



(12)发明专利申请

(10)申请公布号 CN 110562258 A

(43)申请公布日 2019.12.13

(21)申请号 201910938698.4

(22)申请日 2019.09.30

(71)申请人 驭势科技(北京)有限公司

地址 102400 北京市房山区弘安路85号2号楼401室

(72)发明人 徐鲲鹏 马万里 周小成 姜岩

(74)专利代理机构 北京开阳星知识产权代理有限公司 11710

代理人 郭鑫

(51) Int. Cl.

B60W 30/18(2012.01)

B60W 40/00(2006.01)

B60W 40/02(2006.01)

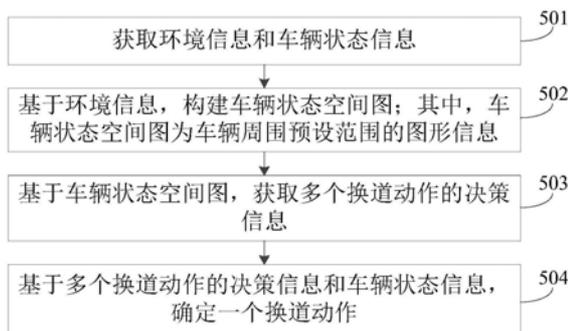
权利要求书2页 说明书15页 附图4页

(54)发明名称

一种车辆自动换道决策的方法、车载设备和存储介质

(57)摘要

本公开实施例涉及一种车辆自动换道决策的方法、车载设备和存储介质,方法包括:获取环境信息和车辆状态信息;基于环境信息,构建车辆状态空间图;其中,车辆状态空间图为车辆周围预设范围的图形信息;基于车辆状态空间图,获取多个换道动作的决策信息;基于多个换道动作的决策信息和车辆状态信息,确定一个换道动作。本公开实施例中,通过基于环境信息构建车辆状态空间图,进而基于车辆状态空间图决策换道动作,考虑了环境的不确定性,保留车辆周围环境信息,没有丢失整体的统计信息,可提高决策的有效性。



1. 一种车辆自动换道决策的方法,其特征在于,所述方法包括:  
获取环境信息和车辆状态信息;  
基于所述环境信息,构建车辆状态空间图;所述车辆状态空间图为车辆周围预设范围的图形信息;  
基于所述车辆状态空间图,获取多个换道动作的决策信息;  
基于所述多个换道动作的决策信息和所述车辆状态信息,确定一个换道动作。
2. 根据权利要求1所述的方法,其特征在于,所述环境信息包括:车辆周围环境的信息;所述车辆状态信息包括:当前车速和当前所在车道。
3. 根据权利要求1所述的方法,其特征在于,所述车辆状态空间图包括:  
所述车辆、所述车辆周围预设范围内的车道以及所述车辆周围预设范围内的障碍物。
4. 根据权利要求3所述的方法,其特征在于,所述车辆状态空间图包括:  
栅格式状态空间图;其中,所述栅格式状态空间图中每个栅格对应预设尺寸的道路区域;  
所述车辆、车道和障碍物均由栅格表示。
5. 根据权利要求1所述的方法,其特征在于,所述车辆周围预设范围,包括:  
车辆前向预设距离、车辆后向预设距离以及车辆侧向预设距离。
6. 根据权利要求1所述的方法,其特征在于,基于所述环境信息,构建车辆状态空间图,包括:  
基于所述环境信息,确定所述车辆周围预设范围内的车道以及所述车辆周围预设范围内的障碍物;  
基于所述车辆周围预设范围内的车道以及所述车辆周围预设范围内的障碍物,构建栅格式状态空间图。
7. 根据权利要求1所述的方法,其特征在于,基于所述车辆状态空间图,获取多个换道动作的决策信息,包括:  
基于所述车辆状态空间图和深度强化学习模型,获取多个换道动作的决策信息。
8. 根据权利要求7所述的方法,其特征在于,所述深度强化学习模型为DQN模型;所述决策信息为所述深度强化学习模型的值函数的Q值。
9. 根据权利要求7所述的方法,其特征在于,所述深度强化学习模型训练过程中的奖励值设置如下:  
若车辆变速行驶,则奖励值基于当前车速和当前所在车道的最小车速确定;  
若车辆直线行驶,则奖励值为预设正值;  
若车辆发送碰撞,则奖励值为预设负值。
10. 根据权利要求7所述的方法,其特征在于,所述深度强化学习模型的记忆回放采样包括:  
确定记忆池中每个样本的贡献值;  
基于每个样本的贡献值,确定每个样本被采样的概率;  
基于每个样本被采样的概率,对记忆池中的所有样本进行采样。
11. 根据权利要求1所述的方法,其特征在于,基于所述多个换道动作的决策信息和所述车辆状态信息,确定一个换道动作,包括:

基于预先设置的车辆行驶约束条件和所述车辆状态信息,确定至少一个有效换道动作;

基于所述多个换道动作的决策信息和所述至少一个有效换道动作,确定一个换道动作。

12. 根据权利要求11所述的方法,其特征在于,基于所述多个换道动作的决策信息和所述至少一个有效换道动作,确定一个换道动作,包括:

从所述至少一个有效换道动作中确定决策值最大的换道动作;或,从所述至少一个有效换道动作中确定满足安全条件且决策值最大的换道动作。

13. 根据权利要求12所述的方法,其特征在于,所述安全条件为:执行换道动作后所述车辆与前车的距离大于或等于安全距离。

14. 根据权利要求13所述的方法,其特征在于,所述安全距离基于所述车辆的车速确定。

15. 一种车载设备,其特征在于,包括:处理器和存储器;

所述处理器通过调用所述存储器存储的程序或指令,用于执行如权利要求1至14任一项所述方法的步骤。

16. 一种非暂态计算机可读存储介质,其特征在于,所述非暂态计算机可读存储介质存储程序或指令,所述程序或指令使计算机执行如权利要求1至14任一项所述方法的步骤。

## 一种车辆自动换道决策的方法、车载设备和存储介质

### 技术领域

[0001] 本公开实施例涉及智能驾驶技术领域,具体涉及一种车辆自动换道决策的方法、车载设备和存储介质。

### 背景技术

[0002] 随着智能驾驶技术的发展,使车辆具备自主换道的功能,车辆在行驶过程中,为了追求更合理的车速和更自由的行驶空间而自主进行车道变换。与跟车行为相比,换道行为需要考虑车辆周围更加复杂的环境信息,例如周围车辆和道路状况。

### 发明内容

[0003] 为了解决现有技术存在的至少一个问题,本公开的至少一个实施例提供了一种车辆自动换道决策的方法、车载设备和存储介质。

[0004] 第一方面,本公开实施例提出一种车辆自动换道决策的方法,所述方法包括:

[0005] 获取环境信息和车辆状态信息;

[0006] 基于所述环境信息,构建车辆状态空间图;所述车辆状态空间图为车辆周围预设范围的图形信息;

[0007] 基于所述车辆状态空间图,获取多个换道动作的决策信息;

[0008] 基于所述多个换道动作的决策信息和所述车辆状态信息,确定一个换道动作。

[0009] 第二方面,本公开实施例还提出一种车载设备,包括:处理器和存储器;处理器通过调用存储器存储的程序或指令,用于执行如第一方面所述方法的步骤。

[0010] 第三方面,本公开实施例还提出一种非暂态计算机可读存储介质,用于存储程序或指令,所述程序或指令使计算机执行如第一方面所述方法的步骤。

[0011] 可见,本公开实施例的至少一个实施例中,通过基于环境信息构建车辆状态空间图,进而基于车辆状态空间图决策换道动作,考虑了环境的不确定性,保留车辆周围环境信息,没有丢失整体的统计信息,可提高决策的有效性。

### 附图说明

[0012] 为了更清楚地说明本公开实施例的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本公开的一些实施例,对于本领域普通技术人员来讲,还可以根据这些附图获得其他的附图。

[0013] 图1是本公开实施例提供的一种智能驾驶车辆的整体架构图;

[0014] 图2是本公开实施例提供的一种智能驾驶系统的框图;

[0015] 图3是本公开实施例提供的一种换道决策模块的框图;

[0016] 图4是本公开实施例提供的一种车载设备的框图;

[0017] 图5是本公开实施例提供的一种车辆自动换道决策的方法流程图;

[0018] 图6是本公开实施例提供的一种车辆周围环境的示意图;

[0019] 图7是基于图6所示的车辆周围环境构建车辆状态空间图的示意图；

[0020] 图8是本公开实施例提供的一种深度强化学习模型示意图。

### 具体实施方式

[0021] 为了能够更清楚地理解本公开的上述目的、特征和优点，下面结合附图和实施例对本公开作进一步的详细说明。可以理解的是，所描述的实施例是本公开的一部分实施例，而不是全部的实施例。此处所描述的具体实施例仅仅用于解释本公开，而非对本公开的限定。基于所描述的本公开的实施例，本领域普通技术人员所获得的所有其他实施例，都属于本公开保护的范畴。

[0022] 需要说明的是，在本文中，诸如“第一”和“第二”等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来，而不一定要求或者暗示这些实体或操作之间存在任何这种实际的关系或者顺序。

[0023] 车辆自动换道决策的方法可分为基于规则(rule-based)的方法和基于统计(statistic-based)的方法。基于规则的方法主要由换道动机和碰撞检测两部分组成。在本车的速度或车头视距低于期望值后，若检测到目标车道具有速度优势或空间优势，则产生换道动机。在产生换道动机后，碰撞检测主要评价目标车道前后是否有足够的间隙允许安全地执行换道行为，需要计算自主性换道能接受的最小间距和碰撞时间从而做出换道决策。

[0024] 基于统计的方法主要包括几种典型的机器学习算法，例如决策树、粗糙集和最近邻模型等，是一种考虑不确定性的决策方法。首先，从数据集中提取车辆行驶过程中影响换道决策的因素作为输入特征，例如本车和周围车辆的位置、速度等状态数据，并在预处理后划分成训练样本和测试样本。然后，选取合理的决策模型和参数阈值，基于训练样本对决策模型进行训练和优化。最后，采用训练后的决策模型对测试样本进行分类，将分类结果与实际的情况进行对比，验证决策模型的精确性。

[0025] 但是，基于规则的方法约束条件众多且规则复杂，无法适用于多种场景。此外，基于规则的方法仅考虑了本车的状态和周围障碍物的状态，没有将道路的几何结构作为决策影响因素，没有考虑环境的不确定性，丢失了整体的统计信息，影响决策有效性。基于统计的方法是一种考虑不确定性的决策方法。

[0026] 与基于规则的方法相比，基于统计的方法需要大量的换道数据样本用于模型训练和测试，然而不同场景下产生的换道动机有很大的差异，很难收集到全面并且完整的换道数据。在决策架构的设计、交通参与者的行为分析、决策模型求解等问题中均面临挑战。而且算法实现的过程较繁琐，数据处理、模型选择和模型训练在时间和空间上都需要占有很多的资源。最重要的是，基于统计的换道模型都具有一定的假阳性率，即在需要车道保持时做出换道决策，显著影响了行车安全。

[0027] 针对现有技术中存在的多个问题，本公开实施例提供一种车辆自动换道决策的方案，基于环境信息构建车辆状态空间图，进而基于车辆状态空间图决策换道动作。

[0028] 相比基于规则的方法约束条件众多且规则复杂，无法适用于多种场景，本方案无需设计复杂的规则及约束条件，避免人为设定规则的缺漏，并可适用于多种场景。另外，相比基于规则的方法没有将道路的几何结构作为决策影响因素，没有考虑环境的不确定性，

丢失了整体的统计信息,影响决策有效性,本方案考虑了环境的不确定性,基于环境信息构建了车辆状态空间图,因此没有丢失整体的统计信息,可提高决策有效性。

[0029] 本公开实施例提供的车辆自动换道决策的方案,可应用于智能驾驶车辆。

[0030] 图1为本公开实施例提供的一种智能驾驶车辆的整体架构图。如图1所示,智能驾驶车辆包括:传感器组、智能驾驶系统100、车辆底层执行系统以及其他可用于驱动车辆和控制车辆运行的部件。

[0031] 传感器组,用于采集车辆外界环境的数据和探测车辆的位置数据。传感器组例如包括但不限于摄像头、激光雷达、毫米波雷达、超声波雷达、GPS(Global Positioning System,全球定位系统)和IMU(Inertial Measurement Unit,惯性测量单元)中的至少一个。

[0032] 在一些实施例中,传感器组,还用于采集车辆的动力学数据,传感器组例如还包括但不限于车轮转速传感器、速度传感器、加速度传感器、方向盘转角传感器、前轮转角传感器中的至少一个。

[0033] 智能驾驶系统100,用于获取传感器组的数据,传感器组中所有传感器在智能驾驶车辆行驶过程中都以较高的频率传送数据。

[0034] 智能驾驶系统100,还用于基于传感器组的数据进行环境感知和车辆定位,并基于环境感知信息和车辆定位信息进行路径规划和决策,以及基于规划的路径生成车辆控制指令,从而控制车辆按照规划路径行驶。

[0035] 在一些实施例中,智能驾驶系统100,还用于进行自动换道决策。在一些实施例中,智能驾驶系统100获取环境信息和车辆状态信息,基于环境信息,构建车辆状态空间图,构建的车辆状态空间图为车辆周围预设范围的图形信息,因此保留了车辆周围环境信息,没有丢失整体的统计信息。进而智能驾驶系统100基于车辆状态空间图,获取多个换道动作的决策信息,并基于多个换道动作的决策信息和车辆状态信息,确定一个换道动作,可提高决策有效性。

[0036] 在一些实施例中,智能驾驶系统100可以为软件系统、硬件系统或者软硬件结合的系统。例如,智能驾驶系统100是运行在操作系统上的软件系统,车载硬件系统是支持操作系统运行的硬件系统。

[0037] 在一些实施例中,智能驾驶系统100,还用于与云端服务器无线通信,交互各种信息。在一些实施例中,智能驾驶系统100与云端服务器通过无线通讯网络(例如包括但不限于GPRS网络、Zigbee网络、Wifi网络、3G网络、4G网络、5G网络等无线通讯网络)进行无线通信。

[0038] 在一些实施例中,云端服务器用于统筹协调管理智能驾驶车辆。在一些实施例中,云端服务器可以用于与一个或多个智能驾驶车辆进行交互,统筹协调管理多个智能驾驶车辆的调度等。

[0039] 在一些实施例中,云端服务器是由车辆服务商所建立的云端服务器,提供云存储和云计算的功能。在一些实施例中,云端服务器中建立车辆端档案。在一些实施例中,车辆端档案中储存智能驾驶系统100上传的各种信息。在一些实施例中,云端服务器可以实时同步车辆端产生的驾驶数据。

[0040] 在一些实施例中,云端服务器可以是一个服务器,也可以是一个服务器群组。服务

器群组可以是集中式的,也可以是分布式的。分布式服务器,有利于任务在多个分布式服务器进行分配与优化,克服传统集中式服务器资源紧张与响应瓶颈的缺陷。在一些实施例中,云端服务器可以是本地的或远程的。

[0041] 在一些实施例中,云端服务器可用于对车辆端进行停车收费、过路收费等。在一些实施例中,云端服务器还用于分析驾驶员的驾驶行为,并且对驾驶员的驾驶行为进行安全等级评估。

[0042] 在一些实施例中,云端服务器可用于获取道路监测单元(RSU:Road Side Unit)和智能驾驶车辆的信息,以及可以发送信息至智能驾驶车辆。在一些实施例中,云端服务器可以根据智能驾驶车辆的信息将道路监测单元中的与智能驾驶车辆相对应的检测信息发送给智能驾驶车辆。

[0043] 在一些实施例中,道路监测单元可以用于收集道路监测信息。在一些实施例中,道路监测单元可以是环境感知传感器,例如,摄像头、激光雷达等,也可以是道路设备,例如V2X设备,路边红绿灯装置等。在一些实施例中,道路监测单元可以监控隶属于相应道路监测单元的道路情况,例如,通过车辆的类型、速度、优先级别等。道路监测单元在收集到道路监测信息后,可将所述道路监测信息发送给云端服务器,也可以发送给通过道路的智能驾驶车辆。

[0044] 车辆底层执行系统,用于接收车辆控制指令,实现对车辆行驶的控制。在一些实施例中,车辆底层执行系统包括但不限于:转向系统、制动系统和驱动系统。转向系统、制动系统和驱动系统属于车辆领域成熟系统,在此不再赘述。

[0045] 在一些实施例中,智能驾驶车辆还可包括图1中未示出的车辆CAN总线,车辆CAN总线连接车辆底层执行系统。智能驾驶系统100与车辆底层执行系统之间的信息交互通过车辆CAN总线进行传递。

[0046] 在一些实施例中,智能驾驶车辆既可以通过驾驶员又可以通过智能驾驶系统100控制车辆行驶。在人工驾驶模式下,驾驶员通过操作控制车辆行驶的装置驾驶车辆,控制车辆行驶的装置例如包括但不限于制动踏板、方向盘和油门踏板等。控制车辆行驶的装置可直接操作车辆底层执行系统控制车辆行驶。

[0047] 在一些实施例中,智能驾驶车辆也可以为无人车,车辆的驾驶控制由智能驾驶系统100来执行。

[0048] 图2为本公开实施例提供的一种智能驾驶系统200的框图。在一些实施例中,智能驾驶系统200可以实现为图1中的智能驾驶系统100或者智能驾驶系统100的一部分,用于控制车辆行驶。

[0049] 如图2所示,智能驾驶系统200可划分为多个模块,例如可包括:感知模块201、规划模块202、控制模块203、换道决策模块204以及其他一些可用于智能驾驶的模块。

[0050] 感知模块201用于进行环境感知与定位。在一些实施例中,感知模块201用于获取传感器数据、V2X (Vehicle to X,车用无线通信)数据、高精度地图等数据。在一些实施例中,感知模块201用于基于获取的传感器数据、V2X (Vehicle to X,车用无线通信)数据、高精度地图等数据中的至少一种,进行环境感知与定位。

[0051] 在一些实施例中,感知模块201用于生成感知定位信息,实现对障碍物感知、摄像头图像的可行驶区域识别以及车辆的定位等。

[0052] 环境感知 (Environmental Perception) 可以理解为对于环境的场景理解能力,例如障碍物的位置,道路标志/标记的检测,行人/车辆的检测等数据的语义分类。在一些实施例中,环境感知可采用融合摄像头、激光雷达、毫米波雷达等多种传感器的数据进行环境感知。

[0053] 定位 (Localization) 属于感知的一部分,是确定智能驾驶车辆相对于环境的位置的能力。定位可采用:GPS定位,GPS的定位精度在数十米到厘米级别,定位精度高;定位还可采用融合GPS和惯性导航系统 (Inertial Navigation System) 的定位方法。定位还可采用SLAM (Simultaneous Localization And Mapping,同步定位与地图构建),SLAM的目标即构建地图的同时使用该地图进行定位,SLAM通过利用已经观测到的环境特征确定当前车辆的位置以及当前观测特征的位置。

[0054] V2X是智能交通运输系统的关键技术,使得车与车、车与基站、基站与基站之间能够通信,从而获得实时路况、道路信息、行人信息等一系列交通信息,提高智能驾驶安全性、减少拥堵、提高交通效率、提供车载娱乐信息等。

[0055] 高精度地图是智能驾驶领域中使用的地理地图,与传统地图相比,不同之处在于:1) 高精度地图包括大量的驾驶辅助信息,例如依托道路网的精确三维表征:包括交叉路口和路标位置等;2) 高精度地图还包括大量的语义信息,例如报告交通灯上不同颜色的含义,又例如指示道路的速度限制,以及左转车道开始的位置;3) 高精度地图能达到厘米级的精度,确保智能驾驶车辆的安全行驶。

[0056] 规划模块202用于基于感知定位模块生成的感知定位信息,进行路径规划和决策。

[0057] 在一些实施例中,规划模块202用于基于感知定位模块生成的感知定位信息,并结合V2X数据、高精度地图等数据中的至少一种,进行路径规划和决策。

[0058] 在一些实施例中,规划模块202用于规划路径,决策:行为(例如包括但不限于跟车、超车、停车、绕行等)、车辆航向、车辆速度、车辆的期望加速度、期望的方向盘转角等,生成规划决策信息。

[0059] 控制模块203用于基于规划模块202生成的规划决策信息,进行路径跟踪和轨迹跟踪。

[0060] 在一些实施例中,控制模块203用于生成车辆底层执行系统的控制指令,并下发控制指令,以使车辆底层执行系统控制车辆按照期望路径行驶,例如通过控制方向盘、刹车以及油门对车辆进行横向和纵向控制。

[0061] 在一些实施例中,控制模块203还用于基于路径跟踪算法计算前轮转角。

[0062] 在一些实施例中,路径跟踪过程中的期望路径曲线与时间参数无关,跟踪控制时,可以假设智能驾驶车辆以当前速度匀速前进,以一定的代价规则使行驶路径趋近于期望路径;而轨迹跟踪时,期望路径曲线与时间和空间均相关,并要求智能驾驶车辆在规定的时间内到达某一预设好的参考路径点。

[0063] 路径跟踪不同于轨迹跟踪,不受制于时间约束,只需要在一定误差范围内跟踪期望路径。

[0064] 换道决策模块204用于获取环境信息和车辆状态信息。在一些实施例中,环境信息包括但不限于车辆传感器采集的车辆周围环境的的信息。车辆传感器可以为图1所示的传感器组。车辆状态信息包括但不限于当前车速和当前所在车道。进而换道决策模块204基于环

境信息,构建车辆状态空间图,构建的车辆状态空间图为车辆周围预设范围的图形信息,因此保留了车辆周围环境信息,没有丢失整体的统计信息,可提高决策有效性。更进一步地,换道决策模块204基于车辆状态空间图,获取多个换道动作的决策信息,并基于多个换道动作的决策信息和车辆状态信息,确定一个换道动作。

[0065] 在一些实施例中,换道决策模块204的功能可集成到感知模块201、规划模块202或控制模块203中,也可配置为与智能驾驶系统200相独立的模块,换道决策模块204可以为软件模块、硬件模块或者软硬件结合的模块。例如,换道决策模块204是运行在操作系统上的软件模块,车载硬件系统是支持操作系统运行的硬件系统。

[0066] 图3为本公开实施例提供的一种换道决策模块300的框图。在一些实施例中,换道决策模块300可以实现为图2中的换道决策模块204或者换道决策模块204的一部分。

[0067] 如图3所示,换道决策模块300可包括但不限于以下单元:第一获取单元301、构建单元302、第二获取单元303和确定单元304。

[0068] 在一些实施例中,换道决策模块300可预先设置车辆换道动作空间和车辆行驶约束条件,便于换道决策模块300在车辆行驶约束条件的约束下,从车辆换道动作空间中决策一个换道动作。

[0069] 在一些实施例中,车辆换道动作空间包括多个换道动作,也可以称为车辆换道动作集合。在一些实施例中,车辆换道动作空间包括5个换道动作:

[0070] (1) N(No-op):无操作。不采取任何行为,保持当前状态。

[0071] (2) A(Accelerate):以恒定的速度加速。

[0072] (3) D(Decelerate):以恒定的速度减速。

[0073] (4) L(Left):向左变道。

[0074] (5) R(Right):向右变道。

[0075] 在一些实施例中,车辆行驶约束条件为车辆行驶过程中应当遵守的规则。在一些实施例中,车辆行驶约束条件包括但不限于以下一个或多个:

[0076] (1) 若车辆处于最左侧车道时,不作出向左变道决策。

[0077] (2) 若车辆处于最右侧车道时,不作出向右变道决策。

[0078] (3) 若行驶道路存在最高限速,且车辆速度达到最高限速临界点时,不作出加速决策。

[0079] (4) 若行驶道路存在最低限速,且车辆速度达到最低限速临界点时,不作出减速决策。

[0080] (5) 若即将经过斑马线时,不作出加速决策。

[0081] (6) 若检测到红灯或黄灯时,强制作出减速决策。

[0082] 第一获取单元301,用于获取环境信息和车辆状态信息。在一些实施例中,环境信息包括但不限于车辆周围环境的信息;车辆状态信息包括但不限于当前车速和当前所在车道。在一些实施例中,车辆周围环境的信息可以理解为图2中感知模块201基于传感器组采集的数据生成的感知定位信息。

[0083] 构建单元302,用于基于环境信息,构建车辆状态空间图;其中,车辆状态空间图为车辆周围预设范围的图形信息。在一些实施例中,由于车辆状态空间图为图形信息,因此,构建单元302基于摄像头、激光雷达等传感器采集的可视的环境信息,也称为外部状态可视

信息,来构建车辆状态空间图。

[0084] 在一些实施例中,车辆状态空间图包括:车辆(也即本车)、车辆周围预设范围内的车道以及车辆周围预设范围内的障碍物。在一些实施例中,车辆周围预设范围,包括:车辆前向预设距离、车辆后向预设距离以及车辆侧向预设距离。车辆侧向预设距离包括车辆左侧预设距离和车辆右侧预设距离。本实施例并非基于传感器可观测范围来构建车辆状态空间图,而是基于预设范围来构建车辆状态空间图,预设范围比观测范围小。

[0085] 在一些实施例中,车辆前向预设距离可为前向距离范围中的任一值,例如前向距离范围为5米至200米;车辆后向预设距离可为后向距离范围中的任一值,例如后向距离范围为2米至50米;车辆侧向预设距离为侧向距离范围中的任一值,例如侧向距离范围为2至30米。需要说明的是,本领域技术人员可根据实际需要设置不同距离范围,并确定不同预设距离的取值,本实施例不限定预设距离的具体取值。

[0086] 在一些实施例中,车辆前向预设距离为20米、车辆后向预设距离为4米以及车辆侧向预设距离为6米,也即车辆左侧预设距离为6米和车辆右侧预设距离为6米。

[0087] 在一些实施例中,车辆前向预设距离、车辆后向预设距离以及车辆侧向预设距离可基于车辆状态信息确定,例如基于车速确定不同预设距离,车速较快时,预设距离的取值较大,车速较慢时,预设距离的取值较小。需要说明的是,本领域技术人员可根据实际需要设置车辆状态信息与不同预设距离之间的对应关系,本实施例不限定车辆状态信息与不同预设距离之间的具体对应关系。

[0088] 在一些实施例中,车辆状态空间图包括栅格式状态空间图;其中,栅格式状态空间图中每个栅格对应预设尺寸的道路区域;且车辆、车道和障碍物均由栅格表示。

[0089] 在一些实施例中,每个栅格大小可为预设道路尺寸范围中的任一值,例如预设道路尺寸范围为1米×1米至10米×10米;车辆可为预设车辆栅格范围中的任一值,例如车辆栅格范围为1×2个栅格至10×10个栅格;车道的宽度可为预设宽度栅格范围中的任一值,例如宽度栅格范围为1个栅格至10个栅格;车辆前向预设距离可为预设前向栅格范围中的任一值,例如前向栅格范围为2个栅格至100个栅格;车辆后向预设距离可为预设后向栅格范围中的任一值,例如后向栅格范围为1个栅格至10个栅格;车辆侧向预设距离可为预设侧向栅格范围中的任一值,例如侧向栅格范围为2个栅格至30个栅格。需要说明的是,本领域技术人员可根据实际需要设置不同范围,本实施例不同范围的具体取值。

[0090] 在一些实施例中,每个栅格大小对应1米×1米的道路区域,车辆对应3×6个栅格,车道的宽度对应5个栅格,车辆前向预设距离对应20个栅格,车辆后向预设距离对应4个栅格,车辆左侧预设距离对应6个栅格,车辆右侧预设距离对应6个栅格。

[0091] 在一些实施例中,构建单元302用于基于环境信息,确定车辆周围预设范围内的车道以及车辆周围预设范围内的障碍物。进而构建单元302基于车辆周围预设范围内的车道以及车辆周围预设范围内的障碍物,构建栅格式状态空间图。

[0092] 图6是本公开实施例提供的一种车辆周围环境的示意图,图7是基于图6所示的车辆周围环境构建的车辆状态空间图。如图7所示,车辆状态空间图为栅格式状态空间图,包括3个车道,本车位于3个车道的中间车道,由于本车前方的障碍物(车辆)部分已超出车辆前向预设距离,因此,该障碍物只有3×2个栅格的区域包含在车辆状态空间图中,可见,车辆状态空间图保留了包括车道在内的完整的环境信息,没有丢失整体的统计信息,可提高

决策有效性。

[0093] 第二获取单元303,用于基于车辆状态空间图,获取多个换道动作的决策信息。在一些实施例中,第二获取单元303,用于基于车辆状态空间图和深度强化学习模型,获取多个换道动作的决策信息。在一些实施例中,深度强化学习模型的输入是车辆状态空间图,输出是多个换道动作的决策信息,由于预先设置了车辆换道动作空间,因此,所述多个换道动作作为车辆换道动作空间中的换道动作。相比基于统计的方法需要大量的换道数据样本用于模型训练和测试,而不同场景下产生的换道动机有很大的差异,很难收集到全面并且完整的换道数据的问题。本公开实施例通过强化学习考虑了车辆与环境的交互过程,将环境因素考虑其中;不需要历史数据,而是通过强化学习让车辆自主学习训练。

[0094] 在一些实施例中,第二获取单元303使用的深度强化学习模型为DQN(Deep Q Network)模型。在一些实施例中,还可以使用其他的深度强化学习模型,例DDPG(Deep Deterministic Policy Gradient)。决策信息为深度强化学习模型的值函数(Value Function)的Q值,也即,DQN模型的输入是栅格式状态空间图,输出是多个换道动作的Q值。DQN是建立在Q-learning的基础上,用神经网络(Neural Network)取代Q-learning中的Q表(Q-table),DQN中有两个结构相同但是参数值不同的神经网络:当前神经网络及目标神经网络,分别用 $\theta$ 和 $\theta^-$ 代表两个神经网络的参数,并以此计算Q值。其迭代更新状态的Q值为:

$$[0095] \quad Q(s, a; \theta) = Q(s, a; \theta) + \alpha [r + \gamma \max_{a'} Q(s', a'; \theta^-) - Q(s, a; \theta)]$$

[0096] 其中,s为当前状态,也即当前的车辆状态空间图。 $Q(s, a; \theta)$ 为当前神经网络计算的当前状态s下执行动作a的Q值; $Q(s', a'; \theta^-)$ 为目标神经网络计算的下一个状态s'下执行动作a'的Q值; $\alpha$ 是学习效率; $\gamma$ 为折扣因子,表示对未来奖励值的衰减值;r是执行动作a后得到的奖励值。DQN通过不断地学习调参,神经网络中的参数最终会收敛到一个稳定的值。

[0097] 目标神经网络的参数值采用延迟更新的方式更新,也即目标神经网络的参数值固定一段时间后被当前神经网络的参数值所替代并再次固定一段时间。在一些实施例中,延迟更新的步数为300。在DQN中,随机抽取记忆库中的数据进行学习,打乱了经历之间的相关性,使得神经网络更新更有效率,上述目标神经网络通过延迟更新参数的方式同样也降低了数据经历间的相关性。此外,使用当前神经网络和目标神经网络进行训练,在训练结束后,仅使用当前神经网络选择决策动作。

[0098] 在一些实施例中,如图8所示,深度强化学习模型的输入层(Input)为栅格式状态空间图(occupancy grid,  $30 \times 15 \times 3$ ),采用卷积层(Convolutional Layer,简称Conv)连接。深度强化学习模型的隐藏层有两层,第一层为前述的卷积层,其中卷积层中卷积核的尺寸为 $5 \times 5 \times 3$ 。栅格通过单个卷积层传递,然后被平展(flat)后与全连接层(FC)连接。第二层为全连接层,与输出层相连,其中神经元数目为10。深度强化学习模型输出为当前状态s下分别执行车辆换道动作空间包括5个换道动作的5个Q值(Q-value),5个换道动作为N(No-op):无操作、A(Accelerate):以恒定的速度加速、D(Decelerate):以恒定的速度减速、L(Left):向左变道、R(Right):向右变道。

[0099] 在一些实施例中,深度强化学习模型训练过程中的奖励值设置如下:若车辆变速行驶,则奖励值基于当前车速和当前所在车道的最小车速确定;若车辆直线行驶,则奖励值为预设正值;若车辆发生碰撞,则奖励值为预设负值。在一些实施例中,加速的奖励值= $0.1 \times (\text{当前车速} - \text{当前所在车道的最小速度})$ ;减速的奖励值= $0.05 \times (\text{当前车速} - \text{当前所在$

车道的最小速度)；车辆直线行驶的奖励值为0.1；碰撞的奖励值为-3。本公开实施例通过改进深度强化学习模型的奖励值设置,提高了学习效率,降低了时间的复杂度。

[0100] 在一些实施例中,深度强化学习模型的记忆池(Replay Memory)大小设置为50000,采样数目为512。在一些实施例中,深度强化学习模型的记忆回放采样包括:确定记忆池中每个样本的贡献值;基于每个样本的贡献值,确定每个样本被采样的概率;基于每个样本被采样的概率,对记忆池中的所有样本进行采样。

[0101] 在一些实施例中,为深度强化学习模型的记忆池(Replay Memory)引入贡献值向量 $w = \{w_i | i=1,2,\dots, \text{Memory\_size}=50000\}$ ,用于样本选择。 $w_i$ 为记忆池中的样本 $i$ 的贡献值,表示对记忆池及神经网络训练的贡献程度,其值越大说明越重要。形式上, $w_i$ 定义如下:

$$[0102] \quad w_i = |r + \gamma \max_{a'} Q(s', a'; \theta^-) - Q(s, a; \theta)|^\beta$$

[0103] 其中,引入参数 $\beta$ 控制多样本加权中的激励强度,并假定 $\beta$ 为已知常数。 $\beta$ 越大,权值分布越平滑。

[0104] 在一些实施例中,考虑到完全按照样本贡献值进行采样会降低样本的多样性,因此,记忆回放采样时在样本贡献值的基础上引入随机性,使样本被抽取的概率与样本贡献值成正比,所有样本均有机会被采样。形式上,定义记忆池中样本 $i$ 被采样的概率为:

$$[0105] \quad p(i) = \frac{w_i}{\sum_k w_k}$$

[0106] 其中, $\sum_k w_k$ 表示记忆池中所有样本贡献值的和。

[0107] 可见,相比传统的记忆回放为随机采样,本公开实施例通过改进深度强化学习模型的采样方式,通过选择更加重要的样本,使得收敛速度更快,提高了学习效率,降低了时间的复杂度。

[0108] 确定单元304,用于基于多个换道动作的决策信息和车辆状态信息,确定一个换道动作。在一些实施例中,确定单元304,用于基于预先设置的车辆行驶约束条件和车辆状态信息,确定至少一个有效换道动作,实现对决策动作硬子空间的划分,得到有效换道动作的子空间,该子空间中的所有动作均为有效换道动作,相应地,非有效换道的动作被划分到无需探索的子空间,提高了DQN学习效率,降低了时间的复杂度。进而确定单元304基于多个换道动作的决策信息和至少一个有效换道动作,确定一个换道动作。

[0109] 在一些实施例中,确定单元304基于预先设置的车辆行驶约束条件和车辆状态信息,可以为车辆划分出有效换道动作的子空间和无需探索的子空间。例如在换道决策问题中,(1)若车辆处于最左侧车道,则采取向左换道的决策会导致车辆离开道路;(2)若车辆已达到行驶道路的最高限速,则采取加速的决策会导致车辆超过最高限速;等等。因此,有效换道动作的子空间不包含向左换道和加速,保证向左换道和加速不会被选中。进而不需要为离开道路设置一个额外负的奖励值,从而简化了奖励功能。此外,由于车辆不探索无需探索的子空间,学习更加快速和高效。

[0110] 在一些实施例中,确定单元304划分有效换道动作的子空间和无需探索的子空间具体为:对深度强化学习模型的输出引入参数 $\lambda = \{\lambda_i | i=1,2,3,4,5. (N,A,D,L,R)\}$ ,参数 $\lambda_i$ 的取值为: $\lambda_i \in [0,1]$ 。若 $\lambda_i = 0$ ,则第 $i$ 个换道动作不需要被探索;若 $\lambda_i = 1$ ,则第 $i$ 个换道动作为有效换道动作。参数 $\lambda_i$ 基于预先设置的车辆行驶约束条件来选择,例如,若车辆处于最左侧车道且速度已达到行驶道路的最高限速,则 $\lambda_1 = 1$ (保持现状), $\lambda_2 = 0$ (加速), $\lambda_3 = 1$ (减

速),  $\lambda_4=0$  (向左换道),  $\lambda_5=1$  (向右换道)。

[0111] 在一些实施例中,确定单元304基于多个换道动作的决策信息和至少一个有效换道动作,确定一个换道动作时,可从至少一个有效换道动作中确定决策值最大的换道动作,或者,从至少一个有效换道动作中确定满足安全条件且决策值最大的换道动作。在一些实施例中,安全条件为:执行换道动作后车辆与前车的距离大于或等于安全距离。其中,安全距离基于车辆的车速确定。在一些实施例中,安全距离 $d_{safe}$ 定义为:

$$[0112] \quad d_{safe} = \frac{v^2}{200\mu}$$

[0113] 其中, $v$ 为车辆的当前速度, $\mu$ 为摩擦系数,设定为0.8。 $d_{safe}$ 与车速成正相关,随着车速的增大, $d_{safe}$ 也随之增大。也即车速较快时,安全距离应更大。计算执行换道动作后车辆与前车的距离 $D$ (若观测范围内无车辆,则 $D=+\infty$ ),若 $D < d_{safe}$ ,则确定该换道动作存在安全隐患,不选择该动作。

[0114] 可见,本公开实施例基于安全条件实现换道安全性分析,增加了车辆换道的安全性。

[0115] 在一些实施例中,换道决策模块300中各单元的划分仅为一种逻辑功能划分,实际实现时可以有另外的划分方式,例如第一获取单元301、构建单元302、第二获取单元303和确定单元304可以实现为一个单元;第一获取单元301、构建单元302、第二获取单元303或确定单元304也可以划分为多个子单元。可以理解的是,各个单元或子单元能够以电子硬件、或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行,取决于技术方案的特定应用和设计约束条件。本领域技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能。

[0116] 图4是本公开实施例提供的一种车载设备的结构示意图。车载设备可支持智能驾驶系统的运行。

[0117] 如图4所示,车载设备包括:至少一个处理器401、至少一个存储器402和至少一个通信接口403。车载设备中的各个组件通过总线系统404耦合在一起。通信接口403,用于与外部设备之间的信息传输。可理解地,总线系统404用于实现这些组件之间的连接通信。总线系统404除包括数据总线之外,还包括电源总线、控制总线和状态信号总线。但为了清楚说明起见,在图4中将各种总线都标为总线系统404。

[0118] 可以理解,本实施例中的存储器402可以是易失性存储器或非易失性存储器,或可包括易失性和非易失性存储器两者。

[0119] 在一些实施方式中,存储器402存储了如下的元素,可执行单元或者数据结构,或者他们的子集,或者他们的扩展集:操作系统和应用程序。

[0120] 其中,操作系统,包含各种系统程序,例如框架层、核心库层、驱动层等,用于实现各种基础业务以及处理基于硬件的任务。应用程序,包含各种应用程序,例如媒体播放器(Media Player)、浏览器(Browser)等,用于实现各种应用业务。实现本公开实施例提供的车辆自动换道决策的方法的程序可以包含在应用程序中。

[0121] 在本公开实施例中,处理器401通过调用存储器402存储的程序或指令,具体的,可以是应用程序中存储的程序或指令,处理器401用于执行本公开实施例提供的车辆自动换道决策的方法各实施例的步骤。

[0122] 本公开实施例提供的车辆自动换道决策的方法可以应用于处理器401中,或者由处理器401实现。处理器401可以是一种集成电路芯片,具有信号的处理能力。在实现过程中,上述方法的各步骤可以通过处理器401中的硬件的集成逻辑电路或者软件形式的指令完成。上述的处理器401可以是通用处理器、数字信号处理器(Digital Signal Processor, DSP)、专用集成电路(Application Specific Integrated Circuit,ASIC)、现成可编程门阵列(Field Programmable Gate Array,FPGA)或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。

[0123] 本公开实施例提供的车辆自动换道决策的方法的步骤可以直接体现为硬件译码处理器执行完成,或者用译码处理器中的硬件及软件单元组合执行完成。软件单元可以位于随机存储器,闪存、只读存储器,可编程只读存储器或者电可擦写可编程存储器、寄存器等本领域成熟的存储介质中。该存储介质位于存储器402,处理器401读取存储器402中的信息,结合其硬件完成方法的步骤。

[0124] 图5为本公开实施例提供的一种车辆自动换道决策的方法流程图。该方法的执行主体为车载设备,在一些实施例中,该方法的执行主体为车载设备所支持的智能驾驶系统。在一些实施例中,可预先设置车辆换道动作空间和车辆行驶约束条件,便于在车辆行驶约束条件的约束下,从车辆换道动作空间中决策一个换道动作。

[0125] 在一些实施例中,车辆换道动作空间包括多个换道动作,也可以称为车辆换道动作集合。在一些实施例中,车辆换道动作空间包括5个换道动作:

[0126] (1) N(No-op):无操作。不采取任何行为,保持当前状态。

[0127] (2) A(Accelerate):以恒定的速度加速。

[0128] (3) D(Decelerate):以恒定的速度减速。

[0129] (4) L(Left):向左变道。

[0130] (5) R(Right):向右变道。

[0131] 在一些实施例中,车辆行驶约束条件为车辆行驶过程中应当遵守的规则。在一些实施例中,车辆行驶约束条件包括但不限于以下一个或多个:

[0132] (1) 若车辆处于最左侧车道时,不作出向左变道决策。

[0133] (2) 若车辆处于最右侧车道时,不作出向右变道决策。

[0134] (3) 若行驶道路存在最高限速,且车辆速度达到最高限速临界点时,不作出加速决策。

[0135] (4) 若行驶道路存在最低限速,且车辆速度达到最低限速临界点时,不作出减速决策。

[0136] (5) 若即将经过斑马线时,不作出加速决策。

[0137] (6) 若检测到红灯或黄灯时,强制作出减速决策。

[0138] 如图5所示,车辆自动换道决策的方法包括如下步骤501至504:

[0139] 501、获取环境信息和车辆状态信息。在一些实施例中,环境信息包括但不限于车辆传感器采集的车辆周围环境的信息;车辆状态信息包括但不限于当前车速和当前所在车道。

[0140] 502、基于环境信息,构建车辆状态空间图;其中,车辆状态空间图为车辆周围预设

范围的图形信息。在一些实施例中,由于车辆状态空间图为图形信息,因此,基于摄像头、激光雷达等传感器采集的可视的环境信息,也称为外部状态可视信息,来构建车辆状态空间图。

[0141] 在一些实施例中,车辆状态空间图包括:车辆(也即本车)、车辆周围预设范围内的车道以及车辆周围预设范围内的障碍物。在一些实施例中,车辆周围预设范围,包括:车辆前向预设距离、车辆后向预设距离以及车辆侧向预设距离。车辆侧向预设距离包括车辆左侧预设距离和车辆右侧预设距离。本实施例并非基于传感器可观测范围来构建车辆状态空间图,而是基于预设范围来构建车辆状态空间图,预设范围比观测范围小。

[0142] 在一些实施例中,车辆前向预设距离为20米、车辆后向预设距离为4米以及车辆侧向预设距离为6米,也即车辆左侧预设距离为6米和车辆右侧预设距离为6米。

[0143] 在一些实施例中,车辆状态空间图包括栅格式状态空间图;其中,栅格式状态空间图中每个栅格对应预设尺寸的道路区域;且车辆、车道和障碍物均由栅格表示。在一些实施例中,每个栅格大小对应1米×1米的道路区域,车辆对应3×6个栅格,车道的宽度对应5个栅格,车辆前向预设距离对应20个栅格,车辆后向预设距离对应4个栅格,车辆左侧预设距离对应6个栅格,车辆右侧预设距离对应6个栅格。

[0144] 在一些实施例中,基于环境信息,确定车辆周围预设范围内的车道以及车辆周围预设范围内的障碍物。进而基于车辆周围预设范围内的车道以及车辆周围预设范围内的障碍物,构建栅格式状态空间图。

[0145] 图6是本公开实施例提供的一种车辆周围环境的示意图,图7是基于图6所示的车辆周围环境构建的车辆状态空间图。如图7所示,车辆状态空间图为栅格式状态空间图,包括3个车道,本车位于3个车道的中间车道,由于本车前方的障碍物(车辆)部分已超出车辆前向预设距离,因此,该障碍物只有3×2个栅格的区域包含在车辆状态空间图中,可见,车辆状态空间图保留了包括车道在内的完整的环境信息,没有丢失整体的统计信息,可提高决策有效性。

[0146] 503、基于车辆状态空间图,获取多个换道动作的决策信息。在一些实施例中,基于车辆状态空间图和深度强化学习模型,获取多个换道动作的决策信息。在一些实施例中,深度强化学习模型的输入是车辆状态空间图,输出是多个换道动作的决策信息,由于预先设置了车辆换道动作空间,因此,所述多个换道动作作为车辆换道动作空间中的换道动作。相比基于统计的方法需要大量的换道数据样本用于模型训练和测试,而不同场景下产生的换道动机有很大的差异,很难收集到全面并且完整的换道数据的问题。本公开实施例通过强化学习考虑了车辆与环境的交互过程,将环境因素考虑其中;不需要历史数据,而是通过强化学习让车辆自主学习训练。

[0147] 在一些实施例中,使用的深度强化学习模型为DQN(Deep Q Network)模型。在一些实施例中,还可以使用其他的深度强化学习模型,例DDPG(Deep Deterministic Policy Gradient)。决策信息为深度强化学习模型的值函数(Value Function)的Q值,也即,DQN模型的输入是栅格式状态空间图,输出是多个换道动作的Q值。DQN是建立在Q-learning的基础上,用神经网络(Neural Network)取代Q-learning中的Q表(Q-table),DQN中有两个结构相同但是参数值不同的神经网络:当前神经网络及目标神经网络,分别用 $\theta$ 和 $\theta^-$ 代表两个神经网络的参数,并以此计算Q值。其迭代更新状态的Q值为:

[0148]  $Q(s, a; \theta) = Q(s, a; \theta) + \alpha [r + \gamma \max_{a'} Q(s', a'; \theta) - Q(s, a; \theta)]$

[0149] 其中,  $s$  为当前状态, 也即当前的车辆状态空间图。 $Q(s, a; \theta)$  为当前神经网络计算的当前状态  $s$  下执行动作  $a$  的  $Q$  值;  $Q(s', a'; \theta)$  为目标神经网络计算的下一个状态  $s'$  下执行动作  $a'$  的  $Q$  值;  $\alpha$  是学习效率;  $\gamma$  为折扣因子, 表示对未来奖励值的衰减;  $r$  是执行动作  $a$  后得到的奖励值。DQN 通过不断地学习调参, 神经网络中的参数最终会收敛到一个稳定的值。

[0150] 目标神经网络的参数值采用延迟更新的方式更新, 也即目标神经网络的参数值固定一段时间后被当前神经网络的参数值所替代并再次固定一段时间。在一些实施例中, 延迟更新的步数为 300。在 DQN 中, 随机抽取记忆库中的数据进行学习, 打乱了经历之间的相关性, 使得神经网络更新更有效率, 上述目标神经网络通过延迟更新参数的方式同样也降低了数据经历间的相关性。此外, 使用当前神经网络和目标神经网络进行训练, 在训练结束后, 仅使用当前神经网络选择决策动作。

[0151] 在一些实施例中, 如图 8 所示, 深度强化学习模型的输入层 (Input) 为栅格式状态空间图 (occupancy grid,  $30 \times 15 \times 3$ ), 采用卷积层 (Convolutional Layer, 简称 Conv) 连接。深度强化学习模型的隐藏层有两层, 第一层为前述的卷积层, 其中卷积层中卷积核的尺寸为  $5 \times 5 \times 3$ 。栅格通过单个卷积层传递, 然后被平展 (flat) 后与全连接层 (FC) 连接。第二层为全连接层, 与输出层相连, 其中神经元数目为 10。深度强化学习模型输出为当前状态  $s$  下分别执行车辆换道动作空间包括 5 个换道动作的 5 个  $Q$  值 ( $Q$ -value), 5 个换道动作为  $N$  (No-op): 无操作、 $A$  (Accelerate): 以恒定的速度加速、 $D$  (Decelerate): 以恒定的速度减速、 $L$  (Left): 向左变道、 $R$  (Right): 向右变道。

[0152] 在一些实施例中, 深度强化学习模型训练过程中的奖励值设置如下: 若车辆变速行驶, 则奖励值基于当前车速和当前所在车道的最小车速确定; 若车辆直线行驶, 则奖励值为预设正值; 若车辆发生碰撞, 则奖励值为预设负值。在一些实施例中, 加速的奖励值  $= 0.1 \times (\text{当前车速} - \text{当前所在车道的最小速度})$ ; 减速的奖励值  $= 0.05 \times (\text{当前车速} - \text{当前所在车道的最小速度})$ ; 车辆直线行驶的奖励值为 0.1; 碰撞的奖励值为 -3。本公开实施例通过改进深度强化学习模型的奖励值设置, 提高了学习效率, 降低了时间的复杂度。

[0153] 在一些实施例中, 深度强化学习模型的记忆池 (Replay Memory) 大小设置为 50000, 采样数目为 512。在一些实施例中, 深度强化学习模型的记忆回放采样包括: 确定记忆池中每个样本的贡献值; 基于每个样本的贡献值, 确定每个样本被采样的概率; 基于每个样本被采样的概率, 对记忆池中的所有样本进行采样。

[0154] 在一些实施例中, 为深度强化学习模型的记忆池 (Replay Memory) 引入贡献值向量  $w = \{w_i | i = 1, 2, \dots, \text{Memory\_size} = 50000\}$ , 用于样本选择。 $w_i$  为记忆池中的样本  $i$  的贡献值, 表示对记忆池及神经网络训练的贡献程度, 其值越大说明越重要。形式上,  $w_i$  定义如下:

[0155]  $w_i = |r + \gamma \max_{a'} Q(s', a'; \theta) - Q(s, a; \theta)|^\beta$

[0156] 其中, 引入参数  $\beta$  控制多样本加权中的激励强度, 并假定  $\beta$  为已知常数。 $\beta$  越大, 权值分布越平滑。

[0157] 在一些实施例中, 考虑到完全按照样本贡献值进行采样会降低样本的多样性, 因此, 记忆回放采样时在样本贡献值的基础上引入随机性, 使样本被抽取的概率与样本贡献值成正比, 所有样本均有机会被采样。形式上, 定义记忆池中样本  $i$  被采样的概率为:

$$[0158] \quad p(i) = \frac{w_i}{\sum_k w_k}$$

[0159] 其中,  $\sum_k w_k$  表示记忆池中所有样本贡献值的和。

[0160] 可见, 相比传统的记忆回放为随机采样, 本公开实施例通过改进深度强化学习模型的采样方式, 通过选择更加重要的样本, 使得收敛速度更快, 提高了学习效率, 降低了时间的复杂度。

[0161] 504、基于多个换道动作的决策信息和车辆状态信息, 确定一个换道动作。在一些实施例中, 基于预先设置的车辆行驶约束条件和车辆状态信息, 确定至少一个有效换道动作, 实现对决策动作硬子空间的划分, 得到有效换道动作的子空间, 该子空间中的所有动作均为有效换道动作, 相应地, 非有效换道的动作被划分到无需探索的子空间, 提高了DQN学习效率, 降低了时间的复杂度。进而基于多个换道动作的决策信息和至少一个有效换道动作, 确定一个换道动作。

[0162] 在一些实施例中, 基于预先设置的车辆行驶约束条件和车辆状态信息, 可以为车辆划分出有效换道动作的子空间和无需探索的子空间。例如在换道决策问题中, (1) 若车辆处于最左侧车道, 则采取向左换道的决策会导致车辆离开道路; (2) 若车辆已达到行驶道路的最高限速, 则采取加速的决策会导致车辆超过最高限速; 等等。因此, 有效换道动作的子空间不包含向左换道和加速, 保证向左换道和加速不会被选中。进而不需要为离开道路设置一个额外负的奖励值, 从而简化了奖励功能。此外, 由于车辆不探索无需探索的子空间, 学习更加快速和高效。

[0163] 在一些实施例中, 划分有效换道动作的子空间和无需探索的子空间具体为: 对深度强化学习模型的输出引入参数  $\lambda = \{\lambda_i | i=1, 2, 3, 4, 5. (N, A, D, L, R)\}$ , 参数  $\lambda_i$  的取值为:  $\lambda_i = [0, 1]$ 。若  $\lambda_i = 0$ , 则第  $i$  个换道动作不需要被探索; 若  $\lambda_i = 1$ , 则第  $i$  个换道动作为有效换道动作。参数  $\lambda_i$  基于预先设置的车辆行驶约束条件来选择, 例如, 若车辆处于最左侧车道且速度已达到行驶道路的最高限速, 则  $\lambda_1 = 1$  (保持现状),  $\lambda_2 = 0$  (加速),  $\lambda_3 = 1$  (减速),  $\lambda_4 = 0$  (向左换道),  $\lambda_5 = 1$  (向右换道)。

[0164] 在一些实施例中, 基于多个换道动作的决策信息和至少一个有效换道动作, 确定一个换道动作时, 可从至少一个有效换道动作中确定决策值最大的换道动作, 或者, 从至少一个有效换道动作中确定满足安全条件且决策值最大的换道动作。在一些实施例中, 安全条件为: 执行换道动作后车辆与前车的距离大于或等于安全距离。其中, 安全距离基于车辆的车速确定。在一些实施例中, 安全距离  $d_{safe}$  定义为:

$$[0165] \quad d_{safe} = \frac{v^2}{200\mu}$$

[0166] 其中,  $v$  为车辆的当前速度,  $\mu$  为摩擦系数, 设定为 0.8。 $d_{safe}$  与车速成正相关, 随着车速的增大,  $d_{safe}$  也随之增大。也即车速较快时, 安全距离应更大。计算执行换道动作后车辆与前车的距离  $D$  (若观测范围内无车辆, 则  $D = +\infty$ ), 若  $D < d_{safe}$ , 则确定该换道动作存在安全隐患, 不选择该动作。

[0167] 可见, 本公开实施例基于安全条件实现换道安全性分析, 增加了车辆换道的安全性。

[0168] 需要说明的是,对于前述的各方法实施例,为了简单描述,故将其都表述为一系列的动作组合,但是本领域技术人员能够理解,本公开实施例并不受所描述的动作顺序的限制,因为依据本公开实施例,某些步骤可以采用其他顺序或者同时进行(例如构建栅格式状态空间图时,可以先构建车辆前向预设距离的栅格,再构建车辆后向预设距离的栅格以及车辆侧向预设距离的栅格;也可以同时构建)。另外,本领域技术人员能够理解,说明书中所描述的实施例均属于可选实施例。

[0169] 本公开实施例还提出一种非暂态计算机可读存储介质,所述非暂态计算机可读存储介质存储程序或指令,所述程序或指令使计算机执行如车辆自动换道决策的方法各实施例的步骤,为避免重复描述,在此不再赘述。

[0170] 需要说明的是,在本文中,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、物品或者装置不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、物品或者装置所固有的要素。在没有更多限制的情况下,由语句“包括……”限定的要素,并不排除在包括该要素的过程、方法、物品或者装置中还存在另外的相同要素。

[0171] 本领域的技术人员能够理解,尽管在此所述的一些实施例包括其它实施例中所包括的某些特征而不是其它特征,但是不同实施例的特征的组合意味着处于本公开的范围之内并且形成不同的实施例。

[0172] 本领域的技术人员能够理解,对各个实施例的描述都各有侧重,某个实施例中沒有详述的部分,可以参见其他实施例的相关描述。

[0173] 虽然结合附图描述了本公开的实施方式,但是本领域技术人员可以在不脱离本公开的精神和范围的情况下做出各种修改和变型,这样的修改和变型均落入由所附权利要求所限定的范围之内。



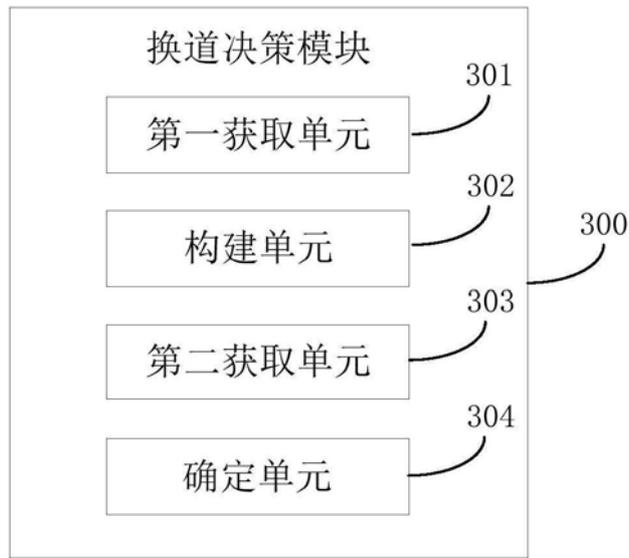


图3

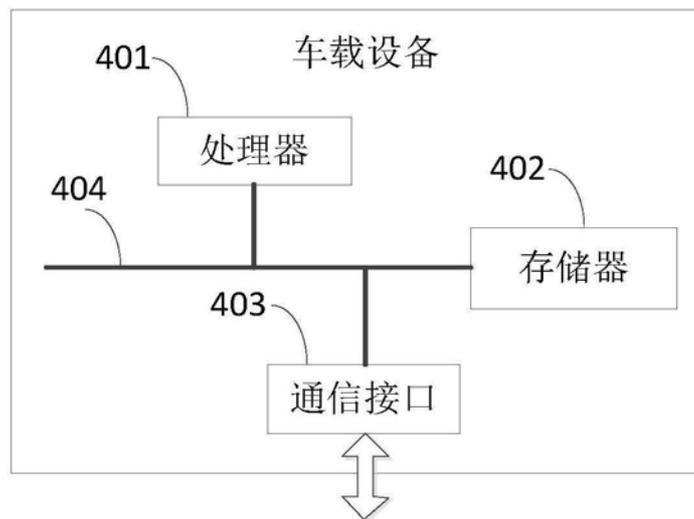


图4

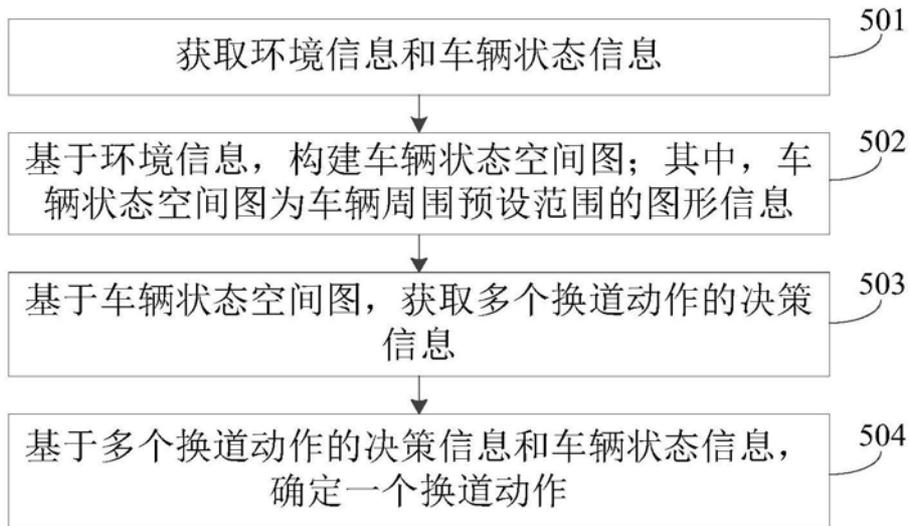


图5

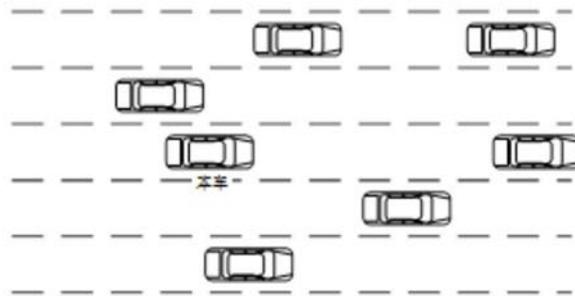


图6

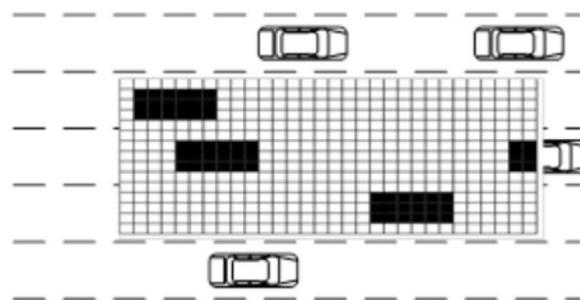


图7

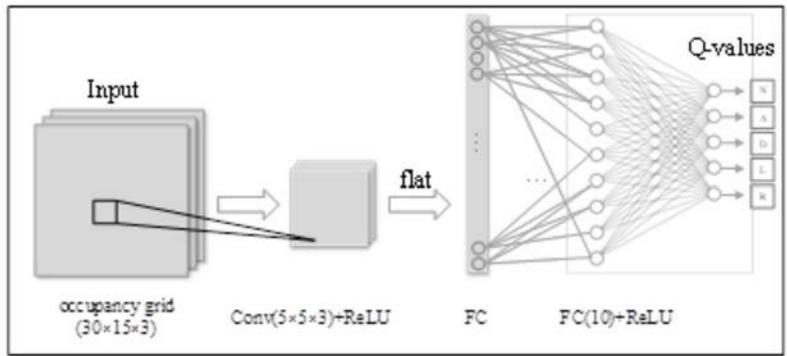


图8