钢铁库区行车智能调度研究综述

[1]张博钰,程银亮,彭功状,刘洋,张学军.钢铁库区行车智能调度研究综述[J].冶金自动化,2022,46(06):48-56.

1 介绍

行车是钢厂库区的核心运输装备，是工序间物流衔接的载体。库区需要在满足上下游生产时间约束和物流空间约束的条件下，合理调度多部行车执行物料出入库、库内倒垛等吊运任务。

行车智能调度的实际应用也面临如下挑战：一是如何高效求解上述多机多任务行车调度问题，满足实际生产需求；二是如何对库区动态扰动进行快速响应以保证调度计划高效稳定 的执行。

本文从不**同库区类型、库区行车调度目标和约束、模型求解方法**以及**集成化调度问题**等维度进行分析行车智能调度研究。

2 钢铁库区分类

2.1 炼钢车间行车调度

李稷、林时敬等[2-3]针对炼钢车间行车调度问题设计了**不同温度策略**下的任务产生方案，并制定不同优先级条件下的天车运行方案，最后开发了完全反映调度策略的钢包调度系统。

2.2 板坯库的行车调度

王旭等[4-5]研究了具有平板非交叉约束的多区间调度问题，提出让全部行车的**整体行驶距离最小**并确保各个行车之间**负载均衡**的问题，并给出相应的**分支定价算法**来解决该问题。

2.3 钢卷库的天车调度

谢谢等[6]在**钢卷库的多行车**调度问题中，在考虑行车空间约束的基础上，建立了**混合整数线性规划模型**，针对问题特征设计了解空间下限

Gabrielan Maschietto等[7]在关于**钢卷配送中心无干扰约束**行车调度问题研究中，为使钢卷仓库卡车运输钢卷的总时间最少，开发了**部分运动规划蝙蝠算法**

2.4 棒线材库的行车调度

彭功状等[8]在**棒材成品库无人行车与货车调度协同优化**的问题中，以订单完工总时间最少为目标，建立了**整数规划模型**，并对不同订单规模下**遗传算法、经验规则和自适应遗传算法**进行了对比试验。

3 行车调度目标分类

3.1 任务完成时间最短

任务完成时间一般用完工时间来表示，其中入库工单的完工时间指的是物料从下线点进入库区到达垛位的时间，出库工单的完工时间指的是物料吊离库区达到货车或货车垛位的时间。

赵国栋等[9]针对热轧厂板坯库可**同时吊运两块板坯**的行车调度问题展开了研究，将**行车调度时间**作为优化总目标建立了**混合整数规划模型**。

雷兆明等[10]在**钢铁企业同轨多行车调度方法**研究中，以**全部调度总时间**最少为目标，采用**改进布谷鸟算法**（improved cuckoo search）设计可传递优先级随变规则，通过加入**动态概率、局部最优解变异机制和自适应动态步长**，提升了整体调度效率。

3.2 行车运行路径最短

行车运行路径的缩短对于提高工作效率和节省运营成本均有较大作用。

王旭等[5]对板坯库的行车分配进行了研究，为使板坯库内**所有行车的总行程最短**，以**库内天车总行程**为目标建立了相关的数学模型，并提出了**分支定价算法**处理多任务间的调度问题，最后结合某钢厂实际数据验证了方法的可行性。

3.3 行车负载均衡率最高

负载均衡率定义了多部行车之间利用率的差异，负载均衡率值越大，表示多部行车在任务分配方面越不均衡。

李稷等[11]在炼钢车间**多行车动态调度问题**中考虑了行车工作量之间差异， 提出一种**滚动调度策略**下的仿真调度方案，结果表明该方案能使各工序平稳运行，并有效减少各行车工作量，极大提高行车负载均衡率，大大提升了性能。

3.4 库区倒垛率最低

倒垛定义为库区物料在入库时将入库垛位上层物料移走或出库时将待出库物料上层物料移走的操作。减少倒垛率能有效提高物料出入库效率和天车运行时间

Byung-In K等[12]研究了钢坯在出入库顺序给定情况下的入库堆垛问题，以**出库过程中倒垛量最小**为目标，将**可能出现的大规模钢坯入库问题**划分为多个可以**使用数学模型进行求解的子问题**

4 行车调度约束分类

4.1 时间约束

1. 行车吊运需要满足上下游生产以及发运任务的时间要求（如钢水在各个工序中加工的开始时间和结束时间）
2. 同一跨相邻天车之间需要考虑运行轨迹的交叉问题

王旭等[13]针对行车行驶过程中受到的**时间空间约束**，设计了基于任务分配规则的Memetic算法， 通过对某钢厂真实数据的仿真测试证明了**该算法的稳定性和收敛性**

4.2 空间约束

交叉约束指的是沿跨长方向的多台天车不能同时在同一工位工作且多台天车的空间位置和顺序不可改变

Tanizaki T等[14]将**行车调度问题**抽象转换为**多阶段job-stop问题**，构建了考虑**不可交叉基础**上行车调度的**整数规划模型**，基于所设计的**结合深度与宽度**的**启发式搜索算法**，进行了**多模式的搜索求解**并通过大量案例证明了该方案的可行性。

