

基于剪边策略的图残差卷积深层网络模型

毛国君,王者浩,黄 山,王 翔

(福建工程学院福建省大数据挖掘与应用重点实验室,福州 350118)

摘要: 图神经网络自2005年以来已经逐步成为图学习中的一个重要的研究分支,其中最为活跃的是图卷积神经网络. 由于图数据在现实世界中广泛存在,因此有效地完成图结构数据的学习具有很大的应用前景. 目前出现的大多数图卷积神经网络模型基本都是浅层结构,过平滑问题成为制约该领域发展的瓶颈问题. 本文提出了一种称为dri-GCN(Graph Convolutional Network via droppedge, residual and identity mapping)的图残差卷积深层网络模型,该模型集成了图剪边、初始残差和恒等映射技术. 主要思想包括:利用图剪边技术增加学习数据的多样性,以防止学习过程中的过拟合现象;构建恒等映射下的初始残差网络,来扩展残差单元的学习路径,以削弱学习过程中的过平滑问题. 实验结果表明,本文提出的dri-GCN模型可以帮助构建深层图卷积神经网络,通过网络层次的加深可以获得优于浅层网络的学习准确率.

关键词: 图神经网络; 图卷积神经网络; 剪边; 残差

中图分类号: TP391

文献标识码: A

文章编号: 0372-2112(2022)09-2205-10

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20210152

A Deep Residual Graph Convolution Network Based on Dropedge Method

MAO Guo-jun, WANG Zhe-hao, HUANG Shan, WANG Xiang

(Fujian Provincial Key Lab of Big Data Mining and Applications, Fujian University of Technology, Fuzhou, Fujian 350118, China)

Abstract: Graph neural network (GNN) has become an important research field since 2005, and the most active branch is the graph convolutional network(GCN). In the real world, many applications are directly related to graphic data, so learning from graphic structure data is becoming more and more important, and can has a huge application prospect. However, most of the existing graph convolution neural network models can only support very few convoluted layers. This is mainly because of the over smoothing problem in GCNs, which has become one of the main bottlenecks in GCNs. In this paper, a novel GCN network: dri-GCN is proposed, which integrates the technologies of the Initial Residual, Identity Mapping and DropEdge into the traditional GCN. The main contributions include: using DropEdge to increase data diversity and prevent over-fitting; constructing the residual convolutional network under the constant mapping to extend the learning paths, which can effectively weaken the over smooth problem in GCNs. Experimental results show that dri-GCN model can help building the deeper graph convolution neural networks. There is no doubt that deeper GCNs can achieve better learning accuracy than shallower networks.

Key words: graph neural network; graph convolutional neural networks; DropEdge; residual network

1 引言

图是描述对象及其相互关系的直观表达,基于图结构的图数据在现实世界中普遍存在而且具有强大的表示能力,已经成为学习的一种重要的数据来源^[1]. 图数据具有以下特征:(1)节点特征:节点自身的属性(作为观察对象需要关注的特征集).(2)结构特征:图数据

中的节点之间构成的拓扑结构(由节点之间的边来刻画). 图数据分析有着广泛的应用价值,如:图分类任务^[2]、图形图像的结构推理^[3-5]、社交网络的链路预测和节点聚类^[6-8]等.

在图数据分析领域,图神经网络(Graph Neural Networks, GNN)已经被提出,近年成为一个重要的研究分

支. 它的一个直接动力来自于深度神经网络的成功. 深度学习作为人工智能第三次浪潮的标志性成果, 已经在包括图像、语音分析应用上取得重要突破. 但是一个不容忽视的事实是: 目前最热门的卷积神经网络(Convolutional Neural Networks, CNN)仍然在传统的欧氏空间来思考和解决问题. 如图1所示, 传统的CNN把观察的对象独立当做输入来迭代分析, 忽略了对象之间的关系对分析的影响, 而近年提出的GNN则可以把对象的结构图作为输入来直接进行分析. 毋庸置疑, GNN突破了传统的图像处理的欧氏空间限制, 有着巨大潜在的应用价值.

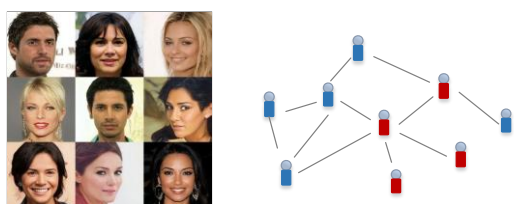


图1 CNN和GNN在图像处理上的输入数据

简言之, 相比于使用CNN, GNN特点主要有: (1) 现实世界中, 一组对象及其关系是不可分割的整体, GNN更符合人类思维模式, 应用范围会更广; (2) 在处理对象关系时, 对象及其关系作为整体来输入, 而传统的CNN没有处理关系的机制, 需要人为设计进而很难避免主观因素的影响. 当然, GNN仍在发展中, 即使是研究最早的图卷积神经网络(Graph Convolutional Network, GCN)在关键的卷积核和池化等操作的有效性上仍然面临挑战.

目前困扰图卷积神经网络的最主要问题是过平滑现象^[9]. 事实上, 最初提出的GCN模型只有2层^[10], 而且发现随着深度加深, GCN模型并不像深度CNN那样推理能力得到有效提升, 甚至效果会变差^[11,12], 这种过平滑现象已经成为GCN发展的主要障碍. 究其原因是图数据结构没有点阵图像所拥有的空间平移等恒定性质, 本质上是多层的拉普拉斯变换, 因此图中的大部分节点的特征值很容易到达某一固定平滑值. 快速的拉普拉斯平滑本身没有错, 问题是由于快速的平滑会导致调整空间变窄, 很难实施精细化地学习, 模型的性能就很难得到持续地提升.

通过加深网络层次来改善图神经网络的表达能力研究已经开始被关注^[13-15], 其中Chen等在ICML2020中提出了GCNII(Graph Convolutional Network via Initial residual and Identity mapping), 是目前已知的最具代表性的深度图神经网络模型^[13]. 简单地说, GCNII就是在恒等连接的学习中通过附加了初始输入信息, 来实现了跨层跳跃的平稳连接. 此外, 之前引入跨层跳转思想来实现深度GCN的一个比较成功的工作是JKNet(Jumping

Knowledge Networks)^[14]. JKNet模型中提出了自适应选择性聚合机制, 利用跨层的跳转链接将浅层的特征值融入到深层中, 使得每个节点可以根据多层信息融合结果来自适应实现聚合半径的动态调整, 从一定程度上有效避免过平滑带来的语义偏差. 简言之, GCNII和JKNet都在深度GCN模型上进行了有效尝试, 可以预见深度的图神经网络将是未来一个重要的研究方向.

