



计算机科学

COMPUTER SCIENCE

基于KAN的双通道图神经网络

王静红, 李鹏超, 王熙熙, 张自立

引用本文

王静红, 李鹏超, 王熙熙, 张自立. [基于KAN的双通道图神经网络](#)[J]. 计算机科学, 2026, 53(3): 188-196.

WANG Jinghong, LI Pengchao, WANG Xixi, ZHANG Zili. [Dual-channel Graph Neural Network Based on KAN](#) [J]. Computer Science, 2026, 53(3): 188-196.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于节点影响力的图遗忘学习近似最差遗忘集构造算法](#)

Node-influence Based Construction Algorithm of Approximate Worst-case Forgetting Set for Graph Unlearning

计算机科学, 2026, 53(3): 64-77. <https://doi.org/10.11896/jsjcx.250700094>

[图神经网络后门攻击与防御综述](#)

Survey of Backdoor Attacks and Defenses on Graph Neural Network

计算机科学, 2026, 53(3): 1-22. <https://doi.org/10.11896/jsjcx.250700093>

[基于异构图注意力网络的智能合约漏洞检测方法](#)

Heterogeneous Graph Attention Network-based Approach for Smart Contract Vulnerability Detection

计算机科学, 2026, 53(2): 423-430. <https://doi.org/10.11896/jsjcx.241200144>

[融合对比学习的掩码图自编码器](#)

Contrastive Learning-based Masked Graph Autoencoder

计算机科学, 2026, 53(2): 145-151. <https://doi.org/10.11896/jsjcx.250100155>

[基于方向感知孪生网络的知识概念先序关系预测方法](#)

Direction-aware Siamese Network for Knowledge Concept Prerequisite Relation Prediction

计算机科学, 2026, 53(2): 39-47. <https://doi.org/10.11896/jsjcx.250600005>

基于 KAN 的双通道图神经网络

王静红^{1,2,3,4} 李鹏超^{1,3,4,5} 王熙照⁶ 张自立^{1,3,4,5}

1 河北师范大学计算机与网络空间安全学院 石家庄 050024

2 河北工程技术学院人工智能学院 石家庄 050020

3 中国科学技术大学认知智能全国重点实验室 合肥 230088

4 河北省网络与信息安全重点实验室 石家庄 050024

5 供应链大数据分析与安全河北省工程研究中心 石家庄 050024

6 深圳大学计算机与软件学院 广东 深圳 518060

摘要 图神经网络(GNNs)是一种专门针对图数据的神经网络模型,近年来被成功应用在各种图学习任务上,如节点分类、链路预测等。然而,目前的图神经网络模型大多基于消息传递范式,无法充分捕捉节点的结构信息与特征信息之间的多维关联关系。此外,传统激活函数容易导致信息丢失和模型解释性不足的问题。为此,提出了一种基于 Kolmogorov-Arnold 网络(KAN)的双通道图神经网络(KDCGNN)。KDCGNN 利用结构卷积和特征卷积,从两个通道分别提取图的结构信息和特征信息,生成节点的结构编码和特征编码,拼接融合后,进一步借助 KAN 对嵌入表示进行特征转换,提升分类性能和模型的可解释性。同时,引入一致性损失函数,鼓励结构编码和特征编码之间的分布一致性,从而增强模型的泛化能力。在 3 个经典引文网络数据集(Cora,Citeseer,Pubmed)上的实验表明,KDCGNN 在节点分类任务中的表现优于现有基准方法。KDCGNN 的提出为图神经网络的可解释性与性能优化提供了新思路。

关键词: 图神经网络; Kolmogorov-Arnold 网络; 双通道机制; 节点分类; 高斯-Dice 相似度

中图分类号 TP391

Dual-channel Graph Neural Network Based on KAN

WANG Jinghong^{1,2,3,4}, LI Pengchao^{1,3,4,5}, WANG Xizhao⁶ and ZHANG Zili^{1,3,4,5}

1 College of Computer and Cyber Security, Hebei Normal University, Shijiazhuang 050024, China

2 College of Artificial Intelligence, Hebei University of Engineering Science, Shijiazhuang 050020, China

3 State Key Laboratory of Cognitive Intelligence, University of Science and Technology of China, Hefei 230088, China

4 Hebei Provincial Key Laboratory of Network and Information Security, Shijiazhuang 050024, China

5 Hebei Provincial Engineering Research Center for Supply Chain Big Data Analytics & Data Security, Shijiazhuang 050024, China

6 College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, Guangdong 518060, China

Abstract GNNs are specialized models designed for graph data and have been successfully applied to various graph learning tasks such as node classification and link prediction. However, most existing GNN models are based on the message-passing paradigm, which fails to fully capture the multi-dimensional relationships between structural information and feature information of nodes. Additionally, traditional activation functions often lead to information loss and lack interpretability in the models. To address these challenges, this paper proposes a novel Kolmogorov-Arnold Network-based Dual-Channel Graph Neural Network (KDCGNN). KDCGNN employs structural convolution and feature convolution in two separate channels to extract structural and feature information from graphs, generating structural and feature encodings for nodes. These encodings are then fused through concatenation and further transformed using the Kolmogorov-Arnold Network to enhance classification performance and model interpretability. Furthermore, a consistency loss function is introduced to encourage distributional alignment between structural and feature encodings, thereby improving the generalization capability of the model. Experiments on three benchmark citation network datasets (Cora, Citeseer, and Pubmed) demonstrate that KDCGNN outperforms existing baseline methods in node classification tasks. KDCGNN provides a novel approach to improving the interpretability and performance of graph neural networks.

收稿日期: 2025-06-11 返修日期: 2025-08-21

基金项目: 河北省自然科学基金(F2024205028); 河北省研究生创新资助项目(CXZZSS2025049); 河北师范大学科技类基金(L2023J05); 认知智能全国重点实验室开放课题(COGOS-2025HE07)

This work was supported by the Natural Science Foundation of Hebei Province(F2024205028), Postgraduate's Innovation Fund Project of Hebei Province(CXZZSS2025049), Hebei Normal University Science and Technology Fund(L2023J05) and Open Fund of State Key Laboratory of Cognitive Intelligence, iFLYTEK(COGOS-2025HE07).

通信作者: 王静红(wangjinghong@126.com)

Keywords Graph neural networks, Kolmogorov-Arnold networks, Dual-channel mechanism, Node classification, Gaussian-Dice similarity

近年来,图神经网络(Graph Neural Networks, GNNs)因其对图结构数据的强大建模能力,在节点分类、链路预测等任务中取得了显著进展。此外,图神经网络广泛应用于分子性质分析、疾病预测领域^[1-2]。其中,基于消息传递的图卷积网络(Graph Convolution Networks, GCNs)^[3]及其变体通过聚合邻居节点的特征信息,将拓扑结构与节点属性融合到统一的嵌入表示中,成为图表示学习的核心方法。然而,现有研究逐渐揭示出 GNNs 在关键环节的局限性:尽管 GCN 通过拉普拉斯平滑实现了局部特征传播,但其对拓扑结构和节点特征的非线性关联建模能力不足;而图注意力网络(Graph Attention Networks, GATs)^[4]虽能自适应调整邻居权重,却因依赖固定激活函数和黑箱注意力机制,导致模型可解释性差且对稀疏标记数据敏感。此外,现有方法对未标记节点信息的利用不足,尤其在低标记率场景下,模型泛化能力显著受限。这些缺陷严重制约了 GNNs 在复杂真实场景中的应用。

