

# 日本全国の位置情報付き仮想の個票合成手法の精緻化

原田 拓弥<sup>1)</sup>、村田 忠彦<sup>2)</sup>

<sup>1)</sup> 青山学院大学 理工学部 経営システム工学科

<sup>2)</sup> 関西大学 総合情報学部 総合情報学科

## 1. はじめに

災害対策や経済対策などの政策決定において、全体的な影響だけでなく、その政策の個々の市民への影響を見る必要性が高まっている。その方法の 1 つとして社会シミュレーションへの関心が集まっている [1]。特に個々の市民のシミュレートを可能にする Agent-Based Simulation (以下、ABS) [2] がよく利用されている。ABS は、意思決定主体 (人や組織、国など) がもつ特徴を属性や振る舞いとして保持するエージェントを定義し、他のエージェントや環境との相互作用による振る舞いをマクロ・ミクロに観察する手法である。ABS を用いることにより、得られたシミュレーション結果が発生する原因や、エージェントの振る舞い及び環境の変化が双方に与える影響の分析が可能となる。

これまでの多くの社会シミュレーションではモデルを単純化せよという “Keep It Simple, Stupid” (以下、KISS 原理) [3] に基づいてモデル化されていた。KISS 原理に基づくモデルは、少数のパラメータからなるシンプルなモデルであり、社会現象の原理的な理解が主な目的である [4]。したがって、特定の地域・事例・状況などの分析を目的としたシミュレーションを行うためには、一般的な帰結から具体事例へのさらなる推論が求められる。また、現実の社会現象と KISS 原理に基づくモデルを用いたシミュレーションには乖離があり [5]、乖離を埋めるために、シミュレーション結果の補完が必要である [6]。このような状況から、現実社会の複雑な現象を理解するためには、KISS 原理に基づくモデルを逸脱する必要がある [7]。そのため、特定の地域・事例・状況に関するパラメータをもつモデルを用いた社会シミュレーションが期待されている。このようなモデルにおいて、モデルの粒度を現実社会に近づけるほ

ど、エージェントの意思決定においても、可能な限り現実の意思決定主体と同様の振る舞いを模倣する必要がある。

現実社会を模倣するモデル上で、エージェントが現実の意思決定主体を模倣した意思決定をモデルに組み込むためには様々な課題がある。その課題の 1 つがエージェントの保持する属性の設定である。エージェントの属性の設定に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、現実社会と整合するエージェントの属性の設定が可能であるが、これらの個人の情報は個人情報保護やプライバシーの観点から利活用が困難である。

このような状況から、政府統計をはじめとする利活用可能な統計情報から、仮想的な属性を持つ個人で構成される人工社会 (以下、仮想都市) を合成し、その仮想都市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いる [8-10]。市川ら [8] は東京都大島町を対象とし、感染症拡大のシミュレーションを実施するに当たり、500m 四方の人口分布や従業員数などの情報を用いて仮想都市を構築している。花岡 [9] は京都市の京町家の取り壊しの分析に当たり、パーソナリティー調査や国勢調査などの情報を用いて仮想都市を構築している。杜ら [10] は日本全国の公的年金制度の所得代替率の考察に当たり、100 分の 1 スケールの仮想都市を合成した。その際に、1 歳階級の人口分布を用いて各エージェントの年齢や性別などの属性を設定した。

これらの研究のように、特定の状況をパラメータとしてもつモデルを用いてシミュレーションを行う場合、研究者はシミュレーションモデルだけでなく、モデルを適用する仮想都市の構築に取り組む必要がある。特に、仮想都市の構築において、エージェン

トが保持する属性の数が増加するほど、妥当な仮想都市の構築が困難となる。仮想都市の構築が研究者の負担となり、新規参入の障壁となっている。

これらの問題を解決するために、統計情報を用いた仮定の個人の属性（以下、仮想個票）を合成する研究が行われている。仮想個票の合成に関する研究の歴史は古く、Synthetic Reconstruction method（以下、SR 法）[11]として知られている。その後、数多くの仮想個票を合成する手法が提案されているが、基本的に SR 法に基づく手法である。Barthelemy ら[12]は、SR 法の弱点として、個人の統計表と世帯の統計表のどちらかに適合する仮想個票の合成ができたとしても、両方に適合する仮想個票の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[13]や Barthelemy ら[12]は、個票データのサンプルを用いない合成手法を提案している。Lenormand ら[14]は、SR 法と個票データのサンプルを用いない手法と比較し、前者がよりよく仮想個票を合成できていることを示した。

