Longitudinal analysis of low-level Web interaction through micro behaviours

Longitudinal analysis of low-level Web interaction through micro behaviours

Aitor Apaolaza, Simon Harper and Caroline Jay
University of Manchester
School of Computer Science
Kilburn Building, Oxford Road, MANCHESTER, M13 9PL, UK
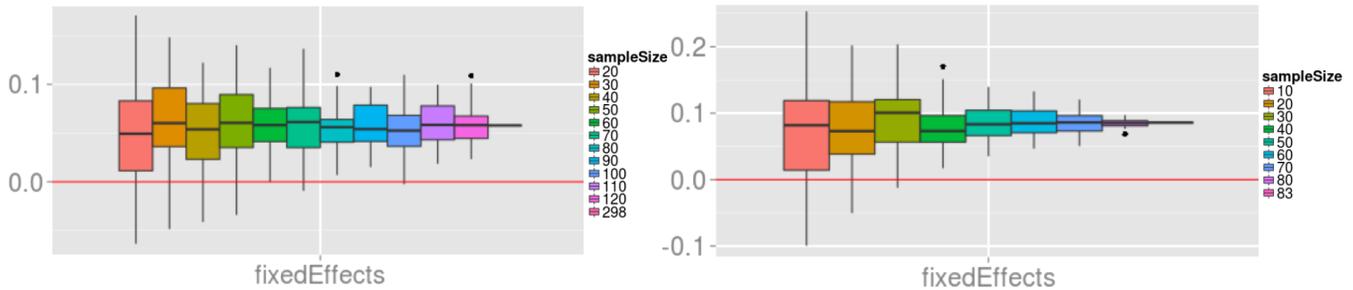[aitor.apaolaza|simon.harper|caroline.jay]@manchester.ac.uk

Figure 1: Evolution of the duration of users' episode over time in each phase.

ABSTRACT

To truly understand how people learn to navigate and use a Web site or application, we need to collect real usage data over extended periods of time. Detailed Web interaction data gathered in the wild (from URLs visited, to keystrokes and mouse movements) has the potential to provide an in-depth, ecologically valid view of interaction, and enable an understanding of how behaviour evolves over time. Interpreting such data is extremely challenging, however. We present a longitudinal data-driven analysis of fine-grained interaction data captured from 14,000 recurrent users over 12 months. At the core of our approach is the aggregation of low-level interaction data into *micro behaviours*. By analysing changes in these behaviours as a function of users' accumulated interaction time, we were able to demonstrate how users' interaction evolves as they become more familiar with a Web page. The results demonstrate that monitoring micro behaviours offers a simple and easily extensible *post hoc* means of understanding how Web-based behaviour evolves over time.

Keywords

Longitudinal; Web ergonomics; usability; behaviour; Web interaction

Categories and Subject Descriptors

H.5.2 [**User Interfaces**]: Evaluation/methodology, Ergonomics; H.5.4 [**Hypertext/Hypermedia**]: User issues

1. INTRODUCTION

Long-term studies are necessary to support the design of interfaces that accommodate interaction over long periods of time. Laboratory studies provide a constrained environment in which confounding variables can be controlled, but are not a feasible way of studying an individual user's interaction over days or weeks. Monitoring behaviour remotely provides a means of obtaining naturalistic, 'in the wild' observations, but analysing data collected in these circumstances is potentially challenging.

Existing approaches to the remote study of behaviour usually rely on prejudgements that help simplify the analysis step. Some investigations have either given participants predefined tasks or made assumptions about those that the user is likely to be performing, to enable the comparison of different interaction sequences [1, 11]. Predefining tasks means prejudging users' interaction goals, however, and making assumptions about the task being performed that may be inaccurate; in both cases, this risks the validity of the study.

We designed a remote capture solution and captured detailed Web interaction data continuously for over two years, and report here on how we are using this approach to understand long term, real world Web usage. Instead of employing a predefined longitudinal interaction metric [7] we were interested in discovering how users' interaction behaviour changes over time. At the heart of our approach is the transformation of low-level data (such as mouse movements) into manageable temporal units of interaction, while avoiding assumptions about users' tasks or goals. These units, which we term *micro behaviours*, describe small samples of user interaction. We describe a number of micro behaviours,

based on evidence from the literature and the application of a data-driven approach to analysis.

