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-wide 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 $S = \{x_1, x_2, …, x_N\}$ where each $x_i$ is a tuple $(t_i, l_i)$ representing a visiting record.

• For each record, $(t_0 \leq t_i \leq t_2)$ denotes the $i^{th}$ timestamp and $l_i$ denotes the location for the $i^{th}$ record.

• It is often intractable to model the joint distribution $P(S)$ especially for long sequences with large $N$. We made the common assumptions to factorize the joint probability $P(S) = P(S_1)P(S_2|S_1)P(S_3|S_2)$, 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 $l_i$ 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 and its various applications.

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?

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 $G_r$ and an MLP discriminator $D_r$ to generate sequences of random variables $z = \{z_1, z_2, …, z_N\}$ and sequentially generates the trajectory by a visiting probability of one location $r$ at time $t$.

$$\nabla_p \mathbb{E}[\mathcal{R}|y] = \sum_{d=1}^{D} \mathbb{E}[G_l(x_i \sim \mathcal{L}) - \mathbb{E}[G_l(x_i \sim \mathcal{L})]]$$

Figure 1: The proposed DeMoGAN includes a continuous-time generator $G_r$ and a time-conditioned location generator $G_l$ with Gaussian noise $z$, together with their discriminators $D_r$ and $D_l$.

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.

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 infected populations (Susceptible, Infected, Recovered).

• We set the first 95% percentiles of visit 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


