



(12) 发明专利

(10) 授权公告号 CN 111966118 B

(45) 授权公告日 2021.07.20

(21) 申请号 202010820394.0

审查员 魏利君

(22) 申请日 2020.08.14

(65) 同一申请的已公布的文献号

申请公布号 CN 111966118 A

(43) 申请公布日 2020.11.20

(73) 专利权人 哈尔滨工程大学

地址 150001 黑龙江省哈尔滨市南岗区南通大街145号

(72) 发明人 孙玉山 吴新雨 张国成 李岳明

王元庆 祁彧 柴璞鑫

(74) 专利代理机构 哈尔滨市阳光惠远知识产权

代理有限公司 23211

代理人 刘景祥

(51) Int. Cl.

G05D 1/06 (2006.01)

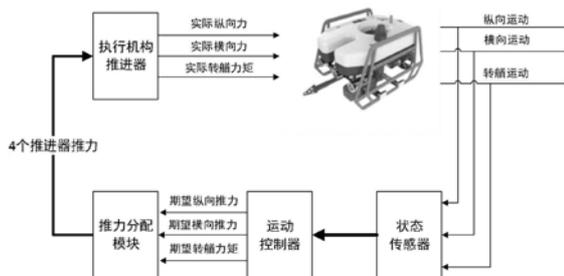
权利要求书5页 说明书14页 附图2页

(54) 发明名称

一种ROV推力分配与基于强化学习的运动控制方法

(57) 摘要

本发明的一种ROV推力分配与基于强化学习的运动控制方法,属于水下机器人技术领域。包括:进行ROV动力学建模,确定ROV推进器空间布置及推进系统模型,提出ROV推力分配方法;在所述ROV推力分配方法的基础上,加入DDPG控制器;在所述DDPG控制器的基础上,加入抗扰控制器。该ROV推力分配方法在进行推力分配的同时降低推进器能耗,从而降低了整个ROV的能耗。DDPG算法能够在动作输出幅度能进行连续值输出,更适用水下机器人控制问题,同时DDPG控制器能够在连续决策问题中进行学习。采用抗扰控制器可以提高该运动控制方法对ROV未建模水动力、外界扰动及其他不确定因素的抗扰能力。



1. 一种ROV推力分配与基于强化学习的运动控制方法,其特征在于,所述方法包括以下步骤:

步骤一、进行ROV动力学建模,确定ROV推进器空间布置及推进系统模型,提出ROV推力分配方法;

步骤二、在所述ROV推力分配方法的基础上,加入DDPG控制器;

步骤三、在所述DDPG控制器的基础上,加入抗扰控制器,

在步骤一中,具体的,建立有利于描述水下机器人运动状态的坐标系,惯性坐标系  $E-\xi\eta\zeta$  以空间任意一点为原点,  $E\xi$  轴正向指向地心,  $E\xi$  轴指向北向、 $E\eta$  指向东向,随体坐标系  $0-xyz$  固连于ROV艇体上,原点取在ROV的对称轴中点或在重心上,取纵轴  $0x$  指向艇首,横轴  $0y$  与  $0x$  轴正交于同一平面并以右舷为正方向,垂轴  $0z$  指向机器人的腹部,机器人的位置和方向由惯性坐标系表示,速度和角速度在随体坐标系中表示,

在惯性坐标系中,ROV位置和姿态表示为:

$$\mathbf{R} = [\mathbf{r}^T, \mathbf{\Lambda}^T]^T, \mathbf{r} = [\xi, \eta, \zeta]^T, \mathbf{\Lambda} = [\varphi, \theta, \psi]^T \quad (1)$$

其中: $\mathbf{R}$ 为ROV的位姿项, $\mathbf{r}$ 为ROV的位置项, $\mathbf{\Lambda}$ 为ROV的姿态项, $\xi$ 为大地坐标系纵向, $\eta$ 为大地坐标系横向, $\zeta$ 为大地坐标系垂向, $\varphi$ 为横倾角, $\theta$ 为纵倾角, $\psi$ 为艏向角,

在随体坐标系中,水下机器人的动力学模型输入为所受环境和控制力的力和力矩,输出为水下机器人的线性加速度和姿态角加速度,在随体坐标系中,载体的速度和加速度为:

$$\mathbf{V} = [\mathbf{U}^T, \mathbf{\Omega}^T]^T, \mathbf{U} = [u, v, w]^T, \mathbf{\Omega} = [p, q, r]^T \quad (2)$$

其中: $\mathbf{V}$ 为ROV的速度项, $\mathbf{U}$ 为ROV的线性速度项, $\mathbf{\Omega}$ 为ROV的角速度项, $u$ 为纵向速度, $v$ 为横向速度, $w$ 为垂向速度, $p$ 为横滚角速度, $q$ 为纵倾角速度, $r$ 为转艏角速度,

水下机器人在惯性系下的位置项与其随体的速度量的动态关系可以由以下转换方程来描述:

$$\dot{\mathbf{R}} = \begin{bmatrix} \mathbf{J}_1 & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} & \mathbf{J}_2 \end{bmatrix} \mathbf{V} \quad (3)$$

式中 $\mathbf{J}_1, \mathbf{J}_2$ 是坐标转换矩阵:

$$\mathbf{J}_1 = \begin{bmatrix} \cos \psi \cos \theta & \cos \psi \sin \theta \sin \varphi - \sin \psi \cos \varphi & \cos \psi \sin \theta \cos \varphi + \sin \psi \sin \varphi \\ \sin \psi \cos \theta & \sin \psi \sin \theta \sin \varphi + \cos \psi \cos \varphi & \sin \psi \sin \theta \cos \varphi - \cos \psi \sin \varphi \\ -\sin \theta & \cos \theta \sin \varphi & \cos \theta \cos \varphi \end{bmatrix} \quad (4)$$

$$\mathbf{J}_2 = \begin{bmatrix} 1 & \tan \theta \sin \varphi & \tan \theta \cos \varphi \\ 0 & \cos \varphi & -\sin \varphi \\ 0 & \sin \varphi \sec \theta & \cos \varphi \sec \theta \end{bmatrix} \quad (5)$$

通过有限元计算或船模实验可以对动力学方程中涉及的水动力参数进行测算;推进系统是由多个处于不同方位的独立推进器所组成,对单个推进器进行数学建模,对于ROV上所使用的螺旋桨,其敞水推力的计算方法如下式所示:

$$T = K_T \rho n^2 D^4 \quad (6)$$

式中: $K_T$ 为螺旋桨推力系数, $\rho$ 为海水的密度 $\text{kg}/\text{m}^3$ , $D$ 为螺旋桨直径 $\text{m}$ , $n$ 为螺旋桨的转速 $\text{r}/\text{s}$ ,

为了获得相应的推力,必须驱动螺旋桨转动推出水体,螺旋桨的转矩的计算公式为:

$$Q=K_q \rho n^2 D^5 \quad (7)$$

式中:Q为螺旋桨扭矩Nm, $K_q$ 为螺旋桨转矩系数,

水下机器人配置有6台矢量布置的螺旋桨推进器,在水平面有四个斜向布置的水平面推进器,垂直面沿艇体纵轴关于质心对称布置有两个垂向推进器,

在水平方向上,4台推进器平行于水平面0-xy布置,按照左上、右上、左下、右下依次编号为:第一推进器、第二推进器、第三推进器和第四推进器,且与随体坐标系纵轴0-x的夹角为 $45^\circ$ ,水平方向上的4台推进器安装位置关于纵轴左右对称,因此水平面的推进器为冗余布置,对于控制器计算的控制量通过各推进器的组合实现,第一推进器的螺旋桨在随体坐标系x0y中的位置为:(237mm, -305mm),第二推进器与第一推进器关于0-xz面对称;第三推进器的坐标为(-300mm, -195mm),第四推进器与第三推进器关于0-xz面对称,

垂直面有两个推进器,其布置在0-xz平面上,依照前后方向编号为第五推进器和第六推进器,其中第五推进器在x0z坐标系中位置为(284mm, 113mm),第六推进器与第五推进器其关于0-yz面对称布置,

遥控水下机器人具有六个推进器,其中要控制的量为纵向横移垂向方向的运动和转艏,其中垂直方向的控制器有两个均匀布置,推力分配时只需将控制力平均分配即可,水平方向有四个推进器,分布在四个角,与随体坐标系成 $45^\circ$ 排布,在分配水平方向的力时,推进器数目大于目标力和力矩的数量,推进器冗余布置,因此在满足推力分配时,可以有无数种分配方式,这要求在进行推力分配的同时,应当兼顾各推进器输出幅值的偏差以及推力分配系统的能耗,

建立水下机器人推力分配的数学模型:

$$\tau=B(\alpha) T \quad (8)$$

式(8)为推力分配问题的等式约束,B表示推力分配的空间位置约束, $\tau=\{\tau_x, \tau_y, \tau_r\}$ 为控制器输出的三个自由度上的力和力矩, $T=\{T_1 T_2 T_3 T_4\}$ 为推进器的推力, $\alpha=\{\alpha_1 \alpha_2 \alpha_3 \alpha_4\}$ 为推进器转角,

$$B(\alpha)=\begin{bmatrix} \sin \alpha_1 & \sin \alpha_2 & \sin \alpha_3 & \sin \alpha_4 \\ \cos \alpha_1 & -\cos \alpha_2 & -\cos \alpha_3 & \cos \alpha_4 \\ \sin \alpha_1 l_{1y} + \cos \alpha_1 l_{1x} & -\sin \alpha_2 l_{2y} - \cos \alpha_2 l_{2x} & \sin \alpha_3 l_{3y} + \cos \alpha_3 l_{3x} & -\sin \alpha_4 l_{4y} - \cos \alpha_4 l_{4x} \end{bmatrix}$$

