

ノイズを含むグラフにおけるノード分類のための対照学習モデルの集団学習

Ensembling Contrastive Learning Models for Node Classification on Noisy Graphs

日野雄斗^{1*} 兼岩憲¹
Yuto Hino¹ Ken Kaneiwa¹

¹ 電気通信大学大学院 情報理工学研究科 情報・ネットワーク工学専攻

¹ Department of Computer and Network Engineering, Graduate School of Informatics and Engineering, The University of Electro-Communications, Tokyo, Japan

Abstract: Graph Neural Networks (GNNs) learn from graph-structured data by aggregating and updating node features based on neighborhood structure. Although GNNs have achieved high accuracy for node and graph classifications, many methods implicitly assume strong homophily and the absence of missing data or noise. Recently, contrastive learning has been explored to obtain useful node embeddings on graphs and is considered to offer a certain degree of robustness to noise. However, existing contrastive approaches still face limitations in representation power and training stability when graphs are noisy. To address this issue, we propose a contrastive framework that combines H2GCN which explicitly separates 1-hop and 2-hop neighborhoods and is effective on heterophilous graphs with NAGphormer, a Transformer based model that captures global dependencies. In addition, we employ an ensemble scheme that integrates multiple contrastive learners. Experimental results on nine benchmark datasets spanning both homophilous and heterophilous settings with injected noise demonstrate that our method consistently outperforms existing baselines.

1 はじめに

近年、グラフ構造で表現されるデータが多く存在し、SNS や分子構造、引用関係など、さまざまな関係性を表現できる。このようなグラフ構造に対して、ニューラルネットワークを適用し、機械学習を実行する手法としてグラフニューラルネットワーク (GNN) がある。GNN はグラフ畳み込みニューラルネットワーク (GCN) や H2GCN などのモデルがあり、ノードの特徴ベクトルを近傍ノードから集約することで、ノード分類やリンク予測などのタスクにおいて高い性能を示している [1, 2]。

しかし、多くの GNN モデルは同種性グラフを前提としており、異種性グラフを学習する時に効果的な特徴の集約ができず、精度が低くなる。この問題を解決するために提案された H2GCN は、集約する特徴ベクトルを深さによって区別することで同種性と異種性の両方に適応できる。また、自然言語処理の分野で優

れた結果を残している Transformer をグラフ構造に応用する研究も進められている。その中で、NAGphormer は自己注意機構を活用してグローバルな関係性を捉えて、長距離グラフのタスクで高い性能を示している [4]。

このようなモデルが登場する一方で、現実で作成されたグラフデータには、しばしばノイズやラベルの誤りが含まれており、これらは機械学習の妨げとなる。また、ラベル付きデータの取得には多大なコストがかかることも多く、少数ラベルで学習を行う手法の重要性が増している。こうした課題に対して、対照学習は近年注目を集めている [3]。対照学習は、ラベル付きデータが少ない場合でも類似データのペアを通じて有用なノード埋め込みを効果的に学習する。特に、異なるタイプの GCN と NAGphormer を組み合わせた GNN-Transformer Co-contrastive Learning (GTC) のような対照学習モデルが考案されている [5]。しかし、対照学習によりビュー間の一貫性を利用して頑健性を高める一方、局所構造と大域構造の双方を十分に活用しつつ、ノイズを含む同種性と異種性のグラフを安定して学習させる枠組みは十分に確立されていない。

*連絡先： 電気通信大学情報理工学研究科
〒 182-8585 東京都調布市調布ヶ丘 1-5-1
E-mail: hino@sw.cei.uec.ac.jp

本研究では、局所構造に着目する H2GCN と大域的依存関係を捉える NAGphormer の相補性に着目する。その組み合わせによって、異種性を持つグラフやエッジの接続不足や特徴ノイズ、ラベルノイズを含むさまざまなグラフデータに対して安定かつ高精度なノード分類を実現する。また、複数の対照学習モデルに対する集団学習を導入することで、さらなる学習の安定化と性能向上を図る。

評価実験では、隣接行列および特徴ベクトルへのランダムドロップアウト、ならびに一部ラベルの置換などによってノイズ条件を制御し、提案手法の有効性を分析する。比較対象として GCN, H2GCN, NAGphormer に加え、従来手法である GTC を用い、提案手法との精度比較を行う。

