

# 世帯マイクロデータの適合度評価における 重みの決定手法

大谷 紀子<sup>1</sup>・杉木 直<sup>2</sup>・宮本 和明<sup>3</sup>

<sup>1</sup>正会員 東京都市大学准教授 環境情報学部 (〒224-8551 神奈川県横浜市都筑区牛久保西3-3-1)  
E-mail: otani@tcu.ac.jp

<sup>2</sup>正会員 株式会社ドーコン 交通部 (〒004-8585 北海道札幌市厚別区厚別中央1条5丁目4-1)  
E-mail: ns1491@docon.jp

<sup>3</sup>正会員 東京都市大学教授 環境情報学部 (〒224-8551 神奈川県横浜市都筑区牛久保西3-3-1)  
E-mail: miyamoto@tcu.ac.jp

マイクロシミュレーションにおける基準年マイクロデータの推定において、異なる推定手法の再現性能を比較するために、推定データ集合の観測データ集合に対する適合度が定義されている。世帯構成員の年齢・性別・世帯主との関係、居住ゾーン、住居タイプ、自動車保有台数、世帯収入を属性とするマイクロデータを対象とし、各属性のデータ間距離の重み付き和の最小値を適合度としている。重みはシミュレーションの目的に応じて設定すべきであるが、目安がない状態での設定は困難である。また、各属性がシミュレーションの結果に与える影響は、シミュレーションの内容によって異なると考えられる。本研究では、重み設定における目安の提示を目的として、対象シミュレーションにおける各属性の影響度、および重み基準量の算出方法を提案する。

**Key Words :** *Microsimulation, Micro-Data, Estimation of Micro-Data, Goodness-of-Fit Evaluation*

## 1. はじめに

マイクロシミュレーションは、都市圏における土地利用と交通の詳細な変化の記述手法として、欧米諸国の複数の研究グループによって都市モデルの開発への活用が進められている<sup>1)2)</sup>。住宅立地モデルのような世帯を対象としたマイクロシミュレーションモデルの場合、各世帯には世帯主との関係および年齢等の各世帯構成員の属性と、世帯収入、自家用車の保有台数、居住地、住宅タイプ等の世帯全体に関わるものに加えて多くの属性が定義される。マイクロシミュレーションを実行するためには、原理的には、基準年における各属性値をすべての世帯に対して求める必要がある。しかし、国勢調査や住民基本台帳などから個人や個別世帯に関するデータを入手することは、わが国はもとより多くの国において禁止されている。したがって、通常マイクロシミュレーションモデルでは、国勢調査などの入手可能な集計データと追加的に個別世帯の属性情報を提供するサンプル調査を組み合わせ、個別世帯等の一般にエージェントと呼ばれる行動主体に対して複数の属性の組合せを設定した

マイクロデータを作成する。マイクロデータの推定手法としては、世帯タイプをまず設定した上で各タイプに属する世帯数を推定するIPF法や、モンテカルロ法により個々のエージェントのマイクロデータを作成する方法などが提案されている<sup>3)~10)</sup>。

本研究グループでは、マイクロデータのより有効な推定方法に加え、推定精度評価のために、マイクロデータ集合の適合度評価手法を開発している。これまでに、世帯の各構成員の世帯主との関係・年齢・性別、世帯の居住ゾーン、住居タイプ、自動車保有台数、世帯収入を属性とするマイクロデータの評価方法を提案した<sup>11)~17)</sup>。各属性のデータ間距離の重み付き和の最小値を適合度としており、重みはシミュレーションの目的に応じて決定すべきであるが、目安がない状態での決定は困難である。本研究では、特定の属性を重視したシミュレーションを行なう際の重み決定手法を提案する。

本稿では、先行研究におけるマイクロデータの適合度評価手法を概観した上で、特定の属性を重視したシミュレーションを行なう際の重み決定手法を提案する。推定データのうち、重視する属性値に対応する観測データの

値に置き換えたときの、シミュレーション結果の変化に従って重みを決定する。道央都市圏パーソントリップ調査に基づくマイクロデータを使用し、提案手法により重みをつけた適合度を比較することで、提案手法の妥当性を検証する。

