Towards Predicting Web Searcher Gaze Position from Mouse Movements

Qi Guo
Emory University
400 Dowman Dr., W401
Atlanta, GA 30322 USA
qguo3@emory.edu

Eugene Agichtein
Emory University
400 Dowman Dr., W401
Atlanta, GA 30322 USA
eugene@mathcs.emory.edu

CHI 2010, April 10–15, 2010, Atlanta, Georgia, USA.
ACM 978-1-60558-930-5/10/04.

Abstract
A key problem in information retrieval is inferring the searcher's interest in the results, which can be used for implicit feedback, query suggestion, and result ranking and summarization. One important indicator of searcher interest is *gaze position* – that is, the results or the terms in a result listing where a searcher concentrates her attention. Capturing this information normally requires eye tracking equipment, which until now has limited the use of gaze-based feedback to the laboratory. While previous research has reported a correlation between mouse movement and gaze position, we are not aware of prior work on automatically inferring searcher's gaze position from mouse movement or similar interface interactions. In this paper, we report the first results on *automatically inferring whether the searcher's gaze position is coordinated with the mouse position* – a crucial step towards predicting the searcher gaze position by analyzing the computer mouse movements.

Keywords
Web search, searcher behavior, eye-mouse coordination

ACM Classification Keywords
H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval – search and selection process.

General Terms
Design, Experimentation, Human Factors

Introduction

A key problem in web search is inferring searcher intent and interest in the results, which is important for tasks such as implicit feedback [1], query suggestion, and result ranking and summarization. An accurate indicator of searcher interest is gaze position [1], which normally requires eye tracking equipment. Unfortunately, eye tracking-based feedback is not yet feasible on web-scale. However, work by Rodden et al. [10] has shown that *sometimes* there is a correlation between mouse and eye position, and described general patterns of mouse/eye coordination. Yet, we are not aware of any reports of *automatically inferring* the gaze position from mouse movements and other interaction features. Furthermore, it has not been shown if it is possible to predict, for a given point in the interaction process, whether the gaze position and mouse position are coordinated -- which is the first step towards predicting the searcher's gaze position. Thus, we explore the basic research question:

> *Is it possible to predict the precise times when mouse and gaze position are closely coordinated, based only on mouse position and movement?*

Our work advances the state of the art by reporting, to our knowledge, the first successful method for *inferring* the times of mouse-eye coordination from interface interactions alone. Specifically, we examine the characteristics of eye-mouse coordination (measured as Euclidian distance between the mouse and gaze position), using data collected from a controlled user study of web search with 10 subjects and 200 instances of search tasks. We consider factors such as basic search task type and result quality – that may have impact on the eye-mouse coordination patterns. Finally, we use machine learning techniques to *infer* mouse-eye coordination, using only the features of the mouse movement. Our results demonstrate that it is indeed possible to predict with over 70% accuracy whether gaze position and mouse position are coordinated. While this is ongoing work, our results may be already useful for search tasks such as implicit feedback, query suggestion, and click prediction.

Related work

Recently, eye tracking has started to emerge as a useful technology for understanding searcher behavior (e.g., [3, 7]). However, eye tracking equipment is not usually available for web search users, and there has been significant interest in *inferring* searcher interest from clicks and other interface actions [8, 9]. In particular, searcher interactions have been used to help disambiguate query intent [5], for implicit feedback [2], and for more accurate prediction of search result clickthrough [6].

However, the work above side-steps the problem of inferring user interest as expressed by gaze position, and focuses on application-specific measures. In contrast, we are interested in the fundamental problem of predicting mouse-eye coordination, which could have applications beyond our specific web search setting. We significantly expand on the previous work by Rodden et al. [10] which described general patterns of eye-mouse coordination, and reported tantalizing statistics of *overall* correlation between mouse and eye position. In contrast, we present, to our knowledge, the first method to *predict* the gaze position from the mouse movements.

