An Analysis of Cross-Device Search*

George D. Montañez
Carnegie Mellon University
Pittsburgh, PA 15213 USA
gmontane@cs.cmu.edu

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

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

ABSTRACT

Background. Ownership and use of multiple devices such as desktop computers, smartphones, and tablets is increasing rapidly. Search is popular and people often perform search tasks that span device boundaries. Understanding how these devices are used and how people transition between them during information seeking is essential in developing search support for a multi-device world.

Aim. The aim of this study is to characterize multi-device search behavior in a search log dataset and to demonstrate that sufficiently non-random behavior patterns emerge for search across multiple devices. We further show that such patterns can be reliably exploited to predict device-to-device transitions by machine learning methods trained on historical search log data.

Data. We analyze a dataset of user search log records drawn from a popular commercial search engine, gathered by the authors during 2013 from the internal database of that search engine. While the dataset is proprietary and agreements prevent us from identifying either the particular search engine or the time period spanned by the data, we can give specifics concerning the form and use of the data. The data consists of roughly two billion English language queries submitted by approximately thirty-three million users in the United States. The data includes query terms, user identifiers, device information, location data, and time of query. In addition, rows have been annotated with topical and historical user information, generated using trained topic classifiers and constructed from simple historical statistics. Three sub-datasets were constructed from the main dataset, one for each of the three learning tasks completed in this study.

Methods. We characterize multi-device search across four device types, including aspects of search behavior on each device (e.g., topics of interest) and characteristics of device transitions. Building on the characterization, we learn models to predict various aspects of cross-device search using existing machine learning methods. Three tasks were completed, which were: (1) predicting the next device used for search, (2) predicting if a device switch occurs, and (3) predicting the next device given a device switch occurs.

Results. Our analysis shows that people use different devices to search for different content, and time of day interacts with device to affect content sought. We found that exploitable patterns emerge for device transitions, with previous device signaling the next device, even when they differ. For two of the prediction tasks (i.e., predict next device and predict next device given switch), predictive accuracy, recall and precision exceeded 90% with a fairly compact feature set. The remaining task (predict device switch) also saw significant gains in accuracy over the baselines, approaching 80% given all features.

Conclusions. Our results give evidence that learnable patterns exist in user device transition behavior. Being able to accurately predict device transition behavior enables many applications. For example, accurately forecasting the device used for the next query lets search engines proactively retrieve device-appropriate content (e.g., short documents for smartphones) for likely topics of interest (which differ by device), while knowledge of the current device combined with device-specific topical interest models may assist in better query-sense disambiguation. The former becomes especially important when combined with search-task continuation prediction systems, which can predict whether a user is likely to continue searching for a specific topic within the near future.

1. INTRODUCTION

Cross-device search is an important emerging domain. The number of people who own and use multiple devices such as desktop computers, smartphones, and tablets has increased rapidly [6]. Search across multiple devices by an individual has become a common usage pattern since people can query search providers almost anytime and from anywhere [34]. In addition, the functional boundaries between different computing devices have become blurred. For example, gaming console users can now conduct Web searches directly from consoles and use applications previously only found on other devices. Figure 1 presents an example of a single user's search activity across four types of device (desktop or laptop computer (referred to as "PC" in this paper), smartphone, tablet, and gaming console) within a single day. This example is drawn from the logs of a large commercial search engine used in our analysis, but query text is replaced with similar alternatives to preserve anonymity. According to our analysis, described later in the paper, at least 5% of searchers are multi-device users, with queries from such searchers accounting for 16% of search volume on the engine studied (i.e., such searchers are highly engaged). Better support for these multi-device searchers is therefore important both for the research community and for search providers.

Despite the importance of these highly-engaged users, supporting cross-device search is a challenging task. From the

*An abridged version of this paper was presented at the 23rd International Conference on Information and Knowledge Management (CIKM 2014) under the title "Cross-Device Search."

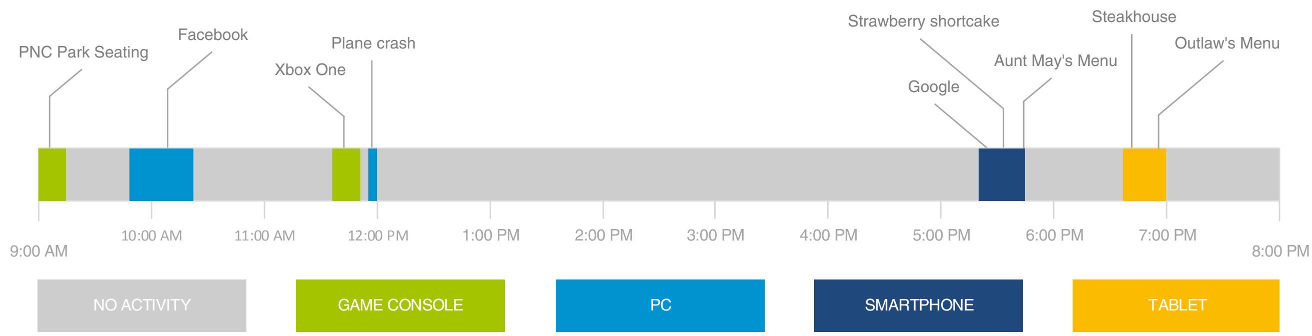


Figure 1: Fictionalized user timeline for one day, based on log data. Queries of interest shown on each device.

example presented in Figure 1 we can observe different topical patterns and different temporal patterns. For this searcher on this day, the PC and the gaming console are used in the morning (perhaps when they are at home), and smartphone and tablet are used in the evening (when they may be at work or commuting). The morning activity involves planning future events and staying updated on events in social and news media. In the evening, we also observe a longer-running search task between smartphone and tablet for dining related topics; the topic and the mobile nature of the devices used suggest that the searcher may not be at home. Recent work has shown that the nature of the current location can be estimated using geolocation data [20]. Figure 1 also suggests some predictability in usage patterns. Aside from the time of day that the devices are utilized, gaming console usage precedes PC usage in both instances. We refer to these switches as *cross-device transitions*. All that being said, this represents only one user's (fictionalized) search behavior on one day. Improving the experience for all searchers as they transition between devices requires a better understanding of multi-device usage patterns over many searchers and queries. We present such an analysis as part of the research described in this paper.

Multi-device behavior has been studied in the human factors community [7, 17], but without an emphasis on search. Search on different devices has also been studied separately with a variety of datasets [13, 14, 29]. However, the transitions between devices were not analyzed. Wang et al. [34] examined cross-device task continuation, but only from PC to smartphone and for a particular definition of search task. We examine a more general scenario, with up to four device types, and target a broader set of prediction tasks including switch detection and the identification of target devices.

The main contributions of our research are:

- Introduce *cross-device search* as an important emerging domain in the field of information retrieval.

- Analyze multi-device usage in Web search on four device types, including gaming consoles, an important emerging platform. Although not our primary focus of this study, our research is the first examination of search behavior on gaming consoles.

- Characterize transitions between devices, which is an area where search engines could provide more direct support for applications such as task continuation. We explore topical, temporal, and historical aspects.

- Develop models to accurately predict the next device from which a searcher will query, to predict if a device-switch will take place, and to predict the next device given that a switch will occur. Among other things, this enables the search engine to provide support such as proactively retrieving device-appropriate content (e.g., favoring shorter articles if the searcher will transition to a smartphone), better assess the scope and semantics of a query, or detect device switches if device information is missing from a request.

The remainder of the paper is structured as follows. Section 2 describes relevant related work. We characterize some key aspects of multi-device use in Section 3 and outline our prediction models in Section 4, along with experimental results. We discuss the findings and their implications in Section 5 and conclude in Section 6.

2. RELATED WORK

Relevant related work falls into the following three areas: (1) large-scale log analysis of search behavior; (2) studies of user behavior, especially characterizations of that behavior, on desktop and mobile, and; (3) user behavior across multiple devices, including their use in domains beyond search.

Behavioral logs from search engines are valuable in understanding how people search in naturalistic settings. Research has focused on the use of automated methods to analyze and predict aspects of search behavior for individual queries [30] and search sessions [2, 35] using logs. Qualitative studies have sought a deeper understanding of the nature and motivations underlying online searching [18]. Other research has focused on tasks that extend over time, but have not considered different devices [1, 19, 22, 23].

