Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 52 (2015) 707 – 712
Procedia Computer Science

The 6th International Conference on Ambient Systems, Networks and Technologies (ANT 2015)

Validation of Activity-Based Travel Demand Model using Smart-Card Data in Seoul, South Korea

Sungjin Cho[a]*, Won Do Lee[b], Jeong-Hwan Hwang[c], Bruno Kochan[a], Luk Knapen[a], Tom Bellemans[a], Keechoo Choi[b] and Chang-Hyun Joh[c]

a Transportation Research Institute (IMOB), Hasselt University, Wetenschapspark 5 bus 6, 3590 Diepenbeek, Belgium
b Department of Transportation Engineering, Ajou University, Suwon, South Korea.
c Department of Geography, Kyung Hee University, Seoul, South Korea

Abstract

This study aims to validate an activity-based travel demand model, FEATHERS, using smart-card data which is collected in Seoul, South Korea, and to discuss some limits and challenges in the prediction of public traffic demands. To achieve the goal, global/local trip pattern indices and a hot-spot analysis were applied for the validation test as a comparison method in this study. Using those methods, the public traffic demands predicted by the simulation in the study area were evaluated comparing with ones in the smart-card data. As a result, FEATHERS Seoul shows the enough performance in predicting the global pattern of the public traffic demands, but a low performance in a local pattern, particularly in some areas with a mixture land-use type and/or a frequent public transit. This is because the current model does not handle such a complicate type of land-use and also a multi-modal trip in the simulation process. In conclusion, this study addressed the limits of the current model through the validation test using smart-card data and suggested some solution to the improvement in the specific models As a future work, we will apply smart-card data for the validation of the models operated in FS, such as a location choice model and a trip mode choice model.

Peer-review under responsibility of the Conference Program Chairs

Keywords: Activity-based model; FEATHERS; Smart-card data; Household trip survey; Validation; Hot-spot analsys; Spatial analysis

* Corresponding author. Tel.: +32-11-269-137; fax: +32-11-269-199.
E-mail address: Sungjin.Cho@UHasselt.be

1877-0509 © 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Peer-review under responsibility of the Conference Program Chairs
doi:10.1016/j.procs.2015.05.080

1. Introduction

To date there are few cases of using an activity-based trip demands model (ABM) in practice though the ABM supports a lot of advantage in the assessment of a transport policy and planning due to some practical limits, for example difficulties in data collection, high complexity of the model and the absence of a detailed assessment of the model. In particular, the absence of the validation seems to be the biggest hindrance in adopting the model to the field[1]. Since last decade numerous studies have been conducted with big data collected from a social media, GPS, a smart phone, smart-card and so on. Among them, the smart-card data provide invaluable information on public transport passenger. The smart-card is typically used as an alternative payment to a cash or credit card with its convenience and security over the existing payments mostly in the US. and European countries. Since the smart-card data contain information about a public transport, such as travel mode, time, location and cost, as well as a card ID (its details will be explained later), it allows us to see the detailed trip patterns of public passengers in a study area. While a number of existing research thus proposed the applicability of the smart-card data in evaluating the prediction the public trip demands using ABMs[2,3,4], there has been no empirical study using the smart-card in the validation of the ABMs up to now. Thus, the purpose of this study is to assess the existing ABM, which is a FEATHERS Seoul (hereafter, FS) developed for traffic demands forecast in Seoul metropolitan area, South Korea, using the smart-card data in the study area. Moreover, this study tries to address the challenges of the public trip demands prediction in the current FS.

2. Methodologies

At first, the required information for the validation test was extracted from the smart-card data. The information extracted consists of a card ID, trip mode (bus, train and subway), location (a stop ID), time (second) and cost. Those trip data were then aggregated into a subzone level, which is the finest spatial resolution (corresponding to the lowest administrative unit, Dong, in Seoul) of FS, in order to derive the public trip frequency in the study area. The public demands in FS were derived from the simulation output. Fig. 1 illustrates the whole process of this study.

