

国勢調査結果を用いた 全ての一般世帯と施設などの世帯を含む全世界帯の合成

○原田拓弥（青山学院大学） 村田忠彦（関西大学）

Synthesize All General and Institution Households using Census

*T. Harada (Aoyama Gakuin University) and T. Murata (Kansai University)

概要—本研究では国勢調査結果を用いて全世界帯を合成する手法を提案する。従来手法では、公開されている統計表の制約のため、国勢調査における世帯の分類のうち、一般世帯に該当する9種類の家族類型に属する世帯を合成対象としていた。本研究では、不詳を含む一般世帯に属するすべての家族類型に加え、入院中や学生寮に居住する個人など施設などの世帯に分類される個人を合成対象とし、対象地域の全世界帯を合成する手法を提案する。実験結果から全世界帯を合成する提案手法においても、従来手法と同程度以下の誤差で人口合成することに成功した。

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

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。LenormandとDeffuan⁷⁾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⁹⁾とサンプルを用いない池田ら¹⁰⁾及びその派生手法^{11)–13)}が提案されている。花岡の手法⁹⁾では国勢調査のサンプルデータを用いているため、合成されたデータは第三者提供できない。従って、各々の研究者が国勢調査のサンプルデータを入手し、花岡の手法⁹⁾を実装し、合成した上でシミュレーションを実施する必要がある。一方、池田ら¹⁰⁾及びその派生手法^{11)–13)}では公開されている情報のみを用いて個票データを合成している。従って、合成した個票データを第三者へ提供することができる。

これらの手法では、探索的解法の1つであるSimulated Annealing (以下、SA法)を用いて人口合成されている。探索的解法にはSA法以外にも、進化計算や粒子群最適化などが存在する。しかし、実規模の人口を合成する場合、進化計算や粒子群最適化のように解個体を複数もつ最適化手法では、メモリ使用量の問題により個体数に制限が生じる。したがって、これらの手法では1つの解個体を探索するSA法が採用されている。

第三者提供可能な個票データを合成する手法^{10)–13)}ではFig. 1に示す国勢調査における世帯の分類のうち、一般世帯に該当するFig. 1(a)の黄色で示す9種類の家族類型を合成の対象としていた。平成27年国勢調査結果によるとこれらの9種類の家族類型に分類される人口は90%、世帯数は95%である。9種類の家族類型に分類される年齢別の人口割合をFig. 2に示す。Fig. 2から高齢になるほど9種類の家族類型に属する人口が減少しており、100歳以上の男性は54%、女性は35%である。これは、従来手法が合成対象としていない施設などの世帯に該当する病院・療養所の入院者や社会施設の入所者に該当する高齢者が多いことが原因である。そのため、従来手法により合成された個票データを用

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個票である。

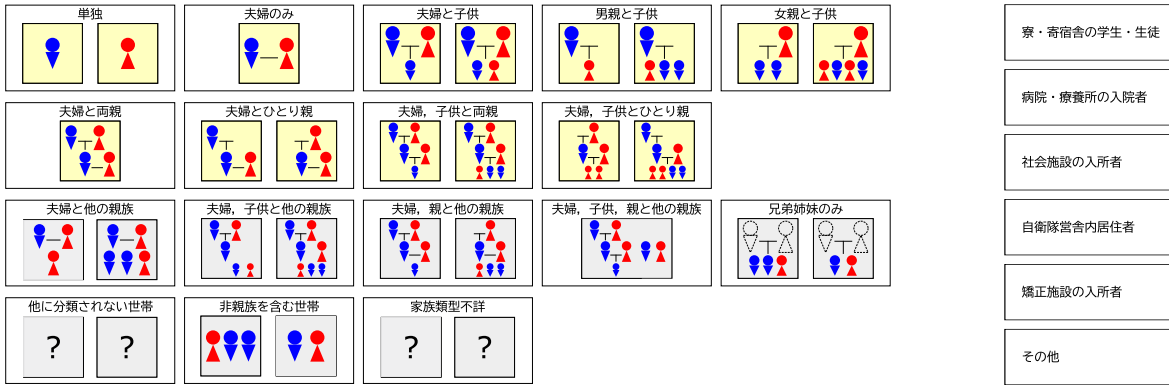


Fig. 1: 国勢調査における世帯の分類

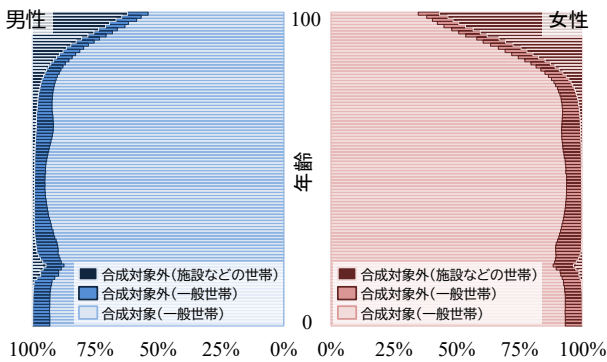


Fig. 2: 従来手法の年齢別人口カバー率^{11)–13)}

いたシミュレーションにおいて、高齢者を分析対象とする場合は留意が必要である。また、16歳から22歳にかけて従来手法の人口カバー率が減少している。これは、施設などの世帯に該当する寮・寄宿舎から高等学校や大学などに通学する学生を合成できていないことが原因である。

