

オープンデータを活用した 建物ごとの構造推定手法の開発

武田 直弥¹・古谷 貴史²・秋山 祐樹³

¹ 学生会員 東京都市大学 工学部都市工学科 (〒158-0087 東京都世田谷区玉堤 1 丁目 28-1)
E-mail: g1818055@tcu.ac.jp

² 学生会員 東京都市大学 大学院総合理工学研究科 (〒158-0087 東京都世田谷区玉堤 1 丁目 28-1)
E-mail: g2081624@tcu.ac.jp

³ 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 (〒158-0087 東京都世田谷区玉堤 1 丁目 28-1)
E-mail: akiyamay@tcu.ac.jp

近年、都市の 3D モデルが防災計画の立案や、避難訓練シミュレーションなどの防災分野において活用され始めている。一方、大規模地震に備えた防災計画や都市計画を立案する際には、建物の形状だけでなく建物ごとの構造に関する情報が重要である。しかし、建物ごとの構造に関する情報は入手が困難である場合が多い。そこで本研究では、東京都世田谷区、渋谷区、足立区を対象に国勢調査などの既存統計やオープンデータとして容易に入手できる様々な空間情報から得られる多様な属性値を建物データを付与した上で、機械学習手法の 1 つである XGBoost を用いて建物の構造推定を行う手法を開発した。また、その信頼性の検証を行った結果、何れの区においても約 80% の精度で構造推定が可能となった。さらに、特徴量の重要度を算出した結果、既存統計を基に作成した特徴量も構造推定に大いに寄与していることも明らかとなった。

Key Words: building structure estimation, disaster prevention plan, open data, machine learning

1. はじめに

2020 年 4 月より、国土交通省は「Project PLATEAU」と呼ばれる日本の都市の 3D 都市モデルの整備を推進するプロジェクトに着手した。これは都市の 3D モデルをオープンデータとして公開し、デジタルツインを推進するための取り組みであり、防災や街づくりなどへの活用が期待されている。すでに水害時の浸水想定を 3 次元的に実施することによる防災計画立案の支援や、避難訓練のシミュレーションなどの活用事例がある^{1,2)}。一方、大規模地震に備えた防災計画や都市計画を立案する際には、建物の形状だけでなく建物ごとの耐震性に直結する建物構造が重要な情報になる³⁾。しかし、建物ごとの構造に関する情報は一般に公開されていない場合が多い。また、日本では課税事務の際に得られる建物に関する情報に構造情報が含まれている場合が多いものの、目的外使用や個人情報保護の観点から使用が困難である⁴⁾。

(1) 既存研究

建物の属性を推定する既存研究としては、条件や確

率に応じて建物属性を配分した研究と、機械学習や深層学習を用いて推定を行った研究が挙げられる。まず、建物属性を配分した事例として、尾崎・盛岡 (2016) は住宅・土地統計調査をクロス集計し、建物属性の条件付確率を算出することで建物ごとの構造を推定した⁵⁾。しかし、同研究では信頼性の検証が行われていないという課題があった。一方、機械学習や深層学習を用いて推定した事例として、Li et al. (2018) による建物画像と深層学習を用いて建物の建築年を推定した事例⁶⁾や、同様に建築年を推定した事例として Rosser et al. (2019) による地形図と数値表層モデルを用いて機械学習により推定した事例がある⁷⁾。また、建物構造を推定した既存研究としては、小川ほか (2020) による不動産データと建物画像データを用いた建物構造推定がある⁸⁾。他にも Zhou and Chang (2021) は建物の形状や POI (Point of interest) などを説明変数にし、12 種類の機械学習手法を用いることで、建物ごとの構造を推定している⁹⁾。同研究では勾配ブースティングのアルゴリズムを用いて高精度の推定が可能であった。しかし、推定には建物の形状に関する詳細な情報が必要であり、また木造の建物は推定できないという課題があった。

これらの既存研究では、建築年や構造といった建物属性の推定において一定の成果が得られている。しかし、これらの先行研究の手法はデータの収集や整備に時間やコストを要する建物外観の画像や、一般的に入手が困難な建物の形状に関する高精細なデータが必要になるといった課題がある。

