



(12) 发明专利

(10) 授权公告号 CN 108803313 B

(45) 授权公告日 2022.07.12

(21) 申请号 201810589190.3

G05D 1/10 (2006.01)

(22) 申请日 2018.06.08

(56) 对比文件

(65) 同一申请的已公布的文献号

CN 106970615 A, 2017.07.21

申请公布号 CN 108803313 A

CN 102799179 A, 2012.11.28

(43) 申请公布日 2018.11.13

CN 107748566 A, 2018.03.02

(73) 专利权人 哈尔滨工程大学

CN 102819264 A, 2012.12.12

地址 150001 黑龙江省哈尔滨市南岗区南通大街145号哈尔滨工程大学科技处知识产权办公室

CN 102819264 A, 2012.12.12

CN 102819264 A, 2012.12.12

CN 102819264 A, 2012.12.12

WO 2006122030 A2, 2006.11.16

(72) 发明人 王卓 姚淑香 冯晓宁 隋炎橙

CN 102175245 A, 2011.09.07

胡磊 徐沈方 张士伟 张佩

董世超. 基于ARIMA\_BP神经网络模型海流流速预测研究.《中国科技信息》.2014, (第02期), 第86-88页第1-4节.

(74) 专利代理机构 哈尔滨市松花江专利商标事务所 23109

曹江丽. 水下机器人路径规划问题的关键技术研究.《中国博士学位论文全文数据库 信息科技辑》.2011, (第02期), 第1140-42页第5章.

专利代理师 张换男

审查员 李朝霞

(51) Int. Cl.

G05B 13/02 (2006.01)

G05B 13/04 (2006.01)

权利要求书3页 说明书6页 附图3页

(54) 发明名称

一种基于海流预测模型的路径规划方法

(57) 摘要

本发明属于水下机器人控制领域,公开了一种基于海流预测模型的路径规划方法,包含如下步骤:根据路径关键点对航行区域进行栅格化处理;利用区域海洋模式对航行区域进行海流预测,拟合计算得到实时的海流信息;利用电子海图信息标记禁航区;将不同深度的禁航信息和起点终点位置信息按照不同深度的平面栅格进行存储栅格各点的经纬度、是否为禁航区、是否到达终点;计算当前位置到终点的方向并在所有下一步行驶方向中确定可选动作;使用Q学习寻求马尔科夫决策过程规划的最优策略并输出路径。本发明充分考虑实时的海流对路径规划的影响,通过BP神经网络和bagging算法进行拟合,使用强化学习来寻求最优解,加快收敛速度,降低运算的复杂度。



CN 108803313 B

1. 一种基于海流预测模型的路径规划方法,其特征在于:包含如下步骤:

步骤(1):根据路径关键点确定航行区域,并对航行区域进行栅格化处理;

根据水下机器人路径的起点终点确定矩形航行区域;水平方向采用正交曲线网格,设定格距范围在2km~30km,垂直方向等深划分20~30层;

垂直方向采用 $\sigma$ 坐标,通过垂向变换函数和拉伸函数将垂向坐标尺度控制在 $[-1, 0]$ ,设置垂向划分的层数;

垂向变换函数:

$$z(x, y, s, t) = \eta(x, y, t) + [\eta(x, y, t) + h(x, y)] \times Z_0(x, y, s);$$

$$Z_0(x, y, s) = \frac{hc \times s + h(x, y) \times C(s)}{hc + h(x, y)}$$

上式中, $z$ 是笛卡尔坐标系的高度, $x$ 是经线坐标值, $y$ 是纬线坐标值, $s$ 是距离水面的垂向距离, $t$ 是时间, $\eta(x, y, t)$ 是随时间变化的自由海面, $h(x, y)$ 是未扰动的水体厚度, $hc$ 是转换参数;

拉伸函数:

$$C(s) = \frac{1 - \cosh(\theta_s \times s)}{\cosh \theta_s - 1}$$

上式中, $\theta_s$ 是表面控制参数, $0 < \theta_s \leq 10$ ;

步骤(2):利用区域海洋模式对航行区域进行时间步长为 $\Delta T$ 的海流预测,根据AUV实时位姿变化、控制指令和对应时刻的海流信息,利用bagging算法和BP神经网络进行拟合,计算得到实时的海流信息;

所述的步骤(2)具体为:

将控制指令对应上一时刻AUV位姿和此时刻AUV位姿及海流信息,使用bagging算法,生成 $T$ 个训练集, $T$ 个BP神经网络基于每个采样集训练出 $T$ 个基学习器,BP神经网络为三层,输入为上一时刻AUV的速度大小与角度、舵和翼及推进器的电压、此时刻AUV的速度大小和角度,输出为海流的经向速度和纬向速度,输入层为7个输入神经元、2个输出神经元,隐层神经元个数取5、8、10、12和15中的一个,通过10折交叉验证法确定隐层个数,并根据误差率按比例求得最终实时海流要素,将此刻求得的海流要素当做下一时刻的海流要素,得到实时的海流信息;

