

日本の全世帯の位置情報付き仮想個票データの統計データからの生成

原田 拓弥¹⁾、杉浦 翔²⁾、村田 忠彦³⁾

¹⁾ 関西大学大学院 総合情報学研究科 総合情報学専攻

²⁾ 関西大学大学院 総合情報学研究科 知識情報学専攻

³⁾ 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。その方法の一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。その課題の一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。しかし、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) を生成し、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きている[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、500 m 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、パーソントリップ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生成した。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想の個人の属性（以下、仮想個票）を生成する研究が行われている。仮想個票の生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。SR 法は、統計調査時に市民や調査員が記入する個票データのサンプルをもとに、**Iterative Proportional Fitting Procedure**（以下、**IPFP**）[12]を用いて仮想個票を生成する手法である。IPFP とは、 N 次元（ N 属性）からなるクロス表の各セルの行や列の総和が統計値に適合するように、繰り返し処理を用いて各セルの値を推計する手法である。その後、数多くの仮想個票を生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[13]は、IPFP 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生成ができたとしても、両方に適合する仮想個票の生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[14]や Barthelemy ら[13]は、個票データのサンプルを用いない生成手法を提案している。Lenormand ら[15]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生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 [14]。本邦の統計表を用いた手法として花岡 [16]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[17-19]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。

本研究では、日本全国、約 1900 市区町村（平成 27 年度国勢調査）において、2000 年、2005 年、2010 年、2015 年における位置情報付きの仮想個票を 10

セット作成する。また、本研究では、図 1 に示す 16 種類の家族類型中、日本の全人口の 95%をカバーする 9 種類の家族類型を対象に、仮想個票を生成する。本研究が生成する仮想個票の例を図 2 に示す。図 2 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。

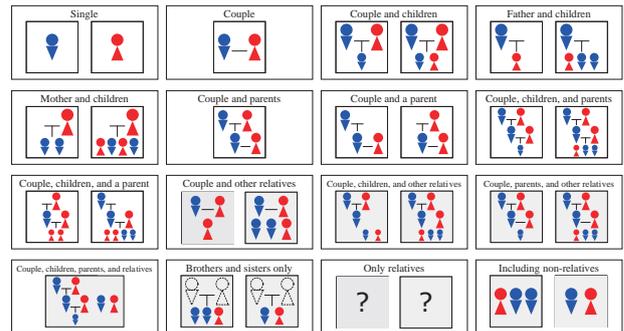


図 1 家族類型一覧（本研究の対象は黄色の 9 種類）

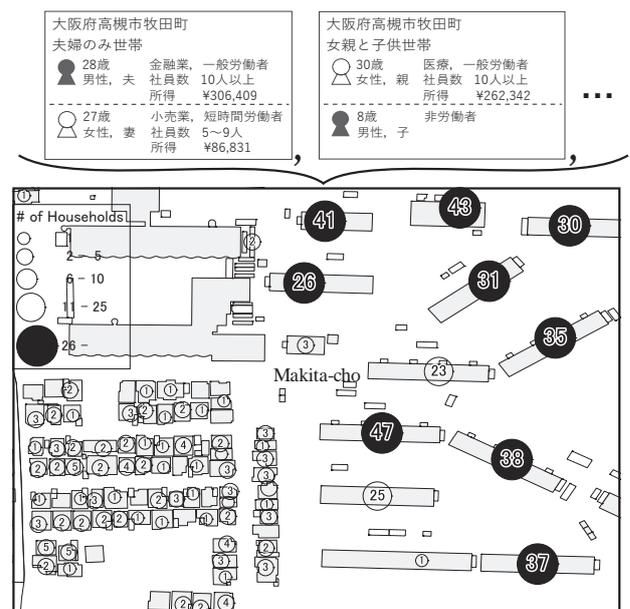


図 2 生成する仮想個票の例

2. 仮想個票生成手法

本研究では 2000 年から 2015 年における日本全国の仮想個票の属性を以下の手順で生成する。

手法1 日本全国の仮想個票の生成（家族類型、年齢、性別、世帯内の役割）[17]

手法2 手法 1 のデータへの SA 法を用いた住宅所有の関係と住宅の建て方の属性の追加[18]

- 手法3 手法2のデータへのSA法を用いた市区町村単位における町丁目の属性の追加[18]
- 手法4 手法3のデータへの建築物情報を用いた建築物の位置情報の追加[18]
- 手法5 手法1のデータへのSA法を用いた就業状態、産業分類、就業形態、企業規模、所得の属性の追加[19]

2.1 手法1：仮想個票の生成[17]

著者らが提案した仮想個票の生成手法[17]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生成を試みている。

著者らは、対象地域と同じ人口規模の仮想個票を生成するために、対象地域の世帯数や人口などの統計表通りに初期世帯を生成している。著者らの初期世帯生成法[17]では、家族類型、世帯人員別に初期世帯を生成し、その後個人に性別と初期の年齢を設定している。初期世帯を生成する際には、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（16区分）、世帯人員（16区分）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世帯生成する。その後、国勢調査 人口等基本集計 表16-1を用いて、各個人の性別と初期の年齢を設定する。

