

面向半监督节点分类的双通道图随机卷积网络

李程鸿,朱小飞

(重庆理工大学 计算机科学与工程学院,重庆 400054)

E-mail: zxf@cqut.edu.cn

摘要:图表示学习在处理图数据中起着至关重要的作用.在现实应用中,图数据中不可避免地存在噪音,然而现有基于图卷积网络的研究工作通常假设原始图数据是可靠的.此外,最近的一些研究工作发现图卷积网络不能很好地融合拓扑结构和节点特征中的信息.针对这些不足,本文提出双通道图随机卷积网络模型,用于解决图数据的半监督节点分类问题.首先,利用数据增强生成多个不同的节点特征表示,用于减小噪音数据的影响.然后使增强的节点特征表示同时在拓扑空间和特征空间中进行传播,并在传播过程中以及传播完成之后,利用注意力机制自适应地融合节点表示.在5个真实数据集上的实验结果表明该模型比基线模型具有更好的分类效果.

关键词:图数据;图表示学习;图卷积网络;节点分类;注意力机制

中图分类号: TP391

文献标识码: A

文章编号: 1000-1220(2023)08-1656-09

Dual-channel Graph Random Convolutional Networks for Semi-supervised Node Classification

LI Cheng-hong, ZHU Xiao-fei

(College of Computer Science and Engineering, Chongqing University of Technology, Chongqing 400054, China)

Abstract: Graph representation learning plays a vital role in processing graph data. In practical applications, there is inevitably noise in graph data, but the existing research work based on graph convolutional networks usually assumes that the original graph data is reliable. In addition, some recent studies have found that Graph Convolutional Networks (GCNs) can't well integrate information from topology structures and node features. To solve these problems, Dual-channel Graph Random Convolutional Network (DC-GRCN) is proposed to solve the classification problem of semi-supervised nodes of graph data. Firstly, DC-GRCN employs data augmentation to generate a number of different node features to reduce the influence of the noise data. Then the augmented node features are propagated within both the topology space and feature space at the same time. It leverages the attention mechanisms to adaptively fuse different representations of nodes during propagation and after the propagation. Experimental results on five real datasets show that this model has better classification performance than baseline model.

Key words: graph data; graph representation learning; graph convolutional network; node classification; attention mechanism

1 引言

图是描述实体之间的成对关系的基本数据结构,如社交网络、学术网络和蛋白质网络等,学习和挖掘图数据可以帮助解决各类现实应用问题.本文的研究重点是图的半监督学习问题,半监督学习的节点分类是图数据挖掘的一项重要任务,通过给定图中一小部分节点的标签,来预测图中其他节点的标签.图卷积网络(Graph Convolutional Network, GCN)通过聚合邻居节点的特征来生成节点的新表示,实现从图数据中提取有效特征.近年来图神经网络已经被广泛应用于许多领域,包括节点分类^[1-4]、图分类^[5,6]、链路预测^[7]和推荐系统^[8,9]等.

最近的一些研究^[10]发现GCN存在一些不足:1)GCN仅仅通过聚合拓扑邻居特征来进行分类任务是远远不够的,由于拓扑图中无法避免地存在噪音边,GCN不能充分融合节

点特征和拓扑结构来提取出相关的信息,影响了最终的分类性能.因此,仅通过节点特征在原始拓扑图上的传播无法学习到可以有效区分节点类别的表示.Wang等人^[1]通过计算节点特征间相似度后得到一个特征图,然后使节点特征分别在拓扑图和特征图上进行消息传递,该方法有效地利用了节点的特征信息,但由于特征图中也会存在噪音边,并且在消息传递过程中忽略了拓扑图和特征图之间的交互关系,因此该方法不能有效地突出拓扑图和特征图在消息传递过程中最有价值的信息;2)当GCN层数叠加过多后,节点的表示都倾向收敛于某一个值,即过度平滑.固定的传播结构使每个节点高度依赖于它的邻居,导致在消息传递过程中容易受到节点特征中潜在的噪音干扰,并且容易受到对抗扰动的影响^[11].数据增强技术现已应用于许多图像领域^[12,13],图像数据增强一般通过改变图像自身的属性来扩充图像,这种方法不会影响到其他图像.但是,图数据和图像数据不同,其中一个显著区

收稿日期:2021-11-12 收修改稿日期:2021-12-24 基金项目:国家自然科学基金项目(62141201)资助;重庆市技术创新与应用发展专项项目(cstc2020jcsx-dxwtBX0014)资助;重庆市教育委员会科学技术研究计划重大项目(KJZD-M202201102)资助. 作者简介:李程鸿,男,1998年生,硕士研究生,CCF会员,研究方向为图神经网络和自然语言处理;朱小飞(通讯作者),男,1979年生,博士,教授,CCF会员,研究方向为自然语言处理、数据挖掘与信息检索.

别是图数据中的节点是相互连接的,而图像是独立的,因此,图像数据增强的技术无法直接用于图数据增强上。目前,已有一些工作对图数据增强技术进行了研究。例如,Wang 等人^[14]提出通过修改节点属性和图结构的数据增强策略,对图中的每个节点独立地并行增强,即分别为每个节点用其邻居的属性来替换自身的部分属性,并添加或删除该节点与其邻居的边。但是,这种通过替换节点部分属性的方法不可避免地会导致节点自身属性信息缺失。Zhao 等人^[15]利用链路预测对原始图有策略的添加边或删除边以提升类内边数并且降低类间边数,从而提高节点分类的性能。Feng 等人^[11]提出一种随机传播的方法,使节点的特征可以随机地部分或全部丢弃,然后将扰动的特征表示在拓扑图中进行传播,然后利用一致性正则化来优化不同节点数据增强表示中未标记节点的预测一致性。这两种方法都没有充分利用节点特征中深层次的信息,因此,图数据增强技术仍有较大的研究空间。

对于图数据的半监督分类问题,为了充分挖掘特征空间中的信息,学习到多方面的节点表示,本文提出一种面向半监督节点分类的双通道图随机卷积网络(Dual-channel Graph Random Convolutional Network, DC-GRCN)。主要思想是生成多个图节点数据增强,使这多个增强的节点特征表示同时在拓扑空间和特征空间中传播。首先,为了充分利用节点特征中的信息并弥补 GCN 中节点特征仅在拓扑图上进行信息传播的不足,除原始的拓扑图外,还利用节点特征生成的 k 近邻图作为特征图,特征图从节点特征空间角度为每个节点选择与拓扑图不同的邻居节点,从而减小拓扑图中噪音边的影响。然后,为了降低噪音数据的影响,本文随机地对节点特征矩阵中的数据进行扰动,并将扰动过程重复多次,从而得到多个增强的节点特征表示。将这些增强的节点特征表示分别在拓扑空间和特征空间中同时传播,使得模型能够关注到节点不同方面的信息,从而达到数据增强的效果。通过数据增强,降低了节点对固定邻居的依赖性,由于同一个节点在不同的增强特征矩阵中学习到的节点表示不同,增加了节点表示的多样性,从而在一定程度上避免了 GCN 的过度平滑问题。此外,在拓扑图 GCN 和特征图 GCN 的传播过程中,引入了一种层级注意力机制,具体做法为,对于拓扑图和特征图的 GCN 在每个对应隐藏层输出的两种节点表示,通过注意力机制对其进行动态融合,生成的新节点表示分别作为拓扑图 GCN 和特征图 GCN 的下一层的输入。层级注意力机制有助于突出节点表示中更有价值的信息,降低噪音特征和噪音边的影响。最后,将拓扑图 GCN 和特征图 GCN 最后一层所有的输出动态地进行结合,得到最终用于分类任务的节点表示。此外,本文还设计了一个一致性约束,以保证 GCN 最后一层的多个节点表示之间保持一致性。

