Review Article

Differences of Tanh, sigmoid and ReLu Activation Function in Neural network

Archana Tomar[1] and Dr. Pradeep Laxkar[2]

[1]*Research Scholar, Computer Science, Mandsaur University*
[2]*Associate Professor, Computer Application and Sciences, Mandsaur University*

ABSTRACT

In a neural network there are more than one Activation Function used for transferring the level of activation into an output. So for transferring activation level multiple common or similar activation functions are used. The selection of the best activation network on the basis of their performance is the main and important task. during this work, we want to propose a comparison of different activation function for enhance their performance in specific data set or the role and responsibility of the different form of activation functions and also mentioned about advantages and drawbacks of specific activation function and where we applied, so we can choose the acceptable and best activation functions to urge the best performance of ANNs.

KEYWORDS

Activation Function, Relu, sigmoid, TanH.

1. INTRODUCTION

For training the neural network in ANN, Activation function perform role. They give the essential nonlinearity of the model to have the option to learn complex portrayals. In ANN multiple neurons work with their corresponding weights and bias and get the outputs through passes the input. So if we don't use activation functions in ANN so here outputs can be produce in the range between infinite to -infinite [-Inf,+Inf] , so here the neurons really don't recognize the limits of the value.

For the estimation of accuracy neural network is work with multilayer and its accuracy is also depends on like what type of function we are using

Nonlinear AF (activation functions) are used in many cases for classification problems in neural network.

If we can't define Activation function in neural network so the neural network works just like a linear regression model. So, the output of this model is the produced as same as the given neurons input.

Generally, Transfer Function are another term of Activation function and this is used for producing better result. Between two neural networks architecture we use Activation function. Means, for getting the accurate outputs in ANNs, we need to update or change the applied weights and/or biases value which is add with neurons and this value is getting by the errors which is obtained through neural network processing. This process is called as Back-Propagation (BP) process.

In this analysis, we examined about a straightforward survey of the ongoing most encouraging enactment work that enhances ReLU or highlight another system who can be useful to utilize more established actuation capacities.

2. LITERATURE SURVEY

In the review, analyze the author Karlik [4] apply back-propagation algorithm for different activation function (AF) for the middle layer called hidden layer (middle layer) and output layer (o/p layer) neurons and examine their performance which is based on generalized MLP architectures. and For experimental comparisons authors use different forms of AF(activation function) like Bipolar sigmoid, sigmoid function , Tanh function and ReLu Function etc.

When we perform the grouping of multilayer perceptron and Generalized Delta rule learning, so for that here author work with five predictable differentiable and monotonic AF (activation functions).

Author also shows that some activation functions are nearest result which is a Tanh hyperbolic function.

Author Ding[6] discuss the performance improvement of neural networks with deep architectures and also discuss the developments of commonly used activation functions specifically, the definitions, the effects on the brain organizations network, and the benefits and weaknesses of quite a few activation functions so here pros and cons of several popular activation function reviewed but not investigated like maxout and softplus so in this paper author

focus on kind of the develop- ment growth ,attributes and most proper Activation function choice.

Author Nwankpa [7] presents a current AF (Activation Functions) utilized in profound learning appli cations and also traces the new patterns within the applications and utilization of those functions in reasonable profound learning organizations against the best-in-class research results. The author also introduced the deep learning concept in this paper, and after that designed different types of activation functions explained and for some specific applications all these functions were used for their improvement or enhancement of learning-based designs and frameworks.

Author Ramachandran [5] proposes a new formula which is applied in neural network (NN) and this is called Swish

function, and this is denoted as f(x) = x.sigmoid(x). Swish activation functions apply the properties of one-sided boundedness and non-monotonicity which is follow the one sided at zero with smoothness, which may assume a role in the noticed adequacy of Swish and comparable AF (activation capacities). In the experimental result the author shows that Swish will in general work in a way that it is more useful/better as compare to ReLU on more profound models across various testing datasets.

M. Lau and K. Lim [9] investigated DNN performance with different classes/categories of AF (activation functions) which is used adaptive activation functions and saturated and/or unsaturated function. for eliminating vanishing gradient issues in neural network saturated function is generally required a pre trained procedure. so if we use two parameter in tanh activation function, problem of the saturation can be resolved. And here author also explain about the Unsaturated activation functions where this AF can resolve vanishing gradient issues and give some sparsity toward the network so, it will not overfitted the network trainable variables or parameters. Due to allowing the network for learning by sample training functions here Adaptive activation functions got the smallest or lowest misclassification rate amongst all types /classes of the AF (activation function). So, here author shows that the saturation issue can be resolved through addon two paramours which is already trained by their experimental results.

