Turkish Journal of Electrical Engineering and Computer Sciences

Volume 30 | Number 1 Article 16

1-1-2022

Automatically classifying familiar web users from eye-tracking data:a machine learning approach

MELİH ÖDER

ŞÜKRÜ ERASLAN

YELİZ YESİLADA

Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2022) 30: 233 – 248

doi:10.3906/elk-2103-6

Automatically classifying familiar web users from eye-tracking data: a machine learning approach

Melih Öder[1], **Şükrü Eraslan**[2,*], **Yeliz Yeşilada**[2]
[1]Department of Information Systems, Graduate School of Informatics,
Middle East Technical University, Ankara, Turkey
[2]Middle East Technical University, Northern Cyprus Campus, Kalkanlı, Mersin 10, Turkey

Received: 01.03.2021 • **Accepted/Published Online:** 09.11.2021 • **Final Version:** 19.01.2022

Abstract: Eye-tracking studies typically collect enormous amount of data encoding rich information about user behaviours and characteristics on the web. Eye-tracking data has been proved to be useful for usability and accessibility testing and for developing adaptive systems. The main objective of our work is to mine eye-tracking data with machine learning algorithms to automatically detect users' characteristics. In this paper, we focus on exploring different machine learning algorithms to automatically classify whether users are familiar or not with a web page. We present our work with an eye-tracking data of 81 participants on six web pages. Our results show that by using eye-tracking features, we are able to classify whether users are familiar or not with a web page with the best accuracy of approximately 72% for raw data. We also show that with a resampling technique this accuracy can be improved more than 10%. This work paves the way for using eye-tracking data for identifying familiar users that can used for different purposes, for example, it can be used to better locate certain elements on pages such as adverts to meet the users' needs or it can be used to do better profiling of users for usability and accessibility assessment of pages.

Key words: Human computer interaction, world wide web, online advertising, machine learning, eye-tracking

1. Introduction

The web plays a crucial role in our daily lives. The design of web pages is important in attracting new users and having users revisiting pages [1]. Eye-tracking is widely used to assess the usability of pages which collect enormous amount of data [2] – where people look at called fixation, how long they look at called fixation duration, and the sequence of fixations called scanpath [3]. This data typically encodes a lot of information about user behaviours and characteristics. There have been studies to analyse, and mine the collected data for different purposes. Some algorithms have been proposed to identify patterns of use in terms of trending scanpaths [4–6]. Some studies focused on differentiating user groups, for example identifying users with dyslexia [7, 8] or users with autism [9, 10] from their eye-tracking data (see Section 2).

In this paper, we focus on identifying the familiarity of users with a web page automatically from their eye-tracking data. Here, we refer to *familiarity* with a web page as "close acquaintance with or knowledge of a particular web page"[1]. This knowledge could be used for different purposes. In particular, it could be used as

*Correspondence: seraslan@metu.edu.tr

[1]Adopted from Oxford Dictionary definition: Oxford English Dictionary (no date). Familiarity [online] Website https://en.oxforddictionaries.com/definition/familiarity [accessed 26 February 2021].

a way of profiling web users and adapting web pages based on their familiarity where necessary. For example, due to banner blindness and familiarity, people tend not to read certain elements (e.g., adverts) on a web page. However, if the system automatically detects whether the user is familiar with a web page, then those elements could be relocated, especially in the areas which are mostly visited by the familiar users, to achieve better engagement and to support loyalty (see Section 5).

When eye-tracking studies are conducted, different users participate in those studies with different demographic information such as age, gender, educational background, experience, etc. Some studies show that user profile affects certain features extracted from eye-tracking data. For example, Eraslan and Yesilada [6] suggest that familiarity affects the common scanpath created for a group of users, Pan et al. [11] show the gender effect on certain eye-tracking features, and similarly Habuchi et al. [12] show that being a regular web user affects certain features. In this paper, we first pull together a set of eye-tracking features and explore these features with different machine learning algorithms to detect whether users are familiar or not with web pages (see Section 3). We use a dataset collected for another study [4]. The dataset was created with 81 participants where each participant visited six different web pages for two different tasks: for spontaneous browsing without any specific task and for searching to complete a specific task [13]. The participants were also asked to rate their familiarity with the web pages before visiting them during the eye-tracking study, therefore the study had a record of familiarity of the participants.

There are different machine learning algorithms and each has different pros and cons. When we look at the literature though for similar tasks, we can see that logistic regression and support vector machine (SVM) are the most widely explored ones. Therefore, in this paper, we mainly explore these two algorithms for familiarity detection and report their results. In overall, our results show that the best accuracy values with the raw data are achieved on the Apple page. The best accuracy with logistic regression is 69.23% and the best accuracy with SVM is 71.79% for both browsing and searching. Besides these, we have also explored KNN and also random forest, and we show that they also give similar and consistent results (see Sections 4 and 5). Given the eye-tracking datasets we have, one can see that the number of instances of training these algorithms might be considered small. Therefore, we also explore a resampling technique, in particular the SMOTE (synthetic minority oversampling technique) technique which creates synthetic instances based on the present instances. With SMOTE, the accuracy values on the Apple page increase to 87.64% for browsing and 84.26% for searching with logistic regression, and to 75.28% for both browsing and searching with SVM. To the best of our knowledge, this paper presents the first work on familiarity detection based on the eye-tracking data.

