

群馬県前橋市における自治体保有データを活用した
将来の空き家分布推定手法の検討
馬場弘樹・秋山祐樹・谷内田修

Predicting the Spatial Distribution of Vacant Housing Taking Advantage of
Municipality-Owned Data in Maebashi

Hiroki BABA, Yuki AKIYAMA, and Osamu YACHIDA

Abstract: Increasing a number of vacant houses causes serious problems such as deterioration of cityscape. This study develops an estimation method for spatial distribution of vacant houses taking advantage of municipality-owned data. With the help of XGBoost, a machine learning method, a future distribution and probability of vacant houses are predicted. As a result, 85.2% of the houses are correctly detected using the test data. Although the method seems to work, we need to develop a predictive model and compare the estimated values to the actual judgement of vacant houses.

Keywords: 空き家 (vacant housing), 分布推定 (distribution estimation), 住民基本台帳 (basic resident register), 水道使用量 (use of tap water), 固定資産税台帳 (property tax ledger)

1. はじめに

我が国で空き家は増加傾向にあり、2018 年住宅・土地統計調査では過去最高の 846 万戸と推計された。過剰な空き家ストックは外部不経済を誘発する恐れがあり（浅見, 2014）、自治体は空家等対策計画の策定をはじめとする対策に腐心している。しかしながら、空き家の地理的分布は時間によって変化するものであり、小規模自治体にとって空き家の分布調査を継続的に行うことは費用や人材の面から容易ではない。

これまでの空き家の地理的分布を把握する手法は、目視による現地調査が主であり、一定の基準に基づき空き家判別を行ってきた。物的状況から空き家の有無を判断した場合、その背景にある居住世帯の状況や建物特性が空き家と相関を持

つことがわかっている（Baba and Hino, 2019）。従って、自治体が保有するデータ、例えば住民基本台帳などは空き家を特定する際に有用であると考えられ、これまでも自治体保有データを用いた空き家推定の研究が行われてきた（秋山ら, 2018; 熊谷ら, 2018）。しかし、既往研究は横断面データを用いており、一部の地域から対象地域全体の空き家分布を推定する際には有用である一方、将来予測を対象としている訳ではない。

そこで本研究は、自治体保有データから空き家分布推定手法を予測するための手法を検討する。具体的な手法は過年度データから空き家推定モデルを構築し、将来時点での空き家予測確率地図を作成する。将来的には空き家現地調査のためのツールとして、その予測地図を実装することを狙いとしている。

