

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

○ 梶井大貴 村田忠彦 (関西大学)

Reconstruct of Agent's Attributes by Simulated Annealing for Minimizing Errors from Statistics

*D. Masui and T. Murata (Kansai University)

Abstract— In social science, it is one of absolute must to design a state in our society precisely. We can re-create a more realistic virtual society by designing properties or characteristics of citizens according to the statistics. Since real individual data is not available usually, we should try to reconstruct individual data from statistics that are open to public by using a method estimating demographic. We employ a simulated annealing method with a heuristic that minimizes the number of citizens that differs from real statistics.

Key Words: Social Simulation, Statistics, Simulated Annealing

1 はじめに

社会科学においては、社会の姿をできるだけ精密に記述することが至上命題の一つとされている¹⁾。社会の中で重要な要素である人を忠実に設計することで、より現実的な仮想社会を再現することができる。人を忠実に表現するには現実のデータを用いることが必要となるが、個人に関するデータを容易に入手することはできない。そこで、必要となる個人のデータを人口推計の手法を用いて公開されている統計データから復元する。そうすることでシミュレーションによる実験から信頼性の高い結果を得ることができる。

調査に基づく個人データを用いる方法の一つに、独立行政法人統計センターが提供している公的統計のマイクロデータ利用²⁾というサービスがある。これは公的機関によって管理されている実統計のマイクロデータを研究などで利用できるものである。利用の申請をすることで個人データを用いてシミュレーションを行うことができる。しかしながら、これらの個人データは当然、厳重に管理する必要がある、また利用期間に制限が設けられている。したがって、このサービスは安易に利用できるものではない。また、実統計は個人データから集計されたデータになっているので、必要となる個人データを利用することはできない。

そこで本研究では、現実社会に存在し得る仮想的な人口社会を再現するために、実統計のデータを選択しそれらに適合するような個人データを推計する。なお、推計する個人データはエージェントを用いて記述する。実統計と、推計された個人データの集合からとった推計統計との誤差を計算する目的関数を設計し、その誤差が最小となるような各エージェントの属性値の組み合わせ探索する。著者らは池田ら³⁾の論文で提案されていた平均誤差に基づく目的関数とは異なる、誤差人数に基づく目的関数を提案した⁴⁾。これにより実統計データに偏ることなく誤差人数を最小化できることを示した。本研究では誤差人数に基づく目的関数を用いる。目的関数値の最小化はSA (Simulated Annealing) を用いて最適化し、年齢の変更によって状態遷移を行う。池田ら³⁾、および福田ら⁵⁾では年齢をランダムに変更していたのに対し、本研究では誤差人数、すなわ

ち人口分布の実統計のデータの誤差の変化を考慮した年齢変更を提案する。提案した年齢変更の手法を既存の手法と比較し、その有効性について考察する。なお、以降で「統計データ」は実統計に基づくデータを指し、「推計データ」とは推計した個人データの集合を表す。

2 モデル

2.1 推計データ

本研究における解となる推計データは多数の世帯から構成されており、先行研究と同様に、統計データを元に Fig. 1 のような9種類の世帯 (単独・夫婦・夫婦と子・父子・母子・夫婦と両親・夫婦と片親・夫婦と子と両親・夫婦と子と片親) を扱う。統計データには他の種類の世帯も存在するが数が少ないことや世帯構成が複雑なことなどから存在しないものとする。推計データの規模は世帯の数によって決定し、初期生成時にあらかじめパラメータ H として規定しておく。世帯には年齢、性別、親族関係を表す情報を持つエージェントが属している。各エージェントが親族関係の情報を持つことで、一体のエージェントの父親や子供などに

