

機械学習の解釈手法を用いた空き家の発生要因分析

-適材適所の空き家対策支援ツールの開発を目指して-

Analysis of Occurrence Factors of Vacant Houses by Interpretability Methods in Machine Learning

-Aiming to Development of Tools that Support Measures to Solve Vacant House Problems, which is the Best Fit for the Place -

水澤克哉*・宮本慧**・田村将太***・田中貴宏***
Katsuya Mizusawa*・Shota Tamura**・Takahiro Tanaka***

In recent years, the declining population and the corresponding increase in vacant houses has become an issue, and efforts to prevent the occurrence of vacant houses are needed. Thus, areal measures are needed against the vacancy of houses in the future for urban planning. Therefore, in this study, we aimed to develop a “vacant house countermeasure support tool” that can spatially identify “which buildings in which areas are likely to become vacant,” and proposed a method for analysing occurrence factor of vacancy of houses using machine learning interpretation techniques. As a result, it became clear that the performance of buildings and the condition of infrastructure affect the occurrence of vacant houses and that the factors that cause the occurrence of vacant houses differ according to the characteristics of each district, and the degree to which these factors affect the occurrence of vacant houses also differs.

Keywords: Vacant House, Prediction Model, Machine Learning, Interpretability Methods

空き家, 予測モデル, 機械学習, 解釈手法

1. はじめに

1-1 研究背景と目的

2018年、わが国の空き家率は、それまでの最高値となる13.6%を記録¹⁾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いる。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²⁾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。

このような状況の中、平成26年に「空家等対策の推進に関する特別措置法（以降、空き家特措法）」が施行され、全国の自治体が法的拘束力をもって空き家の情報収集や利活用を行うことや、空き家対策への財政上、税制上の措置等が可能となった。しかし、空き家特措法に基づく空家等対策計画の多くは、改修・解体補助、マッチングといった個々の建物や所有者に対する「点としての対策」が多く、道路整備や便益施設の適正な配置といった、広がりのあるエリアを対象とした「面としての対策」の位置づけは曖昧な点が多い。しかし、空き家発生には、建物の老朽度や個人属性等の建物属性に加え、空き家周辺の人口や世帯の増減、道路整備の状況といった地区特性が影響することが指摘されており³⁾、適材適所の効果的な空き家対策を行うためには、「点」と「面」の両面から対策を検討する必要があると考えられる。

近年では、対象区域を定めて空き家解体補助金を支出する等、「面」的に空き家対策を行う自治体が見られるようになつた。本研究の対象地である防府市では、空き家や狭あい道路を一体で解消することを目的とした「防府モデル事業」⁴⁾が実施されているが、補助の対象となるエリアの設定根拠を示すことが難しく、制度の運用面に課題を有し

ていると考えられる。このような課題に対応するためには、建物属性と地区特性の両面より空き家の発生要因を明らかにし、「どのエリアで、どのような建物が空き家化しやすいか？」を空間的に把握する必要があるが、現状では、そのような傾向把握のための基礎データの整備が不足していると考えられる。

そこで本研究では、機械学習の解釈手法を用いて、建物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。

1-2 既往研究との関連

空き家の発生要因を扱った研究には一定程度の蓄積があり、山下ら⁵⁾は町丁目単位で空き家率を推定するロジスティック回帰分析を行い、西浦ら⁶⁾は推定空き家数を算出するポアソン回帰分析を行っている。しかし、これらは一定の確率分布を仮定するモデルを用いた分析であり、複雑性と非線形性を有する空き家予測において、精度や適合率の観点から限界があることが指摘されている⁷⁾。このような課題に対して、近年では高精度な機械学習モデルの採用が増えつつある。例えば、GBDT (Gradient Boosting Decision Tree) やDNN (Deep Neural Network) は、線形性を仮定しない表現力の高いモデルであるため、複雑な非線形性を有する問題の予測に適している。秋山ら⁸⁾は、自治体データを活用し、高精度な機械学習手法の一つであるXGBoostを用いて空き家確率を推定するモデルを構築することで、高い予測精度を得ている。また北島ら⁹⁾は、道路情報や公共交通網等の画像データと深層学習を組み合わせることで、空き家の発生予測を行っている。しかし、これらの研究は空き家の分

* 正会員 (株)建設技術研究所 (CTI Engineering Co., Ltd.)

