

仮想都市の統計情報を用いて合成した合成人口の評価

○原田拓弥（青山学院大学） 高橋真吾（早稲田大学） 村田忠彦（関西大学）

Evaluation of Synthetic Population Synthesized From the Statistics of A Virtual City

*T. Harada (Aoyama Gakuin University), S. Takahashi (Waseda University), and T. Murata (Kansai University)

概要— 本研究では、人口個票の評価手法を提案する。従来の人口個票合成手法では、統計表と合成した人口個票から生成する統計表との差により合成された人口個票を評価していた。実個票と比較できれば、実個票と人口個票の類似度による評価ができる。しかし、実個票の利活用は困難である。本研究では、仮想都市における真の個票を作成し、その個票から各合成手法が用いる統計表を生成することで、真の個票と人口個票の類似度により、これまで提案されている合成手法の評価を試みる。これにより、合成に用いる統計表や手法が大きく変化した場合においても従来の手法との比較も可能となる。

キーワード: Real-Scale Social Simulation, Synthetic Population, Statistics

1 はじめに

国家的また国際的な災害対策や経済政策において、より精度が高く、きめの細かい対応が求められている。これらの分析と将来の可能性を可視化する社会シミュレーションへの関心が高まっている¹⁾。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)²⁾に基づいてモデル化されていた。しかし、KISS 原理では現実社会の複雑な現象のモデル化は不可能であると指摘されている³⁾。そのため、可能な限り忠実に現実社会を模倣する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。このようなモデルを作成するためには、環境のデータと市民のデータが必要となる。環境のデータは地理情報や地域メッシュといった地理情報システムの利用が可能である。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、可能な限り現実社会を模倣する必要がある。市民の意思決定モデルを記述する際には、年齢、性別、職業などの市民個人の属性が必要になることが少なくない。

現実社会を可能な限り模倣するエージェントの意思決定の実現には様々な課題がある。その課題の1つがエージェントが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、実データによるエージェントの属性の設定が可能である。しかし、これらの市民のデータ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いる^{4),5)}。

統計情報に基づく個票データの合成¹⁾に関する研究の歴史は古く、Synthetic Reconstruction Method (SR 法)⁶⁾として知られている。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⁷⁾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⁸⁾や Barthelemy ら⁷⁾は、サンプルを用いない合成手法を提案している。Lenormand と Deffuan⁹⁾は、サンプルを用いて合成する SR 法と、サンプルを用いない合成手法とを比較し、後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。これらの海外の研究ではそれぞれの国において利用可能な統計表と特徴に基づいた手法が開発されており、他の地域へ適用する際には留意が必要と指摘されている⁸⁾。日本においても、日本の利活用可能な情報と特徴に基づいた人口個票が合成されている。日本における合成手法として、国勢調査のサンプルを用いた花岡の手法¹⁰⁾とサンプルを用いない池田ら¹¹⁾及びその派生手法¹²⁾⁻¹⁶⁾が提案されている。

しかし、これらの手法¹⁰⁾⁻¹⁶⁾では、統計表と合成された人口個票から生成する統計表との差（以下、統計表との誤差）により合成した人口個票を評価している。これは、統計表の利活用は可能であるものの、統計表を作成する際に用いる実個票データの利活用は困難であることが理由である。そのため、統計表と誤差により評価されていた。従って、現実の集団と合成された人口個票が同じ統計的特徴をもつ集団と言及することはできるが、現実の集団との一致度については言及することができていない。

そこで、本研究では、統計表との誤差以外に真の個票との一致度で合成された人口個票の評価を試みる。真の個票との一致度で合成手法を比較できれば、合成に用いる統計表や手法が大きく変化した場合においても、従来の手法との比較が可能となる。しかし、前述のように現実の個票データは利活用が困難である。そのため、本研究では、まず、コンピュータ上で様々な世帯により構成される仮想の都市を生成する。次に、真の個票データから合成手法で用いる統計表を生成する。生成した統計表をもとに人口個票を合成し、合成した人

¹⁾ここで、個票データの「復元」ではなく「合成」という用語を用いている。復元の場合、実際の人口構成と同一の個票の復元が期待されるが、合成される個票はあくまでも統計的特徴が類似した仮想的な個票である。

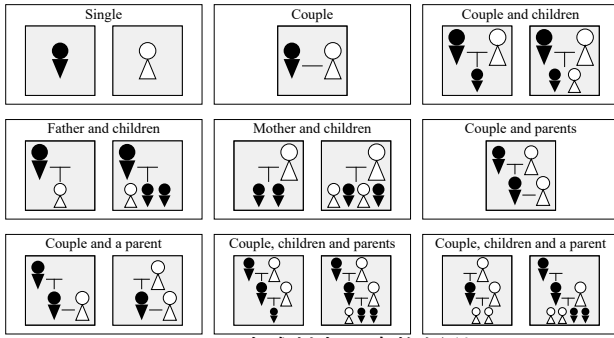


Fig. 1: 合成対象の家族類型

口個票と仮想都市の真の個票を比較することで、本研究では、合成手法と真の個票の一致度により合成手法の評価を試みる。

2 人口個票合成手法

池田らが提案した合成手法¹¹⁾とその改良手法^{12)–16)}は、公開されている複数の統計表に適合するような人口データを合成する手法である。市民の年齢や親子の年齢差についての統計表とコンピュータ上で再現したデータ集合（人口個票）から作成する統計表の差を計算し、探索的解法の1つである Simulated Annealing（以下、SA法）を用いて誤差を最小化している。これらの合成手法では Fig. 1 に示す9種類の家族類型を対象に人口個票を合成している。Fig. 1 の9種類の家族類型に分類される世帯の割合は約95%である。

