



基于改进DDPG算法的高速列车运行调整方法研究

杨林远^{1,2}, 秦勇^{3,4}, 王莉^{1,2}

(1. 北京交通大学 交通运输学院, 北京 100044;

2. 北京交通大学 北京市城市轨道交通信息智能感知与服务工程技术研究中心, 北京 100044;

3. 北京交通大学 先进轨道交通自主运行全国重点实验室, 北京 100044;

4. 北京交通大学 运营主动安全保障与风险防控铁路行业重点实验室, 北京 100044)

摘要: 针对高速铁路网络复杂度提升下列车突发延误引发的时刻冲突、运行秩序紊乱及连锁延误问题, 文章提出一种基于改进深度确定性策略梯度 (DDPG) 算法的列车运行调整方法。针对传统DDPG在列车离散发车顺序优化中存在的收敛困难、决策稳定性不足等缺陷, 本方法将发车顺序建模为连续优先级分数向量, 通过强化学习框架实现发车决策的自适应动态优化; 同时引入状态归一化、奖励缩放机制提升模型收敛速度, 结合行为克隆预训练实现参数的高效初始化, 有效解决了离散排列问题与连续动作空间的适配性难题。为验证方法有效性, 基于京广高铁武广段运行数据构建仿真环境, 并设计两类典型延误场景进行试验, 对比改进后DDPG、去除预训练的改进DDPG、原始DDPG及传统“先到先发”和“先计划先发”调度方法。从算法性能维度, 改进后DDPG在两类场景中均实现快速收敛且运行稳定, 奖励值迭代曲线显著优于其他算法; 场景一中总晚点时长较“先到先发”“先计划先发”分别降低约67.8%、65.9%, 场景二中分别降低约48.0%、67.1%, 算法通过赶点、越行等策略有效降低了延误连锁反应。研究结果表明, 该方法在满足安全约束的前提下, 能在不同延误场景中实现列车群的高效秩序恢复, 为高速铁路运行调整提供了新的智能优化方案。

关键词: 列车运行调整; 深度强化学习; DDPG; 高速铁路; 行为克隆预训练; 高速列车

中图分类号: U292.91⁺4

文献标志码: A

文章编号: 1000-128X(2025)05-0141-08

0 引言

随着我国高速铁路、城际铁路与城市轨道交通的快速发展, 路网规模持续扩大、行车密度不断提高, 由设备故障、恶劣天气等突发事件引发的列车晚点及延误扩散问题愈发突出。面对动态多变、复杂性极强的运营环境, 传统调度方法难以兼顾实时性与优化性, 亟需引入智能化手段以提升调度效率与服务质量。

列车运行调整的核心任务是在扰动发生后, 通过

动态优化列车发车时间、到站时间、运行次序及停站方案等决策, 在满足安全间隔、车站容量、线路能力等硬约束的条件下, 实现总延误最小、准点率最高、旅客等待时间最短等目标^[1-3]。该问题具有高维、动态、随机、多目标与强约束的典型特征, 属于NP-Hard难题^[4-5]。

针对列车运行调整问题, 现有的研究方法主要分为仿真方法、运筹学方法和人工智能方法^[6-7]。仿真方法通过建立基础设施约束和列车运行过程的模拟程序, 在输入不同基础条件和参数进行测试并评估不同

基金项目: “十四五”国家重点研发计划课题 (2022YFB4300603); 中央高校基本科研业务费专项资金资助 (2023JBZY005)

通信作者: 王莉, 女, 博士, 副教授, 博士生导师, 主要从事交通运输现代组织管理理论与方法方面的研究; E-mail: wangli@bjtu.edu.cn

引用格式: 杨林远, 秦勇, 王莉. 基于改进DDPG算法的高速列车运行调整方法研究[J]. 机车电传动, 2025(5): 141-148.DOI:10.13890/j.issn.1000-128X.2025.05.015.

Citation: YANG Linyuan, QIN Yong, WANG Li. Research on high-speed train operation adjustment based on an improved DDPG algorithm[J]. Electric drive for locomotives, 2025(5): 141-148.DOI:10.13890/j.issn.1000-128X.2025.05.015.

决策方案的效果。文献[8]根据实时交通状况和网络拓扑结构,构建列车晚点仿真模型,并利用地铁运营数据对模型进行验证。文献[9]利用仿真技术模拟列车运行的复杂动态过程,通过动态调度优先级规则,最大程度地减少列车之间的冲突和延误,提高整体运行效率。运筹学方法一般构建包含多约束条件的列车运行优化调整模型,并通过优化器或启发式算法对模型求解,得到列车运行的实时调整策略^[7]。文献[10]提出了一种基于混合整数规划的列车运行调整方法。在该模型中,列车运行信息、乘客行为及其出行不便程度都由一组整数线性不等式表示。文献[11]针对地铁列车调度与衔接协同问题,建立了混合整数非线性规划模型,并采用拉格朗日松弛法求解。文献[12]针对疫情条件下的高速列车运行调整问题,建立了以列车总延误时间最小为目标的整数规划模型,并采用量子粒子群优化方法进行求解。