Study

Apparatus

To capture the user interactions, including mouse movements, we use our extension of a LibX plugin for the Firefox browser. The events are encoded in a string and sent to the server as HTTP requests for analysis. We sample mouse movements at every 5 pixels moved, or every 50 ms, whichever is more frequent. Eye tracking was performed using the Tobii T120 system, paired with a 17" LCD monitor set at the resolution of 1280×1024 for all experiments. The Web browser was maximized for all subjects, allowing us to reproduce exactly the same result presentation for all subjects.

Participants

Ten subjects were recruited for this user study. All the subjects were graduate students and staff in the Emory Math & CS department; as such, all subjects were technically savvy and had some experience with web search.

Tasks and Search Results

We used a set of 20 web search tasks, sampled from the web search usage logs collected from public-use machines in the Emory University Libraries. Table 1 reports example search tasks, together with the descriptions provided to the participants. Half of the tasks were Navigational (find a particular web site or page) and the others were Informational (a specific piece of information on any web page).

Query	***Description***
Navigational Tasks	
cnn	Find the official site of CNN
wikipedia	Find the official site of Wikipedia
Informational Tasks	
circuit city ceo 1995	Find out who was the CEO of circuit city in 1995
setting up a gym in your house	Find out how to set up a home gym or work out area

Table 1: Example search tasks (queries and descriptions).

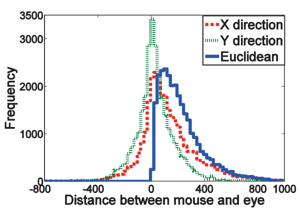

Figure 1: Eye-mouse distance (X, Y, and Euclidean).

For each task, a pre-specified query was provided and the associated first Search Engine Result Page (SERP) was stored by querying the Google search engine. To ensure result consistency, only the stored copies were shown to the participants. Nine of the 20 SERPs contained images while the rest do not. To explore the effects of result quality, the search results for half of the tasks (namely, for 5 navigational and 5 informational tasks) were *randomized*. Thus, our dataset consisted of 20 SERPs, with 10 original result ranking, and 10 with the top search results in random order.

Procedure

To begin a task, the participants were presented a list of queries and task descriptions. For each task, the participants were instructed to find the most relevant result on the cached first Search Engine Result Page (SERP) for the pre-specified query. Once they reached the SERP, they could read the text and find the most relevant result to click. Clicking a chosen result recorded the subjects' decision, and redirected them back to the task description list to start the next task.

Characterizing Eye-Mouse Coordination

Eye-mouse distance characteristics

In order to calculate the set of distances between mouse and eye, we matched each mouse point with the eye *fixation* (if any) in the same time interval. The overall distributions of distances are reported in Figure 1, and are similar to Rodden et al. [10]. The mean Euclidean distance is 225 pixels (σ=178); the median is 180. For the Y direction there is a higher peak around the origin than for the X direction, suggesting that mouse and eye positions coordinate more closely in the vertical direction, which agrees with the findings in [10].

Factors		***X direction***	***Y direction***	***Euclidean***
All		166	115	228
Task	Navigational	147 **	124 **	219
	Informational	183 **	107 **	235
Ranking	Original	152 *	119	219
	Randomized	179 *	111	236
Image?	Yes	159	116	220
	No	173	114	234

Table 2. Mean eye-mouse distance for varying search factors.

Eye-mouse coordination for varying factors

We now analyze the eye-mouse distance according to the task (Informational vs. Navigational), Ranking (Original vs. Randomized) and Result type (Image or No Image). The corresponding means for each pair of variables are summarized in Table 2. All significant differences with two-tailed unpaired t-tests for each variable (i.e., task, result order, image in result) are marked using * for $p < .1$ and ** for $p < .05$. As Table 2 reports, the mean Euclidean distance is relatively consistent across different factors, while X and Y distance exhibit significant differences for task type ($p<0.05$). For navigational tasks, the mean X distance is smaller, while the mean Y distance is higher - which could be explained by "easier to find" results by scanning the result page without moving the mouse, at least in the original ranking. Interestingly, low result quality (simulated by randomized ranking) forced the searchers to examine the results more carefully. We posit that it is in these cases, predicting gaze position and identifying the few relevant results could be particularly helpful. Finally, the presence of images did not have a significant effect on overall mouse-eye coordination.

