doi:10.12068/j.issn.1005-3026.2026.20250130

铁水运输机车调度优化:基于强化学习的Benders分解算法

刘佳斌[1]，蒋忠中[1,2]，赵金龙[1]，易红亮[3]
（1.东北大学 工商管理学院，辽宁 沈阳 110169；2.东北大学 辽宁省服务型制造研究院，辽宁 沈阳 110169；
3.东北大学 数字钢铁全国重点实验室，辽宁 沈阳 110819）

摘 要：铁水运输作为衔接炼铁与炼钢两个工序的关键环节,其机车调度问题具有服务时间窗、后进先出装卸顺序等钢铁生产的关键特征.针对该问题,以机车行驶总时间最小化为目标,提出并构建2种混合整数规划模型,分别为基于节点和基于模式的机车调度模型.针对模型特点,提出一种基于强化学习的Benders分解启发式算法.该算法采用Benders分解算法框架将基于模式的机车调度模型拆分,同时引入Q-learning算法自适应地调整主问题模式决策变量的固定与排除集合以加速收敛.实验结果表明,与求解器相比,所提出的启发式算法在大规模算例中求解质量与计算时间更具优势.本研究可为钢铁企业提升铁水运输系统的运行效能、实现生产工序间的高效协同提供有效决策支持.

关 键 词：铁水运输;机车调度优化;取送货路径优化;Benders分解算法;Q-learning算法
中图分类号：TP 18 **文献标志码**：A **文章编号**：1005-3026(2026)06-0044-10

Locomotive Scheduling Optimization for Molten Iron Transportation: Reinforcement Learning-Based Benders Decomposition Algorithm

LIU Jia-bin[1], *JIANG Zhong-zhong*[1,2], *ZHAO Jin-long*[1], *YI Hong-liang*[3]
(1. School of Business Administration, Northeastern University, Shenyang 110169, China; 2. Liaoning Provincial Institute of Service-Oriented Manufacturing, Northeastern University, Shenyang 110169, China; 3. State Key Laboratory of Digital Steel, Northeastern University, Shenyang 110819, China. Corresponding author: ZHAO Jin-long, E-mail: zhaojinlong@mail.neu.edu.cn)

Abstract: Molten iron transportation serves as a key link connecting ironmaking and steelmaking processes, and its locomotive scheduling problem is characterized by key features of steel production, such as service time windows and a last-in-first-out loading and unloading sequence. To address this problem, with the objective of minimizing the total locomotive travel time, two mixed-integer programming models, namely a node-based locomotive scheduling model and a pattern-based locomotive scheduling model, were proposed and constructed. According to the model characteristics, a reinforcement learning-based Benders decomposition heuristic algorithm was proposed. In the proposed algorithm, the Benders decomposition algorithm framework was adopted to decompose the pattern-based locomotive scheduling model, and simultaneously, the Q-learning algorithm was introduced to adaptively adjust the fixation and exclusion sets of the pattern decision variables in the master problem to accelerate convergence. Experimental results indicate that compared with the solver, the proposed heuristic algorithm possesses greater advantages in solution quality and computational time in large-scale instances. Effective decision support can be provided by this study for steel enterprises to improve the operational efficiency of the molten iron transportation system and achieve efficient coordination between production

收稿日期：2025-11-03
基金项目：国家社会科学基金重大项目(23&ZD050).
作者简介：刘佳斌(1998—),男,河北廊坊人,东北大学博士研究生.
通信作者：赵金龙,E-mail:zhaojinlong@mail.neu.edu.cn.

processes.

Key words: molten iron transportation; locomotive scheduling optimization; pickup and delivery routing optimization; Benders decomposition algorithm; Q-learning algorithm

钢铁工业作为国民经济的重要基础产业[1]，其发展水平深刻影响国家工业化进程与制造业核心竞争力.在钢铁生产全流程中，涉及诸多复杂的工程优化问题，例如矿区运输路径规划[2]、成品钢材物流网络优化[3]、仓储空间配置[4]以及钢板切割方案设计[5]等.其中，铁水运输作为衔接炼铁与炼钢关键生产环节的纽带，其运输效率直接决定前后工序的协同水平与全流程的能耗表现，因而成为影响钢铁生产系统整体效能的重要一环.然而，在实际作业中，铁水运输过程仍面临诸多问题，如运输路径交错复杂、多工序作业时序协调困难、机车资源调配紧张等.上述因素不仅制约生产节奏的连续性，亦极易导致铁水生产与消耗的匹配失衡，为铁水温度维持乃至整个系统的平稳运行带来严峻挑战.因此，亟须对铁水运输中的机车调度问题展开深入分析，通过优化机车运送作业计划，提升运输环节的整体作业效率，进而为保障钢铁生产系统的高效、稳定与安全运行提供理论依据与方法支撑.

铁水需要储存在特制的鱼雷罐中，在实际生产过程中，鱼雷罐通过机车牵引，经由厂内铁路网络在高炉、转炉、缓冲站等固定设施间进行转运，铁水运输场景如图1所示.这一过程不仅要求确保鱼雷罐车的准时送达，还需严格控制运输过程中的安全规范限制，形成极具挑战性的机车调度优化问题.其关键特征主要体现在两个方面：首先，钢铁的连续生产特性使得铁水运输的每个步骤，即鱼雷罐的获取与交付节点都具有严格的服务时间窗，装有铁水的鱼雷罐，称为实罐，超时获取或交付会导致铁水在等待过程中温降超限；未装有铁水的鱼雷罐，称为空罐，超时获取与交付会引发铁路网络节点的拥堵和鱼雷罐体的温度波动问题；其次，受铁路机车作业安全规范限制，机车装卸鱼雷罐必须遵守后进先出规则，实罐和空罐因稳定性要求不能同时运送，机车牵引能力有限，具有明确的鱼雷罐运送容量限制，极大限制机车调度的灵活性.因此，在复杂的生产与安全条件共同约束下，该问题在理论上是一个极具挑战性的NP难问题，有待提出新的求解方法.