** 正会員 (株) パスコ (Pasco Corporation)

*** 正会員 広島大学大学院先進理工系科学研究科 (Hiroshima University)

布推定に主眼を置いた研究であり、様々な空き家の発生要因を網羅的に抽出し、その影響度の大小等を比較した研究の蓄積は少ない。

また、機械学習手法は高い精度を有する一方で、予測結果の解釈が困難というデメリットも存在する。空き家対策を考える上では、それぞれのサンプルの予測値に対して「モデルがなぜそのような判断を行ったのか」を解釈することが重要であり、そのような解釈を行うことで初めて、対策検討に向けた有用な知見が得られると考えられる。高精度な機械学習モデルを利用した空き家関連の既往研究では、このような空き家対策への適用可能性は考慮した提案はなされていないため、本研究ではこの課題に焦点を当て、機械学習のモデル解釈手法を用いることとした。

2. 研究概要

2-1 対象地概要

本研究では、山口県防府市を対象とした。防府市は、市北部から瀬戸内海に向かって一級河川の佐波川が流れおり、河口付近の開けた平野部に市街地が形成されている。人口は 11.4 万人（2021 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^{補注 1}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¹⁰。防府市の人口はさらに減少すると予測されており¹¹、今後も空き家の増加が見込まれる

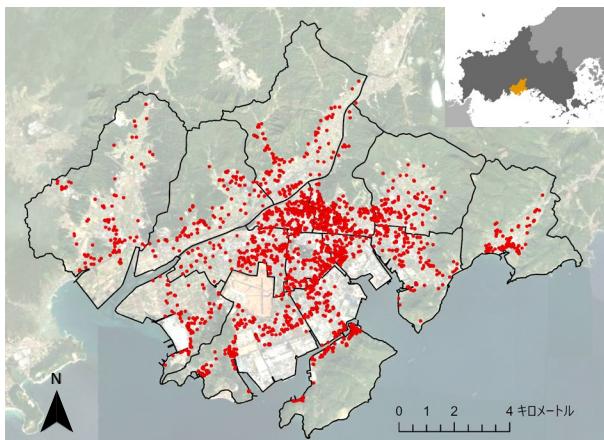


図 1 防府市の空き家の分布



図 2 対象建物

ため、空き家対策が必要と考えられる。なお、空き家発生の傾向を土地利用や人口密度等の観点より把握するため、本研究は防府市都市計画区域内を分析対象とした（図 1）。

2-2 研究の流れ

本研究の流れを以下に示す。

(1) 空き家発生要因の抽出と特徴量の整備

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。

(2) 空き家発生予測モデルの構築

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。

(3) 機械学習の解釈手法を用いた空き家発生要因の分析

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。

本論文では、(1)で抽出した指標を「要因」と仮定し分析を行うが、因果関係の証明には因果推論の手法を用いる必要があり、分析結果について因果関係を保証するものではない点に留意を要する。その上で、既往研究等で述べられている、空き家発生のメカニズムと対応する指標については、一定の影響を有するものとし、考察を行う。

2-3 使用データ

空き家データについては、令和元年度に行われた防府市空き家実態調査の結果を使用した。本研究で対象とする空き家は、防府市都市計画区域内の空き家特措法第 2 条に定める、「空家等及び特定空家等」とし、損傷の有無に関わらず、長期間使用されていないことが常態であるものを対象としているが、公共施設や工場、倉庫等は除外した。よって、令和元年度の空き家 2372 棟のうち、この条件に該当する 2087 棟を空き家として扱うこととし、同都市計画区域内の空き家以外の建物と合わせて 59935 棟の建物を次章以降の分析に用いた（図 2）。

特徴量データの整備にあたって、建物要因や周辺建物要因、道路要因といった特徴量については、都市計画基礎調査の結果や筆界データを用い、人口要因については令和 3 年度 6 月時点の住民基本台帳の情報を用いた。また、地形要因の特徴量については、基盤地図情報の標高 DEM から各建物が位置する標高と傾斜度を GIS 上で計算した。これらの特徴量は、詳細な傾向の把握のために各建物単位で集計を行い、各建物の重心点をデータの起点としている。また、都市計画基礎調査データをベースマップに用いているため、地図精度は 2500 分の 1 程度である。

3. 空き家発生予測モデルの構築

3-1 概要

本章では、2 章で作成した特徴量データから、空き家発生の有無を予測するモデルを構築することとした。まず、

①データセットを学習データとテストデータに分け、②学習データを用いて教師あり学習を行い、交差検証を行うことで、パラメータグリッドの中から最良のパラメータを選択した。その後、③最良のパラメータを用いて再度学習を行い、④テストデータに対するスコアでモデルの評価を行った。

3-2 空き家発生要因の抽出と特徴量の整備

既往研究^{12),13),14),15),16),17),18),19)}で用いられている空き家発生と関連のある指標とともに、構造、階数、敷地面積等の建物に関する点的な情報である建物属性と、周辺の人口や高齢者人口、インフラの整備状況、地形的な条件、土地利用等の面的な情報である地区特性にデータを分類した。次に、説明性・解釈性を考慮して、本研究ではこれらの指標とともに、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した。得られた発生要因カテゴリから、モデルに

表1 発生要因カテゴリと特徴量

発生要因 カテゴリー	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]
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[m]/傾斜度[%]/想定最大浸水深[m]
道路要因	・前面道路情報 接道長さ[m]/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[m]/前面道路幅員-広幅員側[m]/袋小路ダミー変数[-]/袋小路の長さ[m]/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[m]/幅員4m以上15m未満の中幅員道路までの距離[m]/水路・管渠までの距離[m]
土地利用 要因	◆用途地域ダミー変数[-] ◆周辺80m圏内の土地利用面積[m^2] 田/畠/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/その他公の施設用地/その他の大空地/平面駐車場
立地要因	大字ダミー変数[-]

