

世帯を単位とした将来推計における 初期世帯マイクロデータ生成の高速化

阪田 知彦¹・鈴木 温²・杉木 直³・正木 俊行⁴・田 寛之⁵

¹正会員 国立研究開発法人建築研究所住宅・都市研究グループ（〒305-0802 茨城県つくば市立原1）

E-mail: sakata@kenken.go.jp

²正会員 名城大学理工学部社会基盤デザイン工学科（〒468-8502 名古屋市天白区塩釜口1丁目501番地）

E-mail: atsuzuki@meijo-u.ac.jp

³正会員 豊橋技術科学大学 建築・都市システム学系（〒441-8580 豊橋市天伯町雲雀ヶ丘1-1）

E-mail: sugiki@ace.tut.ac.jp

⁴非会員 アカデミックエクスプレス株式会社（〒305-0047 茨城県つくば市千現2-1-6）

E-mail: masaki@academic-express.com

⁵非会員 アカデミックエクスプレス株式会社（〒305-0047 茨城県つくば市千現2-1-6）

E-mail: den@academic-express.com

本研究は、世帯マイクロシミュレーションに基づいた将来都市構造の予測技術の開発のうち、初期マイクロデータ生成に関する実用化の一環として、水流らの方法をベースとしつつ、誤差修正の計算時間を大幅に短縮するためのアルゴリズムを改良・追加し、かつ既存研究よりもさらに精度の高い推計が可能な初期マイクロデータの推計方法を提案することを目的としている。また、この初期マイクロデータ生成のための計算アルゴリズムの有効性を検証するため、富山県富山市を対象として、今回提案する計算アルゴリズムを実装し、オープンデータを用いた検証を実施した。その結果、計算時間が大幅に減少できることを確認した。

Key Words : *householdbased micro-simulation, population synthesis, open data, simulated annealing*

1. はじめに

本格的な人口減少期に入った我が国の今後の都市構造を考えるには、多様な観点からの検討が必要であるが、その1つとして、将来の住宅立地や世帯分布の動向把握が不可欠であると考えている。それを支える技術は様々あるが、その1つに世帯マイクロシミュレーション（HUMS:Household based Urban Micro-Simulation model）に基づいた将来予測¹⁾が提案されている。世帯マイクロシミュレーションを構成する大きな2つの要素として、①現況の個々人の個人属性と世帯構成を再現する初期マイクロデータの生成と、②個々のライフイベント変化をモデル化した将来予測、がある。そのうち、本研究では、①の初期マイクロデータ生成を対象としている。初期マイクロデータの推計は、得られている統計データの集計値の組み合わせにより、個々のマイクロデータの属性と分布を推定しようとするものである。初期マイクロデータは、対象とする母集団のすべての個票データがあれば正確に再現できるが、プライバシー保護の観点から、すべ

ての個票を入手することは難しい。また、独自にアンケート調査等を実施しても一部のサンプルデータに留まるケースが多い。入手可能な統計データから初期マイクロデータを推計する方法は、これまでに様々な方法が提案されており、これらについては、2章で説明する。このうち、水流らは、国勢調査等、一般に公表されているオープンデータのみを用いて、鈴木ら¹⁾の開発した世帯マイクロシミュレーションモデルの初期マイクロデータを推計する方法論を提案している。水流ら²⁾の方法は、オープンデータのみから、ある程度精度の高い初期マイクロデータの推計が可能な方法であるが、誤差を減少させるために、長い計算時間を要していることが課題であった。

本研究は、水流ら²⁾の方法をベースとしつつ、誤差修正の計算時間を大幅に短縮し、かつ既存研究よりもさらに精度の高い推計が可能な初期マイクロデータの推計方法を提案することを目的としている。また、この初期マイクロデータ生成のための計算アルゴリズムの有効性を検証するため、富山市を対象として、計算アルゴリズム

を実装し、オープンデータを用いた検証を実施した。なお本研究は、建築研究所研究課題「将来都市構造の予測・評価手法の高度化による目標管理・推進評価技術の開発（平成28～令和3年度）」の一環で実施した。

2. 既存研究と本研究の位置づけ

入手可能な統計データから、個票データの合成データを作成する方法は、Synthetic Reconstruction Method (SR法)として知られている。また、人口データを合成する方法は、Population Synthesis (人口合成)と呼ばれることもある。Population Synthesisの方法として、これまで最も多く用いられてきた方法は、iterative proportional fitting (IPF)という方法であり、1940年にDeming et al.³⁾が提案しており、非常に古くから使われている手法でもある。しかし、IPFは、属性数が増えるほどゼロセル問題が顕著になり、信頼性が低下するといった問題⁴⁾や、個人の統計と世帯の統計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を行うことが困難であるといった問題⁵⁾が指摘されている。

