

# 基于改进图注意机制的网络嵌入方法研究及应用

韩津津<sup>1</sup>, 李智杰<sup>1</sup>, 李昌华<sup>2</sup>, 张 颖<sup>2</sup>

(西安建筑科技大学 信息与控制工程学院, 西安 710055)

**摘要:** 网络已被广泛用作抽象现实世界系统以及组织实体之间关系的数据结构; 网络嵌入模型是将网络中的节点映射为连续向量空间表示的强大工具; 基于图卷积 (GCN, graph convolutional neural, GCN) 的网络嵌入方法因受其模型迭代过程参数随机优化和聚合函数的影响, 容易造成原始节点特征信息丢失的问题; 为有效提升网络嵌入效果, 针对于图神经网络模型在网络嵌入中节点表征学习的局限性, 提出了一种基于二阶邻域基数保留策略的图注意力网络 (SNCR-GAT, second-order neighborhood cardinality retention strategy graph attention network), 通过聚合二阶邻域特征基数的方式, 解决网络节点潜在特征学习过程中重要信息保留问题; 通过在节点分类和可视化两个网络嵌入应用任务上进行实验, 结果表明, SNCR-GAT 模型在网络嵌入上的性能表现相比较基准方法更具优越性。

**关键词:** 图注意力; 节点分类; 特征学习; 二阶邻域; 基数保留

## Research and Application of Network Embedding Method Based on Improved Graph Attention Mechanism

HAN Jinjin<sup>1</sup>, LI Zhijie<sup>1</sup>, LI Changhua<sup>2</sup>, ZHANG Jie<sup>2</sup>

(Xi'an University of Architecture and Technology, School of Information and Control Engineering, Xi'an 710054, China)

**Abstract:** Networks have been widely used as data structures for abstracting the real-world systems and for organizing the relationships between entities. The network embedding model is a powerful tool to map the nodes in the network into a continuous vector space representation. The network embedding method based on graph convolutional neural (GCN) is easily affected by the random optimization of parameters in the model iteration process and the influence of the aggregation function. The problem of original node feature loss information occurs. In order to effectively improve the network embedding effect, a graph attention network based on the second-order neighborhood cardinality retention strategy graph attention network (SNCR-GAT) is proposed for the limitation of the graph neural network model in the node representation learning in the network embedding, by aggregating the second-order neighborhood feature cardinality, the problem of important information retention in the process of latent feature learning of network nodes is solved; By classifying and visualizing two networks in nodes, the experiments are carried out on the actual task of embedding, and the results show that the performance of the SNCR-GAT model on network embedding is more superior than that of the baseline method.

**Keywords:** graph attention; node classification; feature learning; second-order neighborhood; base retention

## 0 引言

信息网络是最为常见的一种信息载体和形式, 随着互联网的不断发展, 各种应用场景下极为丰富的信息网络数据已成为日常生活中不可或缺的一部分<sup>[1]</sup>。网络数据结构可以自然地表达物体与物体间的联系, 用于对数据信息进行系统的建模。如社交媒体网络中热点话题的深度挖掘<sup>[2]</sup>、移动通信网络中用户群体的分类<sup>[3]</sup>以及论文推荐系统中引文网络的划分<sup>[4]</sup>等。在这些场景中, 用户、项目、分子和知识概念等现实实体被抽象为网络中的节点, 而实体之间的关系被建模为它们之间的链接<sup>[3]</sup>。基于网络嵌入的表示学

习研究旨在探索更好地分析复杂信息网络中节点间的联系, 寻找解决信息网络背景下的各种实际问题的普适方法, 有效融合网络结构与节点外部信息, 形成更具区分性的网络嵌入表示<sup>[6]</sup>。网络嵌入的目的是将每个节点映射到一个低维向量空间, 得到节点的低维向量表示。节点的表示向量可以应用到许多下游分析任务, 如网络节点分类<sup>[7]</sup>、可视化<sup>[8]</sup>等。为获得有效的网络嵌入并在实际任务中获得更好的表现, 需要充分利用图的拓扑结构及节点特征中蕴含的丰富信息。

网络数据信息特征挖掘首先需要面对的挑战是如何恰

收稿日期: 2022-04-27; 修回日期: 2022-05-30。

基金项目: 国家自然科学基金(61373112, 51878536); 陕西省自然科学基金(2020JQ-687); 陕西省住房和城乡建设科技计划项目(2020-K09)。

作者简介: 韩津津(1995-), 男, 陕西安康人, 硕士, 主要从事深度学习、图嵌入等方向的研究。

李智杰(1980-), 男, 河南人, 博士, 副教授, 主要从事模式识别、数字建筑等方向的研究。

李昌华(1963-), 男, 宁夏银川人, 博士, 教授, 主要从事图形图像处理、模式识别、数字建筑等方向的研究。

引用格式: 韩津津, 李智杰, 李昌华, 等. 基于改进图注意机制的网络嵌入方法研究及应用[J]. 计算机测量与控制, 2022, 30(9): 207-212.

