Agent-based Estimation of Household Micro-data with Detailed Attributes for a Real City

Nao Sugiki, Tomoya Muranaka, Noriko Otani, and Kazuaki Miyamoto

Abstract

Intensive developments in land-use microsimulation models have been made in order to forecast changes in household attributes and location. However, a micro-data set of the base year is usually unavailable because of the right to privacy. Thus, this study proposes a system based on a convenient and practical method to estimate the agent-based household micro-data set of the base year in a real city. First, a set of member attributes is generated based on existing methods. Second, other attributes are synthesized based on their similarity, which is defined as the distance between the estimated and the referenced sample households. Finally, we use a case study to demonstrate the utility of the system by applying it to Toyama City in Japan.

N. Sugiki (Corresponding author)
Transportation Department, Docon Co. Ltd., 4-1, 5-chome, 1-jo, Atsubetsu-chuo, Atsubetsu-ku, Sapporo, 004-8585 Japan
Email: ns1491@docon.jp

T. Muranaka
Kyushu Branch, Survey Research Center Co. Ltd., 2-6-26 Hakataeki-Higashi, Hakata-ku, Fukuoka, 812-0013 Japan
Email: muranaka_t@surece.co.jp

N. Otani
Faculty of Informatics, Tokyo City University, 3-3-1 Ushikubo-nishi, Tsuzuki-ku, Yokohama, 224-8551 Japan
Email: otani@tcu.ac.jp

K. Miyamoto
Faculty of Urban Life Studies, Tokyo City University, 8-9-18 Todoroki, Setagaya-ku, Tokyo, 158-8586 Japan
Email: miyamoto@tcu.ac.jp

1. Introduction

In the field of urban modeling, microsimulation is becoming a popular approach to describe the detailed changes in land use and transport in a metropolis. This technique provides the detailed information necessary for decision making on emerging issues at the household or firm level. In this study, we consider the individual household data required for the base year in a residential microsimulation.

An individual household is characterized by many attributes, such as member composition and age, housing type and location, car ownership, and income. Each of these attributes must, in principle, be defined for the base year in the microsimulation. However, these data are usually unavailable because the content of administrative registers and censuses are protected by the right to privacy in most countries. As a result, the data used in microsimulation models are so-called “synthetic populations.” These are created from generally accessible aggregate data provided by the national census, with additional information obtained by conducting sample surveys. Most existing procedures used to generate synthetic populations estimate the number of households by type after categorizing households using the iterative proportional fitting procedure (Pritchard and Miller 2009; Müller and Axhausen 2011; Lee and Fu 2011). However, this approach has several shortcomings when the model has to deal with many types of household attributes. Another approach is to generate a set of individual synthetic households using Monte Carlo sampling by agent (Moeckel et al. 2003). Here, each household has a unique set of attributes (i.e., “micro-data,” in the terminology of this study).

Miyamoto et al. (2010a, 2010b) used the Monte Carlo approach to develop a comprehensive system based on a method to estimate the micro-data set of the base year. Their model handles both continuous and discrete attribute variables by agent (i.e., a household and its members) (see also Sugiki et al. 2012). These authors proposed an indicator to evaluate the goodness-of-fit of the two micro-data sets originally presented in Otani et al. (2010), and subsequently extended in Otani et al. (2011). Then, this indicator was used to develop their system in a rational and objective manner. Moreover, they demonstrated the utility of the system by applying it to person-trip survey data for the Sapporo metropolitan area. However, the system has not yet been applied and validated in a real city.

This study proposes a convenient and practical system to estimate the agent-based household micro-data set of the base year for a land-use microsimulation of a real city. The system consists of two estimation stages. In the first stage, a set of member attributes is generated for a given number of households and household size. The attributes describe each member's relationship with the household head, gender, and age. In the second stage, further attributes are synthesized using the attributes of the most similar sample household, taking spatial proximity into account. These attributes include the housing type and commuting zone. Finally, we confirm the utility of the system in a case study in which we apply the method to 160,000 households in Toyama City. The sample data set is constructed from the results of a questionnaire survey conducted in Toyama City. The data set contains the micro-data, as well as data on the population by age band and by gender, and on the number of households by household size. The latter data are taken from the national census and are used as control totals in each zone.

