

**令和5年度
長野県白馬村における空き家分布推定技術の開発
に関する研究実績報告書**

令和6年3月

東京都市大学建築都市デザイン学部都市工学科

秋山研究室

准教授 秋山 祐樹

修士2年 富田 健人

修士1年 水谷 昂太郎

目 次

1. 本事業の背景と目的	1
1.1. 背景	1
1.2. 本事業の目的	3
2. 現地調査	4
2.1. 現地調査の実施	4
2.2. 現地調査の結果	8
3. 本事業で使用したデータの概要	10
3.1. 自治体保有データの概要	10
3.2. データの結合処理	11
3.3. データの結合結果	12
4. 空き家分布推定モデル	14
4.1 機械学習モデル（XGBoost）	14
4.2 機械学習モデルによる推定結果	15
5. モデルによる空き家分布推定結果	17
5.1 白馬村の現地調査結果から作成した機械学習モデルによる推定結果	17
5.2 和歌山市で構築済みの機械学習モデルを用いた外挿	17
5.3 和歌山市のモデル外挿による推定結果	18
6. 本事業の成果の具体的な活用方法の提案	21
6.1 空き家確率データを活用した現地調査の負担軽減の可能性	21
6.2 水道使用量データを活用した空き家予備軍の特定	24
7. まとめ	26
謝辞	27
参考文献	28
付録	29

1. 本事業の背景と目的

1.1. 背景

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。総務省統計局の最新の住宅・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（総務省統計局，2019）、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）。

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）。また、2023 年 12 月には同法が改正され、特定空家の予備軍となる「管理不全空き家」が新たに設定された。これにより、特定空家化する前に行政による改善の指導や勧告を行うなどの予防的措置を講じられるようになるなど、空き家問題への関心は近年ますます高まりつつある（国土交通省，2023）。

しかしこれらの行動を起こすためには、まず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（益田・秋山，2020）。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、自治体が保有するデータを活用して、空き家の分布調査を迅速かつ安価に実施する手法の研究が近年行われている（秋山ほか，2021；馬場ほか，2021；水澤ほか，2022；Sayuda et al., 2022）。これらの手法の特徴は、調査対象となる自治体において、ある時点に実施された空き家分布調査結果を教師データとし、空き家分布調査結果の実施時期に近い時期に取得した自治体が保有する様々なデータ（以下「自治体保有データ」）を活用することで、建物ごとの空き家確率を推定する機械学習モデルの構築を行う点にある。また、同モデルを他の地域や時点の自治体保有データに外挿し、任意の地域・時点における空き家の空間分布を推定できる点にある。これら

の手法により、地域メッシュや町丁字等の小地域単位で空き家数や空き家率が推定可能となりつつある。なお、「自治体保有データ」には自治体外部の者でもアクセス可能なオープンデータのみならず、住民基本台帳や水道使用量情報、固定資産課税台帳、建物登記情報などの自治体内で利活用されるクローズドなデータも含む。これらのデータは日本全国の自治体が既に保有している場合が多く、また継続的に更新・管理が行われていることから、自治体にとってこれらのデータを用いた手法は、空き家分布調査業務のコスト圧縮につながることで期待される。

一方、以上の既存研究で対象としている自治体の多くは県庁所在地クラスの比較的規模の大きい都市が中心であり、中山間地域や島嶼部に分布する人口規模の小さい自治体を対象に、以上で挙げているような機械学習モデルを構築した事例は殆ど見られない。本事業で研究対象地域とする白馬村も、長野県北部に位置する人口約 9,200 人（2024 年）の小規模な自治体である。また、村の中央に南北に細長く広がる白馬盆地を除けば、村域の多くは山林で構成されている中山間地域である。また、秋山ほか（2024）の「空き家予測マップ」（<https://www.akiyama.jp/wp/>）によると、白馬村においても空き家は今後増加することが予想されており（図 1.1）、白馬村においても空き家の増加は無視できない課題となるものと予想される。こうした背景の中、白馬村では令和 5 年に「白馬村空き家等対策協議会」を設置し、令和 5 年 6 月には「空き家対策講座・相談会」を開催するなど、空き家に対する取り組みを本格的に着手した。

したがって、秋山ほか（2021）や Sayuda et al. (2022)に見られるような自治体保有データを活用して、白馬村全域の空き家分布推定を行うとともに、その推定精度を現地調査により得られた空き家分布調査結果を使って検証することで、白馬村で有用な空き家分布推定モデルの開発や有用な活用方法を検討できるものと期待される。その結果、白馬村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、様々な公的統計が利用可能になることや、和歌山市で過去に実施した類似の取り組みを参考にできること、また和歌山市で開発済みのモデルを活用することができるなど、多大なシナジーが期待される。

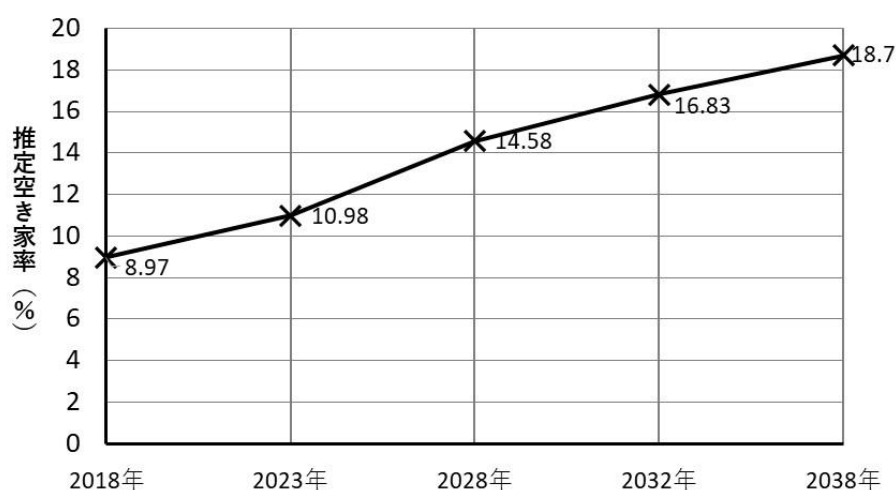


図 1.1 空き家予測マップ（<https://www.akiyama.jp/wp/>）による
白馬村の将来の推定空き家率

1.2. 本事業の目的

以上より、本事業の目的を以下の通り設定する。

1. 白馬村の自治体保有データ（住民基本台帳、水道使用量データ）を説明変数、空き家の現地調査を目的変数とする機械学習モデルを構築し、建物単位で空き家・非空き家を予測するモデルを構築する。
2. 目的1の目的変数を構築するために、白馬村全域を対象とする空き家分布を把握できるデータセットの作成を行う。それに伴い白馬村全域の戸建て住宅に対する現地調査を実施し、戸建て住宅ごとに空き家・非空き家の状態を把握する。
3. 目的1の機械学習モデルを用いて、白馬村全域の空き家分布状況を把握できるデータセットおよび地図の作成を行う。また、モデルの精度検証を行う。
4. 目的3で得られた結果の精度向上のために、他の自治体で構築されたモデルを外挿し、その効果を検証する。
5. 目的3、目的4で得られた結果の具体的な活用方法を提案する。

2. 現地調査

本事業を実施するためには、白馬村の最新の空き家分布状況を把握する必要がある。そこで、本事業では 2023 年 10 月に白馬村全域の戸建て住宅を対象とする悉皆調査を実施した。本章ではその概説および結果について報告する。

2.1. 現地調査の実施

2023 年 10 月に白馬村全域の戸建て住宅を対象に現地調査を実施した（図 2.1、図 2.2）。以下に現地調査の方法を示す。

- ・実施日時：2023 年 10 月 10 日～13 日
- ・調査対象：
白馬村全域の戸建て住宅（2,247 件：2020 年のゼンリン住宅地図により戸建て住宅と判定された建物）。ただし、原則全戸別荘であるみそら野地区は除く。図 2.3 に調査対象地域および、白馬村の地区の分布を示す。また、図 2.4 に各地区における調査対象件数を示す。
- ・調査方法：調査員による外観目視調査
- ・調査員の数：8 名（東京都市大学の教員、研究員、大学院生および学部生）
- ・調査結果の記録方法：専用の現地調査アプリケーション（図 2.5、図 2.6）への結果入力
- ・現地調査の判定基準：
表 2.1 に示す判定基準を設け、戸建て住宅ごとに表 2.1 に示す結果を付与する。なお、「更地または新築（4）」と「戸建て以外（5）」が発生する理由は、調査対象となる戸建て住宅の特定に使用した住宅地図の作成時期が 2020 年であるためである。



