

文章编号:1004-2954(2026)05-0010-10

基于深度强化学习的高速列车运行节能在线优化

黄子凌, 石红国

(西南交通大学交通运输与物流学院, 成都 611756)

摘要:【目的】高速列车运行节能优化研究对于铁路经济效益和节能减排均有重大意义。【方法】首先,采用划分子区间的方法对线路进行预处理,然后在考虑再生制动、分相区、长陡坡、变限速等多个因素的基础上构建列车节能操纵模型,并将模型转换为基于子区间分步优化的马尔可夫决策过程。为求解马尔可夫决策过程,引入深度强化学习算法,提出一种基于SAC(Soft Actor-Critic)算法的在线节能优化方法。【结果】选取实际线路对算法进行有效性验证,结果表明,在满足约束条件的前提下,SAC算法较PPO(Proximal Policy Optimization)算法具有更快的收敛效率,较粒子群算法与PPO算法的节能效果更佳,节能率达到37.4%。同时,选取线路中若干个位置进行仿真试验,结果表明算法均能实时生成剩余路线的节能运行曲线,验证了算法的在线优化能力。【结论】研究成果为高速列车运行节能优化提供依据。

关键词:高速列车; 节能操纵; 深度强化学习; SAC算法; 在线优化

中图分类号:U238; TP181 **文献标识码:**A **DOI:**10.13238/j.issn.1004-2954.202407290004

Energy-Saving Online Optimization of High-Speed Train Operation Based on Deep Reinforcement Learning

HUANG Ziling, SHI Hongguo

(School of Transportation and Logistics, Southwest Jiaotong University, Chengdu 611756, China)

Abstract: [Objective] The research on energy-saving optimization of high-speed trains is of great significance for the economic benefits of railway operations and energy conservation and emission reduction. [Methods] First, the railway line was preprocessed by subdividing it into several subintervals. Then, considering factors such as regenerative braking, neutral sections, long and steep slopes, and variable speed limits, a train energy-saving control model was constructed. This model was then transformed into a Markov decision process based on stepwise optimization of the subintervals. To solve the Markov decision process, a deep reinforcement learning algorithm was introduced, and an online energy optimization method based on the soft actor-critic (SAC) algorithm was proposed. [Results] The algorithm's effectiveness was validated using an actual railway line, and the results showed that, under the given constraints, the SAC algorithm converged faster than the proximal policy optimization (PPO) algorithm and achieved better energy savings compared to both the particle swarm optimization (PSO) and PPO algorithms, with an energy-saving rate of 37.4%. Additionally, simulations conducted at several locations along the route indicated that the algorithm generated energy-efficient speed profiles for the remaining route in real time, confirming its online optimization capability. [Conclusion] The research findings provide a basis for energy-saving optimization a high-speed train operations.

Key words: high-speed train; energy-saving control; deep reinforcement learning; SAC algorithm; online optimization

收稿日期:2024-07-29; 修回日期:2024-09-06; 网络首发日期:2025-03-11

基金项目:国家重点研发计划项目(2022YFB4300504-1);四川省自然科学基金项目(2022NSFSC0418, 2023NSFSC0386)

作者简介:黄子凌(2000—),男,硕士研究生,主要从事城轨列车与高速列车运行节能优化研究工作,E-mail:978291632@qq.com。

通信作者:石红国(1974—),男,教授,工学博士,主要从事列车节能运行优化研究工作,E-mail:shg@home.swjtu.edu.cn。

引言

高速列车具有高速、安全、准时的突出优点,但是能源消耗巨大。在我国推进碳达峰和碳中和的背景下,优化高速列车运行控制过程从而实现节能减排,具有重要的现实意义和理论价值。

列车节能运行问题的研究于20世纪60年代开始兴起, ICHIKAWA^[1]最早使用庞特里亚金极大值原理研究列车节能操纵策略。PHIL^[2]继续以极大值原理为理论基础,将列车节能运行问题简化为寻找列车工况转换点的优化问题,并给出求解方法,同时证明离散控制的列车最优速度曲线应由最大牵引、惰行与最大制动3种工况组成^[3-6]。LU等^[7]使用动态规划、蚁群算法和遗传算法对列车运行曲线进行优化,并通过仿真证明动态规划在三种算法中的优越性,但文中方法较为简单且忽略了太多线路约束条件。TRIVELLA等^[8]建立一个考虑风向风速的列车运行优化模型,并将线搜索与基于距离-速度网格的动态规划方法相结合,对模型进行求解,但该方法涉及太多假设条件,且目前铁路系统缺少风速风向的测量设备,使其难以在实际中应用。WANG等^[9]基于STS(Space-Time-Speed)离散网格建立列车运行优化模型,并使用动态规划和离散微分动态规划算法对模型进行求解,但该方法难以解决因离散网格间隔过小而带来的计算时间长的问题,因此无法进行在线优化。XING等^[10]提出一种改进的暴力搜索算法求解多目标列车操纵节能问题,并通过试验证明该算法较GA算法和PSO算法具有更好的节能效果,不足之处是该方法难以适用于包含复杂路况的线路。LI等^[11]提出一种具有自适应邻域范围的改进萤火虫算法优化高速铁路列车的运行曲线,但该方法无法应用于在线计算。此外,鸡群算法^[12]、改进差分进化算法^[13]、文化基因算法^[14]等启发式算法也被应用到列车运行节能优化领域。除了动态规划和启发式算法以外,近十余年还陆续出现了混合整数线性规划、伪谱法、序列二次规划等求解列车运行节能优化模型的数值方法^[15-18]。

近年来,随着人工智能的兴起,深度学习、强化学习等机器学习算法也被逐渐应用于解决列车运行曲线优化问题。SU等^[19]将列车节能操纵问题转化为子区间能量分配问题,并使用SAC算法对问题进行求解,通过仿真试验证明该方法相较于传统数值方法拥有更好的节能效果。孟建军等^[20]将Sarsa算法与城轨列车控制节能优化相结合,并通过试验证明其算法在满足舒适性和准时性要求的情况下较传统动态规划方法更节能。周敏等^[21]提出一种基于双深度Q网络算法的列车速度曲线优化方法,并通过实例仿真发现该方法在节能与舒

适度方面优于遗传算法,但文中未展示算法实际收敛效果。LIU等^[22]将列车速度曲线优化与深度强化学习算法相结合,提出一种基于PPO(Proximal Policy Optimization)算法的列车节能驾驶策略,针对列车晚点情况设置算法提前终止条件,降低智能体在前期的探索效果。

前人研究可以总结如下。

(1)目前多数研究仅局限于城市轨道交通,高速列车相较于城轨列车具有更长的站间距离以及更复杂的线路环境,如分相区、长陡坡等。不同于城轨列车的连续型牵引、制动方式,高速列车通常使用分级牵引、制动模式。另外,部分研究未考虑再生制动,其计算结果与列车真实运行情况出入较大。针对这类问题,需要从建模角度出发,在考虑分相区、长陡坡、再生制动等情况下建立相应的节能优化模型。

(2)列车节能操纵问题的算法大多基于离线优化,如各类启发式算法、动态规划算法等,计算速度较慢,无法满足高速列车运行调整下在线计算的需要。针对这类问题,需要从算法的角度出发,寻找一种能满足实时计算的算法。

综上所述,首先对线路进行预处理,即根据线路条件和合理计算长度进行分段,然后在考虑再生制动、分相区、长陡坡、变限速等条件的基础上构建高速列车节能优化模型,并将模型转化为分段顺序优化的马尔可夫决策过程。为求解马尔可夫决策问题,引入深度强化学习,提出一种基于SAC算法的列车运行节能在线优化方法。由于强化学习的本质是通过探索所有可能的列车操纵策略,进而训练出能够自主决策的智能体,因而具有“离线训练、在线优化”的优势,可以满足列车运行调整条件下实时生成新的节能运行曲线的要求,具有较强的现实意义。