Studies of search in mobile settings have examined the characteristics of queries issued from mobile devices, analyzing behavior along different dimensions such as geographic location and search interfaces [3, 36]. Others have studied mobile search intent and the effect of contextual factors on behavior. Church and Smith [5] studied mobile search, focusing on their underlying intents, topics, and the impact of contexts such as location and time. Teevan et al. [33] showed that local searches are influenced by geography, time, and social context. Smaller scale studies of online behavior have considered other devices such as tablets [25], and mobile device usage rationales [26, 31].

Research on comparing and contrasting search behavior on multiple devices is also relevant [13, 14, 21]. Kamvar and

Baluja [13] describe a large-scale study of search patterns between phones, personal digital assistants, 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 clickthrough. Kamvar et al. [14] presented a log-based comparison of search patterns on different devices (computers and mobile). They showed that search usage is more focused on mobile than on computer, but behavior on high-end phones resembles computer-based search. Li et al. [21] studied good abandonment of search results on desktop and mobile (where users do not click but are still satisfied) and showed that it is significantly higher in mobile settings. Song et al. [29] compared search behavior on three different platforms—desktop, mobile, and tablet—and developed specialized rankers for each platform separately.

Wang et al. [34] examined cross-device search tasks initiated on a desktop computer and resuming soon thereafter on a mobile device. They performed a detailed analysis on topics and transition times for these tasks and showed, for example, that interdevice time varied by time of day. They developed models to accurately predict task continuation from PC to mobile. We address a more general scenario, with up to four device types, and predict key aspects of cross-device search that were not addressed in [34], including whether a person will switch devices and their next device.

There has been other research on multi-device use in domains beyond search. Studies have shown that user activities tend to span multiple devices [2] and frustrating experiences on mobile devices will drive users to complete tasks on conventional computers [16]. Karlson et al. [17] analyzed the usage log of desktops and mobile phones from a user study and showed that there is little support for carrying over tasks between devices. Kane et al. [15] studied Web browsing usage patterns across devices. Their results indicated sharing browsing information between devices could help improve the effectiveness of browsing on mobile devices. Dearman and Pierce [7] conducted an interview study of multiple device use and showed that current support is inadequate.

The research described in this paper is the first to study cross-device search in a general sense. It also extends previous work in a number of ways. First, we focus on multi-device usage during Web search across four device types, including search on gaming consoles. Previous studies on mobile, desktop, and tablet search have focused on devices independently, and not considered cross-device searching within users (i.e., the same user transitioning between devices over time, as seen in Figure 1). Second, we study key aspects of multi-device usage such as differences in topical interests and usage patterns on these devices. Third, we characterize aspects of the *transitions* between devices, which is an activity where search engines could help directly. Finally, we implement classifiers to predict aspects of cross-device search, demonstrating that this can be accurately forecast using a range of features. Foreknowledge of the next device (especially coupled with task-continuation prediction [1, 34]) can help the search engine select device-appropriate content, feeding into methods to help people search over time [8, 24].

3. CROSS-DEVICE SEARCH ANALYSIS

We analyzed sampled search logs from a large commercial search engine, over a period of several months. Our goal was to understand search behavior across devices along different dimensions, including time and topic. We consider queries on four types of device: PC (desktop or laptop computers, which were indistinguishable in our logs), smartphones, tablets, and gaming consoles. Our dataset comprises 2,271,142,893 records from the United States English language locale (en-US) market, representing queries from 33,221,253 users. We map searchers across devices using unique identifiers obtained from those who queried when signed into the engine (for most searchers, sign-in happened automatically), allowing us to identify the same searcher as they moved across devices. We retained only records with non-empty queries, filtering abnormal records (such as those from bot traffic) and further eliminated records without query classification information (labeling the type and optionally the query type/topic (e.g., navigational) with proprietary classifiers, discussed more later).

Our filtering methods reduced the number of smartphone records, since a significant proportion were missing meta information (e.g., user identifiers). We are still able to reliably track users who were signed in on all devices (the default for most users), allowing us to analyze transitions between devices over time. While the sample of smartphone users is sufficient to justify analysis, the relative proportions of each device type are likely not representative of general multi-device behavior, independent of search engine. The numbers reported here primarily document the sample sizes used for our analyses, and the multi-device usage seen here likely represents a lower bound on actual multi-device usage. Table 1 presents the final query counts from this filtered dataset issued on each of the four devices considered in our study.

Table 1: Query dataset statistics.

Description	Total	%
Queries	2,271,142,893	100.00
Multi-Device User Queries	370,865,428	16.33
PC Queries	2,074,083,054	91.32
Smartphone Queries	53,939,886	2.38
Tablet Queries	137,979,833	6.08
Gaming Console Queries	1,854,422	0.08

3.1 Users and Queries

Of the 33 million unique users in our primary dataset, 1.68 million (5.04%) were observed using more than one device. These users are extremely active searchers; their queries comprise over 16% of the total search volume in our dataset (see Table 1). Understanding and supporting these users' search activity is therefore of critical importance to search providers. Delving into the specifics of observed device usage, Table 2 presents the number of users associated with each type of device, as well as the number of users associated with each combination of device type.

As we can see, the most common device is PC, accounting for roughly 95% of search volume, and most users (95.0%) are associated with a single device. A significant fraction of people use two devices and a small percentage (less than 0.01%) search on all four devices. The most common device pairings are PC-smartphone and PC-tablet, reflecting the still-central role of personal computers in search.

3.2 Multi-Device Locations

In addition to considering aggregate statistics about device usage and the distributions of device combinations, the

Table 2: User device-type statistics.

Device(s)	Users	%
Any Device(s)	33,221,253	100.00
More Than One Device	1,675,272	5.04
One Device	31,545,981	94.96
Two Devices	1,585,018	4.77
Three Devices	89,834	0.27
Four Devices	420	< 0.01
PC	31,770,955	95.63
Smartphone	1,301,717	3.92
Tablet	1,863,783	5.61
Gaming Console	50,744	0.15
PC-Smartphone	696,928	2.10
PC-Tablet	1,022,279	3.08
PC-Console	20,153	0.06
Smartphone-Tablet	111,554	0.34
Smartphone-Console	4,026	0.01
Tablet-Console	2,100	< 0.01
PC-Smartphone-Tablet	86,840	0.26
PC-Smartphone-Console	2,689	< 0.01
PC-Tablet-Console	1,521	< 0.01
Smartphone-Tablet-Console	464	< 0.01

geographic distribution of multi-device search was also investigated. Given that our filtered search queries contain geolocation information (e.g., latitude and longitude, IP address, etc.), we are able to gain a rough understanding of the extent and geographic distribution of cross-device search. As a safeguard against high-variance estimates caused by cities with small sample sizes, we examined those cities with a population of at least 250,000 individuals to ensure that there was sufficient data from each city from which to reliably compute multi-device usage estimates.

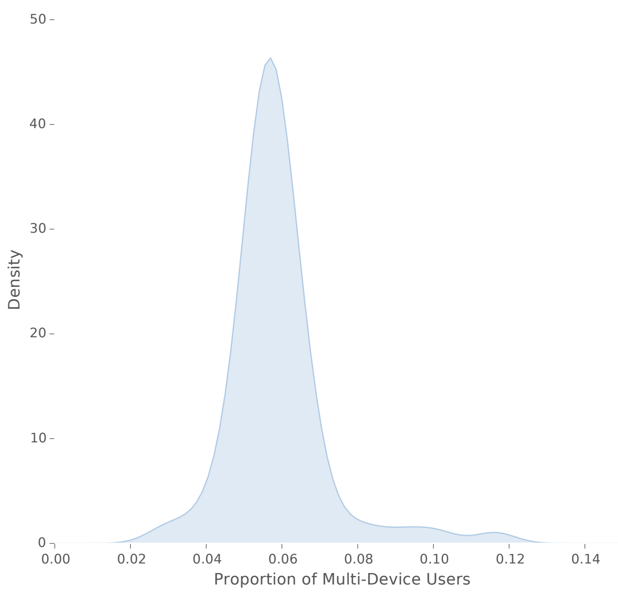


Figure 2: Density estimate of per city multi-device usage proportion, for populous U.S. cities.

Figure 2 shows the estimated distribution of multi-device usage proportion per city, using nonparametric Gaussian kernel density estimation (with bandwidth $h = 0.167$). On average across these cities, 6% of search engine users were multi-device users, ranging from 2.9% in San Jose, CA to 11.6% in Seattle, WA. The top 10 cities and their associated multi-device usage percentages are shown in Table 3, along with the number of unique users of the search engine who queried from that city and were in our dataset, and the population for the city from U.S. census data. Ranking is by descending multi-device proportion, which is the number of unique users in the location who use multiple devices divided by the total number of unique users for the location.

