Author = Song Jian

#### 阅读论文： DropEdge: Towards Deep Graph Convolutional Networks on Node Classification

|  |  |  |
| --- | --- | --- |
| **@article{Etworks2020,**  **author = {Etworks, Tional N and Ode, O N N},**  **title = {{DropEdge: Towards Deep Graph Convolutional Networks on Node Classification}},**  **year = {2020}**  **}** | [1] T. N. Etworks and O. N. N. Ode, “DropEdge: Towards Deep Graph Convolutional Networks on Node Classification,” no. 2019, pp. 1–17, 2020.  源代码：https://github.com/DropEdge/DropEdge. | **腾讯AI实验室**  **清华大学** |

#### 文档由以下几个部分组成：

* 文章简述
* 作者贡献
* 模型介绍
* 实验部分
* 总结

#### 文章简述

摘要

用于节点分类的GCN在层次过深时会遇到两个主要的问题：过度拟合和过度平滑。过度拟合会削弱小型数据集的泛化能力，而过度平滑会通过随着网络深度的增加将输出表示与输入特征隔离开来，从而阻碍模型训练。

本文提出了DropEdge机制，在模型训练时随机删减掉原始图中的边，来缓解这两个问题。此外，作者还从理论上证明了DropEdge既可以降低过平滑的收敛速度，又可以减少由过平滑引起的信息损失。并且DropEdge可以用于许多GNN模型以增强其性能，例如JKNet、GCN、ResGCN、GraphSAGE。实验结果显示DropEdge可以提升多个浅层和深层GCN模型的性能。

简介

过度拟合来自以下情况：当我们利用过度参数模型来拟合训练数据有限的分布时，我们所学习的模型非常适合训练数据，但对测试数据的泛化性很差。

平滑的function可以降低对输入的敏感性，但是随着网络层数的加深，过于平滑的function则会走向极端，将输出与输入特征个离开来。比如Deeper insights into graph convolutional networks for semi-supervised learning以及A comprehensive survey on graph neural networks中都有介绍。在下图中展示了CORA数据集上的4层gcn的过拟合以及8层GCN的过平滑导致的收敛问题

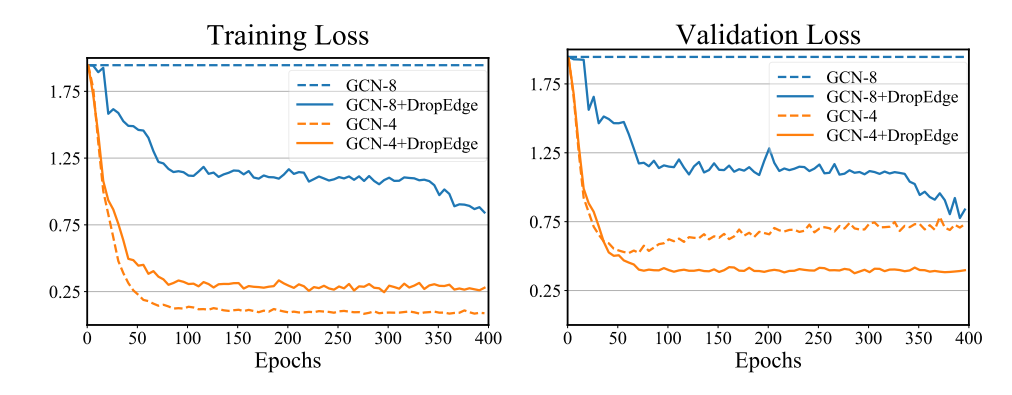


Figure 1: Performance of Multi-layer GCN

作者提出的DropEdge方法可以缓解以上两个问题。

* 首先可以将DropEdge视为一种数据增强技术。通过DropEdge，我们实际上是在生成原始图的不同随机变形副本。这样，我们增强了输入数据的随机性和多样性，因此能够更好地防止过度拟合。
* 其次，DropEdge也可以被视为消息传递缩减器。在GCN中，在相邻节点之间传递的消息是沿着边缘路径进行的。删除某些边缘使节点连接更加稀疏，因此在GCN变得很深时在某种程度上避免了过度平滑。

