TRAIN TRAFFIC SIMULATION ALGORITHM BASED ON HISTORICAL TRAIN TRAFFIC RECORDS

SHUHO WATANABE[2], YUKI MORI[1], YOSHINOBU TAKATORI[2],
KAZUSHIGE YONEMOTO[2] & NORIO TOMII[1]
[1]Chiba Institute of Technology, Japan
[2]Tokyo Metro Co., Ltd, Japan

ABSTRACT

We have developed a train traffic simulator in order to precisely simulate the delay occurrence and propagation which must be different depending on particular "situations." By "situation" we mean the weather, a day of the week, season and so on. It is well known that occurrence and propagation of delays are different depending on these situations. Thus, it is necessary to develop a simulation algorithm which can exactly simulate the difference in occurrence of delays in order to use the simulator to evaluate effectiveness of delay reduction measures. The basic idea of our simulator is to use the historical train traffic records. By constructing regression trees from the historical train traffic records, we can know dwell times, running times and intervals of trains in each situation. By incorporating the results obtained from the regression tree to our macroscopic simulator constructed based on the longest path algorithm, we can construct a train traffic simulator which exactly simulate train traffic reflecting the "situation." We have evaluated the simulator using the actual data and we have confirmed our approach is very promising.

Keywords: delay, simulator, regression tree, datamining.

1 INTRODUCTION

One of the problems of railways in urban areas of Japan is that small delays very often happen during morning rush hours. This is because in urban areas, trains are running very densely to meet the huge amount of demands for transport service by railways. As a matter of fact, in many lines in Tokyo area, 25 to even 30 trains which consist of 10 cars (about 200m long) are running per hour per direction on a double track.

In order to prevent such delays from happening, railway companies are taking various kinds of countermeasures, such as a revision of timetables, an improvement of signalling systems, deployment of more staff on platforms and so on [1], [2].

Railway companies have an intention to evaluate how effective these countermeasures will be before they actually implement these countermeasures. In order to evaluate the effectiveness, simulation is used. The simulator assumes primary delays and simulates the train traffic considering the mechanism of delay propagation so that we can know how the delays will occur and propagate.

One of the issues with regard to the simulation algorithm is that occurrence and propagation of delays are not the same every day. For example, it is known that trains tend to be delayed more severely on rainy days. Another example is seasonal fluctuations. It is believed that in April, delays are larger because there are many freshmen for the schools and companies who have not got accustomed with commuting by trains. In the middle of August, trains are very punctual because many people take days off and trains are less congested. One more example is the midnight of Fridays. Because trains, especially the last train are very congested on Friday evenings, delays of those trains become much larger than the trains on other days of a week.

These observations suggest that we need to develop a simulator which can exactly simulate the train traffic reflecting the situation we are interested in. If we are interested in

WIT Transactions on The Built Environment, Vol 181, © 2019 WIT Press
www.witpress.com, ISSN 1743-3509 (on-line)
doi:10.2495/CR180261

simulating trains on rainy days, it is required to simulate the occurrence and propagation of delays peculiar to rainy days, for example.

In order to develop such a simulator, we introduce an idea to utilize historical train traffic records. More in details, we construct the simulator as follows:

1. We make regression trees, which is one of the most commonly used technique in data mining, for the running times, dwell times and headways of each train at each station from historical train traffic records.
2. From the regression trees, we can know the rules which specify running times, dwell times and headways for each situation.
3. We construct a train traffic simulator based on the longest path method in which the above-mentioned rules are installed.

We have developed the simulator and confirmed it can simulate train traffic for different situations with an enough accuracy and the algorithm is very promising.

