

# 图神经架构搜索综述

张子威 王鑫 朱文武

(清华大学计算机科学与技术系 北京 100084)

**摘要** 图数据可以广泛建模事物之间的复杂关系,小到蛋白质中的分子与氨基酸结构,大到世界范围的物流与交通网络;从人类社会的社交网络,到信息空间的互联网,均可统一表示为图数据的形式.图数据中蕴藏着巨大的研究与应用价值.图神经网络是过去几年中图数据上进行机器学习的主要范式.通过在图数据的链接关系上重新定义神经网络架构并实现端到端的学习,图神经网络可以有效处理节点分类、链接预测、图分类等多种图数据分析与挖掘任务.然而,由于图数据的复杂性、图任务的多样性以及图神经网络的复杂程度,人工设计最优的图神经网络架构变得越来越困难,且无法适应开放变化环境.图神经架构搜索,旨在自动化设计针对特定数据集与任务的最优图神经网络架构,应运而生并逐渐受到了学术界和工业界的关注.在本文中,我们对图神经架构搜索这一快速发展的新兴领域进行综述.特别地,我们系统总结并梳理了目前已公开发表的四十余篇图神经架构搜索算法,并从搜索空间、搜索策略、模型性能评估策略以及其他特点对已有算法进行了全面的分类、对比与评述,并从实验上对上述算法进行了归纳.此外,我们还对近期的图神经架构搜索研究趋势进行了评述.最后,我们分享了对图神经架构搜索未来研究方向的想法.

**关键词** 图神经网络;神经架构搜索;图机器学习;自动机器学习;人工智能

**中图法分类号** TP18 **DOI号** 10.11897/SP.J.1016.2023.01532

## Graph Neural Architecture Search: A Survey

ZHANG Zi-Wei WANG Xin ZHU Wen-Wu

(Department of Computer Science and Technology, Tsinghua University, Beijing 100084)

**Abstract** Graph data can generally model the complex relationships between entities. From as small as the molecular and amino acid structures in proteins, to as large as the worldwide logistics and transportation networks; from the social networks in the human society to the Internet in the information space, all these data can be uniformly represented as graphs. Huge research and application values exist underlying the graph data. Graph neural networks are the dominant paradigm for machine learning on graphs over the past few years. By redefining neural network architectures on the link relationship of graph data and realizing end-to-end learning paradigms, graph neural networks can effectively handle a variety of graph analytical and mining tasks such as node classification, link prediction, and graph classification. However, due to the complexity of graph data, the diversity of graph tasks, and the complex architecture of graph neural networks, it becomes increasingly difficult to manually design the optimal graph neural network architectures, failing to adapt to open and changing environments. Graph neural architecture search, which aims to automate the design of optimal graph neural network architectures for

收稿日期:2022-09-09;在线发布日期:2023-04-09.本课题得到科技创新2030——“新一代人工智能重大项目”(2020AAA0106300)、国家自然科学基金项目(62250008,62222209,62102222,61936011,62206149)、北京信息科学与技术国家研究中心(BNR2023RC01003)、网络多媒体北京市重点实验室、博士后创新人才支持计划(BX20220185)、中国博士后科学基金面上项目(2022M711813)资助.张子威,博士,助理研究员,中国计算机学会(CCF)会员,主要研究方向为图机器学习,包括图神经网络、网络嵌入、图自动机器学习等. E-mail: zwzhang@tsinghua.edu.cn.王鑫,博士,助理研究员,中国计算机学会(CCF)会员,主要研究方向为媒体智能、机器学习、媒体大数据.朱文武(通信作者),博士,教授,中国计算机学会(CCF)会员,主要研究领域为多媒体计算、网络多媒体与大数据分析. E-mail: wwzhu@tsinghua.edu.cn.

specific data sets and tasks, comes into being and has attracted considerable attention from both academia and industry. In this paper, we provide a comprehensive review for the rapidly evolving and emerging field of graph neural architecture search. In particular, we systematically review and summarize more than 40 published graph neural architecture search algorithms, and comprehensively classify, compare and comment on the existing algorithms from the search space, the search strategy, the model evaluation strategy, and other characteristics of the models. We also summarize the above algorithms from the experimental point of view. In addition, we analyze recent trends in graph neural architecture search studies. Finally, we share our insights on the future research direction of graph neural architecture search.

**Keywords** graph neural network; neural architecture search; graph machine learning; automated machine learning; artificial intelligence

## 1 引言

图(Graph)是一种建模事物之间复杂关系的通用数据形式。小到蛋白质中的分子与氨基酸结构,大到世界范围的物流与交通网络;从人类社会的社交网络,到信息空间的互联网,均可统一抽象为图数据的形式<sup>[1]</sup>。在图中,节点表示实体,边代表实体之间的链接关系。图数据无所不在,且其中蕴藏着巨大的研究与应用价值。因此,图数据分析与挖掘是人工智能的一个重要研究方向。然而,相比于语音、图片、文本等数据类型,图数据需要建模事物间的复杂联系,具有独特的结构与性质。如何将机器学习等人工智能算法应用于图数据分析与挖掘是一个重要的挑战,正如图灵奖得主约书亚本吉奥(Yoshua Bengio)指出:“将神经网络扩展到图结构输入是机器学习当前的主要挑战之一”<sup>[2]</sup>。

图神经网络(Graph Neural Network, GNN)是过去几年中图数据上进行机器学习的主要范式<sup>[3]</sup>。通过在图数据的链接关系上重新定义神经网络架构并实现端到端的学习,图神经网络可以有效处理节点分类、链接预测、图分类等多种图数据分析与挖掘任务。早期的图神经网络可以追溯到2010年之前,早于深度学习的兴起,例如基于递归定义的图神经网络<sup>[4]</sup>以及基于上下文定义的图神经网络<sup>[5]</sup>。在深度学习兴起之后,图神经网络从2014年再次受到研究者关注,并历经了从频域<sup>[5]</sup>到空域<sup>[6]</sup>的发展。目前主流的图神经网络往往符合一个消息传递框架<sup>[7]</sup>,即节点间相互传递消息以不断更新自己的表征(在第2节中,我们会对图神经网络进行更详细的介绍)。代表性的图神经网络包括切比雪夫图卷积网络

(ChebNet)<sup>[8]</sup>、图卷积神经网络(Graph Convolutional Network, GCN)<sup>[6]</sup>、图注意力神经网络(Graph Attention Network, GAT)<sup>[9]</sup>、图采样聚合神经网络(Graph Sample and Aggregate Network, GraphSAGE)<sup>[10]</sup>、图同构神经网络(Graph Isomorphism Network, GIN)<sup>[11]</sup>等等。更多关于图神经网络的介绍,可参见相关综述<sup>[12-13]</sup>。图神经网络已经在多个领域取得了优异的效果,包括但不限于推荐系统<sup>[14-15]</sup>、自然语言处理<sup>[16]</sup>、生物信息学<sup>[17-18]</sup>、交通预测<sup>[19-20]</sup>、知识图谱<sup>[21]</sup>等。

然而,考虑到图数据的复杂性和图任务的多样性,在开放变化的场景中,为每个数据集和任务单独设计最优的图神经网络会消耗极大的人力成本且缺乏灵活性。此外,由于图神经网络的研究逐渐增多,图神经网络的结构日益复杂,因此人工设计最优的图神经网络架构变得越来越困难。另一方面,神经架构搜索(Neural Architecture Search, NAS)<sup>[22]</sup>是自动机器学习(Automated Machine Learning, AutoML)<sup>[23-25]</sup>的一个核心研究方向,并已经在计算机视觉等领域<sup>[26-27]</sup>取得了优异的效果,可以达到甚至超越人工设计的神经网络架构<sup>[28-29]</sup>。因此,图神经架构搜索(Graph Neural Architecture Search)应运而生,旨在自动化设计针对特定图数据集与图任务的最优图神经网络架构。图神经架构搜索从2019年以来逐渐受到了学术界和工业界的关注,并涌现出了许多的相关研究。然而,与将其他领域神经架构搜索直接应用于图数据不同,图神经架构搜索在结合图神经网络与神经架构搜索研究时,存在许多重要问题与技术挑战,包括以下几方面:

(1)图数据与图神经网络的特殊性。由于图数据并非像文本、图片一样可以表示为网格(grid)结

构,而是存在于一个非欧几里得空间,因此图神经网络与其他神经网络的架构存在较大区别.如何针对图数据和图神经网络的特殊性设计图神经架构搜索空间,是首先要解决的基础问题.本文将在第 3.1 节与第 4.1 节介绍图神经架构搜索空间的设计.

(2) 图神经架构搜索的高效性. 由于图数据中不同实体耦合在一起,因此如何提升图神经架构搜索策略的高效性也是一个较大挑战,并包含两个问题:如何提升搜索过程的效率以及如何提升搜索得到图神经网络架构的效率. 本文将在第 3.2 节与第 4.2 节介绍图神经架构搜索策略的设计.

(3) 图数据的多样性. 由于图是对关系数据的一般抽象,图数据本身即存在多样性并有丰富的类型,例如有向/无向图、有权/无权图、异构图、动态图等. 如何处理不同类型的图数据更增加了图神经架构搜索的挑战. 此外,当用于解决具体领域问题时,如何容纳进领域知识(domain knowledge),也是图神经架构搜索的一个关键问题. 该挑战在每个图神经架构搜索研究中均会被显式或隐式的涉及到,也促进了不同图神经架构搜索的研究.

在本文中,我们对图神经架构搜索这一快速发

展的新兴领域进行综述. 特别地,我们系统总结并梳理了目前已公开发表的四十余篇图神经架构搜索算法(见表 1),并从搜索空间、搜索策略、模型评估策略以及其它特点这四个角度,对已有算法进行了全面的分类、对比与评述. 总结来看,目前图神经架构搜索空间包括微观搜索空间、宏观搜索空间、池化搜索空间、超参数以及层数五个维度,并且这些维度可以被任意组合;搜索策略上,主要包括强化学习、进化算法和可微分方法三个主要种类,以及它们的一些结合和改进;在模型性能评估策略上,主要包括每个架构单独评估、权重共享以及超网络三类主要方法. 此外,本文也对图神经架构搜索的近期进展进行评述,并着重讨论了搜索空间压缩、其他搜索空间、图结构学习、搜索高效架构、可迁移性、图神经网络系统等六类方法. 本文同时在实验上对上述算法进行了归纳,包括较小规模节点分类数据集、较大规模节点分类数据集以及图分类数据集三类数据集上的结果. 最后,我们分享对图神经架构搜索未来研究方向的观点,并总结了七个可能的研究方向,包括可扩展性、可解释性、鲁棒性、分布外泛化、神经网络的图模型、硬件感知的模型和综合的评价体系.

表 1 图神经架构搜索算法总结

算法	搜索空间					任务		搜索策略	模型性能评估	其它特点
	微观	宏观	池化	超参数	层数	节点级	图级			
GraphNAS <sup>[32]</sup>	✓	✓	×	×	固定	✓	×	强化学习	单独	—
AGNN <sup>[33]</sup>	✓	×	×	×	固定	✓	×	强化学习	权重共享	—
SNAG <sup>[34]</sup>	✓	✓	×	×	固定	✓	×	强化学习	权重共享	简化搜索空间
PDNAS <sup>[35]</sup>	✓	✓	×	×	固定	✓	×	可微分方法	超网络	—
NAS-GNN <sup>[36]</sup>	✓	×	×	✓	固定	✓	×	进化算法	单独	—
AutoGraph <sup>[37]</sup>	✓	✓	×	✓	可变	✓	×	进化算法	单独	—
GeneticGNN <sup>[38]</sup>	✓	×	×	✓	固定	✓	×	进化算法	单独	—
EGAN <sup>[39]</sup>	✓	✓	×	×	固定	✓	✓	可微分方法	超网络	采样子图提升效率
NAS-GCN <sup>[40]</sup>	✓	✓	✓	×	固定	×	✓	进化算法	单独	考虑边特征
LPGNAS <sup>[41]</sup>	✓	✓	×	×	固定	✓	×	可微分方法	超网络	搜索量化选项
GraphGym <sup>[42]</sup>	✓	✓	×	✓	可变	✓	✓	随机搜索	单独	考虑迁移性
SGAS <sup>[43]</sup>	✓	×	×	×	固定	✓	✓	SGAS	单独	—
Peng et al. <sup>[44]</sup>	✓	×	×	×	固定	×	✓	CEM-RL <sup>[45]</sup>	单独	搜索时空模块
GNAS <sup>[46]</sup>	✓	✓	×	×	可变	✓	✓	可微分方法	超网络	—
AutoSTG <sup>[47]</sup>	×	✓	×	×	固定	✓	×	可微分方法	超网络+元学习	搜索时空模块
DSS <sup>[48]</sup>	✓	✓	×	×	固定	✓	×	可微分方法	超网络	动态调整搜索空间
SANE <sup>[49]</sup>	✓	✓	×	×	固定	✓	×	可微分方法	超网络	—
AutoAttend <sup>[50]</sup>	✓	✓	×	×	固定	✓	×	进化算法	超网络	跨层注意力机制
DiffMG <sup>[51]</sup>	✓	✓	×	×	固定	✓	×	可微分方法	超网络	处理异构图
DeepGNAS <sup>[52]</sup>	✓	✓	×	×	可变	✓	×	强化学习	单独	缓解过平滑问题
LLC <sup>[53]</sup>	×	✓	×	×	可变	✓	×	可微分方法	超网络	—
FL-AGCNS <sup>[54]</sup>	✓	✓	×	×	固定	✓	×	进化算法	超网络	联邦学习场景
G-Cos <sup>[55]</sup>	✓	×	×	×	固定	✓	×	进化算法	超网络	搜索加速器
PAS <sup>[56]</sup>	✓	✓	✓	×	固定	×	✓	可微分方法	超网络	—
FGNAS <sup>[57]</sup>	✓	×	×	×	固定	✓	×	强化学习	单独	软硬件共同设计
GraphPAS <sup>[58]</sup>	✓	×	×	×	固定	✓	×	进化算法	共享种群	并行搜索

(续 表)

