

控制与决策

Control and Decision

面向人机物三元数据的热轧调度问题研究

李洪泽, 凌卫青, 刘飞翔

引用本文:

李洪泽, 凌卫青, 刘飞翔. 面向人机物三元数据的热轧调度问题研究[J]. *控制与决策*, 2021, 36(11): 2825–2831.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2020.0551>

您可能感兴趣的其他文章

Articles you may be interested in

基于地标特征和元学习方法推荐最适用优化算法

Recommending best suitable metaheuristic based on landmarking feature and meta-learning approach

控制与决策. 2021, 36(5): 1223–1231 <https://doi.org/10.13195/j.kzyjc.2019.0993>

顺序依赖的调整时间和拖期的无缝钢管热轧批量调度算法

Hot-rolled batch scheduling algorithm for seamless steel tube with sequence-dependent setup times and tardiness

控制与决策. 2021, 36(2): 505–512 <https://doi.org/10.13195/j.kzyjc.2019.0723>

基于深度强化学习与迭代贪婪的流水车间调度优化

Scheduling optimization for flow-shop based on deep reinforcement learning and iterative greedy method

控制与决策. 2021, 36(11): 2609–2617 <https://doi.org/10.13195/j.kzyjc.2020.0608>

带峰值能耗约束流水线调度的协同群智能优化

Cooperative memetic optimization for flowshop scheduling with peak power consumption constraint

控制与决策. 2021, 36(10): 2350–2358 <https://doi.org/10.13195/j.kzyjc.2020.0429>

基于强化学习的小型无人直升机有限时间收敛控制设计

Finite time control based on reinforcement learning for a small-size unmanned helicopter

控制与决策. 2020, 35(11): 2646–2652 <https://doi.org/10.13195/j.kzyjc.2019.0328>

面向人机物三元数据的热轧调度问题研究

李洪泽, 凌卫青[†], 刘飞翔

(同济大学 电子与信息工程学院, 上海 201804)

摘要: 随着钢铁行业的数字化发展, 其订单逐渐趋于多样化和随机化, 这对热轧调度模型的适应性和灵活性等提出了新的要求. 针对热轧调度问题, 当前的主流方法是启发式算法, 但其存在两个问题: 一是没有考虑数据的组织表示; 二是此类算法具有很强的针对性, 当问题发生很小的改变就需要进行复杂的参数调整. 相比之下, 机器学习具有更好的适应性和灵活性, 对此, 采用本体进行人机物三元数据的组织表示, 提出一种指针网络+强化学习的热轧调度求解方法. 采用指针网络来学习序列到序列的映射, 同时为解决指针网络训练困难和性能不高等问题, 通过 actor-critic 网络进行训练, 提高模型的准确性和收敛速度. 最后, 通过设计相应的实验对算法的性能进行仿真并与 LK-H 的局部搜索算法进行对比, 进一步验证了所提出方法的有效性.

关键词: 人机物三元数据; 本体; 热轧调度; 指针网络; 强化学习

中图分类号: TP18

文献标志码: A

DOI: 10.13195/j.kzyjc.2020.0551

开放科学(资源服务)标识码(OSID):



引用格式: 李洪泽, 凌卫青, 刘飞翔. 面向人机物三元数据的热轧调度问题研究[J]. 控制与决策, 2021, 36(11): 2825-2832.

Research on hot rolling scheduling problem oriented to human-cyber-physical data

LI Hong-ze, LING Wei-qing[†], LIU Fei-xiang

(College of Electronics and Information Engineering, Tongji University, Shanghai 201804)

Abstract: With the digital development of the steel industry, the orders become multiple species and random change, which puts forward new requirements for the adaptability and flexibility of the hot-rolling scheduling model. For the hot rolling scheduling problem, the current mainstream method is a heuristic algorithm, which has two problems: one is that it does not consider the organizational representation of data; the other is that this kind of algorithm has strong pertinence. When the problem changes very little, it needs complex parameter adjustment. Compared with machine learning, it has better adaptability and flexibility. Therefore, this paper uses ontology to represent the organization of human-cyber-physical data and proposes a hot rolling scheduling solution method of pointer network + reinforcement learning for the first time. The pointer network is used to learn the mapping from sequence to sequence. In order to solve the problems of the pointer network training difficulty and low performance, the actor-critical network is used to improve the accuracy and convergence speed of the model. Finally, the effectiveness and performance of the algorithm are simulated by designing the corresponding experimental scheme and compared with LK-H's local search algorithm to further verify the effectiveness of the proposed method.