步骤(3):利用电子海图信息,将危及水下机器人安全航行的区域在栅格中标记为禁航区;

步骤(4):将不同深度的禁航信息和起点终点位置信息按照不同深度的平面栅格进行存储,存储栅格各点的经纬度、是否为禁航区、是否到达终点;

步骤(5):计算当前位置到终点的方向并在所有下一步行驶方向中确定可选动作;

所述的步骤(5)具体为:

根据矩形网格结构图,假设矩形网格中间的黑点为水下机器人的当前位置,且当前动作有 $a_1$ 、 $a_2$ 到 $a_{16}$ 共16种可能,可能的动作为从当前水下机器人位置到其外两层位置的动作,假设执行当前某个动作后所在的位置在禁航区;

设 $a_{st}$ 为当前点位置到目标点位置的动作,动作选择公式如下:

$$A_i = \frac{a_i \cdot a_{st}}{\|a_i\| \cdot \|a_{st}\|}$$

上式中,  $i$  为整数, 且  $i \in [1, 16]$ ; 选择  $A_i > 0$  的动作, 若障碍物在当前点最近 8 个格点, 则障碍物所在格点对应的动作和相邻动作都舍弃; 若障碍物在当前点外一层格点处, 则只有存在障碍物的格点对应的动作舍弃;

步骤 (6): 采用重点式学习方式, 使用 Q 学习寻求马尔科夫决策过程规划的最优策略并输出路径;

所述的步骤 (6) 包括如下步骤:

步骤 (6.1): 初始化值函数  $Q(s, a) = 0$ , 初始化原始策略  $\pi(s, a) = \operatorname{argmax}_a Q(s, a)$ ;

步骤 (6.2): 初始化状态  $S_0$  为初始位置, 确定初始时间  $t_0$ ;

步骤 (6.3): 通过神经网络计算当前位置的实时的海流速度;

步骤 (6.4): 运用重点探索策略选择动作  $a$ , 产生奖励  $r_{t+1}$ , 转移到状态  $S_{t+1}$ ;

步骤 (6.5): 根据原始策略  $\pi$ , 在状态  $S_{t+1}$  选择并执行动作  $a_{t+1}$ ;

步骤 (6.6): 更新状态动作值函数的函数值;

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \beta [r_{t+1} + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)];$$

上式中,  $\beta$  表示学习速率, 取值范围为  $[0, 1]$ ;  $\gamma$  表示折扣因子;

立即奖励函数:

$$r = \begin{cases} +50 & \text{目标点} \\ w_d r_d + w_r r_r + w_c r_c & \text{其他} \end{cases}$$

其中,  $w_d$ 、 $w_r$ 、 $w_c$  分别是距离奖惩函数、危险奖惩函数、海流奖惩函数的权重系数;

距离奖惩函数  $r_d = d(t) - d(t+1)$ ,  $d(t)$  表示  $t$  时刻机器人位置到目标点的距离,  $d(t+1)$  表示  $t+1$  时刻机器人位置到目标点的距离;

危险奖惩函数:

$$r_r = \begin{cases} 0 & d_0 > 5 \\ -5 & 2 < d_0 \leq 5 \\ -10 & 1 < d_0 \leq 2 \\ -50 & 0 < d_0 \leq 1 \end{cases}$$

$d_0$  是水下机器人当前位置与障碍物之间的格距;

海流奖惩函数  $r_c = v_c \cos |\alpha_0 - \theta|$ ,  $\alpha_0$  是航向角,  $\theta$  是海流方向;

步骤 (6.7): 利用贪婪策略更新当前策略  $\pi(s) = \operatorname{argmax}_a Q(s, a)$ ;

步骤 (6.8): 判断水下机器人是否到达目标位置的状态, 若未到达转步骤 (6.3), 若到达目标状态则转步骤 (6.9);

步骤 (6.9): 判断是否达到迭代次数或是否所有状态动作值函数收敛, 若未到达迭代次数或状态动作值函数未均收敛则转步骤 (6.2); 若到达迭代次数或所有动作值函数都收敛则输出最优策略, 得到最优规划路径;

所述的重点探索策略  $\mu(x)$  为:

$$\mu(x) = \begin{cases} 0 & \text{非可选动作} \\ a_i & \text{可选动作以概率 } p_i \end{cases}$$

上式中,

$$p_i = \frac{w_1 |A_i| + w_2 |B_i|}{w_1 \sum |A_i| + w_2 \sum |B_i|}$$

上式中,  $w_1$  是距离影响的权重系数,  $w_2$  是海流影响的权重系数,  $B_i = \frac{a_i \cdot v_c}{\|a_i\| \cdot \|v_c\|}$ ,  $v_c$  是  $t$  时刻当前位置所在格点的海流速度,  $a_i$  是概率为  $p_i$  的可选动作。

