Inferring Visual Behaviour from User Interaction Data on a Medical Dashboard

Ainhoa Yera, Javier Muguerza, Olatz Arbelaitz, Iñigo Perona
Faculty of Informatics
University of the Basque Country
Donostia, Spain
{firstname.lastname}@ehu.eus

Richard Keers, Darren Ashcroft
Division of Pharmacy and Optometry
Greater Manchester Patient Safety Translational Research Centre
University of Manchester
Manchester, United Kingdom
{richard.keers|darren.ashcroft}@manchester.ac.uk

Richard Williams, Niels Peek
Division of Informatics, Imaging and Data Sciences
Greater Manchester Patient Safety Translational Research Centre
University of Manchester
Manchester, United Kingdom
{richard.williams2|niels.peek}@manchester.ac.uk

Caroline Jay, Markel Vigo*
School of Computer Science
University of Manchester
Manchester, United Kingdom
{caroline.jay|markel.vigo}@manchester.ac.uk

ABSTRACT

Making medical software easy to use and actionable is challenging due to the characteristics of the data (its size and complexity) and its context of use. This results in user interfaces with a high-density of data that do not support optimal decision-making by clinicians. Anecdotal evidence indicates that clinicians demand *the right amount of information* to carry out their tasks. This suggests that adaptive user interfaces could be employed in order to cater for the information needs of the users and tackle information overload. Yet, since these information needs may vary, it is necessary first to identify and prioritise them, before implementing adaptations to the user interface. As gaze has long been known to be an indicator of interest, eye tracking allows us to unobtrusively observe where the users are looking, but it is not practical to use in a deployed system. Here, we address the question of whether we can infer visual behaviour on a medication safety dashboard through user interaction data. Our findings suggest that, there is indeed a relationship between the use of the mouse (in terms of clickstreams and mouse hovers) and visual behaviour in terms of cognitive load. We discuss the implications of this finding for the design of adaptive medical dashboards.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**;
• **Applied computing** → **Health care information systems**;

---

*Corresponding author.

KEYWORDS

Medical dashboards; interaction analysis; user modelling

1 INTRODUCTION

The responsibility of health care professionals (especially General Practitioners) is shifting from a reactive patient-by-patient role to a proactive manager of population health. This shift requires the availability of health data and information tools that give a population-level view of such data, which allow the identification of individual patients that require intervention. Consequently, the use of medical dashboards is becoming increasingly important in using this data to improve healthcare. While the current wealth of clinical data satisfies the availability premise, it becomes, at the same time, a double-edged sword in that medical dashboards suffer from information overload. What is more, clinicians have varying levels of practical clinical experience, different problem-solving skills, and vary considerably in their IT skills. As the information density in the clinical environment is increasing rapidly and the role of medical dashboards is still at an early stage, it is of paramount importance to build smart adaptive systems that cater for the needs of clinicians and support them in the transition towards a proactive management of population health.

Before we start building smart adaptive systems that cater for the information needs of the users, we need to understand what these information needs are and whether they can be unobtrusively detected while the users interact with medical dashboards. As gaze is well known to be an indicator of interest [4], in this paper we explore if we can approximate gaze through the interactive behaviour of users including mouse clicks and mouse hovers. Previous work has explored this relationship and has found that
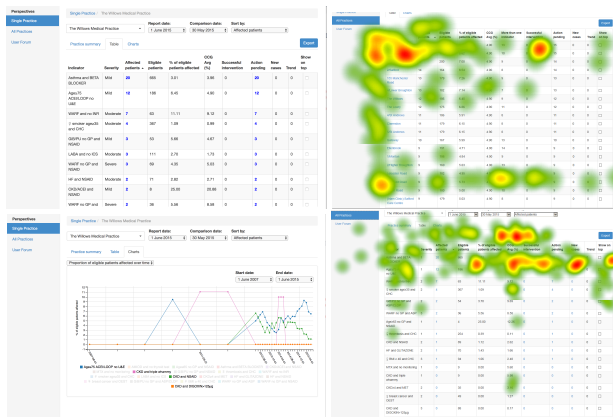
Figure 1: A sample of SMASH views: the table view (top-left) and visualisations (bottom-left). On the right, the heatmap patterns generated on the pilot study