図 2.1 現地調査の様子（岩岳スキー場付近）



図 2.2 現地調査の様子（白馬村通の集落）

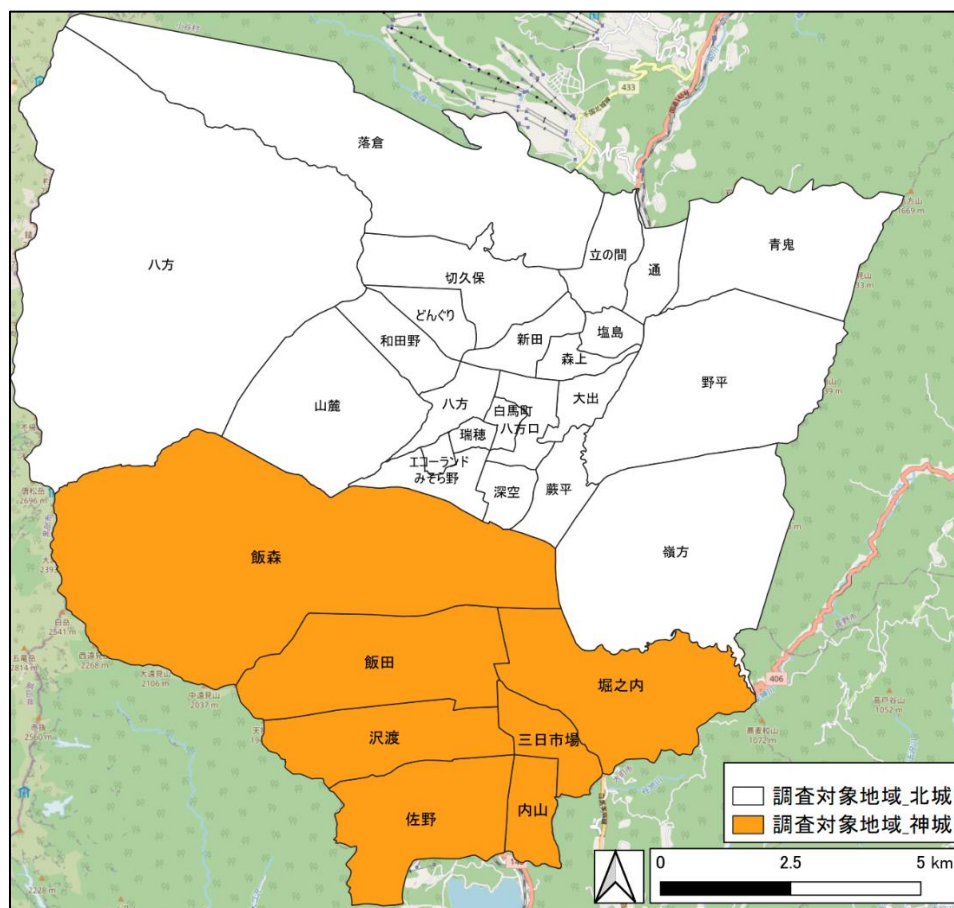


図 2.3 現地調査の対象地域

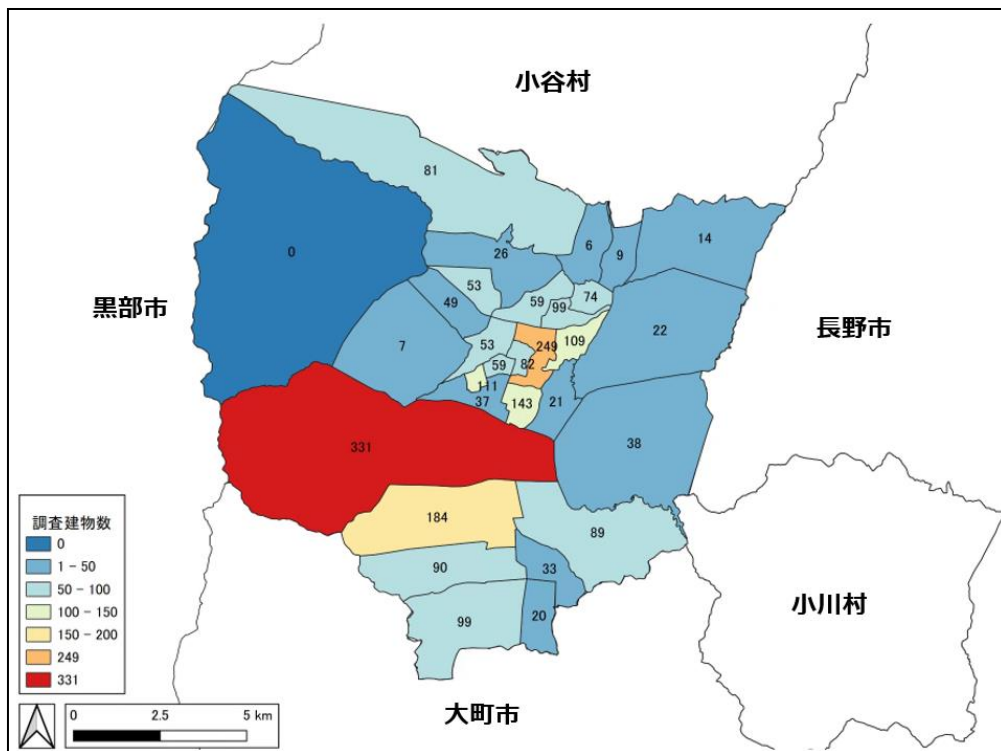


図 2.4 現地調査対象地域ごとの調査対象となる戸建て住宅の件数



図 2.5 本事業で使用した現地調査アプリケーション



図 2.6 本事業で使用した現地調査アプリケーションを使用している様子

表 2.1 本事業の現地調査の判定基準

判定基準		特徴
空き家	A(1)	状態が良く流通可能
	B(2)	流通には補修が必要
	C(3)	特定空き家(補修が困難な常態)
非空き家(0)		住民が住んでいる住宅
別荘(6)		別荘であることが明らかに分かる場合
更地または新築(4)		建物ポリゴン(2020年時点)では存在していた建物が解体され、更地になったまたは新築が建てられていた場合
戸建て以外(5)		倉庫や店舗などの建物であった場合
調査不可(99)		玄関が高台にあるなどの事情で建物を目視できなかった住宅

2.2. 現地調査の結果

表 2.2 に現地調査の結果を示す。現地調査の結果、空き家は 268 件、非空き家は 1,399 件観測された。また、別荘も 286 件観測された。仮にこれらの合計を戸建て住宅の総数とし、別荘を非空き家とカウントした場合、白馬村の最新の空き家率は約 13.7%となる（空き家：268 件 / （空き家：268 件＋非空き家：1,399 件＋別荘：286 件））。なお、268 件の空き家のうち、状態が極めて悪い特定空き家に該当する空き家は 12 件（約 4.5%）に留まり、状態が良く流通が可能な空き家が 156 件（約 58.2%）と半数以上を占めていた。

現地調査の結果を地区ごとに集計した結果を図 2.7 および、図 2.8 に示す。図 2.7 より空き家数が多い地区は主に JR 大糸線の駅を含む地区であり、特に白馬駅を含む白馬町、いい森駅を含む飯森地区、神城駅を含む飯田地区であった。一方、図 2.8 より空き家率が高い地区は飯田地区、落倉地区、嶺方地区であった。特に、飯田地区では空き家数、空き家率ともに高いことが明らかとなり、対策優先度が高い地区であると考えられる。

表 2.2 白馬村における空き家の現地調査結果

判定基準		件数
空き家	A(1)	156
	B(2)	100
	C(3)	12
非空き家(0)		1,399
別荘(6)		286
更地または新築(4)		65
戸建て以外(5)		152
調査不可(99)		77

