

基于 DQN 的列车节能驾驶控制方法

宿帅¹, 朱擎阳¹, 魏庆来², 唐涛¹, 阴佳腾¹

(1. 北京交通大学轨道交通控制与安全国家重点实验室, 北京 100044;

2. 中国科学院自动化研究所复杂系统管理与控制国家重点实验室, 北京 100190)

摘要: 随着轨道交通网络规模的扩大和列车运营间隔的缩短, 列车牵引能耗在快速增加。因此, 通过优化列车的驾驶策略降低牵引能耗, 对于轨道交通系统的节能减排具有重大意义。针对列车的驾驶策略优化问题, 提出一种基于深度 Q 网络 (DQN) 的列车节能驾驶控制方法。首先介绍了传统的列车节能驾驶问题并构造其反问题, 即通过分配最少的能耗达到规定运行时分。进一步将该问题转化为有限马尔可夫决策过程 (MDP), 通过设计状态动作值函数、定义动作策略选取方法等, 构建基于 DQN 方法的列车节能驾驶控制方法。通过实际驾驶数据对 DQN 进行训练, 得到最优的状态动作值函数, 并通过该值函数确定最优的能耗分配方案, 从而得到最优驾驶策略。最后, 以北京地铁亦庄线的实际运营数据设计了仿真算例, 对方法的有效性进行验证, 并对方法参数进行了敏感度分析。提出的方法可充分利用列车的驾驶数据提升驾驶策略, 降低列车牵引能耗, 对未来我国智慧城轨的发展具有一定的借鉴意义。

关键词: 列车节能驾驶; 驾驶策略; 深度 Q 网络

中图分类号: U238.4

文献标识码: A

doi: 10.11959/j.issn.2096-6652.202040

A DQN-based approach for energy-efficient train driving control

SU Shuai¹, ZHU Qingyang¹, WEI Qinglai², TANG Tao¹, YIN Jiateng¹

1. State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China

2. The State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

Abstract: The energy consumption in railway system is growing rapidly due to the expanding scale of the railway network and decreased operational headway. Hence, it is of great significant to apply the energy-efficient operation of the vehicles to cut down the energy cost of the railway system. A method for solving the energy-efficient train driving control based on deep Q-network (DQN) approach was proposed. Firstly, the traditional energy-efficient train driving control problem was presented and its inverse problem was formulated, i.e., distributing the least energy consumption units to achieve the scheduled trip time. Moreover, the problem was reformulated as a Markov decision process (MDP) and a DQN-based approach for energy-efficient train driving control was proposed. A DQN was built to approximate the action value function which determines the optimal energy distribution policy and further obtain the optimal driving strategy. Finally, a numerical experiment based on the real-world operational data was proposed to verify the effectiveness of the proposed method and analyze the performance of the proposed method. The driving data of the trains is applied to improve the driving strategy via the proposed method in the paper which reduces the traction energy consumption. It is of significance for the future development of Chinese intelligent urban railway system.

Key words: energy-efficient train driving, driving strategy, deep Q-network

收稿日期: 2020-11-18; 修回日期: 2020-12-02

通信作者: 阴佳腾, jtyin@bjtu.edu.cn

基金项目: 国家自然科学基金资助项目 (No.61803021, No.U1734210); 北京市自然科学基金资助项目 (No.L191015)

Foundation Items: The National Natural Science Foundation of China (No.61803021, No.U1734210), The Natural Science Foundation of Beijing of China (No.L191015)

1 引言

目前, 轨道交通以其运量大、准时性高、运营间隔小的优势, 成为城市居民日常出行的重要方式, 也成为公共交通体系中不可或缺的重要力量。近年来, 轨道交通系统的规模迅速扩大, 运营间隔也在不断缩短。以北京地铁系统为例, 2002 年北京地铁仅有两条线路, 31 个车站; 截至 2019 年年底, 北京地铁已经扩展到 23 条线路, 总计 405 个车站, 运营总里程达 699.3 km, 总规模比 2002 年扩大了 10 倍以上^[1]; 而北京地铁的运营间隔也已经缩短至高峰时段 2 min、平峰时段 5 min。但随着轨道交通系统运营规模的扩大和运营间隔的缩短, 其运营能耗不断增加。例如, 北京地铁 2018 年的运营能耗高达 19 亿 kWh, 年均增长近 9.2%^[2], 已经成为北京工业用电的消耗大户。据统计, 在地铁系统的运营过程中, 列车的牵引能耗占据运营总能耗的 50% 以上。例如, 北京地铁 10 号线的年均运营总能耗为 8 500 万 kWh 左右, 其中 4 600 万 kWh 的电能被用于列车的牵引^[3]。由此可见, 减少列车牵引能耗对于降低轨道交通系统的运营能耗、减少运营成本以及建设绿色环保的城市公共交通体系有重大意义。

降低列车牵引能耗的方式主要有两种: 从单车控制的角度, 通过优化列车的驾驶策略, 在保证准时性的同时降低牵引能耗; 从行车组织的角度, 通过优化列车运营时刻表, 提升再生制动的利用效率^[3]。本文主要针对前者展开研究, 即设计列车的节能驾驶控制方法, 从而优化列车驾驶策略, 降低牵引能耗。

现有对列车节能驾驶控制的研究主要依赖于最优驾驶控制模型, 在控制模型的基础上提炼目标函数和约束条件, 再使用数值解法或搜索算法求解最优驾驶策略。但在实际运营场景中, 列车的控制模型很难建立甚至无法建立。例如, 针对曲线、分相区、咽喉区等特殊路段的建模十分复杂, 而对于隧道阻力等复杂因素则无法准确进行建模。由此可见, 对于拥有这些特殊路段的区间来说, 传统的基于模型的优化方法无法针对这类问题进行准确求解。此外, 现有对列车节能驾驶控制的研究主要从运行时分的角度入手, 先求解满足运行时分的驾驶策略, 再从中寻找能耗最优的策略, 加大了求解的复杂性。实际上, 对于给定的驾驶策略而言, 其运行时分和牵引能耗是一一对应的, 因此可以从牵引

能耗的角度出发直接寻找最节能的驾驶策略, 由此简化求解过程。

基于以上分析, 本文的创新点概括如下。第一, 本文直接从牵引能耗的角度设计列车节能驾驶控制方法。本文构造了列车节能驾驶问题的反问题, 即如何通过分配最少的能耗来获得满足规定运行时分的驾驶策略, 并提出基于能耗分配的列车节能驾驶控制方法, 直接从牵引能耗的角度求解相应的驾驶策略, 简化了问题的求解过程。第二, 上述反问题符合有限马尔可夫决策过程 (Markov decision process, MDP) 的定义, 可以使用强化学习方法进行求解。本文使用深度 Q 网络 (deep Q-network, DQN) 求解列车的能耗分配策略, 进而得到最优驾驶策略。DQN 结合了强化学习与深度学习的优势, 既突出了决策者与环境的交互, 又能够处理和利用大数据, 并从中获取数据的本质特征^[4-5]。DQN 可以直接从列车的实际驾驶数据或虚拟仿真得到的驾驶数据中学习驾驶经验, 在不依赖于最优驾驶模型的情况下得到最优驾驶策略, 提高求解的准确性。

2 相关研究

列车节能驾驶控制方法的优化问题主要由两个子问题构成: 一是确定列车的节能驾驶工况, 二是确定在规定区间内这些工况间的转换点。因此, 确定列车最优控制方法的过程就是确定列车最优驾驶工况序列的过程。针对此问题, 国内外很多学者展开了大量研究, 根据其研究方法可以分为以下 3 类。

第一类方法是基于列车最优控制模型的数值解法。Ichikawa K^[6]将列车的基本运行阻力线性化, 并忽略线路的坡度阻力和限速, 提出了列车的基本节能控制模型。他使用庞特里亚金极大值原理 (Pontryagin maximum principle, PMP) 对模型进行求解, 得出列车的节能驾驶工况由最大加速 (maximum acceleration, MA)、巡航 (cruising, CR)、惰行 (coasting, C) 和最大制动 (maximum braking, MB) 组成的结论。Howlett P^[7]在平坦线路和固定限速的条件下, 使用 PMP 证明了列车的节能驾驶策略由 MA、CR、C 和 MB 这 4 种工况组成。在此基础上, Howlett P^[8]考虑了在不同坡度、不同限速的线路条件下, 列车采用离散控制档位的节能驾驶问题, 给出了使列车牵引能耗最小的关

