A Non-Parametric Generative Model for Human Trajectories

Kun Ouyang[1,2]**, Reza Shokri**[1]**, David S. Rosenblum**[1]**, Wenzhuo Yang**[2]
[1] National University of Singapore
[2] SAP Innovation Center Singapore
{ouyangk, reza, david}@comp.nus.edu.sg, wenzhuo.yang01@sap.com

Abstract

Modeling human mobility and generating synthetic yet realistic location trajectories play a fundamental role in many (privacy-aware) analysis and design processes that operate on location data. In this paper, we propose a non-parametric generative model for location trajectories that can capture high-order geographic and semantic features of human mobility. We design a simple and intuitive yet effective embedding for locations traces, and use generative adversarial networks to produce data points in this space, which will finally be transformed back to a sequential location trajectory form. We evaluate our method on realistic location trajectories and compare our synthetic traces with multiple existing methods on how they preserve geographic and semantic features of real traces at both aggregated and individual levels. Our empirical results prove the capability of our generative model in preserving various useful properties of real data.

1 Introduction

Modeling human mobility is a challenging task due to many reasons: (1) Human trajectories are extremely high dimensional; (2) the trajectory of each individual is very unique in the geographical space [De Montjoye *et al.*, 2013]; nevertheless, (3) human mobilities are similar in an underlying semantic space, which gives meaning to a trajectory [Bindschaedler and Shokri, 2016]. For example, majority of people leave their respective home in the morning and go to their respective work places, spend time for lunch at their favorite cafe, visit their points of interest, etc. This is not substantially different among individuals; (4) the semantics and geographic features of locations in a trajectory are all correlated with each other. There is also a correlation between the mobility features of different individuals in an area. Thus, it's extremely hard to generate (geographically and semantically) meaningful location trajectories using sequential models, where a location is generated given only a short list of its preceding visited locations. The whole trajectory, instead, needs to be generated in a consistent manner, using a model that captures all its features.

Existing models for the human mobility are mostly sequential. They include Markov Chains [Song *et al.*, 2004; Shokri *et al.*, 2011] and Hidden (semi) Markov Models [Krumm and Horvitz, 2004; Baratchi *et al.*, 2014], and Recurrent Neural Networks [Lin *et al.*, 2017; Song *et al.*, 2016; Gao *et al.*, 2017]. Although such models capture useful patterns in sub-trajectories, they are prone to declining accuracy for large trajectories, when used as generative models. A more fundamental limitations of these methods is that they only model the geographic features of the trajectories. The state of the art in generating location trajectories, while explicitly incorporating their semantic features is the generative model used in the plausible deniability framework [Bindschaedler and Shokri, 2016]. But, this model is also a parametric sequential model, and has a high computational complexity for preserving high-order semantic features.

In this paper, we present a non-sequential non-parametric generative model for human trajectories. We assume locations are discretized, and the map of the considered area is modeled as a two-dimensional matrix, where each element of the matrix is corresponding to one location. We flatten and embed a trajectory in such matrix form, where each cell contains information about the time and duration of visiting that cell in the given trajectory. This could be visualized as a colored trajectory on a map where the color reflects the time of the visit and its intensity captures duration of the visit. We train a generative adversarial neural network to model and then produce data in such representation. We then transfer a generated output to a sequential trajectory. The advantage of this technique is that *all* locations in a trajectory are considered at once by the model, and we do not impose any parametric model that decomposes the locations. The model not only can find the correlation between visited locations in a trajectory, but also can learn the common semantic/geographic patterns of mobilities across the training data.

Via extensive experiments on real data, we demonstrate the efficacy of our model in synthesizing location trajectories that preserve aggregate and individual statistics, such as location inter- and intra-similarity across trajectories, in a real dataset. We show that the model is able to generate trajectories that are indistinguishable from real data.

