Driven by Food: Modeling Geographic Choice

Ravi Kumar   Mohammad Mahdian   Bo Pang   Andrew Tomkins   Sergei Vassilvitskii

Google
Mountain View, CA
{ravi.k53, mahdian, bopang42, atomkins, sergeiv}@gmail.com

ABSTRACT

In this work we study the dynamics of geographic choice, i.e., how users choose one from a set of objects in a geographic region. We postulate a model in which an object is selected from a slate of candidates with probability that depends on how far it is (*distance*) and how many closer alternatives exist (*rank*). Under a discrete choice formulation, we argue that there exists a factored form in which unknown functions of rank and distance may be combined to produce an accurate estimate of the likelihood that a user will select each alternative. We then learn these hidden functions and show that each can be closely approximated by an appropriately parameterized lognormal, even though the respective marginals look quite different. We give a theoretical justification to support the presence of lognormal distributions.

We then apply this framework to study restaurant choices in map search logs. We show that a four-parameter model based on combinations of lognormals has excellent performance at predicting restaurant choice, even compared to baseline models with access to the full (densely parameterized) marginal distribution for rank and distance. Finally, we show how this framework can be extended to simultaneously learn a per-restaurant quality score representing the residual likelihood of choice after distance and rank have been accounted for. We show that, compared to a per-place score that predicts likelihood without factoring out rank and distance, our score is a significantly better predictor of user quality judgments.

1. INTRODUCTION

We consider the problem of modeling geographic choice: how users choose one among a set of objects in a geographic region. We study this problem and postulate a general model by analyzing a large maps query log for directions to the single-largest consumer business type [38]: restaurants. More specifically, for each request for directions to a particular restaurant, we consider that restaurant as the *chosen* one among all possible alternatives in the region.

The list of important covariates on restaurant choice is long and well-studied. Restaurant choice can be affected by cuisine type, price, quality, and other venue-specific metadata. Breaking reviews or word of mouth may lead to local bursts of activity. Group social

dynamics may play a significant role. These factors are similar to ones affecting people's choices over various consumer products and have been relatively well-studied. But choices over geo-entities like restaurants must also take into account the actual physical locations of the candidate restaurants under consideration. The most obvious contribution comes from the distance between the user and a particular candidate. However, the likelihood of selecting a restaurant at a particular distance must also depend critically on the nature of the *other* candidates. A restaurant 5km from the user in Tokyo is vanishingly likely to be selected, while another restaurant at the same distance in Iowa may be the overwhelmingly most likely choice.

Our goal in this paper is to study how these different factors operate, rather than to optimize prediction accuracy by incorporating as many features as possible. While we do give specific functional forms for transformations by which these factors should be incorporated into general place recommendation systems, this work is not a proposal for a place recommender.

1.1 Distance, rank, and popularity

We now describe the three key factors we study in more detail. First, we consider the absolute *distance* to the restaurant. Second, to capture a candidate's role relative to alternatives, we consider the number of these alternatives that are closer to the user; we use *rank* to denote this number. Finally, after a detailed study of rank and distance, we incorporate a per-restaurant residual *popularity* score to improve the prediction; we believe this score has value in its own right as a measure of quality. The score may be viewed as the residual likelihood of an object to be chosen after distance and rank have been accounted for.

There is a significant body of work studying the dynamics of geographic choice. Much of this work focuses on the distance to each object, and in many cases falls under the rubric of "gravity laws," in which the desirability of an object at a particular distance is proportional to an inverse polynomial function of the distance; see the book [10]. We will show that predictions of geographic choice based on distance alone is insufficient.

Alternative proposals based on rank rather than distance, dating back to Stouffer's original proposals from 1940 [34], have shown improvements in predictive power. Noulas et al. [24] have observed a good fit between the rank of a destination and the probability of its selection, showing that the log of the probability drops inversely with the log of the rank of the object. In the context of restaurants, we observe the same form as theirs roughly from rank 150 and above, but for lower ranks, the marginal likelihood drops slower than what they observed. While there may be more sophisticated approaches to predicting based on distance, our baseline measures show rank alone outperforming distance alone. Consistent with previous work, we observe that rank, as an implicit measure of the competitive landscape, provides good predictions of choice.

However, the form of the rank marginal varies when we stratify by the density of restaurants in the area of the query, suggesting that predictions based solely on rank are limited in the accuracy they can attain. We provide evidence to suggest that human decision making involves an interplay of both rank and distance and that simply considering rank, either alone or bucketed by average density, will not suffice. Our proposed model captures this interplay with only four parameters, and significantly outperforms dense models based on the entire marginal distributions of rank or distance.

Thus, a natural approach to predict the likelihood of selecting a particular candidate restaurant is to use its rank, distance, or other features. One could imagine a function of rank, distance, cost, cuisine, etc., that returns the likelihood of a restaurant being selected. However, this approach is problematic, as the likelihood of selecting a restaurant depends on not just the restaurant's features but also the attractiveness of the alternatives. To address this issue, we propose a simple model based on modeling an intrinsic desirability of each option and normalizing across different alternatives via a discrete choice framework.

1.2 A discrete choice model

The appropriate formulation, therefore, is slightly more general: given a set of alternatives with their features, determine a distribution over the alternatives so as to maximize likelihood of the data. We derive such a distribution by making a simplifying assumption. R. Duncan Luce in 1959 [20] formulated the *Choice Axiom*, which states that a user's likelihood to select an object A over object B is unaffected by the nature of the other alternatives. While this axiom is likely not strictly true for geographic choice in general, it provides a good approximation that yields a valuable functional form for our analysis. Luce's Choice Axiom implies that each alternative i may be assigned a score s_i such that $\Pr[\text{user chooses } i] = s_i / \sum_j s_j$, i.e., the user's likelihood to choose i is proportional to s_i, but with a normalization factor that depends on the alternatives. Hence, an object with some set of features might be likely in one setting but unlikely in another, even though the score is unchanged, because the stronger alternatives of one setting provide a larger normalizing constant.

We must therefore determine the form of the score s_i of object i. To begin, we will express this as a function of the rank and distance of i with respect to the user; later, we will introduce the concept of i's overall quality. Consider a situation in which a user selects object i from a set of alternatives. Our challenge will lie in teasing apart the relative contribution of i's distance versus its rank, or some complex element of their interplay. We make a simplifying assumption that will be borne out by our analysis: the score s_i of object i, occurring with rank r at distance d, may be written as $s_i = f_r(r) \cdot f_d(d)$, i.e., the score of a object is the product of some unknown function f_r of its rank and some unknown function f_d of its distance. One may view this as a univariate function of rank capturing the rank-specific component of this mental process, multiplicatively modified by a univariate function of distance capturing the travel cost. This form is fairly general, though not universal.

