教師なし学習を活用した 空き家分布推定モデルの開発

阿部　真大[1]・水谷　昂太郎[2]・冨田　健人[3]・秋山　祐樹[4]

[1]学生会員　東京都市大学　建築都市デザイン学部都市工学科（〒158-8557 東京都世田谷区玉堤 1-28-1）
E-mail:g2032008@tcu.ac.jp

[2]学生会員　東京都市大学　総合理工学研究科建築・都市専攻（〒158-8557 東京都世田谷区玉堤 1-28-1）
E-mail: g2381641@tcu.ac.jp

[3]学生会員　東京都市大学　総合理工学研究科建築・都市専攻（〒158-8557 東京都世田谷区玉堤 1-28-1）
E-mail: g2281640@tcu.ac.jp

[4]正会員　東京都市大学准教授　建築都市デザイン学部都市工学科（〒158-8557 東京都世田谷区玉堤 1-28-1）
E-mail: akiyamay@tcu.ac.jp

日本では全国的に空き家が増加傾向にある．そこで日本全国の自治体には，空き家の分布状況を把握した上で，空き家への早急な対応と中長期的な空き家対策計画の策定が求められている．しかし，これまで空き家の分布調査の方法は現地調査に依存していたため，調査に多大な時間と費用を要するという課題を有していた．そこで，本研究では自治体保有データ（水道使用量，住民基本台帳）などを，教師なし学習のクラスタリングを用いて分析することで，空き家分布予測モデルを構築した．本モデルの適用によって研究対象地域である和歌山市全域において，調査対象となる住宅の約 52%を含み，その中に全空き家の77.5%を含むクラスターを抽出することができた．この結果は，自治体における空き家調査の対象建物数を削減し，空き家把握の効率化や空き家対策計画の検討に貢献することが期待される．

Key Words: *vacant house, distribution prediction , unsupervised learning , clustering , k-means++*

1. はじめに

日本では少子高齢化や産業構造の変化により全国で空き家が増加し続けている．2018年の住宅・土地統計調査によると，空き家数は全国で約 850 万戸，空き家率は13.6%と過去最多となっている[1]．同統計では，空き家をその状態に応じて「①二次的住宅（別荘，その他（たまに寝泊まりする人がいる住宅））」，「②賃貸用又は売却用の住宅」，「③その他の住宅（上記以外の人が住んでいない住宅）」の 3 種類に分類されている．このうち特に問題となりうるのが「③その他の住宅」である．これは①，②に該当する空き家に比べて，比較的利用や流通の機会が少なく，放置されることで管理が十分に行き届かなくなる可能性が高いためである[2]．管理が不十分な空き家は，空き家の周辺環境や景観の悪化，地震や台風などの自然災害発生時の倒壊事故，強風等による落下・飛散事故の発生，放火や不審者侵入による防犯性の低下など，様々な問題を引き起こすとされている[3]．また秋山・柴崎（2015）の推計によれば 2013 年から 2040 年までの間に「③その他の住宅」数は 318 万戸から 617 万戸と倍増すると予測されている[4]．これらの結果からも当該住宅の早急な対策が必要であることは明白である．

そこで，空き家の適正な管理の必要性の高まりを受けて，2015 年 5 月より「空家等対策の推進に関する特別措置法」（以下「空き家対策特措法」）が施行された．これにより，自治体の調査に基づいて管理が不十分と判定された空き家を「特定空き家」に指定し，所有者への指導や改善を促すことが可能となっただけでなく，自治体内の空き家の空間分布を悉皆的に調査し，その結果を管理・更新することが自治体の努力義務と定められた[5]．

また，国土交通省は 2022 年 10 月に「空き家対策小委員会[5]」を設置し，空き家の発生抑制や活用促進，適切な管理・除去のさらなる推進，NPO 等の民間主体の活動やコミュニティ活動の促進に向けて活動している．

以上のように，近年空き家問題への対応が積極的に行われており，その重要性が年々高まりつつある．

(1) 空き家分布調査の課題と既存研究における取組

自治体は空き家の分布調査の結果に基づいて，空き家に対して必要な対策を迅速かつ効率的に行うことが求められている．これを実現するためには，自治体内における空き家の空間分布の状況を効率的に把握する必要がある．多くの自治体では地域住民を対象とした聞き取り調査，アンケート調査，そして現地調査による戸別目視などによって空き家の分布状況を把握している[6]．

