Understanding Human Mobility and Activity in Multiple Non-identifiable Statistics

Ken Hidaka, Hiroshi Ohno and Takahiro Shiga

Abstract

Although the use of location-based services to study human mobility has increased, the spread of their use has been limited due to privacy protection. Population statistics derived from the operational data of mobile networks have also received increasing attention. However, information about human mobility has been eliminated by the aggregation process. Under these conditions, this paper proposes a method to integrate multiple non-identifiable statistics and produce pseudo-data on human mobility. This data consists of numerous virtual agents generated by a stochastic model and fulfills the given statistical features. To evaluate the proposed model's performance, it is applied to an actual data set. As a result, the spatial distribution of people estimated by the model shows a valid correlation to the data estimated using mobile phone data in any time zone. In addition, the comparison of time series of population data shows good correlation in most grids.

K. Hidaka (Corresponding author) • H. Ohno • T. Shiga
Toyota Central R&D Labs., inc., Nagakute, Japan
Email: hidaka@mosk.tytlabs.co.jp

H. Ohno
Email: oono-h@mosk.tytlabs.co.jp

T. Shiga
Email: t-shiga@mosk.tytlabs.co.jp

1. Introduction

Recently, various types of new mobility services such as ride-sourcing (Rayle et al., 2015) and free-floating car sharing (Wielinski et al., 2015) have increased along with urbanization. To design such mobility services and forecast their impact, it is essential to understand human mobility (i.e., who travels, when they travel, where they travel, why they travel, and how they travel).

In the context of traffic engineering, numerous studies have addressed travel-demand estimation using activity-based models, including the utility-maximizing approach (Bowman and Ben-Akiva, 2000; Bhat et al., 2013), rule-based approach (Miller and Roorda, 2003), and mathematical programming approach (Recker, 1995). These approaches require a detailed travel diary based on census data and/or a questionnaire. Although these data have made substantial contributions to the understanding of human mobility, surveys have decreased in frequency due to their high time and cost expenses (Kang et al., 2010; Wu et al., 2014).

Along with the growth in the use of the mobile phones, studies of human mobility using various types of location-based data have increased. These include methods such as GPS trajectories (Fan et al., 2014; Furletti et al., 2013; Hayano and Adachi, 2013), check-in data based on the location of social network services (Hasan et al., 2013; Wu et al., 2014), and anonymous mobile phone call records (Gonzalez et al., 2008; Phithakkitnukoon et al., 2010; Kang et al., 2010). Though location-based services were expected to increase the understanding of human mobility, their use has been limited due to privacy protection. Therefore, a new method to derive small area population statistics based on operational mobile phone data (Terada et al., 2013) has received increasing attention. These statistics are easily collected and updated. However, information about the mobility of people has been eliminated by the process of aggregation to grid square statistics.

Under these circumstances, we propose a method that integrates multiple non-identifiable statistics and produces pseudo-data on human mobility. Data, which fulfill the given statistical features, consist of numerous virtual agents generated by a stochastic model. Section 2 of this paper explains the model in detail. Section 3 describes how the model was applied to an actual data set and the results of evaluation. Conclusions are presented in Sect. 4.

2. Method

2.1 Model Framework

Generation of pseudo-data involved three steps: 1) numerous agents characterized by habitat, age-group, and sex, were generated based on 1-km grid square statistics from the population census, 2) a sequence of activities was assigned to each agent using age-group and sex as input data, and 3) a destination corresponding to each activity was assigned to each agent. In these steps, some attributes (habitat, age-group, sex, activity schedule, destination of each activity) were randomly assigned to each agent. The following sections describe each of these steps.

2.2 Generation of Agents

In this step, numerous agents characterized by habitat, age-group, and sex were generated based on 1-km grid square statistics from the population census.

2.3 Generation of Activity Sequence

An activity sequence was defined as the sequential set of activities in each time zone based on the time of day. The activity in each time zone was assigned to an agent using age-group and sex as input data.

If the transition probability between activities at each time zone was available, the activity sequence was generated using a random number. Specifically, the following relationship was obtained on the assumption of a first-order Markov chain between activities with a transition matrix $\mathbf{A}^{(t)}$:

$$P^{(t)}(i \leftarrow j) = a_{ij}^{(t)} \tag{1}$$

where $P^{(t)}(i \leftarrow j)$ was a choice probability of activity i at time $t+1$, given an activity j at time t, and $a_{ij}^{(t)}$ was the ij element of the matrix $\mathbf{A}^{(t)}$. Therefore, the sequential generation of an activity was possible. Several studies, particularly those on residential electricity demand modeling, have used the Markov chain model to generate residential electricity demand profiles, because they have a high correlation with activity patterns (Torriti, 2014). However, in most cases, the transition probability matrix was estimated by calculating the relative frequency of individual time use data (Widén and Wäckelgärd, 2010). In contrast, our model required aggregated time use data instead of individuals.

The NHK National Time Use Survey (NHK Broadcasting Culture Research Institute, 2015) and the Survey on Time Use and Leisure Activities (Statistics Bureau of Japan, 2015a) provided data suitable for the generation of activity sequences because they had a sufficient number of samples for each age and sex group. Although we could extract the participation rate in each time zone and the participation rate during an entire day, the probability of transition between activities was not available and was thus estimated under the constraint of the participation rate in each time zone. In the following calculation, we formulated this question as a mathematical optimization problem.

