

# アクティビティベースドシミュレーションのための社会経済属性データ作成手法の提案

三谷 卓摩<sup>1</sup>・Thaithatkul Phathinan<sup>2</sup>・日下部 貴彦<sup>3</sup>

<sup>1</sup>正会員 東京大学空間情報科学研究センター特任助教（〒277-8568 千葉県柏市柏の葉5-1-5）

E-mail: mitani@csis.u-tokyo.ac.jp

<sup>2</sup>正会員 東京大学空間情報科学研究センター客員研究員（〒277-8568 千葉県柏市柏の葉5-1-5）

E-mail: t.phathinan@csis.u-tokyo.ac.jp

<sup>3</sup>正会員 東京大学空間情報科学研究センター講師（〒277-8568 千葉県柏市柏の葉5-1-5）

E-mail: t.kusakabe@csis.u-tokyo.ac.jp

本研究は、東京都市圏を対象とした活動ベースのエージェントシミュレーションのための社会経済属性データの作成手法を提案する。従来、交通シミュレーションにおける人口分布および初期の交通需要は、パーソントリップ調査などの約2kmのゾーンベースの行動データから作成されていた。しかしながら、MaaSやライドソーシングといった新たなモビリティサービスが必要とする短トリップや公共交通機関の乗降を分析するにはサイズが大きすぎるといった問題がある。そこで、本研究では500mグリッドの国勢調査と経済センサスにパーソントリップ調査データを組み合わせて、IPF法に基づいたマイクロレベルの社会経済属性データの作成手法を提案する。その結果、東京都市圏という大都市圏を対象にしたうえで、属性別の割合や従業地分布、外出時間分布について、実測値に近い社会経済属性を付与するシミュレーションを行うことが可能となった。

**Key Words :** Activity Model, Person Trip Survey, Iterative proportional fitting

## 1. はじめに

本研究は、東京都市圏を対象とした活動ベースのエージェントシミュレーションのための社会経済属性データの作成手法を提案する。従来、交通シミュレーションにおける人口分布および初期の交通需要は、パーソントリップ調査などの約2kmのゾーンベースの行動データから作成されている。しかしながら、MaaSやライドソーシングといった新たなモビリティサービスに必要とする短トリップや公共交通機関の乗降を分析するにはサイズが大きすぎると考えられる。そこで、本研究では500mグリッドの国勢調査と経済センサスにパーソントリップ調査データを組み合わせた社会経済属性データの作成手法を提案する。

## 2. 本研究の位置づけ

Iterative proportional fitting (IPF)法はDemingら<sup>1)</sup>によって提案された人口データ作成のための一般的な方法である。特徴として、複数の属性間の相関関係を維持したまま各属性の周辺分布に一致するように再帰的に計算すること

で、拡大係数を算出することができる。この特徴を生かした方法として、西田ら<sup>2)</sup>はパーソントリップ調査と国勢調査の結果を世帯単位で拡大する方法を用いている。また、倉内ら<sup>3)</sup>は、属性に世帯だけでなく、世帯と個人の双方を考慮することができるIPU法を用いて、拡大を行っている。これらの研究では、ゾーンサイズのパーソントリップ調査をそのまま拡大しており、グリッド単位の全数調査の周辺分布に対して行っていない。さらに、IPF法の問題点として、個人属性を対象にして拡大を行った場合、世帯構成まで明示的に考慮できないこと、多次元になった場合ゼロセルが多数発生するなどの問題が指摘されている。そこで、本研究では、パーソントリップ調査のデータを用いて、ゾーン単位の世帯構成生成モデルを構築し、個人属性を考慮しつつ、世帯構成も考慮できるようなIPF法を緩和した方法を用いて計算を行うことを試みる。

## 3. 作成手法の提案

本研究では、社会経済属性データを作成するにあたり、パーソントリップ調査、国勢調査、経済センサスのデー

表-1 入力データの特徴

	サンプル調査 (約2%)	全数調査 (100%)	
	パーソントリップ 調査(PT)	国勢調査	経済センサス
世帯構成	○	-	-
エリアの個人属性 (性別,年代,就業状態)	同時分布	周辺分布	周辺分布 (性別)
居住地	Zone (~5km圏)	Grid (500m)	-
居住地と従業地	Zone (~5km圏)	-	-
従業地	Zone (~5km圏)	-	Grid (500m)
外出時間	○	-	-