当的表示网络数据<sup>[9]</sup>。早期的网络嵌入算法基于矩阵特征向量计算,利用相近节点其特征相似的思想,通过对矩阵分解来近似网络嵌入中节点的特征向量表示,例如 GraRep<sup>[10]</sup>、HOPE<sup>[11]</sup>等模型,但这种方式对于大规模的网络嵌入是非常耗时的。随机游走通过近似网络节点中心性和相似性等属性来获得网络嵌入表示,这针对大型网络图是有效的,典型的模型有 DeepWalk<sup>[12]</sup>、node2vec<sup>[13]</sup>等。

图神经网络(GNN, graph neural network)<sup>[14]</sup>的研究解决了非欧氏空间结构网络数据在经过传统离散卷积的过程中无法保持平移不变形的局限性。利用实体之间的内在关系,为网络嵌入学习开辟了一条独特的道路。GNN 通过聚合目标节点的邻域来更新自身新的嵌入表示,大多数 GNN 变体将在其聚合器中分配将和之间的非参数化权重。例如传统的图卷积神经网络(GCN, graph convolutional neural networks)<sup>[15]</sup>因其强大的迭代训练模式和聚合信息的能力,能够实现网络嵌入中聚合一阶邻域信息近似的节点特征更新方式,但是训练过程中由于共享相同的参数化权重,因此无法学习和区分目标节点及其邻居之间的信息,然而考虑网络中节点的不同贡献在现实世界数据中是非常重要的,因为并非所有的边都有相似的影响。一个自然的替代解决方案是使边权重能够训练,以具有更好的表达能力。为了在聚合中分配可学习的权重,图注意力网络(GAT, graph attention network)<sup>[16]</sup>在进行邻域聚合时,采用注意力机制为目标节点的邻域节点分配不同的权重系数,以区分邻居对目标节点的不同贡献程度,以此来聚合网络中的节点信息表征。尽管在 GNN 聚合器中加入注意力机制在各种任务上取得了令人满意的性能,但是存在两个局限性:首先,图卷积过程中只能聚合中心节点的一阶邻域特征,无法获得高阶邻域信息来扩大局部网络结构的感受野,这将造成网络嵌入表征局部节点划分不明确的问题,其次,通过注意力机制筛选节点造成的问题是权重分配虽然将节点特征按照重要程度进行划分,但是当下一层网络训练时会受到权重特征筛选的影响,无法将最原始的基数信息作为新一层的网络训练<sup>[17]</sup>,这些基数信息包含了网络节点的属性特征和局部结构信息,由于注意力权重,这些基数信息会被过滤掉一部分,这致使网络嵌入表征学习过程会丢失掉初始信息,影响了对于网络结构的判别能力。

为了解决这些问题,本文基于图注意力机制为出发点,探索可解决的改进方案,提出了一种二阶邻域基数保留策略的图注意力神经网络模型(SNCR-GAT),综合考虑网络嵌入中的多邻域信息聚合以及模型训练中的基数信息保留问题。本文的贡献有以下 3 点:1) 基于网络节点邻接矩阵的幂实现二阶邻域连通性,以扩大局部信息的收敛感受野来聚合二阶邻域信息;2) 为了保证模型在训练阶段保留更多的基数信息,对注意力机制更新网络节点特征的聚合函数进行了一些改进;以实现基数信息的保留;3) 为验证 SNCR-GAT 模型的网络嵌入效果,我们在 3 个真实的网络数据集 Cora、Citeseer、Pumbed 上进行了节点分类及可

可视化两个实际应用实验,相比较其它基准图神经网络模型具有更高的网络嵌入效果。

## 1 相关工作

图注意力网络是最流行的 GNN 体系结构之一,被认为是最先进的图表征学习体系结构。图注意力机制能够通过计算节点间不同的权重值对节点邻域进行筛选,从而保留有效的局部结构。在 GAT 中,输入为图的邻接矩阵  $A$  和节点的特征矩阵  $h$ ,其中目标节点邻域特征向量集合  $h = \{h_1, h_2, \dots, h_N\}$  以组为单位,  $h_i \in R^F$ ,其中  $N$  是节点个数,  $F$  是单个节点的特征数,将这个集合作为图注意卷积层的输入,通过训练,每次迭代都会将邻域节点特征进行加权求和输出新的节点特征组(可能具有不同的特征数  $F'$ ),  $h' = \{h'_1, h'_2, \dots, h'_N\}$  作为输出,其中  $h'_i \in R^{F'}$ 。同时为了使特征具有更高的表达信息能力,GAT 模型利用权重矩阵  $W \in R^{F' \times F}$  为每一层节点的 embedding 都进行线性变换。

