

# 基于相似度随机游走聚合的图节点分类算法

车翔玖, 孙雨鹏

(吉林大学 计算机科学与技术学院, 长春 130012)

**摘要:** 针对异质图节点分类任务中 MLP、GCN 等方法准确率相对较低的问题, 提出了一种基于相似度随机游走聚合的图神经网络 (SRW-GNN)。为降低异质性对节点嵌入的影响, SRW-GNN 利用节点间的相似度作为概率进行随机游走, 并将采样路径作为邻域以获取更多同质信息。为解决大多数现有 GNN 聚合器对节点顺序不敏感的问题, 本文引入基于循环神经网络 (RNN) 的路径聚合器来同时提取路径中节点的特征和顺序信息。此外, 节点对不同路径有不同的偏好, 为了自适应地学习不同路径在节点编码中的重要性, 采用注意力机制动态地调整各路径对最终嵌入的贡献。在多个常用的异质图数据集的实验结果表明, 本文方法的准确率明显优于 MLP、GCN、H2GCN、HOG-GCN 等方法, 验证了其在异质图节点分类任务中的有效性。

**关键词:** 计算机应用技术; 图神经网络; 异质图; 节点分类

**中图分类号:** TP391.4 **文献标志码:** A **文章编号:** 1671-5497(2025)06-2069-07

**DOI:** 10.13229/j.cnki.jdxbgxb20231043

## Graph node classification algorithm based on similarity random walk aggregation

CHE Xiang-jiu, SUN Yu-peng

(College of Computer Science and Technology, Jilin University, Changchun 130012, China)

**Abstract:** In addressing the issue of relatively low accuracy in node classification tasks on heterophily graphs using methods such as MLP and GCN, a Graph Neural Network based on Similarity Random Walk Aggregation (SRW-GNN) was proposed. To address the impact of heterophily on node embeddings, SRW-GNN employs the similarity between nodes as probabilities for conducting random walks. The sampled paths serve as the neighborhood, enabling the model to gather more homophily-based information. To address the issue of insensitivity to node order in most existing graph neural network (GNN) aggregators, a path aggregator based on recurrent neural networks (RNNs) was introduced to simultaneously extract features and order information of each node in the path. Furthermore, nodes exhibit varying preferences for different paths. To adaptively learn the importance of different paths in node encoding, an attention mechanism was employed to dynamically adjust the contributions of each path to the final embedding. Experimental results on several commonly used heterophily graph datasets demonstrate

**收稿日期:** 2023-10-01.

**基金项目:** 国家自然科学基金项目 (62172184); 吉林省科技发展计划项目 (20200401077GX, 20200201292JC).

**作者简介:** 车翔玖 (1969-), 男, 教授, 博士. 研究方向: 计算机图形学, 大数据可视化. E-mail: chexj@jlu.edu.cn

that the proposed SRW-GNN method achieves significantly higher accuracy compared to the methods such as MLP, GCN, H2GCN, HOG-GCN, validating its effectiveness in heterophily graph node classification tasks.

**Key words:** computer application technology; graph neural network; heterophily graph; node classification

## 0 引言

图作为现实中常见的数据结构,被广泛应用于诸多任务,包括社交网络、推荐系统等领域。图是一种非欧几何空间中的数据结构<sup>[1]</sup>,不要求所有节点之间都有边连接,因此无法直接将处理图像的卷积神经网络(Convolutional neural network, CNN)<sup>[2]</sup>应用在图结构中。

为了更有效地处理图数据结构, Gilmer 等<sup>[3]</sup>提出了信息传递神经网络(Message passing neural network, MPNN)。MPNN 可以将节点与它们邻居之间的信息传递和聚合,充分利用了图的拓扑结构,已经被广泛应用于节点分类、图分类和链路预测等与图数据相关的任务中。

MPNN 是依托于同质性假设的图卷积网络(Graph convolutional network, GCN)<sup>[4]</sup>,同质性是指大多数连接的节点来自同一类别,并且具有相似的属性。在具有强同质性的网络上,GCN 及其变体 GraphSAGE<sup>[5]</sup>、GAT<sup>[6]</sup>表现出卓越的性能。

