

基于相似网络和联合注意力的图嵌入模型

王静红^{1a,b,c}, 李昌鑫^{1a}, 杨家腾², 于富强^{1a}

(1.河北师范大学 a.计算机与网络空间安全学院;b.河北省网络与信息安全重点实验室;
c.供应链大数据分析与数据安全河北省工程研究中心,石家庄 050024;
2.河北工程技术学院 人工智能与大数据学院,石家庄 050020)

摘要:图注意力网络(graph attention network,GAT)将注意力机制与图神经网络融合,但模型只关注节点的一阶邻域节点,缺乏对高阶相似节点的考虑,同时在计算注意力分数时缺乏对节点结构特征的关注.为此提出一种基于相似网络和联合注意力的图嵌入模型.首先计算网络中的节点相似性,并将高相似度且未连接的节点对构建新边以形成相似网络.其次,引入结构相关性和内容相关性的概念,分别用于表征节点之间的结构关系和内容特征.通过融合两种相关性得分计算得到联合注意力分数.最后使用联合注意力分数对节点特征加权聚合,得到最终的节点嵌入表示.将本文所提算法在Cora、Citeseer和Pubmed 3个数据集上进行节点分类任务,准确率分别达到85.70%、74.30%、84.10%,与原始图注意力网络模型相比分别提高了2.70%、3.94%和2.60%.可见,所提出的算法可以得到更好的节点嵌入表示.

关键词:图嵌入;图注意力网络;节点相似性;相似网络;节点分类

中图分类号:TP181

文献标志码:A

文章编号:1000-2367(2024)06-0036-09

图可以表示许多现实世界的数据集,例如蛋白质结构网络,引文网络以及社交网络等.图的节点和边蕴含着丰富的信息,并可适应多个领域的学习任务.在图的各种分析任务上依赖于可用的图表示,获得用于图数据挖掘的特征表示的关键环节就是图表示学习,又称图嵌入.图表示学习目标是节点映射到低维空间,生成低维、稠密的向量并尽可能保留原始图中的信息.最后生成的向量表示应用于各种下游任务,比如节点聚类^[1]和链路预测^[2]等等.

传统的图嵌入算法,例如矩阵分解(matrix factorization,MF)、Deepwalk^[3]等算法,这类算法在生成节点的低维向量表示时容易丢失初始节点的属性特征,在保留图的特征信息方面存在不足.近年来,神经网络逐渐推广到图数据邻域并取得显著成果,研究人员提出许多图神经网络模型(graph neural network,GNN),其中图注意力网络(GAT)^[4]引起了众多的关注,并被应用于解决大量现实世界的问题,例如节点分类、图像分割和社交推荐等.

图注意力网络将神经网络与注意力机制相结合,旨在通过计算每个节点与其邻居节点之间的注意力权重来动态聚合邻居节点的特征,并将其与当前节点的特征结合起来递到下一层.这种方法能够在多层神经架构中传递和整合来自相邻节点的高度相关特征,从而提高网络性能.

但目前GAT在聚合节点特征时仍有缺点.一方面GAT在聚合节点特征时只有一阶邻居节点被关注,

收稿日期:2023-06-16;**修回日期:**2023-07-04.

基金项目:河北省自然科学基金(F2021205014);河北省高等学校科学技术研究项目(ZD2022139);中央引导地方科技发展资金项目(226Z1808G);河北省归国人才资助项目(C20200340).

作者简介(通信作者):王静红(1967-),女,河北石家庄人,河北师范大学教授,博士,主要研究方向为人工智能与大数据、数据挖掘,E-mail:wangjinghong@126.com.

引用本文:王静红,李昌鑫,杨家腾,等.基于相似网络和联合注意力的图嵌入模型[J].河南师范大学学报(自然科学版),2024,52(6):36-44.(Wang Jinghong, Li Changxin, Yang Jiateng, et al. A graph embedding model based on similar networks and joint attention[J]. Journal of Henan Normal University(Natural Science Edition), 2024, 52(6): 36-44. DOI:10.16366/j.cnki.1000-2367.2023.06.16.0001.)

而与节点密切相关的高阶邻居节点会被忽略.但在真实网络数据集中,如图嵌入中常用的数据集 Cora^[5]、Citeseer^[5]等,节点的度通常较小,仅利用一阶邻居节点会导致在特征聚合时可利用的信息过少.同时,在图嵌入中,相似节点在嵌入空间中也更接近,因此利用高阶相似节点可以丰富节点特征信息以完成更好的节点表征.但若在注意力网络中直接利用高阶邻居节点则会出现过度平滑^[6]问题.另一方面 GAT 在计算注意力分数时主要基于节点内容特征,而很少考虑节点的结构特征.以上反映出 GAT 在利用节点结构信息和高阶相似节点信息方面存在弱点.

