

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*

福田 純也[†]・喜多 一[‡]

Evaluation of a Method for Deciding Agents' Attributes for an Agent-based Population Estimation Model*

Junya FUKUTA[†] and Hajime KITA[‡]

In agent-based simulation using actual social statistics, attributes of all the agents have to be decided to meet the statistics. However, they can't be decided uniquely because statistics gives only limited number of aggregated values. To this problem, Hara et al. have proposed a method of formulating the problem as a constrained optimization problem, and solving it using simulated annealing. They applied it to estimation of future population. However, they didn't carry out evaluation of computational performance of their model. In this paper, the authors improved implementation of their model to enhance computational performance, discuss convergence of the optimization process and evaluate computational load of their method. Estimation results are compared with an estimation using conventional cohort-component method.

1. はじめに

社会をシミュレーションする手法にはさまざまなものがある [1] が、人や組織をソフトウェアエージェントとして記述し、社会構造をエージェント間の相互作用としてモデル化してボトムアップにシミュレーションを行うエージェントベースドシミュレーション (Agent-Based Simulation, 以下「ABS」) は、他の手法では困難な政策評価などを含めてさまざまな応用が期待されている。

ABS はその解像度から、社会現象の原理的理解を目的に少数のパラメータでシンプルなモデルとして記述する Abstract Model, 対象をより詳細に記述する Middle Range Model, さらに解像度が高く、特定の状況に限定された社会システムを可能な限り忠実に表現する Facsimile Model に分類される [2] が、現実の問題を定量的に取り扱い、具体的な政策の評価などを行うためには Facsimile Model の構築や評価のための手法の検討

が重要である。

Facsimile Model 型の ABS を実行するためには、個々のエージェントの属性を、関心ある事象に関して現実を一定程度再現できるように決定しなければならない。しかしながら、社会調査の個票に基づいてエージェントを設定できるような場合を除けば、多くの実社会のデータは統計として集計されたものであり、また個々の統計は個別に調査対象が設定されているため、統計相互に必ずしも整合しているわけではない。ABS の実施にあたっては、大規模なエージェントの属性をこのような統計への適合を考慮して決定しなければならない。

これに対し、池田ら [3]、Hara ら [4] は人口推計モデルを例に、エージェントの属性決定問題を拘束条件付最適化として定式化し、その解のランダムサンプリングをシミュレーテッドアニーリング法 (Simulated Annealing, 以下「SA」[5]) により求める手法を提案している。

Hara らは上述した手法を用いた人口推計モデルを提案しているが、SA 法により拘束条件充足問題がどの程度良好に解かれているのか、また、計算量の多い同手法が実際のどの程度の計算時間を要し、どの程度の規模までの人口推計が現実的に扱えるのかは検討されておらず、また将来人口の推計結果についても他の人口推計モデルとの比較などを通じた妥当性の検証は行われていない。そこで本研究では Hara らの人口推計モデルについて、シミュレーションプログラムとしての実装を見直して計

* 原稿受付 2013 年 9 月 2 日

[†] 京都大学 大学院 情報学研究科 Graduate School of Informatics, Kyoto University; Yoshida-Honmachi, Sakyo ward, Kyoto city, Kyoto 606-8501, JAPAN

[‡] 京都大学 国際高等教育院 Institute for Liberal Arts and Sciences, Kyoto University; Yoshida-Honmachi, Sakyo ward, Kyoto city, Kyoto 606-8501, JAPAN

Key Words: agent-based simulation, simulated annealing, constrained optimization, population estimation.

算性能を大幅に向上させたいとこれら点を評価するとともに、人口推計の代表的な手法であるコホート要因法を用いた国立社会保障・人口問題研究所の人口推計との比較評価を行う。

2. 人口推計手法

人口推計の代表的な手法としては性別および年齢別の人の集団（コホート）の人口という集計量についてのダイナミクスとして記述するコホート要因法と、個人を明示的に扱う個人ベースの手法がある。

2.1 コホート要因法を用いた将来人口推計

コホート要因法は人口学の分野での標準的な人口推計の手法である。同じ期間に生まれた男女の集団、つまり性別および年齢別の人の集団を「コホート」とよび、コホート要因法ではコホートについての集計量である人口に対して、「加齢」、「出生」、「死亡」、「人口移動」の計算を行うことで将来人口を推計するマクロレベルの推計手法である。

国立社会保障・人口問題研究所が平成 24 年 1 月に行った日本の将来人口の推計 [6] でも、同手法が用いられている。この推計では、加齢に伴って生ずる年々の人口変動要因（「出生」、「死亡」および「国際人口移動」）に仮定を設け、コホートについて、これらを計算することで日本全体の将来の人口規模、ならびに年齢構成などの人口構造の推移を推計している。各人口変動要因の仮定は、各要因に関する統計指標の実績値に基づいて、人口統計学的な投影を実施することにより設定されているが、将来の出生、死亡などの推移は不確実であるため、それぞれ 3 段階、合わせて 9 種類の仮定に基づく複数の推計を行い、将来の人口推移について一定幅の見通しを与えるものとしている。

なお、コホート要因法は、集計量に対して人口変動を記述しているため、個人や世帯単位での具体的な政策の効果を織り込んだ人口の推計は行いにくいという欠点がある。

2.2 個人ベースの人口推計手法

個人ベースの人口推計手法としては、個人についての出生や死亡などを外生的な確率で適用して人口推計を行うマイクロシミュレーション (Microsimulation, 以下「MS」) や、個人を一定の自律性をもってふるまうエージェントとして記述して人口推計を行うエージェントベースのシミュレーション (ABS) などがある。

2.2.1 INAHSIM

世帯情報解析モデルである INAHSIM は MS を用いたモデルである。稲垣の研究 [7] では、その INAHSIM が用いられており、平成 13 年の「国民生活基礎調査」の個票データから日本の総人口の 1,000 分の 1 の仮想社会を構成し、100 年間のシミュレーションを行っている。個票データには、個々人の「性別」、「年齢」、「家族構成」、

「就業状態」、「健康状態」、「所得」、「年金額」などが含まれており、これらの情報から人口の初期値を構成している。シミュレーションのサイクルは 1 年であり、「出生」、「死亡」、「結婚」、「離婚」、「就業状態の遷移」などの合計 9 種類のイベントをそれぞれの確率に従って発生させている。

2.2.2 Hara らのモデル

Hara らの人口推計モデル [4] は MS と ABS の中間的なモデルとして構成されており、内部ではエージェントとして実装されているが、人口変動の要因となる「出生」、「死亡」、「婚姻」、「離婚」、「世帯分離・統合」の発生ではエージェントの自律的な動作は組み込まれていない。

