

基于图核同构网络的图分类方法

徐立祥¹ 葛 伟¹ 陈恩红² 罗 斌³

¹(合肥学院人工智能与大数据学院 合肥 230601)

²(大数据分析与应用安徽省重点实验室(中国科学技术大学计算机科学与技术学院) 合肥 230027)

³(安徽大学计算机科学与技术学院 合肥 230601)

(xulixianghf@163.com)

Graph Classification Method Based on Graph Kernel Isomorphism Network

Xu Lixiang¹, Ge Wei¹, Chen Enhong², and Luo Bin³

¹(School of Artificial Intelligence and Big Data, Hefei University, Hefei 230601)

²(Anhui Province Key Laboratory of Big Data Analysis and Application(School of Computer Science and Technology, University of Science and Technology of China), Hefei 230027)

³(School of Computer Science and Technology, Anhui University, Hefei 230601)

Abstract Graph representation learning has become a research hotspot in the field of graph deep learning. Most graph neural networks suffer from oversmoothing, and these methods focus on graph node features and pay little attention to the structural features of graphs. In order to improve the representation of graph structural features, we propose a graph classification method based on graph kernel homomorphic network, namely KerGIN. The method first encodes the node features of the graph through graph isomorphism network (GIN), and then uses the graph kernel method to encode the graph structure. The Nyström method is further used to reduce the dimension of the graph kernel matrix. The graph kernel matrix is aligned with the graph feature matrix with the help of MLP, and the feature encoding and structure encoding of the graph are adaptively weighted and fused through the attention mechanism to obtain the final feature representation of the graph, which enhances the ability to express the structural feature information of the graph. Finally, the model is experimentally evaluated on seven publicly available graph classification datasets: compared with the existing graph representation models, KerGIN model is able to improve the graph classification accuracy substantially, and it can enhance the ability of GIN to represent the graph structural feature information.

Key words graph classification; graph neural network; graph kernel; Nyström method; graph attention mechanism

摘 要 图表示学习已成为图深度学习领域的一个研究热点. 大多数图神经网络存在过平滑现象, 这类方法重点关注图节点特征, 对图的结构特征关注度不高. 为了提升对图结构特征的表达能力, 提出了一种基于图核同构网络的图分类方法, 即 KerGIN. 该方法首先通过图同构网络 (graph isomorphism network, GIN) 对图进行节点特征编码, 并使用图核方法对图进行结构编码, 进一步利用 Nyström 方法降低图核矩阵的维度. 其次借助 MLP 将图核矩阵与图特征矩阵对齐, 通过注意力机制将图的特征编码和结构编码进行自适应加权融合, 进而得到图的最终特征表示, 提升了图结构特征信息的表达能力. 最后在 7 个公开的图分类

收稿日期: 2022-12-07; 修回日期: 2023-06-05

基金项目: 国家自然科学基金项目 (62176085); 国家自然科学基金国际合作重点项目 (61860206004)

This work was supported by the National Natural Science Foundation of China (62176085) and the Key International Cooperation Projects of the National Natural Science Foundation of China (61860206004).

通信作者: 陈恩红 (chench@ustc.edu.cn)

数据集上对模型进行了实验评估:与现有图表示模型相比,KerGIN模型能够在图分类准确度上有较大幅度提升,它可以增强GIN对图结构特征信息的表达能力。

关键词 图分类;图神经网络;图核;Nyström方法;图注意力机制

中图法分类号 TP391

现实世界中的大量问题都可以抽象成图模型(graph model),也就是节点和边的集合,包括自然语言处理^[1]、异常检测^[2-3]、学术网络分析、生物医疗、推荐系统等.图是一种不同规则的非欧氏几何数据,图数据的结构错综复杂,包含大量的信息,比如结构信息、节点特征信息等.通过学习基于图形的嵌入表示,可以获取结构化数据的顺序、拓扑、几何和其他关系特征.近年来,图深度学习的研究是学术界和产业界的一个热点,主要集中在节点分类^[4]、链路预测^[5]、图分类等.本文将重点关注的是图分类任务.图分类任务的关键是学习图与对应标签的映射关系.图分类在生物化学方面有着广泛的应用,例如对一些化学分子图进行分类来判断其活性、水溶性、毒性等.因此研究图分类问题有着重要意义.

图分类的重要方法之一是图核方法,它是一种计算图之间相似度的重要方法,把一些在低维空间下非线性不可分的数据转化到高维空间中,使得这些数据线性可分,是专门针对图数据的一种特殊方法.图核函数一般是根据专家知识设计的,它考虑了不同子结构的相似性,例如随机游走核和最短路径核.不同的图核函数之间也能相互融合,例如多图核学习^[6].这就为图分类引入不同的相似度度量方法和不同的偏差,从而生成具有不同性能的图分类模型.然而在缺乏专家知识的情况下,执行图分类任务时很难确定选择哪种图核函数是最好的.

随着深度学习的兴起,图卷积神经网络(graph convolution neural networks, GCN)^[7]成为图数据挖掘领域最重要的方法之一.GCN首次提出了卷积的方式融合图结构特征,提供了一个全新的视角,即在节点嵌入表示中将邻域节点的特征融入其中,与将节点特征直接通过全连接层分类的方法相比,在节点分类准确度上得到了很大提升.然而GCN存在共享权重、灵活性差、可扩展性不强的缺点,此外当网络层数增加时,会出现过平滑现象,导致每个节点的特征结果十分相似.为了解决GCN领域聚合时权值共享问题,带有注意力机制的图注意力网络(graph attention network, GAT)^[8]被提出,GAT具有高效低存储的优点,GAT是基于邻居节点的计算方式,它属于

一种归纳的学习方式.GAT的缺点就是只归纳了1阶邻居,导致GAT的感受野必须依赖深层的网络.为了解决GCN扩展性差的问题,GraphSAGE(graph sample and aggregate)^[9]提出了多种节点采样和聚合的方法,使图的嵌入表示更加灵活,当图中有新的节点加入时,固定的节点采样方式使得GraphSAGE无需对所有的节点进行重新训练,便可获取最新的嵌入.图神经网络(graph neural network, GNN)主要是针对节点特征的更新与提取,图分类要在此基础上增加图池化的操作,图池化主要是在图节点嵌入的基础上得到整个图的嵌入,其中主流的图池化方法有2种,即全局池化和分层池化.全局池化是在叠加图卷积之后运用全局池化操作(如最大池化和平均池化)选出能代表整张图表示的节点信息.分层池化借鉴了CNN中的池化,每次池化都会缩小数据的规模,对于图数据来说,就是通过某种算法减少节点的数目来完成逐层的抽取,从而实现图的池化,这种算法有Top- k ^[10]、图聚类池化等.图神经网络用于图分类的整个过程如图1所示.

为了提升图神经网络的图分类性能,近年来,一些研究人员致力于把图核与图神经进行融合,提出了许多基于图核的图神经网络框架.例如图卷积核网络(graph convolutional kernel network, GCKN)^[11]采用随机游走核提取路径并投影到核空间中,然后把核空间中路径信息聚合到起始节点上.图结构核网络(graph structural kernel network, GSKN)^[12]在GCKN的基础上增加了匿名随机游走核,使得提取局部子结构的能力得到了进一步加强.这2种框架虽然能在一定程度上提升表达能力,但提取路径的操作耗费大量时间.核图神经网络(kernel graph neural network, KerGNN)^[13]也是采用随机游走核与图神经网络进行融合,与先前工作不同的是,它采用可训练隐藏图作为图滤波器,与子图进行结合,并利用图核更新节点嵌入,使得图神经网络具有一定的可解释性,降低了图核计算的时间复杂度,然而对于图分类的性能提升不大.

