A Large-Scale Study of Mobile Search Examination Behavior

Xiaochuan Wang[†], Ning Su[†], Zexue He[‡], Yiqun Liu[†][*], Shaoping Ma[†]

[†]Department of Computer Science & Technology, Tsinghua University, Beijing National Research Center for
Information Science and Technology, Beijing, China
[‡]Beijing Normal University, Beijing, China
yiqunliu@tsinghua.edu.cn

ABSTRACT

With the rapid growth of mobile web search, it is necessary and important to understand user's examination behavior on mobile devices in the absence of clicks. Previous studies used viewport metrics to estimate user's attention. However, there still lacks an in-depth understanding of how search users examine and interact with the mobile SERP. In this work, based on the large-scale real search log collected from a popular commercial mobile search engine, we present a comprehensive analysis of examination behavior. Specifically, we analyze the position bias, the relationship with click behavior, and examination's change as the session continues. The findings shed new light on the understanding of user's examination behavior, and also provide some implication for the improvement and evaluation of mobile search engine.

KEYWORDS

Examination; Viewport; Log Analysis; Mobile Web Search

1 INTRODUCTION

Understanding user's examination behavior in web search is important because it provides valuable insights on snippet attractiveness [5], result relevance [8], ranking performance, and opportunities for advertising and the design of the interaction. In desktop settings, eye-tracking is widely used as an approach to studying user examination behavior. It provides rich details on user attention by giving fairly accurate estimates of eye gaze positions in pixel level. Meanwhile, some previous studies find strong eye-mouse correlations, suggesting that we may reasonably infer user attention by tracking mouse behavior, such as moving, scrolling and hovering.

Recent years have witnessed a rapid growth of mobile web search. According to surveys, mobile web browsing overtook desktop for

---

[*]Corresponding author

the first time in October 2016[1], and in the U.S. nearly 60 percent of searches were made on mobile devices[2]. Considering mobile devices are touch enabled with small displays, user interaction behaviors on mobile devices are different from those on the desktops. For example, touch events have different meanings from cursor events [4]. Using touch coordinates on mobile devices as a translation of cursor coordinates may not be reliable because users may view SERPs without touch. Another difference is that mobile search engines often present answers or relevant information (in the form of knowledge graph) on the SERP in response to user's simple information needs that are popular on mobile [6]. Thus, users can be satisfied by the snippet and end the search session without the need to click on any results (i.e. *good abandonment* [10]). Since clicks have been widely used to infer result relevance and search satisfaction, the absence of clicks makes it difficult to evaluate the quality of both documents and search engines in the mobile environment. In these cases, the examination is the only interaction behavior we can use to understand search users and further make quality evaluation. Therefore, understanding user's examination behavior on mobile devices becomes increasingly necessary and important.

Viewport, the visible portion of a web page, has been widely studied in mobile search. In [6], by simultaneously recording eye gaze and viewport data, it proved that viewport is strongly correlated with user attention, indicating one can track viewport to estimate user's examination behavior. Therefore, many previous studies used viewport data to understand mobile search users, such as measuring user satisfaction [3, 6], understanding user attention with rich ad formats [7] and detecting good abandonment [10]. However, it is still unclear how users examine SERPs on mobile devices, e.g. *How deep will users view? How long will users examine a result? What is the relationship between click and examination?* Meanwhile, most of the previous studies carried out user studies to collect viewport data, which are limited in scale.

In this work, with the help of a popular commercial mobile search engine, we collect large-scale real search log, including user's click and viewport data. Based on the log, we provide a comprehensive analysis of user's examination behavior on the first SERP. In particular, we focus on the following three aspects:

- The position bias in user's examination behavior.
- The relationship between click and examination behavior.
- The change in examination behavior as the session continues.

We hope the findings in this study can give some inspiration on the understanding of search users in mobile search, and better apply studies in web search to the mobile environment.