针对上述问题,本文提出融合高阶相似节点和节点结构信息的图联合注意力嵌入模型,其关键思想是第一步利用节点相似性度量指标计算节点相似性,然后构建相似网络,使节点能够与其高阶相似节点建立起联系.第二步结合节点内容相关性和结构相关性计算联合注意力分数.最后使用联合注意力分数进行加权聚合.更具体地说,本文的贡献总结如下:1)提出了相似网络构建方法.通过相似性度量指标构建相似网络,使节点与其关系密切的高阶邻居节点建立起联系,成为其新的一阶邻居节点,保证模型可以考虑到其相似度较高的高阶邻居节点.2)提出了联合注意力分数计算方法.结合内容相关性和结构相关性计算联合注意力分数.内容相关性由现有的图注意力机制计算,用来表征节点的内容关系.设计一个自适应距离计算函数计算节点的结构相关性,用来表征节点的结构关系,从而使模型在计算注意力分数时不但关注节点内容信息还考虑到节点结构信息.3)提出基于相似网络和联合注意力的图嵌入模型.利用联合注意力分数进行节点特征加权聚合,得到节点的嵌入表示.将模型在真实世界的 3 个数据集上进行实验,并与多个基准算法进行了比较,取得较好的节点分类结果,表明了本文所提算法的优越性及合理性.

1 相关工作

图嵌入旨在将图中的节点映射到低维空间,生成低维稠密的向量,并保留图原始信息.对于图嵌入算法,大致可以分为 3 类:基于分解的方法、基于随机游走的方法和基于深度学习的方法.

基于分解的图嵌入方法有拉普拉斯特征图(laplacian eigenmaps,LE)算法^[7]、局部保留投影算法(locality preserving projections, LPP)^[8],基于随机游走的方法有 Deepwalk^[8], Node2vec^[9]. Deepwalk 算法将 SkipGram 模型^[10]应用到生成的随机游走上,被认为是第一个基于图表示学习的图嵌入算法.上述基于分解和基于游走的两种方法仅考虑网络的拓扑结构信息.而现实世界中图中的节点都带有丰富的属性信息,因此使用这两种方法学习图嵌入时会忽略节点的属性信息从而影响最终的表示.

然后,神经网络开始应用到图嵌入领域.为了更好地学习图结构数据中的低维表征,迄今为止已经提出了许多图神经网络模型(GNNs),例如图卷积网络(graph convolution network, GCN)^[11]、图注意力网络(GAT)等.谱域 CNNs 中用于特征聚合的函数是根据图的谱表示定义的,如 Spectral CNN^[12]将图域中的卷积运算转换为更为简单的拉普拉斯运算.之后使用谱图理论中卷积运算的方法被提出,如 GCN、SGCN^[13].而空域 GNNs 直接利用中心节点的局部结构属性来定义用于特征聚合的卷积运算,因此必须通过各种处理步骤来适应不同的节点结构.如 Graph SAGE^[14]通过固定领域采样大小聚合特征信息,或根据节点度学习一个权重矩阵^[15].最近,图注意力网络(GAT)通过将图神经网络与注意力机制相结合取得了很大的成功^[4].注意力机制允许处理可变大小的输入并关注最相关的部分,目前已经广泛应用于机器翻译^[16]和视觉处理^[17]. GAT 首先会基于中心节点和一跳邻居节点的特征计算节点之间的注意力分数,然后使用注意力分数来获得节点特征的加权聚合,随后传播到下一层.

然而,图注意力网络在计算注意力分数时严重依赖于低阶邻居节点的节点特征,而高阶邻居节点特征和节点拓扑结构特征往往被忽略.由于 GNN 的过度平滑原因^[6],直接利用低阶邻居节点之间的注意力时, GAT 的性能会下降.因此如何利用 GAT 中的结构信息和相关度极高的高阶节点仍是一个挑战.

研究者开始关注结构因素对注意力分数的影响. CATs 算法^[18]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,但是所提模型预测能力取决于结构干预的质量,易受外部噪声影响. ADSF 算法^[19]关键思想是将每一个节点置于由其高阶邻居节点组成的局部感受域内,感受域内将自适应学习图局部结构. SuperGAT^[20]通过自监督学习改进图注意力机制,利用边的自监督任务来提高对节点

关系重要性的理解,并提出了平衡标签一致性和边存在性的注意力形式.GOAT 模型^[21]通过引入部分信息分解和排列敏感的聚合器,利用自注意力机制学习节点的排序并捕捉邻域节点之间的协同和冗余信息.上述方法并没有考虑到节点相似性问题.SiGraC^[22]模型提出使用基于节点相似性的卷积矩阵计算节点嵌入,但其是基于图卷积思想,并没有考虑到不同节点之间相对重要性的问题.

因此针对以上算法存在的不足,本文提出新的图嵌入模型,在算法中增加对高阶相似节点和拓扑结构特征信息的关注,从而得到更好的节点嵌入表示.

2 基础知识

本节介绍论文中所涉及的所有变量以及相关定义.

G 为图, V 是图的节点集, E 是图的边集, N 表示图的节点数量, \mathbf{X} 表示特征矩阵, \mathbf{A} 为图的邻接矩阵, \mathbf{C} 是图的相似度矩阵, α 是相似度阈值, f_{ij} 表示内容相关性, s_{ij} 表示结构相关性, H 表示图嵌入.

定义 1 图.图可以表示为 $G=(V,E)$,其中 V 为图的节点集合, E 为边集.图共有 N 个节点, C 个节点类别, $|E|$ 条边.使用 $A \in \{0,1\}^{N \times N}$ 表示图的邻接矩阵, $\mathbf{X} \in \mathbf{R}^{N \times D}$ 为输入节点特征矩阵, N_i 表示节点 i 及其一跳邻居节点.