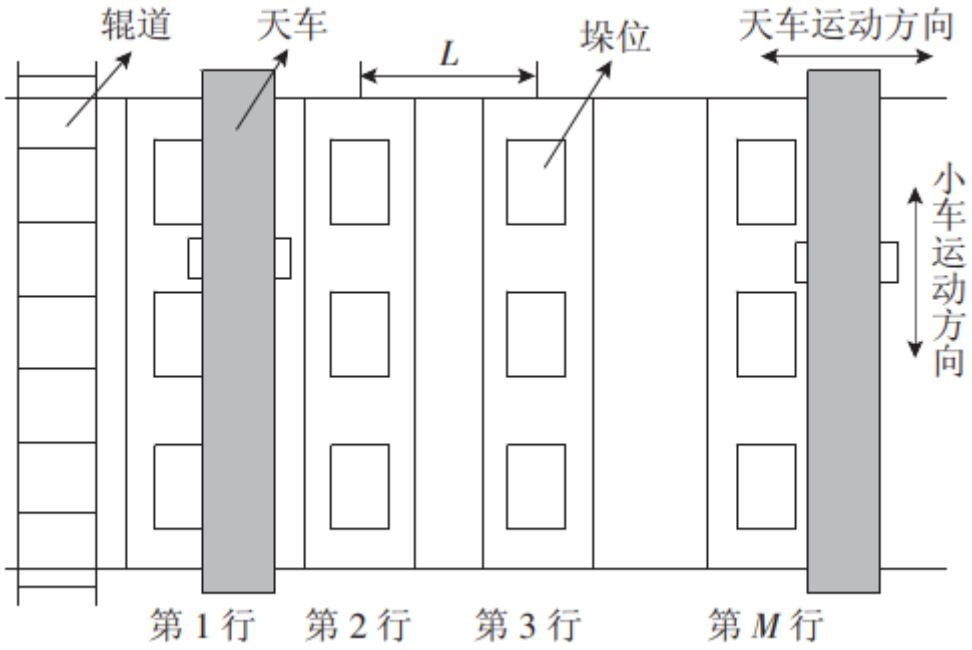

图1 库区行车调度的空间约束示意图

4.3 工艺和操作约束

库区内行车主要任务包括入库、出库、倒垛等，根据库区内生产设备的平面布局和工艺流程的车间管理模式等实际情况，行车的运行需要满足工艺和操作约束。

在炼钢车间，精炼跨行车的任务包括从转炉工位到精炼工位的运输， 以及从精炼工位之间的运输作业；钢水接收跨行车的任务包括从精炼到连铸的作业、从连铸到倒渣的作业、从倒渣到热修的作业以及从热修到转 炉的作业。

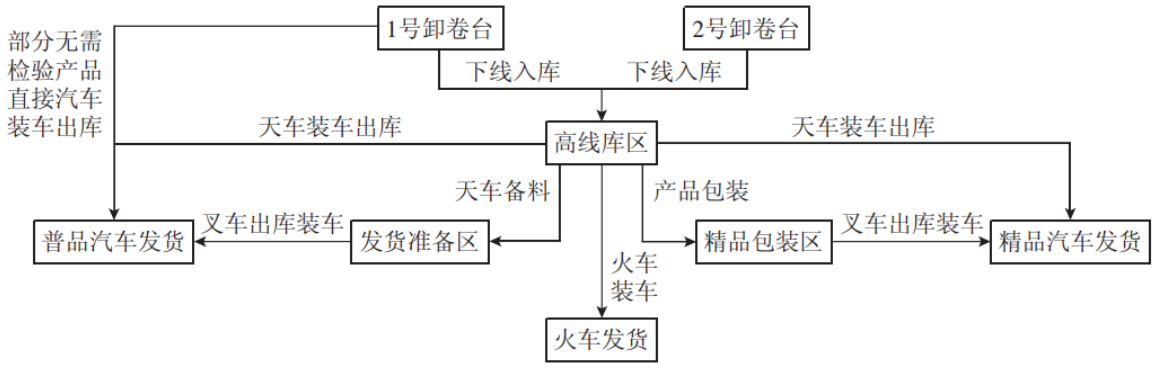

图2 某钢厂高线成品库物流示意图

5 调度模型求解方法

5.1 整数规划方法

整数规划方法作为一种精确算法，在求解问题的最优解方面有很大的优势，用于行车调度问题的整数规划求解算法包括**分支定界法、分支定价法**等。

程旭[15]将**分支定界方法**应用于物流库行车调度问题中，提出了**利用节点处增加有效不等式**来动态提升下界，达到提高算法效率等策略。

王旭等[5]使用**分支定价算法**来求解板坯库的多行车调度模型，通过将**列生成算法求解松弛模型**的最优解，提升了算法效率和整体性能。

5.2 智能优化算法

由于行车调度问题变量和约束规模较大，一般为NP-hard问题，精确算法的求解效率和时间受到很大限制。与传统的精确求解优化方法相比，智能优化算法不以达到某个最优性条件或找到理论上的精确最优解为目标，而是更看重计算的速度和效率，对目标函数和约束函数的要求相对宽松。

郑忠等[16]针对炼钢车间的**多行车多任务调度问题**，采用**遗传算法**求解基于作业工位和行车位置的**任务优先级关系**，有效解决了行车调度作业时的**时空约束问题**。

Sammarra M等[17]将**多岸桥调度问题**划分为**岸桥路径规划和岸桥调度**两部分进行研究，以**最小化卸船完成时间**为目标建立混合整数规划模型，采用**禁忌搜索算法**求解岸桥路径规划问题，**局域搜索算法**求解岸桥调度问题，算法效率优于贪婪自适应搜索算法。

5.3 强化学习方法

强化学习算法通过多智能体与环境的不断交互提高系统自洽性，在应对复杂多变的实际生产环境中具有较大优势，因而近几年被广泛应用于调度领域。

王博[18]针对板坯库内运输利用率低、多种运输方式、入库板坯动态到达等问题，设计出**动态板坯空间定位决策方法**，提出了**基于模型的强化学习算法**对问题进行求解，并开发出动态板坯空间位置智能决策系统。

林时敬等[3]建立了基于**深度强化学习的炼钢车间天车调度架构**，采用**深度Q学习算法(DQN)**算法搭建动作价值网络模型，通过选择不同的动作策略使得整体奖励函数最大化，实现天车智能体与环境的感知交互，为炼钢车间行车调度问题的解决提供了新思路。

5.4 基于仿真的求解方法

基于仿真的方法是通过建立和运行实际系统的仿真模型来模仿系统的运行状态和规律，以实现在计算机上进行试验的全过程。

在车间中炼钢 - 连铸过程天车调度研究中，李霄峰等[19]考虑**行车调度时间约束**，基于**排队论方法**建立了一种较为通用的行车调度仿真模型。

赵宁等[20]针对板坯库行车作业具有随机性的问题，建立了**行车调度仿真模型**，采用基于**Agent技术**的循环仿真方法对模型进行仿真优化。

整数规划方法可以**获得模型的全局最优解**，但是在问题**规模增大时求解时间指数延**长；智能优化算法在求解效率上有较大优势，但是**只能获得局部最优解**；强化学习方法可以通过智能体与环境的交互学习动态响应不同扰动，但是它**需要历史调度数据的训练**；基于仿真的调度方法不需要具体解析模型，但是**仿真模型的建立需要花费大量时间和费用**，且仿真模型难以准确反映实际问题。

