计算机科学
COMPUTER SCIENCE

基于图熵理论的图数据增强研究

富坤，崔静远，党兴，成晓，应世聪，李建伟

引用本文

富坤，崔静远，党兴，成晓，应世聪，李建伟. 基于图熵理论的图数据增强研究[J]. 计算机科学，2025，52(5)：149-160.

FU Kun, CUI Jingyuan, DANG Xing, CHENG Xiao, YING Shicong, LI Jianwei. Study on Graph Data Augmentation Based on Graph Entropy Theory [J]. Computer Science, 2025, 52(5): 149-160.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.240200016

基于图熵理论的图数据增强研究

富　坤[1]　崔静远[1]　党　兴[2,3]　成　晓[2,3]　应世聪[1]　李建伟[1]
1 河北工业大学人工智能与数据科学学院　天津 300401
2 天津航天机电设备研究所　天津 300462
3 天津市宇航智能装备技术企业重点实验室　天津 300462

摘　要　图数据增强是一种通过变换和扩充图结构和节点特征来增加训练数据多样性、提高图神经网络性能的技术。为了应对图数据增强面临的难以综合信息完整性、特征平滑性、图多样性和局部依赖关系的挑战,缓解图神经网络的过平滑和过拟合问题,提高其性能,提出了一种基于物理热力学中的熵理论的图数据增强模型(Neighbor Replacement Based on Graph Entropy,NRGE)。首先,引入了一种新的图熵定义,用于度量数据流形平滑度;基于减少图熵损失的思想,提出了一种新的数据增强策略,用于生成更多合适的训练数据。然后,通过增强节点的采样邻居,以保证数据增强的一致性;采用随机替换节点的一阶邻居为二阶邻居的方式,增加了数据增强的多样性。最后,引入了邻居约束正则化方法,通过约束增强后的邻居之间的预测一致性来提高模型性能。消融实验结果表明,通过保持三角形图案的信息结构,NRGE 模型能够有效降低图熵损失,从而改善学习效果。在 Cora,Citeseer 和 Pubmed 3 个公开数据集上进行了节点分类实验,相较于基准模型,NRGE 模型在 Cora 数据集上提升了 1.1%,在 Citeseer 数据集上提升了 0.8%,在 Pubmed 数据集上略微降低了 0.4%。结果表明,NRGE 模型有效改善了图神经网络的性能,提高了其泛化能力。

关键词:图熵;图数据增强;邻居替换;一致性和多样性;结构增强
中图分类号　TP391

Study on Graph Data Augmentation Based on Graph Entropy Theory

FU Kun[1],CUI Jingyuan[1],DANG Xing[2,3],CHENG Xiao[2,3],YING Shicong[1] and LI Jianwei[1]
1 School of Artificial Intelligence and Data Science,Hebei University of Technology,Tianjin 300401,China
2 Tianjin Institute of Aerospace Mechanical and Electrical Equipment,Tianjin 300462,China
3 Tianjin Key Laboratory of Aerospace Intelligent Equipment Technology,Tianjin 300462,China

Abstract　Graph data augmentation,as a technique aiming to enhance the performance of graph neural networks,involves transforming and expanding the graph structure and node features to increase the diversity and quantity of training data. The integrity of information structures,the smoothness of feature manifold,the diversity of graph,and local dependencies are difficult to comprehensively considered in graph data augmentation. Additionally,over-smoothing and over-fitting problems exist in the training of graph neural networks,which limit their learning capabilities. To address these issues,a graph data augmentation model (NRGE) based on the entropy theory in thermodynamics is proposed. Firstly,a novel definition of graph entropy is introduced to measure the smoothness of the feature manifold. A new data augmentation strategy,whose main idea is to reduce the loss of graph entropy is proposed to generate more appropriate training data. Secondly, the sampling neighbors of the nodes are augmented to ensure the consistency of data augmentation. To increase the diversity of data augmentation,the first-order neighbors of nodes are randomly replaced with their second-order neighbors. Finally,a neighbor-constrained regularization method is introduced,which improves model performance by enforcing prediction consistency between augmented neighbors. Ablation experiments show that the NRGE model effectively reduces the loss of graph entropy by preserving the information structure of triangles,thereby improving learning effect. Three real datasets are trained by the NRGE model. The obtained low-dimensional representation is applied to node classification. Compared with the baseline methods,the NRGE model achieves a performance improvement of 1.1% on the Cora dataset,0.8% on the Citeseer dataset,and a slight decrease of 0.4% on the Pubmed dataset. The experimental results show that the NRGE model can significantly enhance the performance of graph neural networks and improve the generalization ability.

Keywords　Graph entropy,Graph data augmentation,Neighbor replacement,Consistency and diversity,Structural enhancement

到稿日期:2024-02-02　返修日期:2024-05-25
基金项目:国家自然科学基金(62072154);天津市科技计划项目(22JCYBJC01740);河北省重大科技成果转化专项(22280803Z)
This work was supported by the National Natural Science Foundation of China(62072154),Tianjin Science and Technology Plan Project (22JCYBJC01740) and Hebei Province Major Science and Technology Achievement Transformation Special Project(22280803Z).
通信作者:富坤(fukun@hebut.edu.cn)

1 引言

随着社交网络、引文网络、生物网络、化学网络和交通网络等各种网络在现实生活中的普及，它们的应用越来越广泛，对图数据结构和图神经网络的研究变得愈发重要。网络表示学习算法[1]是一种用于学习网络中节点的低维表示的机器学习算法。学习到的表示能够更好地表达节点的特征和语义信息，为后续的分析任务和应用提供基础。图神经网络(Graph Neural Network，GNN)[2]是一种特定的网络表示学习算法，它通过学习节点的表示来捕捉图结构中的信息。图卷积网络(Graph Convolutional Network，GCN)是图神经网络的代表，它通过消息传递或特定邻域聚合的方式，从节点及其邻域中提取高级特征。

过拟合指模型在训练集上表现得很好，但泛化能力较差的情况。过拟合会导致模型过度依赖训练数据的特定噪声或样本分布，而无法准确地推广到新的节点集。特别是在图机器学习中，由于特征数据的不完整性、结构数据的稀疏性以及标记数据的缺乏，而更容易出现过拟合现象。过平滑[3]是图学习面临的另一个问题，具体指随着网络深度的增加，GNN将节点的输出表示从输入特征中过度分离。过平滑的原因在于GNN采用了局部邻居聚合的方式，每一层的节点表示都是通过聚合邻居节点的特征得到的。随着网络深度的增加，邻居节点的信息会被多次聚合，导致节点表示逐渐趋于相似，无法区分不同的节点。

近年来，数据增强(Data Augmentation，DA)技术提升了计算机视觉(Computer Vision，CV)和自然语言处理(Natural Language Processing，NLP)[4]在数据处理领域中基于数据驱动的推理的泛化能力和预测性能，具有有效增强学习模型的性能和减少计算开销的优点。DA技术为解决过拟合问题提供了有效手段，通常通过标签的变换，如图像的平移和反射，在不增加真实标签的情况下有效扩大训练集，从而增强学习模型的泛化能力并抑制数据中的噪声。DA技术可以很好地处理数据中的噪声和数据的稀疏性，以及生成合理的增强样本，而不是引入更多的噪声或丢失关键信息。

图数据增强(Graph Data Augmentation，GDA)技术可以应用于GNN，以提高其准确性和泛化性[5-6]，但GDA面临着与传统的CV和NLP数据增强不同的挑战。数据增强需要考虑图的结构和拓扑关系，对于节点分类任务，增强策略需要保持图的连通性以及节点之间的关系，并且不能改变图的整体特征。GDA通常还需要对图数据的特征和预测结果进行解释和调整，以便更好地理解和分析图神经网络的训练过程。此外，GDA还需考虑在生成新的图样本的过程中引入更多的图局部结构变化和噪声，使得模型可以更好地捕捉节点之间的局部结构信息，以减轻过平滑现象。

物理学中的熵是描述系统混乱程度或无序程度的一个概念，它可以用来衡量系统的不确定性。当系统的状态多样且随机性高时，熵值较大；当系统的状态较为有序或确定时，熵值较小。图熵则是在图理论中引入的一个概念，用于度量图的复杂程度或信息量。图熵可以衡量图中节点或边的分布情况，以及图的结构多样性。在图神经网络中，图熵可以用于评估图的平滑度和多样性。物理学中的熵理论可以引入到图学习中[7]，以此作为度量系统不确定性的特征[8]，再通过基于熵的宏观参数来定量分析真实网络的复杂性；本文通过对比研究不同的增强方法，分析它们在一致性和多样性[9]方面表现出不同优缺点的原因，以此为基础改进现有的图数据增强方法。

图熵综合了整个图中的节点特征信息，能够从全局角度对数据进行分析。在数据增强中，生成的新样本需要尽量与原始样本的整体特征分布和结构保持一致。图熵作为度量指标可以较好地反映图中节点特征的一致性，它还可以直观地度量节点特征在图中的平滑性。通过减小图数据增强过程中的图熵损失，可以确保生成的样本在特征空间上与原始样本的平滑度相似，从而保持数据的连续性。