---

[1]http://gs.statcounter.com/press/mobile-and-tablet-internet-usage-exceeds-desktop-for-first-time-worldwide
[2]http://hitwise.connexity.com/070116_MobileSearchReport_CD_US.html

2 RELATED WORK

There have been some studies focus on user's attention and the usage of viewport data. Lagun et al. [6] first studied whether tracking the viewport on mobile phones could enable accurate measurement of user attention at scale. Through quantitative eye tracking and viewport tracking user study, they found strong correlations between viewport metrics and user attention. Following this work, viewport data begin to be widely used in the study of mobile search in the absence of clicks. Williams et al. [10] extracted a series of gesture features based on the viewport data to detect good abandonment in mobile search. Guo and Song [3] present large-analysis of viewing behavior in mobile proactive systems based on the viewport. They identified a variety of biases may influence the viewing behavior, and demonstrated that the viewport-based metrics are more effective compared to click-based and dwelltime-based metrics in measuring user satisfaction with the proactive systems. Lagun et al. [7] studied how users' eye gaze and satisfaction are impacted by the presence of ads and their formats and improved accuracy of existing viewport-based gaze inference methods.

Compared to these studies, our work focuses on studying examination behavior in traditional mobile web search. Different from most of them, our dataset is collected from real search log, rather than controlled user studies, thus the findings can be generalized on a large scale.

3 DATASET

From a popular commercial mobile search engine, we collect user interaction log recorded viewports and clicks. Viewport logging is enabled through JavaScript inserted into every SERP shown to the users, and the viewport data is buffered and subsequently sent back to the server through HTTP requests. We record the position and the size of the viewport, the bounding boxes of the search results in the viewport, as well as the viewport change events with timestamps. All the events are listed in Table 1. Meanwhile, we also record search results on the first SERP with click timestamps for clicked ones. These records allow us to reconstruct examination and click behavior of users.

We first randomly sample 10% of the daily search users during one week in January 2018. Then, we randomly sample 1% of their interaction log as our dataset. It contains 1,473,919 searches in total. In this work, we only study user examination behavior on the first SERP, so we first count the number of results on the first SERP for each search in the dataset. As we can see in Table 2, about half of the first SERP (45.43%) have 13 results. Therefore, we make the analysis of these log records as representative. It is worth noting that we also analyze interaction log with 11, 12 and 14 results on the first SERP and obtain the consistent conclusions. To save space, we only show the analysis for interaction log with 13 results on the first SERP in the following sections.

Table 1: Viewport change events

Events	Description
Loaded/Leave	Users load or leave SERP.
Scrollstart/Scrollstop	Users start or end scrolling.
Visible/Hidden	SERP is visible or hidden. It is recorded when users switch tabs or Apps.

Table 2: The number of results on the first SERP

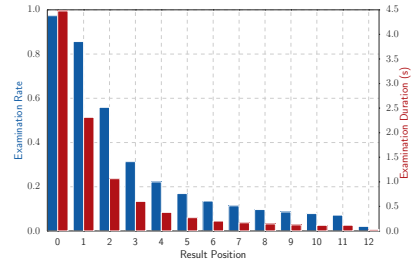| Number	1-5	6-10	11	12	13	14	≥15
%	9.65	10.74	9.10	11.68	45.43	10.80	2.59

Figure 1: Average examination rate and duration by position

4 ANALYZING EXAMINATION BEHAVIOR

In this section, we present a comprehensive analysis of user's examination behavior using collected search log. In particular, we analyze the position bias, the relationship with click behavior, and the examination's change during a search session.

4.1 Examination v.s. Position Bias

The position bias has been widely studied in both desktop and mobile search. It is easy to think that results in the higher positions have a higher probability to be examined. In this section, we compare the examination rate, the examination duration, the moving direction and end probability by position.

