

# Web 不動産データを用いた空物件が入居されるまでの期間に関する データ特性を考慮した統計モデリング

Statistical modeling of the transition time of an occupation of rental rooms by using the housing information website data

渡邊隼史 \*<sup>1</sup> 一藤裕 \*<sup>2</sup> 鈴木雅人 \*<sup>3</sup> 山下智志 \*<sup>4</sup>  
Hayafumi Watanabe Yu Ichifuji Masahito Suzuki Satoshi Yamashita

\*<sup>1</sup>金沢大学 Kanazawa University \*<sup>2</sup>長崎大学 Nagasaki University \*<sup>3</sup>UD アセットバリュエーション UD Asset Valuation Co., Ltd.

\*<sup>4</sup>統計数理研究所  
The institute of statistical mathematics

The apartment loan is a loan for rentals such as for condos, apartments. This loan is a very large loan which is the account for a percentage of more than 10 percents of the whole banks' loan. However, a risk model of the apartment loan with the appropriate accuracy has not been provided in Japan mainly due to the lack of data. Thus, in order to develop the risk model, we analyzed the duration in which a vacant room become occupied by a tenant by using the housing information website data, as a first step. As a result, it was found that (i) This duration can be explained by the geometric distribution, and (ii) The mixture geometric regression model considering nonlinear effects can describe the data properties. In addition, coefficients of this model roughly consistent with the empirical common senses.

## 1. はじめに

アパートローンは、賃貸物件向けの融資であり、銀行が貸し出す全与信額の 10% を超える巨大な融資であるにもかかわらず、これまで十分な精度のリスク計量化モデルが考案されてこなかった。アパートローンのリスクは、(1) 賃貸物件の経営自体による資金不足リスク (2) 賃貸経営以外の事業によるリスクの 2 種類の要因から構成される。(2) については、一般的な信用リスクモデルやデータベースが整っており、それを用いてある程度の精度をもってリスク計量が可能である。一方 (1) については、賃貸不動産の空室データベースが整っておらず、十分な精度のモデルが提供されていなかった。そこで現在 (1) について空占室データベース構築と評価モデルの開発を目指しプロジェクトとして研究を行っている。なお、本研究プロジェクトの最終的な目的の一つは構築したアパートローンリスク計量モデルを CRD 協会 (銀行にリスクデータベースとリスクモデルを提供する組織) を通して、民間金融機関に提供することにある。

前学会より、著者らは、アパートローンリスクの評価法の開発の第一歩として、「物件の埋まりやすさ」や「埋まりにくさ」をアパートのもつ特性から不動産情報サイトのデータ (以下 Web データ) から評価するモデルの構築について報告している。前学会では、(1) Web データは、棟のレベルでは、銀行の融資物件からの層化抽出データである鑑定士による調査データ (後述) と統計的には性質が類似しており、アパートローン用の空室率解析に利用可能性があること。(2) 簡単なロジスティック回帰モデルである程度は空物件が埋まる状況が説明可能なことを報告した。ただし、特に (2) について、簡単な線形ロジスティック回帰モデルでは現象記述にはやはり限界があり、モデルの説明力のうち 9 割以上が「築年数」が占めるという問題点が残った (この結果は、不動産の経験や常識とあまり整合性が高いとはいえない)。

そこで、今学会では、上記の問題を克服するためのモデルの精密化について報告する。まず、築年数の説明力が大きい理由

の考察、次に、埋まる期間の確率的構造の解析を行った。さらに、それらに基づき、ロジスティック回帰モデルよりデータ構造に適合した空室が埋まる期間を記述する統計モデルを提案し、その性質を調べた。結果、データ解析では、築年数は線形の主成分解析においてデータの特性を最も説明する第一主成分であること、データ上の「空物件が埋まる期間」は幾何分布と対応する特性をもつこと; モデル化では、提案モデルは、単純なロジスティック回帰に比べて、説明力が築年数以外にも分散した常識と大きくは矛盾しない回帰係数をもつこと等を示した。また、それらの考察により物件が埋まる期間の変動がどの程度がランダムに決定し、どの程度が物件自身の特性によって決まるかの割合を見積もることが出来る可能性を示唆した。