We hypothesised that the changes in these behaviours over time could be indicators of how users' interaction behaviour evolves as they get more familiar with the Web site or application. We compared our initial findings with an additional four months of captured interaction, supplemented with a survey that asks users about their familiarity with the site. The results support our hypothesis that the manifestation of certain micro behaviours are a valid proxy for users' familiarity with a website.

We present results discovering evolving aspects of different micro behaviours over time. When users scroll in a continuous way (such as scanning a page) the speed increases and the duration of this scroll action reduces as they get more familiar with the Web page. Users' interaction time with the mouse decreases as they become more familiar with the page, which may be caused by the lack of necessity to explore the page. Finally, we found that users' episode length within a Web page increases as they become more familiar, possibly indicating the reuse of a known page as reference.

2. CAPTURE SOLUTION

In order to obtain naturalistic observations, it was important that interaction data was captured unobtrusively. Our solution also needed to be scalable, as it would be required to manage large amounts of data captured over extended periods of time. The data reported here were obtained by monitoring the `cs.manchester.ac.uk` site. This site contains information about the School of Computer Science at the University of Manchester, covering news, events, teaching and research, and is used by a range of people (including current and prospective students and staff). As the site is frequently updated, the number of pages it contains varies; in this study we captured data from a total of 1411 different URLs.

We based our capture solution on the UsaProxy [3] tool, which supports the recording of low-level Web interaction events. The original tool was extensively modified, in particular, to allow deployment by adding Javascript code to the Web pages, rather than requiring users to set their browser to re-route all connections through a proxy server. Proxies have been successfully employed in short remote experiments in the past [5], but we discovered they are not suitable for longer term experiments. In particular, we found out that various bots and random internet requests made the server unstable and led to us closing the proxy connection altogether. More details about the capture tool can be found in [2][1].

Captured data.

Specific users are tracked via a unique, anonymised code stored in a cookie. Events captured include all mouse and keyboard interaction, as well as browser window events, changes to the state of the elements on the page, and other system information. The state of the DOM and its changes during interaction are also stored, allowing the context of interaction to be potentially recreated. All the data is stored in JSON format in a NoSQL database, which allows for extensibility and query scalability and means that further events can easily be included as required. The design also supports the modification or removal of existing events should they become deprecated.

[1]The source code for the capture tool is available in `iam-data.cs.manchester.ac.uk/data_files/12`.

Window events, indicating when the page is loaded, or when the page has lost focus, combined with user specific information, provide context for lower-level events. This information allows the grouping of events according to the specific environment in which they took place, such as the browser tab – identifiable via the page load timestamp – and URL. Logging the platform and browser version supports the control of event compatibility according to the system. The ever-changing nature of the Web and the possibility of enabling partially supported events makes this information useful for discerning problems with the way a given system handles interaction events.

3. ANALYSIS OF MICRO BEHAVIOURS

The focus of our analysis is to provide insight into how users' interaction changes over time. An important initial step when dealing with large amounts of freely instigated low-level interaction data is to group it appropriately from a temporal perspective. We decided to segment the data into 'episodes', which equate to a single, continuous session of Web use. We split data by applying an inactivity threshold of 40 minutes to user interaction [6, 8]. The threshold was determined by looking at the distribution of inactive periods, and calculating the point after which the variance between the length of inactive periods was small enough to be negligible. This helped ensure that the number of episodes containing a long interruption was minimal [12].

For each episode, we grouped low-level events (individual mouse movements, keystrokes etc.) into sequences that represent what we term *micro behaviours*. These behaviours represent a small sample of user interaction, such as clicking a link, typing in a text field, or scrolling down the page. Previous work has shown that it is possible to identify certain behaviours that are common across individuals through behavioural observation and interaction data analysis, and also that these behaviours are important indicators of both what the user is doing, and how successfully they are completing their task [9, 10]. For example, scrolling down quickly to the bottom of a page and then up again indicates that someone is struggling to find the information they are looking for [10].

