

通勤距離を加味した居住地選択モデルの構築

～北海道札幌市の仮想個票を事例に～

○太田奎祐(静岡大学大学院) 李皓(静岡大学)

Development of a model for selecting a place to live that takes into account commuting distance

* K.Ota(Shizuoka University Graduate School) and H.Lee (University of Shizuoka)

概要— 我が国では人口減少が進み、人口推定の研究が行われている中、立地選択モデルを組み込んだマイクロシミュレーションを用い、将来人口予測が行われているが、居住地選択において重要とされている通勤距離を加味したモデルはない。そこで本研究では通勤距離を考慮したモデルを構築する。世帯属性から世帯が求める地域属性を推定し、それを元に居住地選択を行う手法を開発。北海道札幌市の仮想個票データを事例に居住地選択の予測を行った。予測結果は地域属性パラメータ精度の問題で、居住地選択では5000-8000m程度の大きな誤差を生じる結果となった。しかし実績値を用いた居住地選択では精度の高い予測ができ、提案手法の有効性が示唆された。

キーワード: 居住地選択モデル, 通勤距離, 仮想個票, 新規手法, 札幌市

1 はじめに

我が国では、2008年を境に人口減少が現在まで進行し、内閣府の人口推計によると2030年には1億1651万人、2048年には1億人を切る。また地方では人口減少に加え、若年層の流出が深刻となっている。

このような状況下で人口減少時代に対応した政策検討を行う為、人口推定に関する研究が各方面で行われているが、詳細な人口分布を予測できるマイクロシミュレーションに世帯や企業の立地選択モデルを組み込んだ研究もある。しかし既存の居住地選択モデルには職場までの距離が加味されていない。

そこで本研究では、原田ほか(2018)の札幌市仮想個票を事例に通勤距離を加味した居住地選択モデルを提案し、構築と妥当性評価を行う。

札幌市は北海道で一番人口の多い都市であり、道内で札幌一極集中が進んでいる。しかし将来推計人口は札幌市人口ビジョンによると、2010年に191万人いた札幌市の人口が、2035年には175万人に減少。更に2060年に143万人と日本の五大都市圏の一つでありながら、人口減少と若年層流出が進んでいる地域である為、人口減少時代に対応した施策が必要である。また札幌市は電車社会でありながらもFig. 1の清田区のように鉄道のない地域もあり、車社会の側面も強い地域でもある。本研究では札幌市全5,980町丁目から居住地選択するモデルを構築する。

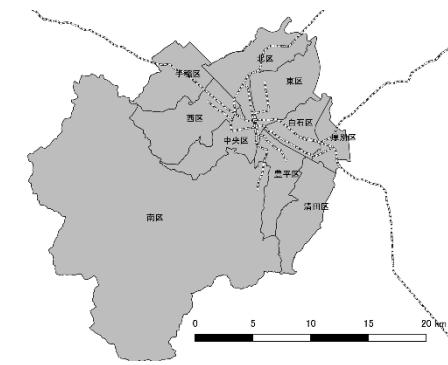


Fig. 1: 札幌市概況

2 先行研究と本研究の位置づけ

2.1 先行研究

人口推定分野において代表的なものとして、国立社会保障・人口問題研究所の人口推計がある。この人口推定では国勢調査による人口を基礎として、出生・死亡、転出入等の人口動向から各年の人口を、コーホート要因法を使って推計している。またコーホート変化率を用いた推計として井上(2017)がある。更に細かく人口の将来分布を予測する為、鈴木ほか(2016)のようなマイクロシミュレーションを使った研究がある。この研究では人口40万人規模の富山市全域を対象とし、将来人口分布を詳細に予測できるマイクロシミュレーションモデルを構築している。

立地選択モデルの研究として、まずWaddell(2002)がある。マイクロシミュレーションにマクロ経済モデルや交通モデル、土地価格モデルなどを組み込んだ上で企業や世帯がある地区に対する支払い意思額から選択確率を求める立地選択モデルを組み込んでいる。

またMargo(1992)では1950年代のアメリカにおいて、所得や人種などの要素から、都心に住むか郊外に住むかという居住地選択モデルを構築し、推定を行っていた。小松ほか(2007)ではMargo(1992)のモデルを採用し、国内の政令指定都市を除く地方都市圏を対象に、世帯属性と地域属性から郊外に住むか都心に住むかというモデルを構築。また桑野(2019)のように移住相談の内容を分析し、鳥取市に住むか鳥取市外に住むかという居住地選択モデル開発した研究もある。

また居住地選択モデルを組み込む際に必要となる地域属性に関する研究としてまず相(2014)があり、東京と周辺3県の人口増減パターンを分析、その結果最寄り駅、病院までの距離、町丁目内の店舗数と事業所数、都市公園の有無、人口密度、平均世帯人員、など人口増減パターンと有意な関係性を示していた。また西山ほか(2011)では、宇都宮都市圏を対象にアンケートを行い、家の広さ、スーパーまでの距離、駅までの距離、価格・家賃について効用が見られた事を明らかにしていた。清水ほか(2017)では奈良市を対象にアンケート調査から、居住地選択において、すべての回答者が住宅価格・家賃、通勤先への距離を重視している事を明らかにしていた。

先行研究では、清水ほか(2017)から人々が居住地を選ぶ際に職場までの距離が関係している事が明らかになっているが、既存の居住地選択モデルには通勤距離を考慮した研究がない。また既存研究には居住地選択モデルによる予測後、予測モデルの精度を確かめる評価が行われていない。

2.2 目的と独自性

清水ほか(2017)のように居住地選択には職場までの距離が考慮される事が明らかにされているが、既存の居住地選択モデルには通勤距離を加味したモデルがない。

そこで本研究では、北海道札幌市の仮想個票データを事例に、通勤距離を加味した居住地選択モデルの構築と推定を行う。

独自性として以下の3点がある。まず原田ほか(2018)の仮想個票データを用い、大量のデータに基づいたモデルを構築して、予測後にモデルの妥当性評価を行う。2つ目に通勤距離を採用する。3つ目に新たな手法を提案し推定を行う事にある。居住地選択には主に二つの手法があり、離散選択モデルを使い選択確率を求める手法やロジットモデルで都心に住むか否か2択のモデルがある。しかし本研究ではそれらの手法ではなく、新たな手法を提案する。