まず，聞き取り調査はどの自治体でも実施可能な手法である．しかし，多大な所要期間と聞き取りに対する協力者が必要であること，回答者の発言を客観的に解釈する必要があること，回答者に不利な情報が隠匿される可能性があることなどの課題を有する．そのため，聞き取り調査は他の手法で得られた結果を補足する手段として用いられる場合が多く，広域調査には不向きであると言える．

次に，アンケート調査は聞き取り調査や現地調査に比べて省力化が可能であり，また所要期間の短縮も期待できる．しかし，設問が予め固定されているため，空き家に関する詳密な情報を得ることには不向きである．また，聞き取り調査と同様に隠匿と選択の問題が存在するため，やはり他手法との併用が多いと言える．

一方，現地調査は聞き取り調査と同様にいかなる地域でも採用可能であり，調査員が建物毎の状態を目視観察するため，他の手法と比べて最も信頼性の高い手法であると言える．ただし，調査に要する時間や人員が他手法に比して大きいため，広域を対象とした悉皆的な調査に用いることは容易ではない [7]．しかし，空き家対策特措法では悉皆的な調査が努力義務と定められており，多くの自治体が多大な時間と費用をかけて調査員の外観目視による悉皆調査を採用しているのが実情である．

このように，空き家の空間分布を把握することは，各自治体が空き家対策を進めていく上で重要であるが，それと同時にその把握業務が自治体にとって極めて大きな負担になっていると言える．

そこで，近年では様々な統計情報や空間情報を利用して，空き家分布状況を把握する手法に関する研究が進展している．石河ら（2017）は，国勢調査を用いて空き家の分布状況を把握しようとした．同研究は，日本全国を対象に国勢調査の小地域集計に基づき，世帯数と建物ポイントデータの住戸数を組み合わせて町丁・字単位の空き家分布推定手法を提案した．ただし，世帯数データが推計値であることやデータ作成年度が異なるため，空き家率が負の値を取る場合もあり，大きな誤差が生じる課題があった [7]．また，山下ら（2015）による水道の閉栓データを用いた空き家分布把握では，栃木県宇都宮市を対象に市域を 16 に区分し，過去の水道の閉栓・休止情報を用いて地域ごとに 31 年分の空き家率を推定し，回帰分析を行うことでその経年変化の性質を明らかにしている [8]．さらに，馬場ほか（2021）は群馬県前橋市の中心市街地を対象に，過年度の住民基本台帳等のデータから将来の空き家分布推定モデルを構築し，将来の空き家予測確率地図を作成している [9]．しかし，同研究も含めて数多くの既存研究では，調査対象地域が限定的な場合が多い．

また，水道の閉栓・休止データだけでなく，水道の使用量やその他の自治体が保有するデータ（住民基本台帳など），さらに各自治体における過去の空き家の分布調査の結果などを用いることで，空き家の分布把握を迅速，安価に実施する手法の研究も行われている [2), 9), 10), 11)]．これらの手法の特徴は調査対象となる自治体で，ある時点において実施された空き家分布調査結果を教師データとして活用していること，空き家調査と可能な限り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を構築する点などが挙げられる．これまで研究では地域メッシュや町字単位などの単位で空き家数や空き家率を推定していたが，これらの手法では個々の住宅に空き家，非空き家の判定を与えることが可能となった．このような結果は，著者らの自治体の空き家担当者への複数のヒアリング調査からも，空き家対策業務の効率化に大いに寄与できるという意見を得ている．特に，自治体においてある特定の地域を対象に空き家対策を行う場合，小地域単位の推定結果の有無に関わらず現地調査を実施する場合がほとんどであるが，その際に予め調査が不要である非空き家の分布を特定できれば，現地調査の負担を大幅に軽減できる．これは，空き家対策特措法によって各自治体の努力義務とされている空き家の悉皆調査においても，その調査の大幅な効率化に貢献できるものである．

しかし，以上の方法を用いて空き家および非空き家の分布を予測した研究では，教師あり学習を用いて構築したモデルの再現率が低く，多くの空き家がモデルの空き家予測から除外されるという課題があった [11)]．

