Location-aware Click Prediction in Mobile Local Search

Dimitrios Lymberopoulos[‡], Peixiang Zhao[†], Christian König[‡], Klaus Berberich[§], Jie Liu[‡]
[†] UIUC, Urbana-Champaign, IL, USA
[‡] Microsoft Research, Redmond, WA, USA
[§] MPI, Saarbrücken, Germany
pzhao4@illinois.edu, {dlymper,chrisko,liuj}@microsoft.com, kberberi@mpi-inf.mpg.de

ABSTRACT

Users increasingly rely on their mobile devices to search, locate and discover places and activities around them while on the go. Their decision process is driven by the information displayed on their devices and their current context (e.g. traffic, driving or walking etc.). Even though recent research efforts have already examined and demonstrated how different context parameters such as weather, time and personal preferences affect the way mobile users click on local businesses, little has been done to study how the location of the user affects the click behavior. In this paper we follow a data-driven methodology where we analyze approximately 2 million local search queries submitted by users across the US, to visualize and quantify how differently mobile users click across locations. Based on the data analysis, we propose new location-aware features for improving local search click prediction and quantify their performance on real user query traces. Motivated by the results, we implement and evaluate a data-driven technique where local search models at different levels of location granularity (e.g. city, state, and country levels) are combined together at run-time to further improve click prediction accuracy. By applying the location-aware features and the multiple models at different levels of location granularity on real user query streams from a major, commercially available search engine, we achieve anywhere from 5% to 47% higher Precision than a single click prediction model across the US can achieve.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

General Terms

Measurement, Human Factors

Keywords

mobile local search, search log analysis, feature extraction

1. INTRODUCTION

The wide availability of internet access on mobile devices, such as phones and personal media players, has allowed users to search, locate and access web information while on the go. Currently, there are 54.5 million mobile internet users and market analysis shows that this number will increase to 95 millions by 2013 [16]. This rapid growth indicates that it is only a matter of time when mobile devices will overtake desktop as the web/search access point of choice.

A significant part of mobile query volume is represented by local search queries, where users search for nearby businesses, parks, attractions and activities. Even though mobile local search is similar to desktop local search, there are two fundamental differences. First, mobile devices provide significantly more accurate location information (e.g. GPS, cell tower and/or wifi triangulation) compared to desktop devices (e.g. reverse IP localization techniques). Accurate location estimation is critical in mobile search since the users are on the go and their range of reach might be limited.

Second, when compared to desktop search, mobile search is more "actionable" in the sense that mobile users usually take an action immediately after a local search session (e.g. visit a restaurant, a grocery store etc.). Because of mobile search's actionable nature, the role of the user's *current* context is particularly important in successfully answering a query. For instance, knowing that a mobile user searching for restaurants is walking in downtown Manhattan during rush hour on a Friday evening, can provide invaluable information such as how much distance this user is willing to travel to visit a business or what type of businesses he might be interested in given the day and time of his search. On the other hand, the context of a desktop user that searches for restaurants from the comfort of his house right before he goes to sleep on a Monday night, might be irrelevant given that the search session might be triggered by a totally different context (i.e. plan a night out with friends for next Saturday).

The context of a mobile query can be defined by a collection of different features such as time of day, day of week, weather, traffic, user preferences and more. Several research efforts have already looked at the importance of context in local search or have attempted to analyze how mobile users click on local businesses [14, 1, 6]. The ultimate goal of these approaches is to learn a ranking function that properly balances the importance of all these different features to provide accurate business rankings. However, the way these features should be balanced might be different across locations. Users at different locations can have different decision processes either due to geographic properties of their regions or demographics of the area they live in. For instance, a mobile user in Manhattan on a Friday evening around 5pm, most probably is willing to travel a short distance to visit a business because of the

Category	Query Volume (%)
Food & Dining	35.36
Shopping	9.31
Arts & Entertainment	7.87
Health & Beauty	7.34
Home & Family	6.64
Automotive & Vehicles	5.61
Travel	5.39
Government & Comm.	4.46
Real Estate & Construction	4.12
Sports & Recreation	3.97
Computers & Technology	3.56
Legal & Finance	2.64
Professional & Services	2.36
Education	1.37

Table 1: Profile of the analyzed mobile search log dataset.

heavy traffic and the difficulty to access the highway. On the other hand, a mobile user in Texas might be willing to use his car and drive a longer distance because of the ease of acess to the highway. Consequently, the relative importance of the different context features might vary across locations. A mobile local search engine can capture these variations by (i) properly leveraging location-aware features to implicitly condition the ranking function on location, (ii) training multiple ranking functions across locations or (iii) by simultaneously combining both approaches.

Properly defining the location granularity at which local search models should be trained to effectively capture location context is quite challenging. While training models at the country level (i.e. US) might mask the importance of local context, training models at low levels of location granularity (i.e. zip code level) can also be inefficient due to data sparsity and data over-fitting issues. In addition, an optimal location resolution at which local search models should be trained might not exist. For instance, even though a high query volume area, such as Manhattan, may provide enough data to accurately train a local model, other nearby locations to Manhattan, such as Bedford, NY, might not. Hence, a more general model obtained at the state or even country level could be used to answer queries that originate from Bedford. In general, the number of local search models to be trained, the location resolution at which these models should be trained and how these models are combined at run-time to answer queries needs to be determined.

2. CONTRIBUTIONS

In this paper we study the impact of location context in mobile local search and make the following contributions. First, we quantify how differently mobile users click across locations by analyzing 2 million local search queries submitted in the United States.

Second, we implement and evaluate two approaches to encode location context in current state-of-the-art models: *implicit* and *explicit* location-aware training. In both approaches we augment the feature space with a set of new location-aware features and allow the model at the training phase to automatically decide the importance of these features. By picking up those features, the trained model implicitly takes location context into account. In the case of *implicit* training a single model is trained across the US, while in the case of *explicit* training, multiple models at different levels of location granularity (e.g. city, state, and country levels) are trained by segmenting the training dataset based on geographical and volume characteristics. At run time, the model at the lowest level of location granularity that is available for the location where the query originated from, is used to rank nearby businesses. By

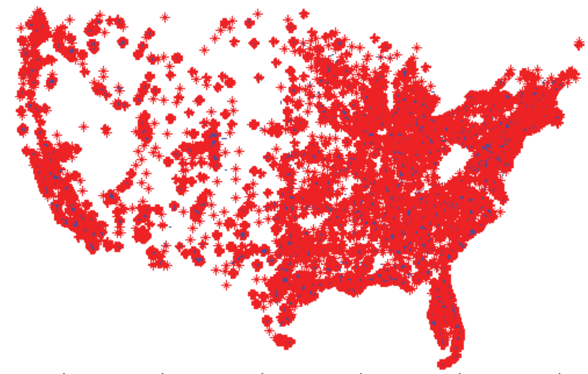


Figure 1: US coverage of the analyzed dataset. Each query is represented as a point at the location where it was submitted.

applying these techniques on real user query streams, we show that Precision can be improved by up to 47% when compared to to the Precision achieved by a single click prediction model for the whole US.

