a different approach in addressing problems and it is both complex and nonlinear. The human brain is a parallel information processing system (Teng & Suga, 2017, p. *350*, 68-81).

It is important to understand the basic mechanisms of the working human brain so that one can understand how this has influenced the complex algorithms behind the working of the neural networks. The brain works by aligning its neurons so that they can perform particular tasks and may be many times quicker in processing than the currently available digital computers (Marr, 1982). Taking the task of vision as an example, the power of the human brain can be brought out. Vision involves the processing of information from the environment surrounding a character or individual and offering the appropriate response that then allows the person to behave or act in a manner that befits the situation at that particular time (Marr, 1985; Churchland & Sejnowski, 1992). Unknown to most individuals, the human brain is in a state of perceiving recognition and does faster than any conventional computer that is currently in use. The human brain can recognize a familiar object or face from unfamiliar environments in approximately between 100 and 200 ms. Suga (1990a,b) goes further to illustrate the capabilities of the brain using a bar's sonar. The function of this sonar is broadcast out messages and receives important pieces of data about the environment such as the distance of surrounding objects, approaching prey, size of the prey, object size as well as the elevation of the object. The bat's sonar is, therefore, an echo-locator or echo-location system in a bat (Suga, 1993, p. 213-231). Surprisingly, the brain of a bat accomplishes all the necessary computations to return the feedback yet it is no bigger than a plum.

When a human baby is born, its brain is already tailored and structured to start learning and automatically generate its own rules from the experiences that the child has in the formative first two years. Essentially, the learning process does not stop at the early age of two years but the brain continues to learn more as the child continues to gain more experiences in the different aspects of human life. During the first two years, the brain is "hardwired" in preparation for further development after "learning" new experiences. This aspect further means that the brain's "developing" neurons are plastic in nature and hence are able to adapt to changes in their surroundings.

Linking the functionalities between the biological neurons and the artificial neurons in neural networks, plasticity is a key component for the artificial neurons so that they can continuously "learn" different experiences and objects as units of information processing (Basheer & Hajmeer, 2000, p. 43). In this light, neural networks can be termed as machines that can replicate the brains capability in terms of solving challenging problems. Unlike the human brain, the artificial neuron network is made possible by the combining the necessary electronic devices of via simulations by the use of a computer (Ekonomou, 2010, p. 35(2), 512-517). The neural network achieves its functionality by the use of immense interconnections among simple computer cells that are referred to as "neurons" or the "processing units".

International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868
Foundation of Computer Science FCS, New York, USA
Volume 12– No.21, June 2019– www.ijais.org

The neural networks store the gathered knowledge in what is called the synaptic weights. The synaptic weights are strengths in the interneuron connections. Neuron networks use sets of procedures in acquiring new knowledge through the process of learning. These procedures are known as learning algorithms. The learning algorithms modify the synaptic weights where acquired knowledge is stored in order so that the required objectives can be met with regards to the network design. This design forms the basis of the earliest techniques for the design of the neural network (Widrow and Stearns, 1985; Haykin, 1996). However, it is a possibility that the neural networks can alter their design in the respect of replicating the brain's capability of replacing dead neurons with new ones and reconnecting the synaptic connections.

Unlike the so-called shallow neural networks, deep neural networks are considered to have at least one hidden layer apart from the input and output layer, which proves to be an advantage in describing complex systems with many parameters (Bengio, 1995, p. 3361(10)). Depending on the application purposes, there are several types of neural networks: feedforward, convolutional (typically applied in image recognition), recurrent, deep belief networks, autoencoder, etc. (Liu et a., 2017; Schmidhuber, 2013).

The focus of this project will be on the feedforward networks (short: FFNN), which are considered the simplest and most general type of neural nets. In the case of FFNN, neurons in one layer are only connected to the neurons of the adjacent layers (Schmidhuber, 2013). The advantage of FFNN is good training algorithms that converge quickly and enable the network to learn, such as back-propagation algorithm and its improvement, the Levenberg-Marquardt algorithm (Lykourentzou et al., 2009). A typical example of a feedforward neural net architecture is given in Fig. 1.

