The Perception of the Perceptron: A History of Artificial Neural Networks

by Grant Yang

Period 4

Introduction

Today, neural networks are ubiquitous. They power a wide array of tools that many people rely on daily, from Google Translate to ChatGPT. Although their arrival into the limelight may seem relatively recent, neural networks have a long history spanning all the way back to the 1940s. By tracing the trajectory of their development through the years, we can better understand the origins and nature of these powerful (and sometimes mysterious) constructs that are so central to today's technologies.

Theory

One of the earliest works laying the foundations for modern neural networks was the 1943 paper "A Logical Calculus of Ideas Immanent in Nervous Activity," a collaboration between Walter Pitts and Warren McCulloch.[1] McCulloch and Pitts set out to study how the individual neuron cells give rise to the complex structure of the brain, and today their paper is recognized for providing solid mathematical and theoretical footing for the theory underlying artificial neural networks.[2]

In their paper, they proposed a simplified model of neurons that now bears their name: the McCulloch-Pitts artificial neuron. These neurons have the property of an "activation" which can only ever be 0 or 1, representing the state of a neuron firing.[3] This type of neuron fires with an activation of 1 only if the weighted sum of its inputs exceeds some threshold value.

More precisely, for n inputs, some threshold t, weights w_i, and input activations a_i, the neuron's activation will be:

$$H\left(\left[\sum_{i=1}^{n} a_i w_i\right] - t\right) \qquad \text{(Eq. 1)}$$

where H is the Heaviside step function

$$H(x) = \begin{cases} 1 \text{ if } x \geq 0 \\ 0 \text{ if } x < 0 \end{cases} \qquad \text{(Eq. 2)}$$

In McCulloch and Pitts's original paper, the duo only considered w_i weights of 0, 1, and -1, but today we consider w_i to be any real value. Modern neural networks contain a feed-forward arrangement of neurons, where inputs are fed into an initial set of neurons (called a "layer"), whose activations are then

---

[1] Michael Marsalli, "McCulloch-Pitts Neurons," in The Mind Project (Illinois State University), last modified 2007, accessed September 13, 2023, https://mind.ilstu.edu/curriculum/mcp_neurons/mcp_neuron_intermed6.html.

[2] Gualtiero Piccinini, "The First Computational Theory of Mind and Brain: A Close Look at McCulloch and Pitts's 'Logical Calculus of Ideas Immanent in Nervous Activity,'" Synthese 141, no. 2 (2004): 175, http://www.jstor.org.harker.idm.oclc.org/stable/20118476.

[3] Marsalli, "McCulloch-Pitts Neurons"

fed forward into the next layer, eventually reaching the output without any cycles or connections skipping between layers.[4]
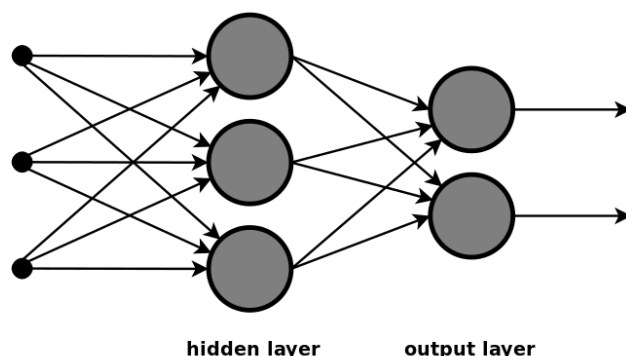

hidden layer   **output layer**

Figure 1:[5] A simple feed-forward neural network.

However, Pitts and McCulloch's research was not limited to only feed-forward networks, which they called networks of zero order. They defined the order of a network as the number of neurons that must be removed in order to remove all cycles from the network, and they proved that higher order networks are much more powerful than simple zero-order networks.[6] Confusingly, another popular definition for the order of a neural network is the number of connections that each hidden layer neuron has to the input layer, obviously unrelated to the definition used by McCulloch and Pitts.

Another influential early work was Donald Hebb's 1949 book *The Organization of Behavior*. His ideas about how learning occurs in brains (often summarized as "neurons that fire together wire together") would inspire a plethora of learning rules, including the family of unsupervised learning algorithms.[7] With unsupervised learning, the expected outputs are not known beforehand and thus cannot be used in computing the cost function.

In contrast to traditional computers using the von-Neumann architecture, this type of "parallel distributed processing" is not a sequential operation. Whereas a Turing machine processes one symbol at a time on a tape, many neurons must act together at the same time and interact with each other to process and store information throughout the network.[8]

Rosenblatt's Perceptron
One of the first physical implementations of these ideas is Frank Rosenblatt's Mark 1 Perceptron, built in 1958 using McCulloch-Pitts neurons. With this analog machine, he hoped to be able to classify images shown to a retina into 8 classes. The machine had only two layers and thus was severely limited, as each hidden layer neuron only had 6 non-zero weights connected to inputs on the retina.

