Characterizing and Supporting Cross-Device Search Tasks

Yu Wang
Emory University
Atlanta, GA 30322 USA
yu.wang@emory.edu

Xiao Huang
Microsoft
Bellevue, WA 98004 USA
xiaohua@microsoft.com

Ryen W. White
Microsoft Research
Redmond, WA 98052 USA
ryenw@microsoft.com

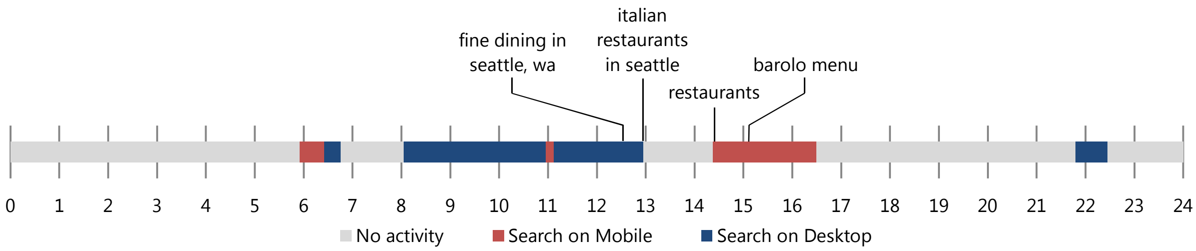

Figure 1. Search activities on mobile and desktop of a fictitious user over the course of a single day. Numbers denote hours from midnight. Queries of interest (relevant to the body of the paper) are included above the figure for reference.

ABSTRACT
Web searchers frequently transition from desktop computers and laptops to mobile devices, and vice versa. Little is known about the nature of cross-device search tasks, yet they represent an important opportunity for search engines to help their users, especially those on the target (post-switch) device. For example, the search engine could save the current session and re-instate it post switch, or it could capitalize on down-time between devices to proactively retrieve content on behalf of the searcher. In this paper, we present a log-based study to define and characterize cross-device search behavior and predict the resumption of cross-device tasks. Using data from a large commercial search engine, we show that there are discernible and noteworthy patterns of search behavior associated with device transitions. We also develop learned models for predicting task resumption on the target device using behavioral, topical, geospatial, and temporal features. Our findings show that our models can attain strong prediction accuracy and have direct implications for the development of tools to help people search more effectively in a multi-device world.

Categories and Subject Descriptors
H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – *search process; selection process*

Keywords
Cross-device search; Slow search; Search tasks; Personalization.

1. INTRODUCTION
Search tasks can involve multiple queries across multiple sessions, and require significant effort to complete, especially if the task is complex [31]. Modern search engines have started shifting their goal from simply achieving better result ranking for individual queries to assisting users in completing such tasks [10][28]. The recent proliferation of mobile devices such as smartphones and tablets allows Web searchers to tackle these search tasks almost anytime, anywhere. Figure 1 presents an example of cross-device behavior for a fictitious (but representative) user searching for information on Italian restaurants on his desktop before continuing the task on his mobile device. Better support for task continuation could help the user resume his restaurant search when mobile. Such *contiguous* cross-device tasks—those resumed soon on the post-switch device—are our focus in this paper. According to our analysis of device switching (shown in Section 5.3.2), around 15% of switches involve contiguous tasks.

Switching between devices may be expensive for contiguous tasks. The user has to remember what he was searching for and what has already been searched, which can be difficult when multiple search tasks are active simultaneously. One solution is sharing all search history across all devices. However, this is insufficient since the user might not necessarily resume a search task with a previously searched query and managing one's search history can be challenging from a smartphone. To provide a smooth transition among search devices and aid users in completing tasks on multiple devices, more sophisticated support is needed.

Cross-device behavior has been studied in the human factors community, but not with an emphasis on search [8]. Cross-session search tasks have been studied, but not across devices [20]. Mobile and desktop search have been studied separately [15], but the transitions between them have not been examined. A detailed study of cross-device searching and the development of tools to support this activity are therefore timely and necessary. This paper makes a significant contribution as the first study in this important area.

The specific research contributions of our work are fourfold:

- Define cross-device search tasks as a key research challenge and an opportunity for search engines to help people better perform search tasks that span devices. We demonstrate via empirical study the prevalence of cross-device searches.
- Characterize cross-device task transitions, including identifying patterns in device transitions and exploring the temporal, geospatial, and topical aspects of cross-device searching.
- Develop predictive models to estimate which search tasks will be resumed following device switching. Prediction occurs at different time points, including before a full device switch is observed (when they abandon the pre-switch device) and after device switching (given that they visit the homepage of the search engine on their post-switch device). The model integrates a rich set of features of cross-device searching behavior.

WSDM'13, February 6–8, Rome, Italy.

- Perform experiments using the search log from a large commercial search engine and show that our model significantly outperforms a task-continuation baseline, based on prior work that lacks access to information on cross-device behavior.

The remainder of the paper is structured as follows. Section 2 describes related work on desktop and mobile search, multi-device usage, and cross-session search tasks. We characterize cross-device searching in Section 3 and present the details of our prediction models in Section 4. The findings of the prediction experiments are presented in Section 5. We discuss the findings and their implications in Section 6 and conclude in Section 7.

2. RELATED WORK

In this paper our focus is on tasks spanning multiple sessions and multiple devices. Related work falls in the following areas: (1) characterizing search behavior on desktop and mobile, (2) multi-device behavior, and (3) cross-session search tasks.

Search behavior has been studied intensively in recent years. Log data from search engines have proven to be extremely valuable in studying how people search in naturalistic settings across a wide variety of different search intents in desktop or mobile search environments. Recent research on search behavior on the desktop has primarily involved automated analysis and prediction of aspects of search behavior for individual queries [29] and single search sessions [1][4][34], using search logs. Other more qualitative studies have focused on a deeper understanding of the nature and motivations behind the pursuit of searchers' information goals [19][21].

Studies of search in a mobile setting have examined the characteristics of search queries submitted from mobile devices, analyzing behavior along different dimensions such as geographic location and search interface used [33]. Others have performed detailed studies to understand mobile search intent and the influence of contextual factors on them. Church and Smith [7] carried out a four-week diary study of mobile information needs, focusing on the intent behind them, the topics users are interested in and the impact of mobile contexts such as location and time. Teevan et al. [32] showed that local searches tend to be highly contextual, influenced by geographic features, temporal aspects, and the searcher's social context. They showed that mobile searchers are often in transit, and tend to seek out information related to their destination rather than their current location.

Research on comparing and contrasting search behavior on multiple devices is also relevant [14][15][21]. Kamvar and Baluja [14] presented a large scale study of search patterns on Google's mobile search interface. They compared search patterns between phones, personal digital assistants (PDAs), and conventional computers and examine the search queries and their categories as well as other aspects of their interaction such as query input speeds and click-through. Kamvar et al. [15] presented a similar log-based comparison of search patterns on different devices, with the explicit goal of understanding how mobile search tasks differ on computers versus mobile devices. Their results suggest that search usage is typically more focused for the average mobile user than for the average desktop computer user, but search behavior on high-end phones resembles desktop search behavior more so than it does mobile. Li et al. [21] studied good abandonment of search results between desktop and mobile (where users do not click but are still satisfied with the results). They show that good abandonment is significantly higher in mobile search than in desktop-based search.

All of these analyses consider search behavior within a single query or search task on different devices independently. They do not consider two key things: (1) transitions between devices, and (2) search tasks that extend over time. We highlight related work in each of these two areas in the remainder of this section.

