



(12) 发明专利

(10) 授权公告号 CN 116552474 B

(45) 授权公告日 2024. 02. 06

(21) 申请号 202310586111.4

(22) 申请日 2023.05.23

(65) 同一申请的已公布的文献号

申请公布号 CN 116552474 A

(43) 申请公布日 2023.08.08

(73) 专利权人 南栖仙策(南京)高新技术有限公司

地址 210000 江苏省南京市江宁区麒麟科技
技创新园天骄路100号江苏南京侨梦苑A栋703室

(72) 发明人 汪娟 周俊杰

(74) 专利代理机构 北京品源专利代理有限公司
11332

专利代理师 赵迎迎

(51) Int.Cl.

B60T 8/32 (2006.01)

B60T 8/34 (2006.01)

B60T 8/172 (2006.01)

B60T 8/176 (2006.01)

(56) 对比文件

JP H05294249 A, 1993.11.09

CN 101311047 A, 2008.11.26

CN 115973179 A, 2023.04.18

CN 116009397 A, 2023.04.25

US 2021171004 A1, 2021.06.10

审查员 吕学昭

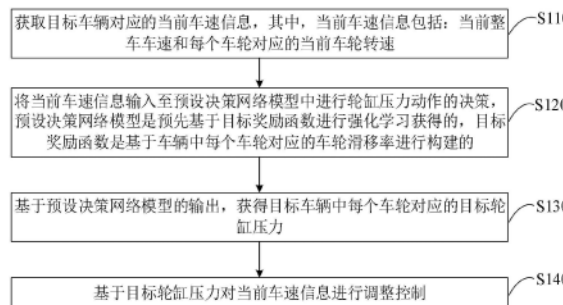
权利要求书2页 说明书11页 附图2页

(54) 发明名称

一种基于强化学习的车速控制方法、装置、设备和介质

(57) 摘要

本发明公开了一种基于强化学习的车速控制方法、装置、设备和介质。该方法包括:获取目标车辆对应的当前车速信息,其中,当前车速信息包括:当前整车车速和每个车轮对应的当前车轮转速;将当前车速信息输入至预设决策网络模型中进行轮缸压力动作的决策,预设决策网络模型是预先基于目标奖励函数进行强化学习获得的,目标奖励函数是基于车辆中每个车轮对应的车轮滑移率进行构建的;基于预设决策网络模型的输出,获得目标车辆中每个车轮对应的目标轮缸压力;基于目标轮缸压力对当前车速信息进行调整控制。通过本公开实施例的技术方案,可以有效控制车速,从而保持车辆制动或加速时的稳定性,进一步提高车辆行驶的安全性和用户的驾驶体验。



1. 一种基于强化学习的车速控制方法,其特征在于,包括:

获取目标车辆对应的当前车速信息,其中,所述当前车速信息包括:当前整车车速和每个车轮对应的当前车轮转速;

将所述当前车速信息输入至预设决策网络模型中进行轮缸压力动作的决策,所述预设决策网络模型是预先基于目标奖励函数进行强化学习获得的,所述目标奖励函数是基于车辆中每个车轮对应的车轮滑移率进行构建的;

基于所述预设决策网络模型的输出,获得所述目标车辆中每个车轮对应的目标轮缸压力;

基于所述目标轮缸压力对所述当前车速信息进行调整控制;

其中,基于车辆中每个车轮对应的车轮滑移率构建所述目标奖励函数,包括:基于车辆中每个车轮对应的车轮滑移率,确定每个车轮对应的车轮奖励函数;所述每个车轮对应的车轮奖励函数是关于车轮滑移率的分段奖励函数;基于各个所述车轮奖励函数,构建所述目标奖励函数;

所述基于车辆中每个车轮对应的车轮滑移率,确定每个车轮对应的车轮奖励函数,包括:若车辆中的当前车轮对应的当前车轮滑动率小于预设最小滑动率,则基于第一权重值对当前车轮滑动率进行加权,获得当前车轮对应的第一分段奖励函数;若当前车轮滑动率大于或等于所述预设最小滑动率,且小于或等于预设最大滑动率,则将当前车轮滑动率作为当前车轮对应的第二分段奖励函数;若当前车轮滑动率大于所述预设最大滑动率,则基于第二权重值对当前车轮滑动率进行加权,获得当前车轮对应的第三分段奖励函数;其中,所述第一权重值小于所述第二权重值。

2. 根据权利要求1所述的方法,其特征在于,基于各个所述车轮奖励函数,构建目标奖励函数,包括:

将各个所述车轮奖励函数进行相加,并将相加结果确定为目标奖励函数。

3. 根据权利要求1所述的方法,其特征在于,基于所述目标轮缸压力对所述当前车速信息进行调整控制,包括:

将每个车轮对应的轮缸压力调整为相应的目标轮缸压力,使得所述目标车辆的当前整车速度调整为所述目标轮缸压力对应的目标整车速度。

4. 根据权利要求1所述的方法,其特征在于,所述预设决策网络模型的强化学习过程,包括:

将样本车辆的样本车速信息输入至待训练的预设决策网络模型中进行轮缸压力动作的决策;

基于待训练的预设决策网络模型输出的样本轮缸压力,对所述样本车速信息进行调整控制,获得控制后的样本车速信息;

基于目标奖励函数和所述控制后的样本车速信息,确定所述样本轮缸压力对应的目标奖励值;

基于所述目标奖励值,调整待训练的预设决策网络模型中的网络参数,直到满足预设收敛条件时训练结束,获得训练结束后的预设决策网络模型。

5. 一种基于强化学习的车速控制装置,其特征在于,包括:

当前车速信息获取模块,用于获取目标车辆对应的当前车速信息,其中,所述当前车速

信息包括:当前整车车速和每个车轮对应的当前车轮转速;

