Exploring Potential Use of Mobile Phone Data Resource to Analyze Inter-regional Travel Patterns in Japan

Canh Xuan Do(✉) and Makoto Tsukai

Hiroshima University, 1-4-1 Kagamiyama, Higashi-Hiroshima, Hiroshima 739-8527, Japan
canh.doxuan@gmail.com, mtukai@hiroshima-u.ac.jp

Abstract. In Japan, Inter-Regional Travel Survey gives rich information to researchers and transportation planners. The current survey data was conducted in 2010, and the newest survey data collected in 2015 will be available soon. This national survey is mainly based on the on-site questionnaire survey which requires an enormous budget and spends so much time to finalize and publish the data result. Recently, ubiquitous mobile computing and the big data give us new opportunities for exploring a new type of data resource besides the traditional survey data. This study clarifies the deviation of cell phone data at aggregated origin-destination level of inter-regional trip flows, compared with the traditional on-site passenger survey. Also, the mechanisms of inter-regional trip generation are explained through travel patterns by a classification tree analysis, one of the big data mining classification algorithms.

Keywords: Mobile phone data · Inter-regional Travel Survey data · Classification tree · Origin-destination trips · Travel patterns · Japan

1 Introduction

The fast growth of expanding cities along with the need for long-distance travel among regions has led a requirement to improve or find a new way to estimate inter-regional travel demand. To meet these challenges, there are various methods such as improving traditional four-step models and more recent activity-based models developed to utilize available computational resources. These models usually use methods of statistical sampling in local [1, 2] or national travel surveys [3–6] to analyze and infer trip characteristics between areas of a city or regions of a country.

Inter-regional travel surveys are typically administered by government or regional planning organizations and are integrated with public data such as national census with detail demographic characteristics of their residents, made available by city, state, and federal agencies. These surveys are designed to select representative samples in population carefully, so they are relatively expensive for surveying, and required much more time in the post-survey processing. As a result, the time between two consecutive surveys is four years or more in even the most developed cities (e.g., 1995 American Travel Survey (ATS) with a new version so-called National Household Travel

Y. Tan et al. (Eds.): DMBD 2017, LNCS 10387, pp. 314–325, 2017.
DOI: 10.1007/978-3-319-61845-6_32

Survey [6], Inter-regional Travel Survey in Japan [4], and so forth). The appearance of ubiquitous mobile phone computing has led to a noticeable increase in new big data resources capturing cell phone user activities in nearly real time and provided solutions to the conventional travel demand models. New data sources are collected by new providers such as large telecommunications companies with their own applications and network providers. Compared to the traditional survey, these big data sources provide large and long-time samples at low cost. Along with new opportunities, however, new mobile phone data (hereafter, called MOBI data) comes with new challenges of estimation, integration, and validation of existing models since they often lack or miss relevant contextual, social demographic information due to privacy reasons and have their interior noise and biases. Although these issues are significant obstacles to accommodate them to existing models, their use for regional transportation planning has the potential to decrease the period of survey conducting, increase survey coverage, and reduce survey costs. Therefore, it is essential to evaluate and propose a new method to integrate the new data sources into the traditional modeling.

To make these new data sources useful for inter-regional planning, we should clarify their biases and limitations. Moreover, we need to evaluate the appropriate level of implementing new data sources as an input data for inter-regional travel demand modeling. Many studies have been explored using these new massive, passively collected data such as individual survey tracking and stay extraction [7], origin-destination flows estimation and validation [8–10], traffic speed estimation [11, 12], and activity modeling [13]. Nevertheless, these studies generally present alternatives for only urban planning, while there is a small number of studies utilizing new data resources into modeling inter-regional travel demand [14]. Thus, here in this study, we clarify the characteristics of MOBI data by several comparisons of origin-destination (hereafter, called OD) pair travel flows and try to explore inter-regional travel patterns.

This paper is organized into following sections. The next section demonstrates the summary of data collection. Section 3 shows the methodology use in this study. The comparisons and the results of exploring travel patterns are presented in the following section. Section 5 illustrates the conclusions, limitations and future research.