2. System Development

2.1 Estimation Procedure

The system proposed in this study consists of the two estimation stages shown in Fig. 1. In the first stage, a set of member attributes is generated for a given number of households and household size. The attributes describe each member's relationship with the household head, gender, and age, and are generated using the same approach as existing methods (Miyamoto et al. 2010a, 2010b; Sugiki et al. 2012).

In the second stage, attributes such as housing type and commuting zone are synthesized using the attributes of the most similar sample household. Here, similarity is defined as the distance between the estimated and the referenced sample households, considering their spatial proximity. Similarity is calculated by household composition, member age, and member gender, and the spatial distance between the zone of the estimated micro-data and referenced sample data. Each estimation or synthesis and adjustment is probabilistic and uses the Monte Carlo approach.

Although this system has been developed to estimate the micro-data set of the base year, it provides an alternative procedure emphasizing convenience and practicality making it applicable to a land-use microsimulation in a real city.

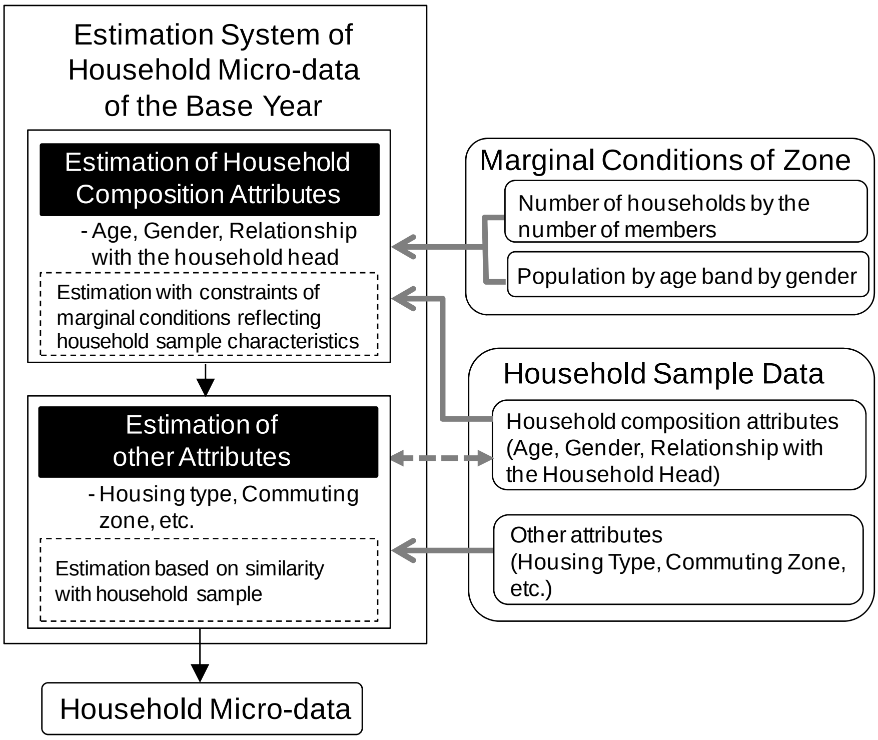

Fig. 1. Procedure of the estimation system

2.2 Propositions

The estimation of the household micro-data set for the base year is based on the following propositions:

- Households are the target agents.
- Households are characterized by members, and members are characterized by their relationship with the household head and their gender and age.
- Households and members are specified by optional attributes utilized in land-use and transportation analyses, such as housing type, commuting zone, and so on.

- The number of households by household size and the number of individuals in five-year age bands are available from census data for the study area.
- Samples are available for which the micro-data are known.
- The study area is divided into zones, and the spatial distance between each pair of zones can be defined.

2.3 Estimation Method for Household Composition Attributes

Fig. 2 shows the flowchart of the estimation procedure for the household composition attributes. The basic estimation principles are as follows:

- The population micro-data set for the base year is estimated from probabilities obtained from a sample micro-data set.
- The relationships between attributes (e.g., household member composition and age of members) are considered if a member composition is common in the sample data set. These relationships are derived from the sample set.
- If a member composition is rare in the sample data set, the age composition is determined to be the same as that of the sample.
- The synthetic population is adjusted to fit the marginal conditions (e.g., population by age band by gender).
- Each estimation or synthesis and adjustment is probabilistic and uses the Monte Carlo approach.

