This is a note for the paper: Thomas N. Kipf, Max Welling, Semi-Supervised Classification with Graph Convolutional Networks (ICLR 2017)

paper and code have been published by the author

问题背景

问题的抽象描述：有一张无向图G（V，E），节点数N，边数E，每个节点各有一个特征向量$x \in R^d$，x表示了某种属性，比如一个词的嵌入向量。现在已有图上一些节点的标签，希望对图上剩余的节点进行分类。

问题的一个实例：文献分类。每个节点代表一篇文献，如果两篇文献有引用关系（不管谁引用谁），则两个节点连一条边。现在已经对一部分文献进行了类别标注（每个类别都标注了一些），希望知道剩余文献的类别。每个节点的特征向量为文献的词向量，维度为词典大小，某一维为1则这个词在此文献中出现过。

图上的前向传播

一次前向传播（作者定义为卷积）：

1、将特征向量经过一个线性变换$X \to XW$，线性变换的参数是待定系数$\in R^{d \times d_1}$

2、对于一个节点，将其周围节点（包括自己）的特征向量加权求和，权重是拉普拉斯矩阵的元素（归一化的邻接矩阵），具体表达式为$\frac{1}{\sqrt{d_i d_j}}$，分母为当前节点和邻居节点的度的几何平均

3、将每个节点的特征向量通过激活函数，比如relu

一次前向传播后，每个节点特征向量的维度改变d_1 。若干次变换后，维度变为K ,K 为分类的类别数量。然后将有标签的节点加入损失函数（比如cross-entropy），然后反向传播更新参数。参数用均匀分布进行初始化。

从图傅里叶变换的角度理解

x为图上的一个信号，为N维向量（N为节点个数） $g * x \approx \sum_{i=1}^{K} \theta_i T_i x$ $*$表示卷积

这个式子是切比雪夫多项式近似的结果（待之后理解和更新）

采用一阶近似，且令$\theta_0 = -\theta_1 = \theta$时 $g * x = \theta \times L \times x$, L为拉普拉斯矩阵

$X \in R^{N \times d}$看作d路信号 一个滤波器将d路信号变为一路信号，共F组滤波器 故输出为$LX\Theta$ 其中$Theta \in R^{d \times F}$ 其中每一列为一组滤波器

可以看出，这近似得太过分了！ 感觉是强行近似。 换一种角度理解，就是将周围节点的特征向量加权求和，权重为$\frac{1}{\sqrt{d_i d_j}}$ di,dj为节点的度