タセットを使用する。パーソントリップ調査は、表-1に示すように、人口の約2%から得られた1日のトリップパターンを含む世帯単位のサンプル調査である。一方、国勢調査は、すべての住民を対象とした全数調査となっている。経済センサスは、事業所および企業を対象とした全数調査で、各業種の事業所数、従業者数などが取得される。これらのセンサスデータでは、各500mグリッドの社会経済属性別の周辺分布が含まれている。ただし、匿名性を保つために、このデータには複数の属性の同時分布は含まれていない。

これらの入力データを用いた作成手法は、国勢調査の居住地をベースに、世帯構成、従業地、外出時間を推定する3つのステップで構成する。

### (1) 世帯構成の推定

最初のステップでは、パーソントリップ調査と国勢調査を入力データとして、世帯構成を推定する。提案手法の特徴は、世帯内人数に対応する属性の条件付き確率関数を推定する世帯構成生成モデルを構築することである。このモデルは、属性に基づいて、最も年上の世帯構成員から最も年下の世帯構成員までを順番に決定する。

まず、ゾーンを  $z \in Z$ 、ゾーン  $z$  中のグリッドを  $g \in G_z$ 、社会経済属性のタイプを  $k \in K$  とする。国勢調査のデータ集合は、グリッド  $g$  ごとの属性  $k$  の属性値を  $s_k \in S_k$  で表し、その人口を  $n_{gs_k}$  とする。グリッド  $g$  ごとの人口の集合は、 $N_g = \{n_{gs_k} | s_k \in S_k, k \in K\}$  で表される。 $S_k$  は、属性  $k$  の上限値であり、 $s$  は離散値である。データ集合は、グリッド  $g$  ごとの世帯内人数が  $l = |u_z|$  である世帯数  $n_{gl}$  も含む。パーソントリップ調査は、情報  $\theta_{ki} \in \theta_i$  または、ゾーン  $z$  の世帯  $u \in U_z$  中の個人  $i \in u$  ごとにのすべての属性も含んでいる。以上の条件のもとに、世帯構成員の条件付き確率を推定する。

そして、確率の初期値は、パーソントリップ調査から導出する。最も年上の世帯構成員のある属性別の確率は、

$$p_{g_0}^{(0)}(\theta_{g_{age}(0,u)}|l)$$

帯構成員の確率は、 $p_{gj}^{(0)}(\theta_{g_{age}(j,u)}|l, \theta_{k^*g_{age}(j-1,u)})$  のように定義される。ここで、 $g_{age}(j,u)$  は、世帯  $u$  の  $j$  番目に年上の構成員である。これらの同時確率は、グリッド  $g$  が属しているゾーンで観測されたパーソントリップ調査の結果から導出される。すべての世帯構成員の同時確率は以下のように記述することができる。

$$\delta(\theta_i, s_k) = \begin{cases} 1 & \text{if } \theta_{ik} = s_k \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

$U$  は、世帯構成のすべての可能なパターン集合とする。世帯構成パターン  $u \in U$  のいずれかが与えられた場合、メンバーのいる世帯で  $j^*$  番目の世帯構成員が属性  $\theta_{g_{age}(j^*,u)}$  を持つ確率は次のように記述できる。

$$p_{g_{age}(j^*,u)}(\theta_{g_{age}(j^*,u)}|l) = \sum_{s_{k^*j^*|j'=0 \in S_{k^*}}} \dots \sum_{s_{k^*j^*|j'=1-1 \in S_{k^*}}} p_{0(0,u)}(\theta_{g_{age}(0,u)}|l) \prod_{j^* \geq j > 0} \delta(\theta_{g_{age}(j-1,u)}, s_{k^*j'}) p_{g_{age}(j,u)}(\theta_{g_{age}(j,u)}|l, s_{k^*j'}) \quad (2)$$

式(2)からグリッド  $g$  の属性  $s_k$  を持つ人数は次のように表すことができる。

$$N_{gs_k} = \sum_{u \in U} \sum_{0 \leq j < |u|} \delta(\theta_{g_{age}(j,u)}, s_k) p_{g_{age}(j,u)}(\theta_{g_{age}(j,u)}|lu) \quad (3)$$

反復計算の重みパラメータは式(4)のように計算することができる。

$$w_{gs_k}^{(\eta)} = \frac{n_{gs_k}}{N_{s_k}^{(\eta)}} \text{ for } s_k \in S_k, k \in K \quad (4)$$

