THE EFFECT OF Z-SCORE STANDARDIZATION ON BINARY INPUT DUE THE SPEED OF LEARNING IN BACK-PROPAGATION NEURAL NETWORK

Mohammed Z. Al-Faiz[1], Ali A. Ibrahim[2] Sarmad M. Hadi[3]
[1] Al-Farahidi University, Baghdad, Iraq
[2] College of Information Engineering, Al-Nahrain University, Baghdad, Iraq
[3] College of Information Engineering, Al-Nahrain University, Baghdad, Iraq
mzalfaiz12@ieee.org[1], aab@sc.nahrainuniv.edu.iq[2], sarmad@coie-nahrain.edu.iq[3]
Received:21/10/2018, Accepted:12/11/2018

Abstract– **The speed of learning in neural network environment is considered as an effective parameter in large data sets. This paper tries to minimize the time required for the neural network to fully understand and learn about the data by standardize input data. The paper showed that the Z-Score standardization of input data significantly decreased the number of epochs required for the network to learn. This paper also proved that the binary dataset is a serious limitation for the convergence of neural network, so the standardization is a must in such case where the 0's inputs simply neglect the connections in the neural network. The data set used in this paper are features extracted from gel electrophoresis images and that open the door for using artificial intelligence in such areas.**

Keywords: **Neural networks, Back-propagation, DNA, gel electrophoresis, standardization, Z-score.**

I. Introduction

Back-propagation neural network has proved itself as a very useful method in artificial intelligent field, due to the speed and accuracy, but in large datasets the story is different, back propagation takes a lot of time to learn about the data especially in low hamming-distance input data. The output itself also plays a role in the speed of learning especially due to the nature of operation in this paper which is input-output mapping where each input must be mapped to a unique and independent output, where this number denotes to a unique identifier for each input. The paper layout is as follows:

- Introduction
- The Z-Score standardization
- Back-Propagation
- Similar work
- Proposed work data
- Simulated results and discussion
- Conclusions

II. The Z-Score Standardization

The standard score (more commonly referred to as a z-score) is a very useful statistic because it (a) allows the programmer to calculate the probability of a score occurring within our normal distribution and (b) enables comparing two scores that are from different normal distributions. The standard score does this by converting (in other words, standardizing) scores in a normal distribution to z-scores in what becomes a standard normal distribution. [1] Each value of the input had been normalized according to Eq.1 below:

$$z(ij) = \frac{a(ij) - \mu}{\alpha} \tag{1}$$

Where:

z(ij) is the new value.

a(ij) is the old value.

μ =the mean of the column of the input value.

α =the standard deviation of the column of the input value.

i= $1 - $ n (n=the number of rows or inputs).

j= $1 - $ k (k=the number of bits that represents each input which are 32 bits).

III. Back Propagation

Back propagation is a tool used in artificial neural networks to sum up errors and learn the network about the input data and calculate error and then update weights in layers between input and output so it is supervised learning.

IV. Similar Work

Several attempts have been done in terms of standardization of input data, this paper introduces using Z-score with binary data-set which can be considered new due to preparing the data to be fed into neural network. According to [2], 6 data standardization methods were applied on a decimal data, in this paper these methods were applied to a simulated binary data but they didn't succeed to improve the speed of learning or decreasing the number of epochs required for the given network to learn. Two methods had been selected from because of the nature of the data where these methods apply for binary data (the simulated binary input) which where the followings: Vector standardization and Manhattan standardization [2] and also bipolar and the raw data (no standardization) had been applied to the proposed network structure and had been compared with the Z-Score standardization. [2] [3] Maximum linear standardization, Weitendorf's linear standardization, Peldschus' nonlinear standardization and Zavadskas and Turskis' logarithmic standardization had been excluded for arithmetic point of view (did not apply for the suggested input dataset). Lin Xiao and Bolin Liao [4] focused on accelerating the convergence in neural network using convergence-accelerated Zhang neural network, but the paper mainly focused on Lyapunov equation solving and did not address large datasets or other platforms. Ibrahim El-Henawy and Kareem Ahmed [5] also focused on the speed of learning in multilayer perceptron neural networks, a case study on character bit-mapped pixel image to ASCII conversion but again no addressing for the 0-1 problem or the nature of data in large datasets. A. Jafarian [6] proposed a model that focuses on scrutiny on the application of diverse learning methods in speed of convergence in neural networks in the solving of linear systems. Mainly, all the above papers did not address the need for a model that can handle large dataset for identification and classification using the artificial intelligence or machine learning and that what this paper addresses and what it is talking about. The proposed work in not new, but it addresses the problem of large dataset and accelerate the learning process, and in the terms of using Z-score with binary data, it is new because of the nature of the input data and the expected size of data-set.