同モデルでは個人だけでなく、ライフスタイルの評価などへの応用を視野に入れて、個人の集合として世帯を明示的に考慮している。また、地域別のモデル化も視野には入れているが、現在は日本全国を一つの地域として扱っている。シミュレーションのサイクルは 1 年であり、初期世帯の設定後は毎年、出生、死亡、婚姻、離婚、世帯分離・統合の 5 種類のイベントを、設定されたパラメータに適合するようにランダムに発生させている。

Hara らの推計では、エージェント数を日本の人口の約 1 万分の 1 程度にとり実施されている。推計のためのデータとしては統計の個票は用いず、公表されている統計値が用いられている。統計値に適合する初期世帯の設定や上述した 5 種類のイベントの発生については、後述の拘束条件充足問題としての定式化と SA によるランダムサンプリングが用いられている。

3. エージェント属性決定手法とアンサンブルシミュレーション

本論文では、Hara らの提案した人口推計モデルの実装改善と評価を目的とするが、本章では同モデルの基本となっているエージェントの属性決定手法とアンサンブルシミュレーションの考え方について紹介する。

3.1 エージェント属性決定手法

ABS を用いた Facsimile Model 型の社会シミュレーションでは、対象となる社会現象をより忠実にモデル化するために、エージェントの属性を既存のデータに適合させる必要がある。先の INAHSIM のように統計調査の個票を用いる場合や、山本らの研究 [8] のように実態の把握しやすい小規模な地域を対象にする場合を除けば、適合させる対象はおもに統計の集計量となる。しかしながら集計量からエージェントの属性を決定する場合、エージェントの属性の自由度の方が高いため、これらを一意には決定できない。そこで池田ら、Hara らは人口推計モデルを例として、仮想社会の住民であるエージェントの属性決定問題を拘束条件付最適化問題として定式化し、SA を用いて解を求めている [3,4]。

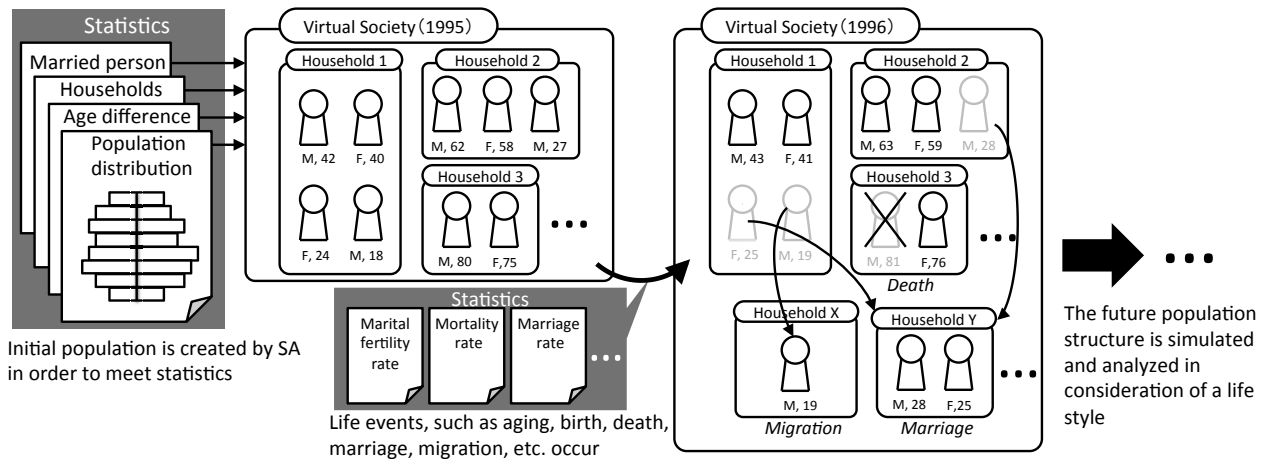


Fig. 1 Simulation model

3.1.1 拘束条件付最適化

池田らの提案した拘束条件付最適化は、対象地域についての統計から個々のエージェントの属性を逆推定する手法である。各種の統計データを、エージェントの属性を集計した値についての拘束条件とし、エージェントの属性決定問題について拘束条件をペナルティ関数とした最適化問題として定式化する。そして、最適解をランダムにサンプリングして、拘束条件を満たす範囲で発生頻度の高い状態を属性値として採用しようというものである。

3.1.2 SA を用いたランダムサンプリング

SA は、高温にした金属を徐冷することで低エネルギー状態を作り出す「焼きなまし」を模した最適化手法である [5]。SA は通常、最適化手法としてとらえられているが、SA の基本となる Metropolis の方法による確率的状態遷移では、状態 x についてのエネルギー関数を $E(x)$ 、エネルギーの増加する状態遷移確率を決める温度パラメータを T としたときに、状態 x についての定常分布 $p(x)$ は $\exp(-E(x)/T)$ に比例することが知られており、十分に低い温度ではエネルギー最小の状態が等確率で得られることを意味する。このことから、拘束条件をペナルティ関数として扱うことにより、SA を拘束条件下でのランダムサンプリング手法として用いることができる。なお、データに適合する ABS のパラメータを推定する手法としては進化計算などを利用する逆シミュレーション [9] が知られているが、Hara らの研究では上述の特性に加え、大規模な問題でも取り扱いの比較的容易な手法として SA が採用されている。

3.2 アンサンブルシミュレーション

上述の属性決定手法を用いてエージェントの属性を決定しても、拘束条件と比べてエージェントの属性値の自由度は圧倒的に高く、属性値は一意に決まらない。そこで、「ランダムなサンプリングを行い、属性値を決定する」という工程を複数回繰り返し、エージェントが構成する仮想社会のアンサンブルを得る。そして、各仮想社

会でシミュレーションを行い、これにより得られた将来の各仮想社会の分布から、将来予測値を考察する。このような手法をアンサンブルシミュレーションとよぶ [4]。

4. 人口推計モデル

本論文では、Hara らによって構築された人口推計モデル（以下、「原モデル」）を用いる。以下では、評価実験を行うにあたっての前提知識として、原モデルの概要を紹介する。

4.1 人口推計モデルの概要

Hara らの人口推計モデルではシミュレーションのサイクルを 1 年としており、シミュレーション開始時にはまず、初期エージェントが生成される。その後、サイクルごとに加齢、婚姻、離婚、出生、死亡、世帯分離・統合の順にライフイベントを所与のパラメータに適合するように確率的に発生させることで人口の推計を行う。Fig. 1 にモデルの概略図を示す。

このうち、初期エージェントの生成に加えて、個人に対して外生的な確率で適用する出生と死亡を除く、婚姻、離婚、世帯分離・統合の各ライフイベントも拘束条件付最適化によりモデル化している。なお、初期エージェント生成においてはエージェントの属性値そのものを決定変数としている一方で、各ライフイベントにおいてはイベントの対象となるエージェントの組合せが決定変数である。

