Modeling Search Behavior in Heterogenous Environments

Position Paper

Ilya Markov† i.markov@uva.nl
Aleksandr Chuklin† a.chuklin@uva.nl
Yiqun Liu‡ yiqunliu@tsinghua.edu.cn

†University of Amsterdam, Amsterdam, The Netherlands
‡Department of Computer Science and Technology, Tsinghua University, Beijing, China

ABSTRACT

A large part of today's information retrieval studies focuses on modeling user search behavior to improve search algorithms, evaluate their quality and simulate users. Most of previous research on user modeling considers homogeneous search environments, where search devices, users, their search tasks and corresponding search results are assumed to be uniform. However, real-world search environments are highly heterogeneous.

In this position paper, we elaborate on heterogeneous search environments and outline challenges in user modeling that arise in this case, e.g., the availability of rich user feedback and the need of device-specific and cross-device models. We also set up an agenda for research on user modeling in heterogeneous search environments, which includes acquisition of user interaction data, development of advanced user models and application-oriented evaluation of these models.

1. INTRODUCTION

Understanding and modeling user search behavior is crucial for the development of search systems. User models help to infer relevance, evaluate the quality of search systems, simulate users, improve ranking algorithms, etc.[1] Most of previous work on modeling user behavior assumed homogeneous search environments, i.e., did not distinguish between various devices, users, search tasks and search results. Moreover, user models appearing in previous studies were quite narrow in a sense that they mainly focused on modeling clicks (a.k.a. click models). However, nowadays the search process is highly heterogeneous [26]: search devices range from desktop computers to mobile phones and dedicated devices; users have different preferences in terms of queries, document relevance, results presentation and behavioral patterns; search tasks span multiple search sessions and involve multiple devices; search results contain objects of different media types, direct answers, entities, etc. Also, users leave rich behavioral traces, such as mouse movements, page scrolls, various gestures on mobile devices, etc.

Recently, researchers attempted to address some of the above issues in user modeling. Several studies considered query intents within click models [5, 12]. Others modeled user click behavior on aggregated SERPs, i.e., SERPs containing web and vertical results [3, 23]. Several studies modeled mouse movements [7, 18], developed personalized click models [22, 24] and modeled multi-query search behavior [8, 25].

However, these are only the first steps towards understanding and modeling user behavior in heterogeneous search environments, while the following problems are still to be addressed. First, various types of implicit user feedback beyond clicks have to be captured and modeled, e.g., mouse movements, page scrolls, touch gestures, etc. Second, user search behavior has to be studied not only on desktop and laptop computers, but also on increasingly more popular mobile devices, devices with limited interaction capabilities (e.g., TVs) and dedicated devices (e.g., GPS navigation systems). Third, cross-session and cross-device models have to be built to capture user behavior within long-term search tasks. Finally, user models have to consider that modern search systems present users not only with web and vertical results, but also with direct answers, entities, quick links, etc., where the composition and layout of these objects is highly personalized. Given recent advances in user modeling, we believe that the time is right to move forward and address the above-mentioned problems.

2. CHALLENGES IN USER MODELING

2.1 Rich User Feedback

Existing models of search behavior (e.g., [2, 9, 10]) focus mainly on result clicks, because this type of implicit user feedback is easy to capture and interpret. However, other types of feedback, such as mouse movements, page scrolls and pagination clicks, provide valuable additional information, which helps to refine existing click models (e.g., [4, 13]) or can be used as a primary source of user feedback (e.g., [7, 18]), especially in cases where a user does not click on any result.

Devices beyond desktop computers provide other means of interaction with search results. Mobile devices mainly use the touch technology, which substitutes mouse movements and page scrolls and supports a number of gestures. Also, it was shown recently that a SERP viewport (visible portion of a web page) is an important signal of searcher's attention on mobile devices [17]. Other devices, such as TV-sets, provide limited interaction capabilities, but may have additional options, such as preview, which are not available in standard search environments.

The main challenges here is to capture and interpret the above-mentioned user signals. For example, when capturing mouse movements and touch gestures, one needs to map the position of a cursor/finger on a screen to an underlying object on a SERP. Also, while some user signals have relatively clear interpretation (e.g., clicks, scrolls, zooming), others do not (e.g., mouse movements).