## 2. 関連研究

本稿で扱う研究は、データを用いた不動産評価の研究の一種といえる。データを用いた不動産評価、特に、不動産価格評価の研究は 1970 年代に主な手法が確立し、現在も大規模データや機械学習等の技術を用いて精度の改善、網羅性やリアルタイム性の性能上の向上の努力が続けられている [清田 17, 清水 17]。

一方、価格ではなく、データを用いた個別物件の空占遷移予測について国内における学術論文はデータ入手の困難もあり多くは見つけられなかった。その例としては、籠, 高辻, 小野らがモンテカルロシミュレーションによる空室率を考慮したアパートリスク計量の理論的な研究を行っている [籠 00]。また、実データを用いた例では小林の研究がある [小林 16]。小林の研究では仲介管理会社 3 件の約 673 件のデータについて空室期間等を実データから推定しそれに基づく将来収益予測モデルの構築を行っている。

本研究は実データ解析のため小林の研究に近い。小林の研究として比べた新たな貢献としては、より様々な物件要因を考慮していること、よりスケーラブルな Web データを利用すること、また不動産鑑定士の現地調査のサーベイデータ (業者による偏りが少ないデータ) と比較することで業者やサービス等

連絡先: 渡邊隼史, hayafumi.watanabe@gmail.com

のバイアスの存在の有無を確認していること等があげられる。

### 3. データの取得

本研究では2つのデータを用いた。一つはメインのWeb不動産サイトデータであり、もう一つは補助的に用いる不動産鑑定士による現地調査データである。不動産鑑定士による調査データを利用する主な目的は、以下の2点である。一点目は、Webでは得にくい情報（既に埋まっている物件の情報、管理の悪さなどネガティブな情報等）の効果の影響調査のため、二点目は、2つのデータを比較することでWebサイトデータのサンプリングバイアスについての情報を得るためである。

Webデータは、大手不動産のサイトのある県のある路線地域について、2014年11月28日から2015年夏まで、各月の8日、18日、28日の情報を取得した。一方、鑑定士によるサーベイデータは対応する地域についてある物件名簿より駅別に層化抽出を行った物件に対して同観測期間に3か月ごとに空室状況や物件の状況等を現地調査して得られたデータである。

本研究のWebデータ解析では「空室が占室がなる」を「Webページから物件が消える」でだまかに近似できるという仮定のもとで研究を行っている。

### 4. データ解析とその結果

#### 4.1 主成分分析による築年数の説明力の高さの考察

なぜ築年数の説明力が高くなるかの原因を考察するためWebデータに関する主成分分析を行った。結果、意味の解釈可能な主成分が6成分とれた。この6成分により、データ変動の意味が解釈できた6成分は、表記を、「第○主成分（寄与率、累積寄与率）：主成分のネーミング（寄与率が大きい要因）」としたとき、第1主成分（6.98%、6.98%）：築年数・築年月系、第2主成分（5.89%、12.8%）：水道光熱系（上下水道、空調等）、第3主成分（4.99%、17.7%）：建物構造・高さ（建築構造、高さ等）、第4主成分（3.42%、21.2%）：各部屋の広さ（面積、賃料等）、第5主成分（3.09%、24.3%）入居条件（楽器、ペット、事務所可否等）、第6主成分（2.24%、26.6%）：周辺環境（スーパー、コンビニ、駅距離、最寄り駅等）であった。以上より、築年数が線形の範囲では、各物件の特徴をもっとも代表する変数ということが確認でき、これが空物件が入居するときに築年数が説明力が高い一つの要因と考えられる。ただし、ほかの変数の説明力も小さくないためこれだけが要因でないことも同時に考察される。