Figure 2: AOIs defined in SMASH. The remaining three AOIs correspond to pop-up dialogues (2) and charts (1)

mouse and gaze are strongly related: dwell times on specific regions are correlated with the likelihood of visiting that region with the mouse [3]. In gaze prediction models for search engine results pages (SERPs), the inclusion of the mouse coordinates, the velocity and direction of the cursor, and the time elapsed since starting to view the results achieves an accuracy of 77% [6]. However, unlike websites and SERPs, medical dashboards are constrained by grid layouts where data is displayed in a tabular fashion, which determines the variability of behaviours that can be exhibited. Given these constraints, we address the following research question: can we determine the users' visual behaviour based on their exhibited interactive behaviour?

We examine this question on the Salford Medication Safety Dashboard (SMASH), one of two vital components of a pharmacist-led information technology intervention for safe prescribing of medications in primary care, from which we collect user interaction data and whose purpose and functionalities we describe later in Section 2.1. First, we ran a user study with six clinicians in the lab where they carried out typical tasks in SMASH while gaze data was collected with an eye-tracker. Using exploratory data mining techniques we clustered the participants based on the collected interactive behaviour and we employed inferential statistics to find relationships between their visual behaviour. Second, we applied the same clustering analysis on the interaction data by adding the logged interaction of 35 clinicians in SMASH on a ten-month observational study. The ultimate objective was to ascertain whether we could reliably extrapolate the lab findings to a setting where no eye-tracking device is deployed.

2 USER STUDIES

We ran one pilot study and two user studies. The pilot study (N = 5) informed the definition of the areas of interest on the SMASH user interface prior to the laboratory study (N = 6), where we collected fixation data generated by the eye-tracker as well as user interaction data. In the observational study (N = 35) that ran for a period of 10 months we collected user interaction data. Next we describe the platform, the data we collected and the metrics we computed.

2.1 Stimulus: the SMASH Dashboard

SMASH is a web dashboard that highlights patients at risk of potentially hazardous medication safety practices in general practices [8]. SMASH incorporates a refined set of 24 prescribing indicators from the pharmacist-led information-technology based intervention (PINCER) [2]. It displays summary statistics for each of the indicators, counting how many patients are currently at risk in a given practice and relating those numbers to previous episodes and other practices. In addition, pharmacists and GPs can look up which patients are currently at risk for each indicator. The dashboard is being deployed in Salford, a city in the Greater Manchester conurbation, comprising a population of 250,000 served by primary care that uses two of the three largest GP computer systems in the UK, with additional linkage to secondary care records. The user interface of SMASH is divided into seven screens or views: (1) a landing page containing a summary of a given practice, (2) a list of patients who are affected by more than one indicator, (3) a table view displaying the indicators and (4) the graph-based visualisations of the indicators. When clicking on the number of patients affected by a given indicator on the table view, further screens are revealed including (5) the list of patients at risk for this specific indicator, (6) the trends for this indicator over time and (7) information about the selected indicator.

Figures 1–2 show the table view, which display a data table containing the number of patients who are affected by the indicators, their severity, the number of eligible patients and the percentage of patients who are affected. Indicators can be contraindications between drugs and conditions (e.g. Chronic Kidney Disease and Non-Steroidal Anti-Inflammatory drugs) or between drugs, habits and demographics. The 'Selection menu' (see Figure 2) allows the user to choose a different view of the data: the landing page, which is a small table containing the size of the practice and the number of patients affected by more than one indicator, and the 'Visualisations' view (see Figure 1, bottom-left), which is the visual representation of the table and displays the incidence of indicators as time-series through line graphs. The 'Selection menu' also contains calendar widgets for selecting reports within specific time periods.

