

# 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

水流 風馬<sup>1</sup>・平野 巧真<sup>2</sup>・鈴木 温<sup>3</sup>

<sup>1</sup>非会員 名城大学 理工学部社会基盤デザイン工学科 (〒468-0073 愛知県名古屋市天白区塩釜口1丁目501)

160448064@ccmailing.meijo-u.ac.jp

<sup>2</sup>非会員 名古屋市役所(〒460-8508 愛知県名古屋市中区三の丸3丁目1-1)

E-mail:150448086@ccmailing.meijo-u.ac.jp

<sup>3</sup>正会員 名城大学教授 (〒468-0073 愛知県名古屋市天白区塩釜口1丁目501)

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

人口減少や単身高齢者増加等、世帯構造変化が多く都市問題と関わりがあるため、世帯構造や世帯分布の将来予測や施策の評価のためのツールとして、筆者らは、世帯マイクロシミュレーションモデル(HUMS: Household based Urban Micro-Simulation Model)の開発を進めてきた。マイクロシミュレーションモデルの構築を行うには、初期の世帯マイクロデータを生成する必要がある。既存のモデルでは、主にアンケートデータをもとに初期マイクロデータを生成していたが、データ収集にかかる費用や手間を削減するため、本研究では、富山市を対象として、国勢調査等のオープンデータをもとに初期世帯マイクロデータを生成する手法を提案することを目的とする。

**Key Words** : *micro simulation, open data, household structure, age relationships, error function,*

## 1. はじめに

現在、我が国では少子高齢化や人口減少が進み、既成市街地の空き家の増加や中心市街地の衰退等による都市機能低下が懸念されている。また、全国的に世帯数が増加傾向にある一方で、三世帯が同居する世帯の減少などにより、平均世帯人員が減少しているという傾向にある。これに高齢化の影響も加わることで高齢者の単身世帯が増加している。さらに、少子化により子供のいない世帯も増加しており、このような世帯構造の変化は、地域社会の構造に影響を与え、今後の都市計画の策定を左右する原因となりうる。このため、持続可能な将来の都市政策決定に際して、地域内の世帯構成変化などの人口変化を予測することは、極めて重要である。マイクロシミュレーションを用いた人口予測は、個人属性の遷移をもとに人口動態をモデル化するものであり、小地域での世帯属性推定に適している。一方で、人口予測に必要な実都市の世帯属性情報は一般には公表されていないため、初年度の世帯属性情報の推定が必要となる。鈴木ら<sup>1)</sup>はアンケート調査を用いて世帯属性情報の推計を行ったが、対象都市ごとに大規模なアンケート調査を要す

るため、費用や手間がかかり汎用性に欠ける。そこで、本研究では富山県富山市を対象として、入手可能な公表データを用いて、同世帯内に含まれる個人同士の年齢関係を考慮した初期世帯マイクロデータを生成するシステムを構築し、生成したデータの妥当性を検証することを目的とする。

## 2. 本研究の位置づけ

本研究では、マイクロシミュレーションの汎用性向上のため、入手可能なオープンデータを用いて、世帯内の個人同士の年齢関係を考慮した初期世帯マイクロデータの生成を行う推計モデルを構築する。本研究で構築する初期世帯マイクロデータは平成22年の国勢調査のデータから周辺分布を作成し、個人・世帯属性の誤差修正を行いながら推定していく。その後、本モデルより得られた結果を用いて、実測値との比較を行いデータの妥当性を検証することを目的とする。

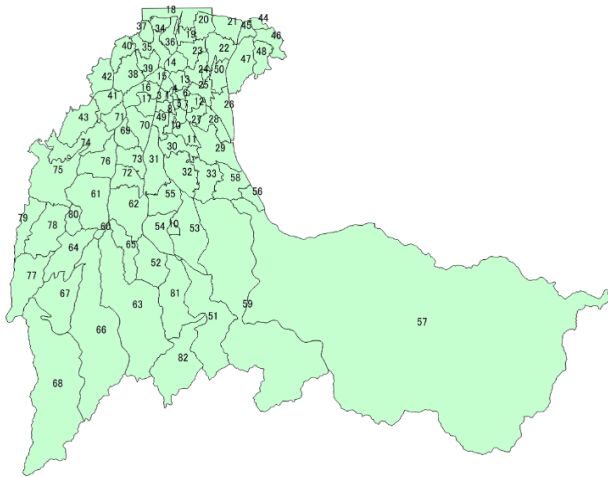


図1 富山市のゾーン区分

3. 対象地域と使用データについて

(1) 対象地域について

対象地域として、コンパクトシティ政策を積極的に進めている富山県富山市を選定した。本研究では対象地域を国勢調査で使用されている 82 ゾーンに分け、初期世帯マイクロデータを推定する。82 ゾーンを図1に示す。

