

基于结构学习和自监督图注意力的网络表示学习

王静红^{1,2,3}, 郑瑞策¹, 米据生⁴, 李昊康^{5*}

1. 河北师范大学 计算机与网络空间安全学院, 河北 石家庄 050024;
2. 河北师范大学 河北省网络与信息安全重点实验室, 河北 石家庄 050024;
3. 供应链大数据分析与安全河北省工程研究中心, 河北 石家庄 050024;
4. 河北师范大学 数学科学学院, 河北 石家庄 050024;
5. 河北工程技术学院 人工智能与大数据学院, 河北 石家庄 050091)

摘要: 网络表示学习是网络分析任务的基础, 对于挖掘和分析真实网络数据具有重大意义。最近, 图注意力网络(Graph Attention Networks, GAT)及其后续变体, 在网络表示学习中表现出了卓越的性能。但是基于注意力的方法存在以下局限性:(1)只考虑节点的一阶邻居信息, 忽略了高阶邻居节点。(2)模型缺乏可解释性。(3)没有考虑到图中噪声边问题。为解决这些问题, 本文提出了一种基于结构学习和自监督图注意力的网络嵌入模型(Structural Learning-based Self-supervised Graph Attention Network, SL-SGAT), 融合节点特征与结构信息, 降低噪声边干扰, 提升模型可解释性。SL-SGAT主要包含三部分: 图结构学习、自监督注意力机制和特征聚合。图结构学习构建全局图结构网络。自监督注意力机制设置一个自监督关系预测任务, 加入噪声边损失。特征聚合利用注意力系数进行加权聚合, 得到最终的节点嵌入表示。本文所提模型在Cora、Citeseer和Pubmed三个数据集上进行节点分类实验, 准确率分别为84.4%、74.4%、81.5%, 与表现优异的GAT及后续变体模型相比, 分别提高1.4%、2.9%、3.2%; 在节点聚类实验中, 聚类精度分别提高3.3%、3.4%、1.2%。可见, 我们提出的算法可以得到更好的嵌入结果。

关键词: 网络表示学习; 图注意力网络; 自监督学习; 图结构学习; 节点分类

中图分类号: TP391 **文献标志码:** A **文章编号:** 0253-2395(2025)01-0029-14

Network Representation Learning Based on Structural Learning and Self-supervised Graph Attention

WANG Jinghong^{1,2,3}, ZHENG Ruice¹, MI Jusheng⁴, LI Haokang^{5*}

1. College of Computer and Cyber Security, Hebei Normal University, Shijiazhuang 050024, China;
2. Hebei Key Laboratory of Network and Information Security, Hebei Normal University, Shijiazhuang 050024, China;
3. Hebei Provincial Engineering Research Center for Supply Chain Big Data Analytics & Data Security, Shijiazhuang 050024, China;
4. School of Mathematical Science, Hebei Normal University, Shijiazhuang 050024, China;
5. College of Artificial Intelligence and Big Data, Hebei University of Engineering and Technology, Shijiazhuang 050091, China)

Abstract: Network representation learning is the foundation of network analysis tasks, that holds significant importance in mining and analyzing the real network data. Recently, Graph Attention Networks (GAT) and their variants have shown exceptional performance. This is particularly evident in the field of network representation learning. However, attention-based methods have the fol-

收稿日期: 2024-06-09; **接受日期:** 2024-10-24

基金项目: 河北省自然科学基金资助项目(F20242050280); 河北省高等学校科学技术研究项目(ZD2022139)

作者简介: 王静红(1967-), 女, 河北石家庄人, 博士, 教授, 研究方向为机器学习、数据挖掘、人工智能。E-mail: wangjinghong@126.com

* **通信作者:** 李昊康(LI Haokang), E-mail: lihaokang217@126.com

引文格式: 王静红, 郑瑞策, 米据生, 等. 基于结构学习和自监督图注意力的网络表示学习[J]. 山西大学学报(自然科学版), 2025, 48(1): 29-42. DOI:10.13451/j.sxu.ns.2024134.

lowing limitations: (1) Only first-order neighbor information of nodes is considered, ignoring higher-order neighbors. (2) The model lacks interpretability. (3) The issue of noisy edges in the graph is not considered. To tackle these issues, this paper proposes a Structural Learning and Self-supervised Graph Attention Network Embedding Model (SL-SGAT), which integrates node features and structural information, reduces noise edge interference, and enhances model interpretability. SL-SGAT mainly consists of three parts: graph structure learning, self-supervised attention mechanism, and feature aggregation. Graph structure learning constructs a global graph structure network. The self-supervised attention mechanism sets up a self-supervised relation prediction task, with noise edge loss added. Feature aggregation utilizes attention coefficients for weighted aggregation to obtain the final node embedding representation. The model proposed in this paper was tested on the Cora, Citeseer, and Pubmed datasets for node classification tasks, achieving accuracies of 84.4%, 74.4%, and 81.5%, respectively. Compared to the high-performing GAT and its subsequent variants, our model shows improvements of 1.4%, 2.9%, and 3.2%, respectively. In the node clustering experiments, the clustering accuracy improved by 3.3%, 3.4%, and 1.2%. These results demonstrate that our proposed algorithm can achieve a better representation of node embedding.

Key words: network representation learning; graph attention network; self-supervised learning; graph structure learning; node classification

0 引言

网络表示学习^[1] (Network Representation Learning), 也称为节点嵌入(Node Embedding)或图嵌入(Graph Embedding), 旨在将图中的节点从高维映射到低维向量空间, 以便进行各种下游任务^[2]。生成的节点向量可用于节点分类^[3]、节点聚类^[4]、节点推荐^[5]、链接预测^[6]等任务。网络表示学习能够将复杂网络结构转换为低维向量表示, 它在降低算法设计难度, 帮助人们更好理解和分析网络数据方面, 具有重要意义。

为了更好地进行网络表示学习, 研究人员开发了许多基于图神经网络方法, 主要分为两类: (1) 基于图卷积神经网络的方法。(2) 基于注意力神经网络的方法。基于图卷积神经网络的方法通过一阶频谱低通滤波器聚合邻域特征。主要方法有图采样与聚合(Graph Sample and Aggregation, GraphSAGE)^[7]、图卷积网络(Graph Convolutional Networks, GCN)^[8]、深度图信息最大化(Deep Graph Infomax, DGI)^[9]、相似度保留的图卷积网络(Similarity Preserving Graph Convolutional Networks, SimP-GCN)^[10], 这些方法在一些数据集上表现出色。但是这些方法是将邻居的信息平均聚合, 未区分不同邻居的重要性。基于注意力神经网络的方法, 考虑了不同节点邻居节点之间的重要性, 通过计算注意力权重, 聚合最相关邻居节点信息。主要有图注意力网络(Graph Attention Networks,

GAT)^[11]、图联合注意力网络(Graph Conjoint Attention Networks, CATs)^[12]、稀疏图注意力网络(Sparse Graph Attention Networks, SGATs)^[13]。GAT模型通过注意力系数聚合邻居节点信息。但是GAT在聚合邻居节点信息时只利用了一阶邻居信息, 并没有充分利用图的结构信息, 另一方面人们对GAT学习到的信息并不了解, 缺乏可解释性, 对图中存在的噪声问题也没有考虑。CATs模型提出联合注意力机制, 结合神经网络内部和外部的异构可学习因素来计算注意力系数, 学习到高阶邻居节点信息, 但易受外部噪声影响。SGATs方法通过注意力系数来剪枝噪声边, 提高图学习任务的准确性, 但其并未考虑高阶邻居节点信息。受文献[11-13]研究的启发, 同时结合节点属性特征和结构特征, 进行图结构学习, 得到全局图结构网络, 使得中心节点可聚合高阶邻居节点信息。通过定义一个自监督关系预测任务, 区分边是否存在, 降低噪声边干扰, 知道节点聚合哪些邻居节点信息, 增加模型可解释性。

因此, 本文提出了基于结构学习和自监督图注意力的网络嵌入模型(Structural Learning-based Self-supervised Graph Attention Network, SL-SGAT)。该模型主要由图结构学习、自监督注意力机制和特征聚合三部分构成。首先, SL-SGAT计算节点特征相似度和结构邻居度, 生成全局图结构矩阵, 用于节点连接和它们高阶相似的节点。通过阈值对生成的全局图结构