本研究の貢献は次のとおりである。

- (i) GCN の 1-hop, 2-hop 分離に基づく H2GCN と Transformer に基づく NAGphormer の 2 つの異なるモデル間に対する対照学習を提案し、ノイズ下での表現安定性を高める。
- (ii) 2 つの対照学習 H2GCN+NAGphormer と GCN+NAGphormer の集団学習を導入し、ノイズ下における対照学習に基づく集団学習の有効性を示す。

2 準備

2.1 半教師あり学習によるノード分類

ノードの集合 V とエッジの集合 E を持つグラフを $G = (V, E)$ とする。ノード集合 $V_L \subset V$ に含まれる各ノード $v \in V_L$ は、ラベル $y_v \in Y$ を持つ。ここで、 Y はラベルの集合である。ノード分類は、ラベルのないノード $v \in V \setminus V_L$ のラベルの予測を行う。また、各ノード $v \in V$ は特徴ベクトル X_v を持つ。

2.2 同種性

homophily ratio とは、グラフの全てのエッジのうち同じラベルを持つノード同士を繋ぐエッジの割合を表していて、以下の式で求められる。

$$h = \frac{|\{(u, v) : (u, v) \in E \wedge y_u = y_v\}|}{|E|} \quad (1)$$

この値が高いグラフは同種性があり、低いグラフは異種性があると判断できる。また、多くの GNN モデルは同種性を持つグラフはノード分類の精度が高く、異種性をもつグラフはノード分類の精度が低くなる傾向がある。

2.3 対照学習

対照学習は、グラフにおける各ノードのラベルの有無にかかわらず、類似するデータ同士を近づけ、異なるデータを遠ざけるように特徴表現を学習する自己教師あり学習手法である。自己教師ありノード表現学習手法は 1 つのグラフ $G = (V, E)$ に対して、構造と特徴の両方にランダムなデータ拡張を適用し、2 つのビュー $G^{(1)} = (V^{(1)}, E^{(1)})$, $G^{(2)} = (V^{(2)}, E^{(2)})$ を生成することで対照学習を行う。各ビューはエンコーダで処理され、以下のように特徴表現 $z_v^{(k)}$ が得られる。

$$z_v^{(k)} = \text{MLP}(\text{Model}[X_v^{(k)}, A^{(k)}]) \quad (2)$$

ここで、 $k \in \{1, 2\}$ はビューのインデックス、 $X_v^{(k)}$ はビュー $G^{(k)}$ におけるノード $v \in V^{(k)}$ の特徴、 $A^{(k)}$ はビュー $G^{(k)}$ における隣接行列、MLP は多層パーセプトロンである。2 つのビューから得られたノード v の特徴表現 $z_v^{(1)}, z_v^{(2)}$ に対して、次の損失 $\mathcal{L}$ を用いて対照的に学習を行う。

$$f(z, z') = \exp(\text{sim}(z, z')/\tau) \quad (3)$$

$$g(z, z') = \sum_{w \in V/\{v\}} \exp(\text{sim}(z, z')/\tau) \quad (4)$$

$$\begin{aligned} \ell(z_v^{(1)}, z_v^{(2)}) = \\ -\log \frac{f(z_v^{(1)}, z_v^{(2)})}{f(z_v^{(1)}, z_v^{(2)}) + g(z_v^{(1)}, z_w^{(2)}) + g(z_v^{(1)}, z_w^{(1)})} \end{aligned} \quad (5)$$

$$\mathcal{L}^{(1+2)} = \frac{1}{2|V|} \sum_{v \in V} (\ell(z_v^{(1)}, z_v^{(2)}) + \ell(z_v^{(2)}, z_v^{(1)})) \quad (6)$$

ここで、 $\text{sim}(\cdot, \cdot)$ はコサイン類似度、 τ は温度パラメータである。対照学習は、同一ノード v の異なるビュー $G^{(1)}, G^{(2)}$ 間の表現を一致させ、他のノード w からは区別されるように学習することで、ラベルなしデータから意味のあるノード表現 $z_v^{(1)}, z_v^{(2)}$ を獲得できる。

