機械学習に基づいた空き家推定モデルの適用性の検証

摂南大学大学院　学生員　○南　宇純
摂南大学大学院　学生員　桃瀬　凌
摂南大学　正会員　熊谷　樹一郎

１．研究の背景

近年, 我が国の人口減少の問題に対して，「コンパクト・プラス・ネットワーク」の取り組みが国土交通省主導で進められている．この取り組みには中長期にわたる都市構造のモニタリングが必要とされており，空き家などの低・未利用空間の分布を把握することの重要性が指摘されている．一方で，既存の市街地内では空き家などの低・未利用空間が増加するといった問題がある．そこで我々は，これまでに広域的な空き家推定モデルを開発してきた．3 時期にわたった小地区での現地調査結果と国や地方自治体が定期的に取得・更新する情報とを統計モデルに適用するアプローチである．一方，近年になって利用可能なデータ量が増加するとともにコンピューターの処理能力も向上し，機械学習・深層学習・AI を適用した分析法が一般的，かつ，実用的になってきた[1)]．新たな空き家推定モデルとしてこれらの分析法を採用することで，推定精度を向上させることが期待できる．

２．研究の目的

本研究では，3 時期にわたる現地調査の結果と国や地方自治体が定期的に取得・更新する情報とを広域的な空き家推定モデルに適用し，推定結果に表れる特徴について調査する．具体的には，新たに導入したランダムフォレストと，これまで用いられてきた順序ロジットモデルとを空き家推定モデルに適用し，その適用効果を比較・検証する．

３．研究の内容

３．１ 検討ケースの設定

ランダムフォレストと順序ロジットモデルとの適用効果の比較には，2016～2018 年度の寝屋川市 5 地区 11 町丁目を対象地区とした現地調査の結果を用いた．本研究では**表**−1 のように用途地域，建物の構造，建物の種類を地理データ，人口密度データや建物のデータなどを共通データと定義し，検討ケースを地理データと共通データとの組み合わせに基づき設定した．具体的には，「共通データのみ：Case1」，「共通データ+建物の構造と用途地域：Case2」，「共通データ+建物の種類と用途地域：Case3」，「共通データ+建物の構造と種類：Case4」，「共通データ+建物の構造と種類と用途地域：Case5」の 5 ケースを設定し，ランダムサンプリングにより生成した 60 セットの基準データセットを整備した．

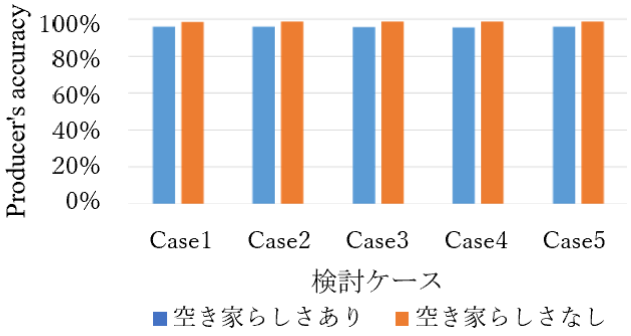
表-1 検討ケース

説明変数		検討ケース番号				
		1	2	3	4	5
共通データ	築年数（年）	全変数を採用				
	水道栓密度（㎡）					
	水道栓密度（栓）					
	水道栓の閉栓日数（日）					
	水道栓の利用状態					
	1995年-2000年の人口密度の変動（人/㎡）					
	2000年-2005年の人口密度の変動（人/㎡）					
	2005年-2010年の人口密度の変動（人/㎡）					
	2010年-2015年の人口密度の変動（人/㎡）					
用途地域			○	○		○
建物の構造			○		○	○
建物の種類				○	○	○

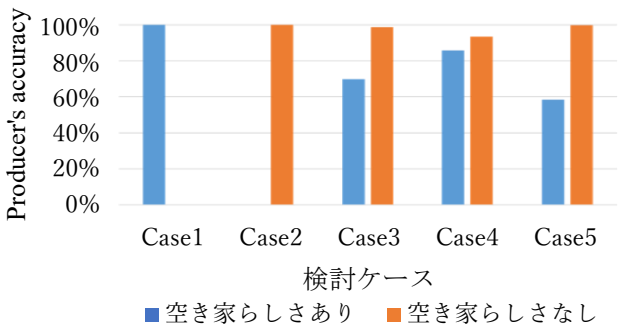
３．２ ２つの空き家推定モデルの比較

ランダムフォレストと順序ロジットモデルでの 5 ケースについて，ある 1 セットの基準データでの Producer’s accuracy の例を**図**−1 に示す．**図**−1(a)を見るとランダムフォレストは Case1 から Case5 まで同じような傾向で高い判別精度を示していることが分かる．一方，順序ロジットモデルの結果を示した**図**−