第一动作决策模块,用于将所述当前车速信息输入至预设决策网络模型中进行轮缸压力动作的决策,所述预设决策网络模型是预先基于目标奖励函数进行强化学习获得的,所述目标奖励函数是基于车辆中每个车轮对应的车轮滑移率进行构建的;

目标轮缸压力获得模块,用于基于所述预设决策网络模型的输出,获得所述目标车辆中每个车轮对应的目标轮缸压力;

当前车速信息调整模块,用于基于所述目标轮缸压力对所述当前车速信息进行调整控制;

目标奖励函数构建模块,用于基于车辆中每个车轮对应的车轮滑移率构建所述目标奖励函数;

其中,所述目标奖励函数构建模块,包括:

车轮奖励函数确定子模块,用于基于车辆中每个车轮对应的车轮滑移率,确定每个车轮对应的车轮奖励函数;所述每个车轮对应的车轮奖励函数是关于车轮滑移率的分段奖励函数;

目标奖励函数构建子模块,用于基于各个所述车轮奖励函数,构建所述目标奖励函数;

其中,所述车轮奖励函数确定子模块具体用于:若车辆中的当前车轮对应的当前车轮滑动率小于预设最小滑动率,则基于第一权重值对当前车轮滑动率进行加权,获得当前车轮对应的第一分段奖励函数;若当前车轮滑动率大于或等于所述预设最小滑动率,且小于或等于预设最大滑动率,则将当前车轮滑动率作为当前车轮对应的第二分段奖励函数;若当前车轮滑动率大于所述预设最大滑动率,则基于第二权重值对当前车轮滑动率进行加权,获得当前车轮对应的第三分段奖励函数;其中,所述第一权重值小于所述第二权重值。

6.一种电子设备,其特征在于,所述电子设备包括:

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计算机程序,所述计算机程序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1-4中任一项所述的基于强化学习的车速控制方法。

7.一种计算机可读存储介质,其特征在于,所述计算机可读存储介质存储有计算机指令,所述计算机指令用于使处理器执行时实现权利要求1-4中任一项所述的基于强化学习的车速控制方法。

一种基于强化学习的车速控制方法、装置、设备和介质

技术领域

[0001] 本发明涉及计算机技术领域,尤其涉及一种基于强化学习的车速控制方法、装置、设备和介质。

背景技术

[0002] 随着汽车的发展,用户越来越注重车辆行驶的安全性。汽车的制动能力是影响车辆行驶安全性的重要因素之一。滑移率可用于表征车辆的制动能力。滑移率是指在刹车或加速时车轮和路面间所产生的滑移距离与车辆移动距离之间的比值。车辆在抓地性最佳的情况下仍会存在5%-10%的滑移率,例如,车轮转动了100m的距离时车子只移动了90m-95m。

[0003] 目前,通常是利用汽车防抱死制动系统(Antilock Brake System,ABS)对车辆制动力进行控制。例如,ABS在汽车制动时防止车轮抱死,从而避免前轮和/或后轮发生侧滑,进而保持制动时的方向稳定性。在车辆进行制动或加速时,ABS会基于车辆出厂时标定的车轮线性特性和该车轮线性特性对应的滑动率控制方式对车速进行控制。然而,车辆在被使用后,每个车轮的磨损程度是不同的,还会存在某个车轮换新的情况,导致了车轮特性发生了非线性变化。此时,仍基于车辆出厂时标定的车轮线性特性和该车轮线性特性对应的滑动率控制方式对车速进行控制,会导致车轮的滑动率变大,并且无法保持车辆制动或加速时的方向稳定性,降低了车辆行驶的安全性和用户的驾驶体验。

发明内容

[0004] 本发明提供了一种基于强化学习的车速控制方法、装置、设备和介质,以有效控制车速,从而保持车辆制动或加速时的稳定性,进一步提高车辆行驶的安全性和用户的驾驶体验。

[0005] 根据本发明的一方面,提供了一种基于强化学习的车速控制方法,该方法包括:

[0006] 获取目标车辆对应的当前车速信息,其中,所述当前车速信息包括:当前整车车速和每个车轮对应的当前车轮转速;

[0007] 将所述当前车速信息输入至预设决策网络模型中进行轮缸压力动作的决策,所述预设决策网络模型是预先基于目标奖励函数进行强化学习获得的,所述目标奖励函数是基于车辆中每个车轮对应的车轮滑移率进行构建的;

[0008] 基于所述预设决策网络模型的输出,获得所述目标车辆中每个车轮对应的目标轮缸压力;

[0009] 基于所述目标轮缸压力对所述当前车速信息进行调整控制。

[0010] 根据本发明的另一方面,提供了一种基于强化学习的车速控制装置,该装置包括:

[0011] 当前车速信息获取模块,用于获取目标车辆对应的当前车速信息,其中,所述当前车速信息包括:当前整车车速和每个车轮对应的当前车轮转速;

[0012] 动作决策模块,用于将所述当前车速信息输入至预设决策网络模型中进行轮缸压

力动作的决策,所述预设决策网络模型是预先基于目标奖励函数进行强化学习获得的,所述目标奖励函数是基于车辆中每个车轮对应的车轮滑移率进行构建的;

[0013] 目标轮缸压力获得模块,用于基于所述预设决策网络模型的输出,获得所述目标车辆中每个车轮对应的目标轮缸压力;

[0014] 当前车速信息调整模块,用于基于所述目标轮缸压力对所述当前车速信息进行调整控制。

[0015] 根据本发明的另一方面,提供了一种电子设备,所述电子设备包括:

[0016] 至少一个处理器;以及

[0017] 与所述至少一个处理器通信连接的存储器;其中,

[0018] 所述存储器存储有可被所述至少一个处理器执行的计算机程序,所述计算机程序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本发明任一实施例所述的基于强化学习的车速控制方法。

[0019] 根据本发明的另一方面,提供了一种计算机可读存储介质,所述计算机可读存储介质存储有计算机指令,所述计算机指令用于使处理器执行时实现本发明任一实施例所述的基于强化学习的车速控制方法。