本文研究问题属于取送货路径优化问题[6]，是车辆路径问题的一个重要分支.该问题核心在于协调取货与送货任务，确保货物在正确的时间和地点被提取并送达，同时满足一系列复杂的实际约束.针对此类问题，现有研究提出了精确算法[7-8]与启发式算法[9-10]，分别从最优性保证与计算效率两个角度提升问题求解能力.在精确算法方面，Alyasiry等[11]针对带时间窗与后进先出约束的取送货问题，构建基于片段的松弛网络流模型，并设计了一种分支切割算法，通过迭代添加割平面来排除不可行解.在此基础上，Aziez等[12]针对多对一的取送货场景，结合分支切割算法提出预处理技术以减少变量数量，并引入有效不等式以增强模型求解能力.在启发式算法方面，Pilati等[13]研究了考虑行驶距离和工作时间的多目标取送货优化问题，提出一种多目标模拟退火算法.Jiang等[14]研究了车辆与无人机协同的取送货问题，提出一种自适应大邻域搜索启发式算法，在大规模实例中取得了高质量解.此外，取送货问题普遍存在于多个行业场景，如产品回收再制造[15]、农业批发市场货物转运[16]、机场行李运输[17].

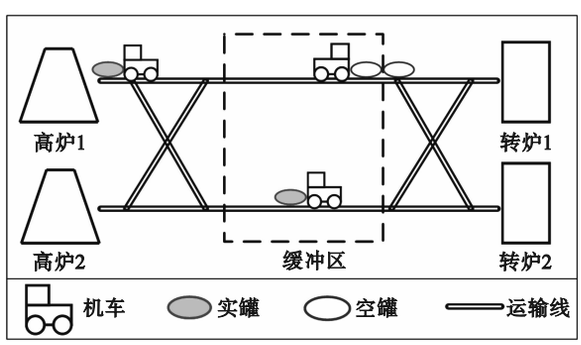


图1　铁水运输场景

Fig. 1　Molten iron transportation scenario

针对铁水运输场景，学者结合钢铁生产特性在机车调度问题中逐步引入更复杂的约束条件和优化目标，并提出具体的启发式算法或精确算法.Lübbecke等[18]针对钢铁行业机车调度问题提出一种基于预生成模式的混合整数规划模型以最小化机车总运送作业时间，引入集合划分模型并设计一种分支定价算法.Wang等[19]研究具有非线性目标函数的机车调度问题，引入机车容量限制和不同类型鱼雷罐不能同时运送的约束条

件.Tang等[20]考虑铁水在等待运输过程中的热能损耗以及排料过程中的性能衰减,提出一种分支定价算法,并设计多维标签结构和标签支配规则以提高定价子问题的求解效率.Cherkesly等[21]考虑装卸的后进先出顺序原则,提出3种分支定价切割算法,通过不同策略处理装载约束.在上述研究基础上,Huang等[22]构建了一个机车调度问题的整数规划模型,综合考虑了硬时间窗、容量限制、类型约束、装卸顺序等实际约束,提出基于重构优化的移除算子和插入算子,并设计了一种自适应大规模邻域搜索算法.Eom等[23]则将研究扩展至熔融材料运输与调度集成问题,在数学模型中考虑了时间效率、质量成本、控制员工作负荷等目标,并设计了生成初始解的启发式算法与一种Benders分解算法.Kweon等[24]进一步将机车运送请求类型细分,提出了钢铁行业中多对多机车路径优化问题,设计了一种基于逻辑Benders分解的启发式算法.

现有研究在铁水运输机车调度问题的模型建立与算法设计方面取得了一定进展,但仍存在明显不足.一方面,精确算法虽能保证理论上的最优性,但在面对大规模、强约束的现实场景时,往往因计算时间过长难以得到应用;另一方面,元启发式算法虽然具有较高的求解效率,但其依赖规则导向的搜索机制相对僵化,在优化过程中常出现过早收敛现象而陷入局部最优,难以获得全局最优解.

为此,本文提出一种融合强化学习和Benders分解算法的启发式算法,用于求解铁水运输机车调度优化问题.Benders分解是一种经典的精确算法框架,通过将原问题分解为主问题与子问题[25],并迭代添加Benders割以逼近最优解.强化学习是一种以环境交互和奖励反馈为核心的机器学习方法[26],尤其适用于复杂动态系统中的决策问题.本文将机车调度模型中的复杂约束通过预处理阶段转化为满足约束条件的机车调度模式,构建一个计算效率更高的基于模式的整数规划模型;在此基础上采用Benders分解框架设计主问题和子问题,引入Q-learning算法自适应地固定或移除主问题中的模式决策变量,从而动态调整变量的搜索空间,加速Benders分解的迭代收敛过程,同时将搜索聚焦于潜力更大的优质解区域,从而提升求解质量.该方法融合了精确算法的收敛性保证与强化学习的动态适应优势,为复杂工业调度问题的求解提供新的思路.

1 铁水运输机车调度优化问题

1.1 问题描述

在钢铁生产过程中,高炉和转炉承担着不同的钢铁生产工序,高炉产生的熔融状态铁水需要运输至转炉进行炼钢工序.铁水运输过程通过专用的鱼雷罐进行储存,并由机车牵引调度完成转运工作.本研究聚焦于多罐场景下的铁水运输机车调度优化问题,这是一类具有硬时间窗、运输类型、容量限制、装卸顺序等复杂约束条件的取送货路径优化问题.具体而言,由于高炉与转炉的生产节奏存在差异,需要通过对鱼雷罐车合理调度以优化铁水运输过程来实现各工序间的衔接,从而保障高炉与转炉的生产需求.机车需要在厂区内执行4类运送请求:①将空罐从缓冲区运送至高炉,②将实罐从高炉运送至缓冲区,③将实罐从缓冲区运送至转炉,④将空罐从转炉运送至缓冲区.为了保障高炉与转炉的连续生产,每个请求均根据生产计划设置严格的时间窗,机车获取与交付每个鱼雷罐需严格在其时间窗内完成.同时,为确保运输期间的作业安全,系统设有一系列严格的作业约束:在运输类型方面,同一辆机车不可混运,即机车不可同时运送实罐与空罐;在容量限制方面,每台机车同时最多牵引2个实罐或4个空罐;在装卸顺序方面,机车执行后进先出的装卸原则,并且,机车访问节点顺序亦须遵守后进先出原则.

针对上述多重复杂约束,一种有效的建模思路是引入“模式”[24]这一概念.模式是一段预定义的、满足上述复杂约束条件的机车行驶路线,可完整描述机车连续执行多个鱼雷罐的获取与交付任务.例如,一个模式可以具体表示为(1^+,2^+,2^-,1^-),指一辆机车依次前往节点1^+和2^+获取请求1和请求2的鱼雷罐,然后依次前往节点2^-和1^-交付请求2和请求1的鱼雷罐.这样的运送模式符合运输类型、容量限制、装卸顺序约束,且每个模式设有运输路线中起点和终点的时间窗.机车只要在模式起点和终点的时间窗内访问,那么就可以在满足模式内各节点时间窗约束下完成访问工作.通过引入模式概念,可以将具有复杂约束的运输问题,转化为对上述已验证满足约束的模式的调度优化问题,大幅度减小求解难度.

