Pursuing Insights about Healthcare Utilization via Geocoded Search Queries

Shuang-Hong Yang*
Twitter Inc.
San Francisco, CA
syang@twitter.com

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

Eric Horvitz
Microsoft Research
Redmond, WA
horvitz@microsoft.com

ABSTRACT

Mobile devices provide people with a conduit to the rich information resources of the Web. With consent, the devices can also provide streams of information about search activity and location that can be used in population studies and real-time assistance. We analyzed geotagged mobile queries in a privacy-sensitive study of potential transitions from health information search to in-world healthcare utilization. We note differences in people's health information seeking before, during, and after the appearance of evidence that a medical facility has been visited. We find that we can accurately estimate statistics about such potential user engagement with healthcare providers. The findings highlight the promise of using geocoded search for sensing and predicting activities in the world.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – *search process, information filtering.*

Keywords

Mobile search; healthcare utilization; query logs.

1. INTRODUCTION AND BACKGROUND

Mobile devices allow people to search the Web as they travel and engage in the physical world. The devices can serve as large-scale sensor networks for population-wide studies of behavior and the environment. Mobile devices that sense location have the ability to provide information on connections between online activities and activities in physical world. In recent work, researchers demonstrated how a log of geocoded mobile queries could be harnessed in a privacy-sensitive manner to make inferences about how health search evolves into transitions in location that appear to represent healthcare utilization (HU) [9]. Other sources such as queries [2], blogs [5], and social media [6] have been used as spatiotemporal sensors, but none offer the clear relationship between virtual and physical spaces as do geocoded mobile queries. We show how such logs to extend prior studies and can provide new insights about healthcare utilization from mobile logs.

Given estimates of HU events derived from mobile queries issued to a Web search engine, we identify and monitor the following:

Search intent dynamics: Previous research has examined patterns of behavior with people's health information seeking over time [8]. We show that mobile query logs can provide insights on search interests and concerns before patients arrive at a healthcare facility, as they wait or receive treatment, and after they depart.

Population-level utilization analyses: Public organizations such as the Centers for Disease Control and Prevention (CDC) compute statistics on healthcare utilization. We explore the possibility of using mobile query logs to complement traditional tracking and monitoring as a low cost means of estimating the statistics of HU.

Patient-specific predictions: We show that signals from mobile search logs can predict the duration of a forthcoming stay at medical facilities, enabling such applications as patient triage.

We first describe the procedure by which we estimate HU events from geocoded mobile query logs. We then present several different uses of these inferred visits. We characterize behavior before, during, and after an inferred visit, estimating visit statistics, and predicting the duration of a user's stay at a medical facility.

2. HEALTHCARE UTILIZATION & LOGS

A key premise of our study is that we can identify when users are likely visiting healthcare facilities and/or engaging with healthcare professionals. We begin by describing the datasets we use in our study and the methods we used to infer utilization.

2.1 Data

We collected datasets with consent via a widely available application provided by a major mobile search provider. The application runs on the Android and iPhone platforms. Users agreed to share their search activities and location information per a published privacy statement, which grants uses of the data to the provider for internal analysis and service improvement. We used automated tools to process the data in the aggregate. In this process, all data were stripped of individually identifiable information. Direct location information was removed and not used in the analysis.

We used a dataset spanning one year of activity from approximately 30K users who had agreed to share their search activities and location information. The log consists of several hundred thousand search entries stored on a secure server. Each entry corresponds to a search interaction with a mobile application issued by an anonymized user, represented by a unique numerical identifier. Entries contain the search query, the time, and GPS location when the query was submitted. Location information was recorded only if and when a search query was issued. The absolute locations of users were removed. As in [9], only the *distance* between the user and the location of medical facilities was stored. To compute the distances, we obtained the GPS coordinates of medical facilities across the United States through a crawl of business listings, which includes data from a variety of sources such as the Yellow Pages (yp.com). The database of facilities contains 34,750 medical facility sites. Site types include emergency medical and surgical services (54.7%), hospitals (34.1%), medical centers (10.5%), and outpatient services (0.6%). As we sought to understand long-term search patterns, we removed users with < 90 days of search history.

* Research done during an internship at Microsoft Research.

SIGIR'13, July 28–August 1, 2013, Dublin, Ireland.

