

# 高精度少量データと低精度大量データの結合: Webアパート空室情報の信頼性評価

山下 智志 データ科学研究系 教授

本研究は渡邊隼史(金沢大学)との共同研究の成果である

## 【1. 背景と目的】

銀行が貸し出すアパートローンは全与信額の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。アパートローンの貸倒は、

- ① 賃貸不動産の収益低下によるキャッシュ不足
- ② 賃貸不動産以外のキャッシュアウトによるキャッシュ不足

によって起こりえる。①に対しては空室率の予測や変動金利の予測によってリスク評価が行われており、②については一般的な信用リスクモデルを活用されることが多い。信用リスクモデルについてはCRD(Credit Risk Database)などの大規模データによる頑健なモデルが構築されており、少なくともデフォルト確率については一定の精度で予測が可能である(期待損失率についてはデフォルト後の損失データが不足しているため必ずしも正確な分析ができていない)。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、実務レベルにおいても研究レベルにおいても精度の高いモデルは供給されていない。

本研究ではWebデータを定時観測することによって大規模データベースを作成し、賃貸住宅の入居化要因を分析する。一方、Webデータの精度を補完するため、不動産鑑定士による現地パネル調査を行った。それぞれのデータを統合することにより、より正確な賃貸住宅の収益予測を行うことを目指している。

## 【2. Web賃貸住宅募集データ】

賃貸住宅募集のホームページに記載されている募集情報を保存している。具体的には楽天不動産とSUUMOのHPのうち特定の都道府県データ(約18000件)を10日ごとに保存することによって、占室化、空室化の情報だけでなく、部屋の属性データも入手されている。属性データには物件の位置情報や募集賃料、面積、部屋割り、階数などの基本情報のほか、インターネット回線の有無やセキュリティ、駐車場などの付帯設備の詳細な情報が含まれている。レコード数、フィールド数とも大きいため、いわゆるビッグデータとなるが、仲介業者の入力精度にやや疑問があり、このデータ単独の分析は実務的に十分とはいえない。

Web大規模データによれば空室が一定期間内に占室となる確率については通常の2項ロジットモデルなどで推定可能である。しかし、現況占室の物件に対してはデータが入手できないため、占室が一定期間内に空室となる確率については推定することができない。そのため占室→空室モデルについては以下のサーベイデータを基本にモデリングを行う。なお、空室→占室モデルについてはWebデータとサーベイデータの両方が利用可能であり、将来的にはWebデータをサーベイデータで補強するモデルを検討している。

