

Measurement and Modeling of Eye-mouse Behavior in the Presence of Nonlinear Page Layouts

Vidhya Navalpakkam[†]

LaDawn Jentzsch[†]

Rory Sayres[‡]

Sujith Ravi[†]

Amr Ahmed[†]

Alex Smola^{†‡}

[†]Google, Mountain View, CA
{vidhyan,ladawn,sayres,sravi,amra}@google.com

[‡]Carnegie Mellon University, Pittsburgh, PA
alex@smola.org

ABSTRACT

As search pages are becoming increasingly complex, with images and nonlinear page layouts, understanding how users examine the page is important. We present a lab study on the effect of a rich informational panel to the right of the search result column, on eye and mouse behavior. Using eye and mouse data, we show that the flow of user attention on nonlinear page layouts is different from the widely believed top-down linear examination order of search results. We further demonstrate that the mouse, like the eye, is sensitive to two key attributes of page elements – their position (layout), and their relevance to the user’s task. We identify mouse measures that are strongly correlated with eye movements, and develop models to predict user attention (eye gaze) from mouse activity. These findings show that mouse tracking can be used to infer user attention and information flow patterns on search pages. Potential applications include ranking, search page optimization, and UI evaluation.

Categories and Subject Descriptors

H.4.m [Informations Systems Applications]: Miscellaneous

General Terms

Design, Experimentation, Human Factors

Keywords

eye, mouse, web search, attention, measurement, prediction

1. INTRODUCTION

A decade ago, search pages were simple, text only and contained a linear listing of documents. Today, search pages are increasingly complex, with interactive elements, images and text in multiple colors, font sizes, and varying indentation; they include new multi-column layouts, and contain various page elements drawn from news, images, documents, maps and facts. *With multiple page elements competing for the user’s attention, and attention being a limited resource, understanding which page elements get more or less atten-*

tion is important, and has applications for ranking, search page optimization, and UI evaluation.

Previous studies of attention on search pages focused on the linear page layout (containing a single column of search results) and showed a Golden Triangle [14] of user attention¹ – where users pay most attention to the top-left of the page, and attention decreases as we move towards the right or bottom of the page. Related studies showed that users tend to scan the search page sequentially from top to bottom, giving rise to popular cascade models and their variants [6, 5]. Given that the search pages have become more complex since (both visually and content-wise), the question of whether the Golden Triangle and other previous findings on attention still hold is open.

Eye tracking has been the favored methodology for studying user attention on the web [9, 8, 7, 2, 3, 20]. It offers rich details on user attention by sampling eye gaze positions every 20ms or more frequently, and providing fairly accurate estimates of user eye gaze (<0.5-1° error, just a few pixels). On the flip side, commercial eye trackers are expensive (\$15K upwards per piece), eye tracking is not scalable (typically performed in the lab with 10-30 participants), and it is not clear to what extent findings from eye tracking studies in controlled lab settings can generalize to user attention in the wild.

Recently, researchers have begun exploring whether a user’s mouse activity can provide approximations to where the user is looking on the search results page. Previous studies have shown reasonable correlations between the eye and mouse for linear page layouts²[22, 12, 15]. In this paper, we conduct a lab study to test whether mouse tracks can be used to infer user attention on complex search pages with nonlinear page layouts. In particular, we test eye-mouse sensitivity to an element’s position on the search page, and its relevance to the user’s task. The main contributions of this paper are outlined below:

1. We present a lab study to test eye-mouse activity on linear search page layouts (containing one column of search results) and new nonlinear search page layouts

¹in this paper, we use the term “user attention” to refer to those aspects of users’ attention that can be *measured by the eye*. Note that attention itself is a more complex, cognitive process.

²We ignore ads in this study, and focus only on search results. Thus, linear layout here refers to a single column of search results.

(containing a rich informational panel on the top-right of the page), and demonstrate that the mouse, like the eye, is sensitive to the element’s position on the search page in both layouts.

2. We demonstrate that both eye and mouse are sensitive to the element’s relevance to the user’s task.
3. We identify mouse measures that are most correlated with eye gaze.
4. We develop models that predict users’ eye gaze reasonably well from their mouse activity (67% accuracy in predicting the fixated result element, with an error of upto one element).
5. We conclude with limitations of mouse tracking and why it may be a weak proxy for eye tracking, but cannot substitute it.

2. RELATED WORK

2.1 Relationship of eye and mouse signals

The relationship of eye and mouse movements have been explored both in lab studies and at scale. Rodden and colleagues [22] measured eye and mouse movements of 32 users performing search tasks in a lab setting. They identified multiple patterns of eye-mouse coordination, including the mouse following the eye in the x and y directions, marking a result, and remaining stationary while the eye inspected results. They found a general coordination between eye and mouse position, where the distribution of eye/mouse distances centered close to 0 pixels in both x and y directions.

Huang et al [15] extended these findings by examining variations in eye-mouse distance over time. They found that eye-mouse distances peaked around 600 ms after page load and decreased over time, and that the mouse tended to lag gaze by ≈ 700 ms on average. They classified cursor behaviors into discrete patterns – Inactive, Reading, Action, and Examining – and measured sizeable differences in mouse-cursor position and time spent engaging in different behaviors.

Because mouse-position signals can be collected at scale more readily than eye movements, recent work has focused on relating large-scale mouse signals to eye movements. One approach proposed by Lagun and Agichtein (“ViewSer”; [18]) involves presenting a search result page in which all elements are spatially blurred except the result containing the mouse cursor. The resulting patterns of mouse movement across results were found to correlate with eye tracking results obtained in a lab setting for the same queries.

Other work from Huang, Buscher and colleagues [17, 16, 4] compare eye and mouse tracking results from lab studies to large-scale logs data from a search engine, deployed both internally [17] and on an external sample of users [16, 4]. This work demonstrated that mouse-based data can be used to evaluate search result relevance, distinguish cases of “good” (user need-satisfying) and “bad” abandonment on web pages, and identify clusters of distinct search task strategies.

2.2 Predictive models

Several studies have developed predictive models of user attention, searcher behavior, or both, based on mouse data. Guo and Agichtein [10] collected mouse-movement data from searches at a university library, and were able to discriminate navigational from informational task intent. The authors also built a binary classifier to distinguish whether the

mouse was within a specified radius from the eye (varied from 100 to 200 pixels), and showed that this model outperformed a simple baseline model which always guessed the majority category [12]. Huang et al [15] used eye and mouse movement data from a lab study to fit eye positions using a linear model based on extracted mouse interaction features. Their model demonstrated an improved eye gaze prediction (decreased RMS error in mouse-cursor distance in each direction, and Euclidean distance) over mouse data alone.

