

172. Deep Neural Network を用いた物件の賃料推定モデルの構築と地域ポテンシャルマップの作成

- 建物属性と地域属性に着目して -

Rent Estimation Model and Visualization of Area Potential Using Deep Neural Network

- A Focus on Attributes of Building and Area -

前田 翠*, 関本 義秀*, 瀬戸 寿一**, 檜山 武浩***

Midori Maeda*, Yoshihide Sekimoto*, Toshikazu Seto**, Takehiro Kashiama***

Both in the Tokyo metropolitan area and rural areas in Japan, the number of development projects being undertaken by private companies and local governments has been increasing. Under these circumstances, there is need for a system to estimate the rent paid by tenants and consider the features of living areas from several perspectives. This information can be used by developers when choosing a site. For both price estimation and an analysis of factors affecting price, we apply three approaches: deep neural networks, the hedonic approach, and random forest regression and clarify the advantages and limitations of them. The results show the importance of data cleaning when using deep neural network. Also, they show that it is possible to create a potential map of areas and visualize time series transformations and the impacts of economic trends on it by using latitude, longitude, the year of registration and deflator as explanatory variables besides building attributes.

Keywords: Price Estimation, Potential Map, Deep Neural Network, Hedonic Approach

賃料推定, ポテンシャルマップ, ディープニューラルネットワーク, ヘドニックアプローチ

1. 研究の背景と目的

近年東京首都圏ならびに地方都市部を中心に都市再生緊急整備地域の指定等により民間企業主導の再開発事業が活発に行われており、特に東京都内では大規模商業ビルや高層オフィスビルの開発が進んでいる。また、全国的にも地方活性化やコンパクトシティ化の流れにより民間企業ならびに行政主導の地域開発が将来的により一層促進され、今後多くの建物が建設されることが考えられる。そういった中で、建設した物件の賃料を推定したい、開発計画策定にあたり地域の持つ特徴を把握し、開発対象地の決定・開発計画の策定を行いたいという事業者のニーズがある。物件の賃料は、物件用途や面積や築年数をはじめとする建物属性と、どんな地域に物件が立地しているのかという地域属性により決まると言える。これを考慮すると、前者は地域属性・建物属性どちらも加味した物件賃料の推定を行いたいというニーズ、後者は地域の特徴を知りたいというニーズであると言える。

また、近年ビッグデータへの関心の高まりから民間企業や行政が保有する様々なデータが公開されてきており、以前よりも多種多様なデータの入手・活用が容易になってきている。後者のニーズに対しては、それらのデータを活用した、地域の特徴を分析・比較するためのデータ可視化システムが国内外で開発されてきている。例えば、MyCityForecast は既存のオープンデータを活用した Web ベースの都市の将来像シミュレーションシステムであり、500m メッシュ単位で将来の人口やアクセシビリティ等 14 の指標から地域の特徴を捉えることができる^{1), 2)}。また、ISO 37120 は World Council on City Data (WCCD) が提供する都市間比較ポータルサイトであり、世界の 20 の都市を 17 カテゴリー、100 指標から比較することができる。これらのシステムは現在・将来の地域指標の分

析のためのツールとしては有用だが、実際に物件を建設した際の賃料の推定という前者のニーズに答えられていない。

物件の賃料推定と賃料を決定づける要因の分析という観点では、従来ヘドニックアプローチが広く用いられてきた^{3), 4), 5)}。一方で、近年オープンデータの増加等に伴い、機械学習手法も多く用いられるようになってきている^{6), 7), 8)}。中でも (Deep) Neural Network は賃料や価格の推定を行うために一般的に用いられている手法である。両者の推定精度の比較を行う研究も数多く行われているが、用いるデータの量やデータの特性、パッケージの種類、被説明変数と説明変数の関係が線形なのか否かによって推定精度の優劣が変わるとされている^{9), 10), 11), 12)}。また、地域属性を考慮した物件の賃料推定という観点では、名古屋市内のマンション・アパートの賃貸物件の賃料推定において、緯度・経度から地域性を学習させることが可能であるということが明らかにされている¹³⁾。さらに、Random Forest Regression (以下 RFR) は説明変数の重要度を算出できるだけではなく連続値の推定を行うことができる機械学習手法として注目されているが^{14), 15)}、賃料推定における使用例はほとんどない。

以上を踏まえて、本研究ではオープンデータを活用した物件の賃料推定・地域指標分析システムを構築し、地域指標の分析のみならず開発物件の賃料の推定ならびに開発による地域への影響の評価を可能にすることを念頭におき、その初段として、開発物件の賃料推定のための建物属性・地域属性両方を加味したモデルの構築、ならびに地域属性が物件の賃料に与える影響の可視化を目的とする。さらに、その目的の下で、東京 23 区を対象地域とし、Deep Neural Network (以下 DNN) と既往手法であるヘドニックアプローチ、比較的新しい手法である RFR の 3 手法で物件の賃料推定を行い、各々の優位点と限界点を示す。さらに、物件の賃料推

* 正会員 東京大学 生産技術研究所 工学系研究科社会基盤学専攻 (Institute of Industrial Science, The University of Tokyo)

** 正会員 東京大学 空間情報科学研究センター (Center for Spatial Information Science, The University of Tokyo)

*** 非会員 東京大学 生産技術研究所 (Institute of Industrial Science, The University of Tokyo)

定モデルを構築するにあたり、緯度・経度、区ダミー変数と緯度・経度、町字ダミー変数による東京 23 区の地域性の学習を試みる。本文では、地域属性はモデル化するにあたって用いる地域の特徴を表す変数を、地域性は地域の持つ特徴を指す言葉として用いる。

