並列計算を用いたSA法による大規模世帯の復元

原田拓弥　○村田忠彦　枡井大貴（関西大学大学院総合情報学研究科）

Reconstructing Large-Scale Household Composition Using Parallel Computing

T. Harada, * T. Murata and D. Masui (Department of Informatics, Kansai University Graduate School)

概要－　本論文では，大都市や国，複数の国家のマイクロシミュレーションやエージェントシミュレーションを行うため，Simulated Annealing（SA）法による大規模な数の世帯構成の復元を，並列計算を用いて行う．社会シミュレーションをより具体的なものにするためには，現実のデータを用いたシミュレーションを行うことが望まれる．本研究では，統計データを用いた大規模な数の世帯構成の復元を行うため，並列計算によるアプローチを実装する．並列計算により，大規模世帯の復元を行うためには，分割して復元を行う個々のスレッドに統計データを分割することが必要になる．単純な分割を行うと，総人口にずれが発生することに着目し，そのずれを小さくするための調整手法を考案する．数値実験により，調整手法により，復元データの統計データに対する誤差が小さくできることを示す．

キーワード:　統計データ, 大規模世帯構成復元, 並列計算, シミュレーティッドアニーリング法.

1　はじめに

計算科学は，理論科学，実験科学と並ぶ第３の科学として認められるようになっている[1]．物理学や化学，天文学などの分野で計算科学は利用されているが，近年では，経済学や政治学，社会学の分野においても，社会シミュレーションという形で活用されるようになっている．人間をモデルに内包する社会シミュレーションは，物理学や化学などと比較して，不確定要素が多く，モデルの妥当性を確認する上で課題が残る場合が少なくないが，多様なシミュレーションを行うことにより，極端な事例（シナリオ）を発見し，それらの極端な事例から，避けなければならないシナリオや，望ましいシナリオを抽出し，なぜそれらのシナリオが発生したのかのメカニズムをモデルの範囲内で調べることにより，実世界で起こりうる現象の説明ができることが期待される．

社会シミュレーションの技法の一つにマイクロシミュレーション[2]やエージェントシミュレーション[3]があり，それらのシミュレーションモデルでは，モデル化する社会における市民を個別に生成することが必要になる．単純化した人工社会における市民の生成の際には，市民の生成においても，単純化された方法を用いて生成されることが多いが，モデルの対象となる社会をより現実的なものにしようとする場合，その人工社会における市民の属性も，実際の社会における市民の属性と同様のものになることが期待される．

市民の属性の再現を行うにあたって，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，それらの情報源に格納されている情報については，正確に再現することができるが，市民のプライバシーへの配慮から，それらの情報源が個票レベルで公開されることはない．また，仮にそれらの情報源が試用可能な場合でも，そのデータを直接用いることにより，ある個人のシミュレーション結果が特定されることが倫理的に考えてふさわしいかどうか検討を進めなければならない．このような状況から，公開されている統計データから，仮想的な市民をもつ人工社会を生成し，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．

統計データに基づく，個々の市民の個票データの復元に関する研究の歴史は古く，Synthetic reconstruction method（SR 法）[4]として知られている．SR 法は，個票データのサンプルをもとに，Iterative Proportional Fitting Procedure（IPFP）[5]を用いて個票データを復元している．その後，数多くの個票データ復元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[6]は，IPFP の弱点として，個人の統計と世帯の統計のどちらかに適合する復元ができたとしても，両方に適合する復元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[7]や Barthelemy ら[6]は，サンプルを用いない復元手法を提案している．Lenormand と Deffuant[8]は，サンプルを用いない復元手法と SR 法と比較し，前者が個人と世帯をよりよく復元できていることを示している．

本研究で用いる復元手法もサンプルを用いない手法である．本研究では，池田ら[9]が提案した SA 法（Simulated Annealing 法）を用いたデータ復元手法をもとに，目的関数を変更した SA 法[10]，特定の統計データの誤差を小さくするヒューリスティックを導入した改良型 SA 法[11]を用いて世帯の復元を行う．

