Developmental Divergence of Artificial Neural Networks

Swopnil N. Shrestha*
Luther College, Decorah, IA
(Dated: May 16 2020)

This paper provides a historical overview of the development of Artificial Neural Networks used the present day research and application of machine learning. From its theoretical conception in the early 1940s out of connectionist neuroscientific models, this paper will highlight significant stages in their development until the modern day application of multilayer perceptron, providing an in-depth path of how the development of ANNs began from the replication of biological neural networks and transformed into representations in vector mathematics, with features beyond those found in nature.

Keywords: Artificial Neural Networks (ANN), Biological Neural Networks (BNN), Linear Threshold Unit (LTU), Perceptron, Adaptive Linear Neuron, Multiple Linear Neuron, Multilayer-Perceptron, Connectionism

I. INTRODUCTION

Attempts of understanding and simulating the function of the human brain have been undertaken for thousands of years, however the first attempts of constructing a model of a human neuron happened in the early 1940s in the United States with flourishing academic interest in connectionist neuroscientific theories. Based on the idea of emergent behaviour, while an artificial neuron is a simple model consisting a few binary digits, multiple connections between these neurons create complex systems that can simulate learning. From the initial attempts to replicate human brain function, developments in Artificial Neural Networks (ANNs) have taken major strides such as the discovery of Linear Threshold Unites (LTUs), Perceptrons, and Multilayer Perceptrons enabling the divergence of ANN research from simulation of Biological Neural Networks (BNNs) to the usage of complex mathematical models to enable the practical application established research in ANNs. Early 21st century ANN models use efficient methods of parallel processing, statistical methods, linear algebra to achieve and ANNs to achieve machine simulated human learning, i.e. machine learning.

II. BIOLOGICAL AND ARTIFICIAL NEURAL NETWORKS

As the movement in cognitive science for connectionism, a movement to understand and model intellectual abilities of the human brain propelled forward the path to research in Artificial Neural Networks. Philosphers found interest in connectionism because "it promised an alternative to the classical theory of mind: the widely held view that the mind is something akin to a digital computer processing a symbolic language" [1]. Thus, early ANNs were developed as non-linear mapping structures originally inspired through the electrical modeling of Biological Neural Networks (BNNs). While there are many different types of ANNs and BNNs, we will explore the fundamental concepts of a BNN, that are relevant to both types of neural networks, which inspired early development of ANNs.

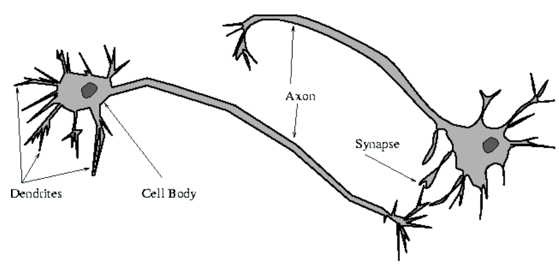

Figure 1. A Biological Neuron
[2]

The human body has evolved to have many systems working cooperatively towards homeostasis, one of which is the nervous system. Through various biochemical processes, the nervous system contains a series of neural networks ultimately connected to the brain. The BNN in the body are responsible for taking sensory input, integrating it with various processes in the brain and creating a reactionary motor output related to the input. For example when a person mistakenly touches a hot stove, the temperature sensitive skin receives the information, which through the process of ionization goes through the nervious system, triggering a series of nerve cells reaching the brain's cerebral cortex. It then produces a reactionary

* Contact Email Address: shresw01@luther.edu

impulse through the neurons of the afferent nervous system to the arms and amygdala, causing a jerking reaction and perception of pain and heat.

Biological Nerve Cells, or neurons are the building blocks of the human nervous system, which also with the principle of emergent behaviour create complex systems that enable abilities like consciousness and learning. Along with a nucleus, DNA, mitochondria and cytoplasm, a biological neuron also contains dendrites which receive information to and from another neurons. Axons send information to other neurons. Synapses, which vary based on the frequency of neural impulses passing through them, are positioned in the cell body. As electrical signals travel through the synapse of a neuron, they "activate", causing the synapse to grow in size and relative significance [3]. This process of synaptic activation and pruning of lesser activated synapses is how learning takes place in the brain.

While neurons are central parts of the nervous system, they work together with many other organelles to provide functionality, such as glial cells, which insulate the neurons and provide defense against foreign bacteria, oligodendrocytes or myelin sheaths which increase the speed at which synapses are conducted. Much of the terminology we use for Artificial Neural Networks have direct parallels to Biological Neural Network. The node in an ANN has the same purpose of a soma, containing the necessary information about its weights, and any attributive information. The node also contains its weight or interconnections, which is parallel to the idea of a synapse in a BNN. Similarly, input of data into the neuron are similar to the functionality of dendrites and output is similar to an axon.