Table 3: Top 10 multi-device searcher cities, ranked by proportion of multiple device users in those cities.

City	State	Proportion	Unique Users	Population
Seattle	WA	0.116	253,618	634,535
Dallas	TX	0.101	576,607	1,241,162
San Francisco	CA	0.093	162,823	825,863
Denver	CO	0.083	185,392	634,265
Albuquerque	NM	0.069	124,704	555,417
Columbus	OH	0.069	196,209	809,798
Boston	MA	0.068	265,207	636,479
Colorado Springs	CO	0.068	106,890	431,834
Philadelphia	PA	0.066	327,850	1,547,607
Portland	OR	0.066	121,546	603,106

Figure 3 plots multi-device user proportions by city, normalized by number of distinct users in the city, with each point scaled by the proportion of users that were multi-device users for each location. We can see that the multi-device "hotspot" cities are distributed throughout the United States and are not confined to any particular region, existing on the West Coast (Seattle and San Fransisco), Gulf Coast (Houston and New Orleans), Great Plains (Denver) and Eastern U.S. (Boston and New York City).

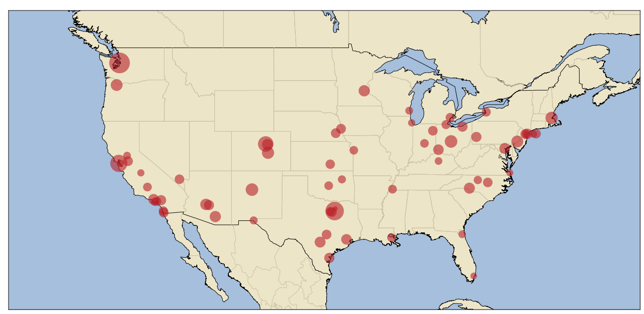


Figure 3: Populous cities in the United States with highest normalized proportion of multi-device users.

To better understand the nature of the locations where multi-device usage was more prevalent, we joined the multi-device user data with the most recent demographic data available from the U.S. census (2012). Using that census data we obtained statistics on the following attributes for each city: (1) percentage of population under 18 years, (2) percentage of females in population, (3) average household size, (4) median per capita income, (5) percentage of population with high school diploma (Edu1), and (6) percentage of population with college degree (Edu2). Note that the census data were only available at the county level, but since the cities we used were large, is it likely that they dominated the county demographics. We then computed the Pearson correlation coefficients (r) between the fraction of users on multiple devices and each of the six demographic statistics listed above. The r-value and p-value associated with the one-tailed t-test is shown in Table 4.

Table 4: Pearson correlation between city multi-device usage and U.S. census demographic data.

	Age	Sex (F)	H. Size	Income	Edu1	Edu2
Pearson r	-0.296	-0.314	-0.469	0.449	0.434	0.509
p-value, $t(65)$	0.025	0.018	0.001	0.001	0.001	< 0.001

We can see from Table 4 that correlations are reasonable (ranging from -0.30 to 0.51), and cities with a greater fraction of multi-device users do contain significantly more men, a more highly educated populace, and a higher income. There are also indications that these cities are more likely to have an older population (at least over 18) and more likely to have smaller households. Although these findings are interesting, more work is needed to understand how demographics contribute to usage and interact with other factors, such as the sources of employment in locations.

Thus far we have considered aggregate usage statistics describing characteristics of multi-device usage, its geographic distribution, and its relationship with demographics. However, to fully realize the vision of supporting users in search across devices, we need to understand user interests on devices and aspects of transitions between devices.

3.3 Query Topic Distributions

We begin by focusing on how searchers' content interests differ among the four device types. Understanding the topics of interest on each device can help establish interest priors to pre-fetch appropriate content or personalization during "cold start" scenarios [28]. To do this, we analyzed the distribution of query topics in the dataset. We classified each query using proprietary classifiers (used by the search engine in determining if support such as instant answers should appear on the result page and characterizing query type for the search engine), corresponding to around 50 query categories. A query could belong to multiple categories. The categories were then grouped by the authors into fifteen higher-level topics, including *Movies and TV*, *Music*, and *Celebrities*.

3.3.1 Topic Distributions Per Device

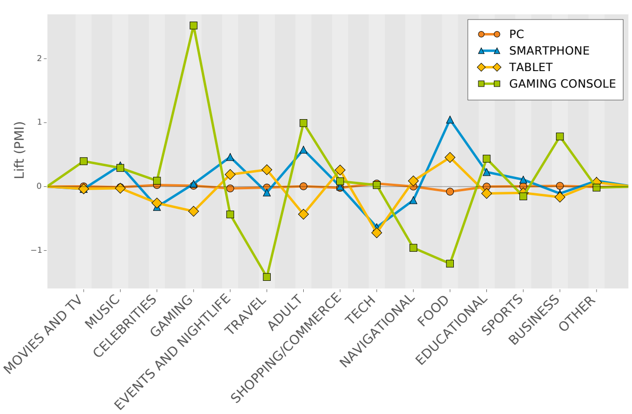


Figure 4: Topic probability lift (PMI) by device.

Figure 4 displays the differences in topical interest between devices, by using the pointwise mutual information (PMI) between $\mathbb{P}(\text{topic}|\text{device})$ and $\mathbb{P}(\text{topic})$, calculated as $\log \frac{\mathbb{P}(\text{topic}|\text{device})}{\mathbb{P}(\text{topic})}$. Positive PMI values indicate an increase in topic popularity on a particular device, with negative values indicating a decrease in popularity. Since the PC contained most query volume, it was most similar to the background model. The most prominent change is the sharp increase in gaming related queries on gaming consoles. Also apparent in Figure 1 is the increase in food-related queries on mobile devices associated with dining (as in Figure 1).

3.3.2 Topic Distributions Per Device Over Time

Rather than assuming a static view of topical interests aggregated over time, we were also interested in how they varied temporally. The queries were partitioned by device and hour of day, taking counts of the number of each type of query in each partition. The counts were then normalized per partition, resulting in a per-device topic distribution over time. We could then investigate how the topic distribution changed throughout the day.

Figures 5-8 show the topic proportions for each device type, over three-hour time segments of the day. We see that navigational queries, where users attempt to locate websites or services, are most popular during working hours for all devices but game consoles. The relatively small amount of data for gaming consoles, however, increases the noise in those topic proportion estimate and obscures trends for that device. Clearly visible is the increase in adult related queries between the hours of midnight and 6am (i.e., hours 0-2, 3-5) for all devices. Also visible are the increase of food related queries on smartphones and tablets as the workday progresses.

Figure 5: Topic proportions by hour of day, PC.

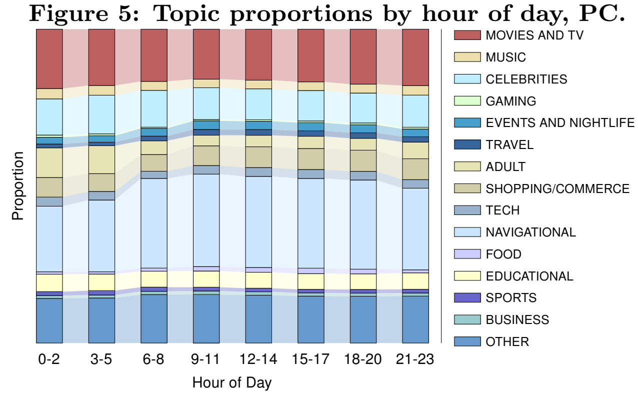
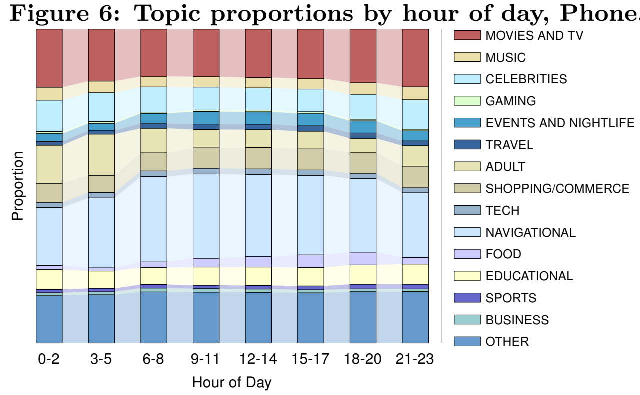


Figure 6: Topic proportions by hour of day, Phone.



