基于双注意力网络深度强化学习的柔性作业车间节能调度方法

曾 丹[1,2] 杨 武[1,2] 杜彦斌[1,2*] 吕 岩[1,2] 李聪波[3]
1.重庆工商大学智能装备绿色设计与制造重庆市重点实验室,重庆,400067
2.重庆工商大学机械工程学院,重庆,400067
3.重庆大学机械与运载工程学院,重庆,400044

摘要:针对柔性作业车间节能调度研究中工序-机器复杂关系建模不足,算法效率与解质量难以兼顾,冗余信息干扰决策等问题,以最小化最大完工时间与机器总能耗为优化目标,提出一种基于双注意力网络深度强化学习的调度优化方法。首先,构建双目标调度问题数学模型并转化为马尔可夫决策过程;设计包含工序消息Agent与机器消息Agent的双注意力网络,用于分别建模工序优先级约束和机器竞争关系与加工资格,精准捕捉工序-机器复杂关联关系并滤除冗余调度信息。其次,在深度强化学习框架下将工序选择与机器分配整合为联合决策问题,通过双目标协同优化机制生成工序-机器对的优先级概率分布,平衡运行效率与解质量。最后,在合成数据集与Mk1~Mk7基准数据集上开展对比实验,结果表明所提方法综合性能优势显著,可有效平衡解质量与运行效率,兼具实用性与优化性能。

关键词:柔性作业车间节能调度问题;深度强化学习;双注意力网络;多目标优化
中图分类号:TP39
DOI:10.3969/j.issn.1004-132X.2026.10.014 **开放科学(资源服务)标识码(OSID)**:

Energy-efficient Flexible Job Shop Scheduling Method via Dual Attention Network Based Deep Reinforcement Learning

ZENG Dan[1,2] YANG Wu[1,2] DU Yanbin[1,2*] LYU Yan[1,2] LI Congbo[3]
1.Chongqing Key Laboratory of Green Design and Manufacturing of Intelligent Equipment, Chongqing Technology and Business University, Chongqing, 400067
2.College of Mechanical Engineering, Chongqing Technology and Business University, Chongqing, 400067
3.College of Mechanical and Vehicle Engineering, Chongqing University, Chongqing, 400044

Abstract: To address the issues in existing research on the EEFJSP, including inadequate modeling of complex operation-machine relationships, the persistent trade-off between algorithm efficiency and solution quality, and decision-making interference from redundant information, this paper proposed a scheduling optimization method based on deep reinforcement learning (DRL) with a dual attention network (DAN), targeting the minimization of makespan and total machine energy consumption. First, the bi-objective scheduling problem was mathematically formulated and further transformed into a Markov decision process, upon which a DAN consisting of an operation message agent and a machine message attention agent was developed to separately characterize operation precedence constraints, machine competition relationships and processing eligibility, enabling accurate capture of intricate operation-machine correlations while eliminating redundant scheduling information. Subsequently, within a DRL framework, operation selection and machine assignment were integrated into a joint decision-making problem, where a bi-objective collaborative optimization mechanism generated the priority probability distribution of operation-machine pairs to achieve a well-balanced trade-off between operation efficiency and solution quality. Finally, comparative experiments on synthetic datasets and the Mk1-Mk7 public benchmark datasets verified that the proposed method achieved prominent comprehensive performance, effectively reconciling solution quality with operation efficiency and delivering both strong practicability and superior optimization capability.

Key words: energy-efficient flexible job shop scheduling problem (EEFJSP); deep reinforcement learning; dual attention network; multi-objective optimization

收稿日期:2025-11-18
基金项目:重庆市自然科学基金(CSTB2022NSCQ-MSX1337);
重庆工商大学高层次引进人才科研启动费(2256014)

0 引言

降低制造环节能耗、推进绿色低碳制造已成为新时代制造业转型升级的必然趋势。柔性作业车间节能调度问题(energy-efficient flexible job shop scheduling problem，EEFJSP)作为更契合现代制造企业生产需求的典型场景，在为工件的每道工序合理分配加工机器时，需综合考虑机器能耗与碳排放等节能目标，从而实现综合效益的提升。该问题是柔性作业车间调度问题(FJSP)的重要延伸，属于典型的NP-hard问题[1]。

近年来，EEFJSP研究持续深化，理论方法创新与工程应用均取得显著进展，其求解范式从传统精确算法、启发式算法逐步演进至元启发式算法，乃至融合机器学习的前沿技术方法[2]。精确算法虽可求得最优解，但时间复杂度较高，仅适用于小规模调度问题。启发式算法以优先级调度规则(如先进先出、最短加工时间)为核心，求解速度较快，但解的质量难以保证，且全局寻优能力较弱。相较而言，以遗传算法[3]等为代表的元启发式算法凭借其强大的全局搜索能力和良好的鲁棒性，能够在可接受的时间内获得质量较高的近似最优解，从而在求解效率与结果精度之间实现有效平衡，因而在EEFJSP研究中得到广泛应用[4-7]。然而，随着生产场景问题规模持续扩大，元启发式方法求解多目标EEFJSP时面临计算复杂度显著上升、收敛速度明显放缓等问题，且在完工时间、能耗与成本等多目标相互冲突的优化场景下难以有效兼顾求解效率与解集质量，制约了其在大规模调度问题中的实际应用。

为突破上述技术瓶颈，基于深度强化学习(DRL)[8]的调度方法逐渐受到关注。该类方法通过与环境持续交互，能够从样本数据中自主学习调度策略，并实现跨问题的知识迁移，从而展现出较强的适应性与泛化能力。针对EEFJSP，现有研究普遍以最小化最大完工时间与机器加工总能耗(或碳排放)为调度目标，在算法层面进行了多样化探索[9-14]。部分研究将深度强化学习策略(如深度Q网络、强化学习协同进化策略等)与传统遗传算法框架或邻域搜索机制相融合，形成混合改进算法，以增强全局搜索能力与求解效率[9-10,12-13]；另有研究基于深度强化学习算法(如深度Q网络、A2C-Gauss等)构建调度决策框架，并通过大规模实例验证算法性能[11,14]。