本文的主要研究贡献如下。

1)使用图熵来衡量图特征流形的平滑度，采用基于信息函数的方法，将节点集映射到正实数集，并构造图熵函数来表示全局特征信息分布的平滑度，从而指导图数据增强策略的设计。

2)引入一致性和多样性指标对数据增强方法进行对比评估。为评估多种基于数据增强策略的图神经网络模型的性能，使用验证集进行预测，通过准确率衡量一致性，通过差异性指标衡量多样性。综合考虑这两个指标，对比增强方法的优劣，以改进现有的图数据增强方法。

3)提出了NRGE图数据增强模型，通过保持基于三角形图案的信息结构的完整性来保持数据流形平滑性，通过对邻居进行增强以保持一致性，使用二阶邻居代替一阶邻居以提高增强的多样性，并使用邻居约束正则化方法在训练中使用未标记的节点，防止模型过拟合。

2 相关工作

本章主要介绍图神经网络以及图数据增强的发展现状。

GNN[10]通常采用递归消息传递的方式对图结构数据进行建模，由于其在节点分类[11-12]、链接预测[13-15]和图分类[16]等下游任务中表现出色，因此被广泛应用于图结构数据的表示学习，成为现代深度图学习(Deep Graph Learning，DGL)模型的主要组成部分。GCN[2]是一种典型的图神经网络，它利用消息传递或特定邻域聚合的方式，从节点及其邻域中提取高级特征。DeepGCNs[17]受深度卷积神经网络在图像分类方面的成功启发，提出了深层次的GCN。GraphSAGE[11]是一种基于图神经网络的算法，其通过对节点和边的信息进行采样和聚合，实现了对图数据的特征学习和分类，它通过捕捉节点之间的局部和全局结构信息，为节点分配全局特征向量。GAT(Graph Attention Network)[12]是一种基于注意力机制的图神经网络算法，它通过引入注意力机制，对节点和边的特征进行加权，提高了图神经网络的学习效率和准确性。GAT算法通过多头注意力机制，将节点和边的特征信息进行聚合，为每个节点生成一个加权特征向量。除此之外，GCAE[18]是一种结合了图卷积网络和自编码器的算法，其通过编码和解码图数据，实现图数据的特征提取和降维。

最近，图数据增强(Graph Data Augmentation，GDA)被

用来提高 GNN 的准确率和泛化能力[19]。GDA 通常是通过对图拓扑结构进行随机重排、子图采样、节点删除或添加等操作，来生成新的图样本。首先，可以引入更多不同的图结构，使得模型能够学习更多不同节点之间的局部关系。其次，可以引入一定程度的随机噪声，例如对节点特征进行随机扰动或添加噪声，这样可以使得节点表示在训练过程中不断变化，增加模型对局部结构的敏感性。此外，数据增强还可以通过标签传播或标签扩展等技术，将标签信息从已标记节点传播到未标记节点，从而增加训练集中的标记数据，这样可以提供更多的节点类别信息，帮助模型更好地区分不同节点。

Dropout[20]是一种在深度学习中常用的正则化方法，通过在每个训练批次中随机丢弃部分神经元来防止过拟合和提高模型泛化能力，也可以作为一种有效的图数据增强方法。DropEdge[21]通过随机删除图中的边来增加图的多样性和复杂性，它简单易行，不需要额外的图结构或节点特征，而且可以与其他数据增强方法结合使用，为解决复杂图问题提供更多的启示和可能性。GRAND[22]通过随机删除图中的节点及其相关的边来增加图的多样性和复杂性。GraphMix[23]是一种基于图神经网络的混合采样方法，通过将不同类型的数据（如节点特征、边信息等）进行混合，生成新的图结构。NodeAug[24]通过随机选择节点及其邻居节点进行替换或增加新的节点和边，以增加图的多样性和复杂性。GraphVAT[25]是一种基于变分自编码器和注意力机制的图数据增强方法，通过学习节点特征的分布和潜在结构，生成具有与原始图相似结构的虚拟图。GAUG[26]通过随机删除、添加和更改图中的节点和边来生成新的图结构。

3　基于图熵的邻居替换算法

图熵可以作为揭示系统相关信息的特征度量方法，使用物理学中熵的方法由宏观参数来定量表征图信息[27]。具体而言，采用基于信息函数的方法将节点集映射到正实数集[28]，并构造一个图熵函数来表示全局特征信息分布的平滑度。此外，还使用了针对图数据增强的度量方法对现有的数据增强方法进行比较和评估。通过对不同增强方法进行对比研究，以及图熵理论的指导，对现有的图数据增强方法进行改进，提出了新的图数据增强方法。通过保留图中特定的主题结构，对其他节点进行采样，并进行邻居替换，以在保持图熵的基础上增强图数据。

3.1　图熵理论

传统的图数据增强方法通常采用局部的邻域信息进行图的变换或扩充，常见的方法包括随机删除或添加节点[29]、随机断开或建立边连接[21]，以及节点属性的随机替换[20,22]等。这些方法都存在一个共同的不足之处，即它们没有充分考虑图中特征的全局分布，而只关注了局部信息。图中的节点往往存在远程的依赖关系，忽视了全局特征分布可能导致无法捕捉到节点之间的远程依赖关系，从而影响对图分析的准确性；此外，忽视全局特征分布可能导致模型泛化能力下降，因为它没有学习到全局的一致信息和模式。因此，综合考虑全局和局部特征分布对改进图数据增强的效果非常重要。

由于真实的图数据网络具有数据量大和结构复杂的特性，从宏观上定量表征图信息非常重要。熵在物理学中被用来度量整个系统的不确定性或混乱程度[30]；在图像分割中也使用熵的概念来量化图像各个区域中纹理的平滑度，高熵表示纹理更平滑并且图形块更少发生突变，即目标图像中包含更多信息，表现出更均匀的分布。在图数据分析中引入熵的概念来量化节点特征，可以更好地表示全图的特征分布。通过计算图中节点上每个特征维度的熵，得到全图特征分布的定量表示，再将这些定量表示用于分析复杂图结构的全局特征。基于熵的图数据处理可以更好地表征图的整体特征分布，将其引入图数据增强方法可以改善增强效果。

图熵被广泛用于根据一般拓扑或特征定量描述和理解图的动态。它最早由 Rashevsky[30]提出，然后 Mowshowitz 等[31]研究图熵来衡量图的结构信息内容，Körner[32]将图熵的不同定义应用到编码理论中 。大多数图熵源自基本的香农熵定义，其详细信息如下。对于离散系统 X，$I(x_i)=-\log p(x_i)$表示 $x_i\in X$ 的自信息，其中 $p(x_i)$为离散系统 X 取值为x_i的概率。当概率分布中的所有概率都相等时，香农熵最大，表示最大的不确定性或信息量；当概率分布中的某些概率接近于 1，而其他概率接近于 0 时，香农熵最小，表示最小的不确定性或信息量。系统 X 的熵为 $H(X)$：

$$H(X)=-\sum_{i=1}^{|X|}p(x_i)\log p(x_i) \tag{1}$$

先将每个节点的特征向量视为一个个体，然后将全体节点的特征集合构成一个特征向量空间，之后对每一个节点计算其概率。在计算概率函数 $p(x_i)$时，先通过基于信息函数 f 的计算和转换将节点集的特征映射到正实数集[28]，即将节点集的特征转换为数值形式，再基于信息函数的计算得到相应的概率函数，其计算式如下：

$$p(x_i)=\frac{f(x_i)}{\sum_{j=1}^{|X|}f(x_j)} \tag{2}$$

信息函数的设计取决于关注的图特征和研究的问题。在研究图中节点的重要性和影响力时，信息函数可以设计为度中心性（节点的度数）、介数中心性（节点在最短路径中频繁出现的次数）或其他中心性指标的函数形式。当研究图中的社区划分和模块化结构时，信息函数可以基于节点所属的社区标签，或者节点与其他社区之间的连接强度等。当研究图的拓扑结构时，信息函数则主要考虑这些特征的统计量。

度量图中全局信息分布的平滑度可以了解图的结构和特征之间的关系，以及节点之间信息的传播和交互情况，这对于理解网络的整体性质、发现重要节点以及预测节点行为具有重要意义，平滑的全局信息分布意味着图中节点之间的信息传播较为均匀和稳定。本文设计了一种基于图全局信息分布的平滑度计算的信息函数。

特征扩散平滑度可以用于度量图神经网络中节点特征在图结构上的连续性[34]。如果邻居节点的特征与当前节点的特征相似或接近，那么特征扩散平滑度就较高；如果邻居节点的特征与当前节点的特征差异较大，那么特征扩散平滑度就较低。

为了度量节点之间的相互作用，可使用节点与其邻居之间的特征距离之和作为相似性度量[8]，以表征节点的局部

特征分布。具体计算如式(3)所示：

$$f(v_i)=\sum_{(v_i,v_k)\in\varepsilon}\langle \boldsymbol{X}_i,\boldsymbol{X}_k\rangle \tag{3}$$

