

## 图分类研究综述\*

王兆慧<sup>1,3</sup>, 沈华伟<sup>1,3</sup>, 曹 婧<sup>1,3</sup>, 程学旗<sup>2</sup>

<sup>1</sup>(中国科学院 计算技术研究所 数据智能系统研究中心, 北京 100190)

<sup>2</sup>(中国科学院 计算技术研究所 网络数据科学与技术重点实验室, 北京 100190)

<sup>3</sup>(中国科学院大学, 北京 100049)

通信作者: 沈华伟, E-mail: [shenhuawei@ict.ac.cn](mailto:shenhuawei@ict.ac.cn)



**摘 要:** 图数据广泛存在于现实世界中, 可以自然地表示复合对象及其元素之间的复杂关联. 对图数据的分类是一个非常重要且极具挑战的问题, 在生物/化学信息学等领域有许多关键应用, 如分子属性判断, 新药发现等. 但目前尚缺乏对于图分类研究的完整综述. 首先给出图分类问题的定义和该领域的挑战; 然后梳理分析了两类图分类方法——基于图相似度计算的图分类方法和基于图神经网络的图分类方法; 接着给出了图分类方法的评价指标、常用数据集和实验结果对比; 最后介绍了图分类常见的实际应用场景, 展望了图分类领域的未来研究方向并对全文进行总结.

**关键词:** 图分类; 图核; 图卷积; 图池化; 图神经网络  
**中图法分类号:** TP18

中文引用格式: 王兆慧, 沈华伟, 曹婧, 程学旗. 图分类研究综述. 软件学报, 2022, 33(1): 171-192. <http://www.jos.org.cn/1000-9825/6323.htm>

英文引用格式: Wang ZH, Shen HW, Cao Q, Cheng XQ. Survey on Graph Classification. Ruan Jian Xue Bao/Journal of Software, 2022, 33(1): 171-192 (in Chinese). <http://www.jos.org.cn/1000-9825/6323.htm>

## Survey on Graph Classification

WANG Zhao-Hui<sup>1,3</sup>, SHEN Hua-Wei<sup>1,3</sup>, CAO Qi<sup>1,3</sup>, CHENG Xue-Qi<sup>2</sup>

<sup>1</sup>(Data Intelligence System Research Center, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China)

<sup>2</sup>(Key Laboratory of Network Data Science and Technology, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China)

<sup>3</sup>(University of Chinese Academy of Sciences, Beijing 100049, China)

**Abstract:** Graph data, as a kind of widely-existing data in the real world, naturally represent complex interactions between elements of composite objects. The classification of graph data is a very important and extremely challenging research topic. There are many key applications in the fields of bio/chemical informatics, such as molecular attribute classification and drug discovery. However, there still lacks a comprehensive review of research on graph classification. This survey first formulates the problem of graph classification and describes the main challenges of this problem; then this survey categorizes graph classification methods into similarity-based methods and graph neural network based methods. Moreover, evaluation metrics for graph classification, benchmark datasets, and comparison results are given. Finally, the application scenarios of graph classifications are summarized, and the research trends of graph classification are also discussed.

**Key words:** graph classification; graph kernel; graph convolution; graph pooling; graph neural network (GNN)

图数据 (graph data) 广泛地存在于我们的生活中, 用于表示复合对象元素之间的复杂关系. 例如社交网络, 引文网络, 生物化学网络, 交通网络等. 不同于结构规则的欧式数据, 图数据的结构复杂, 蕴含着丰富的信息. 近年来,

\* 基金项目: 国家自然科学基金 (61472400, 91746301)

收稿时间: 2020-07-09; 修改时间: 2020-09-17; 采用时间: 2021-02-08; jos 在线出版时间: 2021-04-20

对图数据的研究是学术界的一个热点. 图上的研究问题包括节点分类<sup>[1,2]</sup>, 图分类<sup>[3,4]</sup>, 链路预测<sup>[5]</sup>等, 本文主要关注图分类问题. 给定一组图, 图分类的目标是学习图和对类别标签的映射关系, 并预测未知图的类别标签. 图分类是一个重要的数据挖掘任务, 可以应用在很多领域, 例如化学信息学中, 通过对分子图进行分类来判断化合物分子的诱变性、毒性、抗癌活性等<sup>[6,7]</sup>; 生物信息学中, 通过蛋白质网络分类判断蛋白质是不是酶, 是不是具有对某种疾病的治疗能力<sup>[8,9]</sup>. 从这个角度来看, 图分类研究具有非常重要的意义.

图分类的研究方法主要包括基于图核的方法, 基于图匹配的方法和基于图深度学习的方法. 目前已有一些针对图分类领域中某类特定方法的综述, 如图核方法综述<sup>[10,11]</sup>, 图相似度学习综述<sup>[12]</sup>. 但就我们所知, 当前还没有既包括传统方法又包括近年来快速发展的深度学习方法的图分类研究综述. 为了方便更多的研究人员, 本文梳理总结了图分类的各类研究方法和这些研究之间的相互关系.

本文将现有图分类方法总结为两大类, 第 1 类是基于相似度计算的图分类方法. 基于相似度计算的图分类是通过计算成对图的相似度对图进行分类, 包括图核方法和图匹配方法. 其中, 图核方法主要通过图核的定义来计算图的相似度, 是常见的传统图分类方法. 过去多年中已经有多种基于图核的分类方法被提出<sup>[13-15]</sup>, 它们共同的思想是将图分解为某种子结构, 通过对比不同图上的子结构来计算图的相似度进而进行图分类. 基于图匹配方法的图分类方法, 则是通过考虑一些跨图的因素来计算图之间的相似度分数进而对图分类. 早期的图分类问题主要关注于图核方法, 然而这种方法不够灵活且通常计算代价较大, 图的特征提取过程和图的分类是独立进行的, 因此无法针对具体任务进行优化.

第 2 类是基于图神经网络的图分类方法. 随着深度学习在图像、文本等领域的成功, 研究人员开始关注用深度学习建模图数据. 基于深度学习的图数据建模方法也逐渐被应用于图分类问题<sup>[16-19]</sup>. 其中, 图神经网络应用于图分类问题时, 主要包括卷积算子和池化算子两个重要部分. 卷积算子利用结构和节点特征信息对图的特征进行提取, 池化算子对特征进行汇总得到整个图的表示用于分类. 本文从这两个角度对基于图卷积神经网络的图分类进行了总结分析.

尽管近期已有大量的基于图神经网络的方法应用于图分类任务, 但这个领域仍然存在许多问题和挑战, 例如领域内不同模型的实验设置不同导致的复现困难; 有些模型在特定数据集上表现较好, 但模型泛化能力有限; 此外, 图分类任务中对图结构信息的利用也是一个挑战. 本文从这个角度总结分析了图分类中存在的挑战和未来的研究方向.

本文第 1 节给出图分类问题定义并指出图分类领域中的问题和挑战. 第 2 节梳理了基于相似度计算的图分类方法, 其中包括基于图核方法的图分类和基于图匹配的图分类. 第 3 节介绍并分析了基于图神经网络的图分类方法. 第 4 节关注图分类方法的评价, 包括图分类的数据集, 评价指标和一些典型方法的效果对比分析. 第 5 节总结了图分类在各个领域的应用场景并给出未来可能的研究趋势. 最后一节总结全文.

## 1 图分类问题定义及挑战

本节我们给出图分类问题的定义和全文的符号定义, 并简要总结图分类中的问题与挑战.

### 1.1 图分类问题与符号定义

图分类是图数据挖掘的一个重要研究问题, 其目标是找到图和对类别标签的映射关系. 具体来说, 给定一组图  $\mathbf{G} = \{G_i\}_{i=1}^N$  及其标签  $\mathbf{Y} = \{Y_i\}_{i=1}^N$ , 其中每个图表示为  $G_i = (V, E, X)$ . 图分类任务是指利用图的邻接矩阵  $A$  和特征信息  $X$ , 通过机器学习的方式获取图和对类别标签的映射关系并预测未知图的标签. 为了正确预测图的类别, 需要充分地利用给定的结构信息和节点特征信息. 常见的图分类包括图的二分类及多分类问题<sup>[4]</sup>, 此时, 每个图对应一个类别标签  $Y_i = \{y_{i,j}\}_{j=1}^1$ . 除此之外, 图分类还包括多标签分类问题<sup>[20]</sup>, 多标签是指每个图对应的标签可以有一个或多个 ( $M$ ), 即  $Y_i = \{y_{i,j}\}_{j=1}^M$ .

对于图  $G = (V, E, X)$ , 图标签为  $\mathcal{Y}$ .  $V$  表示图中节点集合,  $|V| = n$  为节点数.  $E$  表示图中边的集合,  $E \subseteq V \times V$ .  $X$  表示图上节点的特征矩阵,  $X \in \mathbb{R}^{n \times D}$ .  $X_i$  表示第  $i$  个节点的特征向量,  $D$  为节点的特征维度. 此外,  $A \in \mathbb{R}^{n \times n}$  表示图的邻接矩阵, 表明图上节点之间的连边关系. 表 1 中汇总了全文中常见的符号及其表示的含义.

表 1 符号定义

符号	含义	符号	含义
$G$	图	$V$	节点集合
$\mathcal{Y}$	图的标签	$n$	节点个数
$\mathbf{G}$	图集	$N(v)$	节点 $v$ 的邻居
$\mathbf{Y}$	图的标签集合	$A \in R^{n \times n}$	邻接矩阵
$E$	连边集合	$X \in R^{n \times D}$	节点特征矩阵

## 1.2 图分类中的问题与挑战

图分类是图领域中一个极具挑战的任务,当前图分类任务上仍然存在许多问题和难点,主要包括以下几个方面.

### (1) 图数据的复杂多样性

生活中有大量的数据都可以用图这种数据结构进行表示.例如社交网络,化学分子结构,生物蛋白质结构等.每种类型的图中都包含不同的特征信息和结构信息.这种多样的信息提高了图数据的分类难度.此外,图数据是非欧空间数据,一般来说,每个图的节点数不同,图中节点连接方式不同,每个节点的邻居个数也不同.卷积、池化等在欧式数据中比较容易定义的操作,很难直接迁移到图数据上.图数据的复杂性和多样性,为图数据的分类带来非常大的挑战.

### (2) 图结构信息的有效建模

作为非欧数据,图的结构信息非常丰富.图数据中的结构信息是指图上节点之间的连接关系,包括节点的一阶连接信息,二阶信息以及高阶信息等<sup>[21]</sup>.图上机器学习的最基础挑战之一就是找到一种可以表示、编码图结构的方法,从而使得图结构信息可以被机器学习方法有效利用<sup>[22]</sup>.图的结构信息对于图分类任务也至关重要.例如,在生物信息学等领域的数据集中,图的属性标签与图上的某些结构模式有着必然的联系.然而 Errica 等人<sup>[23]</sup>在实验中发现,目前基于图神经网络的图分类方法在大部分数据集上并没能有效地利用到图的结构信息,其对于图分类的预测性能甚至不如没有建模图结构信息的方法.因此,如何有效建模并合理利用图结构信息是图分类任务面临的一大重要挑战.

### (3) 强表达能力且高效的模型构建

目前基于信息传递的图神经网络方法都与 1-WL 图同构测试有着紧密的联系. Xu 等人<sup>[24]</sup>已经证明,基于信息传递的图神经网络,其表达能力的上界就是 1-WL (Weisfeiler-Lehman) 图同构测试.近年也有一些对表达能力更强的基于高阶 WL 图同构测试的图神经网络的探索<sup>[25,26]</sup>.但总的来说, WL 测试关注的是对图是否同构的判断.一方面,对图同构的判断还未被证明可以在多项式时间内完成,通常计算复杂度较高.另一方面,在这种标准下,并不能保证表达能力强的模型,也就是对图是否同构的判断准确率高的模型,在图分类问题上也表现得好<sup>[27]</sup>.基于此,探索合适的图分类模型表达能力的判断标准非常重要,这也是对图分类本质的探索过程.如何构建一个具有强表达能力且高效的模型是图分类问题中的一个关键挑战.

## 2 基于图相似度计算的图分类

在很多用图来表示数据的领域,图之间相似度度量是关键问题之一<sup>[12]</sup>,它可以进一步处理一些下游任务,包括图分类,图聚类和相似性搜索等.本节关注利用图的相似度度量进行图分类的方法.给定一组图,基于相似度计算的图分类方法先通过图核或者图匹配的方法获得两个图之间的相似度度量,然后利用机器学习方法,根据已经得到的相似度度量对图进行分类.这类方法隐含的假设是当两个图相似度较高时,它们所属的类别也相同.这类方法的关键是对图之间相似度的计算.本节从相似度计算的角度,将基于图相似度计算的图分类分为基于图核的方法和基于图匹配的方法,分别进行介绍和分析.

### 2.1 基于图核的图分类

核方法是一类把低维空间下非线性可分问题转化为高维空间下线性可分问题的方法,其核心在于度量成对数

据的相似性. 图核 (graph kernel) 是一种特殊的核方法, 其核心在于对图之间的相似度的度量. 基于图核的图分类方法实质是先利用图核方法显式或隐式地计算图之间的相似度, 然后利用核机器通过图相似度进行图分类. 基于图核进行图分类的过程如图 1 所示.