2 Problem Definition

Let $Traj = \{(x,y)_{t_1}, \ldots, (x,y)_{t_T}\}$ be a location trajectory, where $(x,y)_{t_i}$ is a visit to a location with coordinates (x,y) at time t_i. It is very difficult, if not impossible, to directly model and learn the joint probability distribution $p(Traj)$,
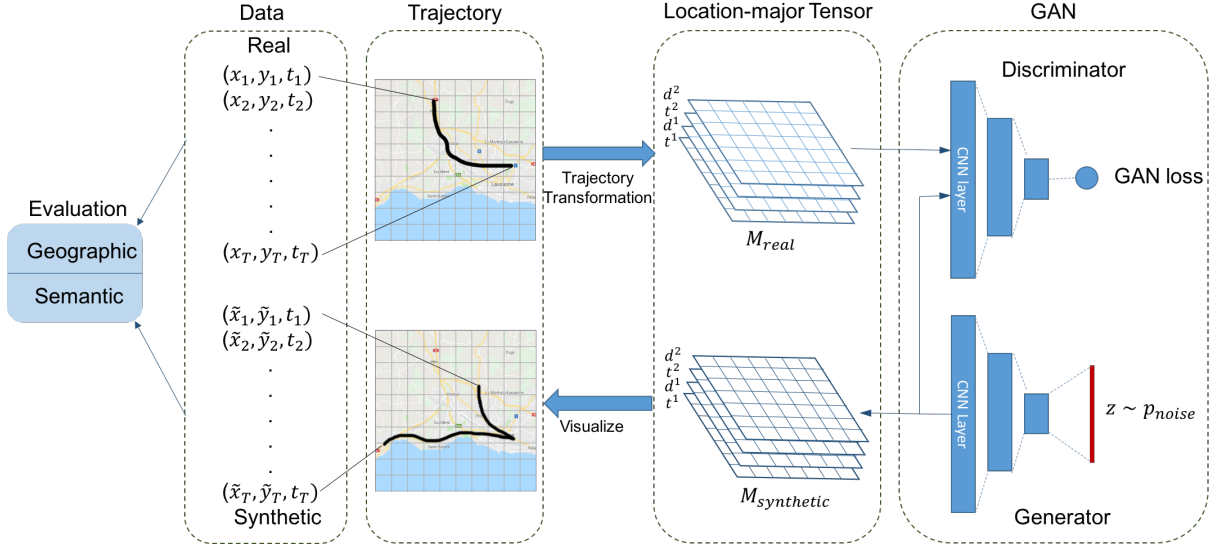
Figure 1: The framework for our generative model. In the training phase, location data is discretized and transformed into a location map tensor following Section 3.1. The discriminator takes M_{real} and $M_{synthetic}$ as input and is simultaneously trained with a generator that is fed with random noise. In the sampling phase, we generate tensors which after post-processing are transformed to location trajectories. We compare the synthetic trajectories with real ones using various geographic and semantic measures.

especially for large T, without making independence assumptions about location visits. One tractable approach is to use a sequential model, which factorizes the joint probability as

$$p((\mathbf{x}, \mathbf{y})_{t=1,\ldots,T}) = p((x,y)_{t_1}) \prod_{t=2}^{T} p((x,y)_t | (x,y)_{<t}) \quad (1)$$

and to use a parametric function to model the conditional probability, $p((x,y)_t | (x,y)_{<t})$. In Markov models (MC, HMM, HSMM), the history $(x,y)_{<t}$ is summarized as $(x,y)_{t-1}$. Recently, Recurrent Neural Network (RNN) models such as the Long Short Term Memory (LSTM) uses a memory matrix and non-linear mappings to learn the conditional probability of moving between locations. However, training LSTM with maximum likelihood estimation suffers from the *exposure bias* problem, where the generated samples might deviate from a realistic path with longer sequences [Ranzato *et al.*, 2015].

In this paper, we represent a trajectory in a slightly different way, by changing the conventional time-major representation to a *location-major* representation. This means that instead of asking *where* a visit happens given the time, we ask *when* and for how long a visit happens (or does not happen) given a location. A mobility sample can be viewed as a "drawing" of a trajectory onto a 2D map, where each pixel represents a specific location, and the value of the pixel is the time and the duration of the visit.