矩阵进行调优,基于修改后的矩阵,生成一个新的图。然后,设计自监督关系预测辅助任务,对图中的边进行编码,在区分正确和错误连接邻居时,降低噪声边干扰,学习到更好的节点特征表示。本文的主要贡献有以下几点:

1)提出了一个新的SL-SGAT模型,包括三部分:图结构学习、自监督注意力机制和特征聚合,考虑高阶邻居节点信息,降低噪声边对模型的干扰,同时增加了模型的可解释性。

2)设计图结构学习方法,为中心节点增加高阶邻居节点,从而聚合到更加丰富的信息。自注意力机制模块中定义关系预测任务,降低噪声边干扰,增加模型的可解释性。

3)在真实世界的三个数据集上进行实验,实验结果表明SL-SGAT模型的优越性与合理性。

第1节介绍了与节点嵌入相关的工作。第2节详细阐述了本文中涉及的关键变量、相关定义以及所使用的符号定义。第3节介绍了模型的整体架构及其实现的原理。第4节对实验结果进行全面分析,除了对SL-SGAT模型结果进行分析外,还分析了超参数对模型性能的影响,设置了消融实验,验证模型的有效性。

1 相关工作

1.1 非图神经网络方法

在早期,许多研究人员提出了基于拉普拉斯特征图和基于图相似度矩阵分解的方法。最早的研究多维尺度分析(Multidimensional Scaling, MDS)^[14]直接考虑两个节点之间的欧式距离,认为任何一对训练节点都是连接的。后续研究(例如等度量映射(Isometric Mapping, Iso-map)^[15]、图的拉普拉斯算子的特征值(Eigenvalues of the Laplacian of a Graph, LE)^[16]、局部保留投影(Locality Preserving Projections, LPP)^[17]等)提出的模型与之不同,根据数据特征构建 k 近邻图 K 邻近算法(K -Nearest Neighbor, KNN),每个节点只与前 k 个相似邻居相连,再根据不同的方法计算相似性矩阵,以尽可能保留所需要的图属性。这些早期的非神经网络嵌入方法,虽然简单高效,但是性能有限。为提高性能,把节点嵌入到潜在的欧几里得空间

中,然后根据节点嵌入来训练模型。比较有代表性的算法有DeepWalk^[18]、node2vec^[19]、大规模信息网络嵌入(Large-scale Information Network Embedding, LINE)^[20]等。这些算法仅考虑图的结构特征,忽视了节点属性信息,无法达到先进模型的性能^[21]。

1.2 图神经网络方法

图神经网络可以捕获高度非线性信息,通过消息传递机制聚集节点邻居特征信息,从而实现节点嵌入。目前图神经网络主要分为两大类:(1)基于图卷积神经网络方法。(2)基于注意力神经网络方法。

基于图卷积神经网络方法中具有代表性的是GCN^[8]模型,GCN将图中节点特征和邻居节点特征进行聚合,从而得到新的节点表示。之后,出现了许多关于GCN的变体,例如具有复杂有理谱滤波器的图卷积神经网络(Graph Convolutional Neural Networks with Complex Rational Spectral Filters, CayleyNet)^[22]、自适应图大小的图卷积网络(Adaptive Graph Convolutional Network, AGCN)^[23]、双图卷积网络(Dual Graph Convolutional Networks, DualGCN)^[24]等。但是所有基于频域的图卷积神经网络无法直接泛化到训练过程中未出现的节点,属于直推式学习。为解决这个问题,研究人员提出了大量基于空间的图卷积方法。GraphSAGE^[7]模型通过对固定邻域局部邻居进行随机采样,将采样得到的节点进行特征聚合。SimP-GCN^[10]通过聚集和转换邻域内的节点特征来提取有效的节点表示。DGI^[9]模型通过最大化局部互信息驱动局部网络嵌入来捕获全局结构信息。

以上图卷积方法在聚合信息时都是同等看待每一个邻居节点,无法对不同节点赋予不同的重要性。为了区分不同邻居节点对目标节点的贡献,基于注意力神经网络的方法GAT^[11]首先提出将注意力机制与图神经网络结合,在图神经网络中使用自注意力机制,根据注意力系数去聚合邻居节点信息,然后将信息传播到下一层。但是GAT只计算目标节点与其周围邻居节点之间的注意力,因此难以聚合到高阶邻居节点信息,并且人们对GAT学习到的信息不了解。针对这些问题,最短路径图注意力网

络 (Shortest Path Graph Attention Network, SPAGAN)^[25]根据最短路径计算中心节点与其高阶邻居之间的注意力分数。但是没有考虑到节点特征信息相似度。CATs^[12]算法提出联合注意力机制,结合神经网络内部和外部的异构可学习因素来计算注意力系数,但是模型易受外部噪声影响。结构学习图注意力网络 (Structure Learning Graph Attention Networks, SLGAT)^[26]计算结构邻近度,为节点增加高阶邻居节点,从而聚合到高阶邻居特征,但没有考虑节点属性特征,没有考虑图噪声问题。自监督图注意力网络 (Self-supervised Graph Attention Network, Super-GAT)^[27]通过降低噪声边对实验的干扰,从而提升网络表示性能。SGATs^[13]方法通过学习系数的注意力系数来剪枝噪声边,提高图学习任务的准确性。但以上两种算法都没有考虑高阶邻居节点信息。

为此,我们提出了SL-SGAT模型,考虑节点特征信息和结构信息,增加高阶邻居节点,通过设置关系预测任务,降低噪声边干扰,又增加了模型的可解释性。

2 相关概念

本节介绍论文中涉及的变量以及相关定义,该方法中涉及的主要符号表示如表1所示。

表1 符号定义

Table 1 Definition of symbols

符号	含义
V	节点集合
E	节点之间的边集合
e_{ij}	节点 v_i 与 v_j 之间的注意力权重
$N(v)$	节点 v 的邻居节点集合
A	邻接矩阵
W^l	第 l 层的权重矩阵
H_i^l	第 l 层,节点 i 的嵌入表示,当 l 为第0层时,表示节点特征矩阵
q	随机游走路径长度
γ	随机游走次数
S_n	负采样概率
S_e	边采样概率, $S_e \in (0,1]$

定义1^[28] 图 $G=(V,E)$,其中 $V=\{v_1, v_2, \dots, v_n\}$ 是图的节点集合,节点数目为 n ; $E=\{e_{ij}\}$ 是图中节点之间的邻接边集合; $A \in \{0,1\}^{n \times n}$ 表示图的邻接矩阵;节点特征的矩

阵表现形式 $H^0=R^{n \times d}$ 是节点属性集合,其中 n 为节点数量, d 为节点特征向量的维度。

定义2 网络表示学习^[11]。给定图 $G=(V,E)$,网络表示学习的目的是学习一个映射函数 $f: v_i \rightarrow h_i \in R^d$ 其中 $v_i \in V$, d 是节点最终的向量表示维度。

定义3 节点分类^[29]。 $Y \in \{0,1\}^{n \times C}$ 表示节点的标签, C 是节点类别数量。对于一个半监督节点分类任务,有 $m(0 < m \ll n)$ 个拥有标签的节点,有 $n-m$ 个没有标签的节点。基于拥有节点特征矩阵 H^0 和节点标签 Y 的图 $G=(V,E)$ 来说,半监督节点分类问题通过一个分类器 $f: (G,X,Y) \rightarrow Y'$ 去预测没有标签节点的标签。

定义4 正点对互信息 (Positive Pointwise Mutual Information, PPMI) 矩阵^[24]。PPMI矩阵是一种用来表示词之间相关性的矩阵,它是基于点互信息 (Pointwise Mutual Information, PMI) 的改进。PMI是一个衡量两个事物相关性的指标,它的计算公式是:

$$I_{\text{PMI}}(x,y) = \log \frac{P(x,y)}{P(x)P(y)}, \quad (1)$$

其中 $P(x)$ 和 $P(y)$ 分别代表词 x 与词 y 出现的概率, $P(x,y)$ 表示词 x 和词 y 同时出现的概率。PMI值越大,说明两个词相关性越强。

PMI矩阵可能存在负值,PPMI将PMI的负值转换为0,避免PMI负值的出现,也保留了正数的PMI值。

$$I_{\text{PPMI}}(x,y) = \max(I_{\text{PMI}}(x,y), 0). \quad (2)$$

3 模型

基于结构学习和自监督图注意力的网络嵌入模型 (SL-SGAT),挖掘节点属性特征和全局结构特征中的内在联系,得到更具有表现力的节点特征表示。SL-SGAT的整体架构如图1所示。首先,根据节点特征相似度和全局结构近似度,构建一个包含节点特征和全局结构信息的图。通过自监督注意力机制,设计一个关系预测辅助任务,区分边的连接是否正确,同时计算注意力系数。最后将注意力系数与节点特征进行加权聚合,使用多头注意力机制,对每一个注意力头进行特征拼接或均值操作,得到最终的节点嵌入表示。