本研究では，統計データを用いた世帯構成復元手法を，大規模な世帯数に対して適用する際に発生する問題点を解決するための手法を提案する．具体的には，都道府県レベルの世帯構成復元を行い，大規模世帯の復元における問題点を明らかにする．先行研究[9,10,11]では，500世帯や1000世帯などの規模の復元が行われていたが，本研究では，約39万世帯をもつ山形県を対象に世帯の復元を試み，統計データとの誤差を最小化するための方法を提案する．また，大規模世帯数を復元するため，本研究では，並列計算を用いた復元手法を提案する．

SA法を用いた世帯構成復元手法[9,10,11]では，日本人口の約95%をカバーする9種類の世帯種類の割合を考慮して生成した世帯の構成員の年齢を，実際の統計データとの誤差を最小化するようにSA法を用いて，世帯構成員の年齢を変更する事で最適化している．これま

での研究では，500世帯や1000世帯など，実際の統計の世帯数よりも少ない世帯数を用いて，最適化手法の開発が行われてきた．その際，小規模化した世帯数に対する目標の統計値を得るため，実際の統計を調整する処理がおこなわれてきた．調整後の人数が小数点以下の項目については，四捨五入で整数化しているため，人口が異なる可能性があった．

世帯数が大規模な場合の世帯構成の復元では，世帯構成の復元に時間がかかるため，より高速な復元手法の開発が望まれる．そこで本研究では，並列計算を用いて，高速化をはかるとともに，復元された世帯構成の誤差が少なくなる調整手法を提案する．

2　SA 法を用いた世帯構成復元手法

池田ら[9]が提案した復元手法は，公開されている複数の統計データに適合するような人口データを復元する手法である．市民の年齢や親子の年齢差についての統計データに対して，コンピュータ上で再現したデータ集合（復元データ）がどの程度ずれているかを計算する目的関数を設計し，SA 法 を用いて最適化している．

復元データは複数の世帯とその構成員である市民によって構成されており，世帯数Hを設定することで復元データの規模が決定する．復元データのモデルをFig. 1に示す．池田らの手法は，全国の統計データに基づく世帯構成の復元を試みられており，世帯の種類は実統計[12]に掲載されているFig. 2の９種類の世帯を用いている．統計データには他の種類の世帯も存在しているが，その割合が少ないので存在しないものとする．なお，これらの９種類の世帯数で日本全体の全世帯数の95%を占めていることが知られている．

世帯の中には構成員として市民が存在している．それぞれの市民は年齢，性別，所属する世帯の種類，世帯の役割，親族関係の５つの属性を持っている．世帯の役割は，所属する世帯の中での役割を表している．例えば，「夫婦と子供」の世帯の場合，夫，妻，子供の３つの役割がある．親族関係は親－子と夫－妻の相互関係を表すための属性であり，一般的な家系図で記述される人と人を繋ぐ線を表現している．例えば，自分の妻の属性を確認する時は，この親族関係を用いて妻の情報にアクセスする．

復元データの初期生成は，規定の世帯数Hだけ世帯をつくることによって行う．世帯をつくる時，統計データ[12]の割合に従って世帯の種類を決定する．Fig. 2の９種類の世帯以外の５%の世帯を除いた後の９ 種類の世帯割合をTable 1に示す．池田らはこれらの割合に従って世帯を生成している．しかし，その具体的な方法が明確ではないため，本研究では確率的に世帯の種類を決定してその割合に従う世帯を生成する．

子供がいる世帯での子供の数は統計データ[13,14]の割合に従って確率的に決定する．Table 2は，３つの世帯で子供の数が１～４人の時のそれぞれの世帯数の割合を表している．９種類の世帯のうち，「夫婦と子供」，「夫婦と子供と両親」，「夫婦と子供と片親」には「夫婦世帯」の割合を用いて子供の数を決定する．「父親と子供」世帯には「父子世帯」の割合を，「母親と子供」世帯には「母子世帯」の割合を用いる．