[0020] 本发明实施例的技术方案,通过获取目标车辆对应的当前车速信息,以便基于获取的当前车速信息,确定出当前时刻目标车辆中每个车轮对应的调整策略,从而对比上一时刻和当前时刻目标车辆中每个车轮对应的调整策略。其中,所述当前车速信息包括:当前整车车速和每个车轮对应的当前车轮转速;将所述当前车速信息输入至预设决策网络模型中进行轮缸压力动作的决策,所述预设决策网络模型是预先基于目标奖励函数进行强化学习获得的,所述目标奖励函数是基于车辆中每个车轮对应的车轮滑移率进行构建的;基于所述预设决策网络模型的输出,获得所述目标车辆中每个车轮对应的目标轮缸压力;基于所述目标轮缸压力对所述当前车速信息进行调整控制,使得每个车轮不会出现抱死的情况,且可以将车轮的滑移率控制在最优滑移率的范围内,从而有效控制车速,并有效缩短车辆的制动距离,进而保持车辆制动或加速时的稳定性,进一步提高车辆行驶的安全性和用户的驾驶体验。

[0021] 应当理解,本部分所描述的内容并非旨在标识本发明的实施例的关键或重要特征,也不用于限制本发明的范围。本发明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

[0022] 为了更清楚地说明本发明实施例中的技术方案,下面将对实施例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发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0023] 图1是根据本发明实施例一提供的一种基于强化学习的车速控制方法的流程图;

[0024] 图2是根据本发明实施例二提供的另一种基于强化学习的车速控制方法的流程图;

[0025] 图3是根据本发明实施例三提供的一种基于强化学习的车速控制装置的结构示意图;

[0026] 图4是实现本发明实施例的基于强化学习的车速控制方法的电子设备的结构示意图。

具体实施方式

[0027] 为了使本技术领域的人员更好地理解本发明方案,下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明一部分的实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都应当属于本发明保护的范围。

[0028] 需要说明的是,本发明的说明书和权利要求书及上述附图中的术语“第一”、“第二”等是用于区别类似的对象,而不必用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换,以便这里描述的本发明的实施例能够以除了在这里图示或描述的那些以外的顺序实施。此外,术语“包括”和“具有”以及他们的任何变形,意图在于覆盖不排他的包含,例如,包含了一系列步骤或单元的过程、方法、系统、产品或设备不必限于清楚地列出的那些步骤或单元,而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它步骤或单元。

[0029] 实施例一

[0030] 图1为本发明实施例一提供了一种基于强化学习的车速控制方法的流程图,本实施例可适用于对车辆的车速进行调整控制的情况,尤其适用于在车辆加速或制动时对车辆的车速进行调整控制的情况。该方法可以由基于强化学习的车速控制装置来执行,该基于强化学习的车速控制装置可以采用硬件和/或软件的形式实现,该基于强化学习的车速控制装置可配置于电子设备中。如图1所示,该方法包括:

[0031] S110、获取目标车辆对应的当前车速信息,其中,当前车速信息包括:当前整车车速和每个车轮对应的当前车轮转速。

[0032] 其中,目标车辆可以是指带有ABS需要进行车速控制的车辆。例如,目标车辆可以是但不限于四轮轿车、六轮客车或十轮货车。目标车辆可以是用户所使用的车,也可以是正在进行ABS调试并没有出厂的车,还可以是车辆动态模拟与分析软体工具(Mechanical Simulation Corporation,carsim/TrucksimMSC)中构建的仿真车辆。Carsim可以用于模拟车辆的实际运动状态,并输出车辆的整车速度和车轮转速等信息。当前整车车速可以由每个车轮对应的当前车轮转速和每个车轮对应的当前车轮滑移率确定的。当前整车车速还可以是由被动轮(不出动力的车轮)的当前车轮转速与该车轮半径确定的。

[0033] 具体地,在目标车辆行驶的过程中,可以实时获取目标车辆对应的当前车速信息,以便基于获取的当前车速信息,确定出当前时刻目标车辆中每个车轮对应的调整策略,从而对比上一时刻和当前时刻目标车辆中每个车轮对应的调整策略。若两个时刻确定出的调整决策一致,则延用上一时刻目标车辆中每个车轮对应的调整策略继续对每个车轮进行调整。若两个时刻确定出的调整策略存在不一致,则使用当前时刻确定出的新的调整策略对每个车轮进行调整控制,从而实现对目标车辆中每个车轮的动态调整,同时有效控制车速,进而从而保持车辆制动或加速时的稳定性,提高车辆行驶的安全性和用户的驾驶体验。其中,可以基于车速传感器获得目标车辆对应的当前整车车速。可以基于轮速传感器获得每

个车轮对应的当前车轮转速。也可以基于ABS传感器单位时间内获得的脉冲信号,确定出每个车轮对应的当前车轮转速。

[0034] S120、将当前车速信息输入至预设决策网络模型中进行轮缸压力动作的决策,预设决策网络模型是预先基于目标奖励函数进行强化学习获得的,目标奖励函数是基于车辆中每个车轮对应的车轮滑移率进行构建的。

[0035] 其中,强化学习可以由智能体 (Agent)、环境 (Environment)、状态 (State)、动作 (Action) 和奖励 (Reward) 组成。本实施例中,目标车辆中每个轮胎可以作为一个智能体。S110中目标车辆对应的当前车速信息可以作为状态。每个车轮对应的轮缸压力可以作为动作或是动作集合。目标奖励函数可以是用于确定预设决策网络模型采取某个动作信息后获得的奖励值。奖励值越大,车辆执行该奖励值对应的动作后车辆越稳定。智能体执行了某个动作后,环境将会转换到一个新的状态,对于该新的状态环境会给出奖励信息 (正奖励或者负奖励),然后智能体根据新的状态和环境反馈的奖励信息,确定新的动作。智能体通过强化学习,可以知道自己在什么状态下,应该采取什么样的动作使得自身获得最大奖励。本实施例可以基于车辆中每个车轮对应的车轮滑移率构建目标奖励函数。