* 業務施設、商業施設、宿泊施設、商業系用途複合施設、住宅、共同住宅、店舗等併用住宅、店舗等併用住宅、店舗等併用共同住宅、作業所併用住宅 で区分

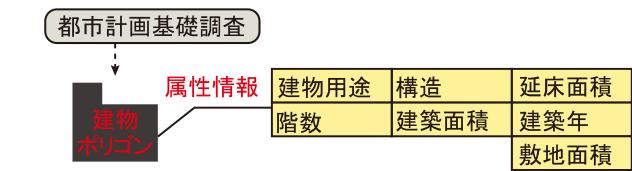
** 木造・土蔵造, 鉄骨鉄筋コンクリート造, コンクリート造
骨造, 軽量鉄骨造, 組積造, 不明, で区分

鉄道、軽量鉄道、組合道、不明 で区分
*** 1項1号、1項2号、1項3号、1項5号、2項(一括指定)、2項(個別指定)、許可済み道路、非道路 で区分

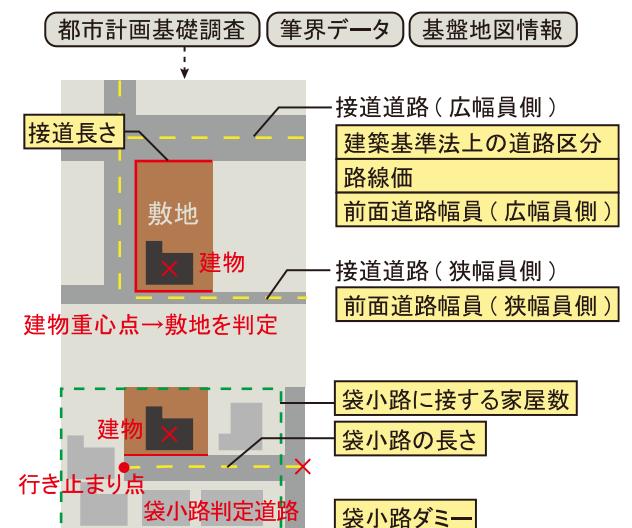
投入する特徴量を整理したものを表1に示す。

各特徴量の整備手法について図3に示す。建物要因の特徴量は、主に建物の性能を示す指標であり、都市計画基礎調査の結果を用いた。人口要因の特徴量は、周辺の人口の特性を示す指標であるが、空き家は住民基本台帳における

(1) 建物属性データを利用



(2) 建物重心点から位置情報を取得



(3) 建物重心点から周辺情報を取得

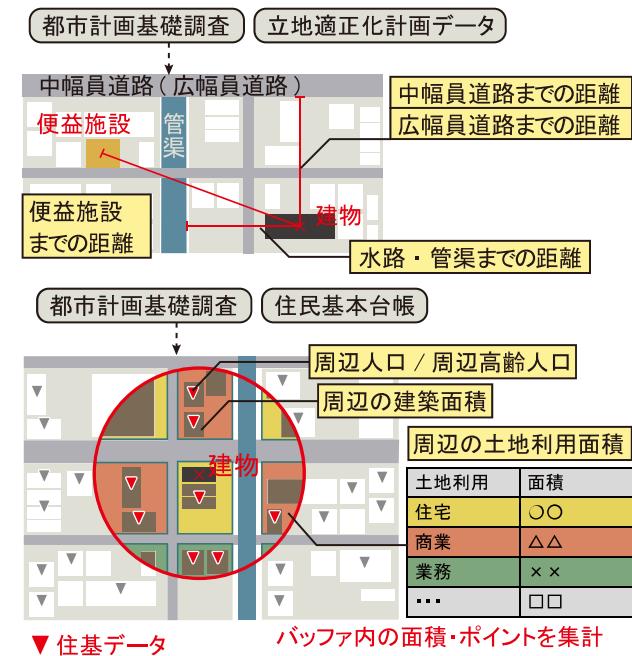


図3 各特徴量の整備手法

現住所として登録されていない場合が多いため、この情報を建物の属性情報に直接付与して予測を行うと「情報の有無」が直接的に空き家か否かの判定に影響を与える可能性がある (Leakage の問題)。このような考えのもと、建物データと住民データを直接紐づけておらず、建物重心点からの 100m バッファ内のポイントデータの値を平均したものを属性情報として付与した。周辺建物要因の特徴量は、周辺の建物の混雑度や老朽度を示す指標である。周辺施設要因の特徴量は、主に建物の利便性を示す指標であり、立地適正化計画のデータをベースに各施設が位置する座標と各建物の座標間の直線距離を属性情報として付与している。地形要因の特徴量は、地形的条件やハザードに関する指標であり、標高や傾斜度のほか、平野部に位置する対象地の特性を考慮し、ハザードマップの情報から想定最大浸水深を属性情報として付与した。道路要因の特徴量は、前面道路に関する指標と周辺道路に関する指標からなり、周辺のインフラの強度を示す指標である。前面道路については、接道長さや前面道路幅員情報のほか、袋小路が多い対象地の特性を考慮し、前面道路が袋小路か否か (袋小路ダメー) 等も追加した。周辺道路については、建物から自家用車でアクセス可能な道幅の道路 (中幅員道路、広幅員道路) までの距離等を属性情報として付与した。土地利用要因の特徴量は、周辺の土地利用状況を示す指標であり、建物が位置する地点の用途地域のほか、都市計画基礎調査データを用いて、対象建物周辺 80m バッファ内の各土地利用面積を集計した値を属性情報として付与した。また、土地利用要因の各指標に関しては、今後の土地利用規制や誘導等で適正化を図ることが可能と捉え、中・長期的にコントロール可能な変数として位置づけている。