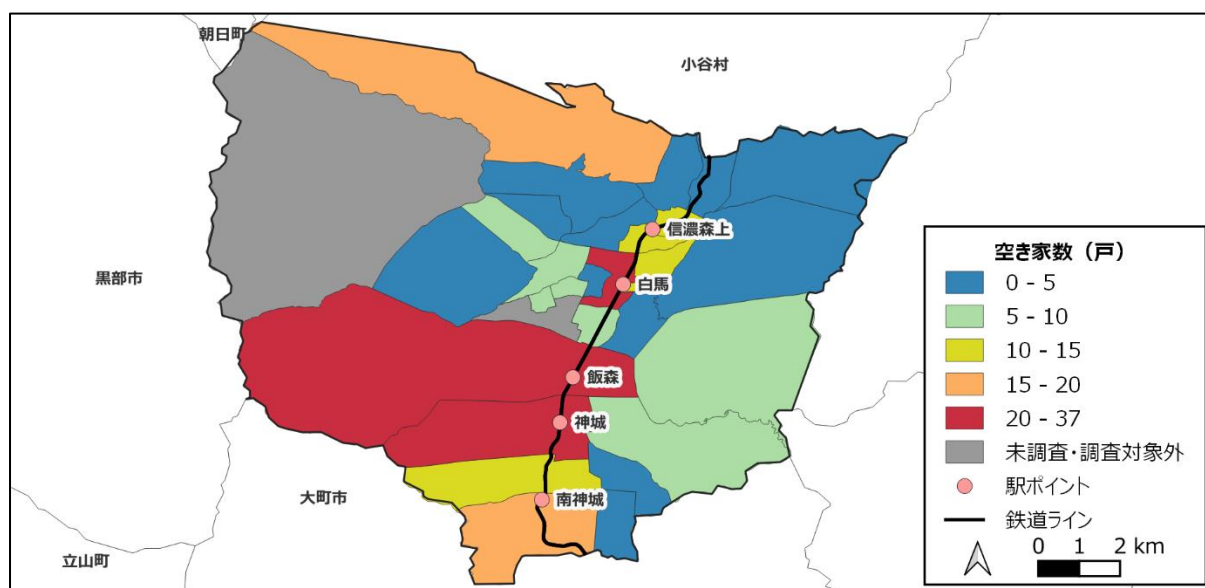


図 2.7 現地調査による小地域単位の空き家数（戸）

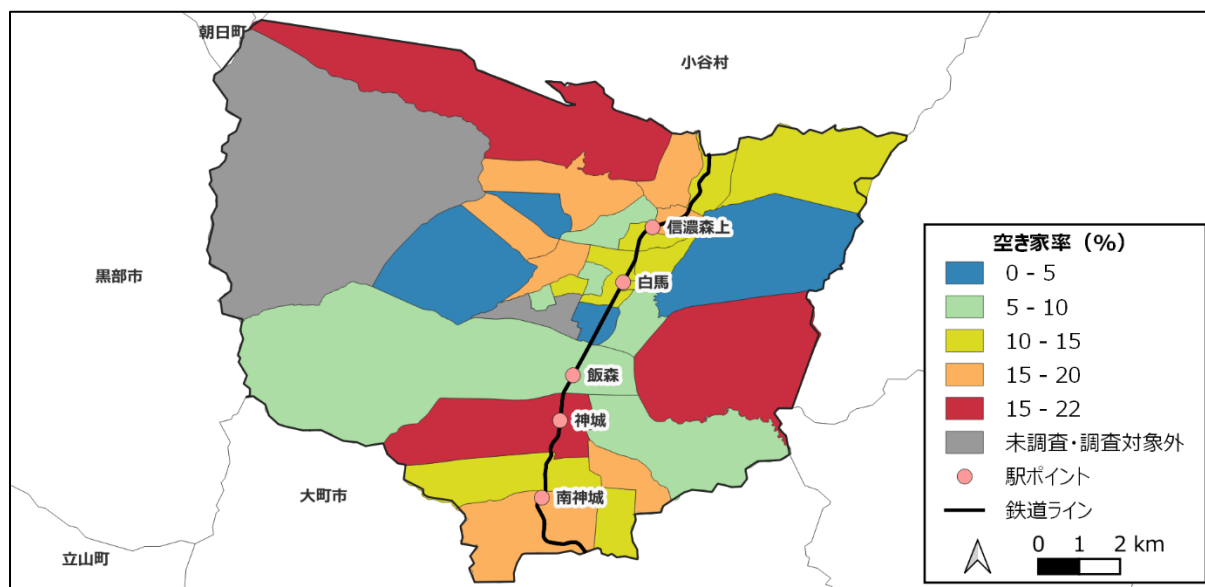


図 2.8 現地調査による小地域単位の空き家率 (%)

3. 本事業で利用したデータの概要

本事業を実施するためには、まず白馬村から提供された自治体保有データ（住民基本台帳、水道使用量データ）の変数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、自治体保有データの変数および構造について概説する。

3.1. 自治体保有データの概要

本事業で利用した白馬村から提供された自治体保有データは、住民基本台帳（以下、「住基データ」）、水道使用量情報（以下、「水道使用量データ」）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。本事業ではこれらのデータを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。

1) 住基データ（2023 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、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った。機械学習モデルに投入した住基データの変数一覧は表 3.1 の通りである。

表 3.1 本研究で利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）	「〇〇〇-〇〇〇〇 大字名××番地△△号」と記載（〇〇〇-〇〇〇〇は郵便番号）
世帯コード（V）	6 桁の世帯ごとに与えられた固有の番号
年齢（V）	居住者の年齢（歳）
性別（C）	居住者の性別（男 または 女）
世帯人数（V）	世帯コードごとの人数

2) 水道使用量データ（2022 年 4 月～2023 年 3 月現在のデータ）

水道使用量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。機械学習モデルに投入した水道使用量データの変数一覧は表 3.2 の通りである。

表 3.2 本研究で利用した水道使用量データの変数一覧

変数名	仕様等
住所（C）	「大字名〇〇番地△△号」と記載
年度ごとの最大水道使用量（V）	単位はm ³ 。 住所ごとの 2022 年度、2023 年度それぞれの 最大使用量を算出した。

3.2. 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、3.1.の自治体保有データ（住基データ、水道使用量データ）と、2章にて前述した空き家現地調査結果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 3.1 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 2020 年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。

なお、住基データと水道使用量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「アドレスマッチング（あるいはジオコーディング）」という。アドレスマッチングを行うことによって、例えば白馬村役場がある「長野県北安曇郡白馬村北城 7025 番地」という住所は、「東経 137.86207206243583、北緯 36.698604825584134」という経度緯度に変換される。

本事業では、Yahoo! Japan が提供する Yahoo! ジオコード API を活用し、住基データと水道使用量データの住所を緯度経度に変換した。なおこれらのデータに付与されている住所の表記揺れなどを理由に、アドレスマッチングの精度は大きく異なってくる。本事業では住宅地図の個々の建物ポリゴンデータ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・枝番レベル」で変換できたデータのみを使用した。

なお、図 3.2 のようにアドレスマッチングの仕様により、建物ポリゴン内に自治体保有データのポイントが落ちないケースがあった。このような場合、各ポイントの最近隣の建物ポリゴンの重心を中心に、ポリゴンの面積と同等のバッファを作成し、そのバッファ内に該当

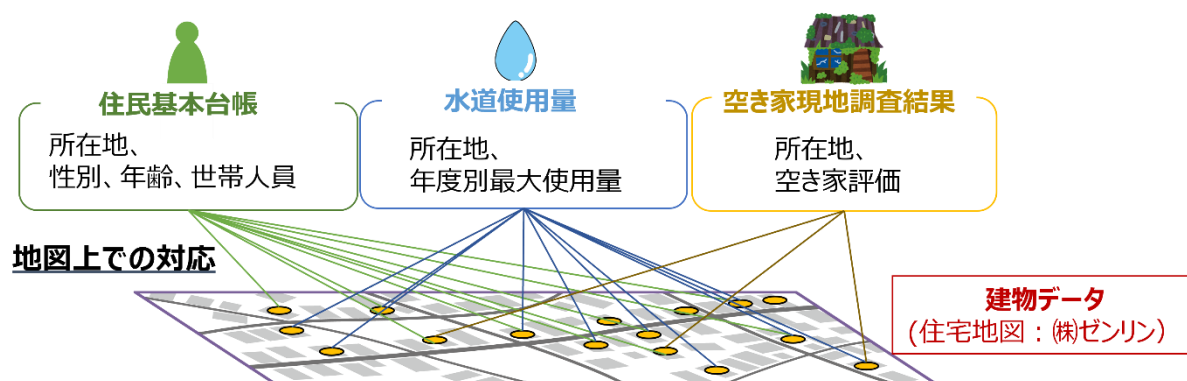


図 3.1 自治体保有データと現地調査結果を空間結合処理する流れ（白馬村の場合）

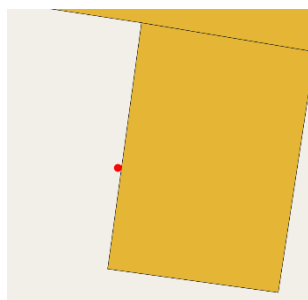


図 3.2 建物ポリゴンと自治体保有データのポイントデータが紐づかない例

ポイントが含まれ、建物データが持つ地番（住宅地図の属性値としては”chiban”と表記）と自治体データの住所表記が一致するデータを、その建物の公共データとして紐付けた。

3.3. データの結合結果

以上の手法により、自治体保有データおよび現地調査結果を住宅地図の建物ポリゴンデータに結合させた。ここでは以下の2点についてその結合結果を整理した。まず「① アドレスマッチングの結果」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。

① アドレスマッチングの結果

前述の通り住居表示の住所を持つ住基データと水道使用量データを、Yahoo!ジオコードAPIを用いて住所を緯度経度座標に変換した。表3.3に白馬村におけるジオコーディングの結果を示す。住基データは2,617件のうち、1,909件（72.9%）に、水道使用量データは4,853件のうち、2,838件（58.4%）に号レベルの緯度経度座標を付与させることができた。その結果、後述のモデル構築において十分な量のデータが得られた。一方、空き家現地調査結果はそもそも住宅地図ベースの調査であり、緯度経度の情報を得ているため、268件（100%）全件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3.4に示す。結合対象となる建物2,654件に対して、住基データは1,225件（46.2%）、水道使用量データは1,518件（57.1%）に少なくとも1件以上が結合した。また空き家現地調査結果は前述のとおり住宅地図ベースの調査であるため、268件（100%）全件が結合した。

③ 本事業の対象となる建物

表2.2に示す現地調査結果のうち、後述する機械学習モデルでは対象を戸建て住宅に限定するため、一般の住宅と使用実態の異なる別荘および戸建て以外の住宅は対象外とした。これは一般の戸建住宅とそれ以外の住宅で利用の実態が大きく異なることや、自治体が介入す