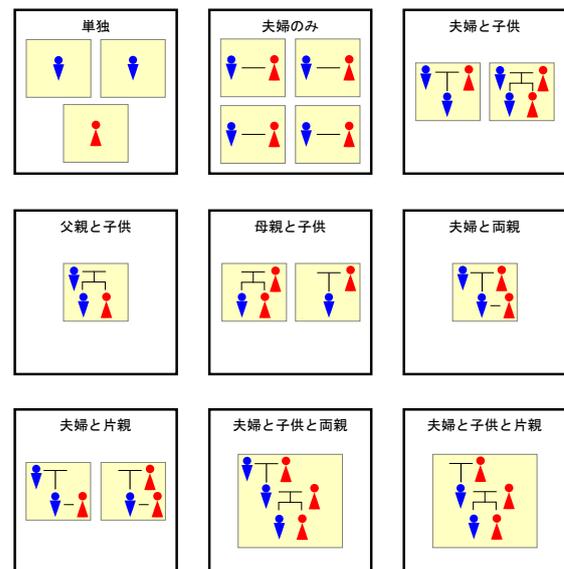


Fig. 1: 世帯の種類

該当するエージェントを多数のエージェントの中から特定することができる。また、エージェントは自身が属する世帯の種類や世帯の中での役割（単独・父・母・夫・妻・子）の情報を持っている。これらは目的関数を用いて統計データとの誤差を計算する時に必要な情報である。最適化によって変更を行うのはエージェントの年齢のみなので、最適化の過程で世帯の種類や世帯での役割が変わることはない。

2.2 目的関数

本研究では、池田ら³⁾によって提案された目的関数をベースに構造を少しだけ変化させた目的関数を用いる。まず、提案された目的関数について記述する。

1. 父子の年齢差（表 4-13, 2013 年⁶⁾）
2. 母子の年齢差（表 4-8, 2012 年⁶⁾）
3. 夫婦の年齢差（表 9-14, 2010 年⁷⁾）
4. 人口分布（男）（表 2-3, 2012 年⁶⁾）
5. 人口分布（女）（表 2-3, 2012 年⁶⁾）
6. ある年齢の人が単独世帯に属する割合（男）（表 7-28, 2013 年⁶⁾）
7. ある年齢の人が単独世帯に属する割合（女）（表 7-28, 2013 年⁶⁾）
8. ある年齢の人が夫婦のみ世帯に属する割合（男）（表 7-28, 2013 年⁶⁾）
9. ある年齢の人が夫婦のみ世帯に属する割合（女）（表 7-28, 2013 年⁶⁾）

適合させる統計データは以上の 9 種類を選択する。統計データのモデルとして 1・3・4・5 の統計データを Table 1~2 に示す。Table 1 は父子関係を持つ人の組み合わせの内父子の年齢差の値がどれくらいの割合かを 5 歳区分で表している。Table 2 は夫婦関係を持つ人の組み合わせの内夫婦の年齢差（夫の年齢 - 妻の年齢）の値がどれくらいの割合かを 1 歳区分で表している。Table 3 は男性の年齢がどれくらいの割合かを 1 歳区分で表している。統計データ 1~5 までは各項目の割合の値を合計すると 100（%）になる。Table 4 は 1 段目の項目に着目すると、年齢が 14 歳以下の男性の内単独世帯に属する人の割合を 5 歳区切りで表している。各項目の割合の値を計算する時の分母が異なっているので合計しても 100（%）にはならない。各統計データの割合の値はある項目 j に着目すると、条件 X_j を満たす人や組み合わせの数の内で条件 X_j と条件 Y_j を満たす人や組み合わせの数の割合を表している。目的関数は一つの統計データに対して一つ設計しているので合計 9 つの目的関数を設計することになる。提案された目的関数を式 (1) に示す。

$$f_s^1(A) = \frac{4}{G_s} \sum_{j=1}^{G_s} (c_{sj}(A) - r_{sj} \cdot m_{sj}(A))^2 \quad (1)$$

目的関数の変数は以下の通りである。

Table 1: 父子の年齢差

条件 X	条件 Y	割合 (%)
父子関係	年齢差 ~14	0.0000
父子関係	年齢差 15~19	0.5469
父子関係	年齢差 20~24	8.4852
父子関係	年齢差 25~29	26.1635
父子関係	年齢差 30~34	33.7867
父子関係	年齢差 35~39	20.8889
父子関係	年齢差 40~44	7.9119
父子関係	年齢差 45~49	2.1035
父子関係	年齢差 50~	0.1107