Author Siddharth Sharma [18] provides a detail description of different activation functions which are generally used in deep learning concept and also explained about the importance/betterment of the AF(activation functions) and improving the presentation of ANN (artificial neural networks). So as my observation the author also explained here about the importance and their need of AF, which is generally used in neural network Architecture and model after that here we also observed that the author define the importance/need of AF for non-linearity in neural networks. At the beginning, the author gives a brief explanation of studied activation functions, then explained what are the need of the AF activation functions in NN and also define the necessity of non-linearity in NN. After that he explained about what kinds of activation functions we used in NN (neural networks) So here the author focuses on how to use activation functions AF on the middle layer called as hidden layer or middle layer of NN neural networks, also explained how conveniently arrange the another domains.

Author Xinru Li [19] found in NN which is some correlative properties between activation functions of Relu and Tanh AF(activation function). Here author also explained about the resultant of Tanh could expand the qualities actuated by Relu units and reduce the qualities cut by Relu units.

The author also shows that the if we use the weighted summation of Relu and Tanh AF activation function instead of Relu (ReLu AF) so resultant the networks can be found a countless enhancement. So for enhancement of the precision of ResNet here the author conducted some experiments on dataset. So the result/output of the Relu value is in positive form, the yield of the TF (Tanh function) which is additionally certain could upgrade the outcome of the Relu value. At the point if the conclusion of the Relu unit value is in zero value, the outcome of the TF Tanh function provides negative value which could introduce the information discarded by Relu AF. So, the conclusion of this that defined properties could improve the CNN convolutional NN (neural networks).

Author, Rahul [20] provides some new theoretical concepts which deal with ReLU function, ReLu have multiple variants like leaky ReLU or some another AF activation functions, so in this author use single hidden feedforward neural networks works. author also shows that the choice of AF activation function indirectly defines regularizer of NN that corresponds with semi norm which is defines as a Banach space.

Author Mercioni, showcase a new novel P-Swish activation function (Parametric Swish) [21], this model is suitable to bring and improvements of performance on object categorization/classification tasks using some datasets like CIFAR-10,100, but here author also used datasets for implementing the NLP. for testing purpose, here author used multiple types of architectures and check P-Swish provide better result as compared with other functions.

3. TYPES OF ACTIVATION FUNCTION

Now a days, ANN is a core component for performing any task related to handwriting recognition concept and some image compression technique or it can be used in stock exchange prediction technique and many more applications.

3.1 Sigmoid Function:-

Sigmoid function could also be a non-linear AF Activation Function that are generally used in feed-forward NN(neural networks).

it's a mathematical differentiable function, which generally distinct for real input parameter values and it also covering the positive (+ve) results everywhere with a selected degree of smoothness so for sigmoid function formula use here

Vanishing Gradient Drawback

In NN value of provided neuron weight is changed so at that time Vanishing gradient issues are occurred so this is happing during the back-propagation implementation. Vanishing problem generally understand by the change or raise the i/p value and after that result shows here the resultant o/p is present on specified range and its also manage the threshold point.

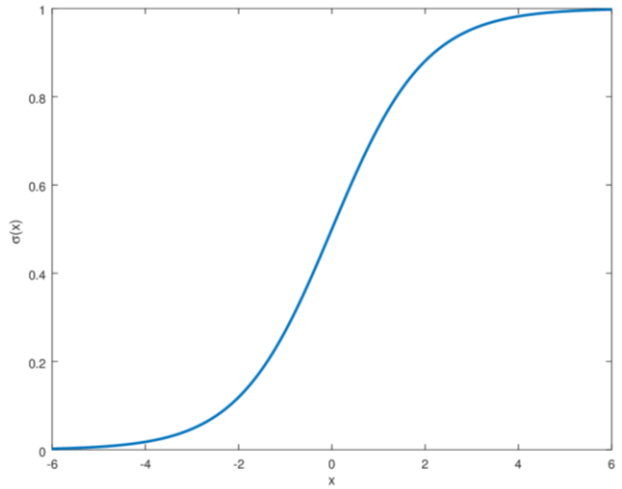


Figure 1. Sigmoid Function

When the sigmoid operate, the vary is from zero to one. we tend to all recognize that the utmost threshold price is one and thus the minimum price is zero. therefore, once we tend to enhance the input values, the anticipated output should lie virtually the higher threshold price that is one. so the anticipated output should be however or virtually the one.