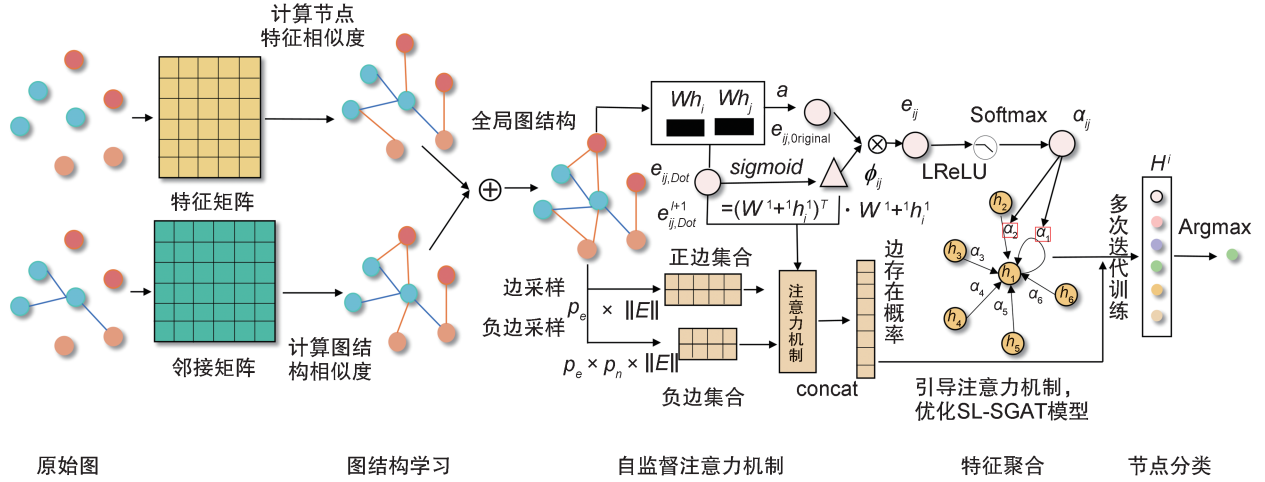


图1 SL-SGAT模型架构图

Fig. 1 Architecture diagram of SL-SGAT model

3.1 图结构学习

目前大多数网络表示学习方法通常只依赖于节点特征,忽略了全局结构相似性。为解决这一问题,我们提出了图结构学习模块。首先,我们需要计算节点之间的特征相似度。现在节点特征相似度度量方法有很多,例如余弦相似度、欧氏距离、注意力机制。根据节点之间的相似度,可学习到丰富的图拓扑结构。节点 i 和节点 j 之间的相似度可以表示为:

$$\kappa_{ij} = \text{Similar}(H_i, H_j), \quad (3)$$

其中 κ_{ij} 是节点 i 和节点 j 之间的相似度, H_i 和 H_j 分别代表节点 i 和节点 j 的向量表示。Similar 是一个相似度度量学习方法,比如余弦相似度和注意力机制相似度:

余弦相似度可以表示为:

$$\kappa_{ij} = \cos(H_i, H_j), \quad (4)$$

注意力机制相似度可以表示为:

$$\kappa_{ij} = \frac{\exp(\text{ReLu}(a^T|H_i - H_j|))}{\sum_{j=1}^n \exp(\text{ReLu}(a^T|H_i - H_j|))}. \quad (5)$$

根据相似度计算方法来计算得到相似度矩阵,矩阵中每一个位置都对应着两两节点的相似度,并根据阈值参数 k 学习稀疏图结构,具体表示为:

$$\kappa_{ij} = \begin{cases} \kappa_{ij} & \text{if } \kappa_{ij} > k, \\ 0 & \text{otherwise,} \end{cases} \quad (6)$$

相似度矩阵 κ 可以转化为邻接矩阵,它在每一层都可能是动态的,并对下游任务有自适应能力。但是它仅包含了节点特征相似信息,并没

有包含全局的拓扑信息。接下来采用 PPMI^[24] 矩阵表示全局结构邻近性,通过全局信息来修改相似度矩阵。

PPMI 矩阵是用来表示词之间相关性的矩阵,在自然语言处理中得到了广泛的应用^[30]。本文将使用 PPMI 矩阵来表示图的全局结构邻近性。首先,以每个节点为根节点进行 γ 次随机游走,每次随机游走的长度为 q 。统计出两个节点在相同随机游走路径中出现的次数,记为频率矩阵 M 。根据频率矩阵 M 我们可以计算出 PPMI 矩阵 I ,具体计算公式如下:

$$I_{ij} = \max \left\{ \log \frac{I_{ij}}{I_{i,*} I_{*,j}}, 0 \right\} = \max \left\{ \log \frac{\frac{M_{ij}}{\sum_{i,j} M_{ij}}}{\sum_j M_{ij} \sum_i M_{ij}}, 0 \right\}, \quad (7)$$

其中 I_{ij} 是节点 i 和节点 j 出现在同一随机游走路径上的估计概率,表现出全局结构中两个节点的邻近性, * 表示 i 或 j 在同一条随机游走路径中的其它节点,PPMI 矩阵 I 反映出全局结构中两两节点之间的邻近性。对矩阵 I 进行归一化操作,并通过阈值参数 k 学习稀疏图结构,具体表现为:

$$I_{ij} = \begin{cases} I_{ij} & \text{if } I_{ij} > k, \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

整合学习相似度矩阵 κ 和 PPMI 矩阵 I , 得到全局图结构 β' 。 β' 反映出节点特征的相似性和全局图结构邻近性。全局图结构 β' 表示为:

$$\beta' = \kappa + I. \quad (9)$$

当处理数据具有图结构时, 还需要考虑邻接矩阵。首先对原始邻接矩阵 A 进行归一化处理, 得到归一化邻接矩阵 A' 。将归一化邻接矩阵 A' 与学习到的全局图结构 β' 结合得到最终的图结构 ρ :

$$\rho = \beta' + A'. \quad (10)$$

对图结构学习内容进行梳理, κ_{ij} 代表两个节点之间的特征相似度, I_{ij} 代表两个节点之间的全局结构邻近度。我们通过阈值 k 来学习稀疏图结构。在原始图 G 上, 节点 i 和节点 j 之间不存在边, 且 $\kappa_{ij} > k$ 或 $I_{ij} > k$ 时, 在节点 i 和节点 j 之间产生一条新边。将权重矩阵转化为邻接矩阵, 在新图中使用图神经网络方法。基于学习到的最终图结构 ρ , 图神经网络可以在每一层聚合到高阶邻居特征, 扩大感受野。

3.2 自监督注意力机制模块

通常图中存在大量的噪声边, 经过图结构学习之后, 这些噪声边的数量可能会增加。现有方法主要是聚合高阶邻居节点信息, 从而提高网络嵌入效果。但是无法明确每个节点聚合了哪些邻居节点, 缺乏可解释性, 且未考虑到噪声的存在。为解决这些问题, 我们提出了自监督注意力机制模块, 设置一个关系预测任务, 计算节点间是否真实存在边, 旨在降低噪声边干扰并提高模型的可解释性。接下来, 我们先介绍三种常见的注意力机制, 然后介绍我们的自监督注意力机制。

3.2.1 注意力机制

原始注意力机制 图 $G=(V, E)$, N 是节点数量, F^l 是第 l 层的节点特征数量, 注意力层的输入为: H^l , 注意力层的输出为: H^{l+1} 。 h_i^{l+1} 的计算方法为: 使用 W^{l+1} 与 h_j^l 相乘, 其中 $W^{l+1} \in R^{F^{l+1} \times F^l}$ 。通过注意力系数 α_{ij}^{l+1} , 线性地组合一阶邻域 ($j \in N_i$) 特征。最后应用一个非线性激活函数 σ 。

首先, 计算节点间的注意力系数, 学习邻居节点的重要性。

$$e_{ij}^{l+1} = \text{attention}(y_i^l, y_j^l) = \bar{\alpha}^T [W^{l+1} h_i^l \oplus W^{l+1} h_j^l], \quad (11)$$

