Uncovering Task Based Behavioral Heterogeneities in Online Search Behavior

Rishabh Mehrotra
Dept. of Computer Science
University College London
London United Kingdom
r.mehrotra@cs.ucl.ac.uk

Prasanta Bhattacharya
Dept. of Information Systems
National University of Singapore
prasanta@comp.nus.edu.sg

Emine Yilmaz
Dept. of Computer Science
University College London
London United Kingdom
emine.yilmaz@ucl.ac.uk

ABSTRACT

While a major share of prior work have considered search sessions as the focal unit of analysis for seeking behavioral insights, search tasks are emerging as a competing perspective in this space. In the current work, we quantify user search task behavior for both single- as well as multi-task search sessions and relate it to tasks and topics. Specifically, we analyze user-disposition, topic and user-interest level heterogeneities that are prevalent in search task behavior. Our results show that while search multi-tasking is a common phenomenon among the search engine users, the extent and choice of multi-tasking topics vary significantly across users. We find that not only do users have varying propensities to multi-task, they also search for distinct topics across single-task and multi-task sessions. To our knowledge, this is among the first studies to fully characterize online search tasks with a focus on user- and topic-level differences that are observable from search sessions.

CCS Concepts

•**Information systems** → **Task models;**

Keywords

Search tasks; Multitasking; User Behavior

1. INTRODUCTION

Users' information search behavior on search engines often span various motivations [2, 14]. While simple informational needs, such as "Adele's latest music" can be satisfied in a single search session, most other informational needs are complex and time-consuming. Consequently, users accomplish more complex information search tasks by issuing a series of search queries spanning multiple search sessions, possibly spread over multiple days. While a major portion of existing work have investigated search behavior using search sessions as the fundamental focus of search activity [6, 18, 19, 20, 24, 25], more recent studies suggest that users often seek to complete multiple search tasks within a single search session [14, 15, 22], while also taking multiple sessions to finish a single task at times. The emergence of multi-tasking behavior within a single search session makes it particularly complex to use user information from search sessions to personalize the user's search activity. This necessitates a shift in focus from search sessions to search tasks as a more accurate unit of analysis of human search behavior.

In our current analyses, we build upon search sessions to identify and characterize search tasks and topics. In characterizing these search tasks across sessions, we consider the possibility of three distinct forms of heterogeneity inherent in the search-task behavior. First, there could be *user-disposition* level heterogeneity wherein some users have a higher propensity to multi-task when searching for information, than other users. Second, there could be *topic* level heterogeneity wherein searchers have a higher (or lower) propensity to multi-task when searching information for specific kind of topics. Third, and finally, there could be *user-interest* level heterogeneity wherein users might have a higher or lower propensity to multi-task when searching for topics they are most or least interested in.

While recent work has highlighted the prevalence of multitasking behavior in online search [14, 15, 22], not much effort has been expended at fully characterizing online search tasks with an emphasis on such user- and topic-level differences. In the current study, we leverage a large dataset of real world search logs to perform a large scale characterization of search tasks with a focus on such differences. Specifically, we find that while most users (>50%) choose to multi-task in their search sessions, there exists significant differences in their choice of topics between single-task and multi-task sessions. Through our analyses we offer the following three insights:

(i) **Users' preference towards multitasking(3.2):** We find evidence that most users multi-task when searching for information with over 50% users completing more than 2 tasks within a single search session, and a minority of users even completing more than 5 tasks within a single session.

(ii) **Topic level heterogeneity(3.3):** For certain type of topics, users prefer to multi-task (e.g. kids, news, shopping etc.), while for certain others, users prefer to single-task (e.g. computers, games, adult etc.).

(iii) **User-interest level heterogeneity(3.4):** Users have different preferences towards multitasking depending on their level of interest in the specific search topic (e.g. some groups of users prefer to search about most-interested topics in single-tasking sessions and least-interested topics in mult-tasking sessions).

SIGIR '16, July 17-21, 2016, Pisa, Italy
© 2016 ACM. ISBN 978-1-4503-4069-4/16/07. . . $15.00
DOI: http://dx.doi.org/10.1145/2911451.2914755

No of Queries	620M
No of Sessions	190M
No of Users	2M
Avg No of queries per session	3.18
Avg no of sessions per user	76.01
Avg no of tasks per session	2.08

Table 1: Data summary

Time	Query	SessionID	TaskID	Topic
05/29/2012 14:06:04	adele songs	1	1	Arts
05/29/2012 14:11:49	wedding venue	1	2	Society
05/29/2012 14:12:01	video download	1	3	Arts
05/29/2012 14:06:04	Obama care	2	4	News
05/29/2012 14:11:49	running shoes	2	5	Shopping
05/29/2012 14:12:01	sports shoes	2	5	Shopping
05/29/2012 14:22:12	wedding cards	2	2	Society

Table 2: Sample search sessions

Analyzing such heterogeneities in online search behavior lends us a better understanding of how users interact with search systems when performing different tasks.