2.4 集団学習

集団学習とは、複数の学習器による予測を組み合わせる最終予測を決める方法である。ノードの集合を V とし、ラベル集合を Y 、ラベル付きノード集合を $V_L \subset V$ とする。 M 個の基学習器 $h_1, \dots, h_M$ を V_L 上で学習し、各学習器がノード v に対してラベル y をとる確率 $p_m(y | v)$ を用いてノード分類予測を以下の式で表す。

$$\hat{y}_v = \arg \max_{y \in Y} \frac{1}{M} \sum_{m=1}^M p_m(y | v) \quad (7)$$

3 対照学習モデルの集団学習

GNN は、隣接ノードからの情報を段階的に集約してグラフの局所構造を捉えるが、長距離のノード間関係を捉えるのが難しい。一方で、Transformer は自己注意機構により、距離に依存せずノード間の依存関係を同時に捉える。このように相補的な GNN と Transformer の組み合わせは、多様な構造情報を活かした頑健なノード表現が学習できる [5]。

GNN モデルの中でも H2GCN は 1-hop および 2-hop の隣接情報を明示的に分離して扱うことで、同種性の低いグラフにおいても有効なノード埋め込みが可能である [2]。このような H2GCN の性質は、グローバルな構造を捉える Transformer との補完関係を構築し、さまざまなグラフの学習で特に有効である。本研究では、それぞれの異なる構造的視点をビューとして対照学習を行うことで、より多面的なノード埋め込みを獲得する。さらに、提案手法と従来手法である GTC の複数の対照学習で得られた予測を組み合わせる集団学習を導入する。これにより、H2GCN の異種性や構造的ノイズに対する頑健性と GCN の同種性の高い領域で平滑化に基づく安定性の両方によって、ノイズ環境下においてさらなる精度向上を実現する。

提案手法では、対照学習で得られた埋め込み表現に対してロジスティック回帰による線形評価を行い、ノード分類をする。また、異なる対照学習モデルから得られた埋め込み表現ごとに独立にロジスティック回帰モデルを学習し、両者の予測確率を平均化することで集団学習によりラベルを決定する。図 1 にグラフデータに対する対照学習モデルの概要図を、図 2 に 2 つの対照学習の結果を用いた集団学習の概要図を示す。

3.1 ノイズを含むグラフの作成

本研究では、ノイズ環境下における対照学習の有効性を検証するために、各データセットの隣接行列に対してランダムなドロップアウトを行い、特徴ベクトルに対してランダムなドロップアウトまたは、ガウシアンノイズを適用し、さらにラベルの一部を他クラスに置き換えることでノイズを導入する。また、実環境を想定し、学習前にデータセット全体に対してノイズを注入する。

- ドロップアウト構造ノイズ

隣接行列 A の非零成分集合を $S = \{(i, j) \mid A_{ij} \neq 0\}$ とする。エッジ削除率 $p_{\text{adj}} \in [0, 1)$ に対し、 $k = \lfloor p_{\text{adj}} \cdot |S| \rfloor$ 個のインデックス集合 $S_{\text{drop}} \subset S$ を選び、対応する成分を 0 にして $\tilde{A}$ を得る。

- ドロップアウト特徴ノイズ

特徴行列 X に対し、特徴次元ごとに独立なベルヌーイ乱数で $m \in \{0, 1\}^d$ を生成する。特徴削除率 q_{feat} を用いて、各要素を $m_j \sim \text{Bernoulli}(1 - q_{\text{feat}})$, ($j = 1, \dots, d$) として $\tilde{X} = X \odot m$ を得る。

- ガウシアン特徴ノイズ

特徴行列 X に対して、各要素に平均 0、標準偏差 σ_{feat} のガウシアンノイズを加える。ノイズ行列 $N \sim \mathcal{N}(0, \sigma_{\text{feat}}^2 I)$ を生成し、ノイズ付き特徴 $\tilde{X} = X + N$ を得る。