そこで本研究では文献¹³⁾の手法を改良し、一般世帯に分類される17種の家族類型と施設などの世帯に分類される6種類を対象とする合成手法を提案する。本研究により国勢調査結果で集計されるすべての人口・世帯を合成することが可能となる。

2 従来手法

従来手法¹³⁾を含む池田らが提案した合成手法¹⁰⁾とその改良手法^{11)–13)}は、公開されている複数の統計表に適合するような個票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、SA法を用いて誤差を最小化している。

従来手法¹³⁾では、合成対象の都道府県下の市区町村を一度に合成することで、都道府県単位で集計されている統計表や市・区を対象に集計される市部の統計表、町・村を対象に集計される郡部、市区町村単位で集計される統計表に整合する個票データを合成している。従来手法¹³⁾は大きく分けて3つの要素からなる。それぞれ、初期世帯合成法、目的関数、近傍解生成である。本節ではこれらの3つの要素を簡単に説明する。

2.1 要素1 初期世帯合成法

従来手法¹³⁾の初期世帯合成法では、合成対象の都道府県下の市区町村を一度に合成している。その際に、人口20万人以上の市区と人口20万人未満の市町村では利用可能な統計表が異なるため、それぞれ異なる手法で初期世帯を合成している。

2.1.1 人口20万人以上の市区の初期世帯合成

従来手法¹³⁾で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統計表を用いて世帯数や人口などの統計情報通りに初期世帯を合成している。従来手法¹³⁾では、家族類型、世帯人員別世帯数の統計表である国勢調査人口等基本集計¹⁴⁾表11を用いている。これにより、合成した世帯数と国勢調査結果と相違なく初期世帯を合成可能である。また、個人の性別と初期の年齢を設定する際には、国勢調査人口等基本集計¹⁴⁾表16-1を用いている。人口等基本集計表16-1は家族類型別、男女別に1歳階級の人口が記されている。従来手法¹³⁾では人口等基本集計表16-1と相違が発生しないように初期世帯の性別を設定し、男女別1歳階級の人口分布と相違が発生しないように年齢を設定する。詳細は文献^{12),13)}を参照されたい。

2.1.2 人口20万人未満の市町村の初期世帯合成

一方、人口20万人未満の市町村では、性別と初期の年齢設定に用いる国勢調査人口等基本集計¹⁴⁾表16-1が公開されていない。そのため、従来手法¹³⁾では、

$$ps_{t,g,a}^Q = \sum_q \hat{p}s_{t,g,a}^q \quad (1)$$

$$ps_{t,g,a5}^q = \sum_a \hat{p}s_{t,g,a}^q \quad (2)$$

を満たす、表16-1の代替となる統計量 $\hat{p}s_{t,g,a}^q$ を求めている。ここで、 $ps_{t,g,a}^Q$ は都道府県 Q に属する人口20万人未満の市町村 q の家族類型 t 、性別 g 、年齢 a の人口、 $\hat{p}s_{t,g,a}^q$ は求める人口20万人未満の市町村 q の家族類型 t 、性別 g 、年齢 a の人口、 $ps_{t,g,a5}^q$ は国勢調査人口等基本集計¹⁴⁾表16-2に集計されている人口20万人未満の市町村 q における家族類型 t 、性別 g 、5歳階級 $a5$ の人口である。 $\hat{p}s_{t,g,a}^q$ は人口20万人未満の市町村

q おける家族類型 t , 性別 g , 年齢 a の係数 $\alpha_{t,g,a}^q$ を使用し次式で算出している.

$$\hat{p}s_{t,g,a}^q = \text{Round}(ps_{t,g,a5}^q \times \alpha_{t,g,a}^q) \quad (3)$$

5 歳階級の人口である $ps_{t,g,a5}^q$ を階級 $a5$ の 1 歳階級の係数である $\alpha_{t,g,a}^q$ を掛け合わせることで 1 歳階級の人口 $\hat{p}s_{t,g,a}^q$ を求めている. ここで Round は四捨五入する関数である². 式 (3) に用いる係数 $\alpha_{t,g,a}^q$ は次式で求めている.

$$\alpha_{t,g,a}^q = \hat{p}s_{t,g,a}^Q / \sum_{a \in a5} \hat{p}s_{t,g,a}^Q \quad (4)$$

