

レイنزのニューラルネットワークを用いた不動産価格査定について

An assessment of real estate price using neural network on REINS

福井 光^{*1} 阪井 一仁^{*1} 南村 忠敬^{*1} 三尾 順一^{*1} 木下 明弘^{*1} 高田 司郎^{*2}
 Hikaru Fukui Kazuyoshi Sakai Tadataka Minamimura Jyunichi Mio Akihiro Kinoshita Shiro Takata

^{*1} 公益社団法人 近畿圏不動産流通機構
 KINKI Real Estate Information Network System

^{*2} 近畿大学
 KINDAI University

The purpose of REINS (Real Estate Information Network System prescribed by Ministry of Land, Infrastructure and Transport) is to distribute real estate and to contribute to improve the house environment of the people so that develop the market utilizing the analysis based on the data of real estate and the agreement real estate. The goal of this paper is to propose the machine learning method for an assessment of real estate price using neural network with agreement real estate on REINS. First, this paper describes the classification and normalization of input data after current state of real estate price assessment. Data normalization for neural network is required to be numeric data from 0 to 1 because most input items are text data. After analyzing the experiment data, we try to realize how the data affect real estate price. Then, we perform error back propagation of neural network with agreement real estate as input data, and classify per million, half a million, 100 thousand. Finally, the experiment enhances feasibility.

1. はじめに

レイنزとは、国土交通省が指定した不動産物件情報を業者同士で登録提供するためのコンピュータ・ネットワーク・システムの略称である。集積される不動産の物件情報や成約物件の価格情報の調査研究を通じて、不動産取引の適正化および流通の円滑化を図り、不動産流通の健全な発達と国民の住生活の安定向上に寄与することを目的としている。

そこで、本稿では、レイنزに大量に集積された成約データを利用して、ニューラルネットワークを用いた不動産価格を査定する学習方式を提案する。

まず、不動産価格査定の現状を述べ、それらを踏まえた上で入力データの分類と正規化について述べる。成約データにはテキストデータが多く、ニューラルネットワークの入力データとするには、数値化する必要がある。また、前処理として入力項目を全て0-1に正規化し、学習パラメータを分析することで、どの項目が価格査定に影響しているか分析することを狙いとする。

次に、レイنزの成約データを入力データとして、ニューラルネットワークの誤差逆伝播法を用いて、100万単位、50万単位、および、10万単位のそれぞれにクラス分類する学習モデルを生成する。

最後に、それら学習モデルを用いて、同じ成約データを入力とした3種類のクラス分類テストを行った結果、成約データの契約価格が分類されたクラスの価格帯にほとんど含まれており、今後、レイنز上での実用化が期待できることを示す。

2. 不動産価格

不動産価格についてはある程度の相場は形成されており、実際に不動産業者が価格を査定する際には、類似物件との比較が一般的である。一方、不動産には全く同じ物件は存在しないという個性を有していることから、明確な根拠を消費者に示し理解してもらうためには多大な労力を要するのが実状である。

●実務的な視点

実務で参考となる指標としては、国土交通省が示している不

動産鑑定評価基準や、公益財団法人不動産流通推進センターが示している価格査定マニュアルがある。価格査定マニュアルの活用は3節の入力データの正規化と分類で後述する。

不動産鑑定評価基準では、不動産価格は、一般に、

- (1) その不動産に対してわれわれが認める効用
- (2) その不動産の相対的稀少性
- (3) その不動産に対する有効需要

の三者の相関結合によって生ずる不動産の経済価値を、貨幣額をもって表示したものである。そして、この不動産の経済価値は、基本的にはこれら三者を動かす自然的、社会的、経済的及び行政的な要因の相互作用によって決定される。不動産価格とこれらの要因との関係は、不動産価格が、これら要因の影響下にあると同時に選択指標としてこれら要因に影響を与えるという二面性を持つものである[国土交通省 14]。

すなわち、不動産価格は、不動産自体の特性、相対的稀少性、買い手側や売り手側の需要供給のバランス、それら諸要因の相互作用によって形成されるが、不動産の持つ要因それ自体も常に変動する傾向を持っている。従って、不動産の鑑定評価に当たっては、価格形成要因を市場参加者の観点から明確に把握し、かつ、その推移及び動向並びに諸要因間の相互関係を十分に分析し、前記三者に及ぼすその影響も考慮する必要がある。このことから、不動産の持つ個性を踏まえた上で査定価格およびその根拠を明確に示すためには、不動産鑑定士のように高度で専門的な知見を持った者でも労力を要する。