To better quantify these trends, we calculated the PMI divergence of $\mathbb{P}(\text{topic}|\text{device, hour})$ from $\mathbb{P}(\text{topic}|\text{device})$. Figure 9 shows how topics with large absolute PMI divergence shift throughout the day. For each device we plot only the

Figure 7: Topic proportions by hour of day, Tablet.

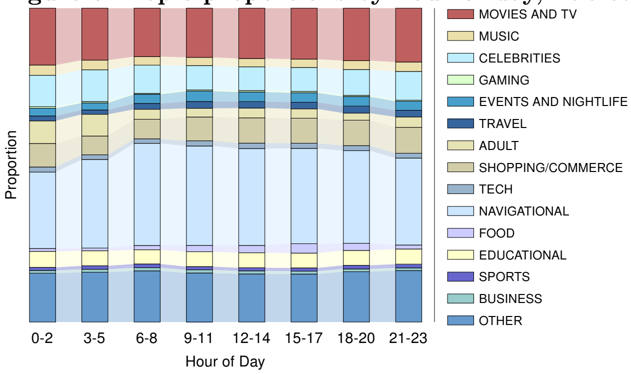
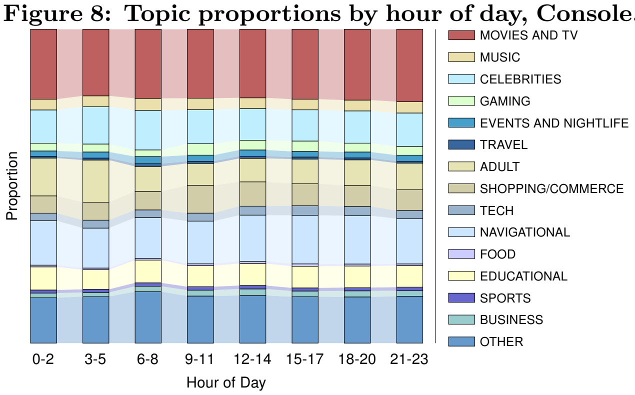


Figure 8: Topic proportions by hour of day, Console.



top three topics, in descending order by largest absolute PMI lift throughout the day. Distributions for gaming console is least smooth given data sparseness.

Across the four sub-figures, we observe some noteworthy trends. First, interest in food and adult content appear to be inversely related for three of the four devices: searching for adult material is common at night and less common during the work day, food searching exhibits the opposite trend. We notice two peaks of interest for food related queries, one around lunch hour and another around dinner time. Querying for gaming related activities (including social media games from facebook.com and zynga.com) is popular on PC and tablet late at night, with interest in gaming on tablets exceeding expectations for most of the day. Halvey et al. [11] studied temporal dynamics in topical interests on early smartphones. However, they used a different content classification and studied Web surfing not Web search, making direct comparisons with this work difficult.

It is clear that there are significant variations in interests across devices and time, and that both (and their interaction) need to be considered by search engines when supporting search on different devices. The analysis can also help generate features for the device identification aspects of the later prediction tasks. However, we focus on *cross-device* behavior, especially device transitions. This is a critical and defining activity in cross-device search and one that needs better support, for a range of information tasks [7, 17].

3.4 Device Transitions

We now describe the characteristics of device transitions within our dataset.

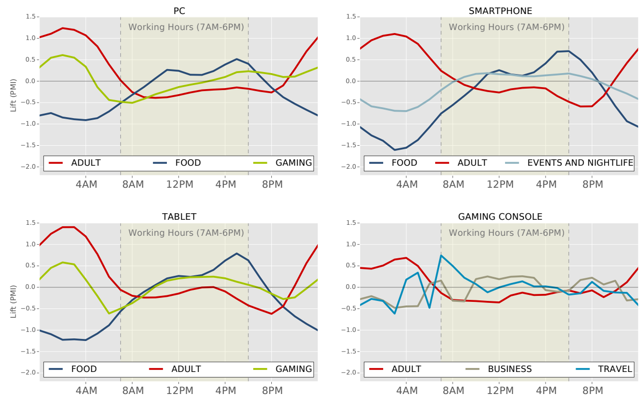


Figure 9: Topic probability lift (pointwise mutual information) by device type and hour of day.

3.4.1 Device-Device Transition Probabilities

We define a *transition* as a pair of consecutive queries issued on the same or different device. *Cross-device* transitions include a device change between consecutive queries. To improve the likelihood that the transitions are meaningful, we limit consideration to transitions with delay times of three hours or less between consecutive queries. Previous work [34] used a six hour threshold, but had a simpler switching scenario (single device pair, single direction). To begin, we computed the maximum likelihood estimates for device transitions as a Markov graph, conditioned on the previous device, which are the empirically observed transition probabilities. Figure 10 shows the probability of the transitions, as percentages, given the previous device.

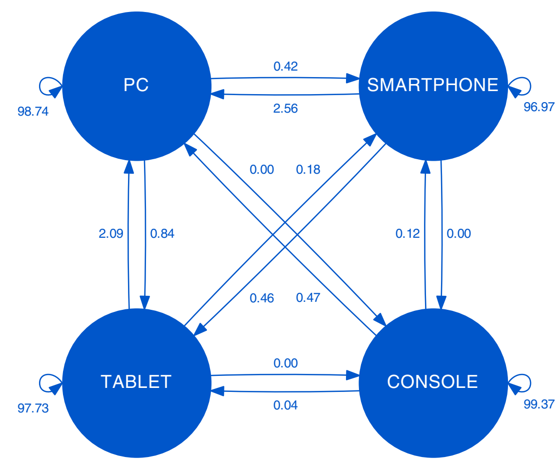


Figure 10: Device transition probabilities (%).

We can see from the figure that the majority of transitions are self-transitions, meaning that the user is likely to continue using the same device. This seems reasonable as searchers are known to search in bursts (as search sessions [35]) and may be unlikely to change device mid-session (although as we show later, cross-device transition delays are also often short). The transition analysis becomes more interesting if we remove self-transitions and only consider transitions between different devices. That is, only cases where the first query in the transition is on one device and the second query is on a different device. This allows us to generate the modified transition graph shown in Figure 11.

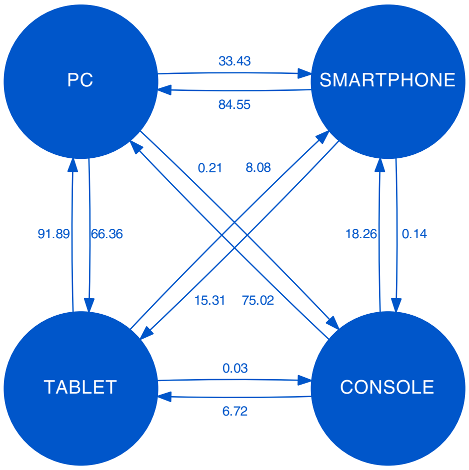


Figure 11: Cross-device transition probabilities (%).

Figure 11 shows that although most transition mass leads to PC devices (and very little leads to gaming console devices), there is also evidence of significant interplay between particular pairs of devices. Specifically, we can observe that PC-smartphone and PC-tablet exhibit a close relationship, with transitions between those devices (in both directions) being particularly common. Our earlier statistics (in Table 2) suggested that these pairs were most likely to be used by the same searchers, but here we show more direct evidence of interaction between them. More work is needed on understanding the nature of the search tasks that people pursue before and after these transitions, perhaps via qualitative user studies, e.g., rapid switches from smartphone to PC may be indicative of a dissatisfying mobile search experience, as has been suggested in a non-search setting [16].

3.4.2 *Previous Topic to Next Device*

We make some progress in understanding task continuation by examining the relationship between the immediately preceding query topic and the device used for the next query. Figure 12 shows that while to-PC transitions continue to dominate, certain topics indicate a shift in next-device probability. The proportions in the figure should be compared to the overall likelihood of transitions into the devices, independent of preceding topic (i.e., PC: 63.9%, Smartphone: 11.2%, Tablet: 24.6%, Console: 0.3%, all shown at the top of Figure 6). It is clear from the figure that the PC dominates given the strong prior. However, there are cases where other devices become more evident depending on the preceding topic. For example, for *Events and Nightlife*-related previous queries, the likelihood of using a gaming console next increases by an extraordinary 870% over the background. Similarly, searches for *Celebrities*-related content signifies an increased likelihood of using a tablet device for the next query (a 34% increase over the background), perhaps signifying that searchers are actively engaged in leisure pursuits.