The terms $\mathbf{y}^{(t)}$ and $\mathbf{y}^{(t+1)}$ denote an n-dimensional vector each and represent a participation rate at time t and $t+1$ respectively. Then, the number of the elements in the transition probability matrix becomes $n \times n$. Therefore, $\mathbf{A}^{(t)}$ cannot be solved uniquely because the number of equations described below is fewer than the number of unknowns:

$$\mathbf{y}^{(t+1)} = \mathbf{A}^{(t)}\mathbf{y}^{(t)} \tag{2}$$

We considered a toy problem as shown in Fig. 1a. Suppose that an activity set consists of only two elements "home" and "work" and a discrete time set consists of two elements "$t=1$" and "$t=2$." We need to consider the manner in which the transition probability matrix $\mathbf{A}^{(t)}$ is estimated under the following condition: $\mathbf{y}^{(1)} = (0.5,0.5)^T$ and $\mathbf{y}^{(2)} = (0.5,0.5)^T$. In this problem, the following three solutions are included: 1) all agents do not change their activity, 2) half of all agents change their activity, and 3) all agents change their activity. Each solution can be seen in Fig. 1b. Each of them corresponds to a different transition probability matrix. For example, solution 1 corresponds to a diagonal matrix, as shown in Fig. 1b.

To find a unique solution, we consider the participation rate during an entire day. The participation rate corresponding to each solution is calculated by counting the number of agents that experienced the activity. The result is shown in the right side of Fig. 1b. Because the participation rate during an entire day can also be extracted from the survey, the most suitable solution is determined by comparing them. However, in general, it was quite difficult to detect all candidates in the solution. Therefore, we formulated the following mathematical optimization problem and derived the solution.

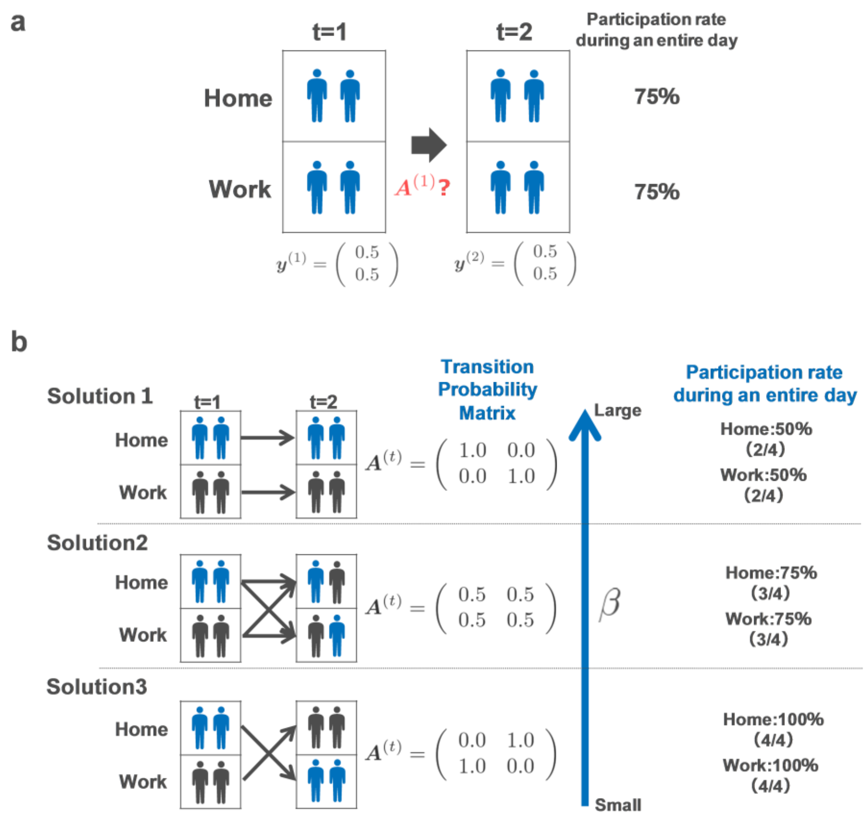

Fig. 1. a) Schematic of the toy problem; **b)** Solution to the toy problem

$$\text{Minimize} \quad \sum_{i,j} a_{ij}^{(t)2}, \tag{3}$$

$$\text{Subject to} \quad \sum_{i,j} a_{ij}^{(t)} = 1, \tag{4}$$

$$a_{ij}^{(t)} > 0 \quad \forall i, j, \tag{5}$$

$$y_i^{(t+1)} - \sum_j a_{ij}^{(t)} y_j^{(t)} = 0, \tag{6}$$

where Eqs. 4 and 5 were constraints on probability. Equation 6 is an important constraint in the given statistical features (i.e., the participation rate in each time zone). Equation 3, which is defined as the sum of the squares of shares, is used as a measure of diversity in the context of economy and ecology known as the Herfindahl-Hirschman index (Hirschman, 1945) and the Simpson diversity index (Simpson, 1949), respectively.