これらの合成手法は大きく分けて4つの要素からなる。それぞれ、初期世帯生成法、最適化に用いる統計、目的関数、最適化の手続きである。本節では、池田ら¹¹⁾、福田ら¹²⁾、柘井ら¹³⁾、原田ら¹⁴⁾の分散手法、原田ら¹⁵⁾、Murate et. al.¹⁶⁾の手法について4つの要素を簡単に説明する。詳細は文献^{11)–16)}を参照されたい。

2.1 池田ら¹¹⁾の手法

2.1.1 要素1 初期世帯生成法

人口個票の初期生成では、世帯数 H だけ世帯を生成する。世帯の生成時、家族類型別世帯数の割合に従って世帯の家族類型を決定する。

世帯の家族類型決定後、子供が存在する家族類型であった場合、家族類型別（夫婦と子供、男親と子供、女親と子供）、子供の数別（1~4人）、世帯の割合の統計表を用いて、世帯に居住する子供の数を確率的に決定する。

家族類型と子供の人数が決定した後、その世帯の構成員の属性を設定する。構成員の年齢は男女別、年齢別人口の割合に従って確率的に設定する。性別は、単独世帯に所属する構成員や子供、ひとり親の構成員をランダムに設定する。それ以外の構成員は、世帯の役割の属性に応じた性別を適切に設定する。

2.1.2 要素2 最適化に用いる統計表

池田らは次のような統計表との差を最小化している。

- (1) 父子の年齢差
- (2) 母子の年齢差
- (3) 夫婦の年齢差
- (4) 男性の人口分布
- (5) 女性の人口分布

- (6) ある年齢の男性が単独世帯にいる割合
- (7) ある年齢の女性が単独世帯にいる割合
- (8) ある年齢の男性が夫婦のみ世帯にいる割合
- (9) ある年齢の女性が夫婦のみ世帯にいる割合

2.1.3 要素3 目的関数

池田らは以下の目的関数を用いて最適化を行っている。

$$f(A) = \sum_{s=1}^S g_s(A) \quad (1)$$

$$g_s(A) = \frac{4}{G_s} \sum_{j=1}^{G_s} (v_{sj}(A) - r_{sj} \times m_{sj}(A))^2 \quad (2)$$

ここで、 A は人口個票、 S は統計表の数、 G_s は統計表 s の項目数、 v_{sj} は統計表 s の項目 j の値、 r_{sj} は統計表 s の項目 j の割合、 $m_{sj}(A)$ は統計値 r_{sj}^c の補正值である。式(2)では、統計表 s の各項目の誤差 $((c_{sj}(A) - r_{sj} \times m_{sj}(A))^2)$ が0.5のとき、 $g_s(A) = 1.0$ となるように正規化している。詳細は文献^{11),12)}を参照されたい。

2.1.4 要素4 最適化の手続き

池田らは以下の手続きにより最適化している。

Step 1 人口個票を初期生成

Step 2 探索回数が規定数に達すれば探索を終了

Step 3-1 人口個票内の市民をランダムに1人選択

Step 3-2 選択された市民の年齢を人口分布に基づきランダムに変更

Step 4 解の遷移判定

Step 5 探索回数を更新してSAの温度を冷却

Step 6 Step 2の処理に戻る

2.2 福田ら¹²⁾の手法

福田ら¹²⁾は池田ら¹¹⁾が用いた統計表(1), (2), (3), (6), (7), (8), (9)に加えて、以下の統計表を用いている。

- (10) ある年齢の男性の有配偶率
- (11) ある年齢の女性の有配偶率
- (12) 年齢別人口分布

その他の3つの要素は池田ら¹¹⁾の手法と同様である。

2.3 柘井ら¹³⁾の手法

柘井ら¹³⁾は池田ら¹¹⁾の目的関数を次のように変更し、最適化している。

$$g'_s(A) = \sum_{j=1}^{G_s} |v_{sj}(A) - \text{Round}(r_{sj} \times m_{sj}(A))| \quad (3)$$