(2) 本研究の目的

そこで本研究では、国勢調査などの既存統計や、オープンデータとして入手できる様々な空間情報を、建物データに付与した上で、建物構造推定に関する既存研究では活用事例の未だ少ない機械学習手法である”XGBoost”を用いて建物ごとの構造推定を行うとともに、推定結果の信頼性検証を行った。なお、研究対象地域は後述する建物ごとの構造情報が公開されている東京都世田谷区、足立区、渋谷区（以下、「研究対象地域」）とする。

2. 研究手法

本研究では、建物用途が住宅（戸建住宅・共同住宅・住商混合ビル）の建物を対象に建物構造の推定を行う。まず、東京都世田谷区を対象に教師データの作成を行う。次に、既存統計から特徴量を作成し建物データと結合する。そして機械学習モデルを構築し、モデルの性能検証と汎化性能の検証を研究対象地域全域と兵庫県神戸市を対象にそれぞれ実施した。

(1) 教師データの作成

本研究では、機械学習に用いる教師データを研究対象地域において建物ごとの構造が地図上で公開されている建物構造別現状図を用いて作成した。具体的な作成手法は、研究対象地域全域の建物データから無作為に抽出した住宅用途の建物4,500件に対して（図-1）、建物構造別現状図から得られる建物構造（図-2）を目視で1棟ごとに付与することで作成した。なお、本研究で使用した建物データは2016年の住宅地図（Zmap TOWN II：株式会社ゼンリン）である。表-1に教師データの建物用途別の構造別建物数を示す。また、建物現状構造図では耐火性能のみが公開されているため、耐火造と準耐火造の建物を非木造、防火木造と木造の建物を木造と分類した。

(2) 既存統計の加工と建物データへの結合

本研究では機械学習により建物構造を推定するため、推定に用いる特徴量を既存統計から町丁字毎に作成した。表-2に本研究で作成した特徴量を示す。これらの特徴量は全国整備されている国勢調査（2015年）や住宅・土地統計調査（2018年）などの既存統計というオ

ープンデータから作成可能なため、データの入手が容易かつコストもかかりにくいという利点がある。

まず「1. 各年齢階級人口の割合」は2015年の国勢調査小地域集計第3表から年少人口、生産年齢人口、老年人口の割合を算出することで作成した。これは居住者が高齢化した地域ほど木造の建物が多い傾向にあると考えられるため追加した。図-3に一都三県における高齢人口の割合の空間分布を示す。

次に「2. 居住期間の割合」は2015年の国勢調査小地域集計第13表から作成した。これは地域全体の居住者の居住期間が長い傾向にある地域では木造の建物が多い傾向にあると考えられるために追加した。図-4に一都三県における居住期間が20年以上の人口割合の空間分布を示す。

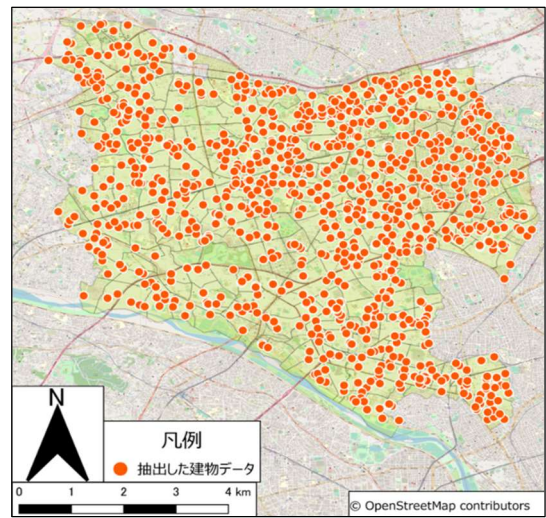


図-1 抽出した建物データ（世田谷区の例）



図-2 世田谷区三軒茶屋周辺における建物現状構造図

表-1 教師データの建物用途別構造別建物数

建物用途	木造[棟]	非木造[棟]
一般住宅	2,254	964
住商混合ビル	18	213
共同住宅	299	726
合計	2,571	1,903

表-2 建物データに付与する特徴量

作成した特徴量	使用した統計資料
1. 年齢階級別人口割合 (3 区分)	国勢調査小地域集計 (2015 年)
2. 居住期間別世帯割合 (6 区分)	国勢調査小地域集計 (2015 年)
3. 年収階級別世帯割合 (9 区分)	国勢調査小地域集計 (2015 年) 住宅・土地統計調査 (2018 年)
4. 建て方別構造割合 (3 区分)	住宅・土地統計調査 (2018 年)
5. 用途地域 (13 区分)	用途地域ポリゴン (国土数値情報)
6. 建物データ	面積, 階数, 延床面積, 建物周長, 用途 (3 区分: 戸建住宅・共同住宅・住商混合ビル), 部屋数