1 列车受力分析与运动方程式

机车牵引力取值可由牵引特性曲线得到。采用分段拟合的方式,对原曲线中的离散点进行三段式仿射变换,进而得到机车牵引力与速度的连续分段函数,拟合过程如图1所示。

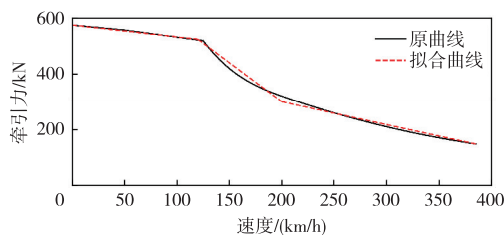


图1 列车牵引特性拟合曲线

Fig. 1 Fitting curves of train traction characteristics

列车运行阻力包含基本阻力和附加阻力,计算公

式如下

$$w_0 = a + bv + cv^2 \quad (1)$$

$$W_f = W_i + W_r + W_j \quad (2)$$

$$W_r = 0.6Mg/R \quad (3)$$

$$W_j = 1.3 \times 10^{-7} L_j Mg \quad (4)$$

$$W_i = Mgsin\theta \quad (5)$$

式中, w_0 为单位基本阻力; a 、 b 、 c 为经验常数; v 为列车运行速度; W_f 为列车附加阻力; W_i 为坡道附加阻力; W_r 为曲线附加阻力; W_j 为隧道附加阻力; M 为列车定员编组质量; g 为重力加速度; R 为曲线半径; L_j 为隧道长度; θ 为坡道夹角。

高速列车制动力通常由空气制动和再生制动两部分构成,其取值可由制动特性曲线得到。空气制动与再生制动切换判断如下

$$W_b = \begin{cases} W_{b1}, & v \leq v_1 \\ W_{b2}, & v > v_1 \end{cases} \quad (6)$$

式中, W_b 为列车制动力; W_{b1} 为空气制动力; W_{b2} 为再生制动力; v_1 为切换制动方式的速度临界值。

根据列车受力分析,可得列车的运动方程式如下

$$v \frac{dv}{dx} = \frac{Q(v, u) - W_b(v, u) - W_0(v) - W_f}{M(1 + \rho)} \quad (7)$$

$$\frac{dt}{dx} = \frac{1}{v} \quad (8)$$

式中, Q 为机车牵引力; u 为牵引/制动极位; ρ 为回转质量系数。

2 节能优化模型构建

2.1 子区间划分

在优化模型构建之前,对区间进行预处理,根据线路条件(限速、附加阻力、分相区)将区间划分成 h 个不等距的子区间,如图 2 所示。子区间附加阻力集合为 $\{W_{f1}, W_{f2}, W_{f3}, \dots, W_{fh}\}$, 限速集合为 $\{v_{lim}^1, v_{lim}^2, v_{lim}^3, \dots, v_{lim}^h\}$, 路程集合为 $\{x_1, x_2, x_3, \dots, x_h\}$ 。同时,对于每个子区间,设定一个列车通过子区间末端时的目标速度,目标速度集合为 $\{v_1, v_2, v_3, \dots, v_h\}$ 。

线路分段的目的在于方便后续构建基于分步推进

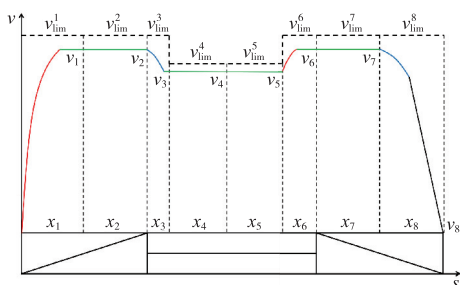


图 2 子区间划分过程

Fig. 2 Process of subinterval divisions

优化的模型以及马尔可夫决策过程,从而引入强化学习方法。另外,基于线路分段的分布推进优化模型,可以使强化学习算法在状态转移过程中捕捉到列车位置变化的信息,从而使得训练出来的模型具有在线优化的能力。例如,文献[19]虽然利用强化学习的方法对列车操纵进行节能优化,但采用的子区间能量分配模型在状态转移过程中无法体现列车位置变化的过程,其本质是对整条速度曲线进行迭代优化,仍属于离线优化的范畴,无法在列车运行途中根据实际情况对节能运行策略进行在线调整。

2.2 子区间操纵策略

对处于分相区的子区间,列车操纵策略只能选择惰行工况。对处于非分相区的子区间,为兼顾不同的线路条件,子区间操纵策略包含以下多种情况。

$$(1) v_{i-1} = v_i$$

列车应选择巡航工况,匀速通过子区间。

$$(2) v_{i-1} < v_i$$

如图 3(a)所示,由于列车运行节能的本质是使列车尽可能惰行,因此,根据列车是否能尽量通过惰行加速至目标速度,可将子区间内的操纵策略设置为“牵引+巡航”“惰行+巡航”和“惰行+牵引”3 种情形,牵引级位的选择应在满足舒适度(加速度变化率)限制的前提下,以惰行距离尽可能长为目标。

$$(3) v_{i-1} > v_i$$

根据列车是否能尽量通过惰行减速至目标速度,可将子区间内的操纵策略设置为“制动+巡航”“惰行+巡航”和“惰行+制动”3 种情形,制动级位的选择应在满足舒适度(加速度变化率)限制的前提下以惰行距离尽可能长为目标,如图 3(b)所示。值得注意的是,“制动+巡航”只适用于列车处于陡下坡的情况,此时由于列车惰行无法降速,因此选择先制动后巡航。

2.3 列车运行时间、距离、能耗计算

将列车分为分级牵引、惰行、巡航、分级制动 4 种工况,并对列车在某一子区间 4 种不同工况下的运行时间、距离和能耗的计算进行分析。

2.3.1 牵引工况

牵引工况计算如下

$$t = \int_{v_{i1}}^{v_{i2}} \frac{M(1 + \rho)}{Q(v, u) - W_0(v) - W_{fi}} dv \quad (9)$$

$$x = \int_{v_{i1}}^{v_{i2}} \frac{M(1 + \rho)v}{Q(v, u) - W_0(v) - W_{fi}} dv \quad (10)$$

$$E = \frac{1}{\eta} \int_{v_{i1}}^{v_{i2}} \frac{vQ(v, u)M(1 + \rho)}{Q(v, u) - W_0(v) - W_{fi}} dv \quad (11)$$

式中, t 为运行时间; x 为运行距离; E 为能耗; W_0 为列车基本阻力; W_{fi} 为第 i 个子区间的附加阻力;