Multi-device usage patterns have been actively studied. Studies have shown that user activities tend to span multiple devices [4] and frustrating experiences on mobile devices will drive users to complete their tasks on desktops [17]. Karlson et al. [18] analyzed the usage log of desktop and mobile phone from a user study, and they showed smartphones had become a primary tool to access the internet. They also pointed out that tasks cannot be easily carried over between devices due to the lack of support. We are try to address in this work, especially in support of Web searching across devices. Kane et al. [16] conducted a study focused on Web browsing usage patterns across devices. Their results indicated sharing browsing information between devices could help improve Web browsing on mobile devices. However, to our knowledge, there has been no work specifically focused on cross-device search tasks.

There is growing interest in using long-term search log data to build models of users' interests. Previous work has tried to automatically identify queries on the same task. Mei et al. [26] proposed a framework to study sequences of search activities and focused on simple prediction and classification tasks, ranging from predicting whether the next click will be on an algorithmic result to segmenting the query stream into goals and missions. Teevan et al. [29] showed, via query log analysis, that nearly 40% of queries were attempts to re-find previously encountered results. Aula et al. [4] studied the search and information re-access strategies of experienced Web users using a survey. They found that people often have difficulty remembering the queries they used originally to discover information of interest. MacKay and Watters [25] explored a variety of Web-based information seeking tasks and found that almost 60% of complex information gathering tasks continued across sessions. Liu and Belkin [23] examined the structure (parallel or dependent) of tasks that extend across different search sessions. Jones and Klinkner [13] proposed methods to partition a query stream into research missions and goals, where each mission corresponds to a set of related information needs and may include multiple search goals.

Some explicit support has also been proposed to help people manage long-running tasks. *SearchBar* [27] is a system that proactively and persistently stores query histories, browsing histories, and users' notes and ratings. SearchBar supports multi-session investigations by assisting with task context resumption and information re-finding. *SearchPad* [10] is a system that automatically identifies research missions and presents a search workspace comprising previous queries and results related to the mission. SearchPad uses measures of topic coherence between pairs of consecutive queries and user engagement to identify such research missions. This work was further extended to group queries into mission-coherent clusters based on search behavior [3]. Kotov et al. [20] modeled cross-session information needs and addressed the challenge of identifying all previous queries in a user's search history on the same task as the current query, and predicting whether a user will return to the task in future sessions. They developed classifiers for these two tasks and through evaluation using labeled data from search logs showed that their classifiers can perform both tasks effectively. We use classifiers trained on similar features a baseline for some of the analysis presented later in the paper. Agichtein et al. [2] perform log analysis to understand, characterize, and automatically detect search tasks that will be continued in the near future. They identify intents, topics, and search behavior patterns associated with long-running sessions that are likely to be continued. They also developed an effective task-continuation prediction algorithm that significantly outperforms state-of-the-art classifiers and humans.

The research described in this paper extends previous work in a number of ways. First, we focus on search tasks spanning multiple devices, an area that is gaining importance but has not been explored in detail. Previous studies on mobile and desktop search have focused on the device types independently. Second, we provide the first characterization of cross-device task transitions, including key statistics on the nature of the switches, such as the role of topic, time, and location in cross-device tasks. Third, we develop models to accurately predict task continuation across devices before the switch and once it has been observed. Importantly, we focus on predicting whether the user will resume the search task immediately on another device. This is a different and more challenging task than predicting resumption at some point in the near future; a problem that others have tackled [2][20].

3. CHARACTERIZING CROSS-DEVICE SEARCH BEHAVIOR

We begin by formally defining cross-device search and providing overview statistics about the data that we used in this study. We then examine several characteristics of search across devices, focusing on temporal, geospatial, and topical dimensions.

3.1 Definition

The search activities of a user are usually represented as a stream of temporally-ordered queries. Query streams provide rich information about users' search interests and search tasks. To better capture the search intent of users, the concept of a search session is employed to segment the query stream into fragmented units for analysis. Here we define a typical search session using a 30-minute timeout as its boundaries in the stream. This definition has been used to identify sessions in previous work [34].

A *device switch* is defined as the act of moving between a pair of devices (e.g., personal computer→smartphone). A search session consists of one or more queries, but in our analysis, we do not permit a search session to span multiple devices. In other words, the queries within the same session always occur on a single device. As a result, switching between devices must involve at least two sessions, the *pre-switch* session and the *post-switch* session.

More precisely, let $\mathcal{Q} = \{q_1, q_2, \dots, q_i, \dots, q_n\}$ be the query stream of a user, where q_i is the i-th query in the stream. For each q_i ($1 \le i \le n$), there is a 3-tuple (t_i, s_i, d_i) associated with it, where t_i is the timestamp of the query q_i, s_i represents the session of q_i, and d_i defines the device where q_i has been issued. $\mathcal{S}$ is the personal search history comprising user search activity from $\mathcal{Q}$ in the time period before s_i. We therefore define cross-device search as:

DEFINITION: *A cross-device search is represented as a 7-tuple,*

$$(\mathcal{S}, q_i, q_{i+1}, s_i, s_{i+1}, d_i, d_{i+1}).$$

And the following conditions need to be satisfied: (1) q_i is the last query in session s_i; (2) q_{i+1} is the first query in session s_{i+1}; (3) d_i and d_{i+1} are two different devices.

The device-switching behavior starts at time t_i and ends at time t_{i+1}. We define q_i as the pre-switch query, s_i as pre-switch session, and s_{i+1} as post-switch session. One of the tasks in this study is to predict whether the search task of q_i will be resumed in the immediate following post-switch session s_{i+1}, referred to as a contiguous cross-device task. All the queries in s_{i+1} are defined as post-switch queries, and q_{i+1} is the first query in the post-switch session. Later in this paper (Section 5.3.2), we will show that q_{i+1} is most likely to be related to continuing the task of q_i among all queries in the post-switch session. Device-switching reveals rich information about the user, including device preferences for searching and the timespan and changing geolocation during the switch.

Table 1. Dataset used in this study.

Number of Days		31
Number of Users		39,081
Number of Sessions	Desktop	709,610
	Mobile	301,028
	Total	1,010,638
Number of Queries	Desktop	3,023,582
	Mobile	667,091
	Total	3,690,673
Number of Switches		158,324

Table 2. Count of switches in two dimensions: direction of the switches across textual relation between switch-related queries.

	Desktop-to-Mobile	Mobile-to-Desktop
Same-query switch	10,480 (6.6%)	5,282 (3.3%)
Different-query switch	69,441 (43.9%)	73,121 (46.2%)

3.2 Dataset Description

In this study, we focus on the switching between two devices: the personal computer (PC) also referred to as "desktop" in this paper[1], and the smartphone. Queries issued on both devices are collected from 39,081 users over one month. Queries are mined from the logs of both modalities respectively and then joined using a persistent user identifier. For each query, we also recorded its timestamp and geolocation (only at the city level), allowing for temporal and geospatial characteristics to be studied. Table 1 shows a basic description of our dataset. Note that all the users in the collection must have used both devices in the period from April 15 to May 15, 2012.

Directionality of device switching: Intuitively, desktop-to-mobile switching may occur when the search is interrupted and the user needs to change his location before resuming (e.g., to catch public transit). Mobile-to-desktop switching may indicate that the user is returning from other events. Since two switching directions imply different search scenarios and suggest different applications, we treat the two directions separately.