本文的主要贡献如下:

- 1) 提出了一种新的 GCN 框架 DC-GRCN,它更好地利用了拓扑空间和特征空间的信息,使 GCN 能够关注到节点特征不同方面的信息。
- 2) 提出了一种用于 GCN 间的信息交互方法,它可以自适应地融合多个 GCN 隐藏层中更有价值的信息。
- 3) 在 5 个基准数据集上的大量实验表明,本文提出的 DC-GRCN 在半监督节点分类任务上性能提升显著。

2 相关工作

现有的图卷积神经网络分为谱方法^[16,17]和空间方法^[3,18]两类,谱方法从谱域定义图卷积,而空间方法从节点域出发,通过定义聚合函数来聚合每个中心节点和其邻近节点^[19]。最早的趋势之一是将图的傅里叶变换应用于谱域^[16]。然后,Defferrard 等人^[17]进一步利用图拉普拉斯的切比雪夫展开来提高效率。后来,Kipf 等人^[18]简化了卷积运算,提出只聚合单跳邻居的节点特征的 GCN 模型:

$$\mathbf{Z}^{(l+1)} = \text{ReLU}(\tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}} \mathbf{Z}^{(l)} \mathbf{W}^{(l)}) \quad (1)$$

其中, $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$ 是添加自连接的无向图 \mathbf{G} 的邻接矩阵, \mathbf{I} 是单位矩阵, $\tilde{\mathbf{D}}$ 是 $\tilde{\mathbf{A}}$ 的度矩阵, $\mathbf{W}^{(l)}$ 是可训练的权重矩阵。基于空间的图卷积神经网络主要思想来源于对图像数据卷积运算的传统卷积网络,将卷积运算从图像数据推广到图形数据中,其核心思想是图传播,通过聚合节点自身的特征和它的近邻节点特征来生成节点的新的表示。

近年来,许多新的图神经网络被提出。Hamilton^[2]等人提出的 GraphSAGE 是一种能够利用节点的特征信息高效产生未知节点表示的一种归纳式学习的框架,其核心思想是学习一个对邻居节点进行聚合的函数来生成目标节点的表示。Veličković 等人^[3]提出 GAT 模型,在 GCN 的基础上引入注意力机制,在图卷积过程中为不同的邻居节点分配不同的权重。Abu-El-Haija 等人^[20]提出 MixHop 模型,针对 GCN 在图卷积过程中只能使用邻居节点的信息的缺点,提出多阶邻域的卷积,对不同的邻域采用不同的聚合方式,然后将节点表示拼接或用其他方式结合起来。虽然 MixHop 没有使用注意力机制来学习邻居的重要性,但其性能依然大幅超过 GAT。Wu 等人^[21]提出一个具体度的图神经网络 DEMO-Net,为不同度的节点使用不同的图卷积,从而保持具体度的图结构。Wang 等人^[1]根据特征与结构之间的相似性对节点标签预测是相互补充的特点,同时从节点特征、拓扑结构及其组合中学习两种特定的节点表示的和一个共同的节点表示,然后利用注意力机制学习这 3 种不同节点表示的重要性权重,最后将这些节点表示自适应地融合为用于分类任务的节点表示。Jin 等人^[10]发现图神经网络的聚集过程往往会破坏原始特征空间中的节点相似性,为了平衡图结构和节点特征的信息,提出一种特征相似性保持聚合方法,它可以自适应集成图结构和节点特征。Shi 等人^[22]利用条件对抗生成网络的特性来加强不同类别之间的区别,同时,通过最小化未标记节点在嵌入空间中的差异,约束未标记节点遵循与已标记节点相似分布。Jin 等人^[23]从局部和全局的角度将拓扑图生成两个增强图,然后使用孪生网络,通过跨视图和跨网络的对比学习损失来提高两个增强图的节点表示之前的一致性。Pei 等人^[24]从节点特征空间中找到离节点距离比较近的节点集合作为邻居节点,并且在消息传递过程中,将二维空间划分成多个区域,对不同区域的节点分块聚合,然后将这些分块信息整合,得到最终的节点表示。Suresh 等人^[25]利用原始拓扑图中的结构信息和节点的邻居信息生成一个新的图,该图增加了邻居结构等价的节点间的关联程度,并对每个节点的结构信息和邻居信息进行编码,然后在生成的新图上进行消息传递。

3 问题定义

本文主要研究图数据的半监督节点分类任务. 设 $G = (\mathbf{A}, \mathbf{X})$ 表示一个图, 其中 $\mathbf{A} \in \mathbb{R}^{n \times n}$ 是有 n 个节点的邻接矩阵, 对于其中的一个元素 $\mathbf{A}_{ij} = 1$ 表示节点 v_i 与节点 v_j 有一条边相连, 否则 $\mathbf{A}_{ij} = 0$, $\mathbf{X} \in \mathbb{R}^{n \times d}$ 是 n 个节点的特征向量, d 是特征向量的维度. 图中的每一个节点 v_i 有它的特征向量 $\mathbf{x}_i \in \mathbf{X}$ 和标签 $\mathbf{y}_i \in \mathbf{Y} \in \{0, 1\}^{n \times c}$, 其中 c 表示标签类别数. 对于半监督分类, 给定 m 个 ($0 < m \leq n$) 节点的标签 \mathbf{Y}_L , 目标是学习一个预测函数 $f: G, \mathbf{Y}_L \rightarrow \mathbf{Y}_U$, 即学习出其余未给出标签节点的标签 \mathbf{Y}_U .

4 模型

本节详细介绍了提出的模型: 双通道图随机卷积网络 (Dual-channel Graph Random Convolutional Network, DC-GRCN). 总体框架如图 1 所示, 主要思想是通过数据增强生成多个节点特征表示, 将这多个节点特征表示同时在拓扑空间和特征空间中传播. \mathbf{A}_f 是原始的拓扑图邻接矩阵, \mathbf{X} 是节点特征矩阵. 首先, 本文基于节点特征 \mathbf{X} 构造了一个特征图的邻接矩阵 \mathbf{A}_f , 然后设计了一种节点数据增强的方式, 根据节点特征矩阵 \mathbf{X} 生成 H 个节点特征增强矩阵 $\tilde{\mathbf{X}}^{(1)}, \dots, \tilde{\mathbf{X}}^{(H)}$,