近年来,尽管多通道机制与自适应融合策略被引入以增强 GNN 的表达能力,但其设计仍存在关键挑战:1)特征与拓扑的融合过程缺乏动态可解释性,依赖人工调参;2)传统多层感知机(MLP)的非线性建模能力受限于固定激活函数,难以适应图数据的高维复杂性;3)节点表示在分布层面未显式对齐,导致模型对稀疏标记数据的适应性不足。这些问题的根源在于现有方法未能充分结合数学可解释性与数据驱动学习的优势,限制了其在复杂图分析任务中的潜力。

针对上述挑战,本文提出了 KDCGNN,该方法聚焦于信息融合与状态感知等关键技术。本文的主要贡献如下:

1)通过结构卷积模块(基于原始邻接矩阵)与特征卷积模块(基于高斯-Dice 相似度 kNN 图)分别建模拓扑结构与节点属性的深层关联,解决单一信息源建模不足的问题;

2)采用 KAN 的可学习 B 样条激活函数替代传统 MLP,动态拟合节点表示的非线性变换,避免信息丢失,并通过 Kolmogorov-Arnold 表示定理提供数学可解释性;

3)设计对称 KL 散度一致性损失,显式对齐结构编码与特征编码的概率分布,显著提升模型在低标记率场景下的鲁棒性。

1 相关工作

1.1 基于 KAN 的图神经网络

Kolmogorov-Arnold Network (KAN) 是一种基于 Kolmogorov-Arnold 表示定理的新型神经网络结构^[5]。与传统神经网络不同, KAN 的核心特点是可学习激活函数,其中原本的权重参数被单变量函数替代。这些函数通常采用样条函数进行参数化,从而提供极高的灵活性,使模型能够以更少的参数逼近复杂函数,同时提升其可解释性。本文将 GNN 中常用的 MLP 分类器替换成 KAN,提升了 GNN 的可解释性,并在节点分类任务上具有很好的性能。

最近,有一些研究人员试图将 KAN 推广到图神经网络中。这是因为大多数最有表现力的 GNN 模型依赖于多层感知器(MLP)作为其主要构建块,所以它们具有通用近似能

力^[6-7]。该定理指出,任何连续函数都可以通过至少包含一个隐藏层的 MLP 进行近似,前提是该层具有足够数量的神经元。然而,在实际应用中,这些模型由于非凸损失函数、缺乏收敛性保证的优化算法以及有限的可解释性,受到一定限制,从而影响了其在多个领域的适用性。

Zhang 等^[8]首次在图卷积任务中应用 KAN,并通过实验验证了 GraphKAN 的有效性,为处理复杂拓扑数据提供了一种新的方法论框架。Kiamari 等^[9]提出了一种图 Kolmogorov-Arnold 网络,与依赖于固定卷积架构的传统图卷积网络(GCNs)不同,该网络在层之间实现了可学习的基于样条的函数,改变了信息在图结构中的处理方式。在此基础上,作者提出了两种不同的方法来将 KAN 层集成到 GKAN 中:一是在特征聚合后应用可学习函数,二是在聚合前应用可学习函数。Bresson 等^[10]比较了 KANs 与 MLP 在图学习任务中的性能,并在节点分类、图分类和图回归数据集上进行了大量的实验。初步结果表明,虽然在分类任务中, KAN 与 MLP 相当,但在图回归任务中, KAN 似乎具有明显的优势。与上面工作不同的是,Carlo 等^[11]将 KAN 应用到 GNN 中,强调了其在可解释性至关重要的领域中的应用潜力,并指出这种设计本质上为模型的决策过程提供了清晰的见解,消除了对事后可解释性技术的需求。在具体应用方面, Ahmed 等^[12]将 KAN 与 GNN 融合,用于预测小分子与蛋白质靶点结合亲和力。虽然该模型没有达到最先进的性能,但结果证明了它的可行性,并强调了它作为计算药物发现的新方法的前景。这项工作开辟了新的研究方向,表明进一步完善和探索 KAN 与 GNN 的融合模型可以显著影响结合亲和力预测的效率和准确性,最终有助于发现新的治疗药物。

综上, KAN 在 GNN 领域的应用仍处于早期阶段,已有研究表明了其在图回归和节点分类任务中的潜力。本文基于 KAN 强大的函数表达能力,将其应用于双通道图神经网络的特征转换模块,以提升模型性能和可解释性,从而为相关领域提供新的研究思路。

1.2 多通道图神经网络

图神经网络在图数据处理领域取得了显著成功。然而,现有基于消息传递机制的 GNN 在特征聚合和更新过程中依赖固定模式,限制了其对邻域结构丰富度的表达能力,从而影响了模型的表示能力和泛化性能。鉴于此,受卷积神经网络(Convolutional Neural Networks, CNNs)的多通道特性启发,引入多通道机制以增强 GNN 的特征表达能力,进一步提升其对复杂图结构的建模效果。

Zhou 等^[13]提出了一种多通道图卷积网络(MuchGCN),它是由卷积神经网络驱动的,在卷积神经网络中,一系列的通道被编码以保持输入图像的综合特征。因此,该研究通过定义特定图卷积算子,在各网络层学习多组图特征通道,并采用迭代图池化操作构建层次化表征。Zhao 等^[14]引入从学者的出版物中获得的更多样化的信息(如职业序列、协作图)来定义学者工作迁移预测任务,以便在没有隐私数据的情况下进行更全面的建模。此外,还设计了一个多通道多分支增强框

架(MMEF),以互补的方式整合异构输入。该框架可以捕捉职业轨迹的时间模式,利用学术合作信息考虑合著者的影响,并处理多余的描述,估计学者和机构之间的相关性分数。Meng等^[15]提出了一种多通道深度图卷积神经网络方法DeepMCGCN,它构造了多个关系子图,并采用多个GCN通道分别学习不同关系子图的特征。利用跨通道连接获取不同关系子图之间的交互关系,可以比单通道GNN学习更丰富、更具判别性的节点嵌入。同时,通过优化卷积函数,增加通道间和通道内的残差连接,缓解了深度模型的过拟合问题。Wang等提出了一种自适应多通道深度图神经网络(AMD-GNN)来自适应和对称地聚合来自深度感受野的信息。该模型保证了深层各节点的接受域是不同的,从而保证了节点表示的可区分性。为了解决GCNs可能无法自适应地学习拓扑结构和节点特征之间的一些深度相关信息的问题,Wang等^[16]提出了一种用于半监督分类(AM-GCN)的自适应多通道图卷积网络。该方法的核心思想是同时从节点特征、拓扑结构及其组合中提取特定的和共同的嵌入,并使用注意机制来学习嵌入的自适应重要权重。该模型既能保留最先进的GCNs的优点,又能大幅增强拓扑结构和节点特征的融合能力。