Eye-mouse coordination for different subjects

The mean eye-mouse distances for each subject across all tasks are reported in Figure 2. Four of the participants exhibit significant differences in mean mouse-eye distance ($p<0.01$, marked by * on the X axis). However, most subjects maintain roughly 200 pixel eye-mouse distance on average. Further investigation with a larger participant group, is needed to characterize individual differences between subjects.

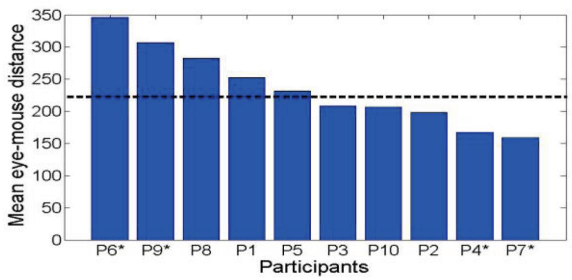

Figure 2. Eye-mouse Euclidean distance for different participants (*=p<.01). The dashed line represents the mean distance over all participants.

Eye-mouse coordination patterns

Previous work has identified three patterns of mouse-eye coordination in search, namely "incidental" mouse usage, "bookmarking", and "follow" [10]. The *follow* and *bookmarking* patterns are particularly important as they could potentially be used to provide evidence for implicit feedback. We illustrate these patterns in Figure 3 (A-F). In panels A, C, and E, the eye fixations (blue circles) and mouse positions (black circles) are overlaid on the original SERPs. Panels B, D, and F plot the corresponding eye-mouse Euclidian distance values over time.

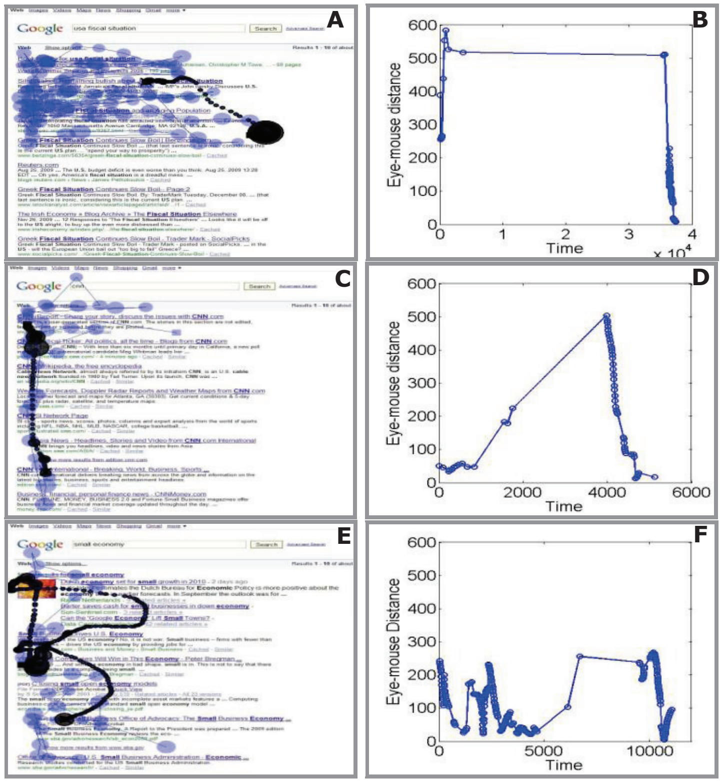

Figure 3. Examples of mouse-eye coordination and corresponding Euclidian distance time series: *incidental* mouse usage (A-B), *bookmarking* (C-D), and *follow* (E-F).

The distance patterns exhibited by the *incidental* (A-B) and *bookmarking* (C-D) mouse usage illustrate the difficulty in automatically detecting eye-mouse coordination: in both cases, the mouse is stationary for a long period, indicated by a relatively larger black circle. However, in the *incidental* case the mouse position is meaningless, while in the *bookmarking* case, the mouse position indicates a result of interest which has been carefully examined. One possible distinguishing characteristic is the X-position of the mouse, which is one of the features we use for prediction.