This system is novel for dealing with the relationships between continuous attributes. In this case, the age of household members $\mathbf{x}$ is considered. The procedure only applies to households whose membership composition $\mathbf{c}$ is sufficiently common, such as sampled data that have over 10 degrees of freedom. However, when the member composition $\mathbf{c}$ is rare, the age composition $\mathbf{x}$ is used for the sample household. First, we use a principal component analysis to convert the original attribute variables $\mathbf{x} = (x_1, x_2, \cdots, x_m)$ (i.e., the ages of household members) for sample households with m members into independent or non-correlated variables $\mathbf{p} = (p_1, p_2, \cdots, p_m)$:

$$p_i = \sum_k^m v_{ik} x_k \,, \tag{2.1}$$

where p_i is a non-correlated variable for member i, x_k is the age of household member k, and v_{ik} is the principal component of i for household member k. Written in matrix form, we have:

$$\mathbf{p} = \mathbf{V}\mathbf{x}\,. \tag{2.2}$$

Then, based on the sample values, the cumulative frequency curve of p_i is drawn for m principal component variables (see Fig. 3). From Eqs. (2.1) and (2.2), we have:

$$\mathbf{x} = \mathbf{V}^{-1}\mathbf{p} = \mathbf{W}\mathbf{p}\,, \tag{2.3}$$

$$x_i = \sum_{k}^{m} W_{ik}\, p_k\,. \tag{2.4}$$

To generate a synthetic household, a random number ran_{is} is generated for member i of household s. Then, p_{is} is obtained from Fig. 3, and the principal component variable x_{is} (or age) is obtained for member i of household s from Eq. (2.4). The procedure is repeated to generate further synthetic households until the total number of households in the study area is attained. Therefore, by introducing the independent variables as intervening variables, the relationships between attributes are easily dealt with and the system becomes straightforward to operate.

