Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2018) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

World Conference on Transport Research - WCTR 2019 Mumbai 26-31 May 2019

Passengers’ OD flow estimation model for the flat fare bus system

Hiroshi Shimamoto[a)] *, Issei Hirai [a)]

[a]Department of Civil and Environmental Engineering, University of Miyazaki, 1-1, Gakuen Kibanadai Nishi, Miyazaki, 889-2192, Japan

Abstract

This paper proposes a methodology to estimate passengers’ OD flows in a flat fare bus system. The proposed methodology is based on two-stage approach. The first stage estimates the leg OD flows, the OD flows within a line without considering transferring behavior, using the number of boarding and alighting passengers and prior information of the leg OD flows. The second stage estimates the journey OD flows, the OD flows between true origins and true destinations considering the transferring behavior, using the leg OD flows estimated at the first stage and the public transportation network information. The estimation accuracy of the first stage is investigated using passengers’ OD flows in a certain line collected by on-board survey. As a result, we confirm that the estimation accuracy of the first stage depends on the observation error of the prior information of the leg OD flows. Then, the estimation accuracy of whole of the proposed methodology is investigated using assumed journey OD flows data in a hypothetical bus network. As a result, we confirmed that the estimation accuracy of whole of the model is as equally good as that of the traditional estimation model of automobile OD flows.

Keywords: Transit OD estimation, Flat fare area, Two stage approach, Entropy model

1. Introduction

It is required for public transportation operators to take effective measures to improve the level of service of public transportation, such as increasing service frequency, expanding vehicles’ capacity, and so on, for the sake of promoting the use of public transportation. Passengers’ OD flows are essential information for taking such effective measures. Therefore, many researchers so far proposed methodologies to estimate passengers’ OD flows.

One of the approaches to estimate passengers’ OD flows is based on inverse problem. For example, William et. al. [1] proposed a bi-level programming approach to estimate passengers’ OD flows whose upper problem minimizes the sum of error measurements in passenger counts and OD matrices, and whose lower problem is a frequency-based SUE

* Corresponding author. Tel.: +81-985-58-7331; fax: +81-98558-7344.
E-mail address: shimamoto@cc.miyazaki-u.ac.jp

transit assignment model. Wu et. al. [2] further considered the elastic line frequencies in the lower problem. Wong et. al. [3] proposed passengers' OD flows estimation model based on entropy maximization approach and Wong et. al. [4] further expand the model so as to estimate OD flows of multimodal public transportation network.

Although above methodologies are similar approach with automobiles' OD flows estimation, the smart card payment system has been widely adopted in the public transportation systems all over the world. The smart card data can be expected to make the use of estimating passengers' OD flows. However, in the public transportation system under the flat fare system, passengers have to tap a smart card only once; when they board or alight a vehicle. Therefore, many researchers have been proposing methods to estimate passengers' OD flows using the smart card data. Barry et. al. [5] estimated passengers' OD flows using entry-only automatic fare collection data. When they inferred alighting location, they made two simple assumptions that most riders start their next trip at or near the destination of their previous trip and that most riders end their last trip of the day at or near the start of their first trip of the day. Gordon et. al. [7] inferred boarding time and location for individual bus passengers in London by combining smart card data and vehicle location data. They further estimated the all passengers' journey so as to satisfy the observed total volume. These methodologies require detailed individual smart card data, however, individual smart card data cannot always used for the transportation planning due to the privacy issues.

Based on these backgrounds, this paper aims to propose a model to estimate passengers' OD flows estimation model for the flat fare bus systems. The proposed methodology assumed to use smart card data which is aggregated in each bus run due to the privacy issue. As described later, the proposed model is based on the two-stage approach. The reminder of this paper is organized as following. Section 2 describes the proposed model. Section 3 evaluates the estimation accuracy of the first stage of the proposed model. Then, Section 4 evaluates the estimation accuracy of the whole of the proposed model. Finally, Section 5 summarizes the conclusions and proposed future works.

