无人驾驶中运用 DQN 进行障碍物分类的避障方法

刘航博，马礼*，李阳，马东超，傅颖勋
（北方工业大学信息学院，北京 100144）

摘　要： 安全是无人驾驶汽车需要考虑的首要因素，而避障问题是解决驾驶安全最有效的手段。基于学习的避障方法因其能够从环境中学习并直接从感知中做出决策的能力而受到研究者的关注。深度 Q 网络（DQN）作为一种流行的强化学习方法，在无人驾驶避障领域取得了很大的进展，但这些方法未考虑障碍物类型对避障策略的影响。基于对障碍物的准确分类提出一种 Classification Security DQN（CSDQN）的车辆行驶决策框架。根据障碍物的不同类型以及环境信息给出具有更高安全性的无人驾驶决策，达到提高无人驾驶安全性的目的。首先对检测到的障碍物根据障碍物的安全性等级进行分类，然后根据不同类型障碍物提出安全评估函数，利用位置的不确定性和基于距离的安全度量来评估安全性，接着 CSDQN 决策框架利用障碍物类型、相对位置信息以及安全评估函数进行不断迭代优化获得最终模型。仿真结果表明，与先进的深度强化学习进行比较，在多种障碍物的情况下，采用 CSDQN 方法相较于 DQN 和 SDQN 方法分别提升了 43.9%和 4.2%的安全性，以及 17.8%和 3.7%的稳定性。

关键词： 无人驾驶；深度 Q 网络；分类避障；评估函数；安全性
源代码链接： https://github.com/Thinkingnote/CSDQN-Obstacle-Avoidance.git
中图分类号： TP391　　**文献标志码：** A　　**DOI：** 10.19678/j.issn.1000-3428.0068769

Obstacle Avoidance Method Using DQN to Classify Obstacles in Unmanned Driving

LIU Hangbo, MA Li*, LI Yang, MA Dongchao, FU Yingxun
(School of Information, North China University of Technology, Beijing 100144, China)

【Abstract】 Safety is the primary factor to consider for unmanned driving vehicles, and obstacle avoidance is the most effective means to ensure driving safety. Learning-based obstacle avoidance methods have attracted the attention of researchers owing to their ability to learn from the environment and make decisions directly from perception. Deep Q-Network (DQN) has made significant progress as a popular reinforcement learning method in obstacle avoidance for autonomous driving; however, it does not consider the impact of the obstacle category on obstacle avoidance strategies. Therefore, we propose a vehicle-driving decision-making framework, the Classification Security DQN (CSDQN), which is based on accurate obstacle classification results. This framework aims to achieve higher safety in autonomous driving by finding safer decision strategies based on different obstacle types and environmental information. First, the detected obstacles are classified according to their safety levels, and then safety evaluation functions are proposed for different types of obstacles. The uncertainties in position- and distance-based safety measures are used to evaluate safety. The CSDQN decision-making framework utilizes obstacle categories, relative position information, and safety evaluation functions for continuous iterative optimization to obtain a final model. Finally, the proposed method is compared with advanced deep reinforcement learning. The simulation results show that in the presence of multiple obstacles, the CSDQN method improves safety by 43.9% and 4.2% and stability by 17.8% and 3.7%, respectively, compared to the DQN and SDQN methods.

【Key words】 unmanned driving; Deep Q-Network (DQN); classification obstacle avoidance; evaluation function; security

0　引言

近年来，随着人工智能技术和通信技术的发展，无人驾驶汽车的研究及应用受到人们越来越多的关注，无人驾驶技术不仅在提高驾驶的便利性和舒适性方面具有潜在优势，更为重要的是，它能够在道路安全性方面发挥积极作用。因此，无人驾驶的应用市场在未来有着广阔的前景[1-2]。

无人驾驶汽车主要的 3 种使能技术，即感知、规划和控制，其中规划算法确保车辆能规划出一条安

收稿日期： 2023-11-03　　**修回日期：** 2023-12-12
基金项目： 国家自然科学基金（62272007，62001007）；北京市自然科学基金（4234083，4212018）。
通信作者 E-mail： * mali@ncut.edu.cn

全的路径并且高效地避开障碍物，对安全驾驶起着重要作用[3]。以往的规划算法大多基于规则，但是该方法严重依赖于如何生成图形，并且没有充分考虑车辆动力学，使得生成的路径可能与动力学不兼容，存在一定的局限性。近年来，研究人员开始更多地研究基于机器学习规划方法。这些方法可以进一步分为有监督学习和深度强化学习。在监督学习领域，研究人员提出一种模仿学习方法，该方法需要收集大量的人类驾驶行为进行人工标注，模仿人类的行驶习惯，进而生成安全的轨迹，但需要大量人类驾驶数据，并且由于不同的驾驶员在面对相同的情况可能做出不同的决定，这可能导致训练的不确定性。在深度强化学习领域，该方法近年来已经在围棋和机器人、机械手等游戏中取得良好的效果[4-5]，其可以学习到复杂的策略并且不需要大量数据集的特性，从而使得深度强化学习在避障领域快速发展，与传统的手动设计规则方法相比，深度强化学习方法可以更好地处理复杂的场景和环境。深度 Q 网络(DQN)是一种经典的基于值的强化学习方法，已应用于无人驾驶的端到端运动规划，在车道跟随、避障、车辆跟随等方面取得了良好的性能[6]。

但上述方法未考虑障碍物类型对障碍策略的影响。对障碍物进行分类可以帮助避免碰撞是因为不同类型的障碍物可能需要不同的应对策略。通过对障碍物进行分类，无人驾驶系统可以识别并理解周围环境中的不同物体或障碍物的类型。这使得系统能够采取特定的行动来避免与障碍物发生碰撞。举例来说，如果系统能够准确地识别出行人，就可以采取减速、停车或远离以确保行人的安全；如果系统能够识别出大型卡车，就可以避免近距离变道造成视野遮挡情况。道路上的障碍物，如小轿车、公交车、行人、骑自行车的人、骑摩托车的人等，它们的尺寸、形态和运动状态可能千差万别。对障碍物准确分类和识别是实现无人驾驶安全行驶的前提。为了保障无人驾驶的安全，现代无人驾驶系统通常配备了多种传感器，如激光雷达、摄像头、毫米波雷达等，用于获取周围环境的信息[7]。利用传感器获取的数据，结合先进的图像处理和深度学习算法，无人驾驶系统可以实时地对道路上的障碍物进行识别和分类[8]。通过分析障碍物的尺寸、形态、运动状态等特征，系统可以做出智能决策，并采取适当的避障策略，从而保证车辆在复杂多变的道路环境中安全行驶。