6 集成化调度问题

6.1 行车调度与垛位优化结合

倒垛是板坯库内常见的问题，垛位优化是在所有的板坯空间位置移动中降低板坯库移动次数，提高物流效率。

董广静等[21]关于**板坯入库堆垛问题**，构建了码垛位置的优化模型，首先对**坯料分类**并构造垛位邻域，然后采用**约束满足算法**对垛位进行优化和分配，达到**行车调度和垛位优化**结合的目的。

张琦琪等[22]为解决板坯入库的问题，以**板坯综合匹配度**、**垛位利用率**和**库存符合平衡度**为目标函数，同时考虑到**垛高限制**、分散性限制以及**出库次序限制**，将**行车调度与垛位优化相结合**，设计了一种**多目标种群与协同粒子群结合**的方法求解该问题。

6.2 行车调度与货车调度结合

可用的天车资源决定了货车停车位的分配，天车调度也受货车分配的影响，为此，统筹调度天车和货车资源变得至关重要，该问题与港口泊位和岸桥的协同调度问题有相近之处。

刘畅[23]在对钢铁企业的**原料入库调度优化**的研究中，建立了以**货车卸货的时间**和**库房的周转时间**为基础的数学模型，并在此基础上单独构建了**具有车辆缓冲区的原料入库调度模式**，设计了**协同引擎搜索算法**和**主从两级引力搜索算法**对该问题进行求解。

6.3 行车调度与生产计划结合

在钢铁物流系统中，生产与调度密不可分、相互耦合，因此在库区行车调度中合理考虑生产计划对于提高调度系统的鲁棒性具有重要作用。

彭功状等[4]在板坯库行车调度问题中同时考虑**倒垛次数和行车运行总路径**，将**发货计划和出坯时间**等生产因素与库区物流因素相结合，建立多优化目标的整数规划模型，并采用**NSGA-Ⅲ算法**对模型进行优化求解，通过实际钢厂数据的试验结果验证了算法求解该类问题的高效性。

7 应用案例 - 无人高线库行车调度系统

本团队在国内某钢厂实现的无人高线库行车调度系统主要由**库区感知、定位控制、库区调度、智慧集控**4个主要部分组成。该系统主要使能技术包括**库区环境感知与三维重构、机器视觉与天车控制深度融合、基于多智能体的库管调度、基于5G与数字孪生的库区集控**等

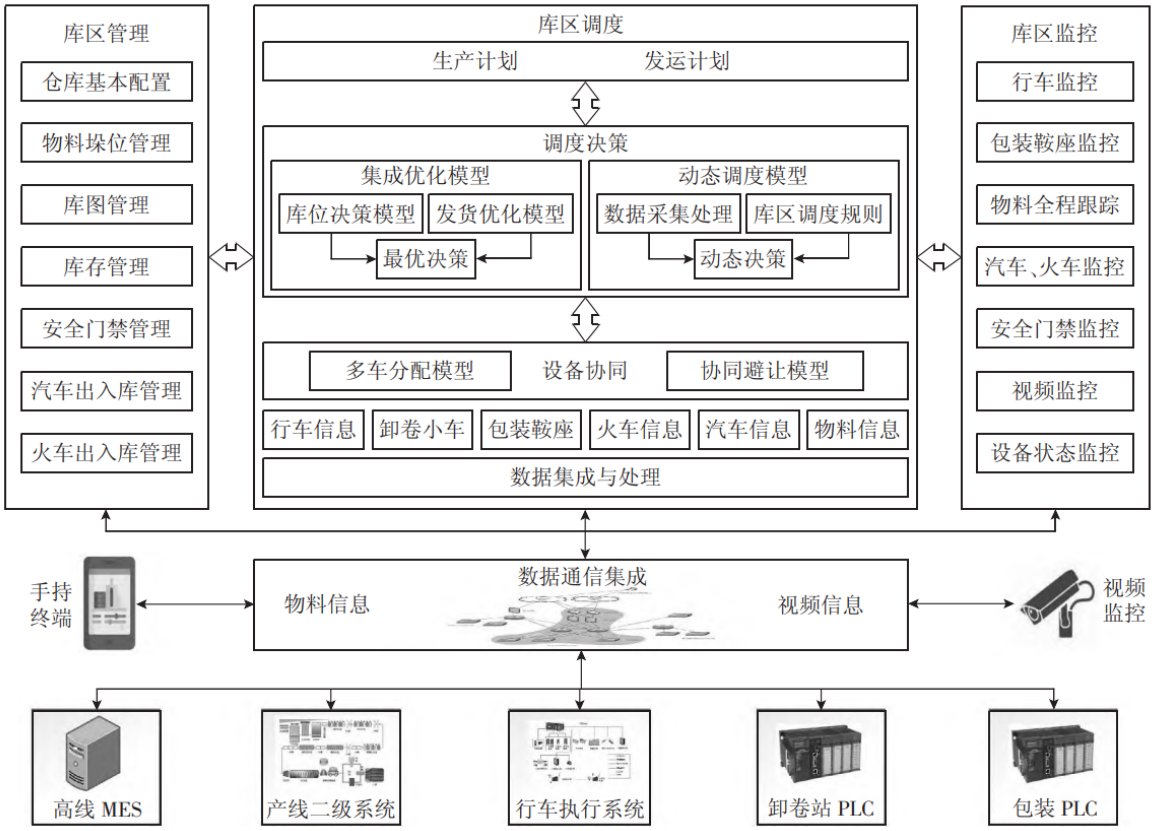

图3 高线库行车调度系统整体架构

其中的库管调度可分为计划层、任务层和运行层。

计划层首先掌握总体设备状态、任务及库存物料信息，通过优化算法选定最优储位，把行 车、卸卷小车的作业任务下发到任务层。

任务层根据工作要求，实现线卷逻辑位置向物理地址转移，把行车复杂的操作分解成易于实施的简单操作，并具体完成行车和多设备之间的协调工作。

无人行车控制系统和无线通信网络为运行层。运行层负责完成行车的每个动作，并对行车进行主钩微摇摆操控、加减速操控等操控，以及对大车、小车的定位精准操控、夹具高度操控，从而达到行车的无人驾驶。