2. Background and related work

Web page familiarity is typically measured based on the subjective assessment. For example, in a study conducted by Eraslan and Yesilada [6], web page familiarity was measured by asking the participants to rate how frequent they visit the pages with a 5 point Likert scale (daily, weekly, monthly, less than a month, never). Even though the study showed the familiarity effects on the results to some extent, it did not conduct an in-depth analysis to check whether the data can be used to distinguish the familiar and unfamiliar users. The same approach was also followed in other studies [14]. However, there are also previous work that investigates how the familiarity of website could be measured more objectively. Specifically, Zhang and Ghorbani [15] propose a measurement to determine the familiarity with a website by taking several factors into account including prior experience, repeated exposure, level of processing and forgetting rate. As the level of processing of the

proposed measurement is mainly associated with the number of different web pages visited within a website, this measurement is not very suitable for individual web pages. The level of processing can also be measured by eye-tracking technology. If the system detects the familiarity by simply tracking eye movements only, this can decrease the time and effort needed to measure the familiarity objectively.

Machine learning and data mining have been widely explored in different research areas. Recently, we have also started to see many applications in human computer interaction (HCI). As it is indicted by Holzinger [16], HCI is interested in the questions of human perception, cognition and intelligence and so it takes place in the centre of supervised learning methods. A recent survey also shows that machine learning approaches have been also explored with eye-tracking data for different purposes [17]. This review in particular highlights that machine learning is an "important aspect in evolving eye-tracking applications owing to their ability to learn from existing data, make better decisions, be flexible [...]". In our study, we also show an application of machine learning to eye-tracking data. In fact, our focus is on user familiarity which is a cognition process and closely related with perception. We also investigate existing work on eye-tracking that uses machine learning techniques to explore user characteristics. Table 1 gives a summary of such related work. This table focuses on presenting related work that focuses on automatically detecting user characteristics, however of course there are other works that use machine learning and eye-tracking [18], but the focus of this work is on understanding user characteristics. Our study is similar to the studies conducted by Rello and Ballesteros [7] and Yaneva et al. [9] as they also perform binary classification based on eye-tracking data. However, the focus of these studies is different. Specifically, Rello and Ballesteros [7] aim to classify people as with or without dyslexia and Yaneva et al. [9, 19] aim to classify people as with or without autism. Rello and Ballesteros [7] use SVM, Yaneva et al. [9] use logistic regression and recently Yaneva et al. [19] use logistic regression, random forest, SVM and naive Bayes for identifying people with autism from eye-tracking data. Furthermore, Klaib et al. [17] show that algorithms such as SVM and regression are widely used and they are robust in understanding data. Hence, we also explore these machine learning algorithms to perform our binary classification (familiar or not familiar) based on eye-tracking data.

Table 1 also shows us that these studies mostly rely on small sample sizes (3-97 participants). This is mainly because eye-tracking studies require user participation and it is costly to perform user studies in terms of time and scope [4, 26]. However, previous studies show that algorithms can be trained in such a way to cope with such small sample sizes.

In summary, there are many studies that bring together eye-tracking and machine learning with data mining. As can be seen from Table 1, different approaches are proposed to characterise users based on their eye-tracking data. However, to best of our knowledge none of them try to detect familiarity of users automatically, and our work in this paper aims to fill this gap.

3. Methodology

The dataset used in our study was collected to evaluate the algorithm called scanpath trend analysis (STA) which aims to create a trending scanpath among a group of users [4, 27]. The eye-tracking study was approved by the University of Manchester Senate Ethics Committee (ref: CS90). The dataset is briefly explained below and its full description can be found in [4, 27].

Participants: We have the data of 81 participants in this dataset and was collected in two different sites: the University of Manchester and Middle East Technical University Northern Cyprus Campus (METU NCC).

Table 1. Related work (#: Number of participants)

Ref	Purpose	Technique	#[2]	Features Related				
				Fixation Duration	Path Angles	Fixation Counts	Fixation Distance	Predefined AOIs
[20]	Investigate relationship between visual memory and gaze features	- DBSCAN (clustering) - Permutation test (nonparametric test)	24	✓	✓	✓	X	X
[7]	Identify if a user is dyslexic or not	Support vector machine (SVM)	97	✓	X	✓	X	✓
[9]	Identify if a user is autistic or not	Logistic Regression	30	✓	X	✓	X	✓
[21]	Determine the relevance of document titles to search tasks	-Principal component analysis (PCA) -Self-organising maps -Linear discriminant analysis (LDA)	3	✓	X	✓	✓	X
[22]	Cluster eye tracking recordings as representation of viewer interest	Mean shift procedure	6	✓	X	X	✓	X
[23]	Assess student learning	Logistic regression	47	✓	✓	✓	✓	✓
[24]	Identify behavioural patterns of use	-Differential sequence analysis -Principal component analysis (PCA)	-	X	X	X	X	✓
[25]	Design information visualisation systems dynamically adapt to user characteristics	-Statistical analysis -Principal component analysis (PCA)	35	✓	✓	✓	✓	✓

Three of the participants were excluded due to some recording problems (such as calibration, improper standing positions, etc.). Thus, we have the data of 78 participants.

Equipment: To record eye movements, a 17" monitor with a built-in Tobii T60 eye tracker was used. Its screen resolution was adjusted as 1280 x 1024.

Materials: The home pages of six websites were used: Apple, Babylon, AVG, Yahoo, Godaddy, and BBC. These websites were selected from the Alexa Top-100[3] and their visual complexity was measured by ViCRAM to ensure that they have different visual complexity [28]. Eraslan and Yesilada [4] also divided these web pages into their elements or areas of interest (AOIs) for their study with the extended and improved version of the vision-based page segmentation (VIPS) algorithm, which automatically segments web pages by using their

[3]Alexa ranks pages based on their popularity: Alexa (no date). The top 500 sites on the web [online]. Website https://www.alexa.com/topsites [accessed 26 February 2021].

source code and visual representations based on different granularity levels [29]. In this work, we also use the same elements to train our machine learning classifiers.

Tasks: The participants were asked to complete two different tasks on each page but one task at a time. In the browsing task, the participants were given 30 s to investigate the web pages independently as they want, in other words, they viewed the web pages spontaneously, while in the searching one, they took a specific task to be completed on the web page in maximum 2 min. For example, they were asked to complete the following two tasks as a searching task on the Apple page: (1) Can you locate the link that allows watching the TV ads relating to iPad mini? (2) Can you locate a link labeled iPad on the main menu?.