[0036] 示例性地,预设决策网络模型的网络架构可以包括但不限于:长短期记忆网络 LSTM (Long Short-Term Memory)、残差网络 ResNet、深度神经网络 DNN (Deep Neural Networks)、卷积神经网络 CNN (Convolutional Neural Network) 或者全连接网络 FC (Fully Connection) 中的至少一种。预设决策网络模型的建立思想可以是为了控制车辆对应的滑移率在最优范围内。

[0037] 具体地,可以基于预设强化学习方式,比如基于策略的强化学习方式或者基于滑移率的强化学习方式,利用目标奖励函数对待训练的预设决策网络模型进行强化学习,将每个车轮对应的滑移率控制在最优滑移率的范围 (5%至10%) 内的奖励值,从而可以保持车辆制动或加速时的方向稳定性,进而提高车辆行驶的安全性和用户的驾驶体验。本实施例可以利用强化学习后获得的预设决策网络模型对目标车辆中每个车轮对应的轮缸压力进行动作决策,从而实现对目标车辆的当前车速信息进行调整控制。

[0038] S130、基于预设决策网络模型的输出,获得目标车辆中每个车轮对应的目标轮缸压力。

[0039] 其中,目标轮缸压力可以是指目标车辆中每个车轮在下一时刻需要达到的轮缸压力。每个车轮对应的目标轮缸压力可以是不同的,从而实现针对每个车轮的轮缸压力动态调整。

[0040] S140、基于目标轮缸压力对当前车速信息进行调整控制。

[0041] 具体地,将每个车轮对应的当前轮缸压力调整为目标轮缸压力,以使每个车轮不会出现抱死的情况,且可以将车轮的滑移率控制在最优滑移率的范围内,从而有效控制车速,进而保持车辆制动或加速时的稳定性,进一步提高车辆行驶的安全性和用户的驾驶体验。在车辆实际行驶过程中,可以利用当前时刻的当前整车车速、制动压力和目标轮缸压力,基于车辆在出厂前预设的车速和车轮转速的计算方式,确定出下一时刻车辆对应的整车车速和每个车轮对应的车轮转速。预设的车速和转速的计算方式不是本发明所涉及的重点,故不进行展开说明。

[0042] 本发明实施例的技术方案,通过获取目标车辆对应的当前车速信息,以便基于获

取的当前车速信息,确定出当前时刻目标车辆中每个车轮对应的调整策略,从而对比上一时刻和当前时刻目标车辆中每个车轮对应的调整策略。其中,当前车速信息包括:当前整车车速和每个车轮对应的当前车轮转速;将当前车速信息输入至预设决策网络模型中进行轮缸压力动作的决策,预设决策网络模型是预先基于目标奖励函数进行强化学习获得的,目标奖励函数是基于车辆中每个车轮对应的车轮滑移率进行构建的;基于预设决策网络模型的输出,获得目标车辆中每个车轮对应的目标轮缸压力;基于目标轮缸压力对当前车速信息进行调整控制,使得每个车轮不会出现抱死的情况,且可以将车轮的滑移率控制在最优滑移率的范围内,从而有效控制车速,并有效缩短车辆的制动距离,进而保持车辆制动或加速时的稳定性,进一步提高车辆行驶的安全性和用户的驾驶体验。

[0043] 在上述技术方案的基础上,S140可以包括:将每个车轮对应的轮缸压力调整为相应的目标轮缸压力,使得目标车辆的当前整车速度调整为目标轮缸压力对应的目标整车速度。

[0044] 具体地,将每个车轮对应的轮缸压力调整为相应的目标轮缸压力,使得每个车轮对应的滑移率保持在最优滑移率的范围内,并基于每个轮胎对应的目标轮缸压力和调整后的滑移率确定每个车轮对应的目标车轮转速,基于所有目标车轮转速,确定出目标车辆基于目标轮缸压力调整后的目标整车速度。

[0045] 实施例二

[0046] 图2为本发明实施例二提供的另一种基于强化学习的车速控制方法的流程图,本实施例在上述实施例的基础上,对目标奖励函数的构建过程进行详细描述。其中与上述各实施例相同或相应的术语的解释在此不再赘述。如图2所示,该方法包括:

[0047] S210、基于车辆中每个车轮对应的车轮滑移率构建目标奖励函数。

[0048] 其中,每个车轮对应的一个车轮奖励函数。目标奖励函数可以是基于所有车轮对应的车轮奖励函数进行整合后确定的。确定出每个车轮对应的车轮奖励函数,再基于所有车轮对应的车轮奖励函数确定出整车对应的目标奖励函数,从而可以针对每个车轮进行细节调整,从而促使整车随着每个车轮的调整而调整,避免了直接对整车进行调整出现较大的调整误差的情况,进而保证基于目标奖励函数进行强化学习训练后的预设决策网络模型的准确性。

[0049] S220、基于样本车速信息和目标奖励函数进行强化学习,获得的预设决策网络模型。

[0050] 其中,样本车速信息可以是获得的历史车辆的车速信息。样本车速信息可以包括样本整车车速和每个车轮对应的样本车轮转速。可以通过样本车速信息、滑移率和期望滑移率进行深度强化学习,实现对车轮的制动力矩(相当于轮缸压力)的优化,从而得到最准确的控制结果。

[0051] 示例性地,将样本车辆的样本车速信息输入至待训练的预设决策网络模型中进行轮缸压力动作的决策;基于待训练的预设决策网络模型输出的样本轮缸压力,对样本车速信息进行调整控制,获得控制后的样本车速信息;基于目标奖励函数和控制后的样本车速信息,确定样本轮缸压力对应的目标奖励值;基于目标奖励值,调整待训练的预设决策网络模型中的网络参数,直到满足预设收敛条件时训练结束,获得训练结束后的预设决策网络模型。