定义 2 相似网络.本文使用节点相似性度量指标计算节点与其邻居节点的相似性,计算方式如下:

$$c_{i,j} = \gamma(i,j), \quad (1)$$

其中, $c_{i,j}$ 表示节点 i 与节点 j 之间的相似度, $c_{i,j}$ 值越大表明节点 i 与节点 j 之间越相似, $\gamma(i,j)$ 为相似度计算函数.在矩阵形式中最终得到相似度矩阵 \mathbf{C} ,其中 $C(i,j)$ 为节点 i 与节点 j 之间的相似度,即 $c_{i,j}$.然后计算相似网络的邻接矩阵,此处超参数 α 为相似度阈值,用来限制相似节点的连接,计算方式如下:

$$A' = \begin{cases} 1, & c_{i,j} \geq \alpha, \\ 0, & \text{其他.} \end{cases} \quad (2)$$

当 $A'_{i,j} = 1$ 时在节点 i 和节点 j 之间构建边 $e_{i,j}$,否则不构建边,最终会形成新的相似网络 G' .

定义 3 图嵌入.给定图 $G=(V,E)$,通过无监督或有监督学习将图中每个节点 v_i 映射到低维空间,并保留原始信息.其任务旨在学习一个映射函数 $f: v_i \rightarrow h_i \in \mathbf{R}^d$,其中 v_i 表示第 i 个节点, $h_i \in \mathbf{R}^d$ 表示节点 i 的低维向量表示.

3 算法框架

本算法首先构建相似网络,其次分别计算相似网络中节点之间的内容相关性和结构相关性,两者结合得到联合注意力分数.最后使用联合注意力分数进行特征加权聚合,得到节点的嵌入表示.算法框架图如图 1 所示.

本算法主要由 3 部分构成:1)相似网络构建模块:首先使用节点相似性度量指标计算网络中节点之间的相似度.根据相似度阈值 α 进行筛选,对相似度较高且未连接的节点对构建新的边,从而使节点与其高阶相似节点建立起联系,最终形成相似网络.此模块解决了原始图注意力网络只关注一阶邻居节点的问题,丰富了聚合时的节点特征信息.2)联合注意力分数计算模块:若计算节点 i 和节点 j 之间的联合注意力分数,需分别计算节点对的内容相关性和结构相关性.从内容上讲,节点内容特征将用来计算节点对的内容相关性;从结构上讲,节点结构特征被用来计算结构相关性.节点内容相关性使用 GAT 中的注意力分数计算方法获得.通过设计一个自适应距离计算函数来计算节点的结构相关性,最后联合节点内容相关性和结构相关性计算得到联合注意力分数.3)图注意力机制模块:使用图注意力机制进行特征提取,在特征聚合过程中根据联合注意力分数对节点特征进行加权聚合,最终得到嵌入表示.

3.1 相似网络构建模块

本模块旨在使高阶相似节点参与到学习过程中.本文使用 4 种节点相似性度量来计算得到网络的相似度矩阵 $\mathbf{C}^{N \times N}$,其中 N 为节点的个数, $\mathbf{C}^{N \times N}$ 中每一个位置对应于原始网络中两个节点之间的相似度,后根据相似度阈值 α 进行筛选相似节点对,若节点对相似度高于相似度阈值则构建边,最终形成相似网络 G' .

4 种有代表性的节点相似性度量指标,分别是 common neighbors(CN)^[23],jaccard index(Jaccard)^[24],

adamic-adar(AA)^[25], hub depressed index(HDI)^[26].下面将分别介绍这 4 种相似性计算方法.

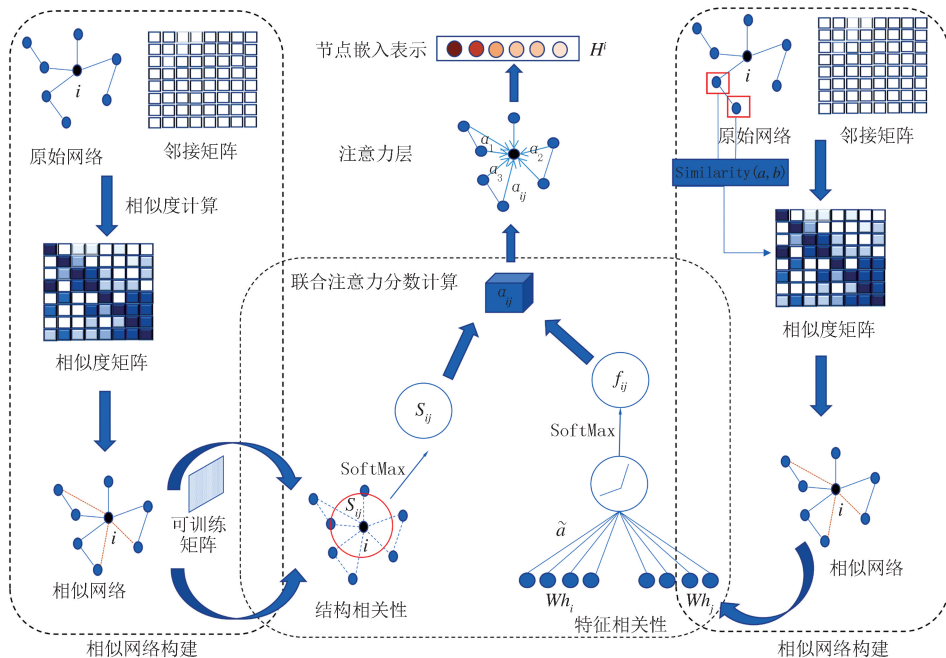


图1 算法整体框架图

Fig.1 Lustrates the comprehensive framework of the algorithm

(1)Common Neighbors(CN):给定节点 $u \in V, \Gamma(u) \subseteq V$ 为节点 u 的邻居节点集合.节点 $u \subseteq V$ 和节点 $v \subseteq V$ 的共同邻域定义如下:

$$c_{u,v} = |\Gamma(u) \cap \Gamma(v)| = |\{\omega \in V \mid (v, \omega) \wedge (u, \omega) \in \epsilon\}|. \quad (3)$$

在矩阵形式中,相似度矩阵可以被表述为:

$$C = \hat{A}^2. \quad (4)$$

(2)Jaccard Index:该指标通过将交集的大小与并集的大小归一化来评估两个节点的邻居之间的重叠情况:

$$c_{u,v} = \frac{|\Gamma(u) \cap \Gamma(v)|}{|\Gamma(u) \cup \Gamma(v)|}. \quad (5)$$

在矩阵形式中,相似度矩阵可以通过以下方式求得

$$C = \hat{A}^2 \cdot (\hat{A}N + M\hat{A} - \hat{A}^2), \quad (6)$$

其中, N 表示与 A 相同大小的全一矩阵, \cdot 表示矩阵点除操作.

(3)Adamic-Adar(AA):该方法通过给连接较少的共同邻居分配更多的权重来完善共同邻居的概念:

$$c_{u,v} = \sum_{w \in |\Gamma(u) \cap \Gamma(v)|} \frac{1}{\log_2(|\Gamma(w)|)}. \quad (7)$$

在矩阵形式中,相似度矩阵可由以下方式求得:

$$C = \hat{A} \log_2(\hat{D}^{-1}) \hat{A}. \quad (8)$$

(4)Hub Depressed Index(HDI):与 Jaccard 指标类似,HDI 的目的是根据节点的度归一化两个节点的邻域的重叠部分,为了关注度数较高的节点:

$$c_{u,v} = \frac{|\Gamma(u) \cap \Gamma(v)|}{\max\{|\Gamma(u) \cup \Gamma(v)|\}}. \quad (9)$$

基于 HDI 构建的相似度矩阵,可以表述为:

$$C = \hat{A}^2 \cdot \max\{\hat{A}N, N\hat{A}\}. \quad (10)$$

3.2 基于联合注意力分数的特征加权聚合模块

本节主要介绍在构建相似网络后,算法中的联合注意力分数计算和图注意力机制模块的应用.

首先,利用 GAT 中的注意力分数计算方法得到相似网络 G' 中所有节点对的内容相关性.计算如下所示:

$$f_{ij} = \frac{\exp(\text{LeakyRelu}(\bar{\alpha}^T (\mathbf{W}\mathbf{h}_i \parallel \mathbf{W}\mathbf{h}_j)))}{\sum_{k \in N_i} \exp(\text{LeakyRelu}(\bar{\alpha}^T (\mathbf{W}\mathbf{h}_i \parallel \mathbf{W}\mathbf{h}_k)))}, \quad (11)$$

其中, $\mathbf{W} \in \mathbf{R}^{F' \times F}$ 是网络中所有节点共享的可训练的权重矩阵, $\bar{\alpha}$ 为单层前馈神经网络的参数向量, \parallel 表示串联函数, F 为节点的初始特征维度, \mathbf{h}_i 和 \mathbf{h}_j 分别表示节点 i 和节点 j 的初始特征向量, N_i 表示节点 i 的一阶邻居节点集合. 通过此过程则能得到节点与其一阶邻居节点之间的特征相关性.

然后,寻求获得节点 i 和节点 j 之间的结构相关性.在矩阵形式中图的结构相关性可用下述方法获得:

$$\mathbf{S}_{ij} = \arg \min \Phi(\mathbf{M}_{ij}, \mathbf{A}_{ij}), \quad (12)$$

其中, \mathbf{S} 为图的结构相关性矩阵, \mathbf{S}_{ij} 表示节点 i 和节点 j 之间的结构相关性. $\Phi(\cdot)$ 为距离函数, \mathbf{A} 为输入图的邻接矩阵, \mathbf{M} 为与 \mathbf{A} 相同纬度的、所有节点可共享的自适应可训练矩阵, 本文中距离函数选择欧氏距离, 则式(12)可以写为:

$$\mathbf{S}_{ij} = \arg \min (\mathbf{A}_{ij} - \mathbf{M}_{ij})^2, \quad (13)$$

之后,归一化结构相关性

$$s_{ij} = \frac{\exp(s_{ij})}{\sum_{k \in N_i} \exp(s_{ik})}. \quad (14)$$

然后将两者结合计算最终的联合注意力分数

$$a_{ij} = \frac{\gamma(\mathcal{F}_{ij})\mathcal{F}_{ij} + \beta(\bar{s}_{ij})\bar{s}_{ij}}{\gamma(\mathcal{F}_{ij}) + \beta(\bar{s}_{ij})}, \quad (15)$$

