

基于蒙特卡洛树搜索的自组织制造车间任务调度优化

王立平, 赵振, 刘长春, 唐敦兵

(南京航空航天大学机电学院, 江苏 南京 210016)

摘要: 自组织制造模式结合现有的工业互联技术、无线网络技术、分布式计算技术和人工智能技术并加以应用, 将传统制造资源封装为具有高自主性、高适应性和高功能性的制造单元, 并通过与其他制造单元的交互, 完成制造任务的自组织协商分配。在这个过程中, 为了构建制造任务智能分配和自组织资源配置的快速响应机制, 实现制造任务与制造资源的高效、动态匹配, 将多智能体合同网协议与蒙特卡洛树搜索算法相结合, 提出了制造车间控制系统自组织运作机制与制造任务调度优化方法。最后, 通过离散车间任务分配案例对所提方法的实际可行性进行了验证。实验结果证明, 该方法更易实现“单步协商、全域寻优”这一目标。

关键词: 多智能体系统; 自组织运作; 蒙特卡洛树搜索; 制造任务分配; 智能制造

中图分类号: TH166; TN05

文献标志码: A

doi: 10.11959/j.issn.2096-3750.2025.00468

Optimization of task scheduling in self organizing manufacturing workshop based on Monte Carlo tree search

WANG Liping, ZHAO Zhen, LIU Changchun, TANG Dunbing

College of Mechanical & Electrical Engineering of Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

Abstract: The self organizing manufacturing model was combined with the existing industrial interconnection technology, wireless network technology, distributed computing technology, and artificial intelligence technology, which can encapsulate traditional manufacturing resources into manufacturing units with high autonomy, adaptability, and functionality. Through interaction with other manufacturing units, self-organizing negotiation and allocation of manufacturing tasks were completed. To develop a rapid response mechanism for the intelligent allocation of manufacturing tasks and self-organizing resource distribution, the multi-agent contract net protocol (CNP) was integrated with the Monte Carlo tree search (MCTS) algorithm. Based on this, dynamic matching between manufacturing tasks and resources was achieved. Apart from this, a self-organizing operational mechanism and an optimization method were accomplished for the scheduling of manufacturing workshop control systems. Finally, the practical feasibility of the proposed method was verified through a discrete workshop task allocation case. The experimental results indicate that the method can better achieve the goal of “one-step negotiation and global optimization”.

Key words: multi-agent system, self organizing operation, Monte Carlo tree search, manufacturing task allocation, intelligent manufacturing

收稿日期: 2024-11-28; 修回日期: 2025-01-23

通信作者: 唐敦兵, d.tang@nuaa.edu.cn

基金项目: 国家自然科学基金重大研究计划项目 (No. 92267109); 国防基础科学研究计划项目 (No. JCKY2023203C026); 国家资助博士后研究人员计划B档项目 (No. GZB20240972); 中国博士后科学基金资助项目 (No. 2024M754122); 江苏省自然科学基金资助项目 (No. BK20241389); 江苏省卓越博士后计划项目 (No. 2024ZB194)

Foundation Items: The Major Research Plan of the National Natural Science Foundation of China (No. 92267109), The National Defense Basic Scientific Research Program of China (No. JCKY2023203C026), The National Postdoctoral Researcher Support Program of China (No. GZB20240972), The China Postdoctoral Science Foundation (No. 2024M754122), The Natural Science Foundation of Jiangsu Province (No. BK20241389), The Outstanding Postdoctoral Program of Jiangsu Province (No. 2024ZB194)

0 引言

近年来,随着工业化和信息化的不断融合,尤其是后疫情时代给全球经济带来的巨大影响,我国制造业正朝着生产规模扩大化、产品订单定制化、决策需求动态化的方向发展。在此背景下,具备自感知、自学习、自决策和自适应能力的“智能制造系统”成为制造业的未来趋势。智能制造的目的是将新一代的传感技术、信息技术、物联技术、人工智能技术与制造技术相结合,进而实现单台设备、单个产线、直至整个工厂等制造服务提供者的自主智能运行。在这个过程中,具备状态感知、数据分析、协商决策和精准控制等特征的多智能体系统(MAS, multi-agent system)逐渐成为主流的研究方向。

然而,尽管MAS在学术界已经成为一种公认的智能制造模式,但在实际使能化过程中,由于缺乏公认统一的体系架构和与之对应的生产过程动态优化方法,MAS在传统离散车间应用中缺乏应对各种不确定因素的灵活性和自适应性^[1]。基于当前的应用基础和实证研究,MAS使能化过程中存在的关键问题可以概括为:现有生产模式面临着生产环境复杂、生产需求多样、生产条件动态和信息不对称等问题,再加上设备间无法通过自主协作完成制造任务的分配,进而导致制造系统组织效率低下、生产柔性不足。针对上述问题,学术界在制造系统建模与制造单元协商等方面做出了大量的探索。

首先,在制造系统建模方面,基于Holon和MAS技术的制造单元封装方式是最常见的两种建模方法。王世进^[2]提出了基于Holon概念模型的、集成强化学习机制的合同网协议(CNP, contract net protocol)和改进过滤定向搜索算法,实现了基于自治与协商机制的柔性作业车间智能调度。Wang等^[3]基于射频识别(RFID, radio frequency identification)与Holon开发了敏捷制造计划与控制系统,以响应动态变化的制造活动与异常。Pujo等^[4]提出了一种无线Holon网络,通过Holon为信息物理系统中的实体进行建模,并通过Holon间合作产生群体智能。Reaidy等^[5]基于“协同竞争=协同+竞争”的新范式提出了一种基于MAS的建模方法,以提高分布式生产系统的按需生产和反应能力。

Barenji等^[6]提出了一种用智能体来部署基于RFID的分布式控制与监控系统体系架构,通过智能体间的相互协作提升系统的可重构性。程涛等^[7]提出了一种基于MAS的数控实体网络化和智能化封装方法,为国内多智能体制造系统的研究奠定了基础。Rodrigues等^[8]提出了一种基于MAS的制造系统服务重构方法,并利用智能体动态识别实现了制造系统重构。Tehrani等^[9]提出了基于智能体的动态工艺规划制造系统体系结构,将启发式搜索算法与协商协议相结合来处理动态干扰。

其次,在制造单元交互协商方面,基于博弈论和合同网协议的交互方式是最常见的两种协商机制。Wang等^[10]提出了一个智能工厂框架,在基于MAS的基础上增加了协调智能体,通过大数据反馈提高生产过程的柔性。Wang等^[11]提出了一种基于MAS的实时调度架构,采用博弈论协调各智能体行为,并根据设备实时状态进行最优分配。Kim等^[12]提出了一种融合MAS与强化学习的智能制造系统,通过智能体决策与交互,提升动态环境下的竞争力。Ma等^[13]通过基于议价博弈的协商机制,研究了基于智能体的物联网制造业柔性作业车间的实时调度问题,实现了任务的最优分配。Guo等^[14]从减少协商沟通、提高多智能体协调能力、改进任务资格评估策略3个方面改进了合同网协议(CNP)。Zhu等^[15]使用强化学习来执行任务分配过程,并改进了CNP实现通信负载均衡。Yeung等^[16]通过受众限制策略提出了一种改进CNP,减少了智能体协商过程的消息拥塞。

从上述研究可以看出,在制造系统建模方面,虽然现有的Holon或者MAS等封装技术可以将RFID、个体自治、群体协商等分布式人工智能应用于传统制造车间,但是对于采用何种方式将车间层的物理设备(机床、机器人等)封装为具有相互通信、自我决策功能的智能体这方面的使能技术却鲜有人深入探究,亟须进行完善。在制造单元交互协商方面,现有研究中以博弈论、合同网为代表的两种协商机制,仍然无法实现策略探索与现有经验知识利用的有效平衡,如何构建设备状态自评估、加工过程自适应、生产管理自组织的制造系统多智能体运行模型,通过MAS间的协作与竞争来实现更高效的资源分配,成为当前亟须解决的另一关键问题。