教師あり学習の主要な評価指標として再現率と適合率が挙げられる．再現率とは正解をどの程度見つけられるかという指標で，再現率が高いと取りこぼしなく正解データを集めることが出来る一方で，誤りのデータを集めることも多くなる．適合率とはどの程度正解できているかを示す指標で，適合率が高いと誤ることなく正解データを集めることが出来る一方で，取りこぼしが多くなる．一般に再現率が高ければ適合率は低くなり，反対に適合率が高ければ再現率は低くなる．

既往研究 [11)]では，空き家数に比べて非空き家数が圧倒的に多いという特徴や，閾値の設定等の理由から再現率が低下した．さらに，調査員の主観による判断がノイズとなりモデルの学習に影響を与える可能性が考えられる．しかし，現地調査結果を教師データとして用いる教師あり学習モデルに依存する限り，これらの問題を解決することは容易ではない．

そこで，現地調査の結果を用いることなく，空き家を予測することができるモデルを構築することができれば，以上の課題を解決することができる可能性がある．

(2) 本研究の目的

Sayuda et al.（2020）や Tomita et al.（2021）の予測モデルでは学習時の特徴量重要度の上位 10 項目に，共通して水道使用量や世帯人員が含まれていた [11),12)]．また，水谷ら（2022）においても，学習時の特徴量の寄与度を示す SHAP によって，推計結果は水道使用量が少ないほど，また世帯人員が少ないほど空き家の傾向にあることが示されてる [3)]．このように空き家となる建物の特徴は以上に挙げた既存研究のいずれのモデルでも共通している．これは益田・秋山（2020）[6)]でも明らかにされている空き家である住宅が有する特徴にも当てはまっていた．すなわち，現地調査の結果に依存すること無く，空き家の傾向を有する建物を抽出できる可能性がある．

そこで本研究では，自治体による空き家の現地調査の結果を用いること無く，水道使用量や住民基本台帳などの自治体保有データおよび，住宅の立地環境を説明するオープンデータから得られる個々の住宅の情報を説明変数とする教師なし学習（クラスタリング）により，空き家の特徴を有する住宅を抽出する手法を開発し，空き家分布推定に活用できる可能性を検討する．これにより，既往研究の課題であった空き家の現地調査の結果における調査員による恣意性を排除できるだけではなく，空き家の悉皆調査が未実施の自治体においても，空き家の可能性が高い住宅を抽出することが可能になる．

(3) 研究対象地域と対象とする建物

本研究の研究対象地域は和歌山県和歌山市全域とする．和歌山市は和歌山県の人口の約 40%を占める都市であり，和歌山県の中心都市となっている．一方で，1985 年をピークに人口が減少に転じ，近年では毎年 2000 人が減少し，都心部も人口減少と空き家問題に悩まされている現状にある．

また，本研究で対象とする建物は，戸建て住宅とした．これは戸建て住宅と集合住宅における空き家発生のメカニズムは，様々な点において違いがあることが指摘されているためである [13)]．また，前述した空き家による外部不経済の解消や，特定空き家およびその予備軍となり得る空き家の分布把握という需要を考慮すると，戸建て住宅を対象とする方が合理的であるためである．

2. データベースの構築

本研究では後述する官民が保有するデータと，国土数値情報の生活利便施設の立地データを使用した．なお，官民保有のデータは個々の建物の情報である一方，国土数値情報などのオープンデータは町丁字等の小地域や地域メッシュなどで集計された，建物周辺の状況（地域特性）を示すデータである．そこで，これらの情報を GIS（地理情報システム）によって空間結合することでデータベースを構築した．以下にデータベースに組み込んだ各データの詳細を記す．

(1) 建物データ（民間保有データ）

2016 年のゼンリン住宅地図（Zmap-TOWNII）の建物ポリゴンを利用した．建物ポリゴンには建物 ID や建物名称などの他に，建物用途に関する情報が付与されている．本研究の対象は戸建て住宅のみであるため，建物用途が戸建て住宅となっている建物のみを対象にモデルを構築した．なお，和歌山市全域の建物数は 211,753 件であり，そのうち本研究の対象となる戸建て住宅は 74,350 件（全建物のうち約 35%）であった．

(2) 住民基本台帳（自治体保有データ）

住民基本台帳とは国内の居住者に関する情報が収録されたものである．本研究では，2017 年 3 月 31 日現在における和歌山市の全居住者 359,979 人のデータを使用した．

