

KDD20一篇将对比学习（contrastive learning）应用于图表示学习任务从而进行迁移的论文

解决的问题

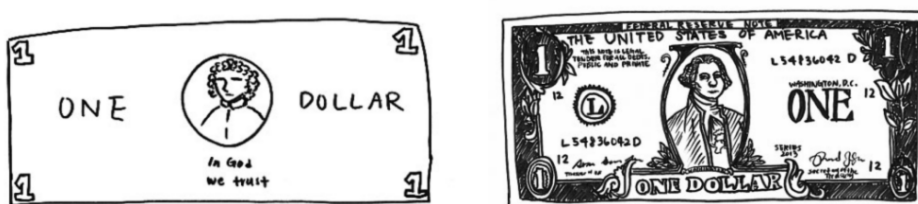
如何将自监督学习的思想应用与图表示学习，通过预训练图神经网络从而仅需要微调就可以应用于新的数据集。

图表示学习目前受到了广泛关注，但目前绝大多数的图表示学习方法都是针对特定领域的图进行学习和建模，训练出的图神经网络难以迁移。

做法及创新

对比学习

对比学习是自监督学习思想的一种典型框架，一个典型的例子如下图所示：



对比学习的思想是：尽管我们已经见过钞票很多次，能够轻易地分辨出一张钞票，我们也很少能画出一张完美无缺的钞票。表示学习算法不需要关注到样本的每一个细节，只要学到的特征能够将用来区分其它样本即可。不需要模型能够生成一匹栩栩如生的马之后它才能去分辨一张图片里的动物是不是马，这就是对比学习和生成对抗网络的一个区别。

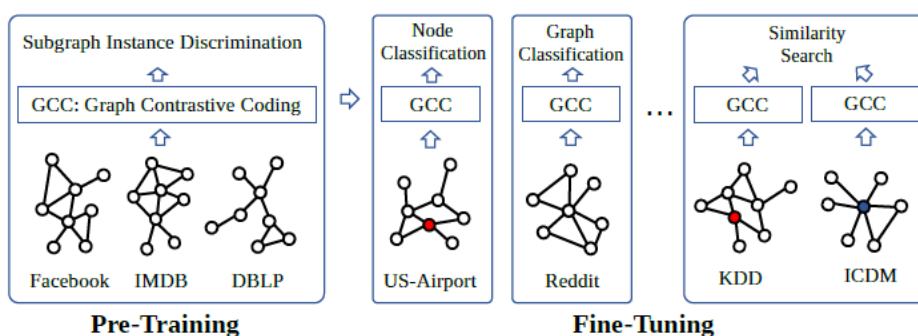
既然是表示学习，核心就是通过一个函数把样本 x 转换成特征表示 $f(x)$ ，而对比学习作为一种表示学习方法，它的思想是满足下面这个式子：

$$s(f(x), f(x^+)) \gg s(f(x), f(x^-))$$

使得类似样本之间的相似度要远大于非类似样本之间的相似度，这样才能够进行区分。

图表示学习

具体到论文的图表示学习任务中，论文的一个重要假设是，具有典型性的图结构在不同的网络之间是普遍存在而且可以迁移的（Representative graph structural patterns are universal and transferable across networks）。受对比学习在计算机视觉和自然语言处理领域的成功应用，论文想把对比学习（contrastive learning）的思想放在图表示学习中。通过预训练一个图神经网络，它能够很好地区分这些典型性的图结构，这样它的表现就不会仅仅局限于某个特定的数据集。



论文首先将现有工作对顶点相似度的衡量分为了三类：

1. 邻域相似度

核心思想：越近的两个顶点之间相似度越高，包括有Jaccard、RWR、SimRank以及LINE、DeepWalk、node2vec。

2. 结构相似度

核心思想：有相似的局部结构的两个顶点之间相似度更高。不同于邻域相似度，结构相似度不需要两个顶点之间有路径相连。常用的局部结构包括vertex degree、structural diversity、structural hole、k-core、motif等。

3. 属性相似度

当数据集中顶点有许多标签信息时，可以将标签作为顶点的特征来衡量它们之间的相似度。

在对比学习中，给定一个查询表示 q 以及一个包含 $K + 1$ 个键表示 k_0, \dots, k_K 的字典，我们希望找到一个能与 q 匹配的键 k_+ 。所以，论文优化的损失函数来自于InfoNCE：

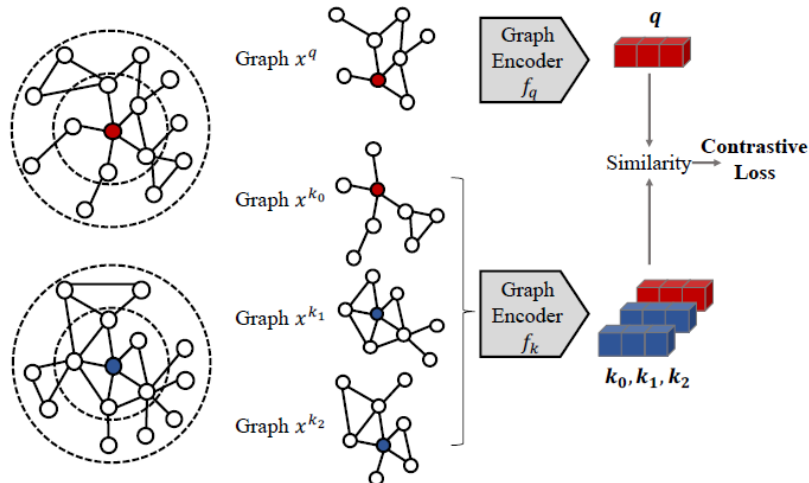
$$L = -\log \frac{\exp(q^T k_+ \tau)}{\sum_{i=0}^K \exp(q^T k_i / \tau)}$$