3-3 モデルのアルゴリズム

これまで空き家推定には、重回帰分析等の線形モデルが多く用いられてきたが、空き家予測は複雑かつ非線形性を有することから、精度や適合率の観点から限界が指摘されている。そこで本研究では、空き家発生確率の推定のために、勾配ブースティング木の実装の一つである XGBoost (eXtreme Gradient Boosting)²⁰⁾を用いることとした。勾配ブースティング木は、複数の決定木を組み合わせることで、より強力なモデルを構築するアンサンブル学習法の一つである。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^{補注21)}。

3-4 サンプリングの手法とパラメータの調整

本研究で使用する空き家データセット (空き家と空き家以外の建物) は、空き家のクラスが空き家以外のクラスと比較して少なく、クラス比に偏りがある (0.97 : 0.03)。このような不均衡データを用いて機械学習によるクラス分類を行うと、データの不均衡性によって少数派クラスの分類

精度が低くなる問題が指摘されている²²⁾。これは、少数派群に誤って判別するよりも、多数派群を誤って判別する方が判別にかかるコストが高いためである。この問題に対する代表的アプローチの一つとして、サンプリングによって多数派群と少数派群のバランスがとれたデータセットで学習させる方法があり、本研究では、アンダーサンプリングを採用した^{補注23)}。また、アンダーサンプリングにおいてランダムサンプリング等を行うと、特徴量の分布が崩れる可能性がある。そこで本研究では、多数派群のサンプリングにおいて、k-means 法によるクラスタリングを行い、各クラスターの大きさ (サンプル数) に等しい割合で各クラスターからサンプリングを行うこととした。

XGBoost は表現力の高いモデルであるため学習データへの過剰適合が発生しやすく、学習に関わる様々なパラメータをチューニングすることで過学習を抑制する必要がある。本研究では、簡易的にグリッドサーチ^{補注24)}によるチューニングを行った (表 2)。また、適合率と再現率がいずれも

表2 パラメータ調整範囲と調整結果

	調整範囲	調整後
max_depth	3 ~ 6	3
min_child_weight	1 ~ 4	2
gamma	0 ~ 1	0
subsample	0.1 ~ 0.9	0.9
colsample_bytree	0.5 ~ 1.0	1.0

高いモデルを重視するため、モデルの評価関数には PR 曲線の AUC (以下 PR_AUC) を用いた。モデルの評価は、5 分割交差検証のスコアを用いて行い、PR_AUC が最善のパラメータセットをチューニングパラメータとして採用した。

3-5 精度検証と結果

テストデータにおけるクラス分類の結果を表 3 に示す。ここでは、予測値に対して F1 スコア (再現率と適合率の調和平均) が最大となる閾値で分類した場合の結果を示している。結果は、空き家以外の判別精度 90.6%、空き家の判別精度が 35.7% であり、F1 スコア = 0.192 であった。空き家の発生には、本研究で扱った要因に加え、個々の事情

表3 テストデータの予測結果

		正解		正解率	0.886
		空き家以外	空き家	再現率	0.358
予測	空き家以外	15687	436	適合率	0.131
	空き家	1615	243	F1スコア	0.192

(例: 所有者の居住地、所有者の家族構成等) の影響も大きく、また本研究の意図が空き家発生の正確な予測ではなく、発生しやすいエリアの特性把握であることから、この結果に大きな問題はないと考えた。