## 一种基于海流预测模型的路径规划方法

### 技术领域

[0001] 本发明属于水下机器人控制领域,尤其涉及一种基于海流预测模型的路径规划方法。

### 背景技术

[0002] 水下机器人也称无人遥控潜水器,是一种工作于水下的极限作业机器人。水下环境恶劣危险,人的潜水深度有限,所以水下机器人已成为开发海洋的重要工具。

[0003] 水下机器人可在高度危险环境、被污染环境以及零可见度的水域代替人工在水下长时间作业,水下机器人上一般配备声呐系统、摄像机、照明灯和机械臂等装置,能提供实时视频、声呐图像,机械臂能抓起重物,水下机器人在石油开发、海事执法取证、科学研究和军事等领域得到广泛应用。

[0004] 由于水下机器人运行的环境复杂,水声信号的噪声大,而各种水声传感器普遍存在精度较差、跳变频繁的缺点,因此水下机器人运动控制系统中,滤波技术显得极为重要。水下机器人运动控制中普遍采用的位置传感器为短基线或长基线水声定位系统,速度传感器为多普勒速度计会影响水声定位系统精度。因素主要包括声速误差、应答器响应时间的丈量误差、应答器位置即间距的校正误差。而影响多普勒速度计精度的因素主要包括声速 $c$ 、海水中的介质物理化学特性、运载器的颠簸等

[0005] 因此路径规划对水下机器人显得尤为重要。路径规划是水下机器人智能导航基本环节之一。水下机器人在大范围海洋环境下航行时,除考虑避障、能耗问题外,还需考虑海洋环境对水下机器人航行的影响。随时间变化的海流给水下机器人的安全及任务实现带来很大挑战,因此在进行路径规划时,需通过预测的海流要素,使水下机器人尽量利用流场中的能量,规划出能耗少的可行的安全路径。

[0006] 从算法策略角度,目前路径规划算法可分为基于智能计算的路径规划,基于行为、学习心理的路径规划和随机采样路径规划。这些算法目的多是提高解空间搜索效率、加快收敛,或是针对未知环境或动态空间提出的,且目前越来越多的学者开始海流影响下的路径规划的研究。从流函数生成海流场到海流历史统计数据库生成海流场,本发明是利用区域海洋模式来预测某地区未来时刻海流场数据和AUV位姿及控制指令,来实时预测海流,使用来做路径规划的海流场更加精确,更具有实时性。

[0007] 申请号为2017110538828.6的专利公开了一种基于探测威胁域的无人潜航器路径规划装置及方法,基于探测威胁域的路径规划算法来解决地形障碍环境下UUV的路径规划问题,能满足UUV本身运动学约束、避碰约束以及隐蔽探测约束。在给定初始位置、终点位置、最大曲率约束、路径离散点分辨率、隐蔽安全指标等,规划出从运动起点到终点的路径,且光滑连续可导,满足UUV的航行转弯曲率约束、隐蔽安全指标等,使其以最短时间安全隐蔽到达终点。本发明首次将探测威胁理论与航行转弯曲率约束的几何理论应用到UUV的路径规划领域中,能快速实现路径规划,方法简单可靠,易于实现,计算量小,实时性较好,能满足路径规划要求,提高了路径规划的实用性,对今后水下路径规划领域的发展有着积极

意义。但该方法应用在水下航行器的路径规划时,存在计算过程过于复杂,实时性不好的问题。

## 发明内容

[0008] 本发明的目的在于公开能耗少,安全性高的一种基于海流预测模型的路径规划方法。

[0009] 本发明的目的是这样实现的:

[0010] 一种基于海流预测模型的路径规划方法,包含如下步骤:

[0011] 步骤(1):根据路径关键点确定航行区域,并对航行区域进行栅格化处理;

[0012] 步骤(2):利用区域海洋模式对航行区域进行时间步长为 $\Delta T$ 的海流预测,根据AUV实时位姿变化、控制指令和对应时刻的海流信息,利用bagging算法和BP神经网络进行拟合,计算得到实时的海流信息:

[0013] 将控制指令对应上一时刻AUV位姿和此时刻AUV位姿及海流信息,使用bagging算法,生成T个训练集,T个BP神经网络基于每个采样集训练出T个基学习器,BP神经网络为三层,输入为上一时刻AUV的速度大小与角度、舵和翼及推进器的电压、此时刻AUV的速度大小和角度,输出为海流的经向速度和纬向速度,输入层为7个输入神经元、2个输出神经元,隐层神经元个数取5、8、10、12和15中的一个,通过10折交叉验证法确定隐层个数,并根据误差率按比例求得最终实时海流要素,将此时刻求得的海流要素当做下一时刻的海流要素,得到实时的海流信息。

