ESTIMATION OF THE NUMBER OF RAILWAY PASSENGERS BASED ON INDIVIDUAL MOVEMENT TRAJECTORIES

Shun Ikezawa¹, Hiroshi Kanasugi², Go Matsubara³, Yuki Akiyama⁴, Ryutaro Adachi⁵, Ryosuke Shibasaki⁶

¹Master, Dept. of Socio-Cultural Environmental Studies, The University of Tokyo, Email:s.ikezawa_0907@csis.u-tokyo.ac.jp
²Project Researcher, Earth Observation Data Integration and Fusion Research Initiative, The University of Tokyo
³Project Researcher, Center for Spatial Information Science, The University of Tokyo
⁴Assistant Prof., Center for Spatial Information Science, The University of Tokyo
⁵ZENRIN DataCom CO., LTD.
⁶Prof., Center for Spatial Information Science, The University of Tokyo

Abstract
In Japan, it has been possible to get information about the number of railway passengers at a station by means of widely popularized traffic IC cards, but it is difficult to know a mid-flow route during railway ride. Although there are some questionnaire surveys such as person trip survey to grasp the detail of route selection, it is difficult to conduct continuous survey due to implementation cost and time-consuming tasks. On the other hand, using positional information automatically acquired from GPS devices, it has been possible to get people flow continuously. In this paper, in order to implement continuous estimation of railway traffic volume including mid-flow routes, we employed movement trajectories obtained from GPS functionality embedded in a smart phone, and firstly extracted railway passengers and their interpolated railway trajectories. Subsequently, we estimated the number of railway passengers, and examined the possibility about the traffic volume estimation of mid-flow routes with taking account of multiple railway lines of different railway companies.

Keywords: mobile phone GPS data, railway traffic volume, smart phone

1. INTRODUCTION

Understanding the number of railway passengers is a social needs to understand traffic condition in city planning, to estimate railroad demand in a short-term event, and to estimate damage at a disaster or an accident. In Japan where most railway operations are managed by private companies, it has been able to get information about the number of railway passengers at a station by means of widely popularized traffic IC cards such as Suica and PASMO. However, it is difficult to understand a passenger’s mid-flow route including a transfer between different companies. Each railway company can understand the number of their own passengers, but it is limited within own railway.

There are some questionnaire surveys such as person trip survey to grasp the detail of route selection, but it is difficult to conduct continuous survey due to implementation cost and time-consuming tasks. Metropolitan Transportation Census by Ministry of Land, Infrastructure and Transport (MLIT) contains the estimation of the number of railway/subway passengers based on “railroad OD survey”. However, the estimation only includes stations where an automatic ticket gates are equipped and the information getting on and off is available.

On the other hand, recently, it is possible to get continuous people flow from the large number of people through GPS equipped mobile phones and smart phones(Sekimoto et.al (2011)). In addition, since GPS functionality equipped in mobile phone can acquire data automatically, users’ workload under a continuous investigation could be drastically mitigated. However, since GPS functionality is unavailable in indoor environment such as subway and tunnels, estimation of railway traffic requires interpolation of mid-flow routes as well as raw GPS positions

In this paper, in order to implement continuous estimation of railway traffic volume including mid-flow routes, we employed movement trajectories obtained from GPS functionality embedded in a smart phone. Firstly, we extracted railway
passengers and their GPS points within railways, and interpolated railway trajectories along with the railway network in cluding subway. Subsequently, we estimated the number of railway passengers at each station in every 1 hour, and examined the possibility about the traffic volume estimation of mid-flow routes with taking account of multiple railway lines of different railway companies.