8 未来研究重点

8.1 行车动态调度方法

时间类扰动是钢铁生产过程中最常见的扰动类型，如连铸加工时间不确定；设备类扰动主要指行车设备的故障；任务类扰动包括新工单插入、紧急工单以及工单取消等。

因此，如何对库区的动态扰动进行快速响应以保证调度计划高效稳定的执行，成为板坯库无人行车调度中的后续研究方向之一。

8.2 基于数字孪生的行车调度研究

由于钢厂实际库区复杂多变，动态影响因素众多，因此行车调度模型应用场景也将变得十分复杂。 数字孪生技术通过仿真系统与实际车间的不断交互和反馈，对不同场景下的行车调度模型及算法进行改进和完善。 通过生产数据实时驱动3D场景，数据可视化融合操控中心的数据或信息，并将生产操作人员、工艺技术人员与管理人员关心的信息及时、准确、一目了然地呈现给相应人员，以提高库区执行、工艺分析与管理决策的实施效率。

8.3 生产物流一体化调度研究

现有的钢厂库区行车调度，尤其是热轧板坯库的行车调度，与上下游工序生产计划的集成不够，板坯出入库垛位决策的不合理会影响下游热轧生产的效率。 因此，考虑生产物流一体化条件下天车的调度具有重要意 义。

8.4 数据知识融合的行车调度高效求解算法

行车动态调度是典型的**NP-hard**问题，具有**非线性、多目标、多约束、解空间庞大**等特点，导致建模和求解困难。在动态调度问题求解方法上，现有的完全反应式调度方法**缺少对历史数据和调度经验的挖掘**；而现有鲁棒调度方法以智能 优化算法及其混合算法为主，缺少对问题特征分析并在此基础上进行算法设计的研究，在**大规模问题求解上缺少精确算法**的研究。 因此，有必要**融合工艺机理与工业数据对库区动态扰动**进行预测，并结合问题需求和特征提出启发式规则和分支策略实现天车调度问题的高效求解。

[1]张博钰,程银亮,彭功状,刘洋,张学军.钢铁库区行车智能调度研究综述[J].冶金自动化,2022,46(06):48-56.

[2]李稷. 炼钢-连铸区段天车调度系统研究[D].北京科技大学,2020.DOI:10.26945/d.cnki.gbjku.2020.000326

[3]林时敬,徐安军,刘成,冯凯,李稷.基于深度强化学习的炼钢车间天车调度方法[J].中国冶金,2021,31(03):37-43.DOI:10.13228/j.boyuan.issn1006-9356.20200402.

[4]Peng Gongzhuang,Wu Youqi,Zhang Chunjiang,Shen Weiming. Integrated optimization of storage location assignment and crane scheduling in an unmanned slab yard[J]. Computers & Industrial Engineering,2021,161.

[5]Xu Wang,MengChu Zhou,Qiuhong Zhao,Shixin Liu,Xiwang Guo,Liang Qi. A Branch and Price Algorithm for Crane Assignment and Scheduling in Slab Yard[J]. IEEE Transactions on Automation Science and Engineering,2020,PP(99).

[6]Xie Xie,Yongyue Zheng,Yanping Li. Multi-crane scheduling in steel coil warehouse[J]. Expert Systems With Applications,2014,41(6).

[7]Gabriela N. Maschietto,Yassine Ouazene,Martín G. Ravetti,Maurício C. de Souza,Farouk Yalaoui. Crane scheduling problem with non-interference constraints in a steel coil distribution centre[J]. International Journal of Production Research,2016,55(6).

[8]彭功状,程银亮,梁越永,何安瑞.轧钢成品库无人天车与货车调度协同优化[J].钢铁,2021,56(09):36-42.DOI:10.13228/j.boyuan.issn0449-749x.20210097

[9]Zhao, Guodong, Jiyin Liu, and Yun Dong. "Scheduling the operations of a double-load crane in slab yards." *International Journal of Production Research* 58.9 (2020): 2647-2657.

[10]雷兆明,王鹏程,廖文喆,赵凡.钢铁企业同轨多天车调度方法研究[J].计算机仿真,2019,36(06):465-470.

[11]李稷,徐安军.炼钢车间多天车动态调度仿真方案[J].东北大学学报(自然科学版),2020,41(12):1699-1707.

[12]Kim, Byung-In, Jeongin Koo, and Hotkar Parshuram Sambhajirao. "A simplified steel plate stacking problem." *International Journal of Production Research* 49.17 (2011): 5133-5151.

[13]王旭. 考虑时空约束的吊机优化调度模型与启发式算法[D].东北大学,2017.

[14]Tanizaki, Takashi, et al. "A Heuristic Scheduling Algorithm for Steel Making Process with Crane Handling (< Special Issue> Advanced Planning and Scheduling for Supply Chain Management)." *Journal of the Operations Research Society of Japan* 49.3 (2006): 188-201.

[15]程旭. Branch-and-Cut方法及其在物流时空调度中的应用研究[D].东北大学,2015.

[16]郑忠,周超,陈开.基于免疫遗传算法的车间天车调度仿真模型[J].系统工程理论与实践,2013,33(01):223-229.

[17]Sammarra, Marcello, et al. "A tabu search heuristic for the quay crane scheduling problem." *Journal of Scheduling* 10.4 (2007): 327-336.

[18]王博. 基于强化学习算法求解动态板坯空间位置决策问题[D].东北大学,2019.DOI:10.27007/d.cnki.gdbeu.2019.000754.

[19]李霄峰,徐立云,邵惠鹤,任德祥.柔性炼钢连铸仿真调度系统及其关键技术[J].系统仿真学报,2002(02):207-210+252.

[20]赵宁,杜彦华,董绍华,李亮.基于循环仿真的钢铁板坯库天车作业优化[J].系统工程理论与实践,2012,32(12):2825-2830.

