



(12)发明专利申请

(10)申请公布号 CN 110347155 A

(43)申请公布日 2019.10.18

(21)申请号 201910562566.6

(22)申请日 2019.06.26

(71)申请人 北京理工大学

地址 100081 北京市海淀区中关村南大街5号

(72)发明人 吕超 于洋 陈昕 龚建伟 杨森

(74)专利代理机构 北京天达知识产权代理事务所(普通合伙) 11386

代理人 侯永帅 庞许倩

(51)Int.Cl.

G05D 1/02(2006.01)

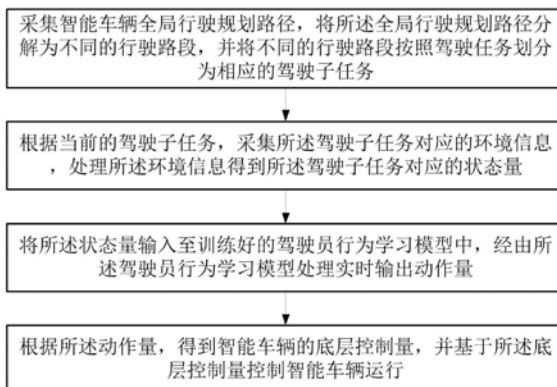
权利要求书3页 说明书8页 附图2页

(54)发明名称

一种智能车辆自动驾驶控制方法及系统

(57)摘要

本发明涉及一种智能车辆自动驾驶控制方法及系统,属于智能驾驶技术领域,解决了现有自动驾驶无法很好地自适应完成在线学习的问题。一种智能车辆自动驾驶控制方法,步骤如下:获取智能车辆全局行驶规划路径,将全局行驶规划路径分解为不同的行驶路段,并将不同的行驶路段按照驾驶任务划分为相应的驾驶子任务;根据当前的驾驶子任务,采集驾驶子任务对应的环境信息,处理环境信息得到所述驾驶子任务对应的状态量;将状态量输入至训练好的驾驶员行为学习模型中,经由所述驾驶员行为学习模型处理实时输出动作量;根据动作量,得到智能车辆的底层控制量,并基于底层控制量控制智能车辆运行。实现了智能车辆自动驾驶的自适应在线学习。



1. 一种智能车辆自动驾驶控制方法,其特征在于,包括如下步骤:

获取智能车辆全局行驶规划路径,将所述全局行驶规划路径分解为不同的行驶路段,并将不同的行驶路段按照驾驶任务划分为相应的驾驶子任务;

根据当前的驾驶子任务,采集所述驾驶子任务对应的环境信息,处理所述环境信息得到所述驾驶子任务对应的状态量;

将所述状态量输入至训练好的驾驶员行为学习模型中,经由所述驾驶员行为学习模型处理实时输出动作量;

根据所述动作量,得到智能车辆的底层控制量,并基于所述底层控制量控制智能车辆运行。

2. 根据权利要求1所述的智能车辆自动驾驶控制方法,其特征在于,将所述全局行驶规划路径分解为不同的行驶路段,并将不同的行驶路段按照驾驶任务划分为相应的驾驶子任务,包括:

根据道路环境的不同,将所述全局行驶规划路径至少分解为直行路段、十字交叉路段;

根据驾驶任务,将所述直行路段的驾驶子任务划分为车道保持子任务或车道变更子任务;将十字交叉路段的驾驶子任务划分为路口左/右转子任务或直行子任务。

3. 根据权利要求2所述的智能车辆自动驾驶控制方法,其特征在于,根据当前的驾驶子任务,采集所述驾驶子任务对应的环境信息,处理所述环境信息得到所述驾驶子任务对应的状态量,包括:

当所述驾驶子任务为所述车道保持子任务时,采集的环境信息为:前车与本车的速度信息、距离信息;此时的状态量为:前车车速与本车车速之间的差值、两车间距与期望车距之间的差值;

当所述驾驶子任务为所述车道变更子任务时,采集的环境信息为:本车和待变更车道上与本车相距预设范围内所有车辆的速度信息、距离信息;此时的状态量信息为:待变更车道上与本车相距预设范围内所有车辆中车速最大值与本车车速之间的差值、待变更车道上距本车最近的车辆与本车之间的距离;

当所述驾驶子任务为所述路口左/右转子任务或直行子任务时,采集的环境信息为:本车和路口处与本车相距预设范围内所有车辆的位置、速度信息;此时的状态量信息为:本车和路口处与本车相距预设范围内所有车辆的位置、速度信息。

4. 根据权利要求3所述的智能车辆自动驾驶控制方法,其特征在于,所述驾驶员行为学习模型包括类型选择层、结构选择层、参数学习层;

所述驾驶员行为学习模型在训练过程中,具体执行以下操作:

根据当前的驾驶子任务,通过类型选择层选择与所述驾驶子任务相适应的神经网络类型、通过结构选择层选择与所述驾驶子任务相适应的神经网络结构参数,并将所述神经网络类型、神经网络结构参数选择结果发送至参数学习层;

所述参数学习层根据所述神经网络类型、神经网络结构参数确定当前神经网络学习模型的结构;

采用基于值函数的强化学习方法,训练所述参数学习层,直至所述神经网络学习模型收敛。

5. 根据权利要求4所述的智能车辆自动驾驶控制方法,其特征在于,所述类型选择层中

可选的神经网络类型至少包括前馈神经网络、深度学习神经网络；

当所述驾驶子任务为车道保持子任务时或车道变更子任务时，所述类型选择层选择的神经网络类型为前馈神经网络；

当所述驾驶子任务为路口左/右转子任务或直行子任务时，所述类型选择层选择的神经网络类型为深度学习神经网络。

6. 根据权利要求5所述的智能车辆自动驾驶控制方法，其特征在于，通过结构选择层选择与驾驶子任务相适应的神经网络结构参数，包括：

根据选用的强化学习方法及当前驾驶子任务确定输入层、输出层的节点数；

根据驾驶子任务的复杂程度确定隐含层层数：当所述驾驶子任务为车道保持子任务或车道变更子任务时，隐含层层数选用单层形式；当所述驾驶子任务为路口左/右转子任务或

直行子任务时，隐含层层数为多层；设定  $H_{xx} = \begin{bmatrix} w_1 & 0 \\ 0 & w_2 \end{bmatrix}$ 、 $H_{xu} = [w_3 \ w_4]^T$ 、 $H_{ux} = [w_3 \ w_4]$ 、 $H_{uu}$