[1] By *user model* we mean a probabilistic model that describes observed user search behavior (e.g., clicks) and can be trained based on this observed data. See [2, 9, 10] for classic examples.

HIA'15, February 6, 2015, Shanghai, China.
ACM 978-1-4503-3317-7/15/02.

2.2 Device Type

Apart from the fact that different devices provide different means of interaction with search results, devices differ in their technical characteristics and main function. Desktops and laptops are general-purpose devices with rich interaction capabilities (large screen, keyboard, mouse, powerful browsers, etc.), so they can be used for most types of search. Smartphones and tablet computers, on the other hand, are much more mobile and have additional components, such as touch screen, microphone and GPS, which make them perfect for specific types of search, e.g., local search. Search systems on devices, such as TV-sets and GPS-navigation devices, are even more specific and may provide only results of a certain type. Thus, the device type determines not only the way of interaction with search results, but also user search intents.

The notion of successful search is also different on different devices. In general search, a successful session is usually the one with many clicks followed by high dwell time. In TV search, however, a successful session could be the one ended with a purchase of a TV program. Similarly, a successful search session on a GPS-navigation device should probably be followed by navigation to a point of interest.

The above shows that user search behavior differs considerably between devices. Therefore, user models have to carefully consider special properties of each device and corresponding user intents and indicators of success.

2.3 Search Tasks

Often, users have long-standing search tasks, e.g., planning a trip or performing research on a certain topic. Such tasks span multiple (not necessarily consecutive) search sessions and may involve multiple devices [15, 19]. Although several studies attempted to model user behavior in multi-query search sessions [8, 25], no work exists on modeling cross-session and cross-device search behavior.

Several challenges arise here. First, search tasks have to be identified and search sessions have to be mapped to corresponding tasks [14]. Cross-device search makes this step particularly difficult [19]. Second, search behavior has to be connected across sessions, changes over time and device types have to be detected. Finally, cross-session and cross-device user models have to be developed.

2.4 Heterogeneous SERPs

Modern search systems present users with rich SERPs, which contain highly heterogeneous results [26]. Objects on a SERP can be classified into three categories: query-related, result-related and general. Query-related objects include the query interface itself (e.g., a text-box), query suggestions, related searches and spelling correction. Result-related objects include but are not limited to web results (usually contain title, URL, snippet, additional URLs and info), quick links, vertical results, other types of diversified results (e.g., fresh content), direct answers, entity pane (usually contains image, snippet, structured info, related entities, etc.), ads and even additional query interfaces for deep web resources. Finally, general objects are tabs, pagination links, settings and account-related objects.

Existing user models mainly deal with web results only [2, 9, 10]. Several studies also considered diversified results [5, 12] and vertical results [3, 23]. However, the actual richness of modern SERPs is largely overlooked. When studying search behavior on such complex SERPs, we need to understand how each component of a SERP affect this behavior, what is the joint effect of multiple components, how the SERP's layout influence the way users perceive search results. Moreover, user models have to take into account that the type, number and layout of components are highly personalized and change from user to user.

3. RESEARCH AGENDA

3.1 Datasets

Most studies on user modeling employ search logs to train and evaluate proposed approaches. Current publicly available datasets include search logs, released by Yandex within the WSCD workshop series, and a click log provided by Sogou. In particular, the WSCD2012 dataset[2] includes clicks and relevance judgements for web search results, WSCD2013[3] additionally contains search engine switching actions [20] and WSCD2014[4] includes user ids [21]. Similarly, the SogouQ dataset[5] contains user ids, clicks and relevance judgements for web search results.

The above datasets substantially advanced research on user modeling. However, they are mostly limited to a homogeneous search environment, namely web search, and to click data. The next step is to mine search interactions in heterogeneous environments, i.e., on different devices, for different search tasks, with various types and layouts of results. Additionally, we need to collect and make available search logs containing user interactions beyond clicks. For example, Liu et al. [18] recently published a dataset, which, in addition to clicks and relevance judgements, contains eye fixation information.[6] One big challenge here is privacy: the more detailed search log is released, the more sensitive it can get. Researchers always have to find a tradeoff between the level of details and the sensitivity of a dataset.

