ORIGINAL RESEARCH

Predicting reading comprehension scores from eye movements using artificial neural networks and fuzzy output error

Leana Copeland*, Tom Gedeon, Sumudu Mendis

Research School of Computer Science, Australian National University, Canberra, Australia

Received: June 29, 2014 **Accepted:** August 7, 2014 **Online Published:** August 25, 2014
DOI: 10.5430/air.v3n3p35 **URL:** http://dx.doi.org/10.5430/air.v3n3p35

Abstract

Predicting reading comprehension from eye gaze data is a difficult task. We investigate the use of artificial neural networks (ANNs) to predict reading comprehension scores from eye gaze collected from participants who read and completed an online tutorial in our lab. Problems such as large feature sets and small highly imbalanced data sets compound to make this task even more complex. We propose using fuzzy output error (FOE) as an alternative performance function to mean square error (MSE) for training feed-forward neural networks to overcome these problems. We show that the use of FOE as the performance function for training ANNs provides significantly better classification of eye movements to reading comprehension scores. ANNs with three hidden layers of neurons gave the best classification results especially when FOE is used as the performance function for training. In these cases we found up to 50% reduction in misclassification rates compared to using MSE. We found that ANNs give optimal classification results in comparison to other classification techniques. When FOE is used as the performance function for training the ANNs the misclassification rates are halved compared to the other techniques. Cluster analysis was performed on one of the more complex data sets. Interesting reading behaviour properties were found within the data set. The intended use of this research is in the design of adaptive online learning environments that use eye gaze to predict user comprehension from reading behavior.

Key Words: Fuzzy output error (FOE), Imbalanced data sets, Artificial neural networks, Performance functions, Eye tracking, Comprehension detection, Adaptive online learning environments

1 Introduction

Eye tracking is becoming more accurate, less obtrusive, and most importantly, more affordable. This gives rise to the possibility of using of eye tracking as a common input to computer systems. Eye tracking has been shown to be an effective way of analysing human behaviour, particularly during reading.[1] Eye gaze can be used to differentiate whether a person is reading or not[2] in addition to how they are reading.[3] Educational material is being offered through online media more frequently with the increased availability of computer technologies at affordable prices. This is especially true for tertiary education, where face-to-face education is now heavily supplemented with material that is available through online learning environments, such as Moodle and Blackboard. It has become common for universities to offer online/off-campus degrees where students never or rarely have face-to-face interaction with their instructors and most likely other students. This raises the question of how eye tracking can be used to make the learning process more effective especially when a teacher or instructor may have little contact with students. The need for additional forms of student monitoring are necessary to detect when a student is under or over-performing so that they can either be given remedial help or advanced material at an early stage prior to summative assessment.

We investigate methods for predicting reading comprehen-

*Correspondence: Leana Copeland; Email: leana.copeland@anu.edu.au; Address: School of Computer Science, Australian National University, Canberra 0200, Australia

sion from eye gaze using machine learning techniques. The eye gaze was recorded from participants while they read and completed an online tutorial and quiz. Little prior work has been done to predict reading comprehension via machine learning. To date prediction of reading comprehension has been made based on eye movement measures that have been derived from the eye gaze signal such as fixation duration[4] and regressions.[5] Current applications of eye tracking in reading analysis only take into account basic assessment of reading behaviour such as using fixation time to predict when a user pauses on a word.[6, 7] Instead, we look at combining eye movement measures to make more complex predictions about reading behaviour. The intended use of reading comprehension prediction from eye gaze is in the design of adaptive online learning environments that use eye gaze to predict comprehension from user reading behaviour. This application poses several obstacles namely restricted size in data sets that are also highly imbalanced. We explore several different methods of increasing prediction accuracy. Initially, we build on previous work by Copeland et al.[8] on improving the classification performance of artificial neural networks (ANNs) using fuzzy output error (FOE) as the performance function for back-propagation training the feed-forward ANNs. We extend this research by exploring different membership function shapes (FMFs) for calculating FOE and compare these results to using mean square error (MSE) as the performance measure for training. We assess whether the use of this performance measure is better suited to this type of problem compared to mean square error (MSE).

1.1 Eye movements and reading

Eye movements can be broadly characterized as fixations and saccades. A fixation is where the eye remains relatively still to take in visual information. A saccade is a rapid movement that transports the eye to another fixation. At the centre of the retina is a special part of the eye called the fovea that sees in fine detail. The foveal region of the eye is very small, being only about 0.2 mm in diameter. Around the point of fixation visual acuity extends about 2.[1]

When reading English fixation duration is generally around 200-300 milliseconds, with a range of 100-500 milliseconds and saccadic movement is between 1 and 15 characters with an average of 7-9 characters.[1] The majority of saccades are to transport the eye forward in the text when reading English; however, a proficient reader exhibits backward saccades to previously read words or lines about 10%-15% of the time.[1] Backward saccades are termed regressions. Short regressions can occur within words or a few words back and may be due to problems in processing the currently fixated word, overshoots in saccades, or oculomotor errors. However, longer regressions occur due to comprehension difficulties, as the reader tends to send their eyes back to the part of the text that caused the difficulty.[1]

Comprehension of the text can have significant effects on the eye movements observed.[1, 5] Studies have shown there are numerous variables that influence eye movements during reading: semantic relationships between words, anaphora and co-reference, lexical ambiguity, phonological ambiguity, discourse factors and stylistic conventions, and syntactic disambiguation. For example, garden-path sentences are syntactically ambiguous and induce regressions to resolve the comprehension problems.[9] Eye movements have also been shown to reflect text difficulty[5] and shown that they can be used to predict reading comprehension.[4] The consequence of these relationships is that eye movements can provide a moment-to-moment understanding of reading comprehension.[4]

One of the challenges when analysing reading behaviour is that there is a high level of variability seen between individuals.[1, 4] Personalisation of eye movement measures has been employed to improve prediction results.[10, 11]

1.2 Eye tracking in adaptive elearning

Eye gaze patterns can be used to detect the kind of task a person is performing[12, 13] such as if a person is reading or not.[2, 14] Eye movements can also be used to detect how a person is performing a task, for instance if they are reading or skimming.[3] Eye gaze patterns can be used to differentiate when individuals are reading different types of content.[15] In that application both support vector machines (SVMs) and ANNs were used to classify eye movement measures as being recording during reading text that was either relevant or irrelevant to answering a set of questions. ANN's have also been used to predict item difficulty in multiple-choice reading comprehension tests.[16] Their analysis took into account the text structure, propositional analysis of the text, and the cognitive demand of the text, but not eye gaze.

Eye gaze can be used to provide feedback about user behaviour such as in ref. 9 where eye gaze is recorded to give implicit perceived relevance of pieces of text in a document. There are several applications that are used in reading assistance. iDict is a reading aid designed to help readers of a foreign language.[6] iDict uses eye gaze to predict when a reader is having comprehension difficulties. If the user hesitates whilst reading a word then a translation of the word is provided along with a dictionary meaning. This is somewhat similar to The Reading Assistant,[7] which uses eye gaze to predict failure to recognize a word. The Reading Assistant then provides auditory pronunciation of the word to aid in reading.

