

事業所系建物の構造推定を行う機械学習モデルの開発

水田京佑*・武田直弥**・秋山祐樹***・佐藤亮吾****・稲村友彦****

Development of a Machine Learning Model for Structural Estimation of Business Facilities

Keisuke MIZUTA*, Naoya TAKEDA**, Yuki AKIYAMA***,
Ryogo SATO****, Tomohiko INAMURA****

It is important to estimate the damage caused by natural disasters not only to residential buildings but also to business facilities such as retail stores and office buildings, which are indispensable to the daily lives of residents to improve the disaster preparedness of the community. However, there are only a few previous studies that have conducted structural estimation for business facilities. Therefore, we developed a method to estimate the structure of business facilities by developing a database that integrates existing building data with building features and various statistical information obtained from insurance policy data, and by building a machine learning model using the data. The estimation accuracy reached approximately 77% for the special wards of Tokyo.

Keywords: 構造推定 (structural estimation), 機械学習 (machine learning), 防災 (disaster prevention), 事業所系建物 (business facilities), 保険契約データ (insurance policy data)

1. はじめに

近年、我が国では自然災害の激甚化に伴い、自然災害による経済的被害が多発している。このような経済的被害の評価における建物被害の評価では、住宅の被害が注目されがちであるが、事業所系の建物への被害も無視できない。国土交通省(2019)によると、平成30年7月豪雨の被害額は約1兆1,580億円と報告されている。また、中小企業庁(2019)によると、同じく平成30年度7月豪雨では中小企業において4,738億円の経済的被害が出ていることが報告されている。すなわち、平成30年7月豪雨の被害額のうち、約4割が中小企業における経済的被害であることが分かる。そのため、自然災害における被害推定では、住宅の被害だけではなく、事業所系の建物についての被害推定も重要である。

1.1 既存研究

自然災害に対する損害額を推定し被害を見積も

る際には、建物ごとの構造に関する情報が重要な要素となる。(佐伯・翠川, 2001) また、生活支障の予測を行う場合にも、建物構造に基づく建物の被害関数を用いて想定を行うことが有効である(村尾・山崎, 2002)。しかし、建物ごとの構造に関するデータは原則非公開であり、国や自治体が課税目的等で徴収した建物構造等に関するデータを入手するには、多大な時間やコストがかかってしまうという問題がある。

そこで、住宅の建物構造を推定する取り組みがこれまでに行われている。例えば、建物の外観画像を用いて、戸建住宅の建物構造の自動分類を行った例がある(井上, 2021)。ただし、建物外観画像を用いた構造推定の課題として、木や車両などといった建物以外の物体が画像内に映り込んでしまうことにより、推定結果が大きく影響を受けてしまう点が挙げられる。また、沖・小川(2021)は外観画像と建物に関する属性情報(建物面積、階数等)を用いて、

* 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28-1 Tel: 03-5707-0104 Email: g1918081@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)

**** 非会員 S O M P O リスクマネジメント株式会社 (Sompo Risk Management Inc.)

居住用賃貸と事業用賃貸の構造推定を行っている。同手法では不動産データベースから得られる建物外観画像と、建築面積や階数に関する属性データを教師データとする機械学習を用いて、木造、S造、RC造の三区分に分けた構造推定を行っている。その結果、様々な建物属性情報を説明変数として用いることによって、建物外観画像だけで属性を推定する手法よりも精度が向上していることが分かった。この結果からも、建物に関する属性情報が建物構造を推定する上で重要な情報であることが分かる。なお、同研究で用いる建物外観画像は、建物を正面から撮影した画像を使用しないと推定精度に影響が出てしまうため、画像の準備に時間を要してしまうという課題がある。

