

不動産ポイントデータと機械学習を用いた建物の構造・築年代推定モデル

小川芳樹・沖拓弥・関本義秀・柴崎亮介

Model for Estimating Structure and Age of Each Building Using Real Estate Point Data and Machine Learning

Yoshiki OGAWA, Takuya OKI,

Yoshihide SEKIMOTO and Ryosuke SHIBASAKI

Abstract: To improve the spatial accuracy of earthquake damage estimation, it is essential to develop a building database with structure and building age information. In this study, we propose a method for estimating the structure and age of a building based on machine learning by using real estate data. First, the training data is created by spatially joining GIS data such as restricted zones and roads with real estate point data (with structure and building age). Next, we build a model that estimates the structure and age of each building using the features learned from the training data. Finally, the accuracy of our models are verified checked by cross-validation.

Keywords: 不動産ポイントデータ (real estate point data), スパースモデリング (sparse modeling), 建物属性 (building attribute), 被害推定 (damage estimation)

1. はじめに

建物の構造と築年代は、大地震時の建物倒壊確率に関わる大きな要因であり、現地調査結果に基づく建物倒壊確率モデル（例えば、村尾・山崎（2000））のパラメータとして用いられている。すなわち、詳細な防災計画を検討する上では、建物1棟ごとの構造や築年代を把握することが重要であるが、そうした詳細な属性情報を都市内の全建物で有するGISデータは一般には存在しない。

そこで本稿では、構造や築年代の情報を含む不動産（賃貸住宅・店舗）ポイントデータを用いて、

建物属性や立地特性から構造と築年代を推定するための機械学習モデルを構築し、推定精度を検証する。具体的には、深層学習などで応用されるスパースモデリング(SpM)を、建物構造と築年代の分析に用いる不動産ポイントデータに適用し、スパースLogitモデルを提案する。

2. 手法の概略

2.1 データ整備

本研究で用いる不動産ポイントデータ(株式会社アットホーム(2015年))は、データが5つのタイプ(販売マンション、賃貸マンション、賃貸アパート、事業所、販売戸建)別に分かれている(表-1)。各データには、以下に示すような112に及ぶ詳細な属性が含まれている。

- ・緯度経度
- ・構造、築年数

小川芳樹：〒151-8505 東京都目黒区駒場 4-6-1

東京大学生産技術研究所

Phone: 03-5452-6412

E-mail: yogawa@iis.u-tokyo.ac.jp

- ・階数，土地利用用途，近隣バス停，鉄道駅など
- ・設備情報（オートロックの有無など 99 種類（以下「機能」として記載））

本稿では，東京都内のデータを分析対象とする（図-1）．また，ZMAP TOWN II データをもとに作成した道路中心線データと各建物ポイントを空間結合し，建物近隣の道路情報（幅員，緊急輸送道路の指定有無，緊急輸送道路までの距離）を付与した．緊急輸送道路の沿道建物は重点的に耐震化が進められており，構造や築年代の分布に影響すると予想される．

以上の前処理の結果，構造・築年代に対して 109 の特徴量を持つデータが得られた．表-1 に東京都における建物データ種別ごとの構造・築年代別の建物棟数を示す．建物の種別により，構造別棟数と築年代別棟数が大きく異なることがわかる．すなわち，本データは不均一データであり，そのままモデリングすると，少数派のクラスのカテゴリ精度が低いモデルになってしまう．これを防ぐために本稿では，後述のようにアンダーサンプリングによるスパースモデリングを試みる．

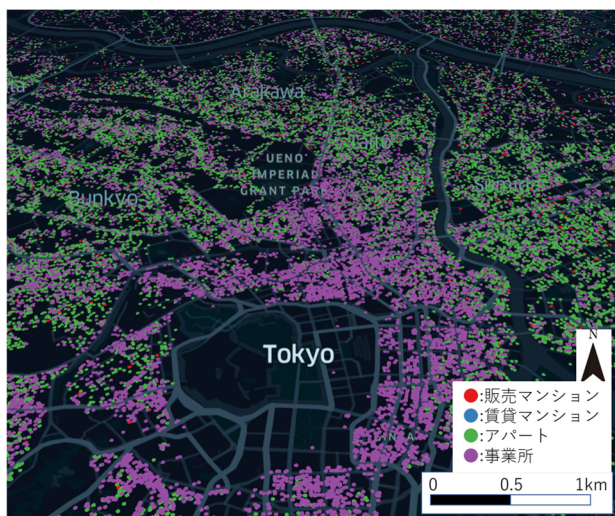


図-1 千代田区周辺の不動産ポイントデータの分布

表-1 データ種別ごとの築年代・構造別棟数

属性		販売マンション	賃貸マンション	アパート	事業所	戸建住宅
築年代	1971年	853	4167	4390	3331	303
	1972-1980年	2506	12930	9253	5740	497
	1981-2000年	7268	64351	49070	14690	1739
	2001年-	4202	29025	20214	2032	4977
構造	RC造	14615	75925	1698	17979	182
	S造	214	34245	24523	6425	370
	木造	0	0	56706	1389	6964

