原著論文

賃料予測モデルにおける間取り図の影響分析[†]

服部 凌典^{*1}・岡本 一志^{*1}・柴田 淳司^{*2}

間取り図が賃料予測に与える影響を明らかにするために、間取り図を考慮した賃料予測モデルを構築し、間取り図の有無による予測誤差を検証する。間取り図の特徴抽出器には、主成分分析、Bag of Features (BoF)、Fisher Vector (FV) を適用し、予測器には線形回帰と LightGBM を採用する。LIFULL HOME'S データセットを使用した賃料予測実験から、線形回帰においては間取り図を考慮することで、全てのカテゴリ（都道府県と間取り規格の組み合わせ）で予測誤差の 95% 信頼区間が短くなり、さらに、カテゴリによっては予測誤差の平均値が改善することを確認している。また、検討した 3 つの間取り図の特徴抽出器の中で BoF が最も予測精度が優れているといえる。

キーワード：賃料予測、間取り図、主成分分析、Bag of Features、Fisher Vector

1. はじめに

間取り図は、間取りに関する情報を幾何学に基づいた図で人工的に表現したものであり、日本の多くの賃貸物件で用意されている。住宅改良開発公社の調査研究によると、賃貸物件の住み替えを考えている人が別の物件に引っ越す 1 番の理由は間取りとされている [1]。また、賃借人は賃貸物件を探す際に間取り図を見て選別する習慣がある [2]。賃貸物件において、全ての属性が同じ物件はほぼ存在せず、例えば、間取りやその規格（1K や 1R など）、物件の外観や内観の状態、立地などの属性は賃貸物件間で異なる。同一の間取り規格であっても、部屋の配置や面積などは物件ごとに異なる。

賃料は一般に、賃貸人側が積算法や収益分析法、賃貸事例比較法などを利用して決定しており [3]、その決定には占有面積や立地などの物件属性が影響を与えていると考えられている [4–6]。特に、賃貸事例比較法は対象物件と類似した物件を選定し、類似物件と対象物件間の属性の比較と補正により賃料を求める手法である。物件属性が賃料の増減に与える影響の推定では、ヘドニック・アプローチが利用される場合がある。ヘドニック・アプローチとは、ある商品の価格を当該商品の属性の価値に関する集合とみなし、商品価格の予測モデルを線形回帰により構築する方法である [7]。賃料を目的変数とした線形回帰モデルを考えることで賃料の予測や、賃料に影響を与えている物件属性の推定ができる。文献 [8] では、賃料予測モデルの説明変数に占有面積や築年数などの物件属性と物件の内観や外観、間取り図などの画像を与えることで、賃料予測の精度向上を目指している。データセットとして京都府の 3K、3DK、3LDK の集合住宅 996 件を使用し、間取り図を考慮しない賃料モデルと考

慮した賃料予測モデルの予測精度を比較している。この研究では、間取り図を考慮した賃料モデルの予測精度の平均絶対誤差が 5,298 円、間取り図を考慮しない賃料モデルの予測精度の平均絶対誤差が 5,985 円であり間取り図が賃料予測モデルの予測精度改善に寄与していることを明らかにしている。文献 [8] では、賃料予測モデルに間取り図を考慮する手法として、間取り図に部屋情報をラベル付けしてグラフ化し、そのグラフの部分グラフの数え上げを行っている。この手法では、間取り図をグラフ化する際に人手でラベル付けする必要がある。そのため、大規模なデータセットを用いることが困難である。本研究ではラベル付けする必要がない間取り図からの特徴抽出方法を適用した賃料予測モデルを構築し、間取り図の利用が賃料予測に与える影響を明らかにする。

間取り図を考慮した賃料予測モデルでは、間取り図から抽出した特徴量を利用する。特徴抽出器として、Principal Component Analysis (PCA)、Bag of Features (BoF) [9]、Fisher Vector (FV) [10] の適用を検討し、賃料予測モデルに適した間取り図の特徴抽出器を明らかにする。また、予測器には線形回帰と LightGBM [11] を用いる。

構築した賃料予測モデルの予測精度は、LIFULL HOME'S データセットの北海道、東京都、愛知県、大阪府、福岡県の賃貸物件 2,244,867 件を用いた実験により評価する。使用データセットは都道府県と間取り規格（K、R、DK、LDK）ごとに分割する。分割したデータセットは学習データとテストデータに別け、学習データに交差検証法を適用し、賃料予測モデルのハイパーパラメータを探索する。探索したハイパーパラメータとテストデータを用いて予測精度を評価する。加えて、線形回帰においては回帰係数と説明変数の内積の値を分析し、LightGBM においては変数重要度を分析することで間取り図の影響を明らかにする。

本論文の構成は次のとおりである。2 節では、賃料予測と分析に関する研究と間取り図を対象とした研究について述べる。3 節では、検討する予測モデルについて説明する。4 節では、予測精度の評価実験の詳細と評価指標を示す。5 節では、実験結果を踏まえ、間取り図が賃料予測に与える影響についての考察を述べる。

[†] Impact Analysis of Floor-Plan Images for Rent-Prediction Model
Hattori RYOSUKE, Kazushi OKAMOTO, and Atsushi SHIBATA

^{*1} 電気通信大学 大学院情報理工学研究所 情報学専攻
Department of Informatics, Graduate School of Informatics and Engineering, The University of Electro-Communications

^{*2} 産業技術大学院大学 産業技術研究所
Graduate School of Industrial Technology, Advanced Institute of Industrial Technology

図 1 賃料予測モデルの予測処理の概要図

2. 関連研究

賃料予測・分析の研究において、賃貸物件の賃料は占有面積や築年数、立地などの物件属性の集合で表され、目的変数を賃料とした回帰モデルの構築により、賃料に与える物件属性の影響を調べた研究が報告されている [12]. 文献 [12] では、西武池袋沿線の 746 件の賃貸物件を占有面積を基準に単身者用、夫婦用、家族用の 3 種類に層別し、層ごとに線形回帰モデルを構築している. 3 つの線形回帰モデルの回帰係数の比較により、層ごとに賃料に影響を与える物件属性の違いを検証している. 説明変数 (物件属性) には、占有面積、築年数、階数に加え、新築フラグ、ペットフラグ、楽器フラグ、ロフトフラグ、リフォームフラグなどの 27 属性が用いられている.