由于图卷积网络(GCN)的信息传播机制是基于同质性假设,并未针对异质性网络进行设计,而异质图中多数连接的节点来自不同类别具有不相似的属性,GCN 可能会将不同类别的信息混合在一起,导致性能受到较大影响。

为了解决同质性限制,有些方法通过调整不同标签邻居之间的注意力权重改善信息聚合,如 CPGNN<sup>[7]</sup>、HOG-GCN<sup>[8]</sup>、BM-GCN<sup>[9]</sup>将邻域中具有相同标签的节点聚合在一起,降低不同标签邻域的聚合程度,以此聚合更多同质性信息,减少异质信息的干扰。然而,此类方法用于半监督学习,标签是预测值,无法准确判断哪些节点具有同种标签。

GGCN<sup>[10]</sup>、GNN-LF/HF<sup>[11]</sup>将注意力权重从 $[0, 1]$ 扩展到 $[-1, 1]$ ,对相同类别节点分配正值以提高同种类别间的信息传递,对不同类别节点分配负值以抑制不同类别间的信息传递。这种设计仅对类别数接近 2 的网络有效,对类别数较多的网络效果不佳。

还有一些方法,如 Geom-GCN<sup>[12]</sup>、H2GCN<sup>[13]</sup>和 GPR-GNN<sup>[14]</sup>,显式聚合高阶邻域节点信息。这些方法假设直接相邻的邻域节点在异质网络中可能以异质性为主导,而高阶邻域节点更有可能以同质性为主导,因此能够为目标节点提供更有价值的信息。然而,聚合过程中的求和算子仍然隐含着同质性假设,并未彻底改变 GCN 的基本聚合机制。

为了解决这一问题,RAW-GNN<sup>[15]</sup>引入了一种新的聚合机制,将随机游走的路径作为邻域进行聚合,通过广度优先和深度优先进行游走分别获取节点附近和更深感受野的信息。然而该类方法同样假设异质网络中目标节点附近的邻域可能以异质性为主导,而高阶邻域更有可能以同质性为主导,没有充分考虑图的异质性,只考虑图的结构性,忽略了节点的特征信息。

针对以上问题,为了更灵活地处理异质图数据,本文提出了基于相似度随机游走聚合的图神经网络,简称 SRW-GNN。不同于以上方法,该方法充分利用了图的结构信息以及图中节点的特征信息,进而提高了异质图中节点分类的准确率。

## 1 相关工作

### 1.1 符号表示

本文将一个无向无权有特征的图网络表示为  $G=(V, E, X)$ ,其中  $V=\{v_1, v_2, \dots, v_n\}$  为  $n$  个节点的集合,  $n$  为图中节点数,  $X \in \mathbf{R}^{n \times f}$  为节点特征的集合,  $f$  为特征维数。  $E=\{e_{ij}\} \subseteq V \times V$  为边的集合,由邻接矩阵  $A=[a_{ij}] \in \mathbf{R}^{n \times n}$  表示,其中如果节点  $v_i$  与  $v_j$  相连,则  $a_{ij}=1$ ,否则  $a_{ij}=0$ 。

本文关注节点分类任务。具体来说,每个节点都属于一个类  $c \in C$ ,  $C$  为所有种类的集合,  $|C|$  为类的数量,给定训练集  $T_V$  中具有已知标签  $y_v$  和特征向量  $x_v$  的节点  $v \in V$ ,目标是推断出所有  $v' \in (V - T_V)$  中的未知标签  $y_{v'}$ 。

### 1.2 MPNN 基本原理

从严格意义上讲,MPNN 不是一种具体的模

型,而是一种空域卷积的形式化框架。它将空域卷积分解为两个过程:消息传递和状态更新。

在MPNN的每一轮迭代中,每个节点都将自己的特征作为信息传递给其邻居节点,然后将来自邻居节点的信息进行聚合,以更新自身的表示。这个过程可以被视为信息在图中的传播,节点通过与其连接的其他节点交换信息,并不断更新自身的特征表示,继而完成节点分类、图分类等下游任务。

