GCN的原始论文，发表于2017年的ICLR会议

解决的问题

如何将神经网络应用在图结构数据上？

问题描述

给定以下输入：

1. 图中顶点的特征矩阵$H \in \mathbb{R}^{n \times F}$，其中$n$为顶点数量，$F$为特征数量
2. 图的结构信息，如邻接矩阵A

输出：

1. 图中顶点新的的特征表示$H' \in \mathbb{R}^{n \times F'}$，即

$$H' = \mathrm{GCN(H)} = g(AHW^T + b)$$

如果套用神经网络模型，每一层可以用一个非线性函数进行表示：

$$H^{(l+1)} = f(H^{(l)}, A)$$

其中$H^{(0)} = X, H^{(L)} = Z$，问题在于如何选取函数$f(.,.)$

做法及创新

对于函数$f(.,.)$的选取，论文中提出了一种可能的函数形式：

$$f(H^{(l)}, A) = \sigma(\hat{D}^{-\frac{1}{2}} \hat{A} \hat{D}^{-\frac{1}{2}} H^{(l)} W^{(l)})$$

其中$\hat{A} = A + I$，因为与矩阵A相乘表示对于每个顶点，我们对除了自身外所有邻居顶点的特征向量进行求和，因此加上单位矩阵是为了引入自环。而正则化是避免与矩阵A相乘改变特征向量的规模。实际在论文中只使用两层网络就达到了很好的效果，表示为：

$$Z_{\mathrm{GCN}} = \mathrm{softmax}\big(\hat{A}' \mathrm{ReLU}\big(\hat{A}' X W_0\big) W_1\big)$$

其中W_0、W_1为这两层网络的参数，$\hat{A}' = \hat{D}^{-\frac{1}{2}} \hat{A} \hat{D}^{-\frac{1}{2}}$，$Z \in R\mathbb{R}n \times c$为预测的顶点标签，$c$为类别数目，毕竟论文解决的就是一个分类问题。

更一般地，使用邻域信息的图神经网络形式可以概括为：

$$h_v^{(l)} = \sigma\Bigg(W_l \cdot \mathrm{AGGREGATE}\Bigg(\{h_u^{(l-1)}, \forall u \in N(v)\}\Bigg)\Bigg)$$

其中W_l是第l层网络的权重矩阵，$\mathrm{AGGREGATE}$是与特定模型相关的聚合函数，$h_v^{(l)}$是顶点v在第l层的隐层特征表示。论文中只是用了一个两层网络就达到了很好的效果。

将论文所提出的函数改写为上述形式，即为：

$$h_v^{(l)} = \mathrm{ReLU}\Bigg(W_l \cdot \sum_{u \in N(v)} (deg(v)deg(u))^{-1/2} h_u^{(l-1)}\Bigg)$$

其中$deg(u)$为顶点u的度。

AS-GCN中对这篇论文的模型形式描述如下：

$$h_{v_i}^{(l)} = \sigma\Big(W_l \cdot \sum_{j=1}^{N} a(v_i, u_j) \cdot h_{u_j}^{(l-1)}\Big), i = 1, \ldots, N$$

这里$A = (a(v_i, u_j)) \in \mathbb{R}^{N \times N}$对应前面一种写法的正则化邻接矩阵$\hat{A}'$，表面上看对于顶点$v_i$，需要考虑将图中剩下的所有顶点的上一时刻的隐层表示做加权和，来作为它当前时刻的隐层表示，因为j的取值范围为$[1, N]$，N就是图中顶点的数量。但实际上，大多数顶点因为与v_i并无边相连，所以邻接矩阵中对应的值为0，意味着在加权和中的权重为0，相当于加权和时只会考虑有边相连的顶点，这同样是考虑邻域，只不过跟上面那种写法不同。