doi：10. 3969/j. issn. 1005-7854. 2022. 02. 017

基于强化学习算法的地下铲运机车速控制

王伯健[1] 战 凯[1] 郭 鑫[1,2] 石 峰[1] 高泽宇[1]
（1. 北京矿冶研究总院，北京 100160；
2. 北京科技大学 机械工程学院，北京 100083）

摘 要：针对铲运机无人驾驶行驶时车速变化难控制的问题，将强化学习算法应用于车速的控制，使车辆在各种状态下车速保持平滑稳定。对比了强化学习算法和经验法、模糊控制、传统 PID 控制、滑膜控制、逆控制、智能优化算法等算法，分析并设计了强化学习策略，推导出了强化学习模型，即控制车速和上一时刻车速、上一时刻航向角偏差、上一时刻位置偏差的关系，计算相关参数，进行了仿真实验并验证了模型的正确性。结果表明，相对于传统的模糊化分级控制车速和经验法，强化学习算法控制可很好地提升车速变化的稳定性，可根据环境和自身状态车速变化，灵活、正确地调节车速变化，很好地提高车辆的动态性能并减少误差。

关键词：地下铲运机；强化学习；无人驾驶；车速控制
中图分类号：TD421 **文献标志码**：A **文章编号**：1005-7854(2022)02-0099-06

Speed control of load-haul-dump(LHD) based on reinforcement learning algorithm

WANG Bo-jian ZHAN Kai GUO Xin SHI Feng GAO Ze-yu
(1. Beijing General Research Institute of Mining and Metallurgy, Beijing 100160, China;
2. School of Mechanical Engineering, University of Science and Technology Beijing, Beijing 100083, China)

Abstract: To solve the problem that it is difficult to control the change of the speed of LHD unmanned driving, the reinforcement learning algorithm is applied to control the speed of LHD unmanned driving, so that the vehicle speed can keep smooth and stable in various states. The reinforcement learning algorithm is compared with the empirical method, fuzzy control, traditional PID control, synovial control, inverse control and intelligent optimization algorithm. The reinforcement learning strategy is analysed and designed, and the reinforcement learning model is deduced, that is, the relationship between the control speed and the last time speed, the last time heading angle deviation and the last time position deviation. The relevant parameters are calculated, and the simulation experiments are carried out to verify the correctness of the model. Experimental results show that compared with traditional fuzzy hierarchical control speed and the empirical method to set the speed, the reinforcement learning algorithm control can improve the stability of the speed change, flexibly and correctly adjust the speed change according to the speed change of the environment and its own state, improve the dynamic performance of the vehicle and reduce the error.

Key words: LHD; reinforcement learning; unmanned driving; speed control

地下铲运机属于铰接车的一种，是灵活、机动的铲装运输设备。KCY-2 型铲运机通过铰接点一侧布置的油缸伸缩来实现转向，在车辆行驶过程中，有左右转向响应效果不同的问题，加上地面摩擦条件的因素，车辆转向的控制较为复杂。为保证

收稿日期：2021-01-12
基金项目：国家重点研发计划项目(2018YFC060402)
第一作者：王伯健，硕士研究生，研究方向为地下矿用铰接车无人操控技术。E-mail：wangbojian@bgrimm. com
通信作者：郭鑫，正高级工程师；E-mail：guoxin@bgrimm. com

无人驾驶过程中转向控制的精确，对车速的控制要求更高，需要车速的控制反应快、稳定性强。

针对刚性车辆的车速控制的研究，当前大多采用专家经验法、模糊化车速、PID 控制、滑膜控制、智能控制算法、智能优化搜索算法等。目前，很多车速控制算法是依据驾驶特征大数据、视觉等经验法控制车速[1,2]，存在局限性。通过模糊理论模糊化分级控制输出车速的方法会使车速分级变化，平稳性、可控性较低[3]。传统的 PID 控制器和控制方法稳定性较高，但存在应用需大量时间调整参数和无法适应变化等系统问题[4-6]。滑膜控制、逆控制、智能优化算法等其他方法对于非线性不稳定系统问题的控制效果不理想，存在不足[7-9]。在实际巷道中，地下铲运机的车速控制需要较高的及时性、可靠性和稳定性，但因为车载控制器硬件配置限制，以及对响应速度的要求较高，复杂的算法无法满足使用条件。而强化学习是从环境获得数据后不断训练，从而获得对环境的精确反应，是一种强学习行为，可以进行离线学习，得到稳定的模型和参数，从而满足控制要求。铲运机是铰接车辆，其转向是通过油缸伸缩改变其铰接角来完成，当油缸动作开始转向时，若铰接角变化过大则需要车辆减速到合适且尽可能大的车速转向，以便提升行驶效率，强化学习控制车速的目的是找到这一时刻的车速。利用专家经验和强化学习方法，设计出符合地下铲运机工况以及硬件设备要求的强化学习策略，再利用智能优化算法离线仿真后得出完整的强化学习模型，进行实车试验，验证可行性和准确性。

