Methods for Analyzing Passenger Flows During Train Traffic Disruption Using Accumulated Passenger Data

Taketoshi KUNIMATSU                  Chikara HIRAI
Transport Operation Systems Laboratory, Signalling and Transport Information Technology Division

When rescheduling train traffic after an operational disruption, train operation companies endeavor to take passenger flows into account. In this paper, a method was developed utilizing accumulated passenger data and records of arrival and departure times of each train to estimate passenger flows in such situations. The first step was to devise a visualization method to understand the relationship between rescheduling arrangements and passenger flows. Multiple regression analysis was also applied to collected data from the previous twelve months in order to develop a model for estimating passenger flows during traffic disruption. The methods are verified by applying them to actual cases of train traffic disturbances, which confirmed their reliability and effectiveness.

Keywords: operational disturbance, timetable, train rescheduling, passengers’ flow, multiple regression analysis

1. Introduction

When there is a train traffic disruption, a series of timetable modifications, such as implementation of shuttle services, cancellation of trains, or change in train departure order, are made to make up for the disruption. This is called train rescheduling. Train operating companies have to recover a disrupted situation as quickly as possible in order to minimize passenger inconvenience. They are always monitoring the state of train traffic to detect an accident on the rail line as soon as possible. If a disruption occurs, the railway company will deploy certain measures to deal with the problem, such as laying on extra shuttle services to preserve the transport capacity on other sections where trains can operate normally. In disrupted conditions, passenger behaviour becomes complicated because they choose alternative routes based on announcements about the traffic situation. Those in the train companies who are responsible for maintaining stable traffic have to predict passenger behaviour based on experience and knowledge in order to make appropriate rescheduling plans, but forecasting such behaviour is difficult to do.

At the same time, a number of train operating companies have already introduced data collection systems in order to analyze daily traffic conditions and passenger flows. In practical terms, such systems collect and record arrival and departure times of each train at each station, and the estimated number of passengers of each train. The accumulated data can potentially be used to determine certain features of passenger flows in disrupted traffic conditions.

This paper proposed a passenger flow analysis methods using the accumulated data which can be applied to support decision-making in train rescheduling. First, a method for visualizing the collected data was devised for days when traffic was disrupted, with a view to understanding the relationship between train rescheduling and passenger flows. The aim of the method is to analyze the difference in passenger flows between ordinary operational conditions and a period when traffic is disrupted. Follow-

2. Train rescheduling and passenger flows during operational disruption

2.1 The disruption and rescheduling during the disruption

In case of a problem which leads to a part of the rail line being blocked, traffic becomes disrupted until the blockage is lifted. The usual ramifications of such a situation include transport reduced services until the blockage is lifted, or passengers having to face overcrowded trains and delays. To solve these problems, train companies try to adapt services by providing temporary shuttle services, or prolonging the dwell time of trains in each station to decrease the congestion on overcrowded trains.

In order to produce an optimal rescheduling plan for passengers it is vital to understand how many passengers are caught up in the disruption in order to know how many trains are required. For example, if there are more passengers needing to travel between Stations A and B than between Stations B and C, and the blockage section is located at Station C, it would be preferable to give priority to ensure train traffic by running extra shuttle services between Stations A and B. Similarly, even if there are many passengers needing to travel between Stations A and D in the same situation, if it can be assumed that most passengers will be able to take a detour using another rail line between Stations A and D, it is better to cancel several trains to recover in a shorter time.
2.2 Passenger flows in disrupted situations

When there is a disruption, passengers may wait for operations to resume, or opt for alternative routes towards their destination, or quit their journey, etc. The general assumption about passenger behaviour however is the drawn from the experience of traffic controllers who have limited information provided by drivers, staff members, or video cameras in stations. As a result, there are some cases where it takes too long to recover normal traffic condition, because the unpredicted increase of the dwell time occurs by the congestion of particular stations or trains, and the rescheduled alternative did not work effectively. It is important to support traffic controllers to understand the passenger flows.

Given that passenger flows during a disruption may be influenced by a whole range of factors, it is difficult to arrive at accurate predictions of movement. These factors include day of the week, period of time, traffic conditions on the day, congestion of trains or stations, or availability of detour routes. This is compounded by the fact that each disruption is quite unique and they rarely occur in the same place, location or for the same duration. Therefore even if a similar disruption has occurred before, passenger flows are unlikely to follow the same pattern since information given to passengers by crew or staff members may be different. Consequently when trying to predict passenger behaviour, content of passenger announcements must be taken into account as well as the type of disruption and the availability of detour routes.

2.3 Related work

Some studies have been conducted on passenger flows during disruptions. Muto [1] conducted a WEB questionnaire survey among passengers faced with operational delays, with questions about whether they used detour routes or waited until operations resumed. Following an analysis of the results, Muto devised a behaviour choice model for detour routes based on estimated time to the destination both with and without detour routes. Kunimatsu et al. [2] proposed a train rescheduling algorithm reflecting passenger flows, by using a train operation and passenger flow simulation technique incorporating the detour route choice model developed by Muto. However, these methods are based on the number of passengers estimated through surveys or simulations. They do not use the number of passengers measured on the target day.

