# Data Encoding and network training on Human Eye Gaze Pattern Recognition Data

**Zixin Wang**

College of Engineering & Computer Science
The Australian National University, Canberra ACT 2600, Australia
Assignment 2 for Neural Networks, Deep Learning and Bio-inspired Computing
u5884895@anu.edu.au

**Abstract.** Data encoding is one of the most important aspects for training neural networks. Choosing appropriate pattern representation is also essential for the training of the neural networks. In this paper, we will use human eye gaze pattern recognition data, analysis different strategy of data encoding and preprocess method to train the simple fully connected neural network. Compared with sigmoidal normalization, min-max normalization and statistical Z function normalization, human eye gaze pattern recognition data perform well on sigmoidal normalization and Z-score normalization. Choosing appropriate preprocessing and data encoding strategy will improve the performance. We also using deep learning method to train a Long short-term memory (LSTM) recurrent neural network and compared the results. LSTM recurrent neural network also perform well on the human eye gaze recognition data.

**Keywords:** Neural networks, data analysis, data encoding, data preprocessing, normalization, LSTM

## 1 Introduction

Human eye gaze and the movement recognition are useful in various research area, it has drawn people's attention in the past few years. In [1] eye gaze detection and tracking can be used in different fields of applications, such as human-computer interaction, driving assistance systems, and assistive technologies. In this paper, we used the same human eye gaze pattern recognition data in [2], do the network training by using different method. Data are collecting from 10 volunteers from the Australian National University community. There are five fixations in the eye gaze patterns. The inputs of the data are the vertical distance between fixations. The task is to figure out whether the eye gaze pattern is recognized as face scanning or text scanning

In this paper, we will first analysis the input data by using the data analysis technique and encoding technique mentioned in [3], The result of analysis is that the data is right skewed and has some large outliers. We then present some decision we made from the analysis, the sigmoidal normalization might be a good method for this particular data. We trained the simple fully connected neural network (with topology 4-10-2) from different methods and discussed the results. It seems that data encoding and preprocessing may not always get better result, the min-max normalization does not perform well on the data. We should do the data analysis and find a appropriate method before training. The sigmoidal normalization and Z-score normalization can improve the performance at this time. Moreover, we trained a Long short-term memory (LSTM) recurrent neural network by using the same data. We use the same training and testing data provided in [2] and the results are slightly improved compared with their simple SVM based classifier.

## 2 Data Analysis

The human eye gaze pattern recognition data has 4 input columns, which was the vertical distance between each fixation. As mentioned in [2], The human eye gaze pattern recognition data are generated from 10 volunteers. Each participant will be asked to view one sets of picture and text documents. The eye movement data will be collected. The original collected gaze points are filtered into fixation which can be easier to use and interpretable. Since it is possible to interpret a plausible eye gaze pattern from the early stage of face viewing, the first five fixation are used. The fixation for text scanning and face viewing will be different as text scanning follows a horizontal pattern and face scanning has more complicated geometrical shape. The input for the human eye gaze pattern recognition data are the distances between five fixations, therefore we analysis the vertical distance between fixation 1&2, 2&3,3&4, and between 4 &5 and try to classify the pattern by using the vertical distance between fixation.

### 2.1 Statistics:

The following table and figures show the statistics and distribution for our input data.

**Table 1.** statistics for input training data.

| | 1&2 | 2&3 | 3&4 | 4&5 |
|---|---|---|---|---|
| Maximum Value | 551 | 660 | 652 | 474 |
| Minimum Value | 0 | 1 | 0 | 0 |
| Average Value | 63.78 | 63.89 | 60.23 | 60.47 |
| Standard Deviation | 85.56 | 85.10 | 85.96 | 76.67 |
| Average Absolute Deviation | 57.27 | 54.20 | 56.15 | 52.41 |

