

# Inter-Category Variation in Location Search

Chia-Jung Lee<sup>†</sup>, Nick Craswell<sup>‡</sup> and Vanessa Murdock<sup>‡</sup>

<sup>†</sup>College of Information and Computer Sciences, University of Massachusetts Amherst, MA, USA

<sup>‡</sup>Microsoft, Redmond, WA, USA

cjlee@cs.umass.edu, {nickcr, vanmur}@microsoft.com

## ABSTRACT

When searching for place entities such as businesses or points of interest, the desired place may be close (finding the nearest ATM) or far away (finding a hotel in another city). Understanding the role of distance in predicting user interests can guide the design of location search and recommendation systems. We analyze a large dataset of location searches on GPS-enabled mobile devices with 15 location categories. We model user-location distance based on raw geographic distance (kilometers) and intervening opportunities ( $n$ th closest). Both models are helpful in predicting user interests, with the intervening opportunity model performing somewhat better. We find significant inter-category variation. For instance, the closest movie theater is selected in 17.7% of cases, while the closest restaurant in only 2.1% of cases. Overall, we recommend taking category information into account when modeling location preferences of users in search and recommendation systems.

## Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval

## Keywords

Mobile local search; category; rank distance; cross entropy

## 1. INTRODUCTION

Users interested in finding a location entity are increasingly using mobile devices to search, rather than desktop machines<sup>1</sup>. On mobile devices, the user's location context is available with greater precision<sup>2</sup>, allowing it to answer questions such as *What is the closest movie theater?* By

<sup>1</sup><http://searchengineland.com/study-consumers-increasingly-turn-mobile-varied-media-combinations-last-mile-200283> visited February 2015

<sup>2</sup>A median accuracy of 75 meters is reported in [12]

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org).

SIGIR'15, August 09 - 13, 2015, Santiago, Chile

Copyright is held by the owner/author(s). Publication rights licensed to ACM.

ACM 978-1-4503-3621-5/15/08 ...\$15.00

DOI: <http://dx.doi.org/10.1145/2766462.2767797>.

contrast, the location of a desktop machine is inferred from its IP address, giving a city-level estimate. This would allow us to list which movie theaters are in the city, but not which is closest to the user. Given this improved location context, it is possible to study specific scenarios such as reranking search results [7] or identifying the specific place a user is currently visiting [12]. While such advances in specific problem areas are valuable, they do not give us an overview of location intent on these new devices. Insights into location intent can guide our design for location-based applications, machine learning systems and evaluation metrics.

Geolocated mobile search logs from a general-purpose search engine provide an opportunity to study what types of location entities users find and select, and how this relates to their location context. Location context includes the location of the user, the distance of the user from the relevant entity, and the density of alternative entities near the user. The relationship between these aspects is complex, and may vary according to the category of entity. For example, if the density of nearby hotels and nearby ATMs is similar, that does not mean that users who search for the nearest ATM will also search for the nearest hotel. Motivated by this intuition, this study focuses on modeling user-entity distance and inter-category differences in location preference.

Our predictive models are based on raw geographic distance (*How many meters is the ATM from me?*), and intervening opportunities (*How many other ATMs are closer?*). We make minimal assumptions about how an end-to-end system would work, by for example using querying, mapping and ranking interfaces to narrow down the search. We analyse user needs in a system-agnostic manner.

Aggregating data on local intent is a challenging problem. Our dataset is from the general-purpose Bing search engine. During the time of our data collection, the system returns up to three location results in the first instance, then the user can opt to view more. We do not employ human relevance assessors to judge the quality of the results, because we do not wish to rely on a third party to understand location intent. Instead we employ information on which results were selected in the touch-based interface, on the assumption that selected results are of interest. Unlike previous studies [7] we do not simply rerank the results to put the clicked item higher. Although reranking offers some benefit to users, it does not give a clear picture of location intent characterization and in-context information needs. For these reasons, we aggregate over queries for studying distance and category, and use cross entropy to evaluate predicted click distributions over location entities.

Cross entropy has been used to characterize the amount of information in personalized search signals [9], and to evaluate click position bias in search results [3]. Similarly, we use cross entropy to measure the predictability of distance models for different business categories. Our results suggest significant inter-category variation, with the intervening opportunity model performing better in most cases.

