计算机科学
COMPUTER SCIENCE

基于社交网络图节点度的神经网络个性化传播算法研究

邵云飞, 宋友, 王宝会

引用本文

邵云飞, 宋友, 王宝会.基于社交网络图节点度的神经网络个性化传播算法研究[J]. 计算机科学, 2023, 50(4): 16-21.

SHAO Yunfei, SONG You, WANG Baohui. Study on Degree of Node Based Personalized Propagation of Neural Predictions forSocial Networks [J]. Computer Science, 2023, 50(4): 16-21.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.220300274

基于社交网络图节点度的神经网络个性化传播算法研究

邵云飞　宋　友　王宝会
北京航空航天大学软件学院　北京 100191
(shaoyunfei_1823@126.com)

摘　要　图是一种重要且基础的数据结构,存在于各种各样的实际场景中。而随着近年来互联网的高速发展,社交网络图数据大量增加,对这些数据进行分析对公共服务、广告营销等实际场景有重要作用。目前已经有不少的图神经网络算法在此类问题中取得了较好的结果,但依然有提升的空间,在很多追求高准确度的场景下,工程师依然希望有性能更好的算法可供选择。文中对神经网个性化传播算法进行了改进,提出了新的可用于社交图网络的图神经网络算法 DPPNP。相比于传统图神经网络算法,在信息于节点之间传播时,该算法会根据节点的度对不同节点按不同比例保留自身信息,以提高准确度。在真实数据集上的实验结果表明,与已有的图神经网络算法相比,该算法拥有更好的性能。

关键词:图结构数据;图神经网络;图卷积神经网络;节点分类
中图法分类号　TP301

Study on Degree of Node Based Personalized Propagation of Neural Predictions for Social Networks

SHAO Yunfei,SONG You and WANG Baohui
School of Software,Beihang University,Beijing 100191,China

Abstract　Graph is an important and fundamental data structure that presents in a wide variety of practical scenarios. With the rapid development of the Internet in recent years,there has been a huge increase in social network graph data,and the analysis of this data can be of great help in practical scenarios such as public services and advertising and marketing. There are already quite a few graph neural network algorithms that can get good results in such problems,but they still have room for improvement,and in many scenarios where high accuracy is pursued,engineers still want to have algorithms with better performance to choose from. This paper improves personalized propagation of neural predictions and proposes a new graph neural network algorithm called degree of node based personalized propagation of neural predictions(DPPNP) that can be used in social graph networks. Compared to traditional graph neural network algorithms,when the information is propagated between nodes,the proposed algorithm will keep its own information for different nodes in different proportions according to the degree of nodes,so as to improve the accuracy. Experiments on real datasets show that the proposed algorithm has better performance compared to previous graph neural network algorithms.

Keywords　Graph structure data,Graph neural networks,Graph convolutional neural network,Node classification

1　引言

图作为图论的基本研究对象,是一种重要的数据表示形式,存在于各种各样的实际场景中。例如,社交媒体网络中的社交图以及电子商务领域中的用户兴趣图、知识图等。对这些图进行分析可以挖掘出隐藏在图中的信息。在过去的几十年中,对该领域的研究受到了极大的关注[1]。有效的图分析可以支持很多其他领域的工作,如节点分类[2]、节点聚类[3]、节点检索/推荐[4]、链接预测[5]等。

在图数据的应用中,有一种十分重要且自然的场景,即社交网络中的图数据分析。例如,通过分析在社交网络中的用户交互(如微博中的转发/评论/关注),我们可以对用户进行分类、检测社区、推荐朋友,并预测两个用户之间是否会发生交互。在这类图中,人或用户被当作节点,而人与人或用户与用户之间的关联被当作边,这样的社交网络图可以在一切人类活动与社会关系当中找到。而随着互联网的高速发展,社交网络图被赋予了新的意义,一些网站中的用户作为节点,用户之间的关系作为边,形成了大量社交网络图数据。对这些数据进行分析对公共服务、广告营销等实际场景有很大帮助。

近年来,基于图数据结构的图神经网络算法得到了长足的发展,这类算法大多会将节点本身的特征与邻近节点的特征结合起来,使得其在图数据结构上相比传统的只考虑节点特征的神经网络算法有着更好的效果。基于图数据结构的图神经网络算法的蓬勃发展不仅促进了基于图的数据挖掘与机器学习等领域的发展,拓展了有关图数据结构的研究广度,也使得在很多实际场景中出现图数据结构时有了更多更可靠

到稿日期:2022-03-30　返修日期:2022-09-10
通信作者:王宝会(wangbh@buaa.edu.cn)

