(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 111547039 B
(45)授权公告日 2021.03.23

(21)申请号 202010401009.9

(22)申请日 2020.05.13

(65)同一申请的已公布的文献号
申请公布号 CN 111547039 A

(43)申请公布日 2020.08.18

(73)专利权人 北京理工大学
地址 100044 北京市海淀区中关村南大街5号

(72)发明人 邹渊 张旭东 孙逢春 邹润楠

(74)专利代理机构 北京高沃律师事务所 11569
代理人 杜阳阳

(51)Int.Cl.
B60W 20/00(2016.01)
B60W 50/00(2006.01)

(56)对比文件
CN 101630144 A,2010.01.20
JP H07329534 A,1995.12.19
JP 2010095067 A,2010.04.30
CN 110834537 A,2020.02.25
CN 109483530 A,2019.03.19
DE 102019110184 A1,2019.10.31

审查员 刘亚运

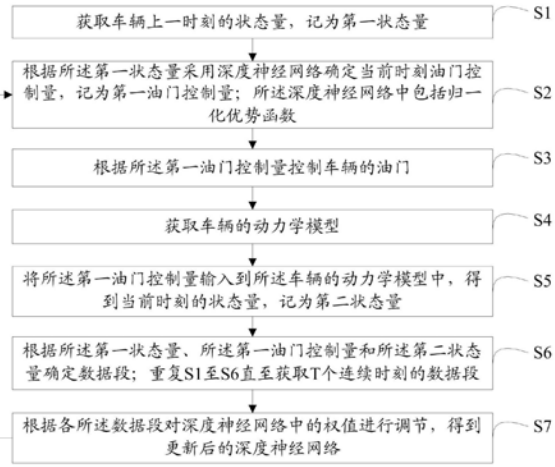
权利要求书2页 说明书7页 附图1页

(54)发明名称

基于深度强化学习的混合动力车辆油门控制方法及系统

(57)摘要

本发明涉及一种基于深度强化学习的混合动力车辆油门控制方法及系统，包括：获取车辆上一时刻的状态量，记第一状态量；根据第一状态量采用深度神经网络确定当前时刻油门控制量，记第一油门控制量；深度神经网络中包括归一化优势函数；根据第一油门控制量控制车辆的油门；将第一油门控制量输入到车辆的动力学模型中，得当前时刻的状态量，记第二状态量；根据第一状态量、第一油门控制量和第二状态量确定数据段；直至获取T个连续时刻的数据段；根据各数据段对深度神经网络中的权值进行调节，得更新后的深度神经网络，采用更新后的深度神经网络确定当前时刻油门控制量，从而精确的对车辆油门进行控制，通过本发明的上述方法提高了对油门的精度控制。



CN 111547039 B

1.一种基于深度强化学习的混合动力车辆油门控制方法,其特征在于,所述混合动力车辆油门控制方法包括:

S1,获取车辆上一时刻的状态量,记为第一状态量;

S2,根据所述第一状态量采用深度神经网络确定当前时刻油门控制量,记为第一油门控制量;所述深度神经网络中包括归一化优势函数;

S3,根据所述第一油门控制量控制车辆的油门;

S4,获取车辆的动力学模型;

S5,将所述第一油门控制量输入到所述车辆的动力学模型中,得到当前时刻的状态量,记为第二状态量;

S6,根据所述第一状态量、所述第一油门控制量和所述第二状态量确定数据段;重复S1至S6直至获取T个连续时刻的数据段;

S7,根据各所述数据段对深度神经网络中的权值进行调节,得到更新后的深度神经网络;并返回S2,采用所述更新后的深度神经网络确定当前时刻油门控制量。

2.根据权利要求1所述的基于深度强化学习的混合动力车辆油门控制方法,其特征在于,所述归一化优势函数为:

$$A\left(s,\ \alpha\middle|\theta^{A}\right)=-\frac{1}{2}\left(\alpha-\mu\left(s\middle|\theta^{\mu}\right)\right)^{T}P\left(s\middle|\theta^{P}\right)\left(\alpha-\mu\left(s\middle|\theta^{\mu}\right)\right);$$

