International Journal of Technology (2017) 6: 1082-1088
ISSN 2086-9614 © IJTech 2017

DEVELOPMENT OF EYE FIXATION POINTS PREDICTION MODEL FROM EYE TRACKING DATA USING NEURAL NETWORK

Boy Nurtjahyo Moch[1*], Komarudin[1], Maulana Senjaya Susilo[1]

[1]*Department of Industrial Engineering, Faculty of Engineering, Universitas Indonesia, Kampus UI Depok, Depok 16424, Indonesia*

(Received: February 2017 / Revised: May 2017 / Accepted: November 2017)

ABSTRACT

Fixation points, as the stopping location of eye movements, can be extracted to generate valuable information about a picture or an object. This information is valuable as it enables the identification of the area/part of the picture that attracts people's attention, which can be used as a consideration when making decisions in the future, for example in marketing. For this reason, in this study, a Neural Network (NN) model was developed to predict the fixation points of a picture. Specifically, the authors experimented with various transfer and training functions in the NN in order to determine which causes the fewest errors. The results show that the method used is applicable in practice since it produces MAPE (Mean Absolute Percent Error) of around 13–15% and MSE (Mean Squared Error) of 0.9–1.1%.

Keywords: Eye tracking; Fixation points; Neural network; MAPE; MSE

1. INTRODUCTION

Most of the information that consciously enters the human brain comes in the form of visible stimuli. By analyzing the direction of human eye motion, we can identify the areas of an image that are considered to be important (Duchowski, 2007). However, the thinking process occurs only in the mind of human in question, while we, as outsiders, can only observe the outcome after the action has been taken. So, what if we want to understand the experience of 10 or 100 people one month ago, or several different events experienced by a person in the past? This can only be achieved by identifying the information people receive through the eyes and recording them so that the information can be processed further whenever the need arises.

One method that can be employed to address these problems is eye tracking (Kenyon, 1985). This method helps researchers to determine where and how many times respondents focus on a certain picture or object, along with their eye movement sequences from one point to another on an object (Drewes, 2010).

Essentially, the eye tracking method recognizes two types of eye movement, i.e. fixation and saccade. Fixation refers to when the eyes focus in observing a certain point on an object; in this case, the eyes do not move. Meanwhile, saccade refers to rapid eye movement between fixations which helps a person to obtain a complete picture of the object he/she sees (Schall & Bergstrom, 2014). Figure 1 illustrates the difference between fixation (red circle) and saccade (red line). Based on fixation points, we can determine which information is considered to be useful or is given more attention by the viewer.

*Corresponding author's email: boymoch@eng.ui.ac.id, Tel: +62-21-78888805, Fax: +62-21-78885656
Permalink/DOI: https://doi.org/10.14716/ijtech.v8i6.717

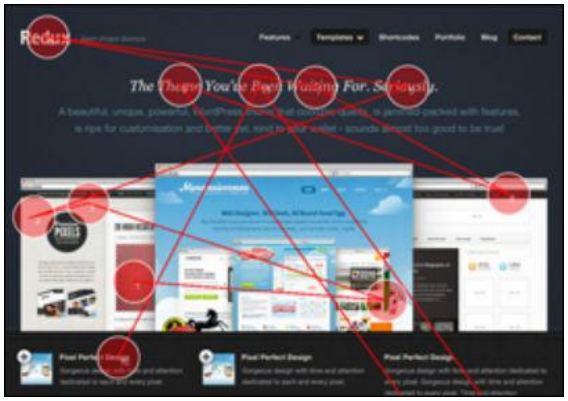

Figure 1 Example of fixation and saccade recording from a respondent

Fixation data from eye tracking experiments are processed using an algorithm stored in the software that directly connects to a PC and a video camera. In addition, a predictive model that can estimate the fixation points for a given picture can be developed. Such a model is useful in evaluating whether a picture is interesting or not. This is particularly critical when designing marketing instruments (Bang and Wojdynski, 2015).

After carrying out the literature review, the authors found that there are two methods that can be used as a predictive model, as required in this study. These are SVR (Support Vector Regression) and NN (Neural Network). Previous research has used the SVR method as the predictive model. In this research, NN is employed as the predictive model. This paper attempts to find whether NN is a more effective predictive model than SVR. This topic is selected because NN is expected to be able to directly handle non-linear relations by employing non-linear functions in the model.

2. METHODOLOGY

In order to study the cognitive aspect of the Human Visual System, one of the common approaches is to build a mathematical model that sufficiently represents the system (Simola et al., 2008; Zhong et al., 2014). Among the most appropriate is the NN model, which is a simplified brain model. It uses mathematical functions to transform inputs to outputs by mimicking brain processes. The NN is composed of many neurons that work together to perform the desired functions. A neuron outputs the result using a weighted sum function with a bias. One of the biases used is a sigmoid function.

