

令和3年度
和歌山県における
空き家分布推定に関する研究成果報告書

令和4年3月

東京大学空間情報科学研究センター	特任准教授	秋山 祐樹
東京大学空間情報科学研究センター	研究支援員	左右田敢太
東京大学空間情報科学研究センター	研究支援員	洪 義定
和歌山県データ利活用推進センター	副主査	徳富 智哉

目 次

1. 空き家分布推定研究の背景と目的	1
1.1. 背景と先行研究	1
1.2. 前年度の研究を踏まえた今年度の研究の位置づけ	2
1.3. 本研究の目的	3
2. 本研究で使用したデータの概要	4
2.1. 公共データの概要	4
2.2. 国勢調査の概要	6
2.3. 住宅・土地統計調査の概要	6
2.4. データの処理	6
2.5. データの結合結果	7
3. 空き家分布推定モデル	9
3.1. XGBoost	9
4. 空き家検出モデルの外挿	11
4.1. 学習データ	11
4.2. モデルの構築	13
4.3. モデルの外挿	15
5. 空き家率回帰モデルの外挿	18
5.1. 学習データ	18
5.2. モデルの構築	18
5.3. モデルの外挿	19
6. 住宅・土地統計調査を統合した国勢調査モデル	20
6.1. 学習データ	20
6.2. モデルの分類	21
6.3. 各モデルの推定結果	22
7. まとめと今後の検討事項	26
7.1. 今年度の成果のまとめ	26
7.2. 未来予測モデルの検討	26
7.3. マニュアル化の検討	26
7.4. 総括	27
謝辞	28
参考文献	29
付表	30
付録	36

1. 空き家分布推定研究の背景と目的

1.1. 背景と先行研究

近年、日本では人口減少や高齢化、都市部への人口移動などを背景に、全国で空き家が増加している。総務省統計局「住宅・土地統計調査」によると、平成30年の日本全国の空き家数は約846万戸、空き家率は13.6%に達しており、空き家数・空き家率ともに過去の調査から比較しても、一貫して増加が続いている状況にある。なかでも、「その他の住宅」（別荘などの一時的に利用実態がある住宅や、賃貸用・売却用の住宅以外の住宅）の増加は著しく、平成20年調査から平成30年調査までの10年間に約268万戸から約347万戸へと約1.3倍に増加している。一部の管理不十分なその他の住宅は、腐朽・破損による倒壊危険性を有するだけでなく、地域の防犯性の低下や景観の悪化にもつながる。このような空き家は「特定空き家」と呼ばれ、近隣住民や地域全体に深刻な影響を及ぼす可能性が高いことから、特定空き家を含む空き家の実態把握はわが国にとって急務となりつつある（浅見, 2014; 高崎経済大学地域科学研究所, 2019）。

こうした背景を受け、平成27年5月から「空家等対策の推進に関する特別措置法（空家等対策特措法）」が全面施行された。同法の施行により、自治体は空き家の所有者への適切な管理の指導や、空き家跡地の活用促進、特定空き家に対する助言・指導・勧告・命令、さらには罰金・行政代執行も可能となり、空き家の活用・除却といった行動を法的根拠に基づいて実施することが可能になった（高崎経済大学地域科学研究所, 2019）。しかしこれらの行動を起こすためには、まずは自治体のどこにどの程度空き家が分布しているのか、という情報を把握する必要がある。しかし空き家の空間的分布を把握する手法は、現状では一棟一棟を個別に訪問し外観を見て判断する戸別目視が中心となっている。また、現地調査を実施する前に、空き家が数多く分布すると考えられる地域を予め把握する手法も確立されていないため、広域の空き家分布を継続的に把握し続けるには多大な労力と時間、そして費用を要する。これが自治体において空き家対策の取り組みを進めていく上で大きな障壁となっている。すなわち、これらの調査を「迅速」、「安価」かつ「継続的」に実施可能な手法の確立が期待されている。

広域を対象とした空き家の分布状況の把握を試みた研究としては、西山（2015）やYamashita et al. (2015) による水道閉栓情報を用いた市域全域の把握の例がある。しかし、これらの手法では水道が閉栓か休止中の建物を全て空き家と定義しており、その根拠が明らかではない点に課題がある。また、秋山ほか（2018）により、水道閉栓の有無のみを使って空き家を特定することは困難であることが指摘されている。この問題を解決するアプローチとして、自治体が所有するデータである住民基本台帳や、建物登記情報、水道使用量の情報などを活用して空き家を特定する方法もある（Akiyama et al., 2020, Baba et al., 2020）。特に、Akiyama et al. (2020) では、鹿児島県鹿児島市と福岡県朝倉市を対象に、上記3つのデータや地図情報等を説明変数とするクロス集計表を使用した分析モデルを提案した。このような公共データを活用した空き家分布推定の手法は、個別目視に頼る従来の手法に比べて、調査費用と調査時間を削減することができるだけでなく、広域を対象に迅速かつ安価な調査を、定期的かつ継続的に実施することができるという長所を持っている。

一方、和歌山市においても今後空き家が増え続けていくことが懸念されている。実際に、平成25年の住宅・土地統計調査によると、和歌山市の空き家率は15.8%（全国平均は13.5%）であったが、平成30年の調査によると和歌山市の空き家率は18.9%（全国平均は13.6%）となっており、全国平均と比べてもその値、また値の増加のペースも高い水準にある。こうした背景の中、和歌山市は空き家対策の取り組みを進めており、平成29年3月には「和歌山市空家等対策計画」を策定し、空き家の空間分布の把握を目的に市内全域を対象とした「和歌山市空家実態調査」を平成29年度に完了させた。和歌山市空家実態調査は現地調査による目視判読で空き家か否かを確認しており、比較的信頼性の高い調査であることが期待される。したがって、Akiyama et al. (2020)に見られるような自治体が保有する各種公共データを活用して、和歌山市全域の空き家分布推定モデルを構築するとともに、その推定精度を和歌山市空家実態調査の空き家データを使って検証することで、和歌山市で有用な分析モデルの選定や、より推定精度の高い分析モデルを開発できるものと期待される。その結果、和歌山市において今後は本研究で開発した手法を採用することで迅速かつ安価な空き家分布調査を、定期的かつ継続的に実施し、空き家対策の取り組みの効果的な推進とその支援を行うことが可能となるものと期待される。さらに、総務省統計局統計データ利活用センターや和歌山県データ利活用推進センターと協働することで、様々な公的統計のマイクロデータ（国勢調査の個票データ等）が利用可能となるため、統計マイクロデータと和歌山市が持つ公共データを融合させることで、より推定精度の高い分析モデルの開発の実現が期待できる。

1.2. 前年度の研究を踏まえた今年度の研究の位置づけ

前年度の研究では、和歌山市が保有する公共データ、すなわち住民基本台帳、建物登記情報、水道使用量情報や、公的統計データである国勢調査を説明変数、同市が実施した空家調査データを目的変数とするデータベースを整備し、機械学習モデルを構築することにより、再現率が6割、適合率が9割を超えるモデルの構築が実現した。一方で、以上の分析を可能にするためのデータ整備やモデルの構築には、それなりの労力と技術が求められることも分かった。またそもそも、迅速な空き家分布予測のニーズは、空き家調査データを整備していない自治体にこそあるのであると言えるだろう。そのため、前年度までの研究は空き家予測モデルへの理解を深め、効果の保証に大きく貢献したといえる一方、和歌山市以外への自治体、さらには広く全国を対象に普及させる手法としては課題が残る結果となった。

そこで今年度の研究では、以上の課題を受けて、以下2つの点について検証を行う。第1に、モデルの外挿性の検証、すなわちある地域で学習したモデルを異なる地域に適用することがどの程度可能であるかということをも明らかにすることを試みる。この際、公共データを用いたモデルによる建物毎の予測と同様に、国勢調査のみを用いたモデルによる基本単位区毎の予測も行う。モデルの外挿性を検証することは、空き家調査データを持たない地域での空き家予測に対して重要な知見を提供することが期待される。

第2に、公的統計を用いた空き家予測モデルを改良するために、国勢調査と住宅・土地統計調査を組み合わせたモデルの予測精度について検証する。住宅・土地統計は国勢調査と異なり、全体の1割程度の調査区を抽出して行われる、統計的調査である。従って全ての調査区でデータが整備されているわけではない。しかし、住宅や土地の詳細な情報を含んでいるため、空き家予測には大きく貢献するものと思われる。住宅・土地統計調査の有用性につい

て検証することで、統計的な調査が母集団の予測精度に与える影響の理解を深めることができるものと期待される。

1.3. 本研究の目的

以上より、本研究の目的は前年度までに構築したモデルの外挿性を検証するとともに、公的統計を用いた空き家予測モデルの予測精度を改良することである。具体的には、以下3つの分析を行う。

1. 和歌山市の公共データを学習したモデルを用いて、田辺市の建物毎の空き家予測を行い、外挿性を評価する。
2. 和歌山市の国勢調査を学習したモデルを用いて、田辺市の基本単位区毎の空き家率予測を行い、外挿性を評価する。
3. 国勢調査に住宅・土地統計調査を加えたモデルを用いて、調査区毎の空き家率の予測精度を評価する

2. 本研究で使用したデータの概要

本研究を実施するためには、まず和歌山市から提供された各種公共データや、政府統計（国勢調査、住宅・土地統計調査）の変数の概要やデータのレイアウト構造を把握する必要がある。そこで、本章では空き家分布推定の際に利用した、公共データと政府統計の変数および構造について概説する。なお、本章は前年度報告書の一部を再掲および再校正した内容となっている。より詳しい情報が必要な場合は昨年度の成果報告書も参考にされたい。

2.1. 公共データの概要

まず、公共データの概要を説明する。和歌山市から提供された公共データは、住民基本台帳（以下、「住基データ」）、建物登記情報（以下、「建物登記データ」）、水道利用量情報（以下、「水道データ」）、和歌山市空き家実態調査（以下、「空き家データ」）である。いずれのデータファイルも CSV形式（コンマ区切りのテキスト形式）となっている。なお、今年度は田辺市においても公共データを使用しているが、使用した情報は同様の内容となっているため、ここでは和歌山市の例のみを掲載する。

1) 住基データ（2019年4月現在のデータ）

住基データは住所がキー変数となる。すなわち、当該データは住所によって他のデータと結び付けられる。一つの住所には、同じ住所を持つ住民が複数存在する場合や、住所に複数の世帯番号が対応している場合がある。そのため、実際の分析の際には世帯ごとに集計し直し、各データに対してユニークな住所を与えている。その際、世帯人数等は合計に直すなどのデータ整形を行う必要がある。住基データの変数一覧は表1の通りである。

表1 本研究で使用した住基データの変数一覧
（(C) はカテゴリデータ、(V) は量的データ。以下同様。）

変数名	仕様等
住所 (C)	「〇〇番地△△号」と記載
年齢 (V)	居住者の年齢（歳）
世帯識別番号 (V)	7桁の世帯ごとに与えられた固有の番号
住定日 (V)	1桁目は元号（1：明治、2：大正、3：昭和、4：平成）、2～3桁目は年、4～5桁目は月、6～7桁目は日を表す。例えば「4040706」は、平成04年07月06日を表す。

2) 建物登記データ（2018年10月現在のデータ）

建物登記データは特殊な形になっており、取り扱いに注意が必要である（表2）。1列目の番号は建物ごとに割り振られた建物番号（項番）である。表2の1行目を見ると、項番、物件情報、物件種別、物件状態…と項目が並んでいる。2行目以降も同じ規則が成立するが、4行

目は主である建物情報の項目が並んでおり、最も重要な部分である。なお各キー変数の5行目において、付属建物情報がついている箇所が見られるが、物置等の居住者が存在しない建物と考えられるため無視する。

3) 水道データ (2019年5月現在のデータ)

