Estimation of passenger origin-destination matrices and efficiency evaluation of public transportation

Mirai Tanaka
Department of Industrial Administration,
Faculty of Science and Technology,
Tokyo University of Science,
mirai@rs.tus.ac.jp

Takuya Kimata
Department of Industrial Administration,
Faculty of Science and Technology,
Tokyo University of Science,
7412042@alumni.tus.ac.jp

Takeshi Arai
Department of Industrial Administration,
Faculty of Science and Technology,
Tokyo University of Science,
tarai@rs.noda.tus.ac.jp

***Abstract*—In analyzing and evaluating the efficiency of a transportation system, origin-destination matrices are important. In this paper, we propose an optimization model to estimate origin-destination matrices. The proposed model is easily handled since it is a convex quadratic optimization model. We validate the proposed model using numerical results from real-world data. In addition, we use our model to evaluate the efficiency of a real-world bus service.**

***Keywords*—OD data estimation, constrained least square method, transportation**

I. Introduction

Analysis and evaluation of the efficiency of a public transportation system is important since the organization that operates that system is often financially troubled. Using the results obtained from such an analysis, the organization might be able to make crucial decision such as whether or not to discontinue the service. In fact, many bus services operated by local governments in Japan are experiencing a financial deficit. An efficiency evaluation of such services contributes to sound financial health.

An origin-destination (OD) matrix is important in analyzing and evaluating a transportation system. For example, predicting levels of congestion and identifying an appropriate location for a facility often require OD matrices. Hence, many researchers have studies the estimation of OD matrices for a long time (see Survey [1]). Historically, OD data has been estimated from a traffic census. More specifically, we count the amount traffic for all pairs of origins and destinations. Recent studies [2]–[4] have proposed that passenger OD data should be collected automatically. However, due to financial constraints some public transportation companies find it difficult to collect accurate OD data because the equipment to do so is extremely costly.

On the other hand, it is easy and more affordable to collect input-output (IO) data. In fact, whereas the collection of OD data requires $O(n^2)$ of cost, the collection of IO data only costs $O(n)$. Thus, the estimation of OD data from IO data has also been investigated. Nihan and Davis [5] proposed a recursive prediction error method to estimate OD matrices from IO data. This method requires historical IO data and returns OD data. However, it takes a significant amount of time and money to collect sufficient historical data. In Japan, it is difficult for a typical local government to collect such massive amounts of data since local governments are experiencing financial pressure. They require approaches that work well without historical IO data.

In this paper, we focus on estimating passenger OD data for a bus line that does not have any branches and loops by only using a small snapshot of historical IO data. In an analysis of a bus service, it is easy to collect IO data since only the number of passengers boarding and alighting passengers at each bus stop needs to be counted. Usually, a bus line's bus stop has two platforms directing passengers in opposite directions—one platform for inbound traffic and one for outbound traffic. Hence, we can collect two boarding-alighting data sets for each direction.

The remainder of this paper is divided as follows. In Section II, we propose an optimization model to estimate a passenger OD matrix using a boarding-alighting data set. The proposed model is written in the form of a convex quadratic optimization problem. Hence, it can easily be solved. In Section III, we investigate the validity of the proposed model. In that section, we present the results of our preparatory numerical experiment in which we used OD data from a real-world subway line. In Section IV, we evaluate the efficiency of a real-world bus line. We quantify the impact that discontinuation of the bus service would have on the following factors: the changes in the bus stops used by passengers and changes in passengers' route choices. Section V presents our concluding remarks.

II. OD data estimation model from IO data

In this section, we propose an OD data estimation model from a boarding-alighting data set of a bus line. We assume that the bus line has no branches and loops. We then assign a number for each bus stop, ranging from 1 to n, where n is the number of bus stops. We focus on the direction corresponding to the ascending order, and the opposite direction can be dealt with in a similar way.