其中,s为车辆状态量,α为油门控制量,μ为在状态量s下的最优动作,$P(s|\theta^{P})=L(s|\theta^{P})L(s|\theta^{P})^{T}$,A(•)为归一化优势函数,$\theta^{A}$为归一化优势函数,$\theta^{P}$为矩阵P的参数,$\theta^{\mu}$为μ的参数,L(•)为下三角矩阵。

3.一种基于深度强化学习的混合动力车辆油门控制系统,其特征在于,所述混合动力车辆油门控制系统包括:

第一状态量获取模块,用于获取车辆上一时刻的状态量,记为第一状态量;

第一油门控制量确定模块,用于根据所述第一状态量采用深度神经网络确定当前时刻油门控制量,记为第一油门控制量;所述深度神经网络中包括归一化优势函数;

车辆油门控制模块,用于根据所述第一油门控制量控制车辆的油门;

车辆的动力学模型获取模块,用于获取车辆的动力学模型;

第二状态量确定模块,用于将所述第一油门控制量输入到所述车辆的动力学模型中,得到当前时刻的状态量,记为第二状态量;

数据段获取模块,用于根据所述第一状态量、所述第一油门控制量和所述第二状态量确定数据段;直至获取T个连续时刻的数据段;

深度神经网络更新模块,用于根据各所述数据段对深度神经网络中的权值进行调节,得到更新后的深度神经网络;并返回所述第一油门控制量确定模块,采用所述更新后的深度神经网络确定当前时刻油门控制量。

4.根据权利要求3所述的基于深度强化学习的混合动力车辆油门控制系统,其特征在于,所述归一化优势函数为:

$$A\left(s,\ \alpha\middle|\theta^{A}\right)=-\frac{1}{2}\left(\alpha-\mu\left(s\middle|\theta^{\mu}\right)\right)^{T}P\left(s\middle|\theta^{P}\right)\left(\alpha-\mu\left(s\middle|\theta^{\mu}\right)\right);$$

其中,s为车辆状态量,α为油门控制量,μ为在状态量s下的最优动作,$P(s|\theta^{P})=L(s|\theta^{P})$

$L(s|\theta^P)^T$,$A(\cdot)$为归一化优势函数,θ^A为归一化优势函数,θ^P为矩阵P的参数,θ^μ为μ的参数,$L(\cdot)$为下三角矩阵。

基于深度强化学习的混合动力车辆油门控制方法及系统

技术领域

[0001] 本发明涉及汽车油门控制技术领域，特别是涉及一种基于深度强化学习的混合动力车辆油门控制方法及系统。

背景技术

[0002] 目前，针对混合动力汽车油门控制问题的主要解决方案有基于规则的方法和以动态规划、强化学习为代表的基于优化的方法。基于规则的方法需要提前知道工程师预设发动机及电池工作模式切换规则，因此对于复杂多变的路况缺乏适应性，难以实现混合动力车辆节能性及高机动性。基于深度强化学习的油门控制方法能有效学习道路工况信息，针对已获取道路信息通过神经网络的训练求得最优策略。但是传统深度强化学习训练中，常将已训练过数据片段储存于经验池中，在训练中随机提取进行再训练以打破数据相关性，随机提取历史经验片段使得训练时间较长且伴随有陷入局部最优解的风险，从而获取的控制量精度低。