We assume that people tend to continue the same activity. To express the ease of the transition, we introduce the weight coefficient β_{ij} $(\forall i, j)$ into each element of Eq. 3. We assume that all the weight coefficients of the transition to another activity are the same and that the probability of the same activity is high. Accordingly, we could reduce the number of the weight coefficients from n^2 to n. In the following, the weight coefficients of the transition to another activity β_{ij} $(i \neq j)$ are equal to 1, and the weight coefficients of the transition to the same activity are redefined as $1-\beta_i$. Then, Eq. 3 is expressed:

$$\text{Minimize} \quad \sum_{i,j} \left(1-\beta_i \delta_{ij}\right) a_{ij}^{(t)2}, \tag{7}$$

where δ_{ij} is the Kronecker delta. In addition, the value of β_i ranges between 0 to 1 and, when the larger value is set (i.e., β_i close to 1), the probability of higher transition is obtained. Equation 7 is a quadratic function of $a_{ij}^{(t)}$, and Eqs. 4–6 are the linear constraints of $a_{ij}^{(t)}$. These equations are known as a quadratic programming problem.

The quadratic programming problem is formulated as follows:

$$\text{Minimize} \quad \frac{1}{2}\mathbf{x}^{\mathbf{T}}\mathbf{Q}\mathbf{x} + \mathbf{p}^{\mathbf{T}}\mathbf{x}, \tag{8}$$

$$\text{Subject to} \quad \mathbf{A}^{\mathbf{T}}\mathbf{x} \geq \mathbf{b}^{\mathbf{T}}. \tag{9}$$

Here, we consider the n^2-dimensional vector $\mathbf{x}$ and arrange the elements as follows:

$$x^T = [a_{11}, a_{12}, \cdots, a_{ij}, \cdots, a_{nn}]. \tag{10}$$

Then, the $n^2 \times n^2$ matrix $\mathbf{Q}$ becomes a block diagonal matrix as follows:

$$\mathbf{Q} = \mathrm{diag}[\mathbf{B_1}, \mathbf{B_2}, \cdots, \mathbf{B_n}], \tag{11}$$

where $n \times n$ matrix $\mathbf{B_i}$ is a diagonal matrix as follows:

$$\mathbf{B_i} = \mathrm{diag}[2\left(1-\beta_i \delta_{i1}\right), 2\left(1-\beta_i \delta_{i2}\right), \cdots, 2\left(1-\beta_i \delta_{in}\right)]. \tag{12}$$

The matrix $\mathbf{Q}$ is diagonal and cannot take a negative value based on the definition of β_i. This means that $\mathbf{Q}$ is a positive-definite matrix. Therefore, this is a convex optimization problem and is guaranteed to have a unique solution.

As mentioned above, we formulated the problem for a transition probability matrix as the quadratic programming problem. Finally, the weight coefficients β_i were adjusted to fit the highest given participation rate for the whole day and the transition probability matrix was calculated to use the coefficients.

2.4 Generation of the Destination Sequence

The destination sequence was defined as the sequential set of destinations in each time zone and showed where each agent visited during the day. The destination corresponding to each activity was assigned to the agent using Wilson's entropy model (Wilson, 1967) as follows:

$$X_{ij} = A_i B_j O_i D_j e^{-\gamma c_{ij}}, \tag{13}$$

where X_{ij} denotes the number of trips between i and j, O_i is the total number of trip origins at grid square i, D_j is the total number of trip destinations at grid square j, c_{ij} is the distance between i and j, and A_i, B_j and γ denote parameters. The equation showed that the total number of trips was proportion to the number of trip origins (i.e., trip demands) and destinations (i.e., attractiveness). The equation also showed that the number of trips became larger when the distance decreased.

The parameters A_i, B_j and γ were determined using the following constraints:

$$O_i = \sum_j X_{ij}, \tag{14}$$

$$D_j = \sum_i X_{ij}, \tag{15}$$

$$\sum_{i,j} X_{ij} c_{ij} = C, \tag{16}$$

where C was the total trip distance. Because the total trip distance was calculated by multiplying the number of trips by the average trip distance, we could determine the parameters under the given average constraint for the trip distance.

Using Eqs. 13–16, the following equations were obtained:

$$A_i = \left(\sum_j B_j D_j e^{-\gamma c_{ij}} \right)^{-1}, \tag{17}$$

$$B_j = \left(\sum_i A_i O_i e^{-\gamma c_{ij}} \right)^{-1}, \tag{18}$$

$$\sum_{i,j} A_i B_j O_i D_j e^{-\gamma c_{ij}} \cdot c_{ij} = C. \tag{19}$$

These equations were typically solved using an iterative procedure.

In the previous step, the activity sequence for each agent was already known. Therefore, we could estimate the number of trips for every activity. The term $X_{ij}^{(k)}$ denoted the number of trips on activity k. Then, $X_{ij}^{(k)}$ was expressed as:

$$X_{ij}^{(k)} = A_i^{(k)} B_j^{(k)} O_i^{(k)} D_j^{(k)} e^{-\gamma^{(k)} c_{ij}}, \tag{20}$$

where superscript k meant "about activity k." Assuming the first-order Markov process on the choice of destination, we obtained the probability for the choice of destination as follows:

$$P\left(\mathrm{pos}^{(t+1)} = j \mid \mathrm{act}^{(t+1)} = k, \mathrm{pos}^{(t)} = i\right) = \frac{X_{ij}^{(k)}}{\sum_j X_{ij}^{(k)}} \tag{21}$$

where the superscript t denoted time, $\mathrm{pos}^{(t)}$ was a position at t, and $\mathrm{act}^{(t)}$ was the activity at t. By calculating the number of trips for every activity, we generated a destination sequence for each agent.

3. Application of Proposed Model

In the following sections, we describe how the model is applied to an actual data set.