4. 空き家発生要因の分析

4-1 概要

機械学習の解釈手法を用いて、3 章で作成したモデルの予測結果の解釈を行った。なお本章で用いた手法は、森下

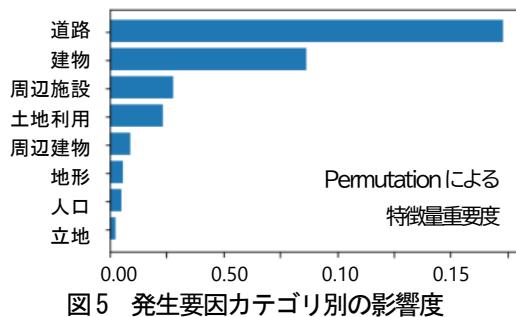
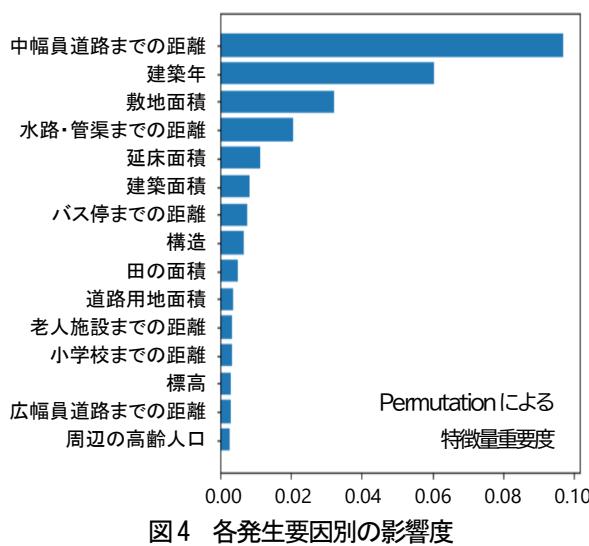
(2021)²³の中で詳しく述べられている。これまでに述べたように機械学習の予測は複雑なプロセスを経るため、説明には解釈手法が必要である。ここでは、各特徴量のモデルの予測に対する影響度の大小を比較するマクロな分析だけでなく、各特徴量と予測値の平均的な関係の把握や、ある特定サンプルの予測値に対する解釈を含むミクロな分析までを一貫して行うことで、より地域の実情に則した分析が可能になると考えられる。

4-2 並び替え特徴量重要度を用いた分析

本節では、各特徴量のモデルの予測に対する影響度を比較するために、特徴量重要度 Permutation Feature Importance (以下 PFI) を用いることとした。PFI 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²⁴である^{補注3}。本研究では、汎化誤差が小さく、精度の高い PFI を算出するために、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。

各特徴量の PFI の算出結果を図 4 に示す。

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。建物固有の特性が空き家発生に影響を与えることは、織田ら³の研究など既に多くの研究で述べられている。周辺のインフラの整備状況が与える影響については、金ら²⁴の研究で住宅へのアクセスの容易さが空き家発生に影響を与えることが述べられており、本研究の対象地においても



自家用車でのアクセス性が影響を与えると考えられる。また、建築基準法上の道路区分が PFI 上位の変数として抽出されなかった理由としては、道路に面していない建物の情報が欠損する点や、道路に面した敷地の建物と旗竿状に面した敷地の建物の区別がつかないなど、「接続性」等は考慮できない点が挙げられる。

次に、モデルの予測に対する各特徴量の発生要因カテゴリ別の影響度を比較するために、特徴量群重要度 Grouped Permutation Feature Importance (以下 GPFI) の指標を用いた。GPFI は、PFI の考え方を特徴量群に拡張したものであり²⁵、各要因のカテゴリ間の影響度の比較が可能である。発生要因のカテゴリ別 GPFI の算出結果を図 5 に示す。空き家発生に対して最も影響が大きい要因は、道路幅員や接道条件等が含まれる道路要因で、建築年等が含まれる建物要因よりも大きく、また、周辺施設要因や土地利用要因の影響は見られたが、相対的に人口要因や地形要因の影響は小さい。これは、狭窄な道路周辺に空き家が多くみられる対象地の特性²⁶を反映していると考えられ、これらの結果から、対象地において空き家対策を検討する際には、建物への対策だけでなく、特に道路との一体的な対策が必要であり、便益施設の適正な配置や、周辺に必要なインフラをバランスよく配置するなどの土地利用の改善も併せて検討する必要性があると考えられる。

4-3 Partial Dependence Plot を用いた分析

本節では、各発生要因と空き家予測値の関係の把握を行うために Partial Dependence Plot²⁷ (以下 PDP) を用いた^{補注4}。

図 6 に、PFI 上位 15 個の特徴量について、対象地の全ての建物サンプルに対して PDP を求めた結果を示す。

上位 1 番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。また、上位 2 番目の建築年は、1990 年より前になると傾きが大きくなり、1970 年以前で高い予測値を示している。これは、戦後から高度成長期にかけて建設された住宅で特に空き化が進行していることを捉えていると考えられる。また、上位 5 番目の延床面積からは延床面積が比較的小さく、上位 6 番目の建築面積からは建築面積が比較的大きい建物で予測値が高い。上位 7 番目のバス停までの距離と、上位 11 番目の老人福祉施設までの距離をみると、距離が小さいほど予測値が高くなる傾向がみられる。一般的に、このような地域交通拠点や福祉施設までの距離は、距離が大きくなるほど居住者の生活利便性に悪影響を及ぼす可能性が考えられるが、公共交通を利用する人口の割合が非常に低い対象地²⁸においては、その影響が小さい可能性が考えられる。