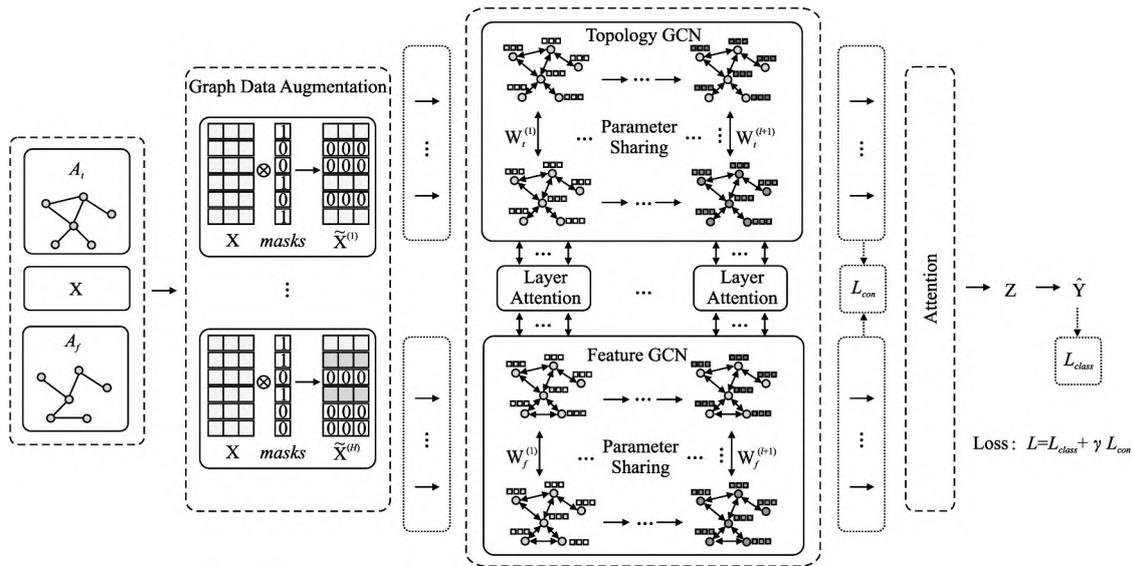


图 1 DC-GRCN 结构图

Fig. 1 Architecture of DC-GRCN

对于每一个特征增强矩阵 $\tilde{\mathbf{X}}^{(h)}$, 使其分别在拓扑图 GCN 和特征图 GCN 中进行传播. 并且, 在每一次传播中, 将拓扑图 GCN 和特征图 GCN 的每个隐藏层对应输出的两种节点表示通过层级注意力动态地融合, 生成新的节点表示分别作为下一次传播的输入. 最后, 自适应地将 GCN 最后一层所有的输出融合, 得到最终的节点表示 \mathbf{Z} 进行节点分类任务. 另外, 对 GCN 最后一层所有的输出进行一致性约束.

4.1 构建特征图

经典的 GCN 通常将邻居节点的特征信息聚合到自身来学习节点表示. 由于原始拓扑图中无法避免存在噪音边, 这种只通过原始拓扑图来学习节点表示的方法远远不能准确的区分出节点的类别. 为了解决这个问题, 自然的想法就是充分地利用节点特征信息. 因此, 基于节点的特征 \mathbf{X} , 构造出特征图 $G_f = (\mathbf{A}_f, \mathbf{X})$, 其中 $\mathbf{A}_f \in \mathbb{R}^{n \times n}$ 是特征图的邻接矩阵.

具体来说, 首先使用距离度量函数, 例如, 余弦相似度、欧氏距离等, 计算 n 个节点的相似度矩阵 \mathbf{S} . 本文统一使用余弦相似度来计算节点间的相似度, 对于节点 v_i 和 v_j , 将它们的特征向量记为 \mathbf{x}_i 和 \mathbf{x}_j , 它们的余弦相似度为:

$$\mathbf{S}_{ij} = \frac{\mathbf{x}_i \cdot \mathbf{x}_j}{\|\mathbf{x}_i\| \|\mathbf{x}_j\|} \quad (2)$$

得到 n 个节点的相似度矩阵 \mathbf{S} 后, 为每个节点选择前 k 个相似节点, 这些节点对组成特征图邻接矩阵 \mathbf{A}_f 的边.

4.2 图数据增强层

对于节点的特征矩阵 \mathbf{X} , 通过随机删除 \mathbf{X} 中的元素来生成 H 个增强的特征矩阵 $\tilde{\mathbf{X}}^{(1)}, \dots, \tilde{\mathbf{X}}^{(H)}$. 具体来说, 在训练期间, 每一次特征增强是随机地将 \mathbf{X} 中的元素按行置为 0, 具体表示为:

$$\tilde{\mathbf{x}}_i = \varepsilon_i \cdot \mathbf{x}_i \quad (3)$$

其中, ε_i 来自 Bernoulli($1-\delta$), δ 是用来控制节点特征被置 0 的概率的超参数, \mathbf{x}_i 表示 \mathbf{X} 的第 i 行.

增强的特征矩阵 $\tilde{\mathbf{X}}^{(1)}, \dots, \tilde{\mathbf{X}}^{(H)}$ 使得 GCN 在聚合过程中每个节点的特征随机地与其邻居的特征结合在一起, 降低了节点对邻居的依赖性, 能够增强模型的鲁棒性. 值得注意的是, 数据增强过程只在训练期间执行, 在推理过程中, 不进行数据增强, 即输入到 GCN 的节点特征为原始的节点特征矩阵 \mathbf{X} .

4.3 双通道图卷积网络

双通道图随机卷积网络由 3 部分组成: 拓扑图 GCN, 特征图 GCN 和层级注意力. 首先将 H 个增强的特征矩阵 $\tilde{\mathbf{X}}^{(1)}, \dots, \tilde{\mathbf{X}}^{(H)}$ 分别输入到拓扑图 GCN 和特征图 GCN. 对于特征图 GCN, 输入图为 $G_f = (\mathbf{A}_f, \mathbf{X})$. 第 h 个增强的特征矩阵 $\tilde{\mathbf{X}}^{(h)}$, 在特征图 GCN 第 l 层的输出可以表示为:

$$\mathbf{Z}_f^{(h)(l)} = \text{ReLU}(\tilde{\mathbf{D}}_f^{-1/2} \tilde{\mathbf{A}}_f \tilde{\mathbf{D}}_f^{-1/2} \tilde{\mathbf{Z}}^{(h)(l-1)} \mathbf{W}_f^{(l)}) \quad (4)$$

其中, ReLU 是激活函数, $\tilde{\mathbf{A}}_f = \mathbf{A}_f + \mathbf{I}_f$, \mathbf{I}_f 是单位矩阵, $\tilde{\mathbf{D}}_f$ 是 $\tilde{\mathbf{A}}_f$ 的度矩阵, $\mathbf{W}_f^{(l)}$ 是特征图 GCN 第 l 层的权重矩阵, $\tilde{\mathbf{Z}}^{(h)(l-1)}$ 是

经过层级注意力更新后得到的节点表示(公式(5)).另外,为了共享信息,特征图 GCN 的每一层各节点特征表示共享相同的权重矩阵 $\mathbf{W}_f^{(l)}$.

对于拓扑图 GCN,输入图为 $G_t = (\mathbf{A}_t, \mathbf{X})$,其中 $\mathbf{A}_t = \mathbf{A}$ 为原始拓扑图.同样地,第 h 个增强的特征矩阵 $\tilde{\mathbf{X}}^{(h)}$,在拓扑图 GCN 第 l 层的输出表示为: $\mathbf{Z}_t^{(h)(l)}$.与特征图 GCN 一样,拓扑图 GCN 的每一层各节点特征表示共享相同的权重矩阵 $\mathbf{W}_t^{(l)}$.

通过上面的方法,可以分别在特征空间和拓扑空间上学习到多种节点表示.为了结合这些表示,并突出更有价值的信息^[26],本文设计了一种将拓扑图 GCN 和特征图 GCN 的隐藏层中对应的节点表示自适应地结合的方法,从而得到更有效的节点表示:

$$\tilde{\mathbf{Z}}^{(h)(l)} = \tilde{\alpha}_t^{(h)(l)} \mathbf{Z}_t^{(h)(l)} + \tilde{\alpha}_f^{(h)(l)} \mathbf{Z}_f^{(h)(l)} \quad (5)$$

其中, $\tilde{\alpha}_t^{(h)(l)}$ 和 $\tilde{\alpha}_f^{(h)(l)}$ 是通过注意力机制得到的注意力权重对角矩阵(公式(9)).

