Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

Simultaneous verses Successive Learning in Neural Networks

Anthony Choi
Electrical and Computer Engineering
Mercer University
Macon, GA, US
choi_ta@mercer.edu

Nithyalakshmi Nareshkumar
Computer Science and System Analysis
Miami University
Oxford, OH, US

***Abstract*— Neural networks were inspired by the human brain, with great hopes that neural networks would capture the vast potential of its biological counterpart. This paper explores the link between neural networks and the human brain in the context of simultaneous vs. successive learning. Learning experiments conducted on human subjects were modeled and repeated using neural networks as test subjects. Neural networks confirmed the conclusion from human subject experiments that simultaneous learning was faster than successive learning. Loess and Duncan [1] further extended their hypothesis without formal experimental evidence that simultaneous would outperform successive as complexity increased beyond the scope of their human experiments. Interestingly, neural networks contradict their hypothesis. The results from neural networks demonstrate an existence of a threshold, after which the effects of simultaneous and successive learning become negligible. Intuitively, when humans are presented with complicated tasks, the type of learning is immaterial, since the complexity of the problem would overwhelm any advantages one method has over the other. Confirmation of this intuition can only be confirmed through future human experiments. Furthermore, this paper demonstrates that neural networks can be used as a rough model and give valuable insight into a problem, before the costly human subject experiments are conducted.**

***Keywords*— Learning, artificial intelligence, neural networks**

I. Introduction

One of the most fascinating areas of research is trying to understand how the brain works. Neuroscientists and psychologists are constantly trying to solve the mystery of the "neuronal matrix." This study has drawn attention from psychologists, cognitive scientists, computer scientists and neurobiologists.

It was a few decades ago, when computer scientists began to show interest in understanding how the brain works. This laid the foundation for various fields such as artificial intelligence, neural networks, and machine learning. Interdisciplinary study has been beneficial in the past. The neural network for example tries to model the working of the brain to classify and is one of the techniques of machine learning. Neural networks have been very useful in a variety fields from Microbiology [2] to Chemical Engineering [3].

A. History of Connectionism

Connectionism [4] is a branch of cognitive science that explains the brain's mechanism using models such as artificial neural networks. Connectionism has drawn attention from many philosophers and psychologists. According to the opposing "classical theory," the processing in the brain is similar to that of a rule based computer process. However, the connectionists believe that information in the brain is stored as "connections" in the neuronal matrix with parallel distributed processing. This debate between the two opposing groups has been going on for the last thirty years [4].

The classicists believe that information in the brain is stored as strings of symbols similar to the way in which information is stored in the memory of a computer. They further believe that cognition is similar to the processing of digital data and the strings of symbols are processed according to some program. The connectionists believe that information is stored non-symbolically similar to the connection weights in a neural network and that cognition is similar to the processing in the nodes in the neural network in which each node's activation depends on the connection weights and the activation of other nodes. The implementational connectionists combine both views and believe that the processing of the brain is symbolic processing and this is done through a structure similar to a neural network. The pure connectionists oppose the theory of the implemetational connectionists as they feel it "fails to explain various activities of human cognition" [4].

Connectionism is important in various fields of study such as cognitive science, artificial intelligence, philosophy, and psychology. Due to its inter-disciplinary nature, the scientists of various fields have contributed towards its growth. Connectionists model the functional properties of the brain that are required for cognition and information processing.

The birth of connectionism dates back to 400 B.C. when the philosopher Aristotle stated that "memory is composed of simple elements linked or connected to each other via a number of different mechanisms (such as temporal succession, object similarity, and spatial proximity). Furthermore, these associative structures could be combined into more complex structures to perform reasoning and memory access" [5].

Later, when psychology branched out from philosophy as a separate field of study, significant contribution was made to connectionism by psychologists such as Spencer, James, and Hull. Spencer [6] laid the foundation and felt that the description of the nervous system was essential for the study of Psychology. He further described the connections not only between neurons but also between ideas and concepts and how

978-1-4244-2794-9/09/$25.00 ©2009 IEEE

neural changes were affected by "psychical" changes. His main idea was that knowledge is stored within the connections in the brain. The most important contribution of James [7] to connectionism is his associative model of memory. In this model of associative memory, ideas are connected in parallel and the recall of one idea will lead to the recall of all/most ideas related to it. Further, when events occur over and over again, the "connection between the relevant brain processes is strengthened." This concept forms the basis for the modern-connectionist theory. Then came Thorndike's [8] contribution by his two laws: The Law of Exercise which states that, when a particular action leads to a certain response, there is a stronger tendency for this to happen in the future given all actions-response pairs have equal strengths initially. According to the Law of Effect, when an action is followed by a reinforcing stimulus, then the connection is strengthened, and if followed by non-reinforcing, stimulus the connection is weakened. Hull's [9] important contribution is his theory for the development of "stimulus-response" habit strength. Hull's basic yet popular learning rule is: "the process of learning consists in the strengthening of certain of these connections as contrasted with others, or in the setting up of quite new connections" [9].