これらの海外の研究では、それぞれの国において利活用可能な統計情報や世帯構成の特徴に基づいた手法が開発されている[13]。日本の統計表を用いた手法として花岡[15]は個票データのサンプルを用いた手法を提案している。一方、著者ら[16, 17]は個票データのサンプルを用いずに、探索手法の 1 つである Simulated Annealing（以下、SA 法）を用いた手法を提案している。著者らが合成した仮想個票の例を図 1 に示す。図 1 では世帯の属性として、家族類型・所属する市区町村及び町丁目・居住地の位置情報・住宅所有の関係・住宅の建て方属性を、各構成員は年齢・性別・世帯内の役割・就業状態・産業分類・就業形態・企業規模・所得属性を保持している。

本研究では、日本全国、約 1700 市区町村において、従来手法[16]より精緻な仮想個票を 10 セット作成合成する。従来手法は都道府県単位で仮想個票を合成する手法を市区町村単位で適用していた。人口 20 万人未満の市町村では、個人が特定されないように、人口 20 万人以上の市区町村と比べ粗い統計情報のみ公開されている。人口 20 万人未満の市町村を合成

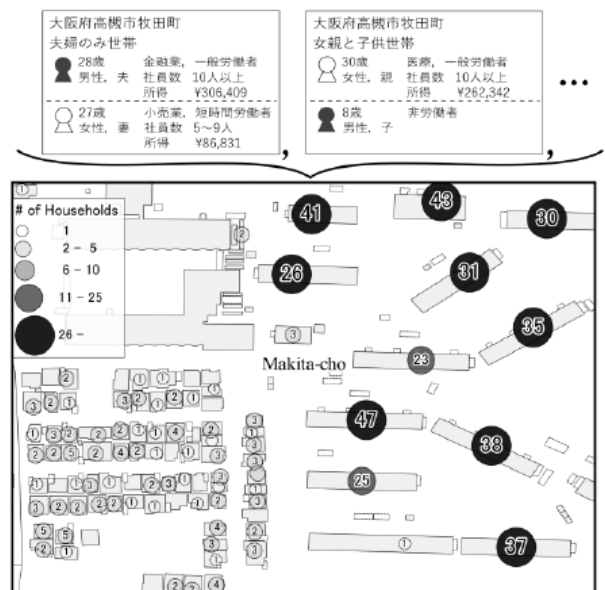


図 1 合成する仮想個票の例

するためには、公開されていない統計表を推計するか、都道府県の統計表を縮小する必要があった。そのため、市区町村単位で合成した仮想個票を統合し、都道府県単位の仮想個票を作成する場合、例えば市区町村単位の仮想個票が使用した統計表と完全に一致している場合においても、統合した仮想個票を都道府県単位の統計表と比較すると差異が発生する恐れがあった。

本研究では、各仮想個票が市区町村属性を保持しつつ、都道府県の統計表にも適合する仮想個票の合成手法を提案する。具体的には、ある都道府県を構成する市区町村の仮想個票を一度に合成する。その際に、統計表の調査範囲と仮想個票の市区町村を対応させることにより、統計表の推計や過度な縮小作業が不要となる。

## 2. 従来手法

従来手法[16]は、統計情報を基に作成した仮想個票を、複数の統計表に適合させる手法である。個人の年齢や親子の年齢差の統計表に対して、コンピュータ上で再現した世帯構成のデータ集合（仮想個票）の統計値との誤差を計算し、SA 法を用いて誤差を最小化している。仮想個票は複数の世帯とその構成員である個人によって構成される。著者らの手法は、統計表の対象地域と同じ人口規模の仮想個票