2. RELATED WORK

There has been a large body of work focused on the problem of segmenting and organizing query logs into semantically coherent structures. Many of these methods use the idea of a *timeout* cutoff between queries, where two consecutive queries are considered as two different sessions or tasks if the time interval between them exceeds a certain threshold, often 30 minutes [5, 9, 20]. However, the experimental results of these methods indicate that the timeouts, whatever their lengths, are of limited utility in predicting whether two queries belong to the same task, and unsuitable for identifying session boundaries. There have been attempts to extract in-session tasks [11, 14, 21], cross-session tasks [12, 23, 13] and hierarchies of tasks and subtasks [16] from query sequences based on classification and clustering methods. While existing research has investigated user behavior separately across sessions [1] and across tasks [10], in this work, we fully characterize search tasks with a focus on user and topic level differences, and jointly model user interactions across sessions, topics and tasks.

3. CHARACTERIZING SEARCH TASKS

We characterize online search sessions with a focus on the underlying user-level and topic level heterogeneities. As mentioned earlier, we investigate the prevalence of three related forms of search heterogeneity viz. (i) *user-disposition* level (do focused users behave differently than multi-taskers), (ii) *topic* level (are some topics more prone to multi-tasking?), and (iii) *user-interest* level differences in search behavior (do users' task behavior vary across their topical interests). We next describe our experimental setup (Section 3.1) based on which we discuss our findings in Section 3.2, 3.3 & 3.4.

3.1 Experimental Setup

3.1.1 Data Context

We use backend search logs for users of a major US-based search engine for a period of 30 days from May 1, 2015 to May 31, 2015 and choose a random sample of over 2 million users where each user is identified by a unique IP address. For our analysis, we filter out inactive users from our dataset

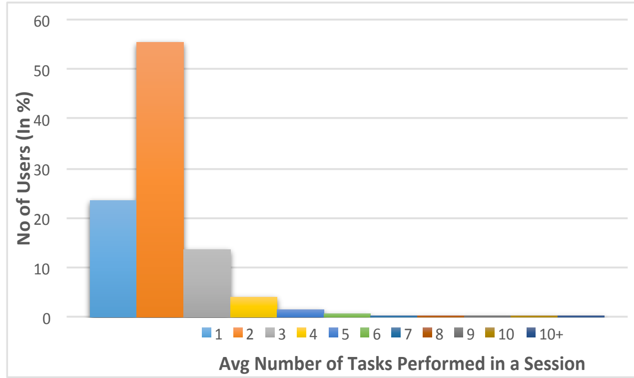


Figure 1: Quantifying the extent of multi-tasking within a session

who participate in <50 sessions, and focus instead on the more active user population. Table 1 presents a summary of the query, session and task information in our empirical context. Our dataset comprises 620 million queries spanning 190 million search sessions, with an average of 3 queries and 2 tasks per search session.

3.1.2 Task Extraction

For our analysis, we make use of the Latent Structural SVM framework [23] for task identification. Given query sequences within sessions, search tasks are identified by clustering queries into tasks by find the strongest link between a candidate query and queries in the target cluster (*bestlink*). This is achieved by making use of a structural learning method with latent variables, i.e., latent structural SVMs, to utilize the hidden structure of query inter-dependencies to explore the dependency among queries within the same task.

