RAILDASH: A DASHBOARD SYSTEM TO ANALYZE EFFECTS OF EVENTS ON RAILWAY TRAFFIC USING BIG GPS DATA

P. Jeph[1,5] , H. Takayasu[2,1], T. Xia[1,3*], H. Kanasugi[1,3], R. Jiang[1], H. Mizuseki[4], R. Shibasaki[1,3]

[1]Center for Spatial Information Science, The University of Tokyo, 277-8568, Kashiwanoha 5-1-5, Kashiwa, Chiba, Japan
{jeph.puneet,xiatianqi,jiangrh,yok,shiba}@csis.u-tokyo.ac.jp
[2]East Japan Railway Company, 1518578, Yoyogi 2-2-2, Shibuya, Tokyo, Japan - h-takayasu@jreast.co.jp
[3]LocationMind Inc., 101-0032, Iwamotocho 3-5-2, Chiyoda, Tokyo, Japan
[4]Blogwatcher Inc., 104-0033, Shinkawa 1-3-17, Chuo, Tokyo, Japan
[5]IRSEE, North Western Railway, 302017, Jaipur, India

Commission IV, WG IV/9

KEY WORDS: Railway Transportation, Visualization System, Big smartphone GPS data, Crowd estimation, Event analysis

ABSTRACT:

The big events cause substantial deviations in the city traffic, especially railway transportation which happens to be the dominant mode of transportation for most of the major cities. Analysis of big events' impact on railway transportation, therefore, is of great importance to urban planning and transport management, yet is quite challenging because of the lack of readily available data about railway passengers' citywide flow and event participants' choice of transportation mode. Previous works have mainly relied on precise but limited data like sensors, AFC (Automated Fare Collection), or smart-card get-on get-off data to estimate the railway passengers, and did not take a holistic multi-dimensional approach to analyze railway traffic congestion caused by the events. To tackle these challenges, we propose a novel interactive Dashboard system that utilizes millions of smartphone GPS records across Japan. The dashboard can estimate and visualize railway passengers for the stations nearby the event venue as well as other relevant event participants' information. We also introduce a Congestion Index to measure the increase in congestion of stations during events. The dashboard can be highly useful for event organizers, railway administrators, and city planners to assess and compare the impact of big events on railway traffic.

1. INTRODUCTION

Rail transportation system plays an important role in urban development as it is energy-efficient, relatively time-saving, and a punctual mode of transportation for transporting passengers and freights. According to the survey of person trips conducted by the Japanese government (Authority et al., 2011), Tokyo has one of the highest rail transit usage ratio of up to 48%, showing that almost half of the Tokyo residents prefer to travel by train or metro services. Therefore, understanding railway traffic is of utmost importance for building an Intelligent Transportation System (ITS) in the development of a Smart City.

An important area of study in railway traffic analysis is the behavior during big events and other incidents, where railway traffic deviates from its regular pattern causing unforeseen and unusual congestion in some of the lines and stations. On the day of big events, a big crowd converges in the venue of the event just before the start of the event and then diverges out of the venue after the end of the event. This causes unusual congestion in railway stations and lines, especially those nearby to the venue. This event-caused congestion is different from the usual congestion observed during peak morning or evening rush hours in terms of the features and exhibits different spatio-temporal characteristics (Ni et al., 2016). Therefore, it becomes important to study and analyze the railway traffic behavior during such kinds of events to understand their effect on railway passengers in particular and railway traffic in general. At the same time, it is important for the event organizers and city planners to understand the features of the audience attending the event, for example, where did the audience come from, and which transportation mode did they use to arrive at the venue? Having this information will help them to better plan and manage future events.

Previous studies have mainly utilized sensors and AFC (Automated Fare Collection) based IC Card records for railway passenger analysis (Filograno et al., 2011, Kusakabe et al., 2010, Sun et al., 2012). Although railway passenger numbers can be more precisely demonstrated by such IC Card or Smart Card data which have exact tap-in and tap-out timestamps, it can neither differentiate between event participants and other passengers nor analyze people's flow outside the origin and destination stations. In addition, the limitations with AFC and Smart Card data are that they are difficult to obtain for the city-wide rail metro system over a longer time duration, and passenger's estimated route is mostly based on shortest path algorithm.

With the recent advent of GPS-enabled smartphones and the popularization of LBS (location-based services), huge amounts of GPS data has become available for human trajectory and traffic analysis. The human mobility data can be used as an effective data source to evaluate the effect and impact of the big events. The traditional data like surveys, smart card data, etc. do not take into account the event participants' mobility pattern. The big smartphone GPS data can overcome these limitations of the traditional data. It can also help assess the mobility flow of event participants and their transportation modes. Though the mobility data is not as accurate as AFC and Smart Card Data and is not completely representative, it still can give a multi-

*Corresponding author

faceted comprehensive analysis and has the potential to be a very useful tool. Over the past decade, a lot of research have focused on the analysis, prediction, and simulation of road traffic using GPS data (Herrera et al., 2010, Sense, 2008) but few have focused on railway traffic analysis and visualization using GPS data.

