

所在地情報と Graph Convolution による 賃貸物件価格推定の一手法

加藤 暢之[†] 新妻 弘崇^{††} 太田 学^{†††}

^{†, †††} 岡山大学大学院自然科学研究科 〒700-8530 岡山県岡山市北区津島中 3-1-1

^{††} 大阪大学データビリティフロンティア機構 〒565-0871 大阪府吹田市山田丘 2-8

E-mail: [†]kato@de.cs.okayama-u.ac.jp, ^{††}niitsuma@ids.osaka-u.ac.jp, ^{†††}tohta@cs.okayama-u.ac.jp

あらまし 多くの賃貸物件のデータは所在地の情報を含んでおり、機械学習や深層学習では所在地の緯度経度情報を使用する際、他の特徴量と同様に扱う場合がほとんどである。一方で近年、グラフ構造を持つデータに対する畳み込みニューラルネットワークの研究が進んだことにより、様々な構造をもつデータをグラフとして学習することが可能となった。本稿では物件の位置関係から作成した物件グラフを利用する賃貸価格推定手法を提案する。また評価実験として、賃貸物件データセットである LIFULL HOME'S データセットにおける岡山県のデータを使って賃貸価格推定を行い、類似した構造のニューラルネットワークによる賃貸価格の推定結果と比較する。

キーワード 深層学習, 地理情報, 機械学習, グラフデータベース

1 はじめに

近年、様々な分野でグラフ構造をもつデータに対する深層学習に関する研究が進んでいる。なかでも Graph Neural Network の拡張が活発に行われており、attention を考慮したグラフの学習 [1] や、巨大なグラフにおける効率的な特徴の学習手法 [2] などが研究されている。

一方、賃貸物件データや不動産データにおいてもニューラルネットワークを用いた研究が進んでいる。不動産の間取り図から画像認識により部屋の構造を表すグラフを構築し応用する手法や、分類した画像から賃貸物件価格を推定する研究 [3] など多様な角度から賃貸物件データを有効活用しようという動きも高まっている。

賃貸物件の価格を決定する際、所在地や立地条件が重要な情報であることは明らかである。そこで本研究では、より詳細に所在地情報を反映させたニューラルネットワークによる学習のために、賃貸物件データの所在地情報から物件グラフを作成し、Graph Convolution により学習する手法を提案する。

本稿では 2 節で関連研究について述べ、3 節で賃貸物件の価格推定について、4 節で提案手法について述べる。また、5 節で提案手法の評価実験とその考察を述べる。最後に 6 節でまとめる。

2 関連研究

グラフ構造をもつデータを学習する手法に Schlichtkrull ら [4] が提案した Relational Graph Convolutional Network (R-GCN) がある。R-GCN は、Kipf ら [5] が提案した Graph Convolutional Network を拡張した、ノードの隣接関係と注目するノード自身を畳み込むことのできる深層学習手法である。

R-GCN が学習の対象とするのはグラフ $G = (V, E, R)$ であ

り V はノード集合、 E はラベル付きエッジ集合、 R はエッジの分類で構成される。 G はループ、多重エッジ、有向エッジをもつグラフを含む。R-GCN 層においてノード v_i についての出力は式 (1) のようになる。

$$\mathbf{h}_i^{(l+1)} = \varphi \left(\sum_{r \in R} \sum_{j \in N_i^r} \frac{1}{c_{i,r}} \mathbf{W}_r^{(l)} \mathbf{h}_j^{(l)} + \mathbf{W}_0^{(l)} \mathbf{h}_i^{(l)} \right) \quad (1)$$

ここで R は隣接関係のラベル集合、 N_i^r はノード v_i と隣接関係 r で隣接するノード集合、 $c_{i,r}$ は正規化定数である。Schlichtkrull ら [4] は正規化定数 $c_{i,r}$ には $|N_i^r|$ などを用いるのが良いとしている。 $\mathbf{W}_0^{(l)}$ は自己ループの重み、すなわちノード v_i 自身の重みであるのに対し、 $\mathbf{W}_r^{(l)}$ はノード v_i と関係 r で隣接しているノードの重みである。これらを集約し活性化関数 φ で変換し出力する。