η 为能耗利用率; v_{i1} 、 v_{i2} 为计算速度上下限。

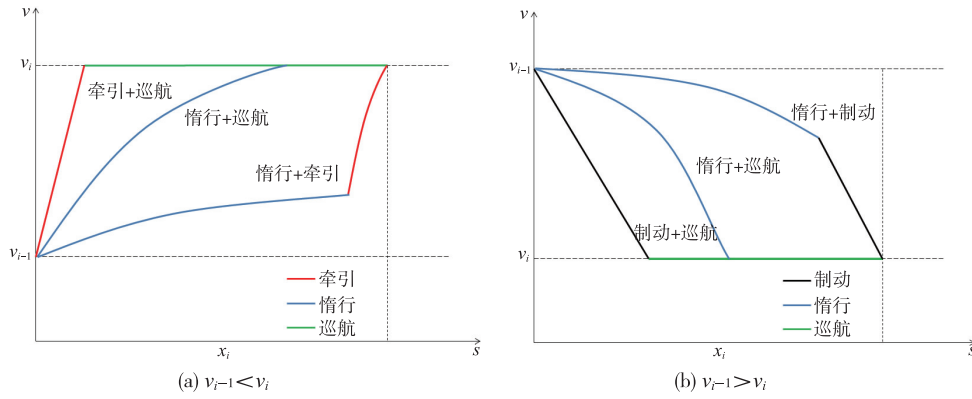


图3 子区间操纵策略

Fig.3 Subinterval control strategy

2.3.2 惰行工况

惰行工况下, 列车运行时间、距离和能耗计算如下

$$t = \left| \int_{v_{i1}}^{v_{i2}} \frac{M(1+\rho)}{W_0(v) + W_{fi}} dv \right| \quad (12)$$

$$x = \left| \int_{v_{i1}}^{v_{i2}} \frac{M(1+\rho)v}{W_0(v) + W_{fi}} dv \right| \quad (13)$$

$$E = 0 \quad (14)$$

2.3.3 巡航工况

巡航工况下, 当巡航距离已知时, 列车运行时间、能耗如下

$$t = \frac{x}{v} \quad (15)$$

$$E = \begin{cases} \left\{ \frac{1}{2} M |v_{i1}^2 - v_{i2}^2| - \int_{v_{i1}}^{v_{i2}} \frac{(W_0(v) + W_{fi})M(1+\rho)v}{W_b(v, u) + W_0(v) + W_{fi}} dv \right\} \lambda, & \min\{v_{i1}, v_{i2}\} > v_l \\ 0, & \min\{v_{i1}, v_{i2}\} \leq v_l \end{cases} \quad (19)$$

式中, λ 为再生制动能耗回收率。

2.4 基于子区间操纵策略的节能优化模型

子区间操纵策略使得在已知子区间目标速度的情况下可获得到该子区间内的列车操纵策略, 因此, 整个区间的列车最优运行策略将转换为所有子区间最优目标速度的集合。

节能模型的目标函数是使所有子区间能耗之和最低, 同时引入运行时间误差、限速、舒适度等其他约束, 建立模型如下

$$\begin{aligned} \min E &= \sum_{i=1}^h E_i(v_i) \\ &\begin{cases} 0 \leq v_i \leq \min(v_{lim}^i - v_{safe}, v_{lim}^{i+1} - v_{safe}) \\ v_h = 0 \\ T - \Delta T \leq \sum_{i=1}^h t_i \leq T + \Delta T \\ \frac{da}{dt} \leq \xi \quad i = 1, 2, \dots, h-1 \end{cases} \end{aligned} \quad (20)$$

$$E = \begin{cases} \frac{1}{\eta} (W_0(v) + W_{fi})x, & W_0(v_i) + W_{fi} \geq 0 \\ 0, & W_0(v_i) + W_{fi} < 0 \end{cases} \quad (16)$$

2.3.4 制动工况

制动工况下, 列车的运行时间、距离和回收能耗计算如下

$$t = \left| \int_{v_{i1}}^{v_{i2}} \frac{M(1+\rho)}{W_b(v, u) + W_0(v) + W_{fi}} dv \right| \quad (17)$$

$$x = \left| \int_{v_{i1}}^{v_{i2}} \frac{M(1+\rho)v}{W_b(v, u) + W_0(v) + W_{fi}} dv \right| \quad (18)$$

式中, E 为列车总能耗; E_i 为第 i 个子区间的列车运行能耗; v_{safe} 为列车运行速度与限速的安全距离; T 为规定运行时间; ΔT 为允许时间误差; t_i 为第 i 个子区间的列车运行时间; a 为列车加速度; $\frac{da}{dt}$ 为乘客舒适度; ξ 为舒适度限制参数。

3 马尔可夫决策过程与强化学习环境

3.1 马尔可夫决策过程构建

列车运动过程是一系列控制决策与运动状态的序列集合。对列车通过的线路进行适当分段, 可以得到唯一的列车控制决策和对应的运动状态。通过构造科学的优化方法, 对这些线路分段内控制决策进行优化, 可以得到列车节能运行状态, 从而获得列车节能运行曲线。

列车在通过某一个子区间时, 其状态变化主要取决于当前状态(位置、速度等)和下一个子区间的目标

速度、限速等参数。列车状态的变化和控制决策类似马尔可夫决策,因此,可将列车运行控制过程转换为马尔可夫决策过程。由于列车运行控制涉及的参数众多,如牵引工况及级位、运行阻力、线路附加阻力、限速等,随着线路长度的增加,列车运行过程决策数量和难度将大幅度增加。传统的马尔可夫决策算法,如动态规划,在求解该类问题时的难度较大,计算时间很难满足高速列车实时运行调整以及ATO驾驶自动快速调整的需要^[10,23]。

综上所述,将强化学习方法运用到马尔可夫决策过程中,建立基于强化学习的马尔可夫决策优化方法,依靠科学的奖惩机制,优化分段决策过程,最终实现列车运行全程的节能操控。基于强化学习的马尔可夫决策过程表示如下

$$MDP = \langle S, A, P, R, \gamma \rangle \quad (21)$$

式中, S 为列车所处的状态; A 为选取的动作,即下一个子区间的目标速度; P 为状态转移概率; R 为奖励值,代表选取动作的好坏; γ 为强化学习中的折扣因子。

3.2 基于强化学习的列车运行环境

3.2.1 列车运行状态和动作

根据强化学习理论构建智能体交互环境,将列车运行状态的调整过程以状态空间和动作空间来描述。其中,状态空间 $S(s_t)$ 设置如下

$$s_t = (n_t, v_t, T_t, E_t) \quad (22)$$

式中,当前状态下列车正处于第 n_t 个子区间起始

处, v_t, T_t, E_t 分别为列车当前速度、运行时间以及累计能耗。设定初始状态 $s_0 = (1, 0, 0, 0)$,终止状态 $s_k = (n_k, v_k, T_k, E_k)$,其中, $n_k = h$ 。

动作空间 $A(a_t)$ 设置如下

$$a_t = (v'_t) \quad (23)$$

式中, v'_t 为下一个子区间的目标速度。通过前文提出的子区间操纵策略,列车选择不同的 v'_t 将转移至不同的状态。

3.2.2 不可行动作的处理

在正常情况下,列车在某一个状态选择某种动作将到达下一个状态,但在实际情况中,某些动作将产生错误行为,例如超速、分相区未惰行等。针对这些不可行动作,以往研究通常在智能体选择这些动作时提前终止当前训练回合,该方法虽然避免了训练过程中错误数据的产生,但也使得智能体在前期训练过程中由于难以产生有效数据而降低算法收敛效率。

因此,针对不可行的动作,采用“暂时容许”的方式以提高数据利用率。对处于非分相区子区间的列车,不可行的动作是指超速的目标速度或无法达到的目标速度,此时智能体将使列车选择能到达的极限目标速度通过下一个子区间,如图4(a)所示。对处于分相区的列车,不可行的动作是指除惰行以外的所有动作,此时智能体将使列车强制惰行通过子区间,如图4(b)所示。图中, v_{i-1} 为当前状态的速度; v_i 为下一个状态的目标速度; v_{lim} 为限速; v_{error} 为错误目标速度,即不可行动作。