尽管上述方法为列车运行调整优化提供了基础性解决方案,但仿真技术难以完全复现多要素耦合的运营场景,运筹学模型受限于约束条件的过度抽象化,而启发式策略则普遍面临计算效率与解空间探索能力的双重瓶颈。列车运行调度调整过程常被作为马尔科夫决策过程来处理^[13]。强化学习方法本质上是一种基于动态规划思想且具有马尔可夫性质的半监督机器学习方法^[14-15],具有试错学习、延迟反馈、时间依赖性和适应环境变化等特点,因而在该领域展现出良好潜力。文献[16]于2016年首次将强化学习(Q-Learning)应用于列车时刻表调整,但受限于维度与收敛速度问题。文献[7]针对高速铁路动态调度,构建以最小化各站列车总晚点时间为目标的混合整数非线性规划模型,采用策略梯度强化学习求解。文献[17]在Q-Learning基础上提出 $Q(\lambda)$ 学习调整算法,借助累积式资格迹的多步奖励更新机制,优化学习效率与收敛性能。文献[18]针对突发事件下高速列车群调整,设计无模型强化学习方法,结合显式静态约束构造启发式规则,缩减搜索空间并降低试错次数。文献[19]基于强化学习近端策略优化(PPO),建立总延误最小化的列车运行调整模型,生成最优时刻表调整方案。文献[15]采用蒙特卡洛树搜索强化学习方法生成列车发车序列,结合启发式规则解决运行冲突。文献[20]针对强化学习在高铁调整中经验利用效率低、收敛慢等问题,提出优先级经验回放可转移深度强化学习调度方法。文献[21]以单列车单站总延误最小为目标,构建动态调度模型并利用Q-Learning求解。

由于列车运行状态信息具有高维连续性,传统表格型强化学习方法难以应对高复杂度场景,DDPG算法虽适用于高维连续控制任务,但直接应用于列车运行调整中的离散排序问题,易出现梯度中断与收敛困难。为此,本文提出一种基于连续优先级建模的改进

型DDPG列车运行调整算法。该算法将离散发车顺序问题转化为连续优先级分数输出,使策略梯度可在连续空间中优化;并引入状态归一化与奖励缩放机制以提升训练稳定性;同时利用启发式行为克隆预训练模块,为智能体提供初始策略,从而显著加快收敛速度。

1 问题建模与约束条件

为简化建模复杂度并贴合高铁实际运营场景,在构建运行调整模型时提出以下假设:

- ① 列车在区间运行时间采用固定基准值,若发生延误,可在区间允许速度范围内微调,但不得低于线路最小允许运行时间;
- ② 车站股道资源充足,不考虑股道分配冲突;
- ③ 突发事件仅影响部分列车初始延误,后续延误通过运行调整逐步传导,不考虑二次突发事件干扰;
- ④ 列车停站方案固定,仅调整停站时长,不新增或取消停站。

基于上述假设进行问题建模,基础参数及含义如表1所示。

表1 模型基础参数及含义表

Table 1 Basic model parameters and meanings

参数名称	含义
$A_{i,j}^{\text{plan}}$	列车 <i>i</i> 在车站 <i>j</i> 的计划到达时刻
$D_{i,j}^{\text{plan}}$	列车 <i>i</i> 在车站 <i>j</i> 的计划发车时刻
$A_{i,j}^{\text{act}}$	列车 <i>i</i> 在车站 <i>j</i> 的实际到达时刻
$D_{i,j}^{\text{act}}$	列车 <i>i</i> 在车站 <i>j</i> 的实际发车时刻
$T_{i,j \rightarrow j+1}^{\text{run}}$	列车 <i>i</i> 在区间 <i>j</i> 的运行时间
$T_{\text{window}}^{\text{start}}$	天窗开始时间
$T_{\text{window}}^{\text{end}}$	天窗结束时间

1.1 列车运行状态

设线路上共有 N 列列车,区间站点数为 M 。列车运行状态通过计划时刻、实际时刻及延误参数进行量化描述,以下进行具体分析。

① 到达延误定义:列车实际到达时刻与计划到达时刻的差值,如式(1)。

$$\delta_{i,j}^{\text{arr}} = A_{i,j}^{\text{act}} - A_{i,j}^{\text{plan}} \quad (1)$$

② 实际停站时间:列车在车站的实际停留时长,如式(2)。

$$T_{i,j}^{\text{stop,act}} = D_{i,j}^{\text{act}} - A_{i,j}^{\text{act}}, T_{i,j}^{\text{stop,act}} \geq 0 \quad (2)$$

③ 发车延误定义:实际发车时刻与计划发车时刻的差值,列车不得早发车,如式(3)。

$$\delta_{i,j}^{\text{dep}} = D_{i,j}^{\text{act}} - D_{i,j}^{\text{plan}}, \delta_{i,j}^{\text{dep}} \geq 0 \quad (3)$$

④ 区间运行:列车从当前车站出发,经区间到达下一车站,如式(4)。

$$A_{i,j+1}^{\text{act}} = D_{i,j}^{\text{act}} + T_{i,j \rightarrow j+1}^{\text{run}} \quad (4)$$