2. The model

2.1. Assumptions

The proposed model in this study is with in mind for applying for a flat fare bus system where passengers pay fare when they alight. Because bus lines in a bus network in general do not connect to all of bus stop pairs, some passengers have to transfer to arrive at their destinations. Figure 1 shows an example of passengers' movement, in which a passenger boards line I at bus stop A, transfers to line II at bus stop B, and alights line II at bus stop C. We define two types of OD pair for the consideration of such transferring behavior. One is the "leg OD pair", which is an OD pair within a line without considering transferring behavior. The other is the "journey OD pair", which is an OD pair between true origin bus stop and true destination bus stop with considering transferring behavior. In the example shown in Figure 1, the leg OD pair corresponds to between A and B-I, and between B-II and C. The journey OD pair corresponds to between A and C.

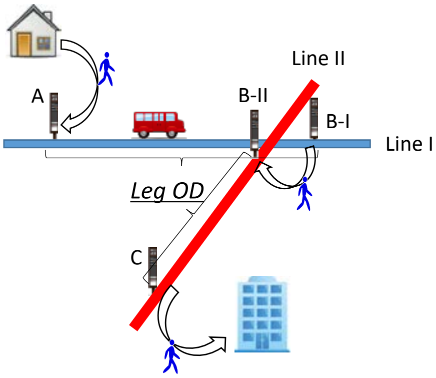

Fig. 1 Example of passengers' movement

- Following data are assumed to be available for the estimation; the number of alighting passengers at each bus stops from the smart data.
- The number of boarding passengers at some of bus stops by the manual count.
- The prior information of leg OD flows between bus stops in the same line.
- The bus network data including travel times between adjacent bus stops and the service frequency of each line.

Note that the penetration rate of the smart card is not necessarily assumed to be high, which means the observation accuracy of the number of alighting passengers may not be good. The number of boarding passengers is assumed to be obtained by manual count at some of bus stops as complementary information for the estimation.

2.2. Outline of the estimation model

We adopt a two-stage approach for estimating the passengers' OD flows as shown in Figure 2. In the first stage, the leg OD flows are estimated based on the observed number of alighting passengers at all of bus stops, the observed number of boarding passengers in some bus stops, and the prior information of log OD flows. In the second stage, the journey OD flows are estimated based on the leg OD flows estimated at the first stage and the bus network information.

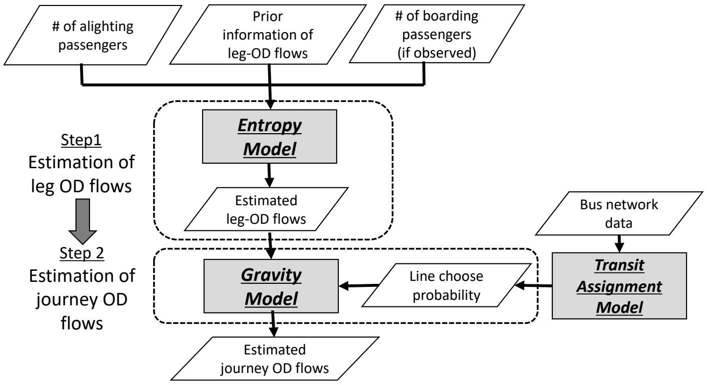
# of alighting passengers
Prior information of leg-OD flows
# of boarding passengers (if observed)
Step1
Estimation of leg OD flows
Entropy Model
Estimated leg-OD flows
Bus network data
Step 2
Estimation of journey OD flows
Gravity Model
Line choose probability
Transit Assignment Model
Estimated journey OD flows