Moeckel et al.⁴⁾や杉木ら^{6),7),8)}は、従来のセルベースの方法では無く、エージェントベースの推計方法を提案している。杉木ら^{6),7),8)}の方法では、アンケートデータを元にしたサンプルデータから、主成分分析によって世帯構造を推定し、さらに、モンテカルロシミュレーションによって、マイクロデータの生成を行っている。

これに対し、Lenormand and Deffuant⁹⁾は、サンプルデータを用いない推計方法 (Sample-free method)の方が、より良く個人や世帯を合成できていることを示した。また、統計から集団を合成する方法としては、Simulation-Based Method (SBM)が有効であると指摘されている¹⁰⁾。SBMの代表的な手法には、マルコフ連鎖モンテカルロ法 (MCMC)¹¹⁾を用いる方法やsimulated annealing(SA)法が提案されている。このうち、simulated annealing(SA)法は、やきなまし法とも呼ばれ、Population Synthesisに関わる多くの研究^{12),13),14),15),16)}で用いられている。Harland et al.¹²⁾は、他のアルゴリズムと比較し、SA法の有効性を示した。池田ら¹⁴⁾は、9区分の世帯タイプを対象に、9つの統計データに適合する世帯集団を合成する手法を提案している。また、榊井ら¹⁵⁾は、池田ら¹⁴⁾の手法を踏襲しつつ、目的関数や合成手法の改良を行った。しかし、これらの研究では、世帯タイプ別世帯数等は、推計されているが、ゾーン別世帯タイプ別世帯数等、空間的な世帯属性の違い等は考慮されていない。これに対し、水流ら²⁾は、前述の世帯マイクロシミュレーション (HUMS)の初期世帯マイクロデータを生成するため、国勢調査等のオープンデータのみを使用して、simulated annealing(SA)法を用いた方法を提案した。水流ら²⁾の方法は、ゾーンごとの年

齢別人口や世帯タイプ別世帯数等のデータを用いることで、ゾーンごとの世帯合成に関する方法論を提示し、初期世帯マイクロデータを推計することに成功したが、長い計算時間や誤差が一定以下には減らないことが課題であった。

そこで、本研究では、水流ら²⁾の方法を踏襲しつつ、高速かつ精度の高い推計が可能な初期マイクロデータの生成や誤差修正のアルゴリズムを提案することを目的としている。

3. 初期マイクロデータ生成の基本構造

高速化にあたっては、SA法を用いた先行研究²⁾の枠組みを継承しつつ、「SA法に inputs する初期解の構築アルゴリズムの導入」、 「SA法における有効入れ替え発生率の改善」の2点を主軸に改良を行った。

SA法は局所探索法と呼ばれる探索アルゴリズムの一種である。既に構築された解を出発点とし、現在の解の近傍にある解が現在の解よりも優れていれば (目的関数がより小さければ)、近傍解を現在の解として次のステップへ移す。先行研究及び本研究においては、個人を世帯に紐付けたマイクロデータを解とし、現在のマイクロデータに対して2つの世帯間で個人を入れ替えを行ったマイクロデータを近傍解と定義する。

また、目的関数として、推定したマイクロデータ (現在の解) と周辺分布との差の絶対値により表される誤差関数を導入する。本研究においては、「夫婦の年齢別夫婦数」および「出生順別母親と子供の年齢差別人口」の2つの周辺分布を用いた。

(1) 初期解構築アルゴリズムの導入

前述の通り、SA法は既にある解を逐次改善していくアルゴリズムであり、実行にあたっては何らかの方法で構築された初期解を必要とする。先行研究においては、初期解の構築を個人と世帯をランダムに結合することで行っているが、ランダムに結合された個人と世帯からなるマイクロデータは周辺分布から大きくかけ離れたものになってしまう。このランダムなマイクロデータから周辺分布に近いマイクロデータを得るには、膨大な反復回数を必要とする。