在MPNN中,邻域被定义为相距一跳或多跳的邻居节点。MPNN的第 $l$ 层定义如下:

$$\begin{cases} m_i^{(l)} = \text{aggregate}^{(l)}(\{g_i^{(l-1)}: i \in N_i\}) \\ g_i^{(l)} = \text{combine}^{(l)}(g_i^{(l-1)}, m_i^{(l)}) \end{cases} \quad (1)$$

式中: $g_i^{(l)}$ 为第 $l$ 层中节点 $i$ 的特征向量; $\text{aggregate}^{(l)}$ 为消息传递操作; $\text{combine}^{(l)}$ 为状态更新操作。

初始向量 $g_i^{(0)}$ 为 $x_i$ ,而 $N_i$ 为节点 $i$ 的邻居节点集合。不同的 $N_i$ , $\text{aggregate}(l)$ , $\text{combine}(l)$ 的选择就意味着不同的模型。通常,节点 $i$ 的邻居节点要么是 $i$ 的相邻节点,要么是 $i$ 的 $k$ 跳邻居节点。

### 1.3 异质图

在早期图神经网络模型中,关注较多的是同质图网络,如GCN、GraphSAGE、GAT等都在同质图上表现出较好的效果。这里的同质性指的是属于相同类别并且拥有相似属性的节点间更倾向于建立连接。

然而,现实世界中存在许多网络,它们不满足同质性假设。相反,这些网络通常表现出异质性或低同质性的特点,其中大多数相连接的节点可能属于不同的类别,并具有不相似的属性。

例如,在社交网络中,一个用户可能不仅与相似的用户相连接,还会与不同年龄、兴趣和职业的其他用户交流;在学术合作网络中,一位科学家可能不仅与同一领域的研究者连接,还会与不同领域的合作伙伴进行合作;在电信通信网络中,不同类型的设备(如手机、计算机、路由器)也可能相互连接,而这些设备具有不同的通信协议和特性。

### 1.4 基于路径的邻域

基于路径的邻域已经在异构图<sup>[16]</sup>中得到了应用。对其进行引申修改就可以用于异质图。具体而言,一个路径 $P$ 的定义采用有序列表的形式, $P = \{v_1, v_2, \dots, v_k\}$ 为长度为 $K$ 的路径,其中 $v_k \in V, k = 1, 2, \dots, K$ ,并且路径上相邻的节点之

间存在边。节点 $i$ 基于路径的邻居被表示为 $P_i$ ,多条以节点 $i$ 为起点游走的路径就构成了节点 $i$ 基于路径的邻域。

### 1.5 聚合后节点相似度

对于如何衡量图的同质性,已经存在许多不同的指标。通常,指标的数值越高表示同质性越强,而数值越低表示异质性越强。然而,Luan等<sup>[17]</sup>指出,现有的同质性度量公式在某些情况下可能会产生误导,根据之前的度量方法,如边同质性比率<sup>[12]</sup>,图1毋庸置疑是一个异质图,具有较强的异质性,相连接的节点都属于不同的类别,然而在均值聚合后类A和类B只是交换了颜色,不同类别的节点并没有变得不可区分。为了更好地衡量图的异质性,文献<sup>[17]</sup>提出了聚合后相似度,并定义了聚合后节点相似度矩阵:

$$S = AX(AX)^T \in R^{n \times n} \quad (2)$$

式中: $A \in R^{n \times n}$ 为一个常规的聚合器,本文直接使用邻接矩阵,也就是聚合节点的一阶邻居。

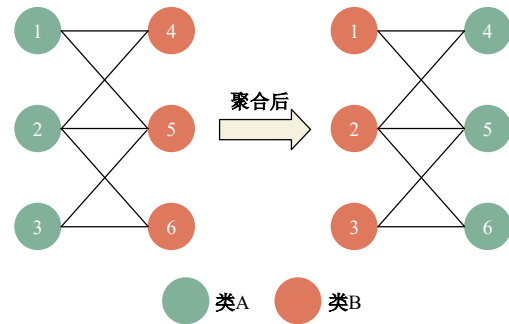


图1 同质还是异质

Fig. 1 Homophily or heterophily

## 2 本文方法

### 2.1 总体框架

为了突破同质性假设的限制,提升异质图节点的分类性能,本文提出了一种新的节点分类方法,包括以下主要组件:基于相似度随机游走生成器、基于RNN的路径聚合器、基于注意力机制的节点编码器以及节点分类器。本文模型示意图如图2所示。