实际上，DropEdge作为一种灵活而通用的技术，能够在多个基准上增强各种流行的骨干网的性能，包括GCN（Kipf＆Welling，2017），ResGCN（Li等，2019），JKNet（Xu等）。等人（2018a）和GraphSAGE（Hamilton等人，2017）。我们在实验中提供详细的评估。

作者还在理论上证明了DropEdge既可以降低过平滑的收敛速度，又可以减少由过平滑引起的信息损失

#### 模型介绍

首先GCN的公式为

每次训练时，DropEdge机制会随机删除掉原始图中固定比例的边。（只在训练集上使用，在测试集和验证集上不使用DropEdge机制）也就是说随机选取邻接矩阵A中Vp个非零的元素，将其置零。其中，V是原始图中的总边数，p是删除率。删除后得到邻接矩阵：

然后对进行re-normalization得到，替换GCN中的。

##### 阻止过拟合

DropEdge对图中的连接带来了扰动，它对输入数据产生了不同的随机变形，可以看成是数据增强。GCNs的核心思想是对每个节点的邻居特征进行加权求和（权重和边有关），实现对邻居信息的聚合。从这一角度出发，可以看出DropEdge在GNN训练时使用的是随机的邻居子集进行聚合，而没有使用所有的邻居。DropEdge删边率为p，对邻居聚合的期望是由p改变的。在对权重进行归一化后就不会再使用p。因此，DropEdge没有改变邻居聚合的期望，是用于GNN训练的无偏的数据增强方法，可以有效防止GNN训练时的过拟合问题。类似于经典的图像数据增强方法：rotation, cropping, flapping。

上述的方式是使用同一个邻接矩阵，也可以对每一层进行DropEdge。

##### 阻止过平滑

过平滑指的是随着网络的不断加深，节点的表示最终会收敛到一个固定点。这种不必要的收敛使得深度GCNs的输出只和图的拓扑结构有关，独立于输入的节点特性，这会损害GCNs的表示能力。Oono和Suzuki同时考虑了非线性（例如ReLu函数）和卷积核，他们将过平滑看成是收敛到一个子空间，而不是收敛到一个固定点。本文使用的是这些学者提出的子空间的概念。

作者从两个角度解释了DropEdge有助于缓解smoothing问题：

1）通过减少节点间的连接，DropEdge可以降低过平滑的收敛速度

2）原始空间和收敛子空间的维度之差（例如 N−M）衡量了信息的损失量，这个差越大说明信息损失越严重。DropEdge可以增加子空间的维度，也就具有减少信息损失的能力。

##### 比较

比较DropEdge和其他相关概念的不同，包括Dropout、DropNode和Graph Sparsification。

**（1）DropEdge vs. Dropout**

Dropout是对特征向量中的某些维度随机置零，可以缓解过拟合，但是不能缓解过平滑，因为它没有改变邻接矩阵。DropEdge可以看成Dropout向图数据的推广，将删除特征维度变换成删除边，有助于缓解过拟合和过平滑问题。

实际上两者是互补关系，实验中对两者进行了比较。

**（2）DropEdge vs. DropNode**

作者将基于节点采样的方法称为DropNode（GraphSAGE, FastGCN, AS-GCN中均有使用）。DropNode的思想是采样子图用于mini-batch训练，和删除的节点相连的边也被删除了，因此它可以看成是删除边的特殊形式。然而，DropNode中删除的边是面向节点（node-oriented）的且无向的。DropEdge是面向边（edge-oriented）的，它可以在训练时保留所有节点的特征，更具灵活性。而且，现有的DropNode的方法通常是低效的，GraphSAGE的layer size是指数级增长的，AS-GCN每层都需要进行采样。DropEdge在加深层数时不会增加layer size，也不需要递归地采样，DropEdge对所有边的采样是并行的。