2 Data Summary

2.1 Net Passenger Transportation Survey in Japan

In Japan, Inter-regional Travel Survey or Net Passenger Transportation Survey (from now on, called NPTS) has been carried out since 1990 in every five years by Ministry of Land, Transportation, Infrastructure and Tourism (MLIT) [4]. This survey collects passenger traffic at certain cross-sections and then aggregates the data into the inter-regional OD tables with its expansion factors. Individual characteristics of travel are captured such as departing point (origin) and arrival point (destination) of each transportation mode on the entire route if travelers transfer. It collects the questionnaire sheet from a respondent, one out of about a million samples, who uses inter-regional trains, express buses, airlines, cars or ships. The purpose of NPTS is to provide basic information for inter-regional transportation infrastructure planning. MLIT provides the

OD tables on MLIT website [15]. After the survey in 2000, trip information at the individual level with the corresponding expansion factor became available. Inter-regional net passenger traffic data used in this study was extracted from 2010's NPTS. The OD table records passenger trips between 207 areas of residential areas (origins) and destinations. The database of regional resources and demographics was compiled from 2011 Japan National Census.

NPTS provides two types of OD tables with a daily OD and an annual OD trips table. For comparison with MOBI data, in this study, we use the daily OD trips table. NPTS data has total 2,323,497 observations with 9,153,592 inter-regional daily trips.

2.2 Mobile Phone Data

The new MOBI data source is provided by NTT DOCOMO, the largest cell phone service provider in Japan. This company had more than 70 million mobile phone subscriptions as of 2016 [16] while the population of Japan was over 127 million as of October 1, 2015, based on the national census in Japan [17]. This company developed a new kind of small area statistics named Mobile Spatial Statistic (MSS), which is used to make estimations of the population of a small area by using the operations records from a mobile terminal network. The MSS's coverage of age is from 15 to 79, which is the generation of active mobile phone users in Japan. It also accounts for around 80% of the total population (2010 Population Census). Another reason why MSS selects this age range is that cell phone penetration rates are dominant in these ages, resulting in enough sample size for MSS to provide accurate estimates.

This study used the MOBI data collected in two days, one in holidays and one in weekdays in October 2015. There are 255,232 observations to be collected, which is equivalent to near 473 million trips over two days. The total number of trips covers intra-regional trips and inter-regional trips in 207 zones.

3 Methodology

The following two subsections review the algorithm for transforming MOBI data into OD matrices, describing some indices in MOBI trip characteristics used to compare with NPTS data regarding the number of inter-regional trips. The last subsection describes a classification method to find inter-regional travel pattern of trip generation.

3.1 OD Inter-regional Travel Flows Matrices

Current methods, which have been employed to estimate the trips between two places, fall into two categories: (1) four-step travel demand estimation methods; and (2) activities based approaches. The MOBI data records the successive activity trajectory through the mobile phone base stations. It provides an opportunity to transform the raw data with billions of points into an OD matrix of flows. Though a mobile phone user should have traveled between two different points at different times, we do not know the precise departure time of their trip. Thus, to extract meaning locations, termed as

stays, we assume that origin points are collected at mid-night time when mobile phone users would stay in their homes or hotels, and destination points are collected at noon of the day after they traveled. In this way, a full OD matrix is made by summing the trip volume computed for all users between all pairs of ODs.

3.2 Indices in MOBI Trip Characteristics

For comparisons between two kinds of data sources, we put the NPTS data as a standard reference. Therefore, from the standard reference, we clarify the deviation of MOBI data at aggregated OD level in the number of trips in the three following cases: (1) traveling between OD pairs; (2) generating from origin areas; and (3) distributing to destination zones. In this study, we use three indices, including *Pearson's correlation coefficient*, *Spearman's rank correlation coefficient*, and *deviation index* to measure these differences.

First, we employ the *Pearson's correlation coefficient*, r, which provides an indication of how closely a set of MOBI data values and a set of NPTS data values agree in relative terms. It is noted that a perfect correlation ($r = 1$) does not imply perfect reliability of MOBI data values – all MOBI data values may be biased in a consistent direction.

Second, *Spearman's rank correlation coefficient*, ρ, provides an indication of similarity between the ranks of two sets of values. It ranges from −1 to +1. The use of ranks means that, if the order of MOBI data is correct, the index will be high.