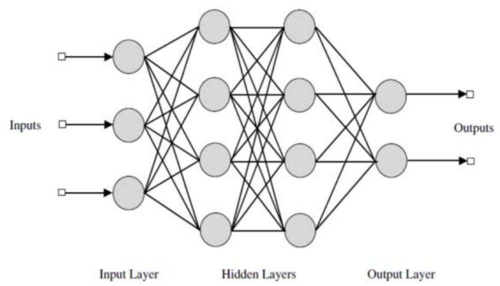


Figure 1. A prototype of architecture for the feedforward neural network. This 4-layer network has a total of 3 input nodes for representing features, 4 neurons per hidden layer, and 2 possible outputs

The neural network model has been previously used in Educational data mining for various purposes: predicting student's grades (Guo et al, 2014), assessment of dropout probability (Kak, Chen & Wang, 2010, p. 16; Papamitsiou & Economides, 2014, p. 17(4), 49-64). An example of a deep learning system based on FNN implemented for EDM is shown in Fig. 2.

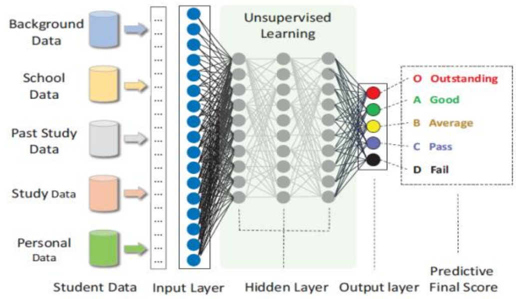


Figure 2. An example of a system used for the prediction of students' success (Guo et al., 2015). The input data are the predictors selected to be relevant for the problem, driven to the hidden layers of a forward-feeding neural network which at output aims to predict student's final mark

These days higher education encounter many challenges from predicting student future learning (Baker & Yacef, 2009), improving institutional effectiveness (Huebner, 1962), predicting student academic path (Abu-Oda & El-Halees, 2015) predicting student dropout, understanding student behavior, improving teaching process and improving existing E-learning system (Anoopkumar & Zubair Rahman, 2015).

3. UNDERSTANDING MODELS OF THE NEURON

As defined earlier in this text, a neuron is a unit of information processing and is paramount in the working of a neural network. Artificial neural network designs are based on the model of the biological neuron model. The biological neuronal model is comprised of three core elements as explained below;

3.0 Synapses / interconnecting links

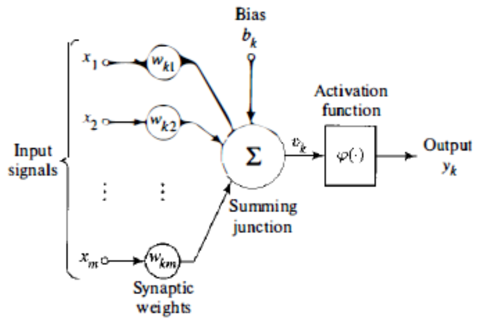
Each synapse and link has a weight. When the signal, take X_j for instance, at an input of a given neuron strength, say j, has some connection to a specific neuron, say n, it's then multiplied against the neuron strength w_{kj} with respect to the order of the subscripts. It is essential to realize that unlike the biological neuron of the brain, an artificial neuron's synaptic strength is denoted as a value in the range between negatives and positives.

The adder that adds all the input signals that have been weighted by some synapse of a neuron respectively.

The activation or squashing function that is responsible for the limitation in the neuron's output amplitude.

The diagram below represents a nonlinear model of a neuron

International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868
Foundation of Computer Science FCS, New York, USA
Volume 12– No.21, June 2019– www.ijais.org



As noted form the figure above, a bias b_k has been included and is responsible for the decrease or increase in the output's activation net accordingly.