---

[4]Mikel Olazaran, "A Sociological Study of the Official History of the Perceptrons Controversy," Social Studies of Science 26, no. 3 (1996): 614, http://www.jstor.org/stable/285702.

[5]Multi-Layer Neural Network Vector, illustration, Wikimedia Commons, April 10, 2015, accessed September 15, 2023, https://commons.wikimedia.org/wiki/File:Multi-Layer_Neural_Network-Vector.svg.

[6]Warren S. McCulloch and Walter Pitts, "A Logical Calculus of the Ideas Immanent in Nervous Activity," Bulletin of Mathematical Biology 52:101, accessed September 13, 2023, https://www.cs.cmu.edu/~./epxing/Class/10715/reading/McCulloch.and.Pitts.pdf.

[7]Richard E. Brown, "Donald O. Hebb and the Organization of Behavior: 17 years in the Writing," Molecular Brain 13 (2020): 1, https://doi.org/10.1186/s13041-020-00567-8.

[8]Olazaran, "A Sociological," 615.

Even then, the hidden layer's weights for its inputs could only be one of three values: +1, -1, or 0. Only the weights of the connections from the hidden layer to the output layer could be adjusted to arbitrary values.[9] Additionally, the output layer had additional functionality that allowed it to feed back information into the hidden layer, inhibiting any neurons that are mutually exclusive so that some outputs always take priority over others.[10]

These design choices very closely follow McCulloch and Pitts's 1943 paper. Taken together with the fact that Rosenblatt called the hidden layer an "association system," they are somewhat reminiscent of convolutional neural networks today. The idea of using the hidden layer to pick out the important features of an image is present, and the sparseness of the weights in the first layer is similar to the "kernels" used for convolution.

In a convolutional network, a small "kernel" (which essentially acts as a matrix multiplication) is applied in a specific pattern over small parts (usually 3x3) of the input image. The many individual outputs are then combined into another image, which is scaled down and fed into either another convolutional layer or a fully connected layer. This convolution operation can effectively utilize the information in an image's spatial layout where it would be completely lost in a fully-connected network. A convolutional network would not work as well if the pixels in the image were randomly scrambled with each other, but a fully-connected network would see no difference, as it has no sense of what order its input pixels are in.[11]

Although the specifics that make these convolutional networks so effective are missing from Rosenblatt's perceptron, some of the basic ideas are still recognizable.

Minsky and the AI Winter

In light of Rosenblatt's research and its extensive coverage in the media, the next critical moment in the history of neural nets was Marvin Minsky's 1969 book *Perceptrons* written in collaboration with Seymour Papert. In it, he provided precise mathematical proofs concerning the capabilities and limitations of neural networks. Popular history recounts that this put a stop to neural network related funding and research, as Minsky had proven that useful neural nets would be impossible to realize. However, in reality, interpretations of his ideas and his book varied. Many researchers in the field saw his work as irrelevant to what they were doing, as Minsky made only very narrow statements about the limited capabilities of single-layer networks and the difficulties in training multi-layer networks. In any case, neural networks were not abandoned completely, and they continued to be developed from psychological and neuroscience perspectives even as hopes for artificial intelligence were somewhat dashed.[12] Eventually, funding and research into neural networks would slowly return in the 1980s.

The Backpropagation Algorithm

The last major component for the neural networks we have today is the backpropagation learning algorithm. This algorithm was first theorized in 1974 by Paul Werbos and implemented in practice by David Rumelhart in 1980. In order for this approach to work, threshold functions other than a step

[9]Olazaran, "A Sociological," 618-19.

[10]Frank Rosenblatt, The Perceptron: A Perceiving and Recognizing Automaton (Cornell Aeronautical Laboratory, 1957), 4, accessed September 13, 2023, https://blogs.umass.edu/brain-wars/files/2016/03/rosenblatt-1957.pdf.

[11]Stanford University, "Convolutional Neural Networks," Deep Learning Tutorial, accessed September 14, 2023, http://deeplearning.stanford.edu/tutorial/supervised/ConvolutionalNeuralNetwork/.

