

特徴グラフを用いた汎用型 CNN 深層学習手法の開発

Development of general purpose CNN deep layer learning method using feature graph

高橋 慧¹, 沼尻 匠², 曾我部 完², 坂本 克好¹, 山口 浩一¹, 横川 慎二³, 曾我部 東馬^{*1,2,3}

¹ 電気通信大学 基盤理工学専攻

² 株式会社 グリッド

³ 電気通信大学 i-パワーエネルギー・システム研究センター

we propose a method of applying Convolutional Neural Networks (CNN) to non - image data. CNN has been successful in many fields such as image processing and speech recognition. On the other hand, it was difficult to adapt CNN to non-image data such as a csv file. The sequence of the data of the low dimensional grid structure such as the image has a meaning, and the CNN recognizes the order as the feature of the image and processes it. Therefore, CNN could not perform feature recognition on non-image data whose structure can be changed. We focused on a method to make CNN applicable by giving meaning to the sequence of non - image data, and demonstrated by adding improvements.

1. はじめに

様々な種類のデータの収集と保存は、あらゆる分野において基礎的かつ重要な活動である。収集され蓄積されるデータの量は日々、驚異的な割合で増加している。このデータを有用な知識にするために、大規模かつ複雑なデータセットを理解する方法を開発しなければならない。

その解析方法の新たな糸口として畳み込みニューラルネットワーク(Convolutional Neural Networks: CNNs)[1]が考えられている。従来のニューラルネットワークは過学習や勾配消失の問題から層を増やすことが容易ではないが、CNNは畳み込み層とプーリング層を交互に繰り返すことで層を増やし、ネットワークを深層化することが可能である。そのため近年、CNNはディープラーニングの主力とされ、画像処理、音声認識、コンピュータビジョン、言語処理など多くの分野で成功を収めてきた。一方で、CNNを変化するグラフ構造を持つデータに適応することは困難とされてきた。なぜならCNNは画像などの低次元グリッド構造の幾何学的特性を利用する畳み込みをしているため、非画像であるグラフ構造やグリッド構造のデータに対しては畳み込みをしても、正しい特徴認識が行えないためである。

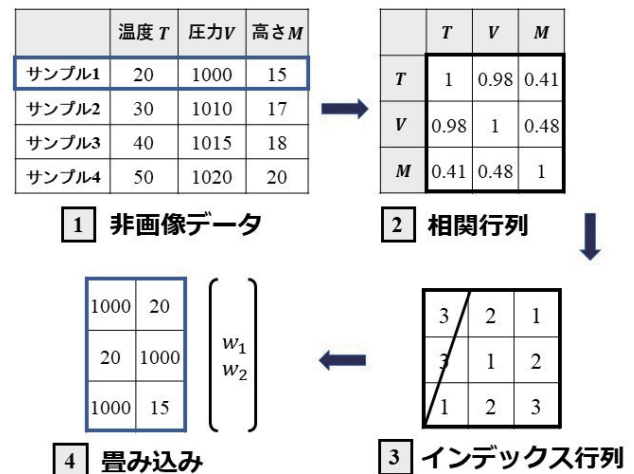
そうした理由から最近、非画像データに対してCNNを一般化する方法が研究されている。本稿では、データの相関を利用して畳み込みを可能とするCorrelation GraphCNNという手法を元に、データ間の距離を利用した非画像データに対するCNNの応用方法を提案する。

2. Correlation GraphCNN (C-GraphCNN)

CNNで最も重要な役割を果たしている畳み込み層は、フィルタと呼ばれる小さな行列を画像全体に渡って適用することで、画素ごとの数値の『順番』を特徴として学習している。CNNが得意とする画像処理の分野では、データの列や行の順番をランダムに入れ換えると、画像が本来の形から崩れて、データの持つ意味は変わってしまう。つまり、画像データの要素の順番には意味があるということであり、その意味がある要素の順番をCNNは特徴と捉えることで、学習を行っている。

図1 GraphCNNの畳み込み手順例

一方、csvファイルのような非画像データは、列順を入れ換えたとしてもデータの持つ意味は変わらないため、データの順番



に意味はないということになる。そのため前述の理由から CNN は非画像データを学習することが出来ない。どうすれば、非画像データを CNN に適応できるかを考えたとき、非画像データの順番にも意味を持たせればよいということが考えられる。