尽管DRL在EEFJSP研究中已取得一定进展，但现有方法仍存在以下局限：①状态表示不够精确，原始调度状态中混入已完工工序等无关信息，造成状态空间冗余，影响策略学习效率与调度性能；②多目标优化场景下目标间存在固有冲突，现有方法难以兼顾求解效率与解质量，易陷入局部最优，优化结果难以匹配实际生产的综合需求；③现有模型对工序间、工序与机器间以及机器间的复杂耦合关系建模不充分，尤其忽视机器资源的分配竞争，削弱了调度策略在实际约束下的适应性与鲁棒性。

为解决上述问题，本文在前期研究成果[15]的基础上，以最小化最大完工时间和机器总能耗为优化目标，构建多目标优化模型，将调度过程形式化为马尔可夫决策过程(MDP)并明确定义其状态空间、动作空间与奖励函数；针对状态表示冗余问题设计紧凑状态表征方法，为化解多目标冲突构建兼顾完工时间与能耗的复合奖励函数；为弥补工序与机器间耦合关系建模的不足，提出基于图注意力机制[16]的双注意力网络(DAN)，通过工序消息Agent与机器消息Agent提取关键特征并动态建模资源竞争关系。最后在DRL框架下将工序选择与机器分配整合为联合决策，输出可行工序-机器组合的优先级概率分布，实现双目标协同优化。

1 EEFJSP模型

1.1 问题描述

车间有n个待加工工件，记工件集$J=\{J_1,J_2,\cdots,J_n\}$，有m台加工机器，记加工机器集$M=\{M_1,M_2,\cdots,M_m\}$，每个工件J_i包含多道工序，每道工序$O_{i,j}$可以在M中的一台或多台机器上加工，同一道工序在不同机器上的加工时间和产生的能耗不同。此外，EEFJSP问题还具有以下假设：①所有工件和机器在零时刻均处于可用状态；②同一机器在某时刻只能加工一道工序；③同一工件在某时刻只能有一道工序处于加工状态且每道工序仅加工一次；④每个工件的加工优先级相同且工序一旦开始加工，不允许中断；⑤加工过程中不考虑机器故障；⑥不考虑机器的启动时间。

1.2 EEFJSP模型建立

本文建立的EEFJSP模型如下：

$$f_1=\min C_{\max}=\min\{C_i\mid i=1,2,\cdots,n\} \tag{1}$$

$$f_2=\sum_{k=1}^{m}[(\sum_{i=1}^{n}\sum_{j=1}^{n_i}X_{i,j,k}t_{i,j,k})E_{c,k}+(C_{\max}-S_t)E_{u,k}] \tag{2}$$

$$\text{s.t.}\quad S_{i,j}+X_{i,j,k}t_{i,j,k}\leqslant C_{i,j} \tag{3}$$

$$C_{i,j}\leqslant S_{i,j+1} \tag{4}$$

$$C_{i,n_i}\leqslant C_{\max} \tag{5}$$

$$S_{i,j}+t_{i,j,k}\leqslant S_{i',j'}+N(1-Y_{i,j,i',j',k}) \tag{6}$$

$$C_{i,j}\leqslant S_{i,j+1}+N(1-Y_{i,j,i',j'+1,k}) \tag{7}$$

$$\sum_{k=1}^{M_{i,j}}X_{i,j,k}=1 \tag{8}$$

$$S_{i,j}\geqslant 0, C_{i,j}\geqslant 0 \tag{9}$$

式中：C_i为工件的完工时间；$C_{\max}$为最大完工时间；n_i为工件J_i的工序总数；$t_{i,j,k}$为工序$O_{i,j}$在机器M_k上的加工时间；$M_{i,j}$为工件J_i的第j道工序可选机器；$X_{i,j,k}$取值0或1，如果工序$O_{i,j}$选择机器M_k加工则为1，否则为0；$E_{c,k}$为机器M_k的加工功率；$E_{u,k}$为机器M_k的待机功率；S_t为系统的开始时间；$S_{i,j}$为工序$O_{i,j}$的开始时间；$C_{i,j}$为工序$O_{i,j}$的完工时间；$Y_{i,j,i',j'+1,k}$取值0或1，如果工序$O_{i,j}$先于工序$O_{i',j'}$在机器M_k上加工则为1，否则为0。

式(1)和式(2)分别为最小化最大完工时间与机器总能耗的目标函数，机器总能耗由机器加工能耗和待机能耗两部分构成。约束式(3)表示工件某道工序的完工时间应不小于该工序的开始时间与加工时间之和。约束式(4)表示同一工件的工序加工存在先后顺序，前一道工序完成后才能开始后续工序。约束式(5)表示各工件的完工时间不能超过总的完工时间。约束式(6)和约束式(7)表示同一机器在同一时刻只能加工一道工序(其中N为足够大的正数)。约束式(8)表示同一时刻工件的某道工序只能由其可用机器中的一台进行加工。约束式(9)表示每个工件某道工序的开始时间和完工时间均为非负实数。

2　方法设计

2.1　马尔可夫决策过程建模

多目标EEFJSP本质上属于一类序列决策过程，体现为在每一决策步中将某工件的特定工序分配至兼容机器，并逐步推进直至所有工序完成调度。本文提出的总体方法框架如图1所示，具体流程如下：①将问题建模为MDP，通过引入更为精确的状态表示，整合与决策相关的工序及机器信息，同时为了便于后续将调度状态嵌入图注意力机制进行处理，本文采用析取图[17]来表示EEFJSP问题，具体表示方式如图2所示；②在此基础上，利用工序间的优先级约束与机器间的竞争关系，将工件工序状态特征向量与机器状态特征向量输入DAN进行深度特征提取；③将提取的特征与每个兼容的工序-机器对状态特征聚合为新的特征向量，并将其输入至决策网络，生成各可选动作的概率分布；④通过采样机制选取当前调度动作，并同步更新系统状态；⑤重复上述过程，直至所有工序调度完毕，从而得到近似最优的调度方案。