Same-query switch vs. different-query switch: When the user issues the same query before and after the switch, it is then reasonable to claim that the search task is resumed after the switch. More formally, if $q_i = q_{i+1}$, the task of q_i is resumed in the post-switch session s_{i+1}. The textual relationship between q_i and q_{i+1} may be a strong indicator of contiguous cross-device tasks.

Table 2 shows the count of cross-device search in two dimensions, directionality and the textual relation within the switch. The substantial volume of same-query switches indicates that contiguous cross-device search tasks may often exist in users' searching activities. One interesting fact is that when users switch from desktop to their mobile phones, they are more likely to resume search tasks with the same query: 6.6% of queries for Desktop-to-Mobile versus 3.3% of queries for Mobile-to-Desktop. Note that distribution of *users* among the four quadrants is similar to that shown in Table 2.

In practice, switching from desktop to mobile may be more challenging for users because typing and resuming search tasks on mobile is likely to be more difficult due to the restricted typing area and high mobility of smartphones. To provide smooth transitions

[1] Note that that our definition of desktops may also include laptops, since we were unable to distinguish them in our log data.

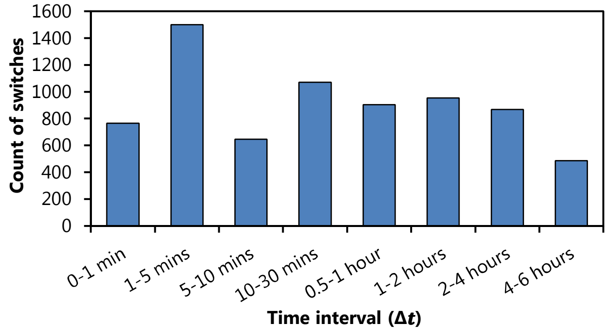

Figure 2. Time interval distribution of same-query switches.

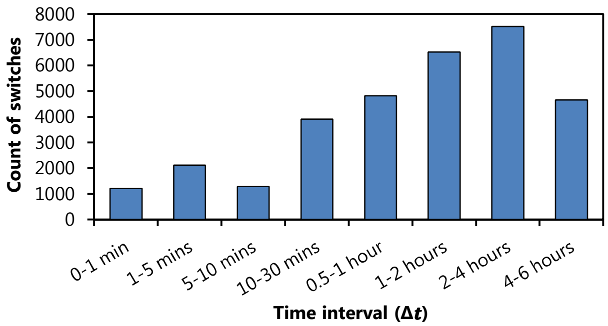

Figure 3. Time interval distribution of different-query switches.

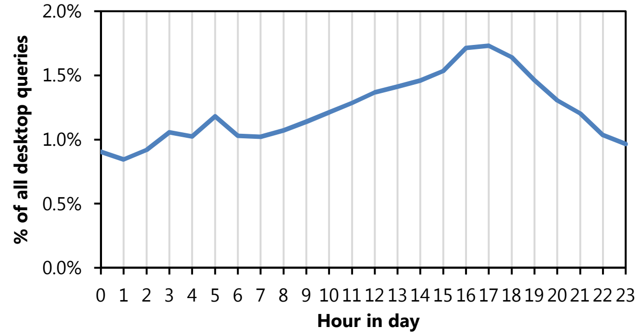

Figure 4. Percentage of pre-switch queries on desktop over time.

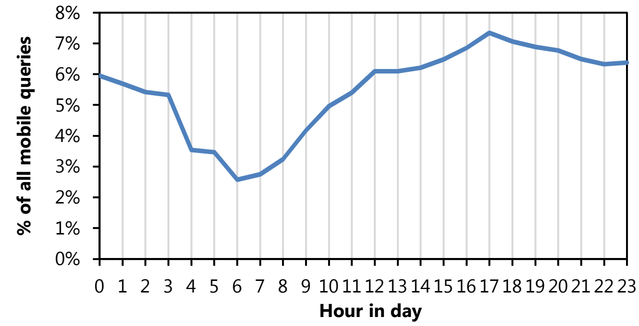

Figure 5. Percentage of post-switch queries on mobile over time.

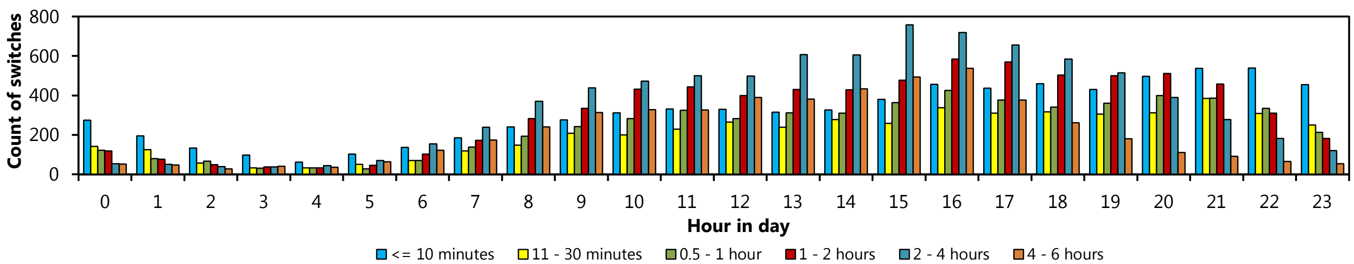

Figure 6. Distribution of switching time interval (Δt) as a function of hour in the day. Order of legend entries corresponds to the order of the bars.

from desktop to mobile as proposed in Section 6, our work tries to predict whether the search task will be resumed post switch. Particularly, we are interested in providing a better mobile search experience by knowing which tasks will be continued when the user leaves the desktop and lands on his phone. For the rest of the paper, we target ***Desktop-to-Mobile*** switches.

3.3 Temporal Characteristics

Cross-device search behavior starts with the pre-switch query and ends with the start of the post-switch session. One intuitive question is the duration of the switch. To answer this question we compute the distribution of the interval Δt, defined as $(t_{i+1} - t_i)$.

***Six-hour limit*:** The time interval exhibits a long-tail distribution, and the longest switch spans several weeks. However, the long-term device-switching is not the focus here, especially since such distal resumptions are more likely to be connected to persistent interests rather than an active search task that the system can directly assist with at the time that the switch occurs. Again, the point of this study is to investigate *contiguous* cross-device search tasks. Therefore we set a six-hour threshold, which covers 50% of the switches, to filter out long-term cross-device searches. Following this filtering, the distribution of switch direction and textual relationship is still similar to Table 2. For the rest of Section 3, all statistics are computed from switches spanning at most six hours.

Figure 2 and Figure 3 show the distributions of Δt for same-query switches and different-query switches. For same-query switches users tend to resume searching on mobile within a short period of time; over 40% of same-query device switches occur within ten minutes of t_i. However, for different-query switches, the transition will take longer perhaps because the user is not actively engaged in a search task. These two figures imply that time interval is another indicator of task-resuming switches.

Another interesting question about the temporal aspect of cross-device search is when users are most likely to leave the desktop (t_i) and what time they will start searching on mobile (t_{i+1}). Intuitively, the searching on desktop might be interrupted around the time when people get off work, and would probably use the smartphone to keep searching (e.g., on transit during the commute). Note that the rates of leaving desktop and starting to use mobile are also affected by the total number of queries issued on both devices. To show the real distribution of t_i and t_{i+1}, the rates are normalized by the search volume on each of the devices over time.

Figures 4 and 5 show the distribution of pre-switch queries and post-switch queries occur over the day. It is evident that users are more likely to begin switching (leaving the desktop) around 4-5PM. Meanwhile, most post-switch queries (starting to use mobile) appear around 5PM. Also, users tend to stop searching on desktop late at night (the rate of leaving desktop is increasing), and the probability of starting search on mobile is very low at late night. These trends align well with people's work-life rhythms.