然而，要实现对多样化障碍物的准确分类与避障并不是一项轻松的任务。不同类型的障碍物之间可能存在相似的外观特征，例如，行人和自行车骑手在某些情况下可能呈现相似的轮廓。此外，道路上的交通环境常常异常复杂，障碍物的运动状态和位置也可能会迅速变化，这为无人驾驶系统的感知和决策能力提出了更高的要求。无人驾驶车辆对道路上多样化障碍物的准确分类与避障是实现驾驶安全的关键一步。

根据以上障碍物分类避障问题，本文提出一种 Classification Security DQN(CSDQN)的车辆行驶决策框架，根据障碍物的不同类型给出具有更高安全性的无人驾驶决策策略，达到提高安全性的目的。

为解决无人驾驶对不同障碍物统一决策带来的安全隐患问题，综合考虑不同障碍物类型，对不同障碍物分配评估系数，促使智能体学习到具有最大安全性的策略。本文主要贡献如下：

1)针对障碍物多样化问题，根据障碍物安全性及特性对障碍物分类，并划定安全等级以便进行细粒度规划。

2)针对安全性问题，提出一种 CSDQN 的车辆行驶决策框架，使得智能体能够找到安全的驾驶策略。

3)针对行驶稳定性问题，设计稳定性约束函数，使智能体能在车道无风险下稳定行驶。

1 相关工作

现有的避障方法主要分为两类：基于运动规划方法和基于机器学习的方法[9]。近年来两个方法在避障领域都取得了巨大的成就。

1.1 基于运动规划的方法

基于运动规划的避障方法涵盖了多种技术和算法，用于在无人驾驶和机器人导航中生成安全、高效的路径并避开障碍物。基于规则的方法在运动规划领域取得了巨大的成就[10-11]，谷歌和 Uber 等许多公司都采用了这些方法。

人工势场法(APF)、三次样条曲线、快速扩展随机树法(RRT)等是一些常见的基于运动规划的避障方法。文献[12]介绍了一种改进的 APF，结合了虚拟参考点的路径规划和障碍物避障。该方法通过引入虚拟参考点以更好地处理复杂环境下的路径规划和避障问题。描述了方法的原理、实验结果以及在移动机器人上的应用，强调了改进方法相对于传统方法的优越性。文献[13]介绍了基于人工势场法在交通中紧急制动和转向的应用，构建了横向变道间距和纵向间距的制动距离模型，引入人工势场法模拟生成安全的行驶路径，并通过模型预测控制(MPC)方法进行控制。文献[14]介绍了基于三次

样条曲线算法在无人驾驶避障问题上的路径规划，通过三次样条曲线的二阶连续性结合车道信息生成平滑可行驶曲线，然而在遇到未知环境或者实时变化的障碍物时，路径规划可能面临数据更新和处理带来的问题。文献[15]介绍了基于快速探索随机树(RRT)算法在赛车的环境中进行决策与控制。然而，在无人驾驶场景中识别相关对象的任务处理，例如车辆、行人和道路环境等[16]，应该在避障决策之前由感知系统完成。这些感知检测是非常耗时的一项任务，可能导致紧急情况发生时反应时间不足。因此，很难在超出规则的情况下应用这些方法，因为它们是基于数学模型设计的，限制于具体定义的规则。并且，上述提到的方法严重依赖于如何生成图形(而不违反物理约束)，并且没有充分考虑车辆动力学，使得生成的路径可能与动力学不兼容，同时也存在一定的局限性，例如在复杂的环境中可能会导致机器人陷入局部极小点或产生震荡。

1.2 基于机器学习的方法

近年来，随着深度学习的不断发展，许多研究人员基于学习的方法解决无人驾驶领域避障问题[17-19]，这些方法可以进一步分为监督学习和强化学习。

在监督学习领域，文献[20]介绍了条件模仿学习模型，在引入了感知输入和控制信号基础上模仿人类驾驶，解决经过端到端训练，以模仿车辆无法在即将到来的十字路口进行特定转弯问题。文献[21]介绍了采用端到端的学习方式，收集大量的人类驾驶行为进行学习，以生成安全的轨迹。上述提到的监督学习方法需要大量人类驾驶的视频数据，网络通过不断学习达到类似于人类驾驶的决策行为。然而数据的收集极为耗时且昂贵，车辆在接近障碍物或碰撞到障碍物的数据收集更加困难。因此，可能导致无人驾驶车辆对危险的感知不足。

在深度强化学习领域，由于深度强化学习可以学习到复杂的策略并且不需要大量数据集的特性，使得深度强化学习在避障领域快速发展，与传统的手动设计规则的方法相比，深度强化学习算法可以更好地处理复杂的场景和环境，从而在躲避障碍物方面表现更好。

近年来深度强化学习已被用于实现无人驾驶车辆的许多决策，包括车道保持[22]、车辆变道[23]、车辆超车[24]。车辆在行驶的过程中可以简单地分为决策控制和车辆控制，决策控制的指令如跟车、变道、超车，结合这些基础决策控制可以实现车辆避障问题。目前，许多研究者对无人驾驶避障问题进行研究，也取得了一定的进展，但是大部分没有对障碍物进行明确分类。因此，这些方法无法根据道路上不同类型的障碍物做出针对性的避障行为，这样的后果就是增加了行驶过程中的风险。

上述方法在自动驾驶避障领域取得了很大的进展，但这些方法未考虑障碍物类型对避障策略的影响。本文基于对障碍物的准确分类结果提出一种 CSDQN 的车辆行驶决策框架，根据障碍物的不同类型找到具有更高安全性的无人驾驶决策策略，达到提高无人驾驶安全性的目的。

