

# KENN: 线性结构熵的图核神经网络\*

徐立祥<sup>1,2,3</sup>, 许巍<sup>1</sup>, 陈恩红<sup>2,3</sup>, 罗斌<sup>4</sup>, 唐远炎<sup>5</sup>

<sup>1</sup>(合肥大学 人工智能与大数据学院, 安徽 合肥 230601)

<sup>2</sup>(中国科学技术大学 计算机科学与技术学院, 安徽 合肥 230027)

<sup>3</sup>(认知智能全国重点实验室, 安徽 合肥 230088)

<sup>4</sup>(安徽大学 计算机科学与技术学院, 安徽 合肥 230601)

<sup>5</sup>(澳门大学 珠海澳大科技研究院, 澳门 999078)

通信作者: 陈恩红, E-mail: [chenh@ustc.edu.cn](mailto:chenh@ustc.edu.cn)



**摘要:** 图神经网络 (graph neural network, GNN) 是一种利用深度学习直接对图结构数据进行表征的框架, 近年来受到人们越来越多的关注。然而传统的基于消息传递聚合的图神经网络 (messaging passing GNN, MP-GNN) 忽略了不同节点的平滑速度, 无差别地聚合了邻居信息, 易造成过平滑现象。为此, 研究并提出一种线性结构熵的图核神经网络分类方法, 即 KENN。它首先利用图核方法对节点子图进行结构编码, 判断子图之间的同构性, 进而利用同构系数来定义不同邻居间的平滑系数。其次基于低复杂度的线性结构熵提取图的结构信息, 加深和丰富图数据的结构表达能力。通过将线性结构熵、图核和图神经网络三者进行深度融合提出了图核神经网络分类方法。它不仅可以解决生物分子数据节点特征的稀疏问题, 也可以解决社交网络数据以节点度作为特征所产生的信息冗余问题, 同时还使得图神经网络能够自适应调整对图结构特征的表征能力, 使其超越 MP-GNN 的上界 (WL 测试)。最后, 在 7 个公开的图分类数据集上实验验证了所提出模型的性能优于其他的基准模型。

**关键词:** 图分类; 结构熵; 图核; 消息传递聚合; 图神经网络

**中图法分类号:** TP18

中文引用格式: 徐立祥, 许巍, 陈恩红, 罗斌, 唐远炎. KENN: 线性结构熵的图核神经网络. 软件学报, 2024, 35(5): 2430–2445. <http://www.jos.org.cn/1000-9825/7039.htm>

英文引用格式: Xu LX, Xu W, Chen EH, Luo B, Tang YY. KENN: Graph Kernel Neural Network Based on Linear Structural Entropy. Ruan Jian Xue Bao/Journal of Software, 2024, 35(5): 2430–2445 (in Chinese). <http://www.jos.org.cn/1000-9825/7039.htm>

## KENN: Graph Kernel Neural Network Based on Linear Structural Entropy

XU Li-Xiang<sup>1,2,3</sup>, XU Wei<sup>1</sup>, CHEN En-Hong<sup>2,3</sup>, LUO Bin<sup>4</sup>, TANG Yuan-Yan<sup>5</sup>

<sup>1</sup>(School of Artificial Intelligence and Big Data, Hefei University, Hefei 230601, China)

<sup>2</sup>(School of Computer Science and Technology, University of Science and Technology of China, Hefei 230027, China)

<sup>3</sup>(State Key Laboratory of Cognitive Intelligence, Hefei 230088, China)

<sup>4</sup>(School of Computer Science and Technology, Anhui University, Hefei 230601, China)

<sup>5</sup>(Zhuhai UM Science & Technology Research Institute, University of Macau, Macao 999078, China)

**Abstract:** Graph neural network (GNN) is a framework for directly characterizing graph structured data by deep learning, and has caught increasing attention in recent years. However, the traditional GNN based on message passing aggregation (MP-GNN) ignores the smoothing speed of different nodes and aggregates the neighbor information indiscriminately, which is prone to the over-smoothing phenomenon. Thus, this study proposes a graph kernel neural network classification method KENN based on linear structural entropy.

\* 基金项目: 国家自然科学基金 (62176085, 62172458); 国家自然科学基金区域 (安徽) 联合基金 (U20A20229)

收稿时间: 2023-04-03; 修改时间: 2023-05-29; 采用时间: 2023-08-26; jos 在线出版时间: 2024-01-10

CNKI 网络首发时间: 2024-01-12

KENN firstly adopts the graph kernel method to encode node subgraph structure, determines isomorphism among subgraphs, and then utilizes the isomorphism coefficient to define the smoothing coefficient among different neighbors. Secondly, it extracts the graph structural information based on the low-complexity linear structural entropy to deepen and enrich the structural expression capability of the graph data. This study puts forward a graph kernel neural network classification method by deeply integrating linear structural entropy, graph kernel and GNN, which can solve the sparse node features of biomolecular data and information redundancy generated by leveraging node degree as features in social network data. It also enables the GNN to adaptively adjust its ability to characterize the graph structural features and makes GNN beyond the upper bound of MP-GNN (WL test). Finally, experiments on seven public graph classification datasets verify that the proposed model outperforms other benchmark models.

**Key words:** graph classification; structural entropy; graph kernel; message passing aggregation; graph neural network (GNN)

图是由节点和边组成的连通网络, 这种类型的图结构数据可以用来描述复杂的关系或网络. 在日常生活的各个方面都有广泛的应用, 如物联网、药物发现、电路设计等. 在机器学习和数据科学中, 基于图的学习和推理一直是最热门的研究课题. 基于图的表示是复杂系统建模的一种强大工具, 数据类型从生物分子化合物到点云再到社交网络等. 在图的数据分析中, 一个基本的挑战是如何将图结构转换为特征向量表示.

过去的 30 年, 最经典的图结构分析方法可以分为两类, 即 (1) 图嵌入方法<sup>[1]</sup>和 (2) 图核方法<sup>[2,3]</sup>. 第 1 类的方法旨在将图转换为向量空间中的元素, 用标准的机器学习算法可直接用于图数据分析. 然而这些嵌入方法在低维模式空间中仅近似表达图的特征属性, 易导致部分结构信息丢失. 为了克服这一缺点, 使用图核方法在高维 Hilbert 空间中刻画图结构, 从而更好地保留结构信息. 但由于当前大多数图核方法无法像深度学习类的方法一样进行多轮迭代训练, 所以在精度上存在一定的劣势.

近年来, 由于深度学习在机器学习中的巨大成功, 人们越来越有兴趣将深度卷积神经网络 (convolution neural network, CNN) 推广到图领域<sup>[4-6]</sup>, 而最经典且最具有代表性的图深度方法当属图神经网络. 目前大多数 GNN 模型主要基于: 谱域<sup>[7-9]</sup>, 空域<sup>[10,11]</sup>, 以及时空域. 最初的空域 GNN 的发展是从谱网络向空间网络发展起来的, 2014 年 Bruna 等人提出了基于卷积定理的 Spectral CNN<sup>[12]</sup>, 他们通过多层叠加, 加入非线性激活函数, 以及在每一层都定义了图卷积核, 进而构造出一个图卷积神经网络, 由于运算都是在节点的谱域上完成, 因此称为 Spectral CNN. 但 Spectral CNN 有着不可忽略的缺点, Spectral CNN 的计算依赖于拉普拉斯矩阵分解, 显式地使用了特征向量矩阵, 而这个分解, 以及矩阵相乘的计算时间复杂度均较高. 又由于是基于拉普拉斯矩阵分解, 导致卷积核不是局部化的, 也即是一个节点的信息聚合不是来自其邻居, 而是所有节点. 为了解决 Spectral CNN 卷积核非局部化问题, 2016 年 NeurIPS 上发表的 ChebyNet<sup>[13]</sup>通过参数化核卷积和特征矩阵的多项式组合, 经过一些简单运算, 大大减少了参数量, 降低了计算的复杂度, 使得谱方法在 GNN 中得到了成功的应用. 谱方法的优势在于它具备平移不变性, 对节点排列的顺序不敏感, 能够处理无序图. 此外, 谱图神经网络基于图信号处理理论, 拥有坚实的数学基础和理论支撑, 可以利用频域变换和滤波器对图上的信号进行处理和学习.

谱域 GNN 通常是处理整个图, 难以并行或扩展到大图上, 且重视节点数据的特征信息, 忽略图数据的空间结构信息. 这对拥有大量特征的节点任务来说是可行的, 而对于具有丰富空间信息且节点特征稀少的图分类任务来说存在不足. 由于谱方法存在抽象性、计算复杂度高和缺乏解释性等挑战, 这导致谱方法的发展相对缓慢. 随着基于空域 GNN 的快速发展. 这些方法通过聚集近邻节点的信息, 直接在图结构上进行卷积. 在 ChebyNet 的基础上, 2017 年 Kipf 等人提出的图卷积神经网络 (graph convolutional network, GCN)<sup>[14]</sup>简化了更多参数, 呈现出轻量化的模型结构, 最终将谱方法与空间方法进行了统一. 基于空间的 GCN 将图卷积表示为从邻域聚合特征信息, 当图卷积网络的算法在节点层次运行时, 图池化模块可以与图卷积层交错, 将图分解为信息表达能力更强的子结构, GCN 是第一个出现的基于空间的图神经网络, 在此后的一段时间, 所出现的 GNN 都以 GCN 为先驱, 在 GCN 的基础上进行探索. 基于 GCN 提出的模型都有一个共同点, 它们都是基于消息聚合的图神经网络, 采用消息传递聚合方案来学习图中节点的低维向量空间表示. 设  $G=(V, E)$  是一个图, 对于每个节点  $v \in V$ , 消息传递聚合方案递归地聚合了  $V$  的邻域中的节点的特征向量, 并将聚合的信息与  $V$  本身的特征向量进行融合获得具有邻居与自身特征的节点表示. 由于节点上没有自然排序, 因此通常要求这种消息传递聚合方案是排列不变的.

