

You Are What You Click: Web Interaction Analysis for User Profile Detection

Leonardo Germán Loza Bonora^{1,2}[0009-0006-4166-4219], Julián Grigera^{1,2,3}[0000-0002-7962-4312], and Alejandra Garrido^{1,2}[0000-0002-5052-705X]

¹ LIFIA, Fac. de Informatica, Univ. Nac. de La Plata, La Plata, Argentina
{lloza, juliang, garrido}@lifia.info.unlp.edu.ar

² CONICET, Argentina

³ CICPBA, Argentina

Abstract. Evaluation of web user interaction serves various purposes like analysis and evolution of UX, but it requires significant resources. To reduce the time and cost associated with such assessments, several automated solutions have been developed to analyze interactions by capturing logs. However, many of these approaches assume that all users interact similarly, disregarding individual characteristics such as mouse and keyboard movement speeds. Additionally, they often overlook the time required for analysis. We propose that identifying a user's interaction profile can enhance the quality of automated log analysis. To achieve this, in this proposal interactive user profiles on the web will be studied and defined. Through experimentation, we will analyze real user behavior and its relationship with these profiles. Finally, the detection of user profiles and interactive characteristics will be automated, considering key variables such as detection speed and prediction accuracy.

Keywords: Web Interaction · User profiling · Log analysis

1 Introduction

The evaluation of user interaction on the web is used for many purposes, such as discovering usability or accessibility problems. By simply observing how users use different standard input devices, such as the mouse and keyboard, diagnostics can be obtained to help improve interaction. This type of evaluations, however, usually require a considerable number of resources. The most commonly used technique for evaluating interaction aspects is user testing, which requires the recruitment of volunteers and specialized professionals to analyze their behavior. To reduce the time and cost that these types of assessments entail, several solutions have emerged that propose to automate interaction analysis by capturing interaction logs in different areas such as usability [1] or accessibility[2].

Many of the works that propose to capture and analyze the interaction in an automated way do not take into account the characteristics of the users that generate the logs, assuming that they all interact in a similar way. This leads to a uniform interpretation of the behavior of users with very different

characteristics, such as mouse or keyboard movement speeds, or waiting times between actions. In addition, in many cases, logs are captured anonymously, which adds the challenge of not knowing the user beforehand [3, 4]. However, users show different behaviors in aspects like clicking, scrolling speed or pause durations. These differences, moreover, cannot always be explained solely on the basis of demographics such as age or computer experience.

Most of the studies on user interaction with interfaces have been conducted in the context of desktop applications. These works highlight the importance of low-level event capture and analysis [5]. They also indicate that the metrics used to classify user behavior vary according to the context studied [6, 7].

Furthermore, there are studies that analyze user interaction in web applications, but they are far from our idea of recognizing and evaluating user profiles. In general, these studies focus on analyzing the importance of captured events [8] or on predicting future actions, such as abandoning a web page [9].

In general, studies in these areas also do not consider the time required to perform this analysis. In some scenarios, user profile classification needs to be done quickly without compromising quality or efficiency. This is the case with adaptive interfaces, which must rapidly recognize user interactions to adjust the interface to their needs [10]. Delaying customization can result in users becoming accustomed to a less efficient or less suitable UI, reducing their satisfaction.

We hypothesize that knowing the user’s interaction profile helps improve the quality of automated log analysis. For example, if a person usually moves the mouse cursor slowly, a sudden and fast movement could indicate the presence of a problem [11]. Moreover, having a deeper knowledge on user interaction behavior would serve various purposes, like helping designers adapt the UIs to different interaction styles for improved usability and accessibility. Our focus is in desktop web applications.

2 Related work

Several authors have worked on the characterization of users based on interaction logs, especially in the area of desktop applications. However, these studies tend to focus on user ability in specific task-oriented programs, leaving aside general interaction. For example, Ghazarian et al. [5] presented methods and experiments to build automatic skill classifiers, in order to adapt interfaces to the user’s needs according to their level of proficiency. Importantly, they note the lack of studies that lead to modeling user skill level from low-level events on the interface, such as mouse movement.

Along the same lines, Grossman et al. [12] explored the viability of software applications, such as AutoCAD, automatically inferring user experience levels based on their in-situ usage patterns. Their results show us that the approaches in other studies of human-computer interaction apply not only to low-level and short-term competence, but also to high-level software competence. It also shows that metrics captured from in-situ data, without knowledge of the user’s environment or task, can be correlated with software proficiency.

Moreover, in the context of web applications, Duran et al. [13] introduced methods to detect collaborative behavior patterns automatically using “Web Usage Mining” techniques to generate collaborative student profiles. These techniques attempt to discover knowledge from data obtained from user interaction with the web, such as web logs, browser logs, user profiles, user sessions, bookmarks, folders, and scrolls. However, these approaches do not consider user interaction while using a web page, as proposed in this plan. They also provide insight into the profiles that are interesting to analyze and various techniques to perform analysis, preprocessing and predictive modeling.

