

適応的なノード埋め込みの伝搬による半教師ありノード分類モデル

小川 裕也[†] 前川 政司[†] 佐々木 勇和[†] 藤原 靖宏^{††} 鬼塚 真[†]

[†] 大阪大学大学院情報科学研究科 ^{††} NTT コミュニケーション科学基礎研究所

1 はじめに

グラフ畳み込みニューラルネットワーク (GCN) は半教師ありノード分類で成功を収めており、広く研究されている。GCN は各層でグラフ上で埋め込みを伝搬 (隣り合うノードの埋め込みを混合) することでグラフ情報を属性情報を組み合わせる。しかし、既存手法は層が浅く伝搬のホップ数が少ないためラベル情報の伝搬が不十分である。その結果、ラベルノードがごく少数の場合大きく精度が劣化する。一方で、より広く伝搬するために層の数を増加させると、パラメータの増加による過適合と多数のグラフ畳み込みによる埋め込みの過平滑化 (全ての埋め込みが似た値に収束する現象) が原因で精度が低下する。

本稿では、過適合と過平滑化を回避しながら埋め込みを広く伝搬する ANEPN を提案する。加えて、最適な伝搬のホップ数がグラフによって異なることに着目して、入力グラフに応じて適応的に伝搬のホップ数を制御する。過適合と過平滑化に対処するために、平滑化損失及び反平滑化損失を導入する。平滑化損失を最小化するために勾配降下法による埋め込みの更新をすることで埋め込みが伝搬される。この損失によって埋め込みの伝搬をすることで、ANEPN では伝搬のホップ数に関わらずパラメータが一定となり、過適合を回避する。一方で反平滑化損失はグラフ上で離れたノードが近い埋め込みを持つことに罰則を科す。よって、過平滑化、すなわち全ての埋め込みが似た値を持つことを回避する。また、適応的に伝搬のホップ数を制御するために、埋め込みと予測ラベルを評価する指標を導入する。その指標に基づいて同じクラス内の埋め込みが近づき異なるクラス間で埋め込みが遠ざかるように伝搬のホップ数を決定する。評価実験において、提案手法が3つの標準的なグラフ上で既存手法を上回る精度を達成することを示した。

2 事前準備

N 個のノードを持つ属性付きグラフは隣接行列 $\mathbf{A} \in \{0, 1\}^{N \times N}$ で表される。また、各ノードの F 種類の属性及び C 種類のクラスラベル情報は属性行列 $\mathbf{X} \in \mathbb{R}^{N \times F}$ 及びクラスラベル行列 $\mathbf{T} \in \{0, 1\}^{N \times C}$ で表される。 \mathbf{X} の行ベクトルは各ノードの属性値を表し、 \mathbf{T} の行ベクトルは各ノードのクラスラベル (one-hot 形式) を表す。半教師ありノード分類の目的は、少数のノードのみがクラスラベルを与えられ多くのノードはクラスラベルが与えられない (行ベクトルが零ベクトル) 条件のもとで、ラベルが与えられていないノードのクラスを予測することである。

GCN [1] は半教師ありノード分類のために提案されたモデルであり、各層においてグラフ畳み込みとニューラルネットワークによる変換を行い、最終的に予測ラベル行列 $\mathbf{Y} \in \mathbb{R}^{N \times C}$ を出力する。本稿では隠れ層の出力をノード埋め込み $\mathbf{Z} \in \mathbb{R}^{N \times H}$ と呼ぶ。ここで H は埋め込みの次元数を表す。以下の損失関数を最小化するようにネットワークパラメータを学習する。

$$L_{ce} = \sum_{i,c=1}^{N,C} -\mathbf{T}_{ic} \log(\mathbf{Y}_{ic}). \quad (1)$$

3 提案手法

提案手法 ANEPN は GCN を拡張して、 α を係数とする平滑化損失 L_{sm} と反平滑化損失 L_{asm} を加えられた損失関数を持つ:

$$L = L_{ce} + \alpha L_{sm} + \alpha L_{asm} \quad (2)$$

平滑化損失 L_{sm} は埋め込みが類似していない隣接し合うノードに罰則を課し、以下のように定義される。

$$L_{sm} = \frac{1}{2H} \sum_{h,i,j=1}^{H,N,N} \mathbf{A}_{ij} \left(\frac{\hat{\mathbf{Z}}_{ih}}{\sqrt{d_i}} - \frac{\hat{\mathbf{Z}}_{jh}}{\sqrt{d_j}} \right)^2 \quad (3)$$

ここで $\hat{\mathbf{Z}}$ は列正規化された埋め込みでありその列ベクトルは $\hat{\mathbf{Z}}_{:,h} = \frac{\mathbf{Z}_{:,h}}{\|\mathbf{Z}_{:,h}\|}$ 。この損失を最小化するように勾配降下法によって埋め込みを更新することで、以下のようにあるノード i の埋め込みは隣接するノード集合 $\Gamma(i)$ と混合される、すなわち伝搬される。

$$\mathbf{Z}_{ih} \leftarrow \beta_i \mathbf{Z}_{ih} + \sum_{j \in \Gamma(i)} \gamma_{ij} \mathbf{Z}_{jh} \quad (4)$$

ここで β_i 及び γ_{ij} はそれぞれノード i 自身への伝搬係数及びノード i に対するノード j からの伝搬係数を

Semi-supervised Node Classification with Adaptive Node Embedding Propagation

[†] Yuya Ogawa

[†] Seiji Maekawa

[†] Yuya Sasaki

^{††} Yasuhiro Fujiwara

[†] Makoto Onizuka

Graduated School of Information Science and Technology, Osaka University

([†])

NTT Communication Science Laboratories (^{††})

表 1: Cora, Citeseer 及び Pubmed 上での分類精度. 太字は各列において最も高い値を表す.