## 2. マイクロデータの適合度評価

観測データ集合により近い推定データ集合を決定するために、観測データ集合に対する推定データ集合の適合度を算出することを適合度評価問題と定義する。ここで、完全な情報を持つ観測データ集合が推定手法の妥当性検証のために入手可能であり、推定データは杉木ら<sup>3)~10)</sup>等のマイクロデータ推定手法により提供されるものとする。本節では、処理対象とするマイクロデータ、適合度の定義および算出方法について概説する。

### (1) 処理対象マイクロデータ

本研究で処理対象とするマイクロデータは、特定のゾーンまたは地域の個別世帯データである。各データには以下の属性情報が含まれている。

#### 【世帯構成員に関する属性】

年齢・性別・世帯主との関係

#### 【世帯全体に関する属性】

住宅タイプ・居住ゾーン・自動車保有台数・世帯収入  
世帯構成員に関する属性のうち、性別と世帯主との関係

を合わせて次の20カテゴリで表現される。

本人・男	本人・女	妻	その他1・男
息子1	娘1	父	その他2・男
息子2	娘2	母	その他1・女
息子3	娘3	息子の妻	その他2・女
孫・男	孫・女	兄弟	姉妹

住宅タイプは、以下の5カテゴリで表現される。

持家一戸建	賃貸一戸建	
持家集合住宅	賃貸集合住宅	その他

観測データ集合  $B$  と推定データ集合  $E_j$  は、それぞれ式(1)および式(2)で表される。

$$B = \{\{\bar{a}_i, \bar{h}_i, r_i, c_i, m_i\} \mid 1 \leq i \leq N\} \quad (1)$$

$$\text{where } \bar{a}_i = (a_{i1}, a_{i2}, \dots, a_{i20}), \bar{h}_i = (h_{i1}, h_{i2})$$

$$E_j = \{\{\bar{a}_i^j, \bar{h}_i^j, r_i^j, c_i^j, m_i^j\} \mid 1 \leq i \leq N\} \quad (2)$$

$$\text{where } \bar{a}_i^j = (a_{i1}^j, a_{i2}^j, \dots, a_{i20}^j), \bar{h}_i^j = (h_{i1}^j, h_{i2}^j)$$

ここで、 $N$  は観測データ数である。両集合における要素は 1 世帯分のマイクロデータであり、2つのベクトル  $\bar{a}_i^j, \bar{h}_i^j$  ( $\bar{a}_i, \bar{h}_i$ ) と 3つのスカラー量  $r_i^j$  ( $r_i$ ),  $c_i^j$  ( $c_i$ ),  $m_i^j$

( $m_i$ )の集合である。 $a_{ik}$  は観測データ集合における  $i$  番目の世帯の  $k$  番目のカテゴリに属する世帯構成員の年齢、 $a_{ik}^j$  は  $j$  番目の推定データ集合における  $i$  番目の世帯の  $k$  番目のカテゴリに属する世帯構成員の年齢を表す。当該カテゴリに属する世帯構成員がない場合、 $a_{ik}$  と  $a_{ik}^j$  の値は 999 となる。 $i$  を世帯番号と呼ぶ。

$h_{i1}^j$  は  $i$  番目の世帯の住宅タイプを表す。 $h_{i1}^j$  の値が 0 のときは持家、1 のときは賃貸であることを意味し、 $h_{i2}^j$  の値が 0 のときは一戸建、1 のときは集合住宅であることを意味する。その他の住宅タイプは  $\bar{h}_i^j = (2, 2)$  で表される。 $r_i^j, c_i^j, m_i^j$  はそれぞれ  $i$  番目の世帯の居住ゾーン、自動車保有台数、世帯収入を表す。

例えば、45歳の父親、42歳の母親、15歳の息子、12歳の娘がゾーンNo.3の持家一戸建に住んでおり、2台の自動車を保有して世帯収入が500万円の場合のデータは、以下のように表される。