另一方面, 图卷积本质是特征聚合问题, 所谓的学习实质上就是寻找聚合收敛的状态, 而且收敛结果只有和领域本身语义相一致才是成功的学习. 事实上, GCN中的输入是信息更丰富的图结构(不仅节点还考虑边), 相邻节点特征的重复利用加速了GCN收敛速度, 极易产生过平滑现象. 从这个角度说, 对输入的图数据进行“裁剪”可以增加图网络的泛化能力, 增强语义的可扩展性, 进而可以预防过平滑现象的过快出现. 裁剪GCN大致有3条路径: 剪输出(DropOut)^[16]、剪点(DropNode)^[17]和剪边(DropEdge)^[18]. 剪输出就是通过适度地裁剪隐含层节点的输出值(使之不到下一层), 它在普通的深度神经网络中就被提出, 而且得到广泛应用^[19]. 顾名思义, 剪点就是裁剪输入图中的节点, 而剪边就是裁剪输入图中的边. 直观上讲, 剪点比剪边力度更大, 即一个节点被裁减不仅它对应的实体不存在, 而且依附它的边也会自然消失. Rong Y等人的研究证实了剪点技术很难被控制, 可能剪边比剪点更适合于GCN网络^[18]; 对应实验也验证了在GCN中适度剪边可以延缓过平滑现象发生, 使构建深度GCN成为可能^[18].

本文的主要工作及贡献包括: 将残差卷积(residual convolution)、恒等映射(identity mapping)和剪边处理融合到图神经网络的设计中, 提出了dri_GCN深度网络模型, 并在多个标准数据集上完成了比较性实验.

2 相关工作

Gori等人最先关注图数据的处理, 在2009年提出了图神经网络模型(GNN)^[20]. 它借鉴了普通神经网络的结构来直接处理图结构数据. 这个模型可以被认为是图神经网络的开创性工作, 为后续图神经网络的发展奠定了基础. 2013年, Bruna等人利用拉普拉斯算子与傅里叶变换等技术, 解决了图数据的卷积计算问题, 提出了第一个图卷积神经网络^[21]. 2016年, Kipf和Welling运用了一阶邻阶聚合思想, 提出了GCN模型^[10], 被之后大多数研究作为基础GCN模型来使用. 当然Kipf和Welling提出的GCN模型的最优性能层数只有2层, 而且之后的许多研究都证实这个GCN模型不适合构建深层网络, 即继续加深模型深度会导致聚合效果急剧下降.

2018年, Xu等人提出了一种改进型的图神经网络模

型JKNet^[14]. JKNet利用跳跃链接和自适应选择性聚合机制,运用跳跃链接在构建深层GCN网络上进行了有效尝试. 2020年,Chen等人设计了GCNII模型^[13],它是目前为止比较成功的一个深层GCN模型. 另一个比较成功的工作来自于Rong等人^[18],他们发现对图结构数据适度进行剪边可以很好地解决GCN的过平滑问题. 事实上,通过删除输入图数据的部分边,一定程度上改变了节点间的连接方式,使节点的聚合产生了差异化,增强了训练数据的多样性,进而提升了模型的泛化能力.

近年对图数据的应用也成为了热门的研究方向. 例如:2020年,郭嘉琰等人把GCN应用到动态网络异常检测中,提出一种基于图神经网络的异常检测算法^[22];2020年,Wang等人利用GCN网络实现了人体姿态估计^[23];2020年,Yu等人利用GCN解决了社交网络中谣言检测问题^[24].

3 dri_GCN: 融入剪边操作的图残差卷积网络设计

首先明确几个GCN相关的名词和表示:

(1) 图: $G=(V,E)$,其中 V 是节点集, E 是边集. 特别地,和CNN不同,GCN的输入就是图本身.

(2) 邻接矩阵: 图 G 的邻接矩阵 $A=(a_{ij})$ 是一个方阵, $a_{ij}=1$ 意味着第 i 个节点和第 j 个节点存在边,否则为0. 很显然,无向图的邻接矩阵是一个对称矩阵.

(3) 度矩阵: 图 G 的度矩阵 $D=(d_{ij})$ 是一个对角阵,对于无向图而言, d_{ii} 就是第 i 个顶点的度,并且 $d_{ij}=0$ ($i \neq j$). 对于有向图,需要通过顶点的出度和入度来分别表示.

(4) 单位矩阵: 是一个特殊对角阵图 $I=(p_{ij}), p_{ii}=1$, 并且 $p_{ij}=0$ ($i \neq j$).

(5) 拉普拉斯矩阵: 图 G 的拉普拉斯矩阵定义为 $L=D-A$,其中 D 和 A 是 G 的度矩阵和邻接矩阵.

3.1 基于谱域的GCN的预备知识

基本的图卷积神经网络是基于谱域(Spectral domain)方法构建的,其中的核心是拉普拉斯矩阵及其计算.

图卷积神经网络在利用拉普拉斯矩阵之前需要使用了两个小技巧来改善其性能:(1)对所有节点增加自环保证节点自身和邻居节点一起参与推理,即 $A=A+I$ (本文仍记为 A);(2)进行归一化以预防梯度爆炸或者消失问题.

定义1(适用拉普拉斯矩阵)^[10] 给定增加自环的图 G ,设 G 的邻接矩阵、度矩阵及单位矩阵分别为 A, D 和 I ,则一种适用的拉普拉斯矩阵归一化方法是:

$$\hat{L} = D^{-1/2} L D^{-1/2} = D^{-1/2} D D^{-1/2} - D^{-1/2} A D^{-1/2} = I - D^{-1/2} A D^{-1/2} \quad (1)$$

定义1被称为对称归一化的拉普拉斯矩阵

(Symmetric-normalized Laplacian). 设 $\hat{A} = D^{-1/2} A D^{-1/2}$,则 \hat{A} 作为一个适用的操作单元被目前大多数GCN模型作为推演的基础(本文亦是).

定义2(拉普拉斯矩阵的傅里叶变换) 给定图 G ,设它的拉普拉斯矩阵 L 的特征值和对应的特征向量分别为 $A=(\lambda_1, \lambda_2, \dots, \lambda_n)$ ($\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$) 和 $U=(u_1, u_2, \dots, u_n)$,对 G 上的任意信号 $x \in \mathbf{R}^n$,定义它的傅立叶变换为 $\bar{x} = f(x) = U^T x$,对应逆傅立叶变换为 $x = f^{-1}(\bar{x}) = U \bar{x}$.

基于谱域分析的卷积一般是在傅里叶域上完成的,依靠定义2的傅里叶变换和逆傅里叶变换,我们就可仿照CNN的卷积方法完成图的卷积计算.

定义3(简化的图卷积)^[10] 一个简化的GCN模型的卷积公式为:

$$H^{(l+1)} = f(H^{(l)}, \hat{A}) = \sigma(\hat{A}, H^{(l)}, W^{(l)}) \quad (2)$$

其中 $\hat{A} = D^{-1/2} A D^{-1/2}$ (A 含自环), $W^{(l)}$ 是第 l 层的学习权重, $\sigma(\cdot)$ 是某个激活函数.