2 DIFFERENCE OF DELAY EMERGENCE

We will show some examples to demonstrate that occurrence and propagation of delays are different depending on "situations". Fig. 1 shows a difference depending on the weather. Fig. 1(a) is a chromatic diagram for a sunny day whereas Fig. 1(b) shows a chromatic diagram for a rainy day. A chromatic diagram is a diagram in which train lines are coloured reflecting the amount of the delay of each train [3], [4]. The colour changes from indigo to blue, green, yellow, orange and red as the delay increases. By comparing Fig.1(a) and Fig. 1(b), we can know that delays are much larger on rainy days.

In Fig. 2, we show the difference depending on the days of a week. Fig. 2(a) and 2(b) are chromatic diagrams for a Monday evening and Friday evening respectively. As you see, trains, especially the last train tend to be delayed more seriously on Friday evening.

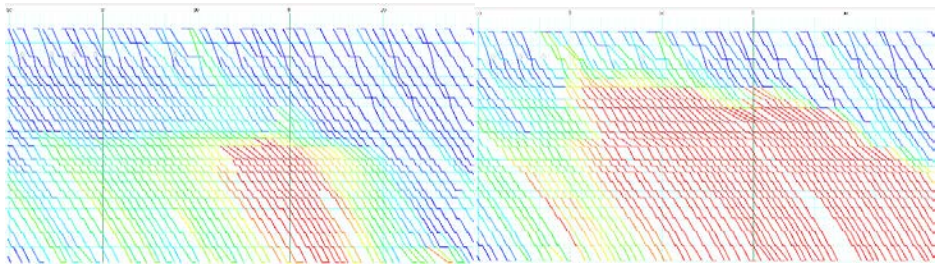
Those observations suggest that we need to develop a simulator which can accurately simulate the difference of delay emergence reflecting the "situation."

3 TRAIN TRAFFIC SIMULATION

3.1 Simulation models

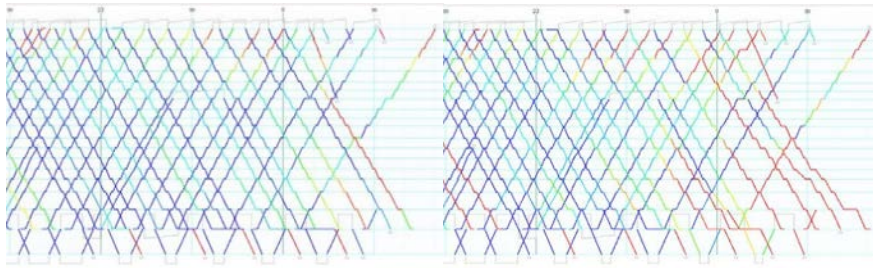
There exist two categories of simulation algorithms [5]. One is a microscopic simulation algorithm in which trains' movement is simulated continuously by solving differential equation. A lot of research has been reported which proposes the microscopic simulation algorithm [6]–[9]. One of the merits of the microscopic simulation is that it can simulate the movement of trains in detail. For example, we can analyze the influence imposed by the signaling systems and energy consumption. In addition, we can analyze drivers' manipulation [10] as well. One of the demerits of the microscopic simulation model is that we need to prepare various kinds of data such as performance data of trains, specification data of signaling systems, profiles of the line which contain curvature, gradient and the exact locations of stations and so on.

The second type of simulation algorithms is the macroscopic simulation. In macroscopic simulation, only the occurrence times of arrival and departure of trains are calculated. One of the merits of the macroscopic simulation algorithm is that we need not prepare various kinds of data and it works much faster than the microscopic model.



(a) (b)

Figure 1: (a) Sunny day; (b) Rainy day.



(a) (b)

Figure 2: (a) Monday; (b) Friday.

Two types of algorithms are reported for the macroscopic simulation model. One is an event driven model [11]. This model has a clock inside the simulation model and the occurring times of the events of trains' departure and arrival are calculated as the clock proceeds by a time unit. The other type is a model based on the longest path algorithm. Train diagram is expressed by an acyclic directed graph and by calculating the weight of the longest path from the imaginary start node to each node, we can get the departure and the arrival times [12], [13]. One of the merits of the longest path model is that it works faster than the event driven model as long as we can assume there is no change in the timetable during the simulation process.