(2) 使用するデータについて

a) 人口について

人口は平成 22 年の国勢調査の「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」のデータを用いた。このデータには個人の、性別、5 歳階級年齢、居住ゾーンが含まれている。年齢については、5 歳階級であるため 15 歳未満と 85 歳以上についてはまとめて、それぞれを年齢 1 と年齢 16 とする。

b) 世帯数・世帯タイプについて

世帯数は平成 22 年の国勢調査の「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 世帯当たり人員、施設等の世帯数)」のデータから算出する。また、世帯タイプは平成 22 年の国勢調査の「第 11 表 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)」の富山県富山市のデータを用いた。Sugiki et al.<sup>2)</sup>の研究では 7 人世帯までを推定対象とし、7 人以上の世帯を統合していたが、年齢不詳人口を扱っていないことから、総人口と世帯人数別世帯数から算出される総人口が一致しておらず、誤差が生じている。そこで本研究では新たに 8 人世帯を作成し 7 人以上世帯による誤差を調整した。本研究ではこの 8 人世帯を含む 30 の世帯人数別の世帯数で推定を行う。作成した世帯人数別の世帯数を表 1 に示す。

c) 周辺分布について

表 1 推定に使用する世帯タイプ

世帯人数	世帯タイプ (0内はコード番号)	世帯人数	世帯タイプ (0内はコード番号)
1人	(11) 単身・男	4人	(40) その他
	(12) 単身・女		(51) 夫婦+子 3人
2人	(21) 夫婦のみ	5人	(52) 夫婦+両親+子 2人
	(22) 男親+子 1人		(53) 夫婦+子 2人+親
	(23) 女親+子 1人		(50) その他
	(20) その他		(61) 夫婦+子 4人
	(31) 夫婦+子 1人		(62) 夫婦+両親+子 2人
3人	(32) 男親+子 2人	6人	(63) 夫婦+子 3人+親
	(33) 女親+子 2人		(60) その他
	(34) 夫婦+親 1人		(71) 夫婦+子 5人
	(30) その他		(72) 夫婦+両親+子 3人
	(41) 夫婦+子 2人		(73) 夫婦+子 4人+親
4人	(42) 女親+子 3人	7人	(70) その他
	(43) 夫婦+両親		(80) その他
	(44) 夫婦+子+親		(90) 施設等世帯
	-		-
	-		-

推定に使用する周辺分布として平成22年の国勢調査の「夫の年齢、妻の年齢別夫婦数」、「単独世帯の年齢別人口」、平成22年の人口動態調査の「出生順位母親と子供の年齢差人口」の3つのデータを用いる。

d) 続柄について

続柄は表2に示された8種類に分類し、世帯内の性別年齢順位の降順と世帯タイプによって個人に付与される。「その他」の続柄はその他世帯の個人全員に付与される。また、続柄なしについては、世帯タイプから続柄をつけられない個人に付与する。例としては、祖父(祖母)と孫のような場合が挙げられる。

4. 初期マイクロデータの生成方法

(1) 生成方法の概要

本研究で用いられる初期世帯マイクロデータ推定システムのフローを図2に示す。まず居住ゾーン、5 歳階級別年齢、性別を含んだ個人データとそのリストを作成する。同時にゾーン、世帯人数別の世帯データを作成する。個人、世帯データは国勢調査の居住ゾーン別、5 歳階級別、性別人口及び、世帯の家族類型別世帯数のデータから作成している。この時国勢調査の居住ゾーン別、5 歳階級年齢別、性別の人口データには年齢不詳の人口データも含まれているため、同ゾーンの 5 歳階級別年齢別、性別人口の割合から、年齢不詳の個人に年齢を確率的に付与する。次に、先程の手順で作成した個人データと世帯データを結合し、個人の年齢の降順に世帯タイプに合った続柄を付与する。その後個人、世帯タイプの入れ替えを繰り返して行く。入れ替えを行った後入れ替え