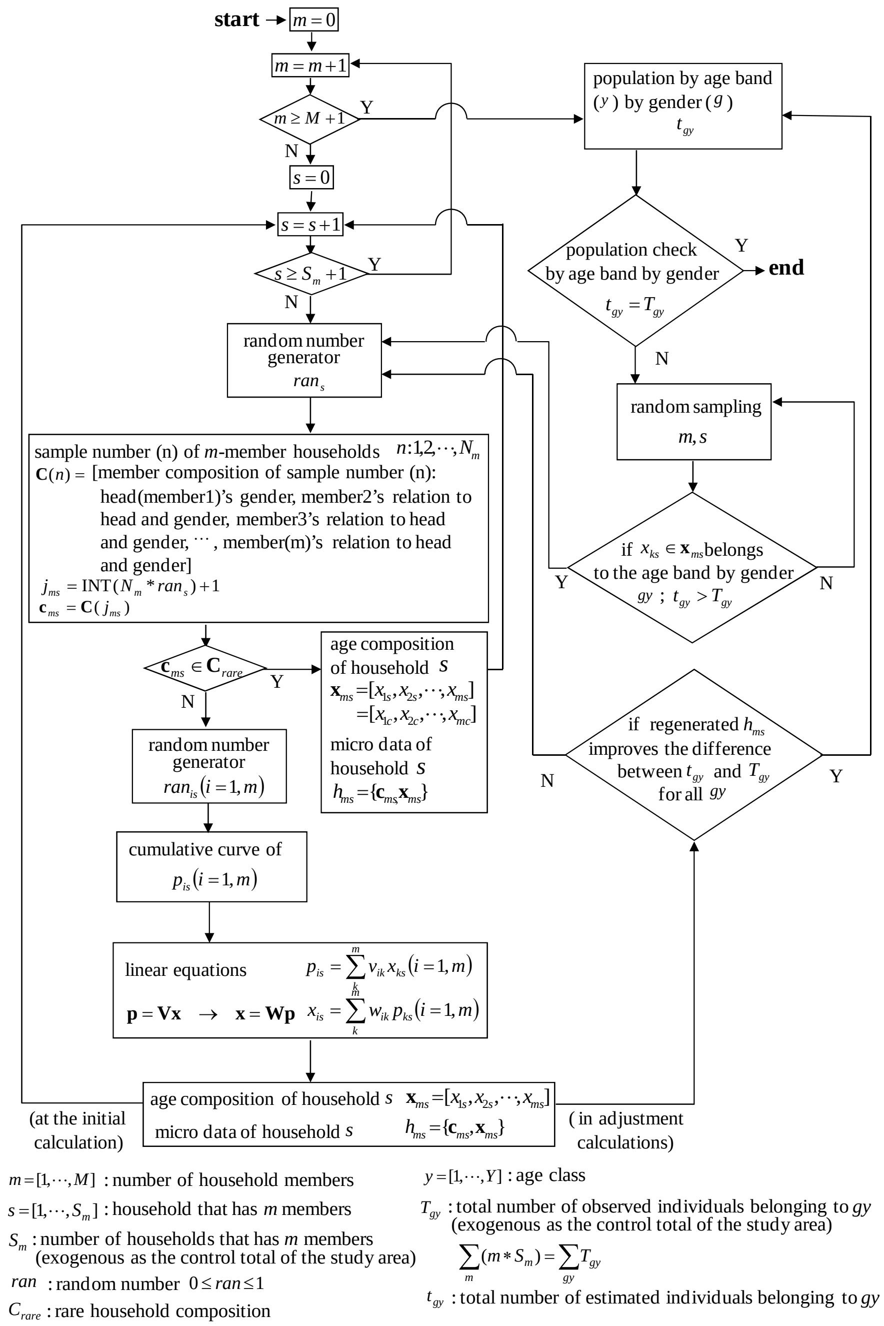

Fig. 2. Flowchart of the estimation procedure for the household composition attributes

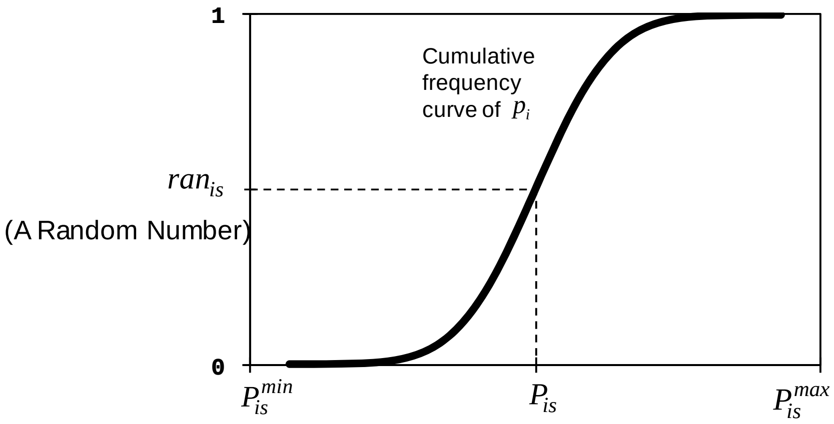

Fig. 3. Correlation determination using the independent variables

2.4 Estimation Method for the other Attributes

To conduct a residential or transport analysis using household micro-data, additional attributes are required for households and members, such as housing type and commuting zone.

In this study, these attributes are synthesized using the attributes of the most similar sample household. The method used to measure the goodness-to-fit between the observed and estimated micro-data sets is presented in Otani et al. (2010, 2011). The measure is defined as the minimum value of the distance summation, and is applied to the separation distance between households.

Ideally, the most similar household sample should be selected from the same zone (i.e., by considering regional characteristics). However, a biased estimation result is assumed in the case of zones with fewer samples, because referable sample data are limited in the same zone. From the viewpoint of system flexibility, we also need stable estimation results, even when the number of samples is low.

Thus, in addition to using the household composition attributes, the separation distance is also defined using spatial proximity in order to determine the attributes of the target micro-data, including the surrounding zones.

The estimation method is shown below. The attributes are first estimated for members and then for households, and refer to the attributes of the sample household with the same number of household members. Target household b and referenced sample household s are represented using the following equations (in vector format):

$$\mathbf{b} = (a_1^b, \cdots, a_K^b, z^b) , \tag{2.5}$$

$$\mathbf{s} = (a_1^s, \cdots, a_K^s, z^s) . \tag{2.6}$$

Here, K is the number of combinations of a member's relationship with the household head and gender, a_k is the age of member k, and z is the zone of the target or referenced household.