2 章ではデータの概要について説明する。3 章では物件用途や面積をはじめとする建物属性に加え、物件の所在地の特徴を表す地域属性を考慮した物件の賃料推定モデルを DNN を用いて構築する。その際、地域属性変数として緯度・経度や区ダミー、町字ダミー、いずれを用いた場合が最も物件の所在地の特徴を学習させることができるのか比較する。さらに、ヘドニックアプローチと RFR と賃料推定精度と地域属性の学習精度を比較する。4 章では 3 章で構築したモデルを元に地域ポテンシャルマップを作成する。地域ポテンシャルの定義については 4 章にて述べる。また、3 章のモデルに物件の登録年度や登録年度に対応するデフレーターを加えて学習させることで地域ポテンシャルマップの時系列変化の表現や、経済動向の反映を試みる。5 章では、4 章で用いた登録年度ならびにデフレーターを元に将来の物件の賃料推定を試みる。最後に 6 章において本研究のまとめを述べる。

2. データの概要とモデル構築環境

物件の賃料推定にあたって今回は athome の不動産データライブラリー、戸データ、全国、1999-2013 データセットの貸店舗・事務所のデータを用いた。このデータは 2008 年より国立大学法人東京大学の空間情報科学研究センター (CSIS) に登録されている共同利用データを用いた。今回対象とする東京 23 区においては 1999-2000 年の登録物件数は 0 件、2001-2013 年の累計登録物件数は 130 万 8311 件である。データ項目を表-1 に、東京 23 区における登録物件数の推移を図-1 に示す。

表-1 データ項目

1. 物件番号	2. 種目名称	3. 新築/中古	4. 県名称
5. 所在地名 1	6. 所在地名 2	7. 沿線名称	8. 駅名称
9. 徒歩(分)	10. バス(分)	11. 停徒(分)*	12. バス停名称
13. 物件賃料	14. 土地面積	15. 専有面積	16. ベランダ面積
17. 土地権利名称	18. 建ぺい率	19. 容積率	20. 地上階数
21. 地下階数	22. 所在階	23. 築年	24. 築月
25. 登録日付	26. 棟 ID		

※徒歩でバス停から要する時間

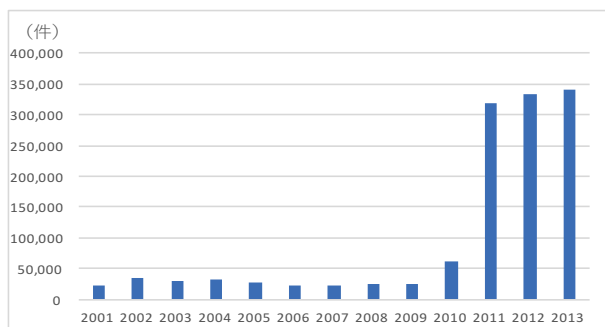


図-1 貸店舗・事務所の登録物件数の推移 (東京 23 区)

図-1 から、2009 年以前には登録物件数が相対的に少ないが、以降データ整備が進むにつれて登録物件数が激増することがわかる。登録年度によってデータ数に大きな偏りがあるが、2008 年以前の登録物件数も絶対数としては大きいことから 2001 年から 2013 年の全てのデータを用いた。

また、DNN と RFR のモデル構築には Jupyter Notebook を、ヘドニックモデルの構築には R を用いた。

3. 地域属性を加味した物件の賃料推定モデルの構築

(1) 建物属性による基本モデルの構築

まず、DNN を用いて建物属性のみを考慮した物件の賃料推定モデルを構築した。被説明変数を物件の賃料(千円/月)とし、表-2 の 9 つの変数を説明変数とした。

表-2 基本モデルの説明変数

1. 貸事務所	二値変数	6. 地下階数	整数
2. 貸店舗(建物一部)	二値変数	7. 所在階	整数
3. 貸店舗・事務所	二値変数	8. 築年数	整数
4. 物件面積	実値	9. アクセシビリティ	整数
5. 地上階数	整数		

ここで、1 から 3 の物件用途は表-1 の 2. 種目名称を元に 3 種の二値変数とした。二値変数としたのは、連続値を用いた場合には例えば貸店舗事務所:1 と貸店舗(建物一部):2 の方が貸店舗事務所:1 と貸店舗・事務所:3 よりも類似しているといったような関係性をモデルが学習してしまう可能性があるためである。また、9 については表-1 の 9 から 11 を元に最寄駅から各物件までの所要時間(分)を、徒歩(分)+バス(分)×15+停徒(分)とし、それをアクセシビリティとして定義した。層数は 3、各層のパーセプトロン数は 500、学習率は 0.001 とし、データの 8 割をランダムサンプリングしてトレーニングデータとし、1 回を 10000 ステップとして 30 回学習させ、残りの 2 割のテストデータでモデルの推定精度を検証した。統計的な検証はされていないが一般的に学習データとテストデータの比率は 8:2 とされることが多い。また、ランダムサンプリングはモデルを構築する度に行い、同じ学習データとテストデータで各モデルの精度を比較しているわけではない。さらに、DNN がエラー値に対応できるか否か(極端に他の分布から外れる値を外れ値と判断して学習ができるか否か)を検証するために、データクリーニング無しと有りでの推定精度の比較を行なった。

(I) データクリーニング無しの場合

貸店舗事務所の登録物件 130 万 8311 件全てのデータを用いた場合、テストデータの実際の物件賃料と推定賃料の相関係数は 0.007、2 乗平均平方根誤差(以下 RMSE)は 87101 (千円/月)、平均絶対誤差(以下 MAE)は 604 (千円/月)、テストデータの平均賃料は 901 (千円/月)であった。全くとってよいほどモデルを学習させることができていないということが分かる。

(II) データクリーニング有りの場合

データの基礎分析を行った結果、物件面積が小・中規模であるにも関わらず 5000(万円/月)以上の非常に高額な物件が存在していたことから、それらの物件84件を除いた130万8227件のデータを用いて同様の分析を行った。その結果、テストデータの実際の物件賃料と推定賃料の相関係数は0.875, RMSEは292(千円/月), MAEは111(千円/月), テストデータの平均賃料は415(千円/月)であった。

以上から、全データのたった0.006%を占める外れ値によりDNNは大きな影響を受けていることがわかる。DNNを用いる際には従来と同様に、もしくはそれ以上に慎重にデータクリーニングを行う必要があることが示唆される。