なお．本データはあらかじめ個人情報に直接抵触する情報を秘匿化処理されたものを使用している．また，本研究では，住民基本台帳が有する情報のうち，世帯内の最高年齢と最低年齢および世帯人員総数を使用した．

(4) 水道使用量データ（自治体保有データ）

2017 年度と 2018 年度の和歌山市内の全水道栓の月ごとの水道使用量と住所が記載されたデータである．本研究では調査対象となる住宅 74,350（件）のうち，71,875 件に水道使用量データが付与された． 残りの 2,475 件は，水道道使用量が空間結合されていなかったため，使用量を 0 として扱った．これは，水道が閉栓，または該当する住宅が井戸から取水するなどを理由として，水道を利用していないなどの理由から水道使用量データが欠損値となったためである．

(5) 生活利便施設の立地データ（オープンデータ）

本研究では様々な生活利便施設と個々の建物との最短距離を変数として使用した．本研究で使用した施設データは，駅，バス停，医療機関，福祉施設，市町村役場の 5 種類で，いずれも国土数値情報が提供しているオープンデータである．表-1 に本研究で使用した生活利便施設の一覧とその作成年度を示す．

なお，既存研究においても，主要な公共交通ネットワークが空き家の発生に影響を及ぼしていることが指摘されている [14)]．そのため，駅とバス停のデータを使用した．また，医療機関や福祉施設に関しても同様に，独居高齢者の生活を支援する役割を担っていることから，空き家の発生に影響を及ぼすとされている [14)]．さらに，市町村役場までの到達時間が一定以上の縁辺性の高い集落では

表-1　本研究で使用した生活利便施設の一覧とその作成年度

種類	利用年度
駅	2018
バス停	2022
医療機関	2020
福祉施設	2015
市町村役場	2022

急激な人口減少や高い高齢化率を背景に世帯減少局面に入っているとされている[15]．そこで，空き家の所在地と市町村役場までの距離の関係を把握するために，市町村役場のデータを利用した．

(6) 空き家現地調査結果（自治体保有データ）

同データは2016年から2017年にかけて和歌山市が実施した外観目視による空き家の現地調査の結果である．本データは後述するクラスタリングの特徴量としては使用していないが，クラスタリング結果の信頼性検証を行うために使用した．

(7) 空き家データベースの整備

住民基本台帳および水道使用量データは住所情報を有する．そこで，本研究では Yahoo! Japan が提供する Yahoo! ジオコーダ API を用いて，これらの住所をアドレスマッチング（住所を緯度経度に変換）した．本研究では住所が住居に対して高精度で一意に定まるデータのみを使用した．なお，住居表示を実施している地域と地番住所の地域において，アドレスマッチングの精度上の顕著な差は見られなかった．

最後に各建物に（2）から（5）のデータを，GIS を用いて空間結合することでデータベースを作成した．表-2 にそれぞれのデータの戸建て住宅への吸着率を示す．なお，吸着率とは戸建て住宅のうち（2）から（5）のデータが吸着した戸建て住宅の割合を指す．その結果，戸建て住宅のうち約 69%に住民基本台帳，約 67%に水道使用量データ，100%に生活利便施設立地データが吸着した．また，現地調査によって空き家と判断された建物件数は 540 件で，これは調査対象となる戸建て住宅の約 0.7%である．

3. 空き家分布推定モデルの開発

本研究では，機械学習の教師なし学習に分類されるクラスタリングを用いて空き家分布推定を行っていく．既往研究では教師あり学習を用いた研究が多く存在するが

表-2　各変数の戸建て住宅への吸着率

変数		吸着率(%)
自治体保有データ	住民基本台帳	69.3
	水道使用量データ	66.9
	現地調査結果	0.7
オープンデータ	各施設までの最短距離	100.0

[2),9),10),11)]，教師あり学習と教師なし学習では，それぞれ適用領域や利点が異なる．教師あり学習とは正解のラベルを有するデータに基づいて予測を行う手法である．モデルはこのラベルと特徴量との関係を学習することで予測モデルが完成する．一方，教師なし学習とは，正解ラベルを有さないデータセットからデータの構造やパターンを抽出する手法であり，データ内の潜在的構造を発見し，特徴を理解することができる．本研究では前述の通り空き家の現地調査の結果という正解ラベルに影響を受けないモデルを構築することが重要であるため，クラスタリング手法を用いてモデルの構築を行う．