1.2 约束条件

列车运行调整方案需满足安全、能力、运营等硬性约束, 否则不具备可行性, 具体约束条件如下:

① 安全到站间隔约束: 同一车站相邻到达的两列车 (列车 k 先到站, 列车 l 后到站), 到达时刻差不小于最小安全间隔 $\Delta t_{\min}^{\text{arr}}$, 如式(5)。

$$A_{l,j}^{\text{act}} - A_{k,j}^{\text{act}} \geq \Delta t_{\min}^{\text{arr}} (k < l) \quad (5)$$

② 安全发车间隔约束: 同一车站相邻发车的两列车 (列车 k 先发车, 列车 l 后发车), 发车时刻差不小于最小安全间隔 $\Delta t_{\min}^{\text{dep}}$, 如式(6)。

$$D_{l,j}^{\text{act}} - D_{k,j}^{\text{act}} \geq \Delta t_{\min}^{\text{dep}} (k < l) \quad (6)$$

③ 最小停站时间: 需满足旅客上下车、安全检查等基础作业需求的最短时间, 如式(7)。

$$T_{i,j}^{\text{stop,act}} \geq T_{i,j}^{\text{stop,min}} \quad (7)$$

④ 运行速度约束: 列车区间运行速度需在安全范围内, 不能小于最小运行时间, 如式(8)。

$$T_{i,j \rightarrow j+1}^{\text{run,min}} \leq T_{i,j \rightarrow j+1}^{\text{run}} \quad (8)$$

⑤ 区间不冲突约束: 相邻列车 (列车 k 先到站, 列车 l 后到站) 在当前车站的发车顺序要和在下一站的

到达顺序相同, 禁止区间越行, 如式(9)。

$$(D_{l,j}^{\text{act}} - D_{k,j}^{\text{act}})(A_{l,j+1}^{\text{act}} - A_{k,j+1}^{\text{act}}) > 0 \quad (9)$$

⑥ 天窗不行车约束: 列车的运行过程不得跨越天窗时间窗口 (即不能部分时段在天窗内、部分时段在天窗外), 需满足“全程天窗前运行”或“全程天窗后运行”, 如式(10)。

$$(A_{i,j}^{\text{act}} \leq T_{\text{window}}^{\text{start}}) \cup (D_{i,j}^{\text{act}} \geq T_{\text{window}}^{\text{end}}) \quad (10)$$

1.3 目标函数

核心目标: 在满足上述约束的基础上, 最小化所有列车在沿线各站点的总有效延误时间, 如式(11)。

$$\min F = \sum_{i=1}^N \sum_{j=1}^M [\max(0, \delta_{i,j}^{\text{arr}}) + \max(0, \delta_{i,j}^{\text{dep}})] \quad (11)$$

2 改进型DDPG算法设计与实现

本文基于深度确定性策略梯度 (DDPG) 算法框架, 设计状态归一化、奖励缩放、行为克隆预训练改进机制, 构建了一套完整的算法实现方案, 如图1所示。

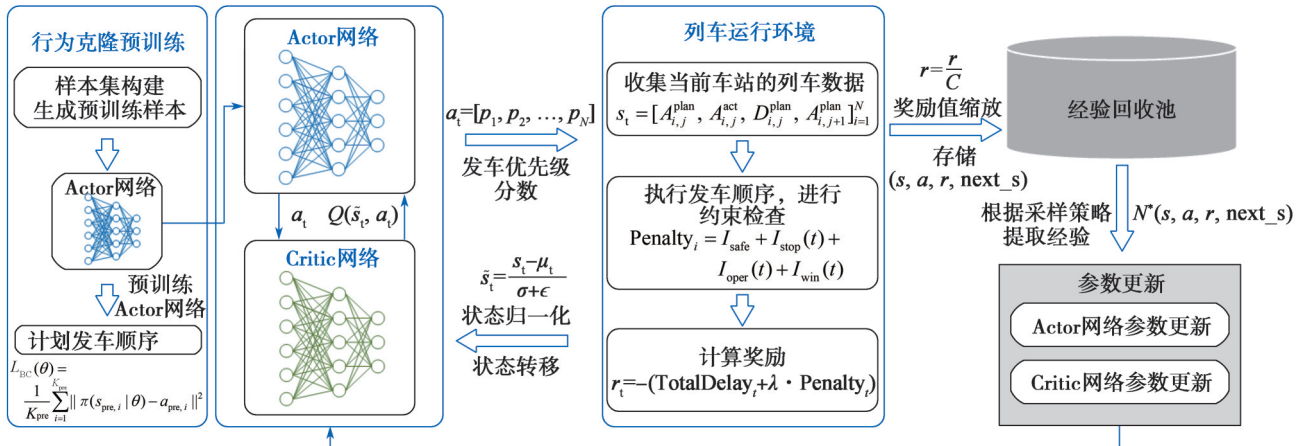


图1 算法架构图

Fig. 1 Architecture diagram of algorithm

整体算法由智能体模块、环境模块、经验回收池及网络更新机制构成, 它们形成了完整的闭环学习体系。在每一阶段, 环境给出状态 s_t , 经状态归一化处理输入智能体。智能体中的Actor网络输出连续优先级动作向量, Critic网络对动作进行价值评估。环境根据Actor输出的优先级顺序执行列车发车与区间运行计算, 得到新的状态和即时奖励, 并将经验存入经验回收池中。在训练过程中, 智能体定期从回收池中采样批量样本进行学习。