Keywords: human-cyber-physical data; ontology; hot rolling scheduling; pointer network; reinforcement learning

0 引言

钢铁行业作为我国重要的基础产业正面临严峻的挑战, 钢铁行业出现利润率低、库存率高、负债率高等问题. 在我国改革的大背景下, 利用数字化转型来优化钢铁生产过程已成为趋势. 热轧通常被认为是整个钢铁生产过程中最为核心的环节, 对产品的质

量、企业生产效率、企业效益有着至关重要的影响. 如何安排订单的生产计划以实现企业的效益最大化就是热轧生产调度要解决的问题, 作为钢铁行业中一个极其复杂的问题, 近年来, 研究人员对该领域进行了深入的研究.

目前, 围绕热轧调度问题的研究大多采用启发式

收稿日期: 2020-05-11; 修回日期: 2020-09-07.

基金项目: 科技创新 2030 新一代人工智能重大项目课题(2018AAA0101801).

责任编辑: 刘士新.

[†]通讯作者. E-mail: lingweiqing@tongji.edu.cn.

算法,如模拟退火算法^[1]、蚁群算法^[2]、混合变邻域搜索算法(HVNS)^[3]、遗传算法^[4]等. Jia等^[5]将热轧生产调度问题表述为一个多目标的收集奖励车辆路线模型(PCVRP),提出了Pareto最大最小蚂蚁系统算法(P-MMAS)来对模型进行求解,同时提出了基于TOPSIS法的多目标决策方法,使用该方法从Pareto最优解中选择轧制批量. Pan等^[6]针对紧凑型带钢(CSP)生产过程的热轧生产调度问题,提出了一个混合整数规划模型,使用该模型将问题抽象为两个耦合子问题的组合,分别为带钢分配问题和带钢排序问题,并提出了一个两阶段启发式算法,即板带分配算法和板带排序算法. Zhang等^[3]在上述问题的基础上,鉴于板带排序算法在平均厚度这个变化目标上表现不佳,提出了一种混合变量邻域搜索算法(HVNS)来求解如何才能使得相邻带钢在轧制过程中的平均厚度变化最小. Pan等^[7]在上述问题基础上,继续研究板带排序问题,提出了一种多目标优化算法来求解板带排序问题的Pareto最优或近似最优解.

从目前的研究现状来看,这些启发式算法在特定的问题上已经取得了不错的效果,并且随着算法的不断改进,局部最优问题也得到了相应改进. 但是,这些算法需要针对不同的热轧调度问题分别设计,当出现新的订单或者订单发生改变时,需要进行复杂的参数修改等操作. 相比之下,随着机器学习的发展,其在组合优化问题方面逐渐表现出良好的性能,此外,也有学者通过机器学习来研究设计新的启发式算法^[8].

强化学习作为机器学习的一个领域,其主要原理是通过代理(agent)执行某个动作(action)后,通过观察(observation)得到环境的奖励(reward),强化学习的目标就是不断最大化奖励函数. 将强化学习应用于组合优化问题起源于pointer network网络^[9]的提出,其作为一种新的网络可以解决从一个序列到另一个序列的映射问题,输出序列的内容与输入序列的内容完全一致,只是序列的顺序发生了改变. 文献[10]借用pointer network网络,另外加上attention mechanism,采用actor-critic进行训练以求解简单的组合优化问题,如旅行商问题(TSP). actor-critic作为强化学习的框架,结合了以值为基础(如Q learning)和以动作为基础(如policy gradients)的两类强化学习算法,既能处理连续的动作,又能单步更新. 也有一些强化学习在调度方面的应用研究,如:文献[11]将强化学习应用到自动化生产线的调度,避免了手工提取特征,克服了结构化数据集的不足;文献[12]将其应用到混合车间调度问题,更有利于实现全局最优解;文献[13]设计了一种深度强化学习方法,并将其应用于急诊科

病人的排班过程中.

强化学习的自学习能力提供了简化问题求解的思路,在实际应用中,它允许不知道具体的模型,但是可以写出奖励函数,然后通过逐渐训练得出答案. 热轧调度问题作为TSP问题的一种变形,其输入为需要进行轧制的单元,输出为轧制单元的轧制顺序. 采用指针网络编码器对输入的轧制单元进行编码;然后利用解码器对轧制单元的编码信息和已经确定的轧制单元序列进行解码,得到下一个进行轧制的单元;最后借鉴文献[10],利用actor-critic对其进行训练.

此外,围绕热轧调度的研究,存在的另一个问题是缺少数据的组织表示等相关的研究. 在钢铁行业数字化转型的关键时期,要重新审视企业数据,从数据的角度出发来研究问题. 已有模型的建立依赖于轧机能力约束、产品工艺约束以及订单信息等,这些都是来自确定的已知数据,但是关于数据的组织表示并没有涉及,只是针对具体的问题. 在数字转型、知识驱动的行业背景下,需要参考已有的一些数据融合处理方法^[14-15],重新考虑数据的组织表示等问题.