其中, $\gamma(\cdot)$ 和 $\beta(\cdot)$ 是转换函数, 用于调整特征相关性和结构相关性. 在本文中使用 Sigmoid 作为转换函数, 有:

$$\gamma(\mathcal{F}_{ij}) = \frac{1}{1 + e^{-\mathcal{F}_{ij}}}, \quad (16)$$

$$\beta(\bar{s}_{ij}) = \frac{1}{1 + e^{-\bar{s}_{ij}}}. \quad (17)$$

在获得联合注意力分数后,执行特征加权聚合用来更新每个节点的特征,并传播到下一层或被用于后续学习任务的最终表征:

$$\mathbf{h}_i^{(l+1)} = \sigma \left(\sum_{j \in N_i} \alpha_{ij} \mathbf{W} \mathbf{h}_j^{(l)} \right), \quad (18)$$

其中, $\sigma(\cdot)$ 为激活函数, α_{ij} 为节点 i 和节点 j 之间的联合注意力分数, $\mathbf{h}_j^{(l)}$ 为节点 j 在第 l 层的向量表示.

具体的算法如算法 1 所示.

算法 1 基于相似网络和联合注意力的图嵌入模型 (SiCAT).

输入 图 $G=(V, E)$, 特征矩阵 \mathbf{X} , 邻接矩阵 \mathbf{A} .

输出 嵌入矩阵 \mathbf{H}

- ① 根据指定的节点相似性度量指标计算节点相似性,得到图的相似度矩阵 \mathbf{S} ,
- ② 根据式(2)计算相似网络邻接矩阵,构建相似网络 G' ,
- ③ 根据式(11)和(13)分别计算节点内容相关性和结构相关性 s_{ij} ,
- ④ 根据式(16)和(17)使用转换函数调整内容相关性和结构相关性,根据式(15)计算最终注意力分数 α_{ij} ,
- ⑤ 根据式(18)进行节点特征加权聚合,得到节点 i 的向量嵌入表示 \mathbf{h}_i ,
- ⑥ 利用随机梯度迭代更新权重,直到收敛到局部最优或达到训练次数上限,
- ⑦ end for.

4 实验分析

本节首先介绍了实验所使用的数据集以及相关实验设置,后使用多个基准数据集对提出的模型进行实证分析,并多个图嵌入学习算法进行对比,对实验结果进行分析,验证本算法的优越性.

4.1 数据集及实验配置

本文使用 3 个经典基准数据集(Cora、Citeseer 和 Pubmed)^[27]进行实验,数据集统计信息如表 1 所示,以上 3 个数据集被广泛应用于各种图嵌入算法评估实验,它们均属于引文网络,其中节点表示论文,连边表示论文之间的引用关系,特征表示论文的属性信息,如作者、年份、研究主题等.本实验的下游学习任务为节点分类,最终通过使用节点分类的准确率评估所有算法的有效性.

实验中,首先分别使用 CA、Jaccard、AA、HDI 4 个相似性度量指标计算节点相似性,初始相似度阈值 α 为 0.本文方法的网络结构遵循 GAT 算法结构设置,采用两层消息传递层和多头注意力机制.第一层,8 个注意力头中的每一个注意力头都学习一个转换矩阵 $W \in \mathbf{R}^{d \times 8}$;第二层,在来自第一层 8 个注意力头产生的级联特征上使用转换矩阵 $W \in \mathbf{R}^{64 \times C}$ (C 为节点的标签数),采用一个注意头后跟一个 softmax 算子.

使用 Adam 优化器来学习参数模型,初始学习率 r 为 0.005,衰减系数为 0.000 5.为了防止模型过度拟合,在实验中引入提前停止策略.模型的输入维度为节点的初始特征维度,隐藏层嵌入向量维数为 8.迭代轮数 e 设置为 1 000,对于每个数据集,所有方法运行 10 次以获得稳定的统计数据,实验结果取平均值记录.

为了验证算法的有效性,本文将模型与具有代表性的图嵌入算法在节点分类问题上进行对比,包括 Deepwalk^[3]、GCN^[11]、GraphSAGE^[14]、GAT^[4]、ADSF^[19]、superGAT^[20]和 GOAT^[21]等.

4.2 实验结果

本文算法及上述基线在节点分类任务上的实验结果如表 2 所示,粗体表示 SiCAT 获得了比其他基线更好的表现.本文提出的基于相似网络的联合注意力图嵌入模型根据所选用的相似性计算方法不同有 4 个变体,分别是 SiCAT-CN、SiCAT-Jac、SiCAT-HDI 和 SiCAT-AA.本文所提算法对节点分类的提升效果较为显著,通过分析数据集节点平均度的变化可以得到解释,原始网络与相似网络节点平均度对比如图 2 所示.