对问题简述如下:已知钢铁生产计划中有一组鱼雷罐运送请求,所有请求运送R^+,每个请求需要从节点$i \in R^+$获取鱼雷罐,并将鱼雷罐交付到节点$i+|R^+|$,其中R^+表示运送请求的获取节点集合,$|\cdot|$表示集合中元素的数量.该过程由一组机车K负责执行,要求在满足节点服务时间窗约束、同类鱼雷罐运输、容量限制与后进先出装卸顺序的前提下,优化机车的调度方案,以最小化机车行驶的总时间.

1.2　符号说明

本文构建了基于节点的铁水运输机车调度模型MIP1和基于模式的铁水运输机车调度模型MIP2,两个数学模型所有的集合、参数、变量符号定义如表1所示.基于节点的MIP1模型以网络节点为基本单元,直接描述机车在各节点间的行驶路径,需综合考虑多种复杂的作业约束.基于模式的MIP2模型则在预处理阶段生成所有满足作业约束的机车运输模式,将问题转化为模式的组合与排序优化,虽引入了较多的决策变量,却显著简化了模型中的约束结构.两个模型分别从节点和模式层面刻画铁水运输机车调度问题,为后续算法设计与对比分析提供了理论基础.

表1　符号说明

Table 1　Notation description

类型	符号	模型	说明
集合	I	MIP1	所有节点集合,$i,j \in I$
	K	MIP1	机车集合,$k \in K$
	R^+	MIP1	鱼雷罐运送请求的获取节点集合,$R^+ \subset I$
	R^-	MIP1	鱼雷罐运送请求的交付节点集合,$R^- \subset I$
	R_{E}^+	MIP1	空罐运送请求的获取节点集合,$R_{\mathrm{E}}^+ \subset R^+$
	R_{L}^+	MIP1	实罐运送请求的获取节点集合,$R_{\mathrm{L}}^+ \subset R^+$
	P	MIP2	所有模式p的集合,$p \in P$
	P_i	MIP2	包含节点i的模式p集合,$P_i \subset P$
参数	$D_{i,j}$	MIP1	机车从节点i到j的行驶时间
	L_i	MIP1	节点i的装卸服务时间
	Q_i	MIP1	机车在节点i的装载容量变化量
	S_i	MIP1	节点i的最早服务时间边界
	E_i	MIP1	节点i的最晚服务时间边界
	U_k	MIP1	机车k的出发节点,$U_k \in I$
	H_k	MIP1	机车k的停车节点,$H_k \in I$
	L_p^*	MIP2	模式p内的所有节点的装卸服务时间之和
	D_p^*	MIP2	机车在模式p内行驶时间之和
	$D'_{p,q}$	MIP2	机车从模式p终点到模式q起点的行驶时间
	S_p^{S}	MIP2	模式p开始服务时间的最早时间边界
	S_p^{E}	MIP2	模式p开始服务时间的最晚时间边界
	E_p^{S}	MIP2	模式p结束服务时间的最早时间边界
	E_p^{E}	MIP2	模式p结束服务时间的最晚时间边界
	$Z_{p,q,k}$	MIP2	如果机车k可以从模式p的终点行驶到模式q的起点,则$Z_{p,q,k}$=1,否则$Z_{p,q,k}$=0
变量	$x_{i,j,k}$	MIP1	如果机车k从节点i行驶到节点j则$x_{i,j,k}=1$,否则$x_{i,j,k}=0$
	$y_{i,k}$	MIP1	机车k离开节点i后的负载量
	$z_{p,q,k}$	MIP2	如果机车k从模式p行驶到模式q则$z_{p,q,k}=1$,否则$z_{p,q,k}=0$
	$s_{p,k}$	MIP2	机车k开始模式p的时间
	$e_{p,k}$	MIP2	机车k完成模式p的时间
	t_i,t_j	MIP1	节点i,j开始服务的时间

1.3　基于节点的机车调度模型

构建一个基于节点的机车调度模型MIP1,模型中的节点用于记录各个请求的获取与交付位置信息以及机车的出发和停车位置信息.节点间的行驶时间与厂内铁路网络物理地点间的行驶时间严格对应.同时,表示机车在节点间行驶的决策变量$x_{i,j,k}$,仅考虑节点i和j对应物理地点真实连接的情形,且该变量只包含调度逻辑上有效的路线,例如不存在从一辆机车的出发点到另一辆机车的停车点的路线.模型MIP1如下:

$$\min \sum_{i\in I}\sum_{j\in I}\sum_{k\in K} D_{i,j}\times x_{i,j,k}. \tag{1}$$

$$\sum_{i\in R^+\cup\{H_k\}} x_{U_k,i,k}=1,\forall k\in K; \tag{2}$$

$$\sum_{i\in R^-\cup\{U_k\}} x_{i,H_k,k}=1,\forall k\in K; \tag{3}$$

$$\sum_{j\in I}\sum_{k\in K} x_{i,j,k}=1,\forall i\in R^+; \tag{4}$$

$$\sum_{j\in I} x_{i,j,k}-\sum_{j\in I} x_{i+|R^+|,j,k}=0,\forall i\in R^+,k\in K; \tag{5}$$

$$\sum_{j\in I} x_{i,j,k}-\sum_{j\in I} x_{j,i,k}=0,\forall i\in R^+\cup R^-,k\in K; \tag{6}$$

$$\sum_{j\in R^-} x_{i,j,k}-x_{i,i+|R^+|,k}=0,\forall i\in R^+,k\in K; \tag{7}$$

$$x_{i,j,k}-x_{j+|R^+|,i+|R^+|,k}=0,\forall i,j\in R^+,k\in K; \tag{8}$$

$$\sum_{j\in R_L^+} x_{i,j,k}+\sum_{j\in R_L^+} x_{j,i,k}=0,\forall i\in R_E^+,k\in K; \tag{9}$$

$$t_i+L_i+D_{i,j}+(x_{i,j,k}-1)\times M\leqslant t_j,\forall i\in I,j\in I,k\in K; \tag{10}$$

$$y_{i,k}-Q_i=y_{i+|R^+|,k},\forall i\in R^+,k\in K; \tag{11}$$

$$y_{i,k}+Q_j+(x_{i,j,k}-1)\times M\leqslant y_{j,k},\forall i\in I,j\in I,k\in K; \tag{12}$$