- ラベルノイズ

ラベル付き訓練ノード集合を $V_L \subseteq V$ 、ラベルノイズ率 $r_{\text{label}} \in [0, 1)$ に対して、ノイズ付与対象の $k_{\text{label}} = \lfloor r_{\text{label}} |V_L| \rfloor$ 個の訓練ノード集合 $S_{\text{label}} \subseteq V_L$ を得る。各 $v \in S_{\text{label}}$ について、真のクラス Y_v を除いた残りのクラスから新ラベル $\tilde{Y}_v$ に置き換える。

3.2 H2GCN と NAGphormer の対照学習

本手法は、局所構造に強い H2GCN と大域依存を捉える NAGphormer の 2 つのビューを対照学習で整合させる。以降、学習に用いる基礎拡張入りデータを $(\tilde{X}, \tilde{A})$ とし、このモデルを H2GTC と記す。

H2GCN のビューを作るための拡張を各ビューのモデルに独立に適用し、 $(\tilde{X}^{(1)}, \tilde{A}^{(1)})$ を作成する。次に、対照学習で異なるビューを作成するためのグラフ変形としてドロップアウトを行い、 $(\tilde{X}'^{(1)}, \tilde{A}'^{(1)})$ を作成する。その後、このビューを入力とした H2GCN(付録 A.2 を参照) の出力表現 $h_v^{\text{H2}}[\tilde{X}'^{(1)}, \tilde{A}'^{(1)}]$ を得る。さらに、投影器 MLP により以下を得る。

$$z_v^{\text{H2}} = \text{norm}(\text{MLP}(h_v^{\text{H2}}[\tilde{X}'^{(1)}, \tilde{A}'^{(1)}])) \quad (8)$$

NAGphormer のビューを作るための拡張を各ビューのモデルに独立に適用し、 $(\tilde{X}^{(2)}, \tilde{A}^{(2)})$ を作成する。次に、対照学習で異なるビューを作成するためのグラフ変形としてドロップアウトを行い、 $(\tilde{X}'^{(2)}, \tilde{A}'^{(2)})$ を作成する。その後、このビューを入力とした NAGphormer(付録 A.3 を参照) の出力表現 $h_v^{\text{N}}[\tilde{X}'^{(2)}, \tilde{A}'^{(2)}]$ を得る。さらに、投影器 MLP により以下を得る。

$$z_v^{\text{N}} = \text{norm}(\text{MLP}(h_v^{\text{N}}[\tilde{X}'^{(2)}, \tilde{A}'^{(2)}])) \quad (9)$$

2 つのビュー間で得られた z_v^{H2} , z_v^{N} を用いて (6) 式の $\mathcal{L}^{\text{H2+N}}$ を最小化する。

対照学習終了後、2 つのビュー間で得られた特徴表現を以下の式で L^2 正規化する。

$$Z_v^{\text{H2+N}} = \frac{z_v^{\text{H2}} + z_v^{\text{N}}}{\|z_v^{\text{H2}} + z_v^{\text{N}}\|_2} \quad (10)$$