键方程,并以此为依据计算不同驾驶工况的转换点。Liu R R 等人^[9]利用 PMP 确定了列车的节能驾驶工况,同时基于汉密尔顿函数的连续性提出了一种计算工况转换点的数值计算方法。实验表明,该方法既可以优化单站间的列车节能驾驶控制策略,也可以优化多站间的运行时分分配,进而优化时刻表。Khmelnitsky E^[10]使用 PMP 计算出连续坡度和限速的线路条件下列车最优的驾驶工况,并设计了一种迭代算法计算最优工况转换点。同时,他发现列车的运行时分越短,其巡航速度越大,并从理论上证明了巡航速度与节能驾驶策略之间的关系。曲健伟等人^[11]对基于 PMP 的驾驶策略求解方法进行简化,引入连接函数和连接误差对 C 工况的连接效果进行表征,并基于最优工况切换条件,提出了给定目标牵引恒速条件下求解最优 C 连接段的数值算法。Aradi S 等人^[12]提出了一种预测优化模型,即在考虑当前位置的工况时,引入对未来一定距离内驾驶策略的预测,并使用逐次二次规划方法求解该模型。Wang Y H 等人^[13]在确定最优驾驶策略的同时考虑了列车的能耗和舒适度,提出了伪谱法和混合整数线性规划方法两种求解列车最优驾驶策略的方法。很多学者在优化单车控制策略的同时还引入了再生制动,进一步降低了能耗。Scheepmaker G M 等人^[14]在列车的节能控制模型中考虑了再生制动,利用 PMP 计算出节能驾驶工况,并通过高斯伪谱方法求解不同工况的转换点。经过实际线路测试发现,引入再生制动后,列车的巡航速度下降,惰行点后移,高速驾驶时会更早使用制动工况以获得更多的再生能。Albrecht A 等人^[15-16]在考虑再生制动的基础上提出适用于连续坡度和限速的线路条件的列车节能控制模型,并指出应该在陡下坡时使用 RB 工况,以此让列车以固定的速度保持巡航的状态。

第二类方法是基于启发式搜索的方法。Morris B 等人^[17]采用遗传算法求解列车的节能驾驶曲线,该方法的目标函数包含了列车能耗和针对晚点的惩罚因子,同时考虑了列车节能和准时性。仿真结果表明,使用该方法可以节约 32.89% 的牵引能耗。Domnguez M 等人^[18]设计了一种计算机辅助选择最优驾驶策略的方法,他根据列车自动控制 (automatic train operation, ATO) 系统的控制性能、线路信息、车辆性能等参数构造一系列可行的驾驶策略,并使用帕累托寻优法选择满足规定运行时分且能耗最

小、舒适度最优的驾驶策略。这种方法还可以扩展到考虑再生制动的场景中^[19],从而优化列车时刻表。Zhang C Y 等人^[20]开发了基于数据驱动的列车驾驶 (data-driven train operation, DTO) 模型,该模型将数据挖掘方法与专家知识结合起来,利用专家知识获得更优的驾驶策略,并设计了启发式停车算法,以保证列车停车精度。Liu X 等人^[21]在考虑线路信息和调度条件的基础上,以列车牵引能耗和准时性为优化指标,建立了列车优化控制模型,并采用布谷鸟搜索算法对模型进行求解。Cheng R J 等人^[22]开发了一种基于数据挖掘技术的高速列车自动驾驶系统。该方法首先对采集的高速列车驾驶数据进行分类和筛选,其次使用分类与回归树提取驾驶数据中的专家经验,进而控制列车输出相应的控制工况。仿真实验表明,该方法求得的驾驶策略的能耗降低了 7.3%。厉高等人^[23]通过分析列车节能最优控制的工况集,将列车运行区间分段,进而使用时间逼近的搜索方法对列车工况转换点的坐标进行求解,然后对不同区段的列车惰行时间进行分配,得到最优控制方法。

随着人工智能技术的不断进步,机器学习的相关算法也被引入列车节能驾驶问题中。Huang Y N 等人^[24]针对长陡下坡中的重载列车安全驾驶问题,提出了使用神经网络计算列车驾驶曲线的方法。他们使用专家司机的驾驶数据对网络进行训练,并使用训练后的神经网络控制重载列车的循环制动,保证了制动过程的安全性和稳定性。Yin J T 等人^[25]提出了一种基于强化学习的智能列车驾驶方法,首先定义了列车节能驾驶场景下的状态、动作和奖励,并基于贝尔曼方程和 ϵ -贪婪策略确定不同状态的值函数及其更新方式。该方法可以在线实时获得驾驶策略,比离线优化方式减少了 10% 的能耗。Zhou R 等人^[26]使用 actor-critic 方法在线求解列车的节能驾驶策略。他使用两个神经网络分别输出控制工况和评估各状态下的值函数,并使用深度确定性梯度下降法对两个网络进行参数更新,进而同时得到最优策略和最优值函数。Huang K 等人^[27]提出了一种基于数据驱动的列车驾驶策略优化模型,基于列车的实际运行数据,使用随机森林回归 (random forest regression, RFR) 算法离线规划最优驾驶策略,使用支持向量机 (support vector machine, SVM) 实时调整列车运行速度,并计算牵引能耗。通过案例分析得知,该算法比普通的节能

驾驶算法平均可以减少 2.84% 的能耗。Huang J 等人^[28]提出了一种 BLSTM-DDQN 方法进行驾驶策略优化。他使用双向长短期记忆 (bidirectional long short term memory, BLSTM) 网络, 从与时间相关的运行信息中提炼列车的前向和反向运行特征, 利用自动编码器提炼列车的性质; 利用双重深度 Q 网络 (double deep Q network, DDQN) 优化列车的驾驶策略。Zhou K C 等人^[29]结合数据挖掘技术和强化学习方法, 以专家司机的驾驶经验为训练数据, 提出了基于深度确定性策略梯度 (deep deterministic policy gradient, DDPG) 方法的智能列车驾驶 (smart train operation based on DDPG, STOD) 方法和基于归一化优势函数的智能列车驾驶 (smart train operation based on normalized advantage function, STON) 方法两种方法, 将它们应用于连续状态和动作空间下的多目标列车驾驶优化问题。实验表明, 与人工驾驶相比, STOD 方法和 STON 方法可分别节约 9.4% 和 11.7% 的能耗。Qi X W 等人^[30]针对现有的列车能耗管理系统无法适用于变化的线路条件, 缺乏自适应性的问题, 提出使用深度 Q 网络学习方法, 在列车与环境的交互过程中自主学习最优的能耗管理策略, 以降低列车的运营能耗。实验表明, 该方法可以节约 16.3% 的牵引能耗。张森等人^[31]综合考虑线路和列车等约束条件, 建立了基于 Q 学习算法的列车节能优化模型。该模型中, 状态为列车的位置和速度, 值函数为准时性和能耗的加权叠加。实验结果表明, 与传统动态规划方法相比, 该方法得到的驾驶策略减少了 8.8% 的牵引能耗。

3 列车节能驾驶控制模型

3.1 问题描述

在列车节能驾驶问题中, 优化目标为最小化指定区间的能耗, 因此目标函数如式 (1) 所示^[32]:

$$\min \int_0^S \frac{u_f(x)F(v)}{\mu(u_f, v)} dx \quad (1)$$

其中, $u_f(x)$ 指列车在位置 x 处输出的相对牵引力, 用来衡量牵引力的输出级位; $F(v)$ 指在列车运行速度 v 下输出的最大牵引力; $\mu(u_f, v)$ 指列车在输出的相对牵引力为 u_f 、运行速度为 v 时从电网吸收电能的传递效率; S 指待优化区间的终点位置。

