

## 图强化学习算法及其在工业领域的应用研究综述

李大宇<sup>1</sup>, 刘子博<sup>1</sup>, 包琰洋<sup>1</sup>, 董才波<sup>1</sup>, 徐昕<sup>2\*</sup>

(1. 北京化工大学 信息科学与技术学院, 北京 100029; 2. 国防科技大学 智能科学学院, 湖南 长沙 410073)

**摘要:**强化学习在决策支持、组合优化及智能控制等领域的成功应用推动了其对复杂工业场景的探索,然而现有强化学习方法难以迁移到非欧几里得空间的图结构数据。图神经网络在学习图结构数据方面表现出卓越的性能,为此,通过将图与强化学习结合将图结构数据引入强化学习任务中,丰富了强化学习的知识表征,为解决复杂工业过程问题提供了新范式。系统梳理了图强化学习算法在工业领域的研究进展,从算法架构层面归纳总结图强化学习算法并提炼出了三大主流范式,探讨了其在生产调度、工业知识图谱推理、工业互联网及电力系统领域的应用进展,并分析了当前该领域面临的挑战与未来的发展趋势。

**关键词:**强化学习;图神经网络;图强化学习;图结构数据

中图分类号:TP391.4 文献标志码:A 文章编号:1001-2486(2025)04-076-15



## Research review of graph reinforcement learning algorithms and their applications in the industrial field

LI Dazi<sup>1</sup>, LIU Zibo<sup>1</sup>, BAO Yanyang<sup>1</sup>, DONG Caibo<sup>1</sup>, XU Xin<sup>2\*</sup>

(1. College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China;

2. College of Intelligence Science and Technology, National University of Defense Technology, Changsha 410073, China)

**Abstract:** Successful application of reinforcement learning in decision support, combinatorial optimization, and intelligent control has driven its exploration in complex industrial scenarios. However, existing reinforcement learning methods face challenges in adapting to graph-structured data in non-Euclidean spaces. Graph neural networks have demonstrated exceptional performance in learning graph-structured data. By integrating graphs with reinforcement learning, graph-structured data was introduced into reinforcement learning tasks, enriching knowledge representation in reinforcement learning and offering a novel paradigm for addressing complex industrial process problems. The research progress of graph reinforcement learning algorithms in industrial domains was systematically reviewed, summarized graph reinforcement learning algorithms from the perspective of algorithm architecture and extracted three mainstream paradigms, explored their applications in production scheduling, industrial knowledge graph reasoning, industrial internet, power system and other fields, and analyzed current challenges alongside future development trends in this field.

**Keywords:** reinforcement learning; graph neural network; graph reinforcement learning; graph structured data

随着大数据和深度学习等人工智能技术的迅猛发展,机器学习的一个重要分支——强化学习也得到了显著的发展。强化学习的理论基础逐渐完善,成为解决实际问题的重要工具之一,在各个领域展现出巨大的应用前景,成为学术界和工业界的关注热点<sup>[1-3]</sup>。强化学习算法的基本思路是将策略优化问题建模为马尔可夫决策过程(Markov decision process, MDP),通过智能体与

环境的交互来优化策略,根据观测到的状态做出相应的决策,以最大化累积奖励函数<sup>[4]</sup>。状态和动作是MDP的重要组成部分,分别体现了强化学习算法的感知与行动能力。强化学习优化策略的好坏与状态和动作的设置密切相关。

强化学习传统上多用于处理欧几里得数据结构的环境,但在非欧几里得数据场景下的应用存在显著瓶颈。由于深度神经网络对非欧几里得

收稿日期:2024-12-15

基金项目:国家自然科学基金资助项目(62273026)

第一作者:李大宇(1970—),女,河南鹿邑人,教授,博士,博士生导师,E-mail:lidz@mail.buct.edu.cn

\*通信作者:徐昕(1974—),男,湖北宜昌人,教授,博士,博士生导师,E-mail:xinxu@nudt.edu.cn

引用格式:李大宇,刘子博,包琰洋,等.图强化学习算法及其在工业领域的应用研究综述[J].国防科技大学学报,2025,47(4):76-90.

Citation:LI D Z, LIU Z B, BAO Y Y, et al. Research review of graph reinforcement learning algorithms and their applications in the industrial field[J]. Journal of National University of Defense Technology, 2025, 47(4): 76-90.

数据(如图结构、流形数据)的特征感知能力有限,导致智能体难以捕捉状态间的复杂关联,限制了决策性能的优化。为此,将图结构引入强化学习框架中,在复杂非欧几里得环境中对状态和动作进行高效建模,捕捉数据之间复杂的关联关系,可以有效提高强化学习在此类场景中的计算效率和学习能力。

许多复杂环境可以自然地建模为图结构。例如,交通网络、社交网络、分子结构等都是天然的图结构,其中节点表示环境中的各个元素,边表示元素之间的关系<sup>[5-6]</sup>。图结构具有很强的层次性,采用图表征不仅能够分层表达系统内在规律,还能整合多维信息(静态属性、时序动态等),为机器学习模型提供更精细的特征提取基础<sup>[7]</sup>。通过图神经网络处理图结构数据,通过聚合局部邻域信息有效捕获节点间依赖关系,已在节点分类、链路预测等任务中展现显著优势<sup>[6-9]</sup>。相较于传统深度学习方法,图结构具备更强的鲁棒性,其丰富的上下文信息可抑制噪声与异常值的干扰,而拓扑可视化特性则增强了模型的可解释性,这对于提高模型的可信度和透明度至关重要<sup>[10]</sup>。

强化学习的状态通常是独立的观察点,然而,状态特征间往往存在跨时序、跨层级的复杂关联。图结构通过构建动态图统一建模此类关联,使智能体能够更好地理解环境<sup>[11]</sup>。同时,基于图结构的局部感知机制允许智能体聚焦邻域信息而非全局状态空间,显著降低了高维环境下的计算复杂度。使用图结构作为强化学习状态和动作的表示形式,能提高智能体对任务环境的感知与交互能力,使得智能体具有学习出更好的动作策略的可能<sup>[12-14]</sup>。

本文对现有的工业领域的图强化学习相关文献进行了全面的综述总结。首先对图强化学习方法的框架进行了归纳总结,随后重点介绍和讨论了图强化学习方法在工业领域的应用进展,最后讨论了图强化学习方法面临的问题和解决思路并对未来的发展趋势进行了展望。

## 1 图强化学习的背景与提出

### 1.1 图学习

图学习旨在使用图神经网络等深度学习技术对图结构数据进行特征提取,获得具有全局结构信息的图或节点表征,其中采用的图神经网络是一类专门用于处理图结构数据的深度学习模型。不同于传统深度学习中的全连接机制或是卷积机制,图神经网络针对图结构数据采用了如图1所

示的消息传播与邻域信息聚合机制,能够有效捕捉图中节点之间的复杂关系,实现对图层面或者节点层面的表示学习<sup>[15]</sup>。图神经网络能够在保留全局图结构信息的同时,逐步更新每个节点的表示,不断学习更远邻居的信息,从而实现单个节点对整个图的信息学习<sup>[16-17]</sup>。

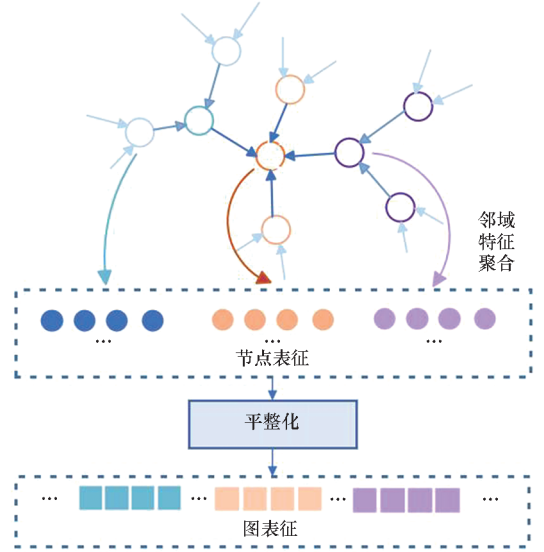


图1 图神经网络机制示意图

Fig.1 Schematic diagram of graph neural network

图卷积神经网络<sup>[18]</sup>被广泛地应用于图学习任务中。其基本思想是基于邻居节点的信息聚合,每一个节点的表示向量通过其邻居节点的表示向量进行更新。这种卷积操作使得图卷积神经网络通过聚合邻居节点的信息有效捕捉图结构中的局部信息,在节点分类和图嵌入等任务中表现出色。

图卷积神经网络的输入为一个图  $G = (V, E)$ ,其中  $V$  是节点集合,  $E$  是边集合。每个节点  $v_i \in V$  有一个特征向量  $x_i$ ,所有节点的特征可以组成特征矩阵  $X \in \mathbb{R}^{N \times F}$ ,  $N$  是节点的数量,  $F$  是每个节点的特征维度。定义邻接矩阵  $A \in \mathbb{R}^{N \times N}$  用于表示图中节点之间的连接关系,  $A_{ij} = 1$  表示节点  $v_i$  与  $v_j$  之间存在连接关系,否则  $A_{ij} = 0$ 。

为了处理不同节点可能具有不同数量的邻居的情况,通常需要对邻接矩阵进行归一化操作。首先计算度矩阵  $D$ ,度矩阵  $D$  是一个对角阵,其中的每个元素计算如下:

$$D_{ii} = \sum_j A_{ij} \quad (1)$$

图卷积神经网络的核心是图卷积操作,图卷积层的计算公式如下:

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

式中:  $\tilde{A} = A + I$ ;  $\tilde{D}$  为  $\tilde{A}$  的度矩阵;  $H^{(l)}$  是第  $l$  层的

节点特征矩阵,初始表示为输入特征矩阵,即  $H^{(0)} = X; W^{(l)}$  是第  $l$  层的权重矩阵;  $\sigma$  表示激活函数。逐层进行图卷积操作,最后一层的输出  $H^{(l)}$  可用于节点分类、链路预测等实际任务。

