

人流ビッグデータを用いた街の賑わいと家賃形成との関係に関する研究

秦桜蘭・秋山祐樹・小川芳樹・柴崎亮介・佐藤大誓

Study of the Relationship Between House Rental Price with Prosperity of Urban Area Defined by Person Flow Big Data

Yinglan QIN, Yuki AKIYAMA, Yoshiki OGAWA,
Ryosuke SHIBASAKI and Taisei SATO

Abstract: House price is affected by house function, surrounding environment and so on. However, there are less research of finding the relationship between house prices with urban people congestion. Therefore, we use location-based big data collected by mobile phone GPS data to define people congestion. In addition, people congestion is divided into daytime, nighttime, weekdays and holidays. We use multiple house related and surrounding environments variables to help for verifying if there are some relationship between house prices with area people congestion in Tokyo. As a result, we found people congestion have a certain influence on house price. Besides that, the results calculated by people congestion is close to the one calculated by surrounding environment.

Keywords: 家賃 (house price), マイクロジオデータ (micro geodata), 街の賑わい (people congestion), ヘドニック価格モデル (hedonic pricing model)

1. はじめに

家賃相場は建物自身の性能（建物面積，構造，築年数，付帯設備の有無等），物件の立地環境，流通市場の景気など様々な要因に影響を受けて決定される。Shimizu et al. (2007)はヘドニック価格モデル (hedonic pricing model) を採用することで日本の住宅価格の決定要因を明らかにする手法を提唱した。同モデルの中では平均通勤時間や最寄り駅までの徒歩時間など7種類の説明変数を採用している。さらに秋山ほか (2018) は建物自身の性能に関する属性以外に建物の立地環境など82種類の説明変数を用いて、スパースモデリ

ングにより家賃形成に寄与する変数を明らかにした。

これらの研究は周辺環境を説明するための変数となるデータが必要となるが，データの収集・整理のプロセスが煩雑な場合がある。また説明変数となるデータが既存統計である場合は更新頻度が低いという問題がある。そのため日々変化する家賃相場を把握していく上では，これらの既存手法では即時性が低いという課題があった。

そこで本研究では，東京23区を対象に携帯電話の移動履歴に基づく滞留人口に関するビッグデータを用いて，「街の賑わい」を定量的に推定するか否かを分析し，既往研究で家賃形成の説明に必要な膨大な説明変数の全て，もしくは一部を街の賑わいで代替できる可能性を検討した。

秦桜蘭 〒277-8568 千葉県柏市柏の葉5-1-5

東京大学 空間情報科学研究センター

東京大学大学院 新領域創成科学研究科 修士課程

Email: 1187170943@edu.k.u-tokyo.ac.jp

2. データ処理

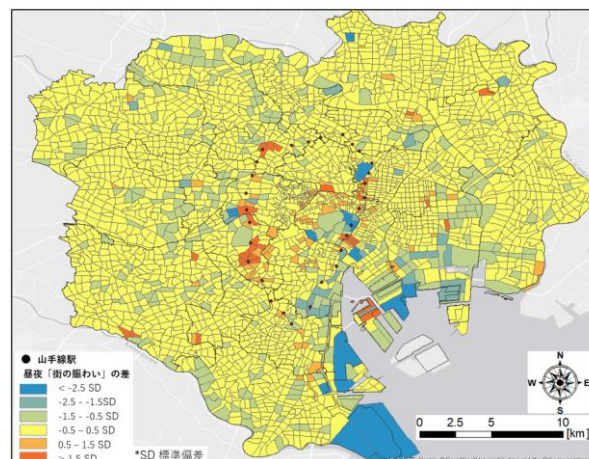
2.1 「街の賑わい」の推定

街の賑わいを定量的に推定するために、2012年の「混雑統計®」を用いてまず東京23区の各町丁目（2,538町丁目）の滞留人口を Akiyama et al.(2016)の手法を採用して推定した。本研究では一年間の総滞留人口を「街の賑わい」と呼ぶ。また日本の一般的な勤務時間に基づき、午前9時から午後6時までを「日中の賑わい」と定義し、残りの時間を「夜間の賑わい」として定義した。さらに平日（250日間）と休日（115日間）の区別も行った。この5つの賑わい同士を比較可能にするために、日ごとに算出された推定滞留人口を日平均した。図-1は、日中の賑わいと夜間の賑わいの差を表わしている。赤い部分は夜間より日中の方が賑わっている。

なお「混雑統計®」データとは、NTTドコモが提供するアプリケーション（ドコモ地図ナビサービス（地図アプリ・ご当地ガイド）等）の一部のアプリの利用者より、承諾を得た上で送信される携帯電話の位置情報を、NTTドコモが総体的かつ統計的に加工を行ったデータである。位置情報は最短5分毎に測位されるGPSデータ（緯度経度情報）であり、個人を特定する情報は含まれない。