Given this functional form, we turn to the techniques of discrete choice modeling to optimize the functions. As we have no preconceptions about the forms of f_r and f_d, we will quantize ranks and distances, then perform an optimization that will empirically estimate the value of f_r and f_d pointwise so as to optimize likelihood. We then study the form of the resulting pointwise estimates in order to model the shape and capture the aggregate behavior in a more parsimonious fashion. We detail the algorithms to perform the pointwise estimation optimally. Our approach has the pleasing property that employing the resulting functions in order to select objects will produce selections that simultaneously exactly reproduce the rank and distance marginals of the original data; it is not in general obvious how to attain this property without such a technique.[1]

1.3 Parametric forms

After our optimization, we have learned non-parametric functions of rank and distance that combine to give us a good approximation of likelihood. We fit a large number of parametric function families to the resulting data, and show that both distance and rank contributions are approximated well by lognormal univariates. Note that these functions of rank and distance are quite different from the marginal distributions in the data; these are instead hidden distributions that are not directly observable, but which emerge as the underlying factors that according to our model drive the process of selecting an object. The lognormal form for the hidden functions of both rank and distance was quite surprising to us, as the rank marginal does not have a good fit to a lognormal distribution. None of our earlier efforts to tease out the relationship between these factors led us to consider such a form. We present some discussion of the implications of this finding below.

We fit the shape parameters for the lognormal functions of rank and distance, resulting in a final likelihood model of choosing an object i with rank r_i and at distance d_i from a set C of alternatives:

$$\Pr[\text{user chooses } i \in C] = \frac{f_r(r_i) \cdot f_d(d_i)}{\sum_{j \in C} f_r(r_j) \cdot f_d(d_j)},$$

where

$$f_r(\cdot) \sim \text{LogNormal}(\mu_r, \sigma_r), \quad f_d(\cdot) \sim \text{LogNormal}(\mu_d, \sigma_d).$$

The parameters $\mu_r, \sigma_r, \mu_d, \sigma_d$ fully specify the model.

We use both the parametric and non-parametric versions of these imputed rank and distance functions, as well as a number of baselines, to compare the likelihoods generated by our model for the restaurant directions case, and show that these two functions outperform other natural alternatives. The model performs well over a variety of densities from urban to rural settings. We expand on this observation by providing some theoretical justification. Under one fairly weak assumption regarding the layout of places, and another commonly-held assumption regarding human perception of the relative cost of growing time intervals, these lognormal functions are in fact the optimal outcomes under an entropy maximization framework (see Section 6).

1.4 Residual as quality

Having established a concise understanding of aggregate choice given rank and distance, we turn to the question of incorporating a quality score for each individual object. We may expect that certain objects will be chosen more often than predicted by rank and distance alone; we would like to explain the residual (positive or negative) by a score representing "quality," or more accurately, whatever attributes of an object cause it to be visited more often than its physical location relative to its clientele would predict.

The machinery developed above applies naturally to this setting, but the optimization problem we must solve now includes a much larger number of parameters. In this new formulation, we assume that each object i is endowed with a quality w_i, and that the likelihood of selecting i is proportional to $f_r(r_i) \cdot f_d(d_i) \cdot w_i$, with

---

[1] Note that this result will hold in expectation if the optimizer is allowed to fit parameters for each individual rank and distance. In our experiments, we run the optimization on a quantized form, and hence recover the marginals for the quantized ranks and distances, which are less precise than the full marginals.

the usual normalization.[2] In this general case, as well as the simplified version above, the gradients of the likelihood function have a simple and easily interpretable form that we will describe later.

After the optimization, we find that for restaurants, the likelihood of selecting any individual restaurant has on average increased by 87% over the best model we had available without per-place scores. This figure is computed as the geometric mean of the increase in likelihood over all candidate places.

2. DATA

We collected a random sample of anonymized direction queries issued at Google Maps (`maps.google.com`) during a period of 170 days in 2012. Each direction query consisted of the timestamp, the query normalized to the form "from: `src` to: `dst`", and the distance of travel as computed by Google Maps. Note that `src` or `dst` can specify a location as a (`lat, lon`) pair or as an explicit address (e.g., `1600 Amphitheatre parkway, Mountain View, CA 94043`). When `src = (lat, lon)`, the direction request is more likely to have originated from a mobile device. Since it is common for users to keep updating directions en route, resulting in a series of decreasing distances for a single trip, the distance distribution in this type of request may not reflect the actual end-to-end distances people travel. Hence, unless otherwise specified, we restrict most of the analysis to directions where `src` is a piece of text (denoted `src = text`).

We obtained a database of business listings and restricted it to a large random subset of US/Canadian restaurants. We then extracted directions where `dst` is recognized (by the system) to be one of the restaurants in our subset.[3] The final dataset consists of 15,454,095 directions where `src = text`; this covered 410,488 restaurants. A small fraction of the traffic (~5%) appeared to be "just checking" (e.g., `src` and `dst` are oceans apart). If we consider 200km to be the threshold of driving distance for the restaurant intent, 405,889 restaurants have at least one directions query within 200km, and 232,230 have at least 10 qualifying queries.

3. RANK VS. DISTANCE

In this section we explore the roles of rank and distance in the geographic choice process, using the data described in Section 2. Intuitively, the rank, or the number of objects closer than a particular object, corresponds to the amount of competition faced by this object in the geographic area. In the context of our case study, we first study the impact of the rank of the restaurant on the decision process. We begin by showing that rank alone does not capture the richness and variability of users' actions.

In Figure 1(a) we show the global distribution of the rank of the restaurant visited by the user. Note that the empirical distribution can be approximated nearly perfectly with the decay proportional to $\log(1/\text{rank})$ for small ranks (under 150) and proportional to $1/\text{rank}$ for ranks above 150. The same phenomenon holds even as we zoom in on specific areas. Figure 1(b) shows the empirical rank distributions for four areas with different restaurant densities. Observe that the specific shapes of the curves differ for different regions; we will consider a generalization of this effect later.

3.1 The effect of density

We now refine our analysis to take into account the density of restaurants in the area of the query. We define the (restaurant) *den-*

---

[2]Note that we do not require a function $f_w(w_i)$, as w_i is inherently non-parametric—the w_i's *are* the function.

[3]The details of these recognition methods are beyond the scope of this work.


(a)

(b)
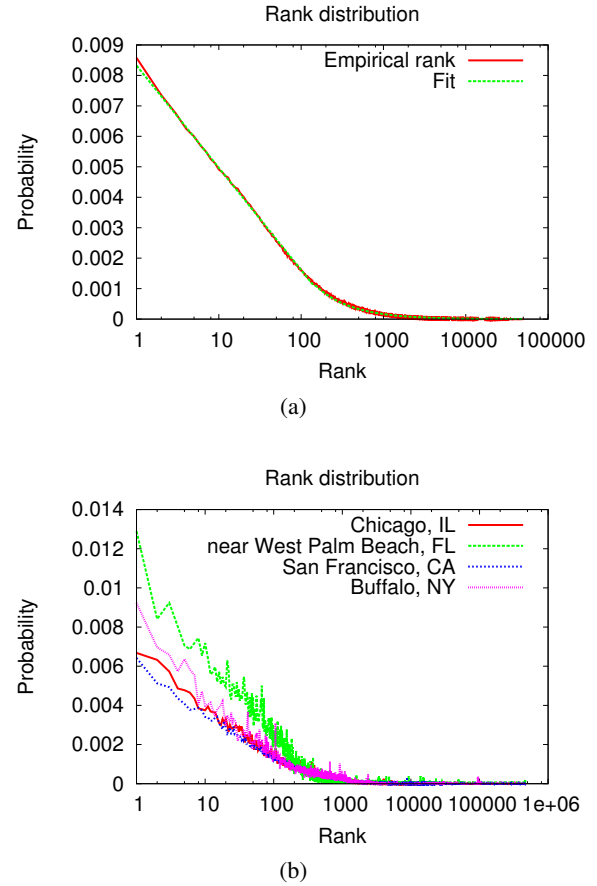
Figure 1: Overall rank distribution and rank distribution for a few urban areas.

sity (more specifically, average density at a rank) of the source location of a query as follows. For a given source location and a given rank r, we compute the distance d_r of the restaurant at rank r to the source location. The average density of the first r restaurants is r divided by the area of the circle with radius d_r, which is proportional to d_r^2; we define $\rho_r = r/d_r^2$. We compute ρ_r at $r = 100$ and use it as the measure for the density of the source location. We then bucket density values on log-scale and put each direction into the appropriate bucket according to its ρ_{100} value. For each density bucket, we can plot the empirical distribution of the rank of the restaurant visited as well as the empirical distribution of distance traveled. The results are shown in Figure 2.