随着图神经网络算法的发展,研究人员已将其应用于各种图结构数据任务,如节点分类<sup>[5]</sup>、图分类<sup>[19]</sup>、链接预测<sup>[20-21]</sup>等。图神经网络的成功激发了研究人员将其与强化学习结合的兴趣,为图强化学习奠定了基础。

### 1.2 深度强化学习

随着深度学习技术的发展,强化学习开始与深度学习结合,形成了深度强化学习<sup>[22]</sup>。在深度强化学习框架中,深度神经网络被用作函数逼近器,能够自动地从复杂、高维度的状态空间中提取有用的特征以近似状态值函数和策略函数,从而帮助智能体在复杂环境中进行决策。DeepMind 研究团队提出了深度  $Q$  网络,在雅达利游戏上达到了人类水平的表现,这是强化学习领域的一个重大突破<sup>[23]</sup>。随后,AlphaGo 的成功进一步展示了深度强化学习在复杂决策问题中的潜力<sup>[24]</sup>。深度强化学习结合了深度学习的状态特征感知能力和强化学习的决策能力,已成为人工智能研究中的一个重要方向。深度强化学习在自动驾驶<sup>[25-26]</sup>、机器人控制<sup>[27-28]</sup>、工业控制<sup>[29-30]</sup>和金融交易<sup>[31]</sup>等领域得到广泛应用。

现有的深度强化学习方法通常在欧几里得空间中对状态特征进行提取,但是在许多复杂的图结构环境中(社交网络、交通网络、电力网络等),实体之间存在复杂的非欧几里得关系。图神经网络能够有效地捕捉和利用非欧几里得数据中的关系信息,这激发了研究者将图神经网络与强化学习结合的兴趣。通过在深度强化学习中引入图神经网络来对图结构环境进行表示和处理,智能体有望更好地利用实体之间的关系和结构化信息,提升决策性能和泛化能力。

### 1.3 图强化学习

图强化学习是深度强化学习的一种特殊形式,通过将图学习技术与强化学习技术融合,解决非欧几里得环境中的决策问题。图 2 所示为一种基于图神经网络的深度  $Q$  学习算法,使用图卷积神经网络对图结构环境状态信息进行感知编码,通过深度  $Q$  网络学习优化策略解决图结构环境决策问题,开创了图强化学习领域的先河<sup>[32]</sup>。

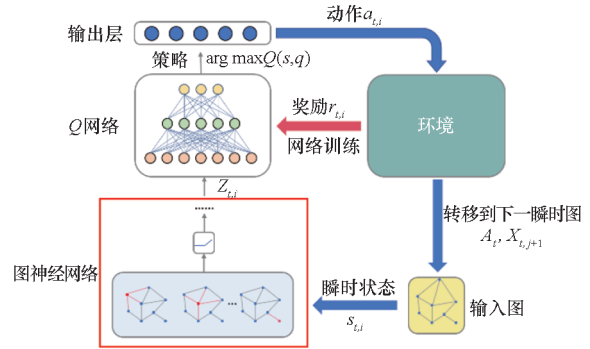


图 2 基于图神经网络的深度  $Q$  学习

Fig. 2 Deep  $Q$ -learning based on graph neural networks

在图强化学习中,所需要解决的任务采用图进行描述,环境由一张或多张图构成,任务信息包含在节点特征与边的连接中,强化学习智能体通过与图中的节点或边的交互来学习最优策略。与其他深度强化学习算法不同,图强化学习算法需要观测并提取图数据的结构信息,将动作反馈到图结构上,并在此基础上进行决策和优化。图神经网络的使用使得图强化学习算法能够更自然地处理图结构数据,学习到图与节点的表征。表 1 展示了图学习、深度强化学习和图强化学习各自算法框架之间的区别与联系。图学习旨在通过图神经网络来学习图或节点的表征,使用消息传递机制与节点更新函数来学习具有全局信息的节点表示,并进行下游的图或节点分类等任务。深度强化学习利用深度神经网络来逼近策略和价值函数,通过智能体在环境中的交互来优化神经网络

表 1 图学习、深度强化学习与图强化学习之间的区别与联系

Tab. 1 Differences and connections between graph learning, deep reinforcement learning, and graph reinforcement learning

算法类型	任务目标	核心机制	网络结构	主要作用	符号表述
图学习	节点/图层级任务	消息传播与节点更新	图神经网络	图上问题表征与特征提取	图表征: $G$
深度强化学习	最优策略	值/策略函数	深度神经网络	环境交互与策略学习	五元组 $M = (S, A, P, R, \gamma)$
图强化学习	最优策略	值/策略函数	图神经网络	二者融合	状态 $S: G$ 动作 $A: G \rightarrow G'$



参数,学习最优策略,以最大化累积奖励。图强化学习是深度强化学习的特殊形式,在深度强化学习的基础上加入了图学习的机制,使用图神经网络对图结构环境进行感知,学习图表征后,将其作为动作选择的依据,采用深度强化学习的方法来学习最优策略。图强化学习中的状态  $S$  直接体现为任务环境的图表征  $G$ ,动作  $A$  则反映为图中节点或边的修改,进而进入新的状态  $S'$ ,即原始图改变后新的图表征  $G'$ 。

## 2 图强化学习方法框架

图强化学习作为一个新兴研究领域,结合了图结构数据与强化学习的优势,受到了学术界和工业界的广泛关注。近年来,研究人员一直在探索将图学习方法与强化学习融合进行决策的方法。在不同的应用场景中,图与强化学习的结合方式多种多样。根据图与强化学习的结合及作用机制,图学习与强化学习结合的方法可分为三类:深度图强化学习、强化学习优化生成图和强化学习探索图。

### 2.1 深度图强化学习

深度图强化学习是三类结合方法中应用最广泛的方法,通常用于图结构化环境中的序列决策问题。在深度图强化学习中,智能体与环境交互的马尔可夫决策过程是在图结构化的环境上完成的,智能体接收图结构化的环境作为状态,并根据策略输出动作来影响环境。深度图强化学习通常用于解决工业领域中的顺序决策问题,应用于工业互联网<sup>[33-42]</sup>、电力系统<sup>[43-50]</sup>、交通控制<sup>[51-55]</sup>、智能装备<sup>[56-61]</sup>领域。基本的深度图强化学习框架如图3所示。深度图强化学习的应用通常包括以下四个步骤:将环境构建为图结构、状态特征嵌入、图神经网络的选择和强化学习方法的选择。

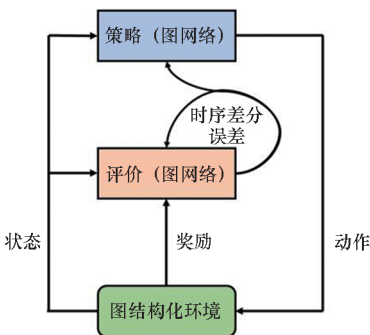


图3 深度图强化学习框架

Fig. 3 Deep graph reinforcement learning framework

在应用深度图强化学习算法时,首先需要构建图结构,需将环境抽象为包含节点、边及其特征

信息的图数据。节点和边根据场景定义:电力系统中节点表示母线,边表示分支;工业互联网中节点为服务器,边为链路连接。大多数场景边仅代表节点之间的连接关系,没有额外的特征信息,一些特定场景(如工业互联网)需引入链路带宽、延迟等边特征。

现有深度强化学习模型通常直接使用全连接神经网络和卷积神经网络对环境状态进行感知,而深度图强化学习接收的是抽象为图结构数据的环境状态,需要使用图神经网络来对输入的状态进行特征信息提取、嵌入等操作。普通神经网络和图卷积神经网络是应用较广泛的两种网络。在一些特殊情况下,为了提高性能会采用其他类型的图神经网络,或者将图神经网络与其他类型的网络进行结合使用。例如有学者使用图注意力网络对异构图进行特征提取<sup>[62]</sup>或解决部分可观测环境中的信息丢失问题,也有学者结合门控循环单元来更好地捕获空间和时间相关性<sup>[52]</sup>。

深度图强化学习采用与深度强化学习相同的方法来学习最优策略。在大部分场景中根据任务需求,基于值的方法(deep  $Q$  network, DQN)、基于策略的方法(策略梯度、近端策略化等)、基于值和策略的演员-评论家(Actor-Critic)方法(优势演员-评论家、异步优势演员-评论家、深度确定性策略梯度等)是常见的选择。深度图强化学习为了能够接收图结构的状态,会将上述选择的图神经网络作为深度强化学习方法中的深度神经网络,例如在DQN中会使用图神经网络作为目标网络和评估网络完成对 $Q$ 值的估计,在Actor-Critic方法中,使用图神经网络替换Actor网络和Critic网络完成对 $Q$ 值和 $V$ 值的拟合。

在智能制造系统、交通信号控制、无人机群等多智能体协同应用场景中有学者使用图神经网络对环境进行感知并训练多个智能体协同完成决策<sup>[41-42, 63-65]</sup>。单智能体方法直接接收整张图结构环境作为状态,感知图结构环境的所有信息并在整张图上进行宏观的决策。而在多智能体方法中,会将每个独立的智能体抽象为图上的节点,环境的信息对单个智能体是部分可观测的。如图4所示,单个智能体只能对图上邻域环境的信息进行感知作为自身的节点特征表示状态,不能感知图上所有的信息。每个智能体通过观察自身的环境状态、分别采取动作共同改变图结构环境,获得奖励来学习各自的策略,以协同实现目标。现有的图上多智能体强化学习的训练方式与普通多智能体强化学习相同。