总的来说,当前的图神经网络在图分类任务中还存在以下问题:(1)图分类数据不同于节点分类数据具有丰富的特征信息,特别在社交网络数据中不含节点标签,这导致即使只使用少量层数的图神经网络,节点区分也并不明显。(2)消息聚合过程中无差化的聚合所有邻居,忽略不同节点平滑速度<sup>[15]</sup>的差异性,导致多层网络后所引起的过平滑问题。为了解决以上问题,同时提高 GNN 的表达能力,本文在节点特征中引入结构熵,通过结构信息编码,以增强节点的表达性,增大节点间的差异性,以及增强节点之间的关联性。同时,本文将图核与图神经网络进行融合,利用图核将相邻节点子图投射至希尔伯特空间,进行结构相似性质编码,从而通过结构相似度为不同节点自适应地赋予平滑权重。

本文的主要贡献包括以下 3 个方面。

贡献 1: 本文将线性冯·诺依曼熵应用到图核方法中,并通过节点特征赋值的方式在 WL 核上进行实现。实验表明,基于线性冯·诺依曼熵的 WL 核与传统的 WL 核相比,具有更高的结构表达性。

贡献 2: 本文将图核应用到图神经网络中,通过图核来表征图的结构特征,并将结构特征运用到传统的消息聚合中,解决了传统图神经网络对图拓扑特征利用率不高问题。

贡献 3: 基于改进后的 WL 核与结构消息聚合框架,本文提出了基于线性冯·诺依曼结构熵的图核神经网络 KENN,并在 7 个公开数据集上进行图分类实验。实验结果表明,KENN 模型要优于其他图分类基准方法,且突破了 MP-GNN 的上界。

## 1 空域图神经网络的相关工作

近几年,图神经网络在节点分类、图分类以及链路预测方面得到了快速发展,并在实际应用中得到了广泛的运用和令人瞩目的成果<sup>[16-19]</sup>。然而,这种快速发展也带来了一些挑战,在图分类领域,过去的研究表明,MP-GNN 在区分图同构问题上的能力有一定的限制,其无法超越 WL 测试。针对如何超越 GNN 性能的上界,人们已经做了大量工作,主要可以分为两个方面,分别是:注入随机属性和注入确定性距离属性。下面将介绍如何达到 GNN 的上界,并探讨一些突破该上界的相关工作。

Xu 等人提出对任意两个非同构图  $G_1$  和  $G_2$ ,如果存在一个图神经网络将图  $G_1$  和  $G_2$  映射到不同的嵌入向量中,那么通过 WL 测试可以确定  $G_1$  和  $G_2$  是非同构图。这表明依赖消息聚合框架的 GNN 结构的上限是 WL 测试。为此作者通过构造单射的聚合函数与读出函数,实现达到图神经网络的上限,即图同构网络 (graph isomorphism network, GIN)<sup>[20]</sup>。但是模型的上限并不是一成不变的,当清晰地认识到模型上限所对应的前提条件并突破它时,相应的上限也就不存在了。

MP-GNN 的表达能力受限制的主要原因是 MP-GNN 不跟踪节点的身份,具有相同属性的不同节点将以相同的向量表示进行初始化,提高 MP-GNN 表达能力的一种方法是为每个节点注入一个独特的属性。Murphy 等人考虑将节点的随机分配顺序作为其额外属性并用  $Z_{RP}$  表示,提出了关系池化 GNN (relational pooling GNN, RP-GNN)<sup>[21]</sup> 模型。 $Z_{RP}$  是从所有可能的排列矩阵中均匀抽样的,随机选取一个双射 (排列)  $\pi: V(G) \rightarrow V(G)$ ,如果  $j=\pi(i)$ ,则设计排列矩阵  $[Z_{RP}]_{ij}=1$ ,否则为 0。而 Sato 等人通过设置从几乎均匀的离散概率分布中抽样每个节点的附加属性来泛化 RP-GNN,并提出了随机图同构网络 (random graph isomorphic network, RGIN)<sup>[22]</sup>。与 RP-GNN 的主要区别在于,两个节点的附加属性被设置为相互独立的。而在 RP-GNN 中,由于排列组合的原因,不同节点的一次性随机属性是相关的。

MP-GNN 在测量不同节点之间的距离和区分有属性的正则图的能力方面是有限的,这些限制基本上是从 WL 测试中继承下来的。如果能将 MP-GNN 与一些距离信息结合起来,那么复合模型必定实现更强的表达能力。例如 Li 等人考虑使用节点之间的距离编码作为节点的额外属性,提出了距离编码 GNN (distance encoding GNN, DE-GNN)<sup>[23]</sup>。而 Wijesinghe 等人提出图结构神经网络 (graph structure neural network, GSNN)<sup>[24]</sup>,它定义了子图同构、重叠同构、子树同构 3 种同构等级,并证明任意两幅图若存在子图同构则必将重叠同构,存在重叠同构则必将子树同构,反之则不成立。同时还提出了满足以上定理的同构系数计算方法,在保证消息聚合、池化,以及图读

出过程满足单射性的基础上, 通过子图重叠的程度引入了具有结构信息的距离属性, 并证明通过此方式构造的图神经网络将突破 WL 测试的上限.

然而, 这些方法大多数是利用浅层的结构信息, 忽略了邻接关系的结构相似度. 针对上面的问题, 本文提出线性结构熵的图核神经网络模型, 从结构信息角度学习图的表征, 既通过熵引入了多层属性信息, 同时利用图核定义了节点间的距离. 具体来说: (1) 利用结构熵提取节点周围结构信息作为节点的特征, 在丰富节点表达的同时, 弥补了图分类数据中节点特征稀疏的问题. (2) 通过图核将相邻子图映射到希尔伯特空间, 利用图核来判断图同构性, 寻找相邻节点的同构系数.

## 2 线性结构熵的图核神经网络

尽管基于消息聚合的 GIN 模型相对于只考虑节点特征的谱 GNN, 加强了节点间的邻接关系对图表征的影响, 使之更具有空间信息, 并且在图分类领域可达到和最优的 WL 测试相同的表现, 但是经历多层消息聚合后的节点特征难以分辨, 且严重破坏了图的结构信息. 本节介绍了一种线性结构熵的图核神经网络, 将图核强大的空间结构表征能力与 GIN 优秀的节点特征提取能力相结合, 同时在图核的基础上引入了冯·诺依曼熵的概念, 弥补了图核在处理原始图数据时节点特征单一的不足.

### 2.1 线性冯·诺依曼图熵

熵是信息论中一个十分重要的概念, 能够用于刻画信源的平均信息量, 和经典的信息论中的香农熵不一样, 量子信息中信源的不确定性需要使用冯·诺依曼熵来进行刻画. 同时, 冯·诺依曼熵在纠缠判别、纠缠度刻画等方面也具有十分重要的地位. 系统的冯·诺依曼熵在量子系统中的定义如公式 (1) 所示:

$$S = -\text{tr}(\rho \ln \rho) \quad (1)$$

其中,  $\text{tr}$  表示迹,  $\rho$  是系统的密度矩阵, 运用密度矩阵的本征态向量分解表达式, 将  $\rho$  定义如下:

$$\rho = \sum_i \omega_i |\psi_i\rangle \langle \psi_i| \quad (2)$$

将  $\rho$  代入公式 (1), 可以得到:

$$S = -\sum_i \omega_i \ln \omega_i \quad (3)$$

对于具有  $n$  个节点的图  $G = (V, E)$ , 其中  $V$  和  $E$  表示其节点集和边集, 进一步可以得到邻接矩阵  $A = \{A_{ij}\}_{(i,j) \in V}^{n \times n}$  和度矩阵  $D = \text{diag}(d_1, d_2, \dots, d_n)$ . 因此, 图  $G$  的拉普拉斯矩阵定义为:

$$LM = D - A \quad (4)$$

冯·诺依曼图熵 (von Neumann graph entropy, VNGE)<sup>[25]</sup> 是具有立方复杂度  $O(n^3)$  的一种结构熵, 公式如下:

$$Hs(G) = -\text{tr}(LM_N \ln LM_N) \quad (5)$$

$$LM_N = LM / \text{tr}(LM) \quad (6)$$

公式 (6) 中  $LM_N$  符合式 (2) 中矩阵  $\rho$  的上述属性. 但因为 VNGE 需要计算拉普拉斯矩阵的所有特征值. 对于生物分子和社交网络数据, 相应的无向图具有大量节点, 并且由于需要为每个节点多次计算图熵, 因此计算的复杂度被进一步扩大. 特别是对于大规模的数据, 庞大的计算使得熵的结构难以有效提取甚至无法完成. 为了减少计算复杂度, 本文应用快速增量冯·诺依曼图熵 (FINGER-H)<sup>[26]</sup>, 该形式将冯·诺依曼熵计算近似为具有线性复杂度的简单表达式. 采用此种方法可以使模型在熵的计算过程中仅使用图的节点数和边数, 使得复杂度降低至  $O(n)$ . 本文定义的线性 VNGE 如下:

$$\tilde{H}s(G_i) = -Q_i \cdot \ln(d_{i,\max}) \quad (7)$$

$$Q_i = 1 - (c_i)^2 \left( \sum_{r=1}^n (d_{i,r})^2 + 2 \sum_{p=1}^n \sum_{q=1}^n (\varepsilon_{i,(p,q)})^2 \right) \quad (8)$$

其中,  $c_i = 1/\text{tr}(D_i)$ ,  $d_{i,r}$  表示度矩阵  $D_i$  中的对角线元素,  $d_{i,\max}$  表示度矩阵  $D_i$  中的最大对角元素,  $\varepsilon_{i,(p,q)}$  表示邻接矩

阵  $A_i$  中的元素.

## 2.2 基于线性冯·诺依曼熵的 WL 图核

基于第 2.1 节的分析, 本文尝试将熵的概念应用到图核中, 通过熵来提取每个节点的子图特征, 进而丰富节点信息, 加强空间结构对图的特征表征的影响. 设节点  $v_i$  原始标签为  $l_{oi}$ , 对节点标签重新赋值. 由节点  $v_i$  的  $k$  阶邻居  $\{V_i^1, V_i^2, \dots, V_i^k\}$  构造子图  $g_i$ . 对  $g_i$  以  $v_i$  为中心通过  $k$  阶邻接矩阵  $\{A_i^1, A_i^2, \dots, A_i^k\}$  求其  $k$  阶图熵  $[\tilde{H}s(g_i^1), \tilde{H}s(g_i^2), \dots, \tilde{H}s(g_i^k)]^T$ , 以图熵表示的节点  $i$  的特征为  $l_{ei}$ , 重新赋值后的节点  $i$  特征为  $l_i$ .  $k$  阶线性 VNGE 的提取方式如图 1 所示.