2.2 Apparatus

The SMASH dashboard logs the user interface events triggered by the users in a database on the server. Because SMASH is a mouse-driven application the collected events are mostly mouse clicks and mouse hovers. A third event logged is the page load event, which signals navigation to a different view (e.g. from the data table to the data visualisation) that does not necessarily entail an update in the URL and is triggered by clicking on the 'Selection menu'. For each event, SMASH collects the user id, the identifier of the session (i.e. every time a user logs in, a new session is established), the timestamp, the URL where the event took place and the specific element on the user interface where the event occurred indicated by an XPATH statement.

The Tobii X2-60 eye-tracker we employed in the laboratory study logged gaze information including fixation coordinates on the screen, duration of the fixations and the saccades (movement of eyes between the fixations).

2.3 Metrics and Data Analysis Method

The gaze activity was computed using the average fixation duration (henceforth fd) in the nine Areas Of Interest (AOIs) that were defined as a consequence of the pilot study described in Section 2.5, $G = \{fd_i \mid i \in \mathbb{N}, i \leq 9\}$. Fixation duration is known to be a proxy for cognitive load [4] so our premise is that, if we want to relate visual behaviour to interactive behaviour on SMASH, cognitive load might well be an indicator to profile participants. We run Pearson correlation analysis between the G vectors in order to find participants with similar fds across the different AOIs. Consequently, a positive correlation between any two participants would entail similar visual behaviours in terms of cognitive load.

Then, we computed interaction metrics for *exploration* and *pace*.

- Exploration (e): median of the number of mouse hovers between two consecutive mouse clicks. This is based on the fact that since mouse location on screen is a proxy of gaze location [5], it can be used to quantify visual exploration.
- Pace (p): median of the elapsed time between two consecutive mouse clicks.

Using these metrics, we created two vectors per participant: $V1$ describes user interaction on all of the screens available and is computed as a vector of two attributes, the overall exploration and pace on SMASH, $V1 = \{(e, p)\}$. The second metric, $V2$, takes into consideration the above metrics in each of the seven screens of SMASH and is represented as a vector of 14 attributes per participant, $V2 = \{(e_i, p_i) \mid i \in \mathbb{N}, i \leq 7\}$. Then we applied different clustering algorithms including k-means and single-linkage method [7] to the two vectors we defined above in order to identify those users who exhibited similar interactive behaviours.

2.4 Tasks

In the pilot and laboratory study participants were assigned a specific practice and were asked to complete nine tasks classified in three ways: a) Identification of patients at risk: i.e. 'List up to three patients at risk for indicator X'; b) Identification of problems in the practice and their evolution over time: i.e. 'Identify the three indicators with the largest number of patients affected'; c) Comparison of problems between practices: i.e. 'Identify three indicators in which your practice performs worse than others'.

In the observational study no tasks were given to the participants since the SMASH dashboard was used for the purpose it was intended: the identification of those patients at risk and the promotion of good prescription practices. We expected that the participants in this study would carry out tasks of a higher ecological validity than those given in the lab setting.

2.5 Pilot Study

Five participants (3 female) who were 39 years old on average (SD = 13.5 and age range = 27–62) took part in the pilot study. All of them were computer savvy and familiar with the domain and terminology of the medication safety dashboard. Two of them were members of the Research User Group, a pool of users who frequently take part in eHealth studies, and of the remaining three: one had a degree in nursing, another one was doing a PhD in nursing and one was a medical microbiologist.

A qualitative analysis of the heatmaps of the pilot study yielded some interesting insights: the visual search strategies on the dashboard followed particular patterns. The C-shaped behaviour in Figure 1 suggests that users look at the data header, the list of indicators on the left and the values in a row belonging to a particular indicator. On the other hand, the paint drop pattern in Figure 1 indicates that users look at the header and the top rows and visual search is restricted to a few columns. This strategy can be explained by the fact that some users discovered that clicking on a header sorted the indicators based on the values of the corresponding column/variable, which was an effective strategy for completing many tasks and reduced the need for visual exploration. While the boundaries between the components of the dashboard are clear, it is difficult to establish the AOIs in a tabular environment. The patterns we found as well as the demarcation of existing user interface elements informed the design of the areas of interest, accounting for nine of them. As depicted by Figure 2 our findings suggest that the column containing the safety indicators, the table header and the remaining rows should constitute independent areas of interest.