In this paper, we use the macroscopic model based on the longest path algorithm because we only need the arrival and the departure times and the longest path algorithm works very fast.

3.2 Estimation of dwell times

In railway lines where trains are running densely, the main cause of delays is an increase of dwell times. Thus, it is crucial to exactly estimate dwell times in particular situation if we want to realize a simulator with a high accuracy. But estimation of dwell times for various kinds of "situation" is very difficult. It is believed that dwell times are dependent on the number of passengers who get off and who get on and the number of passengers are different from one "situation" to another. An existing research proposes an algorithm to estimate the number of passengers for each door of a train and estimate the dwell time of the train [14].

This approach, however, needs several assumptions; such as distribution of passengers in a train, choice of routes, relationship between the dwell time and the number of passengers; not only the passengers who get on/off but the passengers inside the train. It is not easy to prepare reasonable assumptions. Thus, we need to develop an algorithm to estimate dwell times which does not need such assumptions.

4 TRAIN TRAFFIC SIMULATION ALGORITHM BASED ON HISTORICAL TRAIN TRAFFIC RECORDS

4.1 Basic concept – fact-based simulation

The basic concept of our simulation algorithm is illustrated in Fig. 3. We assume that train traffic for some particular situation could be simulated if we properly specify the dwell times, running times and intervals of trains which correspond to the situation.

We introduce an idea to get these data from the historical train traffic records. The basic idea is as follow:

- We estimate the running times, dwell times and intervals of each train at each station from the historical train traffic records.
- We construct a simulator in which running times, dwell times and headways appropriate for the "situation" are used.

We call this simulation model "**a fact-based simulation model**" because "facts" are obtained from historical train traffic records.

4.2 Historical train traffic records

Historical train traffic records are the data which contain train numbers, actual arrival times at each station, actual departure times at each station and so on. Thus, we can calculate the delay by comparing the data with planned times. Historical train traffic data are obtained from a train traffic control system. Hence, we can get the data for every day for a long period.

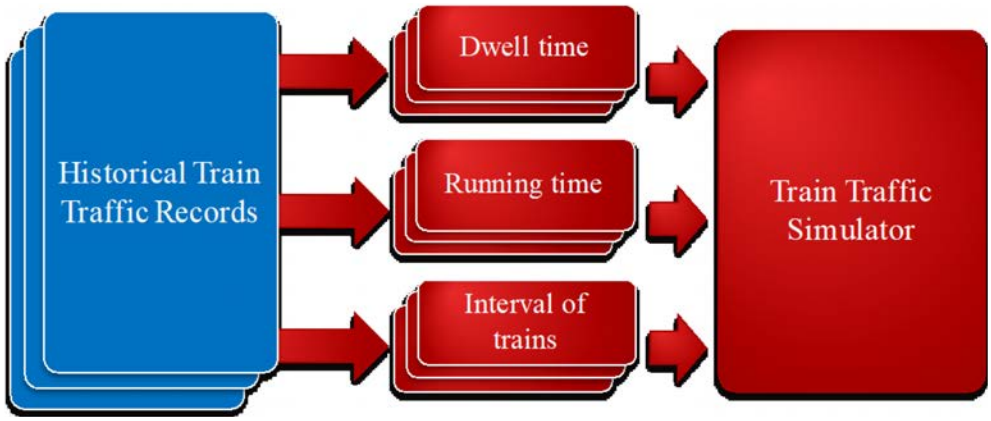


Figure 3: Basic structure – fact-based simulation.

4.3 Estimation of dwell times by regression trees

Decision trees are a non-parametric supervised learning method mainly used for classification. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. Decision trees in which the target variable can take a continuous value are called regression trees. Decision trees and regression trees are widely used in the data mining community [15]. The decision trees and the regression trees are typically used to decide to which class a given datum belongs. By tracing the tree using the conditions of the given datum, we can know to which group the datum should be classified.