Finally, we introduce *deviation index* as an index for quantitatively evaluating the size of different values of two datasets. This index is used to indicate how much they deviate from the ideal state (i.e., when both are equal) (see Fig. 1).

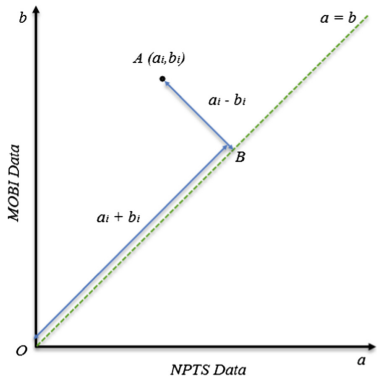

Fig. 1. Intuitive illustration of deviation index

Assume that for each pair of comparative values, i, the value in NPTS dataset is a_i, the value in MOBI dataset is b_i, and the average of the two values is $\mu_i = (a_i + b_i)/2$.

The deviation index, δ_i, for each pair of comparative values is defined as follows

$$\delta_i = \frac{b_i - \mu_i}{\mu_i} \tag{1}$$

Substituting μ_i into Eq. 1 gives us the new equation for deviation index as follow

$$\delta_i = \frac{b_i - a_i}{a_i + b_i} \tag{2}$$

Figure 1 describes an intuitive understanding of the concept, with deviation index, δ_i, represented by the ratio between the distance from the point $A(a_i, b_i)$ to its orthogonal projection point B on the line $a_i = b_i$ to the distance from the origin point O to the point B. As seen from the Eq. 2, the deviation index is normalized from −1 to 1. It can be inferred that deviation index is nearer to zero indicating the smaller difference between two values. Values are near −1 or 1, which means larger differences. Positive deviation indexes suggest that the MOBI data values are greater than the NPTS data values and vice versa. In case of $a_i = b_i = 0$, we also set the deviation index $\delta_i = 0$.

3.3 Decision Tree Analysis – A Classification Method

In the field of data mining, there are many classification methods such as decision tree, rule-based classifiers, and so forth. In this study, decision tree method is selected among other classification methods because of its outstanding advantages over other methods. First, the decision tree is a nonparametric approach which does not require any prior assumptions regarding the type of probability distributions satisfied by the class and other attributes. Second, decision trees are relatively easy to interpret, especially smaller-sized trees. Moreover, the accuracies of the decision trees are comparable to other classification techniques for many simple datasets.

Decision tree method is a technique for classifying data patterns into tree structures and consists of predicting a certain outcome based on a given input. In regression models, the statistical relationship between explanatory and outcome is assumed before starting the analysis, but in this technique, it is not necessary. Ture, Tokatli and Kurt [18] also stated that the outstanding advantage of this method is to discover and give a better explanation of the relationships between the predictors that would make it possible to predict the outcome. Also, the decision tree model often inputs all predictor variables, and then the representation of the outcome and predictor relationship is made with more concise and perspicuous results. A tree structure consisting of three sub-elements: a root node, internal nodes, and terminal nodes - the main output of a decision tree. The development of a decision tree includes three stages: a tree growing, a tree pruning and an optimal tree selection.

Decision tree consists of three groups such as Classification and Regression Tree (e.g., its name is often as CRT, CART, or C&RT), Chi-Square Automatic Interaction Detector (CHAID), and Quick-Unbiased-Efficient Statistical Tree (QUEST) [18]. Of

these groups, the CHAID analysis has not been widely applied in travel and tourism research field [19–23]. Van Middelkoop, Borgers and Timmermans [21] used an Exhaustive CHAID to identify heuristic principles for transport mode choices. The results proved that the methodology could be applied to understand tourist behavior. Bargeman, Chang-Hyeon, Timmermans and Van der Waerden [19] investigated the relationships between holidays choice behaviors and socioeconomic variables to classify respondents into different clusters by using a combination of CHAID and log-linear analyses. Chen [23] concluded that CHAID would be a useful tool to put forward the subdivision methodology in travel and tourism research. Chen [20] made a CHAID analysis to classify each actionable group by demographic and trips characteristics.