(9)

其中 $l_{xi}$ 、 $l_{yi}$ 分别为第i个推进器在x、y方向上距离机器人重心的距离;

遗传算法借鉴生物进化理论的机理,以保留更适应环境的生物体为筛选原则的一种随机算法,先使用随机选取的一组解作为初始种群,该种群中的每个解决方案被称为个体或染色体,种群大小是其中个体数目,遗传算法对变量进行编码,遗传算法中每个染色体的适应值,具有评估新产生的种群与上一代种群关系的能力,为了得到每个染色体的适应值,计算同一染色体的优化函数的值,所述优化函数叫做目标函数,通过遗传算法中主要的选择、交叉和变异这种处理使种群个体进行更新,得到下一次迭代的解群,新产生的种群叫子代,上一代种群叫父代,遗传算法的这个迭代过程直到满足停止条件才会终止,在最后一代中具有最佳适应值的染色体被称为最优解,

根据推进器系统模型,建立遗传算法的优化目标方程:

$$\min \text{fun}(T) = \sum_{i=1}^4 k_i P(T_i) + T^T W T \quad (9)$$

其中P为推进系统的能耗方程,  $k_i$ 为第i个推进器占总能耗的占比权值, 在本目标机器人中, 总功率为各个推进器功率之和, 根据电机功率公式, 其中n为转速, Q为扭矩:

$$P = nQ \quad (10)$$

将上式与螺旋桨的敞水推力的计算公式(11)、螺旋桨的转矩公式(12)结合:

$$T = K_T \rho n^2 D^4 \quad (6)$$

式中,  $K_T$ 为螺旋桨推力系数,  $\rho$ 为海水的密度 $\text{kg/m}^3$ , D为螺旋桨直径m, n为螺旋桨的转速r/s,

为了获得相应的推力, 必须驱动螺旋桨转动推出水体, 其转矩的计算公式为:

$$Q = K_Q \rho n^2 D^5 \quad (7)$$

式中, Q为螺旋桨扭矩Nm,  $K_Q$ 为螺旋桨转矩系数,

计算每个推进器的推力和功率量化公式:

$$P = K_q \frac{T^{\frac{3}{2}}}{K_T^{\frac{3}{2}} \rho^{\frac{1}{2}} D} \quad (13)$$

单个推进器功率与推进器推力的3/2次方成正比,

将控制分配任务抽象成最优化问题进行描述:

$$\begin{aligned} \arg \min_u \text{fun}(u) \\ BT = \tau \end{aligned} \quad (14)$$

开始遗传算法计算, 确定遗传算法中优化变量为四项, 即推进器的推力大小u, 生成初始种群, 设计种群中个体数目为50, 进行500次迭代, 自变量数目4, 交叉概率0.8, 变异概率0.15, 公差为 $1e-8$ , 对于个体的变量上下限为 $[-20, 20]$ , 初始种群的个体随机产生, 随机范围即为变量上下限,

在优化目标函数fun的基础上设计遗传算法的适应度函数V并计算:

$$V = \frac{1}{\text{fun} + (BT - \tau)^T W (BT - \tau)} \quad (15)$$

上式中 $Bu - \tau$ 表示当前个体关于等式约束的偏差值, W为惩罚矩阵, 当个体越接近等式约束的解时, W的适应度也越大, 同样当个体越满足优化条件时, 它的适应度越大, 这样便可以求出种群中更加满足约束条件且适应优化方程的个体,

根据适应度选择当前种群中个体, 根据适应程度对最佳个体进行分类, 记录最佳个体, 并以如下公式选择计算群体中个体的选择概率:

$$P_i = \frac{f_i}{\sum_{i=1}^n f_i} \quad (16)$$

式中 $P_i$ 为第i个个体的选择概率,  $f_i$ 为其适应度, 拥有较大适应度的解有更大的概率被选择继承到下一次迭代, 在获取下一次迭代种群时对选中个体进行交叉变异提高在优值周围的探索,

对遗传算法的终止条件进行判断,如果当前最佳个体的适应度与上代最佳个体适应度之差小于容忍度,则认为达到最优条件,遗传算法结束;若不满足,判断是否达到繁殖代数,若达到则停止计算,

对有交叉权的个体进行染色体交叉,采用浮点数交叉法,即交换父母中某一段浮点位,对有交叉权但不满足交叉概率的保留父代,对于产生的子代,依照编译概率对其某浮点位进行随机数变异,在产生的子代中,如果它的变量值超过了变量上下限,则将其值修改为上下限的值,同时为了保证种群的优选性,将上代最优个体进行保留,将交叉、变异后的产生的新种群回到计算适度函数V的步骤,进行下一轮的计算,

结束计算后,获得的最优个体u即为本次推力分配问题中能耗最小的优化推力;

在步骤二中,针对的水下机器人在水平面操纵上具有纵轴和艏向上的执行机构,通过调整前后方向的推力和转向使机器人趋近目标点,控制原理为获得目标值和实际值的差值,将偏差作为强化学习的状态值,使用策略网络选择所需执行的动作,其中动作空间为归一化以后的x方向推力和绕z轴旋转的力矩,

在控制器设计时,建立Actor网络作为策略网络 $\mu$ ,某一时刻的惯性坐标系位置和随体坐标系速度 $R, V$ 作为状态空间的输入 $s = [x, y, \varphi, u, v, w]$ ,通过神经网络输出控制指令,其包括纵向、横向和转艏方向上归一化的力和力矩,动作空间为: $a = \{\tau_x, \tau_y, \tau_r\}$ 将选择的动作交由推力分配系统生成实际的推力分配和推进器转速控制方案,通过调节机器人纵向和横向的速度和艏向角使机器人向目标点靠拢,

控制器状态选择随体坐标系下的误差 $s = \{e_{xr}, e_{yr}, e_\psi, u, v, r\}$ ,  $e_{xr}$ 为随体坐标系下纵向误差: $e_{xr} = x_{Target} - x_r$ ,  $e_{yr}$ 为横向误差,  $e_\psi$ 为大地坐标系下目标艏向和当前艏向的差值,目标艏向: $\psi_{Target} = \arctan(dy/dx)$ ,因此艏向误差可表示为:

$$e_\psi = \psi_{Target} - \psi \quad (17)$$

以点镇定为目标,设置初始目标艏向角为0,偏差为目标点在机器人艇体坐标系纵向距离,将目标速度和角速度设置为0,则设置强化学习的奖励函数为:

$$r = -\sqrt{\Delta x^2 + \Delta y^2} \quad (18)$$

当智能体越靠近智能体,且智能体的速度越慢时,奖励值就越大,以累积折扣奖励值最大为目标,通过DDPG方法使策略网络选择最优策略,当某一时刻输入智能体状态时,它能输出最佳动作,进而完成机器人点镇定的控制问题,

当传感器获得水下机器人状态信息 $s$ 后,根据状态 $s$ ,策略网络会选择动作控制水下机器人运动,在下一采样时刻获得状态信息 $s'$ ,同时根据奖励函数计算当前状态的奖励值 $r$ ,将 $\langle s, a, r, s' \rangle$ 作为一个采样点储存在记忆库中,在参数训练的过程中,从记忆库小批量采样,使用策略网络进行决策,并用目标评价计算采样动作的目标Q值,通过现实评价估计Q值与目标Q值的均方误差调整评价网络参数更新,同时使用目标Q函数计算更新策略网络,采用软更新的方式,异步的调整目标神经网络的参数,使目标神经网络向接近现实神经网络的方向缓慢更新,在训练过程中使用策略网络以状态信息 $s$ 为输入,输出3个自由度上的力和力矩 $a$ ,并加入高斯噪声以实现对环境的探索,将策略网络选择的动作 $a$ 发送给推力分配模块,进而驱动ROV推进器运转,实现动作控制;

在步骤三中,具体的,抗扰控制器以状态观测器为基础,依据观测器估计的状态和测量

状态误差为输入量调整控制器,水下机器人在控制力的作用下由状态 $s_t$ 转移至 $s_{t+1}$ ,而由于外部扰动和系统不确定性的存在使状态转移至 $s'_{t+1}$ ,将干扰后的状态量与估计状态量的误差作为输入,加入控制力作为线性补偿以改善控制器的抗扰性能,

以随体坐标系下位置量为被控状态量 $s = \{x, y, \psi\}$ ,根据2个采样时刻状态量 $\{s_{k-1}, s_k\}$ 和控制量 $\{\tau_{k-1}, \tau_k\}$ 使用SVR方法状态估计水下机器人状态 $s_{k+1}$ ,在无干扰情况下采集运动状态数据作为数据集训练SVR收敛,并以此来估计理想状态下被控状态,则状态误差为测量的状态与估计状态的差值:

$$e_s(k) = \hat{s}(k) - s(k) \quad (19)$$

根据反馈控制的思想,使用非线性反馈控制器对系统误差进行补偿,对于非线性系统加入非线性补偿控制量:

$$\dot{x} = f(x) + w(t) + \tau(t) + g(e) \quad (20)$$

其中, $x$ 为非线性系统状态量, $w$ 为时域干扰项, $\tau$ 为控制量, $g(e)$ 为以估计误差为输入的补偿控制量,通过调整控制量来调整期望状态与系统输出的误差,对于一阶误差输入,取 $g(\cdot)$ 为fal函数[110],

$$g(e_s) = \beta * fal(e_s, \alpha, \delta) = \begin{cases} |e_s|^\alpha \operatorname{sgn}(e_s) & |e_s| > \delta \\ \frac{e_s}{\delta^{1-\alpha}} & |e_s| < \delta \end{cases} \quad (21)$$