3.4.3 *Previous Hour to Next Device*

We also analyzed the temporal relationship between the previous hour and next device that was used, as shown in Figure 13. We focus on the previous hour and not the current hour for two reasons. First, focusing on current hour simply gives proportions reflective of the background device proportions in most cases, and does not produce informative results. More importantly, any system aiming to predict device transitions (see Section 4) may benefit from exploitable patterns linking past query times to future device choices. Inspecting our results, we find that PC usage is most likely during the workday, when the previous query was issued between 7AM to 5PM. Transitions to tablets and smartphones are most likely when searching in the early morning and late evening (perhaps signifying commuting activities), and transition to a gaming console reaches its peak probability between 12AM and 4AM; late-night gaming practices are well known [9].

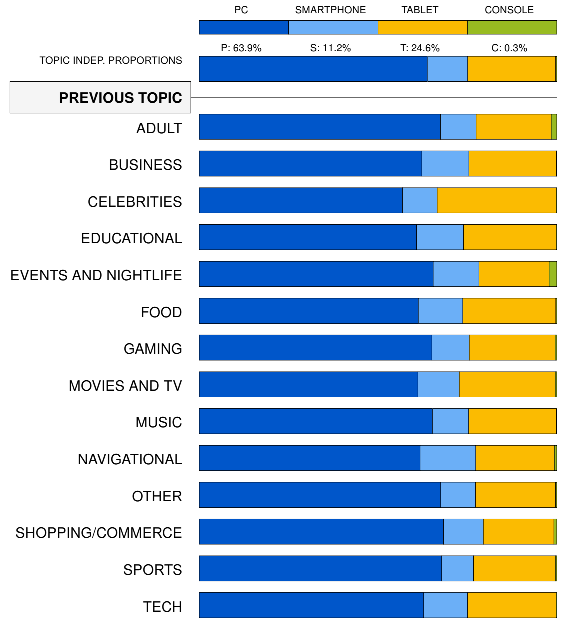


Figure 12: Previous topic to current device proportions.

The variations in future device usage given *previous* topic and query appear promising for the prediction task described later. Before considering that task, we also examine the time between queries, referred to as the *transition delay*, since a regressor to predict transition delay may also have utility, e.g., in determining the amount of time that the engine has to proactively work on the searcher's behalf or employ crowdworkers to help with an ongoing search task [32].

3.4.4 *Device Transition Delay Times*

To study transition delays, we examine all transitions (including self-transitions, which are the majority and take almost no time) and cross-device transitions. As we will show in Section 4.2, the transition delay time (DT) has strong predictive value for whether a device switch occurred. Figure 14 depicts the proportion of delay times for all transitions and all cross-device transitions (times in seconds from 0 to 10,800 (3 hrs)) on a log-log plot. A spike occurs for all transitions at a delay time of approximately five minutes, perhaps associated with session termination [12]. In general, cross-device transitions take longer on average, although the distributions are heavily skewed, with many short delay times (especially for the "all transitions" set) and long tails.

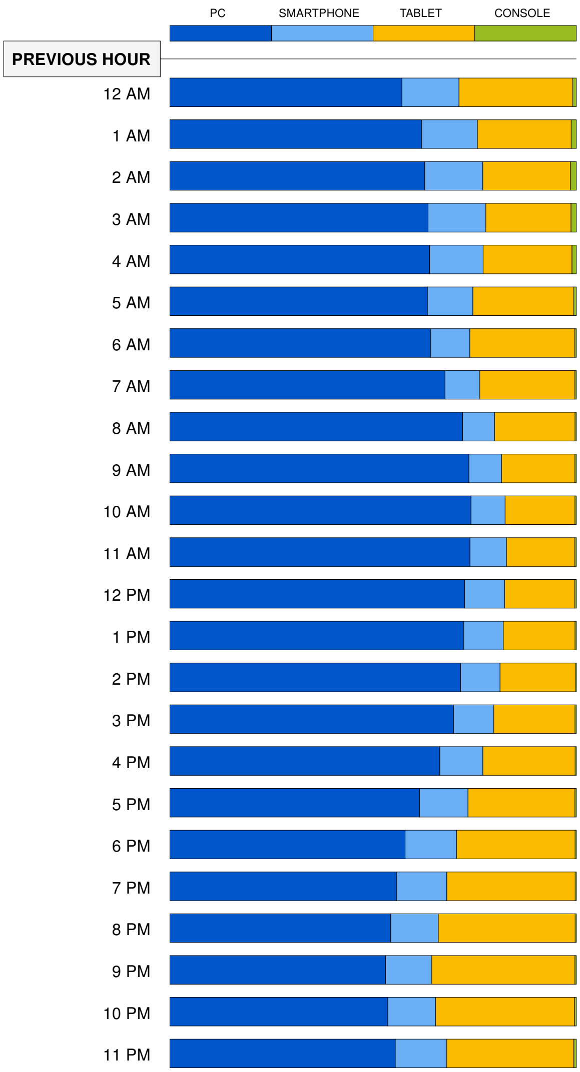


Figure 13: Previous hour to current device proportions.

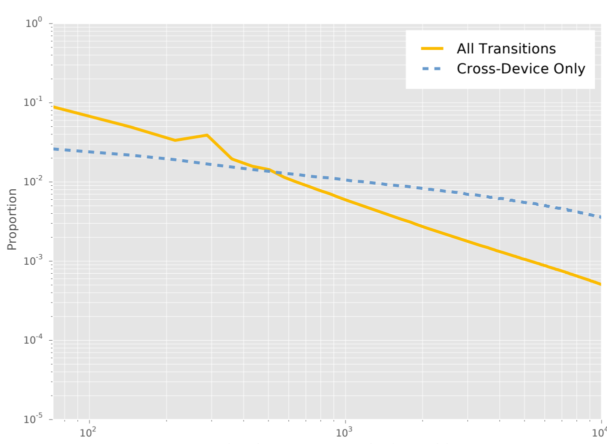


Figure 14: Delay time for all and cross-device transitions. Log-log plot is used to highlight differences.

4. PREDICTING DEVICE TRANSITIONS

We now focus on predicting transitions between devices. Using features such as those outlined in the previous section, including the searcher's previous device, query times, and transition history, we perform three prediction tasks:

1. Predict the next device that a person will search from. This could help engines select device-appropriate information in anticipation of the next device used.

2. Predict whether a device switch has occurred between two consecutive queries. This could help if device information is missing at query time. For example, a personalization system should use a cloud-based user profile (not a client-side profile) if a switch is detected.

3. Predict the next device given a switch. This could help if device information is missing, facilitates selection of device-appropriate information, and may also help with query-sense disambiguation.

4.1 Methodology

4.1.1 Datasets

We predict device transitions using three datasets derived from our primary search log dataset described earlier. From these queries, we created a set of device-device transition data, which was then used to create three datasets of features for training and testing. The first dataset ("Main") contained both same- and cross-device transitions. The second dataset ("Balanced") consisted of an equal 50-50 mix of same-device and cross-device transitions, used in predicting whether or not a change in device would occur during a transition. The final dataset ("Cross-Device Only"), comprised only instances where the next device did not equal the previous device and the user had at minimum 200 transitions observed in the historical dataset (from which user specific transition statistics were computed). For all three datasets, 250,000 instances were selected at random for five-fold cross-validation. It should be noted that sampling was done with respect to filtered transitions rather than with respect to users or queries, so the final datasets do not necessarily reflect user and device type proportions reported earlier.

4.1.2 Features

A set of 177 features were used for training and testing. These include features for the previous device used, probabilities of device-type usage, and device-pair transition probabilities for the current user and globally. Other features such as the previous query topic and the previous hour are directly informed by the analysis presented in the previous section. For historical features, a separate dataset from previous months was used to compute the feature statistics, allowing for no overlap between historical data and training/test data.

Table 5 lists the features used, with counts for each type. A count of "x4" denotes four features of that type. Some of the features, such as the *Previous Query Local Hour Flag*, are represented using one-hot binary vectors, hence creating twenty-four binary flags for that single feature (one per hour). The features are as follows. The *Previous Device Flag* is a one-hot binary representation of which device was used for the previous query. The *Query Length* feature denotes

Table 5: Features used in predictive models.