2.2 LASSO 法

2.1 節で整備したデータは特徴量の数が 100 を超える高次元データである．したがって，未知の

データに対する予測精度を保ちつつ，多重共線性やオーバーフィッティングの問題が生じないように，データの次元を圧縮しながらモデリング（圧縮センシング）する必要がある．本稿での圧縮センシングには，深層学習などの特徴量抽出に応用されるスパースモデリング（SpM）を用い，SpM には Least Absolute Shrinkage and Selection Operator (LASSO) 法 (Tibshirani, 1996) を用いる．LASSO 法は，推定したいパラメータを \mathbf{ax} ，対象データの確率を P_i としたとき， $\log \frac{P_i}{1-P_i} = \mathbf{ax}$ の Logit モデルの離散選択問題において，対数尤度 $\sum_{i=0}^n \log(P_i) + (1 - \delta_i) \log(1 - P_i)^2$ に， \mathbf{x} の 1 次ノルム $\|\mathbf{x}\|_1$ を非スパース性のペナルティー項として付加して最小化 (L1 正則化) させることで解析し，スパース解 \mathbf{ax} を得る (式(1))．

$$\arg \max [\sum_{i=0}^n \{ \log(P_i) + (1 - \delta_i) \log(1 - P_i)^2 \} - \lambda \|\mathbf{x}\|_1] \quad (1)$$

ここで， λ はスパース性の制御パラメータであり， $\lambda = 0$ のとき，式(1)は通常対数尤度と一致する．LASSO 法の L1 正則化では，適切な規範で λ を選択し，説明変数を刈り込むことで，データの特徴構造を適切に抽出したスパース解を得ることが可能である．

建物種別ごとにデータの不均一性を考慮してアンダーサンプリングを行い，教師データを作成した上で，Lasso 法によりモデリングしスパース解を得る．モデルのパラメータチューニングでは，交差検証 (5-fold Cross Validation) を行い，平均二乗誤差根 (RMSD: Root-Mean-Square Deviation) を評価指標として用いる．

3. 結果

モデリング結果を図-2,3 に示す．

両図の左上図は，スパース性を制御するパラメータ λ を横軸にとり，非ゼロ特徴量の数と，得られたスパース解による推定データと不動産ポイントデータの RMSD の変化を示したものである． λ の減少とともに，非ゼロの特徴量の数が増加しつつ，それに反比例して RMSD が減少している．

図-2,3 の右上図は，図-1 に示した不動産ポイントデータの建物種別とスパース解による推定種別を比較した結果である．建物構造については概ね正しく推定できている (図-3)．一方，築年代に関しては，4 クラス (1971 年以前/1972-1980 年/1981 年-2000 年/2001 年以降) のうち 1 クラスもしくは 2 クラス分ずれて誤判定している建物が多い．

図-2,3 の下図は，RMSD が最小となったときの，各建物種別のスパース解におけるパラメータを示したものである．築年代のスパース解について

は、109 の特徴量のうち賃貸マンションは 53 個、販売マンションは 25 個、アパートは 21 個、事業所は 27 個、戸建住宅は 20 個しか非ゼロの特徴量を持たず、それ以外は推定精度に寄与しない。同様に、建物構造のスパース解においては、109 の特徴量のうち賃貸マンションは 53 個、販売マンションは 27 個、アパートは 15 個、事業所は 13 個、戸建住宅は 17 個しか非ゼロの特徴量を持たない。このことから、SpM によって、データのノイズを適切に刈り込んだスパース解が得られていると言える。また非ゼロの特徴量は、建物構造においては建物面積と機能 17 (プロパンガス) の有無がすべてのモデルで抽出され、築年代においては建物面積、階数、機能 8 (トランクルーム) の有無、機能 18 (オートロック) の有無が 4 つのモデルで抽出された。

4. おわりに

本稿では、機械学習を用いて建物の構造・築年代を 1 棟単位で推定する手法を提案した。詳細な属性情報を有する不動産ポイントデータと、用途地域や道路などの GIS データを空間結合させることで、教師データを作成し、SpM の一つである Lasso 法により建物 1 棟単位で建物構造・築年代を推定するモデルを構築した。また、交差検証や GIS 上での可視化により、構築したモデルの精度を検証した。

