

# 遺伝的マッチング法に基づく 人口予測手法の開発

竹内 一貴<sup>1</sup>・川久保 慎二<sup>2</sup>・瀬谷 創<sup>3</sup>

<sup>1</sup>非会員 神戸大学大学院 工学研究科市民工学専攻 (〒657-8501 神戸市灘区六甲台町1-1)

E-mail:196t119t@stu.kobe-u.ac.jp

<sup>2</sup>非会員 株式会社日本工営 交通都市部部 (〒530-0047 大阪市北区西天満1-2-5)

E-mail:a8920@n-koei.co.jp

<sup>3</sup>正会員 神戸大学大学院 准教授 工学研究科市民工学専攻 (〒657-8501 神戸市灘区六甲台町1-1)

E-mail:hseya@people.kobe-u.ac.jp

コーホート要因法を用いた将来人口の推計においては、「出生」、「死亡」、「移動」の仮定値を設定する必要がある。特に市区町村レベルでは、純移動率で与えられる「移動」の仮定値をどのように設定するか推計結果が依存することが知られている。純移動率は、通常過去のトレンドから外生的に与えられる。しかし、日本は人口減少と高齢化が年々進んでおり、特に地方部においては、先行的にこれらの現象が進行している地域も少なくない。したがって、ある自治体の将来の純移動率を設定する際に、自身の過去の純移動率だけではなく、産業構造等が似通った「他の」自治体の過去の純移動率の変化経路の情報が役に立つ可能性があると考えられる。そこで本研究では、ある自治体に対して、産業構造等の属性変数を用いて類似自治体をマッチングさせ、自身の過去の純移動率だけでなく、類似自治体における過去の純移動率の変化経路の情報を利用して純移動率の将来推計値を設定する新たなコーホート要因法の提案を試みる。マッチングの方法としては、遺伝的マッチングを用いることとした。

**Key Words :** *population projection, cohort component method, Genetic Mathing, net migration rate*

## 1. はじめに

### (1) 研究の背景

将来推計人口は、都市計画、産業政策、国土政策などの多くの計画・政策の基礎情報となる。我が国では、国立社会保障・人口問題研究所（以下、社人研）がコーホート要因法を用いた将来推計人口を国・都道府県・市区町村等の単位で公表している。また、いわゆる増田レポート等、いくつかの独自推計の試みも行われている。

コーホート要因法を用いた将来人口推計においては「出生」、「死亡」、「移動」の3人口変動要因の仮定値を設定する必要がある。人口推計を行う単位が空間的に細かいほど、人口移動の影響は出生・死亡と比較して相対的に大きくなるため、市区町村などの詳細な空間領域における推計では、地域間移動の仮定値をどのように設定するか推計結果が依存することが知られている。特に市区町村以下のレベルでは、そもそも年齢階層別の地域間移動のデータを入手することが難しいため、人口推計においては転入と転出を区別しない純移動率を用い

られることが多い。純移動率は、上述の推計人口を含めて、通常過去のトレンドから外生的に与えられる。

しかし、日本は人口減少と高齢化が年々進んでおり、特に地方部においては、先行的にこれらの現象が進行している地域も少なくない。したがって、ある自治体の将来の純移動率を設定する際に、自身の過去の純移動率だけではなく、産業構造等が似通った「他の」自治体の過去の純移動率の変化経路の情報が役に立つ可能性があると考えられる。これが本研究の問題意識である。

### (2) 研究の目的

以上のような背景の下、純移動率を過去の自地域におけるトレンドから推計する手法に対し、産業構造等が似通った過去の「他地域」の純移動率を用いて推計する新たな方法論を提示し、その方法論の特性を把握することを目的とする。従来の純移動率を過去のトレンドから推計する手法は、属性変数を考慮しないため人口減少の「対策」を議論することが難しいが本研究では属性変数を用いて過去の類似自治体とマッチングされるため、例えば産業構造と人口増減の関係性を捉えることが可能で