1 强化学习算法基本理论

强化学习与监督学习不同，是从环境状态获得信息判断执行动作的学习，使执行动作从环境中获得的累积奖赏值最大，通过试错来寻找最优的动作行为。其学习过程是一个试错与评价选择的马尔科夫决策过程，对问题进行建模，将其定义为一个四元组(S，A，p，f)，其中，S 为状态集合，$s_t \in S$ 表示控制对象在 t 时刻的状态量；A 为控制对象可执行动作集合，$a_t \in A$ 表示控制对象在 t 时刻的动作；p 为奖赏函数集合，$r_t \rightarrow p(s_t, a_t)$ 表示控制对象状态 s_t 执行动作 a_t 获得的即时奖励量；f 为概率在 0～1 的状态转移概率分布函数，$s_{t+1} \rightarrow f(s_t, a_t)$ 表示控制对象在状态 s_t 执行动作 a_t 转移到下一状态 s_{t+1} 的概率[10,11]。

强化学习方法是学习一个行为策略 π：$S \rightarrow A$，使学习对象的动作能够获得最大奖赏。即当学习对象由状态 s_t 变化到状态 s_{t+1} 时，所有动作集合能够获得最大的奖赏值，作为一个行为策略[12-14]。奖赏函数形式为：

$$Rs_t = \sum \gamma_t r_t s_t (0 < \gamma_t \leqslant 1, 0 < s_t \leqslant 1) \quad (1)$$

式中，γ_t 是折扣因子，用来平衡未来奖赏对累积奖赏的影响。

根据式 2 目标函数可以计算出最优值，确定最优行为策略[15-19]：

$$\pi = \arg(Rs_t)_{\max}, s_t \in S \quad (2)$$

2 适用于铲运机的强化学习算法

根据控制系统硬件分析，受控制器和传感器的硬件限制，无法对庞大的数据进行实时在线学习，且速度较快时，学习速度来不及会带来车辆行驶风险，因此采用离线学习出强化学习模型，导入到车辆进行车速控制。学习算法逻辑如图 1 所示。

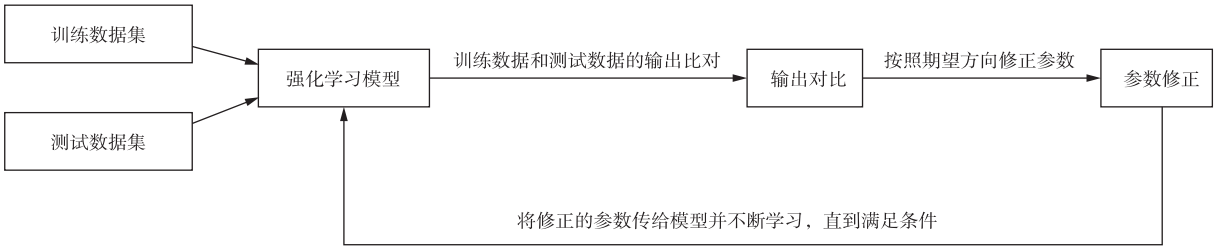

图 1 强化学习算法逻辑图

Fig. 1 Logic diagram of reinforcement learning algorithm

2.1 强化学习数据集

训练数据集主要是平时实车测试的数据包和随机生成的一些数据包，测试数据集主要来源于平时实车测试表现良好的数据包和有经验的司机师傅驾驶铲运机时记录的数据包。训练数据数量和测试数据数量比例大致为 2∶1。数据包的信息主要包含了车辆 GPS 得到的正东和正北方向的坐标、前车头航向角、铰接角角度值、车辆行驶速度以及当前时刻等。通过计算得出车辆坐标变换得到的横向位置误差、绝对航向角误差、铰接角变化量以及上一时刻车速。