Procedure: At the beginning, the participants read an information sheet about the study and then signed a consent form to show that they were a volunteer to participate in the study. After that, they filled a questionnaire about their gender, age group, and educational background. The participants were also asked to rate their familiarity with the web pages as 1 (Daily), 2 (Weekly), 3 (Monthly), 4 (Less than once a month), or 5 (Never). After that, they sat in front of the monitor and completed the given tasks on the web pages. Both the order of the tasks and web pages were randomised to eliminate the ordering effect.

3.1. Dataset preparation

The output from the eye-tracker, visual segmentation of web pages (AOIs) and demographic information of the participants are available from the previous study. For our learning algorithms, we used the familiarity rating of the participants. When the participants rated the web page with 1, 2 or 3, it means that they visit the page at least once a month. Therefore, they are aware of the web page. However, when the participants rated the web page with 4 or 5, it means that they visited the web page very few times or they have not visited the web page yet. Therefore, if the participants rated the web page with 1, 2, or 3, we considered them as familiar with the page. Otherwise, we considered them as not familiar. There are two main reasons for considering familiarity as a binary feature. The first one is the size of our dataset as we do not have sufficient records for each familiarity score. The second one is that these familiarity scores were given by the participants, and by using familiarity as a binary feature, we aim to decrease the subjectivity level. This approach was also followed by Eraslan and Yesilada [6]. This information is then used to label the data of the participants as familiar and unfamiliar. The number of familiar and unfamiliar participants for each web page is given in Table 2. Unfortunately, for the AVG, Babylon and GoDaddy web pages, the number of familiar and unfamiliar users was unbalanced (numbers respectively: 6-72, 5-73 and 1-77). Therefore, we could not continue to use the data for these three pages as this will cause the overfitting problem where the algorithms will learn exactly the data given and not be able to predict future observations [16].

Table 2. Number of familiar users for each web page

Web Page	#Familiar	#Unfamiliar
Apple	22	56
BBC	40	38
Yahoo	30	48
AVG	6	72
Babylon	5	73
GoDaddy	1	77

The data was collected for each participant for each web page. Therefore, in order to run the machine learning algorithms, we needed to convert the data into a format that brought together the data for each web page. This is because we were interested in classifying the participants as familiar or not to a particular web page. Furthermore, the raw eye-tracking data was collected in the format given in Table 3. However, we needed to convert this data into a set of meaningful features. In order to decide which features could be used, we investigated the literature and identified a set of features that could have an impact on familiarity, which are explained in the following section.

In order to compute these features, we developed a tool in Java that takes the eye-tracking output (see Table 3), the generated AOIs (see Table 4) and demographic features of participants (see Table 5) and then automatically computes these features and gives an output table (see Table 6) which is ready to be trained by machine learning algorithms. This tool is publicly available and can be also used by other researchers to generate the features explained in the following section (see *Open data* section). We use AOIs as the input because some of these features require AOI information. This tool enables us to automate feature generation process. It also enables us to add new features and be able to compute them in different ways.

Table 3. Sample eye-tracking data output (F.I: FixationIndex, T.S: TimeStamp, F.D: FixationDuration, M.F.X: MappedFixationPointX, M.F.Y: MappedFixationPointY, S.N: StimuliName)

F.I	**T.S**	**F.D**	**M.F.X**	**M.F.Y**	**S.N**
1	1207	300	629	247	http://emine.ncc.metu.edu.tr/survey/web/pages/http/www.babylon.com/
2	1507	383	677	348	http://emine.ncc.metu.edu.tr/survey/web/pages/http/www.babylon.com/
3	1890	317	746	348	http://emine.ncc.metu.edu.tr/survey/web/pages/http/www.babylon.com/
...	...	...	...	...	...

Table 4. Areas of interest sample records (A.N: AoIName, T.L.X: TopLeft_XCoordinate, W: Width, T.L.Y: TopLeft_YCoordinate, H: Height, S.A.N: ShortAoIName)

A.N	**T.L.X**	**W**	**T.L.Y**	**H**	**S.A.N**
VB.1.1.1	998	179	11	47	A
VB.1.1.2	185	813	11	47	B
VB.1.2.1	207	453	178	212	C
...	...	...	...	...	...

Table 5. Demographic features of sample participants (F.L: Familiarity level - 1, 2, 3: familiar, 4, 5: unfamiliar)

ID	**Gender**	**Age Group**	**Education Level**	**Godaddy (F.L)**	**Apple (F.L)**	**AVG (F.L)**	**Yahoo (F.L)**	**Babylon (F.L)**	**BBC (F.L)**
1	Female	25-34	High School	5	4	3	4	5	3
2	Male	25-34	Master	5	3	4	2	4	4
3	Female	18-24	Bachelor	5	3	4	4	5	1
...	...	...	...	...	...	...	...	...	...