ここで, $\hat{p}s_{t,g,a}^Q$ は都道府県 Q に属する人口 20 万人未満の市町村 q , 家族類型 t , 性別 g , 年齢 a の人口 $ps_{t,g,a}^Q$ を初期値にもつ変数である. この $\hat{p}s_{t,g,a}^Q$ を人口 20 万人未満の市町村 q の初期世帯を合成するたびに次式を用いて値を更新している. この操作により, 都道府県 Q の人口分布と相違ない初期世帯が合成可能となる.

$$\hat{p}s_{t,g,a}^Q = \hat{p}s_{t,g,a}^Q - \hat{p}s_{t,g,a}^q \quad (5)$$

従って, 都道府県 Q の初期世帯合成後には,

$$\sum_a \hat{p}s_{t,g,a}^Q = 0 \quad (6)$$

となる. 詳細は文献¹³⁾を参照されたい.

2.2 要素 2 目的関数

従来手法¹³⁾では以下の目的関数を用いて最適化を行っている.

$$f(A) = \sum_s^S f_s(A) \quad (7)$$

ここで, 式 (7) の A は合成データ, S は統計表の数である. $f_s(A)$ は実統計表の各項目と合成データから集計される人工統計表の各項目の差を定式化した関数であり, 従来手法¹³⁾では次式の絶対誤差を用いている.

$$f_s(A) = \sum_j^{G_s} |v_{sj} - r_{sj}| \quad (8)$$

ここで, G_s は統計表 s の項目数, v_{sj} は人工統計表 s の項目 j の統計量, r_{sj} は実統計表 s の項目 j の統計量である.

また, 従来手法¹³⁾では集計対象が異なる以下の実統計表を用いて目的関数値を計算している.

都道府県単位で集計されている統計表

P-1 父子の年齢差³

P-2 母子の年齢差⁴

²なお, この Round 関数は $ps_{t,g,a5}^q = \sum_a \hat{p}s_{t,g,a}^q$ を満たすように, 四捨五入した要素を ± 1 し調整している.

³人口動態職業・産業別統計 出生¹⁵⁾表 1, 合成対象の都道府県を抽出.

⁴人口動態統計 出生¹⁶⁾表 5-2, 合成対象の都道府県を抽出.

人口 20 万人以上の市区を対象とする統計表

BC_x-1 夫婦の年齢差⁵

BC_x-2 9 種類の家族類型別, 男女別, 人口分布⁶

ここで, x は合成対象の都道府県下のうち人口 20 万人以上の市区である.

人口 20 万人未満の市町村を対象とする統計表

SC_y-1 9 種類の家族類型別, 男女別, 人口分布⁶

SC_y-2 男女別の人口分布⁷

ここで, y は合成対象の都道府県下のうち人口 20 万人未満の市町村である. 人口 20 万人未満の市町村では 9 種類の家族類型別, 男女別, 人口分布が 5 歳階級でのみ公開されている. 1 歳階級の人口分布と整合させるために統計表 SC_y-2 男女別の人口分布を調整して使用している. 統計表 BC_x-1 で用いていた夫婦年齢差の統計表は人口 20 万人未満の市町村では公開されていない. 従って, 都道府県下の市区を対象に集計された市部の統計表と町村を対象に集計された郡部の統計表を用いる.

市部・郡部の市町村を対象とする統計表

A-1 市部を対象とする夫婦の年齢差⁵

A-2 郡部を対象とする夫婦の年齢差⁵

2.3 要素 3 近傍解生成

従来手法¹³⁾では以下の手続きにより近傍解を生成している.

Step 1 合成データを初期生成

Step 2 探索回数が規定数に達するか, $f(A) = 0$ になるまで探索する

Step 3-1 合成データ内の個人をランダムに 1 人選択

Step 3-2 後述する条件に該当する個人をランダムに 1 人選択

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

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

Step 3-2 では, Step 3-1 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 3-1 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 3-2 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 3-1 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.

⁵国勢調査 人口等基本集計¹⁴⁾表 17, -70 歳~+70 歳の 1 歳階級の年齢差毎の夫婦の数を人口 20 万人以上の市区と市部・郡部のみ抽出し, 年齢差を計算. なお, 統計表 A-1 は統計表 BC_x-1 の統計量を取り除いている.

⁶国勢調査 人口等基本集計¹⁴⁾表 16-1 もしくは表 16-2 より年齢不詳の項目を除いて調整し, 家族類型, 男女別の人口を抽出. 人口 20 万人以上の市区は 1 歳階級, 人口 20 万人未満の市町村は 5 歳階級である.

⁷国勢調査 人口等基本集計¹⁴⁾表 3-2, 1 歳階級の人口分布を人口 20 万人未満の市町村のみ抽出し, 同じ市町村の統計表 SC_y-1 の年齢別の人口及び全市区町村の年齢別の人口の総和が都道府県の 9 種類の家族類型の男女別 1 歳階級の人口分布と整合するように調整.

