

2020年度 統計データ分析コンペティション
特別賞（統計分析） [大学生・一般の部]

人口増減している市区町村の特徴と人口規模の影響

三木祐司（J F E スチール株式会社スチール研究所）

論文の概要

市区町村の人口増減について、Random Forest法による説明変数の重要度を算出し、成長の著しい情報、複合サービス業、運輸などの産業を発展させることが人口増加に効果的であることを示している。また、人口規模別の結果により、周辺地域から地方都市への人口流入が起こっている可能性を指摘し、転入者数に加えて学術研究・専門・技術サービス業の就業者数等も影響していることを示している。

論文審査会コメント

人口増減に関する要因を探るためにRandom Forestを使った分析は評価できる。

人口増減している市区町村の特徴と人口規模の影響

三木 祐司^{*1}

*1: JFEスチール株式会社 スチール研究所

1. 研究のテーマと目的

2019年の統計データ解析コンペティションで、張ら¹⁾による地方創生目標指標に関する変化要因ネットワークの推定と地域間連携策に関する大変興味深い研究があり、地方創生には、クラスタリングと各自自治体に共通する変化要因のネットワークが存在している可能性が示唆されている。

張らによると、各市区町村における人口の増減要因として、順に市町村の一般財源、15~64歳人口、課税対象所得、地方税、歳入決算額の変化数が挙げられた。これらは推定されたネットワークでその変化が他の変数におよぼす正の影響が強かった変数であり、媒介中心性の負の値が大きかった変数としては、災害復旧費、実質公債費比率、経常収支比率などが挙げられた。

また、竹内²⁾は、都道府県別に我が国における人口増減の決定要因を変量効果モデルにより解析し、医師数を十分確保でき医療体制を充実させることができたことが人口増加に寄与することを指摘した。地方創生において人口増加の要因を解析することは重要であり、多くの調査・研究がなされ、整理されてきている³⁾。

解析手法については、Random Forest Regressor 法⁴⁾を用いた。張らが用いたLASSOはスパース推定法として非常に有用であるが、説明変数同士の相関が高い場合、その中から1つしか選択されないため、クラスターlasso法を用いるなどの工夫が必要である。多重共線性については、他の解析モデルでも同様に問題となりうるが、Random Forest法は、各分割でサンプリングされた変数の数を通じて正則化パラメーターがあり、各ツリーを他のツリーとより高い相関関係にし、複数のツリーを推定することにより多様化の影響をある程度緩和する。すなわち、回帰分析とは対照的に、Random Forestモデルの一部が高度に共線的な変数によって損なわれにくく、2つの変数が同じ子ノードを提供する場合であっても、結果の品質を低下させることなく1つを選択する傾向にある。

。

2. 研究の方法と手順

Random Forest法は決定木解析を複数回行いその平均値を用いる手法であるが、解析時の決定木の最大深さ、平均化する決定木数などが最適化されないと解析精度が確保できないことがあり、最適化されるとDeep Learning以上により精度が出ることもある。そこで、max depth, n estimators, max featuresについては、kfold法4分割の交差検証により決定係数 r^2 を最大化するように、ベイズ推定によるパラメーターチューニングを実施した。

説明変数の重要性を明らかとするため、重要度と部分依存性を評価した。重要度は、Random Forest法におけるツリー内の決定ノードとして使用される特徴量の相対的ランク（すなわち深さ）を使用し、目的変数の予測可能性に関するその特徴量の相対的重要度を評価した。重要度は「ある特徴量で分割することでどれくらいジニ不純度を下げられるのか」に対応する。部分依存性は、他の特徴量はその変数の平均値とし、対象の変数を仮想的に変化させたときの目的変数を計算しており、当該変数に対する影響を仮想的に予測する。

全体のプログラムはpython3.7で作成、sklearn.ensemble^{5]}のRandom Forest Regressor, パラメータチューニングのためのoptuna^{6]}libraryに加えて、部分依存性はlibraryが無いことから独自にコーディングした。

3. データセットの加工

本論文では、データベース SSDSE-2020A、平成 27 年国勢調査による平成 22 年から 27 年の間の人口増加数^{7]}を e-stat より入手、外国人の増減数については、同じく e-stat で平成 28 年から平成 29 年の 1 年間での増加数^{8]}を用い、地域コードをキーとして、python プログラム上で統合、人口増加数を目的変数として、1720 の市区町村について解析した。なお、福島県富岡町、大熊町、双葉町、浪江町は人口データ欠損のため除外、福島県葛尾村、飯舘村は人口 12 人以下のため除外している。SSDSE-2020A は平成 26 年から 30 年の期間のデータであり、人口増加時期、外国人増加数はほぼ同時期と考え、各市区町村での人口増加の要因を解析した。1 つでも欠損がある場合には、プログラムで当該市区町村を削除している。