以降では全てデータクリーニングを行った上で分析を行っている。3-(1)(II)のデータクリーニングを行った上で構築したモデルを基本モデルとし、以降の3-(2)(I)(II)(III)では基本モデルに用いた説明変数(表-2)のうち9. アクセシビリティの代わりに緯度・経度、緯度・経度と区ダミー、町字ダミーをそれぞれ地域属性変数として用い、モデルに地域属性を学習させるためにはどの地域属性説明変数を用いるのが適当か比較・検討した。以降の(I)(II)(III)の建物属性説明変数と各々で用いた地域属性説明変数を表-3に示す。緯度・経度は表-1の4:県名称, 5:所在地名1, 6:所在地名2に、物件所在地の都道府県名から町丁目までが含まれていたことから、東京大学空間情報科学研究センターが提供するCSVアドレスマッチングサービスを用いて町字単位で付与した。

表-3 建物属性と地域属性説明変数

建物属性	変数種別		地域属性	変数種別
1. 貸事務所	二値変数	(I)	緯度	実数
2. 貸店舗(建物一部)	二値変数		経度	実数
3. 貸店舗・事務所	二値変数	(II)	緯度	実数
4. 物件面積	実値		経度	実数
5. 地上階数	整数	(III)	区ダミー	二値変数
6. 地下階数	整数		町字ダミー	二値変数
7. 所在階	整数			
8. 築年数	整数			

(2) 建物属性と地域属性によるモデルの構築

(I) 緯度・経度による地域性の学習

層数や学習率等のハイパーパラメータは3-(1)(II)の基本モデルと同様に設定し、全体の8割のデータを用いて学習を行い、残りの2割のテストデータで推定精度の検証を行った。その結果、テストデータの実際の物件賃料と推定賃料の相関係数は0.915, RMSEは240(千円/月), MAEは91(千円/月), テストデータの平均賃料は412(千円/月)であった。物件の賃料推定精度自体は3-(1)(II)の建物属性による基本モデルと比較して向上している。

(II) 区ダミーと緯度・経度による地域性の学習

層数や学習率等のハイパーパラメータは3-(1)(II)の基本モデルと同様に設定し、全体の8割のデータを用いて学習を行い、残りの2割のテストデータで推定精度の検証を行った。

その結果、テストデータの実際の物件賃料と推定賃料の相関係数は0.926, RMSEは226(千円/月), MAEは87(千円/月), テストデータの平均賃料は413(千円/月)であった。物件の賃料推定精度自体は3-(1)(II)の建物属性のみを用いた基本モデル、ならびに3-(2)(I)の建物属性に加えて緯度・経度を地域属性として加えた場合と比較して向上している。

(III) 町字ダミーによる地域性の学習

東京都内の町字数は939であり、これと表-3に示した8つの建物属性説明変数を合わせると説明変数は947となるため、層数は5, 各層のパーセプトロン数は1000とした。その他のハイパーパラメータについては3-(1)(II)の基本モデルと同様とした。全体の8割のデータを用いて学習を行い、残りの2割のテストデータで推定精度の検証を行った。その結果、テストデータの実際の物件賃料と推定賃料の相関係数は0.937, RMSEは215(千円/月), MAEは85(千円/月), テストデータの平均賃料は414(千円/月)であった。物件の賃料推定精度自体は3-(1)の建物属性のみを用いた基本モデル、ならびに3-(2)(I), (II)の建物属性に加えてそれぞれ緯度・経度、区ダミーと緯度・経度を地域属性として加えた場合と比較して向上している。物件の賃料と推定賃料の関係を図-2に示す。

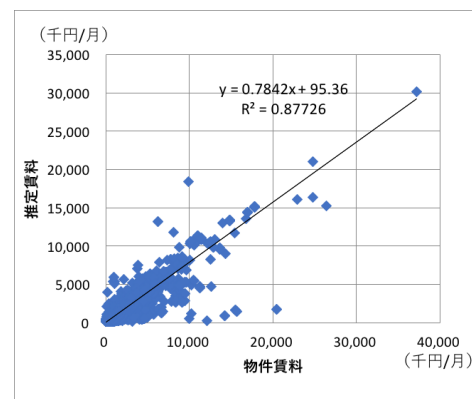


図-2 物件賃料と推定賃料の関係(DNN)

次に、3-(2)(I)(II)(III)それぞれのモデルにおいて、地域属性による物件賃料の違いをどれだけ表現できているのかを確認するため、構築されたモデルを用いて表-4に示すような一定の建物属性条件での元で物件の賃料推定を行なった。すなわち、表-4に示すような建物属性を持つ物件が東京23区内全てに存在する場合の、各地域における物件の賃料を推定した。また、その結果推定された物件の賃料の高低をQGIS上で自然分類でクラス数を50として表現したものを図-3に示す。

表-4 建物属性説明変数の値

建物属性	値	建物属性	値
貸事務所	1	地上階数	8
貸店舗(建物一部)	0	地下階数	1
貸店舗・事務所	0	所在階	1
物件面積	80 m²	築年数	10(年)

