Neural Sheaf Diffusionと交通予測

軽部友裕[1]矢野良輔[2]

[1]東京大学 [2]東京海上ディーアール(株)

グラフは様々なところに現れる。例えば、

1. 分子結合の様子
2. 画像から得られるデータ
3. SNSのフォロー関係
4. 交通道路などのモデル

上のデータに対して「頂点のクラス分け」や「グラフの辺の存在の推定」を行う。

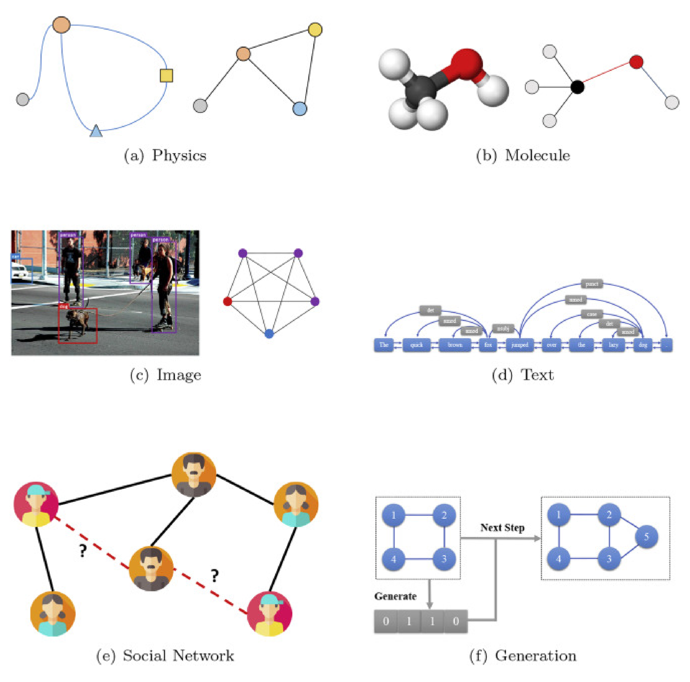


Figure: グラフの例[1]

グラフ畳み込みニューラルネットワーク

X をグラフGのデータとする。1層のグラフ畳み込みニューラルネットワーク(GCN)を以下で定義する。

$$F_\theta(X) = \sigma((I - \underbrace{D^{-\frac{1}{2}}AD^{-\frac{1}{2}}}_{\text{グラフの寄与}})XW_\theta)$$