表 1 解析で用いた説明変数

総人口 (男)	婚姻件数	従業者数 (情報通信業)	小学校教員数	地域コード
総人口 (女)	離婚件数	従業者数 (運輸業、郵便業)	小学校児童数	市区町村
日本人人口 (男)	総面積 (北方地域及び竹島を除く)	従業者数 (卸売業、小売業)	中学校数	人口 平成27年
日本人人口 (女)	可住地面積	従業者数 (金融業、保険業)	中学校教員数	人口 平成22年
15歳未満人口 (男)	事業所数	従業者数 (不動産業、物品賃貸業)	中学校生徒数	平成22年～27年の人口増減数
15歳未満人口 (女)	事業所数 (農業、林業)	従業者数 (学術研究、専門・技術サービス業)	高等学校数	平成22年～27年の人口増減率 (%)
15～64歳人口 (男)	事業所数 (建設業)	従業者数 (宿泊業、飲食サービス業)	高等学校生徒数	面積 (km2)
15～64歳人口 (女)	事業所数 (製造業)	従業者数 (生活関連サービス業、娯楽業)	就業人数 (男)	人口密度 (1km2当たり)
65歳以上人口 (男)	事業所数 (電気・ガス・熱供給・水道業)	従業者数 (教育、学習支援業)	就業人数 (女)	外国人増減
65歳以上人口 (女)	事業所数 (情報通信業)	従業者数 (医療、福祉)	完全失業者数 (男)	
75歳以上人口 (男)	事業所数 (運輸業、郵便業)	従業者数 (複合サービス事業)	完全失業者数 (女)	
75歳以上人口 (女)	事業所数 (卸売業、小売業)	従業者数 (サービス業 (他に分類されないもの))	非労働力人口	
外国人人口	事業所数 (金融業、保険業)	従業者数 (公務 (他に分類されるものを除く))	非労働力人口 (男)	
出生数	事業所数 (不動産業、物品賃貸業)	第 1 次産業従業者数	非労働力人口 (女)	
死亡数	事業所数 (学術研究、専門・技術サービス業)	第 2 次産業従業者数	第 1 次産業就業人数	
転入者数 (日本人移動者)	事業所数 (宿泊業、飲食サービス業)	第 3 次産業従業者数	第 2 次産業就業人数	
人口増加	事業所数 (生活関連サービス業、娯楽業)	農家数 (販売農家)	第 3 次産業就業人数	
人口増加比	事業所数 (教育、学習支援業)	農家数 (自給的農家)	公民館数	
転出者数 (日本人移動者)	事業所数 (医療、福祉)	経常収支比率 (市町村財政)	図書館数	
世帯数	事業所数 (複合サービス事業)	実質公債費比率 (市町村財政)	総人口 (非水洗面人口 + 水洗面人口)	
一般世帯数	事業所数 (サービス業 (他に分類されないもの))	成人決算総額 (市町村財政)	非水洗面人口	
一般世帯員数	事業所数 (公務 (他に分類されるものを除く))	地方税 (市町村財政)	小売店数	
核家族世帯数	第 1 次産業事業所数	歳出決算総額 (市町村財政)	飲食店数	
単独世帯数	第 2 次産業事業所数	民生費 (市町村財政)	大型小売店数	
65歳以上の世帯員がいる核家族世帯数	第 3 次産業事業所数	土木費 (市町村財政)	一般病院数	
高齢夫婦のみの世帯数	従業者数	教育費 (市町村財政)	一般診療所数	
高齢単身世帯数 (65歳以上の者 1人)	従業者数 (農業、林業)	災害復旧費 (市町村財政)	歯科診療所数	
	従業者数 (建設業)	幼稚園数	医師数	
	従業者数 (製造業)	幼稚園在園者数	保育所等数	
	従業者数 (電気・ガス・熱供給・水道業)	小学校数	保育所等在所児数	

4. データ分析の結果

まず、全体間を知るために、都道府県単位で平成 27 年国勢調査による平成 22 年から 27 年の間の人口増加数を調べた。都道府県単位では東京周辺、愛知県、福岡県、宮城県などで人口増加、北海道、福島県などで人口減少が認められる。(図 1) この結果は、張らの解析結果^{1]}とも一致する。