算法	搜索空间					任务		搜索策略	模型性能评估	其它特点
	微观	宏观	池化	超参数	层数	节点级	图级			
ALGNN <sup>[59]</sup>	✓	✓	✗	✓	可变	✓	✗	多目标粒子群优化	单独	考虑计算成本
EGNAS <sup>[60]</sup>	✓	✓	✗	✗	固定	✓	✓	可微分方法	超网络	考虑边特征
AutoGEL <sup>[61]</sup>	✓	✓	✓	✗	固定	✓	✓	SNAS <sup>[62]</sup>	超网络	考虑边特征
GASSO <sup>[63]</sup>	✓	✗	✗	✗	固定	✓	✗	可微分方法	超网络	图结构学习
AutoGRL <sup>[64]</sup>	✓	✓	✗	✓	可变	✓	✗	随机搜索	GBDT 预测器	数据增强和特征工程
AGNAS <sup>[65]</sup>	✓	✗	✗	✗	固定	✓	✓	遗传算法	共享种群	并行系统
PaSca <sup>[66]</sup>	✓	✗	✗	✗	可变	✓	✗	贝叶斯优化	单独	可扩展框架与系统
F2GNN <sup>[67]</sup>	✓	✓	✗	✗	固定	✓	✗	可微分方法	超网络	—
HGNAS <sup>[68]</sup>	✓	✓	✗	✗	固定	✓	✗	强化学习	单独	处理异构图
GEMS <sup>[69]</sup>	✗	✗	✗	✗	固定	✓	✗	遗传算法	GCN 预测器	搜索元结构
Wang 等人 <sup>[70]</sup>	✓	✓	✗	✗	可变	✓	✗	随机搜索	单独	简化搜索空间
ARGNP <sup>[71]</sup>	✓	✓	✗	✗	固定	✓	✓	SNAS <sup>[62]</sup>	超网络	建模关系
GraphNAS++ <sup>[72]</sup>	✓	✓	✗	✗	固定	✓	✗	强化学习	单独	分布式系统
GraTO <sup>[73]</sup>	✓	✓	✗	✗	可变	✓	✗	可微分方法	超网络	缓解过平滑问题
NW-GNN <sup>[74]</sup>	✓	✓	✗	✗	可变	✓	✗	强化学习	单独	为每个节点定制架构
DFG-NAS <sup>[75]</sup>	✗	✓	✗	✗	可变	✓	✗	遗传算法	单独	搜索深层图神经网络
GAUSS <sup>[76]</sup>	✓	✗	✗	✗	固定	✓	✗	遗传算法	超网络	采样以处理大规模图
GRACES <sup>[77]</sup>	✓	✗	✗	✗	固定	✗	✓	可微分方法	超网络	考虑图分布外泛化

相关综述: 图神经网络与神经架构搜索分别都有较多研究, 因此有许多相关综述, 例如图神经网络的综述<sup>[3, 12-13]</sup>以及神经架构搜索的综述<sup>[22, 23]</sup>。然而这些综述均未直接涉及图神经架构搜索。仅有文献[30]部分总结了一些早期的图神经架构搜索。相比之下, 本文极大扩展了所覆盖图神经架构搜索的内容和范围, 并增加了更多讨论。例如, 文献[30]仅讨论了 11 种早期图神经架构搜索方法, 而本文则总结了超过 40 种方法。据我们所知, 本文是图神经架构搜索领域目前最详尽的中文综述。

本文在第 2 节, 我们对图神经网络和神经架构搜索进行简单回顾, 并介绍使用的数学符号; 在第 3 节, 我们对经典图神经架构搜索进行总结, 并详细介绍搜索空间、搜索策略和模型评估策略的分类以及每类的具体内容; 在第 4 节, 我们对图神经架构搜索的近期进展进行评述; 在第 5 节, 我们梳理目前方法的实验结果; 最后, 在第 6 节, 我们进行未来展望。

## 2 图神经网络与神经架构搜索简介

考虑一个图  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ , 其中  $\mathcal{V} = \{v_1, v_2, \dots, v_{|\mathcal{V}|}\}$  代表节点的集合,  $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$  代表边的集合。节点  $v_i$  的邻居节点集合记为  $\mathcal{N}(i) = \{v_j : (v_i, v_j) \in \mathcal{E}\}$ 。若节点包含特征或属性信息, 记作  $\mathbf{F} \in \mathbb{R}^{|\mathcal{V}| \times f}$ , 其中  $f$  是特征的维度。

大多数的图机器学习任务可以分为以下两类:

(1) 节点级任务<sup>①</sup>。任务与单个或一对节点相

关, 例如节点分类、链接预测、异常检测等。

(2) 图级任务。任务与整个图相关, 例如图分类、图回归和图生成。

对于节点级任务, 图神经网络模型通常首先学习节点表征  $\mathbf{H} \in \mathbb{R}^{|\mathcal{V}| \times d}$ ,  $d$  是表征的维度, 然后在节点表征上采用一个分类器或预测器来解决该任务。对于图级任务, 则需要学习整个图的表征  $\mathbf{h}_g \in \mathbb{R}^d$ , 并将该表征输入到分类器或预测器中。

图神经网络在过去几年得到了广泛关注。目前, 图神经网络大都符合一个消息传递(message-passing)框架, 可以表述如下<sup>[7]</sup>。

$$\mathbf{m}_i^{(l)} = \text{AGG}^{(l)}(\{a_{ij}^{(l)} \mathbf{W}^{(l)} \mathbf{h}_j^{(l)}, \forall j \in \mathcal{N}(i)\}) \quad (1)$$

$$\mathbf{h}_i^{(l+1)} = \sigma(\text{COMBINE}^{(l)}[\mathbf{m}_i^{(l)}, \mathbf{h}_i^{(l)}]) \quad (2)$$

其中  $\mathbf{h}_i^{(l)}$  代表节点  $v_i$  在图神经网络第  $l$  层的节点表征,  $\mathbf{m}_i^{(l)}$  是节点  $v_i$  的消息向量,  $\text{AGG}^{(l)}(\cdot)$  是聚合函数,  $a_{ij}^{(l)}$  代表从  $v_j$  到  $v_i$  的权重,  $\text{COMBINE}^{(l)}(\cdot)$  是合成函数,  $\mathbf{W}^{(l)}$  代表可学习权重,  $\sigma(\cdot)$  是非线性激活函数。节点表征一般被初始化为节点特征, 即  $\mathbf{H}(0) = \mathbf{F}$ 。简而言之, 每个节点通过和邻居节点交换信息不断更新自己的表征。每个节点的最终表征由  $L$  层消息传递层学习得到, 即  $\mathbf{H} = \mathbf{H}^{(L)}$ 。对于全图表征, 池化函数(Pooling, 亦称为读出函数 Readout)被用来聚合各个节点的表征:

$$\mathbf{h}_g = \text{POOL}(\mathbf{H}) \quad (3)$$

① 有些文献会区分节点级任务和图级任务。在本文中, 我们将图级任务视为广义的节点级任务, 因为图级任务往往仅涉及两个节点。

即  $\mathbf{h}_G$  是图  $\mathcal{G}$  的表征. 基于上述节点或全图的表征, 再通过一个或若干个任务特定的层作为分类器或解码器, 图神经网络即可进行端到端的学习.

神经架构搜索旨在让算法可以自动找到当前任务和数据集最优的神经网络架构. 其可一般被形式化为如下的双层优化(bi-level optimization)问题<sup>[28]</sup>:

$$\begin{aligned} & \min_{\alpha \in \mathcal{A}} \mathcal{L}_{\text{val}}(\mathbf{W}^*(\alpha), \alpha) \\ \text{s. t. } & \mathbf{W}^*(\alpha) = \underset{\mathbf{W}}{\operatorname{argmin}}(\mathcal{L}_{\text{train}}(\mathbf{W}, \alpha)) \end{aligned} \quad (4)$$

其中,  $\alpha$  是希望得到的最优神经网络架构,  $\mathcal{A}$  是架构的搜索空间,  $\mathbf{W}(\alpha)$  是神经网络里面的可学习参数. 简而言之, 神经架构搜索的目标是在搜索空间中找到一个或若干个神经网络架构, 使得该架构在验证集上取得最优的结果, 而  $\mathbf{W}^*$  则表示神经网络的可学习参数在训练集上被完全优化. 不同的神经架构搜索方法在搜索空间和搜索策略设计和具体实例化上存在重要区别, 并导致最终搜索得到的架构产生不同. 此外, 一般的双层优化问题是 NP 难的<sup>[31]</sup>, 因此直接通过枚举和尝试每个可能的架构来优化式(4)在实际中是不可行的, 需要更有效的评估策略来更高效地寻找到最合适的架构.

### 3 经典图神经架构搜索总结

如图 1 所示, 与一般神经网络架构搜索方法类似<sup>[22]</sup>, 图神经架构搜索方法可以通过三个方面进行比较: 搜索空间(search space)、搜索策略(search strategy)和模型性能评估策略(performance estimation strategy). 所以接下来, 我们从这三个方面回顾并比较不同图神经架构搜索方法. 在下文讨论中, 我们将主要关注于为符合式(1)的消息传递图神经网络设计的图神经架构搜索. 我们在表 1 中总结了不同图神经架构搜索算法的特点.

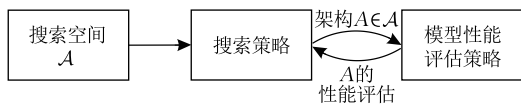


图 1 神经架构搜索方法主要构成元素示意图(包括搜索空间、搜索策略以及模型性能评估策略三部分. 图片改编自文献<sup>[22]</sup>)

#### 3.1 搜索空间

搜索空间决定了神经架构搜索算法在什么范围内进行搜索, 即式(4)中的  $\mathcal{A}$ . 其决定了搜索得到神经网络的可行域和神经网络架构的表达能力. 若搜索空间过小或设计不佳, 即不存在任何有效的架构, 则后续的搜索策略不可能得到有效结果; 反之, 若搜

索空间过大, 即使理论上空间内存在理想的神经网络架构, 也会导致搜索算法难以优化而无法真正找到这些有效架构. 因此, 一个合适的搜索空间需要巧妙设计, 同时避免过小或过大. 因为图神经网络的组成模块与其他深度学习模型, 如卷积神经网络(Convolutional Neural Network, CNN)、循环神经网络(Recurrent Neural Network, RNN)等不同, 图神经架构搜索的第一个挑战与特点即搜索空间的设计. 对于图神经网络, 常见的搜索空间可以分为以下五个类别. 由于这五类空间相互独立, 因此一个算法也可以很容易按照场景与任务需求自由组合不同类别的空间, 即在每个类别空间的笛卡尔积中进行搜索.

##### 3.1.1 微观搜索空间

微观搜索空间定义了节点如何与每层中的其它节点交换消息, 是最细颗粒度的空间设计. 按照式(1)中的消息传递框架, 其中所有的函数, 如  $\text{AGG}(\cdot)$ 、 $\text{COMBINE}(\cdot)$  等, 均可放入搜索空间以进行定制化设计. 常见的图神经网络微观搜索空间包括如下的选项<sup>[32-33]</sup>:

(1) 聚合函数  $\text{AGG}(\cdot)$ . 求和(sum)、求均值(mean)、求最大值(max)以及多层感知机(Multi-Layer Perceptron, MLP).

(2) 聚合权重  $a_{ij}$ . 表 2 中列出了常见的聚合权重搜索空间.

表 2 聚合权重  $a_{ij}$  的代表性搜索空间

类型	公式
CONST	$a_{ij}^{\text{const}} = 1$
GCN	$a_{ij}^{\text{gcn}} = \frac{1}{\sqrt{ \mathcal{N}(i) } \sqrt{ \mathcal{N}(j) }}$
GAT	$a_{ij}^{\text{gat}} = \text{LeakyReLU}(\text{ATT}(\mathbf{W}_a[\mathbf{h}_i, \mathbf{h}_j]))$
SYM-GAT	$a_{ij}^{\text{sym}} = a_{ij}^{\text{gat}} + a_{ji}^{\text{gat}}$
COS	$a_{ij}^{\text{cos}} = \cos(\mathbf{W}_a \mathbf{h}_i, \mathbf{W}_a \mathbf{h}_j)$
LINEAR	$a_{ij}^{\text{lin}} = \tanh(\text{sum}(\mathbf{W}_a \mathbf{h}_i + \mathbf{W}_a \mathbf{h}_j))$
GENE-LINEAR	$a_{ij}^{\text{gene}} = \tanh(\text{sum}(\mathbf{W}_a \mathbf{h}_i + \mathbf{W}_a \mathbf{h}_j)) \mathbf{W}'_a$

注:  $\text{ATT}(\cdot)$  表示用于计算注意力机制(Attention)的函数, 例如一个小的全连接神经网络.

(3) 注意力头数(head). 当使用注意力机制时可选择, 例如 1、2、4、6、8、16 等.

(4) 合成函数  $\text{COMBINE}(\cdot)$ . 拼接(concatenation)、求和以及多层感知机.

(5) 表征向量  $\mathbf{h}'$  的维度. 常用维度包括 8、16、32、64、128、256、512 等.

(6) 非线性激活函数  $\sigma(\cdot)$ . Sigmoid、Tanh、ReLU、Identity、Softplus、LeakyReLU、ReLU6、ELU 等.