在图神经网络中,通常通过邻域聚合来为目标节点获取有用信息。GCN通常将邻居节点作为邻域,然而,异质图中目标节点的邻居节点通常属于与目标节点不同的类别或拥有不相似的属性。一阶邻域通常具有强异质性,计算目标节点嵌入时会聚合太多异质性信息导致节点分类准确

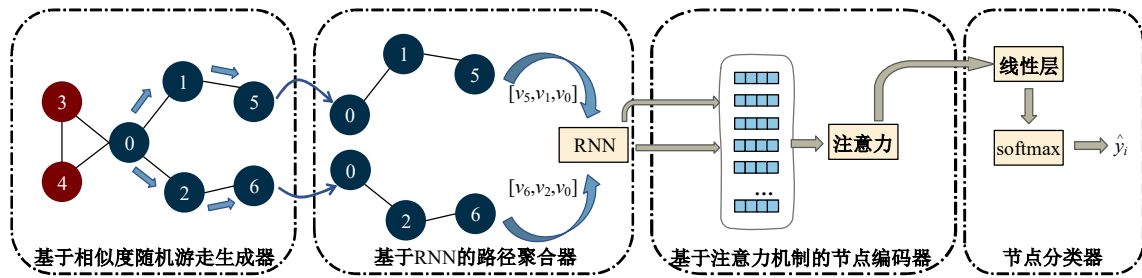


图 2 模型框架

Fig. 2 Model framework

率较低,即便是直接聚合高阶邻域,也会忽略一些同样重要的低阶邻域节点信息,而高阶邻域中同样存在异质信息干扰节点嵌入。于是本文并未采用直接邻居或者高阶邻居节点作为邻域,而是选择路径作为邻域,不局限于目标节点的某阶邻居,路径能记录节点的顺序信息,凭借采样距离获取到更多信息。

一个直接的获取路径的方法是枚举每个节点的所有路径,但在感受野范围内的路径数量呈指数级增长,特别是当中心节点被遍历时。因此,使用所有路径进行计算是不切实际的,尤其是在感受野较大,即采样距离较长时。

为了在搜索空间中以更少的路径获得更多信息,最简单的方法之一是采用随机游走算法。然而,随机游走算法只能通过深度优先搜索或者广度优先搜索来随机采样路径,只考虑了图的结构信息而忽略了节点的特征信息。因此,面临的首要问题是如何进行随机游走,在计算开销可控的情况下更好地获取图中的同质性信息。于是本文设计了基于相似度的随机游走生成器,根据节点间相似度确定下一跳节点,使节点在游走时更有可能跳转到属于相同类别、具有相似特征的节点,也就可以获取到更多的同质性信息。

采样路径之后面临的问题是如何对路径进行编码。在路径中,节点出现的顺序对捕获邻域内的结构信息尤为关键。然而,大多数现有的GNN聚合器对节点顺序不敏感,仅对节点的特征信息进行聚合,而不考虑路径中节点的顺序信息。为了有效地聚合路径信息,本文提出了一种基于循环神经网络(Recurrent neural network, RNN)的路径聚合器。这个聚合器用于处理由一个个节点组成的路径序列,能够充分考虑路径节点的顺序信息和特征信息。

现在,模型已经获取了每个节点通过随机游走生成的路径嵌入。多数现有的方法采用均值聚

合或者求和聚合计算目标节点的嵌入,然而这会忽略路径间的差异,考虑到节点对不同路径具有不同偏好,本文引入了注意力机制。该机制能够自适应地学习邻域中不同路径在节点嵌入中的重要性,允许模型动态地调整不同路径对最终嵌入的贡献,以更好地适应任务的需求。得到节点的最终嵌入后,使用一个常规的分类器对节点进行分类。