Mouse behavior has been used to model patterns of user search goals and strategies. Huang et al [16] incorporated mouse data from search logs into a Dynamic Bayesian Network model of searcher activity, using the positions of results that were hovered over but not clicked to provide a more robust measure of which results were evaluated. The searcher model incorporating these signals performed better (lower click perplexity metrics, measuring unexpected click patterns) than models without the signals. Guo and Agichtein [11] used mouse movements to classify hidden states representing searchers’ search goals (researching versus conducting a purchase) or ad receptiveness (likely or unlikely to click on a relevant ad) and tested the performance of their model against data extracted from a user study, and user-provided data from a library system. The authors also developed a model of document relevance based on a combination of mouse activity on the search result page, and on subsequent post-search pages [13].

2.3 Differences from our work

To summarize, previous research on attention and eye-mouse behavior in search focused on linear page layouts containing a single column of search results, and demonstrated mouse sensitivity to position on page. Our work differs from previous work in at least 3 ways: 1) In addition to linear page layouts, we explore nonlinear 2-column results layouts, with a rich information panel on the right hand side of the page; 2) Apart from position on page, we test whether the mouse is sensitive to important factors such as element’s relevance to user’s task; 3) We systematically compare various user-specific, global and hybrid models for predicting user eye gaze based on mouse signals.

3. EXPERIMENT DESIGN

We recruited 31 participants (15 male and 16 female; age range 18-65, with a variety of occupations and self-reported web search expertise). Data from 5 participants was excluded due to calibration problems with the eye tracker.

Participants were given 4 warm-up tasks to familiarize themselves with the testing procedure. They were then asked to describe all of the elements of the search results page to ensure they were aware of all elements on the page, including the relatively new Knowledge Graph (KG) results feature on the top-right of the page.³ For example, a query on an entity such as a famous celebrity or place triggers a result page with an informational panel (KG) on the right hand side, containing images and facts about that entity.

Participants were provided a list of 24 search tasks. Each task was accompanied by a link to the search results page with a prespecified query (to ensure that all participants saw

³<http://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html>

the same page). For consistency in results sets across users, users could not change the queries. Blocks of images and advertisements were suppressed on the results pages.

We used a 2 x 2 within subject design with two factors: (1) KG present or absent, (2) KG relevant or irrelevant to the user’s task. We varied task relevance of KG by designing 2 tasks for the same query, one for which the answer was present in KG (e.g., "find the names of Julia Robert’s children") and another for which the answer was present in web results, but not KG (e.g. "what does People magazine say about Julia Roberts?"). Each user performed an equal number of tasks with KG present and absent, as well as with KG relevant and irrelevant. The order of tasks, KG presence, and KG task relevance were all randomized within and across users.⁴

For the purposes for this study, we injected custom javascript in search results pages to log mouse activity on search page as users performed their search tasks. In particular, we logged the browser viewport, and sampled mouse x,y positions every 20ms during mouse movements. Eye tracks were simultaneously recorded using a Tobii TX300 eye tracker with 300 Hz tracking frequency and an accuracy of 0.5° visual angle on a 23" monitor (1880 x 1000 browser size). Both eye and mouse tracks were recorded with an origin at the top-left of the document.

We considered data upto the first click on the search page, or until the user terminated the task (by clicking the back-button to revisit the task list). Raw eye and mouse tracks were of the form: <user, task, timestamp, x, y>. The raw eye data was parsed to obtain a sequence of fixations (brief pauses in eye position for around 200-500ms) and saccades (sudden jumps in eye position) using standard algorithms [21, 23]. Eye fixations and their duration are thought to indicate attention and information processing [8]. Thus, our subsequent analysis is performed using eye fixations data. We aligned eye fixation and mouse data using time from page load. Since eye fixations occur every 200-500ms, while mouse movements were logged every 20ms, we aligned them per user, task by assigning the most recently fixated eye position to each mouse event.

4. RESULTS

4.1 Correlations in pixel space

To test eye-mouse correlations in pixel space, we extract the following measures per user, per task: x,y positions of eye and mouse, maximum x, maximum y, minimum x, and minimum y. The scatter plot in figure 1 shows that the maximum y position shows reasonable correlation ($r^2=0.44$). Other pixel measures such as instantaneous x,y positions of eye-mouse, or minimum x, maximum x, and minimum y show poor correlation between eye and mouse ($r^2<0.05$).

4.2 Correlations in Area-of-Interest space

The pixel-level measures described above include eye-mouse activity in white space, and other elements that are not web results (e.g., search bar, navigation menu). Since we mainly care about which result element the user is looking at, we

⁴Another example of a query from the study, along with tasks are: For search [hope floats]: "Who directed the movie 'Hope Floats'" (answer in KG); "What does the Hope Floats wellness center do?" (answer not in KG)

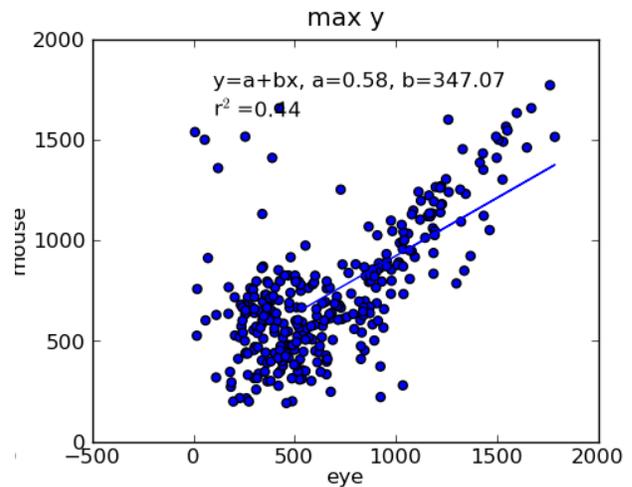


Figure 1: Eye and mouse show correlations in the maximum cursor distance along the vertical axis (y). Each point in the scatter plot denotes a user, task combination.

proceed to analyze the data by defining meaningful areas-of-interest on the search page. We divide the page into Knowledge Graph results (KG), and divide result elements into 3 roughly equal-sized parts (so that their size is similar to KG): Top 3 results (top), middle 3 results (mid), bottom 4 results (bot). In addition, we consider the left navigation menu (left), and group everything else under miscellaneous (misc). Thus, we define 6 areas-of-interest (AoI). The bounding boxes of these AoIs are illustrated in Figure 3, which shows a heatmap visualization of user eye gaze when KG is present and absent.

To analyze the data quantitatively, for each AoI, we extract the following mouse measures: 1) #mouse hovers or visits to the AoI, 2) time to first mouse visit on the AoI (in milliseconds), 3) dwell time (in milliseconds) per mouse position within the AoI, 4) total mouse dwell time (in milliseconds) in the AoI, 5) fraction of page dwell on AoI, 6) fraction of tasks where the last mouse position occurs within the AoI. We also extract corresponding measures for the eye.

Figure 2 shows eye-mouse correlations in AoI space. As seen in the figure, the fraction of page dwell on AoI (dwell time within AoI / total page dwell time) is strongly correlated between the eye and mouse ($r^2=0.89$), followed by dwell per AoI (in seconds, $r^2=0.36$). We believe that the fraction of page dwell time is a more useful measure, as it adjusts for the pace at which users read, while raw dwell times are sensitive to variation in user reading speeds. Interestingly, time-on-page after the eye/mouse visits the AoI is also reasonably well correlated ($r^2=0.45$). We will see later (section 4.5) that this measure is affected by the AoI’s relevance (the user spends more time searching on the page after visiting an AoI if the AoI is irrelevant).

