大きな GPS データを使用して鉄道交通を分析および視覚化するインタラクティブシステム

An Interactive System to Analyze and Visualize Railway Traffic using Big GPS Data

学籍番号　47-206795
氏　　名　JEPH Puneet
指導教員　柴崎 亮介 教授

1 Introduction

Rail transportation system plays an important role in urban development as it is energy-efficient and is relatively time-saving and punctual mode of transportation. A serious issue in railway traffic analysis is the behavior during abnormal events. These can be of two types mainly -(1) Scheduled Events (2) Natural Disasters or Human Accidents. Both types of incidents cause congestion in some stations/lines, or delay/suspension of railway services. Therefore, it becomes important to study and analyze the railway traffic behavior during such kinds of events and unforeseen incidents, to understand their effect on railway passengers.

The human mobility GPS data can be used as an effective data source to evaluate the effect and impact of these abnormal events. The purpose of this research is to analyze railway traffic during such abnormal events [1], and develop a score Index and other indicators to help assess and compare the impact on railway traffic and passengers.

We develop a novel dashboard by integrating both Plotly Dash and Kepler.gl platforms.The dashboard can help visualize the change in congestion of the railway stations during events, and change in railway traffic during an accident, as well as give information about event participants mobility flow and railway commuters.

2 Data and Methodology

2.1 Data sources

For GPS trajectory data, we collaborated with Blogwatcher Inc. to get big human GPS trajectory data that covers almost 5 million people covering 47 prefectures of Japan. The time range of the GPS dataset is 2018-01 to 2022-05. For this study, three popular types of events in Japan are chosen -Horse Race Events, AAA Tokyo Dome Concerts, and Soccer Games. We collected the data from 81 horse race events organized at five different race-course venues and 9 Tokyo Dome events, spanning from 2018 to 2021. Similarly, 741 soccer games from 20 unique venues all across Japan were collected spanning from 2019-02 to 2022-05. Accident data is obtained in collaboration with JR-East (East Japan Railway Company) database.

2.2 GPS Data Preprocessing

The GPS Trajectories labeled as Railway Transportation Mode were extracted from the original dataset. The entry, exit, and transfer stations and their respective time for each user were identi-

fied. Some of the wrongly map matched trajecories where entry/exit stations were mapped to transfer links, were rectified. Then it was aggregated in two manners: (1) Railway Station-wise and (2) Railway Link-wise. Railway station-wise aggregation is basically Entry/Exit aggregation within 1-hour time duration, while railway link-wise aggregation is the aggregation of users traveling through a particular railway link within 30-minutes time duration.

2.3 Event Participant's features extraction

We extract the previous and next trajectory OD information for the passengers entering and exiting the station of interest. The local origin/destination density is defined as the grid-density of the origin/destinations for the passengers entering/exiting the station, respectively. Using GPS trajectories and transportation information, we identify the event participants by detecting the STAY points within the event venue during the event, and also other relevant information like origin prefecture, transportation mode used to arrive at the venue, arrival/departure time, time spent at the venue, etc.

2.4 Railway Commuter Identification

A railway commuter can be defined as a person who travels between a fixed set of home and work stations, and has somewhat set hours of departure and arrival. I have used entropy method [2] to identify railway commuters through their GPS trajectories. Entropy is the measure of disorder in the given distribution. If the there is more disorder, then entropy is higher, otherwise entropy is lower. Then formula for entropy is as follows,

$$E = (-1) \sum \frac{P_i \log_2 P_i}{\log_2 I} \tag{1}$$

where E is the entropy value, P_i is the proportion of i^{th} label in the distribution, and I is the total number of items. The entropy is computed in following steps for the railway passengers:

- Identify home/destination stations, and departure/arrival hours for each railway user
- Compute the Entropy of distribution of home/destination stations and departure/arrival hours for each railway user, denoted respectively as $EStation$ and $EHour$
- K-means clustering based on multi-dimensional features of $EStation$, $EHour$, and $ndaysr$ (normalized no. of days of railway travel in a month)
- Label the clusters having low level of entropy as potential railway commuters

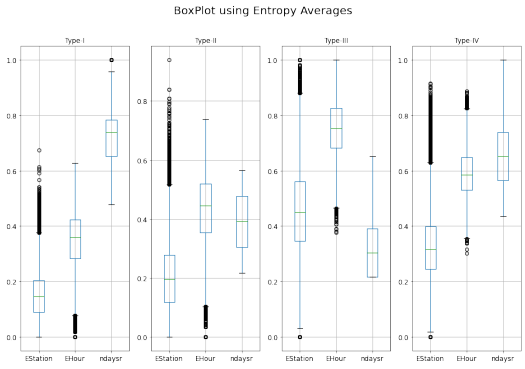

Figure 1: Boxplots of the entropy features for the four clusters of railway passengers

Users having low entropy levels and high days of travel may be considered as railway commuters. Therefore, as shown in fig. 1, I have categorized Type-I and Type-II clusters as Commuter-

A and Commuters-B respectively. While Type-III and Type-IV have been categorized as non-commuters.