1.3 Dealing with imbalanced data sets

Dealing with imbalanced data sets is not a new problem. Performance functions for dealing with imbalance in data sets include increasing the weight updating for the minority

class and decreasing it for the majority class.[17,18] This error function was designed specifically for use in the back-propagation algorithm for training feed-forward artificial neural networks. The error function was used to classify two data sets from the UCI machine-learning repository and showed improved classification results compared to conventional back-propagation training.

Many methods have been used to overcome the problem of imbalanced data sets such as using under-sampling, over-sampling, and other forms of sampling to reduce the imbalance. Cost sensitive methods have also been used to deal with imbalanced learning.[19] Cost sensitive learning centres on the fact that certain misclassifications may be more costly than others. In these cases it is appropriate to optimise a classifier so that it performs well identifying certain cases.[20] An example of a cost sensitive learning algorithm is MetaCost[20] that is based on relabelling of training data with their estimated minimal cost classes. Another way of achieving cost sensitivity is to change the algorithm used by the classifier to utilise the cost matrix, such as with neural networks.[19,21]

Whilst binary cost sensitive classification is relatively simple, multiclass cost sensitive classification is a much less trivial matter.[22] A combination of cost sensitive learning and different sampling methods have been used to train neural networks on imbalanced data sets.[23] This analysis highlights that although these techniques can be used for learning two-class problems they may not be effective in multi-class problems with imbalance.

1.4 Fuzzy Output Error (FOE)

Fuzzy Output Error (FOE)[24] is an extension of FYCLE and SYCLE.[25] FOE uses a fuzzy membership function to quantify the difference between the predicted and the target values, i.e. the error, rather than assign the difference a value of 0, 0.5 or 1 as is done in FYCLE. As opposed to MSE, FOE describes the error in a fuzzy way and then sums the fuzzy errors together to get the total error.

FOE is defined as follows for a data set of n records with matching pairs of target and predicted values for each record 1 to n. See Equation 1.

$$FOE = \sum_{i=1}^{n} 1 - \mu(\hat{y}_i - y_i), where \quad n \in N \tag{1}$$

Where $\mu()$ is the membership function of a desired classification and its complement describes the error. The membership function is termed the FOE Membership Function (FMF). The FMF is used to describe the output of a fuzzy classification (or a regression) in regards to how close that output is to the target output. The membership function itself represents the fuzzy set for "good classification". The value of $\mu(x)$ gives the degree of membership of the error in the good classification fuzzy set and consequently the complement of $\mu(x)$ gives the error measure. Therefore, $\mu(\hat{y} - y) = 1$ and hence there is no error when there is perfect classification. The more $\mu(x)$ tends toward 0 the higher the error since the difference is larger. The FMF shapes used in this analysis will be trapezoidal or triangular membership functions. FMF's can be created in any shape in order to describe the output of a function.

It is important to note that the difference between target and predicted values is not taken as the absolute value of the difference (i.e. $|\hat{y} - y|$). Although this would make the FMF simpler because it would only need one side of a piecewise linear function, it provides more flexibility in describing the types of error. For example, false negatives may be considered a much worse error than false positives when screening for diseases.

Approximation of fuzzy membership functions using squashing functions

There are many different ways to construct membership functions as described in Ref.25, however, commonly piecewise linear functions are used as they are ease to handle.[27] The problem with these functions is that optimisation of parameters via gradient-based methods become complicated, as they do not have continuous derivatives. One of the solutions to this problem is to approximate piecewise linear functions using combinations of sigmoid functions called a squashing function.[27–29]

A sigmoid function is an s-shaped function that is commonly used as an activation function of artificial neurons, as well as in economic and biological models. The definition of a sigmoid function is shown in Equation 2.

$$\sigma_{\alpha}^{\beta}(x) = 1/1 + e^{-\beta(x-\alpha)} \tag{2}$$

The parameter β controls the steepness of the sigmoid curve, that is, varies the function from a shape either close to linear or more like a step function. The parameter α controls where the centre of the curve, $\sigma(x) = 0.5$, is on the horizontal axis. More precisely, $x - \alpha$ will move the centre to α and $x + \alpha$ will move the centre to $-\alpha$. These two parameters play an important role in how the sigmoid function will be shaped to approximate the piecewise linear membership functions.

To approximate one half of a trapezoidal or triangular function, we integrate the difference between two sigmoid functions on an interval [a, b].[27,28] The definition of the squashing function on interval [a, b] is shown in Equation 3.

$$S_{\alpha,\delta}^{\beta} = 1/2\delta \ln(\sigma_{\alpha+\delta}^{-\beta}(x)/\sigma_{\alpha-x\delta}^{-\beta}(x))^{1/\beta} \tag{3}$$

Where α gives the centre of the squashing function and δ gives the steepness of the squashing function. The parameter δ is referred to as the fuzziness parameter and β the approximation parameter. The larger β is the closer the approximation to the trapezoidal function being modelled.

A piecewise linear membership function can therefore be approximated with the combination of two squashing functions using the conjunction operator. The following equation defines the approximation of a trapezoidal membership function.[27,28]

$$S^{(\beta)}_{1/2,1/2}(S^{(\beta)}_{a_1,d_1}(x) + S^{(-\beta)}_{a_2,d_2}(x) - 1) \quad (4)$$

When $a_1 = d_1 = -1/2$ and $a_2 = d_2 = -1/2$ the squashing function approximates a triangular membership function. All FMF shapes are represented in this form throughout the analysis so that gradient descent methods can be used to optimise the error function.

2 Method

2.1 Design

A user study was conducted to collect participants' eye gaze as they read a tutorial and completed a quiz based on the tutorial's content. The tutorial and quiz were coursework from a first year computer science course taken at the Australian National University (ANU). We applied for, and were granted, Ethics Approval for our experiments using the ANU's processes for ethical conduct of research involving humans. The tutorial and quiz was presented to participants in four formats to observe the effect that presentation has on learning and reading behaviour. These presentation formats are described as follows:

Format A: The tutorial content slide Figure 1(a) is first shown to participants followed by questions and the content slide Figure 1(b). The content part of the second slide is identical in all cases to the content in the first slide, being a simple repetition of the material with the same formatting and so on. An example of this sequence of presentation of material is shown in Figure 1. Since there are 9 topics there are 18 slides in total displayed in the study.

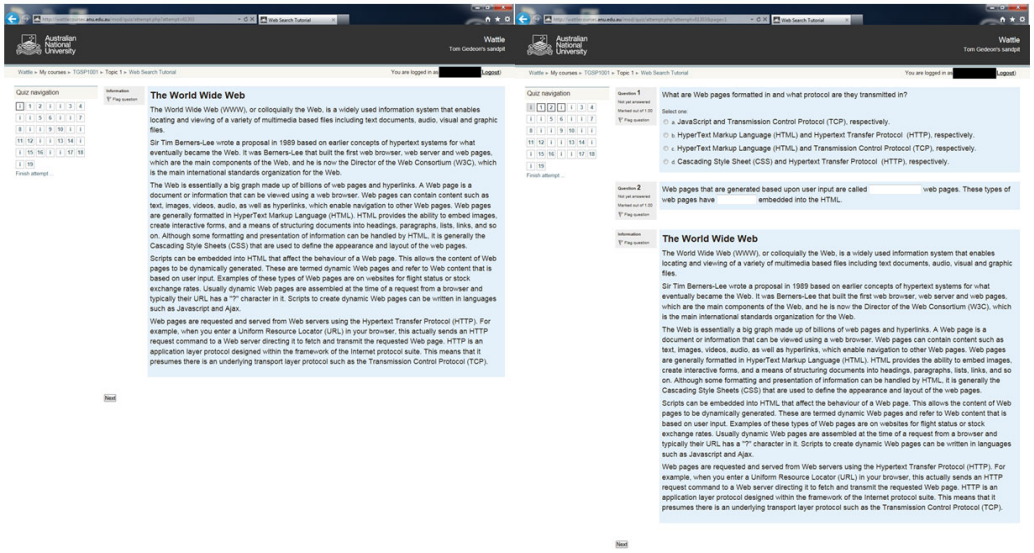

Figure 1: Example of Presentation Variation 1 (a) the tutorial content slide (b) the questions plus the tutorial content slide