2.6 Participants

Six participants (4 male) took part in the lab study. Five were GPs and one was a pharmacist; their average age was 38 (SD = 10 and age range = 30–56). In the observational study we collected interaction data belonging to 35 individuals who participated over a period of 10 months. Since data logging occurs unobtrusively no demographic data was collected for this study. These 35 participants are divided into the roles they play on the National Health Service: 10 pharmacists, 8 members of the Salford Clinical Commissioning Group, 8 GPs, 5 managers, and 4 other including nurses and technicians.

3 RESULTS

### 3.1 Interaction Data Analysis in the Lab

The k-means algorithm (k = 3 and Euclidean distance) clustered clinicians in three pair wise groups for $V1$: P4-P5, P1-P3 and P2-P6. To better understand the structure of these groups we carried out a second clustering procedure including the centroids of the

clusters generated by the k-means algorithm. We then computed the distance matrix for $V1$ and then calculated the centroids using the Euclidean distance again. The resulting distance matrix can be visualized using hierarchical clustering techniques [7]. Thus, we applied the single-linkage and Ward hierarchical clustering algorithms using the Euclidean distance. Figure 3a shows the resulting dendogram for the single-linkage clustering procedure, where the height at which participants are grouped represents the distance between clusters. The arrangement of clusters in the dendogram shows three main branches that group the participants and the centroids together: P4-P5-C1, P1-P3-C2 and P2-P6-C3 which indicates that the groups discovered by k-means applied to the overall interactive behaviour of the participants remain stable.

(a) $V1$ **(b)** $V2$
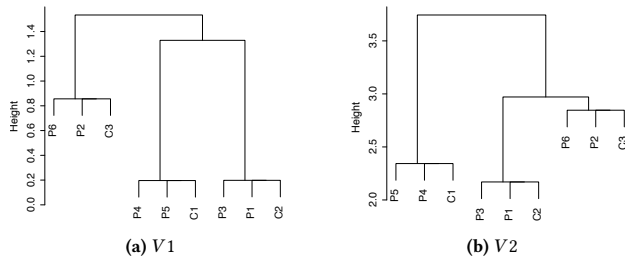
Figure 3: Single-linkage algorithm dendogram for the distance matrix and the computed centroids.

Regarding the closeness or similarity of the patterns, it can be observed that the groups of P1-P3 and P4-P5 are more compact since they are placed at the bottom of the dendogram. To better show the proximity of the six participants, we performed a more exhaustive study of the overall interaction analysis and used the neighbour-joining tree estimation of Saitou and Nei [10] over the distance matrix of $V1$, excluding the centroids. In the resulting tree shown in Figure 4, it can be seen that P2 and P6 are some distance from the remaining participants, and P6 in particular is further than any other. This suggests that the cluster of groups P2 and P6 was not as compact as the other ones.
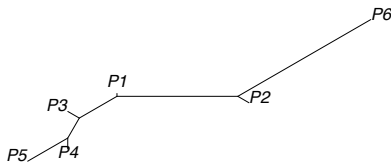
Figure 4: Neighbour joining method for the distance matrix of $V1$

We applied the same pattern discovery method on $V2$, that is, we ran the k-means algorithm (k = 3) using the Euclidean distance[1], then computed the centroids of the resulting clusters. Again, we computed the distance matrix for $V2$ and the centroids calculated in the previous step, using the Euclidean distance. Accordingly, the size of the distance matrix was of 9 x 9 measuring the distance based similarity of the six participants and the three centroids (C1-C3). Finally, we studied the proximity of the patterns discovered in $V2$, using the single-linkage algorithm [7] in the computed distance

matrix. Figure 3b shows the visual output of the single-linkage algorithm and illustrates how participants were distributed in the same form for the two clustering procedures used in $V2$, k-means and hierarchical clustering. These patterns matched the ones found for $V1$ when using the same clustering procedures. When we analysed the proximity of the patterns represented in Figure 4, we found again that P2 and P6 are quite far from the other two groups.