2.2 强化学习模型建立

根据一般的强化学习方法，将信息量全部考虑其中，考虑到实际问题，车辆在行驶过程中，下一时刻的状态信息是未知的，只能通过预测或者经验进行判断，因此在已知当前时刻之前的状态信息前提下，采用强化学习控制下一时刻的车速。初步模型中，记录了之前 10 个时刻的状态信息，将其作为控制因素考虑其中，进行简单学习时发现，铰接角变化量和航向角偏差基本一致，且铰接角变化量的角度变化很小，其绝对值基本小于 3°，所以略去铰接角变化量，确定模型具有其他三个状态信息量，由于车速没有方向，且无需考虑状态量的方向，所有状态量采用绝对值形式，模型见式 3，其中 $n=10$，$1\leqslant t<10$：

$$v_{f+1}=\frac{\sum_{t}^{n}v_t\gamma_t^v r_t^v}{K_v\times\sum_{t}^{n}v_t\gamma_t^v r_t^v+K_x\times\sum_{t}^{n}x_t\gamma_t^x r_t^x+K_\theta\times\sum_{t}^{n}\theta_t\gamma_t^\theta r_t^\theta} \tag{3}$$

式 3 中，v_{t+1} 为输出的期望车速；v_t 为之前 t 时刻的车速，km/h；x_t 为之前 t 时刻横向位置误差绝对值，cm；θ_t 为之前 t 时刻航向角误差绝对值，°。

因设备配置水平和提高算法执行效率，只采用上一时刻的状态信息，即 $n=1$，$t=1$，修改强化学习模型见式 4。

$$v_{t+1}=\frac{v_t\gamma_t^v r_t^v}{K_v\times v_t\gamma_t^v r_t^v+K_x\times x_t\gamma_t^x r_t^x+K_\theta\times\theta\gamma_t^\theta r_t^\theta} \tag{4}$$

此时，所有的 $\gamma_t r_t$ 都为一个常系数，可以系数合并再除去 $\gamma_t^v r_t^v$ 得到最终模型见式 5。

$$v_{t+1}=\frac{v_t}{K'_v\times v_t+K'_x\times x_t+K'_\theta\times l_t} \tag{5}$$

式 5 中，车速系数 $K'_v=\dfrac{K_v\gamma_t^v r_t^v}{\gamma_t^v r_t^v}$；横向误差系数 $K'_x=\dfrac{K_v\gamma_t^x r_t^x}{\gamma_t^v r_t^v}$；航向角误差系数 $K'_\theta=\dfrac{K_v\gamma_t^\theta r_t^\theta}{\gamma_t^v r_t^v}$。

式 5 为最终的强化学习模型，对其进行离线学习训练出系数参数即可。

2.3 强化学习过程

离线训练强化学习模型中的三个系数参数，获得最终完善的强化学习模型。强化学习一般分为基于值函数的方法和基于策略搜索的方法求解最优参数。本文采用基于遗传算法的策略搜索方法，遗传算法其交叉、变异的算子具有打破局部最优特点，且算法灵活多变，应用广泛，适用绝大多数优化求解问题。具体步骤为：

1)随机生成 5 个个体和选取计算后的 5 个训练集数据组成父代种群，即建立 10 行 3 列 n 页的矩阵 $\boldsymbol{A}$，n 为最短数据集的总元素组个数；

2) 对矩阵 $\boldsymbol{A}$ 每页数据随机配对随机选择交叉点进行交叉计算，生成子一代个体；

3)当前时刻元素对矩阵 $\boldsymbol{A}$ 每页数据进行最优策略选择，选取最接近测试数据集的车速，并进行迭代计算搜索；

4)分析计算选取一组适合参数进行参数拟合校验，再进行弯道和直道的加权平均，计算出合理的三个系数参数。

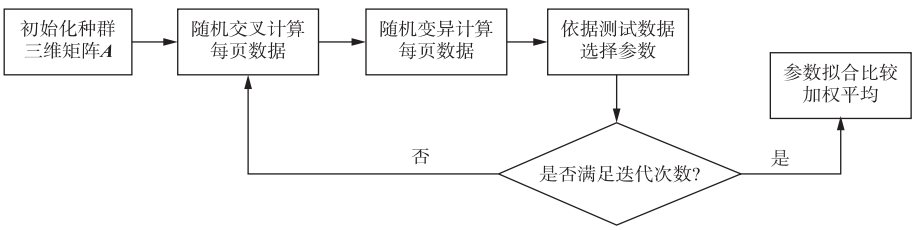

图 2 强化学习遗传算法计算系数参数流程图

Fig. 2 Flow chart of parameters calculated by reinforcement learning genetic algorithm