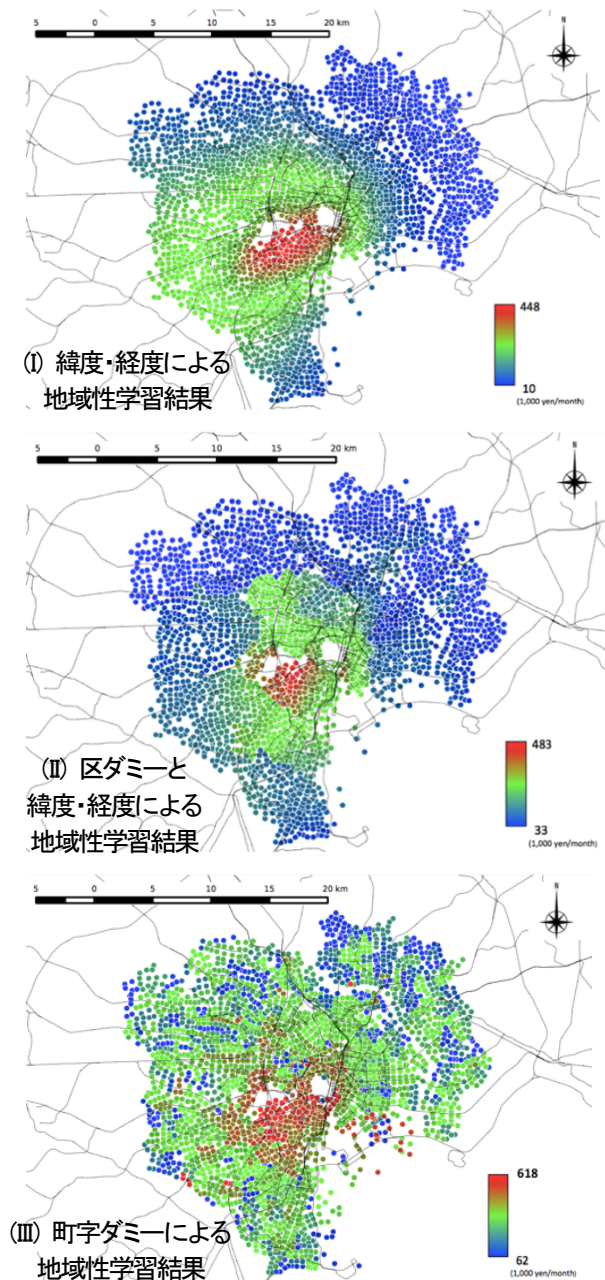


図-3 地域性学習結果比較

図-3から、(I)のモデルでは、東京23区の中心部で賃料が高くなる傾向があることは学習出来ている。また、(II)のモデルでは渋谷、新宿、六本木、銀座周辺で賃料が高くなっているということは学習できている。(III)のモデルは3つのモデルの中で、所在地による物件賃料の違いが最も大きく出ている。(I)(II)のモデルと比較して(III)のモデルの推定賃料と実際の賃料の相関係数が大きく、RMSEやMAEが小さいことを加味すると、(III)のモデルが最も地域性を学習することができていると考察される。また、(I)(II)のモデルにおいて(III)のモデルと同様に層数を5、各層のパーセプトロン数を1000に変更して同様の分析を行った場合にも、地域属性の学習結果には大きな差は見られなかった。名古屋市内のマンショ

ン・アパートの賃貸物件の賃料推定においては緯度・経度から地域性を学習させることが可能であると明らかにされているが¹³⁾、今回対象としている東京23区はよりスケールが大きく、地域の特徴がかなり複雑であるため、町字ダミーを地域属性変数として用いるのが適当であると言える。

(3) 他手法との比較

3-(2)より所在地による物件の賃料の違いを反映するには(III)のモデルのように地域属性変数として町字ダミーを用いることが最適であるという結論に至った。3-(2)(III)で用いた建物属性ならびに地域属性変数(表-3)を用いて、ヘドニックアプローチとRFRで同様の分析を行い、比較する。

(I) ヘドニックアプローチ

まず、分析を行うにあたり用いた全ての説明変数の相関関係を分析したが、多重共線性は生じていなかった。被説明変数を物件の賃料(千円/月)とし、説明変数を3-(2)(III)と同様にした場合の決定係数の値は0.266、被説明変数を単位面積当たりの物件の賃料(千円/㎡・月)とし、同様の説明変数を用いた場合の決定係数の値は0.588であり、いずれも低い値となった。従って、単位面積当たりの物件の賃料(千円/㎡・月)を被説明変数とし、3-(2)(III)と同様の説明変数を用いて単位面積当たりの物件の賃料の線形予測式(式(1))をつくり、予測された単位面積当たりの物件賃料を元に物件賃料(千円/月)の推定を行った(式(2))。

$$z = \sum_{i=1}^8 a_i x_i + \sum_{j=1}^{939} b_j y_j \quad (1)$$

$$x_4 z = x_4 (\sum_{i=1}^8 a_i x_i + \sum_{j=1}^{939} b_j y_j) \quad (2)$$

x_j は表-3の8つの建物属性説明変数に対応し、 y_j は各物件が東京23区内の939町字のうちどの町字に存在しているのかを表す二値ダミー変数である。3-(2)(III)のDNNのモデルとの比較を行うため、全体の8割のデータを用いてパラメータ a_i 、 b_j の推定を行い、残りの2割のテストデータと推定されたパラメータから物件の賃料を推定した。その結果、テストデータの実際の物件賃料と推定賃料の相関係数は0.902、RMSEは256(千円/月)、MAEは98(千円/月)、テストデータの平均賃料は413(千円/月)であった。図-4に物件賃料と推定賃料の関係を示す。図-2のDNNでの推定結果と比較すると推定の誤差が大きいことが分かる。

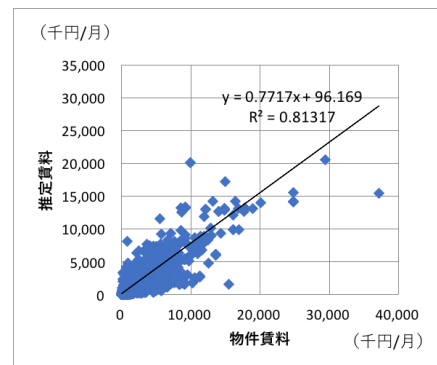


図-4 物件賃料と推定賃料の関係(ヘドニック)

(II) Random Forest Regression

3-(2)(III)と同様の被説明変数と説明変数を用いて分析を行った。RFR は、重要と考えられる説明変数に対して、ある閾値に基づいてデータを2分割するという作業を繰り返して決定木を作成し、それを弱学習器とする集団学習アルゴリズムである。一般的に、学習器を構成する決定木の深さを深くすればモデルの説明力は向上するが、過学習の可能性があることが知られている。今回用いた説明変数の総数は947と非常に多く、与えられた説明変数のうち重要と考えられるものから選択的に学習するRFRでは、町字ダミー変数を均等に学習させることは非常に難しい。従って、過学習の可能性は高くなるが、決定木の深さを100に設定し、町字ダミー変数をどの程度学習できるのかを検証した。過学習が生じているかどうかを確認するため8割のデータを用いて学習させ、残りの2割のテストデータで推定精度を検証するという作業を10回行った。その結果、テストデータの実際の物件賃料と推定賃料の相関係数は0.979、RMSEは120(千円/月)、MAEは20(千円/月)、テストデータの平均賃料は414(千円/月)であった。これらは10回の平均値である。図-5に物件賃料と推定賃料の関係を示す。