Table 2の「夫婦世帯」について，実際の統計データでは子供の人数が０，１，２，３人と４人以上の５つ

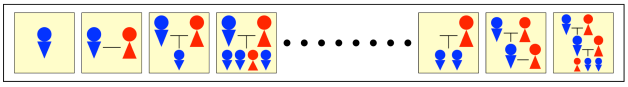

Fig. 1：復元データのモデル

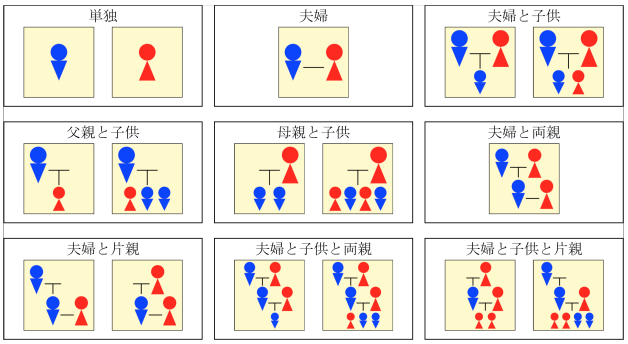

Fig. 2：世帯の種類

Table 1：９種類の世帯の割合

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
合計	99. 99

Table 2：世帯の子供の数の割合

	子供の数			
	1 人	2 人	3 人	4 人以上
夫婦世帯	16. 97	59. 98	20. 70	2. 35
父子世帯	54. 70	36. 00	8. 20	1. 10
母子世帯	54. 70	34. 50	8. 90	1. 90

に分類されている．しかし，子供の数が０人の夫婦世帯は本研究で用いる９種類の世帯の「夫婦のみ」世帯に該当するので，子供の数が０人の値を除いた割合を用いる．また，子供の数が「４人以上」の項目については詳細な人数がわからないので，この項目は「４人」として扱う．したがって，復元データを構成する世帯の子供の数は１～４人となる．

世帯の種類と子供の人数が決定した後，その世帯の構成員である市民の属性を設定する．市民の年齢は人口ピラミッドの統計データ[12]の割合に従って確率的に設定する．これは男女の人数の合計で１歳区切りとなっている．性別については，単独世帯に所属する市民と，子供か片親の市民のみをランダムに設定する．それ以外の市民は，世帯の役割の属性に応じた性別を適切に設定する．具体的には，父親か夫の役割を持つ市民の性別を男性に，母親か妻の役割を持つ市民を女性に設定する．

以上の過程で生成された各世帯に属する市民の年齢を複数の統計データに適合するように最適化することが，池田らの提案した復元手法の目的である．次に，

適合させる統計データについて説明する．

池田らは次のような全国の統計データを用いて市民の属性の復元を行っている．対象の統計データを以下に示す．

(I) 父子の年齢差（表4-13，2013 年[12]）
(II) 母子の年齢差（表4-8，2012 年[12]）
(III) 夫婦の年齢差（表9-14，2011 年[15]）
(IV) 男性の人口分布（表2-3，2012 年[12]）
(V) 女性の人口分布（表2-3，2012 年[12]）
(VI) ある年齢の男性が単独世帯にいる割合（表7-28，2013 年[12]）
(VII) ある年齢の女性が単独世帯にいる割合（表7-28，2013 年[12]）
(VIII) ある年齢の男性が夫婦のみ世帯にいる割合（表7-28，2013 年[12]]）
(IX) ある年齢の女性が夫婦のみ世帯にいる割合（表7-28，2013 年[12]）

統計データ(I),(II) は5 歳区切りの年齢差で,(III) は１歳区切りになっている．統計データ(I)，(II) のデータ形式をTable 3に，(III) のデータ形式をTable 4に示す．統計データ(IV)，(V) は１歳区切りの人口ピラミッドになっている．データ形式をTable 5に示す．統計データ(VI) － (IX) は５歳区切りで世帯の種類に関係している．データ形式をTable 6に示す．Table 3～Table 6の割合の値が実統計の値で，全人口の中で条件Xを満たす数の内，条件Y を満たす割合である．本研究では，基本的に2010 年の状態を表す統計データを用いるが，同じ年のデータがない一部の統計データについては，近い年のものを用いる．

