

智能车辆规划与控制策略学习方法综述

龚建伟¹, 龚乘¹, 林云龙¹, 李子睿^{1,2}, 吕超¹

(1. 北京理工大学 机械与车辆学院, 北京 100081; 2. 代尔夫特理工大学, 荷兰, 代尔夫特 2628CN)

摘 要: 智能车辆相关技术已实现了长足的发展, 并已能够在有限封闭场景中实现自主行驶的基本功能。然而, 实际道路测试结果表明, 目前智能车辆技术仍存在较多局限, 而智能车辆在复杂城市与越野环境的大规模应用仍面临较多挑战。作为智能车辆关键技术之一, 运动规划与控制技术已基本建立了完整的理论体系并得到较多工程应用, 但传统方法在实际应用中仍存在动态复杂场景理解能力弱、场景适应性差、模型复杂度高、参数调整难度大等缺陷。由于机器学习方法具备较强的知识表征与模型拟合能力, 其已经在智能车辆的感知与导航技术中得到了广泛的应用。而为了解决传统运动规划与控制技术存在的泛化性与适用性等问题, 许多研究者近年来也开始探索基于深度学习、强化学习等机器学习方法的运动规划与控制方法。本文将对目前基于机器学习的智能车辆规划与控制方法研究现状进行回顾, 从规划与控制策略基本架构、基本学习范式以及基于学习的规划与控制方法三方面对现有智能车辆规划与控制策略学习方法进行分析, 最后对研究现状与未来发展方向进行总结与展望。

关键词: 智能车辆; 机器学习; 运动规划与控制; 模型预测控制

中图分类号: TP18,U461

文献标志码: A

文章编号: 1001-0645(2022)07-0665-10

DOI: 10.15918/j.tbit1001-0645.2022.095

Review on Machine Learning Methods for Motion Planning and Control Policy of Intelligent Vehicles

GONG Jianwei¹, GONG Cheng¹, LIN Yunlong¹, LI Zirui^{1,2}, LÜ Chao¹

(1. School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100081;

2. Delft University of Technology, Delft, Holland, 2628CN)

Abstract: Intelligent vehicles have achieved a considerable development in technologies and can fulfill the basic functions of autonomous driving in a limited closed environment. However, results of actual road tests show that the current technologies of intelligent vehicles still have many limitations and their large-scale application in complex urban and off-road environments still faces many challenges. As one of the key technologies, the motion planning and control technology has basically formed a complete theoretical system and has been widely applied in engineering. However, the traditional methods still have some defects in practical application, such as the inability of understanding dynamic and complex scenes, poor adaptability for different scenes, high complexity of the model, and difficulty in parameter tuning. Due to the strong ability in knowledge representation and model fitting, machine learning methods have been widely applied in perception and navigation technology for intelligent vehicles. In order to solve the problems of generalization and applicability in traditional motion planning and control techniques, many researchers have also devoted themselves to exploring the usage of deep learning, reinforcement learning, and so on machine learning methods in motion planning and control policy for intelligent vehicles. In this paper, machine learning-based methods were reviewed for motion planning and control in

收稿日期: 2022-04-10

作者简介: 龚建伟(1969—), 男, 博士, 教授, E-mail: gongjianwei@bit.edu.cn.

通信作者: 吕超(1980—), 男, 博士, 副教授, E-mail: chaolu@bit.edu.cn.

intelligent vehicles, analyzing the existing policy learning methods for motion planning and control from three aspects, including basic framework, basic learning paradigms, and different planning and control methods based on learning. Finally, the research status and future development directions were summarized and prospected.

Key words: intelligent vehicles(IV); machine learning; motion planning and control; model predictive control.

智能车辆是集环境感知、行为决策、规划与控制等多项功能于一体的综合智能系统,涵盖了机械、控制、传感器技术、信号处理、模式识别、人工智能和计算机技术等多学科知识.我国智能车辆相关技术的研究近年来取得长足的进展,能够满足智能车辆在小范围、简单城区环境下的低速行驶要求,以及城际公路简单环境中的高速行驶需求.然而,要实现在城区与越野等动态复杂交通环境下的智能车辆自主行驶,仍有很多技术亟待提升与完善.其中,智能车辆运动规划与控制技术,作为智能车辆主要关键技术之一,在复杂动态真实场景中的应用仍面临较大挑战.

智能车辆的规划与控制技术主要包含智能车辆运动规划与智能车辆运动控制两个模块.其中,运动规划技术的主要作用是根据传感器探测的周围环境信息与对场景的理解,规划出一条从起点到终点的安全、平缓、可行的最优运动轨迹.运动规划模块得到的路径、速度等规划信息将传递到控制模块,而运动控制模块再根据当前周围环境和位移、姿态、速度等环境知识按照驾驶任务需求,并分别向油门、制动及转向等执行系统发出解算的控制指令,进而控制车辆的姿态与行为,使得车辆能够尽量精确地跟踪运动规划模块得到的轨迹.

传统规划与控制算法通过构建规划与控制多层框架在相对较低的速度和静态环境中已经得到较为成熟的应用,但在应对复杂动态环境方面仍存在较大局限性.传统运动规划方法通常包括基于搜索的方法、基于采样的方法、基于优化的方法三类,但其分别在路径搜索成本、路径采样结果最优性、模型非线性等方面存在一定的局限性,同时传统运动规划方法在应对动态性较高的场景时仍面临极大挑战.对于运动控制方法,在相对较低的速度下,通常基于车辆的运动学模型设计控制策略,其中比例-积分-导数(PID)控制^[1]、反馈线性化与模型预测控制(model predictive control, MPC)^[2]都是实际工程中常用的方法.然而,在高速下操作或执行激进机动时,这些常用方法则存在较大的误差因而难以应用.非线性最优控制^[3-4]、MPC^[5]或反馈-前馈控制^[6]在智能车辆高

速稳定规划控制中取得了一定的成果,然而其存在的实时性、调参难度、以及场景适应性问题则限制了其应用范围.

