Learning Latent Factor Models of Travel Data for Travel Prediction and Analysis

Michael Guerzhoy[1] and Aaron Hertzmann[1],[2]

[1] University of Toronto {guerzhoy,hertzman}@cs.toronto.edu
[2] Adobe Research

Abstract. We describe latent factor probability models of human travel, which we learn from data. The latent factors represent interpretable properties: travel distance cost, desirability of destinations, and affinity between locations. Individuals are clustered into distinct styles of travel. The latent factors combine in a multiplicative manner, and are learned using Maximum Likelihood.

We show that our models explain the data significantly better than histogram-based methods. We also visualize the model parameters to show information about travelers and travel patterns. We show that different individuals exhibit different propensity to travel large distances. We extract the desirability of destinations on the map, which is distinct from their popularity. We show that pairs of locations have different affinities with each other, and that these affinities are partly explained by travelers' preference for staying within national borders and within the borders of linguistic areas. The method is demonstrated on two sources of travel data: geotags from Flickr images, and GPS tracks from Shanghai taxis.

1 Introduction

Understanding human travel and mobility is a key component of understanding individuals and societies. In the past few years, researchers in many different disciplines have looked to online social media datasets as new sources of information about how we move from place to place. Good models of this data could yield new scientific insights into human behavior [1, 8, 11]. Furthermore, tools for analysis and prediction of movement can be useful for numerous applications. For example, travel models can help in predicting the spread of disease [6, 15]; surveying tourism [9, 10], traffic [2, 14], and special events mobility for urban planning [3]; geolocating with computer vision [4, 16]; interpreting activity from movements [19, 22]; and recommending travel [7, 17]. Analysis of travel data can also play a role within the larger context of analysis of social network data. More recent work has explored the value of social network and taxi trajectory data for predicting mobility [5] and knowledge discovery [23].

Previous work has generally employed two kinds of models to describe human travel probabilities. First, scientists have employed the Lévy flight, a stochastic process model [1, 11]. This model captures the heavy-tailed nature of travel, but ignores the actual locations being visited. Second, one can build probability tables based on empirical travel histograms [4, 6, 9, 10, 15, 16]. Such models are more accurate than simple stochastic process models, but require enormous datasets while revealing little

about the underlying structure of the data. Detailed models have also been developed for some specialized applications (e.g., [14, 19]).

This paper describes an approach to building structured models of human travel, using spatially-varying and individual-varying *latent factors*. The main idea is to model travel probabilities as functions of spatially-varying latent properties of locations and the travel distance. The latent factors represent interpretable properties: travel distance cost, desirability of destinations, and affinity between locations. Moreover, individuals are clustered into clusters with distinct travel models. The latent factors combine in a multiplicative manner, and are learned using Maximum Likelihood. The resulting models exhibit significant improvements in predictive power over previous methods, while also using far fewer parameters than histogram-based methods.

We present our approach as a family of models of increasing generality, similar to the use of multiplicative models in social network analysis [13]. While there are many factors in travel that we have not incorporated into the model, the multiplicative formulation easily lends itself to incorporating additional latent factors and sources of information.

Our model allows us to analyze the different "factors" separately. We visualize and analyze the propensity to travel long distance of different user clusters, independently of the locations to which they travel. We show that the desirabilities of the locations on the map are distinct from the number of transitions to those locations, partly since some desirable locations are more remote. We analyze the affinities between the different locations, independently of the desirability of the locations and the distances between locations. In particular, the affinities our model learns are stronger, on average, for pairs of locations inside national borders, and for some languages, for pairs of locations within the same linguistic areas.

2 Background

Suppose we discretize the world into J locations, and we observe an individual at location i. We wish to predict where that individual is likely to be after a certain time interval τ: $P_{ij\tau} \equiv P(L_{next} = j | L_{current} = i, \Delta T = \tau)$.

Like most previous work, we make the simplifying assumption that human travel can be treated as a Markov process. We also discretize all continuous quantities: τ represents one of a fixed set of time interval bins, and i and j represent location bins (quads) on the map.

To date, two general kinds of models have been applied to human travel. First, scientists have employed the Lévy flight, a stochastic process model [1]. In particular, the marginal probability distribution of traveling a distance d in some fixed time interval is given by a truncated power law:

$$p(d) \propto d^{-b} \tag{1}$$

for a parameter b. This model captures the heavy-tailed nature of travel: long-distance travel is rare but not surprising. A truncated power law [11] can be used to model limits in travel. These simple parametric models can be interpreted to yield insights into travel

behavior; however, they assume that travel probabilities depend only on the travel distance. Destinations that are equidistant from one's current location are equally likely, regardless of whether they are in New York City or in the Arctic.