为了计算层级注意力,首先,使用注意力机制学习对应节点表示相应的重要性:

$$(\alpha_t^{(h)(l)}, \alpha_f^{(h)(l)}) = \text{att}(\mathbf{Z}_t^{(h)(l)}, \mathbf{Z}_f^{(h)(l)}) \quad (6)$$

其中 $\alpha_t^{(h)(l)}, \alpha_f^{(h)(l)} \in \mathbb{R}^{n \times 1}$ 分别表示 $\mathbf{Z}_t^{(h)(l)}$ 和 $\mathbf{Z}_f^{(h)(l)}$ 的 n 个节点的注意力值.

关注于节点 v_i ,它在 $\mathbf{Z}_f^{(h)(l)}$ 中的表示为 $z_{fi} \in \mathbb{R}^{1 \times m}$.首先得到注意力值 w_{fi} :

$$w_{fi} = \mathbf{v}^T \cdot \tanh(\mathbf{W}_1 \cdot (\mathbf{z}_{fi})^T + \mathbf{b}_1) \quad (7)$$

其中, $\mathbf{W}_1 \in \mathbb{R}^{m \times m'}$ 是权重矩阵, $\mathbf{b}_1 \in \mathbb{R}^{m' \times 1}$ 是偏置向量, $\mathbf{v} \in \mathbb{R}^{m' \times 1}$ 是一个共享的注意力向量.同样地, \mathbf{z}_{ti} 的注意力值为 w_{ti} .然后,使用 *softmax* 函数对 w_{fi} 和 w_{ti} 进行归一化,得到最终的权重:

$$\alpha_{fi} = \text{softmax}(w_{fi}) \quad (8)$$

其中, α_{fi} 的值越大,表明对应的节点表示越重要.同样地, $\alpha_{ti} = \text{softmax}(w_{ti})$.对于第 h 个增强的特征矩阵在第 l 层的所有 n 个节点,得到 $\alpha_t^{(h)(l)} = [\alpha_{t1}, \dots, \alpha_{tn}]$ 和 $\alpha_f^{(h)(l)} = [\alpha_{f1}, \dots, \alpha_{fn}]$.最后,注意力权重对角矩阵为:

$$\begin{aligned} \tilde{\alpha}_t^{(h)(l)} &= \text{diag}(\alpha_t^{(h)(l)}) \\ \tilde{\alpha}_f^{(h)(l)} &= \text{diag}(\alpha_f^{(h)(l)}) \end{aligned} \quad (9)$$

其中 $\tilde{\alpha}_t^{(h)(l)}, \tilde{\alpha}_f^{(h)(l)} \in \mathbb{R}^{n \times n}$.

H 个增强的节点特征矩阵 $\tilde{\mathbf{X}}^{(1)}, \dots, \tilde{\mathbf{X}}^{(H)}$ 分别与两种输入图 G_t 和 G_f 进行图卷积操作,拓扑图 GCN 和特征图 GCN 的输出得到 $2H$ 个节点表示 $\mathbf{Z}_t^{(1)}, \dots, \mathbf{Z}_t^{(H)}, \mathbf{Z}_f^{(1)}, \dots, \mathbf{Z}_f^{(H)}$.

4.4 注意力机制

得到 $2H$ 个节点表示 $\mathbf{Z}_t^{(1)}, \dots, \mathbf{Z}_t^{(H)}, \mathbf{Z}_f^{(1)}, \dots, \mathbf{Z}_f^{(H)}$ 后,首先,使用注意力机制来学习他们相应的重要性:

$$(\alpha_t^{(1)}, \dots, \alpha_t^{(H)}, \alpha_f^{(1)}, \dots, \alpha_f^{(H)}) = \text{att}(\mathbf{Z}_t^{(1)}, \dots, \mathbf{Z}_t^{(H)}, \mathbf{Z}_f^{(1)}, \dots, \mathbf{Z}_f^{(H)}) \quad (10)$$

然后,与层级注意力一样,计算出注意力权重对角矩阵 $\tilde{\alpha}_t^{(1)}, \dots, \tilde{\alpha}_t^{(H)}, \tilde{\alpha}_f^{(1)}, \dots, \tilde{\alpha}_f^{(H)}$.最后,通过注意力权重对角矩阵来结合这 $2H$ 个节点表示,得到最终的节点表示 \mathbf{Z} :

$$\mathbf{Z} = \sum_{h=1}^H (\tilde{\alpha}_t^{(h)} \mathbf{Z}_t^{(h)} + \tilde{\alpha}_f^{(h)} \mathbf{Z}_f^{(h)}) \quad (11)$$

4.5 损失函数

根据最终的节点表示 \mathbf{Z} ,可以得到 n 个节点的预测 $\hat{\mathbf{Y}}$:

$$\hat{\mathbf{Y}} = \text{softmax}(\mathbf{Z} \cdot \mathbf{W}_2 + \mathbf{b}_2) \quad (12)$$

其中 $\hat{\mathbf{Y}} \in \mathbb{R}^{n \times c}$, c 的大小为节点的类别数.然后由预测的标签 $\hat{\mathbf{Y}}$ 与真实标签 \mathbf{Y} 计算交叉熵损失 L_{class} :

$$\mathcal{L}_{class} = - \sum_{l \in L} \sum_{i=1}^c \mathbf{Y}_{li} \ln \hat{\mathbf{Y}}_{li} \quad (13)$$

其中 L 是训练集中有标签的数据.

另外,尽管在特征传播过程中层级注意力机制使拓扑图 GCN 和特征图 GCN 的节点表示进行了交互,为了突出共同信息,设计了一个一致性正则化损失,以约束 GCN 最后一层输出的一共 $2H$ 个节点表示 $\mathbf{Z}_t^{(1)}, \dots, \mathbf{Z}_t^{(H)}, \mathbf{Z}_f^{(1)}, \dots, \mathbf{Z}_f^{(H)}$ 的分布距离.首先,对于节点 v_i ,通过取所有分布的平均值来计算节点表示的分布中心:

$$\bar{\mathbf{Z}}_i = \frac{1}{2H} \sum_{h=1}^H (\mathbf{Z}_{Ti}^{(h)} + \mathbf{Z}_{Fi}^{(h)}) \quad (14)$$

然后,使 \mathbf{Z}_{Ti} 和 \mathbf{Z}_{Fi} 与 $\bar{\mathbf{Z}}_i$ 间的距离在一定程度上相似,一致性约束损失为:

$$\mathcal{L}_{con} = \frac{1}{H} \sum_{h=1}^H \sum_{i=0}^{n-1} \|\bar{\mathbf{Z}}_i - \mathbf{Z}_{Ti}^{(h)}\|_2^2 + \frac{1}{H} \sum_{h=1}^H \sum_{i=0}^{n-1} \|\bar{\mathbf{Z}}_i - \mathbf{Z}_{Fi}^{(h)}\|_2^2 \quad (15)$$

结合节点分类任务的交叉熵损失和一致性约束损失,总的损失函数为:

$$\mathcal{L} = \mathcal{L}_{class} + \gamma \mathcal{L}_{con} \quad (16)$$

其中 γ 是一致性约束的超参数.在训练集中有标签数据的指导下,通过反向传播来训练模型,学习用于分类的节点表示.

5 实验

5.1 实验设置

5.1.1 数据集

本文在 5 个真实数据集上评估了提出的方法,数据集统计汇总如表 1 所示.

表 1 数据集统计

Table 1 Statistics of the datasets

	节点个数	边个数	类别数	特征维度	训练集	测试集
ACM	3025	13128	3	1870	60/120/180	1000
BlogCatalog	5196	171743	6	8189	120/240/360	1000
CiteSeer	3327	4732	6	3703	120/240/360	1000
Flickr	7575	239738	9	12047	180/360/540	1000
UA12010	3067	28311	19	4973	380/760/1140	1000

ACM^[27]:该网络是从 ACM 数据集中提取的,其中节点代表论文.如果两篇论文的作者相同,那么它们之间有一条边.论文分为数据库、无线通信、数据挖掘 3 类.其特征是论文关键词的词袋表示.