The next stage in the growth of connectionism is the influence from neuro-psychology. Lashley's view [10] that learning is a distributed process was important towards connectionist research. The most important contribution towards connectionism came from the Canadian psychologist, Donald. O. Hebb [11]. According to Hebb, learning is based on the modifications of synaptic connections between neurons. The Hebbain Rule is stated as follows: "When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased" [11]. This Hebbian learning later formed the underlying principle of the neural network.

Mathematicians such as Pitts and McCulloh [12] laid the foundation of modeling neuronal activity using propositional logic. After the advent of the computer, models such as the pandemonium and perceptron contributed towards the growth of connectionism. Pandemonium proposed by Selfridge [13] was the first model to incorporate parallel processing. The basic unit of the pandemonium is called a demon and consists of four layers: the first layer is the data layer into which the data is fed, the demons of the second layer do computation on the data and are referred to as the computational demons, followed by the third layer that utilizes the results of the computational demons, weights these results, and passes the evidence to the final layer. The final layer consists of decision demons and as the name implies, based on the evidence supplied to it from the cognitive demons, makes a decision. The pandemonium was effective in distinguishing dots from dashes and also in recognizing hand-written characters. The next level of progress was the development of the perceptron model which functionally models the brain closer than that of the pandemonium. Minsky and Parert [14] then proved that the solving capability of perceptrons was restricted to linearly separable problems.

The architectures that were developed after the perceptrons fall under "New Connectionism" [15]. Various connectionist models have been proposed since then:

- The first architecture is the Interactive and Competition Model (IAC) [16] by James McClelland. With the introduction of the IAC, the adaptive resonant theory (ART) networks and Kohenen's self organizing maps [4] were developed.
- The Multi-Layered Neural Network using the back-propagation algorithm, developed by RumelHart [17, 18], is also known as the PDP (Parallel Distributed Processing) architecture. PDP is very popular for various practical applications and is most widely used of all the other "new connectionist" computer models.
- The next model was an improvisation of the PDP architecture [19] in which the activation function is Guassian. This has the advantages of using fewer hidden nodes.
- The Radial Basis Function network is similar to the feed forward network and employs radial functions [20].

B. Psychology

The birth of neural networks occurred as a result of computer scientists' inspiration and their drawing ideas from other fields, specifically neuroscience. In a similar way, inspiration could be drawn from fields such as psychology as it relates closely to the working of the brain and how organisms react to various stimuli.

Extensive research has been done in psychology in understanding simultaneous versus successive discrimination. The period between the 1930's and the early 1950's was the era during which this research was primarily carried out. Two groups with opposing views emerged. As more research was done, people began to support one of the two views based on the results conducted by each of them. One group stated that simultaneous discrimination or learning of tasks resulted in faster, better and more effective learning as opposed to successive learning [1]. Their reasoning was that when an organism or subject has to learn a set of tasks, learning to differentiate between these stimuli simultaneously will help in better learning. While learning to discriminate between the two stimuli at the same time, the learning happens at a faster rate as the subjects can understand the distinct feature of each stimuli by comparison and hence can distinguish better between or classify the stimuli (i.e. when subject to a particular stimuli, the subject can conclude as to which class the stimuli belongs to). The opposing view to this study was that typically there should be no difference between simultaneous and successive discrimination/learning of tasks. If one of the two was better, it had to be successive discrimination [1], since they believed that learning was related to the development of "habit strength" and independent of the type of learning.

It was assumed that simultaneous learning was more difficult than successive. It was not until Weise and Bitterman's [21] data from experiments the opposite was proved. The experiment was to choose the brighter or darker of two alleys. The experiment was conducted on two groups of rats: one group trained in successive learning and the other group in simultaneous. The second group performed far better than the first.

Lashley and Wade [22] proposed that learning to discriminate between patterns occurs only when the correct stimuli are compared against the incorrect stimuli. They claimed that comparison is an important variable for discrimination and the amount of comparison available is directly proportional to the rate of learning. Further, if stimuli are presented successively, learning should be slower than when the stimuli are supplied simultaneously. The reinforcement theory of Hull and Spence [23, 24] states that comparison is not an important variable for discrimination tasks. The differential response to stimuli is due to the development of habit strength for positive stimuli and inhibition of habit strength towards negative stimuli. The development/inhibition of habit strength is related to the type of stimuli and comparison is immaterial. Later however, for pattern recognition tasks, Spence [25] predicted that successive learning would be slower than simultaneous.