水道データは住所をキー変数としたリレーショナルな形式になっているため、取り扱いが容易である。水道データの変数一覧は表3の通りである。

表2 建物登記データのレイアウトと変数一覧 (前年度報告書再掲)

	行	1列	2列	3列	4列	5列	6列	7列	8列	9列
物件情報	1行	項番	物件情報	物件種別	物件状態	地番区域	地番家屋番号	不動産番号		
一般建物表題部登記事項	2行	項番	所在	所在(実際の所在地)				原因及びその日付	登記の日付	その他
	3行	項番	家屋番号						登記の日付	その他
	4行	項番	主である建物の表示		種類	構造	床面積	原因及びその日付	登記の日付	その他
	5行	項番	付属の建物の表示	符号	種類	構造	床面積	原因及びその日付	登記の日付	その他

表3 本研究で使用了水道データの変数一覧

変数名	仕様等
住所 (C)	「〇〇番地△△号」と記載
閉栓区分 (C)	開栓、閉栓の2値データ
開栓日 (V)	1桁目は元号(1:明治、2:大正、3:昭和、4:平成)、2~3桁目は年、4~5桁目は月、6~7桁目は日を表す。例えば「4040706」は、平成04年07月06日を表す。
閉栓日 (V)	閉栓している場合は、開栓年月日を7桁の数字で表す。開栓中の場合は0で表す。
月ごとの水道使用量 (V)	単位はm ³ 。変数名は、429-1のような形となっている。429-1は平成29年1月を表す。

4) 空き家データ (2016~2017年に現地調査実施)

空き家データは、水道閉栓情報、平成24年和歌山県廃墟建築物調査、和歌山市危険家屋台帳（苦情をピックアップした台帳）を基に空き家候補を割り出し、その候補に対して現地調査（外観目視による判定）を行うことにより、空き家を特定したものである。そのため、上記のデータで空き家候補とならなかったものは調査対象となっていないため、過小推計になっている可能性があることに注意が必要である。空き家データはシェープファイル形式であるため、建物の位置を表す座標情報が予め付与されている。空き家実態調査データの中で必要な変数は表4の通りである。

表4 本研究で使用した空き家データの変数一覧

変数	仕様等
位置座標 (V)	建物重心の経度緯度 (世界測地系)
空き家判定 (C)	和歌山市空き家実態調査で空き家と判定されたものは1、空き家でないものは0とする2値の変数

2.2. 国勢調査の概要

本研究で使用した国勢調査（以下、「国調」）は、e-Stat等で公開されているオープンデータではなく、基本単位区単位で集計が行われているもの（統計マイクロデータ）を使用した。なお、本研究で用いる国勢調査の特徴量は数が非常に多いため、**附表1**を参照されたい。

2.3 住宅・土地統計調査の概要

本研究で使用した住宅・土地統計調査（以下、「住調」）は、国調と同じくe-Stat等で公開されているオープンデータではなく、調査区単位で集計が行われているもの（統計マイクロデータ）を使用した。国勢調査と同じく特徴量は数が非常に多いため本研究で使用した変数の一覧は**附表2**を参照されたい。なお、国調は全ての基本単位区を調査対象としている一方、住調はサンプル調査であり、全調査区の概ね10分の1を調査対象としている点に注意が必要である。

2.4. データの結合処理

後述する機械学習モデルを構築するためには、以上に挙げた2.1の公共データ（住基データ、水道データ、建物登記データ、空き家データの4種類）、2.2の国調、2.3の住調の6種類のデータを位置情報に基づいて結合処理を行う必要がある。そこで、**図1**に示すようにこれらの情報を位置情報に基づいて、以上のデータの作成時期に近い2016年版のデジタル住宅地図（株式会社ゼンリン）の建物ポリゴンデータに空間結合処理を行った。また、同時に後述する機械学習モデルに投入するために変数の加工を行った。同処理の詳細については令和2年度報告書を参照されたい。

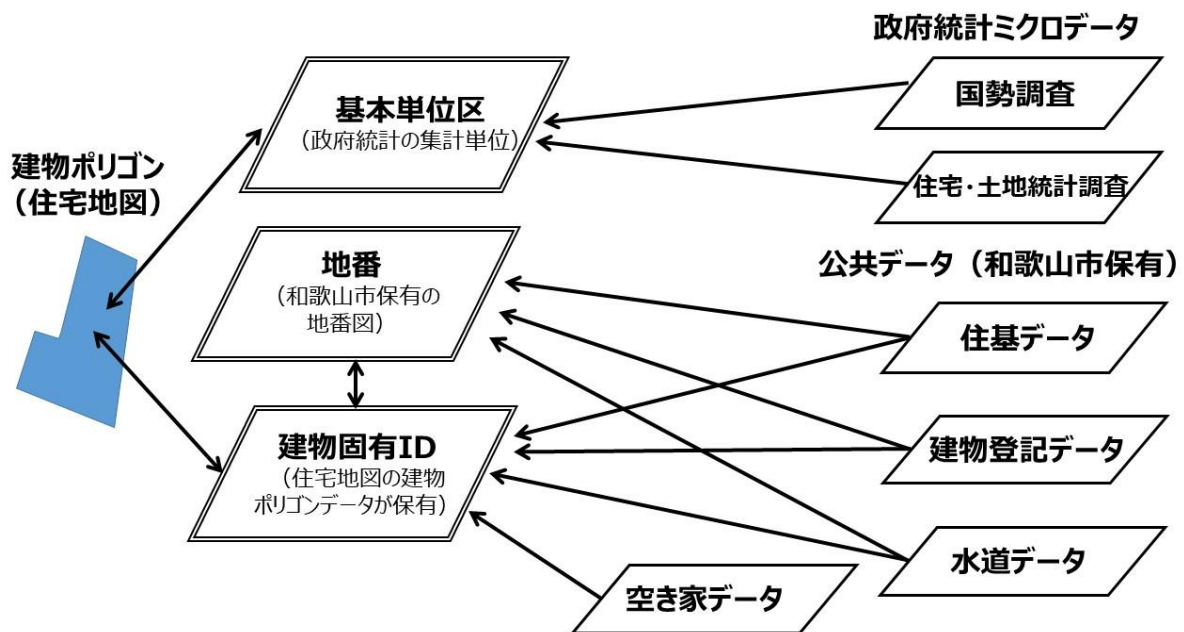


図1 公共データ、国調データ、住調データを空間結合処理する流れ（和歌山市の場合）

なお、住基データと水道データは位置情報を住所（住居表示）というカテゴリデータとして保有している上に表記揺れを含むため、全ての住所を経度緯度という量的データに変換することで、相互に空間結合を行った。このように住所や地名を経度緯度という地理座標に変換することを「ジオコーディング（あるいはアドレスマッチング）」という。本研究ではジオコーディングの中でも研究目的として利用可能な東京大学空間情報科学研究センターが提供している「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を付与した。なお、住所が適切に記述・設定されていない場合や、CSVアドレスマッチングサービスの住所の参照元となる住宅地図が、新規開発などに伴う住所の変更・新設を捕捉しきれていない場合があるため、一定程度の誤差が生じてしまう点には注意が必要である。また、建物登記データの住所は住居表示ではなく地番となっているため、地番図の住所と対応させて地番ポリゴンの重心座標を緯度経度座標として取得することで、経度緯度情報を得た。

2.5. データの結合結果

以上の手法により、公共データ、国調データおよび住調データを住宅地図の建物ポリゴンに結合させた。ここでは以下の2点についてその結合結果について整理した。まず「①ジオコーディングの結果」であり、もう一点は「②建物へのデータ集約の結果」についてである。本報告書では和歌山市の結果について紹介する。

1) ジオコーディングの結果

前述の通り住居表示の住所を持つ住基データと水道データは、CSVアドレスマッチングを用いて住所から緯度経度座標を付与し、地番表示である建物登記データは地番図の住所と対応させることで経度緯度座標を付与した。CSVアドレスマッチングは緯度経度座標の特定した精度（例えば号レベルの精度や丁目・大字レベルの精度など）を出力させることができる。

本研究では住宅地図の建物ポリゴンと空間結合させる必要があるため、建物を完全に特定可能な精度である「号レベル」で位置情報が特定されたデータのみを利用した。表5に和歌山市におけるジオコーディングの結果を示す。住基データは173,769件のうち、141,719件（81.6%）に、建物登記データは142,401件のうち、96,451件（67.7%）に、水道使用量データは160,230件のうち、122,035件（76.2%）に号レベルの緯度経度座標を付与させることができ、後述のモデル構築において十分なデータが得られた。一方、空き家実態調査はそもそも緯度経度の情報を得ているため、5,510件（100%）が緯度経度付きデータとして利用可能である。

2) 建物へのデータ集約の結果

続いて1)の各データおよび、国調が建物データにどの程度結合したかについて集計した結果を表6に示す。結合対象となる建物88,363件に対して、公共データである住基データは67,616件（76.5%）、水道データは69,915件（79.1%）、建物登記データは68,330件（77.3%）、空き家データは4,494件（5.1%）に少なくとも1件以上が結合した。一方、国調は全調査区をカバーしているため88,382件（99.9%）であった。後述する公共データを用いたモデル構築では、公共データのうち何れか1種類が結合していれば空き家確率の推定可能な建物となるが、それに該当する建物は86,060件（97.4%）であることから、和歌山市内の多くの戸建住宅で空き家確率の推定が可能であることが分かる。なお、住調は全調査区の約10分の1のみが調査対象であり、また後述する建物単位の推定には使用していないことから、建物単位でのデータ集約件数を集計する意味は無いため、ここではその結果を割愛する。

表5 和歌山市におけるジオコーディングの結果

データの種類	元データの件数	ジオコーディング成功件数
住基データ	173,769	141,719
水道データ	160,230	122,035
建物登記データ	142,401	96,451
空き家データ	5,510	5,510

表6 和歌山市において1件でも自治体保有データおよび国調が結合する建物数

	自治体保有データ				政府統計
	住基	水道	建物登記	空き家	国調
1件でも結合した建物数	67,616	69,915	68,330	4,494	88,382
結合率[%]	76.5	79.1	77.3	5.1	99.9

3. 空き家分布推定モデル

本章では、本研究における空き家分布推定モデルの基本的な考え方と構築手法を述べる。なお、本章は前年度報告書の一部を再掲および再校正した内容となっている。より詳しい情報が必要な場合は昨年度の成果報告書も合わせて参考にされたい。

3.1. XGBoost

モデルの構築には、Chen and Guestrin (2016)によって提案された機械学習的分類手法のひとつであるXGBoost (eXtreme Gradient Boosting)を用いた。これは、後述する決定木を基本として、それをいくつも作成し、各決定木から得られた値を足し合わせることで最終的な予測値を得る手法である。本節では概説に留めるが、詳しくは付録を参照されたい。

本手法は推定精度が一般的に高く出る点において他手法よりも有利である。例えば、回帰分析は空き家の多寡が延床面積などの変数に対して単調増加（または減少）することを仮定しているが、本研究で扱う現象の場合、それは必ずしも当てはまらない。本手法は変数の変化に応じて柔軟に推定空き家確率を予測できるため、そのような問題は生じない。さらに、本研究は複数種類のデータを統合しており、必然的に欠損値の多いデータとなっている。XGBoostは欠損値を扱うことができるが、他の多くの手法では扱うことができない。以上のような利点を有するものの、当該手法は複雑な事象の場合、上手く予測値を得られない可能性がある。ただしこの点は予測モデルに対して検証データを投入し、予測値の適合度合いから問題の有無を判断することができる。

