

成人口データから生成する統計表との差（以下、統計表との誤差）により合成した合成人口データを評価している。これは、統計表の利活用は可能であるものの、実個票データの利活用は困難であるため、統計表との誤差により評価されていた。したがって、現実の集団と合成された合成人口データが同じ統計的特徴をもつ集団と言及することはできる。しかし、現実の集団の個々の世帯とどの程度類似しているか評価できない。

そこで、本論文では、合成された集団と合成対象の集団の類似度を評価する手法を提案する。従来の評価手法と提案手法の差異を Fig. 1 に示す。Fig. 1 (a) に示す従来の評価手法では、実個票データの利活用は困難であるため、実個票データから集計される統計表（以下、実統計表）と、合成人口データから集計される統計表（合成人口データから作成する統計表を合成統計表と呼ぶ）の各統計量の差を用いて、合成人口データを評価していた。Fig. 1 (b) に示す提案手法では、まず、実統計表から仮想都市の人口（仮想都市における個票データ、以下、仮想個票データ）を作成し、仮想個票データを用いて統計表を集計する。この仮想個票データから集計された統計表（以下、仮想統計表）を用いて人口合成することで、従来の評価手法である統計量の差に基づく評価ができるとともに、合成された個票データと仮想個票データを比較することができる。

本論文では、世帯の構成員同士を比較するために、仮想的な個人によって構成される世帯群を仮想都市として構成する。この手法により、実統計表のもととなる入手が困難な実個票データがなくとも合成手法の一致度による比較が可能になる。

本論文の構成はつぎのとおりである。まず、2章で合成手法について説明する。つぎに3章で本論文が提案する評価手法を説明する。4章で実験結果を示しその考察を述べる。5章は本論文のまとめである。

2. 合成手法

池田らが提案した合成手法¹²⁾とその改良手法^{13)~17)}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（合成人口データ）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. 2 に示す9種類の家族類型を対象に合成人口データを合成している。Fig. 2 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。

これらの合成手法が解く現在の人口合成問題は、拘束条件に該当する統計表の数が少なく、複数の最適解が存在する自由度がある中で、特定の解が探索されている。異なる調査に基づく拘束条件が相互に矛盾する場合があるものの、拘束条件の自由度が高いため、すべての拘束条件を充足する解が得られる問題である。また、親子の年齢差については、子の出生年によって、親子の年齢差の分布は異なっているが、本手

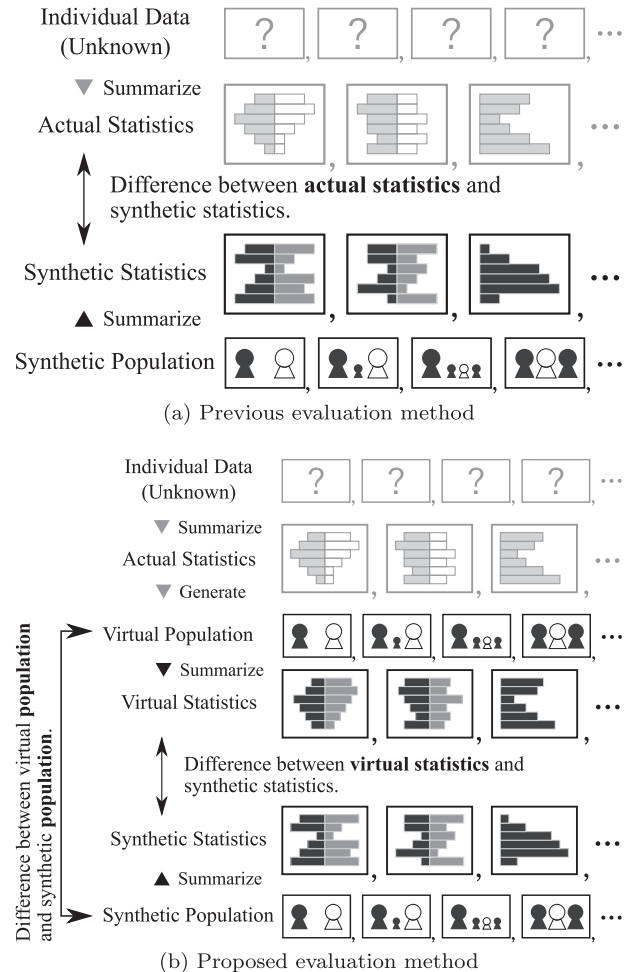


Fig. 1 Difference between previous evaluation method and the proposed evaluation method