Type	Features and Counts
Previous Device	Previous Device Flag (x4)
Query Length	Query Length
Hour	Previous Query Local Hour Flag (x24)
Topic	Previous Query Topic Flags (x15)
	Current Query Topic Flags (x15)
Global Stats	Global Device Probabilities (x4)
	Global Device-Device Transition Probabilities (x16)
	Global Cross-Device Transition Probabilities (x12)
Global Temporal Stats	Global Device Avg. Transition Delay (x4)
	Global Device-Device Avg. Transition Delay (x16)
User Stats	Number of Historical User Samples
	Number of Historical User Cross-Device Samples
	User Device Probabilities (x4)
	User Device-Device Transition Probabilities (x16)
	User Cross-Device Transition Probabilities (x12)
	User Cross-Device Destination Device Probabilities (x4)
	User IsDeviceDominant Flags (x4)
	User IsCrossDeviceDominant Flags (x4)
User Temporal Stats	User Device Avg. Transition Delay (x4)
	User Device-Device Avg. Transition Delay (x16)

the length of the previous query (in characters). The *Previous Query Local Hour Flag* records the hour when the previous query was issued. The *Previous (and Current) Query Topic Flags* are sets of fifteen binary flags showing which topics are assigned to the query, as measured by the topical classifiers discussed in Section 3.3. The *Global Device Probabilities* record how often queries were issued from each device type across all users. Similarly, the *Global Device-Device Transition Probabilities* give the likelihood of transitions between each device-type pair (such as PC-to-PC and Tablet-to-Smartphone transitions). The *Global Cross-Device Transition Probabilities* are similar, but limited to the probabilities for cross-device transitions (thus the normalization of the probabilities differs). The *Global Device Avg. Transition Delay* measures the average delay between queries conditioned on the previous device type. The *Global Device-Device Avg. Transition Delay* measures the average delay time between device-type pairs.

We also have several user specific features and variants of the above. The *Number of Historical User Samples* feature denotes how many transitions were previously seen for a given user, and the *Number of Historical User Cross-Device Samples* feature is similar, but restricted to cross-device transitions. The *User Device Probabilities*, *User Device-Device Transition Probabilities*, *User Cross-Device Transition Probabilities*, *User Device Avg. Transition Delay* and *User Device-Device Avg. Transition Delay* features are all similar to their global counterparts, but are instead computed on a per user basis. The *User Cross-Device Destination Device Probabilities* features measure the probability that a certain device will be the destination of a transition, regardless of the origin device, given that it differs from the destination device. Lastly, The *User IsDeviceDominant Flags* record which of the four device types is the historically dominate destination device, and the *User IsCrossDeviceDominant Flags* are analogous, but restricted to cross-device transitions. There is some feature overlap, with some features derivable from others. We use L1 regularization to prune unnecessary and redundant features.

All *Stats* features in Table 5 are computed from the non-overlapping historical data. We further group the features into sets to measure their effect on predictive accuracy. The baseline feature set consists of the previous device type only. The query length feature gives the number of characters in the previous query, and previous hour gives the hour of day when the previous query was issued. Topical features are grouped by previous and current query and the user transition features are partitioned into six groups, outlined in Table 6. The full list of groupings are given in the ablation results of Table 9 and are discussed in Section 4.2.

Table 6: User-specific historical feature sets.

Set	Description
U1	Number of user transitions seen in historical data (total transitions and cross-device transitions).
U2	Historical probabilities for each device type.
U3	Unconditioned device-device transition probabilities.
U4	Destination device probabilities, i.e., the probability a cross-device transition lands on a given device type.
U5	Conditioned cross-device transition probabilities. (similar to U3, but only for cross-device transitions and conditioned on previous device type).
U6	Average device-device transition delay times.

We next discuss our predictive models for the learning tasks, the experimental set-up and baselines used for performance comparisons.

4.1.3 Learning Algorithms

Using these datasets, we performed multi-class classification to forecast different aspects of cross-device search given the previous device and other contextual features related to the previous query and user history. We evaluated an L1-regularized Logistic Regression model [4] and a Gradient Boosting Tree ensemble [10] through five-fold cross-validation, using the Scikit-learn package for Python [27]. Since the primary objective of this study is to demonstrate the feasibility of developing accurate predictive models of cross-device searching (and not to develop new machine learning algorithms), we limited our evaluation to two representative classes of learning methods. We compared their performance against three baselines: (1) Most Frequent Label, which predicts the most common class label in the training dataset, (2) Uniform Guessing, which predicts class labels uniformly at random, and (3) Stratified Guessing, which predicts class labels randomly, while respecting the observed global class label proportions encountered during training.

4.2 Prediction Results

4.2.1 Predict Next Device

For the task of predicting the next device, both models show significant improvements over the baseline methods, with over 98% accuracy, and high average precision, recall and F1-score. The metrics shown represent weighted (by class proportion) averages of per class scores, uniformly av-

Table 7: (Predict Next Device) Main. Statistical significance is computed with reference to the Most Frequent Label baseline method.

Method	Accuracy	Avg. Precision	Avg. Recall	Avg. F1-Score	Log-Loss
Baseline Method - Most Frequent Label	0.648	0.420	0.648	0.510	12.157
Baseline Method - Stratified	0.489	0.489	0.489	0.489	0.883
Baseline Method - Uniform	0.250	0.488	0.250	0.309	1.386
Gradient Boosting Trees Classifier	0.982***	0.982***	0.982***	0.982***	0.097***
L1 Logistic Regression	0.983***	0.983***	0.983***	0.983***	0.089***

(Statistical significance assessed by two-tailed paired t-test, with: * $< \alpha = .05$, ** $< \alpha = .01$, *** $< \alpha = .001$)

eraged across folds[1]. Table 5 presents the results. Examining the feature contributions (not shown for this dataset), we found that conditioning on the previous device produces the accuracy observed (98.3%), so that the previous device state is the strongest feature for next device prediction, and extra features do not improve performance.

4.2.2 Predict Device Switch

Our next task was to predict whether or not a user switched devices between queries. This is a non-trivially important task in some scenarios, such as when a new query is issued without corresponding device-type information (as was the case for many records excluded from our dataset). To perform the prediction, we trained our classification models on the balanced transition dataset, where transition instances were equally split among the "same-device" and "different-device" class labels. In addition to the features given in Tables 5 and 6, we considered an additional feature for this task, the transition delay time, which is the elapsed time between two consecutive queries. We evaluated the effect of adding this feature to the existing historical features.

The classification results for predicting whether a device switch will occur are given in Table 8. Predictive accuracy is significantly increased over the baseline methods, indicating the existence of exploitable non-randomness in device transition behavior. Table 9 demonstrates the effect of different feature sets on classification accuracy, including the use of current query features. The greatest gains are seen when adding user-specific historical transition features (U1-U6) and the transition delay time feature (see Figure 7).

4.2.3 Predict Next Device Given Device Switch

Given a classifier capable of predicting whether or not a cross-device transition occurred, the next step is to develop a classifier that predicts the next device for cases when it differs from the previous device. In practice the classifiers could be chained, but we do not do that here since we wanted to limit interaction effects. The task of predicting the next device given it differs from the previous device is more difficult than predicting the next device in general, since a learner cannot simply predict the class label of the previous device (as is normally the case). Table 10 demonstrates the increased difficulty of this task, with all methods (except uniform random guessing) suffering decreased performance compared to the next-device prediction task from Table 7. However, the feature-based methods continue to significantly outperform the baselines. Surprisingly, the previous device still produced the strongest signal when predicting the future device, apparent in the ablation results of Table 11. This held even though the next device and previous device were guaranteed to differ. Other features, such as previous topics (PT) and global transition stats (G), had no effect on performance, despite proving useful when predicting the occurrence of a device switch. User transition statistics led to significant increases in performance in combination with the other features, but also on their own (as Baseline+U1-6). Thus, strong, exploitable patterns emerge for searchers when switching to new devices.

4.2.4 Summary

Overall, we observe significant improvements over baselines, with gains in predictive accuracy of over twenty-five percentage points for all three classification tasks evaluated. The previous device and searchers' own transition histories were the primary factors in the prediction. These results are important as they clearly demonstrate successful prediction of various aspects of cross-device search (the first time this has been accomplished), but also that high accuracy can be achieved with compact models comprising only a few key features. Compactness is important for large-scale deployment in search engines.