本研究はまだ発展途上であるが、本稿で提案したモデルを用いることで、建物データに必要な特徴量データを付与すれば、全国の建物に対して構造と築年代を推定することが可能となる。今後は、より多くの特徴量を付与するとともに、防火・準防火地域などの GIS データや衛星画像などから得られる特徴量を考慮することで、精度の向上を

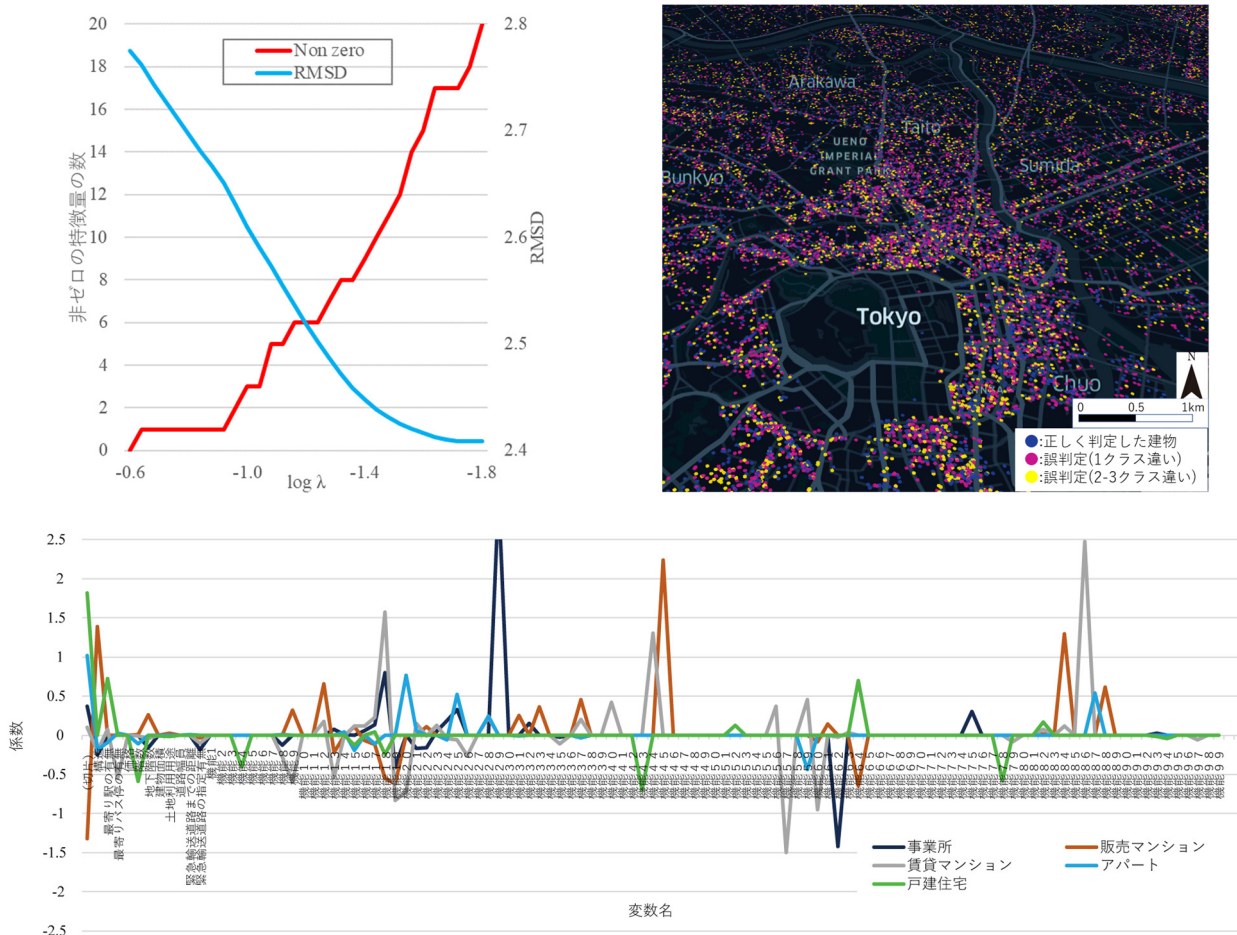


図-2 築年代の SpM によるモデリング結果 (左上図：スパースパラメータ λ に対する非ゼロの特徴量数と RMSD の関係, 右上図：不動産ポイントデータとスパースモデリングによる推定データとの比較, 下図：建物分類別の RMSD が最小となる λ のときの各特徴量のパラメータ)

