精读论文

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

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

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的表示。

图的快速近似卷积

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

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:

$$g_\theta \star x = U g_\theta U^\top x\,, \tag{3}$$

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_θ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:

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

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:

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

式（5）的计算复杂度为O(|ε|)。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

$$g_{\theta'} \star x \approx \theta_0' x + \theta_1' \left(L - I_N\right) x = \theta_0' x - \theta_1' D^{-\frac{1}{2}} A D^{-\frac{1}{2}} x \,, \tag{6}$$

简化参数 $\theta = \theta_0' = -\theta_1'$.

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

使用再规范化技巧：

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

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时整个数据集。
- 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。