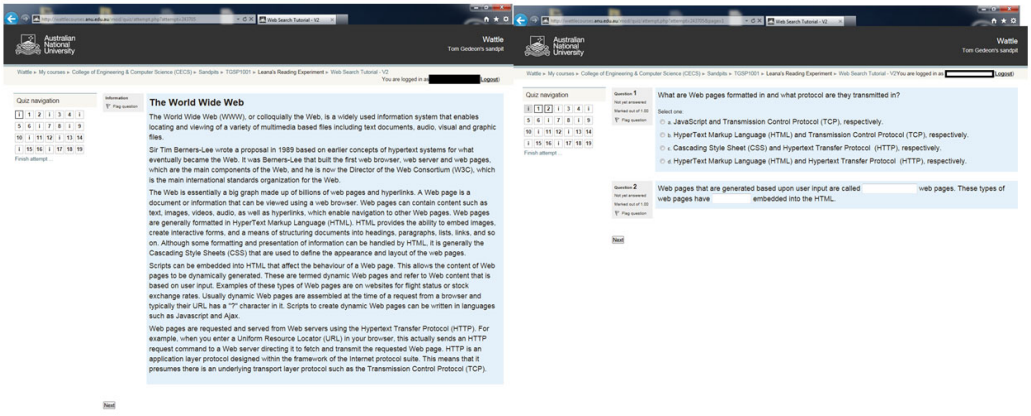

Figure 2: Example of Presentation Variation 3 (a) the tutorial content slide (b) the questions only slide

Format B: The questions and content slide is shown to participants immediately. An example of this is seen in Figure 1(b). Since there are 9 topics there are 9 slides in total displayed in the study.
Format C: The content slide is shown to participants followed by a screen with the questions but with no repeat of the content again Figure 2(b). An example of this sequence of presentation of material is shown in Figure 2. Since there are 9 topics there are 18 slides in total displayed in the study. This format can be considered as a control presentation method.
Format D: The last presentation consisted of displaying a slide with only the questions on it, as seen in Figure 2 (b), followed by the content slide Figure 2(a), and then again presenting them with the questions. Since there are 9 topics there are 27 slides in total displayed in the study. The reasoning behind this format is that we wanted to mimic a situation where the participants knew what the comprehension questions were but had no access to them as they read. The hypothesis is that participants will read the text differently to format A(a) and format C.
Each slide's content section is 400 words long with an average Flesch Kincaid Grade Level of 12. The tutorial topic is Web Search and each slide covered a sub-topic. All participants were university students and therefore had at least high school level education indicating that the readability of the slides should not be above their reading abilities. The tutorial content was accessible via the Wattle online learning environment at ANU (a variant of Moodle). Participants answered two questions at a time to measure their comprehension (18 questions in total); one question in the set is multiple-choice and the other is cloze (fill-in-the-blanks). The two types of questions are to assess different forms of comprehension.[30] The scores that the participants can receive for each question are 0, 0.5 and 1, corresponding to incorrect, half correct and correct respectively.
Once the participant finished the quiz and before being shown their result, participants were asked to subjectively rate their overall comprehension on a scale of 1 to 10 with 10 being complete understanding.

2.2 Demographics

There is different demographic data for each of the studies. The study that used format A was the initial study that can be broken into two demographics (COMP1710 students and others). The follow-up study that consisted of testing the three remaining presentation formats only consists of one demographic, being COMP1710 students. The choice of participants is based on the target user group of the eventual online learning environment, which are university students.
Format A: For the first presentation method 15 (6 female, 9 male) participants aged between 17 and 31, with an average age of 22.3 years old took part in the study. Of the participants, 9 were enrolled in COMP1710, however only 7 participants stated that their degree or major was related to computer science, information technology or software development. The remaining 6 participants were doing degrees such as BSc, BEng, and BA's, and were recruits not from the Computer Science department. No participant stated that they had reading problems such as dyslexia, and 4 of the participants stated that English was not their first language.
Format B: There were 8 participants in this group all of whom were COMP1710 students. All but 1 of the participants was male and had an average age of 21.8 years (standard deviation 7.9 years), age range 18-41 years. All participants had a major related to computer science and where enrolled in either BIT (Bachelor of Information Technology) or BSc. English was not the first language for 3 of the participants.
Format C: There were 9 participants in this group all of whom were COMP1710 students. All but 2 of the participants were male and had an average age of 22.8 years old (standard deviation 6.4 years), age range 18-37 years. All but 1 of the participants had a degree or major related to computer science. English was not their first language of 5 of the participants.
Format D: There were 7 participants in this group all of whom were COMP1710 students. All but 1 of the participants was male and had an average age of 20.1 years (standard deviation 2.8 years), age range 17-24 years. Of the participants, 5 had a major related to computer science and where enrolled in either BIT or BSc. English was not their first language of 3 of the participants.

2.3 Data collection method

The study was displayed on a 1280×1024 pixel Dell monitor. Eye gaze data was recorded at 60Hz using Seeing Machines FaceLAB 5 infrared cameras mounted at the base of the monitor. This eye tracker has a gaze direction accuracy of 0.5-1 rotational error and measures pupil diameter as well as blink events. EyeWorks was the software used to collect the data. The study involved a 9-point calibration sequence for each participant, which takes about 5 minutes at the start of a session, and is done only once for each participant.
As the data recorded is a series of gaze points, EyeWorks Analyze was used to pre-process the data to give fixation points. The parameters used for this were a minimum duration of 0.06 seconds and a threshold of 5 pixels.

2.4 FMF shapes used to calculate FOE

In this analysis we utilise 7 FMF shapes. The first FMF shape presented, from this point called as FMF1 (see Figure 3(a)), is designed to be a cross between FYCLE and the shape of an MSE curve. The difference between the predicted value and target value is within ±0.2 so is not considered an error and therefore considered correct classifi-

cation. Progressing out from ±0.2, the difference between predicted and target value is considered to be more in the error set in a linear fashion. FMF2 (Figure 3(b)), is designed to be a model of FYCLE. FMF3 (Figure 3(c)) is a triangular membership function that is designed to resemble the shape of an MSE curve. Here the difference between target and predicted values is given a lower value for membership in the good classification set the further the different progresses to -1 or 1, the extremities.