The age c_k^m of member m considers the relationship with the household head and gender, and is defined as follows:

$$c_k^m = \begin{cases} a_k : k = k^m \\ 999 : k \neq k^m \end{cases} . \tag{2.7}$$

The member attributes m of target household b are added by searching the individual sample using the minimum separation distance. The member-based separation distance $p\_dis(\mathbf{b}_m, \mathbf{s}_{m'})$ is defined by the consistency of a member's attributes and the zonal spatial proximity, and is represented as follows:

$$p\_dis(\mathbf{b}_m, \mathbf{s}_{m'}) = \sqrt{\sum_{k=1}^{K} (c_k^{b_m} - c_k^{s_m})^2 + \alpha \cdot \sum_{k=1}^{K} (z^{bs})^2} . \tag{2.8}$$

Here, z^{bs} is the spatial distance between the zones of target household b and referenced sample household s. In addition, α is a weighting coefficient of the member's attributes and the spatial distance, and is defined as the inverse of the average of z^{bs}. When $p\_dis(\mathbf{b}_m, \mathbf{s}_{m'})$ is minimized, the similarity of the sample member is adjudged to be greatest.

The attributes of target household b are added by searching the household sample using the minimum separation distance. The household-based separation distance $h\_dis(\mathbf{b}, \mathbf{s})$ is defined using the similarity of the household member composition and the zonal spatial proximity, and is represented as follows:

$$h\_dis(\mathbf{b}, \mathbf{s}) = \sqrt{\sum_{k=1}^{K} (a_k^b - a_k^s)^2 + \sum_{k=1}^{K} (z^{bs})^2} . \tag{2.9}$$

When $h\_dis(\mathbf{b},\mathbf{s})$ is minimized, the similarity of the sample household is adjudged to be greatest.

If two or more samples of the minimum separation distance exist, one is selected probabilistically using the Monte Carlo approach.

3. System Application

3.1 Target Area

Toyama City is a city in Japan with half a million inhabitants. In this section, we estimate the city's household micro-data set as a case study in order to apply the system to a real city. The target area consists of seven major regions, and is divided to 82 middle-sized census zones. This makes it suitable for a land-use microsimulation, as shown in Fig. 4.

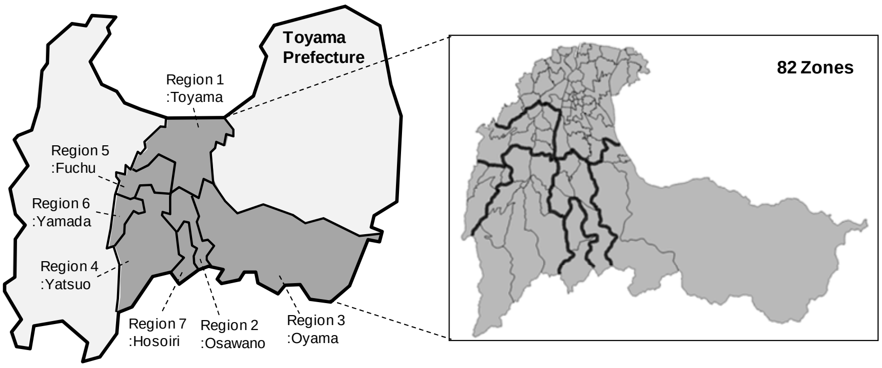

Fig. 4. Target area and zone setting

3.2 Data

The household sample micro-data are constructed from questionnaires collected by the government of Toyama City and our research group. The questionnaires were applied as follows:

- Period: December 2011 to January 2012
- Area: Toyama City in Japan
- Target households: 14,073 households sampled randomly from 140,743 households
- Sent questionnaires: 14,073

- Collected questionnaires: 5,089 (36.2%)
- Sending and collecting: by mail (reply envelopes were enclosed)

Questionnaires with serious missing values were discarded, leaving full information available for 3,649 households and 8,411 individuals.

Data on population by age band by gender and number of households by household size from the 2010 national census were utilized as the control totals in each zone. In the national census, the total population derived from the population by age band by gender is generally not equal to that of the number of households by household size. This is because the categories for households with over seven members are integrated, and the numbers of household members who live in certain facilities are unknown. Thus, the population by age band by gender is adjusted to the total population derived from the number of households in our case study. The total of marginal condition data were 158,833 households and 409,621 individuals.

