Activation Function Using Deep Neural Network

SOBITA ALAM, 19-40113-1, C, CSE, American International University-Bangladesh
TANZIMA ZAHIR, 18-38958-3, C, CSE, American International University-Bangladesh
RIDWAN MANNAN RAHAT, 18-38990-3, C, CSE, American International University-Bangladesh
MD.NAZMUL HASAN LIKHON, 18-38977-3, C, CSE, American International University-Bangladesh

The primary neural networks decision-making units are activation functions. Moreover, they evaluate the output of networks neural node; thus, they are essential for the performance of the whole network. Hence, it is critical to choose the most appropriate activation function in neural networks calculation. Numerous recipes have been developed over the years, though some of them are now considered obsolete due to their inability to function properly under certain conditions. These functions have a variety of characteristics that are thought to be necessary for successful learning. Some of these characteristics include their monotonicity, individual derivatives, and finite range .The commonly used additive functions, such as step function,linear Activation function, ReLU, Sigmoid, and so on, will be evaluated in this research paper. This will be followed by their properties, individual disadvantages and advantages, and specific formula application recommendations. [11]

Additional Key Words and Phrases: neural networks,activation function, step function, linear activation function, ReLU, Sigmoid

ACM Reference Format:
Sobita Alam, Tanzima Zahir, Ridwan Mannan Rahat, and MD.Nazmul Hasan Likhon. 2021. Activation Function Using Deep Neural Network. In *Bashundhara R/A '21: ACM Activation function using Neural Network, August 03–08, 2021, Bashundhara R/A, DHK*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Artificial neural networks and deep learning have enabled enormous progress in basic science applications. There are numerous deep learning neural network applications, including voice analysis, speech or pattern recognition, and object classification. Artificial neural networks make extensive use of activation functions and designing activation functions that allow for quick training of accurate deep neural networks is a contentious issue in research.[5]
When a neural network has multiple hidden layers, training becomes difficult and challenging. Some of these difficulties include zigzagging weight, vanishing gradient problems, overly complicated formulas, and saturation problems in the neural network of the activation function. This results in a lengthy process of continuous learning[11], [12]. Byrd et al.[13] discuss a practical and theoretical comparison of different activation functions presented in this research paper. Softplus, tanh, swish, linear, Maxout, sigmoid, Leaky ReLU, and ReLU are examples of activation functions. Each function's analysis will include a definition, a brief description, as well as its advantages and disadvantages.This will allow us to develop guidelines for selecting the best activation function for each situation.[8]

Thus, this paper is unique since it entails real-world applications of activation functions. Hence, it has a summary of

Manuscript submitted to ACM

the current trends in the usage and use of these functions against the state of the art research findings in practical deep learning deployments. The complication in this paper will enable suitable decisions about how to choose the appropriate and best activation function and implementing it in any given real-world application.[7]

2 LITERATURE REVIEW

2.1 Why we use Neural Network?

To simulate the human brain behavior and develop complex systems based on this Artificial Intelligence, the implementation of Artificial Neural Networks has become the basal stone of complex systems like Natural Language Processing, Autonomous Vehicles, and others trending technologies.[3]

Using a similar model to human neural networks has been possible to built systems with multiple inputs and specialized algorithms that taking advantage of this architecture can take decisions, as a human - with the advantages disadvantages that a machine has.

2.2 Step Function

The first question that the classification of the activation function should answer is whether or not neuron activation should occur. A person can only activate the neuron if the input value is greater than a certain threshold value, or leave it deactivated if the condition is not met. The equation below shows how a step function can be activated or deactivated.
f(x) = 0, for x T 1, for x > T... (1)

2.3 Linear Activation Function

Because gradient/slope descent cannot be progressive due to zero gradients, the step function does not update gradient during back propagation. As a result, rather than using the step function, we can try using a linear function. In the simplest case of a linear equation, the equation output can be equivalent to the input, but in most cases, the value of "a" varies with 1, where "a" is the proportional input activation. The following equation [Eq. 2.] shows a simple linear equation with the required variables.[6]
f(x) = a * x,where a R...(2)
Using linear activation, multiple neurons can be activated at the same time. In this case, if we choose multiple classes, we should prioritize the one with the highest value. However, there is still an activation issue in this situation. Consider the function's derivative, as shown in equation two below [Eq. 3.].[9]
f(x) = a... (3)
Also, it appears that issues with gradient descent for training occurs in this function too.Thus, it has a constant derivative function in all situations.

2.4 Sigmoid Function

A mathematical function with the features of a sigmoid curve is referred to as a sigmoid function. The combinations of the sigmoid function are not linear since this type of activation function is nonlinear. The figure below represents a demonstration of this function.Thus, it makes sense to stack layers. This applies to non-binary activation as well. It also has a smooth gradient value. Hence, this makes it suitable for shallow networks like functions simulation.[10]

2.5 Basic Rectified Linear Unit (ReLU)

The rectifier is an activation that assigns zero to values and value itself is above zero [Eq. 10.]. This is also known as a ramp function and is analogous to half-wave rectification in electrical engineering [Fig. 8.]. However, the activation function was first introduced in a dynamical network with strong biological motivations and mathematical justifications[30].[14]

f(x) = x ax(0,)+ = mx… (10)