仮に初期解の段階である程度、周辺分布に沿った解が得られていれば、SA法の反復回数は劇的に減少すると考えられる。本研究では、周辺分布を考慮した初期解構築アルゴリズムを導入した。

a) SA法における誤差関数の改善傾向

初期解構築アルゴリズムを考案するにあたって、後段のSA法の特性に着目する。図-1に、SA法における周辺

分布誤差の収束速度を示す。本研究では「夫婦の年齢別夫婦数」および「出生順別母親と子供の年齢差別人口」の2つの周辺分布を誤差関数に用いているが、母子の周辺分布誤差に比べ、夫婦の周辺分布誤差の収束速度は著しく遅いことがわかる。SA法開始直後、母子と夫婦の周辺分布誤差は概ね同程度の値を取るが、母子の周辺分布誤差は反復回数が200万回未満の時点で収束しているのに対し、夫婦の周辺分布は反復回数が2000万回を超えても収束しきっていない。

このことから、SA法による改善が難しい夫婦の周辺分布誤差を初期解構築時点で予め小さくできていれば、SA法の収束速度を大幅に改善できると推測される。

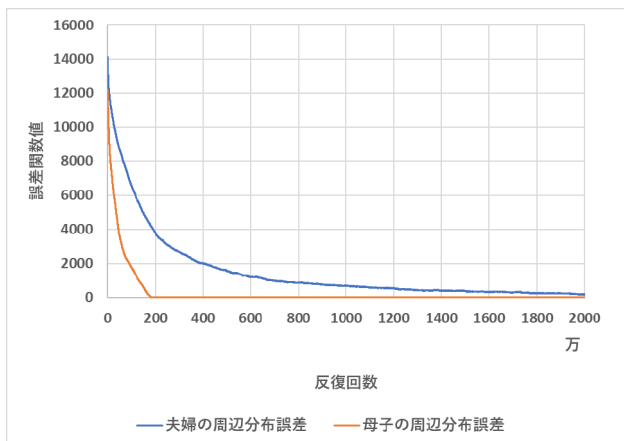


図-1 SA法における周辺分布誤差の収束速度の差

b) 初期解構築アルゴリズムの概略

世帯に個人を割り当てたものがマイクロデータ、すなわちSA法における解であるため、初期解の構築は各世帯に個人を割り当てることに相当する。

世帯への個人の割り当ては、1世帯ずつ順に行う。まだ世帯に割り当てられていない個人を「未割り当て個人」とし、既に世帯に割り当てられた個人を「割り当て済み個人」とする。一度世帯に割り当てられた個人が再度「未割り当て個人」となることはない。

STEP 1 親の年齢候補の決定

未割り当て個人の中から「最も若い大人世代の妻」の年齢（5歳階級）候補をリスト化する。

STEP 2 子供の年齢の決定

世帯内に子供世代が存在する場合、親の年齢候補から逆算して子供の最小年齢を算出し、「未割り当て個人」から最小年齢以上の個人を当該世帯の子供として割り当てる。

STEP 3 親の年齢の決定

STEP 2で決定した子供の年齢を基に、全ての子供の

母親となりうる「最も若い大人世代の妻」の年齢候補を絞り込む。

ここで、現時点で作成済みの世帯を対象に年齢別夫婦数を集計し、周辺分布との差分を計算する（図-2）。この周辺分布との差分から、以下の手順で「夫婦の望ましい年齢組み合わせリスト」を作成する。

- ・ 図-2の表の各列の最小値が負である場合、すなわち、その組み合わせの夫婦が不足している場合には、その組み合わせをリストに追加する。
- ・ 列の最小値が0である場合、すなわち、不足している組み合わせが存在しない場合には、周辺分布の割合に従って行（夫の年齢）を決定し、その組み合わせをリストに追加する。

「妻の年齢候補」と「『夫婦の望ましい年齢組み合わせリスト』に含まれる妻の年齢リスト」の積集合を取り、残った候補の中から妻の年齢をランダムに決定する。妻の年齢と「夫婦の望ましい年齢組み合わせリスト」から、夫の年齢が一意に定まる。