4.2 初期エージェント生成

拘束条件付最適化を用いて統計データを充足するように個票を生成することで初期エージェントを生成する。

4.2.1 拘束条件

拘束条件として、国立社会保障・人口問題研究所の「人口統計資料集」で公表されている以下の八つの統計データを用いた。

- (1) 年齢別人口分布
- (2) ある年齢・性別の人が単身世帯にいる割合
- (3) ある年齢・性別の人が夫婦のみ世帯にいる割合

Table 1 The rate of married male by age group (1995)

Condition X	Condition Y	Rate
Male, 15-19	Married	0.260
Male, 20-24	Married	6.483
Male, 25-29	Married	31.617
⋮	⋮	⋮
Male, 75-79	Married	83.218
Male, 80-84	Married	74.843
Male, 85-	Married	56.433

- (4) 父と子の年齢差の分布
- (5) 母と子の年齢差の分布
- (6) 夫婦の年齢差の分布
- (7) 女性の年齢階級別有配偶率
- (8) 男性の年齢階級別有配偶率 (Table 1 参照)

これらの統計データは、人口構成や世帯構成、婚姻関係といった初期エージェントを生成するうえで重要な要素を拘束条件として取り入れるために選択した。とくに世帯構成に関しては、単身世帯および夫婦のみ世帯の数が全体に占める割合が高いため、それらに関わる統計データを選択している。

例として、統計データ (8) を Table 1 に示す。同表の条件 X はある年齢階級に該当する男性であるかどうか、条件 Y はその年齢階級の男性エージェントが有配偶の状態にあるかどうかを意味している。上記の八つの統計データは、同様に「条件 X を満たす人 (組) の数」に対する「条件 X, Y を満たす人 (組) の数」の割合 (Rate) として表現できる。このことを利用して後述のように一般性をもった評価関数の設計が行える。

4.2.2 評価関数の設計

初期エージェント生成では、まず「人口統計資料集」で公表されている「世帯人員別一般世帯数」および「家族類型別一般世帯数」を利用する。総世帯数を所与として、これらの統計に合うように、総エージェント数 (N とする) および世帯ごとの人数と役割 (夫, 妻, 子どもなど) を伴った世帯構成を決める。

つぎに、あらかじめ世帯の構成が割り付けられているエージェントの年齢, 性別のセットを決定変数と考え、解空間を I として $\vec{a} = \{(age_i, gender_i)\}_i, i = 1, \dots, N, \in I$ とする。そして、八つの統計データそれぞれの \vec{a} に対する誤差を表す関数を評価関数として $f_1^I(\vec{a}), f_2^I(\vec{a}), \dots, f_8^I(\vec{a})$ と定義する。具体的には次のように導入している。

- 評価したいパラメータセット \vec{a} と、統計 l が与えられ、統計データ中の条件 X_l についての分類階級数を G_l とする¹。
- 統計データ中の階級 $j = 1, \dots, G_l$ について、条件 X_l を X_{lj} , 条件 Y_l を Y_{lj} と表すものとし、統計値とし

て示されている割合を r_{lj} とする。また、 \vec{a} について、「条件 X_{lj} を満たす人 (組) の数」を m_{lj} とし、「条件 X_{lj}, Y_{lj} を満たす人 (組) の数」を c_{lj} とする。統計上の割合が r_{lj} であるため、 c_{lj} の理想の数は $m_{lj}r_{lj}$ となる。

- これらにより、誤差関数を、

$$f_l^I(\vec{a}) = \frac{4}{G_l} \sum_{j=1}^{G_l} (c_{lj} - m_{lj}r_{lj})^2 \quad (1)$$

と定義することにする。

この指標では、「実際の人 (組) の数 c_{lj} 」と「理想の人 (組) の数 $m_{lj}r_{lj}$ 」との差の二乗和を最小化しようとする。 c_{lj} が整数であり端数が生じることから各階級で 0.5 程度の差は問題なく、仮にすべての階級で 0.5 の差があった場合に $f_l^I(\vec{a}) = 1.0$ と正規化されるように係数 $4/G_l$ を掛ける形で (1) 式は定義されている。すなわち、この指標が 1 程度であれば従わせるべき統計データとほぼ正確に一致しているといえる。

つぎに、複数の評価指標を考慮するため、 $f_1^I(\vec{a}), f_2^I(\vec{a}), \dots, f_8^I(\vec{a})$ を単一の目的関数として定式化することを考える。各統計データの重みをそれぞれ $\omega_1^I, \omega_2^I, \dots, \omega_8^I$ とし、目的関数を

$$\min_{\vec{a} \in A} \sum_{l=1}^8 \omega_l^I f_l^I(\vec{a}) \quad (2)$$

と定式化し、この最小化問題を SA で解くことで初期エージェントの属性を決定する。

各世帯を構成するエージェントの年齢, 性別の決定後、後述の世帯分離・統合イベントを発生させるため、異なる世帯間で親世帯-子世帯の親子関係が設定される。ここでは、「人口統計資料集」の「女性の年齢別出生数」および「女性の年齢別死亡率」を利用し、ある世帯主のエージェントにおいて、母親の年齢と生存を決定し、生存している場合に「当該年齢女性エージェントの選択」を確率的に行うことで、世帯間の親子関係が設定される。

4.3 婚姻イベント

婚姻イベントによって独身の男女 (男性 18 歳以上, 女性 16 歳以上) が婚姻状態となり、新たな世帯を構成する。その際、拘束条件付最適化を用いて婚姻に関わる統計データを充足するような組合せを決定することで、エージェント同士を婚姻させる。なお、婚姻に伴い必ず新世帯を構成するとしているが、現実にはさまざまな世帯構成が考えられる。この点については後述の世帯分離・統合イベントで調整される。

4.3.1 拘束条件

拘束条件として、厚生労働省の「人口動態統計特殊報告」および「人口動態調査」で公表されている、以下の三つの統計データを用いた。なお、拘束条件の設定に際しては初婚と再婚の区別をせずに用いた。

¹Table 1 のデータは、実際には 15 階級である。

Table 2 Age-specific marriage rate of females (1995)

Condition X	Condition Y	Rate ²
Female, 16	Couples	1.0
Female, 17	Couples	2.6
\vdots	\vdots	\vdots
Female, 25	Couples	96.3
Female, 26	Couples	95.0
\vdots	\vdots	\vdots
Female, 74	Couples	0.1
Female, 75-	Couples	0.5

² Number of marriages per 1,000 females

- (1) 初婚夫妻の年齢差別婚姻率（男性年齢－女性年齢）
- (2) 男性の年齢別婚姻率（人口千対）
- (3) 女性の年齢別婚姻率（人口千対）（Table 2 参照）