ラベル率	Cora			Citeseer			Pubmed		
	0.5%	2%	4%	0.5%	2%	4%	0.03%	0.05%	0.1%
GCN	44.5	68.7	77.0	43.6	61.7	68.6	45.6	55.0	64.9
GAT	41.1	54.2	77.0	40.1	62.8	68.7	50.2	53.0	60.5
M3S	59.9	75.8	79.2	54.2	66.2	70.4	57.0	62.9	68.4
ALaGCN	57.9	73.7	78.5	41.0	59.3	67.2	57.1	63.0	71.4
ALaGAT	48.2	73.5	77.3	38.4	58.6	68.4	56.8	62.4	69.3
ANEPN	66.1	77.6	79.9	60.5	68.8	71.0	60.8	64.4	71.4

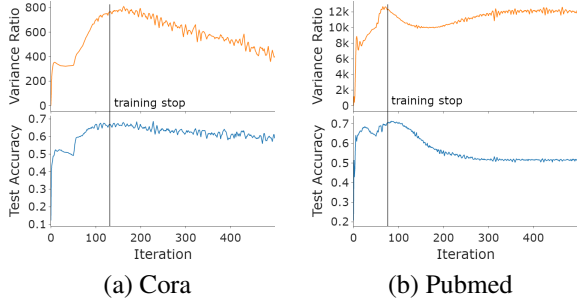


図 1: 分類精度と Variance Ratio

表す. 既存の GCN が層の数を増加させることで伝搬のホップ数を増加させる一方で, 提案手法では埋め込みの更新の回数を増加させることでネットワークパラメータを増加させずに伝搬のホップ数を増加させることが可能となるため過適合を回避する.

次に, 反平滑化損失 L_{asm} はノードの埋め込みが類似している直接隣接していないノードに罰則を科し, 以下のように定義される.

$$L_{asm} = \max(\mu - \frac{1}{2H} \sum_{h,i,j=1}^{H,N,N} \bar{A}_{ij} (\frac{\hat{Z}_{ih}}{\sqrt{d_i}} - \frac{\hat{Z}_{jh}}{\sqrt{d_j}})^2, 0). \quad (5)$$

\bar{A}_{ij} は隣接行列 A の値を反転 ($0 \leftrightarrow 1$) させた行列である. μ は L_{asm} による罰則の強さを決めるマージンパラメータである. 全ての埋め込みが似ることを防ぐため, 過平滑化を回避することが可能となる.

さらに, 適応的に伝搬のホップ数を制御するために, ノード埋め込みと予測ラベルに対してクラス内分散 (クラス内の埋め込み間の距離の二乗和) とクラス間分散 (異なるクラス間の埋め込みの距離の二乗和) を評価する指標 Variance Ratio を導入する. Variance ratio は値が高いほどクラス内の埋め込みが近く, またクラス間の埋め込みが遠いことを示す. 提案手法では, この指標に基づいてクラス内分散を下げクラス間分散を大きくするように伝搬のホップ数を制御する. すなわち, Variance Ratio が高くなるように埋め込みの更新の反復を行い, 下がった時点で停止する.

4 評価実験

提案手法の分類精度を評価するため, GCN [1] と M3S [2], そして ALaGCN [3] との比較を行う. 使用するグラフは Cora (ノード数 2708, 属性数 1433, クラス数 7) と Citeseer (ノード数 3312, 属性数 4732, クラス数 6) 及び Pubmed (ノード数 19717, 属性数 500, クラス数 3) の 3 つである. 表 1 にそれぞれのグラフ上での分類精度の結果を示す. 表が示すように, 提案手法が全ての場合において既存手法を上回るか同等の精度を達成している. 特にラベル率が低い場合 (ラベルノードが少ない場合) に分類精度を大きく向上させている. 次に提案手法が適応的に伝搬のホップ数を制御しているかどうかを検証するために, 図 1 に更新の反復 (iteration) に対する Variance Ratio とテストデータの分類精度を示す. 図中の training stop は反復の停止した点を表している (ただし, 評価のために停止後も反復を続けている). 最適な反復の回数 (すなわち, 伝搬のホップ数) が異なるグラフ (Cora と Pubmed) に応じて, 提案手法では最適に近い回数に制御していることがわかる.

謝辞

本研究は JSPS 科研費 JP20H00583 の助成を受けたものです.

参考文献

- [1] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *ICLR*, 2017.
- [2] Ke Sun, Zhouchen Lin, and Zhanxing Zhu. Multi-stage self-supervised learning for graph convolutional networks on graphs with few labels. In *AAAI*, 2020.
- [3] Y. Xie, Sha Li, Carl Yang, Raymond Chi-Wing Wong, and Jiawei Han. When do gnns work: Understanding and improving neighborhood aggregation. In *IJCAI*, 2020.