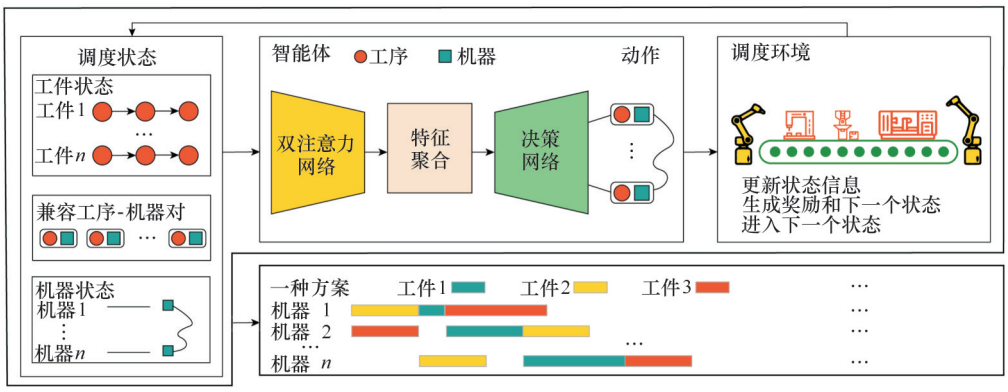
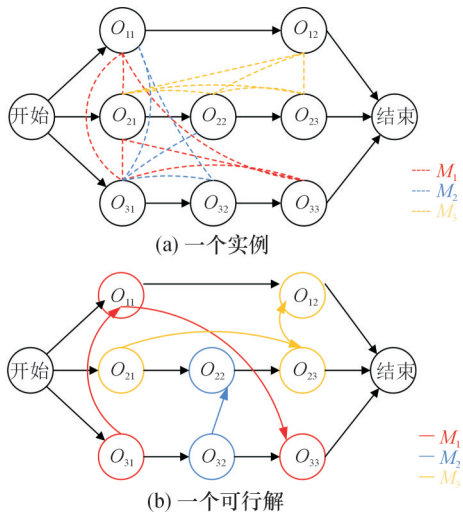


图1　基于双注意力网络深度强化学习的柔性作业车间节能调度方法总体框架

Fig.1　Overall framework for EEFJSP via dual attention network based deep reinforcement learning

与WANG等[18]的方法类似，本文的MDP建模如下。

1)状态。状态是调度环境的主要属性，表示与决策相关的工序和机器信息的特征向量集合，本文中状态主要分为工序状态、机器状态以及每个兼容的工序-机器对状态。每道工序$O_{i,j}$都包含4个静态属性和6个动态属性：①所兼容机器的最短加工时间；②所兼容机器的平均加工时间；③所兼容机器之间加工时间的跨度；④可加工工序$O_{i,j}$的机器比例；⑤调度标签，如果工序$O_{i,j}$被调度，则为1，否则为0；⑥工序$O_{i,j}$预计的完成时间下限$\underline{C}(O_{i,j})$；⑦工件J_i的剩余工序数；⑧工件J_i的剩余未加工工序的平均加工时间之和；⑨工序$O_{i,j}$的等待时间；⑩工序$O_{i,j}$的剩余被加工时间。每台机器M_k都包含2个静态属性和6个动态属性：①所兼容工序的最短加工时间；②所兼容工序的平均加工时间；③机器M_k所兼容的未加工工序数；④机器M_k所兼容的候选工序数；⑤机器M_k的空闲时刻；⑥机器M_k的待机时间；⑦加工标签，如果机器M_k空闲，则为0，否则为1；⑧机器M_k的剩余加工时间。每个所兼容的工序-机器对($O_{i,j}$，M_k)包含2个静态属性和6个动态属性：①工序



(a) 一个实例

(b) 一个可行解

图2 EEFJSP的一个实例和一个可行解

Fig.2 An instance and a feasible solution for EEFJSP

$O_{i,j}$在机器M_k上的加工时间$t_{i,j,k}$；②$t_{i,j,k}$与工序$O_{i,j}$最大加工时间之比；③$t_{i,j,k}$与机器M_k所兼容的候选工序最大加工时间之比；④$t_{i,j,k}$与未加工工序的最大加工时间之比；⑤$t_{i,j,k}$与机器M_k所兼容的未加工工序的最大加工时间之比；⑥$t_{i,j,k}$与所兼容的工序-机器对$(O_{i,j},M_k)$的最大加工时间之比；⑦$t_{i,j,k}$与工件J_i的剩余未加工工序的平均加工时间和之比；⑧工序$O_{i,j}$与机器M_k的等待时间之和。

2)动作。本文将动作定义为工序-机器对的复合决策，所有可行的工序-机器对视为动作空间，随着调度过程的进行，工序被不断地分配给所兼容的机器，动作空间会逐渐减少。

3)状态转换。每一个决策步骤t，从当前状态s_t开始，智能体会从动作空间中选择一个动作，在执行动作后，环境状态会发生改变，从而进入下一个状态s_{t+1}。

4)奖励。本研究的目标是协同优化最小化最大完工时间和机器总能耗，奖励函数r_t设计的目的是指导智能体能够选择有助于减少调度过程中所有工序的最大完工时间和所有机器的总能耗的动作。为使模型能够兼顾效率与解的质量，避免陷入局部最优，本文将奖励函数r_t设置为决策步骤t与$t+1$之间工序的预估最大完工时间之差r_1以及机器总能耗之差r_2的和：

$$r_t=r_1+r_2 \tag{10}$$

$$r_1=\max \underline{C}(O_{i,j},s_t)-\max \underline{C}(O_{i,j},s_{t+1}) \tag{11}$$

$$r_2=f_2(s_t)-f_2(s_{t+1}) \tag{12}$$

5)策略。采用随机策略$\pi(a_t|s_t)$来定义在给定状态s_t下每一个动作a_t的概率分布，该策略的分布是由可训练参数的DRL方法生成的，在训练过程中优化特定参数以最大化积累奖励。最后，使用贪婪算法选取概率最大的动作作为当前状态下需要执行的动作。

2.2 双注意力网络

多目标EEFJSP的调度过程核心在于逐步确定工序与机器的分配关系。图注意力机制能够有效建模序列元素之间的依赖关系，通过捕捉元素间的结构化关联，识别出对优化目标具有关键影响的元素，这一特性使其特别适用于识别高优先级工序，并有效区分机器间隐含的竞争关系，从而为调度决策提供更具判别力的特征表示。为此，本文引入基于图注意机制的DAN系统以捕捉工序和机器之间的复杂交互关系。该网络总体框架如图3所示，由L个结构相同但参数各异的工序消息Agent与机器消息Agent构成。其中，工序消息Agent以工序状态特征为输入，通过建模工序间的优先级约束，增强关键工序的特征表达；机器消息Agent则以机器状态特征为输入，通过分析机器间隐含的竞争关系，综合处理每台机器的加工状态与所兼容待加工工序的加工资格，从而实现对机器资源动态能力的准确建模。