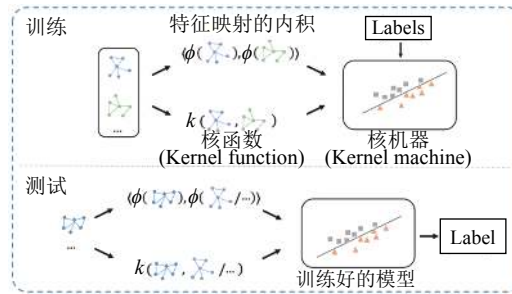


图 1 基于图核的图分类方法

图核是一个定义在图空间  $\mathcal{G}$  上的对称正定函数, 这个函数也可以表示为希尔伯特空间的一个内积. 具体来说, 当给定图  $G$  时, 存在一个图空间到希尔伯特空间的映射  $\phi: \mathcal{G} \rightarrow \mathcal{H}$ , 使得  $k(G_1, G_2) = \langle \phi(G_1), \phi(G_2) \rangle$  [28]. 可以看出, 核函数可以将高维空间的内积运算转换为低维空间的核函数计算 [29]. 这种定义导致了两种计算图核的方法 [30], 一种是显式地定义每个图的特征映射函数  $\phi$ , 然后通过特征映射后的向量内积得到图之间的相似度. 另一种是通过核函数隐式计算, 不需要知道具体的特征映射函数  $\phi$ , 直接通过核函数的定义计算得到任意两个图之间的相似度度量核矩阵 (kernel matrix), 用来连接核函数和用于分类的核机器算法. 图核方法在图分类问题中被广泛使用 [10], 其核心在于定义合适的核函数.

R-卷积核 [31] 是常见图核方法的顶层框架, 直观来看, R-卷积核的框架将复合对象 (例如如图) 按照某种分解方式分解为不同的组件 (例如子图, 子树), 然后通过对比每对组件来计算复合对象的相似度 [11]. 用于图分类的常见图核方法包括 3 类: 基于路径的图核, 基于子图的图核和基于子树的图核. 他们都基于 R-卷积核 [31] 的框架定义的, 都是 R-卷积核的实例. 这 3 类图核方法的主要区别在于图的分解方式. 基于路径, 子图和子树的图核方法分别将图分解为路径, 子图和子树, 然后利用显式或者隐式的方法对图的相似度进行计算. 一般来说, 当显式计算可行, 并且得到的特征向量维度数不是过高, 那么它通常比隐式计算更快、内存效率更高 [10].

基于路径的图核方法中, 两个图的公共路径越多, 它们越相似. 这类方法的核心在于对路径的不同定义, 主要方法有最短路径核 [13] 和随机游走核 [32,33]. 随机游走核存在计算复杂度大, 路径回溯等问题. 本文中我们主要关注最短路径核 [13], 该方法是将图分解为节点之间的最短路径, 然后对比两个图中任意对节点间的最短路径, 从而计算图的相似度. 最短路径核方法先是将原始图转化为最短路径图, 最短路径图中节点集同原图中节点集一样, 不同的是, 最短路径图中的边用两个节点间的最短路径长度标记. Borgwardt 等人 [13] 用弗洛伊德算法 (Floyd-Warshall) 得到最短路径图, 设  $G_1, G_2$  是两个输入图, 最短路径图分别是  $S_1, S_2$ . 然后在  $S_1 = (V_1, E_1)$  和  $S_2 = (V_2, E_2)$  上定义最短路径图核:

$$k_{\text{shortest\_path}}(S_1, S_2) = \sum_{e_1 \in E_1} \sum_{e_2 \in E_2} k_{\text{walk}}^{(1)}(e_1, e_2),$$

其中,  $k_{\text{walk}}^{(1)}$  是边上的半正定核, 可以被定义为狄拉克核函数, 即当  $e_1$  和  $e_2$  长度相同时其值为 1, 否则为 0. 可以看出, 相同长度的最短路径越多, 图的相似度越高. 计算最短路径图的弗洛伊德算法复杂度为  $O(n^3)$ , 考虑两个图中所有边对时, 最短路径核算法的复杂度为  $O(n^4)$ .

基于子图的图核在度量图的相似度时, 先将图转换成子图结构的组合, 再通过比较两个图中的子图分布来度量图相似度 [34]. 主要方法有循环模式核, 最优分配核等 [14,35,36]. 其中的一个经典方法是 Graphlet 图核 [14]. Graphlet 是指由  $m$  个节点组成的子图, 一般而言  $m \in \{3, 4, 5\}$ . 图 2 给出了大小为 4 的所有 Graphlet. Graphlet 图核方法通过对 Graphlet 计数的方式来对比不同的图. 定义  $G = \{\text{Graphlet}(1), \dots, \text{Graphlet}(N_k)\}$  为大小为  $N_k$  的 Graphlet 的集合,



$f_G$ 是一个长度为 $N_k$ 的向量,  $f_G$ 的第 $i$ 个值表示 $Graphlet(i)$ 在图 $G$ 中出现的频次. 给定两个大小为 $n$ 的图 $G_1, G_2$ , 其中 $n > k$ . Graphlet图核定义为:

$$k_{Graphlet}(G_1, G_2) = f_{G_1}^T f_{G_2},$$

其中, 向量 $f_{G_i}$ 可以看作是图 $G_i$ 的特征向量. 可以看出, Graphlet核值是通过两个图的特征向量的内积来显式计算得到的. 由于每个图中有 $\binom{n}{k}$ 个大小为 $k$ 的子图<sup>[11]</sup>, 计算该图的特征向量复杂度为 $O(n^k)$ , 因此, 该方法不适用于大规模的图.

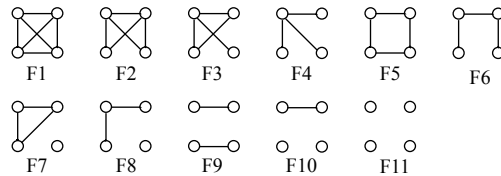


图2 大小为4的所有 Graphlets<sup>[14]</sup>

基于子树的图核度量图的相似度时, 先将图转换成树, 再通过度量树相似度来度量图相似度. 主要方法有基于子树模式的图核, 基于树模式的图核等<sup>[7,15,37]</sup>. 其中, Weisfeiler-Lehman子树核(WL子树核)<sup>[15]</sup>是使用最广泛的基于子树的图核方法. 该方法的主要思想是将图分解为子树, 通过度量子树间的相似度来获得图的相似度<sup>[34]</sup>. 它是基于1维的Weisfeiler-Lehman图同构测试<sup>[38]</sup>提出的一种快速特征提取机制. 给定多个节点标签离散的图, 先将每个节点的邻居聚合并对邻居进行排序, 节点标签和排序后的邻居标签共同组成多重集合(multiset). 然后对这些多重集合进行压缩映射, 得到对应的新标签. 接着将这些新标签赋给节点, 此时完成一次迭代, 如图3中a-d所示. 需要注意的是, 多重集合压缩和重新标记的过程是所有输入图共同进行的, 标签压缩的对应关系也是所有图共享的. 在图 $G_1, G_2$ 上迭代 $h$ 次的WL子树核定义为:

$$k_{WLsubtree}^{(h)}(G_1, G_2) = \langle \phi_{WL}^h(G_1), \phi_{WL}^h(G_2) \rangle,$$

其中,  $\phi_{WL}^h(G)$ 定义为 $h$ 次迭代中节点标签出现次数的特征向量.  $\phi_{WL}^h(G) = (c_0(G, \sigma_{01}), \dots, c_0(G, \sigma_{0|\Sigma_0|}), \dots, c_h(G, \sigma_{0h}), \dots, c_h(G, \sigma_{h|\Sigma_h|}), \dots)$ ,  $c_i(G, \sigma_{ij})$ 表示节点标签 $\sigma_{ij}$ 在图 $G$ 中出现的次数. 可以看出, WL子树核是通过两个图中共同的子树结构进行计数进而计算图相似度的. 在 $N$ 个图迭代 $h$ 次的情况下, 算法复杂度为 $O(Nhm + N^2hn)$ , 其中 $n$ 和 $m$ 表示图中最大节点数和最大边数. 图3是WL子树核在两个图上迭代计算一次的过程<sup>[15]</sup>, 其中, 经过压缩得到的每个标签都代表一个子树模式. 例如标签11, 表示一个高度为1, 根节点为4, 邻居节点为1, 1, 3和5的子树. 表2对本文中介绍的3种代表性传统图核方法进行了对比.

Yanardag等人<sup>[39]</sup>发现, 在常见的图核方法中, 用于计算核矩阵的子结构并不独立. 例如在Graphlet图核方法中, 一个Graphlet可能由另一个Graphlet添加一条边或者一个节点得到(图2中F3可以由F4增加一条边得到). 由于随着Graphlet尺寸增大, 特征空间指数级增加, 而子结构之间的依赖导致这样的特征空间中存在大量冗余. 此外, 特征空间维度的指数级增长会导致核方法中的核矩阵的对角控制问题, 也就是同一个数据集中, 一个图只与自己本身相似, 而与其他图不相似. 为了解决这两个问题, Yanardag等人<sup>[39]</sup>提出了深度图核方法:

$$k_{dgt}(G_1, G_2) = \phi(G_1)^T M \phi(G_2)$$

其中,  $M$ 是一个表示子结构之间关系的矩阵. 关于 $M$ 矩阵的计算有两种方法: (1) 在子结构之间具有清晰的数学关系时, 可以利用它们之间的相似度来计算 $M$ 矩阵; (2) 将子结构当做“单词”, 利用自然语言处理领域的单词嵌入技术学习子结构隐式表达之间的相似度矩阵 $M$ . 上文中提到的基于WL子树核方法和Graphlet图核方法都可以应用于这个深度图核的框架下并在分类效果上取得了一致性的提升.

过去几年中, 有大量的图核方法被提出并应用于图分类任务中, 图核方法结合了图的表示能力和核方法的区分能力<sup>[11]</sup>, 可以解决基于图相似度计算的图分类任务. 但这些方法中特征表示和分类的过程是分开进行的, 无法统一优化. 另外, 基于图核的方法计算复杂度通常较高, 无法应用到大规模图中.

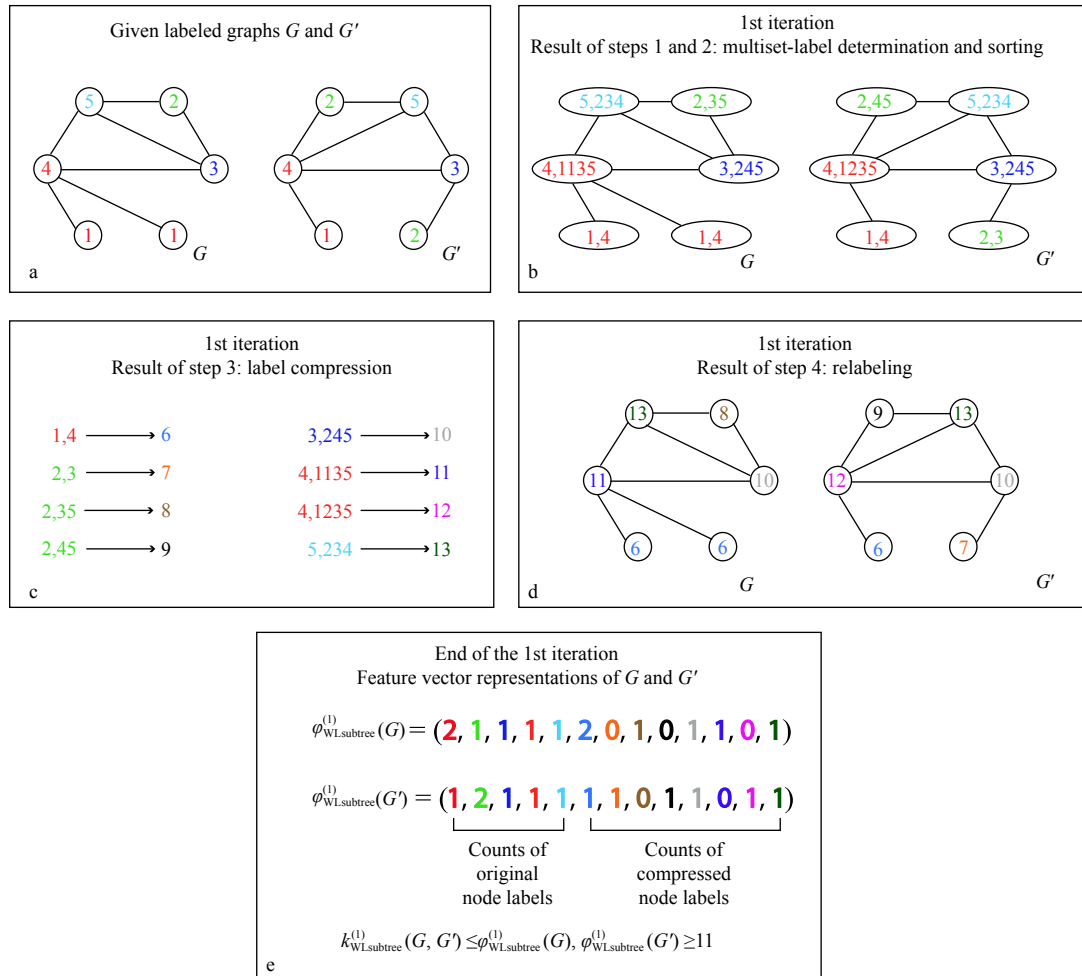


图 3 1-WL 子树核在两个图上迭代计算一次的过程

表 2 典型图核方法详细对比

类别	图核方法	代表方法	公式	计算方式	时间复杂度
R-卷积核	路径核	最短路径核	$k_{\text{shortest\_path}}(S_1, S_2) = \sum_{e_1 \in E_1} \sum_{e_2 \in E_2} k_{\text{walk}}^{(1)}(e_1, e_2)$	隐式	$O(n^4)$
	子图核	Graphlet图核	$k_{\text{Graphlet}}(G_1, G_2) = f_{G_1}^T f_{G_2}$	显式	$O(n^k)$
	子树核	WL子树核	$k_{\text{WLsubtree}}^{(h)}(G_1, G_2) = \langle \phi_{\text{WL}}^h(G_1), \phi_{\text{WL}}^h(G_2) \rangle$	显式	$O(Nhm + N^2hm)$

### 2.2 基于图匹配的图分类

图匹配是一种图相似度度量的方法. 图匹配包括精确图匹配和非精确图匹配<sup>[40]</sup>, 精确图匹配需要图之间节点的映射关系, 应用于图同构, 子图同构, 最大公共子图的判别等问题中. 非精确图匹配多数可以形式化为图编辑距离 GED (graph edit distance) 的计算. 本节主要介绍基于非精确图匹配进行图分类的方法. 这类方法的核心是先通过跨图机制对两个图之间的某种相似度度量进行计算, 例如图编辑距离, 然后通过相似度度量对图进行分类, 常用的分类器是最近邻分类器<sup>[41,42]</sup>. 基于图匹配的图分类过程如图 4 所示.

