Hapori: Context-based Local Search for Mobile Phones using Community Behavioral Modeling and Similarity

Nicholas D. Lane, Dimitrios Lymberopoulos†, Feng Zhao† and Andrew T. Campbell
Dartmouth College, Microsoft Research†
{niclane,campbell}@cs.dartmouth.edu, {dlymper,zhao}@microsoft.com†

ABSTRACT
Local search engines are very popular but limited. We present *Hapori*, a next-generation local search technology for mobile phones that not only takes into account location in the search query but richer context such as the time, weather and the activity of the user. Hapori also builds behavioral models of users and exploits the similarity between users to tailor search results to personal tastes rather than provide static geo-driven points of interest. We discuss the design, implementation and evaluation of the Hapori framework which combines data mining, information preserving embedding and distance metric learning to address the challenge of creating efficient multidimensional models from context-rich local search logs. Our experimental results using 80,000 queries extracted from search logs show that contextual and behavioral similarity information can improve the relevance of local search results by up to ten times when compared to the results currently provided by commercially available search engine technology.

Author Keywords
Local Search, Context-ware Mobile Search, Mobile Phone Sensing

ACM Classification Keywords
H.4.m Information Systems Application: Miscellaneous.

General Terms
Design, Experimentation, Performance.

INTRODUCTION
Location-based services on mobile phones have been standard fare for several years. People on-the-go regularly use a variety of mobile local search applications throughout their day to find, for example, nearby restaurants, gyms and cafes to name just a few of a growing number of local search categories. Local search that takes into account the location of the phone as *context* for the search is perhaps one of the most exciting, popular and useful mobile applications today.

Existing mobile local search engines perform well when used for a particular narrow range of queries where the relevance of different *Points of Interest (POIs)* is clear, such as, finding the nearest coffee shop. As a result local search is limited. We imagine, for example, future search engines capable of taking into account more contextual information than just the location of the user as is the norm today. Mobile phones offer significantly richer context than simply location; for example, future queries could include any or all of the following context in a query: time of day, day of the week, weather conditions, the current activity of the user (e.g., jogging, biking, traveling by car or bus through the use of embedded sensors in smartphones [3]), whether the user is with friends or alone. Other innovations are possible too. We believe that local search can be advanced considerably by enabling search engines to personalize responses for each user by: 1) incorporating behavioral profiles and preferences with respect to contextual factors (e.g., weather); and 2) exploiting the choices made by others in the broader community who have similar behavioral histories.

To illustrate what we think is possible consider the following scenario. Two people, a senior and a teenager are located at the same position in a city and unbeknown to each other use the same service to issue the exact same query (e.g., entertainment) – not surprisingly they are both presented with the exact same geo-driven list of places using existing local search technology. Both are disappointed at their choices. Now, consider a next generation local search engine that is capable of understanding context and building behavioral models – i.e., time, date, weather, as well as taking into account the popular choices of other people in the community with behavioral similarity – then, it is not surprising that the senior and teenager are presented with a ranked list of POIs that are radically different from each other and as a consequence much more appealing to their individual tastes; as a result the teenager heads off in one direction to a free rock concert in Central Park and the senior heads off in the opposite direction to catch a popular foreign movie playing at Lincoln Center Cinemas (which, our senior is happy to be informed is air conditioned) at the exact same time on a very hot Saturday evening in the summer in Manhattan.

In this paper, we present the design, implementation and evaluation of *Hapori* (meaning community in Maori), the first context driven local search framework built on a foundation of community behavioral modeling and similarity. The goal of Hapori is to meet the diverse needs of different people (such as the teenager and senior) taking into account their *context, behavioral profile and behavioral similarities with others in the broader community of local search users*. With

Ubicomp 2010, Sep 26-Sep 29, 2010, Copenhagen, Denmark.

Hapori POI choices of the entire community drive the search results that everyone receives. We model the POI preferences of people, based on the context (e.g., weather, time, location) under which they make these choices. The preferences of people are then linked to form a community model based on behavioral similarity between people which determines how appropriate the preferences displayed by one person are to another person.

The Hapori framework discussed in this paper captures the knowledge of a community to improve the overall POI search relevance by: i) computing features that capture significant aspects of context that effect people's POI decisions; ii) learning customized ranking metrics that emphasize the specific important elements in a POI decision for a particular category of POI (be it for example the popularity and personal preference for a music club or weather and proximity to a running track), iii) modeling differences between people; and finally, iv) adapting to changes in community behavior (e.g., a new restaurant opens up, traffic patterns change, preferences shift). By mining the local search decisions of thousands of people, Hapori recognizes how the appeal of places of interest can change not only from location to location but, for example, change from weekday to weekend, from person to person, from season to season, or from morning to evening within a day.

Hapori makes two important contributions. First, it combines several data mining, information preserving embedding and learning techniques to address the challenges of creating efficient multidimensional models from local search logs. Second, it uses local search logs from Mobile Bing Local [11] containing queries from more than 11,000 users over a period of 6 months to *identify* the different context parameters and to *quantify* their effect on the way people click on businesses. Our experimental results using 80,000 queries extracted from the search logs of a leading commercial mobile search engine show that personalized results based on contextual and behavioral similarity information can improve the relevance of local search results by up to 10 times when compared to the results currently provided by Mobile Bing Local.

In our current prototype implementation of Hapori, we only use context information that we can extract from the mobile search logs (i.e. time of day, day of week, weather, search history etc.). However, the techniques employed in Hapori's design could be applied to any possible context information available. For instance, the available sensors on the phone could provide information about the context of the query such as if the person is inside or outside, if she is alone or with company, if she is driving or walking, or even if she has kids in the car. In practice, any sensing information could be used within the existing design of Hapori to better define the context of the query, narrow the intent of the user and provide more relevant POI information.

CONTEXT AND COMMUNITY BEHAVIOR ARTIFACTS

In this section, we analyze over 80,000 local search queries submitted to Mobile Bing Local [11] by more than 11,000 users over a period of six months beginning January 1st 2009. By analyzing search logs we clearly identify that POI selection of people is noticeably shaped by both context and the preferences within cliques of people – that is, the community preferences. We use this analysis to motivate and guide the design of the Hapori framework presented in next section. The results presented in this section are not meant to be exhaustive but illustrative of the fact that the artifacts due to context and user behavior can already be observed in existing systems. Hapori makes advanced context, behavioral modeling and the computation of similarity a first class citizen in the design of local search.