如图 1 所示,目标节点  $h_i$  通过聚合周围一阶邻域节点的特征,进行合并或平均操作,以生成新的特征表示  $h'_i$ ,其中邻域中每个邻居节点由注意力机制分配一个贡献度权重。

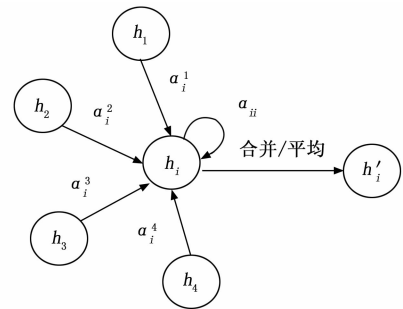


图 1 图注意力信息特征聚合方式

节点特征聚合依赖于 Attention 对邻域重要节点的选择,在  $K$  阶邻域生成的邻接矩阵中,对于每个节点  $x_i$ ,为了表现目标节点对待邻域节点选择的偏向不同,注意力机制  $\alpha: R^{F' \times F} \rightarrow R$  是共享的,它为每个邻域节点计算一个与目标节点的相关度值,这个值决定了该邻域对中心节点的重要程度。计算方式如下:

$$e_{ij} = \text{LeakyReLU}(\alpha^T [Wh_i \parallel Wh_j]) \quad (1)$$

式中,  $e_{ij}$  表示邻居节点  $j$  对目标节点  $i$  的重要度,  $[\cdot \parallel \cdot]$  将  $Wh_i$  和  $Wh_j$  串联起来,作为模型的输入,  $\alpha$  是隐层的参数,是一个  $2 \times F'$  的特征向量,激活单元采用了非线性激活函数 LeakyReLU 进行激活。其中  $W \in R^{F' \times F}$  为单层图注意网络中的权重矩阵,需要利用网络训练学习获得,  $F$  和  $F'$  分别为网络层输入和输出层特征个数,权重矩阵建立了这两个参数之间的映射关系。而  $h_j$  是节点  $i$  邻居节点  $j$  的特征向量表示,为反映了节点  $j$  对于节点  $i$  的重要性程度,通过在节点  $i$  和节点  $j$  之间赋予权重值来表现这种重要性。为了使不同邻居的贡献度可以对比,使用 softmax 归一化,最终得到邻居  $j$  对节点  $i$  生成新特征的贡献度  $\alpha_{ij}$ :

$$\alpha_{ij} = \text{softmax}_j(e(h_i, h_j)) = \frac{\exp(e(h_i, h_j))}{\sum_{j \in N_i} \exp(e(h_i, h_j))} \quad (2)$$

注意力机制 Attention 是一种静态非线性映射, 基于单层结构实现的前馈神经网络, 能够高效完成任意大小样本集的训练。节点间的注意系数  $\alpha_{ij} \in R^{2^F}$  作为网络参数参与训练, 并使用 LeakyReLU 作为非线性激活函数, 其计算方式如下:

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(a^T[Wh_i \| Wh_j]))}{\sum_{k \in N_i} \exp(\text{LeakyReLU}(a^T[Wh_i \| Wh_k]))} \quad (3)$$

在得到目标节点与每个邻居节点的注意系数之后, 再对节点  $i$  的邻居特征按贡献度  $\alpha_{ij}$  进行加权平均, 再进行一个非线性转换, 得到目标节点  $i$  新的特征向量表示  $h'_i$ :

$$h'_i = \sigma(\sum_{j \in N_i} \alpha_{ij} Wh_j) \quad (4)$$

根据公式, 对于节点  $i$  新的特征向量表示聚合方式是: 先将所有邻居节点特征向量  $h_j$  与权重矩阵  $W$  进行线性求和, 同时乘以注意力权重系数, 最后将线性求和进行非线性转换。

## 2 二阶邻域基数信息保留策略

### 2.1 二阶邻域节点信息聚合方案

为了深度挖掘网络数据中局部结构特征, 聚合更多的邻域信息来生成有效的目标节点嵌入表示, 采用了如图 2 所示的二阶邻域信息聚合方案。

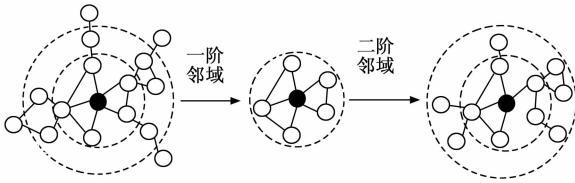


图 2 二阶邻域信息聚合

二邻域信息聚合通过网络节点邻接矩阵的 2 次幂作为屏蔽矩阵实现 2 阶邻域的信息收敛, 从各阶邻域提取局部结构特征, 获得节点的多个中间表示。通过增加邻接矩阵的 2 次幂能够实现目标节点与 2 阶邻居的连通性, 获取到 2 阶邻居的信息传递, 通过邻接矩阵得到目标节点 2 跳邻域的边连接, 如下所示:

$$\begin{array}{c}
 \begin{array}{ccccc}
 v_2 & & v_3 & & \\
 & \diagdown & & \diagup & \\
 & v_1 & & & \\
 & \diagup & & \diagdown & \\
 v_4 & & v_5 & & 
 \end{array}
 \longrightarrow
 \mathbf{A} = \begin{bmatrix}
 0 & 0 & 1 & 1 & 1 \\
 0 & 0 & 1 & 1 & 0 \\
 1 & 1 & 0 & 0 & 0 \\
 1 & 1 & 0 & 0 & 1 \\
 1 & 0 & 0 & 1 & 0
 \end{bmatrix} \\
 \\
 \mathbf{A}^2 = \begin{bmatrix}
 0 & 0 & 1 & 1 & 1 \\
 0 & 0 & 1 & 1 & 0 \\
 1 & 1 & 0 & 0 & 0 \\
 1 & 1 & 0 & 0 & 1 \\
 1 & 0 & 0 & 1 & 0
 \end{bmatrix}
 \begin{bmatrix}
 0 & 0 & 1 & 1 & 1 \\
 0 & 0 & 1 & 1 & 0 \\
 1 & 1 & 0 & 0 & 0 \\
 1 & 1 & 0 & 0 & 1 \\
 1 & 0 & 0 & 1 & 0
 \end{bmatrix}
 = \begin{bmatrix}
 3 & 2 & 0 & 1 & 1 \\
 2 & 2 & 0 & 0 & 1 \\
 0 & 0 & 2 & 2 & 1 \\
 1 & 0 & 2 & 3 & 1 \\
 1 & 1 & 1 & 1 & 2
 \end{bmatrix}
 \end{array}$$

在一阶邻域邻接矩阵  $\mathbf{A}$  中, 当两个点之间存在一条边时, 矩阵中元素  $\alpha_{ij} = 1$ , 不存在边时  $\alpha_{ij} = 0$ , 这表明从点  $v_i$  出发走一跳到点  $v_j$  有 1 条路径。当对邻接矩阵  $\mathbf{A}$  平方之后,

对角线元素  $\alpha_{ii}$  表示了当前节点的度, 并且跳数加一, 每一个元素  $\alpha_{ij}$  表示的是当前目标节点  $v_i$  经过两跳到达邻居节点  $v_j$  的路径有多少, 以节点  $v_i \rightarrow v_j$  为例, 经过两跳可到达的路径数为 2, 分别是  $v_i \rightarrow v_4 \rightarrow v_2$  和  $v_i \rightarrow v_3 \rightarrow v_2$ 。因此可以说  $\mathbf{A}^2$  建立了以中心节点为目标的周围二阶邻域的边连接, 由于节点之间是通过边进行信息传递的, 这样的操作首先增加了图的连通性, 保障了目标节点更新新的节点过程中聚合更多的局部邻域结构信息, 对于目标节点, 扩大了局部信息收敛的感受野, 同时也提高了嵌入向量的有效性。

### 2.2 基数保留策略

图注意力网络训练过程中目标节点特征向量的更新依赖于邻域所有节点特征向量的聚合。对于节点  $i$  其邻域节点集合是  $N(i)$ , 其特征向量集合为  $H(N(i)) = \{h_1, h_2, \dots, h_N\}$ , 其中  $N$  是邻域节点数, 对于邻域中的每个节点共享特征向量集合。Attention 机制通过权重值来筛选相近的邻域节点以保留有效的局部结构, 同时图卷积网络模型在更新第  $l$  层向量表示  $h'_i$  时, 只有具有相同目标节点特征和相同节点特征分布的  $H(N(i))$  才会被聚合。在这个过程中, 由于受到随机优化算法的影响, 导致加权求和过程中无法区分  $H(N(i))$  中共有的相同节点特征分布, 这就忽略了目标节点相关的基数信息, 这些基数信息包含上一层网络输入中未经过 Attention 筛选的原始信息, 经过筛选重要的邻域节点会继续参与网络下一次迭代训练, 这种方式使得基于注意力的聚合器无法区分特定的不同结构 (局部或全局结构), 尽管注意力机制能够判别局部小范围的图结构, 但为了减少注意力权重计算中受随机算法影响的误差, 希望能够在网络训练过程中用到更多的原始基数信息。

为提高图注意力模型在节点嵌入中的结构判别能力, 保证在不改变注意力函数保持原来的注意机制学习能力的情况下, 在式 (4) 目标节点特征聚合方式上优化了每一层网络更新新的节点向量特征的输出方式:

$$h'_i = \sigma^l(\sum_{j \in N_i} \alpha_{ij}^{l-1} h_j^{l-1} + h_j^{l-1}) \quad (5)$$