为进一步提升模型训练的稳定性与收敛性能, 算法在训练初期引入行为克隆预训练机制, 以原计划发车顺序作为先验指导; 在整个训练阶段采用状态归一化与奖励缩放技术, 以保证输入特征与奖励信号的数值平衡。通过多轮交互与参数更新, 智能体能够逐步

学习至最优发车顺序, 从而实现列车运行调整的自适应优化。

2.1 状态与动作空间设计

状态: 对每一列车构造4维特征向量 $[A_{i,j}^{\text{plan}}, A_{i,j}^{\text{act}}, D_{i,j}^{\text{plan}}, A_{i,j+1}^{\text{plan}}]$, 系统状态 s_t 为 N 列列车的拼接, 如式(12)。

$$s_t = [A_{i,j}^{\text{plan}}, A_{i,j}^{\text{act}}, D_{i,j}^{\text{plan}}, A_{i,j+1}^{\text{plan}}]_{i=1}^N \quad (12)$$

状态归一化: 为稳定训练, 本文采用滑动均值标准化方法对每一维状态进行线性归一化, 以提升训练稳定性。具体而言, 针对每一维状态特征, 实时计算每维均值 μ_t 与标准差 σ_t , 在存储与训练输入时使用标准化, 同时为防止分母为零, 设置极小常数 ϵ , 如式(13)。

$$\tilde{s}_t = \frac{s_t - \mu_t}{\sigma_t + \epsilon} \quad (13)$$

动作：动作向量 a_t 定义为 N 列列车的连续优先级分数，Actor 网络以 $\tilde{s}_t$ 为输入，直接输出该连续动作向量，其中 p_i 为列车 i 的优先级分数（数值越大，发车优先级越高），如式(14)。

$$a_t = [p_1, p_2, \dots, p_N] \quad (14)$$

动作衰减：为增强探索性，训练时在动作向量中加入高斯噪声，避免策略陷入局部最优；训练后期需降低探索性，确保策略稳定，因此设计噪声衰减机制，如式(15)、式(16)。

$$a_t = \pi(\tilde{s}_t | \theta) + \mathcal{N}(0, \sigma) \quad (15)$$

$$\sigma = \sigma \times \text{decay rate} \quad (16)$$

式中： $\pi(\tilde{s}_t | \theta)$ 为 Actor 网络（参数）； $\mathcal{N}(0, \sigma)$ 为均值 0、标准差 σ 的高斯噪声；decay rate 为噪声衰减系数，每一个训练步标准差 σ 都按一定的比例进行衰减。

2.2 奖励函数设计

奖励函数需引导算法最小化总延误，同时惩罚约束违规，确保调整方案可行。因此，即时奖励 r_t 由“总延误惩罚 TotalDelay_t”与“约束违规惩罚 Penalty_t”两部分组成，公式如下：

$$r_t = -(\text{TotalDelay}_t + \lambda \cdot \text{Penalty}_t) \quad (17)$$

式中： λ 为惩罚权重。

式(17)中的 TotalDelay_t 为当前站点所有列车的有效到达延误与发车延误之和，其定义与本文第 1 章目标函数一致，即

$$\text{TotalDelay}_t = \sum_{i=1}^N [\max(0, \delta_{i,j}^{\text{arr}}) + \max(0, \delta_{i,j}^{\text{dep}})] \quad (18)$$

Penalty_t 为约束违规惩罚项，其根据第 1 章定义了几类核心约束，设计多维度惩罚项，确保方案可行性，见式(19)。

$$\text{Penalty}_t = I_{\text{safe}}(t) + I_{\text{stop}}(t) + I_{\text{oper}}(t) + I_{\text{win}}(t) \quad (19)$$

式中：若当前步存在安全到发间隔违规， $I_{\text{safe}}(t)$ 取值为 1，否则取值为 0；若当前步存在停站时间违规， $I_{\text{stop}}(t)$ 取值为 1，否则取值为 0；若当前步存在区间运行时间违规， $I_{\text{oper}}(t)$ 取值为 1，否则取值为 0；若当前步存在天窗行车违规， $I_{\text{win}}(t)$ 取值为 1，否则取值为 0。

为避免奖励值过大导致训练不稳定，对即时奖励进行缩放，如式(20)。

$$r'_t = \frac{r_t}{C} \quad (20)$$

式中： C 为奖励缩放系数，用于调节奖励信号的幅度范围，防止因奖励值过大或过小而影响训练稳定性。

2.3 行为克隆预训练

为解决 DDPG 在训练初期因策略随机导致的探索效率低问题，利用启发式规则（按计划发车时间排序）

生成高质量初始策略，通过行为克隆预训练 Actor 网络，步骤如下：

①预训练样本生成。

样本集构建：生成 K_{pre} 个预训练样本；

启发式策略：对预训练样本中的每个状态生成按原计划发车的发车顺序；

动作映射：将发车顺序转化为动作 $a_{\text{pre},i}$ ，生成对应的优先级分数向量 $[p_1, p_2, \dots, p_N]$ 。

②预训练损失优化。

Actor 网络通过最小化“预测动作”与“启发式动作”的均方误差进行预训练，见式(21)。

$$L_{\text{BC}}(\theta) = \frac{1}{K_{\text{pre}}} \sum_{i=1}^{K_{\text{pre}}} \left\| \pi(s_{\text{pre},i} | \theta) - a_{\text{pre},i} \right\|^2 \quad (21)$$