Other relevant experiments include those conducted by Grice [26] on rats to learn to differentiate different sizes of circles. A group of ten rats was subject to successive stimuli and another group of rats was subject to simultaneous stimuli. A specific error was defined by Grice for this experiment. Grice noticed no significant difference in learning between the two groups of rats in terms of the error as well as the number of trials for learning. Hence the conclusions made by Grice did not support the Lashley-Wade theory [22] that learning would be faster when there is room for comparison. One other relevant experiment was that of size discrimination of two groups of rats conducted by Baker and Lawrence [27]. The group of rats that were trained on simultaneous learning performed better.

All experiments conducted so far did not take into account the difficulty of tasks involved. The intuition behind the experiments conducted by Loess and Duncan was that for easy tasks, the difference between simultaneous and successive learning might not be significant. However, for a difficult discrimination task, simultaneously distinguishing between the tasks might prove useful. Loess and Duncan [1] experimented with human subjects and concluded that the type of discrimination would not make a difference with an easy problem, but with a difficult task, the type of discrimination would make a significant difference. The experiment conducted on human subjects was that of visual discrimination. The task was made more difficult by adding an irrelevant feature to the task and since it was a visual discrimination task, the subjects could not easily tell that it was due to an irrelevant feature that the task appeared to be more difficult.

In brief, their experiment is described as follows: the subjects were 140 students chosen from a psychology department which consisted of advanced and elementary students. Two groups were required for each type of learning. For the easy task, each group consisted of 20 elementary and 20 advanced students. For the difficult task, 30 advanced students were put into each group. For each of the tasks, one group was exposed to simultaneous stimuli and the other group was exposed to successive stimuli. The subjects were seated at one end of the table and the supervisor E was seated at the other end of the table.

The stimuli consisted of eight stimuli cards, four of which were white and four of which were gray. Each card was a 3 ½ x 4” cardboard. On the cards were drawn either large or small circles or squares. The eight cards were as follows: a small circle on white, a small circle on gray, a small circle on white, a large circle on white, and similarly for the squares.

As shown in Figure 1, these cards had three features: color of the background (white/gray), type of shape (circle/square), and size of shape (small/large). These cards were divided into two sets, A and B. Set A was comprised of circles on gray cards and squares on white cards. Set B comprised of the remaining cards, i.e. circles on white and squares on gray. The task was to discriminate the cards of the two sets, i.e. which cards belonged to set A and which cards belonged to set B. The cards were randomly selected and presented to the subjects. For simultaneous learning, a card from each set was presented together and the subjects had to classify the cards presented to them and were immediately given feedback by the supervisor E on their outputs. The eight cards were presented two at a time with a total of 32 combinations and the trials were repeated until the subjects learned to classify/discriminate correctly. The other group underwent successive learning for the same set of cards and the primary difference was that instead of displaying two cards at a time as in the former case, the cards were displayed one at a time, i.e. a card belonging to set A followed by a card belonging to set B and the feedback was given by E after each card was presented to the subjects. This was the easy experiment.

The next set of experiments conducted was for the difficult task. An extra feature was added to the same set of eight cards to increase the complexity. This was done by adding an equilateral triangle on the card (an irrelevant feature). The triangle fit within the large circle/square and the triangle enclosed the small circle/square. The triangle was pointing down on the white cards and pointing up in the grey cards. This did not alter the two sets. The rest of the experiment was carried out similar to that of the easy task with immediate feedback. The experiment was repeated until the subjects identified which cards each set contained. For the easier task, if within 75 trials the students did not discover the features that differentiated the two sets, the experiment was stopped. The same applied to the difficult task after 96 trials.

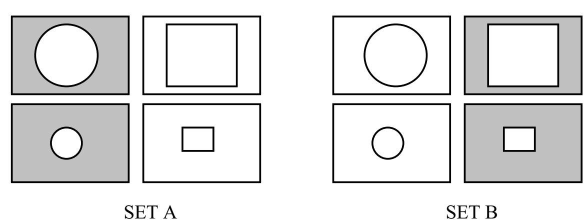

Figure 1. The eight cards used in the experiment as the stimuli for the visual discrimination task. Circles on gray and squares comprise set A and circles on white and squares on grey comprise set B.

The results of the experiments were similar to what Loess and Duncan had expected. With respect to the easy task, for both the types of learning, the number of trials required to learn the task was the same. The difference between the number of trials required to learn the easy task among the two groups (Successive and simultaneous) was about 7 trials (successive requiring fewer trials than simultaneous). However this difference was not statistically significant and hence it was concluded that for an easy task, there is no difference between simultaneous and successive learning. With respect to the difficult task, the number of trials required to learn for the successive was almost twice that required for simultaneous and the difference in the number of trials was around 30. This is a significant difference. With these results the original hypothesis was supported. Further, the number of trials for learning with respect to simultaneous learning, there was no difference for the easy and difficult task but there was a significant difference in the number of trials for successive learning for the easy and difficult task. The significance of difference between the two tests was determined using the Mann-Whitney U test and the Chi-square test. This does not completely prove that simultaneous learning is better than successive learning. However, with respect to pattern discrimination, experimentally and intuitively, it made sense that simultaneous learning is better than successive learning.