1.2 本研究の目的

以上の研究背景および既存研究の課題を踏まえ、本研究では事務所系の建物について建物構造推定を行う手法を開発する。具体的には、建物外観などの画像情報は一切用いることなく、国勢調査や住宅・土地統計調査、経済センサスなどのオープンデータである統計資料と、損害保険会社から提供を受けた損害保険契約情報から得られる建物に関する様々な情報を特徴量とする機械学習を用いることで、事務所系建物の構造推定を行う。なお、研究対象地域は東京23区と兵庫県神戸市とした。

2. 研究手法

本研究では、前述の通り事業所系の系建物に対象を絞って建物の構造の推定を行う。建物構造は建物構造級を読み替えることで得られる鉄筋コンクリート造（以下RC造）、鉄骨造、木造としたモデルを作成した。

図1に構造推定モデルの構築方法の概要を示す。まず、損害保険会社から提供を受けた損害保険契約情報（以下「保険契約データ」）から得られる建物に関する様々な情報（以下「建物特徴量」）に様々な統計資料を付与した。次に全データの2割をテストデータとし、残り8割を訓練用データに分割する。続いて、訓練用データをさらに5分割して、1つを

検証用データとし、残り4つを学習用データとする。そして、検証用データになる部分が被らないように5回学習を行い、それぞれのF値と正解率を求め、5回の結果の平均を取る。最後に学習させたモデルを用いて最初に分けた2割のテストデータでテストを行い、構造推定モデルの精度評価を行う。

2.1 教師データの作成

本研究では損害保険ジャパン社保有の匿名化処理済み保険契約データから得られる建物ごとの構造に関する情報を教師データとした。表1に東京23区と兵庫県神戸市における構造別の建物数を示す。

2.2 説明変数の作成

本研究では、損害保険ジャパン社の保有している保険契約データから得られる建物の階数や延床面積などの建物に関する情報と、国勢調査や住宅・土地統計調査から得られる年代別の人口や居住期間などの居住者に関する情報、経済センサス得られる従業者数や事業所数などの事業所に関する情報などを説明変数に用いた。以上の統計データはいずれもオープンデータであるため、全国に適用可能であるという利点がある。

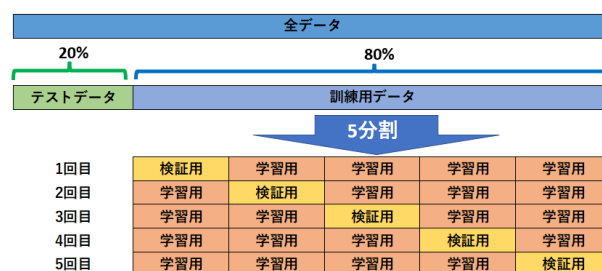


図1 構造推定モデルの構築方法の概要

表1 東京23区と兵庫県神戸市における構造別の建物数

建物構造	東京都の建物数[棟]	兵庫県神戸市の建物数[棟]
RC造	28,359	1,325
鉄骨造	11,218	828
木造	13,215	646
合計	52,792	2,799

表 2 に建物構造推定に用いた説明変数の一覧を示す。以下、説明変数について紹介する。

(1) 年齢階級別人口割合

2015 年の国勢調査小地域集計第 3 表から年少人口割合、生産年齢人口割合、老年人口割合を算出し、作成した。居住者の年齢が高い方がその地域において建物の入れ替えが少なく、木造建物が多いと考え、説明変数に加えた。

(2) 居住期間別世帯割合

2015 年の国勢調査小地域集計第 13 表から作成した。居住期間が長いほど建物の入れ替えが少ないため木造建物が多いと考え、説明変数に加えた。また、居住期間が短い場合にはその地域が再開発された可能性もあり、非木造建物である可能性が高いものと期待される。

(3) 年収階級別世帯割合

2015 年の国勢調査小地域集計第 7 表と 2018 年住宅・土地統計調査第 44-4 表から作成した。住宅・土地統計調査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、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・土地統計調査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し、町丁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、その地域の建物の建て方の傾向に影響が出るものと考えられるため説明変数に加えた。