除此之外, 在列车运行过程中, 还要考虑以下

限制条件。首先, 列车在起点、终点的速度为 0, 同时为保证列车在运行过程中不触发紧急制动, 列车的速度不能超过线路的限速; 其次, 列车输出的牵引力或制动力不得超过列车牵引、制动装置所能提供的最大牵引力或制动力; 最后, 电网传递的电能无法全部转化为列车的机械能, 因此电网的传递效率取值必须满足相应的边界约束。这些限制条件如式 (2) 所示:

$$\begin{cases} v_0 = 0, v_T = 0 \\ 0 \leq v(x) \leq V(x) \\ u_f \in [0, 1], u_b \in [0, 1] \\ \mu \in [\mu_{\min}, \mu_{\max}] \end{cases} \quad (2)$$

其中, v_0 、 v_T 分别表示列车的初速度、末速度, $V(x)$ 表示线路在位置 x 处的限速值, u_b 表示相对制动力, 用来衡量制动力的输出级位, $\mu_{\min}$ 、 $\mu_{\max}$ 分别表示能量传递效率的上界、下界。

根据 PMP, 可以将以上目标函数与限制条件化成一个汉密尔顿函数, 该函数与 u_f 和 u_b 有关:

$$H = \frac{L_1}{mv} (u_f F - u_b B - R - G) - \frac{u_f F}{\mu} + \frac{L_2}{v} \quad (3)$$

其中, H 表示汉密尔顿函数, F 表示列车的牵引力, B 表示列车的制动力, R 表示列车的基本阻力, G 表示列车受到的坡度阻力, m 表示列车的质量, L_1 、 L_2 表示松弛因子。

依据式 (3) 中松弛项 $\frac{L_1}{mv}$ 的不同取值, 可以得到 5 种对应的工况^[33], 见表 1。

表 1 列车节能驾驶控制工况

松弛项取值范围	相对牵引力	相对制动力	工况
$\frac{L_1}{mv} < 0$	$u_f = 0$	$u_b = 1$	MB
$\frac{L_1}{mv} = 0$	$u_f = 0$	$u_b \in [0, 1]$	部分制动 (partial braking, PB)
$\frac{L_1}{mv} \in \left[0, \frac{1}{\mu_{\max}}\right]$	$u_f = 0$	$u_b = 0$	C
$\frac{L_1}{mv} > \frac{1}{\mu_{\min}}$	$u_f = 1$	$u_b = 0$	MA
$\frac{L_1}{mv} \in \left[\frac{1}{\mu_{\max}}, \frac{1}{\mu_{\min}}\right]$	$u_f = 1$ 或 0	$u_b = 0$	MA 或 C

综上所述, 列车的节能驾驶工况由 MA、C、PB 和 MB 组成。

3.2 模型转化

确定列车节能驾驶控制方法的过程就是确定上述节能驾驶工况序列的过程,即根据节能驾驶工况的组成和线路条件、车辆条件等信息,确定不同工况间的最优转换点。通常来讲,当一组驾驶工况序列被确定后,其能耗和运行时分便可以一一对应起来,如图1所示。

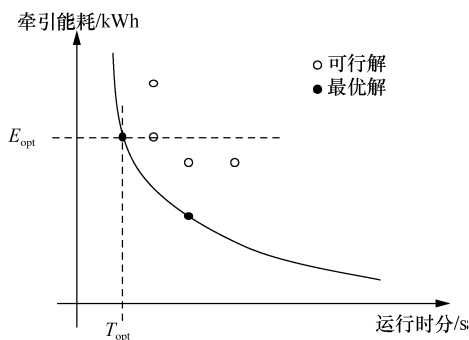


图1 运行时分与能耗的帕累托曲线

图1中的曲线表示列车在规定站间的运行时分和最小牵引能耗,所有满足规定运行时分的驾驶策略均排列在帕累托曲线的右侧,如图1中的空心圆圈所示;而在规定运行时分下的最优驾驶策略则位于帕累托曲线上^[16],如图1中的实心圆圈所示。因此,在规定运行时分下的最优驾驶策略可由列车的牵引能耗唯一确定。本节提出一种基于能耗分配的列车节能驾驶控制方法^[32],从牵引能耗的角度直接确定不同工况的转换点,进而确定符合规定运行时分的最优驾驶策略。

首先,根据站间的线路坡度和限速将区间划分为多个区段,使每个区段内的坡度和限速不变。接着,生成一个仅由MA-C-MB组成的初始策略,该策略的能耗很小但运行时分很长。为了缩短列车的运行时分,需要向线路的不同区间分多次分配若干能量单元,以增加牵引能耗为代价缩短运行时分,直至符合规定运行时分的要求。根据能耗分配确定驾驶策略的原则,具体如下:

- 在某个区段内,若分配到该区段的能量尚未用完且列车并未超速,则列车采取MA工况;
- 在某个区段内,若分配到该区段的能量已经用完,则列车采取C工况;
- 在某个区段内,若分配到该区段的能量尚未用完,但列车的速度已经接近限速值,则列车应采取MA-C组合工况,以避免超速;
- 仅在陡下坡处使用PB工况,以避免列车超速;

依次确定不同区段内的驾驶策略并连接起来,即可得到整个站间的驾驶策略。这里需要注意,前一个区段的末速度是后一个区段的初速度。

由以上分析可知,每分配一个能量单元后,列车的驾驶策略会发生变化,相应的运行时分和牵引能耗也会发生变化,可表示为:

$$\begin{cases} E_{i+1} = E_i + \Delta E \\ t_{i+1} = t_i - \Delta t_i \end{cases} \quad (4)$$

其中, E_i 表示分配第*i*个能量单元后所得驾驶策略的牵引能耗; t_i 表示分配第*i*个能量单元后所得驾驶策略的运行时分; ΔE 表示能量单元的大小, Δt 表示分配能量后运行时分的减少量。

在每一次能耗分配时,能量单元可以被分配到任意一个区段内。由式(4)可知,由于能量单元 ΔE 是相同的,每次分配后带来的牵引能耗增量也是相同的,但运行时分的减少量 Δt_i 是不同的。

不同能耗分配方案下得到的驾驶策略如图2所示,图中的两种驾驶策略是在初始策略的基础上向第3个区段或第7个区段分配一个能量单元后得到的。这两种策略的牵引能耗相同,但分配至第3个区段得到的驾驶策略运行时分更短。因此,在每一轮分配能量单元时,应尽可能将能量单元分配到运行时分减少量最大的区段,从而用最少的能量单元达到规定的运行时分要求,得到最节能的驾驶策略。

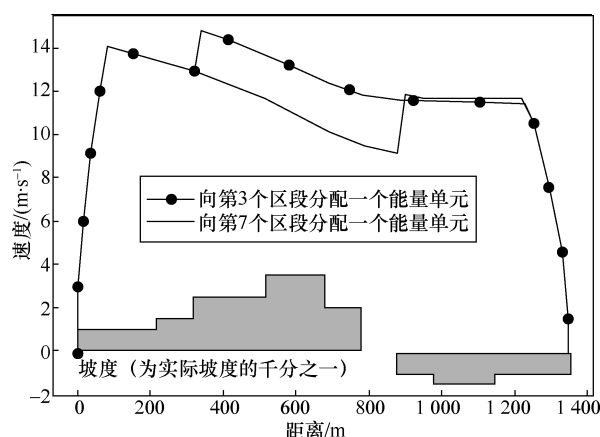


图2 不同能耗分配方案下得到的驾驶策略

基于以上分析可知,在每次选择向哪个区段分配能耗时,只需考虑当前各个区段的能耗分配状态,而不需要考虑之前所有的状态,即^[34]:

$$\pi(a|s) = P(A(t) = a | S(t) = s) \quad (5)$$

其中, $\pi(a|s)$ 表示决策者在状态 s 下决定采取动作 a 的概率, $S(t)$ 、 $A(t)$ 分别表示决策者在 t 时刻各个区段的能量单元分配状态和决定分配能量单元的区段。该决策过程具备马尔可夫性, 因此该方法符合有限 MDP 的定义, 可以应用强化学习方法进行求解。在本文研究的 MDP 中, 决策者需要通过与环境交互学习到一个最优的能耗分配策略, 即根据当前各个区段已经分配的能量单元数量决定向哪个区段继续分配能量, 最终得到最优的各段能量单元分配数, 并以此为依据确定节能驾驶控制方法。