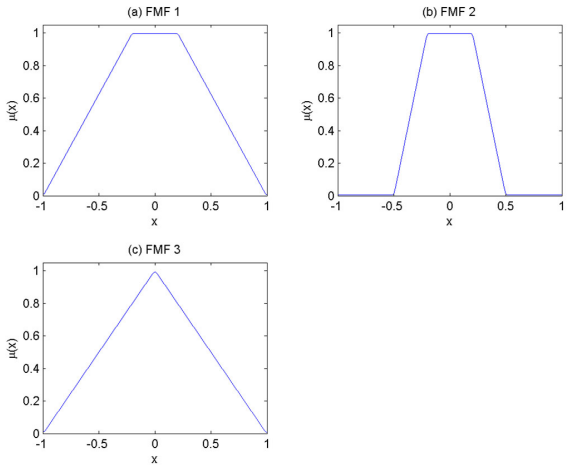

Figure 3: Plots of : (a) FMF1; (b) FMF2; and (c) FMF3

FMF4 (Figure 4(a)) and FMF5 (Figure 4 (b)) are asymmetrical FMFs that are inverses of each other. They are both a combination of half of FMF1 with the opposite half of FMF2, and were trialled to investigate the effect of asymmetric FMFs, which may have benefit in some application, and is not possible using MSE.

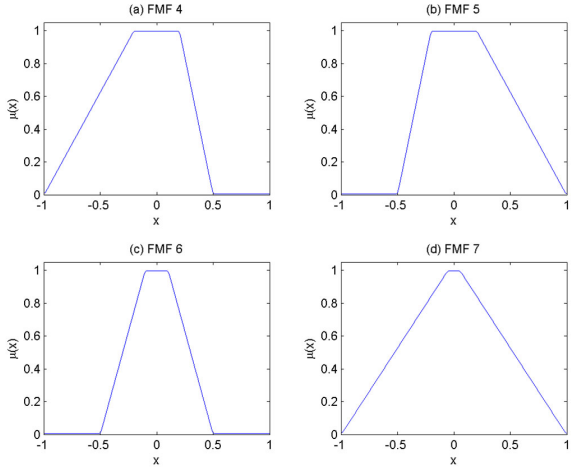

Figure 4: Plots of (a) FMF4; (b) FMF5; (c) FMF6; and (d) FMF7

The shape of FMF6 (Figure 4(c)) is a variant of the FYCLE approximation FMF2. It has a smaller region that defines the difference between the predicted and target values as being completely in the good classification set, i.e. $\mu(\hat{y} - y) = 1$. This region is when the difference is between ±0.1 instead of ±0.2. Again, this is to make the error output closer to zero as described above. FMF7 (Figure 4(d)) is a variation of FMF1 but is also a combination of FMF1 and FMF3. Again the variation is that there is a smaller region that defines the difference between the predicted and target values as being completely in the good classification set, i.e. $\mu(\hat{y} - y) = 1$. This region is when the difference is between ±0.05 instead of ±0.2.

2.5 Description of data sets

The raw eye gaze data consists of x,y-coordinates recorded at equal time samples (60Hz). Fixation and saccade identification was performed on the eye gaze data. From this point many other eye movement measures are derived. The measures used in this analysis are:

Number of fixations: The sum of fixations recorded for each tutorial page. The number of fixations can be affected by the reading behaviour, text difficulty, and reading skill.[1]

Maximum fixation duration (seconds): The maximum duration of the longest fixation recorded for a tutorial page. The length of a fixation can be an indication of difficulties in processing particular words or due to comprehension difficulties.[1]

Average fixation duration (seconds): The sum of the durations of all fixations on a paragraph divided by the number of fixations on that paragraph.

Total fixation duration (seconds): The sum of all fixations on complete text.

Number of regressions and regression ratio: The number of regressions, and that divided by the total number of saccades on a paragraph. There is evidence that when reading harder text more regressions are observed.[5]

Average forward saccade length (pixels): The average length of the left to right saccades. Saccade length is known to be affected by characteristics of the text.[1]

Reading analysis: Using a reading detection algorithm[3] the percentage of fixation transitions classified as being part of reading (read ratio), skimming (skim ratio), and scanning/searching (scan ratio) are found.

Regional Analysis: the fixation-to-word and duration-to-word ratios measured for the paragraphs where the answers are located. These are measures for how long the participant spent in the area containing the answer to the question. The hypothesis is that there is a relationship between the proportion of attention participants give to the answer paragraphs and the answers they provide.

Answer-seeking behaviour: as described in Ref. 30 this is the behaviour of jumping to and fro between the questions and the text to find the answers. This includes the fixation transitions classified as reading whilst jumping in between areas.

The number of inputs varies depending on the presentation method as the inputs are generated from the pages that the

participant viewed. This means that for format A, as the participants view the tutorial content page and then the questions and content page, the inputs are generated from both pages for the scores obtained from the questions and content page. Since there is a large difference in the ranges for each of the inputs they are normalized to a range of [0,1].

Table 1: Properties of each data set

Properties of data set	Format A	Format B	Format C	Format D
Total Number of Inputs	36	20	28	40
Size	135	72	81	63
Multiple choice score class imbalance	109/26	59/13	61/20	56/7
Class Split %: 1/0	*81%/19%*	*82%/18%*	*75%/25%*	*89%/11%*
Cloze Score class imbalance	124/11/0	69/1/2	35/22/24	49/0/14
Class Split %: 1/0.5/0	*92%/8%/0%*	*96%/1%/3%*	*43%/27%/30%*	*78%/22%*

The two outputs for all data sets are the multiple-choice question score and cloze question score. The outputs are assigned based on the answers to the multiple-choice and cloze questions. That is, the multiple-choice score can take values of 0 or 1, corresponding to incorrectly or correctly answered questions. Similarly for the cloze question except that in this case half marks can be achieved so the output that is assigned can take the values 0, 0.5 or 1. This is therefore a classification problem, a binary classification task for the multiple-choice score and a 3-class classification task for the cloze score. However, as shown in Table 1 the ratio of the number of data instances in each class for each problem in considerably imbalanced for each output.

3 Results

This section details the results from the analysis of methods used to predict reading comprehension from eye movement measures. This section begins with a comparison of results from ANNs using different performance functions. We then compare the results from the ANNs to other classification techniques. In these scenarios we integrate the scores from the two questions to give one prediction output. Finally, we cluster the eye movement measures from format C to look for natural clusters in the data and make conclusions about the nature of these clusters. All analyses are performed in Matlab R2013a.

3.1 FOE analysis with all eye movements measures

In previous work we considered the use of FOE as the performance function for training of ANNs.[8] Only one FMF shape was used in that investigation, namely FMF2, which is designed to be a model of Fuzzy Classification Error (FYCLE).[25] We extend this investigation to look at the use of 6 other FMF shapes to calculate FOE. Due to space limitations we only report the average misclassification rates (MCR) and standard deviations for the best FMF shape for calculating FOE for the particular data set.

A set of ANNs was trained using the scaled conjugate gradient algorithm[32] with the performance function is set to be either FOE or MSE. The analysis is performed Matlab R2013a using the Neural Network toolbox. FOE was implemented as a custom performance function. The default training method is the Levenberg-Marquardt algorithm[33] however this training method will not accept custom performance functions. The scaled conjugate gradient algorithm has been shown to perform faster than other methods available.[32] Furthermore, as seen in later parts of the analysis the classification outcomes from using the Levenberg-Marquardt algorithm was used, with MSE as the performance function, were poorer than those obtained when the scaled conjugate gradient algorithm was used with MSE as the performance function. For these reasons we use the scaled conjugate gradient algorithm in our analysis.