这种计算方法可以直观地反映节点在特征空间中的相对位置和分布情况，较好地描述图的平滑性。然而，该方法也存在一些不足之处，如忽略了邻居之间的差异性和相关性、未考虑特征的相关性以及无法区分重要特征和次要特征等。因此，在将节点与其邻居之间的特征距离之和作为相似性度量时，引入邻居之间的加权特征距离可以提供更精确的相似性度量。

给不同邻居之间的特征距离赋予权重可以反映邻居的重要性和贡献度差异，更准确地表达节点与其邻居之间的相似性和局部特征分布，最终能够更好地反映节点特征空间的平滑度。其具体计算式如下：

$$f(v_i)=\sum_{(v_i,v_k)\in\varepsilon}\boldsymbol{W}_k\langle \boldsymbol{X}_i,\boldsymbol{X}_k\rangle \tag{4}$$

其中，$\boldsymbol{W}_k$是邻居节点距离的权重，采用节点间的余弦相似度进行计算。余弦相似度常用于衡量节点之间的相似性，它能够捕捉到节点特征之间的方向性信息。其通过将节点特征向量标准化为单位向量，弱化了向量长度的影响，确保了节点特征的数值幅度不会主导相似度计算，而特征向量的方向性信息则较受关注。通过计算特征向量之间的方向性信息，可以捕捉到多维特征之间的相似性，更准确地衡量节点之间的关系。这种特征的表示在处理高维和复杂数据时，能够提供更深入的特征解释以及更准确的相似性度量。W_k的计算式如下所示：

$$W_k=\frac{\langle \boldsymbol{X}_i,\boldsymbol{X}_k\rangle}{\|\boldsymbol{X}\|\cdot\|\boldsymbol{X}\|} \tag{5}$$

其中，$\langle \boldsymbol{X}_i,\boldsymbol{X}_k\rangle$表示向量$\boldsymbol{X}_i$和向量$\boldsymbol{X}_k$的内积，$\|\boldsymbol{X}_i\|$表示向量$\boldsymbol{X}_i$的长度。

随后，根据信息函数计算图中的每个节点概率值：

$$p(v_i)=\frac{f(v_i)}{\sum_{j=1}^{|V|}f(v_j)} \tag{6}$$

最终将概率函数代入熵的计算式(1)可以得到一种新型的图熵 $I(G)$。其中 G 表示图，其计算式如下所示：

$$I(G)=-\sum_{i=1}^{|v|}p(v_i)\log p(v_i) \tag{7}$$

通过计算 $I(G)$来量化特征分布的随机性，表示局部特征对全局的贡献。如果 $I(G)$达到相对较高的值，则特征信息倾向于更加均匀地分散在图中，即图中特征信息扩散更加平滑，反之亦然。因此，该新型图熵 $I(G)$可以用来较好地计量图中特征信息扩散的平滑度。

在图数据增强过程中，通过最大化图整体的特征扩散平滑度，可以确保增强操作不会破坏原始图的连续性和一致性。这有助于保持图的结构完整性，防止信息的丢失或扭曲；有效提取图的全局结构和属性信息，从而改善节点特征的表达能力[30,32]。

度量特征扩散平滑度的方式主要有图熵、平均绝对差、方差、局部邻域的梯度变化率等[33]。选择图熵来度量特征扩散平滑度主要有 3 个方面的优势。1)图熵对节点特征的变化非常敏感，可以较灵敏地反映节点特征在图结构中的平滑性和连续性。2)图熵是一种全局性的度量指标，它综合了整个图中的节点特征信息。通过计算图熵，可以考虑到图中所有节点的特征传播和扩散情况，从而对整体的特征平滑度进行评估，这有助于提取图的全局结构。3)图熵作为一种度量方式具有直观性，它可以被理解为图中节点特征的分布均匀程度。当图熵较高时，表示节点特征在图上分布较为均匀，即特征扩散较为平滑；而当图熵较低时，表示节点特征在图上分布不均匀，即特征扩散不够平滑。

3.2 基于图熵的图数据增强研究

当前的图数据增强方法主要包括两种：拓扑增强方法和特征增强方法。拓扑增强方法，如 DropNode 和 DropEdge，会破坏特定主题的拓扑结构，同时去除主题结构上的特征；而特征增强方法，如 Dropout 和 GRAND，会破坏附加到主题结构的特征，但并不扰乱图的拓扑结构。为了研究不同增强策略对图熵的影响，对各种增强后的图和原始图进行了图熵对比。结果显示，原始图的图熵最高，原因在于其拓扑和特征都没有受到损坏，而增强后的图因为破坏了图数据的拓扑结构和特征信息，会导致图熵的降低。

当前，邻居替换算法是一种可以较好地提高模型泛化性的图数据增强方法，但常用的邻居替换算法没有考虑其对图特征扩散平滑度的影响。本文基于图熵理论，研究了邻居替换算法对图特征扩散的平滑度的影响。先对每个数据集中节点的邻居进行采样，随机将其一阶邻居替换成二阶邻居，再计算进行邻居替换前后的图熵。通过在 Cora[35]，Citeseer[36] 和 Pubmed[37] 图数据集上进行实验来分析邻居替换算法对图熵的影响，从实验结果可知，邻居替换算法虽然可以有效地对图数据进行增强，但明显降低了图熵，且明显影响了图整体的特征扩散平滑度。

为了得到更好的数据增强效果，在图数据增强的过程中需要尽量保持图熵的稳定。文献[38]指出，作为高阶连接模式的图形结构对于图拓扑的构建非常重要，在特定图形结构中，来自局部密集节点的特征被聚集成一个整体来表达信息，这些图形结构在图熵维持中发挥着重要作用。图数据中存在的典型图形结构如图 1 所示，不同的图形结构对于图熵的维持效果存在差异。研究人员开展了图形结构对图熵影响的相关研究[8]，针对 5 个典型的场景：原始图，仅保留三角形、四边形、五边形或五跳链的图。

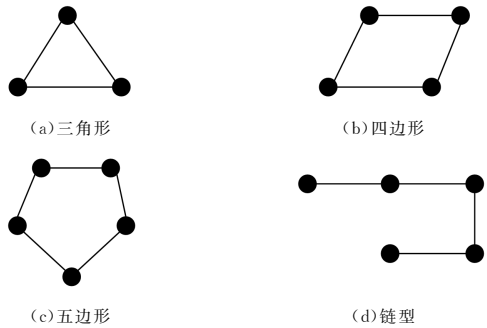

图 1 4 种主体结构

Fig. 1 Four main structures

如图 1 所示，除了图形结构覆盖的节点外，将每个节点上的特征向量置为零向量。不同场景下图熵的计算结果如表 1 所列。结果表明，与四边形、五边形或链相比，图中的三角形

结构在图熵维持方面表现出明显的优势，计算得到的图熵仅略低于原始图熵，但远高于其他图形结构。

表 1　保留不同结构时图熵的计算结果

Table 1　Calculation results of graph entropy when preserving different structures

Datasets	Original	Triangle	Square	Pentagon	5-Hop Chain
Cora	7.4525	7.4016	7.0189	7.0400	6.7788
Citeseer	7.7212	7.4943	6.7188	6.5592	7.0282
Pubmed	9.0150	8.9891	8.4884	8.4661	8.4554

三角形作为图论中的完整子图或聚类算法中的团，具有较好的连通性，在图拓扑结构的构建块中发挥着重要作用[37-39]。为了维持图数据的图熵，在图数据增强时无须考虑其他高阶完整子图。固定三角形结构是一种有效的策略[40]。首先，三角形结构往往代表了一些重要的关联。例如，在社交网络中，三角形可以表示互为好友的关系；在生物网络中，三角形可以表示基因之间的相互作用等。在图数据增强过程中保留三角形结构，可以保持这些重要结构的完整性，从而减少重要图的拓扑结构的改变造成的信息丢失。其次，节点之间的信息可以在三角形中更快地传递，减少信息的丢失和混淆。在图数据增强过程中保留三角形结构，可以最大程度地保留信息传递的最短路径，减少信息的损失。最后，相对于其他高阶完整子图，固定三角形结构也更易于实现。这不仅可以降低算法的复杂性，并且更容易应用于不同的图数据增强方法。三角形结构是高频率出现的，有利于保持增强前后图结构的一致性。因此，保留三角形信息结构的完整性在图数据增强策略设计中具有重要的意义。

基于以上研究，提出了一种基于图熵理论的邻居替换算法，该增强策略能够最大程度地保留图的信息和图熵。两种增强算法对维持图熵的稳定性表现如表 2 所列。

表 2　两种增强算法对图熵的影响

Table 2　Impact of two enhancement algorithms on graph entropy

算法	Cora	Citeseer	Pubmed
原始图	11.51	12.03	14.24
邻居替换	9.78	10.33	12.60
基于图熵的邻居替换	11.11	11.61	13.85

3.3　基于图熵理论的邻居替换算法

基于图熵理论的邻居替换算法（NRGE）模型的细节，如图 2 所示，该模型主要由两个部分组成：增强和正则化。增强部分主要由挖掘三角结构和邻居替换两个算法构成。在增强中，通过保留图中的三角结构以促进增强的一致性，使用二阶邻居代替一阶邻居以促进增强的多样性；在正则化中，使用邻域约束正则化方法来限制增强的预测结果。

