

# 实验报告

## 主要内容:

作者为神经网络模型引入一个简单而且表现良好的层级传播规则，该规则直接在图上运行，利用了谱域图卷积的局部一阶近似。使用了这种传播规则的 GCN 模型在图的边数上线性缩放。通过图结构中部分有标签的节点数据对卷积神经网络模型进行训练，使得网络可以对其余无标签数据进一步分类。

## 前人工作及存在的问题

对于图中节点的半监督分类问题，(Zhu et al., 2003; Zhou et al., 2004; Belkin et al., 2006; Weston et al., 2012)等人在处理该问题时，将标签信息通过显示的基于图正则化而平滑化，例如，通过在损失函数中使用图的拉普拉斯正则项，如下式 (1) 所示

$$\mathcal{L} = \mathcal{L}_0 + \lambda \mathcal{L}_{\text{reg}}, \quad \text{with} \quad \mathcal{L}_{\text{reg}} = \sum_{i,j} A_{ij} \|f(X_i) - f(X_j)\|^2 = f(X)^\top \Delta f(X). \quad (1)$$

式(1)依赖于图中相连节点可能共享相同标签的假设。然而，这种假设可能会限制建模能力，因为图的边不一定需要编码节点相似性，而可能包含其他信息。

## 改进之处

在本文中，直接使用神经网络模型 $f(X, A)$ 对图结构进行编码，并用有监督损失 $L_0$ 对所有带标签的节点进行有监督训练，从而避免在损失函数中进行基于显示的图的正则化。

在图的邻接矩阵上调节 $f(\cdot)$  将允许模型从监督损失中分配梯度信息，并使其能够学习带标签用和不带标签的节点表示。

## 图的快速近似卷积

$$H^{(l+1)} = \sigma \left( \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)} \right). \quad (2)$$

作者设计了具有式 (2) 形式的层级传播规则的多层图卷积网络 (GCN)

## 公式推导

谱域卷积:

$$g_{\theta} \star x = U g_{\theta} U^{\top} x, \quad (3)$$

式 (2) 计算上很昂贵, 如与特征向量矩阵  $U$  相乘, 复杂度为  $O(N^2)$ 。2011 年

Hammond 等人提出  $g_{\theta}$  可以通过一个切比雪夫多项式  $T_k(x)$  的截断展开为第  $K$  阶:

$$g_{\theta'}(\Lambda) \approx \sum_{k=0}^K \theta'_k T_k(\tilde{\Lambda}), \quad (4)$$

其中, 切比雪夫多项式为:

$$\begin{aligned} T_k(x) &= 2xT_{k-1}(x) - T_{k-2}(x) \\ T_0(x) &= 1 \\ T_1(x) &= x \end{aligned}$$

因此, 卷积转化为:

$$g_{\theta'} \star x \approx \sum_{k=0}^K \theta'_k T_k(\tilde{L}) x, \quad (5)$$

式 (5) 的计算复杂度为  $O(|\varepsilon|)$ 。当  $k=1$  时, 此函数就是关于  $L$  的线性函数。可以通过

堆叠多个如式 (5) 的卷积层来建立一个 GCN 模型。对于中心节点, 进行一层卷积

时, 只利用了距离中心节点最近的邻居节点, 那么对第一层卷积结果重复这样的卷积

操作，可以扩展到第二层邻居节点。所以跟以前不同的是，不是一次性进行 k 阶，而是累加 k 次的结果。这样可以避免复杂图结构中的过拟合问题。

进一步近似  $\lambda_{\max} \approx 2$ ，则有

$$g_{\theta'} \star x \approx \theta'_0 x + \theta'_1 (L - I_N) x = \theta'_0 x - \theta'_1 D^{-\frac{1}{2}} A D^{-\frac{1}{2}} x, \quad (6)$$

简化参数  $\theta = \theta'_0 = -\theta'_1$ .

$$g_{\theta} \star x \approx \theta \left( I_N + D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \right) x, \quad (7)$$

