

# 傾向スコア分析と説明可能 AI 技術を用いた 賃料決定に関する重要部屋配置の特定

菱沼 大輝<sup>1</sup> 尾崎 知伸<sup>1,a)</sup>

**概要：**賃貸物件の賃料は様々な要因により決定される。本論文では、様々な要因の中でも部屋配置に着目し、賃料決定に重要視される構造の特定を試みる。そのアプローチとして (1) 類似物件の特定と因果効果推定を行うことで効果的な構造特定が期待できる傾向スコア分析、および (2) 賃料推定を行う機械学習モデルを対象に予測結果への寄与度が高い構造を抽出する説明可能 AI 技術を利用する。

**キーワード：**傾向スコア分析, 説明可能 AI 技術, 部屋配置

## Investigation of the impact of room properties on rent determination using propensity score analysis and explainable AI techniques

HISHINUMA DAIKI<sup>1</sup> OZAKI TOMONOBU<sup>1,a)</sup>

**Abstract:** The rent of a rental property is determined by various factors. In this paper, we focus on the room layout and attempt to identify the structure considered important in determining the rent. As approaches for this purpose, we utilize propensity score analysis and explainable AI techniques. The former can be expected to effectively identify the important structure through the estimation of causal effects excluding other factors. The latter extracts the structures having high contribution on the prediction by machine learning models for rent estimation. Through the experiments using about 1,800 of rental properties, we succeeded to extract some structure having high impact on rent.

**Keywords:** propensity score analysis, explainable AI techniques, room layout

### 1. はじめに

賃貸物件の賃料は、物件面積や部屋数、部屋配置など、様々な要因から決定される。しかし、これらの要因が賃料決定にどれほど影響しているのかは必ずしも明らかではない。これらの要因の賃料への影響が明確化できれば、適切な賃料を決定する際の指標になることや、設定した賃料に合わせた物件設計支援の実現などが期待できる。実際これまでも、様々な側面から各要因が持つ賃料への影響を特定することが試みられている。例えば文献 [1-3] ではそれぞれ、物件に関わる情報 (面積, 最寄り駅までの距離等),

地震耐久度, 物件の見た目から評価される感性評価等の要因による賃料予測を行い、各要因が賃料に与える影響力を分析している。また文献 [4-6] では、部屋を頂点、部屋の繋がりを辺と考えることで間取り図をグラフに変換し、部屋配置に着目した分析を展開している。一方、文献 [7,8] では、グラフ化された間取り図を対象に、住み心地などの魅力度の観点から特徴的な部分構造を特定する研究が行われている。

これら近年の関連研究と同様、本研究では特に部屋配置に着目し、(1) 対象部屋配置の有無による賃料との因果効果の推定を行う傾向スコア分析 [9], (2) 賃料推定モデルを対象とした説明可能 AI 技術 SHAP [10], (3) 賃料をクラスとするグラフ分類器を対象とした説明可能 AI 技術 GNExplainer [11] を用いた分析を展開する。これらを通

<sup>1</sup> 日本大学文理学部情報科学科  
〒156-8550 東京都世田谷区桜上水 3-25-40  
<sup>a)</sup> ozaki.tomonobu@nihon-u.ac.jp

表 1 物件情報に関する属性

属性	説明
property_type	物件の種類
room_type	間取りの種類
area	物件の面積
floor	物件の階数
shape	物件の縦横比
room_area	部屋の面積
room_shape	部屋の縦横比
position_machine	洗濯機の位置
cnt_window	窓の数
cnt_closet	クローゼットの数
time	最寄り駅までの徒歩距離
hospital_distance	病院までの距離
cvs_distance	コンビニまでの距離
super_distance	スーパーまでの距離

表 2 頂点ラベル

ラベル	説明	ラベル	説明
closet	クローゼット	window	窓
entrance	玄関	corridor	廊下
kitchen	キッチン	cooking_space	流し兼コンロ
wc	トイレ	wash_machine	洗濯機
unit_bus	ユニットバス	wash_room	洗面所
wbed	洋室	jbed	和室
balcony	バルコニー	loft	ロフト

じ、どのような部屋配置があることによって賃料が高くなるのか、安くなるのかといった傾向の特定を試みる。

本論文の構成は以下のとおりである。2章で対象とするデータセットについて言及した後、3章で傾向スコア分析(3.1節)、SHAP(3.2節)、GNNExplainer(3.3節)による分析結果をそれぞれ示す。最後に、4章でまとめを行い、今後の課題を述べる。