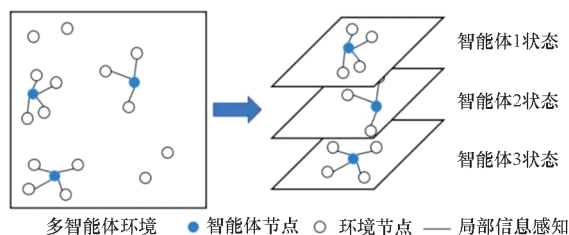


图 4 多智能体环境图上信息部分可观测示意

Fig. 4 Observable schematic of information part on multi-agent environment graph

深度图强化学习的创新之处在于将非欧几里得数据的环境表示为了图结构,图结构化的环境能够更直观地表示实体间的关系和交互,帮助模型更好地聚合和传递信息。同时图神经网络的使用赋予了智能体感知图结构化环境的能力,通过图神经网络有效地从图结构数据中提取和聚合节点特征,使得智能体能够理解和利用环境中的局部和全局信息,从而更好地进行决策。未来,随着算法优化、跨领域应用和理论研究的不断推进,深度图强化学习有望在更多工业领域中发挥重要作用。

## 2.2 强化学习优化生成图

深度图强化学习通过图结构抽象环境提升智能体决策能力,但其图连接关系由环境静态决定,智能体仅能调整节点/边特征而无法改变拓扑结构。使用强化学习优化图结构是强化学习与图学习结合的另一类范式,其核心思想是通过强化学习智能体的决策来主动调整图的连接关系,表现为对图结构中边的添加、删除,从而优化图的结构以实现特定目标,主要应用在工业生产中的组合优化问题<sup>[62-64,66-74]</sup>和分子结构生成任务<sup>[75-79]</sup>。

马尔可夫决策过程完全建立在图上,状态表征为当前图结构及特征,动作定义为对节点/边的增删操作,奖励函数则引导图结构向目标属性演进。大部分情况下该框架采用单步迭代优化机制,智能体在每时间步接收当前图状态,执行局部结构调整,经多步交互逐步逼近最优拓扑。例如在分子生成任务中,智能体逐步接收当前图状态,并逐步调整修改节点/边结构,通过迭代式图演化生成具备目标化学属性的分子结构。图 5 为强化学习单步优化生成分子图的过程示意。

智能体在完成所有图结构优化后才能得到评判优化结果好坏的奖励信号,这使得智能体在学习过程中难以评估其动作的好坏,增大环境探索难度,带来了奖励稀疏的问题。为了应对这个挑战,学者们通常会通过增设中间奖励进行子任务

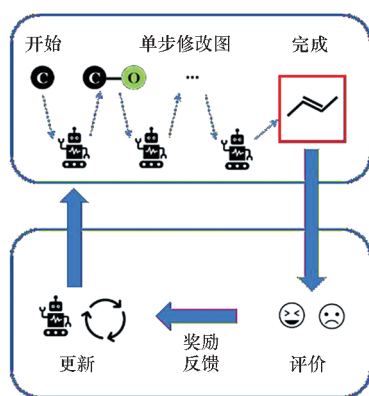


图 5 强化学习单步优化生成分子图过程示意

Fig. 5 Schematic diagram of the process of generating molecular graphs through single step optimization in reinforcement learning

划分或为图神经网络引入序列建模方法(循环神经网络、Transformer 等)跟踪和记忆过去的图修改步骤来指导智能体逐步完成对图结构的构建优化工作<sup>[62,69,79]</sup>。

强化学习优化生成图的基本方法框架与前述的深度图强化学习基本相同,使用图神经网络对图结构环境进行特征提取和聚合后结合深度强化学习进行决策,区别在于两者面向的任务不同,深度图强化学习面向具体的复杂任务,将环境建模为图结构并在图上解决其决策问题,而强化学习优化生成图则用于对具体的图结构对象进行优化或生成操作。

## 2.3 强化学习探索图

强化学习具有很强的探索能力和知识发现能力。使用强化学习对图进行探索,将强化学习用于图上的知识挖掘任务是另一大热门的图与强化学习结合的范式,主要应用于知识图谱的关系推理和推荐系统。

使用强化学习进行关系推理的任务是在知识图谱上的探索过程,智能体从源实体出发,通过实体之间的连接关系移动到相邻的实体并最终到达目标实体,从而得到源实体和目标实体之间的关系,探索过程如图 6 所示。将关系推理问题建模为马尔可夫决策过程,状态集合包括智能体当前所在实体、目标实体和历史轨迹,动作集合为当前所在实体的所有出边信息,状态转移则表现为采取动作后转移到下一实体。使用延迟奖励和即时奖励引导智能体到达目标实体完成推理。

强化学习探索图的另一种应用就是基于知识图谱的推荐系统,推荐系统根据用户的历史交互信息来有针对性地向用户推荐内容。Zhou 等<sup>[80]</sup>

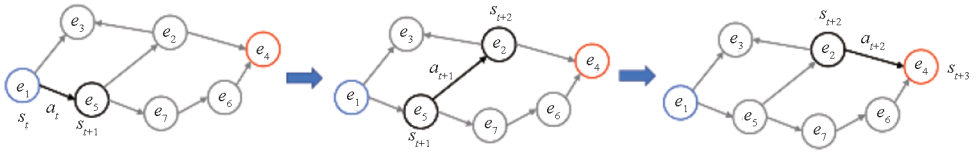


图6 知识图谱关系推理过程示意图

Fig.6 Schematic diagram of knowledge graph reasoning process

提出了用于推荐系统的知识图谱增强  $Q$  学习框架,基于用户的历史交互信息作为种子集,在集成了先验知识的知识图谱上通过邻域信息来动态地生成候选集,用于强化学习最终的决策输出。在知识图谱上进行搜索增强了状态的表示,结合知识图谱的语义关联信息缩小了搜索空间,提高了策略搜索效率。

### 3 图强化学习在工业领域的应用

目前已有一些研究在各个工业场景中应用图强化学习方法,图强化学习在生产调度、工业知识图谱、工业互联网以及电力系统等工业应用中表现出强大的性能。

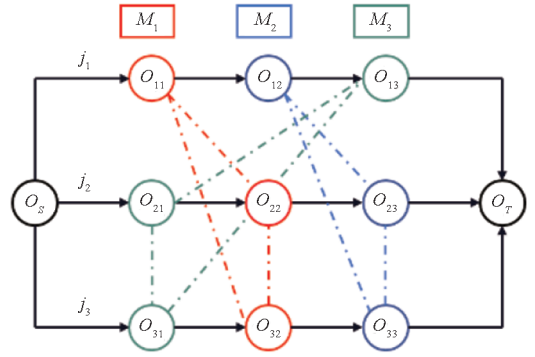
#### 3.1 生产调度

车间调度(生产调度)是运筹学和生产管理中经典的 NP 困难问题,在工业生产中有广泛且重要的应用。解决车间调度问题的核心在于优化生成调度图,由于图强化学习有天然处理图结构数据的优势,近年来图强化学习方法在车间调度问题中的应用引起了研究者的广泛关注<sup>[81]</sup>。

##### 3.1.1 静态车间调度问题

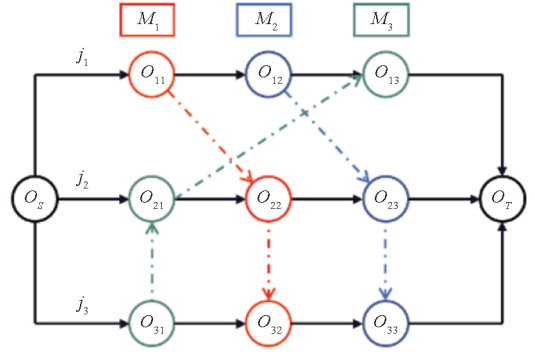
在静态车间调度问题中,每个作业都有一个唯一的操作序列,必须在特定的机器上进行处理。一个标准的静态车间调度问题由一组作业集合和一组机器集合组成,当每个操作仅能在一台特定的机器上完成时为基本的车间调度问题(job shop scheduling problems, JSSPs),当每个操作可以在多个候选机器集合中选择一台完成时为柔性车间调度问题(flexible JSSPs, FJSSPs)。车间调度问题的目标是找到一个调度方案,使总完成时间最短。车间调度问题可以使用如图 7(a)所示的析取图表示,这是一个 3(作业)×3(机器)的 JSSPs 实例。黑色箭头表示同一作业中操作之间优先级的连接弧,虚线是要指定方向的析取弧,颜色相同的析取弧(或操作节点)需要同一台机器进行加工。求解结果是如图 7(b)所示的调度图。

Zhang 等<sup>[66]</sup>首次使用图强化学习方法来解决 JSSPs 问题,使用经过近端策略优化算法训练的图同构网络来学习析取图的嵌入,并选择要调



(a) JSSPs 问题的析取图

(a) Disjunctive graph of JSSPs problem



(b) 生成的调度图

(b) Generated scheduling diagram

图7 静态车间调度问题的析取图表示

Fig.7 Disjunctive graph representation of static JSSPs

度的操作。与 Zhang 等<sup>[66]</sup>的方法类似, Park 等<sup>[67]</sup>使用基于图神经网络的强化学习框架解决车间调度问题,通过图神经网络学习析取图节点特征,将嵌入的节点特征映射到最佳调度动作,具有优秀的泛化性能。类似地,也有学者将柔性制造系统建模为 Petri 网并采用图卷积神经网络来近似深度  $Q$  网络的动作值函数<sup>[68]</sup>。Song 等<sup>[62]</sup>创新性地使用异构图结构来表示作业状态,增加了机器节点及其之间的操作-机器弧,通过降低图密度以及整合机器信息提高了表示效率,使用图注意力网络自动学习不同节点的重要性特征。由于图网络提供的状态质量对强化学习性能有显著影响,Chen 等<sup>[69]</sup>使用 Transformer 融合图嵌入与时序特征增强图网络表征,通过 node2vec 预训练编码器后,采用强化学习从图嵌入中学习车间调



度策略。类似地, Liao 等<sup>[70]</sup>也提出了 Transformer 的方法, 区别在于他们使用近端策略优化算法来训练整个网络架构, 而不需要离线学习。

Chen 等<sup>[71]</sup>将多目标强化学习与图神经网络进行了结合, 他们采用基于凸包的方法来调整环境中不同奖励的权重来对图同构网络进行训练, 为考虑完工时间、完工率和成本的多目标 JSSPs 问题制定了调度策略。