### 2.2 基于相似度的随机游走生成器

为了在计算开销可控的同时更好地获取图中的同质性信息,本文提出了基于相似度的随机游走生成器,并以此采样路径。

根据图中节点的特征和边的连接关系计算相似度矩阵  $S \in R^{N \times N}$ ,其中每一行表示每个节点与其他节点之间的相似度。采用基于相似度的随机游走算法计算有边连接的节点间的相似度,并作为游走概率,以此进行随机选择以确定路径的下一跳节点,从而生成一系列路径,这些路径有助于获取更多同质性信息。

考虑一个长度为  $K$  的采样路径,从节点  $v_i$  到节点  $v_j$ ,路径的生成概率可以表示为:

$$Pro_{ij}^K = s_{i_1, i_2} \cdots s_{i_{k-2}, i_{k-1}} s_{i_{k-1}, i_k} \quad (3)$$

式中:  $s_{ij}$  为相似度矩阵中的一个元素,表示节点  $v_i$  与  $v_j$  之间的相似度。

本文采用聚合后的相似度作为随机游走方法的度量。当两个节点的相似度高时,这意味着它们在邻域聚合后具有更相似的特征。这种相似度不仅考虑了节点自身的特征,还考虑了节点一阶邻居的特征。这使得在游走时更有可能跳转到属于相同类别的节点,从而有望提取更多同质性信息,并减少异质性对任务的影响。

### 2.3 基于RNN的路径聚合器

该聚合器的目的是对路径进行编码,对于给定的一条路径  $P_i$ ,路径聚合器旨在学习路径上所有节点的结构和语义信息。这不仅包括路径的起

点节点和终点节点,还包括路径中间的所有上下文节点。路径聚合器的任务是保留路径上节点之间的有序连接,同时考虑节点特征,以捕获更多的信息对路径进行编码。

具体而言,路径聚合器  $\text{Path}: \mathbf{R}^{K \times f} \rightarrow \mathbf{R}^{d_i}$  的定义为:

$$\mathbf{h}_P = \text{Path}(P) = f_{\theta}(\{\mathbf{x}_{n_1}, \mathbf{x}_{n_2}, \dots, \mathbf{x}_{n_K}\}) \quad (4)$$

式中:  $\mathbf{h}_P \in \mathbf{R}^{d_i}$  为路径  $P = \{v_{n_1}, v_{n_2}, \dots, v_{n_K}\}$  的嵌入向量,其中,  $d_i$  为径聚合器的输出维度;  $\text{Path}(P)$  表示将路径  $P$  输入到路径聚合器中,从而得到路径表示;  $\mathbf{x}_{n_k} \in \mathbf{R}^f$ ,  $k = 1, 2, \dots, K$  为节点  $v_{n_k}$  的初始特征;  $\theta$  为聚合器的所有可学习参数。

为了对路径进行编码,可以使用任何考虑元素顺序的编码器,例如循环神经网络(RNN)中的长短期记忆神经网络(Long short term memory, LSTM)<sup>[18]</sup>、门控循环单元(Gate recurrent unit, GRU)<sup>[19]</sup>等,本文使用了门控循环单元。

考虑到目标节点的特征对最终嵌入影响比较大,在将随机游走产生的路径序列输入路径聚合器之前进行翻转,让目标节点最后进入聚合器,以保留更多目标节点的初始特征。

#### 2.4 基于注意力机制的节点编码器

考虑到不同的路径可能对目标节点嵌入的贡献不同,本文引入了一种类似GAT中的注意力机制,用于学习节点  $i$  的路径邻域  $N_i$  中每个路径嵌入的不同权重,  $N_i$  中包含节点  $i$  基于路径的邻域中所有路径的嵌入。具体实现如下:

$$\begin{cases} \mathbf{e}_P = \text{LeakyReLU}(\mathbf{u}_P^{\top} \cdot \mathbf{h}_P) \\ \alpha_P = \frac{\exp(\mathbf{e}_P)}{\sum_{Y \in N_i} \exp(\mathbf{e}_Y)} \\ \mathbf{o}_i = \sigma \left( \sum_{P \in N_i} \alpha_P \cdot \mathbf{h}_P \right) \end{cases} \quad (5)$$