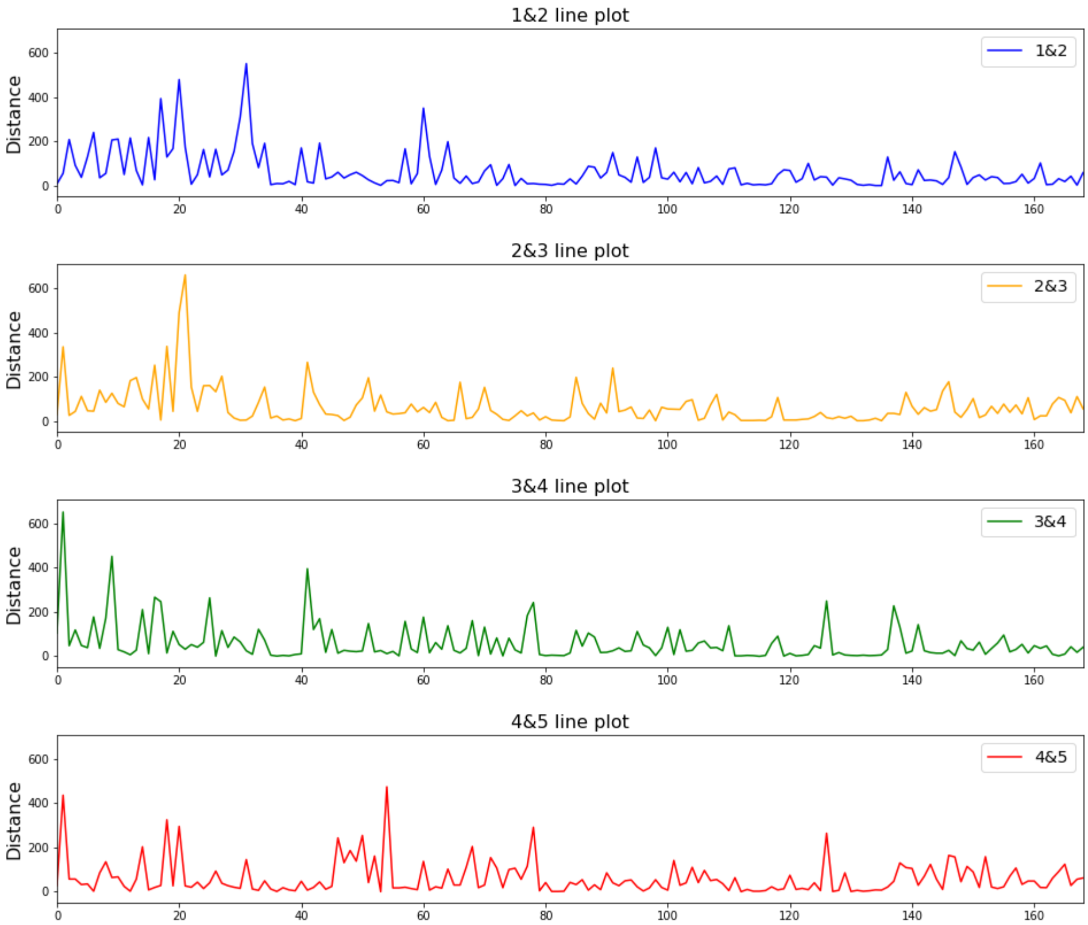


**Figure 1**. line plot for input training data

### 2.2 Encoding Decision

As in figure 2, Data are right skewed, we need to normalize data over the for the network. From the distributions and line plots below, we can see that the outlier of the data points has quite large values. There are a few outliers in the distributions and a few peaks in the line plots.
As in [4], the sigmoidal normalization is an appropriate approach to represent large outlier data, we can normalize our input data by using sigmoidal normalization.
After applied the sigmoidal normalization to the input data, the distribution looks much better in figure 3 compared to the result in figure2. It seems that the outlier has been fixed.
Since the human eye gaze pattern recognition data is a timeseries data, we might also use the moving average method for our input data.