例として、統計データ (3) を Table 2 に示す。条件 Y はその時点で婚姻の組合せとして選ばれているか、ということの意味している。

4.3.2 評価関数の設計

婚姻件数 N_m は Table 2 の統計データと年齢別の女性エージェント数から、

$$N_m = \frac{1}{1000} \sum_{j=1}^{60} m_{3j} r_{3j} \quad (3)$$

と計算できる。当該統計データの階級数は 60 である。

決定変数は男女のエージェントの組である婚姻関係であり、解空間を M として $\vec{m} = \{(M_i, F_i)\}_i, i = 1, \dots, N_m, \in M$ とする。ここで (M_i, F_i) は婚姻関係にある男性エージェント M_i と女性エージェント F_i である。そして、三つの統計データそれぞれの \vec{m} に対する誤差を表す関数を評価関数として $f_1^M(\vec{m}), f_2^M(\vec{m}), f_3^M(\vec{m})$ とする。

- 評価したいパラメータセット \vec{m} と、統計データ l が与えられ、統計データ中の条件 X_l についての分類階級数を G_l とする。
- 統計データ中の階級 $j = 1, \dots, G_l$ について、条件 X_l を X_{lj} 、条件 Y_l を Y_{lj} 、割合を r_{lj} とする。また、 \vec{m} について、「条件 X_{lj} を満たす人（組）の数」を m_{lj} とし、「条件 X_{lj}, Y_{lj} を満たす人（組）の数」を c_{lj} とする。統計上の割合が r_{lj} であるため、理想の数は $m_{lj} r_{lj}$ となる。
- これらにより、誤差関数を、

$$f_l^M(\vec{m}) = \frac{4}{G_l} \sum_{j=1}^{G_l} (c_{lj} - m_{lj} r_{lj})^2 \quad (4)$$

と定義する。

そして各統計データに対する誤差関数を重み ω_l^M で荷重和して目的関数を、

Table 3 The divorce rate of males by age group (1995)

Condition X	Condition Y	Rate ³
Male, -19	Divorce	40.67
Male, 20-24	Divorce	34.76
Male, 25-29	Divorce	18.19
\vdots	\vdots	\vdots
Male, 45-49	Divorce	3.89
Male, 50-54	Divorce	2.62
Male, 55-	Divorce	1.52

³ Number of divorces per 1,000 males

$$\min_{\vec{m} \in M} \sum_{l=1}^3 \omega_l^M f_l^M(\vec{m}) \quad (5)$$

と定式化し、SA で解くことでエージェントの婚姻の組合せを決定する。

4.4 離婚イベント

夫婦が存在する世帯に対して離婚が発生する。夫はそのままその世帯に残り、妻は新世帯を構成するものとする。また、どの世帯の夫婦を離婚させるか決めるだけでなく、子どもの親権をどちらに与えるかも決定変数に加えている。離婚時の世帯構成は婚姻時と同様、単純なものであるが世帯分離・統合イベントで調整される。

4.4.1 拘束条件

拘束条件として、厚生労働省の「人口動態統計特殊報告」から、以下の三つの統計データを用いる。

- (1) 夫の年齢階級別離婚率（人口千対）（Table 3 参照）
- (2) 妻の年齢階級別離婚率（人口千対）
- (3) 夫婦の親権を行う者別離婚率

例として、統計データ (1) を Table 3 に示す。条件 Y は、その時点で離婚する夫婦として選ばれているか、ということの意味している。

4.4.2 評価関数の設計

1 年間の離婚件数 N_d は厚生労働省の「人口動態統計」で公表されている離婚率（人口千対）を用いて、 $N_d = \text{全エージェント数} \times \text{離婚率} / 1000$ と計算される。決定変数は、 N_d 組の夫婦および親権の場合分けのセットであり、解空間を D と表すものとし、 $\vec{d} = \{(H_i, W_i, C_i)\}_i, i = 1, \dots, N_d, \in D$ と書く。 H_i は夫の役割を担うエージェント、 W_i は妻の役割を担うエージェント、 C_i は以下に示す親権の場合分けである。

- (1) 親権を行わなければならない子どもがいない
- (2) 妻がすべての子どもの親権を行う
- (3) 夫妻が分けあって親権を行う
- (4) 夫がすべての子どもの親権を行う

そして、婚姻イベントと同様に 3 種の統計データそれぞれの \vec{d} に対する誤差を表す関数 $f_1^D(\vec{d}), f_2^D(\vec{d}), f_3^D(\vec{d})$ を定義し、重み ω_l^D で荷重和して目的関数を、

Table 4 Marital fertility rate for the second child by age group (1995)

Age group	Rate ⁴	Age group	Rate
15-19	49.7	35-39	10.8
20-24	78.9	40-44	0.9
25-29	84.0	45-49	0.0
30-34	54.3	—	—

⁴ Number of births per 1,000 married females

$$\min_{\vec{d} \in D} \sum_{l=1}^3 \omega_l^D f_l^D(\vec{d}) \quad (6)$$

と定式化し, SA で解くことで離婚する夫婦を決定する.

4.5 出生イベント

夫婦が存在する世帯において, 女性エージェントの属性値に応じた確率で出生し, 0 歳のエージェントをその世帯に追加する. ただし, 一人の女性に何度も出生イベントが発生しないように, 世帯に所属する子どもの数に応じて, 出生確率を変化させる. 利用する統計データは, 「人口動態統計特殊報告」の第 1 子～第 3 子の「年齢階級別有配偶出生率 (有配偶女性人口千対)」である. 例として, 第 2 子のデータを Table 4 に示す. 一つ目の階級は, 1995 年には 15-19 歳の有配偶女性 1,000 人あたり 49.7 人が第 2 子を出生したことを意味する. ここで注意すべきは, この「出生率」の統計データをエージェントの「出生確率」として利用できるように, 換算する必要がある点である. たとえば, 15-19 歳の有配偶女性 1,000 人のうちの 100 人が, 第 1 子だけがいる世帯に所属していたとすると, このうち 49.7 人が第 2 子を出生しなければならないことから, その 100 人が第 2 子を出生する確率は 49.7% としなければならない. ここで, 子どもの数および年齢階級に関する添字をそれぞれ i, j とし, 年齢階級別有配偶女性人口を N_j , 子どもの数および年齢階級別の有配偶女性人口を $N_{i,j}$ とおく. そして, 統計データで示される子どもの数および年齢階級別有配偶女性の出生率 $r_{i,j}$ から, 同出生確率 $p_{i,j}$ を計算すると,

$$p_{i,j} = \frac{r_{i,j} N_j}{N_{i,j}} \quad (7)$$

となり, 換算が可能となる.