Examination Rate. Similar with click through rate, we first analyze the examination rate of each position, which is defined as the ratio of the number of results at a certain position that ever appears in a viewport to the number of total search sessions, with a maximum of 1. In Figure 1, we can see an obvious positional decay. The examination rate of the first position is not equal to 1. It is because users may end the search before the web page is loaded for some reasons (e.g. typing errors, automatic shutdown). The top two positions have relatively close rates because the initial viewport usually contains two results. In addition, we can see a sharp decay between the last two positions. We believe this may be due to the special vertical type of the last results, which is less attractive to search users. Therefore, they often scroll the SERP immediately as soon as they see these vertical results.

Examination Duration. Using the viewport change events with timestamps, we can get the examination duration of each viewport. Following Lagun et al.'s work [6], we then calculate weighted duration on each result based on its exposure and coverage, where exposure is defined as how much of the result area was visible to a user and coverage is defined as how much of the viewport real estate did the result occupy. The examination duration of each result is computed as the sum of weighted duration across all viewports. Through the statistical analysis, 90% of the examination durations are no more than 12.5 seconds. To avoid statistical bias, we set a time threshold of 30 seconds, which is longer than 95% of the durations, and correct the abnormal durations (exceed the threshold) to 30 seconds. Similarly, we can see a positional decay of examination duration in Figure 1. In contrast to examination rate, the difference between the top two positions is much larger, suggesting users tend to examine first result more in the initial viewport.
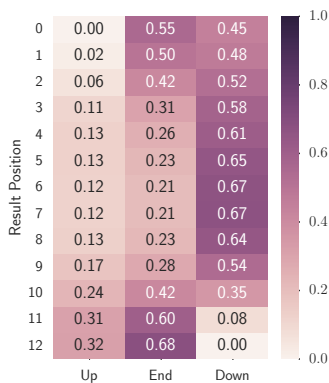
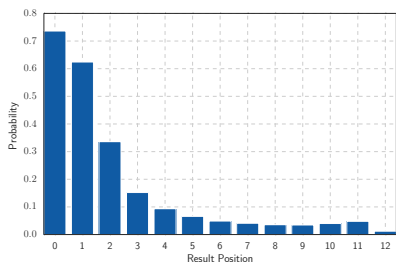**Figure 2: Moving probability in three directions by position**

Figure 3: Probability of the result appearance in the last viewport by position

Due to the special vertical type, the examination duration is also very short (about 28ms).

Moving Direction. We now investigate how the examination moves. When finishing the examination of a viewport, users can swipe up, swipe down or end the search session, and the viewport will move with it. These actions can reflect users' different states, such as dissatisfied or need to revisit previous results. Figure 2 shows the probabilities of three moving directions at each position. The *Up* probability of 0.13 at position 4 means 13% of the viewports containing the result at position 4 will move up. From the figure, We can get some interesting findings. First, both the *End* probabilities and the *Down* probabilities of top two positions are about 50%, indicating in nearly half of the sessions, users only examine the initial viewport. The *End* probabilities of the bottom two results indicate that users have more than 60% tendency to end the search when they examine the last few results. The *Down* probabilities of the middle results (at position 4-8) are high, and almost the same. It means that if a user does not end the search after examining the top results, he or she will have a high probability of constantly examining the following results. The *Up* probabilities are very low except bottom results, indicating the revisit behavior is quite few in mobile search.

End Position. We also investigate where users end examination behavior. Figure 3 shows the probability of the results at each position appear in the last viewport. We can see a similar result with examination rate. However, the probabilities of the bottom half of the positions are almost the same (except the last one, i.e. position 6-11), which can also be seen in examination duration (Figure 1), suggesting that users are not sensitive to the result's position in the bottom half of the SERP.
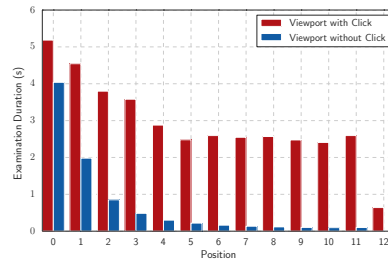