2.2 家賃と物件の属性に関する情報の整備

家賃と物件の属性（面積、最寄り駅までの徒歩時間など）の情報は、アットホーム株式会社の「不動産データライブラリー 全国 1999-2016 データセット」を使用して整備した。なお本研究で扱う賃貸物件は賃貸マンションと賃貸アパートを含む。前処理後（外れ値、欠損値処理）の物件数は東京23区全域（2012年）で195,9627件である。また本研究では街の賑わいに関するデータの集計単位と合わせるために、町丁目ごとで集計（平均）した。図-2は町丁目ごとの物件の単位面積あたりの家賃である。



「混雑統計®」 © ZERIN DataCom CO.,LTD.

図-1 昼夜間の街の賑わいの差

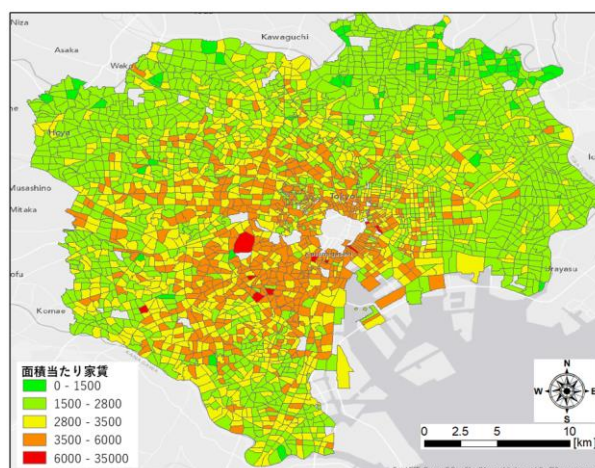


図-2 東京23区における物件の単位面積あたりの家賃（2015年）

2.3 物件の周辺環境に関する情報の整備

既往研究で使われた説明変数のうち、収集が比較的容易で、東京23区全域で利用可能な物件の周辺環境を表わす以下の変数を整備した。なおこれらは既往研究でも用いられている。座標付き電話帳DBテレポイント 法人版（株式会社ゼンリン）（2015年）を用いて、町丁目内のスーパーマーケット、コンビニエンスストア、薬局・ドラッグストア、飲食店、総合病院、専門病院、小学校、中学校、郵便局、酒場（居酒屋・バー・スナック・キャバレー等）、金融機関（銀行・信用金庫等）、デパート、ホームセンター、書店、家電量販店、クリーニング店、衣料品店、理美容店、食料品店、

ガソリンスタンド、幼稚園・保育園、高等学校、大学（全 23 種類）の件数を算出した。本稿ではこれらを「周辺環境変数」と呼ぶ。

3. 分析手法

3.1 重回帰分析

まず多数の説明変数を組み合わせることによって、5つのモデル（表-1）に分けて線形重回帰分析を行った。なお線形重回帰分析で問題となる多重共線性問題は、VIF（Variance Inflation Factor）を使って調べ、非線形問題を説明変数と目的変数の単回帰分析を使い確認した。

3.2 正則化

線形回帰分析の課題を解決するために、正規化法は有効な手法として知られている。そこで本研究では正則化回帰、Ridge 回帰、Lasso 回帰、Elastic Net を用いて線形回帰分析との違いを分析した。また評価関数として RMSE と MAE を用いて結果の評価を行った。

4. 結果

4.1 重回帰分析

表-2 は各モデルによる分析結果を示している。街の賑わいを追加することで決定係数が大きくなった。なお表-2 には示していないが、5つの街の賑わいを説明変数として全て追加することで、決定係数は0.617まで上がったが、それでも model 4 の決定係数の方が高い。この結果から街の賑わいのみ周辺環境変数を代替することは難しいことが分かる。一方で街の賑わい家賃の間にはある程度の相関が見られることも明らかとなった。

以上の結果より、街の賑わいが家賃の説明変数になれるか否かはさらに検討を進める必要がある。一方、周辺環境変数と家賃との間には明らかな相関が見られると言えよう。なお周辺環境変数同士での共線性問題も存在している。例えば専門病院と薬局・ドラッグストア間などが挙げられる。また家電量販店、書店などの P 値は有意水準

表-1 モデルの説明

Model 1	建物属性
Model 2	建物属性+街の賑わい（全体）
Model 3	建物属性+街の賑わい（昼夜）
Model 4	建物属性+周辺環境
Model 5	建物属性+周辺環境+街の賑わい（全体、昼夜、平日、休日）

表-2 各モデルによる重回帰分析の結果

	R ²	RMSE	MAE
model1	0.573	21584.5	13737.0
model2	0.585	21288.4	13679.9
model3	0.610	20659.1	13520.8
model4	0.627	20098.0	13063.2
model5	0.650	19432.0	12793.4

「混雑統計®」 © ZERIN DataCom CO.,LTD.