るべき管理不全が見込まれる空き家、ならびにその予備軍となる空き家の分布の把握を目的としていることを鑑みると合理的であるためである。また調査不可である建物のうち、ゼンリン住宅地図の建物名（住宅地図の属性値としては”house name”と表記）の項目に表札名等が存在する場合を非空き家、存在しない場合を空き家とした。

表 3.3 白馬村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	2,617	1,909
水道使用量データ	4,853	2,838
空き家現地調査結果	268	268

4. 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。

4.1. 機械学習モデル（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 4.1 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 m³の住宅を得た場合、図 4.1 の木構造から、空き家確率 0.23 を得る。

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

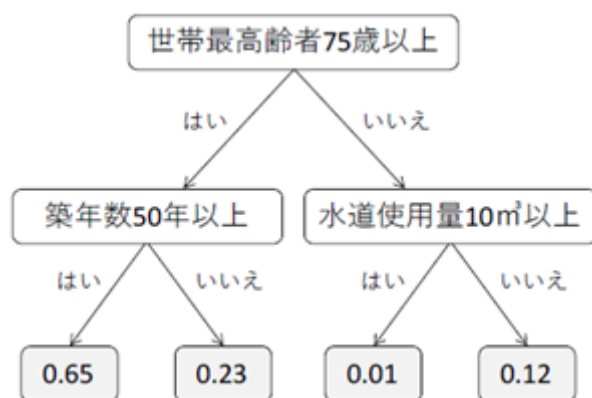


図 4.1 決定木のイメージ

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 4.2 のように表すことができる。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。

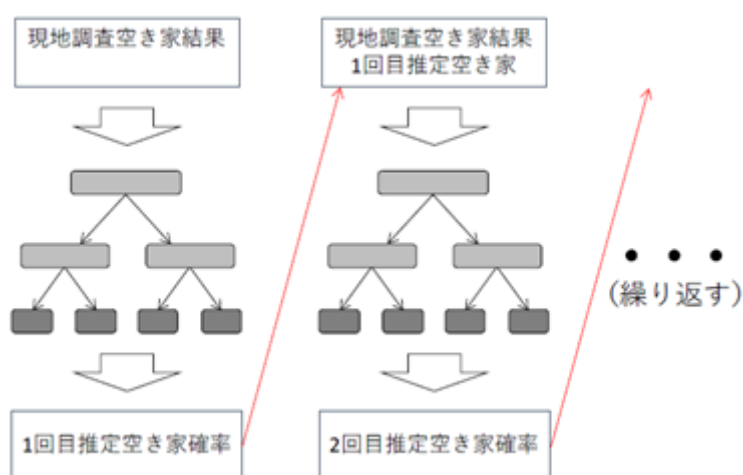


図 4.2 分析手法のイメージ

4.2. 機械学習モデルによる推定結果

以上の手法を用いて自治体保有データから得られる変数を説明変数、現地調査に基づく空き家判定を目的変数とするモデルを構築し、建物ごとの空き家確率を推定した結果を図 4.3 に示す。本事業では空き家確率が 0.5 以上の建物を空き家と定義した。すなわち図 4.3 において凡例が赤である建物が空き家、それ以外の建物が非空き家となる。

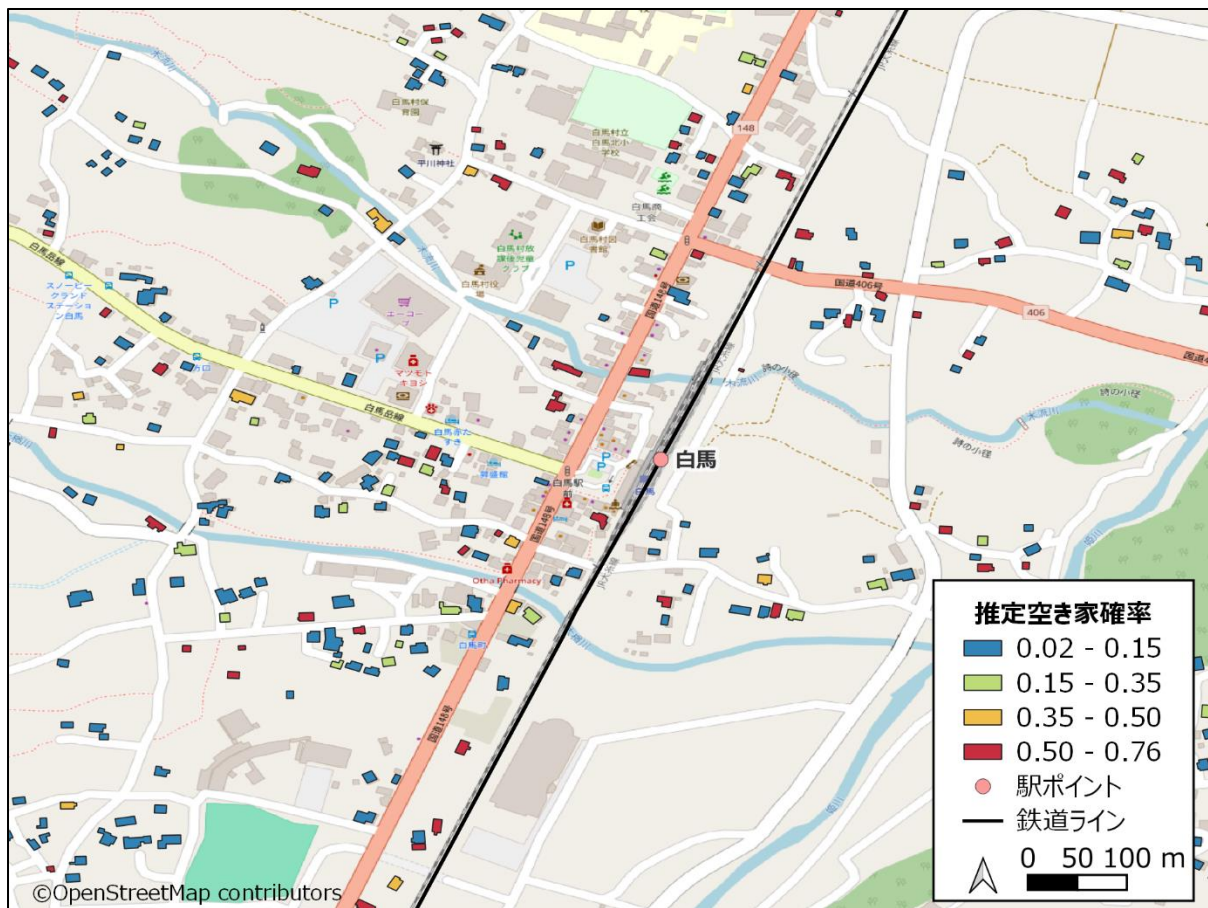


図 4.3 建物ごとの空き家確率

5. モデルによる空き家分布推定結果

本章では、まず白馬村の現地調査結果から作成した機械学習モデルの精度について評価を行う。また、教師データの件数が十分に確保されている自治体である和歌山市のモデルを外挿することで、推定精度の向上を図る。

5.1. 白馬村の現地調査結果から作成した機械学習モデルによる推定結果

表 5.1 に白馬村の現地調査結果のみから作成した機械学習モデルの推定結果を示す。全体の正解率は 82.4% となり、比較的高い精度のモデルを構築することができた。特に、真値が非空き家である建物は、非常に高い精度で予測できた（特異度：96.3%）。一方、空き家は 86 件を予測できたものの、決して高い精度とは言えない結果であった（再現率：24.6%）。ただし、白馬村は村内の空き家数が少ないため、このモデルが空き家をどの程度正しく予測できたかを評価することは難しいといえる。

表 5.1 白馬村の現地調査結果のみから作成した機械学習モデルの推定結果

		推定値	
		非空き家	空き家
真値	非空き家	(A) 1,406	(C) 54
	空き家	(D) 264	(B) 86
正解率		82.4%	
適合率		61.4%	
再現率		24.6%	
F 値		35.1%	
特異度		96.3%	

正解率(全ての推定値のうち正しく推定した割合) =

$$\frac{(A) + (B)}{(A) + (B) + (C) + (D)} = \frac{1,492}{1,810} = 0.824$$

適合率(空き家と推定した中で実際に空き家であった割合) =

$$\frac{(B)}{(B) + (C)} = \frac{54}{140} = 0.614$$

再現率(実際に空き家である中で空き家と推定できた割合) =

$$\frac{(B)}{(D) + (B)} = \frac{86}{350} = 0.246$$

特異度(実際に非空き家である中で非空き家と推定できた割合) =

$$\frac{(A)}{(A) + (C)} = \frac{1,406}{1,460} = 0.963$$

5.2 和歌山市で構築済みの機械学習モデルを用いた外挿

以上のように、白馬村では空き家数が少ないため、空き家の予測するためのモデル構築が困難であることが分かった。この課題は他の小規模自治体で同様のモデル構築を行う際にも共通して発生するものと予想される。