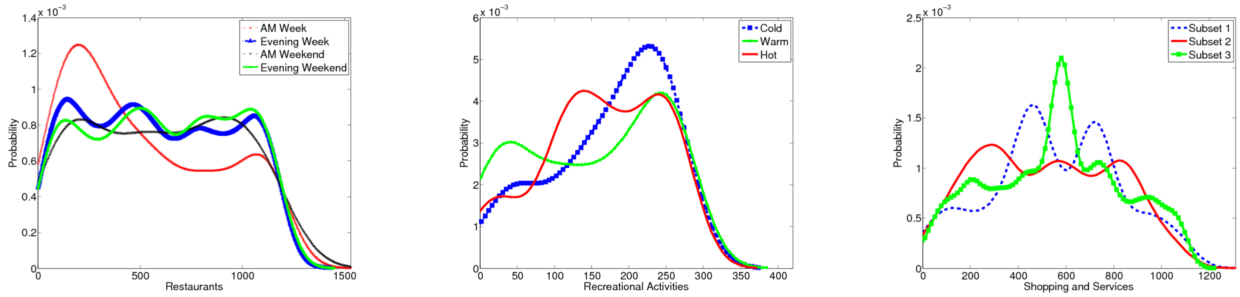
Search Logs

Every entry in the search logs we analyzed contain: query terms, a unique identifier for the POI that is clicked after the query is submitted, the location of the user, the exact date and time the query is submitted and an anonymized user identifier that can be used to link together queries submitted by the same user over time. Note, that to preserve the privacy of users, the exact location of the user (i.e., GPS location) is not recorded. Rather, coarse-grained location information at the resolution level of a map tile is stored for every query.

The analysis presented also uses the information provided in the search logs to reconstruct other pieces of useful information related to submitted queries. In particular, we use a unique POI (e.g., business) identifier to identify the location of the POI that is clicked for every query. This allows us to study the spatial correlations between the location of the query and the location of the clicked POI. In addition, we use the time and date of each query to retrieve weather conditions (e.g., temperature) at the time the query is submitted. This extended context allows us to study how, for example, weather conditions affect the way mobile users click on POIs. In what follows, we discuss our main findings from the analysis of the search logs.

Impact of Temporal Context

We find that temporal patterns play an important role in the decision process of mobile users. People's behavior and the activities they wish to engage in vary depending on if it is a weekday or weekend, or if it is an evening or morning. Figure 1(a) shows the click probability of 1500 different restaurants for 4 different time windows: weekday morning, weekday evening, weekend morning and weekend evening. Note, that during a weekday morning there are about 100 POIs (i.e., businesses positioned on the x-axis between 200 and 300) with very high click probabilities. These POIs mainly consist of fast-food places, informal restaurants and local coffee shops. In other words, eateries that most people visit before they go to work or during their day, e.g., getting a cup of coffee. The click probability for these POIs reduces significantly during weekends and weekday evenings with the highest drop happening during weekend evenings. The results intuitively indicate that mobile users are interested in a different set of POIs during weekend evenings (businesses with an x-axis position from 450 to 550 and from 800 to 1050). Note, also that these POIs show low click probabilities during weekday mornings. On the other hand, popular businesses during evenings seem to be similarly independent of the actual day considered (weekday versus weekend). Fi-



(a) (b) (c)

Figure 1. Popularity of businesses significantly changes across (a) time windows, (b) weather and (c) groups of users. Each number on the x-axis represents a unique business. The y-axis shows the click probability. Note that the same business numbers across the different plots does *not* correspond to the same business.

nally, there are about 250 POIs (businesses positioned higher than 1250) that have low click probabilities across all time windows. As a result, even if a user is searching for a restaurant nearby these POIs they do not seem to interest the user independent of the temporal context.

Impact of Weather Context

Weather condition is another important factor that mobile users take into account when they plan activities; for example, during a sunny day, a walk to the park or the closest lake might sound like a good idea but not when it's raining or snowing. During cold days many people might preferred to stay indoors, for example, visit a bowling facility where they could have fun and stay warm at the same time. It seems very intuitive that such issues play an important role during local search and this is held out by the data. Figure 1(b) shows the click probability of 450 recreational activities across three different weather conditions; that is, cold, warm and hot weather. The first 100 activities on the left of the Figure have significantly higher click probabilities during warm days. The next 50 activities (i.e., activities with positions from 100 to 150) have significantly higher click probabilities during hot days. Note, that there are approximately 100 businesses with equally high click probabilities across all weather conditions. Similarly to the temporal context, there is a set of businesses (i.e., businesses with positions higher than 275) that seem to have low click probabilities across all weather conditions.

Impact of Personal Context

Personal preference plays an important role in the selection of POIs, e.g., the keyword 'recreation' would likely lead to different choices when considering the teenager and senior discussed earlier. Figure 1(c) shows the click-through probability of three different subsets of users for 1300 different shopping and service businesses, each numbered uniquely along the x-axis. These subsets were identified based on a simple search heuristic where approximately even sized subsets were pulled from the community until the distribution of their historical POI clicks were beyond a fixed threshold of pairwise distinctness as measured by KullbackLeibler divergence. Figure 1(c) shows that the popularity of certain businesses are very different for each of the cliques of people. The third subset of users is mainly interested in businesses positioned between 500 and 700 along the x-axis. The first subset is interested in businesses positioned between 400 and 500, as well as 700 and 800, while the second subset of users is mainly interested in businesses positioned between 200 and 400. As a result, knowing the group which a person is implicitly associated with provides valuable information about which POIs she might have interest in. Computing similarity across a community of local search users to capture these group effects is a design principle that drives the Hapori framework, as discussed in the next section.