Predicting Eye-Mouse Coordination

We formulate our prediction task as binary classification: that is, we classify each mouse sample point into two classes: *InFocus* (eye-mouse distance is below a threshold of *T* pixels) and *Away* (eye-mouse distance is above a threshold of *T* pixels). We experimented with *T* values from 200 to 100 pixels, that is, from roughly the mean Euclidean eye-mouse distance to roughly the vertical size of one search result.

Feature	*Description*
x_i	X coordinate
y_i	Y coordinate
t_i	time elapsed
Δt_i	$t_i - t_{i-1}$
Δx_i	$x_i - x_{i-1}$
Δy_i	$y_i - y_{i-1}$
vH_i	$abs(\Delta x_i/\Delta t_i)$
dH_i	$sign(\Delta x_i/\Delta t_i)$
vV_i	$abs(\Delta y_i/\Delta t_i)$
dV_i	$sign(\Delta y_i/\Delta t_i)$
pID	Participant ID

Table 3. Features for representing a mouse sample point *i*

Evaluation Metrics

We use standard IR and classification metrics:

- *Accuracy (Acc):* Fraction of points in the timeline that were correctly predicted, reported in %.
- *Precision (Prec):* For a given class, fraction of predicted instances that were correctly predicted.
- *Recall (Rec):* For a given class, fraction of all true instances that were correctly predicted.
- *Macro-averaged F1 (F1):* F1 measure, computed as 2*Prec*Rec / (Prec + Rec), averaged across both classes.

Note that we are particularly interested in the *InFocus* predictions, as the most useful for implicit feedback and other applications. Thus, we explicitly report the Precision and Recall for the *InFocus class* in addition to the overall *Accuracy* and *F1* performance.

Baseline method

To gauge the performance of our classification methods, we also report performance of a naive baseline, which always guesses the majority class (e.g., the *Away* class for threshold = 100, which is true for almost 74% of the data). Thus, our majority baseline has significantly higher accuracy than a more traditional "random" baseline (50% accuracy).

Our Model and Implementation

We represent each mouse sample point as a feature vector, with values corresponding to the X-coordinate, Y-coordinate, time elapsed since page load, velocity, and others (see Table 3, sidebar, for the full list). We also include a feature for the participant ID to allow for the possibility of "personalizing" the predictions. We experimented with different classification algorithms, including Support Vector Machines (SVMs), Random Forrest (RF), decision trees, and boosting-based algorithms. For this setup, logistic regression-based classifiers, specifically the LogitBoost (*LB*) algorithm [4] performed best, so we report classification results for the *LB* algorithm only.

Experimental Setup

We used four-fold cross validation: in each fold, 75% of the SERPs were used for training and the rest for testing. We report the average of the results across the folds. Five- and Three-fold cross validation setups resulted in similar classifier performance. Note that we always split the data by <SERP, Participant> pair, to avoid putting mouse points from the same SERP into both training and test sets, which could result in artificially high classification accuracy.

Experimental Results

We report the main experimental results in Table 4. The *LB* classifier (our method) significantly outperforms *Baseline* for the *T* value of 200, exhibiting the accuracy of 74% compared to the Baseline system (majority class=*InFocus*) accuracy of 55%. Interestingly, reducing the threshold *T* to 150 and 100 pixels (for more fine-grained prediction), allows our system to maintain the accuracy of 73% and 77% respectively,

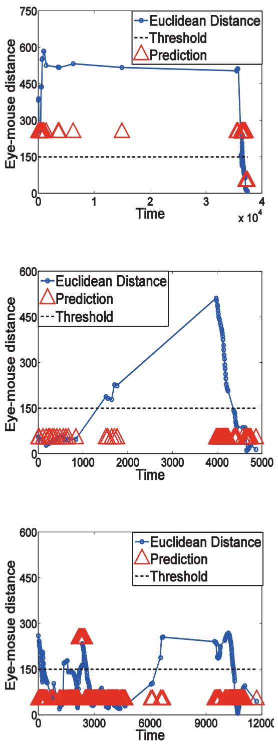

Figure 4. Predictions (red triangles) for *incidental*, *bookmarking*, and *follow* mouse use examples, for *T*=150. The triangles *above* *T* (dotted line) are predictions for the *Away* class; triangles below *T* predict *InFocus* class. Blue lines show the actual eye-mouse distances. Correct predictions are those on the same side of the dotted line as the actual distance.