Table 2: 夫婦の年齢差

条件 X	条件 Y	割合 (%)
夫婦関係	年齢差 ~-4	6.1253
夫婦関係	年齢差 -3	3.1166
夫婦関係	年齢差 -2	4.8194
夫婦関係	年齢差 -1	9.5689
夫婦関係	年齢差 0	20.2292
夫婦関係	年齢差 1	13.7581
夫婦関係	年齢差 2	9.9258
夫婦関係	年齢差 3	7.7380
夫婦関係	年齢差 4	6.0384
夫婦関係	年齢差 5	4.7240
夫婦関係	年齢差 6	3.5213
夫婦関係	年齢差 7~	1.0450

Table 3: 人口分布 (男)

条件 X	条件 Y	割合 (%)
男性	年齢 0	0.8669
男性	年齢 1	0.8660
男性	年齢 2	0.8900
...
男性	年齢 98	0.0110
男性	年齢 99	0.0065
男性	年齢 100	0.0095

Table 4: ある年齢の人が単独世帯に属する割合

条件 X	条件 Y	割合 (%)
男性・年齢 ~14	単独世帯	0.0116
男性・年齢 15~19	単独世帯	7.0119
男性・年齢 20~24	単独世帯	27.9853
...
男性・年齢 75~79	単独世帯	10.2981
男性・年齢 80~84	単独世帯	10.9273
男性・年齢 85~	単独世帯	11.7366

A: エージェントの集合である推計データ

S: 統計データの数 ($S = 9$)

G_s : 統計データ s の項目数

m_{sj} : 統計データ s の条件 X_j を満たす人や組み合わせの数

c_{sj} : 統計データ s の条件 X_j と条件 Y_j を満たす人や組み合わせの数

r_{sj} : 統計データ s の項目 j の割合の値

$m_{sj} \cdot r_{sj}$ を計算することで、推計データの規模に合わせた目標の値を計算することができる。この値と c_{sj}

との差をとることで推計データと統計データの誤差を表している。統計データの値は割合なので、仮に推計データの c_{sj} が適切な値になったとしても $m_{sj} \cdot r_{sj}$ を計算すると、最大で 0.5 の誤差が値として生じる。この 0.5 を二乗することで 0.25 となり、4 を掛けることで 1.0 になる。各項目から生じた 1.0 の和を項目数 G_s で正規化している。したがって、目的関数値が 1.0 程度になれば推計データと統計データがほぼ正確に適合していると考えられる。

$$\text{Min} \sum_{s=1}^S f_s^i(x), (i = 1, 2) \quad (2)$$

それぞれの目的関数値の和を小さくすることを目的とした式 (2) の最小化問題を解くことで統計データに適合している推計データを得ることができる。提案された式 (1) をベースに構造を少し変化させた目的関数を式 (3) に示す。

$$f_s^2(A) = \sum_{j=1}^{G_s} |c_{sj}(A) - \text{Round}(r_{sj} \cdot m_{sj}(A))| \quad (3)$$

式 (1) との違いは 2 点ある。1 点目は各項目ごとの誤差を項目数 G_s で平均化していない点である。各統計データの項目数 G_s は統計データごとに異なっており、1 歳区分と 5 歳区分で大小の差がある。統計データ 3・4 は 1 歳区分の年齢なので 101 項目であるのに対し、統計データ 1・2 は 9 項目、統計データ 3 は 12 項目、統計データ 6~9 は 16 項目となっている。項目数が大きい統計データでは、年齢の変更によってエージェント一体が適切な項目に割り当てられた時に改善する目的関数値の量が小さくなるのが分かっている⁴⁾。これにより、項目数の多い統計データとの誤差が考慮されにくくなり項目数の少ない統計データとの誤差が優先的に最小化されることになる。統計データの項目数に関係なく、全ての統計データに対して均等に誤差を最小化させるために項目数 G_s による平均化をせずに計算する構造に変更した。2 点目は誤差を計算する時に $m_{sj} \cdot r_{sj}$ を四捨五入して絶対値を計算している点である。四捨五入せずに実数値のまま二乗して計算すると、誤差がどの程度生じているかを目的関数値から直感的に読み取ることが難しい。誤差を人数で表すために四捨五入し、項目の過不足を誤差として統一するために絶対値を計算している。以上のことから、式 (3) は各統計データの各項目で生じている誤差の人数を表しているの、目的関数値が 0 になれば推計データと統計データが一致していると考えられる。

