

# 基盤地図情報による合成した世帯構成への位置情報の付加

○原田拓弥 村田忠彦（関西大学総合情報学研究科）

## Addition of Location Attribute to Household Composition Using Fundamental Geospatial Data

\*T. Harada and T. Murata (Department of Informatics, Kansai University Graduate School)

**概要**— 本研究では、複数の統計データを用いて合成した世帯構成に位置情報を付加する手法を提案する。より具体的な社会シミュレーションを行うためには、現実のデータを用いることが望まれるが、現実のデータは個人情報であるため安易に用いることはできない。これまでの世帯合成手法の研究において、世帯構成員が保持する属性は、年齢、性別、所属する家族類型、世帯の役割、親族関係の5つであった。様々なモデルにおいて合成した世帯構成を活用するためには、他の属性を付加しなければならない。本研究では、合成された世帯の属性に位置情報を付加するために、3つの手法を提案する。町丁目の属性を付加する際の誤差は多いものの、合成した世帯構成に建築物が存在する位置情報の属性を付加した結果を示す。

**キーワード**: 世帯合成手法, 統計データ, シミュレーティッドアニーリング法, 位置情報, 緯度経度.

### 1 はじめに

計算科学は、理論科学、実験科学と並ぶ第3の科学として認められるようになってきている<sup>1)</sup>。物理学や化学、天文学などの分野で計算科学は利用されているが、近年では、経済学や政治学、社会学の分野においても、社会シミュレーションという形で活用されるようになってきている。人間をモデルに内包する社会シミュレーションは、物理学や化学などと比較して、不確定要素が多く、モデルの妥当性を確認する上で課題が残る場合が少なくないが、多様なシミュレーションを行うことにより、極端な事例（シナリオ）を発見し、それらの極端な事例から、避けなければならないシナリオや、望ましいシナリオを抽出し、それらのシナリオの発生メカニズムをモデルの範囲内で調べることにより、実世界で起こりうる現象の説明ができることが期待される。

社会シミュレーションの技法の一つにマイクロシミュレーション<sup>2)</sup>やエージェントシミュレーション<sup>3)</sup>があり、それらのシミュレーションモデルでは、モデル化する社会における市民の属性を個別に生成することが必要になる。単純化した人工社会における市民の生成の際には、市民の生成においても、単純化された方法を用いて生成されることが多いが、モデルの対象となる社会をより現実的なものにしようとする場合、その人工社会における市民の属性も、実際の社会における市民の属性と同様のもことになることが期待される。本研究では、政府が公開する統計データを用いて合成された世帯群を地図上に展開する手法を提案する。

市民の属性の再現を行うにあたって、政府や行政が収集している戸籍や納税のデータを用いることができれば、それらの情報源に格納されている情報については、正確に再現することができるが、市民のプライバシーへの配慮から、それらの情報源が個票レベルで公開されることはない。また、仮にそれらの情報源が試用可能な場合でも、そのデータを直接用いることにより、ある個人のシミュレーション結果が特定されないように配慮しなければならない。このような状況から、公開されている統計データから、仮想的な属性を持つ市民で構成される人工社会を生成し、その人工社会の中でどのような事象が発生するかを観察する社会シミュレーションが行われるようになってきている。

レーションが行われるようになってきている。

統計データに基づく、個々の市民の個票データの合成に関する研究の歴史は古く、Synthetic Reconstruction method (SR法)<sup>4)</sup>として知られている。SR法は、個票データのサンプルをもとに、Iterative Proportional Fitting Procedure (IPFP)<sup>5)</sup>を用いて個票データを合成している。その後、数多くの個票データ合成法が提案されているが、基本的にSR法に基づく、個票データのサンプルを用いたアルゴリズムとなっている。Barthelemyら<sup>6)</sup>は、IPFPの弱点として、個人の統計と世帯の統計のどちらかに適合する合成ができたとしても、両方に適合する合成が困難であることを指摘している。この課題を解決するため、GargiuloらやBarthelemyら<sup>7)</sup>は、サンプルを用いない合成手法を提案している。LenormandとDeffuanは<sup>8)</sup>、サンプルを用いない合成手法とSR法と比較し、前者が個人と世帯をよりよく合成できていることを示した。

サンプルを用いない手法として、池田ら<sup>9)</sup>は9種類の家族類型を対象に、SA法（Simulated Annealing法）を用いて9つの統計データへの適合を目的とした世帯集団を合成する手法を提案した。家族類型とは、一般世帯<sup>1</sup>を世帯員と世帯主の続柄により区分した分類である。柘井ら<sup>10, 11, 12)</sup>は、池田らが提案した合成手法を基に目的関数や合成手法の改良を行った。池田らや柘井らは日本全国の統計データを縮小し、500世帯や1,000世帯の世帯構成の合成を行った。一方、著者ら<sup>13)</sup>は対象地域と同じ規模の世帯構成の合成方法を提案した。