2.5 Score Indicators

2.5.1 Event Congestion Index

This index can help in assessing the relative scale of congestion in railway traffic compared to past average congestion. We define the congestion index separately for entry and exit numbers of the stations. They are defined as the following:

$$C_{Entry} = W_i \sum_{i=1}^{N} \left(\frac{I^{o}_{peak}}{\frac{1}{5}\sum_{m=1}^{5} I^{\text{-m}}_{peak}} \right) \quad (2)$$

$$C_{Exit} = W_i \sum_{i=1}^{N} \left(\frac{O^{o}_{peak}}{\frac{1}{5}\sum_{m=1}^{5} O^{\text{-m}}_{peak}} \right) \quad (3)$$

where C_{Entry} and C_{Exit} denote the Congestion Index for entry and exit numbers respectively, W_i is the weight for i^{th} station, I, O denote get-on and get-off numbers, respectively, N denotes the number of stations selected for that venue, and m denotes the number of past days taken for comparison. A congestion index value of 1 will indicate no change in railway congestion compared with past five days, while a value of more than 1 will show an increase in the congestion.

2.5.2 Accident Related Indicators

We define the **Commuter Time Increase Factor** as following:

$$CTIF = \frac{mean\ commute\ time_{(accident\ day)}}{mean\ commute\ time_{(ordinary\ day)}} \quad (4)$$

The **Line Load Drop Factor** is defined as following:

$$LLDF = 1 - \frac{mean\ load\ volume_{(accident\ day)}}{mean\ load\ volume_{(ordinary\ day)}} \quad (5)$$

3 Dashboard for Event Analysis

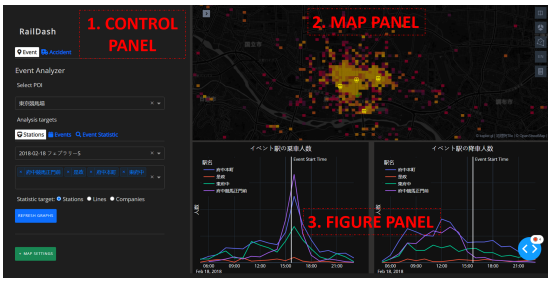

Figure 2: Front-end design layout of Dashboard for Event-analysis

Figure 2 shows the front-end design layout of the browser-side application. The dashboard primarily consists of three panels which are - Control Panel, Map Panel, and Figure Panel. Control panel is mainly for selecting and setting the parameters and to generate desired statistics results and visualizations. Map panel uses an extended Kepler.gl package [1] for map visualization. Figure panel consists of the figures created by Plotly Dash and is visible in all modules.

The dashboard as three primary modules - Station Analysis, Event Comparison, Event Participant's Statistics. One can visualize and analyse following features in the dashboard: station-wise entry/exit plots and comparison with ordinary days, comparison of two or more events, share of railway stations and lines used by event-participants, event-participants' features like transportation mode used to arrive at venue, origin prefectures, time spent at the venue, Con-

[1] https://natsuapo.github.io/keplerjis/

gestion Index values separately for Entry and Exits.

4 Dashboard for Accident Analysis

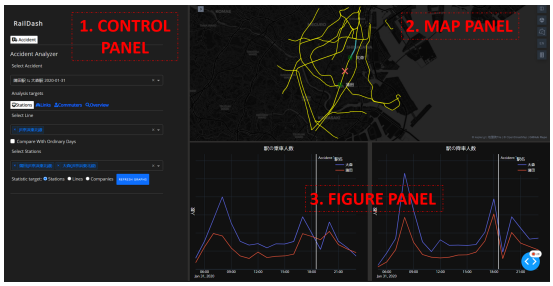

Figure 3: Front-end design layout of Dashboard for Accident-analysis

Figure 3 shows the front-end design layout of the browser-side dashboard application for accident analysis. This accident analysis dashboard also consists of three panels which are Control panel, Map panel, and Figure panel, respectively. The user can interact with the system using four separate modules where each module has unique functionality. These four modules are (i) Station Analysis, (ii) Link Analysis, (iii) Commuters Analysis, and (iv) Accident Overview, respectively.

One can visualize and analyse following features in the dashboard: station-wise entry/exit plots and comparison with ordinary days, link-wise passenger load volume plots and comparison with ordinary days, map visualization of stations and links selected, affected railway commuter's features like commute time increase factor, change in railway lines share, share of new railway lines used on day of accident, home and work/destination cities share, Line load drop factor, signifying how much passenger load volume dropped for the given line, Animation of change in passenger load volume on the day of accident.

5 Conclusion

In this study, we proposed a novel and generic dashboard system for analyzing and visualizing the effect on railway traffic during big events and unforeseen incidents through big GPS trajectory data. By using big GPS trajectory data generated by millions of users, we are able to implement a comprehensive and multi-faceted study for the events/accidents which not only include railway passenger analysis but also mobility study of event participants and affected railway commuters. This dashboard can have a great significance for railway administrators, event organizers, and city planners to measure and estimate the impact of big events and accidents on railway traffic.

References

[1] W. Huimin, H. Zhongwei, G. Jifu, Z. Liyun, and S. Jianping, "Operational analysis on beijing road network during the olympic games," *Journal of Transportation Systems Engineering and Information Technology*, vol. 8, no. 6, pp. 32–37, 2008.

[2] P. Lin, J. Weng, D. Alivanistos, S. Ma, and B. Yin, "Identifying and segmenting commuting behavior patterns based on smart card data and travel survey data," *Sustainability*, vol. 12, no. 12, p. 5010, 2020.