Gaze Pattern and Reading Comprehension

Tan Vo, B. Sumudu U. Mendis, and Tom Gedeon

School of Computer Science, The Australian National University,
Acton, Canberra, ACT 0200, Australia
{tan.vo,sumudu.mendis,tom.gedeon}@anu.edu.au

Abstract. Does the way a person read influence the way they understand information or is it the other way around? In regard to reading of English text, just how much we can learn from a person's gaze pattern? It is known that while reading, we inadvertently form rational connections between pieces of information we pick up from the text. That reflects in certain disruptions in the norms of reading paradigm and that gives us clues to our interest level in reading activities.

In this paper, we validate the above statement and then propose a novel method of detecting the level of engagement in reading based on a person's gaze-pattern. We organised some experiments in reading tasks of over thirty participants and the experimental outputs are classified with Artificial Neural Networks with an approximately 80 percent accuracy. The design of this approach is simple and computationally feasible enough to be applied in a real-life system.

"Your eyes are the windows to your soul"

Keywords: Eye-gaze Pattern, Artificial Neural Network, SVM, Grid-based Clustering.

1 Introduction

Everyone is taught to read (in English) the same way: Read a line form left to right and then drop down to the next line once the end of the current line is reached. As a beginner, we followed this simple rule very closely but as we get more adept in reading English text, it is no longer the case. What we have found is that people develop their own personal behaviours when reading, that they probably do not notice they even do. One aim of this research is to characterise these behaviours and to identify and abstract significant model that can show how engaged a person is in the reading activity.

We conducted an experiment where we capture test participants gaze activities while they perform reading tasks. We analysed a set of features of those data: reading time, fixations time, differences in X and Y coordinates, etc.... to identify the key factors to indelicate user engagement level in reading. With the positive result obtained from that, we would like to introduce a simple but affective approach to measure a person's interest in the materials he is reading.

We use Artificial Neural Network (ANN) method to validate the effectiveness of the proposed approach. By combing our solution with a reasonably simple

K.W. Wong et al. (Eds.): ICONIP 2010, Part II, LNCS 6444, pp. 124–131, 2010.

ANN, we could introduce a very achievable real life system. This solution is also flexible in combining with other classifying techniques. We further strengthen our claim by achieving a comparably accurate results with Support Vector Machine (SVM).

2 Proposed Method

2.1 Background

Comprehending the meaning of words in sentences and paragraphs is a great (unnoticed) strain on a persons cognitive process. In order to comprehend text a person needs to be able to read quickly because a person can, generally, only keep seven pieces of information (± 2) in their short-term memory [6]. Any additional information is quickly lost and cannot be recalled. This general rule stands for many different kinds of information from the very simple (letters, or words) to the very complex (entire sentences or a word and 20 Literature Survey all its associated contexts). This allows a fluent reader to be able to "chunk" related information together so that they can get more words into their short term memories.

The above phenomenon results in certain disruptions in reading patterns. We believe that these stochastic behaviours are the keys to effectively quantify the reading engagement load level of a person. Previously in 2009, a study was organised in our research group to investigate using eye-tracking to analyse reading behaviour. Even still in the preliminary stages, it showed the potential in using machine learning approached to classify eye gaze patterns. The purpose of the research behind this paper is to consolidate the previous studies result and to propose a feasible model for classifying gaze-pattern with machine learning methods such as Neural Networks.

The method we are proposing here is for detecting user engagement in reading and is based on the aforementioned gaze features. We also introduce three design principles to make it a lightweight yet effective method for this purpose.

2.2 Effective Reduction of Data Resolution

The gaze-tracking equipment that we use provides us the gaze points in term of a series of X and Y coordinates. These coordinates identify the locations of the gaze points on the screen and have been used to calculate the horizontal and vertical movements. In previous experiments[2][3], fixation points have been produced by filtering those gaze points, resulting in a more interpretable form for later data processing.

By observation, we found that most of the movements of fixations, i.e. saccades are just small and subtle position changes caused by the fact that the eyes do not actually focus on one place. Those saccades are considered irrelevant for this purpose and we can afford to omit them in the pattern recognition stage, hence further reduce the sample size.