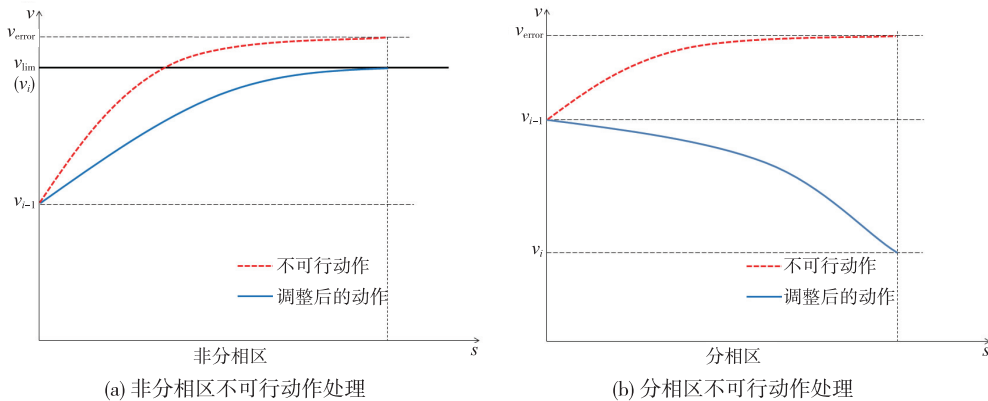


图 4 不可行动作的处理

Fig. 4 Handling of infeasible actions

3.2.3 奖励函数设置

奖励函数是在强化学习环境中用来评估智能体在某一状态下选择某个动作好坏的标准,奖励函数的设置很大程度上影响了算法收敛效率和计算效果。在列车未到达最终状态时,对不可行动作进行惩罚,具体奖励函数设置如下

$$r_t = \begin{cases} 0, & a_t \text{ 可行} \\ -1, & a_t \text{ 不可行} \end{cases} \quad (24)$$

当智能体到达最终状态时,将设置一个额外奖励用来评估列车在整个区间的运行效果。由于最后一个子区间,即第 h 个子区间的目标速度为0,因此,当列车无法在该子区间减速至0时,额外奖励设置为0。

当列车能在该子区间减速至0时,额外奖励设置如下

$$r_{\text{last}} = \begin{cases} \frac{1}{\varepsilon_1 |T_h + t_h - T| + \varepsilon_2 E_h} + \frac{1}{\varepsilon_3 \Delta T}, & |T_h + t_h - T| \leq \Delta T \\ \frac{1}{\varepsilon_3 |T_h + t_h - T|}, & |T_h + t_h - T| > \Delta T \end{cases} \quad (25)$$

式中, r_{last} 为额外奖励; $T_h + t_h$ 为列车实际运行时间; T_h 为列车处于第 h 个子区间起始处时的累计运行时间,可由最终状态得到; t_h 为第 h 个子区间的运行时间; ΔT 为允许运行时间误差; $\varepsilon_1, \varepsilon_2, \varepsilon_3$ 为平衡系数,用来控制奖励大小在整体上为同一量级。由上式可知,当列车运行时间误差较大时,额外奖励只与时间误差大小有关,只有当时间误差小于一定阈值后,才考虑能耗对奖励的影响。这样做的目的是使算法在早期训练过程中能够专注于降低时间误差,尽早得到次优解,从而加快收敛效率。

4 基于 SAC 算法的列车运行节能在线优化

4.1 SAC 算法

SAC 算法是一种基于最大熵思想的离线策略深度强化学习算法,相较于其他强化学习算法, SAC 的目标除了使总体奖励最大以外还会尽可能最大化熵,以增强智能体的探索能力,避免陷入局部最优解^[24]。因此, SAC 算法的目标函数定义为^[24-25]

$$\pi^* = \arg\max_{\pi} \mathbb{E}_{\pi} \left[\sum_t r_t(s_t, a_t) + \alpha H(\pi(\cdot | s_t)) \right] \quad (26)$$

式中, $\pi(\cdot | s_t)$ 为状态 s_t 下的动作概率,代表策略; $H(\cdot)$ 为熵; α 为温度系数,是一个控制熵重要程度的正则系数。在强化学习中, Q 网络用来评估每个状态下对应动作的价值,同时策略网络将根据 Q 网络进行策略更新。值得注意的是, SAC 算法中使用了 2 个 Q 网络,每一次 Q 网络与策略网络的更新都会选取 2 个 Q 网络中的较小值,这样可以避免 Q 值的过高估计。对于任意 Q 网络的损失函数表示如下^[25]

$$L_Q(\omega_j) = \mathbb{E}_{(s_t, a_t, r_t, s_{t+1}) \sim R, a_{t+1} \sim \pi_{\theta}(\cdot | s_{t+1})} \left[\frac{1}{2} (Q_{\omega_j}(s_t, a_t) - (r_t + \gamma (\min_{j=1,2} Q_{\omega_j}(s_{t+1}, a_{t+1}) - \alpha \ln \pi(a_{t+1} | s_{t+1})))^2 \right] \quad (27)$$

式中, R 为经验回放池中的数据; θ 为策略网络参数; $\omega_j (j=1,2)$ 为两个 Q 网络的参数; Q_{ω_j} 为两个 Q 网络对应的目标 Q 网络; γ 为折扣因子。

同时,策略网络将根据两个 Q 网络进行更新,更新方式为最小化损失函数^[24-25]

$$L_{\pi}(\theta) = \mathbb{E}_{s_t \sim R, a_t \sim \pi_{\theta}} [\alpha \ln \pi(a_t | s_t) - \min_{j=1,2} Q_{\omega_j}(s_t, a_t)] \quad (28)$$

最后,对于温度系数 α 也同样存在一个网络,通过该网络对 α 进行不断调整,动态改变熵在训练过程中的重要程度, α 网络的损失函数表示为^[25]

$$L(\alpha) = \mathbb{E}_{s_t \sim R, a_t \sim \pi_{\theta}} [-\alpha \ln \pi(a_t | s_t) - \alpha H^-] \quad (29)$$

式中, H^- 为目标熵。

神经网络将采用梯度下降的方法对参数进行更新^[24-25],具体更新方式如下

$$\omega_j \leftarrow \omega_j + \sigma \nabla_{\omega_j} L_Q(\omega_j) \quad (j=1,2) \quad (30)$$

$$\theta \leftarrow \theta + \sigma \nabla_{\theta} L_{\pi}(\theta) \quad (31)$$

$$\alpha \leftarrow \alpha + \sigma \nabla_{\alpha} L(\alpha) \quad (32)$$

$$\omega_j^- \leftarrow \tau \omega_j + (1 - \tau) \omega_j^- \quad (j=1,2) \quad (33)$$

式中, σ 为神经网络学习率; τ 为一个小于 1 的正数。值得注意的是,不同于其他网络参数, ω_j^- 将通过软更新的方式进行更新^[24-25],以保持训练的稳定性。

4.2 基于 SAC 算法的列车运行节能在线优化

将 SAC 算法与列车运行节能优化相结合,可得到基于 SAC 算法的列车运行节能在线优化算法,如图 5 所示。算法具体步骤如下。

(1) 初始化参数。初始化神经网络参数及强化学习算法参数。

(2) 初始化列车运行状态。

(3) 判断列车是否到达最终状态。若到达最终状态,则转至步骤(8)。

(4) 计算状态转移过程。记录当前列车运行状态 s_t , 根据策略网络选取当前动作 a_t , 列车获取下一个子区间的目标速度; 根据子区间操纵策略, 列车运行通过该子区间, 并转移至新的状态 s_{t+1} ; 根据式(24)与式(25)计算当前动作奖励 r_t 。

(5) 存储状态转移过程。通过经验回放池存储状态转移过程 (s_t, a_t, r_t, s_{t+1}) , 若经验回放池样本数量小于最小样本容量, 则转至步骤(3)。

(6) 经验回放池取样。从经验回放池中随机取出固定数量的样本进行训练。

(7) 更新神经网络参数。根据式(27)~式(29)分别计算各个网络的损失函数, 并通过式(30)~式(33)对每个网络参数进行更新。该步骤完成后转至步骤(3)。

