

附加特征图增强的图卷积神经网络

孙隽姝^{1),2)} 王树徽^{1),3)} 杨晨雪⁴⁾ 黄庆明^{1),2),3)} 郑振刚^{5),6)}

¹⁾(中国科学院计算技术研究所智能信息处理实验室 北京 100190)

²⁾(中国科学院大学计算机科学与技术学院 北京 100049)

³⁾(鹏城实验室 广东 深圳 518055)

⁴⁾(中国农业科学院农业信息研究所 北京 100081)

⁵⁾(香港大学计算机科学系 香港)

⁶⁾(深圳大学粤港澳智慧城市联合实验室 广东 深圳 518060)

摘要 近年来,图卷积网络(Graph Convolutional Network,GCN)凭借其简单的网络结构、在图上任务中展现出的优异性能,受到了学术界和工业界的广泛关注。然而GCN也存在着在浅层时信息传播范围过小、特征提取不充分的缺陷。针对这一问题,本研究提出附加特征图模型(Additional Feature Graph,AFG)。AFG通过引入图的节点结构特征(度特征),对度相同的节点随机增加连边、缩短信息传播距离。AFG并不是独立的图神经网络模型,而是作为一种附加技术与GCN及其相关模型配合使用。实验显示,在Cora、Citeseer、Pubmed数据集上AFG能够对浅层主干模型实现显著性能增益,帮助主干模型性能超越了其他以提升模型特征提取能力、改善欠传播情况为目的进行设计的模型。本研究进一步分析了AFG与DropEdge——一种随机切断原始图连边的附加技术——的区别与联系,并通过实验证明附加特征图模型与DropEdge模型共同使用的可行性,以及两者间存在一定的互补性。结合使用两种附加技术可以实现更大的节点分类准确度增益。

关键词 图表示学习;图神经网络;信息传播;图卷积网络;节点分类

中图法分类号 TP391

DOI号 10.11897/SP.J.1016.2023.0190

Graph Convolutional Neural Networks with Additional Feature Graph

SUN Jun-Shu^{1),2)} WANG Shu-Hui^{1),3)} YANG Chen-Xue⁴⁾ HUANG Qing-Ming^{1),2),3)}
Reynold C. K. Cheng^{5),6)}

¹⁾(Key Lab of Intelligent Information Processing (CAS), Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190)

²⁾(Department of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing 100049)

³⁾(Pengcheng Laboratory, Shenzhen, Guangdong 518055)

⁴⁾(Agriculture Information Institute of Chinese Academy of Agriculture Sciences, Beijing 100081)

⁵⁾(Department of Computer Science, The University of Hong Kong, Hong Kong)

⁶⁾(Guangdong-Hong Kong-Macau Joint Laboratory, Shenzhen University, Shenzhen, Guangdong 518060)

Abstract Graph structures are suitable for the modeling of complex interactions and relations. Therefore, graphs are widely used in data representation such as molecules, chemical compounds, citation networks, social networks, traffic web, and knowledge graphs. In light of the great success of neural networks in image understanding and natural language processing, there

收稿日期:2023-03-6;在线发布日期:2023-03-13. 本研究得到科技创新2030-新一代人工智能重大项目:面向跨媒体内容管理的智能分析与推理(No. 2018AAA0102000)、国家自然科学基金委员会:跨媒体理解与知识推理(No. 62022083)、国家自然科学基金委员会:数据和知识联合驱动的跨媒体语义理解与文本生成(No. 62236008)、中国科学院计算技术研究所创新课题(E161060)、香港大学项目(104005858, 10400599)、粤港澳联合实验室项目(2020B1212030009)、鹏城实验室重大攻关项目:脑眼融合的智能感知计算技术与平台(PCL2023AS6-1)等项目资助。孙隽姝,硕士研究生,主要研究领域为图表示学习、计算机视觉。E-mail: junshu.sun@vipl.ict.ac.cn. 王树徽(通信作者)博士,研究员,中国计算机学会(CCF)高级会员(22482S),主要研究领域为跨媒体分析推理与可信机器学习方法。E-mail: wangshuhui@ict.ac.cn. 杨晨雪,博士,助理研究员,主要研究领域为机器学习、数据挖掘及应用。黄庆明,博士,教授,中国计算机学会(CCF)会士,主要研究领域为多媒体分析、计算机视觉、模式识别理论与方法。郑振刚,博士,教授,主要研究领域为大规模数据管理、知识库、社交网络、众包技术等。

has been a rising interest in Graph Neural Networks (GNN) for the study of learning on graphs. Among the popular GNNs, Graph Convolutional Networks (GCN), highlighted for their simple network structures and excellent performance with graphs, have attracted wide attention and become a promising direction. However, the limitation on message passing distance deteriorates the performance of GCN. To address this problem, people propose constructing deep GCN models to improve propagation. However, as the depth of GCN increases, node features become over-smoothed or over-squashed. This leads to a sharp decrease in the model performance. Though different deep GCN models have been proposed to tackle the over-smoothing or over-squashing problem, the inherent problems of the message passing mechanism are less explored. The problems of message passing in GCN include: 1) having unreliable paths in the original input graphs which pass information with low signal-to-noise ratios, 2) reaching limited message passing extent which gives rise to the less expressive feature representation, and 3) lacking explicit learning fashion on structural features. To this end, some people seek to directly improve the message passing in shallow GCN concerning the limitation in robustness, message passing extent, or feature diversity. But all these studies fail to tackle three problems in a single model. In this paper, we propose a novel model named AFG (Additional Feature Graph), to improve the message passing over robustness, message passing extent, and feature diversity. Specifically, AFG can inject the structural features of the input graph into the message passing process and randomly add edges between node pairs that have the same degree. The degree feature represents the first-order topological structure of individual nodes. Connecting nodes with the same degree allow explicit learning on structural features. As a lightweight and general technique, our AFG model can be easily plugged into GCN and its related models, bringing extra improvements. Experimental results on three datasets demonstrate the efficiency of AFG. AFG-aided models outperform shallow backbones and related GCN-based models on Cora, Citeseer, and Pubmed. AFG achieves 0.69% averaged improvement on 2-layer models, 0.57% on 4-layer models, 1.18% on 6-layer models, and 0.99% on 8-layer models. The improvements are proven to be significant with hypothesis testing. We also provide detailed experimental analysis on AFG with both synthetic datasets and real world datasets. The corresponding experimental results demonstrate that AFG can improve the connectivity and shorten the average path length of the input graphs. With AFG, nodes can get more informative features from their neighbors. Compared to GCN, AFG reaches a broader message passing extent in shallow model structures. Furthermore, we provide experiments verifying that AFG and DropEdge (another plug-and-play technique for GCN models) are complementary to each other and can be combined to achieve better performance.

Keywords graph representation learning; graph neural networks; message passing; graph convolutional networks; node classification

1 引言

现实生活中存在大量的图结构数据,如新浪微博、豆瓣等社交网络数据,生化分子结构数据,引文网络数据等等^[1]。单纯地将图转化为欧几里得结构数据、交由卷积神经网络(Convolutional Neural Networks,CNN)等传统神经网络方法处理会造成

拓扑结构信息的损失,因此更好的图结构数据处理方法——图神经网络(Graph Neural Network,GNN)应运而生。

作为图神经网络的重要分支,由 Kipf 等^[2]提出的图卷积网络(Graph Convolutional Network,GCN)凭借其优异的性能、简单的网络结构,受到了学术界和工业界的广泛关注。目前,GCN 已被成功地应用于推荐系统^[3-5]、药物研发^[6-8]、点云分

割^[9]和交通流预测^[10-12]等任务。尽管在多种任务上性能表现优异,GCN 也存在着众多缺陷。在本文中,我们主要关注 GCN 的浅层网络信息欠传播问题。

根据 Gilmer 等^[13]的研究分析, GNN 的前向传播过程通常可以被划分为信息传播和节点更新。其中,信息传播指以目标节点为中心,聚合邻居节点状态信息(输入层状态信息即为节点特征信息)。GCN 模型通过图空间域上的卷积实现信息传播,且信息传播范围随着卷积层的叠加逐渐扩大。

为了扩大 GCN 模型的信息传播范围、避免可能出现的信息欠传播,一种自然的思路是构造深度 GCN 模型^[14]。这种思路主要面临的困难是,GCN 模型在深度增加时会出现梯度消失、图上节点特征分布过度平滑(即信息过传播)、网络性能快速下降甚至崩溃等现象。对此,研究人员从不同的角度提出了解决方案:在网络中增加跳连,如 JKNet^[15],Deep-GCN^[9,16]等;删除图上连边,如 DropEdge^[17-18](见图

1-b)。

不同于构造深度 GCN 模型的思路,本文从思考如何扩大浅层网络信息传播范围的角度入手,提出一种 GCN 及其相关模型的改进技术——附加特征图模型(Additional Feature Graph, AFG)。AFG 通过引入图的节点结构特征,对度相同的节点对随机增加连边,缩短信息传播距离(见图 1)。度特征是节点拓扑连接结构的体现,度相同的节点拥有相同的一阶拓扑结构。通过令度相同的节点之间建立连边,AFG 可以显式地在一阶拓扑结构相同的节点之间进行信息传递,提取局部拓扑相同的节点之间的共性特征。同时 AFG 能够帮助 GCN 类模型在不增加网络深度的同时,扩大信息传播范围,提取更多图上特征。特别的,AFG 并非一种独立模型,而是作为附加改进策略,嵌入 GCN 等主干模型(backbone)中,实现模型性能的提升。AFG 的卷积层信息传递如图 1-c 所示。本文将在第 3 节对 AFG 进行详细介绍。

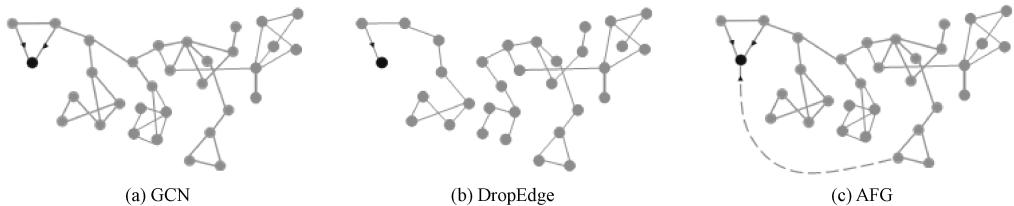


图 1 模型信息传递对比

本文的贡献可以被总结为三个方面:(1)提出附加特征图模型,为 GCN 及其相关模型引入节点的度特征;(2)在节点分类任务上,通过实验证明了附加特征图模型对主干模型的性能增益作用,证明了模型各部分对于整体效果发挥的必要性,揭示了附加特征图模型能够有效缩短平均路径长度,从而扩大信息传播范围;(3)在实验中发现了附加特征图模型与 Rong 等^[17]提出的 DropEdge 对模型性能增益效果的互补性。

2 相关工作

GCN 模型通过网络深度的增加逐渐扩大节点信息的传播范围。因此当 GCN 深度较浅时,模型对于输入数据的特征提取和表达相对不足。针对这种信息欠传播现象的改进思路主要可分为两种,本文将在第 3 节从第二种角度入手尝试解决欠传播问题。