$$\{(45, 15, 999, 999, 42, 12, 999, \dots), (0, 0), 3, 2, 500\}$$

### (2) 適合度の定義

推定データ集合  $E_j$  の適合度  $Fit(E_j)$  は、5種の距離の重みつき和の最小値で表す。定義を式(3)に示す。

$$Fit(E_j) = \min_{\sigma \in S_N} \left\{ \frac{\bar{w}_a}{N} \cdot \sum_{i=1}^N edis(\bar{a}_i, \bar{a}_{\sigma(i)}^j) + \frac{\bar{w}_h}{N} \cdot \sum_{i=1}^N mdis(\bar{h}_i, \bar{h}_{\sigma(i)}^j) + \frac{\bar{w}_r}{N} \cdot \sum_{i=1}^N tdis(r_i, r_{\sigma(i)}^j) + \frac{\bar{w}_c}{N} \cdot \sum_{i=1}^N ldis(c_i, c_{\sigma(i)}^j) + \frac{\bar{w}_m}{N} \cdot \sum_{i=1}^N ddis(m_i, m_{\sigma(i)}^j) \right\} \quad (3)$$

$$\text{where } w = w_a + w_h + w_r + w_c + w_m, \bar{w}_a = \frac{w_a}{w},$$

$$\bar{w}_h = \frac{w_h}{w}, \bar{w}_r = \frac{w_r}{w}, \bar{w}_c = \frac{w_c}{w}, \bar{w}_m = \frac{w_m}{w}$$

ここで、 $S_n$  は集合  $\{1, 2, \dots, N\}$  から集合  $\{1, 2, \dots, N\}$  へのすべての全単射の集合を表し、 $\sigma(i)$  は全単射  $\sigma$  による  $i$  の像を表す。 $w_a, w_h, w_r, w_c, w_m$  はそれぞれ年齢・性別・世帯主との関係、住宅タイプ、居住ゾーン、自動車保有台数、世帯収入に対する重みである。

年齢・性別・世帯主との関係に関するデータ  $\bar{a}_i, \bar{a}_{\sigma(i)}^j$  の距離  $edis(\bar{a}_i, \bar{a}_{\sigma(i)}^j)$  は、正規化されたユークリッド距離として、式(4)により算出する。計算時のオーバーフローと不在世帯構成員の過度の影響を回避するため、成分の差の二乗に上限値  $D_{\max}$  を設ける。

$$edis(\bar{a}_i, \bar{a}_{\sigma(i)}^j) = \sqrt{\frac{\sum_{k=1}^{20} \min((a_{ik} - a_{\sigma(i)k}^j)^2, D_{\max})}{20 \cdot D_{\max}}} \quad (4)$$

住宅タイプに関するデータ  $\bar{h}_i$ ,  $\bar{h}_{\sigma(i)}^j$  の距離  $mdis(\bar{h}_i, \bar{h}_{\sigma(i)}^j)$  は、正規化されたマンハッタン距離として、式(5)により算出する。

$$mdis(\bar{h}_i, \bar{h}_{\sigma(i)}^j) = \frac{1}{4} \sum_{k=1}^2 |h_{ik} - h_{\sigma(i)k}^j| \quad (5)$$

居住ゾーンに関するデータ  $r_i$ ,  $r_{\sigma(i)}^j$  の距離  $tdis(r_i, r_{\sigma(i)}^j)$  は、最長時間距離に対する2ゾーン間の時間距離の割合として、式(6)により算出する。分子はゾーン  $r_i$  とゾーン  $r_{\sigma(i)}^j$  との間の時間距離であり、分母はゾーン間時間距離の最大値を表す。

$$tdis(r_i, r_{\sigma(i)}^j) = \frac{\text{time - distance } r_i - r_{\sigma(i)}^j}{\max_k (\text{time - distance } r_i - r_k^j)} \quad (6)$$

自動車保有台数に関するデータ  $c_i$ ,  $c_{\sigma(i)}^j$  の距離  $ldis(c_i, c_{\sigma(i)}^j)$  は、式(7)により算出する。

$$ldis(c_i, c_{\sigma(i)}^j) = \frac{|\log(c_i + 1) - \log(c_{\sigma(i)}^j + 1)|}{\max_k \log(c_k + 1)} \quad (7)$$

世帯収入に関するデータ  $m_i$ ,  $m_{\sigma(i)}^j$  の距離  $ddis(m_i, m_{\sigma(i)}^j)$  は、所得差の最大値に対する2世帯間の所得差の割合として、式(8)により算出する。

$$ddis(m_i, m_{\sigma(i)}^j) = \frac{|m_i - m_{\sigma(i)}^j|}{\max_k (|m_i - m_k^j|)} \quad (8)$$

