

オープンデータと機械学習を用いた建物ごとの構造推定手法の開発

武田直弥*・古谷貴史**・秋山祐樹***

Development of Estimating Method for Building Structure by Using Open Data and Machine Learning

Naoya TAKEDA Takafumi FURUYA Yuki AKIYAMA

Abstract: In order to proceed with housing earthquake resistant for massive earthquakes, municipalities need to estimate and understand damage caused by the earthquake before occurring them. In this process the structure of each building is important information. However, it is often difficult to obtain information of the structure of each building. In this study, we integrated various geospatial information collected from open data such as the national census to each building data and used “XGBoost”, a method of machine learning, to estimate the structure of building. In addition, we verified estimated result in Setagaya ward, Tokyo. As a result, it showed that our method can estimate the structure of the building with accuracy of about 92%

Keywords: 耐震化 (earthquake resistant) , オープンデータ (open data) , 機械学習 (machine learning) 構造推定 (estimating structure)

1. はじめに

南海トラフや首都直下型地震などの大規模地震が最大クラスの規模で発生した場合、東日本大震災を超える被害が発生することがほぼ確実とされている。しかし国土交通省（2020）の発表によると、平成30年の時点で日本全国の住宅の耐震化率は約87%とされており、令和2年度までに95%を達成するという従来の目標よりも遅れている状況にあり、住宅の耐震化の取り組みを全国的に進めていくことは急務といえる。

このような巨大地震に備えた住宅の耐震化を全国的に進めていくためには、自治体単位での取り組みが重要となる。事実、令和2年の時点で全国の98.0%（1,707自治体）の自治体が耐震改修促進計画を策定済みである（国土交通省、2020）。しかし、このような取り組みを自治体レベルを進めためには、自治体が事前に地震による被害状況を

把握・推定した上で、どの地域から優先的に耐震化を進めていくべきかや、将来的に防災性を考慮しながら居住誘導を進めていくべきか、などの計画を立てていく必要がある。その際に建物の耐震性や防火性を決める重要な要素である建物ごとの構造に関する情報は重要である。（国土交通省、2021）。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く、また自治体内においても建物ごとの構造の情報は目的外使用が困難な場合が多いため、建物構造に関する情報を推定・把握する手法は重要である。

建物構造を知るための既存研究としては、小川ほか（2020）は不動産ポイントデータから得られる建物構造を正解データとし、不動産ポイントデータ及び建物画像データを用いて建物の外観から建物構造の推定を行った。しかし同研究で用いた不動産ポイントデータは一般的に入手が困難

* 学生会員 東京都市大学工学部都市工学科（Tokyo City University）
〒158-0087 東京都世田谷区玉堤1丁目2-8-1 Tel：03-5707-0104 E-mail：g1818055@tcu.ac.jp

** 学生会員 東京都市大学大学院総合理工学研究科（Tokyo City University）

*** 正会員 東京都市大学建築都市デザイン学部都市工学科（Tokyo City University）

であったり、画像データは建物全体が写っている必要があったりするなどデータの入手や整備などに課題があった。他にも、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。しかし、同研究では信頼性の検証が十分でないという課題があった。

以上を鑑み、本研究ではオープンデータである国勢調査などの各種統計から得られた様々な地理空間的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“XGBoost”を用いて建物ごとの構造推定を行うとともに、信頼性の検証を行った。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を対象とする。

2. 研究手法

2.1. 教師データの作成

まず、世田谷区の建物データ(2016年の住宅地図(Zmap TOWN II:株式会社ゼンリン)を使用)から、住宅用途の建物を1,000件抽出した。続いて、抽出した1,000件の建物に建物構造の情報を与える。本研究の対象地域である東京都世田谷区では、「せたがや iMap (世田谷区電子地図情報配信サービス)」により、世田谷区全域の建物の分布と建物ごとの構造や耐火性能などを地図で閲覧することができる。そこで、先に抽出した1,000件の建物の構造をせたがや iMap 上で1棟1棟目視により確認することで、建物ごとの構造の真値を与えた。以上の処理により機械学習に用いる教師データを作成した。表1に教師データの建物用途別の構造別建物数を示す。

2.2. オープンデータの加工の加工と建物データへの属性付与

本研究では構造推定を行うためにオープンデータでから得られる以下に示す属性を建物データに空間結合することで付与した。

1. 各年齢階級人口の割合 (3区分:小地域集計)
2. 居住期間の割合 (6区分:小地域集計)