At the same time, alternative approaches to training and evaluating models of user search behavior should be developed. The recently launched LivingLabs initiative[7] is one of such approaches, which allows researchers to inject their results into a production search system and collect implicit user feedback [1]. Currently, this approach is limited to relatively few queries and specific search systems providing specific feedback, but it opens up possibilities for collecting and using real-world user interaction data in various search environments.

Another direction of research is to develop data mining techniques for capturing user interactions on various devices. Recently, Lagun et al. [16] proposed to mine common motifs from mouse trajectories on desktop computers, while Guo et al. [11] and Lagun et al. [17] captured touch gestures and SERP viewports on mobile devices. Currently, these techniques work on a small scale (especially for mobile devices), but provide a basis for developing scalable data mining methods, which can be used to capture user search behavior on a large scale.

3.2 Models

Next, we need to build models that address challenges outlined in Section 2, i.e., model user interactions beyond clicks, consider search behavior on various devices, model cross-session and cross-device behavior and consider complex SERPs. Here, we will first need to generalize models across devices and interaction types. In particular, models of mouse movement may be suitable to represent certain touch gestures, while click models for web search on

[2] http://imat-relpred.yandex.ru/en/datasets
[3] http://switchdetect.yandex.ru/en
[4] https://www.kaggle.com/c/yandex-personalized-web-search-challenge
[5] http://www.sogou.com/labs/dl/q-e.html
[6] http://www.thuir.cn/group/~yqliu/publications/cikm2014-liu.7z
[7] http://living-labs.net

desktop computers may be adapted to other devices, such as TV sets.

Then, one can extend existing models of search behavior to heterogeneous environments by adding variables representing additional component (e.g., device type). However, we will need to find a trade-off between model complexity and descriptive power in order to keep models interpretable and resistant to overfitting. Finally, new models may be needed for certain devices, cross-session search tasks and particular complex SERPs.

3.3 Evaluation

Models of search behavior are usually evaluated using likelihood and perplexity [9], but a number of application-oriented evaluation techniques are also proposed, such as click-through rate prediction [2] and relevance prediction [8]. However, these methodologies do not capture the possible heterogeneity of search environments and multiple types of implicit user feedback.

Recently, Chuklin et al. [6] proposed a new evaluation methodology for vertical-aware click models, which captures the key aspects of aggregated search, such as vertical selection, item selection, result presentation and vertical diversity. Similar application-oriented evaluation methodologies may be developed for other aspects of heterogeneous search environments, such as search on different devices or cross-session search.

4. CONCLUSIONS

In this position paper we considered the problem of user modeling in heterogeneous search environments, where search devices, users, search tasks and results are not uniform. We showed that while research on user modeling is gradually getting rid of the homogeneity assumption, many major problems are still open. First, rich interaction data, such as mouse movements, page scrolls and device-specific interactions, has to be acquired and modeled. Second, user search behavior on various devices has to be studied, considering the device main function and device-specific indicators of search success. Third, cross-session and cross-device search tasks have to be identified and modeled. Finally, heterogeneous search results, such as verticals and direct answers, have to be considered when modeling user behavior.

Given these problems, we set up the agenda for studying user models in heterogeneous search environments. We argued that a considerable effort has to be put into the development of publicly available datasets, which should contain rich interaction data collected in a variety of search scenarios. Then we outlined that user models addressing the above problems have to be developed. Finally, we argued that evaluation methodologies should be shifted towards specific search scenarios and have to become more application-oriented.

Acknowledgments. This research was partially funded by grant P2T1P2_152269 of the Swiss National Science Foundation.

5. REFERENCES

[1] K. Balog, L. Kelly, and A. Schuth. Head first: Living labs for ad-hoc search evaluation. In *Proceedings of CIKM*, pages 1815–1818, 2014.

[2] O. Chapelle and Y. Zhang. A dynamic bayesian network click model for web search ranking. In *Proceedings of WWW*, pages 1–10, 2009.

[3] D. Chen, W. Chen, H. Wang, Z. Chen, and Q. Yang. Beyond ten blue links: enabling user click modeling in federated web search. In *Proceedings of WSDM*, pages 463–472, 2012.

[4] A. Chuklin, P. Serdyukov, and M. de Rijke. Modeling clicks beyond the first result page. In *Proceedings of CIKM*, pages 1217–1220, 2013.