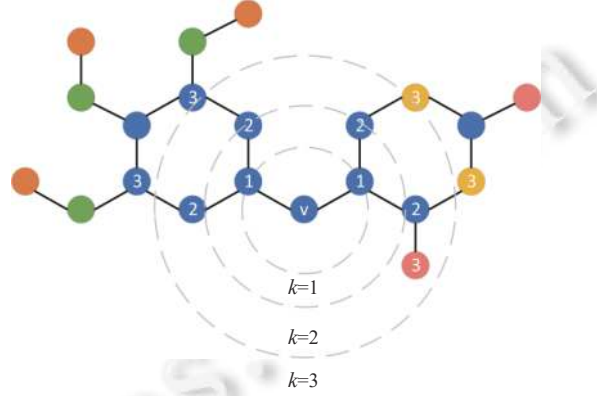


图 1  $k$  阶线性 VNGE 提取图

当图核的输入为向量时, 节点  $i$  的特征向量  $l_i$  定义如公式 (9), 公式 (10) 所示, 通过加权拼接的方式重新构造节点特征:

$$l_i = \text{Connect}(\partial \cdot l_{oi}, (1 - \partial) \cdot l_{ei}) \quad (9)$$

$$l_{ei} = [\tilde{H}s(g_i^1), \tilde{H}s(g_i^2), \dots, \tilde{H}s(g_i^k)] \quad (10)$$

当图核的输入为标量时, 节点  $i$  的特征标签  $l_i$  定义如公式 (11), 公式 (12) 所示, 通过加权求和的方式重新构造节点特征:

$$l_i = \partial \cdot l_{oi} + (1 - \partial) \cdot l_{ei} \quad (11)$$

$$l_{ei} = \tilde{H}s(g_i^1) + \tilde{H}s(g_i^2) + \dots + \tilde{H}s(g_i^k) \quad (12)$$

**定义 1.** 设  $g$  和  $g'$  是图, 定义  $\sum_i \subseteq \sum$  为在 Weisfeiler-Lehman (WL) 算法的第  $i$  次迭代结束时, 在  $g$  或  $g'$  中作为节点标记至少出现 1 次的集合. 设  $\sum_0$  为  $g$  和  $g'$  的节点标记集, 假定所有  $\sum_i$  是两两不相交的, 在不丧失一般性的情况下, 假定每个  $\Sigma_i = \{\sigma_{i1}, \dots, \sigma_{i|\Sigma_i|}\}$  是有序的. 定义一个映射  $c_i: \{g, g'\} \times \sum_i \rightarrow \mathbb{N}$ , 使得  $c_i(g, \sigma_{ij})$  是图  $g$  中字母  $\sigma_{ij}$  的出现次数<sup>[27]</sup>. 图  $g$  与  $g'$  上具有  $H$  层的 WL 子树核为:

$$K_{\text{WLentropy}}^h(g, g') = \langle \phi_{\text{WLentropy}}^h(g), \phi_{\text{WLentropy}}^h(g') \rangle \quad (13)$$

$$\phi_{\text{WLentropy}}^h(g) = (c_0(g, \sigma_{01}), \dots, c_0(g, \sigma_{0|\Sigma_0|}), \dots, c_0(g, \sigma_{h1}), \dots, c_0(g, \sigma_{h|\Sigma_h|})) \quad (14)$$

$$\phi_{\text{WLentropy}}^h(g') = (c_0(g', \sigma_{01}), \dots, c_0(g', \sigma_{0|\Sigma_0|}), \dots, c_0(g', \sigma_{h1}), \dots, c_0(g', \sigma_{h|\Sigma_h|})) \quad (15)$$

即 WL 子树核计算两个图更新后具有熵属性的公共节点标签以及多层迭代后的压缩标签, 基于线性冯·诺依曼熵的 WL 图核结构图后文如图 2 所示<sup>[28]</sup>.

## 2.3 KENN 模型设计

虽然在之前的 GNN 中使用了特征平滑操作对于大型图表示学习是可扩展的, 但它会导致节点的代表不是最优秀的. 由于不同结构属性的节点具有不同的平滑速度, 因此像之前的 GNN 那样对所有节点不加区别地执行特征平滑是不合理的, 必须采用节点自适应特征平滑 (即不同节点具有不同的平滑级别或机制) 来满足每个节点对

平滑级别的不同需求. 本文定义了一种通用的消息传递神经网络框架, 它可以根据相邻节点所在不同子图的结构相关性将局部结构信息注入聚合方案中, 通过图核的方式, 评估所有节点的平滑水平, 增大节点间的区别性, 并根据不同节点间的平滑权重进行邻域聚合. KENN 框架图如图 3 所示.

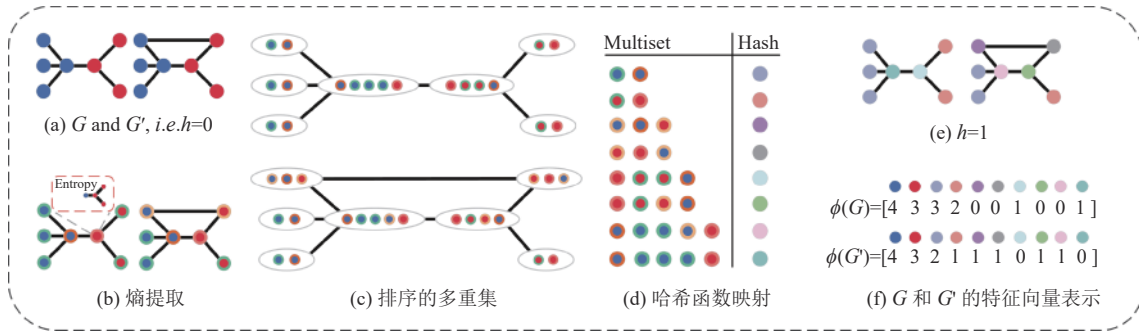


图 2 基于线性冯·诺依曼熵的 WL 图核结构图

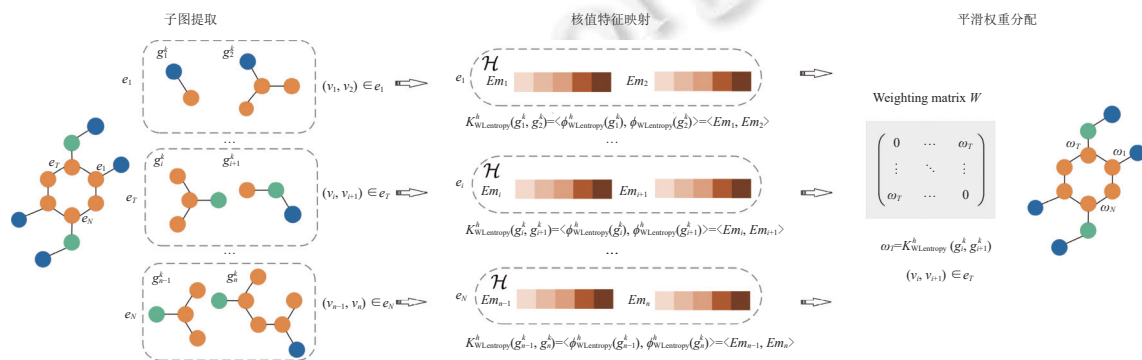


图 3 KENN 结构系数提取框架图

设  $SG = \{g_i^k | v \in V\}$  是提取图的节点对应子图集合. 定义每个顶点  $v$  及其邻居的结构系数  $\psi$ , 使得  $A_{ij} = \psi(v_i, v_j)$ , 对相邻节点需要提取相邻子图对, 而以每个节点为中心的子图  $g_i$  的大小是不固定的, 受与之相邻的邻居节点所在的子图影响. 当  $v_i$  与  $v_j$  互为邻居时, 要确定以  $v_i$  为中心的子图  $g_i$  的阶数大小  $k_i$ , 以及以  $v_j$  为中心的子图  $g_j$  的阶数大小  $k_j$ . 当  $k$  过小时, 无法有效表达节点周围结构信息, 而当  $k$  过大时, 相邻节点结构属性无法区分. 为防止以上现象, 同时也为了避免两个图的能量相差太大, 本文定义了能量函数, 利用邻居节点间的能量差来确定  $k_i$  和  $k_j$  的大小, 构造出总体能量相差较小的两个图.

$$\arg \min_{(k_i, k_j)} \{Energy(g_i^{k_i}) - Energy(g_j^{k_j}) | k_i, k_j \leq k_{\max}\} \quad (16)$$

$$Energy(g) = \sum_{i=1}^n d(v_i) \quad (17)$$

其中,  $g_i^k$  表示以节点  $v_i$  为中心选取  $k_i$  阶邻居所构造的子图,  $v_i$  代表子图  $g_i^k$  内的节点, 通过公式 (16) 和公式 (17) 提取两个图内节点总度数相差较小的子图.

设计  $\psi$  的方法是多样的, 导致具有不同表达能力的 GNN, 本文使用基于熵的 WL 核来设计  $\psi$ . 利用图核良好的结构化表述能力, 寻找相邻节点所构造子图间的结构相关度, 如公式 (18) 所示.

$$\psi(v_i, v_j) = \psi(g_i^{k_i}, g_j^{k_j}) = K_{\text{WLEntropy}}^h(g_i^{k_i}, g_j^{k_j}) \quad (18)$$

这个定义可以为图  $G$  构造一个权重矩阵  $W = (W_{ij})_{i,j \in V} = (\psi(v_i, v_j))_{i,j \in V}$ , 利用权重矩阵  $W$  与邻接矩阵  $A$  构造加权邻接矩阵  $A' = A + \bar{W}$ , 其中  $\bar{W}$  为权重矩阵  $W$  进行 Softmax 归一化处理后的矩阵, 目的是平衡权值与邻接矩

阵的占比. 为了比较不同节点间的结构系数, 继续将  $A'$  通过 Softmax 归一化为  $\bar{A}$ ,  $N(i)$  表示节点  $i$  的一阶邻居集合.

$$\bar{A}_{ij} = \frac{A'_{ij}}{\sum_{j \in N(i)} A'_{ij}} \quad (19)$$

而对于上述的归一化方法也可以使用 Min-Max, Z-score 等方法, 基于结构熵的图核提取结构信息的算法描述见算法 1.