そして最近、相関を用いて非画像データの順番に意味を持たせる Correlation GraphCNN[2] という手法が報告された。C-GraphCNN の畳み込みの手順例を図 1 に示す。C-GraphCNN では、最初に非画像データの特徴同士の相関係数を計算し、相関行列を作成する。相関行列は性質上、並べ替えることが出来ないため、データの順番に意味があるということになる。その相関行列から相関係数を強い順に並べたインデックス行列を作成する。ここで、データに対してあまり相関のない特徴変数を除外するため、インデックス行列を左側から任意の行だけ削除する。そしてサンプル 1 を学習する場合、インデックス行列の各行の要素の順番にサンプル 1 から特徴変数の値を取ることで、順番に意味がある行列を作成される。こうした行列を作成し、畳み込みを可能とするのが C-GraphCNN の手法である。

3. Distance GraphCNN (D-GraphCNN)

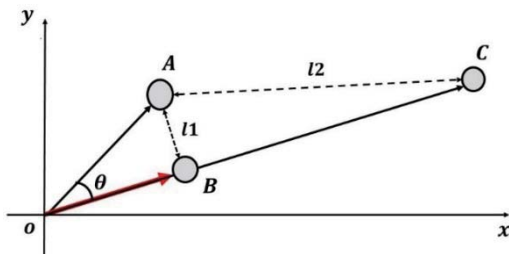
Correlation GraphCNN では、非画像データの順番に意味を与える手段として相関係数を用いていたが、相関係数にはある

欠点がある。ここで図 2 のような 3 つのデータ点がある場合を考える。相関係数 r は 2 つのベクトルのなす角度 θ の余弦で表されるため、ベクトル OA とベクトル OB の相関係数は $\cos\theta$ となる。同様に、ベクトル OA とベクトル OC の相関係数もまた $\cos\theta$ で表される。つまり、相関係数を基準に考えたとき、点 A, B, C は同じクラスタに分類される。

一方、データ同士の距離を基準に考えたとき、点 C は点 A, B から離れているため、一般的に点 A, B と同じクラスタと見なすことは出来ない。相関係数には、こうしたデータ間の実際の距離を考慮せずに関係性を決めてしまう特性があるため、データの関係性を表すのに最適とは言い切れない。そのため、本稿では非画像データの順番に意味を与える手段として相関係数ではなく、距離をコンセプトとした Distance GraphCNN という手法を 2 つに分けて提案する。

図 2 相関と距離によるクラスタリングの違い

(i) Kernel GraphCNN



Kernel GraphCNN では、全てのデータ間の距離を計算し、距離行列を作成することで、データの順番に意味を与える。計算にはガウシアンカーネルを用いる。ガウシアンカーネルは、

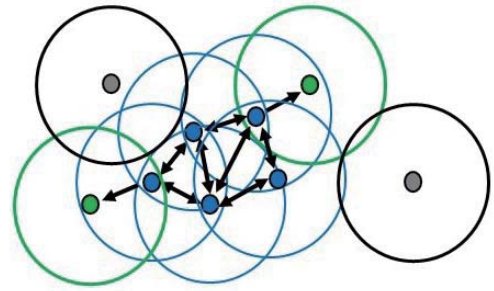
$$k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{\sigma^2}\right)$$

で定義される二変数関数であり、 $k(x_i, x_j)$ は x_i, x_j の距離を表す。 σ はカーネル関数の拡がり制御するパラメータであり、 σ が小さいほど細かいクラスタリングになり、大きいほど単純なクラスタリングになる。

(ii) DBSCAN GraphCNN

DBSCAN GraphCNN では DBSCAN[3] という一般的なクラスタリングアルゴリズムを用いて、近い点同士のみの距離を計算し距離行列を作成する。特徴空間において、多くの点が近接しているような領域を高密度領域といい、DBSCAN はその領域に属する点を見つける。DBSCAN には minPts と eps という二つのパラメータがある。高密度領域内にあるデータをコアサンプルといい、距離 eps 以内に minPts 個以上のデータポイントがあるポイントのコアサンプルとして定義する。また、コアポイントから距離 eps 以内にあるデータを到達可能点、距離 eps 以内に minPts 個以上のデータポイントがない場合、そのデータを外れ点として扱う。図 3 に DBSCAN のクラスタリング例を示す。図 3 において $\text{minPts}=3$ であり、青点はコアポイントである。また緑点は距離 eps 内にコアポイントを含むため到達可能点となる。青点と緑点は単一のクラスタとして扱われ、黒点はコアポイントでも到達可能点でもないため外れ点となる。データに対してこのようなクラスタリングを行い、同クラスタのデータとの距離行列を作成する。