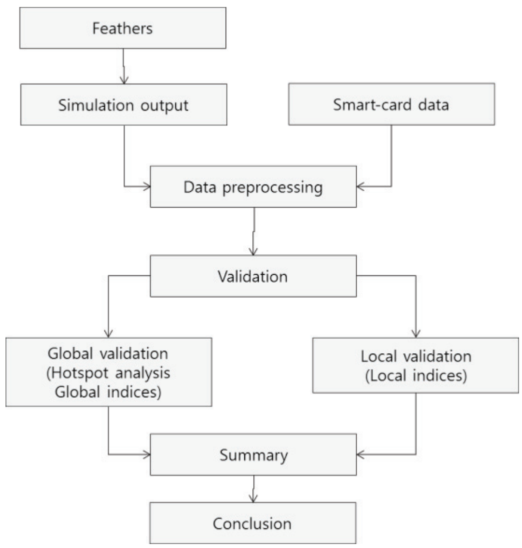

Fig. 1. Research flowchart.

Note that the population target is limited to adults (older than 19 years old) and a public mode is considered as one single mode due to the fact that FS only targets the adult group and the single public mode for the simulation. Thus, the overall public demands for adults over all types of public modes are only counted in each spatial (a subzone level) and temporal (a peak and non-peak time period) dimension. Followed by, the aggregate public demands computed were applied for a global and local test with a purpose of the validation of the simulation result in FS.

In detail, for the global validation test, global trip indices for public modes were calculated in each subzone during a peak (7-9 A.M.) and non-peak (10 A.M. – 2 P.M.) period using a descriptive statistics, for example daily average trip frequency, trip distance and time for public modes. The calculated global indices were compared between the prediction from FS and the smart-card data. In addition to the basic statistics, a hot-spot analysis was used for the global test to identify the spatial similarity of trip patterns in a public mode. A Getis-Ord index (or known as G*), proposed by Getis and Ord (1992) was applied for the hot-spot analysis in this study. The G* can be defined by following as;

$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j}x_j - \bar{X}\sum_{j=1}^{n} w_{i,j}}{S\sqrt{\frac{[n\sum_{j=1}^{n} w_{i,j}^2 - \left(\sum_{j=1}^{n} w_{i,j}\right)^2]}{n-1}}} \tag{1}$$

Where, x_j: the value of the *j* object,
$w_{i,j}$: the spatial weight between the objects *i* and *j,*
n, $\bar{X}$, S : total number, mean and standard deviation of the object.

Thus, the bigger G* is, the more clustered the objects are spatially[5]. For the local validation, we figured out the local trends in public modes, predicted by the model. For the test, we first chosen three Zones (Gu, one upper level of the administrative unit of Dong in Korea), where are representative to three typical types of land-use; residential, commercial and business area considering the impact of land-use constraint in the G* model. Next, we computed local indices for each of those areas using the same measurements as the global test and then compared between the indices from FS and the smart-card data. Table 1 describes the chosen three areas in this study.

Table 1. Description of the regions for comparison study.

Type	Region name	*Population (households/persons)	*Land-use ratio
Residential	*Seodaemun-gu*	135,104/ 324,733	53.2% (for residence)
Commercial	*Jung-gu*	61,546/ 140,807	28.65% (for shopping)
Business	*Gangseo-gu*	223,708/ 573,794	15.81% (for business)

* According to Census 2012 in Korea
* The highest ratio in each of three land-use types

3. Application: FEATHERS

FEATHERS is an activity-based travel demands model system developed for the transport policy assessment in Flanders, Belgium. Thus, it allows us to predict individual activity-trip diary through simulation in the study area, and to produce traffic demands in a network[6]. Like other ABMs, Feather requires several input data; activity-trip diary for training the model, population data for the model prediction, and environment data for describing land-use and network system in the study area. FS has been typically evaluated by using validation set achieved from the same resource as the training data or by comparing the global indices of the FS simulation output with the survey data, named a Household trip survey in Korea, which are used for the training and validation set. However, there are some problems with using the survey data for the validation test. First, the survey might introduce a sampling bias or data inconsistency due to respondents' mistake or misunderstanding. Typically, since the trip survey often possesses such problem, one wonders the applicability of the survey data in the model validation.