4-4 SHAP を用いた分析

本節では、SHAP を用いた予測結果の解釈を行う。前節の手法では全体のサンプルに対する傾向は把握可能だが、

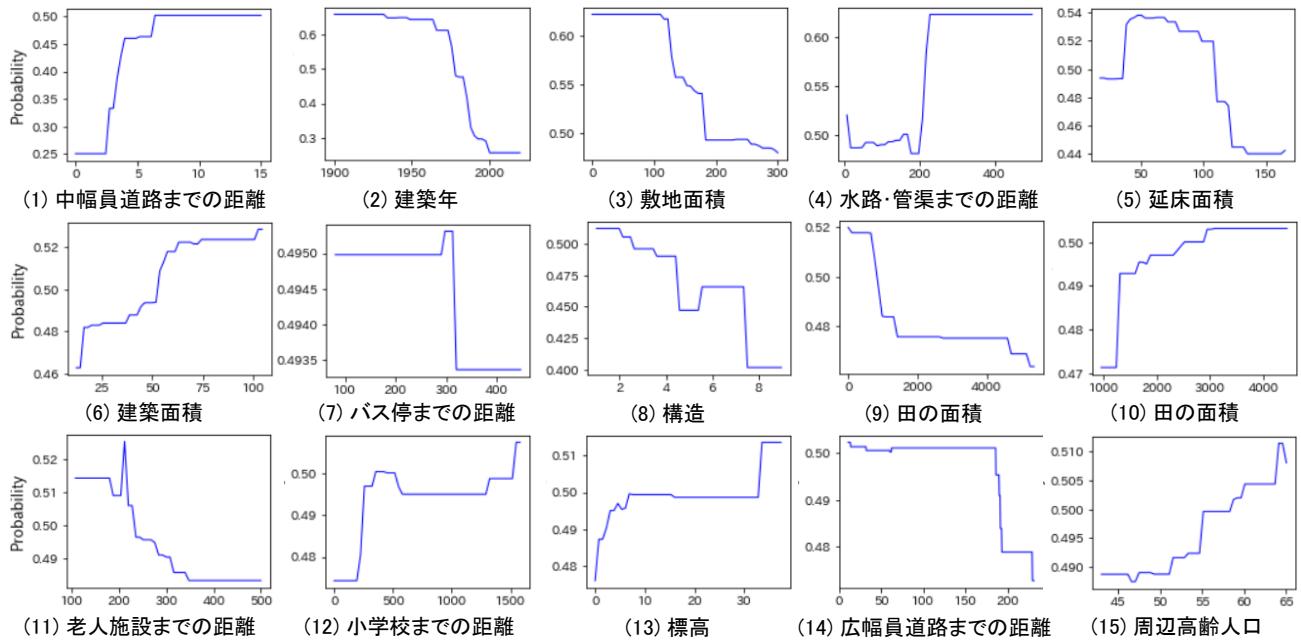


図6 PFI上位15番目までのPDP

特定のサンプル群に対する傾向は把握できない。そこで本節では、ローカルな発生要因を可視化する手法としてSHAP²⁹を導入する。SHAP²⁹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^{補注7}。SHAPを導入する大きなメリットとして、SHAP値テーブル(図7)によって任意のサンプル群に対して、各特微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微量を抽出することが可能となる。

図8に各地区別の発生要因の差異を示す。地区単位で集計した理由としては、建築物や道路等の形態を制限し、空き家対策を行う上で一つの単位となり得ると考えられるためであり、ここでは各地区の特性を各発生要因カテゴリの寄与度を比較することで明らかにする。縦軸は各カテゴリの寄与度をSHAP値で示しており、発生確率に対応している。0(平均値)を基準として発生確率を増加させる方向に寄与するカテゴリは上に、反対に発生確率を減少させる方向に寄与するカテゴリは下に棒グラフを積み上げて可視化した。また、折れ線グラフは各カテゴリの寄与度の合計を示しており、これは各地区における平均の発生確率に対応する。例えば、ある地区において建物要因カテゴリがプラスである場合は、その地区における建物の建築年等が相対的に古く、空き家が発生しやすい状況である可能性が高い。

中央地域(松崎地区、佐波地区、勝間地区、華浦地区)では、建物要因や周辺施設要因、土地利用要因が発生確率を増加させる方向に寄与しており、人口要因は発生確率を減少させる方向に寄与している。また、市街地縁辺部、市街化調整区域(向島地区、大道地区、富海地区、西浦地区)では建物要因、人口要因が発生確率を増加させる方向に寄

予測値	平均値	各特微量(55列)									
		松崎地区					牟礼地区				
全建物	0.9	=	+Σ	+0.1	+0.2	-0.1	...	+0.3	...	+0.1	+0.2
	0.1	=	+Σ	-0.1	+0.2	-0.1	...	-0.1	...	+0.2	+0.2
	0.2	=	+Σ	-0.1	+0.3	+0.1	...	+0.2	...	-0.1	+0.1
	...	=
	0.5	=	+Σ	+0.1	+0.8	-0.2	...	0.2	...	+2.0	-1.1
	...	=
	0.6	=	+Σ	+0.1	+0.2	-0.1	...	-0.1	...	+0.2	+0.2
	0.7	=	+Σ	-0.1	+0.3	+0.1	...	+0.2	...	-0.1	+0.1

