

機械学習と地理空間情報を活用した 周辺の過去の賃料情報を用いた アパート賃料推定

鶴田大 豊島裕樹
株式会社新生銀行

中市新太
富山県立大学 情報システム工学科
u020025@st.pu-toyama.ac.jp

April 14, 2023

研究の背景と目的

2/12

目的

個人のアパートやマンションのローンは、借り手の相続税対策等を目的に、特に地域銀行で増加してきている。先行研究では、築年数の経過にともないアパート収支のみで返済資金を賄えない借り手が増える傾向にあることから、金融機関は金利上昇や空室・賃料低下等のリスクを適切に評価し、わかりやすく伝える必要があるとされている。アパートの収支の評価においては、空室や賃料等のいくつかの要素を検討する必要がある。

データ収集

LIFULLHOME'Sに掲載されている賃貸アパート物件の掲載情報を対象に分析を行う。

分析にあたってマンスリー賃貸および家具家電付き、定期借家権、分譲賃貸の物件を除外する。

アパートを対象とするため、建物階数が3階以下、建物構造が木造と軽量鉄骨および情報を入力した不動産会社が本物件をアパートとみなす物件を対象とする。賃料が30万円以下、平米単価が7000円以下、部屋面積が15m²以上、80m²以下の物件を対象とする。

データの期間は、2009年11月から2019年12月までである。



図 1: LIFULLHOME'S

図 2: 貸貸情報

2 研究の概要

4/12

背景と目的
研究の概要
実験結果
考察と結論

基本変数

設備の情報→面積、構造、築年月、部屋の階数、建物の階数といった情報
バス・トイレ別やオートロック、TVモニター付きインターホン
→0-1変数として取り込む

間取りの変数化→、緯度経度情報を活用した変数

→近隣・周辺の特性として住所、緯度経度、小学校やコンビニ、スーパーからの距離

住所情報を正規表現で分割を行い、都道府県名、市区町村名

交通利便性にかかわる変数→、最寄り駅、最寄り路線、最寄駅からの距離。

この物件のこだわり	<input checked="" type="checkbox"/> 新築	<input type="checkbox"/> 2階以上の物件	<input type="checkbox"/> 南向き	<input type="checkbox"/> 駐車場あり	<input type="checkbox"/> オートロック	<input type="checkbox"/> エアコン
	<input checked="" type="checkbox"/> バス・トイレ別	<input type="checkbox"/> 追焚機能	<input checked="" type="checkbox"/> フローリング	<input type="checkbox"/> ペット相談可		
位置	角部屋					
入居条件	ルームシェア不可、保証人不要、即入居可					
キッチン//バス・トイレ	IHコンロ、コントローラー、システムキッチン、給湯、バス・トイレ別、シャワー付洗面化粧台、温水洗浄便座、浴室乾燥機、洗面所独立					
設備・サービス	防犯カメラ、TVモニタ付インターホン、エアコン、ウォーターサーバー、シューズボックス、CATV、インターネット対応、インターネット使用料無料、プロパンガス、フローリング、室内洗濯機置場、照明器具付き、駐車場あり					
その他	住まい探し特典あり					
備考	【退去時費用 ルームクリーニング費：29,700円、エアコン清浄費1基：11,000円 故意過失が発生した場合、別途請求あり※故意・過失等別途料金】 再契約事務手数料22,000円/2年※バイクの方は駐車場の契約必要					

図 3: 設備情報

2 研究の概要

5/12

賃料計算

周辺賃料は、500m メッシュ、1km メッシュ、5km メッシュ、10km メッシュの過去 1 年、5 年、全期間の賃料の中央値を用いる。

変数は [メッシュ] × [専有面積] × [築年月] として計算する。

例

メッシュ ID が 500 で専有面積 18 平方 m, 築年数 7 年の物件の場合
同じメッシュ ID が 500 内の専有面積が 15-20 平方 m, 築年数が 5-10 年の
住戸の過去 1 年賃料の中央値を算出し、対象物件の説明変数とする。
さらに、緯度経度情報を用いざとも、市区町村や最寄り駅などが同一の物件
の過去賃料を算出することができるため、
これらの単位を基にした過去賃料の中央値も比較対象の変数として活用
する。