上述的微观搜索空间可以包含主流的图神经网络,如图卷积神经网络(GCN)<sup>[6]</sup>、图注意力神经网络(GAT)<sup>[9]</sup>、图采样聚合神经网络(GraphSAGE)<sup>[10]</sup>、图同构神经网络(GIN)<sup>[11]</sup>等作为特例,因此其表达能力有一定保证。然而,直接搜索所有这些组合会使单个消息传递层中产生成千上万个可能的选项,例如上述列出的空间即可枚举出: $4 \times 7 \times 6 \times 3 \times 7 \times 8 = 28224$ 个可能的不同架构。若考虑多个不同的消息传递层,则搜索空间大小会随消息传递层的数量指数增长。并且注意到,这仅仅是微观搜索空间,全部的搜索空间一般是不同搜索空间的笛卡尔积,因此会更加巨大。

### 3.1.2 宏观搜索空间

与微观搜索空间关注于细颗粒度的消息传递机制设计相比,宏观搜索空间更关注于整体图神经网络的粗颗粒度架构。类似卷积神经网络中的残差连接(residual connection)和稠密连接(dense connection),图神经网络中每一层的节点表征不一定仅依赖于其前一层表征<sup>[78-79]</sup>,即式(1)中的 $\mathbf{H}^{(l+1)}$ 不仅仅只基于 $\mathbf{H}^{(l)}$ 进行消息传递而计算。这些不同层之间的连接关系构成宏观搜索空间。具体来说,宏观搜索空间可以被形式化如下:

$$\mathbf{H}^{(l)} = \sum_{j < l} \mathcal{F}_{j,l}(\mathbf{H}^{(j)}) \quad (5)$$

其中, $\mathcal{F}_{j,l}(\cdot)$ 包括式(1)中的消息传递层、ZERO(即不连接)、IDENTITY(即恒等映射,例如为了实现残差连接)或一个全连接层。由于 $\mathbf{H}^{(j)}$ 的维度在不同层可能不尽相同,IDENTITY仅适用于连接维度一致的层之间。若将每层表征视为计算图中的一个节点, $\mathcal{F}(\cdot)$ 视作有向边,则式(5)实际上定义了计算图为一个有向无环图(DAG)<sup>①</sup>。宏观搜索空间产生的一个概念示意图如图2所示,其中不同颜色的图代表节点在不同消息传递层的表征,箭头则表示学习到的层间链接关系。

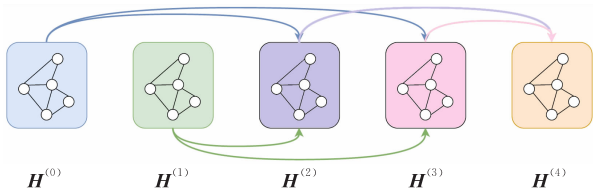


图2 宏观搜索空间可能产生图神经网络层间连接关系的一个概念示意图(不同颜色的图代表节点在不同消息传递层的表征,两个表征之间存在箭头则表示学习到层间具有连接关系(具体采用哪种函数可以通过优化微观搜索空间决定),两个表征之前不存在箭头则表示两层之间不具有连接(即ZERO)。图片改编自文献[48])

通过宏观搜索空间,图神经架构搜索可支持更多更复杂架构的图神经网络。不难发现,若当 $j \neq l-1$ 时,均有 $\mathcal{F}_{j,l} = \text{ZERO}$ ,即每层仅与之前一层相连,则上述宏观搜索空间退化为一个串行的、最朴素的图神经网络;此外,上述宏观搜索空间可以包含带残差连接的图神经网络<sup>[6]</sup>(即当 $j = l-2$ 时, $\mathcal{F}_{j,l} = \text{IDENTITY}$ )、跳跃知识网络(JKNet)<sup>[79]</sup>(即当 $l = L$ 时, $\mathcal{F}_{j,l} \neq \text{ZERO}$ )、深度图卷积神经网络<sup>[78]</sup>等作为特例,因此其表达能力有一定保证。

### 3.1.3 池化操作空间

利用上述微观和宏观搜索空间,可以得到有效的处理节点级任务的图神经网络架构。为了处理图级任务,所有节点的信息需要被聚合以得到全图表征,即式(3)中的池化操作。一个最简单的方案是采用一个预设的池化操作,而仅仅搜索学习节点表征的图神经网络架构。但为了使池化操作更有效且更灵活,图神经架构搜索中也可以定义池化操作空间,以搜索最有效的池化操作。文献[40]提出一个池化操作的搜索空间,包括以下池化操作:

(1) 按节点的求和、平均值或最大值,即

$$\mathbf{h}_g = \text{POOL}(\mathbf{H}) = \mathcal{F}_{\text{pool}}(\{\mathbf{H}_{v,\cdot}, \forall v \in \mathcal{V}\}),$$

其中 $\mathcal{F}_{\text{pool}}(\cdot)$ 是求和、求平均值或求最大值。因此, $\mathbf{h}_g \in \mathbb{R}^d$ 。

(2) 按维度的求和、平均值或最大值,即

$$\mathbf{h}_g = \text{POOL}(\mathbf{H}) = \mathcal{F}_{\text{pool}}(\{\mathbf{H}_{\cdot,i}, \forall 1 \leq i \leq d\}),$$

其中 $\mathcal{F}_{\text{pool}}(\cdot)$ 是求和、求平均值或求最大值。因此, $\mathbf{h}_g \in \mathbb{R}^{|\mathcal{V}|}$ 。由于图表征的维度取决于图中的节点数,不同大小的图表征维度可能不同。

(3) 注意力池化,

$$\mathbf{h}_g = \sum_{v=1}^{|\mathcal{V}|} \sigma(\mathbf{H}_{v,\cdot} \mathbf{W}_1 + \mathbf{b}_1) \odot (\mathbf{H}_{v,\cdot} \mathbf{W}_2 + \mathbf{b}_2),$$

其中 $\mathbf{W}_1, \mathbf{W}_2, \mathbf{b}_1, \mathbf{b}_2$ 是可学习权重。输出维度可通过可学习权重的维度进行调整。

(4) 注意力求和,即在求和时使用注意力权重:

$$\mathbf{h}_g = \sum_{v=1}^{|\mathcal{V}|} \mathbf{b}_v \mathbf{H}_{v,\cdot}, \quad \mathbf{b} = \text{softmax}(\mathbf{H}\mathbf{W}),$$

其中 $\mathbf{W}$ 是可学习参数。因此, $\mathbf{h}_g \in \mathbb{R}^d$ 。

(5) 扁平化操作,即直接将 $\mathbf{H}$ 展开为向量,因此, $\mathbf{h}_g \in \mathbb{R}^{d|\mathcal{V}|}$ 。

通过适当的设计,更多的池化方法,如层次池化<sup>[80]</sup>也可以被添加到搜索空间中。PAS<sup>[56]</sup>提出了一

① 本文中的计算图是将数据(表征)视为节点,神经网络操作视为边。另一种计算图建模是将神经网络操作视为节点,数据是否流通过为边。

个统一的 Top- $k$  池化,即

$$\mathbf{s} = \mathcal{F}_{\text{pool}}(\mathbf{A}, \mathbf{H}), \text{id}x = \text{TOP}_k(\mathbf{s}) \quad (6)$$

其中  $\mathbf{s} \in \mathbb{R}^{|\mathcal{V}|}$  是一个得分向量,表示学习到的每个节点重要程度,  $\mathcal{F}_{\text{pool}}(\cdot)$  是可搜索的池化函数,  $\text{id}x$  表示选中的  $k$  个节点,  $\text{TOP}_k(\cdot)$  则是选择  $k$  个最大值下标的函数. 池化的结果表示为

$$\mathbf{A}' = \mathbf{A}(\text{id}x, \text{id}x), \mathbf{H}' = \mathbf{H}(\text{id}x, \cdot) \quad (7)$$

其中  $\mathbf{A}'$ ,  $\mathbf{H}'$  和  $\mathbf{A}$ ,  $\mathbf{H}$  分别表示池化前后的邻接矩阵和节点表征. PAS 可以涵盖 TopK 池化<sup>[81]</sup>、SAGPool<sup>[82]</sup>、ASAP<sup>[83]</sup> 等方法作为特例.

### 3.1.4 超参数空间

除了上述空间定义的神经网络架构之外,训练神经网络还需要许多超参数,它们往往也是人工通过尝试和经验决定的. 超参数优化(hyper-parameter optimization)<sup>[84]</sup> 本身也是自动机器学习的一个重要研究方向,例如使用随机优化<sup>[85]</sup>、贝叶斯优化<sup>[86]</sup>、基于树结构的核方法<sup>[87]</sup>等. 对于自动化神经网络设计,可以在用神经架构搜索学习到最佳架构,再针对该架构使用一个已有的超参数优化算法,即神经架构搜索-超参数优化的串行优化,也可以将两者交替进行<sup>[88]</sup>,即先找到一个神经网络架构,然后寻找其最合适的超参数,再搜索该超参数下最优的架构,基于该架构重新优化超参数,以此类推. 另一种方案是将这些训练所需超参数合并到搜索空间中,并直接用神经架构搜索策略进行优化. 图神经网络中代表性超参数包括:

- (1) 学习率(learning rate):  $1e-1$ ,  $1e-2$ ,  $1e-3$ ,  $1e-4$  等.
- (2) 训练轮数(epoch): 100, 200, 400 等.
- (3) 批大小(batch size): 16, 32, 64 等.
- (4) 批归一化(batch normalization): 是或否.
- (5) 优化器(optimizer): Adam, RMSProp, SGD 等.
- (6) 丢弃率(dropout):  $0 \sim 0.9$  等.
- (7) 权重衰减(weight decay):  $1e-2$ ,  $1e-3$ ,  $5e-4$ ,  $1e-4$ ,  $5e-5$ ,  $1e-5$  等.

这些超参数可以拼接到上述微观、宏观和池化操作空间,利用搜索策略与神经网络架构共同优化.

### 3.1.5 层数

图神经网络的另一个重要超参数是消息传递层的数量,即参数  $L$ . 理论上来说,由于大多数图神经网络在一层消息传递中仅聚合一阶邻居,因此每个节点的感受野(即能够收发消息的范围)取决于消息传递的层数. 层数越多,每个节点可以聚合更多层邻

居的信息. 此外,图神经网络的表达能力也与其层数(即深度)和每层的神经元数量(即宽度)相关<sup>[89-90]</sup>.

然而,与其它深度学习方法(如卷积神经网络或 Transformer)不同,目前大多数图神经网络的层数一般较少,例如不超过 3~4 层. 当层数越多时,图神经网络的效果反而会有所下降. 研究者一般认为该问题的可能原因是过平滑(over-smoothing)与过拟合(over-fitting)<sup>[78,91]</sup>,以及近期发现的一些其它原因,例如过度挤压(over-squashing). 关于图神经网络层数问题的讨论超出了本文范畴,可参见更多相关论文. 但受这些问题限制,现有的图神经架构搜索方法通常预设一个较小的固定层数,以避免层数过多所产生的问题. 也有少数方法支持枚举若干层数或者自适应调整层数. 后者的一般思路是,首先预设一个初始层数,然后在搜索过程中增加或减少层数,并比较改变层数得到的图神经网络架构与原层数得到的图神经网络架构效果,然后选择其中较优的架构,将其层数作为新的层数默认值<sup>[37]</sup>. 如何自动设计深层图神经网络,例如通过结合过平滑的解决方案,这在很大程度上仍是一个未被探索的方向. DeepGNAS<sup>[52]</sup> 是该方向的一个近期尝试. 受 GCNII<sup>[92]</sup> 工作启发,DeepGNAS 通过增加初始残差连接及恒等映射(即式(5)中,令  $j=0$  和  $j=l-1$  时,  $\mathcal{F}_{jl} = \text{IDENTITY}$ )来解决过平滑问题. 另一方面,通过分析图神经架构搜索自动设计出的图神经网络架构,有望对设计深层图神经网络带来新的启发<sup>[32]</sup>,例如如图 3 所示,DeepGNAS 搜索出的深层图神经网络架构包含一些人工难以设计出的架构.

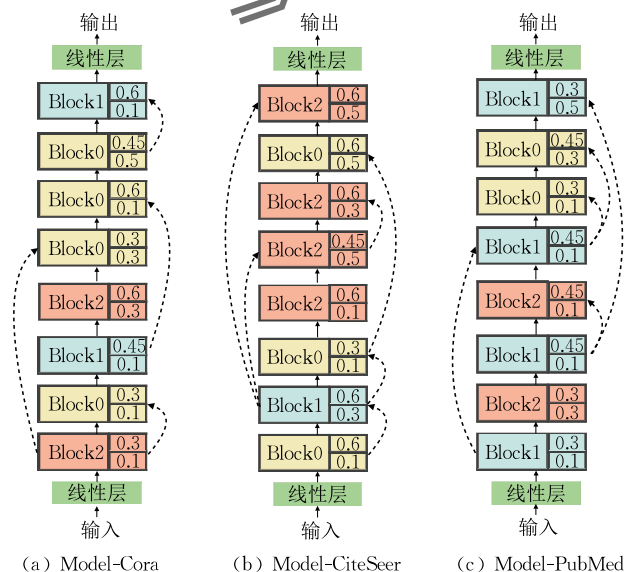


图 3 DeepGNAS 搜索出的针对不同数据集最优的深层图神经网络架构(图片改编自原论文<sup>[52]</sup>)

### 3.2 搜索策略

图神经架构搜索的第二个重要方面是搜索策略,即一个算法如何在搜索空间里面找到最优的神经网络架构.与一般的神经架构搜索方法类似,图神经架构搜索的搜索策略大致可分为三类:使用强化学习(Reinforcement Learning, RL)训练的架构控制器(controller),可微分(differentiable)方法和进化算法(evolution algorithm).

#### 3.2.1 强化学习

神经架构搜索一个被最早提出并广泛采用的搜索策略是使用控制器来生成神经架构的描述,并通过强化学习来训练控制器<sup>[93]</sup>.例如,如果将神经架构描述为一个序列,可以使用循环神经网络作为控制器<sup>[93]</sup>,并生成变长度的神经架构描述,如图 4 所示.强化学习的优化函数则是最大化模型奖励函数(reward)的期望,即

$$\mathcal{J}(\theta_c) = \mathbb{E}_{P(a_{1:T}|\theta_c)} [R] \tag{8}$$

$\theta_c$  是控制器的参数,  $R$  是期望(例如模型在验证集上的性能),  $a_{1:T}$  表示到目前长度生成的架构描述.与合适的搜索空间和模型性能评估策略相结合,这类方法只需调整控制器输出的操作对应于图神经网络的搜索空间,即可以较直接应用于搜索图神经网络架构.

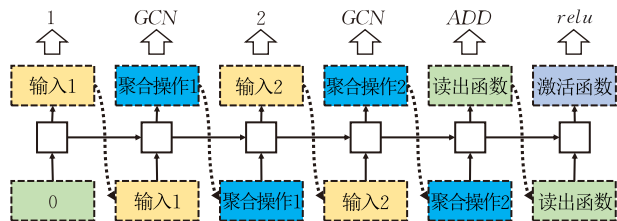


图 4 将图神经网络架构描述为一个序列并用循环神经网络生成架构描述的示意图(图片摘自文献[32])

#### 3.2.2 可微分搜索方法

可微分搜索算法如 DARTS<sup>[28]</sup>和 SNAS<sup>[62]</sup>等近年来受到广泛关注.与强化学习方法中单独处理每个模型不同,可微分方法往往构建一个包含所有可能操作的超网络,也称为单试(one-shot)模型.具体来说,该模型可以写为

$$y = o^{(x,y)}(x) = \sum_{o \in \mathcal{O}} \frac{\exp(\mathbf{z}_o^{(x,y)})}{\sum_{o' \in \mathcal{O}} \exp(\mathbf{z}_{o'}^{(x,y)})} o(x) \tag{9}$$

其中,  $o^{(x,y)}(x)$  是图神经网络中的一个操作,其输入是  $x$ , 输出是  $y$ ,  $\mathcal{O}$  代表所有可能操作组成的集合,  $\mathbf{z}^{(x,y)}$  是可学习的参数,用于控制选择哪个操作.简而言之,每个采用的操作是所有可能操作的一个概率分布叠加,如图 5 所示.通过这种方式,架构中操

作的选择和模型的权重可以通过梯度下降方法共同优化.具体来说,常采用下列迭代的优化方式:

$$\alpha = \alpha - \nabla_{\alpha} \mathcal{L}_{val}(\mathbf{W}(\alpha), \alpha) \tag{10}$$

$$\mathbf{W} = \mathbf{W} - \nabla_{\mathbf{W}} \mathcal{L}_{train}(\mathbf{W}, \alpha) \tag{11}$$

其中  $\alpha$  为所有操作对应可学习概率分布  $\mathbf{z}^{(x,y)}$  的集合,  $\mathbf{W}$  为不同操作中的可学习权重.上述优化过程通过控制其中一个变量(架构或者权重),然后对另一个变量进行优化,并进行反复迭代以得到最优的权重和概率分布.