物件属性として間取り図などの画像を用いる場合、画像から特徴量を抽出する必要があるが、その際に 2 つの手法が考えられる. 1 つはグラフを用いた手法である [8, 13]. 文献 [8, 13] では、玄関や居間などの部屋を頂点とし、そのつながりを辺とするラベル付き無向グラフで間取り図を表現している. このグラフの部分グラフの出現数ベクトルにより間取り図の特徴量表現を行っている. 文献 [8] では、京都府の 3K, 3DK, 3LDK のマンションを対象に、間取り図を前述の方法で利用した線形回帰モデルを構築し、賃料における間取りの影響を調べている. 文献 [13] では、まず占有面積や立地などの物件属性と最寄り駅単位の賃料の相場を説明変数として与えて賃料を予測している. その後、この予測賃料と真の賃料との残差を目的変数とし、部分グラフの出現数を入力とした 2 段階の賃料予測モデルを提案している. もう 1 つは画像に Convolutional Neural Network (CNN) や Speeded Up Robust Features (SURF) [14] を適用し、画像から直接特徴量を抽出する方法である [15, 16]. 文献 [15] では、外観や内観などの最大 5 枚の画像を優先度に基づき、順に 1 枚の画像に結合し、結合した画像と占有面積や築年数などの 8 つの物件属性を入力とする CNN を用いた賃料予測モデルを提案している. 文献 [16] では、SURF を外観や内

観などの 4 枚の画像に適用することで特徴量を抽出し、占有面積や築年数などの物件属性との結合により入力ベクトルとしている. この入力ベクトルをニューラルネットワークや Support Vector Regression に与え、物件の販売価格を出力するモデルを提案している.

予測モデル一般において、大量の学習データを準備することで精度向上が期待できる. しかしながら、文献 [8, 13] のグラフを用いる手法では、事前に間取り図の部屋配置を手でアノテーションする必要があるため学習データの準備コストが高い. 本研究は文献 [15, 16] と同様に、間取り図から抽出した特徴量と物件属性を結合したベクトルを入力とする予測モデルを考える.

3. 構築する賃料予測モデル

本研究では、間取り図を考慮した賃料予測モデルと考慮しないモデルを構築する. その比較により、間取り図の使用が賃料予測に与える影響を明らかにし、賃料予測モデルに適した間取り図の特徴抽出器を検証する. 図 1 は PCA, BoF, FV を用いて構築する賃料予測モデルの概要図である. 概要図において、物件属性 (築年数や階数、占有面積など) を $\mathbf{a} \in \mathbb{R}^m$ 、間取り図を $\mathbf{f} \in \mathbb{R}^n$ 、 y を賃料、関数 ϕ_θ をパラメータ θ を持つ特徴抽出器とし、 g を予測器とする. ここで、本研究では、グレースケール変換を適用した間取り図を $\mathbf{f} \in \mathbb{R}^{n_r \times n_c (=n)}$ 次元の縦ベクトルとして扱う. 例えば、本実験で扱うデータセットでは、 120×120 ピクセルの間取り図が含まれており、本実験ではこれを (120×120) 次元の縦ベクトルで表現する. 本研究では、間取り図を考慮する場合は $\mathbf{x} = [\mathbf{a}, \phi_\theta(\mathbf{f})]$ を、考慮しない場合は $\mathbf{x} = \mathbf{a}$ を説明変数として扱う.

3.1 予測器

本研究では、図 1 の予測器には、説明変数の回帰係数が算出できる線形回帰

$$\hat{y} = \mathbf{a}^T \mathbf{x} + \beta \quad (1)$$

と変数重要度が算出できるアンサンブル決定木

$$\hat{y} = \sum_{i=1}^{|I_t|} t_i(\mathbf{x}) \quad (2)$$

を適用する。ここで、 $\hat{y}$ を予測賃料、 $\alpha \in \mathbb{R}^{m+d}$ を回帰係数、 β を定数項、 t_i を決定木とし、学習データから損失関数を最小化するように推定する。アンサンブル決定木とは、複数の決定木を組み合わせた予測モデルである。決定木の予測値は未知物件が属する葉の重み $t_i(\mathbf{x})$ となる。本研究では、アンサンブル決定木に LightGBM [11, 17] を採用する。LightGBM とは、勾配ブースティングを適用したアンサンブル決定木のひとつである。LightGBM では、モデルの学習過程において、勾配ブースティングに Leaf-Wise を用いることで学習コストの削減を図っている。さらに、学習データの特徴量を階級に分けてヒストグラム化するため、決定木の厳密な分岐点を探索する必要がなく、計算コストの削減が期待できる。また、LightGBM における予測値は未知物件が各決定木に属する葉の重みの合計値となる。線形回帰では、回帰係数と説明変数の内積により間取り図の価値を分析できる。アンサンブル決定木では、間取り図の影響度を変数重要度により分析できる。

予測誤差の理解のしやすさの観点から予測器の損失関数を平均絶対パーセント誤差 (Mean Absolute Percentage Error : MAPE)

$$\frac{100}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (3)$$

とする。ここで、 N はデータ数を表す。

3.2 間取り図からの特徴抽出

間取り図を考慮した賃料予測モデルでは、間取り図を物件属性 (築年数や階数、占有面積など) と組み合わせるために、間取り図の特徴抽出器 ϕ_θ を構築する必要がある。ここで、 θ はパラメータとする。特徴抽出器のパラメータ θ は学習データ $L = \{(a_1, f_1, y_1), \dots, (a_s, f_s, y_s)\}$ を用いて求める。未知の物件 (a_*, f_*) に対する賃料予測処理では、求めた θ を用いて f_* から特徴量 $\phi_\theta(f_*)$ を抽出する。本研究では、間取り図が賃料予測に与える影響を明らかにすることを目的としており、賃料予測に影響を与えている間取り図の特徴を把握できる特徴抽出器 ϕ_θ が望ましいため、特徴抽出器 ϕ_θ として PCA, BoF, FV を適用する。

3.2.1 PCA に基づく特徴抽出

学習データ L の間取り図の集合を $\{f_1, f_2, \dots, f_s\}$, $f_i \in \mathbb{R}^n$ とする。 f_i は縦ベクトルとし、 f_i を並べた行列を $F = [f_1, f_2, \dots, f_s]^T$ とする。 F の平均ベクトルは

$$\bar{f} = \frac{1}{s} \sum_{i=1}^s f_i = \frac{1}{s} F^T \mathbf{1}_s \quad (4)$$

となる。ここで、 $\mathbf{1}_s = (1, 1, \dots, 1)$ は要素がすべて 1 の s 次元ベクトルである。

PCA では、まず F の列ごとの平均が 0 となるように F を $\mathbf{1}_s \bar{f}^T$ だけ移動させた $F - \mathbf{1}_s \bar{f}^T$ を特異値分解することで

(a) 入力画像

(b) 復元画像 (累積寄与率: 0.903)