本研究では，SA法を用いて，上記の統計データに適合する世帯構成の復元を行う．SA法では，以下の目的関数[10] を用いて，最適化を行う．

$$f_s(A)=\sum_{j=1}^{G_s}|c_{sj}(A)-Round(r_{sj}\cdot m_{sj}(A))|, \qquad (1)$$

ここで，A は復元データ，s は統計データの種類，G_s は統計データ s の項目数，c_{sj} は 統計データ s の条件 X_{sj} と条件 Y_{sj} を満たす復元データの市民の値，r_{sj} は統計データ s の項目j の割合，m_{sj} は統計データ s の条件X_{sj} を満たす復元データの市民の値を表している．

上記の目的関数を統計データ$s=1,2,\ldots,S$ に対して行い，その総和の最小化をSA法を用いて行う．本研究で用いるSA法は以下の手続きにより最適化を行う．

Step 1．復元データを初期生成
Step 2．探索回数が規定数に達すれば探索を終了
Step 3．復元データ内の一人の市民の年齢を変更
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

なお,改良型SA法[11]では,Step 3を次のように変更し，統計データ(IV) か(V)について，最小化を行う．

Table 3 (I) 父子の年齢差

条件X	条件Y	割合(%)
父子関係	年齢差～14	0.00
父子関係	年齢差15～19	0.55
父子関係	年齢差20～24	8.49
...	...	...
父子関係	年齢差40～44	7.91
父子関係	年齢差45～49	2.10
父子関係	年齢差50～	0.11

Table 4 (III) 夫婦の年齢差

条件X	条件Y	割合(%)
夫婦関係	年齢差～-4	6.13
夫婦関係	年齢差-3	3.12
夫婦関係	年齢差-2	4.82
...	...	...
夫婦関係	年齢差5	4.72
夫婦関係	年齢差6	3.52
夫婦関係	年齢差7～	10.44

Table 5 (IV) 男性の人口分布

条件X	条件Y	割合(%)
男性	年齢0	0.87
男性	年齢1	0.87
男性	年齢2	0.89
...	...	...
男性	年齢98	0.01
男性	年齢99	0.01
男性	年齢100	0.01

Table 6 (VI) ある年齢の男性が単独世帯にいる割合

条件X	条件Y	割合(%)
男性・年齢～14	単独世帯	0.01
男性・年齢15～19	単独世帯	7.01
男性・年齢20～24	単独世帯	27.99
...	...	...
男性・年齢75～79	単独世帯	10.30
男性・年齢80～84	単独世帯	10.93
男性・年齢85～	単独世帯	11.74

Step 3．解の遷移が5 回連続で行われていない時に，統計データ(IV)か(V)の人口分布との誤差が小さくなるような候補解が存在しているならば，その誤差が小さくなるように市民の年齢を変更，それ以外の時は年齢別の人口分布の割合に応じて，確率的に年齢を変更

本研究では，上記のSA法，改良型SA法を用いることにより,統計データに基づく世帯構成の復元を行う．

3 大規模世帯の復元のための調整手法

本研究では，前節の SA 法[10] と改良型 SA 法[11] を用いて，大規模世帯の復元を行う．上記の SA 法では，実際の統計データと比較して，少数の世帯数（$H=500$ 世帯）を対象として，統計データに適合する世帯の復元を行っていた．世帯数を大規模化するにあたって，エージェント個々の属性数が増えてくると，１台の計算機で取り扱えなくなる状況が考えられる．そこで，複数の計算機に分割して，大規模世帯の復元を行う状況を考える．

複数の計算機で世帯の復元を行う際に，各計算機が担当する世帯数は少なくなるため，前節で示した小規

模世帯向けの SA 法を使用することができる．ただし，小規模世帯向けの SA 法の目的関数では，式(1)に示したように round 関数による四捨五入を行っているため，四捨五入分の誤差が複数の計算機にわたって累積することが考えられる．