其中, $\beta$ 为比例参数,其反映了补偿控制力在总控制力中的比重, $\delta$ 为反馈函数线性段的范围,当 $|e_s| < \delta$ 时,反馈函数处于线性区间,避免了误差在0点附近时的高频振荡,减小稳态误差,取 $\alpha$ 在0~1之间,反映了系统的阶次,在 $|e_s| > \delta$ 的区间中,当误差增幅大时,减小系统增益防止发散, $\delta$ 取值较小但其值过小会导致线性段缩短,使系统出现高频抖震,取值过大会导致线性段延长,使非线性段克服系统稳态误差的特性降低。

## 一种ROV推力分配与基于强化学习的运动控制方法

### 技术领域

[0001] 本发明涉及基于一种ROV推力分配与基于强化学习的运动控制方法,属于水下机器人技术领域。

### 背景技术

[0002] 随着海洋技术的发展,水下探测、水下作业等任务的需求越来越大,无人设备在水下工程的地位愈发重要,其中水下机器人技术的研究与应用受到国内外研究者的重视。其中由于遥控水下机器人(ROV)通常用于水下作业任务,需要ROV在复杂水体环境中长时间稳定工作,因此如何使ROV控制更加精准可靠受到很多研究人员的青睐。

[0003] 通常水下机器人控制器设计时需要建立运动系统模型,根据模型设计控制器形式并调整控制器参数。但由于水下机器人的运动学模型由多方面因素影响,其运动系统是多阶非线性的,同时多自由度交叉耦合现象加大了水下机器人的运动控制的难度,并且水体环境复杂多变,使水下机器人运动学模型难以用于对实际工程的指导。

[0004] 常见的ROV推进分配方法有伪逆分配法、数学规划法等,伪逆分配法不需要考虑执行能力范围和位置饱和约束,会出现计算结果与实际预期不符的现象,出现推进器饱和和欠缺等问题,并且以该方法不能考虑能耗函数或其他优化目标,此方法不能执行在分配的同时进行优化的任务,导致机器人能量浪费等损失。数学规划法求解最优问题较为繁复,计算速度较慢,可能有实时性不足的问题,使用该方法有在局部极值点处进行迭代不能收敛到全局最优的问题,而且在求解时由于放宽了限制条件,可能造成分配后推力无法达到预期效果等限制。

[0005] 该发明提出了一种基于遗传算法的推力分配方法,在进行推力分配的同时降低推进器能耗,之后提出了一种基于强化学习的运动控制方法进行ROV的运动控制,实现了ROV运动的精确控制。

### 发明内容

[0006] 本发明的目的是提出一种ROV推力分配与基于强化学习的运动控制方法,针对传统设计模式的缺点,本文以搭载矢量布置推进器的遥控水下机器人为载体,研究推力分配节能优化方案,和使用强化学习对ROV进行运动控制。

[0007] 一种ROV推力分配与基于强化学习的运动控制方法,所述方法包括以下步骤:

[0008] 步骤一、进行ROV动力学建模,确定ROV推进器空间布置及推进系统模型,提出ROV推力分配方法;

[0009] 步骤二、在所述ROV推力分配方法的基础上,加入DDPG控制器;

[0010] 步骤三、在所述DDPG控制器的基础上,加入抗扰控制器。

[0011] 在步骤一中,具体的,建立有利于描述水下机器人运动状态的坐标系,惯性坐标系  $E-\xi\eta\zeta$  以空间任意一点为原点,  $E\xi$  轴正向指向地心,  $E\xi$  轴指向北向、 $E\eta$  指向东向,随体坐标系  $0-xyz$  固连于ROV艇体上,原点取在ROV的对称轴中点或在重心上,取纵轴  $Ox$  指向艇首,横

轴 $0y$ 与 $0x$ 轴正交于同一平面并以右舷为正方向,垂轴 $0z$ 指向机器人的腹部,机器人的位置和方向由惯性坐标系表示,速度和角速度在随体坐标系中表示,

[0012] 在惯性坐标系中,ROV位置和姿态表示为:

$$[0013] \quad \mathbf{R} = [\mathbf{r}^T, \mathbf{\Lambda}^T]^T, \quad \mathbf{r} = [\xi, \eta, \zeta]^T, \quad \mathbf{\Lambda} = [\varphi, \theta, \psi]^T \quad (1)$$

[0014] 其中: $R$ 为ROV的位姿项, $r$ 为ROV的位置项, $\Lambda$ 为ROV的姿态项, $\xi$ 为大地坐标系纵向, $\eta$ 为大地坐标系横向, $\zeta$ 为大地坐标系垂向, $\varphi$ 为横倾角, $\theta$ 为纵倾角, $\psi$ 为艏向角,

[0015] 在随体坐标系中,水下机器人的动力学模型输入为所受环境和控制力的力和力矩,输出为水下机器人的线性加速度和姿态角加速度,在随体坐标系中,载体的速度和加速度为:

$$[0016] \quad \mathbf{V} = [\mathbf{U}^T, \mathbf{\Omega}^T]^T, \quad \mathbf{U} = [u, v, w]^T, \quad \mathbf{\Omega} = [p, q, r]^T \quad (2)$$

[0017] 其中: $V$ 为ROV的速度项, $U$ 为ROV的线性速度项, $\Omega$ 为ROV的角速度项, $u$ 为纵向速度, $v$ 为横向速度, $w$ 为垂向速度, $p$ 为横滚角速度, $q$ 为纵倾角速度, $r$ 为转艏角速度,

[0018] 水下机器人在惯性系下的位置项与其随体的速度量的动态关系可以由以下转换方程来描述:

$$[0019] \quad \dot{\mathbf{R}} = \begin{bmatrix} \mathbf{J}_1 & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} & \mathbf{J}_2 \end{bmatrix} \mathbf{V} \quad (3)$$

[0020] 式中 $J_1, J_2$ 是坐标转换矩阵:

$$[0021] \quad \mathbf{J}_1 = \begin{bmatrix} \cos \psi \cos \theta & \cos \psi \sin \theta \sin \varphi - \sin \psi \cos \varphi & \cos \psi \sin \theta \cos \varphi + \sin \psi \sin \varphi \\ \sin \psi \cos \theta & \sin \psi \sin \theta \sin \varphi + \cos \psi \cos \varphi & \sin \psi \sin \theta \cos \varphi - \cos \psi \sin \varphi \\ -\sin \theta & \cos \theta \sin \varphi & \cos \theta \cos \varphi \end{bmatrix} \quad (4)$$

$$[0022] \quad \mathbf{J}_2 = \begin{bmatrix} 1 & \tan \theta \sin \varphi & \tan \theta \cos \varphi \\ 0 & \cos \varphi & -\sin \varphi \\ 0 & \sin \varphi \sec \theta & \cos \varphi \sec \theta \end{bmatrix} \quad (5)$$

[0023] 通过有限元计算或船模实验可以对动力学方程中涉及的水动力参数进行测算;

[0024] 推进系统是由多个处于不同方位的独立推进器所组成,由于本论文最后要讨论不同的推力优化算法之间的效果对比,所以要对单个推进器进行数学建模,

[0025] 对于ROV上所使用的螺旋桨,其敞水推力的计算方法如下式所示<sup>[60]</sup>

$$[0026] \quad T = K_T \rho n^2 D^4 \quad (6)$$

[0027] 式中: $K_T$ 为螺旋桨推力系数, $\rho$ 为海水的密度 $\text{kg}/\text{m}^3$ , $D$ 为螺旋桨直径 $\text{m}$ , $n$ 为螺旋桨的转速 $\text{r}/\text{s}$ ,

[0028] 为了获得相应的推力,必须驱动螺旋桨转动推出水体,螺旋桨的转矩的计算公式为:

$$[0029] \quad Q = K_Q \rho n^2 D^5 \quad (7)$$

[0030] 式中: $Q$ 为螺旋桨扭矩 $\text{Nm}$ , $K_Q$ 为螺旋桨转矩系数,

[0031] 水下机器人配置有6台矢量布置的螺旋桨推进器,在水平面有四个斜向布置的水平面推进器,垂直面沿艇体纵轴关于质心对称布置有两个垂向推进器,

[0032] 在水平方向上,4台推进器平行于水平面 $0-xy$ 布置,按照左上、右上、左下、右下依

次编号为：第一推进器、第二推进器、第三推进器和第四推进器，且与随体坐标系纵轴0-x的夹角为45°，水平方向上的4台推进器安装位置关于纵轴左右对称，因此水平面的推进器为冗余布置，对于控制器计算的控制量通过各推进器的组合实现，第一推进器的螺旋桨在随体坐标系x0y中的位置为：(237mm, -305mm)，第二推进器与第一推进器关于0-xz面对称；第三推进器的坐标为(-300mm, -195mm)，第四推进器与第三推进器关于0-xz面对称，

[0033] 垂直面有两个推进器，其布置在0-xz平面上，依照前后方向编号为第五推进器和第六推进器，其中第五推进器在x0z坐标系中位置为(284mm, 113mm)，第六推进器与第五推进器关于0-yz面对称布置，

[0034] 遥控水下机器人具有六个推进器，其中要控制的量为纵向横移垂向方向的运动和转艏，其中垂直方向的控制器有两个均匀布置，推力分配时只需将控制力平均分配即可，水平方向有四个推进器，分布在四个角，与随体坐标系成45°排布，在分配水平方向的力时，推进器数目大于目标力和力矩的数量，推进器冗余布置，因此在满足推力分配时，可以有无数种分配方式，这要求在进行推力分配的同时，应当兼顾各推进器输出幅值的偏差以及推力分配系统的能耗，