Fig. 2 Outline of the proposed model

2.3. Leg OD estimation model

In the first stage, the leg OD is estimated based on the entropy model proposed by Sasaki [7], which can utilize the prior information of OD flows for the estimation. The model of the first stage can be formulated as following;

$$\min_{x_{ij}^{r_l(\tau)};i<j\leq N} \sum_{n=1}^{N}\sum_{m=1}^{n-1}\left(x_{mn}^{r_l(\tau)}\, ln\frac{x_{mn}^{r_l(\tau)}}{q_{mn}^{r_l(\tau)}} - x_{mn}^{r_l(\tau)}\right),\ \ l \in L,\ \forall r \in R_l,\ \tau \in T\square \tag{1}$$

such that

$$\sum_{i\leq n}\sum_{j\geq n+1} x_{ij}^{r_l(\tau)} \leq C_{r_l}\ \ n = 1,2,\dots,N_l - 1 \tag{2}$$

$$\sum_{i\leq n} x_{in}^{r_l(\tau)} = \xi Y_n^{r_l(\tau)}\ , n = 1,2,\dots,N_l \tag{3}$$

$$\sum_{n<j\leq N} x_{nj}^{r_l(\tau)} = X_n^{r_l(\tau)}\ , n \in B_l \tag{4}$$

where L is the set of lines, R_l is the set of bus runs in line $l \in L$, T is the set of time periods, N_l is the set of bus stops (which is labeled from the starting bus stop), B_l is the set of bus stops where the number of boarding passengers are counted, C_{r_l} is the capacity of bus $r \in R_l$, $l \in L$, $X_n^{r_l(\tau)}$ is the observed number of boarding passengers from bus $r \in R_l$, $l \in L$ at bus stop $n \in B_l$ at time period $\tau \in T$, $Y_n^{r_l(\tau)}$ is the number of alighting passengers from bus $r \in R_l$, $l \in L$ at bus stop n at time period $\tau \in T$, ξ is the expansion rate (which is to be estimated from the usage rate of the smart card historical data), $q_{mn}^{r_l(\tau)}$ is the prior passengers' demand between on bus $r \in R_l$, $l \in L$ at time period$\tau \in T$. $x_{mn}^{r_l(\tau)}$ is unknown variable in the model which represents for the passenger demand between on bus $r \in R_l$, $l \in L$ at time period$\tau \in T$.

Equation (2) represents for the capacity constraints condition that the number of passengers between bus stops n and $(\mathrm{n}+1)$ cannot exceed the capacity of the bus. Equation (3) represents that the number of alighting passengers at bus $r \in R_l$, $l \in L$ should coincide with the estimated number of alighting passengers from the smart card historical data. Equation (4) represents that the number of boarding passengers at bus $r \in R_l$, $l \in L$ should coincide with the observed number of boarding passengers.

2.4. Journey OD estimation model

In the second stage, the relationship between the leg OD and the journey OD is formulated under the condition that the line choice probabilities of the journey OD pairs are given. Suppose that the journey OD accords to following gravity model;

$$\hat{T}_{OD}^{\tau} = (NB_O^{\tau})^{\alpha}(NA_D^{\tau})^{\beta}(d_{OD})^{\gamma}(LOS_{OD}^{\tau})^{\delta}, , \forall OD \in \Omega \tag{5}$$

where, T_{OD}^{τ} is the journey OD flow between O and D at time period τ, NB_o^{τ} and NA_D^{τ} respectively represents for the number of boarding and alighting passengers at bus stop O and bus stop D at time period τ, d_{od} is the distance between O and D, LOS_{OD}^{τ} is the generalized cost by bus between O and D at time period τ,and Ω is the set of journey OD pairs. $\alpha, \beta, \gamma, \delta$ are parameters to be estimated by the model.

Furthermore, following relationship should be satisfied between the journey OD and leg OD;

$$\hat{y}_{mn}^{r_l(\tau)} = \sum_{OD \in \Omega} \hat{\mu}_{rs,l}^{OD}(\tau)\hat{T}_{OD}, \forall\, l \in L, mn \in \omega_l, r \in R_l, \tau \in T \tag{6}$$

where ω_l is the set of leg OD pairs, $\mu_{rs,l}^{OD}(\tau)$ is the probability that the journey OD pair OD choose line l between rs at time period τ, and $\hat{A}$ is the estimated value of A. On the other hand, the leg OD by each bus run $\hat{x}_{mn}^{r_l(\tau)}$ (which is estimated at the first stage) can be aggregated with respect to bus runs as following;

$$\tilde{y}_{mn}^{l(\tau)} = \sum_{r \in R_l(\tau)} \hat{x}_{mn}^{r_l(\tau)}, \forall mn \in \omega_l, l \in L \tag{7}$$