4 基于 DQN 的列车节能驾驶控制方法

强化学习是一种针对目标导向任务的机器学习方法, 它不会告诉决策者应该如何行动, 而是通过与环境的交互引导决策者学习到正确的策略^[35]。在强化学习方法中, 决策者根据当前所处环境的状态 s 选择并执行动作 a , 该动作对环境产生影响, 使得当前环境的状态变为 s' , 并反馈奖励 r 给决策者, 决策者再根据新的状态继续与环境进行交互。在列车节能驾驶问题中, 状态指当前各个区段内已分配的能量单元数:

$$s = \{D_k | k = 1, 2, \dots, N\} \quad (6)$$

其中, D_k 表示第 k 个区段已经分配的能量单元数, N 表示所有可被分配能量的区段数。

奖励被定义为分配能量前后列车运行时分的减少量:

$$r_i = t_{i-1} - t_i \quad (7)$$

其中, 下标 i 表示第 i 次能耗分配。

动作定义为向某个特定区段分配能量, 因此动作集空间为 $A \in [1, N]$ 。

在交互过程中, 决策者的目标是使长期奖励总和最大化^[36]。通常使用值函数来评估某个状态或状态-动作对带来的长期奖励总和, 值函数常用贝尔曼方程进行定义^[37]:

$$q(s, a) = \sum_{s'} p(r, s' | s, a) \left[r + \sum_{a'} q(s', a') \right] \quad (8)$$

其中, $p(r, s' | s, a)$ 表示在状态 s 下采取动作 a 后, 得到奖励 r 并转移到新状态 s' 的概率, r 表示奖励, $q(s, a)$ 表示当前待评估的状态-动作对的值函数, a' 表示在 s' 下可执行的动作。

在强化学习方法中, 值函数通常用值函数表的形式进行存储, 将每个状态-动作对其值函数一一对应。但在节能驾驶问题中, 其状态空间大小为 $S = 1 + \sum_{k=1}^G N^k$, 其中 N 表示可分配能量单元的区段数, G 表示分配的能量单元总数。随着规定运行时分的缩短和线路长度的增加, 分配的能量单元总数和线路的区段数会增加, 其状态空间的规模会呈指数增加, 由此带来维数灾难的问题^[38], 严重影响值函数存储、更新和读取的速度。随着深度学习方法获得巨大的成功, 产生了深度学习与强化学习结合的深度强化学习 (deep reinforcement learning, DRL), 其将感知、学习与决策融为一体, 实现了从高维状态到动作决策的“端到端”感知与决策^[39]。DQN 方法是 DRL 方法的典型代表, 它将深度神经网络与 Q 学习方法结合起来, 能适应更大的状态空间, 因此适用于列车节能驾驶问题的求解。

DQN 是一种直接从环境中获取高维状态输入^[40], 并输出相应动作值函数的网络。将 DQN 应用于强化学习问题时的基本架构如图 3 所示。

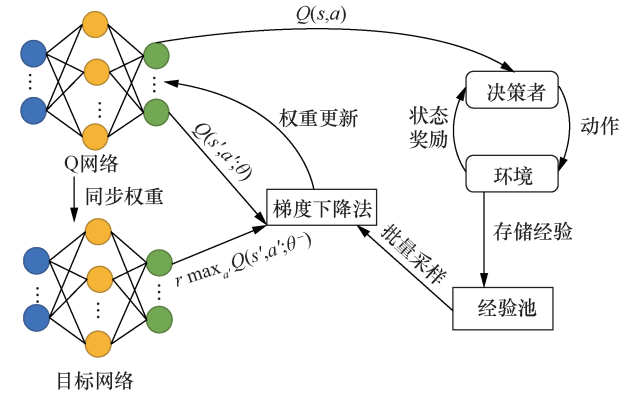


图3 DQN 基本架构

在深度学习问题中, 为保证训练结果的准确性, 通常要求训练数据满足数据间彼此独立的要求, 但在强化学习中, 数据序列之间有很强的关联性^[41], 这会导致神经网络的训练结果出现较大的偏差, 打破 Q 学习方法中的无条件收敛性^[42]。为解决此问题, 在 DQN 中引入经验回放机制。决策者不断地与环境进行交互, 得到一系列状态转移四元组 (s, a, r, s') 并存储到经验池中, 每次更新权重时, 从经验池内随机批量采样一部分数据作为网络的训练样本。经验回放的目的在于打破数据间的关联性, 求得准确的梯度估计值^[43]。

另外,在传统的Q学习或DQN方法中,通常仅将一个值函数表或值函数网络用于选择和评估动作,但单一的值函数表或值函数网络通常会导致决策者选择过估计的动作,导致对状态动作值函数的估计值过高^[44]。因此,在传统方法的基础上额外引入一个目标网络,专门用于评估状态-动作对的值。Q网络每进行若干轮权重更新后,会与目标网络进行一次权值同步,其余时间则保持目标网络的权重不变^[45],从而减小动作值函数的评估误差。

DQN方法中更新网络权重的计算式为^[46]:

$$\theta_{i+1} = \theta_i + \alpha[r_i + \gamma \max_a Q(s_{i+1}, a; \theta^-) - Q(s_i, a_i; \theta_i)] \nabla_{\theta} Q(s_i, a_i; \theta) \quad (9)$$

其中,下标*i*表示权重更新的次数,下标*t*表示当前状态、动作的时序, θ 表示Q网络的权重, θ^- 表示目标网络的权重, α 表示学习率, γ 表示衰减因子,二者的取值范围均为 $[0,1]$ 。

为了在特定状态下选到最优动作,决策者会采取特定状态下值函数最大的动作来执行,即采取贪婪策略,得到最优的动作策略:

$$\pi(a|s) = \arg \max_{a'} q(s, a') \quad (10)$$

但是,仅采取贪婪策略有可能会漏掉可能存在的更优动作,因此在学习的过程中必须对环境进行探索,这也是唯一能发现更好的策略的方法^[47]。本文采用的探索方法为 ε -贪婪策略,即以 ε 的概率随机选取任意一个可选动作,以 $(1-\varepsilon)$ 的概率采取贪婪策略:

$$\pi(a|s) = \begin{cases} (1-\varepsilon) + \frac{\varepsilon}{|A(s)|}, a = A^* \\ \frac{\varepsilon}{|A(s)|}, a \neq A^* \end{cases} \quad (11)$$

其中, $|A(s)|$ 为状态*s*下可选取的动作个数, A^* 为状态*s*下的最优动作。

综上所述,将DQN方法应用于列车节能驾驶问题时的算法如算法1所示。

算法1 将DQN应用于列车节能驾驶问题的求解流程

步骤1: 初始化

(1) 随机初始化Q网络权重 θ 和目标网络权重 θ^- ,且 $\theta = \theta^-$;

(2) 初始化学习率 α 、衰减因子 γ 、探索参数 ε ;

(3) 初始化经验池*D*和采样规模*B*;

(4) 初始化线路数据、车辆数据和规定运行时分*T*;

(5) 指定训练步数*M*、同步次数*C*;

步骤2: 开始训练

令更新次数 $\text{count} = 0$;

当 $\text{count} < M$ 时, 循环:

得到初始驾驶策略,其运行时分 t_0 ;

令列车当前运行时分 $t = t_0$;

当 $t > T$ 时, 循环:

根据当前能耗分配状态*s*,采用 ε -贪婪

策略选取动作*a*并执行,得到新状态*s'*;

根据*s'*更新驾驶策略,其运行时分 t_{new} ;

得到奖励 $r = t - t_{\text{new}}$;