发明内容

[0003] 本发明的目的是提供一种基于深度强化学习的混合动力车辆油门控制方法及系统，提高油门控制精度。

[0004] 为实现上述目的，本发明提供了如下方案：

[0005] 一种基于深度强化学习的混合动力车辆油门控制方法，所述混合动力车辆油门控制方法包括：

[0006] S1，获取车辆上一时刻的状态量，记为第一状态量；

[0007] S2，根据所述第一状态量采用深度神经网络确定当前时刻油门控制量，记为第一油门控制量；所述深度神经网络中包括归一化优势函数；

[0008] S3，根据所述第一油门控制量控制车辆的油门；

[0009] S4，获取车辆的动力学模型；

[0010] S5，将所述第一油门控制量输入到所述车辆的动力学模型中，得到当前时刻的状态量，记为第二状态量；

[0011] S6，根据所述第一状态量、所述第一油门控制量和所述第二状态量确定数据段；重复S1至S6直至获取T个连续时刻的数据段；

[0012] S7，根据各所述数据段对深度神经网络中的权值进行调节，得到更新后的深度神经网络；并返回S2，采用所述更新后的深度神经网络确定当前时刻油门控制量。

[0013] 可选的，所述根据各所述数据段对深度神经网络中的权值进行调节，得到更新后的深度神经网络，具体包括：

[0014] 根据所述数据段获取所述数据段对应的取值概率；

[0015] 根据所述取值概率确定数据段权值；

[0016] 根据所述车辆的动力学模型，采用深度强化学习奖励函数确定数据段所对应的奖

励；

[0017] 根据所述奖励和所述数据段权值确定数据段损失；

[0018] 根据所述数据段损失调节所述深度神经网络中的权值，得到更新后的深度神经网络。

[0019] 可选的，所述根据所述车辆的动力学模型，采用深度强化学习奖励函数确定数据段所对应的奖励，具体包括：

[0020] 根据公式$R(s,a)=-\left[\int_{t_0}^{t}\alpha\,\dot{f}_{rate}(t)dt+\beta[S\dot{O}C(t)-S\dot{O}C(t_0)]^2\right]$确定数据段所对应的奖励；

[0021] 其中，R(s,a)为车辆在状态量s下进行a动作所得的奖励，α，β均为正参数，$\dot{f}_{rate}$为车辆发动机燃油消耗率，[t_0,t]为车辆工作时间段，$\dot{SOC}(t_0)$为t_0时刻电池荷电状态变化率，$\dot{SOC}(t)$为t时刻电池荷电状态变化率。

[0022] 可选的，所述归一化优势函数为：

[0023] $$A(s,a|\theta^A)=-\frac{1}{2}\left(a-\mu(s|\theta^\mu)\right)^T P(s|\theta^P)\left(a-\mu(s|\theta^\mu)\right);$$

[0024] 其中，s为车辆状态量，a为油门控制量，μ为在状态量s下的最优动作，$P(s|\theta^P)=L(s|\theta^P)L(s|\theta^P)^T$，A(•)为归一化优势函数，$\theta^A$为归一化优势函数，$\theta^P$为矩阵P的参数，$\theta^\mu$为μ的参数，L(•)为下三角矩阵。

[0025] 一种基于深度强化学习的混合动力车辆油门控制系统，所述混合动力车辆油门控制系统包括：

[0026] 第一状态量获取模块，用于获取车辆上一时刻的状态量，记为第一状态量；

[0027] 第一油门控制量确定模块，用于根据所述第一状态量采用深度神经网络确定当前时刻油门控制量，记为第一油门控制量；所述深度神经网络中包括归一化优势函数；

[0028] 车辆油门控制模块，用于根据所述第一油门控制量控制车辆的油门；

[0029] 车辆的动力学模型获取模块，用于获取车辆的动力学模型；

[0030] 第二状态量确定模块，用于将所述第一油门控制量输入到所述车辆的动力学模型中，得到当前时刻的状态量，记为第二状态量；

[0031] 数据段获取模块，用于根据所述第一状态量、所述第一油门控制量和所述第二状态量确定数据段；直至获取T个连续时刻的数据段；

[0032] 深度神经网络更新模块，用于根据各所述数据段对深度神经网络中的权值进行调节，得到更新后的深度神经网络；并返回所述第一油门控制量确定模块，采用所述更新后的深度神经网络确定当前时刻油门控制量。