図 3 DBSCAN によるクラスタリング例



4. 実験および評価

Distance GraphCNN が非画像データに対して有効であるか、また C-GraphCNN との精度の違いを比較することを目的に 3 つのデータセットを用いて実験を行った。解析には入力層の後に、GraphCNN による畳み込み層と Dropout 層を 2 回繰り返され、Flatten 層、Dropout 層、全結合層、出力層と続く 9 層からなるニューラルネットワークを用いた。また比較のため 3 つの GraphCNN の他に、4 層のニューラルネットワークを加えた計 4 種の構成を用いた。

4.1 非画像データに対する回帰問題

(i) The Boston Housing Dataset

非画像データの例として、物件の人口統計に関する 12 つの情報から住宅価格を予想する回帰問題を扱う“The Boston Housing Dataset”[4]を用いた。データは 506 サンプルであり、分析の評価には決定係数を用いた。決定係数は

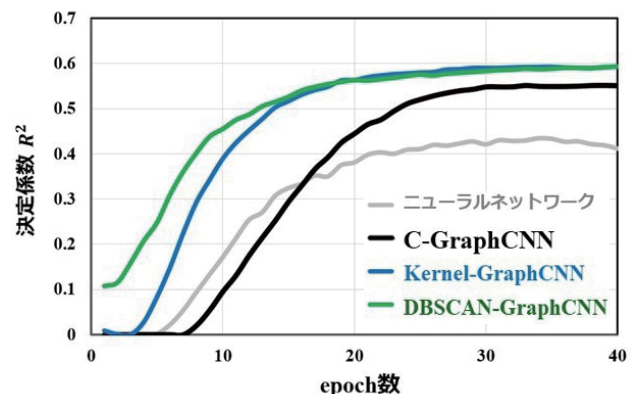
$$R^2 = \text{Corr}(Y, \hat{Y})^2$$

で定義される関数であり、 Y は実際の価格、 \hat{Y} は予測された価格、 R^2 は 1 に近いほど予測の精度が高いことを表す。3 つの GraphCNN の構成において、畳み込み 1 層目は 15 枚、2 層目は 20 枚のフィルタ、全結合層のユニット数は 150 個、Epoch 数は 40 回というパラメータ設定を共通に用いた。

実験結果を図 4 に示す。GraphCNN を用いた 3 つの構成の精度は、従来のニューラルネットワークの精度を大きく上回ることが分かる。また、C-GraphCNN はエポック数が 30 回付近で学習が収束しているのに対して、ガウシアンカーネル・高次元距離行列ではエポック数が 20 回付近で学習が収束している。これにより GraphCNN は相関行列を用いるよりも、距離行列を用いた方が予測の精度が良く、学習の速度も速いということが考えられる。

図 4 回帰問題における決定係数の推移

(ii) Merck Molecular Activity



次に、2153 個のも特徴と 6148 個のサンプルデータを持つ “Merck Molecular Activity Challenge” [5] に挑戦した。このデータセットはデータマイニングコンペティションサイト Kaggle で扱われていた題材であり、分子内の原子間の構造に基づいて、異なる分子の活性レベルを予測するという回帰問題である。

実験結果および様々な構成の R^2 結果を表 1 に示す。DNN と Random Forest は当時の Kaggle コンテストの勝者の 2 人のものであり、Spectral Networks [6] では最先端技術であるスペクトルアプローチを使用している。比較すると GraphCNN は Kaggle コンテストの優勝者よりも優れた性能を発揮していることが分かる。また DBSCAN GraphCNN は最先端技術である Spectral Networks よりもわずかながら予測精度が良いことが分かる。

以上、2 つの実験結果より、GraphCNN は非画像データに対して有効であり、我々の提案手法である D-GraphCNN は C-GraphCNN の精度を上回ることも実証された。

表 1 Merck の分子活性における各手法の予測精度

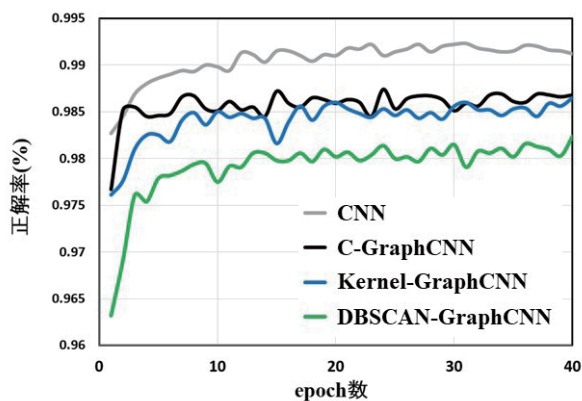
Method	R^2
Random Forest	0.232
Merck winner DNN	0.224
Spectral Networks	0.277
C-GraphCNN	0.264
Kernel GraphCNN	0.235
DBSCAN GraphCNN	0.280

