# 間取り図を用いた賃料予測モデルに関する一検討 

Evaluation of Rent Prediction Models using Floor Plan Images

$* 1$ 服部 凌典 $\quad * 2$ 岡本 一志 $* 3$ 柴田 淳司<br>Ryosuke Hattori Kazushi Okamoto Atsushi Shibata<br>＊1電気通信大学 情報理工学部<br>Faculty of Informatics and Engineering，The University of Electro－Communications<br>＊2電気通信大学 大学院情報理工学研究科<br>Graduate School of Informatics and Engineering，The University of Electro－Communications<br>＊3産業技術大学院大学 産業技術研究科<br>Graduate School of Industrial Technology，Advanced Institute of Industrial Technology


#### Abstract

This study constructs rent prediction models with／without floor plan images in order to validate whether such images contribute the prediction accuracy．In addition，applications of PCA（principal component analysis）and convolutional neural network are considered as a feature extractor from floor plan images．The prediction accuracy is measured using properties of 90,000 rental housings in Tokyo．In the experimental results，the root mean squared error values of the prediction model with floor plan images and PCA tend to be higher than without floor plan images．This suggests that the use of floor plan images contributes to accuracy of rent prediction．


## 1．はじめに

賃貸物件において，属性が全て同じ物件は数少なく，類似し た属性を持つ物件でも築年数や階数，間取りなどのどこかに違 いがある．特に，同一の間取りの規格（ 1 K や 2 LDK など）で あったとしても，各部屋の配置や面積などは物件毎に異なる。 また，日本では賃貸物件を探す際に間取り図を見る習慣があり ［清田 2017］，多くの物件では間取り図が用意されている。

物件属性は賃料に影響を与えると考えられており［河合2008，石島 2010，福井 2018］，賃料決定の際に利用可能な手法として ヘドニック・アプローチがある。間取り図の記載内容も物件属性のひとつと見なすこともできるが，これまでのヘドニック・ アプローチによる賃料予測の事例では考慮されてきていない。

本研究では，間取り図を用いた賃料の線形回帰モデル構築 し，間取り図を用いない線形回帰モデルと予測精度を比較す ることで，間取り図の活用が賃料予測に与える影響を明らか にする。前者の線形回帰モデルでは，間取り図から抽出した特徴量を利用するが，特徴抽出器として，主成分分析（PCA： principal component analysis）と畳み込みニューラルネット ワーク（CNN：convolutional neural network）の適用を検討 する．構築したモデルの予測精度は，LIFULL HOME＇S デー タセットに含まれる東京都の賃貸物件 9 万件を用い，モデル の出力と真の賃料との平均二乗誤差を求めることで検証する。

## 2．構築する賃料予測モデル

ヘドニック・アプローチとは，ある商品の価格をその商品の属性の価値に関する集合とみなし，商品価格を線形回帰モデル で予測する手法である［河合 2008］．本研究では，入力ベクト $ル \boldsymbol{x} \in \mathbb{R}^{d}$ から予測賃料 $\hat{y}$ を

$$
\begin{equation*}
\hat{y}=\boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{x}+\beta \tag{1}
\end{equation*}
$$

連絡先：岡本一志，電気通信大学 大学院情報理工学研究科情報学専攻，東京都調布市調布ヶ丘 1－5－1，Tel．：042－443－5280， E－mail：kazushi＠uec．ac．jp

で求めるモデルを扱う．ここで，$\alpha \in \mathbb{R}^{d}$ を偏回帰係数，$\beta$ を定数項とし，データから予測誤差を最小化するよう推定する。

本研究では，間取り図を用いる場合と用いない場合の予測 モデルを構築する．物件属性（築年数や階数，専有面積など） を $\boldsymbol{u} \in \mathbb{R}^{m}$ ，間取り画像を $\boldsymbol{v} \in \mathbb{R}^{n}$ ，関数 $f_{\theta}$ をパラメータ $\theta$ を持つ画像特徴抽出器としたとき，間取り図を用いる場合は $\boldsymbol{x}=\left[\boldsymbol{u}, f_{\theta}(\boldsymbol{v})\right]$ を，用いない場合は $\boldsymbol{x}=\boldsymbol{u}$ として扱うことに なる。また，特徴抽出器 $f_{\theta}$ には PCA と CNN を適用する。