BlogCatalog^[28]:它是一个博客社区社交网络的数据集,其中包含 5196 个用户节点,171743 条边表示用户交互.用户可以将他们的博客注册到 6 个不同的预定义类中,这些类为节点的标签.

CiteSeer^[18]:这是一个论文引用网络,记录了论文之间引用或被引用信息,节点表示论文,共分为 6 类,节点属性是论文的词袋表示.

Flickr^[28]:这是一个包含 7575 个节点的基准社交网络数据集.每个节点表示一个用户,边代表用户之间的关系.根据

用户的兴趣组将所有节点划分为9类。

UAI2010^[29]:该数据集有3067个节点和28311条边。

5.1.2 基线

为了验证本文提出的模型,将提出的方法与一些最先进的的方法进行了比较。本文使用的基线结果来自AM-GCN^[1]。

DeepWalk^[30]是一种网络嵌入方法,它使用随机游走来获取节点序列,然后借助自然语言处理的思想,将生成的节点序列看作由单词组成的句子,所有的序列可以看作一个大的语料库,最后利用word2vec将每一个顶点表示为一个相同维度的向量。

LINE^[31]是一种大规模的网络嵌入方法,分别保持网络的一阶和二阶邻近性,针对经典随机梯度下降算法的局限性,提出了一种边缘采样算法,提高了推理的效率和效果。

Chebyshev^[17]是利用切比雪夫滤波器在谱域实现图卷积神经网络的方法。

GCN^[18]是一种半监督图卷积网络模型,它通过聚合邻居的信息来学习节点表示。

kNN-GCN^[1]使用由特征矩阵得到的特征图代替传统的拓扑图作为GCN的输入图。

GAT^[3]在GCN的基础上引入注意力机制,在图卷积过程中为不同的邻居节点分配不同的权重。

DEMO-Net^[21]是一个具体度的图神经网络,提出多任务图卷积,即为不同度的节点使用不同的图卷积,从而保持具体

度的图结构。

MixHop^[20]针对GCN在图卷积过程中只能使用邻居节点的信息的缺点,提出多阶邻域的卷积,对不同的邻域采用不同的聚合方式,然后将结果拼接或用其他方式结合。

AM-GCN^[1]同时从节点特征、拓扑结构及其组合中提取两个特定节点表示的和一个共同的节点表示,并利用注意力机制学习这3种不同节点表示的重要性权重,将这些节点表示自适应地融合为最终的节点表示。

5.1.3 参数设置

在实验中,数据集采用AM-GCN相同的划分方式,即为训练集选择3种标签率(每类有20、40、60个标签节点),测试集包含1000个节点。数据增强层中,数据增强次数 $H \in \{1, 2, \dots, 5\}$,伯努利概率 $\delta \in \{0.1, 0.2, \dots, 0.5\}$ 。同时训练两个具有相同隐藏层数 $nhid1$ 和相同输出维数 $nhid2$ 的两层GCN网络,其中 $nhid1 \in \{512, 768\}$ 、 $nhid2 \in \{128, 256\}$,dropout为0.5。Adam优化器的学习率和权重衰减范围为 $\{0.0001, 0.0005, 0.001, 0.005\}$ 。特征图中 k 近邻的 $k \in \{2, 3, \dots, 20\}$ 。另外,一致性约束超参数取值范围为 $\gamma \in \{0.0001, 0.001, 0.01, 0.1, 1\}$ 。对于所有方法,本文使用随机初始化运行实验5次,然后取平均作为最终的结果。

5.2 节点分类

节点分类的结果如表2所示,L/C为每类标记节点的数量。实验结果分析如下:

表2 节点分类结果(%).(粗体:最优结果;下划线:次优结果)

Table 2 Node classification results(%).(The best result on each dataset are in bold. The second-best ones are underlined)												
数据集	度量指标	L/C	DeepWalk	LINE	Chebyshev	GCN	kNN-GCN	GAT	DEMO-Net	MixHop	AM-GCN	DC-GRCN
ACM	ACC	20	62.69	41.28	75.24	87.80	78.52	87.36	84.48	81.08	<u>90.40</u>	91.78
		40	63.00	45.83	81.64	89.06	81.66	88.60	85.70	82.34	<u>90.76</u>	92.36
		60	67.03	50.41	85.43	90.54	82.00	90.40	86.55	83.09	<u>91.42</u>	92.42
	F1	20	62.11	40.12	74.86	87.82	78.14	87.44	84.16	81.40	<u>90.43</u>	91.75
		40	61.88	45.79	81.26	89.00	81.53	88.55	84.83	81.13	<u>90.66</u>	92.30
		60	66.99	49.92	85.26	90.49	81.95	90.39	84.05	82.24	<u>91.36</u>	92.38
BlogCatalog	ACC	20	38.67	58.75	38.08	69.84	75.49	64.08	54.19	65.46	<u>81.98</u>	87.18
		40	50.80	61.12	56.28	71.28	80.84	67.40	63.47	71.66	<u>84.94</u>	89.74
		60	55.02	64.53	70.06	72.66	82.46	69.95	76.81	77.44	<u>87.30</u>	91.06
	F1	20	34.96	57.75	33.39	68.73	72.53	63.38	52.79	64.89	<u>81.36</u>	86.71
		40	48.61	60.72	53.86	70.71	80.16	66.39	63.09	70.84	<u>84.32</u>	89.45
		60	53.56	63.81	68.37	71.80	81.90	69.08	76.73	76.38	<u>86.94</u>	90.76
CiteSeer	ACC	20	43.47	32.71	69.80	70.30	61.35	72.50	69.50	71.40	<u>73.10</u>	74.12
		40	45.15	33.32	71.64	73.10	61.54	73.04	70.44	71.48	<u>74.70</u>	75.26
		60	48.86	35.39	73.26	74.48	62.38	74.76	71.86	72.16	<u>75.56</u>	76.08
	F1	20	38.09	31.75	65.92	67.50	58.86	68.14	67.84	66.96	<u>68.42</u>	69.48
		40	43.18	32.42	68.31	69.70	59.33	69.58	66.97	67.40	<u>69.81</u>	70.76
		60	48.01	34.37	70.31	71.24	60.07	<u>71.60</u>	68.22	69.31	<u>70.92</u>	72.02
Flickr	ACC	20	24.33	33.25	23.26	41.42	69.28	38.52	34.89	39.56	<u>75.26</u>	82.14
		40	28.79	37.67	35.10	45.48	75.08	38.44	46.57	55.19	<u>80.06</u>	85.84
		60	30.10	38.54	41.70	47.96	77.94	38.96	57.30	64.96	<u>82.10</u>	86.46
	F1	20	21.33	31.19	21.27	39.95	70.33	37.00	33.53	40.13	<u>74.63</u>	81.82
		40	26.90	37.12	33.53	43.27	75.40	36.94	45.23	56.25	<u>79.36</u>	85.76
		60	27.28	37.77	40.17	46.58	77.97	37.35	56.49	65.73	<u>81.81</u>	86.46
UAI2010	ACC	20	42.02	43.47	50.02	49.88	66.06	56.92	23.45	61.56	<u>70.10</u>	75.44
		40	51.26	45.37	58.18	51.80	68.74	63.74	30.29	65.05	<u>73.14</u>	76.54
		60	54.37	51.05	59.82	54.40	71.64	68.44	34.11	67.66	<u>74.40</u>	79.24
	F1	20	32.93	37.01	33.65	32.86	52.43	39.61	16.82	49.19	<u>55.61</u>	62.60
		40	46.01	39.62	38.80	33.80	54.45	45.08	26.36	53.86	<u>64.88</u>	66.99
		60	44.43	43.76	40.60	34.12	54.78	48.97	29.05	56.31	<u>65.99</u>	71.22

1) 与所有的基线相比,本文提出的 DC-GRCN 在所有的数据集上均达到了最优水平. 与最优的基线 AM-GCN^[1] 相比,本文提出的模型在 BlogCatalog 数据集上准确率最大提升 6.34 个百分点,macro F1-score 最大提升 6.57 个百分点;在 Flickr 数据集上准确率最大提升 9.14 个百分点,macro F1-score 最大提升 9.63 个百分点. 实验结果表明了 DC-GRCN 的有效性,DC-GRCN 能够更加充分地融合拓扑结构和节点特征间更深层次的信息,进一步有效地利用节点特征中的信息.