$= [w_5]$ ，从而得到Q函数  $Q(S_k, A_k) = \begin{bmatrix} S_k^T & A_k^T \end{bmatrix} \begin{bmatrix} H_{xx} & H_{xu} \\ H_{ux} & H_{uu} \end{bmatrix} \begin{bmatrix} S_k \\ A_k \end{bmatrix}$  的线性表示形式， $w_1$ 、 $w_2$ 、

$w_3$ 、 $w_4$ 、 $w_5$ 通过神经网络模型的Q函数表达式对输入层节点求偏导解得，其中， $S_k$ 表示第k时刻的状态量， $A_k$ 表示第k时刻的动作量。

7. 根据权利要求4-6中任一项所述的智能车辆自动驾驶控制方法，其特征在于，采用基于值函数的强化学习方法，训练所述参数学习层，直至所述神经网络学习模型收敛，包括：

初始化参数学习层中神经网络的权重参数、激活函数参数，初始化  $Q(S_0, A_0) = 0$ 、状态量  $S_0$ 和动作量  $A_0$ ；

对每个步数k，执行：

接收当前时刻状态量  $S_{k+1}$ ，获取上一时刻状态量  $S_k$ 和动作  $A_k$ ；

根据公式  $R_k = S_k^T C S_k + A_k^T D A_k$  计算当前损失  $R_k$ ；

根据公式  $A_{k+1} = -(H_{uu})^{-1} H_{ux} S_{k+1} = L S_{k+1}$  计算当前时刻的动作量  $A_{k+1}$ 并输出；

根据公式  $Q(S_{k+1}, A_{k+1}) = \begin{bmatrix} S_{k+1}^T & A_{k+1}^T \end{bmatrix} \begin{bmatrix} H_{xx} & H_{xu} \\ H_{ux} & H_{uu} \end{bmatrix} \begin{bmatrix} S_{k+1} \\ A_{k+1} \end{bmatrix}$ ，计算当前  $Q(S_{k+1}, A_{k+1})$  函数；

根据公式  $e_k = \alpha [R_k + \gamma Q(S_{k+1}, A_{k+1}) - Q(S_k, A_k)]$  计算当前时刻的残差  $e_k$ ；

当连续20个时间步数内残差的绝对值均小于预设值时，即判定神经网络学习模型收敛，此时停止训练。

8. 根据权利要求7所述的智能车辆自动驾驶控制方法，其特征在于，

当步数k能够被m整除时，根据当前时刻的残差  $e_k$ ，采用梯度下降法调整更新当前神经网络的权重参数和偏置参数：

计算出均方误差： $E_k = \frac{1}{m} \sum_{k=0}^m \frac{1}{2} e_k^2 + \frac{\lambda}{2} \sum_{l=1}^{n_l-1} \sum_{i=1}^{s_{l+1}} \sum_{j=1}^{s_l} (W_{ij}^{(l)})^2$ ；其中，m为批量步长，表示每m

个仿真步长更新一次训练参数； $\lambda$ 为权重衰减系数，该项用于防止神经网络训练过程中过拟

合的发生。 $n_l$ 表示神经网络的层数。 $s_l$ 表示第 $l$ 层的节点数。 $W_{ij}^{(l)}$ 表示第 $l$ 层的第 $j$ 个节点与第 $l+1$ 层的第 $i$ 个节点间的权重参数;

将所述均方误差作为代价函数,采用梯度下降法调整更新当前神经网络的权重参数和偏置参数。

9. 根据权利要求1所述的智能车辆自动驾驶控制方法,其特征在于,根据所述动作量,得到智能车辆的底层控制量,并基于所述底层控制量控制智能车辆运行,包括:

将所述动作量转变为期望速度,将所述期望速度与实际车速的偏差输入至PID控制器,利用PID控制器处理得到节气门开度和制动油压百分比;

将所述节气门开度和制动油压百分比传递至车辆底层控制器,调节节气门开度和制动油压,实现对车辆的速度控制。

10. 一种智能车辆自动驾驶控制系统,其特征在于,包括:

路径分解模块,用于采集智能车辆全局行驶规划路径,将所述全局行驶规划路径分解为不同的行驶路段,并将不同的行驶路段按照驾驶任务划分为相应的驾驶子任务;

状态量生成模块,用于根据当前的驾驶子任务,采集所述驾驶子任务对应的环境信息,处理所述环境信息得到所述驾驶子任务对应的状态量;

动作量求解模块,用于将所述状态量输入至训练好的驾驶员行为学习模型中,经由所述驾驶员行为学习模型处理实时输出动作量;

执行模块,用于根据所述动作量,得到智能车辆的底层控制量,并基于所述底层控制量控制智能车辆运行。

## 一种智能车辆自动驾驶控制方法及系统

### 技术领域

[0001] 本发明涉及智能驾驶技术领域,尤其涉及一种智能车辆自动驾驶控制方法及系统。

### 背景技术

[0002] 随着传感技术、人工智能技术、计算机技术的发展,智能交通系统逐渐兴起。具有高度自动化水平的驾驶员行为学习系统引起了公众和研究机构的持续关注,并取得长足的发展与进步。