それぞれの確率は式(5)(6)のように重みを付けて正規化して更新される。

$$p_{g_0}^{(|K|\eta+k)}(\theta_{g_{age}(0,u)}|l) = \delta(\theta_{g_{age}(0,u)}, s_k) w_{gs_k}^{(\eta)} p_{g_0}^{(|K|\eta+k-1)}(\theta_{g_{age}(0,u)}|l) \quad (5)$$

$$\text{for } s_k \in S_k, k \in K, u \in U$$

$$p_{gj}^{(|K|\eta+k)}(\theta_{g_{age}(j,u)}|l, \theta_{k^*g_{age}(j-1,u)}) = w_{gs_k}^{(\eta)} p_{gj}^{(|K|\eta+k-1)}(\theta_{g_{age}(j,u)}|l, \theta_{k^*g_{age}(j-1,u)}) \quad (6)$$

$$\text{for } s_k \in S_k, k \in K, u \in U$$

以後、式(3)から式(6)を収束条件 $\eta = H$ を満たすまで反復計算し、世帯構成生成モデルの確率関数を推定する。

## (2) 従業地の推定

2 番目のステップでは、パーソントリップ調査の居住地と従業地の同時分布と、国勢調査の居住地、経済センサスの従業地を考慮して、IPF モデルを用いて従業地を推定する。対象は従業者のみを対象とし、男女別で計算する。まず、居住者数と従業者数は調査が異なるため全数を調整する。

$$W = \frac{N_p}{N_e} \quad (7)$$

$N_p$ は国勢調査の総人口であり、 $N_e$ は経済センサスの総人口である。 $W$ は、経済センサスの結果を国勢調査の結果に合わせるための補正係数である。そして、この補正係数を用いて、経済センサスの従業地グリッド $g$ の従業者数 $N_{eg}$ を補正し、 $N_{eg}^*$ を求める。

$$N_{eg}^* = N_{eg}W \quad (8)$$

居住地 $r \in Z$ に住んでいる人々の間で、従業地グリッド $g \in G_z$ の割合は式(5)で表される。 $p_{zr} = p(z|r)$ は、パーソントリップ調査から導出される居住地ゾーン $r \in Z$ の人が従業地ゾーン $z \in Z$ である確率で、 $p_{gz} = p(g|z)$ は、ゾーン $z$ でのグリッド $g \in G_z$ の人口比で計算される確率である。

$$p_{gr}^{(0)} = p_{gz}p_{zr} \quad (9)$$

ここで、 $\tilde{N}_g^{(\eta)}$ はグリッド $g \in G_r$ の従業者数であり、この値は、 $N_{ph}$ は国勢調査のグリッド $h \in G_r$ の居住者数と、 $p_{gr}^{(\eta)}$ との積から計算できる。

$$\tilde{N}_g^{(\eta)} = p_{gr}^{(\eta)} N_{ph} \quad (10)$$

そして、重みパラメータ $w_g^{(\eta)}$ は式(5)のように $N_{eg}^*$ との比により計算することができる。

$$w_g^{(\eta)} = \frac{N_{eg}^*}{\tilde{N}_g^{(\eta)}} \quad (11)$$

重みパラメータを用いて、確率 $p_{gr}^{(\eta+1)}$ は、つぎのように更新される。

$$p_{gr}^{(\eta+1)} = w_g^{(\eta)} p_{gr}^{(\eta)} \quad (12)$$

以後、式(10)から式(12)を収束条件 $\eta = H$ を満たすまで反復計算し、 $p_{gr}^{(H)}$ を求める。

## (3) 外出時間の推定

3番目のステップでは、社会経済属性を考慮した外出時間の分布を推定する。外出時間は、労働時間と通勤時間の合計値として推定する。外出時間、居住地、従業地、個人属性の同時分布を用いることはPTのサンプル数から考えて困難であることから、外出時間と居住地、外出時間と従業地、外出時間と個人属性の3つの同時分布とIPF法を用いることで収束計算を行う。

$d \in D$ を30分間隔で離散化される外出時間とする。

$p_{dr} = p_r(d|r)$ および $p_{dz} = p_z(d|z)$ は、それぞれ居住地および従業地に対応する外出時間の分布である。

$p_{d\theta} = p(d|\theta)$ は、属性が $\theta$ である個人の外出時間の分布である。これらの変数はすべて、パーソントリップ調査びトリップデータから計算できる。 $N_{hg\theta}^*$ は、グリッド $z$ の従業地が $h$ で属性が $\theta$ の従業者数である。