(1) モデルに用いる変数選択

本研究では，自治体保有データ，生活利便施設の立地データを変数として用いている．前述の通り，既往研究[3)]において水道使用量や世帯人員などの自治体保有データが空き家推定に有効であることが明らかになっているため，必須の項目として組み入れた．一方，生活利便施設の立地と各建物間の距離データは，複数組み合わせることによって，戸別の建物に地域特性を付与することが可能である．一方，これらの情報を全て利用することで多重共線性が起こる可能性があり，モデル精度を低下させる恐れがある．そこで本研究の立地情報は，駅，バス停，医療機関，福祉施設，市町村役場それぞれの距離データを単一または複数組み合わせたモデルを複数構築し，最も空き家を抽出可能なモデルを探索した．その結果，本モデルでは市町村役場，医療機関，福祉施設を変数として取り入れた．

(2) クラスタリングで採用した手法

本研究では，クラスタリングを利用して，高頻度に空き家が存在するクラスターを特定する試みを行った．本研究では，クラスタリング手法として，k-means 法をベースとした k-means++法を採用した．k-means 法はクラスタリングを行う手法として広く用いられている手法の一つである[16)]．本手法はまず入力データを空間上にプロットし，初期のクラスター中心をランダムに配置する．続いて，各データに最も近い重心点を探索し，該当のクラスターに所属させる．すべての入力データの所属が決定後，それぞれのクラスターの重心を再計算し，これを新しいクラスターの重心点にする．以上の手順を繰り返し，重心の移動距離が十分に小さくなれば終了とする手法である．

しかし本手法はランダムに初期重心点を配置するため初期位置が偏ってしまう可能性があるという課題を有する．そこで k-means++法では，k-means 法における初期重心点をなるべく広げて取るように改良した手法である [17]．そのため k-means 法よりも安定的な結果を得ることができるという利点がある．そのため本研究におけるクラスタリングの手法として採用した．

(3) エルボー法を用いたクラスター数の検討

クラスタリングを行う上で，クラスター数はクラスタリング結果に大きな影響を与える．例えば，クラスター数を過剰に設定してしまうと，各クラスターに属するデータ数が少なくなってしまい，ノイズがクラスターとして誤って認識される過学習が起きる可能性が生じる．さらに，クラスター数が多すぎる場合，クラスター間の意味の違いを認識することが困難になり，各クラスターの解釈が難しくなってしまう [17]．そのため，クラスター数を適切な値に設定することが重要になる．

そこで，本研究ではエルボー法を用いて適切なクラスター数を設定した．エルボー法はクラスターの重心とクラスターに所属する各データの距離の総和に注目し，クラスター数を 1 つから増やしていく．クラスター数が適切な値まで増えていく過程では，各点がより近いクラスター重心に所属できるため，距離の総和は減少していく．しかし，適切な値を超えてさらにクラスター数が増えていく過程では，この総和の減少の程度が徐々に低下していくという特徴を有する．すなわち，以上の総和の減少のペースが鈍化する直前のクラスター数が最適であると判断できる．

図-1 に本研究のデータに対してエルボー法を適用した結果を示す．この結果は 2017 年度の最大水道使用量，世帯人員，世帯内最高年齢，最寄り市町村役場までの距離，最寄り医療機関までの距離，最寄り福祉施設までの距離を変数として用いたもので，クラスター数が 3 を超えると，減少の程度が小さくなることが分かる．この結果から，今回行ったクラスタリングではクラスター数は 3 あるいは 4 が適当であることが分かる．本研究の成果は主に悉皆調査の効率化など，自治体の空き家対策業務において活用されることを目標としている．そのため，クラスター数をできるだけ少なくすることでクラスタリング結果の解釈を容易にするとともに，対策重要度を「低，中，高」などのように明瞭にできるものと考え，クラスター数は 3 とした．