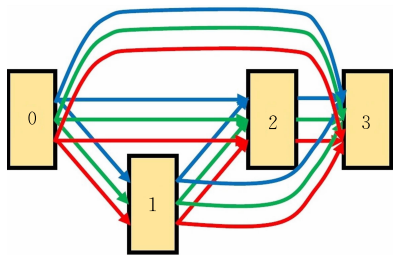


图 5 超网络示意图(每个方块代表一个节点隐层表征,数字代表其层数,不同颜色线段代表不同操作,多条线段代表不同操作的概率叠加.图片摘自文献[63])

对于上述方法,由于超网络中所有参数需同时优化,因此整个超网络均需要被储存在显存中,导致显存开销很大.为解决该问题, GDAS<sup>[94]</sup>提出通过 Gumbel-softmax<sup>[95]</sup>和 Concrete 分布<sup>[96]</sup>来解决该问题.具体来说,其修改式(9)为

$$y = \sum_{o \in \mathcal{O}} \frac{\exp((\mathbf{z}_o^{(x,y)} + g_{x,y})/\tau)}{\sum_{o' \in \mathcal{O}} \exp((\mathbf{z}_{o'}^{(x,y)} + g_{x,y})/\tau)} o(x) \tag{12}$$

其中  $g_{x,y} = -\log(-\log(u))$  是 Gumbel 样本,  $u$  是均匀随机样本,  $\tau$  是温度函数.当  $\tau \rightarrow +\infty$  时,上述 Gumbel-softmax 会近似为严格的最大化函数(即 argmax).在 GDAS 中,在神经网络的前向传播中,仅有概率最大的一个操作会被激活并计算,因此可以降低显存开销;在反向传播中,则会根据式(12)计算所有操作的梯度.当应用于图神经架构搜索时,这些方法可能需要一些修改以考虑图神经网络的搜索空间,但可微分搜索方法的总体思想是一致的.

#### 3.2.3 进化算法

受生物进化过程的启发,进化算法是一类常用的优化算法<sup>[97]</sup>.应用于神经架构搜索时,可以通过随机生成的架构作为初始“种群”(population),每个架构作为一个个体.然后,利用进化过程中的变异(mutation)和交叉(crossover)操作以生成新的架构.这些架构经过评估和选择后形成新的种群,然后重复上述的过程.在更新种群过程中,效果最好的架构会被记录下来.经过反复的迭代更新后,即模拟生



物的进化过程,即可得到最终搜索出的架构.不同进化算法的区别在于如何定义变异、交叉等操作,以及如何筛选出新的种群.例如,正则化演化搜索算法<sup>[27]</sup>的核心思想是一个“老化”机制,即在选择过程中,最老的个体将被移除出种群.

对于图神经架构搜索,进化算法同样需要首先定义架构描述,例如与第 3.2.1 节相同,可以将每个图神经网络架构描述为一个序列.然后则需要基于该描述,定义变异、交叉等操作.例如,如图 6 所示,文献<sup>[37]</sup>中定义变异操作为更换图神经网络微观搜索空间中的一个选项.

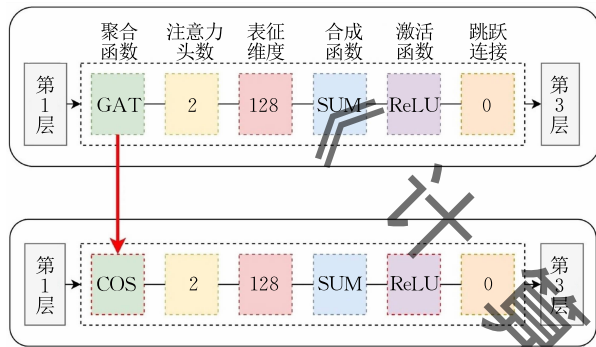


图 6 图神经架构搜索进化算法中的变异操作示意图  
(图片摘自文献<sup>[37]</sup>)

Genetic-GNN<sup>[88]</sup>则提出了另一个进化算法,通过交替地更新图神经网络的架构空间和超参数空间,以找到两者间的最佳组合.

### 3.2.4 讨论与算法组合

上述三类算法实际上均是通用的优化算法,因此其应用于图神经架构搜索时的优缺点尚无共识.本文作者认为,强化学习方法是最早应用于神经架构搜索的经典方法,但由于其往往需要使用策略梯度(policy gradient)进行优化,因此计算效率相对较慢;可微分搜索方法通过直接优化超网络,因此可直接得到架构和对应的权重,计算效率相对较快;进化算法由于可以较容易定义每个架构的“适应度”,因此更加灵活.

此外,上述三种类型的搜索策略也可以被结合使用,以更好解决具体应用场景的需求或者更加复杂的优化问题.例如,AGNN<sup>[33]</sup>提出了一种基于强化学习的保守搜索策略,其在控制器中同时采用循环神经网络和进化算法,并用强化学习训练控制器.通过仅生成略有不同的架构,控制器可以更有效地找到效果较好的图神经网络.文献<sup>[44]</sup>则采用 CEM-RL 算法<sup>[45]</sup>,其结合了进化算法和可微分算法进行架构搜索.具体来说,CEM-RL 将进化算法与

离线强化学习相结合,每次将生成种群中的一半个体用于训练强化学习中的评判员(critic)网络,并用训练后的评判员网络对这一半个体进行策略梯度更新,另一半个体不需要进行此操作.然后,其使用标准的进化算法进化并选择种群.同时,生成的种群样本也会被放入回放缓存(replay buffer)中,并用来训练强化学习的神经网络.除 CEM-RL 外,近期也有许多工作<sup>[50,54-55]</sup>均先通过可微分方法训练超网络,然后基于超网络进行模型性能评估,并使用标准的进化算法搜索出最好的架构.

### 3.3 模型性能评估策略

由于搜索空间中包含海量的可能架构,如何对模型进行性能评估也是图神经架构搜索算法的一项关键因素.接下来,我们将介绍一些模型性能评估策略以高效判断架构的优劣.

最简单也是最直接的模型性能评估策略是对每个架构进行单独的训练,然后在验证集上测试它的效果.这种方法可以准确得到每个架构的性能,是精确度最高的方法.然而,由于每个架构的训练均需耗费可观的计算资源,因此对每个架构进行单独地从零开始训练需要的计算量往往非常高,难以应对大规模网络或较大的搜索空间.一些常用的加快性能评估的技巧是降低精确性<sup>[22]</sup>,例如减少训练的轮数或者训练数据的数量.这些称为低保真度(low fidelity)的模型性能评估策略往往可以被直接使用到图神经架构搜索过程中,但其无法从根本上解决效率与精度间的平衡问题.

另一种神经架构搜索中的常见策略是在不同模型之间共享一些参数,称为参数共享或权重共享<sup>[29]</sup>.由于参数共享可以减少整体所需要训练参数数量,例如若平均每  $k$  个模型共享一组参数,则整体参数量可下降为原来的  $\frac{1}{k}$ ,因此模型的性能评估速度可以被提升.最简单的权重共享通过规则定义哪些架构间可以共享权重.例如,对于卷积神经网络,由于已知许多卷积滤波器是一般的特征提取器,可以提取出视觉信息中的一般特征(例如线段、简单模式等<sup>[98]</sup>),在卷积神经网络中继承之前架构的的权重,尤其是低层卷积滤波器的权重,是可行且有物理含义的<sup>[99]</sup>.相比之下,由于图神经网络中权重代表的物理含义尚不是十分明确,因此参数共享时需要更加谨慎<sup>[34]</sup>.例如,如图 7 所示,AGNN<sup>[33]</sup>提出了参数共享的三个约束,即相同的权重形状(即权重维度要相同),相同的注意力和激活函数,以及批归一化和跳跃连接中不进行权重继承.

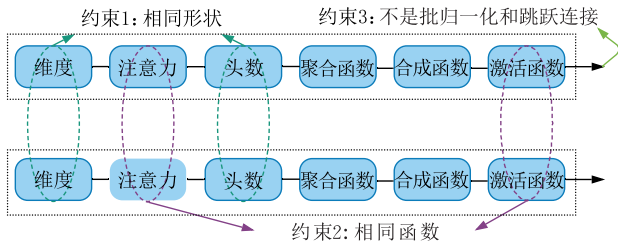


图 7 AGNN 提出的三个参数共享约束示意图  
(图片摘自原文<sup>[33]</sup>)

还有一类方法是通过训练预测器来预测图神经网络架构的性能. 该类方法首先需要构造一个预测器模型作为代理(proxy), 然后训练一些图神经网络架构以得到其对应的效果. 然后, 利用这些图神经网络架构和其真实效果作为数训练据来训练预测器, 并对其他图神经网络模型的性能进行预测. 例如, AutoGRL<sup>[64]</sup> 提出利用梯度提升决策树 (GBDT) 作为预测器模型, GEMS<sup>[69]</sup> 则使用了一个固定架构的图卷积神经网络作为预测器. 由于基于预测器的方法仍需训练一些图神经网络以得到其效果作为预测器的训练数据, 其复杂度依旧相对较高. 此外, 预测器本身的模型设计会极大程度影响其预测的性能.

另一类更常见、更加主流的参数共享机制即通过第 3.2.2 节所介绍的超网络, 即将所有可能的操作通过一个共享的超网络来进行权重共享, 使所有的模型参数可以被联合训练. 但是因为超网络中包含所有可能的操作, 其训练过程可能会变得困难. 为了进一步加快训练过程, 单路径单试模型被提出<sup>[100]</sup>, 其中在每次神经网络的正向和反向传播过程中, 输入和输出对之间只有一个操作被激活. 此外, 若超网络仅用于进行模型性能评估, 而不是需要直接得到最终的图神经架构, 则式(9)中的  $z^{(x,y)}$  参数也可以被省略, 而仅训练不同操作中的参数, 例如文献<sup>[100]</sup>中提出的训练过程中, 对每个操作的选项进行均匀采样, 以充分优化所有操作. 若超网络训练得当, 则其可提供最高效的模型性能评估. 需要注意的是, 在利用超网络作为模型性能评估策略并得到最终架构后, 可以选择重新训练该架构的参数, 也可以选择继承超网络中的架构参数, 作为最终输出.

## 4 图神经架构搜索近期进展盘点

在本节中, 我们梳理并介绍图神经架构搜索的近期进展. 我们将从搜索空间进展、搜索策略进展以

及图神经网络系统三个方面分开介绍.

### 4.1 搜索空间进展

#### 4.1.1 搜索空间压缩

如第 3.1 节所介绍, 虽然大的搜索空间可以容纳更多可能架构以提升图神经网络的表达能力, 其也会导致搜索过程难以优化. 为缓解图神经架构搜索空间过大的问题, 有一些可能的方案对搜索空间进行压缩.

对于微观搜索空间, 借鉴其它神经网络如卷积神经网络的设计, 可以仅搜索一层消息传递层作为单元(cell), 然后堆叠该层形成最终的图神经网络架构, 即不同层的消息传递机制是相同的, 以减少层数带来的组合爆炸问题<sup>[52]</sup>; 此外, 可以根据具体应用和领域知识减小搜索空间, 例如仅关注于一些对最终任务效果最关键的选项, 忽略掉一些对最终结果影响不大的选项<sup>[34,70]</sup>; 最后, 也可将整个微观空间中不同操作的若干可能选项组合作为固定的算子(operator), 然后在整个微观搜索空间中仅预设出少数算子作为算子库, 然后仅考虑如何从算子库中进行选择(例如, 在文献<sup>[32]</sup>中, 作者从 9408 个可能微观搜索空间中选择了 12 个作为算子库).

对于宏观搜索空间, 注意到不同层之间的连接关系是随层数平方增长的, 因此当微观搜索空间与宏观搜索空间叠加, 会导致搜索空间超指数增长. 例如, 假设宏观搜索空间共有  $L$  层, 任意两层间有  $|\mathcal{O}|$  种可能的消息传递函数, 则两者叠加的搜索空间可达到近似  $(|\mathcal{O}|+2)^{\frac{L(L-1)}{2}}$ <sup>①</sup>. 一个常用的方案是限制计算图中边的数量. 若将边的数量限制为  $E_{\max}$ , 则可能的架构数会下降到不高于  $(|\mathcal{O}|+2)^{E_{\max}}$ .

DSS<sup>[48]</sup> 中提出了另一种动态调整搜索空间的策略. 具体来说, 在优化超网络过程中, DSS 每次仅优化每个操作中的  $M$  个选项, 其中包括  $K$  个最优的选项以及从其它所有剩余选项中随机选择  $M-K$  个. 这样, 算法每次需要优化的参数量可以从  $|\mathcal{O}|$  下降到  $M$ .

#### 4.1.2 其他搜索空间

除第 3.1 节介绍的基本搜索空间之外, 不同的图数据和图任务可能需要其它的搜索空间, 下面进行简单介绍.

① 注意到, 这些搜索空间中有些架构是无效的, 例如代表输入数据的  $\mathbf{H}^{(0)}$  和代表输出数据的  $\mathbf{H}^{(L)}$  不连通. 此外, 由于图同构等问题, 计算过程相同的架构可能有不同的对应计算图, 因此存在冗余, 即真正有效的空间会低于此数量, 但仍同一量级.

对于异构图,即图中的节点和边包括多个不同的类型,元路径<sup>[101]</sup>和元结构<sup>[102]</sup>至关重要.因此,GEMS<sup>[69]</sup>和 DiffMG<sup>[51]</sup>均提出考虑元结构的搜索空间,以更好处理异构图. DiffMG 在两个数据集上搜索出的元结构如图 8 所示.