式中, 其中  $l$  是网络层数,  $\sigma^l$  是第  $l$  的非线性激活函数,  $h_j^{l-1}$  是  $l-1$  层网络中目标节点所有邻居节点特征与  $W$  进行点积最后的结果。表达式最终简化为:

$$h'_i = \sigma^l(\sum_{j \in N_i} (\alpha_{ij}^{l-1} + 1) h_j^{l-1}) \quad (6)$$

在更新下一层节点特征的时候, 保留了不经过权重筛选的原始节点特征  $h_j^{l-1}$ , 保证了在进入下一层网络训练时依然存在最原始的节点特征所包含的各种基数信息能够被下一层网络引入训练。

SNCR-GAT 模型训练过程如下:

输入: 网络数据集 Cora、Citeseer 和 Pumbed

处理: 1) 节点特征预处理 (划分节点特征和标签信息);  
2) 应用稀疏矩阵存储节点特征, 划分训练集、验证集和测试集;

3) 构建二阶信息收敛的邻接矩阵。

训练: 1) 初始化权重用于输入特征的线性变化;

- 2) 计算参数化权重向量并初始化;
- 3) 线性变换节点特征;
- 4) 利用公式 (1) 计算注意力系数;
- 5) Softmax 归一化邻居节点特征;
- 6) 根据基数保留策略公式 (6) 生成嵌入向量矩阵。

输出: 节点分类预测精度。

### 3 实验与分析

#### 3.1 数据集

为了验证本文所提模型 SNCR-GAT 的实际效果, 在 3 个引文网络数据集 Cora, Citeseer 和 Pumbed 上进行实验。每个数据集中的节点表示一篇文章, 边构成引用关系, 数据集详细信息见表 1。

表 1 数据集信息描述

数据集	节点数	边数	特征维度	类别数
Cora	2 708	5 429	1 433	7
Citeseer	3 327	4 732	3 703	6
Pumed	19 717	44 338	500	3

#### 3.2 实验设置

实验是基于半监督的训练模式, 主网络采用两层基于注意力机制的 GCN, 第一层用来学习节点的特征表示, 利用 12 个多头注意力端来平衡模型训练, 特征向量聚合后非线性激活采用 LeakRelu 函数, 第二层网络是数据节点分类层, 通过 softmax 函数实现。

为了能够找到适合模型性能的最好参数, 在参考其它 GNN 模型的最优参数的同时也进行了自行调整, 并且为了方便比较模型的实验效果, 采取控制单一变量的原则构建实验所涉及对比的图神经网络模型, 为了验证改进模型在图嵌入时的表现效果, 对多个不同的标准数据集进行了训练, 测试以及验证。实验中初始网络学习率为 0.02, 模型优化采用自适应的 Adam<sup>[18]</sup> 优化器, 统一设置 Dropout 率为 0.6, 模型训练迭代 100 次, 经过测试随着迭代次数的增加, 结果会出现平滑现象, 同时采用多头注意力机制来增加模型训练的稳定性, 实验环境配置见表 2。

表 2 实验环境配置

实验环境	环境配置
操作系统	Windows 系统
处理器	Intel(R) Core(TM) i7-8700
内存	32 GB
编程语言	Python3.6
深度学习框架	Pytorch

#### 3.3 节点分类应用实验

##### 3.3.1 模型对比

实验结果将所提模型 SNCR-GAT 与其它基准方法生成的网络嵌入进行比较, 包括通过随机游走生成节点序列学习网络嵌入的 DeepWalk、学习节点连续特征表示的 node2vec、基于谱分解定义卷积操作的 GCN、通过 Atten-

tion 生成不同权重的基准模型 GAT。前面两个方法都是通过中心节点沿边游走, 利用后续节点出现的概率保持节点的高度相似来生成节点特征表示。而基于神经网络训练的 GCN 和 GAT 则是利用迭代训练聚合不同的邻域信息来生成网络节点嵌入。通过对比这几个典型的基本算法在节点嵌入上的效果来表现 SNCR-GAT 的优势。

##### 3.3.2 结果与分析

为验证 SNCR-GAT 模型相较于在网络嵌入中的整体分类效果, 与文献 [14] 模型进行了对比实验, 同时为保证实验结果的公平准确性, 采用了十折交叉验证<sup>[19]</sup>测试模型, 将数据集分成 10 份, 每一份都作为一次测试集来进行实验, 最终结果以 10 次结果均值为准。10 次实验结果对比如图 3 所示。其中横轴代表每一份测试集, 纵轴表示分类精度 (单位:%)。