C. Neural Networks

Neural networks were inspired from the working of the human brain. The neuron is the functional unit of the brain and its structure is described in [28]. Synaptic Plasticity is an important biological phenomenon that forms the basis for learning and memory. As the brain learns a particular activity, the connections between specific neurons are increased and hence become stronger.

Various connectionist models were developed based on the model of the brain. By connectionist theory, the functional unit of the brain is composed of the following six components [15]:

- Input device that receives signals from other neurons or from the environment.
- Integrating device that integrates and manipulates the input.
- Conducting device that conducts the integrated and manipulated input.
- Output device that sends information to other neurons.
- Computational device that maps one type of information to another.
- Representational device for the formation of internal representations.

In order to model the above functions of the neuron, Rumelhart developed Parallel Distributed Processing (PDP) models. The functions of PDP's that map the properties of neurons are [17]:

- A set of processing units.
- A state of activation.
- An output function for each unit.
- A pattern of connectivity among units.
- A propagation rule for propagating patterns of activities through the network of connections.
- An activation rule for combining the inputs of a unit with the current state of that unit to produce a new level of activation for the unit.
- A learning rule in which patterns of connectivity are modified by experience.
- An environment within which the system must operate.

This led to the development of the multi-layer neural network that uses the back-propagation algorithm for learning.

The representational power of the neural networks increases as the number of hidden layers is increased. The network has a single input layer, one output layer and one or more hidden layers. The number of hidden layers determines the representational power or the expressivity of the neural network. The expressivity of the network based on the number of hidden layers can be summarized as follows [29]:

- One hidden layer: any continuous function can be represented with one hidden layer provided there are a sufficient number of nodes in this layer.
- Two hidden layers: any discontinuous function can be represented with two hidden layers. The downside, however, is that by increasing the number of layers, the speed of the network is dramatically reduced and the chances of getting trapped in a local minima are also increased.

Determining the number of hidden layers is one of the many design issues involved in setting up the neural network. Determining the neural network architecture is one of the most important issues to deal with in artificial neural networks. There is no well-defined procedure or methodology in the literature that clearly defines this. However, there are certain approximate ways through trial and error in which the number of layers and the number of nodes within each layer could be determined. Research has shown that keeping the number of nodes to a minimum, as long as good performance is obtained, is desirable due to the following reasons:

- it significantly reduces the time involved in computation,
- it helps avoid over-fitting (capturing the idiosyncrasies of the data) and thereby improves generalization,
- the trained network can be analyzed easily if the number of connection weights is fewer in number.

Thus multi-layered feed-forward neural networks model the brain and the back-propagation algorithm models the learning. In a broad sense, neural networks can be classified as iterative or batch neural nets [29] based on the frequency at which the connection weights are updated.

II. MODELING THE PSYCHOLOGY (LOESS AND DUNCAN) EXPERIMENT: SIMULTANEOUS VS. SUCCESSIVE LEARNING

It is known that neural networks model the brain, but to what degree they reflect the working of the brain is a part of the question that we seek to answer through this study. Hence this study was carried out by modeling experiments conducted by the psychologists Loess and Duncan [1]. These experiments were then extended to incorporate larger data sets. This section describes in detail all the experiments that were conducted.

The task was to learn to discriminate between two classes A and B, as shown in Figure 1. The same features that were used by Loess and Duncan in their experiment characterized each class. Modeling the input or the input encoding can typically be done in one of two ways: pixel method or feature selection

With pixel method, each image of the card would have to be digitized into pixels and the pixel values of the resulting image given as input to the neural network. With the feature selection method, each feature of the card would be taken as an attribute. Three features characterized each card: background color (white/gray), shape (circle/square), and size of the shape (large/small). As stated earlier, the task was to learn to discriminate between two sets of cards – set A and set B. Circles on gray and squares on white comprised set A and squares on grey and circles on white belonged to set B for the easy task. This task was made difficult by adding a triangle to the existing features. The triangle was an irrelevant feature, which was always present in the data for the difficult task and was pointing up on white cards (0) and pointing down on grey cards (1). As shown in Table I, the values of 0 and 1 were used to represent the values of the attributes. The attribute and class values were binary and hence the values were represented as 0 and 1 respectively.

Human beings have prior knowledge of shapes, colors and sizes that the neural network lacks. Hence by using feature selection to encode the input, the network was given some extra knowledge, thereby trying to implicitly capture the extra knowledge that human beings have. If pixel method was used, then less information would be given to the neural network about the domain and result in much higher complexity. Thus the training data set was eight data points and was ordered as shown in Table I.