3.1 *The mathematical expression of a neuron*

The neuron may be expressed in a mathematical form or representation. The following functions are used for the description of a neuron mathematically;

$$u_k = \sum_{j=1}^{m} w_{kj} x_j$$

$$y_k = \varphi(u_k + b_k)$$

Where

$X_{1}, x_{2}, \ldots, x_m$ - represents the input signals

$W_{k1}, w_{k2}, \ldots, w_{km}$ – represents *synaptic strengths/weights, say k*

U_k – *represents the output* from *linear combiner as a result of the input signals*

V_k – *is the activation potential of the neuron*

$\varphi(.)$ – represents the function of activation for the neuron

y_k – represents output signal

It would be crucial to note that applying the bias b_k results into *fine transformation* in the output of the linear combiner, U_k as illustrated below

$$v_k = u_k + b_k$$

4. ACTIVATION FUNCTION

An activation function $\varphi(v)$ dictates that the neuron output with regards to an induced field v.

4.1 *The threshold functions defined as follows*

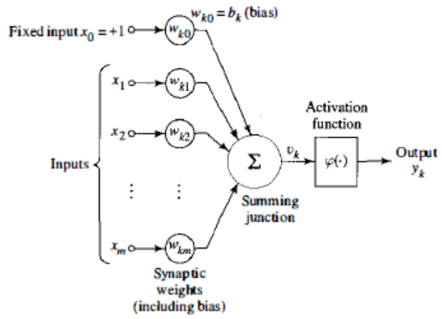
$$\varphi(v) = \begin{cases} 1 & \text{if } v \geq 0 \\ 0 & \text{if } v < 0 \end{cases}$$

The function may also be referred to as a *Heaviside function* in an engineering context. When a neuron *k* applies the threshold function above, the output may be expressed as follows

$$y_k = \begin{cases} 1 & \text{if } v_k \geq 0 \\ 0 & \text{if } v_k < 0 \end{cases}$$

Note: v_k represents the local field of inducement v of the respective neuron in context; that is

$$v_k = \sum_{j=1}^{m} w_{kj} x_j + b_k$$



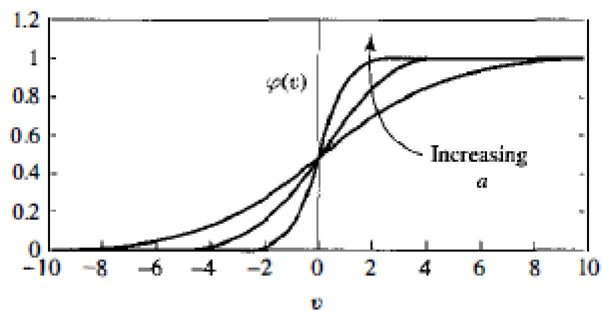
The neuron in this context is referred to as the *McCulloch-Pits model* in the literature. Using the above model, when the local induced field is nonnegative, then the output assumes the value of 1, otherwise, it is 0. This result is due to the *all-or-none property* trait of the *McCulloch-Pits model.*

4.2 *The piecewise-linear function.*

$$\varphi(v) = \begin{cases} 1, & v \geq +\frac{1}{2} \\ v, & +\frac{1}{2} > v > -\frac{1}{2} \\ 0, & v \leq -\frac{1}{2} \end{cases}$$

It is essential to note that in this case, the amplification factor within the area of the performance is taken as unity as an assumption. Furthermore, this function may at time be just an estimation of the non-linear activation function.

Sigmoid Function. This is the most widely used function in the development of artificial neural networks. The function has a graph that has an S shape as illustrated below



The *logistic function* below illustrates a sigmoid function

$$\varphi(v) = \frac{1}{1 + \exp(-av)}$$

Note: *a* is the represents the slope. When the value of *a* is varied, different slopes are obtained from the corresponding different sigmoid functions. When the slope approached infinity within the given limit, the *sigmoid function* then translates into *a* function *of threshold.* As compared to the threshold function which takes a value of either 1 or 0, the sigmoid function is different in that it accepts a range of continuous figures between 0 and 1. Furthermore, sigmoid functions can be easily differentiated unlike the former (threshold) which cannot be differentiable in any case. Differentiation is an important concept in neural network architecture which perhaps explains one of the reasons why the sigmoid function is widely applied in artificial neural network architecture.