4.6 死亡イベント

エージェントの年齢・性別に応じた確率に従ってエージェントが死亡する. エージェントが従う死亡確率は, 国勢調査による「5 歳階級別男女人口」および厚生労働省の「人口動態調査」による「5 歳階級別男女死亡数」の二つの統計データから算出した, 5 歳階級別男女死亡率である. この死亡率のデータを Table 5 に示す.

Table 5 Mortality rate by sex and age group (1995)

Age group	Male ⁵	Female ⁵
0-4	1.3	1.1
5-9	0.2	0.2
⋮	⋮	⋮
65-69	19.9	8.6
70-74	31.4	15.1
75-79	54.3	28.0
80-84	94.5	54.1
85-89	156.0	100.4
90-94	249.0	176.8
95-99	359.9	266.2
100-	530.1	430.5

⁵ Number of deaths per 1,000 people

Table 6 Households by size of household (1995)

Condition X	Condition Y	Rate
Household	1 person	0.2681
Household	2	0.2234
Household	3	0.1778
Household	4	0.1683
Household	5	0.0924
Household	6	0.0700
Household	7-	0

4.7 世帯分離・統合イベント

1 年間で世帯構成の分布に大きな変化はないと仮定し, 介護や独立によって世帯構成が変化することと先の婚姻, 離婚で生じた世帯構成の偏りを調整するイベントである. 世帯分離では独立などを理由に若い世代が新規世帯を生成するものとし, 世帯統合では介護などを理由に子世帯が親世帯に加わるものとする. 本イベントでも拘束条件付最適化を用いて最適な世帯分離および統合を行う.

4.7.1 拘束条件

以下の二つの世帯構成の分布に関して 1 年目は統計データとの一致を, 2 年目以降は前年との一致を拘束条件とした.

- (1) 家族類型別世帯数の割合
- (2) 世帯人員別世帯数の割合 (Table 6 参照)

例として, 1 年目のサイクル開始前における統計データ (2) を Table 6 に示す. 条件 Y は世帯人員数を意味する. 7 人以上の世帯の数が 0 となっているのは, 初期エージェント生成の際に 7 人以上の世帯を考慮していないためであるが, シミュレーションが進むことで世帯構成が変化し, 7 人以上の世帯が新たに構成されることはあり得る.

4.7.2 評価関数の設計

決定変数は世帯であり、すべての世帯が評価対象である。仮想社会の全世帯数を N_h とし世帯を H_i ($i=1, \dots, N_h$) とする。そして解空間を X と表すものとし、決定変数を $\vec{h} = \{H_i\}_i, i=1, \dots, N_h, \in X$ と書くこととする。婚姻イベントなどと同様に、二つの統計データそれぞれについて \vec{h} の誤差を表す関数 $f_1^H(\vec{h}), f_2^H(\vec{h})$ を定義し、それらを重み ω_1^H, ω_2^H で荷重和して目的関数を、

$$\min_{\vec{h} \in X} \sum_{l=1}^2 \omega_l^H f_l^H(\vec{h}) \quad (8)$$

と定式化する。この最小化問題を SA で解くことで、世帯分離・統合が発生する世帯群を決定する。

5. 評価実験

原の研究 [10] はおもに 4. で述べたモデルの提案であり、そのプログラムとしての実装と性能評価については十分に検討されておらず、計算時間に関して「約 12,000 エージェントのモデルで 20 年間のシミュレーションに約 1 時間半を要する」[10] と言及しているのみである。シミュレーションの規模は表現したい人口動態の内容に依存する。個人ベースのモデルではモデルの規模が小さすぎると表現したい内容、すなわち年齢や性別別の人口構成の精度が下がってしまう。現実と同規模であることを最も詳細な場合としつつ、シミュレーションに要するコストを考慮すると表現すべき内容が適切にシミュレーションされる範囲で規模を縮減することが望まれる。個人ベースの人口推計の先行研究である INAHSIM ではわが国の 1/1000 のモデルを用いていることから、全国レベルのモデルとしてはこれを参照しつつ地域を考慮したシミュレーションへの拡張を視野に入れたうえで Hara らのモデルがどの程度の規模の推計まで現実的に扱えるかを検証することは重要である。本研究では、Hara らのモデルによるシミュレーションプログラムの実装を見直してその性能向上を図ったうえで計算量の多い同モデルの性能評価を行う。

以下の実験では、日本全体を対象に将来人口推計を行っている。推計期間は 1995 年から 2045 年までの 50 年間としている。本実験では、1995 年をシミュレーション開始年としているため、利用する統計データも基本的には 1995 年のものとしている。ただし、1995 年のデータが存在しない統計に関しては、近い年のデータを利用した。

本実験の目的は以下の 3 点である。

- (1) SA を用いた最適化プロセスの収束状況の検証 (5.2)
- (2) 実行する計算機上でのメモリーの使用量と所要計算時間 (5.3)
- (3) 人口推計の結果とコホート要因法による推計の比較 (5.4)

次節では、本実験で用いた仮定値について説明する。

5.1 仮定値

50 年間のシミュレーションを行うにあたり、シミュレーションサイクルを構成する各イベントのためのパラメータを設定する必要がある。原モデルでは設定すべきパラメータはシミュレーションにより変化する人口分布そのものにはあまり依存しないものが選ばれている。婚姻率、死亡率などについては将来の変化も想定したシナリオの導入が考えられるが、本実験ではエージェントの属性決定手法の評価が主な目的であるため、シミュレーション開始時の統計データから得られるパラメータは将来も一定であると仮定した。

5.2 最適化状況の分析

ここでは代表として、初期エージェント生成および 1 年目 (1995 年) の婚姻イベントの最適化に関して、世帯数 5,000 の場合について分析する。なお、(2) 式および (5) 式で示した各統計データの重み $\omega_l^k, k=I, M$ はすべて 1 を標準として実施した。

複数の統計データに関する評価指標である誤差関数値について、SA の探索につれてその値がどのように推移・収束するかを観察した。SA のアニーリングスケジュールとしては事前の予備実験での検討のうへ、初期エージェント生成では、初期温度は 1000、探索回数は $1900 \times$ 全人口 ($\approx 1.33 \times 10^7$ 回) で最終温度 0.00171 まで、婚姻イベントでは、初期温度は 100、探索回数は $500 \times$ 男女の独身人口 ($\approx 2.08 \times 10^6$ 回) で最終温度 0.00298 まで一定比率で温度を低下させるものとした。