的工具。很多图神经网络算法都可以直接或间接完成节点分类、节点聚类和链接预测等具体任务，其效果往往远超传统方法。然而对于不同的使用场景，不同具体算法的效率和准确度都有所差别，对于社交网络图这一使用场景，虽然现有算法已经可以完成其中的很多具体任务，但其在效率和准确度等方面都依然有提升空间。无论是对已有算法进行改进，还是探索新的算法，都有重大意义。

神经网络个性化传播[6]（Personalized Propagation of Neural Predictions，PPNP）算法是一种性能较优的图神经网络算法。它将个性化网页排名（Personalized PageRank）算法与图神经网络相结合，让节点信息沿图中的边传播，并且在传播过程中平衡了每个节点自身的信息与来自整张图的信息的比例，在数据实验上取得了不错的效果。然而该算法在每一次信息传递中，对于图中每一个节点，均以一个固定的比例保留节点自身信息，而在实际的图数据特别是社交网络图中，每个节点所适合的保留节点自身信息的比例不一定是相同的。因此，本文在 PPNP 算法的基础上提出了一种新的可用于社交网络图的图神经网络算法（Degree of Node Based Personalized Propagation of Neural Predictions，DPPNP），根据不同节点自身的情况，保留不同的自身信息的比例，以得到更好的效果。

本文的主要创新点如下：

（1）提出了一种在信息传播过程中可以对不同节点按不同比例保留自身信息的可用于社交网络图的图神经网络算法。

（2）探究如何设计保留自身信息的比例，使算法获得更好的性能。

（3）在真实数据集上进行实验，验证了算法具有良好的性能。

本文第 2 节总结了现有的主流的图神经网络算法；第 3 节给出了本文提出的图神经网络算法 DPPNP；第 4 节给出了在真实数据集上的实验结果；最后总结全文并展望未来。

2 相关工作

近年来，图神经网络领域蓬勃发展。最早的图神经网络算法是 Franco 等[7]于 2007 年提出的同名算法图神经网络算法（Graph Neural Network，GNN）。该算法受到了循环神经网络的启发，并将其思路应用在图数据结构上，以得到节点的特征表达。在该算法中，每个节点会重复聚合邻节点的信息以更新自身的特征表达，直至稳定。作为图神经网络的鼻祖，它对之后的各类图神经网络算法都有重要的启发作用。同时，该算法也属于图循环神经网络，不少同类算法沿用了该算法的思路，只是改变了每个节点从邻节点聚合信息时的具体方式。

Thomas 等[8]于 2017 年提出了图卷积网络算法（Graph Convolutional Networks，GCN）。顾名思义，该算法受到了卷积神经网络的启发，将其思路应用在图数据结构上，以得到节点的特征表达。同时它也继承了 GNN 的基本思路：每个节点重复从邻节点聚合信息以更新自身的特征表达。与 GNN 不同的是，这类算法不再使用迭代收缩的方法来使每个节点的特征表达达到稳定状态，而是使用特定数量的卷积层来得到节点特征表达。大量算法都使用了类似的思路，只需要改变其中卷积层的设计，就能提升算法性能，得到新的算法。Johannes 等[6]于 2019 年提出的 PPNP 算法同样也使用了类似的思路，取得了很好的效果。

Wang 等在 2016 年提出的 SDNE[9]算法使用了完全不同的思路。SDNE 并不考虑节点自身的特征，只考虑图的结构，它使用神经网络框架将图中节点在低维向量空间中表示，因此它也属于图嵌入算法。而 Jain 等在 2015 年提出的 Structural-RNN 算法[10]则针对动态图设计了图神经网络算法。近年来国内也有不少针对特定场景、特定需求的基于图神经网络的研究，例如在 2021 年，Wu 等提出了针对异质信息网络的节点分类算法 MHGCN[11]，Wei 等基于图神经网络提出了活动推荐算法 DGNN-AR[12]，Qian 等提出了胶囊图神经网络 GLCapsGNN[13]；2022 年，Huang 等提出了针对会话推荐问题的图神经推荐方法 FA-GNR[14]，Shu 等将图神经网络应用于多跳阅读理解问题[15]。上述算法丰富了图神经网络算法的广度，使得图神经网络算法可以在更多场景中使用。