決定木とは、ある住宅が空き家であるかどうかを予測する際、いくつかの条件を設定して住宅を分岐させた構造をいう。例えば、条件に世帯内最高齢者の年齢、築年数、年間水道使用量を設定すると、**図2**のようなイメージとなる。ここでは、はじめに世帯内最高齢者の年齢で分岐させ、続いて築年数と年間水道使用量で分岐させている。予測用のデータをこの決定木に当てはめて、木の終端の数値がその建物の空き家確率となる。この空き家確率は、木の終端に空き家に該当するデータが多く含まれるほど高い値をとる。例えば、世帯内最高齢者の年齢が80歳、築年数30年、年間水道使用量100 m^3 の住宅を得た場合、**図2**の木構造から、空き家確率0.23を得る。

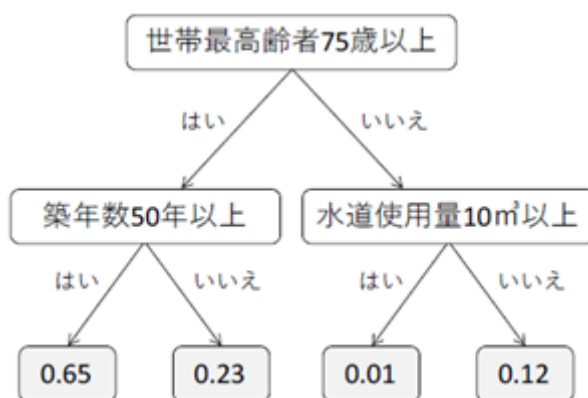


図2 決定木のイメージ

空き家の推定確率を算出する際、上記の決定木を逐次作成していき、各決定木の結果を足し合わせることで予測値を算出する。1番目の決定木では、現地調査空き家のデータに基づいて木の構造を作成する。その際に算出される推定空き家確率は、2番目の決定木の作成に利用される。予測値を改善するため、2番目の決定木では、1番目で生じた誤差を評価してその決定木の重み（重要度）を定める。このような操作を予測値が改善しなくなるまで繰り返す。

以上のような流れは図3のように表せる。モデルの複雑さに依存するが、最終的に何千本もの決定木が構築され、それぞれの決定木から算出された値を足し合わせることで最終的な推定空き家確率を得ることができる。このような構造を採用することで、未知のデータにもうまく適合して予測値を返すことができるようになる。

本研究で構築するXGBoostモデルは、大きく2種類に分類される。第1に、建物毎に空き家か否かを分類する「空き家分類モデル」である。これは、住基や水道、建物登記といった自治体保有のデータを学習データとして用いている。第2に、基本単位区毎または調査区毎の空き家率を回帰する「空き家率回帰モデル」である。これは、国勢調査や住宅・土地統計調査といった、公的統計データを用いている。これら2種類のモデルは、目的変数が2値データであるか、連続値であるかという点で異なる。しかしXGBoostにおいては、両方とも決定木に基づく予測が可能であるため、本質的には同じ手法であると言って良い。なお、空き家率のデータについては、建物毎の空き家データに基づき、基本単位区毎または調査区毎に集計した空き家数を全ての建物数で除した値を用いた。

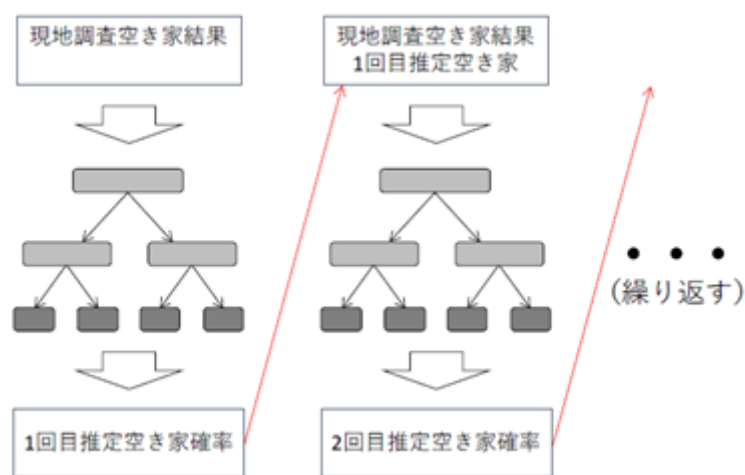


図3 分析手法のイメージ

4. 空き家検出モデルの外挿

本章では、建物毎の空き家検出モデルの外挿性を評価する。具体的には、和歌山市保有の公共データを学習した空き家検出モデルを用いて、和歌山県田辺市における建物毎の空き家予測を行う。また、外挿による予測結果を評価する基準を得るために、田辺市および橋本市保有の公共データを学習した空き家検出モデルによる両市の空き家予測（いわゆる、内挿）の精度も評価する。

4.1. 学習データ

学習に用いたデータは、住基データ、建物登記データ、水道データの3種類である。2章に述べた方法で全てのデータを建物データへ結合した。ただし、和歌山市と田辺市・橋本市では各公共データに含まれる情報が異なるため、最終的なデータは相異なる。表7、表8、表9にそれぞれの自治体におけるデータの特徴量記述統計表を示す。

表7 和歌山市の公共データの特徴量記述統計

属性	Mean	S. D.	Min	Median	Max
空き家ダミー*1	0.051	0.22	0	0	1
住居専用地域ダミー*2	0.114	0.318	0	0	1
住居地域ダミー*2	0.115	0.319	0	0	1
商業地域ダミー*2	0.027	0.163	0	0	1
工業地域ダミー*2	0.022	0.148	0	0	1
指定建蔽率*2	61.034	7.122	30	60	80
指定容積率*2	203.427	58.443	50	200	600
世帯内最高年齢	65.665	16.097	17	68	109
世帯内最低年齢	42.559	26.438	0	43	108
世帯人員	2.932	1.769	1	3	19
住基ダミー	0.765	0.424	0	1	1
住宅ダミー*2	0.973	0.163	0	1	1
木造ダミー	0.712	0.453	0	1	1
鉄骨造ダミー	0.161	0.367	0	0	1
RC造ダミー	0.038	0.192	0	0	1
築年数	28.79	14.432	0	30	94
延床面積	110.983	47.669	10.07	104.49	999
階数	1.909	0.36	1	2	6
水道使用量2018*2	217.995	345.886	0	209	55355
水道使用量2017	214.798	364.351	0	205	63373
開栓ダミー	0.95	0.218	0	1	1
閉栓期間	3.753	23.394	0	0	349.5

*1 この変数は目的変数であり、空き家確率を計算する際の説明変数ではない。

*2 田辺市の公共データに含まれない特徴量。その他は、田辺市と共通する特徴量。

表8 田辺市の公共データの特徴量記述統計

属性	Mean	S. D.	Min	Median	Max
空き家ダミー*1	0.038	0.191	0	0	1
世帯内最高年齢	70.539	15.622	21	72	107
世帯内最低年齢	54.381	26.345	3	61	107
世帯人員	2.314	1.279	1	2	10
住基ダミー	0.623	0.485	0	1	1
木造ダミー	0.914	0.281	0	1	1
鉄骨造ダミー	0.081	0.272	0	0	1
RC造ダミー	0.003	0.057	0	0	1
延床面積	93.416	46.139	1.08	90.79	835.24
階数	1.678	0.535	0	2	3
築年数	40.235	19.537	3	40	109
建物登記ダミー*2	0.828	0.377	0	1	1
開栓ダミー	0.838	0.369	0	1	1
簡易水道ダミー*2	0.181	0.385	0	0	1
閉栓期間	26.443	76.833	0	0	377
水道使用量2016*2	191.661	171.404	0	175	3609
水道使用量2017	189.577	169.200	0	174	4287
水道データダミー*2	0.804	0.397	0	1	1
旧田辺市ダミー*2	0.592	0.492	0	1	1

*1 この変数は目的変数であり、空き家確率を計算する際の説明変数ではない。

*2 和歌山市の公共データに含まれない特徴量。その他は、和歌山市と共通する特徴量。

表9 橋本市の公共データの特徴量記述統計

属性	N	Mean	S. D.	Min	Median	Max
空き家ダミー*1	33410	0.034	0.181	0	0	1
世帯内最高年齢	20013	67.289	16.152	20	70	109
世帯内最低年齢	20013	50.761	27.320	0	58	106
世帯人員	20013	2.389	1.250	1	2	10
住基ダミー	33410	0.599	0.490	0	1	1
木造ダミー	21356	0.866	0.340	0	1	1
延床面積	21356	110.906	46.627	4.95	109.72	700.6
階数	21814	1.740	0.455	1	2	3
築年数	16252	30.473	14.498	0	31	105
建物登記ダミー*2	33410	0.639	0.480	0	1	1
開栓ダミー	24663	0.588	0.492	0	1	1
閉栓期間	9779	202.342	199.291	0	163	1359
水道データダミー*2	33410	0.738	0.440	0	1	1

*1 この変数は目的変数であり、空き家確率を計算する際の説明変数ではない。

*2 和歌山市の公共データに含まれない特徴量。その他は、和歌山市と共通する特徴量。

4.2. モデルの構築

次に、それぞれのデータを学習させた空き家検出モデルを構築する。初めに、和歌山市の公共データから構築した「空き家検出モデル_W」の予測精度を表10に示す。本研究における予測精度とは、全体のデータの内7割を訓練データに用いてモデル構築した際の、残り3割のテストデータに対する予測精度のことを指す。なお、正解率とは予測の当たった建物の数を全体数で除した値である。また適合率が高いほど、予測された空き家が実際に空き家である確率が高く、再現率が高いほど、実際の空き家を予測できた割合が高いことを意味する。F値は適合率と再現率の調和平均である。それぞれの評価指標の詳細な定義については昨年度の報告書4.1を参照されたい。同様に、田辺市の公共データから構築した「空き家検出モデル_T」の予測精度を表11に示す。

表10、表11より、両モデルの正解率はどれも高い一方で、適合率や再現率に異なる傾向が見られる。空き家検出モデル_Wの特徴は、高い適合率（0.910）である。これは予測した空き家の多くが、実際に空き家である可能性が高いことを示す。また再現率も0.626であり、実際の空き家の6割以上を予測することが出来ている。一方、空き家検出モデル_Wに比べると、空き家検出モデル_T（調整なし）の予測精度は高くない。特に再現率は0.131と低く、実際の空き家の多くを予測できなかった。そこで空き家データのオーバーサンプリングや空き家クラスの誤分類に対する重み付けを行ったところ、再現率は6割程度まで改善したものの、適合率は調整前のモデルより低くなってしまった。これは、モデルが実際の空き家をなるべく多く予測しようとした結果、実際には空き家でない多くの住宅を空き家であると誤分類したためである。このように、両モデルはその学習データの違いによって、異なる予測精度の傾向が見られた。

表10 空き家検出モデル_Wの予測精度

正解率	適合率	再現率	F値
0.978	0.910	0.626	0.742

表11 空き家検出モデル_Tの予測精度

	正解率	適合率	再現率	F値
調整なし	0.965	0.702	0.131	0.220
オーバーサンプリング	0.955	0.435	0.588	0.500
重み付け	0.955	0.435	0.644	0.512

注：本モデルでは、田辺市の空き家数が全体の3.8%しかないことによるデータの不均衡性に対処するため、空き家データのオーバーサンプリングと、空き家を非空き家と予測したミスに対する重み付けを行った場合の予測精度も調査した。

続いて橋本市の公共データから構築した空き家検出モデル_Hの予測精度を表12に示す。正解率は0.978と高く、また適合率も1となっており、予測された空き家の全てが実際に空き家と予測できていることがわかる。しかし、再現率は0.338と低く、実際の空き家の多くを予測できていない結果となっている。ところが、モデル構築に用いるデータを減らすことで再現率が上昇した。登記データを除外して、住基データと水道データのみからモデルを構築すると、再現率は0.518に上昇した。そこからさらに水道データも除外すると、すなわち住基データのみからモデルを構築すると、再現率は0.797に上昇した。しかも適合率は1のまま減少していないので、住基データのみを用いて構築したモデルから予測を行えば、予測された空き家のすべてが実際に空き家で、かつ現地調査の約80%をカバーできていることになる。