図2 PCAによる間取り図の次元削減と復元例 (次元数: 1,028)

$$F - \mathbf{1}_s \bar{f}^T = A \Lambda P^T \quad (5)$$

となる行列 A , Λ , P を求める。このとき、 $A \in \mathbb{R}^{s \times s}$, $P \in \mathbb{R}^{n \times n}$, $A^{-1} = A^T$, $P^{-1} = P^T$, $n < s$ を満たし、 Λ は $\Lambda \in \mathbb{R}^{s \times n}$ かつ非対角成分が 0 の行列とする。一方で、 F の共分散行列の固有値分解は $F = P^T \Sigma P$ であり、 $\Sigma = \Lambda^2$ の関係がある。 P は固有ベクトル $u \in \mathbb{R}^n$ を横に並べた行列を意味している。 Σ の d ($d < n$) 個の固有値を $\lambda_1, \lambda_2, \dots, \lambda_d$ とすれば、固有値に対応する固有ベクトル $u_1, u_2, \dots, u_d$ が決まり、 d 次元の主成分空間が得られる。 i 番目の固有値を固有値の総和で割った

$$\frac{\lambda_i}{\sum_{j=1}^n \lambda_j} \quad (6)$$

を寄与率といい、1 から d 番目までの寄与率の総和

$$\frac{\sum_{j=1}^d \lambda_j}{\sum_{j=1}^n \lambda_j} \quad (7)$$

を累積寄与率という。 d 個の固有ベクトルを列に並べた行列を $\theta = [u_1, u_2, \dots, u_d]$, $\theta \in \mathbb{R}^{n \times d}$ とすれば、間取り図の特徴量は

$$\phi_\theta(f) = \theta^T f \quad (8)$$

で得られる。未知の物件では、学習用の行列 F で推定した θ を用いて、図 1(a) の手順で間取り図の間取り図の特徴量 $\phi_\theta(f_*)$ に変換する。一般に、次元削減後のデータは、元データに含まれているノイズを排除した情報を保持していると考えられる。また、 $\theta \phi_\theta(f_*) + \bar{f}$ により、次元削減後の間取り図を図 2 のように近似的に復元でき、予測器から得られた特徴量の重要度との関係を確認できる。これにより、賃料予測に影響を与えている間取り図の特徴量を可視化し分析することができる。例えば、本実験では、 (120×120) 次元の縦ベクトルで表現されている間取り図 f に対して、 $(120 \times 120) \times k$ 次元の θ との内積をとることで間取り図 f を k 次元の特徴量で表現している。

3.2.2 局所特徴量クラスタリングに基づく特徴抽出

間取り図は物件ごとに縮尺に違いがある。加えて、「バルコニーが必ず下側になる」といった描画規則がなく、間取り図の

向きに統一性がない。そこで、局所特徴抽出器を用いる。局所特徴抽出器はスケール変化や、回転に不変な特徴量を抽出することができるため、間取り図は物件ごとに縮尺に違いや間取り図の向きに統一性の問題を解決できる。局所特徴抽出器には SURF と Scale-Invariant Feature Transform (SIFT) [18] があるが、本研究では SIFT の改良版であり、間取り図からの局所特徴量の抽出速度が速い SURF を採用する。SURF により検出した特徴点群と間取り図を対応づけることで、賃料予測に影響を与える間取り図の特徴を分析できる。ただし、検出される特徴点の数は間取り図ごとに異なるため、特徴点を表現するベクトル集合（局所特徴量群）を固定長の特徴量に変換する必要がある。

本研究では、局所特徴量群を固定長の特徴量に変換する手法として BoF と FV を適用する。間取り図 1 枚の局所特徴量群を $D_i = \{d_1, d_2, \dots, d_j, \dots\}$, $d_j \in \mathbb{R}^k$, 学習データ L に含まれる間取り図の局所特徴量群の集合を $D = D_1 \cup D_2 \cup \dots \cup D_s$ とする。ただし、任意の $D_i \cup D_j$ に重複は生じないものとする。

BoF [9] では、 D に k -means 法を適用し、クラスターの割り当て $\psi_\theta: \mathbb{R}^k \rightarrow \{0, 1\}^d$ を学習する。ここで、 d はクラスター数を意味する。学習した ψ_θ を用いて、

$$\phi_\theta(D_i) = \frac{1}{|D_i|} \sum_{d \in D_i} \psi_\theta(d) \quad (9)$$

とすることで間取り図を $\phi_\theta(D_i) \in \mathbb{R}^d$ に変換する。予測処理では、図 1(b) の手順で未知の間取り図 f_* から局所特徴量群 $D_* = \{d_{*1}, d_{*2}, \dots, d_{*j}, \dots\}$ を抽出し、 $\phi_\theta(D_*) \in \mathbb{R}^d$ に変換する。BoF は類似した局所特徴量を最近傍の学習済みの ψ_θ に割り当てている。

FV [10, 19] は、学習データ L に含まれる間取り図の局所特徴量群の集合 D を代表モデルとし、この代表モデルの確率密度関数 $P(D|\theta)$ を混合ガウス分布モデルを用いて学習する。 $\theta = \{\mu_j, \Sigma_j, \pi_j\}_{j=1, \dots, d}$ とし、ガウス分布の混合数を d 、平均ベクトルを $\mu_j \in \mathbb{R}^k$ 、共分散行列を $\Sigma_j \in \mathbb{R}^{k \times k}$ 、混合比を $\pi_j \in \mathbb{R}$ とする。次に全ての D_i について代表モデルとの差分（勾配ベクトル）

$$g_\theta(D_i) = \nabla_\theta \log P(D_i|\theta) \quad (10)$$

を求める。この $g_\theta(D_i)$ をフィッシャー情報行列

$$F_\theta = E_{D_i} [\nabla_\theta \log P(D_i|\theta) \{\nabla_\theta \log P(D_i|\theta)\}^T] \quad (11)$$

で正規化することで、フィッシャーベクトル

$$\phi_\theta(D_i) = F_\theta^{-\frac{1}{2}} \nabla_\theta \log P(D_i|\theta) \quad (12)$$

を得る。予測処理では、図 1(c) の手順で未知の間取り図 f_* から局所特徴量群 $D_* = \{d_{*1}, d_{*2}, \dots, d_{*j}, \dots\}$ を抽出し、学習した θ を用いて $\phi_\theta(D_*) \in \mathbb{R}^{(2k+1)d}$ に変換する。FV によって得られる特徴量は平均と分散の統計量を考慮しており、特徴量の出力次元数は $(2k+1)d$ 次元となる。BoF では 1 つの局所特徴量が k -means 法によって 1 つの学習済みの ψ_θ に分類されるが、FV では局所特徴量と多くの学習済みの ψ_θ の関係を表現することができ、クラスターへの帰属度を考えることができる [20]。

4. 賃料予測モデルの精度評価実験

本研究では、賃料予測に与える間取り図の影響を明らかにするため、間取り図を考慮した賃料予測モデルと考慮しない賃料予測モデルの賃料予測実験を行う。

4.1 使用データセット

本実験では、国立情報学研究所の情報学研究データリポジトリで公開されている LIFULL HOME'S データセットを用いる。このデータセットには、2015 年 9 月時点で不動産住宅情報サイト LIFULL HOME'S に掲載されていた全国 533 万件の物件データが含まれている。賃料や占有面積、築年数などの物件属性に加え、間取り図（120×120 ピクセル）や外観写真、内観写真などの画像ファイルで構成されている。本実験では、北海道・東京都・愛知県・大阪府・福岡県の 2,244,867 物件を対象とし、目的変数を毎月の家賃と共益費の合計とし、説明変数を表 1 の 28 の物件属性と間取り図とする。画像ファイルから、各物件に対応した間取り図を抽出するために、LIFULL HOME'S が提供している画像種別のラベル付けシート（photo_rent.tsv）から、各物件の間取り図に対応するパスを抽出する。この際、間取り図にカラー画像と白黒画像が混在していたため、間取り図にグレースケール変換を適用し白黒画像に統一する。また、データセットを確認した所、ラベル付けシートに誤りがあり、間取り図タグが他の画像種別と結びついているものが見受けられたが本実験ではそのまま使用している。