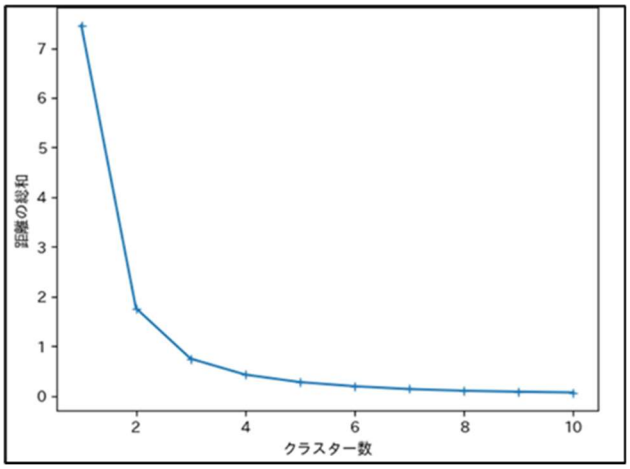


図-1　エルボー法の結果

4. 結果

(1) クラスタリング結果の評価

表-3 にクラスタリング結果を示す．クラスター0 により対象となる戸建て住宅の約 52%を抽出しており，このクラスター内には対象地域の全空き家数のうち 77.5%が含まれていることが分かる．和歌山市を対象とした教師あり学習を行った既往研究 [11]と比較すると精度は落ちるものの，高い水準で空き家が属するクラスターの形成に成功しており，対象となる全住宅の半数を調査することで，既往研究の課題であった空き家の見落としをすることなく，約 80%の空き家を発見することができた．これは，各自治体が行っている空き家の現地調査の対象となる建物数を削減し，調査の効率化に十分に貢献出来ると考えられる．さらに，本成果は現地調査をする際に，外観目視以外の観点から空き家か否かを判断する材料になる可能性や，空き家が多く属するクラスターでは空き家に似た特徴の住宅が多いため空き家になり得る住宅の早期発見などにつながる可能性も考えられる．

(2) 和歌山市全域の建物ごとのクラスター分布

図-2 に和歌山市における現地調査で空き家と判定された住宅の分布を示す．和歌山市では，空き家は住宅が密集している地域に数多く分布しており，市全域を見ても中心市街地に多く空き家が分布していることが分かる．

次に，クラスターごとに含まれる空き家の分布を図-2 に示す．クラスター0 とクラスター1 には，地理的特徴は見られず市全域に広く分布していることが分かる．一方，クラスター2 は市の中心部から離れた郊外や海岸沿いに分布していることが分かる．これは，クラスター2 に分類された住宅が，新興住宅街などの比較的新しい戸建て住宅が多く含まれていることに起因すると考えられる．

(3) 本モデルの特徴量の分布

表-4 に相関係数を示す．他の特徴量と比べて空き家との相関係数が大きい特徴量は，-0.087 の水道使用量と，-0.033 の世帯人員であった．これは，空き家では水道が使用されることが少ないため，水道使用量が極端に少な

表-3　クラスタリング結果

クラスター番号	抽出できた空き家の割合(%)	全建物に対する割合(%)
0	77.5	51.9
1	18.7	37.4
2	3.8	10.7

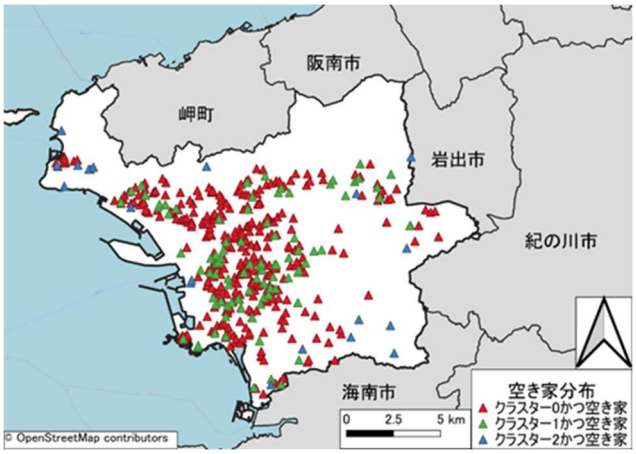


図-2　各クラスターに所属する空き家の分布

表-4　各変数と空き家判定の相関係数

	2017年最大水道使用量	世帯人員	最高年齢	役場までの距離	医療機関までの距離	福祉施設までの距離
空き家判定	-0.087	-0.033	-0.0019	-0.020	-0.021	-0.012