Second, one can build probability tables based on empirical travel histograms [4, 6, 9, 10, 15, 16]. For example, Kalogerakis et al. [16] use counts of actual transitions in a database of 6 million travel records. Letting $N_{ij\tau}$ be the number of observed transitions from location i to location j in time interval τ, and $N_{d\tau}$ be the number of observed transitions over distance d in time interval τ, they set

$$P_{ij\tau} = \frac{N_{ij\tau} + \lambda q_{d_{ij}\tau}}{\sum_\ell (N_{i\ell\tau} + \lambda q_{d_{i\ell}\tau})} \qquad (2)$$

where d_{ij} is the distance between locations i and j. The histogram is regularized by the term $q_{d\tau} = \frac{N_{d\tau}}{\sum_\delta N_{\delta\tau}}$, which corresponds to the probability of traveling a given distance d in time interval τ in the training set, with a constant factor λ set using cross-validation. Such models are more accurate than simple stochastic process models, but require enormous datasets while revealing little about the underlying structure of the data.

Our work aims to combine the advantages of the two approaches described above, while extending them to yield new generality.

3   The Datasets

We consider two datasets. The first dataset comprises the publicly-available Flickr.com image streams of 75,250 distinct individuals, collected by Kalogerakis et al. [16]; there are about 6 million images in total in the dataset. Each image is accompanied by a timestamp, and a geotag specifies the location where the picture was taken. We consider only the geotags and the time intervals between consecutive geotags. As done by Kalogerakis et al., we discretize the Earth into 3186 bins, each of which is approximately $400\text{km} \times 400\text{km}$ in size. An individual's trajectory is the sequence of transitions between the locations of the individual's photos.

The second dataset comprises taxi travel data, collected by Peng et al. [20]. We use the taxi trips recorded on Feb. 1, 2007. A trip (transition) begins or ends when the taxi changes its status from "occupied" to "vacant" or vice versa; we calculate the duration of each trip. The dataset, which we split into a training set and a test set, consists of 2,000 taxis which take about 287,000 trips in total. We discretize the area around Shanghai served by the taxis in the dataset (30°N to 32°N, 120°E to 122°E) into 50×50 quads using a rectangular grid.

The locations on the map differ in popularity (see Fig. 2). In the Flickr dataset, we observe there are more transitions to locations in the coastal United States, Western Europe, Australia, and Japan than to locations elsewhere on the map. In the taxi dataset, there are more transitions to locations in downtown Shanghai. The variation in the probability of transitioning from a given source to a given destination is not determined by the distance between the source and the destination alone. Individuals in our datasets exhibit travel patterns that are distinct from one another; see Fig. 1.
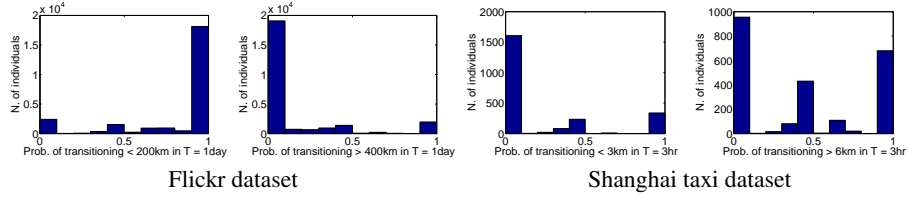
Fig. 1. Histogram of different individuals travel propensities. For a given distance and timeframe, and for a given probability, each bar in a histogram represents the number of individuals who travel the distance within the timeframe with the given probability.

4   Basic Model: Distance Cost and Desirability

The model can be motivated by observing that one is usually more likely to travel to nearby destinations than to further ones, but that certain destinations are more popular than others. For example, New York is almost always a local peak compared to its neighbors, whereas the ocean is never a destination. We note that there is a "logical-AND" relationship between these factors: undesirable destinations (like the ocean) are never visited no matter how close; highly-desirable destinations (like New York) are never visited if they are simply too far to be reached within the given time interval. This relationship suggests the use of a multiplicative model to combine these factors.

In particular, we consider the following multiplicative model:

$$P_{ij\tau} = \frac{r(d_{ij}, \tau)\, a_j}{\sum_\ell r(d_{i\ell}, \tau)\, a_\ell} \tag{3}$$

The model consists of two terms. First, the *distance factor* $r(d, \tau)$ captures the dependence of travel on the distance d_{ij} between locations i and j and the transition time τ. For a given τ, this term mimics the power-law terms used in previous work (Eq. 1). Second, the spatially-varying a_j term represents the *desirability* of the destination j. While not representing literal "desirability," it reflects that fact that some destinations attract more travel than others. For convenience, we define the parameters $\rho(d, \tau) = \log r(d, \tau)$ and $\alpha_j = \log a_j$. And we write the log-probability of traveling to location j from location i as $\log P_{ij\tau} \doteq \rho(d_{ij}, \tau) + \alpha_j$. (The symbol $\doteq$ indicates that we have omitted the normalization term.)