将四元组 (s, a, r, s') 存储进经验池*D*中

$t = t_{\text{new}}$;

如果经验池*D*存储的经验数量大于或等于*B*,则:

从经验池中抽取*B*个四元组;

根据式(9)更新 θ ;

$\text{count} = \text{count} + 1$;

每经过*C*步,则:

$\theta = \theta^-$;

步骤3: 确定驾驶策略

得到初始驾驶策略,其运行时分 t_0 ;

令列车当前运行时分 $t = t_0$;

当 $t > T$ 时, 循环:

根据当前能耗分配状态*s*,采用 ε -贪婪

策略选取动作*a*并执行,得到新状态*s'*;

根据*s'*更新驾驶策略,其运行时分 t_{new} ;

得到奖励 $r = t - t_{\text{new}}$;

最终输出的驾驶策略即所求。

5 案例分析

为验证算法的有效性,并探究超参数对算法性能的影响,本节以北京地铁亦庄线小红门到肖村的实际运行数据为基础设计仿真案例,主要线路数据见表2^[48-49]。

表2 案例分析中的主要线路数据

参数	取值	参数	取值
站间长度	1 350 m	列车质量	200 t
最大牵引力	250 kN	最大制动力	120 kN
线路最大限速	22 m/s	能量单元	0.5 kWh
初始策略运行时分	141.13 s	初始策略牵引能耗	5.62 kWh

5.1 方法有效性验证

本节验证在不同规定运行时分下 DQN 方法的有效性。将站间线路按照坡度和限速分为 9 个区段，因此网络的输入层和输出层均有 9 个单元。首先将网络设置为一个隐藏层，在隐藏层内设置 5 个神经元，令 $\alpha = 0.1$ ， $\gamma = 0.9$ ，并分析不同规定运行时分下 DQN 方法的有效性。对比实验选取参考文献[32]提到的基于能耗分配的数值计算方法（以下简称数值法），该方法已经被证明可以求得最优解。实验结果见表 3。

表 3 DQN 方法与数值法得到的驾驶策略结果对比

规定运行时分/s	求解方法	实际运行时分/s	牵引能耗/kWh
100	DQN 方法	99.14	11.62
	数值法	99.14	11.62
98	DQN 方法	97.80	12.12
	数值法	97.80	12.12
96	DQN 方法	95.50	13.12
	数值法	95.50	13.12

由表 3 数据可知，DQN 方法得到了与数值法相同的结果，证明 DQN 方法可以得到最节能的驾驶曲线。

由 DQN 方法得到的速度-距离曲线如图 4 所示。

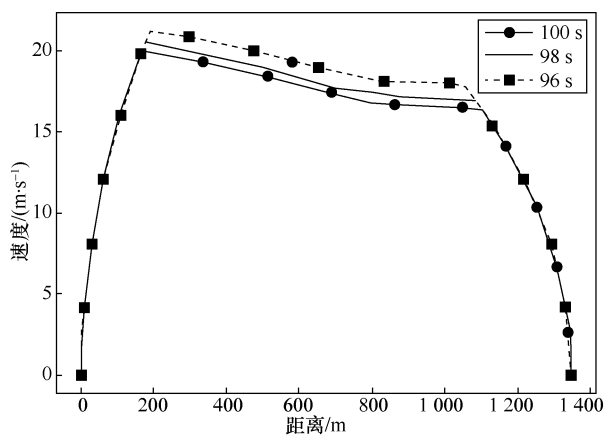


图 4 不同规定运行时分下的速度-距离曲线

由图 4 可发现，所有能量单元均分配在区间的起始部分，因此在区间的起始阶段，列车一直采取 MA 工况，之后便进行长距离的惰行，直到进站制动，且规定的运行时分越短，MA 工况到 C

工况的转换点处对应的速度值越高。这种驾驶策略加长了惰行距离，减少了不必要的制动，因此可充分利用由 MA 工况转化成的机械能，让这部分机械能尽可能多地被用来让列车克服运行阻力向前运行，从而在保证准时性的同时节约牵引能耗。由此可见，在列车运行过程中通过延长惰行距离，减少不必要的制动工况，可以达到节能驾驶的目的^[50]。

使用 DQN 方法进行训练时的收敛过程如图 5 所示。

为了跟踪训练时的状态动作值函数，在训练开始前首先采取随机动作执行一个片段，得到一个状态集。每经过一步训练后，将状态集中每个状态输入 DQN，得到每个状态的最大动作值并取均值^[41]。

由图 5 可知，DQN 方法的训练过程是不稳定的，在训练初期会有较大幅度的波动。但随着训练步数的增加，列车的运行时分逐渐接近规定运行时分，最终在最优运行时分附近小幅度波动；列车消耗的牵引能耗逐渐降低，最终在最优的牵引能耗处小幅度波动；状态动作值函数在训练初期变化较快，最后会收敛到一个稳定值。

以上实验仅考虑了单站间的优化问题，在下面的实验中考虑多站间的优化问题。本实验选取亦庄线宋家庄至肖村、肖村至小红门、小红门至旧宫 3 个站间。根据既有的列车时刻表可知，这 3 个站间的规定运行时分分别为 190 s、108 s、157 s^[49]。使用数值法和 DQN 方法分别对上述 3 个站间的驾驶策略进行优化，其结果见表 4。

从表 4 可以发现，数值法根据每个站间的规定运行时分确定最优能耗分配策略，3 个站间的实际运行时分分别为 185.46 s、105.82 s 和 154.51 s，均符合各个站间的规定运行时分，3 个站间的最优牵引能耗分别为 11.46 kWh、7.81 kWh、12.85 kWh；而使用 DQN 方法时，可以将 3 个站间视为一个整体，依据 3 个站间的总规定运行时分确定最优的能耗分配策略，进而得到列车的节能控制方法。由表 4 的数据可知，虽然在肖村—小红门站间和小红门—旧宫站间的实际运行时分分别为 111.16 s 和 172.77 s，并不符合这两个站间的规定运行时分，但 3 个站间的总运行分为 450.47 s，符合规定的总运行时分 455 s，且比数值法的实际运行时分 445.79 s 更接近规定运行时分，提高了准时性。同时，使用 DQN