和普通CNN一样,卷积运算解决了输入的前向传播问题,但是还要依靠反向传播来进行误差评估. 当然,图卷积需要依据图结构进行,因此反向传播的基础是边导数.

定义4(GCN中的梯度)^[21] 给定图 G, e_{ij} 是 G 中节点 v_i 到 v_j 的一条边, $A=(a_{ij})$ 和 $D=(d_{ij})$ 分别是 G 的邻接矩阵和度矩阵. 设 $J \in \mathbf{R}^n$ 是 G 上一个 n 维向量,则 e_{ij} 在 v_i 节点上的边导数如下计算:

$$\frac{\partial J}{\partial e_{ij}} |_{i} = \sqrt{a_{ij}} (J_i / \sqrt{d_{ii}} - J_j / \sqrt{d_{jj}}) \quad (3)$$

节点相关的梯度表示为

$$\|\nabla\| = \sqrt{\sum_{e_{ij} \in G} \left(\frac{\partial J}{\partial e_{ij}}\right)^2} \quad (4)$$

3.2 残差网络与跨层连接

带残差的图卷积模型主要在一个残差单元中引入浅层到深层的跳转链接. 如图2所示,对于 $l+1$ 层而言,除了 l 层输出作为输入之外,通过增加(l 层)之前的跳转链接来防止计算偏差地蔓延,提升聚合效能.

定义5(带残差的图卷积)^[13] 一个带残差的GCN模型的基本卷积公式为:

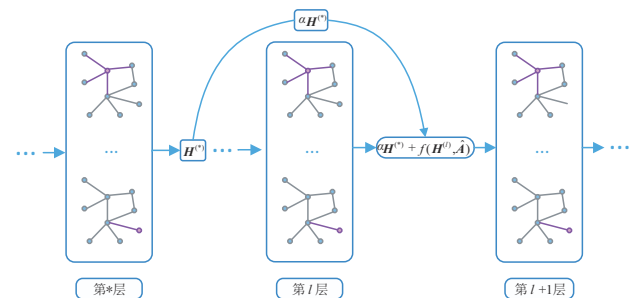


图2 一个残差块的连接结构示意图

$$\mathbf{H}^{(l+1)} = \alpha \mathbf{H}^{(l)} + f(\mathbf{H}^{(l)}, \hat{\mathbf{A}}) \quad (5)$$

其中 $\mathbf{H}^{(*)}$ 表示 * 层或者更浅层次的输出。

引入残差的初始动机是为了防止梯度消失问题^[25]。近年研究也发现:在 GCN 中引入残差进行卷积计算可以加强深层与浅层之间信息互补。此外,引入残差的卷积网络需要恒等映射机制,这不仅保证了浅层和深层信息的兼容性,而且可以使网络结构更加均匀化,更利于通过加深网络层次来提高学习精度。

图 2 给出的残差块只是个示意,浅层跳转到深层的途径可以有許多,需要和实际应用结合起来考虑。如:GCNII^[13]就是采用一种称为初始残差(Initial residual)的技术,即所有的中间层残差块都是直连初始的输入。再如:JKnet^[14]则是仅在最后一层选择之前的卷积层的输出进行融合聚集。从本质上说,通过跨层连接来改善残差卷积网的性能是目前一个探索最多的方法之一,相信未来还会继续得到关注。

图 3 给出了引入残差和恒等映射的 GCN 的工作原理示意。

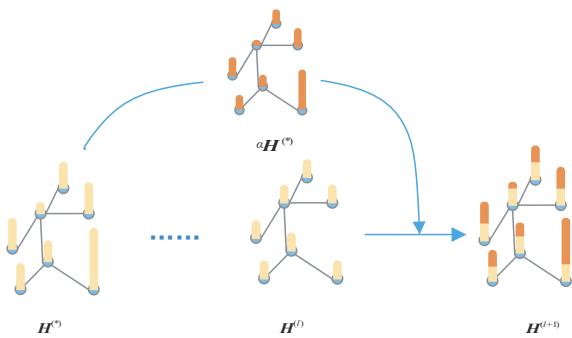


图 3 残差和恒等映射在 GCN 中工作原理示意

如图 3 所示,假如在 GCN 网络中,不使用残差连接,节点的特征值会很快趋于同质(如图 3 的 l 和 $l+1$ 层的黄色部分),即出现过平滑现象。假如加入浅层的第 * 层到深层的第 l 层的跳转通路,那么第 $l+1$ 层的节点就需要综合考虑第 l 层和第 * 层计算结果,这样就利用浅层信息修正了学习中的偏差,延缓了过平滑现象过早发生,为深层 GCN 提供有力支撑。

当然,节点特征的有效叠加,要归功于恒等映射技术。就是说,经过若干层的 GCN 推理后,恒等映射保证了各个节点处于同一量纲上,让引入残差的节点特征能够成功的叠加在到深层单元。

3.3 dri_GCIN 模型设计

集成图剪边、初始网络和恒等映射技术,结合上一节的相关定义,本文提出了一种图神经网络模型 dri_GCIN。

剪边被证明是实现深层 GCN 的一个有效技术^[18],其主要作用概括为:

(1) 剪边可以看作一种数据增强技术,在训练过程中对输入的邻接矩阵进行随机地删减,可以增加模型的输入数据的多样性。

(2) 剪边技术用于训练中,使用得当可以增强节点之间的区分度,减少信息的重复利用率,进而减缓了过平滑现象。

融入剪边和初始残差技术,图 4 给出了对应的恒等卷积网络模型 dri_GCIN 的运行框架,其中 $\hat{\mathbf{A}}_p$ 表示剪边操作后并归一化后的邻接矩阵。

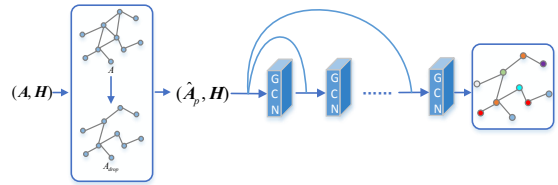


图 4 dri_GCIN 运行框架

定义 6 (dri_GCIN 的图卷积) 对于图 G , 设:剪边概率为 $1-p$, 则用 $\hat{\mathbf{A}}_p$ 表示剪边后归一化的邻接矩阵; α 和 β 为用来控制跳跃连接的两个超参数; σ 为激励函数; $\mathbf{W}^{(l)}$ 为 l 层的权重。dri_GCIN 的图卷积公式表示为:

$$\begin{aligned} \mathbf{H}^{(l+1)} &= f(\mathbf{H}^{(l)}, \hat{\mathbf{A}}) \\ &= \sigma(((1-\alpha)\hat{\mathbf{A}}_p \mathbf{H}^{(l)} + \alpha \mathbf{H}^{(0)})((1-\beta)\mathbf{I} + \beta \mathbf{W}^{(l)})) \end{aligned} \quad (6)$$