The quantities d and τ are discretized, and the functions $\rho(d, \tau)$ and α_j are each represented as 2D look-up tables. This method can be viewed as a form of "logistic PCA" [21] applied to travel data, taking the structure of the problem into account. The model can also be viewed as a special case of logistic regression, because the probability $P_{ij\tau}$ is linear in the parameters r and a. In particular, the model can be written as $P_{ij\tau} \propto \exp(\boldsymbol{\theta}^T \boldsymbol{\phi}_{ij\tau})$, where $\boldsymbol{\theta} = [\text{vec}(\rho)^T, \text{vec}(\alpha)^T]^T$, and $\boldsymbol{\phi}_{ij\tau}$ is a vector of indicator variables.

For our problem, the size of the dataset dwarfs the number of model parameters, which suggests that Maximum Likelihood should be sufficient for learning. We estimate the α and ρ parameters from the data, by minimizing the negative log-likelihood using

Most popular	Most popular
1. London	1. South Shaanxi Rd.
2. New York	2. North Zhongstan Rd.
3. San Francisco	3. Zhongstan Park
4. Paris	4. Xiaonanmen
5. Milan	5. Guilin Rd.
6. Washington, DC	6. Linping Rd.
7. Vancouver	7. Sci. & Tech. Museum
8. Chicago	8. Xincun Rd.

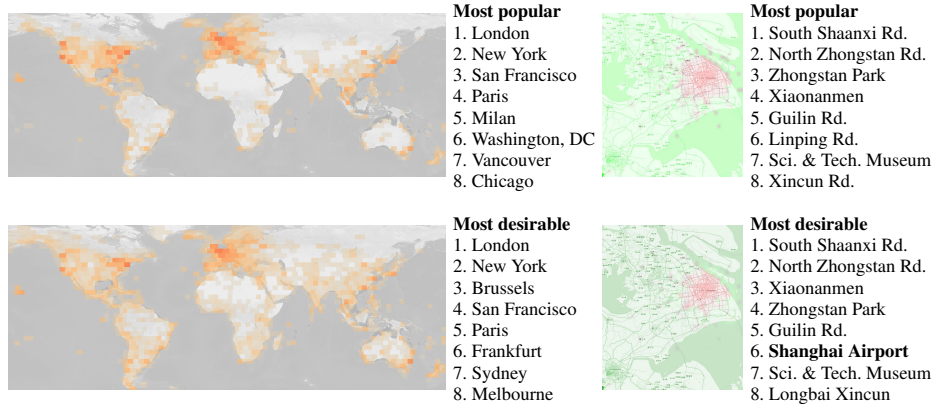| Most desirable	Most desirable
1. London	1. South Shaanxi Rd.
2. New York	2. North Zhongstan Rd.
3. Brussels	3. Xiaonanmen
4. San Francisco	4. Zhongstan Park
5. Paris	5. Guilin Rd.
6. Frankfurt	6. **Shanghai Airport**
7. Sydney	7. Sci. & Tech. Museum
8. Melbourne	8. Longbai Xincun

Fig. 2. The popularity (count of inbound transitions) and desirability (our learned parameter α) of destinations, with the top locations displayed. Desirability and popularity are related but distinct. On the Shanghai map, strong red means very large and strong green means very small. Quads in Shanghai identified by Metro stations, except for Shanghai Airport.

the conjugate gradient method. The negative log-likelihood of the data is:

$$NLL = -\log \prod_{ij\tau} (P_{ij\tau})^{N_{ij\tau}} = -\sum_{ij\tau} N_{ij\tau} \log P_{ij\tau} \tag{4}$$

Here, $N_{ij\tau}$ is the number of observed transitions from i to j in time interval τ. The equivalence to logistic regression shows that the negative log-likelihood is convex.