2.4 算法伪代码

算法详细迭代及训练流程如表 2 所示。

表 2 改进型 DDPG 列车运行调整策略优化算法伪代码

Table 2 Pseudocode of improved DDPG

改进型 DDPG 列车运行调整策略优化算法
输入：计划运行图，区间运行时间，晚点信息
1: 初始化 Actor 网络 $\pi(s \theta)$ 与 Critic 网络 $Q(s,a \phi)$
2: 复制网络参数得到目标网络 π' 和 Q'
3: 初始化经验回收池 D
4: 进行行为克隆预训练
5: 使用启发式排序策略(按计划发车时间)生成目标动作
6: 最小化 MSE 损失预训练 Actor
7: 结束预训练
8: for episode = 1 to MaxEpisode do
9: 重置环境，得到初始状态 s
10: for t = 1 to MaxStep do
11: 归一化状态 s
12: Actor 生成优先级向量
13: 环境执行排序后的调度决策
14: 存储经验至回收池 D
15: 从 D 中随机采样 batch 样本
16: 计算目标值
17: 更新 Critic 参数
18: 更新 Actor 参数
19: 软更新目标网络参数
20: 若所有列车都到达终点，记录当前时刻表，跳出当前循环
21: 动作噪声衰减
22: end for
23: 返回最优策略网络 $\pi(s \theta)$ 和最优调整后时刻表

3 试验设计与结果分析

3.1 试验设置

试验基于京广高铁武广段部分的区间数据，包含

15列列车、11个站点, 线路信息如表3所示。

表3 长沙南至广州区段线路信息表

Table 3 Railway data for Changsha South–Guangzhou section

序号	区间	区间里程/km	最小运行时间/min
1	长沙南—株洲西	52	10
2	株洲西—衡山西	84	15
3	衡山西—衡阳东	41	7
4	衡阳东—耒阳西	55	10
5	耒阳西—郴州西	98	18
6	郴州西—乐昌东	106	19
7	乐昌东—韶关	44	8
8	韶关—英德西	87	16
9	英德西—清远	57	11
10	清远—广州北	36	7

延误场景1: 3号列车在始发站延误20 min。

延误场景2: 2号列车和5号列车在始发站分别延误40 min和30 min。

模型主要参数设置如表4所示。

表4 模型基础参数及含义表

Table 4 Basic model parameters and meanings

参数名称	设置值
Actor学习率	1×10^{-4}
Critic学习率	1×10^{-3}
折扣因子	$\gamma=0.99$
软更新系数	$\tau=0.001$
回收池容量	20 000
批量大小	64
各列车在车站最小停站时间	2 min
最小发车间隔	3 min
最小到达间隔	3 min
奖励缩放常数	$C=100$

3.2 结果分析

1) 场景1结果分析

对于场景1的不同算法奖励值的迭代情况如图2所示, 3类算法(改进后DDPG, 去除预训练改进后DDPG, 原始DDPG)的奖励值迭代呈现显著差异, 分析如下:

改进后DDPG算法: 奖励值快速上升并在迭代前期(约500次内)收敛至较高水平, 后续保持稳定且波动极小。该结果验证了状态归一化、奖励缩放等机制在提升收敛速度与训练稳定性的有效性。

去除预训练的改进后DDPG算法: 奖励值整体低于改进后算法, 且收敛过程中伴随更明显的波动。该结果说明预训练机制可帮助算法初始化更优参数, 降低训练初期的探索盲目性。

原始DDPG算法: 奖励值始终处于3类算法中的最低水平, 且在整个训练过程中未能实现稳定收敛, 验证了原始DDPG难以直接解决列车运行调整决策问题。

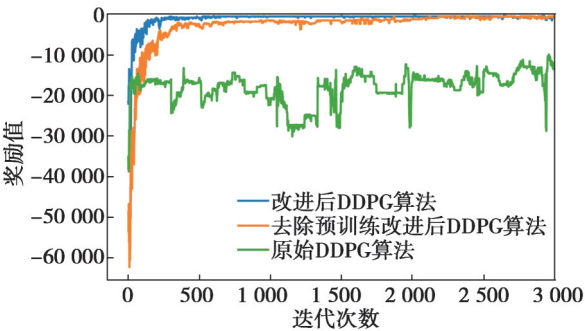


图2 场景1奖励值迭代对比图

Fig. 2 Comparison of reward value iterations in Scenario I

图3为场景1调整后的列车运行图, 实线为计划运行线, 虚线为调整后的运行线。在场景1单列车延误的背景下, 本文算法通过赶点策略使3号列车(虚线)在后续运行中逐步追回延误时间, 并在2号车站越行5号列车, 随后5号列车通过组织赶点在4号车站基本恢复按图运行。整个过程各列车之间在时间维度上保持安全的时间间隔, 避免了冲突。表5为总晚点时间对比表, 从表5可以看出在场景1中, 本文算法相较于传统的“先到先得”和“先计划先发”规则, 总晚点时长分别降低了约67.8%和65.9%, 优化效果显著。