初期エージェント生成と婚姻イベントでの結果を、それぞれ Fig. 2, Fig. 3 に示す。どちらも縦軸を対数表記としている。SA の探索終了時点では双方のどの統計データの誤差関数値も 1 前後に収束しており、誤差関数の構成法より満足できる値に低下しているといえる。特定の誤差関数のみ急激な低下をしている様子もみられず、各評価関数が適切に設計されているといえる。すなわち、本実験で設定した SA の探索回数であれば、拘束条件付最適化を用いてエージェントの属性決定に関する適切な解が得られているものと考えられる。

評価指標である誤差関数値の収束値について、初期エージェント生成および婚姻イベントの結果をそれぞれ Table 7, Table 8 に示す。婚姻イベントでは、三つの統計データとも誤差関数値がいずれも 1 以下に収束しており、婚姻に関するエージェントの属性を統計データとほぼ一致させることができたといえる。他方、初期エージェント生成では誤差関数値が 1 以上となる統計データが三つ存在している。これは、拘束条件としている統計データが八つと多く、さらには人口構成や世帯構成、婚姻関係といった多様な統計データを採用しているため、すべての統計データについてほぼ正確に一致させることが困難となっていると考えられる。

そこで、(2) 式で示した評価指標の重み ω_l^I に着目する。

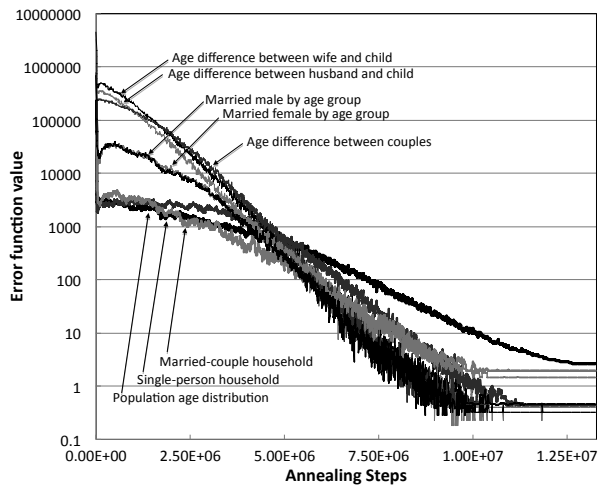


Fig. 2 Change of error function values in creation of initial population (5,000 households)

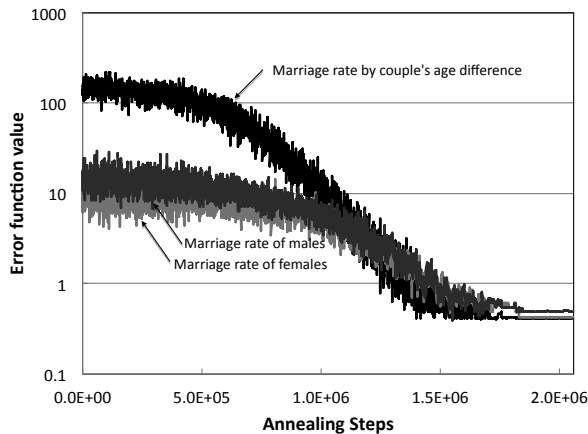


Fig. 3 Change of error function values in marriage event generation (5,000 households, 1st year)

「年齢別人口分布 (Population age distribution)」の誤差関数値が2.5871で最大となっているが、この統計データは人口構成という重要な要素に関わることから、当該評価指標の重みを $\omega_1^I = 5.0$ とすることで他の統計データよりも重要視するよう変更する。その結果、年齢別人口分布の誤差関数値は0.9126となり、1以下まで低下した。他の統計データについての評価値はおおむね増加したが、全体として誤差が均衡したといえる (Table 7 参照)。ここでは当該評価指標の重みを増加させることで誤差を低下させた。本論文では、この方法が最も直感的で調整が容易であると判断したため採用している。

本論文で用いた拘束条件付最適化という手法は、拘束条件としているデータの数や種類を増加させるほど、より現実的なエージェント構成が可能になるが、他方最適化がより難しくなるため、誤差が増加しやすい。したがって、両者のバランスを取って拘束条件を設定する必要がある。

Table 7 The convergence values of error function in creation of initial population (5,000 households)

	$\omega_1^I = 1.0$	$\omega_1^I = 5.0$
Population age distribution	2.5871	0.9126
Single-person households	0.4366	0.5269
Married-couple household	1.9366	1.6776
Age difference between husband and child	0.4047	0.7029
Age difference between wife and child	0.3227	0.8227
Age difference between couples	0.4485	0.4485
Married female by age group	1.4559	1.9444
Married male by age group	0.4518	0.6874

Table 8 The convergence values of error function in marriage event generation (5,000 households, 1st year)

Error Function	Value
Marriage rate by couple's age difference	0.4125
Marriage rate of males	0.4223
Marriage rate of females	0.4896

5.3 計算量の評価

ランダムサンプリングの手法として用いているSAでは、最適解を得るまでに膨大な計算量を必要とする。そこで本節では、空間計算量と時間計算量の二つの計算量から、SAを用いたエージェント属性決定手法が現実的に有用であるかの評価を行う。なお、シミュレータはJavaで実装されており、実行環境にはLinux上で稼働するJDK 1.6.0を用いた。使用した計算機の仕様は以下の通りである。

- OS : Scientific Linux 6.2
- メモリー : 12GB (4GB × 3)
- CPU : Intel Core i7 3.07GHz (Quad Core)

空間計算量の評価としてはマシンのメモリー消費量の最大瞬間値を、時間計算量としてはCPU時間を対象とした。仮想社会の初期世帯数を5,000~500,000 (初期エージェント数に換算すると、それぞれ13,970~1,401,786) とし、50年間のシミュレーションを通して各世帯数における計算量を観察した。実行状況の観察には、JMX (Java Management Extensions) 仕様に準拠した監視ツールであるJConsoleを用いた。

結果をTable 9およびFigs. 4, 5に示す。エージェントベースの人口推計モデルでは、住民であるエージェント単位でモデル化を行うため、世帯数やエージェント数が多いほどそのためのオブジェクトも多数保持することからメモリー消費量も多くなる。Fig. 4より、50,000世

Table 9 Computational complexity according to the model scale

Households	Agents	Memory usage [MB]	CPU time [min]
5,000	13,970	374	4.4
10,000	27,988	530	8.7
20,000	56,014	736	17.5
30,000	84,054	868	27.3
40,000	112,090	1,059	38.0
50,000	140,124	1,010	48.7
100,000	280,306	1,092	110.9
200,000	560,676	1,308	267.1
300,000	841,058	1,425	522.6
400,000	1,121,426	1,568	799.2
500,000	1,401,786	1,711	1,135.1

帯以上でメモリー使用量はおおむね直線的に増加する。500,000 世帯でのメモリー使用量は 1,711 MB となり、この値ならば通常のデスクトップ PC で十分利用可能であるが、これ以上の規模でのシミュレーションの実施には注意を要する。