[0035] 建立水下机器人推力分配的数学模型：

$$[0036] \quad \tau = B(\alpha) u \quad (8)$$

[0037] 式(8)为推力分配问题的等式约束，B表示推力分配的空间位置约束， $\tau = \{\tau_x, \tau_y, \tau_r\}$ 为控制器输出的三个自由度上的力和力矩， $u = \{u_1, u_2, u_3, u_4\}$ 为推进器的推力， $\alpha = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$ 为推进器转角，

$$[0038] \quad B(\alpha) = \begin{bmatrix} \sin \alpha_1 & \sin \alpha_2 & \sin \alpha_3 & \sin \alpha_4 \\ \cos \alpha_1 & -\cos \alpha_2 & -\cos \alpha_3 & \cos \alpha_4 \\ \sin \alpha_1 l_{1y} + \cos \alpha_1 l_{1x} & -\sin \alpha_2 l_{2y} - \cos \alpha_2 l_{2x} & \sin \alpha_3 l_{3y} + \cos \alpha_3 l_{3x} & -\sin \alpha_4 l_{4y} - \cos \alpha_4 l_{4x} \end{bmatrix} \quad (9)$$

[0039] 其中 $l_{xi}$ 、 $l_{yi}$ 分别为第i个推进器在x、y方向上距离机器人重心的距离；

[0040] 遗传算法借鉴生物进化理论的机理，以保留更适应环境的生物体为筛选原则的一种随机算法，先使用随机选取的一组解作为初始种群，该种群中的每个解决方案被称为个体或染色体，种群大小是其中个体数目，遗传算法对变量进行编码，遗传算法中每个染色体的适应值，具有评估新产生的种群与上一代种群关系的能力，为了得到每个染色体的适应值，计算同一染色体的优化函数的值，所述优化函数叫做目标函数，通过遗传算法中主要的选择、交叉和变异这种处理使种群个体进行更新，得到下一次迭代的解群，新产生的种群叫子代，上一代种群叫父代，遗传算法的这个迭代过程直到满足停止条件才会终止，在最后一代中具有最佳适应值的染色体被称为最优解，

[0041] 根据推进器系统模型，建立遗传算法的优化目标方程：

$$[0042] \quad \min \text{fun}(u, \alpha) = \sum_{i=1}^4 k_i P(u_i) + (u - \alpha_0)^T W (u - \alpha_0) \quad (9)$$

[0043] 其中P为推进系统的能耗方程， $k_i$ 为第i个推进器占总能耗的占比权值，在本目标机器人中，总功率为各个推进器功率之和，根据电机功率公式，其中n为转速，Q为扭矩：

$$[0044] \quad P = nQ \quad (10)$$

[0045] 将上式与螺旋桨的敞水推力的计算公式(11)、螺旋桨的转矩公式(12)结合：

[0046]  $T = K_T \rho n^2 D^4$  (11)

[0047] 式中,  $K_T$ 为螺旋桨推力系数,  $\rho$ 为海水的密度kg/m<sup>3</sup>,  $D$ 为螺旋桨直径m,  $n$ 为螺旋桨的转速r/s,

[0048] 为了获得相应的推力, 必须驱动螺旋桨转动推出水体, 其转矩的计算公式为:

[0049]  $Q = K_q \rho n^2 D^5$  (12)

[0050] 式中,  $Q$ 为螺旋桨扭矩Nm,  $K_q$ 为螺旋桨转矩系数,

[0051] 计算每个推进器的推力和功率量化公式:

[0052] 
$$P = K_Q \frac{u^{\frac{3}{2}}}{K_T^{\frac{2}{3}} \rho^{\frac{1}{3}} D} \quad (13)$$

[0053] 单个推进器功率与推进器推力的3/2次方成正比,

[0054] 第二项为推进器推力变化项, 其中 $W$ 为权值,

[0055] 将控制分配任务抽象成最优化问题进行描述:

[0056] 
$$\begin{aligned} & \arg \min_u \text{fun}(u) \\ & Bu = \tau \end{aligned} \quad (14)$$

[0057] 开始遗传算法计算, 确定遗传算法中优化变量为四项, 即推进器的推力大小 $u$ , 生成初始种群, 在本文中设计种群中个体数目为50, 进行500次迭代, 自变量数目4, 交叉概率0.8, 变异概率0.15, 公差为1e-8, 对于个体的变量上下限为[-20, 20]kN, 初始种群的个体随机产生, 随机范围即为变量上下限,

[0058] 在优化目标函数 $\text{fun}$ 的基础上设计遗传算法的适应度函数 $V$ 并计算:

[0059] 
$$V = \frac{1}{\text{fun} + (Bu - \tau)^T W (Bu - \tau)} \quad (15)$$

[0060] 上式中 $Bu - \tau$ 表示当前个体关于等式约束的偏差值,  $W$ 为惩罚矩阵, 当个体越接近等式约束的解时,  $W$ 的适应度也越大, 同样当个体越满足优化条件时, 它的适应度越大, 这样便可以求出种群中更加满足约束条件且适应优化方程的个体,

[0061] 根据适应度选择当前种群中个体, 根据适应程度对最佳个体进行分类, 记录最佳个体, 并以如下公式选择计算群体中个体的选择概率:

[0062] 
$$P_i = \frac{f_i}{\sum_{i=1}^n f_i} \quad (16)$$

[0063] 式中 $P_i$ 为第 $i$ 个个体的选择概率,  $f_i$ 为其适应度, 拥有较大适应度的解有更大的概率被选择继承到下一次迭代, 在获取下一次迭代种群时对选中个体进行交叉变异提高在优值周围的探索,

[0064] 对遗传算法的终止条件进行判断, 如果当前最佳个体的适应度与上代最佳个体适应度之差小于容忍度, 则认为达到最优条件, 遗传算法结束; 若不满足, 判断是否达到繁殖代数, 若达到则停止计算,

[0065] 对有交叉权的个体进行染色体交叉, 本发明中采用浮点数交叉法, 即交换父母中某一段浮点位, 对有交叉权但不满足交叉概率的保留父代, 对于产生的子代, 依照编译概率对其某浮点位进行随机数变异, 在产生的子代中, 如果它的变量值超过了变量上下限, 则将

其值修改为上下限的值,同时为了保证种群的优选性,将上代最优个体进行保留,将交叉、变异后产生的新种群回到计算适度函数V的步骤,进行下一轮的计算,

[0066] 结束计算后,获得的最优个体u即为本次推力分配问题中能耗最小的优化推力。

[0067] 在步骤二中,具体的,针对的水下机器人在水平面操纵上具有纵轴和艏向上的执行机构,通过调整前后方向的推力和转向使机器人趋近目标点,控制原理为获得目标值和实际值的差值,将偏差作为强化学习的状态值,使用策略网络选择所需执行的动作,其中动作空间为归一化以后的x方向推力和绕z轴旋转的力矩,

[0068] 在控制器设计时,建立Actor网络作为策略网络 $\mu$ ,某一时刻的惯性坐标系位置和随体坐标系速度 $R$ 、 $V$ 作为状态空间的输入 $s = [x, y, \varphi, u, v, w]$ ,通过神经网络输出控制指令,其包括纵向、横向和转艏方向上归一化的力和力矩,动作空间为: $a = [T_x, T_y, T_r]$ 将选择的动作交由推力分配系统生成实际的推力分配和推进器转速控制方案,通过调节机器人纵向和横向的速度和艏向角使机器人向目标点靠拢,

[0069] 控制器状态选择随体坐标系下的误差 $s = \{e_{x_r}, e_{y_r}, e_{\varphi}, u, v, r\}$ , $e_{x_r}$ 为随体坐标系下纵向误差: $e_{x_r} = x_{Target} - x_r$ , $e_{y_r}$ 为横向误差, $e_{\varphi}$ 为大地坐标系下目标艏向和当前艏向的差值,目标艏向: $\varphi_{Target} = \arctan(dy/dx)$ ,因此艏向误差可表示为:

$$[0070] \quad e_{\varphi} = \varphi_{Target} - \varphi \quad (17)$$

[0071] 以点镇定为目标,设置初始目标艏向角为0,偏差为目标点在机器人艇体坐标系纵向距离,将目标速度和角速度设置为0,则设置强化学习的奖励函数为:

$$[0072] \quad r = -\sqrt{\Delta x^2 + \Delta y^2} \quad (18)$$

[0073] 当智能体越靠近智能体,且智能体的速度越慢时,奖励值就越大,以累积折扣奖励值最大为目标,通过DDPG方法使策略网络选择最优策略,当某一时刻输入智能体状态时,它能输出最佳动作,进而完成机器人点镇定的控制问题,

[0074] 当传感器获得水下机器人状态信息s后,根据状态s,策略网络会选择动作控制水下机器人运动,在下一采样时刻获得状态信息s',同时根据奖励函数计算当前状态的奖励值r,将 $\langle s, a, r, s' \rangle$ 作为一个采样点储存在记忆库中,在参数训练的过程中,从记忆库小批量采样,使用策略网络进行决策,并用目标评价计算采样动作的目标Q值,通过现实评价估计Q值与目标Q值的均方误差调整评价网络参数更新,同时使用目标Q函数计算更新策略网络,采用软更新的方式,异步的调整目标神经网络的参数,使目标神经网络向接近现实神经网络的方向缓慢更新,在训练过程中使用策略网络以状态信息s为输入,输出3个自由度上的力和力矩a,并加入高斯噪声以实现对环境的探索,将策略网络选择的动作a发送给推力分配模块,进而驱动ROV推进器运转,实现动作控制。

[0075] 在步骤三中,具体的,通常抗扰控制器以状态观测器为基础,依据观测器估计的状态和测量状态误差为输入量调整控制器,水下机器人在控制力的作用下由状态 $s_t$ 转移至 $s_{t+1}$ ,而由于外部扰动和系统不确定性的存在使状态转移至 $s'_{t+1}$ ,将干扰后的状态量与估计状态量的误差作为输入,加入控制力作为线性补偿以改善控制器的抗扰性能,