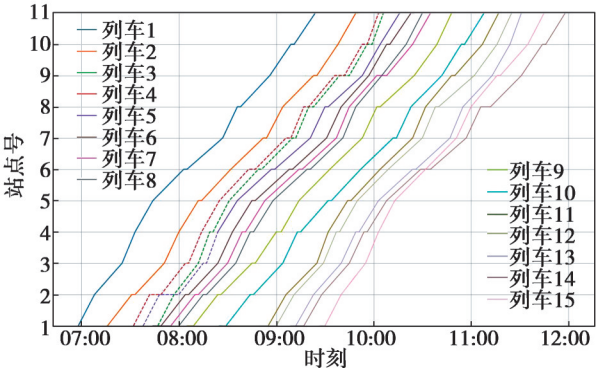


图3 场景1调整后列车运行图

Fig. 3 Adjusted train operation diagram in Scenario I

2) 场景2奖励值迭代分析

对于场景2的不同算法奖励值的迭代情况如图4所示, 相较于场景1, 场景2的晚点情况更为复杂, 3类算法的性能差异进一步凸显, 分析如下:

改进后DDPG算法: 迭代初期快速收敛后保持稳定, 即使在相对复杂的干扰条件下仍能有效学习到接近最优的调整策略。

去除预训练的改进后DDPG算法: 奖励值整体低于改进后算法, 且训练后期仍存在小幅波动, 反映其未能完全收敛。该结果说明预训练机制在复杂场景下对参数初始化的指导作用更关键, 能显著降低高维状态空间下的探索成本。

原始DDPG算法：奖励值在3类算法中持续处于最低水平且始终无法收敛，进一步验证了原始DDPG算法若不经针对性改进，则难以应对高维强约束的列车运行调整问题。

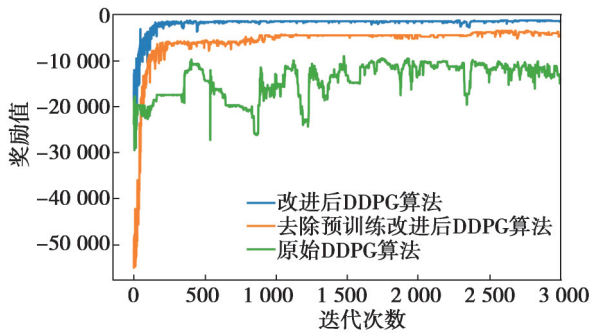


图4 场景2奖励值迭代对比图

Fig. 4 Comparison of reward value iterations in Scenario II

图5为场景2调整后的列车运行图，在多列车延误背景下，本文算法先在1号车站（始发站）安排2号列车优先于6号列车出发，并令6号列车在后续区间赶点，逐步恢复按图运行；与此同时，为配合2号列车和5号列车，3、4、6、7、8、9号列车也在合适区间进行赶点操作，从而在整体上实现了延误的有效吸收与传播抑制，总晚点时间对比见表5。由表5可知，在场景2中相较于“先到先得”和“先计划先发”规则，总晚点时长分别降低了约48.0%和67.1%，进一步验证了其在复杂延误场景下的调整能力。

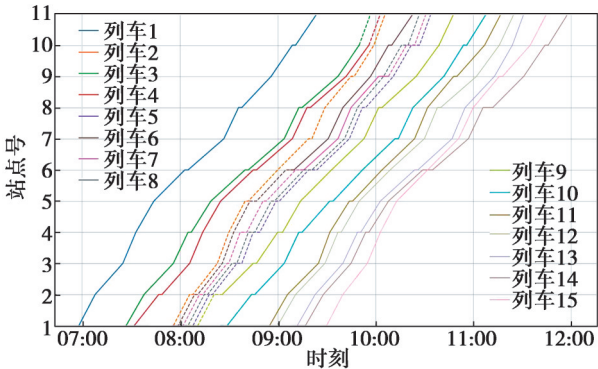


图5 场景2调整后列车运行图

Fig. 5 Adjusted train operation diagram in Scenario II

表5 总晚点时间对比表

Table 5 Comparison of total delay times min

场景	改进DDPG算法	改进DDPG算法 (无预训练)	原始DDPG算法	先到先得	先计划先发
1	375	375	9 967 (不收敛)	1 165	1 101
2	1 221	3 627(未收敛)	9 156 (不收敛)	2 348	3 715