### 2.1 PCAによる間取り図からの特徴抽出

特徴抽出器 $f_{\theta}$ に PCA を用いる場合，学習データに含まれる間取り図の集合から変換行列 $\theta \in \mathbb{R}^{k \times n}$ を推定し，$f_{\theta}(\boldsymbol{v})=\theta \boldsymbol{v}$ とすることで間取り図の特徴ベクトルを得ることができる。変換行列 $\theta$ の推定には教師信号（賃料）は必要ないため， $\boldsymbol{x}=$ $\left[\boldsymbol{u}, f_{\theta}(\boldsymbol{v})\right]$ の算出後に，式（1）の $\boldsymbol{\alpha}, \beta$ を推定することになる． また，$f_{\theta}(\boldsymbol{v})$ の結果を逆変換することで，次元圧縮後の間取り図を近似的に復元でき，次元圧縮の傾向を画像として確認する ことができる。

## 2．2 CNNによる間取り図からの特徴抽出

CNN による間取り図用の特徴抽出器 $f_{\theta}$ ではVGG16を利用する．VGG16 は，ImageNet データセットを用いて学習し た 13 層の畳み込み層と 3 層の全結合層を持つニューラルネッ トワークである．VGG16 の出力は 1,000 クラスの確率分布で あり，層数が増加することにより計算コストも高くなってしま うため，本研究では，VGG16の後ろ 5 層を削除し新たに $k$ 次元の全結合層を追加する。なお，VGG16 の結合係数は固定と し，追加した $k$ 次元の全結合層への結合行列 $\theta$ のみを学習す ることとする。

この特徴抽出器を用いる場合，結合行列 $\theta$ の推定には教師信号が必要となるため，本研究では，間取り図 $v$ をVGG16に入力し得られた特徴ベクトル $f_{\theta}(\boldsymbol{v})$ と物件属性ベクトル $\boldsymbol{u}$ の連結層と賃料の出力層（1 次元）を持つニューラルネットワー クを構築する．そのため，PCA を用いる場合と異なり，賃料 を用いてパラメータ $\boldsymbol{\alpha}, \beta, \theta$ を同時に推定するモデルとなる。

## 3．賃料予測モデルの精度評価実験

## 3.1 使用データセット

本研究では，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 データセットを使用し，構築した賃料予測モデルの予測精度を検証する。このデータ セットは，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 万件の物件データで構成 されており，賃料，面積，立地，築年数，間取り，建物構造，諸設備などの物件属性や，間取り図（ $120 \times 120$ ピクセルの） や外観写真，内観写真などの 8,300 万の画像ファイルが含まれ ている．このうち，東京都の 9 万件の物件について，賃料と共益費を足したものを目的変数とし，市区郡町村•徒歩距離•建物構造•建物面積／専有面積•建物階数（地上）•建物階数（地下）－築年月•新築•未入居フラグ・部屋階数•契約期間（年）•駐車場料金•引渡／入居時期の 12 変数を物件属性として利用 する．なお，賃料が欠損している物件は除外する。

本研究では， 9 万件の物件を K／R／DK／LD／LDK の 5 種類の間取り規格で層別し，間取り規格毎に賃料予測モデル を構築する。ただし，間取りが LD の物件は 35 件のみである ため，間取りが K／R／DK／LDK の物件を使用する。