STEP 4 祖父母、曾祖父母世代の年齢の決定

「最も若い大人世代」の上に世代が存在する場合は、下の世代の夫の年齢をSTEP 3における子供の年齢とみなし、STEP 3の手順で年齢を決定する。

		妻の5歳階級																
		4	5	6	7	8	9	10	11	12	13	14	15	16	17	18		
夫の5歳階級	4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
	6	-1	0	2	-1	-1	0	0	0	0	0	0	0	0	0	0		
	7	0	-1	1	0	-1	0	1	0	0	0	0	0	0	0	0		
	8	0	0	1	-1	-1	-1	0	1	0	0	0	0	0	0	0		
	9	0	0	0	1	-1	0	0	2	0	0	0	0	0	0	0		
	10	0	0	2	1	-1	1	-1	1	0	0	0	0	0	0	0		
	11	0	0	1	0	-1	1	-2	1	0	0	0	0	0	0	0		
	12	0	0	0	0	0	1	0	-1	0	0	0	1	29	0	0		
	13	0	0	1	0	2	0	2	0	-1	-1	2	12	46	0	79		
	14	0	0	0	0	3	0	2	0	-1	-1	2	-8	13	24	0		
	15	0	0	0	0	0	0	0	0	0	0	0	-8	8	28	43		
	16	0	0	0	0	0	0	0	0	0	0	2	-8	-22	30	67		
	17	0	0	0	0	0	0	0	0	0	1	0	5	-7	-23	-24	48	
	18	0	0	0	0	0	0	0	0	0	0	0	-9	-24	-56	-235		

図-2 初期解構築過程における年齢別夫婦数の周辺分布からの誤差

（太枠は「夫婦の望ましい年齢組み合わせ」）

着目すべきは、STEP 3で現状の「夫婦の周辺分布誤差」を参照し、不足している組み合わせの夫婦を生成するようにしていることである。これにより、ランダムに生成した初期解と比較して「夫婦の周辺分布誤差」を飛躍的に低減させることが可能となる。

また、STEP 2で子供の、STEP 3で親の年齢を決定する際に、「母子の周辺分布誤差」を参照していないことも重要である。これは初期解の時点で両方の周辺分布をバランス良く小さくするのではなく、「夫婦の周辺分布誤差」を小さくすることのみに注力しているためである。

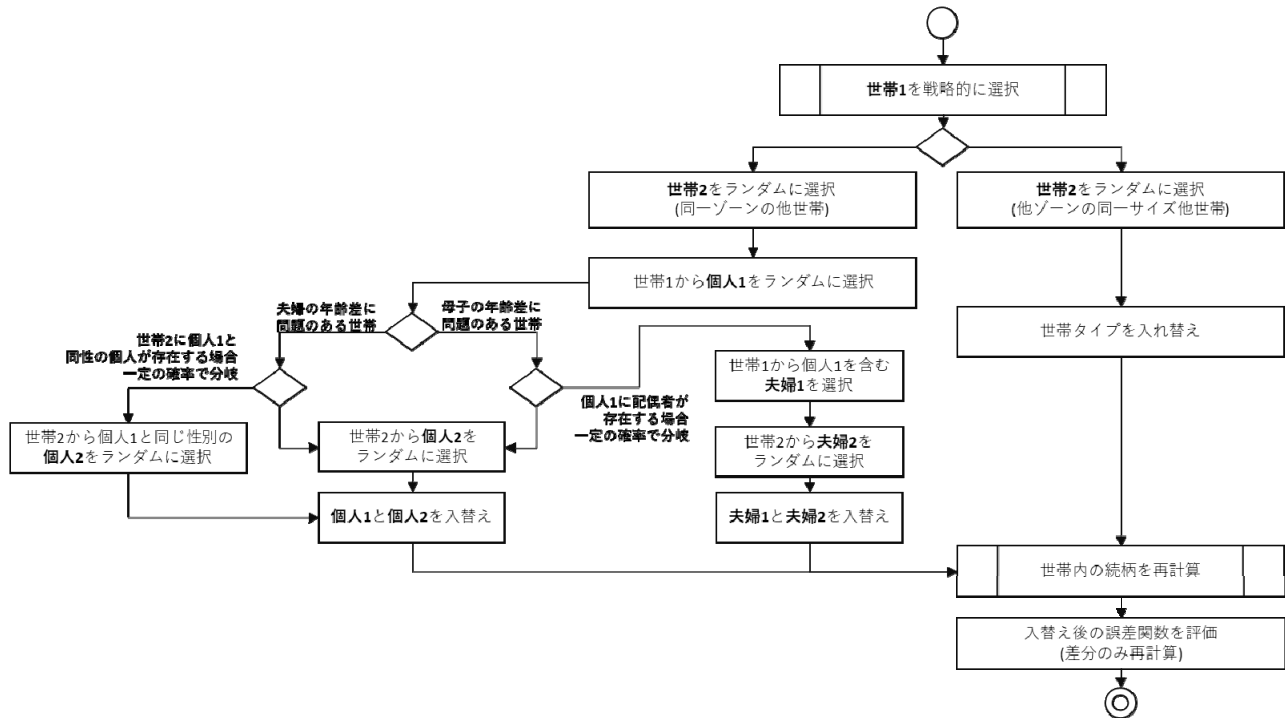


図-3 SA法の処理フロー (1反復分)