図7 SHAP値テーブルの集計方法

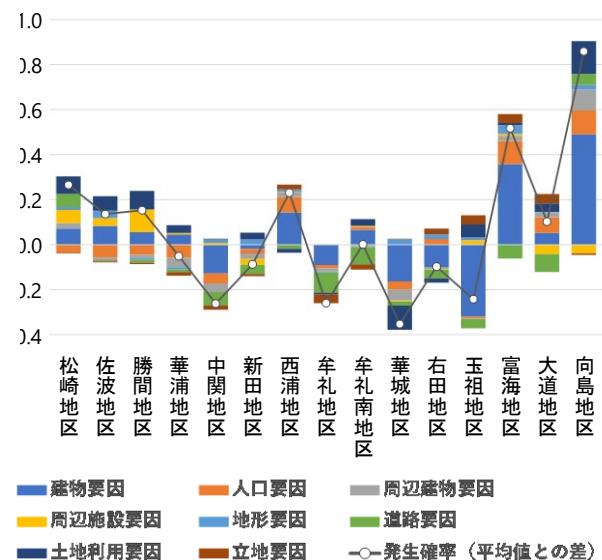


図8 地区別の発生要因分解

与している。また、低層住宅が多い華城地区や、沿道サービス施設等が立地する玉祖地区では建物要因が発生確率を減少させる方向に寄与している。このように、各地区の特

性が空き家発生確率の大小に影響を与えていたと考えられ、特に発生確率を増加させる方向に寄与している要因カテゴリは各地区の課題を示しているといえる。

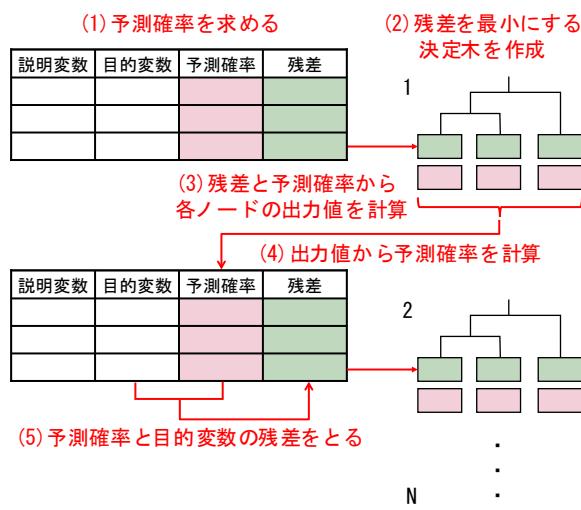
5. 総括

本研究では、今後のまちづくりの中での面的な空き家対策の必要性を踏まえ、機械学習の解釈手法を用いて、建物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った。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、各地区的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。

今後の展望として、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が考えられる。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。

【補注】

- 1) 空き家の定義は、空き家特措法の定義に従う。共同住宅は、全ての賃室が空室となったものを指す。
- 2) XGBoost を用いた分類は、まず(1)予測確率（初期確率は 0.5）を求め、(2)類似度スコア補注 1)を用いて決定木を構築する。(3)次に目的変数と予測確率の残差から各ノードの出力値を求め、(4)各ノードの出力値から予測確率を計算する。(5)予測確率と実測値（目的変数）との残差を計算し、(6)得られた残差を目的変数にとり、類似度スコアを用いて決定木を構築する(2)の手順に戻る。(2)～(5)の手順を繰り返すことで残差を最小化するように逐次的に学習が進む。



- 3) 多数派群を減らす手法をアンダーサンプリング、少数派群を増やす手法をオーバーサンプリングという。本研究では、空き家と空き家以外のサンプル比率が 1:1 になるようにリサンプリングしたデータセットに対して学習を行った。
- 4) 予め指定したパラメータグリッドの中から最適なパラメータセットを総当たりで探索する方法である。
- 5) 仮にモデルがある特徴量を重視して予測を行ったと仮定すると、その特徴量の情報が使えない状態での予測誤差は、使える場合と比較して大きくなると考えられる。PFI では、この「特徴量の情報が使えない」という状態を特徴量の値をシャッフルすることで表現している。
- 6) 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のか」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。
- 7) SHAP では、複雑なモデルを解釈するために線形の関数である説明モデルを導入している。各サンプルの出力値とベースラインの出力値の差は、単純化された各サンプルに対応する効果の線形結合によって表せる。この各サンプルの入力に対応する各特徴量の効果を表す統一的な尺度として SHAP 値を導入しており、協力ゲーム理論の Shapley 値を応用することで求めることができる。