III. LINEAR THRESHOLD GATES AND MCCULLOCH-PITTS MODEL

The first implementations of the idea of Artificial Neural Networks began with several precursors such as threshold logic followed by several attempts to find a non-classical solutions to representations of the human mind. One such attempt which used a purely electronic hardware based framework to explain the process of human neurons was that of McCulloch-Pitts. In 1943, with a simple electrical circuit, neurophysiologist Warren McCulloch and mathematical Walter Pitts work describing how an electrified model of the brain might be represented. Their respective models involved very specific implementation of an artificial neural network which could take boolean input and produce boolean output, containing exhibitory and inhibitory functionality.

The McCulloch-Pitts model was the first example of a linear threshold gate, which became important for later development of ANNs. The model contained a set of inputs $I_1, I_2, I_3, ..., I_m$ and output y. The linear threshold gate classified the set of inputs into two different classes, making y binary. The function can be described as $Sum = \sum_{i=1}^{N} I_i W_i$, and $y = \int(Sum)$.

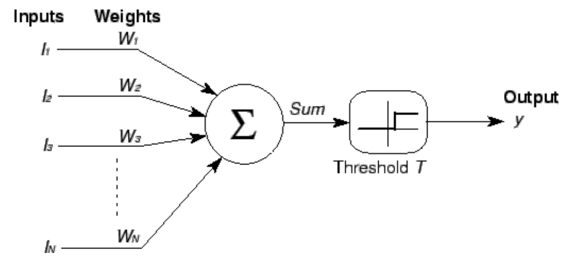

Figure 2. The McCulloch-Pitts Linear Threshold Gate

While very simple, the McCulloch-Pitts model had substancial computing motential and a precise mathematical definition. Figure 2 illustrates a Linear Threshold Gate as demonstrated by McCulloch-Pitts [4]. However, it was too simplistic for doing any practical work since it was only able to generate a binary output.

Research following this model had more flexible and computational features, which included quadratic and polynomial threashold gate models and throughout the 1940s other improvements to the representations and applications of the Model were initiated. Among the most significant works following was that of Donald Hebb, a Canadian neuropsychologist who reinforced the concept of synaptic activation and pruning in 1949. This concept, while not new at the time gave a rise to the concept of activating ANNs and neurons holding values corresponding to their "weight" or significance.

IV. PERCEPTRONS AND THE REDEFINITION OF LEARNING

While the artificial neural networks began moving away from its initial connectionist foundations, it began to adapt theories from cognitive science and pscyhologies, breakthroughs in which contributed significantly to the development and understanding of learning representation in ANN models. As Donald Hebb developed Hebbian Logic, Hebb's work mirrored a series of further achievements such as perceptrons and layered networks of perceptrons such as ADALINE.

Donald Hebb wrote *The Organization of Behavior* in 1949, a paper that pointed out the fact that neural path-ways are strengthened every time they are used, a concept fundamentally essential to the ways in which humans learn. If two nerves are triggered at the same time, he argued, the connection between them increases. As computers became more advanced in the 1950s, it was finally possible to simulate a hypothetical neural network [5]. His work identified thought as the integrated activity of the brain and his views on learning described behavior and thinking in terms of brain function, explaining cognitive processes in terms of connections between sets of neurons. While Linear Threshold Units did not have the ability to learn, Hebbian Logic proposed a mechanism to update weights between neurons. In Hebbian Logic, weight change between neurons was proportional to the product of activation values for neurons. As learning took place, simultaneous or repeated activation of weakly connected neurons incrementally changed the strength and pattern of weights, leading to stronger connections.

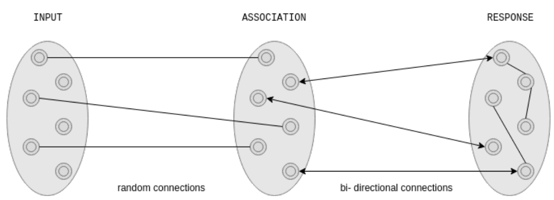

Figure 3. Rosenblatt's Hebbian Learning Model

Hebb's work was a breakthrough in cognitive science, ANNs and psychology and provided a theoretical framework for understnanding learning. Whatsoever, an implementation of Hebbian logic in ANNs were only implemented in 1958 by Frank Rosenblatt. Intrigued by the eye of a fly which does much of the processing telling a fly to in a given direction, Rosenblatt's work led to the invention of the perceptron. A perceptron was found to be useful in classifying a continuous-valued set of inputs into one of two classes. The perceptron computes a weighted sum of the inputs, subtracts a threshold, and passes one of two possible values out as the result. Figure 3 represents Rosenblatt's neuron model following Hebbian logic, consisting of multiple of connections between perceptrons.