V. Proposed Work Data

The data used in the learning process is a simulation based on features extracted from the DNA gel electrophoresis [7] images where sample of DNA is applied to one end of a gel near a negatively charged electrode, and this sample then migrates through the gel for a distance of 20 cm towards a positively charged electrode and then a digital image is taken and that is the input of the proposed case study as shown below in Fig.1 the features are: [8]

- Average intensity
- Standard deviation
- Smoothness
- Skewness
- Uniformity
- Entropy

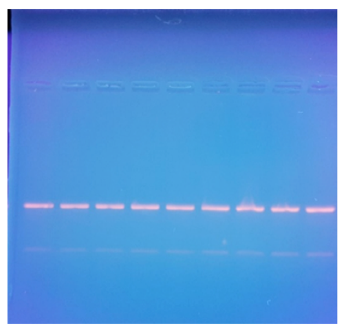

Figure 1: Gel electrophoresis image

The input data had been structures as shown below in Fig.2

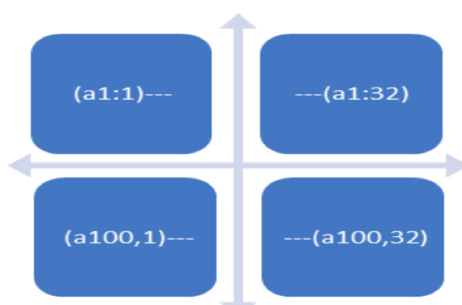

Figure 2: The structure of input data-set

Where one hundred inputs had been used, and each row in the above matrix represents a different input, and then each transposed row was fed into the neural network for learning, each row represents a single input, where each row is unique (had been generated without repetition because of the nature of DNA applications).

VI. Simulated Results and Discussion

The structure of the network used in this paper is as follows: (32-520-1), 32 binary inputs (32 neurons)-520 neurons in the hidden layed-1 decimal output (one neuron), where a transposed row of 32-bit input had been fed into the network at once and 520 neurons single hidden network has been used and a single output neuron , the number of neurons in hidden layer may appears large, but the error goal which is 10-5 made this number a must because of the high level of accuracy that this paper introduce. Different learning rates had been used and a maximum error of 10-5 is allowed as a check condition to complete the learning process, the below table shows the results of applying the standardization and standardization and the network behavior:

TABLE I
The Results of Leaning in the Proposed Neural Network Architecture

method	Learning rate	Time in seconds	epochs	Network
Raw data	0.00005	34.164	2477	yes
Raw data	0.0005	17.313	1140	yes
Raw data	0.005	-	Not applicable	
Raw data	0.05	-	Not applicable	
Raw data	0.5	-	Not applicable	
Manhattan	0.00005	-	>200000	no
Manhattan	0.0005	-	>200000	no
Manhattan	0.005	-	>200000	no
Manhattan	0.05	-	>200000	no
Manhattan	0.5	-	Not applicable	
Vector	0.00005	-	>200000	no
Vector	0.0005	-	>200000	no
Vector	0.005	92.336	7504	yes
Vector	0.05	-	Not applicable	
Vector	0.5	-	Not applicable	
bipolar	0.00005	9.594	696	yes
bipolar	0.0005	1.869	113	yes
bipolar	0.005	-	Not applicable	
bipolar	0.05	-	Not applicable	
bipolar	0.5	-	Not applicable	
Z-Score	0.00005	17.182	1198	yes
Z-Score	0.0005	1.813	102	yes
Z-Score	0.005	-	Not applicable	
Z-Score	0.05	-	Not applicable	
Z-Score	0.5	-	Not applicable	

and the following figures shows the behavior of the neural network where the best result of each standardization method had been chosen.

1) No-Standardization: The original data-set had been fed into the neural network and the learning process is as shown below in fig.3.

2) Manhattan Standardization: As shows in the below equation, the Manhattan standardization works as follows in Fig.4

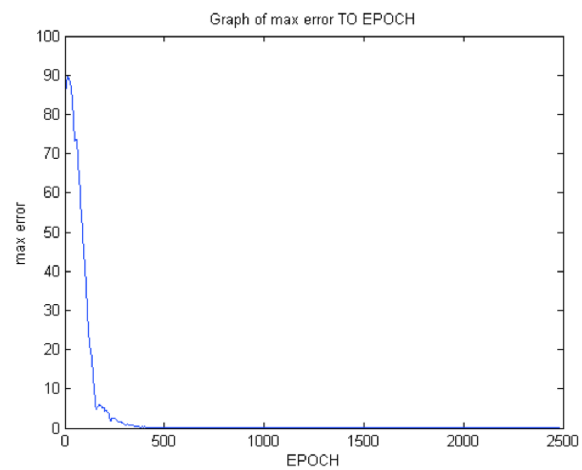

Figure 3: (0.0005) Learning rate without any standardization method

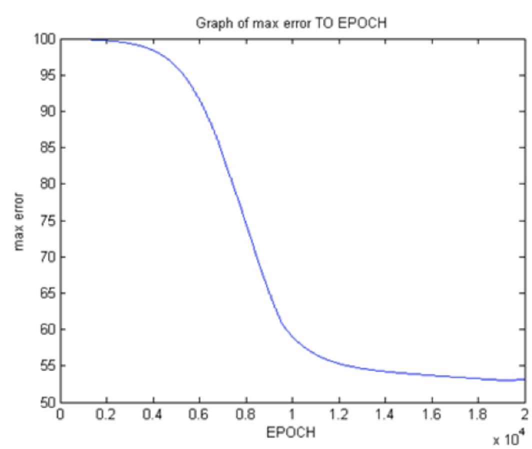

Figure 4: (0.00005) Learning rate using manhattan standardization method

$$z(ij) = \frac{a(ij)}{\sum_{i=1}^{n} a(ij)} \tag{2}$$

3) Vector Standardization: The following equation shows the behavior of this type of standardization along with Fig.5 that shows learning process:

$$z(ij) = \frac{a(ij)}{\sqrt{\sum_{i=1}^{n} a(ij)^2}} \tag{3}$$

4) Bipolar Linear Transformation: As the name explains it all, its either -1 or 1 according to the input data as shown below in the Eq. 4 and Fig.6

$$Z(ij) = -1 if a(ij) == 0 \tag{4}$$

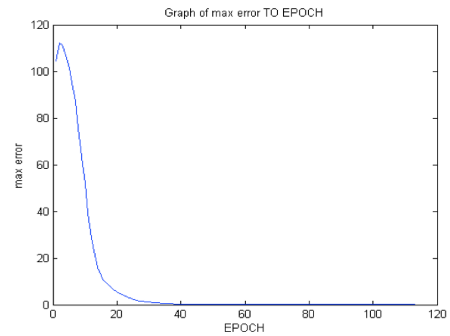

Figure 5: (0.005) Learning rate using vector standardization method

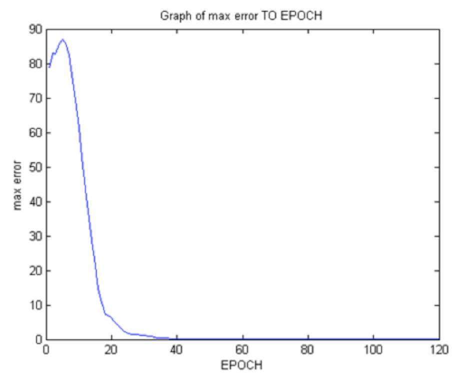

Figure 6: (0.005) Learning rate using vector standardization method

5) Z-score Standardization: As shown in Eq. 1, where this method shows the best results as Fig.7 shows.

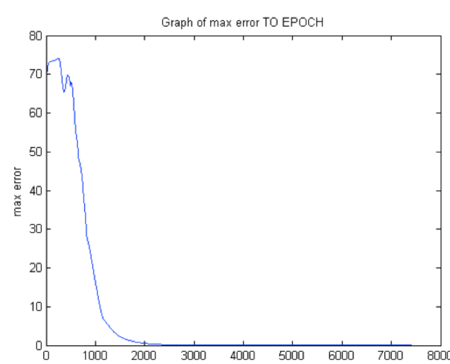

Figure 7: (0.0005) learning rate using Z-score standardization method

The above figures and table clearly show that the Z-score standardization method is the best method in the case of using binary input data. The 2000 epochs limiter had been used because the number of inputs (100) is not that much due to be learned by the neural network, more than 2000 epochs will be a waste of time and resources.

VII. Conclusion

From the results above, the following points had been concluded:

1-The obtained results show that nature of the data can deeply affect the behavior of the network.

2-Gausiisn standardization has been proved to be the best standardization method.

3- The Z-Score shows the best results but bipolar Standardization shows a promising result and might be useful in other platforms.

4-The Results of learning in the case of raw data clearly outframed the need for a standardization because of the nature of input data (binary data), where the (0) value of the input will deeply affect the learning behavior of the neural network.

5-The learning rate can deeply affect the behavior of the network.

References

[1] K. Kuzniar, M. Zajqc., Some methods of preprocessing input data for neural networks, Computer Assisted Methods in Engineering and Science, Institute of Fundamental Technological Research, Polish Academy of Sciences 22: 141−151, 2015.
[2] H. Anysza., A. Zbiciaka., N. Ibadova, The influence of input data standardization method on prediction accuracy of artificial neural networks, Procedia Engineering 153 66−70 , 2016.
[3] C. Timothy, Statistics in Plain English, Routledge, Third edition, 2010.
[4] L. Xiao, B. Liao, A convergence-accelerated Zhang neural network and its solution application to Lyapunov equation, Neurocomputing 193, March 2016.
[5] I. El-Henawy , K. Ahmed, Accelerating convergence of backpropagation for multilayer perceptron neural networks: a case study on character bit mapped pixel image to ASCII conversion, International Journal of Computers and Applications ,Volume 38, 2016 - Issue 1
[6] A. Jafarian, On the convergence speed of artificial neural networks in the solving of linear systems, Int. J. Industrial Mathematics (ISSN 2008− 5621) Vol. 7, No. 1, 2015
[7] R. Calladine , R.. Horace, F. Drew Ben Luisi , A. . Andrew Travers, Understanding DNA, The Molecule and How It Works, Academic Press, Third Edition 2004.
[8] F. Cai1, S. Liu, P. Ten Dijke, F. J. Verbeek, Image Analysis and Pattern Extraction of Proteins Classes from One-Dimensional Gels Electrophoresis, International Journal of Bioscience, Biochemistry and Bioinformatics, Volume 7, Number 4, October 2017, Pages (201 − 212).