上述方法均属于单智能体强化学习, 在处理复杂的调度任务时可能会遇到困难。因此有学者提出将图神经网络与多智能体强化学习结合以解决复杂的 JSSPs 问题。Jing 等<sup>[63]</sup>将 FJSSPs 建模为拓扑图结构预测过程, 将多个作业视为多智能体, 并使用图卷积神经网络从每个智能体执行的动作中提取特征抽象全局动作表示。Zhang 等<sup>[64]</sup>将每个机器和作业视为智能体, 以机器/作业为节点, 边表示操作顺序, 结合深度  $Q$  网络优化 FJSSPs 的作业排序与路径决策, 取得了优秀的性能。

### 3.1.2 动态车间调度问题

动态作业车间调度问题 (dynamic JSSPs, DyJSSPs) 是对静态车间调度问题的扩展, 考虑了新作业的随机到达、机器故障、作业取消或加工时间的变化等实际生产环境中的动态变化情况。

Yang 等<sup>[72]</sup>通过图卷积神经网络学习节点嵌入, 并使用深度双  $Q$  网络开发策略解决具有随机作业到达的 DyJSSPs 问题。Liu 等<sup>[73]</sup>提出了一种基于深度强化学习的方法, 使用图神经网络来学习调度规则。他们扩展了 L2D 的图表示来处理 DyJSSPs 问题, 利用仅到达作业的操作构建析取图表示来处理随机作业到达和随机机器故障, 处理时间、操作完成程度和机器可用性的信息在动态环境中随着时间更新。

Lei 等<sup>[74]</sup>引入了分层强化学习框架来解决具有新工作到达和灵活机器分配的 DyJSSPs 问题。该框架包括一个高级代理和两个低级代理, 高级代理通过决定何时释放在线到达的作业来将 DyJSSPs 分解为静态的 JSSPs 问题, 两个低级代理分别用来处理机器分配和操作调度。高级代理的 MDP 由缓存作业的数量和重新调度后所有机器完成时间的变化来定义。析取图被用作低级作业操作选择代理的状态表示, 机器分配代理的状态由要分配给机器的操作的完成时间和处理时间组成。图同构网络编码作业操作选择代理的析取图状态表示, 多层感知机用于编码高级代理和机器分配任务状态表示。深度双  $Q$  网络训练高级代

理何时释放作业, 并使用多近端策略优化方法训练两个低级代理。实验结果表明, 所提出的框架在性能上优于调度规则和元启发式等现有的动态调度方法。

### 3.1.3 常用数据集

车间调度问题常用的数据集有 Taillard、OR-Library、Brandimarte、Hurink、Demirkol 等, 这些数据集都是公开可用的。Taillard 和 OR-Library 数据集用于研究 JSSPs, Brandimarte 数据集除包含 JSSPs 外还可用于研究 FJSSPs, Hurink 和 Demirkol 数据集考虑了实际生产环境的动态性, 用于研究复杂的 DyJSSPs。

## 3.2 工业知识图谱

知识图谱是图学习中的关键概念之一, 它是一种结构化的知识库, 用于存储实体之间的关系。利用强化学习强大的探索和知识发现能力, 许多学者将强化学习用于知识图谱的关系推理问题中。Xiong 等<sup>[82]</sup>首次将强化学习方法用于知识图谱的多跳推理问题, 取得了优于基于路径排名的算法和知识图谱嵌入方法的性能。Wang 等<sup>[83]</sup>提出引入自注意力机制的深度强化学习模型用于知识图谱推理问题, 自注意力机制帮助模型有效地推断知识图谱中实体和关系的重要性, 提高推理效率。此外, 有学者使用对抗强化学习方法<sup>[84-85]</sup>, 使用路径鉴别器输出奖励信号, 指导生成器的优化过程, 解决了虚假路径问题。Wan 等<sup>[86]</sup>提出了一种模拟人类思维模式的分层强化学习方法, 将整个推理过程分解为强化学习策略的两个步骤。针对时间知识图谱推理场景, Sun 等<sup>[87]</sup>利用相对时间编码函数获取代理的时间信息, 并利用时间形状奖励指导模型学习。

工业 4.0 与智能制造的快速发展催生了海量多源异构数据, 然而传统数据驱动方法在复杂工业场景下面临解释性与可靠性不足的挑战。知识图谱作为结构化知识库, 通过整合多源数据构建语义化知识体系, 为故障诊断、异常检测等关键任务提供可解释的决策支持。

在工业故障诊断和异常检测领域, 知识图谱可以有效地整合故障相关的知识, 帮助快速定位故障原因、提供诊断建议, 并辅助工程师进行维修决策, 知识图谱用于工业故障诊断过程的流程图如图 8 所示。针对构建故障诊断知识图谱面临数据标注耗时和数据噪声问题, Chen 等<sup>[88]</sup>提出基于强化学习的远监督关系抽取方法, 利用关系感知的句子级注意力机制筛选有效信息, 并通过强化学习动态剔除噪声数据, 构建出高精度故障诊

断知识图谱。Han 等<sup>[89]</sup>则针对钢铁生产线复杂设备场景,构建具有进化能力的双层知识图谱架构。底层图谱通过关系导向模型实现基础故障知识表达,上层则结合强化学习推理与图神经网络嵌入技术,动态补全实体关系缺失,形成高层次的知识图谱。Zhang 等<sup>[90]</sup>提出了基于强化学习的知识图谱推理方法,用于钢铁表面缺陷检测。他们通过融合多源异构数据(工艺参数、化学成分、缺陷图像和操作日志等)构建知识图谱,并利用双智能体强化学习框架进行路径推理,通过路径共享机制和综合奖励函数高效地发现缺陷原因。

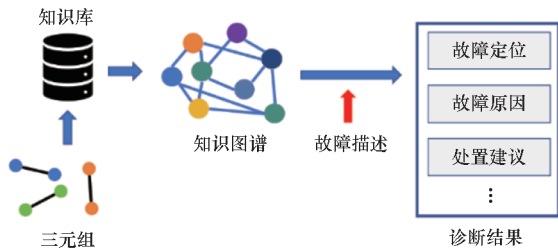


图8 知识图谱用于工业故障诊断过程的流程

Fig. 8 Flowchart of industrial fault diagnosis process knowledge graph

除故障诊断和异常检测之外,也有学者使用基于图强化学习的知识图谱方法对工业过程的因果关系进行推理建模。因果关系揭示了过程变量之间的相互作用机制,可以揭示过程变量之间的关系,准确地反映工业过程的实际运行机制。Zhang 等<sup>[91]</sup>提出的时空因果建模方法,利用图注意力门控循环单元捕捉变量时空演化特征,结合知识引导的强化学习识别真实因果链。通过构建时空因果有向图克服了现有因果建模方法忽视时间延迟和引入虚假因果关联的局限性,在硫回收单元的软测量和氩精馏系统的异常检测两个实际工业案例中验证了其优越性,他们的方法显著提升了模型的预测可靠性和可解释性。

在智能制造系统领域,Zheng 等<sup>[65]</sup>提出了“Self-X”认知制造网络的概念,通过知识图谱为智能体提供任务分解策略与初始配置方案,同时利用智能体优化过程产生的策略反哺知识库更新,通过知识融合和更新进一步丰富知识图谱的内容。通过多机器人协同任务的仿真实验证明了该方法在动态适应任务需求和生成优化解决方案方面的可行性。

### 3.3 工业互联网

随着智能制造技术的快速发展,工业互联网已成为推动工业自动化和智能化的关键技术。工

业互联网通过连接大量的设备、传感器和控制器,实现设备之间的高效通信和协同工作,对网络服务质量(quality of service, QoS)提出了极高的要求。因此,如何对网络资源进行高效的规划和利用已成为急需解决的问题。图强化学习能够自适应地学习网络空间中的复杂关系和动态变化,已经在边缘计算、网络虚拟化和网络规划等场景中取得了优秀的性能。

#### 3.3.1 应用场景

边缘计算是一种分布式计算模式,旨在将计算和数据存储靠近数据源或用户,以提高处理效率并降低延迟,在智能制造系统中有广泛应用。Yan 等<sup>[33]</sup>通过任务调用图来表示移动应用程序中任务之间的复杂依赖关系,并采用深度强化学习来学习最优卸载策略,显著降低计算复杂度。Chen 等<sup>[34]</sup>提出了一种集中式图卷积网络深度强化学习算法,帮助边缘设备在动态环境中做出高效决策。Yao 等<sup>[35]</sup>引入图注意力机制,构建多智能体强化学习框架,有效捕捉边缘节点之间的空间关系,展示了优于现有基准的性能。

当大量异构任务被卸载到边缘服务器时,可能导致资源受限和任务延迟。Gao 等<sup>[36]</sup>提出了一种基于课程注意力加权图循环网络的多智能体方法,使用注意力机制的图神经网络进行信息聚合和通信,提供完全去中心化卸载方案。

网络虚拟化技术(virtual network function, VNF)在互联网架构中扮演了重要角色,通过软件实现网络功能,取代了传统的硬件依赖方式。虚拟网络嵌入(virtual network embedding, VNE)作为网络虚拟化的核心挑战之一,需要将虚拟网络请求有效地映射到物理网络资源上,这是一个NP困难问题。Yan 等<sup>[37]</sup>率先将图卷积网络引入VNE,突破传统卷积神经网络处理非欧几里得数据的局限,其多目标深度强化学习框架显著优化嵌入策略。Sun 等<sup>[38]</sup>提出的DeepOpt算法利用图网络自动学习网络拓扑特征,在VNF部署中展现出优异泛化能力。此外,在动态VNE问题中,Zhang 等<sup>[39]</sup>在前人研究的基础上引入了适应度矩阵和适应度值到图强化学习算法中,可以动态适应虚拟网络拓扑和资源属性的变化。

网络规划是确保网络性能、可靠性和成本效益的关键过程。传统上,这一过程依赖于专家知识或基于整数线性规划和约束编程的求解器,但 these 方法在大规模网络优化问题上存在计算量大、求解时间长等局限性。图强化学习已经被应用在了网络拓扑、无线局域网信道分配和路由优



