



(12) 发明专利申请

(10) 申请公布号 CN 116203964 A

(43) 申请公布日 2023.06.02

(21) 申请号 202310266544.1

(22) 申请日 2023.03.13

(71) 申请人 阿波罗智联(北京)科技有限公司
地址 100176 北京市大兴区经济技术开发区瑞合西二路7号院1号楼1层101

(72) 发明人 于宁 潘屹峰

(74) 专利代理机构 北京猷德知识产权代理有限公司 16084
专利代理师 范继晨

(51) Int.Cl.
G05D 1/02 (2020.01)

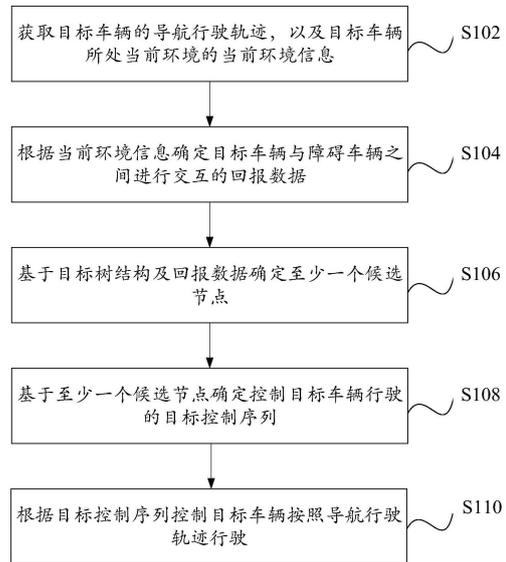
权利要求书2页 说明书13页 附图4页

(54) 发明名称

一种控制车辆行驶的方法、设备和装置

(57) 摘要

本公开提供了一种控制车辆行驶的方法,涉及人工智能领域,尤其涉及自动驾驶等技术领域。具体实现方案为:获取目标车辆的导航行驶轨迹,以及目标车辆所处当前环境的当前环境信息;根据当前环境信息确定目标车辆与障碍车辆之间进行交互的回报数据,其中,回报数据表征了目标车辆在交互场景下的驾驶性能;基于目标树结构及回报数据确定至少一个候选节点,其中,目标树结构由多个节点组成,每个节点表征了目标车辆在每个时刻的控制量;基于至少一个候选节点确定控制目标车辆行驶的目标控制序列;根据目标控制序列控制目标车辆按照导航行驶轨迹行驶。



1. 一种控制车辆行驶的方法,包括:

获取目标车辆的导航行驶轨迹,以及所述目标车辆所处当前环境的当前环境信息,其中,所述当前环境信息至少包括:所述当前环境对应的交通信息、所述目标车辆与所述当前环境中的障碍车辆之间的交互信息;

根据所述当前环境信息确定所述目标车辆与所述障碍车辆之间进行交互的回报数据,其中,所述回报数据表征了所述目标车辆在所述交互场景下的驾驶性能;

基于目标树结构及所述回报数据确定至少一个候选节点,其中,所述目标树结构由多个节点组成,每个节点表征了所述目标车辆在每个时刻的控制量;

基于所述至少一个候选节点确定控制所述目标车辆行驶的目标控制序列;

根据所述目标控制序列控制所述目标车辆按照所述导航行驶轨迹行驶。

2. 根据权利要求1所述的方法,其中,根据所述当前环境信息确定所述目标车辆与所述障碍车辆之间进行交互的回报数据,包括:

根据所述当前环境信息构建所述目标车辆与所述障碍车辆之间进行交互的状态转移函数;

根据所述状态转移函数确定所述回报数据。

3. 根据权利要求2所述的方法,其中,根据所述当前环境信息构建所述目标车辆与所述障碍车辆之间的状态转移函数,包括:

根据所述当前环境信息确定所述目标车辆在多个时刻的第一状态信息、第一控制信息以及所述障碍车辆在所述多个时刻的第二状态信息、第二控制信息;

根据所述第一状态信息、所述第一控制信息、所述第二状态信息以及所述第二控制信息确定所述目标车辆与所述障碍车辆之间的联合状态信息;

基于所述联合状态信息构建所述目标车辆与所述障碍车辆之间的状态转移函数。

4. 根据权利要求3所述的方法,其中,根据所述状态转移函数确定所述回报数据,包括:

获取所述目标车辆对应的专家轨迹信息,其中,所述专家轨迹信息表征了目标对象驾驶所述目标车辆按照所述导航行驶轨迹行驶的过程中,所述目标车辆的状态信息和控制信息;

确定所述专家轨迹信息相对于预设变量的概率分布函数;

计算所述概率分布函数对应的似然函数;

根据所述似然函数的梯度值确定所述预设变量;

根据所述第一状态信息、所述第一控制信息、所述第二状态信息以及所述第二控制信息以及所述预设变量确定所述目标车辆对应的回报数据。

5. 根据权利要求1所述的方法,其中,基于目标树结构及所述回报数据确定至少一个候选节点,包括:

基于所述回报数据计算所述目标树结构中每个节点所对应的平均回报值以及累计回报值;

统计所述目标树结构中每个节点的访问次数;

根据所述平均回报值、所述累计回报值以及所述访问次数从所述目标树结构的多个节点中确定所述至少一个候选节点。

6. 根据权利要求5所述的方法,在基于目标树结构及所述回报数据确定至少一个候选

节点之前,所述方法还包括:

对所述目标车辆的控制信息进行离散化处理,得到离散后的控制信息;

确定所述目标树结构中每个节点所对应的状态信息;

基于所述离散后的控制信息以及所述每个节点所对应的状态信息构建所述目标树结构。

7.根据权利要求5所述的方法,在基于目标树结构及所述回报数据确定至少一个候选节点之前,所述方法还包括:

根据所述当前环境信息以及所述导航行驶轨迹确定初始控制信息;

对所述初始控制信息进行离散化处理,得到离散后的控制信息;

确定所述目标树结构中每个节点所对应的状态信息;

基于所述离散后的控制信息以及所述每个节点所对应的状态信息构建所述目标树结构。

8.一种控制车辆行驶的装置,包括:

获取模块,用于获取目标车辆的导航行驶轨迹,以及所述目标车辆所处当前环境的当前环境信息,其中,所述当前环境信息至少包括:所述当前环境对应的交通信息、所述目标车辆与所述当前环境中的障碍车辆之间的交互信息;

第一确定模块,用于根据所述当前环境信息确定所述目标车辆与所述障碍车辆之间进行交互的回报数据,其中,所述回报数据表征了所述目标车辆在所述交互场景下的驾驶性能;

第二确定模块,用于基于目标树结构及所述回报数据确定至少一个候选节点,其中,所述目标树结构由多个节点组成,每个节点表征了所述目标车辆在每个时刻的控制量;

第三确定模块,用于基于所述至少一个候选节点确定控制所述目标车辆行驶的目标控制序列;

控制模块,用于根据所述目标控制序列控制所述目标车辆按照所述导航行驶轨迹行驶。

9.一种电子设备,包括:

至少一个处理器;以及

与所述至少一个处理器通信连接的存储器;其中,

所述存储器存储有可被所述至少一个处理器执行的指令,所述指令被所述至少一个处理器执行,以使所述至少一个处理器能够执行权利要求1-7中任一项所述的方法。

10.一种存储有计算机指令的非瞬时计算机可读存储介质,其中,所述计算机指令用于使所述计算机执行根据权利要求1-7中任一项所述的方法。

11.一种计算机程序产品,包括计算机程序,所述计算机程序在被处理器执行时实现根据权利要求1-7中任一项所述的方法。

12.一种车辆,其中,所述车辆中包括控制器,所述控制器用于执行权利要求1-7中任一项所述的方法。

一种控制车辆行驶的方法、设备和装置

技术领域

[0001] 本公开涉及人工智能领域,尤其涉及自动驾驶、智能交通、深度学习等技术领域,具体涉及一种控制车辆行驶的方法、设备和装置。

背景技术

[0002] 目前,现有技术中的自动驾驶车辆在求解最佳动作序列或者轨迹时,往往是在给定外部环境的情况下进行,但是,上述方法单纯根据障碍物预测轨迹对自动驾驶车辆做决策规划,缺少了对系统交互博弈的考虑,从而影响了自动驾驶车辆的控制效果。