如今数据的来源与形式也发生了变化,如果还像以往只重视传感器的数据,则会失去企业数字化的意义. 虽然人类在许多方面无法与传感器竞争,但是人具有很强的理解能力和处理不确定信息的能力,可以理解许多复杂的情况,以人为传感器,可以弥补物理传感器的局限性,充分考虑实体概念关系等语义信息. 信息技术的发展,使得互联网数据激增,同时也出现了大数据、云计算等新技术,把互联网中的数据加入进来,以增强感知和决策. 本团队将上述3种来源的数据描述为人机物三元数据,从而出现了面向人机物三元数据^[16]组织和表示的问题. 传统结构化的数据可以很方便地存储和利用,如今要考虑的是多元异构的数据,同时还要保留数据的语义信息.

本体作为一种语义化的数据描述和共享的方法,主要包括4个元素:概念、关系、公理和实例,可以解决语义冗余、数据异构等问题,在知识密集型上下文中扮演着重要的角色^[17],已被广泛用于解决数据的组织表示和融合等问题^[18-19]. 综上所述,本文将热轧调度问题相关数据采用本体进行组织表示,然后应用“指针网络+强化学习”的方法模型(下文称RL模型)对问题进行求解,整体结构如图1所示.

1 问题描述与建模

热轧的生产调度是指在考虑轧制规则和轧机约束的条件下,根据订单安排不同规格钢材的轧制顺序,最大限度地减少更换轧辊的时间,同时节约安装

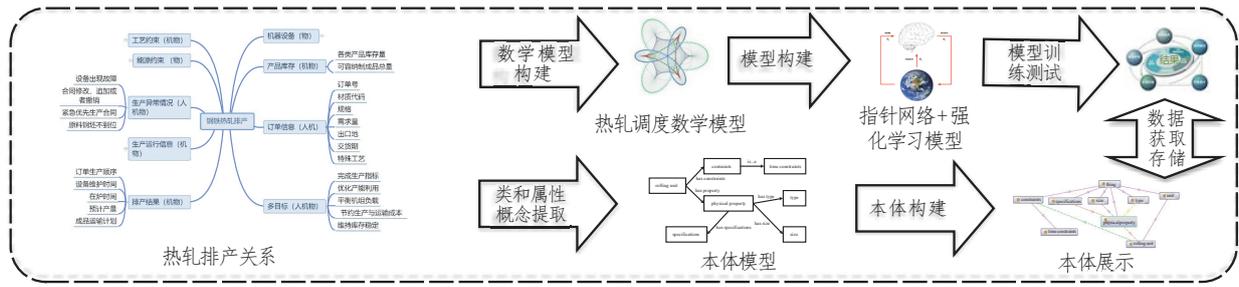


图 1 热轧调度问题整体框架

轧辊的成本. 不同的轧制单元对应于不同的轧制规格, 当发生转换时, 需要额外的调整时间, 所以合理地安排轧制顺序可以提高效率. 同时, 不同的轧制单元可能对应于不同的交货期, 所以在设计顺序时要满足客户的交货时间, 不然会产生延期成本.

本文把热轧的批量调度问题描述为: 需要轧制的单元已经确定, 并且每个轧制单元的规格、交货的时间等属性都是已知的, 需要做的就是对轧制单元进行合理排序, 使得生产过程中设备的调整时间和拖延惩罚最小. 该问题可以归纳为旅行商问题, 需要轧制的单元为需要走的城市, 城市之间的距离代表不同轧制单元设备的调整时间, 同时考虑到交货期限就是某个城市必须在特定的时间内到达, 不然会产生交货延期惩罚. 接下来对问题进行数学模型的构建.

1) 定义符号含义.

- i 为轧制单元编号, 假定有 n 个, 即 $i = 1, 2, \dots, n$;
- k 为不同轧制单元的轧制规格编号, 假定有 m 个, 即 $k = 1, 2, \dots, m$;
- r_i 为轧制单元 i 的轧制规格;
- p_i 为轧制单元 i 的轧制时长;
- d_i 为轧制单元 i 的交货日期;
- s_i 为轧制单元 i 轧制的开始时间;
- c_i 为轧制单元 i 轧制结束时间, 有 $c_i = s_i + p_i$;
- $t_{kk'}$ 为轧制规格从 k 到 k' 的调整时间;
- γ 为延期的单位惩罚值;
- U 为一相对大的常数.