在提高智能车辆的智能化程度方面,机器学习技术应用越来越广泛.近年来,许多研究者开始关注结合机器学习的规划控制算法或利用机器学习框架替代传统规划控制框架以提升智能车辆的规划控制性能.例如,许多互联网企业以及院校利用深度强化学习(deep reinforcement learning, DRL)^[7-8]、模仿学习(imitation learning, IL)^[9-11]等机器学习方法,设计了端对端(End-to-End)的整套规划控制系统,实现了由感知层得到的数据信息直接获得最终的控制指令等功能.另一方面,部分研究者则利用逆强化学习(inverse reinforcement learning, IRL)^[12]、深度卷积神经网络(deep convolutional neural network, DPNN)^[13]、自适应神经网络(adaptive neural network, ANN)^[14]等机器学习方法对现有规划与控制技术进行模块化替代或改进,进而提升传统规划控制方法的效率与自适应性,其中的一些方法还在高速越野等极端工况中取得了较为优异的测试表现^[9].

本文对国内外基于机器学习的规划控制方法的研究现状进行回顾与总结.除了选择合适的学习范式与学习策略外,环境知识的表征方式也将极大影响智能车辆规划与控制策略学习效果.合理、有效的环境知识表征是基于机器学习的规划与控制算法的学习前提,通过选择相关性强的环境感知动静态数据,减低数据输入维度的同时保证环境信息的完备性,进而保证策略学习效率并降低策略学习成本.进一步地,利用场景表征信息,基于机器学习的规划与控制方法将应用不同的机器学习范式对规划、控制策略进行学习,进而最终保证智能车辆安全、高效的行驶表现.

本文第1节介绍环境知识表征的相关内容,第2节介绍两类常用的规划与控制学习范式——模仿学习与强化学习,在前两节的基础上,第3节进一步介绍具体的基于机器学习的规划控制算法,具体地,将分别回顾基于机器学习的端对端方法、规划方法以及控制方法,并评述现有的相关研究,

最后进行总结与展望. 图1所示为本文各节逻辑架构.

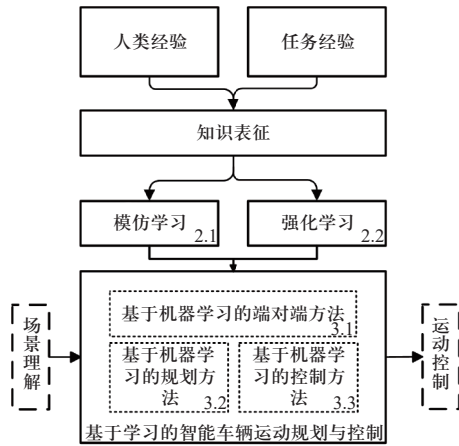


图1 本文各节逻辑架构

Fig. 1 Logic framework of this paper

1 环境知识表征

环境知识表征指的是对感知得到的数据信息进行知识理解与表征。作为连接感知信息与规划控制知识的桥梁,环境表征模块将为规划控制模块提供可用的信息。在环境知识表征中,对于交通场景的理解是其中重要的一个组成部分。许多研究致力于实现对场景中交通参与者的行为进行理解与预测(如车辆与行人的轨迹预测、驾驶意图预测等)。例如文献[15]使用图神经网络对行人与车辆的交互行为进行建模,提出了一种基于图神经网络的分层行为预测方法。此外,文献[16]还提出了基于迁移学习(transfer learning, TF)的驾驶行为建模方法,在数据有限的情况下实现了模型对不同驾驶员间驾驶行为的迁移学习。来自西班牙巴勃罗·德·奥拉维德大学的PEREZ等^[17]等研究者从人类演示中学习了一种使用卷积神经网络的轨迹预测器,并将预测的路径作为粗略的代价地图来校正RRT规划器。德国弗莱堡大学的KRETZSCHMAR等^[12]使用最大熵概率分布来对具有社会行为的智能体轨迹进行建模,并结合轨迹交互建模与道路几何结构作为输入进行路径规划。在上述方法中,提升驾驶行为的预测精度是许多新方法的研究方向,除了设计不同的网络与机制来更好地表征各交通参与者的行为,还有一类方法则通过直接应用机器学习中的集成学习(ensemble learning)范式提升环境知识表征模型的表现。例如南阳理工的XING等^[18],通过集成学习,提高了基于循环神经网络模型

对驾驶行为意图(如换道意图、车道保持意图)的预测表现。类似的,LI等^[19]则提出了一种基于集成学习框架,用于提高交互场景中的轨迹预测模型的表现。此外,KIM等^[20]使用逆强化学习方法,在传统的规划控制系统中,基于IRL在提取出的特征中学习局部代价函数,使得局部路径规划能够使用学习的代价函数来满足规划问题中的人车交互约束。HENRY等^[21],同样使用IRL来学习代价函数表征,使得智能车辆具备在车流中无碰撞移动、避开所处交通场景的拥挤区域的能力。基于机器学习的场景表征,还被用于遵循社会交互之外的其他导航风格。例如,STEIN等^[22],将子目标定义在已知和未知空间的边界上,并利用神经网络从全局目标的某个子目标学习机动成本,如到达该子目标后陷入死路的概率,从而提升智能车辆机动效率。又如WIGNESS等^[23],利用人类演示和最大熵IRL学习局部代价地图,使地面平台能够模仿人类演示者的导航风格。

2 规划与控制的常用学习范式

基于环境、场景的知识表征,现有大多数基于机器学习的运动规划与控制方法常使用模仿学习与强化学习(reinforcement learning, RL)两种学习范式进行规划控制算法的学习。本节将分别介绍模仿学习与强化学习。

2.1 模仿学习

模仿学习也称为示教学习(learn from demonstration),其本质是将先验知识通过知识表征提供给学习模型,使模型能够模仿人类专家的行为。通常地,在模仿学习中,智能体通过寻求一个由专家演示的训练数据集的最佳使用方式来学习一定任务规则,并尽可能执行与专家相近的行为^[24],从而达到“机器代替人”的学习目的。