发明内容

[0003] 本公开提供了一种用于控制车辆行驶的方法、设备和装置。

[0004] 根据本公开的一方面,提供了一种控制车辆行驶的方法,包括:获取目标车辆的导航行驶轨迹,以及目标车辆所处当前环境的当前环境信息,其中,当前环境信息至少包括:当前环境对应的交通信息、目标车辆与当前环境中的障碍车辆之间的交互信息;根据当前环境信息确定目标车辆与障碍车辆之间进行交互的回报数据,其中,回报数据表征了目标车辆在交互场景下的驾驶性能;基于目标树结构及回报数据确定至少一个候选节点,其中,所述目标树结构由多个节点组成,每个节点表征了所述目标车辆在每个时刻的控制量;基于至少一个候选节点确定控制目标车辆行驶的目标控制序列;根据目标控制序列控制目标车辆按照导航行驶轨迹行驶。

[0005] 根据本公开的另一方面,提供了一种控制车辆行驶的装置,包括:获取模块,用于获取目标车辆的导航行驶轨迹,以及目标车辆所处当前环境的当前环境信息,其中,当前环境信息至少包括:当前环境对应的交通信息、目标车辆与当前环境中的障碍车辆之间的交互信息;第一确定模块,用于根据当前环境信息确定目标车辆与障碍车辆之间进行交互的回报数据,其中,回报数据表征了目标车辆在交互场景下的驾驶性能;第二确定模块,用于基于目标树结构及回报数据确定至少一个候选节点,其中,所述目标树结构由多个节点组成,每个节点表征了所述目标车辆在每个时刻的控制量;第三确定模块,用于基于至少一个候选节点确定控制目标车辆行驶的目标控制序列;控制模块,用于根据目标控制序列控制目标车辆按照导航行驶轨迹行驶。

[0006] 根据本公开的另一方面,提供了一种电子设备,包括:至少一个处理器;以及与至少一个处理器通信连接的存储器;其中,存储器存储有可被至少一个处理器执行的指令,指令被至少一个处理器执行,以使至少一个处理器能够执行本公开实施例的控制车辆行驶方法。

[0007] 根据本公开的另一方面,提供了一种存储有计算机指令的非瞬时计算机可读存储介质,其中,计算机指令用于使计算机执行本公开实施例的控制车辆行驶方法。

[0008] 根据本公开的另一方面,提供了一种计算机程序产品,包括计算机程序,计算机程序在被处理器执行时实现本公开实施例的控制车辆行驶方法。

[0009] 根据本公开的另一方面,提供了一种车辆,包括控制器,控制器用于执行本公开实施例的控制车辆行驶方法。

[0010] 由上述内容可知,在本公开中,通过获取目标车辆的导航行驶轨迹,以及目标车辆所处当前环境的当前环境信息确定目标车辆与障碍车辆之间进行交互的回报数据,再基于多个节点组成的目标树结构及回报数据确定至少一个候选节点,基于至少一个候选节点确定控制目标车辆行驶的目标控制序列,最终根据目标控制序列控制目标车辆按照导航行驶轨迹行驶,从而避免了相关技术中对自动驾驶车辆控制精度不高,轨迹规划不准确的技术问题。由此可见,本公开所提供的方案解决了现有技术中在对自动驾驶车辆进行控制时所存在的合理性差的问题,达到了提高自动驾驶车辆在交互博弈场景中控制合理性的技术效果。

[0011] 应当理解,本部分所描述的内容并非旨在标识本公开的实施例的关键或重要特征,也不用于限制本公开的范围。本公开的其它特征将通过以下的说明书而变得容易理解。

附图说明

[0012] 附图用于更好地理解本方案,不构成对本公开的限定。其中:

[0013] 图1是根据本公开实施例的一种控制车辆行驶的方法的流程图;

[0014] 图2是根据本公开实施例的一种自动驾驶车辆核心模块的示意图;

[0015] 图3是根据本公开实施例的一种蒙特卡洛树搜索方法的示意图;

[0016] 图4是根据本公开实施例的一种控制车辆行驶装置的示意图;

[0017] 图5是用来实现本公开实施例的控制车辆行驶方法的电子设备的示意图。

具体实施方式

[0018] 以下结合附图对本公开的示范性实施例做出说明,其中包括本公开实施例的各种细节以助于理解,应当将它们认为仅仅是示范性的。因此,本领域普通技术人员应当认识到,可以对这里描述的实施例做出各种改变和修改,而不会背离本公开的范围和精神。同样,为了清楚和简明,以下的描述中省略了对公知功能和结构的描述。

[0019] 例如,本公开提供的控制车辆行驶的方法,可以应用在控制自动驾驶车辆驾驶的场景下,通过获取目标车辆的导航行驶轨迹和所处当前环境的环境信息来确定目标车辆与障碍车辆之间进行交互的回报数据,再基于多个节点组成的目标树结构及回报数据确定出候选节点,进而根据候选节点确定控制目标车辆行驶的目标控制序列,最终根据目标控制序列控制目标车辆按照导航行驶轨迹行驶,从而为控制自动驾驶车辆行驶的过程提供了一种提高控制合理性的技术手段。该技术方案的关键在于,通过在交互博弈场景中,构建多车交互场景下的回报数据,基于该回报数据确定控制车辆的最优的控制序列,从而,本发明提供的技术方案,实现了可以通过构建交互博弈场景来以确定出目标车辆的最优控制序列,从而可以达到提高自动驾驶车辆在交互博弈场景中控制合理性的技术效果。

[0020] 下面对本公开实施例的一种控制车辆行驶的方法进行介绍。

[0021] 图1是根据本公开实施例的一种控制车辆行驶的方法流程图,如图1所示,该方法可以包括以下步骤:

[0022] 步骤S102,获取目标车辆的导航行驶轨迹,以及目标车辆所处当前环境的当前环

境信息,其中,当前环境信息至少包括:当前环境对应的交通信息、目标车辆与当前环境中的障碍车辆之间的交互信息。

[0023] 可选地,本实施例的执行主体为自动驾驶系统,其中,其他电子设备、处理器也可以作为执行主体,在此不作更多限定。

[0024] 在本公开上述步骤S102提供的技术方案中,如图2所示,目标车辆为自动驾驶车辆,自动驾驶系统中的全局导航模块和感知系统模块可以获取目标车辆的导航行驶轨迹和目标车辆所处当前环境的当前环境信息,其中,目标行驶轨迹为全局导航模块通过全局路径搜索算法得到的符合性能评估指标的最优全局导航路径,环境信息包括当前环境对应的交通信息、目标车辆与当前环境中的障碍车辆之间的交互信息,从而避免了相关技术中未考虑到目标车辆和障碍车辆之间进行交互的情况所产生的问题,达到了为自动驾驶系统提供最优的全局导航路径和准确的环境信息的技术效果。

[0025] 可选地,外部环境具体包括但不限于所有交通参与者的位置、姿态、历史运动轨迹、未来运动轨迹(预测)等,在此不作进一步地限定。

[0026] 可选地,如图2所示,全局导航模块为自动驾驶系统中的特定功能模块,其中,全局导航模块可以根据自动驾驶车辆初始位置和目标位置,结合路网拓扑结构,通过全局路径搜索算法,得到符合性能评估指标的最优全局导航路径。

[0027] 可选地,如图2所示,感知系统模块为自动驾驶系统中的特定功能模块,通过相机、激光雷达、毫米波雷达、超声波雷达等设备结合先进的障碍物检测算法,为自动驾驶车辆提供全方位的环境感知服务。

[0028] 可选地,如图2所示,自动驾驶系统还可以通过高精地图模块和定位系统模块获取自动驾驶车辆的初始位置,其中高精地图模块可以提供高精度地图服务,定位系统模块可以提供高精度(厘米级)的定位服务。

[0029] 可选地,如图2所示,自动驾驶系统中还包括预测系统模块,上述预测系统模块可获取到上游感知系统的数据为输入,通过提取障碍物历史运动参数,结合卡尔曼滤波、神经网络等手段,推理得到障碍物未来时刻运动轨迹,用于下游规划控制模块使用。