CPU 時間を見ると、50,000 世帯でおよそ 49 分であったものが、500,000 世帯でおよそ 1,135 分（19 時間弱）となっている。Fig. 5 から世帯数に対して計算時間は線形以上のオーダで増加していることが見てとれる。処理ごとの詳細な CPU 時間を分析したところ、最も支配的であるのが、初期エージェント生成後の親子関係の構築であり、50,000 世帯で 69 秒、500,000 世帯で 7,179 秒（ ≈ 120 分）かかっている。この処理では世帯統合の際に必要な情報を設定しているが、世帯数 N に対して CPU 時間が $O(N^2)$ となる処理をしていることが原因であり、実装上のさらなる工夫が必要である。なお、原の研究では約 12,000 エージェント 20 年間の推計で 90 分かかっていた [10] のに対し、実装の見直しにより約 14,000 エージェント 50 年間の推計で 5 分以下になるまでに大幅に短縮されている。

つぎに、SA を用いたエージェント属性決定手法の計算時間を評価するために、初期エージェント生成および各ライフイベントにおける SA でのランダムサンプリングに要する CPU 時間を計測した。その結果を、Table 10 に示す。各ライフイベントはシミュレーションサイクルの 1 年目を対象として計測した。SA の CPU 時間は、おおむね世帯数 N に対して $O(N)$ となっている。これは探索回数による部分が大いいためであるが、世帯分離・統合イベントに関しては、解の探索部分で $O(N^2)$ の設計となっているため、規模に対する CPU 時間の増大が急であり、今後の改善課題である。

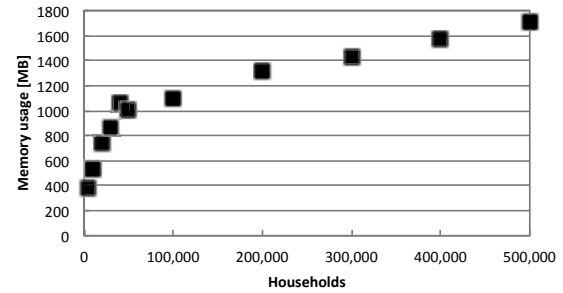


Fig. 4 Memory usage according to the model scale

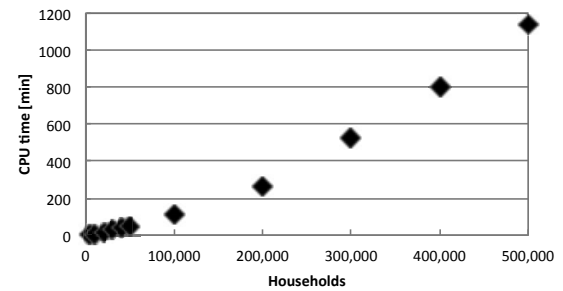


Fig. 5 CPU time according to the model scale

Table 10 CPU time [s] for random sampling with SA

Households	5,000	50,000	500,000
Creation of initial population	29.4	286.7	3,348.8
Marriage	3.5	31.9	353.3
Divorce	1.5	11.6	131.1
Division/integration of households	0.9	6.5	731.0

5.4 将来人口の推計

原モデルは日本全国を対象とした人口推計モデルである。推計においては、世帯数を現実の人口規模に近くするほど精度が得やすいが、他方で計算量が増大する。予備的検討により 5,000 世帯では年齢、性別別の人口分布の推計が個体数の有限性からややばつらくこと、50,000 世帯とすることで一定の精度が得られること、500,000 世帯では推計精度は得られるが計算量がかなり大きくなることを確認した。このことから世帯数を 50,000 とし将来人口の推計を行った。なお、これは個人ベースの人口推計の先行研究である INAHSIM と同程度の規模である。2045 年（50 年後）時点の 5 歳階級別人口分布について、乱数系列が異なる 10 試行のアンサンプルの結果を男女別にそれぞれ Fig. 6, Fig. 7 に示す。比較の対象として、国立社会保障・人口問題研究所（以下、「社人研」）が平成 24 年 1 月に実施した将来人口推計 [6] の結果（出生中位、死亡高位仮定）も示している。ここでいくつかの仮定の中から死亡高位仮定を選択しているのは、本推計の死亡に関する仮定値に極力近づけるためであるが、社人研の死亡中位仮定と死亡高位仮定の推計結果とでは、2060 年時点の人口で 1.6% 程度の差が生じる

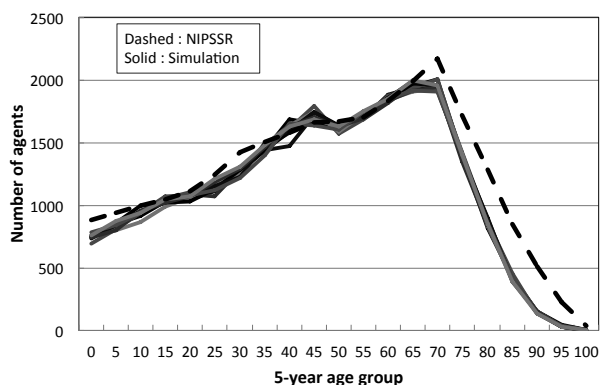


Fig. 6 Population distribution of males by 5-year age group (2045)

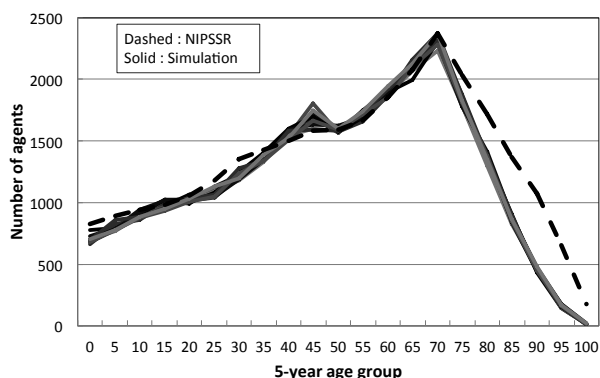


Fig. 7 Population distribution of females by 5-year age group (2045)

のみであり、推計結果にさほど大きな変化はない。なお、社人研の推計結果は、本推計結果と人口変動に関する比較ができるように、社人研の推計開始年である 2010 年の人口を基準としてスケールしている。

アンサンブルシミュレーションで得られた各人口分布を見ると、50 歳未満の人口に比較的大きなばらつきが生じている。この世代は、シミュレーション開始後に新たに出生した世代であるため、初期エージェントの構成の影響を受けやすくエージェントの属性値の自由度の高さによりばらつきが生じたと考えられる。他方、50 歳以上については主な人口変動の要因が死亡であるため、シミュレーションごとのばらつきはかなり少ない。シミュレーション結果については属性決定がランダムに行われることから、このように複数の試行を行いシミュレーションのアンサンブルとして結果を把握する必要がある。