例えば，Fig.3 に，ある統計データを合計 30 人の人数に変換したときのグラフを示す．このとき，項目１，３，５，６，８については，従来の目的関数[10,11]では，式(1)のように目標人数を四捨五入して求めるため，それぞれ，１，５，８，５，０になり，総人口が合計 31 人となる．例えば，従来の目的関数を用いて，300 人のコミュニティを復元するために， 30 人の復元を 10 回行うとすれば，全体の人口が 310 人となる問題が生じる．また，項目８については，目標人数が０人になってしまうため，その項目が復元されなくなることになる．そこで，本研究では， 10 分割で各 30 人のデータを復元する場合，例えば，項目１を 10 個中２つの分割で各２人，その他８つの分割では各１人で扱うことにより，全体の統計に整合する構成員数を復元できるように調整する．

4　実験結果

本研究では，山形県の統計データを用いて，世帯の復元を行った．山形県を選んだ理由は，平均世帯人員が3.01(平成22年国勢調査)と47都道府県で最も高く，世帯人員が多いため，世帯の復元が難しいと考えられたからである．また，全国的にはFig.2の９種類の世帯で95%の世帯をカバーできるものの，山形県では９種類の世帯で90%程度であり，その違いにより，どれだけ最適化における歪みが出てくるかを確認できるからである．

本研究では，512分割，64分割，8分割，１分割（＝分割なし）の４種類の分割方式で行った．今回，最適化の対象とした統計データに対して，個々の市民がもつデータ量は最大でも約300バイトであり，約100万人の人口をもつ山形県の最適化を行うには,264メガバイト程度のメモリがあればよいため，分割なしでの復元を行うことが可能である．なお，世帯復元に使用した計算機のCPUは，Intel Core i7-3930Kで，メモリは64ギガバイトである.また,山形県の統計データとして，Table 7～Table 13の形式のデータを使用した．

９種類の世帯に対応する350,662世帯（山形県統計2010年度．年齢不詳の単独世帯を除く）を512分割した場合，各分割で684世帯または685世帯の復元を行うことになる．一方，64分割の場合は，5,479世帯または5,480世帯，8分割の場合は，43,832世帯または43,833世帯の復元を行うことになる．512分割，64分割，8分割，分割なしにより，約35万世帯の構成の復元を行った実験結果をTable 14に示す．これらの結果はいずれも10回平均の結果である．

Table 14で，単純分割は単純に統計情報を所定の分割数で分割して最適化を行った場合(式(1)の四捨五入を行う場合），調整分割は提案手法により統計的に人口を整合するように調整して最適化を行った場合の，山形県の９つの実統計に対する絶対値誤差の総和を示している．平均誤差は，誤差の10回平均を示し，標準偏差は10回の試行における偏差を表している．この結果からいずれの最適化手法を用いた場合でも，提案した統計データの調整方法を用いることにより，実統計に対する誤差の少ない復元ができていることがわかる．

Table 14の512分割で改良型SA法の結果がSA法と比較して悪くなっている原因は，改良型SA法による改善が統計データ(IV)か(V)を対象としたものであるため，これらの統計データに対する誤差は小さくなったとしても，それ以外の統計データの誤差が悪化し，誤差の総和としては，悪化するためである．本来，他の統計データとも整合的に統計データ(IV)か(V)を最適化できるはずであるが，統計データを分割したために，統計データ(IV)か(V)の最適化を,他の統計データと整合的に行えなくなると考えられる．

Table 14の64分割と8分割では， 512分割の場合と同様に，提案手法を用いて調整を行うことにより，実統計に対する誤差が小さくなることがわかる．

Table 15は，分割なしの結果を示しており，統計データの分割を行う必要がないため，提案手法による調整は行っていない．Table 15においては，改良型SA法を用いることで誤差を小さくできていることがわかる．Table 14とTable 15を比較することにより，分割数が小さい方が，誤差が小さくなることがわかる．すなわち，分割なしの結果がもっとも誤差が小さく，分割数が多くなるほど誤差が大きくなっている．これは，分割数が多くなるほど，複数の統計データを分割した際の,各分割で復元に用いる統計データの分割において，統計データの整合性がとれなくなるからであると考えられる．各分割で復元される世帯数は上記の通り，固定値が与えられるが，世帯タイプごとの世帯数は確率的に与えられるため，世帯タイプの違いにより，子供の数や年齢別人口分布との不整合が出てくることになる．