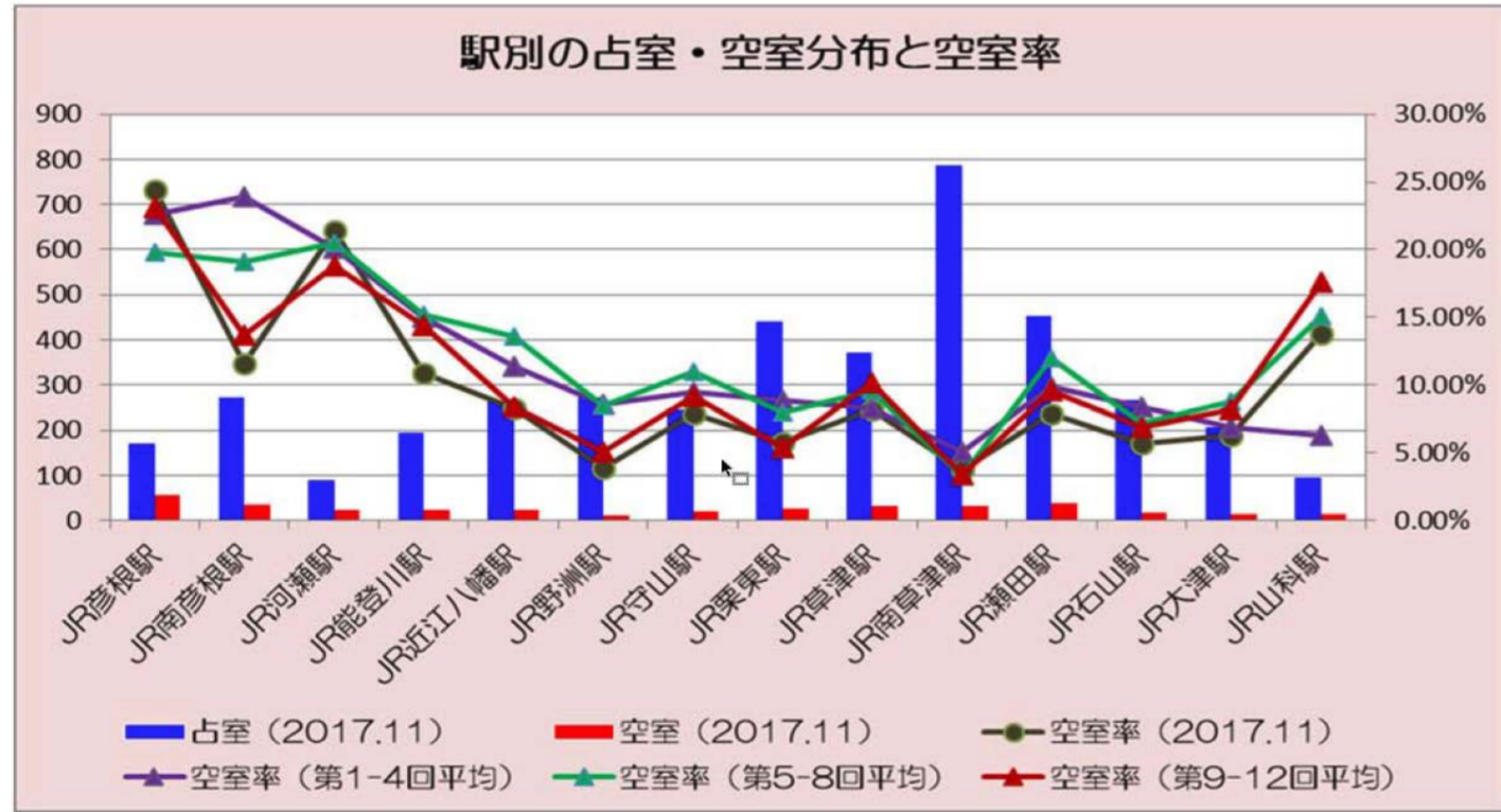


図1: 棟ごとの「埋まるまでの期間」の分布。(a) 累積分布(縦軸は横軸の値以上になる確率)(b) 対応する確率密度関数。黒実線はWebデータ、赤破線は鑑定士データ、緑破点線は指数関数  $\propto \exp(x/143)$ 。図より、Webデータと鑑定士データの分布はよく対応し、指数分布に近似できることがわかる。なお、鑑定士データが不連続の階段状なのは、3か月おきの観測のためである。