の合成を試みている。

著者らは、対象地域と同じ人口規模の仮想個票を合成するために、世帯数や人口などの統計表通りに初期世帯を合成している。著者らの初期世帯合成法 [16]では、家族類型、世帯人員別に初期世帯を合成し、その後個人の性別と初期の年齢を設定している。初期世帯を合成する際には、国勢調査 人口等基本集計 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。例えば、夫婦と子供世帯において世帯人員数が 8 人の世帯が 100 世帯と統計表に記載されている場合、夫婦と子供世帯における子供が 6 人存在する世帯を 100 世帯合成する。その後、国勢調査 人口等基本集計 表 16-1 を用いて各個人の性別と年齢を設定する。

初期世帯合成後、以下の手続きにより仮想個票と統計表との差を SA 法により最小化する。

Step 1 仮想個票内の同じ家族類型かつ性別の個人をランダムに 2 人選択する。

Step 2 選択した 2 人の年齢を入れ替える。

Step 3 仮想個票と統計表との差を計算する。

Step 4 メトロポリス法により解を遷移判定する。

Step 5 探索回数に達するまで Step1 に戻る。

Step 3 では、次式を用いて仮想個票と統計表との差を計算している。

$$f(A) = \sum_{s=1}^S \sum_{j=1}^{G_s} |c_{sj} - R_{sj}| \quad (1)$$

ここで、 $A$  は仮想個票、 $S$  は最適化に用いる統計表の数、 $G_s$  は統計表  $S$  の項目数、 $R_{sj}$  は統計表  $S$  の項目  $j$  における統計値である。 $c_{sj}$  は統計表  $S$  の項目  $j$  における仮想個票から作成した仮定の統計値である。

Step 3 では次の統計表との差を計算している。

- 統計表 1 父子年齢差
- 統計表 2 母子年齢差
- 統計表 3 夫婦年齢差
- 統計表 4 単独世帯の男性の人口分布 (1 歳階級)
- 統計表 5 単独世帯の女性の人口分布 (1 歳階級)
- ⋮
- 統計表 21 夫婦、子供とひとり親世帯の女性の人口分布 (1 歳階級)

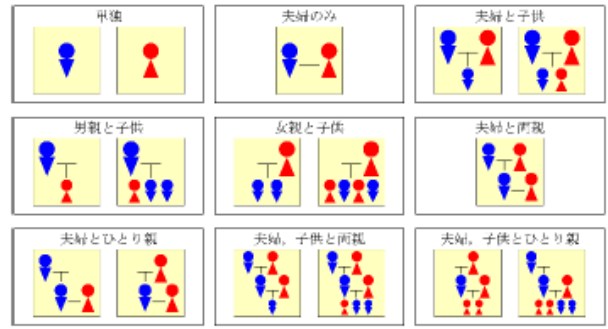


図 2 合成対象の 9 種類の家族類型

なお、統計表 4~21 は図 2 に示す 9 種類の家族類型における男女別の 1 歳階級の人口分布である。人口 20 万人未満の地域においてこれらの統計表は 5 歳階級の人口分布のみ公開されている。そのため、従来手法を用いるためには、5 歳階級を 1 歳階級に推計する必要がある。

### 3. 提案手法

本研究では、市区町村属性を保持しつつ、都道府県下の市区町村を一度に合成する。そのため、従来手法の初期解合成法と SA 法で最小化する統計表、最適化の手続きを変更する。

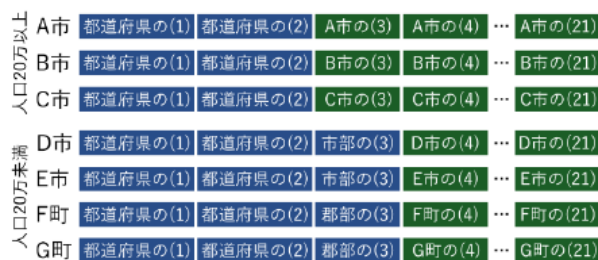
初期世帯を合成する際には、都道府県下のすべての市区町村を合成する。各市区町村の初期世帯を合成する際には従来手法を用いる。

従来手法と提案手法における SA 法により最小化する統計表とその適用範囲を図 3 に示す。図 3 (a) に示す従来手法では、都道府県下の市区町村を独立に合成していた。そのため、統計表 1~3 を縮小して使用していた。また、人口 20 万人未満の統計表 4~21 は 5 歳階級から 1 歳階級へ推計して使用していた。