このような結果が得られた要因を考察するために、各特徴量の重要度を算出した。図4～図6にその結果を示す。図4から、全ての公共データを用いたときに最も重要度が高いのは、建物の階数であった。ところが使用するデータを減らしていくと、図5と図6が示すように、重要度の高い特徴量が住基ダミーと世帯内最高年齢に変化した。これは、公共データと空き家データを結合する際に空き家のレコードに住基データが紐付けされないケースが多いためである。すなわち、住基データの存在しない建物は高い確率で空き家ということになる。しかし、住基データが存在しないことを理由に空き家であると予測することは、本研究のモデルを現地調査が行われていない市町村に対して適用することが不可能となる。なぜなら、そのような市町村では空き家データが存在しないため、住基データのみを用いて予測をしようとしても、住基データの無い建物はそもそもデータベースの中に現れることがなく、空き家予測の対象にならないためである。このように、空き家検出モデル_Hの予測精度は過大評価されている可能性が高い可能性がある点に注意が必要である。

表12 空き家検出モデル_Hの予測精度

	正解率	適合率	再現率	F値
住基・登記・水道データ	0.978	1	0.338	0.505
住基・水道データのみ	0.981	1	0.518	0.683
住基データのみ	0.989	1	0.797	0.887

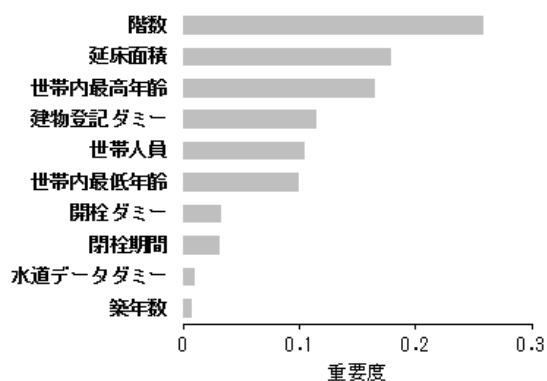


図4 空き家検出モデル_Hの重要な特徴量

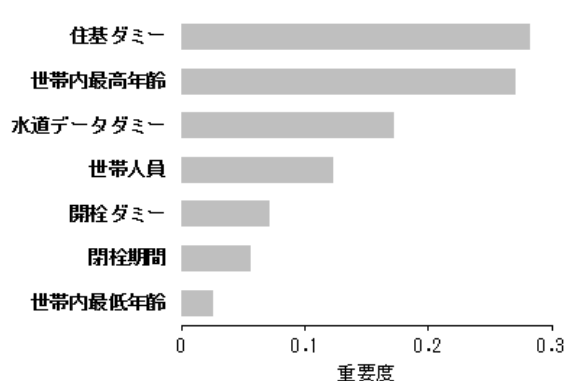


図5 空き家検出モデル_Hの重要な特徴量 (住基・水道データのみ)

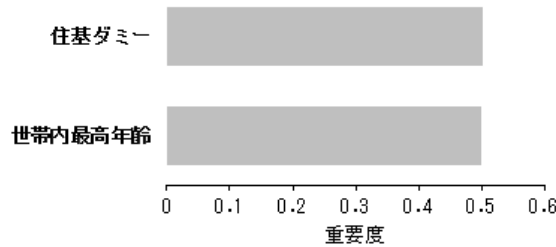


図6 空き家検出モデル_Hの重要な特徴量
(住基データのみ)

4.3. モデルの外挿

本節では、空き家検出モデルの外挿、具体的には空き家検出モデル_Wを用いた田辺市の建物毎の空き家分類を行う。和歌山市の公共データで学習したモデルが、どれ程の精度で田辺市の空き家を検出可能か分析することで、建物スケールの空き家分布予測に対する外挿性が評価できるだろう。

ところで、空き家検出モデルの外挿にはいくらかデータの編集が必要である。なぜなら、表7、表8が示す様に、和歌山市と田辺市における公共データの特徴量は相異なるためである。そこで、以下のように特徴量の補完、削除、もしくは変換を行った。

- 空き家検出モデル_Wに用いられるが、田辺市の公共データに存在しない特徴量がある場合、欠損値として新たに以下の変数を作成する。
 - 指定建蔽率
 - 住居専用地域ダミー
 - 住居地域ダミー
 - 商業地域ダミー
 - 工業地域ダミー
 - 指定容積率
 - 住宅ダミー
- 空き家検出モデル_Wに用いられないが、田辺市の公共データに存在する特徴量がある場合、以下の変数を削除する。
 - 旧田辺市ダミー
 - 簡易水道ダミー
 - 建物登記ダミー
 - 水道データダミー
- 空き家検出モデル_Wに用いられる特徴量と、田辺市の公共データに存在する特徴との取得年度が異なっている場合、年度をずらす。
 - 水道使用量2017 → 水道使用量2016
 - 水道使用量2018 → 水道使用量2017

以上のデータ変換を行い、空き家検出モデル_Wを田辺市の公共データに適用した。その結果を、表13に示す。表10と表13の比較より、空き家検出モデル_Wの外挿精度は、元々の予測

精度より著しく低くなった。また、表11と表13の比較より、空き家検出モデル_Wの外挿結果は、概して空き家検出モデル_Tの予測精度よりも低いことが分かった。これより、和歌山市の公共データを用いたモデルの外挿による、田辺市の建物毎空き家分類は、現時点で十分な精度を担保できないことが分かった。

外挿によって最も精度の低下した指標は、適合率である。空き家検出モデル_W、空き家検出モデル_T（調整なし、オーバーサンプリング、重み付け）の適合率はそれぞれ、0.910、0.702、0.435、0.435であった。しかし、外挿による適合率は0.168である。つまり、モデルの外挿によって予測された空き家の多くは、実際には空き家ではない可能性が高いと言える。

ただし、全ての指標が著しく低下したわけではない。例えば、空き家検出モデル_T（調整なし）の再現率と、空き家検出モデル_Wの外挿による再現率を比較すると、後者の方が高い。これは、外挿による精度低下幅が、不均衡データによる精度低下幅よりも小さいことを示唆している。結果として、空き家検出モデル_T（調整なし）のF値と、空き家検出モデル_Wの外挿によるF値は、ほぼ同等である。

表13 空き家検出モデル_Wの外挿結果

正解率	適合率	再現率	F値
0.916	0.168	0.307	0.217

注：外挿の結果は、田辺市の全データに対する予測結果に基づいて算出されている。これに対して、表10～表12の予測精度は、全データの内3割のテストデータに対する予測結果に基づいて算出されている。テストデータのサンプルサイズは異なるものの、本質的には同じ操作を行っていると考えてよい。

以上の様に建物スケールで空き家を検出するモデルの外挿は、現時点で十分な精度を示さない。この原因としては、3つの点が挙げられよう。第1に、空き家検出モデル_Wの過学習である。しかし、この影響が大きいとは考えにくい。先述のように空き家検出モデル_Wは、テストデータに基づく予測精度で高い値を示しているからである。もし空き家検出モデル_Wが過学習しているのであれば、外挿結果のみならず予測精度も低い値となったであろう。

そこで第2の原因として、和歌山市と田辺市間での特徴量の違いが考えられる。本研究ではモデルを外挿する際、各データのみ存在する変数の補完や削除、変換を行った。そうした操作を行った特徴量の中には、予測上重要なものも含まれる。例えば図7から、「住宅ダミー」や「商業地域ダミー」「指定建蔽率」は、空き家検出モデル_Wの重要な特徴量であることが分かるが、田辺市の公共データにはこれらの特徴量が含まれていないため欠損値として補完せざるを得なかった。こうした特徴量を予測に反映することが出来なかったため、モデルの外挿精度が低下したものと考えられる。

最後に第3の原因として、地域ごとの空き家発生要因の違いが示唆される。つまり、和歌山市と田辺市では空き家の分類に用いられるべき決定木の構造が、そもそも違うという仮説である。この場合、たとえ空き家検出モデル_Wが過学習しておらず、また和歌山市と田辺市の公共データ間に特徴量の違いがなくとも、外挿による精度を担保することは出来ない。ここで参考に、空き家検出モデル_T（重み付け）の予測上重要な特徴量を確認する（図8）。図7と図8の比較より、どちらも水道データが重要な特徴量として共通する一方、「最低年齢」

や「階数」といった特徴量は空き家検出モデル_T（重み付け）でのみ重要となっている。更に、同じ特徴量を用いてもその閾値が異なる場合もあるだろう。

以上のような地域毎の空き家発生要因構造は、その地域の空間的な特徴（大都市であるか、山間部であるか等）によって分類可能であると考えられる。地域の特徴に応じたタイポロジーを用いることで、「地域Aと地域Bは、同じ地域グループに属するため、外挿による妥当な精度が担保される」「地域Aと地域Cは、異なる地域グループに属するため、外挿の精度が担保されない」といった判断が可能になるだろう。外挿の指針となる地域タイポロジーを作成することは、今後の重要な研究課題である。

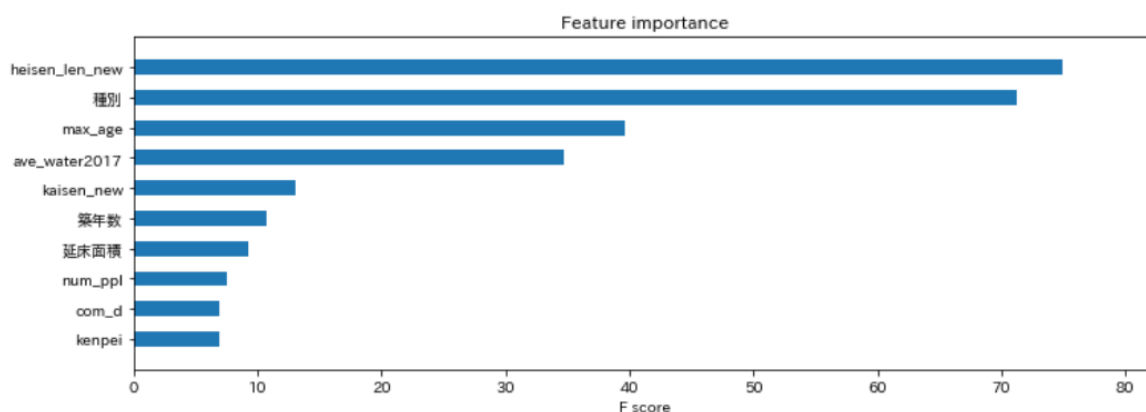


図7 空き家検出モデル_Wの重要な特徴量

注：重要度タイプの引数を” gain” とした時のFeature importanceを、上位10個まで表示してある。” 種別” が「住宅ダミー」、” kenpei” が指定建蔽率、” com_d” が商業地域ダミーを指す。