2. RELATED RESEARCHES

There is "the railroad inside of car congestion estimate that used the search demand to the wireless LAN access point" that Nakano, Numano and others as a precedent study of the estimation of number of railway passenger using a mobile terminal. They focused a mobile terminal with wireless LAN sent a probe demand to search neighboring access point, and they suggested the method of estimate the congestion of the railway inside of car using this probe demand. Their suggestion method is to estimate the portion of the section where each passenger took by classifying a terminal in the railroad inside of car and the terminals outside the railroad car, using a hash value of the MAC address of the dispatch terminal as ID. By the evaluation experiment using this method, there was the result to estimate a part of the section which a passenger took, but was not able to reach the estimate station getting on and off of the passenger. In addition, because there was the MAC address that they used as ID as the ID that was only in the railroad inside of train,

It was not able to understand passenger’s action before getting on train and after getting off train like whether passengers performed the transfer of the train or get out from station. In this research, it is possible to understand user’s action not depend on a railway and a railway company using GPS log data of the cell-phone. This is the big advantage in this study.

3. DATA SPECIFICATION

This chapter describes three data sets in this study.

3.1 Congestion Statistics ®

The data called “congestion statistics ®” provided from ZENRIN DataCom CO., LTD. (ZDC) through the joint research is the statistics data that they added up positional information data, then quantify and visualized flow pattern of people. Positional information data is provided by NTT DOCOMO, INC. (docomo). This data got from “docomo map navi” provided by docomo. Users of “docomo map navi” consented to that. The data used in “congestion statistics ®” (∗) is managed it generally and statistically so that an individual can't identified by docomo requested from ZDC. So, this data don’t include attribution such as sex, age. They extracted the data that related to time and position of start and end move of people from this data by method of Ono (2013). And they added a means of transportation to that data. Comprehensive processing of the congestion statistics® in this paper were performed by not authors but ZDC. In this research, we used all over the Kanto region between July 22, 2013 and July 28, 2013. Figure 1. shows the number of railway users and trip one every day.

![Image](image.png)

Figure 1. Number of daily railway passengers and number of estimated daily trips (∗∗)

ⓒ「congestion statistics ®」 Copyright© ZENRIN DataCom CO., LTD
3.2 Railway Network Data

Railway network is necessary to estimate mid-flow routes between getting on and off stations. In this paper, with the conversion method proposed by Kanasugi et al. (2013), we employed the conversion result from the railway geometry (2011) provided by MLIT. The dataset covers all over Japan, and represents each station as a node and each section between neighboring stations as a link.

3.3 National Survey Data for Evaluation

For the evaluation of our estimation, we used the national survey data about the number of railway passengers in 2013 provided by MLIT. The data include the number of average passengers a day at all stations except for some unknowns, and exclude the separation of weekday or holiday. And, we also used the national survey data about the time distribution of railway users in 2010 by the Metropolitan Transportation Census.

4. ANALYSIS METHOD

4.1 Railway Route Estimation and Interpolation (map matching process)

Since mobile phone GPS data we employed in this paper have insufficient time resolution, and have no observation points in subway areas, GPS points do not always represent all railway lines where passengers used. Therefore, in order to estimate and to interpolate mid-flow routes and stations, we employed sparse map matching proposed by Hara et al. (2013). Sparse map matching can estimate a practical route based on observation points by implementing the combination of the map matching of GPS points to network and the route search between the matching points on the network. Although the original methodology of sparse map matching in their paper employed the nearest network node as the map matching target, we employed the nearest link to mitigate matching error from original observation points (Figure 2). In this paper, sparse map matching is applied to railway trips extracted from mobile GPS data with Ono’s methods (2013).