$$y_{U_k,k}=0,\forall k\in K; \tag{13}$$

$$x_{i,j,k}\leqslant X_{i,j,k},\forall i\in I,j\in I,k\in K; \tag{14}$$

$$Q_i\leqslant y_{i,k}\leqslant 4,\forall i\in R^+,k\in K; \tag{15}$$

$$0\leqslant y_{i,k}\leqslant 4+Q_i,\forall i\in R^-,k\in K; \tag{16}$$

$$S_i\leqslant t_i\leqslant E_i,\forall i\in I. \tag{17}$$

式(1)表示最小化机车行驶时间的总和;式(2)和式(3)分别限制每辆机车从出发节点驶离,并最后回到停车节点;式(4)确保每个请求的鱼雷罐获取节点仅被一辆机车访问一次;式(5)表示每个请求的鱼雷罐获取与交付必须由同一辆机车完成;式(6)确保每辆机车从一个鱼雷罐获取节点或交付节点驶入与驶离流量守恒;式(7)表示请求的交付节点只能在访问获取节点之后才能访问;式(8)表示鱼雷罐的获取节点访问顺序与交付节点访问顺序相反;式(9)表示实罐与空罐运送请求不能由一辆机车同时运输;式(10)保证每辆车行驶到每个节点时间连续;式(11)确保每辆机车驶离一个请求获取节点的负载量与到达该请求交付节点负载量相等;式(12)确保机车连续访问节点时,离开与到达的负载保持一致;式(13)规定每辆机车在原点的负载量为零;式(14)限制每辆机车只能访问铁路网络真实连接和调度逻辑有效的线路;式(15)和式(16)表示每个节点机车容量限制;式(17)表示各节点的硬时间窗约束.

1.4 基于模式的机车调度模型

构建一个基于模式的机车调度模型MIP2.在给出数学模型之前,首先介绍运送模式相关的时间窗参数的计算方法.假设机车在模式p中最先访问获取鱼雷罐的节点为i^+,最后访问交付鱼雷罐的节点为i^-,则模式p开始服务时间的最早时间边界S_p^S为机车在模式p中最先访问节点i^+的最早时间边界S_{i^+},模式p结束服务时间的最晚时间边界E_p^E为机车在模式p中最后访问节点i^-的最晚时间边界E_{i^-}.然后,依次按照模式p内节点访问顺序(i,j)计算其余节点j最早时间边界$s(p,j)=\max(s(p,i)+D_{i,j}+L_i,S_j)$,进而得到模式$p$结束服务时间的最早时间边界$E_p^S=s(p,i^-)$.最后,依次按照模式$p$内节点访问逆序$(j,i)$计算其余节点$i$最晚时间边界$l(p,i)=\min(l(p,j)-D_{i,j}-L_j,E_i)$,得到模式$p$开始服务时间的最晚时间边界$S_p^E=l(p,i^+)$.运送模式生成与上述时间窗的计算在预处理阶段完成.模型MIP2通过选取合适模式组合并优化模式分配调度,以最小化机车行驶的总时间,具体模型表示如下:

$$\min \sum_{p\in P\cup\{U_k\}}\sum_{q\in P}\sum_{k\in K}(z_{p,q,k}\times D_p^*)+\sum_{p\in P\cup\{U_k\}}\sum_{q\in P\cup\{H_k\}}\sum_{k\in K}(z_{p,q,k}\times D'_{p,q}). \tag{18}$$

$$\sum_{p\in P\cup\{U_k\}}\sum_{q\in P_i}\sum_{k\in K} z_{p,q,k}=1,\forall i\in R^+; \tag{19}$$

$$\sum_{p\in P\cup\{U_k\}}\sum_{k\in K} z_{p,q,k}\leqslant 1,\forall q\in P; \tag{20}$$

$$\sum_{p\in P\cup\{H_k\}} z_{U_k,q,k}=1,\forall q\in P,k\in K; \tag{21}$$

$$\sum_{p\in P\cup\{U_k\}} z_{p,H_k,k}=1,\forall q\in P,k\in K; \tag{22}$$

$$\sum_{p\in P\cup\{U_k\}} z_{p,q,k}=\sum_{p\in P\cup\{H_k\}} z_{q,p,k},\forall q\in P,k\in K; \tag{23}$$

$$s_{p,k}+L_p^*+D_p^*+\left(\sum_{q\in P} z_{p,q,k}-1\right)\times M\leqslant e_{p,k},\quad \forall p\in P,k\in K; \tag{24}$$

$$e_{p,k}+L_p^*+D'_{p,q}+(z_{p,q,k}-1)\times M\leqslant s_{q,k},\quad \forall p\in P\cup\{U_k\},q\in P,k\in K; \tag{25}$$

$$S_q^S\times\sum_{p\in P\cup\{U_k\}} z_{p,q,k}\leqslant s_{q,k}\leqslant S_q^E\times\sum_{p\in P\cup\{U_k\}} z_{p,q,k},\quad \forall q\in P,k\in K; \tag{26}$$

$$E_q^S\times\sum_{p\in P\cup\{U_k\}} z_{p,q,k}\leqslant e_{q,k}\leqslant E_q^E\times\sum_{p\in P\cup\{U_k\}} z_{p,q,k},\quad \forall q\in P,k\in K; \tag{27}$$

$$e_{U_k,k}=E_{U_k},\forall k\in K; \tag{28}$$

$$z_{p,q,k}\leqslant Z_{p,q,k},\forall p\in P,q\in P,k\in K. \tag{29}$$

目标函数(18)表示最小化机车行驶的总时间,即机车在模式内行驶时间与模式间行驶时间之和;式(19)限制所有的运送请求都被满足,且

对于每个运送请求仅有一个包含其请求的模式被选择;式(20)表示每个模式最多只能被调用一次;式(21)和式(22)分别限制每辆机车从出发节点驶离,并最后回到停车节点;式(23)表示每个模式驶入与驶离流量平衡;式(24)保证每辆机车在一个模式内行驶到每个节点时间连续;式(25)保证每辆机车在两个模式间行驶时间连续;式(26)和式(27)表示机车行驶到每个模式的开始时间和结束时间满足时间窗限制;式(28)规定每辆机车出发时间;式(29)限制每辆机车只能选择访问铁路网络真实连接和调度逻辑有效的模式.