3 Method

Transforming a trajectory to its location-major representation results in a sparse matrix with a large volume. In this section, we present our transformation process. Then, given location data in the location-major representation, we demonstrate our non-parametric generative model for approximating the joint

probability of $p(Traj)$. Figure 1 presents our overall methodology for modeling and generating location trajectories.

3.1 Representation

We discretize the region into a $N_1 \times N_2$ grid where each visit of a trajectory falls into one of these cells. The human trajectory usually includes stationary activities [Lin *et al.*, 2017] where the location is relatively stable for a certain period of time, for example, having lunch, working, etc. Thus, we define these visits as a *stay* $r_{t_i:t_j} = \{(x,y)_{t_i}, \ldots, (x,y)_{t_j}\}$. Using this, we can transform a trajectory into a sequence of stays R, where each $r \in R$ has a location, a start time, and a duration (denoted as d). Note that an individual might have multiple stays at the same location but at different times (of the day). This happens when multiple activities that happen at different locations are interleaved with each other, e.g., working, having lunch, working. However, the frequency of such interrupted stays is small [Park *et al.*, 2017]. It rarely happen that people leave and return to a location for many times. Let K denote the maximum number of stay repetition for each location that we can model in our trajectory embedding. Let M be a matrix of dimension $N_1 \times N_2 \times K$ to be used for embedding $Traj$. We set $M_{x,y,k} = (t, d)$ for the kth stay at location (x, y) at time t for duration d. In Section 4.1, we show that even a small K is enough to encode almost all locations of all trajectories in our dataset.

3.2 Generative Model

Non-parametric generative models, trained in adversarial process, are able to approximate and sample from the joint distribution over input features of complex data [**?**; Goodfellow *et al.*, 2014]. The non-parametric generative model does not assume an explicit parametric function of the target distribution. Instead, it uses a model (e.g., a deep neural network) to

	$JSD(p_{real}, p_{synthetic})$			
	$p(r)$	$p(r,t)$	$p(r,d)$	$p(r, d_{total})$
First-order MC	0.226045 (2.455)	0.377977 (1.898)	0.617615 (2.158)	0.282112 (2.235)
Time-dep MC	0.244647 (4.208)	0.383154 (2.756)	0.617357 (0.216)	0.299368 (3.789)
HMM	0.274112 (5.654)	0.438551 (3.497)	0.619812 (0.624)	0.420264 (0.819)
LSTM-MLE	0.386385 (14.140)	0.555498 (8.778)	0.642768 (6.427)	0.603640 (0.981)
Our method	**0.066109(1.818)**	**0.355129(1.779)**	**0.213122(7.308)**	**0.245102 (2.188)**

Table 1: The mean (standard deviation) of JSDs of four probabilistic measures between real data and the synthetic data generated using different generative models. The standard deviation is scaled by e^{-3}. The result shows our method achieves a significantly lower JS divergences (i.e., higher similarity to real data) across all metrics as compared to the existing models.

learn the target function.

In this work, we make use of generative adversarial networks to train a non-parametric model that generates location trajectories in their map matrix representation from the joint distribution $p(Traj)$. Our model is constituted of Convolution Layers. In each layer, a rectangular filter is used to convolve the feature map from lower layer to generate higher level representation. This structure, as compared to fully connected layer, is able to detect local patterns and is useful in our case since the visit of a particular location is highly correlated to its adjacent locations.

Let P_{real} be the joint distribution of location trajectories from real data. During the training process, the discriminator network $\mathcal{D}_w$ is trained to distinguish between the samples from P_{real} and from $P_{\mathcal{G}(z)}$, where z is a random seed, and $\mathcal{G}$ is the generative model. The model $\mathcal{G}_\theta$ is trained to generate samples that cannot be distinguish from real data by the discriminator. We train our model using the loss function from WGAN-GP [Gulrajani et al., 2017], as it is known to provide stable gradients. The loss function is

$$L = \mathbb{E}_{z \sim P_{\mathbf{z}}}[\mathcal{D}(\mathcal{G}(z)] - \mathbb{E}_{x \sim P_{real}}[\mathcal{D}(\mathbf{x})]$$
$$+ \lambda \mathbb{E}_{\tilde{x} \sim p_{\tilde{x}}}[(\|\nabla_{\tilde{\mathbf{x}}} \mathcal{D}(\tilde{\mathbf{x}})\| - 1)^2]$$

where $\mathcal{G}_\theta$ and $\mathcal{D}_w$ are two neural networks with stacked convolution layers, and the $\tilde{x}$ is drawn from a convex combination of $P_{\mathcal{G}(z)}$ and P_{real}.