[0033] 可选的，所述深度神经网络更新模块具体包括：

[0034] 取值概率获取单元，用于根据所述数据段获取所述数据段对应的取值概率；

[0035] 数据段权值确定单元，用于根据所述取值概率确定数据段权值；

[0036] 奖励确定单元，用于根据所述车辆的动力学模型，采用深度强化学习奖励函数确定数据段所对应的奖励；

[0037] 数据段损失确定单元，用于根据所述奖励和所述数据段权值确定数据段损失；

[0038] 深度神经网络更新单元，用于根据所述数据段损失调节所述深度神经网络中的权

值，得到更新后的深度神经网络。

[0039] 可选的，所述奖励确定单元具体包括：

[0040] 奖励确定子单元，用于根据公式$R(s,a)=-\left[\int_{t_0}^{t}\alpha\,\dot{f}_{rate}(t)dt+\beta[S\dot{O}C(t)-S\dot{O}C(t_0)]^2\right]$确定数据段所对应的奖励；

[0041] 其中，R(s,a)为车辆在状态量s下进行a动作所得的奖励，α，β均为正参数，$\dot{f}_{rate}$为车辆发动机燃油消耗率，$[t_0,t]$为车辆工作时间段，$S\dot{O}C(t_0)$为t_0时刻电池荷电状态变化率，$S\dot{O}C(t)$为t时刻电池荷电状态变化率。

[0042] 可选的，所述归一化优势函数为：

[0043] $$A(s,a|\theta^{A})=-\frac{1}{2}\left(a-\mu(s|\theta^{\mu})\right)^{T}P(s|\theta^{P})\left(a-\mu(s|\theta^{\mu})\right);$$

[0044] 其中，s为车辆状态量，a为油门控制量，μ为在状态量s下的最优动作，$P(s|\theta^{P})=L(s|\theta^{P})L(s|\theta^{P})^{T}$，A(•)为归一化优势函数，$\theta^{A}$为归一化优势函数，$\theta^{P}$为矩阵P的参数，$\theta^{\mu}$为μ的参数，L(•)为下三角矩阵。

[0045] 根据本发明提供的具体实施例，本发明公开了以下技术效果：

[0046] 本发明提供了一种基于深度强化学习的混合动力车辆油门控制方法机系统，将混合动力车辆状态量输入至深度神经网络中，利用归一化优势函数及经验优先权值回顾对深度神经网络中的权值进行调节，采用更新后的深度神经网络得到高精度控制量，提高对油门的精确控制。

附图说明

[0047] 为了更清楚地说明本发明实施例或现有技术中的技术方案，下面将对实施例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性的前提下，还可以根据这些附图获得其他的附图。