その代償として、生成された初期解における「母子の周辺分布誤差」は「夫婦の周辺分布誤差」と比べて大きくなってしまふものの、(1)-a)で述べたように「母子の周辺分布誤差」は後段のSA法により比較的高速に収束するため、初期解の時点で誤差が大きいことはほとんど問題にならない。

ここでは、誤差関数が改善された入れ替え回数が入れ替え回数に占める割合を「有効入れ替え発生率」と呼ぶことにする。本研究では有効入れ替え発生率を向上させるために、誤差関数が改善される可能性が高い近傍を優先的に選択するルールを導入した。

(2) SA法における有効入れ替え発生率の改善

SA法そのものは近傍解の選択方法を規定しないため、個々の問題に応じて適切な近傍解の選択基準を考える必要がある。最も単純な選択基準は近傍を完全にランダムに選択することであり、先行研究においてもこの方法を採っている。しかしながら、マイクロデータにおいて全くランダムに選択した2人の個人の入れ替えによって誤差関数が改善される可能性は非常に低い。

a) 入れ替え世帯選択ルール

本研究におけるSA法の処理フローを図-3に示す。このフローにおいて最も重要なのは、最上段において「世帯1を戦略的に選択」する部分である。

図-5および図-4に、SA法適用過程における「母子の周辺分布誤差」と「夫婦の周辺分布誤差」の例を示す。両図において、青色のセルは周辺分布に対して数が不足している組み合わせを、赤色のセルは周辺分布に対して過剰な組み合わせを示している。誤差関数を減少させる

		妻の5歳階級															
		4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	
夫の5歳階級	4	-2	3	0	0	0	0	0	0	0	0	0	0	0	0	0	
	5	12	-2	-23	-3	0	0	0	0	0	0	0	0	0	0	0	
	6	-6	6	73	-31	-6	-8	0	0	0	0	0	0	0	0	0	
	7	0	-4	-29	12	-84	-6	-2	0	0	0	0	0	0	0	0	
	8	-4	2	-15	24	92	-39	-4	1	0	0	0	0	0	0	0	
	9	0	-8	2	-4	40	105	37	-1	1	0	0	0	0	0	0	
	10	0	3	-9	6	-30	-78	77	3	-7	1	0	0	0	0	0	
	11	2	-1	6	-4	-15	23	-67	107	-9	20	0	0	0	0	0	
	12	0	0	-3	-5	6	-8	-49	-107	4	70	2	29	0	0	0	
	13	0	0	5	6	-9	10	2	17	13	-16	62	100	140	0	196	
	14	0	0	0	0	6	1	6	-8	0	-63	181	172	255	303	0	
	15	0	0	0	0	0	0	1	-3	3	-8	-148	246	497	675	533	
	16	0	0	0	0	0	0	0	0	-6	-1	-71	-396	260	841	1032	
	17	0	0	0	0	0	0	0	0	0	1	-3	-12	-123	-880	-64	
	18	0	0	0	0	0	0	0	0	0	0	0	-7	-28	-278	-1746	

図-4 SA法の適用過程における年齢別夫婦数の周辺分布との誤差の例

ためには、赤いセルの組み合わせを減らし、青いセルの組み合わせを増加させる必要がある。

したがって、「世帯1」として赤いセルの組み合わせを含む世帯を優先的に選択することで、誤差関数を改善できる見込みは高くなると考えられる。これが「世帯1の戦略的選択」である。

母子の5歳階級差	第1子	第2子	第3子
3	2856	156	-1
4	924	621	-54
5	-1666	-338	-223
6	-1185	-1443	-549
7	-655	247	213
8	-274	757	612

図-5 SA法の適用過程における出生順別母親と子供の年齢差別人口の周辺分布との誤差の例

b) 世帯間での入れ替えを行う個人の選択ルール

その他、先行研究からの改良点として、世帯間でランダムに個人を入れ替えるだけでなく、一定の確率で「世帯間で同じ性別の個人を入れ替える」、さらに「世帯間で個人ではなく夫婦を入れ替える」という操作を行っている点が上げられる。

前者は、子供世代を除いて世帯内の性別比は世帯タイプによって概ね一定（例えば世帯内に3世代の夫婦のみが存在する場合、男女比は必ず3:3となる）であり、性別の異なる個人を入れ替えると「続柄不明」の個人が発生するリスクが高くなってしまったため、これを防ぐ狙いがある。