あり、属性変数の改善（例えば大学卒業者率の増加等）の効果が純移動率に及ぼす影響をマッチング相手の変化として記述することが出来る。この点は地方創生戦略を立案する上で重要な特性であると考えられる。

(3) 研究の方法

最も広く用いられている将来人口推計手法として、コーホート法が挙げられる。コーホート法は大きくコーホート変化率法とコーホート要因法に分けることが出来る。ここで、コーホートとはある特定の同じ人口事象を経験した集団、すなわち同時経験集団である。

一般的にはコーホート要因法が用いられることが多く、本研究でもコーホート要因法を用いる。また、人口推計において重要である純移動率の仮定は産業構造等の属性変数を用いて類似自治体をマッチングさせ、自身の過去の純移動率だけでなく、類似自治体における過去の純移動率の変化経路の情報を利用して純移動率の将来推計値を設定する。マッチングの方法としては遺伝的マッチング (Diamond and Sekhon, 2013) を用いることとした。その後、純移動率のみを変数としてマッチングした場合と産業構造等の属性変数を含めてマッチングをした場合の比較を通して、属性変数の検討を行う。

2. 分析

(1) 分析の概要

本研究における人口推計の対象は日本全国の市区町村とし、2015年に実施された国勢調査のデータを検証用のデータとして使用し、1990年～2010年に実施された国勢調査のデータを基に各コーホートの純移動率を算出し、産業比率や大学卒業者率といった様々な属性変数で自治体を特徴づけし、過去の類似自治体と遺伝的マッチングを用いてマッチングすることで純移動率の仮定値を設定した。3人口変動要因の他の2つである、出生に関しては、市区町村別の出生率は年による変動が大きく不安定であることから、こども女性比および0-4歳性比の仮定値を用いて、死亡に関しては、0歳から100歳程度までの死亡率の推移を表す生命表を用いて生残率を算出する。以上のように設定した3人口変動要因を用いたコーホート要因法により実際に2015年の人口を推計し、実測値と比較する。予測誤差の測度として、平均二乗平方根誤差 (RMSE) と全国の市区町村を対象にしており人口の多い地区ではRMSEは大きくなることから平均二乗平方根誤差率 (RMPSE) も同時に用いることで精度の検証を行う。

(2) Genetic Matching

Genetic Matchingにおける距離は、マハラノビス距離を一般化したものであり、一般化マハラノビス距離は以下のように定義される。

$$GMD_{ij} = \left\{ (x_i - x_j)^t \left( S^{-1/2} \right)^t W S^{-1/2} (x_i - x_j) \right\}^{1/2} \quad (1)$$

$GMD_{ij}$  : 対象者*i*と対象者*j*間の一般化マハラノビス距離

$x_i$  : 対象者*i*の共変量ベクトル ( $k \times 1$ )

$S^{-1/2}$  : 共変量ベクトルの共分散行列  $S$  のコレスキー分解

$W$  : すべての共変量の重み対角行列 ( $k \times k$ )

(3) Genetic Algorithms (GA) に基づく重み行列  $W$  の決定

GAは生物進化の原理に着想を得たアルゴリズムであり、確率的探索の一手法である。図1にGenetic Matching Algorithmのフローチャートを示す。⑧における基準を満たさなかった場合、再びpop.size個の $W$ を作成するがそこでGAが用いられる。GAは「選択 (Selection)」, 「交差 (Crossover)」, 「突然変異 (Mutation)」の遺伝的操作を用いる。そのフローチャートを、図2に示す。具体的には、評価の基準において、評価が上位の個体を複数個選択する。次に、選ばれた個体の書く重み成分を他の個体と一部入れ替えることを行う。この操作を交差という。この操作を行うと一つ前の世代の値に依存した重みのみ生成される。この問題を解決するために一定の確率で重みを変化させることを突然変異という。一連の操作によって算出された $W$ を用いてGenetic Matchingを行うことが出来る。