### (3) 適合度算出方法

式(3)による適合度計算は、 $N!$  種類の全単射から、距離和が最小となる全単射を探索する問題である。すなわち、距離和が最小となるように、観測データ集合の各要素を推定データ集合のいずれかの要素と対応付ける組合せ最適化問題といえる。

先行研究<sup>11)~17)</sup>では、最適解探索アルゴリズムの1つである共生進化を適用することで、マイクロデータ推定の評価に有効な適合度を実時間内に算出できることが示されている。共生進化とは、Moriartyらにより提案された遺伝的アルゴリズムの1モデルであり<sup>18)</sup>、部分解を個体とする集団と、部分解の組合せを個体とする全体解集団を保持する点を特徴とする。両集団を並行して進化させることで、1集団を進化させる単純GAよりも多様な解

候補からの探索が可能であり、局所解への収束を回避しつつ、よりよい解を早期に発見することができる。

## 3. 重み決定指標

マイクロデータの推定において各属性ごとに要求される精度は、シミュレーションの目的によって異なると考えられる。特定の属性が適合する度合いは高めたいが、他の属性は多少のズレが生じて構わない場合がある一方、いずれの属性も同程度で適合させたい場合もある。シミュレーションの目的、および各属性を重視する度合いに応じて式(3)における各重みを設定することで、意図に沿った推定データの評価が可能となる。

しかし、何も目安がない状態での重み設定は困難である。各属性が結果に与える影響の大きさは、シミュレーションの内容によって異なると考えられる。対象シミュレーションにおける各属性の影響力を数値化し、重み基準量を定義することで、重みを設定する際の目安を提示する。

### (1) 属性の影響力

本研究で対象としている推定データ集合の適合度評価では、パーソントリップ調査等で観測データ集合が得られていることを前提としている。ある属性に関して観測データ集合の値を変更して改変観測データ集合を作成し、観測データ集合と改変観測データ集合におけるシミュレーション後の政策変数の相違を調べることで、当該属性の対象シミュレーションにおける影響力の大きさを計る。

観測データ集合  $B$  のすべてのデータにおける属性  $z$  の値を変更した改変観測データ集合を  $B_z$  とし、 $B$ ,  $B_z$  のシミュレーション実行後のデータ集合をそれぞれ  $B'$ ,  $B_z'$  とする。また、 $B'$  と  $B_z'$  における政策変数の相違は、0以上1以下の実数値  $diff(B', B_z')$  で表されるものとする。以上の表記を用いて、属性  $z$  のシミュレーションにおける影響力  $att(z)$  を式(9)のように定義する。

$$att(z) = \frac{diff(B', B_z')}{Fit(B_z)} \quad (9)$$

ここで、 $Fit(B_z)$  はすべての重みを1として算出する。式(9)の分子はシミュレーション後の観測データ集合との相違であり、属性  $z$  の値を変更したことによるシミュレーション結果への影響の大きさを示している。また、分母により、基準年における観測データ集合との相違が与える影響を考慮する。

シミュレーションにおいて、評価対象が特定の政策変数でない場合には、式(10)のように、シミュレーション

後用の適合度  $Fit'(B_z)$  を  $diff(B', B_z')$  の代わりに使用する。

$$att(z) = \frac{Fit'(B_z)}{Fit(B_z)} \quad (10)$$

シミュレーション後用の適合度とは、シミュレーションによるデータ数の変動を考慮して算出する適合度である。いずれの世帯データとも距離が最大、すなわち  $edis$ ,  $mdis$ ,  $tdis$ ,  $ldis$ ,  $ddis$  がすべて1となるようなダミーデータで観測データ集合、もしくは改変観測データ集合を補完し、すべての重みを1として式(3)により算出する。

## (2) 重み基準値

各属性の影響力に基づいて、属性  $z$  の重み基準値  $wbase(z)$  を式(11)のように定める。

$$wbase(z) = \frac{1}{att(z) \sum_z att(z)} \quad (11)$$

すべての属性に対する重みを重み基準値と等しくすることで、いずれの属性も同等に考慮する場合の適合度を算出することができる。また、各属性を重視する度合いの比を重み基準値にかけ合わせて重みとすることで、より意図に沿った推定データ集合の評価が可能となる。