「3. 年収階級別推定世帯数割合」は2015年の国勢調査小地域集計第7表と2018年の住宅土地統計第444表から作成した。まず住宅土地統計から市区町村毎で集計された住宅の所有形態・年収階級別世帯数を割合に変換し、次に国勢調査の住宅の所有形態別の世帯数に対して、先に住宅土地統計から算出した割合に基づいて年収階級別の世帯数を推定することにより、町丁ごとの年収階級別の世帯割合を作成した。これは地域ごとの年収階級によって、その地域の住宅の建て方の傾向に影響が出るものと考えられるため追加した。図-5 に世帯年収 1,000 万円以上の世帯割合の空間分布を示す。

「4. 住宅の建て方別の各構造割合」は国勢調査小地域集計第8表と住宅土地統計第8-3表を用いて作成した。まず、住宅土地統計から地区町村毎の住宅の建て方別の各構造割合を算出した。これは木造の割合が高い地域ほど、建物個別の構造においても木造となる確率が高くなるため追加した。図-6 に市区町村ごとの戸建てかつ木造の建物割合の空間分布を示す。

「5. 用途地域」は国土数値情報より入手した2011年の用途地域データ(ポリゴンデータ)を使用した。なお、用途地域はダミー変数化して付与した。これは、等しい用途地域が指定されている場合には類似した特徴を有する建物が集まっていると考え、類似した建物のグループを表現することができると考え追加した。

最後に「6. 建物データ」は本研究で使用した建物データから得られる変数である。階数, 用途, 部屋数は住宅地図の建物ポリゴンが持つ属性値であり、面積と建物周長はGISソフトウェア(QGIS3.16)(以下「GIS」)を用いて建物の各ポリゴンから取得した。また、延床面積は面積と階数の積により取得した。

なお、本節で作成した既存統計を基に作成した特徴量は、建物データと国勢調査の小地域ポリゴンデータ(2015年)をGISを用いて空間結合を行うことで、

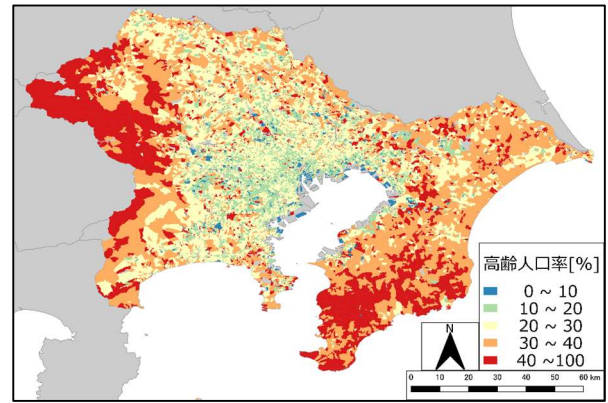


図-3 高齢人口率
(一都三県, 町丁字毎)

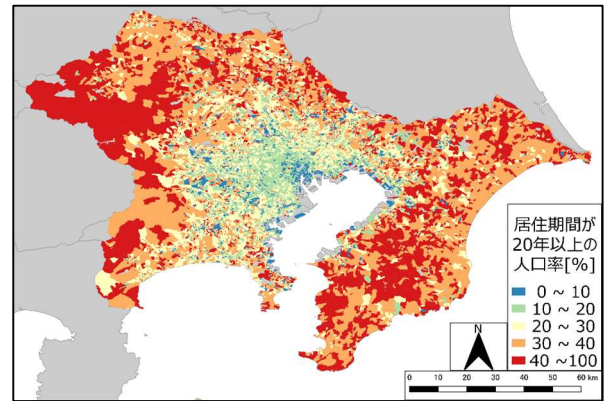


図-4 居住期間 20 年以上の人口割合
(一都三県, 町丁字毎)