- Step 3-1 で選択された個人と同じ市区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、Step 3-1 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

ここで、同じ年代とは、0 歳～4 歳、5 歳～9 歳など、5 歳階級の年齢である。

3 提案手法

本研究では国勢調査の世帯分類で一般世帯に分類される 17 種の家族類型と、施設などの世帯に分類される 6 種類を合成する。そのために、9 種類の家族類型を合成対象とする従来手法¹³⁾を拡張する。

3.1 要素 1 初期世帯合成法

要素 1 では以下の順に初期世帯を合成する。

1. 従来手法¹³⁾が合成対象としている 9 種類の家族類型
2. 世帯内に他の親族が含まれる「夫婦と他の親族」、「夫婦、子供と他の親族」、「夫婦、親と他の親族」、「夫婦、子供、親と他の親族」の 4 種類の家族類型
3. その他の 4 種類の家族類型
4. 施設などの世帯に分類される 6 種類

3.1.1 従来手法¹³⁾の対象である 9 種類の家族類型の合成

従来手法¹³⁾が合成対象としている 9 種類の家族類型は基本的に従来手法¹³⁾と同様に合成する。しかし、従来手法¹³⁾では年齢不詳を除いていたが、本研究では全人口を合成するため、年齢不詳の個人も合成する。そのために、性別を設定する際には 0 歳から 100 歳以上の人口に加え年齢不詳の男女別人口を算出し、統計表と一致するように設定する。初期の年齢は 0 歳から 100 歳以上の人口に加え年齢不詳の人口と一致するように、人口 20 万人以上の市区の場合は 2.1.1 項の手法で、人口 20 万人未満の市町村は 2.1.2 項の手法で各個人へ初期年齢を設定する。

3.1.2 他の親族を含む家族類型の合成

他の親族を含む家族類型は以下の手順で初期世帯を合成する。ここで、他の親族とは、世帯主（または夫婦）からみて、親や子供以外の親族（おじ、おば、成人後の兄弟姉妹、甥、姪など）である。

1. 他の親族を含む 4 種類の家族類型を世帯人員別に規定数世帯を合成する。
2. 各個人に世帯内の役割を設定する。
3. 各個人に性別を設定する。
4. 各個人に初期年齢を設定する。

1. は家族類型別、世帯人員別、世帯数が集計されている国勢調査 人口等基本集計¹⁴⁾表 11 と相違がないように他の親族を含む 4 種類の家族類型を世帯人員別に合成する。

2. では夫婦、親、子供、他の親族の順に個人の世帯内の役割を設定する。なお、他の親族を含む 4 種類の家族類型に上述のいずれかの役割が存在しない場合は、その役割の割り当てをスキップする。まず、夫婦を割

り当てる。他の親族を含む 4 種類の家族類型いずれも夫婦が 1 組存在するため、各世帯に夫と妻を 1 人ずつ設定する。

次に親を設定する。「夫婦、親と他の親族」、「夫婦、子供、親と他の親族」世帯は各世帯にひとり親か両親のいずれかが含まれる。そのため、各世帯の親の人員数を決定する必要がある。国勢調査 人口等基本集計¹⁴⁾表 11 に記述されている「夫婦と両親」、「夫婦、子供と両親」の世帯数の合計と「夫婦とひとり親」、「夫婦、子供とひとり親」の世帯数の合計の割合に従って確率的に親の人員数を決定する。ただし、「夫婦、親と他の親族」世帯の場合は他の親族が最低 1 人、「夫婦、子供、親と他の親族」世帯の場合は子供と他の親族が最低 1 人ずつ存在する。そのため、「夫婦、親と他の親族」世帯における世帯人員数が 4 人の世帯と「夫婦、子供、親と他の親族」世帯における世帯人員数が 5 人の世帯はひとり親と設定する。

次に子供を決定する。子供は国勢調査 世帯構造等基本集計¹⁷⁾表 8-3 を用いる。表 8-3 には 3 区分の家族類型、子供の数別世帯数が集計されている。3 区分の家族類型は「夫婦のいる核家族世帯」、「夫婦のいるその他の世帯（同居の親あり）」、「夫婦のいるその他の世帯（同居の親なし）」が集計されている。それぞれ概ね「夫婦と子供」世帯、「夫婦、子供、親と他の親族」世帯、「夫婦、子供と他の親族世帯」に該当する。「夫婦、子供と他の親族世帯」世帯は表 8-3 の「夫婦のいるその他の世帯（同居の親なし）」の子供の数別世帯数と整合するように世帯内の子供の数を決定する。「夫婦、子供、親と他の親族」世帯は同様に「夫婦のいるその他の世帯（同居の親あり）」を用いて子供の数を決定する。なお、最低 1 人は他の親族が含まれるように設定する。