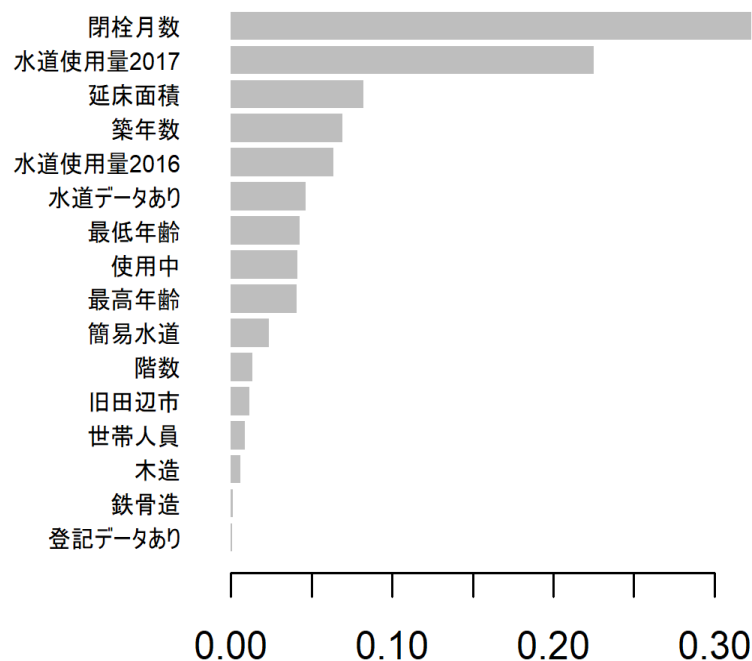


図8 空き家検出モデル_T（重み付け）の重要な特徴量

5. 空き家率回帰モデルの外挿

本章では基本単位区毎の空き家率予回帰モデルの外挿性を評価する。具体的には、和歌山市・田辺市それぞれの国勢調査を学習した2つのモデルを構築し、各モデルをもう一つの地域の空き家率予測に適用する。2方向からの外挿性を評価することで、単なる外挿性の精度分析に留まらず、外挿性の相互関係をも評価することが出来るだろう。

なお前章との違いは、2点ある。第1に、前章は建物毎に空き家か否かを予測する、2値分類モデルの評価であるのに対し、本章は基本単位区毎の空き家率を予測する回帰モデルの評価である。第2に、前章は和歌山市モデルを田辺市に適用する1方向分析であるのに対し、本章ではその逆も行う双方向分析である。これは、国勢調査データの構成が異なる地域間でも共通しているという特徴を生かした分析である。

5.1. 学習データ

学習データには国勢調査を用いた。国勢調査に含まれる変数は全国を通して共通している。そのため、和歌山市と田辺市間のモデルの外挿は容易に行える。

5.2. モデルの構築

次に、和歌山市と田辺市、それぞれのデータを学習させた空き家率回帰モデルを構築する。ここでは、和歌山市の国勢調査から構築したモデルを「空き家率回帰モデル_W」、田辺市の国勢調査から構築したモデルを「空き家率回帰モデル_T」と呼ぶ。両モデルのチューニングされたパラメータと予測精度を表14に示す。なお、R2とは決定係数であり、回帰によって導いたモデルの当てはまりの良さを表現する。またMAEとは平均絶対誤差(Mean Absolute Error)であり、予測した空き家率と実際の空き家率との絶対誤差の平均を意味する。

表14より、チューニングされたパラメータの値は大きく変わらない一方で、精度指標には異なる傾向が見られる。すなわち、空き家率回帰モデル_WはR2が高く（モデルの当てはまりが良い）、空き家率回帰モデル_TはMAEが低い（予測した空き家率の誤差が小さい）。それぞれの原因としては、和歌山市の方が田辺市よりも基本単位区数、つまりサンプルサイズが大きいこと、田辺市の方が平均の空き家率が低いために誤差が小さくなりやすいことが原因であると考えられる。

表14 空き家率回帰モデルのパラメータと予測精度

	パラメータ			精度指標	
	max_depth	sub_sample	n_estimators	R2	MAE
空き家率回帰モデル_W	8	0.85	49	0.918	0.0522
空き家率回帰モデル_T	7	0.8	52	0.780	0.0309

注：“max_depth”は決定木の最大の深さ、“sub_sample”はサブサンプリングの抽出比率、“n_estimators”はラウンド数、すなわち決定木の個数を指す。また“n_estimators”は、early_stoppingの機能を用いてチューニングしている。

5.3. モデルの外挿

続いて両モデルを、もう片方の地域の空き家率回帰に適用した結果を表15に示す。表14と表15の比較より、外挿しても大きく精度が変化しないことが分かった。前章では、建物毎の空き家分類におけるモデルの外挿は、十分な精度を担保出来なかった。しかし本章の結果では、基本単位区毎の空き家率回帰のモデルの外挿が、十分な精度を維持している。すなわち、基本単位区のスケールにおける空き家率は、建物スケールの空き家分類に比べて、地理的な影響を受けづらいということが示唆される。

さらに、空き家率回帰モデルTの予測精度と空き家率回帰モデル_Wの田辺市への外挿結果を比較すると、後者は外挿であるにもかかわらず、指標値を改善している (R2: 0.780→0.867, MAE: 0.0309→0.0293)。この原因について、表14のチューニングされたパラメータ値に着目すると、空き家率回帰モデル_Tのパラメータ値は、max_depth, subsample共に空き家率回帰モデル_Wよりも低い値であることが分かる。このことは、空き家率回帰モデル_Tがデータから得る情報量が、もう一方のモデルのそれよりも少ないことを意味する。従って空き家率回帰モデル_Tは田辺市の国勢調査サンプルサイズが小さいがために、十分な学習が出来なかったのではないだろうか。もしこの仮説が正しければ、近接した地域では一方の地域のサンプルサイズが小さい場合、もう一方のモデルを外挿することで十分な精度の空き家率回帰が可能になると言える。

本章の結果は、異なる地域で学習したモデルの外挿による、基本単位区の空き家率回帰の実効性を支持している。今後の研究は、国勢調査の全国共通のデータ構造を活かし、様々な地域間の外挿結果を検証することで、より詳細な外挿の信頼性や、地域の空間的特徴に応じたタイポロジーの作成等を行うことが期待される。

表15 空き家率回帰モデルの外挿の結果

	R2	MAE
空き家率回帰モデル_W → 田辺市	0.833	0.0293
空き家率回帰モデル_T → 和歌山市	0.867	0.0575

6. 住宅・土地統計調査を統合した国勢調査モデル

自治体によっては水道や住民基本台帳等の公共データを使用することが難しい場合がある。また、自治体のデータは基本的に自治体ごとに仕様が異なるので、他自治体へのモデルの外挿や全国展開のためには自治体ごとに整備されている公共データよりは、全国で一様な規格として整備されている政府統計を使用する手法を用いることで、実施が容易になる。

昨年度の研究では、2015年度に実施された国勢調査の統計データを用いた空き家分布推定が行われた。国勢調査による空き家分布推定の結果として、モデルの予測空き家率と実際の空き家率との誤差を集計すると、基本単位区毎のRMSEは0.18となり、基本単位区単位における空き家率や空き家数の推定において、政府統計は十分に利用価値があることが明らかとなった。

そこで今年度の研究では、昨年度の研究を拡張して国勢調査に加えて住宅・土地統計調査を用いて調査区単位の空き家率を推定することの有効性を検証し、精度を高める手法確立を追求した。なお、今後の検討課題として政府統計を用いた将来の空き家予測が挙げられるが、その検討プロセスとして2015年に実施された国勢調査に加えて、2010年の国勢調査の統計データを追加的に活用することで、複数年度のデータによる推定手法を開発した。

6.1. 学習データ

学習データには2015年実施分の国勢調査（以下、国調2015）と、2010年実施分の国勢調査（以下、国調2010）に加えて、2013年実施分の住宅・土地統計調査（以下、住調2013）を用いた。なお、国勢調査と住宅・土地統計調査の特徴量については第5章と同じく、それぞれ**附表1**と**附表2**を参照されたい。

国勢調査は基本的に全数調査が実施されており、調査の結果は街区もしくは街区に準じた小区画からなる基本単位区ごとに結果が集計されている。なお、国勢調査の調査区は一つまたは二つ以上の基本単位区を組み合わせて設定されている。なお、実施年度によって基本単位区に変動があるため、調査区にも変動ができ、複数年度のデータを用いる際には調査区の変動に注意する必要がある。

一方で、住宅・土地統計調査の場合は、直近の国勢調査の調査区に基づいて調査結果が集計されており、住調2013の場合国調2010と同様な調査区で結果が集計されている。なお、住宅・土地統計調査の場合、全数調査は実施されておらず、一部の調査区に対してサンプル調査が実施されている。**表16**和歌山市における国調2010、国調2015と住調2013のサンプルサイズを示す。

表16 和歌山市における統計データのサンプルサイズ

	国勢調査 (2015)	国勢調査 (2010)	住宅・土地統計調査 (2013)
サンプルサイズ	2582	2545	348

6.2 モデルの分類

昨年度のモデルで用いた国調2015に加えて、新たに国調2010と住調2013を利用することによる精度の差を比較検証するために、利用したデータセットの組み合わせ方により8つのモデルを作成した。表17にモデルごとに利用したデータをまとめた。なお、国調2010と住調2013は、同じ調査区で集計されているが、国調2015では異なる調査区で集計がされているため、国調2015を他のデータと組み合わせる際には調査区の面積で按分する作業を通して国調2010の調査区に国調2015のデータに割り当てた。

表17 政府統計を用いた8つのモデル一覧

モデル 番号	使用データ			学習データ サンプル数 *1	予測データ サンプル数 *1	決定係数 (R2)	平均絶対誤 差 (MAE)
	国調 2015	国調 2010	住調 2013				
1	●			2,582	2,582	0.918	0.0521
2		●		2,545	2,545	0.902	0.0543
3		● (一部)		348	348	0.983	0.0488
4		● (一部)	●	348	348	0.999	0.0366
5		●	○*2	2,545	2,545	0.708	0.0649
6		● (一部)	●	348	348		0.0566
		●		2,197	2,197		
7		● (一部)	●	348	348		0.0513
		●		2,545	2,197		
8	●	●	●	2,593	2,593	0.967	0.0443

*1 実際には3割をランダム抽出して用いた。

*2 住調2013が存在しない調査区においては住調ダミーのみを属性とした。

6.3 各モデルの推定結果

本節では、各モデルの推定結果について述べる。まず、モデル1は、国調2015のみを使用したモデルである。決定係数（R2）が0.918であり、平均絶対誤差も0.0521となっていることから、単年度実施分の国勢調査だけでも十分高い精度でエリアの空き家率を予測できることがわかる。図9にモデル1の調査区別のMAEの分布を表す。

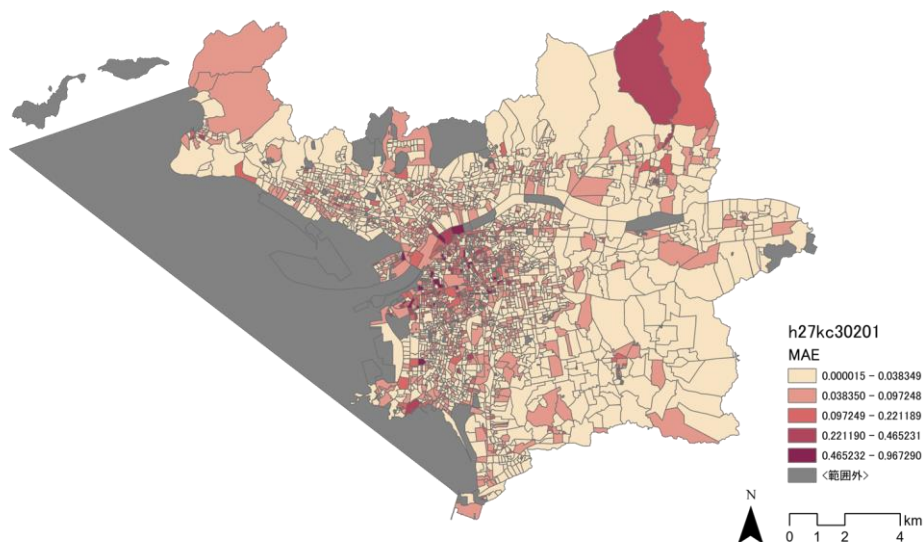


図9 モデル1における調査区別のMAE分布

次に2010年度実施分の国勢調査の結果を用いて予測を行ったモデル2の結果について述べる。本研究で用いる空き家分布は、2017年に調査されたものであり、2010年の国勢調査から2017年の空き家分布を予測することは、7年間の時期のずれがあるため、未来予測を行うという性格ももっている。モデル2は決定係数（R2）0.902、平均絶対誤差（MAE）0.0543という結果であったが、モデル1に比べるとやや精度が落ちるものの、十分に高い精度で2017年時点の空き家分布を推定することができた。図10にモデル2の調査区別のMAEの分布を表す。