We proposed a simple but effective method by dividing the screen into smaller cells using a m-by-n grid. This effectively replaces change in positions of any two fixations with the difference in position of the cells that contain them. We refer to this as **cell movements**. In the cases when the fixation movements are contained within a cell, we consider it a no change in cell position. The benefit of this is it will be less computationally demanding to perform any processing/analysis because the number of data points have been greatly reduced. We can also adjust the resolution of the grid (m and n) for finer or coarser filtering.

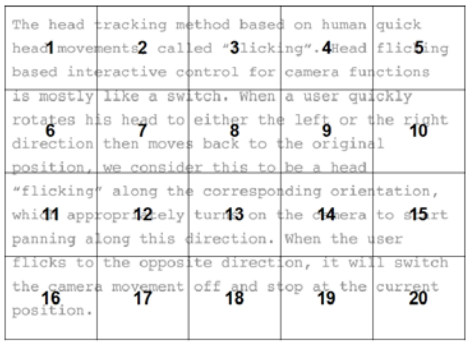


Fig. 1. A paragraph is divided into cells by a 4-by-5 grid. Each identified by a cell number

We examined the data sample from the experiment with and without using this data reducing method. Both yield comparable results except for computational speed.

2.3 Focus on Back Tracking and Forward Tracking

The most significant disruption in reading flow are the skipping forward and back-tracking activities found in the gaze. As participants try to "link" information up, they shift their eyes' focus back-and-forth to achieve a better understanding of the information.

Back tracking and forward tracking are two activities that we would like to qualify as the main factor to detect engagement in reading tasks. To quantify if a gaze movement is a back/forward tracking patterns, we consider if it belongs to the two "extreme" of cell movement groups. If we established a normal distribution of the distances of movement, the "extreme" groups are the one that did not fall within the 68-th confidence interval (a margin of one standard error). Figure 2 below demonstrates this idea:

The figure depicts the distribution of all cell movement distances of a person reading of one paragraph. It shows that if the distance of a cell movement is less than -5 ($\mu - 1\sigma$), that saccade is considered a backtracking. On the other hand, a forward tracking saccade is one that has the distance greater than 5($\mu + 1\sigma$). These thresholds are expected to be different on a case by case basis.

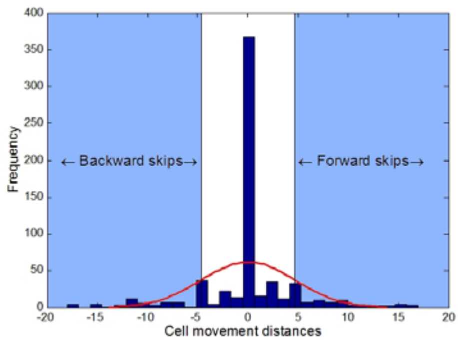


Fig. 2. A distribution of cell movement distances throughout reading activity of a paragraph

2.4 Levenberg-Marquardt Based Neural Network

Previously, an experiment carried out by Zhu et al.[2] to evaluate the performance of Levenberg-Marquardt optimization, combined with fuzzy signature, in classifying gaze patterns. What they found is that this optimization algorithm performs well with the gaze-pattern classification problem and on par with SVM in the two classes test.

In this paper, we evaluate the performance of Levenberg-Marquardt optimization as the training function in a Neural Network to classify eye gaze. The neural network we construct is a two-layer, feed-forward back-propagation that has one single output node. Hence the output value regarding to a pattern T is described as[4],[5],[7]:

$$y_1^T = g_O(b_1 + \sum_j W_{1j} \cdot g_H(b_j + \sum_k w_{jk} \cdot x_k^T)), \tag{1}$$

- b_1, b_j: the bias
- w_{1j} is the weight of the jth hidden neuron to the single output neuron
- w_{jk} is the weight of kth input neuron to the jth hidden neuron
- x_k^T the kth element of the input pattern T
- g_O transfer function on the output layer - linear transfer function
- g_H transfer function on the hidden layers - sigmoid transfer function

We evaluate the training performance of the network with this error function (mean square error):

$$E = \frac{1}{N}\sum_{k=1}^{N}(y_E - y_P)^2, \tag{2}$$

where y_E is the vector of predict outcomes and y_P represents the vector of predicted outcome.