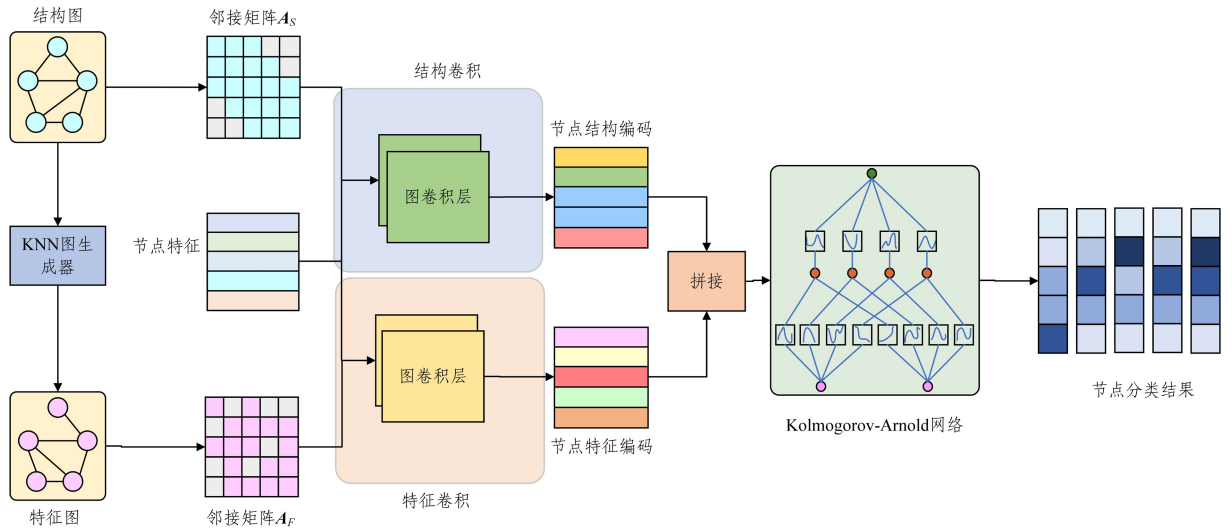


图1 KDCGNN模型框架

Fig.1 Framework of KDCGNN model

2.2 问题建模

本文研究面向属性图 $G=(V, A, X)$ 的半监督节点分类任务,其中: $V=\{v_1, v_2, \dots, v_n\}$ 为有 n 个节点的节点集合; $A \in \mathbb{R}^{n \times n}$ 为有 n 个节点的对称邻接矩阵, $A_{ij}=1$ 表示节点 i 和 j 之间存在一条边,否则 $A_{ij}=0$; $X \in \mathbb{R}^{n \times d}$ 为节点特征矩阵, d 表示节点特征的维数。

假设节点类别集合为 C ,其大小为 $|C|$ 。在训练阶段,仅有部分节点具有已知标签,记为标记节点集合 $V_L \subseteq V$,对应的标签为 $y_i \in \{1, 2, \dots, |C|\}$ 。任务目标是通过 V_L 的监督信息,学习分类模型以预测未标记节点集合 V/V_L 的类别标签。

2.3 基于拓扑信息结构卷积模块

为有效捕捉图数据中节点间的拓扑结构信息,本节提出一种基于拓扑信息结构卷积模块。该模块通过多层图卷积操作逐步聚合邻接节点的特征信息,生成与拓扑强关联的节

在GNNs中,多通道架构是一种有效的方法,用于增强模型在处理图数据时的特征学习能力。多通道GNN通过在不同的通道上并行处理信息,可以捕获图数据中不同方面的特征,从而提高模型的性能。虽然多通道GNN在处理图结构数据时提供了强大的工具,但仍然需要应对一些挑战,如过拟合、过平滑和计算复杂度等问题。未来的研究可能会集中在提高模型的泛化能力、扩展性和解释性上。

2 本文模型 KDCGNN

2.1 KDCGNN模型

本文提出的KDCGNN模型框架如图1所示,包括结构卷积、特征卷积、KAN和特征融合。其关键思想是,KDCGNN包含两个通道,并从这两个通道中提取图的特征编码和结构编码。为此,节点特征通过kNN图生成器构造一个特征图,原始图作为结构图,并通过结构卷积和特征卷积这两个特定的卷积模块,节点特征可以在特征图和结构图上同时传播,分别学习节点结构编码和节点特征编码这两个特定的嵌入,然后将这两部分编码进行拼接融合,再通过KAN进行特征维度的变换,最后得到节点分类的结果。

点表示。具体设计如下:

结构图就是原始图,即 $G_s=(V_s, A_s, X_s)$,其中 $V_s=V$, $A_s=A$, $X_s=X$ 。为确保每个节点在特征传播过程中能够保留自身的特征信息,对邻接矩阵添加自环:

$$\tilde{A}_s = A_s + \mathbf{I}_s \quad (1)$$

其中, $\mathbf{I}_s \in \mathbb{R}^{n \times n}$ 为单位矩阵,用于保留节点自身特征。

基于拓扑图的学习输出嵌入 Z_s 的计算过程如下:

$$Z_s^{(l)} = \text{ReLU}(\tilde{D}_s^{-\frac{1}{2}} \tilde{A}_s \tilde{D}_s^{-\frac{1}{2}} Z_s^{(l-1)} W_s^{(l)}) \quad (2)$$

其中, $W_s^{(l)}$ 为图卷积神经网络中第 l 层的权重矩阵,用于特征变换; $\text{ReLU}(\cdot)$ 为激活函数,引入非线性表达能力; $Z_s^{(l)}$ 为第 l 层的输出特征,初始 $Z_s^{(0)}=X$; \tilde{D}_s 为 \tilde{A}_s 的对角度矩阵。经过 l 层卷积后,最终的结构编码为 Z_s 。

2.4 基于高斯-Dice相似度构建kNN图

为充分挖掘节点属性间的潜在关联,本节提出一种基于

高斯-Dice 相似度的 k 近邻(kNN)图构建方法,旨在生成能够表征特征空间局部相似性的结构引导图。首先,基于节点特征矩阵 $\mathbf{X} \in \mathbb{R}^{n \times d}$, 计算节点间的相似度矩阵 $\mathbf{S} \in \mathbb{R}^{n \times n}$ 。

传统方法通常采用以下两种方法衡量特征相似性,其中 x_i 和 x_j 是节点 i 和 j 的特征向量。

1)线性核:用两个向量的内积来衡量相似度,由于没有将向量映射到高维空间,可以保持数据的原始状态。其计算式为:

$$\mathbf{S}_{ij} = \mathbf{x}_i \mathbf{x}_j^\top \quad (3)$$

2)高斯核:一种具有局部性强的相似性计算方法,其通过将特征向量映射到一个更高维的空间内,从而有效捕捉样本间的非线性相似关系。其计算式为:

$$\mathbf{S}_{ij} = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2) \quad (4)$$

其中, $\|\mathbf{x}_i - \mathbf{x}_j\|$ 表示 x_i 和 x_j 之间的欧氏距离; γ 为缩放因子,用来控制高斯核函数的宽度。因此,当两个节点特征越接近时,高斯核函数的值就越趋近于 1;反之,当两个节点特征越远时,其值就越接近于 0。这种特性使得高斯核函数在区分输入空间中相似的样本时特别有效。

但 these 方法仅依赖特征距离,忽略了拓扑结构的影响。为进一步优化高斯核函数的相似度计算方法,本文引入 Dice 相似度,将其与高斯核函数结合。该方法在考虑特征距离的

同时,也融入了节点的拓扑分布信息,一定程度上反映了节点度对相似性的影响,从而提高了计算结果的鲁棒性和可解释性。Dice 相似度通过量化节点邻居的重叠比例引入拓扑分布信息,公式如下:

$$\mathbf{S}_{\text{Dice}}(\mathbf{v}_i, \mathbf{v}_j) = \frac{2\text{CoN}(\mathbf{v}_i, \mathbf{v}_j)}{N(\mathbf{v}_i) + N(\mathbf{v}_j)} \quad (5)$$