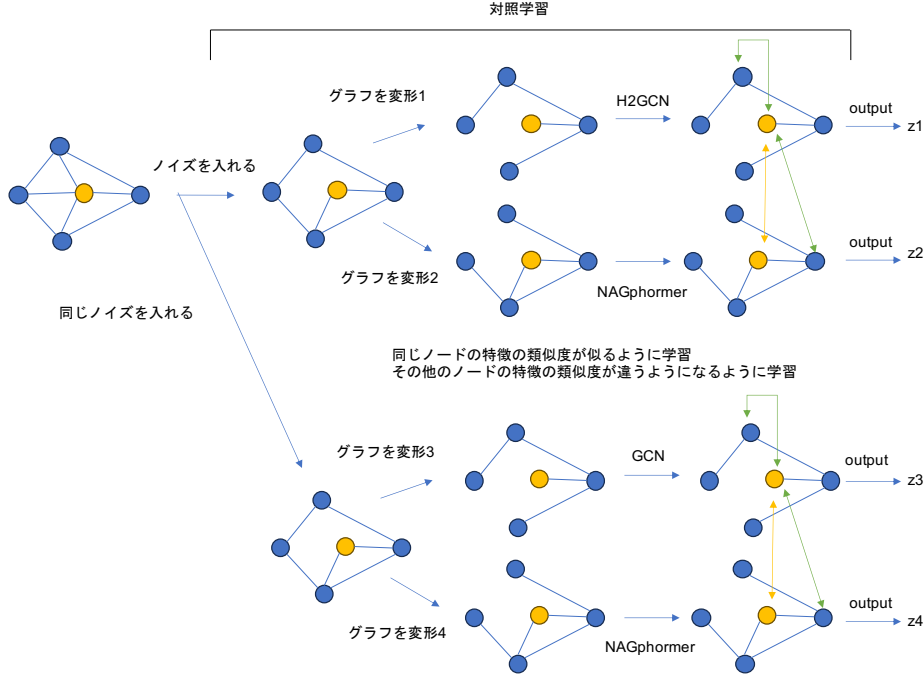


図 1: グラフデータに対する対照学習の概要図

次に、線形評価として、各対照学習の埋め込みごとに独立に分類器を学習し、分類器はロジスティック回帰として実装して以下の結果を得る。

$$\hat{y}_v = \arg \max_{y \in Y} \text{softmax}(\text{LogReg}(Z_v^{\text{H2+N}})) \quad (11)$$

3.3 対照学習に対する集団学習

異種モデル間の対照学習を複数実行して、それらの線形分類器の確率平均による集団学習を行う。まず、対照学習の埋め込みを式 (10) の $Z_v^{\text{H2+N}}$ と以下の式で L^2 正規化する。

$$Z_v^{\text{G+N}} = \frac{z_v^{\text{G}} + z_v^{\text{N}}}{\|z_v^{\text{G}} + z_v^{\text{N}}\|_2} \quad (12)$$

ここで、 z_v^{H2} , z_v^{N} , z_v^{G} はそれぞれ、対照学習によって学習された H2GCN, NAGphormer, GCN の出力である。

次に、線形評価として、各対照学習の埋め込みごとに独立に分類器を学習し、分類器はロジスティック回帰として実装する。最後に、各対照学習の結果に対して、以下の式で予測された結果を集団学習の予測とする。

$$\hat{y}_v = \arg \max_{y \in Y} \sum_{k \in \{\text{H2+N}, \text{G+N}\}} \text{softmax}(\text{LogReg}_k(Z_v^k)) \quad (13)$$

4 評価実験

本節では、提案手法のノード分類性能を従来手法と比較する。評価方法は先行研究に倣い、各データセットに対して、10%, 10%, 80%の割合でトレーニング、検証、テストデータセットに分割する。また、対照学習によって得られた埋め込みを L^2 正規化した後、ロジスティック回帰により線形分類器を学習し、ラベルの正解率を求める。最終的な精度は 3 回の実行の平均値とする。

4.1 データセット

本研究では、以下に示す GNN のベンチマークとして多く用いられている 9 個のデータセットを使用する。各データセットの詳細を以下に示す。

- Cora, Citeseer, Pubmed は引用データセットで、論文間の引用関係をノードを論文、エッジを引用関係として表している [6, 8]。各ノードの特徴ベクトルには、論文中に特定の単語が存在するかどうかを 0 か 1 で表す bag-of-words モデルが用いられている。ラベルは、各論文をいくつかのカテゴリに分類したものを用いている。
- Texas, Wisconsin, Cornell は大学の Web ページ間のリンク関係を表すデータセットで、ノードを大学の Web ページ、エッジをリンク関係として