2.1 模型深度

受深度学习在神经网络中的成功应用启发,人们很自然地开始探索深度 GCN 模型的可行性。然而,GCN 模型深度的增加会导致模型性能的快速下降,甚至出现性能崩溃。目前,学界对这一现象的主要解释包括以下 3 类:

(1)过平滑。主流研究观点称这种现象为“过平滑”。Li 等^[19]推导证明了 GCN 的模型本质为拉普拉斯平滑。基于这一结论可知,GCN 对输入信号进行逐层拉普拉斯平滑,导致高频信息丢失、深层网络提取到的节点特征同质化严重。从信息传播角度分析,过平滑现象也被解释为邻居节点间信息的过度共享、交换。

(2)过拟合。Yang 等^[20]通过分析深度 GCN 的训练过程,以及深度 GCN 在训练、测试中的损失函数曲线提出:GCN 能够在训练中依赖监督信号纠正节点的过平滑表示,而性能崩溃来自深层 GCN 模型对训练样本的过拟合。

(3)过挤压。Alon 和 Yahav^[21]通过对图神经网

络进行分析,提出了过度挤压的观点,认为图神经网络中沿连通节点链路进行的信息传播导致随着传播距离变长,规模以指数级增长的节点信息被挤压入了固定大小的向量中。这种过度挤压带来了大量有效信息的损失,映射到 GCN 网络结构中,即产生深层网络性能崩溃问题。

为了构建深度 GCN 模型,Rong 等^[17]提出一种附加模型 DropEdge,尝试削减节点间信息的过度传播,解决深层 GCN 过平滑问题。从解决过平滑的角度来说,DropEdge 通过断开连边,截断了节点间信息传播的路径,避免了信息过度交换。同时由于断开操作的随机性,DropEdge 有助于抑制噪声信息的交换。从解决过拟合的角度来说,DropEdge 在每轮训练中都生成新的邻接矩阵,相当于对输入数据做了随机变形,起到了扩大训练集的作用。

Li 等^[9]受 ResNet^[22]、DenseNet^[23]的启发,将残差连接和膨胀卷积引入 GCN 模型,提出了 Deep-GCN 模型。该模型成功地在避免性能崩溃的同时,将网络深度叠加到了 112 层。

2.2 信息传递

不同于从模型深度入手的解决方案,第二类方向主要从改善模型信息传递机制出发解决欠传播问题。信息传递机制通过邻域节点信息聚合和更新实现。GCN 通过将信息传播限制在连通的邻居间,显式地提取了图上节点的特征信息,同时隐式地编码了图的拓扑连接信息。这样的信息传播方式使得 GCN 仅仅基于拓扑连通性对中心节点及其邻居进行拉普拉斯平滑。由于非欧式结构数据的特殊性,图上拓扑连通的两个节点尽管在连通路径上临近,其特征并不一定相似。在这种情况下,GCN 的信息传递方式存在以下问题:(1)可能存在信息损失和噪声,消息流的信噪比低;(2)过于关注邻域内节点,特征提取不足;(3)缺乏对拓扑结构信息的显式利用。因此,为改善 GCN 模型存在的欠传播问题,模型需要具备广泛特征的鲁棒学习能力,同时兼顾特征提取的局部性和全局性,即一个更具表达力、更远、更多元的信息传递图模型。

在特征鲁棒学习能力方面,根据邻域中不同节点提供的信息,Veličković 等^[24]引入注意力机制对邻域节点赋权,实现信息的加权聚合。以 Hamilton 等^[25]提出的 GraphSAGE 为代表的邻域采样类^[17,26]模型通过随机采样邻居节点进行特征聚合,帮助聚合函数学习邻域节点更具代表性和决定性的特征。采样模型的聚合过程等价于随机赋二值权重

的加权求和过程,即注意力模型和采样模型都通过对节点邻域的加权聚合来帮助更好地提取节点特征,提升聚合信息的信噪比。但他们却忽视了信息欠传播中远程节点信息缺失的问题,也没有很好地针对图结构信息进行模型设计。

为了扩大浅层信息传播范围,Gasteiger 等^[27]引入个性化随机游走来获取远距离节点信息。Abu-El-Haija 等^[28]提出 MixHop 模型,通过拼接不同幂次的邻接矩阵在同一层中获得多阶邻居信息。Klicpera 等^[29]将个性化随机游走等扩散机制泛化,提出利用扩散机制重新学习图结构的方法。这种方法使得输入图中的节点距离被打破重排,改变了原有信息传播的通路,扩大了节点的信息传播范围。

在多元特征引入方面,部分研究人员尝试通过度特征编码图结构信息。Wu 等^[30]受 Weisfeiler-Lehman 测试^[31]启发,认为图卷积应该具备编码度信息的能力,提出根据各节点的度值编码聚合权重。Corso 等^[32]在模型 PNA 中编码了度信息,并指出度反映了局部节点的连接状态,度的微小变动可能意味着节点信息或梯度的指数级变化。以上研究均从信息传播机制缺陷的某一点入手尝试对模型进行改进,忽视了其他缺陷对模型能力的影响。对此我们提出 AFG 模型来尝试对信息欠传播引起的低鲁棒性、特征提取不充分问题做出综合改进。

3 附加特征图模型

为了解决浅层 GCN 模型的信息欠传播问题,提升模型特征提取能力,本文提出 AFG。AFG 通过引入新的特征信息(度信息)来建立新的拓扑连接,从而扩大信息传播范围,提升浅层 GCN 模型性能。同时,由于不需要叠加 GCN 深度,AFG 规避了深层 GCN 模型可能出现的过平滑、过拟合或过挤压等问题。AFG 的整体结构如图 2 所示。下面首先介绍主干模型 GCN,随后具体介绍 AFG 方法细节。

3.1 GCN 模型

给定图 $G = (V, E)$, 设节点个数 $|V| = N, V = \{v_1, v_2, \dots, v_N\}, e_{v_i v_j} \in E$ 。定义图的邻接矩阵为 $\mathbf{A} \in \mathbb{R}^{N \times N}$, 度矩阵为 $\mathbf{D} \in \mathbb{R}^{N \times N}$, 拉普拉斯矩阵 $\mathbf{L} \in \mathbb{R}^{N \times N}$, 其中

$$A_{ij} = \begin{cases} 1, & e_{v_i v_j} \in E \\ 0, & \text{其他} \end{cases} \quad (1)$$

$$D_{ii} = \sum_{j=1}^n A_{ij} \quad (2)$$

$$\mathbf{L} = \mathbf{D} - \mathbf{A} \quad (3)$$

GCN 模型由 Kipf 等^[2]提出。取输入信号 $X_n \in \mathbb{R}^{N \times C}$, C 为节点特征维度, GCN 模型的单层卷积过程可写作

$$\mathbf{Z} = \widetilde{\mathbf{D}}^{-\frac{1}{2}} \widetilde{\mathbf{A}} \widetilde{\mathbf{D}}^{-\frac{1}{2}} \mathbf{X}_n \mathbf{W} \quad (4)$$

其中, $\widetilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}_N$, \mathbf{I}_N 表示 N 阶单位阵, $\widetilde{D}_{ii} = \sum_j \widetilde{A}_{ij}$, $\mathbf{W} \in \mathbb{R}^{C \times F}$ 为可学习的滤波器参数, F 为滤波器个数。

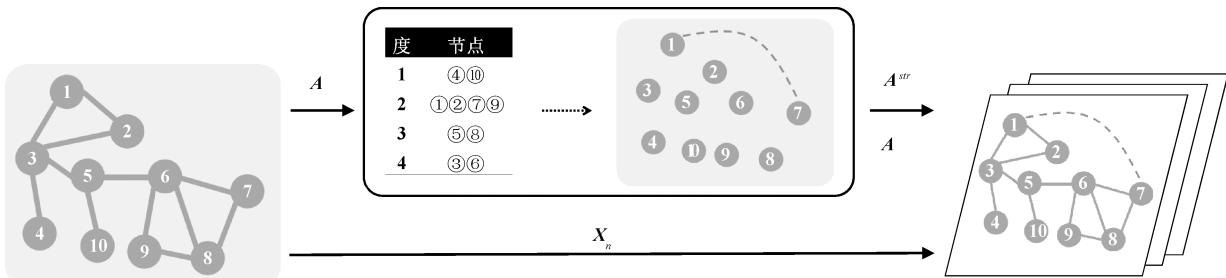


图 2 附加特征图模型整体结构(以相加混合训练为例)

3.2 结构特征图生成

根据 GCN^[2] 模型逐层聚合邻居节点信息的工作特点, 我们基于节点的度特征扩大了节点邻居的定义范围, 即允许度相同节点建立连边。定义度等邻接矩阵 \mathbf{A}_{deg} 如下式

$$\mathbf{A}_{deg} = \mathbf{D}_{oh}^T \mathbf{D}_{oh} \quad (6)$$

式中 $\mathbf{D}_{oh} \in \mathbb{R}^{N \times M}$ 由节点度值的独热(one-hot)表示构成, M 为图上节点度的最大值。

真实网络往往具有稀疏连接的特点。为了保证节点连接的稀疏性和模型的可扩展性, 我们采用随机采样的方式对邻接矩阵 \mathbf{A}_{deg} 进行稀疏化处理。稀疏化后的图称为结构特征图。令输入图为 $G = (V, E)$, 图中节点度的集合为 d , 则结构特征图为 $G^{str} = (V, E^{str})$, 其节点度的集合为 d^{str} 。建立结构特征图的候选边集合为 $E^{str'}, E^{str'}$ 中顶点度的集合为 $d^{str'}, d^{str'} \subseteq d$ 。

假设对任意 $d_n \in d$, 度为 d_n 的节点对 (v_i, v_j) 建立连边的概率为 P_{link} , $d_n \in d^{str'}$ 的概率为 P_{degree} ; 对任意 $e_{v_i v_j} \in E^{str'}$, 定义 $e_{v_i v_j} \in E^{str}$ 的概率为 $P_{link|degree}$ 。基于以上定义与假设, 可列等式如下

$$P_{link} = P_{link|degree} P_{degree} \quad (7)$$

其中 $P_{link|degree} \sim U(0,1)$, $P_{degree} \sim U(0,1)$ 。

设度为 d_n 的节点集合为 V^{d_n} , 则结构特征图生

成两层 GCN 为例, 令 $\hat{\mathbf{A}} = \widetilde{\mathbf{D}}^{-\frac{1}{2}} \widetilde{\mathbf{A}} \widetilde{\mathbf{D}}^{-\frac{1}{2}}$, 网络前向传播过程可以写作下式

$$\mathbf{Z} = \sigma(\hat{\mathbf{A}} \text{ReLU}(\hat{\mathbf{A}} \mathbf{X}_n \mathbf{W}^{(0)}) \mathbf{W}^{(1)}) \quad (5)$$

其中, σ 表示 softmax 函数。

从频域信息传播机制的角度来看, GCN 的本质是对输入信号进行逐层的拉普拉斯平滑滤波。从空间信息传播机制的角度来看, GCN 通过对输入信号左乘拉普拉斯矩阵, 引入了以目标节点为中心一阶邻居的信息。随着 GCN 层数的叠加, 邻居节点的步长范围也在逐层扩大。

成算法的伪代码如算法 1 所示。根据公式 7, 我们将生成新连边的概率拆分成两步, 首先通过 P_{degree} 将候选节点的范围缩小, 随后再从符合度要求的节点中选取节点对建立连边。这样做的目的是缩小节点对的搜索空间, 缩短程序执行时间。

算法 1. 结构特征图生成.