Given a query sequence $Q = q_1, q_2, ..., q_M$, a feature vector for the task partition y is specified by the hidden best-link structure h as $\psi(Q, y, h)$. Based on $\psi(Q, y, h)$, the bestlink SVM is a linear model parameterized by w, and predicts the task partition by,

$$(y*, h*) = argmax_{y,h} w^T \psi(Q, y, h) \quad (1)$$

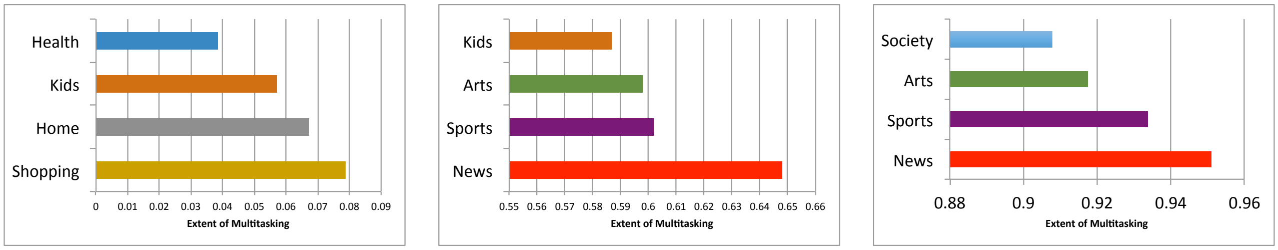
where Y and H represent the sets of possible structures of y and h respectively. $y*$ becomes the output for cross-session tasks and $h*$ is the inferred latent structure. Based on the best-link structure, $h(q_i, q_j) = 1$ if query q_i and q_j are directly connected in h; and otherwise, $h(q_i, q_j) = 0$, with the added clause that a a query can only link to another query in the past, or formally, $\sum_{i=0}^{j-1} h(q_i, q_j) = 1 \ \forall j \geq 1$. The feature vector for any particular task partition y is defined over the links in h as,

$$\psi(Q, y, h) = \sum_{i,j} h(q_i, q_j) \sum_{s=1}^{S} \phi_s(q_i, q_j) \quad (2)$$

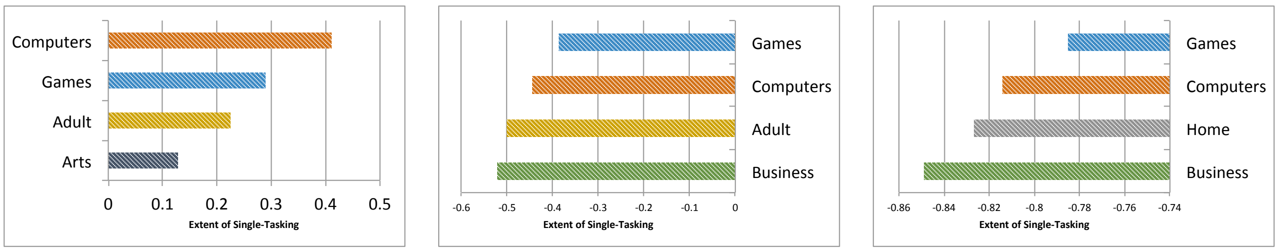
where a set of symmetric pairwise features $\phi_s(q_i, q_j)$ is given to characterize the similarity between query q_i and q_j. Given a set of query logs with annotated tasks, the feature vector design and the directed linkage structure of h can be inferred in an SVM setting. A detailed overview of the approach can be found in Wang *et al.* [23].

3.2 Search Sessions to Search Tasks

Search sessions have been exploited in previous work on information search, as being the major focus for most analy-



(a) Focused: Multitasking (b) Multitasker: Multitasking (c) Supertasker: : Multitasking



(d) Focused: Single-tasking (e) Multitasker: Single-tasking (f) Supertasker: Single-tasking

Figure 2: Top topics prone to multi-tasking (Top) and Single-tasking (Bottom) across different user groups.

sis of search behavior. The context of search activities within the current session has been used to build richer models of interests and improve how the search system interprets the user's current query. Session context has been used for modeling query and click sequences [4, 3], to disambiguate current search query [17], to build topical profiles for future interest prediction [24], to improve search quality [25, 6] to quantify struggling users [18], for understanding learning and expertise development [8] and for detecting atypicality in user behavior [7].

While search sessions are an important and convenient source for analysis, we contend that this conceptualization of sessions as focal units of analysis makes certain assumptions that are quite untenable in the general case. First, there exists no theoretical basis for bounding search sessions, as it is largely a data-driven subject. Previous research on the topic have adopted a time-out based strategy to bound search sessions [5, 9, 20]. However, it remains to be understood if such time-out based techniques have strong external validity across search contexts. Second, and most importantly, evidence from our own analysis of search logs show that users do indeed search for multiple unrelated topics within a single search session. Based on the task extracted within and across search sessions (3.1.2), our analysis hints at the presence of multiple search tasks within single session. This has been shown in Tables 1 and 2. In Table 2 we provide an illustration of a single search session (tagged session ID = 1) and the 2 different tasks that we have been able to identify from within the search session. Further, we also list the ODP topics that we could extract from the associated queries for each of the tasks. Since search sessions are inherently complex and often comprise a combination of multiple search tasks and topics, we shift our focus in the current study to analyze user- and topic-level differences across search tasks within and across sessions.