そこで、本事業ではこの課題の解決方法として、空き家数が十分に多い他の自治体で構築したモデルを外装する方法を試行する。本事業では同様の手法で既にモデルが構築済みである和歌山県和歌山市のモデルを白馬村の自治体保有データに外挿するものとする（図 5.1）。なお、このモデルは和歌山市における空き家分布推定において高い推定精度を発揮していることから（Baba et al., 2020; Sayuda et al., 2022; Tomita et al., 2023）、モデルとして高い性能を有していると言える。

なお、和歌山市のモデルを白馬村に外装するにあたり、以下の条件および制約があることを確認する必要がある。まず、和歌山市の現地調査は全ての戸建て住宅を対象として実施したのではなく、水道使用量が「閉栓」となっている住宅にフィルターを掛けた上で実施されている。すなわち、水道が開栓していても使用量が極めて少なく、空き家の可能性を有するような戸建て住宅だったとしても、現地調査に含まれない。そのため、白馬村の空き家に関しては必ずしも高い精度で予測できない可能性を有する。

また、白馬村の現地調査結果には空き家、非空き家以外の評価も含まれるため、これらについては以下のように扱うものとした。まず、「更地または新築(4)」と判断された建物は「空き家」と定義した。また、「調査不可(99)」と判断された建物は、その建物の住宅地図掲載の表札名を確認し、そこに居住者名の記載があれば「非空き家」、欠損（空白）であれば「空き家」と定義した。さらに、「別荘(6)」と判断された建物は、あらかじめ検証から除外した。

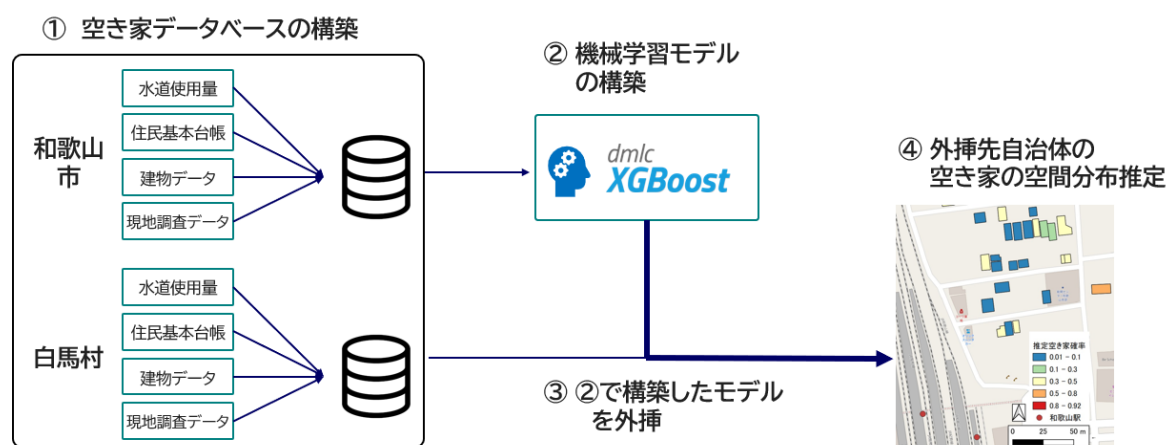


図 5.1 和歌山市で構築済みの機械学習モデルを白馬村の自治体保有データに外挿する流れ

5.3. 和歌山市のモデル外挿による推定結果

表 5.2 に和歌山市で構築済みの機械学習モデルを白馬村の自治体保有データに外挿した結果を示す。全体の正解率は 72.3% となり、白馬村の現地調査のみから構築したモデルの正解率（82.4%）よりは正解率は低下した。ただし、空き家の予測精度は 60.6%（再現率）となり、白馬村の現地調査のみから構築したモデルの再現率（24.6%）と比べると大幅な改善が達成された。一方、真値が非空き家である建物の予測精度は 75.1%（特異度）と、白馬村の現地調査のみから構築したモデルの特異度（96.3%）と比べると低下した。

以上の結果より、空き家数が十分に確保できる他の自治体で構築したモデルを外挿することで、空き家の予測精度を大幅に向上させることができることが確認できた。すなわち、白馬村のような空き家数が非常に少ない小規模な自治体においても、他の自治体で構築したモデルの外挿により、空き家の分布予測をある程度行うことができる可能性を示すことができた。

図 5.2 に白馬村全域の地区ごとの推定空き家数を、図 5.3 に白馬村全域の 500 m メッシュごとの推定空き家数を示す。地区単位でみると、村役場のある白馬地区や村南部の神代地区で空き家数が多く推定されていることが分かった。また、メッシュを用いて空間的により詳細に見ていくと、JR 大糸線沿線で空き家が多く、特に駅周辺で顕著に多くなることが分かった。集落単位では、嶺方地区で特に空き家が多くなることが分かった。

表 5.2 和歌山市で構築済みの機械学習モデルを白馬村の自治体保有データに外挿した結果

		推定値	
		非空き家	空き家
真値	非空き家	(A) 1,097	(C) 363
	空き家	(D) 138	(B) 212
正解率		72.3%	
適合率		36.9%	
再現率		60.6%	
F値		45.8%	
特異度		75.1%	