3.3 Category Setting

Age is categorized into five-year groups (except over 85 years of age). The following 20 general household types were selected based on members' relationships with the household head and their gender: head (male), wife, one child (male), two children (male), three children (male), grandchild (male), brother, father, other (male), two other members (male), head (female), one child (female), two children (female), three children (female), grandchild (female), sister, mother, child's wife, other (female), and two other members (female). Table 1 shows the household types defined by the combinations of relationships with the household head. These are categorized by extracting household type from over 10 samples in the questionnaires.

With regard to attributes other than household composition, the following were considered from the questionnaire data: employment and schooling status, commuting zone of the household head, car ownership, commuting mode of the household head, and housing type. Employment and schooling status is categorized into full-time job, part-time job, housewife, student, unemployed, and other. There were 84 commuting zones, 82 of which are in Toyama City. Then, one zone represents the remaining part of the Toyama Prefecture outside of Toyama City, and one zone is the area outside the Toyama Prefecture. Car ownership is used as a dummy variable, and represents whether the member has free use of a car. The commuting mode of the household head is categorized as car, train, bus, and walk or bicycle. Lastly, housing type is divided into own-detached, own-apartment, rent-detached, and rent-apartment.

Table 1. Categories for household type

Number of Household Members	Household Type (Composition of Household Members)
1	Single (male) Single (female)
2	Couple Head (male) + child (male) Head (male) + child (female) Head (female) + child (male) Head (female) + child (female) Head (male) + mother Two other members
3	Couple + child (male) Couple + child (female) Couple + father Couple + mother Head (male) + child (male) + child (female) Head (male) + sister + mother Three other members
4	Couple + two children (male) Couple + child (male) + child (female) Couple + two children (female) Head (male) + child (male) + mother Head (male) + child (female) + mother Couple + son-generation couple Four other members
5	Couple + two children (male) + child (female) Couple + child (male)+ two children (female) Couple + three children (female) Couple + two children (male) + mother Couple + child (male) + child (female) + mother Couple + son-generation couple + grandson Five other members
6	Couple + son-generation couple + grandson + granddaughter Six other members
7	Seven members

3.4 Calibration

The parameters of B to generate x_i from p_i given by ran_i are generated using parameter matrix A. The latter matrix was obtained using a principal component analysis for household types in the sample household data set with more than 10 degrees of freedom. Table 2 shows the results of B for four-member households as a part of the parameter estimation results.

Table 2. Parameters of B for four-member households

Household Type	b_{ik}	p_1	p_2	p_3	p_4	C
41: Couple + two children (male)	x_1	0.494	0.532	0.504	0.469	10.197
	x_2	0.711	-0.352	-0.556	0.248	4.156
	x_3	-0.438	-0.351	0.030	0.827	3.361
	x_4	0.242	-0.686	0.660	-0.187	9.697
42: Couple + child (male) + child (female)	x_1	0.520	0.490	0.470	0.519	10.108
	x_2	0.546	-0.531	0.457	-0.459	3.823
	x_3	-0.080	0.677	0.168	-0.712	9.601
	x_4	-0.653	-0.138	0.736	0.117	3.704
43: Couple + two children (female)	x_1	0.521	0.463	0.519	0.495	9.923
	x_2	-0.811	0.014	0.471	0.346	9.423
	x_3	-0.259	0.886	-0.271	-0.272	3.872
	x_4	-0.063	0.010	-0.659	0.749	3.231
44: Head (male) + child (male) + mother	x_1	0.482	0.647	0.455	0.376	12.525
	x_2	0.255	-0.660	0.089	0.701	6.200
	x_3	0.449	-0.380	0.553	-0.590	11.975
	x_4	-0.708	-0.037	0.692	0.134	17.075
45: Head (male) + child (female) + mother	x_1	0.540	0.596	0.492	0.333	11.621
	x_2	-0.097	-0.107	-0.381	0.913	11.138
	x_3	-0.462	-0.354	0.779	0.234	5.483
	x_4	0.697	-0.712	0.082	0.024	16.931
46: Couple + son-generation couple	x_1	0.451	0.553	0.435	0.550	15.875
	x_2	0.326	-0.309	0.722	-0.526	10.000
	x_3	0.263	0.658	-0.302	-0.638	14.969
	x_4	0.789	-0.407	-0.446	0.116	9.813

3.5 Results for Household Composition Attributes

Using the data and estimated parameters, the household composition attributes for all micro-data sets were estimated for the 82 zones. Table 3 shows the aggregated number of households by type, and each type's share of the same household size for the observed sample. The table also shows the estimation results for three representative zones with different population sizes, namely 16 (Gofuku district), 69 (Hayahoshi district), and 81 (Hosoirihokubu district).