[0014] 步骤(3):利用电子海图信息,将危及水下机器人安全航行的区域在栅格中标记为禁航区;

[0015] 步骤(4):将不同深度的禁航信息和起点终点位置信息按照不同深度的平面栅格进行存储存储栅格各点的经纬度、是否为禁航区、是否到达终点;

[0016] 步骤(5):计算当前位置到终点的方向并在所有下一步行驶方向中确定可选动作:

[0017] 根据矩形网格结构图,假设矩形网格中间的黑点为水下机器人的当前位置,且当前动作有 $a_1$ 、 $a_2$ 到 $a_{16}$ 共16种可能,可能的动作为从当前水下机器人位置到其外两层位置的动作,假设执行当前某个动作后所在的位置在禁航区;

[0018] 设 $a_{st}$ 为当前点位置到目标点位置的动作,动作选择公式如下:

$$[0019] \quad A_i = \frac{a_i \cdot a_{st}}{\|a_i\| \cdot \|a_{st}\|};$$

[0020] 上式中, $i$ 为整数,且 $i \in [1, 16]$ ;选择 $A_i > 0$ 的动作,若障碍物在当前点最近8个格点,则障碍物所在格点对应的动作和相邻动作都舍弃;若障碍物在当前点外一层格点处,则只有存在障碍物的格点对应的动作舍弃。

[0021] 步骤(6):采用重点式学习方式,使用Q学习寻求马尔科夫决策过程规划的最优策略并输出路径。

[0022] 步骤(6.1):初始化值函数 $Q(s, a) = 0$ ,初始化原始策略 $\pi(s, a) = \operatorname{argmax}_a Q(s, a)$ ;

[0023] 步骤(6.2):初始化状态 $S_0$ 为初始位置,确定初始时间 $t_0$ ;

[0024] 步骤(6.3):通过神经网络计算当前位置的实时的海流速度;

[0025] 步骤(6.4):运用重点探索策略选择动作 $a$ ,产生奖励 $r_{t+1}$ ,转移到状态 $S_{t+1}$ ;

[0026] 重点探索策略 $\mu(x)$ ：

$$[0027] \quad \mu(x) = \begin{cases} 0 & , \text{非可选动作} \\ a_i & , \text{可选动作以概率} p_i \end{cases}$$

[0028] 上式中，

$$[0029] \quad p_i = \frac{w_1 |A_i| + w_2 |B_i|}{w_1 \sum |A_i| + w_2 \sum |B_i|} ;$$

[0030] 上式中， $w_1$ 是距离影响的权重系数， $w_2$ 是海流影响的权重系数； $B_i = \frac{a_i \cdot v_c}{\|a_i\| \cdot \|v_c\|}$ ， $v_c$ 是

$t$ 时刻当前位置所在格点的海流速度， $a_i$ 是概率为 $p_i$ 的可选动作。

[0031] 步骤(6.5)：根据原始策略 $\pi$ ，在状态 $S_{t+1}$ 选择并执行动作 $a_{t+1}$ 。

[0032] 步骤(6.6)：更新状态动作值函数的函数值：

$$[0033] \quad Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \beta [r_{t+1} + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)] ;$$

[0034] 上式中， $\beta$ 表示学习速率，取值范围为 $[0, 1]$ ； $\gamma$ 表示折扣因子。

[0035] 步骤(6.7)：利用贪婪策略更新当前策略 $\pi(s) = \underset{a}{\arg \max} Q(s, a)$ ；

[0036] 步骤(6.8)：判断水下机器人是否到达目标位置的状态，若未到达转步骤(6.3)，若到达目标状态则转步骤(6.9)。

[0037] 步骤(6.9)：判断是否达到迭代次数或是否所有状态动作值函数收敛，若未到达迭代次数或状态动作值函数未均收敛则转步骤(6.2)；若到达迭代次数或所有动作值函数都收敛则输出最优策略，得到最优规划路径。

[0038] 本发明的有益效果为：

[0039] 本发明充分考虑实时的海流对路径规划的影响，通过区域海洋模式预测未来的海洋要素，利用BP神经网络和bagging算法进行拟合，得到实时的海流信息。同时根据马尔科夫决策过程进行规划，使用强化学习来寻求最优解，加快收敛速度，降低运算的复杂度，更好更快地得到规划路径。