●学術的な視点

これまでも不動産価格の分析は統計的なアプローチは元より都市経済学、金融工学等、様々な分野で行われている。ここでは、不動産の価格分析のアプローチとして用いられることが多い、ヘドニック・モデルから不動産価格の特質について言及する。ヘドニック・アプローチとは、ある商品をその商品の様々な属性の価値に関する集合体とみなし、回帰分析を利用して、各々の属性価格を推定する手法であり、不動産の個性を踏まえた上で統計的な分析を行うのに適し、不動産市場分析でよく用いられている[Karato 16]。価格を形成する要因で重要なものが何か、例えば立地や面積、築年数等が重要な要素であることを示せるが、線形的なアプローチになるため、不動産の持つ個性や歪みを必ずしも考慮できるとは限らない[Ishijima 10]。

連絡先: 福井 光, 近畿圏不動産流通機構, 〒540-0036
 大阪市中央区船越町 2-2-1, email: h.fukui@kinkireins.com

AI 分野では、従来からあったニューラルネットワーク技術が進化し、近年、深層学習に関する研究が活発化、多数のパラメータを必要とする非線形の問題に対しても学習が可能になっている[Asoh 13]。不動産価格の自動査定への応用も期待され、不動産にテクノロジーを融合させた「不動産テック」と言われる研究や実用化に向けた動きが活発化している。

3. 入力データの分類と正規化

実験には、大阪府で過去 5 年間(2017 年 3 月以前)に成約登録されたマンションのデータ 40,056 件を利用する。前処理として 500 にも及ぶ全項目を数値 0-1 に正規化。機械学習を効率よく進める上で入力データの正規化は重要な意味を持つ。

3.1 入力データの分類

入力データの正規化を行うに際し、前提としてレインズ項目の分類と具体例を表1に示す。大まかな分類としては非数値と数値に分けられ、また不動産価格と正の相関がある項目と負の相関がある項目に分ける。

表1 レインズ項目の分類と具体例

	分類	具体例
1	非数値(立地)	所在地、沿線、駅
2	非数値(上記以外)	方角、用途地域、建物構造
3	数値(正の相関)	専有面積、所在階
4	数値(負の相関)	築年数、最寄駅からの距離
5	その他	備考欄、未入力項目等

3.2 入力データの正規化

1) 所在地の数値化

「所在地」は不動産がどこに所在するかを表す項目で、不動産価格に強い影響を及ぼす項目である。しかしながら、項目としての所在地は数値化されていないため、何らかの指標を活用して数値に置き換える必要がある。駅ランキングや住みやすさランキング等、市場でも数値化された指標は出回っているが主観を排し切れていない数値であることから、今回は近畿レインズで保有する成約土地データ(過去 5 年分)を活用。所在地ごとの平均坪単価を算出することで、地価そのものがその地域の強さを表すという合理的な考えに基づいている。実際の数値化については、平均坪単価を最大値で除算する。表 2 を参照。

2) 方角の数値化

方角が同じマンション、同じ階であっても、日照の問題から向きにより不動産価格に影響を及ぼすものとされている。一般的に南向きがもっとも日当たりの関係で良いとされている。今回は、不動産流通推進センターの「マンション価格査定マニュアル」で定義される区分(以下)を参考に、正規化した。表 3 を参照。

表2 住所(所在地 1)の正規化(上位 5 位まで)

住所 (所在地 1)	平均坪単価 (単位: 万円)	正規化後の値 (実数/最大値)
大阪府大阪市 中央区	171.3	1.00
大阪府大阪市 西区	169.4	0.99
大阪府大阪市 天王寺区	142.3	0.83
大阪府大阪市 北区	109.7	0.64
大阪府大阪市 浪速区	104.8	0.61

3) 専有面積の価値計算

「専有面積」については広い方がより価値があり、「所在階」についてもより高い方が一般的に価値があるとされるため、単純に最大値で各項目を除算。結果については割愛する。

表3 方角(バルコニー向き)の正規化

方角	正規化後の値	正規化の考え方
南	1.0	(公益財団法人)不動産流通推進センターの「マンション価格査定マニュアル」で定義される区分(以下)を参考に、正規化 ・南 ・南東、南西 ・東、西 ・北東、北西 ・北
南東	0.8	
南西	0.6	
東		
西		
北東	0.4	
北西	0.2	
北		