しかし、これらの世帯構成の合成方法において、付加する市民の属性は年齢、性別、所属する家族類型、世帯の役割、親族関係の5つの属性であった。これらの属性のみを用いた社会シミュレーションは限られている。合成データをより多くの社会シミュレーションに活用するためには、合成データが保持する属性の種類を増やさなければならない。

本研究では従来の合成データに世帯が住む位置情報の属性を付加する手法を提案する。提案手法では、市

<sup>1</sup>一般世帯とは、住居と生計を共にしている人の集まりや一戸を構成して住んでいる単身者、間借り、下宿、独身寮に住んでいる単身者の世帯である。

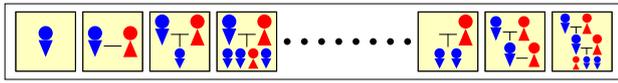


Fig. 1: 合成データのモデル

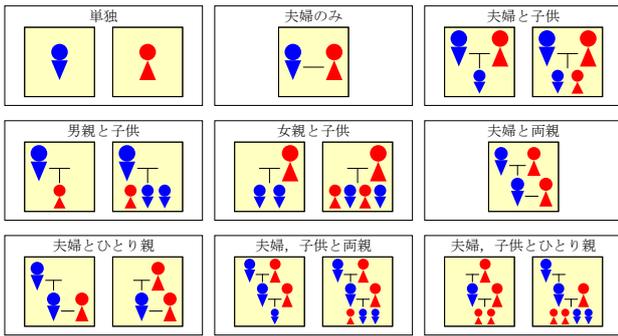


Fig. 2: 合成対象の家族類型

区町村を対象に世帯構成を合成し、各世帯がどの町丁目に属するか SA 法を用いて属性を付加する。町丁目の属性を付加した後、町丁目毎の建築物データを割り当てる。建築物データは現実世界の建築物が存在する位置の緯度経度情報である。これらの操作により合成データの各世帯に位置情報の属性を付加する。

## 2 SA 法を用いた世帯構成合成手法

著者が提案した世帯合成手法<sup>13, 14)</sup>は、統計データを基に作成した仮想の世帯構成を、複数の統計に適合させる手法である。市民の年齢や親子の年齢差の統計データに対する、コンピュータ上で再現した世帯構成のデータ集合(合成データ)の誤差を計算し、SA 法を用いて誤差を最小化している。

合成データは複数の世帯とその構成員である市民によって構成しており、世帯数  $H$  を設定することで合成データの規模が設定する。合成データのモデルを Fig. 1 に示す。著者の手法は、山形県を対象に、山形県の統計データに基づく世帯構成の合成を試みており、家族類型は実統計<sup>15)</sup>に掲載されている Fig. 2 の 9 種類の家族類型を用いている。なお、これらの 9 種類の世帯数で日本全体の全世帯数の約 95%、山形県については約 90% を占めていることが知られている。その他の世帯数の家族類型は、他の親族を含む世帯や兄弟姉妹のみからなる世帯、非親族を含む世帯、他に分類されない世帯である。これらの家族類型は、公開されている複数の統計データに整合させることが困難である。例えば、他の親族を含む世帯において、親族と他の親族を結びつける統計がない。兄弟姉妹のみからなる世帯については、兄弟間の年齢差を設定することは可能だが、基準となる年齢を結びつける統計が存在しない。そのため、著者は 9 種類の家族類型のみを対象とした。

世帯の中には構成員として世帯の構成員が存在している。それぞれの市民は年齢、性別、所属する家族類型、世帯の役割、親族関係の 5 つの属性を持っている。世帯の役割は、所属する世帯の中での役割を表している。例えば、「夫婦と子供」の世帯の場合、夫、妻、子供の 3 つの役割がある。親族関係は親一子と夫一妻の相互関係を表すための属性であり、一般的な家系図で記述される人と人を繋ぐ線を表現している。例えば、自分の妻の属性を確認する時は、この親族関係を用いて