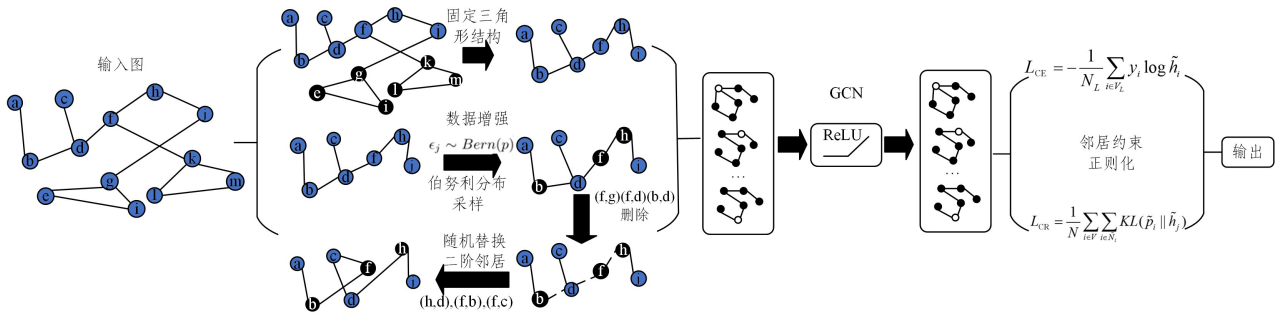

图 2　基于图熵理论的邻居替换算法示意图

Fig. 2　Schematic diagram of neighbor replacement algorithm based on graph entropy theory

邻居替换是一种适用于图数据增强的技术，它在图表示学习、图卷积神经网络和图生成等领域得到广泛应用。邻居替换的基本思想是通过替换图中节点的邻居节点，生成新的图数据。具体而言，对于每个节点改变节点的邻居关系，这样可以生成与原始图具有相似结构但略有差异的新图。

在 NASA[9] 中，使用了邻居替换的算法对图数据进行增强。采用邻居替换算法具有两个优势：首先，交换一阶邻居和二阶邻居可以扰乱图结构，但不会严重损害准确性；其次，邻居替换可以将监督信号传播到更多未标记节点中，从而促进泛化能力的提升。然而，该算法虽然可以有效地对图数据进行增强，但它没有考虑其对图特征扩散平滑度的影响，明显降低了图熵，影响了图整体的特征扩散平滑度。为了获得更好的增强效果，本文基于图熵的研究，采用保留三角形图形结构完整性的增强策略来改进邻居替换算法。改进算法首先挖掘图中的三角形结构，通过遍历每个节点及其邻居节点，检查它们之间是否存在边，进而确定是否存在三角形结构。如果存在三角形结构，则将相关节点添加到集合中，并在后续的增强过程中将其忽略；然后，对除三角形结构外的其他节点的邻居进行伯努利分布采样，再将采样后的一阶邻居节点随机替换为二阶邻居节点。算法伪代码如算法 1－算法 3 所示。

算法 1　基于图熵理论的邻居替换算法

输入：图 $\mathbf{G}=(V,E)$ 和特征矩阵 $\mathbf{X}$

输出：增强图和特征 $\widetilde{\mathbf{G}},\widetilde{\mathbf{X}}$

1. For $v_i \in V$ / * V 是图中的节点集合 * /
2.　{If 节点不在三角形结构中
3.　$\widetilde{\mathbf{G}},\widetilde{\mathbf{X}}$ = 邻居替换算法($\mathbf{G},\mathbf{X}$)
4.　End if}
5. End for
6. return $\widetilde{\mathbf{G}},\widetilde{\mathbf{X}}$

算法 2　邻居替换算法

输入：图 $\mathbf{G}=(V,E)$ 和特征矩阵 $\mathbf{X}$

输出：进行邻居替换后的图和特征矩阵 $\widetilde{\mathbf{G}},\widetilde{\mathbf{X}}$

1. 初始化加边和删边集合：add_E_v_i，add_E_v_j，del_E_v_i，del_E_v_j
2. For $\forall v \in V$
3.　{ $\forall u \in U$ //U 是节点 v 的邻居集合
4.　伯努利采样得到集合U。

```
5.    For ∀ u_i ∈ U_s
6.     {
7.       del_E_v_i += [v_i, u_i]//将要删除的边写到删边集合中
8.       U_i' = 随机选择 u_i一阶邻居(v_i的二阶邻居)
9.       add_E_v_i += [v_i, u_i']//将要添加的边写到加边集合中
10.    }
11.    End for
12.  }
13. End for
14. If del_E 非空:
15.  {根据删边集合移除旧边;
16.   根据加边集合添加新边;}
17. End if
18. return G̃, X̃
```

算法 3 挖掘三角形结构算法

输入:图 G=(V,E)

输出:在三角形结构中节点的集合 S

```
1. S=[ ] // 初始化三角形集合为空
2. For v_i ∈ V //V 是图中的节点集合
3.   neighbors=indices of non-zero elements in adj_matrix[i] // 获取第 i 个节点的邻居节点
4.   triangle=False // 初始化标志变量,用于判断是否存在三角形
5.   For j=0 to length(neighbors)-1 do // 遍历邻居节点
6.     For k=j+1 to length(neighbors)-1 do // 遍历邻居节点中后面的节点
7.     if adj_matrix[neighbors[j]][neighbors[k]] !=0 then// 判断邻居节点之间是否存在边
8.     triangle=True // 存在边,标志变量设为 True
9.     break
10.     If triangle then
11.     break
12.   If triangle then// 判断标志变量
13.     S += v_i// 添加三角形节点集合到集合 S
14. return S
```

3.4 模型优化方法

然而,基于邻居替换算法的增强可能会引入一些噪声,为了防止这些噪声对训练过程产生严重干扰,本文应用了两种技术改进模型:邻居约束正则化和动态训练。正则化架构如图 3 所示。

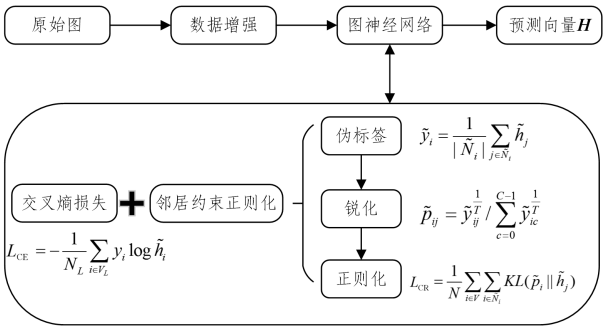

图 3 正则化架构图

Fig. 3 Regularization architecture

3.4.1 邻居约束正则化

邻居约束正则化能够利用未标记节点和其邻居节点之间的关系来提供额外的监督信号,从而提高模型在未标记节点上的预测性能。

对每个节点进行基于邻居替换算法的增强之后,将增强图 $\widetilde{A}$ 和原始节点特征矩阵 $\boldsymbol{X}$ 输入 GNN 中学习,得到节点预测向量 $\widetilde{\boldsymbol{H}}=Trans(Agg\{\widetilde{\boldsymbol{A}},\boldsymbol{X};\Phi\};\Theta)$,其中,$Trans$ 表示特征变换函数;Agg 表示聚合函数;Φ 是聚合函数的参数,它将输入的邻居特征进行组合,生成节点的中间表示;Θ 是图神经网络中权重和偏差等参数,这些参数通过反向传播算法更新,以最小化模型的预测误差。对于标记节点,使用交叉熵损失来监督 GNN 的预测:

$$L_{CE}=-\frac{1}{N_L}\sum_{i\in V_L}y_i\widetilde{\log h_i} \tag{8}$$

其中,y_i 代表节点的真实标签,h_i 代表节点的预测标签,N 代表节点的个数。

伪标签生成是为了在半监督学习中使用已有的标记数据来生成未标记数据的伪标签,从而扩充训练数据并提供额外的监督信号。这些伪标签可以用于训练模型,以增强其性能。利用邻居生成伪标签,可以更好地利用节点之间的相似性和关联性,从而提供更准确的伪标签。因此,本文使用邻居聚合和锐化技巧来改进伪标签生成策略。

首先,将邻居的预测融合为中心节点的伪标签 $\widetilde{y}_i=\frac{1}{|\widetilde{N}_i|}\sum\widetilde{h}_j$。其中,$N$ 是节点 i 的邻居集合,h_j 是节点 j 的预测标签。通过使用邻居预测的平均值作为结果,可以有效地抵消噪声偏大邻居的影响。

然后,利用锐化技巧来强制分类器输出低熵预测 $\widetilde{p}_{ij}=\widetilde{y}_{ij}^{\frac{1}{T}}/\sum_{c=0}^{C-1}\widetilde{y}_{ic}^{\frac{1}{T}}$。其中,$\widetilde{p}_{ij}$ 是锐化后的伪标签;$\widetilde{y}_{ij}$ 是节点的伪标签,表示第 i 个节点属于第 j 个类的概率,i 是节点的编号,取值为[1,N],j 是类别编号,取值为[1,C],C 是类别的个数;T 是一个缩放超参数,取值为(0,1),T 越小,在概率分布中高概率节点的影响将被进一步强化,锐化效果越明显。锐化技巧是一种通过调整分类器输出的置信度来提升分类器输出的准确性的方法,该方法可以使分类器的输出更加明确,减少模糊性,增强分类器对不确定性样本的置信度,有助于提高预测的准确性和可靠性。