3.3 User-disposition and Topic Level Heterogeneity

Recent work on the topic of search multi-tasking has shown that a majority of users perform two or more tasks within a single search session [15]. Consistent with these studies, our analysis also uncovers that close to 55% of users perform two or more tasks within single sessions, with a minority of users even performing 5 or more tasks within the single session, as illustrated in Figure 1. Following Mehrotra *et al.* [15], we term these three discernible classes of users based on their frequency of multitasking behavior viz. focused (i.e. 1 task per session), multitaskers (i.e. 2-5 tasks per session) and supertaskers (i.e. >5 tasks per session). Having established that users vary on their disposition to single-task and multi-task, we now delve deeper into understanding whether users multi-task to varying extents depending on their topics.

To obtain such a topic representation for this study, we labeled each document with a vector of probabilities of categories from the top two levels of the ODP hierarchy using a text-based classifier. Each documents vector was restricted to the three most probable classes. The classifier has a micro-averaged F1 value of 0.60 and is described more fully in [24]. The most prominent topic among the top 3 returned results per query was used as the final tagged topic for that query.

We analyse topic level heterogeneity by investigating the level of multi-tasking in sessions filtered by topics. Our results are illustrated in Figure 2 wherein we highlight the top 4 most prevalent topics across multi-tasking and single-tasking sessions (top to bottom panels), for all three category of users (left to right panels). The length of the bars in each of the charts in the Figure 2 highlights the extent of multitasking (top panels) and the extent of single-tasking (bottom panels). The extent of multi-tasking is defined as $\frac{N_M - N_S}{N_{total}}$, which measures the difference between the proportion of times the topic featured in a multi-tasking session (N_M) and the proportion of times the topic featured in a single-tasking session (N_S). Conversely, the extent of single-tasking was calculated as the difference between the proportion of times the topic featured in a single-tasking session and the proportion of times the topic featured in a multi-tasking session.

We find that focused users primarily multi-task for top-

	Focused		Multi-taskers		Super-taskers	
	Single-Tasking	Multi-Tasking	Single-Tasking	Multi-Tasking	Single-Tasking	Multi-Tasking
Most Interested Topics	0.593	0.407	0.310	0.690	0.105	0.895
Least Interested Topics	0.458	0.542	0.249	0.751	0.081	0.919

Table 3: Relating User's Single/Multitasking Nature with their interest profiles.

ics related to shopping, home, kids, health and recreation. However, both multi- and super-taskers have a shared preference for multi-tasking on topics related to news, sports and arts. We also observe that focused users prefer to single task when searching for topics related to computers, games, adult and arts categories, while multi-taskers and super-taskers do not prefer to single-task when searching for their preferred topics. This is reflected by the negative scores on the extent of single-tasking in the bottom panel of Figure 2. These findings confirm our intuition that indeed certain topics are more prone to multi-tasking (e.g. news, sports) while others (e.g. computers, adult) usually witness single tasking sessions.

3.4 User-interest Level Heterogeneity

We next investigate whether users exercise any specific search preference when searching for topics that are of high vs. low interest to them. To analyze this, we compute the frequency of most and least searched topic categories from the search history of users in each of the three user groups viz. focused, multi-taskers and super-taskers [1]. Following this, we analyze their search behavior during single-tasking and multi-tasking sessions to investigate the distribution of high and low interest topic categories across these sessions. The results from this analysis are described in Table 3, and highlight that users exercise distinct preferences in search sessions for high vs. low interest topics.

Our results show that multi-taskers and super-taskers prefer to multi-task for a large majority of their search sessions (i.e. almost always >70%), irrespective of whether they are searching for high or low interest topics. In contrast, however, focused users prefer to search for high interest topics in single-tasking sessions (i.e. 59% of the time), and low interest topics in multi-tasking sessions (i.e. 54% of the time).