2 研究の概要

6/12

基本変数

位置の近い物件の賃料情報を活用するため, kd-tree を用いた k 近傍法により効率的に各評価対象物件の周辺の N 件の賃料情報を平均し変数化する。メッシュ単位の周辺賃料と同様に, 築年数や面積に基づき学習データをいくつかのカテゴリに分け,
カテゴリごとに緯度経度情報などを説明変数, 過去賃料を目的変数とした k 近傍法の学習を行う。

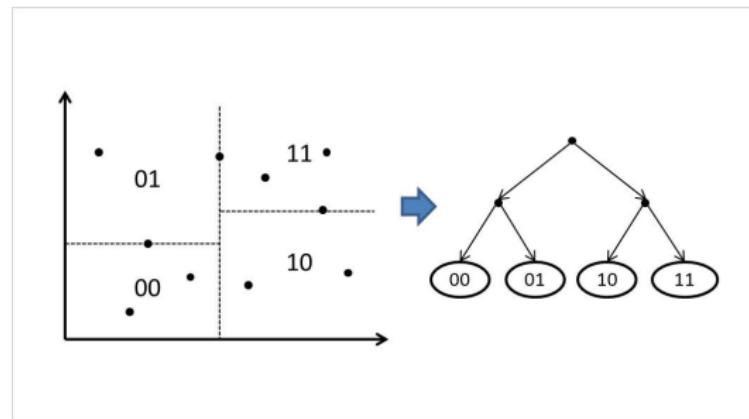


図 4: kd-tree

2 研究の概要

7/12

災害情報

洪水の浸水, 津波の浸水, 土砂災害の警戒区域といったハザードマップの情報を取り、国土交通省国土数値情報から取得し、緯度経度情報を基準に物件に紐づけを行う。

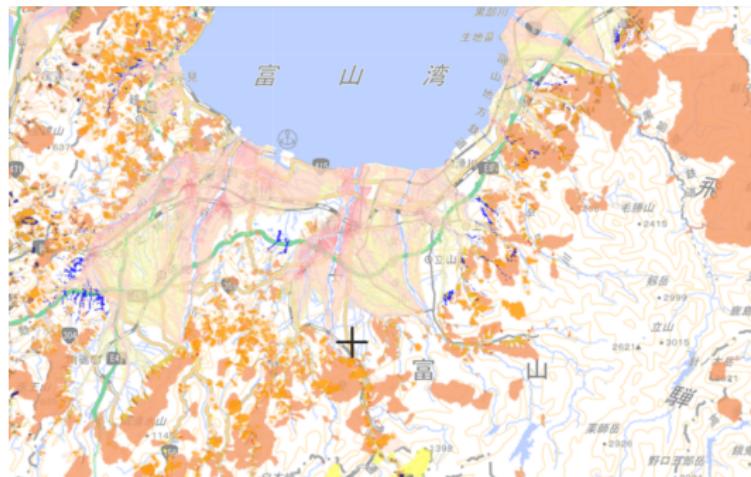


図 5: ハザードマップ

背景と目的

研究の概要

実験結果

考察と結論

2 研究の概要

8/12

機械学習モデル

本研究では、機械学習手法の一種である勾配ブースティング木を用いる。決定木にブースティングと呼ばれるアンサンブルモデルのアプローチを適用した手法である。

ブースティングは、すでに学習済みの予測モデルに対し、正しく分類できていない学習データを正しく分類できるよう新たな予測モデルを追加・構築していく手法である。

ライブラリには LightGBM を使用する。

目的変数は対数賃料として、目的変数を huber 関数とする。

主なハイパーパラメータとしては、学習率を 0.1、葉の数の最大値 (num leaves) を 13、1 つの葉のノードの最小サンプル数 (min data in leaf) を 3 としている。