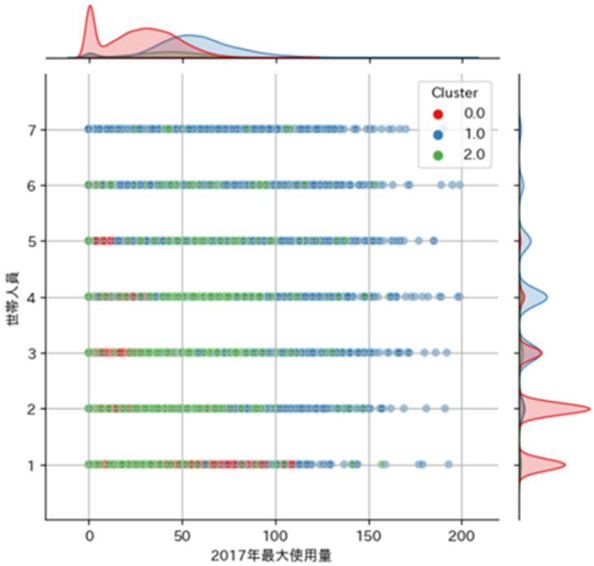


図-3　2017 年水道使用量と世帯人員のクラスター分布

い住宅が空き家である可能性が高いということを示していると考えられる．また，世帯人員が複数人である場合，世帯主が亡くなる等の理由で空き家になる可能性は低くなる一方，一人暮らしの場合は世帯主が亡くなることで空き家になる可能性を有するため，世帯人員と空き家の間には負の相関関係が現れているもと考えられる．

次に，この 2 つの特徴量の分布を図-3 に示す． 2 つの特徴量について，クラスター0 はどちらか一方，もしくはどちらも小さい値を取る住宅が多く分類されている．一方，クラスター1 は少なくとも一方、もしくはどちらも大きい値を取る多い住宅が分類されており，クラスター2はどちらの特徴量もクラスター0 と 1 の間の値を取る住宅が多く分類されている．さらに，クラスターごとに水道使用量，世帯人員の平均を算出すると，クラスター0 は水道使用量 26.56，世帯人員 1.95，クラスター1 は水道使用量 58.32，世帯人員 3.89，クラスター2 は水道使用量 41.08，世帯人員 2.81 となり，水道使用量と世帯人員がクラスタリングに与えている影響が大きいと確認できた．

その一方で市町村役場，医療機関，福祉施設についても負の相関が出ているものの，水道使用量や世帯人員ほど強い相関といえるものではなかった．これは，これらの特徴量が水道使用量などの空き家と非空き家の属性で明確に数値が変わるものではなく，一定の範囲の中に調査対象となる戸建て住宅の多くが存在するためであると考えられる．また，地方都市では住民の多くが移動手段として自動車を用いている場合が多く，各施設までの最短距離は空き家予測にあまり寄与しないものと考えられる．

5. まとめ

本研究では現状の空き家の分布を迅速かつ安価に，精度良く把握する手法として，和歌山市全域を対象として官民保有のデータとオープンデータを用いてデータベースを作成し，教師なし学習によるクラスタリングを行った．その結果，対象住宅の半数の中に全体の約 80%の空き家を含む形で抽出することができた．また，本研究の目的であった他自治体への展開の妨げの要因となっていた，空き家の現地調査の結果を使用することなく一定の精度を得られたため，空き家推定における教師なし学習が有用である可能性が明らかになった．精度の観点から見れば，既往研究で用いられてきた教師あり学習の方が勝るものの，現地調査を行うことなく空き家予測を行うことができる本モデルは，今後の空き家対策業務に一定の貢献することができるものと期待される．

現在，多くの地方自治体では，様々なデータ・統計を活用した EBPM の推進に取り組む一方で，空き家に関する政策は人の主観に大きく左右される現地調査結果に基づいて立案されている状況にある．本研究によって得られるクラスタリング結果から，空き家に関する客観的・定量的な事実に基づいた判断をすることで空き家対策においても EBPM が推進されるきっかけになるものと考えている．

今後は，高精度で予測できる教師あり学習と容易に展開できる教師なし学習の 2 つの特徴を活かすことのできる自己教師あり学習を用いてモデルを作成し，調査対象

建物の削減と空き家把握業務の効率化を目指したいと考えている.