3 手法

3.1 提案手法

本研究の新たな手法として、所得などの世帯属性から、各世帯が求める通勤距離や小売店などの地域属性パラメータを推定し、そこから居住地を選択するというモデルを構築した。

世帯が求める地域属性パラメータ推定には、重回帰、ランダムフォレスト、ディープラーニングの3つの回帰手法を用いる。この3つの手法は複数の目的変数を同時に求められる。算出にはTable 1のようにScikit-learnのライブラリを採用した。

説明変数(世帯属性)は所得、世帯主の年齢、子供数、家族数、独身か否か、目的変数(世帯が求める地域属性)は通勤距離、土地地価、(飲食料品)小売店、病院、小中学校までの最短距離。

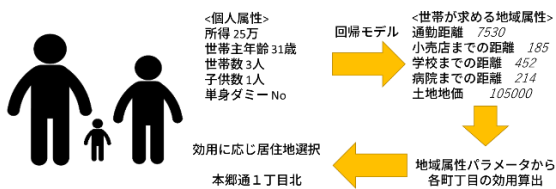


Fig. 2: 本研究の手法のイメージ 居住地選択

Table 1: 使用ライブラリー一覧

重回帰分析	multioutput.MultiOutputRegressor
ランダムフォレスト	ensemble.RandomForestRegressor
ディープラーニング	neural_network.MLPRegressor

3.2 居住地選択の手法

3-1で推定した世帯が求める地域属性パラメータを用いて、居住地選択の予測モデルを作成する。本研究では2種類の効用算出方法を提案する。居住地選択の候補は札幌市全5,980町丁目である。

<1つ目の手法>

まず世帯*i*が町丁目*j*で得られる効用 $DIFF_{ij}$ の計算を式(3)で行う。この式では $A_{i,p}$ は世帯*i*の地域属性*p*に対する希望であり、 $A_{j,p}$ は地域*j*の地域属性*p*の値である。 P は地域属性の集合。つまり各地域属性、希望値と実績値の差分を総和する。次に式(4)では $DIFF_{ij}$ で1を割り、世帯*i*の町丁目*j*に対する効用 U_{ij} を算出する。これらの式を使い、各世帯の5,980町丁目に対する効用を全て算出する。

$$DIFF_{ij} = \sum_{p \in P} |A_{j,p} - A_{i,p}| \quad (3)$$

$$U_{ij} = 1 \div DIFF_{ij} \quad (4)$$

$DIFF_{ij}$: 世帯*i*が求める地域属性と地域*j*の地域属性の差の和

$A_{i,p}$: 世帯*i*が求める地域属性*p*の値

$A_{j,p}$: 地域*j*の地域属性*p*の値

P : 地域属性の集合

U_{ij} : 世帯*i*の地域*j*に対する効用

効用を算出したら、それを元に居住地選択を行う。本研究では居住地選択の方法として以下の三種類を想定した。居住地選択には一番欲しい場所に物件や土地がないなど、現実社会では希望の場所に居住地選択を選択できない可能性を考慮し、②や③のような選択肢も採用した。

- ① 一番効用の高い場所を選ぶ
- ② 上位10からランダムに選ぶ
- ③ 上位100からランダムに選ぶ

<2つ目の手法>

2つ目の手法は効用計算に違いがあり、流れは1つ目の手法と同じである。数式の記号が示すものは手法1の式と同じで、まず式(5)で世帯の希望 $A_{i,p}$ で実際の値 $A_{j,p}$ を割り、割った値に1を引き、比を計算する。次に式(6)で出た $RATIO_{ij}$ で1を割り、世帯*i*の町丁目*j*に対する効用 U_{ij} となる。居住地選択の選択肢は1つ目の手法と同じである。

$$RATIO_{ij} = \sum_{p \in P} \left(A_{j,p} \div A_{i,p} \right) - 1 \quad (5)$$

$$U_{ij} = 1 \div RATIO_{ij} \quad (6)$$

$RATIO_{ij}$: 世帯*i*が求める地域属性と地域*j*の地域属性の比の和

$A_{i,p}$: 世帯*i*が求める地域属性*p*の値

$A_{j,p}$: 地域*j*の地域属性*p*の値

P : 地域属性の集合

U_{ij} : 世帯*i*の地域*j*に対する効用

このように全世帯の各町丁目に対する効用を算出。居住地選択モデルの評価方法は世帯が選んだ居住地と実際に住んでいる居住地の差を直線距離で算出し、誤差でモデルの妥当性評価を行う。

3.3 仮想個票データについて

本研究では原田ほか(2018)の仮想個票データを用いる。この合成データは、5年ごとに実施される国勢調査の結果や国の統計、都道府県の統計、市区町村の統計、町丁目の統計から個票データを復元したもので、現実に近い仮想個票データベースを構築している。格納されている属性はTable 2である。

今回居住地選択をする際は世帯主の家族数や所得などの属性に合わせ選択すると仮定した。しかし仮想個票データには世帯主を示すフラグが存在しない為、各世帯の中で一番所得が高い人を世帯主と仮定した。札幌市世帯数は495,943世帯である。

Table 2: 原田ほか合成人口データ(筆者作成)

	村田, 原田の合成データ
市民	市民番号, 世帯番号, 年齢, 性別, 世帯内の役割, 収入, 産業タイプ, 所属企業の規模, 通勤先
世帯	世帯番号, 世帯類型, 世帯人数, 建物番号, 町丁目コード
建物	建物番号, 緯度, 経度, 建物の構造, 町丁目コード, 建築面積