其中, $\text{CoN}(\mathbf{v}_i, \mathbf{v}_j)$ 表示 \mathbf{v}_i 和 \mathbf{v}_j 的共同邻居数量, $N(\mathbf{v}_i)$ 和 $N(\mathbf{v}_j)$ 分别表示 \mathbf{v}_i 和 \mathbf{v}_j 的邻居数量。

将高斯核函数中的缩放因子 γ 替换成 Dice 相似度,如式(6)所示:

$$\mathbf{S}_{ij} = \exp(-\mathbf{S}_{\text{Dice}}(\mathbf{v}_i, \mathbf{v}_j) \|\mathbf{x}_i - \mathbf{x}_j\|^2) \quad (6)$$

该方法不仅保留特征空间的非线性关联,还通过 Dice 相似度消除节点度偏差,避免度数大的节点在相似性计算中占据主导地位。

这里选择高斯-Dice 相似度来计算各个节点之间的相似度,如图 2 所示。首先,对于原始图的节点特征,通过相似度函数依次得到两个节点的相似度值,从而得到相似度矩阵 \mathbf{S} ; 然后为每个节点选择相似度最高的前 k 个邻居(top- k)设置边,构建 kNN 邻接矩阵 $\mathbf{A}_f \in \{0,1\}^{n \times n}$, 定义为:

$$\mathbf{A}_f[i, j] = \begin{cases} 1, & \text{若 } \mathbf{v}_i \in \text{top-}k(\mathbf{S}_{ij}) \\ 0, & \text{否则} \end{cases} \quad (7)$$

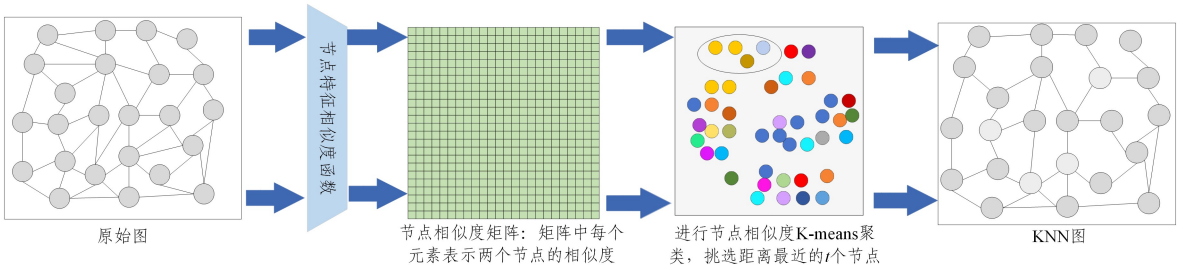


图 2 kNN 图的生成过程

Fig. 2 Generation process of kNN graph

2.5 基于相似度图的特征卷积模块

相似度图是 2.4 节中通过节点相似度构造的 kNN 图,即 $\mathbf{G}_f = (\mathbf{V}_f, \mathbf{A}_f, \mathbf{X}_f)$, 其中, $\mathbf{V}_f = \mathbf{V}$, $\mathbf{A}_f = \mathbf{A}_f$, $\mathbf{X}_f = \mathbf{X}$ 。基于特征图的学习节点嵌入 \mathbf{Z}_f 的计算过程如下:

$$\mathbf{Z}_f^{(l)} = \text{ReLU}(\tilde{\mathbf{D}}_f^{-\frac{1}{2}} \tilde{\mathbf{A}}_f \tilde{\mathbf{D}}_f^{-\frac{1}{2}} \mathbf{Z}_f^{(l-1)} \mathbf{W}_f^{(l)}) \quad (8)$$

其中, $\mathbf{W}_f^{(l)}$ 为图卷积神经网络中第 l 层的权重矩阵; 初始 $\mathbf{Z}_f^{(0)} = \mathbf{X}$; $\tilde{\mathbf{D}}_f$ 为 $\tilde{\mathbf{A}}_f$ 的对角度矩阵; 将最后一层的输出嵌入表示为 \mathbf{Z}_f , 即节点的特征编码。

2.6 KAN 的深层特征融合与分类

通过结构卷积模块和特征卷积模块分别得到节点结构编码 \mathbf{Z}_s 和特征编码 \mathbf{Z}_f , 考虑到节点标签可以与它们的组合相关联, 这里采用拼接的方式进行两部分节点的特征融合。这种方式的计算复杂度比较小, 如式(9)所示:

$$\mathbf{Z} = \text{concat}(\mathbf{Z}_s, \mathbf{Z}_f) \quad (9)$$

其中, $\mathbf{Z} \in \mathbb{R}^{n \times m}$ 为节点的最终嵌入表示, concat 为拼接函数。

为了完成节点分类任务, 需要对节点的最终嵌入表示 \mathbf{Z} 的第二个维度进行特征变换, 即 $\mathbb{R}^m \rightarrow \mathbb{R}^c$, 其中, \mathbb{R}^m 和 \mathbb{R}^c 分别表示 m 维和 c 维的向量。在这里不采用多层感知机, 而是采用 KAN。与 MLP 不同, KAN 引入了一种新颖的方法, 通

过沿网络边缘用基于样条的单变量函数替换线性权重。这些函数被特别构造为可学习的激活函数。KAN 可用于解决 GNN 中 MLP 遇到的效率和可解释性限制。此外, KAN 还解决了使用激活函数而导致的信息丢失问题。适合更深层次架构的 KAN 层由单变量函数矩阵定义, 如式(10)所示:

$$\Phi = \{\phi_{q,p}\}, p=1,2,\dots,n_{\text{in}}, q=1,2,\dots,n_{\text{out}} \quad (10)$$

其中, 每个函数 $\phi_{q,p}$ 都有可训练的参数, n_{in} 和 n_{out} 分别表示单变量函数矩阵的列和行数。由此, 最初被描述为仅由两层此类函数组合构成的 Kolmogorov-Arnold 表示, 可以扩展到更深的网络结构。在本文中, KAN 的激活函数为 B 样条函数。

为了具体而直观地理解 KAN 架构, 将其表示为整数数组 $[n_0, n_1, \dots, n_L]$, 其中 n_i 为第 i 层节点数。激活值及其在层间的转换定义如式(11)所示:

$$\mathbf{x}_{l+1,j} = \sum_{i=1}^{n_l} \phi_{l,j,i}(\mathbf{x}_{l,i}) \quad (11)$$

其中, 将第 l 层的节点 i 与第 $l+1$ 层的节点 j 连接起来。这种设置允许堆叠深层 KAN 层, 便于通过连续变换与求和对复杂函数进行建模。

上述为 KANLayer 的定义,将之前得到的最终嵌入表示通过 KAN 进行维度变换:

$$\hat{\mathbf{Y}} = \text{softmax}(\text{KANLayer}(\mathbf{Z})) \quad (12)$$

其中, $\text{softmax}(\mathbf{x}) = \frac{\exp(\mathbf{x})}{\sum_{i=1}^n \exp(\mathbf{x}_i)}$ 用于分类归一化函数, $\hat{\mathbf{Y}} \in \mathbb{R}^{n \times c}$

为预测的各个节点的标签。

2.7 一致性损失设计与优化目标

为实现结构编码与特征编码的分布对齐,本节提出一种基于 KL 散度的一致性损失函数,通过约束两种编码在概率空间中的相似性,增强模型对未标记节点的泛化能力。具体设计如下:

首先,对于节点结构编码 $\mathbf{Z}_S = [\mathbf{z}_1^s, \mathbf{z}_2^s, \dots, \mathbf{z}_m^s]$, 计算其在消息传递模块嵌入空间中的概率分布:

$$\tilde{\mathbf{z}}_s^i = \frac{\exp(\mathbf{z}_s^i / T)}{\sum_{j=1}^m \exp(\mathbf{z}_s^j / T)} \quad (13)$$

其中, T 为温度参数。这样就可以得到节点结构编码的分布。同理,可以计算得到节点特征编码的分布。

为了鼓励节点结构编码的分布 $\tilde{\mathbf{Z}}_S = [\tilde{\mathbf{z}}_1^s, \tilde{\mathbf{z}}_2^s, \dots, \tilde{\mathbf{z}}_m^s]$ 和节点特征编码的分布 $\tilde{\mathbf{Z}}_F = [\tilde{\mathbf{z}}_1^f, \tilde{\mathbf{z}}_2^f, \dots, \tilde{\mathbf{z}}_m^f]$ 表示之间的一致性,设计了一致性损失来进一步增强它们的通用性并通过对称化设计来避免单方向对齐导致的偏差。一致性损失定义如下:

$$\mathcal{L}_{\text{con}} = \frac{1}{m} \sum_{i=1}^m \frac{1}{2} (D_{\text{KL}}(\mathbf{z}_s^i \parallel \tilde{\mathbf{z}}^i) + D_{\text{KL}}(\mathbf{z}_f^i \parallel \tilde{\mathbf{z}}^i)) \quad (14)$$

其中, \mathcal{L}_{con} 表示一致性损失, D_{KL} 是用来计算两个概率分布的 KL 散度, $\tilde{\mathbf{z}}^i = (\mathbf{z}_s^i + \mathbf{z}_f^i) / 2$ 。为了避免 KL 散度在概率分布不重叠时会出现趋向于无穷大的问题,这里采用 JS 散度 (Jensen-Shannon divergence), JS 散度是 KL 散度的对称化版本。

假设训练集为 \mathbf{L} , 对于每个 $l \in \mathbf{L}$, 真实标签为 \mathbf{Y}_l , 预测标签为 $\hat{\mathbf{Y}}_l$, 则所有训练节点上节点分类的交叉熵损失表示为 \mathcal{L}_l :

$$\mathcal{L}_l = - \sum_{i \in \mathbf{A}} \sum_{a=1}^C \mathbf{Y}_{li}^a \ln \hat{\mathbf{Y}}_{li}^a \quad (15)$$

结合分类交叉熵损失和一致性损失,可以得到模型总的损失函数:

$$\mathcal{L} = \mathcal{L}_l + \alpha \mathcal{L}_{\text{con}} \quad (16)$$

其中, α 为超参数,用于平衡监督学习与分布对齐的强度。通过反向传播联合优化模型参数,确保分类性能与分布一致性同步提升。

2.8 算法流程

KDCGNN 进行节点分类的算法流程如算法 1 所示。

算法 1 KDCGNN 节点分类的算法流程

输入: 图 \mathbf{G} 的邻接矩阵 \mathbf{A} , 节点集合 \mathbf{V} , 节点特征 \mathbf{X} , 层数 l

输出: 节点的标签 $\hat{\mathbf{Y}}$

1. 初始化节点嵌入 $\mathbf{Z}_i^{(0)} = \mathbf{Z}_S^{(0)} = \mathbf{X}$, 拓扑图邻接矩阵 $\mathbf{A}_S = \mathbf{A}$
2. 通过式(5)计算各个节点的相似度矩阵 \mathbf{S}
3. 将节点按相似度进行排序
4. 选择 top-k 节点构建 kNN 图 \mathbf{G}_i
5. for $i=1, 2, 3, \dots, l$ do
6. 通过式(2)得到第 i 层的结构编码 \mathbf{Z}_S
7. 通过式(8)得到第 i 层的特征编码 \mathbf{Z}_F
8. 通过式(14)计算 \mathbf{Z}_S 与 \mathbf{Z}_F 间的一致性损失 \mathcal{L}_{con}
9. end for

10. 将 \mathbf{Z}_S 与 \mathbf{Z}_F 进行拼接得到节点最终嵌入表示 \mathbf{Z}

11. 通过式(12)计算得到节点的标签 $\hat{\mathbf{Y}}$

12. 通过式(15)计算节点标签的交叉熵损失 \mathcal{L}_l

13. 通过式(16)得到最终的损失 \mathcal{L}

3 实验与结果分析

3.1 实验设置

3.1.1 数据集

本节使用了 3 个引文网络数据集 Cora, Citeseer 和 Pubmed^[17], 用于半监督节点分类。3 个数据集的指标统计如表 1 所列。

表 1 节点分类数据集统计信息

Table 1 Node classification dataset statistics

数据集	类型	节点数	边数	类别数	维度
Cora	引文网络	2708	5429	7	1433
Citeseer	引文网络	3327	4732	6	3703
Pubmed	引文网络	19717	44338	3	500

Cora 数据集是一个由 2708 篇机器学习领域的科学论文构成的引文网络数据集, 广泛用于图深度学习研究。这些论文被分为 7 个类别, 分别是案例基础、遗传算法、神经网络、概率方法、强化学习、规则学习和理论。每篇论文通过一个 1433 维的向量表示, 每个维度对应一个词, 取值为 0 或 1, 分别表示该词在论文中未出现或出现。这些词来自于一个包含 1433 个词汇的字典。由于每篇论文至少与其他一篇论文存在引用或被引用的关系, 因此样本点之间相互连接, 形成了一个无孤立点的连通图。

Citeseer 数据集是一个引文网络, 包含 3312 篇论文, 这些论文被归类为 6 个主要领域, 分别是智能体 (Agents)、人工智能 (AI)、数据库 (DB)、信息检索 (IR)、机器学习 (ML) 和人工交互 (HCI)。数据集中记录了论文间的引用关系。在预处理阶段, 移除了停用词以及文档频率低于 10 的词汇, 最终得到了 3703 个独特的词汇。

Pubmed 数据集是一个专注于糖尿病研究的引文网络, 它包含来自 PubMed 数据库的 19717 篇科学出版物。这些出版物被分为 3 个类别, 分别是糖尿病实验 (Diabetes Mellitus, Experimental)、1 型糖尿病 (Diabetes Mellitus Type 1) 和 2 型糖尿病 (Diabetes Mellitus Type 2)。该网络由 44338 个引用链接构成。数据集中的每篇出版物都通过一个由 500 个唯一词汇组成的字典中的 TF/IDF 加权词向量来表示。

3.1.2 基准方法

对于节点分类, 将本文方法与以下基准方法进行了比较:

- 1) 传统的节点嵌入方法, 包括 Node2vec^[18] 和 Deep Walk^[19];
- 2) 基于图神经网络的方法, 包括 GCN, GAT, Graphsage^[20], SGC^[21], CPGNN^[22] 和 HOG-GCN^[23];
- 3) 基于图池的方法, 包括 g-NUnets^[24] 和 SEP-N^[25]。

