图卷积网络

Thomas N. Kipf, Max Welling

Semi-Supervised Classification with Graph Convolutional Networks

<https://arxiv.org/abs/1609.02907>

问题描述：

分类问题上，假设每个样本都有一定的特征，部分样本有标签。如果是普通的神经网络，则训练过程只能用到有标签的样本。如果不同的样本之间有关系，普通网络也无法使用这些关联。

举例：社交网络成员的分类问题。

每个成员有固有的特征属性，如性别、年龄、发帖频率等。

成员之间有关注关系。

一共有N个成员以及他们的关注关系和特征属性

其中M个成员的标签已知，需要预测其他N-M个成员的标签。

构建GCN：

已知：一共有N个节点，特征矩阵，d表示特征的维数。邻接矩阵表示节点之间的联系。对于一般的无向图，如果节点i和j之间存在连接否则。如果边带有权重，那么。

卷积：一种信息的聚合。

考虑把节点邻居的信息聚合起来，

，，即卷积后第i个节点的第j个特征表示节点i所有邻居第j个特征的总和。这样引出了一个问题，因为一般，那么一次卷积后，节点自己的特征就丢失了。

改进：

其中，，I是单位矩阵，这样。所以表示节点i所有邻居第j个特征的总和再加上节点i自身的第j个特征。这样也有问题，即邻居较多节点的特征倾向于较大的值。

改进：

其中，是一个对角矩阵，，即对角线上的每个元素对应1加上第i个节点的度数。线性代数告诉我们，矩阵A乘以一个对角矩阵，其结果等于A的每行乘以对角矩阵对应行的元素。所以结果的每一行求和为1。

卷积神经网络：

代表网络第*l*层的特征图，。这和一般的多层感知机（MLP）的形式是很相似的，和分别为变换矩阵和截距项，只不过我们使用这一项来“卷积”了邻居节点的特征。为激活函数，可以是RELU，softmax等等。

训练（半监督）:

k为类别总数，前一层的激活函数用了softmax。由于只有M个样本有标签，因此损失函数只计算有标签数据的交叉熵

GCN的每一层包含了节点多少个hop邻居的信息？

记并忽略截距项和激活函数，则GCN的第二层可以表示为

卷积了节点邻居(1-hop)的信息，而则用到了邻居信息的邻居节点(2-hop)的信息（因为之后，每个节点的特征就含有其邻居的信息了，就将节点邻居的邻居的信息也囊括了进来）。

GCN的层数：

不宜太深。作者做了实验，发现层数过多时，准确率反而下降。一个直观的解释是，当GCN的层数越来越大，则节点汇聚了(k-hop)邻居的信息，而图论的研究表明，k-hop很容易就囊括了整张图的大多数节点，这表明每个节点的特征都是从相同的大多数节点卷积而来，反而没有了区分度。

注意力池化：通过注意力机制，提取图数据的层级特征

AttPool: Towards Hierarchical Feature Representation in Graph Convolutional Networks via Attention Mechanism