2 架构设计与实现

本节将介绍 CSDQN 行驶决策框架，在多种障碍物的双车道场景中实现安全避障。

2.1 问题描述

车辆的避障可以表示为一个马尔可夫决策过程(MDP)，MDP 被定义为 5 个元素(S,A,R,f,γ)，分别表示状态空间、动作空间、即时奖励、状态转换模型和折扣因子。MDP 是一种数学框架，用于建模和求解具有随机性和决策性质的问题，在无人驾驶领域中有广泛应用。

一个策略$\pi_\theta(a\mid s)\overset{\text{def}}{=}P(a\mid s;\theta)$，表示在特定环境 s 及参数 θ 下，从策略 π 采取动作 $a\in A$ 的概率。智能体与环境交互将产生运动轨迹，称为状态-动作样本，MDP 的目标是找到最佳策略 $\pi^*(s)$，即在每个状态下选择一个动作的函数，以使智能体能够最大化累积奖励或期望价值。

在强化学习模型中，智能体通过一系列观察、动作和奖励与环境交互，并选择可以最大化累积奖励的动作[25]。然而，输入的高维度以及动作和所得奖励之间的延迟给传统强化学习(RL)方法带来了挑战[26]。在深度神经网络的帮助下，深度 RL 方法在解决耦合非线性控制或决策问题方面显示出有较大的潜力[27-30]。

具有离散时间步长 $t\overset{\text{def}}{=}\{1,2,\cdots,n\}$ 被用来制定智能体运动规划问题，提供一个数学架构来建模顺序决策过程。根据 MDP 的 5 个元素(S,A,R,f,γ)定义运动规划问题。

状态空间：采用数值信息表示状态。为了在行驶过程中基于智能体和障碍物(OB)的关系做出安全正确的决策，综合考虑两者之间相对信息、车道环境信息、障碍物信息帮助智能体做出决策。智能体和 OB 之间的关系状态可以表达为：

$$S\overset{\text{def}}{=}[\text{type},\text{la},\text{lo},\text{obstacle\_direction},\text{left\_border},\text{right\_border},\beta,\text{dev},\text{direction}]$$

式中：type 表示障碍物的类型；la 和 lo 分别表示智能体与 OB 之间的横向距离和纵向距离；obstacle_direction 表示障碍物所在车道；left_border 和 right_border 分别表示智能体与车道左边界和有边界的距离；β 表示智能体的转角；dev 表示智能体与车道中心线的距离；direction 表示当前车辆所在的车道。参数的表达如图 1 所示。

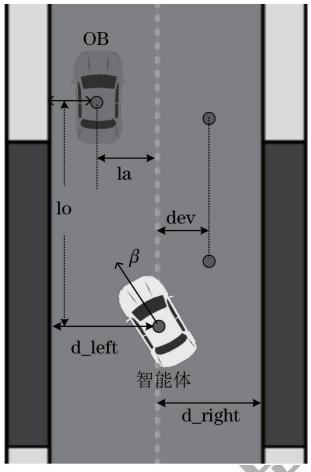

图 1 状态中的参数

Fig. 1 Parameter in the state

行为空间：对所设定的避障需要通过对智能体进行横向控制与纵向控制。考虑到赋予智能体纵向控制权(即加减油门)，智能体很有可能会为了保证更高的安全性采取频繁制动，所以将纵向控制权交给驾驶人员。在时间 t 处使用的转向动作为：

$$\text{Action} \overset{\text{def}}{=} [\text{LSB},\text{LSL},\text{S},\text{RSL},\text{RSB}]$$

式中：LSB 表示大幅度左转；LSL 表示小幅度左转；S 表示保持当前所在方向行驶；RSL 表示小幅度右转；RSB 表示大幅度右转。

奖励：使用深度强化学习的目的是为了找到一个策略，使得预定义奖励函数取得的数值最大化。奖励函数的目的是保证智能体在行驶的过程中能够取得最大的安全性。

安全函数：无人驾驶车辆的安全性是在遵守交通规则的情况下与障碍物保持一定的相对距离。因此，函数设计如下：

$$\text{R\_safe} \overset{\text{def}}{=} \begin{cases} 2 - c1^{-\text{la}} - a^{-\text{lo}}, & \text{当前车道存在 OB} \\ c2, & \text{否则} \end{cases} \tag{1}$$

式中：R_safe 表示安全奖励；la 和 lo 分别表示智能体与 OB 之间的横向距离与纵向距离。式(1)表示当车辆与障碍物不在同一车道的情况下可以获得最大的安全值，若当前车道前方存在障碍物则需要保持一定的车距，该奖励值随着与障碍物的相对距离持续降低。奖励函数 R_type_safe 图形化表示如图 2 所示。

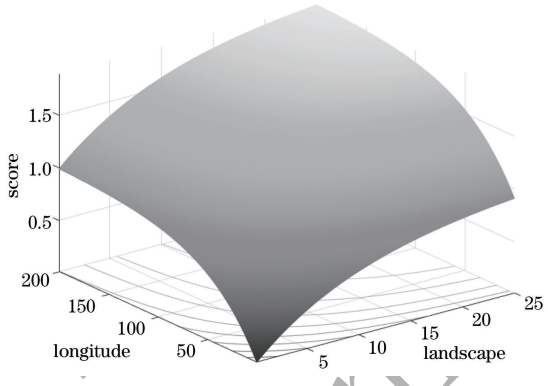

图 2 奖励函数 R_type_safe 图形化表示

Fig. 2 Graphical representations of the reward function R_type_safe

分类安全函数：与安全函数不同，分类安全函数估计不同类型下的具体安全值。根据不同类型，使用 4 个安全性等级 $T \overset{\text{def}}{=} \{\text{lowest},\text{low},\text{high},\text{highest}\}$，其中，$T$ 是安全等级的集合，将 1、2、3、4 的数分别分配给 T 中 lowest、low、high、highest。不同类型的安全等级定义为：$\text{type} \in T \overset{\text{def}}{=} \{1,2,3,4\}$。

在安全函数的基础上增加类型参数，有效增加或减弱智能体对障碍物的感知敏感度。奖励值随着与障碍物的相对距离的缩小持续降低。函数设计如下：

$$\text{R\_type\_safe} \overset{\text{def}}{=} \begin{cases} 0.5 \times \text{type} \times (2 - c1^{-\text{la}} - a^{-\text{lo}}), & \text{当前车道存在 OB} \\ \text{type}, & \text{否则} \end{cases} \tag{2}$$