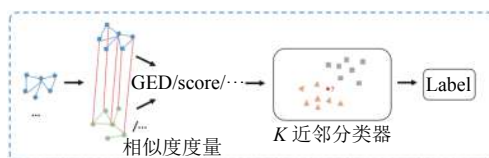


图4 基于图匹配方法的图分类过程

图匹配方法通常利用跨图节点的对比来计算两个图之间的相似度。Li 等人<sup>[43]</sup>首先提出图匹配网络模型 (GMN), 该模型计算图之间相似度时不仅考虑每个图内节点信息, 也考虑跨图节点的信息。输入一对图时, GMN 同时学习两个图的表示, 通过跨图的注意力机制计算跨图节点的相似性, 图的表示不仅依赖于本图中的邻居节点也依赖于另一个图中的节点。GMN 可以保证当两个图匹配度较高时, 学得表示也更相似, 相似度分数更高。类似地, Bai 等人<sup>[44]</sup>提出了基于卷积神经网络的图相似度计算模型 (GSimCNN)。不同于 GMN, GSimCNN 设计了图交互层计算两个图中每对节点的相似度并形成图相似度矩阵。接着用多个卷积神经网络层对相似度矩阵进行特征提取, 计算得到两图的相似度分数。在 GSimCNN 的基础上, Bai 等人<sup>[45]</sup>提出了快速计算图相似度的模型 SimGNN。该模型不仅计算了跨图的节点相似度向量, 而且计算了两个图表示间的相似度向量, 进而利用这两个相似度向量计算得到两个图的相似度分数。虽然这些基于深度学习的图匹配方法得到图的表示或者相似度分数后, 没有显式地进行后续的图分类任务, 但他们本质与图分类任务非常接近, 图的表示和图相似度分数都可以用来完成图分类任务<sup>[41]</sup>, 因此本文将它们归于基于图相似度计算的图分类中。

基于图匹配的图相似度计算方法常常被用于计算机安全领域中二分函数相似度搜索, 检测代码中是否含有已知不稳定结构。也常被用于生物化学领域中, 利用图相似度分数判断化合物属性。具体应用方法详见第5节。

### 3 基于图神经网络的图分类

前文介绍的图核方法很多年来都是图分类中的主导方法, 也取得了不错的分类效果<sup>[25]</sup>。但由于这些方法通常依赖于固定特征, 其特征表示难以有效地适应于新的数据分布。随着图深度学习的发展<sup>[46]</sup>, 一些神经网络方法开始用于解决图分类任务。本节重点关注基于图神经网络的图分类方法, 这类方法通过端到端的方式进行模型的优化学习, 为图分类的准确率带来了较大的提升。

应用于图像分类任务的传统卷积神经网络, 主要包括卷积和池化两个操作, 这两个操作依赖于图像数据的结构规则性和平移不变性。类比于图像分类任务, 图卷积神经网络应用于图分类问题时, 同样需要关注卷积和池化算子。但不同于图像数据, 图数据是非欧空间数据, 同一个数据集中的每个图大小不同, 结构不一。图中的每个节点也具有不同的局部结构, 为图分类中卷积算子和池化算子的设计带来了巨大的挑战。给定一组图, 基于图神经网络的图分类方法通常先通过卷积的方式对这些图进行多次特征变换, 然后在此基础上进行池化操作, 将图的规模缩小。这个过程可以重复多次, 最终得到整个图的表示, 从而进行分类。本节就从图分类任务中的卷积算子和池化算子角度, 对基于图神经网络的图分类方法进行总结和分析。利用图神经网络进行图分类的过程如图5所示。其中, 可选的操作和模块用虚线表示。环形箭头表示操作可以选择重复  $1-n$  次。

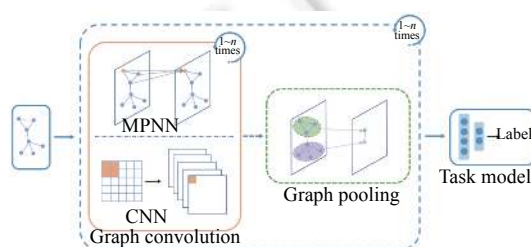


图5 基于图神经网络的图分类过程

### 3.1 卷 积

在图级别的任务中,图卷积操作主要是为了对图进行特征变换和特征提取.在这个过程中需要尽可能多地利用图中节点特征和拓扑信息.当前在图分类任务中,主要的卷积方式有两种,一种是信息传递(massage-passing)式的卷积,即直接在原始图结构中定义由邻居聚合和迭代更新机制所组成的卷积算子.例如图卷积神经网络 GCN,基于注意力机制的图卷积神经网络 GAT 等.另一种是传统卷积神经网络(CNN)式的卷积,先将非欧图转化为规则网格结构,再应用传统卷积神经网络直接进行卷积操作.

#### (1) 信息传递式卷积

Gilmer 等人<sup>[47]</sup>提出了信息传递式(MPNN)的通用框架,将卷积过程形式化为信息传递和节点信息更新两个函数.信息传递是指各个节点将自己邻居的信息聚合到自身节点,节点信息更新是将该节点上一层的节点表示与聚合后的邻居信息进行结合的过程.本文把符合这种框架的卷积方式称为信息传递式(MPNN)卷积.早期的信息传递式卷积主要用于节点分类任务,取得了很好的分类效果.在图分类任务中,信息传递式卷积用来进行特征变换,得到图上节点的表示.以 Kipf 等人<sup>[48]</sup>提出的图卷积网络 GCN 为例:

$$H^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}H^{(l)}W^{(l)}\right),$$

其中,  $H^{(l)}$  是  $l$  层中节点的表示矩阵,  $W^{(l)}$  是可训练的参数矩阵,  $\tilde{A} = A + I_N$  表示包含自连接的邻接矩阵,  $\sigma$  表示非线性激活函数.

采用 MPNN 式卷积的图分类模型通常先利用节点的特征和图的结构求得特征映射后的节点表示,然后利用池化机制得到图的表示并进行分类. Ma 等人<sup>[16]</sup>提出的谱池化(EigenPool)图分类模型, Gao 等人<sup>[49]</sup>提出的 Graph U-net 模型,和 Lee 等人<sup>[19]</sup>提出的自注意力池化(SAG)模型中均采用信息传递式卷积中的 GCN 进行节点表示学习.此外, Zhang 等人<sup>[4]</sup>提出的深度图卷积网络 DGCNN 和 Yuan 等人<sup>[50]</sup>提出的结构池化图分类模型 StructurePool 也采用了类似于 GCN 的信息传递式卷积,具体卷积形式为  $Z = f(\tilde{D}^{-1}\tilde{A}XW)$ , 其中,  $\tilde{A}$  和 GCN 中定义相同,  $\tilde{D}$  是度矩阵, 其中  $\tilde{D}_{ii} = \sum_j \tilde{A}_{ij}$ .  $W \in R^{c \times c'}$  是可训练参数矩阵. Zhang 等人<sup>[4]</sup>证明了这种卷积方式理论上比 GCN 更近似于 WL 算法. Ying 等人<sup>[17]</sup>提出的可微分池化分类模型 DiffPool 中,卷积部分可以采用任意 MPNN 式卷积.

以上提到的各模型中的 MPNN 式卷积,都是基于第 2.1 节中介绍的 1-WL 算法提出的.但 Xu 等人<sup>[24]</sup>指出,基于 1-WL 的任何信息传递式的图卷积神经网络,其表达能力都不会强于 1-WL.而 1-WL 本身的表达能力有限,对很多非同构图并不能准确地分类.也就是说,当 1-WL 算法将两个非同构图判断为“可能同构”时,信息传递式的图卷积网络输出的两个图的表示是相同的.这样的弱区分能力,对图分类的效果会产生负面的影响.近年来,有研究者提出了一些基于 k-WL 算法的图神经网络<sup>[25,26]</sup>.k-WL 算法是一种基于带标签子图的图同构判断算法.在基于 1-WL 的 GNNs 中,信息传递的基本单元是节点.不同于此,k-WL 算法中信息传递的基本单元是大小为  $k$  的子图结构.其中,基于集合的 k-WL 算法 set k-WL 表达能力弱于 k-WL 算法,计算复杂度低于 k-WL<sup>[21,25]</sup>.考虑到内存的限制, Morris 等人<sup>[25]</sup>提出基于 set k-WL 算法的 k-GNNs 模型,该模型的关键是信息传递在大小为  $k$  的节点集合对应的子图结构之间进行.k-GNNs 为每个大小为  $k$  的集合  $s \in [V(G)]^k$  按照如下方式更新他们的特征向量:

$$f_k^{(t)}(s) = \sigma\left(f_k^{(t-1)}(s) \cdot W_1^{(t)} + \sum_{u \in N_t(s) \cup N_G(s)} f_k^{(t-1)}(u) \cdot W_2^{(t)}\right),$$

其中,  $N(s) = \{t \in [V(G)]^k \mid |s \cap t| = k - 1\}$ , 表示集合  $s$  与邻居集合  $t$  的交集元素个数为  $k - 1$ .k-GNNs 的表达能力和 set k-WL 算法一样强大.该模型需要枚举大小为  $k$  的子图,由于计算复杂度较高,只考虑了  $k$  为 1, 2, 3 的情况,并提出一种层次化的 1-2-3GNN 网络.该模型在图分类任务中,取得了一致性好于 1-GNN 的效果. Maron 等人<sup>[26]</sup>提出通过矩阵相乘实现一种和 3-WL 有同样区分能力的 GNN 模型,比 k-GNNs<sup>[25]</sup>具有更低的空间复杂度和更容易部署的优点.

目前,图卷积神经网络模型的表达能力都是通过对图是否同构的判断能力来衡量的<sup>[24,51]</sup>,那么基于 k-WL 算法的图卷积神经网络的表达能力显然强于基于 1-WL 算法的图卷积神经网络.然而,基于 k-WL 算法的图卷积神经网络<sup>[25,26]</sup>需要对带标签的子图结构进行枚举,类似于对带标签信息的 Graphlet 的枚举,因此带来了非常大的计算代价.实际应用中,我们仍需在表达能力和计算消耗之间进行权衡.



## (2) CNN 式卷积

图分类任务中,还有一种卷积方式,即先把图数据按照一定的规则变换为欧式数据,然后按照传统卷积网络的方式进行卷积变换.本文把他们称做 CNN 式卷积.相比 MPNN 式卷积,这类卷积在目前的图分类中应用不多,但取得了较好的分类效果.类比图像分类任务中的 CNN,图分类中的 CNN 式卷积网络,可能也具备抽取图中关键结构的能力.Niepert 等人<sup>[52]</sup>指出,由于图像中像素存在隐式的空间顺序,如图 6(a),因此可以为节点按照顺序从左到右从上到下生成邻居图,也就是感受野.接着,3×3 的感受野读到的值会被转化为序列然后送入卷积层中,如图 6(b)所示.然而,图数据中每个节点结构并不一致,不存在这样统一的空间顺序.

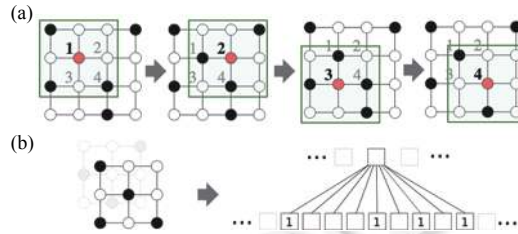


图 6 感受野为 3×3 的 CNN 框架<sup>[52]</sup>

Niepert 等人<sup>[52]</sup>率先提出了 PSCN 模型将非欧图数据转换为欧式数据,利用节点选择的方法解决了不同图数据中节点数不同的问题.具体地,用图标记算法按照某种排序标准对节点排序,使得不同图中结构相似的节点可以处在序列中相似的位置,然后选出指定个数的节点.接着,PSCN 利用图标准化解决了图中每个节点的邻居个数不同导致的节点感受野定义问题.具体地,该模型先用宽度优先的方式聚合邻居并进行排序,将无序图空间映射到具有线性顺序的向量空间,接着在该序列中选择指定个数的节点作为中心节点的感受野.由于最优的图标准化是一个非确定性多项式难(NP-hard)问题,PSCN 并未解决这个问题,而是在众多图标记方法中选择了效果较好的方法并利用了辅助工具<sup>[53]</sup>对邻居图中节点进行定序.得到统一的图表示后,用一维卷积层进行卷积,其中,卷积核大小与节点感受野大小相同.其余卷积和分类部分结构则可以任意选定.