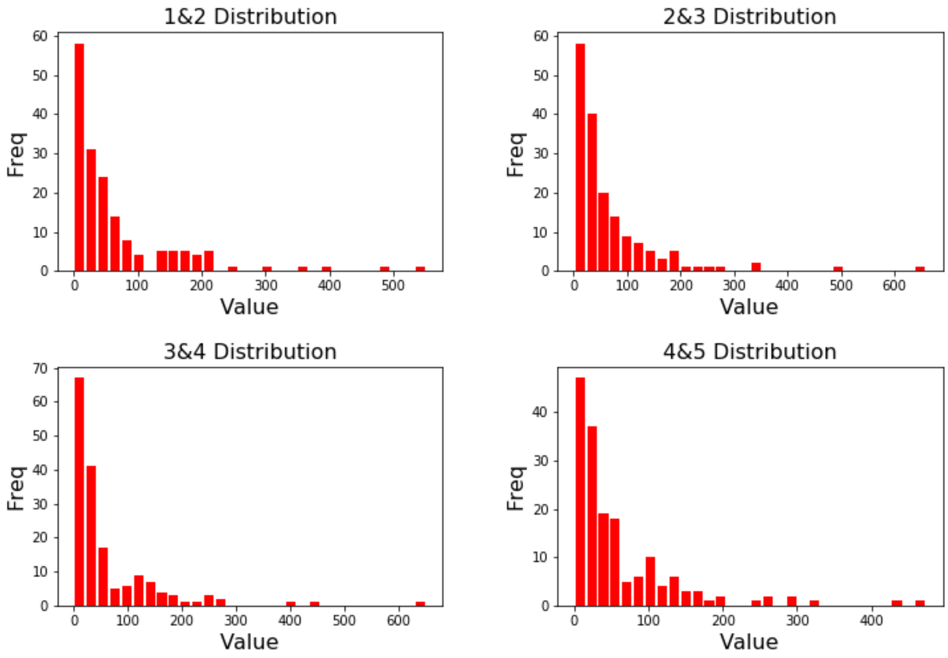


**Figure 2.** distributions for input training data

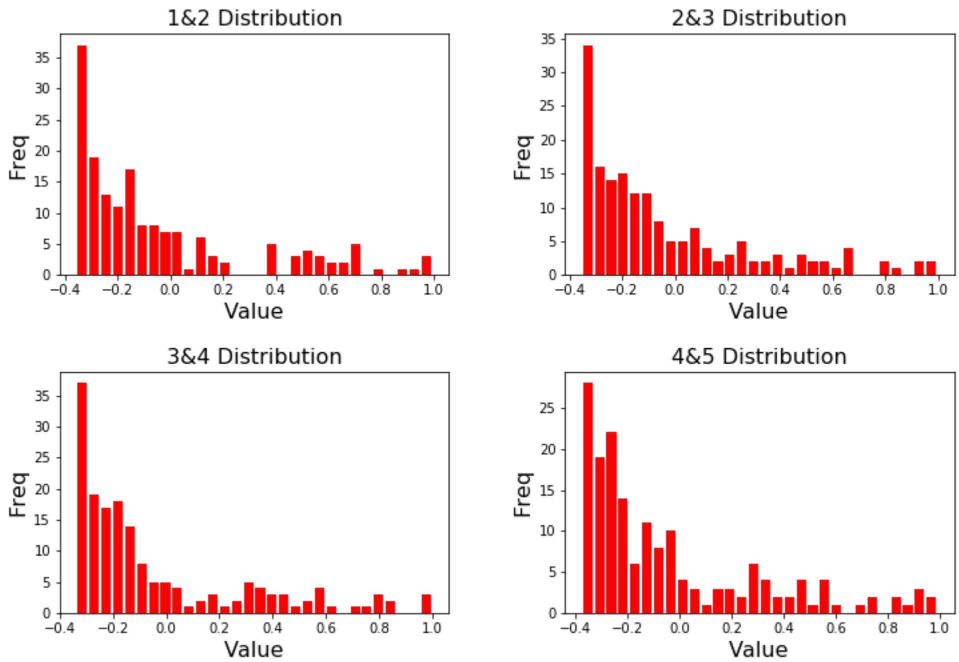


**Figure 3**. distributions for input training data after sigmoidal normalization

## 3 Method

Data encoding and preprocessing is quite useful and important for training neural network, different preprocessing methods may lead to different results. In this study, we will train a simple two-layer fully connected neural network and compare with networks that using three different normalize methods: sigmodal normalization, min-max normalization (simple linear squashing) and statistical Z function mentioned in [3].

We trained the neural network with and without the data encoding to compare the result and the difference between each neural network.
After training a simple fully connected neural network, we try to use the LSTM to train our data and evaluate different results.

### 3.1 Preprocess Function

**Sigmoidal Function:**

$$x' = \frac{1 - e^a}{1 + e^a}$$

where

$$a = \frac{x - \overline{x}}{\sigma}$$

$x'$ is the new convert value and x is the old value, $\overline{x}$ is the mean and $\sigma$ is the standard deviation, the data will be normalized to -1 to 1 by using these functions

**Min-max Function:**

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

where

$x'$ is the new convert value and x is the old value. the data will be convert to the data in range 0-1 by using this functions.

**Z-score Function:**

$$x' = \frac{x - \overline{x}}{\sigma}$$

where

$x'$ is the new convert value and x is the old value, $\overline{x}$ is the mean and $\sigma$ is the standard deviation.