**算法 1.** 基于 VNGE 的图核提取结构信息.

输入: Graph  $G(V, E)$ ;

输出:  $\bar{A}$ .

1. **For**  $e_i$  in  $E$ :
2.    $(v_i, v_j) \in e_i$
3.   Find subgraphs  $g_i, g_j$
4.    $\arg \min_{(k_i, k_j)} \{Energy(g_i^{k_i}) - Energy(g_j^{k_j}) | k_i, k_j \leq k_{\max}\}$
5.    $Energy(g) = \sum_{i=1}^n d(v_i)$
6.   **For**  $v_i$  in  $g_i$  and  $g_j$ :
7.     **Switch** (choice)
8.     **Case 0:**
9.        $l_i = Connect(\partial \cdot l_{ot}, (1 - \partial) \cdot l_{ei})$
10.       $l_{ei} = [\tilde{H}s(g_i^1), \tilde{H}s(g_i^2), \dots, \tilde{H}s(g_i^k)]$
11.     **Case 1:**
12.        $l_i = \partial \cdot l_{ot} + (1 - \partial) \cdot l_{ei}$
13.        $l_{ei} = \tilde{H}s(g_i^1) + \tilde{H}s(g_i^2) + \dots + \tilde{H}s(g_i^k)$
14.       $\psi(v_i, v_j) = \psi(g_i^{k_i}, g_j^{k_j}) = K_{\text{WLentropy}}^h(g_i^{k_i}, g_j^{k_j})$
15.       $K_{\text{WLentropy}}^h(g_i^{k_i}, g_j^{k_j}) = \langle \phi_{\text{WLentropy}}^h(g_i^{k_i}), \phi_{\text{WLentropy}}^h(g_j^{k_j}) \rangle$
16.       $W_{ij} = \psi(v_i, v_j)$
17.    $\bar{W} = \text{Softmax}(W)$
18.    $\bar{A} = \text{Softmax}(A + \bar{W})$

基于以上方法得出的结构加权邻接矩阵和带有结构熵的 GIN 框架, 本文提出了一个通用的消息传递框架, 它能够利用图核对相邻子图进行特征映射到希尔伯特空间, 进而计算结构相似度, 同时将局部结构融入聚合方案中. KENN 的消息传递框架如下所示:

$$a_i^{(t)} = \text{AGGREGATE}^{(t)}(\{\bar{A}_{ij}, h_j^{(t-1)} | j \in N(i)\}) \quad (20)$$

$$h_i^{(t)} = \text{COMBINE}^{(t)}(h_i^{(t-1)}, a_i^{(t)}) \quad (21)$$

$$h_G = \text{READOUT}(\{h_v^{(t)} | v \in G\}) \quad (22)$$

其中,  $h_i^{(t)}$  是节点  $i$  在第  $t$  次迭代的特征向量. 初始化  $h_i^{(0)} = x_i$ ,  $x_i$  为节点  $i$  的输入特征,  $N(i)$  是与  $i$  相邻的节点集合. 在进行了  $T$  轮特征融合后得到每个节点更新后的特征  $h_v^{(T)}$ , KENN 采用求和的 READOUT 函数得到最终的图表征  $h_G$ , 以保证特征映射过程的单射性. 对每个顶点  $v_i \in V$ , 第  $t$  层的特征向量由公式 (23) 计算:

$$h_i^{(t)} = \text{MLP}^{(t)}\left(\gamma^{(t-1)}\left(\sum_{j \in N(i)} \bar{A}_{ij} + 1\right)h_i^{(t-1)} + \sum_{j \in N(i)} (\bar{A}_{ij} + 1)h_j^{(t-1)}\right) \quad (23)$$

其中,  $\gamma^{(l)}$  是一个可学习的标量参数. 由于  $\mathcal{N}(i)$  是指节点  $i$  的一跳邻域, 因此可以堆叠多层来处理多阶. 特别地, 为了防止因结构系数过大而导致消息聚合阶段节点自身特征被忽略, 在公式 (23) 的第 1 项和第 2 项中加 1, 这种设计对于提升 KENN 的特征表达是至关重要的. 总的来说, 节点  $i$  在第  $t$  次迭代中的特征向量  $h_i^{(t)}$  是通过图核构造的加权邻接矩阵将节点  $i$  及其邻居在第  $t-1$  层的特征进行聚合, 并将聚合结果输入到多层感知机 (multi-layer perceptron, MLP) 进行学习而得到的.

#### 2.4 复杂性分析

KENN 在计算上是高效的, 其算法的时间复杂度为  $O(kmfh)$ , 内存复杂度为  $O(m)$ , 与图中边的数量为线性关系. 这里的  $m$  表示图数据中的边数,  $k$  表示图神经网络的层数, 而  $f$  和  $h$  分别表示输入和输出特征向量的维度. 通过使用 WL 图核来计算邻接子图的平滑系数, 其时间复杂度是  $O(dtn^2)$ , 其中  $d$  为图数据的平均度数,  $t$  为 WL 图核的迭代次数,  $n$  为图中的节点数. 而使用线性冯·诺依曼熵为节点赋予结构特征的操作可以以线性时间复杂度在所有节点上并行计算. 平滑系数和熵特征可以预先计算并在 KENN 训练阶段直接调用, 从而提高了计算效率.

### 3 实验结果与分析

在本节中, 将进行实验验证 KENN 在图分类任务中的分类能力. 本节从数据集的介绍、基于不同熵的图核对比试验、KENN 与基准方法的对比实验、同类模型间的对比实验、过平滑分析、熵有效性分析、可视化分析及鲁棒性分析 8 个方面进行实验验证与分析. 所有实验都是在单个 NVIDIA 3090 24 GB GPU 上训练/测试的.

#### 3.1 实验数据

本文使用 7 个公开的图分类数据集: 5 个生物信息学数据集 (MUTAG<sup>[29]</sup>, COX2<sup>[30]</sup>, PROTEINS<sup>[31]</sup>, PTC<sup>[32]</sup>, BZR<sup>[27]</sup>) 和 2 个社交网络数据集 (IMDB\_B, IMDB\_M)<sup>[33]</sup>. 一般情况下, 图核以及图神经网络对生物信息学图数据, 采用节点的类别标签作为节点特征, 而对于社交网络这种无节点标签的数据则一般采用节点度的 one-hot 编码作为节点特征. 这种方法易造成图节点的虚拟种类过于冗余, 网络表征不够准确. 而本文方法对于无节点标签的图数据可提取节点所在  $k$  阶子图的  $k$  层熵作为节点特征进行处理, 与传统的对社交网络数据集采用节点度作为节点特征的方法相比更加有效. 数据集统计信息汇总在表 1 中.

表 1 实验数据集详细信息

| Dataset  | Graphs | Classe | Max node | Avg node | Avg edge | Node labels |
|----------|--------|--------|----------|----------|----------|-------------|
| MUTAG    | 188    | 2      | 28       | 17.93    | 19.79    | 7           |
| COX2     | 467    | 2      | 56       | 41.22    | 43.45    | 35          |
| PROTEINS | 1113   | 2      | 620      | 39.06    | 72.82    | 3           |
| PTC      | 344    | 2      | 64       | 14.29    | 14.69    | 18          |
| BZR      | 405    | 2      | 57       | 35.75    | 38.36    | 53          |
| IMDB_B   | 1000   | 2      | 136      | 19.77    | 96.53    | —           |
| IMDB_M   | 1500   | 3      | 89       | 13       | 72.82    | —           |

#### 3.2 不同图核对比实验

本文在 WL 核中尝试融入不同的熵进行实验它们分别是基于香农熵的 WL 图核 (Sh-WL)<sup>[34]</sup>, 基于瑞丽熵的 WL 图核 (Ré-WL)<sup>[35]</sup> 和基于线性冯·诺依曼熵的 WL 图核 (LVN-WL)<sup>[26]</sup>, 并对基于不同熵的 WL 核和原始 WL 核在 7 个数据集上的分类精度进行了对比分析.

本文使用 LIBSVM 中的标准支持向量机 (C-SVM)<sup>[36]</sup> 来进行 10 折交叉验证, 并计算了分类精度. 实验部分对训练数据进行交叉验证, 为每个模型寻找最优参数. 基于 10 折交叉基础上, 对每种方法和数据集重复实验了 10 次并取平均值作为最终实验结果, 在表 2 中展示了平均分类精度和标准误差. 其中, 表 2 中 WL 核与基于 3 种不同熵的 WL 核均在相同的设备环境下执行实验.

实验结果及分析: 表 2 与图 4 表明, 所提出的 Sh-WL、Ré-WL 和 LVN-WL 在图分类任务中的性能显著优于



原始 WL 核. 具体而言, 在生物网络数据的 PROTEINS 和 MUTAG 上, 基于不同熵的 WL 核与原始 WL 相比, 提升最为明显, 分类精度平均提升了 2.48% 和 2.32%. 而在生物网络数据 BZR、PTC 和 COX2 上, 基于不同熵的 WL 核与原始 WL 核相比也具有一定的提升, 分类精度平均提升了 0.83%、0.46 和 2.32%. 在社交网络数据 IMDB\_B 和 IMDB\_M 上, 基于熵的 WL 核与原始 WL 核相比, 精度分别提升了 0.70% 与 0.75%, 这也证明了通过将熵作为节点特征的方法与传统的采用度作为节点特征的方法相比更加有效. 在 3 种基于熵的 WL 核之间, LVN-WL 在生物网络数据 MUTAG、PROTEINS、PTC、BZR 和所有社交网络数据上的分类精度都超越了 Sh-WL 与 Ré-WL, LVN-WL 仅在 COX2 数据上的分类精度与 Sh-WL 相似并优于 Ré-WL. 总的来说, 基于不同熵的 WL 核与原始 WL 核相比, 更有效性的原因主要有以下两个方面.