3.4 距離データの算出方法

仮想個票データには通勤先が組み込まれている。通勤距離の算出には仮想個票データと統計GISの境界データに付随している経度緯度を組み合わせ、経路検索の2種類を算出した。

通勤距離は距離計算フリーライセンスである Open Source Routing Machine (OSRM) を採用し、経路検索を行った。OSRMは緯度経度に基づく、計算を行う事ができ、電車や車、徒歩による経路距離の計算ができる。ただ本研究で使用した合成人口データは交通手段の推定は行っていない為、全世帯車移動を仮定した。採用理由は本研究のモデルは距離を採用しており、通勤時間ではないため、車の経路は徒歩と大きな差がなく、また中央区を除き札幌市は車社会の側面も強い為、本研究は車移動による経路検索を行った。

495,943 サンプルに対しては、所要時間は約50分である。例えばFig.3のように南24条西8丁目に家があり、宮の森3条10丁目に通勤している世帯があったとする。経路検索結果は7,584.1mとなる。このような計算を全世帯で行った。Fig.4は札幌市全世帯の計算結果で、通勤距離に5km前後の世帯が最も多い。



Fig.3: ある世帯の通勤距離計算結果

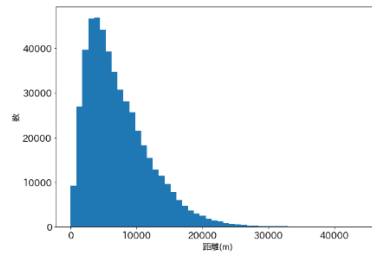


Fig.4: 札幌市仮想個票データ全世帯通勤距離の結果
店や病院、学校までの距離の算出は、境界データに各町丁目の経度緯度が含まれている事から、各データに付随している住所と境界データの経度緯度を結合し算出した。計算リソースの制約で経路検索ではなく、GeoPyを用いて直線距離で算出。ある世帯の札幌市内にある店、学校、病院までの距離をそれぞれ計算し、最短距離を各施設までの距離とした。Table 3はある世帯の各施設までの距離の計算結果で、0行目の一番短い距離を個票データに格納している。また使用した経度緯度の境界データは個票データでも同じく使われており、居住地と施設の町丁目の経度緯度が同じである為、同じ町丁目に施設がある場合は距離が0mとなっている。

Table 3: ある世帯の各施設までの距離一覧
(左:店 中:病院 右:学校)

index	distance	index	distance	index	distance
0	315 128.595584	0	52 938.701658	0	23 422.455522
1	298 169.397647	1	54 950.662524	1	9 438.329571
2	316 233.588393	2	48 956.929369	2	55 1135.251577
3	297 266.160015	3	45 1014.039052	3	99 1172.370641
4	314 271.020114	4	56 1030.310892	4	94 1175.902730
...
1837	1366 19712.153758	307	23 14192.716112	307	270 15055.643050
1838	1363 19765.414519	308	119 14381.181608	308	286 15292.526803
1839	1364 20155.591117	309	253 14604.616761	309	207 18960.239424
1840	1367 20330.390385	310	120 14646.912757	310	196 19198.153378
1841	1361 23690.295236	311	254 15055.643050	311	190 20330.390385

Table 4: データの出典元一覧

データ名	データの出典
土地地価	令和3年度 北海道地価調査・地価公示
小売店	平成24年 経済センサス活動調査
学校(小中)	平成25年 国土数値情報
病院	平成25年 国土数値情報
経度緯度	平成23年 地図で見る統計(国勢調査)

4 推定結果

4.1 世帯が求める地域属性パラメータ推定の結果

数値予測のモデルの良さを測る指標の一つである平均二乗誤差(RMSE)で、モデルの誤差の比較を行った。RMSEの結果はTable 5である。それぞれみると通勤距離は僅差でランダムフォレスト、店や学校、病院までの距離は手法で大きな変化はなく、土地地価はランダムフォレストが一番良く、この中ではランダムフォレ

ストが最もよくパラメータ推定できたと示唆された。

Fig. 5からFig. 9は各推定値と実績値の差をまとめたものである。まず通勤距離は一人一人通勤距離が違うという事も考えられるが、差のばらつきが大きく、推定が上手くできていない事が考えられる。特に実際の通勤時間よりも少なく推定されている。小売店や病院までの距離は差が比較的小さく、パラメータ推定ができていたといえる。しかしこれはそもそも、比較的值が小さい事から、パラメータ推定が外れたとしても比較的小さな範囲で収まっているからという事も考えられる。学校までの距離や土地地価は小売店、病院と比べると、誤差が大きい事が分かる。またどの変数も実際の値よりも低めに推定されてしまっている傾向がある。

すべての地域属性においてばらつきがあり、特に学校や通勤距離は差が大きい。推定の精度が高いとはいえない結果となった。通勤時間は今回の元データで最もばらつきが大きい為、パラメータ推定する事が特に難しかった事が考えられる。また土地地価の安価な場所や高価な場所の差がかなり大きく、それによりパラメータ推定大きな誤差が生じる為、居住地選択に与える影響が大きい可能性がある。

本研究では、scikit-learnの初期設定で行った為、多くの計算リソースを投入しランダムフォレストやディープラーニングではパラメータ探索を行うと、パラメータ推定の結果が改善される可能性も考えられる。今回は計算リソースの制約によりこれらのパラメータを使い、居住地選択の推定を行う。

Table 5: RMSEの結果

	通勤距離	店	学校	病院	土地地価
重回帰	4766.2836	290.55196	301.4905	240.94572	194173.07
ランダムフォレスト	4754.3209	290.08426	301.09812	244.12724	191515.11
ディープラーニング	4818.2415	290.70445	302.22535	244.57136	200673.26