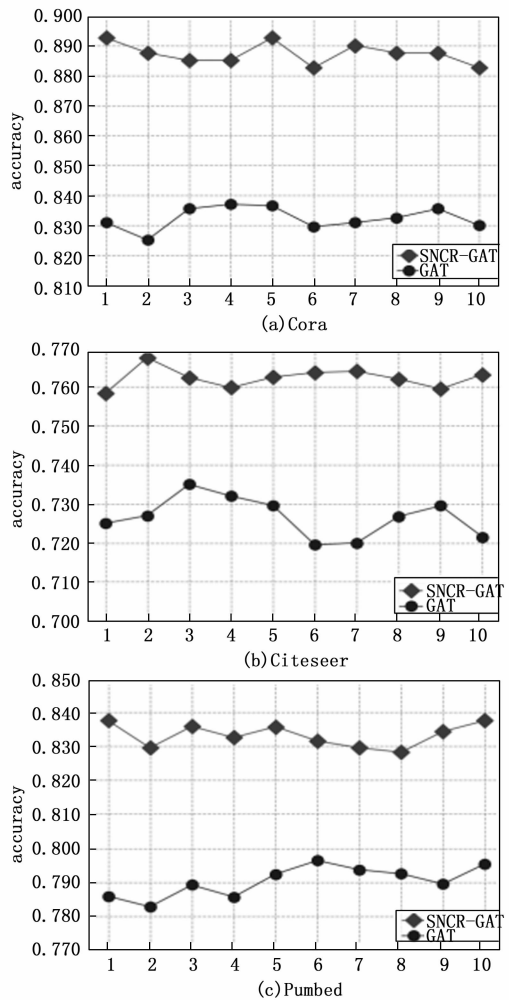


图 3 十折交叉验证节点分类精度对比

实验结果表明, 本文采用的二阶邻域基数保留策略在网络嵌入上的效果的确优于基准 GAT 模型, 分析其原因, SNCR-GAT 在目标节点特征更新上能够聚合更大范围内的邻域信息, 保留更多的结构属性。同时也减小了网络训练中受注意力系数筛选影响导致的部分早期基数信息的丢失,

这为提高节点分类任务准确率方面的表现做出了极大的贡献。总的实验结果对比汇总在表 3 中。

表 3 节点分类精确率对比

模型	Cora	Citeseer	Pumbed
DeepWalk	67.2%	43.2%	65.3%
node2vec	75.7%	64.7%	77.2%
GCN	81.5%	70.3%	79.0%
GAT	83.7%	73.2%	79.3%
SNCR-GAT(ours)	89.3%	76.8%	83.8%

通过比较节点的分类精度, SNCR-GAT 模型在总体上表现良好, 在 Cora、Citeseer、Pubmed 三个数据集上的训练准确率相比较原始的 GAT 网络提升分别为: 5.6、3.6 和 4.5。特别是 Cora 数据集上分类精度提升明显。这也表明 SNCR-GAT 模型在图嵌入过程中进行节点表示向量更新时的, 有效的减少了信息丢失的问题, 重要的是在基于二阶邻域的信息聚合方式对于当前节点更新新的向量表示提供了更广泛的信息, 基数保留信息聚合策略对于重要的邻域信息选择更可靠。

### 3.3.3 复杂度分析

SNCR-GAT 在训练阶段首先会将节点的特征向量做一个维数  $F \rightarrow F'$  的空间映射, 其时间复杂度为  $O(|V| \times F \times F')$ ,  $|V|$  是网络中节点数; 另外注意力机制  $\alpha(\cdot)$  是将  $2F'$  维的特征向量映射为实数的过程, 因此会产生  $O(|E| \times F')$  的计算复杂度, 其中  $|E|$  是边数。二阶邻域聚合过程是一个加权求和的过程, 只涉及到一次邻接矩阵的平方运算, 同时基数保留的聚合函数做了一次特征向量的加法运算, 均不存在高复杂度的乘法运算。采用二阶邻域基数保留策略能够使模型在最少的 epoch 次数下收敛到最好的结果, 综合评价, SNCR-GAT 模型在复杂度上也存在明显的优势。

### 3.3.4 阶邻域分析

为了证明二阶邻域信息收敛的有效性, 在实验中以逐步增长阶数的方式测试采用不同阶邻域对模型整体的影响, 随着阶数增加, 利用节点分类实验结果和算法模型训练时间进行评估, 同样地, 对每一阶的测试中, 模型运行 10 次, 以平均准确率为作为最终结果。实验在 Cora 数据集上测试, 如图 4 所示, 为方便比较, 将节点分类精确度曲线与时间曲线在同一图上进行描述, 横轴代表邻域阶数  $k$ , 纵轴代表时间 (min)。