機械学習の誤差を確認するため、75%の学習データと 25%のテストデータにわけて、人口増加を実際の値と予測値で相関を調べた。図 2 に示すように、本モデルで概ね実測の人口増加数を説明できていると考える。決定係数は市区町村全体の決定係数 r^2 は、学習データ 0.93、テストデータで 0.55 となった。若干過学習気味ではあるが、相関関係が得られていることがわかる。

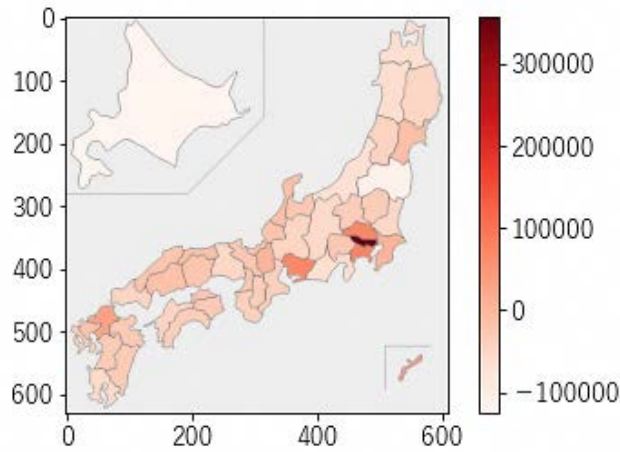


図1 都道府県単位の人口増加数（H22-27年）

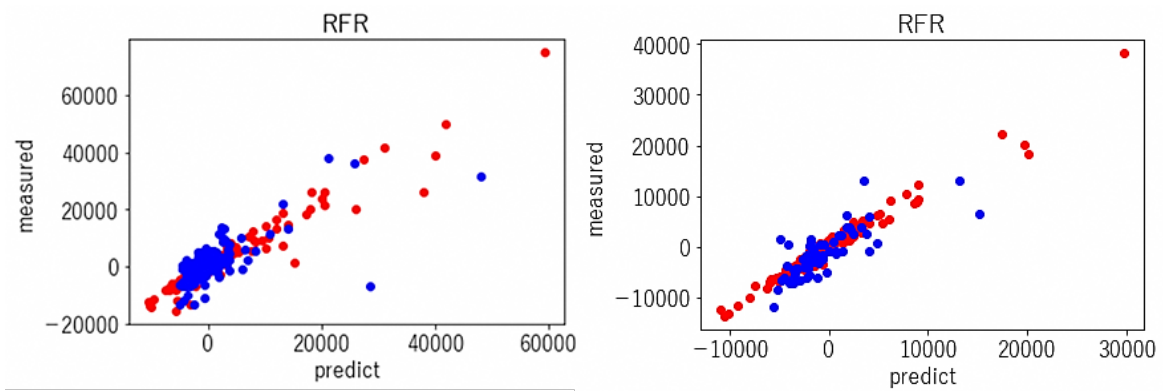


図2 人口増加数の実測値と予測値比較（赤プロット：学習データ、青プロット：テストデータ）左は全市区町村（max_depth:109、n_estimators:147、決定係数 r^2 学習データ 0.93、テストデータ 0.55、右は平成27年の人口が10万人を超え30万人未満の市区町村（max_depth:122、n_estimators:67、決定係数 r^2 学習データ 0.95、テストデータ 0.58）