影响图分类性能主要有2个方面:1)对图节点的特征编码;2)对图的结构编码.在一些化学分子图中,

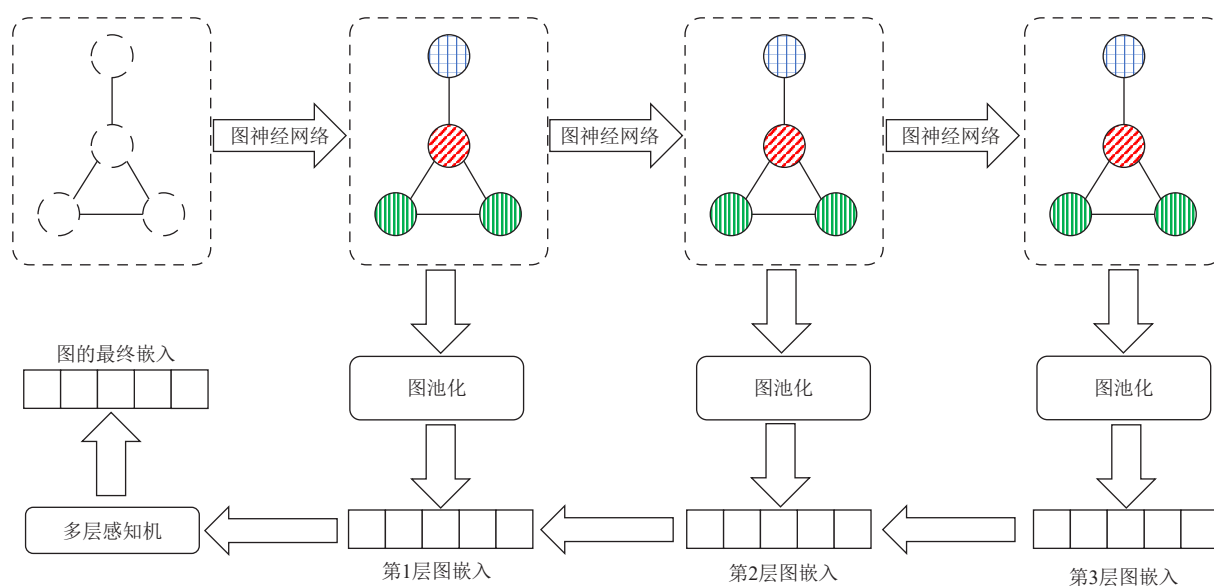


Fig. 1 The process of graph classification based on graph neural networks

图1 基于图神经网络的图分类过程

结构对性能的影响占比很大,这类图的性质与特定子图结构的相关性比较强,对于一些社交网络图,这类图不依赖于特定的局部结构,节点特征分布对图分类性能影响较大.基于图核的方法,对图的结构编码的方法重点关注图之间的结构相似性,本质上来讲也是对图的一种结构相似性编码,因此图核方法在一些化学分子图上表现出良好的性能,但对于一些社交网络图表现出的性能相对较差.而基于图神经网络的模型更加关注节点的特征,其本质上也是基于消息传递的框架.当今的一些图神经网络框架存在3个问题:1)在图神经网络的邻域聚合的过程中,获取了图的树形结构信息和邻域内的节点特征信息,却无法区分例如多元环等高阶子结构;2)为了获得更好的性能,图神经网络会叠加多层特征信息,但是这个层数的设置很难把握,如果设置过大,会产生过平滑问题,也即是,它使得深层的节点嵌入表示都十分相似,因此,把这些相似的节点嵌入堆叠会破坏图节点特征编码与图标签的单射关系,最终会导致图分类性能下降;3)以往图核和图神经网络融合的工作主要是采用随机游走核来提升节点获取邻域内高阶子结构信息的能力,但是这种方法的时间复杂度较大.此外,随机游走核具有不确定性,无法保证每一次游走的路径都包含了高阶子结构信息.为了解决这3个问题,本文将WL(Weisfeiler-Lehman)^[14]核与图神经网络融合起来,WL核对图数据进行结构相似性编码,图同构网络(graph isomorphism network, GIN)^[15]对图的节点特征进行编码,并将这2部分编

码通过注意力加权方式进行融合,相当于在基于消息传递的图神经网络中增加了图的结构相似性信息,提升图神经网络的表达能力.

本文的主要贡献包括4个方面:

1) 提供了一个新的视角,将WL核应用到图神经网络领域中,通过GIN与WL核方法进行融合,丰富图的结构特征和节点特征,提升GIN对高阶图结构的判别能力;

2) 针对不同类型的图数据集,提出基于注意力机制的图结构相似编码和图节点特征编码的融合方法,完成两者权重的自适应学习;

3) 在图核中,使用Nyström方法构建一个低秩矩阵去近似原核矩阵,从而大大降低图核矩阵的维度,解决图核矩阵在计算中运算代价大的问题;

4) 在7个公开的图数据集上,与一些当前已知性能最好的多种代表性的图分类基准模型相比,所提出的模型在大多数数据集上可以表现出更好的性能.

1 相关工作

1.1 基于WL核的图分类方法

WL核是当前应用最广泛的图核方法之一,它是基于子树的图核方法,其主要思想是对图进行子树分解,使用子树间的相似度来代替图的相似度,它是一种基于1-WL图同构测试所提出的一种快速特征提取算法,详细的操作步骤为:对于拥有多个节点标签的离散图,首先对各个节点进行邻域聚合;然后对

邻居节点进行排序,与此同时,节点标签和排序后的邻居标签共同组成多重集,并对这些多重集进行压缩映射,生成对应的新标签,进而将这些新标签赋给节点,这样就完成了一次迭代.在迭代过程中,重新标记的过程以及多重集的压缩是所有输入图同时进行的,并且所有图共享标签压缩的对应关系.

具有 H 次迭代的 2 个图 G_1 和图 G_2 上的 WL 子树核被定义为

$$K_{\text{sub}}^h(G_1, G_2) = \langle \phi_{\text{WL}}^h(G_1), \phi_{\text{WL}}^h(G_2) \rangle, \quad (1)$$

$$\phi_{\text{WL}}^h(G) = (c_0(G, \sigma_{01}), c_1(G, \sigma_{02}), \dots, c_h(G, \sigma_{0|\Sigma_h|})), \quad (2)$$

其中 $\phi_{\text{WL}}^h(G)$ 表示 h 次迭代中节点标签出现的次数, $c_i(G, \sigma_{ij})$ 表示节点 σ_{ij} 在图 G 中出现的次数. WL 核的图同构测试能力被证明是图神经网络的上限,图同构是图论中用来描述 2 个图在拓扑结构上是否完全等价的一个概念,如果 2 个图 G_1 和 G_2 完全等价,那么就称 G_1 和 G_2 是同构的.判断 2 个图是否为同构是一个非常困难的问题,目前还没有一个可以在多项式时间内求解的算法.除了一些极端的情况外, WL 测试可以被用来判断 2 个图是否同构.与一些传统的图核方法,如随机游走核^[16]、最短路径核^[17]或 REGK 核^[18]相比, WL 测试的同构检测能力是该领域的一个重要进步,该方法的运行时间仅与节点和边的数量呈线性关系,同时在各种统计学习任务中表现出良好的性能.

1.2 基于图同构网络 (GIN) 的图分类方法

图神经网络的任务主要是进行图节点表征学习,并基于学习到的节点表征进行下游的任务,例如节点分类或者链路预测等.而 GIN 提出了图级别表示学习,即图表征学习.图表征学习要求根据节点属性、边和边的属性(如果有的话)生成一个向量作为图的表征,基于图表征可以做图的预测.基于 GIN 的图表征学习主要包含 2 个过程:1)计算得到图的节点特征,即每一个节点的特征依次聚合各个邻居节点的特征,常用的图神经网络节点聚集函数有求和函数、平均函数和最大化函数.2)在 GIN 中,节点聚集函数选择的是求和函数,而不是选择平均函数,原因是平均函数不能识别某一节点出现的次数,不能精确描述多重集,它只能捕捉实体的特征分布,而最大化函数适合捕捉具有代表性的元素或“骨架”,而不能区分确切的结构或分布的任务.图神经网络还有其他的节点聚合函数,如加权平均、LSTM 池化等,而对判断同构问题来说,本文使用了基于注意力机制的自适应加权求和方法,此方法具有较强的表征能力.

为了实现图节点特征编码与图标签的单射,GIN 使用加法作为聚合函数,并使用多层感知机来模拟函数的组合,实现每层之间的映射关系,更新 GIN 节点表示函数为:

$$\mathbf{h}_v^{(k)} = \text{MLP}^{(k)} \left((1 + \varepsilon^{(k)}) \mathbf{h}_v^{(k-1)} + \sum_{u \in N(v)} \mathbf{h}_u^{(k-1)} \right), \quad (3)$$

其中 $\mathbf{h}_v^{(k)}$ 表示在第 k 层节点 v 的特征表示, $N(v)$ 表示节点 v 的邻居节点, ε 为一个可学习的参数或者一个固定的标量,进一步对图上各个节点的表征进行图池化(graph pooling)或图读出(graph readout),得到整个图的表征(graph representation),并将所有层读出的图表征拼接起来:

$$\mathbf{h}_G = \text{Concat}(\text{Readout}(\{\mathbf{h}_v^{(k)} | v \in G\}) | k = 0, 1, \dots, K), \quad (4)$$

其中 \mathbf{h}_G 是整个图的嵌入表示,这里的读出函数分别使用了求和、求平均和 MLP.