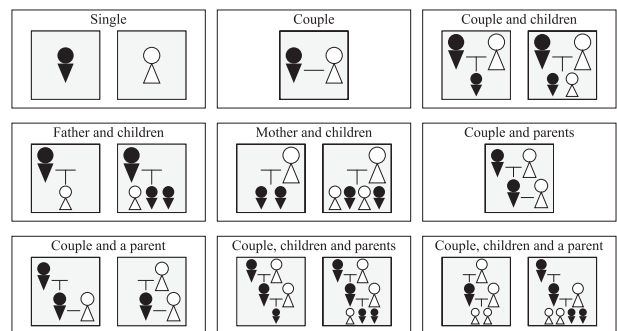


Fig. 2 Nine family types

法では、ある年度に出生した子と親の年齢差の統計表をすべての親子に適用している。したがって、ほかの拘束条件として用いている統計表との矛盾が発生しうる。これは、現在利用可能な公開データが存在しないことから、利用可能な統計データのみを使用せざるを得ないためである。したがって、合成手法により求められた合成人口には、実社会の真の人口との乖離が残っていることに注意が必要である。

これらの合成手法は大きく分けて、初期世帯生成法、最適化に用いる統計表、目的関数、最適化の4つの要素からなる。

本章では、池田ら¹²⁾、福田ら¹³⁾、柁井ら¹⁴⁾、原田ら¹⁵⁾の分散手法、原田ら¹⁶⁾、Murata et al.¹⁷⁾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12)~17)を参照されたい。

2.1 池田ら¹²⁾の手法

2.1.1 要素1：初期世帯生成法

合成人口データの初期生成では、世帯数 H だけ世帯を生成する。世帯の生成時、世帯の家族類型を家族類型別世帯数を出現確率とする乱数によって確率的に決定する（以降、統計表を出現確率とする乱数によって確率的に決定する方法を確率的に決定すると略す）。

世帯の家族類型を決定した後、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を決定した後、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50%の確率でランダムに設定する。それ以外の構成員は、夫であれば男性、妻であれば女性など、性別を世帯の役割の属性に応じて設定する。

2.1.2 要素2：最適化に用いる統計表

池田ら¹²⁾はつぎ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¹²⁾は以下の目的関数を用いて最適化を行なっている。

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

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

ここで、 A は合成人口データ、 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。(2)式では、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12),13)を参照されたい。

2.1.4 要素4：最適化の手続き

池田ら¹²⁾は以下の手続きにより最適化している。

Step 1 合成人口データを初期生成

Step 2 探索回数が規定数に達すれば探索を終了

Step 3-1 合成人口データ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¹³⁾の手法

福田ら¹³⁾は池田ら¹²⁾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

そのほかの3つの要素は池田ら¹²⁾の手法と同様である。

2.3 柁井ら¹⁴⁾の手法

柁井ら¹⁴⁾は池田ら¹²⁾の目的関数をつぎのように変更し、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¹⁵⁾

原田らの分散手法¹⁵⁾では、柁井ら¹⁴⁾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¹⁵⁾では、統計表の統計量を N で除し、 N 個の合成人口データ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¹⁵⁾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¹⁶⁾手法

原田ら¹⁶⁾は柁井ら¹⁴⁾の目的関数を除く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¹⁶⁾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どお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を発生させないように設定する。

2.5.2 要素2：最適化に用いる統計表

原田ら¹⁶⁾は池田ら¹²⁾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¹⁶⁾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.¹⁷⁾の手法

Murata et al.¹⁷⁾は原田ら¹⁶⁾の手法をもとに、最適化の手續きのStep 3を以下のように変更している。

Step 3-1 合成人口データ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. 提案手法

本論文では、Fig. 1 (b)で示した仮想都市を生成する方法と2つの合成人口データの比較手法を説明する。仮想都市の生成と合成手法における初期解の生成は類似している。大きな違いとして、仮想都市の生成では、親より年齢の高い子供や、結婚できない年齢の夫婦が発生しないように、合成手法の初期解生成で用いていない親子年齢差と夫婦年齢差の統計表を用いて、統計上存在する親子や夫婦の年齢差になるように仮想個票データを生成している(注1)。

