Learning Latent Factor Models of Human Travel

Michael Guerzhoy, Aaron Hertzmann

Learning Latent Factor Models of Human Travel

Michael Guerzhoy[1,3] **Aaron Hertzmann**[2,3]
[1]LEAR, INRIA Grenoble [2]Adobe Systems, Inc [3]University of Toronto

Abstract

This paper describes probability models for human travel, using latent factors learned 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. The resulting models exhibit significant improvements in predictive power over previous methods, while also using far fewer parameters than histogram-based methods. The method is demonstrated on travel data from two sources: geotags from a social image sharing site (Flickr), 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, 14]; surveying tourism [9, 10], traffic [2, 13], and special events mobility for urban planning [3]; geolocating with computer vision [4, 15]; interpreting activity from movements [18, 21, 22]; and recommending travel [7, 16]. Analysis of travel data can also play a role within the larger context of analysis of social network data.

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, 14, 15]. 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., [13, 18]). More recent work has explored the value of social network data for predicting mobility [5].

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 [12]. 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.

2 Problem Statement and 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:

$$P_{ij\tau} \equiv P(L_{next} = j | L_{current} = i, \Delta T = \tau) \tag{1}$$

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 Earth.

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:

$$p(d) \propto d^{-b} \tag{2}$$

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, 14, 15]. For example, Kalogerakis et al. [15] 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

$$P_{ij\tau} = \frac{N_{ij\tau} + \lambda q_{d_{ij}\tau}}{\sum_\ell (N_{i\ell\tau} + \lambda q_{d_{i\ell}\tau})} \tag{3}$$

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.

In this work, 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. [15]; 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. Since we are concerned only with the travel, we ignore the images themselves, and 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.

The second dataset comprises taxi travel data, collected by Peng et al. [19]. The dataset includes taxi trips taken on Feb. 1, 2007. A trip 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.

3 Basic Model: Distance Cost and Desirability

The model can be motivated by observing the histograms in Fig. 1, which suggest 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

Paris Honolulu, HI Kansas City, MO

Figure 1: Example travel histogram data ($N_{ij\tau}/\sum_\ell N_{i\ell\tau}$), for three starting points, in a log scale (from [15]). Statistics come from 6 million georeferenced Flickr photographs. Note the significant spatial structure in the histograms: nearby locations are more likely, but are modulated by destination popularity—New York City is a popular destination for each starting location.

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:

$$P_{ij\tau} = \frac{r(d_{ij},\tau)\, a_j}{\sum_\ell r(d_{i\ell},\tau)\, a_\ell} \tag{4}$$

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. 2). 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.

We find it more convenient to parametrize the model in the log domain, particularly when we generalize the model in later sections. In particular, we define the parameters $\rho(d,\tau)$ and α_j as:

$$\rho(d,\tau) = \log r(d,\tau) \tag{5}$$

$$\alpha_j = \log a_j \tag{6}$$

We now write the log-probability of traveling to location j from location i as:

$$\log P_{ij\tau} \doteq \rho(d_{ij},\tau) + \alpha_j \tag{7}$$

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. Our method can be viewed as a form of "logistic PCA" [20] 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 $\boldsymbol{\theta} = [\mathrm{vec}(\rho)^T, \mathrm{vec}(\alpha)^T]^T$, and $\boldsymbol{\phi}_{ij\tau}$ is a vector of indicator variables.