2.4 原田らの分散手法¹⁴⁾

原田らの分散手法¹⁴⁾では、柘井ら¹³⁾の手法を並列分散処理により高速化する手法が提案されている。原田らの分散手法¹⁴⁾では、統計表を N 分割し、 N 個の人口個票を並列に合成することで高速化している。単純に統計表を N 分割すると、小数点以下の統計量を切り捨てや切り上げといった操作により、分割した N 個の統計表の総和がもとの統計表と一致しない。原田らの分散手法¹⁴⁾では実統計表を相違なく分割することで、統計表との誤差を抑えつつ高速化している。

2.5 原田らの¹⁵⁾手法

原田ら¹⁵⁾は柘井ら¹³⁾の目的関数以外の3つの要素を変更している。

2.5.1 要素1 初期世帯生成法

原田ら¹⁵⁾は対象地域と同じ規模の世帯構成を合成するために、対象地域を集計対象とする家族類型、世帯人員別世帯数の統計表を用いて世帯数や人口などの統計表の値通りに初期世帯を生成している。また、個人の性別と初期の年齢を設定する際には、家族類型別、男女別、人口分布の統計を用い、男女の数や年齢別の人口に相違が発生させずに確率的に設定する。

2.5.2 要素2 最適化に用いる統計表

原田ら¹⁵⁾は池田ら¹¹⁾が用いた統計表(1), (2), (3)に加えて、以下の統計表を用いている。

(13) 9種類の家族類型別、男女別人口分布

2.5.3 要素4 最適化の手続き

原田ら¹⁵⁾は池田らの最適化の手続きのStep3-2を以下のように変更している。

Step 3-2 選択された市民の年齢を、家族類型別、男女別、世帯内の役割別、人口分布に基づき確率的に変更

2.6 Murata et. al.¹⁶⁾の手法

Murata et. al.¹⁶⁾は原田ら¹⁵⁾の手法をもとに、最適化の手続きのStep 3を以下のように変更している。

Step 3-1 人口個票内の市民をランダムに1人選択

Step 3-2 Step 3-1で選択された市民と同じ家族類型かつ性別の市民をランダムに1人選択する。

Step 3-3 Step 3-1とStep 3-2で選択した個人の年齢を交換する。

3 仮想都市の生成と個票の比較手法

本研究では、前節で紹介した合成手法と真の個票の一致度の評価を試みる。そのために、仮想都市における真の個票を生成する方法と2つの個票の比較方法を説明する。

3.1 仮想都市の生成

仮想都市における真の個票の生成では、世帯の生成と、世帯構成員の属性の設定の2つの手続きにより個票を生成する。世帯の生成では、世帯数 H だけ世帯を生成する。その際に、家族類型と世帯人員数を順に決定する。生成する世帯の家族類型を設定する際には、国勢調査人口等基本集計¹⁷⁾表11を用いた。表11は世帯の家族類型別、世帯人員別世帯数が集計されている。同表から家族類型別の世帯数を算出し、生成する世帯の家族類型を確率的に設定する。世帯人員数を設定する際においても、国勢調査人口等基本集計¹⁷⁾表11を用いた。同表から家族類型別、世帯人員別の世帯数を算出し、生成する世帯の家族類型に応じて、世帯人員数を確率的に決定する。

世帯の生成後、各世帯の構成員の性別と年齢を設定する。年齢と性別の設定には、国勢調査人口等基本集計¹⁷⁾表16-1と表17、人口動態職業・産業別統計¹⁸⁾

保管表 出生 表1と2を用いた。国勢調査人口等基本集計¹⁷⁾表16-1は、世帯の家族類型別、配偶関係別、男女別、年齢別、人口が同表17は夫の年齢別、妻の年齢別夫婦数が集計されている。表16-1から家族類型別、男女別人口と家族類型別、男女別、年齢別人口を、表17から夫婦の年齢差別夫婦の数を算出する。人口動態職業・産業別統計¹⁸⁾保管表 出生 表1は父の年齢別出生数が、同表2は母の年齢別出生数が集計されている。これらの表から、父の年齢別出生数と母の年齢別出生数を算出する。算出した4つの表を用いて各世帯の構成員の年齢と性別を設定する。

単独世帯の構成員の性別は単独世帯の男女別の人口をもとに確率的に設定する。年齢は単独世帯における確率的に設定した性別の年齢別人口を用いて確率的に設定する。単独世帯以外の世帯は夫婦の両親・ひとり親、夫婦・男親・女親、子供の順に年齢と必要である場合、性別を設定する。世帯に夫婦の両親が存在する場合は夫、妻の順に年齢を設定する。夫の年齢を夫婦のみ世帯における男女別の人口をもとに確率的に設定する。妻の年齢は設定した夫の年齢と夫婦の年齢差別夫婦の数をもとに確率的に設定する。

世帯に夫婦のひとり親が存在する場合は、当該家族類型の男女別、年齢別、人口をもとに確率的に性別を設定し、当該家族類型の設定した性別における年齢別人口を用いて年齢を決定する。世帯に夫婦とその両親、もしくはひとり親が存在する場合は、夫の母の年齢と母の年齢別出生数をもとに夫の年齢を確率的に設定し、妻の年齢を夫婦の年齢差別夫婦の数をもとに確率的に設定する。

