ツイッターデータによる災害時の人間移動性の予測

Predicting Human Mobility under Disaster with Twitter Data

学籍番号　47-196800

氏　　名　汪 昊宇（Wang, Haoyu）

指導教員　柴崎 亮介 教授

1. Introduction

The rapid urbanization in recent years, due to the growth of the urban population, has a huge impact on urban traffic, which may increase travel demand, and the difficulty of traffic management and crowd forecasting and the risk of deteriorating traffic conditions when disasters such as typhoons occur. Therefore, urban traffic forecasting plays an important role in the process of urban development. It can provide insights for urban planning, traffic management and resource allocation, and help improve urban traffic efficiency and human settlements. However, the occurrence of natural disasters such as typhoons has significantly reduced the accuracy of traffic prediction, because when faced with such a different situation, people's movements will be very different from those in the past.

This study aims to improve the accuracy of mobility flow prediction when natural disasters occur by collecting the number of disaster-related tweets in Twitter. As shown in Figure 1, it is the hourly flow of people and the number of COVID-19-related tweets in Tokyo during the period of COVID-19 from December 15th, 2019 to February 28th, 2021. It can be seen that before the infection increases, Tweets will be significantly increased. This study also aims to use disaster-related tweets as a key condition and refer to Curb-GAN's network structure to train neural networks. Using historical mobility flow data to generate future mobility flow graphs to improve human mobility prediction accuracy when disasters such as typhoons occur.

2. Problem definition

There are the definitions that will be used in the study.

Definition 1 (Regions). The whole Japan area is split into 47 prefectures, denoted as $S = \{s_i\}$, where $1 \leq i \leq 47$.

Definition 2 (Target region). The target region R is the prefectures which currently have corresponding disaster-related tweets. Currently 7 prefectures' disaster-related tweets are collected, so $R = \{s_i\}$, where s_i currently is 8,9,10,11,12,13,14.

Definition 3 (Disaster-related Tweets). The number of disaster-related tweets of a prefecture captures how people react with the disaster in a period of time. Thus, I denote number of disaster-related tweets of a region s in time slot t as $d_t^s \in \mathbb{N}$. Moreover, the number of disaster-

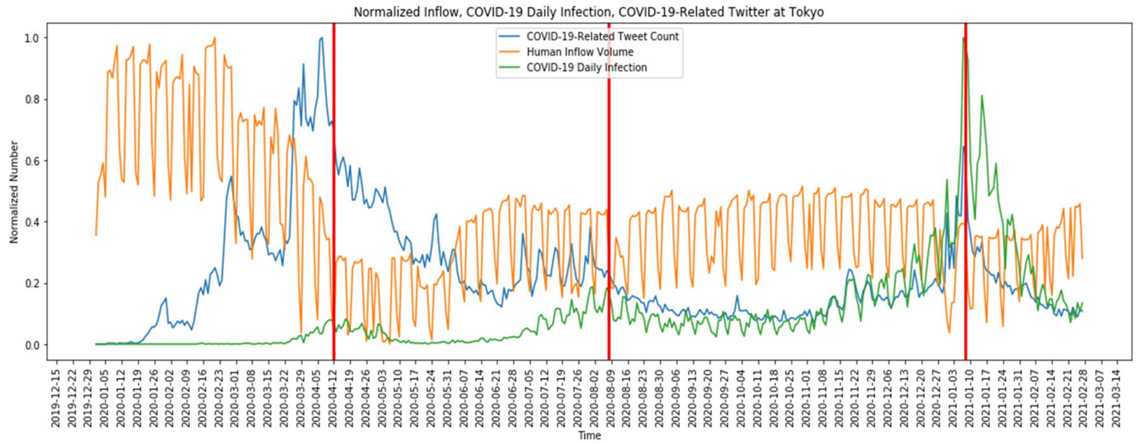

Figure 1: The comparison of Tokyo's daily inflow, tweets and inflection

related tweets of a target region R within a day are denoted as a sequence $D^R = \{d_1^R, \ldots, d_{N_t}^R\} \in \mathbb{N}^{N_t}$, where d_t^R is the sum of number of disaster-related tweets in all prefectures within R in time slot t.

Definition 4 (Traffic status and traffic distribution). Traffic status indicates the quality of traffic, which can be measured by traffic speed, traffic inflow/outflow, traffic volume, etc. I denote m_t^s as the average traffic status of prefecture s in time slot t. The traffic distributions of a target region R within a day is denoted as a tensor $M^R = \{M_1^R, \ldots, M_{N_t}^R\} \in \mathbb{R}^{N_t \times l}$, where we denote the l is the number of all prefectures in the region R. each entry of M_t^R is m_t^s, where $s \in R$.

Definition 5 (Traffic correlation matrix). The traffic correlation matrices of a target region R within a day is denoted as a tensor $A^R = \{A_1^R, \ldots, A_{N_t}^R\} \in \mathbb{R}^{N_t \times l \times l}$, where A_t^R is a traffic correlation matrix of size $l \times l$ in region R in time slot t, A_t^R is non-negative and row-normalized.

Traffic correlations capture the inherent traffic dependencies between a prefecture pair. I use adjacency matrix of each pair of prefectures to quantify their corresponding traffic correlations. Note here I assume that traffic status at different locations are either positively correlated or independent.