The number of inputs for each presentation format is outlined in Table 1 and all networks have 2 outputs. From initial testing it was found that a single layer network performed poorly for all both FOE and MSE. We have chosen two and three layer topologies for the analysis. The following topologies were tested: [10 5], [20 10], [30 15], [12 6 3], [16 8 4], [20 10 5], and [30 20 10]. The notation [X Y Z] indicates neurons in the first hidden layer to the third hidden layer. As a baseline comparison MSE is used as one of the performance functions. Reported are the average misclassification rate (MCR) values from 10-fold cross validation with standard deviations. The topologies that generate the best predictions are [12 6 3] and [16 8 4] for all formats and for both FOE and MSE. Due to space limitations we restrict our presentation of these results to report only average results for all topologies and the two optimal topologies. These results are shown in Table 2.

The topologies that generate the optimal predictions are [12 6 3] and [16 8 4] for all formats and for both FOE and MSE. This confirms the fact that the data set are hard to classify and contain complex relationships, as three layers of hidden neurons are needed to provide decent classification results.

For all formats, on average the MCR produced from using FOE as the performance function for training ANNs to pre-

dict the question scores is lower than that from using MSE as the performance function. These results are an improvement on the results from previous work where FMF2 was used to calculate FOE.[8] In that work, we found that on average the MCR values for formats A and B were 0.28 with an average reduction of MCR of 9% and 21% respectively. However, in that analysis we used Levenberg-Marquardt training for the ANN's when MSE was used as the performance function and we used a different set of eye movement measures. In this analysis we used scaled conjugate gradient descent training for all performance functions to keep for consistency.

Table 2: Misclassification rate (MCR) comparison: FOE versus MSE as the performance function for ANN training

		FOE		MSE	Difference in MCR	% Reduction in MCR when FOE is used
Format		**FMF**	**Mean**	**Mean**		
A	[12 6 3]	5	0.16±0.06	0.20±0.09	0.04	19
	[16 8 4]		0.14±0.05	0.26±0.12	0.11	44
	Average		*0.21±0.09*	*0.24±0.09*	*0.03*	*13.22*
B	[12 6 3]	2	0.11±0.05	0.19±0.14	0.08	42
	[16 8 4]		0.14±0.06	0.14±0.08	0	1
	Average		*0.22±0.12*	*0.23±0.13*	*0.01*	*4*
C	[12 6 3]	6	0.61±0.17	0.67±0.15	0.06	9
	[16 8 4]		0.51±0.12	0.59±0.11	0.07	13
	Average		*0.59±0.15*	*0.66±0.13*	*0.07*	*11*
D	[12 6 3]	7	0.21±0.07	0.41±0.30	0.21	50
	[16 8 4]		0.27±0.14	0.32±0.15	0.06	18
	Average		*0.35±0.15*	*0.40±0.18*	*0.05*	*13*

Table 3: Average Misclassification (MCR) results for decomposition of eye movement measure based on the page they were recorded for

Format	Eye Movement measure from individual pages	FOE		MSE	Difference in MCR	% Reduction in MCR
		FMF	**Mean**	**Mean**		
A	Text page	3	0.22±0.08	0.25±0.09	0.03	11
	Questions and text page	3	0.21±0.11	0.23±0.09	0.02	7
B	Questions and text page	2	*0.22±0.12*	*0.23±0.13*	*0.01*	*4*
C	Text page	2	0.58±0.14	0.64±0.13	0.05	8
	Questions page	2	0.59±0.14	0.68±0.13	0.08	12
D	Questions page (before text)	7	0.35±0.15	0.39±0.15	0.04	11
	Text page	4	0.35±0.18	0.40±0.13	0.04	11
	Questions page (after text)	2	0.34±0.15	0.36±0.14	0.01	5

We believed that Levenberg-Marquardt training would be more appropriate for training the ANNs when MSE was used as the performance function. This proved to be an incorrect assumption on this data. Although when using scaled conjugate gradient there has been an improvement in results when FOE is used there has also been an improvement in MCR results for when MSE is used. Nevertheless, FOE still provides better classification results in this context. In particular, for format A when the [16 8 4] topology is used the highest average classification results that can be achieved is 86% (MCR=0.14). In this case FOE is used as the performance function and this is a 44% reduction in MCR compared to when MSE is used. Similarly, for format B when the [12 6 3] topology is used the highest average classification results that can be achieved is 89% (MCR=0.11). Again, this is when FOE is used as the performance function and is a 42% reduction in MCR compared to when MSE is used.

3.2 Consideration of eye movement measures from page type

For all formats other than format B there are multiple tutorial pages presented to the participant for each set of com-

prehension questions. For example, for format A the tutorial text is shown to the participant and then the quiz questions and tutorial text are presented to the participant. We break down the eye movement measures observed for each of these pages and input them in the ANNs to see if this reduction in inputs can improve classification results.

Scaled conjugate gradient descent is used to train the ANNs and the performance functions are set to be FOE or MSE. The topologies considered in this analysis are the same as in section 3.1. 10-fold cross validation is used for each of the topologies, and only average MCR values are reported for each of the error measures. For each of the other formats the average MCR and standard deviations for each of the pages is presented in Table 3. Given that format B includes only one page of tutorial content per set of questions the results presented in Table 3 are the same as in Table 2 but are reported as a comparison.

The results show that there is no improvement of MCR when the eye movement measures are considered separately depending on which page from which they were recorded. However, it is observed that when FOE is used as the performance measure when training the ANNs, lower MCR values are obtained.

Interestingly, the results for each of the page types are consistent within the formats. This demonstrates that only the measures from one of the pages are needed as inputs for the ANN to achieve the same results as with all pages. It also demonstrates that reading behaviour from each of the pages shown within each format can be used to determine reading comprehension measures.

3.3 Comparison of different classifiers

We look at other classification techniques for predicting reading comprehension from eye gaze. The purpose of this is to investigate whether ANNs are an appropriate classification technique for redacting the reading comprehension measures based on eye gaze.

In the above scenarios we considered the multiple choice and the cloze questions as separate outputs. However, in this part of the analysis we combine (add) the scores so that there is only one output. The outputs possible (i.e. the classes) are now 0, 0.5, 1, 1.5 and 2. The problem is not a 5-class prediction problem. Three supervised learning techniques are used to compare prediction results to using ANNs. These techniques are classification trees, k-nearest neighbour, and discriminant analysis. The results from this analysis are summarized in Table 4.

The results from these classification techniques are suboptimal compared to using ANNs. As can be seen in Table 4 when the ANN is used the MCR values are halved for formats A, B and D compared to using any of the other techniques.

From these results it can once again be observed that formats C and D are difficult to classify as we found above. The eye movements generated from format C are the hardest to classify by far. For this reason we move on to explore these data sets further using clustering.

Table 4: Comparison of Misclassification (MCR) results for predicting combined multiple-choice and cloze scores for all eye movement measures

Format	Classification Tree	k-Nearest Neighbour	Discriminant analysis	Best ANN Result*
A	0.33	0.37	0.33	0.14
B	0.21	0.29	0.38	0.11
C	0.64	0.65	0.73	0.51
D	0.42	0.38	0.40	0.21

3.4 Cluster analysis

Making predictions on the eye gaze data collected for formats C has proven to be quite challenging. Exploration of this data set using clustering is performed to see if there are any natural clusters in the data that we can make conclusions from and apply the classification techniques to. The clusters are compared to see if they are statistically different. In this analysis unpaired two-sided t-tests are used to assess the statistical difference.

