

グラフ表現学習のためのグラフゲート機構付き多層パーセプトロンの提案と有限要素法のサロゲートモデルとしての応用

中井 優, 奥田 洋司

東京大学

1. はじめに

グラフのノードやエッジに対する表現学習や分類・回帰を行うニューラルネットワークは Graph Neural Network (GNN) と呼ばれ, 近年盛んに研究開発が行われている。多くの GNN は GCN¹ などのグラフ畳み込み層を積み重ねることにより構成される。これらのグラフ畳み込み層は, 情報の伝播を近傍ノードに限定することで, グラフの構造を反映したノード間の相互作用を反映し, 特徴量を更新することができる。より広範なノードの作用を考慮して特徴量を抽出するには, より多くのグラフ畳み込み層を積み重ねることが必要である。他方で, GCN などのグラフ畳み込み層では, 総数を増やすことにより性能低下を生じてしまう過平滑化 (over-smoothing)² と呼ばれる問題が発生することが知られていた。この現象は, 各ノードの特徴ベクトルが定常ベクトルに収束してしまうことにより, ノードの見分けが付かなくなってしまうことが原因であるとされている³。この問題によって従来の GNN モデルは, 一定以上遠くのノードの情報を考慮することができず, それにより精度も頭打ちとなってしまう課題を抱えていた。

過平滑化を解決するための手段はこれまでいくつかの研究で提案されてきた。初期の試みとして Residual Connection⁴を導入する手法が提案されたが, この手法は過平滑化の始まるタイミングを遅らせることはできるものの, その発生を防止することはできないことが Huang らにより指摘されている⁵。同じく Huang らによる DropEdge では入力グラフのエッジをランダムに切り落とすことにより汎化性能の獲得を目指すアプローチが, JKNet⁶では各層の出力を集約して最終層に渡すことにより異なるスケールの特徴を抽出するアプローチが提案された。これら 2 つのアプローチは一定の効果を示すものの, 層数が深くなることにより性能低下に対する抑制効果が低下し

¹ Kipf, T., & Welling, M. (2017). Semi-Supervised Classification with Graph Convolutional Networks. In Proceedings of the 5th International Conference on Learning Representations.

² Li, Q., Han, Z., & Wu, X.M. (2018). Deeper Insights into Graph Convolutional Networks for Semi-Supervised Learning. In Proceedings of the 32nd AAAI Conference on Artificial Intelligence (pp. 3538-3545). AAAI Press.

³ Wang, G., Ying, R., Huang, J., & Leskovec, J. (2019). Improving Graph Attention Networks with Large Margin-based Constraints. CoRR, abs/1910.11945.

⁴ He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770-778.

⁵ Huang, W., Rong, Y., Xu, T., Sun, F., & Huang, J. (2020). Tackling Over-Smoothing for General Graph Convolutional Networks. CoRR, abs/2008.09864.

⁶ Xu, K., Li, C., Tian, Y., Sonobe, T., Kawarabayashi, K., & Jegelka, S. (2018). Representation Learning on Graphs with Jumping Knowledge Networks. CoRR, abs/1806.03536.

てしまうことが指摘されている⁷。このように過平滑化を解決するための様々なアプローチが提案されてきたものの、いまだに本質的な問題解決には至っていない。

過平滑化によって十分に離れたノードの情報を考慮することができない課題は、特に大規模なグラフを対象とするタスクにおいて大きな障壁となる。その一例として、有限要素法 (FEM) による応力計算のサロゲートモデルへの応用例が挙げられる。

FEM とは、構造解析などにおいて利用される数値解析手法の 1 つである。FEM の計算量は節点数や要素数の増加に伴い増加するため、サロゲートモデルを構築・利用することで計算の高速化を目指す研究が過去に行われてきた。Amir et al. (2021) はメッシュデータとその各節点に対して FEA 計算により求めた全体剛性マトリクスに基づく特徴量を入力として与え、複数層の GCN によって FEM 計算における平衡方程式を求解する手法を提案した⁸。この研究では、離れたノードの影響を加味するためには GCN の層数を増やす必要がある一方で、8 層を超えたあたりからいわゆる過平滑化を起こして性能低下を生じてしまうという問題があった。そこで、同研究では全ノードの特徴量から PCA によりグローバルな特徴を抽出することで、離れたノードの影響を考慮できるようにするアプローチを採用した。他方で、この方法はノードの位相的な情報を無視した情報の伝播を行ってしまう。