5. DISCUSSION AND IMPLICATIONS

We showed that there are variations in interests across the four different device types and that multi-device searchers were fairly common (5% of users) and generated a significant amount of search engine traffic (16%). Of particular interest were the transitions *between* devices (many of which are more or less immediate, as shown in Figure 14), which provides insight into the context within which devices are used. For example, smartphone/tablet and PC appear to frequently be used consecutively, and more work is needed to understand device interplay (e.g., are there tasks or scenarios for which a device switch should be recommended?) and how best to support it from a search perspective.

If the search system is expected to work proactively, the ability to predict the device that a searcher will use for their next query is important in determining what type of information to seek on the searcher's behalf. Different devices have different display, bandwidth, and interaction constraints, as well as differences in the type of information that people are interested in on each device (as demonstrated in Section 3.3). This information could be coupled with data about cross-session search tasks and searchers' long-term interests [1, 19] to model their interests and intentions. An additional component that would be welcome in this model is the anticipated time until next query (i.e., the transition delay), which could help guide system assistance decisions.

[1]This weighting method, combined with the one-vs-all learner implementation and class label imbalance, tends to produce precision and recall scores near or equal to accuracy.

Table 8: (Predict Device Switch) Balanced Transition Data. Statistical significance is computed with reference to the Most Frequent Label baseline method.

Method	Accuracy	Avg. Precision	Avg. Recall	Avg. F1-Score	Log-Loss
Baseline Method - Most Frequent Label	0.499	0.249	0.499	0.332	17.320
Baseline Method - Stratified	0.498	0.498	0.498	0.498	0.693
Baseline Method - Uniform	0.501	0.501	0.501	0.501	1.386
Gradient Boosting Trees Classifier	0.787***	0.788***	0.787***	0.786***	0.462***
L1 Logistic Regression	0.779***	0.782***	0.779***	0.778***	0.484***

(Statistical significance assessed by two-tailed paired t-test, with: * $< \alpha = .05$, ** $< \alpha = .01$, *** $< \alpha = .001$)

Table 9: (Predict Device Switch - Feature Ablations, Balanced Transition Data) L1 Logistic Regression. Statistical significance is computed with reference to the Baseline feature group.

Feature Group	Accuracy	Avg. Precision	Avg. Recall	Avg. F1-Score	Log-Loss
Baseline	0.589	0.591	0.589	0.587	0.676
Baseline+Q	0.589	0.591	0.589	0.587	0.674***
Baseline+Q+PT	0.599***	0.602***	0.599***	0.596***	0.665***
Baseline+Q+PT+PH	0.608***	0.609***	0.608***	0.607***	0.661***
Baseline+Q+PT+PH+U1	0.660***	0.661***	0.660***	0.660***	0.615***
Baseline+Q+PT+PH+U1-2	0.665***	0.666***	0.665***	0.664***	0.610***
Baseline+Q+PT+PH+U1-3	0.665***	0.666***	0.665***	0.665***	0.608***
Baseline+Q+PT+PH+U1-4	0.668***	0.668***	0.668***	0.668***	0.607***
Baseline+Q+PT+PH+U1-5	0.668***	0.669***	0.668***	0.668***	0.606***
Baseline+Q+PT+PH+U1-6	0.673***	0.674***	0.673***	0.673***	0.601***
Baseline+Q+PT+PH+U1-6+G	0.674***	0.674***	0.674***	0.673***	0.602***
Baseline+Q+PT+PH+U1-6+G+CT	0.675***	0.675***	0.675***	0.675***	0.599***
Baseline+Q+PT+PH+U1-6+G+DT	0.778***	0.782***	0.778***	0.778***	0.485***
All Features	0.779***	0.782***	0.779***	0.778***	0.484***

Baseline = Previous Device
Q = Previous Query Length
G = Global Transition Stats Features
PT = Previous Topic features
CT = Current Topic Features
PH = Previous Query Hour
U1 = Number of historical samples for the user and the number of cross-device samples for the user
U1-2 = U1 and user-specific device probabilities
U1-3 = U1, U2 and device-device pair transition probabilities
U1-4 = U1, U2, U3 and destination device transition probabilities
U1-5 = U1 through U4 and device conditioned transition probabilities
U1-6 = U1 through U5 and user-specific average transition times for device-device pairs
DT = Delay time (in seconds) between previous and current queries
All Features = All of the above feature sets

(Statistical significance assessed by two-tailed paired t-test, with: * $< \alpha = .05$, ** $< \alpha = .01$, *** $< \alpha = .001$)

Table 10: (Predict Next Device Given Device Switch) Min. 200 Historical Samples. Statistical significance is computed with reference to the Most Frequent Label baseline method.

Method	Accuracy	Avg. Precision	Avg. Recall	Avg. F1-Score	Log-Loss
Baseline Method - Most Frequent Label	0.455	0.207	0.455	0.284	18.829
Baseline Method - Stratified	0.370	0.371	0.370	0.371	1.046
Baseline Method - Uniform	0.250	0.370	0.250	0.292	1.386
Gradient Boosting Trees Classifier	0.931***	0.931***	0.931***	0.931***	0.197***
L1 Logistic Regression	0.934***	0.933***	0.934***	0.933***	0.193***

(Statistical significance assessed by two-tailed paired t-test, with: * $< \alpha = .05$, ** $< \alpha = .01$, *** $< \alpha = .001$)

Table 11: (Predict Next Device Given Device Switch - Feature Ablations) L1 Logistic Regression. Statistical significance is computed with reference to the Baseline feature group.

Feature Group	Accuracy	Avg. Precision	Avg. Recall	Avg. F1-Score	Log-Loss
Baseline+U1-6	0.932***	0.931***	0.932***	0.931***	0.196***
Baseline	0.781	0.642	0.781	0.703	0.496
Baseline+Q	0.781	0.642	0.781	0.703	0.496
Baseline+Q+PT	0.781	0.642	0.781	0.703	0.495
Baseline+Q+PT+PH	0.781	0.679	0.781	0.703	0.493
Baseline+Q+PT+PH+U1	0.781	0.716	0.781	0.703	0.491
Baseline+Q+PT+PH+U1-2	0.903***	0.903***	0.903***	0.898***	0.281***
Baseline+Q+PT+PH+U1-3	0.928***	0.927***	0.928***	0.927***	0.203***
Baseline+Q+PT+PH+U1-4	0.932***	0.932***	0.932***	0.931***	0.195***
Baseline+Q+PT+PH+U1-5	0.932***	0.932***	0.932***	0.931***	0.195***
Baseline+Q+PT+PH+U1-6	0.933***	0.932***	0.933***	0.932***	0.194***
Baseline+Q+PT+PH+U1-6+G	0.933***	0.932***	0.933***	0.932***	0.194***
Baseline+Q+PT+PH+U1-6+G+CT	0.933***	0.932***	0.933***	0.932***	0.194***
Baseline+Q+PT+PH+U1-6+G+DT	0.933***	0.933***	0.933***	0.932***	0.194***
All Features	0.934***	0.933***	0.934***	0.933***	0.193***
For Feature Group legend, see Table 9					

(Statistical significance assessed by two-tailed paired t-test, with: * $< \alpha = .05$, ** $< \alpha = .01$, *** $< \alpha = .001$)

With that goal in mind, we attempted to predict the time until the next query. Our efforts at this regression task were met with limited success, even when using predictions from the classification models and device-device transition time statistics as features. Neither of the regression models tested (Lasso Regression and Gradient Boosting Trees Regression) significantly improved performance over the baselines of guessing the mean and median delay times. Transforming the task into a three-way classification task (classes = {delay less than 30 seconds, delay between 30 secs. and 10 mins., delay greater than 10 mins.}) was more promising, with significant improvements of approximately ten percentage points over the best baseline method, but accuracy was still below 50% (results not shown).

There are some limitations that we should acknowledge. First, since the study is log-based, we do not have insight about searchers' rationales for moving between devices or the context of the transitions, which is important for, say, understanding why some device pairings are more tightly coupled than others. Follow-up qualitative studies in the search context are needed to understand: (1) how people transition between devices (both physically and practically), and (2) criteria that they use to select a particular device if multiple devices are accessible in a particular location (e.g., searcher is at home with access to all of their devices). Second, our prediction tasks focused on query-query transitions, however there are other scenarios that should be explored, e.g., predicting the device used for the next search *session*. Finally, the dataset used for our analysis is proprietary and cannot be shared publicly given privacy considerations. Researchers seeking to perform similar studies should consider the need to represent users, their devices, and their inter-device transitions when designing logging mechanisms to collect search data.