[0030] 可选地,如图2所示,预测模块可以将预测得到的不同导航行驶轨迹输入到全局导航模块中,再依据性能评估指标挑选出最优的导航行驶轨迹,其中性能评估指标可以包括目标车辆的行驶距离,目标车辆的耗油量,目标车辆的行驶状态。

[0031] 需要说明的是,上述内容仅为举例说明,此处不对调用目标车辆的导航行驶轨迹以及目标车辆所处当前环境的当前环境信息的具体实现方式进行限定,其他能够调用目标车辆的导航行驶轨迹以及目标车辆所处当前环境的当前环境信息的实现方式均在本公开的保护范围内。

[0032] 步骤S104,根据当前环境信息确定目标车辆与障碍车辆之间进行交互的回报数据,其中,回报数据表征了目标车辆在交互场景下的驾驶性能。

[0033] 在本公开上述步骤S104提供的技术方案中,自动驾驶系统可以根据感知系统模块获取到的环境信息确定目标车辆与障碍车辆之间进行交互的回报数据,从而构建目标车辆与障碍车辆之间的交互场景,其中,回报数据可以表征目标车辆在当前交互场景下的驾驶性能,从而达到了为目标车辆与障碍车辆之间建立交互博弈场景的技术效果。

[0034] 可选地,回报数据可以包括安全指标、高效指标、舒适指标和合规指标等,其中安

全指标表征了目标车辆与障碍车辆之间的安全指数,即表征了目标车辆与障碍车辆或障碍物之间发生碰撞的可能性;高效指标表征了目标车辆与障碍车辆交互后所生成的行驶轨迹是否为目标车辆到目标位置的最近距离,或,高效指标表征了目标车辆行驶上述行驶轨迹的所用时间是否为最短时间;舒适指标表征了目标车辆在行驶上述轨迹的过程中的舒适程度,例如在行驶路线上是否存在道路障碍或视野障碍;合规指标表征了车辆在行驶过程中是否存在违反交通规则的现象,或,合规指标表征了车辆在行驶过程中是否出现了不合理的现象,例如存在车辆行驶路线绕远的问题或车辆行驶路线中存在修路路段的问题。

[0035] 可选地,可以通过环境信息确定目标车辆和障碍车辆进行交互的回报数据包括:依据环境信息对目标车辆和障碍车辆进行交互博弈问题的系统建模。

[0036] 可选地,对目标车辆和障碍车辆进行交互博弈问题的系统建模所采用的算法可以为基于图搜索的方法(Graph search based planners)、基于采样的方法(Sampling based planners)、插值曲线(Interpolation curve planners)以及数值优化方法(Numerical optimization approach),在此不作具体限定。

[0037] 可选地,确定目标车辆与障碍车辆之间进行交互的回报数据可以为确定目标车辆与障碍车辆之间进行交互的回报函数,其中回报函数表征了目标车辆与障碍车辆所构建模型的输出数据。

[0038] 步骤S106,基于目标树结构及回报数据确定至少一个候选节点,其中,目标树结构由多个节点组成,每个节点表征了目标车辆在每个时刻的控制量。

[0039] 在本公开上述步骤S106提供的技术方案中,自动驾驶系统基于目标树的结构和回报数据来确定至少一个候选节点,其中目标树结构中可以有多个节点,其中每个节点表征了目标车辆在每个时刻的运动信息,确定候选节点的过程可以包括对不同节点组成的不同路径计算回报数据值,依据预设的确定条件来确定出候选路径,上述候选路径上的节点为候选节点,从而可以达到使依据回报数据来确定的候选节点更加准确的技术效果。

[0040] 可选地,目标树的类型可以为蒙特卡洛树(MCTS),或者其他具有多个节点的树结构,在此不作具体限定。

[0041] 可选地,目标树中的多个节点可以包括根节点,子节点,叶子节点及终止节点,不同节点对应了目标车辆在不同时刻的运动信息。

[0042] 可选地,确定至少一个候选节点的过程可以为将蒙特卡洛树应用于自动驾驶决策规划的过程,包括了对目标行驶过程中的控制变量进行离散化,确定出多个节点的信息,基于上述信息和节点选择策略确定出最终候选节点。

[0043] 可选地,挑选候选路径的预设条件可以设置为回报数据值最高的路径或者回报数据值满足某个经验数值的路径,只要具有特定的条件即可,在此不作具体限定。

[0044] 步骤S108,基于至少一个候选节点确定控制目标车辆行驶的目标控制序列。

[0045] 本公开上述步骤S108提供的技术方案中,每个候选节点都对应一个目标控制序列,自动驾驶系统基于上述确定的候选节点来确定出对应的控制目标车辆行驶的最优决策序列,从而达到了可以根据确定的候选节点选出最优的目标控制序列的技术效果。

[0046] 可选地,自动驾驶系统中包含决策规划模块,上述目标控制序列可以存储至决策规划模块内,其中决策规划模块主要提供主车避让障碍物、换道决策,路径规划、速度规划服务。

[0047] 步骤S110,根据目标控制序列控制目标车辆按照导航行驶轨迹行驶。

[0048] 本公开上述步骤S110提供的技术方案中,自动驾驶系统根据得到的目标控制序列对车辆进行决策应用,控制目标车辆按照导航行驶轨迹行驶,从而保证了目标车辆可以依据最优控制决策在目标导航轨迹上行驶。

[0049] 可选地,自动驾驶系统中包含控制系统模块,上述控制步骤的执行主体即可以为控制系统模块,其中控制系统模块可以根据决策规划系统提供的驾驶轨迹,进行纵向和横向的跟踪控制。

[0050] 可选地,控制系统模块对目标车辆进行控制时,可以采用多种控制方式,例如可以采用前馈控制、反馈控制或者前馈控制和反馈控制的结合,其中控制的应用方法可以为积分微分比例(PID)控制,在此不作进一步地限定。

[0051] 需要说明的是,上述内容仅为举例说明,此处不对控制目标车辆按照导航行驶轨迹行驶的具体实现方式进行限定,其他能够控制目标车辆按照导航行驶轨迹行驶的实现方式均在本公开的保护范围内。

[0052] 上述步骤S102至步骤S110,在本公开中,通过获取目标车辆的导航行驶轨迹,以及目标车辆所处当前环境的当前环境信息确定目标车辆与障碍车辆之间进行交互的回报数据,再基于多个节点组成的目标树结构及回报数据确定至少一个候选节点,基于至少一个候选节点确定控制目标车辆行驶的目标控制序列,最终根据目标控制序列控制目标车辆按照导航行驶轨迹行驶。

[0053] 容易注意到的是,在本公开实施例中,控制车辆自动驾驶的过程中考虑到了其他障碍车辆和自动驾驶车辆的行为存在交互博弈的情况,即其他障碍车辆和自动驾驶车辆当前时刻所采取的动作会影响下一时刻的状态,而下一时刻的状态也会影响下下一时刻其他车辆的决策,从而避免了相关技术中对自动驾驶车辆控制精度不高,轨迹规划不准确的技术问题。由此可见,本公开所提供的方案解决了现有技术中对自动驾驶车辆进行控制时所存在的合理性差的问题,达到了提高自动驾驶车辆在交互博弈场景中控制合理性的技术效果。

[0054] 下面对该实施例的上述方法进行进一步地详细介绍。

[0055] 作为一种可选的实施方式,步骤S104,根据当前环境信息确定目标车辆与障碍车辆之间进行交互的回报数据,包括:根据当前环境信息构建目标车辆与障碍车辆之间进行交互的状态转移函数,根据状态转移函数确定回报数据。

[0056] 在该实施例中,根据当前环境信息确定目标车辆与障碍车辆之间进行交互的回报数据包括以下步骤,根据当前目标车辆的环境信息构建出目标车辆与障碍车辆之间进行交互的状态转移函数,再基于上述状态转移函数运用自动驾驶车辆状态转移方程来确定目标车辆与障碍车辆的回报数据,从而可以将环境信息运用到回报数据的计算中,达到了提高回报数据准确性的技术效果。