We also find strong correlations between the last AoI seen by the eye and mouse before page click or abandonment ($r^2=0.86$), number of eye and mouse visits to an AoI ($r^2=0.83$). In comparison, there is weaker correlation between eye and mouse time to first noticing an AoI ($r^2=0.33$), and no correlation for time per eye/mouse pause ($r^2=0.07$).

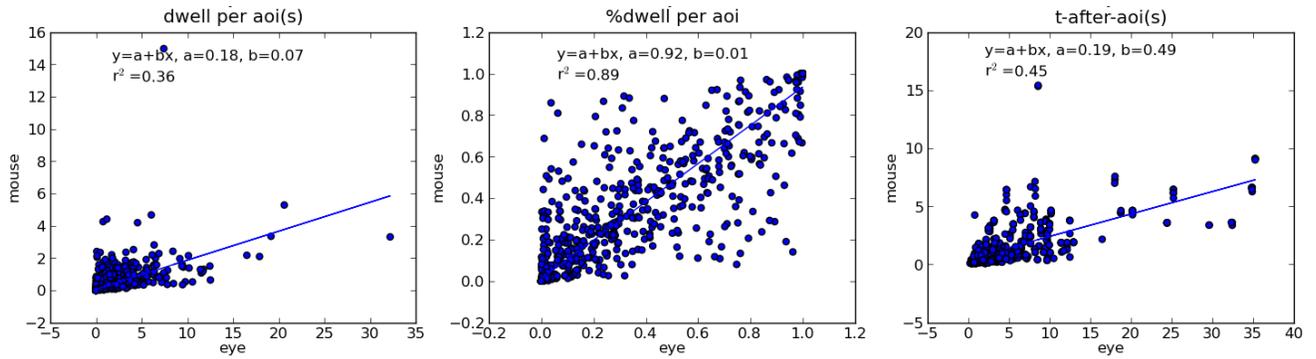


Figure 2: Area of interest (AoI) measures. The measures for eye data are shown in the x axis, and mouse data are shown in the y axis. The left panel shows the dwell time per AoI in seconds, the middle panel shows the fraction of page dwell per AoI, and the right panel shows the time on page after visiting the AoI (in seconds). Each data point reflects one user/task/AoI combination.

4.3 Information flow patterns

Does the different page layout (due to the presence of KG) alter the way users examine the search page? The typical information flow pattern is that users examine the search page linearly from top to bottom (the driving hypothesis behind the popular cascade models of user click behavior on search)[6, 5]. A Markovian analysis of eye tracks shows the following information flow pattern for our study.

- 78% fixations start at the Top (14% on mid, 5.9% on KG and near zero elsewhere), followed by nearly equal probability of switching from top results to KG or middle results.
- majority of incoming fixations on KG come from the top (81%;14% from mid, and 0.7% from bottom)
- majority of outgoing fixations from KG go to the top (78%; 12% to mid, 9.5% to left and 0.4% to bot)

We find strong correlations between eye and mouse-derived transition probabilities across AoIs ($r^2=0.73$) and the starting distribution (of first AoI visited by eye and mouse; $r^2=0.82$). This suggests that users' information flow on the page may be reasonably inferred from mouse tracks as well.

4.4 Sensitivity of eye-mouse to position and page layout

Figure 4 shows a visualization of eye and mouse patterns (superimposed) for 2 different experimental conditions, one containing KG (new page layout), and another without (traditional page layout containing one column of results). Eye patterns shown in green, mouse in blue. Each eye/mouse pause is denoted by a circle, whose size is proportional to the duration of the pause. The bigger blue circles and the smaller green circles show that the mouse pause durations are longer (e.g., mouse may be parked idle for a few seconds) compared to eye fixation durations that are typically 200-500ms. This figure shows clearly that the eye and mouse are sensitive to page layouts and KG presence – in this example, both show activity in the right hand side of the page when KG is present, and no activity when KG is absent. Over all users, the fraction of page time spent by the eye increases from <1% when KG is absent to 13% when KG is present.

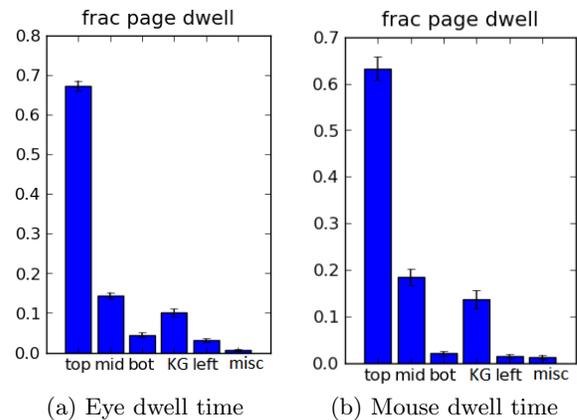


Figure 5: Fraction of page dwell in areas of interest is similar for eye and mouse

The mouse shows a similar increase, although to a smaller extent (from 9%⁵ to 15%).

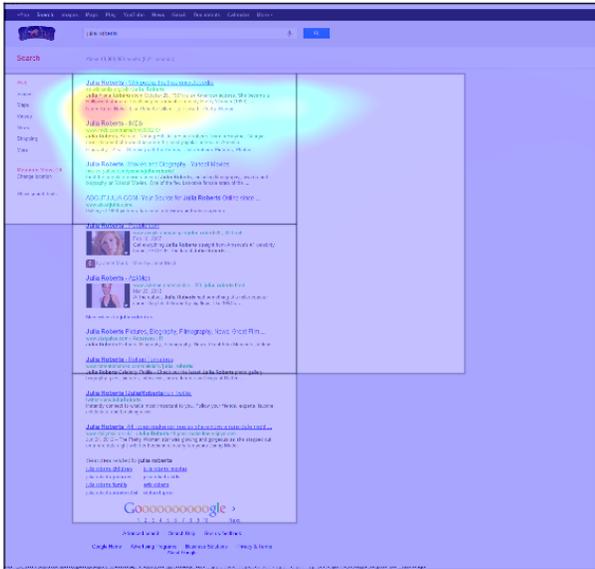
Figure 5 further demonstrates sensitivity to position by quantifying the fraction of page dwell by eye (panel a) and mouse (panel b) for different positions on the search page when KG is present. Both the eye and mouse show that the top results dominate by receiving over 60% page dwell, followed by middle results and KG, each receiving between 10-20% page dwell, followed by the bottom results (< 5%; others are negligible).