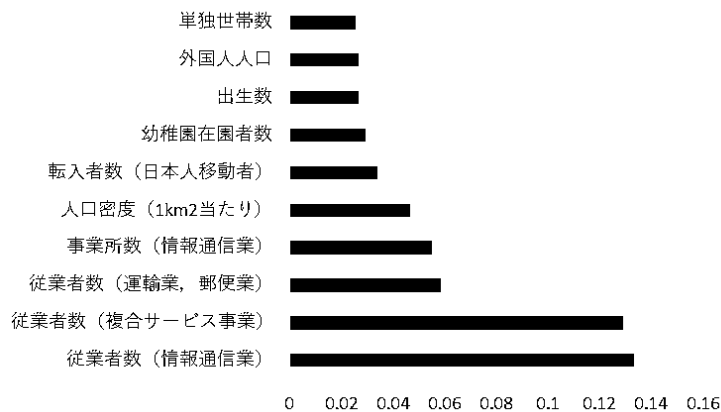


図4 重要度の大きい変数（全1720の市区町村）

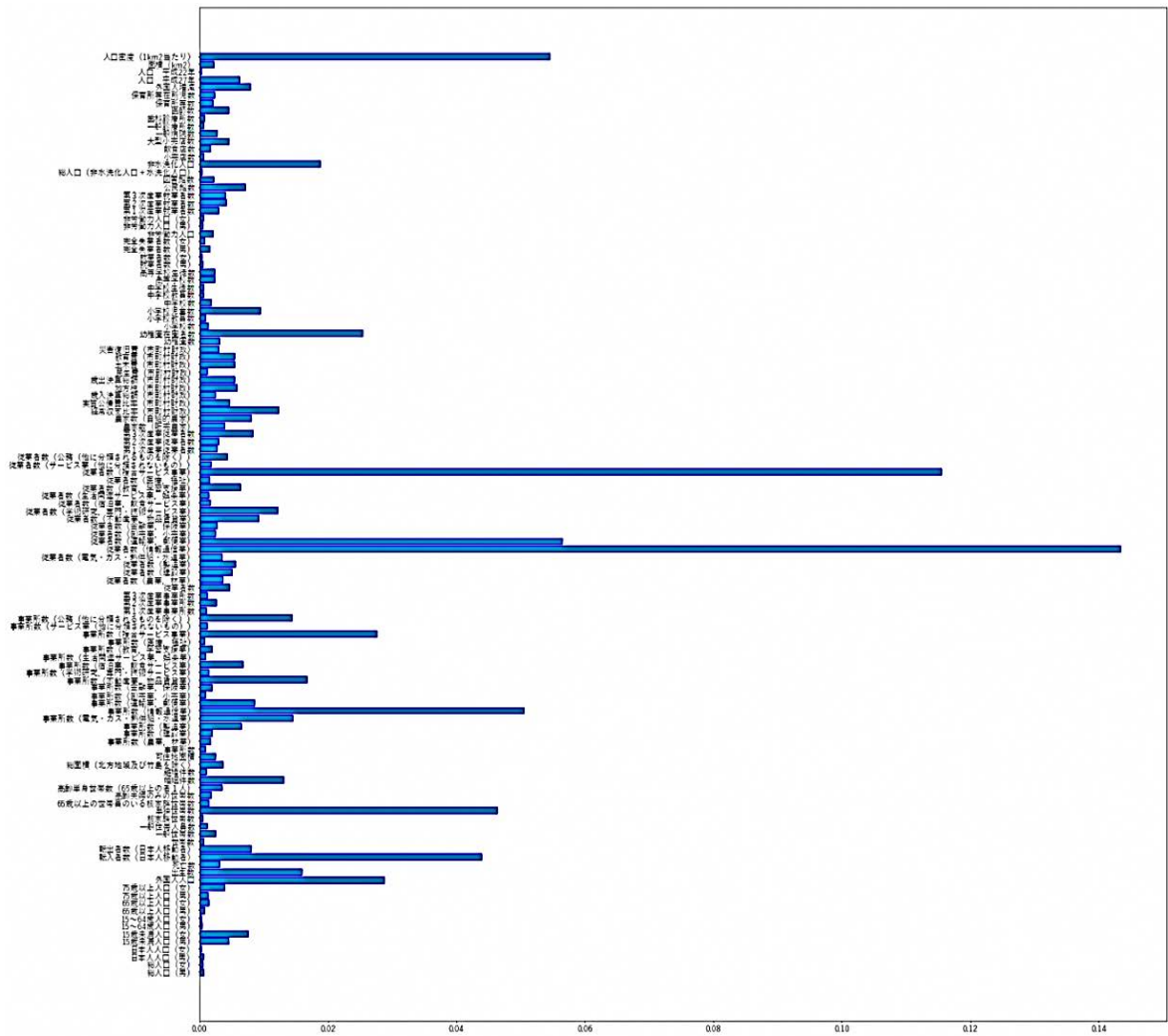


図3 全変数の重要度 (全 1720 の市区町村)

図3に全市区町村における説明変数の重要度を示す。図4には、そのうちの重要度の大きかった10の説明変数の重要度を示した。重要度の大きい変数として、大きい順に、従業員数(情報通信業)、従業員数(複合サービス事業)、人口密度(1m²当たり)、従業員数(運輸、郵便業)、事業者数(情報通信業)、単身世帯数、転入者数(日本人移住者)、外国人人口となった。

重要度の大きい説明変数の部分依存性を計算した。、図5に示すように、第1, 2, 3次産業人口にはほとんど依存せず、また人口増加は総人口にもほとんど依存せず、図6に示すように、従業員数(情報通信業)、従業員数(複合サービス事業)、人口密度(1m²当たり)、従業員数(運輸、郵便業)、事業者数(情報通信業)などの変数で人口増加が説明されることが明らかとなった。