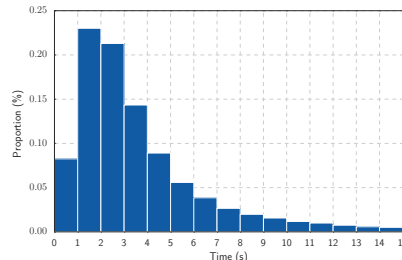**Figure 4: Average examination duration w.r.t click**

Figure 5: The time between the beginning of the viewport and the click event

4.2 Examination v.s. Click

According to *Examination Hypothesis* [9], a user will click on a result when he or she has examined it and consider it as relevant. In this section, we try to give the relationship between examination and click by analyzing the click behavior in the viewport.

Examination Duration. We first compare the examination durations between viewports with or without clicks. The examination durations of viewports with clicks (median 4.0s, IQR 2.0-11.2s) are longer than those without clicks (median 1.6s, IQR 0.8-3.8s) in general. Figure 4 shows that users tend to examine the clicked results longer at each position. And on average, the examination duration at each position (except position 12) is longer than 2.5 seconds. Similarly, the position bias for clicked results' examination duration is not obvious in the bottom half of the SERP.

Click Time. To better understand how long users need to examine a result that may be clicked on, we calculate the time between the beginning of the viewport and click event. From Figure 5, we can see nearly 80% of the time is between 0-5 seconds, suggesting that 5 seconds is basically enough for a user to examine a result. Therefore, it is reasonable to believe that if a user examines a result nearly 5 seconds, he or she may have a high probability to click the result, or the snippet already provides enough information, which may lead to a *good abandonment*.

Click Position. In [6], the authors found that about 70% of users' attention is focused on the top half of the screen, with little attention paid to the bottom 1/3 portion. Here, using the click position in the viewport, we can roughly estimate the position of users' attention. We first normalize the visible height of the result list in each viewport to 100 grids. For each clicked result, we get the occupied grids by its bounding box in the viewport, and increase the number of clicks on these grids by one. Then, using all the clicked results, we can get the click through rate of each grid. As the heatmap in Figure 6 shows, users tend to click results in the top 2/3 portion of the visible result list. This can also roughly indicate user's examination position, and is consistent with the findings in [6].
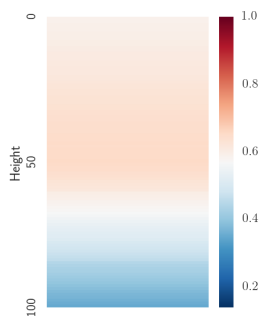
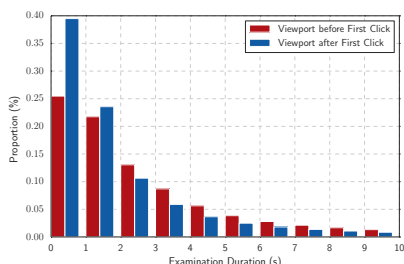**Figure 6: Heatmap for click position in the visible result list**

Figure 7: The distribution of viewport's examination duration before or after the first click

4.3 Examination Change

Users' *benefit* and *cost* are associated with their interactions [1], and are constantly changing with the search session. In this section, we investigate examination's change with these two factors.

Benefit. *Benefit*, also referred to as *gain* or *utility*, has been discussed and introduced in a variety of ways. Based on *Cascade Model* [2], which assumes a linear traversal through the ranking, and that documents below a clicked result are not examined, we use *the first click* to denote user's benefit in the search. We compare the examination durations before and after the first click. As Figure 7 shows, after the first click, nearly 40% of the examination durations are within 1 seconds, and 24% are within 1-2 seconds. In contrast, more than 50% of the examination durations are above 2 seconds before the first click. It means that user's examination duration decreases with the increase in user's benefit, which can also be partly explained by *Cascade Model*.