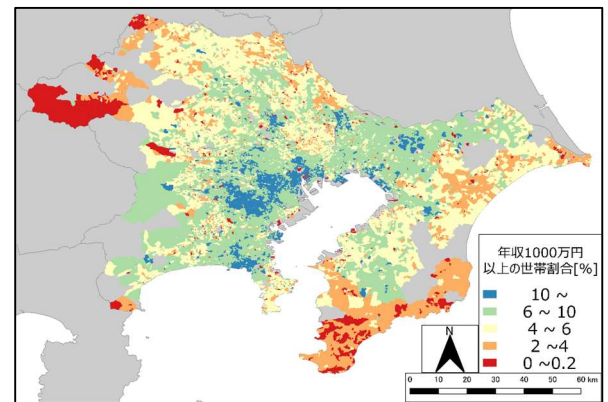


図-5 世帯年収 1000 万円以上の世帯割合
(一都三県, 町丁字毎)

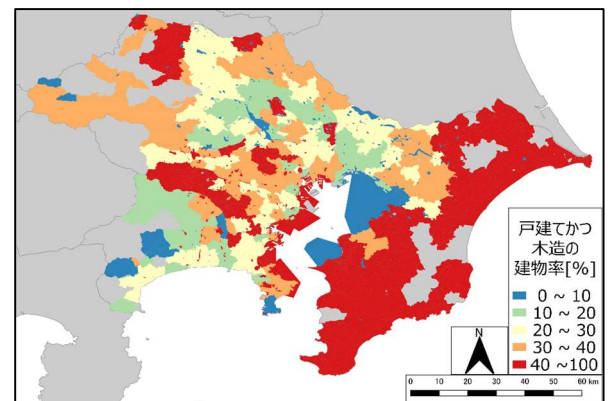


図-6 戸建てかつ木造の建物割合
(一都三県, 町丁字毎)

建物ごとに町丁字固有の11桁もしくは9桁の町丁字コードを付与を行い、付与したコードを基に建物ごとに町丁字に対応する統計値を付与した。また、用途地域はポリゴンデータで公開されているため、GISを用いて建物データと空間結合を行うことで用途地域の情報を建物ごとに付与した。

(3) 構造推定の手法

本研究では、前節で建物データに与えた属性と建物データが保有する属性合わせて37種類の属性を、建物構造を説明するための説明変数とする。本研究のように複数の説明変数から目的変数を推定する際に、これまでに広く用いられている手法としては、重回帰分析やロジスティック回帰分析が知られている^{10,11)}。また、モデルが線形的に振る舞わないことが予想される場合は、ランダムフォレストやサポートベクタマシンなどの分類器を用いた推定方法も用いられている¹²⁾。一方、本研究で扱うデータは、用途地域など一部の説明変数において欠損値が発生する場合がある。前述の手法では欠損値を含むデータを扱うことが困難なため、欠損値を含むデータを扱うことができるモデルを構築する必要がある。そこで、本研究では欠損値を含むデータを扱うことができる機械学習手法の1つであるXGBoost (eXtreme Gradient Boosting)を用いた。XGBoostはChen and Guestrin (2016)¹³⁾により考案された手法であり、決定木を逐次的に作成し、各決定木から得られた値を足し合わせ、多数決を取ることで予測値を推定する手法である。この手法は推定精度が高く、欠損値を含むデータも処理が可能で優れており^{14,15)}、本研究の目的達成に有用であると判断した。

(a) 機械学習モデルの構築

本研究では、モデルの学習を行った後、建物が木造に分類される確率を算出し、その確率が0.5を超えるものを木造に分類する機械学習モデルを作成した。損失関数にはlogloss (Logarithmic Loss)を用いた。本研究の損失関数に用いるloglossは式(1)で定義される。

$$\text{logloss} = -\frac{1}{N} \sum_{i=1}^N y_i \log(p_i) + (1 - y_i) \log(1 - p_i) \quad (1)$$

loglossとは N をデータ数、 p_i を予測確率、 $y_i \in \{0,1\}$ を正解ラベルをする時の、正解値との距離を考慮した評価値であり、同値が小さいほどモデルが優れていると評価できる¹⁶⁾。本研究では、木造に分類される確率を算出するモデルを構築するためこの損失関数を用いた。

また、本研究では汎化性能を高めるためにクロスバ

リレーション (以下「cv」) を用いてモデルの評価と学習を行った。本研究のcvはK-分割交差検証を用いた。同手法ではデータを k 個に分割し、そのうち1つを検証用データ、 $k-1$ 個を学習用データに分割する。この際に k 個のモデルが作成されるため、それぞれのモデルの予測値の平均を予測をしたいデータへの出力とした。本研究では学習用データとテスト用データの比率を8:2とし、 $k=5$ としてK-分割交差検証を用いた。cvの種類には他にも様々なものがあるが、本研究では分類問題を取り扱うことからK-分割交差検証を用いた。