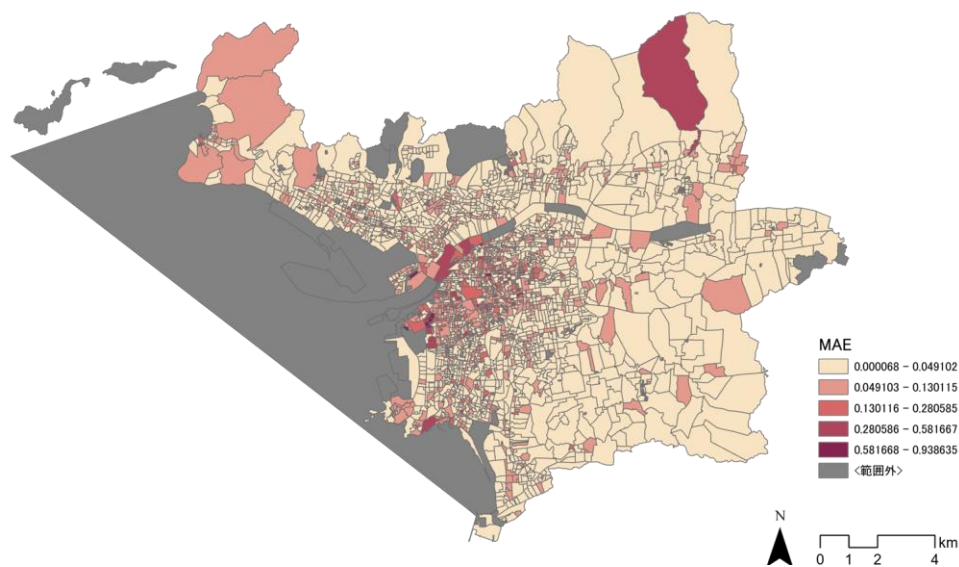


図10 モデル2における調査区別のMAE分布

モデル3は、モデル2の学習データである国調2010のうち、住調2013で調査が行われた348調査区のみを抜粋したものであり、モデル4はモデル3に住調2013のデータを加えて予測したモデルである。モデル3とモデル4の精度を比較することで、住宅・土地統計調査を加えた場合の効果を確認することができる。結果として、モデル4は決定係数(R²)0.999、平均絶対誤差(MAE)0.0366となり、調査区は限られてしまうものの、非常に高い精度で空き家率を予測することが可能になった。なお、モデル3の結果よりもモデル4の方の決定係数が大きく、平均絶対誤差が小さいことから、住宅・土地統計調査が精度向上に寄与していることが確認された。図11と図12にそれぞれモデル3とモデル4の調査区別のMAEの分布を表す。

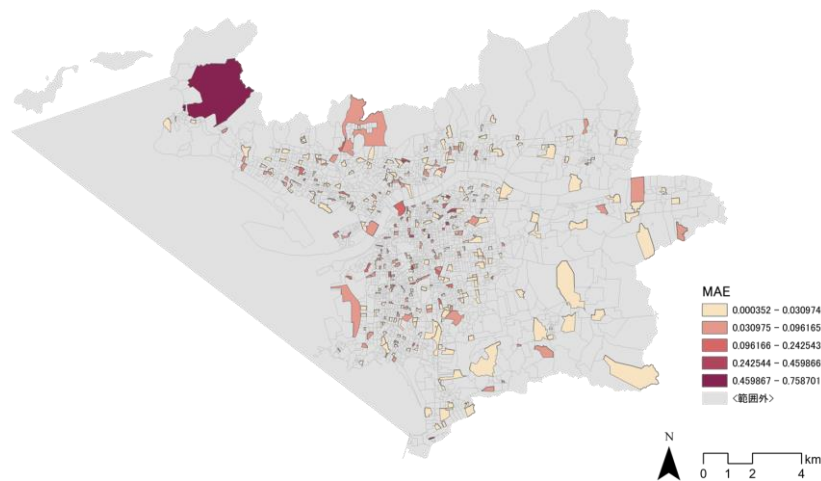


図11 モデル3における調査区別のMAE分布

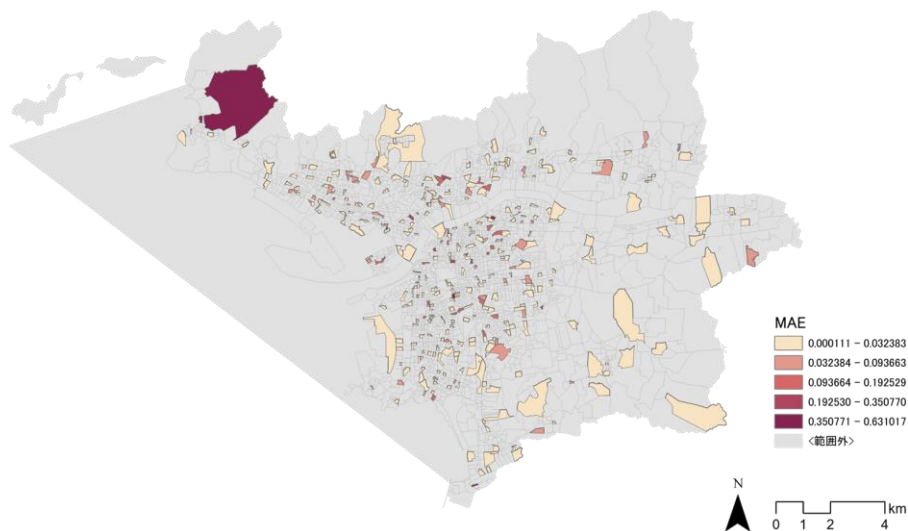


図12 モデル4における調査区別のMAE分布

モデル4では、住宅・土地統計調査のサンプルデータが存在する調査区のみを対象にして推定を行い、住宅・土地統計調査のサンプルが存在する調査区では住宅・土地統計調査を加えることで精度が向上することを確認することができた。続いてモデル5、モデル6、モデル7は、一部サンプルのみ調査がされている住宅・土地統計調査の結果と全数調査が行われている国勢調査の両方のデータを活用して市全域の空き家率を推定したモデルである。各モデルの違いは表18の通りである。

まず、モデル5は、住宅・土地統計調査の調査結果が存在する調査区を1、調査結果がしない調査区を0とするダミー変数を作成し、市全域に亘って国調2010と住調2015のサンプルを一律的に学習したものである。学習の結果として、R2が0.708、MAEが0.0649となり、住宅・土地統計調査をダミー化して入れる方法では、むしろ精度が落ちてしまうことがわかった。

これを改善するために、モデル6では住調2013のデータが存在しない地域では、国調2010のみで学習し、住調2013が存在する地域では、モデル4と同様に国調2010・住調2013をあわせて学習する方法を採用した。しかし、この場合もMAEが0.0566で、絶対値としては悪い数値ではないが、国調2010を単独で用いたモデル2よりも、MAEが上昇する結果が得られた。全体のMAEが上昇した原因として、住宅・土地統計調査の結果が存在しない地域のみで国調2010で学習した結果、モデル2よりサンプル数が減少し精度が悪くなったことが原因であると考えられる。

さらに、モデル6における住宅・土地統計調査の結果が存在しない地域のMAEの高さを補正する方法として、モデル7では市全域において国調2010を学習し、住宅・土地統計調査が存在しない地域のみにもその結果を適用する形でサンプルデータの数を増やすことで、MAEを低くすることができた。その結果、モデル7では市全域におけるMAEが0.0513と測定され、住宅・土地統計調査を加えることで市全域における空き家推定の精度が改善することが確認された。図13にモデル7の調査区別のMAEの分布を表す。

表18 モデル5、モデル6、モデル7の分類

	モデル5	モデル6	モデル7
住宅・土地統計調査の結果が存在しない地域	住宅・土地統計調査存在ダミーの値を0にして、国調2010・住調2013を合わせて学習	住調2013のデータが存在しない地域のみ、国調2010のみで学習 MAE=0.0586	全域にわたって国調2010で学習した結果を使用（モデル2の結果） MAE=0.0537
住宅・土地統計調査の結果が存在する地域	国調2010・住調2013を合わせて学習	国調2010・住調2013を合わせて学習（モデル4の結果） MAE=0.0366	国調2010・住調2013を合わせて学習（モデル4の結果） MAE=0.0366
住調結果存在有無でサンプルを分離して学習したか	X	0	0
市全域におけるMAE	0.0649	0.0566	0.0513

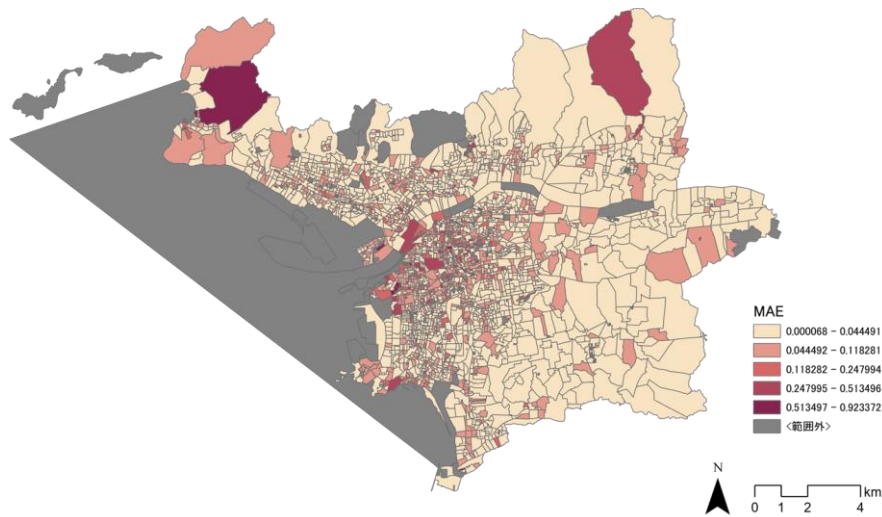


図13 モデル7における調査区別のMAE分布

最後に、モデル8では国調2010、国調2015、住調2013の全てのデータを利用して空き家率を推定した。国調2010と住調2013は同じ調査区で分割されているため、調査区に修正を加える必要がなかったが、国調2015は他のデータとは異なる調査区で結果が集計されているので、国調2015の結果においては、国調2015の調査区を国調2010の調査区に面積比で按分する形で、データの統合を実現した。モデル8はモデル7に国調2015のデータを加える形で推定を行い、 R^2 が0.967、MAEが0.0443という非常に高い精度で空き家率を予測することができた。この結果から、データが集計される調査区の形が異なる複数の統計であっても、面積按分により空き家推定の精度向上が期待できることが明らかになった。図14にモデル8の調査区別のMAEの分布を表す。