4.5 Sensitivity of eye-mouse to relevance

Figure 6 shows how KG relevance affects eye attention. While both irrelevant and relevant KGs get around 1s of at-

⁵The baseline mouse activity when KG is absent is higher than the corresponding eye activity as some users tend to park their mouse in the whitespace corresponding to the KG area, while scanning nearby result elements. Due to such noise, the magnitude of attention measures may differ between the eye and mouse, however, both the eye and mouse show the same direction of trend – an increase in activity – due to KG presence and relevance.

no kp



kp



Figure 3: Effect of KG presence on eye gaze heatmaps. Red hotspots indicate where the user spent most time looking. The left panel shows a heatmap of user eye gaze when KG is absent (the shape resembles a Golden Triangle focused on top-left of the page). The right panel shows the corresponding heatmap when KG is present. The search pages have the page areas of interest (AoIs) outlined, and regions outside AoIs dimmed. For the actual result pages presented to users, the AoIs were not visually marked in this way. Note the increased activity near KG, suggesting a potential second Golden Triangle focused on KG.

tention, there are significant differences in other measures: irrelevant KGs slow down the user by 3.5-4s on average (time on page and time on task increase), while relevant KGs speed up the user by around 4s on average (users spend 2-2.5s less on each of the top and mid as the answer is found in KG). Thus, relevant KGs get a higher fraction of page dwell (18%) than irrelevant KGs (8%), and search terminates faster on average after the user visits a relevant KG compared to an irrelevant KG (0.9 vs 2.8s). Clearly, task relevance is an important factor affecting user attention and task performance.

We tested whether mouse activity is sensitive to changes in relevance. We observe similar trends as the eye, but to a smaller extent. Like the eye, the mouse shows that relevant KGs get a higher fraction of page dwell (17%) compared to irrelevant KGs (12%), and search terminates faster on average after the user visits a relevant KG compared to an irrelevant KG (2.9 vs 6.4s). Figure 7a shows sample mouse tracks when KG is relevant – in this example, the task is “when was the sequel to *Toy Story* released?”, we find that the user finds the answer in KG, hence search terminates soon after user visits KG. Figure 7b shows sample mouse tracks when KG is irrelevant – in this example, the task is “find more about the *Let’s Move* program by Michelle Obama”, we find that the user visits KG, and continues searching on the rest of the page⁶. Thus, mouse activity, like eye gaze, is sensitive to relevance.

⁶The figure shows two different queries to illustrate more examples of pages with KGs. However, for analysis, we used the same set of queries to compare KG-relevant and KG-irrelevant conditions.

5. PREDICTING EYE FROM MOUSE

Given the observed correlations between eye and mouse activity in some measures, we are motivated to ask the following questions:

- How well can we predict eye gaze from mouse activity?
- Can we achieve higher accuracy by predicting elements of interest on the screen rather than estimating the exact eye gaze coordinates?
- To what extent is the relationship between eye gaze and mouse position user-specific and how far can we generalize to unseen users?

To answer these questions we developed a set of regression and classification models to predict the exact coordinates of the eye gaze and the element of interest on the page, respectively. Before describing these models in detail we need a formal definition of our learning problem: We divided the set of eye-mouse readings into a set of points, where each point $d_i = (y_i, e_i, \mathbf{v}_i)$ represents the eye gaze coordinates y_i , a corresponding element of interest on the page e_i , and a covariate vector \mathbf{v}_i comprising of a set of mouse features. We let $D^u = \{d_1^u, \dots, d_{n_u}^u\}$ be the set of n_u points pertaining to user u , and we let $D = \{D^1, \dots, D^U\}$ be the set of all data points across all U users.

5.1 Regression to predict the eye position

As a first step consider the problem of estimating the y -coordinate of eye-gaze directly from mouse activity⁷. This

⁷Predicting the y coordinate of eye gaze is more interesting than the x coordinate, as it can reveal which result element

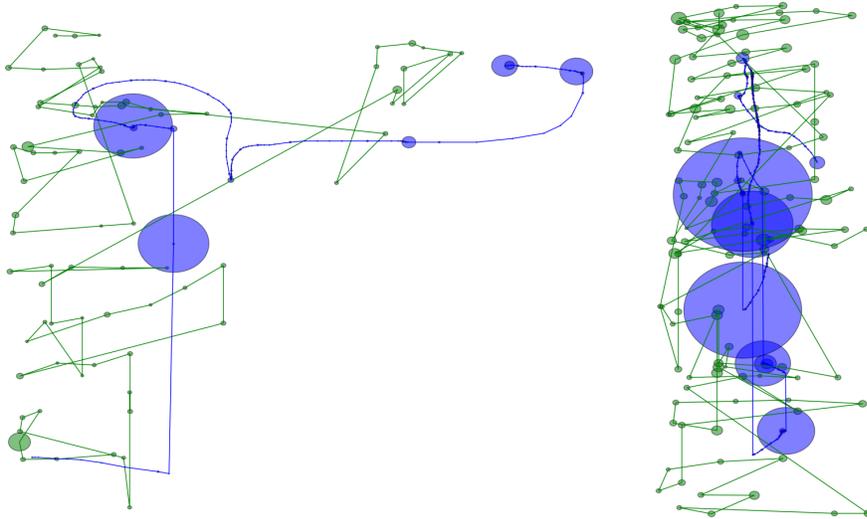


Figure 4: Examples of eye (green) and mouse (blue) tracks when KG is present (left) and absent (right)

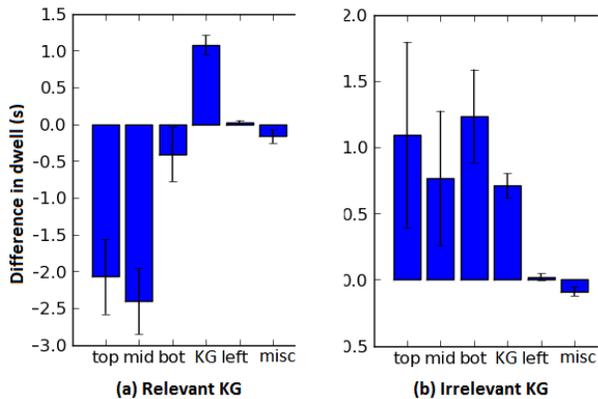


Figure 6: Effect of KG relevance on eye. Consider the left panel. The x axis shows the AoIs, and the y axis shows, for each AoI, the difference in attention (in seconds) when KG is present and relevant vs. when KG is absent (mean \pm standard error). Right panel shows the corresponding plot for irrelevant KG.

is a regression problem where we seek to find a function $f : \mathbf{v} \rightarrow y$ such that the discrepancy between $f(v)$ and the observed eye position y is minimized.

In the following we use a (generalized) linear model to represent the mapping from attributes to eye positions. That is, we seek to learn a regression function

$$f(\mathbf{v}) = \langle \mathbf{w}, \phi(\mathbf{v}_i) \rangle$$

Here f is parametrized by a weight vector \mathbf{w} that we seek to estimate. When $\phi(\mathbf{v}_i) = \mathbf{v}_i$ we end up with a linear regression function in the input covariate space \mathbf{v}_i . When

the user is looking at. Thus we focus on y coordinate in this paper.