输入: $d, P_{degree}, P_{link|degree}, \{V^{d_n} \mid d_n \in d\}$

输出: \mathbf{A}^{str}

Initialize $\mathbf{A}^{str} = O$

FOR each $d_n \in d$ **DO**

$p_1 \leftarrow p \sim U(0,1)$

IF $p_1 < P_{degree}$ **THEN**

FOR each $v_i \in V^{d_n}$ **DO**

FOR each $v_j \in V^{d_n}$ **DO**

IF $i \neq j$ **THEN**

$p_2 \leftarrow p \sim U(0,1)$

IF $p_2 < P_{link|degree}$ **THEN**

$A_{ij}^{str} \leftarrow 1$

END ALL

3.3 组合训练方法

本文给出三种结构特征图与原图的组合训练方式: 相加混合训练、相加并行训练、独立并行训练。三种训练方式的网络结构如图 3 所示。

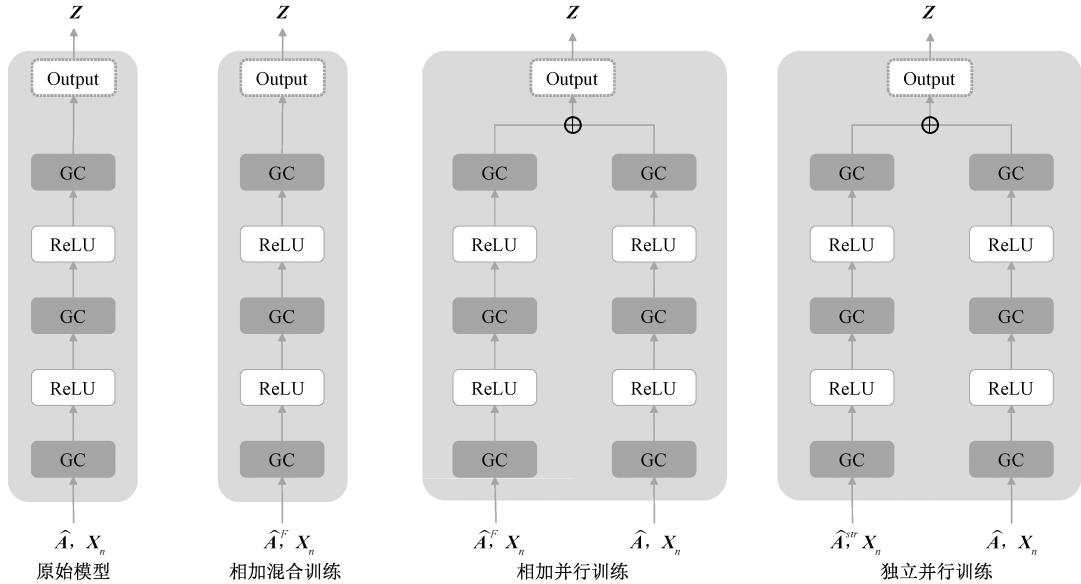


图3 附加结构特征模型训练方式对比

在两类相加训练中,我们对结构特征图做了进一步处理——生成混合特征图。利用得到的结构特征图邻接矩阵 \mathbf{A}^{str} ,与原图邻接矩阵 \mathbf{A} 做“与”运算,可以得到混合特征图邻接矩阵 \mathbf{A}^F 如下式

$$\mathbf{A}^F = \mathbf{A} \cup \mathbf{A}^{str} \quad (8)$$

相加混合训练。利用得到的混合特征图邻接矩阵 \mathbf{A}^F ,替换输入图信号的邻接矩阵 \mathbf{A} ,代入 GCN 或其改进模型的网络中进行训练。两层 GCN 网络的前向传播过程可写作下式

$$\mathbf{Z} = \sigma(\hat{\mathbf{A}}^F \text{ReLU}(\hat{\mathbf{A}}^F \mathbf{X}_n \mathbf{W}^{(0)}) \mathbf{W}^{(1)}) \quad (9)$$

这种训练方式没有改变主干模型的网络结构,模型直接在混合特征图上进行信息传递。由于原始的拓扑连接发生了改变,模型编码提取的是基于度特征重构后的拓扑结构。

独立并行训练。分别利用了原图和结构特征图。在网络中,先分别将原图与结构特征图的邻接矩阵 $\mathbf{A}, \mathbf{A}^{str}$ 代入 GCN 或其改进模型网络中并行完成训练,得到两组逐层提取出的图特征。随后在 softmax 层计算分类输出之前,对两条训练线得到的图特征相加合并。两层 GCN 网络的前向传播过程可写作下式

$$\begin{aligned} \mathbf{Z} = & \sigma(\hat{\mathbf{A}}^{str} \text{ReLU}(\hat{\mathbf{A}}^{str} \mathbf{X}_n \mathbf{W}^{(0)}) \mathbf{W}^{(1)} \\ & + \hat{\mathbf{A}} \text{ReLU}(\hat{\mathbf{A}} \mathbf{X}_n \mathbf{W}^{(0)}) \mathbf{W}^{(1)}) \end{aligned} \quad (10)$$

这种训练方式改变了主干模型的网络结构,模型分别在原图和结构特征图上进行信息传递,可以独立编码原始拓扑结构和基于度特征构建的新拓扑结构。在原图训练分支上,模型提供主干模型的原始

特征提取结果;在结构特征图训练分支上,模型能够提供基于节点度信息的特征提取结果。

相加并行训练。相加并行训练的网络结构与独立并行训练相同,也改变了主干模型的网络结构。两者的区别在于相加并行训练利用的是原图和混合特征图。两层 GCN 网络的前向传播过程如下式

$$\begin{aligned} \mathbf{Z} = & \sigma(\hat{\mathbf{A}}^F \text{ReLU}(\hat{\mathbf{A}}^F \mathbf{X}_n \mathbf{W}^{(0)}) \mathbf{W}^{(1)} \\ & + \hat{\mathbf{A}} \text{ReLU}(\hat{\mathbf{A}} \mathbf{X}_n \mathbf{W}^{(0)}) \mathbf{W}^{(1)}) \end{aligned} \quad (11)$$

相加并行训练模式下,AFG 既保留了主干模型的原始特征提取结果,又提供了原始拓扑连接发生改变后,混合特征的编码信息,可以被看作是相加混合训练和独立并行训练的结合体。本文将在 4.3 节对比三种训练方法的性能。

3.4 讨论

3.4.1 与 DropEdge 的对比

从模型的设计方法与模型在网络中的组织方式上进行对比,AFG 与 Rong 等^[17]的 DropEdge 具备相似性,即

(1) 都采用了随机增加/删除图中连边的方式来改变图的拓扑结构;

(2) 都通过在每轮训练中引入新的图拓扑结构,扩大数据集、防止模型过拟合;

(3) 都作为主干模型的附加技术进行设计,而非独立的网络结构,对 GCN 及其相关模型广泛适用。

同时,AFG 与 DropEdge 间也存在差异性:

(1) AFG 通过构建结构特征图或混合特征图

来参与网络训练,编码了图的节点结构特征信息,弥补了 GCN 及其相关模型中该特征的缺失;

(2) AFG 通过对度相同的节点对随机建立连边,缩短了信息传播距离,使原本的远距离节点对,可在极短步长范围内完成信息传播. AFG 被期望解决 GCN 及其相关模型浅层网络信息欠传播问题,提升浅层网络性能;

(3) DropEdge 通过断开连边,稀释噪声信息、缓解节点间信息的过度交换、防止模型过平滑. DropEdge 删除连边与 AFG 增加连边从功能上并不冲突,在实际操作中两者可以结合使用. 本文 4.5 节对 DropEdge 和 AFG 的互补性进行了分析.

3.4.2 连边依据

在本文中,AFG 基于度的数值比较建立附加特征图,初始候选边集由度相同节点对之间的连边构成. 对此,我们进一步讨论为何选择度相同的节点对,以及放松候选边集约束条件的可行性与必要性.

基于度相同原则构建候选边集合的合理性. 度特征是节点拓扑连接结构的体现,度相同的节点拥有相同的一阶拓扑结构. 如 2.2 节所述,GCN 本质是在邻域节点上进行拉普拉斯平滑,提取节点特征. 这种聚合方式并不能有效地编码拓扑结构特征,且信息传播范围有限. 为解决上述问题,AFG 令度相同的节点之间建立连边,构建了附加特征图. 基于附加特征图,模型可以显式地在一阶拓扑结构相同的节点之间进行信息传递,通过拉普拉斯平滑提取局部拓扑相同的节点之间的共性特征,同时又能有效地扩大节点信息传播范围,促进消息传递. 在实验 4.4 部分,我们在合成数据集和真实世界数据集中验证了这种构建方式的作用效果.

更一般性的候选边集生成机制. 本文研究的目的是通过引入附加的拓扑特征来增强浅层 GCN 模型的信息传递. 其中的关键点在于附加拓扑特征(度或其他拓扑特征)和特征比较方式(相等或相似)的选择. 从比较方式选择来说,度相等或相似主要影响候选边集的规模. 若候选边集规模较小,则无法有效改善信息传递;若候选边集规模较大,则会引入过多噪声信息. 基于不同比较方式约束产生的候选边集大小展示在表 1 中, Δ 表示度的差值阈值. $\Delta = t$ 时,选择度值差值小于等于 t 的节点对. 考虑到真实世界网络连接的稀疏性,仅度相同的节点对就已经可以提供大量的候选边,进一步松弛约束会造成候选边集规模过大. 因此本文仅仅考虑度相同作为边的生成依据. 从特征选择角度来说,除度特征以外常

见的图拓扑特征还包括聚类系数^[33]、中心性^[34-35]等. 相比较其他图拓扑特征,度特征直接反映了节点一阶拓扑结构的硬编码,是只考虑中心节点与邻居节点连接关系的结构特征. 基于度相同连边方法的实验结果证明了 GCN 及其相关模型不必拘泥于图的原始输入拓扑来进行信息传递,附加特征能够有效改善图上的信息传递效果. 我们不否认其他结构特征对于增强图上信息传递的可行性,但直接采用枚举式方法来验证其他特征的意义不大. 我们将在后续工作中探索更加一般的可学习附加特征图增强模型.

表 1 数据集统计数据

名称	Cora	Citeseer	Pubmed
#图	1	1	1
#节点	2,708	3,327	19,717
#边	5,429	4,732	44,338
#特征	1,433	3,703	500
#类别	7	6	3
$\Delta = 0$	1,137,736	2,689,130	98,689,187
$\Delta = 1$	3,107,348	5,956,394	176,778,904
$\Delta = 2$	4,580,542	7,780,972	216,674,244

4 实 验

4.1 实验设定

本文选择 Cora^[36]、Citeseer^[36]、Pubmed^[36] 三个引文数据集作为基准数据集进行节点分类任务. 数据集的具体数据呈现在表 1 中. 本文实验中数据集的划分与 Rong 等^[17] 在 DropEdge 中所用一致.