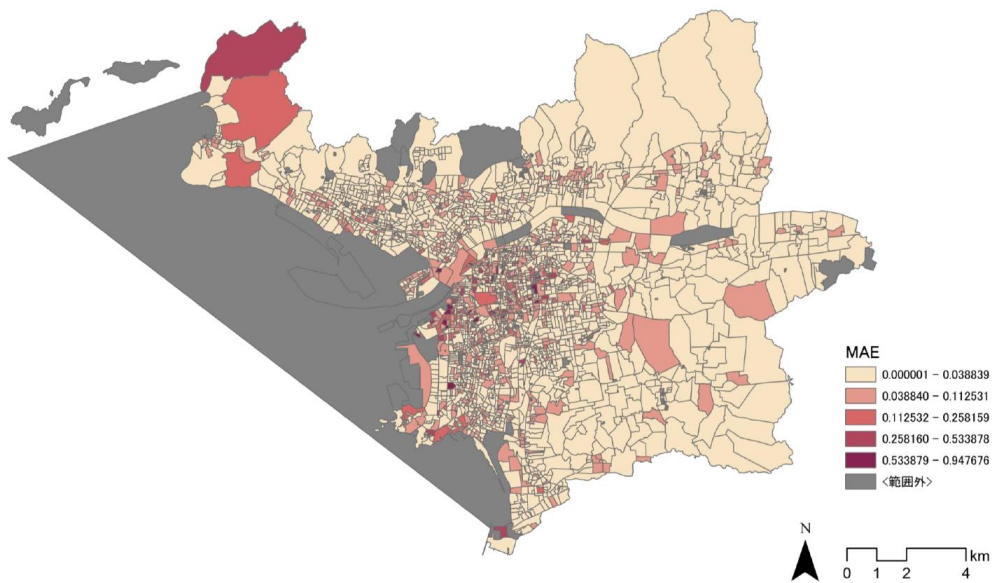


図14 モデル8における調査区別のMAE分布

7. まとめと今後の検討事項

最後に今年度の成果についてのまとめと、今後検討が必要な事項についてまとめる。

7.1. 今年度の成果のまとめ

今年度は前年度の課題を受けて、主に以下の2点についての検証を行った。まず1つ目として、和歌山市において構築した空き家予測モデルの外挿性を明らかにし、また公共データを用いたモデルによる建物毎の予測と同様に、国勢調査のみを用いたモデルによる基本単位区毎予測結果の外挿を行った。その結果、異なる地域で学習したモデルの外挿において、現時点において公共データを用いた建物単位の空き家検出では十分な精度が出ていないものの、基本単位区の空き家率回帰においては地域に関係なく、高い精度を維持していることがわかった。今後の研究は、国勢調査の全国共通のデータ構造を活かし、様々な地域間の外挿結果を検証することで、より詳細な外挿の信頼性や、地域の空間的特徴に応じたタイポロジーの作成等を行うことが期待される。

そして2つ目として、公的統計を用いた空き家予測モデルの改良を行うために、国勢調査と住宅・土地統計調査を組み合わせたモデルの予測精度についての検証を行った。その結果、国勢調査に住宅・土地統計調査を組み合わせることで精度向上が期待でき、さらには複数年度の国勢調査の結果を用いることで、調査区を面積按分法することで空き家率予測の精度向上ができることが確認された。

以上の結果より、前年度までに構築した空き家予測モデルの外挿性が明らかになるとともに、公的統計を用いた空き家予測モデルの予測精度の改良が実現した。以上の成果は、前年度までの成果で課題となっていた、和歌山市以外の自治体に同手法を広く普及させる上で大きな進歩となったといえるだろう。

7.2. 未来予測モデルの検討

2015年の国勢調査から2017年時点の空き家率を推定することに加えて、2010年のデータから2017年という、7年後の空き家率を高い精度で推定することができた。なお、データが集計される調査区が異なるため、統合が困難である複数年度の国勢調査の結果を面積按分法により統合することで、空き家予測の精度が確認されることが確認できた。以上の成果から、2010年の国勢調査から2017年の空き家率を予測するモデルに、2015年の国勢調査のデータを入れることで2022年時点の空き家率予測が可能となることが期待される。また、その時点に空き家調査が実施されるのであれば、未来予測モデルの評価もできると期待される。

7.3. マニュアル化の検討

今年度の成果から、公共データや国勢調査等の政府統計（マイクロデータ）を活用することで、和歌山市だけでなく他の自治体でも高い精度で空き家の分布を予測できることが明らかとなった。そこで次のステップとして、同手法を今年度対象とした自治体以外に適用できる環境構築が考えられる。適用する自治体を増やすことで、同手法の外挿性や課題、限界な

どを明らかにすることができるものと期待されるだけでなく、多くの自治体における空き家関連業務の負担軽減にも貢献できる。

ただし、今年度と歌山市以外に展開する過程で明らかとなったことは、住基データや水道データ、建物登記データといった自治体保有のデータの処理を外部委託することは、これらのデータが個人情報に該当することから、そのハードルは高い場合があるということである。このハードルのため、同手法を他自治体に展開することは必ずしも容易ではないという課題が明らかになった。

そこで、来年度は今年度までに実施してきた手法を自治体内部で再現し実行できるように、作業フロー全体のマニュアル化も検討したい。現時点で想定される内容としては以下のような項目が挙げられる。

- ・分析に使用するデータ（自治体保有データや政府統計など）について
- ・分析に使用するデータの加工方法（分析可能なデータ形式への整形方法など）
- ・機械学習モデルの実装方法（Python環境の構築方法）
- ・機械学習モデルを用いたデータ分析方法（機械学習の実行方法）
- ・予測結果の精度検証方法（混同行列を用いた予測結果の解釈など）
- ・予測結果の可視化方法（QGIS等のGISソフトウェアを活用）

7.4. 総括

本研究を通して和歌山市が保有する公共データ（住基データ、水道データ、建物登記データ）、市による空き家調査データ、また政府統計データ（国勢調査、住宅・土地統計調査）を活用することにより、和歌山市全域の空き家分布状況を迅速に推定するモデルの構築が実現した。また、同モデルの信頼性の検証も実施し、高い信頼性であることも明らかとなった。さらに、複数年度のデータを用いることが可能であり、7年後という未来時点の空き家率予測が高い精度でできることが確認された。加えて、同様の手法で他自治体においてもモデル構築が可能であることや、和歌山市で構築したモデルを他自治体に外挿することも技術的には可能であることも明らかとなった。

今後は、今年度構築したモデルを改善して信頼性の向上を目指すとともに、公共データの活用が困難な自治体向けに政府統計データのみを用いて空き家分布を予測するモデルの開発も検討したい。また、今年度までに得られた成果をより多くの自治体に展開するためのマニュアルの作成も進めていきたいと考えている。

謝辞

本研究で使用した住宅地図（2019年版Zmap TOWNⅡ（株式会社ゼンリン）は、東京大学空間情報科学研究センターとの共同研究（No. 880）により使用したものである。また、分析に当たっては、和歌山市から公共データを、総務省統計局から統計マイクロデータを提供していただいた。田辺市および橋本市の公共データについては、両市の庁舎内で職員立会いのものと閲覧・分析させていただいた。加えて本研究では総務省統計局所管の「国勢調査」及び「住宅・土地統計調査」の調査票情報を利用して独自に集計・分析を行ったものである。ここに記して謝意を表したい。

参考文献

1. 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, 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 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1), 日本建築学会計画系論文集, 744, 275-283, 2018.
2. 浅見泰司, 「都市の空閑地・空き家を考える」, プロGRESS, 2014.
3. 高崎経済大学地域科学研究所, 「空き家問題の背景と対策」, 日本経済評論社, 2019
4. 西山弘泰, 「宇都宮市における空き家の特徴と発生要因-宇都宮市空き家実態調査の結果から-」. 駿台史学153, 55-74, 2015.
5. Akiyama, Y., Ueda, A., Ouchi, K., Ito, N., Ono, Y., Takaoka, H. and Hisadomi, K., Estimating the Spatial Distribution of Vacant Houses using Public Municipal Data, *Geospatial Technologies for Local and Regional Development*, 165-183, 2020.
6. Baba, H., Akiyama, Y., Tokudomi, T. and Takahashi, Y., Learning Geographical Distribution of Vacant Houses Using Closed Municipal Data: A Case Study of Wakayama City, Japan, *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, VI-4/W2-2020, 1-8, 2020.
7. 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*, 785-794, 2016.
8. Yamashita, S. and Morimoto, A., Study on Occurrence Pattern of the Vacant Houses in Local Hub City, *Transactions of CPIJ*, 50, 932-937, 2015.

附表

附表1 国勢調査の特徴量

カラム名	大項目名_日本語	項目名_日本語	
AZATANCODE		基本調査単位区コード	
n		基本調査単位区に集約した標本数	
general_hh	世帯の種類	一般の世帯	
single_person		一人世帯（会社等の独身寮の入居者を含む）	
students		寮・寄宿舎の学生・生徒	
inmates_of_med		病院・療養所の入院者	
inmates_of_soc		社会施設の入所者	
other_type_hh		その他	
defence_dorm		自衛隊営舎居住	
correctional		矯正施設入居者	
detached		建て方の種類	一戸建
row			長屋建
apt	共同住宅		
other_type_residence	その他		
num_stories_1_2	建物の階数	1・2階	
num_stories_3_5		3～5	
num_stories_6_10		6～10	
num_stories_11_14		11～14	
num_stories_15_		15階以上	
num_live_floor_1_2	世帯が住む階	1・2階	
num_live_floor_3_5		3～5	
num_live_floor_6_10		6～10	
num_live_floor_11_14		11～14	
num_live_floor_15_		15階以上	
owned	住宅の所有形態	持ち家	
public_rent		公営の借家	
public_corp		都市機構・公社の借家	
private_rent		民営の借家	
for_employee		給与住宅	
room_rent		間借り	
company_dorm		会社等の独身寮・寄宿舎	
other_type_ownership		その他	
num_residents_hh_1	世帯人員	1人	
num_residents_hh_2		2	
num_residents_hh_3		3	
num_residents_hh_4		4	
num_residents_hh_5		5	
num_residents_hh_6		6	
num_residents_hh_7		7	
num_residents_hh_8		8	
num_residents_hh_9		9	
num_residents_hh_above10		10人以上	
married_couple	家族類型	夫婦のみの世帯	
married_couple_child		夫婦と子供から成る世帯	
male_parent_child		男親と子供から成る世帯	
female_parent_child		女親と子供から成る世帯	
couple_parents		夫婦と両親から成る世帯	
couple_single_parent		夫婦とひとり親から成る世帯	
couple_child_parents		夫婦、子供と両親から成る世帯	
couple_child_single_parent		夫婦、子供とひとり親から成る世帯	

couple_other_relatives		夫婦と他の親族（親、子供を含まない）から成る世帯
couple_child_other_relatives		夫婦、子供と他の親族（親を含まない）から成る世帯
couple_parents_other_relatives		夫婦、親と他の親族（子供を含まない）から成る世帯
couple_child_parents_other_relatives		夫婦、子供、親と他の親族から成る世帯
siblings_only		兄弟姉妹のみから成る世帯
other_family		他に分類されない世帯
include_non_relatives		非親族を含む世帯
single_hh		単独世帯
unknown_hh		不詳（単独世帯以外の基本項目記入不備世帯）
only_65_above_single	65歳以上のいる世帯類型	65歳以上世帯員のみ 世帯員が1人世帯
only_65_above_couple		65歳以上世帯員のみ 世帯員が2人世帯（夫婦のみの世帯）
only_65_above_other		65歳以上世帯員のみ 世帯員が2人世帯（その他の世帯）
only_65_above_above3		65歳以上世帯員のみ 世帯員が3人以上の世帯（夫婦とその親の世帯）
hh_65_above3_others		65歳以上世帯員のみ 世帯員が3人以上の世帯（その他の世帯）
under_65_couple		65歳未満世帯員がいる 世帯員が2人世帯（夫婦のみの世帯）
under_65_other_two		65歳未満世帯員がいる 世帯員が2人世帯（その他の世帯）
above3_son_wife		65歳未満世帯員がいる 世帯員が3人以上の世帯（息子夫婦のいる世帯）
above3_daughter_husband		65歳未満世帯員がいる 世帯員が3人以上の世帯（娘夫婦のいる世帯）
above3_single_child		65歳未満世帯員がいる 世帯員が3人以上の世帯（単身の子供がいる世帯（共にいる場合））
above3_son_only		65歳未満世帯員がいる 世帯員が3人以上の世帯（（うち）息子がいる世帯（息子のみの場合））
above3_daughter_only		65歳未満世帯員がいる 世帯員が3人以上の世帯（（うち）娘がいる世帯（娘のみの場合））
above3_other		65歳未満世帯員がいる 世帯員が3人以上の世帯（その他の世帯）
child_0	子供の数	0
child_1		1
child_2		2
child_3		3
child_4_		4 >
child_min_age_0	最年少子供の年齢	0
child_min_age_1_2		1-2
child_min_age_3_5		3-5
child_min_age_6_8		6-8
child_min_age_9_11		9-11
child_min_age_12_14		12-14
child_min_age_15_17		15-17
child_min_age_18_		18 >
male	男女の別	男
female		女
unmarried	配偶者の有無	未婚（幼児などを含む）
with_spouse		配偶者あり
bereaved		死別
divorced		離別
unknown_haigu		不詳
japanese	国籍	日本
foreign		外国
unknown_nationality		不詳
reside_from_birth	居住期間	出生時から
reside_1		1年未満
reside_1_5		1～5年未満
reside_5_10		5～10年未満
reside_10_20		10～20年未満
reside_20_		20年以上
current_address_before_5y	5年前の常住地	現住所