化问题。Zhu 等<sup>[40]</sup>提出了 NeuroPlan 的模型,结合了图强化学习与整数规划,使用图神经网络和新的域特定节点链路变换来编码网络状态,处理网络拓扑的动态特性,深度强化学习首先用于缩小搜索空间,随后利用整数线性规划求解器在缩小的空间中找到最优解。针对光传输网络的路由优化问题,Almasan 等<sup>[41]</sup>使用图神经网络捕捉网络中的链接和流量关系,通过深度强化学习在不同网络拓扑中进行高效路由决策。

在密集无线局域网信道分配场景中,Nakashima 等<sup>[42]</sup>提出了专用于密集部署的无线局域网信道分配的图强化学习模型,通过图卷积神经网络提取信道特征,该模型能够处理由接入点构成的图结构数据,性能优于仅使用简单神经网络的深度强化学习方法。

### 3.3.2 常用数据集

在工业互联网优化问题的研究中,常用的数据集包括了真实网络流量数据集 Abilene、GÉANT 和 CERNET 等,通过模拟真实网络拓扑和流量而生成的合成网络流量数据集 EBONE (Europe)、Sprintlink (US) 和 Tiscali (Europe) 等,以及通过网络模拟器生成的定制化数据集,比较出名的是 OMNeT++ 网络模拟器,曾在网络规划挑战赛中用于生成模拟数据集来评估不同的解决方案。

## 3.4 电力系统

电力系统为工业生产提供了可靠、稳定的能源支持。随着工业 4.0 和智能制造的兴起,对电力系统的智能化、可靠性和能源质量也提出了更高的要求。电力系统天然可以建模为图结构,使用节点和边来表示电网的拓扑连接,利用图神经网络可以捕捉电力系统中复杂的关系。因此,图强化学习方法在电力系统中具有非常大的应用潜力,为电力系统的智能化动态控制提供了解决方案。目前已有一些研究将图强化学习用于电力系统中输电网络和配电网络的控制。

### 3.4.1 输电网络控制

输电网络是电力系统中用于长距离、大容量电力传输的基础设施,用于将发电厂产生的电力传输到配电网络或直接供应给大型工业用户。图强化学习方法通过动态改变电网拓扑结构来优化输电网络的运行,减少线路过载情况的发生。Xu 等<sup>[43]</sup>使用图神经网络捕捉电网的状态特征,利用双  $Q$  网络来动态调整电网拓扑结构,优化电力流动,减少了线路过载和功率损耗。针对大规模动作空间的问题,Xu 等<sup>[44]</sup>在使用图注意力网络进行状态特征提取的基础上,使用蒙特卡罗树搜索

在大规模动作空间中高效地筛选和评估动作,通过模拟环境结果反馈动态调整搜索方向,缩小了动作搜索空间,提高了决策效率。Van Der Sar 等<sup>[45]</sup>利用输电网络的分层特性提出了三层分级多智能体框架,顶层决策行动触发,中间层协调变电站操作顺序,底层由独立智能体执行母线切换,分层协作实现复杂电网拓扑的协同优化。

### 3.4.2 配电网络控制

在电力系统中,配电网络在电力系统的末端,负责将电力从输电网络分配到最终用户,它直接与电力消费者相连,电压控制是配电网络的关键任务。图强化学习在电压控制中的应用主要分为常规运行控制和紧急情况控制。

在常规运行控制中,图强化学习方法通过优化分布式能源(蓄电池、电容器、光伏逆变器等)的无功功率输出来维持电压稳定。Lee 等<sup>[46]</sup>将电力系统表示为图结构(节点为母线、边为线路),将分布能源和电压调节器的状态编码到节点特征中,利用图卷积网络融合拓扑关系与设备状态进行策略决策。Wu 等<sup>[47]</sup>提出了两阶段电压调节框架来应对可再生能源(光伏)导致的电压波动问题,日前阶段通过整数规划调度变压器、电容器和储能设备,实时阶段采用图强化学习动态调整光伏逆变器无功功率,在不平衡配电系统上表现出色。Mu 等<sup>[48]</sup>使用多智能体强化学习方法,各分布式能源通过局部观测协同调整无功输出,有效处理大规模配电网的复杂拓扑结构。

在紧急情况下,图强化学习方法通过负载削减避免配电网络崩溃。Hossain 等<sup>[49]</sup>提出了一种基于图卷积神经网络的深度强化学习框架,专注于应对电力系统的拓扑变化,并自动执行负载削减策略以确保电压稳定控制。Jacob 等<sup>[50]</sup>使用结合了图神经网络和前馈网络的图胶囊神经网络捕捉图的结构信息,提供了更丰富的状态表示,实现电压稳定控制。

### 3.4.3 常用仿真环境

在电力系统中,IEEE 标准测试系统与 Grid2Op 是两个典型环境。IEEE 系统提供多规模标准案例,可在 MATPOWER、Pandapower 等工具中直接调用,但其固定拓扑和参数难以模拟设备故障、负载波动等动态场景,适用于电力系统静态分析。Grid2Op 则是面向强化学习设计的动态仿真平台,支持自定义电网模型和实时运行模拟,能够反映电网拓扑重构、电压调节、故障恢复及可再生能源波动等复杂工况,为动态控制策略研究提供环境支撑。

## 4 图强化学习面临的问题及解决思路

随着图强化学习的发展,该方法被越来越多地用来解决工业场景中的实际问题。然而,图强化学习仍存在问题,使得在复杂工业场景中部署时仍面临较大困难,如何完善图强化学习的理论,并将图强化学习推广到更广阔的工业应用场景是未来的重要研究方向。

### 4.1 图表征的不准确问题及解决思路

图嵌入质量对下游强化学习性能具有关键影响,但当前图学习存在两个不足:一是嵌入失真问题,欧氏空间的线性特性难以捕捉图的层次关系,导致节点几何关系无法反映内在层级结构<sup>[92-93]</sup>;二是过平滑问题,随着图神经网络层数增加,节点特征通过邻域聚合趋于同质化,降低模型对节点特征的区分能力<sup>[94-95]</sup>。

针对层次结构失真问题,双曲空间图嵌入提供了一种有效解决方案<sup>[96]</sup>。其负曲率特性天然适合建模分层数据,可更精准表征节点间的层级关系,提升嵌入保真度。对于过平滑问题,一种思路是优化模型架构,引入残差连接<sup>[97]</sup>或通过跳跃传播保留浅层特征<sup>[98]</sup>,也有学者改进注意力机制减少冗余信息<sup>[99]</sup>。另一种解决思路是在训练过程中采用随机删边等正则化方法增加邻域多样性,缓解冗余信息聚合。

### 4.2 数知分离问题(冷启动问题)及解决思路

图强化学习虽通过图神经网络增强环境感知能力,但仍面临深度强化学习类似的冷启动问题。现有方法通常直接使用图神经网络替代传统深度神经网络,但未充分融合领域知识,导致智能体需从零开始探索陌生图结构。由于图表示学习对先验信息的利用不足,智能体仅能通过低效交互获取图特征,形成“数知分离”的困境<sup>[6]</sup>。

初始图的提供者通常拥有丰富的先验知识,理论上可以指导图表示的学习过程,提高学习表示的准确性和有效性。为了更好地解决图强化学习中的冷启动问题,可以考虑如何有效融合先验知识与数据驱动的方法。一种思路是开发混合模型,将领域知识编码到图结构中,作为智能体学习过程的指导,提高模型的学习效率。另外,利用图的先验信息进行预训练也是一种有效策略,可以通过专家系统或历史数据提取有价值的先验信息,将其整合到图网络中,作为初始状态或参数进行微调,从而减少模型训练所需的时间和资源。

### 4.3 迁移能力弱问题及解决思路

尽管图强化学习在处理复杂网络空间问题中

展现了巨大的潜力,但其迁移能力较弱的问题仍限制了其广泛应用。当智能体迁移至图结构变化或新网络环境时,往往需要重新训练,无法有效复用已有知识,导致效率显著降低。该问题的核心在于现有方法未能充分挖掘图结构间的共性特征,过度依赖特定环境的数据驱动模式,缺乏跨环境的特征归纳机制。同时,智能体在训练过程中未能建立有效的关键特征识别与适应框架,导致知识迁移受阻。

为了解决这个问题,有学者使用基于图卷积网络的归纳图强化学习,可以推广到新网络拓扑中,无须额外的训练<sup>[53]</sup>。此外,也有学者引入元学习框架来提升图强化学习的迁移能力<sup>[56]</sup>,模型不仅对新环境具有很强的可迁移性,而且也能很好地克服拓扑结构的变化。

### 4.4 数据稀疏问题及解决思路

与传统的强化学习类似,图强化学习也可能面临数据稀疏的问题。数据稀疏问题主要表现在节点和边的稀疏问题、奖励信号稀疏问题两个方面。

图强化学习中节点和边的稀疏问题表现为图数据中节点或边的数量较少,或者节点之间的连边较少,这种情况会影响到学习过程中的信息传播和特征提取,使得模型难以捕捉到有效的结构信息。解决思路是采用图数据增强(对节点和边进行增广)、图嵌入(DeepWalk、node2vec和GraphSAGE)方法,将节点和边嵌入低维向量空间中,捕捉潜在的结构特征和节点关系<sup>[100-102]</sup>。此外,也可以考虑子图采样(使用随机游走或邻域采样来获取更丰富的训练样本)和自监督学习(生成内部标签来提供更多的训练信号)等方法来克服节点和边的稀疏问题<sup>[103-104]</sup>。

在图强化学习优化修改图的任务中,智能体需完成全图修改才能获得反馈,导致训练信号过于稀疏,难以有效学习和优化策略。解决思路是设计高质量的奖励函数,可以通过奖励塑形(设计额外的、密集的奖励信号引导智能体逐步接近目标)<sup>[105]</sup>、层次化任务(将任务分解为多个子任务,利用高层策略和低层策略的协调来解决稀疏奖励问题)<sup>[106]</sup>以及与逆强化学习技术结合(从专家示范行为中推导高质量的奖励函数)等方法,缓解图强化学习中的奖励信号稀疏问题,提高智能体在复杂环境中的学习和决策能力。