1.3 Nyström 方法

图核方法的计算代价比较大,对于一个有 n 个图的数据集,得到的图核矩阵是 n^2 个元素,当 n 非常大时,图的结构编码维度就比较大, Nyström 方法^[19]是作为一种使用简单正交规则离散积分方程的方法而被提出的,是一种广泛使用的降维方法,用于给定的列采样子集逼近核矩阵^[20]. Nyström 常用在核空间的计算问题中,对于一个样本集合 $\{x_1, x_2, \dots, x_n\}$,以及它们的核矩阵 $\mathbf{K} \in \mathbb{R}^{n \times n}$, Nyström 可通过采样的方式,构建一个低秩矩阵去近似表示原核矩阵,降低核矩阵在计算中的运算代价. Nyström 可以作为一种无监督的降维编码.同时,也可以得到核空间中样本的矩阵表示 $\tilde{\mathbf{K}} \in \mathbb{R}^{n \times d}$.

2 图核同构网络 KerGIN

本节将重点介绍 KerGIN,该模型以 GIN 为基础,借助图核方法,将图的结构特征和节点特征进行深度融合.本文提出的模型框架如图 2 所示,整个模型分为 3 个部分:GIN 编码器、图核和注意力模块,下面将对模型的每一部分进行详细的介绍.首先介绍关于图核和相关图同构网络模型的一些基本概念.一个图可以表示为 $g = (V, \mathbf{X}, \mathbf{A})$,其中 $V = \{v_1, v_2, \dots, v_N\}$ 表示图节点的集合, $\mathbf{X} \in \mathbb{R}^{N \times d}$ 表示图中节点的特征,总共有 N 个节点,每个节点的特征维度都是 d , $\mathbf{A} \in \mathbb{R}^{N \times N}$ 表示图的邻接矩阵,本文所研究的图都是无权无向图,如果节点 v_i 与 v_j 之间存在边,则 $A_{ij} = 1$,否则 $A_{ij} = 0$.对于图分类问题,给定一个数据集

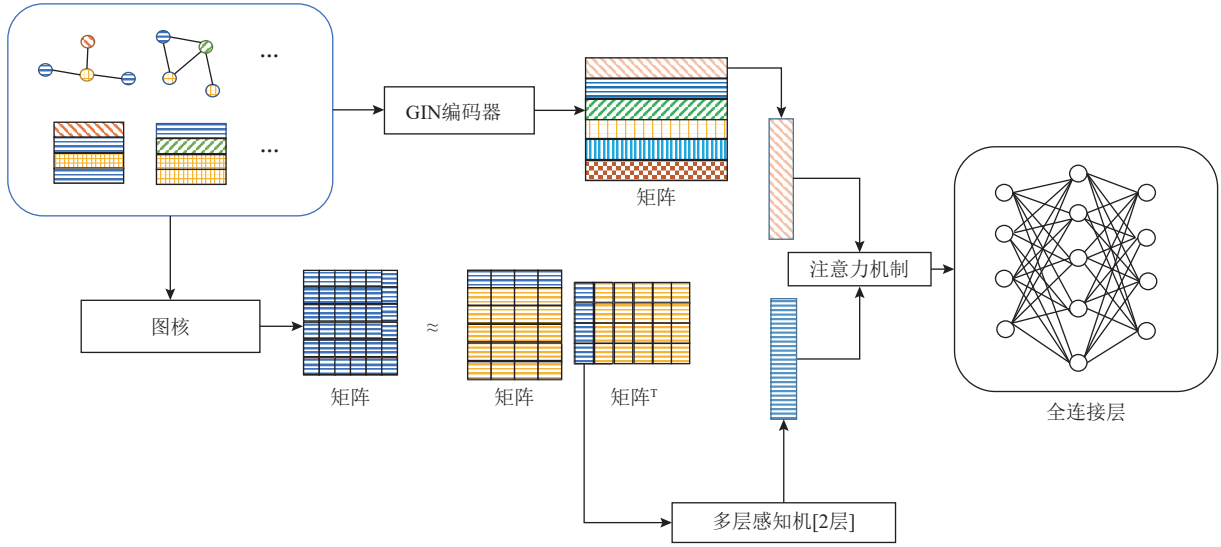


Fig. 2 The overall architecture of KerGIN

图2 KerGIN 总体框架

$\{(g_1, y_1), (g_2, y_2), \dots, (g_n, y_n)\}$, 其中 y 表示图的标签. 图分类任务的目的就是学习到由图 g 到标签 y 的映射函数 $y_g = f(g)$. 本文使用 one-hot 编码处理离散标签. 例如 4 个标签分别由 4 维向量 $(1, 0, 0, 0)$, $(0, 1, 0, 0)$, $(0, 0, 1, 0)$, $(0, 0, 0, 1)$ 来表示.

2.1 GIN 编码

根据图的邻接矩阵 A 和图的特征 X , 使用 GIN 对图进行编码, 首先对每个节点进行邻居的采样和聚合, 采样一个节点的所有邻居, 聚合邻居采用求和函数, 即每个节点的特征加上邻居节点的特征. 节点特征采样函数和聚合函数分别为:

$$a_v^{(k)} = \text{Aggregate}^{(k)}(\{h_u^{(k-1)} : u \in N(v)\}), \quad (5)$$

$$h_u^{(k)} = \text{Combine}^{(k)}(h_u^{(k-1)}, a_v^{(k)}), \quad (6)$$

其中 $\text{Aggregate}()$ 是采样邻居节点函数, $\text{Combine}()$ 是求和函数, 因为 GIN 已经证明了求和函数是单射的. 进一步在 k 层得到的特征向量 $h_v^{(k)}$ 经过一个多层感知机:

$$H_v^{(k)} = \text{MLP}^{(k)}(h_v^{(k)}), \quad (7)$$

这样即可得到每个节点经过消息传递后的特征向量 $H_v^{(k)}$, 然后再将每一层得到的特征向量相加, 即把每一层求和后的特征向量拼接起来, 这样就得到了图的特征编码 H_G , 该过程采用的函数表达式为:

$$H^{(k)} = \sum_{v \in V} H_v^{(k)}, \quad (8)$$

$$H_G = \text{Concat}(\{H^{(k)} \mid k = 0, 1, \dots, m\}). \quad (9)$$

2.2 WL 图核矩阵的生成

在 2.1 节中 GIN 已经对图进行了节点特征编码, 在

本节中重点关注对于图的结构特征编码. 由于 GIN 对图的结构表征能力有限, 所以引入一个图核矩阵即图的结构相似性编码来增强 GIN 的结构表征能力.

图核用于计算 2 个图的相似度, 对于一个图数据集 $G = \{g_1, g_2, \dots, g_N\}$, 计算每 2 个图的核值, 构成核矩阵 $K \in \mathbb{R}^{N \times N}$, 图核矩阵中的 i 行表示的图与其他图 g_i 的结构相似度, 相当于对图 g_i 的结构相似性编码, 本文使用的图核是不带节点标签的 WL 核, 即输入 2 个图的邻接矩阵 $A \in \mathbb{R}^{N \times N}$, 不需要节点的特征或者节点的标签. 如图 3 所示, 对于 2 个原始的图, 聚合其邻居节点, 然后对聚合后的每一个节点进行重新的哈希编码, 即对节点使用新的颜色来表示, 这是进行 1 次迭代所得到的结果, 然后按颜色来统计所有节点的个数, 这样就把图转化为特征向量, 最后 2 个向量之间求内积, 即得到 2 个图的相似度, 求 2 个图的核值函数为:

$$K_{\text{WL}}(g_i, g_j) = \sum_{i=0}^k K_{\text{subtree}}^{(i)}(g_i, g_j), \quad (10)$$

$$K_{\text{subtree}}^{(i)} = \sum_{u \in V_i} \sum_{u' \in V_j} k_{\text{subtree}}^{(i)}(u, u'). \quad (11)$$

在 WL 核的计算过程中, 用内积来度量 2 个图的子树模式向量. 本文也选取了其他常用的图核函数进行了对比实验, 如最短路径 (shortest path, SP) 核、随机游走 (random walk, RW) 核, 详见 3.4 节, 最后采用了效果最优的 WL 核.