随后,使用改进算法得到的伪标签来监督 GNN 的训练过程,邻居约束正则化损失 L_{CR} 设计为:

$$L_{CR}=\frac{1}{N}\sum_{i\in V}\sum_{j\in\widetilde{N}_i}KL(\widetilde{p}_i\parallel\widetilde{h}_j) \tag{9}$$

使用 KL 散度来测量伪标签数据 p 和节点预测标签数据 h 概率分布之间的距离。

最终的损失函数是交叉熵损失和邻居约束正则化的组合:

$$L=L_{CE}+\alpha L_{CR} \tag{10}$$

其中,α 是超参数,用于调整两种损失所占的权重。通过以上邻居约束正则化方法,在训练中引入了未标记节点的信息特征,防止了模型的过拟合。

3.4.2 动态训练

只在训练前对原数据进行一次增强称为静态训练,静态训练很容易受到极端增强的影响,模型学习效果较差。在训

练过程中，在每次迭代前都对数据进行增强，即每次迭代的增强图拓扑结构都是不同的。这种训练方式相对于静态训练被称为动态训练，如图 4 所示。

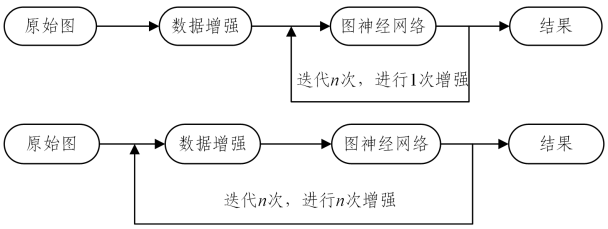

图 4　两种训练方式

Fig. 4　Two training methods

动态训练能提高模型的鲁棒性和泛化能力，使其对图的数据增强具有更好的适应性，避免过拟合，从而在面对不同的拓扑结构时实现更稳定的性能。

3.5　基于数据增强的图神经网络模型评估

接下来，将说明如何对增强策略进行评估。仅使用标记数据无法全面评估增强的质量，为了更好地衡量增强的正确性和泛化性，需要引入额外的数据如验证集，用于评估。

首先分别通过训练数据D_{train}及其增强$\widetilde{D}_{\text{train}}$训练两个模型$F_\theta,\widetilde{F}_\theta:\mathbb{R}^d\to\mathbb{R}^C$，其中 d 是输入特征的维度，C 是类的数量，θ 是参数。之后，使用这两个模型在验证集D_{val}上进行预测。如果增强具有更好的正确性和泛化性，则模型$\widetilde{F}_\theta$在验证集上应该具有更高的准确性。

为了衡量模型预测结果与初始标签的一致性，本文在验证集上使用增强模型进行训练，用预测准确度来量化一致性水平：

$$C=Acc(\widetilde{F}_\theta(D_{\text{val}}),Y_{\text{val}}) \tag{11}$$

其中，Y_{val}表示验证集的标签；Acc 代表准确率，它使用增强模型对验证数据集进行预测，预测其结果与真实标签相比的准确率。一致性较低意味着增强与原始数据不一致，可能会降低模型的准确性。然而，较高的一致性并不意味着增强的质量一定好，因为它可能对模型的泛化性贡献较小。

因此本文提出了另一个评估指标来衡量增强方法的多样性，即使用原始模型F_θ和增强模型$\widetilde{F}_\theta$的预测之间的差异来表示多样性水平：

$$D=\|\widetilde{F}_\theta(D_{\text{val}})-F_\theta(D_{\text{val}})\|_F^2 \tag{12}$$

其中，$\|\cdot\|_F$是矩阵的欧氏范数，表示增强模型和原始模型的预测之间的差异。如果增强模型和原始模型的预测结果相似，则它们之间的差异较小，欧氏范数的值会较低；相反，如果增强模型和原始模型的预测结果差异较大，则欧氏范数的值会较高。多样性水平较低表明增强数据与原始数据具有相似的分布，这可能不利于提高模型的泛化能力；但较高的多样性水平则容易降低增强的准确性。因此，需要将上述两个指标结合起来进行评估。

4　实验和结果分析

本章将通过实验来对 NRGE 进行评估。4.1 节介绍数据集、基线模型和相关的实验设置；4.2 节将 NRGE 与基线模型的实验结果进行对比并对实验结果进行分析；4.3 节对原始图、邻居替换、删除节点等图增强方法的一致性和多样性进行评估；4.4 节对 NRGE 进行消融实验，验证模型邻居替换、动态训练和锐化等相关模块对模型性能的影响；4.5 节对 NRGE 中的超参数进行敏感性分析。

4.1　实验设置

实验使用 Python 3.9，PyTorch 2.0.1，Numpy 1.25.2和 CUDA 12.8 作为计算环境，所有实验均在 Intel(R) Core(TM) i7-8750H CPU(2.2 GHz，16 核)，16 GB 内存和 NVIDIA GeForce GTX 1060 GPU 的服务器上进行。

1)数据集

使用现实世界中的 3 个引文数据集 Cora[35]，Citeseer[36]，Pubmed[37]和一个共同购买数据集 Amazon Photo[41]来评估所提模型，如表 3 所列，其中包括每个数据集中节点、边、特征和类别的数量。

表 3　数据集统计信息

Table 3　Dataset statistics

Dataset	Nodes	Edges	Features	Classes
Cora	2708	5429	1433	7
Citeseer	3327	4732	3703	6
Pubmed	19717	44338	500	3
Photo	7487	119043	745	8

2)基线模型

为了验证 NRGE 的性能，将其与一些主流的图神经网络进行比较，下面介绍用于比较的学习方法的详细信息。当前主流的图神经网络分为 5 大类：图传播算法类、图聚合网络类、图马尔可夫神经网络类、图卷积网络类，以及图数据增强类方法。图传播算法类方法通过在图上进行信息传播和标签传递来推断未标记节点的标签，从而实现半监督学习，如 LP[42]，GLP[43]和 GCN-LPA[44]。图聚合网络类方法通过采样和聚合邻居节点的特征来更新中心节点的表示，如 GraphSAGE[11]。图马尔可夫神经网络类方法通过建模节点之间的马尔可夫转移过程来学习节点的表示，如 GMNN[45]。图卷积网络类方法主要关注在图结构上进行信息传播和聚合，如 GCN[2]，GAT[12]和 MixHop[46]。图数据增强类方法通过在图中添加、修改或删除节点和边的方式来增强图的结构和特征，从而提高模型的鲁棒性和泛化能力，如 DropNode[28]，DropEdge[21]，Dropout[20]，GRAND[22]，GAUG[26]，NodeAug[24]和 NASA[9]。所有方法的超参数设置如下：学习率=0.01，权重衰减=1×10^{-3}，隐藏单元=32。对于基线模型，如果原论文提供了超参数，本文就按照作者的建议进行设置。

4.2　节点分类实验

本文使用 Cora[35]，Citeseer[35]，Pubmed[37]和 Photo[41]数据集进行节点分类实验。采用了不同的基线模型，并进行了 10 次运行，每次运行训练 500 个 epoch。记录了平均准确度和标准偏差，以便进行公平比较。为了确保改进来自于数据增强本身而不是高级 GNN，使用标准的两层 GCN 作为 NRGE 模型的主干，这样做可以确保对比的是数据增强的效果。同时，将 NRGE 应用于 GAT 以验证其在不同图神经网络中的效果。表 4 中，第一列是实验的算法名称，实验结果使用

平均准确率作为评估指标，括号内为标准差，粗体为最好结果。

表 4 节点分类实验结果

Table 4 Node classification experiment results

(%)

算法	Cora	Citeseer	Pubmed	Photo
LP	70.4(±0.0)	50.6(±0.0)	71.8(±0.0)	79.0(±4.8)
GLP	80.3(±0.2)	71.7(±0.6)	78.8(±0.4)	89.6(±0.7)
GCN-LPA	82.8(±0.1)	72.3(±0.2)	78.6(±0.2)	89.4(±1.5)
GCN	81.5(±0.3)	70.3(±0.9)	79.0(±0.2)	90.4(±0.7)
GAT	83.0(±0.7)	72.5(±0.7)	79.0(±0.3)	—
MixHop	81.9(±0.4)	71.4(±0.8)	79.8(±0.6)	—
GMNN	83.7(±0.3)	72.9(±0.5)	79.3(±0.4)	91.0(±2.9)
APPNP	83.8(±0.3)	71.6(±0.5)	79.7(±0.3)	90.6(±2.0)
GAUG	83.6(±0.5)	73.3(±1.1)	79.2(±0.3)	—
GRAND	84.5(±0.3)	74.2(±0.3)	80.0(±4.3)	91.7(±2.2)
NodeAug	84.3(±0.5)	74.9(±0.5)	**80.5(±0.5)**	92.3(±2.2)
DropEdge	84.4(±0.4)	73.4(±0.7)	79.1(±0.4)	89.4(±1.7)
DropNode	83.8(±0.5)	74.2(±0.3)	79.0(±0.4)	—
Dropout	83.7(±0.4)	73.5(±0.9)	78.8(±0.3)	—
GraphSAGE	83.4(±0.5)	72.2(±0.7)	79.5(±0.3)	—
NASA	84.7(±0.3)	74.9(±0.3)	79.4(±0.3)	92.7(±2.9)
NRGE-GCN	**85.8(±0.4)**	**75.7(±0.3)**	80.1(±0.2)	**92.9(±1.9)**
NRGE-GAT	84.9(±0.3)	75.2(±0.2)	79.4(±0.5)	92.9(±2.3)