Now, let us suppose that the residual errors between the leg OD obtained from Eqs (6) and (7) follows the normal distribution with 0 mean. Then, the likelihood function, which correspond with the joint probability density of all of leg OD pairs, is given as following (Hazemoto et, al. [8]);

$$L_{\tau} = \prod_{l \in L} \prod_{mn \in \omega_l} \left[\frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{1}{2} \frac{\left\{ ln\left(\tilde{y}_{mn}^{l(\tau)}\right) - ln\left(\hat{y}_{mn}^{l(\tau)}\right) \right\}^2}{\sigma^2} \right) \right]^{\delta_{mn}^{l}} \rightarrow max \tag{8}$$

where δ_{mn}^{l} takes 1 if leg OD pair mn uses line $l \in L$.

The parameters in Eq (5) and the dispersion parameter in Eq (8) can be estimated so as to maximize the likelihood function shown in Eq (8) for each time interval.

Not that the proposed model is assumed to apply for a city where a complex bus network is formed and high frequency bus service is provided. Therefore, the generalized cost of journey OD pairs are calculated using the transit assignment model [9]. Furthermore, the probability that the journey OD pair OD choose line l between rs at time period τ, $\mu_{rs,l}^{OD}(\tau)$, can also be estimated from the line choice probability obtained from the transit assignment model. Hereafter, the estimation accuracy of the proposed methodology is investigated.

3. Estimation accuracy of the first stage

This chapter evaluates the estimation accuracy of the leg-OD, which is estimated in the first stage, using passengers' demand data based on on-board survey.

3.1. Investigation condition

For the investigation of the estimation accuracy of the first stage, we utilize passengers' OD flows in a certain line which is collected by on-board survey conducted in a certain city in Japan. Flat fare system, where passengers pay fare when they alight, is adopted in that city. The number of bus stops and the number of leg OD pairs are respectively 56 and 1,540 in a line utilized for investigation.

3.2. The effect of the observation errors

Firstly, in order to investigate the effect of observation error onto the estimation accuracy, we set 9 scenarios as shown in Table 1 by combining the error of prior passengers' demand, the error of the number of observed boarding passengers, and the error of the number of observed alighting passengers. For each scenario, 10 sets of the input data is created by generating 10 sets of random number. Figure 3 shows the comparison of the estimation accuracy of each scenario. Although the accuracy of the prior passengers' demand information affects to the estimation accuracy, the accuracy of the number of boarding and alighting passengers does not affect to the estimation accuracy so much.

Table 1. Observation errors for each scenario

Scenario	Observaion Error		
	Prior Info	Boarding	Alighting
1	0.1	0.00	0.1
2	0.1	0.05	0.2
3	0.1	0.10	0.3
4	0.2	0.00	0.2
5	0.2	0.05	0.3
6	0.2	0.10	0.1
7	0.3	0.00	0.3
8	0.3	0.05	0.1
9	0.3	0.10	0.2