4 Evaluation

We evaluate the utility of synthetic data by comparing its similarity to real data on its geographic and semantic features.

4.1 Dataset

We use Nokia Lausanne location trajectories [Kiukkonen et al., 2010], and pre-process them the same way as in [Bindschaedler and Shokri, 2016] to construct the trajectories. Each trace sample contains 288 data points per day (i.e., one location visit per 5 minute). We filter out the region of interest within Lausanne area, which is a rectangular region bounded between $(45.09429, 5.86211)$ and $(47.79493, 10.66721)$ coordinates. We discretized the area into a 64×64 grid. The size of our dataset is $2, 661, 647$ trajectories. We set K, maximum stay repetition, to 4, and transformed trajectories to their map matrix representation.

4.2 Experiments

Baselines

We compare our method with the following baseline models.

- First-order MC [Song et al., 2004]: this model defines the current state as the current location, and assumes $p(r_t|r_0, \ldots, r_{t-1}) = p(r_t|r_{t-1})$ with space-complexity $\mathcal{O}(N^2)$ where N is the number of the total locations, and r the location variable.

- Time-dependent MC model [Shokri et al., 2011]: it assumes the transition probability is time-dependent, by splitting time into multiple time periods (e.g, morning, afternoon, evening).

- HMM [Krumm and Horvitz, 2004]: this model is set up with discrete emission probability and 7 hidden variables according to [Yin et al., 2017] and is optimized using the Baum-Welch [Rabiner, 1989] algorithm.

- LSTM-MLE [Song et al., 2016]: the LSTM model is trained by maximizing log likelihood $log \ p(\theta|Traj)$ where θ is the model parameter, and the joint probability is factorized according to (1). In the training phase, the whole sequence is given, whereas in the generation phrase, the model generates a location given its preceding location (generated by the model).

Geographical Statistics

Instead of directly measuring the quality of the joint distribution of trajectories, which is computationally infeasible for large dimensions, we use three marginal distributions, (i.e. $p(r)$, $p(r,t)$, $p(r,d)$ to evaluate the general performance of the model.

- $p(r)$ measures the visiting probability for a location r, reflecting the popularity of locations.

- $p(r,t)$ measures the visiting probability for a location r at any time t, reflecting the temporal popularity of locations (e.g. bars, homes).

- $p(r,d)$ measures the probability of visiting a location r for a duration d, reflecting the staying patterns in different places.