Impact of Spatial Context

The use of location as context in local search is common practice and the driving factor behind its current success. However, the use of location is typically very simple, such as, limiting the potential POIs used in query responses based on physical distance. Even though displaying the closest POIs to user's location is helpful, in many cases, it is insufficient. An important aspect of attraction to certain businesses is based on their popularity within the community; for example, how people gravitate toward a trendy new cocktail lounge or a restaurant known to have good food. This aspect of the POI decision can override other (perhaps more rational) factors when a person is selecting from a list of candidate POIs, such as, the difficulty of travel or cost and quality. Figure 2 shows the average distance between the location of the query and the location of the business/POI selected as well as the total number of clicks that each business received over the 6 month period. Note, that the closest 20-30 businesses to the query location were not as popular as POIs that were further away. In addition, as the average distance of POIs increases, the more businesses appear to receive a higher number of clicks. In many cases, users click on businesses that might be up to 3 kilometers away from the query location. This data clearly shows that mobile users are not always interested in the closest POIs. Many times mobile users want to visit POIs that are popular across the user population which might result in traveling further.

Limitations in the State of the Art

Figures 1 and 2 show the impact of temporal, personal, weather and spatial context on the way mobile users click on POIs. Many times these results are intuitive and reinforce common wisdom. Other times data reveals new insights. From the data it is clear that time, date, weather, popularity and association with a group (hidden from the user) play an important role in understanding which POIs the user is trying to reach. Local search services today ignore most of this

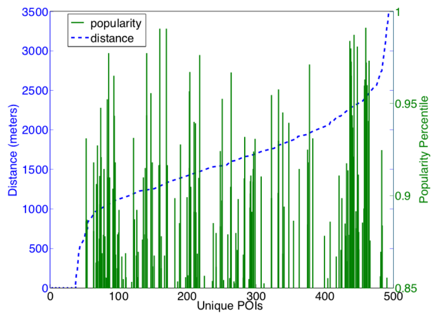


Figure 2. Biasing search results to the closest businesses to the user is not always effective. Mobile users are more interested in popular businesses even if they are further away.

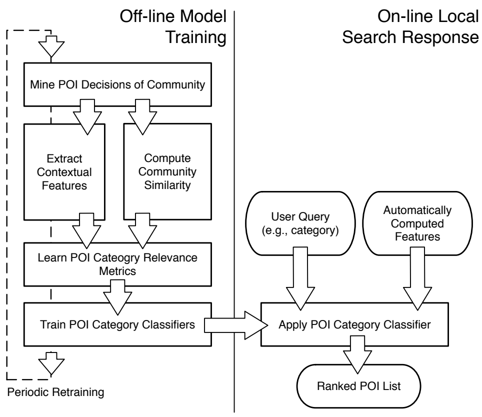


Figure 3. The stages of the Hapori framework which comprises an off-line model training process and an on-line local search request/response.

information and always display the same set of static search results for a given query and a given location: the result that first let down our teenager and senior when presented with static non-personalized POIs. However, as Figures 1 and 2 show, a static answer cannot capture the complex dynamics that govern the decision process of mobile users.

Hapori is designed to explicitly support the context and community artifacts shown in the results presented in this section. Hapori studies the extent to which these issues can be explicitly supported in the design of a new generation of local search services using modeling the decisions of the individual user and cliques of people who behave similarly to the user and the community at large.

Simple context such as temporal patterns, weather (e.g., temperature) can clearly effect the distribution of categories of POI. Changes in temperature are shown to skew the recreational activities chosen. Or changes in the time of day effect the restaurants people are interested in. Clusters of people, even when they are exposed to the same context, will select different POI. Context is therefore insufficient, the connections between groups of people and the POIs selected is as important as context. As a result we need to address context, community modeling and similarity in the design of Hapori, discussed in the next section.

HAPORI

In this section, we present a detailed description of the Hapori framework that is based on a *POI preference model* which reflects the POI selection process of the community and captures: i) contextual information that affect the decision made and ii) the personal preferences and behavior of both the people submitting queries and members of the community with whom they have similarities. Responses to local search queries are based on a model of how real people make their own personal POI selections.

Framework

The Hapori framework comprises an off-line model training process and an on-line local search response, as shown in Figure 3. The off-line model training part of the framework is required to build the POI preference model through a set of stages which we discuss in detail in this section. The resulting POI preference model generated is used by the on-line local search to respond to queries made by users carrying mobile phones. Figure 4 illustrates the different stages in the data flow that occur to construct a POI preference model. The focus of our contribution is to improve the relevance of the list of POIs returned to users by mining the POI decisions made by others in the community users have similarities with. The Hapori framework assumes only a minimal form of query as input where queries are specified by the selection of a simple POI category. The query is augmented with contextual information taken from sensors on mobile phones as well as historical information about the user built up over time by the Hapori. In what follows, we discuss each stage of the Hapori framework describing how the framework models the POI preferences of the community to improve the relevance of local search query responses.

Mining Community POI Decisions

The construction of the preference model begins with the mining of POI decisions from a community, such as the population of a city. We define a *POI decision* as simply a person demonstrating interest in a POI – clicking on one. Conventional local search services that operate on a mobile device (e.g., Google and Bing mobile search applications on iPhone) provide users with a ranked list of POIs that they can choose from. In these applications a user can either manually browse a list of POIs and then click on a specific POI that suits their needs or directly enter the name of a specific POI. We assume that user selection of a POI, in aggregate, is an indication that it is suitable given the behavioral characteristics of the person and the context under which the selection occurred. In our evaluation we constrain ourselves to use only POI decisions based on user clicks. However, POI decisions may be mined directly from the actions of a person, for example, when the person is jogging at a particular running track or shopping at a specific store. Sensors on their phone can be used to automatically classify such by activities by applying sensing and inferencing techniques [3]. This would be an implicit POI decision mined from actions of the user rather than their interactions with a mobile application. The Hapori framework is agnostic to how POI decisions are mined, as long as they are comprised of: sensor data sampled at the time of the decision (e.g., location, time); the ground-truth POI decision (e.g., a business, place,