## 3.2 実験条件

間取り図を予測モデルに用いる場合，特徴抽出器のハイパー パラメータとして，次元数 $k$ を設定する必要がある。このた め，パラメータ調整用に検証用データを，構築した予測モデ ルの評価用にテストデータをそれぞれ用意する。具体的には，各間取り規格において，物件を $4: 1$ の割合で検証用データとテ ストデータに分割し，パラメータ調整は検証用データに 10 分割交差検証法を適用することで行う。テストデータに対する予測では，検証用データ全てを学習に利用する。なお，本研究に おいては，CNN の全結合層の次元数 $k$ の調整は計算時間の関係から行えておらず，$k=64$ に固定している。PCA について は，$\{64,128,256,512,1024,2048\}$ の次元数の中から適切な次元数を探索している．また，モデルの予測精度を評価指標には平均二乗誤差（RMSE：root mean squared error）を用いる。

## 4．予測精度の比較

PCA を用いた特徴抽出器において，検証用データにおけ る最良の平均 RMSE をとるのは，間取り規格が K の物件で $10548.18(k=512)$ ， R の物件で $8789.96(k=2048)$ ， DK の物件で $11069.98(k=1024)$ ，LDK の物件で 19668.87 （ $k=2048$ ）であることを確認している．そのため，PCA の次元数 $k$ にはこれらの値を利用する。

表1に間取り規格が LDK の物件におけるRMSEを示す。表1のテストデータの欄では，間取り図に PCA を適用した予測モデルの RMSE が最も高く，間取り図に CNNを適用し た予測モデルの RMSE が最も低くなっている．同様の傾向は，間取り規格がR／DK の物件についても確認している。 さら に，全ての間取り規格の物件において，PCAを適用する場合 が CNNを適用する場合に比べてRMSEが高くなっている。 このことは，顀料予測に間取り図を使うことよりも，今回の ニューラルネットワークの構造に課題があるためと考える。ま た，間取り規格が LDK の物件について，PCA を適用した間取り図の復元例を図1に示す。図1より，PCA による間取り図の次元圧縮自体は適切に行えていることが確認できる。

PCA を用いた予測モデルと間取り図を用いない予測モデル のテストデータにおけるRMSE の差は，間取り規格が K の物

表 1：間取り規格が LDK の物件に対する RMSE 値（検証用 データについては交差検証法適用時の平均と $95 \%$ 信頼区間）

|  | 検証用データ | テストデータ |
| :---: | :---: | :---: |
| $\mathrm{w} /$ 間取り図 <br> $(\mathrm{PCA}, k=2,048)$ | 19668.87 <br> $[19265.94,20071.80]$ | 19646.72 |
| $\mathrm{w} /$ 間取り図 <br> $(\mathrm{CNN}, k=64)$ | 23590.04 <br> $[23230.41,23949.65]$ | 22878.80 |
| w／o 間取り図 | 20227.00 <br> $[19927.19,20526.87]$ | 20141.84 |


（a）入力画像

（b）復元画像

図 1：PCA（ $k=2048)$ により次元圧縮した間取り図の復元例

件で－86，R の物件で 490，DK の物件で 199 ，LDK の物件で 495 となっている。間取り図の利用が賃料の予測精度の向上に寄与していると考えるが，その効果の分析はまだ詳細には行え ておらず，今後明らかにしていく予定である。

## 5．おわりに

本研究では，間取り図を用いた賃料の線形回帰モデルを構築 し，間取り図を用いない線形回帰モデルとの予測精度の比較を行っている。東京都の賃貸物件 9 万件を利用した賃料予測実験 より，間取り図に PCAを適用した線形回帰モデルが間取り図 を利用しない線形回帰モデルよりも予測精度（RMSE）が高い傾向にあり，間取り図の利用が賃料の予測精度の向上に寄与し ていると考える。加えて，間取り図からの特徴抽出には PCA が VGG16 ベースのニューラルネットワークよりも RMSE が高くなることも確認している。

## 参考文献

［清田 2017］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，Vol．32，No．4，pp．529－535（2017）
［河合2008］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一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 2010］石島博，谷山智彦：個別性と歪みを考慮した住宅亚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 2018］福井 光，阪井 一仁，南村 忠敬，三尾順一，木下 明弘，高田 司郎：レイン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，人工知能学会全国大会論文集，Vol．JSAI2018， No．4A2－03，pp．1－4（2018）

## 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。