[0057] 可选地,自动驾驶车辆状态转移方程的公式如下:

$$[0058] \quad X_{AV}^{t+1} = \mathcal{F}_{AV}(X_{AV}^t, U_{AV}^t)$$

[0059] 其中, \mathcal{F}_{AV} 为自动驾驶车辆状态转移函数。

[0060] 可选地,交互场景自动驾驶车辆回报数据的公式如下:

$$[0061] \quad R_{AV}(X^0, U_{AV}, U_{Obs}) = \sum_{t=1}^N r_{AV}(X^t, U_{AV}^t, U_{Obs}^t)$$

[0062] 可选地,交互场景障碍车回报数据的公式如下:

$$[0063] \quad R_{Obs}(X^0, U_{Obs}, U_{AV}) = \sum_{t=1}^N r_{Obs}(X^t, U_{Obs}^t, U_{AV}^t)$$

[0064] 可选地,回报数据为确定的数值,其中回报数据可以表征目标车辆的驾驶性能。

[0065] 作为一种可选的实施方式,根据当前环境信息构建目标车辆与障碍车辆之间的状态转移函数,包括:根据当前环境信息确定目标车辆在多个时刻的第一状态信息、第一控制信息以及障碍车辆在多个时刻的第二状态信息、第二控制信息,根据第一状态信息、第一控制信息、第二状态信息以及第二控制信息确定目标车辆与障碍车辆之间的联合状态信息,基于联合状态信息构建目标车辆与障碍车辆之间的状态转移函数。

[0066] 在该实施例中,根据当前环境信息构建目标车辆与障碍车辆之间的状态转移函数包括以下步骤,根据当前环境信息来确定目标车辆和障碍车辆在多个时刻的状态信息和控制信息,再根据目标车辆和障碍车辆在多个时刻的状态信息和控制信息并基于交互场景联合状态公式来确定联合状态信息,最终基于联合状态信息来构建目标车辆与障碍车辆之间的状态转移函数,从而限定了在交互博弈场景下交互状态方程的构建,实现了基于状态转移函数确定出回报数据的技术效果。

[0067] 可选地,自动驾驶车辆状态信息可以从如下公式获得:

$$[0068] \quad X_{AV} = (x_1, x_2, \dots, x_T)^T$$

[0069] 其中, T 表示转置, x_1, x_2, \dots, x_T 表示自动驾驶车辆在 t_1, t_2, \dots, t_T 时刻的状态量,比如 x_1 表示在 t_1 时刻自动驾驶车辆的坐标 (x, y, z) ,航向角,速度及加速度等。

[0070] 可选地,自动驾驶车辆控制信息可以从如下公式获得:

$$[0071] \quad U_{AV} = (u_1, u_2, \dots, u_T)^T$$

[0072] 其中, T 表示转置, u_1, u_2, \dots, u_T 表示自动驾驶车辆在 t_1, t_2, \dots, t_T 时刻的控制量,比如 u_1 表示在 t_1 时刻自动驾驶车辆的方向盘转角和加速度。

[0073] 可选地,自动驾驶车辆状态转移方程的公式如下:

$$[0074] \quad X_{AV}^{t+1} = \mathcal{F}_{AV}(X_{AV}^t, U_{AV}^t)$$

[0075] 其中, \mathcal{F}_{AV} 为自动驾驶车辆状态转移函数。

[0076] 可选地,交互场景联合状态公式为:

$$[0077] \quad X = \{X_{AV}, X_{Obs}\}$$

[0078] 其中, X_{AV} 表示自动驾驶车辆的状态信息, X_{Obs} 表示障碍车辆的状态信息。

[0079] 可选地,交互场景联合状态转移方程为:

$$[0080] \quad X^{t+1} = \mathcal{F}_{AV}(X^t, U_{AV}^t)$$

$$[0081] \quad X^{t+1} = \mathcal{F}_{Obs}(\mathcal{F}_{AV}(X^t, U_{AV}^t), U_{Obs}^t)$$

[0082] 其中, \mathcal{F}_{AV} 为自动驾驶车辆的状态转移函数, \mathcal{F}_{Obs} 为障碍车辆的状态转移函数。

[0083] 可选地,自动驾驶系统可以通过自动驾驶车辆状态信息公式和自动驾驶车辆控制

信息公式来获得状态信息和控制信息,再基于上述状态信息和控制信息和交互场景联合状态公式得到目标车辆与所述障碍车辆之间的联合状态信息,最终基于上述联合状态信息和t时刻自动驾驶车辆的状态信息和动作信息获得状态转移函数。

[0084] 作为一种可选的实施方式,根据状态转移函数确定回报数据,包括:获取目标车辆对应的专家轨迹信息,其中,专家轨迹信息表征了目标对象驾驶目标车辆按照导航行驶轨迹行驶的过程中,目标车辆的状态信息和控制信息,确定专家轨迹信息相对于预设变量的概率分布函数,计算概率分布函数对应的似然函数,根据似然函数的梯度值确定预设变量,根据第一状态信息、第一控制信息、第二状态信息以及第二控制信息以及预设变量确定目标车辆对应的回报数据。

[0085] 在该实施例中,根据状态转移函数确定回报数据包括以下步骤,自动驾驶系统可以获取到目标车辆对应的专家轨迹信息,再通过计算专家轨迹信息相对于预设变量的概率分布函数的似然函数的梯度值确定预设变量,再根据目标车辆和障碍车辆的状态信息和控制信息以及上述预设变量来确定目标车辆对应的回报数据,从而可以基于最大熵采样方法确定回报数据,提高了自动驾驶系统基于回报数据控制车辆的合理性。

[0086] 可选地,专家轨迹信息表征了目标对象在驾驶目标车辆按照导航行驶轨迹行驶的过程中目标车辆的状态信息和控制信息,其中专家轨迹信息可以通过采样样本评估得到,例如可以采集不同司机在相应路段的驾驶轨迹,并将多个驾驶轨迹输入至训练模型,通过训练模型对所采集到的样本进行训练得到目标样本,即专家轨迹信息。

[0087] 可选地,上述训练模型的训练参数可以设置为经验数值,例如可以设置为距离最短的驾驶轨迹或耗时最短的驾驶轨迹,在此不作具体限定。

[0088] 可选地,专家轨迹的获取公式如下:

$$[0089] \quad \xi = [x_0, u_0, x_1, u_1, \dots, x_N, u_N]$$

[0090] 可选地,获取概率分布函数的公式如下:

$$[0091] \quad P(\mathcal{D}_E | \theta) = \prod_{i=1}^M \frac{e^{\beta R(\xi_i, \theta)}}{Z_{\xi_i}}$$

[0092] 可选地,获取似然函数的公式如下:

$$[0093] \quad \bar{L}(\theta) \approx \frac{1}{M} \sum_{i=1}^M \log \frac{e^{\beta R(\xi_i, \theta)}}{\sum_m^K e^{\beta R(\tau_m^i, \theta)}}$$

[0094] 其中,似然函数的梯度值确定公式为:

$$[0095] \quad \Delta_{\theta} \bar{L} \approx \frac{\beta}{M} \sum_{i=1}^M (f(\xi_i) - \frac{1}{M} \sum_{i=1}^M \frac{e^{\beta R(\tau_m^i, \theta)}}{\sum_m^K e^{\beta R(\tau_m^i, \theta)}} f(\tau_m^i))$$

[0096] 可选地,根据目标车辆和障碍车辆的状态信息和控制信息获得回报数据还包括基于梯度下降法,将参数 θ_k 逼近最优参数,公式如下:

$$[0097] \quad \theta_k^* \leftarrow \theta_k$$

[0098] 其中,梯度下降法的公式为:

$$[0099] \quad \theta_{k+1} = \theta_k + \Delta_{\theta_k} L = \theta_k + \alpha (\bar{f}(\mathcal{D}_E) - \bar{f}(\mathcal{D}_S))$$