Third, we demonstrate that mobile search logs can be used to extract a model mapping at the zip code level that can guide the application of multiple models at different levels of location granularity at run time to improve click prediction accuracy. In the offline phase, we leverage the search logs to train models at the city, state, and country levels. By applying these models on search log traces and comparing their accuracies at the zip code level, we automatically learn a mapping between zip codes and models for each state in the US. At run-time, we lookup the appropriate model (city, state or country level) to use for each query based on the zip code where the query originates from and the learnt mapping. Through a 10-fold cross validation on real user query streams from a commercially available search engine, we show that we can achieve up to 4% higher Precision compared to the Precision that a single click prediction model at the state level with the location-aware features can achieve.

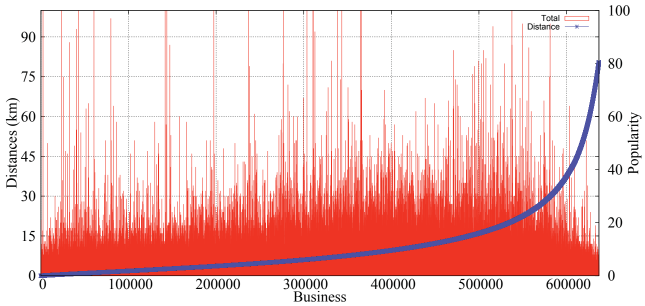
3. DATASET AND TOOLS

In this section we give an overview of the analyzed mobile search log dataset and describe the learning tools we use to train local search click prediction models.

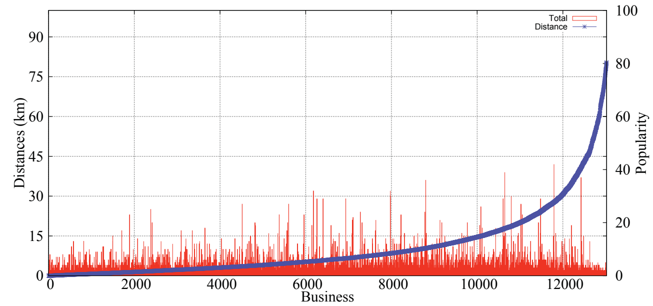
3.1 Data Profile

The dataset used in this paper consists of approximately 2 million local search queries submitted to a major search engine across the United States over a period of 3 months. All the queries were submitted from mobile users that opted to download and install the search engine's mobile application on iPhone and Android devices. The information recorded and analyzed for every query included the GPS location where the query was submitted, the actual query string, the unique ID of the business that was clicked by the user, the unique IDs of the businesses shown to the user, and a timestamp. To protect the privacy of the users, no other information about the user or the query was recorded.

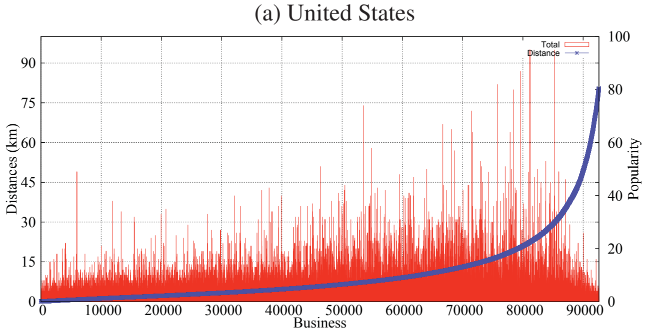
The original information recorded in the search logs was augmented with publicly available demographic data from recent CENSUS reports [17]. Using the GPS location available in every search log entry, we retrieved the zip code from where each query was submitted. Based on this information, we were able to associate zip code level demographic data (i.e. population density, average household size etc.) to every search log entry from the 2000 CENSUS report.



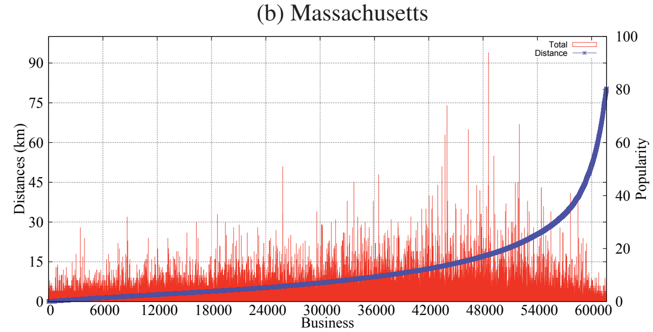
(a) United States



(b) Massachusetts



(c) California



(d) Texas

Figure 2: Distance-popularity relationship across the US and 3 representative states. For every business ID (x-axis) in the search logs, the average distance in km between the location of the query that resulted into a click for that business and the location of the business is shown (y-axis on the left). The business' normalized popularity based on the number of clicks it received in the logs (histogram, y-axis on the right) is also shown. Businesses are sorted based on average distance.

In addition, the unique ID of every business was used to retrieve more detailed information such as business' location and zip code.

Table 1 shows the breakdown of the analyzed query volume based on the type of businesses clicked. More than half of the query volume targeted businesses in the dining, entertainment and shopping categories. Another 20% of the volume targeted businesses around personal, home and automotive care. The remaining 30% of the query volume was distributed over travel and sports related businesses as well as over various public and private services.

Figure 1 shows the geographical coverage of the 2 million search queries analyzed in this paper. The query volume is distributed across the United States with the states of CA, WA, MA, AZ, NY, and TX being among the highest query volume producing states.

3.2 Training Tools

To provide a common training framework on top of which we can compare the performance of different modeling techniques and features, we adopt MART [20], a learning tool based on *M*ultiple *A*dditive *R*egression *T*rees. MART is based on the stochastic gradient boosting approach described in [4, 5, 7]. We formulate the mobile local search problem as a *click prediction* problem[1] and leverage MART to solve it. Each entry in the training and validation data contains a set of n features, $\mathcal{F}_q = \{f_{q_1}, f_{q_2}, \ldots, f_{q_n}\}$, that might be related to a query q (e.g. query zip code), a business object b (e.g. popularity), or both (e.g. distance between query and business) in conjunction with a click label which records the user's response c (1 for click and 0 otherwise). The training data is fed into MART to build a classification model, $\mathcal{M}$, which we use to estimate the probability of clicks $p_{\mathcal{M}}(c|q, b)$ in the test data. In our experiments we use the log-likelihood as the loss function (optimization criterion), steepest-descent (gradient-descent) as the optimization technique, and binary decision trees as the fitting function.

In addition to the trained model $\mathcal{M}$, MART reports a relative ordering of all the features $\mathcal{F}_q$. This ordering indicates the "*relative importance*" of the features during click prediction and provides insight on how the model is built. The most important feature has a relative importance of 1, while the rest of the features have a relative importance between 0 and 1.

For each experiment, we report the relative feature importance, and the Precision achieved by the different models for different Recall values.