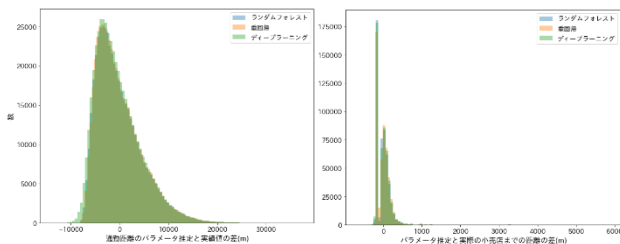


Fig. 5 通勤距離
実績値と推定の差

Fig. 6 小売店までの距離
実績値と推定の差

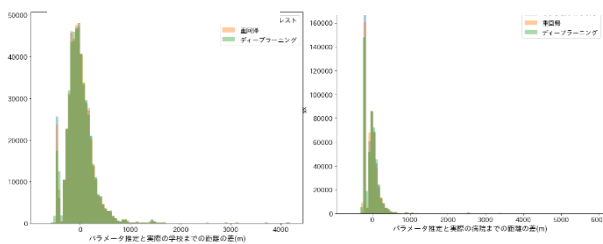


Fig. 7 学校までの距離
実績値と推定の差

Fig. 8 病院までの距離
実績値と推定の差

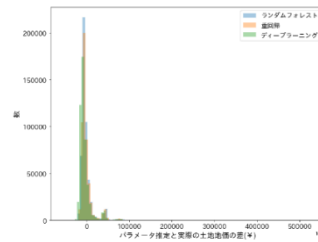


Fig. 9 土地地価 (x 軸は指数表示 1e1)
実績値と推定の差

4.2 居住地選択の結果

4.1で算出したパラメータで居住地選択モデルの推定。例えば二十四軒三条5丁目に住むある世帯を手法1, 回帰モデルはランダムフォレストで計算を行うと、Table 6のようにそれぞれの5,980町丁分の効用が算出される。この世帯が一番効用の高い選択肢①は「月寒西一条2丁目」を選んだこととなる。また選択肢②の上位10から選ぶ場合、今回は仮に3位を選ぶとすると「美園五条1丁目」となる。選択肢③の上位100から選ぶ場合は、仮に93位を選ぶとすると「北二十五条西4丁目」となる。

評価方法は、前述の通り、選んだ町丁目と実際に住んでいる町丁目の直線距離を算出し、その差で評価する。

例えばTable 6の世帯だと選択肢①「月寒西一条2丁目」は「7,760.486928794386m」、選択肢②「美園五条1丁目」は「7,200.677326408346m」、選択肢③「北二十五条西4丁目」は「3,600.918005457705m」となる。

Table 6: ランダムフォレスト手法1のある世帯の効用

Index	household_ID	evaluation_point	town	X_CODE	Y_CODE	town_real	X_CODE_real	Y_CODE_real
3678	656173	0.000902991	月寒西一条2丁目	141.391	43.0361	二十四軒三条5丁目	141.309	43.0717
3679	656173	0.00078496	月寒西一条3丁目	141.391	43.0348	二十四軒三条5丁目	141.309	43.0717
4132	656173	0.000672648	美園五条1丁目	141.381	43.0426	二十四軒三条5丁目	141.309	43.0717
4136	656173	0.000659273	美園五条5丁目	141.386	43.0396	二十四軒三条5丁目	141.309	43.0717
4133	656173	0.000645693	美園五条2丁目	141.382	43.0419	二十四軒三条5丁目	141.309	43.0717
...
193	656173	1.91769e-07	北西条西5丁目	141.348	43.0659	二十四軒三条5丁目	141.309	43.0717
196	656173	1.91768e-07	北西条西8丁目	141.344	43.0652	二十四軒三条5丁目	141.309	43.0717
199	656173	1.91766e-07	北西条西11丁目	141.339	43.0646	二十四軒三条5丁目	141.309	43.0717
194	656173	1.91764e-07	北西条西6丁目	141.347	43.0656	二十四軒三条5丁目	141.309	43.0717
195	656173	1.91759e-07	北西条西7丁目	141.345	43.0654	二十四軒三条5丁目	141.309	43.0717

計算時間の制約により、495943世帯の20%をランダム抽出し、99188世帯で推定を行った。予想結果はTable 7に誤差を合計したもの、Fig. 10からFig. 33で各予想の評価結果をグラフにし、提案モデルと通勤距離を除外したモデルを図表化した。

まずTable 7を見ると、どの手法も誤差の合計に大きな違いが小さいが、もっとも誤差が小さいのはディープラーニング選択肢3手法1, それに続きディープラーニング選択肢2手法1, ディープラーニング選択肢1手法1となっている。推定誤差が大きい通勤距離の結果を見ると、ディープラーニングは他手法に比べると誤差が小さく、その中でもディープラーニング選択肢2手法2, ディープラーニング選択肢2手法2, ディープ

ラーニング選択肢3手法1が良く、通勤距離を加味したモデルも加味していないモデルもどちらもディープラーニングが、一番誤差が小さいという結果となった。また誤差の合計を見る限り、通勤距離の有無で結果に大きな違いはないが、通勤距離を加味していないモデルの方が誤差がやや小さい。

次に全体の結果をまとめたFig. 10からFig. 33を見ていくと、全体的に距離の差が5000m-8000mの間でばらつき、1,000m以内で予測ができていないものがあるが、全体的には予測ができていないとは言えない。また全手法、通勤距離を省いたモデルの方が、誤差が小さくなる傾向があり、通勤距離を加味したモデルの誤差が大きかった。通勤距離を加味したモデルの予測結果が6,000m-8,000mに密集し、他モデルのように誤差が小さいものはないものの、逆に大きく外れているものが少なかった。

また選択肢の違いでは、選択肢1の一番住みたい場所に求める場合、結果が6,000m-8,000mに集まり、選択肢2で少し結果が分散し、選択肢3で更に分散している。特にディープラーニングでその結果が顕著に表れている。つまり一番住みたい場所を選ぶと6,000m-8,000mの誤差となり、選択肢2,3と徐々に検討範囲を広げていくと誤差のばらつきが広がる為、札幌市の面積を考えると、精度が良いとは言えない結果となった。