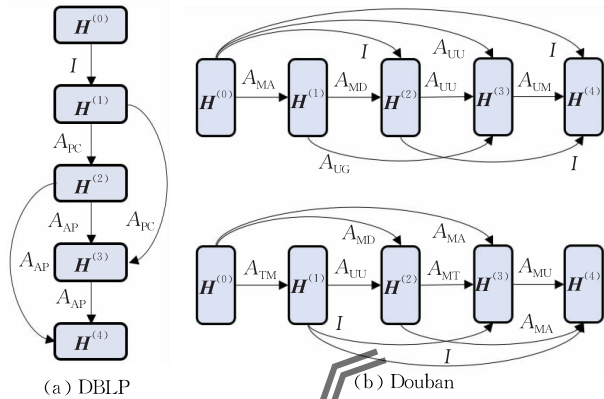


图 8 DiffMG 搜索出的元结构示意图(节点表征间的连线代表不同元路径,并用下标表示.(a)是一个学术网络,其中 A 代表作者(author),P 代表论文(paper),C 代表会议(conference);(b)是一个电影打分网络,其中 U 代表用户(user),G 代表群组(group),M 代表电影(movie),A 代表演员(actor),D 代表导演(director),T 代表电影类型(type). 图片摘自原论文<sup>[51]</sup>)

对于分子图建模,边上的特征是不可或缺的,因此 NAS-GCN<sup>[40]</sup>提出了一考虑边结构的图神经架构搜索,其将式(1)改为

$$\mathbf{m}_i^{(d)} = \text{AGG}^{(d)}(\{a_{ij}^{(d)} \text{MLP}(\mathbf{e}_{ij}) \mathbf{W}^{(d)} \mathbf{h}_i^{(d)}, \forall j \in \mathcal{N}(i)\}) \quad (13)$$

其中  $\mathbf{e}_{ij}$  表示节点  $i$  和节点  $j$  之间边的特征. 其中用于处理边特征的全连接网络  $\text{MLP}(\cdot)$  中的参数(例如层数、隐层数量),也可以被搜索.

对于时空图,图神经架构搜索需要在搜索空间中同时考虑空域和时域信息. 例如, AutoSTG<sup>[47]</sup> 提出针对智慧城市网络的时空图神经网络搜索空间,其包含空间图卷积和时间卷积. 文献<sup>[44]</sup>则提出针对计算机视觉任务中基于骨骼的行为识别的图神经网络的搜索空间,其同样包含时间卷积和空间卷积.

对于其他图任务,设计一个合适的搜索空间通常需要精细的设计以及相应的领域知识,但通常也是图神经架构搜索能够有效的一个关键因素.

## 4.2 搜索策略进展

### 4.2.1 图结构学习

目前绝大多数图神经架构搜索研究均假设图结构已经给定,并作为固定的数据作为输入. 然而,在一些图场景和任务中,图结构可能存在缺失、噪音等,甚至是完全未知的. 此时,图神经架构搜索即需要对图结构进行细致处理,称为图结构学习(graph

structure learning). 一个直接的思路是用现有的图结构学习方法<sup>[103]</sup>作为预处理,然后再进行图神经架构搜索. 例如,可以通过一些先验假设,例如邻接矩阵的低秩假设或相邻节点特征的平滑假设<sup>[103]</sup>,以对输入图结构进行预处理. 对于图结构完全未知的情况,则可以通过人工定义的规则或最近邻算法构建图结构. 然而,由于图结构学习和图神经架构搜索过程是分离的,因此该类方法无法保证有效性.

GASSO<sup>[63]</sup>提出了一种可微分的图架构和图结构联合学习方法,以同时得到最优的图神经网络架构和去噪后的图结构. 具体来说,GASSO 将式(4)的双层优化改为

$$\begin{aligned} \min_{\alpha \in \mathcal{A}} \mathcal{L}_{\text{val}}(\mathbf{W}^*(\alpha), \alpha, \mathcal{G}^*) \\ \text{s. t. } \mathbf{W}^*(\alpha) = \arg \min_{\mathbf{W}} (\mathcal{L}_{\text{train}}(\mathbf{W}, \alpha)) \\ \mathcal{G}^* = \arg \min_{\mathcal{G}'} \mathcal{L}_s(\mathbf{W}^*(\alpha), \alpha, \mathcal{G}') \end{aligned} \quad (14)$$

其中  $\mathcal{G}^*$  表示优化后的图结构,  $\mathcal{L}_s$  表示用于学习图结构的目标函数. 具体来说,GASSO 提出如下的平滑目标函数:

$$\mathcal{L}_s = \lambda \sum_{i,j} \mathbf{A}'_{i,j} \|\mathbf{F}_{i,:} - \mathbf{F}_{j,:}\|_2 + \sum_{i,j} (\mathbf{A}'_{i,j} - \mathbf{A}_{i,j})^2 \quad (15)$$

其中  $\mathbf{A}$  和  $\mathbf{A}'$  代表原图  $\mathcal{G}$  和优化后的图  $\mathcal{G}'$  对应的邻接矩阵,  $\lambda$  是一个超参数. 该目标函数的基本思想是基于图结构应该满足平滑,即同质性(Homophily)假设. 此外,式中的后一项约束则防止优化后的图结构与原图结构差别过大. 通过优化式(14),GASSO 即可实现端到端的图结构与图神经网络架构联合优化.

### 4.2.2 搜索高效架构

除了模型的效果或准确率,图神经网络的运行效率也是值得关注的. 特别是将图神经网络应用到一些资源受限的场景中,例如终端设备等. 如何用图神经架构搜索设计出高效的图神经网络架构在近期也受到了一定关注.

LPGNAS<sup>[41]</sup>提出同时搜索图神经网络架构和神经网络量化(quantization)操作,以便减少模型和缓冲区大小,并保持与原模型相近的精度. 量化操作是指将神经网络的操作进行低比特压缩,例如将 32 位或 16 位的运算压缩为 8 位或 4 位运算,以便在同一批次执行更多的计算. LPGNAS 提出针对图神经网络可学习参数和激活函数的不同量化操作,并从实验上发现,参数使用 4 比特、激活函数使用 8 比特的量化策略是比较适合图神经网络的.

ALGNN<sup>[59]</sup>提出同时在目标函数中考虑搜索图神经网络的架构和复杂度,即将式(4)中的外层优化改为

$$\min_{\alpha \in \mathcal{A}} (\mathcal{L}_{val}(\mathbf{W}^*(\alpha), \alpha), \mathcal{L}_{cost}(\alpha)) \quad (16)$$

$\mathcal{L}_{cost}(\alpha)$ 是模型  $\alpha$  的复杂程度. 具体来说, ALGNN 中提出如下的模型复杂度度量:

$$\mathcal{L}_{cost}(\alpha) = \lambda (KATT(|\mathcal{E}|) + \sum_{v \in \mathcal{V}} AGG(|\mathcal{N}(v)|)) + (1-\lambda) \#Parameters(\alpha) \quad (17)$$

其中  $K$  是注意力的头数,  $ATT(\cdot)$  和  $AGG(\cdot)$  是注意力和聚合函数的计算量,  $\#Parameters$  是模型的参数量,  $\lambda$  是一个超参数. 由于目标函数中有两项, 且其可能存在潜在冲突(即效果好的图神经网络架构可能复杂度也高), 因此 ALGNN 采用一个多目标粒子群优化作为搜索策略.

G-Cos<sup>[55]</sup> 提出同时搜索图神经网络的架构和其在硬件上的合适加速部件, 以最大程度提高模型的准确性和加速效率. 具体来说, G-Cos 提出了一个对于硬件加速的搜索空间, 包括拼贴模式(Tiling Mode)、拼贴大小、内核模式(Kernel Mode)、缓冲区重置比例、权重缓冲区是否共享等. 与式(16)类似, G-Cos 为了在目标函数中同时优化模型的效果和效率, 采用了一个基于遗传算法的搜索策略. 实验结果表明 G-Cos 可以在 FPGA 硬件上搜索出比传统方法更加高效且准确率相当的图神经网络架构, 如图 9 所示.

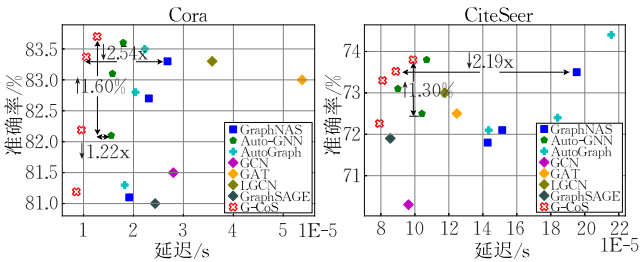


图 9 G-Cos 实验结果图(证明其可实现模型效率和效果的最佳平衡. 图片摘自原论文<sup>[55]</sup>)

### 4.2.3 可迁移性

由于图任务的复杂性和多样性, 在不同的数据集和任务之间迁移图神经网络的架构并非易事. GraphGym<sup>[104]</sup> 中提出使用一组固定的图神经网络作为“锚点”, 利用它们效果的排序作为相关性指标, 以衡量不同任务和数据集的相似度. 然后, 最相似任务中性能最佳的图神经网络架构可以被用来解决目标任务. 他们实验上证明, 目前的图神经网络数据和任务大致可以分为两类, 每一类内部的可迁移性较高, 而两类间的迁移性较低, 如图 10 所示. 如何更系统研究图神经架构搜索在更多不同图数据与图任务的迁移性, 是一个有待深入研究的问题, 并且涉及到

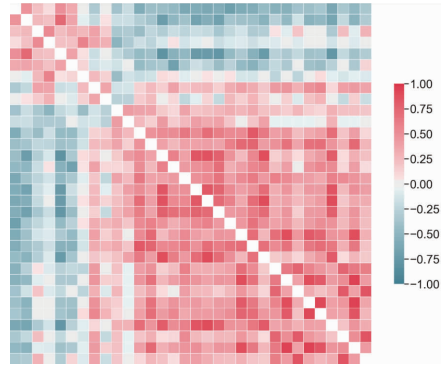


图 10 GraphGym 中关于图神经网络数据和任务的可迁移性实验结果(每行/列代表一个数据集上的一个任务, 颜色代表两个任务间的相关性. 图片摘自原论文<sup>[104]</sup>)

图神经架构搜索的可解释性、鲁棒性与分布外泛化性等(更多讨论见第 6 节).

### 4.3 图神经架构搜索系统

随着图神经架构搜索研究不断深入, 关于图神经架构搜索系统层面的研究也逐渐增多, 并尤其关注在效率方面, 例如通过并行化、分布式等多种方式, 使图神经架构搜索可以处理真实图数据场景.

GraphPAS<sup>[58]</sup> 提出了一种分布式的图神经架构搜索算法, 其采用进化算法作为搜索策略, 并在种群初始化以及进化过程中, 将不同架构的模型性能评估进行分布式计算, 并通过在不同计算节点上共享种群, 使每个计算节点均可利用其他计算节点的信息. 基于该算法, 该论文作者又提出了 AGNAS<sup>[65]</sup> 系统, 以实现高效可并行的图神经架构搜索.

GraphNAS++<sup>[72]</sup> 提出将 GraphNAS 算法扩展为分布式版本, 其中的循环神经网络控制器每次生成一小批次(mini-batch)图神经网络架构样本, 然后在分布式计算环境中对这些样本进行训练, 并用它们的奖励函数来更新控制器的参数, 如图 11 所示. 作者证明, 他们提出的系统满足异步置换不变性, 并且有很好的分布式加速比.

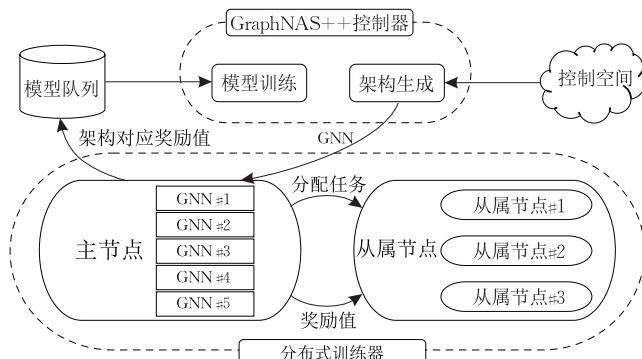


图 11 GraphNAS++ 分布式架构的示意图(图片摘自原论文<sup>[72]</sup>)

PaSca<sup>[66]</sup>提出了一个专为搜索高效图神经网络的图神经架构搜索系统. 他们提出了一个为搜索高效图神经网络架构设计的框架, 包括前处理层、模型训练层和后处理层, 并且将邻居消息传递与模型训练过程解耦, 以此保证整体系统的高效性.

AutoGL<sup>[105]</sup>是首个针对图自动机器学习的开

源库与工具包, 它包含了一个完整的图自动机器学习解决方案 AutoGL Solver, 如图 12 所示, 并将图神经架构搜索作为其中最重要的模块之一. AutoGL 采用了一个模块化的设计理念, 可以便捷的定制化搜索空间、搜索策略、模型性能评估策略等, 以及它们的各种组合, 因此可以方便实现并比较不同的图神经架构搜索算法.

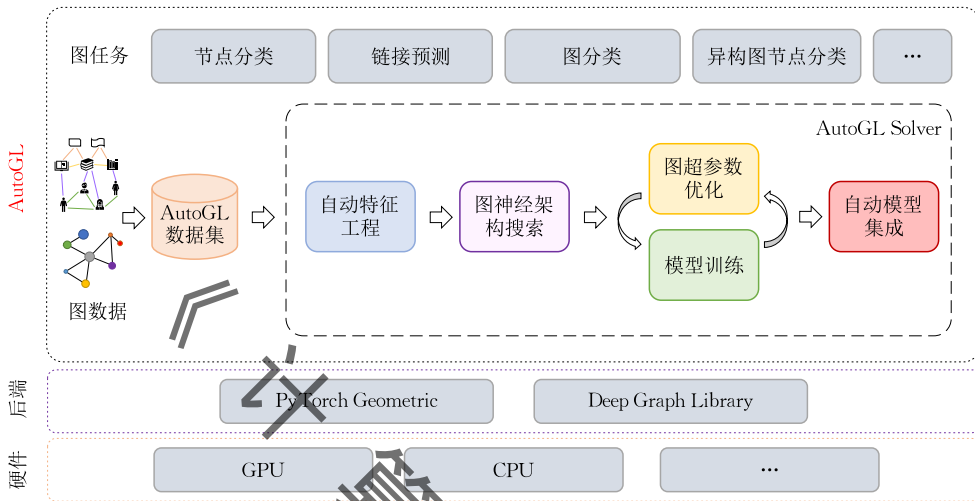


图 12 AutoGL v0.3 版本整体框架示意图

## 5 实验验证

在本节中, 我们对之前图神经架构搜索算法的实验验证进行总结. 具体来说, 我们总结了论文中常用的 17 个数据集, 其统计信息如表 3 所示. 这些数据集可按任务分为两类: 用于节点分类的数据集和用于图分类的数据集. 其中, 对于节点分类任务, 其又可分为 4 个规模相对较小的数据集和 6 个规模相对较

大的数据集. 不同论文的实验结果, 按照上述分类, 分别总结于表 4(较小规模节点分类)、表 5(较大规模节点分类)和表 6(图分类)中. 需要注意, 由于不同方法的假设、实验条件和应用场景不同, 摘取这些实验结果仅旨在对不同方法的实验进行一定概括性总结, 并对验证方法有效性提供一些参考和建议, 而非仅仅通过比较这些实验来证明不同方法的优劣. 此外, 由于许多方法代码未开源或实验条件不同, 我们仅摘取论文中汇报的结果. 从结果中, 我们观察到如下现象.