[0100] 可选地,其中f表示了特征向量,表征了目标车辆和障碍车辆的状态信息和动作信息,还充分考虑了目标车辆在每个时刻周围的地图、交通灯信息。

[0101] 可选地,自动驾驶系统可以对目标车辆和障碍车辆的历史轨迹(例如过去10秒)运用LSTM神经网络分别编码,同时考虑到目标车辆和障碍车辆之间的相关影响,可以通过Cross-Attention机制对这种交互进行编码。除此之外,由于目标车辆和障碍车辆的驾驶行为与地图、交通等设施紧密耦合,所以自动驾驶系统可以对历史时间(例如过去10秒)、当前时刻目标车辆和障碍车辆以及交通环境进行联合编码,其中,上述交通环境可以包括静态地图和动态地图事件(例如道路施工、事故)等。

[0102] 作为一种可选的实施方式,步骤S106,基于目标树结构及回报数据确定至少一个候选节点,包括:基于回报数据计算目标树结构中每个节点所对应的平均回报值以及累计回报值,统计所述目标树结构中每个节点的访问次数,根据平均回报值、累计回报值以及访问次数从目标树结构的多个节点中确定至少一个候选节点。

[0103] 在该实施例中,自动驾驶系统基于目标树结构和回报数据确定至少一个候选节点的过程包括以下步骤,基于获取到的回报数据计算目标树结构中每个节点所对应的平均回报值以及累计回报值,再对其中每个节点的访问次数进行统计,最终根据每个节点的平均回报值和累计回报值及访问次数确定出候选节点,从而实现了通过节点选择策略来确定候选节点的过程,达到了可以为自动驾驶系统确定出候选节点,从而进一步可以获取到最优控制序列的技术效果。

[0104] 可选地,目标树的类型可以为蒙特卡洛树(MCTS),或者其他具有多个节点的树结构,在此不作具体限定。

[0105] 可选地,如图3所示,蒙特卡洛树搜索方法包括了四个关键步骤:

[0106] 1) 选择(Selection):从根节点开始,递归选择最优的子节点,最终到达一个叶子结点。

[0107] 2) 扩展(Expansion):如果当前叶子结点不是终止节点,那么就创建一个或者更多的子节点,选择其中一个进行扩展。

[0108] 3) 模拟(Simulation):从扩展节点开始,运行一个模拟的输出。

[0109] 4) 回溯(Back Propagation):在模拟结束之后,它的父节点以及从根节点到目标节点上的路径上的所有节点都会根据本次模拟的结果来添加自己的累计评分。

[0110] 可选地,根据每个节点的平均回报值和累计回报值及访问次数确定出候选节点的过程为确定节点选择策略的过程。

[0111] 其中,可以通过上界置信算法(UCB算法)来确定候选节点,UCB算法的运用公式如下:

$$[0112] \quad \text{UCB1}(n_i) = \bar{V}_i + C \sqrt{\frac{\log N_{\text{sim}}(n_0)}{N(n_i)}} = \frac{Q_i}{N(n_i)} + C \sqrt{\frac{\log N_{\text{sim}}(n_0)}{N(n_i)}}$$

[0113] $\text{argmax}_1(n_i), s.t. n_i \in \text{children}_{n_0}$

[0114] 其中 n_0 表示根节点, n_i 表示当前节点, V 表示当前目标树节点平均回报值, Q 表示当前节点累计回报值, $N(n_i)$ 表示当前节点被探索或者访问的次数, $N(n_0)$ 表示总的访问次数,等价于根节点的访问次数, C 是一个常量参数,可以控制访问和探索权重。

[0115] 可选地,根据平均回报率、累计回报率以及访问次数从目标树结构的多个节点中确定至少一个候选节点的过程还包括,预设确定条件可以为UCB的值,其中,可以将UCB的值与经验数值相比较进行判断,也可以将挑选出多个节点中UCB值最高的节点作为候选节点,在此不作进一步地限定。

[0116] 作为一种可选的实施方式,在基于目标树结构及回报数据确定至少一个候选节点之前,上述方法还包括:对目标车辆的控制信息进行离散化处理,得到离散后的控制信息,确定目标树结构中每个节点所对应的状态信息,基于离散后的控制信息以及每个节点所对应的状态信息构建目标树结构。

[0117] 在该实施例中,在基于目标树结构及回报数据确定至少一个候选节点之前,还包括以下步骤,自动驾驶系统可以离散化目标车辆的控制信息以得到进行离散化处理后的控制信息,确定目标树结构中每个节点所对应的状态信息,再基于离散后的控制变量和节点对应的状态信息构建出目标树结构,从而为候选节点的选择提供了目标树的结构,达到了将目标树结构应用于自动驾驶决策规划的技术效果。

[0118] 可选地,对目标车辆的控制信息进行离散可以包括加速度离散和方向盘转角离散;

[0119] 其中,加速度离散的初始值为0,步长为 δa (例如: 0.5m/s^2),最大加速为 a_{\max} ,最大减速度为 dec_{\max} ,则可以通过上述变量得到加速度的离散序列为:

[0120] $[\text{dec}_{\max}, \text{ec}_{\max} + \delta a, \text{dec}_{\max} + 2\delta a, \dots, \delta a, 0, a, 2\delta a, \dots, a_{\max}]$

[0121] $[-2\delta a, a_{\max}]$

[0122] 其中,方向盘转角离散的初始值为0,步长为 $\delta \omega$,最大转角为 ω_{\max} ,

[0123] 则可以通过上述变量得到方向盘转角的离散序列为:

[0124] $[-\omega_{\max}, \dots, -\omega_{\max} + \delta \omega, \dots, -\delta \omega, 0, \delta \omega, \omega_{\max} - \delta \omega, \dots, \omega_{\max}]$

[0125] 可选地,确定的节点信息可以包含时间,位置(x,y),航向角,速度,加速度,方向盘转角等状态信息,在此不作进一步的限定。

[0126] 可选地,确定目标树结构中每个节点所对应的状态信息的方式可以为逐层遍历目标树结构中的每个节点,通过预设的变量对每个节点进行筛选,最终遍历到自动驾驶系统控制过程中所需的节点状态信息。

[0127] 需要说明的是,上述内容仅为举例说明,此处不对确定目标树结构中每个节点所对应的状态信息的具体实现方式进行限定,其他能够确定目标树结构中每个节点所对应的状态信息的实现方式均在本公开的保护范围内。

[0128] 作为一种可选的实施方式,在基于目标树结构及回报数据确定至少一个候选节点之前,上述方法还包括:根据当前环境信息以及导航行驶轨迹确定初始控制信息,对初始控制信息进行离散化处理,得到离散后的控制信息,确定目标树结构中每个节点所对应的状态信息,基于离散后的控制信息以及每个节点所对应的状态信息构建所述目标树结构。

[0129] 在该实施例中,在基于目标树结构及回报数据确定至少一个候选节点之前,还包括以下步骤,自动驾驶系统可以根据宏观决策目标车辆当前所处的环境信息和目标导航行驶轨迹确定最开始的控制信息,并对控制信息进行离散处理确定目标树结构中每个节点所对应的状态信息,再基于离散后的控制变量和节点对应的状态信息构建出目标树结构,实现了在基于目标树结构确定候选节点之前,先排除了部分节点的可能性,从而可以降低搜

索耗时,提高了搜索效率。

[0130] 可选地,自动驾驶系统确定最初控制信息的过程可以包括:自动驾驶系统可以基于宏观决策并根据目标车辆导航路径以及周围实时障碍物的信息来确定出宏观决策结果,例如,车辆保持车道不变(车道中心线保持、向左偏离车道中心线绕行、向右偏离车道中心线绕行)、车辆变换车道(左变道、右变道)。

[0131] 可选地,具体的判断宏观决策的方法可以基于规则或者基于神经网络来完成,只要是能输出决策的模型即可,在此不作具体限定。

[0132] 可选地,在自动驾驶系统确定了宏观决策后(例如左变道),则可以结合上述步骤(基于蒙特卡洛树(MCTS)搜索求解最优轨迹),从目标车辆当前位置开始,在左变道的场景下对目标车辆加速度和方向盘转角进行离散处理,再通过蒙特卡洛树搜索算法搜索出最优决策序列,从而可以使目标树结构在搜索时极大的缩小了搜索空间(例如不会去搜索直行、右侧绕行、右侧变道空间等情况的节点),达到了降低了搜索复杂度的技术效果。

