Simulating Continuous-Time Human Mobility Trajectories

Nan Xu*[1] Loc Trinh*[1] Sirisha Rambhatla[1] Zhen Zeng[2] Jiahao Chen[2] Samuel Assefa[2] Yan Liu[1]

[1]University of Southern California [2]J.P. Morgan AI Research

ICLR 2021 Workshop Deep Learning for Simulation (simDL)

Background & Motivation

- Understanding how humans navigate in modern day cities is critical for urban planning and location-based services e.g., traffic congestion, disaster management, network support, and epidemic modeling.
- However, it is difficult to gain access to large-scale city-wise mobility trajectory data of high quality in practice due to **privacy** concerns and **limited** availability.
- To better understand human mobility behaviors, learning to **simulate** realistic mobile trajectories has become a major subject of many recent research efforts.



Problem Definition & Objective

Problem:

- Human mobility data contains spatial-temporal trajectories $S = [x_1, x_2, \ldots, x_N]$ where each x_i is a tuple $[t_i, l_i]$ representing a visiting record.
- For each record, $t_i\,[0 \le t_i < 24]$ denotes the i^{th} time-stamp and l_i denotes the location $[lat, lon]$ of the record.
- It is often **intractable** to model the joint distribution $\mathbb{P}(S)$ especially for long sequences with large N, we made the common assumptions to factorize the joint probability ,$\mathbb{P}(S) = \mathbb{P}(x_1)\prod_{t=2}^{N}\mathbb{P}(x_t|x_{1:t-1})$, treating the modelling approach as a sequential process.
- Following recent work (Feng et al., 2020), we discretize GPS coordinates into an $M \times M$ grid L containing up to 3 digits after the decimal point of coordinates. We do not discretize time.

Objective:

We focus on building a comprehensive understanding of generative models for human mobility synthetization.

1. **Distribution Similarity:** Can the proposed continuous time-conditioned location generator produce more realistic sequences in both spatial and temporal aspects?
2. **Application Utility:** Are the underlying dynamics in human mobility data sufficiently captured by our model and reflected in downstream tasks when compared to real-world data?

Method

Continuous-Time Generation:

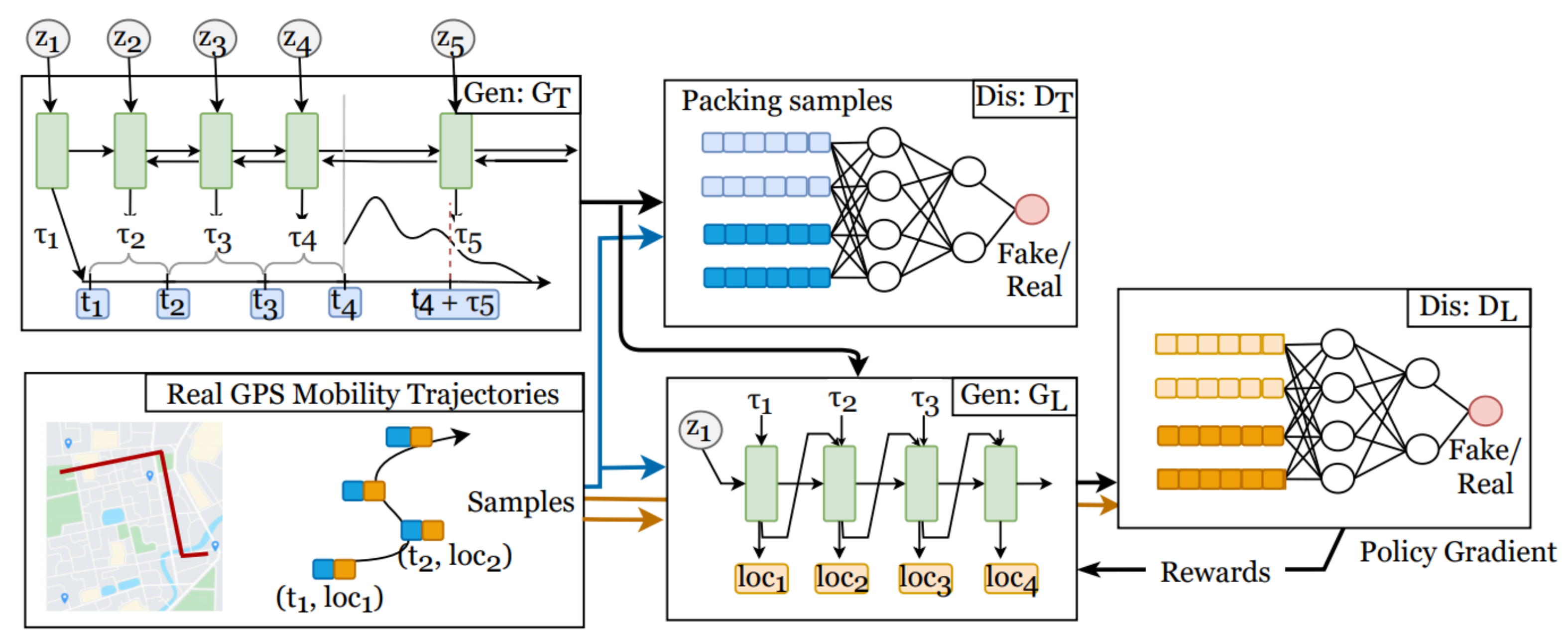
- We view a mobility trajectory as a spatial-temporal point process with each event denoting a person entering a new location. Instead of binning timestamps into large discretized time slots, trajectories are viewed as sequences of events happening at irregular intervals,
- We leverage the off-the-shelf implementation of WGAN-GP with a recurrent generator G_T and an MLP discriminator D_T . G_T takes a sequence of random variables $z = [z_1, \ldots, z_N]$ $z_i \sim \mathcal{N}(0, I)$ and sequentially generates the sequence of duration $[\tau_1, \ldots, \tau_N]$ using a bidirectional LSTM.