其中 attention 为注意力层, \oplus 表示向量拼接, 注意力机制 α 是一个前馈神经网络, $\bar{\alpha} \in R^{2F^{l+1}}$ 。 W^{l+1} 和 $\bar{\alpha}^{l+1}$ 都是要学习的参数。 e_{ij}^{l+1} 表示节点 v_j 对节点 v_i 的重要性。为了使注意力系数在不同节点间更容易比较, 使用 Softmax 函数对其进行归一化操作。

$$\alpha_{ij} = \text{Softmax}(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{t \in N_i} \exp(e_{it})} \frac{\exp(\text{LeakyReLU}(\bar{\alpha} [Wh_i \| Wh_j]))}{\sum_{t \in N_i} \exp(\text{LeakyReLU}(\bar{\alpha} [Wh_i \| Wh_t]))}. \quad (12)$$

计算归一化注意力系数后, 节点根据注意力系数进行特征聚合, 并应用非线性激活函数 σ , 将其结果作为最终节点特征。

为稳定注意力的学习过程, 我们采用多头注意力机制。具体来说, 通过 k 个独立的注意力机制执行以上变换, 并将其串联, 得到最终的输出。

$$h_i^l = \parallel_{k=1}^K \sigma(F^k) = \parallel_{k=1}^K \sigma \left(\sum_{j \in N_i} \alpha_{ij}^k W^k h_j \right) = \sigma \left(\frac{1}{K} \sum_{k=1}^K \sum_{j \in N_i} \alpha_{ij}^k W^k h_j \right), \quad (13)$$

其中 \parallel 表示连接法或平均法, 在最后预测层执行多头注意力, 采用平均法效果更好。应用的非线性函数, 通常是 Softmax 函数或者 Logistic sigmoid 函数。

点积注意力机制 与原始注意力机制使用一个单层前馈神经网络 $\bar{\alpha} \in R^{2F^{l+1}}$ 相比, 点积注意力机制, 采用点积计算方式代替单层前馈神经网络。以下是原始注意力系数和点积注意力系数的表示:

$$e_{ij, \text{Original}}^{l+1} = \text{attention}(y_i^l, y_j^l) = \bar{\alpha}^T [W^{l+1} h_i^l \oplus W^{l+1} h_j^l], \quad (14)$$

$$e_{ij, \text{Dot}}^{l+1} = \text{attention}(y_i^l, y_j^l) = (W^{l+1} h_i^l)^T \cdot W^{l+1} h_j^l. \quad (15)$$

混合注意力机制 将原始注意力系数与经过 sigmoid 函数的点积注意力系数相乘。使用 sigmoid 函数的点积注意力系数表示边存在的概率, 所以混合注意力系数可以软性的放弃不

可能有联系的邻居,同时隐式地赋予其余节点更加重要的系数。其计算表达式如下:

$$e_{ij, \text{Mixed}} = e_{ij, \text{Original}} \cdot \sigma(e_{ij, \text{Dot}}). \quad (16)$$

3.2.2 自监督注意力机制

SL-SGAT模型设置一个自监督关系预测任务,通过预测节点对之间是否存在边,来引导注意力机制,从而降低噪声边对模型性能的干扰,提高模型的可解释性。使用注意力机制预测节点之间的关系,对于一个节点对 (i, j) ,如果存在边,值为1,如果不存在边,值为0。引入一个变量 ϕ_{ij} ,记为节点 i 和节点 j 之间存在边的概率, ϕ_{ij} 的表示为:

$$\phi_{ij} = \sigma(e_{ij}), \quad (17)$$

其中 e_{ij} 为以上三种注意力系数中的一种。本模型采用混合注意力机制,通过混合注意力系数 $e_{ij, \text{Mixed}}$ 以及节点 i 和节点 j 之间存在边的概率 $\phi_{ij, \text{Mixed}}$ 来描述SL-SGAT模型。

$$\phi_{ij, \text{Mixed}} = \sigma(e_{ij, \text{Dot}}), \quad (18)$$

$$e_{ij, \text{Mixed}} = e_{ij, \text{Original}} \cdot \phi_{ij, \text{Mixed}} = e_{ij, \text{Original}} \cdot \sigma(e_{ij, \text{Dot}}). \quad (19)$$

训练样本是一组边 E^+ 和负采样边 E^- ,记为 $E^+ \cup E^-$ 。从边集中任意选择 $S_e \cdot |E|$ 个边作为正样本 E^+ 。本模型使用负采样,从正边集 E^+ 中任意选择 $S_n \cdot |E^+|$ 个边作为负样本 E^- 。其中 $S_e \in (0, 1]$, $S_n \in R^+$ 均为超参数。

计算注意力系数后,进行特征加权聚合更新节点特征,进行消息传递,传递到下一层的表征为:

$$h_i^{l+1} = \sigma\left(\sum_{j \in N_i} \alpha_{ij} W h_j^l\right), \quad (20)$$

其中 $\sigma(\cdot)$ 为非线性激活函数, α_{ij} 为节点 i 和 j 之间的归一化注意力系数, h_j^l 为节点 j 在第 l 层的向量表示。

使用与GAT相同的多头注意力机制,采用 K 个独立的注意力机制对节点特征进行更新,取每个头的均值作为节点嵌入表示,最终节点 i 的嵌入表示为:

$$h_i^{l+1} = \sigma\left(\frac{1}{K} \sum_{k=1}^K \sum_{j \in N_i} \alpha_{ij}^k W^k h_j^l\right). \quad (21)$$

3.2.3 自监督联合损失函数

SL-SGAT有足够多的边采样和负边采样,进行自监督关系预测任务。定义优化目标函数为 L_E^l ,为第 l 层的二元交叉熵损失函数。其中

“1”为指示函数。每次训练使用 $E^+ \cup E^-$ 的一个子集,以概率 $S_e \in (0, 1]$ 进行采样,以保证每次迭代训练,获得来自随机性的正则化效应。最终,我们将来自节点标签交叉熵损失 L_V^l 、所有层的自监督注意力噪声边损失以及L2正则化损失进行整合,得到模型完整的损失函数:

$$\begin{aligned} L_E^l &= -\frac{1}{|E^+ \cup E^-|} \sum_{j, i \in E^+ \cup E^-} 1_{(j, i)=1} \cdot \log \phi_{ij}^l + \\ &\quad 1_{(j, i)=0} \cdot \log(1 - \phi_{ij}^l), \quad (22) \\ L &= L_V + \lambda_E \sum_{l=1}^L L_E^l + \lambda_2 \cdot \|\mathbf{W}\|_2 = \\ &\quad -\frac{1}{N} \left(\sum_{i \in N} \log(h[i, \mathbf{y}[i]]) \right) + \\ &\quad \lambda_E \sum_{l=1}^L L_E^l + \lambda_2 \cdot \|\mathbf{W}\|_2, \quad (23) \end{aligned}$$

其中 λ_E 用来平衡来自节点标签损失和所有层自监督注意力损失, λ_2 是L2正则化损失函数系数,用来控制正则化大小。 $h[i, \mathbf{y}[i]]$ 表示第 i 个节点中目标类别对应的概率值。

3.3 算法流程

具体算法如算法1所示。

算法1 基于图结构学习的自监督网络表示学习模型(SL-SGAT)

输入:图 $G=(V, E)$,特征矩阵 X ,邻接矩阵 A

输出:节点表示矩阵 H

- ① 根据公式(3),(4),(6),采用余弦相似度,计算得到图的节点特征相似度矩阵 κ ;
- ② 根据公式(7)–(8),计算全局图结构中两两节点邻近度,得到PPMI矩阵 I ;
- ③ 通过公式(9)–(10),计算得到最终图结构矩阵 ρ ;
- ④ 将得到的权重矩阵转化成邻接矩阵 ρ' ;
- ⑤ 通过公式(14)–(15),计算邻居节点的原始注意力系数 $e_{ij, \text{Original}}$ 和点积注意力系数 $e_{ij, \text{Dot}}$;
- ⑥ 根据公式(17)–(18),计算邻居节点之间边存在概率 ϕ_{ij} ;
- ⑦ 根据公式(19),计算得到节点的混合注意力系数 $e_{ij, \text{Mixed}}$;
- ⑧ 根据公式(12),计算得到归一化注意力系数 $\alpha_{ij, \text{Dot}}$;
- ⑨ 根据公式(20)–(21),进行特征加权聚合,得到节点 i 嵌入表示 h_i