其中 $\mathbf{H}^{(l)}$ 为 l 层的输出向量, \mathbf{I} 是单位阵。

如式(6)所示, dri_GCIN 只使用初始残差 $\mathbf{H}^{(0)}$, 这主要是我们的实验发现初始残差对该模型更有效。

基于初始残差和恒等映射,融入剪边操作的 dri_GCIN 模型的第 l 层的处理步骤如下。

步骤 1 设定剪边系数 p , 按概率随机丢弃: $\hat{\mathbf{A}}_p \leftarrow \text{re-normalization}(\mathbf{A} - \mathbf{A}')$ 。

步骤 2 设定控制参数 α , 按 α 导入初始输入 $\alpha \mathbf{H}^{(0)}$ 。

步骤 3 融合 $\mathbf{H}^{(l)}$ 和 $\mathbf{H}^{(0)}$, 生成带有残差的 l 层输出: $\mathbf{H}^* \leftarrow (1-\alpha)\hat{\mathbf{A}}_p \mathbf{H}^{(l)} + \alpha \mathbf{H}^{(0)}$ 。

步骤 4 设定控制参数 β , 对权重实施恒等映射: $\mathbf{W}^* \leftarrow (1-\beta)\mathbf{I} + \beta \mathbf{W}^{(l)}$ 。

步骤 5 选定激励函数 σ , 生成层 l 输出: $\mathbf{H}^{(l+1)} \leftarrow \sigma(\mathbf{H}^*, \mathbf{W}^*)$ 。

值得注意的是, dri_GCIN 模型中参数 p 、 α 和 β 以及激励函数 σ 的设定是个经验问题, 需要根据应用背景来逐步尝试获得优化值。我们在实验中获得的经验包括:

(1) p 是图保留边的比例(剪边的比例为 $1-p$), 对应的 $\hat{\mathbf{A}}_p$ 就是正则化的输入矩阵。一般地, p 不宜过小, 过小会导致图结构性的数据不足, 丧失图网络的优势; 当然 p 也不宜过大, 过大会使迭代过程的数据多样性不足, 容易产生过平滑。我们实验的经验 p 值在 0.7~0.9 之间。

(2) 控制初始残差的参数 α 不宜过大, 一般设置在

0.1 或者 0.2 比较适中^[13]. 如式(6)所示, α 过大将直接减弱上层计算对本层计算的作用, 严重影响学习效率.

(3) 式(6)使用 $(1-\beta)I+\beta W^{(l)}$ 来替换普遍使用的权重向量 $W^{(l)}$, 这可以保证与浅层模型具有等价的性能, 有利于跨层信息聚合. 事实上, $(1-\beta)I+\beta W^{(l)}$ 是 $W^{(l)}$ 的正则化结果, 当 β 足够小时, $(1-\beta)I+\beta W^{(l)}$ 的奇异值会接近于 1^[26]. 这样就可以通过合适的 β 值设定, 来达到权重向量修正和网络等价性维护的平衡目的, 提升系统的自适应性. 我们实验中 β 设为随层数 l 变换的变量, 为 $\log(\frac{\lambda}{l} + 1)\lambda$ 为控制系数, 一般设为 0.5.

为了说明 dri_GCN 模型的基本原理, 对 2 层的 dir_GCN 网络学习过程的主要结构进行了跟踪和解析, 其中数据集使用的是标准数据集 Citeseer^[27]. 图 5 给出了 2 层 dri_GCN 网络的处理过程示意.

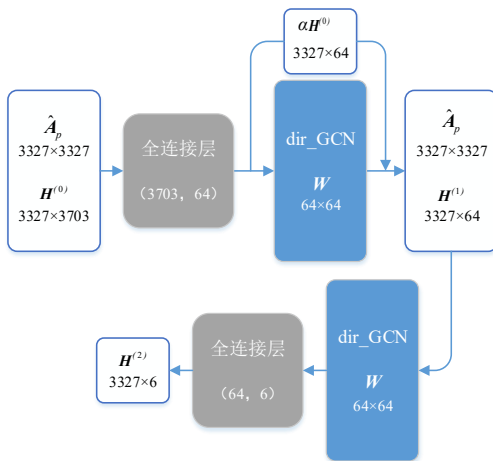


图 5 dri_GCN 网络的学习过程示意

依据图 5, 设定 $p=0.9$ 、 $\alpha=0.1$ 和 $\lambda=0.5$, 2 层 dri_GCN 网络对 Citeseer 数据集处理的基本过程是:

(1) 由 Citeseer 数据集的原始数据, 得到输入特征 $H^{(0)}$ 和剪边后归一化的邻接矩阵 \hat{A}_p (如式(7)). 值得注意的是, 剪边仅仅是将邻接矩阵部分数值置零, 并不改变邻接矩阵维度.

(2) $H^{(0)}$ 经过全连接层后特征维度变为 64, 以便与预先设置的权重矩阵 W 维度相匹配, 其中权重矩阵的值是随机生成的. 经过第一个 dri_GCN 卷积层后与初始残差 $\alpha H^{(0)}$ 叠加得到中间特征向量 $H^{(1)}$.

(3) 将更新后的 $H^{(1)}$ 和 \hat{A}_p 输入到下一个 dri_GCN 卷积层, 再经过全连接层后, 得到最后的输出特征向量 $H^{(2)}$.

式(8)给出了 100 次网络学习后最终输出的 $H^{(2)}$ 与分类结果. 式(9)中分类结果 (3 0 1 ...) 只给出了前三个节点 1、2、3 被分成的类标识“3”、“0”、“1”. 具体得到方式是: 节点对应行在 $H^{(2)}$ 中的最大值所在的位置.

$$\hat{A}_p = \begin{pmatrix} 0.153 & 0 & 0 & \dots \\ 0 & 0.2 & 0 & \dots \\ 0 & 0 & 0.5 & \dots \\ \dots & \dots & \dots & \dots \end{pmatrix} \quad (7)$$

$$H^{(2)} = \begin{pmatrix} -2.462 & -2.148 & -1.916 & \underline{-0.823} & -1.721 & -3.418 \\ \underline{-1.011} & -1.808 & -3.147 & -3.978 & -2.431 & -1.131 \\ -2.439 & \underline{-0.846} & -1.244 & -2.831 & -3.395 & -2.276 \\ \dots & \dots & \dots & \dots & \dots & \dots \end{pmatrix} \quad (8)$$

$$H^{(n)} = (3 \ 0 \ 1 \ \dots) \quad (9)$$

4 实验与分析

4.1 数据集

实验采用了标准数据集 Cora、Citeseer 和 Pubmed 进行评估和比较^[27]. 对于所有数据集, 都按照无向图进行处理, 并且仅考虑最大的连通部分, 即每个数据集都对应一个无向连通图. 表 1 给出了本文实验中的 3 种数据集的基本信息.