（**3）DropEdge vs. Graph-Sparsification**

图稀疏性的目标是去除掉不必要的边，以对图进行压缩，同时尽可能地保留原始输入图中的信息。

DropEdge在每次训练时会随机删除输入图中的一些边，而图稀疏性的方法需要复杂的优化过程来决定删除掉哪些边。

#### 实验部分

1、数据集

（1）对papers的研究领域进行分类：Cora、Citeseer、Pubmed（transductive）

（2）预测社交网络中的帖子属于哪个社区：Reddit（inductive）

2、对比方法GCN; ResGCN; JKNet; INcepGCN; GraphSAGE

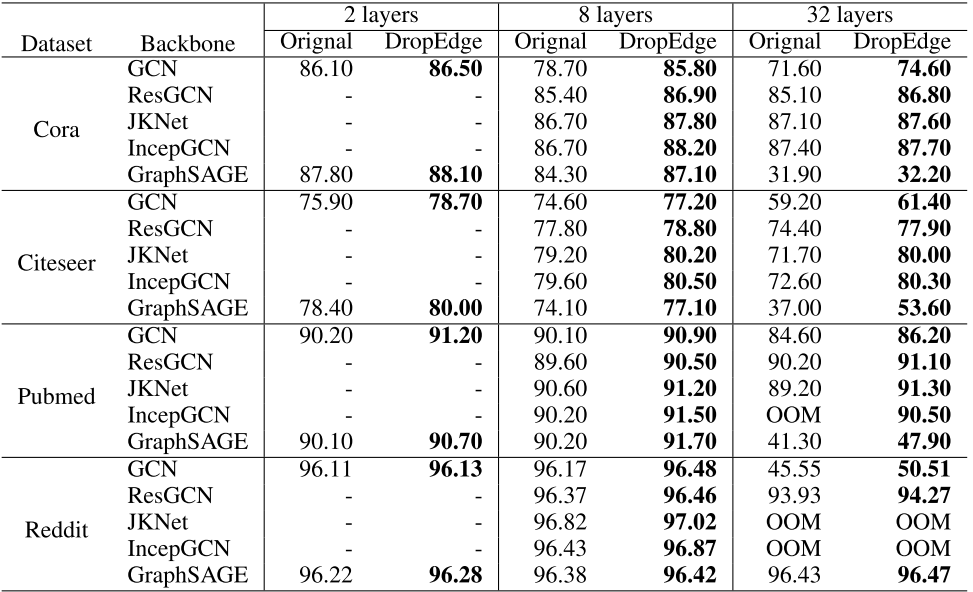


Table 1: Testing accuracy (%) comparison

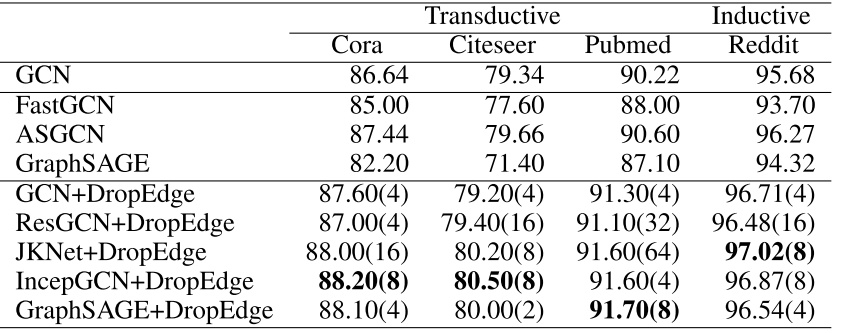


Table 2: Accuracy (%) comparisons with SOTAs. The number in parenthesis denotes the network depth for the models with DropEdge.