⑩ for epoch=1, 2, ..., epochs do

⑪ 利用随机梯度下降更新权重

⑫ 优化联合损失函数 $L = L_V + \lambda_E \sum_{l=1}^L$

$L'_E + \lambda_2 \cdot \|W\|_2$

⑬ end for

3.4 复杂度分析

本文所提出的模型时间效率较高,其主要时间开销主要包括三部分:计算节点对之间的相似性得分、计算 PPMI 矩阵、计算注意力系数。采用公式(4)计算节点对之间的相似度得分,这需要 $O(n^2 F^0)$ 的时间复杂度,其中 n 为节点数量, F^0 为节点输入特征维度。计算 PPMI 矩阵,所需的时间复杂度为 $O(n\gamma q^2)$ 。计算注意力系数的时间复杂度主要涉及两个乘法运算环节:

(1) 节点特征映射,使用 Wh_i 将节点特征维度从 F^0 映射到 F' ,其时间复杂度为 $O(F^0 F')$,所有节点都需要映射,则时间复杂度为 $O(nF^0 F')$ 。

(2) $\alpha(\cdot)$ 以及点乘操作, $\alpha(\cdot)$ 是将 $2 \times F'$ 维度的向量映射到一个实数上,其时间复杂度为 $O(F')$,一共有 m 条边,则时间复杂度为 $O(m \times F')$ 。点乘操作,两个向量维度都为 F' ,其时间复杂度为 $O(F')$,每条边都需要计算,则时间复杂度为 $O(m \times F')$,以上操作可并行计算,其总时间复杂度为 $O(m \times F')$ 。

结合以上两个环节,计算注意力系数所需的时间复杂度为 $O(nF^0 F' + m \times F')$,后续的其他操作不再涉及复杂度高的乘法运算,所以 SL-SGAT 模型的时间复杂度为 $O(n^2 F^0 + n\gamma q^2 + nF^0 F' + m \times F')$ 。

4 实验

本节将介绍在三个真实网络数据集上进行节点分类实验结果,验证 SL-SGAT 的有效性。本文将 SL-SGAT 与流行的网络嵌入模型进行比较。基于随机游走的嵌入算法, DeepWalk^[18] 和 node2vec^[19]。基于卷积神经网络的嵌入方法 GraphSAGE^[7], GCN^[8]、DGI^[9] 和 SimP-GCN^[10]。基于图注意力神经网络的方法 GAT^[11], CATs^[12], SGATs^[13]。为了进行公平比较,保持实验算法跟算法设置一致,且所有

实验均用 Python 语言编写,通过 Pytorch 框架进行构建,并在 GEFORCE RTX 3080 GPU 上运行。

首先介绍数据集以及实验训练参数配置,展现了本模型与其他先进基线模型的在不同实验设置下的实验结果,最后给出超参数灵敏度分析实验结果和消融实验结果。

4.1 数据集

本文使用 Cora^[31]、Citeseer^[31] 和 Pubmed^[31] 三个引文网络数据集进行模型学习和模型测试评估。在这三个数据集中,节点表示论文,节点之间的边表示论文之间的引用关系,节点特征表示每篇论文的属性特征,比如关键词、发表日期、作者、研究主题等等。表 2 列出了所有数据集的详细统计数据。

表 2 数据集统计信息

Table 2 Statistical information of the datasets

数据集	节点数	边数	特征数	标签	训练集	验证集	测试集
Cora	2 708	5 429	1 433	7	140	500	1 000
Citeseer	3 312	4 714	3 073	6	120	500	1 000
Pubmed	19 717	44 338	500	3	60	500	1 000

4.2 训练参数设置

对于随机游走的方法,设置窗口大小为 5,随机游走次数为 10,随机游走长度为 40。所有模型都是用 Adam 优化器学习模型参数,学习率 $l_r \in \{0.005, 0.02, 0.05\}$,在其中选择最佳学习率。对于 GCN、GAT 及其变体模型,采用两层消息传递层和多头注意力机制。第一层注意力层包含 8 个注意力头,每个注意力头有一个权重矩阵 $W \in R^{d \times 8}$;第二层注意力层包含 8 个注意力头,每个注意力头有一个权重矩阵 $W \in R^{64 \times c}$;对于我们的模型,网络架构与 GAT 相同,epoch 为 500。在构建 PPMI 矩阵时,设定随机游走长度 $q = 5$,以每个节点为根进行 $\gamma = 40$ 次随机游走,其他超参数因数据集不同,超参数不同,不同数据集的最优超参数如表 3 所示。

4.3 实验结果

4.3.1 节点分类

使用节点分类任务进行模型性能评估,计算目标节点嵌入向量中最大值,作为分类结果。表 4 显示了 10 次运行的平均结果。在所有

表3 不同数据集在SL-SGAT模型上取得最优结果的参数设置

Table 3 The parameter settings that achieve optimal results on different datasets using the SL-SGAT model

数据集	S_e	S_n	λ_E	k
Cora	0.8	0.5	4	0.2
Citeseer	0.2	0.1	4	0.2
Pubmed	0.6	0.5	10	0.1

模型中,最佳性能用粗体数字表示。与基于随机游走的方法(DeepWalk、node2vec)相比,SL-SGAT的性能提高了10%~20%。这表明基于随机游走的方法仅考虑图的结构特征,忽视了节点属性信息,无法达到先进模型的性能。与基于注意力机制的方法(GAT、CATs、SGATs)相比,SL-SGAT性能提升了1.4%~6.2%,这表明仅考虑噪声边或全局结构信息不能很大程度提高模型性能,而SL-SGAT模型通过使用图结构学习和自监督注意力机制,充分考虑节点属性特征和结构信息,能够聚合到高阶邻居节点信息,降低图中噪声干扰,提高模型的表现。值得注意的是,在Cora数据集上SL-SGAT模型和SGATs模型相差不大,仅提高0.5%,可能原因是Cora数据集中存在的邻居节点较少,对模型最终性能影响不大,两个模型对图中噪声处理的效果相近。此外CATs模型与GAT相比,GAT在Cora和Citeseer数据集上的性能要优于CATs模型,我们猜测可能CATs模型在添加高阶邻居节点时,添加了噪声边,导致模型性能下降。我们还对比了三种注意力机制对模型性能的影响,结果显示混合注意力机制效果最佳。这是因为混合注意力系数通过将原始注

意力系数与节点间边存在的概率相乘,隐式地为更重要的节点分配权重,在计算注意力系数时能够舍弃不重要的节点,降低噪声边对模型性能的干扰。

表4 节点分类对比实验结果,使用准确率Accuracy和F1指标对分类结果进行评估,从而评价模型性能

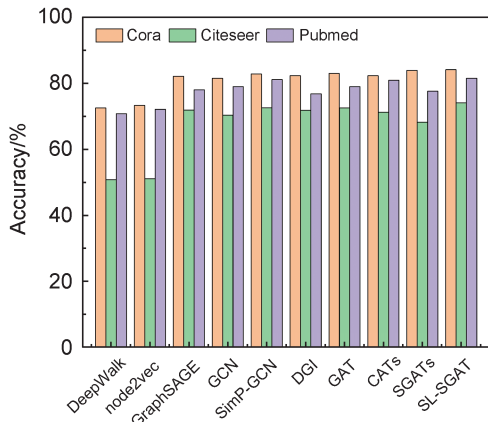
Table 4 The results of the node classification comparison experiment, evaluated using Accuracy (ACC) and the F1 score to assess the classification results, thereby evaluating the model's performance %

算法/数据集	Cora		Citeseer		Pubmed	
	ACC	F1	ACC	F1	ACC	F1
DeepWalk	72.5	71.5	50.8	49.4	70.8	71.1
node2vec	73.3	74.1	51.1	50.1	72.1	72.0
GraphSAGE	76.5	75.9	66.9	62.2	74.5	74.4
GCN	81.5	76.7	70.3	61.5	79.0	80.0
DGI	82.3	82.3	71.8	71.5	76.8	76.5
SimP-GCN	82.8	81.1	72.6	70.9	81.1	79.9
GAT	83.0	83.1	71.5	63.0	77.8	77.1
CATs	82.3	82.1	73.2	72.3	80.9	80.5
SGATs	83.0	83.3	71.5	71.9	78.3	79.6
SL-SGAT _{Original}	83.1	82.3	71.9	70.4	79.9	79.5
SL-SGAT _{DP}	83.3	82.8	72.9	72.3	80.5	80.1
SL-SGAT _{MX}	84.4	84.3	74.4	74.1	81.5	81.1