表 1 三个数据集的基本信息

数据集	节点数	边数	类别数	特征数
Cora	2708	5429	7	1433
Citeseer	3327	4732	6	3703
Pubmed	19717	44338	3	500

按照目前大多数论文的处理惯例和本文需求, 实验前对 3 个数据集做无向图处理, 之后使用它们评估本文模型和对比模型. 图 6 展示了 3 个数据集的部分连通图, 其中对一些节点进行了标识以直观地反映对应数据集情况.

4.2 学习准确率与分类精度的比较性实验

实验中准确率和精度由式(10)和式(11)计算得到:

$$Accuracy = \frac{TP + TN}{P + N} \quad (10)$$

$$Precision = \frac{TP}{TP + FP} \quad (11)$$

其中式(10)和式(11)对应的符号及含义如下:

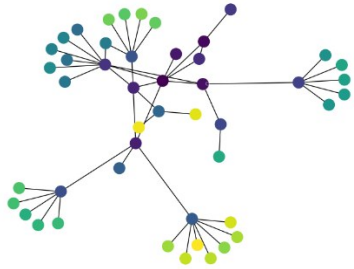
(1) P (Positive) 和 N (Negative): 分别代表训练样本中的正例和负例数目.

(2) TP(True Positives): 被正确划分为正例的数目, 即实际为正例且被模型划分为正例的样本数.

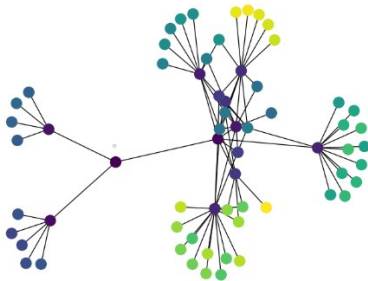
(3) TN(True negatives): 被正确划分为负例的数目, 即实际为负例且被分类器划分为负例的样本数.

(4) FP(False Positives): 被错误划分为正例的数目, 即实际为负例但被分类器划分为正例的样本数.

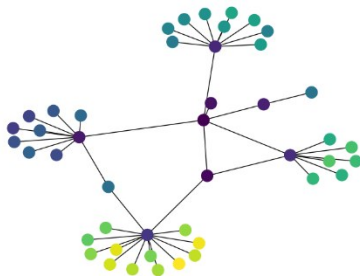
(5) FN(False Negatives): 被错误划分为负例的数目, 即实际为正例但被分类器划分为负例的样本数.



(a) Cora数据集部分无向连通图



(b) Citeseer数据集部分无向连通图



(c) Pubmed数据集部分无向连通图

图6 数据集对应的部分无向连通图

从式(10)、式(11)的计算方式可以看出:准确率主要反映一个学习方法或算法的整体学习效果;精度只关心正例集的分类情况,深入到一个具体的类别进行评价,可以更准确地反映某个类别分类的效果。

首先使用式(10),比较性地评价本文模型的准确率.表2给出了对比实验中使用的参数设置情况,其中的参数是根据数据集和模型特点,在多次实验后获得的优化值。

表2 实验模型所用的参数

数据集	学习率 lr	剪输出系数 dropout	剪边系 数 p	残差系 数 α	权重控 制系数 λ
Cora	0.01	0.6	0.9	0.1	0.5
Citeseer	0.01	0.6	0.9	0.1	0.5
Pubmed	0.01	0.6	0.98	0.1	0.5

根据本文使用的主要技术,从基本的GCN模型(简称 G)、剪边技术(简称 d)和恒等映射下的初始残差技术(简称 r)3个维度来择选比较算法:

(1) 单独考虑 G . 实验中使用的是典型算法GCN^[10];

(2) 考虑 $d+G$. 文献[18]提出了在GCN中进行剪边的理由和部分效果的展示,我们在此基础上进行了完善,并命名为 d_GCN 来用于本文比较性实验;

(3) 考虑 $r+G$. GCNII^[13]属于这类模型的典型代表。

(4) 考虑 $d+r+G$. 本文提出的 dri_GCN 模型属于该类。

按照如上4个方法类别,择选了GCN、 d_GCN 、GCNII和本文的 dri_GCN 模型,在Cora、Citeseer和Pubmed数据集上进行了对比实验.表3至表5给出了实验结果(下划线数据为获得的最高准确率),其中的准确率是100次实验获得的平均值。

表3 不同方法在Cora数据集上的分类准确率(%)

层数	模型			
	GCN (G)	d_GCN ($d+G$)	GCNII ($r+G$)	dri_GCN ($d+r+G$)
2	<u>81.1</u>	<u>82.8</u>	80.4	80.6
4	80.3	82.0	81.3	81.4
8	69.5	75.8	82.6	82.8
16	64.9	75.7	84.0	84.2
32	60.3	62.5	<u>84.5</u>	84.5
64	28.7	49.5	84.5	84.5
128	—	—	84.3	<u>84.6</u>

表4 不同方法在Citeseer数据集上的分类准确率(%)

层数	模型			
	GCN (G)	d_GCN ($d+G$)	GCNII ($r+G$)	dri_GCN ($d+r+G$)
2	<u>70.8</u>	<u>72.3</u>	65.9	66.2
4	67.6	70.6	66.1	66.5
8	30.2	61.4	69.3	70.0
16	18.3	57.2	71.7	72.0
32	25.0	41.6	<u>72.2</u>	<u>72.5</u>
64	20.0	34.4	72.1	72.2
128	—	—	72.0	72.3

表5 不同方法在Pubmed数据集上的分类准确率(%)

层数	模型			
	GCN (G)	d_GCN ($d+G$)	GCNII ($r+G$)	dri_GCN ($d+r+G$)
2	<u>79.0</u>	<u>79.6</u>	78.1	77.8
4	76.5	79.4	78.3	78.4
8	61.2	78.1	79.1	79.1
16	40.9	78.5	79.2	79.3
32	22.4	77.0	<u>79.3</u>	<u>79.7</u>
64	35.3	61.5	79.2	79.3
128	—	—	79.1	79.3

从表3~表5可以看出:

(1) 本文 dri_GCN 模型在 3 个标准数据集上最高准确率均有所提升.

(2) GCN 模型在 2 层网络上还能获得较好的学习准确率,但是随着深度加深,学习准确率会急剧下降,而且简单地使用剪边技术(对应 $d+C$ 模型)也很难支撑深度 GCN 的构建.