$\phi(\mathbf{v}_i)$ comprises a nonlinear mapping, we obtain a nonlinear function in the input covariate space.

To assess the impact of a personalized model we compare the following three models: a global model that estimates the parameter \mathbf{w} common for all users. Secondly we infer a user-specific model that provides an upper bound of how accurately the model can estimate eye positions from mouse activity. Finally, we infer a hybrid model that combines global and user-specific components. This allows us to dissociate both parts, thus allowing us to generalize to users where only mouse movements are available while obtaining a more specific model whenever eye tracking is possible. We describe these three approaches below:

Global model: In this setup we learn a global regression function f_g parametrized by a global weight vector \mathbf{w}_g . The learning goal is to find \mathbf{w}_g that minimize the average prediction error on the *whole* dataset. More formally, our learning problem for the y -coordinate is:

$$\underset{\mathbf{w}_g}{\text{minimize}} \sum_{d_i \in \mathbf{D}} \|y_i - \langle \mathbf{w}_g, \phi(\mathbf{v}_i) \rangle\|_2^2 + \lambda \|\mathbf{w}_g\|_2^2$$

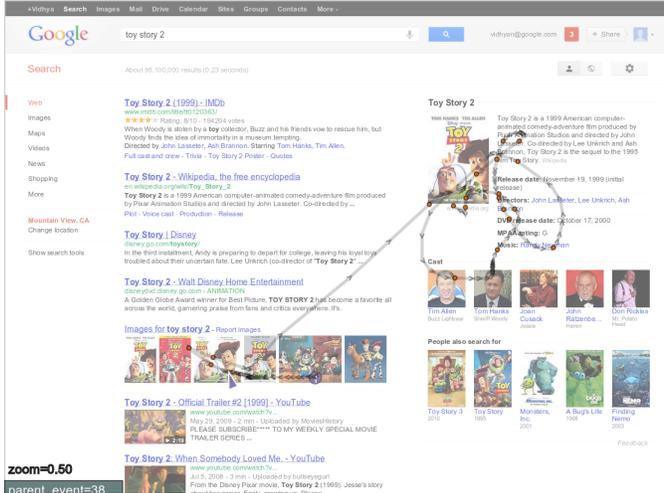
where λ is a regularization parameter to prevent overfitting. This model tests the hypothesis that eye-mouse correlation is a global phenomenon and does not depend on the specific user behaviour.

User-specific models: In this setup we learn regression functions f_u *independently* for each user u . The learning problem for the y -coordinate is:

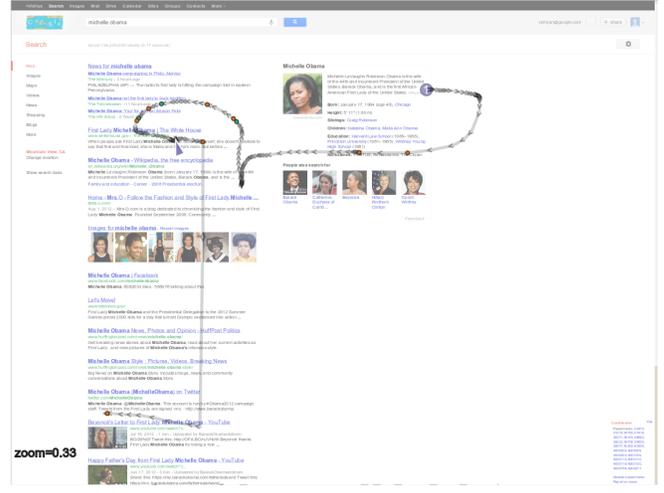
$$\underset{\mathbf{w}_u}{\text{minimize}} \sum_{d_i \in \mathbf{D}} \|y_i^u - \langle \mathbf{w}_u, \phi(\mathbf{v}_i^u) \rangle\|_2^2 + \lambda \|\mathbf{w}_u\|_2^2$$

This model tests the hypothesis that eye-mouse correlation is NOT a global phenomenon and depends on the specific user behaviour.

Hierarchical model: In this setup we still learn a per-user regression model, however we decompose each user-specific regression weight additively into a user-dependent part \mathbf{w}_u and a global part \mathbf{w}_g . More for-



Relevant KG



Irrelevant KG

Figure 7: Effect of KG relevance on mouse. Search tends to terminate soon after visiting a relevant KG. For irrelevant KGs the user visits the KG and then continues examining the rest of the page. See footnote 6.

mally, our learning problem for the y-coordinate is:

$$\begin{aligned} & \text{minimize}_{\mathbf{w}_g, \mathbf{w}_{u_1}, \dots, \mathbf{w}_U} \sum_{u \in U} \sum_{d_i^u \in \mathcal{D}^u} \|y_i^u - \langle \mathbf{w}_u + \mathbf{w}_g, \phi(\mathbf{v}_i^u) \rangle\|^2 + \\ & \lambda \left[\sum_{u \in U} \|\mathbf{w}_u\|_2^2 + \|\mathbf{w}_g\|_2^2 \right] \end{aligned}$$

This model tests the hypothesis that eye-mouse correlation has some global patterns shared across users, as captured by \mathbf{w}_g , in addition to user-specific patterns, as captured by the set \mathbf{w}_u weights.

5.2 Classification for elements of interest

Instead of estimating the absolute position of eye gaze explicitly, one might want to settle for a slightly simpler task — that of estimating which element of interest is being inspected by the user. For this purpose we divide the screen into blocks of pixels that represent special elements on the page (e.g., result element 1,2, etc.). In our experiments, each page is divided into a set of 11 different elements (10 result elements, 1 KG)⁸. The prediction task here involves predicting the particular element that the eye gaze is currently focused on using information from the mouse activity. We treat this as a multi-label classification problem. We use the same terminology from Section 5.1. Our goal is to learn a classification function $h(\cdot, \mathbf{w}) : \phi(\mathbf{v}_i) \rightarrow \mathcal{L}$, where \mathcal{L} is the label space. In analogy to Section 5.1, we are interested in the following three cases:

Global model: In this setup we learn a global classification function h_g parametrized by a global weight vector \mathbf{w}_g . The learning goal is to find \mathbf{w}_g that minimize the misclassification error on the *whole* dataset:

$$\text{minimize}_{\mathbf{w}_g} \sum_{d_i \in \mathcal{D}} I[e_i \neq h(\phi(\mathbf{v}_i), \mathbf{w}_g)] + \lambda \|\mathbf{w}_g\|_2^2$$

⁸Note that the 11 elements for classification are different and more fine-grained than the area-of-interest classes mentioned in section 4.

where I is the indicator function, which is 1 iff its argument is evaluated to be true.

