

ニューラルネットワークを利用した既存一戸建て住宅の建築年代予測*

弘前大学大学院地域共創科学研究科 潘 蔣杰

弘前大学 片岡 俊一

1. はじめに

日本は地震による多くの被害を受けてきた世界有数の地震国である。地震による建築物の被害を評価するには、耐震性能が建築年代によって異なることから、建築年代を知ることが重要である。しかしながら、一般に建築物の建築年代を知ることが容易ではなく、戸田・境 (2011) では外観で建築年代を推定することを試みている。

戸田・境 (2011) の研究では、建築年代が分かっている建築の写真と対象建築の写真を比較することで年代を判別していることを提案している。この作業を AI に置き換えることを考えた。つまり、本研究では図 1 に示す階層型ニューラルネットワークを用い、建築物の外観特徴から建築年代区分を予測するモデルを構築した。

しかしながら、階層型ネットワークは自由度が高く、特に中間層 (又は隠れ層) のユニット数については任意である。また、入力パラメータが結果に与える影響も分かり難い。そこで、本研究ではまず、得られているデータを全て利用したモデルを構築し、それを基本として、入力パラメータを減らすという手順でモデルを作成した。併せて中間層のユニット数の影響も検討した。

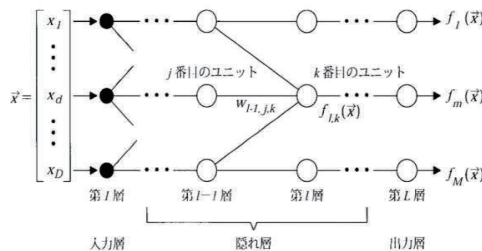


図 1 階層型ニューラルネットワーク (萩原(2011))

2. データの処理

本研究では平成 25 年度 (2015 年度) の弘前大学工学部地球環境学科の卒業研究として長内奈津美が収集したデータを用いた。長内は、不動産サイト「@nifty 不動産」を利用して 2015 年 9 月から 2015 年 12 月の期間に掲載されていた、青森県青森市、福島県福島市、兵庫県明石市の三つの地域の住宅データを収集している。

データセット総数は 270 であり、内訳は青森市が 82 件、福島市が 77 件、明石市が 111 件である。そこでは各住宅について以下の 23 項目を調べ、特徴量を整理している。町名・建築年・建築年数・値段・土地面積・建物面積・構造・階数・外壁・ドアノブ・店舗兼用・玄関フード

*A Neural Networks Model for Predicting the Age of Housings by Pan Jiangjie and Shunichi Kataoka

の有無・屋根の形状・塀・門の有無・庭の草の状況・蔦の有無・基礎が見えるかどうか・外壁の色・外壁の素材・窓のサッシの色・換気口・妻換気口。なお、全ての住宅でこの項目が全て分かっている訳で無い。

データのうち外観の特徴を、特徴が分からないこと（不明）を含めて順序数として数値化した。さらに、ニューラルネットワークモデルに入力する際には、データの数値を0～1の範囲に納める必要があるため、以下の数式により、データの正規化を行った。

$$Y = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

ここで、 X はデータの値で下付添え字の max , min はそれぞれ最大値と最小値を意味する。また、 Y は正規化されたデータである。

3. 初期モデルの構築

プログラミング言語 R の “neural” パッケージを用いて学習モデルを構築した。出力判定は建築物の耐震性能が大きく変わることになる建築基準法改正年に着目した。できれば、1971年より前、1971年から1980年、1981年以降の3区分の分類を行いたいが、最低でも1981年以降とそれより前の区分は行いたい。そこで、3区分の分類が可能かの検討も含め、以下の4パターンのモデルを作成し、モデル間の優劣を検討した。

- ・入力層(全項目)、中間層ユニット数3、出力層2区分(以下、O2N3)
- ・入力層(全項目)、中間層ユニット数4、出力層2区分(以下、O2N4)
- ・入力層(全項目)、中間層ユニット数3、出力層3区分(以下、O3N3)
- ・入力層(全項目)、中間層ユニット数4、出力層3区分(以下、O3N4)