4) 築年数の正規化

「築年数」については建築後、一般的に新しい方がより価値があるとされる。欧米と異なり新築信仰が強い日本ならではの特色と言える。従って、建築後の経過年数で比較した際、築年数が大きければ大きいほど、不動産価格の価値は減少していくと考えられる。しかし、一律に負の相関があるとは必ずしも言えない。理由として、耐震基準等も関係するが、一定の年数を経過した物件については「新しさ」という点では、有意な価値は存在しなくなることが挙げられる。一定の築年数が経過すれば、数年の差は誤差になってしまうという特性が日本の不動産市場には存在する。このような特性を踏まえ、正規化を実施。表4を参照。

表4 築年数の正規化

築年数	正規化後の値	正規化の考え方
1	1.00	築年数については負の相関があることから、考え方としては逆数を取る。 ただし、築 10 年までは価値が緩やかに逓減し、0.05 ずつ減じる。
2	0.95	
3	0.90	
4	0.85	
5	0.80	
6	0.75	
7	0.70	
8	0.65	
9	0.60	築 11 年以降は逓減の幅が一般に大きくなることから、築 16 年までは、0.08 ずつ減じる。 築 17 年以降築 20 年までは、逓減の幅は緩やかになり、0.01 ずつ減じる。
10	0.55	
11	0.47	
12	0.39	
13	0.31	
14	0.23	
15	0.15	
16	0.07	
17	0.06	築 20 年以降は大きな差異はないものとして、全て 0.01 とする。
18	0.05	
19	0.04	
20	0.01	
21	0.01	

5) 備考欄、未入力項目など

「備考欄」については、文字数制限はあるが物件の PR ポイント等を自由に入力することができる。中には不動産価格に影響

響を与える要素がありうるという推察は容易にされるが、任意入力項目であることもあり、今回の実証実験では一律 0 とする。また、未入力項目については、レインズシステムは必須項目が少なく、入力のない項目については一律 0 とする。

4. 実験

ニューラルネットワークを用いて、4 万～3 億円の範囲の不動産価格を 100 万単位、50 万単位、および、10 万単位に、それぞれ、0(0 以上 100 万未満)～301(3 億以上 3 億 100 万未満) 0(0 以上 50 万未満)～601(3 億以上 3 億 50 万未満)、0(0 以上 10 万未満)～3001(3 億以上 3 億 10 万未満)にクラス分類する実験を行う。

4.1 実験方法

レインズの過去 5 年分 40,056 件の成約データを 3 節の方法で数値化した 495 項目を入力データとして、不動産価格査定の実験に使用する。本実験では、大量の実際に成約したデータを用いることができるため、2 節で述べたような不動産価格の特徴関数を求めて不動産価格値を回帰する方法は取らず、ニューラルネットワークの誤差逆伝播法を用いてクラス分類を行う。この方法を用いることには、以下の狙いがある。

- (1) レインズシステムに組み込み後も、オンライン学習が容易であり、時代の趨勢にマッチした学習が進められる。
- (2) クラス分類を並行して行うことで、より精度の高い価格査定を行う。例えば、100 万、50 万、10 単位のクラス分類時の各ソフトマックス値(クラス分類時の確率値)を考慮して、より正確な価格査定を行う。
- (3) 人間は各自の経験に基づいて、価格を査定している。ニューラルネットワークの学習済みパラメータは、価格分類の内部状態表現である。学習パラメータを分析することで、どの項目を重視して査定を行なっているかを、例えば、ネットワークシミュレータを作成することで分析する。

学習方式は以下の通り。ニューラルネットワークには Batch Normalization、更新手法は Adam、活性化関数は ReLU を使用するミニバッチ学習方式。入力データの 80 パーセントを訓練データ、残り 20 パーセントをテストデータとする。

実験環境は、CPU Intel Core i7-7800X、メモリ 64GB。OS は Ubuntu 17.10、言語は Python3.6.3 を使用する。

4.2 学習モデルの生成

図 1、図 2、図 3 はそれぞれ、100 万単位、50 万単位、10 万単位の各学習モデルの認識精度グラフである。表 5 は、ハイパーパラメータと epoch 数(訓練データ件数分の更新数を 1epoch とする)、elapsed time、訓練データおよびテストデータの max 認識精度である。