5 Location Affinities

We expect that similarities between locations would play a role in travel. For example, we expect that people are more likely to stay in their own countries than to cross borders, and, when leaving their country, they are more likely to visit countries where the same language is spoken as in their home country. We can express this by assigning a low-dimensional latent parameter vectors $\mathbf{u}_j$, $\mathbf{v}_j$ to each location j (for a discussion of why an asymmetric factor is useful, see [12]) . A low-dimensional vector is used, typically $\mathbf{u}_j, \mathbf{v}_j \in R^4$. The transition probability is then given by $\log P_{ij\tau} \doteq \rho(d_{ij}, \tau) + \alpha_j + \mathbf{u}_i^T \mathbf{v}_j$. The log-likelihood of a transition is then:

$$\log P_{ij\tau} \doteq \rho(d_{ij}, \tau) + \alpha_j + \mathbf{u}_i^T \mathbf{v}_j \tag{5}$$

The learned models are shown in Figs. 2 and 3.

6 Clustering Individuals

We incorporate the observation that travel varies between individuals [11] by clustering individuals into C clusters. Each cluster has its own distance factor functions $\rho^{(c)}$. We
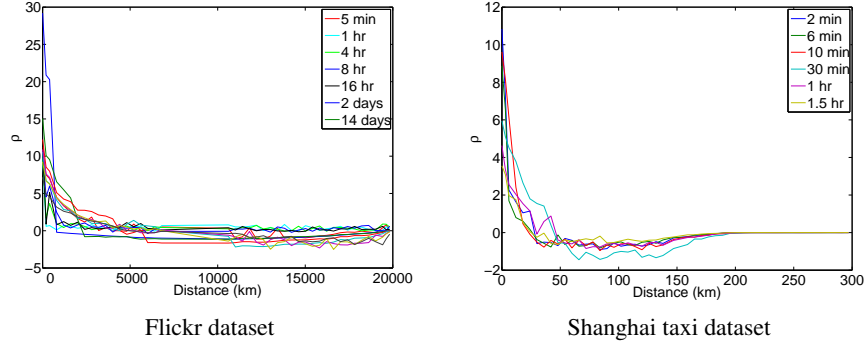
Flickr dataset   Shanghai taxi dataset

Fig. 3. The learned distance factors for the Flickr and Shanghai taxi datasets, for selected time intervals out of a total of 23 and 13 respectively.

assume that each individual travels according to the parameters of one cluster, and that an individual belongs to cluster c with prior probability π_c, such that $\sum_c \pi_c = 1$. The other parameters $\alpha, \mathbf{u}, \mathbf{v}$ are shared across all clusters; we experimented with clustering them as well, but this did not lead to increased performance.

The probability of an individual's travel trajectory visiting locations $L_{1:K}$, given that the individual is a member of cluster c, is $P_{1:K}^{(c)} = P(L_1) \prod_{k=2}^{k} P^{(c)}(L_k | L_{k-1}, \tau_{k-1}, c)$ where $P^{(c)}$ corresponds to $P_{ij\tau}$ in Eq. 5, as parametrized by cluster c. We use a uniform distribution over bins for the start probability $P(L_1)$. The probability of an individual's entire trajectory of K locations, marginalized over the possible cluster assignments is

$$P_{1:K} = \sum_c \pi_c P_{1:K}^{(c)} \tag{6}$$

Some learned ρ values for the different clusters are shown in Fig. 4.

7   Fitting the Models

Our single-cluster model is given by Eq. 5. The Negative Log-Likelihood (NLL) of the data under the single-cluster model is then

$NLL = -\log \prod_{ij\tau} (P_{ij\tau})^{N_{ij\tau}} =$
$-\sum_{\tau d} N_{\tau d}\rho(d, \tau) - \sum_j N_j \alpha_j + \sum_{ij\tau} N_{ij\tau} \log \sum_\ell \exp(\rho(d_{i\ell}\tau) + \alpha_\ell)$.

Taking derivatives with respect to the model parameters yields:

$\frac{\partial NLL}{\partial \alpha_j} = -N_j + \sum_{i\tau} N_{i\tau} \frac{\exp(\rho(d_{ij}, \tau) + \alpha_j)}{\sum_\ell \exp(\rho(d_{i\ell}, \tau) + \alpha_\ell)} = -N_j + \sum_{i\tau} N_{i\tau} P_{ij\tau}$,

$\frac{\partial NLL}{\partial \rho_{\tau d}} = -N_{\tau d} + \sum_{ij} N_{ij\tau} \frac{\sum_{\ell:d_{i\ell}=d} \exp(\rho(d_{i\ell}, \tau) + \alpha_\ell)}{\sum_\ell \exp(\rho(d_{i\ell}, \tau) + \alpha_\ell)} = -N_{\tau d} + \sum_i N_{i\tau} P_{i\tau d}$,

$\frac{\partial NLL}{\partial u_i} = -\sum_{j\tau} N_{ij\tau} v_j + \sum_{j\tau} N_{ij\tau} \sum_\ell v_\ell P_{i\ell\tau}$.