## 2. RELATED WORK

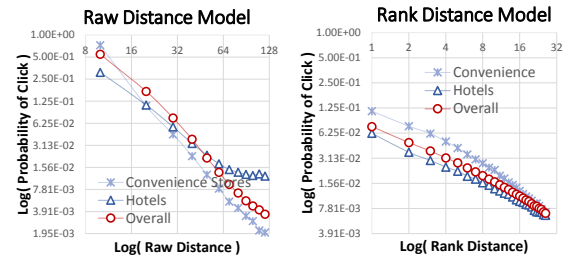
Mobile search is a relatively new field because mobile phones were not as commonly used for search prior to 2007. There are some early studies of mobile information seeking behavior [2, 4], but they do not relate directly to distance modeling or evaluation. Lymberopoulos et al. [8] predicted user clicks on mobile devices by incorporating location at different levels of granularity (e.g., zip code or state), and suggested the state-level models achieves higher accuracy than a US-wide model. Lv et al. [7] noted that the click-through rate of a business in local search sublinearly decreases with the distance to the user, which is evaluated using a click-based reranking framework. Berberich et al. [1] leveraged direction requests from an online mapping service to determine a reasonable travel distance for a user to a business, along with the business density in the area surrounding the business. We consider a superset of these scenarios, since we gather data from a general-purpose system that allows for driving directions, phone calls and informational intent.

One key exploration is to study the notion of intervening opportunities in the context of local search on mobile devices. In 1940, Stouffer [13] proposed that whether a person is willing to move to another place is dependent not on the geographic distance, but on the number of comparable alternatives along the way. This mobility model explained user movements from one neighborhood to another in the city of Cleveland. Our work does not make claims about the movements of people in a city; rather we reference his characterization to patronize a business or visit a point of interest. We adopt Stouffer’s proposal that a person’s willingness to visit a business is related to the number of comparable intervening businesses available to that person, which we denote as *rank distance*. Noulas et al. [10] used intervening opportunities to model Foursquare check-ins, finding that the probability of transitioning to a new location is inversely proportional to a power of its rank. Kumar et al. [5] used a combination of distance and rank distance to model geographic choice. They tested this using driving directions data with a source and destination address, taking cases where the destination address is a restaurant and interpreting the source address as the user’s location.

We base our evaluation approach on cross entropy. Entropy has been used widely to measure the degree of predictability of a probability distribution. Mei and Church [9], for instance, estimated the size of the search space of the web using entropy, and further computed the conditional entropy to characterize the contribution of personalization in web search. Craswell et al. [3] compared cross entropy among various models of click position bias in a manner that is agnostic to the model itself.

## 3. EVALUATING DISTANCE MODELS

*Raw distance model.* Given GPS coordinates, the most intuitive model is to directly measure the raw physical dis-



**Figure 1: The log-log plot for the probability distribution of clicks as a function of raw (left) and rank (right) distance. The distributions of categories Hotels and Convenience stores differ more in the raw distance model than the rank distance model.**

tance between users and places. We compute this using the great circle distance<sup>3</sup>. To avoid false assumptions, we opt for a non-parametric approach that counts the number of times entities are clicked given they were  $x$  kilometers away from users. In the training data, we discretize the raw distance and record the probability of click at each tick. We test different bucketing sizes including 1, 5 and 10 kilometers.

*Rank distance model.* For a place  $v$ , we define its rank distance as the number of places  $w$  that are in the same category as  $v$  but are closer to a user  $u$  than  $v$  is. That is,  $rank_u(v) = |w : d(u, w) < d(u, v)| + 1$ . We count the number of times an entity is clicked at each rank, based on which the probability of click is calculated.

An important difference between rank distance and raw distance is related to business density. Users in a dense area, with many in-category alternatives, may tend to have lower raw distance and users in a sparse area will have higher raw distance. This makes it difficult for a single raw distance model to explain users in both areas. By contrast, a rank distance model naturally predicts lower distances in dense areas because there are many nearby intervening opportunities. For both models, we consider all nearby entities in the same category as alternatives<sup>4</sup>, not just those that were retrieved by the engine. We build both models in a per-category manner. Figure 1 shows the resulting distance models for two example categories. In the raw distance model, the difference in click distributions is evident where we see the probability mass spreads out for the hotel entities, implying that distance may play a less critical role for hotel selection. Interestingly, such variation in categories appears less obvious in the rank distance model, suggesting that rank distance is more invariant of categories, resembling the findings of human mobility in cities [10].