在基于模仿学习的运动规划与控制方法中,可以通过将车辆复杂的运动分解为运动基元(motion primitives, MPs),进而在基元层面进行规划与控制策略的学习。例如,从未标注的驾驶数据中提取出运动基元,并将学习到的运动基元与一个预先建立的基元库进行关联,实现了使用动态移动基元来表征驾驶员轨迹的目的^[25-26]。

此外,也可以采用扩充训练数据的方式提高模型的模仿学习表现,并在损失函数的设计中加入关注事件的损失项,以此提高模型性能。例如,谷歌大脑(Google Brain)及Waymo公司的BANSAL等^[27]则

将感知模块得到的数据与作为专家演示的驾驶行为数据进行融合,进而基于融合的数据对车辆的驾驶策略进行学习.同时,该研究还对模仿学习的损失函数增加了对应处理,如碰撞或驶出道路的不良驾驶情况事件的损失项,避免模型学习人类的错误驾驶经验.类似地,百度自动驾驶技术部门的 ZHOU 等^[28]也在模仿学习的损失函数中引入用于处理碰撞等因素的损失项.此外,他们还设计了一种用于数据扩增的反馈合成器,使得模型可获得更多先前未曾学习过的不同环境中的驾驶数据,通过积累更丰富的驾驶经验,进而提高运动规划策略学习的性能.

另一类模仿学习的方式则是通过训练人工神经网络来替代传统运动规划与控制模块功能或是使用模仿学习方法与传统规控算法结合的形式,实现更加高效的规划与控制策略学习.例如,佐治亚理工学院的 PAN 等^[9],通过训练一个深度神经网络来模仿学习一个装备有高级传感器的模型预测控制器的控制策略,最终神经网络生成一套可由高维度感知数据到连续方向盘转角以及节气门开度的控制指令的映射方法.该方法的优势在于不需要对状态进行估计,也无需使用实时规划算法就可实现对车辆的规划与控制.美国波士顿大学的 ZHANG 等^[29]通过转换主车所观察到的其他车辆信息视角,利用其他车辆的驾驶演示来学习驾驶策略,从而提高了场景数据的利用率.卡耐基梅隆大学的 RHINEHART 等^[30]则使用模仿学习结合目标导向式规划算法,以期望驾驶行为的概率预测模型规划出解释性强的专家式轨迹,以实现到达指定的目标的轨迹规划.

此外,伊朗谢里夫大学的 SAMSAMI 等^[10]提出结合因果分析的模仿学习模型(causal imitative model, CIM)来处理基于模仿学习的自动驾驶车辆算法所产生的不利惯性与碰撞问题.大部分传统的模仿学习范式是基于知识表征、演示直接学习任务,而该研究则将模型输入分解为潜在的因果变量,在传统的模仿学习框架中加入因果分析的过程,最终利用因果分析得到的变量进行运动规划,避免了由于车辆惯性与碰撞引起的问题.还有一类模仿学习与本文下一节介绍的强化学习方法有关——逆强化学习. IRL 一方面与环境交互另一方面从专家演示中获得经验,反向去估计奖励函数,基于所估计的奖励函数,再通过强化学习的方式对模型进行优化.通过对奖励函数进行学习,IRL 能够避免人工设定奖励函数可能带来的偏差.

2.2 强化学习

强化学习主要通过智能体与环境互动并获得反馈评价的机制对策略进行学习,其类似于人类通过“试错”进行某一任务的学习.智能体在某一特定环境下进行不同的行为行动,依据环境的特性或特定规则,将得到不同的环境反馈,通过接收积极的反馈(如人类认知中的“奖励”)与消极的反馈(如人类认知中的“惩罚”),使得智能体在与环境的交互中能够学习到在特定的环境下应该采取怎样的行动是符合期待奖励的,而应该避免哪些行动从而规避惩罚.环境的反馈在强化学习中,一般通过设置奖励函数(reward function)来实现.具体地,强化学习可通过马尔可夫决策过程来进行数学描述.马尔可夫决策过程可通过状态空间 S (state space), 状态空间中所有可能的行动 A (Action), 状态转移概率 P 以及奖励函数 R (reward function), 4 个参数进行描述.首先,智能体观察环境状态空间 S 中的一系列状态.根据所观察的状态,智能体从可能的行动集合 A 中选择一个行动并执行该动作,由于智能体的行动对环境是产生影响的,环境状态依据此变化以及相应的状态转移概率更新环境状态.同时,智能体将收到奖励函数的反馈,函数的大小反映对智能体在观察状态下执行此动作的评价优劣.而智能体的学习目标就是通过对其某一规则映射的学习,使得累积的奖励函数最大化,以此实现智能体与环境互动并获取反馈,最终实现任务的学习.

强化学习依据学习机制通常分为三大类:基于策略梯度(policy gradient)的强化学习,基于价值函数(value-based)的强化学习以及基于执行者-批评者(actor-critic)模型的强化学习.其中,基于策略梯度的方法是对策略进行参数化描述,在训练过程中优化策略参数,使之获得最大的奖励函数值.这个优化过程常以梯度下降法进行损失函数的最小化,计算出的损失函数梯度下降方向称为策略梯度,模型参数以策略梯度进行参数更新.较为典型的基于策略梯度的强化学习方法有 Reinforcement^[31]等.加拿大拉瓦尔大学的 DESJARDINS 等^[32]利用策略梯度强化学习设计了一种自适应巡航(ACC)控制方法,韩国汉阳大学的 CHAE 等^[33]则基于深层确定性策略梯度(deep deterministic policy gradient)方法设计的控制方法降低了实验中的车辆碰撞率.基于价值函数的强化学习方法则通过对价值函数的估计,选择可带来最大奖励值的动作.当环境状态转移概率 P 已知时,

该方法将选择可使当前状态下对应期望奖励最大的行动.当环境状态转移概率未知时,则使用通过贪心算法最大化状态-行动价值函数(state-action value function)来寻找最优的学习策略.此类方法较为典型的算法如Q-Learning^[34]等,在规划控制方面的应用中^[35-37]均将深度学习网络与基于价值的强化学习方法进行结合,实现机器人的导航功能.上述两类方法均具有一定的局限性——基于策略梯度的方法的主要缺点是算法所估计的策略梯度方差较大,基于价值函数的方法主要缺点则在于无法保证学习后的策略是最优策略^[38].而第三类执行者-批评者模型结合价值函数与参数化的策略函数,可视为基于策略梯度与基于价值函数方法的综合,较为典型的算法如A3C等^[39].此类方法较好地权衡了前两类方法的利弊,例如研究基于执行者-批评者模型设计出不同的纵向控制方法^[40-42],在实验中取得了较好的性能表现.