在离线学习过程中，由于一个数据包记录了上百组数据元素，相邻两组数据变化不明显，所以在选取数据时，每格 5 个组数据选取一个作为实验数据，此时特征明显且计算量显著减少。在初步离线学习时，可以大致确定参数变化范围，在保障打破局部最优局限情况下，选取尽可能大的一定变异阈值可以使搜索计算更快地收敛。最终得到参数结果保留 4 位小数为：K'_v 0.006 5，K'_x 0.060 8，K'_θ 0.111 4。

3 实车测试结果

两种控制算法都在直线行驶时的结果如图 3 和图 4 所示。两种控制算法都先行驶一小段直线加速后再转向行驶时的结果如图 5 和图 6 所示。

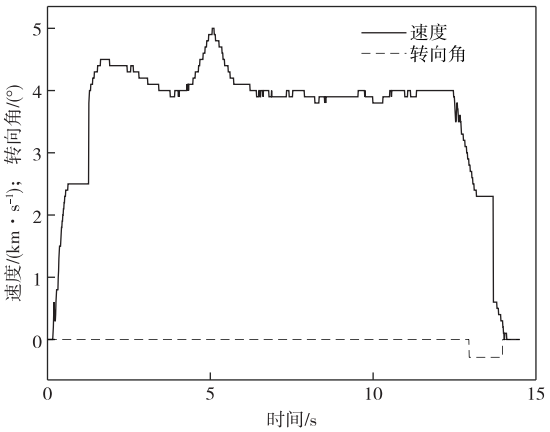

图 3 纯模糊控制直线行驶结果数据图

Fig. 3 Data graph of straight line driving results of pure fuzzy control

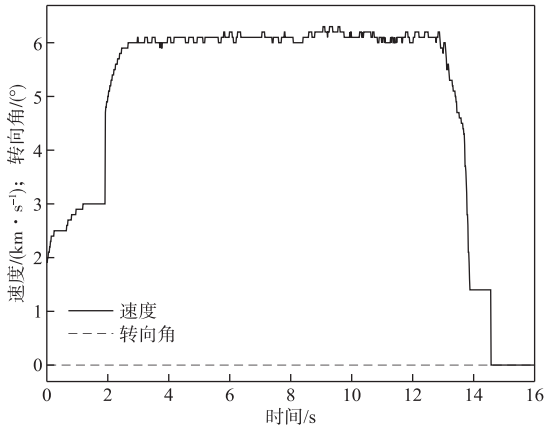

图 4 强化学习控制直线行驶结果数据图

Fig. 4 Data graph of straight line driving results of reinforcement learning control

其中，由于铲运机转向是通过油缸伸缩来实现，由于油缸存在摩擦力，且控制油缸转向的液压阀有死区，所以只有当压力达到一定值时，油缸才会动作，因此在图中转向控制量的绝对值达到 4 以上时，铰接角才会开始变化。