The back-propagation training algorithm, being Levenberg-Marquardt optimization, will be represented by the formula[5]:

$$\delta w = (J^T J + I \cdot \mu)^{-1} J^T e \tag{3}$$

where J is the Jacobian matrix of the error function calculated in equation(2), μ is the learning rate which is updated after iteration. *diag* being the diagonal of $J^T J$.

3 Experiment

3.1 Background

We have selected 35 participants for this experiment. The experiment involves the participant reading some paragraphs from a computer screen while the computer gathers their eyes'(gaze) movements with gaze-tracking equipment.

In total there were ten paragraphs for the participants to read. Seven of the paragraphs were taken from the paper "Keyboard before Head Tracking Depresses User Success in Remote Camera Control" by Zhu et al.[1]. The remaining three paragraphs were extracts from various sources (miscellaneous paragraphs). Five of the paragraphs from the paper were chosen for the amount of useful information that was contained within. The other two paragraphs from the paper and the three miscellaneous paragraphs describing that paper were chosen because of their generality and lack of specific technical information, the paragraphs being introductory in nature. That is, care was taken to make sure that this fact was not obvious.

3.2 Setup

Within the 35 volunteered participants, we divided them into two groups. Group A were people that had been informed that they would have to answer questions about the paragraphs they read toward the end of the experiment session. Group B, however, were allowed to read as if they were just reading any piece of text - and that they would not be questioned at the end. Group A contained 13 people while group B had 22 participants.

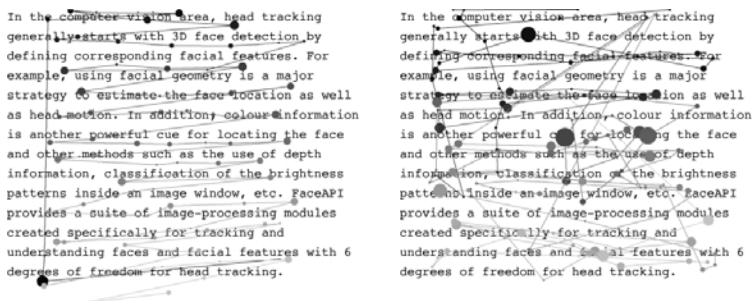
The paragraphs were presented to participants in different orders to prevent any specific paragraph ordering from affecting the results. This was an experiment design choice to help show which participants could look at the bigger picture even when the information is out of sequence and scattered.

The screen used was a 19 inch LCD and was set to a resolution of 1280 by 1024. All the paragraphs are displayed in full-screen. To stabilise the head position, we use a chin rest, which positions the participant faces about 72 cm away from the centre of the screen.

3.3 Data Collection and Preparation

The gaze points are collected at about 1/60th of a second rate, and we produce fixation points from them. We "group" gaze points into clusters with size of 15 pixels radius[3] to define fixations. The fixation length (ms) is worked out by the number of gaze points within each cluster.

Below are a visualisation of gaze data being projected onto their correspondent paragraphs. The solid circles represents the fixation points. The shade of the circles indicates the fixation length - with the darker one indicate a longer fixation point. The lines that connect the fixations points represent the saccades.



(a)

(b)

Fig. 3. An example of 2 read patterns of the same paragraph by two different participant

We further filter out the data using the aforementioned grid-based method with 4-by-5 grid. Based on that we calculate numbers of back-tracking and forward-tracking of each paragraph. For the classification task, we use these three following feature to be evaluated with ANN:

- Average fixation length for each paragraph and each participant
- Back-tracking count for each paragraph and each participant
- Forward-tracking count for each paragraph and each participant

4 Evaluation and Comparison

4.1 Neural Network Results

A two-layer neural network with one output neuron is use for classifying data. The transfer function of the output layer is a linear transfer function while the hidden layer is equipped with a tangent sigmoid transfer function. The hidden layer comprise of 5 neurons. We designed this to be a binary classification problem where the target values were 1 for *relevant paragraph* and 0 for *irrelevant paragraph*. The neural network was back propagation trained with Levenberg-Marquardt optimization. The LM parameters are configured with an initial μ value of 0.001 with the increase and decrease factors as 10 and 0.1 respectively. The network performance is evaluated by Mean of Square Error method.