表 2 推定に使用する続柄

親 (男)	親 (女)	夫	妻
子供 (男)	子供 (女)	その他	続柄なし

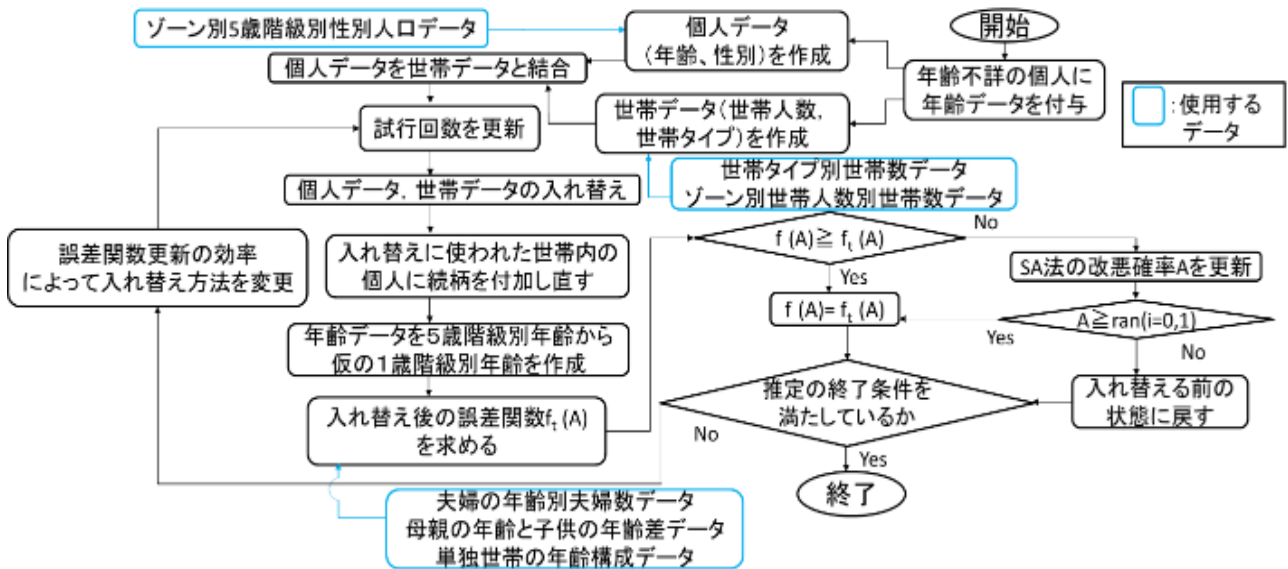


図2 初期マイクロデータ推定のフロー

に関わった世帯内の個人の続柄を付加しなおし、データ全体の誤差関数の計算を行う。この誤差関数は周辺分布の割合と、推定中のデータの周辺分布と同じ条件の割合との差を評価する関数で、誤差が小さくなると誤差関数は小さくなるように設定されている。この誤差関数が入れ替え前より小さくなった場合、誤差関数を更新し次の入れ替えに移る。誤差関数が大きくなった場合、入れ替えを行う前の状態に戻し次の入れ替えに移る。入れ替えによる誤差関数の更新効率が低下したとき入れ替え方法を変更する。以上の操作から誤差関数を更新していき、終了条件を満たしたとき推定を終了する。終了条件は3つの周辺分布において存在する世帯または個人が0%の項目に、推定中のデータの世帯または個人が存在しないことである。年齢不詳の個人に対する年齢付与、入れ替えの対象となる個人、世帯の選択は全てランダムに行う。

## (2) 個人・世帯の入れ替え方法について

### a) 入れ替えの対象について

入れ替え対象の選定方法として所属する世帯タイプでは続柄を付与できない「続柄なし」や周辺分布では0%の組み合わせを優先的に入れ替えの対象とする。入れ替えられる個人は同ゾーン内からランダムに選択し、入れ替えられる世帯に関してはゾーンを越えて世帯人数が一致し世帯タイプが異なる世帯タイプをランダムに選択する。

### b) 入れ替えの制約について

入れ替えの制約として個人の入れ替えについては、使用するデータの特性から居住ゾーンごとに年齢別、性別人口が決まっているため、居住ゾーンを越えて入れ替えを行うことができない。世帯タイプの入れ替えについては、推定対象地域全体で世帯数が決まっているため、居住ゾーンを越えて入れ替えを行うことができる。しか