<https://ieeexplore.ieee.org/document/9009471>

任务：图分类

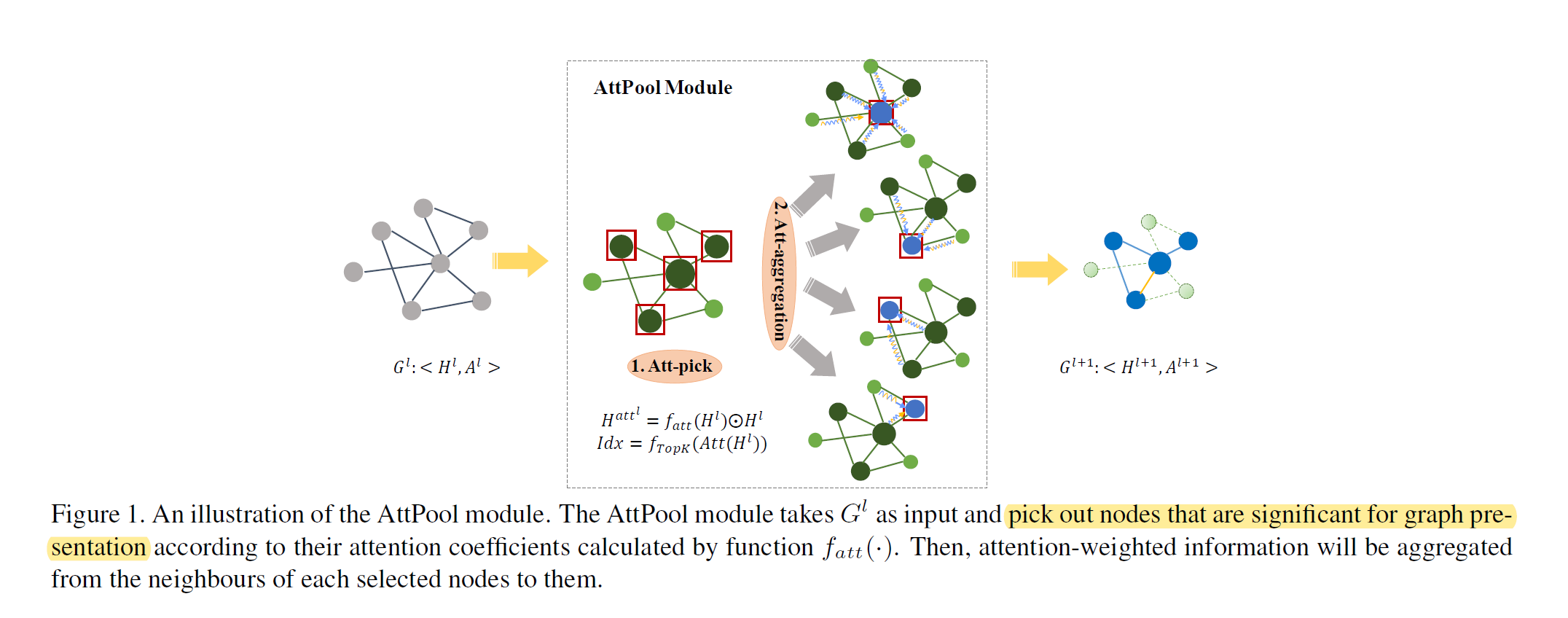
一般操作：通过GCN获得N个节点的特征，随后通过pooling操作将N个节点的特征聚合，一般的pooling操作包括max/average/sum等，得到一个1×d的特征，后接一个MLP来进行分类。

问题：需要一种分层的特征表达来帮助我们。作者举了个有机分子分类的例子，[单个原子、原子间的键] 🡪 [原子团、原子团的互相影响]🡪有机分子。如果可以获得特征的层级表示，这应该是很有帮助的。

CNN中的池化可以让空间域越来越小，卷积核大小不变的情况，感受野越来越大。对应了不同的层级。那么如何在图结构上实现pooling呢？grid形式的数据很好分割（partition），然后池化。**那在图里怎么分割**？

1. 基于聚类的方法，利用已经聚类好的结果(谱聚类)，pooling时基于聚类来分割。不能端到端的学习。
2. 基于学习的方法。Diffpool、graph-u-net，通过学习找到比较重要的节点和与这些重要节点比较相关的节点，划分到一起后提取高层次表达。