Figure 2(a) shows that the empirical distributions of rank are well-separated for users originating from regions with different density values. Indeed, users in less dense regions (e.g., West Palm Beach, Florida) are much more likely to choose restaurants at lower ranks than those in denser regions (e.g., San Francisco, California). The distribution corresponding to the latter is therefore flatter when compared to that of the former. Hence rank alone, when conflated across a wide range of densities, misapprehends user choice significantly. On the other hand, Figure 2(b) shows that employing distance alone is just as bad. The distributions of distances traveled by users in dense and sparse regions are again well-separated, if not worse: users in dense regions do not need to travel long distances to have plenty of options and hence their distance distribution is correspondingly left-shifted.
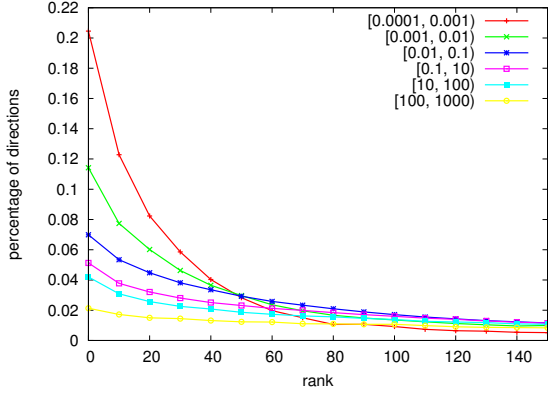
(a)
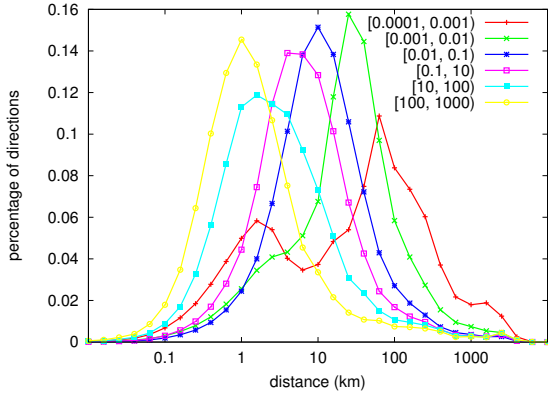
(b)

Figure 2: Distribution of rank and distance for different densities. Each curve corresponds to a set of directions from a source location whose density (ρ_{100}) is in the given range.

Note also that while most curves more or less follow the lognormal shape, the curve for the least dense region (red curve, $\rho_{100} \in [0.0001, 0.001)$) is bimodal. In comparison, the different curves in Figure 2(a) seem to correspond to functions with different parameters, but coming from the same family. This prompts us to ask the following question: can we describe the empirical observations with a set of rank distributions parameterized by density?

3.2 A closer look at density

We consider a region as *smooth* if ρ_r is a constant (relatively speaking) at least for reasonably small r (e.g., ≤ 200). More specifically, we consider a region encircling r restaurants as ϵ-smooth if for a small ϵ and a constant ρ, we have $|\rho_{r'} - \rho| \leq \epsilon$ for all $r' \leq r$. If the local neighborhood of a source is smooth, then we can capture its local density with a single parameter ρ rather than a family of ρ_r's. Hence we can model all directions originating from regions with density ρ with a single rank distribution.

Figure 3 plots ρ_{200} (y-axis) vs ρ_{100} (x-axis). If ρ_r were relatively constant for most regions encircling 200 restaurants, then most points would fall within a narrow band along the diagonal line of $y = x$. Figure 3(a) shows the heat map on a log-log scale (where the color corresponds to log of count for that cell). While this gives us the entire range, it compacts the data in a way that is hard to interpret. Figure 3(b) shows the heat map for density values between 1 and 10. Here we can clearly see that many points fall

(a) log scale   (b) zoom-in (linear scale)
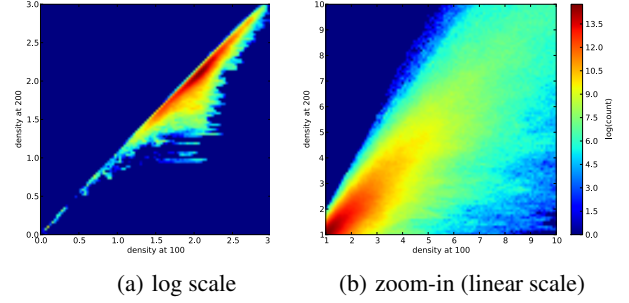
Figure 3: ρ_{200} (y-axis) vs ρ_{100} (x-axis).

outside of a close band around $y = x$. Indeed, if we compute the CDF for $|\rho_{200} - \rho_{100}|/\rho_{100}$, we observe that ρ_{200} is within 20% of ρ_{100} for only about half of the time. This phenomenon also repeats itself at smaller r: e.g., the CDF for $|\rho_{100} - \rho_{50}|/\rho_{50}$ has a similar shape.

One might suspect that the variability in the densities observed is due to the natural tendency of many restaurants to be located in groups close together; zoning ordinances further enforce this phenomenon. However, this hypothesis is not borne out in the data—while restaurants do occur in groups, we observed that these groups are (essentially) uniformly distributed as a function of the distance from a randomly placed user. Therefore, the cumulative distribution is not affected by the natural clustering of eateries.

These observations suggest that it is hard to capture the local density of a source location with a single parameter ρ, since objects may not be uniformly distributed geographically. Indeed, for the restaurant case, we could not find a simple analytic form to capture the relationship between rank and distance; we believe this is true for other types of objects as well. Instead, we propose a unified model of user choices that depends on both rank, distance, and a per-object quality measure.

4. A GEOGRAPHIC CHOICE MODEL

We saw above that multiple factors play a role in determining the probability a user will choose a particular geographic object (i.e., driving to a specific restaurant in the context of restaurants). The rank plays a role, as one is more likely to go to a closer object; the absolute distance is also crucial, as one is not likely to drive hours for it; and it is intuitive that some inherent quality of the object also plays a role. In this section we set up a framework to model how these different factors can be combined to give a prediction of a user's choice, and how we can learn about the quality of objects by estimating the underlying model parameters.