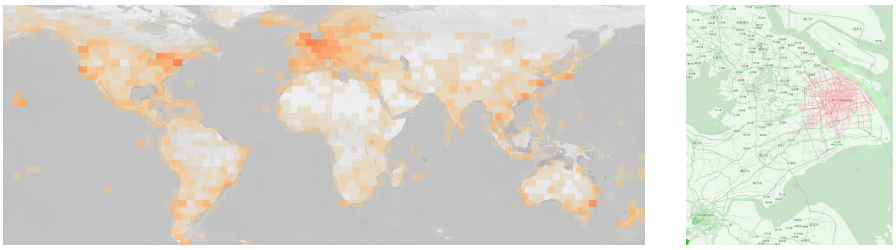
For our problem, the size of the dataset dwarfs the number of model parameters, which suggests that Maximum Likelihood should be sufficient for learning. We estimate the α and ρ parameters from the data, by minimizing the negative log-likelihood with the conjugate gradient method. The negative log-likelihood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

$$NLL = -\ln \prod_{ij\tau} (P_{ij\tau})^{N_{ij\tau}} = -\sum_{ij\tau} N_{ij\tau} \ln P_{ij\tau} \tag{8}$$

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.

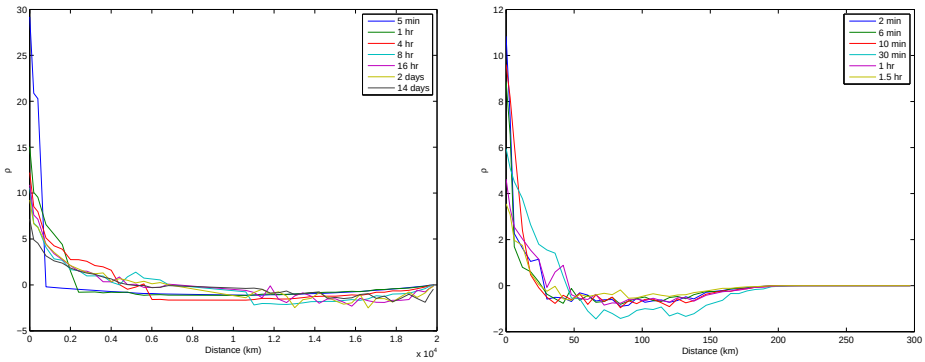
The learned models are shown in Figures 2 and 3[1]. Both the estimated desirability and distance factor functions mimic their histogram counterparts in form: the desirability is similar to the actual

[1]Due to the way we quantize the map, it is sometimes the case that $n(i,d)$, the number of of map quads that are distance d away from i, is not piecewise-monotonic. This introduces artifacts into the plots of ρ. To compensate for this, we plot ρ for the subset of the distances such that $n(i,d)$ is piecewise-monotonic



Flickr dataset Shanghai taxi dataset

Figure 2: The learned desirability of destinations in the Flickr and Shanghai taxi datasets.



Flickr dataset Shanghai taxi dataset

Figure 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.

travel histogram N_i, and the probabilities derived from the distance factor functions exhibit a heavy-tailed shape for large-enough time intervals (note that the probability of traveling distance d from location i is computed by summing the transition probabilities from i to each location j that is a distance d from i for the appropriate time interval, and so the shapes of $r(d, \tau)$ and $P_{d,\tau}$, the probability of traveling distance d in time interval τ, differ). However, the estimated desirabilities are more uniform than the actual travel histogram N_i. The travel histogram measures actual popularity of locations in terms of number of visits, whereas the desirability can be interpreted as a normalized measure of the number of transitions.

4 Location Affinities

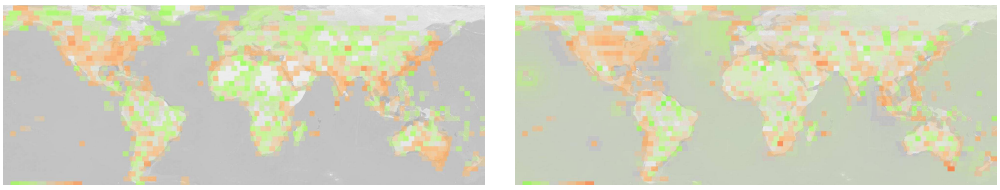
We expect that similarities between locations would play a role in travel. For example, we expect that people are more likely to stay in their own countries than to cross borders, and, when leaving their country, they are more likely to visit countries where the same language is spoken as in their home country. We can express this by assigning a low-dimensional latent parameter vector $\mathbf{x}_j$ to each location j. A low-dimensional vector is used, typically $\mathbf{x}_j \in R^4$. The transition probability is then given by:

$$\log P_{ij\tau} \doteq \rho(d_{ij}, \tau) + \alpha_j + \mathbf{x}_i^T \mathbf{x}_j \tag{9}$$