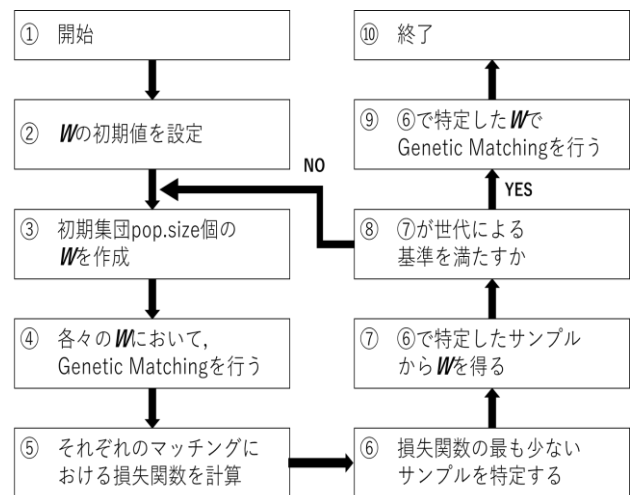


図-1 Genetic Matching アルゴリズムのフローチャート

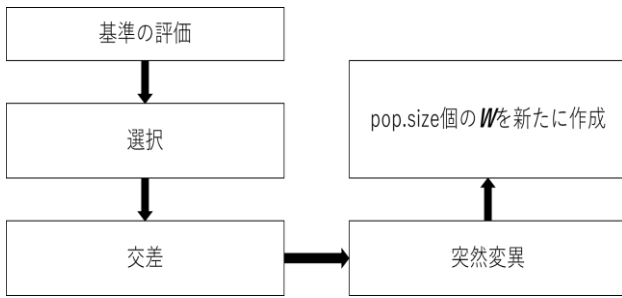


図-2 Genetic Matching 上のGAのフローチャート

### 3. 実データを用いた検証

#### (1) 本研究の分析対象

我が国では、平成の大合併以降で市区町村の数が大きく減少している。市区町村合併後の人口は単純な総和で求めることが出来るが、合併後から合併前の市町村人口を計算することは不可能であるため、本研究の推計対象である日本全国の市区町村とは2015年時点での市区町村1741都市とする。なお、政令指定都市の区に関しては、熊本市や相模原市、岡山市のように近年に新たに指定された場合も多く、その際の過去の人口データが入手不可能であったため、政令指定都市は市単位で、東京特別区のみ、区単位を対象とする。

#### (2) 本研究に用いるデータ

本研究における人口推計を行う際のマッチングに使用する説明変数は表1の通りである。政令指定都市及び中核市は2018年4月1日に指定されている政令指定都市20都市及び、中核市54都市とした。また、東京特別区から新宿区を政令指定都市と同様の扱いとし、各市町村役場から政令指定都市、中核市までの距離を算出する。ここで、自身が政令指定都市及び中核市である場合、距離が0となるため栗田・腰塚（1988）による対面の領域間平均移動距離を用いた。対象年は1990年から2015年に行われた国勢調査、純移動率は5期間のデータを用意した。コーホート要因法で必要となる生残率は社人研の日本版データベースにおける都道府県別5歳×5年死亡率を用いた。

#### (3) マッチングにおける説明変数の設定

本研究では遺伝的マッチング法に基づき、純移動率を決定するが、マッチングを行うにあたって、説明変数の選択が重要になってくる。本研究では1900年から2005年の間を5年ごとのグループに分け1期先の予測を行う。10年前からの2期間の説明変数を用いてマッチングを行った。また、将来人口推計に有効である説明変数を検討するために純移動率のみでマッチングを行った場合と純移動率に各説明変数を加えた場合で実測値との誤差の比較