2) 通过对比 kNN-GCN 和 GCN 的结果,可以发现拓扑图和特征图确实存在结构差异. 并且,对于数据集 BlogCatalog、Flickr 和 UAI2010 这 3 个数据集上,kNN-GCN 的结果要明显优于 GCN,进一步说明了引入特征图的重要性.

3) 在所有数据集上,DC-GRCN 始终优于 GCN 和 kNN-GCN,这说明 DC-GRCN 中注意力机制的有效性,它们可以自

适应地从拓扑空间和特征空间中提取出最有用的信息.

4) 对于同一个数据集,标签率越高,模型的性能越高,表明 DC-GRCN 能够有效地使用节点的标签指导信息.

5.3 消融实验

本节研究层级注意力、一致性约束和图数据增强对 DC-GRCN 的贡献. 主要包括 3 部分:

1) w/o LA. 移除层级注意力模块,在进行 GCN 过程中,不对拓扑图 GCN 和特征图 GCN 的隐藏层节点表示进行交互.

2) w/o CC. 移除一致性约束损失,在损失函数中不对 GCN 最后一层的各节点表示进行一致性约束.

3) w/o DA. 移除图数据增强层,只使用节点原始特征参与训练和推理.

如图 2 为消融实验的结果,其中 ALL 为不做消融的 DC-GRCN. 根据实验结果得到的结论为:

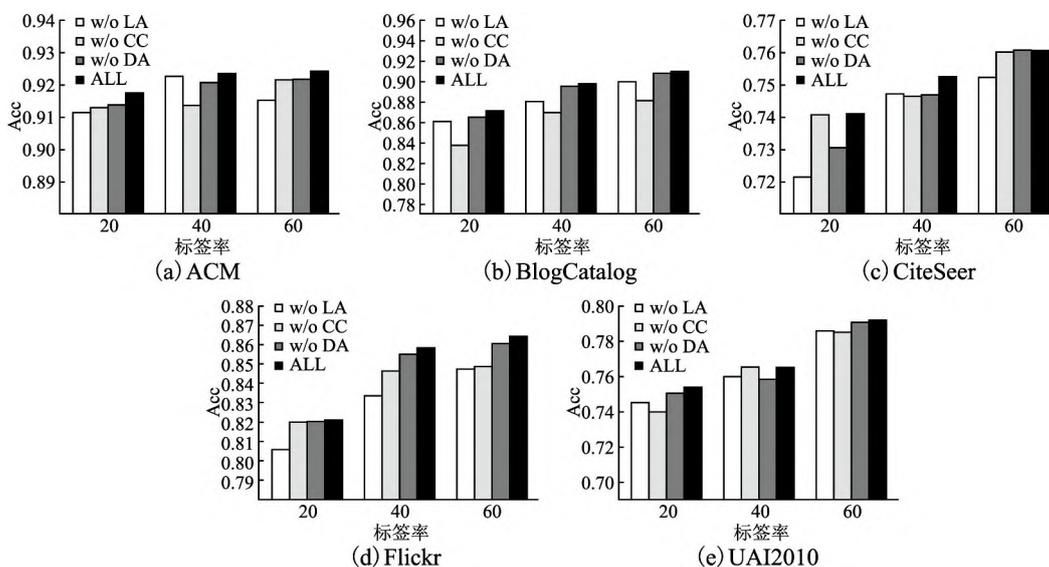


图 2 DC-GRCN 在 5 个数据集上的消融实验
Fig. 2 Ablation study of DC-GRCN on 5 datasets

1) 与完整的 DC-GRCN 相比,所有移除了某组件的 DC-GRCN 变体的性能都明显下降,这表明每个组件都起到了积极的作用.

2) 删除了层级注意力模块后,在 5 个数据集上的实验结果均有较大程度的下降. 可以看出,层级注意力模块可以自适应地融合从拓扑空间和特征空间中得到的信息,降低噪音信息的影响,从而获取更有用的节点表示,对于分类任务帮助较大.

3) 一致性约束能够使多种节点表示尽可能的保持一致,加强了各节点表示的相关性,突出了它们的共同信息.

4) 图数据增强模块将原始节点特征转换成多种表示,在图卷积过程中形成多通道,可以让模型关注到不同方面的信息,有助于模型学习到更丰富的信息.

5.4 可视化

为了更直观的比较本文方法的有效性,在标签率为 20 的 BlogCatalog 和 Flickr 数据集上执行可视化任务. 将最后得到的节点表示使用 t-SNE 进行降维,得到如图 3 所示的可视化图.

从图 3 中可以看出,GCN 学习到的节点表示把不同标签的节点都混到一起,不能很好地区分出各节点的类别;AM-

GCN 学习到的节点表示相对于 GCN 来说虽然能较为明显的

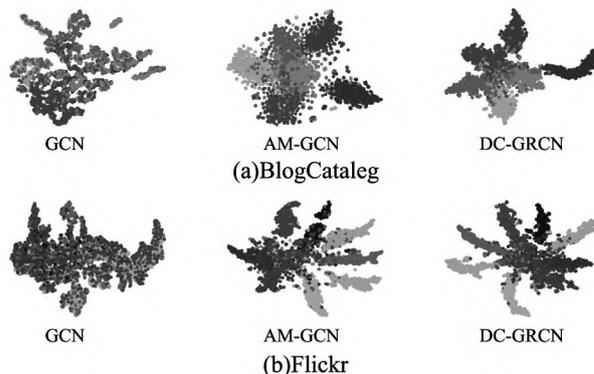


图 3 在 BlogCatalog 和 Flickr 数据集上学习到的节点表示的可视化结果

Fig. 3 Visualization of the learned node embeddings on BlogCatalog and Flickr dataset

区分各类,但类内相似性不够高;DC-GRCN 能够学习到更为

紧凑、类内相似性高、类间边界清晰的节点表示。

5.5 参数敏感性实验

5.5.1 特征图参数的敏感性分析

为了测试特征图中 k 近邻的超参数 k 的影响,实验测试

了 k 从 2 ~ 20 之间的 DC-GRCN 的性能,结果如图 4 所示.可以看出,随着 k 值的增加,DC-GRCN 的性能呈现逐渐上升然后缓慢下降的趋势.例如,在数据集 ACM 标签率为 60 时,随着 k 值从 2 开始增加,准确率逐渐上升,当 k 为 7 时,准确率