Finally, in the area of web interaction capture and analysis, several studies have investigated user behavior characteristics and patterns. Huang et al. [14] analyzed mouse cursor behavior on search engine results pages (SERPs), focusing on clicks, movements, and hovers, and demonstrated that analysis of this data is valuable for estimating the relevance of a result and distinguishing types of search abandonment. Similarly, Kirsh [15] conducted a study focused on mouse movement directions and speeds, suggesting that these factors may be more revealing than simple cursor position. These studies are focused on the importance of individual events or analyzing a user’s behavior to predict their next action, far from our idea of understanding and predicting the characteristics of their interaction. However, these findings are fundamental to identify which events should be captured and how to do it efficiently to predict the user’s actions in the shortest time or with the least amount of data possible.

3 Objectives

This research aims to analyze user interaction to study and define user profiles on the Web. To achieve this, the following objectives are defined:

- Analyze through experimentation the behavior of real users and their relationship with the profiles studied.
- Study and define interactive user profiles on the web, according to specific criteria.
- Automate the detection of user profiles or interactive features, considering different variables such as detection speed and prediction accuracy. The detection is designed to be anonymous, i.e., avoiding user’s personal data or the pages the user goes through.
- Validate the found user’s profiles with other metrics established in the literature.

To study user interaction to the required level of detail, we will conduct different tests and experiments in which we will capture real users interaction.

We also propose to study our metrics related to user profiling with other established metrics, for example, a statistical analysis to evaluate the relationship between the effort a user must make when using widgets in a user interface [16] and the time it takes to use them.

4 Methodology

Due to the nature of the research topic, which is closely related to HCI, the primary method will consist on user studies. For this reason, several experiments will be conducted that require special preparation for working with volunteers. In some cases, it will be necessary to develop test web applications, and in others, to modify existing applications to record interactions or evaluate changes in the interface, such as when comparing different versions of the same interface. In addition, recruitment and experiment execution processes are required that take into account the multiple biases that can affect human studies. We will use user recruiting platforms due to the high volume of users required.

Interaction captures will be performed using tools that we have developed and continue to improve. These consist of configurable browser extensions, which allow to select the low level events to capture in the most used browsers and define to which server to send the collected data.

We aim to define meaningful profiles that help different scenarios, looking for common characteristics to avoid overfitting.

5 Preliminary Results

We aim to create predictive models for web interactive characteristics, such as cursor speed or interaction styles. For this purpose, we proposed the Composite Interaction Speed (CIS) metric that allows us to find how early we can estimate web users interaction speed. CIS is the equally weighted sum of three key velocities in the literature: mouse, scroll, and intra-keyboard speed. Since these use different units, they were normalized using a relative unit based on the average unit of each user's captured session. This allows the speeds to be compared at different times of a user's total interaction and to choose an optimal speed point, looking for the earliest time where the average speed is close to the final speed and there is little standard deviation.

A preliminary study was conducted analyzing the behavior of users on web interfaces with a high interaction in the context of user testing. We used logs from a previous study [17], which contain captures of users performing tasks on various web pages. The data relevant to the study were extracted from these captures, and the CIS metric was calculated. Early results suggest that the optimal speed point is at 500 events or 50 seconds [18]. Even if we worked closely with each participant, we applied anonymization techniques on the captured events, since we intend our detection mechanisms to avoid keeping personal data.

A second long-term experiment was carried out, where interaction of several volunteers was captured over a longer time period. To achieve this, a browser extension was built to capture web events in a private way. Using the captured data, the CIS was calculated and the result was that the optimal speed point is at 2000 events or 8.3 minutes. An alternative criteria result was also proposed that gave us a different optimal speed point at 800 events or 5.7 minutes.

We are capturing new volunteers interactions to ensure statistical validity. Also, we are studying how to optimize the weights for the CIS metric calculation.

Following the objective of validating the user profiles with other metrics, we performed an experiment to evaluate an interaction complexity metric called 'Big I' [19]. This metric allows to evaluate web interface alternatives by estimating the interaction complexity of each of them by measuring the number of steps to complete a task. We recruited 100 users to completed a task on a fictitious web page, which allowed us to measure interaction times, speed, and relate the complexity of the steps to the time required to complete them [20].

6 Contributions

The expected contributions of this work will be:

- An analysis of low-level interaction behavior on the web, and how different users behave differently in general cases.
- One or more taxonomies of web users depending on their interaction habits
- Experiments that demonstrate the different users profiles, and the captured data for other researchers to use
- Detection mechanisms for predicting user profiles based on the studied taxonomies, considering aspects such as detection speed, efficiency and quality.