3 基于机器学习的规划控制方法

基于机器学习的规划、控制算法较之传统的基于规则、优化的方法,在处理更加复杂的场景、任务时具备更好的性能表现.本节将首先介绍基于机器学习的端对端方法,接着分别回顾国内外基于机器学习的规划方法及控制方法研究现状.

3.1 基于机器学习的端对端方法

大多数基于学习的端对端规划方法直接将原始感知数据或经一定预处理的感知数据作为输入,并直接输出规划、控制指令,即将整个智能系统作为黑箱使用,而无需阐明各模块及其子模块间的关系.基于智能车辆导航目标的不同,端对端方法主要分为基于固定目标的规划控制方法与基于移动目标的规划控制方法,其中固定目标的规划控制方法指的是基于设定的确定位置,系统能够使智能车辆安全、高效地到达此确定位置.而移动目标的规划控制方法中,系统则偏于使智能车辆保持一定状态或对移动位置进行跟踪,例如车辆跟驰、车道保持等.

针对固定目标的端对端规划控制,瑞士苏黎世理工的PFEIFFER等^[43]在2017年提出了一种将原始激光雷达数据以及固定目标位置映射到低级别速度指令的端对端神经网络方法,结合传统全局与局部控制的模仿学习方法来训练网络,实现固定目标无碰撞规划.为提高算法的适应能力(如改变目的地、行驶环境),德国弗莱堡大学的ZHANG等^[35]在2017

年结合深度强化学习方法,通过使用深度卷积神经网络在离散动作空间(动作包含停留、左转、右转以及前进动作)上执行导航任务,在仿真与实物实验中均取得了较好的效果.同样,来自弗莱堡大学的ZHELO等^[36],在2018年同样利用深度强化学习方法训练模型进行无先验地图知识的规划控制,他们通过设置内在奖励函数获得更快的学习速率,较大提升了模型训练效率.此外,近年来还有许多研究者提出针对“多智能体”(multi-agents)、固定目标的端对端机器学习技术.如麻省理工学院的CHEN等^[37]使用DRL训练可直接将激光雷达信息及目标位置映射至低级别导航信息的神经网络,该算法考虑环境中多个智能体移动的不确定性(周边的移动机器人),可在有多智能体参与的动态环境下为固定目的地进行无碰撞的速度、路径规划.

虽然移动至设定目的地是多数导航算法的基本目标,但针对车辆驾驶任务而言,仍有大量基于学习的端对端方法尝试去解决更为具体的智能车辆规划控制问题——动目标运动规划控制,即任务目的地并不是固定的位置.与这类问题相关的驾驶行为包含车道保持、避障以及地形适应等.纽约大学的LECUNN等^[13]使用一个六层的卷积神经网络,训练了一个端对端模式的模型,可将原始的图片数据映射输出为车辆速度与方向盘转角,用于车辆在越野环境中的高车速避障巡航.来自英伟达公司BOJARSKI等^[11]则利用更深层的卷积神经网络以及大量的人类驾驶员数据建立了一个名为DAVE-2的自动驾驶系统,DAVE-2表明卷积神经网络可以学习车道和道路跟踪任务,而无需将任务分解为独立的感知、推理和规划步骤.

3.2 基于机器学习的规划方法

相比于端对端的方法,模块化方法具备更好地解释性同时具备更强的实用性,在许多场景已取得实际应用.牛津大学MARKUS WULFMEIER等^[44]设计了一个基于最大熵的非线性逆向强化学习生成同时满足人类驾驶员习惯与系统设定误差要求的运动规划框架,其中使用完全卷积神经网络(fully convolutional networks, FCN)来代表隐藏在成本函数下的人类驾驶员的驾驶行为,并采用大量实车数据验证了该方法的可行性.类似的,加州大学伯克利分校DOROSA等^[45]也采用IRL来获取人机交互复杂系统下人类驾驶员规划的价值函数,当智能车辆在相似环境下行驶时,通过采用此价值函数可逆向规划出

类似该人类驾驶员的操作轨迹. 湖南科技大学肖浩等^[46]基于预测后再规划的思想提出面向实际环境的无人车驾驶系统框架, 并结合深度 Q 学习和深度预测网络技术提出一种快速全局路径规划方法, 减少了车辆的行车时间. 上海理工大学刘磊^[47]等设计了一种基于生存理论的局部路径规划策略学习方法, 实现了快速且高精度的路径优化. 文献 [25] 提出了一种融合运动基元的路径规划算法, 通过机器学习算法实现对人类驾驶员的驾驶经验学习并以基元形式表征, 再对基元进行平滑组合, 实现基于类人经验的运动规划, 并通过基元泛化改善了生成轨迹的可调整性.

在模型自适应方面的研究, 德克萨斯大学奥斯汀分校的 LIU 等^[48]在 2021 年提出了一种具备自演进能力的、在不同环境中具备持续学习能力的机器人规划控制框架. 该研究针对现有的经典规划方法无法在不同的环境种积累经验而提升自规划能力的局限性, 提出了基于持续学习的规划框架, 结合了传统方法, 通过基于对传统方法无法克服的复杂场景规划控制方案进行持续学习, 实现地面无人平台在复杂环境中更快速稳定的行驶. 美国陆军研究院的 GARRETT WARNELL 等^[49], 在 2021 年提出了一种基于学习的自适应规划方法. 他们认为无人地面平台经典的导航系统虽然可以提供点到点的无碰撞路径规划, 但在一些特定场景下往往给出次优的规划路径, 而若加入人工干预修正, 次优结果很容易校正为最优结果. 根据这一思路, 他们提出的基于学习的自适应规划方法 (adaptive planner parameter learning from interventions, APPLI) 能够从人类干预系统的过程中学习, 以提升导航性能. 他们的真实环境试验结果表明, 较之使用固定参数的规划方案, 使用 APPLI 的地面移动机器人有着更好的表现. 同时, 他们还在 300 个不同的仿真环境中做了模拟实验, 表明 APPLI 具有较好的泛化能力.