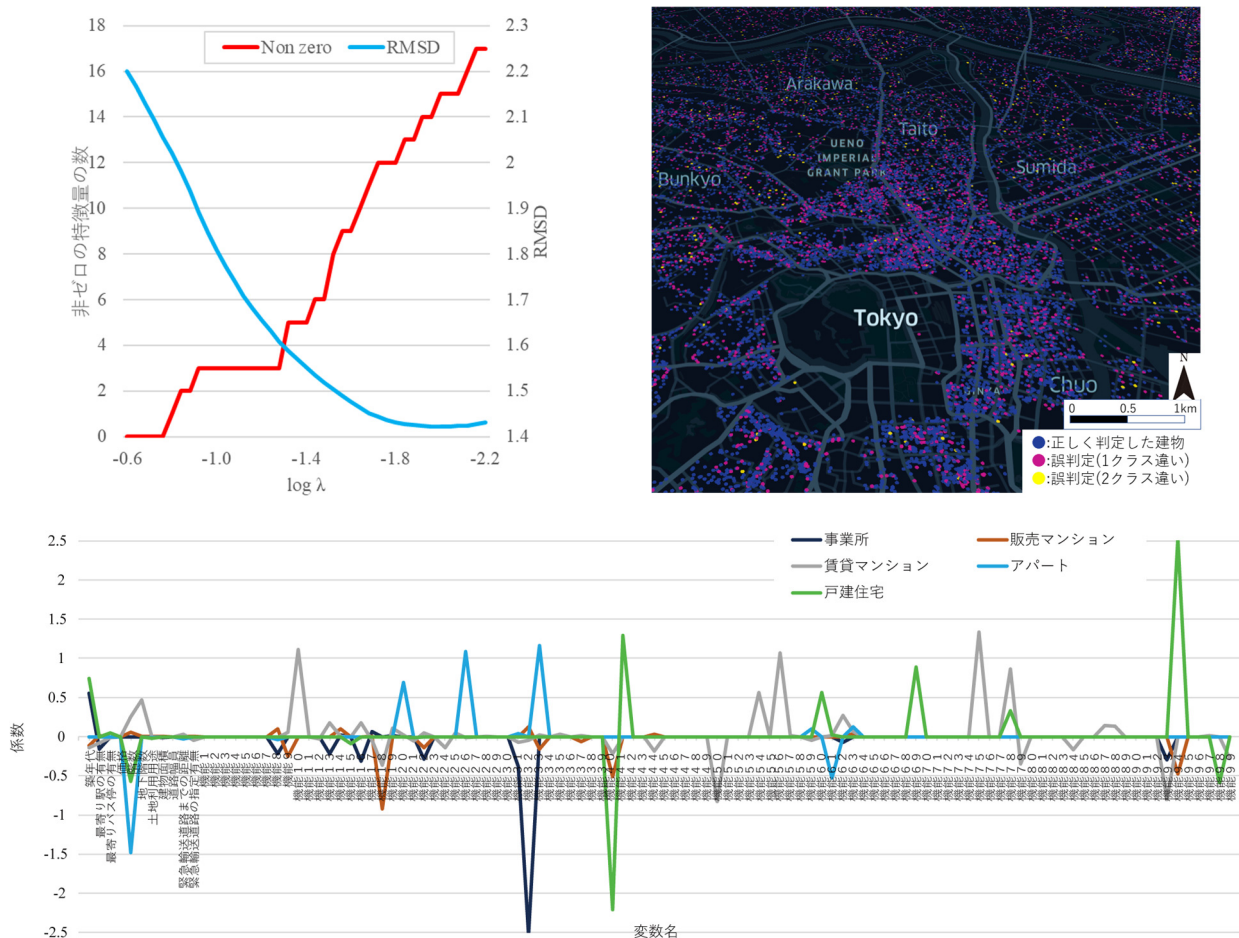


図-3 建物構造の SpM によるモデリング結果（左上図：スパースパラメータ λ に対する非ゼロの特徴量数と RMSD の関係，右上図：不動産ポイントデータとスパースモデリングによる推定データとの比較，下図：建物分類別の RMSD が最小となる λ のときの各特徴量のパラメータ）

図る。また，Finite Element Method (FEM) による建物応答解析など，より詳細なシミュレーションを実現するために，3次元の建物構造の推定手法も提案していきたいと考えている。

謝辞

本研究は，科学研究費の補助を受けた．ここに記して感謝の意を表する．

参考文献

村尾修，山崎文雄（2000）：自治体の被害調査結果に基づく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，65，527，189-196．

Tibshirani, R. (1996): Regression shrinkage and

selection via the lasso, J. Roy. Stat. Soc. B, 58, 267–288.