[0052] 其中, 滑移率可以是基于整车车速和车轮转速确定出来的。具体地, 将样本车辆的样本车速信息输入至待训练的预设决策网络模型中, 以使待训练的预设决策网络模型基于输入的样本车速信息确定每个车轮对应的车轮滑移率。以左前轮为例, 计算方式如下:

$$[0053] \quad S_{L1} = \frac{V - V_{L1}}{V}$$

[0054] 其中, S_{L1} 为左前轮的车轮滑移率, V 为样本整车车速, V_{L1} 为左前轮的车轮转速。可以基于确定出的滑移率, 确定待训练的预设决策网络模型要输出的调整后的车轮滑移率对应的样本轮缸压力。基于待训练的预设决策网络模型输出的样本轮缸压力, 对样本车速信息进行调整控制, 获得控制后的样本车速信息。例如, 在 carsim 仿真软件中设置对开路面场景, 其中, 样本车速为 V_0 , 制动压力为 P_m 。在 carsim 设置的场景中可以根据 V_0 、 P_m 以及预设轮缸压力确定模型确定出每个车轮对应的轮缸压力 P_{L1} 、 P_{R1} 、 P_{L2} 、 P_{R2} , 并通过内部积分运算确定下一时刻的整车车速、每个车轮对应的车轮转速。利用 carsim 和强化学习配合进行车辆仿真训练的好处在于, 无需在真实车辆中进行存在危险性的测试, 该方式能够在大量样本数据和复杂场景下进行高效学习, 从而有效提高制动系统的性能。通过对大量样本数据的学习和分析, 可以确定有效的制动策略, 使得车辆在加速或制动, 尤其是在紧急制动工况下具备更好的响应速度和减速效果。同时, 本强化学习方式, 充分考虑了轮胎纵向力与滑移率之间非线性的关系特性, 通过对这一关系的深入研究和精确建模, 能够实现对制动力 (相当于轮缸压力) 的精确控制。这有助于在紧急制动过程中减小滑移, 提高制动效果, 同时避免因过度制动导致的轮胎损伤和车辆失控的情况。基于控制后的样本车速信息中的样本整车车速和样本车轮转速确定滑移率。基于目标奖励函数和确定出的滑移率, 确定样本轮缸压力对应的目标奖励值。基于目标奖励值, 调整待训练的预设决策网络模型中的网络参数, 直到满足预设收敛条件时训练结束, 获得训练结束后的预设决策网络模型。

[0055] S230、获取目标车辆对应的当前车速信息, 其中, 当前车速信息包括: 当前整车车速和每个车轮对应的当前车轮转速。

[0056] S240、将当前车速信息输入至预设决策网络模型中进行轮缸压力动作的决策。

[0057] S250、基于预设决策网络模型的输出, 获得目标车辆中每个车轮对应的目标轮缸压力。

[0058] S260、基于目标轮缸压力对当前车速信息进行调整控制。

[0059] 本发明实施例的技术方案, 通过基于车辆中每个车轮对应的车轮滑移率构建目标奖励函数, 其中, 每个车轮对应的一个车轮奖励函数。目标奖励函数可以是基于所有车轮对应的车轮奖励函数进行整合后确定的。确定出每个车轮对应的车轮奖励函数, 再基于所有车轮对应的车轮奖励函数确定出整车对应的目标奖励函数, 从而可以针对每个车轮进行细节调整, 从而促使整车随着每个车轮的调整而调整, 避免了直接对整车进行调整出现较大的调整误差的情况, 进而保证基于目标奖励函数进行强化学习训练后的预设决策网络模型的准确性, 从而更加准确地控制车速, 进而保持车辆制动或加速时的稳定性, 进一步提高车辆行驶的安全性和用户的驾驶体验。

[0060] 在上述技术方案的基础上, S210 可以包括: 基于车辆中每个车轮对应的车轮滑移率, 确定每个车轮对应的车轮奖励函数; 基于各个车轮奖励函数, 构建目标奖励函数。

[0061] 示例性地, 将各个车轮奖励函数进行相加, 并将相加结果确定为目标奖励函数。其

中,以四轮车辆为例,目标奖励函数表示如下:

$$[0062] \quad R_t = R_{S_{L1}} + R_{S_{R1}} + R_{S_{L2}} + R_{S_{R2}}$$

[0063] 其中, R_t 为整车对应的目标奖励函数值, $R_{S_{L1}}$ 为左前轮对应的奖励函数值, $R_{S_{R1}}$ 为右前轮对应的奖励函数值, $R_{S_{L2}}$ 为左后轮对应的奖励函数值, $R_{S_{R2}}$ 为右后轮对应的奖励函数值。

[0064] 在上述技术方案的基础上,“基于车辆中每个车轮对应的车轮滑移率,确定每个车轮对应的车轮奖励函数”可以包括:若车辆中的当前车轮对应的当前车轮滑移率小于预设最小滑移率,则基于第一权重值对当前车轮滑移率进行加权,获得当前车轮对应的第一分段奖励函数;若当前车轮滑移率大于或等于预设最小滑移率,且小于或等于预设最大滑移率,则将当前车轮滑移率作为当前车轮对应的第二分段奖励函数;若当前车轮滑移率大于预设最大滑移率,则基于第二权重值对当前车轮滑移率进行加权,获得当前车轮对应的第三分段奖励函数;

[0065] 其中,第一权重值小于第二权重值。每个车轮对应的车轮奖励函数是关于车轮滑移率的分段奖励函数。具体地,将每个车轮作为当前车轮,确定出每个车轮对应的车轮奖励函数。基于车轮滑移率和相应车轮对应的车轮奖励函数,确定出相应车轮对应的奖励值。沿用上例,左前轮对应的车轮奖励函数如下:

$$[0066] \quad R_{S_{L1}} = \begin{cases} cS_{L1} & \text{if } S_{L1} < a \\ S_{L1} & \text{if } a \leq S_{L1} \leq b \\ dS_{L1} & \text{if } S_{L1} > b \end{cases}$$