1 多智能体制造系统

1.1 多智能体制造系统建模

智能体的标准化范式于20世纪被提出，以应对未来工厂的挑战。智能体被定义为具有自主性和协作性的个体，能够做出决策并与其他智能体的交互，完成系统性的目标^[17]。在制造系统领域，多品种、小批量、定制化的制造任务、智能决策和分布式制造结构推动了多智能体制造系统(MAMS, multi-agent manufacturing system)的研究。智能体作为设备级执行和控制单元也被认为是分布式制造系统建模的理想方法，用来实现制造资源的建模与封装^[18]。通过智能体封装技术，可以实现制造设备的自感知、自决策与自学习，进而传统制造系统被建模为MAMS。因此，基于MAS的自组织智能制造系统也是一个反应式的自适应系统，该系统可以动态修改自身各层级行为和结构，持续不断地进行从输入到输出的交互，以应对生产现场的动态性和不确定性因素的持续增长。

通过上述分析，基于MAS的智能制造系统架构如图1所示，依据制造系统的角色划分，MAMS最少需要具备机器智能体、工件智能体、调度智能体和监控智能体4种角色。其中，工件智能体根据工件状态决定是否发起加工协作请求，并请求调度智能体选择合适的机器智能体进行加工；工件智能体用数据标签取代了传统的物联网跟踪方法（如RFID、二维码等），而监控智能体为调度智能体的决策提供实时信息支持。

(1) 机器智能体

传统制造车间多采用分层控制系统，这导致制造设备间无法相互连接。为了实现车间制造资源间的互联，通过MAS技术，将制造资源封装为具备感知、计算、通信和控制等功能的机器智能体。同时，基于MAS的设计理念与制造系统自组织运作机制，每个机器智能体均具备一定的决策能力，它们首先要与对应设备建立连接，然后通过与其他机器智能体的信息交互完成制造任务协商，最终实现对设备的精准控制。因此，机器智能体是物理实体

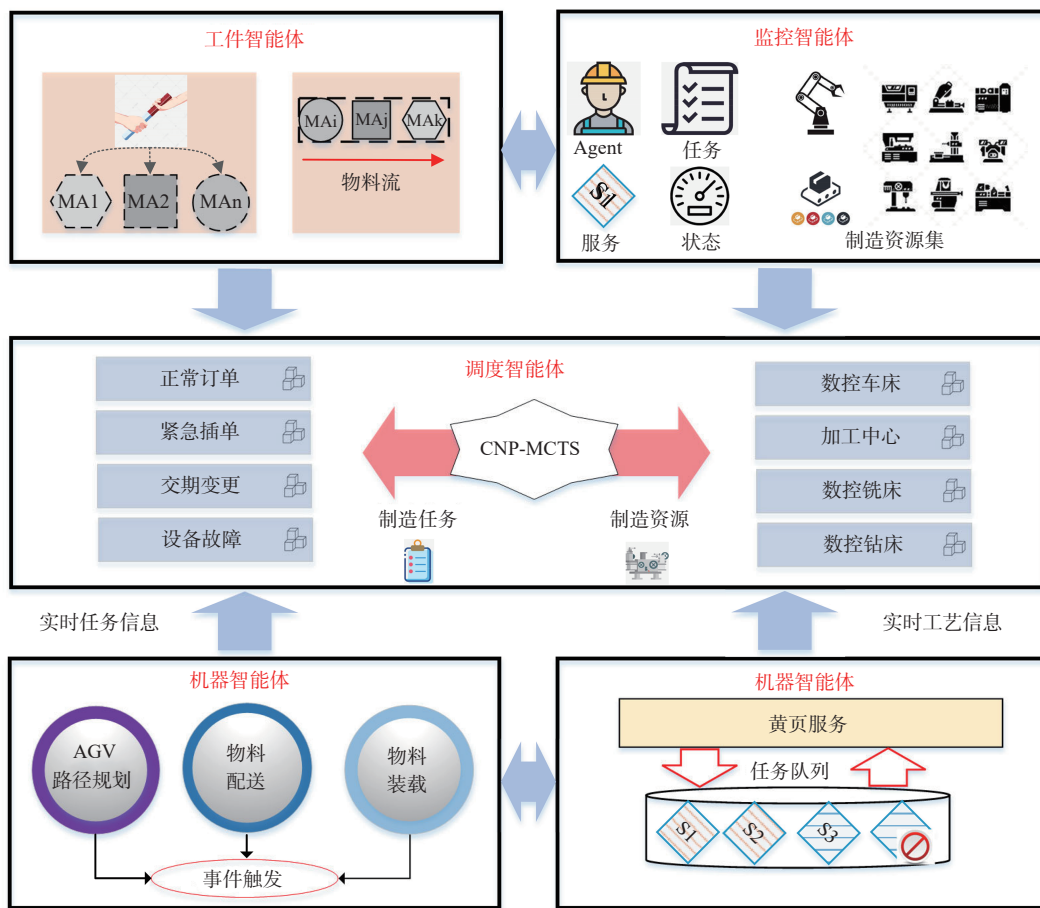


图1 基于MAS的智能制造系统架构

在信息空间的映射, 对应其感知、计算、通信、控制 4 个基本属性, 机器智能体也是由这 4 个主要模块组成, 并分模块进行实现的。

机器智能体主要包括以下两个功能: 首先, 通过统一的通信协议组件实现分布式制造资源的互联; 其次, 通过对来自物理设备的实时信息进行分析处理, 为动态调度和异常处理提供实时信息支持。在这个过程中, 由于机器智能体的计算与通信压力过大, 单个线程往往无法满足所有需求, 机器智能体需要采用多线程模型进行开发、部署。这些线程分别负责制造单元控制、实时状态监控、智能体之间的交互和协商等功能。

(2) 工件智能体

在众多关于 MAS 的研究中, 研究者在深入研究之前通常会假定车间中每个工件创建对应的工件智能体。当工件数量较少时, 这种设定无论从运行逻辑还是系统开发角度都易于实现。然而, 随着工件数量的增加, 容易造成工件智能体间的通信负荷急剧升高, 进而影响系统的正常运行。

虽然工件智能体数量的增加对系统性能有很大影响, 但如果没有独立的控制单元对制造任务进行管理也是不合理的。在此背景下, 本文采用单一工件智能体模式下工件接力协作加工的控制方法对制造任务和生产过程进行统一管理。该控制方法将工件视为一根在制造资源之间连续改变状态的“接力棒”, 这根接力棒记录着完整的生产过程、消耗材料和加工细节等信息, 并由工件智能体进行跟踪。

(3) 调度智能体

调度智能体是智能体协商的发起者与评判者。首先, 由工件智能体向调度智能体发起当前加工任务的招标请求; 然后, 各机器智能体根据自身状态决定是否投标并生成标书; 最后, 调度智能体通过标书评估完成可行投标集的收集与筛选。

调度智能体的决策算法设计是影响制造车间自组织运作效率的核心。当调度智能体对工序 O_j 接收到机器智能体的标书后, 可以根据不同的决策目标进行计算和筛选, 常用决策目标有: 最小完工时间、最小加工成本、负载均衡等。

(4) 监控智能体

在 MAMS 中, 由于单个智能体视野的局限性, 很难独立完成复杂的决策任务, 需要多个智能体协作完成制造任务的优化。然而, 每个智能体对全局

信息的利用率较低, 进而造成协商结果的不确定。针对这个问题, 本文的解决方案是设置一个面向全局制造信息的监控智能体, 该智能体既能提供来自所有制造资源的属性信息, 又能提供全局制造任务信息。

在实际生产过程中, 参与投标的单个机器智能体由于观测视野受限容易做出不合理的决策。而监控智能体的存在保证了整个系统朝着全局优化的方向发展, 调度智能体决策时只需要与监控智能体订阅所需信息即可实现全局决策优化。

1.2 智能体间的协作过程

MAMS 由多个独立、分散的智能体组成, 并通过高效的协商机制相互协作, 最终实现实时动态调度^[9]。这些智能体在物理上是分布式的, 并且在逻辑上是相互隔离的。同时, 在单一智能体行为高度自治的基础上, 如何通过特定的协商机制来组织现有智能体进而完成制造任务分配, 是一个亟须攻克的难题。

MAMS 中多智能体协作过程如图 2 所示, 当制造任务下达后, 智能体之间通过交互协商的方式实现制造任务的分配。然而, 单一智能体收集信息的局限性往往导致决策不够准确, 进而难以获得最优调度结果。在制造资源有限的情况下, MAMS 中的功能智能体 (机器智能体、工件智能体) 和服务智能体 (调度智能体、监控智能体) 协同工作, 实现动态调度。