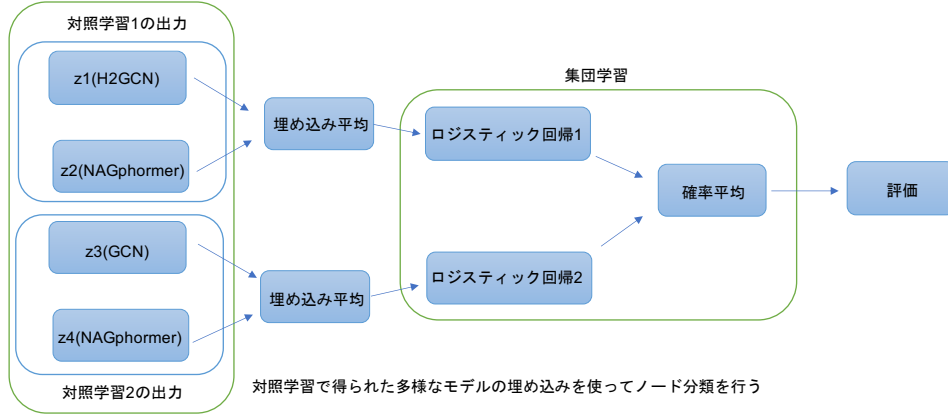


図 2: 2つの対照学習に対する集団学習の概要図

表している [7]. 特徴ベクトルには, bag-of-words モデルが使われている. ラベルは, 各 Web ページをいくつかのカテゴリーに分類したものをを用いている.

- Squirrel, Chameleon は英語版 Wikipedia 間の特定のトピックに関するページ間ネットワークを表す [7]. ノードは Web ページを表し, エッジはリンク関係を表している. 特徴ベクトルは, 記事内の特定の名詞の存在を示す. ラベルはページの月間平均トラフィックの観点から分類されている.
- Actor は, Wikipedia 上の共起を表すデータセットで, ノードは俳優を表し, エッジは Wikipedia 上の共起を表している [7]. ノードの特徴ベクトルは, Wikipedia 上のいくつかのキーワードに対応している. ラベルは, 俳優の Wikipedia の単語に関して分類されている.

4.2 実験設定

本研究では, GNN モデルとして GCN, H2GCN を, Transformer モデルとして NAGphormer を用いる. また, 実験に用いるノイズが含まれたグラフは, エッジ削除率 および特徴削除率を $p_{\text{adj}}, q_{\text{feat}} \in \{0.3, 0.5\}$ または, ガウシアンノイズ $\sigma_{\text{feat}} = 0.1$ とし. ラベルノイズ率を $r_{\text{label}} \in \{0.0, 0.3\}$ と設定する合計 6 パターンを作成する. さらに, 対照学習におけるビューを生成する際のドロップアウト率を $dr \in \{0.0, 0.3, 0.5\}$ とし, 温度パラメータ $\tau = 0.5$, エポック数は 200 とする.

4.3 実験結果

各データセットを同種性を持つグラフと異種性を持つグラフに分けて, 従来手法と下線の提案手法の精度

を表 1, 2 に示す. これらの表は, 引用データセットである Cora, Citeseer, Pubmed の 3 つを同種性を持つグラフ, その他 6 つを異種性を持つグラフとして平均値で表し, 9 つ全てのデータセットの平均値も表している.

表 1 より, 特徴およびグラフノイズ 30% では異種性, 同種性を持つデータの両方に対して提案手法が 5.37% ~ 12.1% 高い精度を得た. また, 特徴およびグラフノイズを 30% から 50% へ増加させても異種性, 同種性を持つデータの両方に対して提案手法が 4.02% ~ 14.3% 高い精度を得た. この結果により, グラフの構造がより崩れた場合にも提案手法が有効であることがわかる. さらに, 特徴およびグラフノイズのようなグラフ構造に加えて, ラベル情報を失った場合でも異種性, 同種性を持つデータの両方に対して提案手法が特徴およびグラフノイズが 30% の場合は 4.31% ~ 14.4%, 特徴およびグラフノイズを 50% の場合は 3.17% ~ 14.4% 高い精度を得た. この結果により, ラベルを用いず類似データのペアを通じて有用なノード埋め込みを行う対照学習のノイズに対する有効性を示した.

表 2 より, 特徴のノイズを別のガウシアンノイズに変更しても提案手法の方がラベルノイズがない場合は 5.50% ~ 17.7%, ラベルノイズが 30% の場合は 3.38% ~ 15.05% 精度が高く, さまざまなタイプのノイズに適応できるモデルであることを示している. ガウシアン特徴ノイズはドロップアウト特徴ノイズと異なり, 各特徴次元ごとにノイズとなる値を加えるという特徴を持ち, この結果は連続的なノイズにも強いということを示す.