最後に他の親族を決定する。1. から 3. で世帯内の役割が未決定の世帯人員をすべて他の親族として設定する。

3. の個人の性別を決定では、まず、世帯内の役割に応じて性別が決定する夫婦と両親の性別を設定する。次に決定した男女別の人口を人口 20 万人以上の市区は国勢調査 人口等基本集計¹⁴⁾表 16-1 を、人口 20 万人未満の市町村は表 16-2 を差し引き、残りの世帯人員へ差し引いた統計表と一致するように性別を設定する。

最後に 4. の初期年齢の決定では、3.1.1 項の手法で各個人へ初期年齢を設定する。

3.1.3 その他の家族類型の合成

その他の「兄弟姉妹のみ」、「他に分類されない核家族以外の親族のみ」、「非親族を含む」、「家族類型不詳」世帯は以下の手順で初期世帯を合成する。

1. その他の 4 種類の家族類型を世帯人員別に規定数世帯を合成する。
2. 各個人に性別を設定する。
3. 各個人に初期年齢を設定する。

1. は家族類型別、世帯人員別、世帯数が集計されている国勢調査 人口等基本集計¹⁴⁾表 11 と相違がないように他の親族を含む 4 種類の家族類型を世帯人員別に合成する。2. では、人口 20 万人以上の市区は国勢調査 人口等基本集計¹⁴⁾表 16-1 を、人口 20 万人未満の市町村は表 16-2 の家族累計別、男女別人口と一致する

ように性別を決定する。3. では、0 歳から 100 歳と不詳の 102 項目の統計表と一致するように、人口 20 万人以上の市区の場合は 2.1.1 項の手法で、人口 20 万人未満の市町村は 2.1.2 項の手法で各個人へ初期年齢を設定する。なお、その他の 4 種類の家族類型では各個人に世帯内の役割を決定することが困難であるため、世帯内の役割なしと設定した。

3.1.4 施設などの世帯の合成

施設などの世帯は一般世帯とは集計単位が異なり、「寮・寄宿舎の学生・生徒」や「病院・療養所の入院者」、「社会施設の入所者」は棟を、「自衛隊営舎内居住者」は中隊または艦船を、「矯正施設の入所者」は建物を、「その他」は 1 人 1 人を世帯としている。100 人以上存在する世帯が存在しうするため、施設などの世帯別、世帯人員別世帯数が集計されている国勢調査 人口等基本集計¹⁴⁾ 表 7 では、世帯人員の階級を「1~4 人」、「5~29 人」、「30~49 人」、「50 人以上」としている。施設などの世帯を合成するためには、世帯数とその人口が統計表と一致するように各世帯の人員を決定する必要がある。そのために本研究では、表 7 に記載されている施設などの世帯別、世帯人員別に以下の条件別に各世帯の世帯人員を決定する。

1. 世帯数が 1 の場合、その世帯人員数をもとに初期世帯を合成
2. 世帯数と世帯人員数が一致する場合、1 人世帯として初期世帯を合成
3. 上記以外の場合、表 7 と一致するように確率的に世帯人員決定し初期世帯を合成
 - (a) 施設などの世帯 i の階級 c の世帯数 H_i^c と世帯人員数 P_i^c を使用し、平均世帯人員数 \bar{p}_i^c を算出する
 - (b) 世帯人員階級の上限 u^c と下限 l^c を使用し、設定する世帯人員の範囲を $r_i^c = \min(u^c - \bar{p}_i^c, \bar{p}_i^c - l^c)$ で決定する
 - (c) $\bar{p}_i^c \pm r_i^c$ の範囲で施設などの世帯 i の階級 c の世帯 n の世帯人員数 $m_{i,n}^c$ を決定する
 - (d) 施設などの世帯 i の階級 c の世帯数 H_i^{c-1} となるまで (c) に戻る
 - (e) 最後の世帯の世帯人員数は $P_i^c - \sum_n^{H_i^{c-1}} m_{i,n}^c$ とする

初期世帯合成後、各個人の性別と年齢は人口 20 万人以上の市区は国勢調査 人口等基本集計¹⁴⁾ 表 16-1 と表 3-2 を、人口 20 万人未満の市町村は表 16-2 と表 3-2 を用いて施設などの世帯に属する男女別人口を算出し、3.1.1 項の手法により設定する。