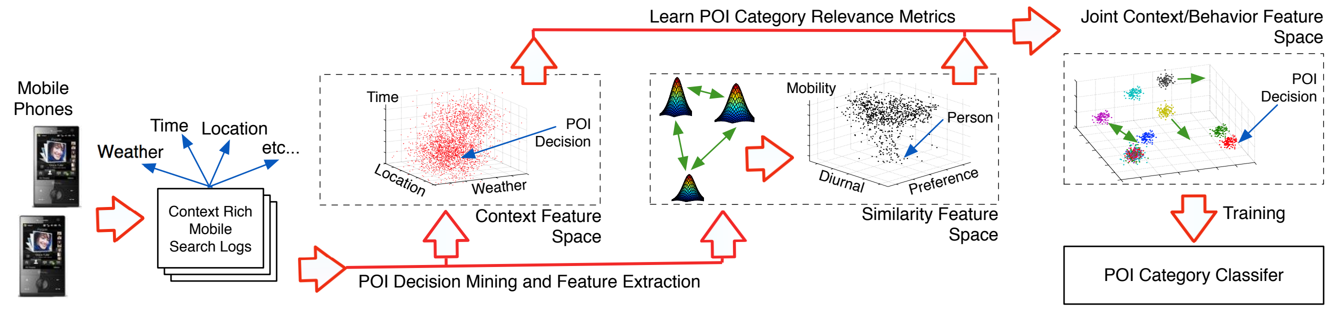


Figure 4. A POI preference model is trained for each POI category. Hapori begins this process with the mining of mobile local search logs from which it constructs context and behavioral similarity feature spaces. Before the model is trained both these feature spaces are fused. During this step a feature space is found that emphasizes the most important elements of both context and behavioral similarity for POI relevance given a POI category.

i.e., the POI selected by the person); and a session identifier or some type of token that can link together the POI decisions of an individual over a period of time.

Extract Contextual Features

For each mined POI decision a series of features are computed which are chosen to capture the contextual information associated with each decision. The extracted features allow the model to learn contextual patterns of POI decisions made by a community. In Figure 4 this is illustrated as the construction of a context feature space based on the mined POI decisions. Currently, Hapori incorporates four types of contextual features: *temporal, spacial, weather and POI popularity features*. We use these features because they represent different types of context that has strong influence over the POI decisions made by people. Most phones with GPS can compute these features. Clearly, this is not an exhaustive set of features and could be extended to cover other categories, such as, traffic levels, air quality, etc., as well as use richer descriptors of locations (e.g., land-use, demographics). These types of extensions to context would likely improve Hapori's ability to capture more interesting contextual patterns. In what follows, we discuss each of the four contextual feature categories in more detail.

Temporal Features. To capture the different types of temporal patterns people exhibit, different temporal features are used corresponding to the time-scales over which people alter their behavior. The following features are calculated based on the time and day when the POI decision occurs: the day of the week; a binary indicator for weekend or weekday; an intra-day count of elapsed four-hour windows; and intra-week versions of the same count of four-hour windows.

Spacial Features. The use of location in local search services is common practice. Typically this leads to simple responses such as scoping the potential POIs used in query responses based on physical distance from the user. Hapori captures more complex spacial relationships by computing features based on both the location where the POI decision (referred to as the source) is made and the location of the corresponding ground-truth POI (referred to as the destination). The following features are extracted for each POI decision: rounded latitude and longitude of both source and destination; the tile of both source and destination given a tessellation of the city area which is repeated for tile counts of $\{10^2, 256^2, 512^2, 1024^2\}$. Spacial patterns operate particularly well in combination with temporal features, for example, combining to capture the fact that people are willing to travel further on certain days (e.g., weekends) because they have more free time than during the week.

Weather Features. The weather has a large impact on the range of activities people will consider when they are searching for POIs. Hapori uses weather statistics from the day on which the POI decision is made, including: rainfall and snowfall totals; and the average temperature of the day. These three statistics are represented as separate features with different levels of discretization applied.

POI Popularity Features. An important aspect of the attraction of certain POIs is based on their popularity within a community. Two forms of such popularity exist: i) sharp spikes of interest in the community when for instance a new POI is created or improved; and ii) longer time-scale stable POI preferences in the community (e.g., perennially popular bars or fishing holes). Hapori measures popularity in terms of the percentile of POI decisions associated with each different POI. To capture the two forms of popularity Hapori computes this percentile relative to different sets of POI decisions. Sets are made up from POIs based on a series of filtering rules that are based on some aspect of the source query, such as: POIs that have the same POI category as the query, POIs within the same spacial tile assuming a 512^2 sized tessellation (as described in the spacial feature subsection) or a POI decision that is less than 3 weeks.

Computing Community Similarity

Hapori uses contextual features (discussed above) to model each POI decision, capturing the circumstances associated with the decision made. Not all POI decisions are given equal influence in effecting the results provided to users, however. A large component of the *relevance of a POI* to a local search query is subjective. This requires that the differences between people contributing POI decisions need to be captured in the model. Hapori uses a community similarity metric computed between all users to address this need. Illustrated in Figure 4 is the process of building a similarity feature space to complement the context feature space. The Figure shows similarity features based on the mined POI decisions of each person being used to compute a metric required for the similarity space.

Similarity Features. Five features mined from the POI decisions of a user are used to compute community similarity: i) two temporal features based on when queries are submitted – an intra-day count of elapsed four-hour windows and the day of the week; ii) one spacial feature – the tile of the source location tessellated assuming 512^2 tiles; and iii) two features based on the POI itself – the POI category (e.g., hair dressing) and the specific POI itself (e.g., joe's hair design) both of which are treated as discrete categorical features.

The temporal and spacial features capture the diurnal and spacial patterns of the life of the individual based on the accumulation of POI decisions over time; as observed from the points in time when a user submits a local search query.These patterns act as a latent observation on aspects of the person such as demographic characteristics (e.g., given the connection to where a person lives, works and spends most of their time) and lifestyle (e.g., regularity of their schedule, early morning or night-owl personality traits). However, two different types of people could potentially live in the same area and have fairly similar schedules (e.g., a parent and their teenage children) but have very divergent POI preferences. For this reason Hapori augments the temporal and spacial features with information about POI preferences based on the frequency at which specific POIs and POI categories occur over time within the stream of POI decisions made by individuals. Category and specific POI information allow Hapori to represent different degrees of similarity; for example, two people may like the same POI category, such as, evening entertainment or dancing, but have different preferences in terms of the type of entertainment that appeals to them. Category only information ignores this difference. If, however, only specific POI information is considered Hapori would ignore similarity based on the common categories. As in the case of the contextual features the current design of Hapori limits itself to only the sources of data which are readily available in mobile phones in use today (e.g., GPS).