2　基于强化学习的Benders分解算法

铁水运输机车调度优化为NP难问题,具有硬时间窗、运送类型、装卸顺序和容量限制等复杂约束.相较于模型MIP1,基于模式的机车调度模型能够在预处理阶段生成满足所有复杂约束的模式,通过优化不同模式的组合来求解该问题,极大地简化了求解模型的约束条件.然而,该模型的缺陷在于生成的模式数量随问题规模呈指数级增长,使得变量规模过大,导致大规模算例无法直接由商业求解器求解.鉴于此,本文提出一种基于强化学习的Benders分解算法,通过基于逻辑的Benders分解算法框架分解基于模式的机车调度模型以降低求解难度,并采用强化学习方法动态固定或排除模式决策变量,从而提高求解效率.

基于逻辑的Benders分解是一种将复杂问题分解为主问题和子问题的精确算法框架.本文将基于模式的机车调度模型进行拆分,其中,主问题为模式选择模型,负责选择一组包含所有运送请求的模式集合;子问题为模式受限模型,负责在给定模式集合下优化具体运输路线,验证其可行性并计算总时间.子问题通过向主问题反馈可行割和最优性割,优化主问题的求解空间.值得注意的是,传统Benders分解算法在处理大规模组合优化问题时,往往因搜索空间过大而收敛缓慢.为此,本文引入强化学习Q-learning算法,该学习机制能够根据历史求解反馈,自适应地动态调整待固定或排除的模式集合,使搜索过程聚焦于具有较高潜力的优质解空间区域,从而有效克服解空间的组合爆炸问题,显著提升算法的求解效率.该算法迭代求解主要步骤如下.

1) 初始解生成:采用启发式算法构造初始可行解,为主问题提供上界并实现热启动;

2) 求解主问题:求解主问题模式选择模型,获得当前迭代的一组模式集合;

3) 子问题求解与反馈:基于主问题选定的模式集合,求解子问题模式受限模型.若子问题可行,则生成最优性割并添加至主问题模型;否则,生成可行性割并添加至主问题模型;

4) Q-learning算法:更新状态和Q值,调整主问题中固定选择或不选的模式集合;

5) 迭代与终止:重复步骤2)至步骤4),直至满足终止条件.

2.1　主问题与子问题模型构建

主问题目标是从所有可行的运送模式中选出一个覆盖全部运送请求的模式集合,以最小化机车的总行驶时间t_{t}.为了引导搜索过程、减少低效模式的重复选择,本文在目标函数中引入一项与模式相关的惩罚项R_p,加速算法收敛.同时,引入集合O和C分别表示当前迭代中强制选中或强制排除的模式集合,从而有效缩小主问题的搜索空间.此外,集合S和B分别用于记录Benders分解的可行性割和最优性割.主问题的模式选择模型如下.

$$\min t_{\mathrm{t}}. \tag{30}$$

$$\sum_{p\in P_i} u_p = 1, \forall i \in R^+; \tag{31}$$

$$u_p = 1, \forall P \in O; \tag{32}$$

$$u_p = 0, \forall P \in C; \tag{33}$$

$$\sum_{p\in S_{\mathrm{sub}}} u_p \leqslant |S_{\mathrm{sub}}| - 1, \forall S_{\mathrm{sub}} \subseteq S; \tag{34}$$

$$t_{\mathrm{t}} \geqslant f_{B_{\mathrm{sub}}} - f_{B_{\mathrm{sub}}} \times \left(|B_{\mathrm{sub}}| - \sum_{p\in B_{\mathrm{sub}}} u_p \right), \forall B_{\mathrm{sub}} \subseteq B; \tag{35}$$

$$t_{\mathrm{t}} \geqslant \sum_{p\in P} u_p \times (D_p^* + R_p); \tag{36}$$

$$u_p \in \{0,1\}, \quad \forall p \in P. \tag{37}$$

目标函数(30)表示最小化机车行驶的总时间;式(31)保证每个请求都有一个包含其请求的模式被选择;式(32)表示集合O中的模式被强制选中;式(33)表示集合C的模式被强制排除;式(34)表示可行性割,即集合S_{sub}中的模式不会被同时选中;式(35)表示最优性割,即集合B_{sub}中的模式同时选中则目标函数值至少为子问题求得结果$f_{B_{\mathrm{sub}}}$;式(36)表示机车行驶的总时间不少于所选模式时间之和,其中考虑模式的惩罚值R_p以减少低效模式的选择,加快算法收敛;式(37)定义模式决策变量的取值范围.

子问题是一个限制模式的机车调度模型,根据主问题所选模式的集合P_{sub},进一步优化机车

的请求分配和具体运送路径,验证解的可行性并计算机车总的行驶时间,返回可行性割或最优性割.子问题的限制模式模型为

$$\min \sum_{p\in P_{\text{sub}}\cup\{U_k\}}\sum_{q\in P_{\text{sub}}}\sum_{k\in K}(z_{p,q,k}\times D_p^*)+\sum_{p\in P_{\text{sub}}\cup\{U_k\}}\sum_{q\in P_{\text{sub}}\cup\{H_k\}}\sum_{k\in K}(z_{p,q,k}\times D'_{p,q}). \tag{38}$$

约束条件见式(21)~(29).

2.2 初始解的生成

为提升算法整体求解效率,本文基于贪心策略[23],结合铁水运输机车调度优化问题中的复杂约束条件,设计一种启发式算法,用于生成高质量的初始可行解,并为基于强化学习的Benders分解算法提供热启动.该算法以运送模式为基本单元,通过逐步将各个运送请求插入至机车路径中可行且距离最短的位置,从而构建完整的机车行驶方案.具体步骤如下:

1) 将全部运送请求按其获取鱼雷罐节点时间窗的最早时间从小到大进行排序;

2) 依次处理每一个请求,尝试将其插入当前机车路径中的4个候选位置:①作为一个新模式,插入在最后一个模式之前;②作为一个新模式,插入在最后一个模式之后;③若最后一个模式仅包含单一请求,则将新请求置于该请求之前构成一个模式;④若最后一个模式仅包含单一请求,则将新请求置于该请求之后构成一个模式;

3) 从上述4个位置中筛选所有符合约束的可行位置,从中选择使总行驶时间增量最小的位置插入该请求;

4) 若当前机车无法找到可行插入位置,则转向下一辆机车重复步骤2)和3),直至该请求被成功插入;

5) 重复步骤2)~4),直至所有请求均被分配完毕.

例如,某机车当前访问序列为[(2,3),(1)],待插入请求为4,则步骤2)所检查的模式分别为[(2,3),(4),(1)],[(2,3),(1),(4)],[(2,3),(4,1)]和[(2,3),(1,4)].