We have explored time interval distributions of cross-device search and the likelihood of starting the switch over time in a day. The combination of these two temporal aspects could give us more insight of device-switching behavior. Particularly, we are trying to learn if Δt is sensitive to t_i, the start time of the switch.

Figure 6 shows the distribution of $(t_{i+1} - t_i)$ over the hour in the day of t_i. We can see there are noteworthy patterns of the duration of the device-switches over time. For example, if the user leaves the desktop around 3-4PM, he will be more likely to resume the search in 2-4 hours. However, the situation would be very different if the user leaves desktop around midnight (hour 0 in the figure), where he is more likely to resume the search within 10 minutes. Being able to estimate the return time based on time of day is potentially useful for search engines to prepare for searches on mobile. The distribution in Figure 6 may offer a way to predict when the user will start searching on mobile given only knowledge of desktop behavior. We leverage time-of-day and other temporal features in the prediction models described later.

3.4 Geospatial Characteristics

One of the most common reasons for users to search on mobile is the limited mobility of desktop. We suspected that cross-device search may involve a change in location. In our dataset, geospatial information is available at town or city level (e.g., Seattle, WA) based on the same information from the internet providers on desktop and on the mobile device. A comparison of the location before and after the switch shows that one third of the switches (33%) involved a shift in location to a different town or city than before the switch. This is less surprising if you consider the large number of conurbations (i.e., extensive urban areas comprising multiple towns or cities) in the United States. Later in this paper (Section 5.5), we will show that geospatial properties of the switching event (such as average speed during the switch) can help predict contiguous cross-device search tasks.

Since we focus on desktop-to-mobile switches in this analysis, we also observe the change of user's location within the post-switch session. To measure the extent of the change in geolocation within the post-switch session, at least two queries and corresponding geo-coordinates are required. Table 3 provides the percentage of post-switch sessions where the user is moving.

Table 3. Mobility of post-switch session (mobile).

Single query session	Multiple query session	
	Moving session	Stationary session
60.6%	5.3%	34.2%

Among all the multi-query sessions, around 13.3% sessions have non-zero moving speed. Such a scenario suggests that the user is resuming their task while travelling, where interaction with the phone may be more difficult than usual and they may want information about their destination not their current location [32].

3.5 Topical Characteristics

Topic level information has been extensively employed to capture users' search intent and construct models of their search tasks [2][35]. We wanted to understand how the post-switch queries are affected by the topic of the pre-switch query. To do so, we study the *sustainability* of query topics during device switching.

Let z_i be the topic of pre-switch query q_i, and z_{i+1} be the topic of post-switch query q_{i+1}. If the topic of post-switch query is affected by the pre-switch query, the conditioned probability $p(z_{i+1}|z_i)$

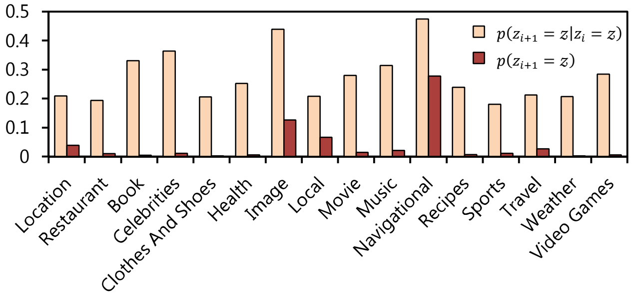

Figure 7. Sustainability and global popularity of query topics.

will be very different from the global probability $p(z_{i+1})$. Otherwise, these two probabilities will be similar. Particularly, we study the difference between $p(z_{i+1} = z|z_i = z)$ and $p(z_{i+1} = z)$ for several popular topic z. We call $p(z_{i+1} = z|z_i = z)$ as the "sustainability" of topic z because it measures the degree to which topic z will be resumed after the device-switching. To operationalize this topic modeling for our experiments, we ran a proprietary classifier on switch-related queries, which could assign multiple topics to a query, resulting in the sum of the probabilities of all topics exceeding one. Figure 7 shows the distribution of the probabilities described above for each of the topics in our dataset.

Figure 7 compares the overall popularity $\left(p(z_{i+1} = z)\right)$ and sustainability $\left(p(z_{i+1} = z|z_i = z)\right)$ of the query topics. It shows that the topic distributions of post-switch queries are strongly dependent on the topics of pre-switch queries. We can also tell that users prefer certain topics. For example, *Book*, *Celebrities* and *Music* are more sustainable than other topics. *Image* and *Navigational* are most likely to be resumed in part because of their high overall popularity.

It is clear that there are a number of important temporal, geolocation, and topical attributes in cross-device search. In the next section, we leverage these insights to explore the prediction of contiguous cross-device search tasks.

4. PREDICTING CROSS-DEVICE SEARCH

Before proceeding, we formally define contiguous cross-device search tasks based on the definition of cross-device search in Section 3.1. After that, the features used to capture the contiguous cross-device search tasks will be introduced.

4.1 Definition

A stated earlier, a cross-device search is defined by a 7-tuple, including the personal search history $\mathcal{S}$, the pre-switch and post-switch queries q_i and q_{i+1}, the pre-switch and post-switch sessions s_i and s_{i+1}, and the devices used before and after the switch. Since we only focus on the desktop-to-mobile switches, the definition could be simplified by excluding the device components.

Given a cross-device search, we are trying to predict whether the task of query q_i is continued by any of the queries in the post-switch session s_{i+1}. The following Boolean function f is introduced to judge if two queries belong to the same search task:

$$f(q_m, q_n) = \begin{cases} True & \text{if } q_m \text{ and } q_n \text{ belong to the same task,} \\ False & \text{else.} \end{cases}$$

Then the ground truth for the prediction will be:

$$\bigvee_{q_j \in s_{i+1}} f(q_i, q_j) \quad (1)$$

Again, q_i is the pre-switch query and s_{i+1} is the post-switch session. By comparing q_i with every query q_j in the post-switch session and take the inclusive OR operation with the results, the final

Table 4. Features for query pair relevance function f.

Name	Description
EditDistance	Editing distance between two queries
NumTermOverlap	Number of overlapping terms in two queries
QueryTermJaccard	Jaccard coefficient of two query term sets
IsSameQuery	Boolean, true if two queries are identical
IsSubsetQuery	Boolean, true if one query contains the other

Table 5. Performance of the query-pair relevance function f.

Accuracy	Positive Precision	Positive Recall	Negative Precision	Negative Recall
0.92	1.00	0.67	0.91	1.00

ground truth for the cross-device search $(\mathcal{S}, q_i, q_{i+1}, s_i, s_{i+1})$ is obtained. Our prediction task is therefore to predict the result of function (1) using features extracted from $(\mathcal{S}, q_i, q_{i+1}, s_i, s_{i+1})$.

In a real setting, the prediction has to be made before post-switch queries are issued, otherwise, the prediction becomes trivial for many applications. In Section 5, we will study the performance of our predictive model when features from different components of the definition are added incrementally.

4.2 The Choice of Function f

Function f measures the relevance between two queries in terms of searching tasks. One option is to ask human labelers to determine the relevance for every query pair. However, this approach is not feasible on a large scale.