## 5 图强化学习的未来技术展望

图强化学习作为人工智能的前沿方向,必将

在未来扮演重要的角色。将最新的思想填充到图强化学习的思想中,弥补算法的缺陷,提高算法的计算效率、学习能力和决策性能是未来研究的重要方向。以下是对图强化学习的未来研究方向的展望。

### 5.1 与强化学习方法的深入结合

目前,图强化学习算法主要是图神经网络和深度强化学习算法的结合。将图学习与逆强化学习、多目标强化学习等先进的强化学习算法结合有望进一步提高图强化学习的性能。

逆强化学习是一种从观察到的行为中推断潜在奖励函数的技术,用于克服强化学习奖励函数设计困难的问题。在复杂的任务中,图强化学习也存在奖励函数设计困难的问题,此外,图强化学习的数据稀疏问题也对奖励函数的设计提出了更高的要求。将逆强化学习与图强化学习结合,为后者寻找出一个高质量的奖励函数,有望显著提升智能体在复杂图结构环境中的学习能力,拓宽应用范围<sup>[107]</sup>。

多目标强化学习是一种用于处理具有多个目标或奖励信号的强化学习方法,在许多工业场景中,单一目标的优化往往不足以描述复杂的任务需求,因此需要同时考虑多个目标,例如在自动驾驶中需要平衡速度和安全性,在资源管理中需要同时优化收益和风险<sup>[108]</sup>。将图强化学习与多目标强化学习结合,智能体可以在图结构中识别和优化多个关键节点和路径,在不同目标之间找到最佳平衡点,从而提高整体系统的性能。Chen等<sup>[71]</sup>将多目标强化学习与图神经网络进行了结合,他们采用基于凸包的方法来调整环境中不同奖励的权重来对图同构网络进行训练,为考虑完工时间、完工率和成本的多目标车间调度问题制定了调度策略。图强化学习与多目标强化学习结合的研究仍然处于起步尝试阶段,有望提升智能体在复杂工业场景中的多目标优化能力,拓宽其应用范围。

### 5.2 与多智能体系统深入结合

多智能体强化学习考虑多个智能体之间的通信,使它们能够有效地协作学习。现有的将图与多智能体强化学习进行结合的研究较少且应用场景有限,处于起步尝试阶段。

现有的方法大多数是将多个智能体置于图结构环境上对邻域进行特征感知并决策,学者们并没有考虑多个智能体之间的混合关系,即在训练过程中,智能体之间的相互作用和信息交换并没

有得到充分的利用。可以考虑将图与多智能体强化学习进行更深入的融合,将从多智能体中学习到特征和这些智能体之间的关系信息融合在一起,充分挖掘实体和关系之间的复杂交互,使其能够处理更复杂和大规模的多智能体环境。图强化学习与多智能体系统的深入结合是未来的研究方向。

### 5.3 与大语言模型相结合

图强化学习与大语言模型的结合为人工智能提供了跨模态推理新范式。以 BERT、DeepSeek 为代表的大语言模型凭借深度语义理解能力,可解析图结构中的文本描述信息;而图强化学习通过构建环境图谱与策略优化,擅长处理复杂关系网络。两者的协同有望突破图文异构数据表征的割裂<sup>[109-110]</sup>。

语义嵌入作为跨模态对齐的核心技术,需兼顾文本语境与图结构特征。传统方法如 word2vec 通过词共现构建语义空间,GraphSAGE、DeepWalk 等图嵌入则聚焦节点拓扑关系,但均存在显著局限,文本嵌入忽视图结构特征,图嵌入缺乏文本上下文感知,且均无法建模图层级语义关联<sup>[111]</sup>。这种割裂导致跨模态信息难以在同一向量空间实现深度融合。

理想的图语义嵌入应构建多层次表征框架。在节点层面,融合文本描述与大语言模型的上下文编码能力,增强局部结构表征;在图层层面上,结合强化学习的策略优化机制,捕捉全局语义拓扑。通过建立图节点-文本描述的双向注意力机制,可将异源信息映射至共享潜在空间,实现节点属性、文本语境与网络拓扑的联合优化。这种融合架构不仅能提升链接预测、节点分类等任务的准确性,更重要的是为复杂系统的推理决策提供可解释的跨模态知识图谱,推动人机协同智能的发展。

## 6 结束语

作为目前人工智能两大方向的前沿交叉,图强化学习旨在解决目前强化学习对非欧几里得结构数据的复杂问题,特别是可以用图表征的问题表征不准确的缺陷。该算法通过将图网络引入强化学习,使强化学习智能体获得从图视角来表征问题的能力,在增强算法环境感知与特征提取能力的同时,拓宽了应用场景。综合考虑图强化学习在人工智能领域的研究意义和应用潜力,对其的研究在解决复杂决策问题、推动智能化设备发展等方面具有重要价值。本文重点介绍了图强化



学习算法在工业领域的研究进展,对现有的图强化学习方法进行了归纳总结并给出了三大范式,讨论了图强化学习在工业中的典型应用场景,总结了图强化学习算法目前存在的不足,并给出了未来可能的进一步研究方向。

未来的研究工作需要加强对图强化学习理论的深入探索,将最新的思想填充到图强化学习算法中,推动其在算法设计和性能优化方面取得更大突破。同时,要将图强化学习算法应用于更广泛更复杂的领域。期待图强化学习算法能够不断改进和完善,为自动化智能化工业的建设带来更多可能性和机遇。

## 参考文献 (References)

- [1] ZHANG Z, WANG D Q, GAO J W. Learning automata-based multiagent reinforcement learning for optimization of cooperative tasks[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, 32(10): 4639–4652.
- [2] ZHAN H X, TAO F, CAO Y C. Human-guided robot behavior learning: a GAN-assisted preference-based reinforcement learning approach [J]. *IEEE Robotics and Automation Letters*, 2021, 6(2): 3545–3552.
- [3] LI J N, DING J L, CHAI T Y, et al. Nonzero-sum game reinforcement learning for performance optimization in large-scale industrial processes [J]. *IEEE Transactions on Cybernetics*, 2020, 50(9): 4132–4145.
- [4] WANG X, WANG S, LIANG X X, et al. Deep reinforcement learning: a survey [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, 35(4): 5064–5078.
- [5] HU S D, XIONG Z, QU M, et al. Graph policy network for transferable active learning on graphs [C]//*Proceedings of 34th Conference on Neural Information Processing Systems (NeurIPS 2020)*, 2020.
- [6] ZHOU J, CUI G Q, HU S D, et al. Graph neural networks: a review of methods and applications[J]. *AI Open*, 2020, 1: 57–81.
- [7] 马帅, 刘建伟, 左信. 图神经网络综述[J]. *计算机研究与发展*, 2022, 59(1): 47–80.  
MA S, LIU J W, ZUO X. Survey on graph neural networks[J]. *Journal of Computer Research and Development*, 2022, 59(1): 47–80. (in Chinese)
- [8] RUSEK K, SUÁREZ-VARELA J, MESTRES A, et al. Unveiling the potential of graph neural networks for network modeling and optimization in SDN [C]//*Proceedings of the 2019 ACM Symposium on SDN Research*, 2019: 140–151.
- [9] SUÁREZ-VARELA J, CAROL-BOSCH S, RUSEK K, et al. Challenging the generalization capabilities of graph neural networks for network modeling [C]//*Proceedings of the ACM SIGCOMM 2019 Conference Posters and Demos*, 2019: 114–115.
- [10] TIEZZI M, CIRAVEGNA G, GORI M. Graph neural networks for graph drawing[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, 35(4): 4668–4681.
- [11] JIN J R, ZHOU S J, ZHANG W N, et al. Graph-enhanced exploration for goal-oriented reinforcement learning [C]//*Proceedings of the Tenth International Conference on Learning Representations*, 2022.
- [12] LIU H, ZHOU S W, CHEN C F, et al. Dynamic knowledge graph reasoning based on deep reinforcement learning [J]. *Knowledge-Based Systems*, 2022, 241: 108235.
- [13] XI J H, ZHU F H, YE P J, et al. Auxiliary network enhanced hierarchical graph reinforcement learning for vehicle repositioning [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2024, 25(9): 11563–11575.
- [14] LI Y, LUO X F, XIE S R. Learning heterogeneous strategies via graph-based multi-agent reinforcement learning [C]//*Proceedings of the IEEE 33rd International Conference on Tools with Artificial Intelligence (ICTAI)*, 2021: 709–713.
- [15] SCARSELLI F, GORI M, TSOI A C, et al. The graph neural network model [J]. *IEEE Transactions on Neural Networks*, 2009, 20(1): 61–80.
- [16] 刘鑫, 梅红岩, 王嘉豪, 等. 图神经网络推荐方法研究 [J]. *计算机工程与应用*, 2022, 58(10): 41–49.  
LIU X, MEI H Y, WANG J H, et al. Research on graph neural network recommendation method [J]. *Computer Engineering and Applications*, 2022, 58(10): 41–49. (in Chinese)
- [17] REN L F, HU R M, LI D S, et al. Dynamic graph neural network-based fraud detectors against collaborative fraudsters [J]. *Knowledge-Based Systems*, 2023, 278: 110888.
- [18] ZHAO Y J, MA Y, HU S L. USV formation and path-following control via deep reinforcement learning with random braking [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, 32(12): 5468–5478.
- [19] SUN Q Y, LI J X, PENG H, et al. SUGAR: subgraph neural network with reinforcement pooling and self-supervised mutual information mechanism [C]//*Proceedings of the Web Conference*, 2021: 2081–2091.
- [20] ZHANG M H, CHEN Y X. Link prediction based on graph neural networks [C]//*Proceedings of the 32nd International Conference on Neural Information Processing Systems*, 2018: 5171–5181.
- [21] TROUILLON T, WELBL J, RIEDEL S, et al. Complex embeddings for simple link prediction [C]//*Proceedings of the 33rd International Conference on International Conference on Machine Learning*, 2016: 2071–2080.
- [22] 赵冬斌, 邵坤, 朱圆恒, 等. 深度强化学习综述: 兼论计算机围棋的发展 [J]. *控制理论与应用*, 2016, 33(6): 701–717.  
ZHAO D B, SHAO K, ZHU Y H, et al. Review of deep reinforcement learning and discussions on the development of computer go [J]. *Control Theory & Applications*, 2016, 33(6): 701–717. (in Chinese)
- [23] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning [J]. *Nature*, 2015, 518(7540): 529–533.
- [24] SILVER D, SCHRITTWIESER J, SIMONYAN K, et al. Mastering the game of go without human knowledge [J]. *Nature*, 2017, 550(7676): 354–359.
- [25] CODEVILLA F, MÜLLER M, LÓPEZ A, et al. End-to-end driving via conditional imitation learning [C]//*Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2018: 4693–4700.
- [26] SHI T Y, WANG P, CHENG X X, et al. Driving decision and control for autonomous lane change based on deep reinforcement learning [EB/OL]. (2019–07–30) [2024–01–10]. <https://arxiv.org/abs/1904.10171>.
- [27] LILLICRAP T P, HUNT J J, PRITZEL A, et al. Continuous control with deep reinforcement learning [EB/OL]. (2019–07–05) [2024–01–10]. <https://arxiv.org/abs/>