With the issues aforementioned, this study proposes an interactive dashboard system to effectively analyze and visualize the railway traffic as well as event participants features during the big events using big human GPS data. Figure 1 shows the system framework of this dashboard system. In this system, the inputs are the raw GPS data, rail-road network data and event information data. In the data preprocessing step, the GPS data are processed to extract the stay points and user trajectories. These user trajectories are further labeled with their transportation mode (WALK/BIKE/CAR/TRAIN/STAY) and then mapmatched. In the next step of event analysis, we first aggregate the identified railway passengers for selected stations nearby the venue (within a 1.5 km. radius buffer) and also compute the local OD (origin-destination) density of passengers for these stations, i.e. destination density for exiting passengers, and origin density for entering passengers. Then event participants are identified using appropriate spatial grids and time windows, along with their other characteristics like arrival/departure time, origin prefecture, transportation mode used to reach the event venue, etc. In the final step, we integrate all the processed results and visualize them using figures, maps, and other statistics in an interactive dashboard browser application developed by a combination of Plotly Dash and Kepler.gl(He, 2018). We also compute a score called Congestion Index to help in the assessment and comparison of the event's impact on railway traffic and passengers.

The main contributions of this paper are listed as follows:

- We develop the first-of-its-kind dashboard by integrating both Plotly Dash and Kepler.gl platforms for comprehensive railway traffic analysis during events using big GPS data generated by millions of smartphone users from entire Japan.
- The dashboard can help visualize the change in congestion of the railway stations nearby the venue on the day of the event, as well as give information about event participants mobility flow like transportation mode used, time of arrival and departure, time spent at the venue, etc., which is generally not possible using traditional methods and data.
- We also present case studies of several big events in Tokyo using our dashboard and demonstrate how those events affected the railway traffic.

The remaining part of this paper is organized as follows: Section 2 introduces the past studies related to railway traffic analysis and the datasets they used. Section 3 explains system design, the data preprocessing, and the methodology used for identifying railway passengers and event participants. Section 4 introduces the dashboard design and its different modules. We also present case studies of some real-world events within this section to further demonstrate the functionalities of various modules of the dashboard, followed by conclusion and future works in section 5.

2. RELATED WORKS

The data used in past researches on Railway Transportation can be mainly categorized into two classes based on their origin. The first type is the information obtained by means of various sensors about railway trains like position, velocity, acceleration, weight, etc. This information can be used to detect the exact status of trains in real-time and can also help in their punctuality analysis. Massimo et al. (Filograno et al., 2011) used sensor information for train identification, speed and acceleration detection, and dynamic load calculation.

The second type is the data primarily about railway passengers, like tap-in, tap-out time collected from AFC (Automated Fare Collection) or Smart Cards. Sun et al. Satoshi et al. (Kawamura et al., 2015) used OD-Matrix of Smart Card Ticket 'PASMO' origin-destination Data of Tokyo Metropolitan Area to examine and visualize railway passenger flow and demands during public events. Baichuan et al. (Mo et al., 2020) also used Smart Card Data collected from AFCs to distribute passengers over the rail network using the Bayesian simulation-based optimization method. However, limitations with AFC and Smart Card data are that they are difficult to obtain for the city-wide rail metro system over a longer time duration, and passenger's estimated route is mostly based on shortest path algorithm.

Some studies have tried to use other novel data sources to analyze railway traffic. Yuki et al. (Maekawa et al., 2014) used Bluetooth RSSI observed by passengers' mobile phones to estimate car-level train congestion using a Bayesian-based likelihood estimator. Kohei et al. (ITO and SAITO, n.d.) used the moving object detection method from web camera images to estimate rail traffic congestion, but passenger information was excluded.

With the advent of smartphone location services, GPS data has been widely used to analyze road traffic, but few studies have focused on railway traffic analysis using the GPS data. The earlier studies utilizing the GPS trajectory data were focused on extracting meaningful place of people (Zhou et al., 2007), understanding people moving pattern (Liao et al., 2006), and predicting movement of people (Ashbrook and Starner, 2003). The recent studies have focused on analysis and prediction of human transportation. GPS log data of taxis was utilized (Huimin et al., 2008) to analyze and detect the road traffic congestion. Xia et al. (Xia et al., 2018) used GPS trajectory information to predict the kernel density in railway network using deep-learning based methodology LSTM (Long Short Term Memory). D'Andrea et al. (D'Andrea and Marcelloni, 2017) designed a system that detects real-time traffic congestion and incidents on roads based on analyzing GPS trajectories of moving vehicles.

Some studies have tried to develop dashboards for traffic visualization. Pathak et al. (Pathak et al., 2015) utilized social network data like Twitter, to design a dashboard for obtaining a real-time view of the traffic data. Mingyu et al. (Pi et al., 2019) used Taxi GPS trajectory data to create a dashboard to find the cause of traffic congestion. Lwin et al. (Lwin et al., 2019) used big data including GPS trajectory data to develop City Geospatial Dashboard as solution provider for disaster management. Li et al. (Li and Zhu, 2016) developed a dashboard to simulate railway passenger flow during train delays using origin-destination data.