3.1 仮想都市の生成

仮想都市の生成では、世帯の生成と、世帯構成員の属性の設定の2つの手續きにより仮想個票データを生成する。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と世帯人員数を設定する際には、国勢調査人口等基本集計¹⁸⁾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Fig. 2の9種類の家族類型別、世帯人員別の世帯数を算出し、生成する世帯の家族類型と世帯人員数を確率的に決定する。なお、Fig. 2の9種類以外の家族類型は本論文が用いる合成手法^{13)~17)}の対象外である。

世帯の家族類型と世帯人員数を決定後、世帯構成員の性別と年齢を設定する。国勢調査人口等基本集計¹⁸⁾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、同表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¹⁹⁾保管表出生表1、同表2から、父の年齢別出生数と母の年齢別出生数を算出する。

単独世帯の構成員の年齢と性別は単独世帯の男女別、年齢別の人口をもとに確率的に設定する。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、子供の順に年齢を設定する。その際に、ひとり親と子供に該当する構成員は性別も設定する。

世帯に夫婦の両親が存在する場合は夫の父、夫の母の順に年齢を設定する。夫の父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。夫の母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、人口をもとに年齢と性別を確率的に決定する。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、妻の順に年齢を設定する。一方で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、人口をもとに年齢と性別を確率的に決定する。

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する。その後、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 合成人口データの比較

合成人口データと仮想個票データの比較手順の概要を以下に示す。

Step 1 比較する合成人口データ(A)と仮想個票データ(B)について、それぞれ以下の処理を実施する

Step 1-1 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 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 合成人口データ(A)と仮想個票データ(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 合成人口データAの i 番目の世帯 a_i との差が d である世帯 b_j を仮想個票データ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 をそれぞれの合成人口データから取り除き、 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。つぎに、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。たとえば、「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。つぎに、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。

(注1) 合成手法では、初期解生成にかかる時間を短縮するため、最適化の過程で、親子や夫婦の年齢差の統計表を考慮することにより、世帯構成における矛盾を解消している。

Table 1 Virtual statistics of # of households and population by family type and gender

Family type	# of households	Male population	Female population
Single	234,646	119,468	115,178
Couple	90,674	90,674	90,674
Couple and children	125,564	229,200	221,983
Father and children	6,044	8,490	5,359
Mother and children	34,379	33,889	46,700
Couple and parents	402	804	804
Couple and a parent	2,454	2,720	4,642
Couple, children, and parents	1,138	3,318	3,286
Couple, children, and a parent	4,699	9,121	12,722
Total	500,000	497,684	501,348

Table 2 Virtual statistics of # of households by family type and member

Family type	Member											
	1	2	3	4	5	6	7	8	9	10	11	12
Single	234,646	0	0	0	0	0	0	0	0	0	0	0
Couple	0	90,674	0	0	0	0	0	0	0	0	0	0
Couple and children	0	0	64,506	49,322	10,303	1,222	167	39	3	1	0	1
Father and children	0	4,545	1,271	202	20	4	2	0	0	0	0	0
Mother and children	0	24,519	8,168	1,460	199	21	10	2	0	0	0	0
Couple and parents	0	0	0	402	0	0	0	0	0	0	0	0
Couple and a parent	0	0	2,454	0	0	0	0	0	0	0	0	0
Couple, children, and parents	0	0	0	0	436	517	159	25	1	0	0	0
Couple, children, and a parent	0	0	0	2,321	1,798	503	66	10	1	0	0	0
Total	234,646	119,738	76,399	53,707	12,756	2,267	404	76	5	1	0	1

Step 2-2 で用いる世帯間の差は次式で計算する。

$$h(a_i, b_j) = \sum_{m=1}^M |a_{i,m}^{age} - b_{j,m}^{age}| \quad (4)$$

ここで、 M は a_i と b_j の世帯人員数、 $a_{i,m}^{age}$ は世帯 a_i の m 番目の構成員の年齢である。 A_i と B_j は世帯の家族類型、世帯人員数、ひとり親・子供の性別の内訳をもとに分類後、年齢や性別、世帯内の役割によって世帯構成員がソートされている。そのため、 $h(a_i, b_j) = 0$ の場合、本論文が比較対象とする年齢、性別、世帯内の役割、家族類型、世帯人員数については、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、合成人口データ A と仮想個票データ 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。なお、合成人口データと仮想個票データで Step 1-1 で細分化した世帯数が異なる場合、超過した世帯は比較不可能な世帯として集計する。