[0133] 图4是根据本公开实施例的一种控制车辆行驶的装置示意图,如图4所示,该控制车辆行驶的装置400可以包括:获取模块401、第一确定模块402、第二确定模块403、第三确定模块404和控制模块405。

[0134] 获取模块401,用于获取目标车辆的导航行驶轨迹,以及所述目标车辆所处当前环境的当前环境信息,其中,所述当前环境信息至少包括:所述当前环境对应的交通信息、所述目标车辆与所述当前环境中的障碍车辆之间的交互信息;

[0135] 第一确定模块402,用于根据所述当前环境信息确定所述目标车辆与所述障碍车辆之间进行交互的回报数据,其中,所述回报数据表征了所述目标车辆在所述交互场景下的驾驶性能;

[0136] 第二确定模块403,用于基于目标树结构及所述回报数据确定至少一个候选节点,其中,所述目标树结构由多个节点组成,每个节点表征了所述目标车辆在每个时刻的控制量;

[0137] 第三确定模块404,用于基于所述至少一个候选节点确定控制所述目标车辆行驶的目标控制序列;

[0138] 控制模块405,用于根据所述目标控制序列控制所述目标车辆按照所述导航行驶轨迹行驶。

[0139] 可选地,第一确定模块402包括:第一构建单元,用于根据当前环境信息构建目标车辆与障碍车辆之间进行交互的状态转移函数;第一确定单元,用于根据状态转移函数确定回报数据。

[0140] 可选地,构建单元包括:第一确定子单元,用于根据当前环境信息确定目标车辆在多个时刻的第一状态信息、第一控制信息以及障碍车辆在多个时刻的第二状态信息、第二控制信息;第二确定子单元,用于根据第一状态信息、第一控制信息、第二状态信息以及第二控制信息确定目标车辆与障碍车辆之间的联合状态信息;构建子单元,用于基于联合状态信息构建目标车辆与障碍车辆之间的状态转移函数。

[0141] 可选地,第一确定单元包括:获取子单元,用于获取目标车辆对应的专家轨迹信息,其中,专家轨迹信息表征了目标对象驾驶目标车辆按照导航行驶轨迹行驶的过程中,目标车辆的状态信息和控制信息;第三确定子单元,用于确定专家轨迹信息相对于预设变量

的概率分布函数；计算子单元，用于计算概率分布函数对应的似然函数；第四确定子单元，用于根据似然函数的梯度值确定预设变量；第五确定子单元，用于根据第一状态信息、第一控制信息、第二状态信息以及第二控制信息以及预设变量确定目标车辆对应的回报数据。

[0142] 可选地，第二确定模块403包括：计算单元，用于基于回报数据计算目标树结构中每个节点所对应的平均回报值以及累计回报值；统计单元，用于统计目标树结构中每个节点的访问次数；第二确定单元，用于根据平均回报值、累计回报值以及访问次数从目标树结构的多个节点中确定至少一个候选节点。

[0143] 可选地，第二确定模块403还包括：第一离散单元，用于对目标车辆的控制信息进行离散化处理，得到离散后的控制信息；第三确定单元，用于确定目标树结构中每个节点所对应的状态信息；第二构建单元，用于基于离散后的控制信息以及每个节点所对应的状态信息构建目标树结构。

[0144] 可选地，第二确定模块403还包括：第四确定单元，用于根据当前环境信息以及导航行驶轨迹确定初始控制信息；第二离散单元，用于对初始控制信息进行离散化处理，得到离散后的控制信息；第五确定单元，用于确定目标树结构中每个节点所对应的状态信息；第三构建单元，用于基于所述离散后的控制信息以及所述每个节点所对应的状态信息构建所述目标树结构。

[0145] 本公开的技术方案中，所涉及的用户个人信息的获取，存储和应用等，均符合相关法律法规的规定，且不违背公序良俗。

[0146] 根据本公开的实施例，本公开还提供了一种电子设备、一种可读存储介质和一种计算机程序产品。

[0147] 图5示出了可以用来实施本公开的实施例的示例电子设备500的示意性框图。电子设备旨在表示各种形式的数字计算机，诸如，膝上型计算机、台式计算机、工作台、个人数字助理、服务器、刀片式服务器、大型计算机、和其它适合的计算机。电子设备还可以表示各种形式的移动装置，诸如，个人数字处理、蜂窝电话、智能电话、可穿戴设备和其它类似的计算装置。本文所示的部件、它们的连接和关系、以及它们的功能仅作为示例，并且不意在限制本文中描述的和/或者要求的本公开的实现。

[0148] 如图5所示，设备500包括计算单元501，其可以根据存储在只读存储器(ROM) 502中的计算机程序或者从存储单元508加载到随机访问存储器(RAM) 503中的计算机程序，来执行各种适当的动作和处理。在RAM 503中，还可存储设备500操作所需的各种程序和数据。计算单元501、ROM 502以及RAM 503通过总线804彼此相连。输入/输出(I/O)接口505也连接至总线504。

[0149] 设备800中的多个部件连接至I/O接口505，包括：输入单元506，例如键盘、鼠标等；输出单元507，例如各种类型的显示器、扬声器等；存储单元508，例如磁盘、光盘等；以及通信单元509，例如网卡、调制解调器、无线通信收发机等。通信单元509允许设备500通过诸如因特网的计算机网络和/或各种电信网络与其他设备交换信息/数据。

[0150] 计算单元501可以是各种具有处理和计算能力的通用和/或专用处理组件。计算单元501的一些示例包括但不限于中央处理单元(CPU)、图形处理单元(GPU)、各种专用的人工智能(AI)计算芯片、各种运行机器学习模型算法的计算单元、数字信号处理器(DSP)、以及任何适当的处理器、控制器、微控制器等。计算单元501执行上文所描述的各个方法和处理，

例如控制车辆行驶的方法。例如,在一些实施例中,控制车辆行驶的方法可被实现为计算机软件程序,其被有形地包含于机器可读介质,例如存储单元508。在一些实施例中,计算机程序的部分或者全部可以经由ROM 502和/或通信单元509而被载入和/或安装到设备500上。当计算机程序加载到RAM 503并由计算单元501执行时,可以执行上文描述的控制车辆行驶的方法的一个或多个步骤。备选地,在其他实施例中,计算单元501可以通过其他任何适当的方式(例如,借助于固件)而被配置为执行控制车辆行驶的方法。

[0151] 本文中以上描述的系统和技术各种实施方式可以在数字电子电路系统、集成电路系统、场可编程门阵列(FPGA)、专用集成电路(ASIC)、专用标准产品(ASSP)、芯片上系统的系统(SOC)、负载可编程逻辑设备(CPLD)、计算机硬件、固件、软件、和/或它们的组合中实现。这些各种实施方式可以包括:实施在一个或者多个计算机程序中,该一个或者多个计算机程序可在包括至少一个可编程处理器的可编程系统上执行和/或解释,该可编程处理器可以是专用或者通用可编程处理器,可以从存储系统、至少一个输入装置、和至少一个输出装置接收数据和指令,并且将数据和指令传输至该存储系统、该至少一个输入装置、和该至少一个输出装置。

[0152] 用于实施本公开的方法的程序代码可以采用一个或多个编程语言的任何组合来编写。这些程序代码可以提供给通用计算机、专用计算机或其他可编程数据处理装置的处理单元或控制器,使得程序代码当由处理单元或控制器执行时使流程图和/或框图中所规定的功能/操作被实施。程序代码可以完全在机器上执行、部分地在机器上执行,作为独立软件包部分地在机器上执行且部分地在远程机器上执行或完全在远程机器或服务器上执行。

[0153] 在本公开的上下文中,机器可读介质可以是有形的介质,其可以包含或存储以供指令执行系统、装置或设备使用或与指令执行系统、装置或设备结合地使用的程序。机器可读介质可以是机器可读信号介质或机器可读储存介质。机器可读介质可以包括但不限于电子的、磁性的、光学的、电磁的、红外的、或半导体系统、装置或设备,或者上述内容的任何合适组合。机器可读储存介质的更具体示例会包括基于一个或多个线的电气连接、便携式计算机盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦除可编程只读存储器(EPROM或快闪存储器)、光纤、便捷式紧凑盘只读存储器(CD-ROM)、光学储存设备、磁储存设备、或上述内容的任何合适组合。