[21]董广静,李铁克,王柏琳,柏亮.管坯入库堆垛问题的模型及算法研究[J].工业工程与管理,2013,18(06):32-39.DOI:10.19495/j.cnki.1007-5429.2013.06.006.

[22]刘畅. 钢铁企业物流原料入库调度优化[D].河北工业大学,2016.

基于深度强化学习的炼钢车间天车调度方法

[1]林时敬,徐安军,刘成,冯凯,李稷.基于深度强化学习的炼钢车间天车调度方法[J].中国冶金,2021,31(03):37-43.DOI:10.13228/j.boyuan.issn1006-9356.20200402.

1 介绍

针对炼钢车间天车任务产生的动态不确定性，提出了基于深度强化学习算法的炼钢车间天车调度方法。

首先，基于强化学习将天车调度问题转化为对天车操作动作序列的求解，采用DPN(Deep Q-network)算法构建**动作价值网络模型**进行求解；

然后，以**某钢厂出钢跨天车调度**为研究对象，以**任务完成总时间最短为目标**，介绍了**基于深度强化学习的天车调度方法**的具体设计；

最后，采用**实际数据**对天车动作价值网络模型进行训练，与目前现场广泛使用的**基于固定分区的天车调度方案**进行仿真试验对比。

天车和台车是主要的运输工具。天车调度[1]主要指导天车运输任务的分配，解决天车分配不合理以及天车间的相互冲突，从而提高天车的运输效率，减少钢水热量损失，稳定钢水温度，对生产工序间物流衔接、顺行以生产节奏的调控具有重要意义。天车作为工序间重要的**物流运输工具**，却主要依靠人工经验进行调度，存在调度优化性差、物流运输效率低等问题。对单体工序控制效果和炼钢过程运行效率产生诸多不利影响。

不同钢铁企业和车间跨由于设备布局与生产工艺的差异，导致天车调度模型方法需要各自单独研究，而**缺少对通用性调度方法的研究**，这不利于天车调度智能化技术快速应用和大规模推广。因此，研究一种高效通用性的天车调度方法对**提高钢厂物流运行效率和实现智能化转型**具有重要意义。

钢厂天车调度的研究主要集中在仓库车间的天车调度针对炼钢车间的天车调度研究较少，本文主要采用的方法是深度强化学习。

1.1 深度强化学习

强化学习是通过智能体与未知环境的交互试错，学会如何在一个动态环境里做出最优决策的过程，如图 1 所示。智能体在t时刻根据观察到的环境状态s_t和策略$\pi(a|s)$，决策选择动作a_t作用于环境，得到环境的下一个状态s_{t+1}和立即奖励r_t。目标是找到一个最优策略$\pi^*(a|s)$，使得结束时刻累计奖励值最高。

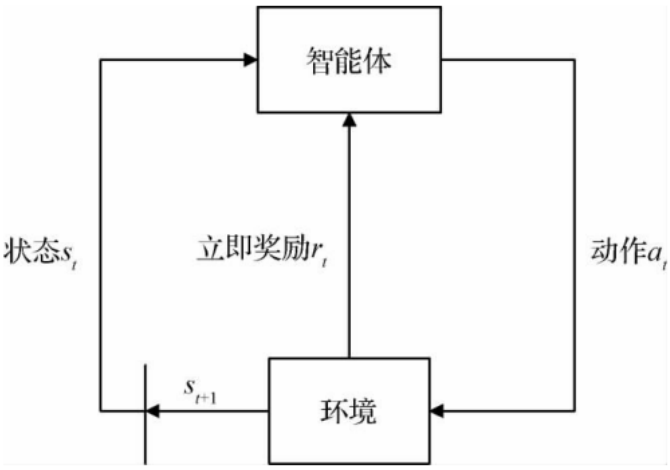

图1 强化学习基本原理

Q-learning是求解最优策略的经典强化学习算法，但其采用Q表格形式存储表示动作价值函数$Q^{\pi}(s,a)$，其刻画了在状态s下，选择动作a的价值大小进行迭代求解，因而无法应用于状态和动作空间复杂的问题。针对该问题，深度强化学习中DQN算法，采用带参数θ的深度神经网络$Q^{\pi}(s,a|\theta)$拟合动作价值函数，同时引入了经验回放和固定目标Q网络两个机制，稳定了网络的训练过程。

1.2 基于深度强化学习的天车调度方法

天车调度是将天车任务合理分配给天车，最终目的是让天车及时高效地完成任务。

天车任务的定义是：将钢包从起点工位运输至重点工位。

天车完成任务的过程：天车移动至起点工位吊起钢包，再移动至终点工位放下钢包。

一个天车任务的完成过程实际为多个时序操作动作的组合序列。求解天车调度方案便可转化为多个时序操作动作的组合序列。求解天车调度方案便可转化为求解每台天车的时序操作动作序列，强化学习是求解此类连贯性序列决策问题的有效方法。

如图2所示，将每台天车抽象成单个智能体，天车所在的车间跨抽象为环境，智能体的动作即为天车的操作动作，智能体观测到的状态即为任务信息和跨内所有天车的状态信息。通过目标函数设计奖励函数，引导天车高效完成天车任务。