[0003] 在军事领域,驾驶员行为学习系统能有效避免危险工况下的人为作业,且对于开展高机动无人平台颇具益处;在民用领域,驾驶员行为学习系统可用于自动驾驶或辅助驾驶,能够有效提高车辆行驶安全性和交通通行能力。但是,现有的驾驶员行为学习系统仍存在许多问题,比如无法完成在线学习,自适应性较差,无法满足实时性需求,车辆平顺性、舒适性差等。

### 发明内容

[0004] 鉴于上述的分析,本发明旨在提供一种智能车辆自动驾驶控制方法及系统,用以解决现有自动驾驶无法很好地自适应完成在线学习的问题。

[0005] 本发明的目的主要是通过以下技术方案实现的:

[0006] 一种智能车辆自动驾驶控制方法,包括如下步骤:

[0007] 获取智能车辆全局行驶规划路径,将所述全局行驶规划路径分解为不同的行驶路段,并将不同的行驶路段按照驾驶任务划分为相应的驾驶子任务;

[0008] 根据当前的驾驶子任务,采集所述驾驶子任务对应的环境信息,处理所述环境信息得到所述驾驶子任务对应的状态量;

[0009] 将所述状态量输入至训练好的驾驶员行为学习模型中,经由所述驾驶员行为学习模型处理实时输出动作量;

[0010] 根据所述动作量,得到智能车辆的底层控制量,并基于所述底层控制量控制智能车辆运行。

[0011] 在上述方案的基础上,本发明还做了如下改进:

[0012] 进一步,将所述全局行驶规划路径分解为不同的行驶路段,并将不同的行驶路段按照驾驶任务划分为相应的驾驶子任务,包括:

[0013] 根据道路环境的不同,将所述全局行驶规划路径至少分解为直行路段、十字交叉路段;

[0014] 根据驾驶任务,将所述直行路段的驾驶子任务划分为车道保持子任务或车道变更子任务;将十字交叉路段的驾驶子任务划分为路口左/右转子任务或直行子任务。

[0015] 进一步,根据当前的驾驶子任务,采集所述驾驶子任务对应的环境信息,处理所述环境信息得到所述驾驶子任务对应的状态量,包括:

[0016] 当所述驾驶子任务为所述车道保持子任务时,采集的环境信息为:前车与本车的速度信息、距离信息;此时的状态量为:前车车速与本车车速之间的差值、两车间距与期望车距之间的差值;

[0017] 当所述驾驶子任务为所述车道变更子任务时,采集的环境信息为:本车和待变更车道上与本车相距预设范围内所有车辆的速度信息、距离信息;此时的状态量信息为:待变更车道上与本车相距预设范围内所有车辆中车速最大值与本车车速之间的差值、待变更车道上距本车最近的车辆与本车之间的距离;

[0018] 当所述驾驶子任务为所述路口左/右转子任务或直行子任务时,采集的环境信息为:本车和路口处与本车相距预设范围内所有车辆的位置、速度信息;此时的状态量信息为:本车和路口处与本车相距预设范围内所有车辆的位置、速度信息。

[0019] 进一步,所述驾驶员行为学习模型包括类型选择层、结构选择层、参数学习层;

[0020] 所述驾驶员行为学习模型在训练过程中,具体执行以下操作:

[0021] 根据当前的驾驶子任务,通过类型选择层选择与所述驾驶子任务相适应的神经网络类型、通过结构选择层选择与所述驾驶子任务相适应的神经网络结构参数,并将所述神经网络类型、神经网络结构参数选择结果发送至参数学习层;

[0022] 所述参数学习层根据所述神经网络类型、神经网络结构参数确定当前神经网络学习模型的结构;

[0023] 采用基于值函数的强化学习方法,训练所述参数学习层,直至所述神经网络学习模型收敛。

[0024] 进一步,所述类型选择层中可选的神经网络类型至少包括前馈神经网络、深度学习神经网络;

[0025] 当所述驾驶子任务为车道保持子任务或车道变更子任务时,所述类型选择层选择的神经网络类型为前馈神经网络;

[0026] 当所述驾驶子任务为路口左/右转子任务或直行子任务时,所述类型选择层选择的神经网络类型为深度学习神经网络。

[0027] 进一步,通过结构选择层选择与驾驶子任务相适应的神经网络结构参数,包括:

[0028] 根据选用的强化学习方法及当前驾驶子任务确定输入层、输出层的节点数;

[0029] 根据驾驶子任务的复杂程度确定隐含层层数:当所述驾驶子任务为车道保持子任务或车道变更子任务时,隐含层层数选用单层形式;当所述驾驶子任务为路口左/右转子

任务或直行子任务时,隐含层层数为多层;设定  $H_{xx} = \begin{bmatrix} w_1 & 0 \\ 0 & w_2 \end{bmatrix}$ 、 $H_{xu} = [w_3 \ w_4]^T$ 、 $H_{ux} = [w_3$

$w_4]$ 、 $H_{uu} = [w_5]$ ,从而得到Q函数  $Q(S_k, A_k) = \begin{bmatrix} S_k^T & A_k^T \end{bmatrix} \begin{bmatrix} H_{xx} & H_{xu} \\ H_{ux} & H_{uu} \end{bmatrix} \begin{bmatrix} S_k \\ A_k \end{bmatrix}$  的线性表示形

式,  $w_1$ 、 $w_2$ 、 $w_3$ 、 $w_4$ 、 $w_5$  通过神经网络模型的Q函数表达式对输入层节点求偏导解得,其中,  $S_k$  表示第k时刻的状态量,  $A_k$  表示第k时刻的动作用量。

[0030] 进一步,采用基于值函数的强化学习方法,训练所述参数学习层,直至所述神经网络学习模型收敛,包括:

[0031] 初始化参数学习层中神经网络的权重参数、激活函数参数,初始化 $Q(S_0, A_0) = 0$ 、状态量 $S_0$ 和动作量 $A_0$ ;

[0032] 对每个步数 $k$ ,执行:

[0033] 接收当前时刻状态量 $S_{k+1}$ ,获取上一时刻状态量 $S_k$ 和动作 $A_k$ ;