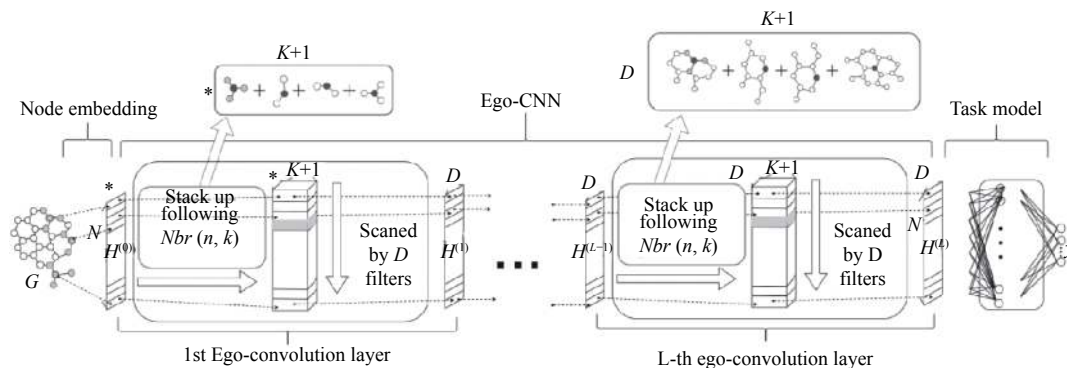
由于 PSCN 模型为了解决不同图节点数不同的问题进行了节点的选择,因此对于节点的卷积过程无法进行层堆叠,只能抽取到一些局部的关键结构.直观的解释就是,如果进行第 2 次卷积,节点的感受野中的邻居节点有可能在第一次卷积前的节点选择中未被选中,因此没有这些节点的表达,无法进行第 2 次卷积,无法抽取一些全局的重要结构.针对这个问题,Tzeng 等人<sup>[54]</sup>提出了以自我为中心的卷积网络 Ego-CNN,可以发现图中的局部和全局关键结构.如图 7 是 Ego-CNN 的模型框架图.

不同于 PSCN,以自我为中心的卷积网络 Ego-CNN 初始并没有进行节点的选择,而是将输入图中的所有节点经过一个网络嵌入算法得到图节点的初始表达.然后为每个节点选择指定个数的邻居并将邻居节点的表达拼接起来作为感受野.接着用与感受野尺寸相同的卷积核进行多层卷积.卷积过程可以公式化为:

$$H_{n,d}^{(l)} = \sigma(E^{(n,l)} \circ W^{(l,d)} + b_d^{(l)}),$$

$$E^{(n,l)} = [H_{n,:}^{(l-1)}, H_{Nbr(n,1),:}^{(l-1)}, \dots, H_{Nbr(n,K),:}^{(l-1)}]^T,$$

其中, $E^{(n,l)}$ 是节点 $n$ 在 $l$ 层邻居的矩阵表示,由节点 $n$ 及节点 $n$ 的 $K$ 个最近邻节点在上一层的表示连接形成, $Nbr(n,K)$ 表示节点 $n$ 的第 $K$ 阶最近邻居,它们是在模型开始前就确定好并且保持不变. $\sigma$ 是激活函数, $b_d$ 是偏置项. $\circ$ 是 Frobenius 内积,定义为 $X \circ Y = \sum_{ij} X_{ij} Y_{ij}$ .与 PSCN 相比,它们的卷积方式相同,主要的不同在于 EgoCNN 中节点 $k$ 阶邻居的固定定义方式和多次卷积的堆叠. $H_{Nbr(n,K),:}^{(l-1)}$ 中已经包含了上一层的节点表示,由于没有进行节点的选择,因此聚合 $K$ 个最近邻时,没有 PSCN 中邻居节点表示可能不存在的问题.对 $E^{(n,l)}$ 进行卷积时,可以重复地利用到上一层节点表示,因此, $l$ 层的 Ego-CNN 可以在以节点为中心 $l$ -跳的网络内发现重要结构,并可以通过反卷积的方式将关键结构可视化.

图 7 以自我为中心的卷积 Ego-CNN 模型框架<sup>[54]</sup>

PSCN 和 EgoCNN 都是在节点上进行卷积, 在卷积的过程中利用结构信息, 不同于此, Peng 等人<sup>[21]</sup>提出的基于 motif 的子图注意力卷积网络 MA-GCNN 在将图表示为矩阵的过程中, 就充分利用了图的结构信息. 具体地, 在对图数据进行预处理时, 该模型先按照某种选择策略选出  $N$  个节点, 并为每个选出的节点选择  $K$  个重要的邻居节点形成  $N$  个子图. 然后将这些子图转换为基于 motif 匹配的矩阵表示, 这个矩阵里的基本单元是一个 2-跳路径的 motif, 经过匹配的多个 motif 形成子图表示, 多个子图的表示连接形成整个图的表示. 可以看出, 该模型在初始表示中用 motif 匹配的方式保留子图上的连边信息, 也就是子图结构信息. 基于此, MA-GCNN 对这个矩阵进行 2 次感受野大小不同的卷积. 然后利用子图级别的注意力机制, 得到由重要子结构组成的图的表示进而进行分类. 直觉上来说, 在图分类中, 子图级别的注意力机制更具有可解释性. 通过验证表明, 基于子图的卷积取得了更好的图分类效果, 这也证明了该模型中对图数据的预处理过程更好地捕捉到了空间结构信息. 但与 PSCN 类似, MA-GCNN 也对节点进行了排序和选择, 节点选择的过程相当于一次池化操作, 直接地丢掉了图中部分结构信息, 节点排序操作则使得后续卷积得到的图上重要结构可能是这种启发式排序规则下的重要结构, 而不是图分类任务所需要的重要结构.

对比这两类卷积方式, MPNN 式的卷积假设相邻的节点标签相同, 因此采用了邻居信息传递机制, 这种机制从节点的角度出发, 更适合于图上节点级别的分类问题. CNN 式卷积则具有发现关键结构的能力, 这对于生物化学信息学等领域的实际应用具有重要的意义, 但当前的 CNN 式卷积中也存在参数过多, 无法准确学习得到所有重要结构的问题. 到目前为止, 关于图分类中适合的卷积方式并无定论, 我们还不知道什么样的卷积方式可以最优地利用图中的结构信息帮助分类, 接下来仍然需要探索.

### 3.2 池化

传统卷积网络中的池化也被称为下采样, 可以降低特征维度, 减少参数量并且扩大感受野. 作为欧式数据, 图像数据中不同像素点的结构统一, 因此传统卷积神经网络中的池化算子考虑结构信息时的方式会比较简单直观, 主要有最大池化、加和池化和均值池化等方法. 在图分类中, 池化操作主要是配合某些类型的卷积操作, 降低表示的维度, 得到整个图的表示. 不同于图像数据, 在图数据中, 每个节点的局部结构不同, 在对图数据节点表示进行池化时, 需要考虑每个节点不同的局部结构信息.

池化是图粗化的过程, 把相似的节点聚类在一起, 因此需要在图上定义有意义的邻居<sup>[55]</sup>. 池化一般包括两个步骤, 第 1 步求分配矩阵, 分配矩阵决定哪些邻居聚类在一起形成聚类簇. 分配矩阵  $S^{(l)} \in \mathbb{R}^{n_l \times n_{l+1}}$  是第  $l$  层节点分配到  $l+1$  层不同的聚类簇中的指示矩阵. 矩阵的行表示节点, 共有  $n_l$  个. 列表示粗化后的聚类簇, 共有  $n_{l+1}$  个. 第 2 步按照分配矩阵进行聚合, 即选定聚类簇中节点的聚合方式并将每个聚类簇聚为一个新的超节点, 所有的超节点构成新的图. 通常在最直接的聚合方式下, 新图的特征矩阵由  $X^{(l+1)} = S^{(l)T} X^{(l)}$  得到, 邻接矩阵由  $A^{(l+1)} = S^{(l)T} A^{(l)} S^{(l)}$  得到. 当然, 很多模型会提出新的更复杂的聚合方式<sup>[16]</sup>. 本小节将从分配矩阵的角度对图分类领域池化方法进行介绍, 并对具体的分配和聚合方式进行介绍分析.

### (1) 节点硬分配与聚合

节点硬分配的方式是将图上的节点分为若干个无重叠的簇. 具体地, 硬分配矩阵是指从行的角度来说, 每一行有一个值为 1, 其他值均为 0 的分配矩阵. 意味着每个节点属于且仅属于一个聚类簇. Ma 等人<sup>[16]</sup>提出的谱池化 (EigenPooling) 利用谱聚类得到硬分配矩阵  $S^{(l)}$ , 并根据该矩阵进行子图划分. EigenPooling 将整个图划分为一些没有重叠节点的子图 (聚类簇), 并把每个子图当做一个超节点. 其中超节点之间的邻接矩阵通过  $A^{(l+1)} = S^{(l)T}A^{(l)}S^{(l)}$  得到. 子图中节点进行聚合时, EigenPooling 并不是直接把分配矩阵  $S^{(l)}$  当做池化算子, 而是将子图特征分解后得到的特征向量组合形成多个池化算子. 接着, 超节点的表示可以利用池化算子  $\Theta_l$  通过  $X_l = \Theta_l^T X$  得到, 多个池化算子得到的多组超节点表示并进行拼接后, 得到最终的超节点表示. 从局域角度来看, 这样的池化过程是子图上图信号的傅里叶变换过程; 从全局的角度来看, 这样的池化算子相当于滤波器, 池化的过程就是对图上信号进行滤波, 这些池化算子保留了很好的性质, 如可以完好恢复图上信号, 最大程度保留图上信息. 图 8 展示了利用信息传递式的图卷积算子和 EigenPooling 的池化算子进行图分类的过程.

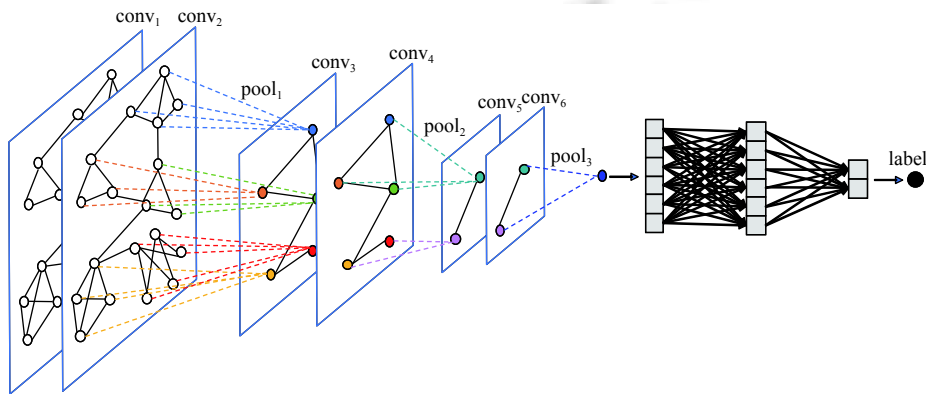


图 8 利用图卷积算子和谱池化算子进行图分类的框架<sup>[16]</sup>

Wu 等人<sup>[3]</sup>则认为度相同的节点可能有相似的结构信息, 为了保留度相关的节点结构信息, 他们提出了度特定的图神经网络模型 (DEMO-Net). 该模型在池化过程中, 将度相同的节点聚合为一个新的超节点并采用了表示拼接的聚合方式, 也属于基于硬分配矩阵的池化. Diehl 等人<sup>[56]</sup>提出的 EdgePool 通过学习边的分数来决定哪两个节点聚合在一起. 具体地, 在进行池化时, EdgePool 先为边打分并排序, 然后选择分数最高的边将其两端的节点聚合在一起形成新的虚拟节点. EdgePool 中没有显式给出分配矩阵, 但这个过程属于节点硬分配的池化. 此外, 值得注意的是, 一些简单的置换不变 (permutation invariant) 函数 (例如加和池化、最大池化等), 也属于特殊的基于节点硬分配的池化方式.

基于节点硬分配的池化模型中, 在得到节点表示后, 一般按照图聚类<sup>[16]</sup>或者特定规则<sup>[3]</sup>将节点聚合为簇, 每个节点属于且仅属于一个簇. 这种池化方式得到的图表示很大程度上依赖于节点聚合规则, 在某些场景下, 可能存在节点聚合规则与分类任务适配不够好的情况, 从而影响图分类的效果.

### (2) 节点软分配与聚合

节点软分配的方式是将图上的节点分为若干个可能有重叠的簇, 也就是每个节点不仅属于一个簇. 具体地, 软分配矩阵是指从行的角度来看, 行中每个值的取值范围为  $0 \leq p_{ij} < 1$ , 表示节点  $i$  以概率  $p_{ij}$  属于第  $j$  个簇. Ying 等人<sup>[17]</sup>提出的可微分池化 (DiffPool) 模型就是通过参数学习的方式得到软聚类分配矩阵  $S^{(l)}$ . 模型框架如图 9 所示. 模型中包含两个图神经网络, 一个嵌入图神经网络用来学习节点在  $l$  层的表示:

$$Z^{(l)} = GNN_{l, \text{embed}}(A^{(l)}, X^{(l)}),$$

另一个池化图神经网络用来学习软分配矩阵:

$$S^{(l)} = \text{Softmax}(GNN_{l, \text{pool}}(A^{(l)}, X^{(l)})).$$



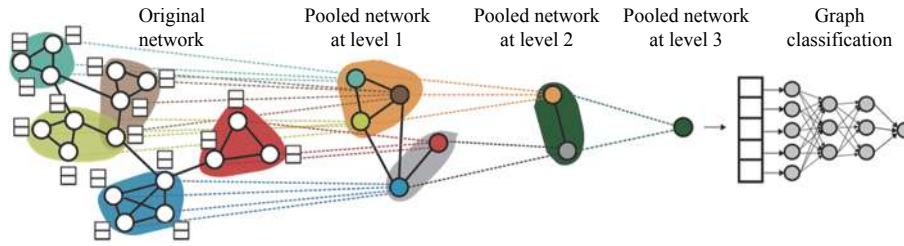


图9 利用图卷积算子和可微分池化算子进行图分类的框架<sup>[17]</sup>

得到  $l$  层的节点表示和软分配矩阵后, 通过  $X^{(l+1)} = S^{(l)T} X^{(l)}$  和  $A^{(l+1)} = S^{(l)T} A^{(l)} S^{(l)}$  分别得到粗化图的节点表示和邻接矩阵. 该模型通过学习的方式得到软分配矩阵, 仅使用图分类任务返回的梯度信号很难训练好可微分池化图神经网络, 因此很难学习到好的软分配矩阵. 为了克服这个问题, 该模型在训练过程中加入了一个辅助的链路预测任务, 假设邻近节点应该被池化在一起. 此外, 为确保每个聚类簇中的元素是确定的, 该模型还对学得的软分配矩阵进行限制, 使得每一行尽量靠近 one-hot 向量 (每行只有一个 1, 其他都是 0 的向量). 可微分池化模型第一次提出通过学习的方式利用结构信息和特征信息得到分配矩阵, 但实际应用中学习到好的分配矩阵难度较大, 且由于软分配矩阵比较稠密, 导致模型计算复杂度较高.