[0048] 图1为本发明所提供的一种基于深度强化学习的混合动力车辆油门控制方法流程图；

[0049] 图2为本发明所提供的一种基于深度强化学习的混合动力车辆油门控制系统的结构示意图。

具体实施方式

[0050] 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，都属于本发明保护的范围。

[0051] 本发明的目的是提供一种基于深度强化学习的混合动力车辆油门控制方法及系统，提高油门控制精度。

[0052] 为使本发明的上述目的、特征和优点能够更加明显易懂，下面结合附图和具体实施方式对本发明作进一步详细的说明。

[0053] 图1为本发明所提供的一种基于深度强化学习的混合动力车辆油门控制方法流程图，如图1所示，本发明所述混合动力车辆油门控制方法包括：

[0054] S1，获取车辆上一时刻的状态量，记为第一状态量。

[0055] S2，根据所述第一状态量采用深度神经网络确定当前时刻油门控制量，记为第一油门控制量；所述深度神经网络中包括归一化优势函数。

[0056] S3，根据所述第一油门控制量控制车辆的油门。

[0057] S4，获取车辆的动力学模型。

[0058] S5，将所述第一油门控制量输入到所述车辆的动力学模型中，得到当前时刻的状态量，记为第二状态量。

[0059] S6，根据所述第一状态量、所述第一油门控制量和所述第二状态量确定数据段；重复S1至S6直至获取T个连续时刻的数据段。

[0060] S7，根据各所述数据段对深度神经网络中的权值进行调节，得到更新后的深度神经网络；并返回S2，采用所述更新后的深度神经网络确定当前时刻油门控制量。

[0061] 所述根据各所述数据段对深度神经网络中的权值进行调节，得到更新后的深度神经网络，具体包括：根据所述数据段获取所述数据段对应的取值概率；根据所述取值概率确定数据段权值；根据所述车辆的动力学模型，采用深度强化学习奖励函数确定数据段所对应的奖励；根据所述奖励和所述数据段权值确定数据段损失；根据所述数据段损失调节所述深度神经网络中的权值，得到更新后的深度神经网络。具体的，根据公式 $R(s,a)=-\left[\int_{t_0}^{t}\alpha\dot{f}_{rate}(t)dt+\beta[S\dot{O}C(t)-S\dot{O}C(t_0)]^2\right]$ 确定数据段所对应的奖励，其中，R(s,a)为车辆在状态量s下进行a动作所得的奖励，α，β均为正参数，α+β＝1，$\dot{f}_{rate}$为车辆发动机燃油消耗率，[t_0,t]为车辆工作时间段，$\dot{SOC}(t_0)$为t_0时刻电池荷电状态变化率，$\dot{SOC}(t)$为t时刻电池荷电状态变化率。

[0062] 所述归一化优势函数为：

[0063] $$A(s,a|\theta^A)=-\frac{1}{2}\left(a-\mu(s|\theta^\mu)\right)^T P(s|\theta^P)\left(a-\mu(s|\theta^\mu)\right);$$

[0064] 其中，s为车辆状态量，a为油门控制量，μ为在状态量s下的最优动作，$P(s|\theta^P)=L(s|\theta^P)L(s|\theta^P)^T$，A(•)为归一化优势函数，$\theta^A$为归一化优势函数，$\theta^P$为矩阵P的参数，$\theta^\mu$为μ的参数，L(•)为下三角矩阵。

[0065] 下面对各步骤进行详细论述：

[0066] 根据所使用车辆搭建车辆的动力学模型：根据所使用混合动力车辆底盘构型、能源动力装置及传动装置进行数学建模，基于python建立车辆动力学模型及车辆各组件数学模型。一般地，建立发动机-发电机模型、动力电池模型、电气驱动系统模型及整车综合控制模型。根据车辆模型确定能量管理状态变量、反馈奖励及控制量，确定发动机转速、电池荷电状态及整车需求功率为能量管理状态变量，具有变速器的车辆将挡位加入状态变量。

[0067] 搭建具有估值网络和评价网络的深度神经网络。

[0068] 根据已确定的状态量确定深度强化学习奖励函数：

$$R(s,a)=-\left[\int_{t_0}^{t}\alpha\dot{f}_{rate}(t)dt+\beta[S\dot{O}C(t)-S\dot{O}C(t_0)]^2\right]。$$

[0069] 搭建包含有两个隐含层一个输出层的深度神经网络，每个网络中都包含有激活函数，输出层中经过先行激活函数处理，分别输出状态动作值，系统控制量μ和归一化优势函数构造下三角矩阵L(s)。此下三角矩阵由神经网络计算得出。估值网络和目标网络结构一致，目标网络参数值由估值网络延迟复制得来。

[0070] 基于深度神经网络搭建归一化优势函数。

[0071] 基于估值深度神经网络输出量，为实现深度强化学习模型直接训练得到控制量，减少运算时间且提高控制精度，搭建归一化优势函数：

[0072] $$A(s,a|\theta^A)=-\frac{1}{2}\left(a-\mu(s|\theta^\mu)\right)^T P(s|\theta^P)\left(a-\mu(s|\theta^\mu)\right)$$

[0073] 其中，s为车辆状态量，a为油门控制量即油门开度，μ为估值网络在状态量s下的最优动作，$P(s|\theta^P)=L(s|\theta^P)L(s|\theta^P)^T$，A(•)为归一化优势函数，$\theta^A$为归一化优势函数，$\theta^P$为矩阵P的参数，$\theta^\mu$为μ的参数，L(•)为下三角矩阵，P为关于系统状态的正定方阵，当a＝μ时，此函数取得最大值，构造正定矩阵P基于正定矩阵唯一Cholesky分解，其中L(•)为下三角矩阵，由估值神经网络输出。