In 1959 Bernard Widrow and Marcian Hoff of Stanford developed models called "ADALINE" and "MADALINE", which represent Adaptive Linear Neuron and Multiple Linear Neuron. ADALINE was developed to recognize binary patterns so that if you were reading transmission bits from a phone line, you could predict the next bit. MADALINE was the first neural network applied to a real-world problem, using an adaptive filter that eliminates echoes on phone lines [6]. It consists of "n" units of input layer and "m" units of ADALINE layer and "1"

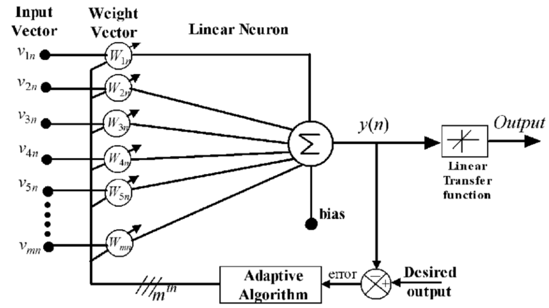

Figure 4. The ADALINE Neural Network
[6]

Unit of the MADALINE layer. Each neuron in the ADALINE and MADALINE layers has a bias of excitation "1". The ADALINE layer is present between the input layer and the MADALINE layer; the ADALINE layer is considered as the hidden layer. ADELINE and MADELINE gave neural networks computational capacity which was not found in single layer networks, however the neural network training process had many complications.

V. HALT IN RESEARCH

After the discovery of perceptrons and ADALINE, critics of connectionism and Artificial Neural Networks increased, highlighting the expenses of development and the infeasibility of perceptrons to be used commercially due to a lack of expertise and hardware limitations. This coupled with ethical criticism and the lack of market demand for ANNs caused a halt in research that lasted almost a decade. Research interest in ANNs was rejuvenated not until 1987.

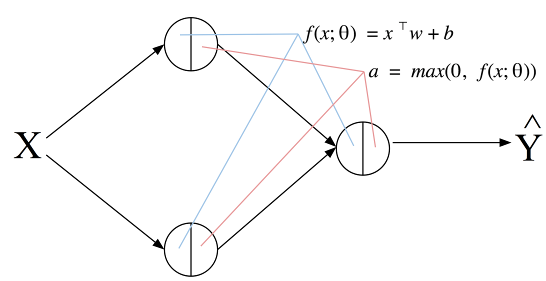

Figure 5. Minsky and Papert's AND / OR Theorem
[7]

One of the more significant sources of criticism was a book by Marvin Minsky and Seymour Papert who proved the Perceptron model to be limited, in their book *Perceptrons: An introduction to computational geometry.* The book included several mathematical proofs which acknowledged some strengths of the perceptron model while showing numerous limitations. The authors in the book imply that, since a single perceptron is not capable of implementing functions such as the XOR logical function, larger networks also have similar limitations, and therefore perceptrons do not hold much practical potential [7]. The book proved that in three-layered feed-forward perceptrons (with a "hidden" or "intermediary" layer), it was not possible to compute predicates unless at least one of the neurons in the first layer of neurons is connected with a non-null weight to each and every input. Illustrated in figure 5 This contradicted the understanding of most researchers who replied mostly on network with few layers of "local" neurons [7]. While a feed-forward machine with "local" neurons was much easier to build and use, however they did not have much practical purpose. This caused respected voices to critique the neural network research. The result was to halt much of the funding. This period of stunted growth lasted through 1981.

Due to many researchers competing for funding on the topic and hopes in application of ANNs diminished, the development of ANNs did not see much growth until 1985, when the Amercian Institute of Physics began what has become an annual meeting - Neural Networks for Computing. By 1987, the Institute of Electrical Engineers' (IEEE) first International Conference saw a replenishing of interest in ANN research again. The following new years led to the improvement building more practical neural networks, that were often layers of multilayer perceptrons.