(4) 建て方別構造割合

国勢調査小地域集計第 8 表と住宅・土地統計調査第 8-3 表を用いて作成した。まず、住宅・土地統計調査から市区町村毎の住宅の建て方別の各構造割合を算出した。算出した割合を町丁字ごとの建物数に乗じることで町丁字ごとの建て方別の各構造割合を算出した。これは各構造割合を出すことによって、地域の建物の特徴を把握することで判定をしやすくするために、説明変数に加えた。例えば、木造の割合が高い場合、その地域の建物は木造である可能性が高いと考えられる。

表 2 建物構造推定に用いた説明変数一覧

説明変数	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)	国勢調査小地域集計 (2015年) 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、建物用途、総部屋数など
7.従業者数割合 (4区分)	経済センサス-基礎調査- (2014年)
8.経営組織別従業者数割合 (7区分)	経済センサス-基礎調査- (2014年)
9.経営組織別事業所数割合 (7区分)	経済センサス-基礎調査- (2014年)
10.廃業・存続・新設別従業者数割合 (3区分)	経済センサス-基礎調査- (2014年)
11.廃業・存続・新設別事業所数割合 (3区分)	経済センサス-基礎調査- (2014年)
12.資本金階級ごとの廃業・存続・新設別従業者数割合 (30区分)	経済センサス-基礎調査- (2014年)
13.資本金階級ごとの廃業・存続・新設別事業所数割合 (30区分)	経済センサス-基礎調査- (2014年)
14.資本金階級別廃業・存続・新設に分けた際の全体総数を分母とした従業者数割合 (30区分)	経済センサス-基礎調査- (2014年)
15.資本金階級別廃業・存続・新設に分けた際の全体総数を分母とした事業所数割合 (30区分)	経済センサス-基礎調査- (2014年)

(5) 用途地域

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、用途地域はラベルエンコーディングをして付与した。これは用途地域による建築物の制限があり、用途地域を知ることによって建物の特徴が似たものを表現できると考え、説明変数に加えた。

(6) 建物データ

保険契約情報から得られる建物の特徴量と本研究で用いた建物データ（建物ポイントデータ：株式会社ゼンリン）から得られる属性値を特徴量として活用した。なお、保険契約情報から得られる属性として建物の階数や面積、建物用途などの情報が挙げられる。

(7) 従業者数割合

2014年の経済センサス-基礎調査-町丁・大字別集計第1-13表を用いて作成した。町丁字ごとの性別従業者の割合と1事業所あたりの従業者数を算出した。これは、従業者の男女比や従業者数をから職業の判定ができるものと考え、説明変数に加えた。

(8) 経営組織別従業者数割合

2014年の経済センサス-基礎調査-の町丁・大字別集計第2-13表を用いて作成した。町丁字単位で経営組織別割合を算出した。これは、経営組織別の従業者数の割合を知ることによって、建物の規模（面積、階数等）を予測でき、建物の構造に違いが出ると考え、説明変数に加えた。

(9) 経営組織別事業所数割合

2014年の経済センサス-基礎調査-の町丁・大字別集計第2-13表を用いて作成した。町丁字単位で経営組織別の事業所数割合を算出した。これは、経営組織別の事業所割合を把握することによって、建物の規模が予測でき、建物の構造の推定ができると考え、説明変数に加えた。

(10) 廃業・存続・新設別従業者数割合

2014年の経済センサス-基礎調査-の事業所に関する集計都道府県別結果第35-2表を用いて作成した。市区町村単位で廃業・存続・新設別の事業所の従業者数割合を算出した。これは、廃業・存続・新設の従業員の割合を知ること、事務所の規模がわかると考え、説明変数に加えた。

(11) 廃業・存続・新設別事業所数割合

2014年の経済センサス-基礎調査-の事業所に関

する集計都道府県別結果第35-2表を用いて作成した。市区町村単位で廃業・存続・新設別の事業所数割合を算出した。これは、廃業・存続・新設の割合を知ることによって建物の新旧がわかると考え、説明変数に加えた。