3. 構造推定の結果と信頼性の検証

(1) 推定結果

前章で作成した機械学習モデルを用いて教師データを作成した世田谷区、渋谷区、足立区の全ての住宅用途の建物の構造推定を行った。図-7に世田谷区全域の推定結果、図-8に渋谷区全域の推定結果、図-9に足立区全域の推定結果を示す。また、図-10に世田谷区下北沢駅周辺の推定結果を示す。これらの結果が示すように研究対象地域全域において建物の構造を建物というミクロな単位で推定することが可能になった。

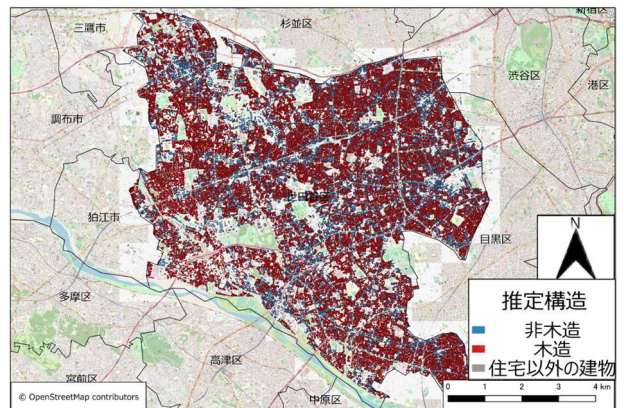


図-7 世田谷区全域の推定結果

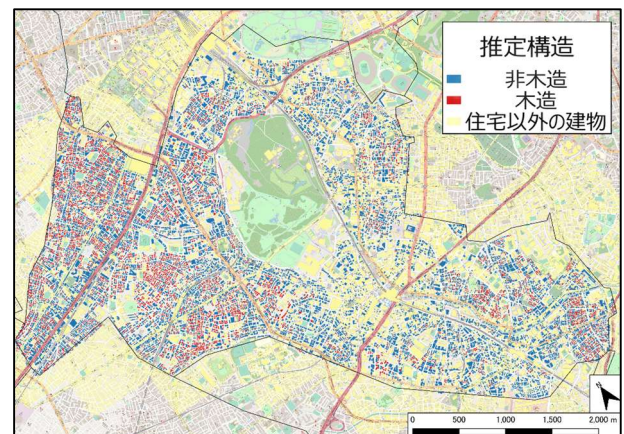


図-8 渋谷区全域の推定結果

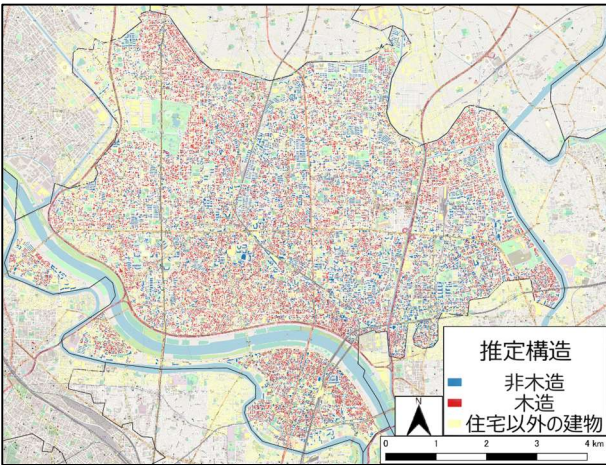


図-9 足立区全域の推定結果



図-10 世田谷区下北沢駅周辺の推定結果

(2) 信頼性の検証

本節では教師データを用いて信頼性の検証を行う。評価指標には正解率 (accuracy) と f 値 (f1_score) を用いて評価を行った。なお、評価に用いる正解率と f 値は、表-3 に示すように実際に木造の建物を木造と推定した個数を TP、実際には非木造の建物を木造と推定した個数を FN、実際には非木造の建物を木造と推定した個数を FP、実際に非木造の建物を非木造と推定した個数を TN としたときに式(2)と式(3)で定義される。

$$accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (2)$$

$$f1score = \frac{2 \frac{TP}{TP+FN} \cdot \frac{TP}{TP+FP}}{\frac{TP}{TP+FN} + \frac{TP}{TP+FP}} \quad (3)$$