在传统的运动规划模块中, 有一部分参数是可调节参数, 调节参数以使模型适应不同的场景环境、规划偏好等, 例如膨胀半径、采样率以及轨迹优化权重等. 一些研究则致力基于机器学习的方法实现参数调节, 从而代替人工, 实现对这些可调参数的自适应调整. 美国佐治亚理工学院的 Bhardwaj 等^[50]在全局规划层面提出了一种可微高斯过程运动规划 (Gaussian process motion planning 2, GPMP2) 算法的可微扩展模块. 该扩展模块旨在通过对专家演示的

学习, 调节障碍协方差这一规划参数. 通过反向传播学习对应参数, GPMP2 可规划出类似人类专家的全局路径. 类似的思路也被应用在局部规划中, 西班牙巴斯克大学的 TESO 等^[51]提出一种可预测 DWA, 通过对传统的 DWA 算法增加一个可预测窗口, 并使用人工神经模糊推理系统 (artificial neuro-fuzzy inference system, ANFIS) 来优化每一个固定的参数值——即 DWA 的优化权重, 以提升 DWA 的性能表现. 来自美国奥斯汀大学及美国陆军研究院的 XIAO 等^[52]提出自适应规划参数学习 (adaptive planner parameter learning, APPL). 该方法可视作一种参数学习范式——所学习的策略不直接用于端对端的运动控制, 而是将所学策略与传统的运动控制器结合, 学习参数其参数调整策略, 并在运行时动态调整规划参数, 以适应不同的驾驶场景, 实现优于固定参数规划策略的效果.

3.3 基于机器学习的控制方法

基于机器学习的类人驾驶控制方法方面, 俄亥俄州立大学 SCOTT^[53]提出由补偿传递函数和基于道路几何的预期分量组成的组合驾驶员模型并于单个驾驶员期望路径相结合, 实现比几何中心线模型更准确的方向盘转角. 麻省理工学院 CHONG 等^[54]提出一种基于规则的神经网络模型模拟驾驶员在跟踪过程和安全检测两种驾驶情况下的驾驶员行为, 从车辆轨迹数据中获取驾驶员驾驶规则, 并运用机器学习方法训练神经网络使其能够模拟单个驾驶员的驾驶行为. 加州大学伯克利分校的 STEPHANE 等^[55]提出了可以从人类示意中学习的自动驾驶框架, 将其应用自动驾驶的纵向控制并通过在线和离线两种方式模拟驾驶员的操纵行为. LU 等^[56]则提出了一种新的自动超车分层强化学习框架, 该框架基于半马尔可夫决策过程和运动基元, 可应用于不同的超车阶段.

在结合预测模型相关的控制方法研究方面, 多伦多大学 CHRIS J. OSTAFEW 等^[57]提出了一种用于自主移动机器人的基于学习算法的非线性 MPC (LB-NMPC) 算法, 以减少沿着参考路径的重复遍历的路径跟踪误差, 该系统可以从通用的先验车辆模型开始, 随后根据经验的学习以减少车辆和轨迹特定的路径跟踪误差. 佐治亚理工 PAN 等^[58]通过训练一个能将原始高维的观测结果映射到连续转向和油门命令的深度神经网络控制策略, 实现了对 MPC 器的模拟. 多伦多大学的 MCKINNON 等^[59]提出了一种随

机 MPC(stochastic model predictive control, SMPC)方法,该方法假设平台动力学控制过程为一个未知的非线性函数表征,通过学习这个未知函数和它的逆函数,进而对控制输入进行优化与表征,实现了更准确且实时的控制。

在基于机器学习的极端工况控制技术方面,麻省理工学院的 RUS 等^[60]关注 F1TENTH 竞速机器人的表现,对比了基于先进的有模型深度强化学习方法与无模型方法在自动驾驶竞速环境应用时的表现,对比结果揭示了基于有模型深度强化学习方法在极端工况中的表现胜于无模型的方法。来自南洋理工大学的 JI 等^[14]提出一种针对车辆驾驶极限工况的、基于 ANN 的横向控制方法。该方法将基于李亚普诺夫稳定性理论的自适应控制机制与径向基神经网络(radial basis function neural network, RBFNN)相结合,设计了用于估计轮胎转弯刚度不确定性的 ANN 逼近器,通过学习近似非线性函数来减少其非线性性带来的不利影响。美国陆军研究院的 MAGGIE WIGNESS 等^[23]针对灾难恢复与救援等动态环境中的快速适应与学习问题,考虑到特殊场景下人员和设备受到的限制,提出了一种基于学习的视觉感知和逆最优控制的有监督学习模型。该模型将视觉感知模型与逆最优控制学习结合,实现地面无人平台控制参数的学习,并使用最大熵理论来学习从视觉感知中提取的给定环境特征的奖励函数,以及由人类在环境中驱动机器人收集的最优轨迹示例,并进行了真实环境中的多个测试,得到了较好的测试结果。

4 结 论

随着计算机科学的发展,机器学习能够实现对复杂场景与任务较强的拟合与学习,已在无人平台感知与定位方面取得了较为广泛的应用。许多国内外研究机构也逐渐开始探索机器学习在运动规划控制算法上的应用,通过将机器学习算法与传统运动规划控制算法结合,或者直接基于机器学习框架取代传统运动规划控制算法框架,以提高对动态复杂环境的适应能力与算法自身的学习能力,突破目前运动规划控制算法的局限。

针对基于机器学习的运动规划方法研究,国外研究开展较早且已经有了一定的基础,包括院校和军方研究所在内的多家机构都进行了相关研究。美国麻省理工学院、卡内基梅隆大学、德国弗莱堡大学以及加拿大麦吉尔大学等院校都先后提出了基于