し世帯人数が異なる世帯タイプを入れ替えると世帯人数に誤差が生じるため、世帯人数の同じ世帯タイプ同士でのみ入れ替えが可能である。

### c) 個人の入れ替え方法について

個人の入れ替え方法について説明する。まず最初に同じゾーンからランダムに2人選択し、入れ替えを行う。次に、それぞれの入れ替えた2人が所属している世帯の性別年齢別順位を付け直す。最後に、その性別年齢別順位と世帯タイプによって続柄を付け直す。

### d) 世帯の入れ替え方法について

世帯の入れ替え方法について説明する。まず最初に所属している世帯の世帯人数が同じ人をランダムに2人選択しする。次にその2人が所属している世帯の世帯タイプの入れ替えを行う。最後に、世帯タイプに合うように続柄を付け直す。

## (3) 誤差関数について

入れ替えた推定データと周辺分布の誤差を評価する誤差関数の式は柘井<sup>3)</sup>の研究で用いられている誤差関数の式をもとに以下の式(1)を使用する。

$$f(A) = \sum_{l=1}^L \sum_{i=1}^{G_l} |C_i(A) - \text{Round}(M(A) * R_i(A))| \quad (1)$$

Lは周辺分布の数、G<sub>l</sub>は周辺分布に含まれる項目数、M(A)は推定中のデータに含まれる人数の合計、C<sub>i</sub>(A)は条件を満たす推定中のデータに含まれる人数、R<sub>i</sub>(A)は周辺分布の条件に含まれる人数の割合を表す。出生順位母親と子供の年齢差の割合の周辺分布の場合、条件は母親の年齢と出生順位の子供の年齢差となり、M(A)は推定中の復元データに含まれる親子の数、R<sub>i</sub>(A)は周辺分布の母親の年齢と出生順位の子供の年齢差の割合を表す。

この時M(A)にRi(A)をかけた値は推定データにおいて条件を満たす理想的な親子の数となるので、推定中のデータとの差の絶対値をそれぞれの項目で合計した値が小さくなる程誤差関数が小さくなる。なお、式1のRound()は、整数値に丸める関数を示しており、誤差を人数で表すために用いている。以上の操作を周辺分布の項目数だけ行った合計を誤差関数とし、誤差関数が小さくなる時その入れ替えを採用する。

(4) SA法について

a) SA法とは

SA法とは、最適化問題への汎用の乱択アルゴリズムで、局所会に収束しないように確率で改悪を行うことで最終的に最適解に導く手法である。SA法の改悪確率については一般的に推定初期は高く、推定が進むにつれて小さくなるように設定されており、本研究でも同様の設定で行う。改悪確率については次項で説明する。本研究では、推定中のデータ誤差関数が大きくなった場合、1から100までの範囲で乱数を発生させ、改悪確率より小さくなった場合、その入れ替えを採用している。

b) 改悪確率について

あSA法を使用する際に試行回数に応じて小さくなる温度Tを設定する必要がある。温度Tは以下の式(2)で表される。式(3)においてff(A)は推定開始時の誤差関数の値、ft(A)は推定中の誤差関数の値を示している。この式より温度Tの最大値は1、最小値は0となり、推定が進み誤差関数が小さくなる程Tは0に近づく。

$$T = \frac{ft(A)}{ff(A)} \tag{2}$$

次に改悪確率Pは先述した温度Tを用いて以下の式 (3) で表される。

$$P = \exp \frac{-1 + B + 0.01}{T} \tag{3}$$

式(4)においてBは改善が連続で行われていない回数を示し、Pが1を超える場合はP=1する。式の構造より改善が連続で行われなくなる程改悪確率は大きくなる。またTの値が小さくなる程改悪確率の増加が遅くなるため、推定が進むにつれて改悪を行う頻度が低くなるように設定されている。

4. 推定結果と精度検証

(1) 夫婦の年齢別夫婦数の割合

推計結果を表3に示す。推計値と周辺分布との差は、割合の高いところでは0.5%以上の結果も出ているが、既存のモデルより少ない。また、ほとんどの年齢で小さい値となった。したがって、夫婦の年齢差を考慮した初期世帯マイクロデータを推計できたと考えられる。