此外, 当制造现场发生扰动事件时, 智能体能够准确识别扰动类型, 并通过协作的方式进行动态处理。MAMS 还会采用分布式控制策略对协商结果进行验证, 从而保证了制造系统自组织运作结果在任务分配生产过程中朝着优化的方向推进。

基于 MAS 在个体自治、群体协商等方面的特性, 与制造系统融合的 MAMS 也为生产计划提供了越来越多更为广泛的功能。然而, 每个自治制造单元的决策问题本质上都是由相应智能体的局部决策驱动的, 目的是优化自身目标函数, 因此很少在全局层面取得最优, 甚至只能产生一个局部较优解^[20-23]。在此背景下, 本文在传统多智能体合同网协商机制的基础上, 引入了自博弈搜索方法中的蒙特卡罗树搜索 (MCTS, Monte Carlo tree search) 算法, 在调度智能体决策时, 通过 MCTS 算法进行全域寻优, 从而保证动态调度结果的优化。

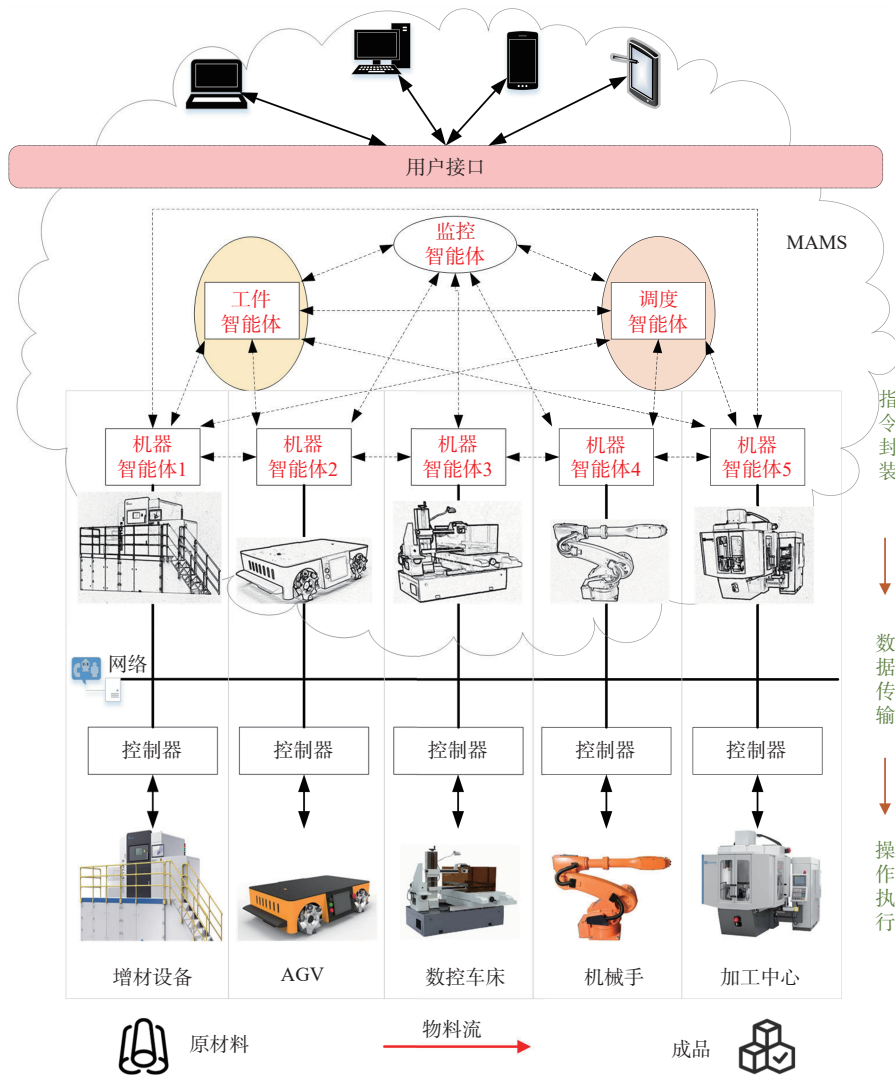


图2 MAMS中多智能体协作过程

2 集成CNP-MCTS的制造车间控制系统自组织运作机制

2.1 CNP-MCTS协商机制模型

单独的CNP协商机制具有单步寻优、无法保障全局优化能力的短板，通过引入MCTS自博弈算法，可以在有限的计算开销内为当前制造任务找到最优加工路线，从而确保订单完工时间、设备负载均衡等目标的优化。本文提出的集成合同网协议与蒙特卡洛树搜索（CNP-MCTS）算法的MAMS协商分配机制，以适应性行为方法为指导，通过强化学习（RL, reinforcement learning）多智能体博弈过程，利用单个智能体强化学习与环境的相互作用，实现智能制造系统内制造任务的实时动态调度。

同时，在智能体决策过程中，利用一系列规则保障了MCTS算法搜索结果精度。CNP-MCTS协商

分配机制基本流程如图3所示，在任务分配过程中，主要协商过程发生在调度智能体与机器智能体之间。智能体间的交互依托CNP，调度智能体决策模块内嵌MCTS算法，可以避免基本合同网协议决策时的单步最优，进而实现“单步协商，全域寻优”这一目标。

在MCTS算法自博弈搜索过程中，首先需要定义搜索树的根节点和叶节点，根节点定义为搜索开始的节点，叶节点定义为没有子节点的节点。在基于CNP-MCTS的自组织协商分配机制中，自博弈树的每个节点代表当前车间环境的生产状态，包含当前状态下价值估计、当前节点被选择次数、当前节点的父节点和子节点。

将上述自博弈树的节点与状态对应到实际制造任务协调场景中，根节点为发起任务招标的调度智

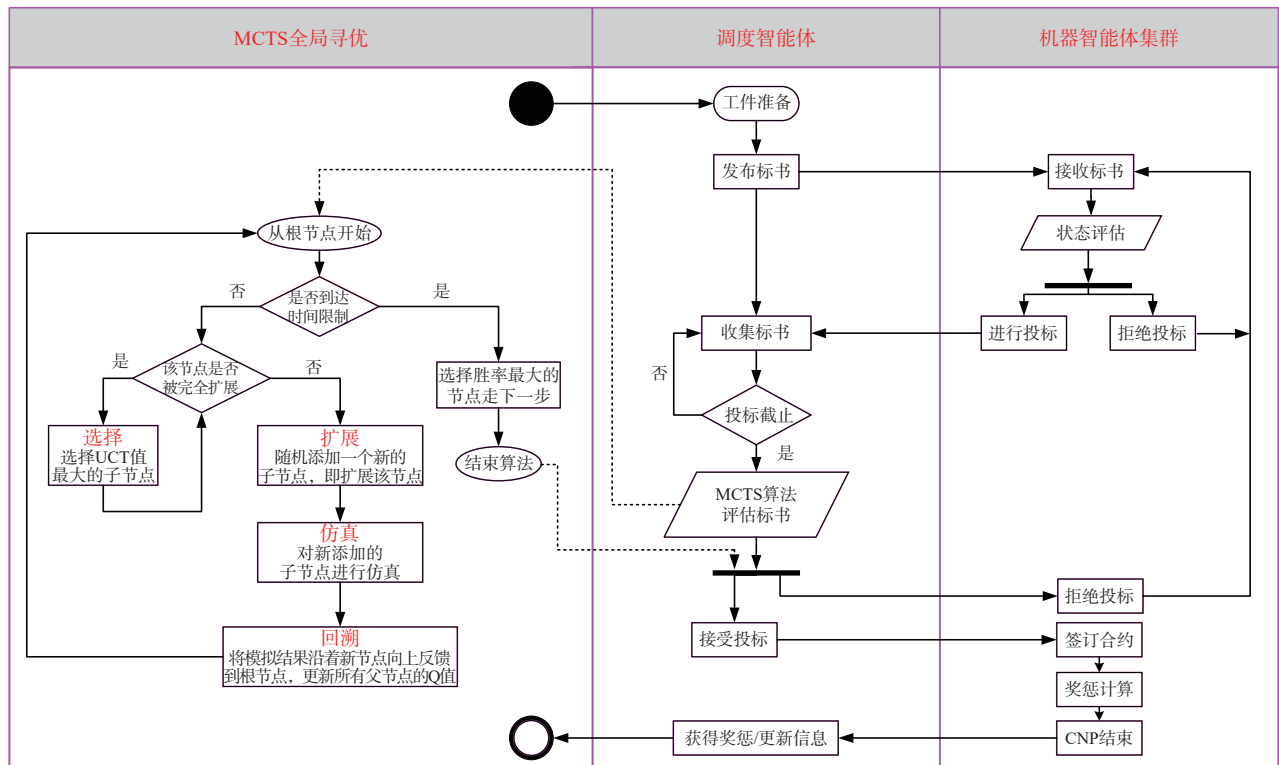


图3 CNP-MCTS协商分配机制基本流程

能体，下面的每个子节点代表具备对应工序加工能力的制造资源，当前状态下价值估计为根据机器智能体投标的标书加权求和得到的投标价格，当前节点被选择次数为经过若干次搜索该制造资源被选择次数，当前节点的父节点和子节点分别对应当前工序的紧前工序和紧后工序。