[0034] 根据公式 $R_k = S_k^T C S_k + A_k^T D A_k$ 计算当前损失 $R_k$ ;

[0035] 根据公式 $A_{k+1} = -(H_{uu})^{-1} H_{ux} S_{k+1} = L S_{k+1}$ 计算当前时刻的动作量 $A_{k+1}$ 并输出;

[0036] 根据公式 $Q(S_{k+1}, A_{k+1}) = \begin{bmatrix} S_{k+1}^T & A_{k+1}^T \end{bmatrix} \begin{bmatrix} H_{xx} & H_{xu} \\ H_{ux} & H_{uu} \end{bmatrix} \begin{bmatrix} S_{k+1} \\ A_{k+1} \end{bmatrix}$ ,计算当前 $Q(S_{k+1}, A_{k+1})$ 函

数;

[0037] 根据公式 $e_k = \alpha [R_k + \gamma Q(S_{k+1}, A_{k+1}) - Q(S_k, A_k)]$ 计算当前时刻的残差 $e_k$ ;

[0038] 当连续20个时间步数内残差的绝对值均小于预设值时,即判定神经网络学习模型收敛,此时停止训练。

[0039] 进一步,当步数 $k$ 能够被 $m$ 整除时,根据当前时刻的残差 $e_k$ ,采用梯度下降法调整更新当前神经网络的权重参数和偏置参数:

[0040] 计算出均方误差: $E_k = \frac{1}{m} \sum_{k=0}^m \frac{1}{2} e_k^2 + \frac{\lambda}{2} \sum_{l=1}^{n_l-1} \sum_{i=1}^{s_{l+1}} \sum_{j=1}^{s_l} (W_{ij}^{(l)})^2$ ;其中, $m$ 为批量步长,表

示每 $m$ 个仿真步长更新一次训练参数; $\lambda$ 为权重衰减系数,该项用于防止神经网络训练过程中过拟合的发生。 $n_l$ 表示神经网络的层数。 $s_l$ 表示第 $l$ 层的节点数。 $W_{ij}^{(l)}$ 表示第 $l$ 层的第 $j$ 个节点与第 $l+1$ 层的第 $i$ 个节点间的权重参数;

[0041] 将所述均方误差作为代价函数,采用梯度下降法调整更新当前神经网络的权重参数和偏置参数。

[0042] 进一步,根据所述动作量,得到智能车辆的底层控制量,并基于所述底层控制量控制智能车辆运行,包括:

[0043] 将所述动作量转变为期望速度,将所述期望速度与实际车速的偏差输入至PID控制器,利用PID控制器处理得到节气门开度和制动油压百分比;

[0044] 将所述节气门开度和制动油压百分比传递至车辆底层控制器,调节节气门开度和制动油压,实现对车辆的速度控制。

[0045] 本发明还提供了一种智能车辆自动驾驶控制系统,包括:

[0046] 路径分解模块,用于采集智能车辆全局行驶规划路径,将所述全局行驶规划路径分解为不同的行驶路段,并将不同的行驶路段按照驾驶任务划分为相应的驾驶子任务;

[0047] 状态量生成模块,用于根据当前的驾驶子任务,采集所述驾驶子任务对应的环境信息,处理所述环境信息得到所述驾驶子任务对应的状态量;

[0048] 动作量求解模块,用于将所述状态量输入至训练好的驾驶员行为学习模型中,经由所述驾驶员行为学习模型处理实时输出动作量;

[0049] 执行模块,用于根据所述动作量,得到智能车辆的底层控制量,并基于所述底层控制量控制智能车辆运行。

[0050] 本发明有益效果如下:本发明公开的智能车辆自动驾驶控制方法,通过将驾驶任

务划分为不同的驾驶子任务,并针对不同的驾驶子任务,通过与之相对应的神经网络模型的训练,得到相应的动作量,并根据该动作量调整智能车辆的运行状态,达到控制智能车辆自动驾驶的目的。能够实现在线学习,自适应性强,可以满足实时性需求,有效提升了车辆平顺性、舒适性。

[0051] 本发明中,上述各技术方案之间还可以相互组合,以实现更多的优选组合方案。本发明的其他特征和优点将在随后的说明书中阐述,并且,部分优点可从说明书中变得显而易见,或者通过实施本发明而了解。本发明的目的和其他优点可通过说明书、权利要求书以及附图中所特别指出的内容中来实现和获得。

## 附图说明

[0052] 附图仅用于示出具体实施例的目的,而并不认为是对本发明的限制,在整个附图中,相同的参考符号表示相同的部件。

[0053] 图1为本发明实施例1中的智能车辆自动驾驶控制方法流程图;

[0054] 图2为本发明中驾驶员行为学习模型训练过程示意图;

[0055] 图3为本发明实施例2中的智能车辆自动驾驶控制系统示意图。

## 具体实施方式

[0056] 下面结合附图来具体描述本发明的优选实施例,其中,附图构成本申请一部分,并与本发明的实施例一起用于阐释本发明的原理,并非用于限定本发明的范围。

[0057] 本发明的一个具体实施例,公开了一种智能车辆自动驾驶控制方法,流程图如图1所示,步骤如下:

[0058] 步骤S1:采集智能车辆全局行驶规划路径,将所述全局行驶规划路径分解为不同的行驶路段,并将不同的行驶路段按照驾驶任务划分为相应的驾驶子任务;具体地,根据道路环境的不同,将所述全局行驶规划路径至少分解为直行路段、十字交叉路段;根据驾驶任务,将所述直行路段的驾驶子任务划分为车道保持子任务或车道变更子任务;将十字交叉路段的驾驶子任务划分为路口左/右转子任务或直行子任务。