表 2 图核实验结果

| Method | MUTAG        | COX2         | PROTEINS     | PTC          | BZR          | IMDB_B       | IMDB_M       |
|--------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| WL     | 84.50        | 78.07        | 73.11        | 59.90        | 83.60        | 72.28        | 49.13        |
| Sh-WL  | 86.07        | <b>78.20</b> | 75.20        | 60.12        | 84.35        | 72.80        | 49.62        |
| Ré-WL  | 86.10        | 78.15        | 75.45        | 60.26        | 84.30        | 73.06        | 49.44        |
| LVN-WL | <b>88.30</b> | <b>78.20</b> | <b>76.12</b> | <b>60.70</b> | <b>84.64</b> | <b>73.10</b> | <b>50.60</b> |
| 平均增量   | 2.32         | 0.11         | 2.48         | 0.46         | 0.83         | 0.70         | 0.75         |

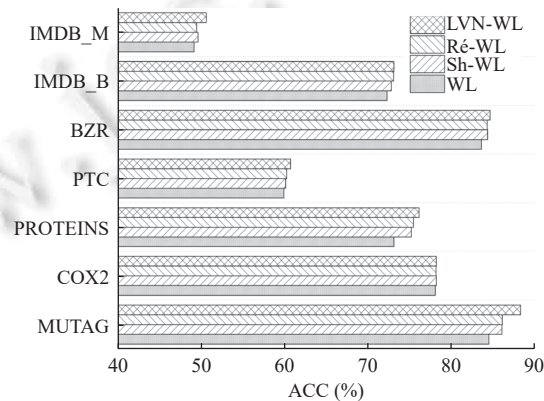


图 4 基于不同熵的图核实验结果

(1) 在生物分子数据集中, 通过对不同数据集的大小和节点标签等基础属性对比发现, MUTAG 和 PROTEINS 的节点标签类别较少, 本文通过将节点子图熵赋值到节点标签中的操作增强了节点类别的表达, 丰富了图数据的结构信息, 所以基于熵的 WL 核在 MUTAG 和 PROTEINS 中具有优异的表现. 相反, 由于 PTC 数据中边的稀疏性, 导致 PTC 数据的图结构单一化, 这导致图熵无法像在其他数据集上一样有效地提取 PTC 的结构信息, 不能区分不同节点子图结构信息的差异.

(2) 在社交网络数据集中, 原有的方案一般是采用节点的度作为节点标签进行处理, 如此操作易引入过多的无效特征, 导致特征的冗余. 而本文采用  $k$  阶图熵提取节点标签的方案使得节点标签代表性更强, 可有效地表征节点周围的结构信息. 社交网络数据集上图中边的数量要远超生物分子数据集, 这样复杂的结构更有利于图熵的结构特征提取, 但同时要求图熵的提取层数要保持一定的数量, 从而防止过平滑现象的出现.

为了探究 3 种不同熵所带来的差异, 将 Sh-WL, Ré-WL 和 LVN-WL 同原始的 WL 核进行对比, 图 5 呈现了基于 3 种不同熵的 WL 图核在 7 种数据集上相对于原始 WL 图核的增长值. LVN-WL 在 MUTAG、PROTEINS、BZR、IMDB\_B 和 IMDB\_M 数据上的精度较原始 WL 核的提升与 Sh-WL 与 Ré-WL 相比最为明显, 而在 COX2 上, LVN-WL 的精度提升与 Ré-WL 相似, 并超越了 Sh-WL, 最后在 PTC 上, 3 种基于熵的 WL 核精度提升基本持平.

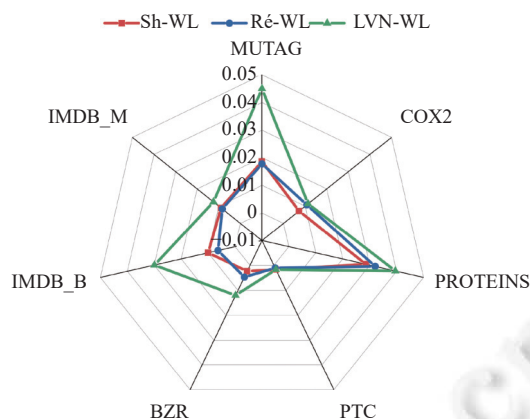


图5 基于不同熵的 WL 图核同原始 WL 图核的精度提升对比分析

### 3.3 不同分类方法的对比实验

本文比较了 12 个基准方法: (1) 基于图核的方法: WL 子树核<sup>[27]</sup>, GNTK<sup>[37]</sup>, WWL<sup>[38]</sup>和 FGW<sup>[39]</sup>; (2) 基于 GNN 的方法: GCN<sup>[14]</sup>, DGCNN<sup>[40]</sup>, GraphSAGE<sup>[41]</sup>, GIN<sup>[20]</sup>, GCKN<sup>[42]</sup>, GSNN<sup>[24]</sup>, SLIM<sup>[43]</sup>和 AdaSNN<sup>[44]</sup>.

为保证实验的公平性, 使用 10 折交叉验证, 选取 90% 作为训练集和 10% 作为测试集, 并展示最佳平均准确率. 遵循与 GSNN 在图分类任务上相同的实验设置, 选取 Adam 优化器, 批处理大小 64, 隐藏维度 64, 0.009 的权重衰减, 具有批处理归一化的 2 层 MLP, 500 代训练测试, 0.6 的节点丢弃, 以及所有数据集上  $\gamma = 0.1$ . 使用 GIN 中的 SUM 读出函数, 拼接所有层的特征以获得最终图的特征.

表 3 展示了 KENN 和其他基准方法的分类精度, 可以看到, 与这些图核以及图神经网络方法相比, 本文的方法表现优异. 在生物分子数据集内对比基准方法中所有的图核模型, 本文的模型在 MUTAG, COX2, PROTEINS, PTC, BZR 上分别比最高精度的图核模型提升了 5.78%, 9.82%, 3.64%, 4.75%, 7.49%. 与不同数据集上最高精度的图神经网络方法相比, 也分别提升了 1.08%, 1.82%, 0.84%, 0.25%, 1.49%. 在社交网络数据集上, 对比表现最优的图核方法 GNTK, 本文的模型在 IMDB\_B, IMDB\_M 上分别提升了 1.90%, 2.33%. 与图神经网络方法中的 GSNN 相比, KENN 在 IMDB\_B 上提升了 0.80%, 而与 SLIM 相比, KENN 在 IMDB\_M 上提升了 1.83%.

表 3 KENN 在图分类数据上运行 10 次的平均分类精度同基准方法的对比结果 (%)

| Method      | MUTAG              | COX2               | PROTEINS           | PTC                | BZR                | IMDB_B             | IMDB_M             |
|-------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|--------------------|
| WL          | 84.5               | 78.0               | 73.1               | 59.9               | 83.6               | 72.7               | 49.1               |
| WWL         | 87.2 ± 1.5         | 78.2 ± 0.4         | 74.2 ± 0.5         | 66.3 ± 1.2         | 84.4 ± 2.0         | 74.3 ± 0.8         | —                  |
| FGW         | 88.4 ± 5.6         | 77.2 ± 4.8         | 74.5 ± 2.7         | 65.3 ± 7.9         | 85.1 ± 4.1         | 63.8 ± 3.4         | 48.0 ± 3.2         |
| GNTK        | 90.0 ± 8.5         | —                  | 75.6 ± 4.2         | 67.9 ± 6.9         | 83.6 ± 2.9         | 76.9 ± 3.6         | 52.8 ± 4.6         |
| GCN         | 85.6 ± 5.8         | —                  | 76.0 ± 3.2         | 64.2 ± 4.3         | —                  | 74.0 ± 3.4         | 51.9 ± 3.8         |
| DGCNN       | 85.8 ± 1.6         | —                  | 75.5 ± 0.9         | 58.6 ± 2.4         | —                  | 70.0 ± 0.8         | 47.8 ± 0.8         |
| GraphSAGE   | 85.1 ± 7.6         | —                  | 75.9 ± 3.2         | 63.9 ± 7.7         | —                  | 68.8 ± 4.5         | 49.9 ± 5.0         |
| GIN         | 89.4 ± 5.6         | —                  | 76.2 ± 2.8         | 64.6 ± 7.0         | —                  | 75.1 ± 5.1         | 52.3 ± 2.8         |
| GCKN        | 91.6 ± 6.7         | —                  | 76.2 ± 2.5         | 68.4 ± 7.4         | —                  | 76.5 ± 5.7         | 53.3 ± 3.9         |
| GSNN        | 94.7 ± 1.9         | 86.2 ± 3.3         | 78.4 ± 2.7         | 70.5 ± 3.1         | 91.1 ± 3.0         | 78.0 ± 2.8         | —                  |
| SLIM        | 93.2 ± 3.3         | —                  | 77.4 ± 4.3         | 72.4 ± 6.9         | —                  | 77.2 ± 2.1         | 53.3 ± 4.0         |
| AdaSNN      | 87.2 ± 5.0         | —                  | 76.5 ± 2.6         | 60.2 ± 6.4         | —                  | 74.2 ± 2.5         | 51.9 ± 4.7         |
| <b>KENN</b> | <b>95.78 ± 2.1</b> | <b>88.02 ± 3.2</b> | <b>79.24 ± 4.4</b> | <b>72.65 ± 3.3</b> | <b>92.59 ± 1.5</b> | <b>78.80 ± 2.8</b> | <b>55.13 ± 4.0</b> |

为探究模型的稳定性, 绘制了模型迭代训练中的损失与精度随训练迭代次数变化的折线图, 如图 6 所示. 通过分析 COX2, PROTEINS, PTC, IMDB\_B 图示可以得出, 训练损失普遍在模型训练的初期高于测试损失, 但随着训

训练迭代次数的叠加, 训练损失大致在 100–200 epoch 间低于测试损失, 随后训练与测试损失将会保持相似的步调缓慢下降, 直到 300–400 epoch 时达到平衡. 同样地, 精度与损失的变化趋势有着相似之处, 在 200 epoch 前精度上升趋势明显, 200–300 epoch 模型逐渐收敛, 精度缓慢提升, 300 epoch 后模型趋于平衡, 精度开始上下浮动. 再分析 MUTAG 与 BZR 数据, 观察图 6 发现, 训练损失虽也在 100–200 epoch 间下降趋势超越测试损失, 但在之后的训练过程中, 测试损失与训练损失存在多次交叉, 这也间接导致精度在 300 epoch 后还存在较大的浮动变化, 这说明模型在 MUTAG 与 BZR 数据上的拟合效果略低于其余 4 个数据集, 但这不影响本文模型优异的特征能力.