![Figure 2. Estimation of railway route with sparse map matching](image)

Since sparse map matching can estimate mid-flow routes, time at each passing station (node) need to be associated with the estimated routes as the next step. Therefore, observation time of GPS point is regarded as the time at the nearest point on the link. And the time at each station can be calculated by assuming that a train runs with uniform velocity between matching points. In addition, the time at a getting on/off station is calculated from the first and last GPS points with average velocity of entire route (Figure 3). Accordingly, we didn’t consider time duration of stoppage time at a station.
4.2 Aggregation of Traffic Volume on Railway

As the estimated traffic volume on railway, we prepared three kinds of aggregation in this paper (Figure 4). The 1st aggregation is time distribution of the number of the hourly railway passengers in each day for understanding situation of railway passengers. 2nd is aggregation of the data extracted information of each passenger’s trip at 3 stations (getting on, getting off, transfer) from after estimation of railway of user’s, and same aggregation at 2 mainly stations in metropolitan area and 2 mainly stations in the commuter belt for understanding time distribution of the number of station user. (Table 5. and Figure 6.) 3rd is aggregation of each railway passengers on each station link using the data of after estimation of user’s railway for understanding time distribution of the number of railway passengers.
5. RESULT AND CONSIDERATION

5.1 Time distribution of the number of railway passengers

The left chart of Figure 7 shows the time distribution of the number of railway passengers.

From the left chart of Figure 7, the peak of the number of railway passengers is located at 7:00, 8:00 and 18:00. These peaks reflect that many passengers use railway for commuting to work and school in the morning, for returning home in the evening. Morning peaks are distributed in 7:00 and 8:00 because start time of most companies and schools is usually 8:00 to 9:00. And evening peak is less than morning because returning time is different among workers and students.

In this paper, it was possible to compare the number of railway passengers between weekday and holiday, that was not included in the estimation of the number of railway passengers in Metropolitan Transportation Census and the data of number of station users of National Land Numerical Information download service, because it is possible to calculate time distribution of railway passengers on holiday. The number of railway passengers on holiday less than the number on weekday because many companies and schools becomes break. That’s why, there are no remarkable peaks on holiday.

The right chart of Figure 7 shows time distribution of the number of the railway passengers of stay in a station on weekday in metropolitan area, extracted the Metropolitan Transportation Census. Because the estimation of the Metropolitan Transportation Census focus railway passengers on weekday, we compared between the time distribution on weekday of the left chart of Figure 7 and the right chart of Figure 7. There are the peak of the daytime passengers of the left chart at 7:00 and 8:00, and there is the peak at 8:00 in the day of the right chart. In addition, there is the peak of the evening in the left chart at 18:00, and there is the peak of the evening in the right chart at 18:30. And, there are fewer passengers both than a peak at daytime around 12:00. That’s why, we understood the time distribution of number of railway passengers seems corresponding.

Figure 7. Left: Time distribution of railway passengers estimated (all, ※) right:Time distribution of the number of the railway passengers of stay in a station (Metropolitan Area, from the Metropolitan Transportation Census)
5.2 Time distribution of station users

Figure 8 shows estimation result of the time distribution of the number of the station users getting on and off and transfer. From Figure 8, there are the peaks of the morning is 8:00, 7:00 and 18:00 like Figure 7.

Figure 9 shows the time distribution of station users at 2 mainly stations located metropolitan area and 2 mainly stations located suburbs. From Figure 9, because there is a peak of station user in the morning at Shinjuku Station located in the metropolitan area most in 4 stations at 8:00, and there is a peak in the morning at Omiya Station located in the suburbs and Kashiwa Station at 7:00, the possibility that many users begin to use a railway from the stations in suburbs at about 7:00 and arrive at the stations in metropolitan area at about 8:00. In addition, the peak of number of station users in the morning at Yokohama station divided 7:00 and 8:00, by there are 2 types users, these are users of arrival for commutation attending school to Yokohama station and users of departure for commutation attending school to Yokohama station, because Yokohama Station is a big station of the scales like Shinjuku Station located near from Shinjuku station in Kanagawa Prefecture.

And, big downtown is located in the neighborhood of Shinjuku Station, and as for the number of station user, there are not weekdays and the big difference even if it is the night of Saturday.

![Figure 8. The Number of station users (all, ※)](https://example.com/image.png)

ⓒ 「congestion statistics ®」 Copyright© ZENRIN DataCom CO., LTD
5.3 Time distribution of railway passengers on each railway

Figure 10 shows mapping result of a part of the time distribution of railway passengers on each railway of weekday and holiday, the time is 8:00, 12:00 and 12:00. From Figure 10, the number of users on weekday more than the number on holiday at 8:00.