この数は、最初と2番目のステップの推定結果から導出される。これらの値に基づいて、従業地、居住地、個人属性の外出時間に対応する人数は、 $N_{dh} = \sum_{g \in G_z, z \in Z} N_{hg\theta}^* p(d|z)$ と $N_{dz} = \sum_{h \in G_z, z \in Z} N_{hg\theta}^* p(d|r)$ と $N_{d\theta} = \sum_{h \in G_z, z \in Z} \sum_{g \in G_z, z \in Z} N_{hg\theta}^* p(d|\theta)$ として導出される。

そして、社会経済属性に対応する各グリッドの外出時間分布の初期値は次のようになる。

$$p_d^{(0)}(d|h, g, \theta) = p(d|\theta) \quad (13)$$

確率関数 $p_d^{(3\eta+1)}(d|h, g, \theta)$ は、対応する居住地の周辺分布によって更新する。

$$\tilde{N}_{dh}^{(\eta)} = \sum_{\theta \in \Theta} \sum_{g \in G_z, z \in Z} N_{hg\theta}^* p_d^{(3\eta)}(d|h, g, \theta) \quad (14)$$

$$w_{dh}^{(\eta)} = \frac{N_{dh}}{\tilde{N}_{dh}^{(\eta)}} \quad (15)$$

$$p_d^{(3\eta+1)}(d|h, g, \theta) = w_{dh}^{(\eta)} p_d^{(3\eta)}(d|h, g, \theta) \quad (16)$$

同様の方法で、対応している従業地の確率を更新する。

$$\tilde{N}_{dg}^{(\eta)} = \sum_{\theta \in \Theta} \sum_{h \in G_z, z \in Z} N_{hg\theta}^* p_d^{(3\eta+1)}(d|h, g, \theta) \quad (17)$$

$$w_{dg}^{(\eta)} = \frac{N_{dg}}{\tilde{N}_{dg}^{(\eta)}} \quad (18)$$

$$p_d^{(3\eta+2)}(d|h, g, \theta) = w_{dg}^{(\eta)} p_d^{(3\eta+1)}(d|h, g, \theta) \quad (19)$$

こちらも同様に、対応している属性によって確率を更新する。

$$\tilde{N}_{d\theta}^{(\eta)} = \sum_{h \in G_z, z \in Z} \sum_{g \in G_z, z \in Z} N_{hg\theta}^* p_d^{(3\eta+2)}(d|h, g, \theta) \quad (20)$$

$$w_{d\theta}^{(\eta)} = \frac{N_{d\theta}}{\tilde{N}_{d\theta}^{(\eta)}} \quad (21)$$

$$p_d^{(3(\eta+1))}(d|h, g, \theta) = w_{d\theta}^{(\eta)} p_d^{(3\eta+2)}(d|h, g, \theta) \quad (22)$$

以後、式(14)から式(22)を収束条件 $\eta = H$ を満たすまで反復計算し、 $p_d^{(H)}(d|h, g, \theta)$ を求める。この推定方法により最終的に従業者の外出時間を推定することができる。

**(4) 推定による出力結果**

この3つの推定を実施することにより、以下の社会経済属性データを出力結果として得ることができる。従業者は、個人属性/居住地/従業地/外出時間を、非従業者は個人属性/居住地を推定することができる。これらの出力結果は、アクティビティベースドシミュレーションの入力データとして利用可能である。

**4. 東京都市圏への適用事例**

前章で示した作成手法を用いて、東京都市圏への適用を行う。最初に使用するデータについて説明し、そのあと適用結果とその考察について示す。

**(1) 使用データ**

表-2に使用データの概要を示す。対象地域は、東京都市圏である。東京都市圏は、第5回東京都市圏パーソントリップ調査をしている東京都、神奈川県、埼玉県、千葉県、茨城県南部の一部の238市区町村1682ゾーンを対象とする。ゾーンに対応する2分の1地域メッシュ(約500m)は、41,833グリッドとなっている。