3.1 Data sets and Processing

To apply the proposed model, the following data sets were used:

- Grid Square Statistics of 2010 Population Census
- The NHK National Time Use Survey (2010)
- Grid Square Statistics of 2009 Economic Census for Business Frame
- Grid Square Statistics of Census of Commerce (2009)
- Establishment and Enterprise Census of Japan (2006)

The Grid Square Statistics divides Japan into a small mesh grid based on latitudes and longitudes (Statistics Bureau of Japan, 2015). In this study, we used a 1-km grid square unit of analysis in accordance with statistical data.

"Mobile Spatial Statistics" is a new type of small area population statistics developed by NTT DOCOMO, which is the largest provider of mobile services in Japan. With a 40% market share of mobile services in Japan, DOCOMO has 60 million subscribers. The statistics provided us with an hourly population for each grid square. Population census statistics provide us with the de jure population (the legally resident population), whereas "Mobile Spatial Statistics" is one of the few statistics that can provide us the de facto population (the population actually present at a given time). Therefore, it was reasonable to evaluate the reproducibility of the de facto population using "Mobile Spatial Statistics." We used the data of the average number of users over one month in each time zone in order to smoothen out day-to-day fluctuations.

The general process for evaluating the model is summarized below:

1. Agents were generated and assigned attributes (habitat, age-group, and sex) based on the Grid Square Statistics of Population Census. Age groups were aggregated into seven groups (15–19, 20s, 30s, 40s, 50s, 60s, and 70+) by sex to fit the division of the NHK National Time Use Survey (hereinafter referred to as the NHK data)

2. Each agent was assigned an activity sequence using the transition probability matrix derived from the NHK data (described in Sect. 3.3)
3. A destination corresponding to each activity was assigned to each agent using destination choice probability derived from several types of census data. We employed a straight line as the distance between the grid squares and defined the inside distance of each grid square as a uniform 500 m. Moreover, we extracted the following features:

- Work: number of employees
- Schoolwork: number of students
- Medical treatment: number of outpatients calculated by multiplying the number of medical institutions (hospitals, clinics, and dental offices) by the average number of outpatients at each institution
- Shopping: annual retail business sales

Further, we assumed that the agents did not move when activities other those outlined above were selected. This meant that the activities occurred in the same place as the previous activity.

The distance coefficient γ was solved numerically when the average trip distance was given. The average trip distance for each activity was calculated from the results of a web questionnaire for 1,600 people living in Nagoya, Japan (described in Sect. 3.4).

3.2 Study Area

We used “Mobile Spatial Statistics” to evaluate the performance of the model. We acquired data covering the urban area of Toyota city, which is located in the middle of the Aichi prefecture. Toyota has a population of 270,000 and includes 309 grid squares. We established a study area that included parts of four cities (Toyota, Okazaki, Miyoshi, and Anjo), as shown in Fig. 2. These cities were chosen because of their frequent intercommunications. The study area contained 576 grid squares (equating to 576 square kilometers) with a population of 717,384.

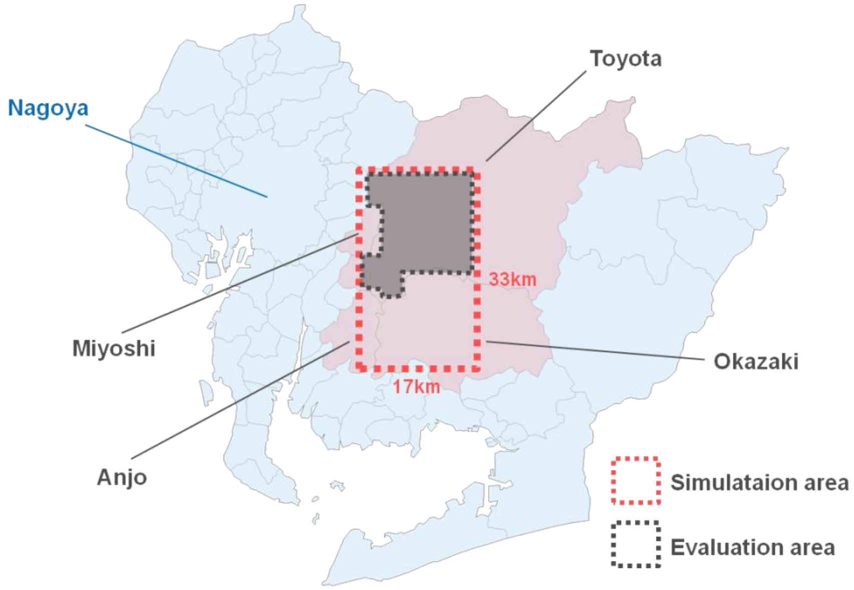

Fig. 2. Study area (source: CraftMap)

3.3 Application of the Activity Sequence Generation Model

As mentioned in Sect. 2.3, we formulated the activity sequence generation model as a convex quadratic programming problem. This section describes how the model was applied to the NHK data.

3.3.1 Definition of Activities

A total of 27 types of activities were defined in the NHK data. Because most of these had no relation to out-of-home activities, we reclassified "at home" and nine other out-of-home activities, as shown in Table 1.