表 5 ニューラルネットのモデルと認識精度

	100 万単位	50 万単位	10 万単位
入力層	485	485	485
出力層	301	601	3001
中間層	(50,100,50)	(50,100,50)	(50,100,50)
学習率	0.000001	0.000001	0.000001
バッチサイズ	4800	4800	12000
epoch 数	5,000	5,000	6,000
elapsed time (sec)	49470	49748	64366
max train accuracy	0.9426	0.9096	0.6994
max test accuracy	0.8693	0.7179	0.5505

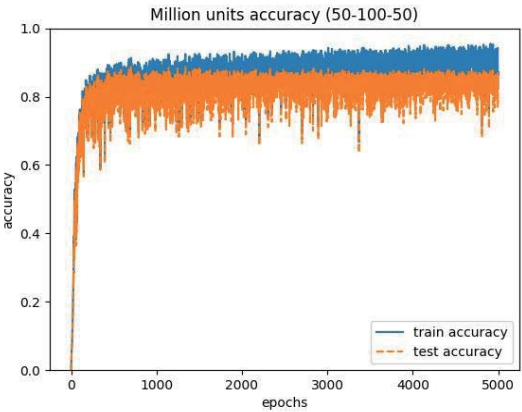


図1 100 万単位モデルの認識精度グラフ

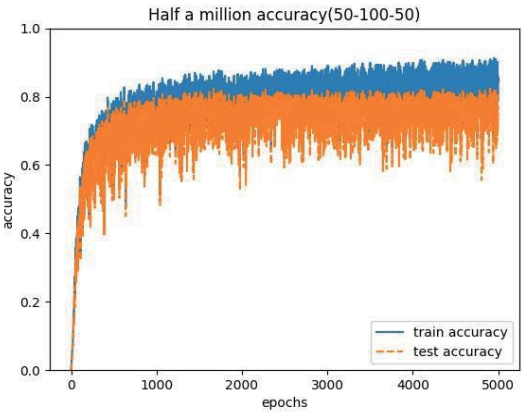


図 2 50 万単位モデルの認識精度グラフ

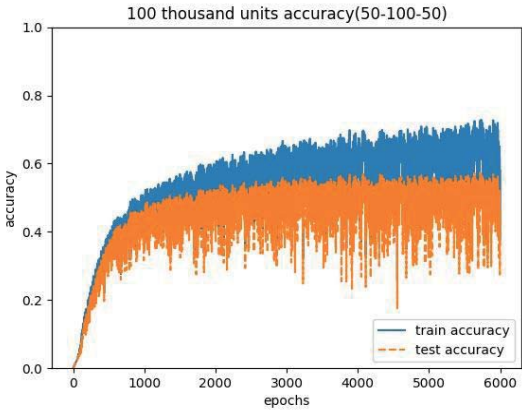


図 3 10 万単位モデルの認識精度グラフ

4.3 推定クラスのテスト結果

表 6 は、テストデータから 15 件を取り出し、作成した 100 万単位、50 万単位、10 万単位モデルを使用して、クラス分類したクラス名とその softmax 値である。カッコ内はクラス分類が間違った場合の正解クラスである。

なお、入力データ 40,057 件の内、40,008 件は 1 億円までの成約データであり、残り 49 件が 1 億より大きく 3 億以下の制約データである。このようなデータのバラツキを考慮し、テスト対象に関しては、1 億円までは 1 千万円単位とし、1 億以上は 5000 万円単位の成約データを選択する。

査定額は、100 万単位、50 万単位、10 万単位にて 2 個以上で同じ金額の場合はその査定金額を記入する。

表 6 各モデルの推定価格と査定額

成約価格	100 万単位	50 万単位	10 万単位	査定額
500	4(5)	10	50	500
	0.9010	0.7874	0.4660	
1500	15	30	150	1500
	0.9683	0.8709	0.8375	
2500	25	50	250	2500
	0.9432	0.8174	0.8457	
3500	34(35)	71(70)	350	---
	0.7020	0.5001	0.6640	
4500	45	90	450	4500
	0.9969	0.9963	0.9418	
5500	55	109(110)	548(550)	---
	0.5670	0.3427	0.4850	
6500	65	130	650	6500
	0.6999	0.9786	0.4100	
7500	75	150	750	7500
	0.4759	0.5579	0.9714	
8500	83(85)	170	828(850)	---
	0.5102	0.9860	0.3805	
9500	95	190	950	9500
	0.9988	0.7826	0.5626	
12500	125	250	1250	12500
	0.9957	0.9983	0.7010	
17500	175	350	1750	17500
	0.9971	0.9999	0.9389	
20000	158(200)	600(400)	1720(2000)	---
	0.9669	0.9606	0.9999	
25000	145(250)	270(500)	1350(2500)	---
	0.9999	0.8624	0.8931	
30000	300	600	3000	30000
	1.0000	1.0000	1.0000	