3. SYSTEM DEVELOPMENT

As introduced in the previous section, this study aims to develop a dashboard system for analyzing the influence of big events on railway traffic via GPS data. The whole system is constructed

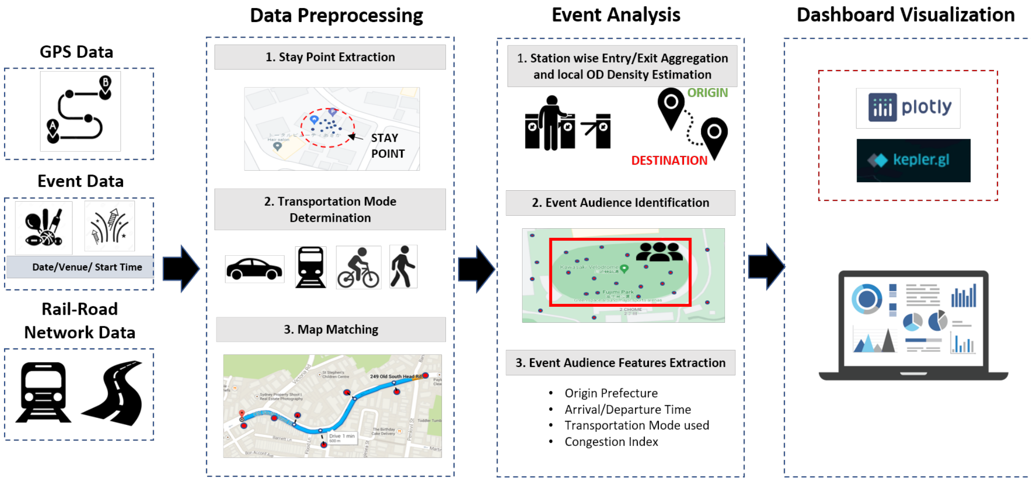


Figure 1. The System Framework of the Dashboard.

based on the architecture illustrated in Figure 2. In this system, data collected from multiple data sources are preprocessed and stored in the PostgreSQL database on the server-side for ease of access and saving the memory. Then the statistic results based on selected events and stations are further processed on the server-side and visualized on the user end in the browser application developed using Plotly Dash. The dashboard consists of several modules with different functionality, which can be visualized and controlled via different panels in the dashboard with either the response from server side, or the callback functions directly in the front-end. In the remainder of this section, we will introduce the panels and modules of the dashboard in detail. Users can interact with each module as per their choice and needs with customized settings.

3.1 Data collection

This system collects three types of data from different data sources, including big human mobility GPS data, event data, and railway network data.

For GPS data, we collaborated with Blogwatcher Inc. to get big human GPS trajectory data that cover 5 million people in the 47 prefectures of Japan. The location data are collected through smartphone applications that have a built-in location-sharing module provided by Blogwatcher Inc. with the user's consent. Any personal information, which could identify individuals, were not collected. Data attributes are anonymized ID, timestamp, longitude, and latitude.

Event data is collected by the following steps: first, we choose several popular event venues and collect their location information from OpenStreetMap. Then for each venue, we collect event information from official websites and several ticket selling platforms, which includes the start time and the number of audiences. For each event venue, we extract the nearby stations by using buffer of 1.5 km radius, for each rail line, we choose the closest station as our research target.

Railway network data is collected from the National Land Numerical Information (NLNI) of Japan and have been adapted by simplifying the network and by adding topology information of transfer stations. This data has a graph-like structure where nodes represent the stations, while edges represent the railway links that connect two stations.

3.2 GPS Data Preprocessing

GPS data generated by human mobility flow represent the spatio-temporal position of people. A GPS points trajectory can be denoted as $P = (p_1, p_2, ..., p_n)$, where *p=(id, timestamp, longitude, latitude)* and *n = total no. of points for that user*. To identify the railway passengers and their trajectories, it is important to first process the raw GPS data to segment them by their transportation modes, followed by map-matching. These steps are explained in detail in the remaining part of this section.

3.2.1 Transportation Mode Detection: In this study, we utilize the travel mode detection method proposed by Witayangkurn et al. (Witayangkurn et al., 2013). The process contains four steps: stay points extraction, change point detection, transportation mode classification, and segment merging.

Stay points are the group of consecutive GPS points where a user remains stationary within a small area and does not move out of it. In the first step, we extract the stay segments based on the distance and time thresholds, and also remove the noise points based on the mean and standard deviation of their Gaussian distribution.

In the second step, moving and stay segments are separated first. Then moving segments are further split into walk and non-walk type segments by detecting change points using a speed threshold. Then we detect the change points in non-walk type segments using VCR (Velocity Change Rate) to identify change of transportation mode. VCR is the rate of change between the average speed of the current segment and the speed of the current point. If VCR is higher than a threshold, then the current point is identified as a change point. We use the ratio of GPS points lying on railway lines in a segment to distinguish between railway and road transportation.

In the third step, we determine the transportation mode for each split segment using random forest classifier. The following features were used for each segment or trajectory: total distance, total time duration, percentage of points in road network and railway lines, and speed features like minimum speed, maximum speed, average speed, maximum acceleration, and velocity change rate. Each moving trajectory is labeled by four types of transportation modes - Walk, Bike, Car, Train.