In order to obtain the ground truth for all cross-device searches in our dataset, we need a lightweight function f. The three authors of this paper labeled 200 randomly chosen pre- and post-switch query pairs. Fleiss' Kappa (κ) was 0.90, signifying almost perfect agreement for this task. We then train a Support Vector Machine (SVM) classifier using the manual labels and features listed in Table 4 chosen based on [20]. We use the trained SVM classifier as function f. The results based on five-fold cross-validation on function f against the human judgments (Table 5) show that f is accurate on the positive class and can capture two thirds of related-query pairs.

4.3 Features

We temporally segment the device-switching process into five stages and design features accordingly. These five stages include: (1) User's searching history $\mathcal{S}$; (2) Pre-switch session s_i; (3) Pre-switch query q_i; (4) The transition, time period from t_i to t_{i+1}; (5) Post-switch session s_{i+1}. Table 7 lists the features for each stage. As we discuss later, features from (4) could be used to predict task resumption once the user is at the search engine homepage on their mobile device. Feature set (5) is included for completeness.

Using function f to group queries into search tasks: since function f is designed to judge the relevance of query pairs, we therefore apply f to every pair of queries in user's search history, and then cluster queries into groups. Each group of queries represents a certain search task. By observing how the group of queries distributes on desktop and mobile, we compute entropy-based features and cross-device features.

Suppose G_i is the i-th group in the user's search history, and it consists of k queries, $G_i = \{q_{i_1}, q_{i_2}, \dots, q_{i_k}\}$. Among these k queries, l of them are issued on desktop and $(k - l)$ queries are on mobile. Then the device entropy of G_i is:

$$\frac{l}{k}\log\frac{k}{l} + \frac{k-l}{k}\log\frac{k}{k-l}$$

Not only entropy-based features are computed from query groups, some cross-device-related features also group queries into search tasks. The purpose is to capture the individual device preferences for different tasks. For example, the history feature *NumOfRelevantCrossDevice* counts the number of tasks spanning both devices; the pre-switch query feature *NumRelatedQuerySwitch* counts the number of contiguous cross-device searches on the tasks of q_i.

Table 6. Features used to predict contiguous cross-device search tasks. "B" indicates the features used in baseline method.

Name	Description
Features from Search History $\mathcal{S}$	
NumOfDesktopQuery[B]	Number of queries issued on desktop
NumOfMobileQuery	Number of queries issued on mobile
PercentageDesktopQuery[B]	Percentage of queries issued on desktop
PercentageMobileQuery	Percentage of queries issued on mobile
PercentageDesktopTime[B]	Percentage of searching time on desktop
PercentageMobileTime	Percentage of searching time on mobile
NumOfSession[B]	Number of search sessions
NumOfContiguousSwitch	Number of contiguous cross-device tasks
NumOfRelevantCrossDevice	Number of search tasks on both devices
EntropyAvg	Average device entropy of same-task queries
EntropySum	Total device entropy of same-task queries
EntropyWeighted	Weighted device entropy of same-task queries
Features from Pre-switch Session s_i	
NumOfQuery[B]	Number of queries within session s_i
TimeSpanPreSess[B]	Temporal length of session s_i (in minutes)
NumOfLocationQuery[B]	Number of location queries in session s_i
AvgDistancePreSess[B]	Average distance from current location to locations mentioned in session s_i
Features from Pre-switch Query q_i	
GlobalFrequency	Historical frequency of q_i in the entire dataset
PersonalFrequency	Frequency of q_i in personal search history $\mathcal{S}$
NumExactQueryDesktop[B]	Number of same queries as q_i on desktop in $\mathcal{S}$
NumExactQueryMobile	Number of same queries as q_i on mobile
NumRelatedQueryDesktop[B]	Number of related queries as q_i on desktop
NumRelatedQueryMobile	Number of related queries as q_i on mobile
NumExactQuerySwitch	Number of switches that pre-switch query and post-switch query are the same as q_i
NumRelatedQuerySwitch	Number of switches that pre-switch query and post-switch query are both relevant to q_i
PreQueryContiguousSwitch	Number of contiguous cross-device tasks of q_i
NumOfRelatedQueryInSess[B]	Number of queries relevant to q_i in session s_i
NumOfTerm[B]	Number of terms in query q_i
PreQueryCategory[B]	The search topic of query q_i
PreQueryHour[B]	The hour component of t_i
PreQueryDayofWeek[B]	The day of week of t_i
IsWeekday[B]	True if t_i is weekday
HasLocation[B]	True if q_i contains location
PreQueryDistance[B]	Distance from current location to q_i location
HasLocalService[B]	Boolean, true if q_i contains local service
Features from the Transition (t_i~t_{i+1})	
TimeIntervalSwitch	The timespan between t_i and t_{i+1}
GeoDistanceSwitch	Distance between where q_i and q_{i+1} are issued
IsSameLocationSwitch	True if q_i and q_{i+1} occur at the same place
AvgSpeedSwitch	Average travelling speed during the switch
Features from Post-switch Session s_{i+1}	
TimeSpanPostSess	The temporal length of session s_{i+1}
PostQueryCategory	The search topic of query q_{i+1}
PostQueryHour	The hour component of t_{i+1}
GeoDistancePostSess	The distance travelled within session s_{i+1}
AvgSpeedPostSess	Average travelling speed within session s_{i+1}

4.4 Baseline Features

Previous work focuses on predicting the resumption of cross-session search tasks on desktop computers [2][20]. In contrast, our work predicts task-resuming in post-switch sessions by considering the search behaviors on *both* desktop and mobile. We use desktop search features (marked as "B" in Table 6) as a baseline, which contains search history-based features, pre-switch session features, and pre-switch query features, making them good comparators for other features in these stages. Additionally, features from transition and post-switch session are evaluated in Section 5.

4.5 Evaluation Metrics

One of the potential applications of our work is to help users resume their search tasks after device-switching in a way that reminds users

Table 7. Statistics of the dataset used in the experiments.

Number of users (switching frequency ≥ 15)	2,125
Number of potential cross-device search tasks (Pre-switch query with personal frequency ≤ 5 and global frequency ≤ 10)	17,235
Number of all cross-device searches (Without pre-switch query frequency limits)	29,839

Table 8. Statistics of human labeling data.

	Labeling Task #1	Labeling Task #2
Number of labels	800	800
Number of positive labels	238	119
Agreement (Fleiss' κ)	0.037	0.509

Table 9. Percentage of the task-resuming queries (first query on task) in post-switch session over positions.

Position in post-switch session	1	2	Other
Percentage	89.2%	8.9%	1.9%

of the search task that they might continue. This requires high precision for positive cases to avoid providing users with irrelevant recommendations. Therefore, we evaluate the performance of our approach against baseline method using accuracy, positive precision, positive recall, and area under the receiver-operator-characteristic curve (AUC) as the metrics of interest.

5. EXPERIMENTS

The dataset and labels used in the experiments are introduced in Sections 5.1–5.3. We then report the performance of baseline method and our approach in Section 5.4. As an additional experiment, the human performance on predicting contiguous cross-device search tasks is also provided (Section 5.4). Finally, we analyze the feature weights in the learned model by measuring their χ^2 value and information gain against labels (Section 5.5).

5.1 Dataset

The dataset used in the evaluation is a subset of the one introduced in Section 3.2, which only involves the device-switching from desktop to mobile. In addition, we also apply several constraints.

Search tasks vs. Long-term search interests: Popular queries, such as [facebook] and [youtube], have better chance to be resumed and are therefore easier to predict. However, they do not necessarily represent search tasks but rather are long-term search interests. Accordingly, we try to distinguish the search tasks from long-term search interests among all cross-device searches by limiting the frequency of pre-switch queries. Particularly, pre-switch queries (q_i) with personal frequencies (for the current user, within one month) of 5 or less, and global frequency (over all users, within one month) of 10 or less will be selected as potential search tasks. Note that we also run an experiment to test the performance of our predictive models if we retain these popular queries.