## 4. 適用例

本節では、提案手法により各属性の影響度と重み基準値を算出する際の改変観測データ作成手法と対象シミュレーションの例を示す。

### (1) 改変観測データ作成例

観測データ集合  $B$  の各データに基づいて、世帯構成員の年齢と性別、住宅タイプ、居住ゾーン、自動車保有台数、世帯収入に関する改変観測データ集合の生成例を以下に示す。

- ◆ 世帯構成員の年齢と性別
  - 1) 半数のデータにおいて、ランダムに選択した世帯構成員1名の年齢から5を減ずる。
  - 2) 残りの半数のデータにおいて、ランダムに選択した世帯構成員1名の年齢に5を加える。
  - 3) ランダムに選択した世帯構成員の性別を変更する。妻や息子の妻など、性別のみ異なる構成員カテゴリがない場合は、妻を兄弟に、息子の妻を息子にするなど、年齢的に近いカテゴリに変更する。変更により本人・男と兄弟の年齢が逆転したり、変更先のカテゴリが既に埋まっていたりする場合に

は、その他のカテゴリに変更する。

- ◆ 住宅タイプ
  - 1) はじめの4分の1のデータにおいて、 $h_{i1}$  の0と1を反転する。また、 $h_{i1}$  と  $h_{i2}$  の2を0または1に変更する。
  - 2) 次の4分の1のデータにおいて、 $h_{i2}$  の0と1を反転する。また、 $h_{i1}$  と  $h_{i2}$  の2を0または1に変更する。
  - 3) 次の4分の1のデータにおいて、 $h_{i1}$  と  $h_{i2}$  の0と1を反転する。また、 $h_{i1}$  と  $h_{i2}$  の2を0または1に変更する。
  - 4) 最後の4分の1のデータにおいて、 $h_{i1}$  と  $h_{i2}$  の0と1を2に変更し、2を0または1に変更する。
- ◆ 居住ゾーン
 

ランダムに選択した居住ゾーンに変更する。
- ◆ 自動車保有台数
  - 1) 自動車保有台数が1台または2台であるデータからランダムに半数を選択し、自動車保有台数から1を減ずる。残りの半数の自動車保有台数に1を加える。
  - 2) 自動車保有台数が3台以上のデータの自動自動車保有台数を2台とする。
  - 3) 自動車保有台数が0台のデータの自動自動車保有台数を1台とする。
- ◆ 世帯収入
  - 1) 半数のデータにおいて世帯収入を0.8倍する。
  - 2) 残りのデータにおいて世帯収入を1.2倍する。

### (2) 対象シミュレーション例

対象シミュレーションとしては、5年後の世帯を予測するマイクロシミュレーションが挙げられる。文献<sup>19)</sup>で示されているマイクロシミュレーションを使用するマイクロデータに合わせて改変することで適用できる。シミュレーションの内容を以下に示す。

- 1) 世帯および世帯構成員の遷移
 

年齢別事象発生確率に基づき、加齢、死亡、離家、結婚、出産による世帯データの新規作成と削除、世帯構成員の属性値の変更を行なう。
- 2) 住み替えの発生
 

世帯タイプ別住み替え率に基づき、住宅タイプと居住ゾーンを変更する。
- 3) 自動車保有台数の設定
 

自動車保有台数を変更する。
- 4) 世帯収入の設定
 

世帯収入を変更する。

以上のように、特定の政策変数を評価するシミュレーションではなく、将来の世帯状況を予測するシミュレー

ションの場合には、式(10)により影響力を算出する。世帯予測をもとに交通発生量等の政策変数を考える場合には、観測データ集合と改変観測データ集合における政策変数の差を0以上1以下で表現し、式(9)により影響力を算出する。

## 5. おわりに

本研究では、マイクロシミュレーションにおける基準年マイクロデータの推定手法の評価において、重み設定の目安を提示するために、対象シミュレーションにおける各属性の影響力、および重み基準量の算出方法を提案した。各属性のシミュレーション結果に対する影響の大きさを考慮することで、目的に応じた推定データの評価が実現される。また、各属性の影響力を数値化することで、土地利用を考えるにあたって重視すべき要素や、将来に大きな変化をもたらす要因など、新たな知見を得ることができる。