2.2 Evidence of Healthcare Utilization

To identify searchers who may be utilizing medical facilities, we consider the spatial context in which a medical query was issued. We only consider whether people searched for medical information from within a certain distance (200 meters in our case) from a medical facility, following methods used and evaluated in [9]. In particular, we consider non-periodic proximity to a medical facility as salient *evidence of healthcare utilization* (EHU). We further characterize engagements with healthcare facilities with information about whether the GPS signal was available; GPS is generally available outside but is often lost when indoors. We assume that searches may occur within a healthcare facility when a GPS signal is not attainable for a period of time (e.g., from 30 minutes to a few days) following observation of EHU. Types of engagements at a medical facility include seeking treatment, visiting an admitted patient, and accompanying a patient. For simplicity we shall refer to all of these activities as *healthcare utilization*. In our data set, 5239 users (14.8% of the total number) queried from within 200 meters of the medical facility at least once, providing evidence of HU. Note that although the proximal query provides a useful way of estimating HU, this mechanism may introduce population bias (e.g., these users may be younger or more tech savvy).

2.3 Error Reduction

Given the nature of the logs, we were unable to distinguish between users visiting the medical facility for treatment, to visit or escort an admitted patient, or simply working at or traveling near the facility but not using its health-related services. We seek to make inferences about HU from the log, which serves as a noisy and sparse sensor; we cannot completely resolve uncertainty about whether the EHUs as defined actually represent a visit or a true engagement. To reduce the likelihood of errors, we require that users had previously searched for at least one of a set of tracked symptoms or synonyms from the Merck medical dictionary in the 90 days before the EHU visit, demonstrating a health interest or concern. In total, 4006 of the 5239 users (76.5%) met this requirement. To improve the likelihood that observed EHU events were related to treatment and consultation, we removed the 885 users (22.1%) who had queried from within 200 meters of the facility in the time period before the first symptom query and who had queried frequently near a facility (where more than 30 of the 90 days of search activities occurred within 200m of a facility). These users may live near the facility, may work at or near the facility or be receiving long-term care at the facility, and would hence not be relevant to our focus on transitions from Web search to in-world healthcare utilization.

3. ANALYZING UTILIZATION

Given an estimate for HU from mobile search logs, we analyze aspects of this activity, specifically studying the dynamics of search intent surrounding the EHU visit. We now characterize intentions and seek to predict the duration for each inferred visit.

3.1 Search Intent Dynamics

We first explore the dynamics of health-related search goals that may be attributed to inferred HU. Specifically, as a user's health condition evolves (e.g., from normal, to ill (or concerned about an illness), to recovered), or engagement status changes (e.g., from approaching a facility, to engaging with a healthcare practitioner, to departing a facility), how is their search behavior affected? We target the presence of the following health signals in mobile queries:

- **Symptoms.** Symptom queries can capture the onset of medical concerns. We label a query as symptom-related if it contains at least one symptom term (e.g., "twitching", "headache"). The list of possible symptoms is extracted from the online Merck medical dictionary. To improve coverage, we also include the synonyms of each symptom, which are identified via a two-step random walk on the click-graph of a search engine [3].
- **Body-part names.** We use a list of major appendages and organs as body-part names obtained from reviews of web sites on human anatomy [4], and track them in queries. Body parts can be related to symptoms (e.g., back pain), treatments (e.g., kidney transplant), or outcomes (e.g., back pain after surgery).
- **Benign and serious conditions.** We also look for the presence of benign or serious medical conditions in queries, which are extracted based on the list of conditions in [4] and their synonyms. Condition-related queries can indicate hypothesis-directed search such as pursuing a diagnosis given symptoms or seeking additional information about certain conditions.
- **Healthcare utilization intentions (HUI).** We also track queries that demonstrate pursuit of in-world healthcare, such as searches for a clinic or hospital. Such HUI searches may arise as information seeking about symptoms or diagnostic hypotheses escalates to the pursuit of in-world resources [7].
- **Drugs.** We define a set of common medications and monitor the presence of them in search queries. Medications are directly related to treatment and remedy of symptoms and disorders.

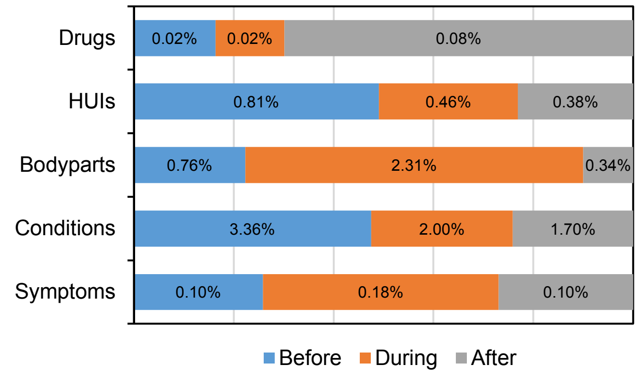

Figure 1. Search intentions before, during, and after assumed engagement with professional healthcare (data presented as normalized stacked bars with percentages as labels).