表-2 使用データの概要

対象地域	<ul style="list-style-type: none"> <li>・エリア 東京都市圏（東京都、神奈川県、埼玉県、千葉県、茨城県南部の一部）</li> <li>・ゾーン:238 市区町村 1682 ゾーン</li> <li>PT 調査小ゾーン</li> <li>・グリッド:41,833 Grids</li> <li>2分の1地域メッシュ(約500m):</li> </ul>
パーソントリップ調査	<ul style="list-style-type: none"> <li>第5回東京都市圏パーソントリップ (調査時期:平成20年10月から11月)</li> <li>・対象人数:578,849</li> <li>・世帯属性 (世帯内人数, 居住地) 世帯内人数:最大7人世帯まで</li> <li>・個人属性 (性別, 年代, 就業状態, 従業地)</li> <li>・外出時間</li> <li>トリップデータから勤務および業務活動時間と通勤時間からの和から外出時間を計算し, 30分単位, 48区間で算出</li> <li>・エリア単位:ゾーン</li> <li>・調査手法:サンプル調査</li> </ul>
国勢調査	<ul style="list-style-type: none"> <li>平成22年国勢調査</li> <li>対象世帯数:16,555,354 世帯</li> <li>対象人数:37,225,544 人</li> <li>世帯属性 (世帯人数, 居住地) 世帯数:最大7人世帯まで</li> <li>個人属性 (性別, 年代, 就業状態)</li> <li>・エリア単位:グリッド</li> <li>2分の1地域メッシュ(約500m):</li> <li>・調査手法:全数調査</li> </ul>
経済センサス	<ul style="list-style-type: none"> <li>平成24年経済センサス</li> <li>・個人属性 (性別, 従業地)</li> <li>※従業者のみ対象</li> <li>・エリア単位:グリッド</li> <li>2分の1地域メッシュ(約500m):</li> <li>・調査手法:全数調査</li> </ul>

つぎに、パーソントリップ調査は平成20年の第5回調査を用いる。外出時間は、トリップデータから勤務時間および業務時間と通勤時間の和から計算し、30分単位で、48区間の結果になっている。国勢調査は平成22年国勢調査を使用しており、グリッド単位の全数調査となっている。経済センサスは、平成24年経済センサス-活動調査を使用しており、こちらもグリッド単位の全数調査となっている。

**(2) 適用結果と考察**

前節で示した東京都市圏のデータを適用して推定を行った。結果を以下に示す。図-1に東京都市圏での属性別の実測値と推定値を示す。全体およびその他属性別のシミュレーションの推定値も国勢調査の実測値に近い値を示していることがわかる。ただし、世帯内人数が7名の場合のみ推定値が少ない値となっている。これは、世