**謝辞：**本論文は、平成23～25年度科学研究費補助金(基盤研究(B), 課題番号: 23360228, 研究課題名: 縮退状況における都市マネジメントのための世帯マイクロシミュレーションシステム)の研究成果の一部を取りまとめたものである。ここに記して謝意を表したい。

## 参考文献

- 1) Wegener, M. : Overview of land-use transport models, Proc. of CUPUM'03, 2003.
- 2) 宮本和明, 北詰恵一, 鈴木温 : 世界における実用都市モデルの実態調査とその理論・機能と適用対象の体系化, 平成 18~19 年度科学研究費補助金(基盤研究(C), 課題番号:18560524)研究成果報告書, 2008.
- 3) Pritchard, D. R. and Miller, E. J. : Advances in agent population synthesis and application in an integrated land use / transportation model, Proc. of 88th TRB'09, 2009.
- 4) 杉木直, 宮本和明, Varameth VICHIANSEN : 土地利用マイクロシミュレーションにおける初期マイクロ世帯データの推定手法, 第 39 回土木計画学研究発表会論文集, 2009.
- 5) Miyamoto, K. and Sugiki, N. : An estimation method of household micro-data for the base year in land-use micro simulation, Proc. of CUPUM'09, 2009.

- 6) 杉木直, 宮本和明, 大谷紀子, Varameth VICHIANSEN : 質的属性を含む初期マイクロ世帯データの推定手法, 第 40 回土木計画学研究発表会論文集, 2009.
- 7) Miyamoto, K., Sugiki, N., Otani, N. and Vichiensan, V. : Agent-based estimation method of household microdata for base year in land use microsimulation, Proc. of 89th TRB'10, 2010.
- 8) 杉木直, 宮本和明, 大谷紀子, Varameth VICHIANSEN : 居住属性を含む初期マイクロ世帯データの推定手法, 第 41 回土木計画学研究発表会論文集, 2010.
- 9) Miyamoto, K., Sugiki, N., Otani, N., and Vichiensan, V. : An agent based estimation method of household micro-data including housing information for the base year in land-use microsimulation, Proc. of 12th WCTR'10, 2010.
- 10) 杉木直, 宮本和明, 大谷紀子, Varameth VICHIANSEN : 総合的属性からなる世帯マイクロデータの推計方法, 第 42 回土木計画学研究発表会論文集, 2010.
- 11) 大谷紀子, 杉木直, 宮本和明 : 土地マイクロシミュレーションにおける観測マイクロデータ集合と推定集合の適合度評価, 第 39 回土木計画学研究発表会論文集, 2009.
- 12) Otani, N., Miyamoto, K. and Sugiki, N. : Goodness-of-fit evaluation method between observed and estimated sets of micro-data in land-use micro-simulation, Proc. of CUPUM'09, 2009.
- 13) 大谷紀子, 杉木直, 宮本和明 : 土地利用マイクロシミュレーションにおける質的属性を含むマイクロデータの適合度評価, 第 40 回土木計画学研究発表会論文集, 2009.
- 14) 大谷紀子, 杉木直, 宮本和明 : 居住情報を含む世帯マイクロデータの推定集合の適合度評価, 第 41 回土木計画学研究発表会論文集, 2010.
- 15) Otani, N., Sugiki, N. and Miyamoto, K. : Goodness-of-fit evaluation method between two sets of household micro-data for land-use microsimulation models, Proc. of 12th WCTR'10, 2010.
- 16) 大谷紀子, 杉木直, 宮本和明 : 総合的属性からなる世帯マイクロデータ推計集合の適合度評価, 第 42 回土木計画学研究発表会論文集, 2010.
- 17) Otani, N., Sugiki, N. and Miyamoto, K. : Goodness-of-Fit Evaluation Method of Agent-Based Household Micro-Data Sets Composed of Generalized Attributes, Proc. of 90th TRB'11, 2011.
- 18) Moriarty, D. E. and Miikkulainen, R. : Efficient reinforcement learning through symbiotic evolution, Machine Learning, Vol.22, pp.11-32, 1996.
- 19) Sugiki, N. and Miyamoto, K. : Spatio-temporal aggregation effects and path-dependence in a land-use micro-simulation system, Proc. of 10th WCTR'04, 2004.