2) 定义决策变量.

$$x_{ijk} = \begin{cases} 1, & \text{表示在轧制规格 } k \text{ 内轧制单元 } i \text{ 先于轧制单元 } j \text{ 生产;} \\ 0, & \text{otherwise.} \end{cases}$$

$$y_{kk'} = \begin{cases} 1, & \text{表示轧制规格由 } k \text{ 切换到 } k'; \\ 0, & \text{otherwise.} \end{cases}$$

$$z_{ik} = \begin{cases} 1, & \text{表示轧制单元 } i \text{ 属于规格 } k; \\ 0, & \text{otherwise.} \end{cases}$$

3) 定义优化目标.

最小化轧制规格变化引起的设备调整时间. 即

$$\min Z_1 = \sum_{k=1}^m \sum_{k'=1}^m \sum_{i=1}^n z_{ik} y_{kk'} t_{kk'}; \quad (1)$$

最小化延期惩罚

$$\min Z_2 = \gamma \sum_{i=1}^n \max(0, c_i - d_i). \quad (2)$$

4) 考虑相关约束.

考虑到轧制单元 i 只能被安排在一个轧制规格, 有

$$\sum_{k=1}^m z_{ik} = 1, \quad i = 1, 2, \dots, n. \quad (3)$$

考虑轧制单元在加工时不允许中断, 有

$$c_i = s_i + q_i, \quad i = 1, 2, \dots, n. \quad (4)$$

考虑不能同时加工两个轧制单元, 有

$$s_j - c_i + U(1 - x_{ijk}) \geq 0, \quad i, j = 1, 2, \dots, n, \quad k = 1, 2, \dots, m. \quad (5)$$

考虑轧制单元间的先后顺序关系, 有

$$(s_i - s_j)x_{ijk} \leq 0, \quad i, j = 1, 2, \dots, n, \quad k = 1, 2, \dots, m; \quad (6)$$

$$(c_i - c_j)x_{ijk} \leq 0, \quad i, j = 1, 2, \dots, n, \quad k = 1, 2, \dots, m. \quad (7)$$

考虑取值约束, 有

$$s_i \geq 0, \quad c_i \geq 0, \quad i = 1, 2, \dots, n; \quad (8)$$

$$x_{ijk} + x_{jik} = 1, \quad i, j = 1, 2, \dots, n, \quad i < j, \quad k = 1, 2, \dots, m; \quad (9)$$

$$x_{ijk} \in \{0, 1\}, \quad i, j = 1, 2, \dots, n, \quad i < j, \quad k = 1, 2, \dots, m. \quad (10)$$

2 数据本体表示

数据的有效组织表示是其可以高效应用的关键, 像单一结构化的数据, 一般直接将其存储到结构数据

3.2 指针网络

指针网络是由 seq2seq 网络模型改进而来, 主要解决输出字典大小不随输入大小改变的问题. 指针网络包括两个递归神经网络 (RNN) 模块, 即编码器和解码器, 两者是长短时记忆 (LSTM) 单元, 可以将输入序列映射为一系列按概率指向输入序列元素的指针, 结构如图 5 所示. 其在 seq2seq 基础上添加了注意力模块, 将 softmax 的结果作为条件概率输出, 公式表示如下:

$$u_j^i = v^T \tanh(W_1 e_j + W_2 d_i), \quad j \in (1, \dots, n); \quad (15)$$

$$p(C_i | C_1, \dots, C_{i-1}, P; \theta) = \text{softmax}(u^i). \quad (16)$$

其中: 向量 u^i 为输入元素的指针; v 、 W_1 和 W_2 是指针网络模型可以训练的参数; softmax 将向量 u^i 进行归一化, 即标准化为输入字典上的输出分布; $p(C_i | C_1, \dots, C_{i-1}, P; \theta)$ 表示选中的条件概率.

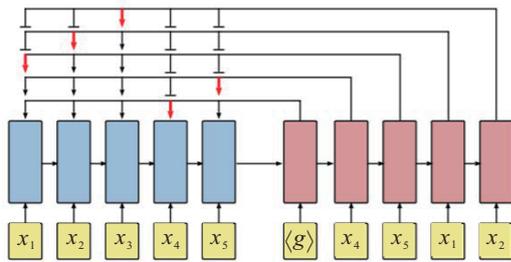


图 5 指针网络结构图^[9]

以图 4 为例进行说明. 模型的编码器为 LSTM 网络, 将需要轧制的 4 个轧制单元信息一次输入 LSTM 网络中, 然后输出整个轧制单元集的信息和重点包含该轧制单元的信息, 将此网络节点称为该轧制单元的节点. 模型的解码器也为 LSTM 网络, 将当前轧制单元作为输入, 将 LSTM 网络输出分别与编码器部分的 n 和轧制单元作为输入, 得到 n 个输出, 再通过 softmax 进行归一化处理, 得到下一个轧制单元的概率分布.