At the heart of our model is a discrete choice model that describes how an agent chooses among a set of alternatives. For this, we use a classical model due to [20], which is based on the following axiom: the relative probability of choosing between two options a and b is independent of the presence of other options. This axiom results in the following probabilistic model. Suppose that every object i has some total score s_i; this score can be a function of many factors. In particular, we consider a model where the score is a function of the distance between the user and the object, the rank of the object among all options available to the user, as well as some inherent metric, which we will refer to as *wonderfulness*. Presented with scores for all objects in a set C of alternatives,

$$\Pr[\text{user selects object } i] = \frac{s_i}{\sum_{j \in C} s_j}.$$

Given such a simple model of user behavior, we ask for the best set of scores s_i that explain the data. Assume we are given a set Q of observations about a user's choice. For each observation, we know the set of alternatives, as well as the alternative chosen by the user. In the context of restaurant direction queries, Q is the set of direction queries to restaurants; the restaurant to which query q is directed is denoted by $r(q)$, and the set of restaurants that the user could have queried for is denoted by $C(q)$. The score of a restaurant i for the query q is denoted by $s_{i,q}$. Given these, the likelihood of these observations in our model can be written as:

$$L(Q) = \prod_{q \in Q} \frac{s_{r(q),q}}{\sum_{j \in C(q)} s_{j,q}}.$$

The log likelihood, $\mathrm{LL}(Q)$ is then:

$$\mathrm{LL}(Q) = \sum_{q \in Q} \log s_{r(q),q} - \sum_{q \in Q} \log \left(\sum_{j \in C(q)} s_{j,q} \right).$$

In order to efficiently find the best scores, we show that $\mathrm{LL}(Q)$ is a concave function in the logarithm of all variables, and therefore gradient ascent on $\mathrm{LL}(Q)$ amounts to a natural multiplicative update step on all of the scores.

LEMMA 1. $\mathrm{LL}(Q)$ is a concave function of all of $\log s_{i,q}$'s simultaneously.

PROOF SKETCH. We show that the second partial derivatives are always non-negative. Let $z_{i,q} = \log s_{i,q}$. Then we can rewrite the likelihood as:

$$\mathrm{LL}(Q) = \sum_{q \in Q} z_{r(q),q} - \sum_{q \in Q} \log \left(\sum_{j \in C(q)} e^{z_{j,q}} \right).$$

Computing the first partial derivative, we have:

$$\frac{\partial \mathrm{LL}}{\partial z_{i,q}} = [i = r(q)] - \frac{e^{z_{i,q}}}{\sum_{j \in C(q)} e^{z_{j,q}}}. \tag{1}$$

The second partial with respect to another variable $z_{k,q'}$ is zero if $q' \neq q$. If $q' = q$, this second partial is equal to:

$$\frac{\partial^2 \mathrm{LL}}{\partial z_{i,q} \partial z_{k,q}} = \frac{e^{z_{i,q}} e^{z_{k,q}}}{\sum_{j \in C(q)} e^{z_{j,q}}}.$$

The rest of the proof follows from the well-known properties of log-partition functions [16]. $\square$

We note two simple, yet powerful consequences of the following concavity result: first, since LL is a concave function, it is possible to efficiently optimize it even when the variables $z_{i,q}$ are subject to additional linear constraints. For example, we can optimize the likelihood if $z_{i,q}$ is a linear combination of a rank factor, a distance factor, and a quality factor. This is precisely what we will use next.

Second, note that in Equation (1), the first term is exactly the number of times a particular object was selected. The second term is exactly the *expected* number of times the object should be selected, given the current setting of the weights. This means that the optimum can be characterized by equations that state that the expected marginal over any variable should be equal to the observed marginal.

ALGORITHM. Given the above observations, we can optimize the function using a gradient ascent algorithm. Let δ_i be the discrepancy between the expected and the observed value (i.e., the value from (1)). Then the gradient update is:

$$s_i^{\text{new}} = \eta \cdot e^{\delta_i} \cdot s_i^{\text{old}},$$

where η is the learning rate. This leads to a particularly simple algorithm: at every iteration, compute the discrepancy between the expected and the actual number of times each object was selected and use that to perform a multiplicative update on the weights. Renormalize the weights to keep the maximum weight at 1 and repeat until convergence.

4.1 Decomposing the scores

We have howed that it is easy to efficiently compute the best scores associated with each object via a simple multiplicative update algorithm, even when the variables are subject to linear constraints. We already know that the distance to the object, and its rank play a large role in the decision process.

We model these three as having multiplicative effects, i.e., for a user u considering an object i, the score s_i can be decomposed as $s_i = f_d(d(i, u)) \cdot f_r(r(i, u)) \cdot w_i$, where f_d is a function of the distance $d(i, u)$ between the user and the object, f_r is a function of the rank $r(i, u)$ of the object for the user, and w_i is the inherent wonderfulness.

Let $g_d(d(i, u)) = \log f_d(d(i, u))$, $g_r(r(i, u)) = \log f_r(r(i, u))$, and $y_i = \log w_i$. The likelihood maximization problem can be written as a concave program:

$$\text{Maximize: } \sum_{q \in Q} z_{r(q),u(q)} - \sum_{q \in Q} \log \left(\sum_{j \in C(q)} e^{z_{i,u(q)}} \right),$$
$$\text{subject to: } z_{i,u} = g_d(d(i, u)) + g_r(r(i, u)) + y_i \qquad \forall i, u.$$

We use the gradient ascent algorithm described below to solve this problem: We proceed in phases, with each phase consisting of three gradient ascent steps. In the first step, we only compute the gradient on the coordinates related to the function f_d, which governs the influence of the distance of the object to the user. Upon updating these coordinates, we then perform a gradient ascent step only on the coordinates related to f_r, which is responsible for the influence of the rank of the object; finally, in the third step, we update the intrinsic quality parameters referred to by w_i.

Note that this simple gradient ascent algorithm is not one of the polynomial-time algorithms for solving a concave maximization problem subject to linear constraints. In fact, it is not hard to see that if we write the objective as a function of the variables f_d, f_r, and w_i, this function is a concave function of individual variables, but it is not necessarily a concave function of the vector of all these variables. This means that the gradient ascent will converge to a local maximum, but not necessarily to a global maximum.

## 5. ANALYZING RESTAURANT CHOICE

We have now defined a geographic choice model that learns unknown functions of rank and distance, and optionally simultaneously learns a per-place wonderfulness score. We have shown how to learn this model, and how to use it for prediction. Model in hand, we now turn to an analysis of restaurant choice via driving directions in map search logs, using the data described in Section 2.

First, to get a good understanding of the effect of both rank and distance, we perform the gradient ascent in Section 4.1 while ignoring the wonderfulness parameter for each restaurant. We factor the score for a restaurant at distance d and rank r as:
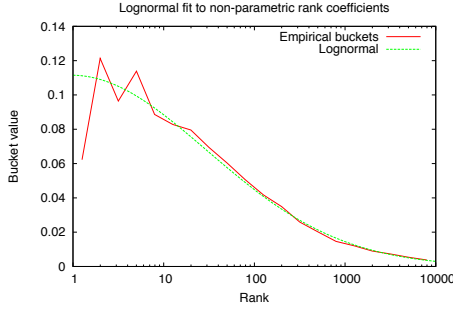
$$s(d, r) = f_d(d) \cdot f_r(r).$$

In the context of restaurants, to estimate f_d and f_r, we discretize both distance (between 1km and 300km[4]) and rank (up to 10K) into

---

[4]Everything below 1km is folded into 1km and distances beyond 300km are thrown out.