Device-switching frequency: We notice that the personal device-switching frequency follows a long-tail distribution, which means many users only switch several times within one month. Users with low switching frequency may not provide much information about daily cross-device search behavior. As a result, we only keep the users with at least 15 switches (within one month).

Table 7 shows the summary of the dataset used in evaluation.

5.2 Experimental Setup

In the experiments, we investigate the problem of predicting contiguous cross-device search tasks in two aspects: (1) the performance of our approach versus the baseline model; (2) the performance of the predictive model by adding features of different stages

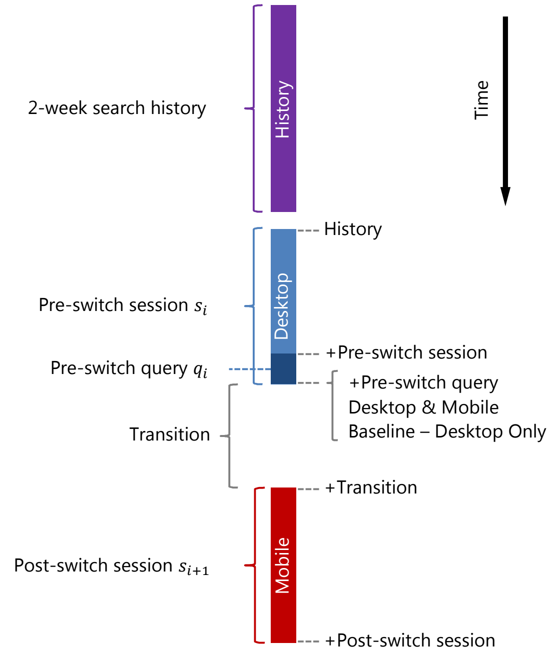

Figure 9. Timeline visualization of the various feature classes that are compared (on the right) and the point where these features are generated in Q in the experiments we perform.

(introduced in Section 4.3) incrementally. We leave two-weeks of search history data (around half of the dataset) for training the history-based features (shown in Table 6), which are designed to capture individual preferences of devices for search tasks. Then five-fold cross-validation is applied on the other half of the data to measure the performance of different classifiers.

Figure 9 shows points in the timeline where each model makes the prediction. Each model builds on the features that are available beforehand. For example, the system *+Pre-switch query* uses features from *History* + *Pre-switch session* + *Pre-switch query*. In the experiments, the model *Desktop & Mobile* and *+Pre-switch query* are identical. The purpose of having *Desktop & Mobile* is to highlight the comparison between our approach and the baseline method. We use Multiple Additive Regression Trees (MART) for classification [11]. The advantages of MART include model interpretability (e.g., a ranked list of features is generated), facility for rapid training and testing, and robustness against noisy labels and missing values.

5.3 Labeling

In Section 3, switches were defined to occur within six hours of the terminal search on the original device. However, if we imposed this temporal threshold on the labeled data, it would reveal something about the future (i.e., that the next query is on a different device and is within six hours) and could bias the labels. Therefore, for labeling, we do not impose the six-hour switch threshold. Instead, we simply label pairs of consecutive queries issued on different devices, irrespective of their inter-query time. Labels are obtained in two ways: automatic labeling and annotation from human labelers. Meanwhile, as a separate prediction task, we also investigate how well the judges can predict contiguous cross-device search tasks.

5.3.1 Automatic Labeling

To obtain labels for all the cross-device searches in our dataset, automatic task labeling is applied. We use function f (introduced in Section 4.2) to annotate the data. According to the performance reported in Table 5, the automatic labeler has a reliable precision. In total, 9.3% of the tasks were labeled as resumed post-switch.

Table 10. Performance of predicting contiguous cross-device search tasks with automatic labels. (indicates statistical significance at $p \leq 0.01$ using paired t-tests compared to the Baseline)**

	Accuracy	Positive Precision	Positive Recall	AUC
Baseline -- Desktop Only	0.903	0.337	0.037	0.646
Desktop & Mobile	**0.907****	**0.504****	**0.145****	**0.757****
History	0.880	0.250	0.142	0.661
+ Pre-switch Session	0.899	0.381	0.130	0.679
+ Pre-switch Query	0.907	0.504	0.145	0.734
+ Transition	**0.910**	0.544	**0.184**	0.781
+ Post-switch Session	**0.910**	**0.568**	0.169	**0.806**

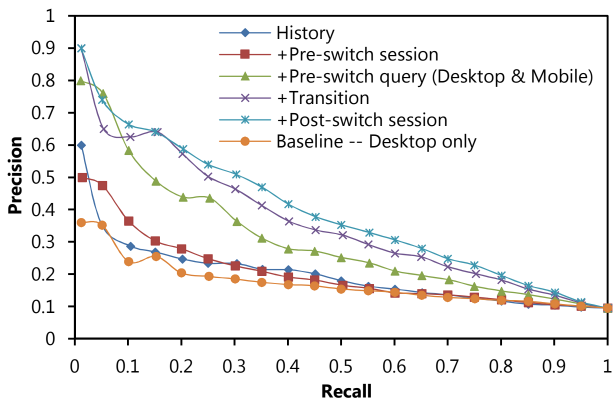

Figure 10. Precision-recall curve of predicting contiguous cross-device search tasks with automatic labels.

5.3.2 Human Labeling

Human annotators are assigned to two labeling tasks as following:

Labeling Task #1: Given the most recent search history (up to 5 queries) and the pre-switch query, annotators are asked to predict if the search task of pre-switch query will be resumed on mobile.

Labeling Task #2: Given the most recent search history (up to 5 queries), the pre-switch query, and the post-switch session, annotators are asked to label every query in the post-switch session if it belongs to the same search task as the pre-switch query.

Note that Task #1 differs from Task #2 in that it requires the judge to predict whether the task will be resumed using only pre-switch behavior. In Task #2, pre- and post-switch behavior is used.

Each of our five judges sees 200 switching instances. Inter-judge agreement is computed from the 50 overlapping labeling instances among the annotators. The remaining 150 instances were unique to each judge. This results in 800 distinct switch labels. Note that to increase the size of our training data, we included the overlapping instances and used the consensus label from all judges to label those instances. All 50 labels could be assigned this way (i.e., no ties).

The ground truth for our predictions is obtained from labeling task #2, and the human performance on predicting contiguous cross-device search tasks is measured from labeling task #1. We hired five annotators for each labeling task and provided them with the same detailed instructions about the judgment task, including example labels. The results are reported in Table 8. Labeling task #1 has poor agreement ($\kappa = 0.037$), suggesting that predicting task continuation is challenging even for human annotators. However, labeling task #2 has moderate agreement ($\kappa = 0.509$), implying that it is less difficult to judge also given the post-switch queries. Note that this task also shows that 14.9% of device switches comprise a contiguous search task, which is higher than the 9.3% reported by automatic labeling, likely because our judges are able to make inferences about task continuations that extend beyond the query overlap features used by the automated approach.