3.2. Features

Eye-trackers typically collect data with many different features. To decide which features to be used, we first look at the literature to see which ones have been commonly used in the studies using machine learning techniques

Table 6. Feature extraction tool sample output (M.S: Mean of sequence based fixation durations (ms), S.S: Sum of sequence based fixation durations (ms), M.P: Mean of page based fixation durations (ms), S.P: Sum of page based fixation durations (ms), S.F: Sequence based fixation counts (num), P.F: Page based fixation counts (num), N.V: Number of viewed AoIs per page based fixations (num), F.F: First fixated AoI, P.F: Percentage of first fixated AoI (ms), D.F: Duration of first fixated AoI (ms), M.D: Mean of distances between page based fixations (pixels), S.D: Sum of distances between page based fixations (pixels), P.A: Mean of path angles between page based Fixations (°), S.P.A: Sum of path angles between page based fixations (°), F.C: Page based fixation counts per task based fixation counts (num)), Fm: Familarity (1: familiar, 0: unfamiliar)

Scanpath	**M.S**	**S.S**	**M.P**	**S.P**	**S.F**	**P.F**	**N.V**	**F.F**	**P.F**	**D.F**	**M.D**	**S.D**	**P.A**	**S.P.A**	**F.C**	**Fm**
CDDCFF..	313.9	8476	302.9	26.9	27	89	0.06	C	0.007	200	176.6	15.5	1.78	157.4	3.29	1
HEHCCC..	332.0	4980	356.5	28.5	15	80	0.07	H	0.035	999	224.5	17.7	-24.8	-1965	5.33	0
CHGDCF..	266.4	3996	355.6	28.0	15	79	0.08	C	0.004	133	170.5	13.3	3.86	301.3	5.26	1

to explore user characteristics. We can see that fixation duration, fixation count, fixation distance, path angles, and predefined AOIs are often used [7, 20–25]. These features are important for our study but we also looked at the wider literature to see if there are other features that could be used to show differences between familiar and unfamiliar users (e.g. [3, 31, 32]). Table 7 shows the features that we selected to explore in this study. Features that are summarised in this table are based on the related work, therefore, allows us to base our work on the previous related work. We group the features into two: sequence-based and page-based ones. Sequence-based features are computed by sequencing fixations which are over AOIs only. Page-based features are computed without taking AOIs into account. Most of the selected features listed in Table 7 have been shown as affected by the familiarity. For example, Marchal et al. [20] show that there is a strong correlation between the relative path angles and the memorised items. Similarly, Eraslan and Yesilada [6] show that there are differences in the scanpaths of familiar and unfamiliar users on web pages. Furthermore, Pan et al. [11] show that "web page viewing behaviour is driven by the gender of subjects, the order of web pages being viewed, and the interaction between site types and the order of the pages being viewed". In their study, they particularly show that mean fixation duration and gazing time are affected by the order of pages viewed. Based on these features, we also selected some other features. For example, we included the first fixated AOI and the percentage of its duration over the total duration in our feature list as the duration of the first fixation was shown as affected by the familiarity [20].

3.3. Algorithms and tools

In this study, we mainly explore logistic regression and SVM because these two classification algorithms are mostly exploited in relevant research (see Section 2). We explore these algorithms for two main reasons: (1) they are also widely used in the literature for identifying user characteristics, and (2) our dataset is small so we need algorithms that can work efficiently with a small dataset. We use Weka 3.8.3 tool[4] for applying these two algorithms. Logistic regression conducts an optimal number of LogitBoost iterations for fitting the logistic models. LogitBoost iterations are cross-validated by default which lead to automatic attribute selection [33]. Logistic regression enables us to optimise it such as setting number of iterations, batchsize and so on. In our case, the maximum iteration number is set to 1 which is a default value. SVM transforms nominal attributes to binary and replaces all missing values, then normalises all attributes by default. It conducts probability

[4]Weka (no date). Weka 3: Machine Learning Software in Java [online] Website https://www.cs.waikato.ac.nz/ml/weka/ [accessed 26 February 2021].

Table 7. Familiarity Related Features (SB: Sequence-based, PB: Page-based, NA: Not Applicable)

Type	Feature	How to Compute	Affected by Familiarity
SB	Scanpath	Shows the sequence of AOIs that a user looks at.	Eraslan and Yesilada [6]
SB	Mean of sequence based fixation durations	Average of fixation durations over AOIs.	Pan et al. [11] , Greene and Rayner [30]
SB	Sum of sequence based fixation durations	Summing fixation durations over AOIs.	Pan et al. [11], Greene and Rayner [30]
SB	Sequence based fixation Counts	Number of AOIs that a participant views.	Marchal et al. [20], Greene and Rayner [30]
SB	First fixated AOI	Determines at which AOI the participant looks at first.	NA
SB	Percentage of first fixated AOI	Calculates the percentage of the first fixated duration according to the whole duration.	NA
SB	Duration of first fixated AOI	Determines the duration when the participant looks at the first AOI.	Marchal et al. [20]
PB	Mean of page based fixation durations	Average fixation durations.	Pan et al. [11], Greene and Rayner [30]
PB	Sum of page based fixation durations	Sum of fixation durations.	Pan et al. [11], Greene and Rayner [30]
PB	Page based fixation counts	Counts the fixations.	Marchal et al. [20], Greene and Rayner [30]
PB	Number of viewed AOIs per page based fixations	Divides the number of AOIs that the participant views in over the number of page based fixations.	NA
PB	Mean of distances among page based fixations	Calculates the average of distances among all points.	Marchal et al. [20]
PB	Sum of distances among page based fixations	Calculates the total distances among all points.	Marchal et al. [20]
PB	Mean of path angles among page based fixations	Calculates the average angle that takes place between sequential points according to horizontal axis.	Marchal et al. [20]
PB	Sum of path angles among page based fixations	Calculates the sum of angles that take place between sequential points according to horizontal axis.	Marchal et al. [20]

estimations and couples predicted probabilities by using pairwise coupling method. While applying SVM, sequential minimal optimization (SMO) is used which is an algorithm to exploit SVM for training data models [34]. In addition to logistic regression and SVM, we also explore KNN and random forest in this study to check the results with similar classification algorithms.

In order to ensure the accuracy of predictions, we applied both of the algorithms with 10-fold cross-validation which divides the dataset into ten parts and repeatedly uses nine parts for training and one part for testing until all parts are used for testing, and then averages the results. The dataset has class imbalances in terms of familiarity. In order to balance familiar – unfamiliar classes, SMOTE (synthetic minority oversampling technique) is used as a resampling technique. SMOTE performs oversampling toward uniformly distributed dataset. It takes each minority class instances into account and creates synthetic examples joining any/all k minority class nearest neighbors [35]. In our study, SMOTE creates five nearest neighbors. Moreover, the

number of synthetic instances depends upon the amount of minority class adjusted as 50%. In this study, SMOTE is applied before the cross-validation. Therefore, raw and SMOTE datasets are trained and tested with logistic regression and SVM separately.