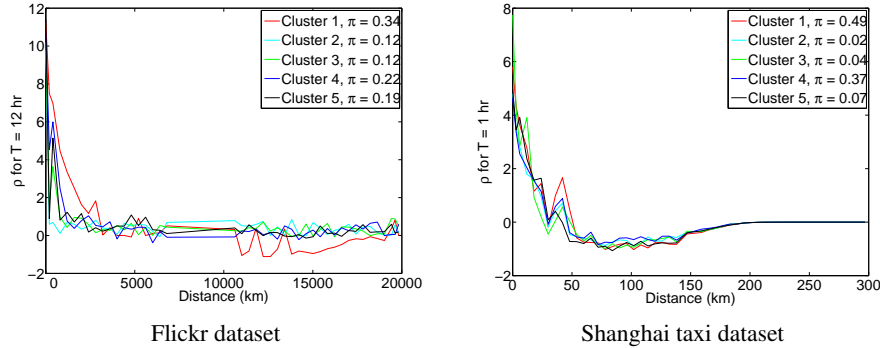We optimize the NLL of the training data using the conjugate gradient method to fit the model.

Fig. 4. Sample learned ρ values for different clusters for the Flickr and Shanghai taxi datasets for a set time period. The prior probability of each cluster is given in the legend. Note the markedly different probabilities for traveling less than 3km for the different clusters for the Shanghai taxi dataset. For the Flickr dataset, Cluster 1 exhibits more propensity for traveling mid-range travel distances than the other models, and Cluster 2 exhibits less propensity for traveling mid-range travel distances than the other models.

For the Cluster model (Equation 6), we optimize the NLL using a generalized EM algorithm [18], using exact E-steps and gradient descent M-steps.

8   Results

We first analyze our models quantitatively and show that they demonstrate improved explanatory power as compared to the baseline model. We then analyze the parameters of the models.

8.1   The Predictive Power of the Models

We test the algorithm on the Flickr dataset by splitting individuals into test and training sets, with 52,769 individuals in the training set and 11,149 individual in the test set; we set the dimensionality of the u and v vectors to 4. We discretize time into 23 intervals. We use 737 taxis from the Shanghai taxi dataset for training and 1,018 taxis for the test set; the dimensionality of u and v is 4, and we discretize time into 13 intervals. We obtain the results shown in Table 1. Throughout, we compute the negative log-likelihood (NLL) of each test individual's transitions according to the learned models. For the Cluster model, we compute the likelihood of each individual trajectory (see Eq. 6); for comparison purposes, we compute the NLL of each transition separately by marginalizing over clusters and present the test NLL obtained by summing the NLLs of all the transitions, as in Eq. 4:

$$NLL = -\sum_{ij\tau} N_{ij\tau} \log \sum_c \pi_c P_{ij\tau}^{(c)} \qquad (7)$$

Note that, for the non-Cluster models, the NLL of an individual's trajectory is always simply the sum of the NLLs of all the transitions in the trajectory (cf. Eq. 4).

Our models yield significantly better NLLs than the histogram models, while using orders-of-magnitude fewer parameters. We see performance improve as we use more sophisticated models, with the Cluster model giving the best predictions. Note that, for the Cluster model, the gain is obtained due to the fact that we marginalize the entire travel trajectory of each individual over cluster assignments. If we ignore the fact that all the transitions in a trajectory were generated by a single individual ("Cluster, indep."), there is no gain in the train likelihood over the Affinity model, and the test NLL is slightly worse. This is to be expected, since we have a non-parametric (binned) ρ, and model averaging does not increase the effective number of parameters much.

Model	Taxi dataset			Flickr dataset		
	Parameters	Train NLL	Test NLL	Parameters	Train NLL	Test NLL
Histogram (Eq. 2)	1,004,692	236.1	511.8	26,332,700	15.76	47.27
Reg. Histogram (Eq. 2)	1,004,692	251.0	279.0	26,332,700	17.36	21.53
Latent model (Eq. 3)	1,578	272.2	276.1	3,370	19.90	20.00
Affinity model (Eq. 5)	3,802	264.6	270.1	11,930	18.95	19.10
Cluster model (Eq. 6)	9,002	232.0	231.8	21,130	18.43	18.66
Cluster, indep. (Eq. 7)	9,002	263.2	271.1	21,130	18.95	19.28

Table 1. NLL per user for the Shanghai and Flickr datasets. The number of parameters is calculated by excluding parameters that pertain to map quads that are never visited.