3 賃貸物件の価格推定

3.1 データセット

本研究では、国立情報学研究所が提供している賃貸物件情報データセットである LIFULL HOME'S データセット [6] を利用する。LIFULL HOME'S データセットには 2015 年 9 月時点での賃貸物件スナップショットデータ、高精細度間取り画像データ、賃貸・売買物件月次データなどが含まれており、本研究では、岡山県内の賃貸物件スナップショットデータ 78,019 件を利用して、賃貸価格を推定する。また、本実験では物件特徴のうち質的特徴を除外し、線形に扱うことのできる量的特徴のみを使用する。使用する量的特徴とその基本統計量を表 1 にまとめる。岡山県の賃貸価格の平均値 53751.97 円に対し標準偏差 17637.50 円である。

表 1 賃貸物件価格推定に使用する岡山県内物件の量的特徴とその基本統計量

特徴名	平均値	中央値	最大値	最小値	標準偏差
賃貸価格 (円)	53751.97	52000.00	648000.00	3000.00	17637.50
バス停までの距離 (最寄り 1 箇所目)(m)	14.78	15.00	90.00	0.00	9.94
バス停までの距離 (最寄り 2 箇所目)(m)	10.09	10.00	60.00	0.00	10.60
建物面積 (m^2)	43.87	44.00	7947.00	0.00	37.59
築年数 (月)	188.23	190.08	1236.32	-1.92	122.97
駐車場までの距離 (m)	2.48	0.00	3150.00	0.00	28.25
小学校までの距離 (m)	784.65	726.00	5500.00	0.00	435.85
中学校までの距離 (m)	1138.33	1015.00	7500.00	0.00	690.61
コンビニまでの距離 (m)	442.81	360.00	9399.00	0.00	369.13
スーパーまでの距離 (m)	656.30	540.00	7785.00	0.00	526.29
総合病院までの距離 (m)	729.87	568.00	9794.00	0.00	605.87

3.2 問題定義

賃貸物件の価値はその賃貸価格に表れるが、賃貸価格は一般的に物件の貸主が決定する。賃貸物件の条件から定量的に賃貸の妥当な価格を推定することができれば、貸主が賃貸価格を決定したり、借主が希望する賃貸物件の価格を見積もる際の助けとなる。そこで本研究では、賃貸物件の特徴からニューラルネットワークにより賃貸価格の妥当な価格を推定することを目的とする。

物件の賃貸価格の推定にはニューラルネットワークを利用する。表 1 に示した物件特徴を表す量的特徴を入力とし、図 1 に示すニューラルネットワークを学習する。図 1 では、活性化関数 ReLU をもつ全結合層を 3 層重ねた中間層をもつニューラルネットワークにより、物件特徴を表す 10 次元の量的特徴から本来の賃貸価格を推定する。

物件特徴

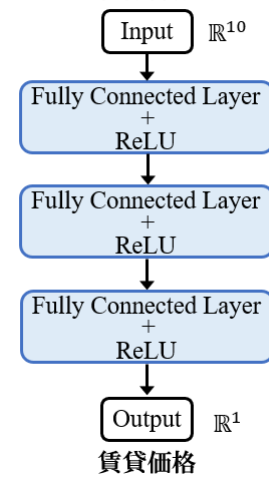


図 1 賃貸物件価格推定モデル

4 提案手法

本研究では、物件の位置関係を表したグラフを学習するために、図 2 に示す構造の Relational Graph Convolutional Network(R-GCN) を利用する。本節では、2 節で述べた R-GCN を拡張し、物件の所在地情報を利用した Graph Convolution による学習の手法について述べる。4.1 節では、図 2 の構造をもつ R-GCN が学習で利用するグラフの作成について述べ、4.2 節では式 (1) に基づく R-GCN の拡張について述べる。

4.1 物件グラフの作成

本研究では物件の所在地情報から物件グラフを作成する。物件グラフでは物件をノード、物件の位置関係をエッジとする。2 節で述べた R-GCN では、ノードの隣接情報から隣接ノードのもつ特性を集約し畳み込みを実現する。本研究で利用する LIFULL HOME'S データセットでは、岡山県内の物件データの所在地は地理的に 31 区分に分類されている。そのため本研究では各地点の位置関係を表す図 3 の地点グラフを作成し、地点グラフから各物件の位置関係を表す図 4 の物件グラフを作成する。