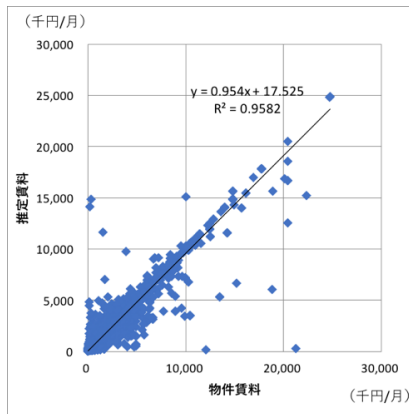


図-5 物件賃料と推定賃料の関係(RFR)

図-5を図-2のDNNでの推定結果と比較すると全体的に推定の誤差が小さく、特に10,000(千円/月)よりも高額な物件の推定精度が高いことが分かる。

また、3-(2)と同様、表-4に示す建物属性を持つ物件が東京23区内全てに存在する場合の、各地域における物件の賃料を3-(3)(I)ヘドニック、(II)RFRモデルで推定しQGIS上で物件の賃料の高低を表したものを図-6に示す。

図-6より、ヘドニックとDNN(図-3(III))では物件の立地する町字による物件賃料の違いを詳細に学習することができているが、RFRでは局所的に高額・低額となる町字のダミー変数が学習されているのみで、全町字ダミーを平均的に学習することはできていないということが分かる。これらの2手法と3-(2)(III)のDNNの結果を比較すると、DNNが賃料推定精度も比較的高く、町字による物件賃料の違いも詳細に学習できており、バランスが取れたモデルとなっていることが

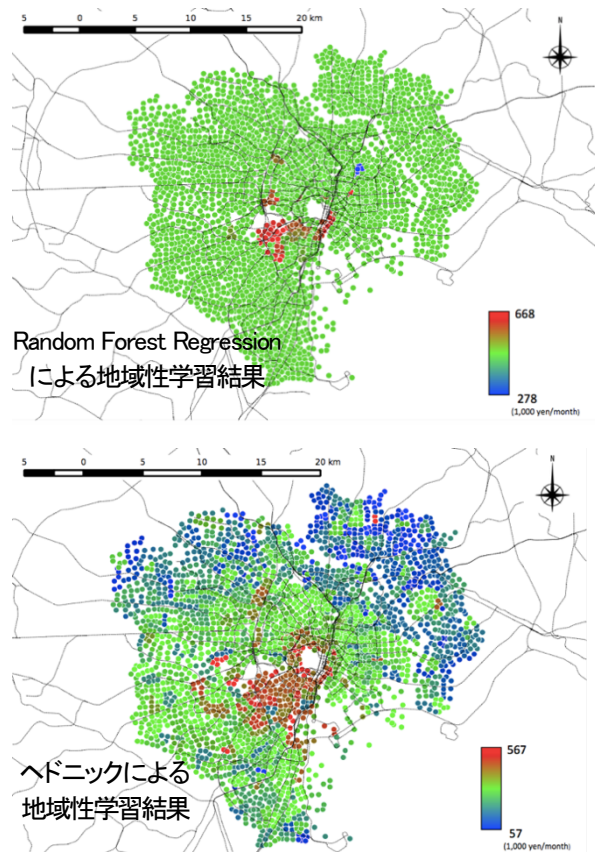


図-6 RFRとヘドニックによる地域性学習結果

分かる。また、ヘドニックとRFRについても3-(2)(I)(II)と同様に緯度・経度、区ダミーと緯度・経度による地域性の学習を試みたが、推定精度・町字による物件賃料の違いの学習の程度もいずれも町字ダミーを用いた本結果よりも劣っていた。さらに、RFRについては重要度(Gini減少量)を算出したところ物件面積の重要度が0.826と非常に高かった。ここから、賃料と面積の関係は線形ではないが、面積に応じてある種のクラスタリングをした上で線形近似をすることで精度の高い賃料の推定を行うことができるということが示唆された。一方、DNNとヘドニックではクラスタリングをしない代わりに様々な説明変数を均等に学習することで推定精度を高くしているという特徴があるということが分かる。

4. 地域ポテンシャルマップの作成

3章で地域属性説明変数を加えたモデルを構築し、表-4に示した建物属性条件を元に各地域における物件の賃料を推定した。これは物件の所在地による同一物件の賃料の違いを表しており、建物属性説明変数では表現できない都心部からの近さ・周辺の人口・賑わい等の各地域の魅力を反映していると考えられる。従って、図-6に示した、ある一定の建物属性条件下で推定された所在地による物件賃料の違いを地域ポテンシャル、それを地図上で可視化したものを地域ポテンシャルマップと本研究では定義する。図-7に表-

4の建物属性説明変数の中で物件面積を400 m²とした場合の地域ポテンシャルマップを示す。

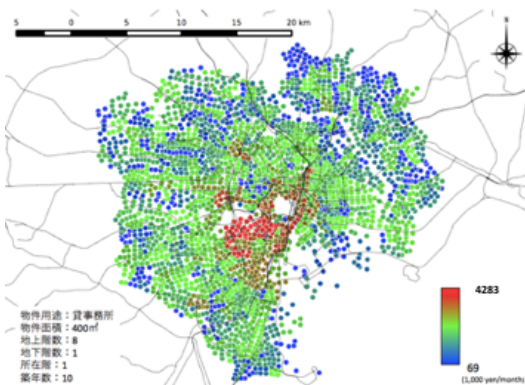


図-7 地域ポテンシャルマップの例

図-7 と図-3(Ⅲ)を比較すると相対的に高額・低額と推定される地域の分布に違いが見られる様に、建物属性説明変数の値を変化させることでその条件下で相対的に高額・低額と推定される地域を可視化することが可能となる。