Similarity Metric. The similarity measurement computed by Hapori uses FINE [4], Fisher Information Non-parametric Embedding. This technique that takes high-dimensional data representing different classes (e.g., different document types) and computes a low-dimensional space based on the probability density functions (PDFs) of the original high-dimensional data for each class. In this low-dimensional space the euclidean distance between the data points representing different classes approximates an information distance (e.g., fisher information) between the original PDFs. This technique is used successfully in recognizing different objects from images, clustering different documents types and classifying human poses from a video sequence. In the Evaluation section, we show that this approach is effective in clustering people with common POI preferences based on readily available measurements of their behavior.

To create the similarity metric the POI decisions from the community are projected into a feature space defined by the similarity features discussed in the previous subsection. For each person a multi-variate PDF is computed based on the POI decisions they contributed to the system. Similarity between people is computed based on the Fisher Information distance [2] between their respective PDFs. We use FINE to approximate these same distances using Hellinger distance [2] but within a lower dimensional space where every person is represented not by a data point for each POI decision but by a single data point. During this process off-the-shelf Multidimensional Scaling [2] is used to perform the dimensionality reduction. The resulting lower dimensional space represents the similarity metric. Between each single point (which represents a single person) the euclidean distance to another point (of another person) is proportional to the original Fisher Information distances from the initial feature space in which all the POI decisions are represented.

Hapori does not use the similarity metric directly, instead the relative coordinates of each data point representing each person become additional features of every POI decision contributed by the person. The addition of these features allow the model to represent POI decisions that have *identical* contextual feature values but yet correspond to different POIs by capturing behavioral differences between the people who make the decision. We use features based on the relative coordinates in the similarity space to avoid making hard assignments of people to specific groups. In practice the behavior of people does not allow them to be neatly placed into a fixed set of groups. Our design allows behavioral tendencies to be captured more naturally, relative coordinates represent how people are similar, to varying degrees, with more than one cluster of people at the same time.

Learn POI Category Relevance Metrics

One of the reasons local search is such a difficult problem is because the criteria for relevancy is difficult to specify and parameterize. Intuitively, it is possible to identify factors that alter the appeal of a POI to an individual, such as, personal preferences, ease of transportation, reputation, popularity. However, it is difficult to formalize these intuitions into a criteria to rank potential POIs in response to a query. This problem is even more difficult given that different POI categories each require their own specific criteria, for example, the most important factors in selecting evening entertainment will be different to those factors important to selecting places to shop. Hapori frames the problem of determining relevance metrics to rank potential POIs as a supervised distance metric learning problem. The general objective of this class of problem is to learn a transformation of a feature space to maximize classification performance. Hapori sets up the classification problem as follows. A feature space is used containing all the context and community similarity features each as dimensions. All POI decisions from people are represented in this space by their respective feature vector with the actual ground-truth POIs selected during these POI decisions acting as a label. The learning problem relates to correctly labeling a new unknown data point based on its features and the examples provided by the POI decisions from the community. In essence, the metric learner computes a transformation of the feature space to cluster those POI decisions (data points) that have identical labels. Figure 4 shows this fusion of context and similarity feature spaces occurring and the application of the metric learner to find an appropriate transformation. The transformation of the feature space rescales those dimensions that have the ability to discriminate those data points with different labels to maximize their benefit. Dimensions that have little significance to discriminate between labels of data points will be rescaled to near zero.

Rescaling feature dimensions has the effect of emphasizing or de-emphasizing components of the POI selection process; for example, if weather and time are a dominant aspect in the selection of a type of POI then similarity of weather conditions and temporal patterns should influence the rankings of POI candidates relative to the power of other features. Hapori uses LMNN [16], Large Margin Nearest Neighbor, a distance metric learner which maximizes k-nearest neighbor classification performance (the classification method Hapori uses to rank POI which we describe later in this section) by formulating the problem as a semi-definite program. Unlike other distance metric learners LMNN does not try to globally cluster all data points of the same label together. Rather, it will form local clusters of the same class (i.e., POI), allowing multiple clusters of data points to exist that are associated with the same POI. This is an important property given the requirements of our model. Each of these different clusters conceptually represents a different set of conditions under which the POI could be suitable. For example, a cafe can be an excellent place to go for breakfast and late at night be an equally suitable place to go for a cocktail. This would have different cafe cliental and temporal patterns and therefore is best modeled as two independent clusters. The final feature space learned by LMNN represents the real patterns of POI selection, as demonstrated by the community, in terms of the features we use to model POI decisions.

POI Category Classifier Training

Hapori learns a completely new metric for every POI category it supports. Only the fraction of POI decisions that result in a POI of the same category are used within the category specific models. When a query is provided to Hapori it is evaluated only using the model that is associated with the POI category specified by the user. The community behavior dictates the importance of different features during the POI ranking phase. Repeating these steps allows Hapori to incorporate changes in community attitudes and physical changes in the area by simply using new POI decision data from the community. We use a simple k-nearest neighbor (KNN) classifier, where the classification decision is based on the most frequent label within a distance radius surrounding the unlabeled data point. Distance is determined by the metric learned by LMNN. We use KNN because it is effective given how the learning problem is set up. However, any one of a number of different classification techniques can be used (e.g., SVM [2]). The two requirements being that it is a supervised learner and can estimate the probability of multiple potential classes (i.e., POIs) based on a new vector of contextual and behavioral features (i.e., query).