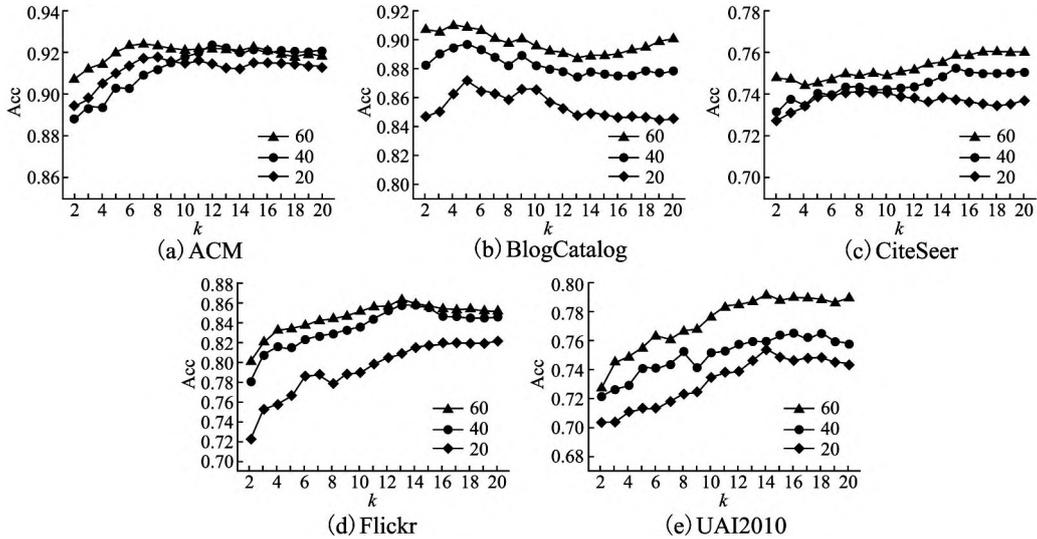


图4 参数 k 的研究

Fig. 4 Analysis of parameter k

达到最高,然后准确率随着 k 值的增加缓慢下降.其他数据集也呈现与 ACM 相同的趋势.这说明,在一定范围内,特征图中增加的大多数是有助于分类的高质量边.但随着边的增加,特征图越来越稠密,引入的噪音边就越多,从而导致分类性能下降.

5.5.2 一致性系数 γ 的敏感性分析

本节分析了公式(16)中一致性约束系数 γ 的影响, γ 越大,一致性约束损失占比就越大.取值范围为 $\gamma \in \{0, 0.0001, 0.001, 0.01, 0.1, 1, 10\}$,结果如图 5 所示.从实验结果中可以观察到,随着 γ 的增加,准确率呈现先缓慢上升然后急速下降

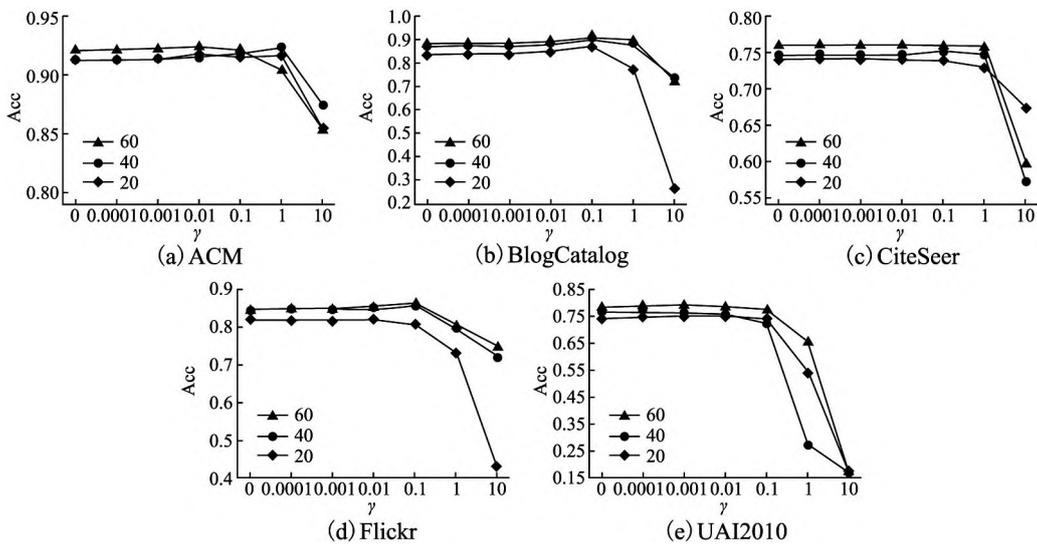


图5 参数 γ 的研究

Fig. 5 Analysis of parameter γ

的趋势.例如,在数据集 BlogCatalog 中,随着 γ 从 0 开始增加,准确率先缓慢上升,当 γ 为 0.1 时模型达到最佳性能,当 γ 的值大于 1 后,模型性能开始下降.其他数据集也呈现与 BlogCatalog 数据集相同的趋势,这说明,一致性约束 \mathcal{L}_{con} 能够在一定程度上进一步提高本文提出模型的表达能力.

5.5.3 图数据增强层超参数敏感性分析

本节分析图数据增强层中的超参数对模型的性能影响,主要的参数一个是特征增强个数 H ,另一个是 Bernoulli 概率 δ .参数敏感性结果如图 6 所示,可以观察到, H 和 δ 对模型的性能影响都起着积极的作用.对于 ACM 数据集,当 δ 固

定时,对于不同的 H 相对来说比较稳定;当 H 固定时,不同的 δ 对模型性能影响较大. 对于 Flickr 数据集来说,当 δ 取值较大

时($\delta \geq 0.3$),不同的 H 对模型性能产生影响较大. 实验结果表明,图数据增强层能够有效地提升模型的性能.

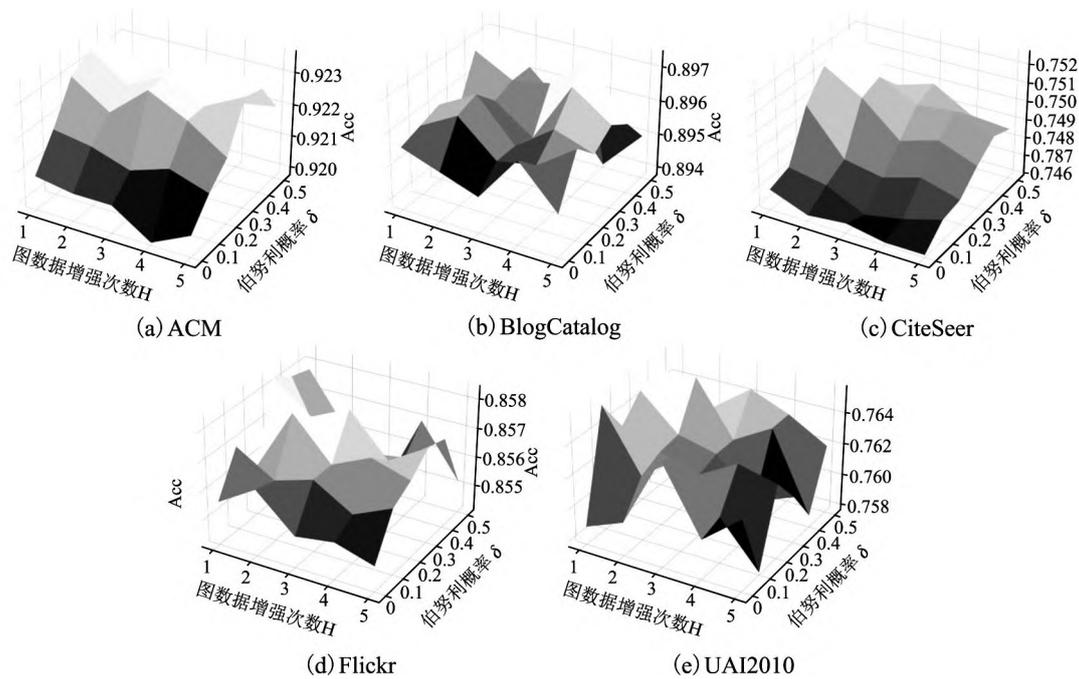


图6 参数 H 和 δ 的研究

Fig. 6 Analysis of parameters H and δ

6 总结

本文提出了一个面向半监督节点分类的图随机卷积网络 DC-GRCN. 首先,该模型利用节点特征矩阵生成一个特征图. 然后,使用图数据增强层生成多个增强的节点特征矩阵,使其分别在拓扑空间和特征空间中进行传播,在传播过程中将两种空间中的节点表示动态地进行信息交互. 最后,对于传播完成后的多个节点表示,使用注意力机制自适应地将它们结合,同时,对这些节点表示进行一致性正则化约束. 在 5 个真实数据集上的大量的实验表明,本文的提出的方法优于最先进的方