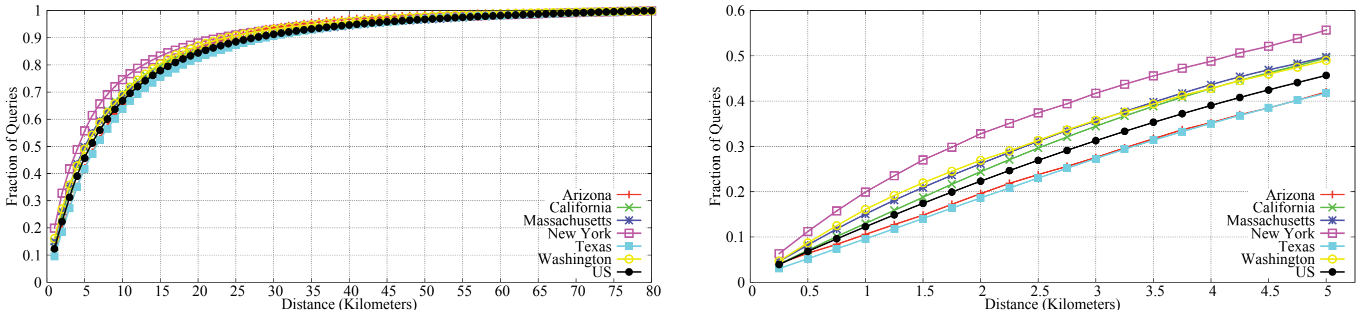
The baseline set of features used for training MART models consisted of 5 representative features that were selected based on previous research studies [14, 1, 6]. For every business click and non-click, we record: *(i)* the position in the search results where the business appeared, *(ii)* the distance between the query and the business locations, *(iii)* the popularity of the business as defined by the number of clicks in the search logs, *(iv)* the time frame within a day (one out of four 6-hour time frames) that the query was submitted, and *(v)* a binary feature that represents if the query was submitted on a workday or over the weekend. For every training experiment, the input dataset is split to training, verification and test data with the data volume ratio being $70\% : 20\% : 10\%$ respectively.

4. MOBILE SEARCH LOG ANALYSIS

In this section we visualize and quantify how differently mobile users click across locations. In particular, we compare click behavior at the country, state and zip code levels.

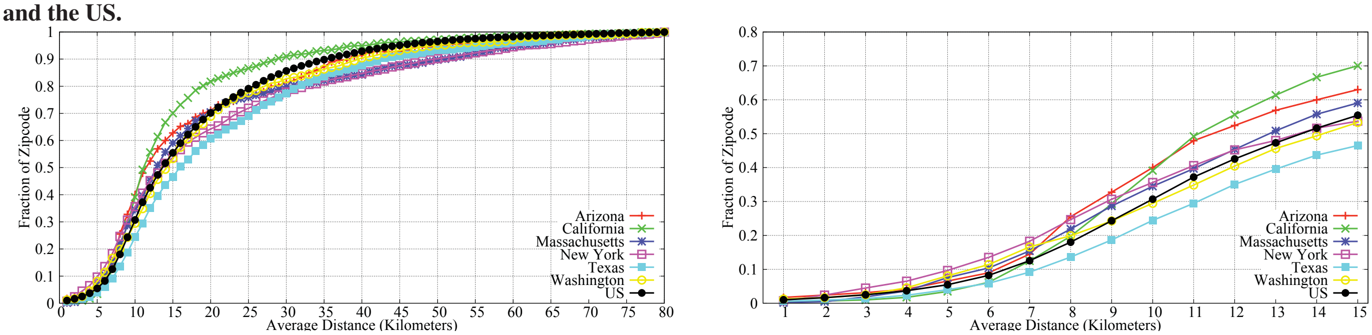
In this analysis, we characterize click behavior using different forms of two features that recent research studies [14, 6] as well as our current experience with 2 million queries has shown to be among the most important features: *traveling distance* (or simply distance) and *business popularity*. The former corresponds to the distance that a mobile user is willing to travel to visit a business

[1]Click prediction is a proxy for ranking quality. Given the close relationship between clicks on a search result and its relevance, it is very likely that features that help improve click prediction will be useful in ranking as well.



(a) (b)

Figure 3: CDF of the distance between the location of the query and the location of the clicked business for 6 representative states and the US.



(a) (b)

Figure 4: CDF of the average distance between the location of the query and the location of the clicked business across zip codes for 6 representative states and the US.

and is calculated as the distance between the location of the query and the location of the clicked business. The latter corresponds to the popularity of a business and is calculated based on the number of clicks each business receives in the search logs.

Due to space constraints, we only present the results for a representative subset of locations, features and demographic data studied.

4.1 Country and State Level Statistics

First, we look at the tradeoff between distance and business popularity. In general, mobile users wish to travel short distances, but at the same time they are willing to visit popular businesses around them. The way users tradeoff these two features is vital during ranking.

Figure 2 shows the tradeoff between distance and business popularity when examining the queries across the whole US, and the states of MA, CA and TX independently. Current techniques use data across the whole US to capture the distance/popularity tradeoff in a single model (Figure 2(a)). According to Figure 2(a), the mostly clicked businesses can be roughly classified into two categories based on the distance the mobile user is willing to travel. About 30% of the popular businesses are clicked within a 2-km radius of the mobile user's location. However, mobile users are willing to travel anywhere between 5km and 15km to reach about 70% of the most clicked businesses.

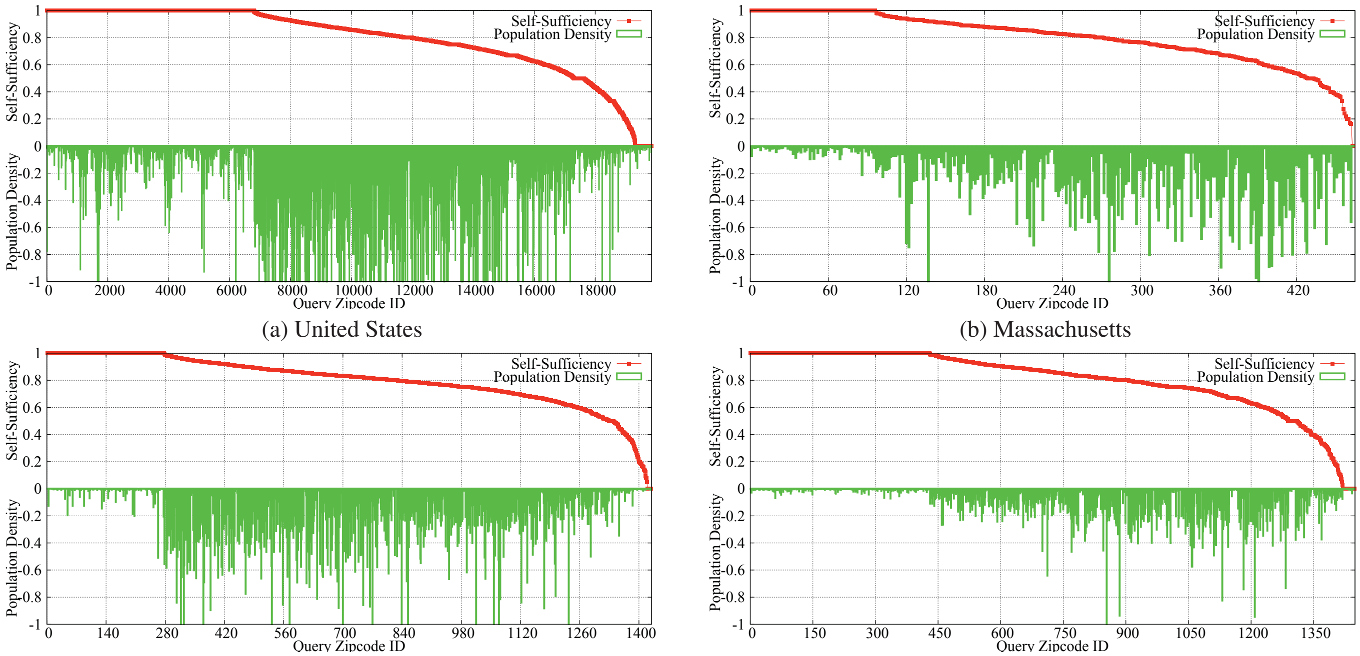
Aside from the fact that mobile users in the US are willing to travel a surprisingly large distance to visit a business, interesting observations can be made by comparing the statistics across the whole US with that of individual states. The states of CA and TX seem to share the same trends with the US as the most clicked businesses tend to be located between 5-15km from the user's location. However, the number of clicked businesses within a 2-km radius from user's location, is lower for the state of CA and significantly lower for the state of TX. Even more astounding is the comparison between Figures 2(a) and 2(b), that shows that mobile users in the state of MA tradeoff distance and business popularity in a different way than people across the US.