3.2 要素 2 目的関数

従来手法¹³⁾ では式 (7) の計算に絶対誤差である式 (8) を用いている。絶対誤差の目的関数である式 (8) は実統計表との誤差が直感的に理解しやすい。しかし、合成データの利活用先を考えると、同じ誤差の合成データにおいて、特定の項目に多くの誤差をもつ合成データより誤差が平均的に発生した合成データの方が好ましい。そのため、本研究では次式に示す 2 乗誤差を用いて最適化する。

$$f_s(A) = \sum_j^{G_s} (v_{sj} - r_{sj})^2 \quad (9)$$

なお、本研究では、統計表との誤差を示す際には式 (8) を用いる。

また、提案手法では従来手法¹³⁾ の実統計表に加え、以下の統計表を用いて目的関数値を計算している。なお、統計表番号の後ろの ' は従来手法¹³⁾ の統計表を変更したことを意味する。

人口 20 万人以上の市区を対象とする統計表

- BC_x-2' 17 種類の家族類型別、男女別、人口分布⁸
- BC_x-3 施設などの世帯の男女別人口分布⁹

人口 20 万人未満の市町村を対象とする統計表

- SC_y-1' 17 種類の家族類型別、男女別、人口分布⁸
- SC_y-2' 男女別の人口分布¹⁰
- SC_y-3 施設などの世帯の男女別人口分布⁹

従来手法¹³⁾ において統計表 SC_y-2 は、対象とする都道府県における合成対象の 9 種類の家族類型の男女別、1 歳階級の人口分布と整合するように人口 20 万人未満の市町村の統計表を調整している。本研究では、すべての人口を対象とするため、従来手法¹³⁾ で実施していた統計表の調整をせずに用いる。

3.3 要素 3 近傍解生成

本研究では 17 種類の家族類型と 6 種類の施設などの世帯を考慮し近傍解を生成するため、近傍解生成法を従来手法¹³⁾ から一部変更する。従来手法¹³⁾ の Step 3-2 では、Step 3-1 で選択した個人によって選択する個人の条件を変更している。その際に従来手法¹³⁾ では 9 種類の家族類型のみを対象としている。本研究では、17 種類の家族類型と 6 種類の施設などの世帯を考慮するために、Step 3-1 で選択された個人が人口 20 万人以上の市区に属する一般世帯の場合、Step 3-2 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。また、Step 3-1 で選択された個人が人口 20 万人以上の市区に属する施設などの世帯の場合、Step 3-2 では、選択されている個人と同じ市区かつ施設などの世帯に属する同じ性別の個人をランダムに選択する。

一方、Step 3-1 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 3-1 で選択された個人が一般世帯に属する場合
 - Step 3-1 で選択された個人と同じ市区町村かつ家族類型かつ性別の個人を選択
 - 人口 20 万人未満の市町村に属する個人のうち、Step 3-1 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を選択

⁸国勢調査 人口等基本集計¹⁴⁾ 表 16-1 もしくは表 16-2 より家族類型、男女別の人口を抽出。人口 20 万人以上の市区は 1 歳階級、人口 20 万人未満の市町村は 5 歳階級である。

⁹国勢調査 人口等基本集計¹⁴⁾ 表 3-2 と表 16-1 もしくは表 16-2 の男女別の人口を抽出し減算。人口 20 万人以上の市区は 1 歳階級、人口 20 万人未満の市町村は 5 歳階級である。

¹⁰国勢調査 人口等基本集計¹⁴⁾ 表 3-2 から男女別の人口を抽出。

Table 1: 目的関数に用いる統計表数と項目数 G_s ¹¹

統計表	従来手法 ¹³⁾		提案手法	
	表数	項目数 G_s	表数	項目数 G_s
P-1	1	24	1	24
P-2	1	24	1	24
BC _x -1	1	241	1	241
BC _x -2	1	1,818	1	3,468
BC _x -3	-	-	1	204
SC _y -1	18	6,804	18	13,464
SC _y -2	18	3,636	18	3,672
SC _y -3	-	-	18	792
A-1	1	241	1	241
A-2	1	241	1	241
合計	42	13,029	61	22,130

Table 2: 式 (8) による目的関数値

探索回数 (回/人)	従来手法 ¹³⁾		提案手法	
	Avg.	S.D.	Avg.	S.D.
初期解	620,730.8	526.9	664,868.2	381.5
1	485,077.2	643.2	518,105.8	732.9
10	187,053.8	513.9	199,858.8	403.7
100	28,703.4	190.8	31,311.8	168.7
1,000	8,936.8	119.7	9,892.4	118.4
10,000	3,700.0	88.2	3,848.8	127.6
100,000	2,312.4	50.2	2,246.0	61.4

- Step 3-1 で選択された個人が施設などの世帯に属する場合

- Step 3-1 で選択された個人と同じ市区町村かつ施設などの世帯に属する同じ性別の個人を選択
- 人口 20 万人未満の市町村に属する個人のうち、施設などの世帯に属する Step 3-1 で選択された個人と同じ性別かつ年代の個人を選択