4. 実験結果

本論文では、合成人口データと仮想個票データの一致度を比較するために、まず、仮想都市を生成し、仮想統計表を集計する。つぎに、各手法により合成された合成人口データと仮想個票データを比較する。本論文が用いた計算機の 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。

仮想都市の生成では、平成 27 年度国勢調査¹⁸⁾と 2015 年の人口動態職業・産業別集計¹⁹⁾における東京都の統計表を用い

た。なお、生成する世帯数は $H = 500,000$ 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集計し、人口を合成する。

本論文では、福田ら¹³⁾の手法、柘井ら¹⁴⁾の手法、原田らの分散手法¹⁵⁾、原田ら¹⁶⁾の手法、Murata et al.¹⁷⁾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¹³⁾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^{14)~17)}は初期温度 1.0 度、収束温度 0.1 度と設定した。冷却関数にはいずれの手法も指数冷却を用いた。原田らの分散手法¹⁵⁾では、分割数 $N = 8, 16$ の 2 種を実験した。

従来手法の評価指標である仮想統計表と合成統計表の差を **Table 3** に示す。Table 3 では池田ら¹²⁾が用いた 9 種類の統計表を対象に、柘井らが提案する (3) 式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、最も採用される統計表が多い池田ら¹²⁾の統計表を用いて再評価した。その際に、(3) 式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.¹⁷⁾と原田ら¹⁶⁾の手法はほかの 4 手法と比べ、誤差が少ない。これらの手法は対象地域の世帯数と男女別人口が一致するように初期世帯を生成したことで、人口分布の誤差が少ない初期解が生成され、探索の多くを父子年齢差、母子年齢差、夫婦年齢差の統計表の誤差

Table 3 Previous evaluation method (difference between virtual statistics and synthetic statistics)

	Fukuta et al. ¹³⁾	Masui. et al. ¹⁴⁾	Parallelization method ¹⁵⁾		Harada et al. ¹⁶⁾	Murata et al. ¹⁷⁾
			$N = 8$	$N = 16$		
Average	149,905.5	55,316.3	55,463.7	55,089.1	55.6	15.2
St. dev.	41,195.1	621.2	1,052.9	1,285.2	11.3	5.6

Table 4 Proposed evaluation method (difference between virtual population and synthetic population)

d	Fukuta et al. ¹³⁾	Masui et al. ¹⁴⁾	Parallelization method ¹⁵⁾		Harada et al. ¹⁶⁾	Murata et al. ¹⁷⁾
			$N = 8$	$N = 16$		
0	72.24%	73.27%	73.27%	73.28%	76.00%	76.28%
1	8.55%	8.46%	8.50%	8.48%	8.05%	8.17%
2	4.19%	4.08%	4.08%	4.07%	3.78%	4.09%
3	2.38%	2.13%	2.11%	2.13%	1.94%	2.18%
4	1.53%	1.32%	1.31%	1.32%	1.18%	1.35%
5	1.10%	0.92%	0.93%	0.92%	0.79%	0.93%
6	0.86%	0.73%	0.72%	0.72%	0.58%	0.70%
7	0.74%	0.63%	0.63%	0.64%	0.46%	0.56%
8	0.61%	0.54%	0.55%	0.54%	0.37%	0.47%
9	0.44%	0.46%	0.46%	0.46%	0.31%	0.40%
10-	4.01%	4.12%	4.06%	4.13%	6.36%	4.71%
Others	3.35%	3.35%	3.39%	3.33%	0.18%	0.17%

の改善に使用できた。加えて、Murata et al. ¹⁷⁾ 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ほかの統計表との誤差を削減することができた。そのため、Murata et al. ¹⁷⁾ の手法は最も統計表との誤差を削減することができた。