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	522	255
住商混合ビル	4	13
共同住宅	72	134
合計	598	402

3. 年収階級別推定世帯割合 (9区分:小地域集計)
4. 住宅の建て方別の各構造割合 (3区分:小地域集計)
5. 用途地域 (13区分:用途地域ごとのポリゴンデータ)

「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した。これは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。

「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。これは居住期間が長い傾向にある地域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。

「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44-4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ほど算出した割合に基づいて年収階級別の世帯数を決定し丁町字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などの傾向が推定できると考えられるため追加した。

「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、次に国勢調査の住宅の建て方別の世帯数に対して先ほど算出した割合に基づいて各構造の建物数を決定し、丁町字ごとの住宅の建て方別の各構造割合

を追加した。

「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。

以上の属性は国勢調査（2015年）と住宅土地統計調査（2018年）の2つのオープンデータを加工することで作成できるため、データの入手が容易であるだけでなく、コストもかかりにくいという点で優れている。

2.3. 構造推定の手法

本研究では2.2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ため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（白木ほか、2007；井城ほか、2012）。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（奥村ほか、2021）。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)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、多数決を取ることによって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、欠損値があるデータも処理が可能という点が優れている。

本研究のモデルではモデルの学習を行った後、建物が非木造に分類される確率を算出し、その確率が0.5を超えるものを非木造と推定する手法で

表 2.世田谷区における検証用データの推定結果

検証データ		推定構造数		
		非木造 [棟]	木造 [棟]	合計
教師 データの 建物数	非木造 [棟]	23	11	34
	木造[棟]	5	161	166
	合計	28	172	200

モデルの作成を行い、また、モデルのパラメータチューニングは交差検証によって実施し、logloss (Logarithmic Loss) を評価指標に用いた。

3. 構造推定の結果と信頼性の検証

3.1. 検証用データを用いた信頼性の検証

まず、本研究で使用したモデルの結果について信頼性の検証を行う。本研究では予め訓練用のデータと検証用のデータを8:2の比率で分割しておき、訓練用のデータのみでモデルの学習を行った。表2は検証用のデータを用いて、モデルの推定結果と2.1で作成した正解データとを集計したものである。この結果表から正答率、F値を算出するとそれぞれ92.00%、0.7519となりこのモデルを用いることで建物構造を全体としては高い精度で推定することができた。しかし、非木造の建物を木造と推定した建物数が多かった。これは学習用データのサンプル数とパラメータチューニングが最適化されていないことが原因と考えられるため今後の課題である。

以上のように世田谷区の建物については正答率、F値ともに高い精度で推定することができたが、モデルの汎用性を検証するため、次節においてほかの自治体において広域で検証を行う。

3.2. 兵庫県神戸市における検証

兵庫県神戸市は神戸市都市計画基礎調査建物構造集計データ（2014年）を公開しており、同データを用いることで、町丁字ごとの木造率および非木造率を把握することが出来る。そこで、まず

3.1 で用いたモデルを使って構造推定を行った兵庫県神戸市の建物データを、住宅用途の建物が存在する町丁字を対象に集計し丁町字ごとの木造率を算出した。続いて、それらを神戸市都市計画基礎調査建物構造集計データから得られる丁町字ごとの木造率と比較するという手順で検証を行った。また、本研究で作成したモデルは住宅用途の建物のみで学習を行っているため、兵庫県神戸市を対象とした構造推定を行う際は住宅用途以外の建物は非木造として集計した。

図1は丁町字ごとに横軸に真値の木造率を、縦軸に推定値の木造率をそれぞれプロットした結果である。相関係数は0.46という結果であり、神戸市全体としてはそれほど高い精度ではなかった。これは、非木造建物が多い丁町字、即ち住宅用途以外の建物が多い丁町字において結果に大きなばらつきが生じてしまっていることが原因であると考えられる。

続いて、図2に、真値の木造率と推定値の木造率の差を取った結果を可視化した。この結果から市街地や住宅地では正しく推定できているが、それ以外の地域では精度に問題があることが分かった。これは本研究で開発したモデルが世田谷区で学習したモデルであり、世田谷区はほぼ全域が市街地及び住宅地で構成されているため、市街地や住宅地以外の地域においては、本研究で構築したモデルでは木造率を正しく推定できなかったためである。また、今回は住宅用途の建物が存在する地域を対象に検証を行ったがモデルの性能をより正確に確かめるためには、住宅用途の建物が一定割合以上の地域に限定して検証を行うなど検証方法にも改善の余地がある。