(12) 資本金階級ごとの廃業・存続・新設従業者数割合

2014年の経済センサス-基礎調査-の事業所に関する集計都道府県別結果第35-2表を用いて作成した。市区町村単位で資本金階級（10区分）に区分した後に区分内での廃業・存続・新設の従業者数割合を算出した。これは、事業所の資本金別に区分することで雇える従業者数が予測でき、必要な事業所の大きさがわかると考え、説明変数に加えた。また、資本金が建物構造に影響を与えるのか確認するため、資本金階級ごとで割合を算出した。

(13) 資本金階級ごとの廃業・存続・新設事業所数割合

2014年の経済センサス-基礎調査-の事業所に関する集計都道府県別結果第35-2表を用いて作成した。市区町村単位で資本金階級（10区分）に区分した後に区分内での廃業・存続・新設の事業所数割合を算出した。これは、事業所の構造が資本金の大きさによって建物構造に影響を与えるのか確認するため、資本金階級ごとで割合を算出した。

(14) 資本金階級別廃業・存続・新設に分けた際の全体総数を分母とした従業者数割合

2014年の経済センサス-基礎調査-の事業所に関する集計都道府県別結果第35-2表を用いて作成した。市区町村単位で資本金階級（10区分）に区分した後に資本金階級別の廃業・存続・新設別の従業員数を分子とし、総数を分母としたときの廃業・存続・新設の従業者数割合を算出した。これは、資本金区分において総数を分母とした際の従業員数の割合を知ること、市区町村単位で地域の事業所についての特徴が出ると考え、説明変数に加えた。

(15) 資本金階級別廃業・存続・新設に分けた際の全体総数を分母とした事業所数割合

2014年の経済センサス基礎調査の事業所に関する集計都道府県別結果第35-2表を用いて作成した。市区町村単位で資本金階級(10区分)に区分した後、資本金階級別の廃業・存続・新設別の事業所数を分子とし、総数を分母としたときの廃業・存続・新設の事業所数割合を算出した。これは、事業所の資本金ごとの廃業・存続・新設の割合を算出することで地域の特徴がわかると考え、説明変数に加えた。

2.3 建物構造の推定手法

続いて、2.2で作成した説明変数を特徴量とした機械学習により、建物構造の推定を行う。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はロジスティック回帰(齊藤, 2021)、線形的に分類することが難しい場合にはサポートベクターマシン(内藤ほか, 2020)を用いることが多い。一方、本研究で作成した説明変数は統計調査が実施されていない地域で欠損値が生じてしまう可能性があるため、これらの既存研究で用いられた手法は適用できない。なお、沖・小川(2021)のように教師あり学習のlasso法を使用して、不要な説明変数を削減する方法もある。しかし、本研究ではどの説明変数が目的変数に対して影響を与えるかを確認するため、使用しなかった。

以上を踏まえて、本研究では欠損値を扱うことのできる機械学習手法の1つである、勾配ブースティングアルゴリズムを用いているLightGBM(Light Gradient Boosting Machine)を用いて学習モデルを構築した(Guolin et al, 2017)。この手法は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。

本研究では各クラス(構造)に分類される確率を算出し、その確率が最も大きかったクラスに分類するモデルを作成した。また、学習を行う際は学習データを目的変数に偏りが生じないように5分割し、そのうち1つを検証用データに用いて学習を行うことで過学習が生じないように対策を行った。

本研究では、まず、東京23区のデータで学習させたデータを作成する。その後、東京23区のデータで学習させたモデルが他の地域でも実装することが可能かどうかの検証を行うために、東京23区で学習させたモデルを兵庫県神戸市のデータに適用させて検証を行った。

3.推定結果と検証

3.1 東京23区の推定結果

表3に学習に用いていないテストデータに対する推定結果を示す。表3から東京23区において建物構造推定の正解率とF値はそれぞれ77.9%と72.1%という精度で推定が可能となった。RC造と木造の推定は高い精度で可能であり、RC造の正解率は89.6%、木造の正解率は81.5%であった。一方で鉄骨造の正解率は44.3%であった。真値が鉄骨造の建物をRC造と木造に誤分類してしまっていることが多かった。鉄骨造に関してより正確に分類できるように特徴量の作成を行う必要がある。