表-3 評価に用いる混同行列

		真値	
		木造	非木造
推定値	木造	TP	FP
	非木造	FN	TN

(a) 研究対象地域全域における信頼性検証

本研究では、学習用データとテスト用データを 8:2 に分割し、更に学習用データを 5 分割し検証用データと学習用データに分けた。表-4 にテスト用データの検証結果を示す。テスト用データに対しては正解率 82.00%、f 値 0.7901 となった。

(b) 世田谷区のみでの検証

教師データは世田谷区、渋谷区、足立区にそれぞれ 1,500 件ある。そこで、まず世田谷区のデータのみで学習を行った結果を表-5 に示す。テストデータに対して正解率 83.00%、f 値 0.8971 となった。

(c) 渋谷区のみでの検証

続いて、渋谷区のデータのみで学習を行った結果を表-6 に示す。テストデータに対して正解率 79.86%、f 値 0.7317 となった。

(d) 足立区のみでの検証

最後に、足立区のデータのみで学習を行った結果を表-7 に示す。テストデータに対して正解率 76.58%、f 値 0.8254 となった。

(e) 信頼性検証結果の考察

表 4~表 8 から研究対象地域全域においても、またそれぞれの区においても、正解率、f 値ともに 80% 程度の精度で推定することができた。また、どの自治体においても実際には非木造である建物を木造と推定している件数が増える傾向が見られた。これは特徴量として用いた市区町村ごとの建て方別の構造割合において、同じ自治体内でも統計値から得られる自治体全体の傾向と差が大きくなる地域で推定精度が低下したものと考えられる。また、本研究では教師データを作成する際に準耐火構造の建物を非木造と仮定したが、実際には準耐火構造の建物の中に木造の建物が含まれているものと考えられる。

表-4 研究対象地域全域におけるテストデータに対する推定結果

検証データ	推定構造数		
	非木造 [棟]	木造 [棟]	合計
非木造 [棟]	261	134	395
木造 [棟]	45	455	500
合計 [棟]	306	589	895

表-5 世田谷区のデータのみでの推定結果

検証データ		推定構造数		
		非木造 [棟]	木造 [棟]	合計
正解値	非木造 [棟]	39	36	75
	木造 [棟]	15	210	215
	合計 [棟]	41	259	300

表-6 渋谷区のデータのみでの推定結果

検証データ		推定構造数		
		非木造 [棟]	木造 [棟]	合計
正解値	非木造 [棟]	142	49	191
	木造 [棟]	11	96	107
	合計 [棟]	153	145	298

表-7 足立区のデータのみでの推定結果

検証データ		推定構造数		
		非木造 [棟]	木造 [棟]	合計
正解値	非木造 [棟]	65	45	191
	木造 [棟]	25	164	107
	合計 [棟]	153	145	299

(2) 構造推定に寄与した特徴量

本研究では 37 種類の属性を、建物構造を説明するための特徴量とした。そこで、これらの特徴量の中から損失関数である logloss の減少に寄与した程度を重要度とした。図-11、図-12、図-13、図-14 にそれぞれ研究対象地域全域、世田谷区のみ、渋谷区のみ、足立区のみの特徴量上位 10 位の変数を示す。なお、各図中の赤色の変数が既存統計に基づいて作成した特徴量、青色の変数が建物データから得られた特徴量を示している。これらの結果から、既存統計に基づいて作成した特徴量も構造推定にかなり寄与していることが分かった。特に延床面積や、建物周長といった建物データから得られる変数だけではなく、町丁字ごとに算出した居住者の居住期間の割合や推定年収など、その地域に住んでいる居住者の属性を表す変数も、建物構造を推定する上で重要な情報であることがわかった。