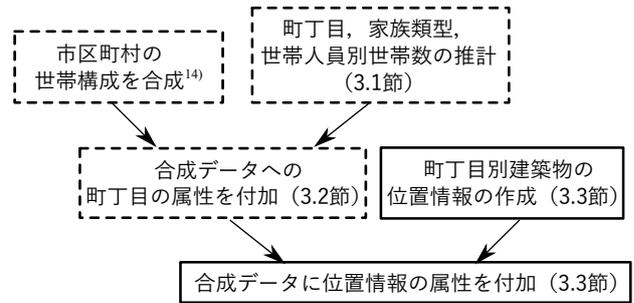


Fig. 3: 位置情報を付加するフロー

妻の情報を参照する。

著者らは、対象地域と同じ規模(以下、実世界スケール)の世帯構成を合成するために、対象地域の世帯数や人口などの統計データ通りに初期世帯を生成する。著者らの初期世帯合成法<sup>13, 14)</sup>は、家族類型、世帯人員別に世帯を合成し、その後市民の性別と初期の年齢を設定する手法である。家族類型、世帯人員別に世帯を合成する際には、国勢調査人口等基本集計<sup>15)</sup>表 7、表 11、表 16-1 を用いて推計した、家族類型、世帯人員別世帯数のデータを用いる。表 7 は 10 区分の世帯人員別世帯数が、表 11 は 16 区分の世帯の家族類型、7 区分の世帯人員別世帯数が記されている。しかし、表 11 には年齢不詳の単独世帯が含まれているため、表 16-1 を用いて、年齢不詳の単独世帯を取り除いた。表 16-1 は 16 区分の世帯の家族類型、3 区分の配偶関係、1 歳階級の年齢、男女別一般世帯人員が記されている。実世界スケールの世帯の合成方法についての詳細は文献<sup>14)</sup>を参照されたい。

## 3 合成世帯への位置情報の属性の付加

本章では合成データに位置属性を付加する手法を説明する。位置属性の付加には 3 つの手法を用いる。合成データに位置属性を付加するフローを Fig. 3 に示す。Fig. 3 の破線により囲まれた手法は SA 法を用いて探索を行う手法である。本手法では、SA 法を 3 段階にわたって適用し、各世帯に位置情報を付加する。市区町村の合成データに位置情報の属性を付加するために、合成データに町丁目の属性を付加する。すなわち、合成された各世帯がどの町丁目に所属し、その後、割り当てられた町丁目の中のどの座標に位置づけられるかを示す。町丁目毎に世帯数や人口が異なるため、「町丁目、家族類型、世帯人員別世帯数」のデータを推計する。「町丁目、家族類型、世帯人員別世帯数」のデータの推計手法を 3.1 節で説明する。推計した「町丁目、家族類型、世帯人員別世帯数」と町丁目の人口分布を用いて合成データに町丁目の属性を付加する手法を 3.2 節で説明する。合成データに位置情報の属性を付加する手法及び、その際に用いる町丁目別建築物の位置情報のデータを作成する手法を 3.3 節で説明する。

### 3.1 町丁目、家族類型、世帯人員別世帯数の推計

本節では市区町村を対象に合成した世帯構成に対して、町丁目の属性を付加する際に必要となる、「町丁目、家族類型、世帯人員別世帯数」のデータを SA 法を用いて推計する手法を説明する。町丁目の統計は国勢調査小地域集計により公開されているが、個人を特定できないように粗い統計のみ公開されている。世帯数に

Table 1: 町丁目  $p$  における世帯数と人口

条件 $X$	世帯数	人口
単独	404	404
夫婦のみ	164	328
夫婦と子供	147	525
男親・女親と子供	39	90
核家族以外の世帯	18	73
非親族を含む世帯	5	12
不詳	1	2

関する統計として、市区町村単位では「家族類型、世帯人員別世帯数」の統計が公開されているが、小地域集計では「町丁目、家族類型別世帯数」や「町丁目、世帯人員別世帯数」と複数の統計に分割されている。

本研究では、これらの統計データに整合的な町丁目単位の家族類型、世帯人員別世帯数を推計する。推計には以下の統計データを用いる。

- 町丁目、世帯人員別世帯数  
(国勢調査 小地域集計<sup>16)</sup> 表 3)
- 町丁目、家族類型別世帯数  
(国勢調査 小地域集計<sup>16)</sup> 表 6)
- 町丁目、家族類型別人口  
(国勢調査 小地域集計<sup>16)</sup> 表 6)

町丁目、世帯人員別世帯数の世帯人員数は  $1, 2, \dots, 7$  人以上である。町丁目  $p$  における世帯数と人口を Table 1 に示す。核家族以外の世帯とは、国勢調査 人口等基本集計<sup>15)</sup> 表 11 に記載されている「II 核家族以外の世帯」の 10 種類である。