## 2. データセット

本研究では、国立情報学研究所のIDRデータセット提供サービスにより株式会社LIFULLから提供を受けた「LIFULL HOME's データセット」[12]を利用する。データセット全体から、東京都世田谷区の桜上水・下高井戸駅を最寄り駅とする1Kもしくは1Rの賃貸物件1,783件を抽出し、分析の対象とする。その際、各物件に対し(1)物件情報に関する属性群 $X_{basic}$ と(2)部屋配置に関する属性群 $X_{graph}$ をそれぞれ準備し、傾向スコア及びSHAPに基づく分析に利用する。また(3)頂点属性付きの間取り図グラフを用い、GNNExplainerによる分析に利用する。以下、それぞれについて説明する。

### 2.1 物件情報に関する属性

本研究で利用する物件情報に関する属性群 $X_{basic}$ を表1に示す。ここで、属性property\_typeは”マンション”もしくは”アパート”，room\_typeは”1K”もしくは”1R”，position\_machineは”部屋の中”もしくは”部屋の外”をそれぞれ表すダミー変数である。また、room\_areaは1k・ワンルームで主要な部屋の面積を、room\_shapeはその部屋の縦横比を表している。なお、対象となる部屋が存在しない場合は、面積・形ともに0とする。

### 2.2 部屋配置に関する属性

本研究では、関連研究[4-8]同様、部屋や空間を頂点、そ

の繋がりを辺とする間取り図グラフ $g = \langle V_g, E_g \rangle$ を用いて部屋配置を表現する。ここで $V_g, E_g$ はそれぞれ、頂点集合、辺集合である。

頂点ラベルの一覧を表2に示す。なお本研究では1Kおよび1Rを対象としており、部屋としてのキッチンと実際に料理を行う空間とを明確に区別し使い分けるため、”kitchen”：料理を行うための部屋(キッチン)と”cooking\_space”：料理器具(流し・コンロ)が配置されている場所とをそれぞれ準備している。また頂点”cooking\_space”には、それを含む頂点(”kitchen”, ”wbed”, ”jbed”など)からの辺を配置する。洗濯機置き場”wash\_machine”も同様であり、それを含む空間・部屋から辺を配置している。

部屋配置に関する属性群 $X_{graph}$ は、間取り図グラフの集合 $G = \{g_1, \dots, g_N\}$ ( $N = 1,783$ )に対する頻出部分グラフパターンを用いて準備する。すなわち、グラフ $p$ がグラフ $g$ の部分グラフであることを $p \sqsubseteq g$ と表記するとき、頻度に関する最小閾値 $\sigma_{min}$ と最大閾値 $\sigma_{max}$ を用い、条件

$$\sigma_{min} \leq sup(p) = |\{g \in G \mid p \sqsubseteq g\}| / |G| \leq \sigma_{max}$$

を満たす頻出部分グラフパターン $p$ の全体集合を $X_{graph}$ とする。なお実際には、頻出部分グラフマイナーgSpan[13]を用いて $\sigma_{min} = 0.2$ を満たすパターンを抽出し、後処理として $\sigma_{max} = 0.9$ を満たさないパターンを排除することで $X_{graph}$ を準備する。また $g \in G$ に対する各属性 $p \in X_{graph}$ の属性値は $\begin{cases} 1 & (p \sqsubseteq g) \\ 0 & (p \not\sqsubseteq g) \end{cases}$ と算出する。

### 2.3 頂点属性付き間取り図グラフ

本研究では間取り図のより自然な表現として、部屋に関連するいくつかの情報を頂点ではなく属性として扱う頂点属性付き間取り図グラフを考える。すなわち、間取り図グラフ $g = \langle V_g, E_g \rangle$ に対し、ラベルが $l \in \{”closet”, ”window”, ”wash_machine”, ”cooking_space”\}$ である頂点とその頂点に連結されている辺を取り除くと共に、各頂点を行、表3に示す属性を列とする頂点属性行列 $X_g$ を追加したグラフ $g' = \langle V'_g, E'_g, X_g \rangle$ を、頂点属性付き間取り図グラフと呼ぶ。なおこのグラフ表現は、深層学習技術の一つの発展であるグラフニューラルネットワークへの入力データ構造に対応するものであり、深層学習を通じた間取り図分析の基礎を与えるものである。

