

論文発表

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

中市新太

富山県立大学
u020025@st.pu-toyama.ac.jp

はじめに

データの概要とモデルの構築

DNN モデルの結果

他手法との比較

地域ポテンシャルマップ

7.3 まとめ

1.1 本研究の背景と目的

2/18

背景

近年首都圏ならび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、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。

目的

本研究ではオープンデータを活用した物件の推定賃料・地域指標分析システムを構築する。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、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。

はじめに

データの概要とモデルの構築

DNN モデルの結果

他手法との比較

地域ポテンシャルマップ

7.3 まとめ

1.2 研究の概要

3/18

概要

東京 23 区を対象地域とし、

Deep Neural Network, ヘドニックアプローチ, Random Forest Regression 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。賃貸推定モデルを構築するにあたり、緯度経度、区ダミー変数と町字（まちあざ）ダミー変数による東京 23 区の地域性の学習を試みる。

ダミー変数

ダミー変数とは数字ではないデータを 0 か 1 かの値をとる説明変数に変換する手法。

例：

東京都内の区、町字数は 23, 939 のため
説明変数はそれぞれ 23, 939 となる。

はじめに

データの概要とモデルの構築

DNN モデルの結果

他手法との比較

地域ポテンシャルマップ

7.3 まとめ

2.1 データの概要

データ元

at home の不動産データライブラリ、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。

Table 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. 停歩(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. 健ぺい率	19. 容積率	20. 地上階数
21. 地下回数	22. 所在階	23. 築年	24. 築月
25. 登録日付	棟ID		

はじめに

データの概要とモデルの構築

DNN モデルの結果

他手法との比較

地域ポテンシャルマップ

7.3 まとめ

2.2 モデルの構築

Table 2: データ項目

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		

モデル計算

1～3 の物件用途は 3 種の二値変数とする。

9 の導入式：

徒歩（分） + バス（分） × 1.5 + 停徒（分）

モデル作成

層数：3 各層のペーセプトロン数：500 学習率：0.001

8割を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 回学習させる。

残りの 2割のテストデータでモデルの精度を検証する。

はじめに

データの概要とモデルの構築

DNN モデルの結果

他手法との比較

地域ポテンシャルマップ

7.3 まとめ

2.3 モデル作成

データクリーニング

物件面積が、小・中規模であるにもかかわらず 5000 (万/月) の物件 84 件を除いて分析をする。

はじめに

データの概要とモデルの構築

DNN モデルの結果

他手法との比較

地域ポテンシャルマップ

7.3 まとめ

Table 3: クリーニング

	相関関係	2乗平均平方根誤差 (千円/月)	平均絶対誤差 (千円/月)	テストデータの平均賃料 (千円/月)
クリーニングなし	0.007	87101	604	901
クリーニングあり	0.875	292	111	415

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)			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			経度	実数
5. 地上階数	実数	(III)	区ダミー	二値変数
6. 地下階数			町字ダミー	二値変数
7. 所在階	整数			
8. 築年数				

図 1: 属性の説明変数

3.1 地域属性による違い

はじめに

データの概要とモデルの構築

DNN モデルの結果

他手法との比較

地域ポテンシャルマップ

7.3 まとめ

Table 4: 地域属性による違い

	相関関係	2乗平均平方根誤差 (千円/月)	平均絶対誤差 (千円/月)	テストデータの平均賃料 (千円/月)
(I)	0.915	240	91	412
(II)	0.926	226	87	413
(III)	0.937	215	85	414

前提条件

貸事務所, 物件面積 m^3 , 地上階数 8, 地下階数 1, 所在階 1, 築年数 10 年の物件の賃料を考える.

3.2 学習結果

はじめに

データの概要とモデルの構築

DNN モデルの結果

他手法との比較

地域ポテンシャルマップ

7.3 まとめ

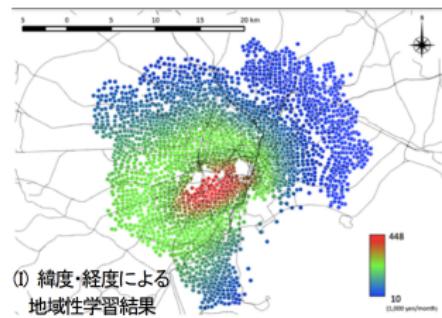


図 2: (I)

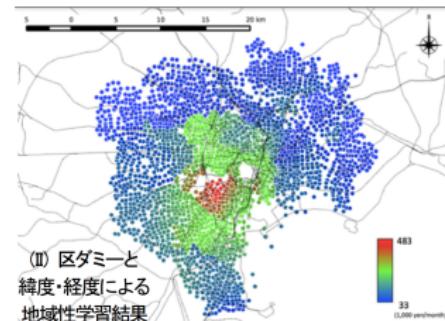


図 3: (II)