キーワード　空き家，空き家推定モデル，機械学習，ランダムフォレスト，順序ロジットモデル
連絡先　〒572-8508　大阪府寝屋川市池田中町 17-8　TEL/FAX：072-839-9122　E-mail：kumagai@civ.setsunan.ac.jp



(a)ランダムフォレスト

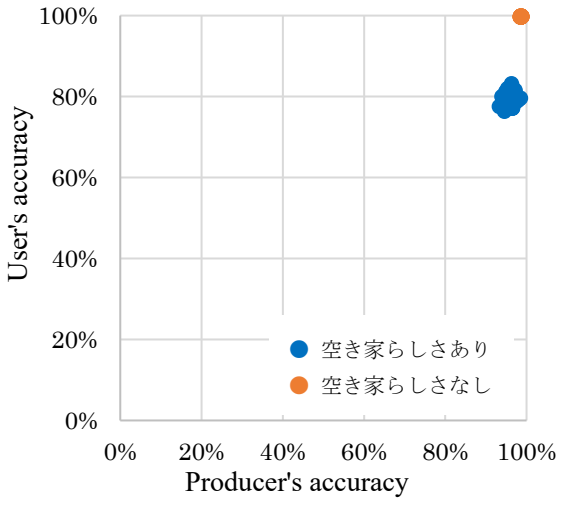


(b)順序ロジットモデル精度

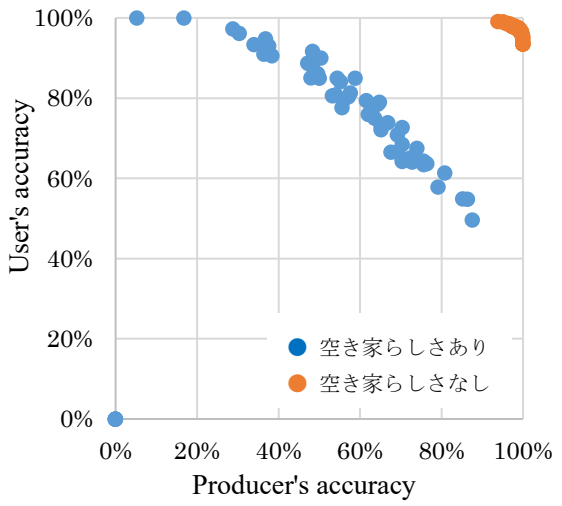
図-1 判別精度

1(b)では Case1 のみ「空き家らしさあり」の判別精度がランダムフォレストを越えているものの，「空き家らしさなし」の判別精度が 0%を示している．また，Case2 から Case5 は「空き家らしさあり」の判別精度は高いとはいえないが，「空き家らしさなし」はランダムフォレストと同等かそれ以上の判別精度を示している．この傾向は User's accuracy でも同様であった．このことから，ランダムフォレストでは地理データの使用・不使用に左右されず，安定した結果が得られている可能性が示唆される．

図-2 には，例として Case5 を取り上げ，ランダムフォレストと順序ロジットモデルそれぞれでの 60 セットの基準データセットの推定精度をプロットした．縦軸を User's accuracy，横軸を Producer's accuracy で表しており，青色が空き家らしさあり，オレンジ色が空き家らしさなしを示す．これらの図においては，点が右上にいくほど精度が高くなっていることを意味する．順序ロジットモデルを採用した場合の**図-2(b)**では，「空き家らしさあり」にばらつきが現れている．一方で，ランダムフォレストを採用した**図-2(a)**では，「空き家らしさあり」「空き家らしさなし」ともにばらつきが小さく，かつ，高い判別精度を示している．なお，順序ロジットモデルでは 1 ケースの計算に要する時間が約 3.5 日であったことに対して，ランダムフォレストでは約 3 秒と大幅な時間短縮となった．このことからランダムフォレストは効率の良い処理が可能となるとともに，地理データの違いや基準データセットの違いに寄らず User's accuracy，Producer's accuracy ともに推定精度が高くなる可能性が示唆された．



(a)ランダムフォレスト



(b)順序ロジットモデル

図-2 Case5 での判別精度の分布

4．まとめ

ランダムフォレストと順序ロジットモデルの判別精度を比較した．結果として，ランダムフォレストでは地理的変数の有無や，基準データセットの違いの影響が少なく，安定して高い判別精度が得られることが分かった．今後は，現地調査の労力軽減を前提に，基準データを取得する範囲を狭めた場合に，同程度の推定精度が得られるかを検証する．

【参考文献】 1) 李勇鶴，布施孝志：小特集「深層学習（その１）」～深層学習の概要と写真測量・リモートセンシング分野での応用状況～，写真測量とリモートセンシング，Vol.61，No.4，pp.56-65，2022.