[0074] 将深度神经网络输出输入至搭建好的归一化优势函数，可得混合动力车辆油门控制量，a＝μ。

[0075] 基于深度神经网络结构搭建经验权值优先回顾模型。

[0076] 搭建SumTree结构储存历史经验数据即N个连续的数据段，历史经验数据为多个数据段的存储空间，每一个数据段(经验)形式为(S_{t-1},a_t,S_t)，其中S_{t-1}为t-1时刻状态，经过t时刻油门a_t的控制，混合动力车辆状态转移至S_t。

[0077] 给予经验池中各数据段取值概率P(j)：$P(j)=p_j^\alpha/\sum p_i^\alpha$，其中，$p_j^\alpha$和$p_i^\alpha$均为各数据段优先值。

[0078] 计算数据段权值：$\omega_j=(N\cdot P(j))^{-\beta}/\max_i \omega_i$，其中，N为经验数量，0＜N＜256，β＝1，$\max_i w_i$为$w_i$中取值最大。

[0079] 计算数据段的TD-error：$\delta_j=R_j+\gamma_j\cdot\hat{Q}(S_j,A_j)-Q(S_j,A_j)$，其中，$\delta_j$为TD-error即数据段损失，$R_j$为该数据段在环境中应用后所得奖励，$\hat{Q}(S_j,A_j)$为目标网络计算所得Q值，$Q(S_j,A_j)$为估值网络计算所得Q值，$\gamma_j$为折扣因子，一个关于期望的常数，在0～1之间，越靠近1就理解为当前结果对最终结果影响越大，S_j为第j个数据段的状态，A_j第j个数据段的动作。

[0080] 计算数据段优先值：$p(j)=|\delta_j|^{0.5}$。

[0081] 根据数据段权值及TD-error计算深度神经网络权值改变量Δ_t：$\Delta_t=\Delta_{t-1}+\omega_j\cdot\delta_j\cdot\nabla_\theta Q(S_{j-1},A_{j-1})$，$\nabla_\theta$为关于theta的梯度。

[0082] 更新深度神经网络权值θ：$\theta_t=\theta_{t-1}+\eta\cdot\Delta_t$，其中，$\Delta_t$为深度神经网络权值改变量。

[0083] 通过定期经验权值优先回顾及网络更新，输出油门控制量，当通过多次迭代，油门控制量收敛(训练变化不大)时，训练完成。具体的，初始化经验池内存空间h，每次回顾数据段大小为n，经验回顾周期T_r，即T个连续时刻，最大训练次数M_{max}，随机初始化归一化估值网络参数，初始化目标网络权重参数，初始化学习率η。

[0084] 针对目标工况时间t，得到此时混合动力车辆状态量s_t，将状态量输入进深度神经网络得到控制量a_t。将控制量输入至混合动力车辆模型得到奖励R_t及下一时刻状态量s_{t+1}。将此状态量存入经验池并计算其取值概率P_t。

[0085] 每过T_r时刻，进入经验回顾模式，更新深度神经网络中的权值。

[0086] 将更新后的深度神经网络用于混合动力车辆能量管理。获取当前车辆工况信息，采用更新后的深度神经网络确定当前时刻油门控制量，得到混合动力车辆能量管理策略。指的是针对一个工况，一个系列的油门控制量，是一个数组。

[0087] 本发明还提供了一种基于深度强化学习的混合动力车辆油门控制系统，如图2所示，所述混合动力车辆油门控制系统包括：

[0088] 第一状态量获取模块1，用于获取车辆上一时刻的状态量，记为第一状态量。

[0089] 第一油门控制量确定模块2，用于根据所述第一状态量采用深度神经网络确定当前时刻油门控制量，记为第一油门控制量；所述深度神经网络中包括归一化优势函数。

[0090] 车辆油门控制模块3，用于根据所述第一油门控制量控制车辆的油门。

[0091] 车辆的动力学模型获取模块4，用于获取车辆的动力学模型。

[0092] 第二状态量确定模块5，用于将所述第一油门控制量输入到所述车辆的动力学模型中，得到当前时刻的状态量，记为第二状态量。

[0093] 数据段获取模块6，用于根据所述第一状态量、所述第一油门控制量和所述第二状态量确定数据段；直至获取T个连续时刻的数据段。

[0094] 深度神经网络更新模块7，用于根据各所述数据段对深度神经网络中的权值进行调节，得到更新后的深度神经网络；并返回所述第一油门控制量确定模块2，采用所述更新后的深度神经网络确定当前时刻油门控制量。