Although the "at home" activities were not included in the classification of the NHK data, we extracted the hourly rate for "at home" activities from another NHK data set. Moreover, we were not able to obtain the participation rate for "eating out" from the NHK data because the NHK data does not distinguish between "eating out" and "eating at home." Therefore, we assumed that the rate for "eating out" could be calculated by multiplying the "meals" and "out of home" rates. The "out of home" rate was calculated by subtracting the "at home" rate from 1. However, because the sum of

Table 1. Type and definition of activities

Symbol	Classification	Definition
H	At home	At home
A_1	Work	Work, work-related association + Commuting to work
A_2	Schoolwork	Classes and school activities + Commuting to school
A_3	Shopping	Shopping
A_4	Exercise and sports	Exercise and sports
A_5	Outings and walks	Outings and walks
A_6	Hobbies, entertainment, cultural activities	Hobbies, entertainment, cultural activities
A_7	Medical treatment or recuperation	Medical treatment or recuperation
A_8	Eating out	Minimum of "meals" times "out-of-home" and "out-of-home" minus the sum of A_1 to A_7 i.e. $\min\left\{\text{Meals} \times (1-H), 1-H-\sum_{i=1}^{7} A_i\right\}$
A_9	Other out-of-home activities	$1-\sum_{i=1}^{8} A_i$

the "at home" rate and "out of home" rate (i.e., the sum of A_1 to A_8 shown in Table. 1) could exceed 1, we added a constraint under which the upper bound of the "eating out" rate was less than the "out of home" rate minus the sum of A_1 to A_7. In addition, the "other out-of-home activity" rate (A_9) represented an out-of-home activity other than A_1 to A_8.

3.3.2 Application of the Model

In the beginning, participation rates of the 10 types of activities defined in each time zone were calculated using the procedure described in Sect. 3.3.1. Participation rates were aggregated to an hourly unit by averaging 15 min of unit data. We then calculated the transition probability matrices by applying the method described in Sect. 2.3 to the hourly participation rate data. To solve the quadratic programming problem, we used "quadprog," which is the quadratic programming package of R language. Because the participation rate for an entire day cannot be obtained analytically, we need to utilize a numerical simulation. Aggregating this result provided the participation rate for the entire day. Finally, based on the comparison of the simulation results and the data, we chose the best-fit parameter β_i. The larger β lowered the participation rate for the entire day.

3.3.3 Application Results

Figure 3 shows some samples of the activity sequences generated by the method shown in the previous section. The figure shows the activity sequence of **a)** men 15–19 years old, **b)** men 40–49 years old, **c)** women 40–

49 years old, and **d)** women 70+ years old. Based on these figures, we can see that the activity patterns generated by the model differ for each attribute.

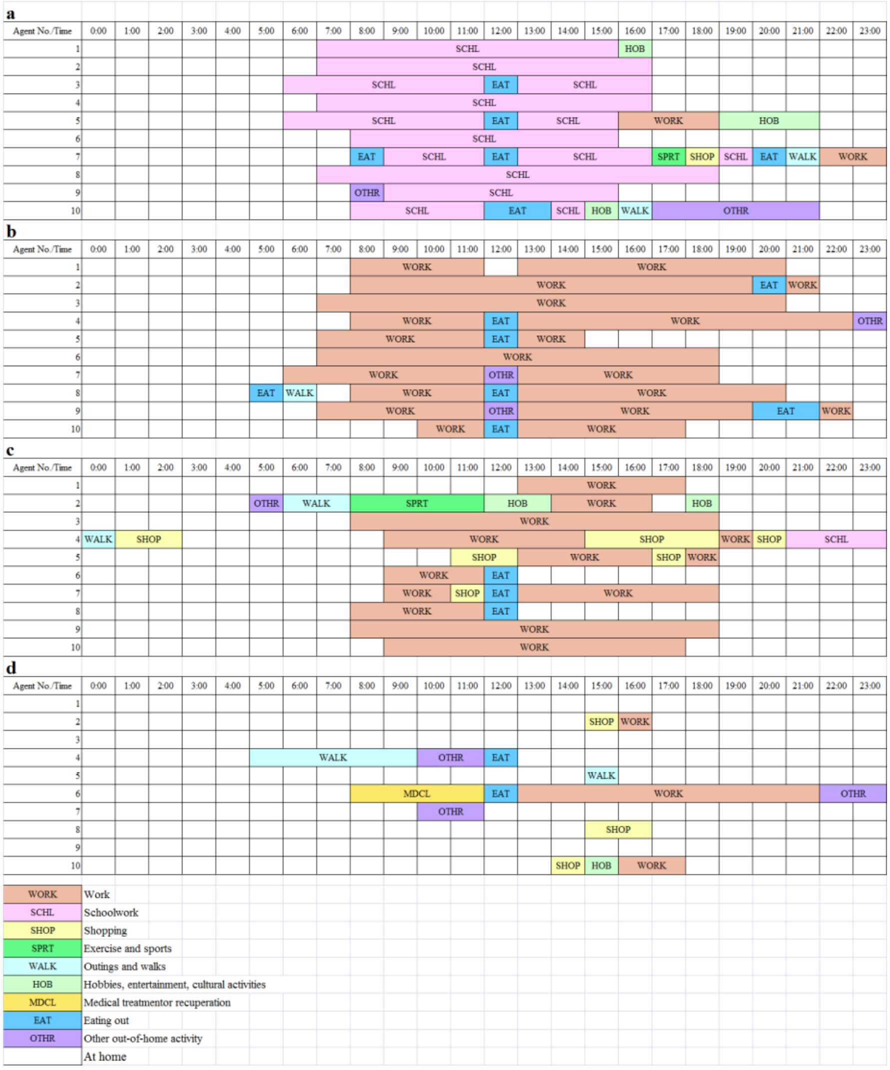

Fig. 3. Examples of activity sequences for **a)** men 15-19 years old, **b)** men 40-49 years old, **c)** women 40-49 years old, and **d)** women 70+ years old