4 実験結果

本研究では、平成 27 年国勢調査結果を用いて鳥取県を対象に世帯構成の合成を行う。従来手法¹³⁾の合成対象は 198,735 世帯、その人口は 493,557 人である。提案手法の合成対象は 213,395 世帯、その人口は 573,441 人である。鳥取県は 19 の市町村からなっており、鳥取市を除く 18 の市町村は人口 20 万人未満である。Table 1 に従来手法¹³⁾と提案手法が目的関数の計算に用いる統計表とその数を示す。

SA 法の設定は従来手法¹³⁾と同様に初期温度を 1.0、収束温度を 0.1 と設定し、冷却関数には指数冷却を用いた。探索回数は 1 人あたり 1 回、10 回、100 回、1,000 回、10,000 回、100,000 回とした。従来手法¹³⁾の総探索回数はそれぞれ、493,557 回、4,935,570 回、49,355,700 回、493,557,000 回、4,935,570,000 回、49,355,700,000 回である。提案手法の総探索回数はそれぞれ、573,441 回、5,734,410 回、57,344,100 回、573,441,000 回、5,734,410,000 回、57,344,100,000 回である。なお、本研究が用いた計算機の CPU は AMD Ryzen 3950X (3.5 GHz, 16 コア) で、メモリは DDR4-3200 32 GB × 2、OS は Microsoft Windows 10 Pro 64 bit である。

Table 3: 式 (8) による統計表別目的関数値¹¹

統計表	初期解		100,000 回/人	
	従来 ¹³⁾	提案	従来 ¹³⁾	提案
P-1	185,384.0	198,434.6	4.6	3.2
P-2	236,493.6	251,562.2	3.8	2.6
BC _x -1	60,008.8	63,787.8	691.8	655.6
BC _x -2	0.0	0.0	0.0	0.0
BC _x -3	-	0.0	-	0.0
SC _y -1	0.0	0.0	0.0	0.0
SC _y -2	19,746.0	22,558.0	0.0	0.0
SC _y -3	-	0.0	-	0.0
A-1	71,863.4	76,062.2	633.2	613.4
A-2	47,235.0	52,463.4	979.0	971.2
合計	620,730.8	664,868.2	2,312.4	2,246.0

Table 4: 処理時間 (秒)

探索回数 (回/人)	従来手法 ¹³⁾		提案手法	
	Avg.	S.D.	Avg.	S.D.
1	4.1	0.3	4.3	0.4
10	19.3	0.9	22.4	1.2
100	181.2	9.1	202.1	8.3
1,000	1,697.8	74.7	1,967.7	77.9
10,000	17,107.8	928.8	20,578.9	834.3
100,000	166,891.5	9,543.8	194,775.1	9,530.9

式 (8) による統計表との誤差を Table 2 に、初期解と探索回数 1 人あたり 100,000 回の統計表別の誤差を Table 3 に示す。なお、試行回数は 10 回であり、Table 2 は平均と標準偏差を、Table 3 は平均を示している。また、Table 3 の統計表 SC_y-1、SC_y-2、SC_y-3 は鳥取県の人口 20 万人未満の 18 市町村の目的関数値を足し合わせている。鳥取県における人口 20 万人以上の市区は鳥取市のみであるため、統計表 BC_x-1、BC_x-2、BC_x-3 は鳥取市の目的関数値である。

Table 2 から探索回数 1 人あたり 10,000 回までは従来手法¹³⁾が誤差を削減できている。Welch の t 検定を実施したところいずれも有意差があった。一方で、探索回数 1 人あたり 100,000 回では提案手法が誤差を削減できており、有意差があった。しかし、従来手法¹³⁾と提案手法では合成対象の人口が増加している。1 人あたりの誤差を算出すると、初期解は従来手法¹³⁾は 1.2577、提案手法は 1.1594、従来手法¹³⁾における探索回数 1 人あたり 1 回から 100,000 回では、それぞれ、0.9828, 0.3790, 0.0582, 0.0047, 0.0047, 提案手法では、それぞれ、0.9035, 0.3485, 0.0546, 0.0067, 0.0039 である。従って、1 人あたりの誤差では提案手法が従来手法¹³⁾より誤差を削減できている。また、十分に探索すると従来手法¹³⁾と同程度以下の誤差の人口を合成できることがわかる。

統計表別の誤差である Table 3 から、十分に探索すると従来手法¹³⁾、提案手法ともに人口 20 万人未満の市町村の 1 歳階級の人口分布である統計表 SC_y-2 と相違なく合成できている。また、1 人あたり 100,000 回探索した場合においても父子年齢差である統計表 P-1、母子年齢差である統計表 P-2、夫婦年齢差である統計表 BC_x-1、A-1、A-2 は誤差が残っている。しかし、従