3 针对社交网络图的图神经网络算法

3.1 文中术语符号

接下来介绍本文使用到的术语及符号的含义，便于后续对算法及实验内容的介绍。术语和符号的名称及含义对照关系如表 1 所列。

表 1 常用符号说明

Table 1 Explanation of common symbols

符号	说明
$G=(V,E)$	图，其中 V 和 E 分别表示图中节点与边的集合
v_i	图 G 中的节点，$v_i\in V$
$deg(v_i)$	节点 v_i 的度
e_{ij}	图 G 中连接节点 v_i 和 v_j 的边，$e_{ij}\in E$
$\boldsymbol{A}$	图 G 的邻接矩阵
$\widetilde{\boldsymbol{A}}$	带自环的邻接矩阵，满足 $\widetilde{\boldsymbol{A}}=\boldsymbol{A}+\boldsymbol{I}$
$\widetilde{\boldsymbol{D}}$	$\widetilde{\boldsymbol{A}}$ 的度矩阵，满足 $\widetilde{\boldsymbol{D}}=diag(deg(v_1),deg(v_1),\cdots,deg(v_{\lvert V\rvert}))$
$\widetilde{\widetilde{\boldsymbol{A}}}$	带自环的标准化邻接矩阵，满足 $\widetilde{\widetilde{\boldsymbol{A}}}=\widetilde{\boldsymbol{D}}^{-\frac{1}{2}}\widetilde{\boldsymbol{A}}\widetilde{\boldsymbol{D}}^{-\frac{1}{2}}\boldsymbol{A}$
f	特征向量维数
$\boldsymbol{X}\in R^{\lvert V\rvert\times f}$	图的特征矩阵，其中每行 $\boldsymbol{x}_i$ 分别是节点 v_i 的特征
$\boldsymbol{H}$	图神经网络的隐藏层
$\boldsymbol{Y}$	图神经网络的输出层
softmax(·)	softmax 函数，对于一行中每一个元素有：$\mathrm{softmax}(x_i)=\frac{e^{x_i}}{\sum_i e^{x_i}}$

3.2 图卷积网络算法

图卷积网络[8]算法，即 GCN 算法，它使用特定数量的卷积层来得到节点特征表达，其中每一层中每个节点都从邻节点聚合信息以更新自身的特征表达。它使用如下公式更新节点的特征表达：

$$\boldsymbol{H}^{(l+1)}=\sigma(\widetilde{\widetilde{\boldsymbol{A}}}\boldsymbol{H}^{(l)}\boldsymbol{W}^{(l)})(l=0,1,\cdots,m-1) \tag{1}$$

其中，$\boldsymbol{H}^{(0)}=\boldsymbol{X}$，$\boldsymbol{W}^{(l)}$ 是需要学习的权重矩阵，σ 是非线性激活函数，m 是卷积层层数。最后一层在进行最终输出前将激活函数 σ 替换为 softmax 函数。

3.3 神经网络个性化传播算法

神经网络个性化传播算法[6]，即 PPNP 算法，它受到了 PageRank 算法[16]的启发，将其中的思路运用到 GCN 上。

PageRank 算法能够使用网页之间的链接关系，计算出网页的重要性得分。节点 v 的重要性得分 $PR(v)$ 表达式如下：

$$PR(v)=\mu\frac{1}{|V|}+(1-\mu)\sum_{(u,v)\in E}\frac{PR(u)}{d_{\text{out}}(u)} \tag{2}$$

其中，μ 表示不根据网页之间链接进行跳转的概率。引入这个变量的原因是其中存在一些只有出度没有入度，或是反之的节点，这会导致网页的重要性被稀释或无限制地积累。将所有节点的重要性得分转化为向量形式表示 $\boldsymbol{w}$ 则得到：

$$\begin{aligned}&\boldsymbol{w}=(1-\mu)\boldsymbol{A}\boldsymbol{w}+\mu\boldsymbol{e}\\&\boldsymbol{w}=\mu(\boldsymbol{I}-(1-\mu)\boldsymbol{A})^{-1}\boldsymbol{e}\end{aligned} \tag{3}$$

其中，$\boldsymbol{A}$ 为邻接矩阵，$\boldsymbol{e}$ 为所有元素均为 $\frac{1}{|V|}$ 的向量。如果将其中的 $\boldsymbol{e}$ 换成代表用户偏好的向量 $\boldsymbol{v}$，将最终的结果导向与用户偏好相关的网页，那么就能得到 Personalized PageRank 算法[16]。