本実験では 2,244,867 件の物件を、物件数の多い 5 つの都道府県と K・R・DK・LDK の 4 種類の間取り規格で層別し、計 20 種類のカテゴリに分ける。各カテゴリに対して賃料予測モデルを構築する。

4.2 実験条件

説明変数には表 1 に記載の前処理を適用する。表 1 において連続変数の物件変数でデータの欠損がある物件はデータセットから除外し、それ以外の物件変数でデータの欠損がある場合は当該変数に欠損カテゴリを付与しデータセットに含める。また、説明変数の値の範囲を統一するため、全ての入力変数が $[0, 1]$ になるようにスケール変換する。間取り図以外の各説明変数 a_i は、学習データ L 中の当該変数の最大値 $\max$ と最小値 $\min$ を用いて、

$$\frac{a_i - \min}{\max - \min} \quad (13)$$

で正規化する。間取り図はグレースケール変換した後、各ピクセルを 255 で割る。また、分割したデータセットにおける賃料の分布は図 3 に示すように、正規分布と異なる分布になる傾向にある。本実験では賃料に対数をとった $\log y$ を目的変数とすることで、賃料の分布を擬似的な正規分布に変換する。

間取り規格ごとに物件を 4:1 の割合で学習データとテストデータにランダムに分割する。学習に使用するデータセットから外れ値を除去するために、学習データに含まれる物件で、賃料が標準偏差の 3 倍以上である物件は除外する。表 2 にカテゴリごとの物件数を示す。

表 1 本実験で用いる説明変数と前処理

物件変数	変数説明	前処理
専有面積	平米（連続変数）	[0, 1] にスケール変換
部屋階数	部屋の所在階数（連続変数）	[0, 1] にスケール変換
築年数	2015 年 9 月時点での築年数（連続変数）	[0, 1] にスケール変換
新築・未入居フラグ	中古, 新築, 未入居	one-hot ベクトル化
物件種別	マンションやアパートなどのフラグ	one-hot ベクトル化
空き物件数	空部屋数	0 件は 0 件として, [1, 55] 件では 5 件毎に one-hot ベクトル化
市区郡町村	渋谷区, 新宿区など	one-hot ベクトル化
駅 1 2	最寄駅名	one-hot ベクトル化
用途地域	近隣商業, 準住居など	one-hot ベクトル化
都市計画	市街化区域, 未線引区域など	one-hot ベクトル化
建物構造	木造, RC など	one-hot ベクトル化
管理人有無	常駐, 日勤, 巡回, 無	one-hot ベクトル化
建物階数（地上）	10 階, 11 階など	1 階は 1 階として, [2, 18] 階は 2 階毎に one-hot ベクトル化
建物階数（地下）	B1, B2 など	1, B1, B2, B3 以上で one-hot ベクトル化
向き	所在部屋の向き	one-hot ベクトル化
引渡／入居旬	上旬, 中旬, 下旬	one-hot ベクトル化
引渡／入居旬	即時, 相談, 期日指定	one-hot ベクトル化
バス時間 1 2	最寄駅までのバス時間	[0, 25] 分で 5 分毎に one-hot ベクトル化
駐車場距離	単位: [m]	50 m 毎に one-hot ベクトル化
徒歩距離 1 2	最寄駅またはバス停まで距離	[0, 2000] m で 400 m 毎に one-hot ベクトル化
総合病院距離	所在物件からの距離	[0, 2000] m で 400 m 毎に one-hot ベクトル化
小学校距離	所在物件からの距離	[0, 2000] m で 400 m 毎に one-hot ベクトル化
中学校距離	所在物件からの距離	[0, 2000] m で 400 m 毎に one-hot ベクトル化
コンビニ距離	所在物件からの距離	[0, 2000] m で 400 m 毎に one-hot ベクトル化
スーパー距離	所在物件からの距離	[0, 2000] m で 400 m 毎に one-hot ベクトル化

図 3 検証データ（東京都の LDK 物件）の賃料分布

賃料予測モデルのハイパーパラメータ探索は、学習データに 5 分割交差検証法（5-fold cross validation : 5-CV）の適用により行う。本実験では予測器として、線形回帰と LightGBM を適用する。線形回帰のハイパーパラメータとして正則化法（Lasso・Ridge）と正則化係数を探索する。また、間取り図を考慮した賃料予測モデルにおいては PCA, BoF, FV の次元数も探索する。テストデータは、予測モデルを評価するために用意し、テストデータの予測時には、学習データの全てを学習に使用する。

間取り図の特徴抽出器として、PCA, BoF, FV を適用する。なお、計算コストの観点から、FV のみ、特徴抽出器のパラメータは抽出した局所特徴量群の 10% をランダムサンプリングしたものをを用いて学習する。

表 3 に賃料予測モデルで探索するハイパーパラメータを示

表 2 カテゴリごとの物件数 [件]

	間取り 規格	学習 データ	テスト データ	合計
北海道	K	13,072	3,268	16,341
	R	38,623	9,655	48,279
	DK	93,728	23,432	117,160
	LDK	18,459	4,614	23,074
東京都	K	121,784	30,446	152,231
	R	141,621	35,405	177,027
	DK	220,154	55,038	275,193
	LDK	377,680	94,420	472,101
愛知県	K	17,528	4,382	21,910
	R	48,259	12,064	60,324
	DK	111,755	27,938	139,694
	LDK	78,536	19,634	98,170
大阪府	K	42,184	10,546	52,731
	R	75,660	18,915	94,575
	DK	3,836	959	4,796
	LDK	167,995	41,998	209,994
福岡県	K	17,826	4,456	22,283
	R	39,953	9,988	49,942
	DK	96,168	24,042	120,210
	LDK	71,065	17,766	88,832

表 3 探索するハイパーパラメーター一覧

PCA	圧縮次元数：514, 1,028, 2,056
BoF	k-means のクラスタ数：514, 1,028, 2,056
FV	混合分布数：2, 4, 8
線形回帰	正則化法：Lasso, Ridge 正則化係数：0, 0.001, 0.01

す。ここで、FV の混合分布数が $2 \cdot 4 \cdot 8$ であるのは、PCA と BoF の出力次元数と合わせるためである。また、損失関数に MAPE を適用するために、線形回帰は Keras で実装し、学習には確率的勾配降下法を採用している。この際、バッチ数は 32、エポック数は 200 に固定している。

5. 間取り図が賃料予測に与える影響の分析

5.1 賃料予測モデルの精度比較

表 4～8 にカテゴリごとのテストデータに対する MAPE と Absolute Percentage Error (APE) の 95% 信頼区間を示す。表 4～8 の FPI は間取り図 (Floor Plan Image) を意味する。また、間取り図を考慮しないモデルよりも MAPE が改善している箇所を太字で示し、間取り図を考慮しないモデルよりも APE の 95% 信頼区間が短いモデルを下線で示している。