(3) dri_GCN 和 GCNII 模型可以支撑深层卷积网的构建,而且在 3 个标准数据集上 dri_GCN 比 GCNII 的学习准确率都有不同程度地提升.

究其原因,dri_GCN 融合了初始残差、恒等映射及剪边技术,整体上减缓了过平滑现象,使图结构的信息聚合更精细.具体地说,初始残差技术将最初的节点特

征叠加在深层模型上,阻止了节点聚合偏差的放大或累加,延长了网络的生命周期;恒等映射技术使初始残差能有效地融入到深层的卷积单元,保证网络加深后的继续学习能力;剪边技术进一步削弱了残差网络退化的可能性,通过图结构数据的变化使得残差单元的数据输入的相异度增加,避免某些节点的特征向量在相邻残差层的快速同化,因而提升了学习的精细度.

此外,依据式(11),对单个类别的分类精度进行了跟踪实验.图7给出了 dri_GCN 在 Cora 数据集上 4 个类别的精度比较实验.图7的横坐标是一次 dri_GCN 网络学习的迭代次数,纵坐标的精度值采用的是 20 次 dri_GCN 网络学习在对应迭代次数上得到平均精度值.

仔细分析图7,可以看出:

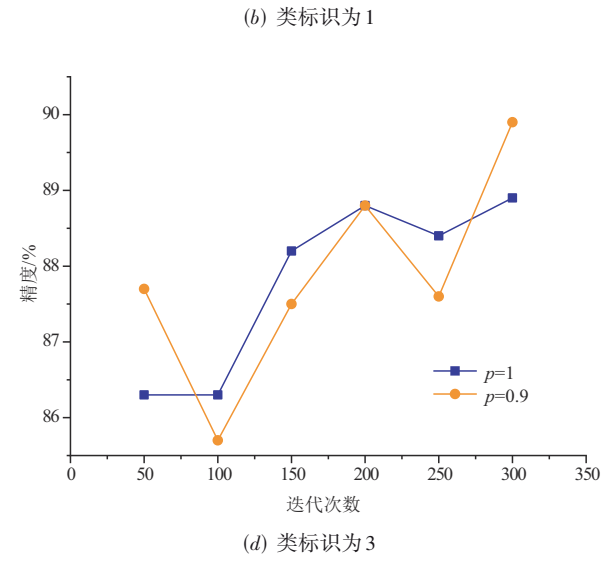
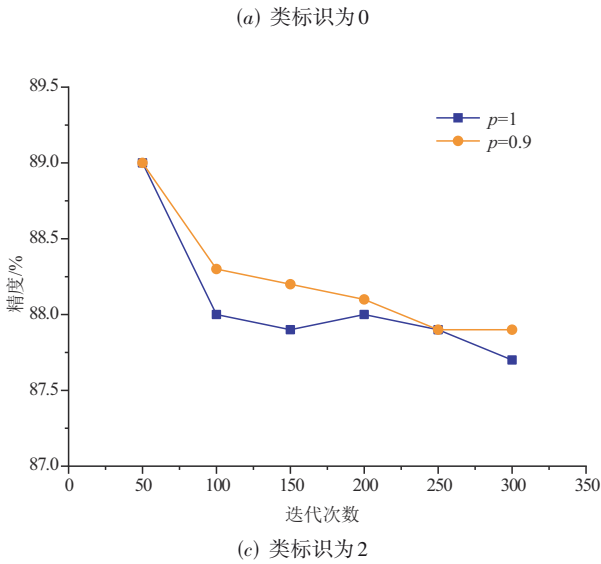
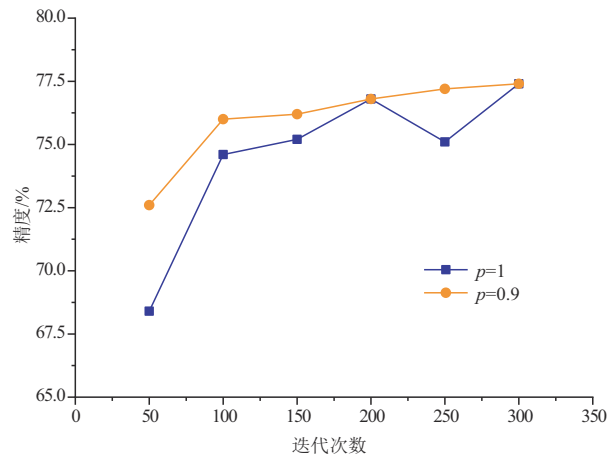
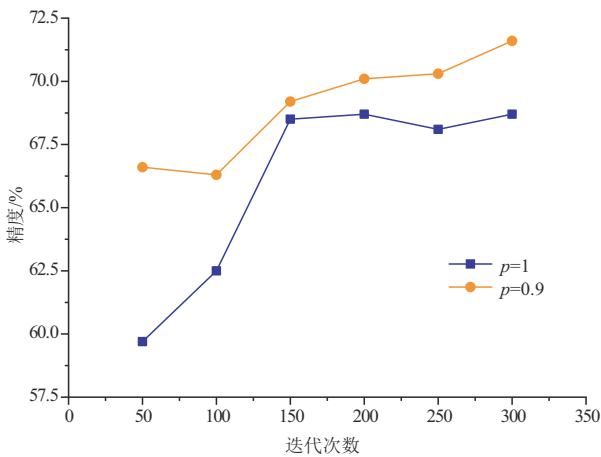


图7 dri_GCN 在 Cora 数据集的精度对比

(1) 在 Cora 数据集上,对于前面 3 个类(类标识为 0, 1, 2),剪边后的分类精度在所有迭代阶段均好于未剪边的图神经网络;

(2) 对于类标识为 3,前面的剪边效果不是很理想,但是迭代 270 次后剪边效果开始显现.可以推断,对于分类精度不佳的类来说,可以通过适当增加学习

过程中的迭代次数来提升分类精度.

4.3 剪边有效性分析

为了更详尽地说明剪边技术对模型的影响,我们通过设置不同的层数和不同 p 值,对模型 dri_GCN 的剪边效果进行了实验.表6给出了在 Cora 数据集上不同 p 值对应的学习准确率(实验的结果也是采用20次网络学习的平均值).

从表6可以看出:

(1) 针对 Cora 数据集,剪边技术在 $p=0.9$ 或 0.8 ,层

表6 Cora数据集 p 参数实验(%)

p 值	层数						
	2	4	8	16	32	64	128
$p=0.9$	80.7	81.4	83.0	84.4	84.9	84.5	84.4
$p=0.8$	81.0	81.1	82.6	84.2	84.2	84.0	83.5
$p=0.7$	80.4	81.6	82.2	83.7	83.9	83.1	83.1
$p=0.6$	81.5	79.5	80.9	82.2	82.7	81.9	82.4
$p=0.5$	81.3	79.4	80.8	81.9	82.1	81.9	82.1
$p=0.4$	80.7	76.5	79.7	81.5	81.4	81.9	81.7
$p=0.3$	80.1	75.4	79.3	79.9	80.9	81.3	80.9
$p=0.2$	80.0	72.4	79.1	79.9	80.0	80.6	80.1

数在16层以上,dri_GCN模型的准确率基本都在84%左右,说明通过设置合适的剪边系数 p ,通过加深模型层次可以获得稳定的更高的学习准确率.