Conditional Spatial Generation:

- We train G_L parameterized by θ via policy gradient with the gradient of the expected end reward R_N where the expected cumulative reward Q^{DL} is the estimated probability of being real or fake by the discriminator.

$$\nabla_\theta \mathbb{E}[R_N|l_0] = \sum_{t=1}^{N} \mathbb{E}_{l_t \sim G_L(l_t|L_{t-1}, d_t)}\left[Q^{D_L}(L_{t-1}, l_t)\nabla_\theta \log P_\theta^{G_L}(l_t|L_{t-1}, d_t)\right]$$

Figure 1: The proposed *DeltaGAN* includes a continuous-time generator G_T and a time-conditional location generator G_L with Gaussian noise z, together with their discriminators D_T and D_L.



Evaluations

Dataset:

- We utilize the GPS trajectory dataset collected by MSRA Geolife project from 182 users in a period of over five years (Zheng et al., 2010).
- We keep the trajectories within the 5-th Ring Road of Beijing (50,652 grids covering 39.85 N ~ 40.00 N, 116.25 E ~ 116.50 E).
- There are 11,375 trajectories with 31.531 records on average, and the average daily traveling duration and distance are 1.945 hours and 9.028 km.

Evaluation Metrics:

We adopt the following individual trajectory and geographical metrics (Ouyang et al., 2018; Feng et al., 2020) to evaluate the distribution similarity (Jensen-Shannon divergence) between real and generated mobility data:

1) *Distance*: the daily cumulative travel distance per trajectory;
2) *Radius*: the radius of gyration for a daily trajectory;
3) *Duration*: the total stay duration of each visited location;
4) *DailyLoc*: the number of unique locations in the daily trajectory;
5) *P(r)*: the visiting probability of one location r;
6) *P(r,t)*: the visiting probability of one location r at time t.

Results

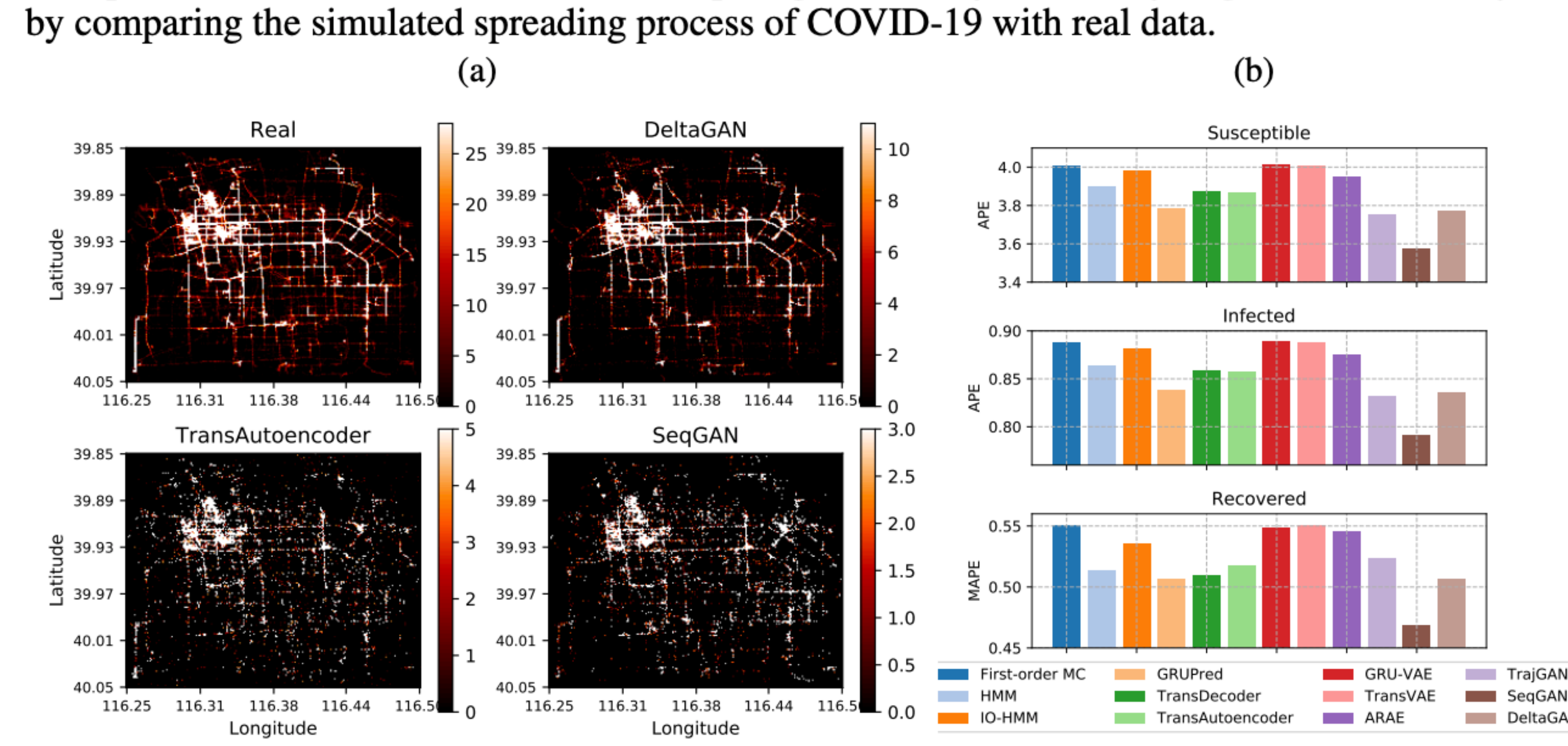
Table 1: Distribution comparison between real and generated mobility data. For all the metrics, lower values indicate more realistic trajectories. We marked the best result with boldface.