CNP-MCTS机制中MCTS算法的4个阶段如图4

所示，CNP-MCTS机制中MCTS算法搜索过程主要分为4步：选择、扩展、仿真和回溯。反复执行这4步操作，直至探索完成或者达到搜索停止条件，再进行最终决策，下面逐一介绍这些步骤的意义和算法的具体细节。

1) 选择。从根节点开始，根据评估函数递归地选择最佳子节点，直至到达可扩展的下一个子节点

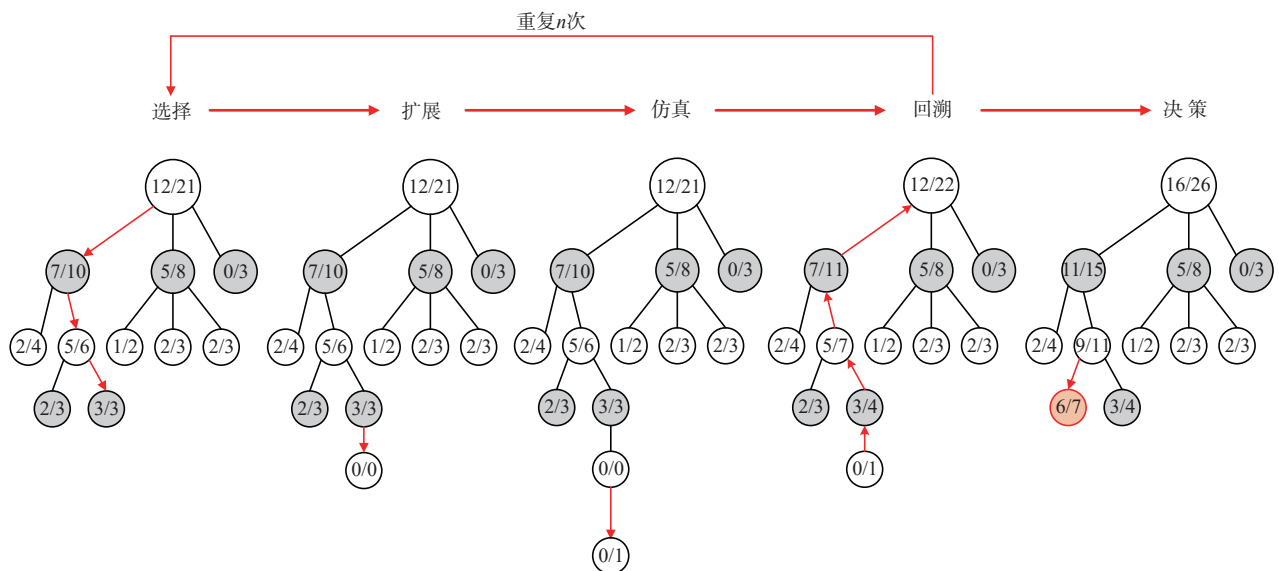


图4 CNP-MCTS机制中MCTS算法的4个阶段

点。这个可扩展的子节点包含一个或多个未访问的子节点，因此它是非终端状态。选择阶段的具体搜索规则采用博弈树上限置信区间（UCT, upper confidence bound apply to tree）算法，根据该算法为每个子节点计算一个分数，然后不断遍历分数最高的子节点，直至到达叶节点，详细计算过程描述如下

$$\text{UCT}(v_i, v) = \frac{Q(v_i)}{N(v_i)} + c \sqrt{\frac{\log(N(v))}{N(v_i)}} \quad (1)$$

其中，第一项 $\frac{Q(v_i)}{N(v_i)}$ 为子节点 v_i 的价值估计，该价值越大，表示探索这个节点越有可能获得更大奖励，其计算式定义为：平均收益=总收益/总次数。另外，第一部分的取值与 CNP 的投标值正相关，在传统 CNP 决策中，通常只用第一项作为决策目标，这也间接导致决策单步最优，无法保证全局性能。为了避免这种贪婪寻优方式的搜索错过最优解，MCTS 算法引入了第二项 $c \sqrt{\frac{\log(N(v))}{N(v_i)}}$ ，这一部分更倾向于那些未被探索过的节点（即 $N(v_i)$ 较小的节点）， c 为系数。

综上， $\text{UCT}(v_i, v)$ 的第一项和第二项分别代表智能体的利用和探索部分，系数 c 控制这两部分之间的平衡，通过调节 c 的大小，可以实现 MCTS 偏向于利用更好的节点和探索尽可能多的节点。特别情况下，当 $c=0$ 时，智能体变为单步寻优，失去全局探索能力。

2) 扩展。在遍历迭代过程中选择一个子节点后，需要根据正向工序关系有向图查询该子节点是否可以展开并继续添加子节点，然后根据可用操作选择所有未访问的子节点，除非在当前节点上无法执行更多操作。

在 MCTS 算法中，子节点代表的是当前自博弈树的一个状态，这个状态下的所有可行步骤都是由强化学习的环境决定的。也就是说，如果当前子节点不是最终节点，可以根据当前的强化学习环境（制造任务工艺进度和制造资源任务队列）获取当前匹配制造资源集合的所有可能动作集合，然后根据这些状态扩展子节点，当新添加的子节点成为叶节点后，当前自博弈树状态更新描述如下

$$s_{t+1} \sim S(s_t, a_t) \quad (2)$$

$$Q(s, a) \leftarrow Q(s, a) + \alpha \left[r + \gamma \max_{a'} Q(s', a') - Q(s, a) \right] \quad (3)$$

其中，在选择过程中，选中的动作（选择制造资源）记为 a_t ，它只是一个假想动作，是在 MCTS 算法搜索过程中“仿真”执行的。同时， s_t 为动作 a_t 执行前的环境状态， s_{t+1} 为更新后的环境状态。

在 MCTS 算法实际搜索过程中，由于刚开始时根节点的所有子节点均未被访问过，可以从子节点中随机选择一个节点进行第一次仿真。

3) 仿真。仿真过程中根据 Rollout 策略的评估函数选择一条随机下行路径以到达叶节点，然后计算该条加工路径下的完工时间与成本。按照博弈论观点，Rollout 策略最适合选出一条未必多目标最优但加工路径最短的策略。因为如果这个策略走得慢，全局优化能力会更强，结果也更准确，但由于耗时多了，在单位时间内模拟次数就少了，间接导致全局优化能力变弱。

所以，MCTS 算法也是一种基于模型的 RL，由两部分组成：一个是状态转换模型，预测 RL 做出动作后的下一个状态；另一个是奖励模型，预测相应状态转换后的预期计算奖励。一旦马尔可夫决策模型构建完成，便可通过 MCTS 仿真计算最优路线或最优值。