3.2 Eye-Tracking Data Analysis in the Lab

In order to discover groups of participants with similar visual behaviours, we paired those participants using the highest value for the Pearson correlation computed in each case – note that data was normally distributed according to the Shapiro-Wilk test ($p > 0.05$). As a result, we identified three groups (with their corresponding fixation duration patterns), which paired P1 and P3 ($\rho = 0.63, p = 0.06$), another one pairing P2 and P6 ($\rho = 0.55, p = 0.11$) and a last one pairing P4 and P5 ($\rho = 0.53, p = 0.13$). It is well known that p values are sensitive to the sample size. Since the G vectors contain 9 items, an alpha value < 0.95 is justifiable so we can say that the moderate-high correlations found show a clear tendency towards significance.

The members of the resulting groups match with the results of the cluster analysis on user interaction data. The emerging gaze patterns and the patterns discovered on user interaction data (see Figure 3) generate the same groupings. That is, those individuals having similar interactive behaviour happen to have related visual behaviour. Specifically, individuals with a similar cognitive load (as indicated by fixation durations) exhibit similar mouse use as captured by the exploration and pace metrics on SMASH ($V1$) and on its seven views ($V2$).

3.3 Interaction Data Analysis in the Observational Study

We carried out the analysis of user interaction data including the data of the lab participants and that of those who took part in the observational study. The purpose of analysing the two datasets together was to ascertain whether the emerging clusters would include the six laboratory participants in the same pairs. If the pairs of users fell again in the same clusters we could speculate that those participants belonging to the same cluster would have similar search behaviour to their lab counterparts. We therefore re-run our analysis (i.e. k-means and Euclidean distance) on $V1$, which this time accounted for 41 participants (i.e. 35 from the observational study + 6 from the lab).

Table 1 below shows the distribution of the lab study participants in the generated clusters[2]. The results indicate that the six laboratory participants are grouped in a similar way. P2 was the only participant falling in a different group, as instead of belonging to the same pair as P6, they were a member of P1 and P3's group. Taking into account the proximity analysis carried out for $V1$ (see Figure 4) this finding was not surprising given that P2 and P6's clustering was unstable. Hence, the fact that P2 switched groups would be understandable.

---

[1]A silhouette analysis on Cluster Validity Indexes [1] indicates that k = 3 is the most appropriate cluster configuration when compared to k = 4 and k = 5. Scores are 0.51, 0.20 and 0.003 respectively for $V1$, and 0.002, -0.011 and -0.046 for $V2$.

[2]A silhouette analysis on Cluster Validity Indexes indicates, again, that k = 3 is the most appropriate cluster configuration for all ks, where $3 \leq k \leq 10$.

Table 1: Results of k-means (k = 3, d = Euclidean) for $V1$ when merging the participants of the lab and the observational study

	Clusters		
	1	2	3
Lab participants	P6	P4, P5	P1, P2, P3
Number of participants from the observational study	2	19	14

4 DISCUSSION AND CONCLUSIONS

The clustering analysis on the vectors containing exploration and pace metrics generated clusters with the same members when the analysis considered the whole platform ($V1$) and when the screens were factored in the analysis ($V2$). This may mean that, within the generated clusters, the users' interactive behaviour was similar across all views of SMASH. Alternatively, if the behaviour was different (i.e. a specific view led to a different behaviour) this also changed similarly across the participants within the clusters. There are implications as far as user modelling is concerned in that this finding suggests that incorporating the specific view in the user model may not make any difference to the way pace and exploration metrics are used in the model.

The analysis of the fixation duration in the nine areas of interest of SMASH yields the same groups as the ones generated by the analysis of user interaction. This indicates that those participants who exhibited a particular interactive behaviour in terms of pace and exploration had a similar visual behaviour in terms of fixation duration on the AOIs. Prior work suggests that gaze is a proxy of attention and, at the same time, attention precedes action: this relationship is transitive and the delay from gaze to cursor action is between 250–700 milliseconds [6]. Therefore we can say that these groupings are not incidental and the exhibition of particular interactive behaviours might be determined, to some extent, by the duration of fixations on specific areas of interest.