表 4 列出了不同方法节点分类的性能，从中可以得出结论：基于标签传播的方法（LP，GLP）准确率低于其他方法（GCN，GAT 等），这是因为标签信息具有稀疏性，标签传播具有不准确性，因此仅利用标签传播的效果不佳。NRGE 在 Cora，Citeseer，Pubmed 和 Photo 数据集上分别将 GCN 的性能从 81.5%，70.3%，79.0%，90.4%提高到 85.8%，75.7%，80.1%，92.9%。同时，在 Cora，Citeseer 和 Pubmed 数据集上分别将 GAT 的性能从 83.0%，72.5%，79.0%提高到 84.9%，75.2%，79.4%。GAT 主要使用自注意力机制来对不同邻居节点的重要性进行加权。而 NRGE 模型在进行邻居替换时可能对 GAT 的自注意力机制产生负面影响，从而降低了数据增强的效果，进而导致对 GAT 增强的效果略弱于对 GCN 增强的效果。NRGE 在数据增强方法这一类模型中也有不错的表现，在 Cora 和 Citeseer 上的准确率分别较之前的最佳结果提高了 1.1%和 0.8%。最后，注意到 NRGE 在 Pubmed 上的性能弱于 NodeAug。这可能是因为相比于其他数据集（如 Cora 和 Citeseer），Pubmed 中节点之间的连接关系更多地反映了文献之间的作者合作关系，而非具体的主题相关性，其节点之间的连接关系并不明确地与节点的标签相关。而 NodeAug 方法采用了更适应 Pubmed 数据集特点的数据增强策略，使得模型能够更好地利用其他信息（如节点特征）进行分类，从而取得更好的性能。综上，NRGE 能够更好地捕捉节点之间的关系和图结构的特征，在多个数据集上都取得了更好的性能。

4.3 图增强方法的一致性和多样性评估

本节是在 Cora[35] 数据集上对几种图数据增强方法进行了一致性和多样性评估实验，其中 Origin 是原始图结果。一致性指标衡量了增强后的数据与原始数据之间的相似程度，多样性指标衡量了增强后的数据的多样性，指标的具体计算方法见 3.5 节，计算结果如表 5 所列。

表 5 一致性和多样性对比

Table 5 Comparison of consistency and diversity indices

方法	一致性指标	多样性指标
Origin	0.810	—
label	0.782	1.4962
dropnode	0.780	1.9999
dropout	0.788	1.8749
NASA	0.792	2.4498
NRGE	0.782	2.7944

由于多样性指标的计算是与基于原始数据的模型进行比较的，因此原始数据计算得到的多样性指标为 0。从实验结果可以看出，相比之前的 label，dropnode，dropout 和 NASA 方法，NRGE 在一致性指标相差一个百分点以内的基础上，多样性指标分别提高了 86.77%，39.73%，49.01%和 14.07%。NRGE 增强在一致性指标上与原始数据相比差别很小，说明增强后的数据与原始数据之间保持了一致性。NRGE 方法在多样性指标上相比其他增强方法有更好的表现，这是因为 NRGE 方法强调保持基于三角形图案的信息结构的完整性。通过保持这种信息结构，NRGE 能够确保增强后的数据与原始数据在全局特征信息分布上保持一致性；基于三角形图案的信息结构也有助于增加数据的多样性，因为不同的三角形图案代表了不同的局部结构和关系。此外，在增强过程中采用随机采样的方式从邻居中选择增强节点，这种随机性的引入有助于增加数据的多样性，使得增强后的数据在特征和拓扑结构上与原始数据有所差异。

4.4 消融实验

为了研究 NRGE 不同模块的有效性，在两个数据集 Cora[34] 和 Citeseer[36] 进行了消融实验：1）图增强方法消融实验，验证了 NRGE 的邻居替换和图熵增强模块，实验结果如表 5 所列；2）正则化方法消融实验，验证了 NRGE 的动态训练和伪标签锐化策略的有效性，实验结果如表 6 所列。实验结果使用平均准确率作为评估指标，括号内为标准差。其中，w/o augmentation 表示去除全部数据增强模块；w/o Motif-au 表示去除图熵增强模块，保留邻居替换模块；w/o NR-au 表示去除邻居替换模块，保留图熵增强模块。dropedge，dropnode，dropout 是 3 种典型的数据增强方法。

表 6 图增强方法消融实验结果

Table 6 Ablation experimental results of graph enhancement method

(%)

模块	Cora	Citeseer
w/ NRGE	85.8(±0.4)	75.7(±0.3)
w/o augmentation	83.8(±0.5)	74.0(±0.4)
w/o Motif-au	84.7(±0.3)	74.9(±0.3)
w/o NR-au	83.8(±0.3)	72.1(±0.6)
w/ dropedge	84.4(±0.4)	73.4(±0.7)
w/ dropnode	83.8(±0.5)	74.2(±0.3)
w/ dropout	83.7(±0.4)	73.5(±0.9)

实验结果表明，进行数据增强的模型相比没有进行数据增强的模型，准确率从 83.8%提高到 85.8%，提高了 2%，这表明数据增强在节点分类任务中起到了正向作用，提高了模型的性能。在 Cora[34] 数据集上，去除图熵增强模块后，准确率从 84.7%下降到了 83.8%；在 Citeseer 数据集上，准确率从 74.9%下降到了 72.1%。这说明图熵增强模块对于模型性能的提升起到了较为重要的作用。这是因为通过保持基于

三角形图案的信息结构的完整性，该模块有效地降低了图熵损失，有助于捕捉节点之间的高阶关系和复杂依赖，从而改善了学习效果。与此同时，邻居替换模块（NR-au）在保持增强数据多样性方面发挥了作用，通过随机替换节点的邻居，增加了训练数据的多样性，提高了模型的泛化能力。此外，图结构上的增强（如 NRGE 和 dropedge）比节点特征上的增强（如 dropnode 和 dropout）对模型的学习效果提升更有效。这是因为图的拓扑结构往往包含了重要的信息，节点特征上的增强方法主要关注节点的属性，无法捕捉到图结构的全局信息。

正则化相关的消融实验结果如表 7 所列，第一列说明了实验内容，其中，w/ dynamic training 表示采用动态训练，w/ static training 表示采用静态训练，w/o sharpening 表示去除伪标签锐化。

实验结果的前两行揭示了动态训练的优势，静态训练的准确率比动态训练低，标准差要高很多，在 Citeseer 数据集上更明显。这表明静态训练很容易受到增强的影响，与之相比，动态训练更稳定。最后一行说明了锐化模块的作用，去掉锐化模块后，准确率有所下降，这表明锐化在一定程度上对于改善模型的性能是有益的，锐化可以使模型对于增强后的数据更加敏感，能够更好地区分不同类别的节点。

表 7　正则化方法消融实验结果

Table 7　Ablation experimental results of regularization method

实验内容	Cora	Citeseer
w/ dynamic training	85.7(±0.2)	75.7(±0.3)
w/ static training	84.9(±0.7)	71.0(±5.1)
w/o sharpening	84.0(±0.5)	73.2(±0.5)

4.5　参数敏感性分析

在 NRGE 模型中，有 4 个主要的超参数：损失平衡的超参数 α（见邻域约束正则化章节）、脱出率 *dropout*、锐化超参数 *temp*（见邻域约束正则化章节）和采样概率超参数 *prob*（见邻居替换章节）。本文讨论了这 4 个超参数在 Cora[35] 数据集上的敏感性，横坐标表示参数的值，纵坐标表示实验准确率。4 个参数的敏感性分析结果分别如图 5 所示。