*Cross Entropy.* Evaluating local search is a challenging problem. The existing Cranfield style evaluation [11] is less appropriate in local search. Unlike web documents, local businesses are represented by metadata, which is sparse. The location relevance heavily depends on various contextual factors such as current user location or time [6, 8], making the assessment non-trivial to recover for offline judges.

Another popular approach to evaluation is to utilize click data in a reranking experiment, where clicks are approxi-

<sup>3</sup>[http://en.wikipedia.org/wiki/Great-circle\\_distance](http://en.wikipedia.org/wiki/Great-circle_distance) visited February 2015

<sup>4</sup>We take into account entities within an area of  $9 \cdot 10^4 \text{ km}^2$  for a better trade-off between efficiency and coverage.

Category	Uniform	TopUni50	RawD10	RawD5	RawD1	RankD	$\Delta_{RankD}$
Movie Theater	13.33	7.54	5.78	5.74	5.73	5.61	57.92%
Convenience Store	16.49	9.87	8.02	7.70	7.58	7.18	56.46%
Department Store	15.84	9.64	7.48	7.38	7.35	7.14	54.90%
ATM	16.84	10.56	8.40	8.15	8.09	7.65	54.56%
Grocery	15.37	9.59	7.60	7.45	7.41	7.11	53.70%
Furniture Store	15.77	10.62	7.85	7.79	7.77	7.61	51.72%
Elementary School	16.08	11.87	8.53	8.37	8.30	8.05	49.94%
Auto Maintenance Repair	17.13	15.93	9.94	9.88	9.87	9.77	42.98%
Healthcare	16.49	15.38	9.72	9.68	9.67	9.68	41.28%
Apartment	16.48	15.05	10.23	10.09	9.96	9.81	40.48%
Government Office	16.68	15.42	10.02	9.98	9.97	9.94	40.40%
Church	17.28	17.02	10.95	10.89	10.88	10.76	37.72%
Lawyer	15.95	15.66	10.08	10.05	10.05	10.14	36.41%
Restaurant	19.23	20.67	14.23	14.08	14.01	13.93	27.53%
Hotels	15.89	16.86	13.39	13.34	13.30	13.28	16.41%
Mean-Model	16.32	13.45	9.48	9.37	9.33	9.18	43.77%

**Table 1: Cross entropy results for different categories and distance models.** RawD<sub>*x*</sub> denotes raw distance model with bucket size *x* and RankD is rank distance model.  $\Delta_{RankD}$  shows the reduction in cross entropy when using RankD compared to the uniform model. RankD performs the best by reducing the most uncertainty.

mated as relevance judgments [1, 8]. The general goal is to improve retrieval metrics such as nDCG@10 or MAP by reranking top results [7]. One problem with this is that the collected responses are biased towards the design of search systems (e.g., ranking or display algorithms). For mobile search, this becomes more of a concern since only a small number of local entities are shown to users. At the time of our data collection, users were presented with three or fewer results in 97% of cases, making most clicks land on only the top three results. Indeed, if we compute MAP as in [7], we get 0.84 for our log sample, leaving it unclear if reranking and evaluating using clicks on top results truly reflects systems’ retrieval ability.

To address these difficulties, we look to an information-theoretic approach to evaluation. In information theory, entropy refers to the average amount of information contained in a sample, while cross entropy measures the average number of bits needed to identify an event when a given distribution  $P'$  is assumed rather than the true distribution  $P$ . We use cross entropy to measure how well the distance models learned from historical data predict users’ responses in the test set. That is, to measure the quality of distance models, we compare, for a test query, the test distribution  $P$  against the distribution estimated from the training data  $P'$  based on the formula  $H_c = -\sum_e p(e) \log p'(e)$ , where  $e$  denotes a local entity. Thus, cross entropy favors a model if  $P$  and  $P'$  are more similar. Although click data is still inherently biased by the user’s choice of query and the ranking of the search system, by aggregating over many such queries and incorporating all nearby opportunities, we hope to better represent the distribution of possible clicks for a given user location and business category.