实验结果表明, 随着邻域阶数的增加, 实验结果没有受到太大的波动, 反观模型收敛时间在增加。分析其原因, 考虑二阶以上的邻域信息收敛会致使 SNCR-GAT 模型可能会从不太相关的邻居那里获取大量信息, 从而无法学习合适的节点表示, 因此影响到网络嵌入节点分类的精度, 同时阶数的增加, 需要计算邻接矩阵的次幂来增加图的  $k$  跳连通性以传递信息, 这会提高模型初期数据特征处理的计算复杂度, 训练时间会更长, 不利于实验的高效进行, 因

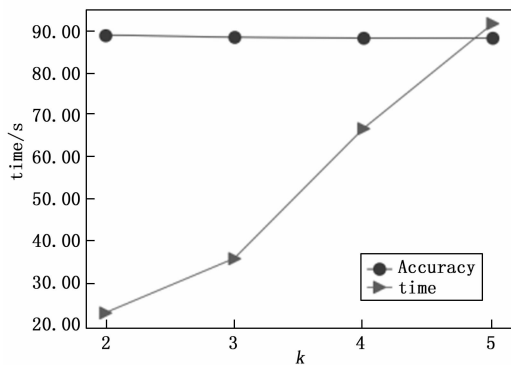


图 4  $k$  阶邻域节点分类准确率对比

此基于二阶邻域下的基数信息保留方案是更有效的, 在保证模型性能的前提下, 控制时间复杂度在可行的范围内是可取的。

### 3.4 网络可视化应用实验

对网络嵌入效果的另一种评估方式是对学习到的网络嵌入表示进行可视化展示。网络嵌入向量作为  $d$  维空间的表示, 无法直观的根据向量特征进行评估模型的嵌入效果。我们以 Cora 数据集进行可视化展示来评估 SNCR-GAT 的网络嵌入效果。对比模型选用了 GCN, GAT 这两个同为聚合收敛的基准模型, 首先将对对比模型及我们的模型锁所学习到的节点嵌入向量输入到  $t$  分布随机邻域嵌入 (t-SNE, t-distributed stochastic neighbor embedding)<sup>[20]</sup> 模型中, t-SNE 能够通过降维的方式将网络嵌入向量表示映射到指定维数空间中进行展示, 为了能够在分类可视化的基础上更好的观察网络嵌入效果的层次结构划分, 我们在三维空间中实现了所有模型的可视化。

图 5 展示了 Cora 数据集在不同模型学习到的网络嵌入表示的三维可视化效果。其中每一个点在三维空间中都有一个坐标, 以实现可视化。每一种类别标签对应一种形状, 具有相同类别标签的节点再二维空间中对应的点具有相同的形状。对比 GCN、GAT 和 SNCR-GAT 在三维可视化上的节点分类效果, 可以通过明显的层次结构清楚的看到节点之间的类别划分, 同时在 GCN 的可视化任务中, 不同类别的簇形成鲜明的对比, 但是簇与簇之间的簇间距很小导致一些相邻边界之间的点划入错误的类别中。比较明显的是, SNCR-GAT 在可视化任务上性能更好, 表现出了清晰的类别和边界距离感, 错分的节点也比较少。因此, 可以认为 SNCR-GAT 在可视化任务上获得了具有竞争性的结果。

## 4 结束语

在本文中, 为解决网络嵌入节点分类不精确, 网络可视化层次结构划分不清晰等问题, 提出了基于二阶邻域基数信息保留策略的改进图注意力机制模型 SNCR-GAT。首先, 通过目标节点的二阶邻域来增大局部感受野以获得更多的网络结构信息, 其次, 使用保留基数信息的聚合方式来避免 GAT 模型训练中节点信息因权重筛选造成的原始特

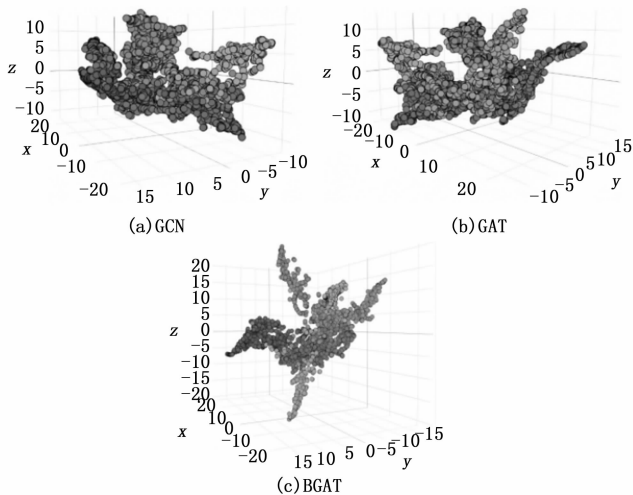


图 5 Cora 数据集 t-SNE 下的三维可视化

征丢失问题, 以此来生成目标节点新的特征向量表示。最后实验结果表明, 与其它基线算法相比本文所改进方案本文展现了更好的优势, 这对于实际的智能应用是可行的, 尤其是在节点分类中模型预测的高精确率更体现了本文方案的优越性以及稳定性。

参考文献:

[1] 孙金清, 周 慧, 赵中英. 网络表示学习方法研究综述 [J]. 山东科技大学学报 (自然科学版), 2021, 40 (1): 117-128.