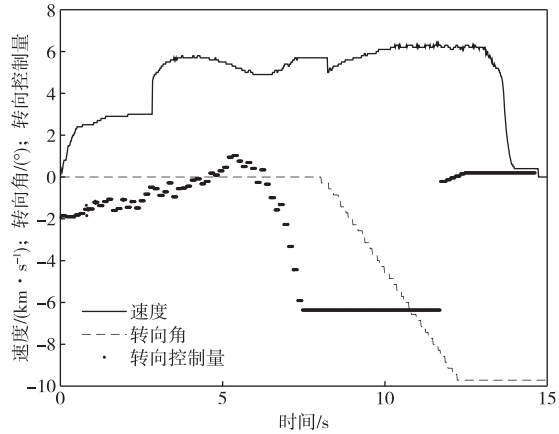

图 5 纯模糊控制直线再转向结果数据图

Fig. 5 Result data graph of straight redirection of pure fuzzy control

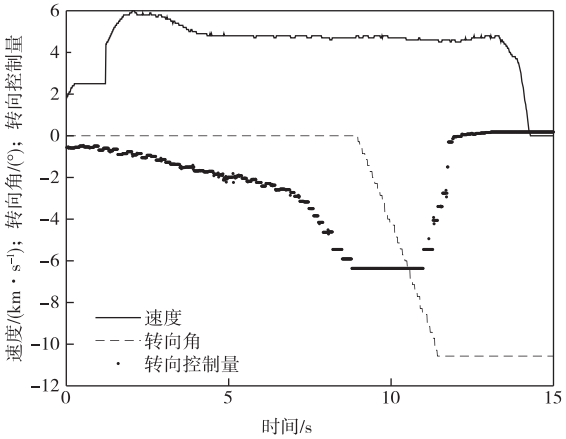

图 6 强化学习控制直线再转向结果数据图

Fig. 6 Data graph of redirection results of reinforcement learning control

实车控制最初采用的是纯模糊控制，控制因素只有航向角误差为主导，不考虑其他因素。实验数据表明，车速控制效果不理想且有明显顿挫感，车辆安全人员十分不适。使用离线学习的理想强化学习模型后，控制因素考虑了横向位置误差、航向角误差以及上一时刻车速，试验数据表明控制曲线较为理想，无明显顿挫感，震荡感不明显。

分析发现，直线试验中，纯模糊控制在加速后，行驶不是很稳定，行驶 6 s 左右时，速度有向上的突变，且速度变化较大不平稳，而强化学习控制在加速后，行驶比较平稳，且速度变化微小，在

0.2 km/h 之内。

转向试验中，纯模糊控制会有相应的减速效果，但是随后偏差变化过快时，其控制有些失控，不再准确，速度会十分不稳定，变化超出 1 km/h，甚至开始加速超出原来直线行驶速度，会给拐弯造成风险，在行驶 13 s 时，人为干预将车辆急停，而强化学习控制车速，在加速直线前进后进入弯道前，会提前及时减速，且速度变化平稳，变化在 0.2 km/h 之内，随后车速一直稳定直到转弯结束。算法优劣对照见表 1。

表 1　纯模糊控制和强化学习控制车速效果对比

Table 1　Comparison of speed control effects between pure fuzzy control and reinforcement learning control

工况	性能	纯模糊控制车速/(km·h^{-1})	强化学习控制车速/(km·h^{-1})
直线	稳定性	较差	较好
	及时性	较差	较好
	正确性	一般	较好
	变化范围	1.1	0.2
直线再转弯	稳定性	较差	较好
	及时性	较差	较好
	正确性	较差	较好
	变化范围	1.0	0.2

4　结论

1)针对铲运机的自身车辆特性和工作工况，对比模糊控制和强化学习控制初步实验，分析推导出的铲运机车速强化学习控制模型，即控制车速和上一时刻车速、上一时刻航向角偏差、上一时刻位置偏差的关系，强化学习算法控制车速可以更好地提高控制效果和行驶平滑性。

2)强化学习算法控制可显著减少车辆行驶过程中的急加、急减速现象，使车速更好地配合转向操作，安全员的舒适度可得到显著提高。在突发情况下，安全员能够更快地接管车辆，提高了车辆无人驾驶的稳定性、可靠性和安全性。

3)由于条件限制，强化学习模型本身化简的相对比较简单，之后还需使其更完善、更具体。例如，除了上一时刻的状态信息，将前几个时刻的状态都考虑其中进行计算分析，还需继续优化控制算法的动态性能和控制指标，甚至具备边行驶边学习的在线学习能力，让其更自动化、智能化，达到更高的目标。

参考文献

[1] 燕荣杰. 基于车联网数据的驾驶行为—车速控制的研究[D]. 济南：山东交通学院，2017.
YAN R J. The research on driving behavior—Speed control based on vehicle networking data[D]. Ji′nan: Shandong Jiaotong University, 2017.

[2] 刘兵. 基于驾驶员视知觉的车速控制和车道保持机理研究[D]. 武汉：武汉理工大学，2008.
LIU B. Mechanism study on the vision-based speed control and lane keeping [D]. Wuhan: Wuhan University of Technology, 2008.

[3] 秦绪情. 自动平行泊车系统定车速模糊控制算法研究[D]. 长春：吉林大学，2007.
QIN X Q. Study on arithmetic of fuzzy control of invariable speed of autonomous parallel parking system[D]. Changchun: Jilin University, 2007.