[0059] 步骤S2:根据当前的驾驶子任务,采集所述驾驶子任务对应的环境信息,处理所述环境信息得到所述驾驶子任务对应的状态量;具体地,

[0060] 当所述驾驶子任务为所述车道保持子任务时,采集的环境信息为:前车与本车的速度信息、距离信息;此时的状态量为:前车车速与本车车速之间的差值、两车间距与期望车距之间的差值;

[0061] 当所述驾驶子任务为所述车道变更子任务时,采集的环境信息为:本车和待变更车道上与本车相距预设范围内所有车辆的速度信息、距离信息;此时的状态量信息为:待变更车道上与本车相距预设范围内所有车辆中车速最大值与本车车速之间的差值、待变更车道上距本车最近的车辆与本车之间的距离;

[0062] 当所述驾驶子任务为所述路口左/右转子任务或直行子任务时,采集的环境信息为:本车和路口处与本车相距预设范围内所有车辆的位置、速度信息;此时的状态量信息为:本车和路口处与本车相距预设范围内所有车辆的位置、速度信息。

[0063] 步骤S3:将所述状态量输入至训练好的驾驶员行为学习模型中,经由所述驾驶员



行为学习模型处理实时输出动作量；

[0064] 步骤S4:根据所述动作量,得到智能车辆的底层控制量,并基于所述底层控制量控制智能车辆运行:将所述动作量转变为期望速度,将所述期望速度与实际车速的偏差输入至PID控制器,利用PID控制器处理得到节气门开度和制动油压百分比;此处的PID控制器还可替换为增量式PID控制器或MPC控制器;将所述节气门开度和制动油压百分比传递至车辆底层控制器,调节节气门开度和制动油压,实现对车辆的速度控制。

[0065] 上述方法的重点在于驾驶员行为学习模型的训练,因此,本发明给出了具体的训练驾驶员行为学习模型的方法。所述驾驶员行为学习模型包括类型选择层、结构选择层、参数学习层;

[0066] 所述驾驶员行为学习模型在训练过程中,具体执行以下操作:

[0067] 步骤S31:根据当前的驾驶子任务,通过类型选择层选择与所述驾驶子任务相适应的神经网络类型、通过结构选择层选择与所述驾驶子任务相适应的神经网络结构参数,并将所述神经网络类型、神经网络结构参数选择结果发送至参数学习层;其中,所述类型选择层中可选的神经网络类型至少包括前馈神经网络、深度学习神经网络;当所述驾驶子任务为车道保持子任务时或车道变更子任务时,所述类型选择层选择的神经网络类型为前馈神经网络;当所述驾驶子任务为路口左/右转子任务时,所述类型选择层选择的神经网络类型为深度学习神经网络。根据选用的强化学习方法及当前驾驶子任务确定输入层、输出层的节点数;根据驾驶子任务的复杂程度确定隐含层层数:当所述驾驶子任务为车道保持子任务时或车道变更子任务时,隐含层层数选用单层形式;当所述驾驶子任务为路口左/右转子

任务或直行子任务时,隐含层层数为多层;设定  $H_{xx} = \begin{bmatrix} w_1 & 0 \\ 0 & w_2 \end{bmatrix}$ 、 $H_{xu} = [w_3 \ w_4]^T$ 、 $H_{ux} = [w_3$   
 $w_4]$ 、 $H_{uu} = [w_5]$ ,从而得到Q函数  $Q(S_k, A_k) = \begin{bmatrix} S_k^T & A_k^T \end{bmatrix} \begin{bmatrix} H_{xx} & H_{xu} \\ H_{ux} & H_{uu} \end{bmatrix} \begin{bmatrix} S_k \\ A_k \end{bmatrix}$  的线性表示形式

$w_1$ 、 $w_2$ 、 $w_3$ 、 $w_4$ 、 $w_5$ 通过神经网络模型的Q函数表达式对输入层节点求偏导解得,其中, $S_k$ 表示第k时刻的状态量, $A_k$ 表示第k时刻的动作量。

[0068] 步骤S32:所述参数学习层根据所述神经网络类型、神经网络结构参数确定当前神经网络学习模型的结构;

[0069] 步骤S33:采用基于值函数的强化学习方法,训练所述参数学习层,直至所述神经网络学习模型收敛。具体地,

[0070] 初始化参数学习层中神经网络的权重参数、激活函数参数,初始化 $Q(S_0, A_0) = 0$ 、状态量 $S_0$ 和动作量 $A_0$ ;

[0071] 对每个步数k,执行:

[0072] 接收当前时刻状态量 $S_{k+1}$ ,获取上一时刻状态量 $S_k$ 和动作 $A_k$ ;

[0073] 根据公式  $R_k = S_k^T C S_k + A_k^T D A_k$  计算当前损失 $R_k$ ;

[0074] 根据公式  $A_{k+1} = -(H_{uu})^{-1} H_{ux} S_{k+1} = L S_{k+1}$  计算当前时刻的动作量 $A_{k+1}$ 并输出;

[0075] 根据公式  $Q(S_{k+1}, A_{k+1}) = \begin{bmatrix} S_{k+1}^T & A_{k+1}^T \end{bmatrix} \begin{bmatrix} H_{xx} & H_{xu} \\ H_{ux} & H_{uu} \end{bmatrix} \begin{bmatrix} S_{k+1} \\ A_{k+1} \end{bmatrix}$ , 计算当前 $Q(S_{k+1}, A_{k+1})$  函

数;

[0076] 根据公式 $e_k = \alpha [R_k + \gamma Q(S_{k+1}, A_{k+1}) - Q(S_k, A_k)]$ 计算当前时刻的残差 $e_k$ 。

[0077] 当步数 $k$ 可以被 $m$ 整除时,计算出均方误差: $E_k = \frac{1}{m} \sum_{k=0}^m \frac{1}{2} e_k^2 + \frac{\lambda}{2} \sum_{l=1}^{n_l-1} \sum_{i=1}^{s_{l+1}} \sum_{j=1}^{s_l} (W_{ij}^{(l)})^2$ ;