### 3.2 Fully Connected Neural Network

We use two-layer network to train the model, there are one input layer, one hidden layer and one output layer in the network, layers are linearly transformed between each layer. We use a sigmoid function as the activation function and cross entropy loss for the loss function. The network is trained by back-propagation and the stochastic gradient descent algorithm are used as our optimizer. After the several training, the ReLU activation function and Adam optimizer are also used for comparing the results.
The input data is the human eye gaze pattern recognition data, which was 4 vertical distance between each fixation. The output will be 0/1, which represent whether the eye gaze pattern is recognized as face scanning or text scanning. It is a binary classification task.
We use the same training data and testing data provided in [2], which originally has 169 training samples and 154 testing samples.

### 3.3 Recurrent Neural Network (RNN)

A recurrent neural network can cycle through the same activation and weight connection for each time steps, it has a feedback connection from its output back to its input. As mentioned in [5], for each time step $t$, the hidden node value $h$ is update by:

$$h^{(t)} = \sigma(W^{hx}x^{(t)} + W^{hh}h^{(t-1)} + b_h)$$

where

$\sigma$ is the logistic sigmoid function, $W^{hx}$ is the conventional weight matrix between input and hidden layer, $x^{(t)}$ is the current data point, $W^{hh}$ is the weight matrix between the hidden layer and itself at adjacent time steps, $b_h$ is a standard bias parameters.

And output $y$ is given by:

$$y^t = softmax(W^{yh}h^{(t)} + b_y)$$

where

$W^{yh}$ is weight matrix and $b_y$ is the bias parameters which allow each node to learn an offset.

### 3.4 Long Short-Term Memory (LSTM) Recurrent Neural Networks

Long short-term memory recurrent neural networks follow the same basic structure as recurrent neural networks except the nonlinear units in the hidden layer are replaced by memory blocks. It can solve vanishing gradient problem in recurrent neural networks [6]. As mentioned in [5] the computation for each time steps are shown below:

$$i^{(t)} = \sigma(W^{ix}x^{(t)} + W^{ih}h^{(t-1)} + b_i)$$
$$f^{(t)} = \sigma(W^{fx}x^{(t)} + W^{fh}h^{(t-1)} + b_f)$$
$$g^{(t)} = \varphi(W^{gx}x^{(t)} + W^{gh}h^{(t-1)} + b_g)$$
$$o^{(t)} = \sigma(W^{ix}x^{(t)} + W^{ih}h^{(t-1)} + b_o)$$
$$s^{(t)} = g^{(t)} \odot i^t + s^{(t-1)} \odot f^{(t)})$$
$$h^{(t)} = \varphi\,(s^{(t)}) \odot o^{(t)}$$

where

$i^{(t)}, f^{(t)}, g^{(t)}, o^{(t)}$ are the input gates, forget gates, cell gates and output gates respectively. $h^{(t)}$ is the hidden state at time step t. $W$ is weight matrix and $b$ is the bias parameters.

The structure of LSTM recurrent neural network is quite similar to our fully connected neural network. We used the same datasets, which originally have 169 training samples and 154 testing samples. We can treat our human eye gaze pattern recognition data as a sentence. Since each fixation are in time series, there should be some relations between each fixation within time. We can treat the distance of the fixation as the word and four distances combined as sentence which describe the different eye gaze patterns. Therefore, for the structure of the Long Short-Term Memory Recurrent Neural Networks, the input size is 4, which represent four input distance between each fixation. The size for output layer is 1, the result for the final output will be in range 0 to 1 after applied the sigmoid functions which represents the probability of either outcome 0 or outcome 1. For values greater than 0.5, the outcome will be 1. For value equal or below 0.5, the outcome will be 0.

## 4 Result and Discussion

Table 2 and Table 3 shows the accuracy of the recognition of the human eye gaze pattern by using the fully connected neural networks. The random seed are set to 1 to ensure consistency. The same training data and testing data provided in [2] are used, which originally has 169 training samples and 154 testing samples. The hyperparameter are set intuitively and changed by experiment. The learning rate is 0.01, numbers of hidden neurons is 10 and num of epochs is set to 2000. Those parameters are same in each trial of training.