【参考文献】

- 1) 総務省統計局：平成30年住宅・土地統計調査, 令和4年1月 15日 13:00 (https://www.stat.go.jp/data/jyutaku/2018/pdf/kihon_gaiyou.pdf)
- 2) NPO法人空家・空地管理センター：放置空き家がもたらす被害について, 令和4年1月 15日 13:00 (<https://www.akiya-akichi.or.jp/what/damage/>)
- 3) 織田峻央, 森本章倫, 浅野周平：地域特性及び物件属性に着目した空き家の発生状況に関する研究, 都市計画論文集, Vol.53, No.3, pp.1074-1079, 2018.10,
- 4) 空き家対策防府モデル事業, 令和4年8月 1日 13:00 (<https://www.cityhofu.yamaguchi.jp/site/akiya/entryou.html>)
- 5) 山下伸, 森本章倫：地方中核都市における空き家の発生パターンに関する研究, 都市計画論文集, Vol.50, No.3, pp.932-937, 2015.10
- 6) 西浦定継, 小林利夫：地域要因からみる空き家発生リスクの試算に関する研究—東京都日野市の空き家調査データを事例に—, 日本建築学会計画系論文集, Vol.82, No.740, pp.2629-2635, 2017.10
- 7) 馬場弘樹, 秋山祐樹, 谷内田修：自治体保有データを活用した空き家の空間分布の将来予測モデル構築—群馬県前橋市を対象として—, 土木学会論文集 D3(土木計画学), Vol.77, No.2, pp.62-71, 2021.
- 8) 秋山祐樹, 馬場弘樹, 大野佳哉, 高岡英生：機械学習による空き家分布把握手法の更なる高度化—自治体の公共データを活用した空き家の分布把握手法に関する研究(その3), 日本建築学会計画系論文集, Vol.86, No.786, pp.2136-2146, 2021.8
- 9) 北島紗恵, 六本木延浩, 富岡秀虎, 森本章倫：画像データを用いた深層学習による空き家の発生予測に関する研究, 都市計画論文集, Vol.56, No.3, pp.1468-1474, 2019.10
- 10) 第2次防府市空家等対策計画, pp.3, 2021.3
- 11) 防府市都市計画マスターplan, pp.9, 2018.4
- 12) 大澤陽樹, 横張真, 雨宮護：都市郊外の住居系用途地域における

- ける空閑地の発生・残存パターンと地形との関係, ランドスケープ研究 Vol.72, No.5, pp.683-686, 2009.
- 13) 阿部正太朗, 中川大, 松中亮治, 大庭哲治: 地方都市中心部の低未利用地における面積変化と居住用地への転換に関する要因分析, 土木学会論文集 D3 (土木計画学), Vol.68, No.5, I_467-I_477, 2012.
- 14) 中井翔太, 嘉名光市, 佐久間康富: 密集市街地における空き家の実態とその「防災空間」としての活用可能性に関する研究: 大阪市鶴橋地区を対象として, 都市計画論文集, Vol.47, No.3, pp.1063-1068, 2012.10
- 15) 金森有子, 有賀敏典, 松橋啓介: 空き家率の要因分析と将来推計, 都市計画論文集, Vol.50, No.3, pp.1017-1024, 2015.10
- 16) 金ドン均, 有馬隆文, 坂井猛: 歩行消費エネルギーから見た斜面市街地における空き家・空き地の発生要因に関する研究, 日本建築学会計画系論文集, Vol.81, No.726, pp.1715-1722, 2016.8
- 17) 坂本慧介, 横張真: 地方中核都市における空き家・空閑地の発生動態—栃木県宇都宮市の中心市街地周辺の住宅地を対象に—, 都市計画論文集, Vol.51, No.3, pp.854-859, 2016.10
- 18) 熊谷樹一郎, 植松恒, 小野裕基, 山本純平: 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), Vol.74, No.2, I_51-I_58, 2018.
- 19) 馬場弘樹, 橋野公宏: 空き家の管理不全要因とその傾向—川口市空き家実態調査の分析—, 日本建築学会計画系論文集, Vol.83, No.749, pp.1263-1271, 2018.
- 20) Chen, T. and Guestrin, C.: Xgboost: A Scalable Tree Boosting System, Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.785-794, 2016.3.
- 21) Chawla, N.V.: Data mining for imbalanced datasets, An overview, Data Mining and Knowledge Discovery Handbook, pp.853-867, Springer, 2005.
- 22) 森下光之助: 機械学習を解釈する技術—予測力と説明力を両立する実践テクニック—, 技術評論社, 2021.
- 23) Breiman, L.: Random forests, Machine Learning, 45, pp.5-32, 2001.
- 24) 金ドン均, 有馬隆文, 坂井猛: 歩行消費エネルギーから見た斜面市街地における空き家・空き地の発生要因に関する研究, 日本建築学会計画系論文集, Vol.81, No.726, pp.1715-1722, 2018.8
- 25) Gregorutti B, Michel B, Saint-Pierre P: Grouped variable importance with random forests and application to multiple functional data analysis, Computational Statistics & Data Analysis, Vol.90, pp.15-35, 2015.10
- 26) 第2次防府市空き家等対策計画, pp.6, 2021.3
- 27) Jerome H. Friedman: Greedy Function Approximation: A Gradient Boosting Machine, Annals of Statistics, Vol.29, No.5, 2001.10
- 28) 防府市地域公共交通網形成計画, pp.11, 2018.3
- 29) Lundberg, Scott M., and Su-In Lee: A Unified Approach to Interpreting Model Predictions, Advances in Neural Information Processing Systems, 2017.