これらの実験結果から, 全てのノイズパターンにおいて, 提案手法である H2GTC と対照学習を用いた集団学習が既存手法より高い平均精度を得た. これは, H2GCN の 1/2-hop を区別して特徴を集約するという性質が 1-hop 近傍に異ラベルが混在している異種性グ

表 1: 特徴・グラフノイズ (30%, 50%) におけるラベルノイズ有無での精度比較

データ種別	ノイズ率	ラベルノイズ	H2GCN	GCN	NAG	GTC	H2GTC	集団学習
同種性グラフ	30%	無し	75.82	75.48	71.86	73.17	78.61	79.76
		30%	66.84	58.08	47.67	67.81	71.88	73.33
	50%	無し	68.91	70.34	69.06	72.86	77.81	78.63
		30%	63.42	55.70	50.11	66.59	69.43	71.31
異種性グラフ	30%	無し	46.89	41.84	39.65	46.39	53.55	53.78
		30%	40.48	37.81	36.32	40.68	45.10	45.01
	50%	無し	46.48	36.17	38.70	46.51	50.08	53.46
		30%	38.66	33.73	33.96	40.64	43.98	45.00
平均	30%	無し	56.53	53.05	50.39	55.31	61.90	62.44
		30%	49.26	44.57	40.03	49.72	54.03	54.45
	50%	無し	53.96	47.56	48.82	55.30	59.32	61.85
		30%	46.91	41.05	39.34	49.29	52.46	53.77

表 2: 特徴にガウシアンノイズとグラフノイズ (30%) におけるラベルノイズ有無での精度比較

データ種別	ラベルノイズ	H2GCN	GCN	NAG	GTC	H2GTC	集団学習
同種性グラフ	無し	55.58	53.83	50.79	72.18	76.39	77.63
	30%	44.94	42.43	40.41	63.56	65.10	69.35
異種性グラフ	無し	44.67	41.19	39.45	46.24	52.39	52.56
	30%	38.27	36.99	36.96	40.12	44.42	45.06
平均	無し	48.31	45.40	43.23	54.89	60.39	60.92
	30%	40.50	38.81	38.11	47.93	51.31	53.16

ラフに有効であり、同種性グラフに対しても有効と言える。また、すべてのノイズパターンで集団学習を用いた手法が精度が高い。これは、全ノイズ条件で単独モデルよりも優れた埋め込みを学習できる対照学習の結果を用いた確率平均の有効性を示す。

5 まとめ

本研究では、ノイズを含むグラフに対する安定かつ高精度なノード分類を目的として、局所構造に強い H2GCN と大域依存関係を捉える NAGphormer を 2 つのビューとして統合した対照学習枠組み、および対照学習モデル間の確率平均による集団学習を提案した。その結果、さまざまなノイズがグラフに含まれた状況下で同種性を持つグラフと異種性を持つグラフの両方に対して精度が向上した。これは、実環境に典型的な構造欠損・特徴劣化・ラベル誤りに対して一貫して有効な、データ特性に頑健な汎用的ノード表現学習基盤であることを示している。

今後は、ノイズのない環境や他種のノイズ条件においても提案手法を適用し、より幅広い条件下で有効に機能する手法となるよう検討を進める。また、本研究で扱った 9 種類に限らず、多様なデータセットへの適用を通じて、一層の汎用性と実用性の向上を目指す。

参考文献

- [1] Kipf, T. N., and Welling, M.: Semi-supervised classification with graph convolutional networks, *In 5th International Conference on Learning Representations (ICLR)* (2017)
- [2] Shu, J., Yan, Y., Zhao, L., Heimann, M., Akoglu, L., and Koutra, D.: Beyond homophily in graph neural networks: Current limitations and effective designs, *Advances in Neural Information Processing Systems*, Vol. 33, pp. 7793–7804 (2020)
- [3] Zhu, Y., Xu, Y., Yu, F., Liu, Q., Wu, S., and Wang, L.: Deep graph contrastive representation learning, arXiv:2006.04131 (2020)
- [4] Chen, J., Liu, C., Gao, K., Li, G., and He, K.: NAGphormer+: A Tokenized Graph Transformer With Neighborhood Augmentation for Node Classification in Large Graphs, *IEEE Transactions on Big Data* (2024)
- [5] Sun, Y., Zhu, D., Wang, Y., and Tian, Z.: GTC: GNN-Transformer co-contrastive learning for self-supervised heterogeneous graph representation, arXiv:2403.15520 (2024)