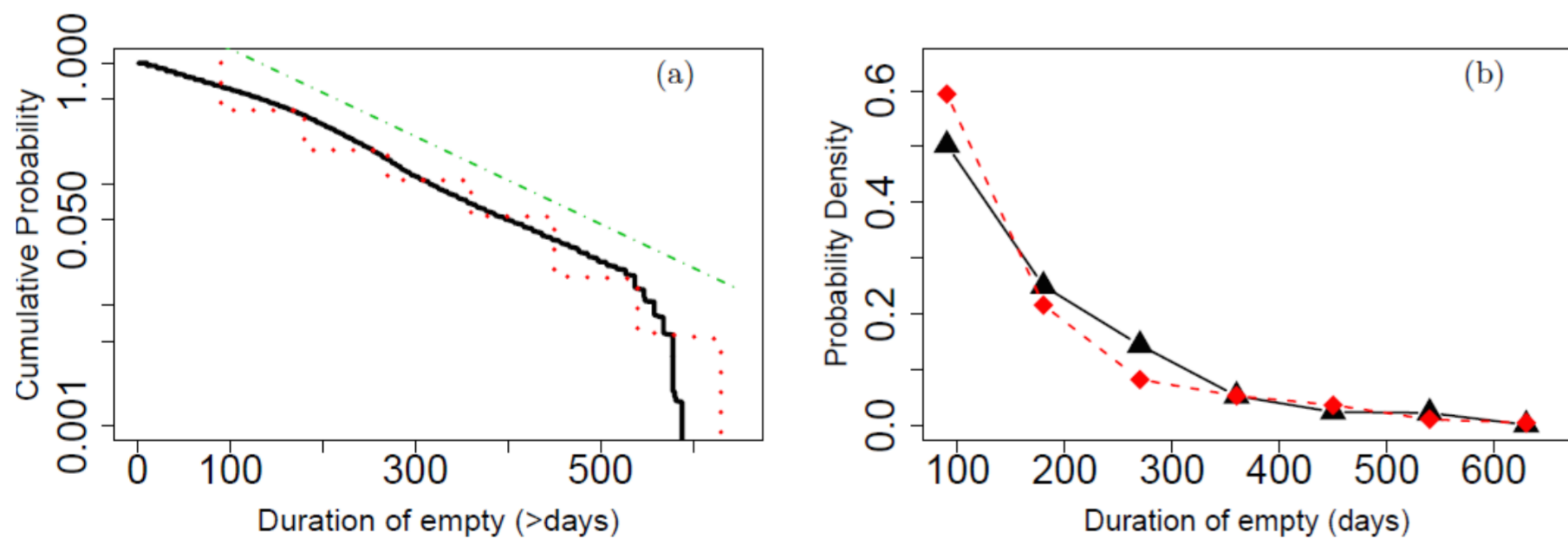


図2: 棟ごとの代表的な物件特性の分布(確率密度関数, 確率質量関数)。(a) 築年数 (b) 棟の高さ (c) 駅からの徒歩時間(分) (d) 賃料。黒三角はWebデータ、赤菱形は対応する鑑定士データ。図より、Webデータと鑑定士データの分布はだまかには類似していることが見て取れる。