Following Loess and Duncan, the data set was ordered such that an instance of A alternated with an instance from class B. The successive learning was modeled using the iterative neural network. In the human successive experiment, a card form set A was presented to the subjects followed by the subjects' guess or answer. Feedback was given immediately to subjects about their answer. Then a card from set B was presented to them and the procedure was repeated. Similarly for neural network, an instance of set A was forward propagated through the network and the error was back propagated, weights were altered and then an instance of set B was forward propagated. Each trial or epoch consisted of all eight instances forward propagated through the network with eight corresponding back-propagations. Simultaneous learning was modeled using a semi-batch neural network. An instance of set A was forward propagated followed by the forward propagation of an instance of set B. The summed error obtained after the forward propagation of the two instances was then back propagated through the network. This corresponded to human simultaneous learning, where the human subjects were presented with a card from set A and one from set B at the same time, followed by their guess to which set each card belonged to and then feed-back was given. Thus the batch size for the semi-batch network was two instances, one belonging to set A and the other to set B.

TABLE I. DATA SET GIVEN AS INPUT TO THE NEURAL NETWORK – MODELING THE PSYCHOLOGY EXPERIMENT CONDUCTED BY LOESS AND DUNCAN.

Color	Shape	Size	Triangle	Class
0	0	0	0	0
0	1	0	0	1
0	0	1	0	0
0	1	1	0	1
1	1	0	1	0
1	0	0	1	1
1	1	1	1	0
1	0	1	1	1

In order to attain good performance with the neural network, it was important to set up the neural network with appropriate values for the parameters. Setting up the neural network is important in any neural network application. This is not an easy task considering the numerous parameters involved. Due to the lack of specific guideline for network and parameter initialization, the neural network was set up through trial and error.

For each learning scenario, custom neural network was developed. First the number of hidden layers to be used was decided through trial and error and the number of nodes per hidden layer was also decided in the same way. Trial and error process started with a single layer with one node. Then the number of nodes were increased. The training error dropped to 0.25 with 2 nodes in the first hidden layer and three nodes in the second. The other parameters of the network were also determined through trial and error in order to determine favorable network architecture. A snippet of the average results (for about 30 trials) obtained for simultaneous learning (for the easy task) is given in the following table. The neural network was set up with 2 and 3 nodes in the two hidden layers and the weights were randomly initialized.

Table II shows the mean and standard deviation for the 30 trials. Tuning the parameters while starting off with momentum = 0.1 and learning rate = 0.1, the neural network was would land in a bad local optima. As the value of momentum was increased, the training error dropped. As stated earlier, the momentum helps in rolling over the bad local optima. With successive learning (for the easy task), the lowest train error obtained was 0.5 with random weight initializations

TABLE II. TRAINING ERROR FOR SIMULTANEOUS LEARNING IN NEURAL NETWORKS WITH RANDOM WEIGHT INITIALIZATIONS FOR DIFFERENT COMBINATIONS OF MOMENTUM AND LEARNING RATE.

Momentum	Learning Rate	Mean over 30 trials	Standard Deviation over 30 trials
0.1	0.1	0.4875	0.0827
0.2	0.2	0.3625	0.0948
0.3	0.5	0.3000	0.0904

Using the same set of parameters for both types of learning, the training error with random weight initialization is given in the Table III. The mean and the standard deviation were also calculated for 30 trials of the experiment and are given below:

The data set was small with only eight data points. With random weight initialization and the tuning of various parameters of the network, the best training error obtained for the simultaneous case was 0.25 and for the successive case was 0.5. The performance of the above experiments was based on the training error.

Due to the surprisingly high training error, another testing method was used to verify the training error. In neural network, typically 2/3 of the data set is used for training and the remaining 1/3rd for testing. Since the data set consists of only eight data points, a separate test set could not be formed. Instead, cross-validation was employed and a pseudo test error was obtained.

Cross-validation is a popular technique used in various machine learning algorithms in order to determine a pseudo test set [29]. It can also be used to determine parameter values of the network. Cross-validation is used to measure test error when the data set is very small. In Cross-validation, the data set is divided into k-partitions. In this case the value of k = 4 was chosen and each partition consisted of two data instances. The neural network was trained on 3 partitions and tested on the fourth partition. This was done four times, each time testing on a different partition. Then the test errors were averaged to get the pseudo test error. The average pseudo test error of 0.5 was obtained for both simultaneous and successive learning, which is just as good as random. Cross-validation was carried out for various network structures and parameters just as previously conducted, but cross-validation produced error measurements similar to train errors shown in Table III.