31 地点の所在地から Google Places API の一部である Geocoding API により緯度経度を取得し、地図上にプロッ

物件特徴

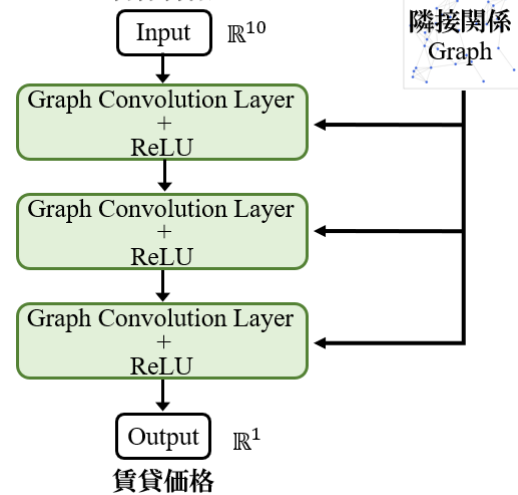


図 2 Graph Convolutional Network モデル

トすると図 5 のようになる。

取得した緯度経度から地点間の距離を算出する。また、地点が近接していると定義する地点間距離の閾値を設定し、閾値以下の距離内にある地点が近接しているものとする。閾値を設定

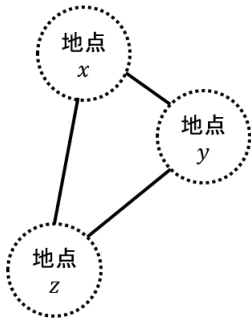


図 3 地点の位置関係

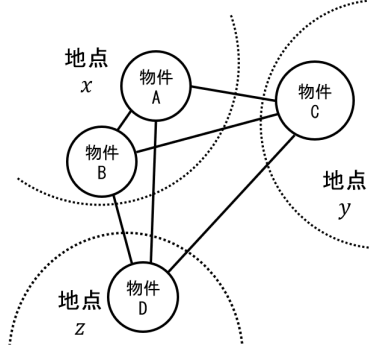


図 4 物件グラフ

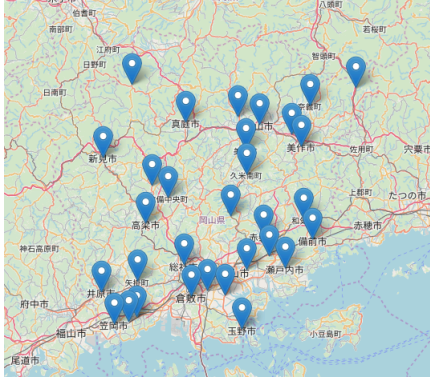


図 5 得られた物件座標

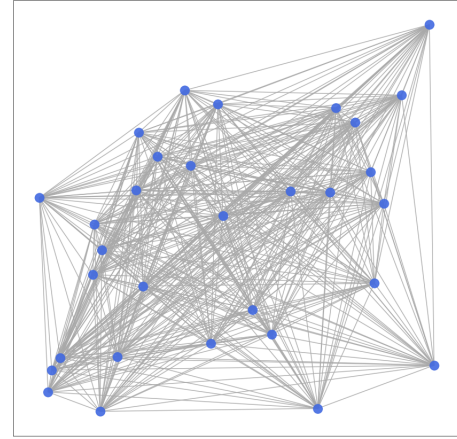


図 6 全結合グラフ

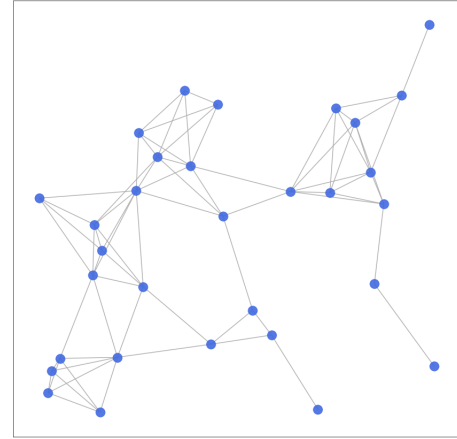


図 7 部分結合グラフ

しないグラフを全結合グラフとし，図 6 に示す．

全結合グラフでは R-GCN において式 (1) で示したノード特性が全ノードから集約される．そのため全結合グラフを使用した場合，遠距離に存在する物件からも特性を学習できる．一方で閾値を $20km$ に設定してエッジを減らした部分結合グラフを図 7 に示す．部分結合グラフを学習する場合，閾値以内の距離にある地点に存在する物件の特性のみを集約するため，より局所的なノード特性が学習される．地点間の近接関係を表す全結合グラフと部分結合グラフのそれぞれから図 4 の物件グラフを作成する．全結合グラフから作成した物件グラフを全結合物件グラフ，部分結合グラフから作成した物件グラフを部分結合物件グラフと呼称する．

4.2 物件間距離による R-GCN の拡張

本節では物件間の距離を用いた R-GCN の拡張について述べる．R-GCN を提案した Schlichtkrull ら [4] は正規化定数として $|N_i^r|$ を使用すると良いとしている． $|N_i^r|$ は隣接ノード数が隣接ラベルごとに異なる場合も，各隣接関係を同等に扱うための正規化定数である．本節の提案手法では $|N_i^r|$ に加え，物件間の距離による重み付けを行った．これにより拡張した R-GCN の中間層の出力は 2 節で示した出力の式 (1) と式 (2) で表せる．