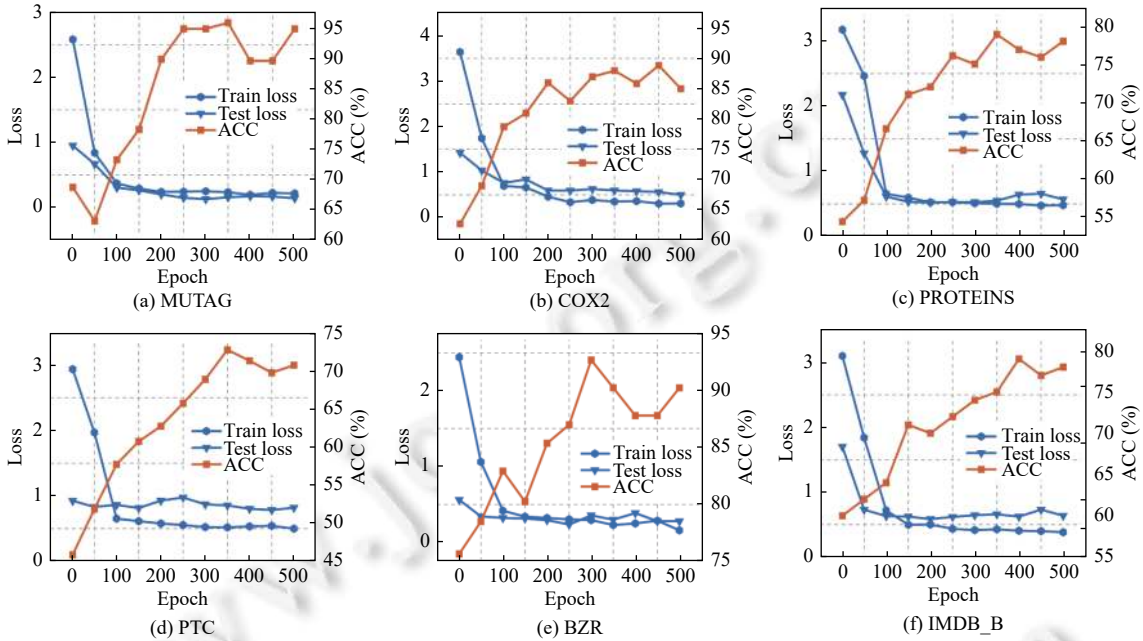


图 6 KENN 在 6 种数据集上的训练和测试的损失精度图

### 3.4 同类模型间的对比实验

由于本文的模型与 GSNN 相似, 同是在邻接矩阵内进行平滑权重的分配, 但通过不同的方式为邻居之间赋予权值, 为了探究 KENN 与 GSNN 相比是否具有更强的表现, 本节绘制了实验对比箱线图如图 7 所示, 图中展示了 KENN 与 GSNN 在 6 个公共数据集上分别运行 10 次后的结果, 在图中 PROTEINS 缩写为 PRO.

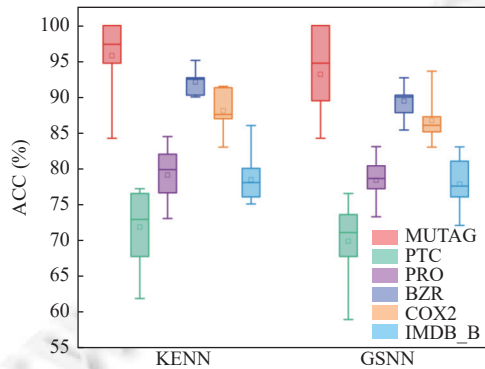


图 7 KENN 与 GSNN 10 折交叉实验结果分布箱线图

由图 7 可以看出, 在 PTC, BZR, IMDB\_B 数据上, 本文的方法无论是最大值还是最小值都超越了 GSNN. 在 MUTAG, PROTEINS 上最大值和最小值虽然与 GSNN 基本持平, 但 KENN 预测的平均准确率更高, 中位线以及

平均值都超越了 GSNN. 而在 COX2 数据中, 虽然 KENN 在精度最大值方面的表现低于 GSNN, 但在平均值以及数据结果的中位线上均优于对比方法. 所以总的来说, KENN 在图分类数据集上表现优异.

### 3.5 过平滑分析

为了探究 KENN 是否会产生过度平滑现象, 本节使用节点分类任务来分析模型深度 (层数) 对节点特征嵌入的影响, 并选取 GCN<sup>[14]</sup>, GIN<sup>[20]</sup>, DFNet<sup>[7]</sup>作为对比模型. 在 Cora (节点分类) 数据集上, 采用了 Adam 优化器, 学习率为 0.001, 训练了 200 个 epoch 在权重衰减的取值范围 {0.001, 0.002, ..., 0.009} 中选择了最佳值, 并使用隐藏单位 {64, 128, 256, 512}. 每个层的 dropout 设置为 0.9. Cora 数据集的划分方法遵循了文献 [14] 中的标准划分, 即每个类别使用 20 个节点进行训练, 500 个节点用于验证, 1000 个节点用于测试.

表 4 表明, 随着模型深度的增加, KENN 除在第 3 层略逊于 DFNet 之外, 在其余的每一层中, 性能都优于基线模型. 这是因为 KENN 利用图核定义的同构系数捕获了目标顶点与邻居之间的结构连接, 强化了消息在邻居之间的传递. 因此, 结构连接较强的邻居能够向目标顶点传递更多的信息, 从而有助于缓解过平滑问题.

表 4 过平滑实验在 Cora 数据集上运行 10 次的平均分类精度 (%)

| GNN layers | GCN               | GIN               | DFNet             | KENN              |
|------------|-------------------|-------------------|-------------------|-------------------|
| 1          | 79.5 ± 0.6        | 74.2 ± 1.4        | 80.5 ± 0.9        | 80.6 ± 0.6        |
| 2          | <b>81.5 ± 0.4</b> | <b>77.6 ± 1.3</b> | 81.6 ± 0.7        | <b>82.9 ± 0.8</b> |
| 3          | 80.9 ± 0.5        | 76.1 ± 1.7        | <b>82.3 ± 0.4</b> | 82.1 ± 0.7        |
| 4          | 78.3 ± 0.8        | 50.6 ± 2.2        | 80.5 ± 0.7        | 81.7 ± 1.2        |
| 5          | 75.2 ± 1.2        | 42.3 ± 1.8        | 74.8 ± 0.4        | 81.2 ± 1.5        |
| 6          | 36.5 ± 1.6        | 38.1 ± 2.3        | 66.3 ± 1.2        | 77.5 ± 1.1        |
| 7          | 32.7 ± 0.9        | 29.5 ± 2.1        | 60.9 ± 1.5        | 75.9 ± 1.3        |
| 8          | 20.6 ± 1.2        | 25.4 ± 1.8        | 54.2 ± 1.4        | 76.7 ± 0.9        |

### 3.6 熵有效性分析

为了验证熵的有效性, 本节采用了一系列实验来评估不同层数的熵对图分类任务的影响. 本节选择与第 3.3 节图分类任务相同的数据集作为测试数据, 并在生物分子数据与社交网络数据集上进行详尽的实验比较. 通过逐步增加熵的层数, 从 0 层到 8 层, 并分别评估每个层数下的模型性能.

表 5 的结果表明 (表中 E layers 代表熵的层数), 在生物分子数据集上, 4 层熵的表现最佳. 这可能是因为在生物分子数据集上的图结构相对简单, 通过堆叠多层熵可以更好地表征顶点周围的结构信息. 相反, 在社交网络数据集上, 2 层熵的效果最好. 这是因为社交网络数据集上的图结构更加复杂, 顶点的邻居数量较多, 而较少的熵层数能更有效地捕捉到局部的图结构信息. 同时, 在部分数据集上提取过多层数的熵会引入冗余信息, 从而导致模型性能下降.

表 5 引入不同层数熵的 KENN 在图分类数据集上运行 10 次的平均分类精度的对比结果 (%)

| E layers | MUTAG             | COX2              | PROTEINS          | PTC               | BZR               | IMDB_B            | IMDB_M            |
|----------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| 0        | 92.2 ± 3.2        | 86.1 ± 3.2        | 77.9 ± 3.4        | 70.2 ± 3.1        | 90.7 ± 3.2        | 77.2 ± 3.0        | 53.8 ± 3.1        |
| 2        | 93.5 ± 2.6        | 87.4 ± 4.3        | 78.5 ± 3.0        | 71.3 ± 4.7        | 92.4 ± 2.8        | <b>78.8 ± 2.8</b> | <b>55.1 ± 4.0</b> |
| 4        | <b>95.7 ± 3.7</b> | <b>88.0 ± 2.3</b> | <b>79.2 ± 4.4</b> | <b>72.6 ± 3.3</b> | <b>92.5 ± 1.5</b> | 78.2 ± 3.2        | 54.5 ± 3.2        |
| 6        | 95.2 ± 4.5        | 87.5 ± 2.8        | 78.1 ± 4.2        | 71.5 ± 5.1        | 91.8 ± 3.1        | 77.4 ± 3.5        | 54.1 ± 3.6        |
| 8        | 93.5 ± 4.2        | 86.0 ± 4.2        | 77.5 ± 5.2        | 69.7 ± 5.2        | 91.2 ± 2.8        | 77.1 ± 2.7        | 53.5 ± 4.2        |

综上所述, 本节的实验结果表明熵在图神经网络分类任务中是有效的. 因为多层熵可以提取数据集的局部结构特征, 所以选择适当的熵层数对不同类型的数据集而言是至关重要的. 在未来的研究中, 可以进一步探索其他图结构属性和熵之间的关系, 以提高图分类任务的性能和鲁棒性.