2.2.1 工序消息Agent

对于每个具有输入特征$\boldsymbol{h}_{O_{i,j}}\in\mathbf{R}^{10}$的工序$O_{i,j}$，工序消息Agent通过计算注意力系数$e_{i,j,p}$，对工序$O_{i,j}$的前继$O_{i,j-1}$和后继$O_{i,j+1}$工序之间的关系建模：

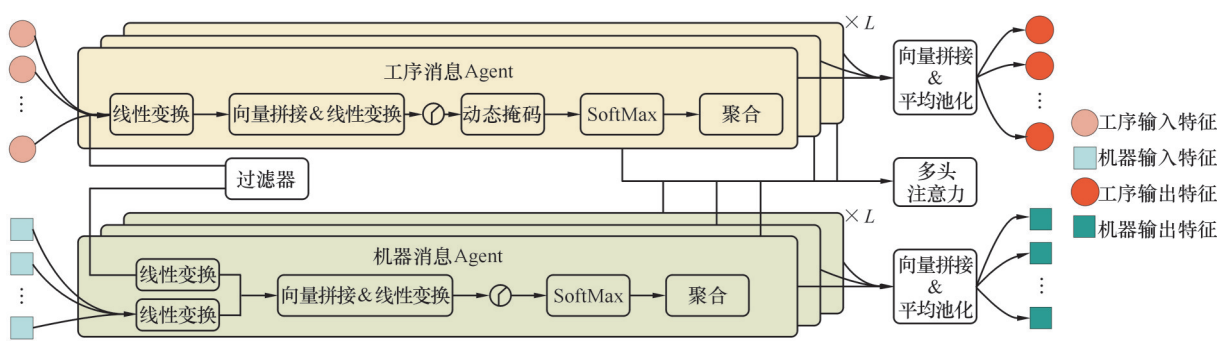


图3 双注意力网络框架

Fig.3 Framework of the dual attention network

$$e_{i,j,p}=LeakyReLU(\boldsymbol{a}^{\mathrm{T}}[(\boldsymbol{W}\boldsymbol{h}_{O_{i,j}})\|(\boldsymbol{W}\boldsymbol{h}_{O_{i,p}})]) \quad (13)$$

其中，$|p-j|\leqslant 1$；$\boldsymbol{a}^{\mathrm{T}}$ 和 $\boldsymbol{W}$ 属于线性变换；*LeakyReLU* 表示激活函数；$\|$表示向量的拼接，由于某些工序不存在前继或者后继工序或将在某个步骤中被删除，因此对它们的注意力系数使用了动态掩码。

然后采用Softmax对注意力系数进行归一化操作，得到归一化后的注意力系数 $\alpha_{i,j,p}$，最后输入特征 $\boldsymbol{W}\boldsymbol{h}_{O_{i,j-1}}$、$\boldsymbol{W}\boldsymbol{h}_{O_{i,j}}$ 和 $\boldsymbol{W}\boldsymbol{h}_{O_{i,j+1}}$ 经过线性变换后与非线性激活函数 σ 进行运算得到工序的输出特征向量 $\boldsymbol{h}'_{O_{i,j}}$：

$$\boldsymbol{h}'_{O_{i,j}}=\sigma\left(\sum_{p=j-1}^{j+1}\alpha_{i,j,p}\boldsymbol{W}\boldsymbol{h}_{O_{i,p}}\right) \quad (14)$$

2.2.2　机器消息 Agent

机器之间可以对所兼容的未加工的候选工序进行竞争，并且随着调度过程的进行，竞争会发生变化。本文定义 $\boldsymbol{C}_{k,q}=\sum_{O_{i,j}\in C_{k,q}\cap J_c}\boldsymbol{h}_{O_{i,j}}$ 来衡量机器 M_k 和机器 M_q 之间竞争的强度，$\boldsymbol{C}_{k,q}$ 的值越大，它们竞争的候选工序重要性越高，二者间的竞争就越激烈，图3中的过滤器专用于获取机器 M_k 和机器 M_q 的 $\boldsymbol{C}_{k,q}$。对于每个具有输入特征 $\boldsymbol{h}_{M_k}\in\mathbf{R}^9$ 的机器 M_k，注意力系数 $u_{k,q}$ 计算式为

$$u_{k,q}=LeakyReLU(\boldsymbol{b}^{\mathrm{T}}[(\boldsymbol{Z}^1\boldsymbol{h}_{M_k})\|(\boldsymbol{Z}^1\boldsymbol{h}_{M_q})\|(\boldsymbol{Z}^2\boldsymbol{C}_{k,q})]) \quad (15)$$

式中：$\boldsymbol{b}^{\mathrm{T}}$、$\boldsymbol{Z}^1$ 和 $\boldsymbol{Z}^2$ 表示线性变换。

机器的输出特征向量 $\boldsymbol{h}'_{M_k}$ 可参照工序注意力块对注意力系数的变换方法得到。

2.2.3　多头注意力和平均池化

为了增强网络的学习能力，工序注意力块和机器注意力块均使用了 H 个多头注意力机制，然后将工序的输出特征向量和机器的输出特征向量经过 L 次平均池化后，通过向量拼接成总的全局特征向量 $\boldsymbol{h}_{\mathrm{G}}^{(L)}$：

$$\boldsymbol{h}_{\mathrm{G}}^{(L)}=\left[\left(\frac{1}{|O|}\sum_{O_{i,j}\in O}\boldsymbol{h}_{O_{i,j}}^{(L)}\right)\|\left(\frac{1}{|M|}\sum_{M_k\in M}\boldsymbol{h}_{M_k}^{(L)}\right)\right] \quad (16)$$

式中：O、M 分别为在当前决策步骤 t 时相关工序和机器的集合；$\boldsymbol{h}_{O_{i,j}}^{(L)}$、$\boldsymbol{h}_{M_k}^{(L)}$ 分别为经过 L 次平均池化后的工序输出特征向量和机器输出特征向量。