福田ら⁵⁾の式 (1) を用いた最適化では統計データ 3・4 の目的関数値が 1.0 を超えており、他の目的関数値に比べて大きな値となっていた。そこで統計データ 3・4 に対応する目的関数に重み付けを行い、他の目的関数値よりも重要視することで指標となる 1.0 以下の目的関数値になっていた。他の目的関数値はわずかに増加したものの、全体として誤差を均衡させることができていた。しかし、目的関数値が 1.0 以下になっても、項目数の多い統計データ 3・4 では適合していな

いエージェントが多く残ることが分かっている⁴⁾。統計データとの誤差がより少ない推計データを得るにはそれぞれの統計データに対して均等に最適化させる必要がある。本研究では、統計データの項目数に関係なく全ての統計データに対して誤差を最小化することができる式 (3) を用いて最適化を行う。

3 提案手法

まず、Table 5 の統計データの割合に基づいて 9 種類の世帯をあらかじめパラメータとして規定された世帯数 H の数だけ確率的に生成していく。その際、夫・妻・父・母のいずれかの役割を持つエージェントは性別を設定し、単独・子の役割を持つエージェントについては不定にしておく。9 種類の世帯には子供がふくまれているものもあるが、世帯の種類に関する統計データには子供の数で区分した割合は掲載されていないので、厚生労働省によって調査されたアンケート結果^{8, 9)}を用いて確率的に子供の数を決定する。世帯の生成と同時にエージェント同士の親族関係、世帯での役割、世帯の種類などの目的関数値の計算に必要な情報を設定していく。

世帯の種類	割合 (%)
単独	33.98
夫婦のみ	20.74
夫婦と子供	29.24
父親と子供	1.34
母親と子供	7.81
夫婦と両親	0.47
夫婦と片親	1.48
夫婦と子供と両親	1.86
夫婦と子供と片親	3.07

規定数の世帯が生成された後、全てのエージェントの年齢を 0~100 の範囲で人口分布の統計データ⁶⁾の数値に基づいて確率的に設定する。この時の人口分布の統計データは男女を総合した割合を用いている。その後、性別が不定になっていたエージェントにランダムに性別を設定する。エージェントが持つ属性値を設定し終えた後、目的関数を用いて SA による最適化の処理を行っていく。SA の温度の値は初期温度 1.0 で収束温度 0.0001 になるように線形的に温度を冷却していく。これにより温度が高い内は目的関数値が悪くなる時でも積極的に変更を行い、温度が低くなるにつれて目的関数値が悪くなる変更を行いくくなる。したがって、収束温度に近づくほど目的関数値が小さい推計データを得ることが期待できる。探索回数は式 (4) に基づいて決定する。

$$\text{探索回数} = H \cdot 2.5 \cdot 100 \cdot N \quad (4)$$

H は推計データの初期生成に用いた規定の世帯数で、2.5 は一世帯当たりの平均人員数である。したがって、 $H \cdot 2.5$ で推計データの人口が概算できる。その後には掛けている 100 はエージェントが取り得る 0~100 の年齢を表しており、101 種類の年齢の概数である 100 を用いている。 N はエージェント一体が最適化の中で何回、年齢変更の対象になるかを表しており、 $N = 1$ の時は全てのエージェントが 0~100 の各年齢を 1 回は設定さ

れるということになる。ただし、あくまで期待値であり必ず年齢を順番に総なめするとは限らない。SA による具体的な最適化の流れは以下の通りである。

- step 1. 推計データの目的関数値を計算する。
- step 2. 探索回数が規定数に達していれば探索を終了する。
- step 3. エージェントを一体ランダムに選択し年齢を変更する。
- step 4. 新たな解の目的関数値を計算し改善している時は step 5-1 へ、そうでないときは step 5-2 に進む。
- step 5-1. 保持している目的関数値を改善した値に更新する。
- step 5-2. SA の温度パラメータに従って変更を許可し、保持している目的関数値を更新する。
- step 6. 変更がなかった時は年齢を変更したエージェントを変更前の年齢に戻す。
- step 7. SA の温度を冷却する。
- step 8. step 2 の処理に戻る。