後者は夫婦の「夫婦の周辺分布誤差」を変化させずに「母子の周辺分布誤差」のみを変化させるための工夫である。例えば世帯内に2世代の夫婦が存在する場合、その中の1人の個人を他の世帯と入れ替えると、必ず「夫婦の周辺分布誤差」が変化してしまう。しかし、構築した初期解は元々「夫婦の周辺分布誤差」が小さいため、SA法の過程で不用意に周辺分布誤差を変化させてしまうことは望ましくない。そこで、一定の確率で夫婦と夫婦の入れ替えを行うことで、「夫婦の周辺分布誤差」を変化させずに「母子の周辺分布誤差」のみを変化させる狙いがある。

(3) 導入の効果

図-6に、本アルゴリズムによって構築された初期解を用いた場合の、SA法の収束傾向を示す。初期解時点で「夫婦の周辺分布誤差」が1,000弱であるのに対して「母子の周辺分布誤差」は10,000強と、10倍以上の開きがあるが、「母子の周辺分布誤差」は急速に減少し、反

復回数が40万回を超えた時点で誤差の大小関係は逆転している。しかしながら「夫婦の周辺分布誤差」も初期値が小さいため、傾きは緩やかであるものの「母子の周辺分布誤差」からさほど遅れることなく収束していることがわかる。

また、表-1に示す通り、近傍選択ルールの改良によって約10倍の有効入れ替え発生率が得られている。

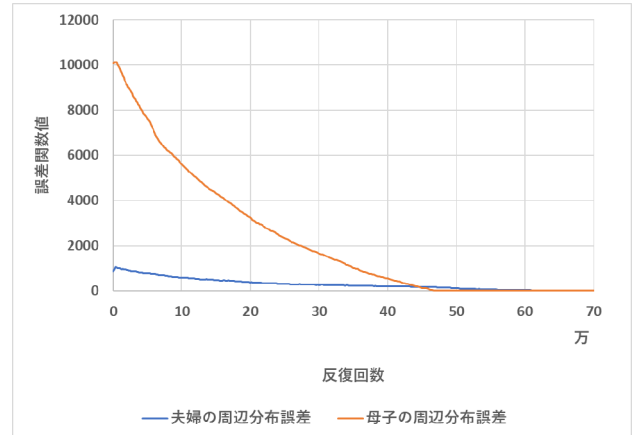


図-6 構築された初期解を用いた場合の、SA法における周辺分布誤差の収束傾向

表-1 SA法改良前後の有効入れ替え発生率の比較

	総反復回数	誤差関数が改善された回数	有効入れ替え発生率
改良前	30,334,061	304,360	1.00%
改良後	749,502	72,654	9.69%

4. オープンデータを用いた検証

前章までの検討を受けて、全体のプログラム化と富山県富山市を対象としたオープンデータを用いた検証を行った。

(1) 全体のプログラム化

プログラム化にあたっては、改良時の可読性等や計算に必要なライブラリの充実さを考慮してpythonでコーディングを行った。開発環境やライブラリ等を表-2に示す。

表-2 開発環境

実装言語	Python 3.5.2
使用パッケージ (非標準のみ記載)	numpy 1.13.1 panads 0.20.3
実動作環境	Windows Subsystem for Linux (Windows10標準搭載のLinux仮想環境)
CPU	Intel® Core™ i7-7500U (ノートPC向け超低消費電力)
メモリ	8GB