We try to construct a regression tree from historical train traffic records to estimate the dwell time of a train at a station in a particular situation. We construct a regression tree for each train for each station.

We will show an example of a regression tree for a dwell time in Fig. 4. This regression tree shows that the dwell time of this train at this station should be clustered into three clusters and the parameters to identify to which cluster a specified situation belongs are: months and the day of the week. So, from this regression tree, we can know that from January to November, the dwell time of this train is about 50 seconds. From Monday to Wednesday in December, the dwell time is about 50 seconds whereas the dwell time significantly increases on Thursday and Friday of December.

The same procedure could be applied to estimate running times and intervals of trains.

4.4 Simulation algorithm based on the longest path algorithm

In a directed acyclic graph used for simulation by the longest path algorithm, a node corresponds to an event (departure / arrival) and an arc expresses the chronological order of the events. A weight of an arc depicts the minimum time necessary between the two events connected by the arc.

In our algorithm, we use the median obtained from the regression tree as the weight of the arc.

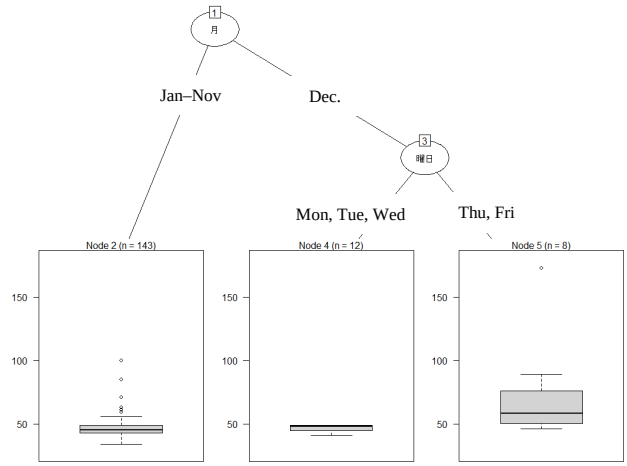


Figure 4: Regression tree – example.

5 NUMERICAL EXPERIMENTS AND DISCUSSIONS

5.1 Numerical experiments – preliminary

We have implemented a simulator based on the ideas we have introduced in the previous section and conducted numerical experiments using actual data. The target railway line is a subway line in the center of Tokyo and many trains go directly from this line into the line of other railway companies of suburban areas. On Friday evening, before midnight in particular, trains are very congested and trains tend to be delayed. We tried to simulate several last trains on Friday evening.

We first constructed regression trees of each train at each station to estimate the dwell times of trains at each station. Some of the results are shown in Fig. 4 and Fig. 5.

We have analyzed the running times and learned that there do not exist big differences for running times and the running times of these trains are a little bit smaller than the planned running times. So, for running times, we did not use regression trees, but we used the values which we get by subtracting 10 seconds from the planned running times.

We gave primary delays of each train when they depart from the first station, which are the same as the actual data.

We show the result in Fig. 6. This is a result we got assuming the target day is Friday in December. Fig. 6(a) shows the result of our simulation and Fig. 6(b) shows the chromatic diagram of the actual data.

We have analyzed the difference between actual data and the simulation results as shown in Fig. 7. Fig. 7(a) is a histogram to show the difference of departure times and Fig. 7(b) is a histogram to show the difference of arrival times.

From Figs 7(a) and 7(b), we may well conclude that our simulation algorithm succeeded to output a result which is very close to the actual data.

But we have to note that there exist a couple of events, which are different by around four minutes. We analyzed the result and learned that a train was intentionally delayed to keep a connection from a train of other line.