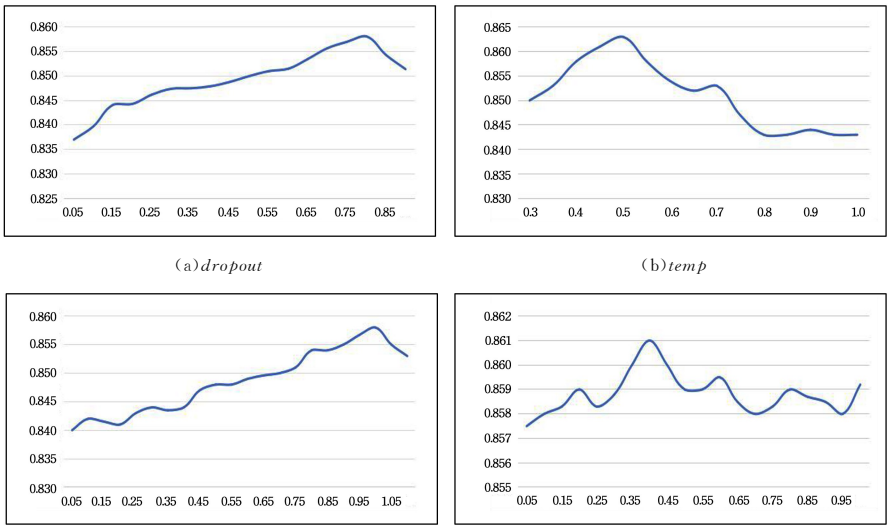

图 5　超参数敏感性

Fig. 5　Sensitivity of hyperparameters

图 5(a)展示了脱出率 *dropout* 从 0.1 到 0.9 时模型的分类准确率。*dropout* 从 0.1 到 0.8 逐步增加时，模型的分类准确率也逐步提升；当 *dropout*=0.8 时，分类准确率达到峰值，约为 85.7%；之后开始缓慢下降。这是因为随着 *dropout* 的增加，随机地丢弃更多神经元的输出，引入一定程度的噪声和随机性，增强了模型的鲁棒性，减少了对特定神经元的依赖，从而能够减少过拟合现象，提高模型的泛化能力。而当 *dropout* 过高时，会过度减少网络的容量，导致模型欠拟合，无法充分拟合训练数据，导致模型丢失一些重要的特征和模式，从而降低准确率。此外，过高的 *dropout* 率还会引入更多的不确定性和噪声，使得模型在每次训练时产生不同的结果。这种不稳定性会导致模型的性能波动。

图 5(b)展示了锐化超参数 *temp* 从 0.4 到 1 时模型的分类准确率。当 *temp*=0.5 时，性能达到峰值，准确率约为 86.3%，然后随着 *temp* 的提高而下降。较低的 *temp* 值会压缩分类器输出的分布，使得模型更加明确地对不同类别进行分类；相反，较高的 *temp* 值会扩展分类器输出的分布，使得模型更加模糊地对待不同类别。但是，过高的 *temp* 值会导致模型过于相信噪声或不重要的特征，从而降低模型的分类准确率；而过低的 *temp* 值可能导致模型过拟合，过度依赖特定的特征或样本，从而降低模型的泛化能力。因此，需要适当的锐化程度使得模型对于增强后的数据更加敏感，能够更好地区分不同类别的节点。

图 5(c)展示了损失平衡的超参数 α 从 0.1 到 1.1 时模型的分类准确率。当 $\alpha=1$ 时，性能达到峰值，分类准确率为 85.7%。较大的 α 值意味着正则化的权重更大，模型更倾向于选择较少的自由参数和较简单的模型结构，这有助于防止模型过度拟合训练数据，提高模型的泛化能力。然而，如果 α 值过大，正则化的惩罚会超过交叉熵损失的影响，导致模型在训练数据上的准确率下降，进一步导致模型分类准确率的下降。实验结果说明模型平衡交叉熵损失和正则化损失时，模型有更好的性能。

图 5(d)展示了采样概率超参数 *prob* 从 0.1 到 1 时模型的分类准确率。*prob* 为利用伯努利分布对节点邻居进行随机采样的概率，1 表示采样全部邻居。通过随机采样节点邻居，可以引入一定的多样性，从而增强模型的鲁棒性和泛化能力。当 *prob* 过低，如 0.1 时，只有很少一部分邻居被采样，模型受采样信息量的限制，分类准确率较低。当 *prob* 逐步增加到 0.4 时，模型的分类准确率逐步提升，在 0.4 时达到峰值，这是因为在该采样概率下，模型能够充分利用节点的邻居信息，同时引入适度的多样性，更好地捕捉数据的结构特征。然而，如果采样概率进一步增加，如接近 1 时，模型可能会过度依赖邻居节点的信息，导致过拟合。

为进一步研究超参数组合对模型的影响，本文在 Cora[35] 数据集上研究了不同的参数组合(*temp* 和 α、*prob* 和 *dropout*)对模型性能的影响，如图 6 所示。图 6(a)中，横坐标代表 α 的值，纵坐标代表 *temp* 的值；图 6(b)中，横坐标代表 *dropout* 的值，纵坐标代表 *prob* 的值。图中颜色越浅表示准确率越高，颜色越深表示准确率越低。

图 6(a)展示了超参数组合 *temp* 和 α 分别从 0.1 到 0.9 和 0.1 到 1.0 时模型的分类准确率。图 6(b)展示了超参数组合 *prob* 和 *dropout* 分别从 0.1 到 0.9 和 0.1 到 0.8 时模型的分类准确率。

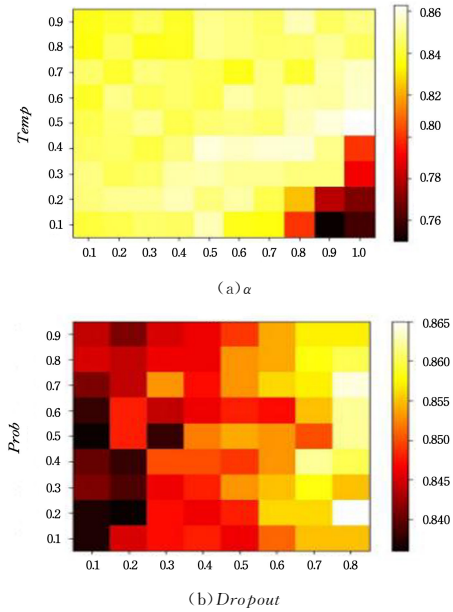

图 6 超参数组合

Fig. 6 Hyperparameter combination

从图中可以观察到，相比于 *temp* 和 α 的组合，*prob* 和 *dropout* 的组合对模型的影响程度更大。*temp* 和 α 的组合只要取值位于适当区间，模型就能有不错的准确率，并且准确率波动较小，而 *prob* 和 *dropout* 的组合对模型性能有较大影响。*temp* 和 α 的组合主要反映了正则化损失对模型的影响，如果 *temp* 和 α 的组合取 *temp* 的较大值和 α 的较小值，模型准确率就会大幅度下降，这是因为这种取值组合使模型对输入数据更加敏感，进而造成分类准确率不稳定。*prob* 和 *dropout* 的组合主要反映了对邻居节点的采样程度，模型通过采样邻居节点，可以扩展节点的上下文信息，更好地学习节点在图结构中的信息。因此，反映邻居节点采样的质量和数量的 *prob* 和 *dropout* 的组合对模型的性能具有更显著的影响，与 *temp* 和 α 的组合相比是需要重点关注的超参数组合。

4.6 可视化实验

本文使用含 NRGE 增强和不含 NRGE 增强的 GCN 在 Cora[35] 上进行图数据学习，再对两种学习的节点表示进行可视化实验，结果如图 7 和图 8 所示。

当不使用 NRGE 增强时，节点嵌入可视化的分布可能会呈现发散的趋势，边界不够清晰。这是因为在嵌入过程中，未经过 NRGE 增强的算法可能没有充分考虑节点之间的邻近关系，导致节点在嵌入空间中的位置分布较为分散。因此，通过使用邻居替换的数据增强算法，使得节点的邻近关系更加明显。在节点嵌入可视化中，增强邻近关系可以更好地展示节点之间的相似性和聚类结构。

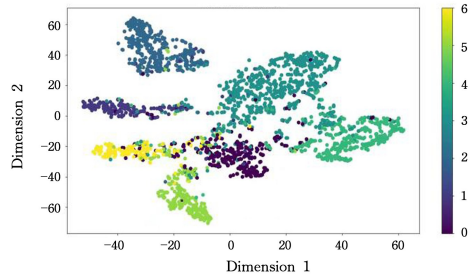

图 7 使用 NRGE 增强的 GCN 训练可视化结果

Fig. 7 Visualization of GCN training results with NRGE augmentation

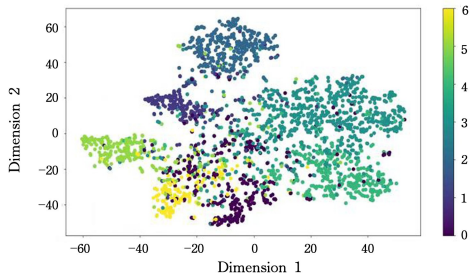

图 8 不使用 NRGE 增强的 GCN 训练可视化结果

Fig. 8 Visualization results of GCN training without NRGE augmentation

结束语 本文针对图神经网络半监督学习中的过平滑和过拟合问题，提出了一种基于图熵理论的图数据增强模型(NRGE)。该模型引入图熵来量化图中节点信息传播的平滑度，充当衡量全局特征信息分布的指标；通过保持基于三角形图案的信息结构的完整性来保持特征信息分布的平滑度，降低了图熵损失；该数据增强策略在保持一致性的同时增加了多样性，提高了图神经网络的性能。在 Cora，Citeseer，和 Pubmed 这 3 个公开数据集上的节点分类实验表明了 NRGE 的学习效果优于基线方法。在 Cora 数据集上设计的消融实验表明了通过保持三角形图案的信息结构，NRGE 模型能够有效地降低图熵损失，有助于捕捉节点之间的高阶关系和复杂依赖，从而改善学习效果。采用随机邻居替换作为增强方法，能够增加训练数据的多样性，提高模型的泛化能力。

然而，提出的模型仍有一些问题需要解决，例如，如何保

证该方法在其他更大、更复杂的图数据集上的表现。模型中的图熵损失是基于三角形图案的信息结构设计的，但不同的图数据集可能存在不同的信息结构特征。因此，进一步研究如何在不同数据集上设计更适合的信息结构，并探索更灵活的图熵损失函数，将是未来研究的一个重要方向。此外，考虑到图神经网络的实际应用场景，如何将 NRGE 模型与图中的动态变化和时序信息相结合，以处理动态图数据的节点分类任务，也是值得进一步研究的方向。

参 考 文 献

[1] ZHANG D，YIN J，ZHU X，et al. Network representation learning：A survey[J]. IEEE Transactions on Big Data，2020，6(1)：3-28.