教師信号について、2区分の場合、1980年以前を0、1981年以後を1とした。3区分の場合、1970年以前を0、1971年から1980年までを0.5、1981年以後を1とした。

また、全270個のデータから180個を学習用データとし、残りの90個のデータをモデルの評価用データとした。

4. モデルの評価方法

各モデルの精度は、次式で示す平均絶対誤差(MAE)と二乗平均平方根誤差(RMSE)及び図2に示す混同行列の正解率と敏感度で評価した。

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (2)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (3)$$

ここで、 y_i : i 番目のデータにおける目的変数の値、 \hat{y}_i : i 番目のデータにおける目的変数の値、 n : データ数である。

また、混同行列の正解率と敏感率は次式で表される。

$$\text{正解率} = \frac{TP + TN}{TP + FP + FN + TN} \quad (4)$$

$$\text{敏感度} = \frac{TP}{TP+FN} \quad (5)$$

		予測	
		1	0
実 際	1	TP	FN
	0	FP	TN

図2 混同行列

5. モデルの比較

学習結果を確認するために、教師信号の値を x 軸に、学習後のモデルの予測値を y 軸にプロットし、図3に示す。図3(a), (b)に示した2区分モデルでは、出力値が0.5未満とそれ以上で区分すれば、完全ではないが、判別できることが分かる。3区分モデルでは、1970年以前と1981年以後の区分の判別精度は高いが、1971年から1980年までの区分の判別精度はあまり良くない。また、中間層ユニット数3のモデルに比べて、中間層ユニット数4（それぞれ(b), (d)）のモデルの方が外れ値は少なくなる。