(8) 判断训练是否结束。若当前训练次数小于最大训练次数, 则转至步骤(2)。

(9)结束训练。

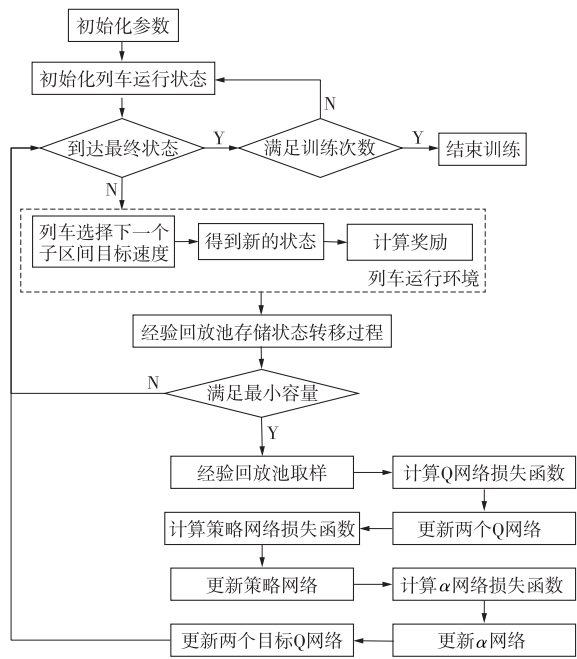


图 5 算法流程

Fig. 5 Flowchart of algorithm

5 算例分析

5.1 基本数据及参数

选取佛坪—洋县西作为实例线路进行仿真,线路全长 53 km,含 8 个下坡道、4 个上坡道、4 段曲线线路、2 个隧道、3 处变限速以及 3 个分相区。子区间数量设置为 30 个,运行时间约束为 1 020 s,允许时间误差设置为 25 s,列车运行速度与限速的安全距离设置为 8 km/h。高速列车选取 CRH380AL,列车参数见表 1,强化学习算法参数见表 2。整个仿真过程在 Python3. 7 运行环境下进行,数学计算部分基于 NumPy、SciPy 等包编程,画图部分基于 pandas、Matplotlib 等包生成,强化学习算法基于 PyTorch 框架进行搭建。

表 1 CRH380AL 参数

Tab. 1 Parameters of CRH380AL

参数	数值
定员质量/t	890
编组方式	14M2T
最高运营速度/(km/h)	385
最大牵引力/kN	640
最大再生制动力/kN	620
单位基本阻力公式/(N/kN)	$0.55+0.0045v+0.00013v^2$
能耗利用率	0.9
再生制动能耗回收率	0.85
回转质量系数	0.06
舒适度限制参数	1
制动方式切换点/(km/h)	15

表 2 算法参数

Tab. 2 Parameters of algorithm

参数	数值
经验回放池容量	1.0×10^5
经验回放池单次采用样本数量	64
经验回放池最小样本容量	1 000
训练回合次数	3 000
神经网络学习率	3.0×10^{-4}
折扣因子	0.99
软更新参数	0.005
奖励平衡系数(ϵ_1)	1.0×10^{-3}
奖励平衡系数(ϵ_2)	1.0×10^{-4}
奖励平衡系数(ϵ_3)	1.0×10^{-3}

5.2 算法有效性验证

使用选取的线路对本文提出的算法进行有效性验证,同时选取 PPO 算法^[22]和粒子群算法(PSO)^[26]与本文算法进行对比分析,PPO 与 PSO 算法的具体流程及参数见文献[22]与文献[26]。算法收敛过程以及计算得到的节能运行曲线如图 6~图 8 所示。从图 6(a)可以看出,SAC 算法在训练初期存在负奖励的情况,这是因为智能体在最初的探索过程中容易出现超速、分相区未惰行等错误行为。随着训练次数的增多,智能体从错误中汲取到教训,慢慢学会避免错误行为的发生,使奖励逐渐变为正数。当训练次数达到 500 次左右时,算法逐渐开始收敛,并能获得较高的奖励回报。图 6(c)记录了 PPO 算法与 SAC 算法 20 次平均奖励值的收敛过程,由于最大熵思想的影响,SAC 算法具有较高的探索能力,相较于 PPO 算法,能够在更少的训练回合内获得更高的奖励值。图 7 单独记录了 SAC 算法计算得到的节能运行曲线,对于线路中的 3 个分相区,列车均准确切换为惰行工况,满足分相区惰行需求。对于 48.85~53 km 的变限速路段,列车做到提前减速以顺利通过线路。此外,列车在 5.7~14.7 km 的长陡下坡路段采用惰行工况,利用下坡势能进行加速,避免了牵引导致的能耗,在 40.2~47.1 km 的两个长陡下坡路段,列车采用先惰行加速再制动回收能耗的方式,使列车在保证速度需求的同时最大程度地节约能耗,符合列车节能操纵驾驶经验。

表 3 为在不同列车预定运行时间下使用 SAC 算法得到的仿真数据,可以看出,对于不同的运行时间需求,本文算法均能在满足时间约束的前提下获得相应的节能操纵策略。表 4 记录了不同算法的仿真结果数据,在运行时间方面,所有方案之间并无太大区别,均满足时间约束;在能耗方面,SAC 计算结果最优,能耗仅有 382.3 kW·h,相较于未优化方案,节能率达到 37.4%;在计算实时性方面,由于 SAC 算法与 PPO 算法均为强化学习算法,能做到离线训练、在线优化,即智能体在离线训练完成后,能在极短时间内通过训练