謝辞：本研究は東大CSIS共同研究（No.880）の一環として実施した．また，本研究は和歌山県統計データ利活用センターおよび和歌山市より，和歌山市の自治体保有データ（住民基本台帳，水道使用量データ，空き家現地調査結果）の提供を受けることで実現した．さらに，東京都市大学総合研究所デジタル都市空間情報研究開発ユニットの成果の一部でもある．ここに記して謝意を表したい.

参考文献

1) 総務省統計局：平成30年住宅・土地統計調査の集計結果（住宅及び世帯に関する基本集計）の概要，<https://www.stat.go.jp/data/jyutaku/2018/pdf/kihon_gaiyou.pdf>，（最終閲覧日 2023 年 10 月 8 日）
2) 水澤克哉，宮本 慧，田村翔太，田中貴宏： 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-， pp. 1377-1384， 都市計画論文集， 2022.
3) 水谷昂太郎，冨田健人，秋山祐樹，谷内田修： 建物周辺の生活利便性を考慮した建物ごとの空き家確率予測モデルの構築， 第 31 回地理情報システム学会講演論文集， C-4-4 , 2022.
4) 秋山祐樹，柴崎亮介： マイクロ将来人口推計データを用いた将来の空き家分布推定，41，「CSIS DAYS 2015 研究アブストラクト集」，2015.
5) 国土交通省：空き家対策小委員会, <https://www.mlit.go.jp/policy/shingikai/s204_akiyataisaku.html>，（最終閲覧日 2023 年 10 月 9 日）
6) 益田理広・秋山祐樹：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，pp.1-26，2020.
7) 石河正寛，松橋啓介，金森有子，有賀敏典：住戸数と世帯数に基づく空き家の詳細地域分布の把握手法. 都市計画論文集，Vol.52，No.3，pp.689-695，2017.
8) 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.
9) 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，No2，pp.62-71，2021.
10) 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.
11) Sayuda, K., Hong, E., Akiyama, Y., Baba, H., Tokudomi, T., Akatani, T. :Accuracy of vacant housing detection models : An empirical evaluation using municipal and national census datasets, Transactions in GIS, 2022
12) Kento, T. Yuki, A. Hiroki, B; Osamu, Y. :Estimating the Spatial Distribution of Vacant Houses with Machine Learning using Municipal Data. IGARSS 2022 - 2022 IEEE International Geoscience and Remote Sensing Symposium., pp1276-1279, 2022.
13) 上田彰紘，秋山祐樹，大野佳哉：空き家発生・分布メカニズムの解明に関する調査研究(その 1)．国土交通政策研究所報，vol.61，pp.24-35，2016.
14) 馬場弘樹，樋野公宏：管理不全空き家の特性と地区レベルでの管理傾向．都市計画論文集，vol.16，pp.108-111，2017.
15) 藤山浩，中山大介：島根県中山間地域における人口減少の GIS 分析-集落 GIS データベースと農村計画への活用性-．農村計画論文集，vol.25, pp.431-436，2006.
16) MacQueen, J. :Some methods for classification and analysis of multivariate observations . Berkeley Symposium on Mathematical Statistics and Probability, vol.5，No1，pp.281-297，1967.
17) David, A . and Sergei, V. :k-means++: The Advantages of Careful Seeding．Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms , pp.1027-1035，2007.
18) 林拓也：Reduced K-means 法によるクラスター分析の社会調査データへの応用．データ分析の理論と応用，vol.7，No1，pp.9-19，2018.

(Received ？ ？, 2023)
(Accepted ？ ？, 2023)

UTILIZING UNSUPERVISED LEARNING DEVELOPMENT OF A MODEL FOR ESTIMATING THE DISTRIBUTION OF VACANT HOUSES

Manato ABE, Kotaro MIZUTANI, Kento TOMITA and Yuki AKIYAMA

The number of vacant houses is increasing nationwide in Japan. Local governments are required to take immediate action on these vacant houses and to formulate long-term plans for vacant house countermeasures. However, vacant houses have been identified only by on-site surveys, a time-consuming and costly process. In this study, we constructed a prediction model for the distribution of vacant houses by clustering, a type of unsupervised learning, using municipal data (water consumption, basic resident registers, etc.). The model was able to identify 77.5% of all vacant houses within approximately 52% of the surveyed houses. These results are expected to reduce the number of target buildings in the vacant house survey and contribute to the efficiency of vacant house identification and the consideration of vacant house countermeasures.