正解率(全ての推定値のうち正しく推定した割合) =

$$\frac{(A) + (B)}{(A) + (B) + (C) + (D)} = \frac{1,309}{1,810} = 0.723$$

適合率(空き家と推定した中で実際に空き家であった割合) =

$$\frac{(B)}{(B) + (C)} = \frac{212}{575} = 0.369$$

再現率(実際に空き家である中で空き家と推定できた割合) =

$$\frac{(B)}{(D) + (B)} = \frac{212}{350} = 0.606$$

特異度(実際に非空き家である中で非空き家と推定できた割合) =

$$\frac{(A)}{(A) + (C)} = \frac{1,097}{1,460} = 0.751$$

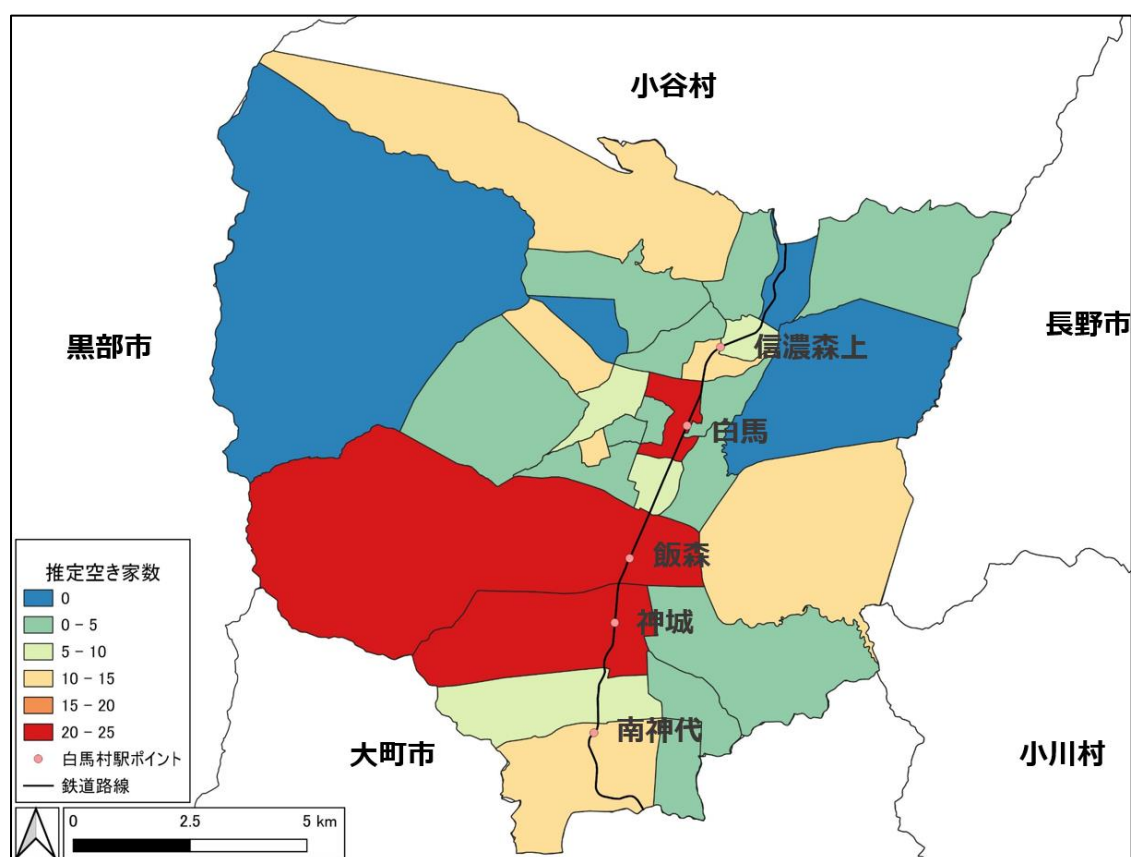


図 5.2 白馬村における推定空き家数（地区単位）

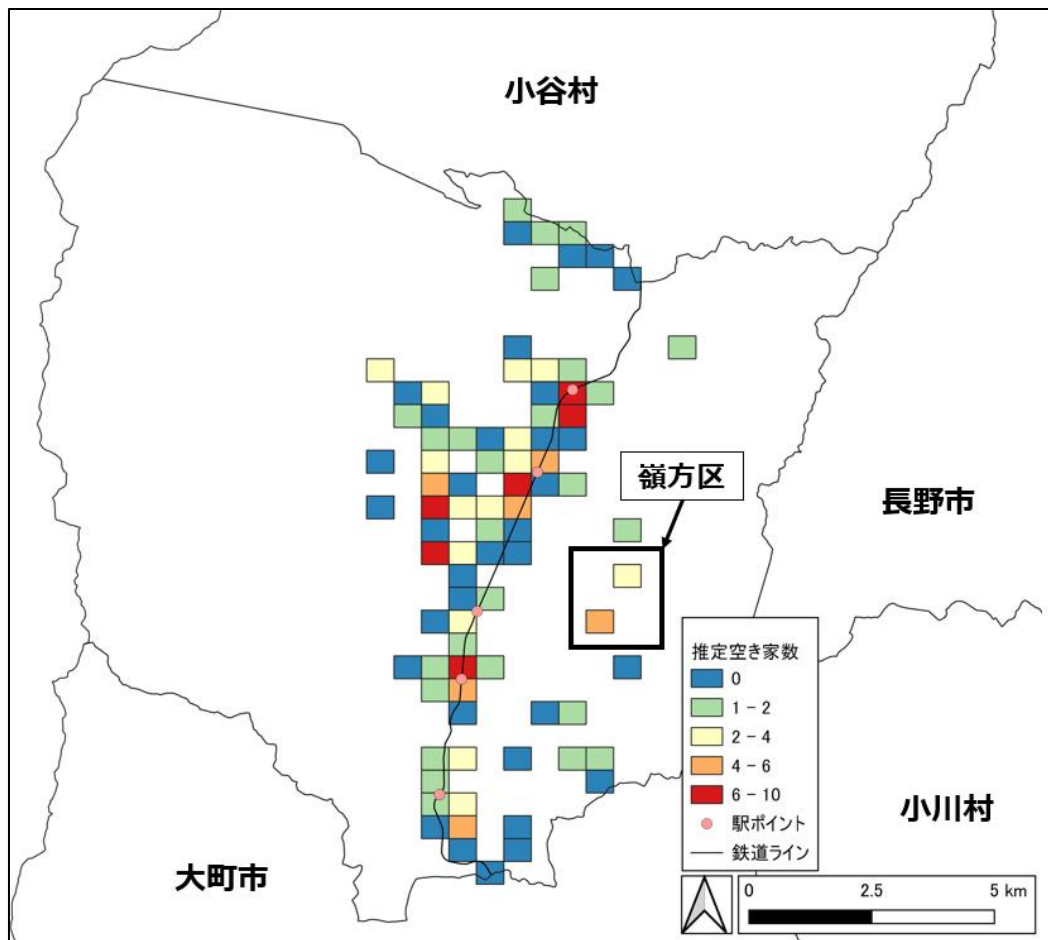


図 5.3 白马村における推定空き家数（500 m メッシュ単位）

6. 本事業の成果の具体的な活用方法の提案

最後に本事業で得られた成果の具体的な活用方法について提案する。本報告書では「空き家確率データを活用した現地調査の負担軽減の可能性」および、「水道使用量データを活用した空き家予備軍の特定」の2つの具体的な活用方法を提案する。

6.1. 空き家確率データを活用した現地調査の負担軽減の可能性

5.1. で得られた白馬村における現地調査結果に基づくモデルの推定結果の特異度は 96.3% に達した。この結果は、このモデルで得られた結果から、非空き家を非常に高精度で予測できることを意味している。この結果を用いることで、空き家の可能性が低い建物、すなわち非空き家の可能性が極めて高い建物を把握できるようになる。その結果、現地調査を実施する前に予め現地調査を行う必要が無い非空き家の特定が可能となる。これは、現地調査時に調査対象となる建物数を大幅に減らし、現地調査の負担を大幅に軽減することが可能になることを意味している。例えば図 6.1 に示すように、これまでは対象地域に存在する全ての建物を悉皆調査する必要があったが（図 6.1 左）、このモデルを用いて予め非空き家と予測された建物を現地調査の対象から外すことで、調査対象となる建物数を大幅に縮減することが可能になる（図 6.1 右）。

5.1. で得られた白馬村における現地調査結果に基づくモデルの場合、空き家確率が 10% 以下となった建物において、真値が非空き家の割合は 96.2% である。そこで、空き家確率が 10% 以下の建物を全て非空き家とした場合、その 96.2% は真に非空き家となる。その結果、推定対象となる 1,810 棟の戸建て住宅のうち、45.3% の建物（820 棟）はほぼ確実に非空き家であるため調査を行う必要がなくなり、現地調査を行うべき建物数を大きく縮減することに繋がる。

そして、非空き家と判定する空き家確率を大きくすることにより、予め調査対象から外すことができる戸建て住宅数を増やすことが可能である（図 6.2・表 6.1）。例えば、非空き家と判定する空き家確率を 30% とした場合、それらが真に非空き家である確率は 91.2% に低下するものの、予め 61.8% の建物（1,118 棟）を調査対象から外すことが可能になる。なお、自治体において同手法を用いて現地調査の負担軽減を検討する際には、この閾値を自治体の事情に応じて設定することで、適切な負担軽減のあり方を検討する必要がある。



図 6.1 空き家確率データを活用し予め現地調査の調査対象を絞り込むイメージ

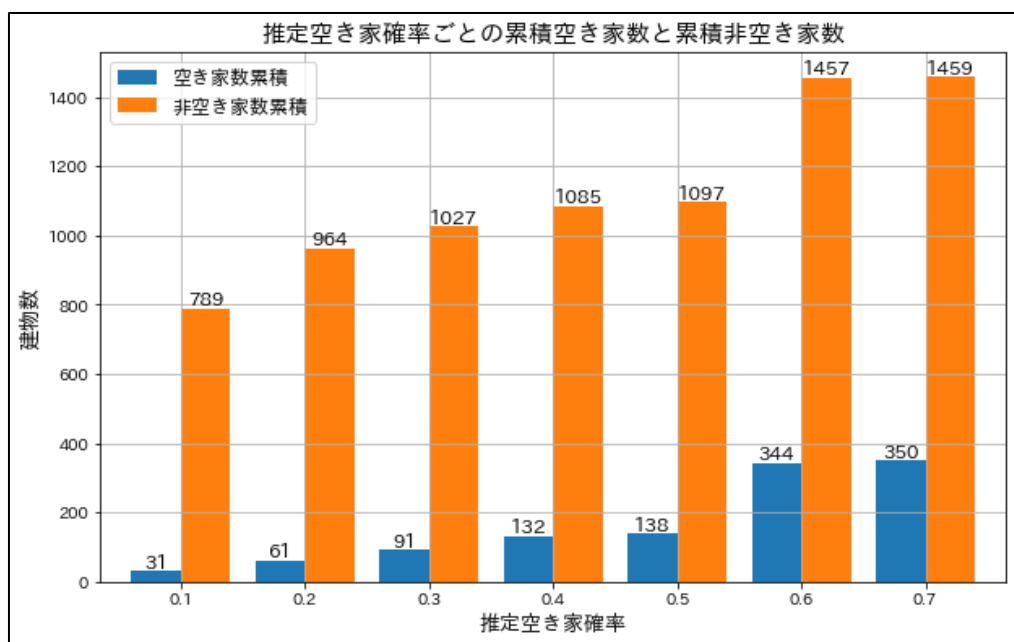


図 6.2 白馬村モデルにおける推定空き家確率ごとの累積空き家数と累積非空き家数

表 6.1 白馬村モデルにおける推定空き家確率ごとの累積空き家数、累積非空き家数、真値非空き家率および推定済み棟数率

空き家確率	空き家数累積 [棟]	非空き家数累積 [棟]	真値非空き家率 [%]	推定済み棟数率 [%]
0.1	31	789	96.22	45.30
0.2	61	964	94.05	56.63
0.3	91	1027	91.86	61.77
0.4	132	1085	89.15	67.24
0.5	138	1097	88.83	68.23
0.6	344	1457	80.90	99.50
0.7	350	1460	80.66	100

以上の結果を用いて、白馬村における現地調査候補マップを作成した。ここでは閾値を空き家確率 30%（真値非空き家率 91.9%）と設定した。この条件の場合、1,810 棟中 692 棟 (38.2%) を調査すれば、ほぼ全ての空き家を見落とすことなく調査できるようになる。また、図 6.3 に白馬村役場周辺における建物単位の現地調査候補マップを示す。地図内の赤色の建物のみを調査すれば良いことを意味している。さらに、図 6.4 に地区ごとの空き家確率 30% 以下、すなわち非空き家である建物の割合を示す。この結果から非空き家の割合が少ない地域、すなわち空き家の割合が高い地域（青色や緑色の地域）の可視化も可能になり、これらの地域は空き家の動向を注視する必要があると言えよう。