3.4 Application of the Destination Sequence Generation Model

Calculation of the parameter γ, which represents sensitivity to distance, is calculated in the following sections. The average of trip distance for each activity obtained from the web questionnaire was as follows: work: 4.14 km, school: 3.42 km, medical treatment: 2.72 km, shopping: 2.08 km. Applying Wilson's entropy model under the given conditions, we could not find a feasible solution for shopping. This is likely because people living in Toyota city have greater distances to travel for shopping than those in Nagoya city. While seeking a solution, we determined that to have a feasible solution, the average trip distance for shopping should be more than 2.4 km. Therefore, we assumed a distance of 2.4 km. Consequently, we obtained the following parameters: work: 0.40, school: 0.70, medical treatment: 0.76, and shopping: 1.22. The parameter γ expresses distance sensitivity. Thus, when compared to other purposes such as medical treatment, school, and work, people who shop tend to choose a closer location. This tendency occurs in the following order: shopping, medical treatment, school, and work.

3.5 Results and Discussion

The proposed model was applied to an area in the middle of the Aichi prefecture. Figure 4 shows the comparison of “Mobile Spatial Statistics” and estimated population in all grids. Therefore, the spatial distribution of people estimated by the model shows a valid correlation ($R > 0.9$) to that estimated using the mobile phone data in any time zone. In addition, Fig. 5 compares the time series of the population in each grid. Some grids such as those shown in Figs. 5a and b show strong correlations, whereas few others do not. To evaluate the reproducibility of time variances quantitatively, we calculated a correlation coefficient for each grid. We also calculated the cumulative distribution of the correlation coefficient. The results showed that almost 50% of the grids have positive correlations ($R > 0.8$) and almost 30% of the grids have negative correlation coefficients ($R < 0$).

According to the “Mobile Spatial Statistics” (Oyabu et al., 2013), data were reliable in grids defined as densely inhabited. Specifically, in 1-km grid squares with a population greater than 3,000, 80% of the area is within a deviation rate of $\pm$10%. Therefore, the evaluation using only the highly reliable grids showed that 90% of the grids have positive correlations ($R > 0.8$).

The performance of the proposed model was confirmed on the basis of the comparison of its results with “Mobile Spatial Statistics”. However, further improvement in the precision and resolution of the model are required. To improve precision, the future study should include:

- Expansion of facilities
- Expansion of the target activity of mobility
- Expansion of the study area
- Use of frequently-updated data

Incorporating frequently updated data utilizing techniques such as data assimilation is necessary to respond to human mobility that varies daily. In addition, incorporating high resolution data is important for improving spatial resolution. A method for estimating detailed population data with household locations, called "Micro-population Census", was recently proposed (Akiyama et al., 2013). Using such data is expected to improve the resolution of the model.

4. Conclusions

In this paper, we proposed a method for generating large amounts of virtual data on human mobility (who travels, when they travel, where they travel, why they travel, and how they travel) without using identifiable data. Data consisted of pseudo-agents with several attributes (habitat, age-group, sex, activities for an entire day, destination for each activity) and fulfills the given statistical features. A significant point to be emphasized is that the proposed model does not require disaggregated data. In contrast, full-fledged activity-based models generally require large amount of detailed disaggregated data. Thus, we are not concerned about privacy protection.

The performance of the proposed model was examined by applying it to a real data set and comparing the results with a time profile of the estimated population based on mobile phone data. The results suggested that the proposed model shows effective reproducibility, particularly in densely populated areas. Positive correlations have indicated 0.8 and more in almost 90% of the grids. One of the limitations of the model is its resolution, which depends entirely on the resolution of the input data. Therefore, using high resolution data is expected to improve resolution. We have proposed to use such high resolution data in future studies.