- 1509.02971.
- [28] DUAN Y, CHEN X, HOUTHOOFT R, et al. Benchmarking deep reinforcement learning for continuous control [C]//Proceedings of the 33rd International Conference on International Conference on Machine Learning, 2016: 1329 – 1338.
- [29] PANDIAN B J, NOEL M M. Tracking control of a continuous stirred tank reactor using direct and tuned reinforcement learning based controllers[J]. Chemical Product and Process Modeling, 2017, 13(3): 20170040.
- [30] ONG K S H, WANG W B, NIYATO D, et al. Deep-reinforcement-learning-based predictive maintenance model for effective resource management in industrial IoT [J]. IEEE Internet of Things Journal, 2022, 9(7): 5173 – 5188.
- [31] HUANG C Y. Financial trading as a game: a deep reinforcement learning approach[EB/OL]. (2018-07-08) [2024-02-02]. <https://arxiv.org/abs/1807.02787>.
- [32] DAI H J, KHALIL E B, ZHANG Y Y, et al. Learning combinatorial optimization algorithms over graphs [C]//Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017: 6351 – 6361.
- [33] YAN J, BI S Z, ZHANG Y J A. Offloading and resource allocation with general task graph in mobile edge computing: a deep reinforcement learning approach[J]. IEEE Transactions on Wireless Communications, 2020, 19(8): 5404 – 5419.
- [34] CHEN J, WU Z L. Dynamic computation offloading with energy harvesting devices: a graph-based deep reinforcement learning approach[J]. IEEE Communications Letters, 2021, 25(9): 2968 – 2972.
- [35] YAO Z X, XIA S C, LI Y, et al. Cooperative task offloading and service caching for digital twin edge networks: a graph attention multi-agent reinforcement learning approach [J]. IEEE Journal on Selected Areas in Communications, 2023, 41(11): 3401 – 3413.
- [36] GAO Z, YANG L, DAI Y. Fast adaptive task offloading and resource allocation in large-scale MEC systems via multiagent graph reinforcement learning [J]. IEEE Internet of Things Journal, 2024, 11(1): 758 – 776.
- [37] YAN Z X, GE J G, WU Y L, et al. Automatic virtual network embedding: a deep reinforcement learning approach with graph convolutional networks [J]. IEEE Journal on Selected Areas in Communications, 2020, 38(6): 1040 – 1057.
- [38] SUN P H, LAN J L, LI J F, et al. Combining deep reinforcement learning with graph neural networks for optimal VNF placement [J]. IEEE Communications Letters, 2021, 25(1): 176 – 180.
- [39] ZHANG P Y, WANG C, KUMAR N, et al. Dynamic virtual network embedding algorithm based on graph convolution neural network and reinforcement learning [J]. IEEE Internet of Things Journal, 2021, 9(12): 9389 – 9398.
- [40] ZHU H, GUPTA V, AHUJA S S, et al. Network planning with deep reinforcement learning [C]//Proceedings of the 2021 ACM SIGCOMM 2021 Conference, 2021: 258 – 271.
- [41] ALMASAN P, SUÁREZ-VARELA J, RUSEK K, et al. Deep reinforcement learning meets graph neural networks: exploring a routing optimization use case [J]. Computer Communications, 2022, 196: 184 – 194.
- [42] NAKASHIMA K, KAMIYA S, OHTSU K, et al. Deep reinforcement learning-based channel allocation for wireless LANs with graph convolutional networks [J]. IEEE Access, 2020, 8: 31823 – 31834.
- [43] XU P D, PEI Y Z, ZHENG X H, et al. A simulation-constraint graph reinforcement learning method for line flow control [C]//Proceedings of the IEEE 4th Conference on Energy Internet and Energy System Integration (EI2), 2020: 319 – 324.
- [44] XU P D, DUAN J J, ZHANG J, et al. Active power correction strategies based on deep reinforcement learning—part I: a simulation-driven solution for robustness [J]. CSEE Journal of Power and Energy Systems, 2022, 8(4): 1122 – 1133.
- [45] VAN DER SAR E, ZOCCA A, BHULAI S. Multi-agent reinforcement learning for power grid topology optimization [EB/OL]. (2023-10-04) [2024-02-16]. <https://arxiv.org/abs/2310.02605>.
- [46] LEE X Y, SARKAR S, WANG Y B. A graph policy network approach for volt-var control in power distribution systems [J]. Applied Energy, 2022, 323: 119530.
- [47] WU H Y, XU Z, WANG M H, et al. Two-stage voltage regulation in power distribution system using graph convolutional network-based deep reinforcement learning in real time [J]. International Journal of Electrical Power & Energy Systems, 2023, 151: 109158.
- [48] MU C X, LIU Z Y, YAN J, et al. Graph multi-agent reinforcement learning for inverter-based active voltage control [J]. IEEE Transactions on Smart Grid, 2024, 15(2): 1399 – 1409.
- [49] HOSSAIN R R, HUANG Q H, HUANG R K. Graph convolutional network-based topology embedded deep reinforcement learning for voltage stability control [J]. IEEE Transactions on Power Systems, 2021, 36(5): 4848 – 4851.
- [50] JACOB R A, PAUL S, CHOWDHURY S, et al. Real-time outage management in active distribution networks using reinforcement learning over graphs [J]. Nature Communications, 2024, 15: 4766.
- [51] NISHI T, OTAKI K, HAYAKAWA K, et al. Traffic signal control based on reinforcement learning with graph convolutional neural nets [C]//Proceedings of the 21st International Conference on Intelligent Transportation Systems (ITSC), 2018: 877 – 883.
- [52] ZHANG Y, YU Z W, ZHANG J, et al. Learning decentralized traffic signal controllers with multi-agent graph reinforcement learning [J]. IEEE Transactions on Mobile Computing, 2024, 23(6): 7180 – 7195.
- [53] DEVAÏLLY F X, LAROCQUE D, CHARLIN L. IG-RL: inductive graph reinforcement learning for massive-scale traffic signal control [J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(7): 7496 – 7507.
- [54] KUYER L, WHITESON S, BAKKER B, et al. Multiagent reinforcement learning for urban traffic control using coordination graphs [C]//Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, 2008: 656 – 671.
- [55] PENG H, DU B W, LIU M S, et al. Dynamic graph convolutional network for long-term traffic flow prediction with reinforcement learning [J]. Information Sciences, 2021, 578: 401 – 416.
- [56] GAMMELLI D, YANG K D, HARRISON J, et al. Graph neural network reinforcement learning for autonomous mobility-on-demand systems [C]//Proceedings of the 60th IEEE Conference on Decision and Control (CDC), 2021: 2996 – 3003.
- [57] CHEN S K, DONG J Q, HA P (Y J), et al. Graph neural network and reinforcement learning for multi-agent cooperative control of connected autonomous vehicles [J]. Computer-Aided Civil and Infrastructure Engineering, 2021, 36(7): 838 – 857.