The main reason for the improvement due to the parametric methods is that the histogram-based methods are very statistically inefficient, whereas the factored models can generalize. For the Flickr dataset, the full histogram contains 231 million entries, which would require enormous datasets to accurately estimate. Indeed, in the learned histogram, only 97,190 of the entries are non-zero. To see the benefit of the parametric model, consider all the possible self-transitions (i.e., transitions from bin i to i) in the test set that do not occur in the training set. The contribution from these transitions to the test NLL is $1.18e3$ under the Regularized Histogram model and 495.08 under the Affinity model. The differences in the transition probabilities can be quite significant. For example, the probability of the Tabriz $\rightarrow$ Tabriz transition in 30 to 60 days is 0.04 under the Regularized Histogram model and 0.14 under the Affinity model. Transitions that take more than 30 days are rare, but our model is able to infer a better transition probability by using data from shorter transitions.

8.2 Popularity vs. Desirability

We learned the *desirability* for each location on the maps. The desirabilities and the popularities (the number of inbound transitions to the location) of a location are related but distinct: see Fig. 2. A partial explanation of the difference is the remoteness of some locations: Sydney and Melbourne are in the top ten most desirable destinations, but are

less popular because of their remoteness. The Shanghai Airport is also remote and more desirable than popular.

8.3 Affinity Factors

Analyzing affinity factors $(u_i^T v_j)$ between pairs of locations allows us to discover affinities between locations that are not accounted for by proximity and desirability. For example, there is obviously more traffic than the average between Russian-speaking cities because some of them are desirable and they are all located in or near Russia. However, by analyzing our learned model we find that there is more traffic between Russian-speaking cities than we would expect based on the proximity between them and their popularity alone.

For the Flickr dataset, we show that our model captures the fact that locations within national borders have higher affinity with each other than the average. We show that the same holds for some linguistic areas. The mean log-affinity factors within and across nation boundaries are plotted in Fig. 5. As expected, the model learns that, on average, locations within the same national boundaries have higher affinities with each other.
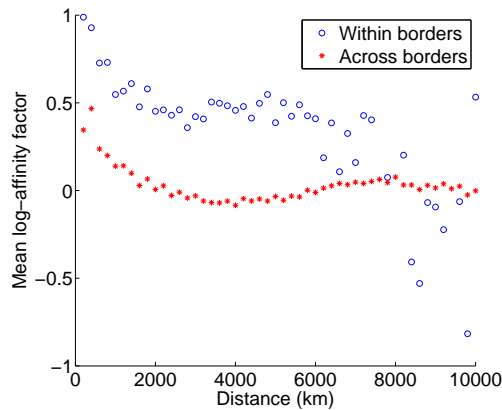

Fig. 5. The mean log-affinity factors within and across national boundaries. For distaces larger than 6000 km, very few location pairs are available.

We explore the affinities for pairs of locations which share an official language, but which are located in different countries. In Fig. 6, we observe a clear pattern where affinities are higher for pairs of Spanish-speaking locations in different countries. There is a clear pattern for Russian. For other languages, there is no clear pattern.

High affinities between locations that are not within the same national or subnational unit may indicate an interesting connection between the locations. For example, the second highest affinity for the Flickr dataset is between Mauritius and Réunion, two nearby islands with similar histories. Other, not as easily explainable, examples from

the list of the highest 100 affinities are Hilo (Hawaii)→Vancouver, Glasgow→Miami, and Tunis→Sassari (Sardinia).

8.4 Travel Distance Cost

The ρs for the different clusters, for a single time interval, are shown in Fig. 4. We have obtained clusters that are different from one another: for the Flickr dataset, one cluster (Cluster 1 in Fig. 4) assigns much higher probability to mid-range distances than the other clusters, and one cluster (Cluster 2) assigns them lower probability than the other clusters. In the Shanghai taxi dataset, the clusters have different probabilities of staying roughly in place: the ρ values for the 5 different clusters for a distance smaller than 3km and a time interval of 1 hour are 5.8, 4.8, 7.7, 4.6, and 5.0; for the location centered at 31.22°N 121.46°E (selected to serve as an example), the corresponding probabilities under the five cluster models are 0.32, 0.25, 0.76, 0.24, and 0.22.

9   Conclusions

We have described several latent factor models of human travel. In contrast to previous approaches to modeling human travel data, our model represents interpretable properties of human travel: travel distance cost, desirability of destinations, and affinity between locations. The resulting models exhibit significant improvements in predictive power over previous methods, while also using far fewer parameters than histogram-based methods.

The parameters of our model are interpretable. We are able to describe the propensity to travel long distances of different clusters of individuals. We contrast the popularity and the desirability of locations, and show that they are distinct. By "factoring out" the desirability of locations and the distance to them, we can analyze the affinities between locations. Our model successfully learns the fact that locations that are located within the same country have higher affinity than the average. We explore the affinities of locations located across national borders, but which share an official language. We see that for pairs of locations where the official language is Spanish and Russian, the affinities are higher than average.