AttPool



步骤1：attention-pick，挑选出对任务最为关键的节点

步骤2：attention-aggregation，聚合信息。

其中，用于排序并选出值最大的k个节点的索引，是下采样率。获得了注意力向量后，可以更新节点的隐藏状态：

结合前面的索引，我们可以提取出一个包含k个被认为是最重要部分的节点组成的子图，它的邻接矩阵可以表示为

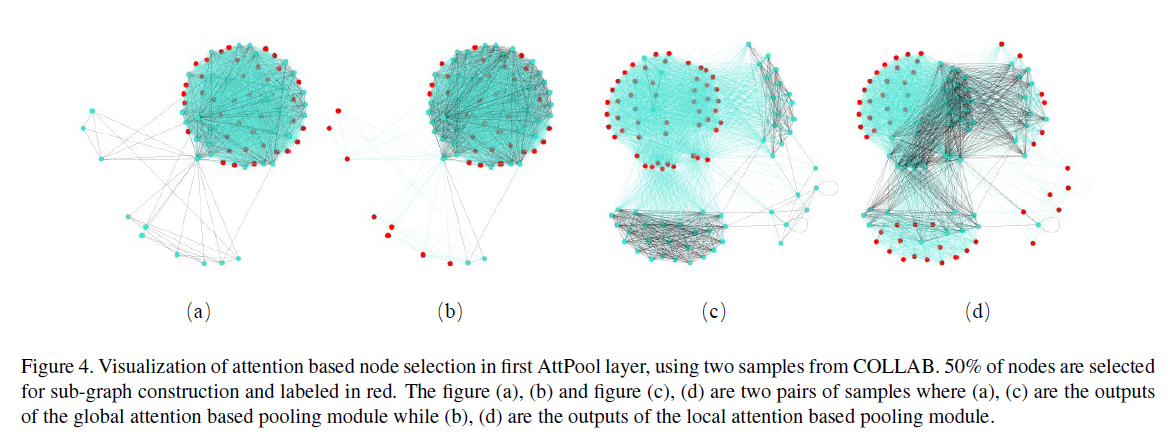
并且对每一行做softmax归一化。然后，对于k个被选中的节点，我们对做池化操作聚合信息，得到下一层的特征

同时，对聚合后的子图结构，更新邻接矩阵

问题：如何选择Att的方式？

Global Attention：衡量在所有节点中，每个节点的重要性。

但分析发现，这样做被选出来的点容易聚集到一起。（Attention is stuck in a certain parts of the graph）。因为本质上图卷积是把相邻节点的信息聚在了一起。而我们希望Attention可以关注到不同的重要parts，而不只是最重要的part。

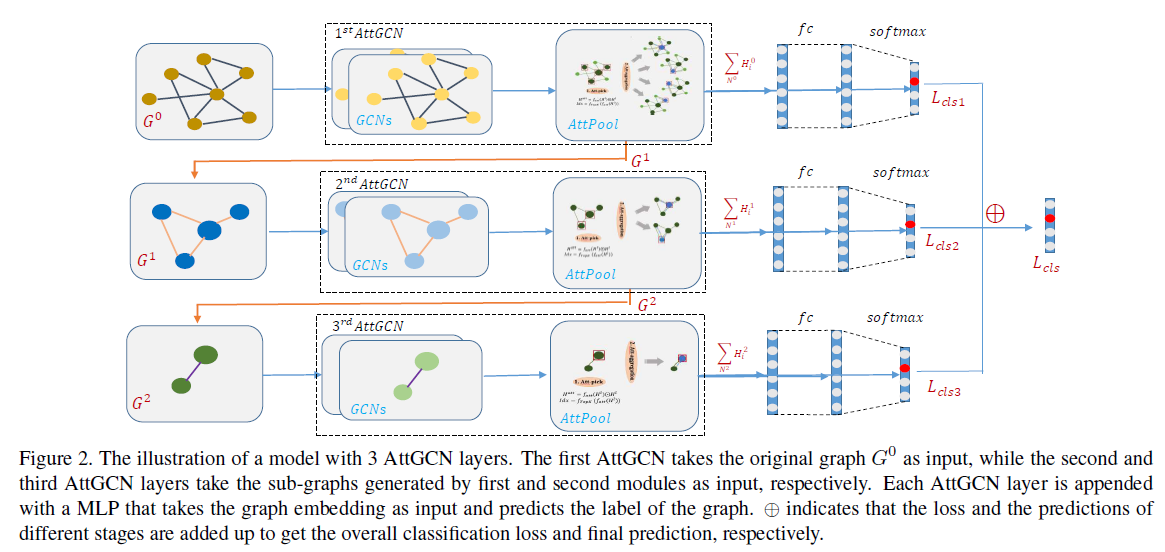


Local Attention

把衡量节点重要性的区域限制在了节点的1-hop邻居，而不是图中所有的节点。是可学习的标量。表示了重要性，表示距离，二者都被考虑到了。后面，为了防止节点越多导致的softmax值越小，用下式更新注意力系数

通过这种机制，就不容易stuck在图中的某一part。

分类模型：层次预测。每个pooling后加了分类分支，Intermediate supervision，训练时梯度直接回传；测试时使三个感受野的预测结果融合。每个层级的准确率还可以反过来证明图中的分层结构。



AttPooling可能对大的图效果比较好，直观来看，对于小的图，我们可能不需要层级特征就能做的比较好了。