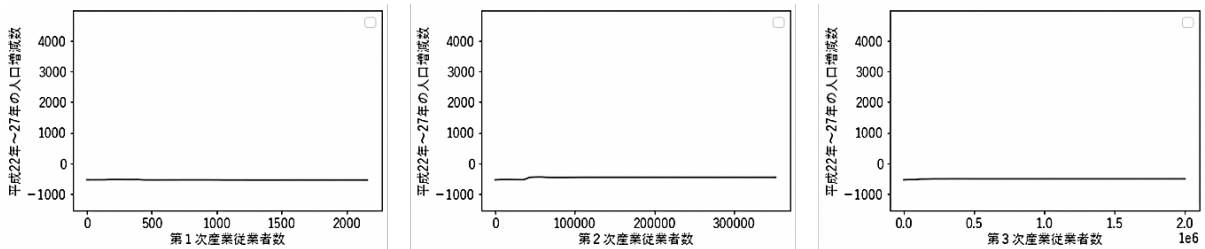


図5 産業別従業員数の影響

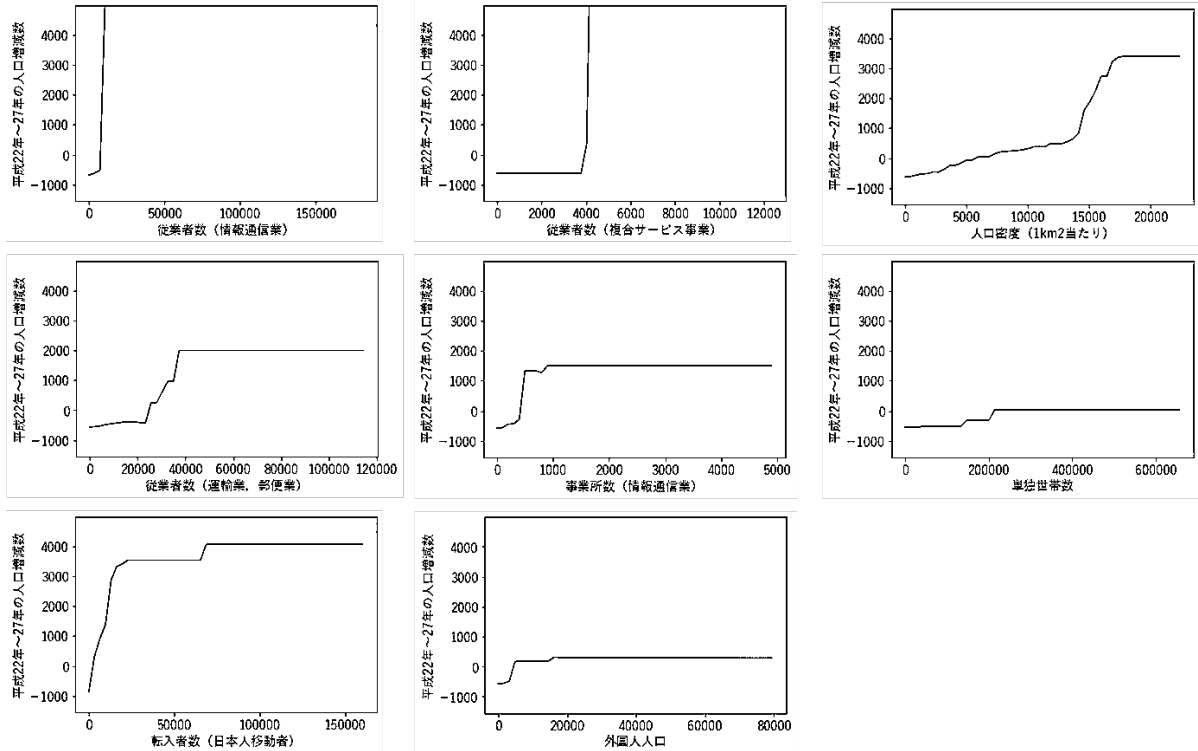


図6 重要度の大きい変数の部分依存性

表2に実際の人口増加数が大きかった地域の90位までの順位を示す。東京都とその周辺都市であることがわかる。このように、人口増加の大きかったこのような大都市・その周辺が全体の要因に大きく影響を与えていることが示唆されたため、次に、平成27年の人口規模ごとに重要度の解析を実施した。

表2 平成22年～27年で人口増加の多かった地区
(本表の分類は平成27年国勢調査による平成22年から27年間の人口増加数⁷⁾ベース)