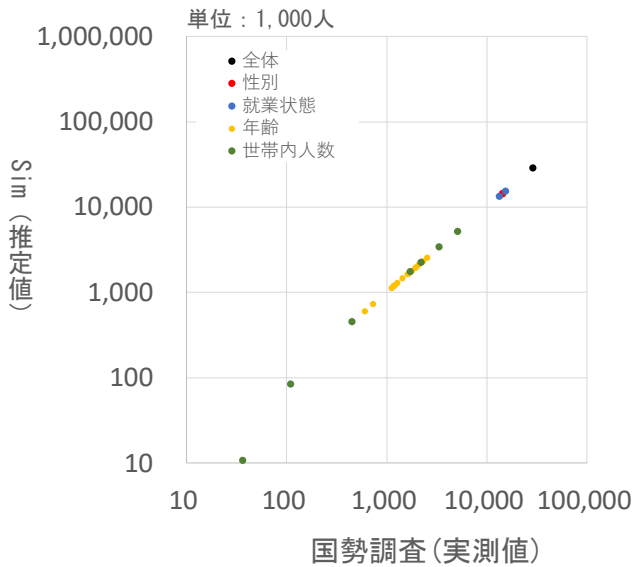


図-1 東京都市圏の属性別実測値と推定値

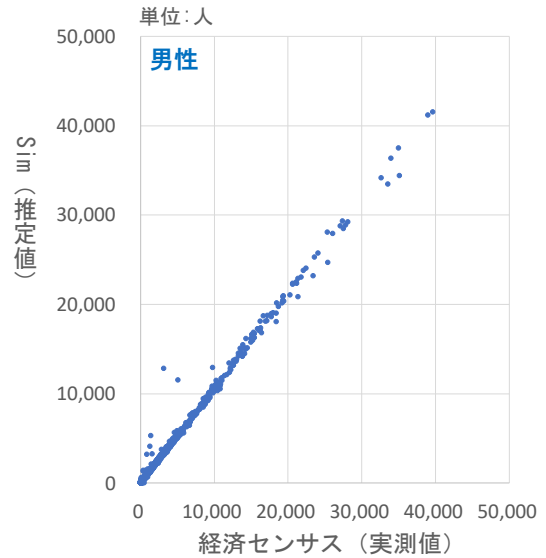


図-3 男性のグリッド別の従業者数の実数値と推計値

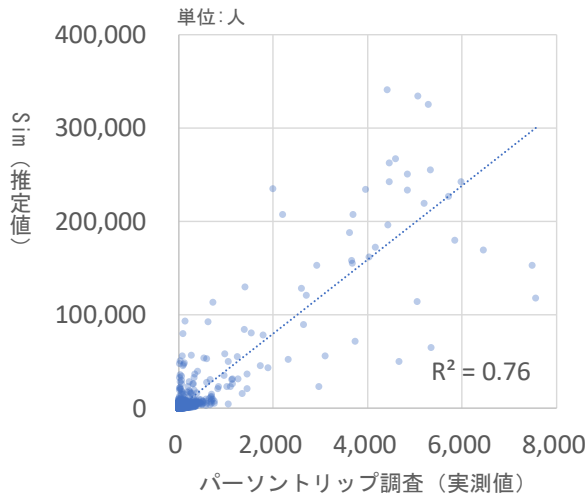


図-2 東京都市圏の世帯構成パターン別実測値と推定値

帯内に7人いるパーソントリップ調査の居住地が必ずしも多くなく、数値がゼロの場合、シミュレーションで生成することができなかつたためと考えられる。

図-2に世帯構成パターン別の実測値と推定値を示す。世帯構成のパターンは、性別（男性、女性）と年代(0-19, 20-39, 40-59, 60-の4つの年代)と5人以下の世帯を対象として組み合わせで分類し、同一の世帯構成パターンとなるものをそれぞれ集計し、どの程度の再現性があるのかを比較した。これは、我々が提案した世帯構成を生成するモデルがどの程度実測に近いものになっているかを示すグラフである。その結果、決定係数は0.76となり、ある程度のばらつきがみられたことから必ずしも再現性が高いとは言えない結果となった。