Cost. We consider temporal efforts as user's cost and investigate the examination duration's change in the first 10 minutes of the search. As the scatter plot in Figure 8 shows, the possibility of longer examination duration begins to reduce since about 2 minutes. It indicates that user's patience decreases over time, thus he or she will spend less time on the later results. And this can also explain the position bias to some extent.

5 CONCLUSION AND DISCUSSION

With the dataset collected in a popular commercial mobile search engine, we present a comprehensive analysis of user's examination behavior on the first SERP, from the aspects of position bias, the relationship with click behavior and its change as the session continues. To summarize, we find user's examination behavior is influenced by position, click behavior and user's benefit and cost. Users tend to examine the few top more carefully. Clicked results
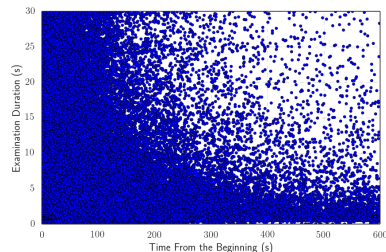

Figure 8: The distribution of viewport's examination duration over time

get longer examination duration, and these results usually appear in the top 2/3 portion of the viewport. The examination duration decrease with the increase of both user's benefit or cost.

Our findings not only enriches the understanding of the examination behavior, but also provide some implications for the improvement and evaluation of mobile search engine. For example, we find that users tend to either end the search with top few results or constantly scroll down to the end, and the position bias is not obvious in the bottom half of the SERP. It suggests that we should design better evaluation metrics in the mobile environment, which consider different position bias in the different part of the SERP. The findings in Sec 4.2 suggest that 5 seconds may be an acceptable threshold to identify good abandonment. Meanwhile, we can also try to regard it as a click when design click models specifically for mobile search to deal with the sparse of clicks.

6 ACKNOWLEDGEMENTS

REFERENCES

[1] Leif Azzopardi and Guido Zuccon. 2016. An analysis of the cost and benefit of search interactions. In *Proceedings of the 2016 ACM on International Conference on the Theory of Information Retrieval*. ACM, 59–68.

[2] Nick Craswell, Onno Zoeter, Michael Taylor, and Bill Ramsey. 2008. An experimental comparison of click position-bias models. In *Proceedings of the 2008 international conference on web search and data mining*. ACM, 87–94.

[3] Qi Guo and Yang Song. 2016. Large-Scale Analysis of Viewing Behavior: Towards Measuring Satisfaction with Mobile Proactive Systems. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*. ACM, 579–588.

[4] Jeff Huang and Abdigani Diriye. 2012. Web user interaction mining from touch-enabled mobile devices. In *HCIR workshop*.

[5] Dmitry Lagun and Eugene Agichtein. 2011. Viewser: 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*. ACM, 365–374.

[6] Dmitry Lagun, Chih-Hung Hsieh, Dale Webster, and Vidhya Navalpakkam. 2014. Towards better measurement of attention and satisfaction in mobile search. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*. ACM, 113–122.

[7] Dmitry Lagun, Donal McMahon, and Vidhya Navalpakkam. 2016. Understanding mobile searcher attention with rich ad formats. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*. ACM, 599–608.

[8] Vidhya Navalpakkam, LaDawn Jentzsch, Rory Sayres, Sujith Ravi, Amr Ahmed, and Alex Smola. 2013. Measurement and modeling of eye-mouse behavior in the presence of nonlinear page layouts. In *Proceedings of the 22nd international conference on World Wide Web*. ACM, 953–964.

[9] Matthew Richardson, Ewa Dominowska, and Robert Ragno. 2007. Predicting clicks: estimating the click-through rate for new ads. In *Proceedings of the 16th international conference on World Wide Web*. ACM, 521–530.

[10] Kyle Williams, Julia Kiseleva, Aidan C Crook, Imed Zitouni, Ahmed Hassan Awadallah, and Madian Khabsa. 2016. Detecting good abandonment in mobile search. In *Proceedings of the 25th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 495–505.