#### 4.2 同じ物件の個別の部屋の埋まるまでの期間の解析

モデル構築するため非説明変数にあたる「埋まるまでの期間」の確率的特性を調べた。具体的には、物件内のすべての部屋が観測できる鑑定士データを用いた分散解析を行った。分散分析では、各棟ごとに埋まるまでの期間の平均 $\lambda$ とその分散 $\sigma^2$ （同じ物件内の埋まるまでの期間の平均からのばらつき）の関係性を調べた。その結果、だまかには、2つの量は $\sigma^2 = \lambda^2 + \lambda$ という関係性を持つことがわかった（加分散が存在する）。この結果は、非常に粗くいうと、全く同じ特性をもつ部屋であっても、埋まる期間のばらつきは、その物件の埋まる期間程度はあるということになる（例えば、平均100日埋まる特性もつ部屋であっても、ばらつきが100日なので、偶然性によって0日で埋まることもあれば200日程度で埋まる可能性もある[ただし、厳密言うと、埋まる期間の分布の形状は対称形をでないで完全にはこの解釈どおりにはならない]）。なお、このばらつき方は同じ棟の各部屋はランダムに埋まるという幾何過程で説明できるため、これをモデル化に利用する。

#### 4.3 データ特性にあったモデル化と結果

上記のデータ解析等をもとに、より現実のデータ構造にあった埋まる期間のモデル（入居期間モデル）を提案した。モデルは前学会と比べて以下の特性をもつ。

- 非説明変数は、埋まる期間（前学会モデルの非説明変数は、3か月で空物件が占か埋まらないかの二値変数）
- モデルは、混合幾何回帰モデル（前学会は、ロジスティック回帰モデル）
- 連続的な説明変数を非線形の効果を記述できるように順序カテゴリカル変数化した（前学会は、線形効果のみ）。また、「隣接するカテゴリの回帰係数の値は近い」という事前知識を事前分布という形で導入した（Bayesian lasso）。

なお、1点目はデータの情報量をより活用するため、2点目は上記のデータ解析の結果得られた構造にモデルが適合するように、3点目は、データ構造をよりモデルが柔軟に記述できるように導入した。

結果、Webデータからモデルの回帰係数を推定したところ、例えば、築年数以外にも、駅やコンビニからの距離が近いほど埋まる期間の回帰係数が小さい[埋まりやすい]、建物が高いほど回帰係数が大きい[埋まりにくい]、クローゼットがあるほうが回帰係数が小さい[埋まりやすい]など、常識との相違が大きい有意に説明力をもつ回帰係数が得られた（c.f.、単純なロジスティック回帰モデルではほとんどの説明力が築年数になる）。また、推定されたモデルパラメータの結果により、非常にだまかな目安として、変動要因別の大きさは(i)Webで可観測の物件特性効果[±50日程度]、(ii)Webでは非観測の物件特性効果[±10日程度]、(iii)物件要因以外のランダム変動[±150日程度]と見積もれる可能性があることを示唆した。

### 5. まとめ

Web不動産賃貸募集データを用いた「空物件が埋まるまでの期間」のデータ特性に合わせた統計モデル（混合幾何回帰モデル）を提示した。結果（1）モデルの回帰係数はだまかには常識とは外れない（2）モデルにより、「Webで観測できる物件特性効果」、「Webでは非観測の物件特性効果」、「物件要因以外のランダム効果」の大きさをそれぞれわけて見積もれる可能性を示した。ただし、まだ、賃料に係る量（要因調整後も高賃料ほど埋まりやすい）などモデル解釈にはいくつかの課題が残るため、今後とも、その原因の解明が必要である。

### 参考文献

- [清田 17] 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, 32巻, 4号, pp. 529-535 (2017)
- [清水 17] 清水千弘: 不動産ビッグデータでみる不動産価格の決まり方, 日本不動産学会誌, 32巻, 1号, pp. 45-51 (2017)
- [籠 00] 籠義樹・高辻秀興・小川祐哉, 空室率と賃料の変動過程を考慮した不動産投資モデルの期待収益率とVaRに関する研究, 麗澤経済研究, 8巻, 2号, pp. 99-113 (2000)
- [小林 16] 小林 秀二: 入退去マイクロ分析による不動産評価における将来予測の捉え方: 生存時間解析データによる家賃キャッシュ・フローの無条件確率遷移, ファイナンシャル・プランニング研究 16号, pp. 18-27(2016)