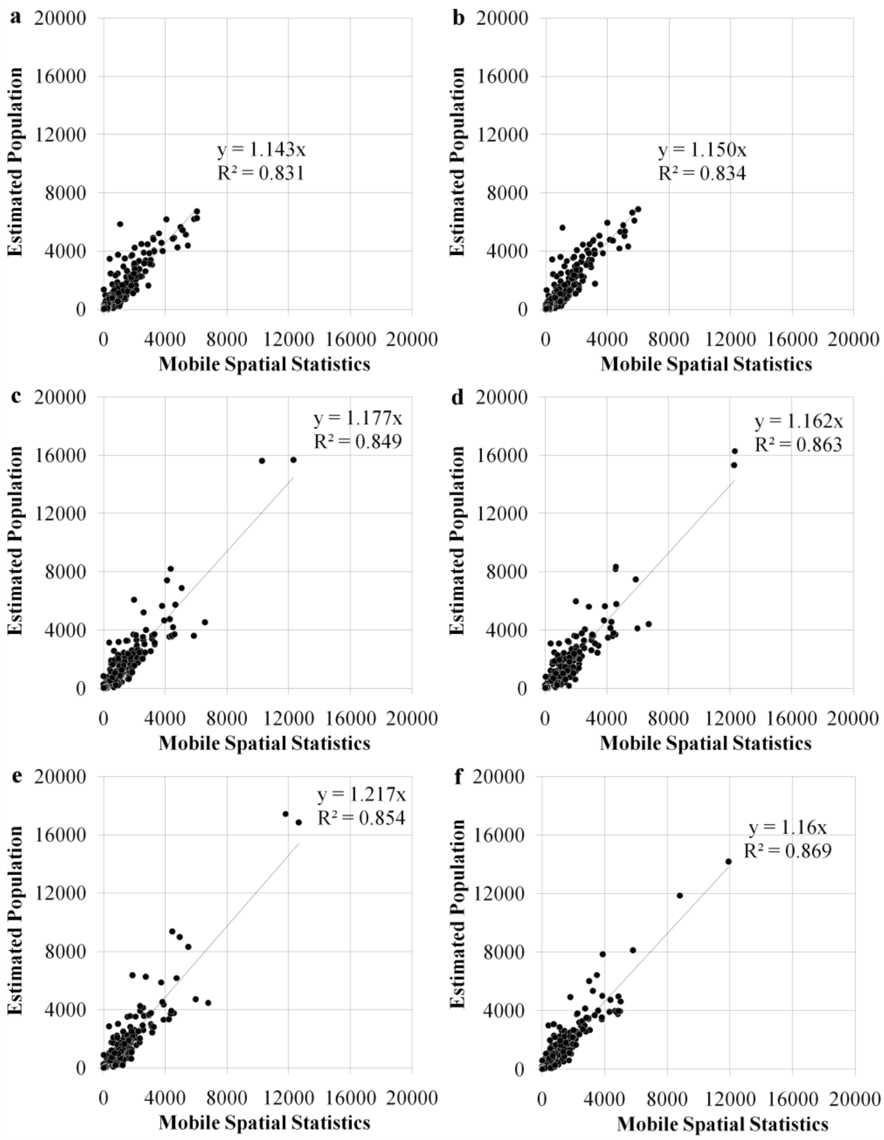

Fig. 4. Comparison of "Mobile Spatial Statistics" and estimated population at **a)** 3:00, **b)** 6:00, **c)** 9:00, **d)** 12:00, **e)** 15:00, and **f)** 18:00

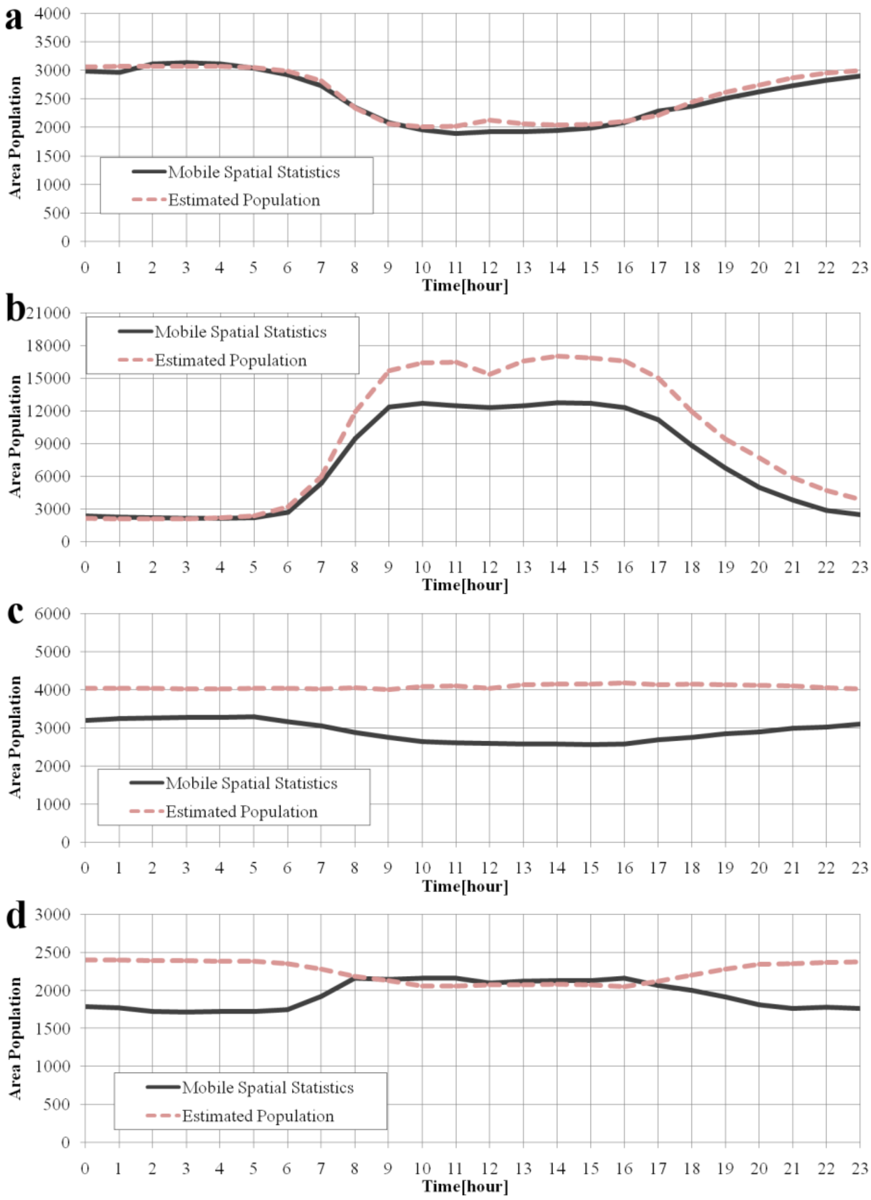

Fig. 5. Comparison of hourly area population for "Mobile Spatial Statistics" and estimated population. Figures **a** and **b** are successful examples. Figures **c** and **d** are failed examples.