Our estimation model requires boarding-alighting data. Let b_i and a_i be the number of passengers boarding and alighting a bus at bus stop i, respectively. For example, $a_1 = 0$ holds since no one alights at the starting bus stop; similarly, $b_n = 0$ also holds.

In our estimation model, x_{ij} (≥ 0) is the decision variable corresponding to the number of passengers from bus stop i to

978-1-4673-8985-3/16 $31.00 © 2016 IEEE
DOI 10.1109/IIAI-AAI.2016.11

bus stop j for $1 \leq i < j \leq n$. The variables must satisfy

$$\sum_{j:i<j} x_{ij} = b_i \tag{1}$$

for $1 \leq i \leq n$ since every passenger boarding at bus stop i must alight at bus stop j $(> i)$. Similarly,

$$\sum_{i:i<j} x_{ij} = a_j \tag{2}$$

for $1 \leq j \leq n$ holds. From these constraints, we can see that $x_{12} = a_2$ and $x_{n-1,n} = b_{n-1}$. However, we cannot determine x_{ij} for another pair (i, j). We then assume passengers alight at bus stop j with a probability proportional to a_j. In other words, we assume

$$x_{ij} = b_i \frac{a_j}{\sum_{k:k>i} a_k} \tag{3}$$

for $1 \leq i < j \leq n$. However, the x_{ij} defined above does not satisfy Constraints (1) and (2). Hence, we minimize the sum of square errors of Equation (3) under those constraints. The optimization model to minimize the sum of squares of the residuals can be written as

$$\begin{array}{lll} \text{minimize} & \displaystyle\sum_{(i,j):1\leq i<j\leq n}\sum \left(x_{ij} - b_i \frac{a_j}{\sum_{k:k>i} a_k}\right)^2 & \\ \text{subject to} & \displaystyle\sum_{j:i<j} x_{ij} = b_i & (1 \leq i \leq n), \\ & \displaystyle\sum_{i:i<j} x_{ij} = a_j & (1 \leq j \leq n), \\ & x_{ij} \geq 0 & (1 \leq i < j \leq n). \end{array} \tag{4}$$

This optimization model is only a convex quadratic optimization model. Hence we can easily solve with a general purpose solver, *e.g.*, the Gurobi Optimizer [6], CPLEX [7], Xpress [8], SCIP [9] *etc.*

III. Model validation

Next, we investigated the validity of our proposed model. Since as far as we know, there are no benchmark OD data for bus services, we used OD data from a subway in London. We downloaded the Rolling Origin & Destination Survey (RODS) data set from the Transport for London website [10]. The data set contains OD data from the Tube, London's underground subway network. We focused on the outbound traffic of the Uxbridge branch of the Piccadilly line from the Acton Town station to the Uxbridge station. Since the structure of the line is simple, we did need not to consider route choice between the two stations. The Uxbridge branch of the Piccadilly line is approximately 20 km and there are 15 stations on that branch. In addition, the fare is based on the zone of the origin and the zone of the destination.

The validation is executed using the following procedure:

1) We computed the number of passengers boarding and alighting at each station from the actual OD data;
2) We estimated OD data by solving Model (4) with the Gurobi Optimizer [6];
3) We compared the estimated OD data with the actual OD data.

We evaluated the quality of the estimate with the correlation coefficient ranging between $\hat{x}_{ij}$'s and x^*_{ij}'s, where x^*_{ij}'s and $\hat{x}_{ij}$'s are the estimated and actual number of passengers. The correlation coefficient was 0.84. Thus, we can see the the strong correlation between them. In addition, Fig. 1 presents a scatter plot of all the x^*_{ij}'s and $\hat{x}_{ij}$)'s. It also verifies the validity of our model.

