

# GCN Notes

## 0 Abstract

任务：对图数据结构的半监督学习

方法：直接作用于图的CNN结构，学习hidden layer representations，能encode局部图结构和nodes的features

思路：来自于图谱卷积的局部一阶近似

## 1 Introduction

- 针对问题：对图上结点进行分类，只有nodes的一个小子集有labels
  - 这个问题能转化为基于图的半监督学习
    - label information通过某种形式的显式基于图的regularization在图上进行平滑(用graph Laplacian regularization term)
$$L = L_0 + \lambda L_{reg}, L_{reg} = \sum_{i,j} A_{ij} \|f(X_i) - f(X_j)\|^2 = f(X)^T \Delta f(X)$$
$$L_0$$
是supervised loss(基于graph上的有label的nodes),  $f(\cdot)$ 可以为NN-like的可微函数,  $\lambda$ 是权重参数,  $X_i$ 是node i的**feature vector**,  $X$ 是 $X_i$ 组成的矩阵, 设 $G = (V, E)$ 有N个nodes  $v_i \in V, (v_i, v_j) \in E, A \in R^{N \times N}$ 为邻接矩阵(binary or weighted),  $D$ 为degree矩阵,  $D_{ii} = \sum_j A_{ij}$
  - GCN方法  
直接用NN model  $f(X, A)$ 来encode图结构, 并且在supervised target  $L_0$ 下训练来避免基于图的显式正则化  
将 $f(\cdot)$ 直接作用于图的邻接矩阵能够让model分配从supervised loss得到的梯度信息, 能够学习nodes的表示(无论有没有label)
- Contribution
  - a simple and well-behaved layer-wise propagation rule for NN,能够直接对图操作并展现如何由图谱卷积的一阶近似推动的
  - 展示了给出的graph-based NN model是如何用于快速、可扩展的graph nodes的半监督分类的

## 2 Fast Approximate Convolutions on Graphs

本节给出graph-based NN model的理论动机,考虑一个多层GCN, 具有以下的分层传播规则

- layer-wise propagation rule(分层传播规则):  $H^{(l+1)} = \sigma(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)})$ 
  - $\tilde{A} = A + I_N$ 是无向图 $G$ 加上自连接后的邻接矩阵
  - $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}, W^{(l)}$ 是layer-specific trainable weight matrix
  - $\sigma(\cdot)$ 是activation function,  $H^{(l)}$ 是第 $l$ 层 activation function的输出,  $H^{(0)} = X$   
这个propagation rule可以用局部图的一阶近似来推动

### 2.1 Spectral Graph Convolutions

- graphs上的spectral convolutions: filter  $g_\theta = \text{diag}(\theta), \theta \in R^N$ 和signal  $x \in R^N$  (a scalar for every node)在傅里叶域的乘法
  - $g_\theta \star x = U g_\theta U^T x$ ,  $U$ 为normalized graph Laplacian算子
$$L = I_N - D^{-\frac{1}{2}} A D^{-\frac{1}{2}} = U \Lambda U^T$$
的特征向量矩阵,  $\Lambda$ 为特征值矩阵,  $U^T x$ 为 $x$ 的图傅里叶变换, 这里可以将 $g_\theta$ 理解为一个关于 $L$ 的特征值的函数, 即 $g_\theta(\Lambda)$

- 这个运算代价很高,  $U$ 的乘法是 $O(N^2)$ , 计算 $L$ 的特征分解对large graphs来说运算代价也很高
- 为了降低计算复杂度,  $g_\theta(\Lambda)$ 可以用切比雪夫多项式 $T_k(x)$ 的 $K^{th}$  order截断表达式来很好的估计:  $g_{\theta'}(\Lambda) \approx \sum_{k=0}^K \theta'_k T_k(\tilde{\Lambda}), \tilde{\Lambda} = \frac{2}{\lambda_{max}} \Lambda - I_N, \lambda_{max}$ 是 $L$ 最大的特征值,  $\theta' \in R^k$ 是切比雪夫参数向量
- 切比雪夫多项式:  $T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x), T_0(x) = 1, T_1(x) = x$
- 最后,  $g_{\theta'} \star x = \sum_{k=0}^K \theta'_k T_k(\tilde{L})x, \tilde{L} = \frac{2}{\lambda_{max}} L - I_N$ , 目前是 $K$ 步, 仅考虑距离central node  $K$ 步的nodes, 那么复杂度就是 $O(|E|)$

## 2.2 Layer-Wise Linear Model

根据 $g_{\theta'} \star x = \sum_{k=0}^K \theta'_k T_k(\tilde{L})x$ , 可以堆叠建立多层convolutional layers