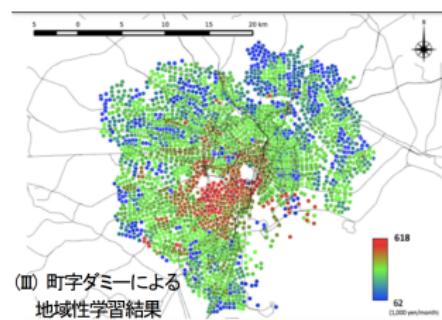


図 4: (III)

- I: 中心部が高くなることが学習できている
- II: 渋谷新宿六本木銀座周辺が高い
- III: 違いが最も大きい
- III の町字ダミーを用いるのが適当といえる

システム設計

単位面積当たりの物件の賃料を被説明変数とし、(III) と同様の説明変数を用いて賃料の線形予測式を作る。

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

予測された単位面積当たりの物件賃料をもとに物件賃料の推定を行った。

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

x_j と y_j は区、町字の二値ダミー変数である。

パラメータ $a_i b_j$ の推定を行う。

はじめに

データの概要とモデルの構築

DNN モデルの結果

他手法との比較

地域ポテンシャルマップ

7.3 まとめ

システム設計

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ため町字を均等に学習させることは難しい。

過学習の可能性が高くなるが、決定木の大きさを 100 に設定する。

はじめに
データの概要とモデルの構築

DNN モデルの結果

他手法との比較

地域ポテンシャルマップ

7.3 まとめ

Table 5: アプローチによる違い

	相関関係	2乗平均平方根誤差 (千円/月)	平均絶対誤差 (千円/月)	テストデータの平均賃料 (千円/月)
DNN	0.937	215	85	414
ヘドニック	0.902	256	98	413
RFR	0.979	120	20	414

4.3 物件賃料と推定賃料の関係

はじめに

データの概要とモデルの構築

DNN モデルの結果

他手法との比較

地域ポテンシャルマップ

7.3 まとめ

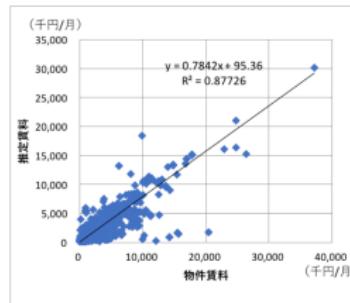


図 5: DNN

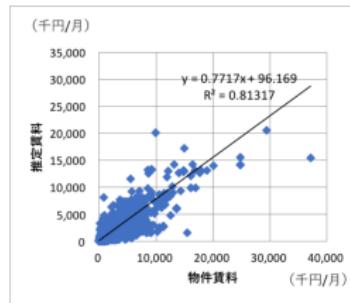


図 6: ヘドニック

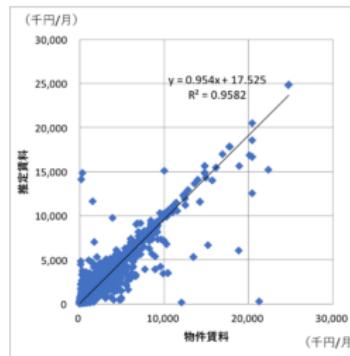


図 7: RFR

ヘドニック：推定誤差が大きい
RFR：推定誤差が小さく、特に 10,000 以上で精度が高い

4.4 他手法の学習結果

はじめに

データの概要とモデルの構築

DNN モデルの結果

他手法との比較

地域ポテンシャルマップ

7.3 まとめ

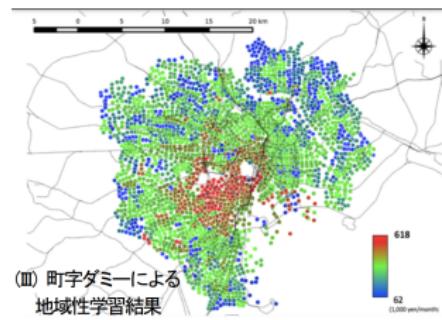


図 8: DNN

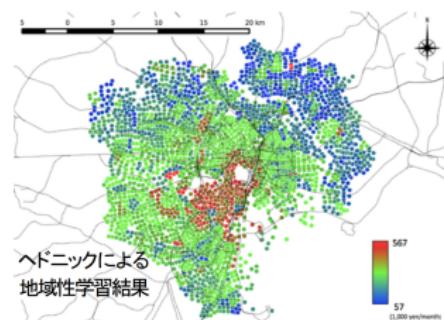


図 9: ヘドニック

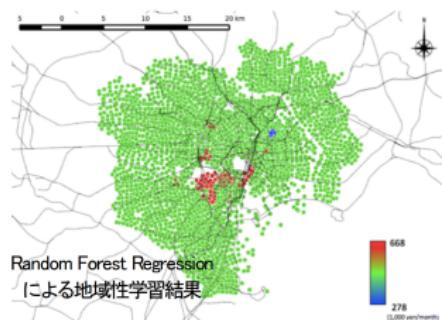


図 10: RFR

他 2 手法と比べると DNN は賃料推定精度も町字による物件賃料の違いもバランスよく学習できていることがわかる。今後は DNN での実験を進めいくこととする。

5.1 時系列変化

時系列変化

ポテンシャルマップが年度によって変化することを表現するため、25. 登録日時から得られる登録年度をデータに加えて学習させる。
地域差をわかりやすくするために物件面積を $400m^3$ に変更。