表 3 頂点属性

属性	説明
closet	付属クローゼット数
window	付属窓数
wash_machine	洗濯機付属の有無
cooking_space	流し兼コンロの有無
area	部屋面積
shape	部屋の縦横比

### 3. 重要部屋配置の推定

#### 3.1 傾向スコア分析による重要部屋配置の推定

傾向スコア分析 [9] は、結果・効果 (目的変数) を  $y$ 、共変量 (説明属性) 集合を  $X$ 、 $y$  に対する因果効果の推定対象属性を  $z$  としたとき、 $z$  の因果効果を

$$E = E_{e(X)} [E(y_1 | z = 1, e(X)) - E(y_0 | z = 0, e(X))] \\ \text{where } e(X) = 1/(1 + \exp(-\sum a_n x_n + \gamma))$$

と推定する。ここで  $E[\cdot]$  は期待値を、 $e(X)$  は傾向スコアと呼ばれる確率  $P(z = 1 | X)$  の推定値をそれぞれ表す。なお  $e(X)$  の実体は、パラメタ  $a_n, \gamma$  を伴う  $z$  に対するロジスティック回帰モデルである。

本研究では、部屋配置に関する各属性  $z \in X_{graph}$  に対し、目的変数  $y$  を賃料、共変量  $X = X_{basic} \cup X_{graph}^{-z}$  をその他の属性とし、 $z$  の因果効果推定を行った。ここで  $X_{graph}^{-z} = \{g \in X_{graph} | V_z \cap V_g = \emptyset\}$  は、 $z$  と同一 (ラベルの) 頂点を持たない部屋配置に関する属性  $g$  の集合である。すなわち、ロジスティック回帰に悪影響を与えると考えられる  $z$  と強い関連を持つ属性を排除している。

また因果効果  $E$  の推定にはいくつかの方法が提案されているが、今回は Caliper 法を採用した。Caliper 法では、 $z = 1$  である各データ  $d_1$  に対し、距離上限を設定した上で傾向スコアの値が最も近い  $z = 0$  であるデータ  $d_2$  を割り当てるとともに、両者の目的変数の差  $y_1 - y_0$  の期待値を推定結果  $E$  とする。詳細については、文献 [9] 等を参照されたい。

表 4 に、大きな因果効果が推定された部屋配置属性 (因果効果絶対値の上位 10 件) を示す。結果より、"unit\_bus" を含む配置の因果効果が大きいこと、特に "unit\_bus-kitchen" を含む物件は賃料が安くなる傾向があることが確認できる。また、同じく台所を含む配置でも、"bathroom-kitchen" を含む物件の賃料が高くなる傾向が認められる。両者より、台所に直結する風呂設備の違いにより、大きな賃料差が生まれることが推察される。

上記に加え、"wbed" と "cooking\_space" や "entrance" の繋がりがあある場合に賃料が安くなる傾向や、"closet" が付属することによって賃料が高くなる傾向が確認できる。

#### 3.2 SHAP による重要部屋配置の推定

説明可能 AI 技術の一つである SHAP [10] は、限界貢献

表 4 傾向スコア分析により推定された重要部屋配置

属性	因果効果 E
closet-kitchen-unit_bus	-7251.55
entrance-kitchen-unit_bus	-6718.66
unit_bus-kitchen-wbed	-4607.48
kitchen-unit_bus	-4311.22
cooking_space-wbed	-4276.95
cooking_space-kitchen-bathroom	4257.02
wbed-entrance	-4162.89
unit_bus	-3713.51
bathroom-kitchen-wc	3648.62
wbed-closet	3093.81

表 5 賃料推定モデルの決定係数

	Random Forest	XGBoost	LightGBM
訓練データ	0.98	0.91	0.97
検証データ	0.91	0.87	0.89
テストデータ	0.90	0.89	0.97

度の期待値

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} (f(S \cup \{i\}) - f(S))$$

を算出することで、機械学習モデル  $f$  に対する特徴量  $i \in F$  ( $F$  は特徴量の全体集合) の影響力を定量化する。

本研究では、属性集合  $F = X_{basic} \cup X_{graph}$  を用いて賃料を推定する機械学習モデル  $f$  として Random Forest [14]、XGBoost [15]、LightGBM [16] の 3 種を採用し、各部屋配置に関する属性  $i \in X_{graph}$  の SHAP 値を算出する。