[5] A. Chuklin, P. Serdyukov, and M. de Rijke. Using intent information to model user behavior in diversified search. In *Proceedings of ECIR*, pages 1–13, 2013.

[6] A. Chuklin, K. Zhou, A. Schuth, F. Sietsma, and M. de Rijke. Evaluating intuitiveness of vertical-aware click models. In *Proceedings of SIGIR*, pages 1075–1078, 2014.

[7] F. Diaz, R. White, G. Buscher, and D. Liebling. Robust models of mouse movement on dynamic web search results pages. In *Proceedings of CIKM*, pages 1451–1460, 2013.

[8] G. Dupret and C. Liao. A Model to Estimate Intrinsic Document Relevance from the Clickthrough Logs of a Web Search Engine. In *Proceedings of WSDM*, pages 181–190, 2010.

[9] G. E. Dupret and B. Piwowarski. A user browsing model to predict search engine click data from past observations. In *Proceedings of SIGIR*, pages 331–338, 2008.

[10] F. Guo, C. Liu, and Y. M. Wang. Efficient multiple-click models in web search. In *Proceedings of WSDM*, pages 124–131, 2009.

[11] Q. Guo, H. Jin, D. Lagun, S. Yuan, and E. Agichtein. Mining touch interaction data on mobile devices to predict web search result relevance. In *Proceedings of SIGIR*, pages 153–162, 2013.

[12] B. Hu, Y. Zhang, W. Chen, G. Wang, and Q. Yang. Characterizing search intent diversity into click models. In *Proceedings of WWW*, pages 17–26, 2011.

[13] J. Huang, R. W. White, G. Buscher, and K. Wang. Improving searcher models using mouse cursor activity. In *Proceedings of SIGIR*, pages 195–204, 2012.

[14] R. Jones and K. L. Klinkner. Beyond the session timeout: Automatic hierarchical segmentation of search topics in query logs. In *Proceedings of CIKM*, pages 699–708, 2008.

[15] A. Kotov, P. N. Bennett, R. W. White, S. T. Dumais, and J. Teevan. Modeling and analysis of cross-session search tasks. In *Proceedings of SIGIR*, pages 5–14, 2011.

[16] D. Lagun, M. Ageev, Q. Guo, and E. Agichtein. Discovering common motifs in cursor movement data for improving web search. In *Proceedings of WSDM*, pages 183–192, 2014.

[17] D. Lagun, C.-H. Hsieh, D. Webster, and V. Navalpakkam. Towards better measurement of attention and satisfaction in mobile search. In *Proceedings of SIGIR*, pages 113–122, 2014.

[18] Y. Liu, C. Wang, K. Zhou, J. Nie, M. Zhang, and S. Ma. From skimming to reading: A two-stage examination model for web search. In *Proceedings of CIKM*, pages 849–858, 2014.

[19] G. D. Montanez, R. W. White, and X. Huang. Cross-device search. In *Proceedings of CIKM*, pages 1669–1678, 2014.

[20] P. Serdyukov, G. Dupret, and N. Craswell. WSCD2013: Workshop on web search click data 2013. In *Proceedings of WSDM*, pages 787–788, 2013.

[21] P. Serdyukov, G. Dupret, and N. Craswell. Log-based personalization: The 4th web search click data (WSCD) workshop. In *Proceedings of WSDM*, pages 685–686, 2014.

[22] S. Shen, B. Hu, W. Chen, and Q. Yang. Personalized click model through collaborative filtering. In *Proceedings of WSDM*, pages 323–332, 2012.

[23] C. Wang, Y. Liu, M. Zhang, S. Ma, M. Zheng, J. Qian, and K. Zhang. Incorporating vertical results into search click models. In *Proceedings of SIGIR*, pages 503–512, 2013.

[24] Q. Xing, Y. Liu, J.-Y. Nie, M. Zhang, S. Ma, and K. Zhang. Incorporating user preferences into click models. In *Proceedings of CIKM*, pages 1301–1310, 2013.

[25] Y. Zhang, W. Chen, D. Wang, and Q. Yang. User-click modeling for understanding and predicting search-behavior. In *Proceedings of KDD*, pages 1388–1396, 2011.

[26] K. Zhou, T. Sakai, M. Lalmas, Z. Dou, and J. M. Jose. Evaluating heterogeneous information access. In *Proceedings of MUBE workshop*, 2013.