Lognormal fit to non-parametric distance coefficients
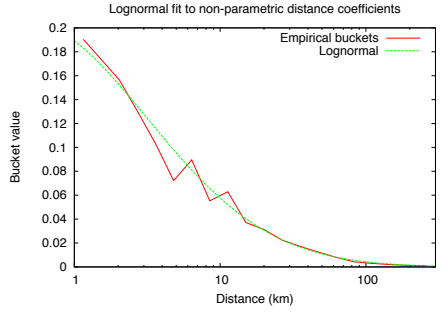
(a) Imputed distance function

Lognormal fit to non-parametric rank coefficients

(b) Imputed rank function

Figure 4: The maximum likelihood functions of impact of distance and rank on user's choice.

20 buckets each, binned on the log-scale. We can compute a non-parametric fit for f_r and f_d in the factored joint model by iteratively re-computing $f_r(r)$ and $f_d(d)$ for the 20 r values and 20 d values on a set of 100K training instances. We refer to the resulting f_d and f_r values as the *Non-Parametric Fit* (or *NPF*) coefficients, which are shown in red solid curves in Figures 4(a), 4(b) respectively.

The figures show that the distributions have a remarkably good fit to a lognormal distribution. Roughly speaking this shows three regimes that people face when deciding on a restaurant—for restaurants that lie in the head of the rank distribution (the closest half dozen establishments), there is little difference between the top few restaurants; for those that lie in the body (about top 250) there is a definite decline in probability as a function of the rank. For those in the tail, the decline is still prominent, but is much less significant. A similar explanation applies to the distance trade-off function: for the closest 20km or so, the score drops linearly with the log of the distance; however, the distance effect flattens out beyond the 20km driving range.

5.1 Comparing against baselines

Next, we examine how well these NPF coefficients (and their parametric fit) capture the data, in comparison to a number of baseline models. In what follows, all models were trained on the same set of 100K training instances as NPF, and the likelihood of each chosen destination was computed for a separate set of 100K test instances. We report the geometric means of these per-instance likelihoods for each model in Table 1.

The scoring function using the NPF coefficients provides the best performance in terms of yielding the highest likelihood over the test data. We then fit a parametric model to these NPF coefficients for $f_r(r)$ and $f_d(d)$ respectively. We experimented with both a lognor-

mal form and a sigmoid form. While the performance is slightly worse than NPF, both yielded significant improvement over modeling rank (or distance) alone, with lognormal fit being the slightly better of the two.

Setting f_r and f_d to other forms can result in useful degenerate scoring functions. If both f_r and f_d are constants, then we get back a uniform scoring function, where each restaurant is equally likely. If we take f_r to be a constant c, then we get back a scoring function based on distance alone (i.e., a distance marginal). We estimate $\hat{f}_d$ as the ratio of the number of times d was chosen to the number of times d was an option. Similarly setting f_d to be a constant yields a rank marginal. While both are significantly better than the uniform baseline, rank marginal is much more effective.

At first glance, it might look like modeling rank alone explains unseen data better than modeling distance alone. However, there is a caveat for using the empirical distance marginal as f_d. While the rank marginal can be interpreted as the probability of observing a candidate restaurant at rank r being chosen, the distance marginal can only be interpreted as the probability of choosing restaurants whose distance is between d and $d + \Delta d$. That is, $P_d(d)$ needs to be shared among restaurants at distance d away (or between d and $d + \Delta d$ for a small Δd). And for larger d, $P_d(d)$ needs to be shared among more instances to yield the probability of going to *one* restaurant at distance d. If we have a function $n(d)$ that captures the number of restaurants at distance d, we could have set $f_d(d) = P_d(d)/n(d)$. By setting $f_d(d) = P_d(d)$, restaurants farther away are given more probability mass than they should.

Nonetheless, for both baselines using empirical marginal distributions, we did not (need to) bucket the data in the same coarse granularity as we did for NPF coefficient estimation. In that sense, NPF is a more parsimonious model, but it still outperforms all the baselines. Indeed, even the model using lognormal fits parameterized by only four parameters (two for f_r and two for f_d) outperforms the marginals. This validated our intuition that it is important to model the rank and distance jointly, and neither sufficiently captures geo-related aspects of restaurant choice.

5.2 Incorporating quality scores

In the above computation we did not include a parameter specific to each restaurant. We run the full optimization with per-restaurant wonderfulness scores. We restrict our attention to those restaurants in the Bay Area (those within 100km of Chez Panisse, Berkeley).

One should interpret the wonderfulness score as follows. Once rank and distance have been taken into account, wonderfulness captures the relative likelihood of a user to visit one place over another. For uniformly distributed users and places, this will exactly equal popularity, but in fact users and places are not uniformly distributed. As an example, consider Figure 5. This shows two popular restaurants: In-N-Out Burger (in Mountain View), and House of Prime Rib (in San Francisco). The figure shows the distances traveled to each. In-N-Out draws traffic from further afield, garnering more visits relative to rank and distance-based likelihood, and hence scores higher under wonderfulness.

To evaluate our approach, we compare the induced ordering of the restaurants to two different ratings: one set of well known critic ratings, which we denote by C, and a set of user-generated ratings U. As a baseline, we take a simple measure of popularity of a particular establishment, based on counts from our data. For each pair of ratings, we compute the Kendall's tau rank correlation; the results are shown in Table 2.

The orderings provided by W and P are highly correlated with each other, this is expected as the wonderfulness scores provide a residual popularity once the rank and distance effects are taken

|	Scoring function	Likelihood
NPF	Non-parametric factored model: red solid curves in Figures 4(a) and 4(b)	5.3
NPF-lognormal	Lognormal fit to NPF coefficients: green dotted curves in Figures 4(a) and 4(b)	**5.1**
NPF-sigmoid	Sigmoid fit to NPF coefficients	4.9
Rank only	$f_r(r)$ = rank marginal, $f_d(d)$ = const	4.6
Distance only	$f_d(d)$ = distance marginal, $f_r(r)$ = const	3.9
Uniform	Each restaurant equally likely: $s(r, d)$ = const	1.1

Table 1: Likelihood (in 10^{-4}) of different estimation procedures.

(a) In-N-Out Burger (b) House of Prime Rib
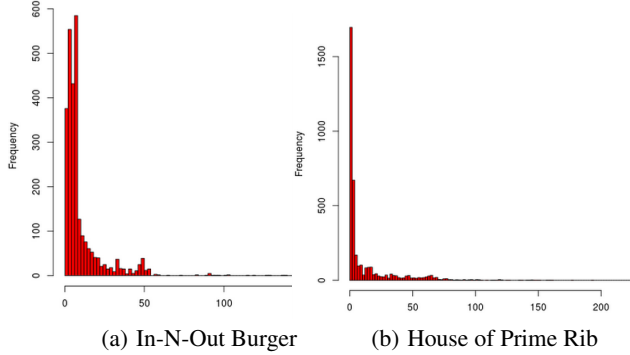
Figure 5: Distribution of distance traveled (km). Note that In-N-Out Burger will have a higher wonderfulness score since it draws a large fraction of its visits from further away.