本研究は、この過平滑化の問題を根本的に解決するための新たな手段を提供することを目的とする。過平滑化を生じることなく多層化を達成することによって高い精度を達成し、大規模なグラフの効果的な特徴表現を可能にすることを目指した。さらに、この大規模なグラフの処理における優位性を活用して、FEM における応力計算を高い再現性で実現することを目指した。また本研究では、後述の考察において過平滑化の要因と考えられる GNN の構造的な問題を指摘し、本研究が提案する新しいモデル構造がその問題の解消に成功していることを定式的に裏付けた。

本研究における実験の結果、提案手法は GNN のノード分類のベンチマークにおいて最も高い精度を達成し、State-of-the-Art を記録した。また FEM のサロゲートモデルとしても高い有効断尾を確認した。さらに、本研究における実験の結果は、提案手法が過平滑化の問題を克服して、層数に対して単調に精度が向上することを定量的に裏付けた。

2. 関連研究

Transformer⁹をベースとする手法が、深層学習の多くの分野で高い性能を発揮している。CNN や RNN などの対象とするドメインに対してその特徴を抽出するのに有効な inductive bias¹⁰を与

⁷ Chen, M., Wei, Z., Huang, Z., Ding, B., & Li, Y. (2020). Simple and Deep Graph Convolutional Networks. CoRR, abs/2007.02133.

⁸ Amir, H., Koichi, S., Tsuguchika, T., & Hiroshi, O. (2021). Surrogate Model for Structural Analysis Simulation using Graph Convolutional Network (Unrefereed Workshop Manuscript). Fujitsu Ltd., Fujitsu Ltd., Fujitsu Ltd., The University of Tokyo.

⁹ Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, L., & Polosukhin, I. (2017). Attention is All you Need. In NIPS (pp. 5998–6008).

¹⁰ Battaglia, P., Hamrick, J., Bapst, V., Sanchez-Gonzalez, A., Zambaldi, V., Malinowski, M., Tacchetti, A., Raposo, D., Santoro, A., Faulkner, R., Gülçehre, Ç., Song, H., Ballard, A., Gilmer, J., Dahl, G., Vaswani, A., Allen, K., Nash, C., Langston, V., Dyer, C., Heess, N., Wierstra, D., Kohli, P., Botvinick, M., Vinyals, O., Li, Y., & Pascanu, R. (2018). Relational inductive biases, deep learning, and graph networks. CoRR, abs/1806.01261.

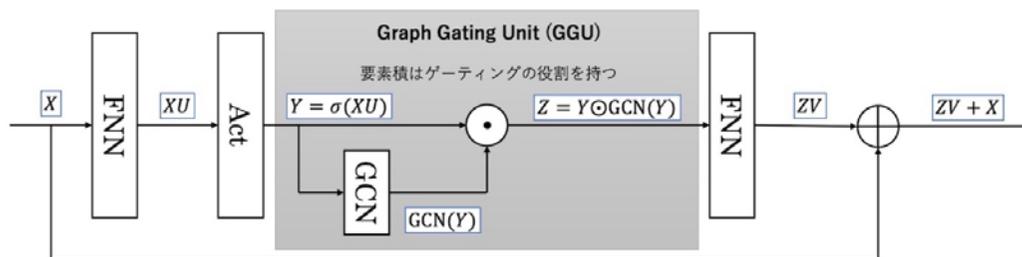
えるネットワークを直接利用するアプローチから、Transformer や gMLP¹¹のように MLP をベースとして付加的にゲートの機構を利用するアプローチへのある種のパラダイムシフトとも捉えることができる。当初は Transformer に利用された Multi-Head Self Attention の持つ inductive bias がこの性能向上に寄与したと考えられ、実際 GNN の分野においても Graph Transformer Networks¹²などの Multi-Head Self Attention を利用した手法が提案された。他方で、この手法の性能向上は特定のタスクに限定された。gMLP を提案した Liu らは、この Transformer や gMLP の優位性は、Multi-Head Self Attention の与える inductive bias ではなく、MLP とそれに対するゲートの機構にあることを指摘した。

本研究においては、これらのゲートの役割を「対象ドメインに有効な inductive bias を与えること」と再定義し、MLP に対してグラフに有効な inductive bias を持つゲート機構を付加した新たなネットワーク構造を提案した。

3. 提案手法

本論文では、MLP をベースとして、Graph Gating Unit (GGU) と名付けた GCN によるゲート構造を付加した新しいネットワークである、Graph Gating Multi-Layer Perceptron (g2-MLP) を提案した。

提案手法のアーキテクチャは、GCN ではなく Feedforward Neural Network (FNN) の積み重ねである Multi-Layer Perceptron (MLP) をベースとする。各層は第 1 図に示すような、各ノードの特徴を独立に 2 層の FNN と勾配消失・爆発を抑制するための Residual Connection からなる基本構造と、ノード間の相互作用を考慮するための GGU から構成される。GGU では、1 層の GCN の入出力の要素積を計算することで、GCN によるゲーティング構造を実現している。GCN はグラフの構造を反映しながらノード間の相互作用を考慮した表現学習を行うことができるため、このゲート構造によりグラフに有効な特徴を抽出することを図っている。