つぎに、社人研の推計結果と比較すると、65 歳未満の人口に関してはおおむね一致していることがわかる。すなわち、本論文で試みたエージェントベースの人口推計モデルにおけるエージェント属性決定手法を用いた人口推計によって、妥当な推計結果が得られたといえる。45-49 歳前後の人口で多少の差がみられるが、50 年前に生まれた世代であることから、先述のように本推計では推計開始後に出生した人口であるのに対し、社人研の推計では推計開始前の実績値から得た人口であるという違

Table 11 Mortality rate⁶ by sex and age (NIPSSR)

Age	2010		2025		2040	
	M	F	M	F	M	F
⋮	⋮	⋮	⋮	⋮	⋮	⋮
65	12.2	5.0	11.1	4.5	10.1	4.0
66	13.2	5.4	12.0	4.9	10.9	4.4
67	14.3	5.8	13.0	5.3	11.8	4.7
⋮	⋮	⋮	⋮	⋮	⋮	⋮
72	22.2	9.6	20.2	8.8	18.1	7.7
⋮	⋮	⋮	⋮	⋮	⋮	⋮
77	39.1	17.7	34.7	15.6	30.4	13.6
⋮	⋮	⋮	⋮	⋮	⋮	⋮
82	68.7	34.0	60.3	29.4	53.0	25.3
⋮	⋮	⋮	⋮	⋮	⋮	⋮
87	115.8	68.7	102.7	58.4	90.6	49.7
⋮	⋮	⋮	⋮	⋮	⋮	⋮
92	183.0	129.0	168.3	111.7	150.6	96.6
⋮	⋮	⋮	⋮	⋮	⋮	⋮
97	264.0	203.4	256.1	193.1	233.4	171.0
⋮	⋮	⋮	⋮	⋮	⋮	⋮

⁶ Number of deaths per 1,000 people

いがある。本論文で扱っているシミュレーションモデルは、パラメータについては将来シナリオを一切設定していないシンプルな人口推計モデルであるため、45-49 歳前後の人口についてはその親の世代 (Figs. 6, 7 における 65-75 歳前後) の人口の多さがそのまま出生数の増加に反映されていることが差異の要因だと考えられる。他方、65 歳未満の人口とは対照的に、70 歳以上の人口はわずかに差がみられる。これは、原モデルと社人研の推計とでの将来死亡率の仮定に差があることが原因と考えられる。ここで、Table 11 に社人研の推計における死亡率 (人口千対) の仮定値を示す。代表として、2010, 2025, 2040 年の仮定値を載せている。最も死亡しやすい仮定である死亡高位仮定値を示しているが、本推計で将来一定として用いられている Table 5 の死亡率と比較すると、2010 年時点のデータでは社人研の死亡率の方が低く、より死亡しにくい仮定となっている。さらに、社人研の推計では近年平均寿命が延びていることなどの傾向を反映させるため、年数が進むにつれて死亡率を低下させている。このような死亡率の仮定値の違いから、高齢者の人口に差が生じていると考えられる。

6. おわりに

本論文では、属性決定手法に拘束条件付最適化を用いる Hara らのエージェントベースの人口推計モデルについて、その実装を見直したうえでその計算性能などを評価した。最適化の収束状況および最適解の評価指標値の

結果から、統計データに従った属性決定が行われており、エージェントの属性決定手法は適切な設計となっていることが確認できた。また、計算量の計測結果からは、属性決定に計算量が多いSAを用いているが、500,000世帯程度の規模であれば現実的な計算量で推計が可能であることが示された。得られた将来人口分布の結果では、コホート要因法を用いた国立社会保障・人口問題研究所の推計との比較で妥当な推計結果が得られている。原モデルは日本全国を対象としているが、地域を明示的に考慮し、転入・転出を加えた地域別モデルに発展させる予定である。また、エージェントベースの人口推計モデルである利点を活かし、人口問題に対する政策の評価を可能にしてシナリオ分析を進めたい。

参考文献

- [1] N. Gilbert and K. G. Troitzch: *Simulation for the Social Scientists*, 2nd. edition, Open University Press (2005)
- [2] 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, Vol. 52, No. 7, pp. 582–587 (2013)
- [3] 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；計測自動制御学会 システム工学部研究会 (2010)
- [4] S. Hara, H. Kita, K. Ikeda and M. Susukita: Configuring agents' attributes with simulated annealing; *Agent-Based Approaches in Economic and Social Complex Systems VII* (T. Murata, T. Terano and S. Takahashi (eds.)), Agent-Based Social Systems, Vol. 10, pp. 45–59 (2013)
- [5] S. Kirkpatrick, C. D. Gelatt and M. P. Vecchi: Optimization by simulated annealing; *Science*, Vol. 220, pp. 671–680 (1983)
- [6] 国立社会保障・人口問題研究所：日本の将来推計人口(平成24年1月推計) (2012)

- [7] 稲垣：日本の将来社会・人口構造分析—マイクロ・シミュレーションモデル (INAHSIM) による推計, 日本統計協会 (2007)
- [8] 山本, 小山, 出口：Soars を用いた村落の人口減少についてのデモグラフィックシミュレーション；計測自動制御学会システム工学部研究会資料 (2005)
- [9] 倉橋：社会システムの研究動向4－評価・分析手法(2)－モデル推定と逆シミュレーション手法；計測と制御, Vol. 52, No. 7, pp. 588–594 (2013)
- [10] 原：世帯構成を考慮したエージェントベースの人口推計モデル, 京都大学大学院情報科学研究科社会情報学専攻平成24年度修士論文 (2013)

著者略歴

ふく た じゅん や
福 田 純 也



1989年7月1日生。2012年3月岐阜工業高等専門学校専攻科電子システム工学専攻修了。同年4月京都大学大学院情報科学研究科社会情報学専攻修士課程に進学, 2014年3月同大学社会情報学専攻修士課程修了。同年4月トヨタ自動車(株)入社, 現在に至る。エージェントベースの人口推計モデルによる社会シミュレーションの研究に従事。

き た はじめ
喜 多 一 (正会員)



1959年11月28日生。1987年京都大学大学院工学研究科電気工学専攻博士後期課程研究指導認定退学。同年京都大学工学部助手。東京工業大学大学院総合理工学研究科助教授, 大学評価・学位授与機構教授, 京都大学学術情報メディアセンター教授を経て, 2013年より京都大学国際高等教育院教授となり現在に至る。社会シミュレーション, 情報教育などの研究に従事。工学博士, 情報処理学会, 電気学会, 計測自動制御学会などの会員。