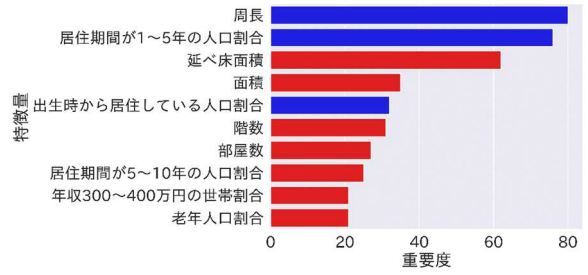


図-11 対象地域全域における特徴量の重要度

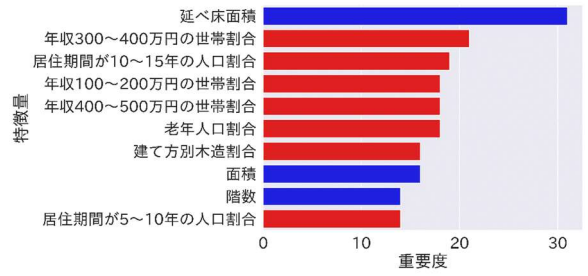


図-12 世田谷区での構造推定における特徴量重要度

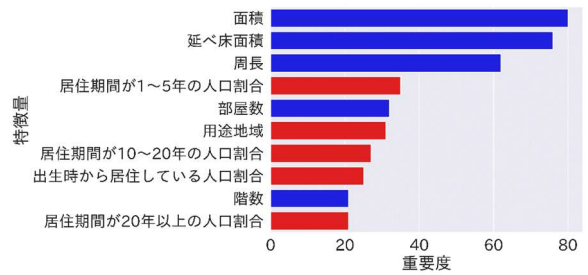


図-13 渋谷区での構造推定における特徴量重要度

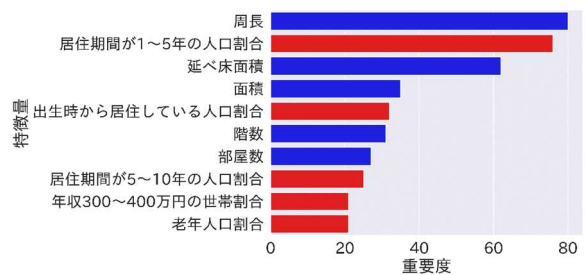


図-14 足立区での構造推定における特徴量重要度

4. 結論と今後の展望

本研究では、機械学習を用いて建物ごとの構造を推定する手法を開発した。まず、世田谷区、渋谷区、足立区で公開されている建物現状構造図を用いて教師データの作成を行った。次に、国勢調査や住宅・土地統計調査という既存統計を用いて推定を行うための特徴量を町丁字単位で作成した。そして、作成した教師データ、特徴量を用いて機械学習モデルを構築し学習を

行うことで建物ごとの構造を推定し、信頼性の検証を行った。その結果、教師データを作成した世田谷区、渋谷区、足立区の 3 区において一定の精度で建物構造を木造・非木造の 2 クラスに分類することができた。また、特徴量の重要度を算出した結果、既存統計を基に作成した特徴量も構造推定に大いに寄与しており、その地域に住んでいる居住者の属性が建物構造を推定する上で有用であることが分かった。さらに、自治体によって推定精度に差が出ることも明らかとなった。

今後は、居住者の属性に関する統計値や建物データの分布などにに基づき、どのような地域において本研究で作成した特徴量が有効なのか、また精度が低かった地域ではさらにどのような情報が必要なのかといった検討を行うことで、研究対象地域全域での精度の向上を目指す。また、本研究では東京都の 3 区のみを対象に研究を行ったが、建物単位の構造に関する情報を入手できる他の地域を調査し、地理的により広域かつ多様な特徴を持つ地域の学習データを構築することで、モデルの性能と汎用性の向上を実現したい。

謝辞：本研究は東大 CSIS 共同研究 (No.884) の一環として実施した (利用データ：Zmap TOWN II (2016 年度 Shape 版) 東京都データセット)。ここに記して謝意を表したい。