対象地は群馬県前橋市（以下、市）中心部とする（図-1）。市は東京都心から約 100 km 圏であり、

馬場 弘樹

所属：東京大学大学院工学系研究科

E-mail: hbaba@csis.u-tokyo.ac.jp

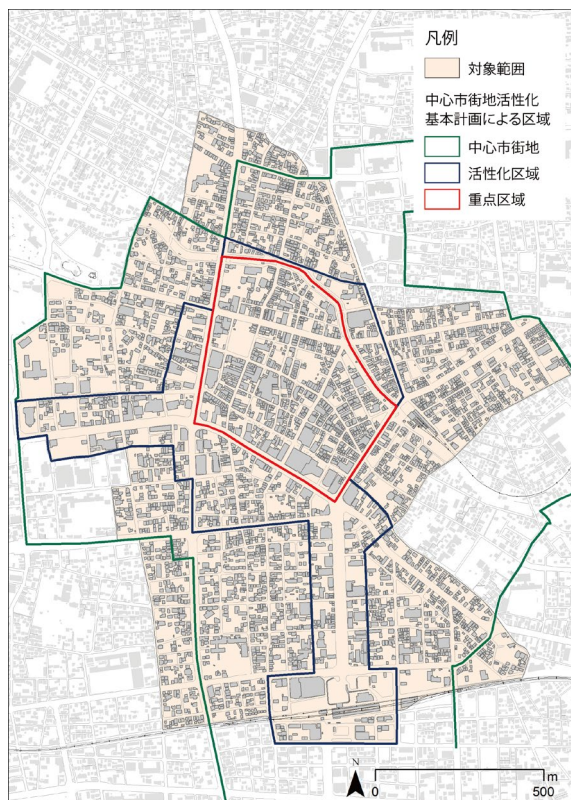


図-1 前橋市における中心市街地活性化区域

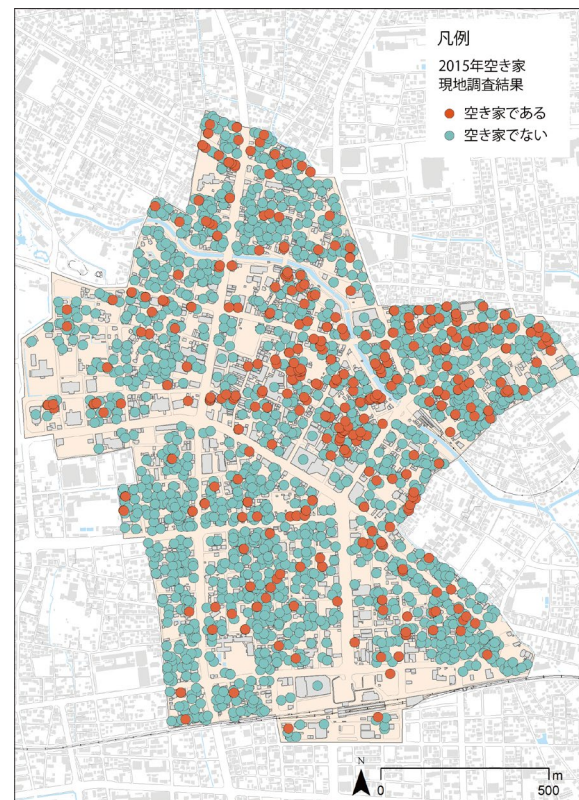


図-2 2015年現地調査での空き家判定

東京中心部への人口流出が顕著な地域である。市はそのような状況を鑑み、中心市街地活性化基本計画において空き家への対応を重要な項目としている。このように、市は大都市圏に隣接する中核市として一般的な特質を有すると考えられ、分析対象として相応しいと考えられる。

2. 手法

2.1 データ

本研究では、市の協力の下、自治体保有データとして住民基本台帳（以下、住基）、水道使用量、家屋固定資産税台帳（以下、固定資産）を用いて分析を行った。住民基本台帳には登録された住民の年齢、性別等が記載されており、2013年11月1日現在、2017年3月31日現在の2時点のデータを使用した。水道使用量は毎月の使用量が原デ

ータで得られたため、それらを2014年度、2018年度でそれぞれ集計して利用した。固定資産税台帳（家屋）は2013年1月1日現在、2018年1月1日現在の2時点を得ており、建物構造、建物用途、建築年の情報を抽出した。自治体保有データには住所が併記されているため、東京大学空間情報科学研究センターが提供するCSVアドレスマッチングサービスを利用して緯度経度座標に変換した。

データの集計は建物単位で行っており、2016年ゼンリン住宅地図の建物ポリゴンを利用し、建物固有IDを用いて自治体保有データをリンクさせた。本研究では戸建住宅に着目するため、建物属性において個人家屋と、名称が存在しない又は名称秘匿に分類される建物を利用した¹。

モデル推定の正解データである空き家の空間

¹ なお、ゼンリン住宅地図上では個人家屋になっている場合でも固定資産税台帳上で共同住宅となっている場合がある（対象サンプル内では27件）。これは、個人住宅

の一部を賃貸に出していると考えられるが、本研究では併用住宅の延長として除外せずに分析対象とした。

表-2 データ集計の結果

	行政データ			
	住基	固定資産	水道	空き家現調
サンプルサイズ	10,572	3,635	9,997	534
建物重複削除後 サンプルサイズ		1,671	2,551	
ジオコーディング 成功サンプル	9,375	1,374	2,418	
補足率[%]	88.7	82.2	94.8	-
建物と結合出来た サンプル	9,093	1,253	2,293	522
全数からみた採用 率[%]	86.0	75.0	89.9	97.8
ゼンリン建物ポリゴンを住所用途に絞り、建物単位で カウントした場合の住宅建物数: 2,587				
結合建物数	1,325	802	1,460	375
建物データ対応率 [%]	51.2	31.0	56.4	14.5

分布は、前橋工科大学が行った中心市街地での現地調査結果を用いた（図-2）。当該調査は2015年7月から9月、2016年2月にかけて行われ、空き家立地の緯度経度座標と空き家の評価が記載されている。同調査では空き家評価を損傷の程度や流通されているか等で分類しているが、本研究では、空き家損傷の有無に関わらず流通されていない物件を空き家として定義し、その有無を推定することとした。空き家の現地調査結果は緯度経度座標を直接得ており、170件が建物ポリゴンと対応した。なお、建物ポリゴンのうち、いずれかの自治体保有データと結合したものは1,509件であり、本研究ではこれらを分析対象とした（表-1）。

2.2 推定手法

本研究では、機械学習的分類手法のひとつであるXGBoost（eXtreme Gradient Boosting）を用いて空き家の分類を行った（Chen and Guestrin, 2016）。これは、アンサンブル学習のひとつで、ランダムフォレストを基本として弱学習器である決定木間の重み付けに勾配ブースティングという手法を合わせたものである。当該手法は推定精

度が高く、欠損値が混入していても処理が可能である一方、パラメータチューニングを行う必要があり、複雑なデータでは汎化性能が下がる、という欠点がある。

本研究で用いた特徴量は、建物人員内最高年齢、最少年齢、建物人員数、住基有無ダミー、建物構造ダミー、建物用途ダミー、建物面積ダミー、水道使用量である。特徴量の二乗項、交差項もモデルに含め、合計144の特徴量を用いた。まず、時間変化する特徴量について2013年住基、2014年度水道使用量を用いてモデルを構築し、2015年空き家現地調査を正解データとしてパラメータチューニングを行った。その後、構築されたモデルを用いて2019年空き家率予測確率を算出した。その際、時間不変の特徴量はそのまま用い、時間変化する特徴量は2017年住基、2018年度水道使用量を用いた。

