PhD Literature Review

Jonathan Carlton

School of Computer Science
University of Manchester
jonathan.carlton@postgrad.manchester.ac.uk

1 Introduction

2 Interaction

In [Atterer et al., 2006] a monitoring system for web-based interactions is defined – called UsaProxy. By requesting the users of the system to re-route all of their connections through a proxy server, HTML pages are modified with JavaScript tracking code before they are delivered to the user. The code collects data on mouse movements, keyboard input, along with other, fine-grained interaction metrics.

The capture solution presented above, in [Atterer et al., 2006], is modified in [Apaolaza et al., 2015] 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. Data; low-level mouse movements, clicks, and keystokes, in this experiement are recorded from a high-traffic website continously for two years. They find that users, rather than interacting with the website quicker as they become more familiar, have increased periods of mouse inactivity. Continually, the users also spend more time on the website as they become more familiar. And finally, they find that there is no need to collect specific information about users, such as any disabilities they may have, as their problems can be indentified through emerging behaviours in the experiements [Apaolaza et al., 2013].

Probematic situations encountered by users with visual impairments and the tactics they employ to overcome them are explored in [Vigo and Harper, 2013b]. Through developing several algorithms, and packaging them together into a web-usage monitoring tool, the employed tactics are identified and isolated automatically, in mouse and keyboard data, and treated as markers to infer the user is having an issue. In [Vigo and Harper, 2013a], more detail is presented about the particular type of tactics and an expansion on the analysis process by going deeper into the tactics the users employed and how they react to problems (do they give up or carry on).

WevQuery, a scalable system to query user interaction logs collected in [Apaolaza et al., 2015, Vigo and Harper, 2013b, Vigo and Harper, 2013a], is presented in [Apaolaza and Vigo, 2017]. It is a hypothesis testing interface with the aim of better understanding user behaviour stored in the data collected (mouse clicks, movement, etc.).

In [Gledson et al., 2016, Bull et al., 2016], a fully-fledged, desktop application with an aim of trying to detect mild cognitive impairment in older computer users, through their interactions with the computer, is presented. The monitoring system collects data on operating system events, web browsers, and applications. Furthermore, mouse movements are collected but the complexity is reduced by only recording dragging movements and the time periods between clicks. They have early evidence that this is a promising method to detect cognitive impairments.

The authors in [Kodagoda et al., 2017] are successfully able to infer the reasoning behind analysts' decision making from low-level user interaction logs. They detail some information about their data processsing techniques; converting features to numeric expressions – a fairly common practise – and a set of categorical labels to a set of integers which worked for Random Forest and Hidden Markov models (scale-invariant models) but not for an SVM, so the integers were standardised to have a mean of zero and a variance of one. To test their models, they used a control model (no information classifier) which provided the baseline for their models to beat.

In [Fu et al., 2017] the aim is to predict user intention from mouse movements, clicks, and positioning – among other features – with the focus on understanding what the user intended when they triggered an interaction and then, from this data, attempting to predict the next action. Two classification models are built; a probability model and a machine learning model (SVM) model. The probability model uses the previous k activities of a user to predict the next activity and a second (probability) model considers the time duration of the previous k activities. The SVM model is trained on the k most recent user activities, the time duration of those activities, and descriptive statisitics about the mouse interations within a set window W. In their results, they find that both model types have a similar performance in their experiments and to achieve the maximum accuracy – for predicting future interactions from historical actions – the combination of both models, into a multimodal approach, achieves the best results (a mean accuracy of 69%). With regards of variable results, they test a range of different values for both k and W and conclude that neither can be too large; the best value for k is 3 and when increasing the value of W, they find that it introduces a large amount of noise.

A system that can assist a user in an answer discovery task is presented in [Dabek and Caban, 2017]. Various options are presented to the user in the form of a graph, they are then able to traverse the graph to find an answer to their question. Throughout the process, an assistive tool provides suggestions on additional variables that could help their search (the number of bedrooms when searching for a house value, for example) and suggest a new path in their search. While not much can be taken from this paper, they make a general point about determining a k value in the k-means algorithm when there is some uncertainty about the optimal value to use; test a range of values between one and the square root of the size of the data divided by two. Adding this upper bound prevents over-fitting the value of k.