表 4～8 より、線形回帰と LightGBM の両方のモデルで PCA 及び BoF を適用した賃料予測モデルの 95% 信頼区間幅が間取り図を考慮しない賃料予測モデルよりも短くなることが確認できる。これは間取り図が賃料予測モデルにとって予測のばらつきを抑える役割があることを示唆している。また、表 4～8 より、BoF を適用した線形回帰が、間取り図を考慮しない線形回帰よりも MAPE が低いカテゴリの数は 15/20 であるため、線形回帰では間取り図を考慮することで MAPE が改善する傾向にあるといえる。ただし、大半のカテゴリにおいて、間取り図の考慮による MAPE の改善度合いは、間取り図の有無での線形回帰の 95% 信頼区間幅が重なっていることから小さいと考える。賃料予測モデルに間取り図を考慮するための計算コストを考えると、間取り図を考慮した線形回帰が間取り図を考慮しない線形回帰よりも優れた賃料予測モデルであるとは必ずしもいいきれない。しかし、東京都の間取り規格 LDK、愛知県の間取り規格 K および、大阪府の間取り規格 LDK の線形回帰において、間取り図を考慮しない線形回帰と BoF を適用した線形回帰を比較すると、95% 信頼区間幅が間取り図を考慮することで短くなり、さらに 95% 信頼区間の重複がなく MAPE が改善していることから、間取り図の有無で予測精度に有意な差があることが分かる。愛知県の間取り規格 K と大阪府の間取り規格 LDK では、間取り図を考慮しない線形回帰モデルの MAPE が 10% を超えているカテゴリであるため、間取り図を考慮しない線形回帰モデルで MAPE が 10% 超えるカテゴリに対して、間取り図を考慮することで MAPE が有意に改善すると考えられる。一方で、LightGBM では、全カテゴリで APE の 95% 信頼区間が重複しており、間取り図の有無で MAPE に有意な差がないといえる。

賃料予測に適した間取り図の特徴抽出器について考察する。

PCA・BoF・FV を適用した 3 つの賃料予測モデルの MAPE をカテゴリごとに比較すると、BoF を適用した賃料予測モデルの MAPE が最も優れているカテゴリの数は、線形回帰で 19、LightGBM で 5 となっており、3 つの特徴抽出器の中で最もカテゴリ数が多い。このことから、間取り図の特徴抽出器には BoF が望ましいと考える。さらに、BoF の適切なクラスタ分割数について、カテゴリごとのクラスタ分割数を比較する。線形回帰ではクラスタ分割数 514 が適切であったモデルが 3 個、1,028 が 2 個、2,056 が 15 個であり、クラスタ分割数の大きい BoF が賃料予測に適している傾向にある。また、LightGBM ではクラスタ分割数 514 が適切であったモデルが 4 個、1,028 が 5 個、2,056 が 11 個であり、適切なクラスタ分割数は線形回帰の場合と比べて偏りが見られない。これは、間取り図の有無での MAPE の差に有意差がほとんどなく、間取り図の特徴量が賃料予測に影響を与えていないことを示唆している。また、本研究において FV を用いた賃料予測モデルの MAPE が優れない原因は、ガウス混合分布の分割数が小さいためであり、間取り図から得られる検出点を分割できていないことが考えられる。

5.2 間取り図の重要度分析

線形回帰の回帰係数および LightGBM の変数重要度の観点で間取り図が賃料予測に与える影響を考察する。考察には、特徴抽出器 PCA・BoF・FV を適用した賃料予測モデルの中で、MAPE が最も優れているカテゴリ数が最も多い BoF を適用した賃料予測モデルを用いる。

5.2.1 線形回帰における間取り図の価値分析

線形回帰モデルは式 (1) と賃料の対数変換から

$$\hat{y} = e^{\alpha_1^T a} \times e^{\alpha_2^T \phi_{\theta}(f)} \times e^{\beta} \quad (14)$$

と表すことができる。ここで、 $\alpha_1 \in \mathbb{R}^m$ を間取り図以外の物件属性の回帰係数、 $\alpha_2 \in \mathbb{R}^d$ を間取り図の回帰係数とする。これは、予測賃料を間取り図以外の価値と、間取り図の価値、バイアスの価値の掛け合わせた表現と解釈できる。

図 4～7 に、BoF を適用した線形回帰の MAPE が、間取り図を考慮しない線形回帰の MAPE と比較して改善したカテゴリ (大阪府の間取り規格 LDK、愛知県の間取り規格 K) と、改悪したカテゴリ (北海道の間取り規格 K、大阪府の間取り規格 DK) に関する物件属性の価値の分布を示す。図 4～7 では、間取り図以外の物件属性の価値 (a) にそれぞれ数件の外れ値があったためそれらは除去している。図 4 に示すように、大阪府の間取り規格 LDK では、間取り図の価値が図 4(b) の価値分布の外れ値を除いて 1.3 から 2.5 であるのに対し、間取り図以外の価値が 0 から 75,000 であるため間取り図の価値の方が小さいことが分かる。しかしながら、式 (14) より予測賃料が間取り図以外の価値と、間取り図の価値、バイアスの価値の掛け合わせで表現されていることを考慮すると、大阪府の間取り規格 LDK の間取り図の価値は、価値分布幅が 1.3 から 2.5 であることから $25,000 \times (2.5 - 1.3) \times 2.22 = 66,600$ 円程度であると考えられる。大阪府の間取り規格 LDK において、66,600 円の間取り図の価値は予測賃料の中央値 99,900 円の 66.7% にあた

表 4 北海道のテストデータにおける MAPE と APE の 95% 信頼区間

間取り 規格	予測モデル	w/oFPI	PCA	w/FPI BoF	FV
K	線形回帰	13.2 [11.4, 15.0]	14.1 [13.6, 14.6]	13.5 [13.0, 14.0]	450 [387, 512]
	LightGBM	9.77 [9.34, 10.2]	9.29 [8.96, 9.62]	11.3 [9.66, 12.9]	9.47 [9.14, 9.80]
R	線形回帰	12.0 [10.9, 13.1]	10.9 [10.6, 11.2]	11.1 [10.7, 11.5]	160 [147, 172]
	LightGBM	7.59 [7.14, 8.04]	7.12 [6.90, 7.34]	7.61 [7.39, 7.83]	7.30 [7.07, 7.52]
DK	線形回帰	14.1 [13.3, 14.9]	14.5 [14.2, 14.8]	13.5 [13.2, 13.8]	222 [207, 238]
	LightGBM	8.62 [8.07, 9.17]	8.48 [8.29, 8.67]	8.72 [8.53, 8.91]	8.56 [8.37, 8.75]
LDK	線形回帰	14.8 [14.2, 15.4]	15.2 [14.9, 15.5]	14.8 [14.3, 15.3]	520 [490, 551]
	LightGBM	7.28 [6.80, 7.76]	7.35 [7.26, 7.44]	7.32 [7.30, 7.34]	7.31 [7.21, 7.40]