各手法の合成人口データと仮想個票データとの世帯間の差別の一致割合を **Table 4** に示す。また、家族類型別、世帯間の差別の一致割合を **Fig. 3** に示す。Table 4 と Fig. 3 の Others は比較不可能な世帯の割合である。比較不可能な世帯を集計している理由は、世帯の家族類型、世帯人員数、ひとり親・子供の性別の内訳が異なる世帯やこれらの数が異なる場合があり、構成員の年齢の差のみで世帯間の差を定義している (4) 式を適用できないためである。Fig. 3 の C., ch., and parents は夫婦、子供と両親世帯、C., ch., and a parent は夫婦、子供とひとり親世帯である。Table 4 から、福田ら ¹³⁾ の手法により合成された合成人口データと仮想個票データにおいて完全に一致する世帯 $d = 0$ の割合は 72.24% である。同様に、柁井ら ¹⁴⁾ および原田らの並列分散手法 ¹⁵⁾ ($N = 8$) は 73.27%、原田らの並列分散手法 ¹⁵⁾ ($N = 16$) は 73.28%、原田ら ¹⁶⁾ の手法は 76.00%、Murata et al. ¹⁷⁾ の手法は 76.28% であった。

Fig. 3 から、世帯間の差 d 別の割合は家族類型によって大きく異なることがわかる。Fig. 3 と Table 2 を家族類型別に比較すると、単独世帯や夫婦のみ世帯など、世帯数が多く世帯人員数が少ない家族類型は仮想個票データと一致する世帯の割合が多く、夫婦、子供と両親世帯のように世帯数が少なく世帯人員数が多い家族類型は仮想個票データと一致する世帯の割合が少ないことがわかる。これは、世帯数が多いほど検索対象世帯数が増加し、世帯人員数が少ないほど 2 つの世

帯が一致する可能性が増加することで、一致する世帯の割合が増加した。

各手法の差を観察するために、Fig. 3 をもとに世帯間の差 $d = 0$ と $d \geq 10$ の平均値を算出した。世帯間の差 $d = 0$ の平均値は福田ら ¹³⁾ の手法では 33.18%、柁井ら ¹⁴⁾ の手法は 33.40%、原田らの並列分散手法 ¹⁵⁾ ($N = 8$) は 33.45%、原田らの並列分散手法 ¹⁵⁾ ($N = 16$) は 13.65%、原田ら ¹⁶⁾ の手法は 36.11%、Murata et al. ¹⁷⁾ の手法は 36.24% であった。世帯間の差 $d \geq 10$ の平均値は福田ら ¹³⁾ の手法では 19.87%、柁井ら ¹⁴⁾ の手法は 20.79%、原田らの並列分散手法 ¹⁵⁾ ($N = 8$) は 20.80%、原田らの並列分散手法 ¹⁵⁾ ($N = 16$) は 20.55%、原田ら ¹⁶⁾ の手法は 21.59%、Murata et al. ¹⁷⁾ の手法は 20.79% と各手法でそれほど大きな差異はない。一方で、比較不可能な世帯割合の平均値は福田ら ¹³⁾ の手法では 13.65%、柁井ら ¹⁴⁾ の手法は 13.65%、原田らの並列分散手法 ¹⁵⁾ ($N = 8$) は 13.65%、原田らの並列分散手法 ¹⁵⁾ ($N = 16$) は 13.78%、原田ら ¹⁶⁾ の手法は 1.05%、Murata et al. ¹⁷⁾ の手法は 1.13% である。

これらのことから、Murata et al. ¹⁷⁾ と原田ら ¹⁶⁾ 手法が仮想個票データとの一致度が高いことがわかる。原田ら ¹⁶⁾ と Murata et al. ¹⁷⁾ 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ことから、仮想個票データと一致する世帯の割合が増加した。また、初期解を統計表どおりに生成する原田ら ¹⁶⁾ と Murata et al. ¹⁷⁾ の手法は類似する世帯を探索することが不可能な世帯の割合が減少しており、仮想個票データと類似する世帯構成を多く生成できている。加えて、Murata et al. ¹⁷⁾ の手法は人口分布と一致する初期解を生成し、年齢を交換し最適化をすることで、人口分布の誤差を発生させずに影響なく父子・