3.4. Evaluation metrics

In order to assess the outcome of these algorithms, we use 10-fold cross-validation with a 90-10% split of the user data and we use the following metrics computed based on true positive (TP), true negative (TN), false positive (FP) and false negative (FN). Correctly classified familiar users are considered TP, correctly classified unfamiliar users are considered as TN, incorrectly classified familiar users are considered FN and incorrectly classified unfamiliar users are considered as FP. Based on these, we computed precision, recall, F-measure, and accuracy.

4. Results

We first conducted an analysis to identify and select the most informative features to classify web users as familiar or not, and then train the model with the selected features by using logistic regression and SVM.

4.1. Feature selection

Measuring the effectiveness of features helps us in understanding how these features have an ability to predict the familiarity. Therefore, we first investigate information gain of our feature set. Table 8 shows the information gain of our features. Average merit shows the value (0-1) which indicates the percentage of relationship between the feature and familiarity objectively. In this study, the feature of which average merit is under 0.5 were removed from the feature set; because according to the study in [36], 0.5 is relevancy threshold; that is, a feature of which average merit is under 0.5 is irrelevant to the target feature. Therefore, we explore the accuracy of the machine learning algorithms without the following four features: page based fixation counts, duration of first fixated AOI, sequence based fixation counts, and first fixated AOI.

Table 8. Feature selection – Information gain

Average Merit	Feature
0.87	Scanpath
0.87	Sum of sequence based fixation durations
0.87	Sum of path angles between page based fixations
0.87	Sum of page based fixation durations
0.87	Percentage of first fixated AOI
0.87	Mean of sequence based fixation durations
0.87	Mean of distances between page based fixations
0.87	Sum of distances between page based fixations
0.87	Mean of path angles between page based fixations
0.87	Mean of page based fixation durations
0.792	Page based fixation counts per task based fixation counts
0.76	Number of viewed AOI per page based fixations
0.438	Page based fixation counts
0.305	Duration of first fixated AOI
0.244	Sequence based fixation counts
0.07	First fixated AoI

4.2. Cross-validation results

Table 9 shows the number of instances that are used to train and test the models. In the raw dataset we have the data of 78 users. For example, in the raw data case for Apple, we have 22 familiar and 56 unfamiliar participants, but with applying SMOTE, we have 33 familiar and 56 unfamiliar participants. In order to cross-validate our results, we have applied 10-fold cross-validation. Specifically, for each fold, we have used 90% of the users for training the model and 10% for testing the trained model. Table 10 and Table 11 show the outputs of the models developed with logistic regression and SVM for the browsing and searching tasks separately.

Table 9. No of familiar and unfamiliar participants

	Raw data		**SMOTE**	
Pages	**Familiar**	**Unfamiliar**	**Familiar**	**Unfamiliar**
Apple	22	56	33	56
BBC	40	38	40	57
Yahoo	30	48	45	48

Table 10. Logistic regression models with selected features

		Browsing			**Searching**		
		Apple	**BBC**	**Yahoo**	**Apple**	**BBC**	**Yahoo**
	Visual complexity	Low	High	Medium	Low	High	Medium
Raw Data	**Accuracy**	69.23%	55.12%	47.43%	69.23%	45.15%	60.25%
	Precision	0.629	0.553	0.370	0.613	0.431	0.533
	Recall	0.692	0.551	0.474	0.692	0.462	0.603
	F-measure	0.638	0.536	0.410	0.624	0.410	0.502
	Number of Instances	78	78	78	78	78	78
SMOTE	**Accuracy**	87.64%	77.31%	70.96%	84.26%	69.07%	78.49%
	Precision	0.876	0.775	0.716	0.844	0.710	0.797
	Recall	0.876	0.773	0.710	0.843	0.691	0.785
	F-measure	0.876	0.774	0.706	0.839	0.693	0.782
	Number of Instances	89	97	93	89	97	93

Table 11. SVM models with selected features

		Browsing			**Searching**		
		Apple	**BBC**	**Yahoo**	**Apple**	**BBC**	**Yahoo**
	Visual complexity	Low	High	Medium	Low	High	Medium
Raw Data	**Accuracy**	71.79%	53.84%	61.53%	71.79%	46.16%	61.53%
	Precision	0.718	0.539	0.615	0.718	0.437	0.615
	Recall	0.718	0.538	0.612	0.718	0.462	0.615
	F-measure	0.836	0.521	0.762	0.836	0.419	0.762
	Number of Instances	78	78	78	78	78	78
SMOTE	**Accuracy**	75.28%	68.04%	69.89%	75.28%	60.82%	67.74%
	Precision	0.823	0.679	0.810	0.823	0.644	0.775
	Recall	0.753	0.680	0.699	0.753	0.608	0.677
	F-measure	0.711	0.680	0.665	0.711	0.609	0.641
	Number of instances	89	97	93	89	97	93

For the browsing tasks, given raw data with logistic regression, the best result is 69.23% (F-measure: 0.638) for the Apple page and the worst result is for the Yahoo page with 47.43% (F-measure: 0.410). With SMOTE, the accuracy results are increased and the accuracy for the Apple and Yahoo pages become 87.64% (F-measure: 0.876) and 70.96% (F-measure: 0.706), respectively. For the searching tasks, given raw data for logistic regression, the best result is 69.23% (F-measure: 0.624) for the Apple page and the worst result is for the BBC page with 45.15% (F-measure: 0.410). With SMOTE, the accuracy results are also increased and the accuracy for the Apple and BBC pages become 84.26% (F-measure: 0.839) and 69.07% (F-measure: 0.693), respectively.