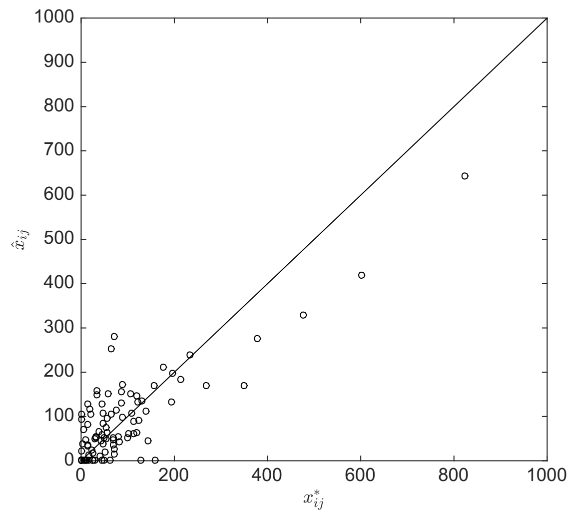

Fig. 1: Scatter plot of all pairs of x^*_{ij}'s and $\hat{x}_{ij}$'s.

IV. Empirical study

In this section, we evaluate the efficiency of a real-world bus service. In our empirical study, we focused on a community bus service operated by Minami Ward in Saitama City, Japan. This bus service has one line without any branches and loops. The line is approximately 10 km, and the fare is based on the distance travelled. The line has 29 bus stops. The first bus stop is Musashi-urawa station west exit and the last stop is Myobana. We numbered each bus stop in ascending order, *e.g.*, the ID number for the Myobana bus stop is 29, and the ID number for the Minami-urawa station west exit bus stop is 14. It is important to note that passengers can transfer to a train at the bus stops 1 and 14.

Bus stops with ID numbers 1–2, 9–14, 18–25, and 28–29 share their platforms with other bus services. Thus, even if the bus service operator discontinued the bus line, these bus stops would still remain; however, bus stops with ID numbers 3–8, 15–17, and 26–27 would be closed. In the text follows, we refer to these bus stops as closed bus stops.

The empirical study is executed in the following manner:

1) We estimated the boarding-alighting data of this community bus line per year from the number of yearly bus service users and daily boarding-alighting data provided by the Saitama City government. We then applied our estimation model to two directions, eastbound and westbound.

2) We quantified the impact of the discontinuation of this bus service based on several assumptions. We evaluated the impact of the discontinuation on the following factors:
   a) Changes in the bus stops: If the bus service operator discontinued the bus service, passengers using closed bus stops would have to use alternatives. We evaluated the number of passengers who would need to switch bus stops and we determined the distances to the alternative bus stops.
   b) Changes in the routes: If the bus line is discontinued, the routes taken by passengers would also be changed. We quantified the effect of the number of transferring passengers, the changes in the bus fare, and the travel time due to the change in the route.

In this analysis, we assume the follow:

- Even after the discontinuation, all passengers will continue to use public transportation, *i.e.*, other bus and train lines.
- A passenger using a closed bus stop would choose the nearest bus stop as an alternative and passengers using other bus stops would continue to use those stops.

Note that the figures in the following results might contain a rounding error.

A. Changes in the bus stops

We evaluated the effect of changes in the bus stops due to the discontinuation of the bus line. Table I presents data on the distance to the alternative bus stops and the number of passengers at each bus stops. The first column of this table shows the ID number of each bus stop. The second column shows the distance from each bus stop to its alternative. The third column shows the number of passenger at each bus stop. From thw information in this table, we can see the following:

- The number of passengers that are using a non-closed bus stop is 30,890. Its percentage of the total number of passengers is approximately 66%. Hence, we can conclude that two-thirds of the passengers would still able to use public transportation if the bus line is discontinued.
- The number of passengers who use closed bus stops is 16,612. In particular, bus stops with ID numbers 5–8 are located more than 300 m from the nearest bus stop and more than 1 km from the nearest station. The number of passengers at these bus stops is 7,285 and it represents approximately 14% of the total number of passengers. We can see that the community bus service is significant for a certain number of passengers. Hence, if the bus service operator discontinues the community bus service, the local government might have to introduce alternative public transportation options, such as shared-taxis.