User-specific models: In this setup we learn classification function h_u independently for each user u :

$$\text{minimize}_{\mathbf{w}_u} \sum_{d_i^u \in \mathcal{D}^u} I[e_i^u \neq h(\phi(\mathbf{v}_i^u), \mathbf{w}_u)] + \lambda \|\mathbf{w}_u\|_2^2$$

Hierarchical models: in this setup we still learn a per-user regression model, however we decompose each user-specific regression weight additively into a user-dependent part \mathbf{w}_u and a global part \mathbf{w}_g :

$$\begin{aligned} & \text{minimize}_{\mathbf{w}_g, \mathbf{w}_{u_1}, \dots, \mathbf{w}_U} \sum_{u \in U} \sum_{d_i^u \in \mathcal{D}^u} I[e_i^u \neq h(\phi(\mathbf{v}_i^u), \mathbf{w}_g + \mathbf{w}_u)] \\ & + \lambda \left(\|\mathbf{w}_g\|_2^2 + \sum_{u \in U} \|\mathbf{w}_u\|_2^2 \right) \end{aligned}$$

Note that for optimization purposes the indicator function for correct labels is replaced by a differentiable loss function. Alternatively, one may use a reduction to binary approach and solve a sequence of associated cost-sensitive learning problems [1].

5.3 Experimental setup

To be able to test our hypothesis we divided the data into train and test as follows. We randomly sampled 30% of the users and reserved them for test. Then for the remaining 70% users, we randomly sampled 70% of their data points to construct the training data set and added the remaining 30% of their data points to the test set. By that, our test set comprises two kind of users: 1) users unseen during the training, and 2) users partially seen via some of their data points in the training set. This allows us to test the generalizability of our models and test our hypothesis of whether or not mouse-eye correlation is user-specific or user-independent. For all the experiments below, we report accuracy on the whole test set and break it as well into ac-

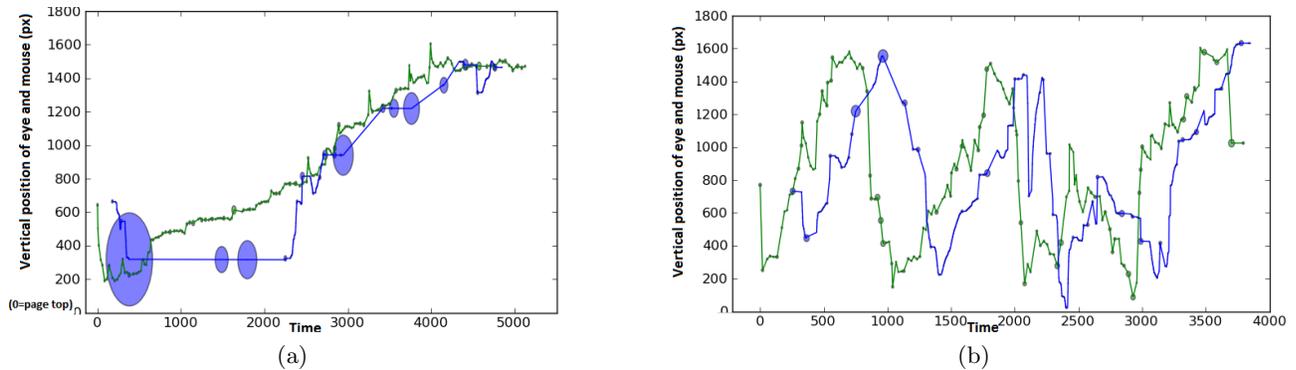


Figure 8: Examples of eye-mouse time course in the y direction. y axis shows the vertical position of eye (green) and mouse (blue) (0 is page-top, and increasing values represent further down the page), and x axis shows time in centiseconds (multiply by 10 to get time in milliseconds). The size of the blue blobs is proportional to the duration of mouse pause at that position. In the example on the left, the mouse is initially parked for around 20 seconds, while the eye examines the nearby elements carefully. The mouse then jumps forward, and then on, correlates better with the eye (task was “Describe the Mozart programming system”). The example on the right shows that the mouse follows eye gaze (with a lag) as the user looks up and down the search page (task was “Describe the koala library for facebook”).

curacy on seen user and accuracy on unseen users.⁹ We perform prediction as follows based on the model:

Global models: We use the same weight vector \mathbf{w}_g on both seen and unseen users.

User-specific models: If the user was seen before we use his own specific weight \mathbf{w}_u . For this model, we do not report results over unseen users.

Hierarchical models: If the user was seen before we use $\mathbf{w}_g + \mathbf{w}_u$, otherwise we use \mathbf{w}_g .

We use the following information from mouse activity as features in our prediction models for each time point t :

1. Time from page load (t)
2. Cursor Position (x_t, y_t)
3. Cursor Velocity: magnitude (v_{xt}, v_{yt}), and direction (s_{xt}, s_{yt})
4. Cursor Distance moved (d_{xt}, d_{yt})
5. Area-of-interest in which cursor lies (aoi_t)
6. Corresponding page element (el_{mt})

In addition to computing these features at time t , we consider past and future values of features 2-4. e.g., we consider the future cursor positions, average future velocities, total cursor distance moved, and number of changes in mouse movement direction for time windows $[t, t+k]$ where $k \in \{1, 2, 4, 8, 16\}$; similarly for the past. This gives us a total of 83 features by the above phase space embedding.

5.4 Eye Prediction Results

Next, we show results from various prediction models on different corpora under the various task settings described

⁹For seen users, it is possible that adjacent data points from the same user may end up being split across the train and test datasets. However, all the methods compared here are provided with identical train/test distributions and therefore this does not introduce bias for any particular method. In addition (as mentioned earlier), we also report results when predicting on test data from completely unseen users that do not overlap with the train dataset.

above. Following Huang et al. [15], we use a baseline model which predicts that the eye gaze is focused exactly at the mouse position. We use two feature map functions $\phi(\mathbf{v})$: linear where $\phi(\mathbf{v}) = \mathbf{v}$ and nonlinear mapping using Nyström approximation for Gaussian RBF kernels [24].

Denote by $k(\mathbf{v}, \mathbf{v}') = \exp(-\gamma \|\mathbf{v} - \mathbf{v}'\|^2)$ a Gaussian RBF kernel as it is commonly used in kernel methods. Then this kernel can be approximated by the following feature map:

$$\tilde{k}(\mathbf{v}, \mathbf{v}') = \langle \tilde{\phi}(\mathbf{v}), \tilde{\phi}(\mathbf{v}') \rangle \text{ where}$$

$$\tilde{\phi}(\mathbf{v}) = K_{nn}^{-\frac{1}{2}} [k(\mathbf{v}_1, \mathbf{v}), \dots, k(\mathbf{v}_n, \mathbf{v})]$$

Here $\mathbf{v}_1, \dots, \mathbf{v}_n$ are 1,000 random observations and K_{nn} is an $n \times n$ matrix obtained by forming the inner products $[K_{nn}]_{ij} = k(\mathbf{v}_i, \mathbf{v}_j)$. The advantage of the mapping $\tilde{\phi}(\mathbf{v})$ is that it can be used to learn a linear function in the transformed feature space that is equivalent to learning a nonlinear function in the input space. That is, in both cases we employ the *VW* package [19] to solve all the optimization problems.