表 3 数据集统计信息

名称	图任务	图数量	节点数	边数	节点特征数	类别数
Cora <sup>[106]</sup>	节点分类	1.000	2708.0	5429.0	1433	7
CiteSeer <sup>[106]</sup>	节点分类	1.000	3327.0	4732.0	3703	6
PubMed <sup>[106]</sup>	节点分类	1.000	19717.0	44338.0	500	3
PPI <sup>[10]</sup>	节点分类	24.000	2372.7	34113.2	50	121
CS <sup>[107]</sup>	节点分类	1.000	18333.0	81894.0	6805	15
Physics <sup>[107]</sup>	节点分类	1.000	34493.0	247962.0	8415	5
Computer <sup>[107]</sup>	节点分类	1.000	13381.0	245778.0	767	10
Photo <sup>[107]</sup>	节点分类	1.000	7487.0	119043.0	745	8
ogbn-arXiv <sup>[108]</sup>	节点分类	1.000	169343.0	1166243.0	128	40
Reddit <sup>[10]</sup>	节点分类	1.000	232965.0	11606919.0	602	41
D&D <sup>[109]</sup>	图分类	1.178	284.3	715.7	—	2
Proteins <sup>[109]</sup>	图分类	1113.000	39.1	72.8	3	2
IMDB-B <sup>[109]</sup>	图分类	1000.000	19.8	96.5	—	2
IMDB-M <sup>[109]</sup>	图分类	1500.000	13.0	65.9	—	3
MUTAG <sup>[109]</sup>	图分类	188.000	17.9	19.8	—	2
COX2 <sup>[109]</sup>	图分类	467.000	41.2	43.4	3	2
NCI109 <sup>[109]</sup>	图分类	4127.000	29.7	32.1	—	2

注: 当图数量大于 1 时, 节点数和边数为所有图的平均值.

表 4 较小规模节点分类任务数据集图神经架构搜索算法效果(评价指标为分类准确率)

数据集	Cora		CiteSeer		PubMed		PPI
	半监督	全监督*	半监督	全监督*	半监督	全监督*	
GraphNAS <sup>[32]</sup>	83.7±0.4	88.5±1.0	73.5±0.3	76.5±1.3	80.5±0.3	90.3±0.8	—
AGNN <sup>[33]</sup>	83.6±0.3	—	73.8±0.7	—	79.7±0.4	—	99.2±0.1
SNAG <sup>[34]</sup>	—	88.95±0.51	—	77.07±0.64	—	89.42±0.10	98.87±0.10
PDNAS <sup>[35]</sup>	84.5±0.6	89.77±0.31	73.5±0.3	76.29±0.25	79.7±0.6	89.25±0.08	99.27±0.03
AutoGraph <sup>[37]</sup>	83.5±0.4	—	74.4±0.4	—	80.3±0.3	—	98.7±0.3
GeneticGNN <sup>[38]</sup>	83.8±0.5	—	73.5±0.8	—	79.2±0.6	—	98.6±0.4
EGAN <sup>[39]</sup>	—	90.0±0.1	—	78.6±1.1	—	90.0±0.1	99.1±0.0
LPGNAS <sup>[41]</sup>	—	89.8±0.0	—	76.2±0.1	—	89.6±0.1	—
DSS <sup>[48]</sup>	83.9±0.3	91.0±0.2	73.3±0.3	81.4±0.4	80.3±0.2	88.2±0.4	—
SANE <sup>[49]</sup>	—	89.26±1.23	—	78.59±1.08	—	90.47±0.91	98.56±1.20
AutoAttend <sup>[50]</sup>	83.9	—	73.0	—	80.6	—	99.3
DeepGNAS <sup>[52]</sup>	85.6	91.23	73.6	84.04	81.0	90.74	—
LLC <sup>[53]</sup>	—	87.77±0.06	—	—	—	89.78±0.23	—
GraphPAS <sup>[58]</sup>	83.9±0.4	—	73.9±0.3	—	80.6±0.3	—	—
ALGNN <sup>[59]</sup>	84.3±0.5	—	74.6±0.4	—	81.2±0.3	—	—
AutoGEL <sup>[61]</sup>	—	89.89	—	77.66	—	89.68	—
GASSO <sup>[63]</sup>	—	87.63±0.29	—	79.61±0.32	—	90.52±0.24	—
AGNAS <sup>[65]</sup>	83.9±0.4	—	73.9±0.3	—	80.6±0.3	—	—
PaSca <sup>[66]</sup>	84.6±0.6	—	73.4±0.5	—	80.8±0.6	—	—
GraphNAS++ <sup>[72]</sup>	83.4±0.5	88.5±1.0	73.6±0.4	76.5±1.0	80.1±0.2	90.3±1.0	—
GraTO <sup>[73]</sup>	81.5	—	65.2	—	78.9	—	—
NW-GNN <sup>[74]</sup>	—	88.03±0.78	—	79.36±0.89	—	91.22±0.38	—
DFG-NAS <sup>[75]</sup>	85.2±0.2	—	74.1±0.4	—	81.1±0.3	—	—

注: \* : 该数据集没有统一的划分,因此不同算法间结果可能不可比较; — : 原论文中未汇报该数据集上的结果。

表 5 较大规模节点分类任务数据集图神经架构搜索算法效果(评价指标为分类准确率)

数据集	CS*	Physics*	Computer*	Photo*	ogbn-arXiv	Reddit
GraphNAS <sup>[32]</sup>	97.1±0.2	98.5±0.2	92.0±0.4	96.5±0.4	—	—
EGAN <sup>[39]</sup>	96.0±0.3	—	92.0±0.3	—	71.5±0.1	95.4±0.0
LPGNAS <sup>[41]</sup>	—	—	90.5±0.0	89.7±0.0	—	—
GraphGym <sup>[42]</sup>	95.2	96.8	91.6	96.1	—	—
DeepGNAS <sup>[52]</sup>	98.6±0.3	99.5±0.3	92.3±0.5	96.6±0.2	—	—
LLC <sup>[53]</sup>	—	96.78±0.02	91.92±0.08	—	—	—
AutoHEnsGNN <sup>[110]</sup>	—	—	—	—	74.3±0.0	—
PaSca <sup>[66]</sup>	92.8±0.5	93.8±0.9	84.8±0.7	92.7±0.8	—	96.7±0.1
F2GNN <sup>[67]</sup>	—	96.92±0.06	91.42±0.26	—	—	—
GraTO <sup>[73]</sup>	—	—	77.3	86.1	—	94.1
NW-GNN <sup>[74]</sup>	—	—	91.27±0.14	95.12±0.23	—	—
GAUSS <sup>[76]</sup>	96.49±0.11	96.76±0.08	—	—	72.35±0.21	—

注: \* : 该数据集没有统一的划分,因此不同算法间结果可能不可比较; — : 原论文中未汇报该数据集上的结果。

表 6 图分类任务图神经架构搜索算法效果(评价指标为分类准确率)

数据集	D&D	Proteins	IMDB-B	IMDB-M	MUTAG	COX2	NCI109
EGAN <sup>[39]</sup>	77.9±3.4	76.1±3.9	—	—	—	—	—
GraphGym <sup>[42]</sup>	75.1	73.9	—	47.8	—	80.9	—
PAS <sup>[56]</sup>	78.96±3.68	76.64±3.29	75.10±5.32	52.20±3.73	—	83.44±6.33	76.84±2.72
AutoGEL <sup>[61]</sup>	—	82.68	81.20	56.80	94.74	—	—

注: — : 原论文中未汇报该数据集上的结果。

首先,绝大多数针对节点分类任务的图神经架构搜索仅在几个较小数据集进行实验,例如 Cora、CiteSeer、PubMed 等。这一方面表明,由于目前图神经架构搜索处理大图的复杂度还相对较高,并对计算资源有较大要求,因此研究者更倾向于使用小数据

集进行评价。然而,从结果中也不难发现,目前这几个小数据集上的结果已经基本饱和,例如在 Cora、CiteSeer、PubMed 的半监督设定下,不同方法的精度差异一般不超过 1%~2%。因此我们认为,在这些小规模数据集上的比较不足以证明不同算法的区别。

对于更大规模的节点分类任务,其中许多数据集不存在统一的数据划分和实验设定,因此不同算法的结果有时不具有可比性.对于 ogbn-arXiv,其属于近期发布的开放图基准测试(Open Graph Benchmark)<sup>[108]</sup>的一部分,因此比较实验结果更权威可信.然而,与榜单上最先进的人工设计的图神经网络效果相比,目前的图神经架构搜索方法效果仍有差距<sup>①</sup>.

除节点分类之外,用于图分类的图神经架构搜索方法相对更少.此外,目前采用的数据集也相对较小,并且其部分数据集本身即可能存在数据划分、设定不公平等问题<sup>[111]</sup>.因此,图分类任务的实验也有进一步提升的空间.

除了常规的节点分类和图分类外,对于其他特殊图数据集和任务,例如异构图、时空图等,也有一些不同的数据集和评测方式.由于其研究相对较少,在此不进行展开介绍.

总结来看,虽然图神经架构搜索取得了一定进展,并在一些情况下比人工设计的图神经网络效果更优,但如何更系统、全面地对图神经架构搜索方法进行评测,是一个需要进一步研究的问题.

## 6 未来方向

虽然图神经架构搜索已经有一定的研究工作并在真实场景中展现了一定的效果,该方向仍存在许多值得探索的挑战和机遇.在本节中,我们将分享值得学术界和工业界进一步研究的未来方向,并归纳为以下几点:

(1)更多图数据类型.由于图数据的多样性,目前图神经架构搜索研究仅能处理部分类型与特性的图数据.例如,目前图神经架构搜索大多关注于同构图,而异构图的研究刚处于起步阶段;静态图的研究相对丰富,而对动态图的关注则较少;目前工作均无法处理更复杂类型的图,例如异配图(heterophilic graph)、超图(hyper-graph)、符号图(signed graph)等.

(2)可扩展性.目前图神经架构搜索可处理的图规模往往较小,例如如第5节所介绍,大部分工作仅考虑不超过几万个节点或边的图.对于大规模图,例如包含几百万甚至上亿节点和边的图,图神经架构搜索的可扩展性是一个尚未解决的重要问题.一个早期的工作 EGAN<sup>[39]</sup>提出在图中采样子图,并在采样的子图上进行神经架构搜索,然后将搜索得到的架构在原图上使用,从而提高搜索效率.该方法在一定程度上提升了图神经架构搜索的效率.然而,

EGAN 忽略了一个重要的问题,即在小图上效果最好的架构不一定能最优地适配原始大图. PaSca<sup>[66]</sup>近期提出了一种可扩展的图神经网络框架和系统,通过将消息传递与模型训练过程相解耦,以提升图神经架构搜索的效果. GAUSS<sup>[76]</sup>提出了一个架构和子图的联合采样方式进行图神经架构搜索,并在一个亿级规模图数据上展现了优异的效果.我们认为,还需要更多、更系统的相关研究来使图神经架构搜索可以高效并有效处理真实大规模的图数据.需要注意图神经架构搜索在图规模上的可扩展性,与其性能评估策略的效率(即在架构数量上的可扩展性)是两个相关但不同的研究角度,因为即使使用了超网络等高效性能评估策略,图神经架构搜索在大规模图上的复杂度仍然很高.

(3)可解释性.现有的图神经架构搜索方法主要基于黑盒优化.因此,目前尚不清楚为什么某些搜索算法比其他算法效果更好,或者什么特点的图数据/图任务需要什么样的图神经网络架构,即图神经架构搜索的可解释性仍然缺乏系统的研究工作.目前关于图神经网络模型的可解释性<sup>[112]</sup>,以及通过超参数重要性去相关的可解释图超参数优化<sup>[113]</sup>已有一些初步研究.然而,对图神经架构搜索的可解释性研究仍然处在未探索阶段.

(4)鲁棒性.由于许多图神经架构搜索的应用场景是风险敏感的,例如金融、医疗等,因此模型的鲁棒性对于其实际落地应用也是一个重要的考虑因素.目前虽然有一些关于图神经网络鲁棒性的研究<sup>[114]</sup>,但如何将这些研究扩展到图神经架构搜索中尚未被探索.一方面来说,需要研究图神经架构搜索过程是否会受到对抗攻击的影响,并提出对应的防御机制;另一方面,可以利用图神经架构搜索自动设计鲁棒的图神经网络架构,以更好抵御对抗攻击.我们近期在该方向进行了一个初步尝试<sup>[115]</sup>,其通过在搜索空间和搜索策略中考虑图的鲁棒性,提升了搜索得到图神经网络架构的鲁棒性.

(5)分布外泛化.虽然图神经架构搜索的最终目标是全自动处理新的图数据集和图任务,但对于目前的图神经架构搜索算法,其仍需要一定的专家经验来设计搜索空间和搜索算法,所以其泛化能力是较为有限的.当训练和测试数据来自不同的分布,即训练和测试数据存在分布迁移(distribution shift)时,该

① 本文写作时,ogbn-arXiv 最先进的不需要额外数据的方法为 GIANT-XRT+DRGAT,其效果为  $76.11 \pm 0.09$ . 比效果最好的图神经架构搜索方法 AutoHEnsGNN 高约 2%.

问题变得尤为严重<sup>[116-117]</sup>. 如何提升图神经架构搜索的泛化性,特别是处理上述数据分布迁移时的分布外泛化(out-of-distribution generalization)能力,甚至可以处理持续变化的图数据和任务,是亟需研究的方向. GRACES<sup>[77]</sup>是我们近期在该方向的首个尝试,其通过自适应定制化针对每个图数据最合适的图神经网络架构,提升了图神经架构搜索在图分类任务上的性能.