(P=0.05) を上回った。すなわち多重共線性問題や非線形問題はここでも表われているといえる。

説明変数を増やすほど決定係数が上昇する傾向があるため、今後より良い結果を出すためにさらに説明変数を増やす方法が考えられる。しかし数多くの説明変数間の共線性や非線形問題を手動で確認することは困難なため、線形回帰分析の課題を解決できる正規化法を採用し、予測精度を機械的に上げられるか否かを確認した。

4.1 正則化

3.2 で説明した三つの正規化回帰の結果と、一般的な線形分析の結果と比較した。ここでは物件の属性、街の賑わい、周辺環境を表わす 33 の変数を用いた。なお全データの 80% をトレーニングデータとして使用し、残りをテストデータとして予測モデルの精度を評価した。R²_[1]、R²_[2] はそれぞれに対する決定係数。また正規化回帰の最適設定は python sklearn ライブラリーの LassoCV、

RidgeCV, ElasticNetCV で行った。

表-3 から Lasso と Ridge 回帰の結果は線形回帰より僅かに上回ることが分かった。またトレーニングデータに対して決定係数は同じだが、テストデータに対する予測は Lasso が一番優れていた。

なお説明変数は 33 種類があるが、Lasso 回帰では平日の賑わいと全体の賑わいが説明変数から排除され、全説明変数は 31 種類となった。また日中の賑わいと夜間の賑わいは目的変数に対して偏回帰係数が高く、前者は正の相関となり、後者は負の相関となった。

続いて表-1 に示す各モデルに対して Lasso 回帰を適用した。その結果、全てのモデルに対して RMSE, MAE の値は線形重回帰分析より小さくなった。また model3 で決定係数が大きくなった。すなわち Lasso 回帰を適用することにより線形回帰よりも精度が良いモデルを構築することができることが明らかになった。

表-3 正規化、線形回帰による分析結果

	RMSE	MAE	$R^2_{[1]}$	$R^2_{[2]}$
Linear	18586.9	13371.6	0.654	0.647
Lasso	18522.6	13346.1	0.654	0.650
Ridge	18552.7	13362.3	0.654	0.648
Elastic Net	18787.5	13443.7	0.648	0.639

「混雑統計®」 © ZERIN DataCom CO.,LTD.

表-4 Lasso 回帰による各モデルの分析結果

Lasso	RMSE	MAE	$R^2_{[1]}$	$R^2_{[2]}$
model1	20010.0	14035.2	0.570	0.591
model2	20042.7	14085.5	0.584	0.590
model3	19275.0	13812.5	0.607	0.620
model4	19123.4	13545.1	0.629	0.626
model5	18522.6	13346.1	0.654	0.650

「混雑統計®」 © ZERIN DataCom CO.,LTD.

5. おわりに

本研究では、東京 23 区を対象に人流ビッグデータで推定した街の賑わいで家賃を定量的に推測できるか否かを分析した。その結果、Lasso 回帰の結果と街の賑わいを入れた model が示すように街の賑わいは家賃をある程度推測できる可能性が示唆された。また分析手法を変更することで予測精度を向上させることができたため、今後分析手法をさらに改善することでより精度の高いモデルを実現できるものと期待される。

謝辞

本研究は、東大 CSIS 共同研究 No. 884 の一環として実施した。また一般財団法人住総研研究助成（選奨報奨枠）、「マイクロジオデータを用いた日本全国の家賃形成メカニズムの解明」

(2018 年度) の一環として実施した。ここに記して謝意を表したい。

参考文献

秋山祐樹・小川芳樹, 2018, マイクロジオデータを用いた家賃形成メカニズムの研究—住まい手からみた住宅の価値と市場価値の乖離に着目して—, 住総研研究論文集・実践研究報告集, 44, 1-12.

Akiyama, Y., Ueyama, S., Shibasaki, R. and Adachi, R., 2016. Event Detection Using Mobile Phone Mass GPS Data and Their Reliability Verification by DMSP/OLS Night light Image, ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, III-2, 77-84.

Shimizu, C., and K.G. Nishimura, 2007. Pricing Structure in Tokyo Metropolitan Land Markets and Its Structural Changes: Pre-Bubble, Bubble, and PostBubble Periods, Journal of Real Estate Finance and Economics 35, 475-496.