References:

- [1] Wang X, Zhu M, Bo D, et al. Am-gcn: adaptive multi-channel graph convolutional networks [C] // Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2020: 1243-1253.
- [2] Hamilton W L, Ying L, Leskovec J. Inductive representation learning on large graphs [C] // Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS), 2017: 1025-1035.
- [3] Veličković P, Cucurull G, Casanova A, et al. Graph attention networks [C] // International Conference on Learning Representations (ICLR), 2018, doi: 10. 48550/ arXiv. 1710. 10903.
- [4] Chen Bin, Li Jin-long. Attention-based network representation learning model using multi-neighboring information [J]. Journal of Chinese Computer Systems, 2021, 42(4): 761-765.
- [5] Yang Y, Feng Z, Song M, et al. Factorizable graph convolutional networks [C] // Advances in Neural Information Processing Systems (NIPS), 2020: 20286-20296.
- [6] Zhang M, Li P. Nested graph neural networks [C] // Advances in Neural Information Processing Systems (NIPS), 2021: 15734-15747.
- [7] Kim D, Oh A. How to find your friendly neighborhood: graph attention design with self-supervision [C] // International Conference on Learning Representations (ICLR), 2021, doi: 10. 48550/ arXiv. 2204. 04879.
- [8] Wu S, Tang Y, Zhu Y, et al. Session-based recommendation with graph neural networks [C] // Proceedings of the Association for the Advance of Artificial Intelligence (AAAI), 2019, 33(1): 346-353.
- [9] Liang Shun-pan, Tu Hao, Wang Rong-sheng, et al. Knowledge graph recommendation algorithm combining importance sampling and pooling aggregation [J]. Journal of Chinese Computer Systems, 2021, 42(5): 967-971.
- [10] Jin W, Derr T, Wang Y, et al. Node similarity preserving graph convolutional networks [C] // Proceedings of the 14th ACM International Conference on Web Search and Data Mining (WSDM), 2021: 148-156.
- [11] Feng W, Zhang J, Dong Y, et al. Graph random neural network for semi-supervised learning on graphs [C] // Proceedings of the International Conference on Neural Information Processing Systems (NIPS), 2020: 22092-22103.
- [12] Yoo J, Ahn N, Sohn K A. Rethinking data augmentation for image super-resolution: a comprehensive analysis and a new strategy [C] // Proceedings of the IEEE/ CVF Conference on Computer Vi-

- sion and Pattern Recognition(CVPR),2020:8372-8381.
- [13] Ghiasi G,Cui Y,Srinivas A,et al. Simple copy-paste is a strong data augmentation method for instance segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR),2021:2918-2928.
- [14] Wang Y,Wang W,Liang Y,et al. Nodeaug:semi-supervised node classification with data augmentation[C]//Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining(KDD),2020:207-217.
- [15] Zhao T,Liu Y,Neves L,et al. Data augmentation for graph neural networks[C]//Proceedings of the Association for the Advance of Artificial Intelligence(AAAI),2021:11015-11023.
- [16] Estrach J B,Zaremba W,Szlam A,et al. Spectral networks and deep locally connected networks on graphs[C]//Proceedings of the International Conference on Learning Representations(ICLR),2014,doi:10.48550/arXiv.1312.6203.
- [17] Defferrar M,Bresson X,Vanderghyest P. Convolutional neural networks on graphs with fast localized spectral filtering[J]. Advances in Neural Information Processing Systems(NIPS),2016:3844-3852.
- [18] Kipf T N,Welling M. Semi-supervised classification with graph convolutional networks[C]//Proceedings of the International Conference on Learning Representations(ICLR),2017,doi:10.48550/arXiv.1609.02907.
- [19] Xu Bing-bing,Cen Ke-ting,Huang Jun-jie,et al. A survey on graph convolutional neural network[J]. Chinese Journal of Computers,2020,43(5):755-780.
- [20] Abu-El-Haija S,Perozzi B,Kapoor A,et al. Mixhop:higher-order graph convolutional architectures via sparsified neighborhood mixing[C]//Proceedings of the International Conference on Machine Learning(ICML),2019:21-29.
- [21] Wu J,He J,Xu J. Demo-net:degree-specific graph neural networks for node and graph classification[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining(KDD),2019:406-415.
- [22] Shi M,Tang Y,Zhu X,et al. Multi-class imbalanced graph convolutional network learning[C]//Proceedings of the 29th International Joint Conference on Artificial Intelligence(IJCAI),2020:2879-2885.
- [23] Jin M,Zheng Y,Li Y F,et al. Multi-scale contrastive siamese networks for self-supervised graph representation learning[C]//Proceedings of the 30th International Joint Conference on Artificial Intelligence(IJCAI),2021:1477-1483.
- [24] Pei H,Wei B,Chang K C C,et al. Geom-gcn:Geometric graph convolutional networks[C]//Proceedings of the International Conference on Learning Representations(ICLR),2020,doi:10.48550/arXiv.2002.05287.
- [25] Suresh S,Budde V,Neville J,et al. Breaking the limit of graph neural networks by improving the assortativity of graphs with local mixing patterns[C]//Proceedings of the 27th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining(KDD),2021:1541-1551.
- [26] Bo D,Wang X,Shi C,et al. Structural deep clustering network[C]//Proceedings of the Web Conference(WWW),2020:1400-1410.
- [27] Wang X, Ji H, Shi C, et al. Heterogeneous graph attention network[C]//Proceedings of the World Wide Web Conference(WWW),2019:2022-2032.
- [28] Meng Z,Liang S,Bao H,et al. Co-embedding attributed networks[C]//Proceedings of the 20th ACM International Conference on Web Search and Data Mining(WSDM),2019:393-401.
- [29] Wang W,Liu X,Jiao P,et al. A unified weakly supervised framework for community detection and semantic matching[C]//Pacific-Asia Conference on Knowledge Discovery and Data Mining(KDD),2018:218-230.
- [30] Perozzi B,Al-Rfou R,Skiena S. Deepwalk:online learning of social representations[C]//Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining(KDD),2014:701-710.
- [31] Tang J,Qu M,Wang M,et al. Line:large-scale information network embedding[C]//Proceedings of the 24th International Conference on World Wide Web(WWW),2015:1067-1077.

附中文参考文献:

- [4] 陈斌,李金龙. 基于注意力机制融合多阶邻域信息的网络表示学习模型[J]. 小型微型计算机系统,2021,42(4):761-765.
- [9] 梁顺攀,涂浩,王荣生,等. 融合重要性采样和池化聚合的知识图推荐算法[J]. 小型微型计算机系统,2021,42(5):967-971.
- [19] 徐冰冰,岑科廷,黄俊杰,等. 图卷积神经网络综述[J]. 计算机学报,2020,43(5):755-780.