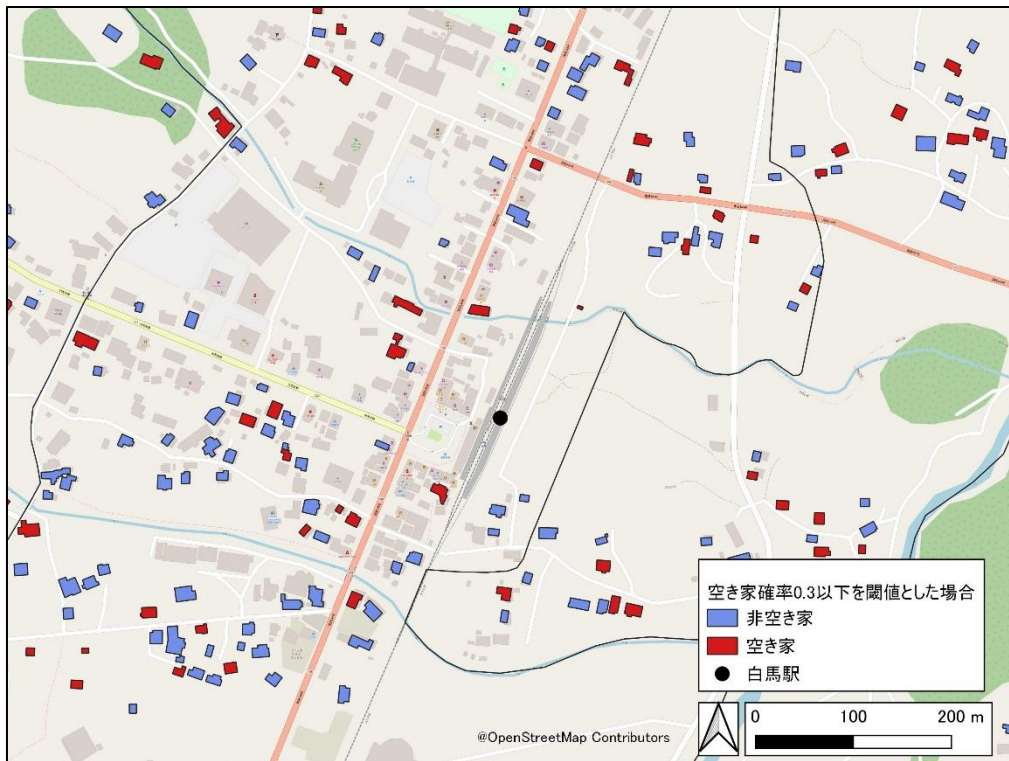


図 6.3 現地調査候補マップ（白马村役場周辺）

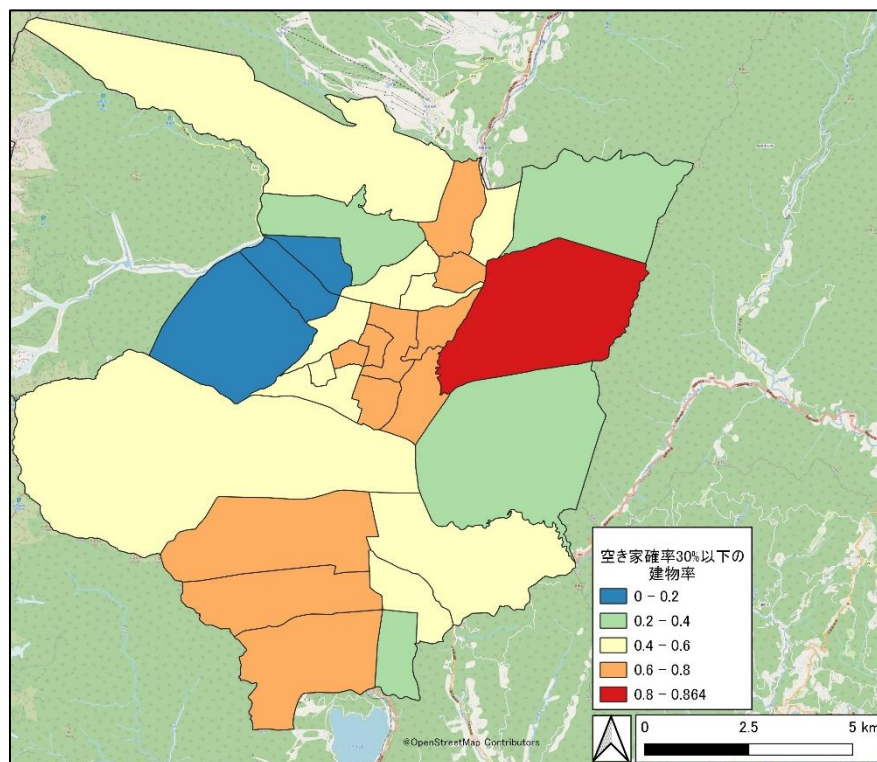


図 6.4 白马村における地区ごとの空き家確率 30%以下（非空き家）建物の割合

6.2. 水道使用量データを活用した空き家予備軍の特定

本事業では水道使用量の時系列データを利用している。水道使用量の時系列データを分析する過程で、今後空き家となる可能性を有する建物（以下「空き家予備軍」）を特定できる可能性が明らかになったため、本節で報告する。

まず、空き家予備軍を特定することの意義について概説する。空き家は長期間放置されると、建物の劣化や倒壊、防犯上の問題などのリスクが高まる。国土交通省が実施した令和元年の空き家所有者実態調査（国土交通省，2020）によると、空き家の約 20.9%が 20 年以上放置されていると報告されており、これが所有者不明の空き家を生む原因にも繋がっている。そのため、もし早期に空き家予備軍を発見し、所有者に適切な管理を促すことができれば、管理不全となる空き家の増加を防ぎ、良好な住環境を維持することにつながるものと期待される。

また、水道使用量データは自治体により検針方法が異なるが、白馬村の場合は毎月検針員が各水道栓の使用量を計測している。すなわち建物の使用実態を経時的に把握できるため、空き家候補の特定に有用である。

そこで、白馬村における①非空き家（図 6.5）、②別荘（図 6.6）、そして③空き家になった可能性のある建物（図 6.7）の水道使用量を分析することで、以下の特徴が明らかになった。

- ① 非空き家は、年間を通して一定の水道使用量が見られた（図 6.5）。
- ② 別荘は、11 月から 3 月までは使用量がない一方、4 月に顕著に大きい使用量が見られ、夏季は一定の使用量となるパターンが見られた（図 6.6）。これは、冬季は積雪の影響によって検針ができず、4 月の検針で冬季の使用量が合算されて検針されるためである。
- ③ 空き家になった可能性のある建物では、2021 年上半期までと下半期以降で水道使用量のトレンドが大きく異なっていた（図 6.7）。実際に当該の建物を調査したところ、この建物は数年前まで貸別荘として利用されていたが、現在は廃業していることが判明した。

以上を踏まえて、水道使用量の時系列的な変化を分析することで、水道使用量の異常検知モデルを構築できるものと期待される。その結果、以下のような空き家の自動検知ができる可能性がある。

- ・年間の使用量が一定以下の住宅を「空き家の可能性あり」として検知
- ・一定期間以上水道の使用がない住宅を「管理不全の空き家の可能性あり」として検知

なお、他分野でも時系列データから異常検知を行う仕組みの実装が進んでいる。例えば、小堀ほか（2022）は漏水量や変形等の各種計測情報の時系列データから、ダムの異常有無を検知するモデルを構築している。また、山口ほか（2023）は患者のバイタルデータの時系列データから患者の異常を検知するバイタルサイン異常検知機能を開発し、実際の病院で導入されつつある。このように本技術は実装レベルにまで至っている技術であることから、実現可能性は十分に高いものと期待される。本モデルの構築により、空き家になった可能性のある建物の早期発見や、住基データや家屋課税台帳等からのみでは判断が難しい、所有者は