表-3 使用したオープンデータの例

パラメータ	設定根拠となるデータ
ゾーン別戸建-集合_持家-賃貸割合	平成27年国勢調査(確定値) 小地域集計結果 住宅 第6表住居の種類・住宅の所有の関係(6区分)別一般世帯数, 一般世帯人員及び1世帯あたり人員 - 町丁・字等 の地区別データ 平成27年国勢調査(確定値) 小地域集計 住宅 第7表住宅の建て方(7区分)別住宅に住む主世帯, 主世帯人員及び1世帯あたり人員 - 町丁・字等 の地区別データ
ゾーン別世帯人数別世帯数	平成 27年国勢調査 第4表 世帯の種類(2区分)、世帯人員(7区分)別一般世帯数、一般世帯人員、1世帯あたり人員、施設等の世帯数及び施設等の世帯人員-町丁・字等
ゾーン別性別5最階級別人口(女性)	平成 27年国勢調査 第2表 年齢(5歳階級)、男女別人口(総年齢、平均年齢及び外国人-特掲) - 町丁・字等
ゾーン別性別5最階級別人口(男性)	平成 27年国勢調査 第2表 年齢(5歳階級)、男女別人口(総年齢、平均年齢及び外国人-特掲) - 町丁・字等
出生順位別母子の年齢差	平成27年人口動態統計4 出生数、2500g未満の出生数(再掲)、性・母の年齢(各歳)・出生順位・都道府県(20大都市再掲)別
世帯タイプ別持家-賃貸割合	平成27年国勢調査 人口等基本集計(男女・年齢・配偶関係、世帯の構成、住居の状態など) 世帯の家族類型(16区分)、住居の種類・住宅の所有の関係(6区分)別一般世帯数及び一般世帯人員(3世代世帯-特掲)
世帯タイプ別世帯数	平成 27年国勢調査 表11 世帯の家族類型(16区分)、世帯人員(7区分)別一般世帯数(3世代世帯及び6歳未満・12歳未満・15歳未満・18歳未満・20歳未満世帯員のいる一般世帯-特掲)

表-4 初期マイクロデータ生成例(世帯テーブル)

ゾーン	世帯ID	世帯人数	世帯タイプ	世帯主年齢	世帯主の個人ID	住居形態	個人ID1	個人ID2	個人ID3	個人ID4	個人ID5	個人ID6	個人ID7	個人ID8
01 富山-0001 総曲輪	0	1	19	93	0	10	0							
01 富山-0001 総曲輪	1	1	19	71	1	10	1							
01 富山-0001 総曲輪	2	1	19	58	2	21	2							
01 富山-0001 総曲輪	3	1	19	75	3	21	3							
01 富山-0001 総曲輪	4	1	19	40	4	21	4							
01 富山-0001 総曲輪	5	1	19	77	5	10	5							
01 富山-0001 総曲輪	6	1	19	54	6	11	6							
01 富山-0001 総曲輪	7	1	19	73	7	21	7							
01 富山-0001 総曲輪	8	1	19	77	8	10	8							
01 富山-0001 総曲輪	9	1	19	90	9	21	9							
01 富山-0001 総曲輪	10	1	19	79	10	11	10							
01 富山-0001 総曲輪	11	1	19	87	11	21	11							
01 富山-0001 総曲輪	991	6	62	69	1844	11	1844	1845	1846	1847	1848	1849		
01 富山-0001 総曲輪	992	6	62	64	1850	11	1850	1851	1852	1853	1854	1855		
01 富山-0001 総曲輪	993	6	62	63	1856	11	1856	1857	1858	1859	1860	1861		
01 富山-0001 総曲輪	994	7	72	72	1862	21	1862	1863	1864	1865	1866	1867	1868	
01 富山-0001 総曲輪	995	7	72	72	1869	10	1869	1870	1871	1872	1873	1874	1875	
01 富山-0001 総曲輪	998	8	89	77	1876	10	1876	1877	1878	1879	1880	1881	1882	1883
01 富山-0002 愛宕	997	1	19	48	1884	10	1884							
01 富山-0002 愛宕	998	1	19	63	1885	10	1885							
01 富山-0002 愛宕	999	1	19	81	1886	21	1886							
01 富山-0002 愛宕	1000	1	19	62	1887	21	1887							
01 富山-0002 愛宕	1001	1	19	40	1888	21	1888							
01 富山-0002 愛宕	1002	1	19	78	1889	21	1889							

(2) 対象地域と使用データ

対象地域として、コンパクトシティ施策を推進している富山県富山市とした。国勢調査の小地域統計では、82ゾーンに分かれているため、初期世帯マイクロデータの推計でもこのゾーン単位を用いることにした。

また、使用した主なオープンデータの例を表-3に示す。今回提案する初期マイクロデータ生成においては、一部のパラメータを除けば、ほぼオープンデータのみで実行できる。

(3) 計算結果例

初期マイクロデータ生成の例として世帯テーブルを表-4に示す。今回のケーススタディでは、170498世帯分の初期マイクロデータを生成するのに、約17分で実行できており、類例よりも圧倒的な計算時間の時間短縮に成功している。

また、表-5に初期マイクロデータ生成収束までのループ回数や誤差(夫婦、母子の年齢差)等の挙動例を示す。ループ回数を経るに従って着実に差の縮小がなされており、これは今回導入した「世帯間で入れ替えを行う個人の選択ルール」(3.2b.)が有効に働いているという証左であると考えられる。