The influence of the estimation results was a concern in some zones. For example, in zone 16, the number of households by household size and the age structure of the population deviated. However, no extraordinary deviation for specific household types appeared in the estimation results compared with the results of the questionnaire sample. The system successfully synthesized the household micro-data set by zone, regardless of population size or number of households. The characteristics of the household sample was reasonably reflected. This confirmed that the system is able to estimate the household composition attributes in small zones.

Table 3. Estimation results for household composition

Member Composition (m: male, f: female)	Observed Sample		ZONE 16		ZONE 69		ZONE 81	
Single (m)	456	57%	3,276	55%	422	55%	41	57%
Single (f)	350	43%	2,639	45%	350	45%	31	43%
Couple	1,067	79%	654	66%	634	71%	81	74%
Head (m) + child (m)	36	3%	1	0.1%	24	3%	-	-
Head (m) + child (f)	49	4%	2	0.2%	35	4%	3	3%
Head (f) + child (m)	12	1%	-	-	11	1%	-	-
Head (f) + child (f)	55	4%	-	-	62	7%	7	6%
Head (m) + mother	74	6%	1	0.1%	72	8%	14	13%
Two other members	52	4%	339	34%	56	6%	4	4%
Couple + child (m)	354	39%	118	18%	333	38%	29	38%
Couple + child (f)	310	34%	111	17%	310	36%	14	18%
Couple + father	17	2%	17	3%	17	2%	1	1%
Couple + mother	127	14%	-	-	83	10%	21	27%
Head (m) + child (m) + child (f)	12	1%	82	13%	17	2%	1	1%
Head (m) + sister + mother	12	1%	85	13%	15	2%	-	-
Three other members	85	9%	239	37%	90	10%	11	14%
Couple + two children (m)	122	22%	153	31%	215	25%	7	13%
Couple + child (m) + child (f)	203	36%	227	46%	350	41%	17	32%
Couple + two children (f)	78	14%	51	10%	141	16%	1	2%
Head (m) + child (m) + mother	40	7%	1	0.2%	40	5%	10	19%
Head (m) + child (f) + mother	29	5%	-	-	30	3%	4	8%
Couple + son-generation couple	32	6%	1	0.2%	18	2%	2	4%
Four other members	60	11%	64	13%	70	8%	12	23%
Couple + two children (m) + child (f)	21	13%	97	62%	68	21%	2	5%
Couple + child (m) + two children (f)	14	9%	18	12%	41	13%	1	3%
Couple + three children (f)	11	7%	11	7%	22	7%	3	8%
Couple + two children (m) + mother	12	7%	1	1%	23	7%	3	8%
Couple + child (m) + child (f) + mother	11	7%	-	-	21	7%	2	5%
Couple + son-generation couple + grandson	21	13%	-	-	21	7%	5	13%
Five other members	73	45%	29	19%	121	38%	23	59%
Couple + son-generation couple + grandson + granddaughter	22	23%	38	57%	46	28%	3	12%
Six other members	32	34%	45	67%	90	54%	16	62%
Seven members	15	100%	30	100%	63	100%	12	100%
Total	3,864	-	8,330	-	3,911	-	381	-

3.6 Results for other Attributes

Employment and schooling status, the commuting zone of the household head, car ownership, the commuting mode of the household head, and housing type were estimated using the same method, taking into account spatial proximity. The spatial distance was set as the straight line distance between zones. The inverse of the average spatial distance between zones was set as a weighting coefficient of a member's attributes. Then, the spatial distance was taken from Eq. (2.8). To compare the estimation results with and without spatial proximity, we also estimated the case in which the separation distance is defined by the household composition attributes only. Then, we compared the results with those of the proposed system.