2.3 Nyström 降秩分解

在 2.2 节中得到了图数据集的核矩阵, 这里的图核矩阵的维度通常较大, 空间复杂度为 $O(N^2)$, 如果

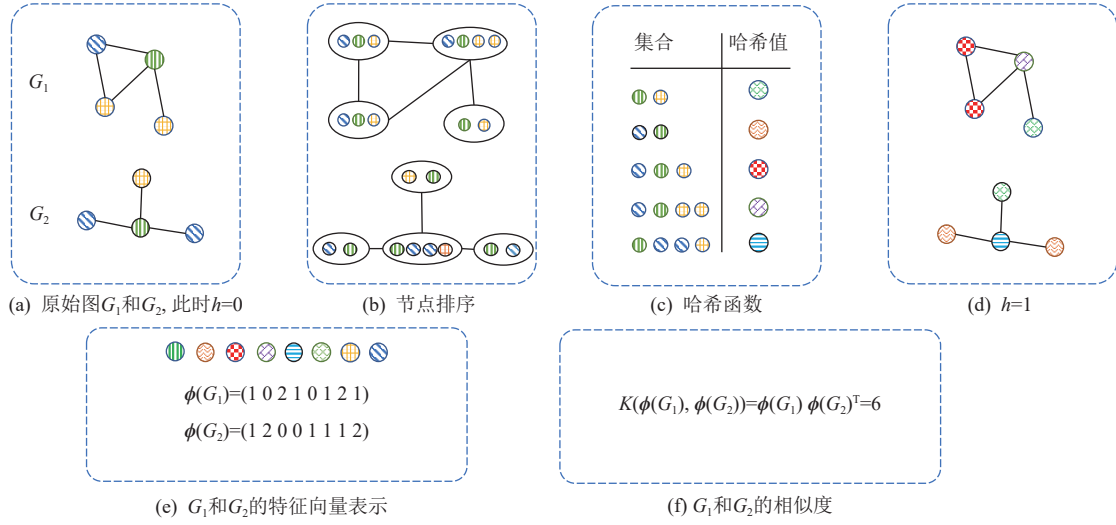


Fig. 3 Diagram of the implementation process of WL kernel

图3 WL核执行过程图

图数据集的数量庞大, 将导致后续的计算代价很大. Nyström方法常用在核空间的计算问题中, 通过降秩分解, 可以显著降低核矩阵的维度. 核矩阵 $\mathbf{K} \in \mathbb{R}^{N \times N}$ 是对称正定矩阵, 核矩阵的分解过程为:

$$\mathbf{K} = \begin{pmatrix} \mathbf{A} & \mathbf{B}^T \\ \mathbf{B} & \mathbf{C} \end{pmatrix}, \quad (12)$$

其中 $\mathbf{A} \in \mathbb{R}^{k \times k}$, $k < n$, 假设 $\mathbf{K} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T$, $\mathbf{A} = \mathbf{U}_A \mathbf{\Lambda}_A \mathbf{U}_A^T$, 令

$$\tilde{\mathbf{U}} = \begin{pmatrix} \mathbf{U}_A \\ \mathbf{B} \mathbf{U}_A \mathbf{\Lambda}_A^{-1} \end{pmatrix}, \quad (13)$$

则

$$\tilde{\mathbf{K}} = \tilde{\mathbf{U}} \mathbf{\Lambda}_A \tilde{\mathbf{U}}^T = \begin{pmatrix} \mathbf{A} & \mathbf{B}^T \\ \mathbf{B} & \mathbf{B} \mathbf{\Lambda}_A^{-1} \mathbf{B}^T \end{pmatrix}, \quad (14)$$

易得

$$\|\mathbf{K} - \tilde{\mathbf{K}}\| = \|\mathbf{C} - \mathbf{B} \mathbf{\Lambda}_A^{-1} \mathbf{B}^T\|, \quad (15)$$

此时得到的核矩阵 $\tilde{\mathbf{K}}$ 近似于原核矩阵 \mathbf{K} , 由矩阵分解可得 $\tilde{\mathbf{K}} \approx \mathbf{Q} \mathbf{Q}^T$, \mathbf{Q} 的维度是 $\mathbf{Q} \in \mathbb{R}^{N \times k}$, 其中 $k \ll N$, 这时核矩阵的空间复杂度由 $O(N^2)$ 降为 $O(Nk)$, 从而降低了运算代价.

由于降维后的核矩阵与2.1节中GIN编码维度不一致, 在这里使用神经网络对核矩阵的维度与GIN编码的向量对齐, 定义一个2层的神经网络, 共享1个隐藏层, 为了防止梯度消失或梯度爆炸现象的出现, 需要对核矩阵 $\tilde{\mathbf{K}}$ 中的每一行进行规范化, 并使用最小值中心化的方法进行归一化, 再经过全连接神经网络得到图核嵌入向量 \mathbf{h}_k , 即图的结构编码, 计算的函数表达式为:

$$\tilde{\mathbf{K}}'[i, j] = \frac{\tilde{\mathbf{K}}[i, j] - \min(r_i(\tilde{\mathbf{K}}))}{\max(r_i(\tilde{\mathbf{K}})) - \min(r_i(\tilde{\mathbf{K}}))}, \quad (16)$$

$$\mathbf{h}_k = \text{Softmax}(\text{ReLU}(\tilde{\mathbf{K}}' \mathbf{W}^0) \mathbf{W}^1), \quad (17)$$

其中 $\tilde{\mathbf{K}}[i, j]$ 表示核矩阵 $\tilde{\mathbf{K}}$ 中归一化后的第 i 行第 j 列元素的值, \mathbf{W} 是一个可学习的权重矩阵, 使得计算更加平滑, \mathbf{h}_k 是经过2层全连接神经网络后所得到的图结构的嵌入表示, 与经过GIN得到的图特征编码 \mathbf{H}_G 维度保持一致, 这为图分类的下游任务做准备.

2.4 GIN编码和WL核编码的融合

在2.2~2.3节中, 通过GIN和图核分别得到了第 i 个图的特征编码 $\mathbf{H}_G^i \in \mathbb{R}^{1 \times d}$ 和图的结构编码 $\mathbf{h}_k^i \in \mathbb{R}^{1 \times d}$ (取核矩阵 $\tilde{\mathbf{K}}$ 的第 i 行), 将 \mathbf{H}_G^i 与 \mathbf{h}_k^i 进行注意力加权求和, 函数表达式为:

$$\mathbf{H}_1 = \begin{pmatrix} \mathbf{H}_G \\ \mathbf{h}_k \end{pmatrix} \mathbf{W}, \quad (18)$$

$$\mathbf{c} = \text{Softmax}(\mathbf{H}_1 \mathbf{a}), \quad (19)$$

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_G \\ \mathbf{h}_k \end{pmatrix}^T \mathbf{c}, \quad (20)$$

其中 $\mathbf{W} \in \mathbb{R}^{d \times d}$ 为权重向量, 它可以使图的嵌入计算更加平滑, $\mathbf{a} \in \mathbb{R}^{d \times 1}$ 为注意力权重向量, $\mathbf{c} \in \mathbb{R}^{2 \times 1}$ 为注意力系数, $\mathbf{H} \in \mathbb{R}^{d \times 1}$ 为图特征编码和图结构编码的注意力加权融合后的向量表示, 进一步将 \mathbf{H} 输入到多层感知机或者支持向量机中进行图分类任务.

KerGIN提取图特征的算法描述见算法1, 输入为一组图数据集以及图的邻接矩阵和节点的特征. 对每个图使用GIN进行特征编码, 即对每一个节点进行邻域聚合, 得到节点的嵌入表示, 然后把所有节点的嵌入表示加起来, 这样就得到了图的表征向量. 与此同时使用WL核求每2个图之间的核值, 即对于每一个节点进行 h 次迭代的哈希编码, 这样就把整

个图映射成一个向量,将2个图的向量表示进行内积运算,这样就得到了2个图的相似度,也即得到了图的结构编码.进一步使用注意力机制将图的特征编码和图的结构编码进行加权求和,进而得到图的特征向量表示,并对所有的图都进行上述的操作,最后将这些图的向量表示输入到多层感知机或支持向量机中进行下游的分类任务.