A careful analysis of the data set given in Table I showed that it could be reduced to a simple XOR problem. This can be observed by eliminating the two irrelevant features (i.e. Size and Triangle). The simplified data set is given in Table IV. Hence, this was the explanation of the high training error rates observed in Table III.

It is known that the neural network can be set up in order to learn the XOR problem with a particular set of initial weights [29]. With this arrangement, the neural network converged to zero error on the train set. This zero error on the training set was not achieved for the repeated trials with initial random weight arrangement. The reason for this could be due to the fact that the random weights end up in a similar point in the hypothesis space after training. It could also be that the hypothesis space is an extended plateau at the various points in which the neural network converges. It is a possibility that a particular combination of weights escapes all the local optima and thereby converges to zero train error. Convergence to the global optima occurred when the neural network was set up to solve the simple XOR problem.

TABLE III. TRAINING ERROR FOR SIMULTANEOUS VERSUS SUCCESSIVE LEARNING WITH THE SAME PARAMETERS FOR BOTH.

Network Structure – 2 layers (# nodes in each layer (a, b))	Momentum	Learning Rate	Training error-simultaneous		Training error-successive	
			Mean	S.D.	Mean	S.D.
(2,3)	0.33	0.2	0.2750	0.5086	0.5000	0.0000
(2,3)	0.5	0.2	0.2667	0.0432	0.5000	0.0000
(2,3)	0.5	0.3	0.2625	0.0381	0.5000	0.0000
(10,5)	0.5	0.3	0.2875	0.0583	0.5000	0.0000
(10,5)	0.4	0.2	0.2917	0.0599	0.5000	0.0000

TABLE IV. **TABLE 4**: MAPPING THE DATA SET TO SIMPLE XOR PROBLEM.

Color	Shape	Class
0	0	0
0	1	1
1	0	1
1	1	0

The training phase of the neural network models the human experiments conducted by Loess and Duncan [1]. In their experiments, no separate testing was done on the students. Their task was to learn to distinguish between the two tasks and the training phase of the neural network does exactly this. Hence its performance was judged by the training error in this domain and not by the test error. While modeling this experiment, there was no difference in the results obtained for the easy and difficult task, as the fourth attribute was irrelevant.

In order to vary the number of attributes, an appropriate data set had to be used in which the number of attributes could be increased progressively. Two sets of experiments were conducted with two types of data sets. One data set was an extension of the psychology experiment (Loess and Duncan) that increased the number of attributes and the other data set was the binary representation of numbers that made the neural network differentiate between odd and even numbers. In the original data set, circles (small or large) on white and squares (small or large) on gray belonged to set A and squares on white and circles on gray belonged to set B. The Loess and Duncan experiment was extended to incorporate more attributes such that circle (small/large with red/black border) on white and squares (small/large with red/black border) on grey belonged to set A and others to set B. The added feature was that the shape could be either red or black and the data set was as follows:

The data set (shown in Table V) had four attributes and could be easily extended to include more attributes. A graph was plotted for simultaneous versus successive learning in order to understand the trend of the performance with a variable number of attributes. In Figure 2, it is seen that as the number of attributes is increased, the performance of the simultaneous tends towards that of successive in this case, i.e. with three or more attributes, the training error for the successive case was 0.5 and with greater than 5 attributes, the training error for simultaneous was also 0.5.

Experiments were conducted to make the neural networks learn odd versus even numbers with the binary representation of the numbers. The number of attributes was increased by increasing the number of bits used to represent the number and thereby also increasing the number of instances in the data set. This experiment differs from the previous experiment presented in Table V, which had features added that affected the classification. The features added (i.e. more bits) were

TABLE V. DATA SET EXTENDED WITH MORE ATTRIBUTES.

Border Color	Background Color	Shape	Size	Class
0	0	0	0	0
0	0	1	0	1
0	0	0	1	0
0	0	1	1	1
0	1	1	0	0
0	1	0	0	1
0	1	1	1	0
0	1	0	1	1
1	0	0	0	0
1	0	1	0	1
1	0	0	1	0
1	0	1	1	1
1	1	1	0	0
1	1	0	0	1
1	1	1	1	0
1	1	0	1	1

irrelevant to classification of even vs. odd, since it only depends on the least significant bit. With respect to this task, both simultaneous and successive learning converged to zero error on the train set, unlike the previous experiments (the Loess and Duncan experiment modeled using neural network) in which the train error did not converge to 0. In Figure 3, the simultaneous and successive learning took a different number of iterations to converge to 0 error. As the number of attributes were increased starting with 3, simultaneous learning took less than half the number of iterations as that of successive learning to converge. As the number of attributes was increased, it was seen that the number of iterations taken by simultaneous to converge was almost as that of successive, and with 7 or more attributes, the number of iterations required to converge was the same for both types of learning. The following graph was plotted to compare the number of iterations taken to converge by simultaneous and successive learning.