式中:  $\mathbf{u}_P \in \mathbf{R}^{d_i}$  为可学习的参数;  $\mathbf{e}_P$ 、 $\mathbf{e}_Y$  为分别路径  $P$ 、 $Y$  的非标准化注意力权重;  $\mathbf{h}_P$  为路径  $P$  的表示;  $\text{LeakyReLU}(\cdot)$  为激活函数;  $\mathbf{o}_i$  为节点  $i$  的最终表示;  $\alpha_P$  为通过 softmax 对  $N_i$  中所有路径进行归一化的路径权重。

在计算节点嵌入的时候不再考虑目标节点本身的特征  $\mathbf{x}_i$ , 因为  $N_i$  中的每条路径都已经包含了  $\mathbf{x}_i$  作为其结束节点的特征。为了进一步稳定学习过程,采用了多头注意力机制,在式(5)中使用了

$H$  个注意力头,然后将它们的嵌入拼接在一起,形式化如下:

$$\mathbf{o}_i = \parallel_{h=1}^H \sigma \left( \sum_{P \in N_i} \alpha_P^h \cdot \mathbf{h}_P \right) \quad (6)$$

式中:“ $\parallel$ ”代表向量拼接操作。

#### 2.5 节点分类器

本文关注的是节点分类任务,采用常用的交叉熵作为损失函数,使用一个线性层和 softmax 层计算预测的标签概率:

$$\hat{\mathbf{y}}_i = \text{softmax}(\mathbf{W} \cdot \mathbf{o}_i + \mathbf{b}) \quad (7)$$

式中:  $\mathbf{W}$  为可学习的权重矩阵;  $\mathbf{b}$  为偏移项;  $\hat{\mathbf{y}}_i$  中每个元素值代表节点属于对应类别的预测概率,其中对应值最大的元素的下标为最终预测的标签,  $\hat{\mathbf{y}}_i \in \mathbf{R}^{|\mathcal{C}|}$ 。

### 3 实验结果及分析

#### 3.1 数据集

为了展示 SRW-GNN 对异质图节点分类的效果,在 3 个公开的异质网络数据集上评估 SRW-GNN 和现有最先进方法的性能。这些数据集的特征如表 1 所示。表 1 中, Cornell、Texas 和 Wisconsin 是从相应大学计算机科学系收集的网页数据集<sup>[12]</sup>,其中节点代表网页,边代表超链接关系,节点特征是网页的词袋表示,节点标签是网页的类别(学生、项目、课程、工作人员和教职员工)。

表 1 数据集统计

参数	数据集		
	Cornell	Texas	Wisconsin
节点数/个	183	183	251
边数/个	295	309	499
特征维数/个	1 703	1 703	1 703
类别数量/个	5	5	5

#### 3.2 基线方法

将本文提出的 SRW-GNN 方法与以下基线方法进行比较:仅使用节点特征信息的多层感知机(Multilayer perceptron, MLP);基于同质性假设的 GNN 模型,如 GCN<sup>[3]</sup>、GAT<sup>[6]</sup>、GraphSAGE<sup>[4]</sup>;针对异质图的模型,如 H2GCN<sup>[13]</sup>、CP-GNN<sup>[7]</sup>、GPR-GNN<sup>[14]</sup>、BM-GCN<sup>[9]</sup>、HOG-GCN<sup>[8]</sup>、RAW-GNN<sup>[15]</sup>。某些方法有多个变种,选择每种方法的最佳结果进行比较。