(1) 時系列変化の表現

地域ポテンシャルマップが年度によってどの様に変化するのかを表現するため、3-(2)(Ⅲ)のモデルに、表-1の25.登録日付から得られる登録年度のデータを加えて学習させた。ハイパーパラメータについては3-(2)(Ⅲ)のモデルと同様とした。全体の8割のデータを用いて学習を行い、残りの2割のテストデータで推定精度の検証を行った。その結果、テストデータの実際の物件賃料と推定賃料の相関係数は0.944、RMSEは212(千円/月)、MAEは86(千円/月)、テストデータの平均賃料は415(千円/月)であった。3-(2)(Ⅲ)の結果と比較すると、登録年度の説明変数を加えることで少し推定精度

が向上していることが分かる。表-4の建物属性と同条件での2001年、2005年、2009年、2013年の地域ポテンシャルマップの変化を図-8に示す。成城学園前では2001年には相対的に地域ポテンシャルが高かったが、2013年にかけて地域ポテンシャルが低下していることが分かる。一方で二子玉川では2013年にかけて徐々に相対的に地域ポテンシャルが上昇していることが分かる。これは再開発による影響が大きいと考えられる。この様に登録年度を説明変数に加えることで物件の賃料精度が向上すること、再開発の影響などによる地域ポテンシャルの変化が表現できることから、登録年度によって物件の賃料に有意な違いが生じていると言えよう。

(2) 経済動向の反映

次に、地域ポテンシャルマップが経済動向によってどの様に変化するのかを表現するため、3-(2)(Ⅲ)のモデルに、デフレーターを加えて学習させた。

具体的には、内閣府が提供する暦年のデフレーター(連鎖方式)を元に登録年度に対応する値を用いた。ハイパーパラメータは3-(2)(Ⅲ)のモデルと同様とした。全体の8割のデータを用いて学習を行い、残りの2割のテストデータで推定精度の検証を行った。その結果、テストデータの実際の物件賃料と推定賃料の相関係数は0.933、RMSEは225(千円/月)、MAEは83(千円/月)、テストデータの平均賃料は414(千円/月)であった。3-(2)(Ⅲ)の結果と比較すると、推定精度が向上しているが、4-(1)と比較すると推定精度が低い。次に、2001年から2013年の間でデフレーターが最小である2009年と最大である2013年について、表-4の建物属性と同条件で各町字における物件の賃料を推定した。2009年のデフレーターの値は-2.9%、2013年のデフレーターの値は1.9%であった。デフレーターの値が-2.9%から1.9%に変化した際