[0076] 以随体坐标系下位置量为被控状态量 $s = \{x_r, y_r, yaw_r\}$ ,根据2个采样时刻状态量 $\{s_{k-1}, s_k\}$ 和控制量 $\{u_{k-1}, u_k\}$ 使用SVR方法状态估计水下机器人状态 $s_{k+1}$ ,在无干扰情况下采

集运动状态数据作为数据集训练SVR收敛,并以此来估计理想状态下被控状态,则状态误差为测量的状态与估计状态的差值:

$$[0077] \quad e_s(k) = \hat{s}(k) - s(k) \quad (19)$$

[0078] 根据反馈控制的思想,使用非线性反馈控制器对系统误差进行补偿,对于非线性系统加入非线性补偿控制量:

$$[0079] \quad \dot{x} = f(x) + w(t) + u_0(t) + g(e) \quad (20)$$

[0080] 其中, $x$ 为非线性系统状态量, $w$ 为时域干扰项, $u$ 为控制量, $g(e)$ 为以估计误差为输入的补偿控制量,通过调整控制量来调整期望状态与系统输出的误差,对于一阶误差输入,取 $g(\cdot)$ 为fal函数[110],

$$[0081] \quad g(e_s) = \beta * fal(e_s, \alpha, \delta) = \begin{cases} |e_s|^\alpha \operatorname{sgn}(e_s) & |e_s| > \delta \\ \frac{e_s}{\delta^{1-\alpha}} & |e_s| < \delta \end{cases} \quad (21)$$

[0082] 其中, $\beta$ 为比例参数,其反映了补偿控制力在总控制力中的比重, $\delta$ 为反馈函数线性段的范围,当 $|e_s| < \delta$ 时,反馈函数处于线性区间,避免了误差在0点附近时的高频振荡,减小稳态误差,通常取 $\alpha$ 在0~1之间,反映了系统的阶次,在 $|e_s| > \delta$ 的区间中,当误差增幅大时,减小系统增益防止发散, $\delta$ 通常取值较小但其值过小会导致线性段缩短,使系统出现高频抖震,取值过大会导致线性段延长,使非线性段克服系统稳态误差的特性降低。

[0083] 本发明的主要优点是:本发明的一种ROV推力分配与基于强化学习的运动控制方法,该ROV推力分配方法在进行推力分配的同时降低推进器能耗,从而降低了整个ROV的能耗。DDPG算法能够在动作输出幅度能进行连续值输出,相较于只能离散输出的强化学习控制更适用水下机器人控制问题,同时DDPG控制器能够在连续决策问题中进行学习,可以一边在线进行控制,一边以在线离线相结合的方式更新神经网络参数,加快了学习进程。采用抗扰控制器可以提高该运动控制方法对ROV未建模水动力、外界扰动及其他不确定因素的抗扰能力。该方法实现了ROV的推力分配,完成了对ROV运动的精确控制,使ROV具备更好完成各种作业任务的基础。

## 附图说明

- [0084] 图1为推力分配系统示意图;  
 [0085] 图2为水平面推进器布置示意图;  
 [0086] 图3为控制系统结构图;  
 [0087] 图4为监督式反馈DDPG控制器的运行逻辑图。

## 具体实施方式

[0088] 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。

[0089] 一种ROV推力分配与基于强化学习的运动控制方法,所述方法包括以下步骤:

[0090] 步骤一、进行ROV动力学建模,确定ROV推进器空间布置及推进系统模型,提出ROV推力分配方法;

[0091] 步骤二、在所述ROV推力分配方法的基础上,加入DDPG控制器;

[0092] 步骤三、在所述DDPG控制器的基础上,加入抗扰控制器。

[0093] 在步骤一中,具体的,强化学习需要智能体采取动作,并在与环境进行交互的过程中,对环境进行充分探索,以调整控制策略,这要求在仿真时建立精确的ROV动力学模型,获取下一时刻机器人状态。

[0094] 本发明以一自制的遥控水下机器人为研究对象,其搭载有6个无刷推进器,其中垂直面搭载2个推进器沿纵向对称分布,水平面推进器以X型矢量布置。同时水下机器人搭载了九轴罗经和声呐等设备用于测量载体的角速度、加速度信息。ROV的深度信息由深度计进行测量,本发明主要研究控制算法和推力分配算法,认为传感器所测数据真实有效。

[0095] 本文所研究ROV参数信息如下:

[0096]	质量	艇长	艇宽	高度	$I_x$	$I_y$	$I_z$
	45kg	75cm	70cm	40cm	$0\text{kg}\cdot\text{m}^2$	$10\text{kg}\cdot\text{m}^2$	$10\text{kg}\cdot\text{m}^2$

[0097] 表1 ROV主尺度

[0098] 建立有利于描述水下机器人运动状态的坐标系,惯性坐标系 $E-\xi\eta\zeta$ 以空间任意一点为原点, $E\zeta$ 轴正向指向地心, $E\xi$ 轴指向北向、 $E\eta$ 指向东向,随体坐标系 $O-xyz$ 固连于ROV艇体上,原点取在ROV的对称轴中点或在重心上,取纵轴 $Ox$ 指向艇首,横轴 $Oy$ 与 $Ox$ 轴正交于同一平面并以右舷为正方向,垂轴 $Oz$ 指向机器人的腹部,机器人的位置和方向由惯性坐标系表示,速度和角速度在随体坐标系中表示,

[0099] 在惯性坐标系中,ROV位置和姿态表示为:

$$[0100] \quad \mathbf{R} = [\mathbf{r}^T, \mathbf{\Lambda}^T]^T, \quad \mathbf{r} = [\xi, \eta, \zeta]^T, \quad \mathbf{\Lambda} = [\varphi, \theta, \psi]^T \quad (1)$$

[0101] 其中: $\mathbf{R}$ 为ROV的位姿项, $\mathbf{r}$ 为ROV的位置项, $\mathbf{\Lambda}$ 为ROV的姿态项, $\xi$ 为大地坐标系纵向, $\eta$ 为大地坐标系横向, $\zeta$ 为大地坐标系垂向, $\varphi$ 为横倾角, $\theta$ 为纵倾角, $\psi$ 为艏向角,

[0102] 在随体坐标系中,水下机器人的动力学模型输入为所受环境和控制力的力和力矩,输出为水下机器人的线性加速度和姿态角加速度,在随体坐标系中,载体的速度和加速度为:

$$[0103] \quad \mathbf{V} = [\mathbf{U}^T, \mathbf{\Omega}^T]^T, \quad \mathbf{U} = [u, v, w]^T, \quad \mathbf{\Omega} = [p, q, r]^T \quad (2)$$

[0104] 其中: $\mathbf{V}$ 为ROV的速度项, $\mathbf{U}$ 为ROV的线性速度项, $\mathbf{\Omega}$ 为ROV的角速度项, $u$ 为纵向速度, $v$ 为横向速度, $w$ 为垂向速度, $p$ 为横滚角速度, $q$ 为纵倾角速度, $r$ 为转艏角速度,

[0105] 水下机器人在惯性系下的位置项与其随体的速度量的动态关系可以由以下转换方程来描述:

$$[0106] \quad \dot{\mathbf{R}} = \begin{bmatrix} \mathbf{J}_1 & \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} & \mathbf{J}_2 \end{bmatrix} \mathbf{V} \quad (3)$$

[0107] 式中 $\mathbf{J}_1, \mathbf{J}_2$ 是坐标转换矩阵:

$$[0108] \quad \mathbf{J}_1 = \begin{bmatrix} \cos \psi \cos \theta & \cos \psi \sin \theta \sin \varphi - \sin \psi \cos \varphi & \cos \psi \sin \theta \cos \varphi + \sin \psi \sin \varphi \\ \sin \psi \cos \theta & \sin \psi \sin \theta \sin \varphi + \cos \psi \cos \varphi & \sin \psi \sin \theta \cos \varphi - \cos \psi \sin \varphi \\ -\sin \theta & \cos \theta \sin \varphi & \cos \theta \cos \varphi \end{bmatrix} \quad (4)$$

$$[0109] \quad \mathbf{J}_2 = \begin{bmatrix} 1 & \tan \theta \sin \varphi & \tan \theta \cos \varphi \\ 0 & \cos \varphi & -\sin \varphi \\ 0 & \sin \varphi \sec \theta & \cos \varphi \sec \theta \end{bmatrix} \quad (5)$$

[0110] 通过有限元计算或船模实验可以对动力学方程中涉及的水动力参数进行测算,本文所研究的水下机器人因条件有限,未能使用相关试验获得水动力系数,因此使用“WL-2”的水动力参数进行研究和仿真:

水动力参数		
$X_{\dot{u}} = -5.9 \times 10^{-3}$	$Y_{\dot{v}} = -1.6687 \times 10^{-1}$	$Y_{\dot{v}r} = 0$
$Y_{\dot{v}r} = 1.258 \times 10^{-2}$	$N_{\dot{v}r} = -1.2432 \times 10^{-1}$	$N_{\dot{v}v} = 0$
$Y_{\dot{r}} = 9.4196 \times 10^{-4}$	$X_{\dot{u}} = -1.5777 \times 10^{-3}$	$N_{\dot{v}v} = 0$
$Y_{\dot{v}} = -3.0753 \times 10^{-2}$	$N_{\dot{r}} = -1.012 \times 10^{-1}$	