2.3　决策网络

如图4所示，本文设计了一个基于演员-评论家结构的强化学习决策网络框架，使用多层感知机（multilayer perceptron，MLP）来代表演员网络和评论家网络，参数分别由 θ 和 ϕ 表示。对每个决策步骤 t 处的每个可行动作 a_t，将DAN输出的特征向量 $\boldsymbol{h}_{O_{i,j}}^{(L)}$、$\boldsymbol{h}_{M_k}^{(L)}$、$\boldsymbol{h}_{\mathrm{G}}^{(L)}$ 以及每个兼容的工序-机器对特征向量 $\boldsymbol{h}_{(O_{i,j},M_k)}$ 输入到演员网络中，演员网络计算动作空间中所有动作的概率分布：

$$\pi_\theta(a_t|s_t)=Softmax(MLP_\theta(\boldsymbol{h}_{O_{i,j}}^{(L)}\|\boldsymbol{h}_{M_k}^{(L)}\|\boldsymbol{h}_{\mathrm{G}}^{(L)}\|\boldsymbol{h}_{(O_{i,j},M_k)})) \quad (17)$$

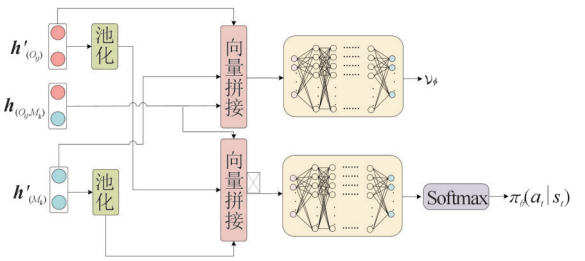


图4　决策网络

Fig.4　Decision network

评论家网络是一种估计器 ν_ϕ，目的是用来预测每个动作的得分，从而为演员网络提供更好的反馈和指导，以改进其策略并获得更高的奖励。

2.4　训练

使用近端策略优化（proximal policy optimization，PPO）[19]算法来训练决策网络，并引入广义优势估计（generalized advantage estimation，GAE）[18]以提高训练过程的稳定性。在参数初始化阶段，采用Xavier方法[20]对神经网络参数进行初始化。动作采样方面，本文使用贪婪策略从当前策略中选取动作，该方法不引入探索性随机行为，而是完全依据当前策略对状态的价值评估做出确定性决策。训练过程的具体算法流程如表1所示。整个训练共执行 I 次迭代，每次迭代期间，智能体处理一个包含MaxStep个实例的训练批次，每完成20次迭代后更新该批次数据。模型参数在每轮迭代中基于所收集样本进行更新。为有效监控训练进展，本文每10次迭代在一组独立验证实例上评估当前策略的性能表现，以确保训练过程的可靠性与策略的泛化能力。

3　实验

3.1　数据集

本文模型在合成数据集上进行训练，并在公开基准数据集上测试其性能。合成数据集的生成遵循文献[18]的方式，每个工件的工序数及其在各机器上的加工时间均设定在区间[1，20]内。公开基准数据集采用7个Mk实例（Mk1～Mk7）[21]，该实例涵盖不同规模的任务场景。实验中所有机器数量设定为10台，各机器的加工功率 $E_{\mathrm{c},k}$ 和待机功率 $E_{\mathrm{u},k}$ 取值均与文献[7]保持一致。

3.2　实验设置

本文所有训练与测试实验均在同一台计算机上完成，该设备硬件配置如下：搭载AMD Ryzen 5 5600型号6核中央处理器（CPU）、NVIDIA Ge-

表1 PPO算法

Tab.1 PPO algorithm

算法1 PPO算法流程
输入：DAN，演员网络和评论家网络以及它们的参数θ和ϕ；
初始化经验池Memory并取样一批MaxStep个EEFJSP实例；
1 for iteration = 1，2，…，I do
2 获取环境的初始状态s_t；
3 for Step = 1，2，…，MaxStep do
4 采样$a_t \sim \pi_\theta(\cdot\|s_t)$；
5 得到奖励r_t并进入下一个状态s_{t+1}；
6 Memory←(s_t, a_t, r_t, s_{t+1})；
7 $s_t \leftarrow s_{t+1}$；
8 end for
9 for Step = 1，2，…，MaxStep do
10 计算每一步的广义优势估计$\widehat{A}_t$；
11 end for
12 for k = 1，2，…，K do
13 计算损失函数；
14 优化并更新参数θ和ϕ
15 end for
16 if iteration mod 20 = 0 then
17 采样新一批MaxStep个EEFJSP实例；
18 end if
19 if iteration mod 10 = 0 then
20 验证策略；
21 end if
22 清空经验池Memory；
23 end for
输出策略$\pi_\theta(a_t\|s_t)$和估计器v_ϕ；

Force RTX 4060型号图形处理器(GPU)，系统环境为Windows10 64位操作系统。本文所提方法基于PyTorch 1.6深度学习框架与Python 3.9编程语言实现。DAN的层数L设置为2，每层的注意力头数H为4。其中，第一层各注意力头的输出维度为32，第二层为8。策略网络与价值网络的结构保持一致，均包含两个隐藏层(隐藏层维度为64，激活函数采用tanh函数)。PPO算法中，策略系数(裁剪系数ε=0.2)、价值系数与熵系数β分别设为1.0、0.5与0.01；GAE参数λ与折扣系数γ分别设为0.98与1。网络训练过程采用Adam优化器，训练周期K设为4，学习率l_r设为3×10^{-4}。

3.3 参数的敏感性分析

为确定最优训练配置，本文在合成数据集(工件数10、机器数5)上对4个关键超参数(学习率、DAN层数、熵系数和裁剪系数)进行了敏感性分析，结果如图5所示。

如图5a所示，当学习率为0.01时，训练过程

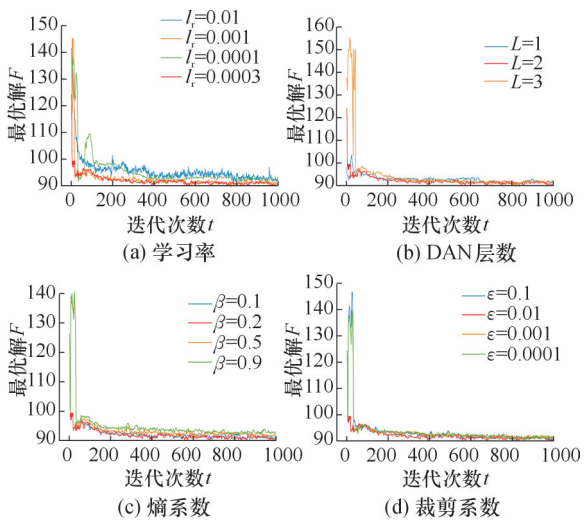


图5 关键超参数敏感性分析

Fig.5 Sensitivity analysis of key hyperparameters

波动显著，收敛性能较差；随着学习率逐步降低至0.0003，结果的收敛性能显著提高，最终收敛至更优解。因此，本文选定0.0003作为最优学习率。图5b～图5d进一步表明，当DAN层数为2、熵系数为0.01、裁剪系数为0.2时，优化结果不仅达到最佳性能，同时也具备良好的收敛稳定性。