3.1.3 参数设置

通过测量 KDCGNN 模型在节点分类任务中的准确性来评估其有效性。为验证模型的有效性性与鲁棒性, 所有实验均基于以下参数配置。

在训练过程中,采用 Adam 优化器进行参数优化。每种方法和数据集均运行 300 个 epoch,并根据验证集上最高准确率的 epoch 作为最终的测试准确率。为便于模型调优,设置了以下超参数:dropout 率为 0.5,权重衰减系数为 5×10^{-4} ,一致性损失权重 α 为 0.6,温度参数 T 为 10,隐藏层维度设置为 $\{64, 32, 16\}$,top- k 构建特征图中参数 k 的取值为 $[0, 9]$ 的整数。

在数据集划分方面,训练集、验证集和测试集的比例为 6:2:2。由于本文实验为半监督节点分类任务,训练集中的标记数据比例分别设置为 30% 和 50% 进行实验。所有实验均基于 PyTorch 框架实现,重复运行 10 次取其均值,固定随机种子以保证可复现性。

3.1.4 实验条件

在本文的所有实验中,所使用的显卡为 NVIDIA GeForce RTX 3090Ti,其中显存大小为 24 GB,操作系统为 Ubuntu 16.04.1 LTS 64 位,代码实现采用 Python 3.9,深度学习框架为 PyTorch 2.1。

3.2 实验结果与分析

3.2.1 节点分类准确率分析

在 3 个节点分类数据集上进行实验,并将 KDCGNN 与基准方法节点分类准确率进行了对比,结果如表 2 所列。在标记数据占比 30% 和 50% 的情况下分别进行了实验,即针对训练集中 30% 和 50% 节点的嵌入表示进行梯度的反向传播。

本文 KDCGNN 方法在 3 个数据集上均达到了最优的效果。其中,在 Cora 数据集上比次优的 CPGNN 高出了 0.9 个百分点和 0.36 个百分点;在 Citeseer 数据集上,比次优的 HOG-GNN 高出了 1.31 个百分点和 0.34 个百分点,Citeseer 数据集 50% 标记率下 KDCGNN 分类性能与 30% 时相差不大,这是因为 Citeseer 数据集的图结构相对稀疏,节点间关联性不强,可能会影响 GNN 的消息传递效率;而在 Pubmed 数据集

上,KDCGNN 比次优的 HOG-GNN 高出了 1.23 个百分点和 0.62 个百分点。

从以上结果可以看出,KDCGNN 在标记数据占比 30% 时,其准确率较次优方法所提升的幅度更大。这表明,KDCGNN 能够充分利用未标记的数据,在标记占比较少时,仍然能够保持更好的性能,并且增加特征图模块有助于捕获节点在特征空间的相似性,从而提升模型的表征能力。

表 2 KDCGNN 与基准方法在不同数据集上的分类对比

Table 2 Classification comparison between KDCGNN and baseline methods on different datasets

数据集	Cora		Citeseer		Pubmed	
	30%	50%	30%	50%	30%	50%
Node2vec	75.45	77.92	52.12	53.13	75.32	76.74
Deep Walk	74.83	78.58	45.26	48.16	74.54	75.21
GCN	85.65	86.67	72.35	75.86	86.33	87.27
GAT	85.73	86.92	71.36	74.35	84.77	86.45
Graphsage	76.34	78.62	66.46	67.33	76.21	77.83
SGC	80.76	82.13	71.33	72.24	77.67	78.07
CPGNN	<u>86.54</u>	<u>87.89</u>	74.76	76.15	86.41	88.76
HOG-GCN	86.13	87.56	<u>75.44</u>	<u>76.53</u>	<u>87.09</u>	<u>88.80</u>
g-N-Unets	85.73	86.16	73.45	74.87	80.23	81.33
SEP-N	85.26	86.17	73.49	75.03	81.25	83.47
KDCGNN	87.44	88.25	76.75	76.87	88.32	89.42

注:每个数据集上最优的结果用粗体标注,次优的结果用下划线标识。

3.2.2 模型的稳定性分析

为了更全面地统计节点分类的分类准确率,表 2 中的 KDCGNN 结果是 10 次实验中测试集分类准确率的平均值。然后针对 3 个节点分类数据集每一次实验的测试集分类准确率进行了统计,如图 3 所示。从图中可以看出,标记数据占比 30% 时,整体的节点分类的准确率不如标记数据占比 50% 时,与此同时,在各个数据集上的分类准确率的波动总体不大,证明了 KDCGNN 的稳定性。

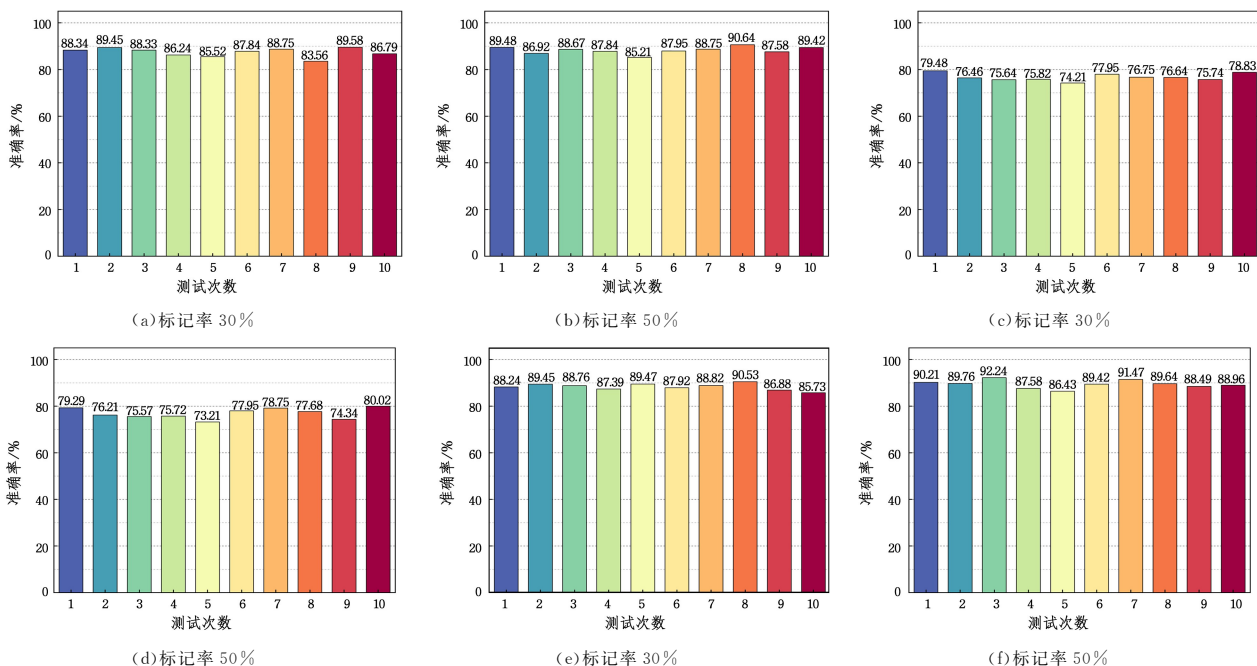


图 3 10 次实验中测试集的分类准确率

Fig. 3 Classification accuracy on the test set across ten experiments

3.2.3 模型的收敛性分析

为了研究 KDCGNN 训练过程的稳定性,分析其收敛情况,如图 4 所示,将每一个数据集在训练集、测试集上的损失值和准确率的变化情况进行了统计。在 Cora 数据集上,训练集在训练轮次为 120 时开始收敛,其准确率和损失值变化的幅度很小。同样地,测试集也大概在训练轮次为 120 处开始收敛。在 Citeseer 数据集上,训练集在训练轮次为 180 时开始收敛,其准确率和损失值变化的幅度很小。测试集和训练集的损失值和准确率变化趋势也是类似的。在 Pubmed 数据集上,训练集在训练轮次为 200 时开始收敛,其准确率和损失值变化的幅度很小。测试集的损失值仍然处在下降阶段,这是因为 Pubmed 在 3 个数据集中规模是最大的,因此收敛比较慢,需要更多的训练轮次。综上,可以得出如下结论:这 3 个数据集在训练和测试的过程中,损失值平稳下降,准确率平稳上升,没有出现异常的波动,这表明了 KDCGNN 在训练过程中没有出现拟合或者梯度爆炸的情况。此外,KDCGNN 通过双通道能够提取更多的结构和特征信息,有效地提高了节点分类的性能。