4.2 Variation Across States

Figure 3 provides deeper insight on how much distance mobile users are willing to travel and how this distance varies across different states. Surprisingly, only 12.5% of the mobile queries across the US result into clicking a business that is within a 1-km radius of the user's location. Looking at individual states, only 10% of the queries in the state of TX result into clicking a business within 1km of user's location. However, this percentage doubles (20%) when considering the sate of NY. The gap across states increases as the distance increases. For instance, the percentage of queries that results in a business click within 5km of user's location is 56% in NY; approximately 32% higher than TX (42.5%). In other words, mobile users in the state of NY are willing to travel shorter distances when compared with users in the state of TX. These differences could be caused, among others, by the geographical differences between Manhattan and TX. Given the heavy traffic conditions in Manhattan and the excellent subway network, mobile users in Manhattan are more likely to want to travel a short distance. On the other hand, mobile users in TX that have easy access to a well built highway networks might be more willing to travel longer distances to visit a business. A click prediction model that can effectively capture this information can have the flexibility to provide more location-aware, and thus more accurate, click predictions.

4.3 Variation Across Zip Codes

To quantify the variation across zip codes, we grouped the queries for every state based on the zip code that they originated



(a) United States (b) Massachusetts

(c) California (d) Texas

Figure 5: The fraction of queries where the zip codes of the query and the clicked business match. A value of 1 means that all the queries from the given zip code resulted into clicking businesses in a different zip code. The histogram at the bottom part of the figure shows the normalized population density for the given zip code across all zip codes.

from and computed the average traveling distance for every zip code. Figure 4 shows the CDF of the average traveling distance across zip codes for 6 representative states and the US. At a high level, users click on businesses within 5km from their location from only 5% of the zip codes. Conversely, for 30% of the zip codes, mobile users tend to click on businesses that are more than 20km away from their current location. The average traveling distance for the remaining 65% of the zip codes varies significantly across states. For instance, the percentage of zip codes in the state of TX where the average distance is less than 20km is 60%, as opposed to 82% for the state of CA. This difference represents a gap of approximately 37% across states (60% vs. 82%), indicating that even at the zip code level granularity mobile users exhibit different click behaviors.

Figure 5 sheds more light on the variation across zip codes. For every zip code we compute the *self-sufficiency* value, that is the fraction of queries where the query and clicked business zip codes are different. A value of 0 on the y-axis means that all the queries from the given zip code resulted into clicking businesses in the same zip code, and therefore we call that zip code *self-sufficient*. A value of 1 means that all the queries from that zip code resulted into clicking businesses in a different zip code and therefore we call that zip code *self-insufficient*. Figure 5 shows a clear segmentation of zip codes. Approximately 35% of the zip codes across the US are self-insufficient, while only 5% of the zip codes are self-sufficient. The rest 60% of the zip codes have a *self-sufficient* value between 0 and 1, with the majority having a value above 0.7. In other words, mobile users are willing to move away from their immediate vicinity in about 70% of the zip codes across the US.

In addition, the distribution of zip codes in terms of their *self-sufficient* value varies across states. For instance, in the state of TX, 40% of the zip codes are *self-insufficient*, while in California less than 20% of the zip codes are *self-insufficient*. Figure 5 also shows the normalized population density for each zip code. From the individual states figures, it is apparent that zip codes with low population density are usually *self-sufficient* or *self-insufficient*. However, Figure 5(a) suggests that about 25% of the *self-insufficient* zip codes across the US exhibit medium to high population densities.

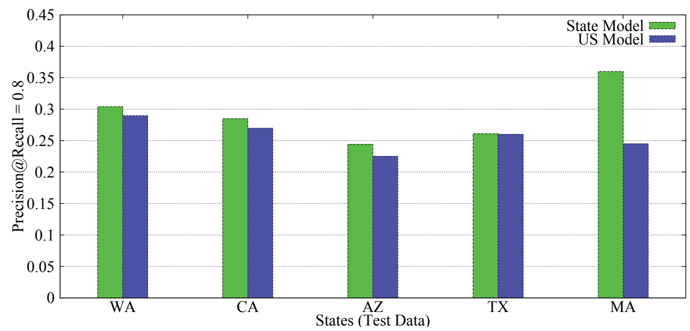


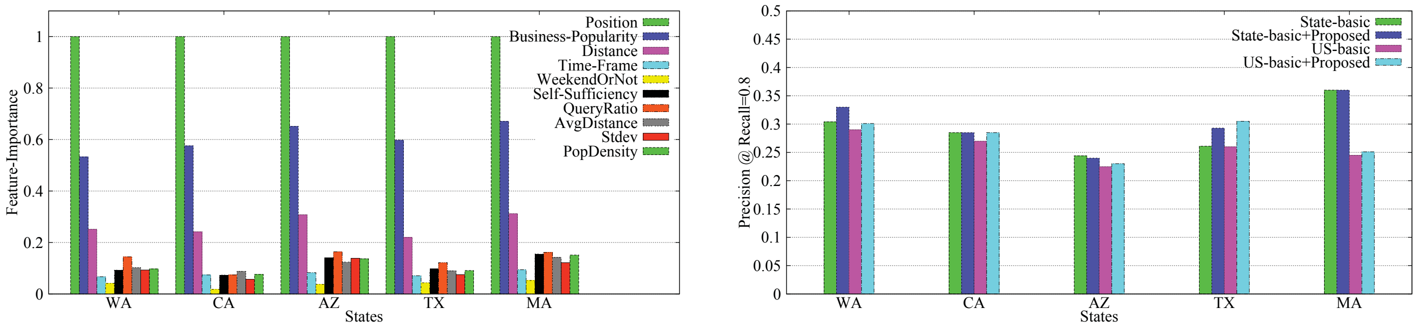
Figure 6: Precision achieved for Recall value of 0.8 when the state and US models are applied on each state's test dataset for 5 representative states.