其中, $m$ 为批量步长,表示每 $m$ 个仿真步长更新一次训练参数; $\lambda$ 为权重衰减系数,该项用于防止神经网络训练过程中过拟合的发生。 $n_l$ 表示神经网络的层数。 $s_1$ 表示第1层的节点数。 $W_{ij}^{(l)}$ 表示第1层的第 $j$ 个节点与第 $l+1$ 层的第 $i$ 个节点间的权重参数;将所述均方误差作为代价函数,采用梯度下降法调整更新当前神经网络的权重参数和偏置参数。

[0078] 与现有技术相比,本实施例提供的智能车辆自动驾驶控制方法,通过将驾驶任务划分为不同的驾驶子任务,并针对不同的驾驶子任务,通过与之相对应的神经网络模型的训练,得到相应的动作量,并根据该动作量调整智能车辆的运行状态,达到控制智能车辆自动驾驶的目的。能够实现在线学习,自适应性强,可以满足实时性需求,有效提升了车辆平顺性、舒适性。

[0079] 实施例2

[0080] 本发明实施例2给出了另一种训练驾驶员行为学习模型的方法,如图2所示。

[0081] 当驾驶子任务为车道保持子任务或车道变更子任务时,由于该类学习任务较为简单,模型并不复杂。因此选择结构简单,求解速度快的前馈神经网络。输入层、输出层节点数根据所采取的强化学习方法而定。隐含层层数优先选用单层形式,以简化模型,避免出现过拟合。隐含层节点数可根据经验公式( $m = \sqrt{nl}$ ,  $m$ 为隐含层节点数,  $n$ 为输入层节点数,  $l$ 为输出层节点数)确定。输出层的激活函数通常选择线性函数以简化训练,输入层和隐含层的激活函数通常选择tanh型激活函数。例如,采取Neural Q-Learning的强化学习方法时,将神经网络模型所逼近的Q函数进行参数简化假设,得到Q函数的线性形式,进而确定输入层节点个数,输出层节点为Q值。当驾驶子任务为路口左/右转子任务或直行子任务时,由于该类学习任务需要考虑多名交通参与者的交互情况,整个过程较复杂。因此选择隐含层数较多的深度神经网络。输入层、输出层节点数根据所采取的强化学习方法而定。隐含层层数优先选用双层形式,以简化模型,避免出现过拟合。隐含层节点数可根据经验公式( $m = \sqrt{nl}$ ,  $m$ 为隐含层节点数,  $n$ 为输入层节点数,  $l$ 为输出层节点数)确定。

[0082] 选择好神经网络类型及神经网络输入层节点数、输出层节点数、隐含层个数及激活函数的形式后,即可初始化训练参数值并建立神经网络模型来逼近Q函数。

[0083] 以驾驶子任务为车道保持子任务为例,Q函数及神经网络模型建立的示例如下:

[0084] 步骤1、根据强化学习方法,定义状态量、动作量及Q函数。将Q函数定义为

$$Q(S_k, A_k) = \begin{bmatrix} S_k^T & A_k^T \end{bmatrix} \begin{bmatrix} H_{xx} & H_{xu} \\ H_{ux} & H_{uu} \end{bmatrix} \begin{bmatrix} S_k \\ A_k \end{bmatrix}。状态量定义为S_k = [s_1 \ s_2]^T。其中, s_1 = v_{1k} -$$

$v_{hk}$  ( $v_{1k}$ 为前车车速,  $v_{hk}$ 为本车车速)和 $s_2 = d_k - d_E$  ( $d_k$ 为两车间距,  $d_E$ 为期望车距)。动作量定义为 $A_k$ 。 $H_{xx}$ 、 $H_{xu}$ 、 $H_{ux}$ 、 $H_{uu}$ 是与系统损失相关的二次型矩阵。

[0085] 步骤2、对Q函数相关参数进行适当简化假设,构建线性函数模型。假定

$H_{xx} = \begin{bmatrix} w_1 & 0 \\ 0 & w_2 \end{bmatrix}$ 、 $H_{xu} = [w_3 \ w_4]^T$ 、 $H_{ux} = [w_3 \ w_4]$ 、 $H_{uu} = [w_5]$ ，代入上述Q函数得到线性模型：Q

$$= w^T x = [w_1 \ w_2 \ w_3 \ w_4 \ w_5] [s_1^2 \ s_2^2 \ 2s_1a \ 2s_2a \ a^2]^T。$$

[0086] 步骤3、根据线性Q函数模型，建立神经网络模型。根据步骤S203，采用一个前馈神经网络模型来近似Q函数。该前馈神经网络模型是一个具有单个隐含层的三层全连接式的神经网络。输入层具有五个节点 ( $x = [x_1 \ x_2 \ x_3 \ x_4 \ x_5]^T = [s_1^2 \ s_2^2 \ 2s_1a \ 2s_2a \ a^2]^T$ )，隐含层具有三个节点，输出层具有一个节点 (Q(x))，训练参数具体包括从输入层各个节点到隐含层各个节点、从隐含层各个节点到输出层各个节点的共18个权重参数和隐含层各个节点处的3个偏置参数。初始训练参数取[-2,1]范围内的随机数。

[0087] 由训练数据集，在每个学习步中由奖励函数计算得到奖励值。根据选择的基于值函数的强化学习方法中的Q值更新公式，进而计算上一学习步（一定状态、一定动作对应）的期望Q值。期望Q值与实际Q值的差值，即残差。进而计算均方误差 (MSE)，使用梯度下降法对神经网络学习模型训练参数进行更新训练。

[0088] 利用梯度下降法更新参数的具体步骤如下：

[0089] 步骤1、计算奖励值。由训练数据集，在每个时刻由奖励函数计算得到奖励值，奖励函数的定义为  $R_k = S_k^T C S_k + A_k^T D A_k$  (k表示不同时刻，C和D为正定矩阵，分别表示状态量和动作量对 $R_k$ 的权重)。