4 结论

本文提出了一种基于改进型深度确定性策略梯度

(DDPG)的高速列车运行调整方法，通过连续优先级动作输出突破传统离散发车顺序优化的局限，结合状态归一化、奖励缩放提升算法稳定性，并引入行为克隆预训练机制加速收敛过程。在算法性能上，改进后的DDPG算法在两类场景中均实现快速收敛且运行稳定，显著优于原始DDPG算法和去除预训练的改进算法，验证了预训练机制在降低探索成本、提升泛化能力方面的关键作用。与“先到先得”“先计划先发”的传统方法相比，场景1总晚点时长分别降低约67.8%、65.9%，场景2分别降低约48.0%、67.1%，充分体现了其在高速铁路运行调整场景的实用性与高效性。

综上所述，本研究为高速列车运行调整提供了一种新的解决方案，未来可进一步探索多智能体协同调整等方向，以适应日益复杂的铁路运输需求。

参考文献：

[1] 林博,俞胜平,刘子源,等.基于改进粒子群算法的高铁列车动态调度[J].控制工程,2021,28(7):1334-1341.
LIN Bo, YU Shengping, LIU Ziyuan, et al. High-speed train dynamic scheduling method based on improved particle swarm optimization algorithm[J]. Control engineering of China, 2021, 28(7): 1334-1341.

[2] 张璞.扰动不确定条件下高速铁路列车运行调整研究[D].北京:北京交通大学,2024.
ZHANG Pu. Research on train timetable rescheduling under uncertain disturbances in a high-speed railway[D]. Beijing: Beijing Jiaotong University, 2024.

[3] 杜心怡,邵长虹.基于强化学习的高速铁路列车运行调整方法研究[J].铁道技术标准(中英文),2024,6(6):35-43.
DU Xinyi, SHAO Changhong. Research on high-speed railway timetable rescheduling based on reinforcement learning [J]. Railway technical standard(Chinese & English), 2024, 6 (6): 35-43.

[4] 段少楠,戴胜华.离散萤火虫算法在高速列车运行调整中的应用[J].计算机工程与应用,2018,54(15):209-213.
DUAN Shaonan, DAI Shenghua. Application of discrete firefly algorithm in high-speed train operation adjustment [J]. Computer engineering and applications, 2018, 54(15): 209-213.

[5] 牛晋财,李茂青,张雁鹏.基于人工鱼群算法的列车运行调整方法研究[J].计算机工程与应用,2019,55(7):259-264.
NIU Jincai, LI Maoqing, ZHANG Yanpeng. Train operation adjustment based on artificial fish swarm algorithm[J]. Computer engineering and applications, 2019, 55(7): 259-264.

[6] 汪臻.基于遗传模拟退火算法的高速列车运行调整问题研究[D].北京:北京交通大学,2019.
WANG Zhen. Study on operation adjustment of high-speed trains based on genetic simulated annealing algorithm[D].