We performed 9-folds cross-validation and obtained the average of classification accuracies from every fold. Due to the relatively small sample size available

(35 participants - with 10 paragraphs per participant), 9-fold cross-validation is preferred to the conventional 10-fold method. For each fold, the training set is divided as followed: 60% for data training, 20% was used to generalise the network and prevent over-fitting and the last 20% was used as the test data.

Table 1. ANN Results for Eye-gaze Feature Pattern Recognition

Experiment	Classification Error Rate	Sensitivity	Specificity
Group A	0.2586	0.7241	0.7586
Group B	0.1717	0.7879	0.8687
Both Groups	0.1975	0.7898	0.8153

With this ANN setup, We were able to achieve about 80% classification accuracy as seen in Table 1. This is encouraging because we only provided three gaze parameters as classification categories.

As we can see the classification performance achieved with Group A data is slightly lower than with group B data. The hypothesis behind that is that with group A, where participants were expected to answer questions about the paragraph, that lead to a more "careful" reading behaviour for all paragraphs. That would results to a less disruptive forward, backward movements in the reading patterns. But nevertheless, the classification results in all cases are positive considered the small number of classification features.

4.2 Support Vector Machine Comparison

Support Vector Machines (SVM) are well-established method for this type of classification problems[4].

We constructed a conventional Support Vector Machine with a linear kernel. We used the same dataset as we did with ANN. The labels we chose are "1" for *relevant paragraph* and "0" for *irrelevant paragraph* . To have a fair comparisons, we also cross-validated the results using 9-folds and obtains the average Classification Error Rate (CER) after 9 iteration.

Table 2. SVM Results for Eye-gaze Feature Pattern Recognition

Experiment	SVM Classification Error Rate(CER)	ANN CER	SVM Sensitivity	SVM Specificity
Group A	0.2538	0.2586	0.7077	0.7846
Group B	0.1773	0.1717	0.8091	0.8364
Both Groups	0.2086	0.1975	0.7600	0.8229

Table 2 compares the results we got by using SVM technique with the previous results by ANN. We found that ANN and SVM performance in term of accuracy almost exactly match each other. Both methods (SVM and ANN) result in a very high accuracy rate and with further optimisations on both, we believe we can attain even more positive results.

5 Conclusion

In this paper, we demonstrated the effectiveness of ANN in recognising gaze-patterns. The findings are encouraging because ANN combined with our proposed method for data preprocessing has resulted in a low computational model that achieves highly accurate results: An ANN being trained with only three classification categories is able to achieve 80% accuracy is very encouraging. It has also consolidate the outcome of our previous experiment[3] as well as the use of Levenberg-Marquardt optimization as the training algorithm for this types of problem[2].

References

1. Zhu, D., Gedeon, T., Taylor, K.: Keyboard before Head Tracking Depresses User Success in Remote Camera Control. In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) INTERACT 2009. LNCS, vol. 5727, pp. 319–331. Springer, Heidelberg (2009)
2. Zhu, D., Mendis, B.S., Gedeon, T., Asthana, A., Goecke, R.: A Hybrid Fuzzy Approach for Human Eye Gaze Pattern Recognition. In: Köppen, M., Kasabov, N., Coghill, G. (eds.) ICONIP 2008. LNCS, vol. 5507, pp. 655–662. Springer, Heidelberg (2009)
3. Fahey, D.: A Preliminary Investigation into using Eye-Tracking to Analyse a Persons Reading Behaviour. B.S. thesis. The School of Computer Science Australian National University (2009)
4. Byvatov, E., Fechner, U., Sadowski, J., Schneider, G.: Comparison of Support Vector Machine and Artificial Neural Network Systems for Drug/Nondrug Classification. Journal of Chemical Information and Computer Sciences, 1882–1889 (2003)
5. Hagan, M.T., Menhaj, M.B.: Training feedforward networks with the Marquardt algorithm. IEEE Transactions on Neural Networks 5(6), 989–993 (1994), doi:10.1109/72.329697
6. Miller, G.A.: The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. In: The Psychology of Communication: Seven Essays. Penguin Books, Inc. (1970)
7. Mendis, B.S.U., Gedeon, T.D., Koczy, L.T.: Learning Generalized Weighted Relevance Aggregation Operators Using Levenberg-Marquardt Method. In: International Conference on Hybrid Intelligent Systems (2006)