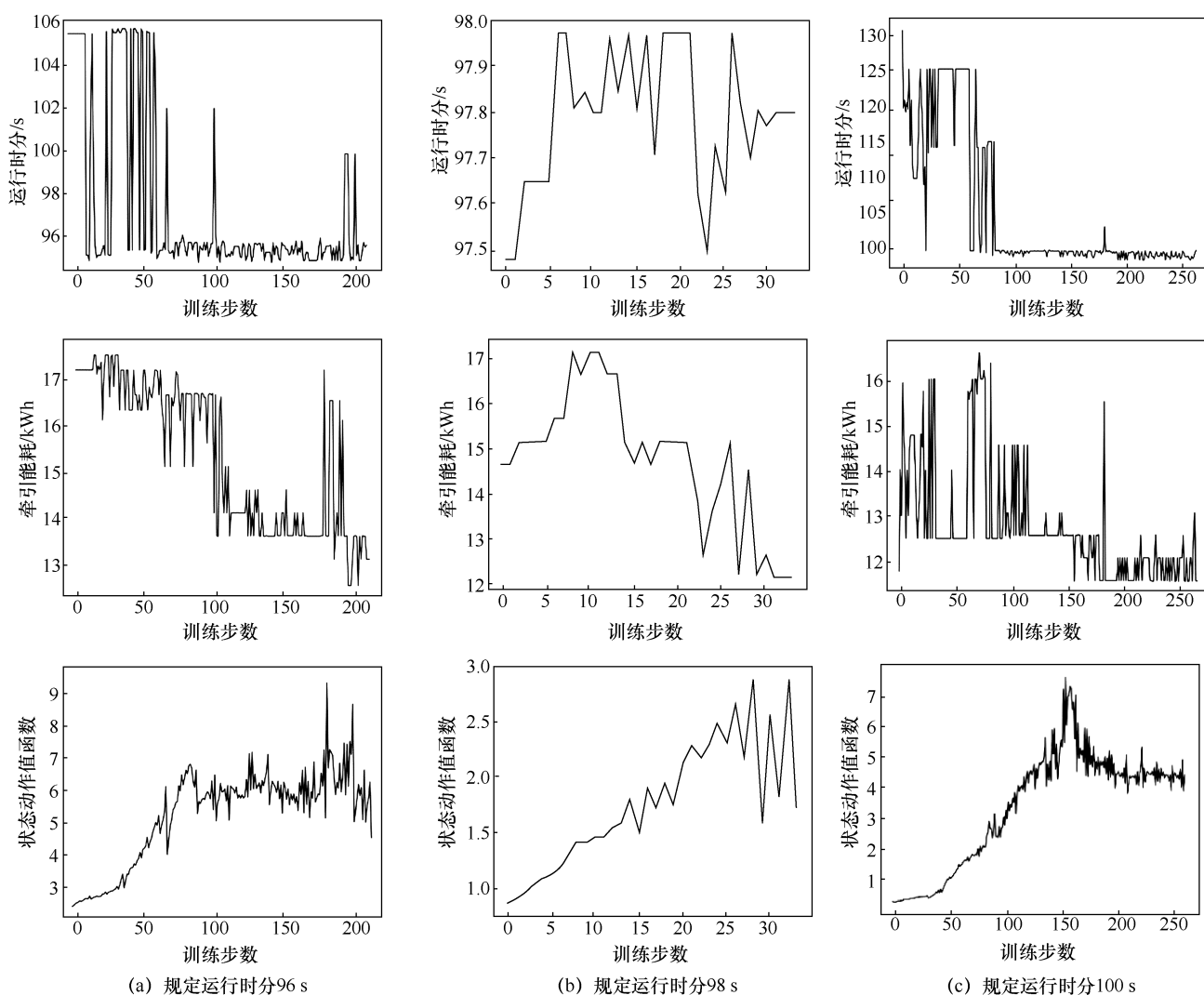


图5 使用DQN方法进行训练时的收敛过程

表4 3个站间场景下使用数值法和DQN方法得到的驾驶策略结果对比

站间	规定运行时分/s	实际运行时分/s		实际牵引能耗/kWh	
		数值法	DQN 方法	数值法	DQN 方法
宋家庄—肖村	190	185.46	166.54	11.46	14.51
肖村—小红门	108	105.82	111.16	7.81	6.92
小红门—旧宫	157	154.51	172.77	12.85	8.94
总计	455	445.79	450.47	32.12	30.37

方法得到的节能控制方法在3个站间所需的总牵引能耗为30.37 kWh,与数值法的总牵引能耗32.12 kWh相比,节约了5.4%。因此,DQN方法不仅可以节约牵引能耗,还可以重新分配各个站

间的运行时分,提高准时性。

图6显示了使用DQN方法得到的从宋家庄到旧宫之间3个站间的最优驾驶策略。从图6可以发现,列车在每一个站间内的驾驶策略与图4

所示的单站间驾驶策略相似, 均在起始阶段采取 MA 工况, 经过长距离的惰行后进入制动状态。同时还可以发现, 宋家庄—肖村站间执行 MA 工况的距离比肖村—小红门和小红门—旧宫站间执行 MA 工况的距离更长, 这表明在多站间场景下, 列车依然会在靠近起点的站间尽可能快速通过, 为后续的列车运行作业节约时间, 在提高准时性的同时, 也可以在后半程降低运行速度, 延长运行时分, 降低牵引能耗, 从而弥补列车在起始站间多消耗的能量。

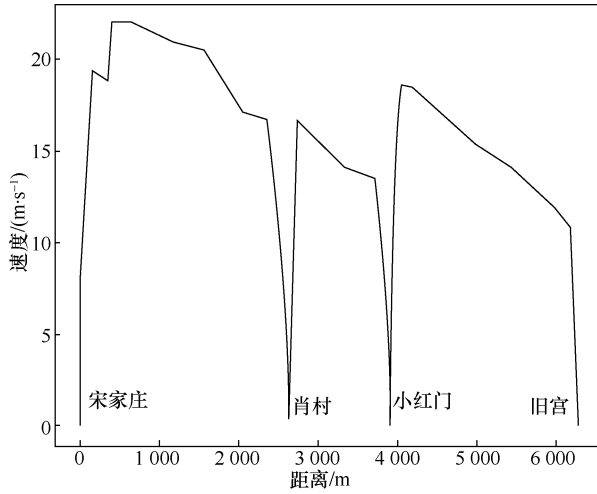


图6 DQN方法在多站间场景下得到的最优驾驶策略

5.2 超参数敏感性分析

本节探究算法关键超参数对 DQN 训练速度的影响。在强化学习中, 影响训练速度的两个关键超参数为学习率 α 和衰减因子 γ , 前者决定了每步更新时当前更新值向目标值靠近的步长, 后者决定了对未来长期奖励的重视程度。同时, 神经网络的结构也会影响训练的速度。根据深度学习相关理论^[51], 神经网络的深度对神经网络性能有更显著的影响, 因此针对神经网络结构的讨论将聚焦于网络深度对训练速度的影响。

图7显示了给定衰减因子时, 不同学习率对 DQN 训练速度的影响。在本实验中, 衰减因子取值为 $\gamma = 0.9$, 神经网络中每层神经元的个数均为5个。由图7可知, 当学习率较小时, DQN 的训练速度较慢, 当学习率 $\alpha \in [0.3, 0.5]$ 时, 可以实现较快的训练速度。同时, 由图7还可发现, 在学习率相同的情况下, 3层神经网络的训练速度通常比单层神经网络更快, 这证明增加神经网络的深度可以提高 DQN 的训练速度, 提高算法性能。

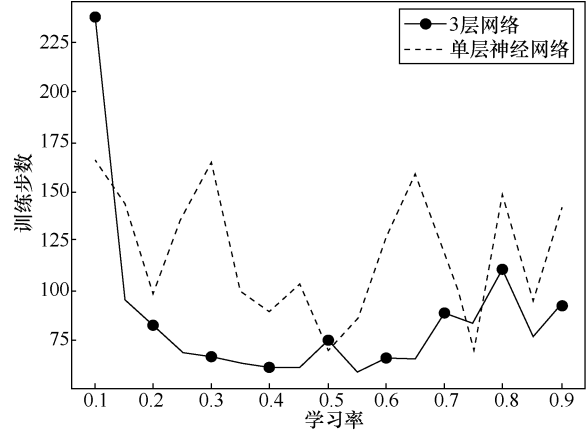


图7 衰减因子一定时不同学习率对 DQN 训练速度的影响

图8显示了给定学习率时, 不同衰减因子对 DQN 训练速度的影响。在本实验中, 学习率设置为 $\alpha = 0.1$, 神经网络中每层神经元个数设置为5。由图8可知, 当使用单层神经网络时, 不同衰减率对训练速度的影响不显著; 当增加网络深度 (本实验为3层) 时, 随着衰减因子的增大, 训练速度显著加快。其原因在于, 在能量单元分配的过程中, 每分配一个能量单元, 即可得到一个新的驾驶策略, 该驾驶策略将会成为后续继续进行能量单元分配的基础, 因此决策者不仅需要关注当前驾驶策略下分配区间的选择, 还要考虑本次分配能量单元后对后续决策的影响。而衰减因子衡量了决策者对未来长期奖励的关注程度, 当衰减因子取值增大时, 可以让决策者更加关注长期的奖励总和, 达到一个片段内奖励总和的最大化, 从而更快地学到每个状态下的最优动作。根据实际实验可知, 衰减因子不能取值太小, 其取值范围最好为 $[0.3, 1)$ 。