背景と目的

研究の概要

実験結果

考察と結論

3 実験結果

9/12

賃金誤差

洪水の浸水、津波の浸水、土砂災害の警戒区域といったハザードマップの情報を取り、緯度経度情報を基準に物件に紐づけを行う。

MER の低下幅が大きいものとしては、市区町村等の過去賃料、メッシュ単位の過去賃料と k 近傍法による変数を加えた場合である。

表 1: 変数別のモデル精度：誤差率中央値。市区町村等の過去賃料は市区町村や最寄り駅単位の過去賃料の中央値、緯度経度の過去賃料はメッシュ単位に基づく過去賃料の中央値、k 近傍は k 近傍法により算出した物件周辺の過去平均賃料をそれぞれ変数として使用。

モデル	市区町村等	緯度経度	公示地価	k 近傍	災害	埼玉県	千葉県	東京都	神奈川県	長野県	静岡県	福岡県	全国平均
1	0-1					0.0772	0.0835	0.0898	0.0886	0.0619	0.0683	0.0621	0.0608
2	0-1			○		0.0758	0.0780	0.0858	0.0856	0.0610	0.0672	0.0619	0.0596
3	0-1	原数値				0.0464	0.0540	0.0492	0.0556	0.0521	0.0525	0.0507	0.0489
4	0-1	原数値	○			0.0457	0.0526	0.0480	0.0544	0.0528	0.0521	0.0501	0.0482
5	0-1	原数値	○	○		0.0456	0.0525	0.0479	0.0544	0.0526	0.0524	0.0496	0.0482
6	過去賃料	原数値	○			0.0432	0.0491	0.0451	0.0522	0.0449	0.0493	0.0421	0.0434
7	過去賃料	過去賃料	○			0.0342	0.0397	0.0392	0.0422	0.0337	0.0414	0.0352	0.0373
8	過去賃料	過去賃料	○	○		0.0297	0.0340	0.0318	0.0363	0.0285	0.0376	0.0298	0.0339
9	過去賃料	過去賃料	○	○	○	0.0294	0.0337	0.0317	0.0364	0.0287	0.0376	0.0300	0.0338

図 6: モデルの誤差

3 実験結果

10/12

重要度

SHAP を用い計測した東京都の変数重要度の結果を示す。それによれば k 近傍法により捉えた周辺賃料が相対的に高い結果となっている。それ以外には、面積、部屋階数、公示地価やバストイレ別といった過去周辺賃料では捉えられない変数が上位に位置する。さらに、市区町村別の過去賃料は、重要度が上位であるが、メッシュ単位の過去賃料は上位ではない結果となっている。



図 7: 重要度

3 実験結果

11/12

背景と目的
研究の概要
実験結果
考察と結論

災害による影響

災害情報が比較的 MER の低下に効果のあるモデル 1 と、最も MER が低いモデル 9 の場合で確認する。

全国の平均的な傾向を確認するため、災害情報が無い場合の SHAP 値が各都道府県で 0 になるように引き算の調整を行い、この調整後 SHAP 値を各浸水ランクや浸水の深さ毎に全国で平均する。

この結果を示したのが図 2 である。モデル 1 の場合、洪水の浸水ランクが 11 から 13 の場合、SHAP 値が正であり、浸水が無い場合と比較し賃料へプラスの効果がある結果となった。

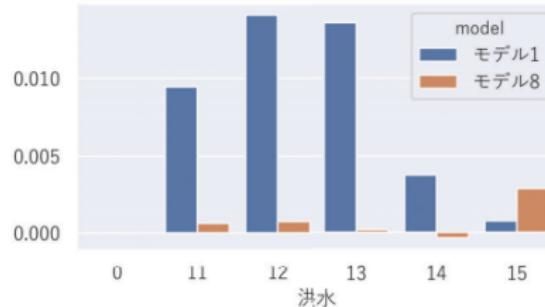


図 8: 洪水

4 考察と結論

12/12

考察

活用した周辺の過去の賃料情報をメッシュ単位や k 近傍法を用い捉えることで、全国で予測精度が向上することを示した。特に k 近傍法を用い捉えた対象物件の周辺過去賃料がモデル上重要な変数となった。一方で、地理空間情報に基づき、災害情報を物件に紐づけることで、物件ごとの災害情報に基づく賃料への影響を確認し、予測精度が改善するかを検証したが、災害情報の予測精度の改善効果は低い結果となった。