In the final step, we merge the segments based on certain principles to reduce uncertain trajectories. Small and uncertain seg-

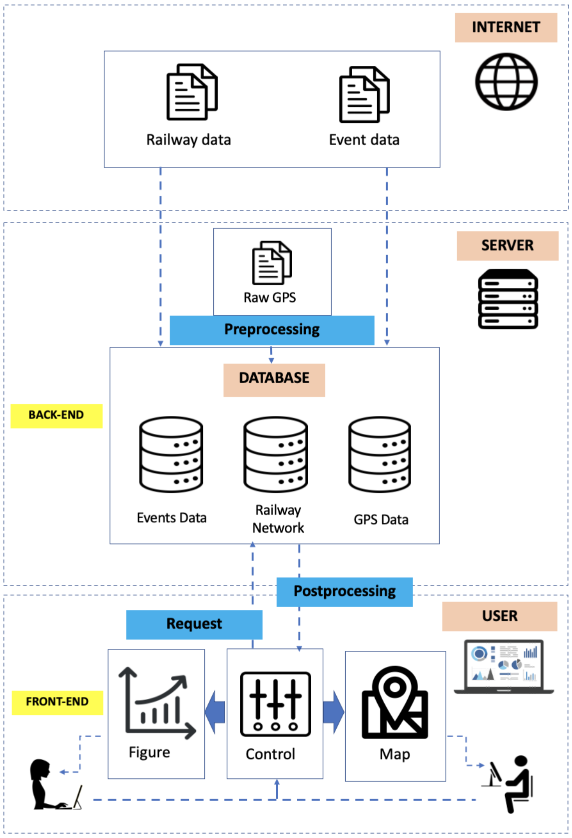


Figure 2. System Architecture

ments are merged with neighboring segments. Consecutive segments of the same type of transportation mode are also merged. For other complex cases, classifiers with training data were used to merge the segments. Thus we obtain segmented trajectories labeled with their transportation mode by implementing the aforementioned transportation mode detection methodology. The next task is map-matching in order to get precise routes of the users identified as railway passengers.

3.2.2 Map Matching:

Map matching is the process that serves the purpose of recovering the original route on a transportation network from a sequence of noisy GPS points. In this study, we referred to the methodology used by Shun et al. (Ikezawa et al., 2016) to map match the Train mode GPS Trajectories. We first use sparse map matching technique on extracted railway mode trajectories, whereby each GPS point is projected to several nearby nodes with the probability reverse to the distance. Then the path with the nodes in maximum likelihood is chosen as the candidate route. For each neighboring node pair in the route, we use Dijkstra's shortest path algorithm to find out the railway route. Then timestamp of each en-route station is computed by assuming that the train runs with a uniform velocity between two map-matched points. In addition, the time of get-on/get-off is computed from the first and last GPS points using the average velocity of the entire trajectory. Accordingly, we don't take into account the stoppage time for ease of computation.

3.3 Event Analysis and Railway Traffic Estimation

Using the map-matched railway mode trajectories, we can estimate the time and location of get-on get-off for each identified railway passenger, with their entire route including transfer stations along with timestamps. For this study, we mainly focus on estimating the railway traffic near the venue of the events to understand the congestion and deviations.

3.3.1 Railway Passenger Aggregation:

To estimate the congestion at stations, we prepare two kinds of railway passenger aggregation in a given time period. We count number of railway passengers getting-on and getting-off separately for each railway station during 1-hour time period. Thus, we obtain two aggregations for each station - passengers entering the station to start the journey, and passengers exiting the station after completing the journey. It is to be noted that we don't include get-on/get-off numbers at transfer stations for aggregation since the passenger is still en-route and the journey is not yet finished. In this study, we take 1-hour time duration as the temporal unit of the aggregation, considering that it is not so short as to have a very low passengers count for small stations, and not so big as to ignore any significant short temporal changes.

3.3.2 Local OD Density Estimation:

For big events, crowd management plays a very important role in the smooth organization of the event. This is not only true inside the event venue, but also for the surrounding area of the venue. Therefore, it becomes important for event organizers and city administrators to have an estimation of nearby areas the audience may visit before and after the event for an effective crowd management. We try to solve this problem using the processed GPS trajectory data. 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. In simpler terms, it is simply the grid-based density of where the passengers are going to just after exiting the station, and where the passengers are coming from before entering the station.

3.3.3 Event Participants Features Extraction:

Using GPS trajectories and transportation information, we try to identify the event participants 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.

To identify the event participants, we assume that the user will have at least one GPS trajectory labeled as *STAY Mode* in the event venue area. So we extract all the *STAY* type GPS points within the event venue area during the one-hour time period from the start of the event and identify them as event participants. After extracting the event participants, we find out their arrival time, and departure time by tracing their GPS trajectories. Similarly, we find out the transportation mode they used to arrive at the event venue by tracing their last *MOVE* type trajectory. We also find their origin prefecture by extracting the location of the first *STAY* point on the day of the event.