(2) 当 p 在0.6以下,dri_GCN模型的准确率不高,说明剪边效果不理想.这是因为过多的剪边丢失了图数据结构,也从一个侧面说明图神经网络的科学性,即合理利用节点的关联信息能够提升节点的评估效果.

为了更直观地表达剪边在GCN上的有效性,实验跟踪了4层dri_GCN模型在3个标准数据集上的损失值.图8给出了随着迭代训练次数增多时剪边($p=0.9$)和不剪边($p=1$)的损失值变化趋势.

图8表明:融入剪边技术可以使验证集的损失值平稳而迅速地下降(相对于不剪边).说明剪边技术很好利用了数据多样性,可以及时修正迭代过程中的偏差,更有利于实现精细化学习.

图神经网络中的验证集的准确率反映了所使用模型的聚合能力.实验跟踪了3个数据集在剪边前后($p=1, p=0.9$)的验证集准确率变化.实验的准确率也是采用每组数据进行20次网络学习取平均值.实验结果如图9所示.

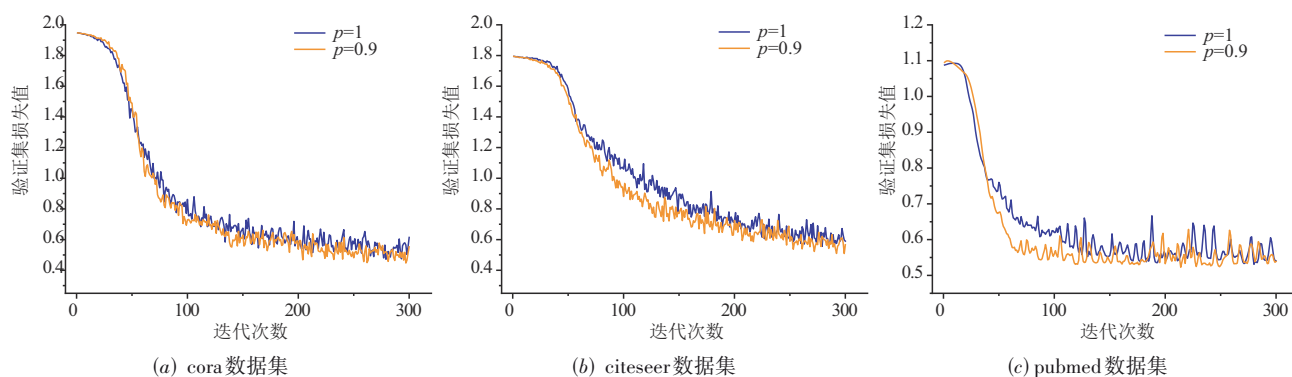


图8 dri_GCN 损失函数对比

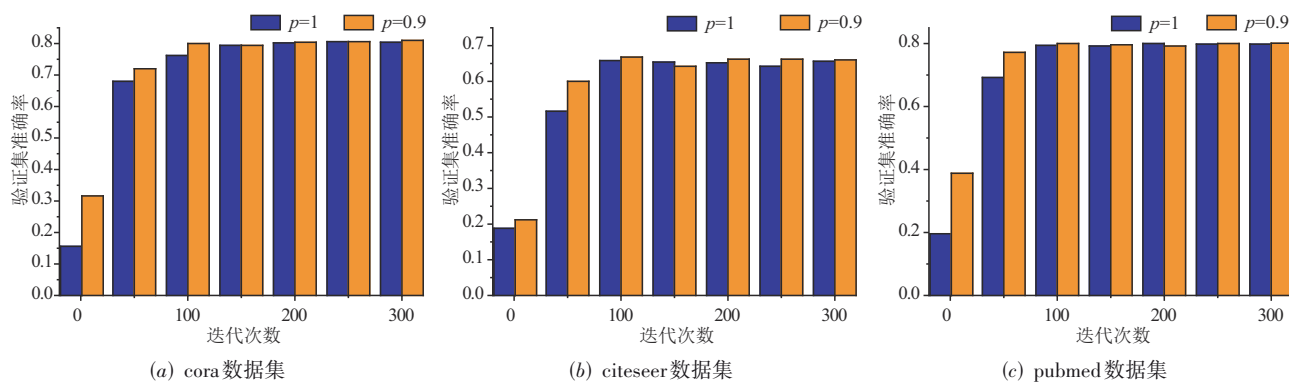


图9 dri_GCN 验证集准确率对比

从图9可以看出:在维持学习准确率的前提下,剪边可以加速聚合速度.在迭代100次之前,剪边后3个数据集的在验证集的准确率快速提升(相比未剪边),说明剪边技术的确可以提升聚合效率.此外,迭代100次之后,剪边模型的准确率也和剪边前相当,说明剪边技术不会减弱模型的学习能力.

5 总结与展望

图神经网络已经成为图学习中的一个重要而活跃的研究分支.现存的图神经网络的关键瓶颈是过平滑问题.本文集成了剪边、初始残差和恒等映射技术,提出了一种图神经网络模型dri_GCN.本文提出的模型直接面向于图神经网络研究中亟待解决的过拟合和过平滑问题,通过多种技术集成来提升图学习的性能,在深层图残差卷积的结构及其计算方法等方面进行了相关工作.实验结果表明,dri_GCN可以提升学习数据的多样性,减少训练数据的过拟合现象;可以更精细地完成基于图结构的信息聚合,有效防止过平滑现象;具备构建深层图神经网络的能力,为实现深层图学习提供一个有效途径.

本文模型为构建深层GCN网络提供新的解决途径.从结果来看,所提模型在标准数据集上的学习准确率和分类精度方面得到有效提升.未来的工作包括:本文模型在实际应用场景上的应用研究;融合其他技术进一步改进深层图神经网络性能等.