We call this the symmetric affinity model. Note that this model generalizes some forms of spectral graph analysis, which decompose probabilities as $\log P_{ij} \doteq \mathbf{x}_i^T \mathbf{x}_j$. The model is now log-bilinear, and thus no longer a special case of logistic regression.

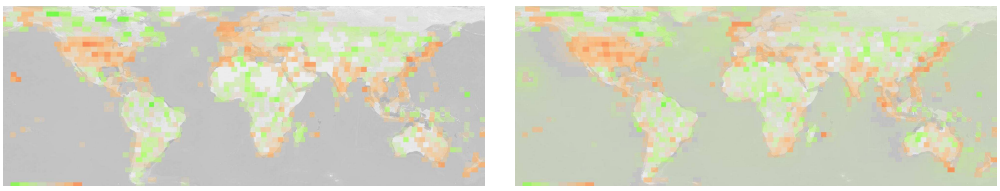
More generally, one can use an asymmetric model, with separate latent parameters $\mathbf{u}$ and $\mathbf{v}$ for each location:

$$\log P_{ij\tau} \doteq \rho(d_{ij}, \tau) + \alpha_j + \mathbf{u}_i^T \mathbf{v}_j \tag{10}$$



Transitions to Paris | Transitions from Paris

Figure 4: The log of the affinity factors ($\mathbf{u}_i^T \mathbf{v}_j$) for transitions from and to the Paris region. Green means the log-affinity factor is negative and the transition probability is downweighted. Orange means the log-affinity factor is positive. Note that London and most of France, the transitions to and from which are very frequent, are nevertheless downweighted, presumably because Paris attracts travelers from far away.



Transitions to Kansas City | Transitions from Kansas City

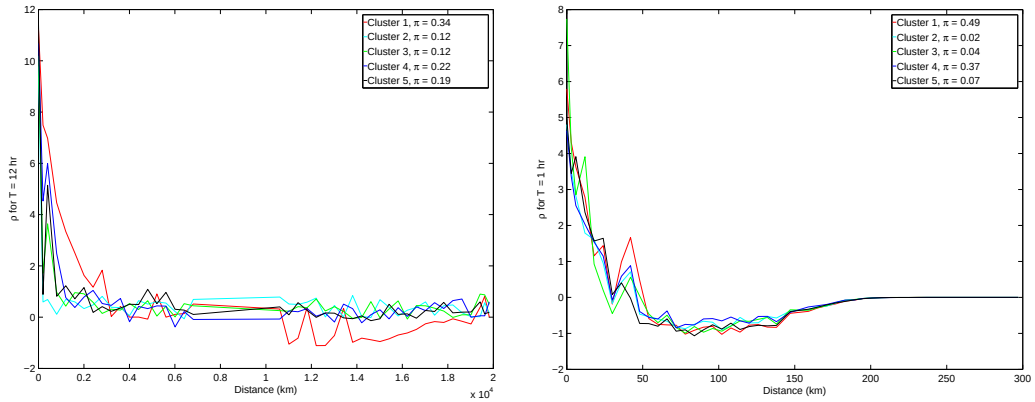
Figure 5: The log of the affinity factors ($\mathbf{u}_i^T \mathbf{v}_j$) for transitions from and to the Kansas City, MO region. Green means the log-affinity factor is negative and the transition probability is downweighted. Orange means the log-affinity factor is positive.

This model is similar to the ones used for social network modeling [12]. Note that the desirability parameter α is not strictly speaking required, because it is redundant with the log-affinity factor term $\mathbf{u}_i^T \mathbf{v}_j$, since it can be mimicked by adding a 1 entry to all $\mathbf{u}$ vectors. The learned affinity factors for Paris are shown in Fig. 4. The learned affinity factors for Kansas City, MO are shown in Fig. 5.