### 3.3 超参数设置

为了保证公平性,本文采用了与 HOG-GCN<sup>[8]</sup> 相同的数据集划分方法,对所有数据集生成 10 个随机划分。在每个数据集中,48% 的节点用作训练集,32% 的节点用作验证集,其余用作测试集。所有基线方法的参数都设置为其作者使用值。对于 SRW-GNN,使用的采样路径长度为 4,每个节点采样了 12 条路径。对于基于循环神经网络(RNN)的聚合器,使用 128 个隐藏单元的 GRU,并将注意力头数设置为 2。采用了 Adam 优化器和 PyTorch 的默认初始化

方式,学习率设置为 0.05。将测试集中节点分类正确率作为准确率以比较不同模型的节点分类精度。

### 3.4 实验结果及分析

在 3 个公开数据集上进行实验,比较不同模型在节点分类任务上的性能,结果如表 2 所示。由表 2 可以看出:本文 SRW-GNN 方法的效果较基于同质性假设的图神经网络有明显提升,与现有的针对异质图设计的模型相比,准确率也有一定提升。实验结果证明了本文方法在异质图上的有效性。

表 2 实验结果

Table 2 Experiment results

方 法	数据集			
	Cornell	Texas	Wisconsin	
基线方法	MLP	84.32±6.14	83.24±5.77	86.47±3.33
	GCN	55.14±7.57	55.68±9.61	53.73±7.65
	GraphSAGE	78.38±6.84	85.41±5.16	85.49±3.53
	GAT	54.59±7.33	57.30±3.38	54.31±5.62
	H2GCN	78.92±5.24	82.16±8.21	82.57±3.21
	CPGNN	63.51±5.83	74.32±7.38	81.76±6.74
	GPR-GNN	82.97±5.68	84.59±4.37	83.92±3.14
	BM-GCN	79.14±8.44	85.13±4.64	82.82±8.89
	HOG-GCN	84.32±4.32	85.17±4.40	86.67±3.36
	RAW-GNN	84.86±5.43	85.95±4.15	88.24±3.72
消融实验	COS-SRW	83.78±5.67	83.78±4.36	87.06±2.35
	SRW-MLP	56.76±6.62	57.57±8.47	44.12±9.62
	SRW-MEAN	84.32±5.64	82.97±6.74	86.47±2.83
本文方法	SRW-GNN	86.49±6.40	86.76±4.43	88.63±3.70

本文 SRW-GNN 方法的主要优点是:首先,利用相似度进行随机游走采样,且计算相似度时不仅考虑了节点自身的特征,还考虑了节点一阶邻居的特征,使得路径邻域拥有更多的同质性信息。其次,使用循环神经网络对路径进行编码,同时考虑了路径中节点的顺序信息和特征信息。最后,每个节点使用注意力机制学习到更适合的路径权重,以得到最终的节点嵌入。

由于异质图的公开数据集目前还都比较小,本文也只验证了在 Cornell、Texas 和 Wisconsin 上的表现效果,下一步会考虑在更大的数据集上进行验证。

### 3.5 消融实验

为了验证本文所提出模块的有效性,本文进行了一系列消融实验,对使用的各个模块进行替换。实验包括采用节点间的余弦相似度代替本文

所提出的聚合后相似度(COS-SRW);将路径中节点特征拼接后使用多层感知机 MLP 代替本文的路径聚合器(SRW-MLP);对所有路径嵌入进行均值聚合代替基于注意力机制的节点编码器(SRW-MEAN)。

相关消融实验结果如表 2 所示,COS-SRW 仅考虑节点间特征余弦相似度,未充分考虑邻居特征,导致性能略微下降;SRW-MLP 没有考虑路径中每个节点的顺序,没有充分利用路径邻域的优势,性能显著下降;SRW-MEAN 没有考虑不同节点对路径的偏好,导致性能稍有下降。实验结果充分说明了本文所提出模块的有效性。

## 4 结束语

本文针对异质网络提出了一种基于相似度随机游走的图神经网络,用于提高异质图节点分类

任务的准确率。该方法首先基于聚合后相似度对全图进行随机游走,计算相似度时不仅考虑了目标节点特征,还考虑了其一阶邻居的特征。其次,将产生的路径输入基于RNN的路径聚合器得到路径嵌入,充分考虑了路径中每个节点的顺序和特征。最后,使用注意力机制自适应学习节点对不同路径的偏好,得到节点的嵌入。模型在3个公开数据集上进行了实验,与MLP、GCN以及现有的针对异质图的模型(如H2GCN)进行了比较。结果表明,本文模型对异质图节点分类任务有更高的准确率。