DiffPool 的目标是学习得到一个软分配矩阵进行图分类, 不同于此, Amir 等人<sup>[18]</sup>提出了基于记忆层的图神经网络, 目标是通过学习的方式得到多种分配方式, 然后利用卷积将多个分配矩阵聚合为一个最终的分配矩阵用于池化. 这种学习多个分配矩阵的方式有效地提升了模型能力, 取得了较好的分类效果.

对比前文介绍的基于节点硬分配的池化方法, 基于节点软分配的池化不要求每个节点唯一的属于一个簇, 这种宽松的限制更符合实际场景中的情况, 每个节点可能同时在多个簇中扮演角色. 此外, 在目前的基于节点软分配的池化方法中, 通常都采用了学习的方式来得到分配矩阵<sup>[17,18]</sup>, 通过端到端优化在任务依赖的场景中取得了较好的效果.

### (3) 节点选择与聚合

基于节点选择的池化方法是指从图中选择一些节点来代表整个图, 其中被选出的每个节点都代表一个簇. 一般来说, 节点选择型分配矩阵在每一列中有一个值为 1, 其他均为 0. 也就是在  $n_l$  个节点中选择  $n_{l+1}$  个节点作为粗化后图节点. 这类方法的核心在于设计一种为节点打分和选择节点的方式. Lee 等人<sup>[19]</sup>首先提出了基于自我注意力的池化 (SAGPool) 模型, 该模型在进行池化时, 同时考虑了节点特征和图拓扑信息. 具体地, SAGPool 利用图卷积网络为每个节点学得一个自注意力分数, 这个分数表示节点在图中的重要性:

$$Z = \sigma\left(\bar{D}^{-\frac{1}{2}} \bar{A} \bar{D}^{-\frac{1}{2}} X \Theta_{att}\right),$$

其中,  $Z \in R^{N \times 1}$ ,  $\bar{A}$  的定义和 GCN 中相同,  $\Theta_{att} \in R^{F \times 1}$  是 SAGPool 模型中唯一的参数. 得到节点重要性分数后降序排列, 按照排序结果保留一定比例的重要节点, 当作粗化图的节点:

$$idx = top\_rank(Z, \lceil kN \rceil), Z_{mask} = Z_{idx},$$

其中,  $top\_rank$  函数返回值最大的  $\lceil kN \rceil$  个节点的标号  $idx$ , 在进行池化时, 保留被选择的重要节点的节点特征和它们之间的连接方式:

$$X' = X_{idx}, X_{out} = X' \odot Z_{mask}, A_{out} = A_{idx, idx},$$

其中,  $X_{out}$  和  $A_{out}$  是池化后图的节点特征矩阵和邻接矩阵. 图 10 给出 SAGPool 的结构示意图<sup>[19]</sup>.

Gao 等人<sup>[49]</sup>提出了 GraphU-Net 模型. 借鉴图像中的 U-Net 模型思想, 该模型包括图上的池化算子 gPool 和上采样算子 gUnpool. 其中, 池化算子 gPool 也采用了节点选择的方式进行, 与 SAGPool 的主要不同在于节点的选择策略. 具体地, 该模型利用一个可训练的映射向量  $p$ , 把所有的节点特征映射到一个 1 维标量, 这个标量大小代表节点中可以保留下来的信息的多少. 为了尽可能多地保留原图中的信息, 模型选择了那些标量值较大的对应节点来组成新图. 计算过程如下:



$$y = \frac{X^l p^l}{\|p^l\|}, \text{idx} = \text{rank}(y, k), \bar{y} = \text{Sigmoid}(y(\text{idx})),$$

其中,  $y$  向量的值表示每个节点在  $p$  上可以保留的信息多少.  $\bar{y}$  中提取出了被选择节点中可以保留信息比例的值.  $\text{idx}$  表示被选择出的节点标号. 选出节点后, 用类似 SAGPool 的方式, 保留被选择节点的连接方式作为新图的邻接矩阵, 并按照比例保留节点特征作为新图的特征矩阵. Cangea 等人<sup>[57]</sup>也利用了 GraphU-Net 的节点选择策略进行图池化. Ranjan 等人<sup>[58]</sup>为了解决: (1) DiffPool 不能应用于大图; (2) GraphU-Net 等池化方法对结构信息的利用不够充分的问题, 提出了自适应的结构感知池化方法 (ASAP), 该方法先将每个节点当做聚类中心进行局部聚类形成簇, 然后提出一种卷积方法对这些簇打分并排序, 选择分数最高的  $[kN]$  个簇形成池化后的图. 实际上节点聚类形成簇的过程中, 簇的个数等于节点个数, 簇的表示相当于加入邻居信息的节点表示. 因此, 整体上来看, ASAP 的池化过程属于基于节点选择的池化.

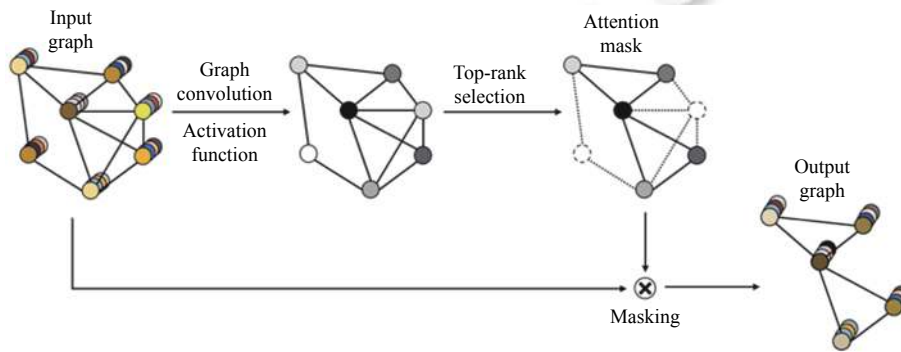


图 10 自注意力池化层示意图

这些基于节点选择的图池化方法中存在一些问题: 第一, 节点选择策略的角度单一, 缺少目标性; 第二, 未被选中的节点特征直接被丢弃, 特征信息未利用充分. 针对这些问题, Zhang 等人<sup>[59]</sup>提出了一种基于结构和特征的图自适应池化方法. 该方法提出了新的节点选择策略, 包括两个分别从结构, 特征方面对节点的重要性进行排序的模块和一个综合结构和特征信息的排序模块. 除此之外, 该模型在节点选择之前, 对所有节点信息进行聚合, 使得池化后的图中节点不仅包含被选中的节点信息, 也包含被丢弃节点的信息. 这个模型在一些生物数据集上取得了较好的分类效果. 不同于其他方法, Zhang 等人<sup>[60]</sup>提出的结构学习的层次图池化方法中, 给出了意义明确的节点选择策略, 他们认为, 当一个节点可以由它的邻居节点们计算得到时, 意味着该节点在池化图中可以被删掉, 并且不会带来信息损失. 这个非参的池化算子在多个图分类数据集上取得了较好的分类效果, 证明了这种节点选择策略的有效性.

在目前的分类体系中, 我们将池化方法分成了基于节点硬分配, 基于节点软分配和基于节点选择的池化. 但从另一个角度来看, 本文中提到的所有图上的池化方法又可以分为两类, 一类是全局池化, 得到节点表示后, 做一次池化得到整个图的表示, 如 DGCNN<sup>[4]</sup>. 另一类是层次池化, 在每个卷积层后面, 进行不同规模的池化以缩小图的尺寸, 最终得到整个图的表示, 如 DiffPool<sup>[17]</sup>, SAGPool<sup>[19]</sup>. 由于卷积中节点沿着图的边进行信息传递, 得到的节点表示是平的, 通过层次池化, 可以得到层次化的图的表示. 层次池化方法的效果在很多模型中得到了验证<sup>[16-18]</sup>. 表 3 对当前基于图神经网络的图分类方法进行总结和对比.

#### 4 图分类方法评价

为了便于对图分类方法的效果进行对比分析, 本节首先介绍图分类领域常用数据集和评价指标, 然后对一些代表性图分类方法的实验结果进行对比分析.

#### 4.1 常用数据集

目前图分类领域常用的数据集主要包括用于二分类/多分类的生物蛋白质数据集, 化学化合物数据集和社交网络类数据集, 以及用于多标签图分类的气味数据集. 数据集相关统计信息见表 4.

表 3 基于图神经网络的图分类方法总结对比

卷积	池化	方法名称	池化复杂度	优缺点
MPNN	基于硬划分的层次池化	EigenPool <sup>[16]</sup>	—	聚合时考虑子图结构; 子图上谱分解复杂度较大
		DEMO-Net <sup>[3]</sup>	$O( V + E )$	度数特定的图卷积; 度数特定池化
	基于软划分的层次池化	DiffPool <sup>[17]</sup>	$O(k V ^2)$	可微分池化得到的分配矩阵比较稠密, 模型不易优化
	基于节点选择的层次池化	SAGPool <sup>[19]</sup>	$O( V + E )$	SAGPool中节点分数通过GCN学习得到
Graphu-net <sup>[49]</sup>		$O( V + E )$	gPool中节点分数通过映射向量学习得到	
CNN	基于节点选择的全局池化	DGCNN <sup>[4]</sup>	—	利用学习得到的表示排序并进行节点选择式池化, 在等大小的图上进行CNN卷积
		PSCN <sup>[52]</sup>	—	通过节点选择和规则排序得到等大小子图表示, 在子图上进行CNN卷积
		MA-GCNN <sup>[21]</sup>	—	子图的表示中包含图上连边关系; 子图表示基于节点排序
	—	EgoCNN <sup>[54]</sup>	—	包含所有节点为中心的子图表示, 通过规则排序, 形成等大小子图表示, 接着进行CNN卷积

表 4 图分类领域常用数据集

任务	领域	数据集	Graph数	Graph类别	正例	负例	平均节点数	节点类别	节点属性数
二分类/多分类	化学化合物	MUTAG	188	2	125	63	17.7	7	—
		Mutagenicity	4337	2	2401	1936	30.3	13	—
		NCII	4110	2	2057	2053	29.9	37	—
		NCII09	4127	2	2079	2048	29.7	38	—
	生物蛋白质	ENZYMES	600	6	100/类		32.6	3	18
		PROTEINS	1113	2	663	450	39.1	3	1
		D&D	1178	2	691	487	284.3	89	—
	社交网络	COLLAB	5000	3	2600/775/1625		74.5	—	—
		REDDIT-BINARY	2000	2	1000	1000	429.6	—	—
		IMDB-BINARY	1000	2	500	500	19.8	—	—
多标签分类	领域	数据集	Graph数	Graph标签个数		数据集中气味属性总数			
	计算机嗅觉	GoodScents	3786	1-15		580			
		Leffingwell PMP 2001	3561	$\geq 1$		—			
		Flavornet	738	$\geq 1$		197			

化学化合物数据集中, 通常每个图表示一个化合物, 图中节点表示原子, 边表示原子之间真实存在的化学键.

(1) MUTAG 数据集<sup>[61]</sup>由 188 个化学化合物结构图组成, 根据它们对细菌的诱变作用分为 2 个类别. 图中的节点表示原子, 节点标签表示原子种类, 包括 C, N, O, F, I, Cl, Br. 边表示化学键, 边标签包括芳香键, 单键, 双键和三键.

(2) Mutagenicity<sup>[62]</sup>是诱变剂化合物数据集, 图的标签分为诱变和非诱变 2 类. 图中节点表示原子, 节点标签表示原子种类, 包括 C, O, Cl, H, N, F, Br, S, P, I, Na, K, Li, Ca. 有 3 个种类的边分别表示不同化学键.

(3) NCII<sup>[63]</sup>是对非小细胞肺癌活性筛选的化合物数据集, 图的标签分为 2 类, 表示具有/不具有抗癌活性. 共包含 4110 个化合物图.

(4) NCII09<sup>[63]</sup>是对非小细胞卵巢癌活性筛选的化合物数据集, 图的标签分为 2 类, 表示具有/不具有抗癌活性.