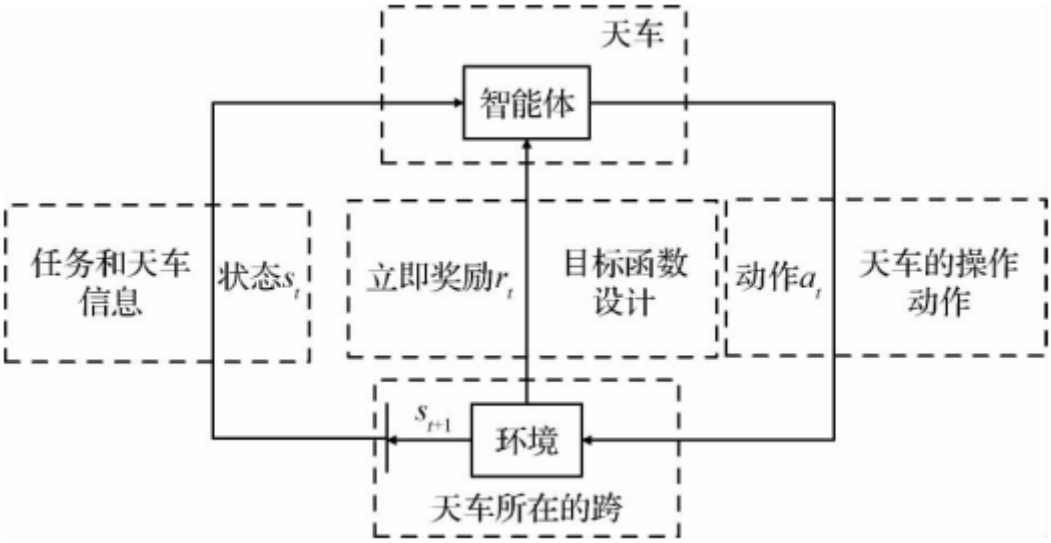

图2 基于强化学习的天车调度机制

由于天车单步动作不仅收到当前动作影响，还会收到之前时刻累计的状态和奖励影响。因此得到以下公式。

$$Q\left(s_t,a_t\right)\leftarrow Q\left(s_t,a_t\right)+\alpha\cdot\left[r_t+\gamma\max_{\pi}Q\left(s_{t+1},a_t\right)-Q\left(s_t,a_t\right)\right]\quad(1)$$

如果当状态空间过大时候，会产生维度灾难以及存储和检索困难，需要使用DQN进行近似值处理。因此整个过程的核心变为如何确定θ来近似值函数，最经典的做法就是采用梯度下降最小化损失函数来不断的调试网络权重θ，Loss function定义为：

$$L_i\left(\theta_i\right)=E_{(s,a,r,s^i)\sim U(D)}\left[\left(r+\gamma\max_{a'}Q\left(s',a^i;\theta_i^-\right)-Q\left(s,a;\theta_i\right)\right)^2\right]\quad(2)$$

2 天车调度方法具体设计

该钢厂出钢跨布局如图3所示，跨全长300m、共有2座转炉、2座双工位ＲＨ 精炼炉、 1座双工位LF精炼炉、2座CAS精炼炉、2台连铸机、2个烘烤位、2个热修位、1个翻渣位和2台天车。天车调度方法的核心是训练天车的Q网络模型，需要对状态空间，动作空间，奖励函数、Q网络结构等基本要素和训练算法进行设计。

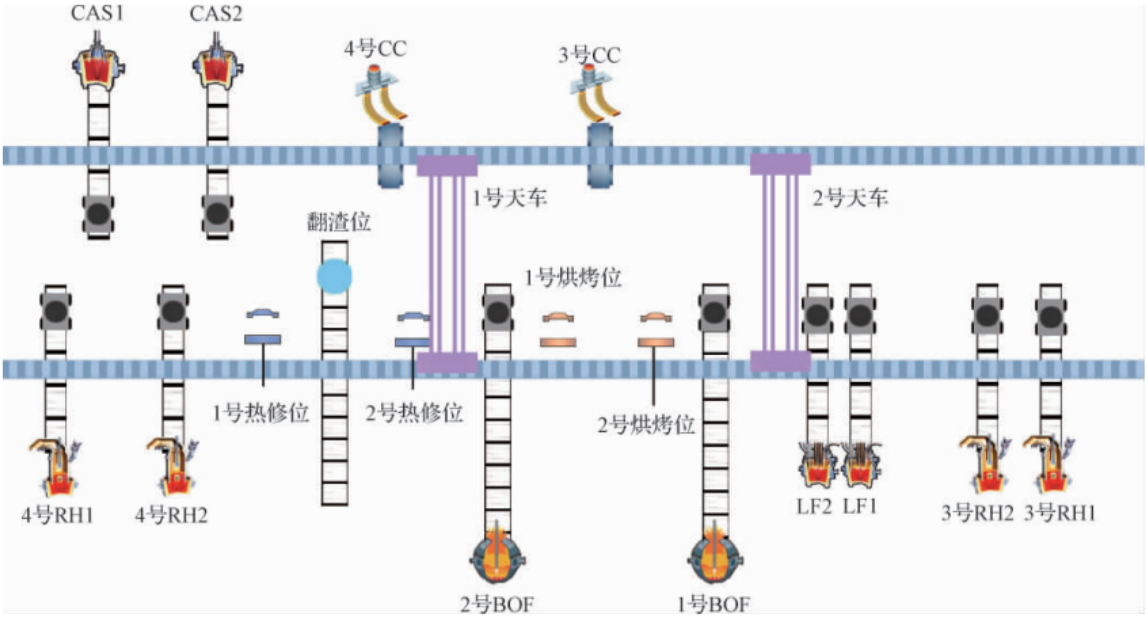

图 3　某钢厂出钢跨布局图

2.1 状态空间

对于天车调度问题，所需要的环境信息为任务起点工位和终点工位在跨中的位置、同一跨中所有天车的位置及状态。为了减少模型网络的复杂度，将车间跨中的工位相对位置等比例缩小10倍， 用3*30的矩阵表示环境的状态。

	0	1	2	3	4	5	6	7	…	26	27	28	29
起点工位	0	0	1	0	0	0	0	0	…	0	0	0	0
天车	0	0	0	1	0	0	2	0	…	0	0	0	0
终点工位	0	0	0	0	0	1	0	0	…	0	0	0	0

图4 环境的状态矩阵

由于工位的纵向位置只影响天车小车的纵向移动，假设天车小车的纵向移动能够在大车横向移动过程中完成，对任务完成时间影响较小，因此忽略，仅**考虑工位和天车的横向位置**，用矩阵列坐标进行表示。第一、二和三行分别为任务起点工位、天车和任务终点工位在跨中的相对位置。任务产生时，矩阵中工位对应的位置的值为1。

图 4 表示当前状态有一个任务，起点工位位于位置2，终点工位位于位置5，1号天车位于位置3，2号天车位于位置6。两台天车当前状态为空闲，若天车负载钢包时，则值变为1.1、2.1。

2.2 动作空间

动作空间是智能体与环境交互的所有动作集合。假设在大车移动过程中，小车有足够时间移动，不考虑小车的移动，因此天车的动作空间包含**左移、右移、静止、吊包、放包**共5个动作，用0、1、2、3、4分 别表示。