[4] 窦宝华，郭璧玺，张旭. 基于智能电动汽车的纵向车速跟随控制策略[J]. 汽车实用技术，2021，46(5)：27-30.
DOU B H, GUO B X, ZHANG X. Longitudinal speed following control strategy based on intelligent electric vehicle[J]. Automobile Practical Technology, 2021, 46(5): 27-30.

[5] 李洪硌. 无人驾驶汽车高速工况智能决策、轨迹规划与跟踪研究[D]. 广州：华南理工大学，2020.
LI H L. Research on intelligent decision-making, trajectory planning and tracking of driver less vehicle in high speed driving conditions[D]. Guangzhou: South China University of Technology, 2020.

[6] 张宁. 基于实时动力学参数辨识的纵向车速控制算法研究[D]. 天津：天津大学，2017.
ZHANG N. Research on longitudinal vehicle speed control algorithms based on real-time dynamic parameter identification [D]. Tianjin: Tianjin University, 2017.

[7] 陈刚，吴俊. 无人驾驶机器人车辆非线性模糊滑模车速控制[J]. 中国公路学报，2019，32(6)：114-123.
CHEN G, WU J. Nonlinear fuzzy sliding mode speed control for unmanned driving robotic vehicle[J]. China Journal of Highway and Transport, 2019,

32(6)：114-123.

[8] 储灿灿，王东，张为公，等. 基于逆控制策略模型的电动车驾驶机器人车速控制[J]. 汽车工程，2020，42(9)：1166-1173.
CHU C C，WANG D，ZHANG W G，et al. Vehicle speed control of electric vehicle driving robot based on inverse control strategy model [J]. Automotive Engineering，2020，42(9)：1166-1173.

[9] 王晖年. 基于网联的信号交叉口下自动驾驶车辆生态驾驶车速控制策略[D]. 厦门：厦门理工学院，2021.
WANG H N. Ecological driving speed control method of automatic driving vehicle at signalized intersection based on network connection [D]. Xiamen：Xiamen University of Technology，2021.

[10] 肖华. 无线通信中的马尔科夫决策过程研究[D]. 成都：电子科技大学，2013.
XIAO H. On markov decision processes in wireless communications[D]. Chengdu：University of Electronic Science and Technology of China，2013.

[11] WHITE C C，WHITE D J. Markov decision processes[J]. European Journal of Operational Research，1989，39(1)：1-16.

[12] 马骋乾，谢伟，孙伟杰. 强化学习研究综述[J]. 指挥控制与仿真，2018，40(6)：68-72.
MA C Q，XIE W，SUN W J. Research on reinforcement learning technology：A review [J]. Command and Control and Simulation，2018，40(6)：68-72.

[13] 高阳，陈世福，陆鑫. 强化学习研究综述[J]. 自动化学报，2004(1)：86-100.
GAO Y，CHEN S F，LU X. Research on reinforcement learning technology：A review [J]. Automatica，2004(1)：86-100.

[14] WOO J，YU C，KIM N. Deep reinforcement learning-based controller for path following of an unmanned surface vehicle [J]. Ocean Engineering，2019，183：155-166.

[15] POLVARA R，SHARMA S，WAN J，et al. Autonomous vehicular landings on the deck of an unmanned surface vehicle using deep reinforcement learning[J]. Robotica，2019，37(11)：1867-1882.

[16] CHU Z，SUN B，ZHU D，et al. Motion control of unmanned underwater vehicles via deep imitation reinforcement learning algorithm[J]. The Institution of Engineering and Technology，2020，14(7)：764-774.

[17] 万里鹏，兰旭光，张翰博，等. 深度强化学习理论及其应用综述[J]. 模式识别与人工智能，2019，32(1)：67-81.
WAN L P，LAN X G，ZHANG H B，et al. A review of deep reinforcement learning theory and application [J]. Pattern Recognition and Artificial Intelligence，2019，32(1)：67-81.

[18] 赵星宇，丁世飞. 深度强化学习研究综述[J]. 计算机科学，2018，45(7)：1-6.
ZHAO X Y，DING S F. Research on deep reinforcement learning[J]. Computer Science，2018，45(7)：1-6.

[19] 刘全，翟建伟，章宗长，等. 深度强化学习综述[J]. 计算机学报，2018，41(1)：1-27.
LIU Q，ZHAI J W，ZHANG Z C，et al. A survey on deep reinforcement learning [J]. Journal of Computers，2018，41(1)：1-27.

（编辑：王爱平）