2.6 Article Selection

From the initial phase of the search process, a total of 14 major research lists were produced. Then, the full text of the 14 major studies was analyzed. At the time of analysis, the relevance of the data to the study topic and the research's quality were considered.[9]

2.6.1 Keywords and Search String. By using those keywords we search in Science Direct, ACM, Google Scholar. a broadly automated keyword search to get an initial set of articles. After each check, the keywords and also the title then read the abstract .digital libraries are open source for data collection. using those keywords find out the suitable paper for this paper.[15]

2.6.2 Digital Libraries to Search. In digital libraries we search paper by using keyword and time duration. Looking advanced libraries alludes to looking and recovering data from inaccessible databases of digitized or computerized objects. These databases may hold either the metadata for an protest of intrigued (e.g. creator and title), or a total question such as a book or a video.[1]

2.6.3 keyword search and Manual Selection. The list of search that was used is "Activation Function "," Neural Network". From those process ended up with 14 articles there 9 journal articles and 4 web articles

2.6.4 Final set of Articles. Finally ended up with those process we select 14 articles (10 journal articles and 4 web articles). Those are suitable for the paper topic. Those paper keywords and abstract are mostly similar to this paper .so for research methodology or background study written based on those papers. All those articles are used in references .[13]

3 DISCUSSION

This section will compare various activation function functions and properties. It will involve a number of properties of the AFs discussed in the preceding summary, which will lead to an appropriate conclusion on the paper. The criteria for this analysis will be based on training speed and classification accuracy. It summarizes the analysis of the various AFs, along with their equations and ranges. With the exception of the final layer, which always uses softmax AF, all hidden nodes have the same activation function for each series of tests of the given formula.[4] The network was built with Python 3's keras and tensorflow frameworks. All operations were carried out on a single GPU unit, the Nvidia GeForce GTX 860M.[2]

Populate the following table with the required information.

FUNCTIONS	Comment	When to use
Step	Does not work with propagation algorithm	Rather Never
sigmoid	Prone to vanishing the gradient function and zizzagging during training due to not being zero centred	Can't into boolean get simulation
ReLU) First to go choice	The most popular function for hidden layer.Although under are circumtances	prone to the dying ReLU problem
Linear Activation function	Far more advance than activation function	Use as last resort

Table 1. Recommendation when to use which activation function in deep neural network

4 CONCLUSION

This paper shows that there is no ultimate answer for questions like "which activation function should I choose?" However, after this comprehensive summary of activation functions used in deep learning, we can make a few but certain recommendations based on the provided theory as shown in the table above. Thus, this paper summarizes comprehensively the activation functions applied in deep learning (DL) highlighting the current trends in the application of these functions, which is most essential and has never been published in any literature. It started by a presentation of a brief introduction on the activation function and deep learning, which was then followed by analysis of different AFs and a discussion of some applications fields of these functions can be used based on systems and architectures in the development of deep neural networks.[12]

5 CONTRIBUTION RECORD

Detail each group member contribution according to the following tables.

5.1 Paper writing contribution

Populate the following table with the required information.

Student id & name	Section No	Section Title
Sobita Alom,19-40113-1	1	Introduction
TANZIMA ZAHIR,(18-38958-3)	2.6	Article Selection,Abstract
RIDWAN MANNAN RAHAT,(18-38990-3)	2,3	Literature Review,Discussion
MD NAZMUL HASAN LIKHON,(18-38977-3)	4	Conclusion

Table 2. Section(s) Written in the paper by the group member

REFERENCES

[1] Adrish Banerjee, Akash Dubey, Abhishek Menon, Shubham Nanda, and Gora Chand Nandi. 2018. Speaker recognition using deep belief networks. *arXiv preprint arXiv:1805.08865* (2018).

[2] Giorno C. Richardson P. J. W. P.-O. Boone, L. 1998. Stock Market Fluctuations and Consumption Behaviour. *F. E. C., DEPARTMENT, D. E.* 3, 3, Article 208 (Nov. 1998).

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[4] Robert Hecht-Nielsen. 1992. Theory of the backpropagation neural network. In *Neural networks for perception*. Elsevier, 65–93.

[5] John Hertz, Anders Krogh, and Richard G Palmer. 2018. *Introduction to the theory of neural computation*. CRC Press.

[6] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* 25 (2012).

[7] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. 1989. Backpropagation applied to handwritten zip code recognition. *Neural computation* 1, 4 (1989), 541–551.

[8] Yann LeCun, Leon Bottou, Genevieve B Orr, Klaus-Robert Müller, et al. 1998. Neural networks: Tricks of the trade. *Springer Lecture Notes in Computer Sciences* 1524, 5-50 (1998), 6.

[9] Michael A Nielsen. 2015. *Neural networks and deep learning*. Vol. 25. Determination press San Francisco, CA, USA.

[10] Herbert Robbins and Sutton Monro. 1951. A stochastic approximation method. *The annals of mathematical statistics* (1951), 400–407.

[11] Joseph Scientist. 2009. The fountain of youth. Patent No. 12345, Filed July 1st., 2008, Issued Aug. 9th., 2009.

[12] Bhuriya D. Singh U. Sharma, A. 2017. Survey of stock market prediction using machine learning approach. *ICECA* 42, 1 (Jan. 2017), 259–268.

[13] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[14] Stephen M Stigler. 1981. Gauss and the invention of least squares. *the Annals of Statistics* (1981), 465–474.

[15] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1–9.