5. NEURAL NETWORK ARCHITECTURE

The neural architecture of a network has a direct link to the kind of algorithm that is applied in the network for learning purposes. This aspect means that the learning algorithm/s in a neural network is *structured.* There are three core classes of neural network architectures. This text shall specifically provide in-depth knowledge of the second architectural framework, *multilayer feedforward networks.*

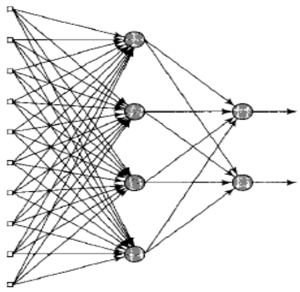
5.1 Single-layer feedforward networks

Neurons are organized to form layers in *layered neural networks.* A basic layered-*network* consists of an input layer (source codes) which then maps to an output layer strictly in this order of events and not in a vise-versa manner. This type of network can be termed as strictly *an acyclic or feedforward* network type.

5.2 Multilayer feedforward networks

This class of neural networks is characterized by the presence of a single or multiple layers of *hidden layers.* The computations nodes in the hidden layers are known as *hidden units.* The nodes may also be referred to as *hidden neurons.* These hidden neurons have some special function whereby they are responsible for interventions between external and network input and output respectively. The hidden layers further enable the network to perform advanced statistics at higher-orders. It may be crucial to note that the neural network gains a global perspective when with regards to connectivity as a result of the available extra synaptic strengths from the hidden neurons (Churchland and Sejnowski, 1992). Higher-order statistics extraction by the neurons in neural networks is paramount when dealing with massive volumes of data.

The nodes at the source of the input layer in the neural network are responsible for supplying corresponding activation pattern elements (input vector), which forms part of the input with regards to the signals passed to the computational neuron nodes within the second neural network area (basically the first hidden layer). Note that the corresponding output signal from the second neural network layer is substituted as inputs passed to the third neural network layer. This pattern is then replicated throughout the entire neural network. This former means that for each layer in this network structure has the output of a preceding layer as its input signals. It then follows that the outputs from the final layer within the network are comprised of all the preceding responses to the activation pattern obtained from the initial input nodes. The following figure illustrates the multilayer feedforward neural network with a single hidden layer output graph.



The neural network illustrated by the output graph above as known as a 10-4-2 neural network. The logic behind this naming is that it has 10 nodes at the sources, a layer of 4 hidden neurons and output neurons. The general rule that is used for naming a feedforward neural network is m-h_1-h_2-q; where *m* is feedforward network source nodes, h_1 refers to a number of neurons within the initial hidden layer, h_2 is the number of neurons from second invincible network layer while *q* represents the neurons present at the output layer. Furthermore, the above network may be termed as *fully connect* in that all the nodes in the layers within the network are linked to all the nodes within the next layer. On the same note, the network is referred to as *a partially connected network* if some of the links between the nodes are missing.

5.3 The recurrent networks

This network has one or more feedback loops which distinguish it from the feedforward networks. The network may be made up of one neuron; where every one of these neurons feeds back to other neuron's outputs in the neural network. The recurrent networks did not generate self-feedback loops within the network.

5.4 Perceptrons

The computational models of a neuron are called perceptrons. The perceptions are also known as the feedforward neural networks and distribute information from the front to the back. Feeding the perceptrons new information through training demands the use of *back –propagation* where massive network paired datasets of input and out are given. After the input has been received, it is processed and the result stored as the output.

5.5 Convolutional networks

Convolutional neural networks are distinctive from most of the available neural network architectures. Their primary application leans towards image processing. The approach may also be used in the analysis of voice inputs where the audio files are fed as the inputs into the network. CNN's basically start with scanner input which parses all the available training data in a single instance. To understand how this scanner works, take the case of a square image with dimensions of 100 x100. Rather than using a layer with 10 000 nodes, a simple 10 x 10 input scanner can be used to feed at least 10 x 10 pixels of the image and the process continues until all the pixels from the image have been scanned (Le, 2018). The data is then fed through convolutional layers rather than using the normal layers where not all the nodes may be connected.