[0090] 步骤2、计算残差及均方误差。在第k时刻，期望Q值与实际Q值的差值，即为残差  $e_k = \alpha [R_{k+1} + \gamma Q(S_{k+1}, A_{k+1}) - Q(S_k, A_k)]$  ( $\alpha$ 为学习率， $\gamma$ 为折扣系数)。进而可计算出均方误差

$$(MSE) E_k = \frac{1}{m} \sum_{k=0}^m \frac{1}{2} e_k^2 + \frac{\lambda}{2} \sum_{l=1}^{n_l-1} \sum_{i=1}^{s_{l+1}} \sum_{j=1}^{s_l} (W_{ij}^{(l)})^2 \quad (\text{其中, } m \text{ 为批量步长, 表示每 } m \text{ 个仿真步长更新}$$

一次训练参数。 $\lambda$ 为权重衰减系数，该项用于防止神经网络训练过程中过拟合的发生。 $n_l$ 表示神经网络的层数。 $s_l$ 表示第l层的节点数。 $W_{ij}^{(l)}$ 表示第l层的第j个节点与第l+1层的第i个节点间的权重参数)。

[0091] 步骤3、将MSE作为代价函数，使用梯度下降法对神经网络学习模型训练参数进行更新训练。首先要明确的是，以Neural Q-Learning强化学习方法为例，神经网络学习模型的参数训练过程实现的目标是第k时刻Q值的函数逼近。神经网络权重参数和偏置参数的更新公式分别为 ( $\alpha$ 为学习率)：

$$[0092] \quad W_{ij}^{(l)} = W_{ij}^{(l)} - \alpha \frac{\partial}{\partial W_{ij}^{(l)}} E(W, b) = W_{ij}^{(l)} - \alpha \left( \frac{1}{m} \sum_{k=1}^N \frac{\partial}{\partial W_{ij}^{(l)}} E_k(W, b) + \lambda W_{ij}^{(l)} \right)$$

$$[0093] \quad b_{ij}^{(l)} = b_{ij}^{(l)} - \alpha \frac{\partial}{\partial b_{ij}^{(l)}} E(W, b) = b_{ij}^{(l)} - \alpha \frac{1}{m} \sum_{k=1}^m \frac{\partial}{\partial b_{ij}^{(l)}} E_k(W, b)$$

[0094] 此外，训练结果还可以对神经网络结构优化提供依据，例如比较不同隐含层节点数、层数或激活函数类型的神经网络模型的收敛训练速度、收敛值等。当连续20个时间步数内残差的绝对值均小于预设值时，即判定神经网络学习模型收敛，此时停止训练。

[0095] 步骤4、动作量求解模型的具体实现过程。

[0096] Q函数的真值函数定义为  $Q(S_k, A_k) = [S_k^T \quad A_k^T] \begin{bmatrix} H_{xx} & H_{xu} \\ H_{ux} & H_{uu} \end{bmatrix} \begin{bmatrix} S_k \\ A_k \end{bmatrix}$ ，本示例选

择使用贪心算法来进行动作量的求解。贪心算法的本质为选择某一时刻一定状态下最大Q值对应的动作量作为输出动作量。对于神经网络模型来说，在某一时刻一定状态下，求解Q函数的极值，极值所对应动作量即为输出动作量。对上述真值函数求解极值，可推导动作量

求解模型表达式为  $A_k = -(H_{uu})^{-1} H_{ux} S_k = -\frac{w_3 S_1 + w_4 S_2}{w_5}$ 。w<sub>3</sub>、w<sub>4</sub>、w<sub>5</sub>的值可由神经网络模型

表达式分别对x<sub>3</sub>、x<sub>4</sub>、x<sub>5</sub>求偏导解得。

[0097] 实施例3

[0098] 本发明还公开了一种智能车辆自动驾驶控制系统，结构示意图如图3所示，包括：路径分解模块，用于采集智能车辆全局行驶规划路径，将所述全局行驶规划路径分解为不同的行驶路段，并将不同的行驶路段按照驾驶任务划分为相应的驾驶子任务；状态量生成模块，用于根据当前的驾驶子任务，采集所述驾驶子任务对应的环境信息，处理所述环境信息得到所述驾驶子任务对应的状态量；动作量求解模块，用于将所述状态量输入至训练好的驾驶员行为学习模型中，经由所述驾驶员行为学习模型处理实时输出动作量；执行模块，用于根据所述动作量，得到智能车辆的底层控制量，并基于所述底层控制量控制智能车辆运行。

[0099] 本发明中系统实施例的具体实施过程参见上述方法实施例即可，本实施例在此不再赘述。由于本实施例与上述方法实施例原理相同，所以本系统也具有上述方法实施例相应的技术效果。

[0100] 本领域技术人员可以理解，实现上述实施例方法的全部或部分流程，可以通过计算机程序来指令相关的硬件来完成，所述的程序可存储于计算机可读存储介质中。其中，所述计算机可读存储介质为磁盘、光盘、只读存储记忆体或随机存储记忆体等。

[0101] 以上所述，仅为本发明较佳的具体实施方式，但本发明的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本发明揭露的技术范围内，可轻易想到的变化或替换，都应涵盖在本发明的保护范围之内。