Therefore, the problem can be described as that, the selected Japan prefectures are partitioned into regions R, given $\tau_{in} \times \| R \|$ samples of D and M, for one specific target region R, estimate the traffic distributions $\widehat{M}^R = \{\widehat{M}_1^R, \ldots, \widehat{M}_{N_t}^R\}$ in consecutive time slots based on a given number of disaster-related tweets sequence $\widehat{D}^R = \{\hat{d}_1^R, \ldots, \hat{d}_{N_t}^R\}$.

3. Methodology

To solve the conditional mobility flow estimation problem, I refer to the Curb-GAN model structure. Using the number of tweets as a condition, the mobility flow distribution of a region in one time slot can be treated as an image and the flow of each prefecture can be seen as pixel value. Thus, Curb-GAN structure could be potentially used to solve the conditional human

Figure 2: The structure of Curb-GAN

flow estimation problem.

3.1 Dynamic Convolutional Layer (DyConv)

In prefecture areas, the strength of traffic spatial auto-correlations is often heterogeneous, which mostly relies on the spatial relationship between prefecture and prefecture. Based on the First Law of Geography, nearby locations and closely connected prefectures often have stronger spatial auto-correlations. Therefore, applying dynamic convolutional layers in both generator and discriminator, which can better capture the spatial auto-correlations of mobility flow.

3.2 Self-Attention Mechanism (SA)

After applying the dynamic convolutional layer to capture the spatial auto-correlations of region, in order to capture the temporal dependencies, self-attention mechanism, achieves excellent performance when dealing with language modeling and machine translation problems. Self-attention mechanism handles sequential data including text, audios and videos, and learn the temporal dependencies from it. Compared with LSTM and GRU, self-attention mechanism is computed in parallel, and thus requires less time to train and results in higher training quality.

3.3 Curb-GAN Architecture

To provide daily consecutive mobility flow estimations conditioned on expected tweets number, I refer to Curb-GAN structure to make it possible to control the estimations by different travel demands. The structure of Curb-GAN is shown in Figure 2. Curb-GAN contains a generator G and a discriminator D. The generator G aims to generate sequences of mobility flow distributions in consecutive time slots which are similar to the real ones so that the discriminator D cannot distinguish the generated mobility flow distribution sequences from the real sequences well.

4. Evaluation

I first describe the real-world spatio-temporal datasets and then introduce method I use and the evaluation metrics.

4.1 Dataset Descriptions

・ **Japan prefecture-level mobility flow data.**

The inflow/outflow and ODflow datasets between prefectures are collected from Jan 1st, 2018 to Feb 28th, 2021. The datasets are hourly collected, which means over 25000 time slots are included.

- **COVID-19-related tweets data**

We also collect the number of tweets every hour of 47 prefectures from Jan 1st, 2020 to Feb 28th, 2021. We count the number of COVID-19-related tweets per hour and record the dataset by prefecture.

4.2 Evaluation Metric

I use rooted mean square error (RMSE) to evaluate the performance:

$$RMSE = \sqrt{\frac{1}{N_s N_t} \sum_{s=1}^{N_s} \sum_{t=1}^{N_t} (y_{s,t} - \hat{y}_{s,t})^2}$$

where $y_{s,t}$ is the ground-truth mobility flow observed in the s-th prefecture and t-th time slot, and $\hat{y}_{s,t}$ is the corresponding prediction.

4.3 Results

First, I tried the effect of the number of heads of self-attention mechanisms on the results. I tried the results of 1 head, 2 heads and 7 heads respectively. The result shows that when the number of heads of the self-attention mechanism is the same as the number of prefectures trained, the model can better learn the changes in the human mobility in the time dimension.

Then, I took the data of adding 6 dimensions with all 0 dimensions first, until the RMSE of the model to be low enough, then took the method of transfer learning and used the real Twitter data with 6 dimensions added to finetune the model. In the finetune stage, the result shows that a discriminator that is too strong cannot train a model, but if it is too weak it cannot generate a sufficiently realistic human mobility series.

Finally, the model achieves the best RMSE with 7 self-attention heads, 2 discriminator building blocks, and 6 one-hot data tweets. After the introduction of Twitter data, the model also has very good prediction results for the mobility flow decline during the COVID-19 epidemic. What's more, compared with the model without Twitter data, the model with Twitter data was able to predict the decline in mobility flow due to the increase in the number of infected people during this period.

Number of heads	Number of D blocks	Include twitter or not	Testing RMSE
2	1	no	0.2436
2	2	no	0.3501
7	1	no	0.1814
7	2	no	0.2678
7	1	yes	0.1455
7	**2**	**yes**	**0.1363**

Table 1: Performance with different model settings

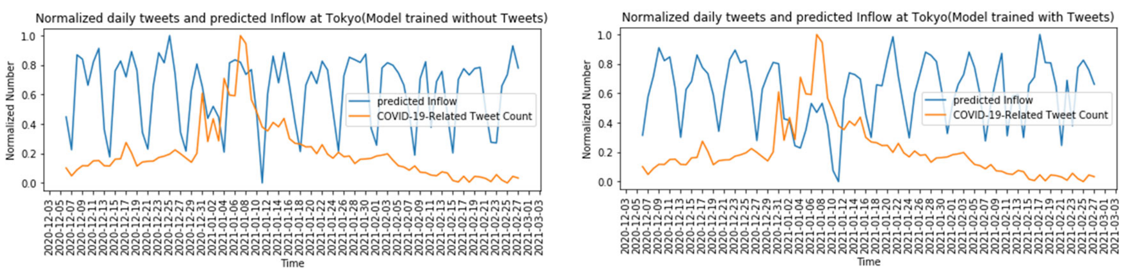

Figure 3: The results of models trained with and without tweets