机器学习的运动规划算法的框架,通过结合机器学习方法对运动规划参数进行调整,实现算法在不同环境的更强适应性,并进行了相关实验。除院校以外,美国陆军研究院也对机器学习在无人平台运动规划方法上的应用产生了较大兴趣,提出了基于机器学习的自适应规划算法以提升智能车辆在复杂环境中的规划能力。总体上来讲,目前基于机器学习方法的运动规划相关研究仍处于起步阶段,且现有方法多关注于考虑较为简单场景中的运动规划策略学习方法,难以应用到更复杂的城市与越野场景。同时,现有研究多只针对环境进行场景知识表征,缺乏针对更复杂的智能车辆驾驶任务知识表征相关研究,难以同时考虑不同环境与任务中的场景特征与任务知识,如舒适性、经济型、越野通过性等需求,因而不同环境与不同任务中的最优运动规划策略构建与学习方面仍存在较大研究空间。

相较于基于机器学习的运动规划方法研究,基于机器学习的运动控制算法的相关研究取得了更多的关注且已取得了一定实际应用成果。谷歌、英伟达、百度等企业积极开发基于机器学习的端对端控制技术,力图从大数据中学习得到最优的智能车辆控制策略。美国麻省理工学院、卡内基梅隆大学、加拿大多伦多大学等院校各自利用机器学习优化了运动控制模块,提升了智能车辆运动控制性能。此外,美国军方和等企业也在积极开展相关研究,美国陆军研究院针对无人车动态环境下的快速适应和学习任务,提出了使用机器学习的视觉感知和最优控制模型。总体上来讲,基于端对端的智能车辆控制策略学习相关研究开展较早且取得了一定进展。针对端对端方法存在的数据饥渴、可解释性差、学习成本高等局限,许多研究者也开始探索对模块化的运动控制策略进行学习。然而,国内外现有方法多通过优化学习模型与增大数据量实现运动控制策略的提升,较少关注数据稀缺场景中的运动控制模型学习技术,同时,多数方法多以离线模型学习为主,较少关注运动控制模型的持续学习与在线优化问题,难以实现模型在更复杂与动态环境中的自适应能力与自学习能力。

综上所述,基于机器学习的智能车辆运动规划与控制方法尽管能够一定程度解决传统方法存在的一些问题,但仍存在一定局限性,未来发展方向主要包括:

(1)研究动态复杂场景中的运动规划与控制策

略学习方法.藉由机器学习较强的环境知识表征与场景理解能力,基于学习的智能车辆运动规划与控制方法在复杂动态场景中已经初步表现出了更强的泛化与适应能力.因此,未来相关研究应进一步利用机器学习方法的优点,深入研究复杂动态场景中的运动规划与控制策略学习方法,将研究与实际应用相结合,尤其应当更多考虑具有更多周围动态交通参与者的城市场景以及具有变化复杂地形的越野场景.

(2)研究具有更强解释性、模块化的运动规划与控制策略学习方法.早期基于机器学习的运动规划与控制方法多关注于端对端的学习由感知数据至运动控制的策略模型,然而由于数据维度高、模型复杂等因素导致该类策略学习方法往往具备较差的可解释性,限制了策略的调整与优化,同时一定程度上可能导致鲁棒性与安全性问题.提高基于学习的运动规划与控制策略方法的可解释性是未来相关研究的重要方向,一方面可以通过合理知识表征与可解释网络构造实现,另一方面可以通过模块化学习运动规划与控制策略实现对传统方法的优化与替代,进而保留传统分层框架的可解释性.

(3)研究具有持续学习与持续进化能力的运动规划与控制策略学习方法.相比于传统方法,基于学习的算法能够基于数据实现策略模型参数的自主调整.然而现有策略学习方法大多依赖于先验知识或预先训练实现模型的参数学习,而无法在运行过程中利用实际操作经验实现模型参数的在线调整与优化.因此,研究策略模型基于实际操作经验的自主学习和进化方法,实现智能车辆在操作过程中在实际运行环境中性能的不断提升与进化,也是未来运动规划与控制策略学习重要方向.

参考文献:

- [1] NARANJO J E, GONZALEZ C, GARCIA R, et al. Lane-change fuzzy control in autonomous vehicles for the overtaking maneuver[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2008, 9(3): 438 – 450.
- [2] KLANČAR G, ŠKRJANC I. Tracking-error model-based predictive control for mobile robots in real time[J]. *Robotics and Autonomous Systems*, 2007, 55(6): 460 – 469.
- [3] VERSCHUEREN R, ZANON M, QUIRYNEN R, et al. Time-optimal race car driving using an online exact hessian based nonlinear mpc algorithm[C]//Proceedings of 2016 European Control Conference (ECC), [S.l.]: ECC, 2016: 141 – 147.
- [4] VERSCHUEREN R, FERREAU H J, ZANARINI A, et al. A stabilizing nonlinear model predictive control scheme for time-optimal point-to-point motions[C]//Proceedings of 2017 IEEE 56th Annual Conference on Decision and Control (CDC), 12–15 Dec. [S.l.]: IEEE, 2017: 2525 – 2530.
- [5] 刘凯, 龚建伟, 陈舒平, 等. 高速无人驾驶车辆最优运动规划与控制的动力学建模分析[J]. *机械工程学报*, 2018, 54(14): 141 – 151.
LIU Kai, GONG Jianwei, CHEN Shuping, et al. Dynamics modeling analysis of optimal motion planning and control of high-speed unmanned vehicles[J]. *Journal of Mechanical Engineering*, 2018, 54(14): 141 – 151. (in Chinese)
- [6] KAPANIA N R, GERDES J C. Design of a feedback-feedforward steering controller for accurate path tracking and stability at the limits of handling[J]. *Vehicle System Dynamics*, 2015, 53(12): 1687 – 1704.
- [7] WURMAN P R, BARRETT S, KAWAMOTO K, et al. Outracing champion gran turismo drivers with deep reinforcement learning[J]. *Nature*, 2022, 602(7896): 223 – 228.
- [8] CHEN J, YUAN B, TOMIZUKA M. Model-free deep reinforcement learning for urban autonomous driving[C]//Proceedings of 2019 IEEE Intelligent Transportation Systems Conference. [S.l.]: ITSC, 2019: 2765 – 2771.
- [9] PAN Y, CHENG C-A, SAIGOL K, et al. Imitation learning for agile autonomous driving[J]. *The International Journal of Robotics Research*, 2019, 39(2-3): 286 – 302.
- [10] SAMSAMI M R, BAHARI M, SALEHKALEYBAR S, et al. Causal imitative model for autonomous driving [J]. arXiv preprint arXiv: 211203908, 2021.
- [11] BOJARSKI M, DEL TESTA D, DWORAKOWSKI D, et al. End to end learning for self-driving cars[J]. arXiv preprint arXiv: 160407316, 2016.
- [12] KRETZSCHMAR H, SPIES M, SPRUNK C, et al. Socially compliant mobile robot navigation via inverse reinforcement learning[J]. *The International Journal of Robotics Research*, 2016, 35(11): 1289 – 1307.
- [13] MULLER U, BEN J, COSATTO E, et al. Off-road obstacle avoidance through end-to-end learning[J]. *Proceedings of the 18th International Conference on Neural Information Processing Systems*, [S.l.]: MIT Press, 2005: 5-8.
- [14] JI X, HE X, LV C, et al. Adaptive-neural-network-based robust lateral motion control for autonomous vehicle at driving limits[J]. *Control Engineering Practice*, 2018, 76: 41 – 53.
- [15] LI Z, GONG J, LU C, et al. Interactive behavior prediction for heterogeneous traffic participants in the urban road: a graph-neural-network-based multitask learning framework[J]. *IEEE/ASME Transactions on Mechatronics*, 2021, 26(3): 1339 – 1349.
- [16] LU C, HU F, CAO D, et al. Transfer learning for driver model