3.3 强化学习

采用 actor-critic 进行训练, 其主要包括: 一个行动者网络, 用于预测给定决策步骤下一个动作的概率分布; 一个评论家网络, 用于估计给定状态下问题的回报. 可以结合值优化和策略梯度优化的优点, 进行单步更新, 具有效率高、收敛快等特点.

参考文献 [10] 基于策略梯度来优化指针网络的参数 θ . 对于给定的热轧集合 S , 网络的优化目标是 minimized 跳跃惩罚和延期惩罚期望

$$J(\theta | S) = E_{\pi \sim p_{\theta}(\cdot | S)} L(\pi | S), \quad (17)$$

其中 $L(\pi | S)$ 表示轧制单元集合 S 确定的情况下, 依照策略 π 行动得到的期望累计损失, 前面已经求出. 本文直接参考文献 [20] 提出的 REINFORCE 算法对行动者网络进行优化, 即

$$\nabla_{\theta} J(\theta | S) = E_{\pi \sim p_{\theta}(\cdot | S)} [(L(\pi | S) - b(S)) \nabla_{\theta} \log_{p_{\theta}}(\pi | S)]. \quad (18)$$

评论家网络以均方误差为目标, 采用随机梯度下降进行训练, 即

$$l(\theta_r) = \frac{1}{B} \sum_{i=1}^B \|b_{\theta_r}(S_i) - L(\pi_i | S_i)\|_2^2. \quad (19)$$

综上所述, 首先采用指针网络进行序列到序列的学习, 由于其是监督学习, 需要标签数据; 然后引入强化学习算法, 采用策略梯度进行训练; 最后, 为了提高收敛速度, 进一步添加评论家网络. 算法流程如图 6 所示.

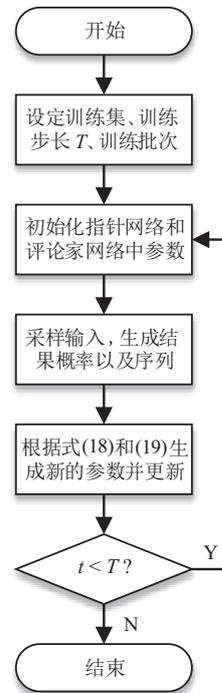


图 6 算法流程

4 实验分析

为了验证上述方法的有效性, 进行了多次实验.

首先, 由数学模型可知, 热轧调度问题是一个双目标问题, 采用式 (12) 和 (13) 进行计算. 对于每个目标分别设置一个权重, 如式 (14) 所示, 根据计算结果, 选择合适的调度顺序. 其次, 实验分为两个部分: 第 1 部分将时间惩罚的权重设置为 0, 其变为典型的 TSP 问题; 第 2 部分将权重各设置为一半, 也就是更改第 1 部分中的奖励函数, 由原来的 $L'_1(\pi | S)$ 更改为 $L(\pi | S)$. 整体实验步骤如图 7 所示.

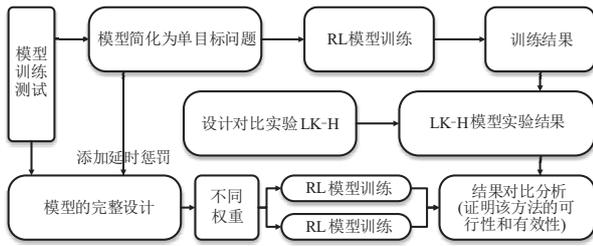


图7 实验步骤

实验采用 python 语言和 PyTorch 框架, 运行环境为 Windows 10 操作系统、8 G 内存、Intel Core i5-5200U/2.2 GHz.

4.1 第1部分

首先只考虑跳跃惩罚这一个优化目标. 假设轧制单元为坐标对应的点, 是从一个固定的分布中随机产生的, 并且其范围为 $[0, 1] \times [0, 1]$. 轧制单元的时间

属性设为零, 这种情况更接近 TSP 问题, 主要用来验证模型的正确性和有效性. 设定序列长度为 18, 也就是随机生成 18 个坐标点. 实验过程中部分点的坐标如表 1 所示, 部分结果如图 8 所示.