	P	C	U
W	0.86	0.14	0.16
P	–	0.12	0.06

**Table 2: Kendall tau rank correlation between rankings induced by different restaurant rating systems.**
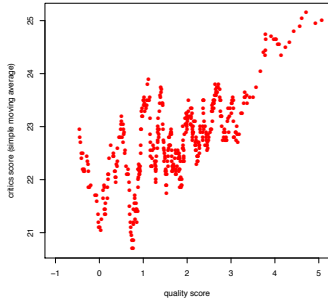
Figure 6: Scatter plot of wonderfulness versus critic score.

into consideration. In that sense W provides a refinement to P. As the correlations with U and C show this refinement significantly improves the observed correlation and wonderfulness is a better estimate of user generated scores. The result is significant with $p < 0.05$.

Figure 6 shows a scatter plot of wonderfulness versus the critic score. Kendall's tau is unforgiving, but the scatter plot shows that the score does correlate with the critics.

While it is encouraging that wonderfulness shows good correlation with both critics and users, our actual goal in introducing the measure is to improve likelihood, as this represents a baseline model for personalization and other extensions. For our Bay Area
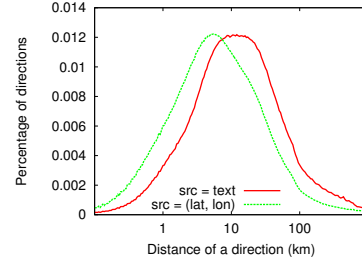
Figure 7: Distribution of individual-direction distances; $\log_{10}(\text{distance})$ is bucketed in units of 0.02.

dataset, we recomputed likelihood scores of restaurant choice using our most effective model of rank and distance. We also measured the improvement in likelihood by incorporating wonderfulness scores. Recall that our empirical measurement of the rank and distance univariates require 40 parameters to fit, and are well-approximated by the four-parameter lognormal fit. Wonderfulness, on the other hand, incorporates a parameter for each restaurant, so we would expect higher likelihood. The result shows that incorporating these scores further improves the likelihood by a multiplicative factor of 1.87.

5.3 Other trends

Before switching to a theoretical discussion of the lognormal forms we have observed for rank and distance, we close our empirical analysis with a brief exploration of the spatial, temporal, and quality trends in restaurant choice. For the analysis, in addition to `src = text`, we used a sample of the driving distances for `src = (lat, lon)` directions but with `src` information removed.

Spatial trends: 12km to food. Figure 7 shows the distribution of direction queries at varying distances plotted on a log scale. Similar to previous works on human mobility, we observe that as the distance goes up, the fraction goes down. For `src = text`, we observe that the median distance is 12km. The empirical observation over distances is symmetric along this point: the fraction of directions with distance between $(1.2, 12)$ is 42.9%, and the fraction of directions with distance between $(12, 120)$ is 42.3%. One may argue that this distribution is underestimating the actual foot traffic at smaller distances—clearly one does not need to look up for directions for a trip to the main cafe at work. More generally, our dataset may under-represent "neighborhood" traffic where people already knew how to reach a nearby venue. Nevertheless, 50% of the directions are trying to reach a destination within a small radius of 12km. For `src = (lat, lon)`, we observe higher fraction of traffic at shorter distances as expected—refreshing directions en route alone would have caused the distance distribution for `src = (lat, lon)` to shift to the left.

Temporal trends: TGIF. Figure 8(a) plots the median distance traveled on each day (relative to Monday), which clearly peaks
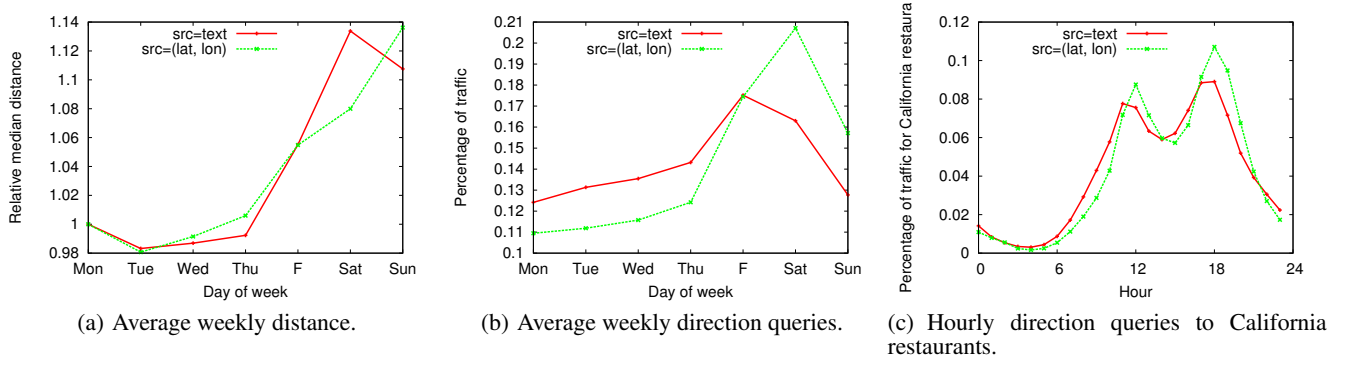
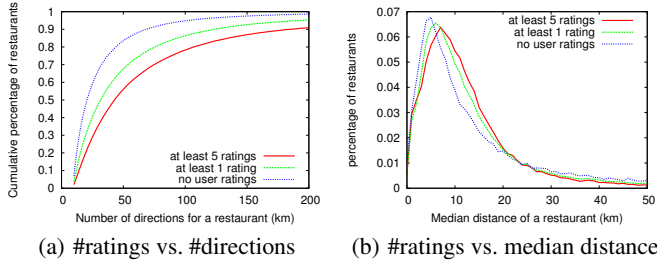(a) Average weekly distance.   (b) Average weekly direction queries.   (c) Hourly direction queries to California restaurants.

Figure 8: Temporal trends.

(a) #ratings vs. #directions   (b) #ratings vs. median distance

Figure 9: Quality trends.

during the weekend for both `src = text` and `src = (lat, lon)`. Figure 8(b) plots the fraction of direction queries associated with different days of the week. We observe that `src = text` directions peak on Fridays, while `src = (lat, lon)` directions peak on Saturdays. One hypothesis is people are more likely to go to unfamiliar and farther away places during the weekends, for which direction queries en route are sometimes inevitable.

Lastly, we plot the hourly trend of direction queries in Figure 8(c). To avoid any complication due to timezones, we restricted to directions ending in a California restaurant. There is clearly a lunch peak and a dinner peak. Compared to the `src = text` directions, the distribution of `src = (lat, lon)` direction traffic is slightly right-shifted, reflecting a more real-time direction need.

Quality trends: Star value. We examine the relationship between a restaurant's popularity measure and the number of directions it gets as well as the median distance users are willing to drive. To this end, we aggregate `src = text` directions by their destinations and pick restaurant with at least 10 directions. We then obtain the number of user ratings associated with each restaurant (from a popular online rating service), and group the restaurants into those with no ratings at all, and those with at least one rating. We also extract a subset of the second group as restaurants with at least five ratings. Figure 9(a) plots the cumulative fraction of restaurants with different number of directions. If a restaurant has no user ratings, then it is more likely to have fewer number of directions directed at it. And a restaurant with more user ratings is more likely to receive more directions. More interestingly, as we can see from Figure 9(b), a restaurant with more ratings is also more likely to have a higher median distance than one with fewer or no ratings. In other words, restaurants with (more) ratings are more likely to attract people from further away.