参考文献

- [1] 白铂,刘玉婷,马驰骋,等.图神经网络[J].中国科学:数学,2020,50(3):367-384.
BAI B, LIU Y T, MA C C, et al. Graph neural network[J]. Scientia Sinica (Mathematica), 2020, 50(3): 367-384. (in Chinese)
- [2] 钱榕,张茹,张克君,等.融合全局和局部特征的胶囊图神经网络[J].吉林大学学报(工学版),2021,51(3):1048-1054.
QIAN R, ZHANG R, ZHANG K J, et al. Capsule graph neural network based on global and local features fusion [J]. Journal of Jilin University (Engineering and Technology Edition), 2021, 51(3): 1048-1054. (in Chinese)
- [3] 高留杰,赵文,张君福,等.G2S:基于语义块的知识图谱问答语义解析[J].电子学报,2021,49(6):1132-1141.
GAO L J, ZHAO W, ZHANG J F, et al. G2S: semantic segment based semantic parsing for question answering over knowledge graph[J]. Acta Electronica Sinica, 2021, 49(6): 1132-1141. (in Chinese)
- [4] HAMILTON W L, YING R, LESKOVEC J. Inductive representation learning on large graphs[C]//NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook, NY: Curran Associates Inc, 2017: 1025-1035.
- [5] WANG X L, YE Y F, GUPTA A. Zero-shot recognition via semantic embeddings and knowledge graphs[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 6857-6866.
- [6] LIN W Y, JI S X, LI B C. Adversarial attacks on link prediction algorithms based on graph neural networks[C]//ASIA CCS'20: Proceedings of the 15th ACM Asia Conference on Computer and Communications Security. New York: ACM, 2020: 370-380.
- [7] 车向北,康文倩,邓彬,等.一种基于图神经网络的SDN路由性能预测模型[J].电子学报,2021,49(3):484-491.
CHE X B, KANG W Q, DENG B, et al. A prediction model of SDN routing performance based on graph neural network[J]. Acta Electronica Sinica, 2021, 49(3): 484-491. (in Chinese)
- [8] ZITNIK M, AGRAWAL M, LESKOVEC J. Modeling polypharmacy side effects with graph convolutional networks[J]. Bioinformatics(Oxford, England), 2018, 34(13): i457-i466.
- [9] PEROZZI B, AL-RFOU R, SKIENA S. DeepWalk: online learning of social representations[C]//KDD'14: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2014: 701-710.
- [10] KIPF T N, WELLING M. Semi-supervised classification with graph convolutional networks[EB/OL]. (2016-09-09). <https://arxiv.org/abs/1609.02907>.
- [11] KARPATHY A, TODERICI G, SHETTY S, et al. Large-scale video classification with convolutional neural networks[C]//CVPR'14: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus:IEEE, 2014: 1725-1732.
- [12] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence. Piscataway: IEEE, 2016, 39(4): 640-651.
- [13] CHEN M, WEI Z, HUANG Z, et al. Simple and deep graph convolutional networks[C]//International Conference on Machine Learning. Vienna: ICML, 2020: 1725-1735.
- [14] XU K, LI C, TIAN Y, et al. Representation learning on graphs with jumping knowledge networks[C]//International Conference on Machine Learning. Stockholm: ICML,

- 2018: 5453-5462.
- [15] LI G H, MÜLLER M, THABET A, et al. DeepGCNs: can GCNs go as deep as CNNs?[C]//2019 IEEE/CVF International Conference on Computer Vision(ICCV). Piscataway: IEEE, 2019: 9266-9275.
- [16] WU H B, GU X D. Towards dropout training for convolutional neural networks[J]. Neural Networks, 2015, 71: 1-10.
- [17] DO T H, NGUYEN D M, BEKOULIS G, et al. Graph convolutional neural networks with node transition probability-based message passing and DropNode regularization[J]. Expert Systems With Applications, 2021, 174: 114711.
- [18] RONG Y, HUANG W B, XU T Y, et al. DropEdge: towards deep graph convolutional networks on node classification[EB/OL]. (2019-07-25). <https://arxiv.org/abs/1907.10903>.
- [19] SRIVASTAVA N, HINTON G E, KRIZHEVSKY A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. Journal of Machine Learning Research, 2014, 15(1): 1929-1958.
- [20] SCARSELLI F, GORI M, TSOI A C, et al. The graph neural network model[J]. IEEE Transactions on Neural Networks, 2009, 20(1): 61-80.
- [21] BRUNA J, ZAREMBA W, SZLAM A, et al. Spectral networks and locally connected networks on graphs[EB/OL]. (2013-12-21). <https://arxiv.org/abs/1312.6203v3>.
- [22] 郭嘉琰, 李荣华, 张岩, 等. 基于图神经网络的动态网络异常检测算法[J]. 软件学报, 2020, 31(3): 748-762.
GUO J Y, LI R H, ZHANG Y, et al. Graph neural network based anomaly detection in dynamic networks[J]. Journal of Software, 2020, 31(3): 748-762. (in Chinese)
- [23] WANG R, HUANG C Y, WANG X Y. Global relation reasoning graph convolutional networks for human pose estimation[J]. IEEE Access, 2020, 8: 38472-38480.
- [24] YU K, JIANG H, LI T, et al. Data fusion oriented graph convolution network model for rumor detection[J]. IEEE Transactions on Network and Service Management, 2020, 17(4): 2171-2181.
- [25] HE K M, ZHANG X Y, REN S Q, et al. Identity Mappings in Deep Residual Networks[C]//European Conference on Computer Vision. Cham: Springer, 2016: 630-645.
- [26] OONO K, SUZUKI T. Graph neural networks exponentially lose expressive power for node classification[EB/OL]. (2019-05-27). <https://arxiv.org/abs/1905.10947>.

- [27] SEN P, NAMATA G, BILGIC M, et al. Collective classification in network data[J]. AI Magazine, 2008, 29(3): 93.

作者简介



毛国君 男, 1966 年月出生于内 蒙古 省 赤 峰 市. 现 为 福 建 工 程 学 院 计 算 机 与 数 学 学 院 教 授. 在 国 内 外 发 表 学 术 论 文 100 余 篇, 知 网 引 用 3000 以 上.

E-mail: 19662092@fjut.edu.cn



王者浩 男, 1995 年 3 月 出 生 于 福 建 省 福 清 市, 现 为 福 建 工 程 学 院 在 读 研 究 生. 主 要 研 究 方 向 为 电 气 工 程 人 工 智 能.

E-mail: wangzhehaofly@163.com



黄 山 男, 1987 年 5 月 出 生 于 福 建 省 武 夷 山 市. 现 为 福 建 工 程 学 院 计 算 机 科 学 与 数 学 学 院 讲 师. 研 究 方 向 为 人 工 智 能 与 深 度 学 习、固 体 发 光.

E-mail: 19872162@fjut.edu.cn



王 翔 男, 1992 年 3 月 出 生 于 福 建 省 三 明 市. 毕 业 于 北 京 大 学, 获 博 士 学 位. 现 为 福 建 工 程 学 院 计 算 机 科 学 与 数 学 学 院 副 教 授, 硕 士 生 导 师. 主 要 研 究 方 向 为 人 工 智 能, 数 字 信 号 处 理.

E-mail: wxsyhw1@fjut.edu.cn