初期世帯生成後、著者らは以下の手続きにより、仮想個票と統計表との差をSA法を用いて最小化している。

- Step 1-1 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 1-2 選択した2人の年齢を入れ替える。
- Step 1-3 仮想個票と統計表との差を計算する。
- Step 1-4 メトロポリス法により解を遷移判定する。

Step 1-5 探索回数が規定数以下の場合、Step 1-1に戻る。

Step 1-3 では、以下の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想の統計値である。

2.2 手法2～5：仮想個票へ新たな属性の追加[18, 19]

2.1節で生成した仮想個票が保持する属性は4種類であり、これらの属性のみを用いて実現できるシミュレーションは限られている。仮想個票が多様な属性を保持することにより、様々なシミュレーションが実現できる。

著者らは、以下の手順により、仮想個票へ新たな属性を追加している。

Step 2-1 既存の属性と新たに追加する属性とを関連付ける統計表を複数見つける。

Step 2-2 新たに追加する属性を各世帯の構成員にランダムに割り当てる。

Step 2-3 SA法を用いて、Step 2-1の統計表と一致するように割り当てた属性を調整し、統計表と仮想個票との差を最小化する。

上記の手順を用いて、文献[18]では、住宅所有の関係や住宅の建て方、町丁目属性を仮想個票へ追加している。文献[19]では、就業状態・産業分類・就業形態・企業規模・所得属性を仮想個票へ追加している。

3. 生成結果

本研究では、日本全国の約1900市区町村における、2000年、2005年、2010年、2015年を対象に仮想個票を10セット生成する。生成対象の人口及び世帯数の一例を表1に示す。本研究において政令指定都市を対象とする場合、政令指定都市下の行政区を

対象とする。これは、政令指定都市下の行政区は仮想個票を生成する上で必要となる統計表が公開されているからである。表 1 の市区町村数は政令指定都市下の行政区を含み、政令指定都市と図 1 の 9 種類の家族類型に属する個人が存在しない地域を除いた。

仮想個票を生成する際に発生する統計表との差の一例を表 2 に示す。生成時の設定として、探索回数を手法 1 は 100,000 回 / 人、手法 2 は 1,000 回 / 世帯、手法 3 は 10,000 回 / 世帯、手法 5 は 500 回 / 人とし、SA 法のパラメータは初期温度 1.0、収束温度 0.1、指数冷却とした。表 2 から、手法 1 における統計表との誤差は 9,500 前後である。これは、合成対象の約 1 億人に対して、約 9,500 人ほど、統計表と生成した仮想個票に相違があることを示している。手法 2 では約 22 万世帯、統計表と仮想個票に相違がある。手法 3 では統計表と仮想個票に約 400 万人の相違があり、手法 1、2 と比べ大きな誤差が発生している。手法 3 は手法 1 と手法 2 の後に町丁目属性の追加をしている。そのため、手法 3 の統計表との誤差は手法 1 と手法 2 の統計表との誤差の影響を受ける。また、町丁目属性の追加に用いる統計表は粗い統計表しか公開されていない。そのため、手法 1 と手法 2 に比べ統計表の調整や推計が必要となり、仮想個票を統計表に適合させることが困難であった。手法 5 の統計表との誤差は 2010 年に比べ 2015 年が多い。これは、手法 1 の影響を受けた結果、2010 年と比べ 2015 年の統計表との誤差が増加したと考えられる。

表 1 対象の市区町村数、世帯数及び人口

年度	市区町村数	世帯数	人口
2010	1,901	49,389,597	116,161,897
2015	1,891	50,962,785	115,552,530

表 2 統計表との誤差 (10 試行の平均値)

手法	年度	
	2010 年	2015 年
手法 1 世帯構成の生成	9,205.8	9,845.8
手法 2 住宅関連属性の追加	226,151.6	219,943.8
手法 3 町丁目属性の追加	4,165,598.1	4,582,970.5
手法 5 所得関連属性の追加	56,599.0	65,233.0

4. おわりに

本研究では、日本全国の約 1900 市区町村における、2000 年、2005 年、2010 年、2015 年の統計情報を用いた、図 2 に示す属性をもつ仮想個票を 10 セット作成した。本研究により生成した仮想個票の提供を予定しており、希望される研究者がおられれば、関西大学 村田 (murata@kansai-u.ac.jp) まで連絡をいただきたい。

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) W. E. Deming and F. F. Stephan, The Annals of Mathematical Statistics, **11**, 428–444, (1940).
- (13) J. Barthelemy and P. L. Toint, Transportation Science, **47**, 2, 266–279, (2012).
- (14) F. Gargiulo, *et. al.*, PLoS One, **5**, 1, 266–279, (2010).
- (15) M. Lenormand and G. Deffuant, Journal of Artificial Societies and Social Simulation, **16**, 4, 1–9, (2013).
- (16) 花岡和星, 人文地理, **64**, 3, 195–211, (2012).
- (17) T. Murata, *et. al.*, SICE JCMSI, **10**, 6, 513–519, (2017).
- (18) T. Harada and T. Murata, SICE JCMSI, **10**, 6, 505–512, (2017).
- (19) T. Murata, *et. al.*, Proc. of IEEE SSCI, 471–476, (2017).