- $p(r, d_{total})$ measures the probability of visiting a location r for (multiple times) during a day and staying for a total duration of d_{total}, reflecting the overall importance of a location.
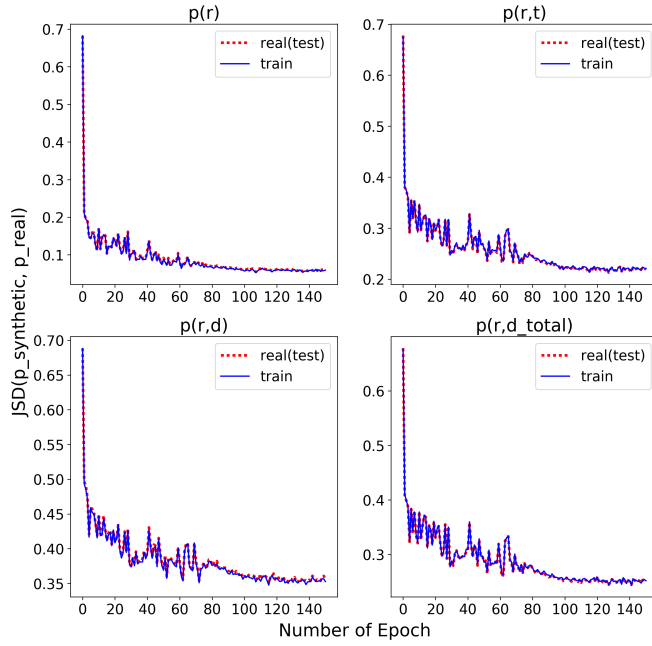
Figure 2: JSD convergence w.r.t. the training epochs. From top left to bottom right, we compute four aggregate statistics, namely $p(r)$, $p(r,t)$, $p(r,d)$, $p(r,d_{total})$ from synthetic samples, and calculate the JSD of these statistics from synthetic samples to real data (test set), and from synthetic samples to training set respectively. We run the experiment for 150 epochs, and notice the convergence at the end. Since the training and testing set are sampled in a balanced way, the difference between JSD score of both sets is small.

From the original dataset, we randomly select 50% trajectories of each user to construct the training set, and another 50% as the test set. To evaluate the distance between the synthetic data and the original real data, we use Jensen-Shannon Divergence (JSD), which is a symmetric measure of the distance between two probability distributions. The JS divergence of two probability P and Q is computed as

$$JS\ divergence(P||Q) = \frac{1}{2}\mathbb{E}_P[\log \frac{P}{X}] + \frac{1}{2}\mathbb{E}_Q[\log \frac{Q}{X}],$$

where $X = \frac{1}{2}(P + Q)$.

During each evaluation period, we generated 2000 synthetic trajectories to compute the $p(r)$, $p(r,t)$, $p(r,d)$ and $p(r,d_{total})$ and the JS divergence between the aggregate statistics over the synthetic data and the real data. Figure 2 shows the results of our comparison. As the optimization (in training the generative model) progresses, the JSD between the distributions of generated data and the real data decreases significantly. The scores at the convergence point are presented in Table 1, where we also compare the quality of our synthetic traces with that of existing methods. The results show that the geographic features of our synthetic data is significantly more similar to real data at the aggregate level, compared to any other sequential model.

Absolute Semantics

Human trajectories reflect the abundant semantics of different locations, which are important features for various machine
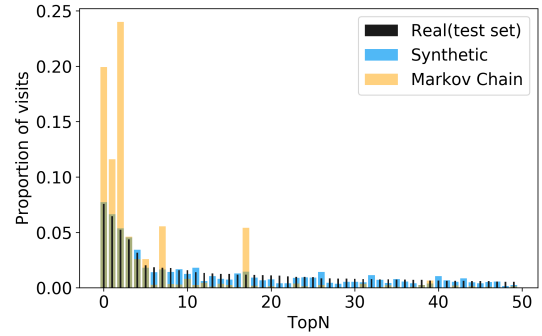
(a) Histogram for proportion of visits in top-50 visited location w.r.t real data
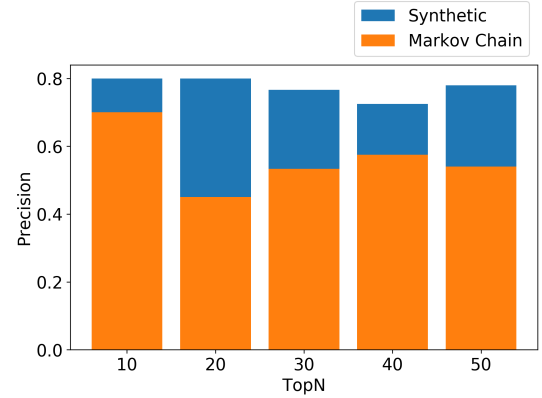