The comparison results of household share by employment and schooling status and by housing type in zone 74, where the number of responses was relatively low, are shown in Tables 4 and 5, respectively. For employment and schooling status, the differences in the estimation results with and without spatial proximity are relatively small. In contrast, the results for housing type do not seem to be replicable. Table 6 shows the detached and apartment shares compared with the data obtained from the national census. Here, the estimated share is different from the previous results, indicating that the estimation accuracy improves when using samples from surrounding zones. In summary, the utility of the system after taking spatial proximity into account is confirmed.

Table 4. Estimation results by employment and schooling status (zone 74)

Employment and schooling status	Observed Sample	Estimated (without spatial proximity)	Estimated (with spatial proximity)
Full-time job	34.2%	20.6%	23.8%
Part-time job	8.7%	8.2%	6.5%
Housewife	8.1%	13.5%	11.5%
Student	17.4%	17.3%	14.1%
Unemployed	27.5%	36.4%	40.7%
Other	4.0%	4.0%	3.4%

Table 5. Estimation results by housing type (zone 74)

Housing Type	Observed Sample	Estimated (without spatial proximity)	Estimated (with spatial proximity)
Own-detached	93.8%	94.3%	79.9%
Own-apartment	0.0%	0.0%	1.6%
Rent-detached	0.0%	0.0%	2.6%
Rent-apartment	6.3%	5.7%	15.9%

Table 6. Comparison with the national census for detached and apartment housing types (zone 74)

Housing Type	National Census		Estimated (with spatial proximity)	
Detached	1,075	84.8%	1,061	82.5%
Apartment	193	15.2%	225	17.5

4. Concluding Remarks

This study proposed a convenient and practical system to estimate the agent-based household micro-data set of the base year for a microsimulation in a real city. The system consists of two estimation stages. In the first stage, a set of member attributes is generated for a given number of households and household members using existing methods. These attributes are the relationship with the household head, gender, and age for each household member. In the second stage, other attributes are synthesized using the attributes of the most similar sample household. These attributes include the housing type and commuting zone. Here, similarity is defined as the distance between the estimated and referenced sample households, taking spatial proximity into account.

To test the system in a real city, we estimated the household micro-data set for Toyama City. The city was divided into 82 zones, and sample data were collected using a questionnaire survey. Then, data from the national census were used as control totals in each zone. The system successfully synthesized the household micro-data set by zone, regardless of population size or number of households. Moreover, by comparing the estimation results with and without spatial proximity, we confirmed the utility of the proposed system.

Acknowledgements

This study was supported by a Grant-in-Aid for Scientific Research (23360228) from the Japan Society for the Promotion of Science.

References

Lee DH, Fu YF (2011) Cross-entropy optimization model for population synthesis used in activity-based microsimulation models. In: TRB 90th Annual Meeting Compendium of Papers.

Miyamoto K, Sugiki N, Otani N, Vichiensan V (2010a) An agent based estimation method of household micro-data for the base year in land-use microsimulation. In: TRB 89th Annual Meeting Compendium of Papers.

Miyamoto K, Sugiki N, Otani N, Vichiensan V (2010b) An agent based estimation method of household micro-data including housing information for the base year in land-use microsimulation. In: Selected Proceedings of 12th WCTR.

Moeckel R, Spiekermann K, Wegener M (2003) Creating a synthetic population. In: Proceedings of 8th International Conference on Computers in Urban Planning and Urban Management.

Müller K, Axhausen KW (2011) Population synthesis for microsimulation: State of the art. In: TRB 90th Annual Meeting Compendium of Papers.

Otani N, Sugiki N, Miyamoto K (2010) Goodness-of-fit evaluation method between two sets of household micro-data for land-use microsimulation model. In: Selected Proceedings of 12th WCTR.

Otani N, Sugiki N, Miyamoto K (2011) Goodness-of-fit evaluation method for agent-based household micro-data sets composed of generalized attributes. Transp Res Record 2254:97–103.

Pritchard DR, Miller EJ (2009) Advances in agent population synthesis and application in an integrated land use/transportation model. In: TRB 88th Annual Meeting Compendium of Papers.

Sugiki N, Vichiensan V, Otni N, Miyamoto K (2012) Agent-based household micro-datasets: An estimation method composed of generalized attributes with probabilistic distributions from sample data and available control totals by attribute. Asian Transp Stud 2/1:3–18.