- Beijing: Beijing Jiaotong University, 2019.
- [7] 俞胜平, 韩忻辰, 袁志明, 等. 基于策略梯度强化学习的高铁列车动态调度方法[J]. 控制与决策, 2022, 37(9): 2407-2417.
- YU Shengping, HAN Xincheng, YUAN Zhiming, et al. A policy gradient reinforcement learning algorithm for high-speed railway dynamic scheduling[J]. Control and decision, 2022, 37(9): 2407-2417.
- [8] MURALI P, DESSOUKY M, ORDÓÑEZ F, et al. A delay estimation technique for single and double-track railroads [J]. Transportation research part E: logistics and transportation review, 2010, 46(4): 483-495.
- [9] SHAKIBAYIFAR M, SHEIKHOLESAMI A, CORMAN F. A simulation-based optimization approach to rescheduling train traffic in uncertain conditions during disruptions [J]. Scientia Iranica: international journal of science & technology, 2017, 25(2): 646-662.
- [10] SATO K, TAMURA K, TOMII N. A MIP-based timetable rescheduling formulation and algorithm minimizing further inconvenience to passengers[J]. Journal of rail transport planning & management, 2013, 3(3): 38-53.
- [11] LIU Renming, LI Shukai, YANG Lixing. Collaborative optimization for metro train scheduling and train connections combined with passenger flow control strategy[J]. Omega, 2020, 90: 101990.
- [12] 孟学雷, 程晓卿, 高如虎, 等. 突发传染性疫情条件下高速列车运行调整研究[J]. 铁道科学与工程学报, 2023, 20(4): 1211-1222.
- MENG Xuelei, CHENG Xiaoqing, GAO Ruhu, et al. High-speed train rescheduling under circumstance of infectious epidemic emergency[J]. Journal of railway science and engineering, 2023, 20(4): 1211-1222.
- [13] 庄河, 何世伟, 戴杨铨. 高速铁路列车运行调整的模型及其策略优化方法[J]. 中国铁道科学, 2017, 38(2): 118-126.
- ZHUANG He, HE Shiwei, DAI Yangcheng. Train operation adjustment model and strategy optimization method for high speed railway[J]. China railway science, 2017, 38(2): 118-126.
- [14] 秦智慧, 李宁, 刘晓彤, 等. 无模型强化学习研究综述[J]. 计算机科学, 2021, 48(3): 180-187.
- QIN Zhihui, LI Ning, LIU Xiaotong, et al. Overview of research on model-free reinforcement learning[J]. Computer science, 2021, 48(3): 180-187.
- [15] 王荣笙, 张琦, 张涛, 等. 基于蒙特卡罗树搜索-强化学习的列车运行智能调整方法[J]. 中国铁道科学, 2022, 43(5): 146-156.
- WANG Rongsheng, ZHANG Qi, ZHANG Tao, et al. Intelligent adjustment approach for train operation based on Monte Carlo tree search-reinforcement learning[J]. China railway science, 2022, 43(5): 146-156.
- [16] ŠEMROV D, MARSETIČ R, ŽURA M, et al. Reinforcement learning approach for train rescheduling on a single-track railway[J]. Transportation research part B: methodological, 2016, 86: 250-267.
- [17] 吴越, 袁志明, 代学武, 等. 强化学习算法在高速铁路运营调度中的应用[J]. 铁路计算机应用, 2022, 31(7): 92-98.
- WU Yue, YUAN Zhiming, DAI Xuewu, et al. Application of reinforcement learning algorithm in high speed railway operation scheduling[J]. Railway computer application, 2022, 31(7): 92-98.
- [18] 代学武, 程丽娟, 崔东亮, 等. 基于强化学习的高速列车群运行调整方法[J]. 中国科学(信息科学), 2022, 52(5): 890-906.
- DAI Xuewu, CHENG Lijuan, CUI Dongliang, et al. Rescheduling of high-speed trains: a reinforcement learning approach[J]. Science in China (information sciences), 2022, 52(5): 890-906.
- [19] 庞子帅, 王丽雯, 彭其渊. 基于强化学习的干扰条件下高速铁路时刻表调整研究[J]. 交通运输系统工程与信息, 2023, 23(5): 279-289.
- PANG Zishuai, WANG Liwen, PENG Qiyuan. High-speed railway timetable rescheduling under random interruptions based on reinforcement learning[J]. Journal of transportation systems engineering and information technology, 2023, 23(5): 279-289.
- [20] 代学武, 吴越, 石琦, 等. 基于优先经验回放可迁移深度强化学习的高铁调度[J]. 控制与决策, 2023, 38(8): 2375-2388.
- DAI Xuewu, WU Yue, SHI Qi, et al. A transferable deep reinforcement learning high-speed railway rescheduling method based on prioritized experience replay[J]. Control and decision, 2023, 38(8): 2375-2388.
- [21] 韩忻辰, 俞胜平, 袁志明, 等. 基于Q-learning的高速铁路列车动态调度方法[J]. 控制理论与应用, 2021, 38(10): 1511-1521.
- HAN Xincheng, YU Shengping, YUAN Zhiming, et al. High-speed railway dynamic scheduling based on Q-learning method[J]. Control theory & applications, 2021, 38(10): 1511-1521.

Research on high-speed train operation adjustment based on an improved DDPG algorithm

YANG Linyuan^{1,2}, QIN Yong^{3,4}, WANG Li^{1,2}

(1. School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China; 2. Beijing Engineering Research Center of Urban Traffic Information Intelligent Sensing and Service Technologies, Beijing Jiaotong University, Beijing 100044, China; 3. State Key Laboratory of Advanced Rail Autonomous Operation, Beijing Jiaotong University, Beijing 100044, China; 4. Key Laboratory of Railway Industry for Operational Active Safety Assurance and Risk Prevention and Control, Beijing Jiaotong University, Beijing 100044, China)

Abstract: To address timetable conflicts, operational disorder, and cascading delays caused by sudden train delays in increasingly complex high-speed railway networks, this paper proposes a train operation adjustment method based on an improved Deep Deterministic Policy Gradient (DDPG) algorithm. To overcome convergence difficulties and insufficient decision stability of traditional DDPG in optimizing discrete departure sequences, the proposed method models the departure order as a continuous priority score vector, enabling adaptive and dynamic optimization of departure decisions within a reinforcement learning framework. Furthermore, state normalization and reward scaling mechanisms are introduced to enhance the convergence rate of the model, while a behavior cloning pretraining module is designed for efficient parameter initialization. These improvements allow effective compatibility between discrete permutations and continuous action spaces. To verify the effectiveness of the proposed method, a simulation environment was constructed based on operational data from the Wuhan-Guangzhou section of the Beijing-Guangzhou high-speed railway and two typical delay scenarios were introduced. Comparative experiments were conducted among the improved DDPG, the improved DDPG without pretraining, the original DDPG, and two traditional scheduling strategies, namely "departure by arrival order" and "departure per schedule". The improved DDPG achieved faster convergence and more stable operation in both scenarios than the other algorithms, with significantly superior reward value iteration curves. In Scenario I, the total delay time was reduced by approximately 67.8% and 65.9% compared with the traditional scheduling strategies, respectively; in Scenario II, reductions of about 48.0% and 67.1% were achieved. The algorithm effectively mitigates cascading delays through tactics such as catching up to schedule and overtaking. The results demonstrate that the proposed method can efficiently restore train operation order under safety constraints in various delay scenarios, providing a new intelligent optimization approach for operation adjustment on high-speed railways.

Keywords: train operation adjustment; deep reinforcement learning; DDPG; high-speed railway; behavior cloning pretraining; high-speed train