We extracted several micro behaviours from the data, based on past research, and manual observation of individual user's interaction. As opposed to the aggregation of interaction data for each episode – e.g. total amount of scroll – extraction of micro behaviours allows us to group interactions into smaller units. Analysing finer grained groups of interactions provides a more detailed insight into users' behaviour, and supports comparisons that would not otherwise be possible. For example, if a user scrolls down fast for a short period of time, it would go unnoticed if data were aggregated for each episode. However, if the scroll actions are combined into micro behaviours, this interaction would stand out as a specific "fast scroll" micro behaviour.

We analysed the characteristics of these behaviours to find meaningful changes over time. In order to account for the uniqueness of each user, we employed linear mixed models, which allow the inclusion of random effects. If a linear mixed model is based on certain independent variables – in our case a time metric and a micro behaviour characteristic – the random effects represent the residual that cannot be accounted for by these variables. When random effects are added to the model, they explicitly model the inter-subject variability. In this case, adding users as random effects tells the model to consider their values as different subjects. The result is a more sensible interpretation of the data. The model
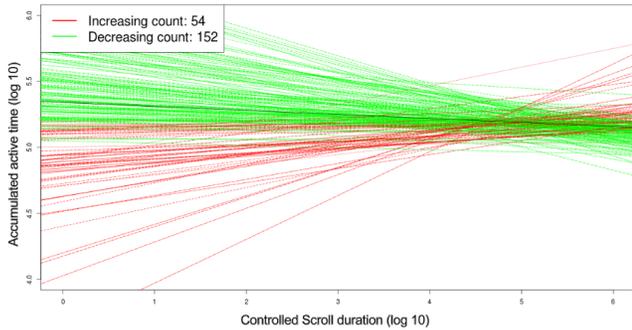
Figure 2: Mixed linear model showing the relation of users' aggregated active time against the duration of the "controlled scroll" behaviour in the cs.manchester.ac.uk page

is robust against unbalanced data [4] and calculates different intercepts and slopes for each user. It also takes into account how noisy the data from each user is and calculates a user specific intercept and slope taking into consideration data from other users.

In Figure 2 a plot of this model is shown, indicating the correlation of users' aggregated active time (aggregated amount of time the user was active on the page over the entire period of data capture) against the duration of the "controlled scroll" behaviour.

The black line represents the model prediction for the overall population, indicating that the value of this feature decreases over time. Despite this prediction, it can be seen that this same feature was found to be increasing over time for some users. Furthermore, these lines are the result of the prediction for those particular users taking the entire population's data into account, i.e. each line represents the model for that user "conditioned" by the rest of the available data.

Due to the linear mixed model's noise handling behaviour, the application of these models can be sensitive to certain users exhibiting particularly strong correlations. In this case, even if those correlations are not shared by other users, the model could still be biased towards them. To tackle this, we applied the model to different user sample sizes chosen at random from the population. We then observed the distribution of the model's fixed effect values – i.e. the value of the slope, indicating that particular feature's tendency to change over time. This approach results in a more conservative and reliable prediction of changes. In order to avoid disparity in the number of data points per user, we also enforce thresholds in the sampling of users. More precisely, we set a minimum amount of interaction for the user to be included in the analysis, as well as removing long-term interaction beyond which users became scarce, to prevent a small number of highly-recurrent users overpowering the model. The examples below include users who have interacted for at least 5 episodes, or have 4 minutes of aggregated active time, and exclude any data after 20 episodes, or 20 minutes of aggregated active time.

4. FINDING LONGITUDINAL FACTORS

We applied the technique introduced in Section 3 to look for characteristics that showed a correlation with the time the user spent in the Web site. Two different metrics were employed to measure time: episode count and users' aggregated active time. Aggregated active time is a precise measurement of the total amount of time the user spent active in the page. We aggregated the duration of all peri-

ods containing continuous interaction without an interruption longer than a stipulated threshold of 50 seconds. This threshold was obtained by applying the same technique described earlier, used to differentiate "episodes", changing the algorithm to use a scale of seconds.