2.3 奖励函数和判定规则

奖励函数是环境对智能体当前采取的动作评价规则，是智能体进行学习改善策略的重要指引信号。 奖励函数应根据问题的目标函数进行设计，目前尚未有相关的理论指导，仅能通过**经验和试验**确定。**天车空载和负载钢包时的奖励分开**计算。详细的奖励设置见表1和表2。

表1 天车空载时的奖励设置表

动作	状态	奖励值
吊包	处在起始工位	5
	不在起始工位	0
放包	处在任意位置	-0.1
静止	没有待完成的任务	0.1
	存在待完成的任务	0
移动	将与其他天车发生碰撞	-1
	将超出车间跨的边界	-10
	其他情况	0

表2 天车负载钢包时的奖励设置表

动作	状态	奖励值
吊包	处于任意位置	-0.1
放包	处于目标工位	30
	不在目标工位	-0.1
静止	任意位置	-0.2
移动	将与其他天车发生碰撞	-1
	不与其他天车发生碰撞且靠近目标工位	0.1
	不与其他天车发生碰撞且远离目标工位	-0.2
	将超出车间跨的边界	-10

为了达到完成所有任务总时间最小的目标，天车没执行一次动作后，奖励额外减少`0.5`。因此为了使累计奖励最高，天车需要用最少的动作完成所有任务，任务完成总时间最小。

天车移动动作的奖励中，存在状态交叉覆盖，需要进行奖励判定顺序设置。首先，优先判断天车是否会移动超出车间跨的边界，若不满足，再进行是否会发生天车碰撞的判定。前两者都不满足时，进行天车远离目标工位或靠近目标工位和其他的情况判定。

2.4 Q网络结构

每台天车各有一个结构相同的Q网络，每个神经网络模型的输入为3*30的矩阵，输出为5个动作的价值，具体网络如下图5所示。

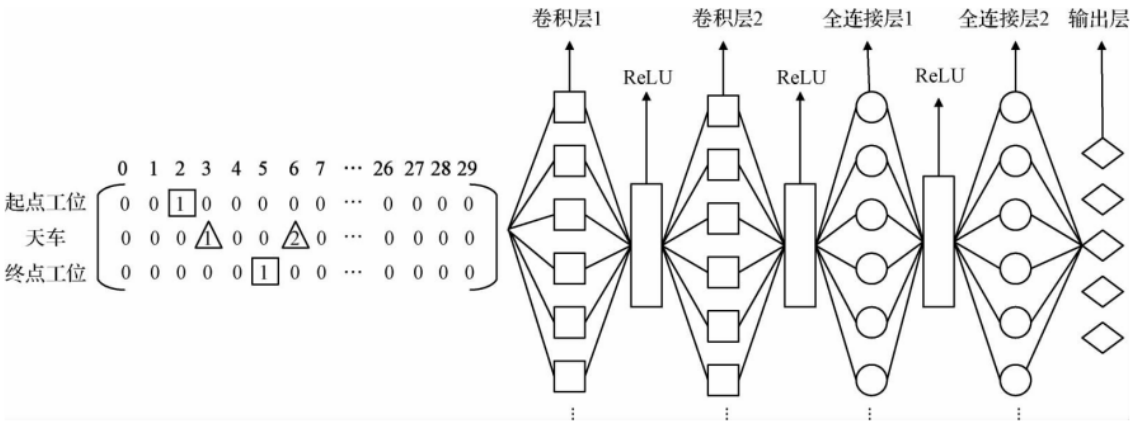

图5 神经网络结构示意图

该神经网络由2层卷积层和2层全连接层构成，其中卷积层的卷积核大小为3*30，全连接层神经元个数1024个，激活函数采用ReLU函数。

2.5 Q网络训练

采用DQN算法对Q网络模型进行训练，两台天车的Q网络与出钢跨之间的交互过程如图6所示。训练时，一个单位时间内，两个网络一次根据环境(出钢跨)的状态选择一个动作，先后作用于环境，得到各自奖励值进行求和后作为两个网络模型得到的立即奖励。天车出现路径冲突时，由于他们各自的奖励是两个天车网络决策得到的动作奖励之后，因此他们会选择最佳的避让方式，获得最高奖励值。

为了让网络收敛至全局最优，训练过程的策略采用$\varepsilon - greedy$策略，为探索率。以ε的概率从动作空间A中随机选择动作，以$1 - \varepsilon$的概率选择Q网络输出值最大对应的动作。