式中：R_type_safe 表示分类安全奖励，智能体与障碍物不在同一车道，固定获得当前安全等级的奖励值。当前车道前方存在障碍物，智能体与障碍物的相对距离通过安全等级调整，安全等级越高，相对距离越大。该奖励值随着与障碍物的相对距离的缩小持续降低。

为保障智能体在车道内安全行驶，智能体和车道之间设立函数如下：

$$\text{R\_border} \overset{\text{def}}{=} \begin{cases} -\text{e}^{c3\times\text{left\_border}}, & \text{左车道} \\ -\text{e}^{c3\times\text{right\_border}}, & \text{右车道} \end{cases} \tag{3}$$

式中：left_border 和 right_border 分别表示智能体与左车道边界和右车道边界之间的距离，智能体和车道边界之间的相对距离越小，惩罚越大。R_border 的物理意义是保证智能体与车道边界保持一定的距离，由于车道总是存在一定的安全隐患，因此设置惩罚函数抑制向车道边界行驶。

为了遵循人类的驾驶习惯，如人类在驾驶车辆时通常是将车辆行驶在当前车道的中心，这样不仅

视野更宽阔，而且也有更大的容错空间，因此设立如下奖励函数：

$$R\_dev \overset{def}{=} \begin{cases} 0, & \text{当前车道存在 OB} \\ e^{c3\times dev}, & \text{否则} \end{cases} \tag{4}$$

式中：dev 表示智能体与车道中心线的距离，智能体与车道中心线的距离越小，奖励越大。R_dev 的物理意义是鼓励车辆沿着车道中心行驶。

在训练的过程中，智能体可能会出现害怕碰撞到障碍物，因而采取不可控行为减少犯错误的可能，陷入局部最优，显然这样的结果并不符合预期。因此，引入行驶距离因素，鼓励智能体持续向前行驶。

$$R\_distance \overset{def}{=} c4^{\frac{total\_dis}{max\_dis}}/c4 \tag{5}$$

式中：total_dis 表示行驶的距离；max_dis 表示全程距离，全程距离可以通过导航获取。R_distance 将随着智能体行驶距离的增长而增加。

$$R\_collision \overset{def}{=} c5^{\frac{total\_dis}{max\_dis}} \times c6 \tag{6}$$

R_collision 作为一个惩罚函数，将随着智能体行驶距离的增长而减小。

$$R \overset{def}{=} R\_border + R\_dev + R\_type\_safe + R\_distance + R\_collision \tag{7}$$

奖励函数在制定时最好避免奖励函数之间相互冲突，如果当前车道发现障碍物就忽略掉 R_dev 对决策的影响，这样更利于智能体稳健学习。

2.2 Classification Security DQN 决策框架

MDP 为 DQN 提供了一个重要的理论框架和方法，并且为算法提供了基础。DQN 引入了深度神经网络估计 Q 值函数。该网络为多层的前馈神经网络输入状态，输出每个可能动作的 Q 值。为了打破样本之间的相关性，减少训练的方差，提高训练的效率和稳定性，采用经验回放改善训练稳定性。具体是将智能体在环境中的历史经验存储在一个回放缓冲区中，并从中随机抽取小批量样本进行训练。为了进一步提高稳定性，DQN 使用了两个神经网络：主网络（用于估计 Q 值）与目标网络。目标网络的参数较慢地更新到主网络的参数，以减小 Q^- 值目标的变化，从而提高训练的稳定性。

DQN 算法基于 Q-learning 方法，用深度神经网络来估计 Q 值函数。因此，将算法定义为：

$$Q(s,a) \overset{def}{=} Q(s,a) + \alpha \times (r + \times \max(Q(s',a')) - Q(s,a)) \tag{8}$$

式中：$Q(s,a)$ 表示在状态 s 下采取动作 a 的 Q 值；α 表示学习率，用于控制 Q 值的更新步长；r 是在状态 s 下采取动作 a 后获得的即时奖励；γ 是折扣因子，用于权衡即时奖励和未来奖励的重要性；$\max(Q(s',a'))$ 表示下一个状态 s' 下可以采取的最优动作的最大 Q 值。目标 Q 值的计算公式如下：

$$Target\ Q(s,a) \overset{def}{=} r + \gamma \times \max(Q\_target(s',a')) \tag{9}$$

式中：Q_target 表示目标网络估计的 Q 值；$\max(Q\_target(s',a'))$ 表示目标网络在下一个状态 s' 下可以采取的最优动作的最大 Q 值。

损失函数使用均方误差（MSE）损失函数来最小化 Q 值的预测误差。损失函数的计算公式如下：

$$Loss \overset{def}{=} MSE(Q(s,a), Target\ Q(s,a)) \tag{10}$$

式中：$Q(s,a)$ 是网络估计的 Q 值；$Target\ Q(s,a)$ 是目标 Q 值。

CSDQN 的目的是对一个端到端的无人驾驶系统进行避障。CSDQN 体系结构如图 3 所示。在 t

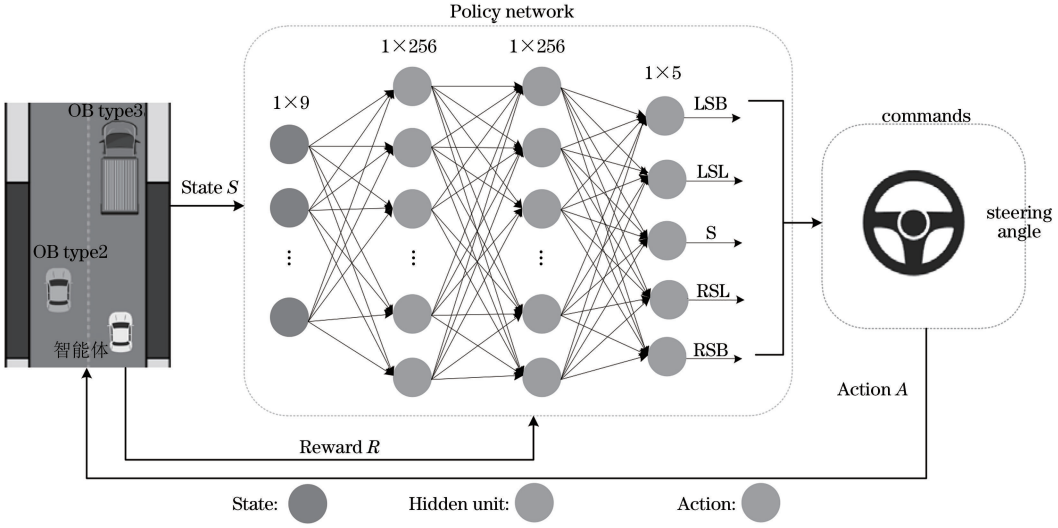

图 3 CSDQN 体系结构

Fig. 3 CSDQN architecture

时刻智能体对获取到的信息预处理为数字信息，并作为模型中的状态馈送。模型接收来自环境 s_t 的状态和全局路径信息，通过运动命令 a_t 作用于车辆，并且在下一次迭代中获得奖励 r_t 作为监督信号。