3.3.4 Event Congestion Index:

To analyze the railway traffic congestion caused by the big events, we develop a score index to help in the assessment and comparison of the impact on railway traffic. We name this index as Congestion Index. This index can help in assessing the relative scale of congestion in railway traffic compared to past average congestion. Usually the events cause big increase in exit numbers and entry numbers of the nearby stations before the start of the event, and just after the end of the event, respectively. Therefore we define the congestion index separately for entry and exit numbers of the

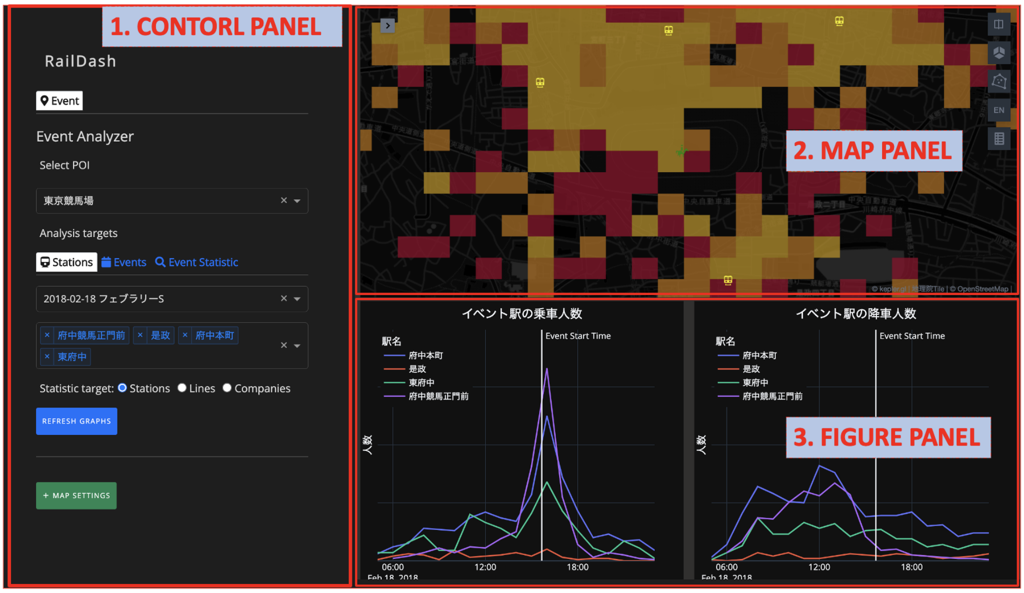


Figure 3. The Layout-design of the Dashboard.

stations. They are defined as the following:

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

$$C_{Exit} = W_i \sum_{i=1}^{N} \left(\frac{O^{o}_{peak}}{\frac{1}{5}\sum_{m=1}^{5} O^{-m}_{peak}} \right) \tag{2}$$

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. We first compute ratio of peak value on the day of the event to the average peak value observed in the last m days, for hourly get-on and get-off numbers separately, and then take the weighted average of these ratios station-wise. Equation 1 and Eq. 2 compute a weighted average of station-wise computed values which are defined as the Entry/Exit Congestion Index. The weights are determined by the share of the passengers for each station. We take value of m as five days for the purpose of computational efficiency. Moreover, these past five days are chosen in a manner to have same weekday type as the event-day to avoid any inherent weekday bias in railway traffic. 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.

4. DASHBOARD DESIGN AND CASE STUDIES

4.1 Layout introduction

Figure 3 shows the front-end design layout of the browser-side application. The layout is inspired by the official map example provided by Plotly Dash. The dashboard primarily consists of three panels which are - control panel, map panel, and figure panel, respectively.

Control panel is mainly for selecting and setting the parameters and to generate desired statistics results and visualizations inside the figure and map panel. Users can select the events and stations that they want to analyze and visualize from the control panel. In addition, a collapsible component called Map Setting is provided at the bottom of the control panel for more customizable interaction with the map visualization panel.

Map panel uses an extended Kepler.gl package [1] for map visualization. Compared to other map visualization tools, Kepler.gl is much better suited for big data visualization and map animation, and thus is more powerful and faster in visualizing passenger density and human mobility using big GPS trajectory data. However, Kepler.gl cannot interact with Plotly Dash in its default form. Moreover, the controller in Kepler.gl can be difficult for people who are not familiar with GIS to manipulate. To solve these limitations, this study improves upon the communication part of Kepler.gl and makes it intractable with Plotly Dash in the front-end part by rewriting some parts of the source code.

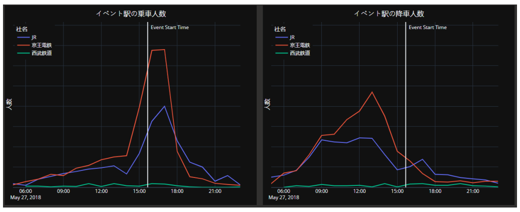
Figure panel consists of the figures created by Plotly Dash and is visible in all modules. The figures are refreshed based on the conditions selected by users from the control panel. And just like all other Plotly Dash applications, the figures are interactive to set the filters and zoom levels directly by the user. It is to be noted that due to data security issues, we do not visualize the ticks of y-axis data during the demonstration of the dashboard.

4.2 Modules and User-Interaction Introduction

As mentioned in the system introduction, this application provides three separate modules for users to interact. The user can first select the venue of the events from a drop-down menu. Once a venue is selected, the dashboard loads the data of the events and stations belonging to that venue. Then the user can interact with the system using three separate modules where each module has separate functionality. These three modules are Railway Passenger Analysis, Events Comparison, and Event-Participants Statistics, respectively. The functionalities of these modules are now explained in detail along with some real-world case-studies for better understanding.

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

(a) Station-wise Get-on and Get-off numbers figure.



(b) Railway Company wise Get-on Get-off numbers figure.