CHAID analysis has a variety of advantages. First, it can produce non-binary trees, which is different with CRT and QUEST. Another advantage of CHAID is, at each node, the procedure of splitting and merging pairs of categories of the predictor variables considering their differences from the outcome and using Chi-square test to measure these differences. A modification of CHAID method, called Exhaustive CHAID, was originally proposed by Biggs, De Ville and Suen [24]. Compared to the earlier CHAID, the optimal split procedure of Exhaustive CHAID was improved by continuously testing all possible category subsets. Fundamentally, it is a decision tree based on the Chi-square test, which is built by repeatedly splitting a parent node into two or more child nodes.

4 Empirical Study

4.1 Trip Weight

In order to focus on long-distance trips, the intra-zonal trips within each prefecture and three metropolitan areas (e.g., Tokyo, Osaka, and Nagoya metropolitans) are excluded to eliminate. In this paper, the number of study zones is 194 out of 207 zones since isolated islands with no choice in mode are also excluded. Thus, both data utilized in this study has a total of 36,346 OD pairs among 194 areas. Moreover, both datasets include trips in weekday and holiday, therefore, to calculate the number of daily trips, we introduce a daily trip weight to keep the balance between holiday trips and weekday trips in the dataset. The trip weight, m, is calculated as follow

$$m = \frac{N_h}{N_w} \tag{3}$$

where N_h is the number of holidays in Japan (i.e., national holidays and non-working days) and N_w is the number of weekdays in the autumn (from September to November 2010 for NPTS data, and from September to November 2015 for MOBI data).

Since the coverage age of MOBI data ranges from 15 to 79, all observations in NPTS data which are out of this age range are removed. Thus, NPTS data has 2,202,466 observations with near 7.8 million inter-regional trips, and MOBI data has 91,742 observations with over 8.2 million inter-regional trips.

4.2 Comparison of MOBI Data with NPTS Data

As mentioned in Subsect. 3.2, to clarify overall characteristics in MOBI data for aggregated OD trips, we first computed *Pearson's correlation index* and *Spearman's rank correlation index* as in Table 1. Compared to two other cases, the r value of trips among OD pair zones is the lowest, which means there is different between two data. This finding would be clarified in the following part of this subsection. The values of ρ are near one indicating the small difference of the number of trips in MOBI data compared to that of NPTS data.

Table 1. Summary of Pearson's correlation and Spearman's rank correlation coefficient

Aggregated trips	r	ρ
(1) OD pair zones	0.602	0.918
(2) From origin zones	0.955	0.942
(3) To destination zones	0.961	0.947

To be more detail comparison, the deviation indexes are calculated to exam the differences in other aspects. In order to understand the difference of OD pair trips, the heat map is made in Fig. 2a. In this figure, the direction of left-to-right or bottom-to-top represented for the north-to-south direction of Japan. The blue points mean that the number of OD pair trips in MOBI data is smaller than that of NPTS data (i.e., deviation index is near or equal minus one). Also, the opposite is true for the red points (i.e., deviation index is near or equal one). These two types of points mean that there is big different between two datasets while white points indicate that there is no different or the difference is very small.

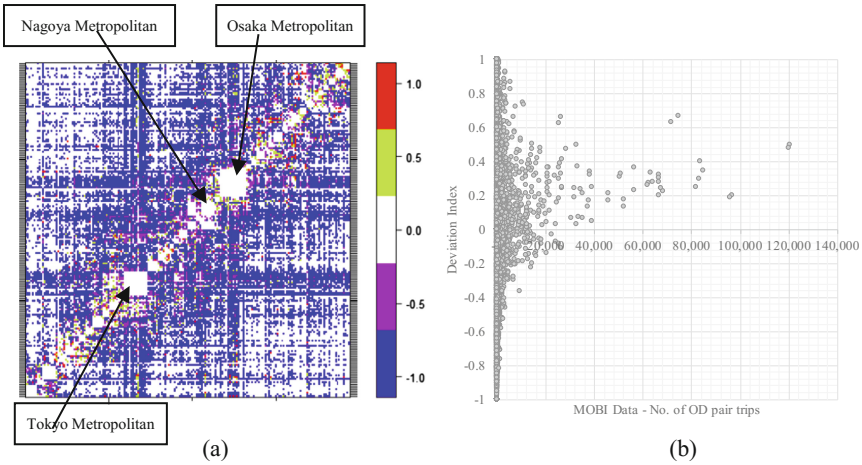

Fig. 2. Heat map (a) and density (b) of deviation indexes for the differences of OD pairs trips. (Color figure online)