参考文献

- 1) 国土交通省：屋内外をシームレスに繋ぐ避難訓練シミュレーション<<https://www.mlit.go.jp/plateau/use-case/disaster-management/2-004/>>, (最終閲覧日 2022 年 2 月 28 日)
- 2) 国土交通省：時系列浸水シミュレーションデータの 3D 可視化による防災計画立案・防災意識啓発<<https://www.mlit.go.jp/plateau/use-case/disaster-management/2-003/>>, (最終閲覧日 2022 年 2 月 28 日)
- 3) 岡田 成幸, 高井 伸雄：地震被害調査のための建物分類と破壊パターン, 日本建築学会構造系論文集, Vol.64, 524 号, pp.65-72, 1999
- 4) 総務省：個人情報適切な取り扱い<https://www.soumu.go.jp/main_sosiki/gyoukan/kanri/question05.html#q5-3>, (最終閲覧日 2022 年 3 月 4 日)
- 5) 尾崎平, 盛岡通：建物属性の推定に基づく地震・津波被害による廃棄物発生量の推計, 廃棄物資源循環学会研究発表会講演集, Vol.27, pp.153, 2016
- 6) Li, Y., Chen, Y., Rajabifard, A., Khoshelham, K. and Aleksandrov, M. :Estimating Building Age from Google Street View Images Using Deep Learning, *10th International Conference on Geographic Information Science (GIScience 2018)*, pp. 40:1-40:7, 2018.
- 7) Rosser, J.F., Boyd, D.S., Long, G., Zakhary, S., Mao, Y. and Robinson, D. :Predicting residential building age from map data *Computers, Environment and Urban Systems*. Vol.73, pp. 56-67, 2019.
- 8) 小川芳樹, 沖拓弥, 関本義秀, 柴崎亮介：不動産に関する地理空間情報と建物画像を用いた建物の構造・築年代推定手法の検討, 「第 34 回人工知能学会全国大会論文集」, 2P6-GS-13-01, 2020.
- 9) Zhou, P. and Chang, Y. :Automated classification of building structures for urban built environment identification using machine learning, *Journal of Building Engineering*, Vol.43, 103008, 2021.
- 10) 白木洋平, 近藤昭彦, 一ノ瀬俊明：GIS とリモートセンシングを用いた地表面構造が都市の温度形成に及ぼす影響評価, 環境科学学会誌, Vol.20, pp.347-358, 2007.
- 11) 井城衣真, 須藤倫, 吉野博, 村上周三, 坊垣和明, 松縄堅, 亀谷茂樹, 高口洋人, 半沢久, 奥宮正哉, 浅野良晴, 下田吉之, 村川三郎, 依田浩敏：事務所建物のエネルギー消費原単位の推定に関する研究：非住宅建築物の環境関連データベースにおける平成 21 年度調査データによる分析, 「日本建築学会環境論文集」, Vol.77, pp203-211, 2012
- 12) 奥村 航太, 有村 幹治, 浅田 拓海：マイクロジオデータと機械学習手法を用いたメッシュ間 OD 交通量の推定, 土木学会論文集 D3 (土木計画学), Vol.76, pp. I_919-I_926, 2021.
- 13) 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.
- 14) Rusdah, D.A. and Murfi, H.: XGBoost in handling missing values for life insurance risk prediction, *SN Appl. Sci.* 2, 1336, 2020.
- 15) Latief, M. A., Bustamam, A. and Siswantining, T. :Performance Evaluation XGBoost in Handling Missing Value on Classification of Hepatocellular Carcinoma Gene Expression Data, *2020 4th International Conference on Informatics and Computational Sciences (ICI-CoS)*, pp. 1-6, 2020.
- 16) Scikit-learn : scikit-learn 1.0.2 documentation,< https://scikit-learn.org/stable/modules/generated/sklearn.metrics.log_loss.html> (最終閲覧日 2022 年 3 月 4 日)
- 17) Rodriguez, J. D., Perez, A. and Lozano, J. A. :A study of cross-validation and bootstrap for accuracy estimation and model selection, *Proceeding of the international joint conference on artificial intelligence*. Vol. 32, pp. 569-575, 2010.

(Received ?)
(Accepted ?)

DEVELOPMENT OF A STRUCTURAL ESTIMATION METHOD OF INDIVIDUAL BUILDING USING OPEN DATA

Naoya TAKEDA, Takafumi FURUYA and Yuki AKIYAMA

In recent years, 3D models of cities have begun to be used for disaster prevention planning and simulation of evacuation drills and so on. On the other hand, information on not only the shape of buildings but also the structure of each building is important for disaster prevention planning and urban planning in preparation for a large-scale earthquake. However, information on the structure of individual buildings is often difficult to obtain. In this study, we developed a method for estimating building structure using XGBoost, a machine learning method, with various attribute values obtained from existing statistics such as the population census and various spatial information readily available as open data for the Setagaya, Shibuya, and Adachi wards of Tokyo. The reliability of the estimation was verified, and the results showed that the structure could be estimated with an accuracy of approximately 80% in all wards. Furthermore, the importance of the features was calculated, and it became clear that the features created based on existing statistics also contributed greatly to the structure estimation.