[0095] 优选的，所述深度神经网络更新模块7具体包括：

[0096] 取值概率获取单元，用于根据所述数据段获取所述数据段对应的取值概率。

[0097] 数据段权值确定单元，用于根据所述取值概率确定数据段权值。

[0098] 奖励确定单元，用于根据所述车辆的动力学模型，采用深度强化学习奖励函数确定数据段所对应的奖励。

[0099] 数据段损失确定单元，用于根据所述奖励和所述数据段权值确定数据段损失。

[0100] 深度神经网络更新单元，用于根据所述数据段损失调节所述深度神经网络中的权值，得到更新后的深度神经网络。

[0101] 优选的，所述奖励确定单元具体包括：

[0102] 奖励确定子单元，用于根据公式$R(s,a)=-\left[\int_{t_0}^{t}\alpha\,\dot{f}_{rate}(t)dt+\beta[S\dot{O}C(t)-S\dot{O}C(t_0)]^2\right]$确定数据段所对应的奖励；

[0103] 其中，R(s,a)为车辆在状态量s下进行a动作所得的奖励，α，β均为正参数，$\dot{f}_{rate}$为车辆发动机燃油消耗率，$[t_0,t]$为车辆工作时间段，$\dot{SOC}(t_0)$为t_0时刻电池荷电状态变化率，$\dot{SOC}(t)$为t时刻电池荷电状态变化率。

[0104] 优选的，所述归一化优势函数为：

[0105] $A(s,a|\theta^A)=-\frac{1}{2}\left(a-\mu(s|\theta^\mu)\right)^T P(s|\theta^P)\left(a-\mu(s|\theta^\mu)\right)$；

[0106] 其中，s为车辆状态量，a为油门控制量，μ为在状态量s下的最优动作，$P(s|\theta^P)=L(s|\theta^P)L(s|\theta^P)^T$，A(•)为归一化优势函数，$\theta^A$为归一化优势函数，$\theta^P$为矩阵P的参数，$\theta^\mu$为μ的参数，L(•)为下三角矩阵。

[0107] 本说明书中各个实施例采用递进的方式描述，每个实施例重点说明的都是与其他实施例的不同之处，各个实施例之间相同相似部分互相参见即可。对于实施例公开的系统而言，由于其与实施例公开的方法相对应，所以描述的比较简单，相关之处参见方法部分说明即可。

[0108] 本文中应用了具体个例对本发明的原理及实施方式进行了阐述，以上实施例的说明只是用于帮助理解本发明的方法及其核心思想；同时，对于本领域的一般技术人员，依据本发明的思想，在具体实施方式及应用范围上均会有改变之处。综上所述，本说明书内容不应理解为对本发明的限制。