As shown in Fig. 2a, most OD pairs are dominated by white and blue pixel points. Note that there are several large white areas in the diagonal because OD trips in three metropolitans are excluded in both NPTS data and MOBI data. Blue pixels concentrate around three metropolitans or in OD pairs where trips are generated from or distributed to the three metropolitans. Also, the red, green, and purple points are scattered along the diagonal. It can be inferred that MOBI data seems to over- or under-estimate at for trips around the densely inhabited areas and short-distance trips.

Figure 2b illustrates the density of deviation indexes and shows the relationship between the number of OD pair trips in MOBI data and deviation indexes. The graph indicates that the density is negatively skewed, which provides another evidence of over-estimation as seen in Fig. 2a. This figure, also, shows that scattering in deviation indexes decrease to near 0.2 as the number of OD pair trips increases. In other words, there is an average of 20% differences between two datasets. This reflects that MOBI data can capture more inter-regional trips in urbanized or the densely inhabited areas.

For evaluating the differences regarding trip generation and trip distribution, Fig. 3a and b illustrates the density of deviation indexes by origin and destination zones, respectively. This graph shows that there is a small difference between two datasets in term of trip distribution at origins and destinations. As seen these figures, most of the zones have smaller deviation indexes ranging from −0.2 to 0.2, indicating the small difference between two datasets in term of aggregated trips at origin or destination zones. However, some large cities such as Tokyo, Osaka, and Sapporo have large deviation indexes, indicating the less reliability of MOBI data in those areas.

4.3 Travel Patterns by a Decision Tree Analysis

To explore travel patterns, the Exhaustive CHAID was employed with 36,346 OD pair flows observations. The number of observations is consistent with Van Middelkoop, Borgers and Timmermans [21] because a CHAID-based algorithm requires a dataset with approximately 150 to 200 observations per one predictor variable. To put it simply, with 68 predictors, we need only 13,600 observations for CHAID analysis. As in Figs. 4 and 5, the objective variable is represented into two categories (i.e., “YES” = there is OD pair travel flow, and “NO” = there is no OD pair travel flow).

At a glance, travel pattern of trip generation is different. This may be because the number of zero OD travel flows in MOBI data is much higher. This is consistent with the previous result in Subsect. 4.2 (i.e., many blue points are seen in the OD pairs with deviation index that is near or equal negative one).

In general, most of the child nodes have low trip generation rate in MOBI data, while this is opposite in NPTS. More specifically, there is equivalent or similar at the strongest predictors between two trees in term of travel distance. The strongest predictor in the tree of NPTS data is the rail travel time variable, while rail travel cost variable is chosen in the tree of MOBI data. These two variables are closely related to travel distance. Therefore, based on these two variables, we defined three groups by the levels of travel distance (see detail in Figs. 4 and 5). However, in the second level of both trees, the difference appears. In the tree of NPTS data, the division is mostly depended on regions’ demographic characteristics such as the number of workers of the