same_neighborhood_before_5y		自市区町村内	
same_district_before_5y		自市内他区	
same_pref_before_5y		県内他市区町村	
another_pref_before_5y		他県	
abroad_before_5y		転入（国外から）	
employed	労働力状態	就業者	
unemployed		完全失業者	
housework		家事	
educated		通学	
other_work		その他	
neither_employed_educated	常住地による従 通	従業も通学もしていない（労働力状態が、5、6、8の者）	
work_at_home		自宅で従業（就業者で自宅または住み込みの者）	
commute_same_district		自宅外の自市区町村で従業・通学（従通者で同じ市区町村への者）	
commute_another_district		自市内他区で従業・通学（従通者で他区への者）	
commute_another_municipality		県内他市区町村で従業・通学（従通者で他市区町村への者）	
commute_another_pref		他県で従業・通学（従通者で他県への者）	
unknown_commute	他市区町村で従業・通学（従通地コード不詳・外国）		
regular_employee	従業上の地位	雇用者（正規の職員・従業員）	
temp_employee		雇用者（労働者派遣事業所の派遣社員）	
part_time_worker		雇用者（パート・アルバイト・その他）	
officer		役員	
employer_w_employee		自営業主（雇人のある業主）	
self_employed		自営業主（雇人のない業主）	
family_employee		家族従業者	
domestic_worker		家庭内職者	
agriculture		産業	A 農業
forestry			U 林業
fishery	B 漁業		
mining	C 鉱業、採石業、砂利採取業		
construction	D 建設業		
manufacturing	E 製造業		
electricity_gas_heat_water	F 電気・ガス・熱供給・水道業		
information_communication	G 情報通信業		
transportation	H 運輸業、郵便業		
wholesale_retail	I 卸売業、小売業		
finance	J 金融業、保険業		
real_estate	K 不動産業、物品賃貸業		
academic_research	L 学術研究、専門・技術サービス業		
accommodation_food	M 宿泊業、飲食サービス業		
lifestyle_entertainment	N 生活関連サービス業、娯楽業		
education	O 教育、学習支援業		
medical	P 医療、福祉		
composite_service	Q 複合サービス事業		
service_other	R サービス業（他に分類されないもの）		
public_other	S 公務（他に分類されるものを除く）		
industry_other	T 分類不能の産業		
ave_stories		平均建物階数	
ave_live_floor_num		平均居住階数	
ave_hh_num		世帯当たり平均居住者数	
ave_hh_male_num		世帯当たり平均男性居住者数	
ave_hh_female_num		世帯当たり平均女性居住者数	
ave_age		世帯当たり平均年齢	

附表2 住宅・土地統計調査の特徴量

カラム名	大項目名_日本語	項目名_日本語
Built_1	建物の種類	住宅で居住世帯のある建物
Built_2		住宅以外の建物
Built_3		住宅で居住世帯のない建物
Built_etc		以外
vacant_first	居住世帯のない住宅	一時居住者のみの住宅
vacant_second		二次的住宅（別荘）
vacant_second_etc		二次的住宅（その他）
vacant_rent		空き家（賃貸用）
vacant_sell		空き家（売却用）
vacant_etc		空き家（その他）
vacant_construction		建築中の住宅
house_exclusive		住宅の種類
house_mixused	店舗その他の併用住宅	
house_norecord	記入なし	
house_type_detached	建て方	一戸建
house_type_row		長屋建（テラスハウスを含む）
house_type_apartment		共同住宅
house_type_etc		その他
structure_wood	建物の構造	木造（防火木造を除く）
structure_fireproof		防火木造
structure_rc		鉄筋・鉄骨コンクリート造
structure_s		鉄骨造
structure_etc		その他
dilapidated	腐朽・破損の有無	有
dilapidated_no		無
elevator	エレベータの有無	エレベータ無し
elevator_no		エレベータ有り
autolock	オートロックの別	オートロック式
autolock_no		オートロック式ではない
silver	高齢者対応型住宅の別	高齢者対応型住宅である
silver_no		高齢者対応型住宅ではない

area	建築面積	m ²
occupancy_2013	現住居への入居時期	平成25年
occupancy_2012		平成24年
occupancy_2011		平成23年
occupancy_2010		平成22年
occupancy_2009		平成21年
occupancy_2006		平成18年～平成20年
occupancy_2001		平成13年～平成17年
occupancy_1996		平成8年～平成12年
occupancy_1991		平成3年～平成7年
occupancy_1981		昭和56年～平成2年
occupancy_1971		昭和46年～昭和55年
occupancy_1961		昭和36年～昭和45年
occupancy_1951		昭和26年～昭和35年
occupancy_before1950		昭和25年以前
ownership_owned	持ち家・借家の別	持ち家
ownership_private_rent		民営の賃貸住宅
ownership_public_rent		都道府県・市区町村営賃貸住宅
ownership_public_corp		都市再生機構・会社などの賃貸住宅
ownership_for_employee		給与住宅
ownership_room_rent		借間
ownership_non_house_rent		住宅以外の建物_賃貸など
ownership_other_type_ownership		住宅以外の建物_所有
ownership_V		不詳
monthly_rent	賃料	1ヶ月の家賃又は間代
monthly_management		共益費又は管理費
GFA	面積	床面積
1FA		1階の面積
Construction_year_2013	建築時期	平成25年
Construction_year_2012		平成24年
Construction_year_2011		平成23年
Construction_year_2010		平成22年

Construction_year_2009		平成21年
Construction_year_2006		平成18年～平成20年
Construction_year_2001		平成13年～平成17年
Construction_year_1996		平成8年～平成12年
Construction_year_1991		平成3年～平成7年
Construction_year_1981		昭和56年～平成2年
Construction_year_1971		昭和46年～昭和55年
Construction_year_1961		昭和36年～昭和45年
Construction_year_1951		昭和26年～昭和35年
Construction_year_before1950		昭和25年以前
handrail_gate	バリアフリー施設	手すり_玄関
handrail_toilet		手すり_トイレ
handrail_bath		手すり_浴室
handrail_dressing		手すり_脱衣所
handrail_hall		手すり_廊下
handrail_stair		手すり_階段
handrail_room		手すり_居住室
handrail_etc		手すり_その他
bath		浴槽
hallwidth		廊下幅
barriorfree		段差のない屋内
approach		道路から玄関の通行
none_silver		設備なし

付録

XGBoostによる推定値の算出方法（令和2年度報告書再掲）

XGBoostは決定木の中でも回帰木を利用しており、建物ごとの空き家か否かの判定は「空き家確率」として最終的に出力される。本モデルは決定木を逐次的に学習させていくものであり、 t 番目の木を学習させるためには、 $t - 1$ 番目までの全ての木の情報を用いる。ただし、木の数が大きくなるにつれ、誤差が小さくなるため、その改善の余地が少なくなっていく。

空き家の推定確率を算出する際、大まかな流れは以下の通りである。

- 決定木を T 本作成する。そのため、以下の流れを繰り返し行う。
- 決定木は、分岐を繰り返すことで作成し、その際に特徴量（例えば、住基の建物内人員数など）の閾値を設定する。
- 閾値設定は全ての候補を調べ、分岐させた際に最適な葉の重みを設定したとき、損失関数の減少が最大になるものを選択する。
- 上記の決定木の作成により予測値を更新する。

続いて、損失関数に基づく最適な重み付け値について簡潔に述べる。当モデルにおいて、 t 番目の決定木を f_t とおくと、 t 番目の単純な形式での損失関数は $\sum_{i=1}^I l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right)$ と表せる。ただし、 x_i は i 番目の入力データ、 y_i は i 番目の出力データ、 $\hat{y}_i^{(t-1)}$ は $t - 1$ 番目までの木を用いた i 番目出力データの予測値（分類の場合には確率値）である。式の意味するところは、 $t - 1$ 番目までの決定木をベースとして、 $f_t(x_i)$ を加味することで t 番目の損失をさらに減少させることを意味する。ただし、上記の損失関数では過学習してしまう恐れがあるため、罰則項を加えて修正し、 t 番目の損失関数 $L^{(t)}$ を、

$$L^{(t)}(f_t) = \sum_{i=1}^I l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \gamma T + \frac{1}{2} \lambda \|w\|^2$$

と表す。ただし、 T は最終ノードの数を、 w は最終的な出力値の集合を示している。ここで、最終ノード数が増えるほど、出力値の種類が多くなるほど学習データに対応出来るようになるが、それによる過学習を抑えるため、 γ, λ はそれぞれ罰則項の調整パラメータとして設定する必要がある。 $\sum_{i=1}^I l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right)$ の推定に際して、 $f_t(x_i)$ まわりのTaylor展開を二次の項まで行うことで近似解を求められる。最終的に、 $L^{(t)}$ を最小化することにより、 j 番目の最適な重み付け値 w_j^* を求めることができる。

XGBoostはパラメータチューニングを必要とするため、グリッドサーチによるチューニングを行う。木の深さの最大値を表すmax_depthは2-8、子ノードでのデータの重み付け合計値の最小値を表すmin_child_weightは1-3、各木でのランダム抽出される列の割合を表すcolsample_bytreeは0.5-1.0、各木でのランダムな抽出を表すsubsampleは0.5-1.0の範囲でそれぞれ試行する。

モデルの評価関数としてはエラー率 $error_i$ を採用しており、下記の式で表される。

$$error_i = \frac{1}{N} \sum_{i=1}^I |y_i - [\hat{y}_i]|$$

ただし、 y_i : 空き家ダミー(空き家である場合1、そうでない場合0)、 \hat{y}_i : 推定空き家確率、 $|\cdot|$: 括弧内の絶対値、 $[\cdot]$: 括弧内が0.5以上であれば1、そうでなければ0を返す関数、 N : データ数である。パラメータを変化させた際に、訓練データと検証データでクロスバリデーションを行い、検証データのエラー率が改善しなくなるまで行う。その後、エラー率の最も低い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する。

このように、XGBoostは一般の関数を想定しており、チューニングパラメータを適切に設定することで精度の高い予測値を実現している。なお、XGBoostでは欠損値の有無も決定木の分岐条件に含むことができるため、複数のデータを組み合わせることで欠損値が多くなる問題にも対応している。

令和3年度

和歌山県における空き家分布推定に関する研究成果報告書

令和4年3月

東京大学空間情報科学研究センター

特任准教授 秋山 祐樹

研究支援員 左右田敢太

研究支援員 洪 義定

〒277-8568 千葉県柏市柏の葉5-1-5 総合研究棟 4階

和歌山県データ利活用推進センター

副主査 徳富 智哉

〒640-8203 和歌山県和歌山市東蔵前丁3番17 南海和歌山市駅ビル5階