Another tool exists for predicting the number of passengers using recorded data. Myojyo et al. [3] analyzed passenger Origin-Destination data collected from automatic ticket checkers at stations, and proposed a method for predicting the number of passengers, by using decreasing rates of passengers recorded in a similar disruption occurred in the past. By this method, we can surely understand the tendency of the passengers’ volume for the particular pairs of origin and destination stations. Nevertheless, the latter is not sufficient for planning train rescheduling because this method does not provide clear insight into the volume of passengers in each section and yet this is vital for the train rescheduling when the number of temporary shuttle services and trains to be suspended have to be determined.

2.4 Requirements for passenger flow analyses

From the above, the following requirements were identified to carry out a passenger flow analysis, and an analysis method was then designed to satisfy them.

1) We can understand the number of passengers on each section.
2) We can analyze the relationship between the passenger flows and the details of the disruption, including passengers’ availability of detour routes, announcements of the estimated time to resume operation, and so on.
3) We can easily analyze the relationship between the passengers’ flow and the train rescheduling arrangements.

3. Recorded data and passenger flow analysis methodology

3.1 Recorded data about daily train operations

Today, traffic control systems, train cars, and automatic ticket checkers use IT making it possible to collect and accumulate various data about daily train operation. Some examples of the recorded data are as follows.

1) Actual arrival/departure time

This is the recorded data of actual arrival/departure time of each train for each station on a specific day. The values used for this data are calculated from records of the track circuits which give train position, and information about timetables held in the main system. Data is then accumulated in the central system. All necessary data about trains and stations within the area covered by the main system can be retrieved.

2) Number of passengers on each train and in each section

This data is information about the number of passengers aboard each train in each section between two adjacent stations. This is estimated by the sum of the weight of each passenger on board measured by the load compensating device installed on the train. The data is transmitted to the ground data centre every time a measurement is sent through wireless communication devices. The main function of the load compensating device is adjustment of the braking power according to the load of the train. It detects the load from the sway of the helical springs on the trucks, or the pressure in the air springs. The sum of the weight of each passenger is estimated using information about the load of the train, and the number of passengers is calculated dividing it by the average weight of passengers.

3) Description of the disruption

Information about the disruption such as time of day, day of the week, location, type of problem, time, and time at which operations resumed.

4) Log of passenger announcements

Information about passenger announcements which were made at the time, including estimated time before operations were supposed to resume, and accuracy of this delay.
3.2 Research methodology

This paper directly analyzes the recorded data with statistical methods, and extracts the characteristics of passenger flows during a disruption.

First, a method was devised to visualize passenger flows on the day. Using this method, lines indicating rescheduled trains on the train diagram are coloured differently according to the number of passengers on board each train. The next step was to predict passenger flows during the delay. This is done using data collected for the target rail line over the previous year, and then building a prediction model to calculate the number of passengers passing each section between two adjacent stations. By using multiple regression analysis, different prediction models can be designed to correspond to each time period before or after resumption of operations.

The feature of the proposed method is that, collecting recorded data over a long period would provide sufficient samples for statistical analysis, on both days with or without operational disruptions. This in turn would make it possible to analyze the difference between the passenger flows in ordinary situations with those when there is a problem.

4. Visualization method

To analyze the relationship between train rescheduling and passenger flows, regular volume of onboard passengers for each train and section must be calculated. After that, both train operation and passenger flows for the target day with disruption can be visualized using recorded data.

4.1 Calculation of regular volume of passengers

It is important to understand the regular value for the number of passengers on board in normal circumstances, in order to be able to recognize the change in disrupted conditions. Even so it is not easy to determine what a typically regular value is, because the number of passengers on board may change to some extent as a result of the weather, the day of the week, or a small train delay of several minutes.

In this paper, the median of the recorded values within the target period was used as the regular value for each day of the week (except national holidays). The reason for having a value for each day of the week was that it was found that the difference in the recorded values was mainly affected by the day of the week. Furthermore, average values were not used because of the possibility of them being distorted by abnormal records such as those collected on days when a large event was being held. Recorded data over about a month was used to calculate the regular values.

4.2 Visualization of data of actual arrival/departure time and number of passengers on board

Recorded data was visualized as follows. First, the position of lines indicating trains on the time-space diagram were defined, according to actual arrival/departure time data for each train and section. Next, the colour of lines according to the number of passengers on board was defined on the target day. The colour varied according to the level of congestion rate on each train. Blue was used for low

Fig. 1 Example of the visualization of passenger flows
congestion, and became redder as crowding increased. The colour changed for every 25%, from 0%. A congestion rate of over 200% was shown in pure red. Bold lines showed crowded trains with over 100% congestion rate. The data of some trains could not be retrieved due to the type of car. These cases were represented by grey dotted lines, to indicate trains for which there was no data. Finally, a symbol “●” was added to the train line when congestion differed over 20% with the regular value. When the value was exceeded the symbol was coloured red, where it was lower, the symbol was coloured blue.