Table 11. Performance of predicting contiguous cross-device search tasks with human labels. (indicates statistical significance at $p \leq 0.01$ using paired t-tests compared to the Baseline).**

	Accuracy	Positive Precision	Positive Recall	AUC
Human	0.676	0.203	0.407	–
Baseline -- Desktop Only	0.827	0.162	0.025	0.548
Desktop & Mobile	**0.829**	**0.347****	**0.100****	**0.587****
History	0.823	0.271	0.084	0.560
+ Pre-switch Session	0.835	0.281	0.083	0.601
+ Pre-switch Query	0.829	0.347	**0.100**	0.587
+ Transition	0.837	0.324	0.083	**0.600**
+ Post-switch Session	**0.841**	**0.420**	0.099	0.592

Table 12. Performance of predicting contiguous cross-device search tasks (and long-term search interests) with automatic labels. (and * indicates statistical significance at $p \leq 0.01$ and $p \leq 0.05$ paired t-tests compared to the Baseline)**

	Accuracy	Positive Precision	Positive Recall	AUC
Baseline -- Desktop Only	0.911	0.770	0.472	0.924
Desktop & Mobile	**0.918****	**0.785***	**0.530****	**0.931****
History	0.878	0.558	0.390	0.911
+ Pre-switch Session	0.887	0.622	0.385	0.911
+ Pre-switch Query	0.918	0.785	0.530	0.931
+ Transition	0.919	0.779	0.542	0.933
+ Post-switch Session	**0.924**	**0.812**	**0.558**	**0.935**

As part of the human labeling, judges also indicate *which* queries in the post-switch session represent a continuation of the pre-switch search task. Table 9 shows that 89% of the task-resuming queries appear as the first query in post-switch session. This suggests that task-resumption support could be useful to searchers if we can accurately predict whether they are likely to resume. For example, if resumption is predicted when they are on the search engine homepage, we could populate the search box with their pre-switch query. Doing this for all queries (without prediction) could annoy users.

5.4 Results

We now present the results of the prediction experiments. Three groups of results are reported in this section. We begin by comparing the different feature classes on how well they predict the cross-device search tasks (with constrained pre-switch query frequency) identified using the automatically-generated labels.

Table 10 shows that *Desktop & Mobile* outperforms the baseline system. The big gap on both positive precision and recall between our system and the baseline system indicates the usefulness of the proposed features. The precision-recall curve in Figure 10 provides strong evidence that using cross-device search behavior can more accurately predict the task-resuming on mobile than the features of desktop search only. Also, the performance of the predictive model grows steadily as the feature classes are added.

The results in Table 11 are from human labels rather than the automatic labels described in Section 5.3.2. Similarly, the proposed model outperforms the baseline. However, we see a decrease in the performance of most systems by using the human labels. One of the reasons is that there are not enough positive cases in human labeled data for the classifier to learn a good decision boundary. The results of the human prediction task are also included in Table 11 (alongside *Human*). Unlike the mechanism of a classifier, human annotated labels are binary and there is no threshold to sweep, therefore AUC for human performance is not computed. Interestingly, our system attains higher positive precision than the human labelers, but lower positive recall. This implies the use of different prediction strategies by the annotators than encoded in the classifiers. Further analysis is needed to understand the nature of these differences.

Table 12 reports the performance on predicting all cross-device searches, which does not constrain the frequency of pre-switch queries (i.e., not excluding popular queries such as [facebook] and [youtube]). As expected, the prediction becomes less challenging because it is easier to capture users' long-term search interests and the frequent queries have better chance to be searched again, or resumed, in the future. But we can still see that the proposed system outperforms the baseline, especially on positive recall.

5.5 Feature Analysis

We now turn our attention to the features which contribute most to the prediction task. In order to show the effectiveness of the proposed features, we compute the χ^2 values and information gain of all the features against the automatic labels.

Table 13 ranks the top 10 features according to their χ^2 value. It clearly shows that users' search behavior on mobile (e.g., *NumRelatedQueryMobile*) and the cross-device search behavior features (e.g., *PreQueryContiguousSwitch*) are very effective in prediction. In addition, the temporal and geospatial features *during* the switch also show substantial weights for prediction. For comparison, we also analyze features against the automatic labels on all cross-device searches (including popular queries). Table 14 lists the top 10 features from that analysis. Not surprisingly, several baseline features appear on the list (e.g., *PersonalFrequency*). However, some cross-device behavior features (e.g., *NumOfContiguousSwitch*) and some transition-related features (e.g., *AvgSpeedSwitch*) still exhibited high evidential weights.

Table 13. The χ^2 value and info gain of features, ranked by χ^2 value.

Features	χ^2	Info Gain
NumRelatedQueryMobile	491.23172	0.031685
TimeIntervalSwitch	378.00423	0.024016
PreQueryContiguousSwitch	342.97317	0.020403
NumRelatedQuerySwitch	315.34983	0.020822
AvgSpeedSwitch	295.76765	0.022745
GeoDistanceSwitch	270.39483	0.020108
NumOfContiguousSwitch	235.27217	0.020083
EntropyAvg	221.82945	0.015446
NumRelatedQueryDesktop	172.73978	0.015161
IsSameLocationSwitch	95.81413	0.009118

Table 14. The χ^2 value and info gain of features on all cross-device searches (including frequent pre-switch queries), ranked by χ^2 value.

Features	χ^2	Info Gain
PreQueryContiguousSwitch	4225.69191	0.149665
NumRelatedQueryMobile	3969.11066	0.147362
NumExactQueryMobile	3408.70987	0.116981
NumOfContiguousSwitch	2755.01591	0.107953
PersonalFrequency	2711.70148	0.106203
NumExactQuerySwitch	2125.13168	0.074743
EntropyAvg	1716.86159	0.063697
PercentageMobileTime	1706.9318	0.059069
PercentageDesktopTime	1706.9318	0.059069
AvgSpeedSwitch	1545.90011	0.067262

6. DISCUSSION AND IMPLICATIONS

We have defined and presented a characterization of cross-device search and developed predictive models that can estimate whether a user will immediately resume on a mobile device the last search task they were attempting on their desktop computer. Our characterization shows interesting variations in switching over time, some topics are more likely to be associated with a switch, and that switching locations is often part of device switching. We also show that we can accurately predict contiguous cross-device task resumption, including more accurately than humans.

Although this is the first study, to our knowledge, to investigate cross-device search tasks specifically, the current trend toward multi-device use [8], suggests a need for further research in this area. For example, the findings that we present in this paper are focused on transitions from desktop to mobile search. However, switches in the opposite direction are also popular and need further exploration. Our early analysis of the Mobile-to-Desktop switches shows that they happen in much less time than Desktop-to-Mobile switches. One explanation for this is that switches in that direction are more related to searcher dissatisfaction with the mobile search results or general search experience on mobile (an assertion supported by non-search studies [17]). Additionally, since we do not interact with the users directly, we cannot be sure that users are actually continuing the same task. We need to work with users to fully understand task resumption, as well as directly measure other factors such as search satisfaction and device-switch motivations.

Beyond further research to better understand and characterize cross-device search tasks, we can also develop support to help searchers perform cross-device searching at different points in the switch, including support for so-called "slow search" (where an instant search-engine response may not be required). The predictive models described in this paper can have direct utility here. We can offer help at two particular points: (1) immediately following the session on the pre-switch device, and (2) on visiting the search engine homepage on the post-switch device.

Immediately following pre-switch session: Accurately predicting future task resumption at this point means that the search engine can perform actions on the users' behalf to maximize the downtime between task termination and resumption. Examples of what the search engine could do during this time include:

- Proactively save recent session state into server memory or disk cache for rapid access when the task is resumed.
- Try different ranking algorithms that may be less efficient but more effective, or issue multiple related queries and blend or summarize the results. The engine could also re-run recent abandoned queries, favoring quality over speed.
- Start a reconnaissance agent [23] to proactively retrieve content from the Web that pertains to the user's current task.
- Pose the query to a question answering site such as Yahoo! Answers (answers.yahoo.com), if the query is sufficiently descriptive to be posed as a question.