PPNP 算法将以上思路应用在 GCN 中。将用户偏好向量 $\boldsymbol{v}$ 替换为与图特征矩阵 $\boldsymbol{X}$ 相关的矩阵 $\boldsymbol{H}$，将邻接矩阵 $\boldsymbol{A}$ 替换为带自环的标准化邻接矩阵 $\tilde{\boldsymbol{A}}$，再带入 GCN 算法中，就得到了 PPNP 算法：

$$\begin{aligned}&\boldsymbol{Z}_{\text{PPNP}}=\text{softmax}(\alpha(\boldsymbol{I}-(1-\alpha)\tilde{\boldsymbol{A}})^{-1}\boldsymbol{H})\\&\boldsymbol{H}=f_\theta(\boldsymbol{X})\end{aligned} \tag{4}$$

其中传送概率参数 α 为超参数，它代替了 PageRank 算法中的 μ；$f_\theta(\cdot)$ 表示一个神经网络。相比 GCN 算法，该算法直接模拟特征的完整传播过程，从而计算出矩阵，不再需要多层卷积层来模拟多次传播，并且传播过程中有概率跳回初始节点，从而平衡了不同大小的邻域特征。

3.4 基于节点度的神经网络个性化传播算法

在 PPNP 算法中，引入了传送概率参数 α，该参数表示在信息传播过程中使用初始信息的概率，从而平衡了不同大小的邻域特征对算法所得到的节点隐藏特征 $\boldsymbol{H}$ 的作用，具体算法如式(4)所示。然而在该算法中，α 是一个全局标量，这意味着对不同节点 v_i，使用初始信息的概率 α_{v_i} 都是相同的，即都为 α。但在社交网络图中，节点代表的用户之间常常有很大差异，不同用户适合不同大小的使用初始信息的概率 α_{v_i}。如果使用不同的初始概率，我们能从节点中提取的信息，除了节点特征以外，还有图中节点的邻近结构。其中对节点邻近结构最直接的体现就是节点的度，故我们应考虑对于节点 v_i，令 α_{v_i} 为节点 v_i 的度的函数 $g(deg(v_i))$。于是我们将 PPNP 算法中的 α 替换成一个矩阵 $\boldsymbol{B}$，以得到基于节点度的神经网络个性化传播算法(DPPNP)。

$$\begin{aligned}&\boldsymbol{Z}_{\text{DPPNP}}=\text{softmax}(\boldsymbol{B}(\boldsymbol{I}-(\boldsymbol{I}-\boldsymbol{B})\tilde{\boldsymbol{A}})^{-1}\boldsymbol{H})\\&\boldsymbol{H}=f_\theta(\boldsymbol{X})\\&b_{ij}=\begin{cases}g(deg(v_i)), & i=j\\0, & i\neq j\end{cases}\end{aligned} \tag{5}$$

其中，$f_\theta(\cdot)$ 表示神经网络，$g(\cdot)$ 表示函数，我们称之为传送概率矩阵函数；b_{ij} 为矩阵 $\boldsymbol{B}$ 中第 i 行第 j 列的元素。算法示意图如图 1 所示。这样，对于不同节点而言，其 α_{v_i} 就与节点的度有关，而且其具体关系可以通过不同的函数 $g(\cdot)$ 来调整，算法变得更加灵活。

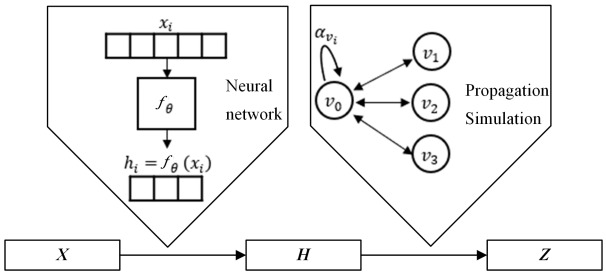


图 1 DPPNP 算法示意图

Fig. 1 Illustration of DPPNP

3.5 传送概率矩阵函数

在 DPPNP 算法中，传送概率矩阵函数 $g(\cdot)$ 决定了每个节点使用初始信息的概率，故选择合适的 $g(\cdot)$ 对最终算法的性能至关重要。经过实验，我们选用以下传送概率矩阵函数：

$$g(x)=\frac{\alpha_1-\alpha_\infty}{\ln(x)+1}+\alpha_\infty,x\in N_+ \tag{6}$$

其中，α_1 和 α_∞ 为超参数，满足 $\alpha_1,\alpha_\infty\in[0,1]$。当 $\alpha_1<\alpha_\infty$ 时，$g(x)$ 单调递增且值域为 $[\alpha_1,\alpha_\infty)$；当 $\alpha_1>\alpha_\infty$ 时，$g(x)$ 单调递减且值域为 $(\alpha_\infty,\alpha_1]$；当 $\alpha_1=\alpha_\infty=\alpha$ 时，有 $g(x)\equiv\alpha$。由于 $g(x)$ 中会先对 x 取对数，这意味着当 x 有数量级上的变化时，$g(x)$ 才会有较为明显的变化，该函数的图像如图 2 所示。