4. まとめと今後の展望

本研究では、国勢調査や住宅土地統計調査といったオープンデータから作成した地理空間的情報を建物データに付与し、機械学習を用いることで建物1棟ごとの建物構造の推定を行った。その結果、東京都世田谷区では入手が容易なオープンデータから高い精度で高い精度で建物構造を推

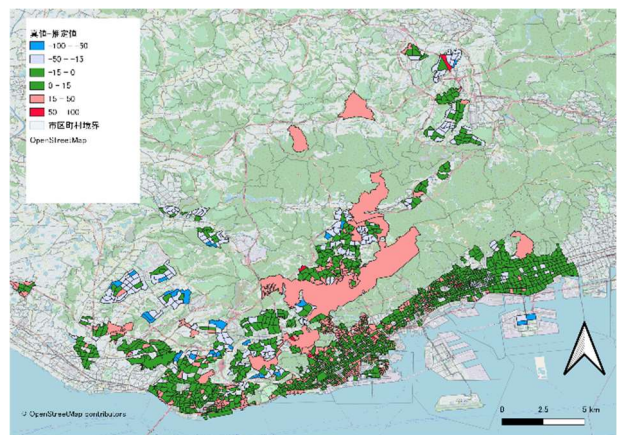


図1 兵庫県神戸市の木造率

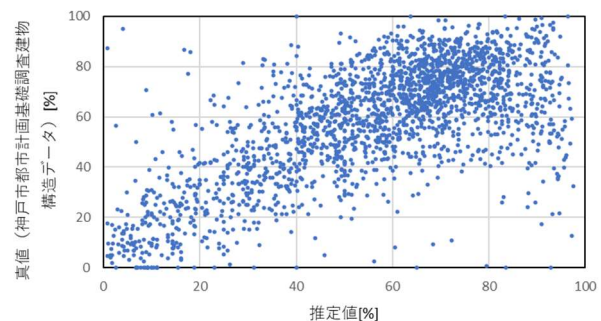


図2 兵庫県神戸市の推定結果

定することが可能になった。

一方、兵庫県神戸市における検証では地域によって推定精度に大きく差が出る結果になったため、モデルの汎化性能を高める方法の検討が必要である。そのための改善点として本研究では訓練データと検証用データを8:2に分割したが、この割合を変化させ様々な割合を試していくことや、各構造に分類される確率の閾値の検討などが考えられる。また、神戸市における推定で正しく推定できていなかった地域は用途地域が指定されていない地域が多かったためそのような地域に対して精度を上げる方法の検討が求められる。

今後は、先ほど述べた点を改善していくと共に地方都市などでも世田谷区で作成したモデルの信頼性を検証していく。また、世田谷区以外の地域で建物単位の構造に関する情報を入手することで、都市部や住宅地以外の学習データを構築し、モデルの性能向上を実現する予定である。

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・兵庫県データセット). ここに記して謝意を表したい.

参考文献

- 井城衣真・須藤愉・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
(2012) 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」, 77(673), 203-211
- 小川芳樹・沖拓弥・関本義秀・柴咲亮介 (2020) 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
- 尾崎平・盛岡通 (2016) 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, 「第 27 回廃棄物資源循環学会研究発表会講演集」, A11-3
- 国土交通省 (2018) 耐震診断義務付け建築物に係る耐進化の促進について<<https://www.mlit.go.jp/common/001232674.pdf>> (最終閲覧日 2021 年 8 月 28 日)
- 国土交通省住宅・建築物の耐震化率のフォローアップのあり方に関する研究会 (2020) 住宅・建築物の耐震化率の推計方法及び目標について<<https://www.mlit.go.jp/common/001345338.pdf>> (最終閲覧日 2021 年 8 月 28 日)
- 国土交通省住宅局建築指導課 (2020) 参考 1. 市区町村の耐震改修促進計画の策定状況 (都道府県別総括表) <<https://www.mlit.go.jp/report/press/content/001375549.pdf>> (最終閲覧日 2021 年 8 月 28 日)
- 白木洋平・近藤昭彦・一ノ瀬俊明 (2007) 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」, 20(5), 347-358

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