(6) 神经网络的图模型. 在本文中,我们主要关注于图神经架构搜索,即在图数据上利用神经架构搜索方法来构建模型. 反过来,即利用图结构来辅助神经架构搜索研究,同样是可行和有效的. 例如,可以将神经网络计算过程建模为一个有向无环图来分析其架构特性<sup>[104,118]</sup>,或者将图神经网络用到神经架构搜索的性能评估过程中<sup>[88,120-122]</sup>. 最终,我们期待图数据和神经架构搜索(或者更广义来说,深度学习研究)能够形成更紧密的联系和闭环,并进一步相互促进.

(7) 硬件感知的模型. 为进一步提高图神经架构搜索的可扩展性和效率,硬件感知模型是一个可能关键步骤,特别是将提出的算法应用在真正的生产环境中. 硬件感知的图机器学习模型<sup>[122]</sup>和硬件感知的 AutoML 模型<sup>[123-125]</sup> 都已经被研究过,但将这些方法结合起来仍处于研究阶段的早期,并带来了许多挑战. G-Cos<sup>[55]</sup> 是其中的一个早期尝试.

(8) 综合的评价体系. 已有大多数图神经架构搜索仅在小的传统基准图数据集上进行测试,例如第 5 节所介绍三个引用数据集,即 Cora、CiteSeer 和 PubMed<sup>[106]</sup>. 然而,这些数据集已经被说明无法充分比较不同的图神经网络模型<sup>[107]</sup>,更不用说比较图神经架构搜索算法. 因此,该领域需要更全面和综合的评价体系,例如在更大规模图数据集<sup>[108,126]</sup> 上进行测试,或类似于 NAS-bench 系列研究<sup>[127]</sup> 的专为图神经架构搜索定制化的评价方案. 我们近期在该方向上进行了一定尝试并提出了 NAS-Bench-Graph<sup>[128]</sup>,图神经架构搜索上的首个基准数据集,有兴趣的读者可以参考.

## 参 考 文 献

- [1] Newman M. Networks. New York, USA: Oxford University Press, 2018
- [2] Veličković P, Fedus W, Hamilton W L, et al. Deep graph infomax//Proceedings of the International Conference on Learning Representations. New Orleans, USA, 2019
- [3] Zhang Ziwei, Cui Peng, Zhu Wenwu. Deep learning on graphs: A survey. IEEE Transactions on Knowledge and Data Engineering, 2022, 34(1): 249-270
- [4] Scarselli F, Gori M, Tsoi A C, et al. The graph neural network model. IEEE Transactions on Neural Networks, 2008, 20(1): 61-80
- [5] Bruna J, Zaremba W, Szlam A, LeCun Y. Spectral networks and locally connected networks on graphs//Proceedings of the International Conference on Learning Representations. Banff, Canada, 2014
- [6] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks//Proceedings of the International Conference on Learning Representations. Toulon, France, 2017
- [7] Gilmer J, Schoenholz S S, Riley P F, et al. Neural message passing for quantum chemistry//Proceedings of the International Conference on Machine Learning. Sydney, Australia, 2017: 1263-1272
- [8] Defferrard M, Bresson X, Vandergheynst P. Convolutional neural networks on graphs with fast localized spectral filtering//Advances in Neural Information Processing Systems. Barcelona, Spain, 2016, 29: 3844-3852
- [9] Veličković P, Cucurull G, Casanova A, et al. Graph attention networks//Proceedings of the International Conference on Learning Representations. Vancouver, Canada, 2018
- [10] Hamilton W L, Ying Zhitao, Leskovec J. Inductive representation learning on large graphs//Advances in Neural Information Processing Systems. Long Beach, USA, 2017, 30: 1025-1035
- [11] Xu K, Hu W, Leskovec J, Jegelka S. How powerful are graph neural networks?//Proceedings of the International Conference on Learning Representations. New Orleans, USA, 2018
- [12] Wu Zonghan, Pan Shuui, Chen Fengwen, et al. A comprehensive survey on graph neural networks. IEEE Transactions on Neural Networks and Learning Systems, 2020, 32(1): 4-24
- [13] Zhou Jie, Cui Ganqu, Hu Shengding, et al. Graph neural networks: A review of methods and applications. AI Open, 2020, 1: 57-81
- [14] Ying Rex, He Ruining, Chen Kaifeng, et al. Graph convolutional neural networks for web-scale recommender systems//Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. London, UK, 2018: 974-983
- [15] Ma Jianxin, Zhou Chang, Cui Peng, et al. Learning disentangled representations for recommendation//Advances in Neural Information Processing Systems. Vancouver, Canada, 2019, 32: 5711-5722
- [16] Wu Lingfei, Chen Yu, Shen Kai, et al. Graph neural networks for natural language processing: A survey. Foundations and Trends® in Machine Learning, 2023, 16(2): 119-328



- [17] Su Chang, Tong Jie, Zhu Yongjun, et al. Network embedding in biomedical data science. *Briefings in Bioinformatics*, 2020, 21(1): 182-197
- [18] Zitnik M, Leskovec J. Predicting multicellular function through multi-layer tissue networks. *Bioinformatics*, 2017, 33(14): i190-i198
- [19] Li Yaguang, Yu R, Shahabi C, Liu Yan. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting // *Proceedings of the International Conference on Learning Representations*, Vancouver, Canada, 2018
- [20] Yu Bing, Yin Haoteng, Zhu Zhanxing. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting // *Proceedings of the 27th International Joint Conference on Artificial Intelligence*. Stockholm, Sweden, 2018: 3634-3640
- [21] Wang Quan, Mao Zhendong, Wang Bin, Guo Li. Knowledge graph embedding: A survey of approaches and applications. *IEEE Transactions on Knowledge and Data Engineering*, 2017, 29(12): 2724-2743
- [22] Elsken T, Metzen J H, Hutter F. Neural architecture search: A survey. *The Journal of Machine Learning Research*, 2019, 20(1): 1997-2017
- [23] He Xin, Zhao Kaiyong, Chu Xiaowen. AutoML: A survey of the state-of-the-art. *Knowledge-Based Systems*, 2021, 212: 106622
- [24] Yao Quanming, Wang Mengshuo, Chen Yuqiang, et al. Taking human out of learning applications: A survey on automated machine learning. *arXiv preprint arXiv:1810.13306*, 2018
- [25] Elshawi R, Maher M, Sakr S. Automated machine learning: State-of-the-art and open challenges. *arXiv preprint arXiv:1906.02287*, 2019
- [26] Zoph B, Vasudevan V, Shlens J, Le Q V. Learning transferable architectures for scalable image recognition // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. London, UK, 2018: 8697-8710
- [27] Real E, Aggarwal A, Huang Yanping, Le Q V. Regularized evolution for image classifier architecture search // *Proceedings of the AAAI Conference on Artificial Intelligence*. Hawaii, USA, 2019, 33: 4780-4789
- [28] Liu Hanxiao, Simonyan K, Yang Yiming. DARTS: Differentiable architecture search // *Proceedings of the International Conference on Learning Representations*. New Orleans, USA, 2018
- [29] Pham H, Guan M, Zoph B, et al. Efficient neural architecture search via parameters sharing // *Proceedings of the International Conference on Machine Learning*. Stockholm, Sweden, 2018: 4095-4104
- [30] Oloulade B M, Gao Jianliang, Chen Jiamin, et al. Graph neural architecture search: A survey. *Tsinghua Science and Technology*, 2021, 27(4): 692-708
- [31] Sinha A, Malo P, Deb K. A review on bilevel optimization: From classical to evolutionary approaches and applications. *IEEE Transactions on Evolutionary Computation*, 2017, 22(2): 276-295
- [32] Gao Yang, Yang Hong, Zhang Peng, et al. Graph neural architecture search // *Proceedings of the 29th International Joint Conferences on Artificial Intelligence*. 2020, 20: 1403-1409
- [33] Zhou Kaixiong, Song Qingquan, Huang Xiao, Hu Xia. AutoGNN: Neural architecture search of graph neural networks. *arXiv preprint arXiv:1909.03184*, 2019
- [34] Zhao Huan, Wei Lanning, Yao Quanming. Simplifying architecture search for graph neural network. *arXiv preprint arXiv:2008.11652*, 2020
- [35] Zhao Yiren, Wang Duo, Gao Xitong, et al. Probabilistic dual network architecture search on graphs. *arXiv preprint arXiv:2003.09676*, 2020
- [36] Nunes M, Pappa G L. Neural architecture search in graph neural networks // *Proceedings of the Brazilian Conference on Intelligent Systems*. Rio Grande, Brazil, 2020: 302-317
- [37] Li Yaoman, King I. AutoGraph: Automated graph neural network // *Proceedings of the International Conference on Neural Information Processing*. Bangkok, Thailand, 2020: 189-201
- [38] Shi Min, Wilson D A, Zhu Xingquan, et al. Genetic-GNN: Evolutionary architecture search for graph neural networks. *Knowledge-Based Systems*, 2022, 247: 108752
- [39] Zhao Huan, Wei Lanning, Yao Quanming, He Zhiqiang. Efficient graph neural architecture search. <https://openreview.net/forum?id=IjIzIOkK2D6>, 2021
- [40] Jiang Shenli, Balaprakash P. Graph neural network architecture search for molecular property prediction // *Proceedings of the IEEE International Conference on Big Data*. Atlanta, USA, 2020
- [41] Zhao Yiren, Wang Duo, Bates D, et al. Learned low precision graph neural networks // *Proceedings of the 1st Workshop on Machine Learning and Systems*. Edinburgh, UK, 2021
- [42] You Jiaxuan, Ying Zhitao, Leskovec J. Design space for graph neural networks. *Advances in Neural Information Processing Systems*, 2020, 33: 17009-17021
- [43] Li Guohao, Qian Guocheng, Delgadillo I C, et al. SGAS: Sequential greedy architecture search // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA, 2020: 1620-1630
- [44] Peng Wei, Hong Xiaopeng, Chen Haoyu, Zhao Guoying. Learning graph convolutional network for skeleton-based human action recognition by neural searching // *Proceedings of the AAAI Conference on Artificial Intelligence*. New York, USA, 2020, 34: 2669-2676
- [45] Pourchot A, Sigaud O. CEM-RL: Combining evolutionary and gradient-based methods for policy search // *Proceedings of the International Conference on Learning Representations*. New Orleans, USA, 2019

- [46] Cai Shaofei, Li Liang, Deng Jincan, et al. Rethinking graph neural architecture search from message-passing//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021; 6657-6666
- [47] Pan Zheyi, Ke Songyu, Yang Xiaodu, et al. AutoSTG: Neural architecture search for predictions of spatio-temporal graph//Proceedings of the Web Conference 2021. Ljubljana, Slovenia, 2021; 1846-1855
- [48] Li Yanxi, Wen Zean, Wang Yunhe, Xu Chang. One-shot graph neural architecture search with dynamic search space//Proceedings of the AAAI Conference on Artificial Intelligence. 2021, 35; 8510-8517
- [49] Zhao Huan, Yao Quanming, Tu Weiwei. Search to aggregate neighborhood for graph neural network//Proceedings of the 2021 IEEE 37th International Conference on Data Engineering. Chania, Greece, 2021; 552-563
- [50] Guan Chaoyu, Wang Xin, Zhu Wenwu. AutoAttend: Automated attention representation search//Proceedings of the International Conference on Machine Learning. 2021; 3864-3874
- [51] Ding Yuhui, Yao Quanming, Zhao Huan, Zhang Tong. DiffMG: Differentiable meta graph search for heterogeneous graph neural networks//Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. Singapore, 2021; 279-288
- [52] Feng Guosheng, Wang Chunnan, Wang Hongzhi. Search for deep graph neural networks. arXiv preprint arXiv: 2109.10047, 2021
- [53] Wei Lanning, Zhao Huan, He Zhiqiang. Learn layer-wise connections in graph neural networks. arXiv preprint arXiv: 2112.13585, 2021
- [54] Wang Chunnan, Chen Bozhou, Li Geng, Wang Hongzhi. FL-AGCNS: Federated learning framework for automatic graph convolutional network search. arXiv preprint arXiv: 2104.04141, 2021
- [55] Zhang Yongan, You Haoran, Fu Yonggan, et al. G-Cos: GNN-accelerator co-search towards both better accuracy and efficiency//Proceedings of the International Conference on Computer Aided Design. Munich, Germany, 2021; 1-9
- [56] Wei Lanning, Zhao Huan, Yao Quanming, He Zhiqiang. Pooling architecture search for graph classification//Proceedings of the 30th ACM International Conference on Information and Knowledge Management. Queensland, Australia, 2021; 2091-2100
- [57] Lu Qing, Jiang Weiwen, Jiang Meng, et al. FGNAS: FPGA-aware graph neural architecture search. OpenReview, <https://openreview.net/forum?id=cq4FHzAz9eA>, 2020
- [58] Chen Jiamin, Gao Jianliang, Chen Yibo, et al. GraphPAS: Parallel architecture search for graph neural networks//Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval. Canada, 2021; 2182-2186
- [59] Cai Rongshen, Tao Qian, Tang Yufei, Shi Min. ALGNN: Auto-designed lightweight graph neural network//Proceedings of the Pacific Rim International Conference on Artificial Intelligence. Hanoi, Vietnam, 2021; 500-512
- [60] Cai Shaofei, Li Liang, Han Xinzhe, et al. Edge-featured graph neural architecture search. arXiv preprint arXiv:2109.01356, 2021
- [61] Wang Zhili, Di Shimin, Chen Lei. AutoGEL: An automated graph neural network with explicit link information. Advances in Neural Information Processing Systems, 2021, 34; 24509-24522
- [62] Xie Sirui, Zheng Hehui, Liu Chunxiao, Lin Liang. SNAS: Stochastic neural architecture search//Proceedings of the International Conference on Learning Representations. New Orleans, USA, 2018
- [63] Qin Yijian, Wang Xin, Zhang Zeyang, Zhu Wenwu. Graph differentiable architecture search with structure learning//Proceedings of the Annual Conference on Neural Information Processing Systems 2021. 2021, 34; 16860-16872
- [64] Sun Junwei, Wang Bai, Wu Bin. Automated graph representation learning for node classification//Proceedings of the International Joint Conference on Neural Networks. Shenzhen, China, 2021; 1-7
- [65] Chen Jiamin, Gao Jianliang, Chen Yibo, et al. Auto-GNAS: A parallel graph neural architecture search framework. IEEE Transactions on Parallel and Distributed Systems, 2022, 33(11); 3117-3128
- [66] Zhang Wentao, Shen Yu, Lin Zheyu, et al. PaSca: A graph neural architecture search system under the scalable paradigm//Proceedings of the Web Conference 2022. Lyon, France, 2022
- [67] Wei Lanning, Zhao Huan, He Zhiqiang. Designing the topology of graph neural networks: A novel feature fusion perspective//Proceedings of the Web Conference 2022. Lyon, France, 2022
- [68] Gao Yang, Zhang Peng, Li Zhao, et al. Heterogeneous graph neural architecture search//Proceedings of the 2021 IEEE International Conference on Data Mining. Auckland, New Zealand, 2021; 1066-1071
- [69] Han Zhenyu, Xu Fengli, Shi Jinghan, et al. Genetic meta-structure search for recommendation on heterogeneous information network//Proceedings of the 29th ACM International Conference on Information & Knowledge Management. Ireland, 2020; 455-464
- [70] Wang Zhenyi, Zhao Huan, Shi Chuan. Profiling the design space for graph neural networks based collaborative filtering//Proceedings of the 15th ACM International Conference on Web Search and Data Mining. Tempe, USA, 2022
- [71] Cai Shaofei, Li Liang, Han Xinzhe, et al. Automatic relation-aware graph network proliferation//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA, 2022; 10863-10873