3.2 Style Hyperbolic Tangent Activation Function (Tanh)

Another kind of AF Activation function called as hyperbolic tangent Activation function. Basically, Tanh is kind of smoother function or its also a or zero centered with range of (-1 ,1). So the formula for tanh function is: $2 * logistic(2x) - 1$

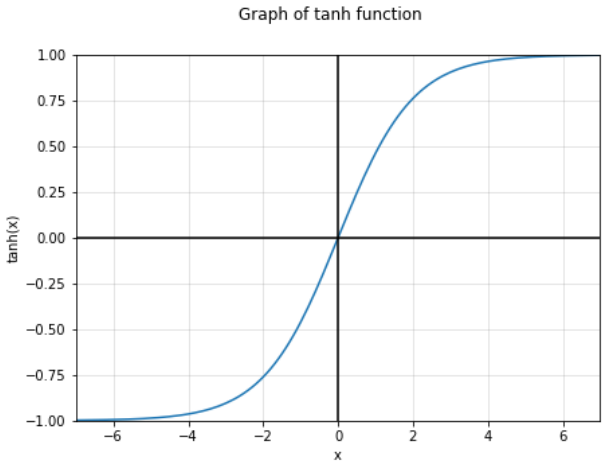


Figure 2. Tanh Function

3.2.1 Benefits of TanH AF: -

In the TanH function result of negative values also used for providing resultant and the lowest range of Tanh is -1.

Additionally, Tanh also referred to as the zero centered activation function.

3.2.2 Limitation of TanH AF:-

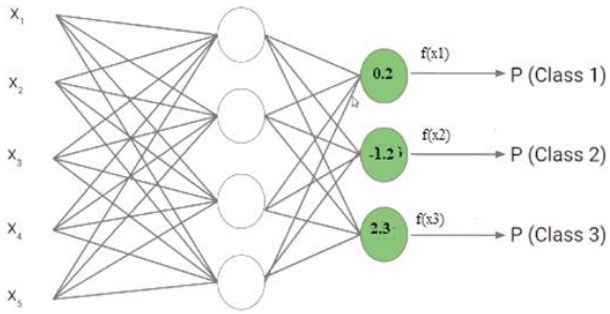
Vanishing Gradient issues occur also in TanH Activation function.

3.3 Softmax Activation Function

Softmax is another type of AF Activation function. It's basically a mathematical equation that is used in NN neural networks output layer to calculate and evaluate the probability distribution. And it also used in classification problem resolutions. And so the output of this function is represented between 0-1. So as our observation This function generally applied in every class neuron. So here is the equation of softmax function:

$$\boldsymbol{F(Xj) = e(x)/\Sigma e(x)}$$

Example: SoftMax applied in multilayer



Class1 exponential (0.2)/ exponential (0.2)+ exponential (-1.2)+ exponential (2.3)

Class2 exponential (-1.2)/ exponential (0.2)+ exponential (-1.2)+ exponential (2.3)

Class3 exponential (2.3)/ exponential (0.2)+ exponential (-1.2)+ exponential (2.3)

3.4 Rectified Linear Unit (ReLU)Function

Relu is mostly used function in network or deep learning concept. In neural network Relu work as a fast learning AF which is provide +ve or increasing order performance. Relu provide better result if we compared with sigmoid and tanh functions in terms of performance and generalization.

For making easier to optimize through with gradient descent method we use Relu function that hold the belongings of linear models.

The ReLU AF apply in each and every input neuron element and the resultant output of ReLu is same as the input values if it is (x>0) greater than zero otherwise it is set as 0. So, the ReLU equation is represented as:

$$\boldsymbol{f(x) = Max(0, Xj)}$$

$$\{\boldsymbol{Xj, if\ Xj >= 0,}$$

$$\boldsymbol{0, if\ Xj < 0}\}$$

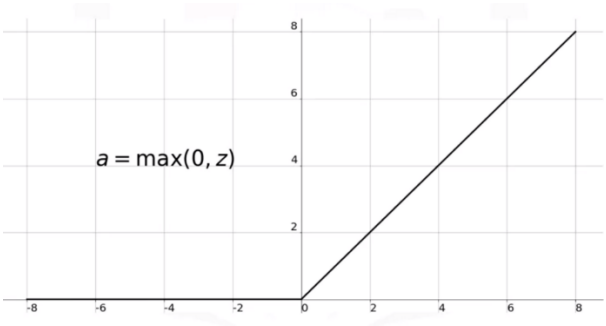


Figure 3. ReLu Function

3.5 **Comparison of Sigmoid, Tanh and Relu properties**

Table 1. Comparison

Activation Function	*sigmoid*	*Tanh*	*Relu*
Range	0-1	-1 to1	0 to inf
Vanish Problem	***more***	more	no
Type	non linear	non linear	linear
Accuracy	good	very good	excellent