[0067] 其中, $R_{S_{L1}}$ 为左前轮的奖励函数值, S_{L1} 为左前轮的车轮滑移率, c 为第一权重值, d 为第二权重值, a 为预设最小滑移率, b 为预设最大滑移率。例如,第一权重值可以为-2,第二权重值可以为-0.2,预设最小滑移率可以为0.1,预设最大滑移率可以为0.3。

$R_{S_{L1}} = cS_{L1}$ 为第一分段奖励函数。

[0068] $R_{S_{L1}} = S_{L1}$ 为第二分段奖励函数。 $R_{S_{L1}} = dS_{L1}$ 为第三分段奖励函数。

[0069] 以下是本发明实施例提供的基于强化学习的车速控制装置的实施例,该装置与上述各实施例的基于强化学习的车速控制方法属于同一个发明构思,在基于强化学习的车速控制装置的实施例中未详尽描述的细节内容,可以参考上述基于强化学习的车速控制方法的实施例。

[0070] 实施例三

[0071] 图3为本发明实施例三提供的一种基于强化学习的车速控制装置的结构示意图。如图3所示,该装置包括:当前车速信息获取模块310、第一动作决策模块320、目标轮缸压力获得模块330和当前车速信息调整模块340。

[0072] 其中,当前车速信息获取模块310,用于获取目标车辆对应的当前车速信息,其中,当前车速信息包括:当前整车车速和每个车轮对应的当前车轮转速;第一动作决策模块320,用于将当前车速信息输入至预设决策网络模型中进行轮缸压力动作的决策,预设决策网络模型是预先基于目标奖励函数进行强化学习获得的,目标奖励函数是基于车辆中每个

车轮对应的车轮滑移率进行构建的;目标轮缸压力获得模块330,用于基于预设决策网络模型的输出,获得目标车辆中每个车轮对应的目标轮缸压力;当前车速信息调整模块340,用于基于目标轮缸压力对当前车速信息进行调整控制。

[0073] 本发明实施例的技术方案,通过获取目标车辆对应的当前车速信息,以便基于获取的当前车速信息,确定出当前时刻目标车辆中每个车轮对应的调整策略,从而对比上一时刻和当前时刻目标车辆中每个车轮对应的调整策略。其中,当前车速信息包括:当前整车车速和每个车轮对应的当前车轮转速;将当前车速信息输入至预设决策网络模型中进行轮缸压力动作的决策,预设决策网络模型是预先基于目标奖励函数进行强化学习获得的,目标奖励函数是基于车辆中每个车轮对应的车轮滑移率进行构建的;基于预设决策网络模型的输出,获得目标车辆中每个车轮对应的目标轮缸压力;基于目标轮缸压力对当前车速信息进行调整控制,使得每个车轮不会出现抱死的情况,且可以将车轮的滑移率控制在最优滑移率的范围内,从而有效控制车速,并有效缩短车辆的制动距离,进而保持车辆制动或加速时的稳定性,进一步提高车辆行驶的安全性和用户的驾驶体验。

[0074] 可选地,该装置还包括:

[0075] 目标奖励函数构建模块,用于基于车辆中每个车轮对应的车轮滑移率构建目标奖励函数;

[0076] 目标奖励函数构建模块,可以包括:

[0077] 车轮奖励函数确定子模块,用于基于车辆中每个车轮对应的车轮滑移率,确定每个车轮对应的车轮奖励函数;

[0078] 目标奖励函数构建子模块,用于基于各个车轮奖励函数,构建目标奖励函数。

[0079] 可选地,每个车轮对应的车轮奖励函数是关于车轮滑移率的分段奖励函数。

[0080] 可选地,车轮奖励函数确定子模块具体用于:若车辆中的当前车轮对应的当前车轮滑移率小于预设最小滑移率,则基于第一权重值对当前车轮滑移率进行加权,获得当前车轮对应的第一分段奖励函数;若当前车轮滑移率大于或等于预设最小滑移率,且小于或等于预设最大滑移率,则将当前车轮滑移率作为当前车轮对应的第二段奖励函数;若当前车轮滑移率大于预设最大滑移率,则基于第二权重值对当前车轮滑移率进行加权,获得当前车轮对应的第三分段奖励函数;其中,第一权重值小于第二权重值。

[0081] 可选地,目标奖励函数构建子模块具体用于:将各个车轮奖励函数进行相加,并将相加结果确定为目标奖励函数。

[0082] 可选地,当前车速信息调整模块340具体用于:将每个车轮对应的轮缸压力调整为相应的目标轮缸压力,使得目标车辆的当前整车速度调整为目标轮缸压力对应的目标整车速度。

[0083] 可选地,该装置还包括:

[0084] 第二动作决策模块,用于将样本车辆的样本车速信息输入至待训练的预设决策网络模型中进行轮缸压力动作的决策;

[0085] 样本车速信息获得模块,用于基于待训练的预设决策网络模型输出的样本轮缸压力,对样本车速信息进行调整控制,获得控制后的样本车速信息;

[0086] 目标奖励值确定模块,用于基于目标奖励函数和控制后的样本车速信息,确定样本轮缸压力对应的目标奖励值;

[0087] 预设决策网络模型训练模块,用于基于目标奖励值,调整待训练的预设决策网络模型中的网络参数,直到满足预设收敛条件时训练结束,获得训练结束后的预设决策网络模型。

[0088] 本发明实施例所提供的基于强化学习的车速控制装置可执行本发明任意实施例所提供的基于强化学习的车速控制方法,具备执行基于强化学习的车速控制方法相应的功能模块和有益效果。

[0089] 值得注意的是,上述基于强化学习的车速控制装置的实施例中,所包括的各个单元和模块只是按照功能逻辑进行划分的,但并不局限于上述的划分,只要能够实现相应的功能即可;另外,各功能单元的具体名称也只是为了便于相互区分,并不用于限制本发明的保护范围。