Results are shown in Table 1. We draw the following conclusions:

- The observed mouse-eye correlation function is highly nonlinear and that is why nonlinear models outperformed their linear counterparts especially in classification settings. For example, the best results achieved on unseen users using nonlinear features is 62.3% prediction error compared to 80.7% error for the linear counterpart (which amounts to 23% reduction in classification error). This is a natural consequence of the fact that we need to specify nonlinear boundaries in order to distinguish between different blocks in the result set. Here nonlinear basis functions are much more advantageous.
- Our best model (nonlinear hierarchical) provides a 24.3% improvement over the baseline (in $RMSE_y$ for pixel space). In comparison, the best model from Huang et

al[15] achieved an improvement of 10.4% over the same baseline.¹⁰

- The observed mouse-eye correlation function is clearly user-dependent because users exhibit different patterns while navigating the page (for example some users tend to align the mouse with their gaze as their attention shifts around the page, so called mouse readers). From our results, it is clear that the user-specific models overwhelmingly outperform all other models.
- While building a user-specific model is advised whenever abundant training data is available from individual users, it is not a scalable approach. We hypothesized that hierarchical models would help in generalizing eye-gaze prediction to unseen users. As evident from our experiments using a simple additive hierarchical model, the results over unseen users slightly improved compared to the global models. The reason is that additive models separate user-specific patterns (via the \mathbf{w}_u weight vector) from global patterns shared across users (captured via the \mathbf{w}_g vector) which are then used in predicting eye-gaze from mouse activity over unseen users. We believe that this is a promising direction and we plan to investigate further into using more advanced transfer learning approaches for transferring eye-mouse patterns from seen users to unseen users.
- Note that at a first glance, the total classification error over seen and unseen users from our best model seems rather high (nonlinear hierarchical, 60.3% error). However, this impression is misleading: firstly, it amounts to a 14.8% reduction over its counterpart in the baseline (70.8%), and the error reduction is bigger (28%) for user-specific models on seen users. Secondly, the current method of computing error penalizes adjacent elements as much as far away page elements, leading to high values of error (for example, predicting that the user is looking at search result 1 while the ground truth is that he is looking at search result 2 would result in the same classification error as between results 1 and 10). A cost sensitive loss function taking the page layout into account could be used to address this issue. Indeed, we find that the result elements were mostly confused with adjacent result elements, and KG was confused with the first result element. If we ignore errors due to adjacent elements on the page, the total classifier error drops dramatically from 60.3% to 33.1% (nearly halved). For example, the error for the first result element drops from 23 to 12% (91% of this error reduction was from ignoring result 1-2 confusion), the error for the second element drops from 79 to 8% (97% error reduction was from ignoring result 2-1, 2-3 confusion), the error for KG drops from 60 to 12% (75% error reduction was from ignoring KG-result 1 confusion). To summarize, these results suggest that the nonlinear hierarchical model can predict the result

element that the user is looking at (with an error of upto one element) at reasonably high accuracy (67%) from mouse tracks only.

6. DISCUSSION

In this paper, we demonstrate through carefully designed experiments in the lab that the mouse, like the eye, is sensitive to two key attributes of page elements – their position on the page, and their relevance to the user’s task. Using a 2x2 within subject design, we systematically varied the page layout (2 column content layout – KG present on the top-right of the page, vs. linear 1 column content layout with web results only and KG absent), and relevance (KG either relevant or not to user task). Eye and mouse tracks were recorded simultaneously as users performed the tasks. We discuss the key findings and potential applications.

6.1 Mouse tracking aids layout analysis

Our analysis shows that the eye and mouse are strongly correlated in some measures such as: 1) fraction of page time spent within the area-of-interest, 2) last area-of-interest visited before user clicks/abandons, and 3) transition probabilities between areas of interest. The first measure is particularly interesting as it is sensitive to both the position and relevance of an AoI: comparing different experimental conditions, we observe that like the eye, the fraction of page time on KG as measured by the mouse, is higher when KG is present vs. absent, and higher when KG is relevant vs. irrelevant. In addition, we find that the page time *after* the mouse visits KG is shorter when KG is relevant than when it is irrelevant, suggesting that users terminate search faster after visiting a relevant KG. Together, these mouse measures can provide useful signals about user attention to, and task relevance of, page elements even in the absence of clicks. Potential applications of tracking these mouse measures at scale include ranking, search page optimization and UI evaluation.

The second finding of strong eye-mouse correlations in the last area-of-interest visited before click/abandonment is consistent with Huang et al’s finding that eye and mouse are strongly correlated at the time of click (user’s decision) than at the beginning of the task (where users tend to explore the page).

As search pages become increasingly complex (visually and content-wise), understanding how users consume the page, or how information flows through the search page is an important question that has applications for ranking and improving user satisfaction (by providing the high quality answers where they are most likely to look at). For example, in linear page layouts containing a single column of search results, the dominant flow pattern is that users scan the page from top to bottom. The corresponding eye gaze heatmaps resemble a Golden triangle with more attention being paid to the top-left of the page, and decreasing attention towards the right or bottom of the page. In contrast, in the new page layout containing 2 columns (one column of search results, and a second column with visually attractive content on the top right of the page corresponding to the Knowledge Graph results), we find that while majority of the users start by viewing the top search results, as a next step, they are equally likely to see the KG on the right, or middle results below. The corresponding eye gaze heatmaps in Figure 3 have 2 Golden triangles, one focused at the top-

¹⁰Huang et al. report results using a slightly different set of features which makes a direct comparison difficult for this particular study. However, we use the same baseline method as theirs and although the feature sets may differ, our global linear model simulates their method. The results presented in this paper indicate that our best approach clearly outperforms these methods and yields a significant improvement in prediction performance for different test settings.

Model		RMSE _y (pixels)			Classification Error (%)		
		Total	Seen users	unseen users	Total	Seen users	unseen users
Baseline	<i>mouse position</i>	270.1	276.9	263.9	70.8	72.0	69.7
Linear	Global	218.2	217.0	219.3	77.3	72.4	81.6
	Hierarchical	216.5	215.1	218.2	76.0	71.9	80.7
	User-specific	–	193.8	–	–	55.9	–
Nonlinear	Global	211.7	210.0	213.2	63.9	64.8	63.0
	Hierarchical	204.5	201.3	207.5	60.3	58.7	62.3
	User-specific	–	179.7	–	–	51.8	–

Table 1: Comparison of models for predicting eye gaze from mouse activity. Size of training data $m_{\text{train}} = 20788$, test data $m_{\text{test}} = 19000$ (comprised of 8899 points from previously seen users and 10101 points from new users). Our best model (nonlinear hierarchical) provides a 24.3% improvement over the baseline (in pixel space) and 14.8% improvement in element space.

left of the search results, and a new triangle focused at the top-left of the KG. Our finding of eye-mouse correlations for the starting distribution (first area-of-interest visited by the eye and mouse), and transition probabilities between pairs of area-of-interest suggests that we may reasonably infer user information flow on search pages by analyzing their mouse tracks. This could be potentially useful as search pages evolve to new UIs, layouts, and richer results, which may not be parsed in a conventional linear manner.