在具体实验设定方面,所有模型均使用 Adam^[37] 优化器进行优化,并使用 Dropout 方法避免过拟合. 模型中所用激活函数为 ReLU. 所有模型均从头训练,epoch 设为 500. 实验结果通过平均准确率呈现. 为了追求实验数据的真实与稳定性,我们对每组实验重复进行 50 次,取结果均值作为最终结果. 特别的,由于本文提出的 AFG 模型设计初衷是希望改善 GCN 类模型的浅层信息欠传播问题,扩大浅层 GCN 类模型的信息传播范围,提升模型特征提取能力,从而在不需要堆叠模型深度的前提下提升模型性能,因此本文实验只针对较浅层模型(2~8 层)进行.

4.2 附加特征图模型性能

4.2.1 对主干模型的性能增益

本文选择 GCN^[2]、ResGCN^[9]、SGC^[38] 作为主干模型,同时引入 DropEdge^[17] 作为对比. 实验结果选择测试集准确率最高的结果进行呈现,各结果的

具体实验参数见本文附录 C 表 10. 特别地, 附加特征图结果取三种训练方法中最优者进行呈现.

表 2 平均分类准确率对比

数据集	层数	模型								
		GCN			ResGCN			SGC		
		原始模型	+DropEdge	+AFG	原始模型	+DropEdge	+AFG	原始模型	+DropEdge	+AFG
P		0.005929			0.000408			0.000408		
Cora	2	0.8615	0.8675	0.8738	0.8615	0.8697	0.8743	0.8636	0.8701	0.8738
	4	0.8619	0.8656	0.8674	0.8627	0.8635	0.8703	0.8679	0.8715	0.8755
	6	0.8459	0.8478	0.8525	0.8438	0.8464	0.8492	0.8602	0.8654	0.8683
	8	0.8598	0.8664	0.8690	0.8302	0.8401	0.8323	0.8584	0.8591	0.8625
Citeseer	2	0.7909	0.7973	0.7955	0.7933	0.7995	0.7966	0.7903	0.7942	0.7948
	4	0.7727	0.7766	0.7769	0.7798	0.7834	0.7814	0.7799	0.7845	0.7831
	6	0.7434	0.7418	0.7644	0.7491	0.7501	0.7621	0.7651	0.7682	0.7703
	8	0.7102	0.7393	0.7433	0.7335	0.7292	0.7386	0.7601	0.7637	0.7626
Pubmed	2	0.8895	0.8894	0.8920	0.8927	0.8930	0.8952	0.8931	0.8935	0.8932
	4	0.8641	0.8646	0.8658	0.8791	0.8796	0.8888	0.8607	0.8621	0.8630
	6	0.8517	0.8534	0.8564	0.8728	0.8738	0.8895	0.8479	0.8485	0.8525
	8	0.8464	0.8461	0.8499	0.8730	0.8732	0.8784	0.8475	0.8477	0.8509

从表 2 可以看出, 本文提出的 AFG 能够对主干模型实现性能提升, 最高达到 4.66%, 对 2 层模型节点分类准确率的平均增益为 0.69%, 4 层模型的平均增益为 0.57%, 6 层模型的平均增益为 1.18%, 8 层模型的平均增益为 0.99%.

相较于 DropEdge, AFG 对三种主干模型在两个数据集上均能实现正向性能提升, 而且在模型的不同深度实现的性能提升更大, 对不同模型在不同层深下实现的性能增幅更加稳定. 尽管 AFG 的一般表现优于 DropEdge, 在部分设定情况下我们也观察到 DropEdge 对主干模型性能增益高于 AFG 的情况. 我们认为这是由于 AFG 增加连边的随机性导致新建立连边存在冗余或引入噪声导致的. 在本文第 3.4.1 节中, 我们分析了 DropEdge 与 AFG 模型的互补性. 实验结果表明一般情况下当 DropEdge 作用于结构特征图时, 模型性能增益低于 AFG 模型或 DropEdge 模型单独作用效果; 但在部分情况下 DropEdge 作用于结构特征图提升了 AFG 模型

的作用效果. 这一现象佐证了部分情况下新增连边中存在噪声连边的观点. 具体结果见附录 E 表 12~表 14.

我们采用假设检验的方式对模型性能增益的显著性进行了验证. 在显著性水平 $\alpha=0.05$ 的情况下, 假设使用 AFG 的主干模型和原始模型性能无显著差异. 我们使用方差分析计算得到: 不同主干模型下 p 值均小于 0.05, 拒绝原假设, 即使用 AFG 的主干模型性能增益显著.

对于使用 AFG 模型引起的计算时间增长, 我们在实验中对比了不同概率下 AFG 生成连边和 DropEdge 删除连边的计算时间. 由于 AFG 模型对于前向预测过程运行时间并无显著影响, 我们并不考虑模型的前向预测过程运行时长. 对不同概率运行时长结果取均值, 运行时长对比如表 3 所示, 完整结果见本文附录 A 图 9. 从结果可以看出 AFG 模型的额外计算时间与 DropEdge 维持在同一尺度, 属于可以接受的计算时间增长范围.

表 3 消融实验结果

实验分类	模型	2 层		4 层		6 层		8 层		并行结构	新连接	原始拓扑连接特征	度特征	
		准确率	增益	准确率	增益	准确率	增益	准确率	增益					
对照	GCN	0.8615	\	0.8619	\	0.8459	\	0.8328	\				✓	
	并行 GCN	0.8609	-0.07%	0.8659	0.46%	0.8420	-0.46%	0.8414	1.03%	✓			✓	
结构	GCN	相加混合	0.8738	1.43%	0.8662	0.50%	0.8525	0.78%	0.8690	4.35%			✓	✓
	+ AFG	相加并行	0.8652	0.43%	0.8674	0.64%	0.8478	0.22%	0.8677	4.19%	✓	✓	✓	✓
		独立并行	0.8744	1.50%	0.8730	1.29%	0.8451	-0.10%	0.8422	1.13%	✓	✓	✓	✓
特征	GCN ($\mathbf{A} = \mathbf{A}^{rand}$)	0.8692	0.89%	0.8677	0.67%	0.8501	0.50%	0.8464	1.63%				✓	
	GCN ($\mathbf{A} = \mathbf{A}^{str}$)	0.8681	0.77%	0.8679	0.70%	0.8348	-1.31%	0.6159	-26.04%				✓	

4.2.2 与相关模型对比

我们选择本文 2.2 中提及的部分代表性模型与附加特征图增强的 GCN 类模型进行性能对比。具体对比模型包括:采样类模型 GraphSAGE^[25];注意力模型 GAT^[24];扩大浅层信息传播范围的拓扑结构学习类模型 APPNP^[27]、MixHop^[28];引入度信息的 DEMO-Net^[30]、PNA^[32]。

如表 4 中对比结果所示,AFG 帮助主干模型性能超越了其他以提升模型特征提取能力、改善欠传播情况为目的进行设计的模型。实验结果证明,AFG 模型通过扩大浅层信息传播范围,引入结构特征,并借助随机化方法提升模型鲁棒性,综合改善浅层模型欠传播方案的有效性。

表 4 运行时间对比 单位:s

	Cora	Citeseer	Pubmed
AFG	0.0208	0.0213	0.0977
DropEdge	0.0401	0.0413	0.0518

4.3 消融实验

为证明模型各模块的必要性、分析本文 3.3 中所提出的三种训练方法,我们设计了消融实验来进行对比。消融实验选用 GCN 作为主干模型,网络层数依次设为 2、4、6、8,隐藏神经元个数取为 64,其它参数设置详见本文附录 C 表 11。在 Cora 数据集上得到的实验数据如表 5 所示,其他数据集结果见附录 B 表 8 和表 9。

表 5 相关模型性能对比

	Cora	Citeseer	Pubmed
GraphSAGE	0.822	0.714	0.871
GAT	0.874	0.786	0.895
APPNP	0.875	0.795	0.894
MixHop	0.874	0.744	0.892
PNA	0.842	0.733	0.851
DEMO-Net	0.812	0.702	0.8201
GCN+AFG	0.8738	0.7955	0.892
ResGCN+AFG	0.8743	0.7966	0.8952
SGC+AFG	0.8755	0.7948	0.8932

4.3.1 结构对比

本文 3.3 所提的三种 AFG 训练方法中,两种并行训练方法设计了并列训练的网络结构(见图 3):一条串行训练路径用原图邻接矩阵 \mathbf{A} 训练,另一条用结构特征图 \mathbf{A}^{str} (即独立并行训练)或混合特征图 \mathbf{A}^F (即相加并行训练)。为证明这两种训练方式带来的增益并不只来自并行结构,我们与两条训练路径均使用原图邻接矩阵 \mathbf{A} (即表 5 中的并行 GCN)的模型进行分类准确率对比。可以发现两种

并行训练方案分类准确率增益高于 GCN 并行产生的增益。

对比三种训练方式,可以看到在网络结构较浅时(2、4 层),并行训练方式的整体表现优于混合训练方式;当模型深度变深时,相加混合训练得到的增益变大、平均准确率最高。独立并行训练在 6 层时出现负增益,相加并行训练在各种深度下表现稳定。

对于相加混合训练和独立并行训练在浅层模型不同深度的增益性能差异,我们可以从两者信息传播路径的不同来进行分析。从信息传播路径来看,原图和结构特征图上的信息传递分别提取了邻域节点特征和拓扑近似节点特征。相加混合训练和独立并行训练的区别在于,两种特征是在信息传递阶段就进行聚合,还是在输出层进行聚合。输出层聚合(独立并行训练)虽然可以为每个节点提供拓扑近似节点特征,但丧失了信息传递范围扩大的普遍性。在结构特征图上,只有新增拓扑连接的节点对可以与远端节点进行信息传递,节点对的邻居则与远端节点没有连通路径,无法有效扩大信息传播范围。而在混合特征图上,新增拓扑连接节点的邻居通过多跳连接与远端节点建立了联系,即使邻居节点本身没有新增连边也能有效扩大信息传播范围。这种随着浅层模型层数加深,信息传播受阻与否的差异,导致了独立并行训练的增益能力退化。而相加并行训练从功能上居于两者之间,可以看作是原模型和相加混合训练 AFG 模型的折中,因此可在不同网络深度中表现稳定。

4.3.2 特 征

本文所提 AFG 为原始的 GCN 模型及其相关模型引入了度特征信息,并基于该信息建立了新的连接。为了验证度特征、原始拓扑连接特征以及建立新连接的必要性,我们将结构特征图邻接矩阵 \mathbf{A}^{str} 代入 GCN 模型,以此消除原始拓扑连接特征;将完全随机建立连边的邻接矩阵 \mathbf{A}^{rand} 代入 GCN 模型,验证不含有原始拓扑连接特征和节点度特征的新连接作用。

通过对原始 GCN,GCN+AFG,GCN($\mathbf{A}=\mathbf{A}^{rand}$),GCN($\mathbf{A}=\mathbf{A}^{str}$)的分类准确率,可以看出只有同时具备原始拓扑连接特征、度特征并建立新连边的 GCN+AFG 组合能实现最好的节点分类结果。因此,网络的信息传递既需要图的拓扑连接特征信息,也需要节点结构特征信息(本文为度特征)。本文所提 AFG 可以帮助 GCN 及其相关模型实现更好的信息传递效果。

