

機械学習の解釈手法を用いた空き家の 発生要因分析

水澤克哉 宮本慧 田村将太 田中貴宏

中市新太
富山県立大学 情報システム工学科
u020025@st.pu-toyama.ac.jp

January 17, 2023

背景

2018 年、日本の空き家率は、それまでの最高値となる 13.6 %を記録した。特に人口減少や高齢化が進む地方都市 では、空き家増加の傾向が顕著となっており、空き家は老朽化による倒壊、景観の悪化、放火による火災等の様々なリスクを伴うとされており、さらにはインフラ管理の非効率化にもつながるため、空き家発生を未然に防ぐ取組や、空き家の放置を防ぐ取組が求められている。

目的

本研究では機械学習の解釈手法を用いて建物の属性や地区特性の影響を可視化する手法を構築することを目的とし、空き家発生要因の分析を行う。最終的には、空き家発生が予想されるエリアをその要因とともに把握することが可能な「適材適所の空き家対策支援ツール」の提案につなげることを意図している。

対象地

本研究では、山口県防府市を対象とする。防府市の人口は 11.4 万人（21 年 12 月）で、近年は人口減少傾向にあるが、住宅棟数は増加傾向にある。空き家の棟数は、平成 27 年度は 1538 棟であったが、令和元年度には 2372 棟と、約 1.5 倍に増加した。防府市の人口はさらに減少すると予測されており、今後も空き家の増加が見込まれるため、空き家対策が必要と考えられる。

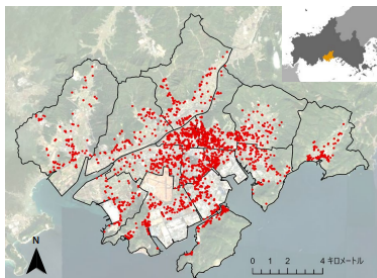


図 1: 防府市の空き家分布

研究の流れ

- 既往研究を参考に、建物属性や地区特性を含む、空き家発生と関連がある指標を抽出し、予測に用いる特徴量データを *GIS* データとして整備する.
- 空き家データ (*GIS* データ) と、(1) で整備した特徴量データを用いて機械学習アルゴリズムの一つである *XGBoost* を用いた空き家発生予測モデルを構築する.
- 機械学習モデルの解釈手法を用い、空き家の各発生要因の影響度の大小の比較や、各発生要因と空き家予測値との関係を把握、エリア別の影響度の大小を比較し、予測結果の解釈を行う.

特徴量抽出

下図の指標をもとに、発生要因を「建物」「人口」「周辺建物」「周辺施設」「地形」「道路」「土地利用」「立地」の8つの要因カテゴリに整理する。

発生要因 カテゴリ	特徴量			
建物要因	建物用途(ダミー変数)*[-]/階数[-]/構造(ダミー変数)*[-]/建築面積[m ²]/延床面積[m ²]/建築年[年]/敷地面積[m ²]	道路要因	<ul style="list-style-type: none"> ・前面道路情報 接道長さ[n]/建築基準法上の道路区分(ダミー変数)*[-]/路線価[千円]/前面道路幅員-狭幅員側[n]/前面道路幅員-広幅員側[n]/袋小路ダミー変数[-]/袋小路の長さ[n]/袋小路に接する家屋数[棟] ・周辺道路情報 幅員15m以上の広幅員道路までの距離[n]/幅員4m以上15m未満の中幅員道路までの距離[n]/水路・管渠までの距離[n] 	<ul style="list-style-type: none"> ◆用途地域ダミー変数[-] ◆周辺80m圏内の土地利用面積[m²] 田/畑/山林(開発可能地)/山林(開発不適地)/水面/自然地(開発可能地)/自然地(開発不適地)/住宅用地/商業用地/工業用地/農林漁業施設用地/公共施設用地/道路用地/交通施設用地/公共空地/その他の公的施設用地/その他の空地/平面駐車場
人口要因	周辺人口[人]/周辺高齢人口[人]			
周辺建物要因	周辺100mの新築数(H24/H28)[棟]/周辺100mの建物数[棟]/周辺100mの建築面積[m ²]/周辺50mの建築面積[m ²]			
周辺施設要因	◆最寄りの施設までの距離[m] 食料品店/行政施設/バス停/小学校/老人福祉施設/幼保施設			
地形要因	標高[n]/傾斜度[%]/想定最大浸水深[n]			立地要因 大字ダミー変数[-]