第 1 図: 1 層の g2-MLP のネットワーク構造。

第 1 図で示した構造を複数層積み重ねた構造を、新たに g2-MLP と名付けて提案した。

4. 数値実験

¹¹ Li, Q., Han, Z., & Wu, X.M. (2018). Deeper Insights into Graph Convolutional Networks for Semi-Supervised Learning. In Proceedings of the 32nd AAAI Conference on Artificial Intelligence (pp. 3538-3545). AAAI Press.

¹² Yun, S., Jeong, M., Kim, R., Kang, J., & Kim, H. (2019). Graph Transformer Networks. In NeurIPS (pp. 11960-11970).

本研究では、PPI データセット¹³と新たに作成した FEM データセットを用いて、既存手法と g2-MLP の性能を比較した。PPI データセットはノード分類の代表的なデータセットであり、タンパク質から構成される 24 のヒト組織について、タンパク質の特徴量を入力として各タンパク質における 121 の遺伝子オントロジーを予測するタスクである。本研究において新規に作成したデータセットである FEM データセットは、複数の節点から構成される 229 の有限要素メッシュについて、各節点の座標情報を入力として、各節点における応力を予測するタスクである。

PPI データセットでは、比較対象とした全ての手法においてその分類精度が論文中で示されていたため、提案手法による実験の結果と、各論文で提示されている結果との間で精度の比較を行った。なお評価指標としては、先行研究に合わせて Micro-F1 値を採用した。FEM データセットでは、比較対象のモデルとして GCN と GAT¹⁴のそれぞれのモデルについて、学習・推論を行った。評価指標としては、回帰のタスクであることから平均二乗誤差を採用した。

5. 結果

第 1 表に示すように、提案手法である g2-MLP は 99.71%の最高精度を達成して、State-of-the-Art を更新することに成功した。既存手法の多くが 5 層から 9 層で高い精度を発揮しているのに対して、本手法は 20 層の構造で最も高い精度を達成した。

第 1 表: PPI データセットにおける精度の比較

Model	Micro-F1 Score
g2-MLP (20 層)	99.705(14)
GCNII (9 層)	99.56(2)
Cluster-GCN (5 層)	99.36
GaAN (5 層)	98.71(2)
GraphSAINT (5 層)	98.10(40)
JKNet (3 層)	97.60(7)
GAT (2 層)	97.30(2)
GraphSAGE (2 層)	61.2

さらに次の第 2 表で示すように、g2-MLP は FEM データセットでも最も高い精度を示した。

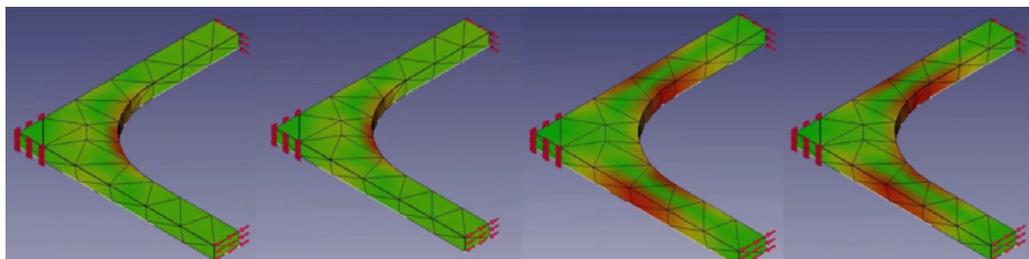
第 2 表: FEM データセットにおける精度の比較

Model	Displacement	Stress
g2-MLP	0.637(286)	0.879(370)
MLP	0.857(160)	1.481(6)
GAT	0.971(145)	1.598()
GCN	0.989(11)	1.602(5)

¹³ Hamilton, W., Ying, R., & Leskovec, J. (2017). Inductive Representation Learning on Large Graphs. CoRR, abs/1706.02216.

¹⁴ Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liù, P., & Bengio, Y. (2017). Graph Attention Networks. In ICLR 2018.

