GNN 谱图分析

刘闯
chuangliu@whu.edu.cn
2019.03.08

Reference

1. Deeper Insights into Graph Convolutional Networks for Semi-Supervised Learning, 2018, AAAI (P3 - P8)
2. Revisiting Graph Neural Networks: All We Have is Low-Pass Filters Node Classification, 2019 (P9 - P11)
3. Graph Neural Networks Exponentially Lose Expressive Power for Node Classification.

Laplacian Smoothing

Deeper Insights into Graph Convolutional Networks for Semi-Supervised Learning, 2018, AAAI

GCN vs FCN

FCN: $$H^{(l+1)} = \sigma\left(H^{(l)}\Theta^{(l)}\right)$$

GCN: $$H^{(l+1)} = \sigma\left(\boxed{\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}}H^{(l)}\Theta^{(l)}\right)$$

One-layer GCN

1. 通过图卷积生成特征 Y

$$Y = \tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}X$$

2. 将特征矩阵 Y 输入到全连接网络

Laplacian Smoothing

- The Laplacian smoothing on each channel of the input features is defined as:

$$\hat{\mathbf{y}}_i = (1-\gamma)\mathbf{x}_i + \gamma \sum_j \frac{\tilde{a}_{ij}}{d_i}\mathbf{x}_j \quad (\textbf{ for } 1 \le i \le n)$$

γ 控制自身特征和邻居特征之间权重

- Matrix form :

$$\hat{Y} = X - \gamma \tilde{D}^{-1}\tilde{L}X = \left(I - \gamma \tilde{D}^{-1}\tilde{L}\right)X$$

- $\gamma = 1$ ： 只利用邻居节点特征

$$\hat{Y} = \tilde{D}^{-1}\tilde{A}X$$

- 归一化改为对称归一化

$$\hat{Y} = \tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}X$$

$\boxed{\text{实际就是卷积生成特征}}$

1. A signal processing approach to fair surface design. Taubin, G. 1995.

Laplacian Smoothing

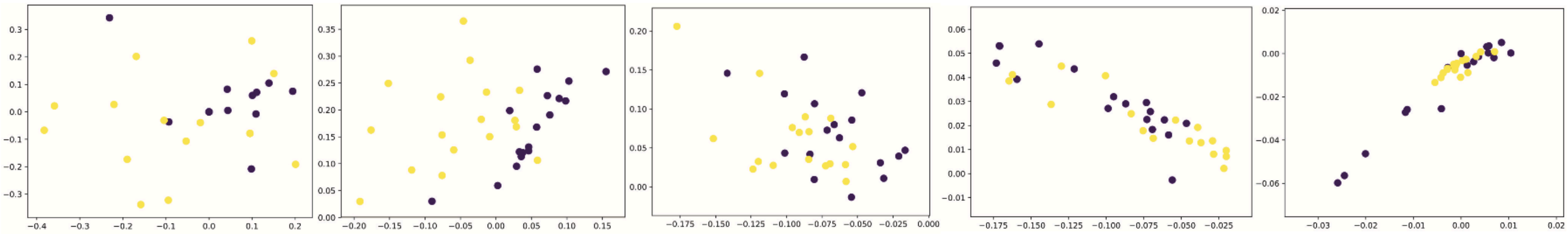
Graph convolution a special form of Laplacian smoothing

1. Laplacian smoothing 计算邻居节点的平均特征作为新特征
2. 由于同一 cluster 中的顶点往往紧密相连，平滑使得它们的特征相似，这使得后续的分类任务更加容易。

How many convolutional layers

层数不是越多越好：

1. 多层难于训练
2. 层数太多， Laplacian smoothing 将不同 cluster 的顶点特征趋于相同



(a) 1-layer (b) 2-layer (c) 3-layer (d) 4-layer (e) 5-layer

Figure 2: Vertex embeddings of Zachary's karate club network with GCNs with 1,2,3,4,5 layers.