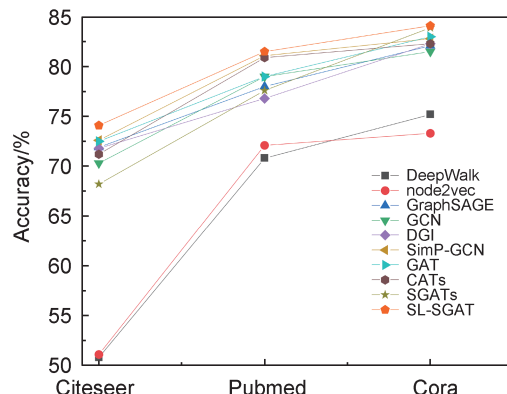
图2显示了本文提出的模型与其他9个模型进行对比的结果。从图中可以看出,本文所提出的模型具有显著的效果。

4.3.2 节点聚类

我们将所有模型输出的节点表示输入K-Means算法中进行聚类。K-Means算法中聚类数目设置为相应节点类别数目,具体为:Cora



(a) SL-SGAT与其他模型实验结果统计



(b) SL-SGAT与其他模型实验结果纵向对比

图2 SL-SGAT与其他模型进行比较的统计结果

Fig. 2 Statistical results of SL-SGAT in comparison with other models

为7类,Citeseer为6类,Pubmed为3类。我们对每个数据集进行10次聚类,取其均值作为结果。实验结果如表5所示。尽管节点聚类存在挑战,但是SL-SGAT仍超越了所有基线模型。与其他最佳基线模型相比,SL-SGAT的节点聚类能力提高了1.2%~3.4%。SL-SGAT在聚合高阶邻居节点信息,降低噪声干扰方面的能力,是其性能提升的重要原因。这些发现表明,SL-SGAT通过图结构学习和自监督注意力机制能够有效学习到节点特征表示。

4.4 可视化分析

为了更直观地了解我们提出的模型性能,将对SL-SGAT学习到的节点嵌入进行定性研究。利用t分布随机邻居嵌入(t-Distributed Stochastic Neighbor Embedding, t-SNE)^[32]将学习到的节点嵌入投影到二维空间中。在图3中,每一个节点代表一篇已发表的论文,节点颜色相同表示相同研究领域的论文。根据图3可知,节点分布展示出良好的聚类效果,节点根据类别形成了明显的簇。各类颜色节点聚集在一起,显示出类别的强一致性。类别之间边界较为清晰,表明节点嵌入效果优异。

图4通过可视化结果,展示了不同模型的效果存在显著差异。由于篇幅有限,只展示了

表5 节点聚类对比实验结果,使用归一化互信息(Normalized Mutual Information, NMI)来评价聚类效果

Table 5 The results of the node clustering comparison experiment, evaluated using NMI (Normalized Mutual Information)

算法/数据集	NMI/%		
	Cora	Citeseer	Pubmed
DeepWalk	32.8	8.8	10.5
node2vec	35.6	10.1	37.9
GCN	54.4	35.1	41.2
DGI	53.5	45.7	30.2
GAT	44.1	37.5	37.5
CATs	60.0	43.7	34.8
SGATs	61.1	44.2	38.0
SL-SGAT	64.4	47.6	39.2

注:NMI衡量了聚类结果与真实标签之间信息共享程度。

Note: NMI measures the degree of information shared between the clustering results and the true labels.

在Cora数据集上的可视化结果。GCN的效率最低,不同颜色的节点混在一起。GAT与SGATs的可视化效果要优于GCN,说明为不同节点分配不同的权重,能够提高模型性能。SGAT虽然可以更好地进行节点分类,但是颜色相同的节点分布稀疏。而我们提出的SL-SGAT模型与其他模型相比,具有更好的节点嵌入效果。说明考虑图节点特征和结构信息,

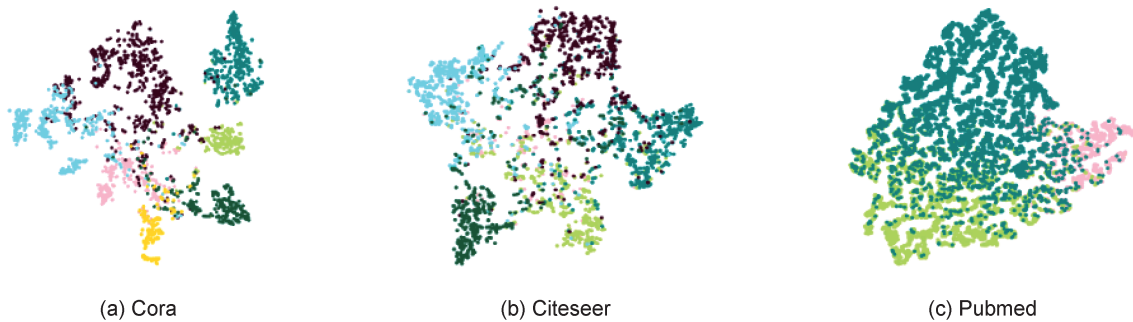


图3 SL-SGAT模型在Cora、Citeseer和Pubmed三个数据集上的可视化效果

Fig. 3 The visualization results of the SL-SGAT model on the Cora, Citeseer, and Pubmed datasets



图4 GCN,GAT,SGAT和SL-SGAT模型对Cora节点嵌入的t-SNE可视化结果

Fig. 4 Results of t-SNE visualization of Cora node embedding by GCN, GAT, SGATs and SL-SGAT models

降低噪声边干扰,可提高节点嵌入效果。

4.5 超参数分析

为了验证所有参数对模型的影响,本节分析了学习率 l_r 、损失混合系数 λ_E 、阈值 k 、负采样率 S_n 、边采样率 S_e 的参数灵敏度。在图5—图9中分别描述了各种超参数对节点分类性能的影响。

图5是不同损失混合系数 λ_E 的实验结果。对于 λ_E 来说,改变 λ_E 值,能够影响模型性能。 λ_E 值越大,关系预测损失对模型的影响也就越大。实验结果表明,当 λ_E 为4时 Citeseer 和 Cora 数据集性能达到最大值,而 Pubmed 数据集则是在 λ_E 为10时达到性能最大值。根据图5可以发现每个数据集都有一个最佳的边缘监督水平,如果 λ_E 过大,则会降低节点分类性能。由此我们可以推断,不同数据集中,噪声对模型的影响不同。通过降低噪声边干扰,可以提高模型性能。

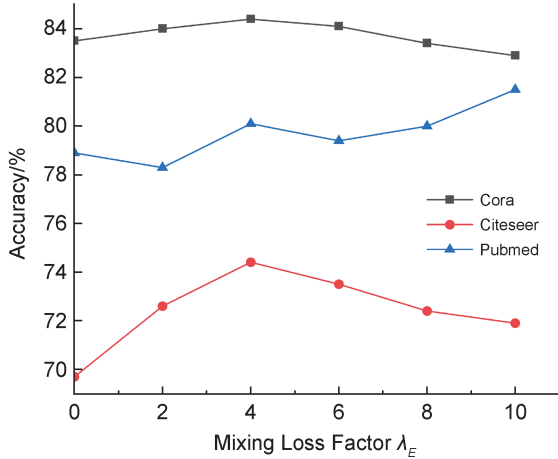


图5 使用不同损失混合系数的节点分类结果

Fig. 5 Node classification results using different loss mixing coefficients

图6是负边采样率 S_n 的实验结果。采样负边作为关系预测的负样本。最佳的负样本数量因数据集而异,但是都应小于正样本数量。根据图6可以发现,随着 S_n 增加,在Cora和Pubmed数据集上的性能呈现凸函数曲线关系。而Citeseer随着 S_n 的增加而降低。这是因为适度地进行负边采样,进行关系预测任务,降低噪声边干扰,但是过多地进行负采样会影响关系预测效果。

通过边缘采样率 S_e ,进行正边采样,作为关系预测的正样本。随 S_e 变化,节点分类性能也会随之发生变化。但是不同数据集变化形式不

同。从图7中可以看出,Cora和Pubmed数据集的性能呈现凸曲线关系。Citeseer数据集性能在 S_e 刚开始增加时,性能提高,到达性能最优时,开始逐步下降。我们推测,过多地进行正边采样可能会影响关系预测效果,从而无法有效降低噪声边干扰。