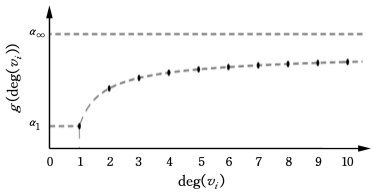


图 2 传送概率矩阵函数图像

Fig. 2 Image of propagation probability matrix function

在 DPPNP 算法中，$x=\deg(v_i)$，因此两个超参数 α_1 和 α_∞ 分别表示节点度为 1 和度趋于无穷时其 α_{v_i} 的值，即有：

$$\alpha_1=\alpha_{v_i}\mid_{\deg(v_i)=1},\alpha_\infty=\lim_{\deg(v_i)\to+\infty}\alpha_{v_i} \tag{7}$$

在社交网络图中，节点的度代表着用户的社交关系数量，若两个节点 v_i 和 v_j 的度拥有相同的数量级，则它们对应的 α_{v_i} 也会比较接近。特别地，因为当 $\alpha_1=\alpha_\infty=\alpha$ 时，有 $g(x)\equiv\alpha$，所以 $\boldsymbol{B}=\alpha\boldsymbol{I}$，此时 DPPNP 算法退化为 PPNP 算法。

4 实验与分析

4.1 数据集介绍

本文的实验使用 3 个公开数据集，每一个数据集的任务都是节点分类。3 个公开数据集的统计信息如表 2 所列，而它们的详细介绍如下：

表 2 数据集信息表

Table 2 Dataset information

Datasets	Nodes	Edges	Classification
Facebook	22470	171002	4
GitHub	37700	289003	2
LastFM Asia	7624	27806	18

(1)Facebook 数据集[17]：社交网站 Facebook 的社交网络

图。其中每一个节点对应着一个 Facebook 官方机构页面，每一条边对应的是一组相应机构之间互相标记喜欢。节点特征提取自其官方机构页面，在分类任务中的目标类型则是在页面中其官方机构的类型，包含政府、政治家、公司、TV 秀共 4 种。

(2)GitHub 数据集[17]：代码托管平台 GitHub 的社交网络图。其中每一个节点对应一个至少有 10 个代码仓库的开发者，每一条边对应一组相应开发者之间的互相关注。节点特征提取自开发者的位置、代码仓库、所属公司以及电子邮件地址，在分类任务中的目标是区分每一个开发者是 web 领域的开发者还是机器学习领域的开发者。

(3)LastFM Asia 数据集[18]：在线音乐流媒体网站 LastFM 中的亚洲用户组成的社交网络图。其中每一个节点对应一名网站的亚洲用户，每一条边对应一组相应用户之间的互相关注。节点特征提取自用户喜欢的音乐人，在分类任务中的目标是区分每一个用户的国籍。

4.2 算法有效性实验

我们首先验证利用 DPPNP 算法中基于节点的度来对传送概率进行优化的有效性。我们对比了当两个超参数α_1和α_∞取多种不同的组合时，DPPNP 算法在以上 3 个数据集上进行节点分类任务的准确率。将训练集、验证集、测试集的比例设定为 6∶2∶2，对于每个数据集中每一组超参数α_1和α_∞的取值组合，我们均使用相同的数据集划分，重复 20 次实验。在 Facebook，GitHub，LastFM Asia 数据集上的实验结果如表 3—表 5 所列，其可视化表示如图 3 所示。

表 3 DPPNP 算法在 Facebook 数据集上的节点分类准确率

Table 3 Accuracy of node classification of DPPNP on Facebook dataset

α_∞	α_1					
	0.05	0.10	0.20	0.30	0.50	0.75
0.05	0.939±0.001	0.940±0.002	0.942±0.001	0.941±0.002	0.941±0.001	0.938±0.002
0.10	0.941±0.001	0.942±0.001	**0.943**±0.001	0.942±0.001	0.941±0.001	0.938±0.001
0.20	0.942±0.001	0.941±0.001	0.942±0.001	0.942±0.001	0.940±0.001	0.937±0.001
0.30	0.942±0.002	0.942±0.001	0.942±0.001	0.941±0.001	0.940±0.001	0.936±0.001
0.50	0.940±0.001	0.940±0.001	0.940±0.001	0.939±0.002	0.937±0.001	0.933±0.001
0.75	0.936±0.001	0.936±0.002	0.935±0.001	0.934±0.001	0.932±0.001	0.927±0.002