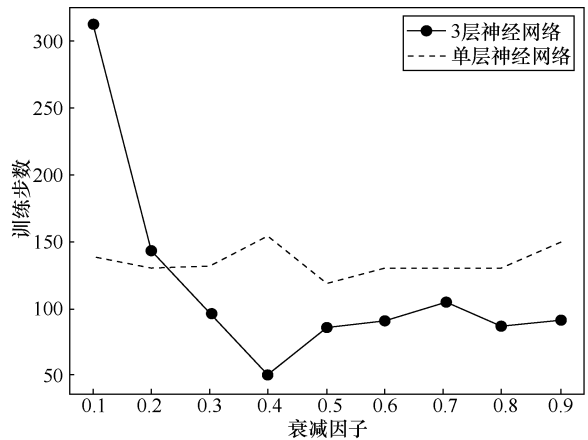


图8 学习率一定时不同衰减因子对 DQN 训练速度的影响

从图8还可以发现, 当 $\gamma > 0.2$ 后, 使用3层神经网络的训练速度快于单层网络的训练速度, 表明

加深网络深度可以很大程度地加速训练过程, 提高收敛速度。

6 结束语

本文提出了一种基于 DQN 的列车节能驾驶控制方法。首先, 本文构造了列车节能驾驶问题的反问题, 即通过分配最少的能耗得到满足规定运行时分的驾驶策略, 并使用强化学习方法求解此问题。针对此问题状态空间较大的特点, 提出使用 DQN 方法来近似表示各个状态-动作对的值函数。使用实际驾驶数据对 DQN 进行训练, 使其得到最优值函数, 并依据最优值函数得到最优的能耗分配方案, 进而得到最优驾驶策略。仿真实验表明, 该方法获得的驾驶策略能够得到满足规定运行时分的节能驾驶策略, 并且使用 DQN 方法得到的节能驾驶策略有很长的惰行距离, 而很少采取制动工况, 因而避免了频繁地加速-制动所带来的牵引能耗损失, 达到了节能驾驶的目的。

目前的工作依旧存在一些缺陷和不足。首先, 能量单元的大小是离散值, 导致驾驶策略的求解结果不够精确; 其次, 本文仅研究了单车的节能优化方法, 没有考虑多车运行时的多角色博弈过程^[52], 无法达到协同优化的目的; 最后, 由于深度学习存在难以应对非结构化数据的缺点^[53], 本文所提出的方法仍然无法充分利用列车运行过程中产生的大量非结构化数据, 降低了数据的利用率。这些问题将在未来的工作中进行研究和解决。

参考文献:

- [1] 中国城市轨道交通协会. 2019 年度中国内地城轨交通线路概况[R]. 2020.
China Association of Metro. An overview of urban rail transit lines in the mainland of China in 2019[R]. 2020.
- [2] 褚心童, 张亚东, 郭进, 等. 基于蚁群算法的列车节能驾驶策略优化算法研究[J]. 铁道标准设计, 2020, 65(2): 1-7.
CHU X T, ZHANG Y D, GUO J, et al. An optimization algorithm for train energy-saving driving strategy based on ant colony algorithm[J]. Railway Standard Design, 2020, 65(2): 1-7.
- [3] YANG X, LI X, NING B, et al. A survey on energy-efficient train operation for urban rail transit[J]. IEEE Transactions on Intelligent Transportation Systems, 2015, 17(1): 2-13.
- [4] 张钹. 人工智能进入后深度学习时代[J]. 智能科学与技术学报, 2019, 1(1): 4-6.
ZHANG B. Artificial intelligence is entering the post deep-learning era[J]. Chinese Journal of Intelligent Science and Technology, 2019, 1(1): 4-6.
- [5] 袁小锋, 王雅琳, 阳春华, 等. 深度学习在流程工业过程数据建模中的应用[J]. 智能科学与技术学报, 2020, 2(2): 107-115.
YUAN X F, WANG Y L, YANG C H, et al. The application of deep learning in data-driven modeling of process industries[J]. Chinese Journal of Intelligent Science and Technology, 2020, 2(2): 107-115.
- [6] 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.
- [7] HOWLETT P. Optimal strategies for the control of a train[J]. Automatica, 1996, 32(4): 519-532.
- [8] HOWLETT P. An optimal strategy for the control of a train[J]. The Anziam Journal, 1990, 31(4): 454-471.
- [9] LIU R R, GOLOVITCHER I M. Energy-efficient operation of rail vehicles[J]. Transportation Research Part A: Policy and Practice, 2003, 37(10): 917-932.
- [10] KHMELNITSKY E. On an optimal control problem of train operation[J]. IEEE Transactions on Automatic Control, 2000, 45(7): 1257-1266.
- [11] 曲健伟, 王青元, 孙鹏飞. 基于极大值原理的地铁列车节能驾驶简化算法[J]. 铁道科学与工程学报, 2019, 16(6): 1577-1586.
QU J W, WANG Q Y, SUN P F. A simplified algorithm of energy-efficient driving for metro trains based on the maximum principle[J]. Journal of Railway Science and Engineering, 2019, 16(6): 1577-1586.
- [12] ARADI S, BECSI T, GASPAR P. A predictive optimization method for energy-optimal speed profile generation for train[C]//14th IEEE International Symposium on Computational Intelligence and Informatics. Piscataway: IEEE Press, 2013: 135-139.
- [13] WANG Y H, SCHUTTER D, NING B, et al. Optimal trajectory planning for trains - a pseudo spectral method and a mixed integer linear programming approach[J]. Transportation Research Part C: Emerging Technologies, 2013, 29: 97-114.
- [14] SCHEEPMAKER G M, GOVERDE R M P. Energy-efficient train control including regenerative braking with catenary efficiency[C]//2016 IEEE International Conference on Intelligent Rail Transportation. Piscataway: IEEE Press, 2016: 116-122.
- [15] ALBRECHT A, HOWLETT P, PUDNEY P, et al. The key principles of optimal train control—part 1: formulation of the model, strategies of optimal type, evolutionary lines, location of optimal switching points[J]. Transportation Research Part B: Methodological, 2016, 94: 482-508.
- [16] ALBRECHT A, HOWLETT P, PUDNEY P, et al. The key principles of optimal train control—part 2: existence of an optimal strategy, the local energy minimization principle, uniqueness, computational techniques[J]. Transportation Research Part B: Methodological, 2016, 94: 509-538.
- [17] MORRIS B, FEDERICA F, MICHELA L. Application of genetic algorithms for driverless subway train energy optimization[J]. International Journal of Vehicular Technology, 2016: 1-14.
- [18] DOMNGUEZ M, FERNÁNDEZ A, CUCALA A P, et al. Optimal design of metro automatic train operation speed profiles for reducing energy consumption[J]. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 2011, 225(5): 1-14.