The Statistics Toolbox in Matlab R2013a is used to perform the cluster analysis. We employed agglomerative hierarchical clustering which starts with every observation in its own cluster and then merges the groups together until they are in the same group.[34] Using agglomerative hierarchical clustering with distance measure set to correlation and linkage set as average, the following hierarchical clustering is obtained Figure 5. The eye movement measures from both the text page and the questions page were used in the clustering.

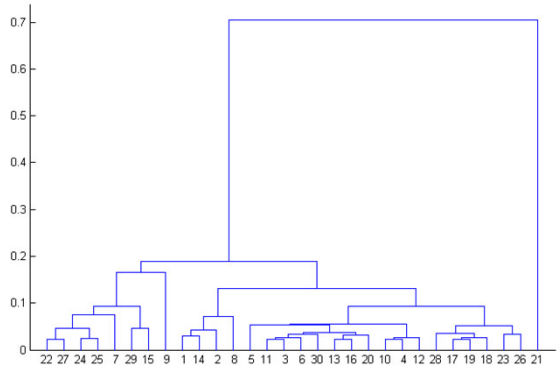

Figure 5: Hierarchical clustering for eye movement measures from format C

From the clustering there is evidently an outlying point. The outlying point had only 59 fixations recorded for reading the tutorial page and a total fixation time of 9.1 seconds. This is well below what is expected given that the text contains 400 words. Additionally, the participant answered both the multiple choice and the cloze questions correctly for this part

of the tutorial. Notably, this particular participant does have lower on average fixations and total fixation time recorded for the entire quiz; however this is by far the lowest. This could mean that the participant has a high amount of prior knowledge about the topic and only needed to skim the text before being confident about being capable of answering the questions. Alternatively, as this recording is an outlier even within the participant's data it could be due to equipment error, i.e. the eye tracker failed to record the participant's eye movements correctly and as a result this is an outlying point.

When the rest of the data set is considered, there are two unequal clusters of data at a high level. Comparing the averages of eye movement measures for these clusters are shown in Table 5.

Table 5: Comparison of average eye movement measures for clusters obtained from hierarchical clustering of format C data

Cluster	Number of fixation		Average fixation duration (sec)		Total fixation duration (sec)		% Fixations for reading		Multiple choice score	Cloze Score
	Text	Questions	Text	Questions	Text	Questions	Text	Questions		
1	325	117	0.23	0.24	74.8	28.7	71%	50%	0.72	0.54
2	127	121	0.18	0.23	22.5	27.1	53%	53%	1	0.72

In format C the participants are presented with the tutorial text and then with the questions about the text in the following page. The reason they are both shown is to show the contrast in average measures between the clusters. There is a clear difference is the reading behaviour of the tutorial between the clusters but that difference no longer exists when it comes to reading the questions. The average number of fixations, average fixation duration, average total fixation duration and percentage fixations as part of reading behaviour are all significantly lower in cluster 2 compared to cluster 1. All of this is evidence that the participants skimmed or at least did not thoroughly read the text in these data instances. The differences between these measures are all statistically significant using two-sided unpaired t-tests with a significance level of $p<0.001$ for number of fixation and total fixation time, $p<0.05$ for the remaining measures. However, if we look at the eye movement behaviour when the participants read the questions this difference disappears. There is no longer any statistical difference between the two clusters. Therefore, the clustering has been dominated by how the text was read.

On average there were high scores were achieved for both the multiple-choice and the cloze score for cluster 2, however there is no statistical difference between the scores obtained between the two clusters.

The outlier was removed from the data set and the remaining data was split into the two clusters. Then the classification techniques were applied to the separate clusters to see if the predictions could be improved. The results did not improve, with MCR values that are similar to those above.

We then used k-means clustering on the eye movements into 3 clusters. All the eye movement measures were used in the clustering. The results of this clustering are shown in Table 6.

Table 6: Comparison of average eye movement measures for clusters obtained from k-means clustering of format C data

Cluster	Number of fixation		Average fixation duration (sec)		Total fixation duration (sec)		% Fixations for reading		Multiple choice score	Cloze Score
	Text	Questions	Text	Questions	Text	Questions	Text	Questions		
1	295	116	0.24	0.25	70	28	76%	54%	0.70	0.42
2	144	90	0.18	0.55	27	39	52%	45%	1	0.81
3	489	150	0.22	0.24	109	37	67%	44%	0.63	0.72

The clusters show quite large differences in the average numbers of fixations and total fixation durations for both the text page and the questions page. There is a strong statistical difference between the number of fixations for each of the clusters for the text page of the tutorial (p=0.0000<0.01). There is a statistical difference between the number of fixations between clusters 1 and 2 (p=0.04<0.05) as well as between clusters 2 and 3 (p=0.007<0.01). There is also a strong statistical difference between the average fixation duration between clusters 1 and 2 (p=0.0002<0.001) as well as between clusters 2 and 3 (p=0.0002<0.001) when considering the text page. All clusters differ in average forward saccade length for the text page. This is strongest when comparing clusters 1 and 2 (p=0.0000<0.01) where we can see that the average forward saccade length is longer in cluster 2. What this points to is that the clustering has clustered the reading behaviours for the text quite strongly. This is concordant with the hierarchical clustering finding.

Additionally, there are now three distinct clusters of reading behaviour as compared the clusters found from hierarchical clustering. These clusters inform more about reading comprehension, as there are differences in the answers obtained between the clusters. Cluster 2 has the highest on average scores for both questions. There is a statistical difference between the multiple-choice (p=0.004<0.01) and cloze (p=0.004<0.01) scores between cluster 1 and 2. Also there is a statistical difference between the cloze score (p=0.02<0.05) for cluster 1 and 3 where as cloze scores in cluster 1 are lower than those in cluster 3, however no difference in multiple-choice score. There is less difference between the other two clusters, however, there is also a statistical difference between the multiple-choice score (p=0.002<0.01) cluster 2 and 3 where the multiple choice scores are lower in cluster 3, but no difference in cloze score.

Given these differences, the data set was divided into 3 sub data sets and the classification techniques were used to predict the reading comprehension scores. The misclassification results remained equally poor.

Finally, clustering was performed based on English speaking ability. Participants had been asked in the pre-experiment questionnaire if English is their native language. The four participants who are native English speakers were grouped together and the five non-native English speakers were in the other group. The comparison of average eye movement measures for the two groups is shown in Table 7.

Table 7: Comparison of average eye movement measures between native and non-native English speakers for format Cnoise

Cluster	Number of fixation		Average fixation duration (sec)		Total fixation duration (sec)		% Fixations for reading		Multiple choice score	Cloze Score
	Text	Questions	Text	Questions	Text	Questions	Text	Questions		
Native English	248	107	0.22	0.40	57	34	68%	50%	0.86	0.72
Non-native English	341	125	0.22	0.24	77	31	70%	50%	0.67	0.44

The results from this clustering are interesting as there is a discrepancy in averages of the eye movement measures and question scores between two groups. The native English speakers showed statistically fewer fixations (p=0.0006<0.001), lower total fixation duration (p=0.0006<0.001) and obtained higher scores for both multiple choice (p=0.04<0.05) and cloze (p=0.003<0.01) questions. There is no statistical difference between the average percentages of fixation classified for reading or average fixation duration. There is however a difference in average fixation duration when reading the questions, but this difference is not statistically significant (p=0.28).