各モデル  $f$  の決定係数  $R^2$  を表 5 に、また各モデルにおける SHAP 値の上位 10 属性を表 6 に示す。

結果より、"bathroom" 単体、もしくは "bathroom" を含む部屋配置が各モデルの上位に現れており、その影響が大きいことが分かる。また "kitchen" は、"unit\_bus" や "wc" と共に現れる傾向が認められ、賃料に関連が強いことが伺える。同様に "balcony" を含む属性も影響力が大きいことが確認できる。一方、Random Forest モデルを採用した場合、"unit\_bus" を含む構造は常に賃料に負の影響を与えていることが分かる。

#### 3.3 GNNExplainer による重要部屋配置の推定

GNNExplainer [11] は、グラフニューラルネットワーク (GNN) に対する局所的な説明を与える手法であり、各入力事例において予測に寄与する部分構造や特徴を抽出する。具体的には、入力グラフ  $G$  に対する学習済み GNN モデルの予測値を  $Y$  としたとき、予測値との相互情報量

$$I(Y, (G_s, X_s)) = H(Y) - H(Y | (G_s, X_s))$$

を最大化する部分グラフ構造  $G_s$  と頂点特徴量  $X_s$  を算出する。

本研究では、頂点属性付き間取り図グラフに対してその

表 6 SHAP により推定された重要部屋配置

Random Forest		XGBoost		LightGBM	
属性	SHAP 値	属性	SHAP 値	属性	SHAP 値
bathroom	1878.78	bathroom	2274.98	bathroom	1555.72
bathroom-washroom	1579.53	bathroom-kitchen	1262.67	bathroom-kitchen	795.74
cooking_space-kitchen-wc	688.43	cooking_space-kitchen-wc	480.93	balcony-wbed-kitchen	379.32
entrance-kitchen-wc	457.28	balcony-wbed-closet	471.23	entrance-wbed	-329.20
balcony	322.39	bathroom-washroom	315.46	wbed-balcony	290.05
wbed-entrance	-316.17	balcony-wbed-kitchen	239.93	cooking_space-kitchen-wc	281.45
kitchen-unit_bus	-321.70	wbed-balcony	234.05	bathroom-washroom	240.33
cooking_space-kitchen-unit_bus	-302.13	entrance-kitchen-wc	-221.14	balcony	228.76
cooking_space-kitchen-bathroom	289.92	kitchen-entrance-closet	192.76	entrance-kitchen	202.42
balcony-wbed-kitchen	205.77	wash_machine	167.62	kitchen-entrance-closet	179.01

表 7 賃料区分

クラス	区分範囲
0	5万5千未満
1	5万5千以上6万未満
2	6万以上7万5千未満
3	8万以上10万未満
4	10万以上

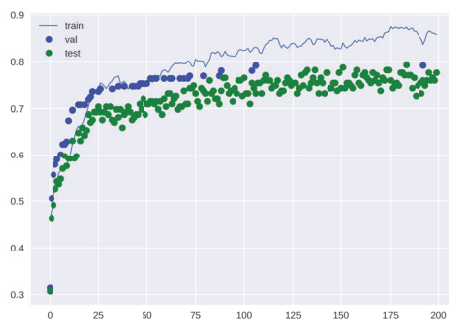


図 1 GNN モデルの精度

賃料区分 (表 7 参照) を予測する GNN モデルを構築し、各賃料区分に対してそれぞれ 5 件の間取りを象に重要属性の特定を試みた。なお図 1 に示す通り、GNN モデルの精度はおよそ 0.8 程度である。

実験結果より、賃料区分が 0~2 である間取りを対象とした場合、“entrance-wbed”や“unit\_bus-kitchen”、“unit\_bus-wbed”を含む構造が重要構造として推定されることが多いことが確認された。

一方、賃料区分が 3~4 である間取りを対象とした場合、“kitchen”を含む構造が数多く抽出される結果となった。特に“kitchen-bathroom”は、高賃料区分への分類根拠とされることが多く、この構造を含む低賃料物件が、高賃料区分へと誤分類されることも確認された。

3 つの手法を比較し、共通点として、“bathroom”、“unit\_bus”の違いが賃料決定に大きな影響を与えている

ことが示唆された。特に、“bathroom”、“unit\_bus”それぞれが“kitchen”と繋がっている部屋配置が賃料決定に大きな影響を与えていることが確認された。また、SHAP では“balcony”を含む構造が重要部屋配置であると特定されたが、傾向スコア分析・GNNExplainer では“balcony”を含む構造が重要部屋配置であると特定されなかった。このことから、“balcony”は賃料決定に大きな影響を与えるが、水回りに関する構造ほどの影響力がないことが確認できる。