(b) Precision of top-N visited location w.r.t. real data

Figure 3: Normalized histograms showing the absolute semantics preserved by the synthetic data. In (a), we compute the proportion of visit across all locations in the real data, and sort them by popularity. Among the top visited locations computed from real data, we present the respective visit proportion retrieved from the synthetic data and the samples drawn from MC. In (b), we compute the precision for the Top-N visited locations w.r.t. the synthetic set and the MC samples. The precision measures among the Top-N visited locations in real set, i.e., the coverage rate of Top-N visited locations as in the other dataset.

learning and data mining tasks. Due to the high cost of annotation for locations, the labels are not available in our dataset. Nevertheless, some semantic of the location can be told by the way it is visited. We measure the top visited locations, which tells the significance of a specific location [Do and Gatica-Perez, 2014; Zheng *et al.*, 2009]. As is shown in Figure 3, we compare the distribution of visits for top-50 most popular locations, and precision of top-N locations using the synthetic set, real dataset and the samples from a Markov Chain (as the baseline). Both metrics suggest our model outperforms the baseline in capturing popular locations.

Relative Semantics

Various studies show that two trajectories that might differ geographically, in many cases, are semantically similar [Jebara, 2009; Bindschaedler and Shokri, 2016]. The relative semantic similarity relates to the intuition that two trajectories that are semantically similar have close mobility pattern between
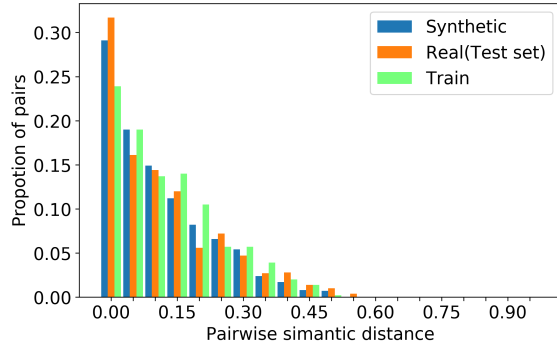
(a) Histograms for pairwise semantic distance within the synthetic, training and testing dataset.
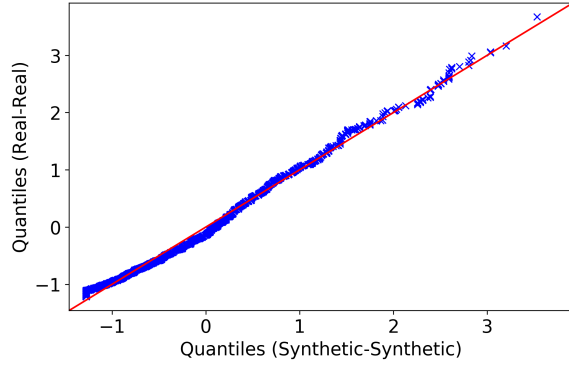
(b) QQ-plot for pairwise semantic distance for synthetic set and test set.

Figure 4: For each dataset (Synthetic, Training and Real set), we randomly select 1000 samples, compute the pairwise semantic distance within the sample set and plot the normalized histogram as in (a). As is expected, the majority of the real (synthetic) trajectories are semantically similar as the pairwise semantic distances are small. In (b), QQ-plot visualizes the relative semantic relationship between the synthetic set and the test set, where the points are aligned at the 45-degree reference line, which illustrates the strong correlation between two distributions.

locations with similar semantics. Since labels are not available in the dataset, we adopt the semantic distance metric in [Bindschaedler and Shokri, 2016], which is:

$$D(Traj_1, Traj_2) =$$
$$\mathbb{E}_{r^1 \sim Traj_1, r^2 \sim Traj_2}[m_d(p_r^{r'_{t+1}}(r^1), p_{\sigma(r)}^{\sigma(r''_{t+1})}(r^2))]$$

where σ is the optimal permutation function that maps semantically similar locations, and m_d is the Mallows distance function. As finding the optimal σ is computationally prohibitive, we use a greedy method to find it, as follows.

$$\sigma(r) = f_{2,j}^{-1}(f_{1,j}(r)),$$
$$where \ f_j(r) = \arg_{r=r_j} \text{sort } p_{d_{total}}(r) \tag{2}$$