### 3.7 可视化分析

为了更好地验证 KENN 在识别图结构信息方面的有效性, 本节探讨 KENN 与 GIN 对多层平滑后的 MUTAG 数据节点特征的差异性. 如图 8 所示, 通过图核为节点融入结构信息, 定义不同邻居间平滑权重后的节点特征差异

性要明显高于不带图核的网络. 这表明, KENN 一方面增强了节点的特征表达能力, 另一方面, 节点特征差异性增大后也会减少多层网络带来的过平滑现象.

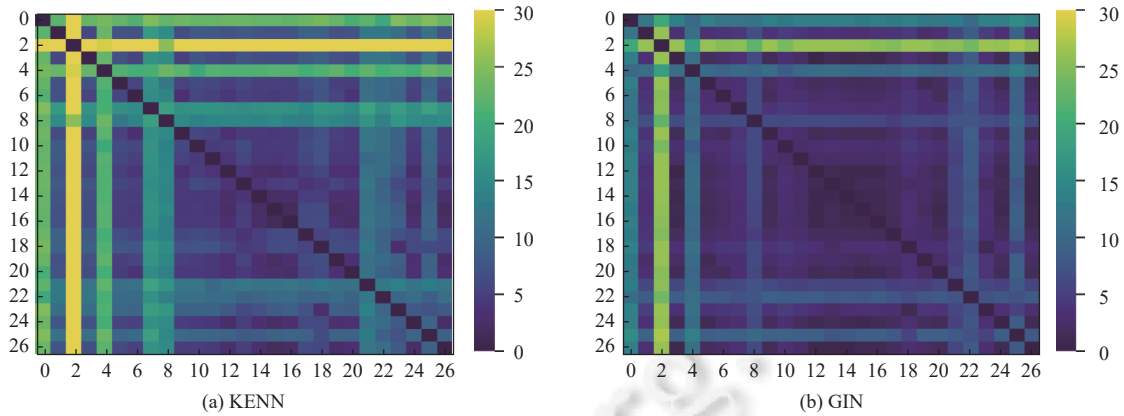


图 8 利用图核融入结构信息的 MUTAG 数据集

如图 9 所示, 给出两个非同构的 6-正则图 (图 9(a)). 传统信息聚合图神经网络通过标签子树对节点进行编码的过程将对应节点映射到相同的子树上 (图 9(b)), 导致节点  $v$  和  $v'$  的不同无法被区分. 同样地, 经过信息聚合后图 (图 9(a)) 中的其余对应节点均形成相同的特征编码. 这说明传统图神经网络没有区分类似的正则图的能力. 相反, KENN 将图熵赋予到节点特征中, 并利用图核判断邻接子图同构性的方式构造邻接权重矩阵 (图 9(c)). 在信息聚合阶段, 节点  $v$  和  $v'$  利用邻接权重生成权重标签子树进行编码 (图 9(d)), 从而生成不同的节点特征. 同理, 当对应的正则图内所有节点经过信息聚合后, 生成的图特征是可以被区分的.

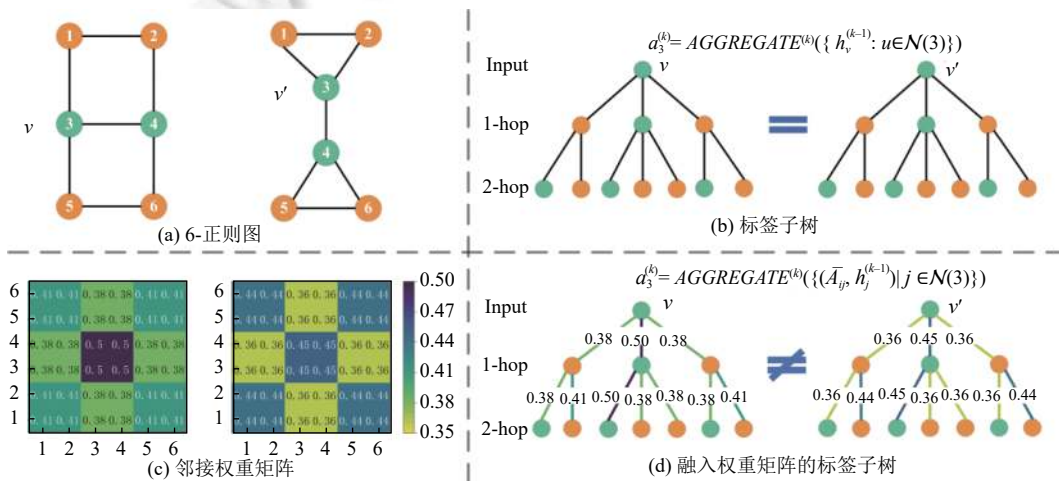


图 9 KENN 与传统信息聚合图神经网络的比较

### 3.8 鲁棒性分析

为了评估 KENN 的鲁棒性, 本实验通过在数据集 IMDB\_B 中分别删除边和增加边来生成新数据集. 具体来说, 对于每个图, 随机删除 (如果有边) 或增加 (如果无边) 25%、50% 和 75% 的边. 图 10 展示了普通 GCN、最具竞争力的基线 GSNN 和本文方法的结果, 其中实线是平均精度, 阴影区域是 5 次运行的标准差, x 轴表示删除或增加边的比例.

如图 10 所示, 所有方法的准确率都随着噪声程度的增加而下降, 其中 GCN 的准确率下降最为明显. 由于

GSNN 和 KENN 可以构造带有结构信息的边, 它们比 GCN 具有更好的鲁棒性, 本文的 KENN 在性能退化较少的情况下仍能取得较好的结果.

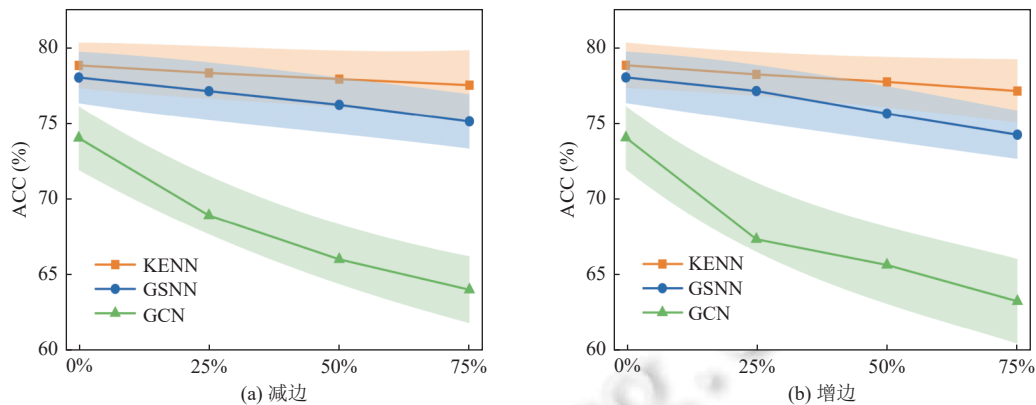


图 10 在 IMDB\_B 数据上, 边干扰情况下的测试精度 ( $\pm$ 标准差)

## 4 总结

本文提出线性结构熵的图核神经网络模型 (KENN). 首先将线性冯·诺依曼熵应用到图核方法中, 加强了图核对图特征提取的结构表达性, 解决了生物分子数据节点特征稀疏的问题, 以及社交网络数据以节点度为特征所产生的信息冗余问题. 然后将基于线性结构熵的图核与消息聚合图神经网络相结合, 利用图核强大的图结构信息表征能力为图中不同邻居节点间定义不同的平滑权重, 丰富了图节点嵌入后的特征信息. 在真实数据集上的对比实验表明, 本文模型对比所有的基准方法在图分类任务上表现优异. 将来, 可考虑利用随机游走算法对子图进行提取, 通过学习来提取出图中例如苯环等特殊结构的完整表示, 加强节点结构属性的准确度.

## References:

- [1] Ou MD, Cui P, Pei J, Zhang ZW, Zhu WW. Asymmetric transitivity preserving graph embedding. In: Proc. of the 22nd ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. San Francisco: ACM, 2016. 1105–1114. [doi: 10.1145/2939672.2939751]
- [2] Xu LX, Bai L, Xiao J, Liu Q, Chen EH, Wang XF, Tang YY. Multiple graph kernel learning based on GMDH-type neural network. Information Fusion, 2021, 66: 100–110. [doi: 10.1016/j.inffus.2020.08.025]
- [3] Bai L, Zhang ZH, Wang CY, Bai X, Hancock ER. A graph kernel based on the Jensen-Shannon representation alignment. In: Proc. of the 24th Int'l Conf. on Artificial Intelligence. Buenos Aires: AAAI Press, 2015. 3322–3328.
- [4] Zhang XY, Zou JH, He KM, Sun J. Accelerating very deep convolutional networks for classification and detection. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2016, 38(10): 1943–1955. [doi: 10.1109/TPAMI.2015.2502579]
- [5] Dong C, Loy CC, He KM, Tang XO. Image super-resolution using deep convolutional networks. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2016, 38(2): 295–307. [doi: 10.1109/TPAMI.2015.2439281]
- [6] You JX, Ying R, Leskovec J. Position-aware graph neural networks. In: Proc. of the 36th Int'l Conf. on Machine Learning. Long Beach: JMLR.org, 2019. 7134–7143.
- [7] Wijesinghe WOK, Wang Q. DFNet: Spectral CNNs for graphs with feedback-looped filters. In: Proc. of the 33rd Int'l Conf. on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2019. 6009–6020.
- [8] Bianchi FM, Grattarola D, Alippi C. Spectral clustering with graph neural networks for graph pooling. In: Proc. of the 37th Int'l Conf. on Machine Learning. Online: JMLR.org, 2020. 874–883.
- [9] Liu PJ, Li HY, Wang TY, Liu H, Sun LM, Ren YF, Li CP, Chen H. Multi-stage method for online vertical data partitioning based on spectral clustering. Ruan Jian Xue Bao/Journal of Software, 2023, 34(6): 2804–2832 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/6496.htm> [doi: 10.13328/j.cnki.jos.006496]
- [10] Zhou CY, Zeng C, He P, Zhang Y. GKCI: An improved GNN-based key class identification method. Ruan Jian Xue Bao/Journal of Software, 2023, 34(6): 2509–2525 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/6846.htm> [doi: 10.13328/j.cnki.jos.006846]