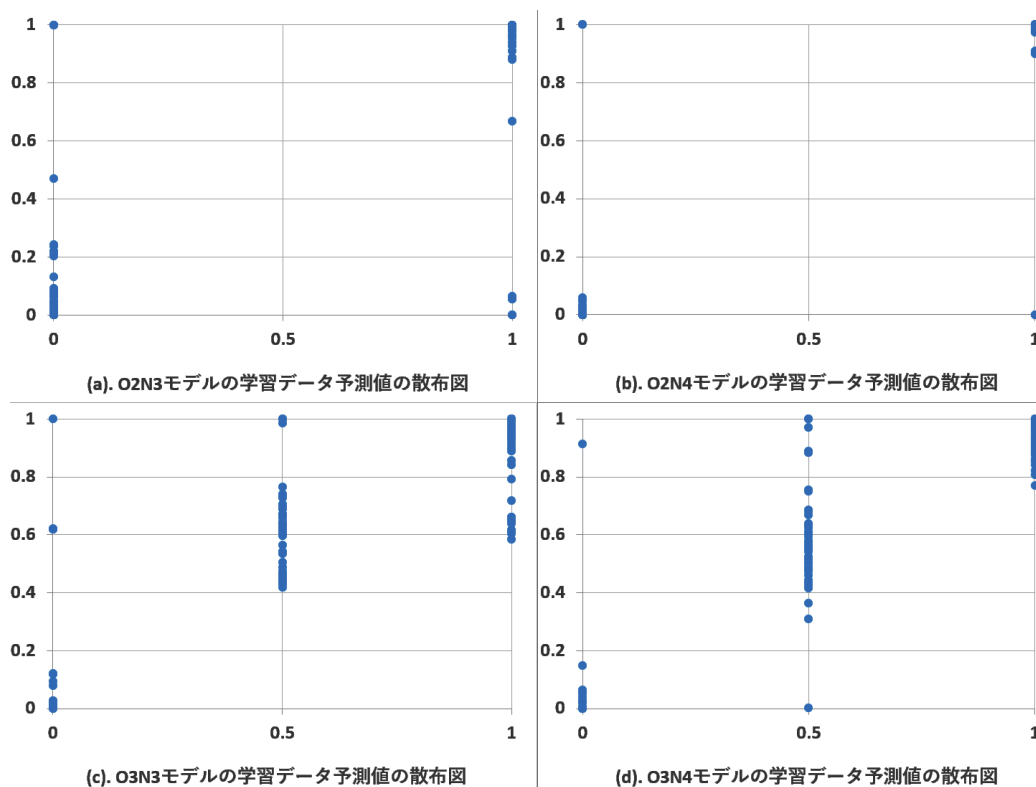


図3 学習結果の状況

学習させた結果を表 1 に示す。各モデルの学習データ（表中の Train）の MAE は全て 0.1 以内に収まっており、学習は順調であった。そこで、検証データに適用した。その結果も表 1 に記されているが（表中の Test）、3 区分モデルの MAE が 2 区分モデルの MAE より小さい、つまり、3 区分モデルの方が高精度で結果を予測できると言える。また、3 区分モデルの RMSE を見ると、中間層ユニット数を 3 としたモデル（O3N3）の方が 3 としたモデル（O2N4）より小さくなりがる。さらに、図示はしないが O3N3 は O3N4 より外れ値が出にくいことが分かった。なお、2 区分モデルと 3 区分モデルのデータセット（出力データ）が異なるため、RMSE を用いたモデルの間の比較は適切でない。

表 1 学習データ及び検証データの MAE と RMSE

	Train MAE	Train RMSE	Test MAE	Test RMSE
O2N3	0.07805	0.24037	0.28089	0.51524
O2N4	0.02869	0.15000	0.26769	0.50144
O3N3	0.09503	0.17966	0.21214	0.33085
O3N4	0.07424	0.14046	0.22389	0.36619

検証データにおける正解率と感度を計算した結果を表 2 に示す。表から、2 区分モデルの正解率が 70%以上になり、3 区分モデルの正解率が 64%前後にとどまっていることが分かる。一方、3 区分モデルの感度が 2 区分モデルのそれより高くなっており、特に O3N4 の感度が 90%を越えている。

表 2 検証データにおけるモデルの正解率と感度

	正解率	感度
O2N3	0.71111	0.77419
O2N4	0.74444	0.74194
O3N3	0.64444	0.79032
O3N4	0.63333	0.90323

以上の結果を踏まえ 3 区分モデルの適用が可能と判断した。次章では中間層ユニット数に関する検討は行うものの、3 区分モデルを改良することとする（以下では初期モデルと呼ぶ）。

6. 説明変数の選定及びモデルの改良

初期モデルの説明変数は 16 個あるが、前述のようにこれらが全て分かっている訳ではない。また、年代判別には関係ないものと考えられるものも含まれている。さらに、説明変数が少ない方が調査は容易になる。以上のことから、説明変数を減らすことを念頭に、各説明変数を表 2 のように特徴ごとにグループ分けし、説明変数の影響を考えた。

以下の検討では、原則として利用するグループ以外の 3 つのグループから、一つずつを抜いたもの（計 3 モデル）、二つを除いたもの（計 3 モデル）、3 つを全て除いたものの（1 モデル）の 7 パターンを考え、さらに中間層ユニット数 3 と 4 とした合計 14 パターンでモデルを構築し、

モデルの性能を調べた。

表 2 説明変数のグループ分け

グループ名	説明変数							
原則として使用	階数	外壁	玄関フード	屋根	塀	庭の草	門	基礎
不明データが多い	ドアノブ	換気口						
特定の特徴に集中	構造	蔦	店舗兼用					
年代判別に影響を及ぼさないと考えられる	壁の色	壁素材	窓のサッシの色					

7. 改良モデルの比較と最終モデルの決定

構築した 14 パターンのモデルの検証データにおける MAE と RMSE を表 3 にまとめる。MAE で比較すると、中間層ユニット数によらず「不確定なしモデル」、「不明・集中・不確定なしモデル」の MAE が検討対象のなかでは小さく、更に初期モデルの値より小さくなった。このうちユニット数を 4 とした「不明・集中・不確定なしモデル」の MAE が最小となった。