4. DISCUSSION AND CONCLUSION

We illustrate in this paper how a shift of focus from the idea of a search session to a search task raises a number of important questions. The most important of these is about fully characterizing the extent and underlying heterogeneities surrounding single-task and multi-task search sessions. While we draw on previous work as well as our own set of analyses to show that multi-tasking within a search session is fairly common, we also emphasize that the extent and nature of multi-tasking is strongly influenced by user dispositions (i.e. whether a user is naturally disposed to single vs. multi-tasking), topic preferences (i.e. users might prefer to multi-task when searching for certain topics than others), and interest preferences (i.e. users might prefer to multi-task about topics they are more or less interested in. Our findings have implications for understanding user preferences which in turn could impact the design of better personalization services for searchers.

[1]Note that this is different from the identification of top topics in the previous section which were identified at a session-level and not at a user-level.

5. REFERENCES

[1] E. Agichtein, E. Brill, and S. Dumais. Improving web search ranking by incorporating user behavior information. In *SIGIR 2006*.

[2] P. Bailey and Chen. User task understanding: a web search engine perspective. In *NII Shonan Meeting on Whole-Session Evaluation of IIR Systems*, 2012.

[3] H. Cao, D. H. Hu, D. Shen, D. Jiang, J.-T. Sun, E. Chen, and Q. Yang. Context-aware query classification. In *SIGIR 2009*.

[4] H. Cao, D. Jiang, J. Pei, Q. He, Z. Liao, E. Chen, and H. Li. Context-aware query suggestion by mining click-through and session data. In *KDD 2008*.

[5] L. D. Catledge and J. E. Pitkow. Characterizing browsing strategies in the world-wide web. *Computer Networks and ISDN systems*, 1995.

[6] M. Daoud, L. Tamine-Lechani, M. Boughanem, and B. Chebaro. A session based personalized search using an ontological user profile. In *Proceedings of the 2009 ACM symposium on Applied Computing*. ACM, 2009.

[7] C. Eickhoff, K. Collins-Thompson, P. N. Bennett, and S. Dumais. Personalizing atypical web search sessions. In *WSDM 2013*.

[8] C. Eickhoff, J. Teevan, R. White, and S. Dumais. Lessons from the journey: A query log analysis of within-session learning. In *WSDM 2014*.

[9] D. He, A. Göker, and D. J. Harper. Combining evidence for automatic web session identification. *Information Processing & Management*, 2002.

[10] J. Jiang, D. He, and J. Allan. Searching, browsing, and clicking in a search session: changes in user behavior by task and over time. In *SIGIR 2014*.

[11] R. Jones and K. L. Klinkner. Beyond the session timeout: automatic hierarchical segmentation of search topics in query logs. In *CIKM 2008*.

[12] Kotov, Bennett, White, Dumais, and Teevan. Modeling and analysis of cross-session search tasks. In *SIGIR 2011*.

[13] L. Li, H. Deng, A. Dong, Y. Chang, and H. Zha. Identifying and labeling search tasks via query-based hawkes processes. In *KDD 2014*.

[14] C. Lucchese, S. Orlando, R. Perego, F. Silvestri, and G. Tolomei. Identifying task-based sessions in search engine query logs. In *Proceedings of ACM WSDM 2011*.

[15] R. Mehrotra, P. Bhattacharya, and E. Yilmaz. Characterizing users' multi-tasking behavior in web search. In *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval*, 2016.

[16] R. Mehrotra and E. Yilmaz. Towards hierarchies of search tasks & subtasks. In *WWW 2015*.

[17] L. Mihalkova and R. Mooney. Learning to disambiguate search queries from short sessions. In *Machine Learning and Knowledge Discovery in Databases*. 2009.

[18] D. Odijk, R. W. White, A. Hassan Awadallah, and S. T. Dumais. Struggling and success in web search. In *CIKM 2015*.

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

[20] C. Silverstein, H. Marais, M. Henzinger, and M. Moricz. Analysis of a very large web search engine query log. In *ACm SIGIR Forum*, volume 33, pages 6–12. ACM, 1999.

[21] Spink, Koshman, Park, Field, and Jansen. Multitasking web search on vivisimo. com. In *ITCC 2005*.

[22] A. Spink, M. Park, B. J. Jansen, and J. Pedersen. Multitasking during web search sessions. *Information Processing & Management 2006*.

[23] H. Wang, Y. Song, M.-W. Chang, X. He, R. W. White, and W. Chu. Learning to extract cross-session search tasks. In *WWW 2013*.

[24] White, Bennett, and Dumais. Predicting short-term interests using activity-based search context. In *CIKM 2010*.

[25] B. Xiang, D. Jiang, J. Pei, X. Sun, E. Chen, and H. Li. Context-aware ranking in web search. In *SIGIR 2010*.