In this case the number of iterations to converge, was plotted, since there was significant difference between the two types of learning in terms of the number of iterations taken to converge. This was not done with the previous experiments as there was little or no difference between the number of iterations taken to converge.

Series of experiment conducted validate the result published (simultaneous learning is faster than successive learning) by Loess and Duncan [1] for small number of threshold. However, it disputes their assertion that it will hold true for more complex classification problems (i.e. increased number of features). Neural network indicates that there exists a threshold, with respect to the number of attributes, beyond which the effect of simultaneous learning vs. successive learning is immaterial.

III. CONCLUSION

This paper compares the behavior and performance of neural networks against human subject experiment conducted by researchers in psychology. Experiments conducted on human subjects to gauge learning in a classification problem was modeled using neural networks. The results of these experiments did not completely parallel those obtained by the psychologists, Loess and Duncan [1]. Unlike the human

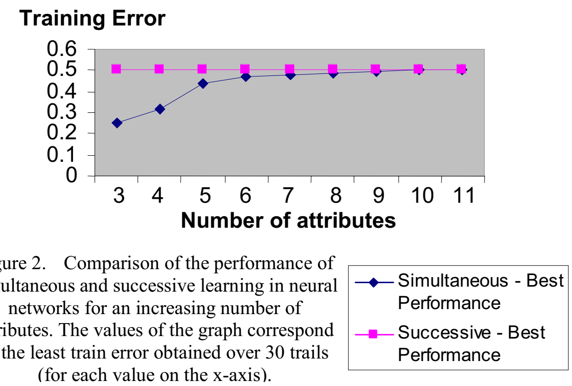

Figure 2. Comparison of the performance of simultaneous and successive learning in neural networks for an increasing number of attributes. The values of the graph correspond to the least train error obtained over 30 trails (for each value on the x-axis).

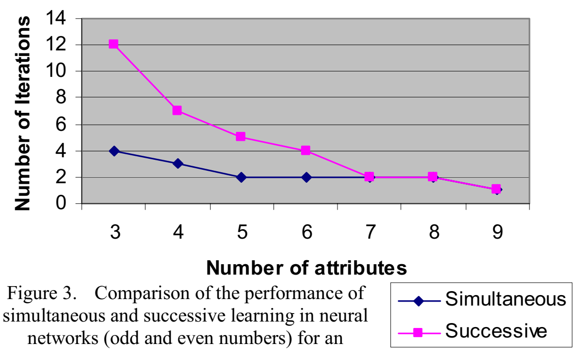

Figure 3. Comparison of the performance of simultaneous and successive learning in neural networks (odd and even numbers) for an increasing number of attributes

subjects, neural networks were not able to fully discriminate between two classes. This phenomenon arises from the fact that Loess and Duncan's experiment simplified into an XOR problem. Neural networks have difficulty converging to a solution in an XOR problem with random initialization of weights, but initial weights can be adjusted to learn the XOR problem. When initial weights were adjusted similar to XOR solution, neural network learn to fully learn the classification problem as human subjects in Loess and Duncan. More importantly, neural networks confirmed Loess and Duncan's main hypothesis that simultaneous learning was faster than successive learning as complexity increased (attribute range of 3 – 4).

Loess and Duncan further extended their hypothesis without formal experimental evidence that simultaneous would outperform successive as complexity increased beyond the scope of their human experiments. Interestingly, neural networks disagreed with their hypothesis.

Experiments conducted with an increasing number of attributes showed that as the number of attributes was increased, the two types of learning attained similar performances. This was seen with two different data sets: one extending the psychology experiment conducted by Loess and Duncan and one with odd/even numbers. The former showed that the two were similar with respect to train error when fully learned, but simultaneous learning required half the number of iterations compared to successive learning (attribute range: 3 – 5). However, when the number of attributes was increased beyond seven, relative performance was negligible. Odd/even number discrimination experiment mirrored the results seen in the extended Loess and Duncan experiment with attribute range of three to nine.

These experiments reflect the presence of a threshold with respect to the number of attributes, beyond which there seems to be no difference between simultaneous and successive learning. Loess and Duncan concluded that simultaneous was better than successive as complexity of classification with increased with irrelevant attribute. However, neural networks contradict their conclusions about more complex problems. As the task was made more difficult in terms of number of attributes and number of data instances, there was no difference between the two types of learning in terms of the train and test error, beyond a certain threshold with respect to the number of attributes. These types of experiments have not yet been conducted on human beings but it intuitively makes sense that when human beings are presented with complicated tasks with many attributes, the type of learning is immaterial. The reason is that with more attributes there is more room for confusion and the complexity of the problem would overwhelm any advantages one method has over the other.