For the browsing tasks, given raw data with SVM, our best result is 71.79% (F-measure: 0.836) for the Apple page and the worst result is for the BBC page with 53.84% (F-measure: 0.521). With SMOTE, the accuracy for the Apple and BBC pages become 75.28% (F-measure: 0.711) and 68.04% (F-measure: 0.680), respectively. For the searching tasks, given raw data, for SVM, our best result is 71.79% (F-measure: 0.836) for the Apple page and the worst result is for the BBC page with 46.15% (F-measure: 0.419). With SMOTE, the accuracy values for the Apple and BBC pages become 75.28% (F-measure: 0.711) and 60.82% (F-measure: 0.609), respectively.

Even though we mainly explore both logistic regression and SVM for developing a model to predict familiarity, we also explore whether we can achieve consistent results with our methodology using similar classification algorithms. The results achieved by KNN and random forest are presented in Table 12 and Table 13 respectively, and they are consistent with the results of logistic regression and SVM.

Given raw data with KNN, the best result is 67.08% (F-measure: 0.604) for the Apple page and the worst result is 54.43% (F-measure: 0.469) for the Yahoo page for the browsing tasks. When we look at the searching tasks, the best result is 72.15% (F-measure: 0.686) for the Apple page and the worst result is 45.57% for the BBC page (F-measure: 0.386). Using raw data with random forest, the best result is 70.88% (F-measure: 0.830) for the Apple page and the worst result is 53.16% (F-measure: 0.382) for the BBC page for the browsing tasks. For the searching tasks, the best result is 70.89% (F-measure: 0.830) for the Apple page and the worst result is 49.37% (F-measure: 0.343) for the BBC page. Similar to logistic regression and SVM, the accuracy results are increased with SMOTE in all cases.

Table 12. K-nearest-neighbor models with selected features

		Browsing			**Searching**		
		Apple	**BBC**	**Yahoo**	**Apple**	**BBC**	**Yahoo**
	Visual complexity	Low	High	Medium	Low	High	Medium
Raw Data	**Accuracy**	67.08%	55.69%	54.43%	72.15%	45.57%	60.76%
	Precision	0.585	0.564	0.445	0.690	0.400	0.559
	Recall	0.671	0.557	0.544	0.722	0.456	0.608
	F-measure	0.604	0.525	0.469	0.686	0.386	0.537
	Number of instances	78	78	78	78	78	78
SMOTE	**Accuracy**	88.89%	67.35%	86.17%	92.22%	60.20%	87.23%
	Precision	0.889	0.713	0.864	0.923	0.601	0.873
	Recall	0.889	0.673	0.862	0.922	0.602	0.872
	F-measure	0.889	0.628	0.862	0.922	0.524	0.872
	Number of instances	89	97	93	89	97	93

Table 13. Random forest models with selected features

		Browsing			**Searching**		
		Apple	**BBC**	**Yahoo**	**Apple**	**BBC**	**Yahoo**
	Visual complexity	Low	High	Medium	Low	High	Medium
Raw Data	**Accuracy**	70.88%	53.16%	62.02%	70.89%	49.37%	60.76%
	Precision	0.709	0.754	0.620	0.709	0.263	0.382
	Recall	0.709	0.532	0.620	0.709	0.494	0.608
	F-measure	0.830	0.382	0.766	0.830	0.343	0.469
	Number of instances	78	78	78	78	78	78
SMOTE	**Accuracy**	83.33%	58.16%	88.30%	83.33%	58.16%	85.10%
	Precision	0.869	0.582	0.904	0.869	0.582	0.874
	Recall	0.833	0.582	0.883	0.883	0.582	0.851
	F-measure	0.820	0.735	0.881	0.820	0.735	0.848
	Number of instances	89	97	93	89	97	93

5. Discussion

This paper explores SVM and logistic regression to detect whether users are familiar or not with a web page from their eye-tracking data. The results are promising and our best accuracy is more than 70% with raw data and 85% with SMOTE. Our results show some web pages are better in guiding the classification of users. For example, with both algorithms, we see that the Apple page is producing higher results in both browsing and searching. This could be due to the underlying design or the complexity of the page. However, further studies need to be conducted with more pages and more users, to make such conclusions.

In order to check the results with similar other algorithms, we have conducted further experiments with k-nearest neighbors (KNN) and random forest. The browsing and searching datasets of Apple, BBC and Yahoo are again trained and tested (10-fold cross-validation) with these two algorithms. The KNN results and the random forest results are consistent with the findings of SVM and logistic regression.

In this paper, we also explored feature selection. We looked at the information gain of the features used and removed the ones which were not reliable based on the information gain metric. When we look at the features that have the merit score below 0.5 that were removed are the following features: 1) page based fixation counts, 2) sequence-based fixation counts, 3) duration of first fixated AOI and 4) first-fixated AOI. The first one counts the fixations and the second one counts the AOIs. The results showed us that these two features do not have significant effect as they are likely to be similar for all pages. When we look at the last two features, they are about the first AOI fixated and also the duration. These also have no significant effect and this could also be explained as from the first entry to the page, the familiarity is probably not a discriminating factor. However, features that are about the overall processing of the page such as scanpths, etc. have higher merit scores. Compared to SVM, with logistic regression, we have slightly better results, however the results in overall are consistent. We also show that by using SMOTE as a resampling technique, we can also improve the accuracy.

The work proposed here can have many applications. It has been shown that people can easily become banner-blindness when they are familiar with the content [37]. That means certain elements on the page are not actually visited by familiar users. Therefore, if the critical content or content like advertisement is located in those areas, there is a possibility that these users will not engage with them, and knowing if they are familiar, the pages can be better personalised and the critical content can be relocated in other parts of the page. Similarly,

it has been shown that website similarity is an effective factor for the loyalty of customers [38]. If the website is designed to sell products and the unregistered familiar web users are detected automatically, some special offers can be provided for them to increase their loyalty.