この SA による最適化では、人口分布の統計データの基づいて確率的に年齢の変更を行うが、おおむねランダムな変更なので解が改善・改悪の両方の可能性がある。最適化が進むにつれて目的関数値の改善はおだやかになっていく。乱数を用いた変更では無駄な探索が行われている可能性が十分にある。そこで本研究では目的関数値が改善されやすいようなエージェントの年齢の変更処理を新たに追加する。効率よく探索を行うためには解の状態が変わらずに探索回数の 1 回の消費を避けることが有効と考えられる。それを実現するためには、統計データに対応する目的関数値がどのエージェントの年齢を何歳に変更することで目的関数値が改善されるかを考慮する必要がある。まず、提案した新たな年齢変更の際の対象となる統計データについては、人口分布の統計データとする。他の統計データでは父子、夫婦、特定の年齢など、目的関数の対象となるエージェントが限定されている。人口分布の統計データは 1 歳区切りのデータであるので、エージェントを一体でも異なる年齢を変更すると必ず目的関数値に影響する。どのエージェントを選択しても目的関数値に影響するということは目的関数値の改善が行われやすいということになる。以上のことから、考慮する統計データは人口分布のデータとする。

具体的な手法については、統計データとの誤差が存在する年齢の情報を過剰か不足かを区別して保持することが必要となる。年齢を変更するエージェントを選択する時に、統計データとの誤差が存在し特定の年齢が過剰になっている年齢のエージェントを選択する。次にエージェントの年齢を統計データとの誤差が存在し特定の年齢が不足している年齢に変更する。この変更により、過剰な年齢のエージェントが一体減少し、不足している年齢のエージェントが一体増加するので、人口分布の統計データとの誤差は必ず改善される。したがって、改善できる誤差に関しては効率よく少なくしていくことができる。ただしこの変更方法は「改悪する状態への変更」ができないので、探索が偏る可能性があり SA の利点が消えることになる。したがって、最適化の中で目的関数値の改善が 5 回滞った時に提案した新たな年齢の変更方法を用いる。また、最適化が進み

探索回数が規定回数に近づくと、人口分布の統計データとの誤差がなくなっていく。誤差がなくなった時は提案した新たな変更方法ではなく通常の人口分布に基づく確率的な変更を行うものとする。

4 実験

本章では、既存の SA による最適化と提案した新たな年齢の変更方法での実験を行い、その結果を比較しその有効性について考察する。実験は次節の統計データとの誤差がより小さい推計データの効率的な発見方法を用いて行う。世帯の数は 500 世帯とした。

4.1 誤差がより小さい推計データ

解の組み合わせが極めて多いことから、推計データは、世帯の初期生成の状態がその後の目的関数値の改善に影響することが考えられる。さらに、探索回数を増加させてもその探索回数に見合うような目的関数値の改善は見られなくなることも起こり得る。したがって、他の研究に应用できるレベルの推計データを得るには工夫が必要となる。まず、世帯の初期生成のシードを調べて、潜在的に目的関数値が小さくなりやすい世帯の組み合わせの推計データを発見する。多くの組み合わせを調べる必要があるので、式 (4) で $N = 10$ とし探索回数を 1.25×10^6 回に設定する。その探索回数の下で世帯の初期生成時のシードを 1~500 の間で順番に変更し、500 試行の最適化を実行する。500 試行の最適化結果を目的関数値を元にソートする。さらに、式 (4) で $N = 100$ とし探索回数を先ほどの 10 倍の 1.25×10^7 回に設定して、ソートされた推計データの目的関数値が上位 30 位までのシードの推計データに対して同様の最適化を行う。これで得られた推計データの上位 5 位までのシード用いて、式 (4) を $N = 500$ にした 6.25×10^7 回の探索回数で最適化を行う。こうして得られた推計データの目的関数値は、最初の $N = 100$ の 500 試行で得られた下位の目的関数値に比べて小さな値となっていると考えられる。