「町丁目、家族類型、世帯人員別世帯数」の推計を行う SA 法では、以下の目的関数を用いて最適化を行う。

$$g_s = \sum_{j=1}^{G_s} \sum_{p=1}^T |d_{sj}^p - R_{sj}^p| \quad (1)$$

ここで、 $G_s$  は統計データ  $s$  の項目数、 $T$  は町丁目数、 $d_{sj}^p$  は統計  $s$ 、項目  $j$  における町丁目  $p$  において、条件  $X_{sj}$  を満たす推計データの市民の数もしくは世帯数、 $R_{sj}^p$  は統計データ  $s$  の項目  $j$  における町丁目  $p$  の実数である。上記の目的関数を統計データ  $s = 1, 2, \dots, S$  に対して行い、その総和の最小化を SA 法を用いて行う。

「町丁目、家族類型、世帯人員別世帯数」の推計は、以下の手続きで行う。なお、以下の  $H_{t,m}^p$  とは、町丁目  $p$  の家族類型  $t$  の世帯人員数  $m$  の世帯数である。

**Step 1** 推計データを初期生成

**Step 2** 探索回数が規定数に達すれば探索を終了

**Step 3**  $H_{t,n}^p \neq 0$ ,  $H_{t,m}^q \neq 0$  の条件を満たす町丁目  $p, q$ 、家族類型を 1 つ ( $t$ )、世帯人員数を 2 つ ( $m, n$ ) ランダムに選択する。

**Step 4** 以下の手続きより解を更新する。

- $H_{t,m}^p \leftarrow H_{t,m}^p + 1$
- $H_{t,n}^p \leftarrow H_{t,n}^p - 1$
- $H_{t,m}^q \leftarrow H_{t,m}^q - 1$
- $H_{t,n}^q \leftarrow H_{t,n}^q + 1$

なお、 $\leftarrow$  は代入である。

**Step 5** 解の遷移判定

Table 2: 町丁目  $p$  における年齢、男女別人口

条件 $X$	男性	女性
0～4 歳	41	32
5～9 歳	22	28
...		
95～99 歳	0	1
100 歳～	0	0

**Step 6** 探索回数を更新して SA の温度を冷却

**Step 7** Step 2 の処理に戻る

Step 1 の推計データの初期生成では、国勢調査 人口等基本集計<sup>15)</sup> 表 11 を基に推計した、市区町村の「家族類型、世帯人員別世帯数」のデータを用いる。各町丁目  $p$  に市区町村の「家族類型、世帯人員別世帯数」のデータを Table 1 の 7 つの家族類型別世帯数を満たす世帯数となるようにランダムに配分する。また、Step 3 において選択する家族類型は Table 1 の 7 つの家族類型である。

### 3.2 合成データへの町丁目の属性の付加

前節において推計した「町丁目、家族類型、世帯人員別世帯数」のデータを用いて合成データの各世帯に町丁目の属性を付加する。町丁目の属性を設定するには以下の統計を用いる。

- 町丁目別男性の人口分布 (5 歳階級)  
国勢調査 小地域集計<sup>16)</sup> 表 3
- 町丁目別女性の人口分布 (5 歳階級)  
国勢調査 小地域集計<sup>16)</sup> 表 3

町丁目別人口分布を Table 2 に示す。

市区町村単位で合成された各世帯に町丁目の属性の付加する SA 法では、以下の式を用いる。

$$f_s(A) = \sum_{j=1}^{G_s} \sum_{p=1}^T |c_{sj}^p(A) - \text{Round}(r_{sj}^p \times m_{sj}^p(A))| \quad (2)$$

なお、 $A$  は合成データ、 $c_{sj}^p$  は推計データにおける統計  $s$ 、項目  $j$  における町丁目  $p$  において、条件  $X_{sj}$  を満たす推計データの市民の数もしくは世帯数、 $m_{sj}^p$  は統計データ  $s$  の条件  $X_{sj}$  を満たす合成データの市民もしくは世帯数である。また、 $r_{sj}^p$  は統計データ  $s$  の項目  $j$  における町丁目  $p$  の割合であり、以下の式を用いて統計データから算出する。

$$r_{sj}^p = R_{sj}^p / \sum_{k=1}^{G_s} R_{sk}^p \quad (3)$$

また、以下の手続きにより最適化を行う。

**Step 1** 個々の世帯に町丁目の属性を設定

**Step 2** 探索回数が規定数に達すれば探索を終了

**Step 3**  $H_{t,m}^p \neq 0$ ,  $H_{t,m}^q \neq 0$  の条件を満たす町丁目  $p, q$ 、家族類型を 1 つ ( $t$ )、世帯人員数を 1 つ ( $m$ ) ランダムに選択する。