4.2 画像の分類問題とフィルタの可視化

通常、CNN は学習後のフィルタの重みや畳み込み層から出力される値を可視化することで、データの何を特徴として学習しているかを知ることが出来る。しかし、非画像データは画像データのように、出力値を可視化しても何を表しているかを読み取ることは出来ない。そこで GraphCNN は非画像データだけではなく、画像データに対しても適用可能なため、画像データに GraphCNN を応用し、フィルタの可視化を試みた。

まず画像データに対して GraphCNN を用いて学習した場合と、9 層の標準的な CNN で学習した場合の違いを比較した。学習サンプルには MNIST データセット [7] を用いた。テストデータ 60000 サンプルを用いて画像の特徴を学習し、完成した識別機を用いてテストデータ 10000 サンプルに対して 10 クラスの分類問題を行い、正解率の比較およびフィルタの可視化を行った。

図 5 分類問題における正解率の推移



実験結果を図 5 に示す。標準的な CNN の正解率 99.12% には及ばなかったものの、GraphCNN はどの構成においても正解率 98% を超えており、画像データに対しても GraphCNN は有効的であると判断できる。標準的な CNN はデータを 28×28 ピクセルの画像データとして学習しているのに対して、GraphCNN は 784×1 の数値データとして学習している。そのため、CNN が高い性能を発揮する画像データにおいては GraphCNN の精度は CNN よりもやや劣ると考えられる。

また、標準的 CNN と Kernel-GraphCNN の学習後の畳み込み 1 層目のフィルタの重み及び出力値の可視化したものを図 6 に示す。フィルタには 0 から 255 の数値が各ピクセルに割り当てられており、特徴として認識する形を緑色で表している。標準的 CNN が学習したフィルタからは、数字の横線や斜め線という特徴を抽出していることが分かり、それは出力される画像からも読み取ることが出来る。一方、GraphCNN で同じ画像データを学習した場合、フィルタはそれぞれ異なる特徴を抽出していることが分かる。出力された画像を見ると、数字の輪郭を認識する傾向があると考えられる。しかし、非画像データにおいては、出力値を画像として可視化することは出来ないため、フィルタが何を学習しているかはまだ分かっておらず、今後の課題として残っている。

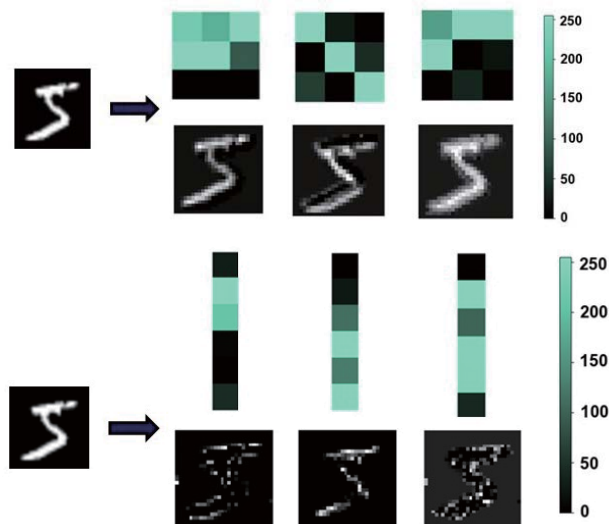


図 6 フィルタと出力値の可視化
(上図: 標準的 CNN、下図: GraphCNN)

5. 強化学習への応用

強化学習とは試行錯誤を通じて、環境に適応する機械学習の一つである。強化学習が扱うデータは非画像データが多く、一般的にステータスの更新にはニューラルネットワークが用いられている。この強化学習のニューラルネットワークにおいても、勾配消失等の問題により層を増やすことが容易ではなく、4 層などの浅いネットワークでステータスの更新を行っていた。