Data was collected in two phases. "Phase 1" contains data captured over a year from 14,000 different recurring users. "Phase 2" occurred over the subsequent four months, and includes data from over 3,000 different recurrent users. Throughout phase 2, we concurrently carried out remote surveys of both first time users, and those who had visited more than once. More precisely, we compared self-reported levels of familiarity between first-time visitors and recurrent users. With 95 answers from first time visitors and 154 answers from recurrent visitors, a Wilcoxon test demonstrates that the familiarity level of recurrent users was significantly higher than first time visitors (p-value = 0.021). Therefore our hypothesis that users' familiarity increases over time was supported. We also limited the analysis to the most visited Web page in the site to remove any variance introduced by page dissimilarity. Lastly, data has been transformed using a logarithmic function to respect the necessary assumptions to apply Mixed Linear Models.

For each example, we present the results for each data set. The graphs shown in Figures 1, 3, 4 and 5, show the variance between different iterations of random sampling of users. We applied the analysis technique 50 times for each sample size and report the distribution of results. For each figure, sample sizes start at 10, and increase left to right by 10, until 120 or the maximum number of available users has been reached. In order to increase readability, the sample size legend has only been included in Figure 1. The most relevant aspect of these visualisations is the possibility of comparing if the result of applying the model to the entire population – last sample size, not showing any variance – differs drastically from using smaller samples.

We found that changes present in phase 1 were also found in phase 2. Some features were not found to have any temporal correlation, showing a normal distribution of results across the origin. For example, we found no change in the time from hovering over an element to clicking on it, while using the mouse.

Controlled scroll speed.

Figure 3 shows the occurrences of users scrolling down for a prolonged period of time at a controlled speed. As opposed to users scrolling down fast to the bottom of a Web page, this behaviour may occur when users are slowly scanning a Web page. We found that the speed of this scrolling action increased as users visited the page again.
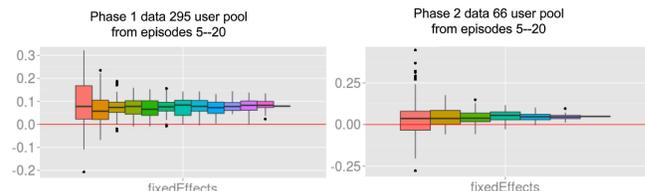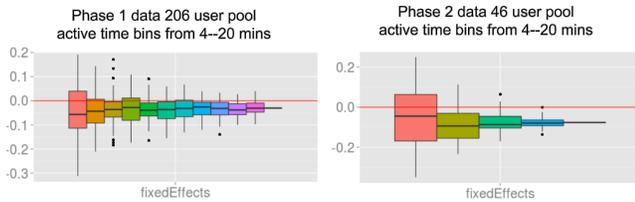

Figure 3: Evolution of the speed of controlled scroll over time in each phase.

Controlled scroll duration.

When examining the duration of the controlled scroll micro behaviour in Figure 4, we found that it was negatively correlated with users' active time. Taking into account the increasing speed of this micro behaviour (see Figure 3), users

may scroll faster, and for less time, to find what they were looking for in the Web page.

Figure 4: Evolution of the duration of controlled scroll over time in each phase.

Mouse inactive time.

For each episode, we extracted different micro behaviours concerning mouse interaction. In the example shown in Figure 5 we looked at the amount of time the user spent on a page without moving the mouse. In this case we analysed the median of all mouse inactive periods per episode. Our initial belief was that as they got more familiar with the Web page users would interact quicker, and take less time to decide what to do when presented with the page, resulting in shorter inactive times. Contrary to this initial belief, we found that users' mouse inactive time increased over time. This phenomenon could be the result of users not having to explore a page they are already familiar with – e.g. first-time users might want to hover over elements to determine what the interaction options are.
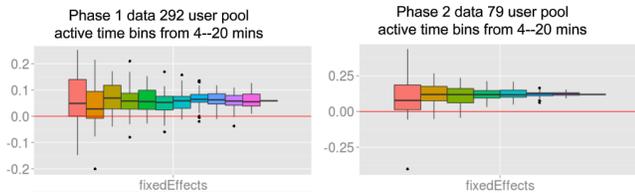
Figure 5: Evolution of the inactive time of mouse interaction over time in each phase.

Episode duration.