障碍物分类：模型对感知到的障碍物进行分类，障碍物包括公交、大卡车、小轿车、行人、骑自行车、骑手、狗、猫 8 种。根据安全性可以分为最高、高、中等、低 4 个等级。行人、骑自行车、骑手为最高，公交、大卡车为高，小轿车为中等，狗和猫为低。type $\in T=\{1,2,3,4\}$，取值为 4 表示安全性最高，其余依次类推。不同类型障碍物的避障预期：在进行障碍物精确分类之后需要根据不同类型的避障方案。本文从 2 个方面评估避障方式：1）在智能体与障碍物同车道场景中，不同障碍物确定不同紧急避障范围；2）智能体与障碍物在不同车道场景中根据周围障碍物形状大小以及安全等级确定回到原车道或允许换道避障的时机。对于第 1 个方面的类型 1 障碍物，该类障碍物在智能体角度将给予最低安全等级，所以在检测到障碍物之后降低了智能体与障碍物之间相对距离的敏感度，使得智能体在未进入紧急换道避障范围内首先要考虑周围环境因素，其次进行换道避障。由此，对 4 种障碍物类型的紧急范围分别定义为 10 m、14 m、14 m、21 m；对于第 2 个方面的类型 1 障碍物，由于其体型较小且安全等级低，因此在障碍物中心点 2 m 范围内不允许采取换道决策。其余 3 类禁止换道，范围分别为 4 m、4 m、9 m。

策略网络：由环境在时间步 t 提取的特征转化为数字状态信息馈送到策略网络中，策略网络是上述提到的 Q 网络。在该网络中，一个三层全连接的神经网络被用来组成 Q 函数。这种具有有限数量隐藏单元的神经网络可以一致地近似连续函数。如图 3 中的 Policy netword 所示，因为设计了 9 个输入信息量和 5 个运动命令，所以输入层、隐藏层、隐藏层和输出层的单元编号分别被设计为 9、256、256 和 5。

根据上述的框架，构建方法所对应的代码展示，算法 1 描述了 CSDQN 分类避障的计算过程。

算法 1 分类安全驾驶框架决策的训练

输入 环境 E，动作空间 A，折扣因子 γ，执行轮次 $M=12\ 000$

输出 最优网络参数 θ

1. 初始化记忆库 D=N，表示可以容纳的数据条数 N

2. 随机初始化网络参数 θ，利用 θ 初始化当前行为值 Q_θ。

3. 令目标网络参数 $\theta^-=\theta$，利用 θ^- 初始化目标行为值函数 Q_{θ^-}。

4. For episode = 1，…，M do//执行 for 循环语句，进行//M 次循环

5. 初始化环境 $E=E_k$

6. If M%10 等于 0 do E = evaluate_E//每 10 轮执行//一次评估环境

7. If 完成环境 E 或上一环境为 evaluate_E THEN 环境 $E=test\_E_{k+1}$

8. END IF

9. 从环境 E 中初始化状态 s

10. While 未到终点且未发生碰撞 do

11. 以概率(1−ε)获取行为 $a_t=\arg\max_a Q_\pi(s,a,\theta)$，否则随机选择行为 a_t

12. 在环境中执行行为 a_t，使用式(7)计算奖励 r_t 和下一步的状态 s_{t+1}

13. 将当前序列$(s_t,a_t,r_{t+1},s_{t+1})$存入记忆库 D 中

14. 从记忆库中 D 中随机抽取最小批数据 B

15. 算法的目标 $y_j=r_i+\gamma\max_{a'}Q(s_{j+1},a',\theta^-)$

16. B 个数据序列的损失函数为：$J(\theta)=\frac{1}{B}\sum_{i=1}^{B}(y_i-Q(s_j,a_j,\theta))^2$

17. 对 J(θ)做梯度下降法：$\nabla\theta=\frac{\partial J(\theta)}{\partial J\theta_j}=\frac{2}{B}\sum_{i=1}^{B}(y_i-Q(s_j,a_j,\theta))\nabla Q(s_j,a_j,\theta)$

18. 使用∇θ 对网络参数进行更新：$\theta=\theta+\nabla\theta$

19. End while

20. End for

3 实验测试

通过 Python pygame 开发库实现避障环境仿真测试。pygame 开发库可用于开发 2D 应用程序场景，能够表达设计想法，该开发库展现了环境的高兼容性，可以仿真大量车辆模型、行人、动物等。硬件环境为 Intel Corei9-12900HX 2.30 GHz，16.0 GB RAM。所有的代码都是在 PyTorch 框架下编写完成。

3.1 训练设置

使用 pygame 模拟一个椭圆形环道，道路中存在不同类型的障碍物如图 4 所示，目的是绕开途中存在的障碍物抵达终点，目标是提高分类躲避障碍物的安全能力。考虑到智能体可能存在保守制动问题，因此在环境中采用匀速行驶，障碍物采用静态形式。首先智能体对场景行学习，环境中设置小汽车、行人、动物、大卡车、公交车。pygame 开发底层是通过像素开发而成的，通过这一理论可以仿真出驾驶员的视角，作为获取环境的传感工具，获取到的信