算法 1. KerGIN 提取特征.

输入: 图数据集 $G = \{g_1, g_2, \dots, g_n\}$, 其中 $g = (V, E, X)$;

输出: 图的嵌入表示 $\phi(G) \in \mathbb{R}^{n \times d}$.

- ① 初始化节点嵌入 $h_0 = X$, 核矩阵 $K \in \mathbb{R}^{n \times n}$, 标量参数 $\varepsilon = \varepsilon_0$;
- ② for $k = 1, 2, \dots, m$ do
- ③ 聚合邻居节点的特征 $h_v^{(k)}$; /*式(3)*/
- ④ 将节点特征相加得到图的嵌入表示 $h^{(k)}$;
- ⑤ 将 $h^{(k)}$ 进行拼接得到图的嵌入表示 H_G ;
- ⑥ end for
- ⑦ for $i = 1, 2, \dots, n$ do
- ⑧ for $j = 1, 2, \dots, n$ do
- ⑨ 计算图 g_i 和 g_j 的结构相似度 $K[g_i, g_j]$;
- ⑩ end for
- ⑪ end for
- ⑫ 借助 Nyström 方法对 K 进行分解得到 \tilde{K} ;
- ⑬ 将 H_G 与 \tilde{K} 通过注意力机制加权求和得到图的嵌入表示 $\phi(G)$. /*式(18)~(20)*/

3 实 验

3.1 数据集的介绍和实验设置

1) 数据集. 本文使用7个公开的图分类数据集进行实验, 分别为 MUTAG^[21], PTC^[22], PROTEINS^[23], NCI1^[24], IMDB-B^[25], IMDB-M^[25], COLLAB^[25]. 前4个为化学分子数据集, 后3个为社交网络数据集. 7个数据集的简介为:

① MUTAG^[21]. 该数据集包含了188个化合物结构图, 依据它们对细菌的诱变作用, 可被分为2类. 图中的节点和节点标签分别表示原子和原子种类, 包括 C, N, O, F, I, Cl, Br.

② PTC^[22]. 该数据集全称是预测毒理学挑战, 用来发展先进的 SAR 技术预测毒理学模型. 这个数据集包含了针对啮齿动物的致癌性标记的化合物. 图中有2个类别的标签, 分别表示有致癌性和无致癌性.

③ PROTEINS. 该数据集^[23]中有1113个蛋白质

结构图, 图的标签分为2类, 分别表示酶或者非酶. 节点表示蛋白质的2级结构, 根据2级结构在氨基酸序列或者蛋白质3维空间中是否为邻居来确定节点之间边的存在性.

④ NCI1^[24]. 该数据集是一个关于化学分子的数据集, 是根据非小细胞肺癌活性筛选的, 图的标签分为2类, 表示具有或不具有抗癌活性, 共包含4110个化合物的图结构.

⑤ IMDB-B^[25]. 该数据集是一个电影合作数据集, 来源于互联网电影数据库IMDB. 图中的节点表示演员, 如果2个演员在同一部电影中出现, 则他们对应的节点之间就存在一条边, 这些合作图分为动作和浪漫2种类型. 合作图是以每个演员为中心的网络图, 图分类的任务是判断这些自我中心网络图属于动作类型还是浪漫类型. 此外该数据集还有一个多类型版本.

⑥ IMDB-M^[25]. 该数据集的任务也是对演员子网络图按电影类型进行分类.

⑦ COLLAB^[25]. 该数据集是一个关于科研合作的数据集, 涵盖了高能物理、凝聚态物理和天文物理3个领域中生成的不同研究人员的自我中心网络(ego-network)图. 对应的图标签为研究人员所属的研究领域. 分类的任务是确定这些自我中心网络图对应的研究人员所属的研究领域.

实验中使用的7个图数据集的信息统计结果如表1所示.

Table 1 Information Statistic of Datasets

表1 数据集的信息统计

数据集	类型	图数量	平均节点数	平均边数	种类	节点属性	属性维度
MUTAG	化学分子	188	18	20	2	Disc.	1
PTC	化学分子	344	26	51	2	Disc.	1
PROTEINS	化学分子	1 113	39	73	2	Disc.	1
NCI1	化学分子	4 110	30	65	2	Disc.	1
IMDB-B	社交网络	1 000	20	97	2	No	2
IMDB-M	社交网络	1 500	13	66	3	No	3
COLLAB	社交网络	5 000	74	2 458	3	No	5

2) 基准方法. 本文将选择当前已知性能最好的多种代表性图分类方法作为基准方法进行实验对比, 分别为基于图核的方法、基于图神经网络的分类方法、基于图池化的分类方法以及近几年一些与图核结合的图神经网络方法, 以此来证明本文模型的有效性. 基于图核的图分类方法有 WL 核^[14]和 DGK 核^[25], 基于图神经网络的图分类方法包括 GIN^[15], DCNN^[26],

PATCHY-SAN^[27]. 基于图池化的图分类方法有 SUGAR^[28]、AVCN(H)^[29], SLIM^[30]. 基于图核与图神经网络融合的图分类方法有 GCKN^[11], GSKN^[12], GSNN^[31]. 本文的分类任务属于有监督学习.

3) 参数设置. 模型训练过程中采用常用的参数设置, 设置学习率 $lr=0.0001$, 训练批次 $batch_size=16$, $epoch=600$, Nyström 方法^[19]中对核矩阵降秩分解后的维度 d 为数据集中图数量的 $1/2$, 全连接层神经网络中隐藏层维度分别为 16 和 8. WL 核的迭代次数 $h=3$. 数据集中 90% 作为训练集, 其余的 10% 作为测试集.

3.2 实验结果

本节将在 7 个数据集上对 KerGIN 和其他所有基准方法进行分类评估. 本文采用 10 次 10 交叉验证, 即将数据集分成 10 份, 每次取 1 份作为测试集, 剩下的 9 份作为训练集; 然后对这 10 次的结果求平均. 每个数据集的分类准确度如表 2 所示.

表 2 展示所有方法在 7 个公开数据集的测试准确度. KerGIN 在大多数数据集上的表现优于基准方法. 其中 MUTAG 数据集的平均准确度为 95.2%, 高于除 SUGAR 外的所有基准方法. 与 GSKN 相比, KerGIN 的准确度在 PTC 数据集上提升了 3.3 个百分点. 在 PROTEINS 数据集上, KerGIN 的准确度相比 GSKN 方法提升了 6.1 个百分点. 对于 NCI1 数据集, KerGIN 的准确度比 GCKN 提升了 4.8 个百分点. 在 IMDB-B 和 IMDB-M 数据集上, KerGIN 的准确度比准确度排名第 2 的 GSKN 方法分别提升了 1.7 个百

分比和 0.8 个百分点. 在 COLLAB 数据集上, KerGIN 准确度接近于最先进的 GCKN 方法. 特别是在一些化学分子数据集上 KerGIN 表现更突出, 与最新的 2 个基于图核的图神经网络方法相比, KerGIN 具有更优越的性能.

为了比较不同方法的综合性能, 分别统计了不同方法的平均排名, 即对每一个方法, 求出其在各个数据集上分类准确度的平均排名情况, 相关的计算公式为:

$$\bar{R} = \frac{1}{n} \left(\sum_{i=1}^n rank(d_i) \right), \quad (21)$$

其中 \bar{R} 表示平均排名, $rank(d_i)$ 表示在第 i 个数据集上的分类准确度排名, n 表示数据集的数量.

在图 4(a) 中, 与所有基准方法比较, KerGIN 在 7 个公开的图分类数据集上, 相比较最优的基准方法, 准确度的平均排名为 1. 由此可知 KerGIN 的图分类性能要优于大多数基准方法. 如图 4(b) 所示, KerGIN 相较于最优的基准方法在 6 个数据集上的分类准确度都有不同程度的提升. 由于 KerGIN 在 MUTAG 数据集上没有提升, 所以在图 4(b) 中只选择了有提升的 6 个数据集进行展示. 其中, 在 PROTEINS 数据集上提升了 7.5%, 在 IMDB-B 数据集上提升了约 2.1%, 在 PTC 数据集上提升了约 3.8%, 在 NCI1 数据集上提升了 0.93%, 在 IMDB-M 数据集上提升了 1.34%, 在 COLLAB 数据集上提升了 0.36%.