- [6] Namata, G. M., London, B., Getoor, L., and Huang, B.: Query-driven active surveying for collective classification, *In 10th International Workshop on Mining and Learning with Graphs*, (2012)
- [7] Pei, H., Wei, B., Chang, K. C.-C., Lei, Y., and Yang, B.: Geom-GCN: Geometric graph convolutional networks, *In 8th International Conference on Learning Representations (ICLR)*, (2020)
- [8] Sen, P., Namata, G., Bilgic, M., Getoor, L., Gallagher, B., and Eliassi-Rad, T.: Collective classification in network data, *AI Mag*, Vol. 29, No. 3, pp. 93–106 (2008)

A 付録

A.1 GCN

GCN は、グラフの畳み込みを利用した GNN のモデルの 1 つであり、以下の式で表される。

$$h'_v{}^{(k)} = \sum_{u \in N_1(v) \cup \{v\}} h_u^{(k-1)} (d_{v,1} + 1)^{-1/2} (d_{u,1} + 1)^{-1/2} \quad (14)$$

$$h_v^G = \sigma(h'_v{}^{(k)} W_k) \quad (15)$$

ここで、 $N_i(v)$ は隣接行列 A においてノード v から i ホップ離れた位置にある隣接ノードの集合であり、 $h_v^{(0)} = X_v$ かつ $d_{1,v} = |N_1(v)|$ である。 W_k は第 k 層の学習可能な重み行列で、 σ は活性化関数である。

A.2 H2GCN

H2GCN は、集約する特徴ベクトルを深さによって区別することで同種性と異種性の両方に適応できる GNN のモデルの 1 つであり、以下の式で表される。

$$h_v^{(0)} = \sigma(X_v W_e) \quad (16)$$

$$h_{v,i}^{(k)} = \sum_{u \in N_1(v)} h_u^{(k-1)} d_{i,v}^{-1/2} d_{i,u}^{-1/2} \quad (17)$$

$$h_v^{(k)} = \text{CONCAT}(h_{v,1}^{(k)}, h_{v,2}^{(k)}) \quad (18)$$

$$h_v^{\text{H2}} = \text{softmax}(\text{CONCAT}(h_v^0, h_v^1, \dots, h_v^K) W_c) \quad (19)$$

ここで、 $N_i(v)$ は隣接行列 A においてノード v から i ホップ離れた位置にある隣接ノードの集合であり、 $d_{1,v} = |N_1(v)|$ である。 W_e と W_c は学習可能な重み行列で、 σ は活性化関数である。

A.3 NAGphormer

NAGphormer は、各ノードについて異なるホップ距離から近傍情報を集約し、トークン化して Transformer で処理することで、情報量の多いノード表現を学習できるモデルであり、以下の式で表される。

$$x_v^k = \phi(N^k(v)) \quad (20)$$

$$h_v^{(0)} = [x_v^0 E; x_v^1 E; \dots; x_v^K E] \quad (21)$$

$$h_v^{(\ell)} = \text{MSA}(\text{LN}(h_v^{(\ell-1)})) + h_v^{(\ell-1)} \quad (22)$$

$$h_v^{(\ell)} = \text{FFN}(\text{LN}(h_v^{(\ell)})) + h_v^{(\ell)} \quad (23)$$

$$\alpha_{v,k} = \frac{\exp([Z_{v,0} \parallel Z_{v,k}] W_a^\top)}{\sum_{i=1}^K \exp([Z_{v,0} \parallel Z_{v,i}] W_a^\top)} \quad (24)$$

$$h_v^N = Z_{v,0} + \sum_{k=1}^K \alpha_{k,v} Z_{v,k} \quad (25)$$

ここで、 $N^k(v)$ は隣接行列 A においてノード v の k ホップ以内の近傍集合、 ϕ は集約演算子、 E, W_a は学習可能な重みである。 また、 $Z_{v,k}$ は $x_v^k E$ である。 MSA は Multi-Head Self-Attention, LN は Layer Normalization, FFN は Feed-Forward Network を表す。