6. FEED FORWARD NETWORKS

Deep learning is known as an evolving field within the machine-learning domain (Bengio, 2013). Recently, deep artificial neural networks (also called neural nets) have shown considerable success in discovering knowledge in data (KDD) and machine learning competitions (Chen et al., 2017). Since neural networks are designed to resemble the networks neurons build in the human brains, the basic processing units of a neural net that form the layers are called *neurons*, and the connections between them *synapses* (Barnes, & Beck, n.d., p. 8-17). The importance of the connection is described with the so-called *weight*, and every neuron has an additional bias. The weight and bias are the parameters that are "learned" during the training phase of the network (Lykourentzou et al., 2009, p. *53*(3), 950–965).

A feedforward network may also be referred to as a feedforward neural network or a multilayer perceptron (Gupta, 2017). It is paramount to state here that the main goal of an FFNN is to find the approximation of some given function. Consider the following example; for the classifier, say $y = f^{*}(x)$, the input x from the function is mapped to some category, y. The feedforward neural network goes further to define the mappings of y in different variations using the equation $\mathbf{y} = f(\mathbf{x};\boldsymbol{\theta})$. The function shall then learn the value of θ that shall result in the best possible approximation functions. As observed from the equations above, the feedforward neural networks cannot feed their output back to the sources that are themselves. When the connections to feedback the output to the individual function is available, then the feedforward networks are referred to as recurrent neural networks. Neural networks are directly linked to perceptrons in that they were originating from these algorithms.

7. REFERENCES

[1] Barnes, & Beck, J. (n.d). *Proceedings of the 1st International Conference on Educational Data Mining.*

[2] Basheer, I. A., & Hajmeer, M. (2000). Artificial neural networks: fundamentals, computing, design, and application. *Journal of microbiological methods*, *43*(1), 3-31.

[3] Churchland, P. S. (1989). Neurophilosophy: Toward a unified science of the mind-brain. MIT press.

[4] Ekonomou, L. (2010). Greek long-term energy consumption prediction using artificial neural networks. *Energy*, *35*(2), 512-517.

[5] Guo, X., Singh, S., Lee, H., Lewis, R. L., & Wang, X. (2014). Deep learning for real-time Atari gameplay using offline Monte-Carlo tree search planning. In *Advances in neural information processing systems* (pp. 3338-3346).

[6] Gupta, T. (2017). Deep Learning: Feedforward Neural Network. Towards Data Science. Retrieved from https://towardsdatascience.com/deep-learning-feedforward-neural-network-26a6705dbdc7

[7] Haykin, S. (1996). Adaptive filter theory 3 rd edition Prentice-Hall. http://sci2s.ugr.es/keel/pdf/keel/congreso/Data%20Mining%20Algorithms%20to%20Classify%20Students.pdf, pp. 8–17

[8] Kak, S. C., Chen, Y., & Wang, L. (2010, August). Data Mining Using Surface and Deep Agents Based on Neural Networks. In *AMCIS* (p. 16).

[9] Le, J. (2018). The 10 Neural Network Architectures Machine Learning Researchers Need To Learn. Medium. Retrieved from https://medium.com/cracking-the-data-science-interview/a-gentle-introduction-to-neural-networks-for-machine-learning-d5f3f8987786

[10] LeCun, Y., & Bengio, Y. (1995). Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, *3361*(10), 1995.

[11] Lykourentzou et. al., (2009a). Dropout prediction in e-learning courses

[12] Marr, D., 1982. Vision, New York: W.H. Freeman and Company.

[13] Papamitsiou, Z., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Journal of Educational Technology & Society*, *17*(4), 49-64.

[14] Schmidhuber, J. (2013). Deep Learning in Neural Networks: An Overview. IDSA. Retrieved from http://www.idsia.ch/˜juergen/DeepLearning8Oct2014.tex

[15] Suga, N. (1990). Cortical computational maps for auditory imaging. *Neural networks*, *3*(1), 3-21.

[16] Suga, N. (1993, October). Computations of velocity and range in the bat auditory system for echo location. In *Computational neuroscience* (pp. 213-231). MIT Press.

[17] Teng, H., & Suga, N. (2017). Differences in velocity-information processing between two areas in the auditory cortex of mustached bats. *Hearing research*, *350*, 68-81.

[18] Widrow, B., & Stearns, S.D., 1985. Adaptive Signal Processing, Englewood Cliffs, NJ: Prentice-Hall.