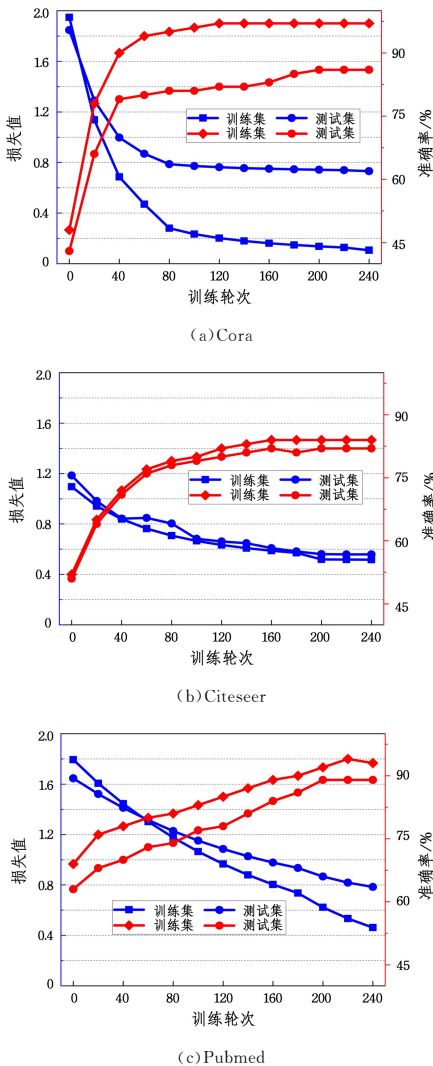


图 4 3 个数据集上的训练和测试损失值、准确率的变化情况
Fig. 4 Variation of training and testing losses and accuracy across three datasets

3.3 消融实验

为了验证双通道机制的有效性,即特征卷积、结构卷积和一致性损失函数是否起到了作用,进行了模块的消融实验。将原本的 KDCGNN 依次删除一些组件,主要包括结构卷积、特征卷积和一致性损失函数。由此可以得到 KDCGNN 的以下 3 种变体:

- 1) KDCGNN(w/o 结构卷积): KDCGNN 删除结构卷积模块,只保留特征卷积模块。
- 2) KDCGNN(w/o 特征卷积): KDCGNN 删除特征卷积模块,只保留结构卷积模块。
- 3) KDCGNN(w/o 一致性损失): KDCGNN 删除一致性损失函数,只保留交叉熵损失函数。

图 5 展示了 KDCGNN 及其 3 种变体在 Cora, Citeseer 和 Pubmed 上的性能。其中,图 5(a)为标记占比 30% 时的节点分类准确率,图 5(b)为标记占比 50% 时的节点分类准确率。从图中可以观察到,具有双通道的 KDCGNN 比其他 3 种变体表现出更好的性能。此外, KDCGNN 始终优于 KDCGNN (w/o 结构卷积),并且在 3 个数据集上与 KDCGNN 的差异约为 1 个百分点,这表明结构卷积在提升节点嵌入的表征能力方面起着重要的作用。KDCGNN 比 KDCGNN (w/o 特征卷积)高 2 个百分点左右,这表明节点的结构图比特征图对节点分类的性能影响更大,进一步证实了在图神经网络中引入特征图的必要性。KDCGNN 始终优于 KDCGNN (w/o 一致性损失),说明一致性损失在该框架中起着更为重要的作用。由此可以得出如下结论: KDCGNN 的各个模块都发挥了作用,其中特征编码对总体性能的影响最大。

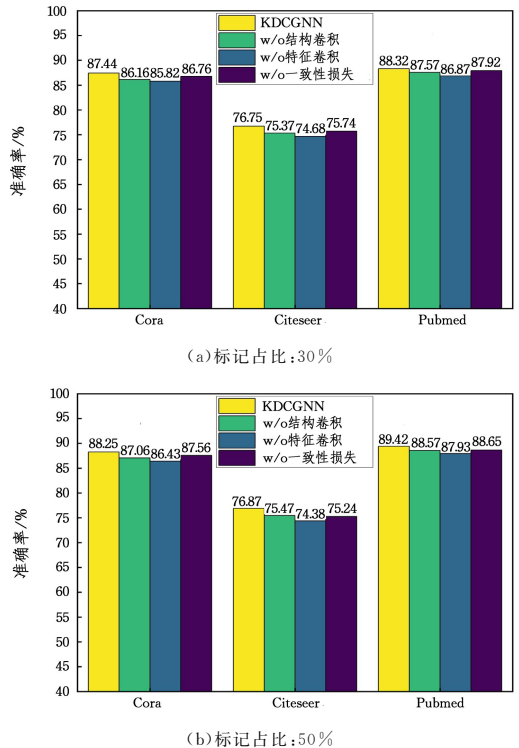


图 5 KDCGNN 不同变体的分类准确率对比
Fig. 5 Comparison of classification accuracy among different variants of KDCGNN

3.4 超参数敏感性实验

本研究中涉及了一种新的图相似度计算方法,为了验证其合理性,将相似度函数依次采用高斯核、线性核和高斯-Dice,然后在标记占比 30% 和 50% 的情况下比较其在 3 个数据集上分类的准确率。结果如图 6 所示,相似度函数采用高斯-Dice 能够取得更好的性能,分类准确率比线性核高 2%~3%,比高斯核函数高 1%~2%,这表明高斯-Dice 能够更准确地衡量节点间的相似程度。这主要是因为高斯-Dice 相似度能够在一定程度上消除节点度的影响,将高斯核函数中的缩放因子替换成 Dice 相似度,所得到的高斯-Dice 相似度同时包含了节点的特征和分布的相似程度。

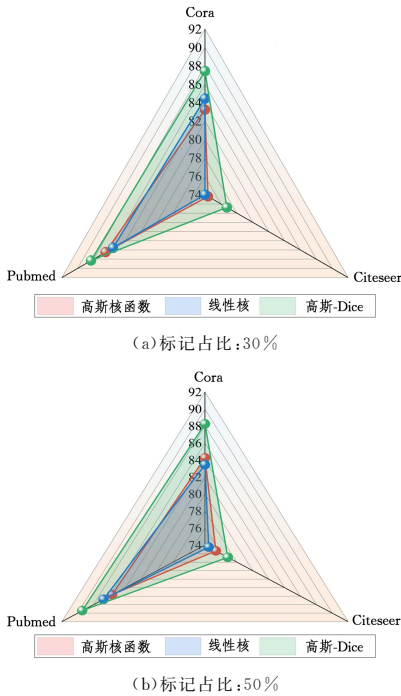


图 6 不同相似度计算函数下的分类准确度对比

Fig. 6 Comparison of classification accuracy under different similarity calculation functions