表1 坐标点

序号	1	2	3	4	5	6
x	0.9964	0.8854	0.2985	0.4704	0.7028	0.5946
y	0.909	0.6762	0.5969	0.0275	0.7572	0.6545
序号	7	8	9	10	11	12
x	0.2570	0.5345	0.9657	0.5663	0.8494	0.1889
y	0.1703	0.2001	0.6486	0.1410	0.6559	0.4480
序号	13	14	15	16	17	18
x	0.7216	0.8901	0.7288	0.3917	0.8546	0.6316
y	0.7683	0.0392	0.7622	0.4616	0.2299	0.8633

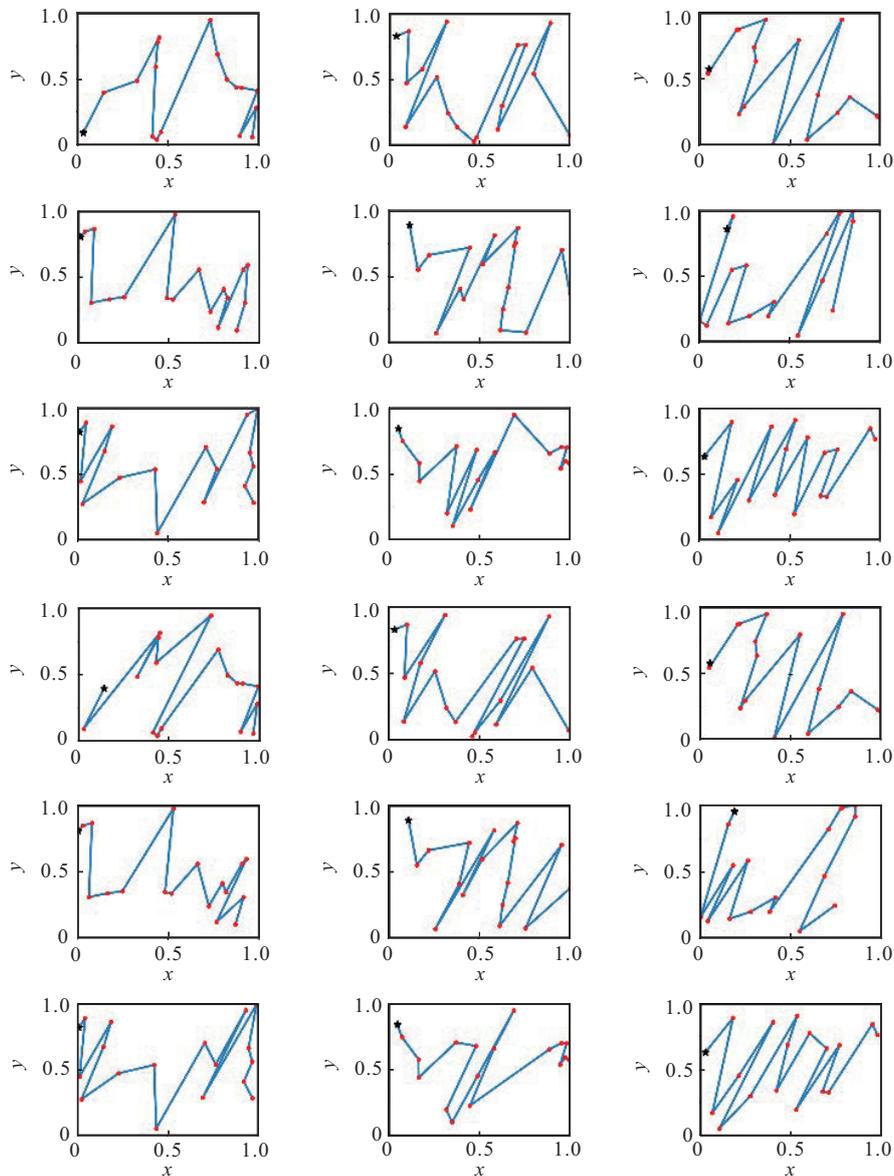


图8 模型训练过程

通过实验可以得到训练过程中排产结果. 其中: 不同的点对应于需要加工的轧制单元, 其距离对应于惩罚值. 从图 8 中可以直观地看出训练的变化情况. 为了验证结果的精确性, 另外使用开源 LK-H^[21] 的局部搜索算法进行对比验证. 设定二维随机分布的数据集分别为 1 000、2 000、3 000、4 000 个, 分为训练集和测试集进行实验, 得到结果如表 2 所示.

表 2 随机数据结果平均值对比表

数据样本	RL	LK-H
1 000	3.97	3.88
2 000	3.95	3.76
3 000	3.80	3.81
4 000	3.81	3.77

如表 2 所示, 在 $x-y$ 二维随机生成数据测试中, 本文方法 (RL) 在少量训练时表现差于 LK-H 算法, 在训练量增加后接近甚至优于 LK-H 算法, 说明了该方法的有效性. 除此之外, 当问题更改后, 启发式算法要进行复杂的模型设计, 采用强化学习 + 指针网络的模型, 可以更方便地进行调整. 例如在调度中考虑延时惩罚, 只需要更改奖励计算函数, 在原来距离惩罚的基础上再加上延时惩罚.