References

Akiyama, Y., Takada, H., and Shibasaki, R. (2013). Development of Micropopulation Census through Disaggregation of National Population Census. In *Proceedings of the 13th International Conference on Computers in Urban Planning and Urban Management.* Utrecht.

Ben-Akiva, M., and Bowman, J. (2000). Activity based disaggregate travel demand model system with daily activity schedules. *Transportation Research Part A*, *35*(1), 1–28.

Bhat, C. R., Goulias, K. G., Pendyala, R. M., Paleti, R., Sidharthan, R., Schmitt, L., and Hu, H. H. (2013). A household-level activity pattern generation model with an application for Southern California. *Transportation*, *40*(5), 1063–1086.

Fan, Z., Song, X., and Shibasaki, R. (2014). CitySpectrum : A Non-negative Tensor Factorization Approach. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing.* Seattle.

Furletti, B., Cnr, K. I., Cintia, P., Cnr, K.-I., and Spinsanti, L. (2013). Inferring human activities from GPS tracks. In *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing.* Chicago.

González, M. C., Hidalgo, C. a., and Barabási, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, *453*(June), 779–782.

Hasan, S., Lafayette, W., and Ukkusuri, S. V. (2013). Understanding Urban Human Activity and Mobility Patterns Using Large-scale Location-based Data from Online Social Media. In *Proceedings of the 2nd ACM SIGKDD InternationalWorkshop on Urban Computing.* Chicago.

Hayano, R. S., and Adachi, R. (2013). Estimation of the total population moving into and out of the 20 km evacuation zone during the Fukushima NPP accident as calculated using "Auto-GPS" mobile phone data. *Proceedings of the Japan Academy. Series B, Physical and biological sciences*, *89*(5), 196–199.

Hirschman, A. O. (1945). *National Power and Structure of Foreign Trade*. Berkeley, California: University of California Press.

Kang, C., Gao, S., Lin, X., Xiao, Y., Yuan, Y., Liu, Y., and Ma, X. (2010). Analyzing and geo-visualizing individual human mobility patterns using mobile call records. In *Proceedings of the 18th International Conference on Geoinformatics.* Beijing.

Miller, E. J., and Roorda, M. J. (2003). A Prototype Model of Household Activity/ Travel Scheduling. In *Proceedings of the 82nd Annual Meeting of the Transportation Research Board.* Washington D.C..

NHK Broadcasting Culture Research Institute. (2015). *NHK Broadcasting Culture Research Institute, publication.* http://www.nhk.or.jp/bunken/english/publications/index.html (Cited 19 February 2015).

Oyabu, Y., Terada, M., Yamaguchi, T., Iwasawa, S., Hagiwara, J., and Koizumi, D. (2013). Evaluating Reliability of Mobile Spatial Statistics. *Docomo Technical Journal*, *14*(3), 16–23.

Phithakkitnukoon, S., Teerayut, H., Giusy, D. L., Ryosuke, S., and Carlo, R.

(2010). Activity-aware map: identifying human daily activity pattern using mobile phone data. In A. A. Salah, T. Gevers, N. Sebe, & A. Vinciarelli (Eds.), *Proceedings of the 1st InternationalWorkshop on human behavior understanding* (pp. 14–25). Istanbul: Springer Berlin Heidelberg.

Rayle, L., Shaheen, S., Chan, N., Dai, D., and Cervero, R. (2015). App-based, ondemand ride services: Comparing taxi and ridesourcing trips and user characteristics in San Francisco. In *Proceedings of the 94th Annual Meeting of the Transportation Research Board.* Washington D.C..

Recker, W.W. (1995). The household activity pattern problem: General formulation and solution. *Transportation Research Part B*, *29*(1), 61–77.

Simpson. (1949). Measurement of Dirversity. *Nature*, *163*, 688.

Statistics Bureau of Japan. (2015a). *Grid Square Statistics.* http://www.stat.go.jp/english/data/mesh/index.htm (Cited 19 February 2015).

Statistics Bureau of Japan. (2015b). *Survey on Time Use and Leisure Activities.* http://www.stat.go.jp/english/data/shakai/ (Cited 19 February 2015).

Terada, M., Nagata, T., & Kobayashi, M. (2013). Population Estimation Technology for Mobile Spatial Statistics. *Docomo Technical Journal*, *14*(3), 10–15.

Torriti, J. (2014). A review of time use models of residential electricity demand. *Renewable and Sustainable Energy Reviews*, *37*, 265–272.

Widén, J., and W¨ackelgård, E. (2010). A high-resolution stochastic model of domestic activity patterns and electricity demand. *Applied Energy*, *87*(6), 1880–1892.

Wielinski, G., Centre-ville, S., Centre-ville, S., and Morency, C. (2015). What about free-floating carsharing ? A look at the Montreal case. In *Proceedings of the 94th Annual Meeting of the Transportation Research Board Annual Meeting of the Transportation Research Board.* Washington D.C..

Wilson, A. G. (1967). A statistical theory of spatial distribution models. *Transportation Research*, *1*(3), 253-269.

Wu, L., Zhi, Y., Sui, Z., and Liu, Y. (2014). Intra-urban human mobility and activity transition: Evidence from social media check-in data. *PLoS ONE*, *9*(5), 1–13.