好的神经网络在线计算得到结果,二者的在线优化能力均远超 PSO 算法。

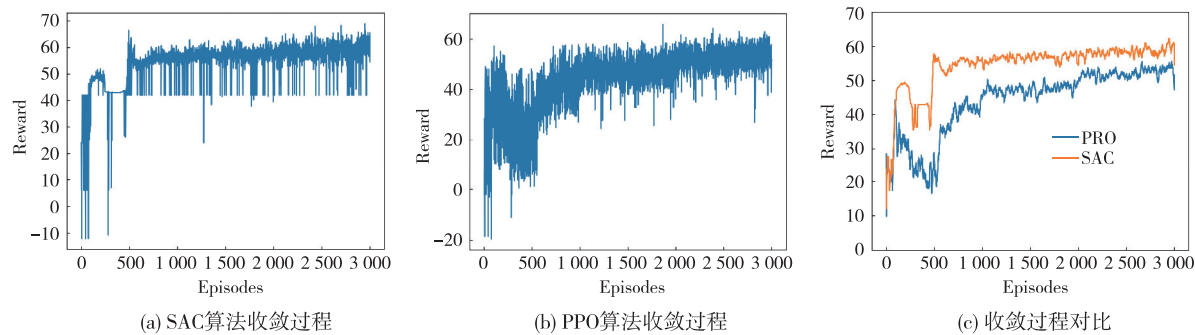


图 6 算法收敛过程
Fig. 6 Convergence process of algorithm

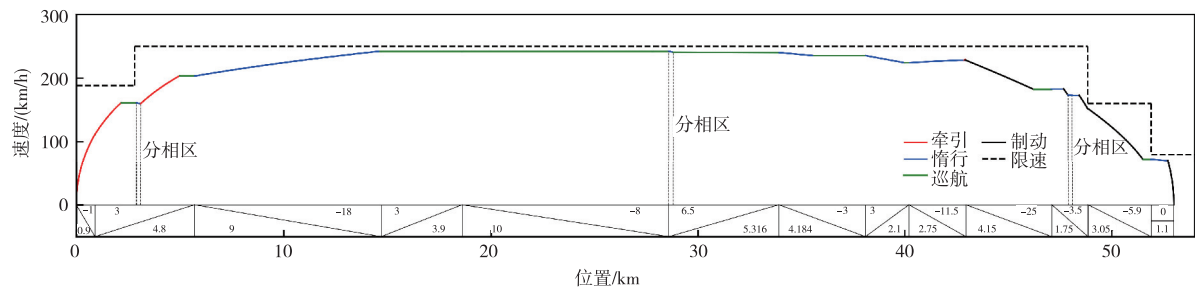


图 7 节能运行曲线
Fig. 7 Energy-saving speed profile

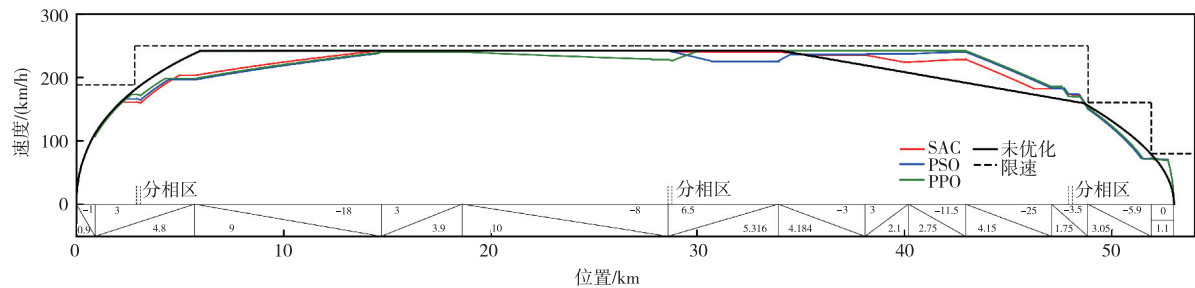


图 8 不同算法计算得到的节能运行曲线
Fig. 8 Energy-saving speed profiles calculated by different algorithms

表 3 基于不同预定运行时间的计算结果

Tab. 3 Results based on different scheduled trip times

预定运行时间/s	实际运行时间/s	能耗/(kW·h)
1 000	1 000.5	438.3
1 010	1 010.4	412.7
1 020	1 020.6	382.3
1 030	1 031.1	363.2
1 040	1 040.2	350.8

表 4 基于不同算法的仿真结果

Tab. 4 Simulation results based on different algorithms

算法	列车运行时间/s	能耗/(kW·h)	节能率/%	计算时间/s
未优化	1 020.0	610.3	0	—
SAC	1 020.6	382.3	37.4	0.1
PPO	1 023.0	411.1	32.6	0.1
PSO	1 021.9	458.9	24.8	53.2

5.3 算法在线性能验证

列车在实际运行过程中,由于司机无法完全精准地按照预定速度曲线进行驾驶,导致实际操作与预先设定的节能操纵策略发生偏差,此时需要根据当前速度、剩余时间、剩余路程等已知信息,在线生成新的节能运行曲线。以往研究往往是基于离线的角度对速度曲线进行优化,无法满足人工驾驶和自动驾驶的实时切换,而强化学习的本质是探索所有的操纵策略,当智能体训练结束后,算法能通过已经训练好的神经网络,针对不同状态生成当前状态下的最优动作,从而能够做到离线训练、在线优化。

首先,在选取的实际线路环境中通过本文提出的深度强化学习算法进行离线训练,训练结束后,任意选取 3 个线路中的位置作为人工转自动驾驶的切换点,

在切换点之前的操纵策略由人为模仿司机驾驶得到,切换点之后则通过已经训练好的网络在线生成剩余路线的节能运行曲线,如图 9 所示,计算所得数据如表 5 所示。

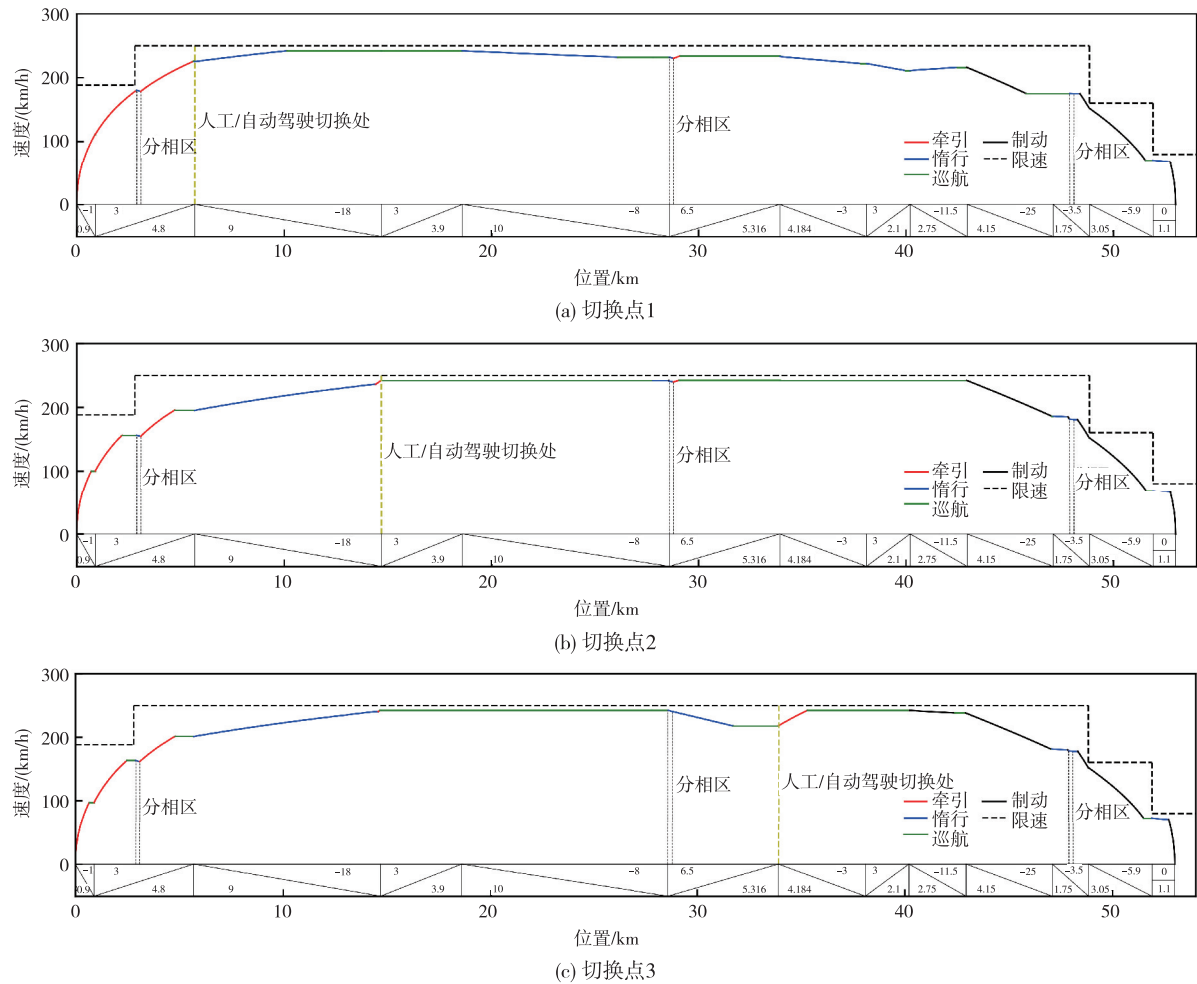


图 9 基于不同人工/自动驾驶切换点的节能运行曲线

Fig. 9 Energy-saving speed profiles based on different manual/autonomous driving switching positions

表 5 基于不同人工/自动驾驶切换点的仿真结果

Tab.5 Simulation results based on different manual/autonomous driving switching points