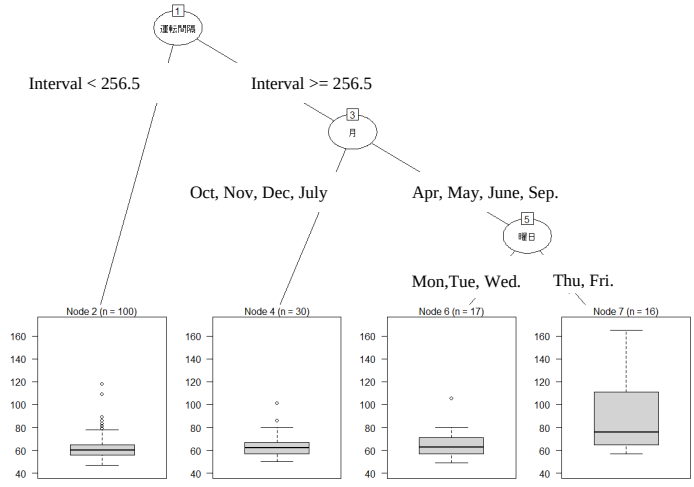
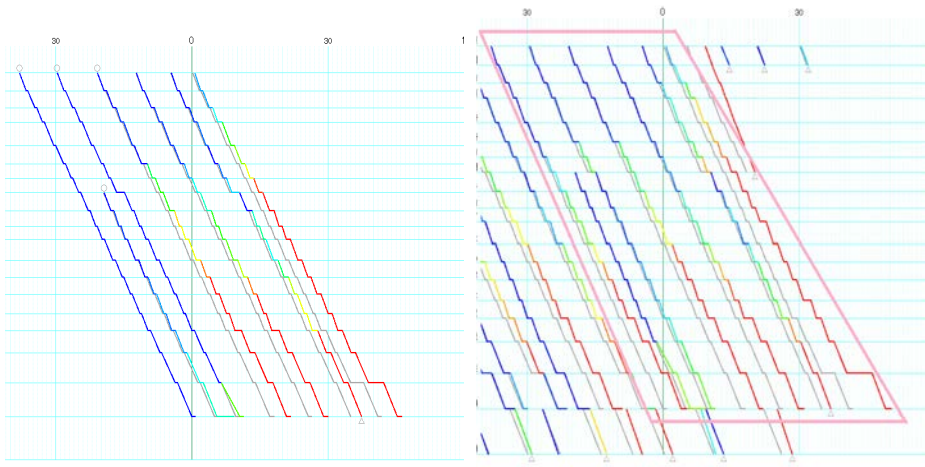
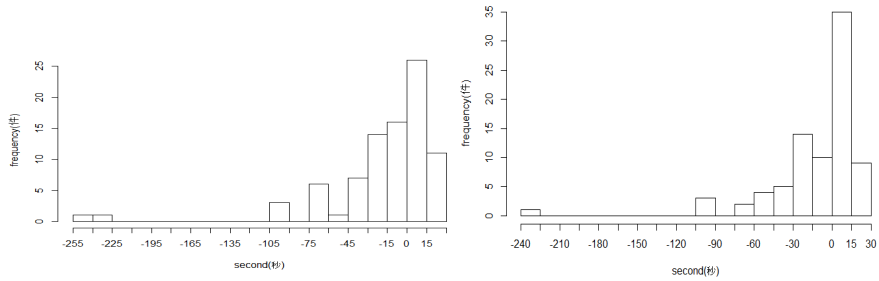


Figure 5: Regression tree-2.



(a) (b)

Figure 6: (a) Simulation result; (b) Actual.



(a) (b)

Figure 7: (a) Difference – departure; (b) Difference – arrival.

5.2 Discussions

What we did in this experiment was to try to simulate the train traffic of particular one day. But because the purpose of this research is to examine effectiveness of delay reduction measures, we should apply this process repeatedly to other days of the same situation and compare a median or an average with those of actual data. In that process, we should give as an input of simulation an irregular delay (such as the delay by keeping a connection) as well as the primary delays.

The results of Fig. 7 show our approach is very promising and we will apply this algorithm to other situations such as rainy days and so on.