2.3 加速策略

式(29)考虑了机车行驶中不同模式在物理空间上的连接关系,在子问题求解过程中,可进一步引入时间约束以减少变量$z_{p,q,k}$的数量.具体而言,若模式p结束的最早时间晚于模式q开始的最晚时间,即$E_p^{\text{S}}>S_q^{\text{E}}$,则在一辆机车的行驶路线中,模式$p$不可能紧邻于模式$q$之前,即$Z_{p,q,k}=0,\forall k\in K$.该策略从时间维度上进一步减少连接变量的数量,从而加快模型求解速度.

另一方面,为减少模式变量的数量,引入节点集合H.假设集合内任意两个节点到集合外任意一个节点的距离相等,即$D_{i,j}=D_{i',j},\forall i\in H,i'\in H,j\in I/H$.本文所考虑的距离具有对称性,因此,集合外任意节点到达集合内任意两个节点的距离相等,即$D_{j,i}=D_{j,i'},\forall i\in H,i'\in H,j\in I/H$.

定理1 若模式p和q由相同的节点构成,且这些节点均属于集合H,则从任意其他模式到模式p和q的行驶时间相等,且从模式p和q到任意其他模式的行驶时间相等,即$D'_{m,p}=D'_{m,q},D'_{p,m}=D'_{q,m},\forall m\in P/\{p,q\}$.

证明 设模式m最后服务的节点为j,模式p和q开始服务的节点分别为i和i',由于$D_{j,i}=D_{j,i'}$,因此$D'_{m,p}=D'_{m,q}$,同理可证$D'_{p,m}=D'_{q,m}$.

定理2 若模式p和q由相同的节点组成,且节点均属于集合H,并满足:$S_p^{\text{E}}\leqslant S_q^{\text{E}},E_p^{\text{S}}\geqslant E_q^{\text{S}},D_p^*\leqslant D_q^*$,则在模式集合中删去模式$q$,不会改变原问题的最优解.

证明 考虑任意包含模式q的可行解.由于模式p与模式q包含相同节点,可满足相同运送请求,且时间窗上$S_p^{\text{E}}\leqslant S_q^{\text{E}},E_p^{\text{S}}\geqslant E_q^{\text{S}}$,因此,将解中模式$p$替代为模式$q$仍满足所有约束,保持解的可行性.根据定理1可得,将解中模式p替代为模式q后模式间行驶时间不变;同时,由于$D_p^*\leqslant D_q^*$,且服务节点相同,模式内的行驶时间与服务时间不会增加,因此,将解中模式p替代为模式q,不会导致机车总行驶时间增加.综上,任意包含模式q的可行解均可被一个不劣于它的包含模式q的解所替代,因此在模式集合中删去模式q,不会改变原问题最优解.

2.4 Q-learning算法

为进一步加速Benders算法收敛进程,本文引入Q-learning算法,通过自适应学习机制动态调整主问题中模式的固定与排除集合.具体而言,Q-learning算法的状态d分为3种:状态$d=1$表示当前子问题更新了主问题上界解;状态$d=2$表示当前子问题可行但没有更新主问题上界;状态$d=3$表示当前子问题无可行解.Q-learning算法的动作集合A包含3种动作:动作$a=1$表示需要向强制选取的模式集合O添加一个模式;动作$a=2$表示需要向强制排除的模式集合C添加一个模式;动作$a=3$表示不需要改变当前集合O和C.Q-learning算法通过不断评估不同状态下固定或排除特定模式所带来的收益,逐步优化模式决

策策略,提高求解效率.Q-learning算法如表2所示.

表2　Q-learning算法

Table 2　Q-learning algorithm

输入:被选的模式集合S_{sub},惩罚值向量$\boldsymbol{R}$,强制选取的模式集合O,强制不选的模式集合C,当前状态d,历史最优解f_{best},当前子问题最优解f

For $p \in S_{\text{sub}}$
　计算模式p的惩罚值r_p;
　If $r_p < 0$ then
　　$R_p \leftarrow R_p - \alpha \times r_p + 1$;
　Else
　　$R_p \leftarrow 0$;
reward $\leftarrow (f_{\text{best}} - f)^+ - \sum_{p \in \mathrm{P}_{\text{sub}}} R_p / |P_{\text{sub}}|$ \\计算奖励值
根据子问题求解情况判断新状态d';
$Q(d,a) \leftarrow Q(d,a) + \alpha \times (\text{reward} + \gamma \times \max\limits_{a' \in A} Q(d',a') - Q(d,a))$; \\更新$Q$值
$\varepsilon \leftarrow \varepsilon \times \beta$;
If random $\geqslant \varepsilon$:
　$a' := \arg\max\limits_{a'' \in A} Q(d',a'')$;
Else
　$a' :=$ randomchoice(A); \\随机选择新动作
根据动作a'更新集合O和C;

输出:$\boldsymbol{R}, O, C$

首先参考Kweon等[24]的方法计算各模式的惩罚值r_p.具体而言,针对当前解中的某一模式p,在已知最优解中找出一个能满足相同运送请求的对应模式,随后计算两者所对应运输方案的总时间之差,并将此时间差定义为该模式的惩罚值,用以更新惩罚值向量$\boldsymbol{R}$.然后,计算当前动作下的奖励值,并更新状态与Q值,其中,学习率α表示对新动作奖励的重视程度,折扣因子γ表示对动作未来奖励的重视程度,衰减因子β控制学习过程的收敛速度.最后,根据ε-贪婪策略决定采取的动作.具体地,当动作$a=1$时,通过轮盘赌选择机制,以惩罚值越小选择概率越大的原则,选择一个模式加入到集合O;动作$a=2$时,通过轮盘赌选择机制,以惩罚值越大选择概率越大的原则,选择一个模式加入到集合C;动作$a=3$时,则不对集合O和C进行更新.随着迭代次数增加,Q值矩阵逐步收敛,算法1能够稳定地识别出高效的模式组合,从而使算法整体在有限迭代内获得高质量的可行解.

3　实验与结果分析

3.1　算例设计

为系统验证所提出模型与算法的有效性及其求解性能,本文基于钢铁厂真实环境分别生成小规模算例与大规模算例,以模拟不同生产强度下的鱼雷罐运送请求调度场景,从而评估本文模型与算法在不同规模问题中的求解效率.每个算例命名为"Ins-|R^+|".