表 2 不同数据集上各算法节点分类任务实验结果

Tab.2 Experimental results of classification task for each algorithm node on different datasets

算法	Cora 数据集	Citeseer 数据集	Pubmed 数据集	算法	Cora 数据集	Citeseer 数据集	Pubmed 数据集
Deepwalk	67.20%	43.20%	65.30%	GOAT	84.90%	69.50%	81.40%
GCN	74.25%	63.36%	77.83%	SiCAT-Jac	84.20%	74.40%	82.10%
GraphSAGE	81.12%	71.06%	79.04%	SiCAT-AA	85.70%	74.30%	84.10%
GAT	83.00%	70.36%	81.50%	SiCAT-CN	84.90%	74.30%	82.60%
ADSF	84.00%	73.50%	81.20%	SiCAT-HDI	83.80%	72.20%	81.10%
SuperGAT	83.70%	72.60%	81.70%				

根据表 2 可见,SiCAT-Jac、SiCAT-AA、SiCAT-CN 3 种算法在 3 个数据集集中的表现全部优于所有基线方法,其中 SiCAT-AA 算法综合结果最优,与原始 GAT 相比分别提高了 2.70%、3.94% 和 2.60%,因此在后续的分析研究中均以 SiCAT-AA 为例进行研究分析.

Cora、Citeseer 和 Pubmed 3 个数据集原始网络中节点的平均度分别为 3.9、2.8 和 4.5.数据集集中的节点的平均度较低,即每个节点连接的邻居节点数量较少,从而导致在特征聚合过程中每个节点可利用的邻居节点特征信息较少.引入相似网络后 3 个数据集中节点平均度分别增加 4.4、3.4 和 5.9,有效提高了节点的平均度.每个节点的一阶邻居节点数量增多,使模型可

表 1 数据集统计信息

Tab.1 Dataset statistical information

数据集	节点数	边数	特征数	类别数
Cora ^[27]	2 708	5 429	1 433	7
Citeseer ^[27]	3 312	4 714	3 703	6
Pubmed ^[27]	19 717	44 338	500	3

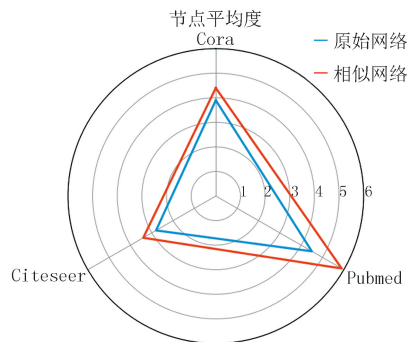


图2 原始网络与相似网络中节点平均度对比图

Fig.2 Comparison of average degree of nodes in original and similar networks

以充分地利用更多的节点特征信息.

4.3 可视化

本小节进行可视化任务.以 SiCAT-AA 算法为例, Cora、Citeseer 和 Pubmed 3 个数据集上的节点利用本算法得到节点嵌入向量,将嵌入向量作为 TSNE(T-distributed stochastic neighbor embedding)的输入,进行降维转化为二维向量表示.同一类节点用相同的颜色进行表示,可视化结果如图 3 表示.可见在 3 个数据集中属于同一类的节点大多能分配到一个簇中,体现了本文所提算法的有效性.

4.4 对比实验

为了进一步验证所提模型的有效性,设置本文所提算法与 GAT 对照实验.对比实验结果选用分类准确率 acc、损失函数与准确率收敛趋势演变进行对比.在图 4 中绘制了 GAT 和 SiCAT 的节点分类准确率和损失函数的演变,可以看出随着 epoch 的增加, SiCAT 的准确率 accuracy 和损失函数 loss 都逐渐收敛到稳定的区域并达到更优值,且与 GAT 相比收敛速度更快且收敛效果更好.



图3 本文算法在节点分类任务上的结果可视化展示图

Fig.3 Visualization of the results of this paper's algorithm on the node classification task