4.4 附加特征图模型作用效果

为了分析附加特征图的作用,我们构建了节点数为 200 的随机图进行实验。受限于真实世界数据集规模对计算时长的影响,我们选择在 4.4.2 和 4.4.3 部分对 Cora、Citeseer 和 Pubmed 数据集进行分析以保证结论的可靠性。

4.4.1 度分布

本节分析了 AFG 作用于主干模型,对图的度分布产生的变化。可视化结果见图 4。各子图的横坐

标为度、纵坐标为 $P_{link|degree}$,图中颜色指示度为横坐标值时对应的节点数量,颜色越浅对应节点越多。子图(a)~(c)依次对应 P_{degree} 取值 0, 0.5, 1。从图 4 中我们可以看到,随着 $P_{link|degree}$ 、 P_{degree} 的增大,图上节点度的分布逐渐向右偏移(即节点度普遍增大),且呈现弥散分布。这说明 AFG 增加了图连接的稠密度,促进了输入图上度相同的节点间的信息传递。这种变化有利于浅层模型扩大信息传递范围,进而改善欠传播问题。

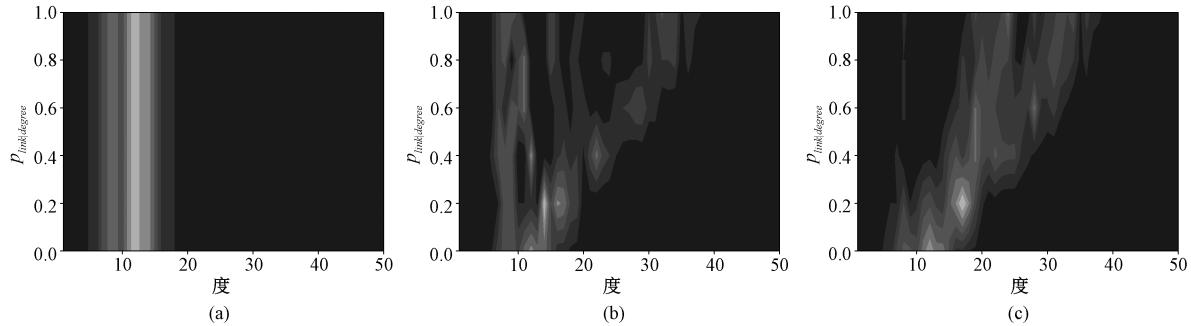


图 4 混合特征图度分布变化

4.4.2 平均路径长度

我们进一步分析了在 AFG 作用下输入图信号的平均路径长度变化,结果如图 5 所示。图中不同颜色指示在该参数组合下,使用 AFG 后输入图信号的平均路径长度。颜色越亮,平均路径长度越长。由图 5 可知,随着新增连边概率的增加,平均路径长度逐渐减小,节点间的信息传播距离得以缩短。因此 AFG 能够帮助远距离的节点对在更短的步长范围内完成信息传播,从而起到扩大浅层模型信息传播范围的作用。我们也在真实世界数据集 Cora 和 Citeseer 上对网络的平均路径长度和网络密度进行了分析,得到与合成数据集一致的结论,结果详见附录 D 图 10。

4.4.3 信息增益

我们使用 Hou 等^[39]提出的特征平滑度和标签平滑度指标来评价 AFG 模型为主干模型带来的信息增益。其中特征平滑度 λ_f 指示中央节点与一阶邻居节点的特征平均差异。 λ_f 越大,意味着中央节点可从邻居中获得的信息越多。标签平滑度 λ_l 指示邻居节点提供的信息质量。 λ_l 越大,中央节点从邻居中获得的信息质量越高。

AFG 在 Cora、Citeseer 和 Pubmed 数据集上作用前后 λ_f 和 λ_l 变化如表 6 和图 6、图 7 所示。表 6 中混合特征图数据为不同连边概率下平滑度结果的均值。图 6、图 7 展示的是不同连边概率下混合特征图

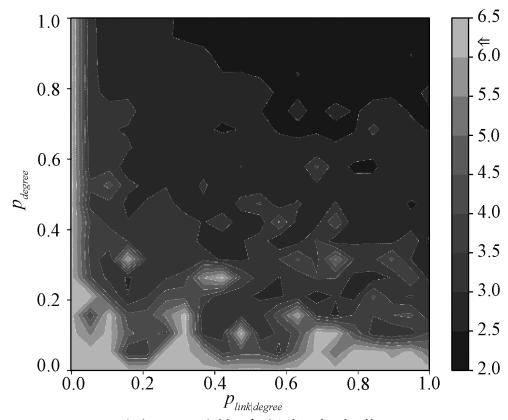


图 5 平均路径长度变化

(\Leftarrow 指示原图平均路径长度,约为 6.2)

平滑度数值的分布情况。从实验结果可以看出使用 AFG 能够极大提升原始数据集的特征平滑度和标签平滑度,即提升中央节点从邻居节点中获得的信息体量和质量。

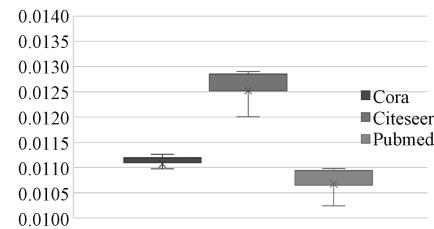


图 6 特征平滑度

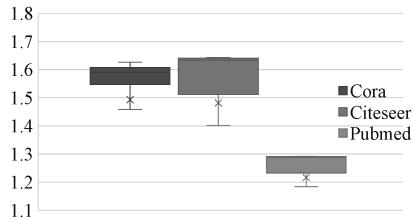


图 7 标签平滑度

表 6 平滑度对比

	原始图		混合特征图	
	λ_f	λ_l	λ_f	λ_l
Cora	0.0102	0.3801	0.0111	1.4920
Citeseer	0.0108	0.5290	0.0125	1.4800
Pubmed	0.0087	0.3952	0.0107	1.2166

4.4.4 信息传播范围

在图 8 中, 我们以黑色节点为中心, 研究使用 AFG 对浅层 GCN 模型信息传播的作用。从图中可以看出, AFG 能够帮助中心节点与更多远距离节点完成信息交换, 扩大信息传播范围。4 层 GCN + AFG 的信息传播范围已明显超过 8 层 GCN。这一实验结果验证了 AFG 能够有效缓解浅层模型信息欠传播问题。

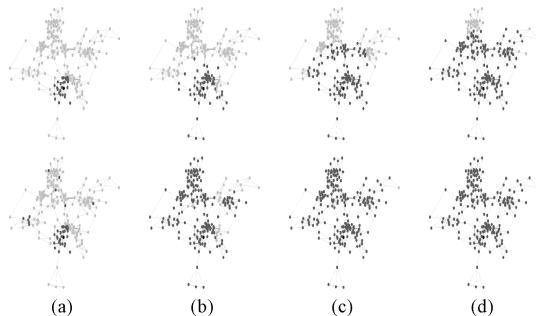


图 8 信息传递范围((a)-(d)依次对应网络层数 2、4、6、8,
第一行为 GCN, 第二行 GCN+AFG)

4.5 与 DropEdge 的互补性分析

本文 3.4.1 节讨论了理论上 DropEdge 与 AFG 之间存在的联系和差异。为了进一步分析两者在实验中的表现, 我们尝试将 AFG 与 DropEdge 同时作为附加技术与主干模型结合, 观察实验结果产生的变化。我们发现: 某些场景下, 同时使用 DropEdge 和 AFG 所产生的增益超过了两种附加技术单独对主干模型实现的增益, 甚至高于两者单独使用时产生的增益之和。也就是说, DropEdge 与 AFG 呈现了互补性。

为了深入研究这种互补现象, 本文设计了对照实验: 分别令 DropEdge 对原图、结构特征图、混合

特征图进行边的随机断连。考虑到上文实验中各种组合训练方法引入拓扑特征和扩大信息传播范围的能力, AFG 训练模式采用相加混合训练。主干模型选用 GCN, 网络层数 2 层。Dropout 概率设为 0.5, 学习率设为 0.005, 隐藏层神经元个数取为 64。在 Cora、Citeseer、Pubmed 数据集上 Adam 算法的权重下降率分别设为 0.0005、0.001 和 0.008。在 Cora、Citeseer 和 Pubmed 数据集上得到的实验数据如表 7 所示, 为节省空间在此仅列出部分参数下的结果, 完整结果见本文附录 E 表 12~表 14。

从表 7 结果可以看出, 当 DropEdge 作用于原图时, 网络的平均分类准确率对原模型分类准确率的增益最高。具体增益在 Cora 数据集上表现为 1.34%, 在 Citeseer 数据集上为 1.41%, 在 Pubmed 数据集上为 0.09%。特别的, 在 Cora 数据集上, 这一数值高于 DropEdge 和 AFG 单独作用实现的增益之和 (1.25%), 这说明两模型在 Cora 数据集的节点分类任务中出现了一定的互补效应。

当 DropEdge 作用于结构特征图时, 增益最低, 甚至低于 DropEdge 和 AFG 单独作用时的增益; 当 DropEdge 作用于混合特征图时, 增益居于作用于原图和作用于结构特征图的增益之间。基于实验结果本文给出推测, DropEdge 仅在对原图作用时作用发挥最为充分。产生这种现象的原因可能在于删除结构特征图中的连边会造成 AFG 增益效果损失。而混合特征图是原图和结构特征图的叠加, 因此当 DropEdge 作用于混合特征图时, 有一定概率删除原图连边, 也有可能得到比模型单独作用时更高的准确率。

基于上述推测, 我们认为 DropEdge 通过随机删除连边, 减少了原图中的噪声连边或噪声信息交换, 弥补了 AFG 没有对原图本身进行修正的缺失; AFG 则为 DropEdge 带来了可利用的节点结构特征信息, 并缩短了信息传播距离。两模型可通过组合实现更好的性能增益。

5 总结与展望

本文提出了针对 GCN 及其相关模型的改进技术——附加特征图模型 (AFG)。AFG 对度相同的节点对随机建立连边, 借助相加并行训练、独立并行训练或相加混合训练与主干模型结合, 从而帮助 GCN 等主干模型提升浅层网络性能。