The main reasoning between clustering the participants based on whether they are native English speakers or not is the assumption that the native English speakers would have an intrinsically highly level of reading skill; i.e. those readers have had more practice and therefore will be more proficient. It has been shown that there is a relationship between fixation duration and reading skill.[4] Our results either have not confirmed this relationship or the assumption that the native English speakers are more skilled readers is wrong. We cannot make conclusions either way concerning this as their reading ability was not measured in this study but should in future be investigated. What can be seen is that there is a clear difference between how the two groups of participants read the text and also their reading comprehension performance.

Given this difference, once again the classification techniques were used to predict the reading comprehension question scores for the two groups. However, there is no improvement on misclassification results. The cluster analysis of format C data has revealed some interesting properties about the data set. The clustering was an attempt to improve classification results of reading comprehension scores, however, no relationship between the reading comprehension scores and these eye movement measures could be shown here. Although no relationship could be shown there are interesting relationships in the data that can be explored further. We end the final section with a clear hypothesis relating to the difficulty with format C.

4 Discussion

Generally, the results reflect that fact that the data sets are quite hard to classify, especially the format C data set. The best classification results were obtained using three layers of hidden neurons, which indicates quite complex relationships between the eye movement measures and reading comprehension scores. The best classification results that could be achieved for format A is an average of 86% classification (MCR=0.14), for format B, an average of 89% classification (MCR=0.11) and 79% (MCR=0.21) for format D. These best results were obtained from the ANN when FOE was used as the performance function for training. We have shown that the use of FOE as a performance function for training feed-forward ANNs using back-propagation training provides better classification results then using MSE. In these cases the use of FOE as the performance function

for training gave up to a 50% reduction in misclassification compared to using MSE as the performance function for certain ANN topologies. These are promising results and show that when dealing with a small data set with a large imbalance in classes MSE is not the optimal performance function to use for training neural networks. Further work is needed to generalize to other data sets as well as with other classifiers.

One of the advantages of using FOE is that it is a flexible error function that can be tailored to data sets and problems. Specifying the shape of the FMF used to calculate FOE does this. However, there is no simple way of constructing an FMF. In this analysis we only investigated 7 predefined FMFs, however, a beneficial approach would be to learn the most appropriate FMF shape from the data set. An area of further exploration is how to apply the learning of FMF shape when using other classifiers such as neural networks.

As observed from the results shown throughout the analysis, the format C data set is quite challenging to make predictions from. There are many reasons why this could be so. This is the baseline format where participants had no knowledge of the reading comprehension questions and therefore no goals during the reading process. Instead participants had to read the text to a point that they thought they understand it well enough. Although the questions are designed to assess whether they understand the material or not this cannot be separated from memory processes. In all other formats the participants have access to the questions before or during the reading process. This means that the participants have set goals for the information that they need to extract from the text. This is most likely why formats A and B have the lowest MCR values as these formats present the text with the questions. An important point to take from this investigation is that different predictive methods are required for making predictions depending reading circumstances, in particular, the reader's goals. In all formats other than format C the readers had goals and as such predictions about those goals could be made.

As just stated, the participants would have read the text to the point at which they deemed they understood the text which is completely subjective and dependent on a number of factors including prior knowledge, familiarity with the subject matter, current state (mood, arousal, etc.) as well as their motivations. This could account for the variability in eye movement measures and the reading comprehension outcomes. This data set does not include their subjective reading comprehension, so future work is to record this information and explore relationships between eye movements and subjective comprehension.

From the hierarchical cluster analysis it can be seen that there are relationships in reading behaviours between participants. In particular, the findings from the hierarchical cluster analysis are concordant with previous findings from Underwood et al.[4] who showed that there was no relationship between reading speed and reading comprehension and there is high variability in reading styles between readers. Indeed there is also high variability in reading style exhibited by an individual, which is often induced by lexical, syntactical and grammatical factors about the text.[1] From the hierarchical clustering analysis we can see that this is indeed true; some of the participants alter their reading behaviour based on how they perceive the text. Cluster 2 is composed of 9 data points with 4 each from two participants and 1 entry from another participant. In particular, those two participants with 4 points each in that cluster are native English speakers. These two readers may have a high level of reading skill, in fact, one of those participants achieved 100% on the quiz and the other achieved one of the highest grades in the group. These participants changed their reading style to be closer to skimming the text in the parts of the tutorial that they deemed they could skim and still achieve decent marks. In all other cases they read the text more thoroughly, just as all other participants did. The k-means clustering also confirms these findings where there are clear changes in reading behaviour observed for participants. Each of the k-means clusters were composed of data samples from different participants which shows all that participants to some extent changed their reading behaviour to reflect the text. Ref. 4 note that there is a difference between readers with different skill levels and that is predominantly in that skilled readers are those that can extract information effectively but not necessarily do it quickly. Future work is to investigate reader skill by measuring the participants' reading abilities. In our analysis the native English speakers were separated from the non-native English speakers but this does not guarantee that the reading skill in consistent within those groups. Although we could not find relationships between reading behaviour and the observed reading comprehension, this could be due to many factors and including variability in personal reading behaviour. It has been shown that personalisation of reading measures improves results.[10,11] Future work is also to investigate whether reading comprehension prediction can be improved by personalisation of eye movement measures.

Our immediate future work is to look at predicting whether a reader is a native English speaker or not. Given that there are differences in the reading behaviour of the text it is plausible that this can be achieved. This would be an important feature for an online learning environment whereby if a reader is not native English speaking then the learning material can be altered to reflect this so that they do not contain complicated grammatical constructions and rare words or have difficult readability.

5 Conclusion

We found that there are differences in reading behaviour that can be used to make inferences about the reader. The application of predicting reading comprehension from eye gaze is in adaptive online learning environments. Prediction of

comprehension would allow a system to adaptively change to a student's knowledge level making the learning process more streamlined and more targeted toward their capabilities. If a student is observed to have poor reading behaviour (low number of fixations, low total reading time and so on) then it can be assumed that the student is skimming the text either due to prior knowledge or lack of interest. The educational material can be altered to stimulate the student by providing them with more challenging material or simpler material if the reasons for skimming can be inferred.

The eye movements measures we used may actually relate to how well participants think they understand which is close to how well they really understand for formats A, B and D where the questions are visible and hence we obtained outstanding prediction results for those formats. We need to re-test and ask for participants' subjective belief in their understanding in our future work.

Acknowledgements

Thank you to the proofreaders and comments from reviewers.

