基于强化学习的 TCU 换挡策略及实时控制算法实现

戴鹏清
杭州聚晟节能科技有限公司　浙江　杭州　310000

摘　要：本研究围绕基于强化学习的 TCU 换挡策略与实时控制算法展开探究，剖析强化学习基础理论，以及其在汽车传动控制范畴的独特优势，构建涵盖车辆动力学特性、实际驾驶工况等多元要素的状态空间体系。以燃油经济性、驾乘舒适性等为关键考量，精心设计奖励函数，进而完成换挡策略模型架构，深入钻研深度 Q 网络、策略梯度等算法于实时控制环节的具体应用，着重探讨算法优化路径与硬件部署要点。此项研究成果为车辆传动系统性能提升、智能化换挡控制实现给予理论依据与技术支持。

关键词：强化学习；TCU；换挡策略；实时控制算法；车辆传动系统

引言

伴随汽车产业朝着智能化与节能化趋势演进，自动变速器控制单元（TCU）换挡策略的改进已然成为增强车辆性能的核心要点，过往的换挡策略大多依据固定准则或经验算式构建，在应对错综复杂的驾驶场景以及各式各样的驾驶诉求时存在局限性，常出现燃油经济性欠佳、换挡平稳性缺失等状况。强化学习作为机器学习范畴内的关键组成部分，拥有于动态环境下自主研习、完善决策的特性，为 TCU 换挡策略的探索开辟崭新路径。把强化学习融入 TCU 换挡策略以及实时控制算法当中，能够达成更为智能、高效的换挡操控，优化车辆整体性能表现，对于促进汽车传动系统技术革新有着深远意义。

1 强化学习理论基础与应用优势

1.1 强化学习基本原理

强化学习的关键在于搭建起智能体和环境彼此交互的学习架构，智能体如同一位决策者，持续对周围环境加以观测，获取当下环境的状态信息，依据自身所掌握的策略，智能体从一系列可行的动作中挑选出一个来执行。动作作用于环境之后，环境状态会随之改变，并且会向智能体反馈一个奖励信号，这个奖励信号是智能体对自身决策好坏进行判断的关键因素，其代表着智能体执行某一动作后在当前状态下所得到的即时收益，智能体的最终目的是经过长时间与环境的交互，积累尽可能多的奖励，进而学习到能使长期累计奖励最大化的最优策略。从数学模型的角度来讲，一般会运用马尔可夫决策过程对强化学习问题进行描述，马尔可夫决策过程具备无记忆特性，也就是说当前状态涵盖了做出决策所需要的所有信息，未来状态仅仅由当前状态和所采取的动作决定，和之前的状态与动作并无关联。

以马尔可夫决策过程为基础，利用贝尔曼方程能够求解出最优价值函数，从而确定最优策略，在实际的应用情形中，常见的强化学习算法大体可划分为三类[1]。一类是基于价值的方法，诸如 Q - learning 和深度 Q 网络（DQN），其重点在于对状态-动作价值函数展开学习，凭借评估不同状态下各个动作的价值高低，来引导智能体做出决策；另一类是基于策略的方法，例如策略梯度算法，该方法直接针对策略的参数实施优化，通过对参数的调整促使策略朝着获取更多奖励的方向发展；还有一类是基于模型的方法，其致力于对环境模型的学习，借助对环境的预测来协助智能体进行决策。

1.2 在 TCU 换挡控制中的应用优势

相较传统换挡策略，强化学习于TCU换挡控制中彰显显著优势，传统换挡策略多依托预先制定的换挡MAP图，这种基于固定转速 - 车速二维坐标的静态模式，一经标定便难以灵活变更。于实际驾驶场景内，遭遇突发交通管制、临时施工路段等特殊状况，或是车辆因乘客数量变动、货物装卸致使负载改变，传统策略常难迅速应对。像城市潮汐车道转换时，车辆频繁处于低速缓行与短时加速交替状态，传统策略因机械执行既定逻辑，引发不必要的挡位升降，使得发动机在低效区域运转，进而加重燃油损耗与机械磨损。

强化学习可将车辆速度、发动机转速、油门开度、道路坡度、车辆负载等诸多影响换挡决策的要素，统一纳入状态空间加以综合分析，借由动态感知特性，在暴雨天气湿滑路面行车时，系统可融合轮胎附着力传感器数据，延缓降挡操作，防止因扭矩突变造成打滑；于高速公路长距离巡航过程中，依据实时风速风向调节升挡时机，维持发动机处于经济转速区间。无论是城市道路频繁启停、高速公路平稳巡航，还是山区道路连续爬坡

等各类复杂驾驶情境，强化学习均能使智能体通过自主学习，探寻适配各场景的换挡策略，强化学习在多目标优化层面具备独特灵活性。合理设计奖励函数，便能对燃油经济性、驾驶舒适性、动力性等重要性能指标进行平衡，着眼燃油经济性，对促使发动机于高效油耗区域运转、适时提前升挡、减少无谓降挡的换挡操作予以正向激励；为保障驾驶舒适性，对有效降低换挡冲击、规避频繁换挡的动作给予奖励，针对电动车辆，还可将电池SOC状态融入奖励函数，电量不足时优先选取动能回收效率高的挡位，以此方式，强化学习突破传统策略单目标优化的桎梏，推动车辆综合性能全方位提升。

2 基于强化学习的 TCU 换挡策略建模

2.1 状态空间构建

搭建精确完备的状态空间，是运用强化学习构建TCU换挡策略模型的关键起点，状态空间应完整囊括所有左右换挡决策的核心要素，这些要素大体归为车辆动力学参数与外部环境参数两类[2]。就车辆动力学参数而言，发动机转速直观呈现发动机运行状态，其数值高低决定发动机动力输出规模，继而作用于车辆加速效能与行驶动力；变速器输入、输出轴转速对传动比计算意义重大，传动比的适宜程度直接关乎换挡时机抉择，恰当的传动比可使发动机于高效区间运转，增强燃油经济性与动力表现；车速作为车辆行驶状况的直观表征，与发动机转速、变速器传动比相互作用，共同影响车辆动力匹配；油门开度彰显驾驶员动力诉求，开度大小反映驾驶员期望车辆加减速的程度；离合器接合状况影响发动机至变速器的动力传递效率，其接合平稳与否直接关系换挡平顺性及车辆动力响应。

外部环境参数同样至关重要，道路坡度显著影响车辆行驶阻力，上坡时车辆需更大动力克服重力产生的阻力，因而需延后升挡、提前降挡，确保车辆具备充足扭矩爬坡；车辆负载变化会改变动力需求特性，重载时动力需求增大，要求变速器提供更大传动比，输出更强扭矩驱动车辆。对上述参数进行恰当量化与编码，转化为多维状态向量。如此智能体便能借由该状态向量，全面精准感知车辆当前运行状态，为后续科学换挡决策筑牢数据根基。

2.2 奖励函数设计

强化学习智能体探寻最优换挡策略进程里，奖励函数发挥关键指引效能，构建奖励函数需紧扣车辆性能提升诸多核心诉求，达成燃油经济性、驾乘舒适性、动力性能等参数的均衡调配。燃油经济性维度，将发动机工况是否处于燃油消耗高效区间作为奖励赋值关键考量。一旦发动机于经济转速范围平稳运转，且挡位配置适宜，确保发动机维持高效工作态势，便对智能体予以正向激励，借此褒奖此类节能换挡操作；反之若发动机运行状态脱离高效区间，致使燃油消耗攀升，即施以负向激励，推动智能体优化换挡策略。

换挡策略优劣评判中，驾驶舒适性占据重要地位，换挡环节，冲击度充当核心衡量尺度，冲击度过高表明换挡过程颠簸，易给驾乘者带来不良乘坐感受，这种情形下对智能体施以负向奖励；而达成平稳换挡，将冲击度妥善控制在合理范畴内的操作，予以正向奖励，驱动智能体采用更舒适的换挡模式。奖励函数设计时，动力性保障同样是不可忽视的要点，车辆遭遇加速需求，像超车、爬坡等状况，若能迅速降挡，为车辆供给充沛扭矩，让车辆快速响应驾驶者动力要求的换挡行为，便给予正向奖励；一旦出现降挡迟缓，致使车辆加速性能下滑，即刻实施负向奖励，促使智能体在动力需求迫切时作出更及时的换挡抉择。

换挡频次过高不仅损害驾乘舒适性，亦会加剧变速器磨损，缩短其使用周期，针对频繁换挡行为，可给予负向激励，促使智能体在保障车辆性能的基础上，最大限度减少非必要换挡动作。各指标权重经科学调配后，将相关要素融合构建综合奖励函数，以此引导智能体习得兼顾多目标的理想换挡策略。