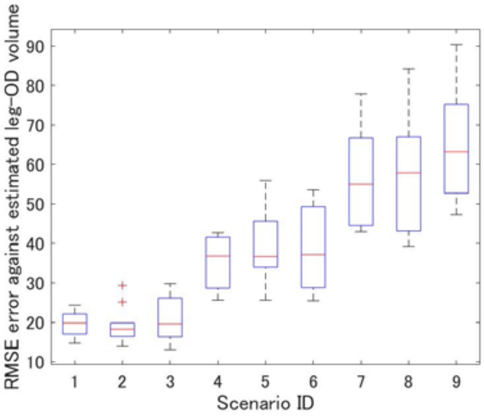

Fig. 3. Relationship between the observation errors and the estimation accuracy

3.3. The ratio of observing number of boarding passengers

Secondly, the relationship between the ratios of bus stops of observing the number of boarding passengers and the estimation accuracy is compared. Note that 10 sets of input data for each scenario is created by generating 10 sets of random numbers with the average error of both of the prior passengers' demand information and the number of alighting passengers as 10%. The bus stops which is assumed to observe the number of boarding passengers is selected in the descending order of the true number of boarding passengers. Figure 4 shows the estimation errors. We can confirm that the estimation accuracy with the lower observation ratios is high. However, the estimation result does not improve as the observation ratio increase. This is because the number of equality conditions becomes too larger to satisfy all of them.

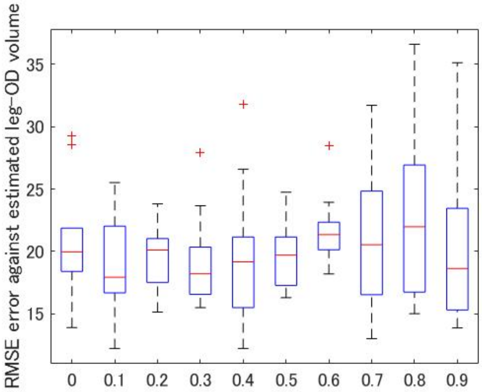

Fig. 4. Relationship between the observation bus stop rate and the estimation accuracy

4. Estimation accuracy of the whole of the model

This chapter investigates the estimation accuracy of whole of the methodology is investigated using artificial data.

4.1. Investigation condition

Because the true journey OD flows in a city mentioned in the previous chapter is not unknown, we apply the proposed methodology to the Sioux Falls network, which is often utilized as a benchmark of transportation network analysis. The Sioux Falls network with assumed bus lines is shown in Figure 5. We assumed the buses are operated based on frequency-based service; frequency of lines 1, 2, 5 and 6 is 1/5(1/minute) and that of line 3, 4, 7, 8 is 1/10 (1/minute). We utilized the OD flows published on-line [10] as the true journey OD flows. Because we confirmed in the previous chapter that the observation ratio of the number of boarding passengers does not affect to the estimation accuracy of the leg-OD flows, we assume hereafter that the number of boarding passengers are not observed at any bus stops. We generate the true leg OD flows (prior information) and the true number of passengers by the traffic assignment model [9] with infinity capacity. Against these true data, we assume that both of the average observation error of the prior information and that of the number of alighting passengers is 10%.

4.2. Estimation accuracy of the model

First of all, a set of input data is created by generating a set of random numbers with the average error of the prior information and the number of alighting number of passengers as 10%. Then, the leg-OD patters are estimated for each line. Table 2 shows the estimation accuracy of each line. Because the coefficient of determination of all of the lines is more than 0.97, we can estimate the leg-OD flows accurately in this case.