4. Study area and data

In this study, we considered Seoul, where is the capital city and the biggest metropolis in South Korea with almost 10 million population, as the study area for the validation test. This study used the public trip information derived from the FS simulation output and the smart-card data, which were collected in Seoul in 2012. The smart-card data have a high confidence compared to the survey data, because it is automatically collected by a smart-card reader installed in a public mode[7]. Since there are some differences between FS and the smart-card data in a spatial and temporal unit, it should be identical at the same level, which are subzone (Dong) and hour, prior to the validation process. Moreover, the population target and the public mode in the smart-card data are mismatched with ones in FS, so that we re-defined or aggregated the target and the single public mode in this study. Table 2 and 3 depict the simulation results from FS and the smart-card data, respectively.

Table 2. FEATHERS output.

Activity/Trip \ Personal		Household ID	Person ID	Other household/ person info. (i.e. composition, age, gender and so on)
Activity	Type	·	·	·
	Location/ Time		·	·
Trip	Mode	·	·	·
	Origin/ Destination	·	·	·

Table 3. Smart-card data list.

Field	Type	Description
PCARD_NO	Character	Card ID
TRAF_FREQ	Number	Frequency of transfer
TRANSP_METHOD_CD	Character	Transport mode
BUS_ROUTE_ID	Character	Bus route ID
PASGR_NUM	Number	Number of passengers
RIDE_AMT	Number	Fare
RIDE_STA_ID	Character	Station ID
USE_DIST	Number	Trip distance
RIDE_DTIME	Character	Departure time
ALIGHT_DTIME	Character	Arrival time

5. Experiment

For the validation test, we first measured the global indices and executed a hot-spot analysis on the public trip patterns of both the FS simulation and the smart-card data. The global indices and the result of a hot-spot analysis are then compared between two resources to evaluate how well FS captures the global pattern of the public passengers' trip in the study area. Table 4 represents the global indices calculated in the FS output and the smart-card data.

Table 4. Global indices of a public trip pattern.

Index	FS	Smart-card
Average trip frequency	2.20	2.52
Average trip duration (minute)	60.54	58.34
Average trip distance (km)	29.08	22.01

As you can see in table 4, the overall trends of public trips predicted by FS shows a similar pattern with that in the smart-card data. When you closely look at the results. An average trip duration and distance are overestimated while an average trip frequency is underestimated by FS, compared to ones in the smart-card data. It might be resulted from the fact that FS considers a householder and its partner as a study target, but on the other hand, the Smart-card data are collected from each of household members. Although the adolescents were excluded from the test, other members, like grandparents who typically account for a travel behavior with relatively high frequency of using public mode, and a short-distance and -term travel, are still considered in the experiment. Fig. 2 shows the hot-spot of public trips in the study area.

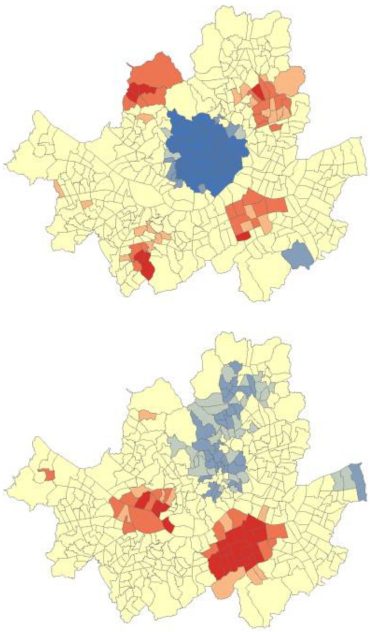

Fig. 2. Hot-spot analysis (Upper left: FS-peak, upper right: smart-card-peak, bottom left: FS-non-peak, bottom right: smart-card-non-peak).

In Fig. 2, the red spot indicates a high-clustered public traffic volumes in that area, but the blue spot a low-clustered volumes in public trip. Thus, it allows us to figure out where a hot-spot and cold-spot of public transports occur in the study area. As can be seen, the hot-spots in FS are a bit different with ones in the smart-card data in both peak and non-peak periods in the study area. This could be resulted from that FS shows a low performance of the prediction of a local traffic trend. To clarify the findings, we are going to look into the validation in a finer spatial scale. Furthermore, we also assessed the local pattern of pubic passengers' trip using the local indices in the study area. For the local validation, we selected three zones (Gu level) considering three types of land-use; residential, commercial and business zone. The local indices were calculated for each of three zones in the FS output and the smart-card data, and checked the similarity of each index in a zone. Table 5 describes the local indices in the three zones.