Apply POI Category Classifier

The off-line model training process discussed above is performed prior to user query processing. In our current design, a mobile client application is required to provide a general POI category to restrict the search of candidate POIs. The

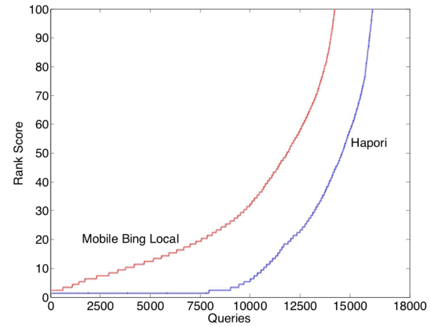


Figure 5. Rank Score comparison between Mobile Bing Local and Hapori

design of Hapori allows for flexibly defined POI categories. A category can be quite broad, such as, a category for recreational activities containing many different POIs (e.g., golf courses, jogging tracks, bowling alleys, karaoke bars). Similarly, categories can be quite narrow, such as, a category of different types of restaurants (e.g., italian, indian, chinese, vegan). A local search client provides this category selection along with features required by the POI preference model. These features are computed based on either: the instantaneous measurements from on-board sensors (e.g., location), longitudinal measurements of the user behavior (e.g., PDFs of temporal patterns when the person uses their device, or POI preferences from the application); and indirectly gathered data, such as, weather and traffic conditions based on the location and time. The user selected POI category determines the POI relevancy metric and the associated model to use when ranking POIs. One complication is that although we use KNN classification, if unmodified it only returns a single prediction. However, Hapori needs to output a relevance ordered list of POIs that match the query. To provide this, we make a simple modification to KNN. After each prediction all the training examples of the predicted label are temporarily removed before KNN is applied again. Each repetition produces the next most probable, and therefore relevant, POI based on the query.

EVALUATION

In this section, we evaluate a prototype implementation of the Hapori framework using real search query streams extracted from the mobile search logs of Mobile Bing Local [11]. Our evaluation aims to: i) Quantify how relevant the results generated by Hapori are and compare its performance to that of Mobile Bing Local and ii) Quantify the impact of the individual context and behavioral similarity parameters modeled by Hapori on the overall relevance performance.

Experimental Methodology

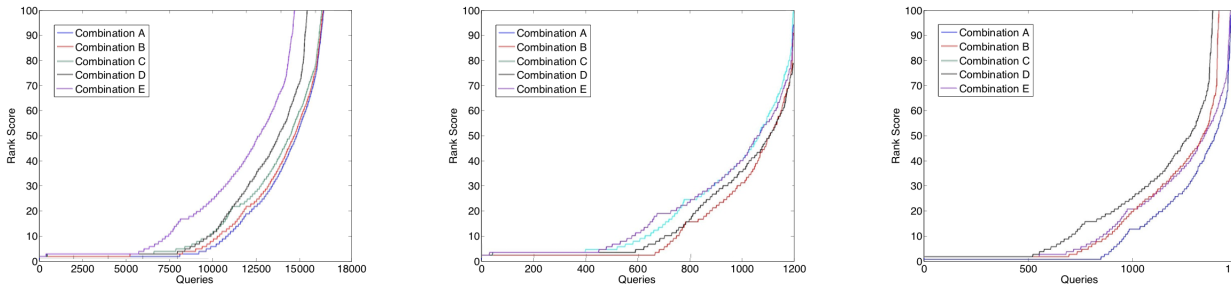
We study the effectiveness of Hapori using the same 6-month mobile local search logs discussed earlier in this paper. For the experiments we use a data set containing approximately 4,000 unique POIs and 80,000 queries generated by 11,000 different users. This data comes from a community of users in the Seattle, WA area and spans a six month period from January to July 2009. Hapori assumes a simple categorical query model in which users provide the system with a request based on a fixed set of POI categories, such as: 'restaurants', 'men's clothing', 'recreational activities' and

	Rank Score	
	Hapori	Mobile Bing Local
Tourist Activities	2.6	4.3
Indian Restaurants	3.5	9.2
Mens Apparel	7.8	13.2

Table 1. Rank Score comparison using narrow POI categories

	Rank Score	
	Hapori	Mobile Bing Local
Recreational Activities	8.5	20.3
Restaurants and General Food	6.1	31.2
Shopping and Services	9.3	17.9

Table 2. Rank Score comparison using broad POI categories



(a) All Categories (b) Fast Food Category (c) Outdoor Recreation Category

Figure 6. Feature sensitivity for all queries and within two specific POI categories. In these Figures the key labels refer to features being withheld during the experiment, the labels have the following interpretation. Combination A: All features. Combination B: Combination A without weather. Combination C: Combination B without community similarity. Combination D: Combination C without popularity. Combination E: Combination D without time.

so on. We ignore searches for a specific POI such as 'starbucks' because searching for specific POIs is handled efficiently by existing local search services. As a result, the evaluation focuses on categorial queries. Categories may be either broad (e.g., recreation) or narrow (e.g., indian restaurants). The Hapori prototype system uses a POI category hierarchy available from Mobile Bing Local. In our experiments we use a representative stream of local search queries where we extract all the categorical queries (80,000 queries in total). For each of these queries we know the query term as well as the actual POI that is clicked after the query is submitted. The clicked POI for every query serves as the ground truth for our evaluation. 75% of the queries are used to train the POI preference model as discussed in the previous section. The remaining 25% of the queries (20,000 queries in total) are used as our test data query stream. To quantify the ability of Hapori to return relevant search results we replay these 20,000 categorical queries where we record the query category, the user selected POI and other contextual information Hapori requires (e.g., location, time, weather and a collection of previous user queries and POI selections). We evaluate the relevance of query replies using a metric called *rank score, which is the position of the ground-truth POI selection within the ranked list of POI the search service provides*. Ideally all queries will have a low rank score with the ground-truth POI near the top of the list. To provide a benchmark to compare the performance of Hapori we replay the same 20,000 queries against Mobile Bing Local [11]. Each time we present rank score in this section the query stream is sorted, for each service, based on the rank score for each query. While this prevents using a Figure to directly compare a specific query between the two services it allows the comparison of aggregate performance between Hapori and Mobile Bing Local to be very clear. We use the rank score metric over more commonly used information retrieval metrics (e.g., Normalized Discounted Cumulative Gain [8]) because none of these are yet accepted as being appropriate for local search queries. Standard metrics that are based on web searches are ill-suited to cope with estimating relevance for local search requests. This is because they do not have the necessary context associated with the query (e.g., location etc.) nor a ground-truth POI selection. Irrespective of the evaluation metric there are fine-grain user satisfaction issues that analysis based on replaying queries can not capture. In the future we expect to examine such issues with small scale user studies to complement the results of the rank score based analysis we present here.