参数	切换点 1	切换点 2	切换点 3
切换点前运行时间/s	150.8	309.4	602.0
切换点后运行时间/s	871.3	712.1	429.4
总运行时间/s	1 022.1	1 021.5	1 031.4
切换点前能耗/(kW·h)	642.3	529.6	814.9
切换点后能耗/(kW·h)	-186.1	-114.3	-381.7
总能耗/(kW·h)	456.2	415.3	433.2

由图 9 可知,3 种情况下的列车运行曲线均满足分相区惰行以及未超限速的要求。由图 9(a)可知,在切换点 1 之前,由于司机选择最快速操纵策略,切换点之后存在充足的剩余运行时间,因此,在线生成的操纵策略使得列车在之后的运行过程中充分利用下坡惰行,在满足运行时间约束的前提下尽可能地节约能耗。

由图 9(b)可知,在切换点 2 之前,司机选择多个巡航以及惰行工况,导致切换点前列车平均速度较低,切换点之后的剩余运行时间较少,因此,在线策略为满足列车的运行时间约束,操纵策略为尽可能地以最高速保持巡航。由图 9(c)可知,在切换点 3 之前,由于司机存在惰行降速的行为,切换点后,算法为了提速选择先牵引加速然后再惰行与制动交替降速,这样可以尽可能地减少运行时间误差。由表 5 可知,3 种切换点情况下,列车均满足运行时间约束。从能耗的角度来看,最优的操纵策略并非是前期全力牵引,因为前期的过度牵引会减少 5.7~14.7 km 处长陡下坡路段的惰行距离,从而无法充分利用惰行节能,因此,切换点 1 对应的操纵策略在 3 种情况中能耗最高。从运行时间角度来看,由于在切换点 3 之前,列车进行了惰行降速,因此,即使列车在之后的运行过程中立刻进行牵引加速,但总运行时间仍然较预计运行时间增加 11.4 s。

6 结论

在列车实际运行过程中,会因临时限速、运行时分改变等突发情况进行列车运行调整,导致实际操纵与预定运行曲线发生偏差,此时,列车的操纵方案或ATO控制曲线需要进行快速调整。运用传统算法,如遗传算法、动态规划等,很难满足实时计算的需要。

本文提出的基于SAC算法的列车运行节能在线优化方法,能够充分发挥强化学习“离线训练、在线优化”的优点,当列车运行发生调整时,可基于当前运行条件快速生成新的列车节能驾驶曲线。以佛坪—洋县西为实例线路进行验证,本文算法能适应不同路况(分相区、变限速、长陡坡)的线路。同时,将算法与PSO算法和PPO算法进行比较,结果表明SAC算法收敛效率高、节能效果最优,节能率达到37.4%,计算速度可以满足实时输出最优运行曲线的需要。

在未来的研究中,可将本文的建模方法与求解算法拓展至多区间的列车运行节能优化。同时,进一步探讨子区间划分距离及子区间操纵策略对算法求解精度的影响,通过对二者的合理设置,在尽可能保证算法训练效率的前提下进一步提升求解精度。

参考文献:

- [1] ICHIKAWA K. Application of Optimization Theory for Bounded State Variable Problems to the Operation of Train[J]. Bulletin of JSME, 1968,11(47):857-865.
- [2] PHIL H. An Optimal Strategy for the Control of a Train[J]. The Journal of the Australian Mathematical Society Series B Applied Mathematics, 1990,31(4):454-471.
- [3] PHIL H. Optimal Strategies for the Control of a Train[J]. Automatica, 1996,32(4):519-532.
- [4] CHENG J X, HOWLETT P. A Note on the Calculation of Optimal Strategies for the Minimization of Fuel Consumption in the Control of Trains[J]. IEEE Transactions on Automatic Control, 1993,38(11):1730-1734.
- [5] HOWLETT P G, CHENG J. Optimal Driving Strategies for a Train on a Track with Continuously Varying Gradient[J]. The Journal of the Australian Mathematical Society Series B Applied Mathematics, 1997,38(3):388-410.
- [6] PHIL H. The Optimal Control of a Train[J]. Annals of Operations Research, 2000,98(1):65-87.
- [7] LU S F, HILLMANSEN S, HO T K, et al. Single-train Trajectory Optimization[J]. IEEE Transactions on Intelligent Transportation Systems, 2013,14(2):743-750.
- [8] TRIVELLA A, WANG P L, CORMAN F. The Impact of Wind on Energy-efficient Train Control[J]. EURO Journal on Transportation and Logistics, 2021,10:100013.
- [9] WANG X K, TANG T, SU S, et al. An Integrated Energy-efficient Train Operation Approach Based on the Space-time-speed Network Methodology[J]. Transportation Research Part E: Logistics and Transportation Review, 2021, 150: 102323.
- [10] XING Z Y, ZHANG Z Y, GUO J, et al. Rail Train Operation Energy-saving Optimization Based on Improved Brute-force Search[J]. Applied Energy, 2023, 330: 120345.
- [11] LI W, ZHAO S Z, LI K, et al. GSOANR-based Multi-objective Train Trajectory Optimization[J]. International Journal of Rail Transportation, 2024,12(4):733-748.
- [12] HE D Q, LU G C, YANG Y J. Research on Optimization of Train Energy-saving Based on Improved Chicken Swarm Optimization[J]. IEEE Access, 2019,7:121675-121684.
- [13] LU G C, HE D Q, ZHANG J L. Energy-saving Optimization Method of Urban Rail Transit Based on Improved Differential Evolution Algorithm[J]. Sensors, 2022,23(1):378.
- [14] LIU K, WANG X, WANG L. An Improved Memetic Algorithm for Urban Rail Train Operation Strategy Optimization[J]. International Journal of Innovative Computing Information and Control, 2020,16(1):241-256.
- [15] WANG Y H, DE SCHUTTER B, VAN DEN BOOM T J J, et al. Optimal Trajectory Planning for Trains-a Pseudospectral Method and a Mixed Integer Linear Programming Approach[J]. Transportation Research Part C: Emerging Technologies, 2013,29:97-114.
- [16] FENG M L, WU C X, LU S F, et al. Notch-based Speed Trajectory Optimisation for High-speed Railway Automatic Train Operation[J]. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 2022,236(2):159-171.
- [17] WEI G D, ZHU S W, WANG Y H, et al. Energy-efficient Automatic Train Operation for High-speed Railways: Considering Discrete Notches and Neutral Sections[J]. Transportation Research Part C: Emerging Technologies, 2022,145:103884.
- [18] GU Q, TANG T, CAO F, et al. Energy-efficient Train Operation in Urban Rail Transit Using Real-time Traffic Information[J]. IEEE Transactions on Intelligent Transportation Systems, 2014, 15(3):1216-1233.
- [19] SU S, ZHU Q Y, LIU J Q, et al. A Data-driven Iterative Learning Approach for Optimizing the Train Control Strategy[J]. IEEE Transactions on Industrial Informatics, 2023,19(7):7885-7893.
- [20] 孟建军,蒋小一,陈晓强,等. 基于Sarsa算法的城轨列车节能控制策略研究[J]. 铁道标准设计,2024,68(8):8-14.
MENG Jianjun, JIANG Xiaoyi, CHEN Xiaoqiang, et al. Intelligent Control Strategy of Urban Rail Train Based on Sarsa Algorithm[J]. Railway Standard Design, 2024,68(8):8-14.
- [21] 周敏,董海荣,周学影,等. 临时限速下基于强化学习的高速列车速度曲线优化[J]. 铁道学报,2023,45(2):84-92.
ZHOU Min, DONG Hairong, ZHOU Xueying, et al. Reinforcement Learning-based Optimization of Speed Profile for High-speed Train with Temporary Speed Restriction[J]. Journal of the China Railway Society, 2023,45(2):84-92.
- [22] LIU X H, ZHANG Y D, GUO J. Research on Optimization of Energy-saving Driving Strategy of High-speed Train Based on PPO[C]//2023 3rd International Conference on Computer, Control and Robotics (ICCCR). Shanghai, China. IEEE, 2023:258-264.