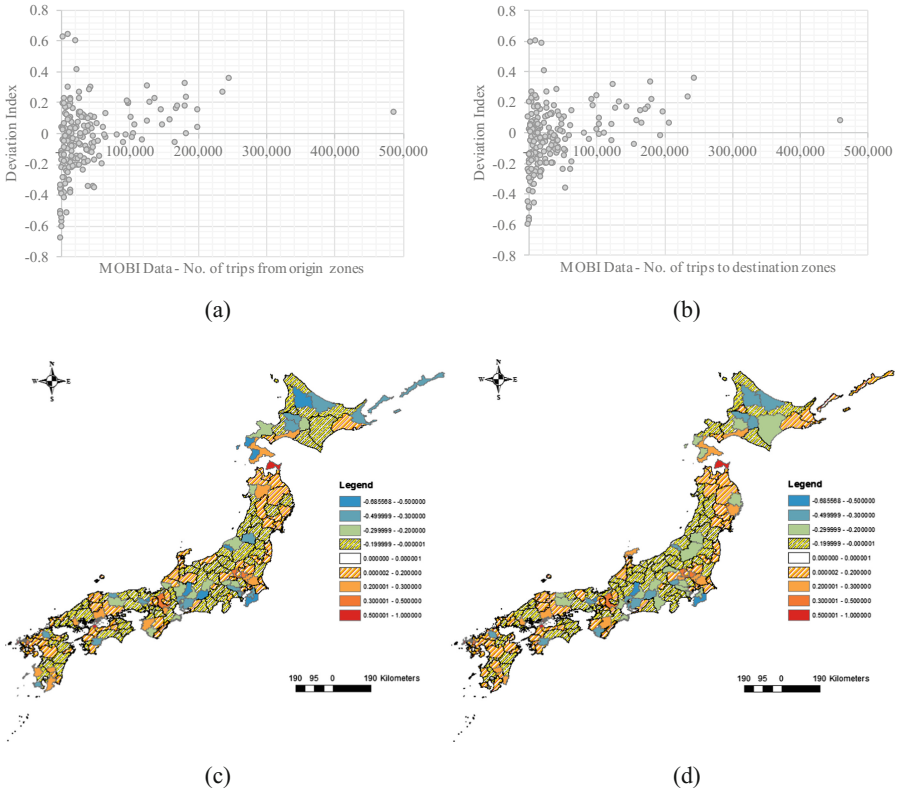

Fig. 3. Density and spatial distribution of deviation indexes by (a; c) origin zones and (b; d) destination zones.

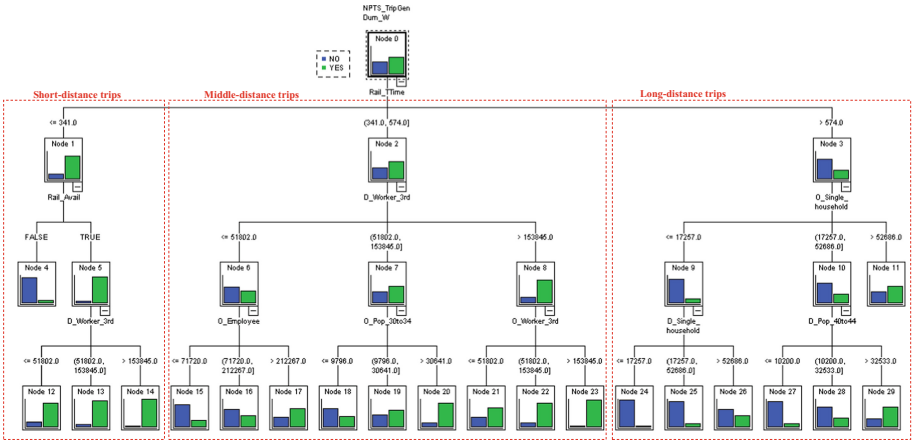

Fig. 4. Travel patterns of trip generation in NPTS data.

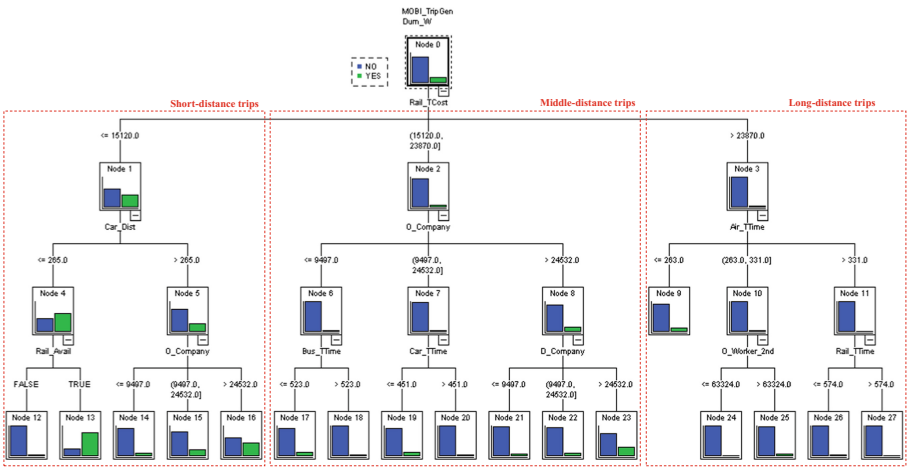

Fig. 5. Travel patterns of trip generation in MOBI data.