4) 回溯。把仿真结果加到该节点的所有父节点上，在一次搜索的选择-扩展-仿真阶段，得到从根节点到叶节点的路径，当路径的叶节点更新以后，需要考虑叶节点的扩展对父节点的影响，这里就要从叶节点出发，回溯对应的决策路径并更新对应的价值函数描述如下

$$Q(s, a) = \frac{1}{N(s, a)} \sum_{s' | s, a \rightarrow s'} V(s') \quad (4)$$

其中， $V(s')$ 为最终到达的叶节点价值函数值，在更新价值函数的同时，也需要更新回传路径上每个节点的访问次数。

在跟踪仿真结果后，从仿真开始的节点到根节点的路径上更新所有树节点的统计信息，MCTS 算法常用的两种回溯统计是每个节点的仿真奖励和每个节点的访问次数。

5) 决策。在 MCTS 算法中，“选择→扩展→仿真→回溯”这个过程要重复执行很多次，常用终止条件有两个：达到一定迭代次数和达到规定搜索时

间,当搜索程序结束时,最佳制造资源对应访问次数最多的节点,也是UCT最大的节点。该节点访问次数越高,代表当前节点价值越高,也越应该被选择。

MCTS算法在许多规划问题中具有出色的性能表现,同时,随着问题规模的增大会出现巨大的状态空间,使得MCTS在在线任务规划中需要花费巨大的计算开销。然而,在CNP-MCTS任务规划背景下,参与每道工序投标的智能体数量相对较少,因此能够在有限时间内达到较好的任务动态调度性能。

2.2 CNP-MCTS机制中调度智能体决策过程

基于CNP-MCTS机制的制造任务协商分配包含CNP协商与MCTS决策两个过程:首先由调度智能体根据CNP规则发起首轮加工招标,等待满足当前工序加工能力的制造资源进行投标;然后调度智能体根据MCTS算法通过对当前任务加工路径进行全局搜索,进而确定当前投标集合中最适合的制造资源。

在调度智能体进行下一工序的全局搜索过程中,各层级制造资源需要返回包括该任务的开始加工时间、预计完工时间和加工成本等信息,而计算开始加工时间最常用的策略有3个:先进先出(FIFO, first input first output)、最短作业优先(SJF, shortest job first)和最长作业优先(LJF, longest job first)^[24]。

根据CNP机制,在调度智能体选择合适的制造资源时,容易造成基于当前状态生成过多劣质解,进而影响最大完工时间等优化目标^[25]。本文针对遗传算法等贪婪竞价和“短视”的缺点,提出了一种新的初始解生成策略——在调度智能体决策过程中,将全局信息与制造资源负载均衡相结合,选择合适的制造资源。该过程详细描述如下

$$M_k = f_{\min} \quad (5)$$

$$L_{\min} = L_{\min} + t_{ik} \quad (6)$$

其中,假设当前最小加工负载设备为 M_k ,调度时要求其负载值为 L_{\min} ,同时,假设在设备 M_k 上加工 O_i 的处理时间为 t_{ik} ,当满足以下条件时,上述步骤生成的初始解是负载平衡的

$$|L_a - L_b| < \max\{t_{ik}\} \quad (7)$$

假设对于任意不同的制造资源 a 和 b 在添加第 i 个任务后满足负载平衡,则对于除 f_{\min} 外的任何制

造实体 b ,满足 $L_{\min} - L_b \leq 0$,第 k 次操作后, f_{\min} 中添加的任务 j 的负载与其他任何制造资源 b 的负载差描述如下

$$|L_{\min} + t_{ik} - L_b| \leq t_{ik} \leq \max\{t_{ik}\} \quad (8)$$

其中,在添加新任务后,资源 f_{\min} 与其他制造资源之间的负载差满足负载平衡,则其他设备之间的负载差不受新添加任务的影响。

在经过上述过程的计算评估后,形成的备选竞价集合描述如下

$$\begin{cases} S_i = \{U_i(s^1), U_i(s^2), \dots, U_i(s^k)\} \\ U_i(s^k) = \sum_{i=1}^{N_i} \sum_{j=1}^{N_j} V_{ijk} T(O_{ij}, M_k) \end{cases} \quad (9)$$

其中, S_i 表示考虑全局设备负载平衡的备选竞价集,假设 $U_i(s^k)$ 为备选竞价集中的最大投标值, $U_i(s^k)$ 定义为设备 M_k 根据合同网协议对工序 O_{ij} 的完工时间,该值包含了设备 M_k 任务队列中所有待加工工件的加工时间。 $T(O_{ij}, M_k)$ 表示设备 M_k 对工序 O_{ij} 的加工时间, V_{ijk} 取值描述如下

$$V_{ijk} = \begin{cases} 1, & \text{工序 } O_{ij} \text{ 在设备 } M_k \text{ 上加工} \\ 0, & \text{其他情况} \end{cases} \quad (10)$$

其中,当工序 O_{ij} 在设备 M_k 上加工时, V_{ijk} 取值为1。然后,基于工序约束的接力协作模型可以通过CNP求出 $U_i(s^k)$ 。

最终,为了分配制造任务 O_i ,定义决策函数描述如下

$$\arg = \min_{M \in M_{\phi(j)}} \left(\underbrace{C_{M_i M_j}^L}_{\text{LogisticsCosts}} + \underbrace{C_{M_j}^P}_{\text{ProcessingCosts}} \frac{\sum_{j \in J^M} x_{jM} \text{dur}_j}{\sum_{M \in M_{\phi(j)}} \sum_{j \in J^M} x_{jM} \text{dur}_j} \right) \quad (11)$$

其中, M_i 是制造任务 O_i 的时序工艺约束中紧前工序的加工设备, M_j 是投标当前工序的加工设备,如果当前工序为该任务的第一道工序,则 M_i 为原料仓库; $M_{\phi(j)}$ 是制造任务 O_i 当前工序具备加工能力的所有可供选择的制造资源集合。该决策函数同时考虑了实时物流时间成本(LogisticsCosts)和加工时间成本(ProcessingCosts)。同时,引入了历史权重因子(MachineUtilizationWeight),将竞标设备当前的相对利用率与所有同属性设备的利用率进行了权衡,从而防止只使用特定设备的作业分配,

确保设备间的负载均衡。这个权重因子的分子 $\sum_{j \in J^M} x_{jM} \text{dur}_j$ 表示投标设备 M_j 的相对工作时长，分母 $\sum_{M \in M_{\{j\}}} \sum_{j \in J^M} x_{jM} \text{dur}_j$ 表示所有同属性设备的总工作时长，该权重是队列长度估计的一种变体。

在调度智能体通过 CNP 完成一轮招标后，还需要借助 MCTS 进行标书评估。通过建立的树结构正向工序关系有向图，MCTS 以此为路径地图进行自博弈搜索，基于 MCTS 的决策优化过程如图 5 所示。一般情况下，调度智能体根据任务 J_x 的工艺特征将其分解为具有工序约束的制造任务集 $\{O_x^1, O_x^2, \dots, O_x^y\}$ ，每道工序可由一个或多个可替换单元集 $\{M_{x1}^y, M_{x2}^y, \dots, M_{xk}^y\}$ 完成。CNP-MCTS 机制的决策过程主要在调度智能体接收到各制造单元的加工投标后进行评估，确定当前状态下最适合的加工单元。其中，车间内各设备状态的可考虑任务发出时刻，各制造单元的加工特性和该任务在各单元上的加工特性，最终决定该任务的加工路径。

在图 5 中，任务 O_x^y 可由 3 个单元 M_{x1}^y ， M_{x2}^y 和 M_{xk}^y 进行加工，则任务 O_x^y 在当前状态 $S1$ 下，有动作集 $A(S1) = \{a1(s1), a2(s1), a3(s1)\}$ ，利用 MCTS 算法的 UCT 搜索策略，决定加工该任务的制造资源为 M_{xk}^y 。在任务 J_x 加工完成时刻，利用同样 UCT 策略决定后续加工任务 J_{x+1} 的制造资源为 $M_{(x+1)k}^y$ 。

为了实现优化决策，必须结合 MCTS 算法的回溯和迭代过程。首先，工件（或任务）在当前状态 s_1 选择特定动作 a_1 （即确定加工单元）后，由当前局面建立根节点，根据正向工序关系有向图生成全部子节点，分别进行模拟；然后，从根节点开始，进行最佳优先探索，在探索过程中利用 UCT 算法