设 $K = 1$ , 那么关于 $L$ 是线性的, 在图拉普拉斯谱上有线性函数

通过这种方式, 仍能通过堆叠多个这样的层来回复丰富的convolutional filter functions, 不受限于切比雪夫多项式给出的显式参数

在GCN的线性公式中, 进一步近似 $\lambda_{max} = 2$ , 然后(5)就简化到

$$g_{\theta'} \star x \approx \theta'_0 x + \theta'_1 (L - I_N)x = \theta'_0 x - \theta'_1 D^{-\frac{1}{2}} A D^{-\frac{1}{2}} x$$

$\theta'_0, \theta'_1$ 为参数, 可以被整张图共享

连续应用这种filter可以有效卷积一个node的 $k$ 阶邻居,  $k$ 为model中连续filter操作或卷积层的数目

实际上可以简化参数数量来防止overfitting并减小每层的操作数, 即

$$g_{\theta} \star x \approx \theta (I_N + D^{-\frac{1}{2}} A D^{-\frac{1}{2}}) x, \theta = \theta'_0 = -\theta'_1$$

注意到 $I_N + D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$ 现在有 $[0, 2]$ 的特征值, 这会导致数值不稳定/梯度爆炸, 因此需要归一化:

$$I_N + D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \rightarrow \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}}, \tilde{A} = A + I_N, \tilde{D}_i = \sum_j \tilde{A}_{ij}$$

- 推广: 特征映射公式

考虑具有 $C$ 个input channel(每个结点 $C$ 维特征向量)的信号 $X \in R^{N \times C}$ , 和 $F$ 个filters, 则feature maps(特征映射)为 $Z = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} X \Theta$

$\Theta \in R^{C \times F}$ 是filter参数矩阵,  $Z \in R^{N \times F}$ 为卷积信号矩阵, filtering operation复杂度为 $O(|E|FC)$

## 3 Semi-Supervised Node Classification

现在有了model  $f(X, A)$ , 可以在图上有效的传播信息

通过调整model  $f(X, A)$ , 可以放松通常在半监督学习中做的假设, 希望邻接矩阵 $A$ 包含的信息( $X$ 没有表达出来的)可以在这种情况下提供更多帮助

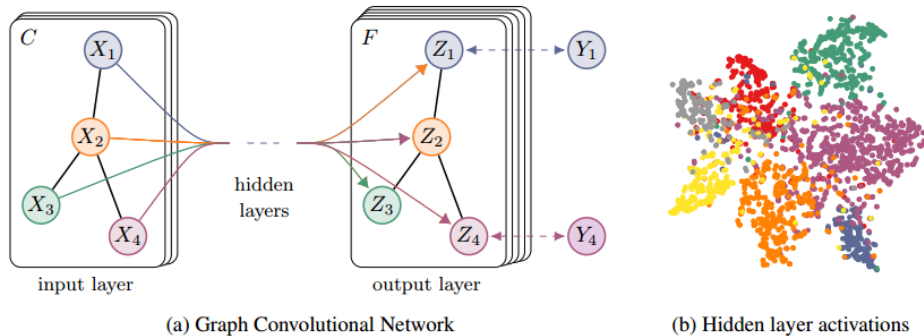


Figure 1: *Left*: Schematic depiction of multi-layer Graph Convolutional Network (GCN) for semi-supervised learning with  $C$  input channels and  $F$  feature maps in the output layer. The graph structure (edges shown as black lines) is shared over layers, labels are denoted by  $Y_i$ . *Right*: t-SNE (Maaten & Hinton, [2008]) visualization of hidden layer activations of a two-layer GCN trained on the Cora dataset (Sen et al., [2008]) using 5% of labels. Colors denote document class. [n.ne/q4\\_41727666](https://n.ne/q4_41727666)

左边是C个输入，中间若干隐藏层，输出层有F个特征映射，标签为 $Y_i$

### 3.1 例子

考虑一个用于semi-supervised node classification on a graph的两层GCN，邻接矩阵 $A$ 是对称的

- 预处理：计算 $\tilde{A} = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}}$
- forward model:  $Z = f(X, A) = \text{softmax}(\tilde{A} \text{ReLU}(\tilde{A} X W^{(0)})) W^{(1)}$ 
  - $W^{(0)} \in R^{C \times H}$ 为输入层-隐藏层的权重矩阵，隐藏层有H个特征映射
  - $W^{(1)} \in R^{H \times F}$ 为隐藏层-输出层的前中矩阵
  - softmax函数作用在每一行上,  $\text{softmax}(x_i) = \frac{\exp(x_i)}{\sum_i \exp(x_i)}$
- cross-entropy Loss:  $L = - \sum_{l \in Y_L} \sum_{f=1}^F Y_{lf} \ln Z_{lf}$ ,  $Y_L$ 为带标签的结点的集合
- 训练  
网络中权重 $W^{(0)}, W^{(1)}$ 通过SGD训练，矩阵A稀疏表示，通过Dropout引入训练过程中随机性。