世帯が求める地域属性のパラメータ推定で大きな誤差が生じた結果、居住地選択の誤差につながった大きな原因である可能性がある。また通勤距離を加味すると予想結果が大きく外れたが、本研究の予測元のデータのばらつきが一番大きいのは通勤距離であり、パラメータ推定が上手くできなかった事により、通勤距離が一番大きく外れている事が考えられる。一方で本研究の重回帰モデルの中ではディープラーニングが一番良い結果となった。

Table 7: 居住地選択したエリアと実際に住んでいる居住地の距離の誤差の合計
(左: 通勤距離あり 右: 通勤距離なし)

重回帰モデル_手法	誤差の合計	重回帰モデル_手法	誤差の合計
重回帰分析_選択肢1_手法1	776,905,916	重回帰分析_選択肢1_手法1	775,247,108
重回帰分析_選択肢2_手法1	768,982,301	重回帰分析_選択肢2_手法1	767,254,747
重回帰分析_選択肢3_手法1	759,833,069	重回帰分析_選択肢3_手法1	763,040,934
重回帰分析_選択肢1_手法2	765,427,558	重回帰分析_選択肢1_手法2	758,125,771
重回帰分析_選択肢2_手法2	760,342,350	重回帰分析_選択肢2_手法2	752,344,747
重回帰分析_選択肢3_手法2	755,739,162	重回帰分析_選択肢3_手法2	749,971,626
ランダムフォレスト_選択肢1_手法1	772,971,566	ランダムフォレスト_選択肢1_手法1	765,635,739
ランダムフォレスト_選択肢2_手法1	763,844,313	ランダムフォレスト_選択肢2_手法1	757,608,954
ランダムフォレスト_選択肢3_手法1	755,394,695	ランダムフォレスト_選択肢3_手法1	757,953,996
ランダムフォレスト_選択肢1_手法2	763,525,041	ランダムフォレスト_選択肢1_手法2	752,767,348
ランダムフォレスト_選択肢2_手法2	759,805,636	ランダムフォレスト_選択肢2_手法2	752,589,884
ランダムフォレスト_選択肢3_手法2	755,164,052	ランダムフォレスト_選択肢3_手法2	747,551,282
ディープラーニング_選択肢1_手法1	734,246,695	ディープラーニング_選択肢1_手法1	736,406,519
ディープラーニング_選択肢2_手法1	725,046,114	ディープラーニング_選択肢2_手法1	725,785,106
ディープラーニング_選択肢3_手法1	720,983,022	ディープラーニング_選択肢3_手法1	727,128,962
ディープラーニング_選択肢1_手法2	756,273,219	ディープラーニング_選択肢1_手法2	733,612,753
ディープラーニング_選択肢2_手法2	746,465,554	ディープラーニング_選択肢2_手法2	722,838,415
ディープラーニング_選択肢3_手法2	744,845,550	ディープラーニング_選択肢3_手法2	727,786,454