More concretely, for an individual trajectory $Traj_i$, we compute $p_{d_{total}}(r) = \frac{d_{total}(r)}{\sum_{r'} d_{total}(r')}$, where $d_{total}(r)$ is the total stay duration in location r, and we interpret d_{total} as the semantic importance of a location. We sort r according to the

weight $p(r)$ in decreasing order, then $\sigma(r)$ is the j-th important location in one trajectory w.r.t. to the other. Then the semantic equivalence between r^1 and $\sigma(r^2)$ lies in that the two locations are both of the j-th importance as for $Traj_1$ and $Traj_2$, respectively. Another difference in our evaluation is that, in [Bindschaedler and Shokri, 2016] the authors use the Earth Mover Distance as a measure of the distance for M_d, instead we use JSD for the reason of consistency.

We denote the relative semantic distance of a pair of trajectory sample as *pair-wise semantic distance*, and the evaluation results are diagrammed in Figure 4. In summary, our model correctly preserves the relative semantic property of trajectories even without explicitly modeling it.

5 Related Work

Various generative probabilistic models have been developed in recent years for approximately modeling the mobility distribution, which allows to simulate human behavior and moving patterns.

The most widely-used models are state-space models. Markov Chain models the location as the state, and the probability of the current state only depends on the previous K-step visits, which results in the order-*K* MC [Song *et al.*, 2004] and time dependent MC [Shokri *et al.*, 2011]. Since native MC models require exponential growth of state space when one expects a finer-grain model, people resorts to Hidden Markov Chain for computational efficiency. HMM assumes the data sequence can be encoded by a sequence of hidden variables **z**, and the transition between hidden variables to be Markovian. Using hidden variables reduces the magnitude of state space, and allows complicated modeling of the posterior $p(\mathbf{z}|x)$ and the likelihood $p(\mathbf{x}|z)$. Recently, Yin et al. [Yin *et al.*, 2017] use Input-Output HMM to model the Call Deal Records, by adding the contextual condition to the transition probability. Bindschaedler et al. [Bindschaedler and Shokri, 2016] categorize the state as the semantics of the visit, and model the transition between different semantics with HMM. The main drawback of HMMs is that they don't completely model the temporal dependency of states. As such, Baratchi et al. [Baratchi *et al.*, 2014] use HSMM to improve the HMMs by including the duration of the state into the hidden variables. Overall, the MC or H(S)MM based requires pre-defined parametric probability of hidden variables, while the learned functions do not necessarily correspond to the real data distribution.

To tackle this problem, [Lin *et al.*, 2017; Song *et al.*, 2016; Gao *et al.*, 2017; Varshneya and Srinivasaraghavan, 2017] take advantage of the modern deep neural networks, and build LSTM models to capture the complicated transition between locations. Deep neural networks are universal approximators, with hierarchal structures to abstract high-level representations, which turn out to outperform H(S)MM-models. However, these works are learned by maximizing log-likelihood, which suffer from the problem of *exposure bias* when generating samples by self-unrolling. In contrast, our model does not depend on likelihood optimization, and is able to generate samples (as a whole) with high fidelity.

6 Conclusion

In this paper we have presented a non-parametric generative model of human mobility for synthesizing realistic human trajectories. To the best of our knowledge, this is the first such work on trajectory synthesis to employ a non-parametric generative model. We have presented experimental results demonstrating that the model preserves not only the statistical characteristics of the trajectories used for training, but also their intrinsic semantics.

In future work we will continue to investigate techniques for incorporating structural and semantic information into the model in order to improve its fidelity to the training data. We also will investigate improvements that will allow the model to be used in applications requiring differential privacy.

Acknowledgements

References

[Baratchi *et al.*, 2014] Mitra Baratchi, Nirvana Meratnia, Paul JM Havinga, Andrew K Skidmore, and Bert AKG Toxopeus. A hierarchical hidden semi-markov model for modeling mobility data. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2014.

[Bindschaedler and Shokri, 2016] Vincent Bindschaedler and Reza Shokri. Synthesizing plausible privacy-preserving location traces. In *Security and Privacy (SP), 2016 IEEE Symposium on*, 2016.