Figures 4 and 5 show clearly that mobile click behavior varies at the zip code level across states as well as within a state. Thus, knowing the zip code from where a query originated from, can provide significant insight on how to answer the query.

4.4 The Effect of Location on Click Prediction Accuracy

To quantify the impact of our findings on the local search click prediction accuracy, we leveraged the MART training tool, described in Section 3.2, to train, validate and test different click prediction models on the recorded search log data.

Using the query volume from 5 representative states, a local search click prediction model was trained for every state. In a similar way, a US click prediction model was trained using the search log data across the whole US. To quantify how differently these



(a) Feature Importance (b) Precision @ Recall= 0.8

Figure 7: (a) Feature importance and (b) Precision performance for the US and 5 representative state models with and without the proposed location-aware features. All models were evaluated against the test dataset of each state (x-axis).

models perform across locations, we applied both the US and state models on each state's test dataset.

The results are shown in Figure 6. The state models consistently achieve higher Precision compared to the US model indicating the importance of location context. The increase in Precision by the use of state models varies anywhere from 0.4% to 47% depending on the individual state. In general, the state models for the states for which click behavior is similar to the click behavior across the whole US (*i.e.*, TX and CA states as shown in Figure 2) achieve minor or modest Precision improvements (0.4% and 4.8% respectively). On the other hand, states for which click behavior is significantly different than that across the US (*i.e.*, MA state as shown in Figure 2), the increase in Precision can be as high as 47%.

Table 2 provides more insight on the performance of models at different levels of location granularity by comparing the achieved Precision for different Recall values. *Columns 2 and 4 in Table 2 show that the state model achieves on average 3%, 6%, 11%, and 13% higher Precision than the US model for a Recall value of 0.5, 0.6, 0.7, and 0.8 respectively.* Only in the case of CA and TX states (that according to Figure 2 share the same click behavior as the whole US) and for high Recall values, the US and state models achieve identical Precision. Overall, simply training models at the state level, can provide significant improvement in Precision performance indicating the impact of location content in click prediction.

	Model(Features) (B): Basic Features, (P): Proposed Features			
Recall Value	WA(B)	WA(B+P)	US(B)	US(B+P)
50%	0.509	**0.52**	**0.485**	0.5
60%	0.434	**0.462**	**0.42**	0.44
70%	0.35	**0.401**	**0.326**	0.375
80%	0.304	**0.33**	**0.29**	0.301
	MA(B)	MA(B+P)	US(B)	US(B+P)
50%	0.47	**0.47**	**0.436**	0.437
60%	0.434	**0.432**	**0.355**	0.36
70%	0.389	**0.387**	**0.276**	0.29
80%	0.36	**0.36**	**0.245**	0.251
	AZ(B)	AZ(B+P)	US(B)	US(B+P)
50%	0.37	**0.371**	**0.369**	0.378
60%	0.325	**0.328**	**0.317**	0.32
70%	0.272	**0.273**	**0.26**	0.267
80%	0.244	**0.245**	**0.225**	0.23
	CA(B)	CA(B+P)	US(B)	US(B+P)
50%	0.43	**0.43**	**0.427**	0.43
60%	0.349	**0.369**	**0.349**	0.36
70%	0.294	**0.297**	**0.294**	0.3
80%	0.285	**0.285**	**0.272**	0.285
	TX(B)	TX(B+P)	US(B)	US(B+P)
50%	0.423	**0.423**	**0.415**	0.42
60%	0.346	**0.37**	**0.335**	0.36
70%	0.28	**0.322**	**0.28**	0.33
80%	0.261	**0.293**	**0.26**	0.305

Table 2: Precision for different Recall values for the state and US models with and without the proposed location-aware features in the case of 5 representative states in the US.

5. LOCATION-AWARE CLICK PREDICTION

To effectively capture the variance of mobile click behavior across locations, we propose and evaluate two different approaches: *implicit* and *explicit* location-aware modeling. In the case of *implicit* location-aware modeling, we train a single click prediction model for the whole US but we augment the feature set described in Section 3.2 with new location-aware features. In the case of *explicit* location-aware modeling, we still expand the feature space with the same location-dependent features, but this time a different click prediction model is built for every state or even city (given that it generates a high-enough search log volume (e.g. Manhattan, NY)).

Based on the analysis presented in the previous sections we introduce 5 new location-aware features:

1. *Self-sufficiency* value of the zip code the query originated from.

2. Average traveling distance within the zip code where the query originated from.

3. Standard deviation of the distance within the zip code where the query originated from.

4. Population density of the zip code where the query originated from.

5. Fraction of the state's query volume generated from the zip code the query originated from.

Note that all the features introduced encode information about the zip code where the query originated from. Providing these features can guide the training process to identify local properties at the zip code level and properly capture variance across zip codes in the trained model. Also, by training individual state models, we explicitly take into account the variance across states.

Figure 7 shows the results for both implicit and explicit location-aware modeling. Figure 7(a) shows the relative feature impor-

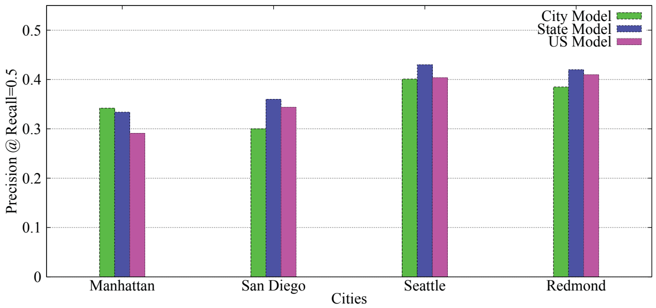


Figure 8: Precision achieved (for Recall=0.5) by models at the city, state and country levels for 4 representative high query volume cities. All models were evaluated against the test dataset of each city.

Recall Value	Model		
	Manhattan	NY	US
50%	0.342	0.334	0.291
60%	0.33	0.284	0.281
70%	0.32	0.262	0.282
80%	0.295	0.262	0.263
	Model		
	Seattle	WA	US
50%	0.401	0.43	0.404
60%	0.35	0.365	0.355
70%	0.32	0.281	0.302
80%	0.281	0.252	0.26
	Model		
	Redmond	WA	US
50%	0.385	0.42	0.41
60%	0.335	0.34	0.354
70%	0.278	0.245	0.263
80%	0.21	0.135	0.15
	Model		
	San Diego	CA	US
50%	0.3	0.36	0.344
60%	0.258	0.329	0.265
70%	0.233	0.267	0.201
80%	0.214	0.238	0.211

Table 3: Precision for different Recall values for the city, state, and US models in the case of 4 representative high query volume cities. Each model was applied on the test dataset of the individual city.

tance reported by the MART tool for every input feature across all state models and the US model. All the proposed location-dependent features are leveraged in the training process across all models. As expected, features such as the traveling distance still remain the most important features in terms of click prediction. However, location-dependent features are shown to be more important than temporal features such as the time window, or the weekday/weekend feature that have been previously shown to have a significant impact on mobile click prediction [14, 1, 6]. Furthermore, even though the feature importance trends are similar across all models, the importance of each location-dependent feature varies across different state models demonstrating the variance in mobile click behavior across locations.