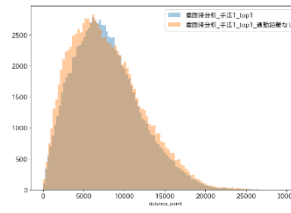


Fig.10 重回帰 手法1 選択1

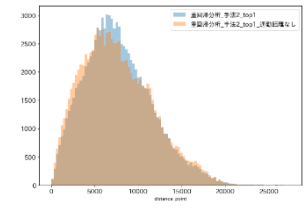


Fig.13 重回帰 手法2 選択1

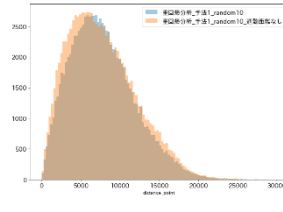


Fig.11 重回帰 手法1 選択1

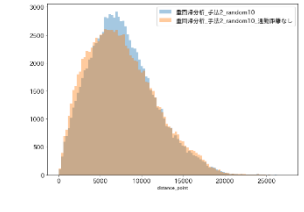


Fig.14 重回帰 手法2 選択2

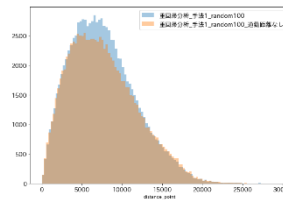


Fig.12 重回帰 手法1 選択3

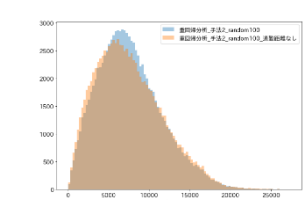


Fig.15 重回帰 手法2 選択3

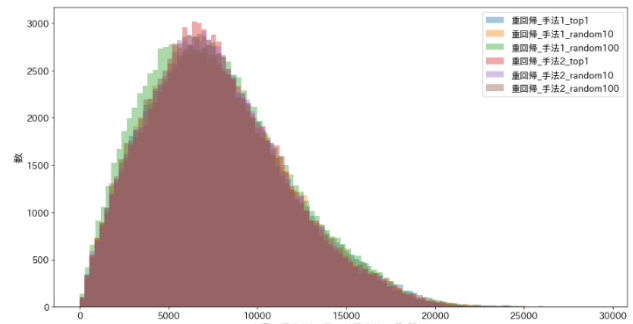


Fig. 16: 居住地重回帰分析_結果一覧

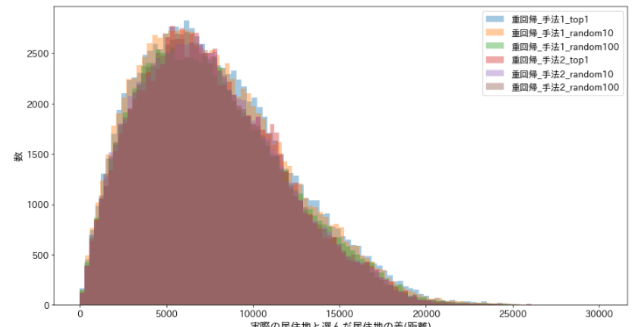


Fig. 17: 居住地重回帰分析_通勤距離なし_結果一覧

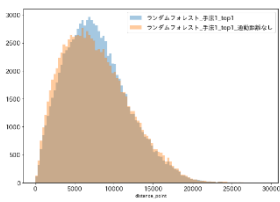


Fig.18 ランダムフォレスト
手法1 選択 1

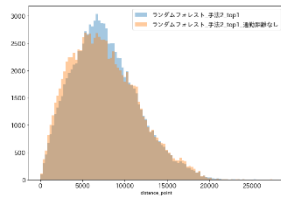


Fig.21 ランダムフォレスト
手法2 選択 1

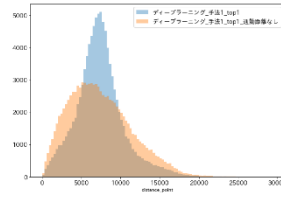


Fig.26 ディープラーニング
手法1 選択 1

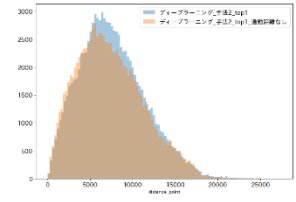


Fig.29 ディープラーニング
手法2 選択 1

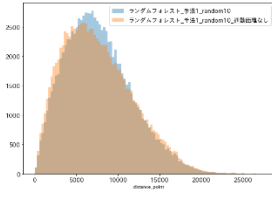


Fig.19 ランダムフォレスト
手法1 選択 1

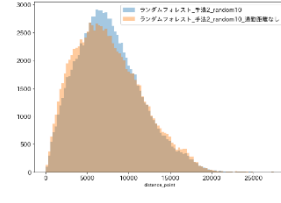


Fig.22 ランダムフォレスト
手法2 選択 2

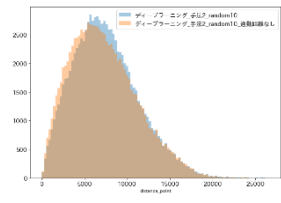


Fig.27 ディープラーニング
手法1 選択 2

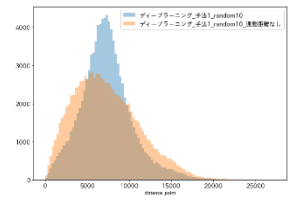


Fig.30 ディープラーニング
手法2 選択 2

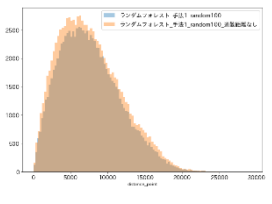


Fig.20 ランダムフォレスト
手法1 選択 3

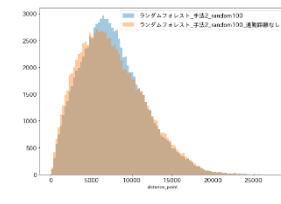


Fig.23 ランダムフォレスト
手法2 選択 3

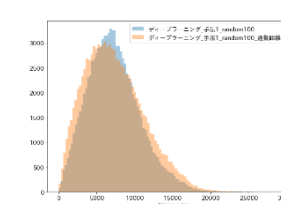


Fig.28 ディープラーニング
手法1 選択 3

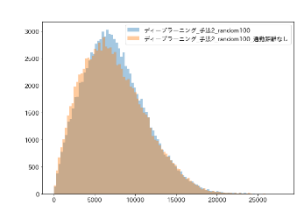


Fig.31 ディープラーニング
手法2 選択 3

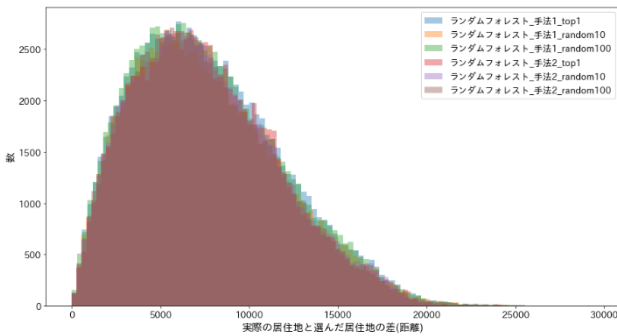


Fig.24: 居住地ランダムフォレスト_結果一覧

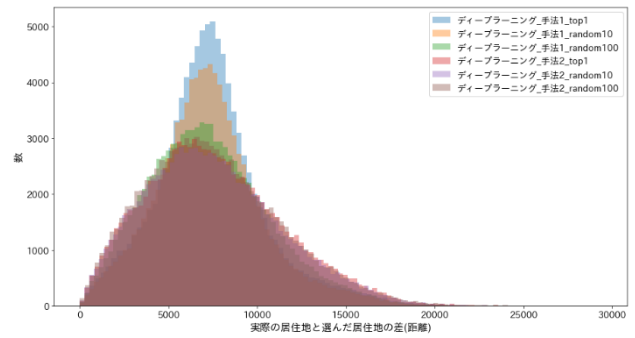


Fig.32: 居住地選択_ディープラーニングの結果一覧

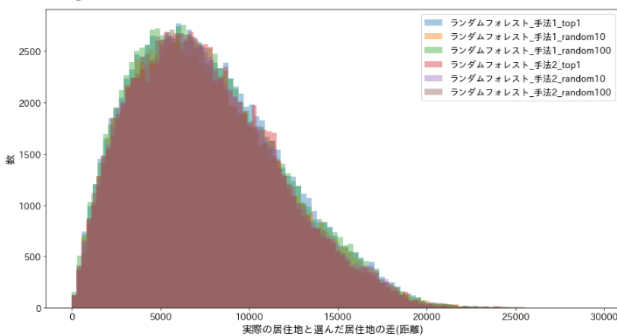


Fig.25: 居住地ランダムフォレスト通勤距離なし_結果一覧

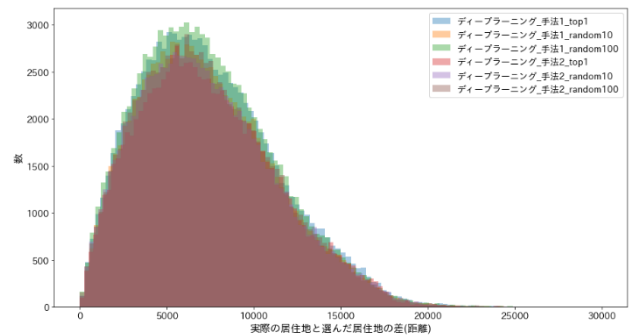


Fig.33: 居住地選択_ディープラーニングの結果一覧
(通勤距離なし)

4.3 実績値を用いた居住地選択の結果