**Table 2.** result for test accuracy using SGD algorithm with different activation functions

| | WITHOUT PREPROCES | SIGMOIDAL | MIN-MAX | Z-SCORE |
|---|---|---|---|---|
| SIGMOID | 94.81% | 96.10% | 92.21% | 96.10% |
| RELU | 95.45% | 96.10% | 94.16% | 96.10% |

We then use Adam algorithm to optimize out model.

**Table 3.** result for test accuracy using Adam algorithm with different activation functions.

| | WITHOUT PREPROCES | SIGMOIDAL | MIN-MAX | Z-SCORE |
|---|---|---|---|---|
| SIGMOID | 94.16% | 96.10% | 92.86% | 96.75% |
| RELU | 94.81% | 95.45% | 91.56% | 96.10% |

Compared with the two tables, we can see that data encoding is quite useful for training the neural networks. Different activation function and different optimize algorithm will also have impact on the result of the network. Despite of the any of the activation functions and optimize algorithm, we can see that Sigmoidal normalization and Z-score normalization get a better performance on the human eye gaze pattern recognition data. The data analysis and data encoding technique mentioned in [2] is also useful for training the neural network. From the result, as the data are right skewed and have some large outliers, sigmoidal and Z-score normalization is more appropriate this time. However, min-max normalization does not work quiet well on the data. It seems that data encoding might not always improve the results. We should analysis our data before training the neural network, as the improper data encoding might lead to a worse neural network.

For the Long Short-Term Memory (LSTM) Recurrent Neural Networks, the learning rate is set to 0.001 and numbers of epochs is 2000. The testing result are shown in table 4.

**Table 4.** test result for Long Short-Term Memory (LSTM) Recurrent Neural Networks.

| EXPERIMENT | MEAN SQUARED ERROR(MSE) | CLASSIFICATION ERROR (CLE) | ACCURACY |
|---|---|---|---|
| Test | 0.02062 | 3.9% | 96.10% |

As mentioned in [2], the simple SVM based classifier constructed by using a linear kernel gives highly accurate classification results which has Test CLE 4.55%. Compared with the result shown in Table 4, our result CLE is slightly better, the CLE is 3.9% which is less than 4.55%.

## 5 Conclusion and future work

In conclusion, data analysis and preprocessing is quite useful before training the neural networks. The pre-processing techniques increase the robustness of the proposed algorithm and improve the performance of the neural networks [7]. We should choose the appropriate normalization method which suit our training data. As we presented, for the human eye gaze pattern recognition data, Sigmodal normalization and Z-score normalization might performant well since the data is right skewed and has some large outliers. As the human eye gaze pattern recognition data is a time series data, The Long Short-Term Memory (LSTM) Recurrent Neural Networks also perform well on the human eye gaze recognition data. The test result for LSTM Recurrent Neural Network is slightly better than the result in [2].
The parameter we used for training are intuitively selected and changed by numbers of experiments. In the future, we may apply evolutionary algorithms like genetic algorithm to find better parameters for the neural networks.

## References


1. Al-Rahayfeh, A., & Faezipour, M. (2013). Eye tracking and head movement detection: A state-of-art survey. IEEE journal of translational engineering in health and medicine, 1, 2100212-2100212.
2. Zhu D., Mendis B.S.U., Gedeon T., Asthana A., Goecke R. (2009) A Hybrid Fuzzy Approach for Human Eye Gaze Pattern Recognition. In: Köppen M., Kasabov N., Coghill G. (eds) Advances in Neuro-Information Processing. ICONIP 2008. Lecture Notes in Computer Science, vol 5507. Springer, Berlin, Heidelberg
3. Bustos R.A., Gedeon T.D. (1995) Decrypting Neural Network Data: A Gis Case Study. In: Artificial Neural Nets and Genetic Algorithms. Springer, Vienna
4. Li, H., Chen, C. P., & Huang, H. P. (2000). Fuzzy neural intelligent systems: Mathematical foundation and the applications in engineering. CRC Press.
5. Lipton, Zachary. (2015). A Critical Review of Recurrent Neural Networks for Sequence Learning.
6. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.
7. Nazri Mohd Nawi, Walid Hasen Atomi, M.Z. Rehman, (2013). The Effect of Data Pre-processing on Optimized Training of Artificial Neural Networks, Procedia Technology,Volume 11, Pages 32-39, ISSN 2212-0173,

8 Karim, F., Majumdar, S., Darabi, H., & Chen, S. (2017). LSTM fully convolutional networks for time series classification. IEEE access, 6, 1662-1669.