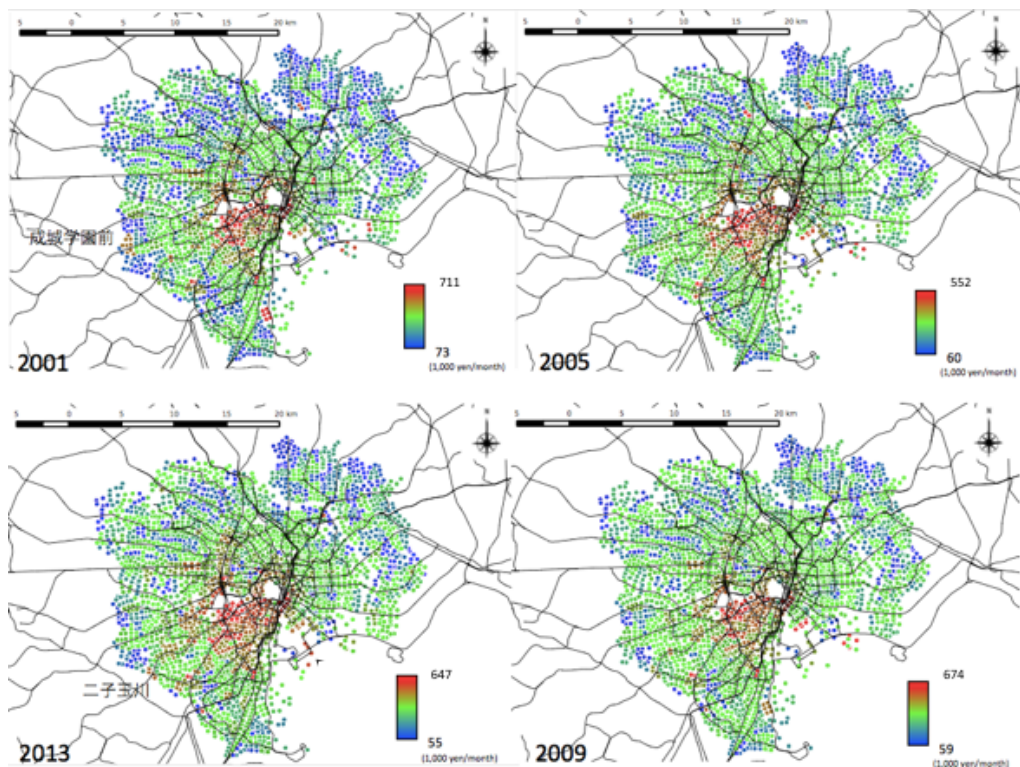


図-8 地域ポテンシャルの時系列変化

の賃料の変化率, 1.9%から-2.9%に変化した際の変化率, それぞれを地図上で自然分類で 10 クラスで表したものを図-9, 図-10 に示す。

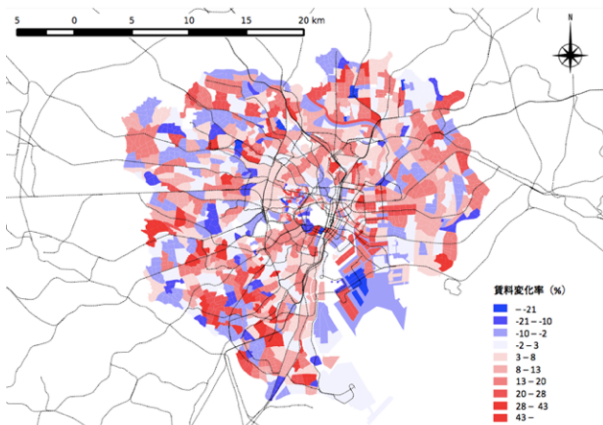


図-9 デフレーターが増加した場合の賃料変化率

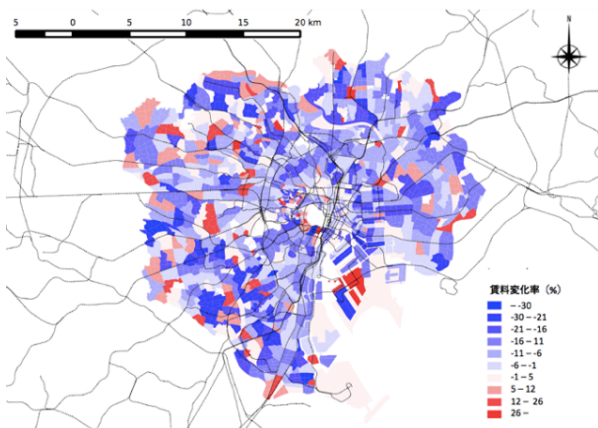


図-10 デフレーターが下落した場合の賃料変化率

図-9, 図-10 を比較すると, 図-9 で暖色となっている町字で図-10 では寒色となっている町字が多くなっていることが読み取れる。逆に図-9 で寒色となっている町字では図-10 では暖色となる傾向にあることが分かる。ここから, 景気が改善した際に賃料が高くなる町字では景気が悪化した際に賃料が低くなり, 逆も同様になるといった傾向があることが伺える。この様にデフレーターを説明変数に加えることで経済動向によって物件の賃料に有意な違いが生じていると言える。

5. 将来の物件の賃料推定の試み

4-(1)(2)から物件の登録年度ならびに経済動向による物件賃料の違いをモデルが学習することができているということが明らかになった。登録年度のみ, 登録年度ならびにデフレーターを所与として将来の物件の賃料推定が可能かどうかその可能性を探る。登録年度を所与として推定をするモデルをモデル(i), 登録年度とデフレーターを所与として推定をするモデルをモデル(ii)とする。いずれのモデルにおいても建物属性ならびに地域属性説明変数は共通である。表-5 に両モデルで用いる説明変数を示す。

建物属性	変数種別	地域属性	変数種別
貸事務所	二値変数	町字ダミー	二値変数
貸店舗(建物一部)	二値変数		
貸店舗・事務所	二値変数	所与とする変数	変数種別
物件面積	実値	(i) 登録年度	整数
地上階数	整数	(ii) 登録年度	整数
地下階数	整数	デフレーター	実数
所在階	整数		
築年数	整数		

表-5 将来の物件の賃料推定に用いた説明変数

学習データの登録年度と物件の賃料推定をする年度が近いと賃料の変化が小さく, 見かけ上推定精度が高くなる可能性がある。従って, 学習データの年度と推定する年度の間隔を空けるため, 2001 年から 2005 年のデータを学習データとしてモデルを学習させ, 2008 年から 2013 年の物件の賃料を推定することとする。さらに, 同時点での物件の賃料推定精度と将来の物件の賃料推定精度を比較するために, 2001 年から 2005 年のデータのうち 8 割を学習させ, 2 割をテストデータとしてモデルの同時点での推定精度を検証した後, 同モデルを用いて 2008 年から 2013 年までの物件賃料を推定し, 将来賃料推定精度を検証する。ハイパーパラメータについては 3-(4)のモデルと同様とした。表-6 に学習データならびに各年度のデータ数を示す。

表-6 データ数

	登録年度	データ数	詳細
学習用	2001-2005	78741 (8 割)	
賃料推定・精度検証用	2001-2005	19685 (2 割)	同時点賃料推定・精度検証用
	2008	19958	将来賃料推定・精度検証用
	2009	22081	将来賃料推定・精度検証用
	2010	62541	将来賃料推定・精度検証用
	2011	341033	将来賃料推定・精度検証用
	2012	358631	将来賃料推定・精度検証用
	2013	310187	将来賃料推定・精度検証用 データクリーニング後※

※データクリーニング前のデータ数は367870

2013 年のデータは, 物件面積が 10 m²以下もしくは 400 m²以上のもの, 築年数が 50 年以上の物件が多く含んでおり, 他の年度のデータと比較して登録物件の属性が大きく偏っていた。従って, それらのデータを除いた上で推定を行った。モデル(i)ならびにモデル(ii)で各年度の物件の賃料を推定し, その推定値と実際の物件の賃料の相関係数と RMSE を算出した。相関係数を図-11 に, RMSE を図-12 に示す。

図-11 より, 2008-2010 年ではモデル(i)とモデル(ii)の間に, 相関係数に大きな差はないが, 2011-2012 年ではモデル(ii)の値が大きくなっていると分かる。図-12 より, 2008-2010 年ではモデル(ii)の方が RMSE が大きくなっているが, 2011-2013 年ではモデル(ii)の値が小さくなっていると分かる。