市区町村	平成22年～27年の人口増減数	市区町村	平成22年～27年の人口増減数	市区町村	平成22年～27年の人口増減数
市部	357594	川口市	16606	目黒区	9292
東京都	355854	福岡市 博多区	15914	沖縄市	9030
神奈川県	77912	横浜市 港北区	14701	荒川区	8968
福岡市	74938	杉並区	14428	広島市 安佐南区	8779
愛知県	72409	大阪市 中央区	14382	神戸市 中央区	8760
埼玉県	71978	福岡市 中央区	14259	福岡市 南区	8701
川崎市	49701	藤沢市	14237	岡崎市	8694
さいたま市	41545	船橋市	13850	墨田区	8668
沖縄県	40748	福岡市 東区	13816	川越市	8075
札幌市	38811	川崎市 中原区	13604	大阪市 浪速区	8021
港区	38152	福岡市 西区	13588	いわき市	7988
江東区	37290	中野区	13465	市川市	7813
仙台市	36173	大阪市 北区	13275	黒川郡	7711
横浜市	36071	横浜市 鶴見区	13178	新宿区	7251
名古屋市	31744	文京区	13098	宇都宮市	6855
市部	31201	戸田市	13071	岡山市 北区	6799
福岡県	29588	つくば市	12373	川崎市 宮前区	6727
世田谷区	26208	名古屋市長区	12230	川崎市 幸区	6678
板橋区	26092	千代田区	11291	朝霞市	6608
大阪市	25871	越谷市	11185	札幌市 北区	6540
大田区	23709	糟屋郡	11018	札幌市 豊平区	6534
台東区	22145	川崎市 高津区	10781	豊島区	6489
品川区	21553	横浜市 都筑区	10480	さいたま市 緑区	6404
広島市	20191	流山市	10389	千葉県	6377
渋谷区	20041	千葉市	10133	草津市	6373
仙台市 青葉区	18747	柏市	9942	熊本市	6348
吹田市	18670	岡山市	9890	福岡市 早良区	6324
中央区	18421	さいたま市 浦和区	9630	仙台市 太白区	6267
札幌市 中央区	17438	大阪市 西区	9372	日野市	6231

図7に、10万人越え30万人未満の197市区町村における重要度を示す。重要度の大きい順に、転入者数（日本人移動者）、就業者数（学術研究・専門・技術サービス業）、非水洗化人口が挙げられた。図8に部分依存性を示す。人口10万人～30万人の都市では、転入者数（日本人）の増加が必要であるが、事業者数（学術研究など）の相関や比水洗化人口で逆相関傾向がみられ、全体の解析で見られた傾向とは大きく異なっている。