Statistical modeling of human society has become a major theme in recent years, yielding both new insights and predictive abilities. Travel is one of the pieces of this puzzle, and the statistical models here can be used in concert with models of other properties. For example, when modeling photo-collections in online datasets, the speed and location of travel will be correlated to the types of pictures taken, as well as to individual identity, both of which can be useful cues for recognition and tagging of image content (see, e.g., [16]). Furthermore, the use of latent models can be useful for making predictions in new situations. For example, the effects of world events on world travel could be predicted by adjusting the latent factors accordingly, which could be useful in epidemic forecasting [6, 15].
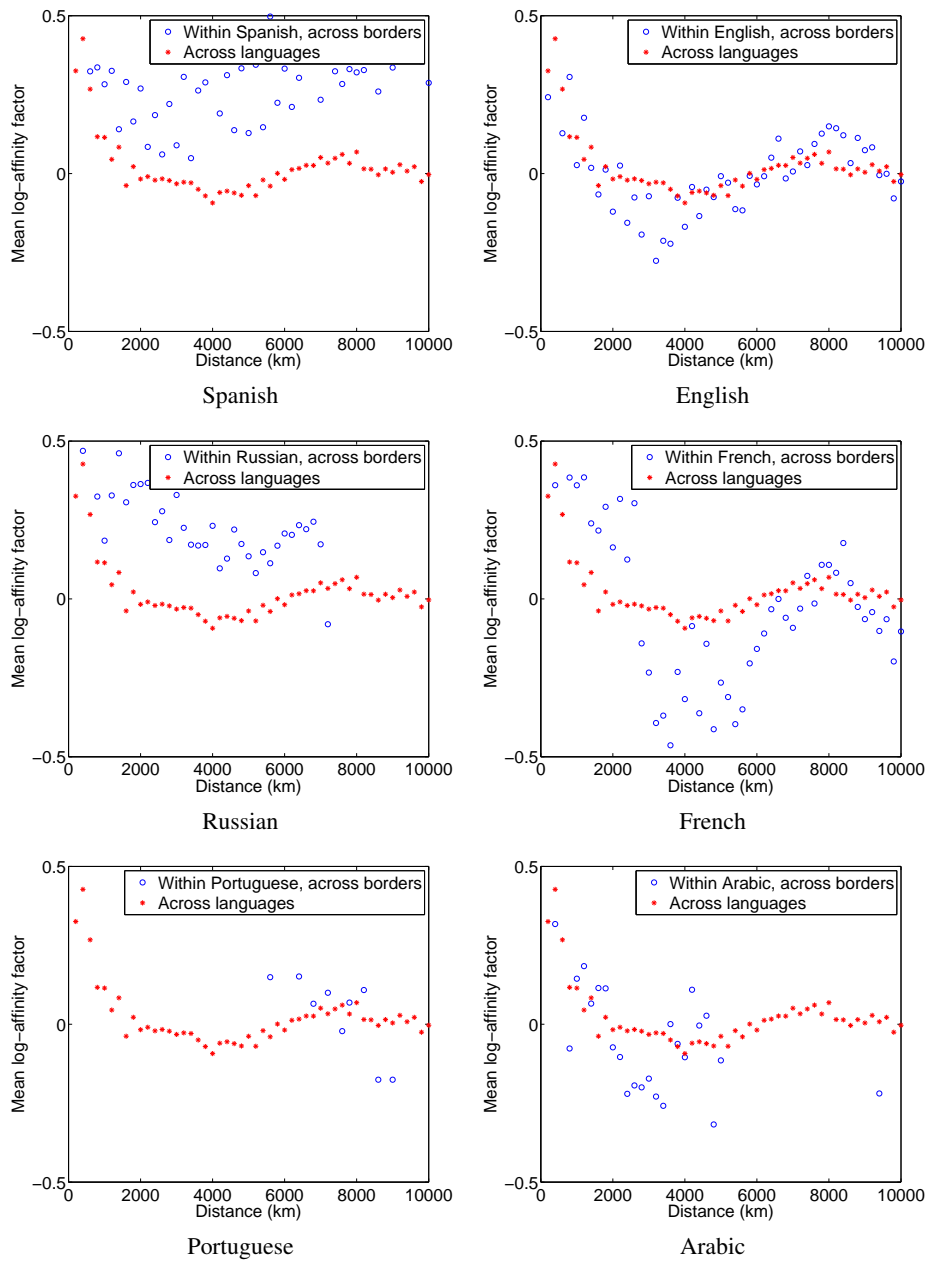
Fig. 6. Mean log-affinity factors for individual official languages, compared to the mean log-affinity factors for locations where the official languages are different. Note that the data is very noisy for larger distances because the mean is based on very few datapoints.