References

[1] Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. Psychological Bulletin, 372-422. http://dx.doi.org/10.1037/0033-2909.124.3.372

[2] Campbell, C. S., & Maglio, P. P. (2001). A robust algorithm for reading detection. Paper presented at the Proceedings of the 2001 workshop on Perceptive user interfaces. 1-7. http://doi.acm.org/10.1145/971478.971503

[3] Buscher, G., Dengel, A., & Elst, L. v. (2008). Eye movements as implicit relevance feedback. Paper presented at the CHI '08 Extended Abstracts on Human Factors in Computing Systems, Florence, Italy. 2991-2996. http://doi.acm.org/10.1145/1358628.1358796

[4] Underwood, G., Hubbard, A., & Wilkinson, H. (1990). Eye fixations predict reading comprehension: The relationships between reading skill, reading speed, and visual inspection. Language and speech, 33(1), 69-81. http://dx.doi.org/10.1177/002383099003300105

[5] Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. Scientific Studies of Reading, 10(3), 241-255. http://dx.doi.org/10.1207/s1532799xssr1003_3

[6] Hyrskykari, A., Majaranta, P., Aaltonen, A., & Räihä, K.-J. (2000). Design issues of iDICT: a gaze-assisted translation aid. Paper presented at the Proceedings of the 2000 symposium on Eye tracking research & applications. 9-14. http://doi.acm.org/10.1145/355017.355019

[7] Sibert, J. L., Gokturk, M., & Lavine, R. A. (2000). The reading assistant: eye gaze triggered auditory prompting for reading remediation. Paper presented at the Proceedings of the 13th annual ACM symposium on User interface software and technology, 101-107. http://doi.acm.org/10.1145/354401.354418

[8] Copeland, L., & Gedeon, T. ("in press"). Fuzzy Output Error as the Performance Function for Training Artificial Neural Networks to Predict Reading Comprehension from Eye. Paper submitted to the 21st International Conference on Neural Information Processing (ICONIP), 2014.

[9] Frazier, L.,& Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. Cognitive Psychology, 14(2), 178-210. http://dx.doi.org/10.1016/0010-0285(82)90008-1

[10] Buscher, G., Dengel, A., Biedert, R., & Van Elst, L. (2012). Attentive Documents: Eye Tracking as Implicit Feedback for Information Retrieval and Beyond. ACM Transactions on Interactive Intelligent Systems, 1(2), Article 9:1-30. http://doi.acm.org/10.1145/2070719.2070722

[11] Martínez-Gómez, P., & Aizawa, A. (2014). Recognition of understanding level and language skill using measurements of reading behavior. Paper presented at the Proceedings of the 19th international conference on Intelligent User Interfaces. 95-104. http://doi.acm.org/10.1145/2557500.2557546

[12] Iqbal, S. T., & Bailey, B. P. (2004). Using Eye Gaze Patterns to Identify User Tasks. The Grace Hopper Celebration of Women in Computing 2004.

[13] Salojarvi, J., Puolamaki, K., Simola, J., Kovanen, L., Kojo, I., & Kaski, S. (2005). Inferring relevance from eye movements: Feature extraction. Tech. Rep. A82. Helsinki University of Technology.

[14] Gustavsson, C. J. (2010). Real Time Classification of Reading in Gaze Data (Masters Thesis). School of Computer Science and Engineering. Royal Institute of Technology. Stockholm, Sweden.

[15] Vo, T., Mendis, B. S. U., & Gedeon, T. D. (2010). Gaze Patterns and Reading Comprehension. Neural information processing. Models and applications, 6444, 124-131. http://dx.doi.org/10.1007/978-3-642-17534-3_1

[16] Perkins, K., Gupta, L., & Tammana, R. (1995). Predicting item difficulty in a reading comprehension test with an artificial neural network. Language testing, 12(1), 34-53. http://dx.doi.org/10.1177/026553229501200103

[17] Oh, S.-H. (2011). Error back-propagation algorithm for classification of imbalanced data. Neurocomputing, 74(6), 1058-1061. http://dx.doi.org/10.1016/j.neucom.2010.11.024

[18] Oh, S.-H. (2012). Improving the Error Back-Propagation Algorithm for Imbalanced Data Sets. International Journal of Contents, 8(2), 7-12.

[19] He, H., & Garcia, E. A. (2009). Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering, 21(9), 1263-1284. http://dx.doi.org/10.1109/TKDE.2008.239

[20] Domingos, P. (1999). MetaCost: a general method for making classifiers cost-sensitive. Paper presented at the Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining. 155-164. http://doi.acm.org/10.1145/312129.312220

[21] Kukar, M., & Kononenko, I. (1998). Cost-Sensitive Learning with Neural Networks. Paper presented at the 13th European Conference on Artificial Intelligence. 445-449.

[22] Elkan, C. (2001). The foundations of cost-sensitive learning. Paper presented at Proceedings of the 17th International Joint Conference on Artificial Intelligence. 2, 973-978. http://dl.acm.org/citation.cfm?id=1642194.1642224

[23] Zhou, Z.-H., & Liu, X.-Y. (2006). Training cost-sensitive neural networks with methods addressing the class imbalance problem. Knowledge and Data Engineering, IEEE Transactions on, 18(1), 63-77. http://dx.doi.org/10.1109/TKDE.2006.17

[24] Gedeon, T., Copeland, L., & Mendis, B. S. (2012). Fuzzy Output Error. Australian Journal of Intelligent Information Processing Systems, 13(2), 37-43.

[25] Mendis, B. S. U., & Gedeon, T. D. (2008). A comparison: Fuzzy signatures and Choquet Integral. Paper presented at the Fuzzy Systems, 2008. FUZZ-IEEE 2008. (IEEE World Congress on Computational Intelligence). IEEE International Conference on. 1464-1471. http://dx.doi.org/10.1109/FUZZY.2008.4630565

[26] Dombi, J. (1990). Membership function as an evaluation. Fuzzy sets and Systems, 35(1), 1-21. http://dx.doi.org/10.1016/0165-0114(90)90014-W

[27] Dombi, J., & Gera, Z. (2005). The approximation of piecewise linear membership functions and Łukasiewicz operators. Fuzzy sets and Systems, 154(2), 275-286. http://dx.doi.org/10.1016/j.fss.2005.02.016

[28] Dombi, J., & Gera, Z. (2008). Rule based fuzzy classification using squashing functions. Journal of Intelligent and Fuzzy Systems, 19(1), 3-8.

[29] Gera, Z., & Dombi, J. (2005). Genetic Algorithm with Gradient Based Tuning for Constructing Fuzzy Rules. Publications of International Symopsium of Hungarian Researchers of Computational Intelligence. 86-95.

[30] Fletcher, J. M. (2006). Measuring Reading Comprehension. Scientific Studies of Reading, 10(3), 323-330. http://dx.doi.org/10.1207/s1532799xssr1003_7

[31] Copeland, L., & Gedeon, T. (2013). Measuring reading comprehension using eye movements. Paper presented at the Cognitive Infocommunications (CogInfoCom), 2013 IEEE 4th International Conference on. 791-796. http://doi.acm.org/10.1109/CogInfoCom.2013.6719207

[32] Moller, M. F. (1993). A scaled conjugate gradient algorithm for fast supervised learning. Neural Networks, 6(4), 525-533. http://dx.doi.org/10.1016/S0893-6080(05)80056-5

[33] Hagan, M.T., and M. Menhaj. (1994) Training feed-forward networks with the Marquardt algorithm, IEEE Transactions on Neural Networks, 5(6), 989–993. http://dx.doi.org/10.1109/72.329697

[34] Xu, R., & Wunsch, D. (2008). Clustering (Vol. 10). John Wiley & Sons.