使用再规范化技巧：

$$Z = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} X \Theta, \quad (8)$$

其中：

$$I_N + D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \rightarrow \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} \\ \tilde{A} = A + I_N \text{ and } \tilde{D}_{ii} = \sum_j \tilde{A}_{ij}.$$

## 模型限制

- 内存需求：模型进行梯度下降优化时使用的整个数据集。
- 定向的边和边特征：框架没有利用边的特征，而且限制图为无向图。
- 假设限制：文中隐式地假设局部性（依赖于 k 阶邻居）以及假定自连接的边和与相邻节点的边的同等重要。

## 复现论文

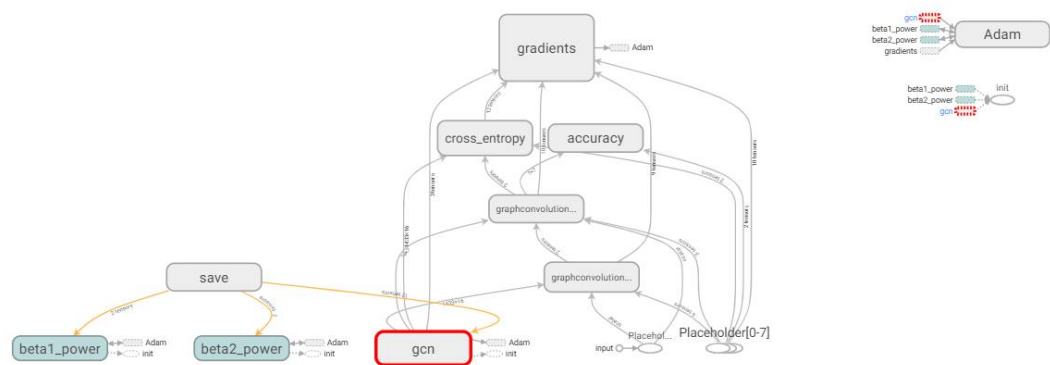
代码在 gcn 文件夹中，在 cora 数据集上的运行结果如下图所示，cora 是文献引用网络，共六种类别。图中 loss 表示交叉熵损失，acc 表示准确率，模型的训练只使用了引用网络中 150 个节点的标签，比例是 0.052，可以看出模型的半监督学习效果良好：

```

Epoch: 0190 train_loss= 0.5095 train_acc= 0.9500 val_loss= 0.8398 val_acc= 0.7967 time= 0.0358
Epoch: 0191 train_loss= 0.5064 train_acc= 0.9500 val_loss= 0.8375 val_acc= 0.7933 time= 0.0379
Epoch: 0192 train_loss= 0.5029 train_acc= 0.9429 val_loss= 0.8364 val_acc= 0.7867 time= 0.0389
Epoch: 0193 train_loss= 0.4998 train_acc= 0.9500 val_loss= 0.8363 val_acc= 0.7933 time= 0.0379
Epoch: 0194 train_loss= 0.4971 train_acc= 0.9500 val_loss= 0.8362 val_acc= 0.8000 time= 0.0370
Epoch: 0195 train_loss= 0.4946 train_acc= 0.9500 val_loss= 0.8364 val_acc= 0.7967 time= 0.0368
Epoch: 0196 train_loss= 0.4926 train_acc= 0.9500 val_loss= 0.8368 val_acc= 0.7900 time= 0.0369
Epoch: 0197 train_loss= 0.4904 train_acc= 0.9500 val_loss= 0.8356 val_acc= 0.7933 time= 0.0389
Epoch: 0198 train_loss= 0.4882 train_acc= 0.9500 val_loss= 0.8339 val_acc= 0.7867 time= 0.0379
Epoch: 0199 train_loss= 0.4861 train_acc= 0.9500 val_loss= 0.8308 val_acc= 0.7900 time= 0.0379
Epoch: 0200 train_loss= 0.4845 train_acc= 0.9500 val_loss= 0.8281 val_acc= 0.7900 time= 0.0369
Test set results: loss= 0.8847 accuracy= 0.7900

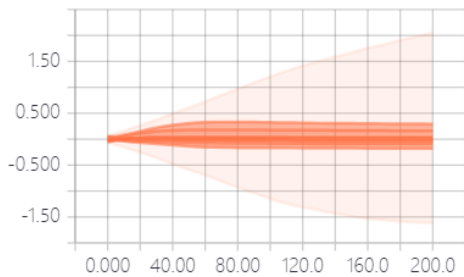
```