References

1. D. Brockmann, L. Hufnagel, and T. Geisel. The Scaling Laws of Human Travel. *Nature*, 439(7075):462–465, 2006.
2. F. Calabrese, G. Di Lorenzo, L. Liu, and C. Ratti. Estimating Origin-Destination Flows Using Mobile Phone Location Data. *Pervasive*, 10(4):36–44, Apr. 2011.
3. F. Calabrese, F. Pereira, G. DiLorenzo, L. Liu, and C. Ratti. The geography of taste: analyzing cell-phone mobility and social events. In *Proc. Pervasive*, 2010.
4. C.-Y. Chen and K. Grauman. Clues from the Beaten Path: Location Estimation with Bursty Sequences of Tourist Photos. In *Proc. CVPR*, 2011.
5. E. Cho, S. A. Myers, and J. Leskovec. Friendship and Mobility: User Movement in Location-Based Social Networks. In *Proc. KDD*, 2011.
6. V. Colizza, A. Barrat, M. Barthelmy, A.-J. Valleron, and A. Vespignani. Modeling the Worldwide Spread of Pandemic Influenza: Baseline Case and Containment Interventions. *PLoS Med*, 4(1), 2007.
7. D. Dearman, T. Sohn, and K. N. Truong. Opportunities Exist: Supporting the Development of Spatial Knowledge with Continuous Place Discovery for Activities. In *Proc. CHI*, 2011.
8. N. Eagle and A. Pentland. Reality mining: sensing complex social systems. *Pers Ubiquit Comput*, 10:255–268, 2006.
9. F. Girardin, F. Calabrese, F. Fiore, C. Ratti, and J. Blat. Digital Footprinting: Uncovering Tourists with User-Generated Content. *Pervasive Computing*, 7(4):36–43, 2008.
10. F. Girardin, F. D. Fiore, C. Ratti, and J. Blat. Leveraging explicitly disclosed location information to understand tourist dynamics: a case study. *J. of Location Based Services*, 2(1):41–56, Mar 2008.
11. M. C. González, C. A. Hidalgo, and A.-L. Barabási. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, Jun 2008.
12. M. Guerzhoy and A. Hertzmann. Learning latent factor models of human travel. In *NIPS Workshop on Social Network and Social Media Analysis: Methods, Models and Applications*, 2012.
13. P. D. Hoff. Multiplicative latent factor models for description and prediction of social networks. *Comp. & Math. Org. Theory*, 15:261–272, 2009.
14. E. Horvitz, J. Apacible, R. Sarin, and L. Liao. Prediction, Expectation, and Surprise: Methods, Designs, and Study of a Deployed Traffic Forecasting Service. In *Proc. IJCAI*, 2005.
15. L. Hufnagel, D. Brockmann, and T. Geisel. Forecast and control of epidemics in a globalized world. *PNAS*, 101(24):15124–15129, Oct. 2004.
16. E. Kalogerakis, O. Vesselova, J. Hays, A. Efros, and A. Hertzmann. Image Sequence Geolocation with Human Travel Priors. In *Proc. ICCV*, 2009.
17. T. Kurashima, T. Iwata, G. Irie, and K. Fujimura. Travel route recommendation using geotags in photo sharing sites. In *Proc. CIKM*, 2010.
18. R. M. Neal and G. E. Hinton. A view of the EM algorithm that justifies incremental, sparse, and other variants. In M. I. Jordan, editor, *Learning in Graphical Models*, pages 355–368. Kluwer Academic Publishers, 1998.
19. D. Patterson, L. Liao, D. Fox, and H. Kautz. Inferring High-Level Behavior from Low-Level Sensors. In *Proc. UBICOMP*, 2003.
20. C. Peng, X. Jin, K.-C. Wong, M. Shi, and P. Liò. Collective Human Mobility Pattern from Taxi Trips in Urban Area. *PLoS ONE*, 7(4), 2012.
21. A. I. Schein, L. K. Saul, and L. H. Ungar. A Generalized Linear Mmodel for Principal Component Analysis of Binary Data. In *Proc. AISTATS*, 2003.
22. T. Sohn et al. Mobility Detection Using Everyday GSM Traces. In *Proc. UBICOMP*, 2006.
23. J. Yuan, Y. Zheng, and X. Xie. Discovering regions of different functions in a city using human mobility and pois. In *Proc. KDD*, 2012.