In [Gotz and Zhou, 2008] the authors attempt to understand a users' analytical goal and reasoning through high-level interaction metrics collected from a controlled study. They had two main aims in the study; determine common structures of visual analysis behaviour and characterise those structures with regards to what they represent and how they impact a users' performance. In their results and evaluation, they were able to find two behaviour structures of visual analysis activity: patterns (short, ad-hoc sequences of visual actions performed iteratively through the task) and trails (chains of user visual interaction activity leading to insight).

A new family of interactions to extend and enrich the input experience for users of smartphones are introducted in [Zhang et al., 2016]. Various additional inputs for a smartphone are presented, these include the ability to detect in-pocket and on-table touches through the detection of different physical interactions with the smartphone. The authors use both machine learning (classification) and rule-based (segementation) approaches to analyse characteristics of the interactions and determine the action the user is attempting to perform. The rule-based approach required no specific training for the users of the smartphone due to looking for particular patterns and spikes in the microphone recording, accelerometer and gyroscope on the phone, however, the machine learning method (kNN) did. They adopted a user-independent model for the implementation of the kNN, which applies a pre-collected training data set and does not require any data collection from each user. In their evaluation, they found that the machine learning approach achieved a range of results between 70% and 90% accuracy for the recognition of interactions and the rule-based interactions can achieve over 92% accuracy.

3 Engagement

A broad review of measuring and defining user engagement in a range of scenarios is presented in [Brown and Glancy, 2015]. The focus is on the understanding of initial reactions to media-based content and what engagement means in this context. They find that if the audience is emotionally invested in the content then their level of engagement is subsequently higher.

In [Jennett et al., 2008], the authors perform an investigation to test if immersion can be defined quantitatively through experiments. They devise three experiments; switching from an immersive to a non-immersive task, changes in eye movements during an immersive task, and measuring the effect of an externally imposed pace of interaction which alters the flow of the participant. Immersion is well defined here and has three features; lack of awareness of time, loss of awareness of the real world, and a sense of being in the task environment (a video game). They find that immersion can be measured subjectively, through questionnaires given to the participants before and after the event, and objectively, through task completion and eye movement. The authors apply Spearman's Rank-Order correlation on the mouse click data (the mean number of mouse clicks vs the mean number of fixations on the non-immersive conidition), relying heavily on this in their analysis.

4 Sequential Data Mining

A review of machine learning approaches, specific to a sequential data analysis context, is presented in [Dietterich, 2002]. There is a primary focus on text- and natural langauge-based solutions and the paper presents algorithms such as; sliding window methods, Hidden Markov models, and graph transformer networks. The point is made that these approaches are not exclusive to text or natural language solutions and have a place in other types of sequential data analysis. An example of these approaches being used in other domains is shown in [Kodagoda et al., 2017] where Hidden Markov models are used in an interaction data context.

Using annotated sensor data, collected in smart home environments, in [Carolis et al., 2015] the authors aimed to describe how process mining can be used to learn users' daily routines. Through utilizing a First-Order learning approach (symbolized reasoning in which each sentence, or statement, is broken down into a subject and predicate), they propose a solution that can automatically learn from daily routines from examples of other people's behaviour. They find, through experimentation with toy and publicly available datasets, that their approach is effectively able to learn and model the daily routines of people within smart home environments. In addition to their work, the authors provide definitions for three terms: activity recognition; aims at identifying the occurence of an acitivity based on suitable combinations of events, task modelling; aims at identifying which combinations of events determine the occurence of an activity, and process mining; starts from the activities performed by some agent to carry out a given process and identifies the valid patterns of activities that support that process. Finally, they also make the point that Hidden Markov model-based approaches are more suitable for representing sequential activities from a stream of sensor events – this backs up the point made previously about the various applications the models can have.

References

Apaolaza et al., 2013. Apaolaza, A., Harper, S., and Jay, C. (2013). Understanding users in the wild. In *Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility*, page 13. ACM.