- [11] Long QQ, Jin YL, Wu Y, Song GJ. Theoretically improving graph neural networks via anonymous walk graph kernels. In: Proc. of the 2021 Web Conf. Ljubljana: ACM, 2021. 1204–1214. [doi: 10.1145/3442381.3449951]
- [12] Bruna J, Zaremba W, Szlam A, LeCun Y. Spectral networks and deep locally connected networks on graphs. In: Proc. of the 2nd Int'l Conf. on Learning Representations. Banff, 2014.
- [13] Defferrard M, Bresson X, Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering. In: Proc. of the 30th Int'l Conf. on Neural Information Processing Systems. Barcelona: Curran Associates Inc., 2016. 3844–3852.
- [14] Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. arXiv:1609.02907, 2017.
- [15] Zhang WT, Sheng ZA, Yang MY, Li Y, Shen Y, Yang Z, Cui B. NAFS: A simple yet tough-to-beat baseline for graph representation learning. In: Proc. of the 39th Int'l Conf. on Machine Learning. Baltimore: PMLR, 2022. 26467–26483.
- [16] Yan Z, Xiang XG, Li ZC. Item correlation modeling in interaction sequence for graph convolutional session recommendation. *Scientia Sinica Informationis*, 2022, 52(6): 1069–1082 (in Chinese with English abstract). [doi: 10.1360/SSI-2020-0383]
- [17] Hu YT, Wang SY, Wu YM, Zou DQ, Li WK, Jin H. Slice-level vulnerability detection and interpretation method based on graph neural network. *Journal of Software*, 2023, 34(6): 2543–2561 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/6849.htm> [doi: 10.13328/j.cnki.jos.006849]
- [18] Li ZC, Tang JH. Semi-supervised local feature selection for data classification. *Science China Information Sciences*, 2021, 64(9): 192108. [doi: 10.1007/s11432-020-3063-0]
- [19] Li ZC, Tang JH, Mei T. Deep collaborative embedding for social image understanding. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2019, 41(9): 2070–2083. [doi: 10.1109/TPAMI.2018.2852750]
- [20] Xu KYL, Hu WH, Leskovec J, Jegelka S. How powerful are graph neural networks? arXiv:1810.00826, 2019.
- [21] Murphy R, Srinivasan B, Rao V, Ribeiro B. Relational pooling for graph representations. In: Proc. of the 36th Int'l Conf. on Machine Learning. Long Beach: PMLR, 2019. 4663–4673.
- [22] Sato R, Yamada M, Kashima H. Random features strengthen graph neural networks. In: Proc. of the 2021 SIAM Int'l Conf. on Data Mining (SDM). Virtual Event: Society for Industrial and Applied Mathematics, 2021. 333–341. [doi: 10.1137/1.9781611976700.38]
- [23] Li P, Wang YB, Wang HW, Leskovec J. Distance encoding: Design provably more powerful neural networks for graph representation learning. In: Proc. of the 34th Int'l Conf. on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2020. 4465–4478.
- [24] Wijesinghe A, Wang Q. A new perspective on “how graph neural networks go beyond Weisfeiler-Lehman?” In: Proc. of the 10th Int'l Conf. on Learning Representations. OpenReview.net, 2022. 1–23.
- [25] Passerini F, Severini S. The von Neumann entropy of networks. arXiv:0812.2597, 2012.
- [26] Chen PY, Wu LF, Liu SJ, Rajapakse I. Fast incremental von Neumann graph entropy computation: Theory, algorithm, and applications. In: Proc. of the 36th Int'l Conf. on Machine Learning. Long Beach: PMLR, 2019. 1091–1101.
- [27] Shervashidze N, Schweitzer P, van Leeuwen EJ, Mehlhorn K, Borgwardt KM. Weisfeiler-Lehman graph kernels. *The Journal of Machine Learning Research*, 2011, 12: 2539–2561.
- [28] Borgwardt K, Ghisu E, Llinares-López F, O'Bray L, Rieck B. Graph kernels: State-of-the-art and future challenges. *Foundations and Trends® in Machine Learning*, 2020, 13(5–6): 531–712. [doi: 10.1561/22000000076]
- [29] Debnath AK, Lopez de Compadre RL, 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]
- [30] 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]
- [31] Sutherland JJ, O'Brien LA, Weaver DF. Spline-fitting with a genetic algorithm: A method for developing classification structure-Activity relationships. *Journal of Chemical Information and Computer Sciences*, 2003, 43(6): 1906–1915. [doi: 10.1021/ci034143r]
- [32] Kriege NM, Giscard PL, Wilson RC. On valid optimal assignment kernels and applications to graph classification. In: Proc. of the 30th Int'l Conf. on Neural Information Processing Systems. Barcelona: Curran Associates Inc., 2016. 1623–1631.
- [33] Yanardag P, Vishwanathan SVN. Deep graph kernels. In: Proc. of the 21st ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. Sydney: ACM, 2015. 1365–1374. [doi: 10.1145/2783258.2783417]
- [34] Lin J. Divergence measures based on the Shannon entropy. *IEEE Trans. on Information Theory*, 1991, 37(1): 145–151. [doi: 10.1109/18.61115]
- [35] Xu LX, Bai L, Jiang XY, Tan M, Zhang DQ, Luo B. Deep Rényi entropy graph kernel. *Pattern Recognition*, 2021, 111: 107668. [doi: 10.1016/j.patcog.2020.107668]
- [36] Oku K, Nakajima S, Miyazaki J, Uemura S. Context-aware SVM for context-dependent information recommendation. In: Proc. of the 7th Int'l Conf. on Mobile Data Management (MDM 2006). Nara: IEEE, 2006. 109–109. [doi: 10.1109/MDM.2006.56]

- [37] Du SS, Hou KC, Póczos B, Salakhutdinov RR, Wang RS, Xu KYL. Graph neural tangent kernel: Fusing graph neural networks with graph kernels. In: Proc. of the 33rd Int'l Conf. on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2019. 5723–5733.
- [38] Toginalli M, Ghisu E, Llinares-López F, Rieck B, Borgwardt K. Wasserstein Weisfeiler-Lehman graph kernels. In: Proc. of the 33rd Int'l Conf. on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2019. 6439–6449.
- [39] Titouan V, Courty N, Tavenard R, *et al.* Optimal transport for structured data with application on graphs. In: Proc. of the 36th Int'l Conf. on Machine Learning. Long Beach: PMLR, 2019. 6275–6284.
- [40] Zhang MH, Cui ZC, Neumann M, Chen YX. An end-to-end deep learning architecture for graph classification. Proc. of the AAAI Conf. on Artificial Intelligence, 2018, 32(1): 4438–4445. [doi: 10.1609/aaai.v32i1.11782]
- [41] 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. Vancouver: Curran Associates Inc., 2017. 1025–1035.
- [42] Chen DX, Jacob L, Mairal J. Convolutional kernel networks for graph-structured data. In: Proc. of the 37th Int'l Conf. on Machine Learning. Virtual Event: JMLR. org, 2020. 1576–1586.
- [43] Zhu YK, Zhang K, Wang J, Ling HB, Zhang J, Zha H. Structural landmarking and interaction modelling: A “SLIM” network for graph classification. Proc. of the AAAI Conf. on Artificial Intelligence, 2022, 36(8): 9251–9259. [doi: 10.1609/aaai.v36i8.20912]
- [44] Li JX, Sun QY, Peng H, Yang BN, Wu J, Yu PS. Adaptive subgraph neural network with reinforced critical structure mining. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2023, 45(7): 8063–8080. [doi: 10.1109/TPAMI.2023.3235931]

#### 附中文参考文献:

- [9] 刘鹏举, 李好洋, 王天一, 刘欢, 孙路明, 任逸飞, 李翠平, 陈红. 基于谱聚类的在线数据库垂直分区多阶段生成方法. 软件学报, 2023, 34(6): 2804–2832. <http://www.jos.org.cn/1000-9825/6496.htm> [doi: 10.13328/j.cnki.jos.006496]
- [10] 周纯英, 曾诚, 何鹏, 张龔. GKCI: 改进的基于图神经网络的关键类识别方法. 软件学报, 2023, 34(6): 2509–2525. <http://www.jos.org.cn/1000-9825/6846.htm> [doi: 10.13328/j.cnki.jos.006846]
- [16] 闫昭, 项欣光, 李泽超. 基于交互序列商品相关性建模的图卷积会话推荐. 中国科学: 信息科学, 2022, 52(6): 1069–1082. [doi: 10.1360/SSI-2020-0383]
- [17] 胡雨涛, 王溯远, 吴月明, 邹德清, 李文科, 金海. 基于图神经网络的切片级漏洞检测及解释方法. 软件学报, 2023, 34(6): 2543–2561. [doi: 10.13328/j.cnki.jos.006849]



徐立祥(1981—), 男, 博士, 教授, CCF 高级会员, 主要研究领域为机器学习, 数据挖掘, 推荐系统.



罗斌(1963—), 男, 博士, 教授, 博士生导师, CCF 专业会员, 主要研究领域为模式识别, 机器学习, 数据挖掘.



许巍(1998—), 男, 硕士生, 主要研究领域为机器学习, 数据挖掘.



唐远炎(1943—), 男, 博士, 教授, 主要研究领域为模式识别与智能系统, 信号处理, 图像处理, 小波和多尺度的理论及其应用, 并行处理, 超大规模集成电路结构, 自然语言处理.



陈恩红(1968—), 男, 博士, 教授, 博士生导师, CCF 会士, 主要研究领域为机器学习, 数据挖掘, 教育大数据分析, 个性化推荐.