- [58] MAW A A, TYAN M, NGUYEN T A, et al. iADA\* -RL: anytime graph-based path planning with deep reinforcement learning for an autonomous UAV [J]. *Applied Sciences*, 2021, 11(9): 3948.
- [59] YE Z H, WANG K, CHEN Y N, et al. Multi-UAV navigation for partially observable communication coverage by graph reinforcement learning [J]. *IEEE Transactions on Mobile Computing*, 2023, 22(7): 4056–4069.
- [60] CHEN F F, MARTIN J D, HUANG Y W, et al. Autonomous exploration under uncertainty via deep reinforcement learning on graphs [C]//*Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020: 6140–6147.
- [61] LI R, JABRI A, DARRELL T, et al. Towards practical multi-object manipulation using relational reinforcement learning [C]//*Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, 2020: 4051–4058.
- [62] SONG W, CHEN X Y, LI Q Q, et al. Flexible job-shop scheduling via graph neural network and deep reinforcement learning [J]. *IEEE Transactions on Industrial Informatics*, 2023, 19(2): 1600–1610.
- [63] JING X, YAO X F, LIU M, et al. Multi-agent reinforcement learning based on graph convolutional network for flexible job shop scheduling [J]. *Journal of Intelligent Manufacturing*, 2024, 35(1): 75–93.
- [64] ZHANG J D, HE Z X, CHAN W H, et al. DeepMAG: deep reinforcement learning with multi-agent graphs for flexible job shop scheduling [J]. *Knowledge-Based Systems*, 2023, 259: 110083.
- [65] ZHENG P, XIA L Q, LI C X, et al. Towards Self-X cognitive manufacturing network: an industrial knowledge graph-based multi-agent reinforcement learning approach [J]. *Journal of Manufacturing Systems*, 2021, 61: 16–26.
- [66] ZHANG C, SONG W, CAO Z G, et al. Learning to dispatch for job shop scheduling via deep reinforcement learning [C]//*Proceedings of the 34th International Conference on Neural Information Processing System*, 2020: 1621–1632.
- [67] PARK J, CHUN J, KIM S H, et al. Learning to schedule job-shop problems: representation and policy learning using graph neural network and reinforcement learning [J]. *International Journal of Production Research*, 2021, 59(11): 3360–3377.
- [68] HU L, LIU Z Y, HU W F, et al. Petri-net-based dynamic scheduling of flexible manufacturing system via deep reinforcement learning with graph convolutional network [J]. *Journal of Manufacturing Systems*, 2020, 55: 1–14.
- [69] CHEN R Q, LI W X, YANG H B. A deep reinforcement learning framework based on an attention mechanism and disjunctive graph embedding for the job-shop scheduling problem [J]. *IEEE Transactions on Industrial Informatics*, 2023, 19(2): 1322–1331.
- [70] LIAO Z J, CHEN J B, ZHANG Z Z. Solving job-shop scheduling problem via deep reinforcement learning with attention model [C]//*Proceedings of 36th International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, 2023: 201–212.
- [71] CHEN S R, TIAN Y M, AN L L. Multi-objective order scheduling via reinforcement learning [J]. *Algorithms*, 2023, 16(11): 495.
- [72] YANG Z, BI L, JIAO X G. Combining reinforcement learning algorithms with graph neural networks to solve dynamic job shop scheduling problems [J]. *Processes*, 2023, 11(5): 1571.
- [73] LIU C L, HUANG T H. Dynamic job-shop scheduling problems using graph neural network and deep reinforcement learning [J]. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2023, 53(11): 6836–6848.
- [74] LEI K, GUO P, WANG Y, et al. Large-scale dynamic scheduling for flexible job-shop with random arrivals of new jobs by hierarchical reinforcement learning [J]. *IEEE Transactions on Industrial Informatics*, 2024, 20(1): 1007–1018.
- [75] DONG C B, LI D Z, LIU J. Glass transition temperature prediction of polymers via graph reinforcement learning [J]. *Langmuir*, 2024, 40(35): 18568–18580.
- [76] CHEN Z D, LI D Z, LIU M H, et al. Graph neural networks with molecular segmentation for property prediction and structure-property relationship discovery [J]. *Computers & Chemical Engineering*, 2023, 179: 108403.
- [77] YOU J X, LIU B W, YING R, et al. Graph convolutional policy network for goal-directed molecular graph generation [C]//*Proceedings of the 32nd International Conference on Neural Information Processing Systems*, 2018: 6412–6422.
- [78] ATANCE S R, DIEZ J V, ENGVIST O, et al. De novo drug design using reinforcement learning with graph-based deep generative models [J]. *Journal of Chemical Information and Modeling*, 2022, 62(20): 4863–4872.
- [79] KHEMCHANDANI Y, O'HAGAN S, SAMANTA S, et al. DeepGraphMolGen, a multi-objective, computational strategy for generating molecules with desirable properties: a graph convolution and reinforcement learning approach [J]. *Journal of Cheminformatics*, 2020, 12(1): 53.
- [80] ZHOU S J, DAI X Y, CHEN H K, et al. Interactive recommender system via knowledge graph-enhanced reinforcement learning [C]//*Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2020: 179–188.
- [81] WANG Y S, TONG Y X, LONG C, et al. Adaptive dynamic bipartite graph matching: a reinforcement learning approach [C]//*Proceedings of the IEEE 35th International Conference on Data Engineering (ICDE)*, 2019: 1478–1489.
- [82] XIONG W H, HOANG T, WANG W Y. DeepPath: a reinforcement learning method for knowledge graph reasoning [EB/OL]. (2018–07–07) [2024–03–01]. <https://sites.cs.ucsb.edu/~william/papers/DeepPath.pdf>.
- [83] WANG Q, HAO Y S, CAO J. ADRL: an attention-based deep reinforcement learning framework for knowledge graph reasoning [J]. *Knowledge-Based Systems*, 2020, 197: 105910.
- [84] WANG Q, JI Y D, HAO Y S, et al. GRL: knowledge graph completion with GAN-based reinforcement learning [J]. *Knowledge-Based Systems*, 2020, 209: 106421.
- [85] CUI H, PENG T, HAN R D, et al. Path-based multi-hop reasoning over knowledge graph for answering questions via adversarial reinforcement learning [J]. *Knowledge-Based Systems*, 2023, 276: 110760.
- [86] WAN G J, PAN S R, GONG C, et al. Reasoning like human: hierarchical reinforcement learning for knowledge graph reasoning [C]//*Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, 2020: 1926–1932.
- [87] SUN H H, ZHONG J L, MA Y P, et al. TimeTraveler: reinforcement learning for temporal knowledge graph forecasting [EB/OL]. (2021–09–09) [2024–03–15]. <https://arxiv.org/abs/2109.04101>.



- [88] CHEN C, WANG T, ZHENG Y, et al. Reinforcement learning-based distant supervision relation extraction for fault diagnosis knowledge graph construction under industry 4.0[J]. *Advanced Engineering Informatics*, 2023, 55: 101900.
- [89] HAN H H, WANG J, WANG X W, et al. Construction and evolution of fault diagnosis knowledge graph in industrial process [J]. *IEEE Transactions on Instrumentation and Measurement*, 2022, 71: 3522212.
- [90] ZHANG Y F, WANG H W, SHEN W M, et al. DuAK: reinforcement learning-based knowledge graph reasoning for steel surface defect detection [J]. *IEEE Transactions on Automation Science and Engineering*, 2025, 22: 557 – 569.
- [91] ZHANG X R, SONG C Y, ZHAO J, et al. Spatial-temporal causality modeling for industrial processes with a knowledge-data guided reinforcement learning[J]. *IEEE Transactions on Industrial Informatics*, 2024, 20(4): 5634 – 5646.
- [92] NICKEL M, KIELA D. Poincaré embeddings for learning hierarchical representations [C]//*Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017: 6341 – 6350.
- [93] BRONSTEIN M M, BRUNA J, LECUN Y, et al. Geometric deep learning: going beyond euclidean data[J]. *IEEE Signal Processing Magazine*, 2017, 34(4): 18 – 42.
- [94] LI Q M, HAN Z C, WU X M. Deeper insights into graph convolutional networks for semi-supervised learning [J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018, 32(1): 3538 – 3545.
- [95] OONO K, SUZUKI T. Graph neural networks exponentially lose expressive power for node classification [EB/OL]. (2019 – 05 – 27) [2024 – 03 – 16]. <https://arxiv.org/abs/1905.10947>.
- [96] CHAMBERLAIN B P, CLOUGH J, DEISENROTH M P. Neural embeddings of graphs in hyperbolic space[EB/OL]. (2017 – 05 – 29) [2024 – 03 – 16]. <https://arxiv.org/abs/1705.10359>.
- [97] CHEN M, WEI Z W, HUANG Z F, et al. Simple and deep graph convolutional networks[C]//*Proceedings of the 37th International Conference on Machine Learning*, 2020: 1725 – 1735.
- [98] GASTEIGER J, BOJCHEVSKI A, GÜNNEMANN S. Predict then propagate: graph neural networks meet personalized pagerank [C]//*Proceedings of International Conference on Learning Representations (ICLR)*, 2019.
- [99] HAN J Q, HUANG W B, RONG Y, et al. Structure-aware DropEdge toward deep graph convolutional networks [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, 35(11): 15565 – 15577.
- [100] HAMILTON W L, YING R, LESKOVEC J. Inductive representation learning on large graphs[C]//*Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017: 1025 – 1035.
- [101] PEROZZI B, AL-RFOU R, SKIENA S. DeepWalk: online learning of social representations [C]//*Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2014: 701 – 710.
- [102] GROVER A, LESKOVEC J. Node2vec: scalable feature learning for networks [C]//*Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016: 855 – 864.
- [103] HU W H, LIU B W, GOMES J, et al. Strategies for pre-training graph neural networks [C]//*Proceedings of the 8th International Conference on Learning Representations*, 2020.
- [104] CHEN J, MA T F, XIAO C. FastGCN: fast learning with graph convolutional networks via importance sampling [C]//*Proceedings of the 6th International Conference on Learning Representations*, 2018.
- [105] SANG J H, WANG Y L, DING W P, et al. Reward shaping with hierarchical graph topology [J]. *Pattern Recognition*, 2023, 143: 109746.
- [106] KULKARNI T D, NARASIMHAN K R, SAEEDI A, et al. Hierarchical deep reinforcement learning: integrating temporal abstraction and intrinsic motivation [EB/OL]. (2016 – 05 – 31) [2024 – 10 – 10]. <https://arxiv.org/abs/1604.06057>.
- [107] 宋莉, 李大字, 徐昕. 逆强化学习算法、理论与应用研究综述[J]. *自动化学报*, 2024, 50(9): 1704 – 1723.
- SONG L, LI D Z, XU X. A survey of inverse reinforcement learning algorithms, theory and applications [J]. *Acta Automatica Sinica*, 2024, 50(9): 1704 – 1723. (in Chinese)
- [108] LI D Z, GU W T, SONG T H. Multi-objective reinforcement learning in process control: a goal-oriented approach with adaptive thresholds [J]. *Journal of Process Control*, 2023, 129: 103063.
- [109] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding [C]//*Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2019: 4171 – 4186.
- [110] UENO L, LYNN T. GPT-4o: the comprehensive guide and explanation [EB/OL]. (2024 – 05 – 14) [2025 – 01 – 16]. <https://blog.roboflow.com/gpt-4o-vision-use-cases>.
- [111] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality [C]//*Proceedings of the 27th International Conference on Neural Information Processing Systems*, 2013: 3111 – 3119.