		Individual Trajectory Metrics				Geographical Metrics	
	Models	*Distance*	*Radius*	*Duration*	*DailyLoc*	$P(r)$	$P(r,t)$
Markov	First-order MC	0.56113	*0.10059*	0.58858	0.37374	0.43219	0.81836
	HMM	0.45217	0.52043	*0.10166*	0.39246	0.38329	0.82717
	IO-HMM	0.30730	0.15118	0.72849	0.66639	0.60712	0.82690
Deep Prediction Models	GRUPred	0.11441	0.17767	0.25546	0.55544	0.48476	0.82401
	TransDecoder	**0.09735**	0.16273	0.28388	0.56912	0.51261	0.82423
	TransAutoencoder	0.16209	0.22480	0.22952	0.54911	0.47934	0.82441
Deep Generative Models	GRU-VAE	0.82830	0.57407	0.15602	0.71901	0.58838	0.82190
	TransVAE	0.83198	0.67098	0.20954	0.62373	0.51397	0.82079
	TrajGAN	0.82075	0.72006	0.16102	0.42136	0.47586	**0.79298**
	ARAE	0.67968	0.57447	0.60294	0.44594	0.50957	0.82129
	SeqGAN	0.11074	0.16360	0.27096	0.57523	0.57125	0.82806
	DeltaGAN (Ours)	0.10553	**0.06677**	**0.00561**	**0.35276**	**0.30523**	0.80262

COVID-19 Spreading Simulation:

- We run simulations with human mobility data, and calculate (mean) absolute percentage error between real and generated data on the number of different populations (Susceptible, Infected, Recovered).

Figure 2: (a) Geographical visualization of 10,000 real and generated trajectories. We set the first 95% percentile of visit times as the colormap range for clarity. (b) Utility of generated mobility data by comparing the simulated spreading process of COVID-19 with real data.



References

1. Jie Feng, Zeyu Yang, Fengli Xu, Haisu Yu, Mudan Wang, and Yong Li. Learning to simulate human mobility. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 3426–3433, 2020.
2. Oleksandr Shchur, Marin Bilos, and Stephan Gunnemann. Intensity-free learning of temporal point ¨ processes. arXiv preprint arXiv:1909.12127, 2019.
3. Kun Ouyang, Reza Shokri, David S Rosenblum, and Wenzhuo Yang. A non-parametric generative model for human trajectories. In IJCAI, pp. 3812–3817, 2018.
4. Yu Zheng, Xing Xie, Wei-Ying Ma, et al. Geolife: A collaborative social networking service among user, location and trajectory. IEEE Data Eng. Bull., 33(2):32–39, 2010.
5. Lantao Yu, Weinan Zhang, Jun Wang, and Yong Yu. Seqgan: Sequence generative adversarial nets with policy gradient. In Proceedings of the AAAI conference on artificial intelligence, volume 31, 2017.
6. Sepanta Zeighami, Cyrus Shahabi, and John Krumm. Estimating spread of contact-based contagions in a population through sub-sampling. arXiv preprint arXiv:2012.06987, 2020.