## 附图说明

[0040] 图1是一种基于海流预测模型的路径规划方法流程图；

[0041] 图2是矩形网格结构图；

[0042] 图3是动作选择示意图；

[0043] 图4是马尔科夫决策过程规划流程图。

## 具体实施方式

[0044] 下面结合附图来进一步描述本发明：

[0045] 实施例1：

[0046] 如图1，一种基于海流预测模型的路径规划方法，包含如下步骤：

[0047] 步骤(1)：根据路径关键点确定航行区域，并对航行区域进行栅格化处理；

[0048] 根据水下机器人路径的起点终点确定矩形航行区域；水平方向采用正交曲线网

格,设定格距范围在2km~30km,垂直方向等深划分20~30层。

[0049] 步骤(2):利用区域海洋模式对航行区域进行时间步长为 $\Delta T$ 的海流预测,根据AUV实时位姿变化、控制指令和对应时刻的海流信息,利用bagging算法和BP神经网络进行拟合,计算得到实时的海流信息:

[0050] 垂直方向采用 $\sigma$ 坐标,通过垂向变换函数和拉伸函数将垂向坐标尺度控制在 $[-1, 0]$ ,设置垂向划分的层数;

[0051] 垂向变换函数:

[0052]  $z(x, y, s, t) = \eta(x, y, t) + [\eta(x, y, t) + h(x, y)] \times Z_0(x, y, s);$

[0053]  $Z_0(x, y, s) = \frac{hc \times s + h(x, y) \times C(s)}{hc + h(x, y)};$

[0054] 上式中, $z$ 是笛卡尔坐标系的高度, $x$ 是经线坐标值, $y$ 是纬线坐标值, $s$ 是距离水面的垂向距离, $t$ 是时间, $\eta(x, y, t)$ 是随时间变化的自由海面, $h(x, y)$ 时未扰动的水体厚度, $hc$ 是转换参数;

[0055] 拉伸函数:

[0056]  $C(s) = \frac{1 - \cosh(\theta_s \times s)}{\cosh \theta_s - 1};$

[0057] 上式中, $\theta_s$ 是表面控制参数, $0 < \theta_s \leq 10$ 。

[0058] 区域海洋模式初始条件由四维同化实现,边界条件由全球模式的预报场经差分而来,空间上采用中央差分格式,时间上采用蛙跳格式,设置时间步长为5min,对航行区域的海流场进行预测并保存在文件中。

[0059] 将控制指令对应上一时刻AUV位姿和此时刻AUV位姿及海流信息,使用bagging算法,生成 $T$ 个训练集, $T$ 个BP神经网络基于每个采样集训练出 $T$ 个基学习器,BP神经网络为三层,输入为上一时刻AUV的速度大小与角度、舵和翼及推进器的电压、此时刻AUV的速度大小和角度,输出为海流的经向速度和纬向速度,输入层为7个输入神经元、2个输出神经元,隐层神经元个数取5、8、10、12和15中的一个,通过10折交叉验证法确定隐层个数,并根据误差率按比例求得最终实时海流要素,将此时刻求得的海流要素当做下一时刻的海流要素,得到实时的海流信息。

[0060] 步骤(3):利用电子海图信息,将危及水下机器人安全航行的区域在栅格中标记为禁航区;

[0061] 步骤(4):将不同深度的禁航信息和起点终点位置信息按照不同深度的平面栅格进行存储存储栅格各点的经纬度、是否为禁航区、是否到达终点;

[0062] 步骤(5):计算当前位置到终点的方向并在所有下一步行驶方向中确定可选动作:

[0063] 如图2,根据矩形网格结构图,假设矩形网格中间的黑点为水下机器人的当前位置,且当前动作有 $a_1$ 、 $a_2$ 到 $a_{16}$ 共16种可能,可能的动作为从当前水下机器人位置到其外两层位置的动作,假设执行当前某个动作后所在的位置在禁航区;

[0064] 如图3,设 $a_{st}$ 为当前点位置到目标点位置的动作,动作选择公式如下:

[0065]  $A_i = \frac{a_i \cdot a_{st}}{\|a_i\| \cdot \|a_{st}\|};$

[0066] 上式中, $i$ 为整数,且 $i \in [1, 16]$ ;选择 $A_i > 0$ 的动作,若障碍物在当前点最近8个格



点,则障碍物所在格点对应的动作和相邻动作都舍弃;若障碍物在当前点外一层格点处,则只有存在障碍物的格点对应的动作舍弃。

[0067] 步骤(6):采用重点式学习方式,使用Q学习寻求马尔科夫决策过程规划的最优策略并输出路径:

[0068] 马尔科夫决策过程由五元组(S,A,P,R, $\gamma$ )描述,其中:

[0069] S为有限状态集,A为有限动作集,P为状态转移概率,R为回报函数, $\gamma$ 为折扣因子,用来计算累计回报。

[0070] 强化学习的目标是给定一个马尔科夫决策过程,寻求最优策略。所谓策略是指状态到动作的映射,常用符号 $\pi$ 表示。水下机器人通过对未知环境用策略探索动作根据回报优化策略来规划最优路径,当一个动作产生正回报时,动作被加强,下次再出现相同状态时则会有大概率选择该动作,反之动作被削弱,通过不断与环境交互来寻求最优策略。由于其固有的适应能力、反应能力和在线学习能力多被用在未知环境的路径规划中,其中Q学习应用最广泛。具体步骤如图4所示。