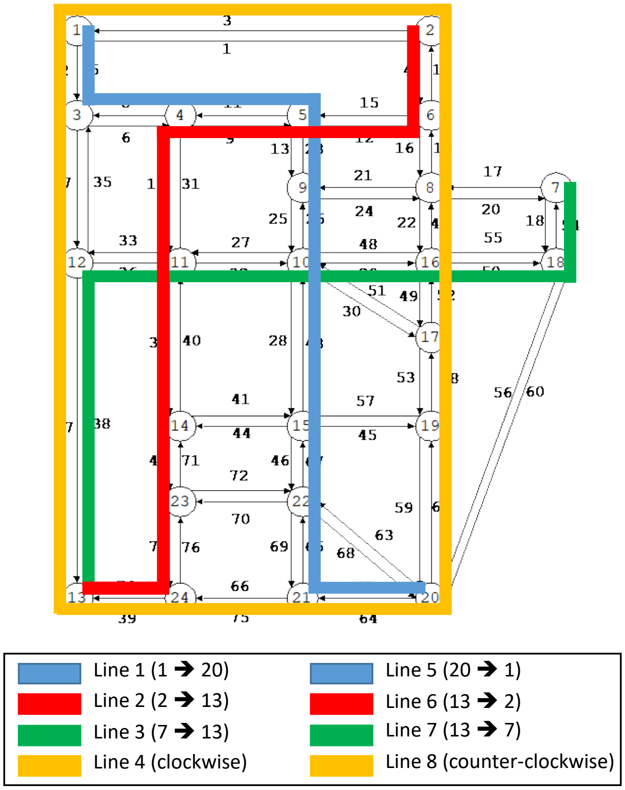

Fig. 5. Sioux Falls network with assume bus lines

Table 2. Estimation accuracy of the leg OD flows

Line	1	2	3	4	5	6	7	8
RMSE	1721	1094	231	1791	1911	1290	888	1210
CoD	0.987	0.981	1.000	0.983	0.985	0.978	0.995	0.993

CoD: Coefficient of Determination

Secondly, the journey OD flows are estimated using the leg-OD flows estimated previously. The distribution of the true OD flows may affect to the accuracy of the estimation result. Therefore, the traditional gravity model for automobile is also estimated for the comparison. Table 3 shows the estimated results of both models. The generalized cost is not included as an explanatory variable in the gravity model for automobile because it is expected to be strong correlation with the direct distance. (Note that the generalized cost of the public transportation network would not be strong correlation with the direct distance due to the existing of the service frequency.) All of the parameters in the model are statistically significant at the 0.01 level. The signs of the parameters of the number of boarding passengers and that of alighting passengers take positive. This implies that the journey OD flows between bus stops with larger number of boarding or alighting passengers tend to be larger. Also, the signs of parameters of the direct distance and the generalized cost is negative. This implies that the journey OD flows between longer distance and the larger generalized cost tend to be smaller. Therefore, the signs of all of the parameters are reasonable. Figure 6 compares the true journey OD flows and the estimated journey OD flows. Most of the plots are scattered along 45-degree line, and hence the coefficient of determination is very large. Based on above consideration, we can estimate the journey OD flows as well as the leg OD flows very accurately.

Table 3. Estimation result of the models

	Journey OD estimation model			Vehicle OD estimation model		
	Parameter	t-value	P-value	Parameter	t-value	P-value
# of boarding passegers	0.8500	38.396	0.000	0.8171	37.813	0.000
# of alighting passegers	0.8399	38.674	0.000	0.8201	37.978	0.000
direct distance	-0.5821	-29.426	0.000	-0.7665	-39.065	0.000
generalized cost	-0.4230	-4.907	0.000	-	-	-
ρ^2	0.1273	12.591	0.000	-	-	-