3. 結果

はじめに、構築したモデルと2015年空き家現地調査データを対照させ、パラメータチューニングを行った。その結果、チューニングしたパラメータは、木の深さ(max depth) = 8.0、学習率(eta) = 0.01、子ノードでの必要な最小の重み(min child weight) = 2.0、各ステップでの列数サンプリング比率(colsample by tree) = 1.0、各ステップでのサンプリング比率(subsample) = 1.0、ブースティング反復回数(nrounds) = 1,644となった。表-2は現地調査結果と検証データによる予測空き家の有無をクロス集計し、モデルの適合度を検証したものである。当該表によれば正当率は83.7%であるが、空き家と予測されていても現地調査で空

表-1 検証データによるモデルの適合度

検証データ		2015年時点予測空き家	
		Yes	No
2015年 現地調査	Yes	12	37
	No	37	367

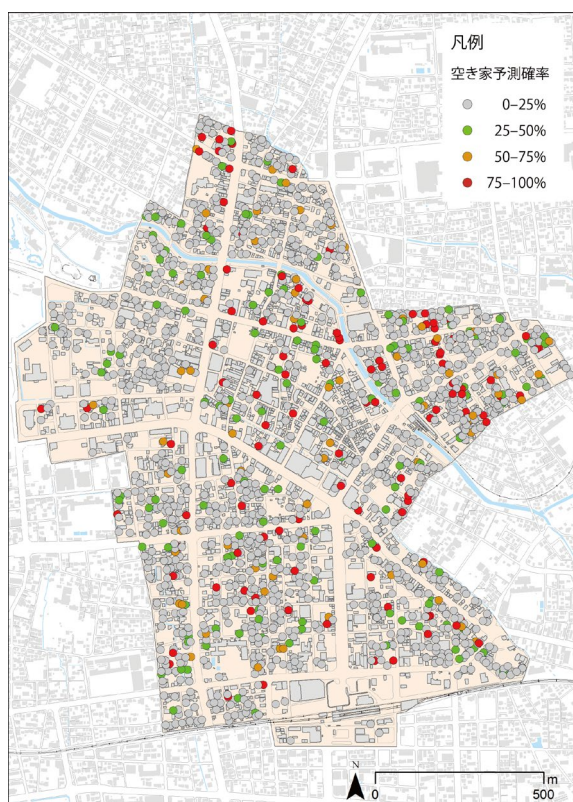


図-3 2019年時点空き家予測確率の分布

き家ではない割合は 12/49 (24.5%) であり、高いとはいえない。これは、データ数が十分でないことや、過学習を防ぐためにパラメータチューニングしていることが原因であり、今後の課題である。なお、表-2 は全サンプルの内 7 割を訓練データとしてモデルを構築し、残り 3 割の検証データに対する適合度を集計しており、予測モデルに対する適合性をみている訳ではないことに注意が必要である。

続いて、構築されたモデルを用いて 2019 年時点の空き家予測確率の推定を行った。図-3 はその地理的分布を示している。空き家予測確率の高い建物は地理的に分散しており、必ずしも空き家の空間的分布が地理的法則性をもつとは言えない。これは、住基データから抽出される年齢等の情報が有効に作用する一方で、地理的な規則を持って分布しないためであると考えられる。図-3 の分布を図-2 と比較すると、対象地東部では 2015 年時点の空き家分布を反映しているように見えるが、

南東部では必ずしも空き家予測の分布が対応しているとは言えない。これは、対象地南東部で建物人員内最高年齢などの特徴量に変化があったためだと考えられる。

4. おわりに

本研究では、自治体保有データから空き家分布を予測するため、XGBoost を用いて 2015 年空き家分布を教師データとして 2019 年空き家分布を推定した。本研究では、横断面データでのモデル検証しかできず、全体の正答率は 83.7% であるものの、空き家と予測されたものの正答率は 24.5% と高いとはいえなかった。今後、モデルの精緻化を図るとともに、対象地において現地調査を行う予定である。現地調査で 2019 年の空き家教師データを作成するとともに、空き家予測確率のデータを調査の際に参照し、実務における予測データの有用性について議論する予定である。

謝辞

公共データの利用にあたり、群馬県前橋市未来の芽創造課をはじめ、関係者の皆様に厚く御礼申し上げます。

参考文献

- 秋山祐樹, 上田章紘, 大野佳哉, 高岡英生, 木野裕一郎, 久富宏大. 2018. 鹿児島県鹿児島市における公共データを活用した空き家の分布把握. 日本建築学会計画系論文集, 83(744), 275-283.
- 浅見泰司. 2014. 都市の空閑地・空き家を考える. プログレス.
- 熊谷樹一郎, 植松恒, 小野裕基, 山本純平. 2018. 低未利用空間のモニタリングを目的とした空き家推定モデルの構築. 土木学会論文集 F3 (土木情報学), 74(2), I_51-I_58.
- Baba, H. and Hino, K. 2019. Factors and tendencies of housing abandonment: An analysis of a survey of vacant houses in Kawaguchi City, Saitama. *Japan Architectural Review*, 2(3), 367-375.
- Chen, T. and Guestrin, C. 2016. Xgboost: A scalable tree boosting system. *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785-794.