It should be noted that although neural network and intuition support that effect of simultaneous and successive learning becomes negligible for very complex problems, it can only be verified by extending the human subject experiments by increasing the complexity. Furthermore, this paper has demonstrated that neural networks can be used as a rough model and give valuable insight into the problem, before the costly human subject experiments are conducted.

Similar future work comparing neural networks performance to existing experiments in psychology and neuroscience is needed to strengthen the promising link between behavior in neural network and biological systems. If this link can be created and verified, researchers would gain a tremendous tool in our endeavor to understand the human brain.

References

[1] H, G. Loess and C, P. Duncan, “Human discrimination learning with simultaneous and successive presentation of stimuli,” Journal of Experimental Psychology 44, 1952, 215-221.

[2] S. Jeyamkondan, D.S. Jayas, and R.A. Holley, “Neural Networks for modeling microbial growth,” North Central Inter-Sectional ASAE & CSAE Conference, Manitoba, 1990.

[3] G.M. Scott, J.W. Shavlik, and W.H. Ray, “Refining PID controllers using Neural Networks,” Neural Computation, 4(5), 1992, 746-757.

[4] D.A. Medler, “Brief history of Connectionism,” Biological Computation Project, Department of Psychology, University of Alberta, 1998.

[5] J.A. Anderson, A. Pellionisz, and E. Rosenfeld, editors. Neurocomputing 2, (MIT Press, Cambridge, MA., 1990).

[6] H. Spencer, The Principles of Psychology 1, (D. Appleton and Company, New York, 3rd edition, 1910).

[7] W. James, The Principles of Psychology 1, (Dover, New York, 1950).

[8] E.L. Thorndike, Selected Writings from a Connectionists Psychology, (Greenwood Press, New York, 1949).

[9] C.L. Hull, Principles of Behavior, (Appleton-Century-Crofts, New York, 1943).

[10] K. Lashley, “In search of the engram,” Symposia of the Society for Experimental Biology, 4, 1950, 454-482.

[11] D.O. Hebb, The Organization of Behaviour. (John Wiley & Sons, New York, 1949).

[12] W.S. McCulloch & W. Pitts, “A logical calculus of the ideas immanent in nervous activity.” Bulletin of Mathematical Biophysics, 5, 1943, 115-133.

[13] O.G. Selfridge, Pandemonium: “A paradigm for learning,” Proceedings of the Symposium on Mechanisation of Thought Processes, London, 1959, 511-529.

[14] M. Minsky and S.A. Papert, Perceptrons: An Introduction to Computational Geometry, (MIT Press, Cambridge, MA, expanded edition, 1988/1969).

[15] Stanford Encyclopedia of Philosophy (http://plato.stanford.edu).

[16] J.L. McClelland, “Retrieving general and specific information from stored knowledge of specifics,” Proceedings of the Third Annual Meeting of the Cognitive Science Society, 1981, 170-172.

[17] D.E. Rumelhart, J.L. McClelland, and the PDP Research Group, editors. Parallel Distributed Processing, 1, Foundations, MIT Press, Cambridge, MA., 1986.

[18] D.E. Rumelhart D, G.E. Hinton, and R.J. Williams, “Learning representations by back-propagating errors,” Nature 323, 1986, 533-536.

[19] M.R.W. Dawson and D.P. Schopflocher, “Modifying the generalized delta rule to train networks of non-monotonic processors for pattern classification,” Connection Science, 4, 1992, 19-31.

[20] J. Moody and C. J. Darken, “Fast learning in networks of locally tuned processing units,” Neural Computation, 1, 1989, 281-294.

[21] P. Weise and M.E. Bitterman, “Response-selection in discriminative learning,” Psychology Review, 58, 1951, 185-195.

[22] K.S. Lashley and M. Wade, “The Pavlovian theory of generalization,” Psychological Review, 53, 1946, 72-87.

[23] C.L. Hull, “The problem of stimulus equivalence in behaviour theory,” Psychological Review, 46, 1939, 9-30.

[24] K.W. Spence, “The nature of discrimination learning in animals,” Psychological Review, 43, 1936, 427-449.

[25] K.W. Spence, “The nature of response in discrimination learning,” Psychological Review, 59, 1952, 89-93.

[26] G.R. Grice, “Visual discrimination learning with simultaneous and successive presentation of stimuli,” Journal of Comparative Physiological Psychology, 42, 1949, 365-373.

[27] R.A. Baker and D.H. Lawrence, “The differential effects of simultaneous and successive stimuli presentation on transposition,” Journal of Comparative Physiological Psychology, 44, 1951, 378-382.

[28] R.E. Kandel, J.H. Schwartz, and T.M. Jessell, Principles of Neural Science, Fourth Edition, (McGraw-Hill Medical, 2000).

[29] M.T. Mitchell, Machine Learning, (McGraw Hill, 1997).