NN algorithms can be set up using several specifications. There are various inter-layer transfer functions and training functions that can be used to train the model used (Bhardwaj et al., 2016), such as tansig (tangent-sigmoid), logsig (log-sigmoid), and traingdx (gradient descent backpropagation). Because there are many algorithms that can be used, it is necessary to conduct experiments to find the best combination of these functions, so that the predicted results could be of a good quality with the smallest possible error value. In order to achieve this objective, the aim of this paper is, through testing, to determine which combination will give more accurate results, which will involve the smallest number of iterations, and which will require the shortest duration of training.

This study requires four types of data, i.e.: (i) the coordinates of fixation data from eye tracking; (ii) the intensity; RGB, and orientation feature of each fixation point; (iii) the Euclidean distance (the distance of each point of fixation from the other fixation points of observed images); and (iv) the transition probability matrix of eye movement. In order to obtain the first

type of data, the authors conducted a study with 20 respondents (comprising 10 men and 10 women). The technical matters were as follows:

- The respondents of this experiment are Industrial Engineering students with eye defection rate ≤ (-)4.00, not using contact lenses and not having *strabismus*. It should be noted that eye deficiencies, such as those mentioned above, can cause the results to be unreliable because the device used in the experiments is not calibrated for deficient eyes.
- Each respondent will be shown 100 pictures, each with a 5-second duration. The fixation coordinates of the respondents will be the reference to extract the RGB values, and the Euclidean distance. These data will become the input of the NN model that will be tested.

After the coordinates of the fixation data were obtained, the remaining three types of data were obtained using the following formulae:

- Intensity and RGB data obtained from RGB value of each fixation point
  - $I = \frac{r+g+b}{3}$
  - $R = r - \frac{g+b}{2}$
  - $G = g - \frac{r+b}{2}$
  - $B = b - \frac{r+g}{2}$
- Orientation feature obtained from the Gabor Filter formula

  $$O = \exp\left(\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right)\cos(2\pi\frac{x'}{\lambda} + \emptyset)$$

  x’ = x cos θ + y sin θ

  y’ = − x sin θ + y cos θ

  θ : orientation

  σ : standard deviation

  γ : *spatial aspect ratio*

  λ : wavelength

  x : horizontal coordinate of fixation point of the picture

  y : vertical coordinate of fixation point of the picture
- Euclidean Distance

  $$D_{ij} = \sqrt{(i_n - j_n)^2 + (i_m - j_m)^2}$$

  n : horizontal coordinate of fixation point of the picture

  m : vertical coordinate of fixation point of the picture

  i : initial fixation point

  j : target fixation point
- Markov Chain Probability Matrix

  $$P_{ij} = \frac{1 + |N_j - N_i|}{|G| + |E|} \cdot e^{-\frac{(i_n - j_n)^2 + (i_m - j_m)^2}{2\sigma^2}}$$

N_j : number of fixation in j

G : number of states on a picture (equal to the number of fixation points)

E : the number of fixation points

n : horizontal coordinate of fixation points

m : vertical coordinate of fixation points

i : initial fixation point

j : target fixation point

σ : free parameter with a value range of 0.1–0.2 times from picture length

After the data were collected, the next step involved with normalizing the data. The normalization process changes the data range to −1 to +1 intervals. As a result, the input data consisted of 12 columns; the first 5 columns are features of fixation point A, the next 5 columns are features of fixation point B, the 11th column is the Euclidean distance (D_{ij}) from A-B point and the 12nd is the Markov Chain Probability Matrix (P_{ij}). Examples of the normalized input data can be seen in Table 1.

Table 1 Normalized input data

I	R	G	B	O	D_{ij}	P_{ij}
1	−1	−0.12791	−1	−1	−1	−0.99139
0.25582	−1	−0.86898	−1	−1	1	−0.99832
1	−0.98788	−0.98788	−1	−1	1	−0.99894
0.56482	−1	−0.97681	−1	−1	1	−0.99936
0.99960	−1	−0.91575	−1	−1	1	−0.99960

After the required input data was obtained, the training session was carried out with the aim of achieving the NN model. Subsequently, the test data set was utilized to determine the predicted probability Markov Chain matrix, such that these results could be compared with the actual results using several measurement errors (Chai and Draxler, 2014; de Myttenaere et al., 2015).

3. RESULTS AND DISCUSSION

After carrying out the training and testing sessions for the 27 combinations of NN functions, the results are as follows:

- *The smallest MAPE value*
  The combination that produces the smallest MAPE value is: logsig-tansig, trainscg with MAPE value = 13.71%. The results can be seen in Table 2 below.

Table 2 The results of NN function combination for the smallest MAPE value

Combination	MAPE Value (%)
logsig-tansig,trainscg	13.71
purelin-purelin,trainscg	14.25
purelin-purelin,trainbfg	14.71
tansig-tansig,trainscg	14.81
purelin-tansig,traingdx	14.84

The logsig-tansig, trainscg combination produces the smallest absolute value (actual result – testing result) = 0.0002–0.1648 for the data with the same (x,y) coordinates. Therefore, when its absolute value is divided by the actual result, it also produces the smallest MAPE value.

- *The smallest MSE value*
  The combination that produces the smallest MSE value is purelin-purelin, trainscg with MSE value = 0.952%. The results can be seen in Table 3 below.