$$\frac{1}{c_{i,r}} = \frac{1}{|N_i^r|} \frac{\max(\mathbf{D}_i) - \mathbf{D}_{i,j}}{\max(\mathbf{D}_i)} \quad (2)$$

$\mathbf{D}_{i,j} \in \mathbb{R}^{|V| \times |V|}$ は物件ノード i, j 間の距離であり， \mathbf{D}_i は物件ノード i と隣接している物件との距離の集合である．式 (1)

と式 (2) を図 2 のように学習する．

物件グラフの一部を図 8 に示す．図 8 では，地点 x に存在する物件 A と物件 B ，地点 y に存在する物件 C ，地点 z に存在する物件 D がある．物件 A と物件 B は地点 x に存在しており，物件 A, B ともに地点 x の近接関係を用いる．また，地点 x と地点 z は設定した閾値以上の距離にあり，近接していないものとする．物件 C に着目すると，地点 y のもつ隣接関係ラベルは r_{xy}, r_{yz} であり，物件 C へ集約されるノード特性は h_A, h_B, h_D ，自己ループの h_C である． h_A と h_B は同一の隣接関係に属するため， $\frac{1}{|N_i^r|}$ により正規化される．また物件 D に着目すると，地点 z のもつ隣接関係ラベルは r_{yz} であり，物件 D に集約されるノード特性は h_C ，自己ループによる h_D となる．

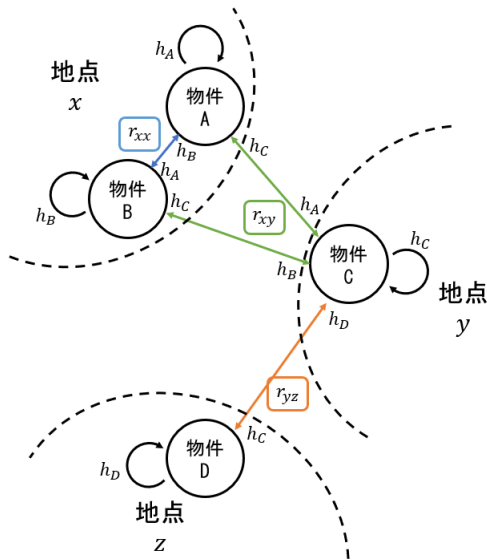


図 8 ノード特性の集約例

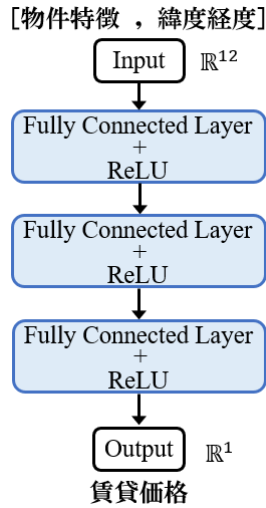


図 9 ベースラインモデル

5 評価実験

5.1 評価手法

本節では、4 節で説明した所在地情報を利用する R-GCN を入力特徴に加えて賃貸物件価格を推定する手法を、ベースラインとする図 9 の手法と実験により比較する。ベースライン手法では量的特徴である 10 次元の物件特徴に所在地の緯度経度を 2 次元の特徴として加え、12 次元の特徴を入力とする。