(2) 単独世帯の人口

推計結果を表4に示す。推計値と周辺分布との差はすべての年齢で0.00%となった。したがって、単独世帯人員を考慮した初期世帯マクロを推計できたと考えられる。

(3) 出生順位母親と子供の年齢差別人口

推計の結果を表5に示す。既存モデルでは、推定値と周辺分布との差がほとんどの年齢差で0.5%以上の誤差があり、最大で30%以上も誤差があった。つまり、あり

表 3 夫婦の年齢別夫婦数の割合 (下段は周辺分布との差)

		妻の年齢 (歳)						
		20~29	30~39	40~49	50~59	60~69	70~79	80~
夫の年齢 (歳)	20~29	2.58%	1.03%	0.60%	0.00%	0.00%	0.00%	0.00%
		0.02%	0.40%	0.58%	0.00%	0.00%	0.00%	0.00%
	30~39	1.88%	11.50%	1.15%	0.02%	0.02%	0.00%	0.00%
		0.00%	1.80%	0.00%	0.01%	0.02%	0.00%	0.00%
	40~49	0.16%	4.40%	10.46%	0.45%	0.10%	0.00%	0.00%
		0.00%	0.00%	1.36%	0.00%	0.08%	0.00%	0.00%
	50~59	0.02%	0.33%	4.94%	11.36%	0.63%	0.06%	0.01%
		0.00%	0.00%	0.15%	0.60%	0.03%	0.04%	0.00%
	60~69	0.02%	0.13%	0.40%	7.11%	15.95%	0.43%	0.10%
		0.01%	0.08%	0.12%	0.00%	0.05%	0.08%	0.10%
	70~79	0.01%	0.03%	0.12%	0.21%	6.55%	8.58%	0.43%
		0.01%	0.02%	0.11%	0.14%	0.01%	0.00%	0.27%
	80~	0.00%	0.00%	0.00%	0.08%	0.28%	3.44%	3.33%
		0.00%	0.00%	0.00%	0.07%	0.20%	0.01%	0.44%

表 4 単独世帯人員の年齢構成

年齢	構成割合	周辺分布との差
0-14歳	0.00%	0.00%
15-19歳	3.88%	0.00%
20-24歳	13.17%	0.00%
25-29歳	8.44%	0.00%
30-34歳	6.51%	0.00%
35-39歳	6.61%	0.00%
40-44歳	5.67%	0.00%
45-49歳	5.41%	0.00%
50-54歳	5.20%	0.00%
55-59歳	6.59%	0.00%
60-64歳	8.11%	0.00%
65-69歳	6.96%	0.00%
70-74歳	6.51%	0.00%
75-79歳	7.20%	0.00%
80-84歳	5.68%	0.00%
85歳以上	4.06%	0.00%
合計	100.00%	0.00%

えない数の母親と子供の年齢差の組み合わせが推計されていた。しかし、本研究の推計値と周辺分布の差はほとんど0.5%以内で最大でも5.77%の誤差に収まった。したがって、母親と子供の年齢差を考慮した初期世帯マイクロデータを推計できたと考えられる。

#### 4) 誤差関数の途中経過

試行回数の増加に対する誤差関数の変異を図3に示す。推定初期は個人や世帯のデータがほぼランダムであるためデータの入れ替えによって誤差関数は急激に減少していった。しかし、推定後期になり周辺分布に近づいてい

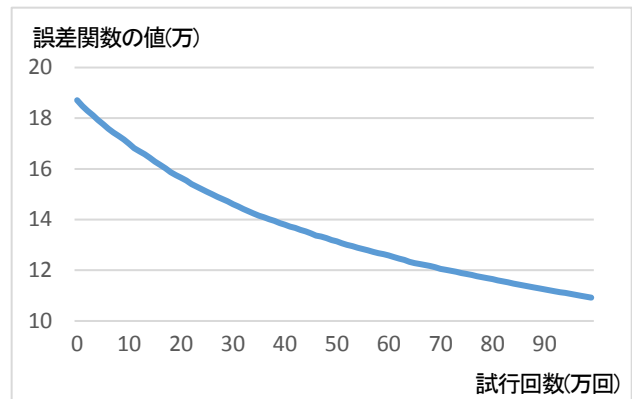


図 3 誤差関数の減少の途中経過

くと変動が緩やかに進むようになった。今回の推定時間は約50時間となった。

#### 5. おわりに

本研究では、公表されているオープンデータを用いて実在する都市である富山市を対象に、人口約40万人規模の初期世帯マイクロデータの推定システム構築と、推定データと周辺分布との誤差及び、既存モデルと比較し、有効性を検討することを目的とした。その結果居住ゾーン別、5歳階級別、性別人口及び、世帯タイプ別世帯数が実測値と一致し、同世帯の個人同士の年齢関係を考慮した初期世帯マイクロデータを作成するシステムを構築することができた。また既存モデルと比較して概ね精度が向上した初期世帯マイクロデータを推定することができた。また、他の都市に関しても本研究で構築したシステムを同様に用いて周辺分布に近い推定データを作成で