2-layer 的时候分的较好，之后开始混合在一起

How many convolutional layers

By repeatedly applying Laplacian smoothing many times, the features of vertices within each connected component of the graph will **converge** *to the same value*

Suppose that a graph $\mathcal{G}$ has k connected components $\{C_i\}_{i=1}^k$, and the indication vector for the i-th component is denoted by $\mathbf{1}^{(i)} \in \mathbb{R}^n$. This vector indicates whether a vertex is in the component C_i, i.e.,

$$\mathbf{1}_j^{(i)} = \begin{cases} 1, v_j \in C_i \\ 0, v_j \notin C_i \end{cases} \tag{11}$$

Theorem 1. *If a graph has no bipartite components, then for any* $\mathbf{w} \in \mathbb{R}^n$, *and* $\alpha \in (0, 1]$,

$$\lim_{m \to +\infty} (I - \alpha L_{rw})^m \mathbf{w} = [\mathbf{1}^{(1)}, \mathbf{1}^{(2)}, \ldots, \mathbf{1}^{(k)}]\theta_1,$$

$$\lim_{m \to +\infty} (I - \alpha L_{sym})^m \mathbf{w} = D^{-\frac{1}{2}}[\mathbf{1}^{(1)}, \mathbf{1}^{(2)}, \ldots, \mathbf{1}^{(k)}]\theta_2,$$

where $\theta_1 \in \mathbb{R}^k, \theta_2 \in \mathbb{R}^k$, *i.e., they converge to a linear combination of* $\{\mathbf{1}^{(i)}\}_{i=1}^k$ *and* $\{D^{-\frac{1}{2}}\mathbf{1}^{(i)}\}_{i=1}^k$ *respectively.*

Solutions

Graph convolution is a localized filter

- ☐ 层数少了，在 label 数据少的时候，难以将 Label 传播出去

☑ Co-Train a GCN with a Random Walk Model

Random walk 可以捕捉网络的全局特征

1. 先 rw 每一类有标签节点最接近的一些节点，
2. 加入训练集，进行训练

Algorithm 1 Expand the Label Set via ParWalks

1: $P := (L + \alpha\Lambda)^{-1}$
2: **for** each class k **do**
3: $\quad \boldsymbol{p} := \sum_{j \in \mathcal{S}_k} P_{:,j}$
4: $\quad$ Find the top t vertices in $\boldsymbol{p}$
5: $\quad$ Add them to the training set with label k
6: **end for**

☑ GCN Self-Training

更好的利用训练样本

1. 训练 GCN， 得到结果
2. 将最可信的结果看作 label 数据
3. 重复训练

Algorithm 2 Expand the Label Set via Self-Training

1: $\boldsymbol{Z} := GCN(X) \in \mathbb{R}^{n \times F}$, the output of GCN
2: **for** each class k **do**
3: $\quad$ Find the top t vertices in $Z_{i,k}$
4: $\quad$ Add them to the training set with label k
5: **end for**

☑ Union + Intersection

GNN: Low-Pass Filter

Revisiting Graph Neural Networks: All We Have is Low-Pass Filters

Our results indicate that graph neural networks only perform low-pass filtering on feature vectors and do not have the **non-linear manifold** learning property. (SGC)