ここで、σは活性化関数、W_θは学習される重みである。行列AとDはグラフGから定まる隣接行列と次数行列である。

GCNはグラフの情報を含んだ特徴量を取り出すために使われ、与えられたデータを扱いやすい形に変形する。

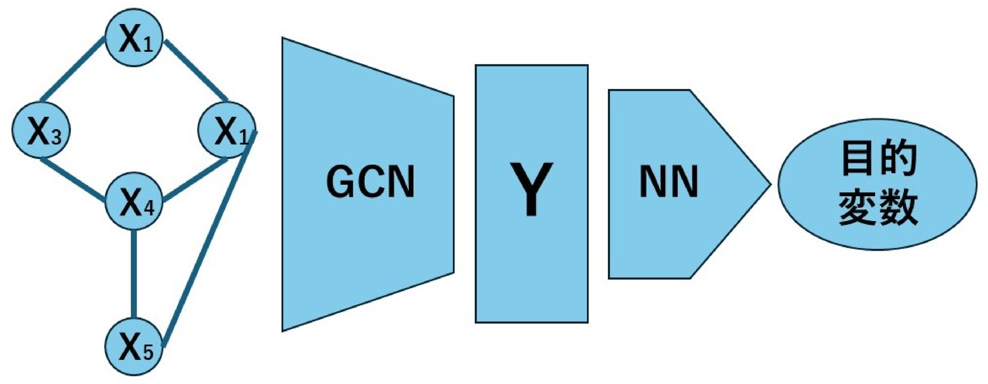


Figure: GCNの概念図

課題

GCN には、層を深くすると推定の精度が落ちる過平滑（oversmoothing)現象が起きる。

下図のように８層のGCNでは４層よりも精度が落ちる。

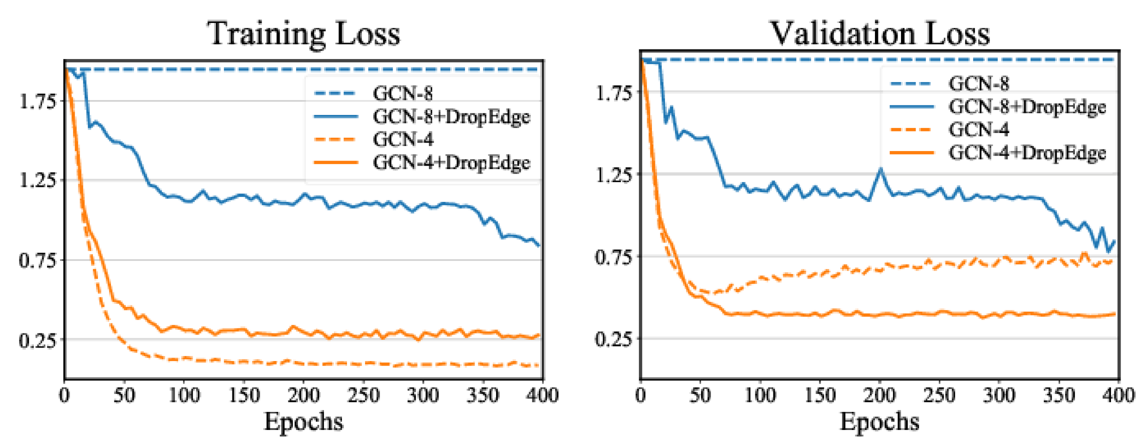


Figure: GCNのlossの様子[2]

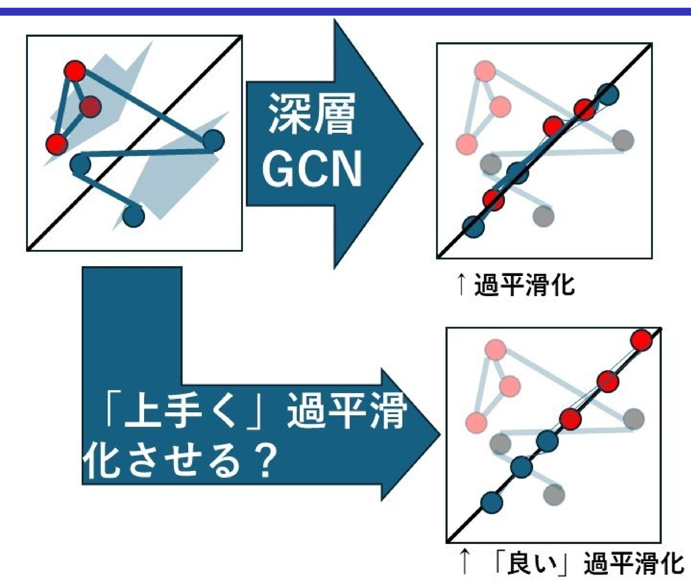
数学的な原因

連結グラフGの頂点の個数をNとし、G上のデータ$X \in \mathbb{R}^{N\times C}$を考える。このとき、GCN$F_\theta$と定数$\varepsilon < 1$、$N \times 1$次元の部分ベクトル空間$H \subset \mathbb{R}^{N\times C}$が存在し、以下が成り立つ。

$$d_H(F_\theta^r(X)) < \varepsilon^r d_H(X)$$

ここで$d_H(X)$は部分空間HとデータXとの距離を表す。

データに対してGCNを作用させると右の図のように指数関数的に1次元の部分ベクトル空間に近づく。この定理はGCNのグラフの寄与の行列$D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$の固有値がほとんど1未満であることが本質である。



課題

過平滑化はグラフの情報を特徴量に乗せるためには避けられないが過平滑化後のデータから情報を取り出すことはできないか？

シーフ(層)

グラフG上のシーフ$\mathcal{F}$とは以下のデータからなる。

1. 各頂点vにベクトル空間$\mathcal{F}(v) = \mathbb{R}^d$
2. 各辺eにベクトル空間$\mathcal{F}(e) = \mathbb{R}^d$
3. 各辺$e = (v, u)$に行列$\mathcal{F}_{ve} \colon \mathcal{F}(v) \to \mathcal{F}(e)$