図 2: 特徴量

XGBoost

勾配ブースティング木の実装の一つである XGBoost を用いることとした。XGBoost を扱う利点としては、決定木の分岐により欠損値を扱うことができるため、欠損値を含む可能性がある自治体データを用いた分析に適合する点、モデルに予め一定の確率分布を仮定しないため、空き家推定のような複雑かつ非線形性を有する予測に有用である点が挙げられる。

並び替え特徴量重要度 (PFI) を用いた分析

6/9

PFI

PFI とは、ある特徴量の値をシャッフルすることで得られる予測誤差の増加分を特徴量の重要度とする手法である。シャッフルを 100 回繰り返して得られた予測誤差の平均値を用いた。その結果、建築年や敷地面積、延床面積など、建物の性能に関する指標のほか、中幅員道路までの距離や水路・管渠までの距離などの指標の影響が大きいことが明らかとなった。

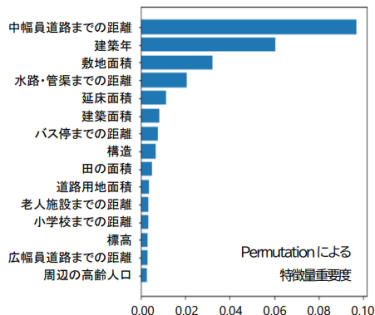


図 3: 要因別の影響度

PDP

PDP は、機械学習におけるモデル解釈手法の一つであり、「ある特徴量が大きくなると予測値がどのように変化するか」という関係を捉えたい場合に有効であり、興味のある特徴量だけを動かした場合の予測値の変化をプロットすることで得られる。下図から、上位 1 番目の中幅員道路までの距離を見ると、距離が 5m より大きくなると予測値が高くなる傾向がみられる。これは自動車が通行可能な程度の幅員の道路に、建物からアクセスできるか否かが大きく影響することを示していると考えられる。

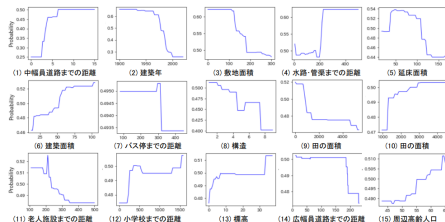


図 4: 上位 15 番目までの PDP

SHAP

SHAP は、2017 年に提案されたツリー構造や DNN 等の複雑なモデルを解釈するための統一的なフレームワークの一つである。SHAP を導入する大きなメリットとして、SHAP 値テーブルによって任意のサンプル群に対して、各特徴量の寄与度を求められることが挙げられる。これによって、ある特定のエリアにおいて影響度の強い特徴量を抽出することが可能となる。

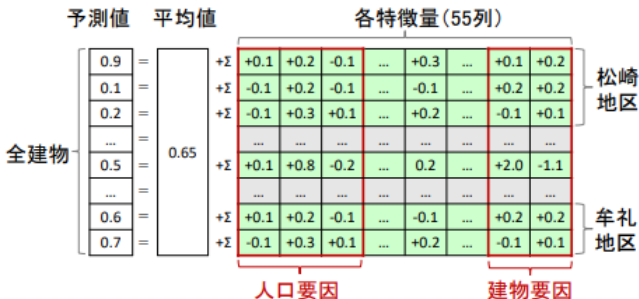


図 5: 上位 15 番目までの PDP

総括

PFI を用いた分析からは、道路要因や建物要因の影響度が強いことが明らかとなり、PDP を用いた分析からは、自動車が通行可能な程度の幅員の道路に、建物からアクセスできるか否かが大きく影響することや、戦前から高度成長期にかけて建設された住宅で空き家が多いこと等が明らかとなった。SHAP を用いた分析では、各地区別に各発生要因（カテゴリ）の影響度を整理し、各地区の特性によって空き家発生の要因の傾向が異なることを示した。

今後

本分析結果を地図上にマッピングし、空間的に空き家発生リスクや空き家発生要因を把握可能なツールの開発をし、最終的には、各自治体の職員が比較的容易に扱うことができ、自治体に保管されている各種データベースから、空き家の発生に関連するデータを自動的に抽出・可視化できるツールの開発が必要と考えられる。