

通过标记过的节点之间的引用关系图，来预测未被标记的节点分类。

总结分析

输入了什么

经过了什么变换

输出了什么

训练阶段:

经过处理后的节点特征矩阵、邻接矩阵、经过mask后带标签的训练实例、训练实例对应的mask张量（都是经过placeholder的数据，可减小运算操作和存储空间，下同）

验证阶段:

经过处理后的节点特征矩阵、邻接矩阵、经过mask后带标签的验证实例、验证实例对应的mask张量

测试阶段:

经过处理后的节点特征矩阵、邻接矩阵、经过mask后带标签的测试实例、测试实例对应的mask张量

两层图卷积

卷积操作被定义为：

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

Θ 是带过滤参数的矩阵， Z 是卷积信号矩阵。对邻接矩阵进行对称归一化后得到 A' 。 $A'X$ 是邻接矩阵（稀疏矩阵）和特征矩阵（稠密矩阵）的乘积。最后乘以一个过滤参数矩阵得到卷积后的矩阵。（代码中有具体实现）

第一层的输入是特征矩阵，卷积得到第一层输出

第一层输出作为第二层输入，卷积得到第二层（最后）输出输出的是节点对应的七类之一的预测标签。

卷积的最终输出是对节点的预测标签
程序的最终输出是预测的损失值和精度

问题总结

- 训练集是在每个类别上挑选20个样本得到的。本文使用拥有500个样本的验证集来优化参数，比如dropout比率、L2正则项权重和学习率。训练好模型后，对于每个数据集，本文在1000个随机挑选的测试样本上评估预测的正确率。
- 作者进行了1-10层的图卷积实验对比发现：2~3层图卷积的效果最好，这是因为over smoothing，随着层数的增加，远处的节点和近处的节点相似而更难以区分，效果反而降低。
- 为什么都有标签叫做半监督呢？半监督体现在文中的哪里？查询了一些资料发现，作者在计算损失函数时只考虑有标签的样本，先用有标签的计算损失求参数，然后再正向传播一次得到无标签节点的标签。半监督是因为用了数据集的所有特征加上了训练集的所有标签来预测的测试集的损失值。