4.2 比較実験

SA の最適化で提案した年齢変更の手法が有用であるか確かめるために既存の SA と比較する。最適化で用いる目的関数は式 (3) とした。まず、 $N = 100$ の探索回数で得られた上位 30 位の目的関数値の平均値と標準偏差を Table 6・7 に示す。式 (3) の値は、実験で使用した目的関数値を表している。式 (1) の値は、目的関数値としては使われていないが得られた推計データを式 (1) で計算した時の値である。式 (3) の値で見

Table 6: 既存の SA ($N = 100$)

	平均値	標準偏差
式 (3) の目的関数値	21.43	4.62
式 (1) の目的関数値	11.45	5.02

Table 7: 提案した SA ($N = 100$)

	平均値	標準偏差
式 (3) の目的関数値	18.87	3.42
式 (1) の目的関数値	11.35	6.11

ると、平均値と標準偏差の両方で提案した新たな年齢変更を追加した手法の方が小さな値となっている。式 (1) の値では、平均値はほぼ等しく標準偏差は既存の SA の方が小さな値となっている。これらの 30 試行の

目的関数値について平均値の検定を行ったところ、式 (3) の値は有意水準 5% で有意な差が見られたが、有意水準 1% では有意な差は見られなかった。式 (1) の値については有意水準 5% で有意な差が見られなかった。したがって、SA の最適化に提案した新たな年齢変更を追加することで、統計データごとの誤差人数を最小化させるためには有効であるといえる。次に 30 試行の上位 5 位までを $N = 500$ の探索回数で最適化した結果を Table 8・9 に示す。

Table 8: 既存の SA ($N = 500$)

	目的関数値	
	式 (3) の値	式 (1) の値
シード 370	12	5.85
シード 210	14	6.35
シード 401	14	6.52
シード 489	16	5.69
シード 470	17	10.84

Table 9: 提案した SA ($N = 500$)

	目的関数値	
	式 (3) の値	式 (1) の値
シード 370	7	2.97
シード 457	12	5.97
シード 362	13	5.14
シード 119	13	5.24
シード 470	15	9.83

最適化には式 (3) の目的関数値を使用しているため、その値でソートしている。Table 8 から、既存の SA による最適化ではシード 370 で世帯の初期生成が行われた推計データが最小の目的関数値となっている。式 (1) の値も指標となる 9.0 ($1.0 \times$ 目的関数の数) を下回っているため統計データとの適合度合いは十分であると考えられる。Table 9 からは、式 (3) の値が 7 で式 (1) の値が 2.97 という推計データが得られたことがわかる。これは既存の SA で最良の結果であった推計データよりも、二つの目的関数値が小さな値となっていることから統計データとの誤差がより小さな推計データであるといえる。二つの結果で目的関数値が最も小さな値になっているのは、それぞれで同じシード 370 であった。 $N = 100$ の探索回数行った 30 試行の結果からは平均値に統計的に有意な差は見られなかったが、探索回数を増加させることで既存の SA に比べて目的関数値がより小さい推計データが得られていることが確認できる。

4.3 各目的関数値

今回行った実験で得られた推計データの中で目的関数値が最小になっている結果について、9 種類の統計データに対応するそれぞれの目的関数値に着目する。比較している二つの SA による最適化で目的関数値が最小となったのは世帯の初期生成時のシードがそれぞれで同じシード 370 となり、Table 8・9 の一番上の試行である。その推計データを各目的関数値ごとに示したのが Table 10 である。提案した年齢変更による SA を用いた最適化で目的関数値が最小となったのは、世帯の初期生成時のシードが 370 で Table 9 の一番上の試行である。同様に目的関数値ごとに示したのが Table 11 である。

Table 10: 既存の SA ($N = 500$ ・シード 370)

	式 (3) の値	式 (1) の値
f_1	4	3.03
f_2	0	0.30
f_3	1	0.48
f_4	4	0.52
f_5	2	0.36
f_6	1	0.44
f_7	0	0.32
f_8	0	0.27
f_9	0	0.15
合計	12	5.85

Table 11: 提案した SA ($N = 500$ ・シード 370)

	式 (3) の値	式 (1) の値
f_1	0	0.36
f_2	0	0.30
f_3	1	0.48
f_4	4	0.52
f_5	2	0.35
f_6	0	0.19
f_7	0	0.34
f_8	0	0.29
f_9	0	0.14
合計	7	2.97