Acknowledgments. This study was funded by the Argentinian National Agency for Scientific and Technical Promotion (ANPCyT), grant number PICT-2019-02485

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Tonio Carta, Fabio Paternò, and Vagner Figueredo De Santana. Web usability probe: a tool for supporting remote usability evaluation of web sites. In *Human-Computer Interaction—INTERACT 2011: 13th IFIP TC 13 International Conference, Lisbon, Portugal, September 5-9, 2011, Proceedings, Part IV 13*, pages 349–357. Springer, 2011.
2. Hernisa Kacorri, Sergio Mascetti, Andrea Gerino, Dragan Ahmetovic, Hironobu Takagi, and Chieko Asakawa. Supporting orientation of people with visual impairment: Analysis of large scale usage data. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*, pages 151–159, 2016.
3. Maximilian Speicher, Andreas Both, and Martin Gaedke. Tellmyrelevance! predicting the relevance of web search results from cursor interactions. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, pages 1281–1290, 2013.
4. Julián Grigera, Alejandra Garrido, José Matías Rivero, and Gustavo Rossi. Automatic detection of usability smells in web applications. *International Journal of Human-Computer Studies*, 97:129–148, 2017.
5. Arin Ghazarian and S Majid Noorhosseini. Automatic detection of users' skill levels using high-frequency user interface events. *User Modeling and User-Adapted Interaction*, 20:109–146, 2010.

6. Amy Hurst, Scott E Hudson, and Jennifer Mankoff. Dynamic detection of novice vs. skilled use without a task model. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 271–280, 2007.
7. Christiane Attig, Ester Then, and Josef F Krems. Show me how you click, and i'll tell you what you can: Predicting user competence and performance by mouse interaction parameters. In *Intelligent Human Systems Integration 2019: Proceedings of the 2nd International Conference on Intelligent Human Systems Integration (IHSI 2019): Integrating People and Intelligent Systems, February 7-10, 2019, San Diego, California, USA*, pages 801–806. Springer, 2019.
8. Dmitry Lagun, Mikhail Ageev, Qi Guo, and Eugene Agichtein. Discovering common motifs in cursor movement data for improving web search. In *Proceedings of the 7th ACM international conference on Web search and data mining*, pages 183–192, 2014.
9. Abdigani Diriye, Ryen White, Georg Buscher, and Susan Dumais. Leaving so soon? understanding and predicting web search abandonment rationales. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pages 1025–1034, 2012.
10. Nilanka Rathnayake, Dulani Meedeniya, Indika Perera, and Anuradha Welivita. A framework for adaptive user interface generation based on user behavioural patterns. In *2019 Moratuwa Engineering Research Conference (MERCOn)*, pages 698–703, 2019.
11. Javier Hernandez, Pablo Paredes, Asta Roseway, and Mary Czerwinski. Under pressure: sensing stress of computer users. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 51–60, 2014.
12. Tovi Grossman and George Fitzmaurice. An investigation of metrics for the in situ detection of software expertise. *Human-Computer Interaction*, 30(1):64–102, 2015.
13. Elena B Duran and Analia Amandi. Web usage mining approach to detect student's collaborative skills. *Journal of Web Engineering*, pages 093–112, 2009.
14. Jeff Huang, Ryen W White, and Susan Dumais. No clicks, no problem: using cursor movements to understand and improve search. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1225–1234, 2011.
15. Ilan Kirsh. Directions and speeds of mouse movements on a website and reading patterns: a web usage mining case study. In *Proceedings of the 10th International Conference on Web Intelligence, Mining and Semantics*, pages 129–138, 2020.
16. Juan Cruz Gardey, Julián Grigera, Andrés Rodríguez, Gustavo Rossi, and Alejandra Garrido. Predicting interaction effort in web interface widgets. *International Journal of Human-Computer Studies*, 168:102919, 2022.
17. Julián Grigera, Juan Cruz Gardey, Andres Rodriguez, Alejandra Garrido, and Gustavo Rossi. One metric for all: Calculating interaction effort of individual widgets. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–6, 2019.
18. Leonardo Germán Loza Bonora, Julián Grigera, Juan Cruz Gardey, and Alejandra Garrido. Web user interaction speed study. In *2022 41st International Conference of the Chilean Computer Science Society (SCCC)*, pages 1–4. IEEE, 2022.
19. Helmut Degen. Big i notation to estimate the interaction complexity of interaction concepts. *International Journal of Human-Computer Interaction*, 38(16):1504–1528, 2022.
20. Leonardo Germán Loza Bonora, Julián Grigera, and Helmut Degen. A study on interaction complexity and time, 2025.