[0071] 步骤(6.1):初始化值函数 $Q(s,a)=0$ ,初始化原始策略 $\pi(s,a)=\operatorname{argmax}_a Q(s,a)$ ;

[0072] 步骤(6.2):初始化状态 $S_0$ 为初始位置,确定初始时间 $t_0$ ;

[0073] 步骤(6.3):通过神经网络计算当前位置的实时的海流速度;

[0074] 步骤(6.4):运用重点探索策略选择动作a,产生奖励 $r_{t+1}$ ,转移到状态 $S_{t+1}$ ;

[0075] 重点探索策略:

$$[0076] \quad \mu(x) = \begin{cases} 0 & \text{,非可选动作} \\ a_i & \text{,可选动作以概率} p_i \end{cases};$$

[0077] 上式中,

$$[0078] \quad p_i = \frac{w_1 |A_i| + w_2 |B_i|}{w_1 \sum |A_i| + w_2 \sum |B_i|};$$

[0079] 上式中, $w_1$ 是距离影响的权重系数, $w_2$ 是海流影响的权重系数; $B_i = \frac{a_i \cdot v_c}{\|a_i\| \cdot \|v_c\|}$ , $v_c$ 是t

时刻当前位置所在格点的海流速度, $a_i$ 是概率为 $p_i$ 的可选动作;

[0080] 立即奖励函数:

$$[0081] \quad r = \begin{cases} +100 & \text{,目标点} \\ w_d r_d + w_r r_r + w_c r_c & \text{,其他} \end{cases};$$

[0082] 上式中, $w_d$ 是距离奖惩函数的权重系数, $w_r$ 是危险奖惩函数的权重系数, $w_c$ 是海流奖惩函数的权重系数;

[0083] 距离奖惩函数 $r_d = d(t) - d(t+1)$ , $d(t)$ 表示t时刻机器人位置到目标点的距离, $d(t+1)$ 表示t+1时刻机器人位置到目标点的距离。

[0084] 危险奖惩函数:

$$[0085] \quad r_r = \begin{cases} 0 & d_o > 5 \\ -5 & 2 < d_o \leq 5 \\ -10 & 1 < d_o \leq 2 \\ -50 & 0 < d_o \leq 1 \end{cases};$$

[0086] 上式中,  $d_0$  是水下机器人当前位置与障碍物之间的格距;

[0087] 海流奖惩函数  $r_c = v_c \cos |\alpha - \theta|$ ,  $\alpha$  是航向角,  $\theta$  是海流方向;

[0088] 根据重点探索策略选择的动作  $a$ , 产生奖励  $r_{t+1}$ , 转移到状态  $S_{t+1}$ ;

[0089] 步骤(6.5): 根据原始策略  $\pi$ , 在状态  $S_{t+1}$  选择并执行动作  $a_{t+1}$ 。

[0090] 步骤(6.6): 更新状态动作值函数的函数值:

[0091]  $Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \beta [r_{t+1} + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)]$ ;

[0092] 上式中,  $\beta$  表示学习速率, 取值范围为  $[0, 1]$ ;  $\gamma$  表示折扣因子。

[0093] 步骤(6.7): 利用贪婪策略更新当前策略  $\pi(s) = \underset{a}{\arg \max} Q(s, a)$ ;

[0094] 步骤(6.8): 判断水下机器人是否到达目标位置的状态, 若未到达转步骤(6.3), 若到达目标状态则转步骤(6.9)。

[0095] 步骤(6.9): 判断是否达到迭代次数或是否所有状态动作值函数收敛, 若未到达迭代次数或状态动作值函数未均收敛则转步骤(6.2); 若到达迭代次数或所有动作值函数都收敛则输出最优策略, 得到最优规划路径。

[0096] 与现有技术相比, 本发明充分考虑实时的海流对路径规划的影响, 通过区域海洋模式预测未来的海洋要素, 利用BP神经网络和bagging算法进行拟合, 得到实时的海流信息。同时根据马尔科夫决策过程进行规划, 使用强化学习来寻求最优解, 加快收敛速度, 降低运算的复杂度, 更好更快地得到规划路径。

[0097] 以上所述并不用于限制本发明, 对于本领域的技术人员来说, 本发明可以有各种更改和变化。凡在本发明的精神和原则之内, 所作的任何修改、等同替换、改进等, 均应包含在本发明的保护范围之内。