[2] 江务学. 社交网络中热点话题深度挖掘方法及实验测试 [J]. 计算机测量与控制, 2017, 25 (2): 174-176, 180.

[3] 王 磊, 李 勇, 付国庆. 物联网云平台 ZigBee 无线网络节点设计 [J]. 计算机测量与控制, 2017, 25 (5): 203-206.

[4] 沈小烽, 刘柏嵩, 吴俊超, 等. 边信息嵌入的学术论文推荐算法研究 [J/OL]. 计算机工程与应用: 1-10 [2022-04-27]. https://cnki.net/.

[5] 马 煜. 基于拓扑感知节点排序的虚拟网络嵌入 [J]. 计算机测量与控制, 2019, 27 (7): 236-241.

[6] LIU N, TAN Q, LI Y, et al. Is a single vector enough? exploring node polysemy for network embedding [C] // Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019: 932-940.

[7] HE Z C, LIU J, ZENG Y Y, et al. Content to node: Self-translation network embedding [J]. IEEE Transactions on Knowledge and Data Engineering, 2021, 33 (2): 431-443.

[16] 康朝虎, 刘 宁, 田永瑞, 等. Anubis 在 GNSS 数据质量可视化分析中的应用 [J]. 导航定位学报, 2018, 6 (4): 105-111.

[17] 陈佳清, 易卫兵, 刘熙添, 等. 基于 G-Nut/Anubis 的 GNSS 数据质量检核可视化分析 [J]. 测绘地理信息, 2018, 43 (5): 50-54.

[18] 朱 璇, 王 峰, 周 巍, 等. 西藏 CORS 站 GNSS 观测数据质量评估 [J]. 导航定位学报, 2021, 9 (2): 109-113.

[8] 李 杨, 郝志峰, 肖燕珊, 等. 差分隐私 DPE k-means 数据聚合下的多维数据可视化 [J]. 小型微型计算机系统, 2013, 34 (7): 1637-1640.

[9] 陈湘辉. 基于朴素贝叶斯算法的社交网络数据挖掘技术研究 [J]. 计算机测量与控制, 2017, 25 (6): 199-202.

[10] CAO S, LU W, XU Q. Grarep: Learning graph representations with global structural information [C] // Proceedings of the 24th ACM international on conference on information and knowledge management, 2015: 891-900.

[11] OU M, CUI P, PEI J, et al. Asymmetric transitivity preserving graph embedding [C] // Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining, 2016: 1105-1114.

[12] 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.

[13] 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.

[14] SHAO SHENG C, LU W and QIONG KAI X. Deep neural networks for learning graph representations [C] // Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI' 16), 2016: 1145-1152.

[15] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks [J]. ArXiv preprint arXiv: 1609.02907, 2016.

[16] VELIKOVI P, CUCURULL G, CASANOVA A, et al. Graph Attention Networks [Z]. arXiv preprint arXiv: 1710.10903, 2018.

[17] ZHANG S, XIE L. Improving attention mechanism in graph neural networks via cardinality preservation [C] // IJCAI: proceedings of the conference. NIH Public Access, 2020, 2020: 1395.

[18] KINGMA D, BA J. Adam: A Method for Stochastic Optimization [Z]. arXiv preprint arXiv: 1412.6980, 2014.

[19] 牛晓太. 基于 KNN 算法和 10 折交叉验证法的支持向量选取算法 [J]. 华中师范大学学报 (自然科学版), 2014, 48 (3): 335-338.

[20] 陈婉杰, 盛益强. 基于网络表示学习的非单一维度的社区发现发现算法 [J]. 计算机应用, 2019, 39 (12): 3467-3475.

[19] 陈 浩. 上海 CORS 站点的数据质量检测及可视化表达 [J]. 导航定位学报, 2020, 8 (6): 118-123.

[20] 薛世强, 牟浩嘉, 胡文建. 基于 ANUBIS 的 SSCORS 数据质量分析 [J]. 测绘通报, 2021, (S1): 213-217.

[21] 张 宁, 何正斌, 赵春梅, 等. 北京房山 IGS 站观测数据质量评估 [J]. 导航定位学报, 2021, 9 (3): 73-79.

[22] 丁志鹏, 贺凯飞, 邱立杰, 等. 基于 Anubis 的北斗观测数据质量分析 [J]. 大地测量与地球动力学, 2022, 42 (2): 143-147.