図-3に男性のグリッド別の従業者数の実数値と推計値を、図-4に女性のグリッド別の従業者数の実測値と推定値を示す。実測値と推定値を比較するとどちらも再現性

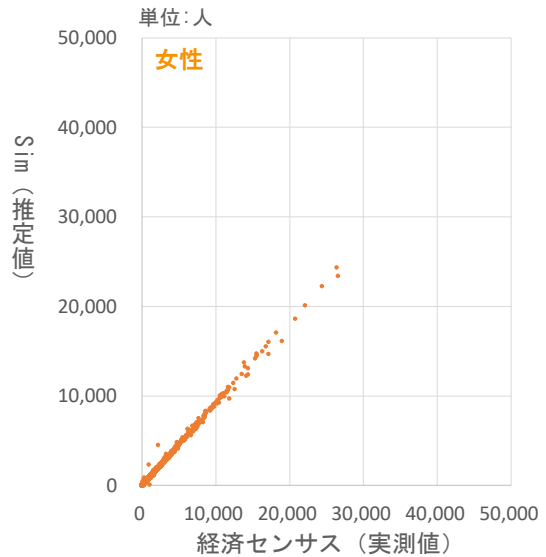


図-4 女性のグリッド別の従業者数の実数値と推計値

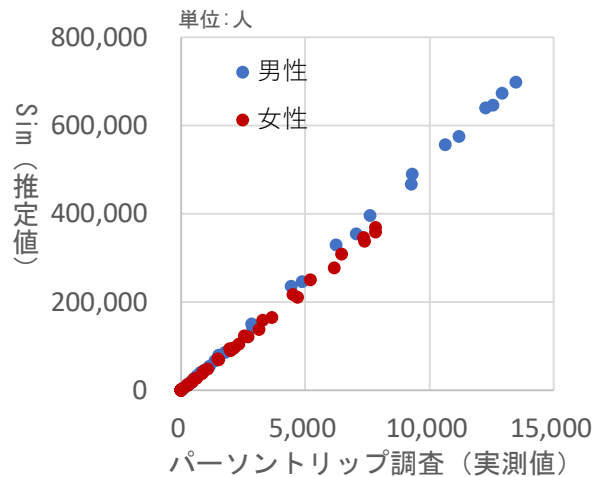


図-5 男女別外出時間別の従業者数のPT 実測値と推定値

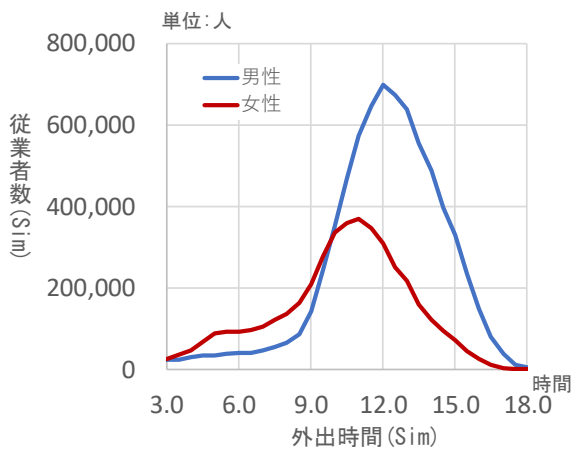


図-6 男女別外出時間分布の推定値

の高い値を示している。従業地の推定が従業地のばらつきを再現できていると考えられる。つぎに、性別の違いで比較すると、男性は少し多く、女性は少し少なく出る傾向があることがわかった。図-5に男女別外出時間別の従業者数のPT実測値と推定値を示す。実測値と推定値を比較すると性別の違いなく外出時間の長さに関わらず、再現性の高い値を示している。

図-6に男女別外出時間分布の推定値を示す。男女で外出時間の長さが異なり、男性では、外出時間が12時間あたりがピークとなり分散が小さいこと、それに対して、女性では、外出時間が10時間あたりが大きい、分散も大きいことがわかる。

## 5. まとめ

本研究では、アクティビティベースドシミュレーションのための社会経済属性データの作成手法について提案し、東京都市圏への適用を行った。その結果を以下にまとめる。

- ・ 既存のパーソントリップ調査、国勢調査、経済センサスを組み合わせることで、世帯構成、従業地、外出時間の3つのステップからなる世帯と個人属性を考慮した社会経済属性データを推定する手法について提案を行った。
- ・ 属性別の割合や従業地分布、外出時間分布について、東京都市圏という大都市圏において適用し、実測値に近い社会経済属性を付与するシミュレーションを行うことができた。

さらに、今後は、以下に示すようないくつかの課題について、取り組むことも考えたい。

- ・ 提案手法の評価には、世帯構成パターンをどの程度精度よく再現しているかを検証する必要がある。そこで、パーソントリップ調査を全数調査、パーソントリップ調査の小ゾーンをグリッド、パーソントリップ調査の大ゾーンをゾーン、パーソントリップ調査をサンプリングしたものをパーソントリップ調査と見立てて、提案手法の精度検証をおこなう。
- ・ IPF法では、対象とする属性のサンプルがゼロの場合、拡大することができないゼロセル問題が発生することから、その影響について把握する。
- ・ 提案手法の活用方法は、アクティビティベースドシミュレーションの入力データである。提案手法の出力結果を効率よく利用できるアクティビティモデルを構築し、拠点開発、シェアリング、Maasなどの交通施策の導入に対する新たなアクティビティの発生を表現できるようにする。

謝辞：本研究の実施にあたり、国土交通省から「東京都市圏パーソントリップ調査に係る調査票情報」を提供していただいた。ここに感謝の意を表す。

## 参考文献

- 1) Deming, W. E. and Stephan, F. F.: On a least squares adjustment of a sampled frequency table when the expected marginal totals are known, *Annals of Mathematical Statistics*, Vol.11, pp.427-444, 1940.
- 2) 西田悟史, 山本俊行, 藤井聡, 北村隆一: 非集計交通需要分析のための将来世帯属性生成システムの構築, 土木計画学研究・論文集, No.17, pp.779-787, 2000.
- 3) 倉内慎也, 萩尾龍彦, 石村龍則, 吉井稔雄: 世帯及び個人属性分布を考慮した PT 調査データの拡大係数算出手法の適用, 土木学会論文集 D3 (土木計画学), Vol.67, No.5, pp. I\_759-I\_767, 2011.

(2019.10.4 受付)

## SOCIOECONOMIC MICRODATA FOR ACTIVITY-BASED AGENT SIMULATION IN THE GREATER TOKYO AREA

Takuma MITANI, Phathinan THAITHATKUL, Takahiko KUSAKBE