POI Model Performance

Figure 5 shows the rank score across the entire 20,000 categorical queries for both Hapori and Mobile Bing Local [11]. This result combines different types of POI categories to give an appreciation of overall relevance. Mobile Bing Local is able to display the business that is clicked by the user in the first ten search results for approximately 3,000 queries. Hapori achieves the same performance for approximately 12,000 queries; an improvement of 4x. Mobile Bing Local displays the business that is clicked by the user in the first 2 search results for approximately 900 queries, whereas Hapori achieves the same performance for approximately 9,000 queries; an improvement of 10x. The results show the benefit of using contextual and community based patterns of POI selection. Hapori is able to significantly outperform Mobile Bing Local by leveraging the user and query context to provide a more dynamic set of search results. By ranking businesses based on who is submitting the query, the time that the query is submitted, the weather conditions at the user's location and the popularity of nearby businesses, Hapori is able to better understand the decision process of real people and therefore better predict what the user actually wants to find. Hapori also benefits from having a specific metric for each POI category which allows Hapori to rank potential POIs based on criteria tuned for the different needs of each category. The Figure 5 does not show the category level differences because all categories are combined. Table 1 and Table 2 show the mean rank score of queries from six different POI categories. Three of these are narrow categories that are fairly specific as to the type of POI, while the other three categories are much broader. Hapori delivers a better mean rank score for all categories in comparison to Mobile Bing Local. Some categories are very broad, for instance more than 1500 POIs are considered within the 'restaurants and general food' category, which makes the large rank score within this category more understandable, others, such as, 'indian restaurants' only have 93 different POIs which makes the 9.2 rank score less impressive. Hapori outperforms the baseline search service by different degrees depending on the category. The variation of out performance is dependent on what degree the POI category benefits from the additional factors incorporated into the selection process. For instance, the 'recreational activities' and 'restaurants and general food' are sensitive to the weather, popularity and temporal features. The large improvement is due to these important factors being incorporated effectively into the search process. In contrast, 'shopping and services' and 'men's apparel' show little benefit from these same features. However, these two shopping categories offer improved performance through personalization and use of community similarity.

Feature Sensitivity

To understand the impact of the variety of different types of features used by Hapori we group together the different features we use into feature families, these being: popularity, spacial, temporal, weather and community similarity. We then perform sensitivity analysis on rank score with different query loads, as illustrated in Figure 6. In this experiment, we repeatedly replay the same query load but each time remove one of the feature families. The amount by which the rank score decreases indicates how important the feature family is to generating the rank score. All three subfigures in Figure 6 show the same baseline result which uses all feature families, this is the line on the furthest lower right indicating the curve has the best rank score performance. Figure 6(a) shows feature sensitivity across all the queries. We find that the most significant features in decreasing order are: temporal features, community similarity features, popularity and weather. Temporal features contribute significantly to rank score with the average rank score increasing by approximately 11 when temporal features are added back to the system. This has a greater benefit than location and weather features with community similarity being responsible for a change in rank score of about half that of temporal features. The large impact of temporal features on people's choices is intuitive and we find time is critical for all categories.

Figures 6(b) and 6(c) repeat the same experiment but for two specific POI categories i.e., fast food and outdoor recreation. By examining two POI categories in isolation we observe the variation in sensitivity to different features across different categories. For the POI category 'outdoor recreation' the weather feature family becomes the second most important influencer. Community similarity is dominant here because 'outdoor recreation' is sensitive to personal preference. In contrast, Figure 6(b) shows an example of a category in which many of the features of Hapori are not useful

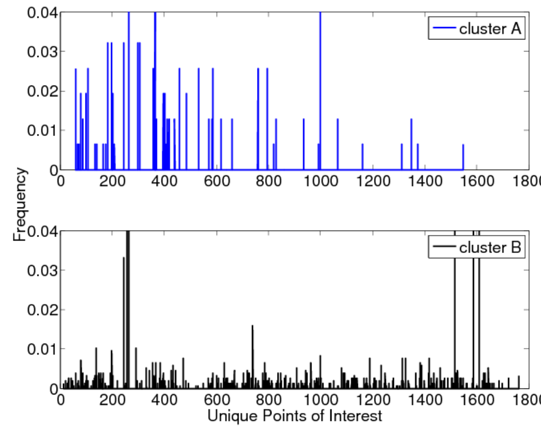


Figure 7. POI bias within groups of people clustered using behavioral similarity

in effecting rank score. Figure 6(b) shows different feature families are tightly bunched. This indicates that none of the feature families play a dominate role in rank score performance. The 'fast food' category is fairly homogenous, for example, the weather and temporal patterns effect the majority of the POIs in this category in similar ways so little is gained by its inclusion in the process. Moreover, the features based on similarity do not help significantly here since a large fraction of the population have weak preferences in terms of the type of fast food or the particular fast food vendor used. This analysis demonstrates the strong need for POI category specific metrics. Our results on feature sensitivity consider only a single community. We leave for future work the study of multiple communities that would identify how much inter-community variation in feature sensitivity exists. Nevertheless it is clear from our results that one size does not fit all when it comes to POI ranking.

Community Similarity Results

Evaluating the performance of community similarity is challenging given the anonymization of data. In the previous subsection we saw community similarity features have strong discriminative power, particularly for certain POI categories. However, ideally the effectiveness of the clustering of people during this process could be checked given the availability of additional meta-data about even a fraction of the people. With such 'out-of-band' data the clustering of people could be verified to check it is driven by expected similarity within latent behavioral characteristics (e.g., similar demographics, similar attitudes, preferences towards POIs). Without such information, we must rely on probing the Hapori model to test the power of community similarity. Figure 7 shows the result from an experiment to verify that community similarity is indeed an important performance consideration. The objective is to see if people who are dissimilar based on these community similarity features are in fact selecting different POI under the same context. In this Figure the user population is split into groups based on their similarity metric scores. This is done using simple k-means clustering of the community similarity feature values. Figure 7 shows the distribution of POI selections by two groups based on assignments we find by applying k-means clustering. This Figure shows virtually none of the POIs are popular with both clusters and in some cases specific POI with high user click-throughs for one cluster are found to have no user interest from the other cluster. We observe the distribution of POIs are distinctly different from each cluster even though the context is the same. In one cluster there is a wide dispersal in the distribution spread over a large fraction of all POIs. In contrast the other cluster has a much narrower focus in the distribution. These observations are inline with the expectations of how the similarity mechanism is functioning.