Table 5. Local indices of a pubic trip pattern.

Area	Index	FS	Smart-card
Residential area (Seodaemun-Gu)	Average trip frequency	-	-
	Average trip duration (min)	29.65	13.56
	Average trip distance (km)	15.34	3.65
Commercial area (Jung-Gu)	Average trip frequency	-	-

	Average trip duration (min)	26.57	17.29
	Average trip distance (km)	15.44	5.11
Business area (Gangseo-Gu)	Average trip frequency	-	-
	Average trip duration (min)	32.36	11.42
	Average trip distance (km)	15.74	3.29

In Table 5, the trip patterns of public passengers are all overestimated in FS, compared to the public trips derived from the smart-card data. Considering the similarity of the global patterns in between two data sets, the local patterns indicate that FS shows a poor performance in forecasting public traffic demands at a local level, even though a good performance at a global level. This might be an issue on the assessment of a regional policy and planning.

6. Conclusion

In this study, we applied the smart-card data for the validation of the public traffic demands predicted by FEATHERS in Seoul, South Korea, as a study area. For the validation, we conducted the global test using global indices and a hot-spot analysis for identifying the global patterns of public trips and a local test using the local indices for checking its local patterns in the study area. As a result of the experiment, FS shows the enough performance in predicting the global pattern of the public traffic demands, but a low performance in a local pattern, particularly in some areas with a mixture land-use type and/or a frequent public transit due to the fact that the current model does not consider both such a land-use type and a multi-modal trip in the simulation process. This study contributes to prove the applicability of the smart-card data in the validation of the ABMs, and the findings from the experiment will be helpful to clarify challenges and solution to the improvement of the existing model. In line with the findings, we will apply the smart-card data for the validation of the specific model components to improve the model system, for example location choice model and mode choice model in FEATHERS. This future work will support an empirical evidence of the improvement of the model components through a validation test using the smart-card data.

Acknowledgements

This work was supported by the National Research Foundation of Korea grant funded by the Korea government (MSIP) (NRF-2010-0028693). The smart-card data used in this study was supplied by the KTDB (Korea Transport DataBase) center.

References

1. Pendyla, Ram M. and Bhat, Chandra R. Validation and Assessment of Activity-Based Travel Demand Modeling Systems. In: Transportation research board conference proceedings. 2008.
2. Pelletier, M.-P., Trépanier, M., Morency, C. Smart card data use in public transit: A literature review. Transportation Research Part C: Emerging Technologies 2011. 19: 557–568.
3. P.C. Bouman, M. Lovric, T. Li, E. van der Hurk, L.G. Kroon, and P.H.M. Vervest. Recognizing demand patterns from smart card data for agent-based micro-simulation of public transport. In Proceedings of the Seventh Workshop on Agents In Traffic and Transportation, 2012.
4. F. Devillaine, M. Munizaga, and M. Tr´epanier. Detection of activities of public transport users by analyzing smart card data. Transportation Research Record: Journal of the Transportation Research Board 2012. 2276(1): 48–55.
5. Getis, A. and J.K. Ord. The Analysis of Spatial Association by Use of Distance Statistics. Geographical Analysis 1992. 24(3): 189-206.
6. Bellemans T, Janssens D, Wets G, Arentze T, Timmermans H. Implementation framework and development trajectory of FEATHERS activity-based simulation platform. Transportation Research Record: Journal of the Transportation Research Board 2010. 2175: 111–119.
7. Shin, S. I., Lee, C. J., and Cho, Y. C. A methodology to evaluate regional public transportation services using public transport use data. Transportation Technology and Policy 2008. 5(1): 111-127.Van der Geer J, Hanraads JAJ, Lupton RA. The art of writing a scientific article. *J Sci Commun* 2000;163:51-9.