Figure 1 shows a positive correlation of episode duration with users' active time. This means that users' episodes get longer as they get more familiar with the page. Our initial belief was that users would need less time to navigate away from the page as their level of familiarity increases. One possible reason for this phenomenon is that users could be coming back to a page they already know to use it as reference. If the time between visits is smaller than the inactivity threshold set to separate episodes, both visits could be considered as part of the same episode. Therefore, rather than longer episodes, this tendency could be showing that the use of a reference page increases as the user becomes more familiar with it.

5. DISCUSSION

We present a technique that provides a means of identifying changes in behaviour that occur over long-term use of a Website from remotely captured Web interaction data. We tackle the problem of handling high levels of fine-grained data – such as mouse movement and scroll action – for extended periods of time, by grouping interaction data into *micro behaviours* representing particular aspects of user interaction. New micro behaviours can be included in the analysis, including data from different interfaces, or new interaction abstractions. Our contributions are threefold: we provide the means to achieve a **naturalistic observation of Web interaction**, remotely capturing data from

users' environments; we designed an **extensible longitudinal analysis technique**, enabling the addition of any new micro behaviour or feature *post hoc* in order to find new tendencies; and we present a **broadly applicable solution**, as it can be easily deployed in any Web site or application.

We present results of the application of our approach. Some of them supported our initial preconceptions about Web interaction, showing that the particular action of scrolling to scan the page becomes quicker over time. Other results challenged our initial beliefs. We found that users' interaction time with the mouse decreases as they become more familiar with the page. We also found that users' episode length within a Web page increases as they become more familiar – when using our interpretation of "episode".

6. ACKNOWLEDGEMENTS

7. REFERENCES

[1] D. Akers, M. Simpson, R. Jeffries, and T. Winograd. Undo and erase events as indicators of usability problems. In *Proceedings of the 27th international conference on Human factors in computing systems*, CHI '09, pages 659–668. ACM, 2009.

[2] A. Apaolaza, S. Harper, and C. Jay. Understanding users in the wild. In *Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility*, W4A '13, pages 13:1–13:4. ACM, 2013.

[3] R. Atterer, M. Wnuk, and A. Schmidt. Knowing the user's every move: user activity tracking for website usability evaluation and implicit interaction. In *Proceedings of the 15th international conference on World Wide Web*, WWW '06, pages 203–212, New York, NY, USA, 2006. ACM.

[4] D. M. Bates. lme4: Mixed-effects modeling with r. 2010.

[5] J. P. Bigham, A. C. Cavender, J. T. Brudvik, J. O. Wobbrock, and R. E. Lander. WebinSitu: a comparative analysis of blind and sighted browsing behavior. In *Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility*, Assets '07, pages 51–58. ACM, 2007.

[6] Google. How visits are calculated in analytics, 2014. [Online `https://support.google.com/analytics/answer/2731565?hl=en-GB`; accessed 8-April-2015].

[7] T. Grossman, G. Fitzmaurice, and R. Attar. A survey of software learnability: metrics, methodologies and guidelines. In *Proceedings of the 27th international conference on Human factors in computing systems*, CHI '09, pages 649–658. ACM. 00083.

[8] D. He and A. Göker. Detecting session boundaries from web user logs. In *Proceedings of the BCS-IRSG 22nd annual colloquium on information retrieval research*, pages 57–66, 2000.

[9] D. Lunn, S. Harper, and S. Bechhofer. Identifying behavioral strategies of visually impaired users to improve access to web content. *ACM Trans. Access. Comput.*, 3(4):13:1–13:35, 2011.

[10] M. Vigo and S. Harper. Considering people with disabilities as überusers for eliciting generalisable coping strategies on the web. In *Proceedings of the 5th Annual ACM Web Science Conference*, pages 441–444, 2013.

[11] J. Vlasenko. Exploring developer's tool path. In *Proceedings of the ACM international conference companion on Object oriented programming systems languages and applications companion*, SPLASH '11, pages 219–220. ACM, 2011.

[12] N. Zakay and D. G. Feitelson. On identifying user session boundaries in parallel workload logs. In W. Cirne, N. Desai, E. Frachtenberg, and U. Schwiegelshohn, editors, *Job Scheduling Strategies for Parallel Processing*, number 7698 in Lecture Notes in Computer Science, pages 216–234. Springer Berlin Heidelberg, 2013.