6. RATIONALE FOR LOGNORMALS

There are a few instances of lognormally-decaying functions of distance in the literature. A notable example is the log-distance path loss model for radio signal strength as a function of distance [27]; this process however applies to the distribution of signal strength at fixed distance from the transmitter, which is not applicable to our situation. On the other hand, lifetime distributions of biotic (longevity analysis) and abiotic (failure analysis) entities are frequently modeled by lognormal distributions ever since the classical work of [17]. According to these models, under *multiplicative degradation* a set of small multiplicative shocks apply over the evolution of the process, which results in a lognormal distribution of total degradation at fixed time t, and also a lognormal distribution of time to failure.

We may however speculate another root cause for the lognormal shape of the hidden rank and distance functions that our model learns. We had shown that the density is non-uniform and complex. For the sake of modeling, however, let us make a weaker assumption: the number of candidates at distance d for most individual instances of choice is reasonably approximated by some polynomial of d. For uniform density, this polynomial would be a quadratic, but we will allow any polynomial. Say object i will occur at some polynomial distance $\alpha(i)$, and will hence attain score

$$f_r(i) \cdot f_d(\alpha(i)) = \frac{1}{i\alpha(i)2\pi\sigma_d\sigma_r}e^{-\frac{(\ln i - \mu_r)^2}{2\sigma_i^2} - \frac{(\ln \alpha(i) - \mu_d)^2}{2\sigma_d^2}}.$$

This function follows the general form of a lognormal in rank, as do the resulting likelihoods, as the normalizing constant is the same for all candidates. We may therefore ask why a lognormal distribution is a natural function. A possible explanation comes from work of [13] and [33], who introduce an explanation based on entropy maximization. In our domain, the explanation proceeds as follows.

First, we argue that a user selecting an object incurs two costs. First, there is a *consideration cost*: the amount of time spent deliberating about where to go. We again make a simplifying assumption that users begin by considering closer places, and then only later consider more distant alternatives. Assuming a constant cost of consideration for each object, this results in cost linear in i to consider the first i places. Second, the user incurs a *travel cost*, which is proportional to the distance.

The next step of the derivation is to appeal to a well-studied law in human behavior: the *Weber–Fechner law*, which states that for a broad range of external stimuli, human perception of the intensity of the stimulus is proportional to the logarithm of the intensity [21, 23]. The law has been observed broadly to hold for human perception of time. Thus, assuming that a user examines candidate

places until finding an acceptable one, the number of distinguishable buckets of places in terms of travel and consideration time grows as $(\log r)(\log d)$, which is proportional to $\log^2 r$ according to our simplifying assumption.

Assuming the user has some upper bound on how much cost they are willing to incur in choosing an object, we may now ask for the maximum entropy distribution under this cost function. By an application of Lagrange multipliers, the resulting distribution is seen to be lognormal in distance or rank. Indeed, the problem is to

$$\text{Maximize: } \sum_i p(i) \log p(i),$$

$$\text{subject to: } \sum_i p(i) \log^2 i \le B.$$

The resulting Lagrangian is

$$\sum_i p(i) \log p(i) + \lambda \left(\sum_i p(i) \log^2 i - B \right).$$

Taking derivatives gives

$$\log p(i) = -1 - \lambda \log^2 i,$$

which yields our lognormal:

$$p(i) \propto \exp\left(-\lambda \log^2 i\right).$$

To summarize, we began this work plotting likelihood to choose a place as a function of rank, and separately of distance. These marginal distributions showed particular forms that were not well approximated by lognormal functions. We then applied our choice model in order to factor likelihood of choice into a product of unknown functions of rank and distance. To our surprise, even though the marginals have different forms, both unknown functions are well-approximated by lognormals. In this section we have shown that under two mild assumptions, these forms are in fact predicted by an entropy maximization framework. This suggests that the surprisingly good fits we saw in Figures 4(a), 4(b) are no coincidence, but capture an underlying process with a formal justification.

7. OTHER RELATED WORK

We already discussed the past work on geographic choice in Section 1. Other related work includes the work on human mobility patterns, rank-based models, geographic data mining, and mining directions in map search logs.

Human mobility patterns have been studied for decades; we list a few recent ones. Brockmann et al. [2] study the circulation of bank notes in the United States and found that the distribution of travel distances follows a power law; they conclude that human travel is a Lévy flight (see also [28]). In contrast, by tracking the location of mobile phone users, Gonzalez et al. [12] argue that human movements exhibit high degree of spatial and temporal regularity. The gravity model has been a popular model for explaining migration patterns; see the work of Simini et al. [32] and Jung et al. [15], the book [10], and the survey on urban trip distribution [8]. Levy [18] studies the migration patterns of humans and uses that to explain the gravitational law of social interaction. Social ties have also been used to explain human movements [4]. Taplin and Qiu [35] model the tourist attractions in Australia using population, travel times, and traffic. All these works deal with studying and modeling human travel patterns. Our work, on the other hand, is focused on directions with distances at finer granularities. In addition, rather than conflating different types of human mobilities, we pick a particular and important type of human travel: to restaurants.[5]

Rank-based models have been rediscovered several times in the past. They have been proposed as a better alternative to distance-based models since they can handle regions of varying densities in a mathematically clean manner. The related notion of intervening opportunities was first proposed by [34]: "the number of persons going a given distance is directly proportional to the number of opportunities at that distance and inversely proportional to the number of intervening opportunities." It has subsequently been studied in various settings [11, 14, 22, 37]. Liben-Nowell et al. [19] use rank-based friendships to explain the small-world phenomenon in social networks. Using FourSquare checkin data, Noulas et al. [24] propose a rank-based model for explaining urban mobility. They argue that rank-distance is a much better measure than physical distance, when it comes to mobility. Our work also considers rank as an alternative measure to distance. However, our findings are far more intricate: in a geographic choice context, we argue that neither suffices by itself.

The use of geographic information in various search and data mining tasks has been a hot area of research. Backstrom et al. [1] and Crandall et al. [6] study the role of IP-based geography in queries and geotagged images. Using geographic information in tweets has also been explored in a variety of topics including linguistics, entity matching, and geolocation [9, 25, 7, 29, 5, 31]. Data mining of GPS is also a well-studied topic. Cao et al. [3] consider mining significant semantic locations from GPS data and Zheng et al. [40] use GPS data for location-based recommendation; see also [41, 26, 30, 39]. Our work does not rely on IP/tag/GPS geolocation but uses the online direction queries as a powerful intent for physical visits: a user issuing a direction query to a place is more likely to visit the place than a user merely searching for the place. Furthermore, GPS data applies to a lot of locations from a few number of users, whereas direction queries correspond to a much broader range of users.

To the best of our knowledge, there has been very little work on studying directions in map query logs. Venetis et al. [36] use map queries to determine the importance of places of interest; their work is more on the efficient indexing of direction queries and a ranking methodology based on it. Xiao et al. [38] study map search logs, but their analysis is mainly focused on its general statistical properties rather than on directions. However, they also find that restaurant is the most popular map query category.