There is also reasonable eye-mouse correlation on maximum y distance on the page, which reflects how far down the page did the user visit (with their eye or mouse). This could be useful in inferring which page folds were attended by the user, and how far down the page they read.

6.2 On mouse tracking as a proxy for eye tracking

While mouse tracking is gaining in popularity, it remains a poorly understood science. Future work will involve analyzing mouse usage statistics, understanding how mouse usage behavior varies as a function of users (intent, age, gender, reading fluency), page (UI, images, result quality) and device properties (e.g., trackpad vs. traditional mouse).

In particular, there exists a rich body of statistical literature to analyze models involving partially observed data, such as Hidden Markov Models and Conditional Random Fields. Analysis of users’ mouse activity (even in the absence of eye data) can provide information about prototypical mouse behaviors of users (e.g., parking, reading, skimming). In turn, such information can help us estimate eye positions more accurately in spite of the paucity of eye tracking data.

limitations of mouse tracking. Unlike eye gaze patterns, there is wide variability in mouse behavior between and within users. For example, some users read with their mouse, which as Rodden et al. [22] note is a rare event. Other users tend to park their mouse idle while scanning the page with their eye. Some users tend to mark interesting results with their mouse, others simply use their mouse for scrolling and clicking. Given the wide variability and noise in mouse usage behavior, inferring attention from mouse is a hard problem. Further noise in mouse behavior may be introduced depending on the type of mouse device (e.g., a trackpad may involve more scrolling and smooth mouse movements than a linear mouse). Due to the above reasons, we need to rely on aggregate statistics of mouse behavior over several users to infer

their attention on page. In contrast, eye tracks from a single user can already reveal a lot about their attention on page.

This leads us to the conclusion that while mouse tracking can offer valuable signals on user attention on areas-of-interest or page elements at an aggregate level, it cannot yet match the millisecond temporal and pixel-level spatial resolution of eye tracking on a per user, per page level. Despite this limitation, mouse tracking has much promise as it offers a scalable methodology to infer user attention on web pages, especially when clicks are absent or few.

7. CONCLUSIONS

We demonstrate through carefully designed lab studies that the mouse, like the eye, is sensitive to 2 key attributes of page elements – their position on page, and their relevance to the user’s task – both for linear one-column page layouts and increasingly popular two-column page layouts. Despite the noise in mouse activity due to wide variability in mouse usage behavior within and between users, we find strong eye-mouse correlations in measures such as the fraction of page time on result elements, and transition probabilities between elements, suggesting that one may reasonably infer user attention and information flow over elements on the search page, using mouse tracks. This is further validated by the reasonably high accuracy (67%) in predicting the fixated result element from mouse activity (with an error of upto one element). Potential applications include ranking, search page optimization, and UI evaluation both in the presence, and absence of clicks.

8. REFERENCES

- [1] N. Abe, B. Zadrozny, and J. Langford. An iterative method for multi-class cost-sensitive learning. In *KDD*, pages 3–11, 2004.
- [2] G. Buscher, E. Cutrell, and M. Morris. What do you see when you’re surfing?: using eye tracking to predict salient regions of web pages. In *Proceedings of the 27th international conference on Human factors in computing systems*, pages 21–30. ACM, 2009.
- [3] G. Buscher, S. Dumais, and E. Cutrell. The good, the bad, and the random: an eye-tracking study of ad quality in web search. In *Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, pages 42–49. ACM, 2010.

- [4] G. Buscher, R. W. White, S. Dumais, and J. Huang. Large-scale analysis of individual and task differences in search result page examination strategies. In *Proceedings of the fifth ACM international conference on Web search and data mining*, WSDM '12, pages 373–382, New York, NY, USA, 2012. ACM.
- [5] O. Chapelle and Y. Zhang. A dynamic bayesian network click model for web search ranking. In *Proceedings of the 18th international conference on World wide web*, pages 1–10. ACM, 2009.
- [6] N. Craswell, O. Zoeter, M. Taylor, and B. Ramsey. An experimental comparison of click position-bias models. In *Proceedings of the international conference on Web search and web data mining*, pages 87–94. ACM, 2008.
- [7] E. Cutrell and Z. Guan. What are you looking for?: an eye-tracking study of information usage in web search. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 407–416. ACM, 2007.
- [8] A. Duchowski. *Eye tracking methodology: Theory and practice*, volume 373. Springer, 2007.
- [9] L. Granka, T. Joachims, and G. Gay. Eye-tracking analysis of user behavior in www search. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 478–479. ACM, 2004.
- [10] Q. Guo and E. Agichtein. Exploring mouse movements for inferring query intent. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '08, pages 707–708, New York, NY, USA, 2008. ACM.
- [11] Q. Guo and E. Agichtein. Ready to buy or just browsing?: detecting web searcher goals from interaction data. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '10, pages 130–137, New York, NY, USA, 2010. ACM.
- [12] Q. Guo and E. Agichtein. Towards predicting web searcher gaze position from mouse movements. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '10, pages 3601–3606, New York, NY, USA, 2010. ACM.
- [13] Q. Guo and E. Agichtein. Beyond dwell time: estimating document relevance from cursor movements and other post-click searcher behavior. In *Proceedings of the 21st international conference on World Wide Web*, WWW '12, pages 569–578, New York, NY, USA, 2012. ACM.
- [14] G. Hotchkiss, S. Alston, and G. Edwards. Eye tracking study. *Research white paper, Enquiro Search Solutions Inc*, 2005.
- [15] J. Huang, R. White, and G. Buscher. 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, pages 1341–1350, New York, NY, USA, 2012. ACM.
- [16] J. Huang, R. W. White, G. Buscher, and K. Wang. Improving searcher models using mouse cursor activity. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '12, pages 195–204, New York, NY, USA, 2012. ACM.
- [17] J. Huang, R. W. White, and S. 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*, CHI '11, pages 1225–1234, New York, NY, USA, 2011. ACM.
- [18] D. Lagun and E. Agichtein. Viewer: enabling large-scale remote user studies of web search examination and interaction. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, SIGIR '11, pages 365–374, New York, NY, USA, 2011. ACM.
- [19] J. Langford, L. Li, and A. Strehl. Vowpal Wabbit, 2007.
- [20] J. Nielsen and K. Pernice. *Eyetracking web usability*. New Riders Pub, 2010.
- [21] A. Olsen. Tobii i-vt fixation filter - algorithm description white paper. 2012.
- [22] K. Rodden, X. Fu, A. Aula, and I. Spiro. Eye-mouse coordination patterns on web search results pages. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '08, pages 2997–3002, New York, NY, USA, 2008. ACM.
- [23] D. Salvucci and J. Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*, pages 71–78. ACM, 2000.
- [24] A. J. Smola and B. Schölkopf. Sparse greedy matrix approximation for machine learning. In *ICML*, pages 911–918, 2000.