模型的结构使用 tensorboard 展示如下，使用两个 gcn 层：

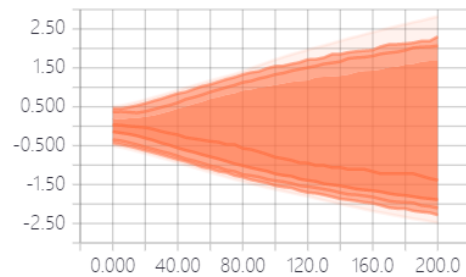


训练过程中各层权重的变化如下图：

gcn/graphconvolution\_1/vars/weights\_0

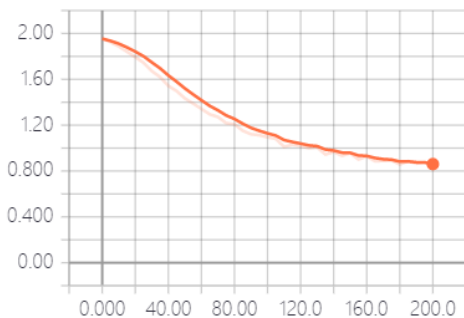


gcn/graphconvolution\_2/vars/weights\_0

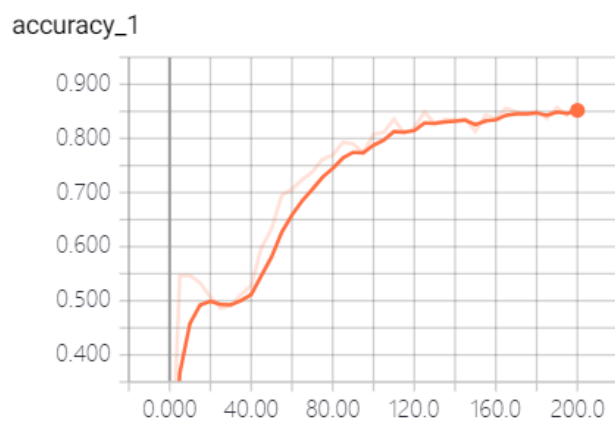


模型有两个评估指标，作为损失函数的交叉熵随着迭代的变化如下图：

cross\_entropy\_1

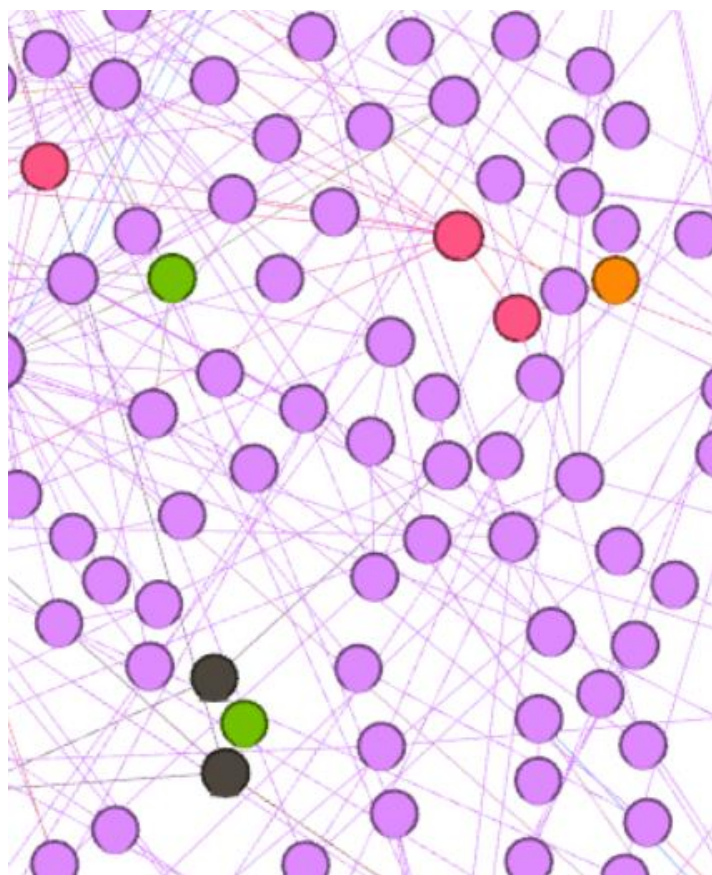


还有可以更加直观地感受分类结果好坏的准确率，如下图：