除上述参数外，批量大小与折扣因子也对训练过程具有重要影响。较大的批量有助于加速收敛，但可能导致内存不足；较小的批量虽能提高训练策略泛化能力，但易增加训练不稳定性。为兼顾收敛效率与训练稳定性，本文将批量设为20。此外，为完整考虑长期累积奖励，将折扣因子设为1，以确保各时间步奖励的公平性。

3.4 结果分析

本文首先在20×10规模的数据集上训练得到基础模型，随后基于4种不同规模的训练数据开展对比实验，生成Pareto非支配解集，并与4种在实践中表现优异的经典优先级调度规则(priority dispatching rules，PDRs)[22]方法进行对比，包括先进先出(FIFO)、剩余最多工序(MOR)、最短加工时间(SPT)以及剩余工作量最多(MWKR)。

表2所示为本文方法与上述4种PDRs方法在各自解集中完工时间最小时的对比结果。由表2可知，本文方法在4种不同规模实验场景下的性能均显著优于对比方法，在最大完工时间与机器总能耗两大核心评价指标上均表现出明显优势。为进一步直观比较性能差异，图6展示了30×10和40×10两种规模场景下的非支配前沿解分布。由图6可知，随着问题规模扩大，4种PDRs方法的性能均出现不同程度下降，其中机器总能耗指标急剧上升，反映其在大规模问题中适配能力不

足，而本文方法在规模扩大后未出现性能衰退，反而在最大完工时间与机器总能耗上均实现进一步降低。该结果不仅验证了本文方法生成解的质量更高，也表明其在不同问题规模下均能保持稳定优越的优化性能，具备良好的泛化能力。由表2可知，本文方法与4种PDRs方法的运行时间差异最大不超过2.73 s。对于实际EFFJSP，该时间差处于可接受范围内，表明本文方法在解质量与运行效率之间取得了良好平衡，兼具优化性能与实际应用价值。

表2　所提方法与4种PDRs方法在训练集上的结果对比

Tab.2　Comparison of the proposed method and four PDRs on the training dataset

规模		本文方法	FIFO	MOR	SPT	MWKR
15 × 10	目标	[**112.67,631.30**]	[148.72,793.6]	[147.77,1001.41]	[118.43,672.83]	[149.72,936.97]
	时间/s	1.25	0.23	0.24	0.23	0.24
20 × 10	目标	[**124.97,783.20**]	[202.41,1424.44]	[198.45,1423.88]	[136.39,853.8]	[199.43,994.35]
	时间/s	1.70	0.32	0.32	0.32	0.32
30 × 10	目标	[**157.18,964.70**]	[295.62,1667.05]	[296.52,2185.90]	[188.22,1138.12]	[298.02,1704.31]
	时间/s	2.56	0.50	0.50	0.50	0.50
40 × 10	目标	[**202.01,1061.30**]	[398.61,2541.69]	[390.81,2112.67]	[245.19,1413.84]	[396.21,2091.47]
	时间/s	3.44	0.71	0.71	0.71	0.71

注：方括号中第一个数字代表最大完工时间，第二个数字代表机器的总能耗。

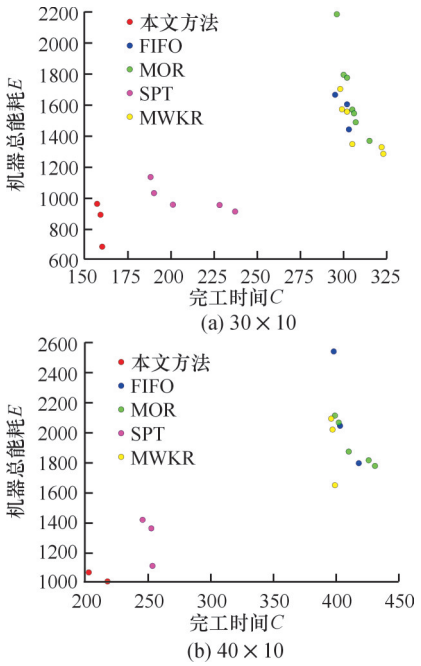


图6　所提方法与4种PDRs方法在不同规模训练集上的非支配前沿解

Fig.6　Non-dominated frontier solutions of the method and four PDRs on the training datasets of different scales

为充分验证本文方法在解决EEFJSP时的优越性与泛化能力，选取7组Mk基准数据集，与两种非支配排序遗传算法（NSGA-Ⅱ和INSGA-Ⅱ）[7]、一种自适应遗传算法（AGA-IS）[23]及一种强化学习算法（Gaussian-based advantage actor-critic，A2C-Gauss）[14]进行对比实验。表3所示为各方法在各自解集中完工时间最小时的对比结果。结果表明，在大多数测试案例中，本文方法在最大完工时间与机器总能耗两个关键指标上均优于所对比的4种算法。

图7进一步比较了本文方法与两种非支配排序遗传算法在解集数量上的差异。由图7可知，本文方法不仅在各指标上表现更优，其生成的Pareto解集还具有更丰富的多样性，反映本文方法在全局收敛与局部搜索能力上均具备更突出的优势。为更细致评估解集质量，图8展示了在Mk2和Mk4数据集上，本文方法与INSGA-Ⅱ所获得的全部非支配解分布情况。由图8可知，本文方法所得解分布更为集中，且整体更接近理想的最优前沿，说明其生成解在质量与收敛性方面均具有明显优势。