[12]Olazaran, "A Sociological," 634.

function must be used so that its derivative would not be zero everywhere. Popular choices of threshold function include the logistic sigmoid activation function $\sigma(x)$:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{Eq. 3}$$

Also popular is the hyperbolic tangent function, $\tanh(x)$. This function has the exact same shape as the logistic sigmoid function and can be expressed as a scaled and shifted version of sigmoid:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} = 2 \cdot \sigma(2x) - 1 \tag{Eq. 4}$$

More recently, however, the ReLU activation function (which stands for **Re**ctified **L**inear **U**nit) has gained popularity.

$$\text{ReLU}(x) = \begin{cases} x \text{ if } x \geq 0 \\ 0 \text{ if } x < 0 \end{cases} \tag{Eq. 5}$$

For deep networks with many layers, using the ReLU function allows them to train significantly faster than sigmoid, but neurons do have a tendency to "die" by wandering into the zero-derivative parts of the activation function.[13] Tweaks like "leaky ReLU" and "GELU" fix this by adding a small slope to the area where $x < 0$ so that the gradient does not disappear.

$$\text{leaky ReLU(x)} = \begin{cases} x \text{ if } x \geq 0 \\ 0.01x \text{ if } x < 0 \end{cases} \tag{Eq. 6}$$
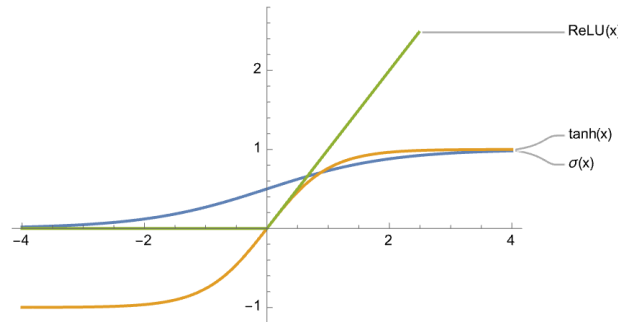

Figure 2: Graph of the various activation functions discussed above.

Conclusion

Neural networks as we know them today began their development nearly eighty years ago. From the idea of the McCulloch-Pitts artificial neuron and Frank Rosenblatt's Mark 1 Perceptron sprung the amazing results we have today. Although Marvin Minsky's 1969 book *Perceptrons* did have an impact on research in the field of artificial intelligence, research did not stop as these concepts still proved interesting to neuroscience. After research bounced back by 1980, modern architectures like the convolutional network and transformer would go on to conquer fields like machine translation and image recognition.

---

[13] Ayush Thakur, "ReLU Vs. Sigmoid Function in Deep Neural Networks," entry posted May 11, 2022, accessed September 14, 2023, https://wandb.ai/ayush-thakur/dl-question-bank/reports/ReLU-vs-Sigmoid-Function-in-Deep-Neural-Networks--VmlldzoyMDk0MzI.

Bibliography

1. Brown, Richard E. "Donald O. Hebb and the Organization of Behavior: 17 years in the Writing." Molecular Brain 13 (2020): 1-28. https://doi.org/10.1186/s13041-020-00567-8.

2. Hunt, Earl. "Review: 'Perceptrons' by Marvin Minsky." The American Journal of Psychology 84, no. 3 (1971): 445-47. https://doi.org/10.2307/1420478.

3. Marsalli, Michael. "McCulloch-Pitts Neurons." In The Mind Project. Illinois State University. Last modified 2007. Accessed September 13, 2023. https://mind.ilstu.edu/curriculum/mcp_neurons/mcp_neuron_intermed6.html.

4. McCulloch, Warren S., and Walter Pitts. "A Logical Calculus of the Ideas Immanent in Nervous Activity." Bulletin of Mathematical Biology 52. Accessed September 13, 2023. https://www.cs.cmu.edu/~./epxing/Class/10715/reading/McCulloch.and.Pitts.pdf.

5. Olazaran, Mikel. "A Sociological Study of the Official History of the Perceptrons Controversy." Social Studies of Science 26, no. 3 (1996): 611-59. http://www.jstor.org/stable/285702.

6. Piccinini, Gualtiero. "The First Computational Theory of Mind and Brain: A Close Look at McCulloch and Pitts's 'Logical Calculus of Ideas Immanent in Nervous Activity.'" Synthese 141, no. 2 (2004): 175-215. http://www.jstor.org.harker.idm.oclc.org/stable/20118476.

7. Rosenblatt, Frank. The Perceptron: A Perceiving and Recognizing Automaton. Cornell Aeronautical Laboratory, 1957. Accessed September 13, 2023. https://blogs.umass.edu/brain-wars/files/2016/03/rosenblatt-1957.pdf.

8. Stanford University. "Convolutional Neural Networks." Deep Learning Tutorial. Accessed September 14, 2023. http://deeplearning.stanford.edu/tutorial/supervised/ConvolutionalNeuralNetwork/.

9. Thakur, Ayush. "ReLU Vs. Sigmoid Function in Deep Neural Networks." Entry posted May 11, 2022. Accessed September 14, 2023. https://wandb.ai/ayush-thakur/dl-question-bank/reports/ReLU-vs-Sigmoid-Function-in-Deep-Neural-Networks--VmlldzoyMDk0MzI.

10. Multi-Layer Neural Network Vector. Illustration. Wikimedia Commons. April 10, 2015. Accessed September 15, 2023. *https://commons.wikimedia.org/wiki/File:Multi-Layer_Neural_Network-Vector.svg*