We find that the symmetric model behaves somewhat poorly in experiments, and we instead recommend the asymmetric model. We give two explanations as to why the asymmetric model might do better.

Consider first fitting the baseline symmetric model (Eq. 9). One possible imperfection of the model is that, while we fit ρ in a location-insensitive way, in fact, we would expect the ρ to be somewhat different depending on the location: it is easier to get farther away faster if there is an airport nearby, for example. The affinity factor can compensate for the effectively different ρs. Consider a major airport hub and a town within driving distance of it. For trips starting at the hub, we would like to downweight the probability of nearby locations, and upweight the probability of faraway locations. For the town, the opposite is the case. Asymmetric affinity factors can compensate for the different travel costs.

A second intuition is as follows. We observe that travel between locations is balanced, in the sense that, for locations i and j, very nearly the same amount of transitions occurs between i and j and between j and i. Consider two locations, i and j, with desirability factors $a_i > a_j$ and total outgoing traffic O_i and O_j, considering only a single time interval. Under the symmetric model, if there are several thousands of locations, the normalizing factors for the transition probabilities are similar for locations with similar characteristics, and so, under the symmetric model, we have that the expected number of transitions from i to j is proportional to $O_i a_j$ and the number of transitions from j to i is proportional to $O_j a_i$, so that the expectations are roughly equal if $O_i/O_j = a_i/a_j$. We find that O and a are not perfectly correlated (the correlation coefficient is roughly 0.9 for the Flickr dataset), and so we may want for the affinity factors to correct the model. The affinity factors must be asymmetric in order to be able to help.



Flickr dataset Shanghai taxi dataset

Figure 6: 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.

5 Clustering Individuals

Travel varies between individuals [11]. We incorporate this observation by clustering individuals into C clusters. Each cluster has its own distance factor functions $\rho^{(c)}$. We 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:

$$P^{(c)}_{1:K} = P(L_1) \prod_{k=2}^{k} P^{(c)}(L_k | L_{k-1}, \tau_{k-1}, c) \tag{11}$$

where $P^{(c)}$ corresponds to $P_{ij\tau}$ in Eq. 10, as parametrized by cluster c. We use a uniform distribution over bins for the start probability $P(L_1)$.

Because the cluster assignment is a hidden variable, we must now consider the probability of an individual's entire trajectory of K locations, marginalized over the possible cluster assignments:

$$P_{1:K} = \sum_c \pi_c P^{(c)}_{1:K} \tag{12}$$

We again learn the model using Maximum Likelihood to estimate the parameters of each cluster, along with the cluster probabilities π. Learning is performed using a generalized EM algorithm [17], using exact E-steps to update cluster assignment probabilities, and gradient descent M-steps to update all other parameters. We also experimented with using hard assignments (i.e., using ICM instead of EM), but found that this produces inferior performance. Some learned ρ values for the different clusters are shown in Fig. 6.

In order to apply the learned model to a new individual, we must first observe a sequence of their transitions, in order to determine their cluster probabilities. In particular, given a sequence of transitions L, the cluster c probability for the individual is

$$P(c | L_{1:K}) = \frac{P^{(c)}_{1:K}}{\sum_x P^{(x)}_{1:K}} \tag{13}$$

Then, the transition probability for that individual's next transition is given by:

$$P(L_{K+1}|L_{1:K}) = \sum_c P(L_{K+1}, c|L_{1:K}) = \sum_c P^{(c)}(L_{K+1}|L_K)P(c|L_{1:K}) \tag{14}$$

6 Results

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. Throughout, we set the dimensionality of the u and v vectors to 4. We discretize time into 23 intervals. 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. 12); 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. 8:

$$NLL = -\sum_{ij\tau} N_{ij\tau} \ln \sum_c \pi_c P_{ij\tau}^{(c)} \tag{15}$$

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. 8).