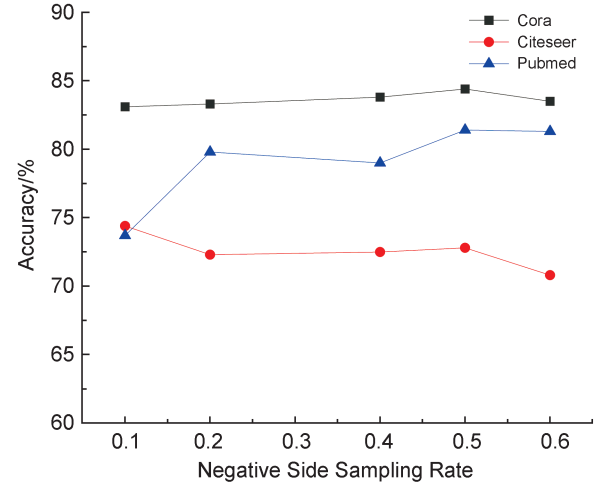


图6 使用不同负边采样率的节点分类结果

Fig. 6 Node classification results using different negative edge sampling rates

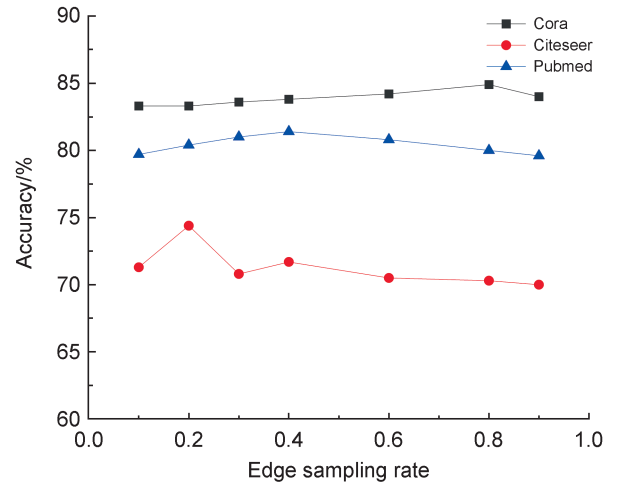


图7 不同边采样率的节点分类结果

Fig. 7 Node classification results for different edge sampling rates

图6和图7是关系预测中正边与负边的采样率,通过选择足够数量的正样本(真实存在的边)和负样本(假设不存在的边),构建一个关系预测模型。模型核心任务区分哪些节点之间应当存在边,哪些节点之间不存在边,从而减少噪声边对模型影响。通过关系预测,我们不仅能够提高图数据的质量,减少数据中的

不确定性,还能够进一步理解节点间的潜在关系。特别是在确定节点之间是否存在连接关系的过程中,我们可以明确每个节点在聚合过程中的具体邻居节点,增加模型的可解释性。

对于阈值 k 来说,不同的阈值对不同数据集影响不同。如图 8 所示,经过多次测试,当 $k=0.1$ 时,SL-SGAT 模型在 Pubmed 中取得了最佳性能。当 $k=0.2$ 时,在 Cora 和 Citeseer 数据集上取得了最佳性能,此时图 Cora 中的边数为 13 296,比原始边数 10 556 增加了 2 740 条边。值得注意的是,全局图结构矩阵的每个元素在归一化处理都小于或等于 1。当 $k=1$ 时,表示我们不使用图结构信息。 k 值越大,可以利用的图结构信息越少。当 $k>0.2$ 时,模型的性能要低于 $k\leq 0.2$,这表明图结构学习能够提高模型的性能,但是过多地考虑图结构信息会影响模型性能。从图中还可以发现,当 $k=0$ 时,学习到的矩阵是一个稠密矩阵,但是真实世界的图是一个稀疏矩阵,所以在这种情况下,模型的性能比较低。

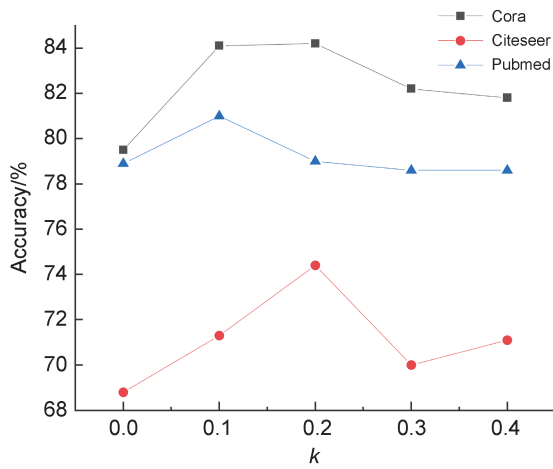


图8 使用不同阈值的节点分类结果

Fig. 8 Node classification results using different thresholds

在三个数据集上进行节点分类实验,观察学习率 $l_r \in \{0.005, 0.05\}$ 时对实验的影响,其他参数均为最优值。图 9 显示学习率对三个实验数据集实验结果的影响。结果表明,最佳学习率因数据集而异,当 $l_r = 0.005$,在 Citeseer 数据集上表现最好,而在 Cora 数据集上, $l_r = 0.02$ 时性能最好,在 Pubmed 数据集上,当 $l_r = 0.05$ 时,效果最好。

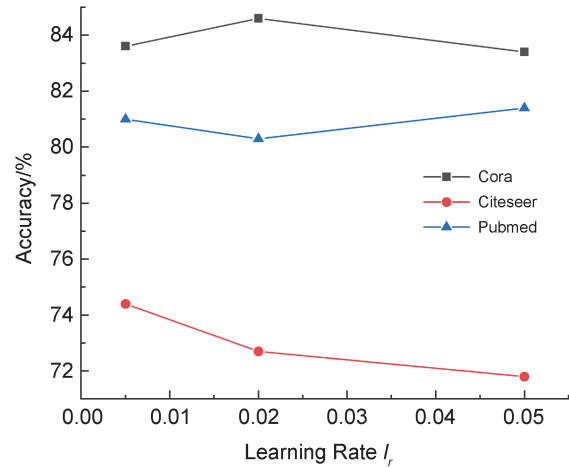


图9 使用不同学习率的节点分类结果

Fig. 9 Node classification results using different learning rates

4.6 消融实验

为验证本文所提算法的有效性,进行消融实验,设置模型变体,进行节点分类任务,实验参数设置与本模型在各个数据集最优表现相同,在 Cora、Citeseer、Pubmed 数据集上的实验结果如表 6 所示。

表6 SL-SGAT 模型的不同变体之间进行节点分类对比实验

Table 6 Comparative experiments of node classification among different variants of the SL-SGAT model

模型	ACC/%		
	Cora	Citeseer	Pubmed
Structure-only	83.1	72.3	78.5
Self-supervised-only	84.0	71.1	80.5
Str+Self	84.4	74.4	81.5

Structure-only 是 SL-SGAT 的变体模型,仅使用图结构学习模型进行网络表示学习。

Self-supervised-only 是 SL-SGAT 的变体模型,仅使用自监督注意力机制模块进行网络表示学习。

Str+Self 表示基于图结构学习和自监督注意力机制的模型。

对于 Citeseer 数据集,结构学习与自监督注意力网络模型(Str+Self)的节点分类精度为 74.4%,变体下的分类精度分别为 72.3%, 71.1%,在本文所提模型下,节点分类精度提高了 2 个百分点左右。同时,只考虑全局结构信息的模型变体要优于只进行自监督注意力网络的模型变体。我们猜测,由于 Citeseer 数据集中边

数较少,存在噪声边也偏少,所以为其增加高阶邻居节点,有利于提高模型性能。对于数据集 Cora 和 Pubmed 来说,只使用自监督注意力的模型 (Self-supervised-only) 分类精确度为 84.0%, 80.5%, 相比于结合结构学习和自监督注意力机制下的分类精确度 84.4%, 81.5% 性能要低一些,但是要优于只考虑图结构学习的模型变体,这是因为在 Cora 和 Pubmed 数据集中,存在着大量的边,噪声边也偏多,自监督注意力机制模块能有效降低噪声边影响,提升模型性能。我们能够很容易发现,图结构学习和自监督注意力机制在影响 SL-SGAT 方面发挥了巨大的作用。