Figure 1 is an example of the visualization of a real disruption. Two different railway lines are operated in parallel between Stations 11 and 20. On this day, the disruption occurred at Station 26, and traffic going through Station 26 was suspended for about an hour. In the diagram in Fig.1, the traffic on the disrupted rail line is indicated at the bottom, and that of the other rail line is at the top. This figure reveals certain trends. First is the collection of red symbols at the top of the diagram. This indicates that many passengers on the disrupted line abandon this route and opt to take a train on the other line. Next, as it took a long time to resume operation, the timetable was rescheduled to set three turn backs of the outbound trains at Station 24, and three extra shuttle services of the inbound trains before operations resumed. These extra trains are indicated with bold black circles at the bottom of the diagram. The bold lines indicated that these trains had a congestion rate of over 100%. This trial shows that the proposed visualization method makes it easier to understand passenger flows and the effects of the rescheduling.

5. Passenger flow prediction model

5.1 Aim and Target

Next, in order to understand how many trains should be set for each section before or after operations resumed, a passenger flow prediction model was designed with recorded data from the few days in the run up to the disruption. In the same way as in the previous chapter, both the data of the actual arrival/departure time and the number of passengers on board was collected. In order to ensure that the model had information from a sufficient number of disruptions data from the target rail line was gathered for over a year. Information was therefore collected from 28 cases, included time during which operations were suspended and when they resumed.

The target number to be predicted is the increasing/decreasing rate of the number of passengers passing each section between two adjacent stations in the period before or after resumption of operations, compared to that in the ordinal operation. The reason not to estimate each number of passengers on board is that, we want to exclude the effects of the daily variation of passenger flows and the local variation of train intervals. We also want to know whether the number of trains provided as a result of rescheduling is enough or not, from a comprehensive point of view. By calculating the increasing/decreasing rates, we can use the recorded data of both cases where the operational disturbance occurs at 11a.m. and 2p.m., to build the same prediction model for the daytime.

When we calculate the increasing/decreasing rate, the regular value of passenger volume is necessary. So, we define the median of passenger volume in a month as the regular value for each day of the week (except national holidays).

5.2 Model building with multiple regression analysis

Multiple regression analysis was applied to build the prediction models. The increasing/decreasing rate of the number of passengers passing each section between two adjacent stations in the period before or after resumption of operations was set as the explained variable. The following 18 items grouped into four categories were set as explaining variables (Fig. 2).

1) Information about the disruption location, blocked section, time of disruption and block duration
2) Structural features of the rail line and each section whether another rail line is running in parallel, whether each side of the section has a station with a connection to another rail line
3) Whether operations resumed or not at the predicted time (prediction models for after resumption)
4) Actual result number of trains set before resumption (prediction models after resumption)

To choose only significant explaining variables, we used a stepwise method, and discarded insignificant variables.

Table 1 shows a part of prediction models for the same rail line as section 4.2. This table shows the formulae for predicting the increasing/decreasing rate of passenger volumes for each section, time period and direction. The value in the table indicates the partial regression coefficient for each prediction formula and explaining variable. When the explaining variable is insignificant for the formula, the value is given as “0.0.” Explaining variables with “D” are dummy variables. They take 1 when the condition of the day fits and 0 when they do not. The multiple correlation coefficients for each formula of between 0.80 and 0.88, indicates good accuracy.
5.4 Validation of the prediction model

To validate the applicability of the proposed method to other operational disruptions, five cases of disruption which had not been used for the modeling were selected. The actual volume of passengers was then compared with the estimated volume obtained using the proposed method to check for correlation.

Figure 3 shows the result from two cases. This figure and the result of quantitative discussion about the error range show that the method has acceptable accuracy and can therefore be applied for train rescheduling.

6. Conclusions

Two methods were proposed for analyzing passenger flows using recorded data, in order to improve rescheduling. The visualization method offers a clear view of the relationship between train rescheduling and the increase/decrease in number of passengers aboard each train. It can be applied to the evaluation of the results of train rescheduling in the control room. On the other hand, the prediction models reflect information on not only operational disturbances and availability of detour routes, but also the history of announcements and the number of trains before resumption of operations, to predict the volume of passengers precisely. It can be used to estimate the number of trains required before or after operations are resumed, with a view to coping with the predicted volume of passengers.

It would be useful to apply the method to more disruptions in the future in order to validate the accuracy of the prediction model. It is also hoped that the models could be applied to other rail lines to ascertain the applicability of the proposed methods in different conditions.

References


Authors

Taketoshi KUNIMATSU
Assistant Senior Researcher, Transport Operation Systems Laboratory, Signalling and Transport Information Technology Division
Research Areas: Transport Behaviour Modeling, Scheduling, Simulation

Chikara HIRAI, Dr. Sc.
Laboratory Head, Transport Operation Systems Laboratory, Signalling and Transport Information Technology Division
Research Areas: Transport Scheduling, Optimization, Simulation