¹¹ 提案手法では統計表に BC_x-2', SC_y-1', SC_y-2' を用いているが、紙面の都合上、' を省略している。

来手法¹³⁾、提案手法ともいずれの統計表も同程度の誤差となっており、提案手法は従来手法¹³⁾と同程度の探索性能を有している。

従来手法¹³⁾と提案手法の処理時間を Table 4 に示す。なお、従来手法¹³⁾と提案手法は総探索回数が異なることに留意されたい。Table 4 から従来手法¹³⁾の1秒あたりの探索回数は295,735.3回、提案手法の1秒あたりの探索回数は294,411.8回である。提案手法は従来手法¹³⁾と同程度の処理速度である。

5 おわりに

本研究では国勢調査の全世帯を合成する手法を提案した。従来手法¹³⁾では Fig. 1(a) の黄色で示す9種類の家族類型を対象に合成されていた。従来手法¹³⁾で世帯は95%、人口は90%カバーできている。しかし、Fig. 2 に示す年齢別のカバー率から、従来手法¹³⁾は高齢者になるほど合成できおらず、高齢者を分析対象とするシミュレーションを実施する場合には留意が必要であった。そこで、本研究では国勢調査で集計されている全世帯を合成するために、従来手法¹³⁾を構成する3つの要素である初期世帯生成法、目的関数、近傍解生成法を改良した。これにより、世帯、人口ともに100%カバーする人口合成に成功した。したがって、提案手法の年齢別人口カバー率も全ての年齢において100%の人口をカバーできている。

実験結果から、提案手法は合成対象の人口が増加しているが、十分に探索すると従来手法¹³⁾と同程度以下の誤差で人口合成することができた。また、処理時間についても提案手法は従来手法¹³⁾と同程度の時間で最適化できている。

今後の課題として合成データへ新たな属性を追加する手法^{18),19)}を改良する必要がある。合成データへ所得属性を追加する手法¹⁸⁾や居住地属性を追加する手法¹⁹⁾では、従来手法¹³⁾が対象とする9種類の家族類型に特化した手法が開発されている。提案手法では全世帯を合成しているため、提案手法により合成されたデータへ所得属性や居住地属性を追加するためには、これらの手法^{18),19)}を改良する必要がある。しかし、提案手法による合成データにより、これらの手法^{18),19)}の9種類の家族類型に合わせた統計表の調整が不要となり、より精緻な属性の追加に期待できる。

謝辞

本研究の一部は、JSPS 科研費 19K23229, JST 未来社会創造事業 JPMJMI20B3, 青山学院大学総合研究所, 2020 年度関西大学研究拠点支援経費研究課題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**, 407/412 (2018)

- 5) Y. Goto: Stylized Fact Analysis of Cash-for-Work Programs in the Disaster Reconstruction Process, in *Proc. of 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) 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)
- 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) 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).
- 10) 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)
- 11) 原田, 村田, 栴井: 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)
- 12) T. Murata, T. Harada, and D. Masui: Comparing Transition Procedures in Modified Simulated-Annealing-Based Synthetic Reconstruction Method without Samples, *SICE JCMSI*, **10-6**, 513/519 (2017)
- 13) T. Murata and T. Harada: Synthetic Method for Population of A Prefecture Using Statistics of Local Governments, in *Proc. of 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 1175/1180 (2018)
- 14) 総務省統計局: e-Stat 平成 27 年度国勢調査 人口等基本集計, <https://www.e-stat.go.jp/stat-search/files?page=1&layout=datalist&toukei=00200521&tstat=000001080615&tclass1=000001089055> (2016)
- 15) 総務省統計局: e-Stat 人口動態職業・産業別統計 保管表 出生年度次 2015 年度, <https://www.e-stat.go.jp/stat-search/files?page=1&layout=datalist&toukei=00450011&tstat=000001028897&cycle=8&year=20151&month=0&tclass1=000001053122&tclass2=000001053131&tclass3=000001053132> (2018)
- 16) 総務省統計局: e-Stat 人口動態統計 確定数保管統計表 都道府県編 (報告書非掲載表) 出生年次 2015 年, <https://www.e-stat.go.jp/stat-search/files?page=1&layout=datalist&toukei=00450011&tstat=000001028897&cycle=7&year=20150&month=0&tclass1=000001053058&tclass2=000001053061&tclass3=000001053074&tclass4=000001053084> (2016)
- 17) 総務省統計局: e-Stat 平成 27 年度国勢調査世帯構造等基本集計 (母子・父子世帯, 親子の同居など), <https://www.e-stat.go.jp/stat-search/files?page=1&layout=datalist&toukei=00200521&tstat=000001080615&tclass1=000001104835> (2017)
- 18) 杉浦, 村田, 原田: 賃金構造基本統計調査に基づく合成人口の労働者への就業属性別の所得の割当て, システム制御情報学会論文誌, **32-2**, 70/79 (2019)
- 19) T. Harada and T. Murata: Projecting Households of Synthetic Population on Buildings Using Fundamental Geospatial Data, *SICE JCMSI*, **10-6**, 505/512 (2017)