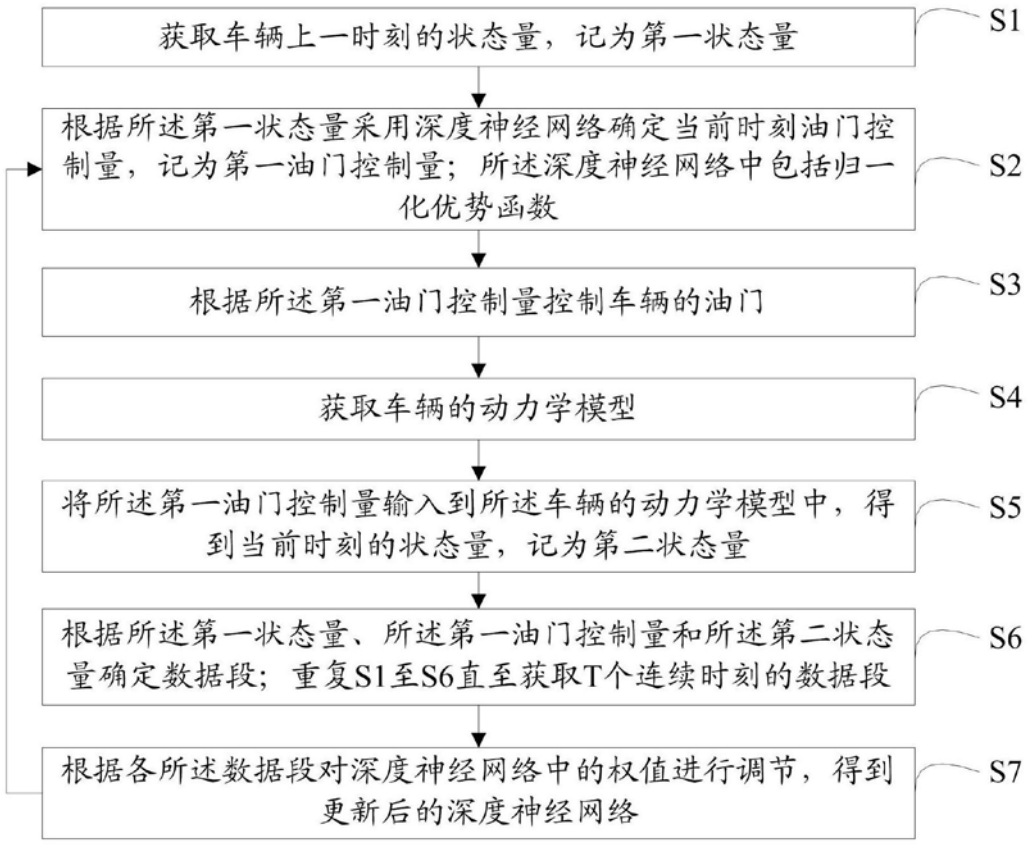

获取车辆上一时刻的状态量，记为第一状态量
S1
根据所述第一状态量采用深度神经网络确定当前时刻油门控制量，记为第一油门控制量；所述深度神经网络中包括归一化优势函数
S2
根据所述第一油门控制量控制车辆的油门
S3
获取车辆的动力学模型
S4
将所述第一油门控制量输入到所述车辆的动力学模型中，得到当前时刻的状态量，记为第二状态量
S5
根据所述第一状态量、所述第一油门控制量和所述第二状态量确定数据段；重复S1至S6直至获取T个连续时刻的数据段
S6
根据各所述数据段对深度神经网络中的权值进行调节，得到更新后的深度神经网络
S7

图1

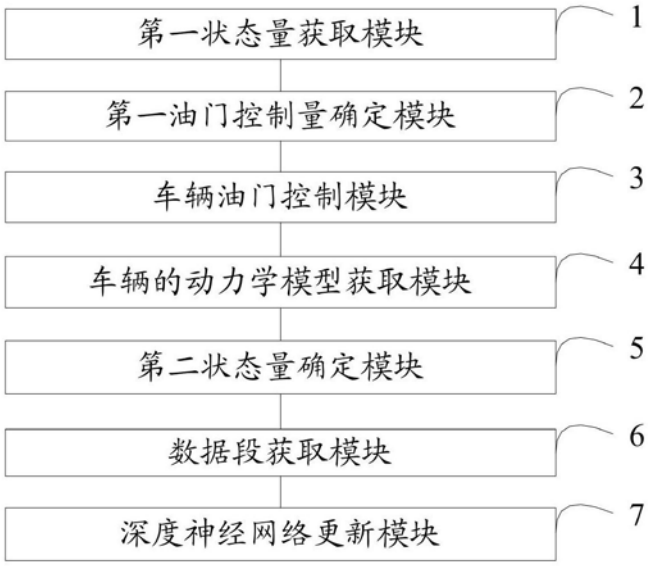

第一状态量获取模块
1
第一油门控制量确定模块
2
车辆油门控制模块
3
车辆的动力学模型获取模块
4
第二状态量确定模块
5
数据段获取模块
6
深度神经网络更新模块
7

图2