在实验中, 我们验证了 AFG 作为 GCN 及其相关模型的附加技术, 能够缩短信息传播所需的步长、

扩大浅层网络的信息传播范围、对模型性能实现显著增益。同时,我们也在实验中发现了附加特征图模型与 DropEdge 模型的互补性。在文章最后,我们通

过实验证明在 DropEdge 仅对原图作用时,DropEdge 与附加特征图模型可以组合使用,实现大于两者单独作用产生增益之和的效果。

表 7 与 DropEdge 的互补性分析

数据集	模型	$P_{link degree}/P_{drop}$	结果					最优	增益
			0.1/0.1	0.2/0.2	0.3/0.3	0.5/0.5			
	GCN				0.8652			\	\
	+DropEdge		0.8669	0.8674	0.8675	0.8671	0.8675	0.26%	
Cora	+AFG(相加混合训练)	$P_{degree} = 0.5$	0.8692	0.8705	0.8717	0.8738	0.8738	0.99%	
	+DropEdge, AFG(DropEdge 作用原图)		0.8687	0.8713	0.8732	0.8768	0.8768	1.34%	
	+DropEdge, AFG(DropEdge 作用结构特征图)		0.8653	0.8668	0.8674	0.8671	0.8674	0.25%	
	+DropEdge, AFG(DropEdge 作用混合特征图)		0.8690	0.8707	0.8729	0.8757	0.8757	1.21%	
	GCN				0.7790			\	\
Citeseer	+DropEdge		0.7842	0.7853	0.7865	0.7891	0.7891	1.30%	
	+AFG(相加混合训练)	$P_{degree} = 0.2$	0.7885	0.7885	0.7894	0.7884	0.7894	1.33%	
	+DropEdge, AFG(DropEdge 作用原图)		0.7887	0.7893	0.7900	0.7890	0.7900	1.41%	
	+DropEdge, AFG(DropEdge 作用结构特征图)		0.7882	0.7888	0.7887	0.7878	0.7888	1.26%	
	+DropEdge, AFG(DropEdge 作用混合特征图)		0.7886	0.7883	0.7892	0.7893	0.7893	1.32%	
Pubmed	GCN				0.8895			\	\
	+DropEdge		0.8894	0.8895	0.8896	0.8897	0.8897	0.02%	
	+AFG(相加混合训练)	$P_{degree} = 0.3$	0.8901	0.8909	0.8911	0.8907	0.8911	0.18%	
	+DropEdge, AFG(DropEdge 作用原图)		0.8910	0.8920	0.8915	0.8915	0.8920	0.28%	
	+DropEdge, AFG(DropEdge 作用结构特征图)		0.8907	0.8909	0.8896	0.8904	0.8909	0.16%	
	+DropEdge, AFG(DropEdge 作用混合特征图)		0.8910	0.8904	0.8920	0.8912	0.8920	0.28%	

本文所提的附加技术 AFG 从扩大浅层 GCN 模型信息传播范围的角度,尝试解决了 GCN 模型信息传播不足、特征提取不完全的问题。然而,本文方法仅仅尝试了基于图上节点度信息随机建立连边的方案。新连边对性能增益的贡献是否存在一定的统计规律,是否可以基于这种潜在的统计规律更高效地提升模型性能等问题尚待研究。此外,相比较通过人工筛选可行的附加特征,我们计划在未来针对以度特征为代表的拓扑特征建立可学习编码,进而构建更加一般的附加特征图模型。

致谢 郑振刚受到香港大学项目(104005858, 10400599)、粤港澳联合实验室项目(2020B1212030009)等项目资助,特此表示感谢!

参 考 文 献

- [1] Zhou J, Cui G, Hu S, et al. Graph neural networks: a review of methods and applications. *AI Open*, 2020, 1:57-81
- [2] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks//Proceedings of the 5th International Conference on Learning Representations. Toulon, France, 2017: 1-14
- [3] Wu L, Sun P, Fu Y, et al. A neural influence diffusion model for social recommendation//Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York, USA, 2019: 235-244
- [4] Jin B, Gao C, He X, et al. Multi-behavior recommendation with graph convolutional networks//Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York, USA, 2020: 659-668
- [5] Wu L, Yang Y, Zhang K, et al. Joint item recommendation and attribute inference: an adaptive graph convolutional network approach//Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. New York, USA, 2020: 679-688
- [6] Zitnik M, Agrawal M, Leskovec J. Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*, 2018, 34(13): i457-i466
- [7] Liu K, Sun X, Jia L, et al. Chemi-Net: a molecular graph convolutional network for accurate drug property prediction. *International Journal of Molecular Sciences*, 2019, 20(14): 3389
- [8] Sun M, Zhao S, Gilvary C, et al. Graph convolutional networks for computational drug development and discovery. *Briefings in Bioinformatics*, 2020, 21(3): 919-935
- [9] Li G, Muller M, Thabet A, et al. DeepGCNs: Can gcns go as deep as cnns? //Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Seoul, South Korea, 2019: 9266-9275
- [10] Guo S, Lin Y, Feng N, et al. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting//Proceedings of the AAAI Conference on Artificial Intelligence, 2020: 1024-1030

- gence. Honolulu, USA, 2019: 922-929
- [11] Zhao L, Song Y, Zhang C, et al. T-GCN: A temporal graph convolutional network for traffic prediction. *IEEE Transactions on Intelligent Transportation Systems*, 2020, 21(9): 3848-3858
- [12] Chen W, Chen L, Xie Y, et al. Multi-range attentive bicomponent graph convolutional network for traffic forecasting// Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York, USA, 2020: 3529-3536
- [13] Gilmer J, Schoenholz S S, Riley P F, et al. Neural message passing for quantum chemistry//Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia, 2017: 1263-1272
- [14] Loukas A. What graph neural networks cannot learn: Depth vs width//International Conference on Learning Representations. Addis Ababa, Ethiopia, 2020: 1-17
- [15] Xu K, Li C, Tian Y, et al. Representation learning on graphs with jumping knowledge networks//Proceedings of the 35th International Conference on Machine Learning. Stockholm, Sweden, 2018: 5453-5462
- [16] Li G, Müller M, Qian G, et al. DeepGCNs: Making gcns go as deep as cnns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 1. doi: 10.1109/TPAMI.2021.3074057
- [17] Rong Y, Huang W, Xu T, et al. DropEdge: Towards deep graph convolutional networks on node classification//Proceedings of the 8th International Conference on Learning Representations. Addis Ababa, Ethiopia, 2020: 1-18
- [18] Huang W, Rong Y, Xu T, et al. Tackling over-smoothing for general graph convolutional networks. arXiv, 2021, 2008.09864
- [19] Li Q, Han Z, Wu X M. Deeper insights into graph convolutional networks for semi-supervised learning//Proceedings of the 32nd AAAI Conference on Artificial Intelligence. New Orleans, USA, 2018: 3538-3545
- [20] Yang C, Wang R, Yao S, et al. Revisiting over-smoothing in deep gcn. arXiv, 2020, 2003.13663
- [21] Alon U, Yahav E. On the bottleneck of graph neural networks and its practical implications//Proceedings of the 9th International Conference on Learning Representations. 2021: 1-16
- [22] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. //Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 770-778
- [23] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks. //Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 2261-2269
- [24] Veli? ković P, Cucurull G, Casanova A, et al. Graph attention networks//Proceedings of the 6th International Conference on Learning Representations. Vancouver, Canada, 2018: 1-12
- [25] Hamilton W L, Ying R, Leskovec J. Inductive representation learning on large graphs//Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook, USA, 2017: 1025-1035
- [26] Huang W, Zhang T, Rong Y, et al. Adaptive sampling towards fast graph representation learning//Proceedings of the 32nd International Conference on Advances in Neural Information Processing Systems. Montréal, Canada, 2018: 4563-4572
- [27] Gasteiger J, Bojchevski A, Günnemann S. Predict then propagate: graph neural networks meet personalized pagerank//Proceedings of the 6th International Conference on Learning Representations. Addis Ababa, Ethiopia, 2018: 1-15
- [28] Abu-El-Haija S, Perozzi B, Kapoor A, et al. MixHop: Higher-order graph convolutional architectures via sparsified neighborhood mixing//Proceedings of the 36th International Conference on Machine Learning. Long Beach, USA, 2019: 21-29
- [29] Klicpera J, Wei? enberger S, Günnemann S. Diffusion improves graph learning//Proceedings of the 33rd International Conference on Advances in Neural Information Processing Systems. Vancouver, Canada, 2019: 13333-13345
- [30] Wu J, He J, Xu J. DEMO-Net: degree-specific graph neural networks for node and graph classification//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York, USA, 2019: 406-415
- [31] Weisfeiler B, Lehman A A. A reduction of a graph to a canonical form and an algebra arising during this reduction. Nauchno-Technicheskaya Informatsia, 1968, Ser. 2 (N9): 12-16
- [32] Corso G, Cavalleri L, Beaini D, et al. Principal neighbourhood aggregation for graph nets//Proceedings of the 34th Advances in Neural Information Processing Systems. 2020: 13260-13271
- [33] Watts D J, Strogatz S H. Collective dynamics of ‘small-world’ networks. *Nature*, 1998, 393(6684):440-442
- [34] Sabidussi G. The centrality index of a graph. *Psychometrika*, 1966, 31 (4):581-603
- [35] Negre C F A, Morzan U N, Hendrickson H P, et al. Eigenvector centrality for characterization of protein allosteric pathways. *Proceedings of the National Academy of Sciences*, 2018, 115(52): E12201-E12208
- [36] Sen P, Namata G, Bilgic M, et al. Collective classification in network data. *AI Magazine*, 2008, 29(3):93
- [37] Kingma D P, Ba J. Adam: A Method for stochastic optimization//Proceedings of the 3rd International Conference for Learning Representations. San Diego, USA, 2015: 1-15
- [38] Wu F, Souza A, Zhang T, et al. Simplifying graph convolutional networks//Proceedings of the 36th International Conference on Machine Learning. Long Beach, USA, 2019: 6861-6871
- [39] Hou Y, Zhang J, Cheng J, et al. Measuring and improving

the use of graph information in graph neural networks//Proceedings of the 7th International Conference on Learning Rep-

resentations. New Orleans, USA, 2019: 1-16

附录 A 运行时间

为了分析添加 AFG 后主干模型运行时间变化,我们在 Cora、Citeseer、Pubmed 数据集上对 AFG 与 DropEdge 的运行时间进行了对比。由于 AFG 模型对于前向预测过程运行时间并无显著影

响,我们只考虑了每轮循环中前向过程之前的连边修改过程:AFG 生成混合特征图(结构特征图近似),DropEdge 删除连边。对比结果见图 9。子图 a-c 依次对应 Cora、Citeseer 和 Pubmed,横轴指示连边/删边概率。图例中 AFG 各项对应 p_{degree} 取值。

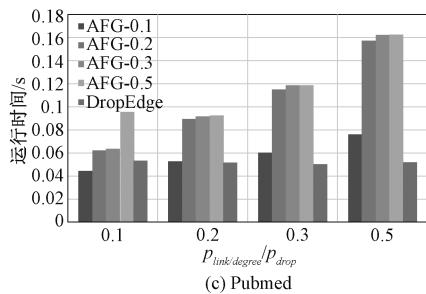
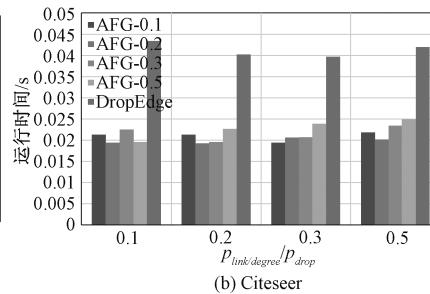
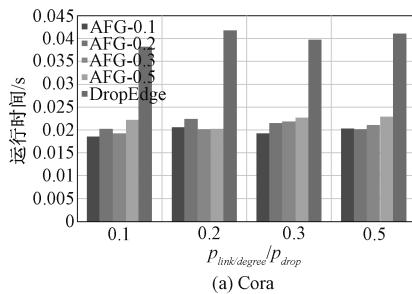


图 9 模型运行时间对比

附录 B 消融实验

正文 4.3 节中,Citeseer 上的实验数据见表 8, Pubmed 上的实验数据见 0.