Apaolaza et al., 2015. Apaolaza, A., Harper, S., and Jay, C. (2015). Longitudinal analysis of low-level web interaction through micro behaviours. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media*, pages 337–340. ACM.

Apaolaza and Vigo, 2017. Apaolaza, A. and Vigo, M. (2017). Wevquery: Testing hypotheses about web interaction patterns. *Proceedings of the ACM on Human-Computer Interaction*, 1(1):4.

Atterer et al., 2006. Atterer, R., Wnuk, M., and Schmidt, A. (2006). 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*, pages 203–212. ACM.

Brown and Glancy, 2015. Brown, A. and Glancy, M. (2015). Engagement: what is it, and how can we measure it? Internal BBC Research and Development Technical Note. Internal Access Only.

Bull et al., 2016. Bull, C., Asfiandy, D., Gledson, A., Mellor, J., Couth, S., Stringer, G., Rayson, P., Sutcliffe, A., Keane, J., Zeng, X., et al. (2016). Combining data mining and text mining for detection of early stage dementia: the sams framework. In *Proceedings of LREC 2016 Workshop. Resources and Processing of Linguistic and Extra-Linguistic Data from People with Various Forms of Cognitive/Psychiatric Impairments (RaPID-2016), Monday 23rd of May 2016*, number 128. Linköping University Electronic Press.

Carolis et al., 2015. Carolis, B. D., Ferilli, S., and Redavid, D. (2015). Incremental learning of daily routines as workflows in a smart home environment. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 4(4):20.

Dabek and Caban, 2017. Dabek, F. and Caban, J. J. (2017). A grammar-based approach for modeling user interactions and generating suggestions during the data exploration process. *IEEE transactions on visualization and computer graphics*, 23(1):41–50.

Dietterich, 2002. Dietterich, T. (2002). Machine learning for sequential data: A review. *Structural, syntactic, and statistical pattern recognition*, pages 227–246.

Fu et al., 2017. Fu, E. Y., Kwok, T. C., Wu, E. Y., Leong, H. V., Ngai, G., and Chan, S. C. (2017). Your mouse reveals your next activity: Towards predicting user intention from mouse interaction. In *Computer Software and Applications Conference (COMPSAC), 2017 IEEE 41st Annual*, volume 1, pages 869–874. IEEE.

Gledson et al., 2016. Gledson, A., Asfiandy, D., Mellor, J., Ba-Dhfari, T. O. F., Stringer, G., Couth, S., Burns, A., Leroi, I., Zeng, X., Keane, J., et al. (2016). Combining mouse and keyboard events with higher level desktop actions to detect mild cognitive impairment. In *Healthcare Informatics (ICHI), 2016 IEEE International Conference on*, pages 139–145. IEEE.

Gotz and Zhou, 2008. Gotz, D. and Zhou, M. X. (2008). An empirical study of user interaction behavior during visual analysis. *Technical Report RC24525*.

Jennett et al., 2008. Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., and Walton, A. (2008). Measuring and defining the experience of immersion in games. *International journal of human-computer studies*, 66(9):641–661.

Kodagoda et al., 2017. Kodagoda, N., Pontis, S., Simmie, D., Attfield, S., Wong, B. W., Blandford, A., and Hankin, C. (2017). Using machine learning to infer reasoning provenance from user interaction log data: Based on the data/frame theory of sensemaking. *Journal of Cognitive Engineering and Decision Making*, 11(1):23–41.

Vigo and Harper, 2013a. Vigo, M. and Harper, S. (2013a). Coping tactics employed by visually disabled users on the web. *International Journal of Human-Computer Studies*, 71(11):1013–1025.

Vigo and Harper, 2013b. Vigo, M. and Harper, S. (2013b). Evaluating accessibility-in-use. In *Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility*, page 7. ACM.

Zhang et al., 2016. Zhang, C., Guo, A., Zhang, D., Li, Y., Southern, C., Arriaga, R. I., and Abowd, G. D. (2016). Beyond the touchscreen: An exploration of extending interactions on commodity smartphones. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 6(2):16.