表 4 DPPNP 算法在 GitHub 数据集上的节点分类准确率

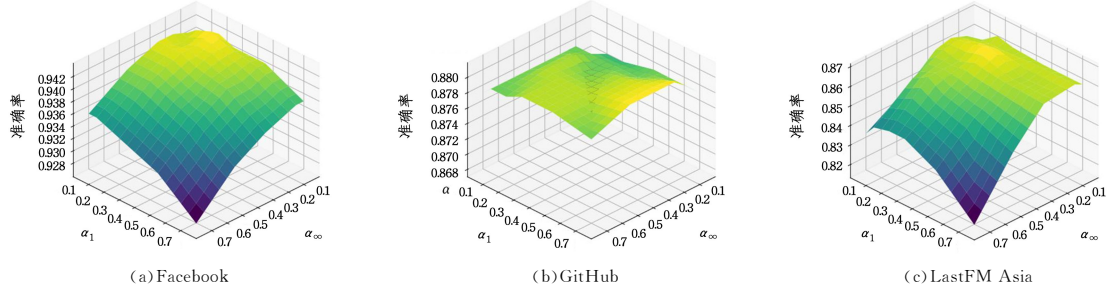
Table 4 Accuracy of node classification of DPPNP on GitHub dataset

α_∞	α_1					
	0.05	0.10	0.20	0.30	0.50	0.75
0.05	0.868±0.001	0.872±0.001	0.876±0.002	0.876±0.002	0.878±0.001	0.878±0.001
0.10	0.874±0.002	0.876±0.001	0.878±0.002	0.878±0.001	0.879±0.002	0.879±0.002
0.20	0.878±0.001	0.879±0.002	0.879±0.001	0.879±0.001	0.879±0.001	**0.881**±0.001
0.30	0.878±0.002	0.879±0.002	0.880±0.001	0.879±0.001	0.880±0.0010	**0.881**±0.001
0.5	0.878±0.001	0.879±0.002	0.880±0.001	0.880±0.001	0.880±0.001	**0.881**±0.002
0.75	0.878±0.002	0.879±0.001	0.879±0.001	0.880±0.002	0.880±0.002	0.879±0.001

表 5 DPPNP 算法在 LastFM Asia 数据集上的节点分类准确率

Table 5 Accuracy of node classification of DPPNP on LastFM Asia dataset

α_∞	α_1					
	0.05	0.10	0.20	0.30	0.50	0.75
0.05	0.852±0.002	0.858±0.002	0.864±0.003	0.862±0.002	0.861±0.002	0.861±0.002
0.10	0.860±0.002	0.861±0.002	0.862±0.003	0.862±0.002	0.863±0.002	0.863±0.003
0.20	0.863±0.002	0.864±0.002	0.866±0.002	**0.868**±0.002	0.861±0.003	0.862±0.002
0.30	0.861±0.002	0.863±0.002	0.861±0.002	0.864±0.002	0.862±0.002	0.861±0.002
0.50	0.848±0.002	0.856±0.002	0.861±0.003	0.856±0.002	0.852±0.003	0.844±0.003
0.75	0.836±0.002	0.841±0.002	0.843±0.003	0.843±0.002	0.835±0.002	0.817±0.002



(a)Facebook (b)GitHub (c)LastFM Asia

图 3 DPPNP 算法节点分类准确率

Fig. 3 Accuracy of node classification of DPPNP

由于当$\alpha_1=\alpha_\infty=\alpha$时，DPPNP 算法会退化为 PPNP 算法，表 3—表 5 中的对角线单元格上表示的即是当超参数α取不同值时，PPNP 算法在这 3 个数据集上的准确率。因此从以上实验结果中能比较容易得出 DPPNP 算法与 PPNP 算法的性能差别。