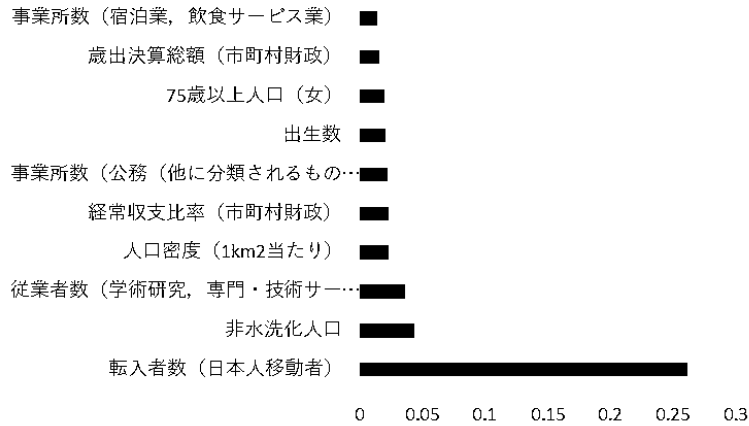


図7 重要度の比較（人口10万人越え30万人未満の197の市区町村）

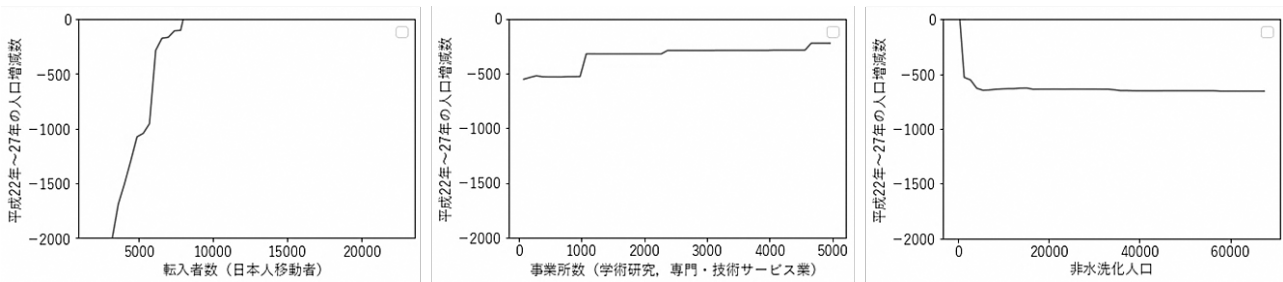


図8 部分依存性（人口10万人越え30万人未満の197の市区町村）

表3 人口増加の大きかった人口10万人越え30万人未満の市区町村

市区町村	人口増減数	市区町村	人口増減数	市区町村	人口増減数
港区	38152	沖繩市	9030	札幌市 東区	6039
台東区	22145	荒川区	8968	武蔵野市	5996
渋谷区	20041	広島市 安佐南区	8779	千葉市 中央区	5706
中央区	18421	神戸市 中央区	8760	仙台市 泉区	5615
札幌市 中央区	17438	福岡市 南区	8701	川崎市 麻生区	5597
福岡市 博多区	15914	墨田区	8668	横浜市 神奈川区	5537
川崎市 中原区	13604	川崎市 宮前区	6727	調布市	5468
福岡市 西区	13588	川崎市 幸区	6678	島尻郡	5452
大阪市 北区	13275	朝霞市	6608	安城市	5449
横浜市 鶴見区	13178	札幌市 北区	6540	さいたま市 大宮区	5376
文京区	13098	札幌市 豊平区	6534	札幌市 白石区	5325
戸田市	13071	豊島区	6489	ふじみ野市	5275
つくば市	12373	さいたま市 緑区	6404	茨木市	5211
名古屋市 緑区	12230	豊津市	6373	熊本市 南区	5169
糟屋郡	11018	福岡市 早良区	6324	さいたま市 南区	5164
川崎市 高津区	10781	仙台市 太白区	6267	三郷市	5106
横浜市 都筑区	10480	日野市	6231	千葉市 緑区	4927
流山市	10389	広島市 中区	6158	木更津市	4829
目黒区	9292	川崎市 川崎区	6050	さいたま市 見沼区	4817

表3に、人口増加が大きかった人口が10万人越え30万人未満の市区町村を示す。大都市の区部のほか、戸田市、つくば市、糟屋郡、流山市などの中堅都市がランク入りしている。ここで、つくば市は学術研究都市として有名であり、また、糟屋郡の人口増加要因を調べたところ、博多駅から最短2駅というアクセスの良さが人口増加につながっていることがわかった⁹⁾。

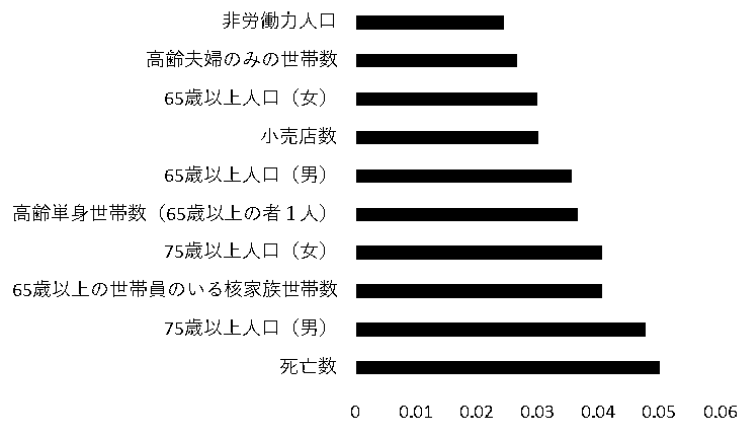


図9 重要度の大きい変数（人口1万人未満）

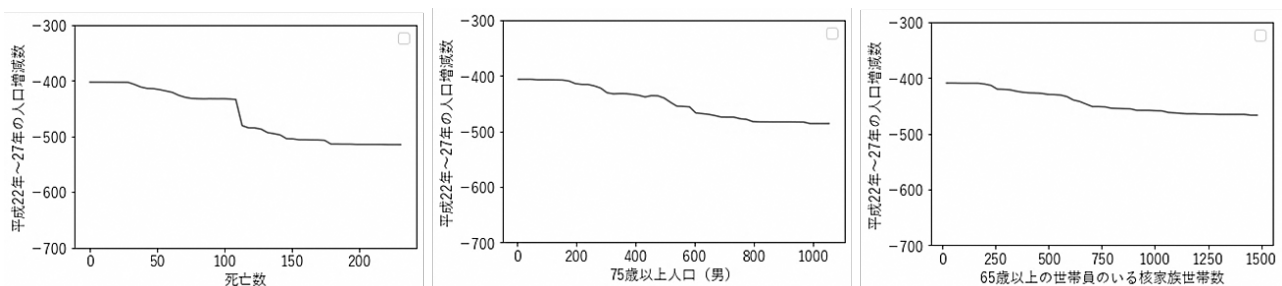


図10 部分依存性（人口1万人未満）

次に、人口1万人未満の市区町村について、同様の解析を行った。重要度を図9に、図10に部分依存性を示す。これまでの結果と大きく異なり、人口1万人未満の市区町村では、死亡数による人口減少が人口変化の主要因であることが明らかとなった。