The number of users on every railway was not many, the number of users on railway for metropolitan area from the suburbs is many tends that there is particulary many it. Reason of result becomes such distribution, was dividing number of users, By the number of subways in the metropolitan area. Around 12:00, number of users on holiday was small by many companies and schools becomes break, number of users on weekday’s around 12:00 by many users stay at companies or schools. That’s why, the number of railway passengers around 12:00 less than it in the morning. Around 18:00, the number of users on holiday less than weekday. However, the number of users on weekday around 18:00 less generally than the number of users on weekday around 8:00, because Many users’s movement time in the morning is overlapping, but the time of returning home is various.
Figure 10. the number of railway passengers on each railway(※)

ⓒ『congestion statistics ©』Copyright© ZENRIN DataCom CO., LTD

Figure 11 shows a comparison result between the number of the passengers getting on and off according to the station in the estimated result and the number of the passengers getting on and off according to the station in this research served from numerical land information downloading service. From Figure 11, most of stations located in the lower left, and it is revealed that four stations greatly deviate from other groups. That these stations are Shinjuku Station, Ikebukuro Station, Shibuya Station and Yokohama Station, has much of the number of the getting on and off and are unbalanced in a regression line by these stations.

Figure 12 shows a comparison result between the number of the passengers getting on and off according to the station in the estimated result and the number of the passengers getting on and off according to the station in this research served from numerical land information downloading service without 4 extreme stations.

The correlation’s value of Figure 11 is 0.828, the correlation’s value of Figure 12 is 0.759. From correlation’s value, both had a strong association. That’s why, estimation of number of station users was effective using method of this research.
Figure 11. Comparison result between the number of estimated passengers and MLIT data (※)

Figure 12. Comparison result between the number of estimated passengers and MLIT without 4 extreme stations (※)

6. CONCLUSION AND FUTURE WORK

By the estimation of number of railway passengers using a smartphone GPS log data, we could made the time distribution of railway passengers of day, and it followed that this closely resembled it with the estimation that Ministry did. And, we was able to understand time distribution of railway passengers in each railway. In addition, we also understood that it was associated between the data of the number of the passengers getting on and off according to the station in the numerical land information downloading service and the data estimated of the number of the passengers getting on and off according to the station using smartphone GPS log data. The estimate of the railroad ridership using the personal movement history was very near to an actual feeling as a result of this study and I forced it and was associated with the estimation that a country went to and was very effective.
From the above, the result of the estimation of the number of railway passengers using individual movement trajectory was near to our feeling, there was a strong association with the estimation by the Ministry, the estimation using method of this research was very effective.

In this estimation, we couldn’t evaluate numerically about the time distribution of railway passengers and the time distribution of railway passengers on each railway.

The future works are evaluation numerically about the estimation result of this research using the data of the number of passengers on each railway and the data of the number of passengers of origin station from destination station, and expanding the estimation area to all of Japan because this research limited only Tokyo metropolitan area.

ACKNOWLEDGMENT

This paper was supported by ZENRIN DataCom CO., LTD.

REFERENCES


Ono Natsumi, 2013. Using Large-Scale, Long-Term GPS Data from Mobile Phones to Identify Transportation Modes and Analyze Mobility in the Tokyo Metropolitan Area, The University of Tokyo, Graduate School of Frontier Sciences Department of Socio-Cultural Environmental Studies, Master’s Thesis


National Land Numerical Information download service http://nlftp.mlit.go.jp/ksj/gml/datalist/KsjTmplt-S12-v2_1.html, last access at 2016-04-25

Hara, Kuwahara (2013) , Study on map matching technique and frequency change type pinpointing for the data which are sparse, The 33th Japan Society of Traffic Engineering proceedings , pp. 415-418