We assemble a test query to be a tuple containing a user  $u$ , a clicked entity  $e$  of category  $c$ , all surrounding entities  $e'$  of the same category  $c$  ( $e$  is considered as in  $e'$ ), and the geo-information of  $u$  and  $e'$ . Given such tuple, we are able to compute the raw and rank distances from the current user to the clicked entity in a test case. This requirement of exact one clicked entity ensures that we hold no assumptions about the probability of click in the test set (i.e.,  $P$ ). That is, the test distribution will assign probability 1 to the clicked entity and 0 to the rest. Since our evaluation is based

on the clicked entity, we must use the distance model to predict which entity will be clicked. We do this separately for each test case, using either the raw or rank distance model to firstly assign a probability to each entity in  $e'$  based on its distance from the user, then renormalizing to give a probability distribution over the set of those entities (i.e.,  $e'$ ).

## 4. EXPERIMENTAL RESULTS

### 4.1 Experimental Setup

*Data Collection.* We evaluate using Bing mobile search log of users in the U.S. The training data for building the distance models is sampled from a period of four months in 2014, and the evaluation set is sampled from the following two months. We complement the mobile search data with an entity repository that includes location information (i.e., latitude and longitude) and category information (i.e., whether an entity is a hotel or a convenience store etc). Our resulting test set contains approximately 1.1 million queries.

*Baseline Approaches.* The naive baseline assumes that no distance information is given between a place and a user, as is the case when users do not allow their location to be known. In this case, a uniform model assigns equal probability to every in-category place entity (Uniform in Table 1). To improve the pure uniformity, the second baseline assigns equal probability to the user’s  $k$  closest in-category entities with a total probability mass 0.99, with the remaining 0.01 uniformly distributed to the much larger number of more distant alternatives. We denote this as a top  $k$  uniform model, and we report the results for the 50 closest entities (TopUni50 in Table 1). These two approaches are used as comparison basis for the raw and rank distance models.

### 4.2 Cross Entropy Results

We take the average of the cross entropy of test queries in each category and show the results in Table 1. The naive uniform model gives an average cross entropy of 16.32, meaning that the search space for a category is approximately  $2^{16}$  when given no distance information. Under the top 50 nearby places, the average number of bits needed is reduced to 13.45, shrinking the search space about 8 times. The average cross entropy for user clicks is approximately 9.48

if we bucket the physical distance by 10 kilometers, which is a huge improvement compared to the top uniform model. The quantity becomes 9.37 and 9.33 respectively when using bucket sizes of 5 and 1 kilometers. This suggests that a finer granularity may lead to lower entropy but the improvement may saturate quickly. Finally, predicting user clicks with the rank distance model performs the best with the lowest cross entropy 9.18, which is a 10% reduction in the search space compared to the best performing raw distance model (i.e., at 1 km granularity). This suggests that using rank distance can explain users' selections towards local places better in our test set, despite that the notion of rank distance, in fact, discards the geographic distance information.

In addition to distance, the categories of entities play an important role for predicting user clicks. The cross entropy  $H_c$  of the uniform model is directly influenced by the number of entities in a category. The more entities a category contains, the higher  $H_c$  is since it is harder to predict which entities users may click without prior knowledge. In our dataset, the size of Restaurant category is the largest, resulting in the largest  $H_c$  19.23 compared to others.

Table 1 shows interestingly that  $H_c$  for the Hotel category drops, at relatively smaller scale, from 15.89 in the uniform model to 13.28 in the rank distance model, resulting only a reduction of 16.41% in uncertainty. On the contrary, when the searches were intended for entities of ATMs or Convenience stores,  $H_c$  drops significantly with more than 50% reduction in the rank distance model. This suggests that although the distance factor plays a strong role in predicting ATMs searches, the clicks on hotel entities are less explainable using only distance from the searching user.