为更精准地评估本文方法与强化学习方法A2C-Gauss[14]的性能差异，从两组优化结果中各提取非支配前沿解进行可视化对比，结果如图9所示。结合表3数据和图9中解集分布特征可知：在多数实验场景下，本文方法所得解在质量上具有明显优势。具体表现为在相同机器总能耗下最大完工时间更短，或在相同最大完工时间下机器总能耗更低。该优势主要源于本文所采用的深度强化学习框架能够有效提取工序与机器的关键特征，过滤冗余调度信息，避免其对决策产生干扰；同时模型充分考虑了机器间竞争关系对后续决策的影响，从而使智能体能够输出更符合多目标优化目标的决策。在少数案例中，本文方法与A2C-Gauss的性能差异较小，这一现象可能源于本文方法在整个优化过程中始终以最大完工时间与机器总能耗双目标协同优化为导向，确保两大指标处于全局平衡状态，但在算法探索过程中难

表3 所提方法与其他先进算法在公开基准数据集上的结果对比

Tab.3 Comparison of the method and other advanced algorithms on the public benchmark datasets

规模		本文方法	NSGA-Ⅱ[7]	INSGA-Ⅱ[7]	AGA-IS[23]	A2C-Gauss[14]
Mk1	目标	[43.0,243.20]	[45,279.0]	[46,276.4]	—	[42,480]
	时间/s	0.45	—	37	—	0.79
Mk2	目标	[32.0,177.70]	[42,378.5]	[39,366.4]	[34,356.3]	[31,348]
	时间/s	0.48	—	32	—	0.82
Mk3	目标	[216.0,1453.80]	[264,3312.7]	[257,3310.2]	[205,2566.1]	[206,2763]
	时间/s	1.34	—	164	—	2.12
Mk4	目标	[68.0,415.80]	[91,1004.8]	[86,1003.1]	[85,901.3]	[70,878]
	时间/s	0.76	—	65	—	0.77
Mk5	目标	[186.0,1393.10]	[193,1260.2]	[192,1256.5]	[185,1295.2]	[182,1243]
	时间/s	0.89	—	110	—	0.94
Mk6	目标	[81.0,792.20]	[122,2103.3]	[117,1960.6]	[87,1344.5]	[73,1256]
	时间/s	1.31	—	76	—	2.45
Mk7	目标	[152.0,1403.90]	[193,1844.1]	[188,1832.0]	[156,1636.5]	[155,1538]
	时间/s	0.88	—	129	—	0.90

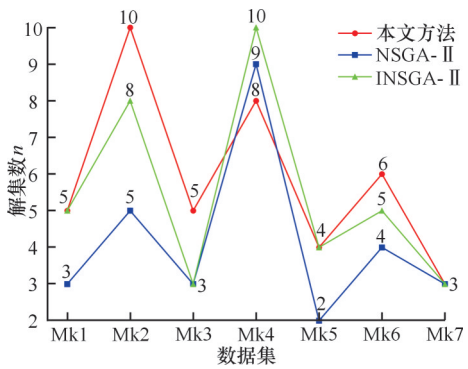


图7 不同方法解集数对比

Fig.7 Comparison of the number of solution sets of the different methods

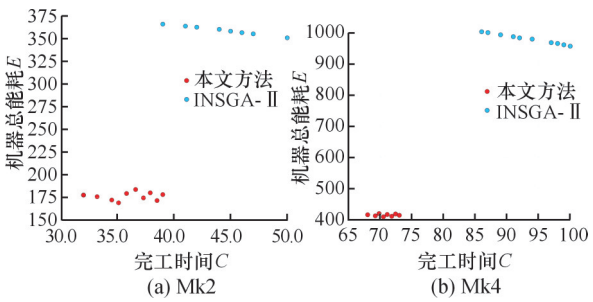


图8 不同方法在Mk2和Mk4测试数据集的所有非支配解散点图

Fig.8 Scatter plot of all non-dominated solutions of the different methods on the test datasets of Mk2 and Mk4

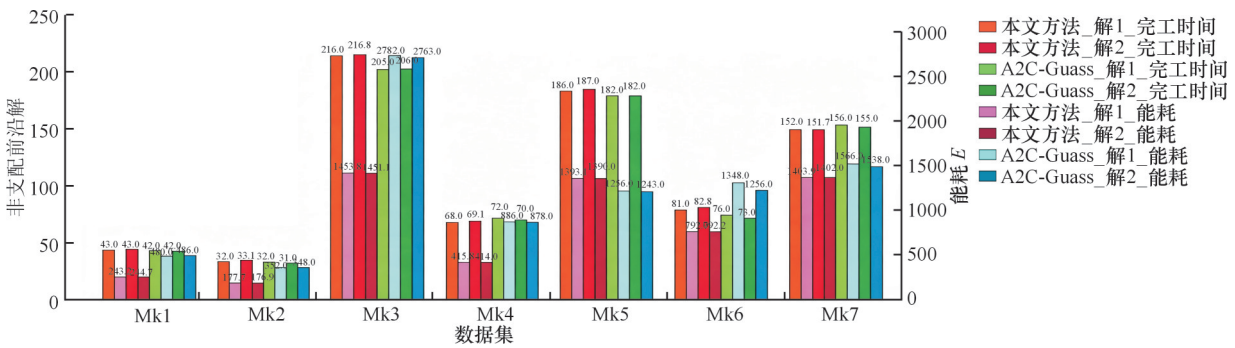


图9 两种方法在测试集中的两组非支配前沿解对比

Fig.9 Comparison of two non-dominated frontier solutions by two methods on the test dataset

免出现目标间协调不足或局部优化偏向，导致在个别场景下性能略逊。由表3算法运行时间对比结果可知，本文方法具有更快的收敛速度，能更早逼近高质量解，表明其在解质量与运行效率之间取得了良好平衡，兼具优越的优化性能与实际应用可行性。

图10为本文方法在Mk1数据集上获得的最小化最大完工时间甘特图，纵轴表示加工机器编号，横轴表示最大完工时间，标注$O_{9,1}$代表第9个工件的第1道工序，其余工序标识依此类推。分析该调度方案可知，本文方法能够合理协调工序选择与机器分配。