RELATED WORK

Surprisingly little is known about local search, with few proposed systems or analysis. Our log analysis represents to the best of our knowledge some of the most interesting local search specific characterization performed today. It complements well the small but growing body of work on web focused search from mobile devices (e.g.,[7]). These papers routinely report that for mobile devices the typical user experience is poor when performing general web search. In recognition of this researchers are pursuing approaches to improve mobile search (e.g., [17]). However very few of these proposals are applicable to the problem we address with Hapori since local search, the problem of finding personally relevant POIs, has very different characteristics compared to those found in web searching from a mobile device.

The benefit of personalization for search, particularly to overcome poorly specified user search intent has been an active area of investigation [15] for desktop web search. Researchers have sought to not only improve results based on data from the user in isolation but from communities of desktop web searchers [14]. However, this body of work provides limited guidance as to how to apply personalization and leverage the community effectively for a local search service rather than desktop web search. A promising source of information looks to be query logs, which researchers are beginning to explore. For instance, [6] uses query logs to improve local search results however the author's propose to essentially expose these logs to other users within a navigable map interface. This leaves all the work in the hands of the users as to how to interpret this community sourced information. In [9] a more transparent method of identifying user preferences for POIs is explored. User POI preference is shown to be correlated to how often they visit these places.

Hapori has commonalities with systems that provide recommendations [13] to people such as tourists or shoppers. In each of these cases the target user often does not know precisely what they are searching for and lack the knowledge needed to provide a fully specified query. One example, MovieLens Unplugged [12], provides movie recommendations while on the go. Such systems fall into the category of collaborative filters. Conceptually collaborative filters and our use of community similarity have parallels in that both attempt to leverage a collection of people to improve a suggestion provided to the user. However, our novel approach to modeling user intent and preferences as captured by a combination of context and behavioral similarity reaches beyond proposed collaborative filtering systems.

Our system heavily exploits context information and in this respect Hapori is building upon a long line of projects such as the pioneering systems of: Cyberguide [1] and GUIDE [5], which developed location-based intelligent guides and recommendation systems. We do not claim to be the first to identify the value of context within a search service, many have already been proposed (e.g., [10]). Still, Hapori is clearly novel in how it uses context to capture the important influences that impact the POI choices of people as a key component within a community guided local search service.

CONCLUSION

We believe that a major transformation of local search services is underway. Local search services are shifting their goal from answering static questions,such as, 'Where is the closest Dennys?' to answering more dynamic questions, such as, 'Where is a little martini bar around here, that is popular with people like myself and that has live jazz?'. The Hapori search service introduced in this paper, takes the first two steps in this direction. First, it demonstrates the effect that different query and personal context parameters have on the way mobile users make their POI selections, using a 6-month long search query stream from a major, commercially available search engine, Mobile Bing Local. Second, Hapori is the first local search service that leverages contextual factors, personal preferences as well as similarity across mobile users to deliver dynamic, highly contextual and personalized information in response to queries. Under Hapori responses to local search queries are generated based on a model of how *real* people make their own personal POI selections.

REFERENCES

1. G. D. Abowd, et al. Cyberguide: a mobile context-aware tour guide. *Wirel. Netw.*, 3(5):421–433, 1997.
2. C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, 2006.
3. A. T. Campbell, et al. The rise of people-centric sensing. *IEEE Internet Computing*, 12:12–21, 2008.
4. K. M. Carter, et al. Fine: Fisher information nonparametric embedding. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(11):2093–2098, 2009.
5. K. Cheverst, et al. Developing a context-aware electronic tourist guide: some issues and experiences. In *CHI '00: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 17–24, New York, NY, 2000. ACM.
6. K. Church and B. Smyth. Who, what, where & when: a new approach to mobile search. In *IUI '08: Proceedings of the 13th international conference on Intelligent user interfaces*, pages 309–312, New York, NY, 2008. ACM.
7. K. Church, et al. Mobile information access: A study of emerging search behavior on the mobile internet. *ACM Trans. Web*, 1(1):4, 2007.
8. B. Croft, D. Metzler, and T. Strohman. *Search Engines: Information Retrieval in Practice*. Addison-Wesley Publishing Company, USA, 2009.
9. J. Froehlich, et al. Voting with your feet: An investigative study of the relationship between place visit behavior and preference. In *Ubicomp*, volume 4206 of *Lecture Notes in Computer Science*, pages 333–350. Springer, 2006.
10. S. Hattori, T. Tezuka, and K. Tanaka. Context-aware query refinement for mobile web search. In *SAINT-W '07: Proceedings of the 2007 International Symposium on Applications and the Internet Workshops*, page 15, Washington, DC, 2007. IEEE Computer Society.
11. Microsoft. Mobile Bing Local. `http://m.bing.com/`.
12. B. N. Miller, et al. Movielens unplugged: experiences with an occasionally connected recommender system. In *IUI '03: Proceedings of the 8th international conference on Intelligent user interfaces*, pages 263–266, New York, NY, 2003.
13. P. Resnick and H. R. Varian. Recommender systems. *Commun. ACM*, 40(3):56–58, 1997.
14. B. Smyth, et al. Exploiting query repetition and regularity in an adaptive community-based web search engine. *User Modeling and User-Adapted Interaction*, 14(5):383–423, 2005.
15. J. Teevan, S. T. Dumais, and E. Horvitz. Personalizing search via automated analysis of interests and activities. In *SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 449–456, New York, NY, 2005. ACM.
16. K. Q. Weinberger and L. K. Saul. Distance metric learning for large margin nearest neighbor classification. *J. Mach. Learn. Res.*, 10:207–244, 2009.
17. J. Yi, F. Maghoul, and J. Pedersen. Deciphering mobile search patterns: a study of yahoo! mobile search queries. In *WWW '08: Proceeding of the 17th international conference on World Wide Web*, pages 257–266, New York, NY, 2008. ACM.