息皆为数字信息。为了能够预留避障通道，实验设计两条同向车道。将车道分为 6 个间隔段设置障碍物，障碍物的位置通过 Python 随机生成，每个间隔段生成一个障碍物，所以不可能出现障碍物并排现象。在实验中，模型被训练了上万轮，以学习最佳的避障策略。在每一轮中，车辆从起点开始行驶，完成路线或撞车（与小汽车、卡车、行人、动物）、或与道路发生碰撞，则结束此轮。在训练阶段，车辆以 1 的概率随机选择车辆运动决策，之后由所提出的 CSDQN 生成车辆运动决策。在这个阶段，概率从 1 线性下降到 0.05，之后保持 0.05。在测试阶段，按照所划定的区域随机生成障碍物。为了验证 CSDQN 的有效性，分别选择了 2 种典型的强化学习算法的代表性的方法作为比较方法以及本文设计的安全模型 SDQN 方法。一种是文献[31]提出的 DQN 方法，另一种是文献[32]提出的原始 Dueling DQN（DDQN）方法，它是 Q-网络的变体，属于策略梯度的范畴。它们是两类强化学习中最受欢迎的模型，并在无人驾驶方面取得了良好的效果。另外，增加 SDQN 是为了体现其安全性，以此作为对比是为了体现出 CSDQN 分类的性能提升。因此，使用此 3 个模型作为比较方法。

对于所提出的奖励函数中的系数 $c1=1.1$，$c2=2$，$c3=-0.25$，$c4=5$，$c5=-0.2$，$c6=30$，当 type 为 1 时，$a=1.03$，当 type 为 2、4 时，$a=1.02$，当 type 为 3 时，$a=1.01$。赛道全长 2 600 m。在实验的过程中，为了保证实验对比的客观性，挑选出

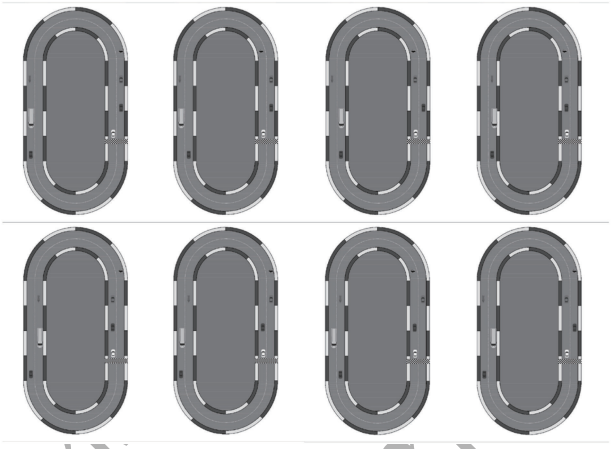

图 4 评估场景快照

Fig. 4 Evaluate scene snapshots

8 个障碍物固定位置的地图用于评估，如图 4 所示。实验设置从开始训练每经过 10 轮，随机抽取一个评估地图进行评估。

3.2 学习性能

本文提出一个测试学习性能的方法，并与其他方法进行比较，以此证明 CSDQN 在避障方面的优势。由于奖励是评价强化学习方法学习性能的常用标准，因此使用奖励相对于迭代步长的曲线（步长-奖励曲线），如图 5(a)所示，展示不同方法的学习性能[33]。此外，由于训练时的障碍物是随机产生，可能产生避障的难易程度不同，因此训练可能缺乏客观性。为解决此问题，采用图 4 中的评估地图作为训练评估。实验结果如图 5(a)和 5(b)所示，基于 CSDQN 的避障方法具有比其他方法更好的学习性能。

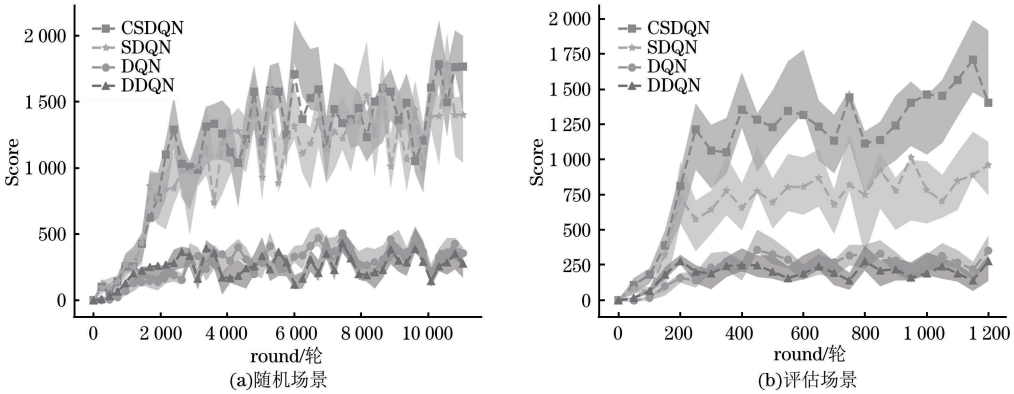

图 5 不同方法训练场景分数对比

Fig. 5 Comparison of training scenario scores using different methods

图 5(a)所示的曲线表示算法进行避障训练回报。横坐标表示训练轮次，纵坐标表示智能体所在环境中获得的总奖励值。本文所提出的 CSDQN 和 SDQN 能够更快地收敛，并且获得的奖励相对更高，也就意味智能体行驶的距离更远。线条下部的阴影部分表示置信区间。由于智能体在前 2 000 轮之前是处于探索阶段（从 1 到 0.05），因此所显示的方法在 2 000 轮前都是处于快速上升阶段，到 2 000 轮后逐步优化趋于平稳。

图 5(b)所示的曲线表示算法进行避障评估的回报。共进行 1 250 轮评估，评估是抽取 8 个固定环境中的任意一个，并且评估环境在训练环境中并

未给出，因此该图更具有客观性。从图中可以看出，CSDQN 相较于其他 3 种方法是更优的，阴影区域表示评分的标准差。

3.3 安全性能

安全性是无人驾驶最为重要的标准。通常安全性的评估是通过当前车辆与障碍物的横向距离与纵向距离所决定的，本文目标是考虑障碍物的类型，需要将障碍物类型加入评估体系，式(2)中包含无人驾驶车辆与障碍物车辆的相对距离以及障碍物类型的影响[9]。所以，使用该公式的取值更加合理。图 6(a)和图 6(b)所示为不同方法的安全性能。本文采用最新 4 种方法的训练模型分别对随机场景和评估场景进行安全性测试。对随机场景和评估场景进行 200 轮和 8 轮测试。