世帯に夫婦の両親が存在しない場合は、夫婦の両親と同様に夫、妻の順に年齢を決定する。世帯に男親もしくは女親が存在する場合は、当該家族類型の男女別、年齢別、人口をもとに確率的に性別を設定し、当該家族類型の設定した性別における年齢別、人口を用いて年齢を順に決定する。

世帯内に子供と妻もしくは女親が存在する場合、母の年齢別出生数を用いて年齢を設定し、当該家族類型の男女別人口をもとに性別を決定する。一方、世帯に子供と夫もしくは男親が存在する場合、父の年齢別出生数を用いて年齢を設定し、当該家族類型の男女別人口をもとに性別を決定する。世帯構成員の年齢を設定した際に、年齢が0歳未満や、101歳以上となった場合は、構成員1人目から再度年齢と性別を設定する。

3.2 個票の比較

2つの個票の比較の手順の概要を以下に示す。

Step 1 比較する2つの個票(A, B)それぞれ以下の処理を実施する

Step 1-1 個票内の各世帯を世帯の家族類型、世帯人員数、ひとり親・子供の性別の内訳をもとに分類する

Step 1-2 個票の各世帯内の構成員を年齢、性別、世帯内の役割を用いてソートする

Step 2 個票A, Bにおいて類似する世帯を以下の処理により探索する

Step 2-1 探索する世帯間の差 d を0(世帯間の差なし)と設定する

Step 2-2 個票 A の i 番目の世帯 A_i との差が d である世帯を個票 B から探す

Step 2-3 世帯 A_i との差が d である世帯 B_j が存在する場合、世帯 A_i と B_j をそれぞれの個票から取り除き、 i に 1 加算し、Step 2-2 に戻る

Step 2-4 探索する世帯間の差 d が既定値以下の場合、 d に 1 加算し、 i を 0 に設定し、Step 2-2 に戻る

Step 2-5 Step 1-1 で分類されたすべての分類を探索するまで Step 2-1 に戻る

Step 1-1 の世帯の分類では、まず、合成対象の 9 種類の家族類型に加えて、ひとり親が存在する「夫婦とひとり親」世帯と「夫婦、子供とひとり親」世帯はひとり親の性別で細分化する。次に、子供が存在する家族類型は子供の男女の数によって細分化する。例えば、「夫婦と子供」世帯において、子供が 2 人存在する世帯は、男性の子供 2 人、男女 1 人ずつ、女性の子供 2 人の 3 種類に細分化する。Step 1-2 の世帯内の構成員のソートでは、まず、世帯内の役割が夫、妻、夫の父、夫の母、子供となるように並べ替える。次に、子供の役割をもつ構成員を性別が男性・女性となるように並べ替える。最後に、子供の役割をもつ構成員を男女別に年齢の降順となるように並べ替える。

Step 2-2 で用いる世帯間の差は次式で計算する。

$$h(x, y) = \sum_{m=1}^M |x_{m,age} - y_{m,age}| \quad (4)$$

ここで、 x と y は比較する世帯を、 M は x と y の世帯人員数、 $x_{m,age}$ は世帯 x の m 番目の構成員の年齢である。 x と y に指定する世帯 A_i と B_j は年齢や性別、世帯内の役割によってソートされているため、 $h(x, y) = 0$ の場合、世帯 A_i と B_j は完全に一致する。Step 2 の操作を Step 1-1 のすべての分類に対して実施することで、個票 A と B の一致度を世帯間の差 d ごとに算出することができる。

4 実験結果

本研究では、人口個票と真の個票の一致度を比較するために、まず、真の個票となる仮想都市を生成し、仮想都市から各合成手法で用いる統計表を出力する。次に、各手法から合成された人口個票と真の個票の一致度を算出する。本研究が用いた計算機の CPU は、AMD Ryzen Threadripper 1950X (3.4 GHz, 16 コア) で、メモリは DDR4-1866 16 GB × 8、OS は Microsoft Windows 10 Pro 64 bit である。

仮想都市の生成では、平成 27 年度国勢調査¹⁷⁾ と 2015 年の人口動態職業・産業別集計¹⁸⁾ における東京都の統計表を用い、生成する世帯数 $H = 500000$ とした。生成された仮想都市の家族類型別世帯数と男女別の人口を Table 1 に、家族類型別、世帯人員別世帯数を Table 2 に示す。500,000 世帯、999,032 人の仮想都市から各合成手法で用いる統計表を作成し、人口個票を合成する。

本研究では、福田ら¹²⁾ の手法、柘井ら¹³⁾ の手法、原田らの分散手法¹⁴⁾、原田ら¹⁵⁾ の手法、Murata et.

al.¹⁶⁾ の手法の 5 種の手法を比較する。探索回数は 1 人あたり 10,000 回とした。総探索回数は 9,990,320,000 回である。SA 法の初期温度と収束温度は各手法で用いた値を設定した。福田らの手法¹²⁾ は初期温度 1000 度、収束温度 0.00171 度とした。福田ら以外の手法^{13)–16)} は初期温度 1.0 度、収束温度 0.1 度と設定した。冷却関数には何れの手法も指数冷却を用いた。原田らの分散手法¹⁴⁾ では、分割数 N を 8 と 16 に設定した。