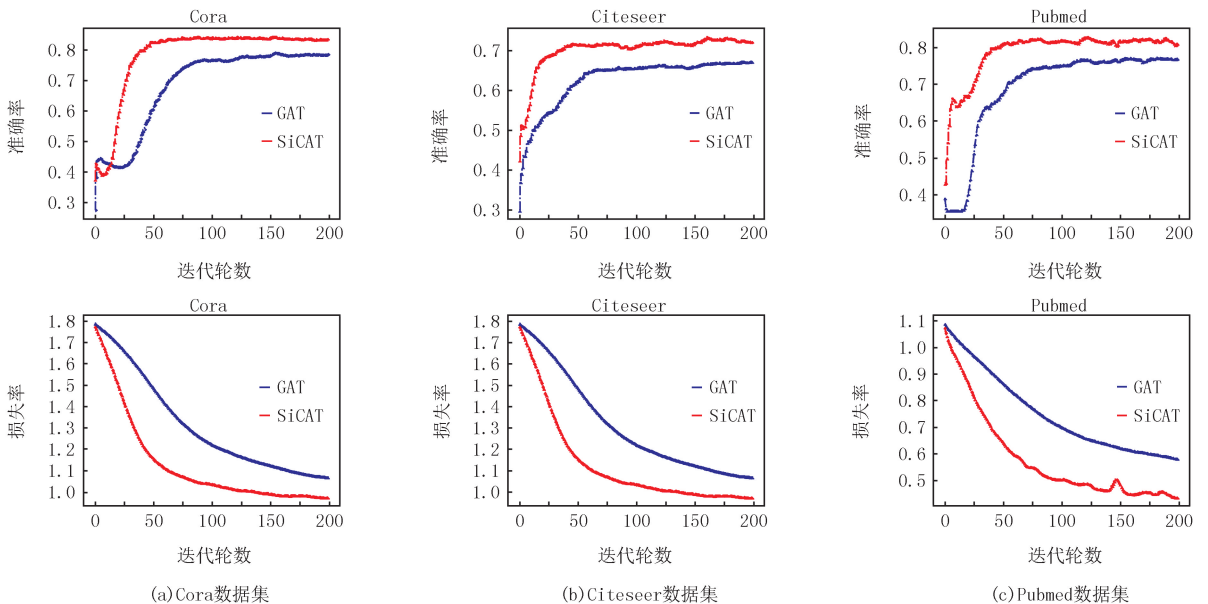


图4 本文所提算法和GAT方法准确率和损失函数的收敛情况

Fig.4 Convergence of the accuracy and loss of the proposed method and GAT method

4.5 超参数分析

本节研究超参数相似度阈值 α 对实验结果的影响,引入相似度阈值 α 旨在过滤掉相似度值较低的节点对.在图注意力网络中,增加邻居节点的数量不一定能够有效提取节点的结构特征,反而可能会产生噪声节点,因此对于相似节点边的构建需要进行一定的限制.本文使用相似度阈值解决这一问题,设置初值为 0,不断增加其值,观察分析实验结果.在 Cora、Citeseer 和 Pubmed 3 个数据集上节点分类准确率跟随相似度阈值的变化趋势分别如图 5 所示.

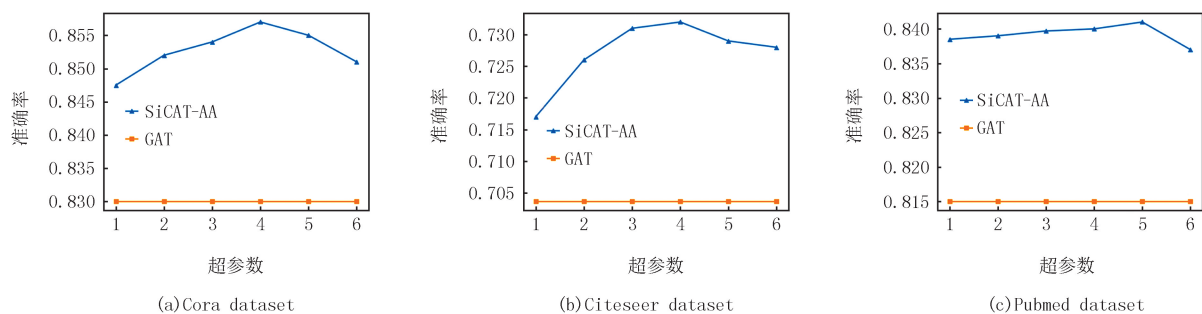


图5 Cora、Citeseer和Pubmed数据集在不同超参数下的分类准确率

Fig.5 The classification accuracy of Cora, Citeseer and Pubmed under different hyperparameters

对于 Cora 和 Citeseer 数据集,当相似度阈值为 4 时,节点分类准确率达到最高;对于 Pubmed 数据集,当相似度阈值为 5 时,节点分类准确率达到最高。可以看出,对于不同的数据集,相似度阈值的选取一般也不同。对于出现此结果的原因,本文分析认为 Pubmed 数据集节点数远远大于 Cora 和 Citeseer 数据集,且原始网络节点的平均度也是最大,因此原始网络上的节点拥有更多的一阶邻居节点可以参与到特征聚合的过程中,对构建新的一阶邻居节点的需求较小。

5 总 结

本文提出一种新的图嵌入算法,考虑了在利用图注意力网络时如何利用节点的高阶相似节点和节点结构特征以获得更好的嵌入表示。本文算法一方面利用高阶相似节点补充特征聚合中可利用的节点信息,解决了实际网络数据集中节点度较小、只使用节点一阶邻居信息而造成特征聚合不足的问题。另一方面解决了注意力分数计算过程中对节点结构特征考虑不足的问题。通过以上方法获得更加全面准确的注意力分数,最终得到高质量的嵌入表示。此外,本文所提算法在 3 个基准数据集上的节点分类任务的实验结果均好于基准,验证了模型的有效性,能够学习到更好的嵌入向量。未来,计划研究表示学习中的图数据增强问题,解决图神经网络模型中普遍存在的过渡平滑、非鲁棒性等问题。