しかし本実験で、非画像データに対しても GraphCNN は有効的であり、なおかつニューラルネットワークよりも優れた結果を出すことが実証された。そのため、強化学習におけるニューラルネットワークを GraphCNN に置き換えることでネットワークをより深層化し、学習精度の向上を試みる実験を行った。実験は Open AI gym のカートポールを利用したシミュレーター環境で行った。エピソードは 1000 回に設定し、4 層のニューラルネットワークと 8 層の C-GraphCNN の二つの構成を用いて、精度の違いを比較した。

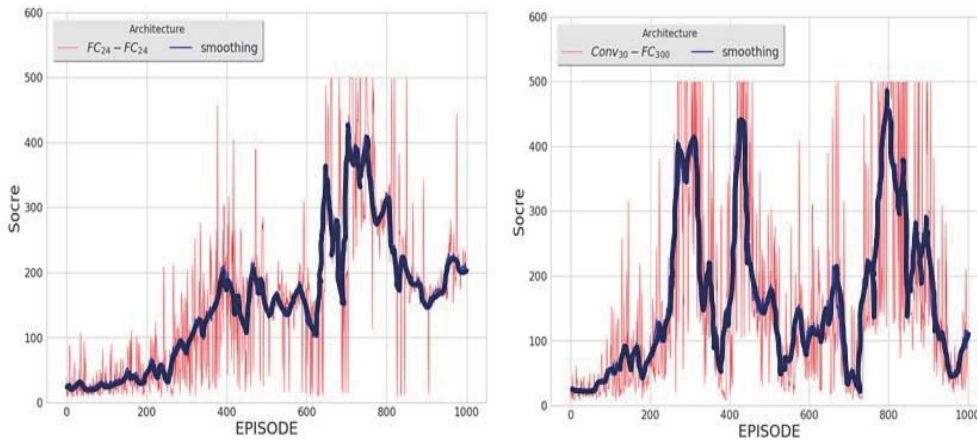


図7 カートポールにおけるスコアの推移
(左図:ニューラルネットワーク、右図:C-GraphCNN)

図7にエピソードごとのスコアの推移を示す。左の図がニューラルネットワーク、右図がC-GraphCNNで学習した結果である。赤線はエピソードごとのスコア、青線は10エピソードごとに平均してスムージングしたスコアを表す。GraphCNNで学習した場合は、ニューラルネットワークに比べて、スコアが最高値に到達するのが早く、また最高値を維持する回数も多いことが分かる。これより、強化学習においてもGraphCNNは有効的であると考えられる。しかし、GraphCNNは行動ごとのスコアの差がやや激しいため、スコアを安定させることが改善点として残っている。

- [4] Harrison, D. and Rubinfeld, D.L. 'Hedonic prices and the demand for clean air', J. Environ. Economics & Management, vol.5, 81-102, 1978.
- [5] <https://www.kaggle.com/c/MerckActivity>
- [6] Mikael Henaff, Joan Bruna, Yann LeCun, "Deep Convolutional Networks on Graph-Structured Data", arXiv preprint, arXiv:1506.05163 (2015)
- [7] Yann LeCun, Corinna Cortes, Christopher J.C. Burges, "THE MNIST DATABASE of handwritten digits", <http://yann.lecun.com/exdb/mnist/>

6. まとめ

本研究では、非画像データに対してのCNNの応用手法としてCorrelation GraphCNNに注目し、その手法の改善案として二つの距離行列を用いたDistance GraphCNNを提案した。提案手法の有効性を検証するため、非画像データの回帰問題に対する実験を行った。その結果、GraphCNNの予測精度は従来のニューラルネットワークよりも優れており、非画像データにもCNNが応用できることが明らかになった。また、提案手法であるDistance GraphCNNは、Correlation GraphCNNの予測精度を上回り、距離行列の有用性が証明された。しかし、GraphCNNが非画像データに対して有効的であるのは証明されたが、データの何を特徴として捉え、学習しているのかは現段階では分かっておらず、今後の課題として残っている。また、GraphCNNは強化学習の一連の学習過程にも応用可能であり、学習精度の向上が期待される。

参考文献

- [1] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L.D. Jackel, "Backpropagation applied to handwritten zip code recognition", Neural Computation, vol.1, pp.541-551, 1989.
- [2] Yotam Hechtlinger, Purvasha Chakravarti, Jining Qin: A Generalization of Convolutional Neural Networks to Graph-Structured Data, arXiv preprint, arXiv:1704.08165 (2017)
- [3] Martin Ester, Hans-Peter Kriegel, Jorg Sander, Xiaowei Xu..A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. Procceedings of 2ndInternational Conference on Knowledge Discovery and Data Mining (KDD-96),1996.