表-5 初期マイクロデータ生成収束までの挙動例

経過時間	ループ回数	夫婦の年齢差	母子の年齢差	ペナルティ	合計
0時間00分	0	874	10,076	34,245	45,195
0時間05分	240,066	330	2,508	0	2,838
0時間10分	438,735	190	240	0	430
0時間17分(収束)	749,502	0	0	0	0

5. まとめ

マイクロシミュレーションを用いた将来予測手法の検討の一環として、初期マイクロデータ生成にかかる高速化について、そのアプローチから実証までを概説した。今回は、特にSA法を用いた先行研究の枠組みを継承しつつ、「SA法に投入する初期解の構築アルゴリズムの導入」、「SA法における有効入れ替え発生率の改善」の高速化への寄与が高い2点を主軸にした改良について述べたが、このほかにも誤差項の改良等についても行っており、総じて現状で考え得る改良は実装できたものと考えている。

今後の課題としては、ケーススタディの蓄積によるパラメータのチューニングや、生成結果の精度検証等があげられる。これらについては、機会を改めたい。

参考文献

- 1) 鈴木温, 杉木直, 宮本和明: 空間的マイクロシミュレーションを用いた都市内人口分布の将来予測—40万人規模の富山市を対象として—, 都市計画論文集, Vol.51 No.3, 2016
- 2) 水流風馬, 平野巧真, 鈴木温: オープンデータを用いた初期マイクロデータ生成方法に関する研究, 土木計画学研究・講演集, Vol.60, CD-ROM, 2019
- 3) Deming, E. and Stephan, F.F.: On a Least Squares Adjustment of a Sampled Frequency Table When the Expected Marginal Totals are Known, The Annals of Mathematical Statistics, Vol. 11, No. 4, pp. 427-444, 1940.
- 4) Moeckel, R., Spiekermann, K., and Wegener, M.: Creating a Synthetic Population, Proceedings of CUPUM '03, Sendai, CD-ROM, 2003.
- 5) Barthelemy, J., and P. L. Toint. Synthetic Population Generation Without a Sample. Transportation Science, Vol. 47, No. 2, 2013, pp. 266–279.
- 6) 杉木直, 宮本和明, Varameth VICHENSAN: 土地利用マイクロシミュレーションにおける初期マイクロ世帯データの推定手法, 第 39 回土木計画学研究発表会論文集, CD-ROM, 2009.
- 7) 杉木直, 宮本和明, 大谷紀子, Varameth VICHENSAN: 居住属性を含む初期マイクロ世帯データの推定手法, 第 41 回土木計画学研究発表会論文集, CD-ROM, 2010
- 8) Nao Sugiki, Tomoya Muranaka, Noriko Otani, Kazuki Miyamoto: Agent-based Estimation of Household Microdata with Detailed Attributes for a Real City, Proceedings of the 14th International Conference on Computers in Urban Planning and Urban Management, pp.231-1 - 231-18 2015.
- 9) Lenormand, M., Deffuant, G., 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, 2013.
- 10) Choupani, A.A. and Mamdoohi, A.R.: Population synthesis using iterative proportional fitting (IPF): A review and future research; Transportation Research Procedia, Vol.17, pp.223-233, 2016.
- 11) Farooq, B., Bierlaire, M., and Hurtubia, R.: Gunnar Flötteröd, Simulation based population synthesis, Transportation Research Part B: Methodological, Volume 58, pp.243-263, 2013.
- 12) Harland, K. Heppenstall, A. Smith, D. and Birkin, M.: Creating realistic synthetic populations at varying spatial scales, A comparative critique of population synthesis techniques, Journal of Artificial Societies and Social Simulation, Vol.15, No.1, pp.1-24, 2013.
- 13) Hara, S., Kita, H., Ikeda, K. and Susukita, M.: Configuring agent's attributes with simulated annealing; Agent-Based Approaches in Economic and Social Complex Systems VII, Agent-Based Social Systems, Vol.10, pp.45-59, 2013.
- 14) 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, pp.11-14, 2010.
- 15) 橋井大貴, 村田忠彦: 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, Vol.30, No.6, 2017 年
- 16) 福田純也, 喜多一: 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, Vol.27, No.7, pp.279-289, 2014.

(2020.3. 受付)

SPEEDING UP OF PRODUCING INITIAL MICRO DATA OF HOUSEHOLDS IN HOUSEHOLD BASED ESTIMATION

Tomohiko SAKATA, Atsushi SUZUKI, Nao SIGIKI, Toshiyuki MASAKI and Hiroyuki Den