Table 2 Classification Accuracy on Each Public Dataset

表 2 在各个公开数据集上的分类准确度

%

方法	数据集						
	MUTAG	PTC	PROTEINS	NCI1	IMDB-B	IMDB-M	COLLAB
WL	90.4(8)	59.9(8)	75.0(9)	86.0(2)	73.8(7)	50.9(5)	78.9(5)
DGK	82.6(11)	57.3(9)	71.6(10)	62.2(9)	66.9 (10)	44.5(7)	73.1(7)
GIN	89.4(9)	64.6(6)	76.2(6)	82.7(4)	75.1(6)	52.3(4)	80.2(4)
DCNN	67.0(12)	56.6(10)	61.3(11)	62.6(8)	49.1(9)	33.5(8)	52.1(9)
PATCHY-SAN	92.6(6)	60.0(7)	75.9(7)	78.6(7)	71.0(8)	45.2(6)	72.6(8)
SUGAR	96.7(1)	77.5(3)	81.3(3)	84.3(3)			
GCKN	91.6(7)	68.4(5)	76.2(6)	82.0(5)	76.5(5)	53.3(3)	82.9(2)
GSKN	93.3(4)	85.2(2)	82.3(2)		79.9(2)	59.3(2)	81.8(3)
GSNN	94.7(3)		78.4(4)		78.1(3)		
AVCN(H)	89.3(10)	62.3(8)	75.7(8)		73.4(8)	50.9(5)	80.2(4)
SLIM	93.2(5)	72.4(4)	77.4(5)	80.5(6)	77.2(4)	53.3(3)	78.2(6)
KerGIN (本文)	95.2(2)	88.5(1)	88.4(1)	86.8(1)	81.6(1)	60.1(1)	83.2(1)

注: 加粗数字表示最优结果, 括号里的数字表示该方法在每个数据集上的分类准确度排名.

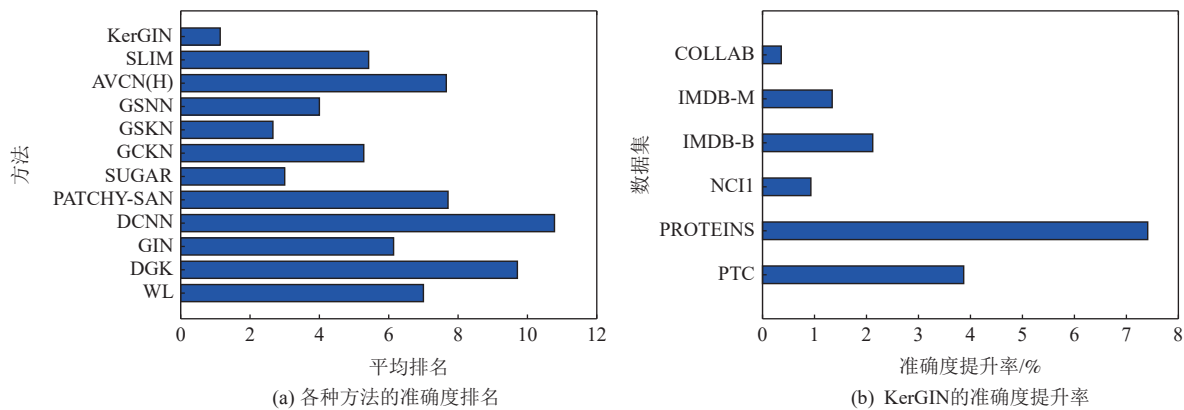


Fig. 4 Average rank and accuracy improvement rate of various methods

图 4 各种方法的平均排名和 KerGIN 的准确度提升率

3.3 消融实验

为了验证图核模块是否在整个模型中起关键作用以及 MLP 对分类准确度的影响,设计了一组消融实验,即将 KerGIN 模型中的图核模块 (GIN-MLP),与本文方法 KerGIN 和基准模型 GIN 进行比较.从表 3 可以看出,GIN-MLP 和 GIN 的图分类准确度差异并不大,说明在 KerGIN 中起到关键作用的不是 MLP.比较 GIN-MLP 和 KerGIN 模型的图分类准确度可以得出:图核模块在整个模型中起了关键作用.

为了研究注意力机制对图分类结果的影响及注意力机制的作用,选择了 2 种常见的融合策略进行对比实验,分别为拼接和求和,即把图结构编码和图特征编码拼接或者求和.如表 4 所示,KerGIN-con 表示采用拼接策略,KerGIN-sum 表示采用求和策略,KerGIN-att 表示采用注意力机制策略.可以看出,采

用拼接策略的分类效果不及求和策略和注意力机制策略,而采用注意力机制策略的分类准确度在这 3 种策略中最高.因此,在本文中注意力机制是较适宜的融合策略.这是因为不同数据集对于图结构编码和图特征编码的偏重程度不同,因此简单拼接和求和很难获得好的实验效果.

3.4 实验分析

本节将分析模型训练的过程以及图结构编码和图特征编码在不同类型数据集上的变化情况.图 5 展示了 MUTAG, PTC, PROTEINS, NCII, IMDB-B, IMDB-M 这 6 个数据集在训练和测试过程中的损失值随训练轮数的变化情况.为便于排版,本文选择了前 6 个数据集的损失变化情况进行图例展示.这 6 个数据集整体在 100 个训练轮数时损失下降得比较快,在 400 个训练轮数时损失下降的幅度较少,在 600 个训练轮数时基本趋向于收敛.其中 MUTAG 数据集

Table 3 Ablation Experiment Based on MLP and Graph Kernel

表 3 基于 MLP 与图核的消融实验

方法	MUTAG	PTC	PROTEINS	NCII	IMDB-B	IMDB-M	COLLAB
WL	90.4	59.9	75.0	86.0	73.8	50.9	78.9
GIN	89.4	68.4	76.2	82.7	75.1	52.3	80.2
GIN-MLP	88.7	68.5	76.8	81.9	75.4	52.6	79.8
KerGIN (本文)	95.2	88.5	88.4	86.8	81.6	60.1	83.2

注:最优结果用加粗数字标识.

Table 4 Ablation Experiment Using with Different Fusion Strategies

表 4 使用不同融合策略的消融实验

策略	MUTAG	PTC	PROTEINS	NCII	IMDB-B	IMDB-M	COLLAB
KerGIN-con	94.7	85.2	86.8	86.3	78.9	56.8	79.6
KerGIN-sum	94.9	86.3	87.5	86.3	79.4	58.6	81.7
KerGIN-att (本文)	95.2	88.5	88.4	86.8	81.6	60.1	83.2

注:最优结果用加粗数字标识.

在 500 个训练轮数时收敛, PTC 数据集在 450 个训练轮数时收敛, PROTEINS 数据集在 200 个训练轮数时开始收敛, NCI1 数据集在 100 个训练轮数时收敛. IMDB-B 数据集大约在 600 个训练轮数时收敛,

IMDB-M 在 200 个训练轮数时开始收敛. 6 个数据集在训练的过程中, 训练集损失值和测试集损失值之间的差距很小, 所以在训练过程中没有过拟合或欠拟合现象.