4.2では居住地選択の結果をまとめ、求める地域属性パラメータ推定の誤差により居住地選択の予測能力を低下させた可能性があるとして述べた。そこで仮に求める地域属性のパラメータ推定の誤差が小さい場合、どのような結果となるか確認を行う為、実際の地域属性の値を用いて、居住地選択の推定を行った。実績値を用いる際、最善選択の選択肢1だと高確率に現在の居住地を選択する為、トップ10からランダムに選ぶ選択肢2とトップ100からランダムに選ぶ選択肢3を採用し、前項と同じ20%の世帯データを使い、居住地選択を行った。

実績値を用いた予想結果の誤差の合計はTable 8、誤差はFig. 34である。その結果、パラメータ推定による居住地選択と異なり、現居住地から3000m以内で居住地選択の推定が行えている。また手法1を用いた選択肢2が最も予測精度が高く、手法2の選択肢2よりも高い精度で予測できている。ただ推定が大きく外れてしまっている世帯もある。

実績値の居住地選択の結果から、パラメータ推定の精度を高める事が出来たら、提案した手法の居住地選択モデルは採用可能であり、特に手法1の選択肢2がもっとも誤差が小さかった。

Table 8: 居住地選択したエリアと実際に住んでいる居住地の距離の誤差の合計

手法	誤差の合計
実績値_選択肢2_手法1	516578365
実績値_選択肢3_手法1	610565915.8
実績値_選択肢2_手法2	634261165.5
実績値_選択肢3_手法2	664596036.6

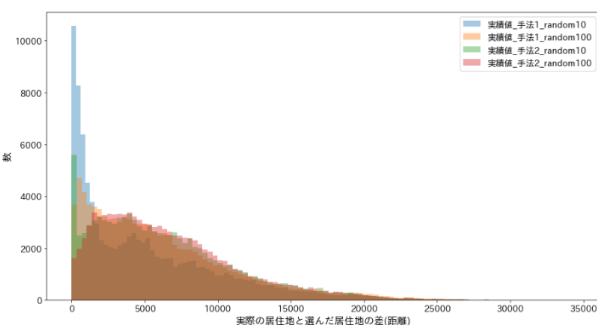


Fig. 34: 実績値を用いた居住地選択の結果

4.4 予測誤差の大きい世帯の傾向分析

推定を行った結果、現住所と推定住居の間に5,000m-8,000mの誤差が生じた世帯が多かったが、その中でも世帯によっては20,000m以上の誤差が見られた。そこで大きな誤差が生じた世帯の性質を調べる為、20,000m以上の誤差が生じた世帯と1,000m以内で収まった世帯を比較した。

推定誤差が大きい世帯の現住所の町丁目をFig. 35にまとめた。Fig. 35から推定誤差が大きい世帯の多くは、札幌市の郊外に居住している事が分かる。特に左下部、札幌市南区はほぼ森林区域であるが、ここに住む世帯で20,000m以上の誤差が生じている。20,000m以上の誤差が生じる原因として、居住地選択の予測で都心を選択しながらも、実際には札幌市郊外に居住している事から大きな誤差に繋がった事が考えられる。

また1,000m以内で予測できた世帯と20,000m以上で誤差が生じた世帯を世帯属性や地域属性など様々な観点で比較した所、明確な違いは無かったが、世帯主年齢で60歳以上の世帯の誤差がやや大きかった。そこで一つの仮説としてFig. 35では、20,000m以上誤差が生じた世帯は札幌市の郊外に分布しているが、本研究の居住地選択モデルでは、遠く離れている都心を選択している。本来なら居住地を移動するが、実際には何らかの理由で高齢者が都心に移動せず、郊外に住み続けている。特に札幌市南区のような衰退している過疎地域に残って生活している高齢世帯に対し、提案手法では大きな推定誤差が生じた可能性がある。合理的に考えると、生活利便性の為、高齢者は都心に移動するが、提案手法で採用した世帯属性や地域属性には表れない、戸建ての家を手放したくない、代々受け継いだ土地を守りたいなどの市民の拘りがあると推測できる。予測精度を高める為、このような各地域特性を調査し、モデルに組み込む必要がある。

このことから、衰退地域のコンパクトシティ構想が思う通りに進まない事と、本モデルの予測精度の低さと同様の原因がある可能性を指摘した。

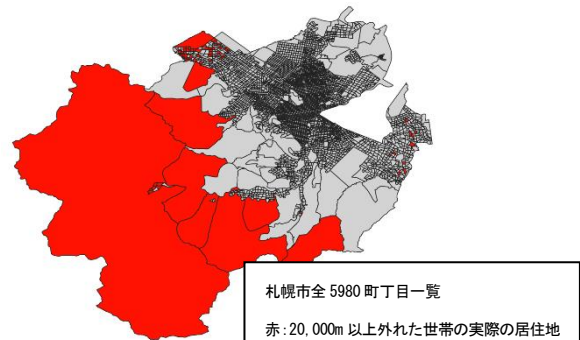


Fig. 35: 予測が20000m以上外れた世帯が実際に住んでいる町丁目

5 おわりに

本研究では、既存の居住地選択モデルには通勤時間が加味されていない事から、通勤時間を考慮した居住地選択モデルを提案した。また先行研究の居住地選択モデルで多く採用されている多項ロジットモデルではなく、新しい手法を提案し、北海道札幌市の仮想個票データを用いて、居住地選択の予測を行った。個票データには原田ほか(2018)により合成された仮想個票データを用い、選択結果は各世帯の嗜好と仮定した。ま