Queries are classified into one of these categories via a white-list based string-match classifier. When a query has overlap with multiple categories, we assign it to the category with highest overlap.

For each searcher demonstrating EHU, we divide search sequences into three phases, i.e., before, during (i.e., user queries near a medical facility and the GPS signal is unavailable or "dark," suggesting that they are indoors), and after the assumed HU. We track the dynamics of the searcher's intent by computing the fraction of their search queries on the aforementioned five categories in each phase. The average dynamics over all users with EHU are depicted in Figure 1. The figure shows the fraction of all observed queries that fall into each phase before, during, and after the visit. For example, 0.1% of queries observed before and after the EHU contain symptoms, but 0.18% of the queries during the visit are associated with symptoms. There are a number of noteworthy patterns in the terminological shifts between the three phases:

1. More searching symptoms and human anatomy when users are close to a medical facility (within 200m and GPS dark) than far away from it, suggesting that searchers have stronger interest and concerns about health when searching near a hospital.
2. There is a stronger interest in HUIs prior to the visit to the facility and this diminishes dramatically once the user reaches the facility (dropping to nearly half the pre-visit level). Since the EHU is

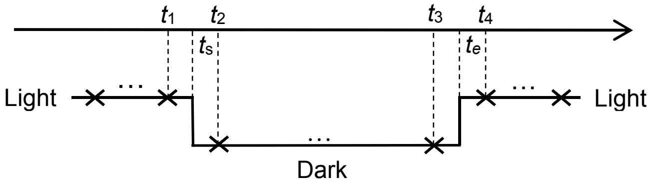

Figure 2. Illustration of start (t_s) and end time (t_e) of an engagement, with measurement error based on queries (×).

based on the GPS location and not the query content, this transition provides strong evidence of HU.

3. We observe changes in the mix of health-related search goals as people transition to different phases of engagement. In particular, users were observed searching significantly more often for conditions (doubled) *before* the assumed engagement than in the other phases, anatomical names (tripled) and symptoms (doubled) *during* engagement than other phases, and drug names (seven times higher) *following* the visit.
4. We believe that the transition from conditions and HUI (before engagement), to symptoms and body-parts (during engagement), and eventually to treatments and drug names (after engagement) reflects shifts in information needs over the course of healthcare engagement episodes. The mobile query logs appear to reflect patients moving from concerns and questions about their conditions, to those of diagnosis and symptoms, and eventually to treatments and post-visit recovery.

Although these findings may be unsurprising, they show that sensing is to some extent reliable and highlight the potential value of geocoded query analysis for linking online intent with in-world use.

We also examined how healthcare engagement influences searchers' medical concerns. As an example, we sought to understand how evidence of engagement with a healthcare professional might lead to signs of resolution of concerns. One way to do this is by examining if and how the frequency of question words (i.e., what, which, when, why, how, and who) in medical queries changes with the phase of the engagement. One might expect to see that users have more uncertainty and questions before, and perhaps during, the early portion of healthcare engagement than after it. We did not observe significant changes in the use of question words in mobile medical queries over the course of the engagement. One explanation is that while the engagement may address some concerns, it also raises others. We observed that queries issued during engagement were on average more technical than in other scenarios, suggesting that users may adopt terminology from discussions with healthcare professionals, or seek the meaning of specialist terms encountered. We shall investigate query technicality in more detail.

3.2 Estimating Visit Durations

We now consider estimations of durations of engagements at medical facilities. The primary focus in this analysis is in comparing the statistics obtained from the mobile search log to ground truth provided by the CDC. The CDC statistics are based on outpatient records, collected from a range of medical facilities across the US.

Since we only have access to location information when users enter queries, we are uncertain about exact time a user enters or leaves a facility. Figure 2 shows an example of a visit to a medical facility, defined in the figure as time period where the user's GPS signal is dark (i.e., they are likely indoors). Assume the last query before engagement that the searcher issued when the GPS signal is available (light) was at time t_1, the first query when their GPS is unavailable (dark) was at time t_2, the last query issued at dark was at time t_3 and the first query after engagement issued when their GPS first received a signal was at t_4. Our estimate of start-time $t_s = (t_1 + t_2)/2$, estimated end-time $t_e = (t_3 + t_4)/2$, and estimated duration $\Delta t = t_e - t_s$. The measurement error $e = (t_2\text{-}t_1)/2 + (t_4\text{-}t_3)/2$. The relative measurement error $E = e/\Delta t$. To mitigate the measurement error, we perform explicit filtering by removing all EHU events whose relative measurement errors exceed 20%.