参 考 文 献

- [1] TONG N, TANG Y, CHEN B, et al. Representation learning using Attention Network and CNN for Heterogeneous networks[J]. Expert Systems with Applications, 2021, 185: 115628.
- [2] CHAMI I, YING R, RÉ C, et al. Hyperbolic graph convolutional neural networks[J]. Advances in Neural Information Processing Systems, 2019, 32: 4869-4880.
- [3] PEROZZI B, AL-FOU R, SKIENA S. DeepWalk: online learning of social representations[C]// Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. [s.l.]: ACM, 2014.
- [4] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph attention networks[EB/OL]. [2023-05-13]. <http://arxiv.org/abs/1710.10903>
- [5] GAO H C, HUANG H. Deep attributed network embedding[C]// Twenty-Seventh International Joint Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2018.
- [6] LI Q M, HAN Z C, WU X M. Deeper insights into graph convolutional networks for semi-supervised learning[C]// Proceedings of the AAAI conference on artificial intelligence. Palo Alto: AAAI Press, 2018.
- [7] BELKIN M, NIYOGI P. Laplacian eigenmaps and spectral techniques for embedding and clustering[M]. Cambridge: The MIT Press, 2002.
- [8] HE X F, NIYOGI P. Locality preserving projections[C]// Proceedings of the 16th International Conference on Neural Information Processing Systems. [s.l.]: ACM, 2003.
- [9] GROVER A, LESKOVEC J. node2vec: scalable feature learning for networks[J]. KDD: Proceedings International Conference on Knowledge Discovery & Data Mining, 2016, 2016: 855-864.
- [10] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space[J]. ArXiv e-Prints, 2013. arXiv: 1301.3781.
- [11] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks[J]. ArXiv e-Prints, 2016. arXiv: 1609.02907.
- [12] BRUNA J, ZAREMBA W, SZLAM A, et al. Spectral networks and locally connected networks on graphs[J]. ArXiv e-Prints, 2013. arXiv:

1312.6203.

- [13] WU F, SOUZA A, ZHANG T, et al. Simplifying graph convolutional networks[C]//International conference on machine learning. Lille: PMLR, 2019.
- [14] 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. [s.l.]: ACM, 2017.
- [15] DUVENAUD D, MACLAURIN D, AGUILERA-IPARRAGUIRRE J, et al. Convolutional networks on graphs for learning molecular fingerprints[C]//Proceedings of the 28th International Conference on Neural Information Processing Systems-Volume 2. [s.l.]: ACM, 2015.
- [16] LUONG T, PHAM H, MANNING C D. Effective approaches to attention-based neural machine translation[J]. arXiv preprint, 2015. arXiv:150804025.
- [17] XU K, BA J L, KIROUS R, et al. Show, attend and tell: neural image caption generation with visual attention[C]//Proceedings of the 32nd International Conference on International Conference on Machine Learning-Volume 37. [s.l.]: ACM, 2015.
- [18] HE T, ONG Y S, BAI L. Learning conjoint attentions for graph neural nets[J]. Advances in Neural Information Processing Systems, 2021, 34:2641-53.
- [19] ZHANG K, ZHU Y, WANG J, et al. Adaptive structural fingerprints for graph attention networks. [EB/OL]. [2023-05-09]. <https://openreview.net/forum?id=BJxWx0NYPr>.
- [20] KIM D, OH A. How to find your friendly neighborhood: graph attention design with self-supervision [EB/OL]. [2023-05-12]. <http://arxiv.org/abs/2204.04879>
- [21] CHATZIANASTASIS M, LUTZEYER J, DASOULAS G, et al. Graph ordering attention networks[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2023, 37(6):7006-7014.
- [22] 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. [s.l.]: ACM, 2021.
- [23] NEWMAN M E. Clustering and preferential attachment in growing networks[J]. Physical Review E, Statistical, Nonlinear, and Soft Matter Physics, 2001, 64(2 Pt 2):025102.
- [24] JACCARD P. The distribution of the flora in the alpine zone. 1[J]. New Phytologist, 1912, 11(2):37-50.
- [25] LIBEN-NOWELL D, KLEINBERG J. The link prediction problem for social networks[C]//Proceedings of the twelfth international conference on Information and knowledge management. [s.l.]: ACM, 2003.
- [26] ZHOU T, LYU L Y, ZHANG Y C. Predicting missing links via local information[J]. The European Physical Journal B, 2009, 71(4):623-630.
- [27] SEN P, NAMATA G, BILGIC M, et al. Collective classification in network data[J]. AI Magazine, 2008, 29(3):93-106.

A graph embedding model based on similar networks and joint attention

Wang Jinghong^{1a,b,c}, Li Changxin^{1a}, Yang Jiateng², Yu Fuqiang^{1a}

(1. a. College of Computer and Cyber Security; b. Hebei Key Laboratory of Network and Information Security; c. Hebei Provincial Engineering Research Center for Supply Chain Big Data Analytics & Security, Hebei Normal University, Shijiazhuang 050024, China; 2. Artificial Intelligence and Big Data, Hebei Polytechnic Institute, Shijiazhuang 050020, China)

Abstract: The Graph Attention Network (GAT) incorporates the attention mechanism into graph neural networks. However, the model only considers the first-order neighborhood nodes of nodes, neglecting the consideration of higher-order similar nodes, and fails to account for the structural features of nodes when calculating the attention score. To address this issue, this paper propose a graph embedding model based on higher-order similar nodes and joint attention. Specifically, our approach first computes node similarities in the network and subsequently constructs new edges between pairs of nodes that are highly similar but not directly connected, thus forming a similar network. Secondly, we introduce the notions of structural relevance and content relevance to respectively characterize structural relationships and content features among nodes. Finally, we perform weighted aggregation of the node features using joint attention scores to obtain the final node embedding representations. The accuracy improvement over traditional models was found to be 2.70%, 3.94%, and 2.60%, respectively. These results demonstrate that the proposed method yields a better representation of node embedding.

Keywords: graph embedding; graph attention network; node similarity; similar network; node classification