Modeling User Interactions for Complex Visual Search Tasks

Category: Research

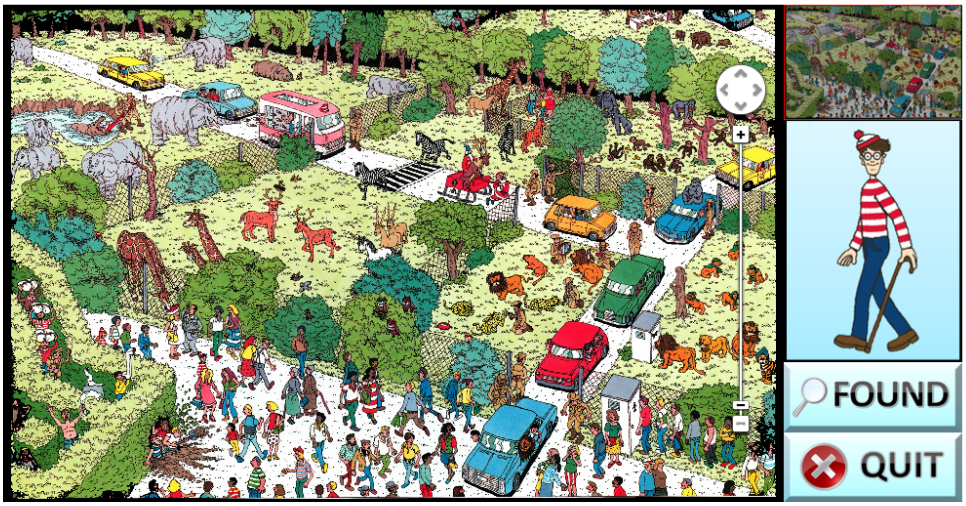

Figure 1: Screenshot of Where's Waldo [2] search task and interface used in the study.

ABSTRACT

Modeling a user's interactions is intimately tied to many areas of research in the fields of HCI and Visual Analytics. Most notably, developing adaptive visual interfaces and effectively prefetching for large datasets, first requires understanding the user's behavior and analytical process. In this work, we demonstrate the potential of using a user's mouse movements and clicks to achieve this goal. In an online study, we gather users' interactions as they perform a complex visual search task. Our results indicate a significant difference between the search strategies employed by users who were quick at completing the task and those who were slow.

1 INTRODUCTION

Thomas and Cook stated that "the goal of visual analytics is to create software systems that will support the analytical reasoning process" [5]. Understanding users' analytic processes is an integral part of developing successful analytic tools. While the community has developed general models of users' analytic processes [4], modeling the processes of individual users or even groups of users has proven to be a difficult challenge. Ideally, this would mean being able to monitor and model user's cognitive process as they solve problems using visual analytics systems. In lieu of directly monitoring cognitive processes, the community proposed the use of interactions as a means of analytic provenance - understanding users' analytic process through interaction analysis [3]. In this work, we demonstrate the potential of using a user's mouse movements and clicks to achieve this goal. By developing mappings between cognitive process and mouse interactions, we posit one can unobtrusively develop models of users' analytical processes.

2 EXPERIMENT

We performed an online experiment, collecting interaction data as users play *Where's Waldo*, a complex visual search task. *Where's Waldo* is a famous children's entertainment series comprised of illustration spreads in which children are asked to locate the character Waldo. While Waldo is usually attired in a distinct red and white striped pattern, he is sometimes hidden behind objects and the illustrations are filled with *red herrings*, specifically designed to mislead and distract the user.

To successfully locate Waldo, users have to visually filter unimportant data, making the task a very complex one. We chose this task because it is easy to understand yet analogous to many other typical visual search task such as locating an item of interest on a map or identifying a data point in a series of data. In this experiment, strategies that result in the user successfully completing the task in 500 seconds or less are defined as *quick*, whereas strategies that require more than 500 seconds before leading to successful completion are defined as *slow*. The purpose of this study is to determine whether patterns in click events and mouse movements can be used to classify a user's search strategy as either *quick* or *slow*.

During the experiment, participants were presented with a Where's Waldo poster and were asked navigate with the image by clicking the interfaces' control bar (Figure 1). The control bar afforded six interactions: zoom in, zoom out, pan left, pan right, pan up and pan down. The interface also included two function buttons, **Found** and **Quit**. When the target is found, the participant is instructed to first click on the target and then click **Found**.

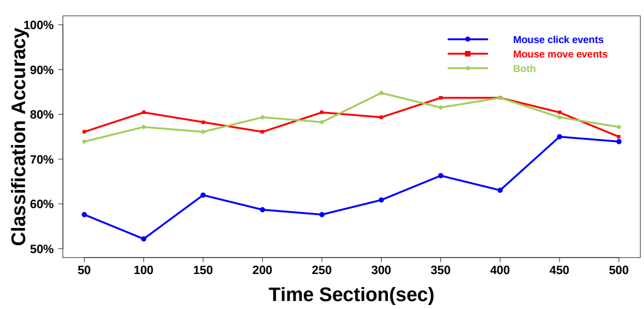

Figure 2: Accuracy of classification in different time intervals using leave-one-out cross validation.