表 5 出産順位別母親と子供の年齢差 (下段は周辺分布との差)

		第1子	第2子	第3子	第4子	第5子
年齢差	0~14歳	0.00%	0.00%	0.00%	0.00%	0.00%
		0.05%	0.00%	0.00%	0.00%	0.00%
	14~19歳	1.46%	0.17%	0.08%	0.00%	0.00%
		0.13%	0.10%	0.02%	0.00%	0.00%
	20~24歳	13.58%	5.87%	1.93%	1.31%	1.92%
		0.01%	0.00%	0.00%	0.04%	1.92%
	25~30歳	38.40%	25.89%	17.75%	8.96%	7.69%
		0.00%	0.00%	0.09%	0.09%	5.77%
	31~34歳	30.79%	42.98%	40.64%	35.45%	21.15%
		0.00%	0.71%	0.15%	0.00%	0.00%
	35~40歳	13.39%	22.08%	34.79%	43.10%	46.15%
		0.00%	0.00%	0.02%	0.06%	0.00%
41~45歳	2.39%	3.02%	4.82%	11.19%	23.08%	
	0.02%	0.67%	0.38%	0.43%	3.85%	
45~49歳	0.00%	0.00%	0.00%	0.00%	0.00%	
	0.10%	0.06%	0.10%	0.63%	0.00%	
合計		100.00%	100.00%	100.00%	100.00%	100.00%

きた。しかし、本研究では母親と子供の年齢差を子供の出生順位ごとに考慮できていたが、子供同士の年齢関係を考慮できていない。そのため、一部世帯では子供同士の年齢差が過度に大きくなってしまっている世帯が存在している。したがって、子供同士の年齢差の周辺分布を用いて推計を行わなければいけないが、国勢調査などのオープンデータでは子供同士の年齢関係を詳細に掲載したデータが存在しなかった。そのため、様々なオープンデータを集計し組み合わせて、子供同士の年齢関係を詳細に示した周辺分布を作成する必要がある。また、このように周辺分布を増やしていけば、推計データの精度向上を見込めるが、推定時間の増加や誤差関数が収束にくくなると考えられる。この問題の解決のためにも推定速度を高めるためのアルゴリズムの改良を考える必要がある。また、世帯タイプの「その他世帯」に含まれる個人の続柄を推定する必要がある。「その他世帯」に該当する世帯構造として祖母と孫の2人世帯や兄弟のみの組み合わせの世帯などが考えられる。実際にマイクロシミュレーションにより将来の世帯マイクロデータを推定する場合、個人の続柄データは全個人に対し必ず必要にな

るが、現在の初期マイクロデータにはその他世帯に属する個人に続柄が付されていない。そこで続柄を付加するために世帯人数や個人の年齢から続柄を決定する定義を考える必要がある。

**謝辞：**本研究はJSPS科研費の助成を受けたものです。

#### 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—人口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年
- 2) Nao Sugiki, Tomoya Muranaka, Noriko Otani, Kazuki Miyamoto: Agent-based Estimation of Household Micro-data 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年
- 3) 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

(?)

## STUDY ON ESTIMATION OF INITIAL MICRO-DATA FOR HOUSEHOLD WITH OPEN DATA

Fuma TSURU, Takuma HIRANO and Atsushi SUZUKI

The authors have been developing Household based Urban Micro Simulation Models as a tools for future prediction of household structure and household distribution and measures evaluation because changes in the household structure, such as population decline and increase in single elderly, are related to many urban problems. To build a micro simulation model, it is necessary to generate the initial household micro data. In the existing model, the initial micro data was generated mainly based on the questionnaire. Therefore, this study aims to propose a method for generating initial household microdata based on open data such as the national census in Toyama City.