分别计算每个节点的 UCT 值并选择具有最大值的子节点，如果该子节点不是叶节点，则以此节点作为根节点，重新开始进行探索，直至找到叶节点；搜索到叶节点后，如果叶节点未曾被模拟加工，则对叶节点进行模拟加工，否则需要为这个叶节点根据正向工序关系有向图继续生成子节点，并进行模拟加工；根据模拟加工的时间与成本更新该节点及各级父节点，同时增加该节点以上所有节点的访问次数并回溯回根节点；最后，根据预先设定的搜索时间或者次数进行停止，从当前局面的子节点中挑选完工时间或完工成本最低的给出决策。

针对当前任务 t 和状态 s ，在给定的决策过程和默认 Rollout 策略的情况下，CNP-MCTS 在计算开销（如最大决策时间 100 ms、最大迭代次数 100 次）内反复调用选择、扩展、仿真、回溯 4 个过程，最后根据返回值选择当前状态下最优机器智能体完成任务分配。

上述探索学习过程的最终目的是在确定 MCTS 因素（选择、扩展、仿真、回溯）的情况下，控制智能体在探索和利用之间的平衡，进而确定可加工单元的动态选择以最优化既定优化指标（如最短完工时间、最小加工成本等）。

3 基于 MAS 的制造车间自组织运作验证

3.1 案例描述

基于 MAS 的制造车间自组织运作实验场景布置如图 6 所示。该系统主要由以下制造实体组成：1 个用于原材料存储的自动化立体仓库、1 个用于成品存储的自动化立体仓库、2 台磁导引 AGV、3 台六自由度机械手、2 台数控车床、2 台数控加工

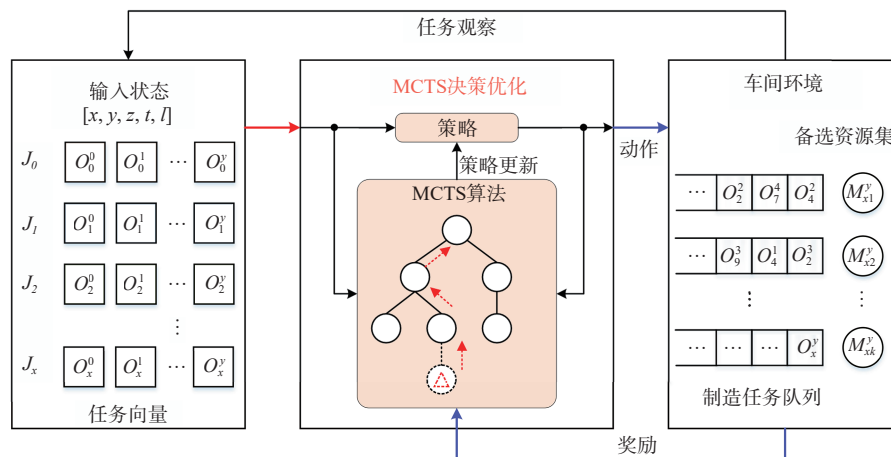


图5 基于MCTS的决策优化过程

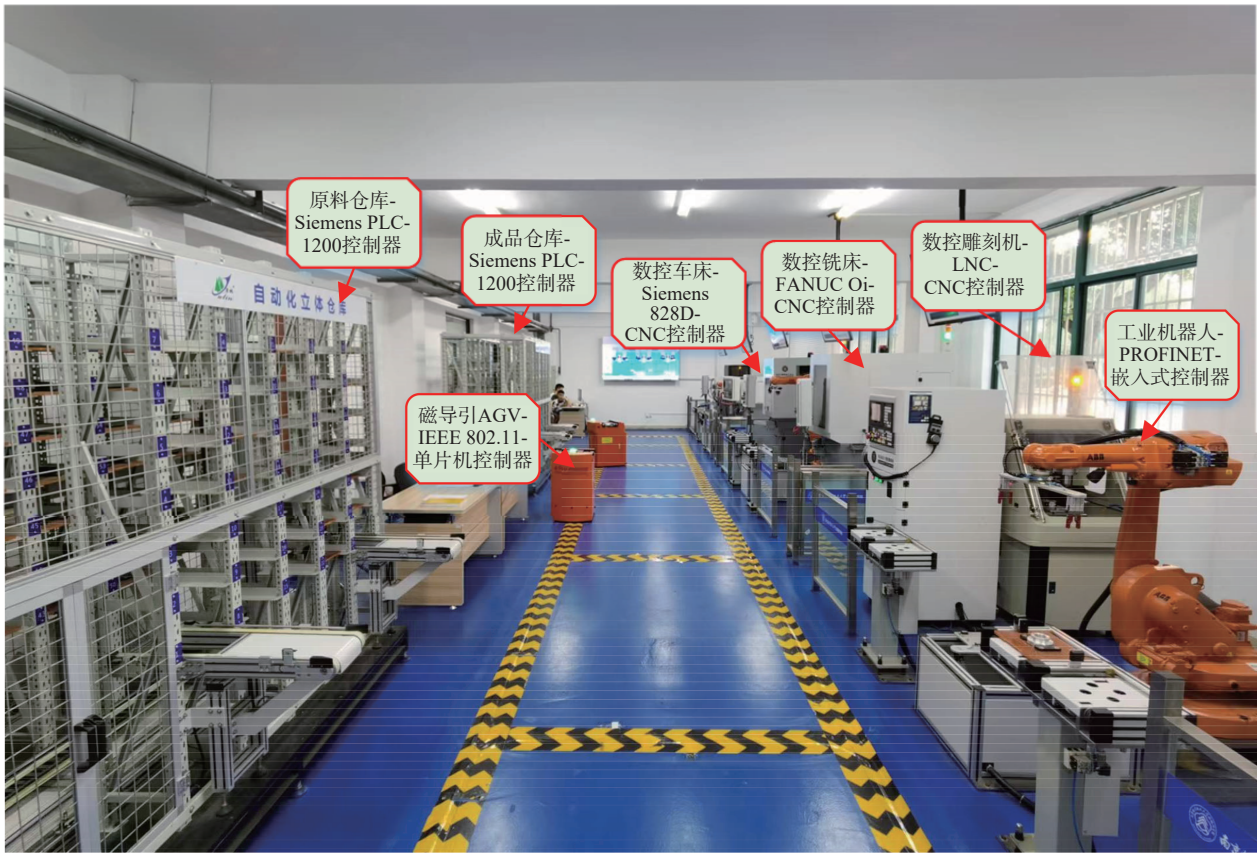


图6 基于MAS的制造车间自组织运作实验场景布置

中心、2台数控雕刻机，上面提到的每个加工单元都配有2个物料等待区。

传统车间调度问题的求解主要用元启发式方法或静态调度算法，一般不使用协商交互等动态调度的方式减小搜索空间^[26]。在这些静态调度算法中，有多种基准调度规则，如FIFO、SJF、LJF等^[27]。由于其简单的结构和易于决策的优势，这些基准规则被广泛用于车间现场以解决调度问题。此外，在MAMS运行期间，由工件智能体管理生产执行的信息跟踪，取代传统的RFID、条形码等物理标签，自组织制造车间加工任务实例见表1。该任务实例总共由6台加工设备、3个制造订单、12个加工任务、28道具体工序组成。此外，根据2台加工设备的距离将AGV的物料运输时间也考虑了进来。传统静态调度的目的是根据这些约束关系将这28道工序进行排序、重组，直至找到最小完工时间。而在自组织制造车间中，以加工任务作为调度的发起方，由各制造单元相互竞争，直至完成所有制造任务的加工。

3.2 实验结果与分析

基于MAS的制造车间自组织任务调度优化属于柔性车间作业调度问题（FJSP, flexible job-shop scheduling problem），对于存在缓冲区的FJSP根据资源选择条件不同，调度过程可分为工件选机床和机床选工件两个阶段，通过这两个阶段的选择可大致确定该任务的预计完工时间，进而确定整个订单的交货期^[28]。

本文提出的CNP-MCTS方法与3个最常用的基线调度技术（FIFO、SJF和LJF）进行对比，以寻找最佳策略来实现智能制造系统的动态调度。CNP-MCTS协商分配机制通过接口与智能制造系统环境集成，实现了探索最优调度策略的目标。实验结果根据以下3个性能标准进行评估。

- 1) 最大完工时间：用于计算制造任务的总完工时间，体现了算法的整体性能。
- 2) 工件平均等待时间：用于评估制造任务（工件）平均等待时间是否减少，体现了算法的实时性。
- 3) 平均设备利用率：用来判断所应用的调度方法是否提高了制造系统的整体效率。