テストデータの1つについて、各モデルによる応力の予測結果を第2図に図示した。

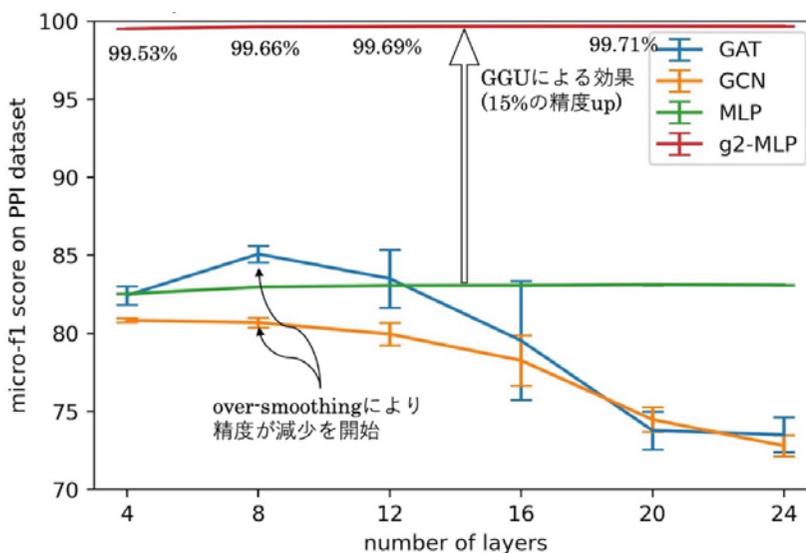


第2図: 左から GCN, GAT, FEM (正解データ), g2-MLP によるミーゼス応力の推論結果

g2-MLP の推論結果では、他のモデルと比較してもより高い再現性で FEM の計算結果を近似できていることが定性的に確認できる。例えば、GCN や GAT には予測することのできていないフィレット部分外側やその裏側における応力を予測できていることが確認された。

6. 考察

層数を変えた時の各モデルの分類精度の変化を第3図にプロットした。

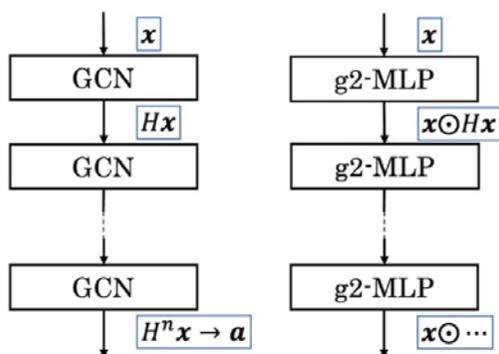


第3図: PPI データセットにおける層数と精度の関係

GAT や GCN では 4 から 8 層で精度が低下するのに対して、g2-MLP は 20 層に至るまで層数に対して単調に精度を向上することが確認できる。この結果から、g2-MLP は確かに過平滑化の問題を解消できていると結論付けられる。

このように g2-MLP が過平滑化の問題を解消することができたのは、GCN と g2-MLP の構造上の違いが理由であると考察した。1次元の入力ベクトル \mathbf{x} を GCN と g2-MLP のそれぞれに入力した時の特徴ベクトルの更新の様子は、グラフの形状に依存するある行列を H として、第4図のように

表される。



第 4 図: GCN と g2-MLP における入力特徴ベクトルの更新の様子

左側の GCN の計算からは、 $H^n x$ という出力が得られるが、このベクトルは十分大きい n に対して行列 H の絶対値最大固有値に対応するある固有ベクトルに収束することが線形代数のべき乗法によって示される。すなわち、層数を増やすことにより入力はその特徴に依存しないベクトルに収束し、入力時の情報を失ってしまうことになる。他方で右の g2-MLP においては、行列 H はゲート内部にあるため H^n を生じず、出力においても入力特徴の情報を失うことなく変換を行うことができる。

このように g2-MLP は、定式的に過平滑化の原因と考えられる特徴更新の問題を原理的に解消し、確かに過平滑化を起こさずに層数に対して精度を向上できることを定量的に確認することができた。

7. 結論・課題

本研究では、過平滑化の問題を解消して高い精度を達成するため、また GNN における大規模なグラフの処理を可能にするための手法として、g2-MLP と名付けた手法を提案した。g2-MLP は PPI ベンチマークにおいて既存手法を上回る精度を達成し、State-of-the-Art を更新することに成功した。また g2-MLP は従来の GNN モデルの構造的問題を解消し、分野の課題であった過平滑化の問題を根本から解消して多層化を実現することに成功した。これらの結果は、今後 GNN モデルの利用・開発において本手法が新たなベースラインとなりうる可能性を示唆する。また FEM のサロゲートモデルとしても、高い有効性が確認できた。

今回はノード分類に焦点をおいて研究を行ったが、グラフそのものを分類するグラフ分類への応用には、研究の余地があると考えられる。いくつかのグラフ分類の実験における g2-MLP の推論結果では、ノード分類ほど高い効果を確認することはできなかった。このようなタスクにおいては、g2-MLP のより適切な利用方法を画策していく必要があると考えられる。