図 3 (b) に示す提案手法では、統計表 1~3 は同じ調査範囲で使用する。また、統計表の推計をさける。そのため、人口 20 万人未満の市町村の統計表 4~21 は 5 歳階級の人口分布を用いる。この場合、5 歳階級内の 1 歳ごとの人口に偏りが生じる恐れがあるため、本研究では、次の統計表を追加する。

- 統計表 22 男性の人口分布 (1 歳階級)
- 統計表 23 女性の人口分布 (1 歳階級)

提案手法の最適化の手続きでは Step 1 を次のように変更する。



(a) 従来手法



(b) 提案手法

図3 従来手法と提案手法における統計表の適用範囲（図中の(1)~(23)は統計表の番号である）

Step 1-1 個人をランダムに 1 人選択 する。

Step 1-2 後述する条件に該当する個人を選択する。

Step 1-2 では、Step 1-1 で選択した個人によって選択する個人の条件を変更する。Step 1-1 で選択された個人が人口 20 万人以上の市区に属している場合、Step 1-2 では、選択されている個人と同じ市区かつ家族類型かつ性別の個人をランダムに選択する。

一方、Step 1-1 で選択された個人が人口 20 万人未満の市町村に属している場合、以下の候補からランダムに選択する。

- Step 1-1 で選択された個人と同じ市町村かつ家族類型かつ性別の個人
- 人口 20 万人未満の市町村の個人のうち、Step1-1 で選択された個人と同じ家族類型かつ性別かつ年代の個人

#### 4. 実験結果

本研究では、日本全国の約 1900 市区町村における 2010 年 と 2015 年を対象に仮想個票を 10 セット合成する。式(1)により算出した統計表との誤差を表 1 に示す。なお、表 1 の算出には 47 都道府県の統計表 1~21 を用いた。これは、従来手法と提案手法で最小化する統計表の種類と数が異なるためである。

表 1 から統計表との誤差を 92.3%削減できている。これは、統計表の推定を避けたことと統計表と同じ調査範囲で使用したことにより、都道府県の統計表との誤差を削減できた。

表 1 統計表との誤差 (10 試行の平均値)

	従来手法	提案手法
統計表との誤差	3,427,773.0	260,563.4
1 人当たりの誤差	0.0295 / 人	0.0022 / 人

#### 5. おわりに

本研究では、日本全国の約 1900 市区町村における、2010 年 と 2015 年の統計情報を用いた、図 1 に示す属性をもつ仮想個票を 10 セット作成した。これらの仮想個票は、次の Web サイトで提供を開始している (<http://www.res.kutc.kansai u.ac.jp/~murata/rsss distribution/>)。

#### 参考文献

- (1) J. M. Epstein and R. L. Axtell, MIT Press, (1996).
- (2) K. M. Carley and W. A. Wallace, Springer US, (2001).
- (3) R. Axelrod, Princeton University Press, (1997).
- (4) 高橋真吾, 計測と制御, **52**, 582-587, (2013).
- (5) 出口弘, 計測と制御, **52**, 574-581, (2013).
- (6) 高橋大志, 計測と制御, **52**, 641-647, (2013).
- (7) 寺野隆雄, 人工知能学会誌, **18**, 6, 710-715, (2003).
- (8) 市川学, 出口弘, 計測自動制御学会論文集, **49**, 11, 1012-1019, (2014).
- (9) 花岡和星, 地学雑誌, **118**, 4, 646-664, (2009).
- (10) 杜逆索, 村田忠彦, システム制御情報学会論文誌, **29**, 9, 422-431, (2016).
- (11) A. G. Wilson and C. E. Pownall, Area, **8**, 4, 246-254, (1976).
- (12) J. Barthelemy and P. L. Toint, Transportation Science, **47**, 2, 266-279, (2012).
- (13) F. Gargiulo, et. al., PLoS One, **5**, 1, 266-279, (2010).
- (14) M. Lenormand and G. Deffuant, Journal of Artificial Societies and Social Simulation, **16**, 4, 1-9, (2013).
- (15) 花岡和星, 人文地理, **64**, 3, 195-211, (2012).
- (16) T. Murata, et. al., SICE JCMSI, **10**, 6, 513-519, (2017).
- (17) T. Harada and T. Murata, SICE JCMSI, **10**, 6, 505-512, (2017).