8. CONCLUSIONS

In this work we studied the dynamics of geographic choice. We analyzed the role of the geometric distance to an object as well as the rank-distance and found intricate connections between the two notions. We used these insights to propose a model of user behavior that captures the effects of rank, distance, and quality. We applied this model to study the properties of direction queries to restaurants on Google maps. Our work opens up an array of future research directions, e.g., using these insights to improve the recommendation and studying mobility patterns via direction queries as a proxy for non-routine adventure-seeking human behavior.

Acknowledgments

---

[5]Directions to restaurants forms the largest category among all direction queries; see [38].

9. REFERENCES

[1] L. Backstrom, J. M. Kleinberg, R. Kumar, and J. Novak. Spatial variation in search engine queries. In *WWW*, pages 357–366, 2008.

[2] D. Brockmann, L. Hufnagel, and T. Geisel. The scaling laws of human travel. *Nature*, 439(7075):462–465, 2006.

[3] X. Cao, G. Cong, and C. Jensen. Mining significant semantic locations from GPS data. *VLDB*, 3(1-2):1009–1020, 2010.

[4] E. Cho, S. Myers, and J. Leskovec. Friendship and mobility: User movement in location-based social networks. In *KDD*, pages 1082–1090, 2011.

[5] D. J. Crandall, L. Backstrom, D. Cosley, S. Suri, D. Huttenlocher, and J. Kleinberg. Inferring social ties from geographic coincidences. *PNAS*, 107(52):22436–22441, 2010.

[6] D. J. Crandall, L. Backstrom, D. P. Huttenlocher, and J. M. Kleinberg. Mapping the world's photos. In *WWW*, pages 761–770, 2009.

[7] N. Dalvi, R. Kumar, and B. Pang. Object matching in tweets with spatial models. In *WSDM*, pages 43–52, 2012.

[8] S. Easa. Urban trip distribution in practice. I: Conventional analysis. *Journal of Transportation Engineering*, 119(6):793–815, 1993.

[9] J. Eisenstein, B. O'Connor, N. A. Smith, and E. P. Xing. A latent variable model for geographic lexical variation. In *EMNLP*, pages 1277–1287, 2010.

[10] S. Erlander and N. Stewart. *The Gravity Model in Transportation Analysis: Theory and Extensions*. CRC Press, 1990.

[11] R. Freymeyer and P. Ritchey. Spatial distribution of opportunities and magnitude of migration: An investigation of Stouffer's theory. *Sociological Perspectives*, pages 419–440, 1985.

[12] M. Gonzalez, C. Hidalgo, and A. Barabasi. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, 2008.

[13] K. C. Gros and D. G. Markovic. Neuropsychological constraints to human data production on a global scale. *Eur Phys J B*, 85(28), 2012.

[14] K. Haynes, D. Poston, and P. Schnirring. Intermetropolitan migration in high and low opportunity areas: Indirect tests of the distance and intervening opportunities hypotheses. *Economic Geography*, 49(1):68–73, 1973.

[15] W. Jung, F. Wang, and H. Stanley. Gravity model in the Korean highway. *Europhysics Letters*, 81(4):48005, 2008.

[16] D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, 2009.

[17] A. Kolmogorov. On a logarithmic normal distribution law of the dimensions of particles under pulverization. *Dokl. Akad Nauk*, 31(2):99–101, 1941.

[18] M. Levy. Scale-free human migration and the geography of social networks. *Physica A: Statistical Mechanics and its Applications*, 389(21):4913–4917, 2010.

[19] D. Liben-Nowell, J. Novak, R. Kumar, P. Raghavan, and A. Tomkins. Geographic routing in social networks. *PNAS*, 102(33):11623–11628, 2005.

[20] R. D. Luce. *Individual Choice Behavior: A Theoretical Analysis*. Wiley, 1959.

[21] S. Masin, V. Zudini, and M. Antonelli. Early alternative derivations of Fechner's law. *J. History of the Behavioral Sciences*, 45:56–65, 2009.

[22] E. Miller. A note on the role of distance in migration: Costs of mobility versus intervening opportunities. *Journal of Regional Science*, 12(3):475–478, 1972.

[23] R. S. Moyer and T. K. Landauer. Time required for judgements of numerical inequality. *Nature*, 215(5109):1519–1520, 1967.

[24] 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):e37027, 05 2012.

[25] B. O'Connor, J. Eisenstein, E. P. Xing, and N. A. Smith. Discovering demographic language variation. In *Workshop on Machine Learning for Social Computing at NIPS*, 2010.

[26] D. Quercia, N. Lathia, F. Calabrese, G. Di Lorenzo, and J. Crowcroft. Recommending social events from mobile phone location data. In *ICDM*, pages 971–976, 2010.

[27] T. Rappaport. *Wireless Communications: Principles and Practice*. Prentice Hall PTR, 2nd edition, 2001.

[28] I. Rhee, M. Shin, S. Hong, K. Lee, S. Kim, and S. Chong. On the Lévy-walk nature of human mobility. *IEEE/ACM Transactions on Networking*, 19(3):630–643, 2011.

[29] A. Sadilek, H. Kautz, and J. Bigham. Finding your friends and following them to where you are. In *WSDM*, pages 723–732, 2012.

[30] S. Scellato, C. Mascolo, M. Musolesi, and J. Crowcroft. Track globally, deliver locally: Improving content delivery networks by tracking geographic social cascades. In *WWW*, pages 457–466, 2011.

[31] S. Scellato, A. Noulas, R. Lambiotte, and C. Mascolo. Socio-spatial properties of online location-based social networks. In *ICWSM*, pages 329–336, 2011.

[32] F. Simini, M. González, A. Maritan, and A. Barabási. A universal model for mobility and migration patterns. *Nature*, 484(7392):96–100, 2012.

[33] P. Sobkowicz, M. Thelwall, K. Buckley, G. Paltoglou, and A. Sobkowicz. Lognormal distributions of user post lengths in Internet discussions - a consequence of the Weber–Fechner law? *EPJ Data Science*, 2(2), 2013.

[34] S. Stouffer. Intervening opportunities: A theory relating mobility and distance. *American Sociological Review*, pages 845–867, 1940.

[35] J. Taplin and M. Qiu. Car trip attraction and route choice in Australia. *Annals of Tourism Research*, 24(3):624–637, 1997.

[36] P. Venetis, H. Gonzalez, C. Jensen, and A. Halevy. Hyper-local, directions-based ranking of places. *VLDB*, 4(5):290–301, 2011.

[37] W. Wadycki. Stouffer's model of migration: A comparison of interstate and metropolitan flows. *Demography*, 12(1):121–128, 1975.

[38] X. Xiao, Q. Luo, Z. Li, X. Xie, and W.-Y. Ma. A large-scale study on map search logs. *TWEB*, 4(3), 2010.

[39] M. Ye, P. Yin, W. Lee, and D. Lee. Exploiting geographical influence for collaborative point-of-interest recommendation. In *SIGIR*, pages 325–334, 2011.

[40] V. W. Zheng, Y. Zheng, X. Xie, and Q. Yang. Collaborative location and activity recommendations with GPS history data. In *WWW*, pages 1029–1038, 2010.

[41] Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma. Mining interesting locations and travel sequences from GPS trajectories. In *WWW*, pages 791–800, 2009.