まず、Table 10 の f_1 の目的関数値を見ると式 (1) の値が 3.03 となっているので指標となる 1.0 を上回っている。式 (3) の値も 4 となっているので f_1 は改善の余地があることがわかる。他の $f_2 \sim f_9$ は式 (1) の目的関数値がそれぞれ 1.0 を下回っているが、式 (3) の値では $f_3 \sim f_6$ が 0 より大きい値になっている。次に Table 11 では、式 (1) の値で見ると全ての目的関数値で 1.0 を下回っている。式 (3) の値で見ると $f_3 \sim f_5$ が 0 より大きい値となっているので改善の余地があることがわかる。二つの結果を比較すると、各目的関数値の合計が Table 10 の式 (1) の値が 5.85、式 (3) の値が 12 となっており、Table 11 の式 (1) の値が 2.97、式 (3) の値が 7 となっている。したがって、提案した新たな年齢変更を追加した SA の方が統計データとの誤差が少ない推計データであることがわかる。

それぞれの目的関数値に着目すると、Table 10 では f_1 と f_4 の式 (3) の値が同じ 4 であるのに、式 (1) の値が 3.03 と 0.52 で目的関数値に大きな差がある。これは式 (1) の構造が統計データの項目数 G_s で割って正規化していることが要因であると考えられる。各目的関数に対応する統計データが項目数が異なっており、 f_1 の統計データの項目数は 9、 f_4 の統計データの項目数は 101 となっている。つまり、統計データの項目数が少ないと少しのズレから大きな誤差の値が計算されるということである。逆に統計データの項目数が多いと多少のズレが生じていても計算される誤差の値は小さくなる。また、式 (1) と式 (3) の値が等しくても同じ項目で誤差が生じているとは限らない。式 (1) の値はその値同士の比較がしやすいように実数を小数第二位まで丸めて表示している。

最後に、各統計データとの誤差がいずれの項目で生じているかを確認するために、Table 11 の f_3 の目的関

数を計算する時に用いている各変数を Table 12 に示す。Table 12 の m_{sj} は「夫婦関係であるか」という条件に当てはまるエージェントの組の数であるので推計データには 301 組の夫婦が存在していることを表す。 c_{sj} の値は 301 組の夫婦の中で特定の年齢差に当てはまる数を表しているのので、例えば、一番上の項目は 301 組の夫婦の中で 18 組の夫婦が年齢差-4 以下であることを意味する。目標の値は $\text{Round}(m_{sj} \cdot r_{sj})$ で計算した c_{sj} に対する目標の値である。誤差は c_{sj} と目標の値の差を表しており、式 (3) で絶対値を計算する前の値なので、正の数であれば過剰であることを表し負の数であれば不足していることを表す。年齢差が-3 の項目の誤差が 1 なので、夫が妻より 3 歳年下の夫婦が 1 組過剰であることを示している。この誤差の値を 0 にするには、年齢差が-3 の 10 組の夫婦の中で 1 組の夫婦の夫か妻のどちらかの年齢を変更することで誤差の値は 0 になる。しかし、年齢を変更した夫婦はいずれかの項目に属することになるので別の項目の誤差が +1 されることになる。Table 12 の状態では年齢差が-3 の項目以外が全て誤差 0 になっているので、年齢差-3 から 1 組の夫婦を移動させても必ず誤差 1 が生じることになる。したがって、エージェントの年齢を変更しても f_3 の目的関数値をこれ以上小さくすることはできない。Table 11 で他の f_4, f_5 についても同様の理由で誤差が生じていた。ただし、 $f_6 \sim f_9$ は目的関数値の計算に必要な m_{sj} と c_{sj} を求める時の条件 X と条件 Y の構造が $f_1 \sim f_5$ と異なっている。よって、 $f_1 \sim f_5$ は年齢の変更では改善できない誤差が生じる可能性がある。これらのことから、Table 11 の結果は、年齢の変更のみを行う SA の最適化で発見することができる、最適解の中の一つであるといえる。