た妥当性評価は実際に住んでいる町丁目と予測で選んだ町丁目の直線距離とした。

居住地選択の予測は全体的に5,000m-8,000mの誤差が多く、精度が高いとはいえないが、実績値を用いた居住地選択は精度の高い結果が出たことから、求める地域属性パラメータ推定の精度を上げる事ができれば、提案した手法で精度の高い居住地選択ができる可能性が示唆された。

札幌市の通勤時間は電車社会の側面が強い中央区が15~30分が多く、車社会の側面が強い清田区は30~1時間だが、通勤時間と照らし合わせると、清田区の手移動を考えるとある程度誤差の範囲内といえるかもしれないが、中央区を考えると、予測の誤差は3,000m以内が妥当な値であり、本研究の居住地選択は通勤時間から考えると予測は外れていると結論づけられる。

今後の課題を述べる。

1つ目に、本研究ではすべての世帯で居住地選択の予測を行う事ができなかつた為、全世界帯で居住地選択を行う事である。妥当性を高める為、より多くの計算機リソースを確保する必要がある。

2つ目に、求める地域属性のパラメータ推定の精度を上げる事である。今回採用した3つの回帰モデルのうち、ランダムフォレストやディープラーニングはパラメータをチューニングせず、初期設定で行ったが、計算リソースを確保しパラメータ探索を行えば精度が上がる可能性がある。また他の回帰モデルの検討も行う必要がある。特に通勤時間の誤差を抑える手法を開発する必要がある。

3つ目に、札幌市の地域特性を捉え、より詳細なモデルを組み込む事である。特に推定結果では20,000m以上も誤差が生じる世帯がある為、モデルの精度を高める為、各地域特性を考慮したモデルを構築する必要がある。

4つ目に、交通手段を推定である。本研究で採用した仮想個票データには交通手段はない為、車を採用したが、交通手段を生成すれば、距離ではなく通勤時間を組み込む事ができる。

5つ目に変数の再検討である。通勤時間はアンケート調査から重視すると明らかになったものであり、数量的にどの程度影響を与えるのか明らかにしたものではない為、通勤距離をどこまで重視するか不明な所がある。また通勤距離を考慮する際、勤務地から求める通勤距離圏内であれば良く、例えば求める通勤距離が6kmの世帯にとって、本来なら5kmでも3kmでも1kmでも範囲内であるため、候補地となる場所は多い。しかし本研究の提案手法では、6kmを3kmや2kmより高く評価するモデルになっている為、通勤距離を再検討する必要がある。また本研究では採用を見送ったが、先行研究のモデルで採用されている女性タミーや戸建て率など、新たな変数も検討しなければならない。また家賃相場の

オープンデータがない為、土地地価を採用し、公示時価を用いたが、このデータは一つの区に対し、30~50地点のみのデータの為、範囲がかなり広く、詳細なデータではない。その為路線価を採用した方が高い精度を出せる可能性がある。

6つ目に選択肢の再検討である。3つの選択肢による予測結果から人間は必ずしも一番住みたい場所に住めるわけではない事が示唆された。本研究では最善選択やランダム選択を採用したが、ルーレット選択をいれるなど別の選択肢の採用も検討の余地がある。

本研究では仮想個票データを用いた。妥当性評価を行う為、実データを用いた予測を行い、比較する必要がある。

参考文献

- 1) 原田, 村田: 市区町村の統計表を考慮した都道府県単位の仮想個票の合成, 第15回社会システム部会研究会資料, pp-30-50 (2018)
- 2) Waddell: UrbanSim: Modeling Urban Development for LandUse, Transportation, and Environmental Planning, Journal of the American Planning Association, pp 297-314 (2007)
- 3) Margo: Explaining the postwar suburbanization of population in the United States: The role of income, Journal of Urban Economics, pp 301-310 (1992)
- 4) 小松, 中川: マイクロデータを用いた居住地選択による郊外居住の実証分析, 都市住宅学 57号, pp52-60 (2007)
- 5) 桑野: 移住相談内容を用いた居住地選択行動の要因分析, 都市計画論文集, 54巻3号, pp848-855 (2019)
- 6) 相: 複数の住環境指標が町丁目の人口増減パターンに与える影響東京圏1都3県の都市地域を対象に, 都市計画論文集, 49巻3号 p. 567-572 (2014)
- 7) 西山, 中谷, 栗栖, 荒巻, 花木: 居住地属性の住民選好に基づく類型化による居住地選択行動の解析, 環境システム研究論文集, 第39巻6号, p. II_1-II_10 (2011)
- 8) 清水, 中山, 土佐野: 若年層の転居理由別に見た居住地選択要因に関する研究, 日本建築学会計画系論文集 第82巻 732号 423-432 2017年
- 9) 鈴木, 杉木, 宮本: 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 51巻3号 p. 839-846 (2016)
- 10) 井上: 「全国小地域別将来人口推計システム」バージョン2.0の公開について 2017年度人文地理学会 一般研究発表 (2017)
- 11) https://www5.cao.go.jp/keizai-shimon/kaigi/special/future/sentaku/s2_1.html
- 12) https://www.city.sapporo.jp/kikaku/mirai-sousei/2nd/documents/miraisousei2nd_04.pdf
- 13) <http://project-osrm.org/>
- 14) <https://geopy.readthedocs.io/en/stable/>