compared to the *Baseline* (majority class=*Away*) with the accuracy of 58% and 74% respectively. In particular, with the distance threshold *T*=100 pixels, *LB* is able to correctly identify when the eye and mouse positions are coordinated (*InFocus*), with nearly 60% precision; with *T*=150, the Precision and Recall of predicting coordination increase to 68% and 65%, respectively.

Threshold (T)	*Class* / *Method*	*Both*		*InFocus*	
		Acc	*F1*	*Prec*	*Rec*
100	Baseline (*Away*)	73.8	42.5	0	0
	LB	**76.7**	**66.1**	**58.1**	**39.6**
150	Baseline (*Away*)	58.2	36.8	0	0
	LB	**72.5**	**71.6**	**67.9**	**64.8**
200	Baseline (*InFocus*)	55.1	35.5	55.1	**100**
	LB	**74.3**	**73.6**	**74.4**	81.0

Table 4. Classification results for predicting eye-mouse distance

To understand the contribution of the different features, we computed the Information Gain (IG) for each. For the prediction task with the threshold *T*=100, we find that x_i (mouse X-coordinate) contributes the most, followed by t_i (time elapsed since starting to view the results), *pID* (participant ID), y_i (mouse Y-coordinate), Δt_i (time since last mouse move), followed by the velocity and direction features. We illustrate prediction performance for the earlier examples in Figure 4 (sidebar). Our approach appears to be most effective for detecting the *incidental* and *follow* patterns; however, detecting the *bookmarking* cases may require richer feature representation, which is the subject of our future work.

Conclusions and Future Work

We explored the feasibility of predicting eye-mouse coordination using the features derived from mouse movements. Our experiments show that we can predict the regions where the eye and mouse position are within 100 pixels of each other, with the accuracy of nearly 77%. However, there is still much room to improve the accuracy and robustness of our prediction methods.

As the natural next steps, we are exploring a richer feature representation of mouse movement and other result-page interactions. We then plan to use our predictions to identify behavioral patterns such as bookmarking. Finally, we plan to incorporate our predictions as more accurate feedback into real search tasks.

Acknowledgements

We thank Yahoo! Research for partially supporting this work through a faculty research grant, and Charles L. Clarke for valuable discussions.

References

[1] Buscher, G., Dengel A., van Elst, L. Query expansion using gaze-based feedback on the subdocument level. *In SIGIR 2008*

[2] Buscher, van Elst, L., and G., Dengel A. Segment-level display time as implicit feedback: a comparison to eye tracking. In *SIGIR 2009*

[3] Cutrell, E., and Guan, Z. What are you looking for? an eye-tracking study of information usage in web search. In *CHI 2007*

[4] Friedman, J., Hastie, T., and Tibshirani, R. Additive logistic regression: A statistical view of boosting. *In Ann. of Stat. 1998*

[5] Guo, Q., and Agichtein, E. Exploring mouse movements for inferring query intent. In *SIGIR 2008*

[6] Guo, Q., Agichtein, E., Clarke C., and Ashkan, A. In the mood to click? Towards inferring receptiveness to search advertising. In *WI 2009*

[7] Joachims, T., Granka, L., Pan, B., Hembrooke, H., and Gay, G. Accurately interpreting clickthrough data as implicit feedback. In *SIGIR 2005*

[8] Piwowarski, B., Dupret, G., and Jones, R. Mining user web search activity with layered bayesian networks or how to capture a click in its context. In *WSDM 2009*

[9] Radlinski, F. and Joachims, T. Query chains: learning to rank from implicit feedback. In *KDD 2005*

[10] Rodden, K. , Fu, X., Aula, A., and Spiro, I. Eye-mouse coordination patterns on web search results pages, In *Ext. Abstracts, CHI 2008*