tertiary industry sector in destination zone, while that of the tree of MOBI data are variables of level of travel services (e.g., car distance, air travel time). This phenomenon is also seen in the lowest level of two trees. Regarding NPTS data, in the group of short-distance trips, trip generation is higher when there are railway connections between OD pairs and the number of tertiary sector workers in destinations is greater. Moreover, in the group of middle-distance trips, the more number of tertiary sector workers in destinations as well as in origins, and the more number of employees and people aged from 30 to 34 is, the more OD pair trips are made. In the group of long-distance trips, the number of single households in origins and destinations and the number of people aged from 40 to 44 are significant effects on trip generation rate. In the tree of MOBI data, trip generation rate is so low in both the group of middle-distance trips and long-distance trips. In the group of short-distance trips, the highest trip generation rate is seen in Node 13 where travel distance by care is lower than 265 km, and railway connection is available between two areas.

In conclusion, compared to trip generation rate at Node 0, the higher trip generation rate is seen in the three groups of OD travel distance in NPTS data (e.g., Node 12 to 14, Node 17, Node 20 to 23, Node 11, and Node 29), while it is only seen in groups of short-distance and middle-distance trips (e.g. Node 13,15,16, and Node 23). Furthermore, in the tree of MOBI data, trip generation rate drops approximately zero in the nodes for long-distance trips. It also happens in Node 12 in the short-distance trips group as well as in Node 18 and 20 in the middle-distance trips groups.

In order to evaluate the performance of two decision trees (or two prediction models), re-substitution and cross-validation test are used. In cross-validation test, our data is divided into ten portions stratified on dependent variable levels. The training error rates are estimated and shown in Table 2. The training error estimate of around 0.2 in case of NPTS data indicates that the category predicted by the NPTS tree model is wrong for 20% of the case or the risk of misclassifying trip generation is around

20%. In case of MOBI data, the risk value is around 10%. This result proves that the tree model of MOBI data is more productive than that of NPTS data.

Table 2. Summary of training error rate estimations

Method	NPTS data		MOBI data	
	Estimate	Std. Error	Estimate	Std. Error
Re-substitution	0.187	0.002	0.100	0.002
Cross-validation	0.223	0.002	0.107	0.002

Table 3 shows that the tree of MOBI data classifies approximately 96.8% of the "NO" cases correctly, which is better than that of NPTS data with 73.3%. This is not true in case of classification "YES". Overall percentage correct is 90% in the tree of MOBI data compared to 81.3% in the tree of NPTS data, indicating that the classification tree of MOBI data is better than the tree of NPTS data.

Table 3. Summary of decision tree coincidence matrix

Method	Tree result for NPTS data			Tree result for MOBI data		
	NO	YES	Correct (%)	NO	YES	Correct (%)
Actual NO	11,560	4,201	73.3%	30,523	1,024	96.8%
Actual YES	2,855	19,020	86.9%	2,731	3,358	55.1%
Overall (%)	38.3%	61.7%	81.3%	88.4%	11.6%	90.0%

5 Conclusions

In this paper, we clarify the characteristics of MOBI data by comparing with 2010 NPTS data as a reference. MOBI data is much similar with NPTS data in case of the number of aggregated trips from origin zones and to destination zones than in case of the number of OD pairs trips and travel pattern of trip generation.

However, the significant different between two data may be caused by the difference in the year of two surveys collected; MOBI data is observed in 2015 while NPTS data was collected in 2010. Thus, to clarify the difference more accurately, we should make another comparison with the 2015 NPTS data to come. Also, in the future, we will find new ways of accurate comparisons by controlling some side effects such as differences in socio-demographic characteristics of respondents and then finding another appropriate method to estimate origin-destination location more precisely.