[0090] 实施例四

[0091] 图4示出了可以用来实施本发明的实施例的电子设备10的结构示意图。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(如头盔、眼镜、手表等)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、以及它们的功能仅仅作为例,并且不意在限制本文中描述的和/或者要求的本发明的实现。

[0092] 如图4所示,电子设备10包括至少一个处理器11,以及与至少一个处理器11通信连接的存储器,如只读存储器(ROM)12、随机访问存储器(RAM)13等,其中,存储器存储有可被至少一个处理器执行的计算机程序,处理器11可以根据存储在只读存储器(ROM)12中的计算机程序或者从存储单元18加载到随机访问存储器(RAM)13中的计算机程序,来执行各种适当的动作和处理。在RAM 13中,还可存储电子设备10操作所需的各种程序和数据。处理器11、ROM 12以及RAM 13通过总线14彼此相连。输入/输出(I/O)接口15也连接至总线14。

[0093] 电子设备10中的多个部件连接至I/O接口15,包括:输入单元16,例如键盘、鼠标等;输出单元17,例如各种类型的显示器、扬声器等;存储单元18,例如磁盘、光盘等;以及通信单元19,例如网卡、调制解调器、无线通信收发机等。通信单元19允许电子设备10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。

[0094] 处理器11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。处理器11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处理器、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。处理器11执行上文所描述的各个方法和处理,例如基于强化学习的车速控制方法。

[0095] 在一些实施例中,基于强化学习的车速控制方法可被实现为计算机程序,其被有形地包含于计算机可读存储介质,例如存储单元18。在一些实施例中,计算机程序的部分或者全部可以经由ROM 12和/或通信单元19而被载入和/或安装到电子设备10上。当计算机程序加载到RAM 13并由处理器11执行时,可以执行上文描述的基于强化学习的车速控制方法的一个或多个步骤。备选地,在其他实施例中,处理器11可以通过其他任何适当的方式(例如,借助于固件)而被配置为执行基于强化学习的车速控制方法。

[0096] 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。

[0097] 用于实施本发明的方法的计算机程序可以采用一个或多个编程语言的任何组合来编写。这些计算机程序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器,使得计算机程序当由处理器执行时使流程图和/或框图中所规定的功能/操作被实施。计算机程序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。

[0098] 在本发明的上下文中,计算机可读存储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计算机程序。计算机可读存储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。备选地,计算机可读存储介质可以是机器可读信号介质。机器可读存储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。

[0099] 为了提供与用户的交互,可以在电子设备上实施此处描述的系统和技术,该电子设备具有: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电子设备。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。

[0100] 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)、区块链网络和互联网。

[0101] 计算系统可以包括客户端和服务器。客户端和服务器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器的关系。服务器可以是云服务器,又称为云计算服务器或云主机,是云计算服务体系中的一项主机产品,以解决了传统物理主机与VPS服务中,存在的管理难度大,业务扩展性弱的缺陷。

[0102] 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,本发明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,只要能够实现本发明的技术方案所期望的结果,本文在此不进行限制。

[0103] 上述具体实施方式,并不构成对本发明保护范围的限制。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。任何在本发明的精神和原则之内所作的修改、等同替换和改进等,均应包含在本发明保护范围之内。