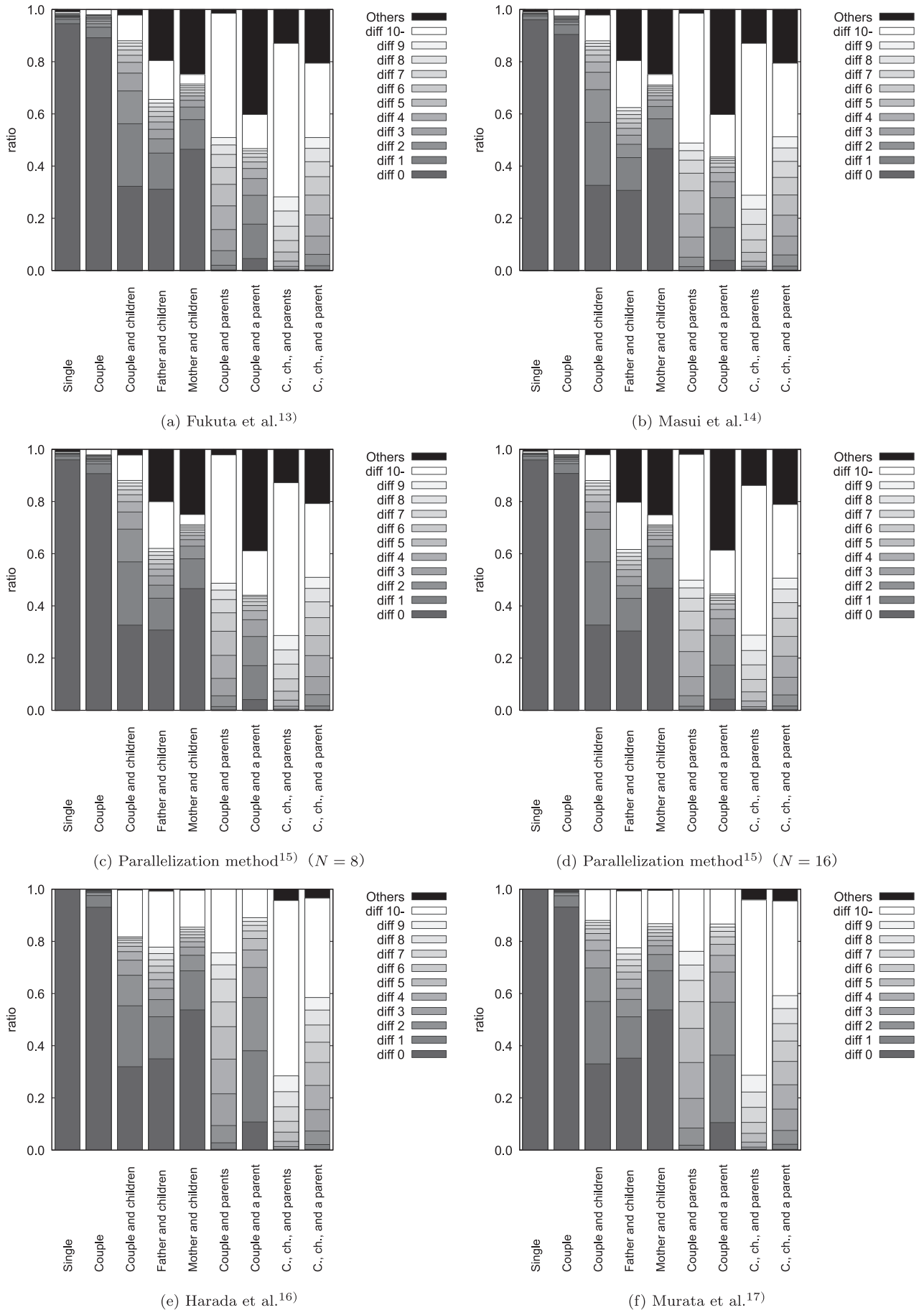


Fig. 3 Difference between virtual population and synthetic population by family type and synthesized method

母子・夫婦年齢差の統計表の誤差の削減が可能である。これにより、仮想個票データに類似する世帯の割合がさらに増加した。

Table 3 の統計表との誤差と Table 4, Fig. 3 の仮想個票データとの一致度から、統計表との誤差が少ない手法ほど仮想個票データとの一致度が高い傾向にある。そのため、統計表との誤差が少ない合成人口データを合成する手法を開発できると仮想個票データとの一致度も向上すると考えられる。