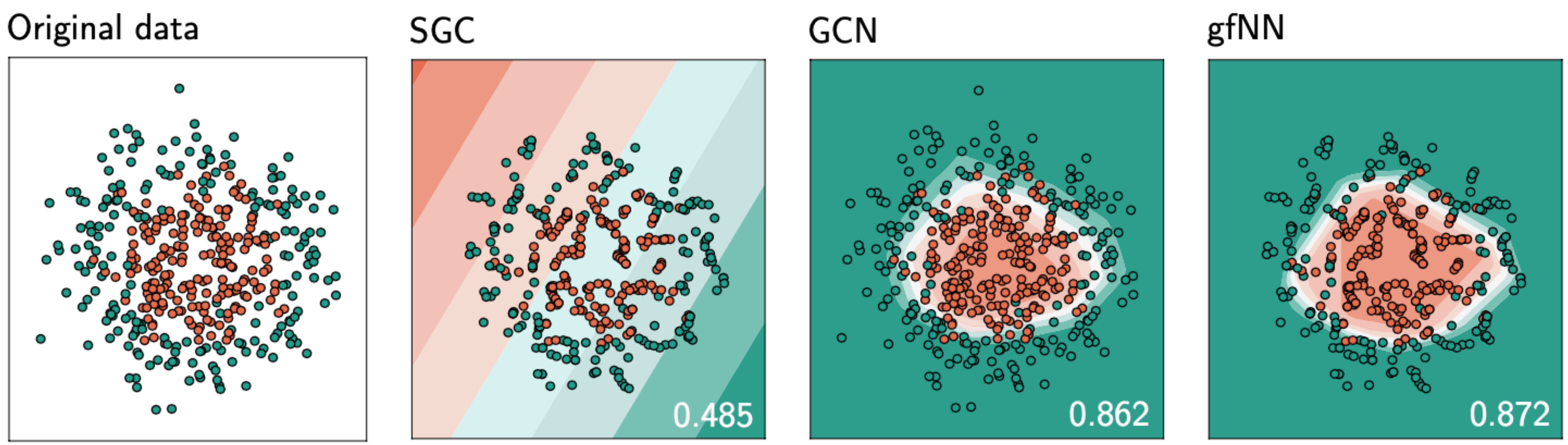


Figure 5: Decision boundaries on 500 generated data samples following the two circles pattern

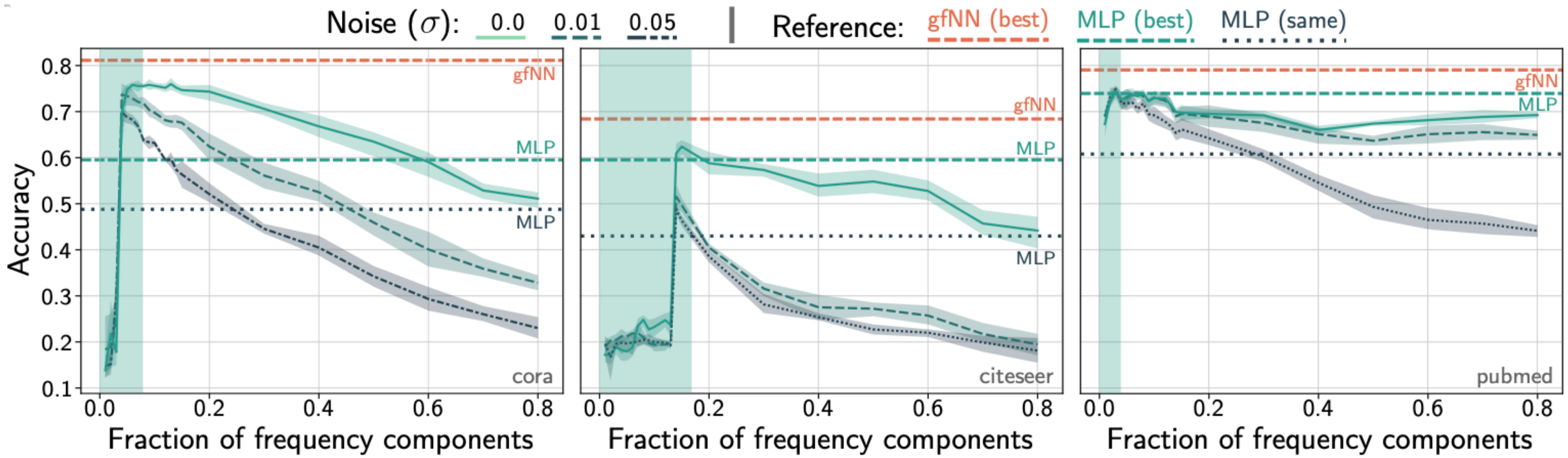
从 graph signal processing 角度来分析 GNN (挑选了 GCN(kipf) 和 SGC 两个简单的模型)

GSP

GSP 将节点上的数据看作信号， 应用信号处理技术理解信号的特性。

在一个标准的信号处理问题中，通常假设观测结果包含一些噪声，并且底层的“真实信号”是低频的。

Assumption 1. *Input features consist of low-frequency true features and noise. The true features have sufficient information for the machine learning task.*



Compute the first k-frequency component: $\hat{\mathcal{X}}_k = U[:k]^\top \tilde{D}^{1/2}\mathcal{X}$

1. 只有少量的频率有用
2. 加入噪声之后预测效果变差，但是低频部分鲁棒性较好

Convolution

将图信号与传播矩阵相乘对应于低通滤波，表明图卷积层只是低通滤波 (low-pass filtering)。因此，不需要学习图卷积层的参数。

Theorem 2 (Informal, see Theorem 7, 8)**.** *Under Assumption 1, the outcomes of SGC, GCN, and gfNN are similar to those of the corresponding NNs using true features.*

剔除噪声之后， 表现相同