**Article Title:** Deep ROI-Based Modeling for Urban Human Mobility Prediction

**Link:** <https://sci-hub.tw/10.1145/3191746>

**Summary:**

Using big GPS trajectory data to predict human mobility for transportation scheduling, urban regulation and emergency management. Most urban human behaviors related to a small number of ROI. Then, urban mobility prediction is modeled as a sequence classification problem for each type of label. Then RNN is used as a good sequence classifier. ROI (Region of interest) is defined as a region with high crowd density, therefore at a higher risk of accident. There is therefore a need to allocate more resources to some ROI to prevent emergency situations.

**Key outcomes of this paper:**

* Given a few steps of observed mobility from one person, predicting where this person will go next.
  + Will this person enter any ROI?
  + If yes, which ROI?

**Data characteristics:**

* Collected historical trajectory data over a long period of time.
* Data of large temporal scale
* This data was noisy and sparse

**Aim of the model:**

* Training Input: Multiple days of trajectory data (eg: 1 month)
* Model: goal is to build a 24-h mobility prediction system on these historical data, which iteratively runs in an online updating mode that takes observed human mobility within the current hour as input and predicts how the mobility will be in the next hour.
* Testing Input: Human mobility within the current hour
* Output: Prediction of mobility in the next hour

**Steps of data processing:**

1. An effective mining algorithm to discover urban ROIs given historical raw trajectory data- ROI found for each hour, for each day- 24-hour data- merged to form the ROI for one day.
2. RNN used for mobility prediction

**The data:**

* People flow data sent from mobile phones (latitude, longitude)- collected from y 1.6 million mobile phone users in Japan over a three-year period
* Does not include information like gender/age to specify individuals.
* Contains 30 billion GPS records, and the total size of the data is more than 1.5 terabytes
* Each record contains

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| User ID | Latitude | Longitude | Altitude | Time-Stamp | Positioning accuracy level |

* Position accuracy level: (there are three levels due to different satellite’s signal strength, correspondingly the positioning error would be within 100m, 200m or 300m

Types of data we need: