**名称**

一种基于强化学习的智能网联汽车协同增强模型及其应用方法

**摘要**

本发明公开一种基于强化学习的智能网联汽车协同增强模型及其应用方法，主要针对智能网联汽车领域下自动驾驶车辆在复杂交通环境中的协同决策问题。该模型通过实时监控和识别交互失效场景，结合强化学习方法优化决策过程，确保车辆在多变的交通环境中实现高效协同决策，并在仿真平台上进行验证测试，显示出良好的决策准确性、响应速度和系统稳定性。最终，该模型可以集成至智能网联汽车控制系统中，提升自动驾驶技术的实用性和可靠性，为高等级自动驾驶汽车的商业化应用提供坚实的技术支持。

**技术领域**

本发明涉及智能网联汽车的协同驾驶技术领域，特别是基于强化学习与博弈论的协同增强决策模型。

**发明内容**

本发明的目的是解决自动驾驶技术在复杂交通环境中遇到的决策瓶颈问题，提出了一种基于强化学习的智能网联汽车协同增强模型及其应用方法。通过实现智能网联汽车的信息互通与协同合作，优化车辆间的协同决策机制，提升车辆在动态复杂交通环境中的应对能力，确保高效、稳定的协同驾驶。通过结合博弈论中的Level-k模型与强化学习算法，本发明有效优化了自动驾驶系统的协同决策过程，提升了系统的响应速度、决策准确性及系统稳定性。

本发明的目的可以通过以下技术方案来实现：

一种基于强化学习的智能网联汽车协同增强模型及其应用方法，包括以下步骤：

S1交互机理研究与失效场景识别：首先，分析智能网联汽车与人类驾驶汽车之间的交互机制。通过大量路测数据的收集与分析，识别出自动驾驶车辆在与人类驾驶车辆交互时的失效场景，并通过支持向量机（SVM）算法构建分类模型，识别与监控这些交互状态。该步骤能够为后续的协同决策模型提供数据支持和理论依据。

S2基于博弈论的Level-k模型设计：根据交互失效场景的分析结果，构建基于博弈论的Level-k模型。此模型用于描述智能网联车辆与其他交通参与者之间的协同决策过程，其中通过定义不同层级（k值）的博弈行为，确定智能网联车辆的最佳决策策略。Level-k模型将为车辆在复杂交通场景中如何选择最优决策提供理论框架。

S3结合强化学习优化协同决策过程：在本步骤中，采用强化学习算法来选择Level-k模型中的k值。强化学习通过与环境的交互不断更新k值的选择策略，使得车辆能够根据交通环境中的动态变化，灵活选择不同的决策层级（k值）。通过强化学习的训练，系统能够在多变的交通环境中调整k值，从而使得车辆能根据当前交通状况做出最合适的决策动作。

S4结合Level-k模型与强化学习进行决策动作控制：在选择合适的k值后，通过Level-k模型进行具体的决策动作控制。此时，车辆根据已选择的k值，确定与其他交通参与者的协同行为，并根据博弈论框架执行相应的决策动作。该过程确保智能网联汽车能够在复杂的交通情境中迅速做出高效的决策，并与其他交通参与者协同合作，确保行车安全和交通流畅。

S5集成优化模型并进行实际应用：将优化后的协同增强模型集成至智能网联汽车的控制系统中，在实际交通环境中进行应用。通过实时监控与协同决策，提高自动驾驶车辆在复杂交通环境中的稳定性、可靠性及适应性，从而推动高等级自动驾驶汽车的商业化应用。

与现有技术相比，本发明有以下优点：

（1）提升协同决策效率：本发明通过结合博弈论的Level-k模型与强化学习算法，优化了车辆间的协同决策过程。通过强化学习选择合适的k值，使得智能网联汽车能够在复杂交通场景下灵活调整决策层级，从而提高决策效率和响应速度。

（2）增强自动驾驶系统的可靠性：通过实时识别与监控智能网联汽车与人类驾驶车辆的交互失效场景，本发明能够有效避免自动驾驶系统在复杂场景中的失效问题，增强了系统的稳定性和可靠性，减少了潜在的交通安全风险。

（3）提升模型的适应性与实用性：本发明在多种复杂交通环境下进行了仿真测试，并根据实验反馈优化决策模型，确保其在实际交通环境中的可行性和实用性。优化后的模型集成到车辆控制系统中，能够更好地应对现实世界中的各种交通情况，推动高等级自动驾驶技术的商业化应用。

（4）增强智能网联汽车与人类驾驶汽车的协同能力：本发明通过研究自动驾驶车辆与人类驾驶车辆的交互机制，建立了智能网联汽车与人类驾驶汽车的协同决策模型。这种协同机制能够提高自动驾驶车辆在与人类驾驶车辆共同参与交通时的协作能力，改善整体交通流畅性和安全性。

附图说明

图1为本发明应用基于强化学习的智能网联汽车协同增强模型的流程图;

图2为无保护左转情景决策时域图；

图3为PPO算法训练PPO-OK模型过程；

具体实施方式：

下面结合图1对本发明作进一步的详细说明。本实施方式所述的一种交叉口事故影响下的基于强化学习的智能网联车辆协同增强方法，所述方法具体包括以下步骤：

步骤一、使用路测数据集与仿真环境训练PPO-OK(PPO-output k)模型。该模型将Level-k与强化学习PPO算法结合，优化协同决策过程，采用选择Level-k模型中的k值。

该强化学习优化的目标函数为：

其中，表示策略网络的参数，S为状态空间，A为动作空间。为在S状态下执行动作A获得的预期回报。

在PPO算法中，当前策略与旧策略之间的比率通过重要性采样进行计算，以衡量当前策略在特定状态下选择某一动作的相对重要性。基于此比率，PPO算法评估当前策略更新与旧策略的差异，并据此调整策略。为了确保训练稳定性，算法还采用剪切操作（clip）限制比率的最大值和最小值，防止策略更新过大，避免训练过程中出现剧烈波动。当比率变化超出预设阈值（通常为）时，进行裁剪，保持更新幅度平稳、渐进。

此机制有效避免过大波动，确保策略的稳定收敛，从而训练出高效的 PPO-OK 模型。

步骤二、智能网联车辆通过车载和环境感知设备，例如摄像头、雷达、激光雷达等，获取交叉口的环境信息和车辆状态信息。

交叉口环境信息具体包括交叉口的出入口位置、出入口车道数、车道宽度及出入口形状尺寸等。

车辆状态信息包括周围车辆的位置、速度、智能网联汽车的行驶目的等

步骤三、基于自动驾驶汽车与人类驾驶汽车交互的argo开源数据集，利用支持向量机（SVM）训练得到失效场景分析模型，其原理如下：

该模型的基础为“同理心”原则：在决策过程中，车辆会考虑其决策对车辆的影响，并推测车辆会如何回应，车辆的决策不仅基于自身的运动状态，还要考虑对方的反应和可能的动作。本研究用协作加速度来量化车辆间的互动程度，并预测各自的意图，对车辆行为进行建模。

车辆视角下的协作加速度为：

其中，是车辆到达冲突点的距离；是车辆的速度；是车辆到达冲突点的时间。

车辆视角下的车辆加速度为：

其中，是车辆到达冲突点的距离；是车辆的速度；是车辆到达冲突点的时间。

基于协同加速度模型，车辆能够当前运动状态和对方运动状态做出决策。引入支持向量机（SVM）处理非线性分类问题，以区分不同的决策类型。通过对数据集的学习，SVM找到最优超平面将不同的决策类别分开。

SVM的输入特征向量为：

X = [ ， ，]

其中，是车辆自身到达冲突点的时间；是对方车辆到达冲突点的时间；是对方车辆的协作加速度。

输出标签Y定义为：

Y = { 0,预测对方先行 1,预测对方让行}

SVM分类器的决策函数为：

+++b=0

在交互过程中，车辆根据当前的运动状态评估对方的优先权。将分界面在TTCP时间域中进行映射，变形得到公式：

引入图1“决策时域图”以进一步描述车辆在不同时间点的决策过程，将交互过程分为不同区域。以无保护左转情景为例：区域A与区域B表示均有一方车辆让行，单车智能有效；区域D表示两方车辆均让行，陷入死锁；区域C与区域E表示两方车辆均未选择让行，存在碰风险。若车辆状态点位于区域C、D、E，单车智能失效，需要进行车辆协同增强

基于车辆运动状态与周围环境信息，利用失效场景分析模型进行分析，识别交互失效状态，及时开启协同增强以保证智能网联汽车交互过程的高效、安全。

若车辆交互处于“单车智能失效”状态，则执行步骤四；

否则，继续保持单车智能驾驶状态；

步骤四、使用强化学习训练得到的PPO-OK模型，将车辆状态信息作为输入，输入为环境中各车的特征向量，特征向量包含；输出为各智能网联汽车k值的排列。

步骤五、利用level-k动作执行模型根据最优k值排列，选择不同车辆的决策动作。第k级车辆的决策公式为：

在时间步,车辆依据k级博弈推理所做出的最优决策可以表示为以下集合：

其中分别表示车辆𝑖和其他车辆在时间步时的状态。，表示车辆𝑖和其他车辆在当前时间步上的动作。奖励函数R依赖于当前状态、动作以及其他车辆依据k-1级推理所采取的行为。折扣因子用于调整奖励的时间衰减，使得较近的未来决策权重更高。

对于任何车辆，通过顺序迭代的计算过程，在每一轮迭代中，车辆会根据其他车辆在上一轮迭代中的决策，即。通过这种逐步迭代，车辆可以逐渐逼近更高层次的博弈均衡。对于每个，通过这种顺序迭代的计算，最终获得 级推理下的最优决策，其中是预定的最大推理层级。

通过level-k的迭代过程最终将不同车辆最优K值排列转化为不同车辆的最优决策动作，在复杂的交通环境中实现高效、安全决策。

步骤六、根据得到的动作决策更新各智能网联汽车位置、速度等车辆行驶状态，重新获得各车辆的行驶状态信息，判断是否通过冲突点：

若通过冲突点，则退出协同增强，恢复单车智能驾驶状态；

否则，继续步骤四；

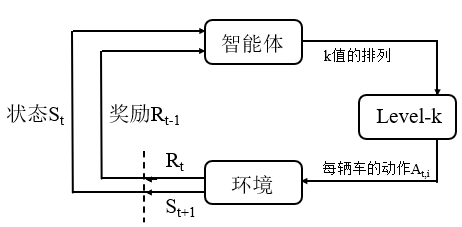


图 3

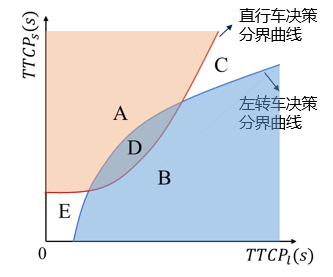


图 2

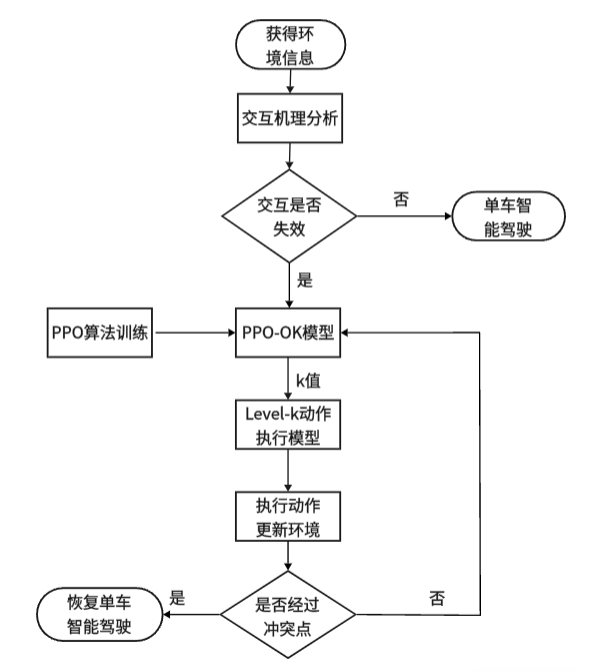


图 1