[0112] 表2水动力参数

[0113] 对于诸如水面航行器或是水下机器人这种结构的载体,它的运动控制对象一般可以分为冗余驱动、完全驱动或欠驱动三种类型,但基本上是基于控制器选择的控制量和执行机构执行数目的关系区分。在水下机器人进行精度较高的运动操纵时,使用完全驱动或冗余驱动的载体更能满足任务要求。由于搭载有多余执行器的艇体在接受到控制指令时,有足够的执行能力以满足其控制要求,并且在出现故障或单一推进器能力不足时可以使用额外推进器辅助推进。本文以搭载有6个推进器的遥控水下机器人为研究对象,其推进器能够满足水平面三自由度运动和升沉运动的控制要求,在水平定位控制中为冗余驱动类型。在满足推进器空间约束的前提下可以执行控制器任意自由度控制指令,因此有必要对其冗余推进器特点进行控制分配,实现控制目标的同时对其推进系统进行优化。控制器在设计时不考虑水下机器人对其指令的执行能力,依据目标位置和当前状态决策出当前机器人应当执行的力或力矩,采用推力分配模块对上述力或力矩进行解析分配给对应的各个推进器,使机器人按照控制指令运动。因此,设计推力分配模块是实现水下机器人运动控制的基础和前提。

[0114] 图1给出了所研究的作业型ROV推力控制系统结构,可以看出,ROV的推力控制分配模块位于运动控制器和推进器之间,完成输入的期望控制指令转化为推进器推力的分配与优化。

[0115] 推进系统是由多个处于不同方位的独立推进器所组成,由于本论文最后要讨论不同的推力优化算法之间的效果对比,所以要对单个推进器进行数学建模,

[0116] 对于ROV上所使用的螺旋桨,其敞水推力的计算方法如下式所示<sup>[60]</sup>

$$[0117] \quad T = K_T \rho n^2 D^4 \quad (6)$$

[0118] 式中: $K_T$ 为螺旋桨推力系数, $\rho$ 为海水的密度 $\text{kg}/\text{m}^3$ , $D$ 为螺旋桨直径 $\text{m}$ , $n$ 为螺旋桨的转速 $\text{r}/\text{s}$ ,

[0119] 为了获得相应的推力,必须驱动螺旋桨转动推出水体,螺旋桨的转矩的计算公式为:

$$[0120] \quad Q = K_q \rho n^2 D^5 \quad (7)$$

[0121] 式中: $Q$ 为螺旋桨扭矩Nm, $K_q$ 为螺旋桨转矩系数,

[0122] 水下机器人配置有6台矢量布置的螺旋桨推进器,在水平面有四个斜向布置的水平面推进器,垂直面沿艇体纵轴关于质心对称布置有两个垂向推进器,其中水平的四台推进器、垂直的两台推进器型号分别相同,

[0123] 对于本文所研究的水下机器人,其配置有6台矢量布置的螺旋桨推进器,在水平面有四个斜向布置的水平面推进器,垂直面沿艇体纵轴关于质心对称布置有两个垂向推进器。其中水平的四台推进器、垂直的两台推进器型号分别相同。

[0124] 参照图2所示,在水平方向上,4台推进器平行于水平面 $O-xy$ 布置,按照左上、右上、左下、右下依次编号为:第一推进器、第二推进器、第三推进器和第四推进器,且与随体坐标系纵轴 $O-x$ 的夹角为 $45^\circ$ ,水平方向上的4台推进器安装位置关于纵轴左右对称,因此水平面的推进器为冗余布置,对于控制器计算的控制量通过各推进器的组合实现,第一推进器的螺旋桨在随体坐标系 $xOy$ 中的位置为:(237mm, -305mm),第二推进器与第一推进器关于 $O-xz$ 面对称;第三推进器的坐标为(-300mm, -195mm),第四推进器与第三推进器关于 $O-xz$ 面对称,

[0125] 垂直面有两个推进器,其布置在 $O-xz$ 平面上,依照前后方向编号为第五推进器和第六推进器,其中第五推进器在 $xOz$ 坐标系中位置为(284mm, 113mm),第六推进器与第五推进器关于 $O-yz$ 面对称布置,

	峰值推力	额定功率	重量
[0126]	5.2kgf	400w	3kg

[0127] 表3推进器参数

序号	描述	$x/(mm)$	$y/(mm)$	$z/(mm)$	夹角	
[0128]	水平	1	237	-305	0	与 $x$ 轴 $45^\circ$
		2	237	305	0	与 $x$ 轴 $45^\circ$
		3	-300	-195	0	与 $x$ 轴 $45^\circ$
		4	-300	195	0	与 $x$ 轴 $45^\circ$
[0128]	垂直	5	0	284	113	无
		6	0	-284	113	无

[0129] 表4推进器布置参数

[0130] 遥控水下机器人具有六个推进器,其中要控制的量为纵向横移垂向方向的运动和转舵,其中垂直方向的控制器有两个均匀布置,推力分配时只需将控制力平均分配即可,水平方向有四个推进器,分布在四个角,与随体坐标系成 $45^\circ$ 排布,在分配水平方向的力时,推进器数目大于目标力和力矩的数量,推进器冗余布置,因此在满足推力分配时,可以有无数种分配方式,这要求在进行推力分配的同时,应当兼顾各推进器输出幅值的偏差以及推力

分配系统的能耗,

[0131] 建立水下机器人推力分配的数学模型:

$$[0132] \quad \tau = B(\alpha) u \quad (8)$$

[0133] 式(8)为推力分配问题的等式约束,  $B$ 表示推力分配的空间位置约束,  $\tau = \{\tau_x, \tau_y, \tau_r\}$ 为控制器输出的三个自由度上的力和力矩,  $u = \{u_1, u_2, u_3, u_4\}$ 为推进器的推力,  $\alpha = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$ 为推进器转角,

$$[0134] \quad B(\alpha) = \begin{bmatrix} \sin \alpha_1 & \sin \alpha_2 & \sin \alpha_3 & \sin \alpha_4 \\ \cos \alpha_1 & -\cos \alpha_2 & -\cos \alpha_3 & \cos \alpha_4 \\ \sin \alpha_1 l_{1y} + \cos \alpha_1 l_{1x} & -\sin \alpha_2 l_{2y} - \cos \alpha_2 l_{2x} & \sin \alpha_3 l_{3y} + \cos \alpha_3 l_{3x} & -\sin \alpha_4 l_{4y} - \cos \alpha_4 l_{4x} \end{bmatrix} \quad (9)$$

[0135] 其中 $l_{xi}$ 、 $l_{yi}$ 分别为第 $i$ 个推进器在 $x$ 、 $y$ 方向上距离机器人重心的距离;

[0136] 本发明提出了一种基于遗传算法的推力分配优化方法。为了在进行推力分配的同时降低推进器能耗,确定了所有推进器推力之和最小的优化目标,并使用遗传算法进行优化。结合水下机器人空间运动的特点,实现了推力分配优化。

[0137] 遗传算法借鉴生物进化理论的机理,以保留更适应环境的生物体为筛选原则的一种随机算法,随机算法的一个显著特征是给出一组解而不是一个解,因此,先使用随机选取的一组解作为初始种群,该种群中的每个解决方案被称为个体或染色体,种群大小是其中个体数目,遗传算法对变量进行编码,遗传算法中每个染色体的适应值,具有评估新产生的种群与上一代种群关系的能力,为了得到每个染色体的适应值,计算同一染色体的优化函数的值,所述优化函数叫做目标函数,通过遗传算法中主要的选择、交叉和变异这种处理使种群个体进行更新,得到下一次迭代的解群,新产生的种群叫子代,上一代种群叫父代,遗传算法的这个迭代过程直到满足停止条件才会终止,在最后一代中具有最佳适应值的染色体被称为最优解,

[0138] 根据推进器系统模型,建立遗传算法的优化目标方程:

$$[0139] \quad \min \text{fun}(u, \alpha) = \sum_{i=1}^4 k_i P(u_i) + (u - \alpha_0)^T W (u - \alpha_0) \quad (9)$$

[0140] 其中 $P$ 为推进系统的能耗方程,  $k_i$ 为第 $i$ 个推进器占总能耗的占比权值,在本目标机器人中,总功率为各个推进器功率之和,根据电机功率公式,其中 $n$ 为转速,  $Q$ 为扭矩:

$$[0141] \quad P = nQ \quad (10)$$

[0142] 将上式与螺旋桨的敞水推力的计算公式(11)、螺旋桨的转矩公式(12)结合:

$$[0143] \quad T = K_T \rho n^2 D^4 \quad (11)$$

[0144] 式中,  $K_T$ 为螺旋桨推力系数,  $\rho$ 为海水的密度 $\text{kg}/\text{m}^3$ ,  $D$ 为螺旋桨直径 $\text{m}$ ,  $n$ 为螺旋桨的转速 $\text{r}/\text{s}$ ,

[0145] 为了获得相应的推力,必须驱动螺旋桨转动推出水体,其转矩的计算公式为:

$$[0146] \quad Q = K_Q \rho n^2 D^5 \quad (12)$$

[0147] 式中,  $Q$ 为螺旋桨扭矩 $\text{Nm}$ ,  $K_Q$ 为螺旋桨转矩系数,

[0148] 计算每个推进器的推力和功率量化公式:

$$[0149] \quad P = K_Q \frac{u^{\frac{3}{2}}}{K_T^{\frac{3}{2}} \rho^{\frac{1}{2}} D} \quad (13)$$

[0150] 单个推进器功率与推进器推力的3/2次方成正比，

[0151] 第二项为推进器推力变化项，其中W为权值，该项限制了在推力分配中各推进器的推力变化不会太大，保证了推进器的输出能力，

[0152] 将控制分配任务抽象成最优化问题进行描述：

$$[0153] \quad \begin{aligned} & \arg \min_u \quad \text{fun}(u) \\ & Bu = \tau \end{aligned} \quad (14)$$