为了检验构建特征图中 top- k 邻域的影响,如图 7(a)所示,探究了 k 值在 2 到 10 之间的 KDCGNN 的性能,并且每个数据集训练了 10 次取准确率平均值作为最终的结果。对于 3 个数据集,准确率变化趋势都是先增加再下降,从图中每个数据集曲线的有颜色的带状区域可以看出,上下变化的幅度总体比较稳定。其中 Cora 数据集的准确率在 k 为 3 时达到最大值,Citeseer 数据集的准确率在 k 为 4 达到最大值,而 Pubmed 数据集则是在 k 为 5 时达到最大值。规模越大的数据集,其准确率最大时对应的 k 值就越大。这可能是由于图更密集,特征更容易平滑,同时,更大的 k 值可能会引入更多的噪声边缘,而节点的数量越多,抗干扰能力越强。

为了探究 KDCGNN 中 KAN 替换 MLP 是否能提升节点分类的准确率,在图 7(b)中对比了使用 MLP 和 KAN 两种情况下各个数据集的准确率变化情况。可以看出,在采用 KAN 的情况下,节点分类的准确率更高,在各个数据集上平均提升了 1%~2%。这表明 KAN 的学习能力更强,能够用极少的参数模拟复杂函数。KAN 解决了使用激活函数导致的信息

丢失问题,并且增强了模型的可解释性。

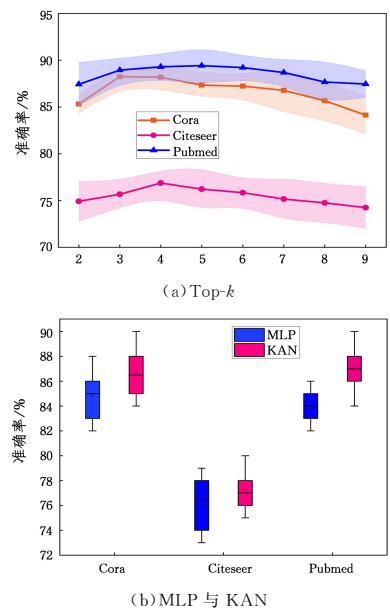


图 7 top- k 邻域与 KAN 对模型性能的影响

Fig. 7 Influence of top- k neighborhoods and KAN on model performance

结束语 本文提出了 KDCGNN,其通过结构卷积和特征卷积提取图的结构信息和特征信息,并利用 KAN 提升嵌入表示的分类性能和可解释性。此外,引入一致性损失函数,鼓励结构编码和特征编码之间的分布一致性,从而进一步增强了模型的泛化能力。实验结果表明,KDCGNN 在节点分类任务中优于现有基准方法,尤其在标记数据较少的情况下表现出显著优势。未来工作将重点研究 KAN 激活函数的选择及优化,并探索模型在图分类、链路预测等任务中的应用潜力,以进一步提升模型的性能和泛化能力。

参考文献

- [1] NIU M, WANG C, ZHANG Z, et al. A computational model of circRNA-associated diseases based on a graph neural network: prediction and case studies for follow-up experimental validation [J]. BMC Biology, 2024, 22(1): 24.
- [2] NIU M, ZOU Q, LIN C. CRBPDL: Identification of circRNA-RBP interaction sites using an ensemble neural network approach [J]. PLoS Computational Biology, 2022, 18(1): e1009798.
- [3] KIPF T N, WELLING M. Semi-supervised classification with graph convolutional networks[J]. arXiv:1609.02907, 2016.
- [4] VELICKOVIC P, CUCURULL G, CASANOVA A, et al. Graph attention networks[J]. arXiv:1710.10903, 2017.
- [5] LIU Z, WANG Y, VAIDYA S, et al. Kan: Kolmogorov-arnold networks[J]. arXiv:2404.19756, 2024.
- [6] CYBENKO G. Approximation by superpositions of a sigmoidal function [J]. Mathematics of Control, Signals and Systems, 1989, 2(4): 303-314.
- [7] HORNIK K, STINCHCOMBE M, WHITE H. Multilayer feed-forward networks are universal approximators[J]. Neural Networks, 1989, 2(5): 359-366.

- [8] ZHANG F, ZHANG X. GraphKAN: Enhancing Feature Extraction with Graph Kolmogorov Arnold Networks[J]. arXiv:2406.13597, 2024.
- [9] KIAMARI M, KRISHNAMACHARI B. GKAN: Graph Kolmogorov-Arnold Networks[J]. arXiv:2406.06470, 2024.
- [10] BRESSON R, NIKOLENTZOS G, PANAGOPOULOS G, et al. Kagans: Kolmogorov-arnold networks meet graph learning[J]. arXiv:2406.18380, 2024.
- [11] DE CARLO G, MASTROPIETRO A, ANAGNOSTOPOULOS A. Kolmogorov-arnold graph neural networks[J]. arXiv:2406.18354, 2024.
- [12] AHMED T, SIFAT M H R. Graphkan: Graph kolmogorov arnold network for small molecule-protein interaction predictions [C]// ICML'24 Workshop ML for Life and Material Science: From Theory to Industry Applications. 2024.
- [13] ZHOU K, SONG Q, HUANG X, et al. Multi-channel graph convolutional networks[J]. arXiv:1912.08306, 2019.
- [14] ZHAO R, SHAO Z, ZHANG W, et al. A multi-channel multi-tower GNN model for job transfer prediction based on academic social network[J]. Applied Soft Computing, 2023, 142: 110300.
- [15] MENG L, YE Z, YANG Y, et al. DeepMCGCN: Multi-channel Deep Graph Neural Networks[J]. International Journal of Computational Intelligence Systems, 2024, 17(1): 41.
- [16] WANG R, LI F, LIU S, et al. Adaptive Multi-Channel Deep Graph Neural Networks[J]. Symmetry, 2024, 16(4): 406.
- [17] SEN P, NAMATA G, BILGIC M, et al. Collective classification in network data[J]. AI Magazine, 2008, 29(3): 93-93.
- [18] GROVER A, LESKOVEC J. node2vec: Scalable feature learning for networks[C]// Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016: 855-864.
- [19] 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. 2014: 701-710.
- [20] HAMILTON W, YING Z, LESKOVEC J. Inductive representation learning on large graphs[C]// Advances in Neural Information Processing Systems. 2017.
- [21] WU F, SOUZA A, ZHANG T, et al. Simplifying graph convolutional networks [C] // International Conference on Machine Learning. PMLR, 2019: 6861-6871.
- [22] ZHU J, ROSSI R A, RAO A, et al. Graph neural networks with heterophily[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2021: 11168-11176.
- [23] WANG T, JIN D, WANG R, et al. Powerful graph convolutional networks with adaptive propagation mechanism for homophily and heterophily[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2022: 4210-4218.
- [24] GAO H, JI S. Graph U-Nets[C]// International Conference on Machine Learning. PMLR, 2019: 2083-2092.
- [25] WU J, CHEN X, XU K, et al. Structural entropy guided graph hierarchical pooling[C]// International Conference on Machine Learning. PMLR, 2022: 24017-24030.



WANG Jinghong, born in 1967, Ph.D. professor, is a member of CCF (No. 58341S). Her main research interests include artificial intelligence, pattern recognition, machine learning and data mining.

(责任编辑:柯颖)