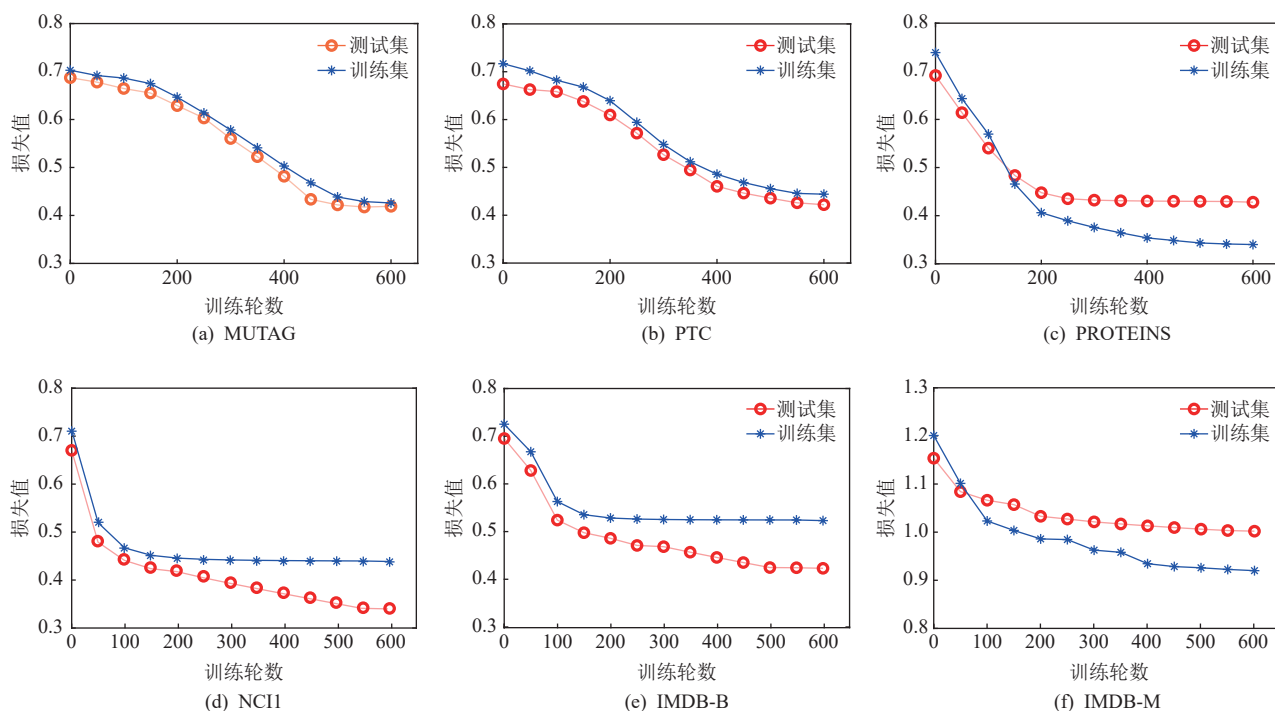


Fig. 5 Variation of loss values for training and testing on six datasets

图 5 6 个数据集上训练与测试的损失值变化

任何图神经网络编码器都能用作图特征编码, 本文除了实验中使用的 GIN 编码器, 还使用了 2 种流行的图神经网络框架: GCN 和图注意力网络 GAT 进行了对比实验. 实验结果如图 6(a) 所示, 可以观察到 GIN 的编码效果要好于 GCN 和 GAT. 这可能是因为 GIN 的图表达能力更加强大, 更适用于图的特征

编码. 此外, 又研究了图核编码器的长度对图分类性能的影响, 图 6(b) 显示了在 7 个公开数据集上 16~160 的不同图核编码长度的 KerGIN 的准确度. 可以看出, 在一定范围内, 图分类的准确度会随图核编码器长度的增加而增加, 当图核编码器长度大于 160 时, 图核编码器长度对图分类准确度的影响较小,

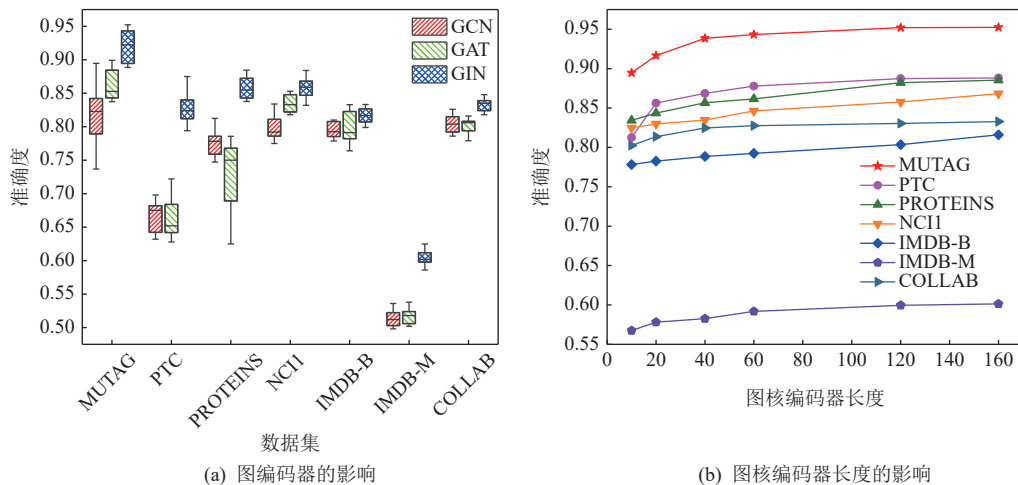


Fig. 6 KerGIN based on different graph encoders and KerGIN based on different length kernel encoders

图 6 基于不同图编码器的 KerGIN 和基于不同长度图核编码器的 KerGIN

所以在一定范围内核编码器长度对分类准确度有重要影响,因此适当降低核编码器的维度不会影响图分类的准确度.

如图 7(a)所示,7 个数据集权重系数都不相同,但是可以明显看出前 4 个数据集的图结构编码的权重大于图特征编码,后 3 个数据集的图结构编码的权重小于图特征编码.由于前 4 种图数据集为化学分子、后 3 种图数据集为社交网络,因此在这 2 种数据集上,图结构编码和图特征编码的权重系数不同.进一步,又探讨了不同图核函数对 KerGIN 分类准确度

的影响,因为图核函数的选择通常是根据专家经验进行选取,很难直接确定 KerGIN 更适用于哪种图核函数,因此,本文选择了 3 种类型的图核函数 RW, SP, WL, 并通过实验对比展示了哪种图核更适用于 KerGIN 的图结构编码.在图 7(b)中,由于 RW 核在大数据集上运行时间过长,所以选择了在 4 个小数据集上进行实验.可以清楚地看出 WL 核在 4 个数据集上综合表现最好,其次是 SP 核, RW 核实验效果最差,此外 RW 核时间复杂度比较大,在一定的时间内无法完成一些大规模图数据集的实验.

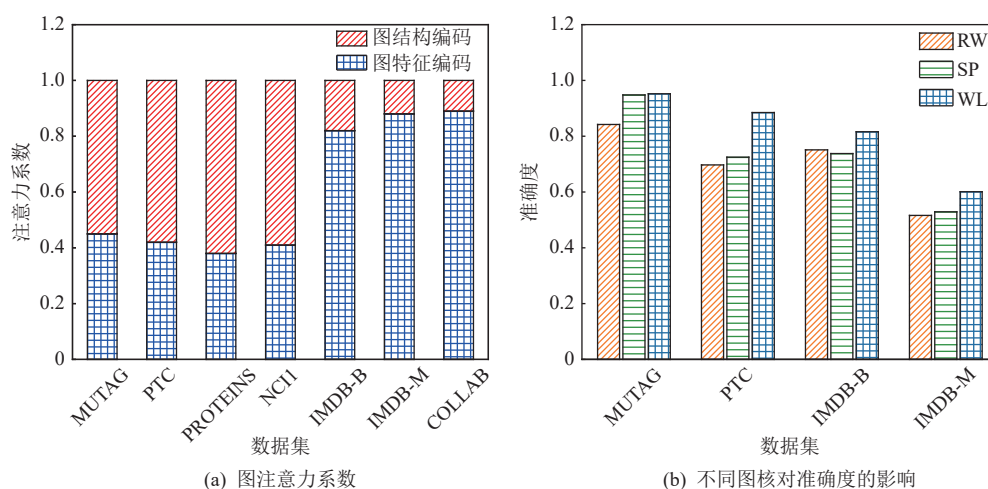


Fig. 7 The weight ratio of graph feature coding and structure coding, and the classification accuracy under different graph kernels

图 7 图特征编码和结构编码的权重占比和不同图核下的分类准确度

通过以上实验分析可以得出,本文方法在化学分子数据集上的分类准确度比在社交网络数据集上具有明显的优势,因此本文方法更适用于化学分子的图分类任务.此外,化学分子的特性受特定局部子结构的影响较大,所以图结构编码对其分类准确度起至关重要的作用,甚至 1 个局部子结构就可以成为其分类的主要依据,比如官能团(functional group).而对于社交网络数据集来说,它对图特定子结构的依赖性相对较小,而对图节点特征依赖性较大.

4 结 论

本文提出了一种基于 GIN、图核以及注意力机制相融合的图表征学习和图分类的方法,该方法提升了 GIN 对图中特定结构的判别能力.实验结果表明,图结构编码对图分类结果影响较大,将图核作为图结构编码,在一定程度上解决了基于消息传递的图神经网络无法识别图中高阶信息的问题,本文方法能够自适应调节图特征编码与图结构编码的权重,

对图分类的准确度有较大的提升,在分类准确度上优于所选的一些基准方法.