在 PPNP 算法中，超参数α设置为 0.1，但在这 3 个数据

集上，使用其他的超参数设定会得到更好的效果：在 Facebook 数据集上，超参数 α 取值为 0.1 和 0.2 时结果最好，准确率最高为 0.942；在 GitHub 数据集上，超参数 α 取值为 0.5 时结果最好，准确率均为 0.880，同时 α 在 0.2～0.75 的范围内均能达到不错的准确率；在 LastFM Asia 数据集上，超参数 α 取值为 0.2 时结果最好，准确率为 0.866，同时 α 在 0.2～0.3 的范围内均能达到不错的准确率。而 DPPNP 算法能找到更好的超参数 α_1 和 α_∞ 的组合，在这 3 个数据集上均能得到比 PPNP 算法更高的准确率。在 Facebook 数据集上，超参数为 $\alpha_1=0.2$，$\alpha_\infty=0.1$ 时结果最好，准确率为 0.943，而 0.2 和 0.1 也恰好是该数据集上 PPNP 算法效果最好时的超参数 α 的取值；在 GitHub 数据集上，超参数设定为 $\alpha_1=0.75$，$\alpha_\infty=0.2$，0.3 或 0.5 时结果最好，准确率为 0.881，而 0.2～0.75 恰好是该数据集上 PPNP 算法效果较好时的超参数 α 的取值范围；在 LastFM Asia 数据集上，超参数设定为 $\alpha_1=0.3$，$\alpha_\infty=0.2$ 时结果最好，准确率为 0.868，而 0.2～0.3 恰好是该数据集上 PPNP 算法效果较好时超参数 α 的取值范围。可以看出，在选取合适的超参数组合时，DPPNP 算法可以达到比 PPNP 算法更好的效果，这说明 DPPNP 算法在 PPNP 算法基础上进行的传送概率优化是有效的。特别地，DPPNP 算法和 PPNP 算法的超参数取值存在关联性，在一个数据集上，如果首先找到 PPNP 算法中较好的超参数 α 的取值范围，在这个范围中选取适当数值分别设置为 DPPNP 算法中超参数 α_1 和 α_∞ 的值，且使 $\alpha_1>\alpha_\infty$，那么就很可能得到比 PPNP 算法更好的结果。

4.3 不同算法对比实验

接下来，在以上 3 个数据集上，将本文提出的 DPPNP 算法与多种算法进行比较，以说明 DPPNP 算法相比已有的多种图嵌入与图神经网络算法在社交网络图上的性能提升。参与对比实验的算法包括 DeepWalk[19]，Node2Vec[20]，LINE[21] 这 3 种图嵌入算法以及 GCN[7]，PPNP[6]，MixHop[22]，ADA-UGNN[23] 这 4 种图神经网络算法。对于图嵌入算法，我们提取其中的图结构信息，将节点转化为向量表达，然后对其使用逻辑回归算法进行节点分类；对于图神经网络算法，我们均令训练集、验证集、测试集的比例为 6∶2∶2，且其数据集划分方式与 4.2 节实验相同。同时，对于 DPPNP 算法和 PPNP 算法，其超参数均使用 4.2 节实验中效果最好的组合。以上 8 种算法在 3 个实验数据集上的准确率如表 6 所列。

表 6 不同算法社交网络图节点分类准确率

Table 6 Accuracy of node classification of different algorithms on social networks

Algorithms	Facebook	GitHub	LastFM Asia
DPPNP	**0.943±0.001**	**0.881±0.001**	**0.868±0.002**
PPNP	0.942±0.001	0.880±0.001	0.866±0.002
GCN	0.932±0.001	0.864±0.001	0.862±0.002
MixHop	0.940±0.002	0.862±0.001	0.861±0.004
ADA-UGNN	0.934±0.004	0.868±0.003	0.826±0.006
DeepWalk	0.759±0.004	0.843±0.002	0.762±0.004
Node2Vec	0.824±0.003	0.843±0.002	0.783±0.004
LINE	0.624±0.006	0.805±0.003	0.619±0.005

从实验结果可以看出，对于有节点特征的数据集来说，图嵌入算法由于不提取节点特征，其效果整体上不如提取节点特征的图神经网络算法。而在参与对比的 4 种图神经网络算法中，DPPNP 算法在以上 3 个数据集上的准确率均超过了现有其他算法。可以看出，DPPNP 算法在社交图网络上有着较为优秀的表现。

结束语 本文提出了一种新的图神经网络算法 DPPNP，该算法主要针对社交图网络数据集，可用于节点分类等任务。该算法在 PPNP 算法的基础上进行了改进，引入了传送概率矩阵和传送概率矩阵函数，针对不同度的节点调整了信息聚合方式，使得算法更加灵活，性能也更加优秀，这种优化算法的思路可以给其他研究者带来一定的启发。本文在真实数据集上进行了实验，结果表明，在社交网络图节点分类问题上，DPPNP 算法相比其他图神经网络和图嵌入算法有着更高的准确度。

在接下来的工作中，我们将研究其他传送概率矩阵函数在 DPPNP 算法中的效果，另外也会考虑如何将 DPPNP 算法中的优化思路与其他已有的图神经网络结合起来，以达到更好的效果。

参 考 文 献

[1] CAI H, ZHENG V W, CHANG K C C. A comprehensive survey of graph embedding: Problems, techniques, and applications[J]. IEEE Transactions on Knowledge and Data Engineering, 2018, 30(9): 1616-1637.