**Step 4** 町丁目  $p, q$  について、家族類型  $t$ 、世帯人員  $m$  を満たす 1 世帯をそれぞれ選択し、所属する町丁目を入れ替える。

**Step 5** 解の遷移判定

Table 3: 基盤地図情報における建築物の種類

種類	説明 <sup>19)</sup>
普通建物	3階相当未満の建物並びに3階以上の木造建築の建物
堅ろう建物	3階相当以上 60m 未満の非木造建物
高層建物	高さが 60m 以上の非構造建物
普通無壁舎	地上 3 階相当未満の壁がない建物
堅ろう無壁舎	地上 3 階相当以上の壁がない建物
総描建物	建物が密集している場合、個別の建物の区別が困難な場合に総合して取得された建物 (25,000 分の 1 の縮尺のみ)
その他 不明	競技場や野球場など特殊な建築物 上記に分類が不可能な建築物

**Step 6** 探索回数を更新して SA の温度を冷却

**Step 7** Step 2 の処理に戻る

Step 1 の町丁目の属性の初期設定では、前節で推計した「町丁目、家族類型、世帯人員別世帯数」のデータに基づき、合成データの各世帯に町丁目をランダムに割り当てる。

### 3.3 合成データに位置情報の付加

町丁目の属性が付加された合成データに位置情報を付加する際には、国土地理院 基盤地図情報<sup>17)</sup>を用いる。基盤地図情報とは、地理空間情報を効率的に利活用するために整備された電子地図における位置の基準である。基盤地図情報では以下の項目が整備されている。

- 測量の基準点
- 海岸線
- 公共施設の境界線 (道路区域界)
- 公共施設の境界線 (河川区域界)
- 行政区画の境界線及び代表点
- 道路線
- 軌道の中心線
- 標高点
- 水涯線
- 建築物の外周線
- 市町村の町若しくは字の境界線及び代表点
- 街区の境界線及び代表点

基盤地図情報の提供範囲は 25,000 分の 1 の縮尺の場合、日本全国を網羅している。また、基盤地図情報では 2,500 分の 1 の縮尺におけるデータも公開している。2,500 分の 1 の縮尺における提供範囲は国土地理院 基盤地図情報サイト<sup>17)</sup>の基盤地図情報の整備状況より確認できる。基盤地図情報のデータ形式は GML である。一般的な GIS (Geographic Information System; 地理情報システム) ソフトで利用されている Shape 形式のデータは提供されていない。しかし、基盤地図情報を Shape 形式に変換する機能を持つ基盤地図情報ビューアが国土地理院 基盤地図情報サイト<sup>17)</sup>で公開されている。

本研究では、基盤地図情報の建築物の外周線のデータを用い、各世帯に建築物の位置情報を付加する。基盤地図情報の建築物の外周線のデータはメッシュ単位で提供されている。町丁目の属性を付加した合成デー

Table 4: 大阪府高槻市の合成結果

シード	誤差	処理時間 (秒)
1	30	17,072.98
2	36	16,304.97
3	22	17,381.28
4	36	16,172.43
5	30	16,979.44
6	24	16,336.99
7	30	17,289.77
8	20	16,632.09
9	16	17,203.38
10	36	16,295.65
平均	28	16,766.90
標準偏差	7.18	468.10

タに位置情報を付加するために、メッシュ単位のデータを町丁目単位に変換する。本研究では町丁目単位の建築物の外周線のデータを作成するために、e-Stat 上で公開されている市区町村毎の境界データ<sup>18)</sup>を用いる。このデータは対象の市区町村について、町丁や字などの地域単位の境界データである。この境界は国勢調査 小地域集計<sup>16)</sup>表 2 以降に用いられている町丁・字等別集計と同じ境界である。メッシュ単位から町丁目単位に変換する際には、境界データと建築物の内外判定により、建築物がどの町丁目に属するか判定した。

基盤地図情報の建築物の外周線のデータでは、建築物の緯度経度の他に建築物の種類<sup>19)</sup>のデータが付加されている。付加されている建築物の種類を Table 3 に示す。本研究では、合成した世帯構成に建築物の位置情報を付加する。付加する位置情報は住宅であることが望ましい。そのため、世帯に付加する建築物の種類として、普通建物、堅ろう建物及び高層建物の 3 種を用いた。また、町丁目  $p$  に属する各世帯に位置情報を付加する際は、町丁目  $p$  に属する建築物の中から重複を許してランダムに 1 軒選択した。重複を許した理由は、共同住宅が存在する町丁目について、建築物数 < 世帯数となる町丁目が存在するからである。

## 4 実験結果

本研究では、大阪府高槻市を対象に、世帯構成の合成と位置情報の付加を行う。高槻市の合成世帯数は 139,749 世帯、その人口は 336,891 人である。世帯構成の合成を 10 回行ったときの統計との誤差と処理時間を Table 4 に示す。探索回数は 1 人あたり 100,000 回、初期温度を 0.5、収束温度を 0.1 と設定し、冷却関数は指数冷却を用いた。なお、本研究が用いた計算機の CPU は、Intel Core i7 3930K (3.2 GHz, 6 コア) で、メモリは DDR3-1600 8 GB×8 である。