作者贡献声明:徐立祥提出了算法思路和实验方案,并撰写论文;葛伟负责完成实验验证,并整理论文;陈恩红和罗斌提出指导意见并修改论文.

参 考 文 献

- [1] Peng Hao, Li Jiangxin, He Yu, et al. Large-scale hierarchical text classification with recursively regularized deep graph-CNN[C] //Proc of the 27th World Wide Web Conf. New York: ACM, 2018: 1063-1072
- [2] Noble C, Cook D. Graph-based anomaly detection[C] //Proc of the 9th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2003: 631-636
- [3] Akoglu L, Tong H, Koutra D. Graph based anomaly detection and description: A survey[J]. *Data Mining and Knowledge Discovery*, 2015, 29(3): 626-688
- [4] Xie Xiaojie, Liang Ying, Wang Zisen, et al. Heterogeneous network node classification method based on graph convolution[J]. *Journal of*

- Computer Research and Development*, 2022, 59(7): 1470–1485 (in Chinese)
(谢小杰, 梁英, 王梓森, 等. 基于图卷积的异质网络节点分类方法[J]. *计算机研究与发展*, 2022, 59(7): 1470–1485)
- [5] Meng Xuying, Zhang Qijia, Zhang Hanwen, et al. Personalized privacy preserving link prediction in social networks[J]. *Journal of Computer Research and Development*, 2019, 56(6): 1244–1251 (in Chinese)
(孟绪颖, 张琦佳, 张瀚文, 等. 社交网络链路预测的个性化隐私保护方法[J]. *计算机研究与发展*, 2019, 56(6): 1244–1251)
- [6] Xu Lixiang, Bai Lu, Xiao Jin, et al. Multiple graph kernel learning based on GMDH-type neural network[J]. *Information Fusion*, 2021, 66: 100–110
- [7] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks[C] // Proc of the 7th Int Conf on Data Mining. Piscataway, NJ: IEEE, 2018: 875–880
- [8] Veličković P, Cucurull G, Casanova A, et al. Graph attention networks[J]. arXiv preprint, arXiv: 1710. 10903, 2017
- [9] Hamilton W, Ying Zhitao, Leskovec J. Inductive representation learning on large graphs[J/OL]. *Advances in Neural Information Processing Systems*, 2017[2022-11-25]. <https://proceedings.neurips.cc/paper/6703>
- [10] Gao Hongyang, Ji Shuiwang. Graph U-Nets[C] // Proc of the 36th Int Conf on Machine Learning. Cambridge, MA: MIT, 2019: 2083–2092
- [11] Chen Dexiong, Jacob L, Mairal J. Convolutional kernel networks for graph-structured data[C] // Proc of the 37th Int Conf on Machine Learning. Cambridge, MA: MIT, 2020: 1576–1586
- [12] Long Qingqing, Jin Yilun, Wu Yi, et al. Theoretically improving graph neural networks via anonymous walk graph kernels[C] // Proc of the 31st World Wide Web Conf. New York: ACM, 2021: 1204–1214
- [13] Feng Aosong, You Chenyu, Wang Shiqiang, et al. KerGNNs: Interpretable graph neural networks with graph kernels[C] // Proc of the 36th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2022: 6614–6622
- [14] Shervashidze N, Schweitzer P, Jan E, et al. Weisfeiler-Lehman graph kernels[J]. *The Journal of Machine Learning Research*, 2011, 12(3): 2539–2561
- [15] Xu Keyulu, Hu Weihua, Leskovec J, et al. How powerful are graph neural networks?[C/OL] // Proc of the 7th Int Conf on Learning Representations. 2019[2022-11-25]. <https://openreview.net/forum?id=ryGs6iA5Km>
- [16] Sugiyama M, Borgwardt K. Halting in random walk kernels[J/OL]. *Advances in Neural Information Processing Systems*, 2015, 28[2022-11-25]. <https://openreview.net/pdf?id=HyZEFubd-H>
- [17] Borgwardt K M, Krieger H P. Shortest-path kernels on graphs[C] // Proc of the 5th Int Conf on Data Mining. Piscataway NJ: IEEE, 2005: 8
- [18] Xu Lixiang, Bai Lu, Jiang Xiaoyi, et al. Deep Rényi entropy graph kernel[J]. *Pattern Recognition*, 2021, 111: 107668
- [19] Prenter P M. The numerical treatment of integral equations[J]. *SIAM Review*, 1981, 23(2): 266–267
- [20] Williams C, Seeger M. Using the Nyström method to speed up kernel machines[J/OL]. *Advances in Neural Information Processing Systems*, 2000 [2022-11-25]. <https://proceedings.neurips.cc/paper/2000/file/19de10adbaa1b2ee13f77f679fa1483a-Paper.pdf>
- [21] Debnath A K, Lopez R L, Debnath G, et al. Structure-activity relationship of mutagenic aromatic and heteroaromatic nitro compounds. Correlation with molecular orbital energies and hydrophobicity[J]. *Journal of Medicinal Chemistry*, 1991, 34(2): 786–797
- [22] Toivonen H, Srinivasan A, King R D, et al. Statistical evaluation of the predictive toxicology challenge 2000–2001[J]. *Bioinformatics*, 2003, 19(10): 1183–1193
- [23] Borgwardt K M, Ong C S, Schönauer S, et al. Protein function prediction via graph kernels[J]. *Bioinformatics*, 2005, 21(1): 47–56
- [24] Wale N, Watson I A, Karypis G. Comparison of descriptor spaces for chemical compound retrieval and classification[J]. *Knowledge & Information Systems*, 2008, 14(3): 347–375
- [25] Yanardag P, Vishwanathan S V N. Deep graph kernels[C] // Proc of the 21st ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2015: 1365–1374
- [26] Atwood J, Towsley D. Diffusion-convolutional neural networks [J/OL]. *Advances in Neural Information Processing Systems*, 2016[2022-11-25]. <https://proceedings.neurips.cc/paper/2016/file/390e982518a50e280d8e2b535462ec1f-Paper.pdf>
- [27] Niepert M, Ahmed M, Kutzkov K. Learning convolutional neural networks for graphs[C] // Proc of the 33rd Int Conf on Machine Learning. Cambridge, MA: MIT, 2016: 2014–2023
- [28] Sun Qingyun, Li Jianxin, Peng Hao, et al. Sugar: Subgraph neural network with reinforcement pooling and self-supervised mutual information mechanism[C] // Proc of the 31st World Wide Web Conf. New York: ACM, 2021: 2081–2091
- [29] Cui Lixin, Bai Lu, Bai Xiao, et al. Learning aligned vertex convolutional networks for graph classification[J/OL]. *IEEE Transactions on Neural Networks and Learning Systems*, 2021[2022-11-25]. <https://ieeexplore.ieee.org/abstract/document/9646437>
- [30] Zhu Yaokang, Zhang Kai, Wang Jun, et al. Structural landmarking and interaction modelling: A “SLIM” network for graph classification[C] // Proc of the 36th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2022: 9251–9259
- [31] Wijesinghe A, Wang Qing. A new perspective on “How graph neural networks go beyond Weisfeiler-Lehman?”[C/OL] // Proc of the 5th Int Conf on Learning Representations. Cambridge, MA: MIT, 2021[2022-11-25]. https://openreview.net/pdf?id=uxgg9o7bI_3



Xu Lixiang, born in 1981. PhD, professor. His research interests include machine learning, data mining, and personalized recommendation system.
徐立祥, 1981年生. 博士, 教授. 主要研究方向为机器学习、数据挖掘、个性化推荐系统。



Ge Wei, born in 1996. Master candidate. His interests include graph neural network and graph representation learning.
葛伟, 1996年生. 硕士研究生. 主要研究方向为图神经网络、图表示学习。



Chen Enhong, born in 1968. PhD, professor. His research interests include data mining, education big data analysis, and personalized recommendation system.

陈恩红, 1968 年生. 博士, 教授. 主要研究方向为数据挖掘、教育大数据分析、个性化推荐系统.



Luo Bin, born in 1963. PhD, professor. His research interests include pattern recognition, machine learning and data mining.

罗 斌, 1963 年生. 博士, 教授. 主要研究方向为模式识别、机器学习、数据挖掘.