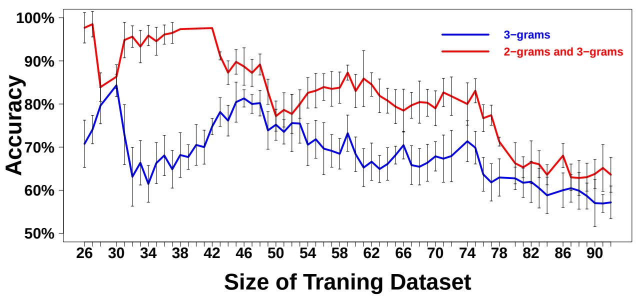

Figure 3: Average accuracy of decision trees applied to 3-grams, and combination of 2-grams and 3 grams for different size of traning dataset using ten runs of 10-folds cross validation.

2.1 Participants

We recruited 130 online volunteers to participate in our pilot study. Participants were required to have basic skill to operate a computer and had not seen the poster in experiment before. Additionally, to avoid repetition, they were only allowed to complete the task once. Participants' age ranged between 18 and 30. Of the 130 recruited participants, 92 successfully completed the task. The average time taken the complete the task was 486.3 seconds and ($\sigma = 316.9$). The results presented in the remainder of this paper are based on the data collected from participants that successfully completed the task.

2.2 Data Collection

In order to build model based on user mouse interaction, we recorded both mouse click and mouse move events. In our study, mouse click events happened when a user clicked a mouse during the searching task. When this occurred, we recorded the position of mouse cursor in relation to the interface and the timestamp. Similarly, we also recorded the interface coordinates of a mouse cursor and the timestamp for ever mouse move event. We then calculated the features shown in Table 1 below and used these calculated features to performed our analysis.

Click Event Features	Mouse Movement Features
Avg. # of clicks	Average # of movements
Avg. time between clicks	Pairwise Euclidean distance (μ, σ, μ_3')
% Left	Pairwise x distance (μ, σ, μ_3')
% Right	Pairwise y distance (μ, σ, μ_3')
% Up	Pairwise speed (μ, σ, μ_3')
% Down	Pairwise angle (μ, σ, μ_3')
% Zoom in	
% Zoom out	
% Found	
% Quit	
% Canvas	

Table 1: Features extracted for analysis.

3 Results

For our analysis, we viewed mouse click events and mouse movements as distinct behavioral measures. When users click on the buttons on the control bar, they saw a different region of the illustration and could gather new information by exploring these regions. Thus we view mouse clicks as a means of getting information about a user's search strategy, while mouse movements can be used as a proxy for user's eye gazes [1]. Therefore we used two different methods to analyze the features from mouse click events and mouse movements dataset, and performed our analysis as such. In the following sections, we will discuss our analysis of the models generated using these features, and demonstrate their utility in partitioning successful strategies.

3.1 Data Analysis of Extracted Mouse Features

Users' behavior varies with time, so we decided to use the extracted features in time intervals of increasing size to model the user interaction. We started the analysis with the features extracted in the first 50 seconds and repeated the process by increasing the size of the time interval by 50 seconds till we reach 500 seconds. We built a Support Vector Machine (SVM) based classification model for each of these time intervals to classify the users into two groups: users who finished the task in 500 seconds and users who did not. The accuracy of SVM classification of these features using Leave-one-out cross-validation is showed in Figure 3.

3.2 Data Analysis of Sequential Data

Another method we employed to build models based on the mouse click events is n-grams, which is usually applied for analyzing a sequence of text. We created text of click events by concatenating characters that represent the each of 8 click interactions (zoom in, zoom out, pan left, pan right, pan up, pan down, quit and found). We then calculated the n-grams for each user and also calculated the number of occurrences for each n-gram and created a feature vector based on these pairs of data.

In this case, we applied decision trees to find subsequences that can be used to distinguish the users into two groups based on the mean completion time. Using the normal distribution as an approximation of the sample distribution, we first classified the most 26 extreme users i.e. users with completion times greater than one standard deviation away from the mean. We performed the classification using 2-grams and 3-grams, which yielded an average accuracy (out of 100 runs) of 97.69% and 70.77% respectively. We then adjusted our cutlines by rate of 1% of the standard deviation. The results of these classification is demonstrated in figure 3.

4 Conclusion

We demonstrated the potential of using a user's mouse movements and clicks to separate users into two groups: those who were quick at completing the task and those who were slow at completing the task. We believe that this lays the foundation for using mouse interaction data as a means of modeling user strategies for visual search tasks.

References

[1] M. C. Chen, J. R. Anderson, and M. H. Sohn. 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*, pages 281–282. ACM, 2001.

[2] M. Handford. *Where's Waldo?* Little, Brown Boston, 1987.

[3] W. A. Pike, J. Stasko, R. Chang, and T. A. O'Connell. The science of interaction. *Information Visualization*, 8(4):263–274, 2009.

[4] P. Pirolli and S. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of International Conference on Intelligence Analysis*, volume 5, 2005.

[5] J. Thomas and K. Cook. *Illuminating the path: The research and development agenda for visual analytics*, volume 54. IEEE, 2005.