Table 1. Estimated duration of engagements with healthcare facilities via mobile logs vs. CDC statistics. Log estimates are mean average durations (± standard errors).

	CDC statistics	Mobile search estimation
Urgent care	3.70 hours	4.26 hours (±1.21 hours)
Hospital	4.90 days	4.29 days (±0.65 days)

Table 2. Features used in predicting duration of engagement.

HUI	**Prior Indoor Engagement**
Query has HUI	*Has previous indoor event*
# HUI	*# previous indoor events*
Time between HUI	*Time since last indoor event*
Has HUI refinement	*Has previous indoor-EHU event*
Medical Search	*# previous indoor-EHU events*
# symptom searches	*Time since last indoor-EHU event*
# serious condition searches	**Prior Engagement Dynamics**
# benign condition searches	*Avg duration of previous indoor events*
# unique symptoms	*Avg duration of previous indoor-EHU events*
# unique serious conditions	
# unique benign conditions	**Others**
Prior Utilization	*Facility type*
Has previous EHU event	*Time since first EHU event*
# previous EHU events	*Query technicality*
Time since last EHU event	*Frequency of user's EHU events*

The duration estimates that we obtain from our mobile search log analysis are provided in Table 1. Since the time that people spend at a facility is influenced by the type of facility (e.g., people may tend to spend more time in inpatient facilities than outpatient centers), we split our analysis into urgent care facilities and hospitals. For reference purposes, the CDC statistics on durations at each of the two facility types are also provided [1]. We can clearly see from Table 1 that estimates of durations of engagement times from the mobile search logs resonate well with the CDC statistics.

3.3 Forecasting Visit Durations

Given the accuracy with which we were able to estimate durations, we explored the use of predictive models to estimate the duration of an upcoming healthcare engagement based on a searcher's pre-engagement search behavior. These predictive models might be used one day for assistance in clinically-relevant forecasting tasks, such as helping medical facilities triage patients before their arrival. The features we used in our predictions are listed in Table 2. We breakout EHU into any EHU (query proximal to a medical facility) and *indoor* EHU (query proximal to medical facility and GPS signal has gone dark, suggesting that they are inside the facility at query time). Note that prior utilization or engagement features are computed using search activity > 12 hours before each engagement to reduce the likelihood of overlap with the engagement itself.

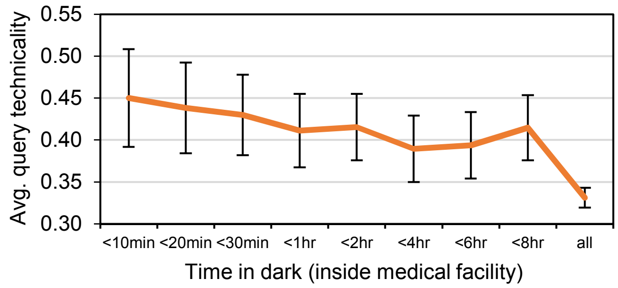

Figure 3. Query technicality over time during healthcare engagement. Error bars denote standard error (n=179).

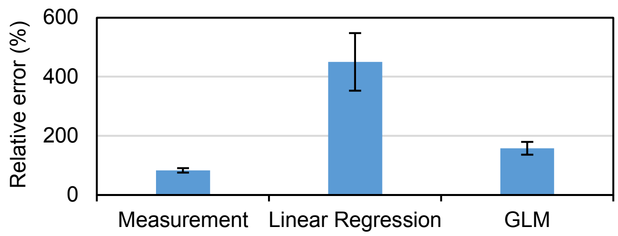

Figure 4. Prediction errors with linear regression and generalized linear model regression (± standard error).