また，Table 16とTable 17に，それぞれの分割の際の計算時間を示す．今回，512分割，64分割，8分割，分割なしのそれぞれにおいて，エージェント個々の年齢変更を行う探索回数の総数を51億2千万回に設定している．これは，各市民の年齢が平均して約5,120回程度変更される探索回数となっている.分割数が512の場合，分割なしと比較して，計算時間が約７倍強早くなっていることがわかる．512分割することにより，理想的には512倍速くなることが期待されるが,本研究では，８並列で実験を行ったため，最大８倍の速度向上であり，妥当な高速化が実現できているといえる．このことから計算機の並列数を多くすればするほど，高速化が期待できることがわかる．

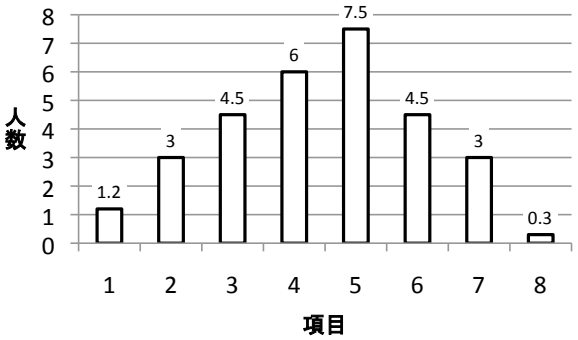

Fig.3：対象人口を小規模化した統計データ

Table 7：９種類の世帯の割合

世帯の種類	割合(%)
単独	24.71
夫婦のみ	18.87
夫婦と子供	24.97
父親と子供	1.37
母親と子供	8.18
夫婦と両親	1.84
夫婦と片親	3.81
夫婦と子供と両親	7.99
夫婦と子供と片親	8.27
合計	100.01

Table 8：世帯の子供の数の割合

	子供の数			
	1 人	2 人	3 人	4 人以上
夫婦世帯	50.79	38.45	10.03	0.72
父子世帯	81.35	16.09	2.31	0.25
母子世帯	75.31	20.80	3.43	0.46

Table 9：(I)父子の年齢差

条件X	条件Y	割合(%)
父子関係	年齢差～14	0.00
父子関係	年齢差15～19	0.31
父子関係	年齢差20～24	9.90
...	...	...
父子関係	年齢差65～69	0.00
父子関係	年齢差70～74	0.02
父子関係	年齢差75～	0.00

Table 10：(II)母子の年齢差

条件X	条件Y	割合(%)
母子関係	年齢差15～19	1.08
母子関係	年齢差20～24	15.03
母子関係	年齢差24～29	35.11
...	...	...
母子関係	年齢差40～44	2.43
母子関係	年齢差45～49	0.08
母子関係	年齢差50～	0.00

Table 11：(III)夫婦の年齢差（夫年齢 − 妻年齢）

条件X	条件Y	割合(%)
夫婦関係	年齢差-42	0.00033
夫婦関係	年齢差-41	0.00
夫婦関係	年齢差-40	0.00
...	...	...
夫婦関係	年齢差45	0.00
夫婦関係	年齢差46	0.00
夫婦関係	年齢差47	0.00033

Table 12：(IV)男性の人口分布

条件X	条件Y	割合(%)
男性	年齢0	0.69
男性	年齢1	0.71
男性	年齢2	0.75
...	...	...
男性	年齢98	0.0097
男性	年齢99	0.0040
男性	年齢100～	0.0050

Table 13：(VI)ある年齢の男性が単独世帯にいる割合

条件X	条件Y	割合(%)
男性・年齢0	単独世帯	0.00
男性・年齢15	単独世帯	0.61
男性・年齢16	単独世帯	0.92
...	...	...
男性・年齢98	単独世帯	10.87
男性・年齢99	単独世帯	15.79
男性・年齢100～	単独世帯	4.17