[2] KIPF T N，WELLING M. Semi-Supervised Classification with Graph Convolutional Networks[J]. arXiv：1609.02907，2016.

[3] SUN K，LIN Z，ZHU Z. Multi-stage self-supervised learning for graph convolutional networks on graphs with few labeled nodes [C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2020：5892-5899.

[4] CUBUK E D，ZOPH B，MANE D，et al. Autoaugment：Learning augmentation strategies from data [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019：113-123.

[5] ADJEISAH M，ZHU X，XU H，et al. Towards data augmentation in graph neural network：An overview and evaluation[J]. Computer Science Review，2023，47：100527.

[6] ZHAO T，JIN W，LIU Y，et al. Graph Data Augmentation for Graph Machine Learning：A Survey [J]. arXiv：2202.08871，2022.

[7] EBRAHIMZADEH A，GISKI Z E，MARKECHOVÁ D. Logical entropy of dynamical systems－A general model[J]. Mathematics，2017，5(1)：4.

[8] LIU X，SUN D，WEI W. A Graph Data Augmentation Strategy with Entropy Preservation[J]. arXiv：2107.06048，2021.

[9] BO D，HU B，WANG X，et al. Regularizing Graph Neural Networks via Consistency-Diversity Graph Augmentations[J]. Proceedings of the AAAI Conference on Artificial Intelligence，2022，36(4)：3913-3921.

[10] HAMILTON W L，YING R，LESKOVEC J. Representation Learning on Graphs：Methods and Applications[J]. arXiv：1709.05584，2017.

[11] HAMILTON W，YING Z，LESKOVEC J. Inductive representation learning on large graphs[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017：1025-1035.

[12] DASOULAS G，SCAMAN K，VIRMAUX A. Lipschitz Normalization for Self-Attention Layers with Application to Graph Neural Networks[J]. arXiv：2103.04886，2021.

[13] ZHANG M，CHEN Y. Link prediction based on graph neural networks[C] // Proceedings of the 32nd International Conference on Neural Information Processing Systems. 2018：5171-5181.

[14] SCHLICHTKRULL M，KIPF T N，BLOEM P，et al. Modeling Relational Data with Graph Convolutional Networks[C]// ESWC. Cham：Springer International Publishing，2018：593-607.

[15] KIPF T N，WELLING M. Variational Graph Auto-Encoders [J]. arXiv：1611.07308，2016.

[16] YING Z，YOU J，MORRIS C，et al. Hierarchical graph representation learning with differentiable pooling[C]// Proceedings of the 32nd International Conference on Neural Information Processing Systems. 2018：4805-4815.

[17] LI G，MÜLLER M，THABET A，et al. DeepGCNs：Can GCNs go as deep as CNNs? [C]// 2019IEEE/CVF International Conference on Computer Vision(ICCV). IEEE，2019：9266-9275.

[18] XUE W，LI T. Aspect Based Sentiment Analysis with Gated Convolutional Networks[J]. arXiv：1805.07043，2018.

[19] DING K，XU Z，TONG H，et al. Data Augmentation for Deep Graph Learning：A Survey [J]. ACM SIGKDD Explorations Newsletter，2022，24(2)：61-77.

[20] SRIVASTAVA N，HINTON G，KRIZHEVSKY A，et al. Dropout：a simple way to prevent neural networks from overfitting [J]. The Journal of Machine Learning Research，2014，15(1)：1929-1958.

[21] RONG Y，HUANG W，XU T，et al. DropEdge：Towards Deep Graph Convolutional Networks on Node Classification[J]. arXiv：1907.10903，2019.

[22] FENG W，ZHANG J，DONG Y，et al. Graph random neural networks for semi-supervised learning on graphs[J]. Advances in Neural Information Processing Systems，2020，33：22092-22103.

[23] VERMA V，QU M，KAWAGUCHI K，et al. Graphmix：Improved training of GNNs for semi-supervised learning[C]// Proceedings of the AAAI Conference on Artificial Intelligence. 2021：10024-10032.

[24] 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 & Data Mining. ACM，2020：207-217.

[25] FENG F，HE X，TANG J，et al. Graph adversarial training：Dynamically regularizing based on graph structure [J]. IEEE Transactions on Knowledge and Data Engineering，2019，33(6)：2493-2504.

[26] PARK H，LEE S，KIM S，et al. Metropolis-Hastings Data Augmentation for Graph Neural Networks[J]. arXiv：2203.14802，2022.

[27] AMIGÓ J M，BALOGH S G，HERNÁNDEZ S. A brief review of generalized entropies[J]. Entropy，2018，20(11)：813.

[28] HUANG W，ZHANG T，RONG Y，et al. Adaptive sampling towards fast graph representation learning [J]. arXiv：1809.05343，2018.

[29] JIZBA P，KORBEL J. Maximum Entropy Principle in Statistical Inference：Case for Non-Shannonian Entropies[J]. Physical Review Letters，2019，122(12)：120601.

[30] RASHEVSKY N. Life，information theory，and topology [J].

The Bulletin of Mathematical Biophysics, 1955, 17(3): 229-235.
[31] MOWSHOWITZ A, DEHMER M. Entropy and the complexity of graphs revisited[J]. Entropy, 2012, 14(3): 559-570.
[32] KORNER J. Coding of an information source having ambiguous alphabet and the entropy of graphs[C]// 6th Prague Conference on Information Theory. Academia, Prague, 1971: 411-425.
[33] HUANG X, QI G, WEI H, et al. A novel infrared and visible image information fusion method based on phase congruency and image entropy[J]. Entropy, 2019, 21(12): 1135.
[34] WANG Y Q, WU M H, GENG F Q, et al. An Underwater Image Enhancement Algorithm Based on Image Entropy Linear Weighting[J]. Journal of Chongqing Technology and Business University(Natural Science Edition), 2024(4): 69-76.
[35] MCCALLUM A K. Automating the Construction of Internet Portals with Machine Learning[J]. Discover Computing, 2000, 3: 127-163.
[36] GILES C L, BOLLACKER K D, LAWRENCE S. CiteSeer: an automatic citation indexing system[C]// Proceedings of the third ACM conference on Digital libraries(DL'98). ACM, 1998: 89-98.
[37] SEN P, NAMATA G, BILGIC M, et al. Collective classification in network data[J]. AI Magazine, 2008, 29(3): 93-93.
[38] YIN H, BENSON A R, LESKOVEC J. Higher-order clustering in networks[J]. Physical Review E, 2018, 97(5): 052306.
[39] BENSON A R, GLEICH D F, LESKOVEC J. Higher-order organization of complex networks[J]. Science, 2016, 353(6295): 163-166.
[40] BURKHARDT P. Triangle Centrality[J]. arXiv: 2105. 00110, 2021.
[41] SHCHUR O, MUMME M, BOJCHEVSKI A, et al. Pitfalls of Graph Neural Network Evaluation[J]. arXiv: 1811. 05868, 2018.
[42] ZHOU D, BOUSQUET O, LAL T, et al. Learning with local and global consistency[C]// Proceedings of the 32nd International Conference on Neural Information Processing Systems. 2003.
[43] LI Q, WU X M, LIU H, et al. Label efficient semi-supervised learning via graph filtering[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 9582-9591.
[44] WANG H, LESKOVEC J. Unifying Graph Convolutional Neural Networks and Label Propagation[J]. arXiv: 2002. 06755, 2020.
[45] QU M, BENGIO Y, TANG J. GMNN: Graph Markov Neural Networks[C // /Proceedings of the 36th International Conference on Machine Learning. PMLR, 2019: 241-5250.
[46] ABU-EL-HAIJA S, PEROZZI B, KAPOOR A, et al. MixHop Higher-Order Graph Convolutional Architectures via Sparsified Neighborhood Mixing[C]// Proceedings of the 36th International Conference on Machine Learning. PMLR, 2019: 21-29.

FU Kun, born in 1979, Ph. D, associate professor. Her main research interests include network representation learning and social network analysis.

(责任编辑:何杨)