図 6 の全結合グラフと図 7 の部分結合グラフによって定義された 31 地点間の近接関係を利用し 1000 個のノードを持つ物件グラフを構築する。評価実験では訓練データ 800 件、検証データ 100 件とテストデータ 100 件の計 1000 件の物件データから物件グラフを構築する。本節では図 9 に示したベースラインによる推定、全結合グラフから構築した全結合物件グラフ、部分結合グラフから構築した部分結合物件グラフのそれぞれについて R-GCN と 4.2 節で述べた拡張をした R-GCN による学習

表 2 価格推定結果 (MAE, 円)

ベースライン		6430.0
R-GCN	全結合物件グラフ	-
	部分結合物件グラフ	6034.9
拡張 R-GCN	全結合物件グラフ	10350.2
	部分結合物件グラフ	5681.1

した価格推定の計 5 手法で物件価格を推定し比較する。各手法の比較は平均絶対誤差 (Mean Absolute Error, MAE) により行う。

ベースラインであるニューラルネットワークモデルと提案手法を含む R-GCN モデルによる賃貸価格の推定結果を表 2 に示す。全結合グラフを用いた R-GCN による全結合による学習では検証誤差の値が収束せず、正確な推定結果を得ることができなかった。拡張 R-GCN による全結合物件グラフの学習は検証誤差は収束したが、ベースラインより推定誤差が大きい。一方、R-GCN による部分結合グラフの学習と拡張 R-GCN による部分結合グラフの学習では、ベースラインより推定誤差が小さい。

また、学習過程の検証誤差の推移を図に示す。R-GCN による全結合物件グラフの学習を図 10 に R-GCN による部分結合物件グラフの学習を図 12 に示す。また、拡張 R-GCN による全結合物件グラフの学習を図 11 に、拡張 R-GCN による部分結合物件グラフの学習を図 13 に示す。

図 10 では検証誤差が収束しなかった。図 11、図 13 はどちらも、学習初期の 50 epoch で検証誤差が 8,000 円程度まで小さくなり、その後の 500 epoch で検証誤差 5,000 円程度まで小さくなった。図 12 と図 13 の相違点は、学習が収束した時点の検証誤差である。図 12 は 100epoch 付近で検証誤差が 10,000 円に収束した。一方で図 13 は 100epoch 以降で検証誤差が小さくなった。

5.2 考察

実験結果から、物件間の位置情報に基づいてエッジを削減した部分結合グラフと、R-GCN の拡張手法は、本実験において学習を阻害せず不要な特徴を削減できたといえる。また R-GCN による全結合グラフの学習は検証誤差が収束しなかったが、拡張 R-GCN による全結合 R-GCN の学習では検証誤差が収束しているため提案した R-GCN の拡張はグラフの学習に有効であることがわかる。

6 まとめ

本稿では物件の所在地情報を活用した Relational Graph Convolutional Network の拡張手法を提案した。提案手法による賃貸価格の推定精度はベースラインによる推定精度を上回った。また、物件グラフにおけるエッジの削減と距離による重み付けはいずれも学習において有効であることを確認した。

今後の課題は、モデルの訓練方法や特徴量の構築の工夫による推定精度向上である。

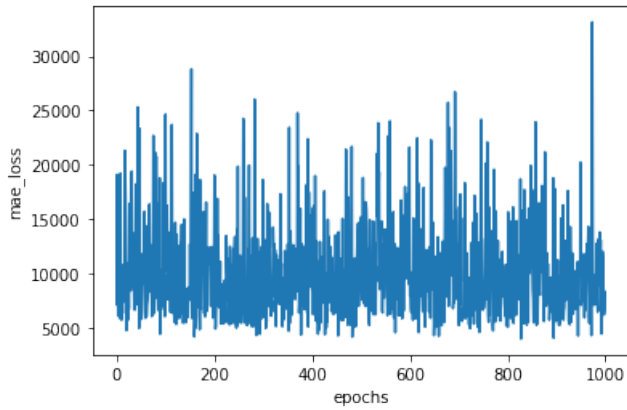


図 10 全結合 R-GCN 検証誤差 (円)