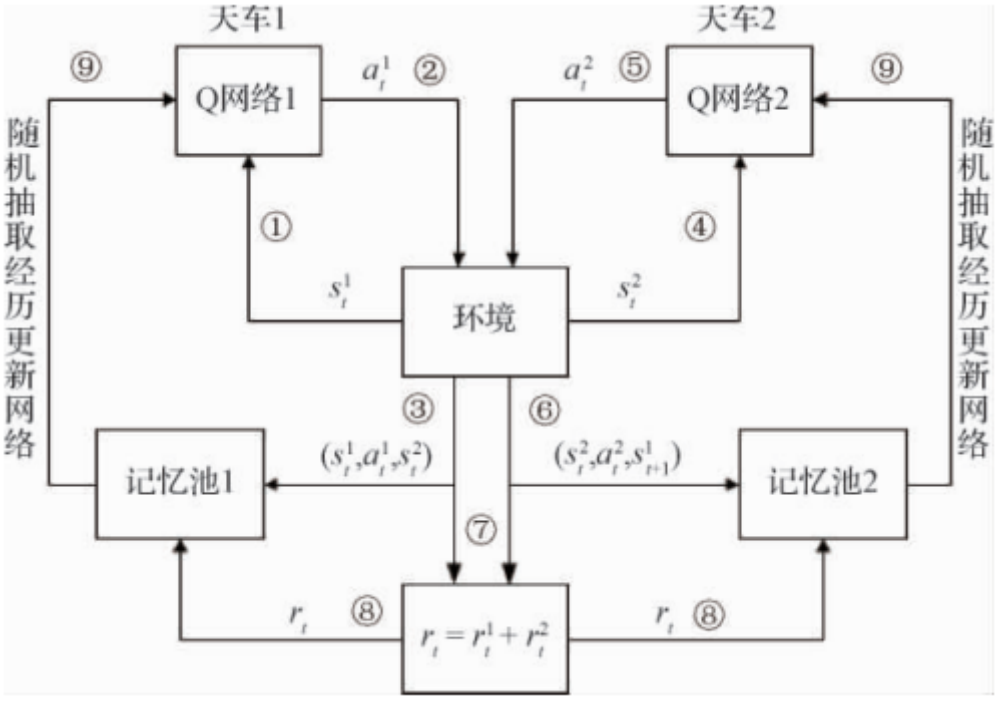

图6 双天车网络与环境交互示意图

算法Q网络训练算法：

初始化记忆池M^1, M^2

随机初始化$Q^{\pi 1}, Q^{\pi 2}$网络的参数θ^1, θ^2

随机初始化目标$\hat{Q}^{\pi 1}, \hat{Q}^{\pi 2}$网络的参数$\hat{\theta}^1, \hat{\theta}^2$

初始化奖励折扣因子γ

初始化探索率ε

初始化动作空间A

```
For episode = 1,2,...,N do
```

初始化环境e得到初始状态s_0^1;

```
For t=0,1,2,...,T do
```

$$a_t^1 = \begin{cases} \underset{a}{\operatorname{argmax}}\, Q^{\pi^1}\left(s_t^1, a \mid \theta^1\right), \\ a \in A, \text{ 概率 } 1-\varepsilon \\ \text{随机选择动作 } a, \\ a \in A, \text{ 概率 } \varepsilon \end{cases} ; \tag{3}$$

执行动作a_t^1得到立即奖励r_t^1和下一个状态s_t^2;

$$a_t^2 = \begin{cases} \underset{a}{\operatorname{argmax}}\, Q^{\pi^2}\left(s_t^2, a \mid \theta^2\right), a \in A, \\ \text{概率 } 1-\varepsilon \\ \text{随机选择动作 } a, \\ a \in A, \text{ 概率 } \varepsilon \end{cases} \tag{4}$$

执行动作a_t^2得到立即奖励r_t^2和下一个状态s_{t+1}^1;

$r_t = r_t^1 + r_t^2$;

将经历$(s_t^1, a_t^1, r_t, s_t^2)$,$(s_t^2, a_t^2, r_t, s_{t+1}^1)$分别存入记忆池$M^1, M^2$中;

从M^1中随机抽取经历(s_j, a_j, r_j, s_{j+1});

$$y_j = \begin{cases} r_j, s_{j+1} \text{ 是结束状态} \\ r_j + \gamma \max_a \hat{Q}^{\pi 1}\left(s_{j+1}, a \mid \hat{\theta}^1\right), \text{其他} \end{cases} \tag{5}$$

计算损失函数值$l_1 = smoothL_1[y_j - Q^{\pi 1}(s_j, a_j|\theta^1)]$;

采用梯度下降法更新$Q^{\pi 1}$网络参数θ^1;

从M^2中随机抽取经历(s_i, a_i, r_i, s_{i+1});

$$y_j = \begin{cases} r_j, s_{j+1} \text{ 是结束状态} \\ r_j + \gamma \max_a \hat{Q}^{\pi 2}\left(s_{j+1}, a \mid \hat{\theta}^2\right), \text{其他} \end{cases} \tag{6}$$

计算损失函数值$l_2 = smoothL_1[y_j - Q^{\pi 2}(s_j, a_j|\theta^2)]$;

采用梯度下降法更新$Q^{\pi 2}$网络参数θ^2;

$s_t^1 = s_{t+1}^1$;

每过C步迭代更新目标网络$\hat{Q}^{\pi 1} = Q^{\pi 1}, \hat{Q}^{\pi 2} = Q^{\pi 2}$;

end for

end for

输出$Q^{\pi 1}, Q^{\pi 2}$

在网络模型训练过程汇总，误差函数采用$smooth L_1(x)$函数，如式(7)所示。

$$\text{smooth}\, L_1(x) = \begin{cases} 0.5x^2, & |x| < 1 \\ |x| - 0.5, & \text{其他情况} \end{cases} \tag{7}$$

3 仿真试验与分析

基于强化学习平台OpenAI Gym搭建性训练环境，采用钢厂实际生产数据对网络进行训练，模型训练参数设置：学习率α为0.0002，奖励扣除因子γ为0.8，搜索率ε初始值为0.99，经过100万步迭代衰减到0.1，记忆池M大小为100万。

表3 部分天车调度方案

序号	任务派发时间-计划结束时间	实际起吊时间-卸载时间	起点工位	终点工位	天车编号
1	00:00:30-00:10:10	00:00:50-00:02:50	1号BOF	3号RH1	2
2	00:15:05-00:23:05	00:15:40-00:17:40	2号BOF	4号RH1	1
3	00:55:05-01:04:10	00:55:20-00:56:10	3号CC	翻渣位	2
4	00:59:15-01:24:10	01:00:30-01:02:50	3号RH1	3号CC	2
5	01:12:05-01:37:00	01:13:30-01:15:00	4号RH1	4号CC	1
6	01:46:15-02:10:00	01:46:20-01:48:20	翻渣位	2号热修	2
7	02:04:15-02:12:00	02:05:00-02:06:10	1号BOF	LF1	1
8	02:12:05-02:20:00	02:12:10-02:12:30	4号CC	翻渣位	2
9	02:23:10-02:20:00	02:23:30-02:24:00	2号热修	2号BOF	2

3.1 基于固定分区的天车调度

其原理是根据工位之间任务产生频率大小和就近原则，将跨区划分为与天车数量相等的区域，每台天车负责由该区域产生的任务。天车发生冲突时，通常以无任务天车避让有任务天车，运输空包的天车以避让运输重包的天车的方式来解除冲突。基于固定分区的天车调度方法能在一定程度上减少天车之间的冲突，且实现方法简单，适用于现场任务的随机性和突发性。

[1]俞侠. 炼钢–精炼–连铸生产过程天车调度问题研究[D].东北大学,2012.