3.2 建物構造の推定に寄与した特徴量

3.1の建物構造の推定に寄与した特徴量の分析を行った。具体的には本研究で作成したモデルに用いた式(1)で定義される損失関数を、各決定木の分岐ごとにどのくらい減少させたかにより、特徴量の重要度を算出した。

表3 テストデータにおける東京23区の建物構造推定の結果

テストデータ		推定値			
		RC造	鉄骨造	木造	合計
真値	RC造	5080	360	232	5672
	鉄骨造	801	993	450	2244
	木造	229	260	2154	2643
	合計	6110	1613	2836	10559

$$ml = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M y_{ij} \log(p_{ij}) \quad (1)$$

N はデータ数、 M はクラス数、 y_{ij} は i 番目のデータがクラス j に含まれているかを表現する変数 (含まれている場合には 1, そうでない場合には 0), p_{ij} はクラス j に分類される確率を示しており, 本研究では各クラスに分類される確率を占めます. 本研究では各クラスに分類される確率を算出するモデルを作成するために, この損失関数を用いた.

図 2 に東京 23 区の建物構造の推定に寄与した上位 10 個の特徴量を示す. 基本保険金額と建物の階数が特徴量として特に重要であることが分かった. また, 延べ床面積や総部屋数, 建築年月などの特徴量が重要であることがわかった. 住宅・土地統計調査から作成した特徴量である建て方別構造割合も重要であることがわかった.

3.3 東京 23 区のデータを用いて作成したモデルを兵庫県神戸市に外挿した検証結果

続いて, 東京 23 区を用いて作成したモデルを兵庫県神戸市に外挿した結果を表 4 に示す. 表 4 から東京 23 区のデータを用いて学習させたモデルで兵庫県神戸市を外挿した際の建物構造推定の正解率と F 値は, それぞれ 67.8% と 64.1% であった. 東京 23 区の推定結果と比較すると正解率で約 10%, F 値で約 8% 低い結果となった. これは, 兵庫県神戸市の場合, 阪神・淡路大震災の影響で, 東京 23 区と比較して築年代が古い建物の建て替えが進んでいるため, 東京 23 区と兵庫県神戸市の間で建物の特徴が異なることが影響しているものと考えられる. さらに, 東京 23 区のデータ数が 52,792 棟であるのに対し, 兵庫県神戸市のデータ数は 2,799 棟であった. そのため兵庫県神戸市ではデータが少なかったことから推定結果が安定しなかった可能性も考えられる.

3.4 まとめ

東京 23 区と兵庫県神戸市の F 値および正解率を表 5 にまとめた. 東京 23 区における事業所建物の構造推定は高精度で推定することができた. ただし,

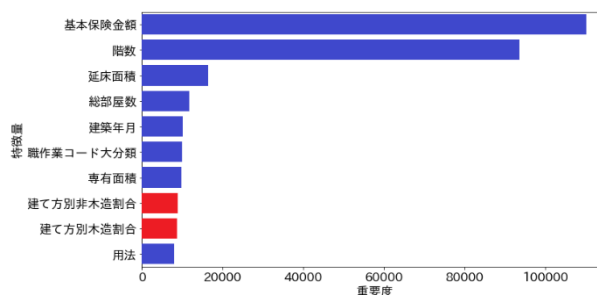


図 2 東京 23 区における建物構造推定における特徴量重要度

表 4 東京 23 区のデータを用いて学習したモデルを兵庫県神戸市に外挿した結果

テストデータ	推定値			
	RC造	鉄骨造	木造	合計
RC造	1051	117	157	1325
鉄骨造	227	309	292	828
木造	46	61	539	646
合計	1324	487	988	2799

表 5 東京 23 区および兵庫県神戸市における建物構造推定の結果 (F 値および正解率)

対象地域	F 値 (%)	正解率 (%)
東京都23区	72.1	77.9
兵庫県神戸市	64.1	67.8

鉄骨造の推定においては課題が残る結果となった. 今後, 鉄骨造の推定の精度向上を目指すために新たな説明変数の作成を行う必要がある.