Figure 4. Visualization of Get-on and Get-off numbers for the event of 27th May, 2018 at Tokyo Race Course venue.

4.2.1 Railway Passenger Analysis Module: We utilize computed get-on and get-off numbers of stations for analyzing and visualizing railway passenger volume during the big events. In this module, users can select one event from the selected venue and can choose multiple stations which are to be analyzed and visualized. The aggregated get-on and get-off passenger numbers from the selected stations on the event date are visualized in the figure panel, while the passenger's local origin and destination location density are encoded to the standard grids of 125-meter Japanese industrial standardized (JIS) mesh code [2] and visualized via the extended JIS mesh code layer in the map panel. Users can also choose to visualize either origin or destination density of the selected stations from the control panel. In addition, we also provide the user the option of choosing aggregated get-on and get-off numbers using railway lines or railway companies in place of stations.

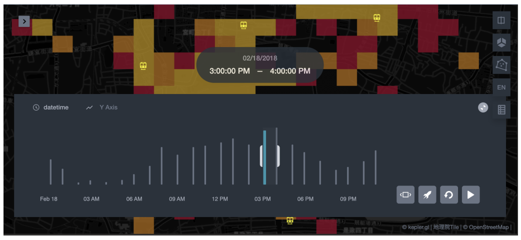
Case-Study: To further demonstrate the functionality of this module, we choose a very popular horse race event as the case study. The event took place on 27th May, 2018 at Tokyo Race Course venue and an estimated 126,767 people attended the event as per the official website. This event is known as Japanese Derby and is an international Grade-1 flat horse race in Japan. There are four stations which fall inside the 1.5 km buffer area around the venue, which are - Fuchukeibaseimon-mae, Higashi-fuchu, Fuchuhommachi, Koremasa. The start time of the event was 15:40 hrs.

Figure 4a shows the plots for get-on and get-off numbers for all these four stations separately. The peak observed in the get-on numbers around and after the event is quite evident from the figure, which shows that the departure pattern of most of the audience is concentrated just after the end of the event. Similarly, in the get-off numbers plot, the distributed peaks before the start of the event indicate the distributed arrival patterns of the audience and are not concentrated around a particular time. The user can also see which station gets the most passengers and can compare the congestion of four stations temporally. Figure 4b, on the other hand, shows the get-on get-off numbers using railway company-wise aggregation instead of station-wise aggregation.

Figure 5 shows the local OD density estimation of railway passengers in the map panel. This map utilizes three layers to respectively visualize railway stations and horse racing fields in icon layers, and OD density in a Japanese Standard Mesh layer. By looking at the color-graded density map, the user can estimate the crowd condition at various places nearby the venue. Moreover, the user can also change density map visualization via the control panel by adding temporal filters, by changing the density type (origin density by default), or visualize both origin and destination density via a dual-map view.

[2]https://nlftp.mlit.go.jp/ksj/old/old_data_mesh.html

(a) OD density maps in dual-maps view: the left and the right maps are respectively the origin and destination density of all railway passengers from selected stations

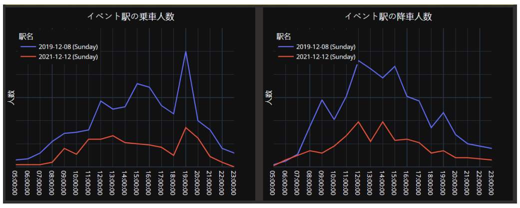


(b) Origin density visualization with one hour filter and animation bar

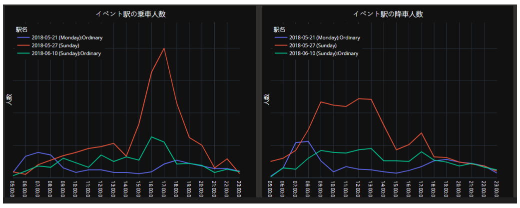
Figure 5. Visualization of density maps in the map panel

4.2.2 Event Comparison Module: Event comparison module gives the user option of comparing railway passengers volume for different events on different dates which were organized at the same venue. The user can choose multiple events and the station to be analysed from the drop-down menu in this module. The figure panel will show the comparison of get-on and get-off numbers for different events for the selected station. We also give the user the option of comparing the get-on get-off numbers on the day of the event with a regular non-event day. The non-event days are chosen in a manner to have both weekday and weekend dates. Although this module is quite similar to the previous Railway Passengers Analysis module in terms of visualization, it is more useful in analyzing and comparing the temporal changes in passenger volume for different events on different dates for a particular station of interest.

Case-Study: We choose Tokyo Dome live events as the case study to demonstrate the functionality of this module. Here we try to demonstrate two different scenarios - comparison of the same type of event organized in pre-COVID and post-COVID times, and comparison of an event with an ordinary non-event day. For the first scenario, we choose two live concert events performed by the same band at the same time of the day, organized on 8th December, 2019, and on 12th December, 2021, respectively. The selected station is Korakuen station. Figure 6a shows the comparison of get-on and get-off numbers for Korakuen station for the two events. It is evident that the station was more crowded during the pre-covid time event. On the other hand, fig. 6b shows the comparison of the Tokyo Horse



(a) Comparison of two different events.



(b) Comparison of an event with ordinary day.

Figure 6. Visualizations in event-comparison module.