5 结论与展望

本文提出了基于结构学习的自监督注意力网络 (SL-SGAT), 通过整合节点特征相似度矩阵和全局结构邻近度来构建新图, 在新图上使用新的注意力机制, 这样能充分利用图的拓扑结构和节点的特征信息。同时设计一个链接预测的自监督注意力辅助任务, SL-SGAT 在区分边链接是否正确时, 会学习到更具有表现力的节点表示, 降低噪声对节点表示的影响。解决了图注意力网络只使用图的一阶邻居节点而特征聚集不足的问题, 同时也降低噪声对节点表示的影响。根据关系预测任务, 可以知道节点与其邻居节点之前是否存在边, 从而知道聚合哪些邻居节点特征, 增加了模型的可解释性。该模型整合了注意力机制和图结构学习的优势, 既适用于有结构图, 也适用于无结构图。为验证模型的有效性, 在三个真实世界数据集上进行节点分类实验。结果表明, 与先进的基线模型进行对比, 本文所提出的模型可获得更好的性能或者与之相当的性能。未来, 将把本模型应用到无图结构的数据集上, 设计更好的注意力机制, 提高模型性能, 增强模型的鲁棒性。

参考文献:

- [1] ZHOU J Y, LIU L, WEI W Q, *et al.* Network Representation Learning: From Preprocessing, Feature Extraction to Node Embedding[J]. *ACM Comput Surv*, 2023, **55**(2): 1–35. DOI: 10.1145/3491206.
- [2] CAI H Y, ZHENG V W, CHANG K C C. A Comprehensive Survey of Graph Embedding: Problems, Techniques, and Applications[J]. *IEEE Trans Knowl Data Eng*, 2018, **30**(9): 1616–1637. DOI: 10.1109/TKDE.2018.2807452.
- [3] WU Q, ZHAO W, LI Z, *et al.* Nodeformer: A Scalable Graph Structure Learning Transformer for Node Classification[J]. *Adv Neural Inf Process Sys*, 2022, **35**: 27387–27401. DOI: 10.48550/arXiv.2306.08385.
- [4] TONG N, TANG Y, CHEN B, *et al.* Representation Learning Using Attention Network and CNN for Heterogeneous Networks[J]. *Expert Syst Appl*, 2021, **185**: 115628. DOI: 10.1016/j.eswa.2021.115628.
- [5] YU W H, LIN X, LIU J F, *et al.* Self-propagation Graph Neural Network for Recommendation[J]. *IEEE Trans Knowl Data Eng*, 2022, **34**(12): 5993–6002. DOI: 10.1109/TKDE.2021.3076772.
- [6] CHAMI I, YING R, RE C, *et al.* Hyperbolic Graph Convolutional Neural Networks[C]//Proceedings of the 33rd International Conference on Neural Information Processing Systems. Red Hook, NY, United States: MIT, 2019: 4868–4879.
- [7] HAMILTON W L, YING R, LESKOVEC J. Inductive Representation Learning on Large Graphs[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook, NY, United States: MIT, 2017: 1025–1035.
- [8] KIPF T N, WELING M. Semi-supervised Classification with Graph Convolutional Networks[EB/OL]. (2017-02-22) [2024-04-23]. <https://doi.org/10.48550/arXiv.1609.02907>.
- [9] VELICKOVIC P, FEDUS W, HAMILTON W L, *et al.* Deep Graph Infomax[EB/OL]. (2018-12-21) [2024-04-23]. <https://doi.org/10.48550/arXiv.1809.10341>.
- [10] JIN W, DERR T, WANG Y Q, *et al.* Node Similarity Preserving Graph Convolutional Networks[C]//Proceedings of the 14th ACM International Conference on Web Search and Data Mining. New York: ACM, 2021: 148–156. DOI: 10.1145/3437963.3441735.
- [11] VELICKOVIC P, CUCURULL G, CASANOVA A, *et al.* Graph Attention Networks[J]. *Stat*, 2017, **1050**(20): 10–48550. DOI: 10.17863/CAM.48429.
- [12] HE T, ONG Y S, BAI L. Learning Conjoint Attentions for Graph Neural Nets[J]. *Adv Neural Inf Process Syst*, 2021, **34**: 2641–2653. DOI: 10.48550/arXiv.2102.03147.
- [13] YE Y, JI S H. Sparse Graph Attention Networks[J]. *IEEE Trans Knowl Data Eng*, 2023, **35**(1): 905–916. DOI: 10.1109/TKDE.2021.3072345.
- [14] HOFMANN T, BUHMANN J. Multidimensional Scaling and Data Clustering[J]. *Adv Neural Inf Process*

- Sys, 1995, 7:459. DOI: 10.5555/2998687.2998744.
- [15] BALASUBRAMANIAN M, SCHWARTZ E L. The Isomap Algorithm and Topological Stability[J]. *Science*, 2002, **295**(5552): 7. DOI: 10.1126/science.295.5552.7a.
- [16] ANDERSON W N, MORLEY T D. Eigenvalues of the Laplacian of a Graph[J]. *Linear Multilinear A*, 1985, **18**(2): 141-145. DOI: 10.1080/03081088508817681.
- [17] HE X, NIYOGE P. Locality Preserving Projections[C]//Advances in Neural Information Processing Systems 16: Proceedings of the 2003 Conference. Cambridge: MIT Press, 2004, **16**: 153.
- [18] PEROZZI B, AL-ROUFI R, SKIENA S. DeepWalk: Online Learning of Social Representations[C]//Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2014: 701-710. DOI: 10.1145/2623330.2623732.
- [19] 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. New York: ACM, 2016: 855-864. DOI: 10.1145/2939672.2939754.
- [20] TANG J, QU M, WANG M Z, *et al.* LINE: Large-scale Information Network Embedding[C]//Proceedings of the 24th International Conference on World Wide Web. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee, 2015: 1067-1077. DOI: 10.1145/2736277.2741093.
- [21] WANG J, ZHANG D, WEI J, *et al.* Node Embedding with Capsule Generation-Embedding Network[J]. *Int J Mach Learn Cybern*, 2023, **14**(7): 2511-2528. DOI: 10.1007/s13042-023-01779-9.
- [22] LEVIE R, MONTI F, BRESSON X, *et al.* CayleyNets: Graph Convolutional Neural Networks with Complex Rational Spectral Filters[J]. *IEEE Trans Signal Process*, 2019, **67**(1): 97-109. DOI: 10.1109/TSP.2018.2879624.
- [23] LI R, WANG S, ZHU F, *et al.* Adaptive Graph Convolutional Neural Networks[C]//Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence. New Orleans Louisiana USA: AAAI, 2018: 3546-3553.
- [24] ZHUANG C Y, MA Q. Dual Graph Convolutional Networks for Graph-based Semi-supervised Classification [C]//Proceedings of the 2018 World Wide Web Conference on World Wide Web-WWW '18. New York: ACM, 2018: 499-508. DOI: 10.1145/3178876.3186116.
- [25] YANG Y D, WANG X C, SONG M L, *et al.* SPAGAN: Shortest Path Graph Attention Network[EB/OL]. (2021-01-10) [2024-04-23]. <https://doi.org/10.48550/arXiv.2101.03464>.
- [26] YUAN J L, CAO M, CHENG H, *et al.* A Unified Structure Learning Framework for Graph Attention Networks[J]. *Neurocomputing*, 2022, **495**: 194-204. DOI: 10.1016/j.neucom.2022.01.064.
- [27] KIM D, OH A. How to Find your Friendly Neighborhood: Graph Attention Design with Self-supervision[EB/OL]. (2022-04-11) [2024-04-23]. <https://doi.org/10.48550/arXiv.2204.04879>.
- [28] THANOU D, DONG X W, KRESSNER D, *et al.* Learning Heat Diffusion Graphs[J]. *IEEE Trans Signal Inf Process Netw*, 2017, **3**(3): 484-499. DOI: 10.1109/TSPN.2017.2731164.
- [29] COMMAULT C, VAN DER WOUDE J. A Classification of Nodes for Structural Controllability[J]. *IEEE Trans Autom Contr*, 2019, **64**(9): 3877-3882. DOI: 10.1109/TAC.2018.2886181.
- [30] TURNEY P D, PANTEL P. From Frequency to Meaning: Vector Space Models of Semantics[J]. *Jair*, 2010, **37**: 141-188. DOI: 10.1613/jair.2934.
- [31] SEN P, NAMATA G, BILGIC M, *et al.* Collective Classification in Network Data[J]. *AI Mag*, 2008, **29**(3): 93-106. DOI: 10.1609/aimag.v29i3.2157.
- [32] VAN DER MAATEN L, HINTON G. Visualizing Data Using T-SNE[J]. *J Mach Learn Res*, 2008, **9**: 2579-2625.