#### 4. まとめ

本研究では、賃料へ影響を与える部屋配置の特定を目的とし、1K・1R の賃貸物件を対象とした傾向スコア分析・SHAP・GNNExplainer による分析を展開した。その結果、特に水回りに関する構造が賃料への影響が高いことが示唆された。

今後の課題としては、2LDK、3LDK などのより大きな間取りを持つ物件を対象とした分析が挙げられる。加えて、GraphSVX [17] に代表される、グラフ構造を対象とした説明可能 AI の最新技術の利用も重要な課題の一つである。

#### 謝辞

本研究では、国立情報学研究所の IDR データセット提供サービスにより株式会社 LIFULL から提供を受けた「LIFULL HOME'S データセット」を利用した。

#### 参考文献

- [1] 阿部 成治, 石崎 幸司: 首都圏における民間賃貸住宅家賃の重回帰分析, 都市住宅学, 19:39-44, 1997.
- [2] 山鹿 久木, 中川 雅之, 齊藤 誠: 地震危険度と家賃耐震対策のための政策的インプリケーション, 日本経済研究, 46:1-21, 2002.
- [3] 瀧澤 重志, 材木 敦史, 加藤 直樹, 具 源龍: 新橋に立地するオフィスビルの感性評価を考慮した賃料分析, 日本建築学会計画系論文集 (627):1053-1059, 2008.
- [4] 瀧澤 重志, 吉田 一馬, 加藤 直樹: グラフマイニングを用いた室配置を考慮した賃料分析: 京都市郊外の 3LDK を中心とした賃貸マンションを対象として, 日本建築学会環境系論文集 (623):139-146, 2008.



- [5] 尾崎 知伸, 小黒 淳斗: 頻出部分グラフを用いた賃料分析, 人工知能学会第 111 回知識ベースシステム研究会, pp.13–16, 2017.
- [6] 長谷川 優也, 尾崎 知伸: 部屋配置とその出現数に着目した二段階賃料推定, 人工知能学会 第 21 回インタラクティブ情報アクセスと可視化マイニング研究会, pp.80–87, 2019.
- [7] 橋原 太郎, 汪 雪, 山崎 俊彦: 不動産間取り図を対象にした物件魅力度データセットの構築と分析, 日本人工知能学会第 34 回大会, 2P6-GS-13-05, 2020.
- [8] 橋原 太郎, 汪 雪, 山崎 俊彦: 不動産間取り図の魅力度データを使った「住み心地」とグラフ構造の分析, 日本人工知能学会第 35 回大会, 4F3-GS-10n-04, 2021.
- [9] P. R. Rosenbaum and D. B. Rubin: The central role of the propensity score in observational studies causal effects, *Biometrika*, 70(1):41–45, 1993.
- [10] S. M. Lundberg and S.-I. Lee: A Unified Approach to Interpreting Model Predictions, *Advances in Neural Information Processing Systems 30*, 2017.
- [11] Z. Ying, D. Bourgeois, J. You, M. Zitnik, and J. Leskovec: GNNExplainer: Generating Explanations for Graph Neural Networks, *Advances in Neural Information Processing Systems 32*, 2019.
- [12] 株式会社 LIFULL: LIFULL HOME'S 賃貸物件スナップショットデータ, 国立情報学研究所情報学研究データリポジトリ (データセット), 2015. (<https://doi.org/10.32130/idr.6.1>)
- [13] X. Yan and J. Han: gSpan: graph-based substructure pattern mining, *Proc. of the 2002 IEEE International Conference on Data Mining*, pp.721–724, 2002.
- [14] L. Breiman: Random Forests, *Machine Learning*, 45(1):5–32, 2001.
- [15] T. Chen and C. Guestrin: XGBoost: A Scalable Tree Boosting System, *Proc. of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.785–794, 2016
- [16] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu: LightGBM: A Highly Efficient Gradient Boosting Decision Tree, *Advances in Neural Information Processing Systems 30*, 2017.
- [17] A. Duval, F. D. Malliaros: GraphSVX: Shapley Value Explanation for Graph Neural Networks, *Machine Learning and Knowledge Discovery in Databases. Research Track. ECML PKDD 2021*, 2021.