生成算例的机车运送网络使用真实的工厂布局,包含5个高炉、2个转炉和1个缓冲区,各设施间的连接关系与行驶时间见表3.同时,为确保安全生产与连续作业,鱼雷罐车须在出铁开始前至少5 min就位.实罐或空罐在完成服务后须在20 min内被机车取走并运往指定交付节点,机车装卸鱼雷罐的服务时间设为2 min.算例中运送请求的时间窗生成方法参考文献[24]:首先生成高炉与转炉的生产计划,确定运送请求单节点的需求时间窗;再根据各设施计划随机生成运送请求剩余节点的时间窗;最后随机设置机车初始状态,生成机车出发节点时间窗,从而系统构建出符合实际调度逻辑的测试算例.

表3　行驶时间矩阵

Table 3　Travel time matrix

项目	转炉1	转炉2	缓冲区	高炉1	高炉2	高炉3	高炉4	高炉5
转炉1	0	1	15	20	21	22	23	24
转炉2	1	0	15	20	21	22	23	24
缓冲区	15	15	0	15	16	17	18	19
高炉1	20	20	15	0	1	2	3	4
高炉2	21	21	16	1	0	1	2	3
高炉3	22	22	17	2	1	0	1	2
高炉4	23	23	18	3	2	1	0	1
高炉5	24	24	19	4	3	2	1	0

3.2　结果分析

本研究所用计算机硬件配置为AMD Ryzen 9 5900HX 3.30 GHz CPU,32.0 GB内存.数学模型MIP1,MIP2以及Benders分解框架中的主问题和子问题模型均通过商用求解器Gurobi 11.0.0进行求解,算法程序均采用Python编程实现.在实验参数设置中,本文使用Gurobi直接求解数学模型MIP1与MIP2的计算时间上限,以及未引入强化学习的传统Benders分解算法的计算时间上限,统一设定为1 800 s;而基于强化学习的Benders分解算法以连续50次迭代未改善最优解作为终止条件.

为验证所提出模型与算法的有效性,首先选取了5组小规模算例开展实验,其中运送请求数

量从20逐步增加至60.实验结果如表4所示.其中,T_{MIP1}和t_{MIP1}表示直接用Gurobi求解数学模型MIP1的机车行驶总时间和运行时间;T_{MIP2}和t_{MIP2}表示直接用Gurobi求解数学模型MIP2的机车行驶总时间和算法计算时间;T_{LB}和t_{LB}为无强化学习的Benders分解算法求得的机车行驶总时间和算法计算时间;T_{QB}和t_{QB}为基于强化学习的Benders分解算法求得的机车行驶总时间和算法计算时间.

表4 小规模算例实验结果

Table 4 Experimental results of small-scale instances

算例	T_{MIP1}/min	t_{MIP1}/s	T_{MIP2}/min	t_{MIP2}/s	T_{LB}/min	t_{LB}/s	T_{QB}/min	t_{QB}/s
Ins20	299	0.42	299	0.20	299	4.80	299	1.03
Ins30	322	1.33	322	0.66	322	1 800	322	4.51
Ins40	452	47.16	452	3.64	454	1 800	454	8.85
Ins50	404	204.42	404	15.38	404	1 800	412	11.52
Ins60	580	1 800	580	215.17	580	1 800	580	15.33

由小规模算例的实验结果可知,采用Gurobi直接求解两种数学模型均能在规定时间内获得最优解.相比之下,基于模式建模的MIP2模型将复杂约束移至预处理阶段,其计算时间显著少于基于节点建模的MIP1模型,且随着请求数量增加,两者在求解效率上的差距更为明显,验证了基于模式构建的机车调度优化模型的优势.无强化学习的Benders分解算法的收敛速度低于Gurobi直接求解模型,在算例Ins40上未获得最优解,且在4个算例上未能在限定时间内完成最优性证明.本文提出的基于强化学习的Benders分解算法在算例Ins20,Ins30和Ins60中获得了经Gurobi验证的最优解,从而验证了该启发式算法的有效性.

本文进一步构建了一组大规模算例,请求数量范围为100至140个,旨在对比分析本文所提算法在更高计算复杂度下的求解性能.大规模算例实验结果如表5所示.

表5 大规模算例实验结果

Table 5 Experimental results of large-scale instances

算例	T_{MIP1}	t_{MIP1}/s	T_{MIP2}/min	t_{MIP2}/s	T_{LB}/min	t_{LB}/s	T_{QB}/min	t_{QB}/s
Ins100	—	1 800	1 088	1 800	1 081	1 800	1 081	76.23
Ins110	—	1 800	—	1 800	1 094	1 800	1 087	114.02
Ins120	—	1 800	—	1 800	1 241	1 800	1 229	165.29
Ins130	—	1 800	—	1 800	1 199	1 800	1 178	181.56
Ins140	—	1 800	—	1 800	1 454	1 800	1 434	204.54

实验结果表明,在表5所列的5个大规模算例中,采用Gurobi直接求解数学模型在多数情况下难以在有限时间内获得可行解.相比之下,无强化学习的Benders分解算法在所有算例中均能在相同限制时间内得到可行解.本文提出的基于强化学习的Benders分解算法在所有大规模算例中均取得了更优的结果,同时在计算时间方面也表现良好,与无强化学习的Benders分解算法相比,体现出本文引入Q-learning算法的改进效果.综上,本文算法在处理大规模算例时,在求解质量与计算时间方面均具有优势.

4 结 语

本文针对铁水运输过程中的机车调度问题展开系统研究.首先,综合考虑机车节点到达的硬时间窗约束、鱼雷罐“后进先出”的装卸顺序规则,以及实罐与空罐不能同载的类型限制等铁水运输典型的生产特性,以最小化机车总行驶时间为优化目标,分别构建了基于节点和基于模式的两种混合整数规划模型.其次,为求解机车调度优化问题,设计了一种融合强化学习Q-learning算法与Benders分解算法的启发式算法框架:通过将原问题分解为模式选择主问题和模式受限子问题,有效降低了模型求解的复杂度;进一步结合贪心策略设计了基于模式的启发式算法以快速生成高质量的初始解;并且引入Q-learning算法的自适应学习机制,动态调整主问题中模式决策变量选择机制,引导算法在具有高效模式的解空间探索,提高算法搜索效率.最后,通过两组

不同规模算例的实验结果验证了所提出模型与算法的有效性,并通过对比分析阐明了基于强化学习的Benders分解算法在求解质量与时间上的优势.