Situationally-induced impairments and disabilities (SIIDs) is a well-known phenomenon that shows that people experience temporary disabilities due constraints in the environment typically caused by their context such as screen size [39]. Therefore, if the system can detect that the user is familiar with a page, in different context, the page can be automatically reformatted such that the context limitations is better addressed.

Web pages have started to be used as a material to diagnose a number of disabilities such as dyslexia [40] and autism [19]. However, these studies tend to focus on the subjective measurements of users' familiarity. Therefore, the data analysis is based on the assumptions that people critically labeled their familiarity. However, if a system can automatically identify whether the user is familiar or not, fairer data analysis could be made. Trust and familiarity have a relationship even though it does not necessarily mean that when a user is familiar with a website, s/he trusts the website [15, 41]. Besides, it has been suggested that web users tend to like a website if it is similar to the ones that they are familiar with [15]. Therefore, if the system is developed to automatically detect whether the user is familiar with the system, then it can be used by web designers to initially assess different web interfaces to understand which one would be more appealing to end-users.

Finally, our study is not without limitations. We use a dataset from another study which meant the data is not collected for this purpose and therefore it is not tailored. However, it also meant that we could not use the full dataset and we could only use the data for three web pages. In future, we would like to conduct more eye-tracking studies with a better balance of familiarity of users and web pages. This will ensure that we will validate our findings here with another study. In this paper, we report results with Weka but in future studies, other platforms such as Phyton-based platforms can also be explored. Furthermore, in this paper we explore four different machine learning algorithms including KNN, SVM, logistic regression and random forest as we have a small dataset, however, there are many other algorithms and approaches that could be explored. For example, with a much bigger dataset, deep learning approaches can be explored.

6. Conclusion

Eye-tracking studies typically collect enormous amount of data that encodes a lot of information about the users behaviour and characteristics on the web. The main goal of this paper is to explore machine learning algorithms to automatically detect users characteristics in particular familiarity with a web page. We present a study conducted with two machine learning algorithms: logistic regression and SVM. Their results are reinforced and empowered by conducting two different machine learning algorithms - k-nearest neighbors and random forest. In this study, we use the eye-tracking dataset of 81 participants on three web pages collected as part of another study. The results with these two algorithms show that with using familiarity-related eye movement features, we are able to classify people from their eye-tracking data whether they are familiar or not with a web page with the best accuracy of approximately 72% for raw data. To the best of our knowledge, this is the first study exploring machine learning algorithms in automatically identifying familiar users from their eye-tracking data.

Open data

Our data converter written in Java can be found in our external repository at https://github.com/melihoder/TJEECS.git. All the materials used for the evaluation are also available in this repository.

References

[1] Deng T, Zhao L,Feng L. Enhancing web revisitation by contextual keywords. In: the 13th International Conference on Web Engineering 2013; 323–337.

[2] Tüzün H, Akıncı A, Kurtoğlu M, Atal D,Pala FK. A study on the usability of a university registrar's office website through the methods of authentic tasks and eye-tracking. The Turkish Online Journal of Educational Technology 2013; 12 (2):26–38.

[3] Poole A, Ball LJ. Eye tracking in human-computer interaction and usability research: Current status and future. In: Ghaoui C (editor). Encyclopedia of Human-Computer Interaction. Pennsylvania: Idea Group Inc 2005; 211–219.

[4] Eraslan S, Yesilada Y,Harper S. Eye tracking scanpath analysis on web pages: How many users?. In: the Ninth Biennial ACM Symposium on Eye Tracking Research & Applications, ETRA'16 2016; 103–110.

[5] Eraslan S, Yesilada Y, Harper S. Eye tracking analysis techniques on web pages: A survey, evaluation and comparison. Journal of Eye Movement Research 2015; 9 (1): article 2.

[6] Eraslan S,Yesilada Y. Patterns in eyetracking scanpaths and the affecting factors. Journal of Web Engineering 2015; 14 (5-6):363–385.

[7] Rello L,Ballesteros M. Detecting readers with dyslexia using machine learning with eye tracking measures. In: the 12th Web for All Conference W4A'15 2015; 16:1–8,

[8] Matsuda N,Takeuchi H. Frequent pattern mining of eye-tracking records partitioned into cognitive chunks. Appl. Comp. Intell. Soft Comput. 2014; article 10.

[9] Yaneva V, Ha LA, Eraslan S, Yesilada Y,Mitkov R. Detecting autism based on eye-tracking data from web searching tasks. In: the Internet of Accessible Things W4A'18 2018; 16:1–10.

[10] Eraslan S, Yesilada Y, Yaneva V, Harper S. Autism detection based on eye movement sequences on the web: A scanpath trend analysis approach. In: the 17th International Web for All Conference W4A'20 2020; 11:1–10.

[11] Pan B, Hembrooke HA, Gay GK, Granka LA, Feusner MK et al. The determinants of web page viewing behavior: An eye-tracking study. In: the 2004 Symposium on Eye Tracking Research & Applications ETRA'04 2004; 147–154.

[12] Habuchi Y, Takeuchi H, Kitajima M. The influence of web browsing experience on web-viewing behavior. In: the 2006 Symposium on Eye Tracking Research & Applications ETRA'06 2006; 47–47.

[13] Marchionini G. Exploratory search: from finding to understanding. Communications of the ACM 2006; 49 (4):41–46.

[14] Eraslan E, Yesilada Y, Yaneva, Ha LA "Keep it simple!": an eye-tracking study for exploring complexity and distinguishability of web pages for people with autism. Universal Access in the Information Society 2021; 20:69-–84.