Table 14: 512分割，64分割，8分割の世帯の復元結果

		単純分割		調整分割	
		SA法	改良型	SA法	改良型
512分割	平均誤差	76,879.1	78,144.3	47,486.2	49,254.6
	標準偏差	420.7	352.97	476.4	591.0
64分割	平均誤差	19,780.6	19,812.0	16,953.5	16,876.5
	標準偏差	512.0	491.6	564.3	594.3
8分割	平均誤差	9,272.4	8,908.2	8,844.1	8,472.7
	標準偏差	616.12	567.43	543.7	494.9

Table 15: 分割なしの世帯の復元結果

	SA法	改良型
平均誤差	5,949.7	5,463.9
標準偏差	830.5	845.8

Table 16: 512分割，64分割，8分割の計算時間（秒）

		単純分割		調整分割	
		SA法	改良型	SA法	改良型
512分割	平均時間	1,208.1	1,206.7	1,209.8	1,201.7
	標準偏差	15.3	30.4	18.6	22.5
64分割	平均時間	1,307.3	1,307.4	1,302.5	1,274.8
	標準偏差	17.7	52.8	42.3	9.6
8分割	平均時間	1,412.3	1,455.3	1,395.1	1,412.9
	標準偏差	18.9	53.4	19.6	25.5

Table 17: 分割なしの計算時間（秒）

	SA法	改良型
平均時間	8,809.4	9315.0
標準偏差	105.6	106.9

Table 18: 調整手法の有効性
(512分割，(V)女性人口分布)

年齢	人口	分割なし	単純分割	調整分割
96	264	264	398	156
97	184	182	0	103
98	122	122	2	67
99	75	75	0	35
100	103	103	1	47

高速化と誤差の関係について，512分割の場合，分割なしと比較して，12倍程度悪くなっていることがわかる．また，8分割と分割なしを比較すると，計算時間は６倍強速くなっているのに対して，誤差は２倍弱程度の悪化にとどまっている．誤差が大きくなることと速度向上にトレードオフがあるものの，誤差を大きくせ

ずに，速度向上をはかることのできる分割数があることがわかる．今回の８並列の実験環境では，８分割が最も高速化と誤差のバランスのよい分割数となる．

最後に，提案した調整手法の有効性を示すために，統計データ(V)の女性の年齢別人口分布における復元例をTable 18に示す．512分割をする場合，単純分割では，表中の人口が0.5未満になる項目については，四捨五入されて目標値が０人になってしまうため，人口がほとんど復元されていないことがわかる．一方，調整分割では，まだ誤差は残っているものの，ある程度復元できていることがわかる．

5 おわりに

本研究では，大規模な数の世帯構成の復元を行うにあたり，課題となる点を明らかにした．分散コンピューティングを使って，世帯構成を復元する場合には，各計算機に分散する統計データを整合的に分割する必要があるが，本研究では，まず人口に誤差がおこらないような分割を行う調整手法の影響を調べた．数値実験の結果，提案した調整法を用いることにより，誤差を小さくできることが確認できたが，分割数が大きくなるほど，分割された統計データの整合性が問題となり，誤差が大きくなることがわかった．また，筆者らが既に提案していた改良型手法が，分割数が多い場合に，一部の統計データに特化した最適化を行うために，不整合を起こしているデータの誤差を拡大してしまうことがわかった．

本論文で示した実験結果から，できるだけ分割数を少なくして，世帯構成の復元を行うことが望ましいことがわかる．今回は，山形県の復元を行ったが，現在，用いているエージェントの属性データ（300バイト程度）であれば，東京都の約1200万人，600万世帯の復元においても約3.4ギガバイト程度のメモリで済むため，今回，用いた64ギガバイトのメモリ環境の計算機であれば，分割なしの手法を適用することが可能である．したがって，日本国内であれば，都道府県単位の統計データがそろっていれば，個々の都道府県の復元を統合することで，一国の世帯構成の復元が可能であることがわかる．