References

1. Smith, M.E.: Design of small-sample home-interview travel surveys. Transp. Res. Record **701**, 29–35 (1979)
2. Daganzo, C.F.: Optimal sampling strategies for statistical models with discrete dependent variables. Transport. Sci. **14**, 324–345 (1980)

3. Stopher, P.R., Greaves, S.P.: Household travel surveys: where are we going? Transport. Res. A Pol. **41**, 367–381 (2007)
4. Ministry of Land, Infrastructure, Transport and Tourism (MLIT). www.mlit.go.jp/common/001005633.pdf
5. Bureau of Transportation Statistics - U.S. Department of Transportation. http://www.transtats.bts.gov/DatabaseInfo.asp?DB_ID=505&Link=0
6. Federal Highway Administration - U.S. Department of Transportation. https://www.nationalhouseholdtravelsurvey.com/
7. Asakura, Y., Hato, E.: Tracking survey for individual travel behaviour using mobile communication instruments. Transport. Res. C Emer. **12**, 273–291 (2004)
8. Caceres, N., Wideberg, J., Benitez, F.: Deriving origin destination data from a mobile phone network. IET Intell. Transp. Sy. **1**, 15–26 (2007)
9. Jing, W., Dianhai, W., Xianmin, S., Di, S.: Dynamic OD expansion method based on mobile phone location. In: Fourth International Conference on Intelligent Computation Technology and Automation, pp. 788–791. IEEE (2011)
10. Iqbal, M.S., Choudhury, C.F., Wang, P., González, M.C.: Development of origin–destination matrices using mobile phone call data. Transport. Res. C Emer. **40**, 63–74 (2014)
11. Bar-Gera, H.: Evaluation of a cellular phone-based system for measurements of traffic speeds and travel times: a case study from Israel. Transport. Res. C Emer. **15**, 380–391 (2007)
12. Zhan, X., Hasan, S., Ukkusuri, S.V., Kamga, C.: Urban link travel time estimation using large-scale taxi data with partial information. Transport. Res. C Emer. **33**, 37–49 (2013)
13. Reades, J., Calabrese, F., Ratti, C.: Eigenplaces: analysing cities using the space-time structure of the mobile phone network. Environ. Plann. B Plann. Des. **36**, 824–836 (2009)
14. Bindra, S.: Using cellphone OD data for regional travel model validation. In: 15th TRB Planning Applications Conference (2015)
15. Ministry of Land, Infrastructure, Transport and Tourism (MLIT). http://www.mlit.go.jp/sogoseisaku/soukou/sogoseisaku_soukou_fr_000018.html
16. NTT Docomo, Inc. https://www.nttdocomo.co.jp/english/corporate/ir/binary/pdf/library/annual/fy2015/p05_e.pdf
17. Japan Statistics Bureau. http://www.stat.go.jp/english/index.htm
18. Ture, M., Tokatli, F., Kurt, I.: Using Kaplan-Meier analysis together with decision tree methods (C&RT, CHAID, QUEST, C4.5 and ID3) in determining recurrence-free survival of breast cancer patients. Expert Syst. Appl. **36**, 2017–2026 (2009)
19. Bargeman, B., Chang-Hyeon, J., Timmermans, H., Van der Waerden, P.: Correlates of tourist vacation behavior: a combination of CHAID and loglinear logit analysis. Tourism Anal. **4**, 83–93 (1999)
20. Chen, J.S.: Market segmentation by tourists' sentiments. Ann. Tourism Res. **30**, 178–193 (2003)
21. Van Middelkoop, M., Borgers, A., Timmermans, H.: Inducing heuristic principles of tourist choice of travel mode: a rule-based approach. J. Travel. Res. **42**, 75–83 (2003)
22. Welte, J.W., Barnes, G.M., Wieczorek, W.F., Tidwell, M.C.: Gambling participation and pathology in the United States - a sociodemographic analysis using classification trees. Addict. Behav. **29**, 983–989 (2004)
23. Chen, J.S.: Developing a travel segmentation methodology: a criterion-based approach. J. Hosp. Tour. Res. **27**, 310–327 (2003)
24. Biggs, D., De Ville, B., Suen, E.: A method of choosing multiway partitions for classification and decision trees. J. Appl. Stat. **18**, 49–62 (1991)