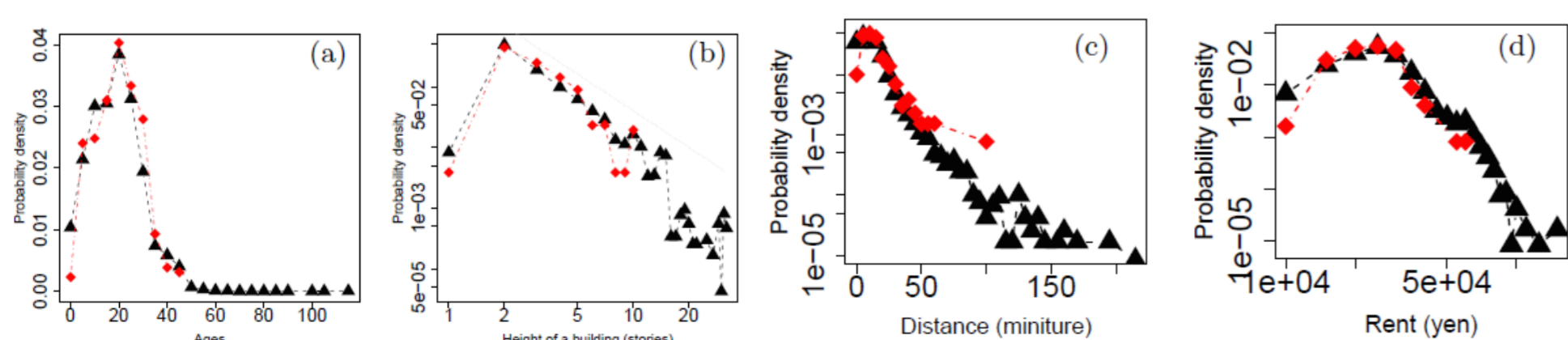
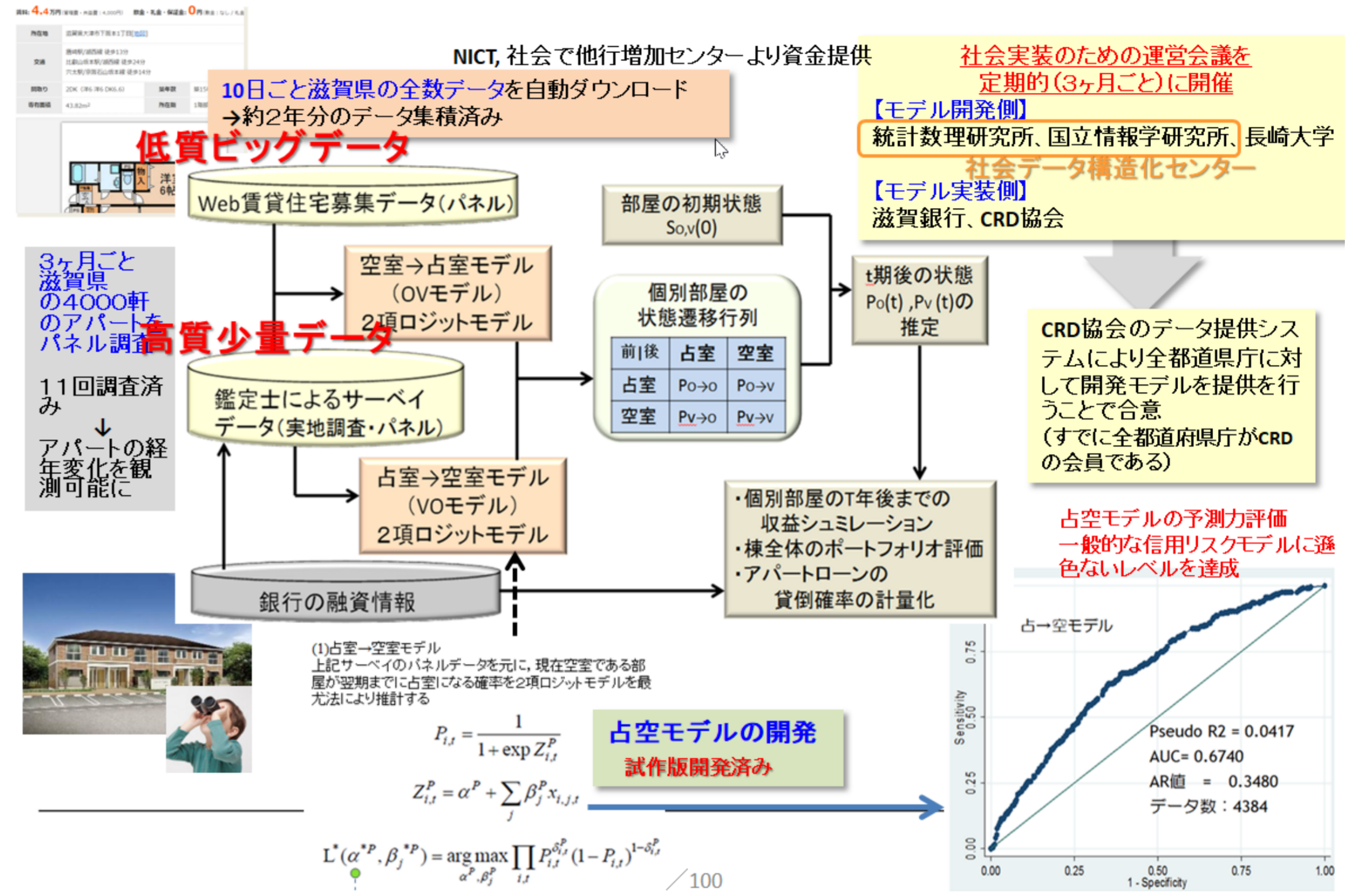