一方，64ギガバイトで，東京都1200万人を表現した場合の個々の市民のもちうる属性数としては，一つの属性に32ビット（約43億通り）を割り当てると，約1500属性をもつことができる．どんなシミュレーションを行うかによって，モデルの中で個人が持つべき属性数は変化するが，将来的にこれらの属性数で足りなくなった場合，どのように統計データを分割して，分散コンピューティングにより，整合的に個人の属性を復元するかが課題になる．

統計データを整合的に分割できるようになれば，分散コンピューティングの技術を使って，高速に世帯の復元を行うことが可能となる．山形県100万人の復元に分割なしで約２時間半の時間がかかっており，東京都1200万人の場合，その12倍の約30時間がかかると推定される．緊急のシミュレーションが求められている場合，この30時間のロスは致命的になりかねないため，高速に世帯構成を復元する方法の開発が必要となる．富士通はスーパーコンピュータ「京」を用いた津波の予測を最短10分で行うモデルを開発している[16]．災害現場の避難シミュレーションでも，発生した災害に対して，その場にいる市民がどのような行動をとるべきか，危険な市民の存在をシミュレーションによりいち早く知ることができれば，災害救助における効果性を一段と向上させることが可能となる．このような社会を実現するため，引き続き，整合的な統計データの分割手法に関する研究が求められている．

謝辞

本研究の一部は，JSPS科研費26380277の助成を受けたものです．

参考文献

1) 小柳義夫：計算科学とシミュレーション，http://olab.is.s.u-tokyo.ac.jp/~oyanagi/reports/cs-and-sim.txt（2003年9月執筆）．
2) 矢田晴那：政策分析ツールとしてのマイクロ・シミュレーション，ファイナンス，6月号，35/40（2010）
3) 山影進：社会科学とマルチエージェントシミュレーション ―シミュレータ開発と事例提供の課題―，情報科学，No. 27，1/10（2007）
4) A. G. Wilson and C. E. Pownall: A new representation of the urban system for modeling and for the study of micro-level interdependence, *Area*, Vol. 8, No. 4, pp. 246/254 (1976)
5) W. E. Deming and F. F. Stephan: A least squares adjustment of a sampled frequency table when the expected marginal totals are known, *The Annals of Mathematical Statistics*, Vol. 11, 428/444 (1940)
6) J. Barthelemy, P. L. Toint: Synthetic population generation without a sample, *Transportation Science*, 266/279 (2012)
7) F. Gargiulo, S. Ternes, S. Huet, G. Deffuant: An iterative approach for generating statistically realistic populations, of households, *PLoS One*, Vol. 5, No. 1, e8828. doi:10.1371/journal.pone.0008828 (2010)
8) M. Lenormand, G. Deffuant: Generating a synthetic population of individuals in households: Sample free vs sample-based methods, *Journal of Artificial Societies and Social Simulation*, Vol. 16, No. 4, 9 pages (2013)
9) 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）
10) 枡井大貴，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，計測自動制御学会第５回社会システム部会研究会，121/126（2014）
11) 枡井大貴，村田忠彦：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，計測自動制御学会第７回社会システム部会研究会，47/52（2014）
12) 国立社会保障・人口問題研究所：人口統計資料集，http://www.ipss.go.jp/syoushika/tohkei/Popular/Popular2013.asp?chap=0（2012）
13) 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，http://www.ipss.go.jp/ps-doukou/j/doukou14/doukou14.asp（2010）
14) 厚生労働省：平成23年度全国母子世帯等調査結果報告19，http://www.mhlw.go.jp/seisakunitsuite/bunya/kodomo/kodomo_kosodate/boshi-katei/boshi-setai_h23/(2011)
15) 厚生労働省：e-Stat 人口動態調査 9-14，http://www.e-stat.go.jp/SG1/estat/List.do?lid=000001101829（2010）
16) 富士通：地震発生から最短10分で津波予測！スパコンで世界を津波から救う，Fujitsu Journal，http://journal.jp.fujitsu.com/2015/04/08/01/（2015/4）