其中 f_q 、 f_k 是两个图神经网络，分别将样本 x^q 和 x^k 转换为低维表示 q 与 k 。

正负样本获取

因为查询和键可以是任意形式，具体到本论文里，定义每一个样本都是一个从特定顶点的 r 阶邻居网络中采样的子图，这里的子图定义和其它论文一致： $S_v = \{u : d(u, v) < r\}$ ，距离顶点 v 最短路径距离小于 r 的顶点构成的集合。既然是最短路径，给定 r 那么这个集合也基本确定了，这种情况下得到的子图数量有限，在计算机视觉领域，当输入用于训练的图片数量有限时，往往会使用反转、旋转等方式对图片进行变换，以扩充训练图片的数量，这里论文也想采取类似的做法，对得到的子图 x 进行变换，来得到对比学习中的类似 x^+ 与非类似样本 x^- ，具体做法如下：

1. 带重启的随机漫步。首先从子图的中心顶点 v 开始随机漫步，每一步时都有一定概率重新回到中心顶点，而漫步到任一邻居顶点的概率与当前顶点的出度有关。
2. 子图推演。随机漫步可以得到一系列顶点，它们构成的集合记为 \tilde{S}_v ，所形成的子图记作 \tilde{G}_v ，它可以看作子图 S_v 的一个变换。
3. 匿名化。重新定义 \tilde{G}_v 中的顶点的标签，将 $\{1, 2, \dots, |\tilde{S}_v|\}$ 的顺序随机打乱作为重新定义后的标签。



论文对于每个子图都进行两次上述变换，而变换后的子图显然会与原子图相似，这样就有了一组相似的子图 (x^q, x^{k+}) 。要得到不相似子图也很容易，不是同一个子图变换得到的子图就定义为不相似：

$(x^q, x^k), k \neq k_+$ 。在上图的例子中， x^q 和 x^{k_0} 是从红色的中心顶点采样得到的子图，我们认为它是一对正样本，而 x^{k_1} 和 x^{k_2} 作为从蓝色的中心顶点采样得到的子图，则被作为负样本。在变换时之所以要做最后一步，是为了防止图神经网络在判断两个子图是否相似时，仅仅是通过判断对应顶点的标签是不是一样，这样显然没有学到任何有用的结构信息。这里有一个小结论：

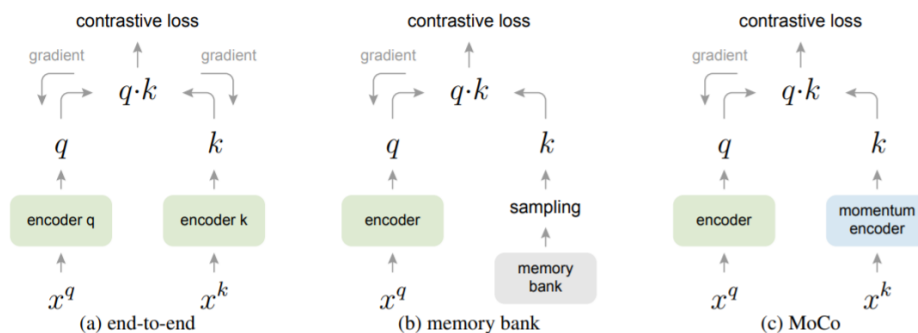
绝大多数图神经网络对于输入图中顶点的顺序的随机扰动有稳定性

现在有了正样本和负样本，下一步就是训练一个图神经网络对它们加以区分了，论文选取的是GIN。这就是自监督学习的思想，对比学习就是这种思想的一种典型框架。因为现有的图神经网络框架都需要额外的顶点特征作为输入，论文提出了一种位置embedding来作为其中特征：

$I - D^{-1/2} A D^{-1/2} = U \Lambda U^T$ ，矩阵 U 中排序靠前的特征向量作为embedding。其它特征还包括顶点度的one-hot编码和中心顶点的指示向量。

模型学习

在模型学习时采用了何凯明组的MoCo框架的思想：



在对比学习中，我们需要维护一个大小为 K 的字典和编码器，要计算上面定义的损失函数，理想的情况是把所有负样本加入字典中进行计算，这会导致 K 很大字典难以维护。在MoCo的方法中，为了增大字典大小 K ，需要维护一个负样本的队列，队列中包含此前训练过的batch的样本作为负样本。在更新参数时，只有 q 的编码器图神经网络 f_q 中的参数通过反向传播进行更新，而 k 的编码器 f_k 中的值通过一种动量法进行更新： $\theta_k \leftarrow m\theta_k + (1 - m)\theta_q$ 。

数据集

Academia、DBLP(SNAP)、DBLP(NetRep)、IMDB、Facebook、LiveJournal