[0154] 为了提供与用户的交互,可以在计算机上实施此处描述的系统和技术,该计算机具有:用于向用户显示信息的显示装置(例如,CRT(阴极射线管)或者LCD(液晶显示器)监视器);以及键盘和指向装置(例如,鼠标或者轨迹球),用户可以通过该键盘和该指向装置来将输入提供给计算机。其它种类的装置还可以用于提供与用户的交互;例如,提供给用户的反馈可以是任何形式的传感反馈(例如,视觉反馈、听觉反馈、或者触觉反馈);并且可以用任何形式(包括声输入、语音输入或者、触觉输入)来接收来自用户的输入。

[0155] 可以将此处描述的系统和技术实施在包括后台部件的计算系统(例如,作为数据服务器)、或者包括中间件部件的计算系统(例如,应用服务器)、或者包括前端部件的计算系统(例如,具有图形用户界面或者网络浏览器的用户计算机,用户可以通过该图形用户界面或者该网络浏览器来与此处描述的系统和技术实施方式交互)、或者包括这种后台部件、中间件部件、或者前端部件的任何组合的计算系统中。可以通过任何形式或者介质的数字数据通信(例如,通信网络)来将系统的部件相互连接。通信网络的示例包括:局域网

(LAN)、广域网(WAN)和互联网。

[0156] 计算机系统可以包括客户端和服务端。客户端和服务端一般远离彼此并且通常通过通信网络进行交互。通过在相应的计算机上运行并且彼此具有客户端-服务器关系的计算机程序来产生客户端和服务端的关系。服务器可以是云服务器,也可以为分布式系统的服务器,或者是结合了区块链的服务器。

[0157] 应该理解,可以使用上面所示的各种形式的流程,重新排序、增加或删除步骤。例如,本发公开中记载的各步骤可以并行地执行也可以顺序地执行也可以不同的次序执行,只要能够实现本公开公开的技术方案所期望的结果,本文在此不进行限制。

[0158] 上述具体实施方式,并不构成对本公开保护范围的限制。本领域技术人员应该明白的是,根据设计要求和因素,可以进行各种修改、组合、子组合和替代。任何在本公开的精神和原则之内所作的修改、等同替换和改进等,均应包含在本公开保护范围之内。