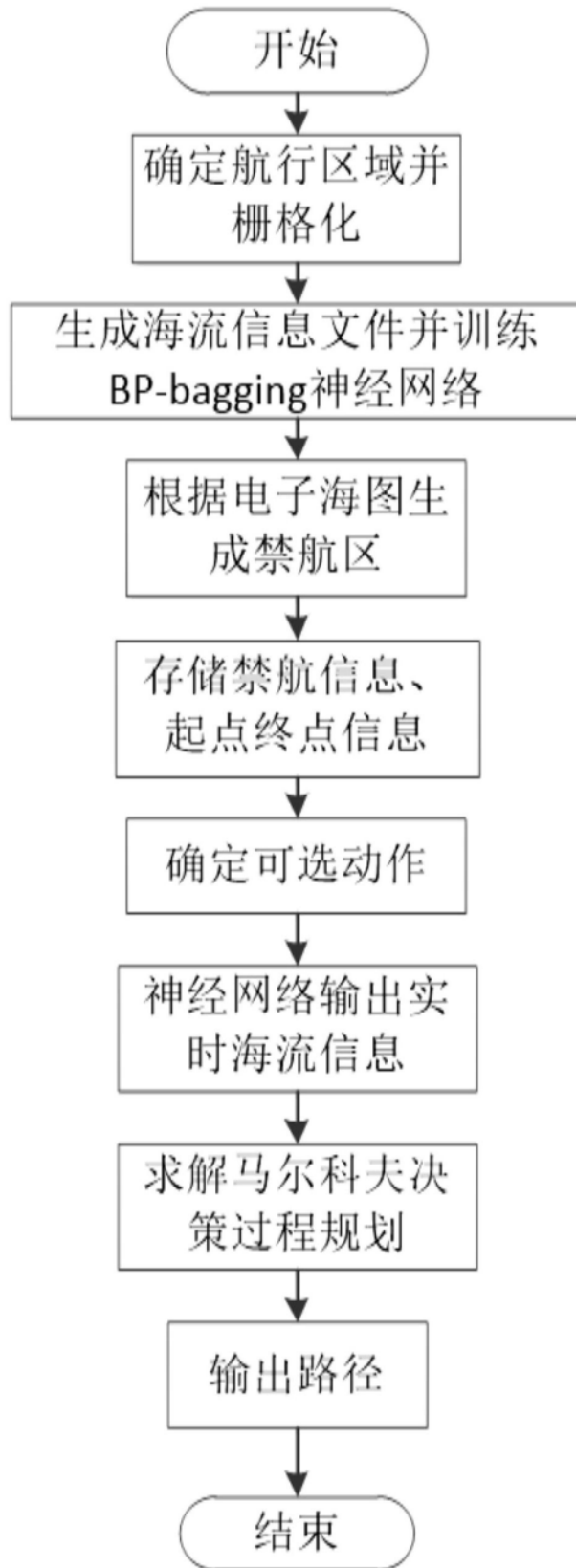


图1

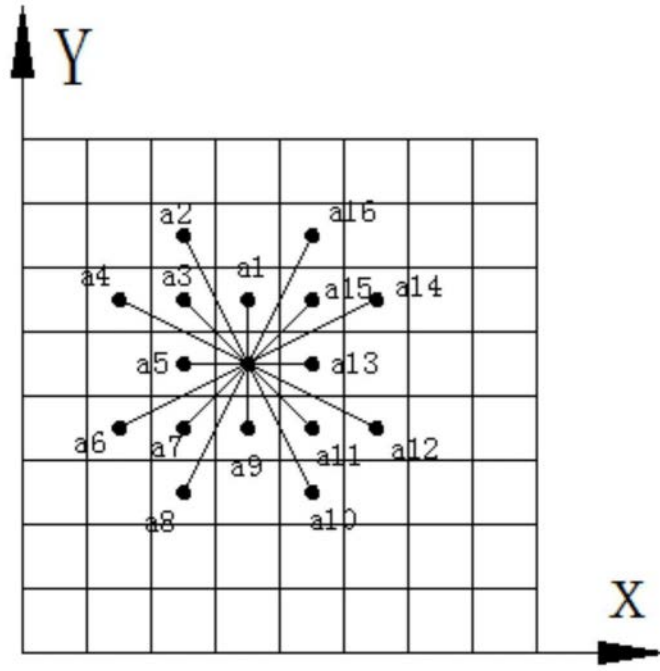


图2