図3: 物件特性と埋まるまでの期間の相関(物件特性で条件づけした埋まるまでの期間の平均値)。信頼区間は両側10パーセントブートストラップ信頼区間。(a) 築年数 (b) 棟の高さ、サポートラインは  $\propto x^{-2}$ 。(c) 駅からの徒歩時間(分) (d) 賃料。黒三角はWebデータ、赤菱形は対応する鑑定士データ。図より、Webデータと鑑定士データはだまかには対応していることが見て取れる。なお、ここでの平均値は、今回のWebデータの観測期間である約700日における平均値であることに注意されたい(詳細は5.3節を参照)。



## 【3. 鑑定士によるサーベイデータ(実地調査)】

3ヶ月ごとに賃貸集合住宅240棟、4300室をパネルウォッチを行っている。空室状況の確認だけでなく、物件の管理状態のチェックや不動産仲介業者の募集情報を確認し、説明変数として有効な情報を収集している。特に不動産鑑定士による「管理状態」「住環境の良否」「修繕の状態」について計9項目のエキスパートジャッジを行っている。なお、調査対象となった住宅棟はある銀行のアパートローン融資対象物件であり、不動産情報のほか融資実態の情報が入手可能である。専門家による実地調査のためデータの信頼性は極めて高いが、費用等の問題で大規模に実施することはできない。

**不動産鑑定士によるサーベイデータ**

**調査名:** 滋賀県周辺の賃貸住宅空室状況パネル調査に関する調査、並びに追加調査

**調査目的:** 賃貸住宅の収益性に関する統計モデルを構築するための基礎データとして、ある県内のJR東海道本線沿線の賃貸住宅について、その占有・空室状況並びに物件の属性等を戸単位で調査し、データ化する。

**調査対象物件:** 統計数理研究所が指定したJR東海道本線沿線地域の集合住宅249棟、4,318戸。(銀行より入手したアパートローン与信ポートフォリオのデータから層化抽出したもの。実際には4,425戸)

**調査項目:**

**A) 物件属性情報**  
棟単位: 総階数, 交通アクセス, 構造, 築年数, 総戸数, 管理人の有無, オートロックの有無, 駐車場の有無, 設備の内容, 管理の良否・修繕の状態・住環境の良否に関するエキスパートジャッジメントなど  
戸単位: 方位, 所在階, 広さ, 間取りなど

**B) 占有空室情報**  
戸単位で占有調査を行う

**調査時点:** 2015年2月より3ヶ月毎に定点観測を行う。

## 【4. Webデータ類似性についての検証】

Webデータがその用途に対してどのような性質を持っているかの概況を明らかにすることを目的に調査を行った。具体的には、まず、Webデータの収集・整理をとおこない(第二節)、その性質の調査を行った。特に、不動産鑑定士による銀行の賃貸物件リストに基づく現地調査データとの比較が本研究の特徴である。その結果、アパートローンリスク計量化モデルの構築への利用に

関連するWebデータの特性として以下のことがわかった。

• Webデータには複数の業者が一つの物件の広告を出す場合があり、統計モデルに利用するためには物件の名寄せをする必要がある。

• Webデータには部屋階数201階や賃料503万円/月など、入力間違え等による異常値が含まれる。ただし、その量は平均値を明確に変動させるほどは大きくない。

• Webデータの各物件に主成分分析を適用すると、第一主成分(最も物件特徴を説明する要因)は築年数であり、その他、水道光熱、建物の高さ・構造、部屋の広さ、入居条件、周辺環境等と解釈できる主成分が得られた。

• Webデータと鑑定士データでは、物件特性の統計特性はだまかには共通している。しかし、物件特性によっては明確に異なるものもある。

• Webデータと鑑定士データで空室数の分布は異なる。これらの相違は、「名寄せによる欠損(名寄せ欠損)」と「すべての空室報告しない欠損(一部公開欠損)」でよく説明できる。

• 埋まるまでの期間の分布は、観測期間の打ちきりにより、平均等の要約統計量が影響を受けるため、観測期間が短いときは、それを考慮して解析する必要がある。