表 5 東京都のテストデータにおける MAPE と APE の 95% 信頼区間

間取り 規格	予測モデル	w/oFPI	PCA	w/FPI BoF	FV
K	線形回帰	13.1 [12.5, 13.7]	12.9 [11.5, 14.3]	12.4 [12.2, 12.6]	273 [232, 313]
	LightGBM	9.26 [8.32, 10.2]	9.17 [9.07, 9.27]	9.26 [9.18, 9.34]	9.24 [9.15, 9.34]
R	線形回帰	10.2 [9.90, 10.5]	11.1 [11.0, 11.2]	10.0 [9.98, 10.0]	219 [209, 228]
	LightGBM	8.43 [8.18, 8.68]	8.46 [8.39, 8.53]	8.49 [8.44, 8.54]	8.51 [8.46, 8.56]
DK	線形回帰	7.67 [7.36, 7.98]	8.54 [8.52, 8.56]	7.44 [7.25, 7.63]	114 [111, 117]
	LightGBM	9.63 [9.32, 9.94]	9.76 [9.55, 9.97]	9.60 [9.51, 9.69]	9.65 [9.56, 9.74]
LDK	線形回帰	7.79 [7.47, 8.11]	7.91 [7.90, 7.92]	6.89 [6.83, 6.95]	265 [256, 274]
	LightGBM	10.2 [9.60, 10.8]	10.2 [10.0, 10.4]	10.2 [10.0, 10.4]	10.3 [10.2, 10.4]

表 6 愛知県のテストデータにおける MAPE と APE の 95% 信頼区間

間取り 規格	予測モデル	w/oFPI	PCA	w/FPI BoF	FV
K	線形回帰	21.6 [20.7, 22.5]	12.7 [12.5, 12.9]	11.8 [11.6, 12.0]	419 [388, 451]
	LightGBM	8.07 [7.67, 8.47]	8.33 [8.19, 8.47]	8.43 [8.32, 8.54]	8.48 [8.34, 8.62]
R	線形回帰	8.63 [7.79, 9.47]	8.74 [8.61, 8.87]	7.80 [7.76, 7.84]	477 [163, 792]
	LightGBM	7.34 [6.87, 7.81]	13.3 [12.7, 13.9]	7.08 [6.88, 7.28]	7.04 [6.94, 7.14]
DK	線形回帰	8.42 [7.94, 8.90]	8.93 [8.77, 9.09]	8.08 [7.93, 8.23]	249 [221, 276]
	LightGBM	7.76 [7.51, 8.01]	7.86 [7.72, 8.00]	7.93 [7.80, 8.06]	7.84 [7.71, 7.96]
LDK	線形回帰	7.62 [7.25, 7.99]	8.26 [8.16, 8.36]	7.32 [7.26, 7.38]	204 [194, 214]
	LightGBM	7.62 [7.18, 8.06]	7.64 [7.55, 7.73]	7.62 [7.55, 7.69]	7.66 [7.57, 7.74]

表 7 大阪府のテストデータにおける MAPE と APE の 95% 信頼区間

間取り 規格	予測モデル	w/oFPI	PCA	w/FPI BoF	FV
K	線形回帰	15.1 [14.4, 15.8]	16.0 [15.7, 16.3]	15.1 [14.8, 15.4]	443 [398, 488]
	LightGBM	9.89 [9.67, 10.1]	10.0 [9.94, 10.1]	9.94 [9.78, 10.1]	10.1 [9.88, 10.2]
R	線形回帰	10.4 [9.87, 10.9]	10.7 [10.6, 10.8]	10.0 [9.92, 10.1]	424 [355, 494]
	LightGBM	7.30 [7.04, 7.56]	7.29 [7.22, 7.36]	7.28 [7.21, 7.35]	7.31 [7.23, 7.38]
DK	線形回帰	8.63 [8.19, 9.07]	9.69 [9.08, 10.3]	8.90 [8.47, 9.33]	280 [267, 294]
	LightGBM	8.36 [8.05, 8.67]	8.42 [8.30, 8.54]	8.40 [8.28, 8.52]	8.47 [8.35, 8.59]
LDK	線形回帰	53.5 [47.2, 59.8]	46.3 [42.6, 50.0]	40.3 [39.6, 41.0]	572 [388, 755]
	LightGBM	21.9 [20.1, 23.7]	22.1 [20.2, 24.0]	23.0 [21.1, 24.9]	22.1 [20.2, 24.0]

表 8 福岡県のテストデータにおける MAPE と APE の 95% 信頼区間

間取り 規格	予測モデル	w/oFPI	PCA	w/FPI BoF	FV
K	線形回帰	10.6 [9.90, 11.3]	10.9 [10.5, 11.3]	10.5 [10.2, 10.8]	858 [705, 101 × 10]
	LightGBM	7.56 [7.42, 7.70]	7.71 [7.48, 7.94]	7.82 [7.59, 8.05]	7.74 [7.50, 7.97]
R	線形回帰	9.26 [8.80, 9.72]	9.60 [9.42, 9.78]	8.87 [8.69, 9.05]	366 [342, 390]
	LightGBM	7.17 [6.72, 7.62]	7.11 [7.00, 7.22]	7.22 [7.12, 7.32]	7.14 [7.04, 7.25]
DK	線形回帰	9.54 [8.98, 10.1]	9.69 [9.47, 9.91]	8.94 [8.73, 9.15]	498 [439, 557]
	LightGBM	7.03 [6.77, 7.29]	7.07 [6.93, 7.21]	7.07 [6.93, 7.21]	7.09 [6.95, 7.23]
LDK	線形回帰	8.45 [7.50, 9.40]	9.62 [7.32, 11.9]	8.53 [8.28, 8.78]	231 [221, 241]
	LightGBM	6.59 [6.36, 6.82]	6.65 [6.57, 6.73]	6.65 [6.57, 6.73]	6.69 [6.61, 6.78]

図 4 大阪府の間取り規格 LDK における物件属性の価値の分布 (バイアス項: 2.22)

図 5 愛知県の間取り規格 K における物件属性の価値の分布 (バイアス項: 4.30)

図 6 大阪府の間取り規格 DK における物件属性の価値の分布 (バイアス項: 15.8)

図 7 北海道の間取り規格 K における物件属性の価値の分布 (バイアス項: 4.37)

るため賃料予測に十分な影響を与えていると考える。ここで、25,000 は大阪府の間取り規格 LDK の間取り以外の価値の中央値を、2.22 はバイアス項を指す。同様に、図 5 から愛知県の間取り規格 K において間取り図の価値の分布幅が 0.95 から 1.02 であり、間取り図の価値が 3,010 円程度であると考えられる。これは予測賃料の中央値 43,430 円の 6.93% にあたる。愛知県の間取り規格 K における予測賃料に占める間取り図の価値 6.93% は大阪府の間取り規格 LDK の予測賃料に占める間取り図の価値 66.7% よりも小さいが、予測賃料が間取り図と 28 個の説明変数で構成されることを考慮すると、間取り図の価値が賃料予測に十分な影響を与えていると考える。MAPE が改悪したカテゴリにおいても、図 6 と 7 から北海道の間取り規格 K では間取り図の価値が 4.47 から 5.75 で間取り図の価値分布幅が 1.28、大阪府の間取り規格 DK では間取り図の価値が 1.50 から 3.00 で価値分布幅が 1.28 であることが確認できる。これらの結果から、BoF を適用した賃料予測モデルでは、間取り図を考慮しない線形回帰の MAPE と比較して改善したカテゴリと改悪したカテゴリのどちらも、予測賃料における間取り図の価値は間取り図以外の価値より小さいものの賃料予測に十分な影響を与えていると考える。