年齢の変更で改善できない誤差が生じるのは、推計データの規模に合わせるために統計データの割合を用いて目標の値を設定していることが原因として考えられる。目標の値の和は m_{sj} と等しくなるべきであるが統計データの割合を用いて計算しているのので等しくならない可能性がある。Table 12 の例では、目標の値の和が 300 であるのに対して推計データには 301 組の夫婦が存在している。したがって、1 組の夫婦はどの年齢差の項目に当てはめても誤差の値 1 が必ず生じることになる。

Table 12: f_3 の目的関数値の詳細

j	夫婦の年齢差	m_{sj}	c_{sj}	目標の値	誤差
0	年齢差 ~ -4	301	18	18	0
1	年齢差 -3	301	10	9	1
2	年齢差 -2	301	15	15	0
3	年齢差 -1	301	29	29	0
4	年齢差 0	301	61	61	0
5	年齢差 1	301	41	41	0
6	年齢差 2	301	30	30	0
7	年齢差 3	301	23	23	0
8	年齢差 4	301	18	18	0
9	年齢差 5	301	14	14	0
10	年齢差 6	301	11	11	0
11	年齢差 7~	301	31	31	0
合計	—	—	301	300	—

5 考察

本研究では、新たな年齢の変更方法を SA の最適化に追加した手法を提案し既存の SA との比較を行った。様々な組み合わせの世帯を持つ推計データに対する試行実験では有意差 5% で平均値に有意な差がみられたが、有意水準 1% ではみられなかった。また、統計データとの誤差が少ない推計データを効率よく発見する方法を用いた実験では、目的関数値が比較的小さな推計データを得ることができた。その推計データは式 (3) の目的関数値が 7 なので、いずれかの統計データで誤差が生じていることになる。しかし、この目的関数値は年齢の変更では小さくすることのできない誤差であった。よって、年齢の変更のみを行う SA による最適化で得ることのできる最適解の中の一つを発見できたといえる。SA を用いた最適化で年齢の変更以外の状態変更を処理に追加することでより解空間を広範囲にわたって探索を行うことができる。例えば、性別の変更は対象となるエージェントを制限すれば容易に実装が可能である。しかし、世帯を減らしたり世帯の種類が変わるような状態変更では一回の探索に処理時間の負担がかかることになる。SA を用いた最適化では、一回の探索に要する処理時間を減らさなければ現実的な時間で最適化を終えることが困難になる可能性がある。よって、年齢と性別以外で状態変更を追加する場合は世帯の管理に影響が出ない方法を検討していく。今回の実験では世帯の数を 500 に設定していたので、日本全体を対象とした統計データを用いていることを考えるとかなり規模が小さい仮想的な人口社会の推計であった。今後、大規模な世帯数を設定した時に提案した手法が適用できるかについては、実験を行い検証していくことが必要である。

参考文献

- 1) 三上達也：マルチエージェントモデルによる社会シミュレーション，政策科学，14-3，121/134 (2007)
- 2) 独立行政法人 統計センター <http://www.nstac.go.jp/>
- 3) 池田心，喜多一，薄田昌広：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，計測自動車制御学会第 43 回システム工学部会研究会，11/14 (2010)
- 4) 柘井大貴，村田忠彦：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，第五回社会システム部会研究会，121/126 (2014)
- 5) 福田純也，喜多一：シミュレーテッドアニーリングによるエージェント属性決定手法を用いた人口推計モデルの評価，第四回社会システム部会研究会，35/40 (2013)
- 6) 国立社会保障・人口問題研究所：人口統計資料集 (2012, 2013) <http://www.ipss.go.jp/syoushika/tohkei/Popular/Popular2013.asp?chap=0>
- 7) 厚生労働省：e-Stat 人口動態調査 9-14 (2010) <http://www.e-stat.go.jp/SG1/estat/List.do?lid=000001101829>
- 8) 国立社会保障・人口問題研究所：第 14 回出生動向基本調査 2-1 (2010) <http://www.ipss.go.jp/ps-doukou/j/doukou14/doukou14.asp>
- 9) 厚生労働省：平成 23 年度全国母子世帯等調査結果報告 19 (2011) http://www.mhlw.go.jp/seisakunitsuite/bunya/kodomo/kodomo_kosodate/boshi-katei/boshi-setai_h23/