本文所提出的算法能够为钢铁企业铁水运输系统提供高效、可行的机车调度方案,有助于企业优化厂内物流、提升资源利用效率并降低运营成本.但本文亦存在一定的局限性,如问题模型基于固定生产计划的假设,未充分考虑生产现场的动态扰动因素.未来研究工作将探索设备突发故障、生产计划临时调整等不确定环境下的鲁棒优化问题,以增强铁水运输系统的抗干扰能力,推动钢铁行业向数智化方向迈进.

参考文献:

[1] 柴立元, 王云燕, 孙竹梅, 等. 绿色冶金创新发展战略研究[J]. 中国工程科学, 2022, 24(2): 10-21.
(Chai Li-yuan, Wang Yun-yan, Sun Zhu-mei, et al. Innovative development strategy of green metallurgy [J]. *Strategic Study of CAE*, 2022, 24(2): 10-21.)

[2] Moradi A A, Askari-Nasab H. Mining fleet management systems: a review of models and algorithms [J]. *International Journal of Mining, Reclamation and Environment*, 2019, 33(1): 42-60.

[3] 黄肖玲, 任宇婷, 张佳安, 等. 基于低碳环保因素的"前港后厂"钢铁产成品运输网络优化[J]. 运筹与管理, 2021, 30(8): 59-66.
(Huang Xiao-ling, Ren Yu-ting, Zhang Jia-an, et al. Optimizing transportation network of steel products under "port before factory" based on low-carbon environmental protection factor[J]. *Operations Research and Management Science*, 2021, 30(8): 59-66.)

[4] Sun D F, Meng Y, Tang L X, et al. Storage space allocation problem at inland bulk material stockyard [J]. *Transportation Research Part E: Logistics and Transportation Review*, 2020, 134: 101856.

[5] Jiang Z Z, Liu J B, Yi H L, et al. A multi-objective evolutionary algorithm based on incremental support vector regression for the irregular strip packing problem[J]. *Annals of Operations Research*, 2026,359(2):2149-2187.

[6] Cherkesly M, Desaulniers G, Irnich S, et al. Branch-price-and-cut algorithms for the pickup and delivery problem with time windows and multiple stacks[J]. *European Journal of Operational Research*, 2016, 250(3): 782-793.

[7] Zhu Z Q, Chen Y R, Wahab M I M. An exact algorithm for simultaneous pickup and delivery problem with split demand and time windows [J]. *Computers & Operations Research*, 2024, 170: 106761.

[8] Luo H Y, Ma T, Li Z D. An exact branch-price-and-cut algorithm for the time-dependent cold chain pickup and delivery problem with incompatibility constraints [J]. *Computers & Operations Research*, 2025, 178: 107007.

[9] Cavaliere F, Accorsi L, Laganà D, et al. An efficient heuristic for very large-scale vehicle routing problems with simultaneous pickup and delivery [J]. *Transportation Research Part E: Logistics and Transportation Review*, 2024, 186: 103550.

[10] Gao Z, Deng F Q, Fu Z H, et al. A problem reduction based memetic algorithm for the vehicle routing problem with discrete split deliveries and pickups[J]. *Computers & Operations Research*, 2025, 182: 107106.

[11] Alyasiry A M, Forbes M, Bulmer M. An exact algorithm for the pickup and delivery problem with time windows and last-in-first-out loading[J]. *Transportation Science*, 2019, 53(6): 1695-1705.

[12] Aziez I, Côté J F, Coelho L C. Exact algorithms for the multi-pickup and delivery problem with time windows[J]. *European Journal of Operational Research*, 2020, 284(3): 906-919.

[13] Pilati F, Tronconi R. Multi-objective optimization for sustainable few-to-many pickup and delivery vehicle routing problem[J]. *International Journal of Production Research*, 2024, 62(9): 3146-3175.

[14] Jiang J, Dai Y, Yang F, et al. A multi-visit flexible-docking vehicle routing problem with drones for simultaneous pickup and delivery services [J]. *European Journal of Operational Research*, 2024, 312(1): 125-137.

[15] Habibi M K, Hammami R, Battaia O, et al. Simultaneous pickup-and-delivery production-routing problem in closed-loop supply chain with remanufacturing and disassembly consideration [J]. *International Journal of Production Economics*, 2024, 273: 109290.

[16] Li J, Cang L, Wu Y S, et al. Two-echelon collaborative many-to-many pickup and delivery problem for agricultural wholesale markets with workload balance [J]. *Omega*, 2025, 130: 103164.

[17] Zhang Z Z, Che Y X, Liang Z. Split-demand multi-trip vehicle routing problem with simultaneous pickup and delivery in airport baggage transit[J]. *European Journal of Operational Research*, 2024, 312(3): 996-1010.

[18] Lübbecke M E, Zimmermann U T. Engine routing and scheduling at industrial in-plant railroads[J]. *Transportation Science*, 2003, 37(2): 183-197.

[19] Wang G S, Tang L X. A column generation for locomotive scheduling problem in molten iron transportation[C]//2007 IEEE International Conference on Automation and Logistics. Jinan, 2007: 2227-2233.

[20] Tang L X, Wang G S, Liu J Y. A branch-and-price algorithm to solve the molten iron allocation problem in iron and steel industry[J]. *Computers & Operations Research*, 2007, 34(10): 3001-3015.

[21] Cherkesly M, Desaulniers G, Laporte G. Branch-price-and-cut algorithms for the pickup and delivery problem with time windows and last-in-first-out loading [J]. *Transportation Science*, 2015, 49(4): 752-766.

[22] Huang B B, Tang L X, Baldacci R, et al. A metaheuristic algorithm for a locomotive routing problem arising in the steel industry [J]. *European Journal of Operational Research*, 2023, 308(1): 385-399.

[23] Eom M, Kim B I. Combinatorial Benders decomposition for melted material blending systems considering transportation and scheduling[J]. *International Journal of Production Research*, 2023, 61(10): 3481-3503.

[24] Kweon O, Kim B I. Many-to-many locomotive routing problem for the steel industry[J]. *International Journal of Production Research*, 2024, 62(23): 8373-8396.

[25] Martínez K P, Adulyasak Y, Jans R. Logic-based benders decomposition for integrated process configuration and production planning problems[J]. *INFORMS Journal on Computing*, 2022, 34(4): 2177-2191.

[26] Karimi-Mamaghan M, Mohammadi M, Pasdeloup B, et al. Learning to select operators in meta-heuristics: an integration of Q-learning into the iterated greedy algorithm for the permutation flowshop scheduling problem [J]. *European Journal of Operational Research*, 2023, 304(3): 1296-1330.