3 实时控制算法实现与优化

3.1 算法选择与应用

强化学习实现TCU实时控制算法实践中，深度Q网络（DQN）与策略梯度算法是备受青睐的有效路径，深度Q网络（DQN）创新融合深度学习和Q - learning，在应对高维状态空间难题上极具优势[3]。于TCU换挡控制情境内，车辆行驶时发动机转速、车速、油门开度、道路坡度、车辆负载等多元状态信息彼此耦合，塑造出复杂高维状态空间，凭借卷积神经网络（CNN）卓越的特征提取效能，DQN可自主辨识各状态向量中的核心特征，从海量数据里敏锐抓取车辆爬坡时发动机转速与油门开度的内在联系。对状态-动作价值函数Q(s,a)展开学习，智能体借由比对当前状态下不同动作的价值高低，择取价值最优动作作为换挡决策，达成智能化换挡操控。

策略梯度算法另辟蹊径，直接针对策略参数实施优化，实际运用中，此算法可依据各异驾驶工况，迅速调整换挡策略，对于山区道路连续弯道叠加陡坡的复杂工况，策略梯度算法能实时察觉车辆动力匮乏状态，即刻优化策略，完成连续降挡操作，确保车辆爬坡动力充足[4]。车辆处于急加速工况时，该算法同样可快速优化策略，实现

迅速降挡，助力车辆及时获取充足动力，契合驾驶员加速诉求，深度Q网络与策略梯度算法各具优势，实际应用中，可依具体需求灵活搭配。训练初期，借助DQN开展广泛策略探索，促使智能体于大量交互里积累经验，初步构建换挡策略；策略优化阶段，运用策略梯度算法对既有策略精细调校，加速策略收敛进程，令换挡策略更高效精准。

3.2 算法优化与硬件部署

强化学习算法于TCU实现实时控制，需对算法专门优化，并充分考量硬件部署适配性，经验回放机制可有效处理算法优化中的数据相关性难题，智能体与环境交互生成的数据通常存在关联，若直接用于训练，或影响算法学习成效。经验回放机制将数据存于回放缓冲区，训练时随机采样数据，如此既破除数据相关性，又提升数据利用率，增强算法稳定性，针对深度神经网络训练易出现的过拟合状况，可运用正则化技术与Dropout等方式。正则化技术在损失函数中引入正则化项，约束网络参数，避免其过度拟合训练数据；Dropout于训练期间随机舍弃部分神经元，促使网络学习更具鲁棒性的特征表达，提升模型泛化能力，使其在实际应用中更好适应各类驾驶工况。

硬件部署层面，高性能车载计算平台的选型不可或缺，多核处理器与GPU加速模块等设备皆在考量范畴，此类硬件凭借强劲计算性能，可满足强化学习算法实时运算要求，助力算法快速完成复杂计算任务，及时输出换挡决策结果。模型压缩与量化处理同样关键，通过削减模型参数数量、降低参数数据精度，大幅减少计算负荷与内存占用，使算法得以在有限硬件资源中迅速作出决策[5]。而算法与TCU硬件间高效通信接口的构建亦不容忽视，唯有确保控制指令传输及时准确，方能实现换挡策略实时调控，保障车辆于各类工况下灵活调整换挡操作，优化整体性能表现。

结语

本研究聚焦于基于强化学习的TCU换挡策略以及实时控制算法的实现，深入阐释强化学习的理论根基和应用优势，成功完成换挡策略的建模工作，并设计出实时控制算法。构建状态空间并精心设计奖励函数，使得强化学习智能体能够学习并适应复杂工况下的换挡策略。合理地挑选算法并进行优化，同时结合硬件部署来达成实时控制，本研究的成果为提升车辆传动系统的性能提供了全新的思路和途径，然而在算法计算效率方面，以及对极端工况的适应性等方面，仍有进一步深入研究和改进的必要。在未来的研究中，可以探索更为先进的强化学习算法，将其与车联网、自动驾驶技术相结合，从而推动TCU换挡控制朝着更高的智能化水平迈进。

参考文献

[1]张坤.多挡AMT电动汽车智能换挡规律研究[D].聊城大学,2022.

[2]郭国栋,龚雁峰.电力市场环境下基于深度强化学习的微网能量管理系统实时自动控制算法[J].电测与仪表,2021,58(09):78-88.

[3]张嘉祺.基于强化学习的混合动力汽车能量管理策略研究[D].燕山大学,2021.

[4]周楠,陈刚.机器人驾驶车辆深度强化学习换挡策略[J].汽车工程,2020,42(11):1473-1481.

[5]张元侠.基于SVM学习模型的换挡决策研究[D].吉林大学,2019.