Table 4 のシードは、乱数を異ならせて合成したときのシードを示している。シード 1、シード 5、シード 7 の誤差は全て 30 である。同じ誤差を持つ合成データがどの程度類似しているか、世帯構成が完全に一致している世帯数を調べた。世帯構成の完全一致とは、2 つの世帯について、家族類型、世帯人員数、世帯構成員の年齢、性別、世帯内の役割が全て同一である状態を指す。シード 1 とシード 5 の合成データについて、単独世帯を除く完全一致した世帯数は 100,318 世帯中 53,744 世帯、シード 1 とシード 7 については 53,940 世帯、シ

Table 5: 大阪府高槻市における町丁目、家族類型、世帯人員別世帯数の推計結果

	誤差	処理時間 (秒)
平均	51.2	617.76
標準偏差	10.08	31.79

Table 6: 町丁目の属性の付加結果

合成データ シード	誤差	平均	標準偏差
1	30	4,528.8	40.09
2	36	4,549.7	53.51
3	22	4,512.9	50.31
4	36	4,530.5	38.01
5	30	4,551.0	33.61
6	24	4,525.7	36.75
7	30	4,552.6	40.96
8	20	4,531.1	40.48
9	16	4,511.6	37.63
10	36	4,537.2	29.08
全試行		4,533.1	41.20

ド5とシード7については54,026世帯であった。また、54,000前後の完全一致した世帯構成のうち、約60%の32,000~33,000世帯は夫婦のみ世帯であった。これらの結果から、同じ誤差となる試行において、約半数は異なる世帯構成となっている。

次に、「町丁目、家族類型、世帯人員別世帯数の推計」の推計結果 ( $\sum_s g_s$ ) を Table 5 に示す。大阪府高槻市の町丁目の数  $T$  は 384 である。なお、探索回数は 1 世帯あたり 1,000 回、初期温度を 1.0、収束温度を 0.1 と設定し、冷却関数は指数冷却を用いた。Table 5 から、町丁目あたりの誤差は 0.133 である。

次に、合成データへの町丁目の属性の付加結果を Table 6 に、処理時間を Table 7 に示す。探索回数は 1 世帯あたり 1,000 回、初期温度を 1.0、収束温度を 0.1 と設定し、冷却関数は指数冷却を用いた。また、実験は 10 個の合成データに対して、それぞれ 10 回、合計 100 回行った。

Table 6 から、合成データの差異による町丁目の属性の付加する際の統計との誤差の影響は少ない。Table 6 について、Welch の  $t$  検定を行ったところ、5%の有意差があった試行は、シード5とシード9 ( $p$  値 = 0.0239) 及びシード7とシード9 ( $p$  値 = 0.0317) であった。

合成データに町丁目の属性を付加した際の誤差の傾向を調べるために、男性の人口分布に対する誤差の 100 回平均を Fig. 4 に、女性の人口分布に対する誤差の 100 回平均を Fig. 5 に示す。集計では、ある試行に対する人口分布の項目  $j$  について全ての町丁目の誤差を、式 (2) の  $c_{s_j}^p(B) - \text{Round}(r_{s_j}^p \times m_{s_j}^p(B))$  の符号別に総和した。すなわち、Fig. 4, Fig. 5 の各項目はプラスの誤差とマイナスの誤差が相殺した結果ではない。

Fig. 4 と Fig. 5 から項目  $j$  の誤差がプラスかマイナス、どちらかに偏っている。そのため、2つの町丁目に属する世帯を交換する SA 法ではこれ以上の最適化は困難である。偏りが発生した原因として、合成データは 9 種類の家族類型を対象にしたが、町丁目別人口分布は全ての人口を対象にしている。そのため、合成データに含まれない家族類型の人口分布に偏りが存在