As shown in Table 1, 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 model, indep. transitions"), 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	Parameters	Train NLL	Test NLL
Histogram (Eq. 3)	26,332,700	8.32e5	5.27e5
Regularized Histogram (Eq. 3)	26,332,700	9.16e5	2.40e5
Latent model (Eq. 4)	3,370	1.05e6	2.23e5
Affinity model (Eq. 10)	11,930	1.00e6	2.13e5
Cluster model (Eq. 12)	21,130	9.73e5	2.08e5
Cluster model, indep. transitions (Eq. 15)	21,130	1.00e6	2.15e5

Table 1: Flickr dataset results. The number of parameters is calculated by excluding parameters that pertain to map quads that are never visited.

We use 737 taxis from the Shanghai taxi dataset for training and 1018 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 2. As before, our parametric models outperform the histogram models, with the Cluster model giving the best results.

Model	Parameters	Train NLL	Test NLL
Histogram (Eq. 3)	1,004,692	1.74e5	5.21e5
Regularized Histogram (Eq. 3)	1,004,692	1.85e5	2.84e5
Affinity model (Eq. 10)	3,802	1.95e5	2.75e5
Cluster model (Eq. 12)	9,002	1.71e5	2.36e5
Cluster model, indep. transitions (Eq. 15)	9,002	1.94e5	2.76e5

Table 2: Shanghai taxi dataset results. 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.

The ρs for the different cluster, for a single time interval, are shown in Fig. 6. We have obtained clusters that are different from one another: for the Flickr dataset, one cluster (Cluster 1 in Fig. 6) 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.

Inspecting the affinity factors $\exp(\mathbf{u}_i^T \mathbf{v}_j)$ for source i and destination j, we find that they do indeed seem to capture some of the differences between locations. For example, consider the location pairs that have at least 10 transitions between them in the Flickr dataset. Out of the 100 pairs with the highest affinities, 91 are within the same country and 29 are within the same sub-national entity (e.g., the same US state). Out of the 100 location pairs with the lowest affinities, only 14 are within the same country and 3 within the same sub-national entity. The lowest affinities indicate travelers' tendency to stay within the same country. The probabilities of transitioning from Toronto to New York and from Paris to London are downweighted by the affinity factors (even though they are still quite large; in fact, Paris $\rightarrow$ London is one of the most common transitions in the dataset). Note, however, that this is not a universal tendency. For example, the affinities of Paris and most of the other locations in France are negative, presumably because Paris attracts many travelers from far away.

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).

7 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.

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., [15]). 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, 14].

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] P. D. Hoff. Multiplicative latent factor models for description and prediction of social networks. *Comp. & Math. Org. Theory*, 15:261–272, 2009.

[13] 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.

[14] L. Hufnagel, D. Brockmann, and T. Geisel. Forecast and control of epidemics in a globalized world. *PNAS*, 101(24):15124–15129, Oct. 2004.

[15] E. Kalogerakis, O. Vesselova, J. Hays, A. Efros, and A. Hertzmann. Image Sequence Geolocation with Human Travel Priors. In *Proc. ICCV*, 2009.

[16] T. Kurashima, T. Iwata, G. Irie, and K. Fujimura. Travel route recommendation using geotags in photo sharing sites. In *Proc. CIKM*, 2010.

[17] 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.

[18] D. Patterson, L. Liao, D. Fox, and H. Kautz. Inferring High-Level Behavior from Low-Level Sensors. In *Proc. UBICOMP*, 2003.

[19] 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.

[20] 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.

[21] T. Sohn et al. Mobility Detection Using Everyday GSM Traces. In *Proc. UBICOMP*, 2006.

[22] J. Yuan, J. Luo, , and Y. Wu. Mining Compositional Features From GPS and Visual Cues for Event Recognition in Photo Collections. *Trans. Multimedia*, 12(7), Nov. 2010.