VI. TODAY'S ARTIFICIAL NEURAL NETWORKS

Around the late 80s and early 90s significant strides were made in the development of the field of ANNs. While exact replication of biological neural networks were prevented by significant gaps in understanding of the brain and limitations in hardware, many strides were made towards a practically implementable ANN. This drove research forward as institutions saw the economic benefit of ANNs taking us to today when ANNs are researched by governments, massive corporation and universities, with complex neural network structures such a Boltzmann Machines, Generalized Delta Rule, and Generative Adversarial Networks,. While there are many different types of neural network which can be specifically constructed based on the need a few of the examples of modern neural networks are Recurrent Neural Networks (RNNs), Long Short Term Memory (LSTM),

In 1990, Jeffrey Elman's introduced recurrent neural networks (RNNs) which are similar to perceptrons however, unlike perceptrons which are stateless, RNNs have connections between passes and through time [8]. RNNs are very powerful, because they combine the properties of distributed hidden state that allows them to store a lot of information about the past efficiently and non-linear dynamics that allows them to update their hidden state in complicated ways [8]. RNNs they could potentially learn to implement many small programs that each gather chunks of information and run in parallel, interacting together to produce very complicated implementations.

Hochreiter and Schmidhuber solved the problem of memory by getting an RNN to remember content for long time by building what known as long-short term memory networks (LSTMs). LSTMs networks try to combat the vanishing gradient problem [9] by introducing gates and an explicitly defined memory cell. The memory cell stores the previous values and holds onto it unless a "forget gate" tells the cell to forget those values. LSTMs also have an "input gate" which adds new information to the cell and an "output gate" which decides when to pass along the vectors from the cell to the next hidden state.

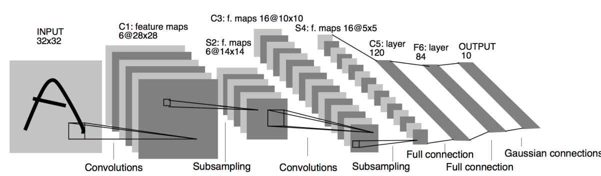

Figure 6. LeNet Convolutional Neural Network Model [10]

In 1998, Yann LeCun and his collaborators developed a recognizer for handwritten digits called LeNet [10]. It used back propagation in a feedforward network with many hidden layers and maps of replicated units in each layer, pooling of the outputs of nearby replicated units, with the ability to train a complete system, not just a recognizer It was later generalized and given a formal definition called convolutional neural networks (CNNs).

Several other steps have been taken to get us to where we are now; today, neural networks discussions are prevalent. Currently most neural network development is simply proving that the principal works. This research is developing neural networks that, due to processing limitations, take weeks to learn. While these models contain thousands of neural networks, the idea of the neuron has not deviated much from the original works of McCulloch and Pitts. However what has remained prevalent is the idea of emergent behavior, as layers of thousands of neural networks are implemented, these create complex behaviours such as learning, image and speech recognition. While reaaching the potential of biological neural networks are still far off due to biological limitations, these neural networks solve the hardware limitations by using complex mathematics to create a more practical and usable neural network structure.

VII. CONCLUSIONS

The first formal attempts of constructing Artificial Neural Networks stemmed from an interest in "connectionist" theories and the attempt to use computational modeling to replicate human brain function. Throughout the late 20th century, developments in cognitive science and the design and implementation of Linear Threshold Units caused an acceleration in ANN research. Hebbian logic enabled the foundation of the ideas such as a perceptron, and multilayer-perceptrons were modeled to simulate the brain's cortex function. In the 1980s research in ANNs were temporarily halted due to low funding and demand. In the 90s several researchers applied sigmoid, rectilinear, and other functions developing a concept of memory followed by gradient based methods in the 1980s. While the foundation of ANNs have a origin of attempting to replicate biological and brain function, the focus of today's ANN research has diverged towards building faster, more efficient models of parallel processing, with features that do not exist in biological systems due.

ACKNOWLEDGEMENTS

I am grateful for the support of my professors Dr. Shafqat Shad and Dr. Roman Yasinovskyy for offering educational content, reviewing my work and providing feedback to make this paper possible.

[1] C. Buckner and J. Garson, "Connectionism," (1997), [Online; accessed 16. May 2020].
[2] A. Schmidt, "Biological Neural Networks," (2000), [Online; accessed 16. May 2020].
[3] P. Lorrentz, *Artificial Neural Systems: Principle and Practice* (Bentham Science Publishers, 2015).
[4] W. S. McCulloch and W. Pitts, The bulletin of mathematical biophysics **5**, 115 (1943).
[5] L. D. Harmon, Science **129**, 962 (1959).
[6] P. Mishra, Medium (2019).
[7] M. Minsky and S. A. Papert, *Perceptrons: An introduction to computational geometry* (MIT press, 2017).
[8] J. L. Elman, Cognitive science **14**, 179 (1990).
[9] S. Hochreiter and J. Schmidhuber, Neural computation **9**, 1735 (1997).
[10] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, Proceedings of the IEEE **86**, 2278 (1998).