- adaptation in lane-changing scenarios using manifold alignment[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2020, 21(8): 3281 – 3293.
- [17] PÉREZ-HIGUERAS N, CABALLERO F, MERINO L. Learning human-aware path planning with fully convolutional networks[C]//Proceedings of 2018 IEEE International Conference on Robotics and Automation, [S.l.]: IEEE,2018: 5897 – 5902.
- [18] XING Y, LÜ C, WANG H, et al. An ensemble deep learning approach for driver lane change intention inference[J]. *Transportation Research Part C: Emerging Technologies*, 2020, 115: 102615.
- [19] LI Z, LIN Y, GONG C, et al. An ensemble learning framework for vehicle trajectory prediction in interactive scenarios[J/OL]. 2022, <https://arxiv.org/abs/2202.10617>.
- [20] KIM B, PINEAU J. Socially adaptive path planning in human environments using inverse reinforcement learning[J]. *International Journal of Social Robotics*, 2016, 8(1): 51 – 66.
- [21] HENRY P, VOLLMER C, FERRIS B, et al. Learning to navigate through crowded environments[C]//Proceedings of 2010 IEEE International Conference on Robotics and Automation. [S.l.]: IEEE, 2010: 981 – 986.
- [22] STEIN G J, BRADLEY C, ROY N. Learning over subgoals for efficient navigation of structured, unknown environments [R]. Conference on Robot Learning. [S.l.]:PMLR,2018: 213 – 222.
- [23] WIGNESS M, ROGERS J G, NAVARRO-SERMENT L E. Robot navigation from human demonstration: learning control behaviors[C]. 2018 IEEE International Conference on Robotics and Automation (ICRA). [S.l.]:IEEE,2018: 1150 – 1157.
- [24] ATTIA A, DAYAN S. Global overview of imitation learning[J]. arXiv preprint arXiv: 180106503, 2018.
- [25] WANG B, GONG J, CHEN H. Motion primitives representation, extraction and connection for automated vehicle motion planning applications[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2020, 21(9): 3931 – 3945.
- [26] 王博洋, 龚建伟, 张瑞增, 等. 基于真实驾驶数据的运动基元提取与再生成[J]. *机械工程学报*, 2020, 56(16): 155 – 165.
WANG Boyang, GONG Jianwei, ZHANG Ruizeng et al. Motion primitives extraction and regeneration based on real driving data[J]. *Journal of Mechanical Engineering*, 2020, 56(16): 155 – 165.
- [27] BANSAL M, KRIZHEVSKY A, OGALE A. Chauffeurnet: learning to drive by imitating the best and synthesizing the worst[J]. arXiv preprint arXiv: 181203079, 2018.
- [28] ZHOU J, WANG R, LIU X, et al. Exploring imitation learning for autonomous driving with feedback synthesizer and differentiable rasterization[C]//Proceedings of 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). [S.l.]: IEEE, 2021: 1450 – 1457.
- [29] ZHANG J, OHN-BAR E. Learning by watching[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. [2021]: <https://arxiv.org/abs/2106.05966>.
- [30] RHINEHART N, MCALLISTER R, LEVINE S. Deep imitative models for flexible inference, planning, and control[J]. arXiv preprint arXiv: 181006544, 2018.
- [31] WILLIAMS R J. Reinforcement-learning connectionist systems[M]. [S.l.]: College of Computer Science, Northeastern University, 1987.
- [32] DESJARDINS C, CHAIB-DRAA B. Cooperative adaptive cruise control: a reinforcement learning approach[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2011, 12(4): 1248 – 1260.
- [33] CHAE H, KANG C M, KIM B, et al. Autonomous braking system via deep reinforcement learning[C]//Proceedings of 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC). [S.l.]:IEEE,2017: 1 – 6.
- [34] WATKINS C, DAYAN P. Q-Learning[M]//Machine Learning, Boston: Kluwer Academic Publishers, 1992, 8: 279 – 292.
- [35] ZHANG J, SPRINGENBERG J T, BOEDECKER J, et al. Deep reinforcement learning with successor features for navigation across similar environments[C]//Proceedings of 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). [S.l.]: IEEE, 2017: 2371 – 2378.
- [36] ZHELO O, ZHANG J, TAI L, et al. Curiosity-driven exploration for mapless navigation with deep reinforcement learning[J]. arXiv preprint arXiv: 180400456, 2018.
- [37] CHEN Y F, LIU M, EVERETT M, et al. Decentralized non-communicating multiagent collision avoidance with deep reinforcement learning[C]//Proceedings of 2017 IEEE International Conference On Robotics And Automation (ICRA). [S.l.]: IEEE, 2017: 285 – 292.
- [38] KUUTTI S, BOWDEN R, JIN Y, et al. A survey of deep learning applications to autonomous vehicle control[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2020, 22(2): 712 – 733.
- [39] MNIH V, BADIA A P, MIRZA M, et al. Asynchronous methods for deep reinforcement learning[C]//Proceedings of the 33rd International Conference on International Conference on Machine Learning. [S.l.]: ICML, 2016: 1928 – 1937.
- [40] SALLAB A E, ABDOU M, PEROT E, et al. End-to-end deep reinforcement learning for lane keeping assist[J]. [2016]. <https://arxiv.org/abs/1612.04340>.
- [41] ZHAO D, HU Z, XIA Z, et al. Full-range adaptive cruise control based on supervised adaptive dynamic programming [J]. *Neurocomputing*, 2014, 125: 57 – 67.
- [42] WANG B, ZHAO D, LI C, et al. Design and implementation of an adaptive cruise control system based on supervised actor-