を行う。

表-1 マッチングに用いる説明変数

説明変数名	データソース
農業就業者率	国勢調査
林業就業者率	国勢調査
漁業就業者率	国勢調査
鉱業就業者率	国勢調査
建設業就業者率	国勢調査
製造業就業者率	国勢調査
電気、ガス、熱供給、水道業就業者率	国勢調査
運輸、情報、通信業就業者率	国勢調査
第1次産業就業者数比率	国勢調査
第2次産業就業者数比率	国勢調査
第3次産業就業者数比率	国勢調査
自市区町村外への通勤者率	国勢調査
持ち家世帯数比率	国勢調査
大卒率	国勢調査
課税対象所得	内閣府市区町村別人口・経済関係データ
高齢者率	国勢調査
政令指定都市までの距離	国土数値情報
中核市までの距離	国土数値情報

表-2 コーホート数

	若者世代	生産年齢人口	高齢者世代
コーホート数	6	20	8
第1次産業就業者数比率	3	11	2
第2次産業就業者数比率	2	12	1
第3次産業就業者数比率	3	15	3
農業就業者率	2	16	2
林業就業者率	4	13	1
漁業就業者率	4	11	1
鉱業就業者率	2	8	0
建設業就業者率	3	13	0
製造業就業者率	2	10	2
電気、ガス、熱供給、水道業就業者率	3	15	1
運輸、情報、通信業就業者率	5	13	0
政令市調整距離	6	19	7
中核市調整距離	1	0	0
自市区町村外への通勤者率	3	8	1
持ち家世帯数比率	3	10	1
大卒率	3	11	0

#### (4) 分析結果

表1の全ての属性変数を取り入れた場合と純移動率のみの場合で誤差の比較を行ったところp値 = 0.05801となり、絶対誤差ではp値が0.1未満で有意傾向にあり属性変数を取り入れることが将来人口推計に有効であるといえる。表2に純移動率に各属性変数を1つ加えて遺伝的マッチングを行った場合に純移動率のみでの遺伝的マッチングに比べ実測値とのRMSPEが小さくなるコーホート数を若者世代、生産年齢人口世代、高齢者世代の3世代ごとに集計した結果を示す。なお、2行目は各世代におけるコーホート数である。結果から、「政令指定都市までの

距離」を属性変数として取り入れることで純移動率のみでの遺伝的マッチングに対して多くのコーホートで RMSPE が小さくなっている。また対応のある t 検定を行うと p 値が 0.0094 と有意水準 5% を下回っており、属性変数として「政令指定都市までの距離」を取り入れることの有効性が示唆された。

#### 4. おわりに

今回の分析により純移動率に加えて属性変数として政令指定都市までの距離などを取り入れることにより純移動率のみのマッチングと比較して誤差が小さくなっていることから、政令指定都市までの距離が似通っている都市部と都市部、郊外部と郊外部によるマッチングがなされ構造が類似している場所が取り出されたことで精度が向上したと考えられる。このように地域特性が似通った地域とマッチングすることで純移動率を自地域の過去のトレンドから推計する方法や純移動率のみでのマッチングと比較しより精度の高い推定が可能であると考えられる。

今後の課題として、今回の分析では短期的な人口予測に留まっており、長期的な予測への適用についての検討

が出来ていないことが挙げられる。今後説明変数の検討を行い長期的な予測モデルの作成を試みる。また自治体ごとに特徴が異なると考えられるので、自治体ごとにどの変数を入れるべきかを検討し、手法の改善と信頼性の向上を試みたい。

#### 謝辞

本研究は、JSPS 科研費 18H03628 および 17K14738 の助成を得たものである。また、本研究は東大 CSIS 共同研究の成果の一部である。

#### 参考文献

- 1) Diamond, A., and Sekhon, J.S.: Genetic matching for estimating causal effects: A general multivariate matching method for achieving balance in observational studies. *Review of Economics and Statistics*, Vol.95, 932-945, 2013
- 2) 国立社会保障・人口問題研究所：日本版死亡データベース(<http://www.ipss.go.jp/p-toukei/JMD/index.asp>)
- 3) 栗田治，腰塚武志：領域間平均距離の近似理論とその応用，日本都市計画学会学術研究論文集，Vol.23, pp.43-48, 1988.