#### 参考文献:

- [1] 徐冰冰,岑科廷,黄俊杰,等. 图卷积神经网络综述[J]. 计算机学报, 2020, 43(5): 755-780.  
Xu Bing-bing, Cen Ke-yan, Huang Jun-jie, et al. A survey on graph convolutional neural network[J]. Chinese Journal of Computers, 2020, 43(5): 755-780.
- [2] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [3] Gilmer J, Schoenholz S S, Riley P F, et al. Neural message passing for quantum chemistry[C]//Proceedings of the International Conference on Machine Learning, Sydney, Australia, 2017: 1263-1272.
- [4] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks[J/OL]. [2023-08-11]. <https://arxiv.org/abs/1609.02907>
- [5] Hamilton W L, Ying R, Leskovec J. Inductive representation learning on large graphs[J/OL]. [2023-08-11]. <https://arxiv.org/abs/1706.02216>
- [6] Veličković P, Cucurull G, Casanova A, et al. Graph attention networks[J]. Stat, 2017, 1050(20): No. 48550.
- [7] Zhu J, Rossi R A, Rao A, et al. Graph neural networks with heterophily[J]. AAAI Technical Track on Machine Learning V, 2021, 35(12): 11168-11176.
- [8] Wang T, Jin D, Wang R, et al. Powerful graph convolutional networks with adaptive propagation mechanism for homophily and heterophily[C]//Proceedings of the AAAI Conference on Artificial Intelligence, Philadelphia, USA, 2022, 36(4): 4210-4218.
- [9] He D, Liang C, Liu H, et al. Block modeling-guided graph convolutional neural networks[C]//Proceedings of the AAAI Conference on Artificial Intelligence, Philadelphia, USA, 2022: 4022-4029.
- [10] Yan Y, Hashemi M, Swersky K, et al. Two sides of the same coin: heterophily and oversmoothing in graph convolutional neural networks[C]//2022 IEEE International Conference on Data Mining (ICDM), Chennai, India, 2022: 1287-1292.
- [11] Zhu M, Wang X, Shi C, et al. Interpreting and unifying graph neural networks with an optimization framework[C]//Proceedings of the Web Conference, New York, NY, USA, 2021: 1215-1226.
- [12] Pei H, Wei B, Chang K C C, et al. Geom-GCN: geometric graph convolutional networks[J/OL]. [2023-08-11]. <https://arxiv.org/abs/2002.05287>
- [13] Zhu J, Yan Y, Zhao L, et al. Beyond homophily in graph neural networks: current limitations and effective designs[C]//Advances in Neural Information Processing Systems, Long Beach, USA, 2020, 33: 7793-7804.
- [14] Chien E, Peng J, Li P, et al. Adaptive universal generalized pagerank graph neural network[J/OL]. [2023-08-11]. <https://arxiv.org/pdf/2006.07988>
- [15] Jin D, Wang R, Ge M, et al. Raw-gnn: random walk aggregation based graph neural network[C]//Proceedings of the 31st International Joint Conference on Artificial Intelligence, Shenzhen, China, 2022: 2108-2114.
- [16] Fu X Y, Zhang J N, Meng Z Q, et al. Magnn: meta-path aggregated graph neural network for heterogeneous graph embedding[C]//Proceedings of the Web Conference, Taipei, China, 2020: 2331-2341.
- [17] Luan S, Hua C, Lu Q, et al. Revisiting heterophily for graph neural networks[J]. Advances in Neural Information Processing Systems, 2022, 35: 1362-1375.
- [18] Hochreiter S, Schmidhuber J. Long short-term memory[J]. Neural Computation, 1997, 9(8): 1735-1780.
- [19] Chung J, Gulcehre C, Cho K H, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[J/OL]. [2023-08-11]. <https://arxiv.org/abs/1412.3555>