Figure 7(b) shows the Precision achieved by the state and US models on the state's test dataset for all 5 states. To facilitate the discussion, we also show the Precision achieved by the state and the US models when none of the proposed location-dependent features is used. When examining the US model's performance across states, the use of location-dependent features leads to a Precision improvement that lies anywhere between 2.2% (AZ test dataset) and 17% (TX test dataset) depending on the state considered. When examining the performance of individual state models, the Precision improvement varies between 0% (CA and MA test datasets) and 12% (TX test dataset) when the proposed location-dependent features are used. *Comparatively, the Precision performance of state click prediction models is anywhere between 6.5% (AZ state) and 43.4% (MA state) better when compared to the US model, and when both basic and proposed features are leveraged.* The only exception are the CA and TX states, for which similar Precision is achieved for both the state and US models when all the features are leveraged. This is expected given the similarity of click behavior in these states and across the US (Figure 2).

Most importantly, the Precision achieved by the state model when both basic and proposed features are used is significantly higher compared to the US model that leverages only the basic features and represents the current state-of-the-art approach. *In particular, the improvement in Precision performance is anywhere between 4.8% (CA state) and 46.7% (MA state).*

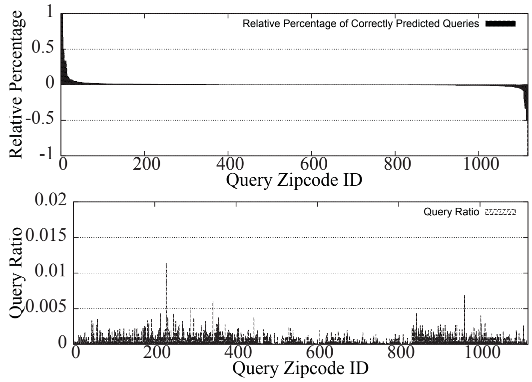
Note that the introduced location-dependent features have greater impact on the US model (implicit location-aware training) when compared to individual state models (explicit location-aware training). This trend is expected since the latter modeling methodology already leverages the location context by explicitly training different models for every state.

Table 2 quantifies in more detail the effect of the proposed location-dependent features on the Precision/Recall performance of the state and US models for the 5 representative states in Figure 7(b). Columns 3 and 5 in Table 2 show that both the state and US models achieve consistently higher Precision for a given Recall value when leveraging the proposed location-dependent features. *Specifically, in the case of WA state's dataset, state model's Precision increases anywhere between 2.2% and 14.6%, while US model's Precision increases anywhere between 3% and 15%, demonstrating the importance of the proposed features.* In addition, columns 3 and 4 show the improvement of leveraging models at the state level and incorporating the proposed location-aware features. Depending on the state and Recall value, an improvement of anywhere between 0.5% and 46.7% in Precision is achieved when compared to the Precision achieved by the US model when only the set of basic features is leveraged.

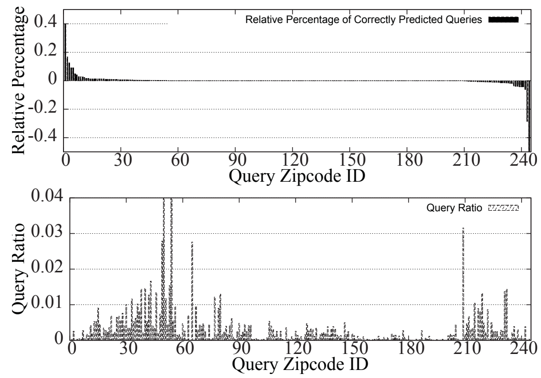
5.1 City-level Modeling

Training click prediction models at lower levels of location granularity can be extremely challenging due to data sparsity and data over-fitting problems. Even though the search log data provide a good indication of high volume areas within a state that one can successfully train click predictions models, the performance of these models might not be the best possible. For instance, Figure 8 shows the Precision performance of models at the city, state, and country levels for 4 of the cities that were among the highest query volume producing cities in the dataset. Surprisingly, city-level models don't always achieve the best Precision. Similar trends can also be seen in Table 3 where the Precision/Recall performance of the city, state and US models are shown on the test dataset of a representative set of cities. Different models achieve the highest Precision across different Recall values.

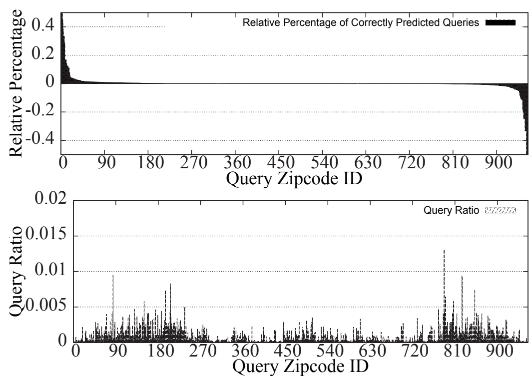
This observation can be explained by the fact that location-independent features, such as the position of a business in the search results list, are among the most important features during click prediction (Figure 7(a)). When we train models at lower lev-



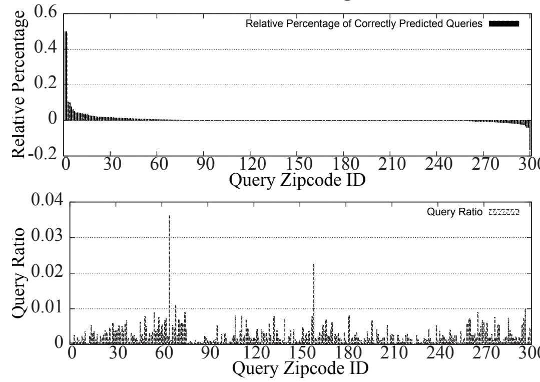
(a) California



(b) Washington



(c) Texas



(d) Massachusetts

Figure 9: Relative percentage of correctly predicted queries per zip code. A positive value, say 0.5, indicates that the state model is able to successfully predict 50% *more* clicks at the given zip code when compared to the US model. Conversely, a negative value, say -0.5, indicates that the state model is able to successfully predict 50% *fewer* clicks at the given zip code when compared to the state model. The histogram at the bottom part of the figure shows the normalized query volume for the given zip code across all zip codes.

els of location granularity (*e.g.*, city level), we enable better modeling of the location-aware features. However, because the volume of data available for training is naturally lower at lower levels of granularity (e.g. city-level vs. state level), the modeling of location-independent features might not be as accurate as when data a coarser level of location granularity is used. Given that location-independent features are significantly more important than location-dependent features (Figure 7), it is possible that models at the city level can achieve worse click prediction.