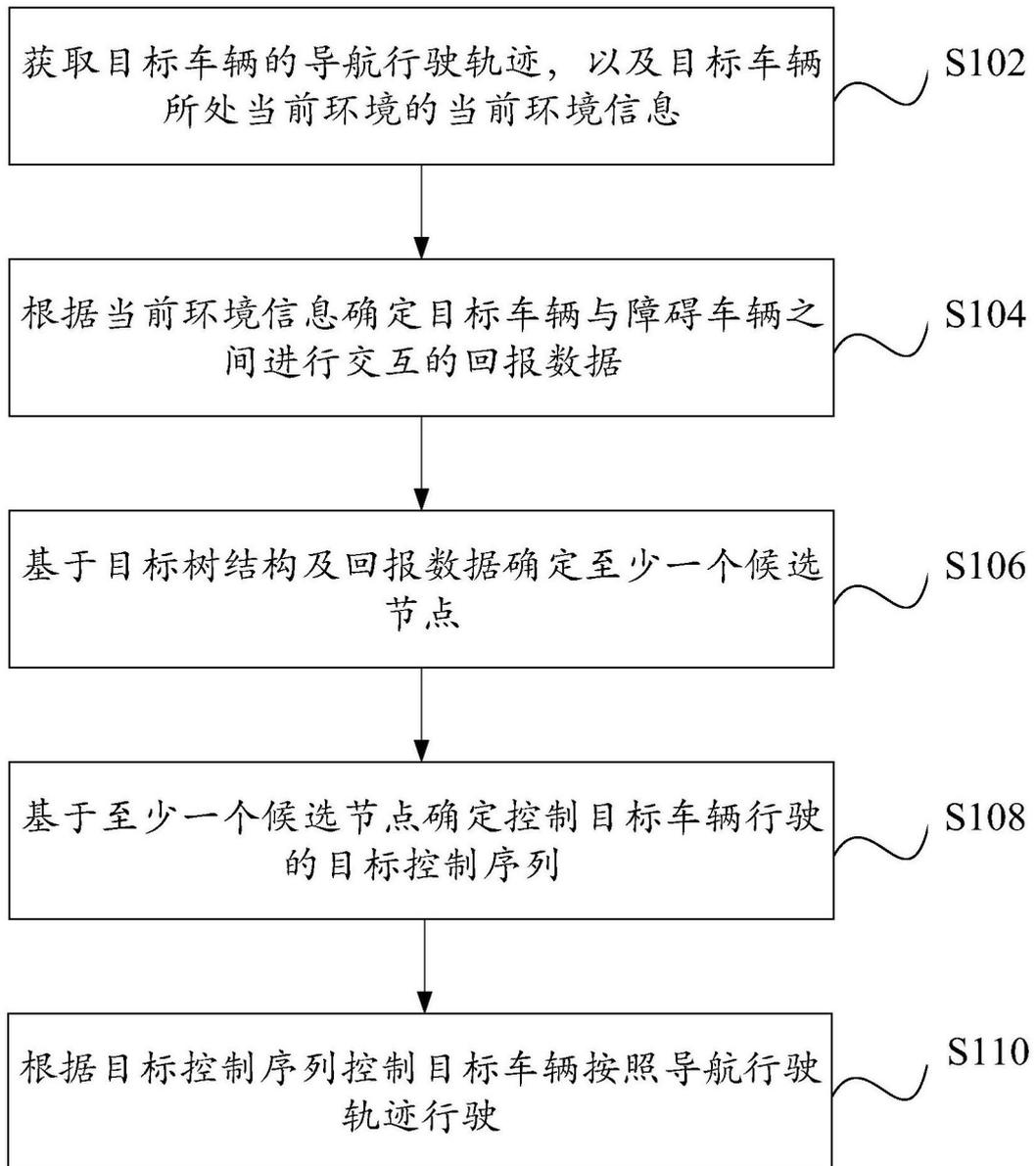


图1

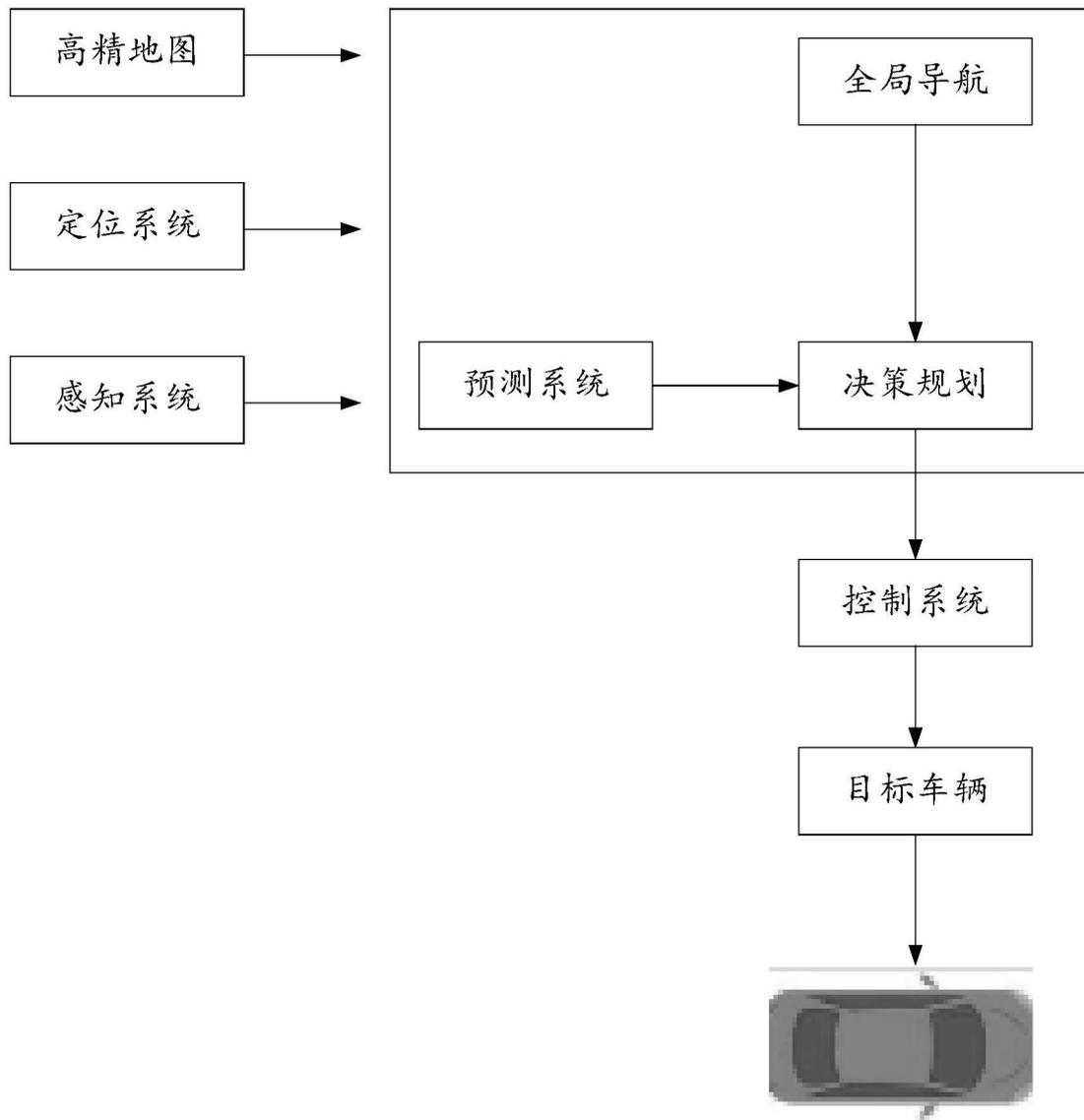


图2

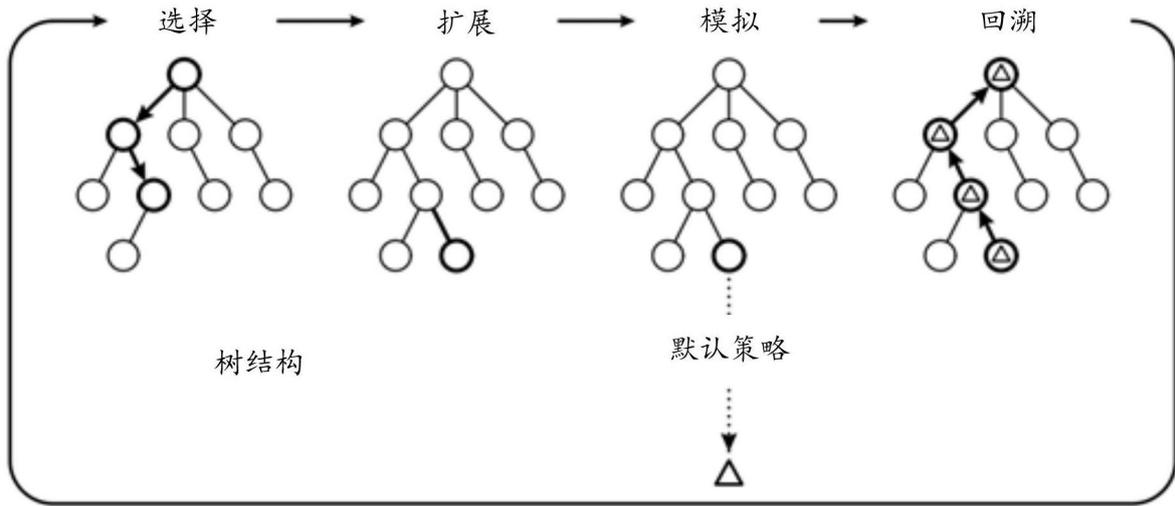


图3

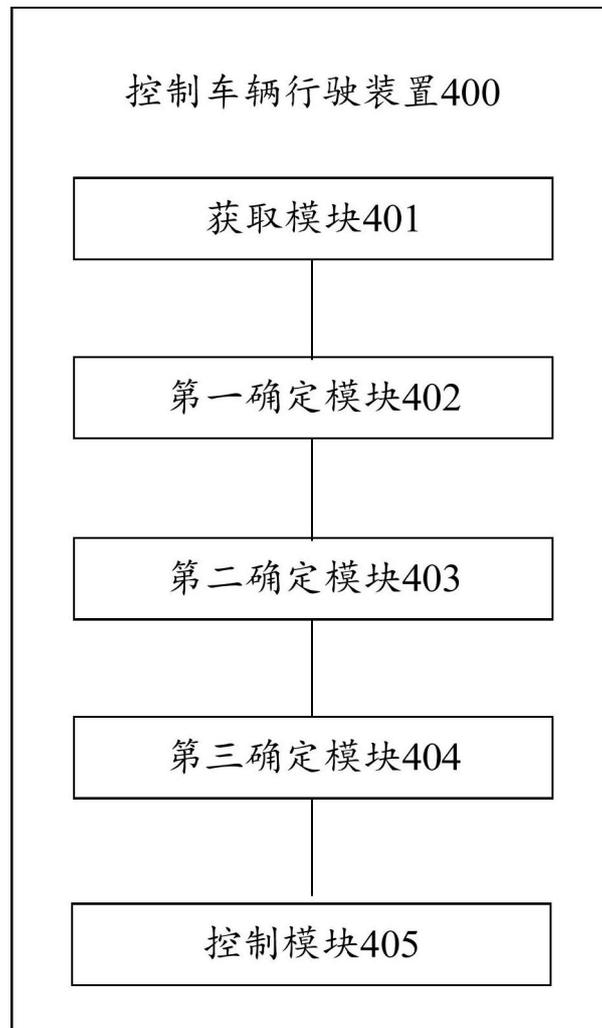


图4

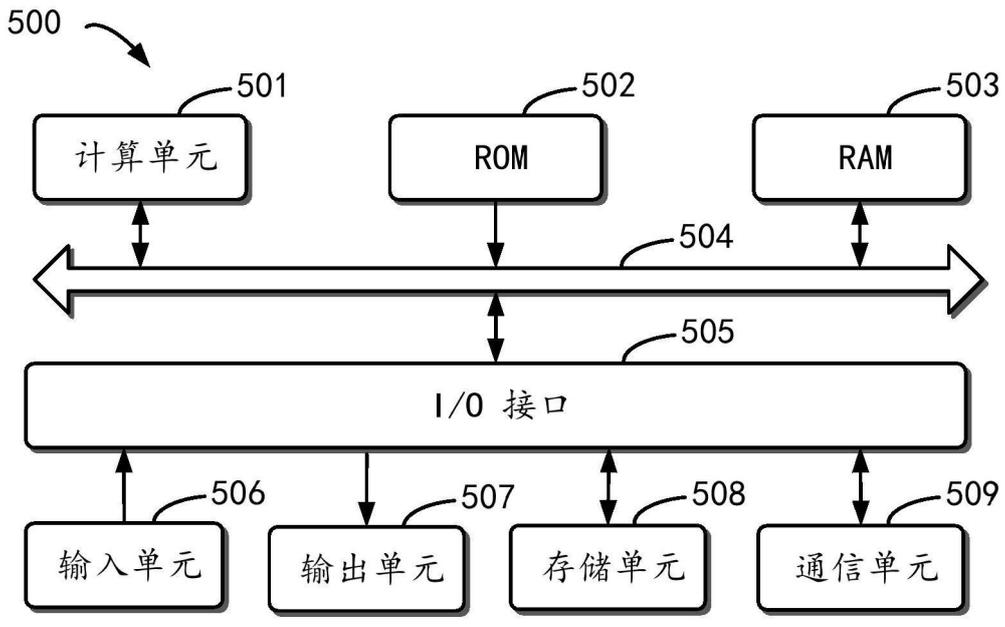


图5