Our classifiers are important here because many of these actions are resource intensive and we do not want to perform them for all queries. One alternative to the prediction is to provide users with a way to tell the system that they will resume soon. This requires an additional action from users, which they may forget or be unwilling to perform, and it may not always be clear to the user at termination time that resumption is likely. A combination of such a capability plus prediction may work best, but testing is needed.

On visit to homepage: Once the user visits the homepage of the search engine on the pre-switch device we have access to features about the transition between devices that we show are useful in the task continuation prediction. Examples of support include:

- Provide the user with the option to explicitly resume their task. This would restore their state to that before the switch.
- Prefer pre-switch queries in auto-completion drop-downs or automatically populate the search box on homepage. Given

the difficulty that users can experience with typing on mobile devices [12], such support may speed up query entry.

We could also provide support on the basis of the queries on the post-switch device, which our results suggested could lead to even more accurate task-continuation prediction. Previous studies have shown that leveraging signals from recent session behavior can yield significant performance gains [5][35]. Extending a mobile search session back to the immediately-preceding desktop session may help address the "cold start" problem of insufficient context to personalize early queries on mobile devices [5].

7. CONCLUSIONS AND FUTURE WORK

As people's use of search technology transitions from a single device to multiple devices, understanding and supporting search tasks across those devices is becoming increasingly important. In this paper we have presented the first study of cross-device search tasks, focusing on users' continuing tasks immediately between devices. We analyzed a large set of device switches from logs of a search engine to understand the characteristics of switching. We showed that there were interesting patterns in the temporal, geospatial, and topical aspects of cross-device searching. We also developed classifiers capable of predicting whether users were going to resume a recently-terminated search task on their mobile device at different points in time, including immediately after they terminate the task and once they visit the search engine homepage on the device they are switching to. This affords a range of different types of search support that could be employed to help users tackle tasks that span different devices. Future work will focus on improving the accuracy of our classifiers, and designing and deploying task-continuation support to help users as they search across multiple devices.

REFERENCES

[1] E. Agichtein, E. Brill, and S. Dumais. Improving web search ranking by incorporating user behavior information. *SIGIR '06*, 19–26, 2006.

[2] E. Agichtein et al. Search, interrupted: understanding and predicting search task continuation. *SIGIR '12*, 315–324, 2012.

[3] L.M. Aiello et al. Behavior-driven clustering of queries into topics. *CIKM '11*, 1373–1382, 2011.

[4] A. Aula, N. Jhaveri, and M. Käki. Information search and re-access strategies of experienced Web users. *WWW '05*, 583–592, 2005.

[5] P.N. Bennett et al. Modeling and measuring the impact of short and long-term behavior on search personalization. *SIGIR '12*, 185–194, 2012.

[6] H. Cao et al. Context-aware query suggestion by mining click-through and session data. *SIGKDD '08*, 875–883, 2008.

[7] K. Church and B. Smyth. Understanding the intent behind mobile information needs. *IUI '09*, 245–256, 2009.

[8] ComScore. Digital omnivores: how tablets, smartphones and connected devices are changing U.S. digital media consumption habits. ipmark.com/pdf/Omnivoros.pdf (July 4, 2012).

[9] D. Dearman and J.S. Pierce. "It's on my other computer!": computing with multiple devices. *CHI '08*, 767–776, 2008.

[10] D. Donato et al. Do you want to take notes? identifying research missions in Yahoo! Search Pad. *WWW '10*, 321–330, 2010.

[11] J.H. Friedman, T. Hastie, and R. Tibshirani. *Additive logistic regression: A statistical view of boosting*. Technical Report, Department of Statistics, Stanford University, 1998.

[12] N. Henze, E. Rukzio, and S. Boll. Observational and experimental investigation of typing behavior using virtual keyboards on mobile devices. *SIGCHI '12*, 2659–2668, 2012.

[13] R. Jones and K. Klinkner. Beyond the session timeout: automatic hierarchical segmentation of search topics in query logs. *CIKM '08*, 699–708, 2008.

[14] M. Kamvar and S. Baluja. A large scale study of wireless search behavior: google mobile search. *SIGCHI '06*, 701–709, 2006.

[15] M. Kamvar et al. Computers and iphones and mobile phones, oh my!: a logs-based comparison of search users on different devices. *WWW '09*, 801–810, 2009.

[16] S.K. Kane et al. Exploring cross-device web use on PCs and mobile devices. *INTERACT*, 722–735, 2009.

[17] A.K. Karlson et al. Mobile taskflow in context: a screenshot study of smartphone usage. *SIGCHI '10*, 2009–2018, 2010.

[18] A.K. Karlson et al. Working overtime: patterns of smartphone and PC usage in the day of an information worker. *Pervasive '09*, 398–405, 2009.

[19] M. Kellar, C. Watters, and M. Shepherd. A field study characterizing Web-based information-seeking tasks. *JASIST*, 58(7): 999–1018, 2007.

[20] A. Kotov et al. Modeling and analysis of cross-session search tasks. *SIGIR '11,* 5–14, 2011.

[21] J. Li, S. Huffman, and A. Tokuda. Good abandonment in mobile and PC internet search. *SIGIR '09*, 43–50, 2009.

[22] Y. Li and N.J. Belkin. A faceted approach to conceptualizing tasks in information seeking. *IP&M*, 44(6): 1822–1837, 2008.

[23] H. Lieberman, C. Fry, and L. Weitzman. Exploring the web with reconnaissance agents. *CACM*, 44(8): 69–75, 2002.

[24] J. Liu and N.J. Belkin. Personalizing information retrieval for multi-session tasks: The roles of task stage and task type. *SIGIR '10*, 26–33, 2010.

[25] B. MacKay and C. Watters. Exploring multi-session web tasks. *CHI '08*, 1187–1196, 2008.

[26] Q. Mei et al. An analysis framework for search sequences. *CIKM '09*, 1991–1994, 2009.

[27] D. Morris, M.R. Morris, and G. Venolia. SearchBar: a search-centric Web history for task resumption and information re-finding. *SIGCHI '08*, 1207–1216, 2008.

[28] A. Singla, R.W. White, and J. Huang. Studying trailfinding algorithms for enhanced web search. *SIGIR '10*, 443–450, 2010.

[29] A. Spink et al. From e-sex to e-commerce: Web search changes. *IEEE Computer*, 35(3): 107–109, 2002.

[30] J. Teevan et al. Information re-retrieval: repeat queries in Yahoo's logs. *SIGIR '07*, 151–158, 2007.

[31] J. Teevan et al. The perfect search engine is not enough: a study of orienteering behavior in directed search. *SIGCHI '04*, 415–422, 2004.

[32] J. Teevan et al. Understanding the importance of location, time, and people in mobile local search behavior, *MobileHCI '11*, 77–80, 2011.

[33] J. Yi, F. Maghoul, and J. Pedersen. Deciphering mobile search patterns: a study of Yahoo! mobile search queries. *WWW '08*, 257–266, 2008.

[34] R.W. White and S.M. Drucker. Investigating behavioral variability in Web search. *WWW '07*, 21–30, 2007.

[35] R.W. White, P.N. Bennett, and S.T. Dumais. Predicting short-term interests using activity-based search context. *CIKM '10*, 1009–1018, 2010.