また, 東京 23 区のデータを用いて学習したモデルを兵庫県神戸市に外挿した結果, 正解率, F 値ともに東京 23 区の結果と比べて低下した. この結果から, 東京都と兵庫県神戸市の建物の特徴が大きく異なる可能性が示唆された. 今後は兵庫県神戸市以外への外挿による外挿性の検証や, 兵庫県神戸市のデータ数を増やした外挿性の検証を行う必要がある.

4. おわりに

本研究では、保険契約情報から得られる建物の特徴量と、国勢調査をはじめとする各種統計情報から得られる建物周辺の人口動態等および事業所に関する情報を用いて、機械学習を活用することで事業所系の建物構造を推定する手法を開発した。

その結果、東京 23 区における事業所系の建物の構造推定は高精度で実施可能になったが、鉄骨造の推定においては課題が残る結果となった。今後、鉄骨造の推定の精度向上を目指すために新たな説明変数の検討を行う必要がある。また、モデルの外挿性については外挿先である兵庫県神戸市の建物数の拡充や、兵庫県神戸市以外への外挿を行うことで、外挿性の確認やその課題を明らかにしていく必要がある。

謝辞

本研究で用いた保険契約データは SOMPO リスクマネジメント株式会社との共同研究の枠組みにおいて提供を受けたものである。また、SOMPO リスクマネジメント株式会社にはデータセットの整備や分析環境の準備など多方面においてご協力頂いた。ここに記して謝意を表したい。

参考文献

経済産業省中小企業庁 (2019) 中小企業白書 第 3 部 第 2 章防災・減災対策 <https://www.chusho.meti.go.jp/pamflet/hakusyo/2019/PDF/chusho/05Hakusyo_part3_chap2_web.pdf> (最終閲覧日 2022 年 8 月 27 日)

国土交通省 (2019) 平成 30 年 7 月豪雨が統計開始以来最大の被害額に～平成 30 年の水害被害額 (暫定値) を公表～ <https://www.mlit.go.jp/report/press/mizukokudo03_hh_000985.html> (最終閲覧日 2022 年 8 月 27 日)

佐伯琢磨・翠川三郎 (2001) 住宅建物の経済被害評価のための被害関数の作成—兵庫県南部地震の際の地震保険支払データに基づいて—, 「日本建築学会構造系論文集」, 545, 79-85.

村尾修・山崎文雄 (2002) 震災復興都市づくり特別委員会調査データに構造・建築年をふかした兵庫県南部地震の建物被害関数, 「日本建築学会構造系論文集」, 555, 185-192.

井上雅史・森谷悠樹 (2021) 外観画像による戸建住宅の建物構造自動分類, 「2021 年度人工知能学会全国大会 (第 35 回)」, 4F3-GS-10n-02.

沖拓弥・小川芳樹 (2021) 不動産データベースの外観画像と属性情報を用いた建物構造・築年代推定モデル, 「日本建築学会計画系論文集」, 86(779), 357-367.

齊藤千尋, アクセシビリティ、独立性、最短路の分布による建物出現位置の予測-典型的な街区における建物配置のロジスティック回帰分析-, 「日本都市計画学会都市計画論文集」, 46(3), 397-402.

内藤昌平・友澤弘充・森悠史・門馬直一・中村洋光・藤村広行 (2020) 複数の地震における航空写真を用いた深層学習による建物被害判別モデルの開発, 「日本地震工学会論文集」, 20(7), 7_177-7_216.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T. (2017) LightGBM: A Highly Efficient Gradient Boosting Decision Tree, “NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems”, 3149–3157.