包含 4127 个化合物图。

生物蛋白质数据集中,通常每个图表示一个蛋白质的高级结构,图中的节点表示一个氨基酸分子,边表示氨基酸分子之间的结构邻近性,当氨基酸之间距离小于某个阈值时,节点之间有边存在。

(1) ENZYMES<sup>[64]</sup>是一个蛋白质三级结构数据集,共包含 600 个酶,每个图表示一个蛋白质结构,图标签是 6 种酶的等级类别。

(2) PROTEINS 数据集<sup>[64]</sup>中的 1113 个图表示蛋白质,图的标签分为 2 类,表示酶或者非酶。节点是蛋白质的二级结构,如果二级结构在氨基酸序列或者蛋白质三维空间中是邻居,那么节点之间有边存在。

(3) D & D 数据集<sup>[65]</sup>由 1178 个蛋白质结构图组成,图的标签分为 2 类,表示酶或者非酶。节点表示氨基酸,如果两个节点间隔小于 6 埃,它们之间有边存在。

社交网络数据集中,一般每个图表示对象之间的互动关系,节点表示一个对象,边表示对象之间存在某种互动。

(1) COLLAB<sup>[39]</sup>是一个科学合作数据集,包括从高能物理,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图 (Ego-network),对应的标签是研究人员所属的研究领域。分类的任务是判断这些自我中心网络图对应研究人员的研究领域。

(2) REDDIT-BINARY 数据集<sup>[39]</sup>中包含 2000 个 Reddit 网站上用户的社交互动图,每个图表示一个在线讨论线程,节点表示用户,如果两个用户之间有过消息回应,则他们对应的节点之间有边相连。这些互动图包括基于问题-答案互动图和讨论互动图 2 类,图分类的任务就是判断给定社交图是来自基于问答的社区还是基于讨论的社区。另外,REDDIT-MULTI-5K, REDDIT-MULTI-12K 均是该数据集的更大的变体,包含更多的来自不同社区的互动图,任务也是将互动图分类到对应社区。

(3) IMDB-BINARY<sup>[39]</sup>是一个电影合作数据集,数据来自于互联网电影数据库 (IMDB)。每个图中节点表示演员,如果两个演员出现在同一部电影中,则他们对应的节点之间会有一条边。这些合作图均来自动作和浪漫这 2 种类型,在合作图中为每个演员衍生出自我中心网络图,图分类的任务就是判断给定的自我中心网络图属于动作类型还是浪漫类型。另外 IMDB-MULTI 是该数据集的多类型版本,任务也是将演员子网络图分类到对应的电影类型。

气味数据集中,一般每个图有多个标签,属于多标签分类问题。具体来说,在计算机嗅觉中的气味数据集 GoodScents<sup>[66]</sup>、Leffingwell PMP 2001<sup>[67]</sup>和 Flavornet<sup>[68]</sup>中,每个图表示一个化学分子,节点表示原子,边表示化学键,图的标签表示该分子的一种或多种气味属性。其中,GoodScents 数据集<sup>[66]</sup>、Leffingwell PMP 2001 数据集<sup>[67]</sup>和 Flavornet 数据集<sup>[68]</sup>中分别包含 3786、3561、738 个分子,每个分子由嗅觉专家使用 1 个或多个气味描述符对分子进行标记,也就是分子的多个类别标签。分类任务是预测每个气味分子所具有的所有类别标签。

## 4.2 图分类方法效果对比

### 4.2.1 评价指标

图分类方法的评价指标主要包括分类准确率,精准率,召回率,F1 值和 AUC,下面分别介绍。

(1) 分类准确率 (Accuracy), 衡量了被正确分类的图在所有图中的占比。

$$Accuracy = \text{正确分类样本数} / \text{总样本数}$$

(2) 精准率 (Precision), 衡量预测为正例的样本中,真正的正例样本的比例。

$$Precision = \text{预测正确的正例数} / \text{所有预测为正的样本数}$$

(3) 召回率 (Recall), 衡量所有样本中的正例,被正确预测出来的比例。

$$Recall = \text{预测正确正例数} / \text{样本中所有正例数}$$

(4) F1 值是精准率和召回率的调和平均数,用来衡量当样本类别不均衡时的分类器性能。

$$F1 = \frac{2Precision \cdot Recall}{Precision + Recall}$$

(5) AUC (the area under the receiver operating characteristic). ROC (receiver operating characteristic) 曲线用来衡量当分类阈值变化时,分类器系统的判别能力。ROC 曲线对样本是否平衡不敏感,适用于样本类别不均衡时对分

类器性能的评价. 而 AUC 指的是 ROC 曲线下的面积, 取值一般在 0.5 到 1 之间. 分类器的 AUC 等于将随机选择的正例排在随机选择的负例前面的概率<sup>[69]</sup>.

#### 4.2.2 效果对比

由于本文提到的图匹配方法没有显式地进行图分类任务, 本小节只给出图核方法和图神经网络方法在图分类上的实验结果, 表 5 是部分具有代表性的图核方法在生物化学数据集和社交网络数据集上的分类结果. WL 子树核方法在超过一半的数据集上取得了最高的分类准确率. 对于深度图核, 这里对比的是深度最短路径图核方法 (deep SP graph kernel). 与最短路径图核方法相比, 取得了更好或者类似的结果. 此外, Graphlet 图核方法在社交网络数据集 REDDIT-B 上取得了最好的分类结果. 对于在具体情境下对图核方法的选择需要理解其特点. 相比随机游走核, 最短路径核方法更适合出现路径回溯情况的场景中; 而当图的规模比较大或者复杂时, 更适合用基于子图和子树的图核方法. 但需要注意的是, 使用基于子图和子树的两类图核时需要确保语义信息不会在整个图上传播, 也就是, 这两类方法不适合图上距离较远的点之间有较大的语义重要性的情境<sup>[70]</sup>. 此外, 深度图核可以用于需要缓解核矩阵对角控制问题的场景.

表 5 图核方法的图分类准确率 (%)

方法名称	生物化学数据集					社交网络数据集	
	D & D	NCHI	NCHI09	ENZYMES	MUTAG	REDDIT-B	COLLAB
ShortestPath <sup>[13]</sup>	78.45	73.47	73.07	41.68	87.28	64.11	59.10
Graphlet <sup>[14]</sup>	78.59	66.00	66.59	32.70	75.61	<b>78.04</b>	64.66
WL subtree <sup>[15]</sup>	<b>79.78</b>	<b>82.19</b>	<b>82.46</b>	52.22	82.05	68.20	<b>78.61</b>
Deep SP Graph Kernel <sup>[39]</sup>	—	73.55	73.26	<b>41.65</b>	<b>87.44</b>	—	—

由于当前基于图表示学习的图分类模型的实验设置和数据划分不同, 我们很难直接横向比较各个方法. 表 6 给出了部分具有代表性方法在统一的实验设置下的分类结果. 这些方法中模型都经过严格的模型评估和模型选择的过程, 模型的输入特征和数据划分都统一公平, 采用十折交叉检验的方式进行实验, 实验结果均可复现<sup>[23]</sup>.

表 6 部分基于图神经网络的模型图分类准确率 (%)

方法名称	生物化学数据集				社交网络数据集					
	D & D	NCHI	PROTEINS	ENZYMES	IMDB-B	IMDB-M	REDDIT-B	REDDIT-M	COLLAB	
DGCNN <sup>[4]</sup>	<b>76.6±4.3</b>	76.4±1.7	72.9±3.5	38.9±5.7	69.2±3.0	45.6±3.4	87.8±2.5	49.2±1.2	71.2±1.5	
DiffPool <sup>[17]</sup>	75.0±3.5	76.9±1.9	<b>73.7±3.5</b>	59.5±5.6	68.4±3.3	45.6±3.4	89.1±1.6	53.8±1.4	68.9±2.0	
GIN <sup>[24]</sup>	75.3±2.9	<b>80.0±1.4</b>	73.3±4.0	<b>59.6±4.5</b>	<b>71.2±3.9</b>	<b>48.5±3.3</b>	<b>89.9±1.9</b>	<b>56.1±1.7</b>	<b>75.6±2.3</b>	
GraphSAGE <sup>[71]</sup>	72.9±2.0	76.0±1.8	73.0±4.5	58.2±6.0	68.8±4.5	47.6±3.5	84.3±1.9	50.0±1.3	73.9±1.7	

从表 6 可以看到, 在严格十折交叉验证、相同的实验场景下, GIN<sup>[24]</sup>在多数图分类数据集上取得了最好的效果, 在基于邻居信息传递的图卷积方法中, GIN 的表达能力最接近 WL 算法. 实验表明 GIN 在获取图表示时, 使用加和池化的方式可以比平均池化及最大池化更好的捕获结构信息. DGCNN<sup>[4]</sup>和 DiffPool<sup>[17]</sup>分别在 D & D 蛋白质数据集和 PROTEINS 蛋白质数据集上取得了最好的分类准确率. 但在实验中, DiffPool 不容易优化, 需要一些辅助任务来帮助训练, 如链路预测. 此外, DiffPool 学习得到的分配矩阵比较稠密, 在实验中存在超出内存的风险. 对于 GraphSAGE 方法<sup>[71]</sup>, 由于原文中并没有进行图分类任务, 这里的对比实验中采用的是最大池化的全局聚合方法得到图表示, 分类准确率低于那些专门为图分类设计的模型. 虽然现有的图神经网络模型在图分类任务上取得了不错的效果, 但 Errica 等人<sup>[23]</sup>在实验中发现, 当前这些基于图神经网络的分类方法对结构的利用仍不充分. 而在生物化学领域, 结构信息与某些分子属性类别紧密相关, 未来仍需更进一步探究图神经网络模型对图结构信息的利用.



## 5 图分类应用与未来研究方向

图分类不仅在实验中取得了较好的结果,而且在生物化学领域,网络分析领域,计算机安全等领域有着非常重要的应用和广阔的前景.本节给出图分类的已有应用场景和一些未来可能的研究方向.

### 5.1 图分类的应用场景

#### (1) 化学信息学、生物信息学

传统的图分类主要应用于生物和化学领域.它们天然地提供了很多图结构数据.通过实验判断分子属性或蛋白质功能的方式代价较大,因此机器学习的方法被广泛应用于生物化学信息学中.在化学信息学中,化合物被建模为图,该领域常见的问题是判断化合物是否具有某些性质.图分类方法已经被用于判断分子是否具有诱变性、抗癌活性、毒性等任务中<sup>[6,7]</sup>.图分类在药物开发领域,也有着非常重要的应用,通过图机器学习的方法对药物的安全性等性质进行判断,同时帮助化学家深入理解不断增长的药物发现数据<sup>[72]</sup>.此外,在多标签图分类场景下,图分类方法也被用于计算机嗅觉领域中定量结构气味关系(QSOR)建模问题.此时,分子有一个或多个气味属性标签,任务是预测分子的气味属性标签<sup>[20,68]</sup>.

同样的,在生物信息学领域,对蛋白质的探索<sup>[9]</sup>也是一项重要任务.蛋白质的高级结构被建模为图.常见的应用包括蛋白质属性判断,如蛋白质是酶或者非酶,通过蛋白质交互网络预测疾病<sup>[8]</sup>等.

#### (2) 社交网络分析

在社交网络分析领域,最常见的数据之一是引用网络,如第4.1节中描述的COLLAB数据集.数据集中的图是研究人员的自我中心网络图,也就是以研究人员为中心的引用关系图.该场景下常见的分类任务是给定训练集中自我网络图的类别标签,模型经训练后对测试集中自我网络图的类别进行判断.

#### (3) 计算机安全

图分类常被应用于计算机安全领域,例如软件剽窃的检测、恶意软件检测、软件漏洞检测<sup>[73-75]</sup>等重要安全问题.该场景下的图一般是经过一些转化方式得到的控制图,通过控制图结构判断是否存在安全问题.如在漏洞检测中,当无权访问源代码时,我们需要分析二进制文件,结合反汇编程序和代码分析器,提取代码的控制流图.控制流图以结构化的形式包含二进制函数中所有信息<sup>[43]</sup>.控制流图中的节点表示汇编指令的基本块,当两个基本块之间有跳转,循环或者返回等控制流时,对应节点之间有边,图标签是有无漏洞.当前,主要是基于图相似度计算的图分类方法应用于计算机安全领域,这些方法的假设是,当未知控制流图的结构和已知有漏洞的控制流图相似度较高时,判断该未知程序可能存在漏洞.

#### (4) 自然语言处理

图分类的方法应用于自然语言处理的第一步就是图的构建.一种常见的方法是构建文本的单词共现图<sup>[76-78]</sup>,节点表示单词等有意义的语言实体,边表示在固定大小的滑动窗口中的共现关系.与传统的词袋表示文本的方法相比,图不仅建模了单词等实体,也对他们之间的远距离依赖关系进行了建模.图分类的方法在自然语言处理领域已经被应用于文档相似性计算,文本分类的重要任务中.例如,Nikolentzos等人<sup>[77]</sup>用共现的方式将文档构建为无向无权图,然后利用最短路径核计算文档的相似性,取得了较好的效果.Peng等人<sup>[76]</sup>将文档构建为词共现图,然后用对单词图进行图卷积操作,提取单词图特征进而对文档进行分类,相比于传统的文本分类方法,该模型取得了较大的提升.

#### (5) 计算机视觉

有些基于图核和基于图神经网络的方法被用于计算机视觉领域的图像分类,语义分割,点云图的形状分类等应用中<sup>[79-82]</sup>.为了进行人体活动识别,Wu等人<sup>[79]</sup>首先构建了2个图模型建模人体活动的空间特征和时序关系,然后提出了上下文相关的图核来衡量图之间的相似性,进而对人体活动进行识别.Wang等人<sup>[80]</sup>在点云图上使用边卷积的方式提取几何特征,然后利用全局池化的方式得到整个图的表示进而进行形状分类任务,取得了较好的效果.

### 5.2 未来研究方向

虽然图分类问题已有很长的研究历史,并在近年取得了较大的进步.但该领域仍然有很多需要注意的问题和

值得继续探索的研究方向.

#### (1) 图分类中图结构信息的充分利用

图中的结构信息,即图上节点的连接信息,如一阶连接信息,二阶信息和其他高阶信息等,对于图分类有着非常重要的作用,例如生物信息数据集中,某些结构模式与分子功能属性有着必然的联系.但当前图分类领域中很多基于图神经网络的方法并没有有效地利用到图结构信息<sup>[23]</sup>,例如,在基于信息传递的图神经网络中,节点之间的连接关系仅用来指导节点之间的信息传递,并没有直接对结构信息建模.对于在图分类中如何更好地利用结构信息和判断模型对结构的利用程度上,我们并无定论.对于图结构信息的合理利用和对结构利用程度的表示是图分类领域重要的研究方向.

#### (2) 图分类方法的可解释性

基于图神经网络的图分类方法的提出,使得图的表示和分类过程可以统一地进行优化,取得了较好的分类效果.但是,这类模型通常比较复杂且不够透明,人类无法直观地理解它们的预测结果.对图分类模型的预测能力进行直观解释,探索这些模型中各个组件对图分类的作用不仅可以增加我们对 GNN 模型的信任,促进 GNN 模型应用于涉及到公平,隐私和安全的领域中,也可以增进研究人员对于网络特征的理解,进一步提升模型效果<sup>[27,83]</sup>.对图卷积神经网络的可解释性已有一些初步的尝试<sup>[24,83]</sup>,但当它们应用于图分类问题时的可解释性,仍然值得进一步探索.