Race event of 18th May, 2018 with ordinary days of 2018-05-21 and 2018-06-10 for the Fuchukeibaseimon-mae station. We can clearly see that the event caused quite large peaks compared to ordinary day numbers.

4.2.3 Event-Participants Statistics Module: This module focuses on the analysis and visualization of event participants' information as a whole instead of just railway passengers. In this module, the extracted event participants' information is further aggregated and processed to provide a more comprehensive statistical result about the event. For each event, various statistical results about event participants are displayed in the figure panel. These general event statistics help users to understand event participants' information as a whole, for example, the event's start time, visiting numbers from different regions, the transportation mode ratio, railway usage ratio, etc. The map panel is not required in this module.

The details of the available statistics are as follows:

Congestion Index: It is the relative scale of the congestion caused by the event in get-on and get-off numbers compared to past 5 days of same weekday type. A value of higher than 1 will show the relative increase in the congestion. For example, a value of 2 means 100% increase in the congestion. This index assigns a single value to the event for its effect on railway traffic, therefore, being valuable in assessing and comparing different events.

Railway Usage Statistic: It shows the ratio of the railway stations used by those event participants who used railway service to reach the event venue, in the form of a pie-chart. This figure provides three further options to the user to display the usage ratio result - station wise, line wise, company wise.

Transportation Mode Ratio: Here we display the ratio of transportation mode used to reach the venue by the event participants in form of a pie chart. The transportation modes available are- *WALK, BIKE, CAR, TRAIN.*

Origin Prefecture Ratio: Origin prefecture is the prefecture where the event participant comes from, to attend the event. This is again displayed in the form of a pie chart. For better visualization performance, we only show the top five prefectures, while sum up the remaining ratio as the ratio from other prefectures (others).

Duration of Visit: This is a histogram figure that shows distribution of the time spent at the venue by the event participants.

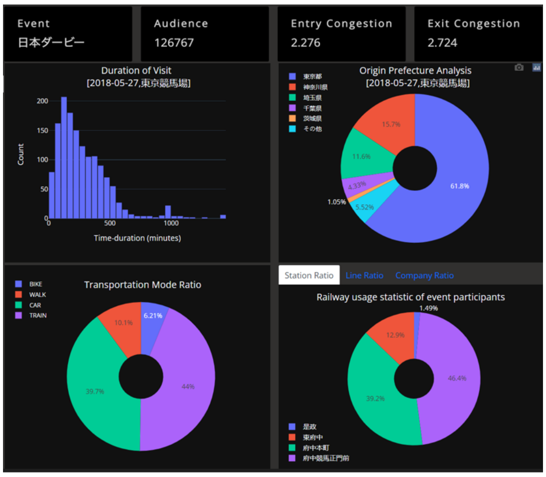


Figure 7. Event-statistics results.

Case-Study: For the case-study of this module, we once again choose the big event of the Tokyo Horse Race on 27th May, 2018. Figure 7 shows the query result of event participant statistic module. The query result includes two part, the upper part includes four cards which are respectively the event name, the audience number which collected from the internet, the entry congestion index and the exit congestion index. The lower part includes various of statistics figures. As is shown in the card part, the entry and exit congestion index were found to be 2.276 and 2.724, respectively, showing an increase of almost 150% in the congestion. From the figure part, We can also infer that almost 44% of identified event participants used railway transportation to reach the venue. Among the participants, who used railway transportation, almost 47% got off at Fuchukeibaseimon-mae station to reach the venue. Almost 62% of the participants were from Tokyo prefecture, while Kanagawa and Saitama prefectures were distant second and third. Another interesting statistic is that most of the participants spent up to 5-6 hours at the venue, which indicates that the participants did not only enjoy Japan Derby, but also participated in the previous and post horse racing games held on the same day.

5. CONCLUSION

In this paper, we proposed a novel and generic dashboard system for analyzing and visualizing the effect on railway traffic during big events through big GPS trajectory data. This is the first study to use big smartphone GPS data to analyze the railway traffic congestion during events to the best of our knowledge. 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 which not only include railway passenger analysis but also mobility study of event participants. This dashboard can have a great significance for railway administrators, event organizers, and city planners to measure and estimate the impact of big events on railway traffic and also compare different big events using the Congestion Index score. It can also help them to make advance preparation, in terms of resource planning or crowd management, to manage such events in the future. Furthermore, we also demonstrate the successful integration of Plotly Dash with the Kepler.gl platform in our dashboard, enabling the user to directly interact with Kepler maps from within the dashboard console. We also showed the utility of the dashboard by providing some case studies of real-world events.

The limitation of using GPS dataset is the potential inaccuracies in transportation mode detection and map matching. The sparsity of the GPS data also adds to this problem. Another point of discussion can be the identification criteria of the target stations, which can be subjective and may differ for different venues. In addition, the interaction between the dashboard and the Kepler map is one-way at present, i.e. the user cannot use the map panel to control graphs and figures in other panels.

In the future, we will further enhance the map visualization part for better map interaction and create a bigger database of different types of big events. We plan to provide railway-link wise passenger aggregation also in addition to present station-wise get-on get-off aggregation for better analysis. We will also include another module to study unforeseen incidents like accidents, earthquakes, etc. in the dashboard.