Another important topic is to find what kind of "situations" exist. We already know that occurrence of delays is different on Friday evening and rainy days but there must exist other "situations." So, it is also useful to develop an algorithm to find such "situations."

6 CONCLUSION

We have introduced a train traffic simulator which exactly simulate delay occurrence and propagation in some particular situation. We proposed an idea to estimate dwell times etc. in each situation using regression trees constructed from historical train traffic records. Then we introduced an idea to incorporate the obtained dwell times etc. in the simulator which is designed based on the longest path algorithm.

We have implemented the simulator and confirmed it can simulate train traffic for different situations with an enough accuracy and the algorithm is very promising.

Using this simulator, it becomes possible to analyze and estimate with a high accuracy how delay reduction measures such as revision of timetables which takes long time and improvement of facilities which is usually very expensive are effective to improve robustness well in advance. We are planning to take measures to decrease delays based on the results of this simulator.

ACKNOWLEDGEMENTS

The fifth author is partly supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Scientific Research (C) 15K01199.

REFERENCES

[1] Yamamura, A., Koresawa, M., Adachi, S. & Tomii, N., How we have succeeded in regaining punctuality in Tokyo Metropolitan Railway Network? *10th World Congress on Railway Research (WCRR)*, Sydney, Australia, 2013.

[2] Yamamura, A., Koresawa, M., Adachi, S. & Tomii, N., Taking effective delay reduction measures using delay elements of indices on the Tokyo metropolitan railways. *COMPRAIL2014*, Rome, Italy, 2014.

[3] Ushida, K., Makino, S. & Tomii, N., Increasing robustness of dense timetables by visualization of train traffic record data and Monte Carlo simulation. *9th World Congress on Railway Research (WCRR)*, Lille, France, 2011.

[4] Tomii, N., Beyond the wave of "big data" – How we can realize robust train operation using train operation record data? *Japanese Railway Engineering*, No.179, **53**(2), 2013.

[5] Hansen I.A. & Pachl, J. (eds), *Railway Timetabling & Operations: Analysis – Modelling – Optimization – Simulation – Performance Evaluation*, 2nd ed., DVV Media Group/Eurailpress, 2014.

[6] OpenTrack. http://www.opentrack.ch. Accessed on: 1 Mar. 2018.

[7] Hürlimann, D. & Nash, A., Railway simulation using Opentrack. *COMPRAIL2004*, Dresden, Germany, 2004.

[8] Janecek, D. & Weymann, F., LUKS – Analysis of lines and junctions. *Proceedings of 12th World Conference on Transport Research*, Lisbon, 2010.

[9] RailSys. http://www.rmcon.de. Accessed on: 1 Mar. 2018.

[10] Ochiai, Y. & Tomii, N., Punctuality analysis using a microscopic simulation in which drivers' behaviour is considered. *Journal of Rail Transport Planning and Management*, **5**(3), pp. 128–145, 2015.

[11] Sato, A., Ikeda, H. & Ono, K., Traffic simulation system of Shinkansen (in Japanese). *Journal of the Society of Instrument and Control Engineers*, **19**(7), 1980.

[12] Abe, K. & Araya, S., Train traffic simulation using the longest path method (in Japanese). *Journal of Information Processing Society of Japan*, **27**(1), 1986.

[13] Tomii, N. & Ikeda, H., A train traffic rescheduling simulator combining PERT and knowledge-based approach, ESS'95. *European Simulation Symposium*, Elrangen, Germany, Nov. 1995.
[14] Kanai, S., Shiina, K., Harada, S. & Tomii, N., An optimal delay management algorithm from passengers' viewpoints considering the whole railway network. *Journal of Rail Transport Planning & Management*, **1**(1), pp. 25–37, 2011.
[15] Berry, M.J.A. & Linhoff, G., *Data Mining Techniques – For Marketing, Sales and Customer Support*, John Wiley & Sons, Inc. 1997.