4.2 第 2 部分

在验证了本文方法模型的可行性和正确性后, 添加最小化延期惩罚优化目标, 同样在 $[0, 1] \times [0, 1]$ 之间随机生成点. 此外, 再生成 $[5, 10]$ 之间随机数, 表示轧制单元需要的轧制时间, 然后再随机生成 $[5, 10]$ 之间的随机数, 依次累加表示轧制的交付时间, 此次设定序列长度为 20, 其他参数不变进行实验.

首先设定跳跃惩罚和延时惩罚权重对半进行模型的训练, 得到训练过程的惩罚值变化曲线如图 9 所示.

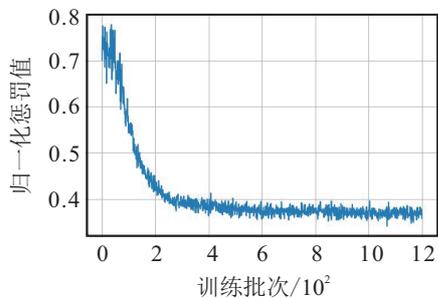


图 9 惩罚值变化曲线(1)

为了分析算法的性能, 从实验数据中取出训练过程惩罚值, 见表 3.

然后设定跳跃惩罚和延时惩罚权重分别为 0.8 和 0.2, 再次进行模型的训练, 得到训练过程的惩罚值

变化曲线, 如图 10 所示.

表 3 训练过程归一化惩罚值

reward	0.4	0.36	0.38	0.37	0.35	0.35	0.34	0.34
--------	-----	------	------	------	------	------	------	------

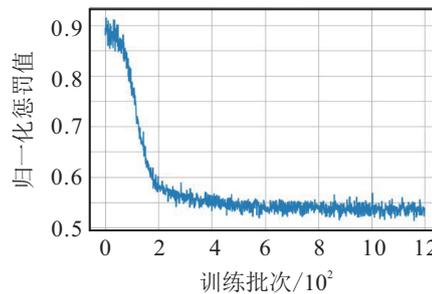


图 10 惩罚值变化曲线(2)

为了分析算法的性能, 从实验数据中再次取出训练过程惩罚值, 见表 4.

表 4 训练过程归一化惩罚值

reward	0.60	0.55	0.54	0.54	0.53	0.53	0.52	0.52
--------	------	------	------	------	------	------	------	------

由不同权重下训练过程中的曲线和惩罚值可以发现, 随着训练批次增加, 惩罚值慢慢变小并逐渐趋于稳定, 说明热轧的调度顺序逐渐趋于最优, 而且在不同的权重下, 惩罚值是不一样的, 这说明不同权重下的调度顺序也有所不同. 最后, 结合实验分析结果进一步表明了该方法的有效性.

将求解之后得到的结果, 也就是轧制顺序, 再加入到本体 rolling unit 的 sequence 属性中, 通过本体便可直接查看或者进行更进一步的知识推理等.

5 结 论

本文提出了一种基于本体的人机物三元数据组织表示方法. 该方法可以保留其概念与概念之间的关系, 并针对实际钢铁热轧调度中的三元数据表示问题, 采用所提出方法完成了本体构建. 同时, 还提出了指针网络 + 强化学习的方法来求解组合优化问题. 该方法不同于传统的启发式算法, 其可以通过更改奖励函数计算方法方便高效地对钢铁热轧调度问题进行求解, 通过算法模拟和性能对比验证了所提出方法求解组合优化问题的有效性. 该方法是机器学习方法用于解决组合优化问题的一次尝试, 还需要进行深入研究. 未来研究可从以下两点进行: 一是解决实际的调度问题还有诸多影响因素需要考虑, 在后续的研究中可以考虑如何将这些因素加入到模型的动态变量中; 二是充分发挥本体的推理能力以及强化学习的自学习能力, 进一步推动钢铁生产各个阶段调度和排产智能化.

参考文献(References)