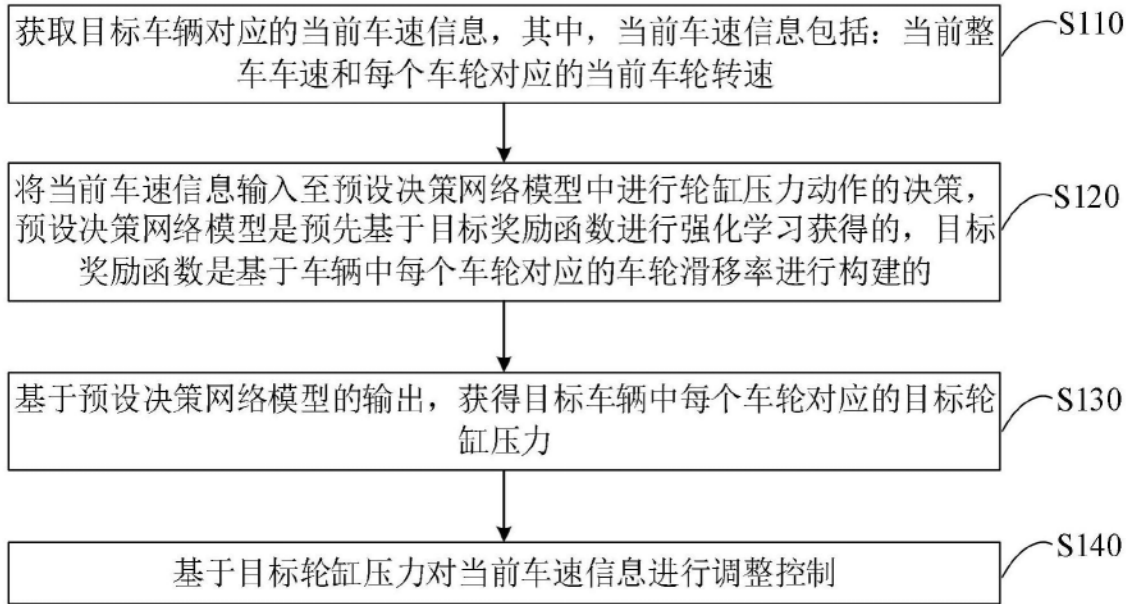


图1

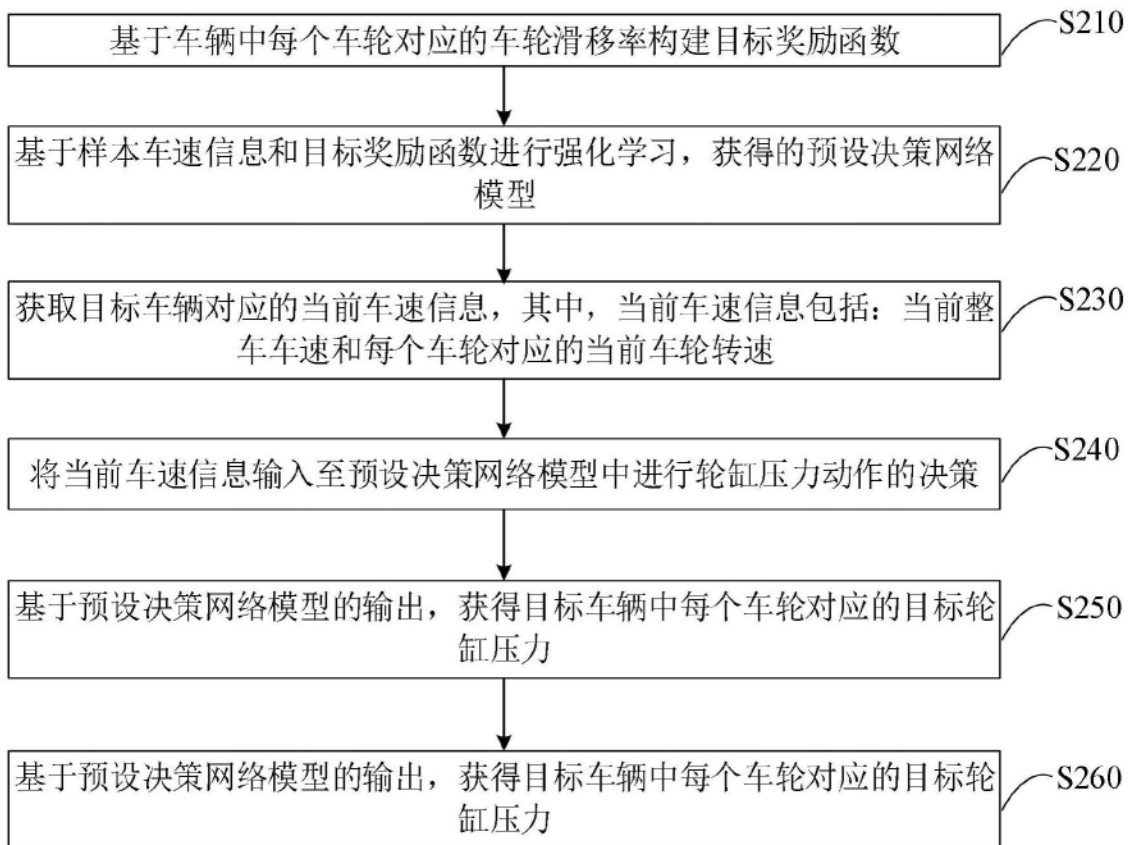


图2

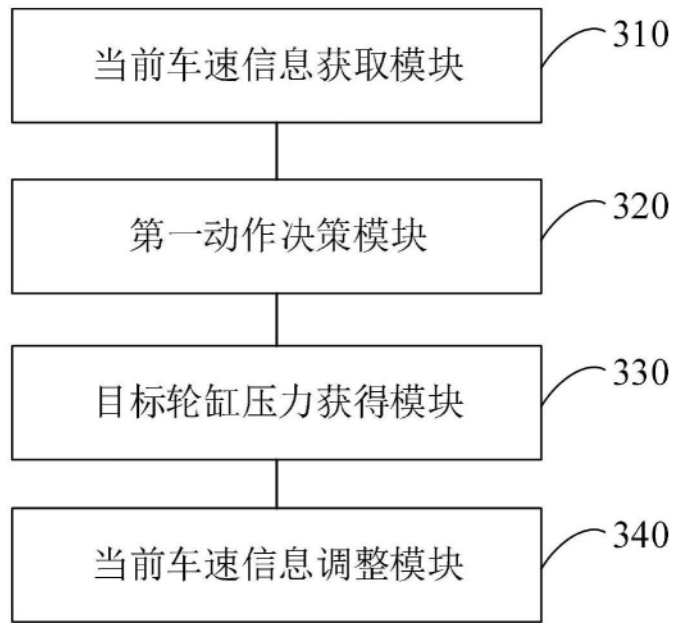


图3

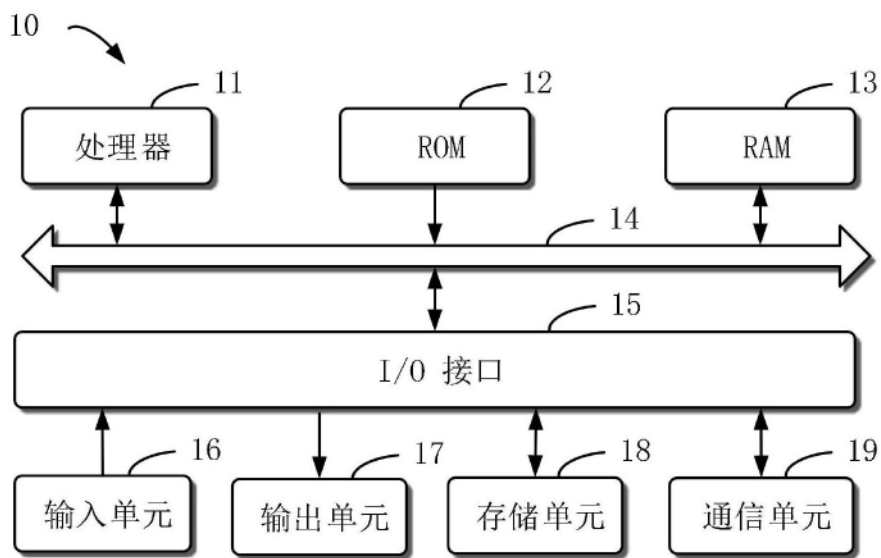


图4