

This is a note for the paper: Inductive Representation Learning on Large Graphs (NIPS 2017)

## 问题

对图上的每个节点输出一个嵌入向量，无监督学习。

作者的核心假设：图上相邻的节点应该具有相近的嵌入向量。由此引出无监督学习的损失函数。 $-\log(\sigma(\langle v_i, v_j \rangle))$ ，其中 $v_i, v_j$ 为相邻节点的嵌入向量，尖括号代表内积， $\sigma$ 为激活函数，保证输出大于0。作者在损失函数里还引入了负采样(negative sampling)。

## 算法

这个算法可以算是对之前很多GNN算法的总结，想法都是将邻域内节点的特征聚集到当前节点，作者把这个聚集相邻节点特征的过程统一称为aggregator。aggregator有不同的实现方式，理想情况是这是一个置换不变函数（因为邻域节点是无序的）。满足这一条件，可以对邻域特征取平均，邻域特征通过MLP取平均或取每个维度的最大值。作者也采用了LSTM，即先随意将邻居节点排序，再输入LSTM单元。为了应对很大的图反向传播很费时，作者提出了一个minibatch的方法，就是每次只采样一些节点进行传播更新，这是比较通用的方法。

采样的方法：分为K步进行（实际中取 $K=2$ ，权衡了效果和时间开销）。第一步采样当前节点的邻居，比如采 $S_1$ 个。第二次采当前节点和那 $S_1$ 个点的邻居，比如采 $S_2$ 个。第 $k$ 步采样的点距离当前节点至多为 $k$ 。

## 理论分析

作者在附录中的证明完全错误，理论本身也不合理。计算clustering coefficient是一个计数问题，证明过程的核心完全不是神经网络，而是说网络可以映射得到邻接矩阵，这个没意义。而且作者探讨的是Inductive learning，一旦加入新的节点，邻接矩阵的维数都在改变，神经网络的维度却是不变的。