TABLE I: The distance to alternative bus stops

BS	dist. to alt. [m]	# passengers
3	380	1,989
4	376	1,210
5	541	2,731
6	420	966
7	780	2,992
8	654	597
15	360	427
16	473	1,357
17	417	838
26	242	1,587
27	210	1,918
others	0	30,890
total	—	46,502

B. Changes in passengers' route choices

We evaluated the following aspects related to changes in passengers' route choices: changes in the number of transferring passengers, changes in the bus fare, and changes in the travel time.

1) Changes in the number of transferring passengers: Our analysis shows how the change in the routes impacts the number of transferring passengers. In this analysis, we assumed that passengers chose routes that had the least number of transfers. Table II shows data on the difference between the number of transferring passengers for the original routes and the number for the alternative routes. The first column depicts the change in the number of transferring passengers. The figures in each row depict the number of passengers with and without a bus stop change for each of the transferring passengers. That is, the former figure refers to the number of passengers who were using a bus stop that has changed, and the latter refers to the number of passengers who are using the bus stop that has not changed. As seen in Table II:

- The number of passengers whose number of transfers has not changed is 29,011. This represents approximately 61.1% of the total. Since the fare structure of this community bus service and the alternative bus service are both based on the distance travelled, the fares these passengers pay have not changed.
- The percentage of passengers who have to transfer twice is 5.6%. If the local government regards that percentage as being sufficiently small, based on the limited number of passengers affected, we can conclude that it can discontinue the community bus service.

TABLE II: Changes in the numbers of transferring

# transferring	# passengers	
	w/ BS change	w/o BS change
0	9,086	19,926
1	6,269	9,557
2	1,256	1,408

2) Changes in the bus fare: Our analysis shows how changes in the bus routes impact the cost of the bus fare. In this analysis, we assumed passengers chose the most economic route. Table III presents the difference between the fare for the original routes and the fare for the alternatives routes. The first and the second columns indicate the lower and upper bounds of the change in the bus fare. The negative value means the fare is reduced. The third and the fourth columns are similar to the second and third columns in Table II. For example, the number of passengers whose were using a bus stop that has changed and whose bus fare is reduced by less than 50 JPY is 1,149. We can interpret this result as follows:

- The change in the fare for most passengers is less than 50 JPY. This could be because the fare structure of the community bus service is the same as the fare structure of the alternative bus service. In addition, the fare for 14.4% of the passengers has been reduced because they take a train between bus stops 1 and 14 as an alternative.
- On the other hand, 35.3% of all passengers have to pay more than 100 JPY for their bus and train fare. This difference is due to the need for additional transfers. These passengers account for a substantial fraction of all of the passengers. Hence, we can conclude that the discontinuation heavily affects the bus fare. If the local government discontinues the community bus, it should introduce discount tickets for connecting services.

TABLE III: Changes in the bus fare

add. fare [JPY]		# passengers	
LB	UB	w/ BS change	w/o BS change
−100	−50	0	37
−50	0	1,149	5643
0	50	8,674	14412
50	100	0	842
100	150	1,254	3358
150	200	2,083	4832
200	250	2,954	0
250	300	108	1672
300	350	1	35
350	400	390	59

3) Changes in the travel time: Our analysis shows how changes in the bus routes impacts the travel time. In this analysis, we assumed that passengers chose the fastest route. In addition, we assumed that the waiting time due to transferring from one bus to another is 20 minutes and the waiting time due to transferring from a bus to a train is 10 minutes, since they run every 20 minutes and every 10 minutes, respectively. Table IV shows the difference between the travel time for an original route and the travel time for an alternative route. The meaning of the figures is similar to those presented in Table III. We can interpret this result as follows:

- There is little change in the travel time for most of the passengers. In addition, the travel time is reduced for 63.0% of the passengers. One reason for this is that some of the passengers used trains as alternatives to the buses.
- On the other hand, the travel time for 20.5% of passengers increased by more than 10 minutes due to the waiting time required for additional transfers. In particular, the travel time for 3.6% of passengers increased by more than 30 minutes. This is because a number of passengers were forced to take detour routes due to the discontinuation of their bus stops. If the bus service operator discontinues the bus service, the local governments should introduce an alternative transportation option for those areas that are affected. In addition, we found that waiting time affects travel time. Thus, it is important to revise the timetables to reduce waiting time by using the method proposed by Akahoshi *et al.* [11] *etc.*

TABLE IV: Changes in the travel time

add. time [min]		# passengers	
LB	UB	w/ BS change	w/o BS change
−30	−20	255	0
−20	−10	311	7,068
−10	0	6,242	15,425
0	10	5,182	3,269
10	20	1,258	2,928
20	30	1,899	1,965
30	40	1,067	177
40	50	103	0
50	60	295	0
60	∞	0	59

V. Conclusion

We proposed an optimization model to estimate OD matrices from IO data. The validity of the proposed model was established by numerical results from experiments using real-world data. In addition, we evaluated the efficiency of a real-world bus service. From this empirical study, we quantified the effect of the discontinuation of the bus service. In addition, we identified the inconvenience that discontinuation would have on passengers in several areas, including bus fare, number of transfers, waiting time, and travel time.

In our empirical study, we did not quantify the effect of the change of using a bus stop. In the area (Saitama City, Japan) examined in this study, the percentage of elderly passengers is high. Thus, the effect of closing a bus stop is significant. The quantification of such an effect should be addressed in future work.

Acknowledgment

The authors would like to thank the Saitama City government for providing the passenger boarding-alighting data used in this study. The first author is supported by Grant-in-Aid for Research Activity Start-up 26882039.

References

[1] T. Abrahamsson, “Estimation of origin-destination matrices using traffic counts—a literature survey,” IIASA Interim Report IR-98-021, 1998.

[2] A. Cui, “Bus passenger origin-destination matrix estimation using automated data collection systems,” Master’s thesis, Massachusetts Institute of Technology, Department of Civil and Environmental Engineering, 2006.

[3] J. Zhao, A. Rahbee, and N. H. M. Wilson, “Estimating a rail passenger trip origin-destination matrix using automatic data collection systems,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 22, pp. 376–387, 2007.

[4] D. Li, Y. Lin, X. Zhao, H. Song, and N. Zou, "Estimating a transit passenger trip origin-destination matrix using automatic fare collection system," in *Database Systems for Advanced Applications*, J. Xu, G. Yu, S. Zhou, and R. Unland, Eds., vol. 21. Springer Berlin Heidelberg, 2011, pp. 502–513.

[5] N. L. Nihan and G. A. Davis, "Recursive estimation of origin-destination matrices from input-output counts," *Transportation Research Part B: Methodological*, vol. 21, pp. 149–163, 1987.

[6] Gurobi Optimization, Inc., *Gurobi Optimizer Reference Manual*, 2016.

[7] IBM, *IBM ILOG CPLEX Optimization Studio*, 2016, http://www-01.ibm.com/software/commerce/optimization/cplex-optimizer/.

[8] FICO, *Xpress Optimization Suite*, 2016, http://www.fico.com/en/Products/DMTools/Pages/FICO-Xpress-Optimization-Suite.aspx.

[9] T. Achterberg, "SCIP: solving constraint integer programs," *Mathematical Programming Computation*, vol. 1, pp. 1–41, 2009.

[10] Transport for London, "Open data users," https://tfl.gov.uk/info-for/open-data-users/.

[11] K. Akahoshi, M. Takamatsu, A. Taguchi, N. Ishii, and T. Kosaka, "A study on method for the regional evaluation of low-frequency public transport accessibility: An analysis of relationship between public transport network and urban structure using time-space network," *Journal of the City Planning Institute of Japan*, vol. 47, pp. 847–852, 2012, in Japanese.