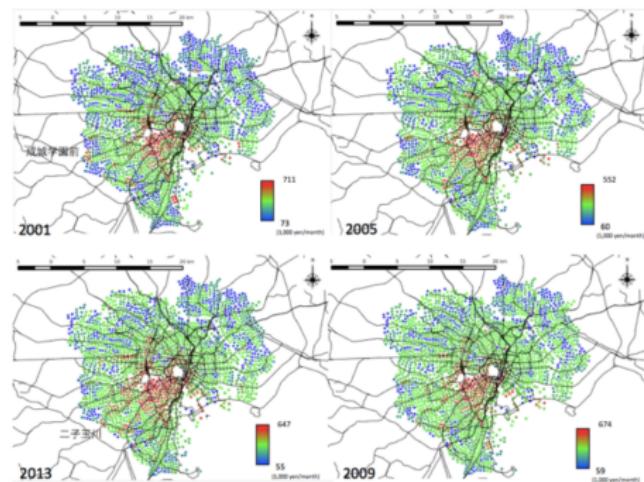


図 11: 時系列変化

はじめに

データの概要とモデルの構築

DNN モデルの結果

他手法との比較

地域ポテンシャルマップ

7.3 まとめ

5.2 経済状況変化

経済状況変化

ポテンシャルマップが経済状況によって変化することを表現するため、登録年度のデフレーターを加えて学習させる。

デフレーターが 1.9 から-2.9 に変化した際の変化率をマップに示す。

はじめに

データの概要とモデルの構築

DNN モデルの結果

他手法との比較

地域ポテンシャルマップ

7.3 まとめ

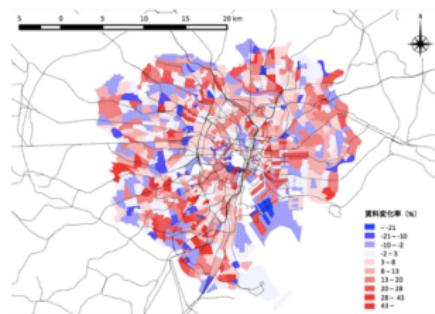


図 12: デフレーター上昇時

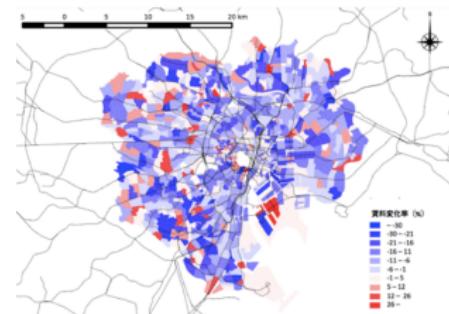


図 13: デフレーター下降時

5.3 変化

はじめに

データの概要とモデルの構築

DNN モデルの結果

他手法との比較

地域ポテンシャルマップ

7.3 まとめ

Table 6: 変化

	相関関係	2乗平均平方根誤差(千円/月)	平均絶対誤差(千円/月)	テストデータの平均賃料(千円/月)
変化なし	0.937	215	85	414
時系列	0.944	212	86	415
経済状況	0.933	225	83	414

将来推定

2001-2005 のデータを使い、2008-2013 を推定する。

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		

図 14: 説明変数

はじめに

データの概要とモデルの構築

DNN モデルの結果

他手法との比較

地域ポテンシャルマップ

7.3 まとめ

5.4 将来推定

将来推定

デフレーターの有無による大きな違いはみられなかった。

はじめに

データの概要とモ
デルの構築

DNN モデルの
結果

他手法との比較

地域ポテンシャル
マップ

7.3 まとめ

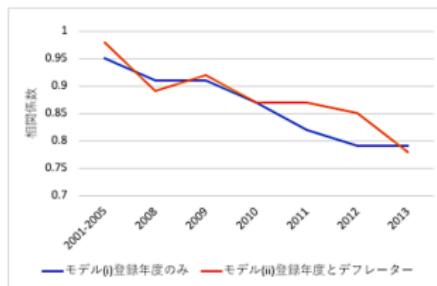


図 15: 相対誤差

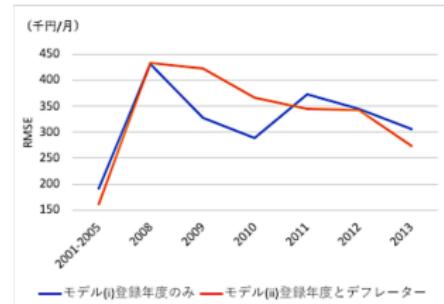


図 16: RMSE

結果

本研究では町字ダミーを利用した DNN モデルが適当であることを示した。
デフレーターによる変化はみられなかった。

今後の展開

RMSE を小さくするため学習器を分けるなどの工夫を行う。
DNN のハイパーパラメータの設定

はじめに

データの概要とモデルの構築

DNN モデルの結果

他手法との比較

地域ポテンシャルマップ

7.3 まとめ