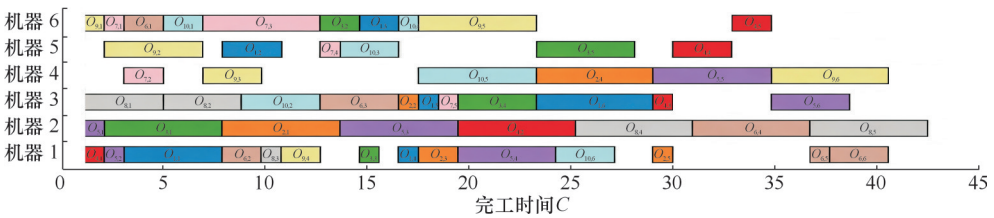


图 10　本文方法在 Mk1 数据集上获得的最小化最大完工时间甘特图

Fig.10　Gantt chart of minimizing maximum makespan obtained by the method on the mk1 dataset

4　结语

针对 EEFJSP，本文构建了以最大完工时间和机器总能耗为优化目标的多目标调度模型，并提出基于双注意力网络的 DRL 求解算法。通过将 EEFJSP 建模为 MDP，并设计工序消息 Agent 与机器消息 Agent 协同工作的双注意力网络架构，实现了对工序间优先级约束和机器间竞争关系的深度特征提取与关联建模。在此基础上，将工序选择与机器分配整合为联合决策问题，通过输出工序-机器对的优先级概率分布，实现了完工时间与机器能耗的双目标协同优化。在合成数据与多个公共基准数据集上的实验表明，相较于经典优先级调度规则、改进遗传算法及现有强化学习方法，本文方法在解质量、收敛速度与泛化能力方面均表现出明显优势。然而，本文方法尚未考虑机器故障、紧急订单插入等实际生产中的动态扰动因素。未来将重点探索该方法在动态调度环境下的扩展应用，研究基于深度强化学习的快速响应与重调度策略，以期进一步提升其在复杂工业场景中的实用价值。

参考文献：

[1] XIE J, GAO L, PENG K, et al. Review on Flexible Job Shop Scheduling[J]. IET Collaborative Intelligent Manufacturing, 2019, 1(3): 67-77.

[2] 李开心，尹瑞雪，周鹏. 面向能耗的柔性作业车间调度研究方法综述[J]. 制造业自动化，2024，46(9)：150-158.
LI Kaixin, YIN Ruixue, ZHOU Peng. Review of Research Methods for Flexible Job-shop Scheduling Based on Energy Consumption [J]. Manufacturing Automation, 2024, 46(9): 150-158.

[3] 李敬花，闫恒山，杨博歆，等. 改进遗传—和声搜索算法求解海工装备制造车间调度问题[J]. 计算机集成制造系统，2022，28(12)：3921-3934.
LI Jinghua, YAN Hengshan, YANG Boxin, et al. Improved Genetic-harmony Search Algorithm for Solving Workshop Scheduling Problem of Marine Equipment[J]. Computer Integrated Manufacturing Systems, 2022, 28(12): 3921-3934.

[4] 李聪波，雷焱绯，肖溱鸽，等. 面向广义能耗的柔性作业车间调度优化模型[J]. 计算机集成制造系统，2018，24(12)：3050-3059.
LI Congbo, LEI Yanfei, XIAO Qinge, et al. Flexible Job Shop Scheduling Optimization Model for Generalized Energy [J]. Computer Integrated Manufacturing Systems, 2018, 24(12): 3050-3059.

[5] 姜一啸，吉卫喜，何鑫，等. 基于改进非支配排序遗传算法的多目标柔性作业车间低碳调度[J]. 中国机械工程，2022，33(21)：2564-2577.
JIANG Yixiao, JI Weixi, HE Xin, et al. Low-carbon Scheduling of Multi-objective Flexible Job-shop Based on Improved NSGA-Ⅱ[J]. China Mechanical Engineering, 2022, 33(21): 2564-2577.

[6] 栾飞，蔡宗琰，吴书强，等. 求解低碳车间调度问题的改进鲸鱼算法[J]. 机械科学与技术，2020，39(5)：721-728.
LUAN Fei, CAI Zongyan, WU Shuqiang, et al. Improved Whale Optimization Algorithm of Scheduling Problem for Low Carbon Workshop[J]. Mechanical Science and Technology for Aerospace Engineering, 2020, 39(5): 721-728.

[7] 徐明，张剑铭，陈松航，等. 柔性作业车间调度问题的多目标优化算法[J]. 计算机与现代化，2021(12)：1-6.
XU Ming, ZHANG Jianming, CHEN Songhang, et al. Multi-objective Optimization Algorithm for Flexible Job Shop Scheduling Problem[J]. Computer and Modernization, 2021(12): 1-6.

[8] WANG L, PAN Z, WANG J. A Review of Reinforcement Learning Based Intelligent Optimization for Manufacturing Scheduling [J]. Complex System Modeling and Simulation, 2021, 1(4): 257-270.

[9] 郑国梁，张朝阳，吉卫喜，等. 基于 DQN 的改进 NSGA-Ⅱ求解多目标柔性作业车间调度问题[J]. 现代制造工程，2025(9)：1-11.
ZHENG Guoliang, ZHANG Chaoyang, JI Weixi, et al. Solving Multi-objective Flexible Job Shop Scheduling Problems with an Improved NSGA-Ⅱ Based on DQN[J]. Modern Manufacturing Engineering, 2025(9): 1-11.

[10] 张国辉，李志霄，张利平，等. 基于强化学习协同进化算法求解柔性作业车间节能调度问题[J]. 计算机应用研究，2025，42(7)：2039-2047.
ZHANG Guohui, LI Zhixiao, ZHANG Liping, et al. Reinforcement Learning Based Co-evolutionary Algorithm for Solving Flexible Job Shop Energy Efficient Scheduling Problem[J]. Application Research of Computers, 2025, 42(7): 2039-2047.

[11] CHENG W, ZHANG C, MENG L, et al. Deep Reinforcement Learning for Solving Efficient and Energy-saving Flexible Job Shop Scheduling Problem with Multi-AGV [J]. Computers & Operations Research, 2025, 181: 107087.

[12] 张国辉，伦伟航，李亮，等. 基于混合学习策略进化算法的柔性作业车间节能分批调度研究[J]. 控制与决策，2025，40(12)：3655-3666.
ZHANG Guohui, LUN Weihang, LI Liang, et al. A Hybrid Learning Strategy Evolutionary Algorithm for Energy-efficient Flexible Job Shop Batch Scheduling Problem[J]. Control and Decision, 2025, 40(12): 3655-3666.

（下转第 2565 页）