附录 C 实验参数

表 2 的详细实验参数见表 10. 表 5 的详细实验参数见表 11.

附录 D 真实世界数据集 AFG 作用效果

图 10 展示了随着 $p_{link/degree} \times p_{degree}$ 数值增大,

网络平均路径长度和网络密度的变化情况。深色散点对应左侧坐标轴,指示网络平均路径长度;浅色散点对应右侧坐标轴,指示网络密度变化。

附录 E 与 DropEdge 的互补性分析

正文 4.5 节中,Cora 数据集上的完整实验结果见表 12,Citeseer 数据集上的完整实验结果见表 13, Pubmed 数据集上的完整实验结果见表 14.

表 8 Citeseer 消融实验结果

实验分类	模型	2 层		4 层		6 层		8 层		并行结构	新连接	原始拓扑连接特征	度特征
		准确率	增益	准确率	增益	准确率	增益	准确率	增益				
对照	GCN	0.7909	\	0.7727	\	0.7434	\	0.7102	\				✓
结构	并行 GCN	0.7920	0.14%	0.7481	-3.18%	0.7322	-1.51%	0.7127	0.35%	✓			✓
	GCN 相加混合	0.7925	0.20%	0.7752	0.32%	0.7644	2.82%	0.7433	4.66%		✓	✓	✓
	GCN 相加并行	0.7918	0.11%	0.7742	0.19%	0.7531	1.30%	0.7186	1.18%	✓	✓	✓	✓
	+ AFG 独立并行	0.7955	0.58%	0.7769	0.54%	0.7130	-4.09%	0.6728	-5.27%	✓	✓	✓	✓
特征	GCN ($\mathbf{A} = \mathbf{A}^{rand}$)	0.7812	-1.23%	0.7096	-8.17%	0.6880	-7.45%	0.5299	-25.39%				✓
	GCN ($\mathbf{A} = \mathbf{A}^{str}$)	0.7729	-2.28%	0.7183	-7.04%	0.5349	-28.05%	0.3590	-49.45%				✓

表 9 Pubmed 消融实验结果

实验分类	模型	2 层		4 层		6 层		8 层		并行结构	新连接	原始拓扑连接特征	度特征
		准确率	增益	准确率	增益	准确率	增益	准确率	增益				
对照	GCN	0.8895	\	0.8641	\	0.8517	\	0.8464	\				✓
结构	并行 GCN	0.8904	0.10%	0.8647	0.07%	0.8440	-0.90%	0.8427	-0.44%	✓			✓
	GCN 相加混合	0.8891	-0.04%	0.8646	0.06%	0.8564	0.55%	0.8467	0.04%		✓	✓	✓
	GCN 相加并行	0.8920	0.28%	0.8658	0.20%	0.8552	0.41%	0.8499	0.41%	✓	✓	✓	✓
	+ AFG 独立并行	0.8911	0.18%	0.8643	0.02%	0.8496	-0.25%	0.8399	-0.77%	✓	✓	✓	✓
特征	GCN ($\mathbf{A} = \mathbf{A}^{rand}$)	0.8648	-2.78%	0.8482	-1.84%	0.7417	-12.92%	0.5793	-31.56%				✓
	GCN ($\mathbf{A} = \mathbf{A}^{str}$)	0.8679	-2.43%	0.8568	-0.84%	0.7693	-9.67%	0.6705	-20.78%				✓

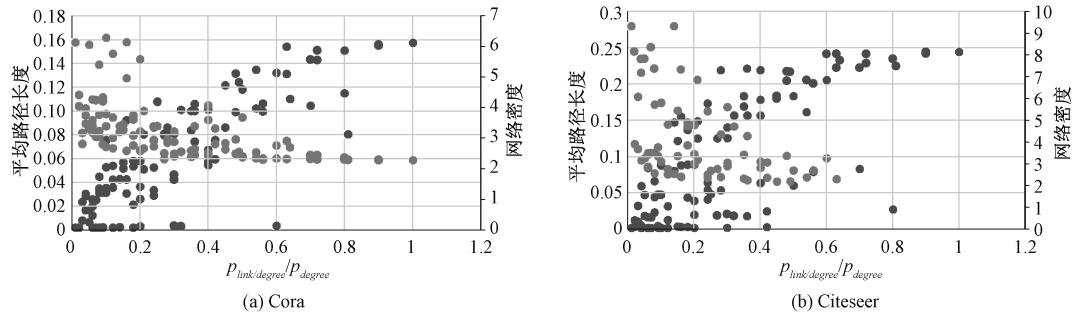


图 10 平均路径长度与网络密度变化

表 10 表 2 实验参数

模型 数据集		GCN											
		Cora				Citeseer				Pubmed			
层数		2	4	6	8	2	4	6	8	2	4	6	8
学习率		0.005	0.001	0.1	0.001	0.005	0.0008	0.001	0.01	0.001	0.01	0.01	0.01
权重下降率		5E-04	0.0009	0.0008	0.002	5E-04	0.0008	0.008	0.0001	1E-03	0.008	0.0008	0.001
隐藏单元数		64	64	64	64	64	64	64	128	64	32	32	32
dropout		0.5	0.4	0.5	0.3	0.5	0.5	0.5	0.4	0.5	0.5	0.5	0.5
P_{degree}		0.5	0.5	0.1	0.2	0.2	0.3	0.5	0.5	0.3	0.1	0.1	0.1
$P_{link degree}$		0.5	0.5	0.3	0.2	0.2	0.3	0.1	0.3	0.5	0.5	0.3	0.3
P_{drop}		0.3	0.3	0.2	0.3	0.5	0.5	0.2	0.5	0.2	0.1	0.2	0.1
训练方法		1	3	1	1	3	3	1	1	2	2	1	2
模型 数据集		ResGCN											
		Cora				Citeseer				Pubmed			
层数		2	4	6	8	2	4	6	8	2	4	6	8
学习率		0.005	0.001	0.1	0.002	0.005	0.0003	0.0001	0.0001	0.1	0.1	0.001	0.001
权重下降率		1E-03	0.001	0.0001	0.0008	1E-03	0.002	0.0001	0.004	1E-04	0.0001	0.001	0.008
隐藏单元数		64	64	64	64	128	128	128	128	64	64	64	64
dropout		0.5	0.5	0.5	0.5	0.4	0.3	0.5	0.3	0.5	0.5	0.5	0.5
P_{degree}		0.5	0.2	0.5	0.2	0.5	0.3	0.3	0.1	0.5	0.3	0.3	0.1
$P_{link degree}$		0.5	0.2	0.2	0.2	0.5	0.3	0.2	0.1	0.2	0.5	0.3	0.3
P_{drop}		0.3	0.3	0.3	0.3	0.5	0.5	0.2	0.1	0.2	0.3	0.3	0.1
训练方法		3	3	1	1	1	2	1	2	1	1	1	1
模型 数据集		SGC											
		Cora				Citeseer				Pubmed			
层数		2	4	6	8	2	4	6	8	2	4	6	8
学习率		0.005	0.0007	0.01	0.008	0.005	0.0008	0.0001	0.002	0.01	0.001	0.001	0.01
权重下降率		5E-04	0.0009	0.001	0.003	5E-04	0.0008	0.008	0.008	8E-04	0.001	0.001	0.001
隐藏单元数		32	64	128	64	64	64	128	128	64	32	32	32
dropout		0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
P_{degree}		0.5	0.1	0.1	0.3	0.3	0.1	0.5	0.2	0.1	0.5	0.2	0.1
$P_{link degree}$		0.5	0.1	0.1	0.3	0.3	0.1	0.3	0.2	0.2	0.3	0.1	0.1
P_{drop}		0.2	0.5	0.1	0.1	0.5	0.5	0.3	0.2	0.2	0.5	0.5	0.5
训练方法		1	3	1	2	3	1	1	1	2	3	2	2

注: 训练方法: 1 是相加混合; 2 是相加并行; 3 是独立并行。

表 11 表 5 实验参数

层数	2	4	8
学习率	0.005	0.001	0.001
权重下降率	0.0005	0.0009	0.002
隐藏单元数	64	64	64
dropout	0.5	0.4	0.3
P_{degree}	0.1	0.1	0.1
$P_{link degree}$	0.1	0.1	0.1
P_{drop}	0.1	0.1	0.1

表 12 表 7 Cora 数据集的完整数据

$P_{link degree} / P_{drop}$	模型	参数	结果			
			0.1/0.1	0.2/0.2	0.3/0.3	0.5/0.5
$P_{degree} = 0.1$	GCN	准确率		0.8652		
	+ DropEdge	准确率	0.8669	0.8674	0.8675	0.8671
	+ DropEdge	增益	0.19%	0.25%	0.26%	0.22%
	+ AFG(相加混合训练)	准确率	0.8669	0.8669	0.8677	0.8682
	+ DropEdge、AFG(DropEdge 作用原图)	准确率	0.8676	0.8685	0.8682	0.8703
	+ DropEdge、AFG(DropEdge 作用结构特征图)	增益	0.27%	0.38%	0.34%	0.58%
	+ DropEdge、AFG(DropEdge 作用混合特征图)	准确率	0.8647	0.8661	0.8656	0.8663
	+ DropEdge、AFG(DropEdge 作用结构特征图)	增益	-0.06%	0.09%	0.04%	0.12%
	+ DropEdge、AFG(DropEdge 作用混合特征图)	准确率	0.8667	0.8668	0.8684	0.8700
	+ DropEdge、AFG(DropEdge 作用混合特征图)	增益	0.17%	0.18%	0.37%	0.55%
$P_{degree} = 0.2$	+ AFG(相加混合训练)	准确率	0.8670	0.8684	0.8696	0.8704
	+ DropEdge、AFG(DropEdge 作用原图)	增益	0.20%	0.37%	0.50%	0.59%
	+ DropEdge、AFG(DropEdge 作用原图)	准确率	0.8681	0.8693	0.8707	0.8728
	+ DropEdge、AFG(DropEdge 作用结构特征图)	增益	0.33%	0.47%	0.63%	0.87%
	+ DropEdge、AFG(DropEdge 作用结构特征图)	准确率	0.8652	0.8657	0.8654	0.8667
	+ DropEdge、AFG(DropEdge 作用混合特征图)	增益	0.00%	0.06%	0.02%	0.17%
	+ DropEdge、AFG(DropEdge 作用混合特征图)	准确率	0.8675	0.8685	0.8704	0.8725
	+ DropEdge、AFG(DropEdge 作用混合特征图)	增益	0.26%	0.37%	0.60%	0.84%
$P_{degree} = 0.3$	+ AFG(相加混合训练)	准确率	0.8677	0.8690	0.8701	0.8719
	+ DropEdge、AFG(DropEdge 作用原图)	增益	0.28%	0.43%	0.56%	0.77%
	+ DropEdge、AFG(DropEdge 作用原图)	准确率	0.8680	0.8696	0.8709	0.8734
	+ DropEdge、AFG(DropEdge 作用结构特征图)	增益	0.32%	0.50%	0.66%	0.95%
	+ DropEdge、AFG(DropEdge 作用结构特征图)	准确率	0.8658	0.8653	0.8667	0.8664
	+ DropEdge、AFG(DropEdge 作用混合特征图)	增益	0.06%	0.01%	0.17%	0.13%
	+ DropEdge、AFG(DropEdge 作用混合特征图)	准确率	0.8679	0.8692	0.8710	0.8728
	+ DropEdge、AFG(DropEdge 作用混合特征图)	增益	0.31%	0.46%	0.67%	0.88%
$P_{degree} = 0.5$	+ AFG(相加混合训练)	准确率	0.8692	0.8705	0.8717	0.8738
	+ DropEdge、AFG(DropEdge 作用原图)	增益	0.45%	0.61%	0.75%	0.99%
	+ DropEdge、AFG(DropEdge 作用原图)	准确率	0.8687	0.8713	0.8732	0.8768
	+ DropEdge、AFG(DropEdge 作用结构特征图)	增益	0.40%	0.71%	0.92%	1.34%
	+ DropEdge、AFG(DropEdge 作用结构特征图)	准确率	0.8653	0.8668	0.8674	0.8671
	+ DropEdge、AFG(DropEdge 作用混合特征图)	增益	0.01%	0.18%	0.25%	0.21%
	+ DropEdge、AFG(DropEdge 作用混合特征图)	准确率	0.8690	0.8707	0.8729	0.8757
	+ DropEdge、AFG(DropEdge 作用混合特征图)	增益	0.44%	0.63%	0.89%	1.21%