[0154] 开始遗传算法计算，确定遗传算法中优化变量为四项，即推进器的推力大小u，生成初始种群，在本文中设计种群中个体数目为50，进行500次迭代，自变量数目4，交叉概率0.8，变异概率0.15，公差为1e-8，对于个体的变量上下限为[-20, 20]kN，初始种群的个体随机产生，随机范围即为变量上下限，

[0155] 在优化目标函数fun的基础上设计遗传算法的适应度函数V并计算：

$$[0156] \quad V = \frac{1}{\text{fun} + (Bu - \tau)^T W (Bu - \tau)} \quad (15)$$

[0157] 上式中Bu-τ表示当前个体关于等式约束的偏差值，W为惩罚矩阵，当个体越接近等式约束的解时，W的适应度也越大，同样当个体越满足优化条件时，它的适应度越大，这样可以求出种群中更加满足约束条件且适应优化方程的个体，

[0158] 根据适应度选择当前种群中个体，根据适应程度对最佳个体进行分类，记录最佳个体，并以如下公式选择计算群体中个体的选择概率：

$$[0159] \quad P_i = \frac{f_i}{\sum_1^n f_i} \quad (16)$$

[0160] 式中P<sub>i</sub>为第i个个体的选择概率，f<sub>i</sub>为其适应度，拥有较大适应度的解有更大的概率被选择继承到下一次迭代，在获取下一次迭代种群时对选中个体进行交叉变异提高在优值周围的探索，

[0161] 对遗传算法的终止条件进行判断，如果当前最佳个体的适应度与上代最佳个体适应度之差小于容忍度，则认为达到最优条件，遗传算法结束；若不满足，判断是否达到繁殖代数，若达到则停止计算，

[0162] 对有交叉权的个体进行染色体交叉，本发明中采用浮点数交叉法，即交换父母中某一段浮点位，对有交叉权但不满足交叉概率的保留父代，对于产生的子代，依照编译概率对其某浮点位进行随机数变异，在产生的子代中，如果它的变量值超过了变量上下限，则将其值修改为上下限的值，同时为了保证种群的优选性，将上代最优个体进行保留，将交叉、变异后的产生的新种群回到计算适应度函数V的步骤，进行下一轮的计算，

[0163] 结束计算后，获得的最优个体u即为本次推力分配问题中能耗最小的优化推力。

[0164] 在步骤二中，具体的，参照图3所示，姿态镇定在水下机器人的运动控制系统设计中占据重要地位。镇定是指设计反馈控制器，使闭环系统全局或局部渐进收敛于稳定平衡状态。姿态镇定是在扰动条件下维持期望的稳定平衡姿态，它涉及到位置与姿态的稳定性

控制,属于定点控制范畴。

[0165] 本发明设计了一个基于DDPG算法的水下机器人运动控制系统,该系统对水下机器人水平面点镇定进行控制,将目标点位置作为控制器的输入,其输出为归一化以后的控制力,控制力交由推力分配系统实现水下机器人运动控制。

[0166] 该强化学习算法所需的状态空间为水下机器人大地坐标系下与目标点的相当位置和随体坐标系下的速度,通过传感器测得的位置和速度计算得到相对位置和实时速度信息,本发明不讨论传感器信息处理的问题。

[0167] 针对的水下机器人在水平面操纵上具有纵轴和艏向上的执行机构,通过调整前后方向的推力和转向使机器人趋近目标点,控制原理为获得目标值和实际值的差值,将偏差作为强化学习的状态值,使用策略网络选择所需执行的动作,其中动作空间为归一化以后的x方向推力和绕z轴旋转的力矩,

[0168] 在控制器设计时,建立Actor网络作为策略网络 $\mu$ ,某一时刻的惯性坐标系位置和随体坐标系速度 $R, V$ 作为状态空间的输入 $s = [x, y, \varphi, u, v, w]$ ,通过神经网络输出控制指令,其包括纵向、横向和转艏方向上归一化的力和力矩,动作空间为: $a = [T_x, T_y, T_r]$ 将选择的动作交由推力分配系统生成实际的推力分配和推进器转速控制方案,通过调节机器人纵向和横向的速度和艏向角使机器人向目标点靠拢,

[0169] 控制器状态选择随体坐标系下的误差 $s = \{e_{x_r}, e_{y_r}, e_\varphi, u, v, r\}$ ,  $e_{x_r}$ 为随体坐标系下纵向误差: $e_{x_r} = x_{T_{target}} - x_r$ ,  $e_{y_r}$ 为横向误差,  $e_\varphi$ 为大地坐标系下目标艏向和当前艏向的差值,目标艏向: $\varphi_{T_{target}} = \arctan(dy/dx)$ ,因此艏向误差可表示为:

$$[0170] \quad e_\varphi = \varphi_{T_{target}} - \varphi \quad (17)$$

[0171] 以点镇定为目标,设置初始目标艏向角为0,偏差为目标点在机器人艇体坐标系纵向距离,将目标速度和角速度设置为0,则设置强化学习的奖励函数为:

$$[0172] \quad r = -\sqrt{\Delta x^2 + \Delta y^2} \quad (18)$$

[0173] 当智能体越靠近智能体,且智能体的速度越慢时,奖励值就越大,以累积折扣奖励值最大为目标,通过DDPG方法使策略网络选择最优策略,当某一时刻输入智能体状态时,它能输出最佳动作,进而完成机器人点镇定的控制问题,

[0174] 当传感器获得水下机器人状态信息 $s$ 后,根据状态 $s$ ,策略网络会选择一个动作控制水下机器人运动,在下一采样时刻获得状态信息 $s'$ ,同时根据奖励函数计算当前状态的奖励值 $r$ ,将 $\langle s, a, r, s' \rangle$ 作为一个采样点储存在记忆库中,在参数训练的过程中,从记忆库小批量采样,使用策略网络进行决策,并用目标评价计算采样动作的目标Q值,通过现实评价估计Q值与目标Q值的均方误差调整评价网络参数更新,同时使用目标Q函数计算更新策略网络,采用软更新的方式,异步的调整目标神经网络的参数,使目标神经网络向接近现实神经网络的方向缓慢更新,在训练过程中使用策略网络以状态信息 $s$ 为输入,输出3个自由度上的力和力矩 $a$ ,并加入高斯噪声以实现对环境的探索,将策略网络选择的动作 $a$ 发送给推力分配模块,进而驱动ROV推进器运转,实现动作控制。

[0175] 在步骤三中,具体的,强化学习控制器在训练时使用的是无干扰的仿真环境,并且推进器的合推力与控制器输出完全一致,由于水下机器人系统较为复杂,使用水动力方程

模型与其受力的真实情况有所差异,所以在训练时使用的仿真环境与实际环境有所差异,并且在运动过程中存在水流、波浪及其他未知干扰因素等不确定因素,并且推进器的推力曲线会出现测不准或由于供电电压变化使推力与预期不符,使控制器的控制效果不如预期,因此需要根据控制效果反馈调整控制律以增强控制系统的抗扰性能。通常抗扰控制器以状态观测器为基础,依据观测器估计的状态和测量状态误差为输入量调整控制器,水下机器人在控制力的作用下由状态 $s_t$ 转移至 $s_{t+1}$ ,而由于外部扰动和系统不确定性的存在使状态转移至 $s'_{t+1}$ ,将干扰后的状态量与估计状态量的误差作为输入,加入控制力作为线性补偿以改善控制器的抗扰性能,

[0176] 以随体坐标系下位置量为被控状态量 $s = \{x_r, y_r, yaw_r\}$ ,根据2个采样时刻状态量 $\{s_{k-1}, s_k\}$ 和控制量 $\{u_{k-1}, u_k\}$ 使用SVR方法状态估计水下机器人状态 $s_{k+1}$ ,在无干扰情况下采集运动状态数据作为数据集训练SVR收敛,并以此来估计理想状态下被控状态,则状态误差为测量的状态与估计状态的差值:

$$[0177] \quad e_s(k) = \hat{s}(k) - s(k) \quad (19)$$

[0178] 根据反馈控制的思想,使用非线性反馈控制器对系统误差进行补偿,对于非线性系统加入非线性补偿控制量:

$$[0179] \quad \dot{x} = f(x) + w(t) + u_0(t) + g(e) \quad (20)$$

[0180] 其中, $x$ 为非线性系统状态量, $w$ 为时域干扰项, $u$ 为控制量, $g(e)$ 为以估计误差为输入的补偿控制量,通过调整控制量来调整期望状态与系统输出的误差,对于一阶误差输入,取 $g(\cdot)$ 为fal函数[110],

$$[0181] \quad g(e_s) = \beta * fal(e_s, \alpha, \delta) = \begin{cases} |e_s|^\alpha \operatorname{sgn}(e_s) & |e_s| > \delta \\ \frac{e_s}{\delta^{1-\alpha}} & |e_s| < \delta \end{cases} \quad (21)$$

[0182] 其中, $\beta$ 为比例参数,其反映了补偿控制力在总控制力中的比重, $\delta$ 为反馈函数线性段的范围,当 $|e_s| < \delta$ 时,反馈函数处于线性区间,避免了误差在0点附近时的高频振荡,减小稳态误差,通常取 $\alpha$ 在0~1之间,反映了系统的阶次,在 $|e_s| > \delta$ 的区间中,当误差增幅大时,减小系统增益防止发散, $\delta$ 通常取值较小但其值过小会导致线性段缩短,使系统出现高频抖震,取值过大会导致线性段延长,使非线性段克服系统稳态误差的特性降低。

[0183] 参照图4所示,在控制系统中加入误差补偿控制以消除外界扰动对系统的误差,考虑到强化学习可以根据环境变化自适应调节控制律,因此将预测误差作为状态量传入控制器,强化学习算法可以在状态误差作用下调整输出以弥补系统扰动。