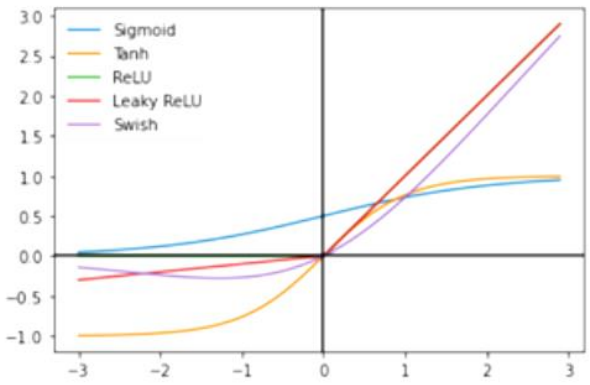


Figure 4. Comparison

4. CONCLUSION

When we compare in neural network NN all of the activation function like ReLU, ELU etc Activation function has performed better results compared with others AF because of this function we can resolved the issue of the vanishing gradient issues which generally a major issue for trained the network. And/or also decrease quality and working efficiency.

REFERENCES

[1]. K. B. and K. S., "Why Tanh: Choosing a Sigmoidal Function," Electronics Letters, 0-7803-0559-0, 192, 1992 IEEE.
[2]. K. H. and, "Simple Sigmoid-Like Activation Function Suitable for Digital Hardware Implementation," Electronics Letters, Vol. 28 ,1992.
[3]. S. Marra, M. Lachino, and F. Morabito, "Tanh-like Activation Function Implementation For High-performance Digital Neural Systems", C2006 IEEE.
[4]. B. Karlik ,A. Olgac, "Performance Analysis of Various Activation Functions in Generalized MLP Architectures of Neural Networks", IJAE International Journal of Artificial Intelligence And Expert Systems, no. 1, 2011.
[5]. P. Ramachandran, B. Zoph, Q. Le, "SWISH: A SELF-GATED ACTIVATION FUNCTION," 2017.
[6]. B. Ding, H. Qian, and J. Zhou, "Activation functions and their characteristics in deep neural networks", 2018.
[7]. C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, "Activation Functions: Comparison of Trends in Practice and Research for Deep Learning" , arXiv:1811.03378v1 [cs.LG], 8 Nov 2018.
[8]. J. Si, S. Harris, and E. Yfantis, "A Dynamic ReLU on Neural Network," IEEE 13th Dallas Circuits and Systems Conference (DCAS), 2018.
[9]. M. Lau and K. Lim, "Review of Adaptive Activation Function in Deep Neural Network," Biomedical Engineering and Sciences, 2018
[10]. K. Kirana, S. Wibawanto, N. Hidayah, G. Cahyono, and K. Asfani, "Improved Neural Network using Integral-RELU based Prevention Ac- tivation for Face Detection," 2019.
[11]. O. Sharma, "A New Activation Function for Deep Neural network", "International Conference on Machine Learning", Big Data, Cloud and Parallel Computing (Com-ITCon). IEEE, 2019.
[12]. Y. Guo1, Y. Sun1, Z. Zhang, H. He1, "Algorithm Research on Improving Activation Function of Convolutional Neural Networks", 2019.
[13]. Z. Qiumei, D. T. Fenghua, and W, "Improved Convolutional Neural Network Based on Fast Exponentially Linear Unit Activation Function", Digital Object Identifier , ACCESS 2019-2948112.
[14]. D. Misra, "Mish: A Self Regularized Non-Monotonic Activation Func- tion," 2019.
[15]. J. Feng and S. Lu, "PERFORMANCE ANALYSIS OF VARIOUS AC- TIVATION FUNCTIONS IN ARTIFICIAL NEURAL NETWORKS", J. Phys.: Conf. Ser, vol. 1237, pp. 22 030–22 030, 2019.
[16]. A. Rasamoelina, F. Adjailia, and P. S. c´ak, "A Review of Activation Function for Artificial Neural Network," SAMI 2020.
[17]. Y. Y. Adu, K. Tashi, N. Anokye, P. W. X, and A. M, "RMAF: Relu- Memristor-Like Activation Function for Deep Learning", VOLUME 8, 2020.
[18]. S. Sharma, S. Sharma, and A. Athaiya, "ACTIVATION FUNCTIONS IN NEURAL NETWORKS", International Journal of Engineering Applied Sciences and Technology, vol no. 04, pp. 310– 316, 2020. [Online]. Available: 10.33564/ijeast.2020.v04i12.054;https: //dx.doi.org/10.33564/ijeast.2020.v04i12.054
[19]. X. Hu, Z. Huang, and X, "Combine Relu with Tanh," IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference, vol. 2020, 2020.
[20]. R.Parhi,R.D. Nowak, "The Role of Neural Network Activation Functions", IEEE SIGNAL PROCESSING LETTERS, VOL. 27, 2020 1779.
[21]. M.A Mercioni and S. Holban, "P-Swish: Activation Function with Learnable Parameters Based on Swish Activation Function in Deep Learning", EEE Xplore, 978-1-7281-9513-1/20, 2020.