## 可视化模型

首先可视化问题，如下图所示，紫色节点表示训练时不适用其标签信息，其他颜色代表在训练中使用标签信息的不同类别的节点，最后训练好的网络要能够对图中的所有节点进行准确分类。

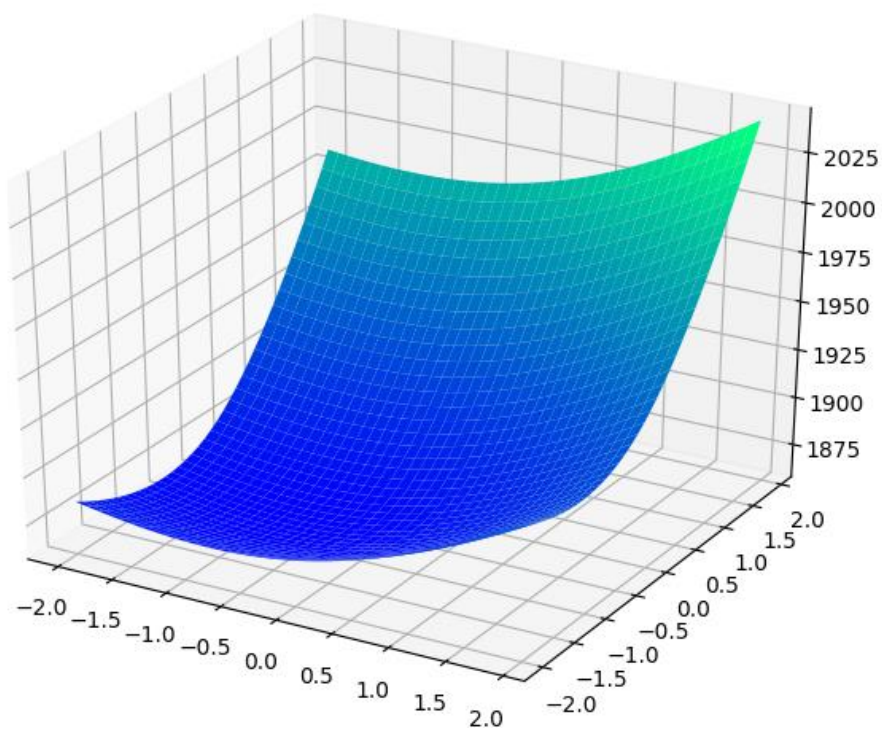


将模型中间隐藏层的输出使用 t-SNE 降到 3d，可视化展示如下图，可能是 t-SNE 的参

数没有调好，模型有部分区域并不能很好的分类：



通过简单的可视化模型的交叉熵，可以看出该模型是凸的，所以可以使用梯度下降的方法进行优化：



## 总结:

这篇文章介绍了一种基于图结构数据的半监督分类方法。GCN 模型使用了一个有效的层级传播规则，该规则基于图上谱卷积的一阶近似，有效的利用了一个节点的领域信息。实验结果表明，该 GCN 模型能够同时编码图结构和节点特征，对半监督分类有一定帮助。