表 13 表 7 Citeseer 数据集的完整数据

$P_{link degree} / P_{drop}$	模型	参数	结果			
			0.1/0.1	0.2/0.2	0.3/0.3	0.5/0.5
$P_{degree} = 0.1$	GCN	准确率		0.7790		
	+ DropEdge	准确率	0.7842	0.7853	0.7865	0.7891
	+ DropEdge	增益	0.66%	0.81%	0.96%	1.30%
	+ AFG(相加混合训练)	准确率	0.7873	0.7869	0.7874	0.7874
	+ DropEdge、AFG(DropEdge 作用原图)	增益	1.07%	1.01%	1.08%	1.07%
	+ DropEdge、AFG(DropEdge 作用结构特征图)	准确率	0.7881	0.7887	0.7887	0.7906
	+ DropEdge、AFG(DropEdge 作用混合特征图)	增益	1.17%	1.25%	1.24%	1.49%
	+ DropEdge	准确率	0.7868	0.7869	0.7871	0.7878
	+ DropEdge、AFG(DropEdge 作用结构特征图)	增益	1.01%	1.01%	1.04%	1.13%
	+ DropEdge、AFG(DropEdge 作用混合特征图)	准确率	0.7874	0.7874	0.7874	0.7888

续表

模型 $P_{link degree} / P_{drop}$		参数	结果			
			0.1/0.1	0.2/0.2	0.3/0.3	0.5/0.5
$P_{degree} = 0.2$	+ AFG(相加混合训练)	准确率	0.7885	0.7885	0.7894	0.7884
		增益	1.21%	1.21%	1.33%	1.21%
	+ DropEdge、AFG(DropEdge 作用原图)	准确率	0.7887	0.7893	0.7900	0.7890
		增益	1.25%	1.32%	1.41%	1.29%
	+ DropEdge、AFG(DropEdge 作用结构特征图)	准确率	0.7882	0.7888	0.7887	0.7878
		增益	1.18%	1.26%	1.25%	1.12%
	+ DropEdge、AFG(DropEdge 作用混合特征图)	准确率	0.7886	0.7883	0.7892	0.7893
		增益	1.23%	1.19%	1.30%	1.32%
	+ AFG(相加混合训练)	准确率	0.7886	0.7893	0.7893	0.7880
		增益	1.23%	1.33%	1.32%	1.15%
$P_{degree} = 0.3$	+ DropEdge、AFG(DropEdge 作用原图)	准确率	0.7898	0.7893	0.7883	0.7880
		增益	1.39%	1.32%	1.20%	1.16%
	+ DropEdge、AFG(DropEdge 作用结构特征图)	准确率	0.7882	0.7885	0.7881	0.7877
		增益	1.18%	1.21%	1.17%	1.11%
	+ DropEdge、AFG(DropEdge 作用混合特征图)	准确率	0.7889	0.7890	0.7894	0.7890
		增益	1.28%	1.28%	1.34%	1.28%
	+ AFG(相加混合训练)	准确率	0.7879	0.7868	0.7865	0.7862
		增益	1.14%	1.00%	0.96%	0.92%
	+ DropEdge、AFG(DropEdge 作用原图)	准确率	0.7871	0.7865	0.7849	0.7801
		增益	1.04%	0.96%	0.76%	0.14%
$P_{degree} = 0.5$	+ DropEdge、AFG(DropEdge 作用结构特征图)	准确率	0.7879	0.7879	0.7870	0.7865
		增益	1.14%	1.14%	1.02%	0.96%
	+ DropEdge、AFG(DropEdge 作用混合特征图)	准确率	0.7883	0.7871	0.7859	0.7854
		增益	1.20%	1.03%	0.89%	0.82%

表 14 表 7 Pubmed 数据集的完整数据

模型 $P_{link degree} / P_{drop}$		参数	结果			
			0.1/0.1	0.2/0.2	0.3/0.3	0.5/0.5
$P_{degree} = 0.1$	GCN	准确率			0.8895	
	+ DropEdge	准确率	0.8894	0.8895	0.8896	0.8897
		增益	-0.01%	0.00%	0.01%	0.02%
	+ AFG(相加混合训练)	准确率	0.8914	0.8912	0.8914	0.8911
		增益	0.21%	0.19%	0.21%	0.18%
	+ DropEdge、AFG(DropEdge 作用原图)	准确率	0.8902	0.8904	0.8906	0.8911
		增益	0.08%	0.10%	0.12%	0.18%
	+ DropEdge、AFG(DropEdge 作用结构特征图)	准确率	0.8905	0.8896	0.8890	0.8905
		增益	0.11%	0.01%	-0.06%	0.11%
	+ DropEdge、AFG(DropEdge 作用混合特征图)	准确率	0.8911	0.8906	0.8908	0.8909
$P_{degree} = 0.2$		增益	0.18%	0.12%	0.15%	0.16%
	+ AFG(相加混合训练)	准确率	0.8920	0.8910	0.8909	0.8913
		增益	0.28%	0.17%	0.16%	0.20%
	+ DropEdge、AFG(DropEdge 作用原图)	准确率	0.8910	0.8912	0.8923	0.8915
		增益	0.17%	0.19%	0.31%	0.22%
	+ DropEdge、AFG(DropEdge 作用结构特征图)	准确率	0.8902	0.8910	0.8899	0.8904
		增益	0.08%	0.17%	0.04%	0.10%
	+ DropEdge、AFG(DropEdge 作用混合特征图)	准确率	0.8910	0.8927	0.8908	0.8915
		增益	0.17%	0.36%	0.15%	0.22%

续表

模型	$P_{link degree} / P_{drop}$	$P_{degree} = 0.3$	参数		结果		
					0.1/0.1	0.2/0.2	0.3/0.3
+ AFG(相加混合训练)	+ DropEdge、AFG(DropEdge 作用原图)	$P_{degree} = 0.3$	准确率	0.8901	0.8909	0.8911	0.8907
			增益	0.07%	0.16%	0.18%	0.13%
	+ DropEdge、AFG(DropEdge 作用结构特征图)	$P_{degree} = 0.5$	准确率	0.8910	0.8920	0.8915	0.8915
			增益	0.17%	0.28%	0.22%	0.22%
+ DropEdge、AFG(DropEdge 作用混合特征图)	+ AFG(相加混合训练)	$P_{degree} = 0.3$	准确率	0.8907	0.8909	0.8896	0.8904
			增益	0.13%	0.16%	0.01%	0.10%
	+ DropEdge、AFG(DropEdge 作用结构特征图)	$P_{degree} = 0.5$	准确率	0.8910	0.8904	0.8920	0.8912
			增益	0.17%	0.10%	0.28%	0.19%
+ DropEdge、AFG(DropEdge 作用原图)	+ AFG(相加混合训练)	$P_{degree} = 0.3$	准确率	0.8912	0.8911	0.8910	0.8911
			增益	0.19%	0.18%	0.17%	0.18%
	+ DropEdge、AFG(DropEdge 作用结构特征图)	$P_{degree} = 0.5$	准确率	0.8929	0.8906	0.8929	0.8921
			增益	0.38%	0.12%	0.38%	0.30%
+ DropEdge、AFG(DropEdge 作用混合特征图)	+ AFG(相加混合训练)	$P_{degree} = 0.3$	准确率	0.8918	0.8904	0.8904	0.8909
			增益	0.26%	0.10%	0.10%	0.15%
	+ DropEdge、AFG(DropEdge 作用结构特征图)	$P_{degree} = 0.5$	准确率	0.8919	0.8928	0.8928	0.8928
			增益	0.27%	0.37%	0.37%	0.37%



SUN Jun-Shu, M. S. student. Her research interests include graph representation learning and computer vision.

WANG Shu-Hui, Ph. D., full professor. His research interests include cross-media analytics and trustable machine learning.

YANG Chen-Xue, Ph. D., assistant professor. Her research interests include machine learning, data mining and

their appl.

HUANG Qing-Ming, Ph. D., chair professor. His research interests include multimedia analysis, computer vision and pattern recognition.

Reynold C. K. Cheng, Ph. D., professor. His research interests include large-scale data management, knowledge base, social networks and crowdsourcing.

Background

Known for excellent performance and simple network structure, GCN (Graph Convolutional Networks) has attracted wide attention and become a popular and promising area. However, there are some limitations in GCN, *i.e.*, limitation in message passing and incomplete feature extraction while the network structure is shallow. To address the problem of the information propagating distance limitation, one popular approach is to build deep GCN models. However, stacking multiple layers leads to performance crushing. Popular explanations for this problem include three types: over-smoothing, over-fitting, and over-squeezing. Among all the explanations, people mainly focus on over-smoothing. Solutions include adding jump connections between layers in GCN, *e.g.*, JKNet, DeepGCN, and removing edges, *e.g.*, DropEdge. Different from stacking multiple layers, we

try to tackle this issue by broadening message passing in shallow network structures and propose AFG (Additional Feature Graph). AFG randomly connects nodes with the same degree. As a plug-and-play technique, AFG can help GCN-type networks extract more features without an increase in depth. Experiments show that AFG can improve the performance of GCN-type networks on node classification. Furthermore, we analyze and compare differences and connections between AFG and DropEdge, which is also a plug-and-play technique designed to randomly remove the edges. We prove that there is complementarity between AFG and DropEdge, which turns out these two techniques can be combined to achieve better performance on backbone models. This work was supported in part by the National Key R&D Program of China under Grant 2018AAA0102000 and in part

by the National Natural Science Foundation of China: 62022083 and 62236008, and the ICT Innovation Project: E161060. Reynold C. K. Cheng was supported by the University of Hong Kong (Projects 104005858 and 10400599) and the Guangdong-Hong Kong-Macau Joint Laboratory Pro-

gram 2020 (Project No: 2020B1212030009). It is the key research achievement of learning and reasoning with data and knowledge in the whole cross-modal/heterogeneous data analytics framework supported by these research projects.