

Web不動産データを用いた空物件が埋まる遷移に関する多変量解析 —アパートローンリスク計量モデル構築のための予備解析—

Multivariate analysis of the transition probability of an occupation of rental rooms by using the housing information website data: Preliminary analysis to develop the risk model of the rental home financing

渡邊隼史 *1
Hayafumi Watanabe

一藤裕 *2
Yu Ichifuji

鈴木雅人 *3
Masato Suzuki

山下智志 *4
Satoshi Yamashita

*1情報・システム研究機構
Research Organization of Information and Systems

*2長崎大学
Nagasaki University

*3UD アセットバリュエーション
UD Asset Valuation Co., Ltd.

*4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 preliminarily analyse and compare two types of data set: the survey data which is made by the real estate appraiser and the housing information website data. As a result, it was found that (i)the web data is approximately corresponding to the survey data which was made by experts with respect to statistical properties, (ii)The AR value of the simple multivariate regression model developed in this study which explains the transition of rooms from vacancy to occupation takes about 0.4 and (iii)this transition probability of the model is mainly explained by an age of a building.

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。

本稿では、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評価するプロトタイプのモデル構築について報告する。空室から占有への遷移や物件情報など研究に必要なデータは不動産情報サイトのデータ(以下Webデータ)を用いた。本研究では、まず、Webデータと不動産鑑定士による現地調査データを比較することでWebデータがどのような特色をもつデータかを確認し、次に結果に基づきWebデータによる「埋まりやすさ」評価のモデルの構築可能性検証のため、簡単な多変量解析モデルの構築と評価を行った。

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]。最近年においては、実務においては、2015年前後から、ソニー、ネクスト(現FIFULL, HOMES)、マンションリサーチ等の企業により不動産価格評価のWebサービスが公開・提供されている[谷山16]。また研究レベルでは、画像情報を用いた不動産価値評価[清田17]、アンケート情報やセンサー等を用いた人の流量データを不動産や都市評価に利用する試み[荒川17]、など近年まで利用が難しかった様々な情報を不動産評価に加える試みもおこなわれている。

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. データの取得および前処理

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。

解析で用いたWebデータについては以下の前処理を行った。具体的には、(1)取得データのスプレイピング(2)単位等の統一(3)表記ゆれ等の統一(4)欠損値補完(5)物件の名寄せである。なお、前処理の結果、観測期間の14か月で解析可能な部屋数は35,807件[空室のみ]となった(不動産鑑定士データは、246棟,4333戸[うち空室は400戸程度])。

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

4. 解析とその結果

4.1 不動産鑑定士データとWebデータの基本的な統計の比較

不動産鑑定士データとWebデータを基礎的な統計量な簡単な比較した結果以下のことがわかった。

- 2つのデータ共に、空室から占室になる期間の分布は共に平均150日程度の指数分布によく近似できる([小林16]も指数分布と報告)。
- 築年数、駅からの距離、建物の高さの存在率の分布はWebと鑑定士データでほぼ対応する。ただし、Webデータのほうが築浅物件がやや多い傾向にある。

つまり、いくつかの相違はあるものの大まかには、Webデータはほぼ鑑定士によるデータと対応していることが確認できた。

4.2 3か月後遷移確率の多変量解析

次に空室から占室への遷移が物件がどの程度説明できるかを確認するため簡単な多変量解析を行った。具体的には、簡単のためロジスティック回帰モデル(二項ロジットモデル)を用いた。モデルでは、被説明変数を「空室だった物件が3か月に占室になる場合のフラグ(空→空が0, 空→占が1)」, 説明変数をWebデータに掲載されている項目(築年数, 最寄り駅, コンビニからの距離, 面積, 募集賃料, 保証人の有無など数十項目程度)としてモデルを構築した。変数選択は多重共線性の影響を避けるため精度指標のAR値($AR=2AUC-1$) [山下11]が改善がとまるまで1変数ずつ回帰モデルに変数を追加する方法を用いた(AICステップワイズ法に近い方法, 「追加」なのは異常値等の影響を目視で確認しやすいため)。なお、一時期だけ空占に効く変数は特定の事業者がまとめて不動産サイトから物件を除去した効果など本質的でない要因が多かったため、観測期間で常に効果がある変数のみを採用するようにした。

結果、以下の7要因採用され、モデルの49観測点でのバックテストAR値の中央値0.378だった(国産小企業の倒産予測モデル程度の精度)。採用された項目は以下の通り: 築年数(0.318)[-], 即入居可(0.357)[-], 保証人不要(0.361)[+], 洗髪洗面化粧台(0.370)[+], 駅から徒歩距離(0.372)[-], 追い炊き有(0.376)[+], コンロ(なし, IH, ガス...) (0.378)[+]。ここで表記は、項目(その要因より前に記述がある要因郡を変数

として採用したモデルの累積AR値)[効果の正負]であった。結果より、築年数が空占遷移の説明力の大きな部分を占めていることがわかった。

5. まとめと考察

本研究ではアパートローンリスク計量化のうち、アパート経営の事業リスク評価の第一段階の準備研究として、Web不動産サイトデータの性質の検証と「空→占」の簡単な状態遷移モデルの作成を行った。具体的には、まず、Webデータの収集と前処理の手法を整備した。次に、不動産鑑定士による現地調査と比較し、Webから得られたデータの性質の検証を行った。最後に、簡単な多変量空占遷移モデルを構築した。(このとき、異常値などWebデータの特有の問題に対応するため、時間に依存せず安定した説明要因をひとつずつ採用する変数選択などWebに合わせた工夫を行った)。

解析の結果以下のことがわかった。(1)Webデータは不動産鑑定士データと統計的性質が類似している[Webデータ使用の妥当性確認]。(2)多変量二項ロジットモデルを用いた「空→占の確率遷移モデル」はバックテストAR値で0.4程度(3)モデルで主要な説明力を持つ要因は築年数。

考察 本研究の多変量解析で採用された変数には、「募集賃料」や「最寄り駅」など主要な変数を含んでいない。これは以下の2つが主な理由と考えている(1)賃料調整効果(2)市場価格より大きく離れた物件データの不足(同データから構築した賃料回帰モデルの説明要因は、今回採用された変数は築年数と駅からの距離を除き含まれなかった)。つまり、本研究の結果は賃料に反映されていない物件の需要要因と解釈できるかもしれない(例えば、築年数はニーズ比べて割安など)。モデルを信頼性をもって利用するためには、このような空占室遷移と賃料との関係を今後詳細に検討する必要があると考えられる。

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. 529535 (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. 4551 (2017)
- [谷山16] 谷山俊彦: 人工知能とビッグデータが返る不動産市場, Financial Information Technology Focus 2016, 6号, pp. 8-9 (2016)。
- [荒川17] 荒川周造・諏訪博彦・小川祐樹・荒川豊安本慶一・太田敏澄, 暗黙知に基づく飲食店用不動産の賃料推定システム, SIG-SAI, 28巻, 1号, pp. 16 (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. 99113 (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)
- [山下11] 山下智志・三浦翔, 信用リスクモデルの予測精度: AR値と評価指標, 朝倉書店 (2011)。