- [72] Gao Yang, Zhang Peng, Yang Hong, et al. GraphNAS++: Distributed architecture search for graph neural networks. *IEEE Transactions on Knowledge and Data Engineering*, 2022; 1-1
- [73] Feng Xinshun, Wan Herun, Feng Shangbin, et al. GraTO: Graph neural network framework tackling over-smoothing with neural architecture search//*Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. Atlanta, USA, 2022
- [74] Wang Zhen, Wei Zhewei, Li Yaliang, et al. Graph neural networks with node-wise architecture//*Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. Washington, USA, 2022; 1949-1958
- [75] Zhang Wentao, Lin Zheyu, Shen Yu, et al. Deep and flexible graph neural architecture search//*Proceedings of the International Conference on Machine Learning*. Baltimore, USA, 2022; 26362-26374
- [76] Guan Chaoyu, Wang Xin, Chen Hong, et al. Large-scale graph neural architecture search//*Proceedings of the International Conference on Machine Learning*. Baltimore, USA, 2022; 7968-7981
- [77] Qin Yijian, Wang Xin, Zhang Ziwei, et al. Graph neural architecture search under distribution shifts//*Proceedings of the International Conference on Machine Learning*. Baltimore, USA, 2022; 18083-18095
- [78] Li Guohao, Muller M, Thabet A, Ghanem B. DeepGCNs: Can GCNs go as deep as CNNs?//*Proceedings of the IEEE/CVF International Conference on Computer Vision*. Seoul, Korea, 2019; 9267-9276
- [79] Xu Keyulu, Li Chengtao, Tian Yonglong, et al. Representation learning on graphs with jumping knowledge networks//*Proceedings of the International Conference on Machine Learning*. Stockholm, Sweden, 2018; 5453-5462
- [80] Ying Zitao, You Jiaxuan, Morris C, et al. Hierarchical graph representation learning with differentiable pooling//*Advances in Neural Information Processing Systems*. Montréal, Canada, 2018; 4805-4815
- [81] Gao Hongyang, Ji Shuiwang. Graph U-nets//*Proceedings of the International Conference on Machine Learning*. Long Beach, USA, 2019; 2083-2092
- [82] Lee J, Lee I, Kang J. Self-attention graph pooling//*Proceedings of the 36th International Conference on Machine Learning*. Long Beach, USA, 2019; 3734-3743
- [83] Ranjan E, Sanyal S, Talukdar P. ASAP: Adaptive structure aware pooling for learning hierarchical graph representations//*Proceedings of the AAAI Conference on Artificial Intelligence*. New York, USA, 2020; 5470-5477
- [84] Tu Ke, Ma Jianxin, Cui Peng, et al. AutoNE: Hyperparameter optimization for massive network embedding//*Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. Anchorage, USA, 2019; 216-225
- [85] Bergstra J, Bengio Y. Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 2012, 13(2): 281-305
- [86] Snoek J, Larochelle H, Adams R P. Practical Bayesian optimization of machine learning algorithms//*Advances in Neural Information Processing Systems*. Lake Tahoe, USA, 2012, 25: 2951-2959
- [87] Bergstra J, Bardenet R, Bengio Y, Kégl B. Algorithms for hyper-parameter optimization//*Advances in Neural Information Processing Systems*. Granada, Spain, 2011, 24: 2546-2554
- [88] Shi Han, Pi Renjie, Xu Hang, et al. Bridging the gap between sample-based and one-shot neural architecture search with BONAS. *Advances in Neural Information Processing Systems*, 2020, 33: 1808-1819
- [89] Sato R. A survey on the expressive power of graph neural networks. *arXiv preprint arXiv:2003.04078*, 2020
- [90] Andreas L. What graph neural networks cannot learn: Depth vs width//*Proceedings of the International Conference on Learning Representations*. Addis Ababa, Ethiopia, 2020
- [91] Li Qimai, Han Zhichao, Wu Xiao-Ming. Deeper insights into graph convolutional networks for semi-supervised learning//*Proceedings of the 32nd AAAI Conference on Artificial Intelligence*. New Orleans, USA, 2018
- [92] Wei Zhewei, Chen Ming, Ding Bolin, et al. Simple and deep graph convolutional networks//*Proceedings of the 37th International Conference on Machine Learning*. 2020
- [93] Zoph B, Le Q V. Neural architecture search with reinforcement learning//*Proceedings of the International Conference on Learning Representations*. Toulon, France, 2017
- [94] Dong Xuan, Yang Yi. Searching for a robust neural architecture in four GPU hours//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, USA, 2019; 1761-1770
- [95] Jang E, Gu Shixiang, Poole B. Categorical reparameterization with gumbel-softmax//*Proceedings of the International Conference on Learning Representations*. Toulon, France, 2017
- [96] Maddison C, Mnih A, Teh Y. The concrete distribution: A continuous relaxation of discrete random variables//*Proceedings of the International Conference on Learning Representations*. Toulon, France, 2017
- [97] Eiben A E, Smith J E. *Introduction to Evolutionary Computing*. New York, USA: Springer, 2003
- [98] Zeiler M D, Fergus R. Visualizing and understanding convolutional networks//*Proceedings of the European Conference on Computer Vision*. Zurich, Switzerland, 2014; 818-833
- [99] Real E, Moore S, Selle A, et al. Large-scale evolution of image classifiers//*Proceedings of the International Conference on Machine Learning*. Sydney, Australia, 2017; 2902-2911
- [100] Guo Zichao, Zhang Xiangyu, Mu Haoyuan, et al. Single path one-shot neural architecture search with uniform sampling//

- Proceedings of the European Conference on Computer Vision. Glasgow, UK, 2020: 544-560
- [101] Sun Yizhou, Han Jiawei, Yan Xifeng, et al. PathSim: Meta path-based top- $k$  similarity search in heterogeneous information networks. Proceedings of the VLDB Endowment, 2011, 4(11): 992-1003
- [102] Huang Zhipeng, Zheng Yudian, Cheng R, et al. Meta structure: Computing relevance in large heterogeneous information networks//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco, USA, 2016: 1595-1604
- [103] Zhu Yanqiao, Xu Weizhi, Zhang Jinghao, et al. Deep graph structure learning for robust representations: A survey. arXiv preprint arXiv:2103.03036, 2021
- [104] You Jiaxuan, Leskovec J, He Kaiming, Xie Saining. Graph structure of neural networks//Proceedings of the International Conference on Machine Learning. 2020: 10881-10891
- [105] Guan Chaoyu, Zhang Ziwei, Li Haoyang, et al. AutoGL: A library for automated graph learning//Proceedings of the ICLR 2021 Workshop on Geometrical and Topological Representation Learning. 2021
- [106] Sen P, Namata G, Bilgic M, et al. Collective classification in network data. AI Magazine, 2008, 29(3): 93-99
- [107] Shchur O, Mumme M, Bojchevski A, Günnemann S. Pitfalls of graph neural network evaluation//Proceedings of the Relational Representation Learning Workshop (NeurIPS 2018). 2018
- [108] Hu Weihua, Fey M, Zitnik M, et al. Open graph benchmark: Datasets for machine learning on graphs. Advances in Neural Information Processing Systems, 2020, 33: 22118-22133
- [109] Morris C, Kriege N M, Bause F, et al. TUDataset: A collection of benchmark datasets for learning with graphs//Proceedings of the ICML 2020 Workshop on Graph Representation Learning and Beyond (GRL+2020). 2020
- [110] Xu Jin, Chen Mingjian, Huang Jianqiang, et al. AutoHENS-GNN: Winning solution to AutoGraph challenge for KDD cup 2020//Proceedings of the 2022 IEEE 38th International Conference on Data Engineering. Kuala Lumpur, Malaysia, 2022
- [111] Errica F, Podda M, Bacciu D, Micheli A. A fair comparison of graph neural networks for graph classification//Proceedings of the International Conference on Learning Representations. Addis Ababa, Ethiopia, 2020
- [112] Yuan Hao, Yu Haiyang, Gui Shurui, Ji Shuiwang. Explainability in graph neural networks: A taxonomic survey. arXiv preprint arXiv:2012.15445, 2020
- [113] Wang Xin, Fan Shuyi, Kuang Kun, Zhu Wenwu. Explainable automated graph representation learning with hyperparameter importance//Proceedings of the International Conference on Machine Learning. 2021: 10727-10737
- [114] Sun Lichao, Dou Yingtong, Yang Carl, et al. Adversarial attack and defense on graph data: A survey. arXiv preprint arXiv:1812.10528, 2018
- [115] Xie Beini, Chang Heng, Zhang Ziwei, et al. Adversarially robust neural architecture search for graph neural networks//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver, Canada, 2023
- [116] Li Haoyang, Wang Xin, Zhang Ziwei, Zhu Wenwu. OOD-GNN: Out-of-distribution generalized graph neural network. IEEE Transactions on Knowledge and Data Engineering, 2022: 1-14
- [117] Li Haoyang, Wang Xin, Zhang Ziwei, Zhu Wenwu. Out-of-distribution generalization on graphs: A survey. arXiv preprint arXiv:2202.07987, 2022
- [118] Xie Saining, Kirillov A, Girshick R, He Kaiming. Exploring randomly wired neural networks for image recognition//Proceedings of the IEEE/CVF International Conference on Computer Vision. Seoul, Korea South, 2019: 1284-1293
- [119] Zhang C, Ren M, Urtasun R. Graph hypernetworks for neural architecture search//Proceedings of the International Conference on Learning Representations. New Orleans, USA, 2018
- [120] Dudziak L, Chau T, Abdelfattah M, et al. BRP-NAS: Prediction-based NAS using GCNs. Advances in Neural Information Processing Systems, 2020, 33: 10480-10490
- [121] Qin Yijian, Wang Xin, Cui Peng, Zhu Wenwu. GQNAS: Graph Q network for neural architecture search//Proceedings of the 2021 IEEE International Conference on Data Mining. Auckland, New Zealand, 2021: 1288-1293
- [122] Auten A, Tomei M, Kumar R. Hardware acceleration of graph neural networks//Proceedings of the 2020 57th ACM/IEEE Design Automation Conference. San Francisco, USA, 2020: 1-6
- [123] Cai Han, Zhu Ligeng, Han Song. ProxylessNAS: Direct neural architecture search on target task and hardware//Proceedings of the International Conference on Learning Representations. New Orleans, USA, 2018
- [124] Tan Mingxing, Chen Bo, Pang Ruoming, et al. MnasNet: Platform-aware neural architecture search for mobile//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 2019: 2820-2828
- [125] Jiang Yuhang, Wang Xin, Zhu Wenwu. Hardware-aware transformable architecture search with efficient search space//Proceedings of the 2020 IEEE International Conference on Multimedia and Expo. London, UK, 2020: 1-6
- [126] Dwivedi V P, Joshi C K, Luu A T, et al. Benchmarking graph neural networks. Journal of Machine Learning Research, 2023, 24(43): 1-48
- [127] Ying C, Klein A, Christiansen E, et al. Nas-bench-101: Towards reproducible neural architecture search//Proceedings

of the International Conference on Machine Learning. Long Beach, USA, 2019; 7105-7114

- [128] Qin Yijian, Zhang Ziwei, Wang Xin, et al. Nas-benchmark: Benchmarking graph neural architecture search// Proceedings of the 36th Conference on Neural Information Processing Systems Datasets and Benchmarks Track. New

Orleans, USA, 2022

- [129] Zhang Ziwei, Wang Xin, Zhu Wenwu. Automated machine learning on graphs: A survey//Proceedings of the 30th International Joint Conference on Artificial Intelligence. Montreal, Canada, 2021



**ZHANG Zi-Wei**, Ph. D. , assistant researcher. His research interests focus on machine learning on graphs, including graph neural network, network embedding, and automated graph machine learning.

**WANG Xin**, Ph. D. , assistant researcher. His research interests include multimedia intelligence, machine learning and its applications in multimedia big data.

**ZHU Wen-Wu**, Ph. D. , professor. His current research interests are in the areas of multimedia computing and networking, and big data.

## Background

Graph neural architecture search lies at the forefront of machine learning and data mining, which fuses the advantages of graph neural networks and automated machine learning. Graph neural architecture search has aroused considerable attention since 2019, with dozens of research papers published in prestigious conferences and journals, such as ICMML, NeurIPS, KDD, AAAI, IJCAI, etc. However, very little effort has been paid to systematically summarizing graph neural architecture search and its recent trends. To the best of our knowledge, only<sup>[30]</sup> has partially surveyed early graph neural architecture search methods. In comparison, our survey has greatly extended the scope, contents, and discussions, e. g. , only 11 graph neural architecture search methods are covered by<sup>[30]</sup> while we cover more than 40 methods. This

paper also has overlaps with our previous survey for automated graph machine learning<sup>[129]</sup>. The difference is that in this paper, we focus on graph neural architecture search by extending the contents as well as providing more in-depth reviews and discussions. This work was supported in part by the National Key Research and Development Program of China (No. 2020AAA0106300), the National Natural Science Foundation of China (Nos. 62250008, 62222209, 62102222, 61936011, 62206149), the Beijing National Research Center for Information Science and Technology (BNRist) (No. BNR2023-RC01003), the Beijing Key Lab of Networked Multimedia, China National Postdoctoral Program for Innovative Talents (No. BX20220185), and the China Postdoctoral Science Foundation (No. 2022M711813).