5.2.2 LightGBM における重要度変数のランキン

本実験では、間取り図を特徴量 $\phi_{\theta}(f) \in \mathbb{R}^d$ で表現しており、間取り図に対応する説明変数の数が d 個となる。特徴抽出器として FV を用いる場合、抽出する特徴量に平均と分散の統計量を保持し、特徴量と間取り図上の座標との対応関係を把握できないため、特徴量の任意の要素 $d_i \in \phi_{\theta}(f)$ についての解析は難しい。そこで、LightGBM における間取り図の価値を、変数重要度上位 k 件に含まれる間取り図の特徴量の数で評価する。表 9 に、各都道府県のテストデータに対して、BoF を適用した LightGBM の間取り図の特徴量の変数重要度が高い 100 件・200 件・300 件・400 件に含まれる個数を示す。表 9 から、400 位以内に含まれる間取り図の特徴量の数 0 から 356 であることが確認できる。これは、間取り図の特徴量の次元数 514・1,028・2,056 と比べると少なく、賃料予測に影響を与える間取り図の特徴が、間取り図の特徴量の特定の要素 d_i に集約されているためだと考える。また、表 9 から、LightGBM における間取り図の重要度順位は 300 位以上 400 位以下の部分に集中

表 9 LightGBM における変数重要度上位 k 件に含まれる間取り図の特徴量数

都道府県	間取り規格	top- k			
		100	200	300	400
北海道	K	54	136	226	315
	R	45	133	216	308
	DK	40	113	200	291
	LDK	17	55	124	214
東京都	K	0	4	11	34
	R	0	0	2	13
	DK	0	0	1	10
	LDK	1	4	8	14
愛知県	K	66	152	238	325
	R	0	19	67	141
	DK	4	26	63	124
	LDK	0	5	13	54
大阪府	K	6	44	88	145
	R	2	12	37	74
	DK	2	13	28	92
	LDK	79	170	267	356
福岡県	K	32	101	181	264
	R	16	68	133	203
	DK	20	57	113	188
	LDK	4	36	86	173

していることが確認できる。このことを、専有面積や部屋数などの物件属性ベクトル約 400 次元と比較すると、間取り図の変数重要度の順位は、間取り図以外の物件属性と比べて低いといえる。

5.3 検出された特徴点と予測誤差の分析

SURF で得られる間取り図の特徴点を可視化し、具体的な間取り図の特徴について考察する。図 8 は大阪府の間取り規格 DK に含まれる物件の中で、間取り図の有無で APE が 10% 以上改善した物件 (a) と低下した物件 (b) である。図 8 から、改善の有無に関わらず、SURF を用いた特徴点検出では壁となる部分が間取り図の特徴として検出されていることが確認できる。また、間取り図に描画されている洋室や LDK などの文字も間取り図の特徴として捉えていることが分かる。

図 8 から、APE が 10% 以上改善した物件 (a) と低下した物件 (b) では、間取り図の特徴検出点数が異なっており、検出点数の差が賃料予測精度に影響を与えていることが考えられる。そこで、間取り図の検出点数ごとの APE を比較する。表 10 に、間取り図を考慮した線形回帰と考慮しない線形回帰の APE 差と間取り図の検出点数の相関係数を示す。表 10 から BoF・FV とともに、間取り図の検出点数と APE の改善に相関は見受けられないことを確認している。これは、間取り図から抽出する検出点数を限定したとしても APE の改善度に影響を与えないことを示唆している。つまり、本実験で使用した間取り図より解像度を高くしても、間取り図から特徴抽出する処理速度と予測機の学習時間に影響を与えないと考えられる。

(a) APE が 10% 以上改善した間取り図

(b) APE が 10% 以上低下した間取り図

図 8 SURF による間取り図の特徴点検出例

表 10 間取り図の有無による APE 差と間取り図の検出点数の相関係数

都道府県	間取り規格	BoF	FV
北海道	K	0.02	0.13
	R	0.03	0.10
	DK	-0.01	-0.01
	LDK	0.01	-0.00
東京都	K	0.01	0.01
	R	0.03	0.01
	DK	0.01	0.12
	LDK	-0.01	0.03
愛知県	K	-0.00	0.04
	R	-0.01	0.01
	DK	0.04	0.04
	LDK	0.01	0.06
大阪府	K	0.03	0.03
	R	0.00	0.00
	DK	0.00	0.10
	LDK	0.04	0.08
福岡県	K	-0.02	0.10
	R	0.01	0.08
	DK	0.03	0.03
	LDK	-0.00	0.06

6. おわりに

本研究では、間取り図を考慮した賃料予測モデルと間取り図を考慮しない賃料予測モデルを構築し、間取り図の有無での予測精度の差の比較により間取り図が賃料予測に与える影

響を明らかにしている。予測精度は、LIFULL HOME'S データセットの北海道、東京都、愛知県、大阪府、福岡県の賃貸物件 2,244,867 件を用いた実験により評価している。間取り図を考慮した賃料予測モデルでは、間取り図の特徴抽出器として、PCA、BoF、FV を適用している。また、予測器には線形回帰と LightGBM を使用している。

賃料予測精度の評価実験から、次の 6 点を明らかにしている。(i) 予測アルゴリズムに線形回帰を採用する場合、間取り図を考慮しない賃料予測モデルで MAPE が 10% を超えるカテゴリに対して間取り図を考慮することで MAPE を改善できる。(ii) 間取り図を考慮することで、APE のばらつきを軽減できる。(iii) 賃料予測モデルに適した間取り図の特徴抽出器は BoF であり、線形回帰では、BoF の次元数は 2,056 次元が望ましい。(iv) BoF を適用した線形回帰では、予測賃料における間取り図の価値は間取り図以外の価値より小さいものの賃料予測に十分な影響を与えている。しかしながら、間取り図が他の物件属性よりも賃料予測に与える影響が小さい原因については明らかにできていないので、今後解明したい。(v) 間取り図を考慮した LightGBM における間取り図の特徴量の変数重要度は、上位 300–400 位に含まれる傾向にあり、間取り図が賃料予測に与える影響は他の物件属性よりも小さい。(vi) SURF で得られる間取り図の検出点数と賃料予測モデルの APE の改善度に相関はないことから、間取り図の解像度が間取り図の特徴量抽出の処理速度と予測器の学習時間に影響を与えない。