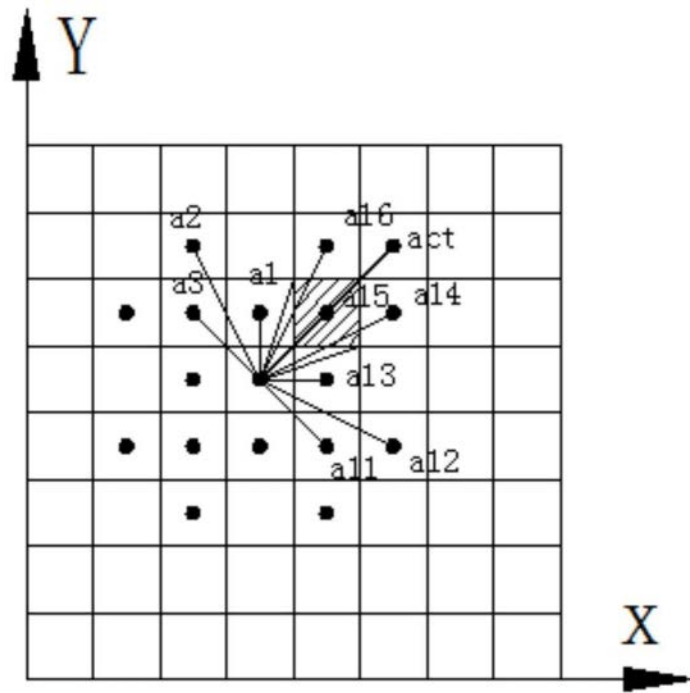


图3

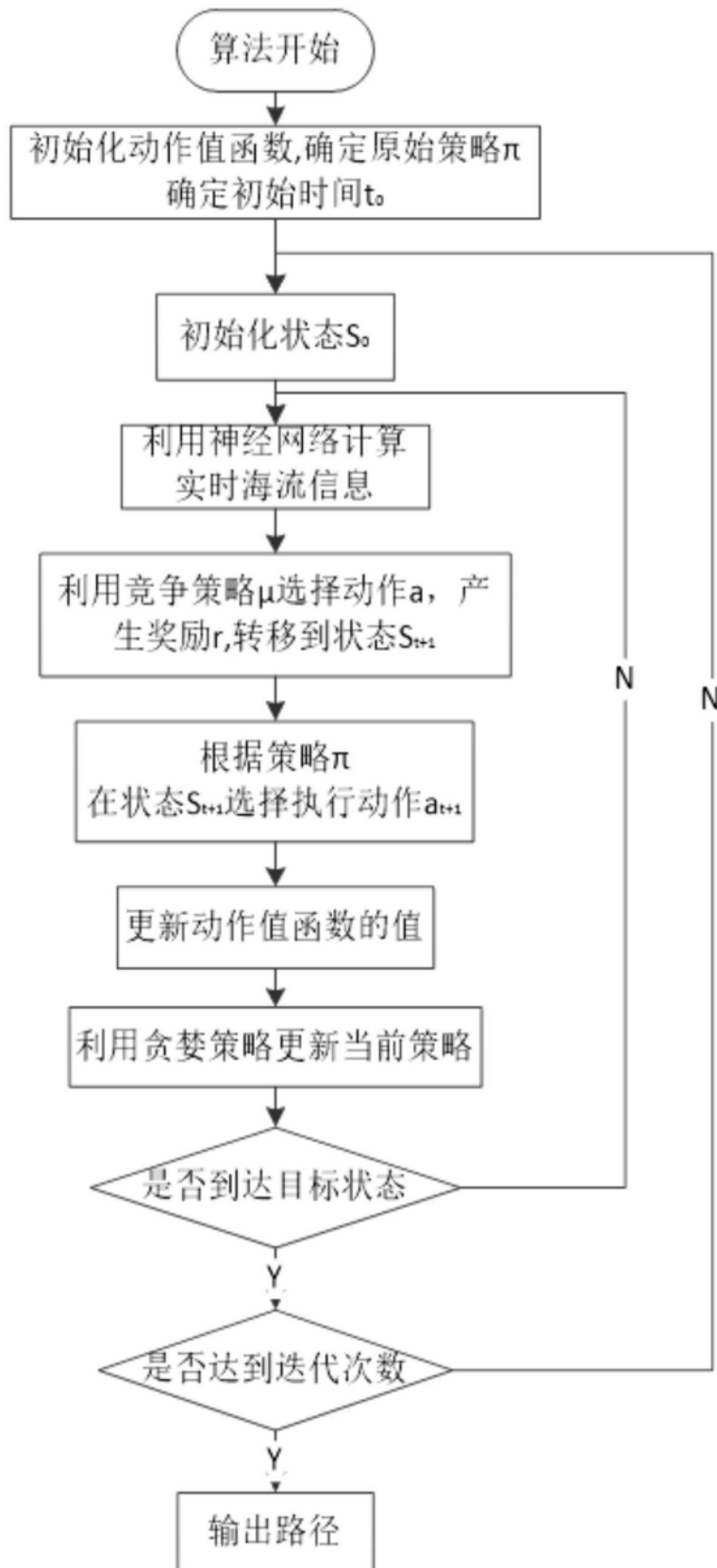


图4