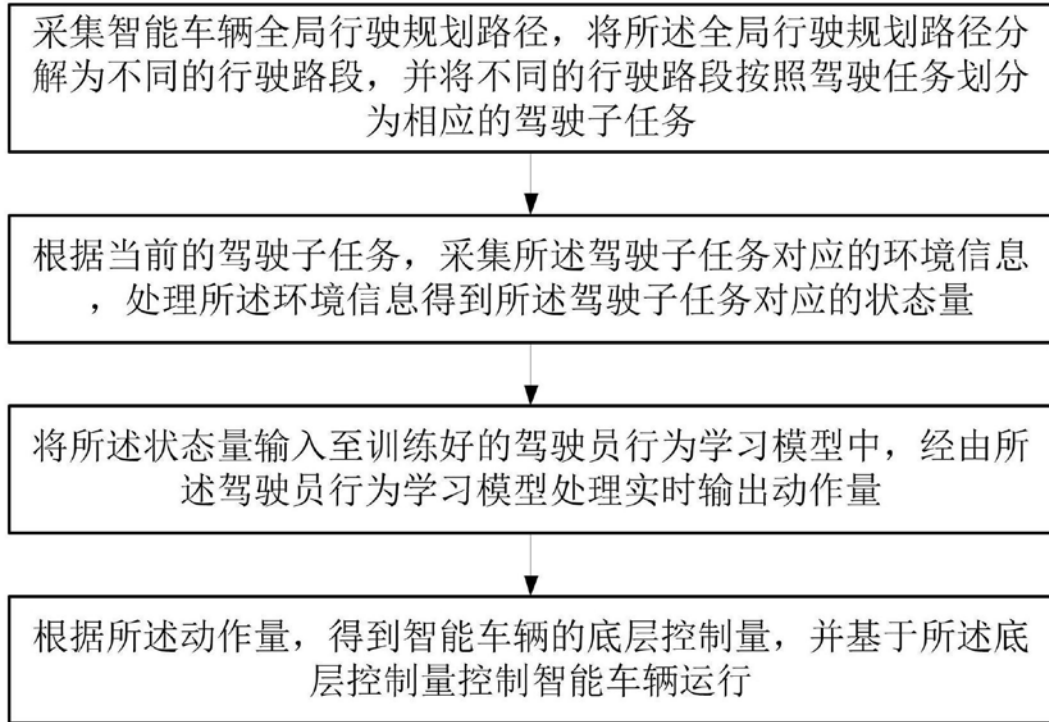


图1

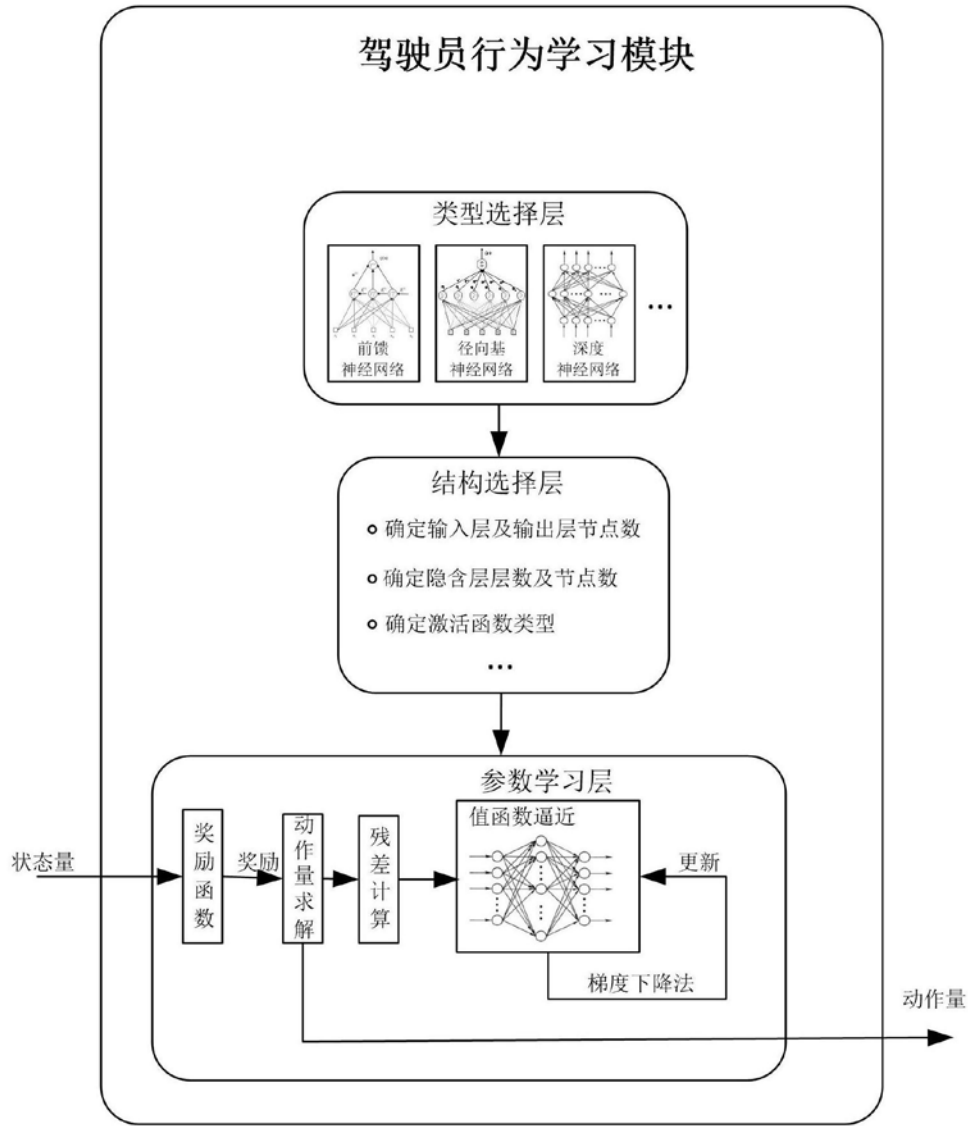


图2

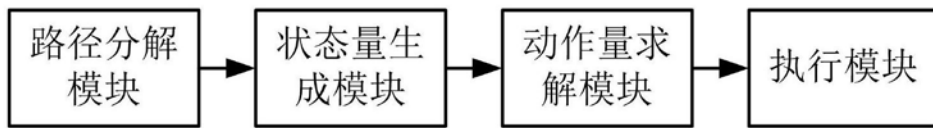


图3