統計表との誤差を Table 3 に示す。Table 3 は人口個票を合成後に池田ら¹¹⁾ が用いた 9 種類の統計表との誤差を、柘井らが提案する式 (3) の絶対値誤差を用いて算出した。これは、合成手法により用いる統計表が異なるため、最も使用される池田ら¹¹⁾ が用いた 9 種類の統計表で再評価した。また、式 (3) の絶対値誤差を用いることで、統計表との誤差を直感的に評価することができる。

Table 3 から、Murata et. al.¹⁶⁾ の手法が最も統計表との誤差が少ない。Murata et. al.¹⁶⁾ の手法は 2 人の市民の年齢を入れ替えることで、人口分布を変更せずに他の統計表との誤差を削減することができるため、最も誤差を削減することができた。一方、福田らの手法¹²⁾ が最も誤差が多い。福田らの手法¹²⁾ では統計表 (4)、(5) の男性の人口分布と女性の人口分布を用いていないため、池田ら¹¹⁾ が用いた統計表で再評価すると誤差が増加した。

各手法が合成した人口個票と真の個票との世帯間の差別の一致世帯数を Table 4 に、その割合を Table 5 に示す。また、家族類型別、世帯間の差別の一致度を Fig. 2 に示す。Table 4 と Table 5 から、福田ら¹²⁾ の手法により合成された人口個票と真の個票において完全に一致する世帯の割合は 53.86% である。同様に、柘井ら¹³⁾ の手法は 58.66%、原田らの並列分散手法¹⁴⁾ ($N = 8$) は 58.70%、原田らの並列分散手法¹⁴⁾ ($N = 16$) は 58.69%、原田ら¹⁵⁾ の手法は 73.48%、Murata et. al.¹⁶⁾ の手法は 73.59% であった。原田ら¹⁵⁾ と Murata et. al.¹⁶⁾ の手法は家族類型別世帯数や男女別人口が統計表と一致するように合成する手法を提案している。そのため、確率的に世帯や性別を設定する福田ら¹²⁾ や柘井ら¹³⁾、原田らの並列分散手法¹⁴⁾ と比べ、真の個票と一致する世帯の割合が増加した。

Fig. 2 から、何れの手法においても単独世帯、夫婦のみ世帯は真の個票と完全に一致する世帯数が多い。世帯人員数が少ない世帯の場合、世帯が一致する確率は高くなるからである。例えば、単独世帯において、市民の年齢が 0 歳から 100 歳の場合、約 1% の確率でランダムに生成された世帯と一致する。本研究が用いた合成手法では、統計表との誤差の最小化を目的としている。そのため、単独世帯や夫婦のみ世帯など世帯人員数が少ない世帯の一致度が高い。

単独世帯を除いた場合における真の個票と完全に一致する世帯の割合について、福田ら¹²⁾ の手法は 28.20%、柘井ら¹³⁾ の手法は 36.43%、原田らの並列分散手法¹⁴⁾ ($N = 8$) は 36.46%、原田らの並列分散手法¹⁴⁾ ($N = 16$) は 36.51%、原田ら¹⁵⁾ の手法は 50.03%、Murata et. al.¹⁶⁾ の手法は 50.23% であった。単独世帯と夫婦のみ世帯を除いた場合における真の個票と完全に一致する世帯の割合について、福田ら¹²⁾ の手法は 16.83%、柘井ら¹³⁾ の手法は 26.10%、原田らの並列分散手法¹⁴⁾ ($N = 8$)

Table 1: 生成した仮想都市における家族類型別世帯数と男女別人口

	世帯数	男性の人口	女性の人口
単独	234,646	119,468	115,178
夫婦のみ	90,674	90,674	90,674
夫婦と子供	125,564	229,200	221,983
男親と子供	6,044	8,490	5,359
女親と子供	34,379	33,889	46,700
夫婦と両親	402	804	804
夫婦とひとり親	2,454	2,720	4,642
夫婦、子供と両親	1,138	3,318	3,286
夫婦、子供とひとり親	4,699	9,121	12,722
合計	500,000	497,684	501,348

Table 2: 生成した仮想都市における家族類型別、世帯人員別世帯数

家族類型	世帯人員数											
	1	2	3	4	5	6	7	8	9	10	11	12
単独	234,646	0	0	0	0	0	0	0	0	0	0	0
夫婦のみ	0	90,674	0	0	0	0	0	0	0	0	0	0
夫婦と子供	0	0	64,506	49,322	10,303	1,222	167	39	3	1	0	1
男親と子供	0	4,545	1,271	202	20	4	2	0	0	0	0	0
女親と子供	0	24,519	8,168	1,460	199	21	10	2	0	0	0	0
夫婦と両親	0	0	0	402	0	0	0	0	0	0	0	0
夫婦とひとり親	0	0	2,454	0	0	0	0	0	0	0	0	0
夫婦、子供と両親	0	0	0	0	436	517	159	25	1	0	0	0
夫婦、子供とひとり親	0	0	0	2,321	1,798	503	66	10	1	0	0	0
合計 500,000 世帯	234,646	119,738	76,399	53,707	12,756	2,267	404	76	5	1	0	1