- [6] 雷智敏. 客运专线铁路无砟轨道支承层试验与研究[J]. 铁道标准设计, 2009, 53(7): 26-28.
LEI Zhimin. Experiment and Research on Bearing Layer for Ballastless Track of Passenger Dedicated Lines[J]. Railway Standard Design, 2009, 53(7): 26-28.
- [7] 胡华军, 白舰兵, 关飞. 无砟轨道路基支承层施工技术[J]. 铁道建筑技术, 2008(S2): 237-239, 282.
HU Huajun, BAI Jianbing, GUAN Fei. Construction Technology of Subgrade Supporting Layer of Ballastless Track[J]. Railway Construction Technology, 2008(S2): 237-239, 282.
- [8] 董冲锋. 水硬性混凝土支承层材料及滑模摊铺施工技术研究[J]. 铁道工程学报, 2009, 26(4): 47-51.
DONG Chongfeng. Research on the Material for HGT and Construction Technology of Slip Form Paving[J]. Journal of Railway Engineering Society, 2009, 26(4): 47-51.
- [9] 于学义. 无砟轨道水硬性混合料支承层施工技术研究[J]. 铁道建筑技术, 2021(4): 156-159.
YU Xueyi. Research on Construction Technology of Hydraulic Mixture Supporting Layer for Ballastless Track[J]. Railway Construction Technology, 2021(4): 156-159.
- [10] 张俊庆. 无砟轨道低塑性混凝土支承层施工技术[J]. 高速铁路技术, 2011, 2(1): 57-59.
ZHANG Junqing. Construction Technology of Low Plastic Concrete Supporting Bed for Ballastless Track[J]. High Speed Railway Technology, 2011, 2(1): 57-59.
- [11] 易忠来, 杨志强, 温浩, 等. 高速铁路无砟轨道支承层服役现状及损伤机理[J]. 铁道建筑, 2023, 63(2): 7-11.
YI Zhonglai, YANG Zhiqiang, WEN Hao, et al. Service Status and Damage Mechanism of Ballastless Track Supporting Layer of High Speed Railway[J]. Railway Engineering, 2023, 63(2): 7-11.
- [12] 温浩, 李化建, 易忠来, 等. 严寒地区高速铁路混凝土结构耐久性防护技术现状及对策[J]. 中国铁路, 2018(4): 77-81.
WEN Hao, LI Huajian, YI Zhonglai, et al. Concrete Structure Durability Enhancement Technologies for High Speed Railway in Extremely Cold Areas[J]. China Railway, 2018(4): 77-81.
- [13] 杨怀志, 刘学文, 杨志强. 无砟轨道支承层混凝土损伤机理分析[J]. 铁道建筑, 2021, 61(8): 119-122.
YANG Huaizhi, LIU Xuewen, YANG Zhiqiang. Analysis of Concrete Damage Mechanism of Ballastless Track Bearing Layer[J]. Railway Engineering, 2021, 61(8): 119-122.
- [14] 杨建锋, 朱玉雪, 张帅, 等. CRTS II 型板式无砟轨道支承层混凝土粉化成因分析及修复方案[J]. 散装水泥, 2018(6): 47-50.
YANG Jianfeng, ZHU Yuxue, ZHANG Shuai, et al. Analysis and Repair Scheme of Pulverization of Concrete with Type CRTS II Slab Ballastless Track Support Layer[J]. Bulk Cement, 2018(6): 47-50.
- [15] 黄传岳. CRTS II 型板式无砟轨道支承层斜裂损伤修复方法[J]. 铁道建筑, 2018, 58(11): 139-141.
HUANG Chuanyue. Repairing Method of Supporting Layer Inclined Crack Damage of CRTS II Slab Ballastless Track[J]. Railway Engineering, 2018, 58(11): 139-141.
- [16] 张雅琴, 高亮, 钟阳龙, 等. 夏季高温下支承层斜裂缝诱发纵连板上拱规律研究[J]. 铁道学报, 2024, 46(3): 165-175.
ZHANG Yaqin, GAO Liang, ZHONG Yanglong, et al. Study on Arching Pattern of Continuous Ballastless Slab Track Induced by Oblique Crack in Concrete Base under High Temperature in Summer[J]. Journal of the China Railway Society, 2024, 46(3): 165-175.
- [17] YAO G W, SONG A X, ZHANG G F, et al. Experimental Study on Interface Performance of CRTS II Slab Ballastless Track under Temperature Loading[J]. Structures, 2024, 62: 106199.
- [18] LIU W X, WANG B Y, WANG Z X, et al. Mechanical Properties and Damage Evolution Features in Uniaxial Compression with Non-coincident Interlayer Contact in Double-block Ballastless Track[J]. Construction and Building Materials, 2024, 426: 136178.
- [19] 刘薇. CRTS II 型板式无砟轨道支承层断裂影响及修复效果研究[D]. 北京: 北京交通大学, 2014.
LIU Wei. Research on the Influence and Repairing Effect of [D]. Beijing: Beijing Jiaotong University, 2014.
- [20] 颜华, 翟婉明, 李春霞, 等. 土质路基双块式无砟轨道混凝土支承层设计分析[J]. 铁道工程学报, 2007(S1): 30-34.
YAN Hua, ZHAI Wanming, LI Chunxia, et al. Analysis of Designing Concrete Supporting Layer of Bi-block Sleeper Track in Soil Foundation[J]. Journal of Railway Engineering Society, 2007, 24(S1): 30-34.
- [21] 石现峰, 王澜, 宣言. 支承层参数变化对双块式无砟轨道结构力学性能的影响[J]. 铁道建筑, 2007(12): 97-100.
SHI Xianfeng, WANG Lan, XUAN Yan. Influence of Change of Supporting Layer Parameters on Mechanical Ballastless of Twin-block Type Ballastless Track Structure[J]. Railway Engineering, 2007(12): 97-100.
- [22] 陈欣, 郑建岚, 王国杰. 矿物掺合料对再生混凝土干燥收缩性能的影响[J]. 混凝土, 2016(10): 16-20, 24.
CHEN Xin, ZHENG Jianlan, WANG Guojie. Influence of Mineral Admixture on the Drying Shrinkage Performance of Recycled Concrete[J]. Concrete, 2016(10): 16-20, 24.

(编辑: 邓帅)

(上接第 19 页)

- [23] ZHANG Z Y, CHENG X Q, XING Z Y, et al. Pareto Multi-objective Optimization of Metro Train Energy-saving Operation Using Improved NSGA-II Algorithms[J]. Chaos, Solitons & Fractals, 2023, 176: 114183.
- [24] HAARNOJA T, ZHOU A, ABBEEL P, et al. Soft Actor-Critic: Off-policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor [EB/OL]. 2018: arXiv: 1801.01290. <https://arxiv.org/abs/1801.01290>.
- [25] HAARNOJA T, ZHOU A, HARTIKAINEN K, et al. Soft Actor-critic Algorithms and Applications [EB/OL]. 2018: arXiv: 1812.05905. <https://arxiv.org/abs/1812.05905>.
- [26] FERNÁNDEZ-RODRÍGUEZ A, SU S, FERNÁNDEZ-CARDADOR A, et al. A Multi-objective Algorithm for Train Driving Energy Reduction with Multiple Time Targets[J]. Engineering Optimization, 2021, 53(4): 719-734.

(编辑: 邓帅)