For categories such as Department store or Furniture store, the reduction in  $H_c$  is evident in the rank distance model. This implies for queries with shopping intent, users may have a tendency to click entities that are closer to them. Entities of these categories may be more replaceable when the store items are not significantly different for the users<sup>5</sup>. People also have a tendency to search for nearby movie theaters, where using rank distance can account for 57.92% uncertainty. The category Elementary school is another example where the closer entities seem preferable, consistent with our intuition that children go to the elementary school in their neighborhood in most districts.

When are users more willing to travel farther even in the presence of entities that are nearby? Relatively speaking, for categories such as Healthcare, Lawyer and Auto Maintenance, the entropy reduction degree is smaller compared to ATM or Convenience store, but not as small as in Hotel. People may skip closer places and seek their more desired choices in the cost of travel. This can be because these entities usually provide services that require specific professional knowledge, and nearby places might not meet a user's exact need. As the cost increases, the overhead of travel may suppress the desire, which is different from the case of Hotel where distance plays even a less important role. Interestingly, we observe that searchers skipped closer intervening opportunities and clicked restaurants farther away. Indeed, people go to restaurants not only for food, but also for factors such as the atmosphere, the chef and nearby attractions or friends. The results for categories such as Church, Apartment and Government office also imply that using distance

<sup>5</sup>We note that there should be exceptions when users look for things only available in specific stores.

alone is not sufficient for predicting user clicks.

So far we focus our discussion on the comparison *within* the rank distance model. Similar trends can be observed if we compare results within the top uniform or the raw distance model, with varying degree in entropy reduction. It is interesting to note that in the top 50 uniform model, the assumption that closer entities should receive more clicks fails for the case of hotels and restaurants; the uncertainty in fact increases compared to pure uniformity.

## 5. CONCLUSION

This paper explored several ways that distance can be modeled in the context of mobile local search. Rather than merely reranking the top results, we considered intervening entities that were physically near the users, shedding light on user choices beyond those on the result page. The cross entropy results suggest that raw geographic and rank distance are both helpful, with the rank distance model outperforming in most cases. This suggests that the task of mobile local search resembles general human mobility in that it is influenced by intervening opportunities. Our analysis identifies significant inter-category variation. We find that locations that are highly replaceable (e.g., convenience stores) are more predictable based solely on distance, whereas user clicks on categories such as hotels and restaurants are dependent on complex factors. Overall, we recommend taking category information into account when modeling location preferences of users in search and recommendation systems.

## 6. REFERENCES

- [1] K. Berberich, A. C. Koenig, D. Lymberopoulos, and P. Zhao. Improving local search ranking through external logs. In *Proc. of SIGIR*, 2011.
- [2] 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), 2007.
- [3] N. Craswell, O. Zoeter, M. Taylor, and B. Ramsey. An experimental comparison of click position bias models. In *Proc. of WSDM*, 2008.
- [4] M. Kamvar and S. Baluja. Deciphering trends in mobile search. *Computer*, 40(8):58–62, 2007.
- [5] R. Kumar, M. Mahdian, B. Pang, A. Tomkins, and S. Vassilvitskii. Driven by food: Modeling geographic choice. In *Proc of WSDM*, 2015.
- [6] 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 *Proc. of UbiComp*, 2010.
- [7] Y. Lv, D. Lymberopoulos, and Q. Wu. An exploration of ranking heuristics in mobile local search. In *Proc. of SIGIR*, 2012.
- [8] D. Lymberopoulos, peixiang Zhao, C. Koenig, K. Berberich, and J. Liu. Location-aware click prediction in mobile local search. In *Proc. of CIKM*, 2011.
- [9] Q. Mei and K. Church. Entropy of search logs: How hard is search? with personalization? with backoff? In *Proc. of WSDM*, 2008.
- [10] A. Noulas, S. Scellato, R. Lambiotte, M. Pontil, and C. Mascolo. A tale of many cities: Universal patterns in human urban mobility. *PLoS One*, 7(5), 2012.
- [11] M. Sanderson. Test collection based evaluation of information retrieval systems. *Foundations and Trends in Information Retrieval*, 4(4):247–375, 2010.
- [12] B. Shaw, J. Shea, S. Sinha, and A. Hogue. Learning to rank for spatiotemporal search. In *Proc. of WSDM*, 2013.
- [13] S. A. Stouffer. Intervening opportunities: A theory relating mobility and distance. *American Sociological Review*, 5(6), 1940.