さらに表 4 に上述の 2 つのモデルにおける正解率と敏感度を整理した。表から、正解率は中間層ユニットが 3 の場合は初期モデルと同等以下であるが、ユニット数を 4 とすると初期モデルよりも良い結果となっていることが分かる。敏感度は一例を除いて低下している。

表 3 改良モデルの MAE と RMSE

	モデル	MAE	RMSE
中間層ユニット数 3	初期モデル	0.21214	0.33085
	集中なし	0.21784	0.37614
	不明なし	0.22377	0.34602
	不確定なし	0.20314	0.32698
	不明・不確定なし	0.25900	0.40887
	不明・集中なし	0.23619	0.34692
	集中・不確定なし	0.22502	0.36473
	不明・集中・不確定なし	0.21137	0.31775
中間層ユニット数 4	初期モデル	0.22389	0.36619
	集中なし	0.23676	0.38134
	不明なし	0.25213	0.40735
	不確定なし	0.22082	0.37129
	不明・不確定なし	0.22125	0.35607
	不明・集中なし	0.21804	0.34598
	集中・不確定なし	0.25899	0.38903
	不明・集中・不確定なし	0.19572	0.30577

以上から中間層ユニット数を 4 とした「不明・集中・不確定なしモデル」が最も適切と判断した。このモデルは、初期モデルに比べて入力項目が半減しているため、学習時の計算時間が大幅に短縮できた。著者の PC (CPU : I7-4790K, 4.4GHz, メモリ容量 : 16GB, グラフィック

ボード: GTX980) で 10 万回の学習を行った結果, 初期モデルの所要時間は 190 分であったが, 改良モデルの所要時間は 132 分となり所要時間が約 30%短縮された。

表 4 改良モデルの正解率と敏感度

	モデル	正解率	敏感度
中間層ユニット数 3	初期モデル	0.64444	0.79032
	不確定なし	0.64444	0.74194
	不明・集中・不確定なし	0.63333	0.82258
中間層ユニット数 4	初期モデル	0.63333	0.90323
	不確定なし	0.65556	0.82258
	不明・集中・不確定なし	0.65556	0.83871

8. まとめ

本研究では, ニューラルネットワークを用いて, 住宅の外観特徴から建築物が建築基準法の改正以前か以後に建てられたのかを予測するモデルを構築した。さらに, 初期モデルに基づき, 精度を損なうことなく, 入力データ数を半分にした改良モデルの構築に成功した。入力データ数を半分にできたことによって, 計算時間の大幅短縮が期待され, 建築物の建築年数調査の迅速化に役立つと考えている。

しかしながら, 本研究に使用したデータは, 青森県青森市, 福島県福島市, 兵庫県明石市の限定された 3 つの地域から収集したものである。そのため, 本研究で構築したモデルが全国的に利用可能であるかについては, 疑問が残る。また, 今回使用したデータ数は 270 件であり, ニューラルネットワークモデルを構築するには少ないため, モデルの予測精度はそこまで高く感じない。より多くの住宅データを収集し, 汎用性の高いかつ予測精度の高いモデルを構築できるようにすることが今後の課題である。

謝辞

本論文で使用したデータは文中にあるように, 平成 25 年度 (2015 年度) の弘前大学理工学部地球環境学科の卒業研究として長内奈津美さんが収集したものである。データの利用に当たっては, 長内さんの指導教員であった元弘前大学の津村浩三先生にお世話になった。記して謝意を示す。

参考文献

- 1) 戸田和徳・境有紀 (2011), 強震観測点周辺における木造建築物群の実被害データの補正 (その 1) 建物写真を用いた建築年代判定の検討, 日本建築学会大会学術講演梗概集 (関東), 構造 II, 21451, 901-902
- 2) 萩原克幸 (2006), ニューラルネットワークの基礎と理論的に重要な課題, J. Plasma Fusion Res. Vol82, No.5, pp. 282-286.