[2] WANG X, CUI P, WANG J, et al. Community preserving network embedding[C]// AAAI. 2017: 203-209.

[3] NIE F, ZHU W, LI X. Unsupervised large graph embedding [C]// Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. 2017: 2422-2428.

[4] ZHOU C, LIU Y, LIU X, et al. Scalable graph embedding for asymmetric proximity [C] // Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. 2017: 2942-2948.

[5] WEI X, XU L, CAO B, et al. Cross view link prediction by learning noise-resilient representation consensus[C] // Proceedings of the 26th International Conference on World Wide Web. 2017: 1611-1619.

[6] JOHANNES K, ALEKSANDAR B, STEPHAN G. Predict then Propagate: Graph Neural Networks meet Personalized PageRank [C] // International Conference on Learning Representations (ICLR). 2019.

[7] FRANCO S, MARCO G, AH CHUNG T, et al. The Graph Neural Network Model [J]. IEEE Transactions on Neural Networks, 2009, 20(1): 61-80.

[8] THOMAS N K, MAX W. Semi-Supervised Classification with Graph Convolutional Networks[J]. arXiv: 1609.02907, 2016.

[9] WANG D, CUI P, ZHU W. Structural deep network embedding [C] // Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016: 1225-1234.

[10] JAIN A, ZAMIR A R, SAVARESE S, et al. Structural-rnn: Deep learning on spatio-temporal graphs[C] // Proceedings of

the IEEE Conference on Computer Vision and Pattern Recognition. 2016:5308-5317.

[11] WU Y,WANG Y,WANG X,et al. Motif-Based Hypergraph Convolution Network for Semi-Supervised Node Classification on Heterogeneous Graph[J]. Chinese Journal of Computers, 2021,44(11):2248-2260.

[12] WEI X H,SUN B Y,CUI J X. Recommending activity to users via deep graph neural network[J]. Journal of Jilin University (Engineering and Technology Edition),2021,51(1):278-284.

[13] QIAN R,ZHANG R,ZHANG K J,et al. Capsule graph neural network based on global and local features fusion[J]. Journal of Jilin University(Engineering and Technology Edition), 2021, 51(3):1048-1054.

[14] HUANG Z H,LIN X L,SUN S L,et al. Feature Augmentation based Graph Neural Recommendation Method in Session Scenarios[J]. Chinese Journal of Computers,2022,45(4):766-780.

[15] SHU C,OUYANG Z,DU N,et al. Research on Multi-Hop Reading Comprehension Based on Graph Neural Network with Improved Graph Nodes [J]. Computer Engineering, 2022, 48(1):99-104.

[16] PAGE L,BRIN S,MOTWANI R,et al. The PageRank Citation Ranking: Bringing Order to the Web[R]. Stanford InfoLab, 1999.

[17] BENEDEK R,CARL A,RIK S. Multi-Scale attributed node embedding[J]. Journal of Complex Networks, 2021, 9(2): cnab014.

[18] ROZEMBERCZKI B,SARKAR R. Characteristic functions on graphs:Birds of a feather,from statistical descriptors to parametric models[C]// Proceedings of the 29th ACM International Conference on Information & Knowledge Management. 2020: 1325-1334.

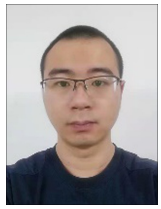
[19] PEROZZI B,AL-RFOU R,SKIENA S. Deepwalk:Online learning of social representations[C]// Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2014:701-710.

[20] 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.

[21] TANG J,QU M,WANG M,et al. Line:Large-scale information network embedding[C]// Proceedings of the 24th International Conference on World Wide Web. 2015:1067-1077.

[22] ABU-EL-HAIJA S,PEROZZI B,KAPOOR A,et al. MixHop: Higher-Order Graph Convolution Architectures via Sparsified Neighborhood Mixing[C]// International Conference on Machine Learning. PMLR,2019:21-29.

[23] MA Y,LIU X,ZHAO T,et al. A unified view on graph neural networks as graph signal denoising[C]// Proceedings of the 30th ACM International Conference on Information & Knowledge Management. 2021:1202-1211.



SHAO Yunfei, born in 1994, postgraduate. His main research interests include graph neural networks and graph embedding.

WANG Baohui, born in 1973, senior engineer, master supervisor. His main research interests include software architecture, big data, artificial intelligence, etc.

(责任编辑:何杨)