##### 过平滑的测试

计算当前层输出和上一层输出的欧氏距离，来衡量过平滑的程度。距离越小说明过平滑现象越严重。图1a可以看出，随着层数的加深，过平滑现象越来越严重。但是使用DropEdge（p=0.8）时过平滑的收敛速度明显减缓了。图1b显示，在训练之后，没有使用DropEdge的GCN第5层和第6层输出的差别为0，而使用了DropEdge的GCN随着层数的加深，不同层之间的差别并没有减为0。

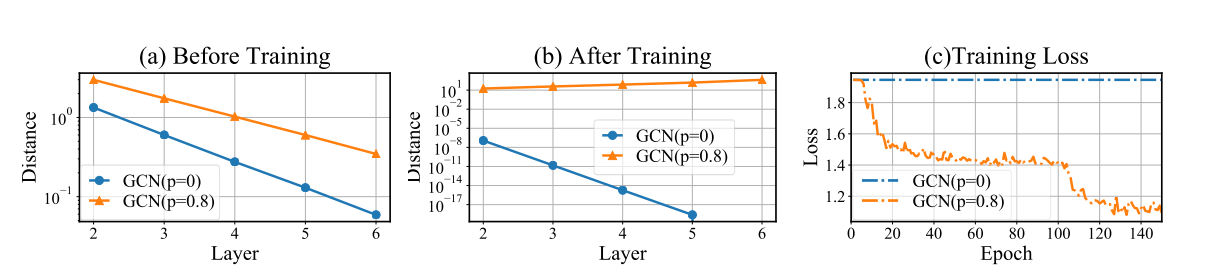


Figure 1

##### 和DropOut的兼容性

在4层的GCN上进行消融实验。验证集上的损失如图2 a所示，结果表明若同时使用Dropout和DropEdge有助于GCN的训练。

如图2b所示，layer-wise的DropEdge比原始的DropEdge在训练时的损失更小，两者在验证集上表现差不多。这表明layer-wise的DropEdge有助于训练的进行。但是为了避免过拟合的风险，以及减小计算复杂度，作者倾向于使用原始的DropEdge而不是layer-wise的DropEdge。

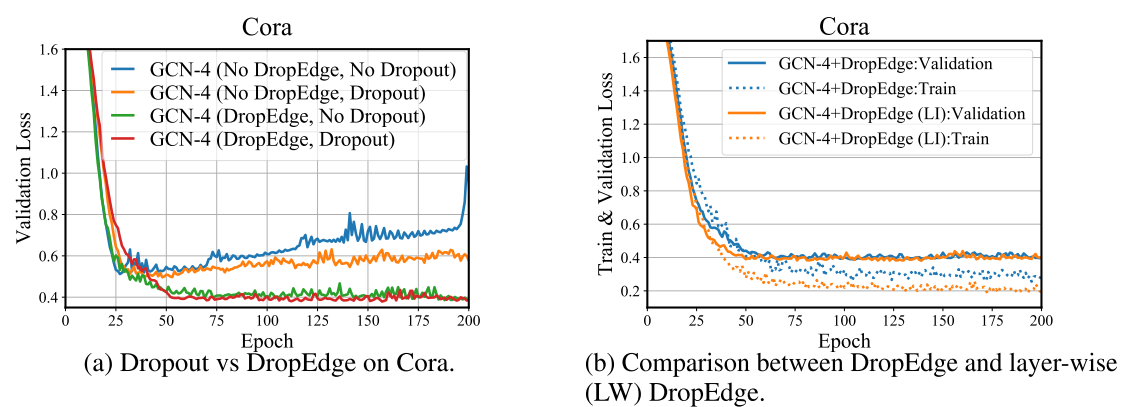


Figure 2

#### 模型介绍

本文提出DropEdge机制，有助于图卷积网络的加深，以用于节点分类任务。具体操作是在每次训练时随机删除图中固定比例的一些边。DropEdge增加了输入数据的多样性，从而缓解了过拟合问题；还减少了图卷积时信息的传递，从而缓解了过平滑问题。在多个数据集上进行实验，证明了在已有的GCN模型上使用DropEdge，可以提升原有模型的性能。