

非マンション要素を考慮したLightGBMによる中古マンションの取引価格予測

明治大学 総合数理学部 現象数理学科 池田研究室

1. 背景

マンションの取引において**妥当な価格帯を予測することは重要**

影響を受ける人々

- ・ 買い手、売り手（不動産会社）
- ・ 投資家
- ・ 銀行（ローン査定）

価格の決まり方

- ・ 当事者の話し合い、交渉により決まる。
 - ・ 取引する時期の景気、社会情勢に左右される
- 取引額の規模は大きい

現在、**データ**を用いたマンション価格の予測が行われている
最も精度の良いモデルは**LightGBM**だということは既知

現状

全体的な予測精度は良好

高価格物件の予測が不安定

理想

どの価格帯も予測精度が良好

どの価格帯も予測精度が10%以下

課題

2つ目は全体の予測精度が上がれば、高価格物件に対する予測精度も上がるという考え

- ・ 高価格物件に対する予測の安定
- ・ 全体的な更なる予測精度の向上

- ・ 高価格物件の取引件数が少ない
- ・ オーナー意向による規則性のない値動き
- ・ マンションブランドなどの付加価値
- ・ 元のスケールに戻す際に誤差が顕著になる

2. 先行研究で行われている方法

事例1 モデルはLightGBMを使用

機械学習とルールベースを組み合わせた予測アルゴリズム

1. 予測対象物件と同一マンション内の物件の取引事例が学習セットに含まれる場合以下を行う
 - a. 予測対象と同じ専有面積かつ、同じ所在階の物件の取引事例があれば、そのような事例のうち最も新しい事例の平米単価を予測値とする
 - b. 予測対象と同じ専有面積の物件の取引事例があれば、そのような事例のうち最も新しい事例の平米単価を予測値とする
2. そうでなければ、機械学習アルゴリズムによる予測器で予測を行う

➡ 学習セットに予測対象と同じマンションの物件が1つでもあればその取引事例を使って予測できる。（マンションブランドによる付加価値と物件数の少なさに対応）

事例2

- ・ 「ブランド」という変数を導入
- ・ 右の表のペアで**交互作用項**を作成

2次の交互作用項の追加

- 「総階数」と「部屋の階数」
- 「専有面積」と「総階数」
- 「専有面積」と「部屋の階数」
- 「最寄り駅までの距離」と「専有面積」

➡ 単独では説明できない影響を捉える

LightGBMは自動的に非線形効果を捉えるが、取って交互作用項を明示している

課題 上記の方法ではまだ改善の余地がある

➡ 本研究では先行研究とは異なるアプローチを試みる

3. 目標

非マンション要素を考慮した場合の**予測精度の変化を調べる**
非マンション要素の取引価格に対する**寄与を調べる**

「所在階」や「総階数」など、先行研究で用いられているいくつかの重要な特徴量を含めできていない
その為、先行研究より精度が悪化する可能性があるが、本研究では独自に追加する非マンション要素が予測精度に大きく影響していることが示されればよい。

使用する非マンション要素

- 平均地価** : 23区ごとの土地の平均価格
- 不動産価格指数** : 四半期ごとの住宅価格の平均的な変化
- 金利** : 金銭の貸借に伴う利息の割合

4. 方法

モデル0 : 非マンション要素を**考慮しない**LightGBMモデル

モデル1 : 非マンション要素を**考慮した**LightGBMモデル

使用する精度指標 : MAPE

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100$$

MAPEは外れ値にロバスト

y_i : 実際の価格、 \hat{y}_i : 予測価格

5. 結果

モデル0とモデル1の予測精度を比較

全体の予測精度、価格帯別の予測精度

全体の予測精度	MAPE
モデル0	0.092363
モデル1	0.090663

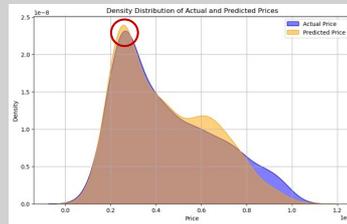
どちらのMAPEが10%未満

モデル0で既に実務で使用できる精度

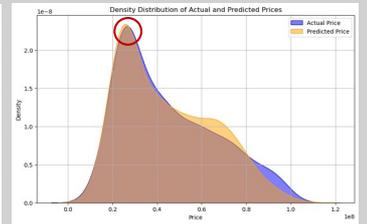
価格帯別の予測精度	モデル0	モデル1
~20,000,000	0.10129	0.08610
20,000,000~50,000,000	0.09088	0.09272
50,000,000~75,000,000	0.08700	0.08321
75,000,000~100,000,000	0.10455	0.09050

結果 モデル1 : 全価格帯でMAPEが10%未満である

モデル0の密度分布



モデル1の密度分布



結果 モデル0 : 中程度の価格帯を重視しすぎている

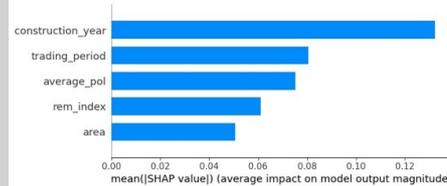
モデル1 : 低価格帯や高価格帯をより重視している

モデル1の特徴量重要度を調査

SHAP値 : 各変数が予測結果に与える影響の平均的な大きさを示す指標

split importance : 分割にどれだけ利用されたか、gain importance : 損失関数をどれだけ小さくしたか

SHAP値に基づく、重要度の高い特徴量5つ



Split Importance	Gain Importance
地区名	地区名
最寄りの駅名	最寄りの駅名
取引時期	建築年
建築年	平均地価
不動産価格指数	専有面積

average_pol : 平均地価、rem_index : 不動産価格指数

結果 平均地価、不動産価格指数が上位にランクイン

6. 結論

- ・ 非マンション要素を加えた事で、どの価格帯もMAPEが10%を切り**全体の予測精度が安定した**
- ・ 平均地価や不動産市場指数といった変数はマンションの取引価格を予測する上で**重要な役割**を果たしている