#### (3) 图分类模型表达能力的衡量

当前图分类模型主要是基于图神经网络的模型.一方面,基于图神经网络模型的表达能力都是用判断图是否同构的能力来衡量的<sup>[24,51]</sup>.但我们并不能保证在这样的衡量标准下,对图是否同构的区分能力在图分类任务中可以泛化得好<sup>[27]</sup>.在图分类问题中,模型表达能力的衡量方法是一个重要的需要考虑的问题.另一方面,由于基于神经网络的模型依赖于充足数据,需要通过大量的数据进行训练.而当前图分类领域的常见数据集通常规模较小,不能很好地体现出方法的优势,限制了基于图神经网络的模型的代表能力.构建更好的图分类数据集成为亟待解决的问题.

#### (4) 图分类新技术

虽然已经有很多经典的图神经网络方法在图分类任务上取得了较好的效果,但仍面临着标签数据获取昂贵、模型迁移能力不足等诸多挑战,需要通过合理引入新技术来解决.具体来说,一方面,图神经网络的训练过程需要大量的任务相关的标签数据,标签数据的获取代价高昂<sup>[84]</sup>.另一方面,实际中,有时我们需要具有迁移能力的模型应用于不同的场景中.类比于自然语言处理和图像处理领域,图上也可以通过先在数据丰富的任务上对模型预训练,然后在目标任务上进行微调来解决这些问题.目前已有一些图上预训练的初步尝试<sup>[84-86]</sup>,未来图上的预训练仍是值得探索的问题.此外,当前图分类主要关注同质图,而实际场景中有很多异质图存在,已有的关于异质图的研究主要集中在节点分类问题<sup>[87,88]</sup>上,未来,关于异质图的分类也是值得关注的方向.

#### (5) 实验可复现性和学术社区的健康发展

在机器学习领域,实验的可复现一直是一个非常关键的议题<sup>[23]</sup>.当前用图神经网络处理图分类的工作中,实验程序通常不够严格且很难复现.不同方法中的实验设置也不尽相同,使得我们很难横向的对不同方法进行比较. Errica 等人<sup>[23]</sup>对 5 个图分类模型在统一的评估框架下做了对比.同样的数据划分和实验设置条件下,用 10 折交叉验证的方法进行模型的评估和选择,保证了实验的公平性.未来图分类领域的工作,应该延续这种做法,详细地给出方法的实验设置,方便公平对比和对问题的深入理解,推进图分类学术社区的健康发展.

## 6 总结与讨论

图分类问题不仅是很多研究领域的基础问题,而且有着广泛的应用,具有重要的研究价值.本文总结了近年来图分类领域的进展.将现有图分类的方法分为基于相似度计算的图分类和基于神经网络的图分类两大类.第 2 节对基于图相似度计算的图分类方法进行梳理,这类方法又包括图核方法和图匹配方法.传统图核方法中特征表示和分类的过程分别进行,存在无法根据任务进行优化,且计算复杂度较高的问题.而图匹配方法也不仅仅面向图分类任务.最近几年,随着神经网络在图领域的成功应用,越来越多的基于图神经网络的方法被应用于图分类问题中

并取得了较好的效果. 因此, 第3节对基于图神经网络的图分类方法进行总结和分析. 此外本文还总结了图分类的重要应用场景并讨论了存在的问题和未来可能的研究方向. 总的来说, 本文对近年来图分类领域的研究进行综述, 充分总结了已有方法并指出了存在的问题, 希望为进一步的研究提供一定的指导和帮助.

## References:

- [1] Xu BB, Shen HW, Cao Q, Qiu YQ, Cheng XQ. Graph wavelet neural network. In: Proc. of the Int'l Conf. of Learning Representation (ICLR). 2019.
- [2] Xu BB, Shen HW, Cao Q, Cen KT, Cheng XQ. Graph convolutional networks using heat kernel for semi-supervised learning. In: Proc. of the 28th Int'l Joint Conf. on Artificial Intelligence. 2019. 1928–1934.
- [3] Wu J, He JR, Xu JJ. Demo-net: Degree-specific graph neural networks for node and graph classification. In: Proc. of the 25th ACM SIGKDD Int'l Conf. on Knowledge Discovery & Data Mining. Singapore: ACM, 2019. 406–415.
- [4] Zhang MH, Cui ZC, Neumann M, Chen YX. An end-to-end deep learning architecture for graph classification. In: Proc. of the 32nd National Conf. on Artificial Intelligence. New Orleans: AAAI, 2018. 4438–4445.
- [5] Cen KT, Shen HW, Gao JH, Cao Q, Xu BB, Cheng XQ. ANAE: Learning node context representation for attributed network embedding. arXiv: 1906.08745, 2019.
- [6] Smalter A, Huan J, Lushington G. Graph wavelet alignment kernels for drug virtual screening. *Journal of Bioinformatics and Computational Biology*, 2009, 7(3): 473–497. [doi: 10.1142/S0219720009004187]
- [7] Mahé P, Vert JP. Graph kernels based on tree patterns for molecules. *Machine Learning*, 2009, 75(1): 3–35. [doi: 10.1007/s10994-008-5086-2]
- [8] Borgwardt KM, Kriegel HP, Vishwanathan SVN, Schraudolph NN. Graph kernels for disease outcome prediction from protein-protein interaction networks. *Pacific Symposium on Biocomputing*, 2007: 4–15.
- [9] Fout AM. Protein interface prediction using graph convolutional networks[MS. Thesis]. Colorado: Colorado State University, 2017. 6530–6539.
- [10] Kriege NM, Johansson FD, Morris C. A survey on graph kernels. *Applied Network Science*, 2020, 5(1): 6. [doi: 10.1007/s41109-019-0195-3]
- [11] Nikolentzos G, Siglidis G, Vazirgiannis M. Graph kernels: A survey. arXiv: 1904.12218, 2019.
- [12] Ma GX, Ahmed NK, Willke TL, Yu PS. Deep graph similarity learning: A survey. *Data Mining and Knowledge Discovery*, 2021, 35(3): 688–725. [doi: 10.1007/s10618-020-00733-5]
- [13] Borgwardt KM, Kriegel HP. Shortest-path kernels on graphs. In: Proc. of the Fifth IEEE Int'l Conf. on Data Mining (ICDM '05). Houston: IEEE, 2005. 8. [doi: 10.1109/ICDM.2005.132]
- [14] Shervashidze N, Vishwanathan SVN, Petri TH, Mehlhorn K, Borgwardt KM. Efficient graphlet kernels for large graph comparison. In: Proc. of the 12th Int'l Conf. on Artificial Intelligence and Statistics. Clearwater: MIT Press, 2009. 488–495.
- [15] Shervashidze N, Schweitzer P, Van Leeuwen EJ, Mehlhorn K, Borgwardt KM. Weisfeiler-lehman graph kernels. *Journal of Machine Learning Research*, 2011, 12: 2539–2561.
- [16] Ma Y, Wang SH, Aggarwal CC, Tang JL. Graph convolutional networks with eigenpooling. In: Proc. of the 25th ACM SIGKDD Int'l Conf. on Knowledge Discovery & Data Mining. Singapore: ACM, 2019. 723–731. [doi: 10.1145/3292500.3330982]
- [17] Ying R, You JX, Morris C, Ren X, Hamilton WL, Leskovec J. Hierarchical graph representation learning with differentiable pooling. In: Proc. of the 32nd Int'l Conf. on Neural Information Processing Systems. Montréal: Curran Associates Inc., 2018. 4805–4815.
- [18] Khasahmadi AH, Hassani K, Moradi P, Lee L, Morris Q. Memory-based graph networks. In: Proc. of the 8th Int'l Conf. on Learning Representations. arXiv: 2002.09518, 2020.
- [19] Lee J, Lee I, Kang J. Self-attention graph pooling. In: Proc. of the 36th Int'l Conf. on Machine Learning. Long Beach: IMLS, 2019. 6661–6670.
- [20] Sanchez-Lengeling B, Wei JN, Lee BK, Gerkin RC, Aspuru-Guzik A, Wiltschko AB. Machine learning for scent: Learning generalizable perceptual representations of small molecules. arXiv: 1910.10685, 2019.
- [21] Peng H, Li JX, Gong QR, Ning YX, Wang SZ, He LF. Motif-matching based subgraph-level attentional convolutional network for graph classification. Proc. of the AAAI Conf. on Artificial Intelligence, 2020, 30(4): 5387–5394.
- [22] Hamilton WL, Ying R, Leskovec J. Representation learning on graphs: Methods and applications. arXiv: 1709.05584, 2017.
- [23] Errica F, Podda M, Bacciu D, Micheli A. A fair comparison of graph neural networks for graph classification. In: Proc. of the Int'l Conf. on Learning Representations (ICLR). 2020.

- [24] Xu K, Hu W, Leskovec J, Jegelka S. How powerful are graph neural networks?. In: Proc. of the Int'l Conf. on Learning Representations (ICLR). New Orleans: OpenReview.net, 2019.
- [25] Morris C, Ritzert M, Fey M, Hamilton WL, Lenssen JE, Rattan G, Grohe M. Weisfeiler and leman go neural: Higher-order graph neural networks. Proc. of the AAAI Conf. on Artificial Intelligence, 2019, 33(1): 4602–4609. [doi: 10.1609/aaai.v33i01.33014602]
- [26] Maron H, Ben-Hamu H, Serviansky H, Lipman Y. Provably powerful graph networks. In: Proc. of the 33rd Conf. on Neural Information Processing Systems. Vancouver: NIPS, 2019. 2153–2164.
- [27] Cai C, Wang YS. A simple yet effective baseline for non-attributed graph classification. arXiv: 1811.03508, 2018.
- [28] Shawe-Taylor J, Cristianini N. Kernel Methods for Pattern Analysis. Cambridge: Cambridge University Press, 2004.
- [29] Bai L, Xu LX, Cui LX, Jiao LX, Cui YH, Wu YF, Pan YY. Graph kernels on machine learning: Recent works and future developments. Journal of Anhui University (Natural Science Edition), 2017, 41(1): 21–28 (in Chinese with English abstract). [doi: 10.3969/j.issn.1000-2162.2017.01.004]
- [30] Kriege NM, Neumann M, Morris C, Kersting K, Mutzel P. A unifying view of explicit and implicit feature maps of graph kernels. Data Mining and Knowledge Discovery, 2019, 33(6): 1505–1547. [doi: 10.1007/s10618-019-00652-0]
- [31] Haussler D. Convolution kernels on discrete structures. Santa Cruz: University of California at Santa Cruz, 1999.
- [32] Gärtner T, Flach P, Wrobel S. On graph kernels: Hardness results and efficient alternatives. In: Proc. of the 16th Annual Conf. on Learning Theory and Kernel Machines. Washington: Springer, 2003. 129–143. [doi: 10.1007/978-3-540-45167-9\_11]
- [33] Mahé P, Ueda N, Akutsu T, Perret JL, Vert JP. Extensions of marginalized graph kernels. In: Proc. of the 21st Int'l Conf. on Machine Learning. New York: ACM, 2004. [doi: 10.1145/1015330.1015446]
- [34] Wang X. Research on the key techniques of graph similarity measure [Ph.D. Thesis]. Changchun: Jilin University, 2017 (in Chinese with English Abstract)
- [35] Horváth T, Gärtner T, Wrobel S. Cyclic pattern kernels for predictive graph mining. In: Proc. of the 10th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. Singapore: ACM, 2004. 158–167. [doi: 10.1145/1014052.1014072]
- [36] Fröhlich H, Wegner JK, Sieker F, Zell A. Optimal assignment kernels for attributed molecular graphs. In: Proc. of the 22nd Int'l Conf. on Machine Learning. Bonn: ACM, 2005. 225–232. [doi: 10.1145/1102351.1102380]
- [37] Ramon J, Gärtner T. Expressivity versus efficiency of graph kernels. In: Proc. of the 1st Int'l Workshop on Mining Graphs, Trees and Sequences. Croatia: Lirias, 2003. 65–74.
- [38] Weisfeiler B, Lehman AA. A reduction of a graph to a canonical form and an algebra arising during this reduction. Nauchno-Technicheskaya Informatsia, 1968, 2(9): 12–16.
- [39] Yanardag P, Vishwanathan SVN. Deep graph kernels. In: Proc. of the 21st ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. Singapore: ACM, 2015. 1365–1374. [doi: 10.1145/2783258.2783417]
- [40] Foggia P, Percannella G, Vento M. Graph matching and learning in pattern recognition in the last 10 years. International Journal of Pattern Recognition and Artificial Intelligence, 2014, 28(1): 1450001. [doi: 10.1142/S0218001414500013]
- [41] Neuhaus M, Bunke H. Automatic learning of cost functions for graph edit distance. Information Sciences, 2007, 177(1): 239–247. [doi: 10.1016/j.ins.2006.02.013]
- [42] Schenker A, Last M, Bunke H, Kandel A. Classification of Web documents using graph matching. Int'l Journal of Pattern Recognition and Artificial Intelligence, 2004, 18(3): 475–496. [doi: 10.1142/S0218001404003241]
- [43] Li YJ, Gu CJ, Dullien T, Vinyals O, Kohli P. Graph matching networks for learning the similarity of graph structured objects. In: Proc. of the 36th Int'l Conf. on Machine Learning. California: PMLR, 2019. 3835–3845.
- [44] Bai YS, Ding H, Sun YZ, Wang W. Convolutional set matching for graph similarity. arXiv: 1810.10866, 2018.
- [45] Bai YS, Ding H, Bian S, Chen T, Sun YZ, Wang W. Simgnn: A neural network approach to fast graph similarity computation. In: Proc. of the 12th ACM Int'l Conf. on Web Search and Data Mining. Phoenix: ACM, 2019. 384–392. [doi: 10.1145/3289600.3290967]
- [46] Xu BB, Cen KT, Huang JJ, Shen HW, Cheng XQ. A survey on graph convolutional neural network. Chinese Journal of Computers, 2020, 43(5): 755–780 (in Chinese with English abstract). [doi: 10.11897/SP.J.1016.2020.00755]
- [47] Gilmer J, Schoenholz SS, Riley PF, Vinyals O, Dahl GE. Neural message passing for quantum chemistry. In: Proc. of the 34th Int'l Conf. on Machine Learning. Sydney: JMLR, 2017. 1263–1272.
- [48] Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. In: Proc. of the 5th Int'l Conf. on Learning Representations (ICLR). Toulon: OpenReview.net, 2016.
- [49] Gao HY, Ji SW. Graph U-Nets. In: Proc. of the 36th Int'l Conf. on Machine Learning. Long Beach: PMLR, 2019. 2083–2092.
- [50] Yuan H, Ji SW. StructPool: Structured graph pooling via conditional random fields. In: Proc. of the 8th Int'l Conf. on Learning Representations (ICLR). Addis Ababa: OpenReview.net, 2020.