- critic learning[C]. 2015 5th International Conference on Information Science and Technology (ICIST). IEEE: 243 – 248.
- [43] PFEIFFER M, SCHAEUBLE M, NIETO J, et al. From perception to decision: a data-driven approach to end-to-end motion planning for autonomous ground robots[C]// Proceedings of 2017 IEEE International Conference On Robotics and Automation (ICRA). [S.l.]: IEEE: 2017: 1527 – 1533.
- [44] WULFMEIER M, RAO D, WANG D Z, et al. Large-scale cost function learning for path planning using deep inverse reinforcement learning[J]. *The International Journal of Robotics Research*, 2017, 36(10): 1073 – 1087.
- [45] SADIGH D, LANDOLFI N, SASTRY S S, et al. Planning for cars that coordinate with people: leveraging effects on human actions for planning and active information gathering over human internal state[J]. *Autonomous Robots*, 2018, 42(7): 1405 – 1426.
- [46] 肖浩, 廖祝华, 刘毅志, 等. 实际环境中基于深度Q学习的无人车路径规划[J]. *山东大学学报(工学版)*, 2021, 51(1): 100 – 107.
- XIAO Hao, LIAO Zhuhua, LIU Yizhi, et al. Unmanned vehicle path planning based on deep q learning in real environment[J]. *Journal of Shandong University (Engineering Science)*, 2021, 51(1): 100 – 107.
- [47] 刘磊, 杨晔, 刘赛, 等. 基于生存理论训练机器学习的智能驾驶路径生成方法[J]. *控制与决策*, 2020, 35(10): 2433 – 2441.
- LIU Lei, YANG Ye, LIU Sai, et al. Path generation method for intelligent driving based on machine learning trained by viability theory[J]. *Control and Decision*, 2020, 35(10): 2433 – 2441.
- [48] LIU B, XIAO X, STONE P. A lifelong learning approach to mobile robot navigation[J]. *IEEE Robotics and Automation Letters*, 2021, 6(2): 1090 – 1096.
- [49] WANG Z, XIAO X, LIU B, et al. APPLI: Adaptive planner parameter learning from interventions[M]. 2021 IEEE International Conference on Robotics and Automation (ICRA). 2021: 6079 – 6085.
- [50] BHARDWAJ M, BOOTS B, MUKADAM M. Differentiable Gaussian process motion planning[C]//Proceedings of 2020 IEEE International Conference on Robotics and Automation (ICRA). [S.l.]:IEEE,2020:10598 – 10604.
- [51] TESO-FZ-BETOÑO D, ZULUETA E, FERNANDEZ-GAMIZ U, et al. Predictive dynamic window approach development with artificial neural fuzzy inference improvement[J]. *Electronics*, 2019, 8(9):
- [52] XIAO X, WANG Z, XU Z, et al. APPL: Adaptive planner parameter learning[J]. *arXiv Preprint arXiv: 210507620*, 2021,
- [53] SCHNELLE S, WANG J, SU H-J, et al. A personalizable driver steering model capable of predicting driver behaviors in vehicle collision avoidance maneuvers[J]. *IEEE Transactions on Human-Machine Systems*, 2016, 47(5): 625 – 635.
- [54] CHONG L, ABBAS M M, MEDINA FLINTSCH A, et al. A rule-based neural network approach to model driver naturalistic behavior in traffic[J]. *Transportation Research Part C:Emerging Technologies*, 2013, 32: 207 – 223.
- [55] LEFEVRE S, CARVALHO A, BORRELLI F. A learning-based framework for velocity control in autonomous driving[J]. *IEEE Transactions on Automation Science and Engineering*, 2015, 13(1): 32 – 42.
- [56] LU C, WANG H, LV C, et al. Learning driver-specific behavior for overtaking: a combined learning framework[J]. *IEEE Transactions on Vehicular Technology*, 2018: 1 – 1.
- [57] OSTAFEW C J, SCHOELLIG A P, BARFOOT T D. Robust constrained learning-based nmpe enabling reliable mobile robot path tracking[J]. *The International Journal of Robotics Research*, 2016, 35(13): 1547 – 1563.
- [58] PAN Y, WANG J. Model predictive control of unknown nonlinear dynamical systems based on recurrent neural networks[J]. *IEEE Transactions on Industrial Electronics*, 2012, 59(8): 3089 – 3101.
- [59] MCKINNON C. Learning-based path-tracking control for ground robots with discrete changes in dynamics[D]. Canada: University of Toronto, 2021.
- [60] BRUNNBAUER A, BERDUCCI L, BRANDSTÄTTER A, et al. Latent imagination facilitates zero-shot transfer in autonomous racing[J]. *arXiv Preprint arXiv: 210304909*, 2021.

(责任编辑: 孙竹凤)