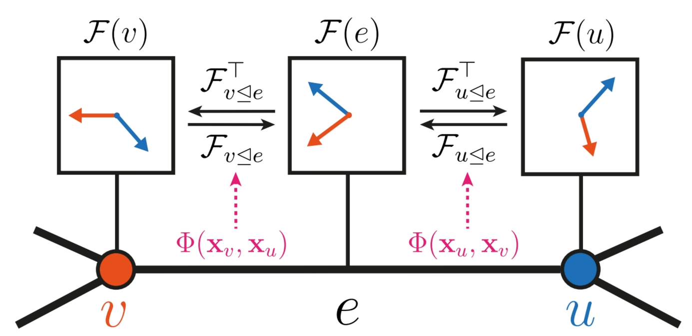


Figure: シーフの様子[3]

シーフは代数幾何などの幾何分野でも重要な概念である。幾何学において接ベクトル空間はシーフの例であり、グラフのシーフは図形の「骨組み」を取ったものに対応する。このとき、行列$\mathcal{F}_{ve}$は右図の接ベクトルを「制限」することに対応する。

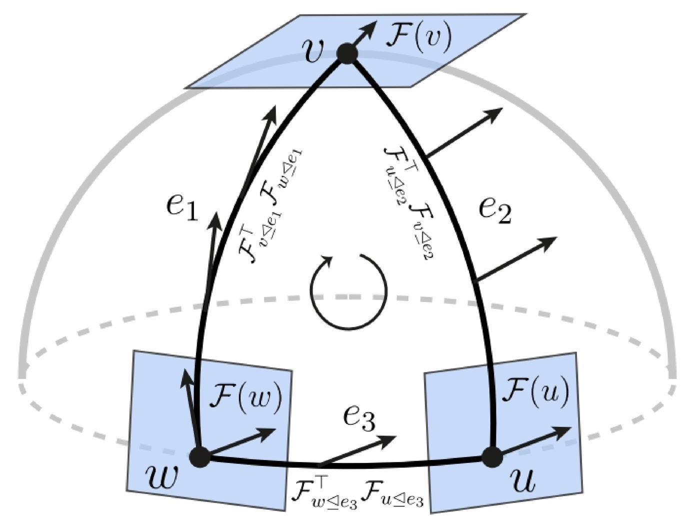


Figure: 図形とその骨組み[3]

シーフ拡散モデル（NSD）[3]

グラフGとシーフ$\mathcal{F}$に対して、シーフ拡散モデルが定まる。

$$F_\theta(X) = \sigma((I - \underbrace{\Delta_{\mathcal{F}_\theta}}_{\text{シーフの寄与}} XW_\theta)$$

アイデア

シーフはグラフの辺ごとに「重要度・寄与度」を与える。推論において重要な辺をうまく反映できているシーフを学習させる。

タスク

道路の複数地点における交通速度の時系列データが与えられたとき、将来の交通速度を予測する。

- in 約200地点における1時間の時系列データ
- out 約200地点における次の15分（45分）の時系列データ

結果

評価を比較すると以下のように先行研究より良い結果になった。RMSE(二乗平均平方根誤差)をlayer数ごとに比較すると過平滑化も抑制されてるように観察される。

Table: 15分予測の比較

モデル	MAE	MAPE(%)	RMSE
先行研究モデル [4]	2.25	5.26	4.04
NSD(ours)	2.08	4.75	3.45
NSD(4layers)			5.77
NSD(8layers)			5.74
NSD(16layers)			5.67

先行研究との比較・発見

- 層拡散モデルの「射影」としての意味づけを与えた。
- クラス分類に対して考察されていたが、回帰問題にも適用できることを発見した。
- 時系列データに対しても十分良い精度を出すことができると分かった。

参考文献

- Zhou, Jie et al. “Graph Neural Networks: A Review of Methods and Applications.” ArXiv abs/1812.08434 (2018): n. pag.
- Huang, Wenbing et al. "Tackling Over-Smoothing in Graph Convolutional Networks With EM-Based Joint Topology Optimization and Node Classification," in IEEE Transactions on Signal and Information Processing over Networks, vol. 9, pp. 123-139, 2023,
- Bodnar, Cristian, et al. "Neural sheaf diffusion: A topological perspective on heterophily and oversmoothing in gnns." Advances in Neural Information Processing Systems 35 (2022): 18527-18541.
- Yu, Bing et al "Spatio-temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting." Proceedings of the 27th International Joint Conference on Artificial Intelligence(2018).