表1 自组织制造车间加工任务实例

订单号	任务号	工艺路线	每道工序对应加工时间/s	运输及准备时间/s
01	T1	铣削→车削→雕刻	12.3→15.6→13.3	70
	T2	铣削→车削→雕刻	12.3→15.6→13.3	
	T3	车削→铣削	20.0→5.3	
	T4	雕刻→铣削	10.6→7.3	
	T5	车削→铣削	20.0→5.3	
02	T6	铣削→车削→雕刻	12.3→15.6→13.3	75
	T7	雕刻→铣削	10.6→7.3	
	T8	雕刻→铣削	10.6→7.3	
	T9	雕刻→铣削	10.6→7.3	
03	T10	雕刻→铣削	10.6→7.3	80
	T11	铣削→车削→雕刻	12.3→15.6→13.3	
	T12	雕刻→铣削	10.6→7.3	

3.2.1 CNP-MCTS 机制自组织协商分配效率

常见任务队列下的不同算法对比实验如图7所示。其中，CNP-MCTS确定了12个作业的时间表，其最大完成时间为300 s。图7(a)中FIFO、SJF和LJF的值分别为310 s、340 s和360 s，本文所提方法明显优于所有基线调度规则，这是因为CNP-MCTS通过应用UCT评估函数来仿真提前获取进站操作的处理时间和相应可用机器的信息。本文所提CNP-MCTS方法降低了所有工件在整个生产过程中的平均等待时间，当与基准调度规则相比时，平均等待时间只有42.6 s，而这些基准调度方法的平均等待时间都超过50 s，详细结果如图7(b)所示。由于平均等待时间的减少，到达作业在调度系统中的任务完成时间也相应减少，导致整体计划提前完成。相反，CNP-MCTS通过最大化资源利用率获得了所有机器中最高的平均设备利用率，如图7(c)所

示，相较于FIFO、SJF和LJF 3种基线调度规则，CNP-MCTS方法平均设备利用率高达86.94%，提高了制造系统的整体效率。

4种策略下的对比实验数据见表2，针对3种不同场景比较了CNP-MCTS算法与基线调度规则，这3个场景包含具有不同数量的任务和工序，但可用机器数量相同的示例问题。通过不同场景下的对比，可以检查本文所提算法在不同大小样本问题中的有效性。实验结果表明，在所有3个场景中，CNP-MCTS算法的表现优于所有基线调度规则，并且3个评估指标均实现了最佳效果。

任务数量增加情况下4种调度算法的最大完工时间对比如图8所示，可以看出，在开始时间到CNP-MCTS探索前16个作业时，它的性能与所有基线调度规则持平，这是由于这些基准调度规则采用贪婪搜索方法。然而，随着每个任务数量的增

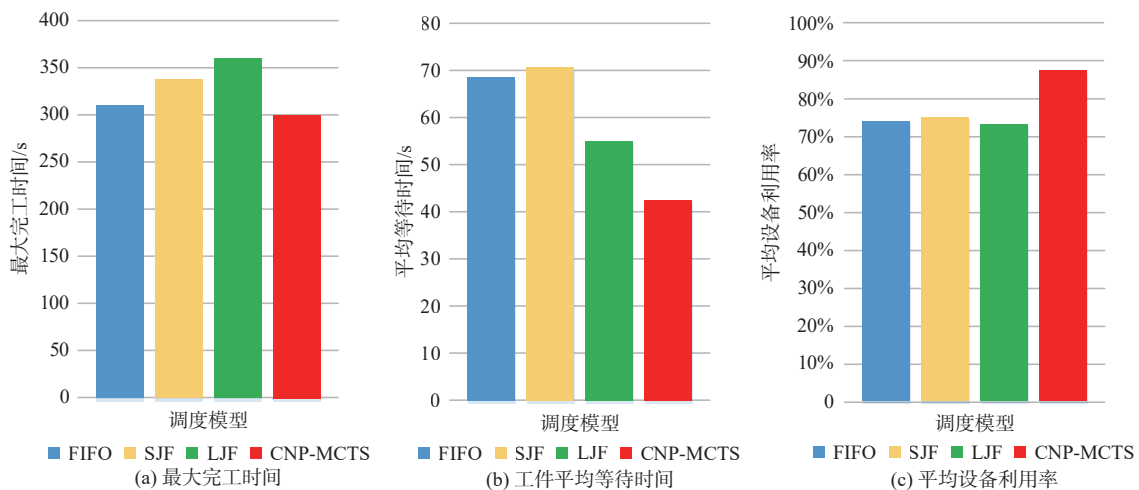


图7 常见任务队列下的不同算法对比实验

表2 4种策略下的对比实验数据

场景	$n \times m \times o$	算法类型	最大完工时间/s	平均等待时间/s	平均设备利用率
1	6×6×18	FIFO	310	68.3	74.05%
		SJF	340	70	75.19%
		LJF	360	55	72.39%
		CNP-MCTS	300	42.6	86.94%
2	12×6×36	FIFO	590	229.17	77.91%
		SJF	690	202.5	79.84%
		LJF	650	196.67	65.99%
3	24×6×72	CNP-MCTS	540	132.36	89.72%
		FIFO	1 090	459.58	81.58%
		SJF	1 190	423.33	83.01%
		LJF	1 210	395.42	68.51%
		CNP-MCTS	970	206.78	90.62%

加，CNP-MCTS的探索能力也在增强，并且性能越来越好，当最终达到第37个作业时，它优于所有基线方法。这是因为随着任务的增加，MCTS应用更多的仿真次数，并在每个周期访问额外的子节点以改进搜索策略。因此，CNP-MCTS更适合于求解较大规模的实时动态调度问题。

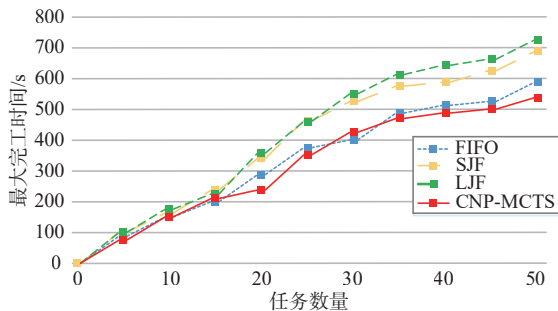


图8 任务数量增加情况下4种调度算法的最大完工时间对比

3.2.2 CNP-MCTS 利用与探索的平衡

CNP-MCTS算法在动态调度实验中取得较好的效果，相较于调度结果来说，MCTS算法本身还能够实现利用与探索的平衡。在每一轮算法迭代中，只要选择胜率高的节点进行下一步仿真即可，但贪婪搜索方法会使得搜索很快结束，导致搜索不充分，进而错过最优解^[29]。因此UCT算法增加了第二项探索单元，这个单元更倾向于未被探索的节点（奖励值较小的节点）。CNP-MCTS机制下UCT算法探索单元取值趋势如图9所示，该值可以调整CNP-MCTS机制在“探索”与“利用”之间的平衡，进而避免模型早熟。从图9可以看出，在算法初期，调度智能体需要探索更大的解空间，因此权重 c 取值趋于1；在迭代次数至上限 N 的80%时，权重 c

取值趋于稳定且低于0.2；当迭代次数达到搜索上限 N 时，UCT算法基本退化为一个贪心算法，贪心度量为过去平均表现最好的节点，算法很快收敛。这种趋势与决策优化过程是相似的：初期探索时需要更大的搜索空间，后期决策时需要算法收敛。因此，CNP-MCTS补充了传统CNP没有任何探索能力的短板。