We evaluated the robustness of the laboratory analysis by including the interaction data of another 35 participants who used SMASH as part of their daily activities. The results show that the resulting clusters are stable across the two settings despite the fact that different tasks were conducted. This adds further evidence to the initial research question about whether we can infer visual behaviour from interactive behaviour. Without having gaze data from the participants of the observational study, we can hypothesise that their visual behaviour might be similar to that of the laboratory participants who were grouped in their respective clusters. Future work will certainly pursue this lead.

This has promising consequences in that visual behaviour could be inferred using interactive data alone. Interaction data analysis can be carried out in real time in the browser. The processing and analysis of interaction data is straightforward –provided that the right metrics are being monitored– and the deployment of eye-tracking devices beyond laboratory settings is not to be expected in the medium term. Since we found that there is a relationship between exploration and pace, and fixation durations (and, consequently, cognitive load), these metrics can be used 1) to infer usability problems in real-time and 2) to inform adaptations on the user interface. These adaptations could be implemented as views on the data to reduce the information overload of medical dashboards. This could be accomplished by enabling the direct manipulation of data tables [11] and the use of transient visualisations [9].

REFERENCES

[1] Olatz Arbelaitz, Ibai Gurrutxaga, Javier Muguerza, Jesús M. Pérez, and Inigo Perona. 2013. An extensive comparative study of cluster validity indices. *Pattern Recognition* 46, 1 (2013), 243 – 256. https://doi.org/10.1016/j.patcog.2012.07.021

[2] Anthony J Avery, Sarah Rodgers, Judith A Cantrill, Sarah Armstrong, Kathrin Cresswell, Martin Eden, Rachel A Elliott, Rachel Howard, Denise Kendrick, Caroline J Morris, et al. 2012. A pharmacist-led information technology intervention for medication errors (PINCER): a multicentre, cluster randomised, controlled trial and cost-effectiveness analysis. *The Lancet* 379, 9823 (2012), 1310–1319.

[3] Mon Chu Chen, John R. Anderson, and Myeong Ho Sohn. 2001. What Can a Mouse Cursor Tell Us More?: Correlation of Eye/Mouse Movements on Web Browsing. In *CHI '01 Extended Abstracts on Human Factors in Computing Systems (CHI EA '01)*. ACM, New York, NY, USA, 281–282. https://doi.org/10.1145/634067.634234

[4] Claudia Ehmke and Stephanie Wilson. 2007. Identifying Web Usability Problems from Eye-tracking Data. In *Proceedings of the 21st British HCI Group Annual Conference on People and Computers (BCS-HCI '07)*. British Computer Society, Swinton, UK, 119–128.

[5] Qi Guo and Eugene Agichtein. 2010. Towards Predicting Web Searcher Gaze Position from Mouse Movements. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems (CHI EA '10)*. ACM, New York, NY, USA, 3601–3606. https://doi.org/10.1145/1753846.1754025

[6] Jeff Huang, Ryen White, and Georg Buscher. 2012. User See, User Point: Gaze and Cursor Alignment in Web Search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 1341–1350. https://doi.org/10.1145/2207676.2208591

[7] Anil K. Jain and Richard C. Dubes. 1988. *Algorithms for clustering data.* Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

[8] RN Keers, R Williams, C Davies, N Peek, and DM Ashcroft. 2015. Improving medication safety in primary care: developing a stakeholder-centred electronic prescribing safety indicator dashboard. *Pharmacoepidemiology and Drug Safety* (2015). https://doi.org/10.1002/pds.3812 Online Supp.

[9] Charles Perin, Romain Vuillemot, and Jean-Daniel Fekete. 2014. A Table!: Improving Temporal Navigation in Soccer Ranking Tables. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 887–896. https://doi.org/10.1145/2556288.2557379

[10] N Saitou and M Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4, 4 (1987), 406–425. https://doi.org/10.1093/oxfordjournals.molbev.a040454

[11] Romain Vuillemot and Charles Perin. 2015. Investigating the Direct Manipulation of Ranking Tables for Time Navigation. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 2703–2706. https://doi.org/10.1145/2702123.2702237