Even though Figure 8 suggests that city-level click prediction models are ineffective, this is not the case. As we show in the next sections, city-level models are valuable when applied intelligently on the test dataset.

5.2 Leveraging Multiple Models at Run-time

Our data analysis so far has shown that training models at the state level and leveraging the proposed location-dependent features, allows for singificantly more accurate click prediction in mobile local search.

However, a more careful comparison of Precision performance results in Figure 7, and in Table 3 indicates an interesting inconsistency. According to Figure 7 the click prediction of state-level models is consistently better than the click prediction of the US model. However, Table 3 suggests that this is not true when examining subsets of a state's test dataset. For instance, even though the WA state model achieves the best Precision on the state's test dataset, in the case of Redmond, the US model is able to achieve lower Precision compared to the WA state model. This inconsistency indicates that the relative performance of the models might be different across locations even within a state.

Motivated by the observed variance across zip codes in Section 4 and to further investigate this inconsistency, we study the click prediction performance for every model at the zip code level. In particular, for every state and for both the state and US models [2], we compute the percentage of correctly predicted clicks for every zip code that appears in the test dataset. Figure 9 shows the relative percentage of correctly predicted clicks at the zip code level for 4 representative state datasets. A positive value, say 0.5, indicates that the state model is able to successfully predict *50% more* clicks at the given zip code when compared to the US model. Conversely, a negative value, say -0.5, indicates that the state model is able to successfully predict *50% fewer* clicks at the given zip code when compared to the state model. Surprisingly, for the queries that originate from approximately 80% of the zip codes the state and US models exhibit very similar performance. However, for approximately 10% of the zip codes, the state model can successfully predict up to 60% more clicks, while for another 10% of the zip codes, the US model can successfully predict up to 40% more clicks.

As a result, in order to achieve the highest number of successful click predictions (and thus the highest Precision) over a state's dataset, we would like to apply the state model only on these zip codes for which it can successfully predict more clicks than the US model (these are the zip codes with *positive* relative percentage val-

[2]The exact same trends apply when considering other combinations of models such as city and state models, or city and country models. The results are not shown due to space constraints.

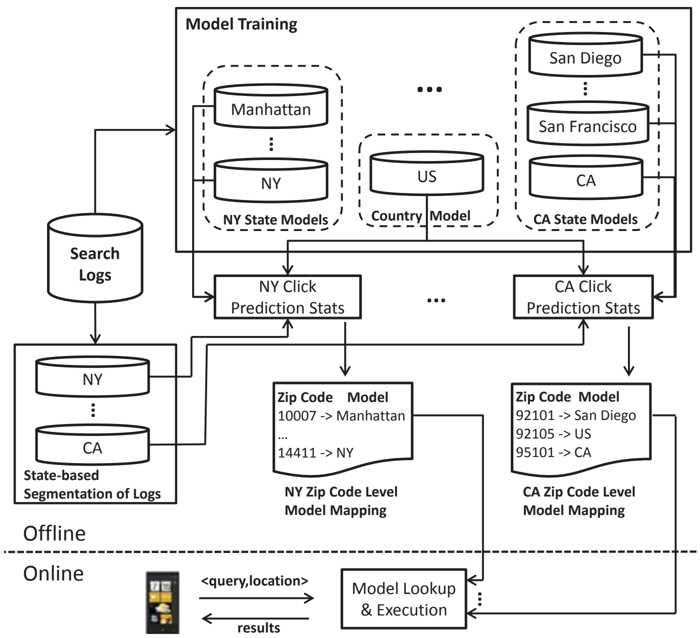


Figure 10: A mapping between zip codes and click prediction models at different levels of location granularity is automatically learned from the search logs.

ues in Figure 9). Similarly, the US model should only be applied on these zip codes for which it can successfully predict more clicks than the state model (these are the zip codes with *negative* relative percentage values in Figure 9). As Figure 9 suggests such a mapping between zip codes and click prediction models exists, and can be automatically learned from the mobile search logs.

Based on this observation, we implement a data-driven technique for leveraging multiple models at different levels of location granularity. The proposed scheme is described in Figure 10. Initially, the search logs are geographically segmented to different datasets; one for each state. At the same time the search log data across the US, individual states, and high query volume cities are used to train click prediction models at the country, state and city levels. The click prediction models available for a given state (country, state and available city models) are evaluated against the dataset of the state. *By computing and comparing the click prediction accuracy for each model at the zip code level (similarly to Figure 9), we create a mapping between zip codes and models so that each zip code is associated to the model that achieves the highest click prediction accuracy on this zip code's query trace*. When a query is submitted at run time, we leverage the learnt mapping to choose the appropriate model to use based on the zip code that the input query originates from.

5.3 Evaluation

The performance of the approach described in Figure 10 depends heavily on the ability to accurately learn the mapping between zip codes and click prediction models. To evaluate the effectiveness and stability of the approach, we perform a 10-fold cross validation on the mobile search log dataset.

Each dataset is split to 10 different collections of training, validation and test datasets with the data volume ratio being 70%:20%:10% respectively. For each collection of training, validation and test data, we leverage the training data to automatically learn the mapping between models and zip codes as described in the previous section (Figure 10). Then, we leverage the learnt mapping to properly apply the trained models on the actual test dataset.

	Model		
	WA	US	Combo
50%	0.455	0.41	0.457
60%	0.4	0.351	0.405
70%	0.347	0.292	0.348
80%	0.29	0.253	0.29
	MA	US	Combo
50%	0.444	0.351	0.445
60%	0.418	0.29	0.418
70%	0.392	0.257	0.391
80%	0.365	0.245	0.364
	AZ	US	Combo
50%	0.372	0.339	0.38
60%	0.33	0.295	0.34
70%	0.29	0.255	0.293
80%	0.272	0.223	0.272
	CA	US	Combo
50%	0.367	0.365	0.37
60%	0.315	0.298	0.32
70%	0.266	0.256	0.27
80%	0.235	0.234	0.235
	TX	US	Combo
50%	0.365	0.365	0.375
60%	0.311	0.31	0.32
70%	0.274	0.272	0.281
80%	0.25	0.25	0.26

Table 4: Precision for different Recall values for the state, US, and combination of city, state and country models (Combo) on the test datasets of 5 representative states.

The click prediction models were generated once and independently of the 10-fold cross validation experiments. We trained models at the country, state and city levels using the corresponding segments of the search log data and splitting them into 70% training, 20% validation and 10% test data. All models leverage both the basic set of 5 features described in Section 3.2 as well as the location-dependent features introduced in Section 5. For each experiment, we report the Precision/Recall performance for every model and test dataset combination.

Table 4 shows the achieved Precision for different Recall values that the state, US and the combination of state, US and city models achieve for 5 representative states' test datasets. Consistently across all Recall values and test datasets, the US model achieves the lowest Precision. When compared to the US model, the Precision achieved by the combo model is anywhere between 0.5% (CA @ Recall=0.8) and 52% (MA @ Recall=0.7) higher. On average, across all states and recall values, the combo model achieves 16% higher Precision compared to the US model.