Table 3 The results of NN function combination for the smallest MSE value

Combination	MSE Value (%)
purelin-purelin,trainscg	0.952
purelin-purelin,trainbfg	1.007
purelin-tansig,traingdx	1.023
logsig-tansig,trainscg	1.045
logsig-tansig,trainbfg	1.079

The combination produces the smallest (Xmax – Xmin) range data. Therefore, the $(X - \bar{X})$ value is also small. As such, it results in the smallest MSE value.

- *The smallest number of iterations*
  The combination with the smallest number of iterations is dominated by the training function trainbfg. More detailed results can be seen in Table 4 below.

Table 4 The results of NN function combination for the smallest number of iteration

Combination	Iteration
logsig-logsig,trainbfg	1
logsig-tansig,trainbfg	1
purelin-logsig,trainbfg	1
purelin-tansig,trainbfg	1
tansig-logsig,trainbfg	1

These results are in accordance with Sharma and Venugopalan (2014), who state that, of the three types of training function, trainbfg achieves the fastest optimum value gradient error and the smallest number of iterations. However, this function uses more computation efforts in each iteration than the other functions. Therefore, even though it produces the smallest number of iterations, trainbfg results in the largest ratio of training duration to number of iterations.

- *The shortest training duration*
  In relation to the previous point, trainbfg gives the fastest training duration result (192 seconds). More detailed results can be seen in Table 5 below.

Table 5 The results of NN function combination for the fastest training duration

Combination	Training Duration
purelin-purelin,trainbfg	192
purelin-tansig,trainbfg	200
purelin-tansig,trainscg	236

4. CONCLUSION

From several analyses that have been performed, there are several conclusions. In terms of the accuracy of the prediction model, as measured by the smallest error value, the best combination of functions is purelin-purelin, trainscg. The combination of these functions ranks first in MSE and second in MAPE calculations. In addition, in terms of the computing performance of the prediction model, as measured by the smallest number of iterations and the shortest training duration, the best combination of functions is purelin-purelin and trainbfg. Moreover, the trainscg training function produces a smaller range of MAPE values than traingdx or trainbfg. Lastly, the trainbfg training function involves a shorter training duration and a smaller number of iterations than traingdx or trainscg.

For future research, several future works can be proposed. Instead of viewing a picture, respondents may be more attracted to seeing human faces. Therefore, future work could use pictures that include a human face as the research object. In this situation, the testing would be more accurate if additional tools were employed, for example the Viola Jones face detector and the Felzenszwalb person detector.

In addition, the combination of the training and testing carried out in this research is limited only to a combination of inter-layer transfer and training functions. Therefore, more research should be performed to investigate different combinations of numbers of layers and network types.

5. ACKNOWLEDGEMENT

This study was financially supported by Hibah PITTA 2017 from the Directorate of Research and Community Engagement, Universitas Indonesia.

6. REFERENCES

Bang, H. and Wojdynski, B.W., 2015. Tracking Users' Visual Attention and Responses to Personalized Advertising based on Task Cognitive Demand. *Computers in Human Behavior*, Vol. 55, pp. 867–876.

Bhardwaj, A., Tiwari, A., Bhardwaj, H. and Bhardwaj, A. 2016. A Genetically Optimized Neural Network Model for Multi-class Classification. *Expert Systems With Applications*, Vol. 60, pp. 211–221

Chai, T., and Draxler, R.R. 2014. Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, Vol. 7, pp. 1247–1250.

de Myttenaere, A., Golden, B., Le Grand, B. and Rossi, F. 2015. Mean Absolute Percentage Error for Regression Models. *Neurocomputing*, Vol. 192, pp. 38–48

Drewes, H, 2010. *Eye Gaze Tracking for Human Computer Interaction.* PhD dissertation der Ludwig-Maximilians-Universität München.

Duchowski, A., 2007. *Eye Tracking Methodology – Theory and Practice*. London: Springer-Verlag.

Kenyon, R.V., 1985. A Soft Contact Lens Search Coil for Measuring Eye Movements. *Vision Research*, Vol. 25, No. 11 pp. 1629-1633.

Schall, A. and Bergstrom, J.R. 2014. *Eye Tracking in User Experience Design.* Amsterdam: Elsevier.

Sharma, B. and Venugopalan, K. 2014. Comparison of Neural Network Training Functions for Hematoma Classification in Brain CT Images. *IOSR Journal of Computer Engineering*, Vol. 16, No. 1, pp. 31–35.

Simola, J., Salojärvi, J. and Kojo, I, 2008. Using Hidden Markov Model to Uncover Processing States from Eye Movements in Information Search Tasks. *Cognitive Systems Research* Vol. 9, pp. 237–251.

Zhong, M., Xinbo, Z., Xiao-chun, Z., Wang, J. Z. and Wenhu, W. 2014. Markov Chain Based Computational Visual Attention Model that Learns from Eye Tracking Data. *Pattern Recognition Letters*, Vol. 49, pp. 1–10