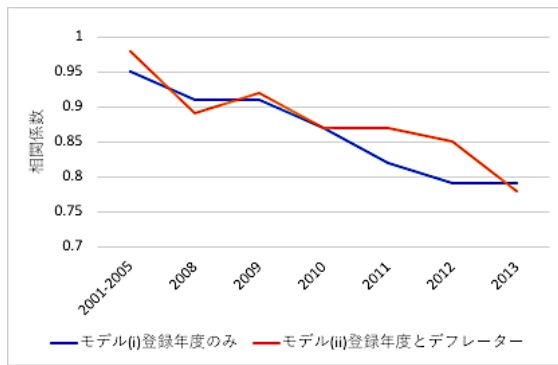


図-11 モデル(i)とモデル(ii)の相関係数の比較

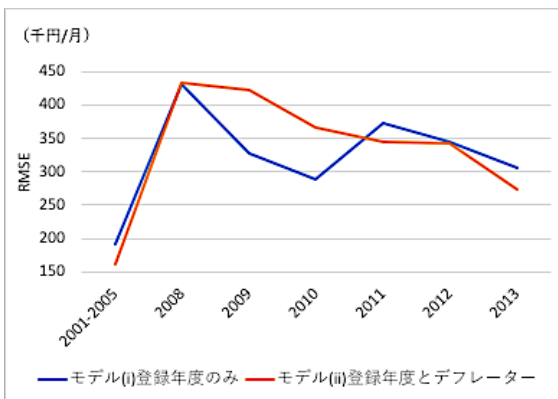


図-12 モデル(i)とモデル(ii)のRMSEの比較

以上から、今回の結果ではデフレーターの有無によって推定精度に有意な差は見られず、デフレーターを説明変数に加えることで推定精度が向上するとは言えない。4(2)で2001-2013年のデータを用いた際にはデフレーターによる物件賃料の違いが学習されていたことを鑑みると、2001年から2005年の5年間の学習データのみでは、モデルに年度とデフレーターの説明変数をうまく区別して学習させられないということが示唆される。それ故、デフレーターがモデルの推定精度向上に寄与しなかったのではないかと考察される。従って、より長期的なデータを用いて再度同様の分析を行うことが今後の課題として挙げられる。また、相関係数についてはモデル(i)、モデル(ii)いずれでも右肩下がりであるが、推定賃料に学習データに対して賃料推定・精度検証用データが極めて膨大なことを加味すると必ずしも推定精度が低いとは言えない。今後、各年度にデータ量に偏りが無い状態で同様の分析を行い、学習データの量が推定精度に及ぼす影響を分析する必要がある。

6. まとめと今後の課題

本研究では建物属性に加え地域属性を加味した賃料推定モデルを構築するにあたり、3章と4章ではDNN、RFRとヘドニックの比較を行い、各々の特徴を明らかにし、5章では時系列変化・経済動向を考慮した地域ポテンシャルマップの作成と将来の物件の賃料の推定を試みた。

その結果、3章と4章ではDNNは外れ値に大きな影響を受けやすくデータクリーニングの必要があること、東京23区

を対象とした賃料推定モデルを構築するには、町字ダミーを地域属性変数として与えたDNNモデルが適当であることを示した。今後は、規模や用途によって学習器を分ける等RMSEの値を小さくするための工夫を施すこと、ベイズ最適化によりDNNのハイパーパラメータの設定を行うこと、安定した量/長期間のデータを用いて長期の物件賃料推定を試みることで、住宅用物件のデータを用いた同様の分析が必要である。さらに、本論文では同一の建物属性条件下で予測した賃料を地域のポテンシャルと見なしたが、モデルが示す地域ポテンシャルには予測誤差も含まれている。今回はモデル自体の賃料推定精度を高めることでモデルが示す地域ポテンシャルに含まれる予測誤差を小さくしようと試みたが、今後さらに予測誤差を分離する工夫を施す必要がある。

<謝辞>

本研究を行うにあたり、株式会社日建設総合研究所、住友商事株式会社の皆様より有益なアドバイスを頂戴した。ここに記して謝意を申し上げる。

参考文献

- 1) Y. Fukushima, Y. Hasegawa, M. Maeda, T. Seto, H. Omata & Y. Sekimoto. (2017). The development of Japanese city's future simulation system: My city forecast. Free and Open Source Software for Geospatial (FOSS4G) Conference Proceedings, 17:Article 27.
- 2) T. Seto, Y. Hasegawa, Y. Sekimoto & Y. Fukushima. (2015). How will compact city affect me? : urban planning simulation for citizens. CUPUM (Computers in Urban Planning and Urban Management), conference papers, 14:1-15.
- 3) 早川玲理, & 浅見泰司. (2004). 居住者タイプに応じた住環境適地マップの構築. 都市住宅学, 2004(47), 148-153. 4) 鈴木雅智, & 浅見泰司. (2017). 東京大都市圏郊外の中古住宅市場に置ける需給バランス. 都市計画論文集, 52(3), 514-520.
- 5) Mok, H. M., Chan, P. P., & Cho, Y. S. (1995). A hedonic price model for private properties in Hong Kong. The Journal of Real Estate Finance and Economics, 10(1), 37-48.
- 6) Wörzala, E., Lenk, M., & Silva, A. (1995). An exploration of neural networks and its application to real estate valuation. Journal of Real Estate Research, 10(2), 185-201.
- 7) Liu, J. G., Zhang, X. L., & Wu, W. P. (2006, May). Application of fuzzy neural network for real estate prediction. In International Symposium on Neural Networks (pp. 1187-1191). Springer, Berlin, Heidelberg.
- 8) Din, A., Hoesli, M., & Bender, A. (2001). Environmental variables and real estate prices. Urban studies, 38(11), 1989-2000.
- 9) Limsombunchai, V. (2004, June). House price prediction: hedonic price model vs. artificial neural network. In New Zealand Agricultural and Resource Economics Society Conference (pp. 25-26).
- 10) Nghiep, N., & Al, C. (2001). Predicting housing value: A comparison of multiple regression analysis and artificial neural networks. Journal of real estate research, 22(3), 313-336.
- 11) Selim, H. (2009). Determinants of house prices in Turkey: Hedonic regression versus artificial neural network. Expert Systems with Applications, 36(2), 2843-2852.
- 12) Wörzala, E., Lenk, M., & Silva, A. (1995). An exploration of neural networks and its application to real estate valuation. Journal of Real Estate Research, 10(2), 185-201.
- 13) 小林裕治. (2017). 深層学習を用いて地域内のアパート・マンションを同一基準で賃料推定し地図上で可視化する試み. GIS理論と応用 Theory and Applications of GIS, 25(2):64.
- 14) Loh, W. Y. (2011). Classification and regression trees. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1(1), 14-23.
- 15) Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., ... & Zhou, Z. H. (2008). Top 10 algorithms in data mining. Knowledge and information systems, 14(1), 1-37.