When compared to the state model, the Precision improvement that the combo model achieves is more modest. On average, across all states and Recall values, the combo model achieves 1.2% higher Precision compared to the state model. Note that this improvement is additional to the one introduced by the location-dependent features that was quantified in Table 2. For individual states, such as AZ and TX, the Precision improvement is on average 2.5%, while for other states, such as MA, the combination of models at run time seems to have a negligible effect. This is expected given that the performance of the US model in the MA state's test dataset has been drastically lower compared to the state model's performance (Table 2). In this case, the performance of the state model dominates the performance of the other models, causing the combo model to become very similar to the state model.

6. RELATED WORK

Search logs have been studied extensively in the past for web search query profiling [3, 10, 11, 21, 2, 12, 13], topic or result clustering [18] and query intent detection [9, 15, 19].

The analysis closest to our work is the one on web search query profiling [3, 10, 11, 21, 2, 12, 13]. These efforts have analyzed search query volumes that vary from hundreds of thousands to several tenths of millions of queries across the US, Canada, Europe, and Asia. However, these efforts mainly focus on analyzing mobile web search trends such as repeatability of queries across users, length of queries and types of queries submitted. Few of these search log analysis efforts [10, 21] have provided insight about mobile local search, but only in terms of reporting the breakdown of business types that mobile users typically click on. Our work is among the first to analyze millions of mobile local search queries to understand how mobile users click on local businesses and how their click behavior varies across locations. In addition, our work goes beyond reporting general statistics. We propose new location-aware features to capture the variance of mobile click behavior across locations and quantify their impact in click prediction by leveraging state-of-the-art learning techniques.

In [19], Weber and Castillo show how demographic information can be leveraged to improve web search relevance and query suggestions. Conversely to [19], our work focuses on mobile local search, and most importantly, it goes beyond incorporating demographics into the ranking process and studies the variability of mobile local click behavior across locations.

Close, but complimentary to our work, is the work presented in [14]. Lane et al. analyzed approximately $80,000$ queries in the state of WA to understand how the context of a mobile query (i.e. traveling distance, business popularity, time, weather etc.) affects the quality of business ranking. They propose a set of context-aware features (several of these features were included in the basic set of features used in our experiments) and demonstrate their effectiveness by training click prediction models and evaluating them against real user query traces. Our work differs in two ways. First, we analyze mobile local search logs across the whole US and not within a specific state, and report on how mobile click behavior varies across locations. Second, we propose and evaluate new features that can efficiently capture the variance of mobile click behavior across locations in the click prediction models. This is something that the work in [14] does not focus on.

Researchers in the area of mobile context-aware systems have also performed user studies to highlight the importance of context in mobile local search. Amin et al. [1] and Froehlich et al. [6] highlight the importance of temporal context as well as the strong correlation between traveling distance and business popularity in mobile local search through long term user studies. In addition, Issaacman et al. [8] analyze call logs to understand specific aspects of mobile behavior across locations. In particular, they demonstrate that mobile users' behavior varies across locations by comparing the traveling distance of AT&T cellular customers in Manhattan and Los Angeles. However, none of these efforts has explicitly studied and quantified how mobile click behavior changes across locations, let alone evaluate its impact on click prediction models.

7. CONCLUSIONS

We have presented an in-depth analysis of mobile local search click behavior across the US. Our contributions were threefold. First, we analyzed 2 million local search queries to better understand mobile local search click behavior and to quantify how this behavior varies across locations. Second, based on our analysis we proposed a set of location-aware features to efficiently capture the variance of mobile click behavior across locations in the click prediction model. By leveraging these features and training models at different levels of location granularity, we showed, using real query traces, that Precision can be improved anywhere from 5% up to 47% across states. Third, we demonstrated that mobile search logs can be used to extract a model mapping at the zip code level that can guide, at run time, the application of multiple models at different levels of location granularity. By applying this approach on real query traces, we showed that Precision can be *additionally* improved by up to 4%.

8. REFERENCES

[1] A. Amin, S. Townsend, J. Ossenbruggen, and L. Hardman. Fancy a drink in canary wharf?: A user study on location-based mobile search. In *Proceedings INTERACT*, 2009.

[2] K. Church, B. Smyth, K. Bradley, and P. Cotter. A large scale study of european mobile search behaviour. In *Proceedings of MobileHCI*, 2008.

[3] K. Church, B. Smyth, P. Cotter, and K. Bradley. Mobile information access: A study of emerging search behavior on the mobile internet. *ACM Trans. Web*, 1(1):4, 2007.

[4] J. H. Friedman. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29:1189–1232, 1999.

[5] J. H. Friedman. Stochastic gradient boosting. *Comput. Stat. Data Anal.*, 38(4):367–378, 2002.

[6] J. Froehlich, M. Y. Chen, I. E. Smith, and F. Potter. Voting with your feet: An investigative study of the relationship between place visit behavior and preference. In *Proceedings of Ubicomp*, 2006.

[7] J. Gao, Q. Wu, C. Burges, K. Svore, Y. Su, N. Khan, S. Shah, and H. Zhou. Model adaptation via model interpolation and boosting for web search ranking. In *Proceedings of EMNLP*, 2009.

[8] S. Isaacman, R. Becker, R. Cáceres, S. Kobourov, J. Rowland, and A. Varshavsky. A tale of two cities. In *Proceedings of HotMobile'10*.

[9] B. J. Jansen, D. L. Booth, and A. Spink. Determining the user intent of web search engine queries. In *Proceedings of WWW*, 2007.

[10] M. Kamvar and S. Baluja. A large scale study of wireless search behavior: Google mobile search. In *Proceedings of CHI*, 2006.

[11] M. Kamvar and S. Baluja. Deciphering trends in mobile search. *Computer*, 40(8):58–62, 2007.

[12] 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 WWW*, 2009.

[13] E. Koukoumidis, D. Lymberopoulos, K. Straus, J. Liu, and D. Burger. Pocket cloudlets. In *Proceedings of ASPLOS (to appear)*, 2011.

[14] N. D. Lane, D. Lymberopoulos, F. Zhao, and A. T. Campbell. Hapori: context-based local search for mobile phones using community behavioral modeling and similarity. In *Proceedings of Ubicomp'10*.

[15] Y. Lu, F. Peng, X. Wei, and B. Dumoulin. Personalize web search results with user's location. In *Proceedings of SIGIR*, 2010.

[16] Mobile Search Trends Report, `http://www.marketingcharts.com/`.

[17] U.S. Census Bureau, `http://www.census.gov/`.

[18] X. Wang and C. Zhai. Learn from web search logs to organize search results. In *Proceedings of SIGIR*, 2007.

[19] I. Weber and C. Castillo. The demographics of web search. In *Proceedings of SIGIR*, 2010.

[20] Q. Wu, C. J. C. Burges, K. M. Svore, and J. Gao. Ranking, boosting, and model adaptation. Technical Report MMSR-TR-2008-109, Microsoft Research, 2008.

[21] J. Yi, F. Maghoul, and J. Pedersen. Deciphering mobile search patterns: a study of yahoo! mobile search queries. In *Proceeding of WWW*, 2008.