Table 7: 町丁目の属性の付加する際の処理時間

処理時間 (秒)	
平均	591.77
標準偏差	37.15

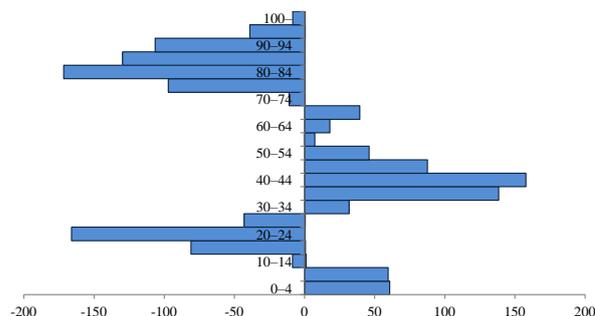


Fig. 4: 男性の人口分布の符号別誤差

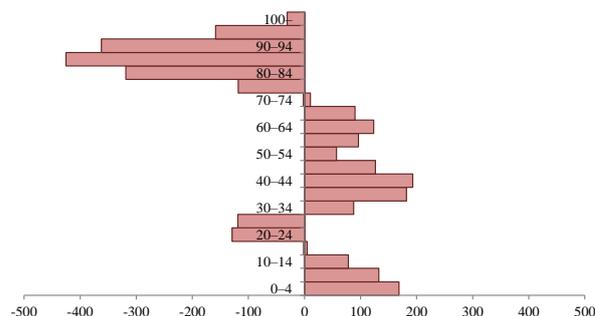


Fig. 5: 女性の人口分布の符号別誤差

する。Fig. 4 から、9 種類の家族類型に属する男性は 10 歳未満と 30 歳から 69 歳の人口が多い。Fig. 5 から、9 種類の家族類型に属する女性は 20 歳未満と 30 歳から 74 歳の人口が多い。

最後に、合成データに位置情報を付加した結果を Fig. 6 に示す。Fig. 6 の赤い円は世帯の位置、灰色の多角形は壁がある建築物、灰色の破線は無壁舎の外周線である。なお、Fig. 6 の灰色の多角形及び灰色の破線は国土地理院の基盤地図情報の建築物の外周線を、町名及び境界線は e-Stat 地図で見る統計 (統計 GIS) を用いた。本研究では、町丁目  $p$  の建築物に対して一様に世帯を割り当てた。そのため、Fig. 6 の多くの建築物に対して世帯の割り当てが可能である。

1 軒の建築物に対して何世帯割り当てたか調べたところ、多くの町丁目において数世帯未満であった。しかし、1 軒に 10 世帯以上割り当てた地域があった。1 軒に 10 世帯以上割り当てた地域の例を Fig. 7 に示す。Fig. 7 は牧田町 (Fig. 6 の南) の北西部である。Fig. 7 の灰色の多角形は壁がある建築物、灰色の破線は無壁舎の外周線、数値は建築物に割り当てた世帯数である。なお、Fig. 6 と同様に、Fig. 7 の灰色の線は国土地理院の基盤地図情報の建築物の外周線を、町名及び境界線は e-Stat 地図で見る統計 (統計 GIS) を用いた。

Fig. 7 から一軒家と思われる南西の建築物に 10 世帯前後割り当てている。牧田町は多くの共同住宅が存在する町丁目である。Fig. 6 南の細長い建築物は共同住宅である。本研究では建物に対して一様に世帯を割り当てたため、対象となる町丁目の世帯数に対して建物の数が少ない場合、一軒家に多くの世帯を割り当てる



Fig. 6: 位置情報の付加結果 (34.822746, 135.603156 付近)

この地図は国土地理院の基盤地図情報の建築物の外周線を使用したものである

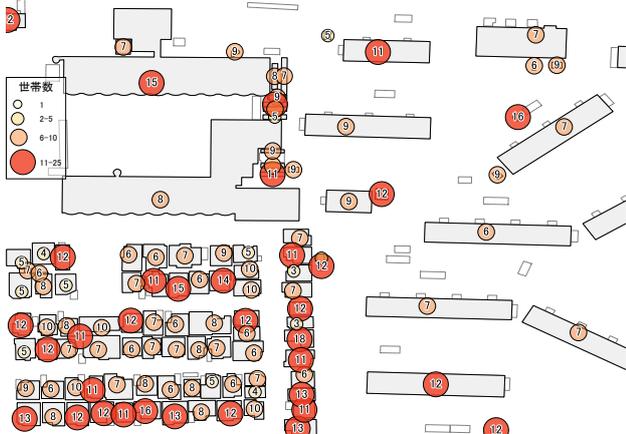


Fig. 7: 建築物毎に割り当てた世帯数 (34.821779, 135.599901 付近)

この地図は国土地理院の基盤地図情報の建築物の外周線を使用したものである

結果となった。

## 5 おわりに

本研究では合成データに位置情報の属性を付加する手法を提案した。提案手法では合成データに町丁目の属性をSA法を用いて付加し、町丁目別の建築物の位置情報を用いて合成データに位置情報の属性を付加した。実験結果から、「町丁目、家族類型、世帯人員別世帯数」の推計する際の誤差は良好な結果が得られた。しかし、世帯に町丁目の属性を付加する際には多くの誤差が発生し、提案手法では誤差の削減が困難である。また、合成対象の人口が増加すると共に処理時間が増加すると考えられる。これは、2つの世帯が所属する町丁目を交換する手法において、合成対象の人口が増加すると解空間が複雑になるからである。