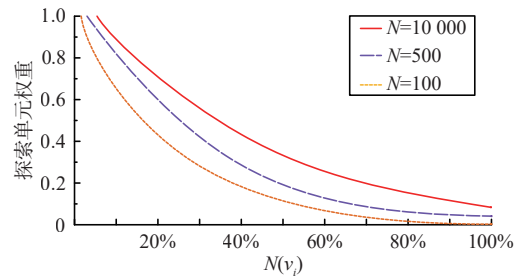


图9 CNP-MCTS机制下UCT算法探索单元取值趋势

4 结束语

本文将多智能体合同网协议与蒙特卡洛树搜索算法相结合，提出了制造车间控制系统自组织运作机制与制造任务调度优化方法。首先，基于MAS技术对车间制造资源进行了封装，使其成为具备通信功能的自治实体，在此基础上对传统CNP协商分配策略进行优化，在智能体决策时调用MCTS算法对标书进行评估，以确保决策的全局有效性。然后，在MAMS框架下，各制造单元通过智能体间的协商机制实现制造单元间的互联。最后，通过实验表明，CNP-MCTS算法较传统的先进先出、最短作业时间优先和最长作业时间优先具有更好的全局优化效果。

参考文献:

- [1] ACATECH. Cyber-Physical systems: driving force for innovations in mobility, health, energy and production[M]. Berlin: Springer International Publishing, 2012.
- [2] 王世进. 基于自治与协商机制的柔性制造车间智能调度技术研究[D]. 上海: 上海交通大学, 2008.
WANG S J. Intelligent scheduling technology inflexible manufacturing shop floor based on mechanism of autonomy and coordination[D]. Shanghai: Shanghai Jiao Tong University, 2008.
- [3] WANG L C, LIN S K. A multi-agent based agile manufacturing planning and control system[J]. *Computers & Industrial Engineering*, 2009, 57(2): 620-640.
- [4] PUJO P, OUNNAR F, POWER D, et al. Wireless Holon network for job shop isoarchic control[J]. *Computers in Industry*, 2016, 83: 12-27.
- [5] REAIDY J, MASSOTTE P, DIEP D. Comparison of negotiation protocols in dynamic agent-based manufacturing systems[J]. *International Journal of Production Economics*, 2006, 99(1/2): 117-130.
- [6] BARENJI R V, BARENJI A V, HASHEMPOUR M. A multi-agent RFID-enabled distributed control system for a flexible manufacturing shop[J]. *The International Journal of Advanced Manufacturing Technology*, 2014, 71(9): 1773-1791.
- [7] 程涛, 吴波, 杨叔子, 等. 支持分布式网络化制造的智能数控系统的研究[J]. *中国机械工程*, 2004, 15(8): 688-693.
CHENG T, WU B, YANG S Z, et al. Study on intelligent CNC system for distributed networked manufacturing[J]. *China Mechanical Engineering*, 2004, 15(8): 688-693.
- [8] RODRIGUES N, OLIVEIRA E, LEITÃO P. Decentralized and on-the-fly agent-based service reconfiguration in manufacturing systems[J]. *Computers in Industry*, 2018, 101: 81-90.
- [9] TEHRANI N N H, SUGIMURA N, IWAMURA K, et al. Multi agent architecture for dynamic incremental process planning in the flexible manufacturing system[J]. *Journal of Intelligent Manufacturing*, 2010, 21(4): 487-499.
- [10] WANG S Y, WAN J F, ZHANG D Q, et al. Towards smart factory for industry 4.0: a self-organized multi-agent system with big data based feedback and coordination[J]. *Computer Networks*, 2016, 101: 158-168.
- [11] WANG J, ZHANG Y F, LIU Y, et al. Multiagent and bargaining-game-based real-time scheduling for Internet of Things-enabled flexible job shop[J]. *IEEE Internet of Things Journal*, 2019, 6(2): 2518-2531.
- [12] KIM Y G, LEE S, SON J, et al. Multi-agent system and reinforcement learning approach for distributed intelligence in a flexible smart manufacturing system[J]. *Journal of Manufacturing Systems*, 2020, 57: 440-450.
- [13] MA S Y, DING W, LIU Y, et al. Industry 4.0 and cleaner production: a comprehensive review of sustainable and intelligent manufacturing for energy-intensive manufacturing industries[J]. *Journal of Cleaner Production*, 2024, 467: 142879.
- [14] GUO C, XIONG W, LIU C X. Prospects and current researches on improvement of contract net protocol[J]. *Journal of Equipment Academy*, 2016(27): 82-89.
- [15] ZHU Y F, LOU P H, QIAN X M, et al. Research on job shop scheduling method based on improved contract net protocol[J]. *Machine Design and Manufacturing Engineering*, 2018(47): 97-102.
- [16] YEUNG W L. Agent-based manufacturing control based on distributed bid selection and publish-subscribe messaging: a simulation case study[J]. *International Journal of Production Research*, 2012, 50(22): 6339-6356.
- [17] 李浩, 邢志远, 李琳利, 等. 基于多智能体的工业数字孪生系统云边端架构与关键技术[J]. *计算机集成制造系统*, 2024, 30(11): 3755-3770.
LI H, XING Z Y, LI L L, et al. Cloud-edge-device system architecture and key technologies of industrial digital twin system based on multi-agent[J]. *Computer Integrated Manufacturing Systems*, 2024, 30(11): 3755-3770.
- [18] WANG C, LV Y, WANG Q, et al. Service-oriented real-time smart job shop symmetric CPS based on edge computing[J]. *Symmetry*, 2021, 13(10): 1839.
- [19] GUI Y, ZHANG Z Q, TANG D B, et al. Collaborative dynamic scheduling in a self-organizing manufacturing system using multi-agent reinforcement learning[J]. *Advanced Engineering Informatics*, 2024, 62: 102646.
- [20] WANG X H, ZHANG L, LIN T Y, et al. Solving job scheduling problems in a resource preemption environment with multi-agent reinforcement learning[J]. *Robotics and Computer-Integrated Manufacturing*, 2022, 77: 102324.
- [21] ZHANG N M, SHEN Y L, DU Y, et al. Counterfactual-attention multi-agent reinforcement learning for joint condition-based maintenance and production scheduling[J]. *Journal of Manufacturing Systems*, 2023, 71: 70-81.
- [22] ZHOU T, TANG D B, ZHU H H, et al. Multi-agent reinforcement learning for online scheduling in smart factories[J]. *Robotics and Computer-Integrated Manufacturing*, 2021, 72: 102202.
- [23] WASEEM M, CHANG Q. From Nash Q-learning to Nash-MADDPG: advancements in multiagent control for multiproduct flexible manufacturing systems[J]. *Journal of Manufacturing Systems*, 2024, 74: 129-140.
- [24] YE X F, DENG Z Y, SHI Y J, et al. Toward energy-efficient routing of multiple AGVs with multi-agent reinforcement learning[J]. *Sensors*, 2023, 23(12): 5615.
- [25] WANG L P, TANG D B, SUN H W, et al. Enabling technology of multiagent manufacturing system: a novel mode of self-organizing IoT manufacturing[J]. *Transactions of Nanjing University of Aeronautics and Astronautics*, 2021, 38(5): 876-892.
- [26] PULIKOTTIL T, ESTRADA-JIMENEZ L A, UR R H, et al.

Agent-based manufacturing: review and expert evaluation[J]. The International Journal of Advanced Manufacturing Technology, 2023, 127(5): 2151-2180.

[27] ZHOU C M, HUANG B D, FRÁNTI P. A review of motion planning algorithms for intelligent robots[J]. Journal of Intelligent Manufacturing, 2022, 33(2): 387-424.

[28] LEFRANC G, LOPEZ-JUAREZ I, GATICA G. Enhancing FMS performance through multi-agent systems in the context of industry 4.0[J]. Studies in Informatics and Control, 2024, 33(2): 5-14.

[29] HELLIWELL T, MORGAN B, VINCENT A, et al. Reconfigurable scheduling as a discrete-event process: Monte Carlo tree search in industrial manufacturing[C]//Proceedings of the 2nd International Conference on Innovative Intelligent Industrial Production and Logistics. Setúbal: SCITEPRESS, 2021: 151-162.

[作者简介]



王立平(1990-), 男, 博士, 南京航空航天大学机电学院助理研究员, 主要研究方向为自组织智能制造系统及其关键使能技术。



赵振(1998-), 男, 南京航空航天大学机电学院博士生, 主要研究方向为工业大模型驱动的多智能体智能制造系统。



刘长春(1995-), 男, 博士, 南京航空航天大学机电学院助理研究员, 主要研究方向为基于深度强化学习的设备状态分析与人机协作。



唐敦兵(1972-), 男, 博士, 南京航空航天大学机电学院教授, 主要研究方向为工业互联网与数字孪生。