One important feature that we mentioned earlier is query technicality, which captures the sophistication of the medical terminology used by a searcher. We implemented a query technicality estimator based on a statistical language model similar to the approach described in [10]. The technicality model was learned from a large corpus comprising medical documents from domains of different technicality (e.g., PubMed, MeSH, CDC, WebMD, and medical search queries issued on Bing.com). Technicality estimations for queries are not accepted if the confidence is low. The technicality estimator enables us to track the specificity of a searcher's medical concerns over the course of HU. Figure 3 plots the temporal dynamics of a searcher's query technicality during an engagement. As a reference, the normal level of query technicality for each user (independent of the visit) is also given as "all" in Figure 3. We can see the visit to the facility influences query technicality, suggesting that there may be exposure to technical terminology at the facility that emerges in query streams. Technicality also trends slightly downward over time, suggesting that most of the effect on query behavior occurs within a short time of the visit beginning. One explanation is diagnoses may be rendered soon after arrival of the patient.

A key challenge in our prediction is a lack of supervised data. We do not have access to the ground truth (e.g., the exact duration) of the potential engagements. For example, assuming a healthcare engagement for which the log-based estimation of duration is four hours with a measurement error of three hours, the true duration could be any value between one and seven hours. We know that we are grappling with substantial and unavoidable measurement errors. We cannot overcome the errors with our methodology. To make our prediction tractable, we removed engagement sessions with excessive measurement errors. Particularly, we require that for each engagement, the start distance and end distance, i.e., the distances at which the searcher issued a query when the engagement started/ended, are at most 150m and one mile respectively from the facility. This restriction results in 179 EHU sessions, with an average measurement error (defined in previous section) of 83.86%.

The duration prediction task can be formulated naturally as a regression problem, i.e., given the features extracted from the pre-engagement search activities, we seek to predict the duration of an engagement. We experimented with two regression models with the 25 features (Table 2). We used a linear regression model with L2 regularization and a generalized linear model (GLM) with a reciprocal link function (i.e., assuming that the duration follows an exponential distribution rather than normal) and L2 regularization. The models were trained and tested based on a leave-one-out cross-validation, where the relative error compared to the estimated duration was used as the evaluation metric. The prediction results are reported in Figure 4. Promisingly, they show that the GLM attains a prediction accuracy close to the measurement error. To better understand the engagement duration, Table 3 lists the top 10 most important features and their weights (normalized against the most important feature). The analysis shows that query technicality and previous EHUs are the strongest predictors, but other features such as facility type and general medical searching are also important.

Table 3. Top 10 features for duration prediction.

Feature Name	Weight
Query technicality	1.00
# previous EHU events	0.98
Time since last indoor-EHU event	0.48
Time since last EHU event	0.41
Avg duration of previous indoor-EHU events	0.40
Has previous EHU event	0.40
Facility type	0.39
# unique benign conditions (>12hr)	0.36
Avg duration of previous indoor event	0.32
Time since first EHU event	0.15

4. CONCLUSIONS

Using geotagged mobile searches as sensors for healthcare utilization, we explore health seeking behavior and seek to identify engagements with professional health providers. We also use search activity to estimate durations of healthcare utilization for different facility types and find strong agreement with CDC statistics. We use mobile search logs to predict engagement durations, considering medical search histories and prior engagement information. We are enthusiastic about the promise of the methods described in this paper, and envision privacy-sensitive applications that could provide value in the healthcare domain and beyond.

REFERENCES

[1] FastStats from CDC: http://www.cdc.gov/nchs/fastats/
[2] Backstrom L et al. (2008). Spatial variation in search engine queries. *WWW*, 357–366.
[3] Beeferman D, Berger A. (2000). Agglomerative clustering of a search engine query log. *KDD*, 407–416.
[4] Cartright M, White RW, Horvitz E. (2011). Intentions and attention in exploratory health search. *SIGIR*, 65–74.
[5] Mei Q et al. (2006). A probabilistic approach to spatiotemporal theme pattern mining on weblogs. *WWW*, 533–542.
[6] Sakaki T, Okazaki M, Matsuo Y. (2010). Earthquake shakes twitter users: Real-time event detection by social sensors. *WWW*, 851–860.
[7] White RW, Horvitz E. (2010). Web to world: Predicting transitions from self-diagnosis to the pursuit of local medical assistance in web search. *AMIA*, 882–886.
[8] White RW, Horvitz E. (2012). Studies of the onset and persistence of medical concerns in search logs. *SIGIR*, 265–274.
[9] White RW, Horvitz E. (2013). From Web search to healthcare utilization: Privacy-sensitive studies from mobile data. *JAMIA*, 20(1): 61–68.
[10] Yang SH, Crain S, Zha H. (2011). Bridging the language gap: topic adaptation for documents with different technicality. *AI Stats*, 823–831.