- 463-474.
- [19] DOMNGUEZ M, FERNÁNDEZ A, CUCALA A P, et al. Energy savings in metropolitan railway substations through regenerative energy recovery and optimal design of ATO speed profiles[J]. IEEE Transactions on Automation Science and Engineering, 2012, 9(3): 496-504.
- [20] ZHANG C Y, CHEN D W, YIN J T, et al. Data-driven train operation models based on data mining and driving experience for the diesel-electric locomotive[J]. Advanced Engineering Informatics, 2016, 30(3): 553-563.
- [21] LIU X, DAI S H. Optimization of train control strategy for energy saving and time precision using multi-objective cuckoo search algorithm[C]//The 2nd International Conference on Computer Science and Application Engineering. [S.l.:s.n.], 2018: 1-5.
- [22] CHENG R J, CHEN D W, GAI W L, et al. Intelligent driving methods based on sparse LSSVM and ensemble CART algorithms for high-speed trains[J]. Computers & Industrial Engineering, 2019, 127: 1203-1213.
- [23] 厉高, 林建辉, 庄哲, 等. 基于再生制动的城市轨道交通列车节能控制研究[J]. 铁道运输与经济, 2019, 41(3): 121-126.
- LI G, LIN J H, ZHUANG Z, et al. A research on energy-saving train control of urban mass transit based on regenerative brake[J]. Railway Transport and Economy, 2019, 41(3): 121-126.
- [24] HUANG Y N, TAN L T, CHEN L, et al. A neural network driving curve generation method for the heavy-haul train[J]. Advances in Mechanical Engineering, 2016, 8(5): 1-14.
- [25] YIN J T, CHEN D W, LI L X. Intelligent train operation algorithms for subway by expert system and reinforcement learning[J]. IEEE Transactions on Intelligent Transportation Systems, 2014, 15(6): 2561-2571.
- [26] ZHOU R, SONG S J. Optimal automatic train operation via deep reinforcement learning[C]//The 10th International Conference on Advanced Computational Intelligence. [S.l.:s.n.], 2018: 103-108.
- [27] HUANG K, WU J J, YANG X, et al. Discrete train speed profile optimization for urban rail transit: a data-driven model and integrated algorithms based on machine learning[J]. Journal of Advanced Transportation, 2019, 4: 1-17.
- [28] HUANG J, ZHANG E D, ZHANG J R, et al. Deep reinforcement learning based train driving optimization[C]//2019 Chinese Automation Congress. [S.l.:s.n.], 2019.
- [29] ZHOU K C, SONG S J, XUE A K, et al. Smart train operation algorithms based on expert knowledge and reinforcement learning[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2020: 1-12.
- [30] QI X W, LUO Y D, WU G Y, et al. Deep reinforcement learning enabled self-learning control for energy efficient driving[J]. Transportation Research Part C: Emerging Technologies, 2019, 99: 67-81.
- [31] 张淼, 张琦, 张梓轩. 基于 Q 学习算法的高速铁路列车节能优化研究[J]. 铁道运输与经济, 2019, 41(12): 111-117.
- ZHANG M, ZHANG Q, ZHANG Z X. A study on energy-saving optimization for high-speed railways train based on Q-learning algorithm[J]. Railway Transport and Economy, 2019, 41(12): 111-117.
- [32] SU S, TANG T, LI X. Driving strategy optimization for trains in subway systems[J]. Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit, 2018, 232(2): 369-383.
- [33] SCHEEPMAKER G M, GOVERDE R M P, KROON L G. Review of energy-efficient train control and timetabling[J]. European Journal of Operational Research, 2017, 257(2017): 355-376.
- [34] 唐伦, 贺小雨, 王晓, 等. 基于迁移演员-评论家学习的服务功能链部署算法[J]. 电子与信息学报, 2020, 42(11): 2671-2679.
- TANG L, HE X Y, WANG X, et al. Deployment algorithm of service function chain based on transfer actor-critic learning[J]. Journal of Electronics & Information Technology, 2020, 42(11): 2671-2679.
- [35] YUAN Y L, YU Z L, GU Z H, et al. A novel multi-step Q-learning method to improve data efficiency for deep reinforcement learning[J]. Knowledge-Based Systems, 2019, 175: 107-117.
- [36] HEINRICH J, SILVER D. Deep reinforcement learning from self-play in imperfect-information games[J]. arXiv preprint, 2016, arXiv: 1603.01121.
- [37] SUTTON R, BARTO A. Reinforcement learning: an introduction[M]. Cambridge: MIT Press, 1998.
- [38] 王飞跃, 曹东璞, 魏庆来. 强化学习: 迈向知行合一的智能机制与算法[J]. 智能科学与技术学报, 2020, 2(2): 101-106.
- WANG F Y, CAO D P, WEI Q L. Reinforcement learning: toward action-knowledge merged intelligent mechanisms and algorithms[J]. Chinese Journal of Intelligent Science and Technology, 2020, 2(2): 101-106.
- [39] 孙长银, 穆朝絮. 多智能体深度强化学习的若干关键科学问题[J]. 自动化学报, 2020, 46(7): 1301-1312.
- SUN C Y, MU C X. Important scientific problems of multi-agent deep reinforcement learning[J]. Acta Automatica Sinica, 2020, 46(7): 1301-1312.
- [40] VOLODYMYR M, KORAY K, DAVID S, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529-533.
- [41] VOLODYMYR M, KORAY K, DAVID S, et al. Playing atari with deep reinforcement learning[J]. arXiv preprint, 2013, arXiv: 1312.5602.
- [42] 张俊杰, 张聪, 赵涵捷. 重复利用状态值的竞争深度 Q 网络算法[J]. 计算机工程与应用, 2020, 56(21): 1-10.
- ZHANG J J, ZHANG C, ZHAO H J. Dueling deep Q network algorithm with state value reuse[J]. Computer Engineering and Applications, 2020, 56(21): 1-10.
- [43] YANG Z R, XIE Y C, WANG Z R, et al. A theoretical analysis of deep Q-learning[J]. Learning for Dynamics and Control, 2020: 486-489.
- [44] ZHAO W Y, GUAN X Y, LIU Y, et al. Stochastic variance reduction for deep Q-learning[J]. arXiv preprint, 2019, arXiv:1905.08152.
- [45] VAN HASSELT H, GUEZ A, SILVER D. Deep reinforcement learning with double Q-learning[J]. arXiv preprint, 2015, arXiv: 1509.06461.
- [46] OSBAND I, BLUNDELL C, PRITZEL A, et al. Deep exploration via bootstrapped DQN[J]. Advances in Neural Information Processing Systems, 2016: 4026-4034.
- [47] HASSELT H V, WIERING M A. Reinforcement learning in continuous action spaces[C]//2007 IEEE International Symposium on Ap-

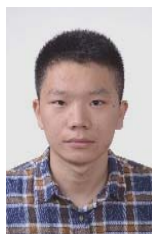
proximate Dynamic Programming and Reinforcement Learning. Piscataway: IEEE Press, 2007: 272-279.

- [48] SU S, LI X, TANG T, et al. A subway train timetable optimization approach based on energy-efficient operation strategy[J]. IEEE Transactions on Intelligent Transportation Systems, 2013, 14(2): 883-893.
- [49] 宿帅. 城轨列车运行图和速度曲线一体化节能方法[D]. 北京: 北京交通大学, 2016.
- SU S. An energy-efficient approach by integrating the train timetable and speed trajectory in urban rail[D]. Beijing: Beijing Jiaotong University, 2016.
- [50] LIU W T, TANG T, SU S, et al. Energy-efficient train driving strategy with considering the steep downhill segment[J]. Processes, 2019, 7(2): 77.
- [51] GOODFELLOW I, BENGIO Y, COURVILLE A. Deep learning[M]. [S.l.]: MIT press, 2016.
- [52] 沈宇, 韩金朋, 李灵犀, 等. 游戏智能中的 AI-从多角色博弈到平行博弈[J]. 智能科学与技术学报, 2020, 2(3): 205-213.
- SHEN Y, HAN J P, LI L X, et al. AI in game intelligence-from multi-role game to parallel game[J]. Chinese Journal of Intelligent Science and Technology, 2020, 2(3): 205-213.
- [53] 郑南宁. 人工智能新时代[J]. 智能科学与技术学报, 2019, 1(1): 1-3.
- ZHENG N N. The new era of artificial intelligence[J]. Chinese Journal of Intelligent Science and Technology, 2019, 1(1): 1-3.

[作者简介]



宿帅 (1987-), 男, 博士, 北京交通大学轨道交通控制与安全国家重点实验室副教授, 主要研究方向为列车节能控制、列车智能调度、强化学习算法等。



朱擎阳 (1996-), 男, 北京交通大学轨道交通控制与安全国家重点实验室硕士生, 主要研究方向为列车节能驾驶控制、强化学习算法等。



魏庆来 (1979-), 男, 博士, 中国科学院自动化研究所研究员, 复杂系统管理与控制国家重点实验室副主任, 中国科学院大学岗位教授, 青岛智能产业技术研究院智能技术创新中心主任, 主要研究方向为自学习控制、平行控制自适应动态规划、智能控制、最优控制及其工业应用。



唐涛 (1963-), 男, 博士, 北京交通大学轨道交通控制与安全国家重点实验室主任, 电子信息工程学院院长、教授, 主要研究方向为高速铁路控制、智能控制理论、智能交通理论等。



阴佳腾 (1992-), 男, 博士, 北京交通大学轨道交通控制与安全国家重点实验室副教授, 主要研究方向为列车智能控制与调度、机器学习、列车节能驾驶控制等。