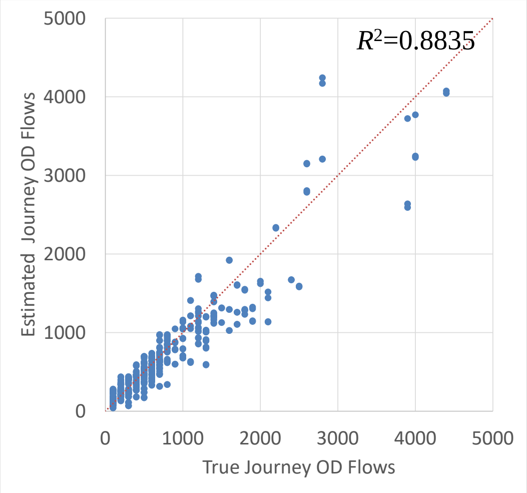

Fig. 6. Comparion between true and estimated journey OD flows

4.3. Effect of the average error onto the estimation accuracy

So far, we only investigate the estimation accuracy using one set of input data with the average error as 10%. This section investigates the effect of the average errors onto the estimation accuracy by assuming 4 scenarios with different average error of the priori information and that of the number of alighting passengers respectively as following; i) (10%, 10%), ii) (10%, 20%), iii) (20%, 10%) and iv) (20%, 20%). In order to take consideration of the effect of random numbers, 10 sets of input data for the estimation of the leg-OD flows of each line are created for each scenario; i.e. the leg-OD flows of each line are estimated 10 times for each scenario. Then, the journey OD flows are estimated using the combination of the estimated results of leg-OD flows with the highest accuracy cases (hereafter called “best case”). We further estimate the journey OD flows using the combination of the estimated results of leg-OD flows with the lowest accuracy cases (hereafter called “worst case”). Table 4 shows the coefficient of determination of both of the best case and worst case for each scenario. Because the coefficient of determinations in both of the best case and the worst case take close value for each scenario, the random number does not affect to the estimation accuracy in this case. Also, the coefficient of determination keeps high value even if the average observation errors become higher. This may be because the size of the network used in this chapter is smaller compared with the real network size.

Table 4. Estimation accuracy of the proposed model

Observaion Error		Coefficint of Determination	
Prior Info	Alighting	Best Case	Worst Case
0.1	0.1	0.8837	0.8804
0.1	0.2	0.8814	0.8841
0.2	0.1	0.8869	0.8873
0.2	0.2	0.8879	0.8877

5. Conclusion

This paper proposed the methodology for estimating passengers' OD flows in a flat fare bus service using prior OD flows information, the number of boarding and alighting passengers' and the bus network information. The proposed methodology is based on two-stage approach, where the leg OD flows are estimated in the first stage and then the journey OD flows are estimated in the second stage. As the results of confirming the estimation accuracy of the proposed model, followings are confirmed;

- The estimation accuracy of the first stage depends on the observation error of the prior information of the leg OD flows

- The estimation accuracy of whole of the model is as equally good as that of the traditional estimation model of automobile OD flows.

However, due to the data limitation, we evaluated the accuracy of whole of the model using assumed journey OD flows data in a hypothetical bus network, whose size is smaller than existing bus networks. Therefore, it is required to evaluate the estimation accuracy of the proposed model in a real size network.

References

[1] W. H. K. Lam, Z. X. Wu, and K. S. Chan., "Estimation of transit origin-destination matrices from passenger counts using a frequency-based approach", Journal of Mathematical Modelling and Algorithm, 2003, pp. 329-348.

[2] Z. X. Wu., and W. H. K. Lam, "Transit passenger origin-destination estimation in congested transit networks with elastic line frequencies", Annuals of Operation Research 144, 2006, pp. 363-378.

[3] S. C. Wong, and C. O. Tong., "Estimation of time-dependent origin destination matrices for transit networks", Transportation Research Part B, 32, 1998, pp. 35-48.

[4] K. I. Wong, S. C. Wong, C. O. Lam, W. H. K. Lam, H. K. Lo., H. Yang., and H. P. Lo, "Estimation of origin-destination matrices for a multimodal public transit network", Journal of Advanced Transportation, 39(2), 2005, pp. 139-168

[5] J. Barry, R. Freimer, and H. Slavin, "Use of entry-only automatic fare collection data to estimate linked transit trips in New York City", Transportation Research Record 2112, 2009, pp. 53-61.

[6] J. Gordon, H. Koutsopoulos, N. Wilson, and J. Attanucci, "Automated inference of linked transit journeys in London using fare-transaction and vehicle location data", Transportation Research Record, 2343, 2013, pp. 17-24

[7] T. Sasaki, "Probability methods to estimate trip distributions", Proceedings of 6th International Symposium on the Theory of Traffic Flow, 1968

[8] J. Hazemoto, M. Tsukai, and M. Okumura. "Reproduction of net passenger trips from gross traffic data", Infrastructure Planning Review 21, 2004, pp. 83-90 (in Japanese)

[9] F. Kurauchi, M. G. H. Bell, and J. D. Schmöcker, "Capacity-constraint transit assignment with common lines", Journal of Mathematical Modelling and Algorithms, 2(4), 2003, pp. 309-327.

[10] Transportation Networks for Research Core Team, Transportation Networks for Research. https://github.com/bstabler/TransportationNetworks, accessed July, 20, 2018