本研究では、町丁目の属性を付加した合成データに位置情報を付加する際に、建物に対して一様に割り当てた。そのため、共同住宅が多い地域において、一軒家に複数の世帯を割り当てる結果となった。このような合成データを用いて社会シミュレーションを行うと、シミュレーション結果について誤った解釈をする危険がある。例えば、共同住宅に住む世帯が少なく、一軒家に複数の世帯が住む合成データを用いて避難のシミュレーションを行う場合、シミュレーション結果から得

られる避難経路の混雑状況と実際の混雑状況が異なる可能性がある。そのため、世帯に位置情報を付加する手法の改良が必要である。

## 参考文献

- 1) 小柳義夫: 計算科学とシミュレーション <http://olab.is.s.u-tokyo.ac.jp/~oyanagi/reports/cs-and-sim.txt> (2003)
- 2) 矢田晴那: 政策分析ツールとしてのマイクロ・シミュレーション, ファイナンス, 35/40 (2010)
- 3) 山影進: 社会科学とマルチエージェントシミュレーション—シミュレータ開発と事例提供の課題—, 情報科学, 27, 1/10 (2007)
- 4) A. G. Wilson, C. E. Pownall: A new representation of the urban system for modeling and for the study of micro-level interdependence, *Area*, 8-4, 246/254 (1976)
- 5) W. E. Deming, F. F. Stephan: A least squares adjustment of a sampled frequency table when the expected marginal totals are known, *The Annals of Mathematical Statistics*, 11, 428/444 (1940)
- 6) J. Barthélemy, P. L. Toint: Synthetic population generation without a sample, *Transportation Science*, 266/279 (2012)
- 7) F. Gargiulo, S. Ternes, S. Huet, G. Deffuant: An iterative approach for generating statistically realistic populations, of households, *PLoS One*, 5-1, 266/279 (2010)
- 8) M. Lenormand, G. Deffuant: Generating a synthetic population of individuals in households: Sample free vs sample-based methods, *Journal of Artificial Societies and Social Simulation*, 16-4, 1/9 (2013)
- 9) 池田, 喜多, 薄田: 地域人口動態シミュレーションのためのエージェント推計手法, 計測自動制御学会第43回システム工学部会研究会, 11/14 (2010)
- 10) 柘井, 村田: SAを用いた統計データからのエージェント属性復元のための目的関数の影響, 計測自動制御学会第5回社会システム部会研究会, 121/126 (2014)
- 11) 柘井, 村田: 統計データとの誤差最小化のためのSAによるエージェント属性復元, 計測自動制御学会第7回社会システム部会研究会, 47/52 (2014)
- 12) 柘井, 村田: エージェント属性復元における Simulated Annealing を用いた世帯構成の最適化, 計測自動制御学会第8回社会システム部会研究会, 167/172 (2015)
- 13) T. Murata, T. Harada, D. Masui: Modified SA-based Household Reconstruction from Statistics for Agent-Based Social Simulations, 2016 IEEE International Conference on Systems, Man, and Cybernetics, 3600/3605 (2016)
- 14) 原田, 村田: 家族類型と世帯内の役割を考慮したSA法による日本全体の世帯の合成, 計測自動制御学会第12回社会システム部会研究会, 6 pages (2017)
- 15) 厚生労働省: 平成22年度国勢調査 人口等基本集計 全国結果, [http://www.e-stat.go.jp/SG1/estat/GL08020103.do?\\_toGL08020103\\_&tclassID=000001034991](http://www.e-stat.go.jp/SG1/estat/GL08020103.do?_toGL08020103_&tclassID=000001034991) (2011)
- 16) 総務省統計局: e-Stat 平成22年度国勢調査 小地域集計 27 大阪府, [http://www.e-stat.go.jp/SG1/estat/GL08020103.do?\\_toGL08020103\\_&tclassID=000001036634](http://www.e-stat.go.jp/SG1/estat/GL08020103.do?_toGL08020103_&tclassID=000001036634) (2012)
- 17) 国土交通省 国土地理院: 基盤地図情報サイト, <http://www.gsi.go.jp/kiban/>
- 18) 総務省統計局: e-Stat 地図で見る統計 (統計GIS), <http://www.e-stat.go.jp/SG1/estat/toukeiChiri.do?method=init>
- 19) 国土地理院: 電子国土基本図 地図情報 ファイル仕様書 第1.0版, <http://www.gsi.go.jp/common/000093949.pdf> (2012)