は 26.19%，原田らの並列分散手法¹⁴⁾ ($N = 16$) は 26.24%，原田ら¹⁵⁾ の手法は 30.33%，Murata et. al.¹⁶⁾ の手法は 30.50%であった。

Table 3 の統計表との誤差と Table 4, Table 5, Fig. 2 の真の個票との一致度から、統計表との誤差が少ない手法ほど真の個票との一致度が高い。そのため、統計表との誤差が少ない人口個票を合成する手法を開発できると真の個票との一致度も向上すると考えられる。しかし、統計表との誤差と真の個票との一致度は非線形の関係である。合成手法を評価するためには統計表との誤差だけでなく真の個票との一致度の確認も必要である。

5 おわりに

本研究では、仮想都市を生成することで、合成された人口個票と仮想都市の真の個票との一致度により合成手法の評価を試みた。合成手法として、福田ら¹²⁾ の手法、柘井ら¹³⁾ の手法は、原田ら¹⁵⁾ の手法、Murata et. al.¹⁶⁾ の手法を用いた。500,000 世帯、999,032 人の仮想都市から合成手法が用いる統計表を生成し、人口個票を合成した。

各合成手法が合成した人口個票と真の個票との一致度について、福田ら¹²⁾ の手法は 53.86%，柘井ら¹³⁾ の手法は 58.66%，原田らの並列分散手法¹⁴⁾ ($N = 8$) は 58.70%，原田らの並列分散手法¹⁴⁾ ($N = 16$) は 58.69%，原田ら¹⁵⁾ の手法は 73.48%，Murata et. al.¹⁶⁾ の手法は 73.59%であった。確率的に世帯を生成する福田ら¹²⁾ や柘井ら¹³⁾、原田らの並列分散手法¹⁴⁾ の手法に比べ、家族類型別世帯数や男女別人口等、統計表と整合するように合成する原田ら¹⁵⁾ と Murata et. al.¹⁶⁾ の手法は、真の個票との一致度が高い結果であった。

本研究が生成した仮想都市の市民がもつ属性は、年齢・性別・世帯内の役割・親族関係である。他の合成手法と比較する場合、仮想都市がもつ属性数の増加が必要である。また、5 歳階級や 3 区分の年齢を生成するなど、合成手法によってある属性のカテゴリ数が異なる場合、複数の手法間で真の個票との比較手法を検討する必要がある。これらは今後の課題とする。

謝辞

本研究の一部は、JSPS 科研費 17K03669, 19K23229 の助成を受けたものです。

参考文献

- 1) J. M. Epstein and R. Axtell: *Growing Artificial Societies: Social Science from the Bottom Up*, The MIT Press, 1st edition (1996)
- 2) R. Axelrod: *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration*, Princeton University Press, (1997)
- 3) 寺野：エージェントベースモデリング：KISS 原理を超えて (<特集>複雑系と集合知)，人工知能学会誌，**18-6**，710/715 (2003)
- 4) 市川：医療分野におけるリスクマネジメント 地理情報分析と社会シミュレーション技術を用いた検討，計測と制御，**57-6**，407412 (2018)
- 5) Y. Goto: Stylized Fact Analysis of Cash-for-Work Programs in the Disaster Reconstruction Process, in 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 1144/1149 (2018)
- 6) A. G. Wilson and C. E. Pownall: A New Representation of the Urban System for Modelling and for the Study of Micro-Level Interdependence, *Area*, **8-4**，246/254 (1976)
- 7) J. Barthelemy and P. L. Toint: Synthetic Population Generation Without a Sample, *Transportation Science*, **47-2**，266/279 (2013)

Table 3: 合成手法別, 統計表との誤差の総和 (10 回平均)

	福田ら ¹²⁾	柘井ら ¹³⁾	原田らの分散手法 ¹⁴⁾		原田ら ¹⁵⁾	Murata et. al. ¹⁶⁾
			$N = 8$	$N = 16$		
平均	840,159.10	250,164.90	250,467.00	251,314.80	4,505.40	4,321.40
標準偏差	2,330.69	408.31	663.62	839.97	55.86	46.47

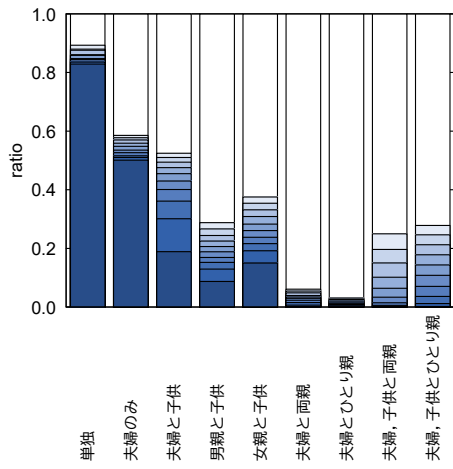
Table 4: 合成手法別, 世帯間の差別, 真の個票との一致世帯数 (10 回平均)