5. 考察

不動産査定は、現場では、100 万単位での査定結果ができれば実用可能であり、表 6 の結果から 100 万単位の誤差の範囲に収まっている。

また、表 6 より、40,008 件の 1 億円までの成約データに対しては、100 万単位、50 万単位、10 万単位の各学習モデルを用いてクラス分類された価格帯に、実際の契約価格がほぼ含まれている。そこで、レインズに大量に集積された成約データを利用して、ニューラルネットワークを用いた不動産価格を査定する本学習方式は、今後のレインズシステム上の実用化に寄与すると考えられる。

ただし、このテスト結果は、テストデータから 15 件を選択した結果であり、全てのテストデータにて、テストを行う必要がある。また、表 5 より 10 万単位の学習モデルのテストデータの認識精度は 0.5505 と 100 万単位と 50 万単位のテストデータの認識精度より低いにも関わらず、表 6 でのクラス分類のヒット率はほぼ同等である。これは、学習モデルのニューラルネットワーク構造は最後の出力層のみの数が異なるだけであり、出力層への分類数が多ければ、逆に精度よくクラスに分類できるのと考えられるのか、これに関しても大量のテストデータで再度、検証する必要がある。

6. 今後の課題

6.1 査定額の方法

5 節では、3 種のモデルの分類クラスから計算される査定金額が 2 種類で同じなら、査定額として採用した。その他として、100 万単位、50 万単位、10 万単位で、金額の包含関係で矛盾がない場合も査定できるとも考えられる。例えば、100 万単位、50 万単位、10 万単位にて、それぞれ、クラス 6、クラス 13、クラス 66 と分類された場合には、金額的には、それぞれ 600 万以上 700 万未満、650 万以上 700 万未満、660 万以上 670 万未満の査定額となり、これらの積をとると、660 万以上 670 万未満と査定することができると考えられる。大量のテストデータで、このような包含関係をはじめとする計算方法を今後、見つける必要がある。

6.2 実証実験に向けて

実証実験に向けては、ハイパーパラメータ、特に、中間層の素子数に関しては十分な組み合わせを考慮して学習を行う必要がある。特に、10 万単位の認識精度の向上が重要である。

また、本研究では大阪で過去 5 年に成約されたマンションを入力データとして用いているが、今後は近畿二府四県へと地域の拡大は元より、マンション以外にも土地、戸建て等、対象物件を広げていく。機械学習を通じて不動産価格査定精度を向上させ、不動産価格査定を行うシステムを構築し、レインズに組み込むことにより不動産価格査定サービスの提供を目指す。

7. おわりに

本稿では、レインズの成約データに対してニューラルネットワークの誤差逆伝搬法を用いてクラス分類を行うことで、レインズ上での不動産査定価格の実用化が期待できることを提案した。

また、不動産の価格査定の実用化に向けては、定量化されていないデータ項目を定量化し、保持していない情報を組み込んでいく等、様々な課題が残されている。合理的且つ迅速に不動産価格を個々の鑑定でなく、システムで算出することができれば、不動産業者の利便性向上だけでなく、消費者と不動産業者間に存在する、情報の非対称性が解消され、不動産市場の活性化につながることが期待される。

8. 謝辞

本実験に当たり、近畿大学理工学部情報学科マルチエージェント研究室の多田涼馬君、長野和人君、谷口翔君に感謝する。

参考文献

- [Hideki Asoh 2013] Hideki Asoh: Deep Representation Learnig by Multi-Layer Neural Networks , Journal of Japanese Society for Artificial Intelligence ,2013.
- [Hiroshi Ishijima 2010] Hiroshi Ishijima,Tomohiro Taniyama: House price fluctuations due to its individuality and market imperfections: implications for personal finance, Journal of Japanese Society for Financial Plannning ,2010.
- [Hiroshi Karato 2016] Hiroshi Karato: ヘドニック・アプローチを利用した不動産価格指数の推定方法とその問題点, Urban housing sciences ,2016.
- [国土交通省 2014] 国土交通省: 不動産鑑定評価基準, 2014.
- [不動産流通推進センター 2015] 不動産流通推進センター: 不動産価格査定マニュアル, 2015.