We considered three important prediction tasks where we see significant improvements over the three baselines: (1) predict next device, (2) predict if a device switch occurred, and (3) predict next device given that a device switch occurred. The predict-next-device classifier (#1) was highly accurate, although primarily because people frequently stay on the same device. The device prediction task (#3) relies on foreknowledge of a device switch, which could be provided by a device switch classifier (#2). We assessed #2 and #3 separately, and future work may include chaining them together.

6. CONCLUSIONS

We introduced cross-device search and drew several data-supported conclusions about search across devices. We show that people use different devices to search for different content, and time of day interacts with device to affect content sought. Exploitable patterns emerge for device transitions (especially from historic user features), with previous device signaling the next device, *even when the devices differ*.

We analyzed device-specific and temporal aspects of cross-device search, and were able to successfully predict the next device from which a searcher will query, even on datasets limited to cross-device transitions. For two of the prediction tasks (i.e., predict next device and predict next device given switch), predictive accuracy, recall and precision exceeded 90% with a fairly compact feature set. The remaining task (predict device switch) also saw significant gains in accuracy over the baselines, approaching 80% given all features. We will continue work in this important emerging area, focused on developing more accurate predictive models, including further investigating delay time prediction. We will also integrate these models into search systems, enabling capabilities such as proactively locating device-appropriate information in advance of anticipated device transitions.

7. REFERENCES

[1] E. Agichtein, R. W. White, S. T. Dumais, and P. N. Bennet. Search, interrupted: understanding and predicting search task continuation. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '12, pages 315–324, New York, NY, USA, 2012. ACM.

[2] A. Aula, N. Jhaveri, and M. Käki. Information search and re-access strategies of experienced web users. In

Proceedings of the 14th international conference on World Wide Web, WWW '05, pages 583–592, New York, NY, USA, 2005. ACM.

[3] R. Baeza-Yates, G. Dupret, and J. Velasco. A Study of Mobile Search Queries in Japan. In *World Wide Web Conference Series*, 2007.

[4] C. M. Bishop. *Pattern recognition and machine learning*. Springer, 2007.

[5] K. Church and B. Smyth. Understanding the intent behind mobile information needs. In *Proceedings of the 14th international conference on Intelligent user interfaces*, IUI '09, pages 247–256, New York, NY, USA, 2009. ACM.

[6] ComScore. Digital omnivores: how tablets, smartphones and connected devices are changing u.s. digital media consumption habits. http://ipmark.com/pdf/Omnivoros.pdf. Accessed: 2013-08-19.

[7] D. Dearman and J. S. Pierce. It's on my other computer!: computing with multiple devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 767–776, New York, NY, USA, 2008. ACM.

[8] D. Donato, F. Bonchi, T. Chi, and Y. Maarek. Do you want to take notes?: identifying research missions in yahoo! search pad. In *Proceedings of the 19th international conference on World wide web*, WWW '10, pages 321–330, New York, NY, USA, 2010. ACM.

[9] S. Eggermont and J. Van den Bulck. Nodding off or switching off? The use of popular media as a sleep aid in secondary-school children. *Journal of Paediatrics and Child Health*, 42(7-8):428–433, 2006.

[10] J. H. Friedman. Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, pages 1189–1232, 2001.

[11] M. Halvey, M. Keane, and B. Smyth. Time based patterns in mobile-internet surfing. Proc. CHI, pages 31–34, 2006.

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

[13] M. Kamvar and S. Baluja. A large scale study of wireless search behavior: Google mobile search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pages 701–709, New York, NY, USA, 2006. ACM.

[14] M. Kamvar, M. Kellar, R. Patel, and Y. Xu. Computers and iphones and mobile phones, oh my!: a logs-based comparison of search users on different devices. In *Proceedings of the 18th international conference on World wide web*, WWW '09, pages 801–810, New York, NY, USA, 2009. ACM.

[15] S. K. Kane, A. K. Karlson, B. R. Meyers, P. Johns, A. Jacobs, and G. Smith. *Exploring Cross-Device Web Use on PCs and Mobile Devices*. 2009.

[16] A. K. Karlson, S. T. Iqbal, B. Meyers, G. Ramos, K. Lee, and J. C. Tang. Mobile taskflow in context: a screenshot study of smartphone usage. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 2009–2018, New York, NY, USA, 2010. ACM.

[17] A. K. Karlson, B. R. Meyers, A. Jacobs, P. Johns, and S. K. Kane. *Working Overtime: Patterns of Smartphone and PC Usage in the Day of an Information Worker*. 2009.

[18] M. Kellar, C. R. Watters, and M. A. Shepherd. A field study characterizing Web-based information-seeking tasks. *Journal of The American Society for Information Science and Technology*, 58:999–1018, 2007.

[19] 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 the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, SIGIR '11, pages 5–14, New York, NY, USA, 2011. ACM.

[20] J. Krumm and D. Rouhana. Placer: semantic place labels from diary data. Proc. UbiComp, pages 163–172, 2013.

[21] J. Li, S. Huffman, and A. Tokuda. Good abandonment in mobile and pc internet search. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '09, pages 43–50, New York, NY, USA, 2009. ACM.

[22] J. Liu and N. J. Belkin. Personalizing information retrieval for multi-session tasks: the roles of task stage and task type. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '10, pages 26–33, New York, NY, USA, 2010. ACM.

[23] B. Ma Kay and C. Watters. Exploring multi-session web tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 1187–1196, New York, NY, USA, 2008. ACM.

[24] D. Morris, M. Ringel Morris, and G. Venolia. Searchbar: a search-centric web history for task resumption and information re-finding. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 1207–1216, New York, NY, USA, 2008. ACM.

[25] H. Müller, J. Gove, and J. Webb. Understanding tablet use: a multi-method exploration. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*, MobileHCI '12, pages 1–10, New York, NY, USA, 2012. ACM.

[26] S. Nylander, T. Lundquist, and A. Brännström. At home and with computer access: why and where people use cell phones to access the internet. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 1639–1642, New York, NY, USA, 2009. ACM.

[27] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[28] A. Schein, A. Popescul, L. Ungar, D. Pennock, and D. Ungar. Methods and metrics for cold-start recommendations. Proc. SIGIR, pages 253–260, 2002.

[29] Y. Song, H. Ma, H. Wang, and K. Wang. Exploring and exploiting user search behavior on mobile and tablet devices to improve search relevance. In *Proceedings of the 22nd international conference on World Wide Web*, WWW '13, pages 1201–1212, Republic and Canton of Geneva, Switzerland, 2013. International World Wide Web Conferences Steering Committee.

[30] A. Spink, B. J. Jansen, D. Wolfram, and T. Saracevic. From E-Sex to E-Commerce: Web Search Changes. *IEEE Computer*, 35:107–109, 2002.

[31] C. A. Taylor, O. Anicello, S. Somohano, N. Samuels, L. Whitaker, and J. A. Ramey. A framework for understanding mobile internet motivations and behaviors. In *CHI '08 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '08, pages 2679–2684, New York, NY, USA, 2008. ACM.

[32] J. Teevan, K. Collins-Thompson, R. White, S. Dumais, and Y. Kim. Slow search: information retrieval without time constraints. HCIR '13, 2013.

[33] J. Teevan, A. Karlson, S. Amini, A. J. B. Brush, and J. Krumm. Understanding the importance of location, time, and people in mobile local search behavior. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, MobileHCI '11, pages 77–80, New York, NY, USA, 2011. ACM.

[34] Y. Wang, X. Huang, and R. W. White. Characterizing and supporting cross-device search tasks. In *Proceedings of the sixth ACM international conference on Web search and data mining*, WSDM '13, pages 707–716, New York, NY, USA, 2013. ACM.

[35] R. W. White and S. M. Drucker. Investigating behavioral variability in web search. In *Proceedings of the 16th international conference on World Wide Web*, WWW '07, pages 21–30, New York, NY, USA, 2007. ACM.

[36] J. Yi, F. Maghoul, and J. Pedersen. Deciphering mobile search patterns: a study of yahoo! mobile search queries. In *Proceedings of the 17th international conference on World Wide Web*, WWW '08, pages 257–266, New York, NY, USA, 2008. ACM.