5. おわりに

本論文では、合成人口データと仮想個票データとの一致度による合成手法の評価手法を提案した。500,000 世帯、999,032 人の仮想統計表を生成し、福田ら¹³⁾の手法、柁井ら¹⁴⁾の手法、原田ら¹⁶⁾の手法、Murata et al.¹⁷⁾の手法を比較した。

各合成手法の合成人口データと仮想個票データとの一致度について、福田ら¹³⁾の手法は 72.24%、柁井ら¹⁴⁾および原田らの並列分散手法¹⁵⁾ ($N = 8$) は 73.27%、原田らの並列分散手法¹⁵⁾ ($N = 16$) は 73.28%、原田ら¹⁶⁾の手法は 76.00%、Murata et al.¹⁷⁾の手法は 76.28%であった。確率的に世帯を生成する福田ら¹³⁾や柁井ら¹⁴⁾、原田らの並列分散手法¹⁵⁾の手法に比べ、家族類型別世帯数や男女別人口など、統計表と整合するように合成する原田ら¹⁶⁾と Murata et al.¹⁷⁾の手法は、仮想個票データとの一致度が高い結果であった。

本論文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、家族類型、世帯人員数である。これらの属性以外の属性を生成する合成手法と比較する場合、対応する属性を含む仮想都市の生成と、3.2 節で説明した比較手法や (3) 式を変更する必要がある。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、複数の手法間で仮想個票データ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 本論文の一部は、JSPS 科研費 19K23229, JST 未来社会創造事業 JPMJMI20B3, 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ある。

参考文献

- 1) J.M. Epstein and R. Axtell: *Growing Artificial Societies: Social Science from the Bottom Up*, 1st edition, The MIT Press (1996)
- 2) N. Gilbert: *Agent-Based Models*, **153**, SAGE Publications (2008)
- 3) 高橋●●: 社会システムの研究動向 3—評価・分析手法 (1)—モデルの解像度と妥当性評価, 計測と制御, **52-7**, 582/587 (2013)
- 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, *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 1144/1149 (2018)
- 6) 村田, 原田, 杉浦: リアルスケール社会シミュレーションの実現に向けて, **36**, 58/62 (2017)
- 7) 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)
- 8) J. Barthélemy and P.L. Toint: Synthetic Population Generation Without a Sample, *Transportation Science*, **47-2**, 266/279 (2013)
 - 9) 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)
 - 10) 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**, 112 (2013)
 - 11) 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)
 - 12) 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)
 - 13) 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)
 - 14) 柁井, 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)
 - 15) 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)
 - 16) 原田, 村田, 柁井: 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)
 - 17) 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)
 - 18) 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)
 - 19) 総務省統計局: e-Stat 人口動態職業・産業別統計 2015 年度, <https://www.e-stat.go.jp/stat-search/files?page=1&layout=datalist&toukei=00450011&tstat=000001028897&year=20151&tclass1=000001053122> (2018)

[著者紹介]

原田 拓 弥 (正会員)



2018 年関西大学大学院総合情報学研究科博士課程後期課程修了。博士 (情報学)。青山学院大学理工学部経営システム工学科助教を経て現在、芝浦工業大学システム理工学部電子情報システム学科助教。大規模な社会シミュレーションの研究に従事。2019 年度計測自動制御学会論文賞受賞。経営情報学会の会員。

村田 忠彦 (正会員)



1997 年大阪府立大学大学院工学研究科博士後期課程修了。博士(工学)。現在、関西大学総合情報学部教授。シカゴ大学計算研究所客員研究員、IEEE SMCS 日本支部長、IEEE SMCS Soft Computing 技術委員長、IEEE SMCS Awareness Computing 技術委員長、IEEE CIS Neural Networks 技術委員、IEEE SMCS Associate Vice President、日本進歩計算学会会長などを歴任。システム制御情報学会奨励賞、IEEE SMCS 最優秀技術委員会賞を受賞。現在、社会シミュレーションの研究に従事。システム制御情報学会、知能情報ファジィ学会、情報処理学会、IEEE などの会員。

高橋 真吾 (正会員)



1989 年東京工業大学大学院総合理工学研究科システム科学専攻博士課程修了、理学博士。現在、早稲田大学理工学術院教授(創造理工学部経営システム工学科)。システム論、社会シミュレーション、ソフトシステムアプローチなどの研究に従事。

.....