世帯間の差 d	福田ら ¹²⁾	柘井ら ¹³⁾	原田らの分散手法 ¹⁴⁾		原田ら ¹⁵⁾	Murata et. al. ¹⁶⁾
			$N = 8$	$N = 16$		
0	269,308.2	293,283.5	293,482.2	293,445.9	367,397.1	367,933.2
1	18,285.4	26,603.9	26,561.6	26,629.6	33,732.6	34,499.5
2	9,789.2	14,445.8	14,411.1	14,509.7	15,938.1	17,259.3
3	8,403.3	8,172.3	8,190.2	8,214.6	8,882.9	9,684.1
4	6,251.3	6,972.7	6,942.1	6,899.9	6,129.1	6,641.8
5	7,990.6	6,065.3	6,068.0	6,022.4	4,726.3	5,293.0
6	5,270.5	5,774.1	5,858.1	5,806.0	4,189.1	4,516.8
7	8,030.9	5,623.8	5,505.3	5,585.7	3,869.8	4,407.2
8	4,770.3	5,311.6	5,286.6	5,253.6	3,300.1	3,638.6
9	6,697.1	4,563.6	4,638.4	4,537.0	3,229.0	3,758.1
10-	155,203.2	123,183.4	123,056.4	123,095.6	48,605.9	42,368.4

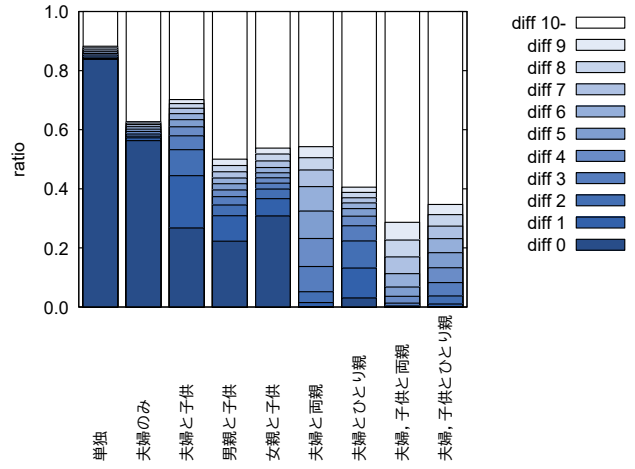
Table 5: 合成手法別, 世帯間の差別, 真の個票との一致世帯の割合 (10 回平均)

世帯間の差 d	福田ら ¹²⁾	柘井ら ¹³⁾	原田らの分散手法 ¹⁴⁾		原田ら ¹⁵⁾	Murata et. al. ¹⁶⁾
			$N = 8$	$N = 16$		
0	53.86%	58.66%	58.70%	58.69%	73.48%	73.59%
1	3.66%	5.32%	5.31%	5.33%	6.75%	6.90%
2	1.96%	2.89%	2.88%	2.90%	3.19%	3.45%
3	1.68%	1.63%	1.64%	1.64%	1.78%	1.94%
4	1.25%	1.39%	1.39%	1.38%	1.23%	1.33%
5	1.60%	1.21%	1.21%	1.20%	0.95%	1.06%
6	1.05%	1.15%	1.17%	1.16%	0.84%	0.90%
7	1.61%	1.12%	1.10%	1.12%	0.77%	0.88%
8	0.95%	1.06%	1.06%	1.05%	0.66%	0.73%
9	1.34%	0.91%	0.93%	0.91%	0.65%	0.75%
10-	31.04%	24.64%	24.61%	24.62%	9.72%	8.47%