[0184] 在强化学习控制器的基础上与反馈控制器串联,使用训练好的强化学习控制器作为基础控制,使用SVR状态估计器对系统运行状态进行估计,使用非线性补偿控制器对该系统误差进行补偿,则扰动情况下串联控制器的输出为:

$$[0185] \quad u(t) = u_r(s, t | \theta) + g(e_s) \quad (22)$$

[0186] 使用串联控制器作为实际输出对水下机器人系统进行控制,以改善在无扰动条件下强化学习控制器的性能。同时考虑到监督强化学习控制器的特性,控制器可以从有监督示范的数据中调节控制策略,因此加入的非线性误差反馈控制器将作为监督信号引导DDPG控制器的控制律调整。

[0187] 借鉴有监督机器学习的方法,在Actor网络参数调整时,采用加入标签的形式,以非线性误差反馈控制器为专家策略,通过行为克隆的方法,通过调整神经网络更新梯度的方式使控制策略向有反馈补偿的示范控制器转移。

[0188] 由于水下机器人的非线性动力学特征,其水动力和各向速度有关,所以其所示水动力之间耦合,并且在运行过程中水动力和推进器推力之间也是动态耦合的,水下机器人运动系统是多自由度耦合的。而基于状态估计误差的补偿控制器是解耦设计的,其在抗扰能力上有所缺陷,由于其水动力特性比较复杂,模型参数也难以确定。而强化学习为多输入多输出(MIMO)控制器,并且其采用无模型方法,控制量由输入状态经过高度非线性的耦合神经网络输出,强化学习方法能作为隐含量处耦合状态,使用强化学习控制器能更好的弥补扰动条件下系统误差。

[0189] 使用非线性误差反馈控制器作为示范策略进行预训练,将Actor网络输出的策略和串联控制策略的误差作为损失,使DDPG策略向反馈控制器学习,并采用经验回放机制调整Critic网络。随着训练的进行降低反馈控制器所占比重,最终调整神经网络参数以适应有环境变量的控制环境。

[0190] 扩展强化学习输入的维度,加入预测误差作为输入,并将加入补偿以后的控制力作为输出值,设计以估计误差为被控量、控制力为控制量的强化学习控制器,以并行的方式结合监督控制,强化学习控制器的输出动作 $a_0$ 、监督控制器的输出动作 $a_E$ 以及外部噪声 $\delta$ ,即:

$$[0191] \quad a_s = ka_0 + (1-k)a_E \quad (23)$$

[0192] 采样分布采样和分离储存的方式将监督控制轨迹记录在经验回放池中,通过并行采样和调整Actor损失函数的方法使控制策略向监督策略学习,通过这种方法建立监督反馈DDPG控制器。

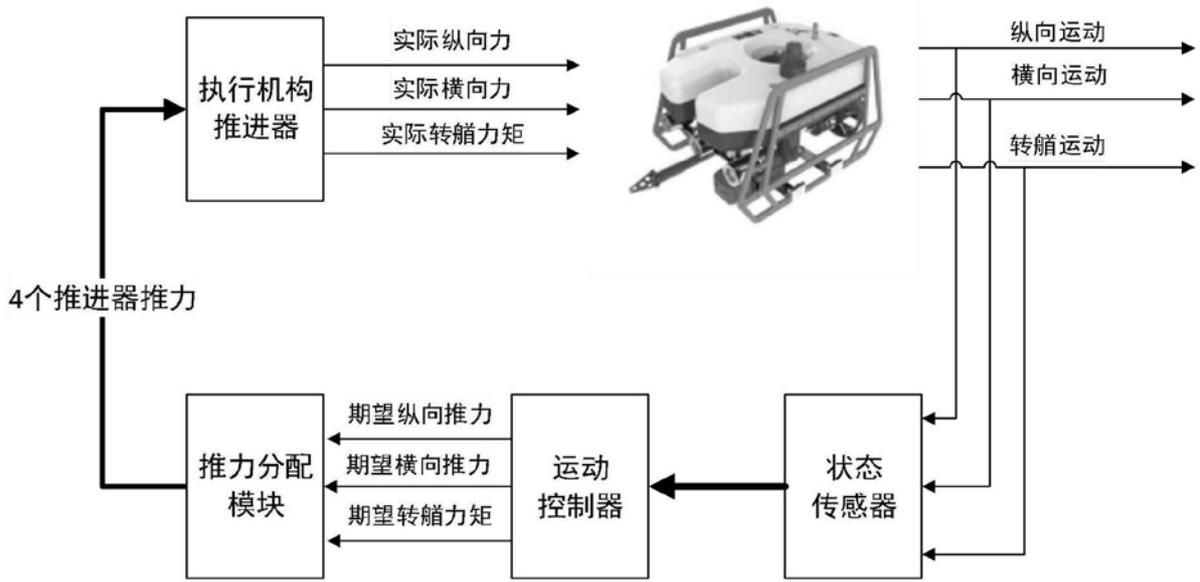


图1

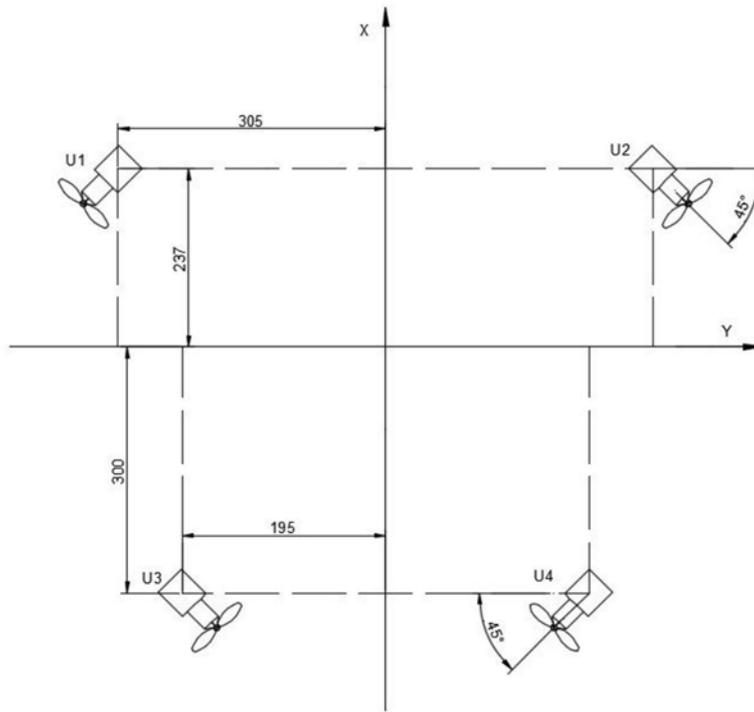


图2

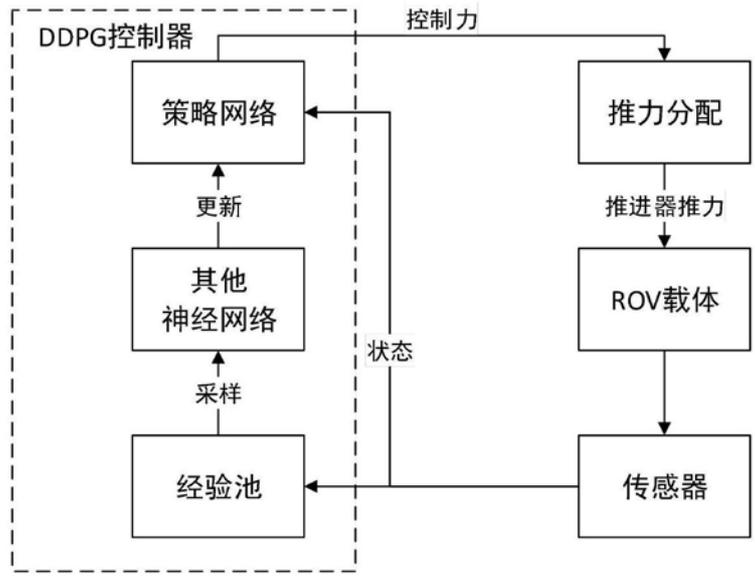


图3

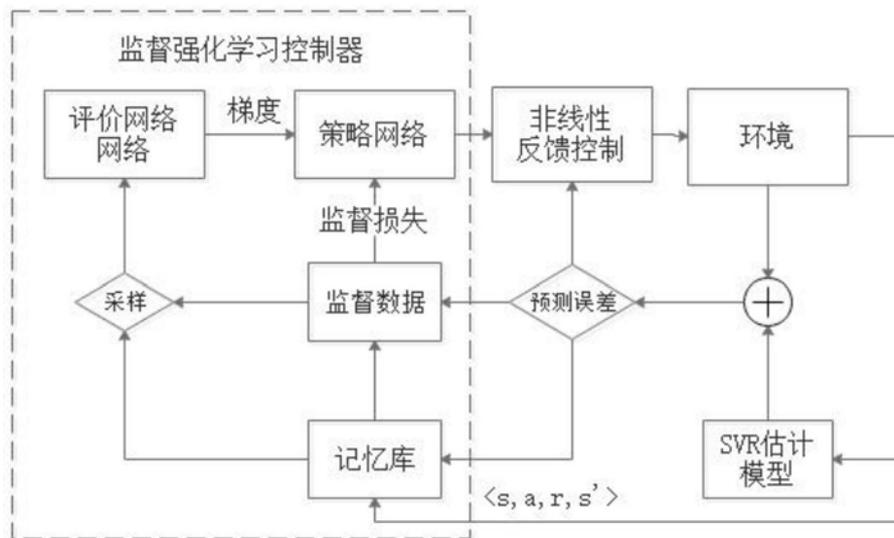


图4