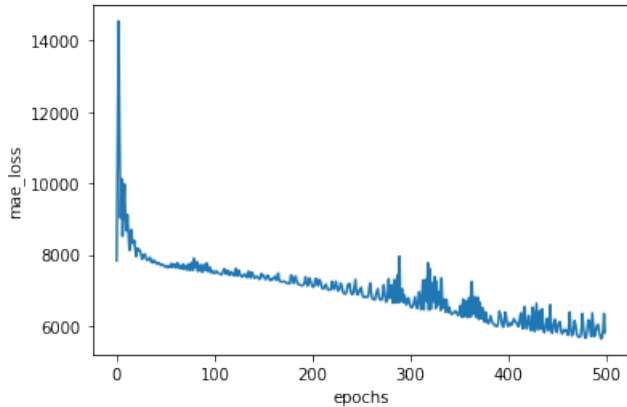


図 11 部分結合 R-GCN 検証誤差 (円)

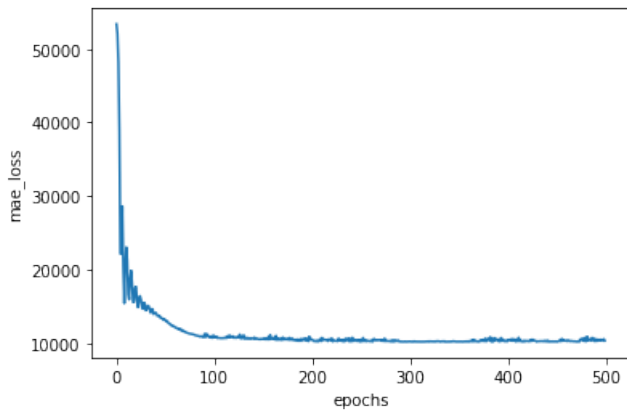


図 12 全結合拡張 R-GCN 検証誤差 (円)

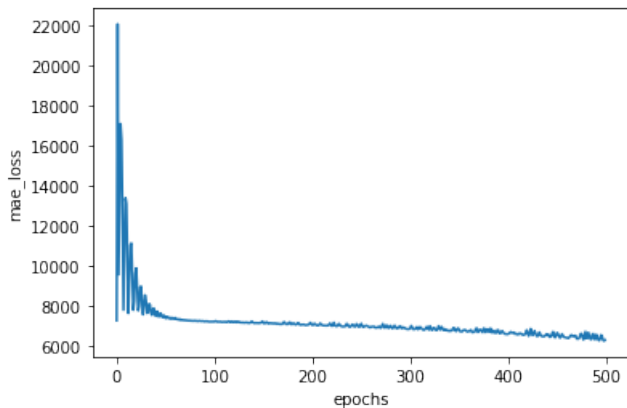


図 13 部分結合拡張 R-GCN 検証誤差 (円)

謝 辞

本研究では、国立情報学研究所の IDR データセット提供サービスにより株式会社 LIFULL が提供する「LIFULL HOME'S データセット」を利用した。ここに記して感謝する。

文 献

- [1] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, “Graph attention networks,” *arXiv preprint arXiv:1710.10903*, 2017.
- [2] W. Hamilton, Z. Ying, and J. Leskovec, “Inductive representation learning on large graphs,” in *Advances in neural information processing systems*, pp. 1024–1034, 2017.
- [3] O. Poursaeed, T. Matera, and S. Belongie, “Vision-based real estate price estimation,” *Machine Vision and Applications*, vol. 29, no. 4, pp. 667–676, 2018.
- [4] M. Schlichtkrull, T. N. Kipf, P. Bloem, R. Van Den Berg, I. Titov, and M. Welling, “Modeling relational data with graph convolutional networks,” in *European Semantic Web Conference*, pp. 593–607, Springer, 2018.
- [5] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” *arXiv preprint arXiv:1609.02907*, 2016.
- [6] 株式会社 LIFULL(2015), “LIFULL HOME'S 賃貸物件スナップショットデータ.” 国立情報学研究所情報学研究データリポジトリ (データセット). <https://doi.org/10.32130/idr.6.1>.