本研究では、間取り図と賃貸物件を題材として「図的な情報」を価格予測モデルについて議論している。価格予測モデルに図情報を考慮することで、価格予測での影響度を定量的に評価することができる。さらに、図情報から直接特徴抽出するため、大規模データセットをラベル付けをする必要がない。加えて、PCA・BoF では、特徴量に図中の座標情報が保持されている。そのため、PCA または BoF を適用した線形回帰では、図情報のどの部分が予測に影響を与えているのかを算出することができる。また、BoF を適用したモデルでは入力画像の解像度が予測モデルの処理に直接影響を与えない利点もある。ただし、本研究のように直接図情報から特徴抽出をする場合、特徴量の要素についての解析は困難であるため、特徴量全体によって図情報の価値を評価する必要がある。

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。本研究は JSPS 科研費 JP18K18159 の助成を受けたものです。

参考文献

- 一般財団法人住宅改良開発公社: 「賃貸住宅市場の現況と中長期見通し」に関する調査研究, ニッセイ基礎研究所, 2014.
- 清田陽司, 山崎俊彦, 諏訪博彦, 清水千弘: “不動産と AI,” 人工知能学会誌, Vol.32, No.4, pp. 529-535, 2017.
- 大野喜久之輔: 継続賃料鑑定評価を再考する, 住宅新報出版, 2006.
- 河合伸治: “ヘドニック・アプローチによる賃貸住宅価格の価格決定要因の推定—西武池袋線の賃貸住宅を事例として,” ソシオサイエンス, Vol.14, pp. 49-63, 2008.
- 石島博, 谷山智彦: “個別性と歪みを考慮した住宅価格分析とパーソナルファイナンスへの応用,” ファイナンシャル・プランニング研究, Vol.10, pp. 4-17, 2010.
- 福井光, 阪井一仁, 南村忠敬, 三尾順一, 木下明弘, 高田司郎: “レイノズのニューラルネットワークを用いた不動産価格査定について,” 人工知能学会第 32 回全国大会論文集, 6 月 5-8 日, 鹿児島, p. 4A203, 2018.
- S. Rosen: “Hedonic Prices and Implicit Markets: Product Differentiation In Pure Competition,” *J. of Political Economy*, Vol.82, No.1, pp. 34-55, 1974.
- 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, Vol.73, No.623, pp. 139-146, 2008.
- G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray: “Visual Categorization with Bags of Keypoints,” *Proc. of ECCV Workshop on Statistical Learning in Computer Vision (SLCV)*, May 5, Prague, 2004.
- K. Simonyan, O. M. Parkhi, A. Vedaldi, and A. Zisserman: “Fisher Vector Faces in the Wild,” *Proc. of British Machine Vision Conf.*, Sep. 9-13, Bristol, pp. 8.1-8.11, 2013.
- G. Ke, Q. Meng, T. Finely, 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: Proc. of the 31st Annual Conf. on Neural Information Processing Systems (NIPS2017)*, Dec. 4-9, Long Beach, pp. 3146-3154, 2017.
- 河合伸治: “ヘドニック・アプローチによる賃貸住宅価格の価格決定要因の推定—西武池袋線・東武東上線・西武新宿線・小田急線・京王線沿線の単身者用賃貸住宅を事例として,” 社学研論集, No.10, pp. 285-294, 2007.
- 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, 3 月 6 日, 京都, pp. 80-87, 2019.
- H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool: “Speeded-Up Robust Features (SURF),” *Computer Vision and Image Understanding*, Vol.110, No.3, pp. 346-359, 2008.
- 濱田雄斗: “不動産賃貸価格における画像を用いた価格推定,” 人工知能学会誌, Vol.32, No.4, pp. 569-571, 2017.
- E. Ahmed and M. Moustafa: “House Price Estimation from Visual and Textual Features,” *Proc. of the 8th Int. Joint Conf. on Computational Intelligence (IJCCI2016)*, Nov. 9-11, Porto, pp. 62-68, 2016.
- LightGBM 3.1.1: <https://pypi.org/project/lightgbm/> [accessed Oct. 5, 2020]
- D. G. Lowe, “Object Recognition from Local Scale-Invariant Features,” *Proc. of the 17th IEEE Int. Conf. on Computer Vision*, Sep. 20-25, Corfu, pp. 1150-1157, 1999.
- jonasrothfuss/fishervector: <https://github.com/jonasrothfuss/fishervector> [accessed Oct. 5, 2020]
- 原田達也: “大規模データを用いた一般物体・シーン認識の潮流と理論,” 情報処理学会研究報告, Vol.2011, No.5, 2012. (2020 年 8 月 17 日 受付)
(2021 年 2 月 10 日 採録)

[問い合わせ先]

〒182-8585 東京都調布市調布ヶ丘 1-5-1

電気通信大学 大学院情報理工学研究科 情報学専攻

岡本 一志

TEL: 042-443-5280

E-mail: kazushi@uec.ac.jp

— 著 者 紹 介 —

はっとり りょうすけ
服部 凌典 [非会員]

2019年電気通信大学情報理工学部先端工学基礎課程卒業。現在、同大学大学院情報理工学研究科情報学専攻在学中。

おかもと かずし
岡本 一志 [正会員]

2006年高知工科大学工学部情報システム工学科卒業。2008年同大学院工学研究科基盤工学専攻修士課程修了。2011年東京工業大学大学院総合理工学研究科知能システム科学専攻博士後期課程修了。同年、千葉大学アカデミック・リンク・センター特任助教。2015年電気通信大学大学院情報理工学研究科総合情報学専攻助教。2020年同大学院情報理工学研究科情報学専攻准教授。現在に至る。博士（工学）。データサイエンスの研究に従事。

しばた あつし
柴田 淳司 [正会員]

2009年青山学院大学理工学部電気電子工学科卒業。2012年東京工業大学大学院総合理工学研究科知能システム科学専攻修士課程修了。2015年同大学院総合理工学研究科知能システム科学専攻博士課程修了。同年、岐阜大学電気電子情報学科産官学連携研究員。同年、東京都立産業技術大学院大学助教。現在に至る。博士（学術）。データサイエンス研究に従事。人工知能学会、情報処理学会、日本知能情報ファジィ学会各会員。

Impact Analysis of Floor-Plan Images for Rent-Prediction Model

by

Hattori RYOSUKE, Kazushi OKAMOTO, and Atsushi SHIBATA

Abstract:

In order to evaluate effects of usage of floor-plan images (FPIs) in a rent prediction, we construct rent-prediction models with and without FPIs and compare them in terms of prediction error. Principal component analysis, Bag of Features (BoF), and Fisher Vector (FV) are used as feature extractors, and linear regression and LightGBM are used as a regressor. A rent prediction experiment with the LIFULL HOME'S dataset is performed. For the linear regression models, the experimental results suggest that usage of FPIs improves the means of prediction errors in some categories (combination of a prefecture and a floor-plan standard) and reduces the 95% confidence intervals of prediction errors for all categories. In addition, BoF achieves the best prediction accuracy in the three feature extractors for FPIs.

Keywords: rent prediction, floor-plan image, principal component analysis, Bag of Features, Fisher Vector

Contact Address: **Kazushi OKAMOTO**

Graduate School of Informatics and Engineering, The University of Electro-Communications
1-5-1 Chofugaoka, Chofu, Tokyo 182-8585, Japan
TEL: +81-42-443-5280
E-mail: kazushi@uec.ac.jp