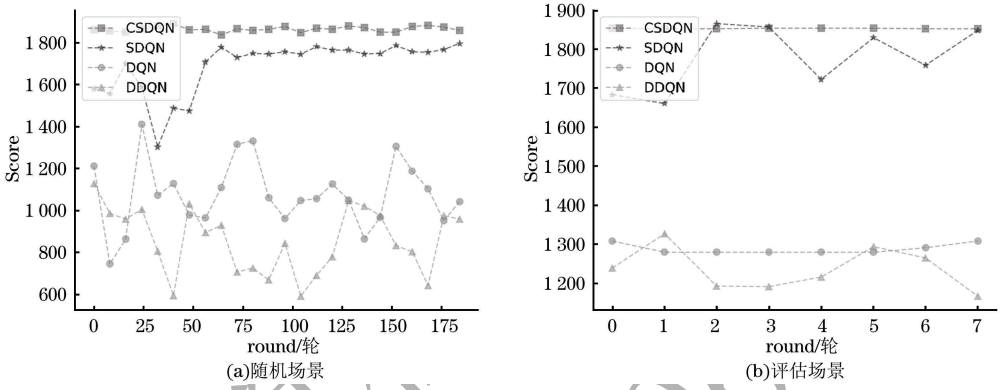

图 6 不同方法训练场景的安全性分数对比

Fig. 6 Comparison of security scores for training scenarios using different methods

从图 6(a)和图 6(b)可以看出，本文提出的 CSDQN 在不同的评估场景中均能获得稳定且较高的安全分数，SDQN 也能够完成大部分评估场景，而 DQN、DDQN 则在评估场景中的安全性较低。在面向安全的无人驾驶场景中，CSDQN 具有更大的安全性能。

3.4 驾驶距离以及稳定性能

结合本文所给出的仿真场景，无人驾驶车辆能够以稳定的状态行驶到终点，这也是评价无人驾驶的重要指标[34]。结合仿真场景使无人驾驶车辆能够行驶至终点，也可以通过行驶距离来评判。此外，面对不同障碍物摆放位置，无人驾驶车辆依然能展现稳定行驶的能力。图 7 所示为不同方法的驾驶距离及稳定性能力。该实验通过对 200 组不同障碍物的随机摆放位置的场景进行测试。

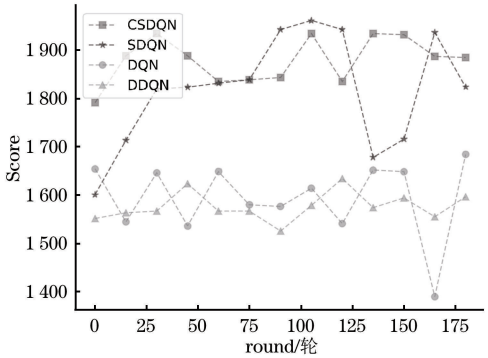

图 7 不同方法随机场景行驶距离对比

Fig. 7 Comparison of driving distances in random scenes using different methods

从图 7 可以看出，本文提出的 CSDQN 在不同的评估场景中行驶了较远距离，并且从整体产生的数据曲线可以看出是比较稳定的。SDQN 能够完成大部分场景，其中主要是遇到大型障碍物时导致不稳定发生。而 DQN、DDQN 在评估场景中的驾驶距离以及稳定性能较差。

通过上述三组对比实验可知，无人驾驶在多类型障碍物的场景中，本文提出方法与深度强化学习 DQN、DDQN 以及增加安全函数的 SDQN 相比，具有更好的学习性能以及较高的安全性及稳定性。

4 结束语

本文提出一种 CSDQN 的车辆行驶决策框架，根据障碍物的不同类型得到一个具有更高安全性的无人驾驶决策策略，达到提高无人驾驶安全性的目的。首先，决策框架对检测到的障碍物根据障碍物的安全性等级进行分类。然后，根据不同类型障碍物对无人驾驶车辆的影响提出安全评估函数，无人驾驶车辆基于障碍物与自身的横向距离以及纵向距离来评估安全性。最后，CSDQN 决策框架利用障碍物类型、相对位置信息以及安全评估函数经过不断训练获得最终模型。仿真结果表明，本文提出方法与深度强化学习 DQN、DDQN 以及增加安全函数的 SDQN 方法相比，在多种障碍物的情况下具有更好的学习性能以及较高的稳定性及安全性。

参考文献

[1] TENG S Y, HU X M, DENG P, et al. Motion planning for autonomous driving: the state of the art and future perspectives[J]. IEEE Transactions on Intelligent Vehicles, 2023, 8(6): 3692-3711.

[2] 章军辉,陈大鹏,李庆. 自动驾驶技术研究现状及发展趋势[J]. 科学技术与工程,2020, 20(9): 3394-3403.
ZHANG J H, CHEN D P, LI Q. Research status and development trend of technologies for autonomous vehicles [J]. Science Technology and Engineering, 2020, 20(9): 3394-3403. (in Chinese)

[3] CHEN Y, CHEN S Z, REN H B, et al. Path tracking and handling stability control strategy with collision avoidance for the autonomous vehicle under extreme conditions[J]. IEEE Transactions on Vehicular Technology, 2020, 69(12): 14602-14617.

[4] SILVER D, HUANG A, MADDISON C J, et al. Mastering the game of Go with deep neural networks and tree search [J]. Nature, 2016, 529: 484-489.

[5] GU S X, HOLLY E, LILLICRAP T, et al. Deep reinforcement learning for robotic manipulation with asynchronous off-policy updates[C]//Proceedings of IEEE International Conference on Robotics and Automation. Washington D. C., USA: IEEE Press, 2017: 3389-3396.

[6] MNIH V, BADIA A P, MIRZA M, et al. Asynchronous methods for deep reinforcement learning[EB/OL]. [2023-09-30]. https://arxiv.org/pdf/1602.01783.

[7] BADUE C, GUIDOLINI R, CARNEIRO R V, et al. Self-driving cars: a survey [J]. Expert Systems with Applications, 2021, 165: 113816.