[De Montjoye *et al.*, 2013] Yves-Alexandre De Montjoye, César A Hidalgo, Michel Verleysen, and Vincent D Blondel. Unique in the crowd: The privacy bounds of human mobility. *Scientific reports*, 2013.

[Do and Gatica-Perez, 2014] Trinh Minh Tri Do and Daniel Gatica-Perez. The places of our lives: Visiting patterns and automatic labeling from longitudinal smartphone data. *IEEE Transactions on Mobile Computing*, 2014.

[Gao *et al.*, 2017] Qiang Gao, Fan Zhou, Kunpeng Zhang, Goce Trajcevski, Xucheng Luo, and Fengli Zhang. Identifying human mobility via trajectory embeddings. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, 2017.

[Goodfellow *et al.*, 2014] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, 2014.

[Gulrajani *et al.*, 2017] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron Courville. Improved training of wasserstein gans. *arXiv preprint arXiv:1704.00028*, 2017.

[Jebara, 2009] Tony Jebara. Learning networks of places and people from location data. *MIT media lab*, 2009.

[Kiukkonen *et al.*, 2010] Niko Kiukkonen, Jan Blom, Olivier Dousse, Daniel Gatica-Perez, and Juha Laurila. Towards rich mobile phone datasets: Lausanne data collection campaign. *Proc. ICPS, Berlin*, 2010.

[Krumm and Horvitz, 2004] John Krumm and Eric Horvitz. Locadio: Inferring motion and location from wi-fi signal strengths. In *mobiquitous*, 2004.

[Lin *et al.*, 2017] Ziheng Lin, Mogeng Yin, Sidney Feygin, Madeleine Sheehan, Jean-Francois Paiement, and Alexei Pozdnoukhov. Deep generative models of urban mobility. *IEEE Transactions on Intelligent Transportation Systems*, 2017.

[Park *et al.*, 2017] Souneil Park, Joan Serrà, Enrique Frías-Martínez, and Nuria Oliver. Mobinsight: A framework using semantic neighborhood features for localized interpretations of urban mobility. *CoRR*, abs/1709.10299, 2017.

[Rabiner, 1989] Lawrence R Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 1989.

[Ranzato *et al.*, 2015] Marc'Aurelio Ranzato, Sumit Chopra, Michael Auli, and Wojciech Zaremba. Sequence level training with recurrent neural networks. *arXiv preprint arXiv:1511.06732*, 2015.

[Shokri *et al.*, 2011] Reza Shokri, George Theodorakopoulos, Jean-Yves Le Boudec, and Jean-Pierre Hubaux. Quantifying location privacy. In *Security and privacy (sp), 2011 ieee symposium on*, 2011.

[Song *et al.*, 2004] Libo Song, David Kotz, Ravi Jain, and Xiaoning He. Evaluating location predictors with extensive wi-fi mobility data. In *INFOCOM 2004. Twenty-third AnnualJoint Conference of the IEEE Computer and Communications Societies*, 2004.

[Song *et al.*, 2016] Xuan Song, Hiroshi Kanasugi, and Ryosuke Shibasaki. Deeptransport: Prediction and simulation of human mobility and transportation mode at a citywide level. In *IJCAI*, 2016.

[Varshneya and Srinivasaraghavan, 2017] Daksh Varshneya and G Srinivasaraghavan. Human trajectory prediction using spatially aware deep attention models. *arXiv preprint arXiv:1705.09436*, 2017.

[Yin *et al.*, 2017] Mogeng Yin, Madeleine Sheehan, Sidney Feygin, Jean-François Paiement, and Alexei Pozdnoukhov. A generative model of urban activities from cellular data. *IEEE Transactions on Intelligent Transportation Systems*, 2017.

[Zheng *et al.*, 2009] Yu Zheng, Lizhu Zhang, Xing Xie, and Wei-Ying Ma. Mining interesting locations and travel sequences from gps trajectories. In *Proceedings of the 18th international conference on World wide web*, pages 791–800. ACM, 2009.