5. 結果の解釈

図4、図5の結果から、平成22年から平成27年の間、情報、運輸、複合サービスなど、成長分野の従業者数の多い市区町村で人口増加が顕著となっていた。この傾向は、第1、2、3次産業という大ぐくりな分類では説明できず、張ら¹⁾が指摘している歳出決算額も必ずしも大きな影響因子ではない結果となった。また、人口密度が大きい市区町村で人口増加数が多いことから、都市化と過疎化が進んでいたこと、また、外国人増加数も、全体の人口増加には、未だそれほど大きい影響を与えていないことも明らかとなった。すなわち、中村良平¹⁰⁾の提案している「稼ぐ力」により、人口増加は、情報、複合サービスといった成長産業分野企業（従業者数）の貢献が大きかったことが示唆される。

また、人口 10 万人以上 30 万人未満の都市では、転入者数の影響をもっとも受けており、糟屋郡⁹⁾の例からもわかるように、周辺から地方都市への流入が起こっている可能性も考えられ、経済地理学的な検討も必要であると思われる。また、学術研究などの事業者数の依存性も高いことが大変興味深く、他の人口規模とは異なる発展の形があることが示唆された。

一方で、人口 1 万人未満の市区町村では、死亡が主要因で人口減少が起こっており、このような状況は、総務省が平成 14 年～19 年の詳細データを解析・整理¹¹⁾しているが、本解析の平成 22 年から平成 27 年の期間でも同様の現象が起こっているものと思われる。

以上をまとめると、本解析により、各市区町村の規模に応じて人口増加の要因が大きく異なることが明らかとなった。すなわち、市区町村全体では、成長分野である情報、複合サービス業、運輸など、成長の著しい産業を発展させることが人口増加には効果的ではあるが、中規模の市区町村でも、その規模に応じた発展の形があることが示唆された。岡田は、人口増減について年代別の詳しい解析¹²⁾を実施しているが、本論文では、市区町村の規模ごとの解析の必要性が示された。また、地理的な観点からの検討も必要であり、都市間の人口移動形態^{13),14)}なども、市区町村の人口増減の解明には有効であると思われる。

参考文献

[1]張 瀚天、白鳥 友風：地方創生目標指標に関する変化要因ネットワークの推定とそれに基づく地域間連携の提案、(2019) 統計データ分析コンペティション

[2]竹内太郎：我が国における人口増減の決定要因、(2019) 統計データ分析コンペティション

[3] 例えば、内閣府：人口をめぐる現状と課題、<https://www5.cao.go.jp/keizaishimon/kaigi/special/future/sentak/pdf/p030108>.

[4] Leo Breiman：“Random Forests”，Machine Learning 45(2001)5-32.

[5] <https://scikit-learn.org/stable/>

[6] <https://optuna.org/>

[7]<https://www.e-stat.go.jp/statsearch/files?page=1&toukei=00200521&tstat=000001080615>

[8] <https://www.e-stat.go.jp/stat-search/database?page=1&toukei=00250012&tstat=000001018034>

[9] 粕屋町：粕屋町人口ビジョン（案），(2015, 12)

<https://www.town.kasuya.fukuoka.jp/s005/010/020/020/040/H2712jinkovison.pdf>

[10] 中村良平：第 4 回まちの稼ぐ力、(2018)，<https://www.rieti.go.jp/jp/papers/contribution/hyogo-jichi/04.html>、あるいは、『ひょうご自治』(2017. 08)

[11] 総務省：過疎地域の現状、これまでの過疎対策、過疎地域の課題等について（各種データ）

https://www.soumu.go.jp/main_sosiki/jichi_gyousei/c-gyousei/2001/kaso/pdf/kasokon19_03_s5.pdf

[12] 岡田豊：二極化する地域別人口と人口減少都市の在り方、みずほ総研、(2014)2, 31-47.

[13] 国土交通省：地域間の人口移動について

<https://www.mlit.go.jp/singikai/kokudosin/keikaku/lifestyle/9/07.pdf>

[14] 登石文夫：都市間人口移動と人口移動モデル、社会学評論、20(1969-1970), 2-20.