[8] 茅智慧,朱佳利, 吴鑫, 等. 基于 YOLO 的自动驾驶目标检测研究综述[J]. 计算机工程与应用, 2022, 58(15): 68-77.
MAO Z H, ZHU J L, WU X, et al. Review of YOLO based target detection for autonomous driving [J]. Computer Engineering and Applications, 2022, 58(15): 68-77. (in Chinese)

[9] LI G F, QIU Y F, YANG Y F, et al. Lane change strategies for autonomous vehicles: a deep reinforcement learning approach based on transformer[EB/OL]. [2023-09-30]. https://arxiv.org/pdf/2304.13732.

[10] HE Y, LIU Y, YANG L, et al. Deep adaptive control: deep reinforcement learning-based adaptive vehicle trajectory control algorithms for different risk levels[EB/OL]. [2023-09-30]. https://arxiv.org/pdf/2023.03408.

[11] 钱玉宝,余米森,郭旭涛,等. 无人驾驶车辆智能控制技术发展[J]. 科学技术与工程, 2022, 22(10): 3846-3858.
QIAN Y B, YU M S, GUO X T, et al. Development of intelligent control technology for unmanned vehicle [J]. Science Technology and Engineering, 2022, 22(10): 3846-3858. (in Chinese)

[12] WANG W R, ZHU M C, WANG X M, et al. An improved artificial potential field method of trajectory planning and obstacle avoidance for redundant manipulators [J]. International Journal of Advanced Robotic Systems, 2018, 15(5): 172988.

[13] FENG S, QIAN Y B, WANG Y. Collision avoidance method of autonomous vehicle based on improved artificial potential field algorithm [J]. Journal of Automobile Engineering, 2021, 235(14): 3416-3430.

[14] 周慧子,胡学敏, 陈龙, 等. 面向自动驾驶的动态路径规划避障算法[J]. 计算机应用, 2017, 37(3): 883-888.
ZHOU H Z, HU X M, CHEN L, et al. Dynamic path planning for autonomous driving with avoidance of obstacles [J]. Journal of Computer Applications, 2017, 37(3): 883-888. (in Chinese)

[15] FERACO S, LUCIANI S, BONFITTO A, et al. A local trajectory planning and control method for autonomous vehicles based on the RRT algorithm[C]//Proceedings of AEIT International Conference of Electrical and Electronic Technologies for Automotive. Washington D. C., USA: IEEE Press, 2020: 1-6.

[16] HNEWA M, RADHA H. Object detection under rainy conditions for autonomous vehicles: a review of state-of-the-art and emerging techniques[J]. IEEE Signal Processing Magazine, 2021, 38(1): 53-67.

[17] BEN ELALLID B, BENAMAR N, MRANI N, et al. DQN-based reinforcement learning for vehicle control of autonomous vehicles interacting with pedestrians [C]// Proceedings of International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies. Washington D. C., USA: IEEE Press, 2022: 489-493.

[18] BIN ISSA R, DAS M, RAHMAN M S, et al. Double deep Q-learning and faster R-CNN-based autonomous vehicle navigation and obstacle avoidance in dynamic environment [J]. Sensors, 2021, 21(4): 1468.

[19] SAXENA D M, BAE S, NAKHAEI A, et al. Driving in dense traffic with model-free reinforcement learning[C]// Proceedings of IEEE International Conference on Robotics and Automation. Washington D. C., USA: IEEE Press, 2020: 5385-5392.

[20] CODEVILLA F, MULLER M, LOPEZ A, et al. End-to-end driving via conditional imitation learning[C]//Proceedings of IEEE International Conference on Robotics and Automation. Washington D. C., USA: IEEE Press, 2018: 4693-4700.

[21] SADAT A, CASAS S, REN M Y, et al. Perceive, predict, and plan: safe motion planning through interpretable semantic representations [C]// Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2020: 414-430.

[22] WANG Z, YAN Z H, NAKANO K. Comfort-oriented haptic guidance steering via deep reinforcement learning for individualized lane keeping assist[C]//Proceedings of IEEE International Conference on Systems, Man and Cybernetics. Washington D. C., USA: IEEE Press, 2019: 4283-4289.

[23] ZHOU W, CHEN D, YAN J, et al. Multi-agent reinforcement learning for cooperative lane changing of connected and autonomous vehicles in mixed traffic [J]. Autonomous Intelligent Systems, 2022, 2(1): 5.

[24] LI X X, QIU X Y, WANG J, et al. A deep reinforcement learning based approach for autonomous overtaking[C]// Proceedings of IEEE International Conference on Communications. Washington D. C., USA: IEEE Press, 2020: 1-5.

[25] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning [J]. Nature, 2015, 518: 529-533.

[26] MOUSAVI S S, SCHUKAT M, HOWLEY E. Deep reinforcement learning: an overview[M]. Berlin, Germany: Springer, 2017.

[27] KIRAN B R, SOBH I, TALPAERT V, et al. Deep reinforcement learning for autonomous driving: a survey[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(6): 4909-4926.

[28] HOEL C J, WOLFF K, LAINE L. Automated speed and lane change decision making using deep reinforcement learning[C]//Proceedings of the 21st International Conference on Intelligent Transportation Systems. Washington D. C., USA: IEEE Press, 2018: 2148-2155.

[29] WANG J J, ZHANG Q C, ZHAO D B, et al. Lane change decision-making through deep reinforcement learning with rule-based constraints[C]//Proceedings of International Joint Conference on Neural Networks. Washington D. C., USA:

IEEE Press, 2019: 1-6.
[30] CYBENKO G. Approximation by superpositions of a sigmoidal function[J]. Mathematics of Control, Signals and Systems, 1989, 2(4): 303-314.
[31] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Playing atari with deep reinforcement learning[EB/OL]. [2023-09-30]. https://arxiv.org/pdf/1312.05602.
[32] WANG Z Y, SCHAUL T, HESSEL M, et al. Dueling network architectures for deep reinforcement learnings[C]// Proceedings of the 33rd International Conference on Machine Learning. Washington D. C., USA: IEEE Press, 2016: 2939-2947.
[33] WANG G, HU J M, LI Z H, et al. Harmonious lane changing via deep reinforcement learning [J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(5): 4642-4650.
[34] CHEN L, HU X M, TANG B, et al. Conditional DQN-based motion planning with fuzzy logic for autonomous driving[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(4): 2966-2977.

编辑 索书志