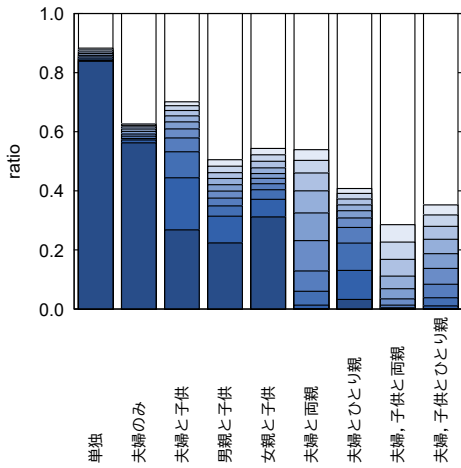
- 8) F. Gargiulo, S. Ternes, S. Huet, and G. Deffuant: An Iterative Approach for Generating Statistically Realistic Populations of Households, *PLOS ONE*, **5-1**, 1/9 (2010)
- 9) M. Lenormand and G. Deffuant: Generating a Synthetic Population of Individuals in Households: Sample-Free Vs Sample-Based Methods, *Journal of Artificial Societies and Social Simulation*, **16-4**, 12 (2013)
- 10) 花岡: 全国版の小地域マイクロデータの構築と災害分析への活用, 地域安全学会論文集, **29**, 247/255 (2016)
- 11) 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, 第43回システム工学部研究会, 11/14 (2010)
- 12) 福田, 喜多: エージェントベースの人口推計モデルにおける属性決定手法の評価, システム制御情報学会論文誌, **27-7**, 279/289 (2014)
- 13) 柘井, 村田: 統計データからの市民の属性復元のための進化計算と SA による 2 段階最適化, システム制御情報学会論文誌, **30-6**, 216/227 (2017)
- 14) 原田, 村田: 並列計算を用いた SA 法による都道府県レベルの大規模世帯の復元, 計測自動制御学会論文集, **54-4**, 421/429 (2018)
- 15) 原田, 村田, 柘井: 家族類型と世帯内の役割を考慮した SA 法による大規模世帯の合成, 計測自動制御学会論文集, **54-9**, 705/717 (2018)
- 16) T. Murata, T. Harada, and D. Masui: Comparing Transition Procedures in Modified Simulated-Annealing-Based Synthetic Reconstruction Method without Samples, *SICE Journal of Control, Measurement, and System Integration*, **10-6**, 513/519 (2017)
- 17) 総務省統計局: e-Stat 平成 27 年度国勢調査 人口等基本集計 (2016)<https://www.e-stat.go.jp/stat-search/files?page=1&layout=datalist&toukei=00200521&tstat=000001080615&tclass1=000001089055>
- 18) 総務省統計局: e-Stat 人口動態職業・産業別統計 2015 年度 (2018)<https://www.e-stat.go.jp/stat-search/files?page=1&layout=datalist&toukei=00450011&tstat=000001028897&year=20151&tclass1=000001053122>



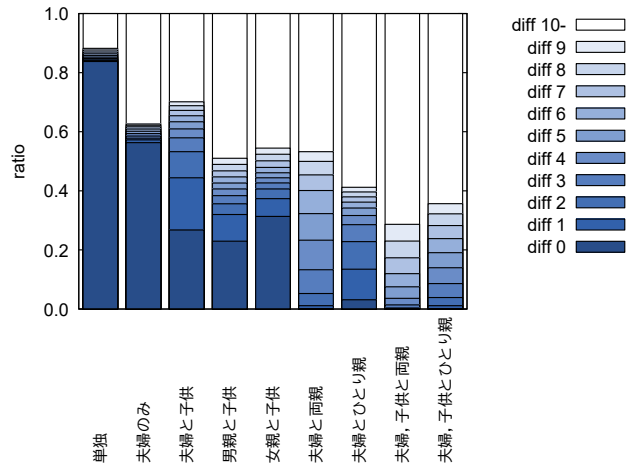
(a) 福田らの手法¹²⁾



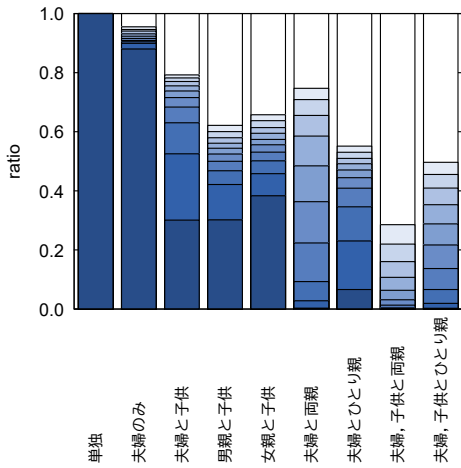
(b) 柘井らの手法¹³⁾



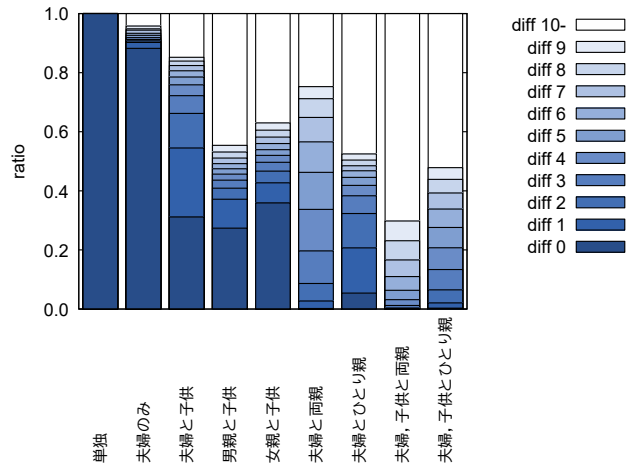
(c) 原田らの分散手法¹⁴⁾ (N = 8)



(d) 原田らの分散手法¹⁴⁾ (N = 16)



(e) 原田らの手法¹⁵⁾



(f) Murata et. al. の手法¹⁶⁾

Fig. 2: 合成手法別, 家族類型別, 世帯間の差別, 真の個票の一致度