- [51] Sato R. A survey on the expressive power of graph neural networks. arXiv: 2003.04078, 2020.
- [52] Niepert M, Ahmed M, Kutzkov K. Learning convolutional neural networks for graphs. In: Proc. of the 33rd Int'l Conf. on Machine Learning. New York: PMLR, 2016. 2014–2023.
- [53] McKay BD, Piperno A. Practical graph isomorphism, II. *Journal of Symbolic Computation*, 2014, 60: 94–112. [doi: [10.1016/j.jsc.2013.09.003](https://doi.org/10.1016/j.jsc.2013.09.003)]
- [54] Tzeng RC, Wu SH. Distributed, Egocentric representations of graphs for detecting critical structures. In: Proc. of the 36th Int'l Conf. on Machine Learning. Long Beach: PMLR, 2019. 6354–6362.
- [55] Defferrard M, Bresson X, Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering. In: Proc. of the 30th Conf. on Neural Information Processing Systems. Barcelona: NIPS, 2016. 3844–3852.
- [56] Diehl F, Brunner T, Le Truong M, Knoll A. Towards graph pooling by edge contraction. In: Proc. of the ICML Workshop on Learning and Reasoning with Graph-Structured Data. 2019.
- [57] Cangea C, Veličković P, Jovanović N, Kipf T, Liò P. Towards sparse hierarchical graph classifiers. arXiv: 1811.01287, 2018.
- [58] Ranjan E, Sanyal S, Talukdar P. ASAP: Adaptive structure aware pooling for learning hierarchical graph representations. Proc. of the AAAI Conf. on Artificial Intelligence, 2020, 34(4): 5470–5477. [doi: [10.1609/aaai.v34i04.5997](https://doi.org/10.1609/aaai.v34i04.5997)]
- [59] Zhang L, Wang XD, Li HS, Zhu GM, Shen PY, Li P, Lu XY, Shah SAA, Bennamoun M. Structure-feature based graph self-adaptive pooling. In: Proc. of the Web Conf. ACM, 2020. 3098–3104. [doi: [10.1145/3366423.3380083](https://doi.org/10.1145/3366423.3380083)]
- [60] Zhang Z, Bu JJ, Ester M, Zhang JF, Yao CW, Yu Z, Wang C. Hierarchical graph pooling with structure learning. arXiv: 1911.05954, 2019.
- [61] Debnath AK, De Compadre RLL, Debnath G, Shusterman AJ, Hansch C. Structure-activity relationship of mutagenic aromatic and heteroaromatic nitro compounds. correlation with molecular orbital energies and hydrophobicity. *Journal of Medicinal Chemistry*, 1991, 34(2): 786–797. [doi: [10.1021/jm00106a046](https://doi.org/10.1021/jm00106a046)]
- [62] Kazius J, McGuire R, Bursi R. Derivation and validation of toxicophores for mutagenicity prediction. *Journal of Medicinal Chemistry*, 2005, 48(1): 312–320. [doi: [10.1021/jm040835a](https://doi.org/10.1021/jm040835a)]
- [63] Wale N, Watson IA, Karypis G. Comparison of descriptor spaces for chemical compound retrieval and classification. *Knowledge and Information Systems*, 2008, 14(3): 347–375. [doi: [10.1007/s10115-007-0103-5](https://doi.org/10.1007/s10115-007-0103-5)]
- [64] Borgwardt KM, Ong CS, Schönauer S, Vishwanathan SVN, Smola AJ, Kriegel HP. Protein function prediction via graph kernels. *Bioinformatics*, 2005, 31(1): i47–i56. [doi: [10.1093/bioinformatics/bti1007](https://doi.org/10.1093/bioinformatics/bti1007)]
- [65] Dobson PD, Doig AJ. Distinguishing enzyme structures from non-enzymes without alignments. *Journal of Molecular Biology*, 2003, 330(4): 771–783. [doi: [10.1016/S0022-2836\(03\)00628-4](https://doi.org/10.1016/S0022-2836(03)00628-4)]
- [66] The Good Scents Company. The good scents company information system. <http://www.thegoodscentscompany.com>
- [67] Leffingwell JC. Leffingwell & associates, 2005.
- [68] Tran NB, Kepple DR, Shuvaev SA, Koulakov AA. DeepNose: Using artificial neural networks to represent the space of Odorants. In: Proc. of the 36th Int'l Conf. on Machine Learning. Long Beach: PMLR, 2019. 6305–6314.
- [69] Fawcett T. An introduction to ROC analysis. *Pattern Recognition Letters*, 2006, 27(8): 861–874. [doi: [10.1016/j.patrec.2005.10.010](https://doi.org/10.1016/j.patrec.2005.10.010)]
- [70] Ghosh S, Das N, Gonçalves T, Quaresma P, Kundu M. The journey of graph kernels through two decades. *Computer Science Review*, 2018, 27: 88–111. [doi: [10.1016/j.cosrev.2017.11.002](https://doi.org/10.1016/j.cosrev.2017.11.002)]
- [71] Hamilton WL, Ying R, Leskovec J. Inductive representation learning on large graphs. In: Proc. of the 31st Int'l Conf. on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017. 1025–1035.
- [72] Xiong ZP, Wang DY, Liu XH, Zhong FS, Wan XZ, Li XT, Li ZJ, Luo XM, Chen KX, Jiang HL, Zheng MY. Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. *Journal of Medicinal Chemistry*, 2019, 63(16): 8749–8760. [doi: [10.1021/acs.jmedchem.9b00959](https://doi.org/10.1021/acs.jmedchem.9b00959)]
- [73] Xu XJ, Liu C, Feng Q, Yin H, Song L, Song D. Neural network-based graph embedding for cross-platform binary code similarity detection. In: Proc. of the 2017 ACM SIGSAC Conf. on Computer and Communications Security. Seoul: ACM, 2017. 363–376. [doi: [10.1145/3133956.3134018](https://doi.org/10.1145/3133956.3134018)]
- [74] Feng Q, Zhou RD, Xu CC, Cheng Y, Testa B, Yin H. Scalable graph-based bug search for firmware images. In: Proc. of the 2016 ACM SIGSAC Conf. on Computer and Communications Security. Seoul: ACM, 2016. 480–491. [doi: [10.1145/2976749.2978370](https://doi.org/10.1145/2976749.2978370)]
- [75] Fyrbiak M, Wallat S, Reinhard S, Bissantz N, Paar C. Graph similarity and its applications to hardware security. *IEEE Trans. on Computers*, 2020, 69(4): 505–519. [doi: [10.1109/TC.2019.2953752](https://doi.org/10.1109/TC.2019.2953752)]
- [76] Peng H, Li JX, He Y, Liu YP, Bao MJ, Wang LH, Song YQ, Yang Q. Large-scale hierarchical text classification with recursively regularized deep graph-CNN. In: Proc. of the 2018 World Wide Web Conf. Lyon: Int'l World Wide Web Conferences Steering

- Committee, 2018. 1063–1072. [doi: [10.1145/3178876.3186005](https://doi.org/10.1145/3178876.3186005)]
- [77] Nikolettos G, Meladinos P, Rousseau F, Stavarakas Y, Vazirgiannis M. Shortest-path graph kernels for document similarity. In: Proc. of the 2017 Conf. on Empirical Methods in Natural Language Processing. Copenhagen: Association for Computational Linguistics, 2017. 1890–1900. [doi: [10.18653/v1/D17-1202](https://doi.org/10.18653/v1/D17-1202)]
- [78] Gao HY, Chen YJ, Ji SW. Learning graph pooling and hybrid convolutional operations for text representations. In: Proc. of the World Wide Web Conf. San Francisco: ACM, 2019. 2743–2749. [doi: [10.1145/3308558.3313395](https://doi.org/10.1145/3308558.3313395)]
- [79] Wu BX, Yuan CF, Hu WM. Human action recognition based on context-dependent graph kernels. In: Proc. of the 2014 IEEE Conf. on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014. 2609–2616. [doi: [10.1109/CVPR.2014.334](https://doi.org/10.1109/CVPR.2014.334)]
- [80] Wang Y, Sun YB, Liu ZW, Sarma SE, Bronstein MM, Solomon JM. Dynamic graph CNN for learning on point clouds. ACM Trans. on Graphics (TOG), 2019, 38(5): 1–12. [doi: [10.1145/3326362](https://doi.org/10.1145/3326362)]
- [81] Qi XJ, Liao RJ, Jia JY, Fidler S, Urtasun R. 3D graph neural networks for RGBD semantic segmentation. In: Proc. of the 2017 IEEE Int'l Conf. on Computer Vision (ICCV). Venice: IEEE, 2017. 5209–5218. [doi: [10.1109/ICCV.2017.556](https://doi.org/10.1109/ICCV.2017.556)]
- [82] Jiang QR. Research on graph kernels and their application in pattern recognition [Ph.D. Thesis]. Beijing: Beijing University of Technology, 2012 (in Chinese with English Abstract)
- [83] Ying R, Bourgeois D, You JX, Zitnik M, Leskovec J. GNNExplainer: Generating explanations for graph neural networks. arXiv: 1903.03894, 2019.
- [84] Hu ZN, Dong YX, Wang KS, Chang KW, Sun YZ. GPT-GNN: Generative pre-training of graph neural networks. In: Proc. of the 26th ACM SIGKDD Int'l Conf. on Knowledge Discovery & Data Mining. Singapore: ACM, 2020. 1857–1867. [doi: [10.1145/3394486.3403237](https://doi.org/10.1145/3394486.3403237)]
- [85] Hu WH, Liu BW, Gomes J, Zitnik M, Liang P, Pande V, Leskovec J. Strategies for pre-training graph neural networks. In: Proc. of the 8th Information Conf. of Learning Representation (ICLR). Addis Ababa: OpenReview.net, 2020.
- [86] Qiu JZ, Chen QB, Dong YX, Zhang J, Yang HX, Ding M, Wang KS, Tang J. GCC: Graph contrastive coding for graph neural network pre-training. In: Proc. of the 26th ACM SIGKDD Int'l Conf. on Knowledge Discovery & Data Mining. Singapore: ACM, 2020. 1150–1160. [doi: [10.1145/3394486.3403168](https://doi.org/10.1145/3394486.3403168)]
- [87] Wang X, Ji HY, Shi C, Wang B, Ye YF, Cui P, Yu PS. Heterogeneous graph attention network. In: Proc. of the World Wide Web Conf. San Francisco: ACM, 2019. 2022–2032. [doi: [10.1145/3308558.3313562](https://doi.org/10.1145/3308558.3313562)]
- [88] Zhang CX, Song DJ, Huang C, Swami A, Chawla NV. Heterogeneous graph neural network. In: Proc. of the 25th ACM SIGKDD Int'l Conf. on Knowledge Discovery & Data Mining. Anchorage: ACM, 2019. 793–803. [doi: [10.1145/3292500.3330961](https://doi.org/10.1145/3292500.3330961)]

#### 附中文参考文献:

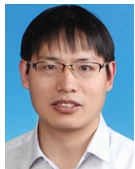
- [29] 白璐, 徐立祥, 崔丽欣, 焦宇航, 吴宇帆, 潘逸逸. 图核函数研究现状与进展. 安徽大学学报(自然科学版), 2017, 41(1): 21–28. [doi: [10.3969/j.issn.1000-2162.2017.01.004](https://doi.org/10.3969/j.issn.1000-2162.2017.01.004)]
- [34] 王旭. 图相似度度量的关键技术研究[博士学位论文]. 长春: 吉林大学, 2017.
- [46] 徐冰冰, 岑科廷, 黄俊杰, 沈华伟, 程学旗. 图卷积神经网络综述. 计算机学报, 2020, 43(5): 755–780. [doi: [10.11897/SP.J.1016.2020.00755](https://doi.org/10.11897/SP.J.1016.2020.00755)]
- [82] 蒋强荣. 图核及其在模式识别中应用的研究[博士学位论文]. 北京: 北京工业大学, 2012.



王兆慧 (1994—), 女, 博士生, 主要研究领域为图神经网络, 图表示学习.



曹琦 (1992—), 女, 博士, 助理研究员, CCF 专业会员, 主要研究领域为社交媒体分析, 信息传播建模与预测.



沈华伟 (1982—), 男, 博士, 研究员, 博士生导师, CCF 高级会员, 主要研究领域为社会计算, 网络数据挖掘, 数据智能.



程学旗 (1971—), 男, 博士, 研究员, 博士生导师, CCF 会士, 主要研究领域为大数据分析, 数据科学, 信息检索与挖掘.