[15] Zhang J,Ghorbani AA. Familiarity and trust: Measuring familiarity with a web site. In: the 2nd Annual Conference on Privacy, Trust and Security PST'04 2004; 23–28.

[16] Holzinger A. Human-Computer Interaction and Knowledge Discovery (HCI-KDD): What Is the Benefit of Bringing Those Two Fields to Work Together? In: Availability, Reliability, and Security in Information Systems and HCI: IFIP WG 8.4, 8.9, TC 5 International Cross-Domain Conference CD-ARES 2013 2013; 319–328.

[17] Klaib AF, Alsrehin NO, Melhem WY, Bashtawi HO, Magableh AA. Eye tracking algorithms, techniques, tools, and applications with an emphasis on machine learning and internet of things technologies. Expert Systems with Applications 2021; 166:114037.

[18] Dong J, Zhang Y, Yue J, Hu Z. Comparison of random forest, random ferns and support vector machine for eye state classification. Multimedia Tools and Applications 2016; 75 (1):11763–11783.

[19] Yaneva V, Ha LA, Eraslan S, Yesilada Y, Mitkov R. Detecting high-functioning autism in adults using eye tracking and machine learning. IEEE Transactions on Neural Systems and Rehabilitation Engineering 2020; 28 (6):1254–1261.

[20] Marchal F, Castagnos S, Boyer A. Tell me what you see, i will tell you what you remember. In: The 2016 Conference on User Modeling Adaptation and Personalization UMAP'16 2016; 293–294.

[21] Salojarvi J, Kojo I, Simola J, Kaski S. Can relevance be inferred from eye movements in information retrieval? In: Workshop on Self-Organizing Maps WSOM'03 2003; 261–266.

[22] Santella D. Robust clustering of eye movement recordings for quantification of visual interest. In: the 2004 Symposium on Eye Tracking Research & Applications ETRA'04 2004; 27–-34.

[23] Bondareva D, Conati C, Feyzi R, Harley JM, Azevedo R et al. Inferring Learning from Gaze Data during Interaction with an Environment to Support Self-Regulated Learning. In: Artificial Intelligence in Education: 16th International Conference AIED 2013; 229–238.

[24] Steichen B, Wu MMA, Toker D, Conati C,Carenini G. Te,Te,Hi,Hi: Eye Gaze Sequence Analysis for Informing User-Adaptive Information Visualizations. In: User Modeling, Adaptation, and Personalization: 22nd International Conference UMAP 2014; 183–194.

[25] Toker D, Conati C, Steichen B, Carenini G. Individual user characteristics and information visualization: Connecting the dots through eye tracking. In: the SIGCHI Conference on Human Factors in Computing Systems CHI '13 2013; 295–304.

[26] Eraslan S, Yesilada Y, Harper S. Less users more confidence: How aois don't affect scanpath trend analysis. Journal of Eye Movement Research 2017; 10 (4): article 6.

[27] Eraslan S, Yesilada Y,Harper S. Scanpath trend analysis on web pages: Clustering eye tracking scanpaths. ACM Transactions on the Web 2016; 10 (4): article 20.

[28] Michailidou E, Eraslan S, Yesilada Y,Harper S. Automated prediction of visual complexity of web pages: Tools and evaluations. International Journal of Human-Computer Studies 2021; 145:102523.

[29] Akpinar ME, Yesilada Y. Vision based page segmentation algorithm: Extended and perceived success. In: the ICWE 2013 International Workshops on Current Trends in Web Engineering - Springer-Verlag 2013; 8295:238–252.

[30] Greene HH, Rayner K. Eye movements and familiarity effects in visual search. Vision Research 2001; 41 (27):3763–3773.

[31] Jacob RJK, and Karn KS. Eye tracking in Human-Computer interaction and usability research: Ready to deliver the promises. In: Hyona J, Radach R, and Deubel H. The Mind's eye: Cognitive The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research, Elsevier Science 2003; 573–603.

[32] Goldberg JH, Kotval XP. Computer interface evaluation using eye movements: methods and constructs. International Journal of Industrial Ergonomics 1999; (6):631–645.

[33] Sumner M, Frank E, Hall M. Speeding Up Logistic Model Tree Induction. In: Knowledge Discovery in Databases: PKDD 2005: 9th European Conference on Principles and Practice of Knowledge Discovery in Databases 2005; 675–683.

[34] Platt JC. Fast Training of Support Vector Machines Using Sequential Minimal Optimization. In: Scholkopf B and Burges CJC and Smola AJ (editors). Advances in Kernel Methods - Support Vector Learning MIT Press, Cambridge, MA, USA 1999; 185–208.

[35] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. Smote: Synthetic minority over-sampling technique. J. Artif. Int. Res. 2002; 16 (1):321–357.

[36] Demisse GB, Tadesse T, Bayissa Y. Data Mining Attribute Selection Approach for Drought Modelling: A Case Study for Greater Horn of Africa. International Journal of Data Mining & Knowledge Management Process 2007; 7:1–16.

[37] Owens JW, Chaparro BS, Palmer EM. Text advertising blindness: The new banner blindness?. J. Usability Studies 2011; 6 (3):172–197.

[38] Kaya B, Behravesh E, Abubakar AM, Kaya OS, Orús C. The moderating role of website familiarity in the relationships between e-service quality, e-satisfaction and e-loyalty. Journal of Internet Commerce 2019; 18 (4):369–394.

[39] Akpinar E, Yesilada Y, Temizer S. The effect of context on small screen and wearable device users' performance - a systematic review. ACM Comput. Surv. 2020; 53 (3): article 52.

[40] Rello L, Ballesteros M. Detecting readers with dyslexia using machine learning with eye tracking measures. In: The 12th Web for All Conference W4A '15 2015; 16:1–8.

[41] Gefen, D. E-commerce: the role of familiarity and trust. Omega 2000; 28 (6):725–737.