いるものの数年間使用実態のない可能性がある住宅を容易に特定し、その状況を網羅的に把握することが可能になるものと期待できる。

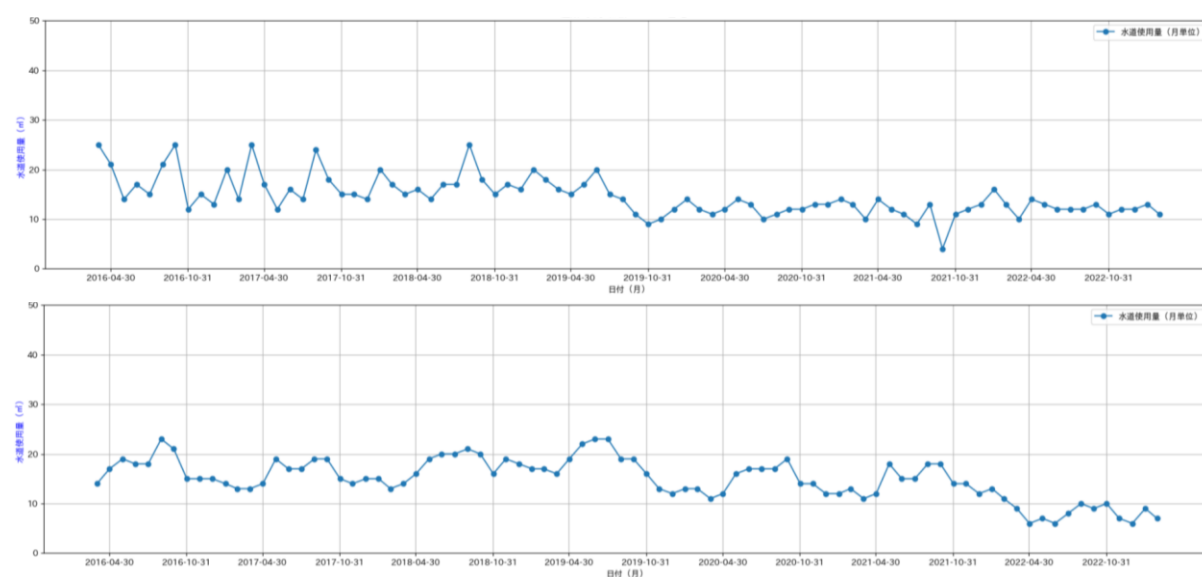


図 6.5 非空き家における水道使用量の時系列変化の例

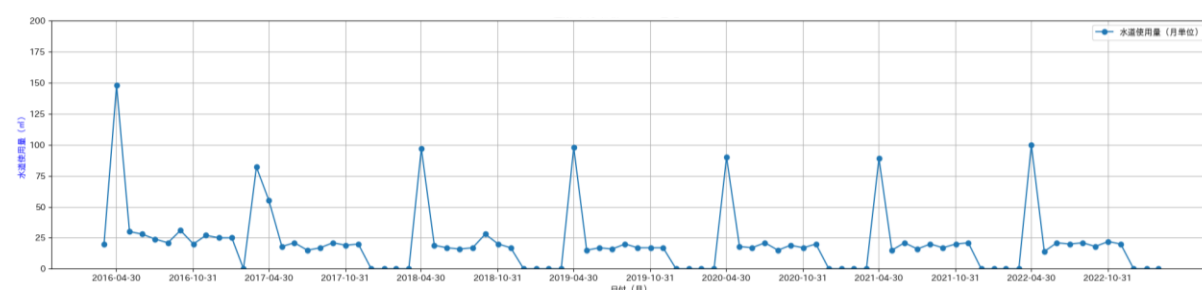


図 6.6 別荘と考えられる建物における水道使用量の時系列変化の例

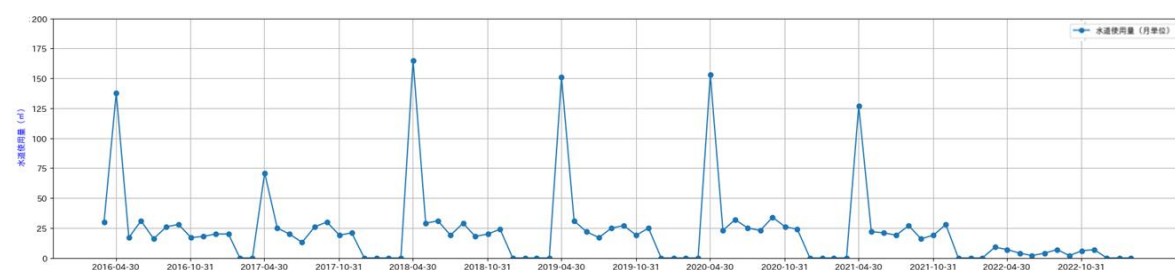


図 6.7 別荘から空き家に変化したと考えられる建物における水道使用量の時系列変化の例

7. まとめ

本事業では長野県白馬村を対象に統計データ利活用推進事業として、自治体保有データ（住民基本台帳、水道使用量データ）を活用した空き家分布推定を実施した。具体的には、以下の内容を実施することができた。

1. 白馬村の自治体保有データを説明変数、空き家の現地調査を目的変数とする機械学習モデルを構築し、建物単位で空き家・非空き家を予測するモデルを構築した。
2. 機械学習モデルの目的変数を構築するために、白馬村全域を対象とする空き家分布を把握できるデータセットの作成を行った。それを達成するために、白馬村全域の戸建て住宅に対する現地調査を実施し、戸建て住宅ごとに空き家・非空き家の状態を把握した。
3. 機械学習モデルを用いて、白馬村全域の空き家分布状況を把握できるデータセットおよび地図の作成を行った。また、モデルの精度検証を行った。
4. モデルの精度向上のために、他の自治体で構築されたモデルを外挿し、その効果を検証した。
5. 本事業で得られた結果の具体的な活用方法として、「空き家確率データを活用した現地調査の負担軽減の可能性」および、「水道使用量データを活用した空き家予備軍の特定」の2つの具体的な活用方法を提案した。

以上の結果より、既存研究で対象としている人口規模の比較的大きい都市ではなく、中山間地域や島嶼部に分布する人口規模の小さい自治体においても、空き家分布推定を行うための機械学習モデルを構築することができることを示すことができた。また、人口規模が大きく地理的条件が大きく異なる他自治体（和歌山市）のモデルを外挿することで、空き家の予測精度を大幅に向上させることができることも確認することができた。さらに、これらの結果は、同手法を全国の様々な自治体に広く普及させる上でも大きな進歩となったといえるだろう。

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、日本全国の他の自治体においても引き続き同様のアプローチを展開していくことで、本手法の有用性や課題、限界などを明らかにしていきたい。また、白馬村における空き家対策に引き続き貢献できる成果を得られるよう、白馬村と協議を続けながら研究を続けていきたいと考えている。

謝辞

本事業で使用した住宅地図（2020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、白馬村から自治体保有データとして住民基本台帳および水道使用量データを提供いただくとともに、事業全体を通して多大なるご支援を頂いた。さらに、本事業の推進においては総務省統計局 統計データ利活用センターおよび、長野県企画振興部総合政策課統計室からも様々なご支援を頂いた。以上、ここに記して謝意を表したい。

参考文献

1. 秋山祐樹・馬場弘樹・大野佳哉・高岡英生（2021）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」, 86（786）, 2136-2146.
2. 秋山祐樹・水谷昂太郎（2024）政府統計を活用した将来の空き家分布予測手法の開発と空き家予測マップの公開, 日本地理学会発表用要旨集, 105, 292.
3. 浅見泰司（2014）「都市の空閑地・空き家を考える」, プロGRESS.
4. 国土交通省住宅局（2020）令和元年空き家所有者実態調査 報告書. <<https://www.mlit.go.jp/report/press/content/001377049.pdf>>（最終閲覧日：2024/03/26）
5. 国土交通省（2023）空家等対策の推進に関する特別措置法の一部を改正する法律（令和5年法律第50号）について. <https://www.mlit.go.jp/jutakukentiku/house/jutakukentiku_house_tk3_000138.html>（最終閲覧日：2024/03/26）
6. 小堀俊秀・二階堂良平・松下智祥・金銅将史（2022）ダムの安全管理用計測データに基づく異常検知へのAI活用に向けた試行検討. インフラメンテナンス実践研究論文集, 1（1）, 363-371.
7. 高崎経済大学地域科学研究所（2019）「空き家問題の背景と対策」, 日本経済評論社.
8. 馬場弘樹・秋山祐樹・谷内田修（2021）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）」, 77（2）, 62-71, 2021.
9. 益田理広・秋山祐樹（2020）日本国内における近年の空き家研究の動向. 「地理空間」, 13（1）, 1-26, 2020.
10. 水澤克哉・宮本慧・田村将太・田中貴宏（2022）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-. 「都市計画論文集」, 57（3）, 1377-1384.
11. 山口庸子・内野滋彦・齋藤慎二郎（2023）集中治療患者情報管理システムのバイタルサイン異常検知機能導入の効果：前後比較研究. 日本集中治療医学会雑誌, 30（6）, 459-460.
12. https://www.jstage.jst.go.jp/article/jsicm/30/6/30_30_459/_article/-char/ja
13. Baba, H., Akiyama, Y., Tokudomi, T., and Takahashi, Y. (2020) 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.
14. 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.
15. Sayuda, K., Hong, E., Akiyama, Y., Baba, H., Tokudomi, T. and Akatani, T. (2022) Accuracy of Vacant Housing Detection Models: An Empirical Evaluation Using Municipal and National Census Datasets. *Transactions in GIS*, 26(7), 3003-3027.
16. Tomita, K. and Akiyama, Y. (2023) Development of an extrapolation model for estimation of vacant house distribution using municipally owned data and machine learning, *ACRS2023 Conference Proceedings*, ACRS2023213.

付録

XGBoost による推定値の算出方法

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 $\sum_{i=1}^I l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i))$ 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、

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

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。 $\sum_{i=1}^I l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i))$ 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5年度

長野県白馬村における

空き家分布推定技術の開発に関する研究実績報告書

令和6年3月

東京都市大学建築都市デザイン学部都市工学科 秋山研究室

准教授 秋山 祐樹

修士2年 富田 健人

修士1年 水谷 昂太郎

〒158-8557 東京都世田谷区玉堤1丁目2-8-1

東京都市大学世田谷キャンパス 10号館3階317号室 秋山研究室