- [1] Liu L L, Wan X, Gao Z G, et al. Research on modelling and optimization of hot rolling scheduling[J]. *Journal of Ambient Intelligence and Humanized Computing*, 2019, 10(3): 1201-1216.
- [2] Zhang R, Song S J, Wu C, et al. Robust scheduling of hot rolling production by local search enhanced ant colony optimization algorithm[J]. *IEEE Transactions on Industrial Informatics*, 2020, 16(4): 2809-2819.
- [3] Zhang B, Pan Q K, Gao L, et al. A hybrid variable neighborhood search algorithm for the hot rolling batch scheduling problem in compact strip production[J]. *Computers & Industrial Engineering*, 2018, 116: 22-36.
- [4] 崔志华, 张茂清, 常宇, 等. 基于平均距离聚类的NSGA-II[J]. *自动化学报*, 2021, 47(5): 1171-1182. (Cui Z H, Zhang M Q, Chang Y, et al. NSGA-II with average distance clustering[J]. *Acta Automatica Sinica*, 2021, 47(5): 1171-1182.)
- [5] Jia S, Yi J, Yang G, et al. A multi-objective optimisation algorithm for the hot rolling batch scheduling problem[J]. *International Journal of Production Research*, 2013, 51(3): 667-681.
- [6] Pan Q K, Chen Q D, Meng T, et al. A mathematical model and two-stage heuristic for hot rolling scheduling in compact strip production[J]. *Applied Mathematical Modelling*, 2017, 48: 516-533.
- [7] Pan Q, Gao L, Wang L, et al. A multi-objective hot-rolling scheduling problem in the compact strip production[J]. *Applied Mathematical Modelling*, 2019, 73: 327-348.
- [8] 陈思远, 林丕源, 黄沛杰. 指针网络改进遗传算法求解旅行商问题[J]. *计算机工程与应用*, 2020, 56(19): 231-236. (Chen S Y, Lin P Y, Huang P J. The improved genetic algorithm of pointer network is used to solve the traveling salesman problem[J]. *Computer Engineering and Applications*, 2020, 56(19): 231-236.)
- [9] Vinyals O, Fortunato M, Jaitly N. Pointer networks[C]. *Proceedings of the 28th International Conference on Neural Information Processing Systems-Volume 2*. Cambridge, 2015: 2692-2700.
- [10] Bello I, Pham H, Le Q V, et al. Neural combinatorial optimization with reinforcement learning[J]. 2016, arXiv: 1611.09940.
- [11] Shi D M, Fan W H, Xiao Y Y, et al. Intelligent scheduling of discrete automated production line via deep reinforcement learning[J]. *International Journal of Production Research*, 2020, 58(11): 3362-3380.
- [12] Han W, Guo F, Su X, et al. A reinforcement learning method for a hybrid flow-shop scheduling problem[J]. *Algorithms*, 2019, 12(11): 222.
- [13] Lee S, Lee Y H. Improving emergency department efficiency by patient scheduling using deep reinforcement learning[J]. *Healthcare*, 2020, 8(2): 77.
- [14] 崔志华, 张春妹, 时振涛, 等. 基于蝙蝠算法的观测矩阵优化算法[J]. *控制与决策*, 2018, 33(7): 1341-1344. (Cui Z H, Zhang C M, Shi Z T, et al. Measurement matrix optimization algorithm with bat algorithm[J]. *Control and Decision*, 2018, 33(7): 1341-1344.)
- [15] 罗俊海, 杨阳. 基于数据融合的目标检测方法综述[J]. *控制与决策*, 2020, 35(1): 1-15. (Luo J H, Yang Y. An overview of target detection methods based on data fusion[J]. *Control and Decision*, 2020, 35(1): 1-15.)
- [16] Chen S, Wang J, Li H, et al. Top-down human-cyber-physical data fusion based on reinforcement learning[J]. *IEEE Access*, 2020, 8: 134233-134245.
- [17] Reda R, Piccinini F, Carbonaro A. Towards consistent data representation in the IoT healthcare landscape[C]. *Proceedings of the 2018 International Conference on Digital Health*. New York, 2018: 5-10.
- [18] Wang S, Kang D, Li Y, et al. Research on ontology-based data fusion[J]. *Lecture Notes in Electrical Engineering*, 2018, 474: 714-720.
- [19] Sekkal N, Benslimane S M, Mriassa M, et al. Proactive and reactive context reasoning architecture for smart web services[J]. *International Journal of Data Mining, Modelling and Management*, 2020, 12(1): 1-27.
- [20] Williams R J. Simple statistical gradient-following algorithms for connectionist reinforcement learning[J]. *Machine Learning*, 1992, 8(3/4): 229-256.
- [21] Keld Helsgaun. LK-H[EB/OL]. [2018-08]. <http://akira.ruc.dk/keld/research/LKH/>, 2018.

作者简介

李洪泽(1995—), 男, 硕士生, 从事数字孪生、工业大数据的研究, E-mail: 1832964@tongji.edu.cn;

凌卫青(1973—), 男, 副研究员, 博士, 从事知识图谱、数字孪生、工业大数据等研究, E-mail: lingweiqing@tongji.edu.cn;

刘飞翔(1997—), 男, 硕士生, 从事工业大数据、知识图谱的研究, E-mail: 1930723@tongji.edu.cn.

(责任编辑: 李君玲)