REFERENCES

Ashbrook, D., Starner, T., 2003. Using GPS to learn significant locations and predict movement across multiple users. *Personal and Ubiquitous computing*, 7(5), 275–286.

Authority, L. T. et al., 2011. Passenger transport mode shares in world cities. *Journeys. Singapore: Land Transport Authority*.

D'Andrea, E., Marcelloni, F., 2017. Detection of traffic congestion and incidents from GPS trace analysis. *Expert Systems with Applications*, 73, 43–56.

Filograno, M. L., Guillén, P. C., Rodríguez-Barrios, A., Martín-López, S., Rodríguez-Plaza, M., Andrés-Alguacil, Á., González-Herráez, M., 2011. Real-time monitoring of railway traffic using fiber Bragg grating sensors. *IEEE Sensors Journal*, 12(1), 85–92.

He, S., 2018. From Beautiful Maps to Actionable Insights: Introducing Kepler. gl, Uber's Open Source Geospatial Toolbox. *Uber Engineering*, 29.

Herrera, J. C., Work, D. B., Herring, R., Ban, X. J., Jacobson, Q., Bayen, A. M., 2010. Evaluation of traffic data obtained via GPS-enabled mobile phones: The Mobile Century field experiment. *Transportation Research Part C: Emerging Technologies*, 18(4), 568–583.

Huimin, W., Zhongwei, H., Jifu, G., Liyun, Z., Jianping, S., 2008. Operational analysis on Beijing road network during the Olympic Games. *Journal of Transportation Systems Engineering and Information Technology*, 8(6), 32–37.

Ikezawa, S., Kanasugi, H., Matsubara, G., Akiyama, Y., Adachi, R., Shibasaki, R., 2016. Estimation of the number of railway passengers based on individual movement trajectories. *6th INTERNATIONAL CONFERENCE ON CARTOGRAPHY AND GIS*, 249.

ITO, K., SAITO, T., n.d. INFORMATION EXTRACTION ABOUT TRAFFIC SITUATION OF RAILWAY WITH WEB CAMERA IMAGES.

Kawamura, S., Tomita, Y., Itoh, M., Yokoyama, D., Toyoda, M., Kitsuregawa, M., 2015. An effective use of Tokyo metro passengers flow by visualization of smart card ticket 'PASMO'origin-destination data for public transport network to be sustainable. *Proc. WECC*.

Kusakabe, T., Iryo, T., Asakura, Y., 2010. Estimation method for railway passengers' train choice behavior with smart card transaction data. *Transportation*, 37(5), 731–749.

Li, W., Zhu, W., 2016. A dynamic simulation model of passenger flow distribution on schedule-based rail transit networks with train delays. *Journal of Traffic and Transportation Engineering (English Edition)*, 3(4), 364–373.

Liao, L., Patterson, D. J., Fox, D., Kautz, H., 2006. Building personal maps from GPS data. *Annals of the New York Academy of Sciences*, 1093(1), 249–265.

Lwin, K. K., Sekimoto, Y., Takeuchi, W., Zettsu, K., 2019. City geospatial dashboard: Iot and big data analytics for geospatial solutions provider in disaster management. *2019 International Conference on Information and Communication Technologies for Disaster Management (ICT-DM)*, IEEE, 1–4.

Maekawa, Y., Uchiyama, A., Yamaguchi, H., Higashino, T., 2014. Car-level congestion and position estimation for railway trips using mobile phones. *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 939–950.

Mo, B., Ma, Z., Koutsopoulos, H. N., Zhao, J., 2020. Capacity-constrained network performance model for urban rail systems. *Transportation Research Record*, 2674(5), 59–69.

Ni, M., He, Q., Gao, J., 2016. Forecasting the subway passenger flow under event occurrences with social media. *IEEE Transactions on Intelligent Transportation Systems*, 18(6), 1623–1632.

Pathak, A., Patra, B. K., Chakraborty, A., Agarwal, A., 2015. A city traffic dashboard using social network data. *Proceedings of the 2nd IKDD Conference on Data Sciences*, 1–4.

Pi, M., Yeon, H., Son, H., Jang, Y., 2019. Visual cause analytics for traffic congestion. *IEEE transactions on visualization and computer graphics*, 27(3), 2186–2201.

Sense, T., 2008. Rich monitoring of road and traffic conditions using mobile smartphones. Technical report, Microsoft Research, Tech. Rep. MSR-TR-2008-59.

Sun, L., Lee, D.-H., Erath, A., Huang, X., 2012. Using smart card data to extract passenger's spatio-temporal density and train's trajectory of mrt system. *Proceedings of the ACM SIGKDD international workshop on urban computing*, 142–148.

Witayangkurn, A., Horanont, T., Ono, N., Sekimoto, Y., Shibasaki, R., 2013. Trip reconstruction and transportation mode extraction on low data rate gps data from mobile phone.

Xia, T., Song, X., Fan, Z., Kanasugi, H., Chen, Q., Jiang, R., Shibasaki, R., 2018. Deeprailway: a deep learning system for forecasting railway traffic. *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, IEEE, 51–56.

Zhou, C., Frankowski, D., Ludford, P., Shekhar, S., Terveen, L., 2007. Discovering personally meaningful places: An interactive clustering approach. *ACM Transactions on Information Systems (TOIS)*, 25(3), 12–es.