

一种基于 Q-Learning 的高超声速飞行器动态路径规划方法

技术领域:

本发明属于高超声速飞行器轨迹设计领域，具体涉及一种基于 Q-Learning 的高超声速飞行器动态路径规划方法。

背景技术:

近年来，高超声速飞行器在航空航天领域发展迅速，以其独特的优势成为世界关注的焦点。飞行器具有飞行速度快、反应时间短、作战半径大、隐蔽性好、穿刺力强等特点。路径规划是执行飞行任务的重要技术，特别是面对来自地面或空间的未知威胁。飞行器路径规划的目的是在自机动性能、敌人威胁和飞行时间等约束条件下，寻找最优或次最优路径以有效规避威胁。

随着空天任务难度的增加，特别是在未知环境下，路径规划需要考虑不确定性因素的影响，要求具有学习能力，以适应环境变化的不确定性。

强化学习技术发展使得路径规划可以不再依赖于环境模型，也不需要环境的先验知识。因此，利用强化学习进行路径规划以提高飞行器对未知环境的适应性的方法得到高度重视。Q-learning 算法是强化学习的典型方法，目前用 Q-learning 为飞行器规划路径的相关研究十分少见。

发明内容:

针对以上背景，为解决现有的路径规划算法不能为未知环境中的飞行器规划避障路径的问题，本发明提供了一种基于 Q-learning 算法的动态路径规划算法，基于 Q-learning 算法，结合飞行器的动态特性，能够为飞行器在未知环境中规划出一条可飞的避障路径。

一种高超声速飞行器的动态路径规划设计，其技术方案包括以下步骤:

1) 场景建立:

在飞行器巡航阶段，进入定高定速飞行的飞行空间，需要一种有效的状态空间划分方法，该方法必须可以清晰地描述飞行器所处的环境。考虑到网格方法在表示二维环境信息方面简单、方便、高效，采用网格方法对环境进行建模。将飞行器的巡航区域划分为 $n \times n$ 个网格，每个网格的长度为 m km。 m 的值应大于或等于飞行器的最小转弯半径 R ，转弯半径由飞行器过载决定，以便于飞行器可以正常机动。

2) 动作选择集:

由于网格的尺寸大小符合飞行器的最小转弯半径的要求,所以可以认为飞行器够在它周围的 8 个方向自由移动的,如图 3 所示。但是为了加快整个学习过程,缩短飞行器到达目的地的时间。利用当前位置点坐标与目标点位置坐标的关系建立新的飞行器动作集。具体方案如下:设当前位置点坐标为 (x_c, y_c) ,目标点位置坐标为 (x_g, y_g) 。

若 $x_c < x_g$ 且 $y_c < y_g$, 则 $a = rand(0, 1, 2)$;

若 $x_c > x_g$ 且 $y_c > y_g$, 则 $a = rand(4, 5, 6)$;

若 $x_c < x_g$ 且 $y_c > y_g$, 则 $a = rand(2, 3, 4)$;

若 $x_c > x_g$ 且 $y_c < y_g$, 则 $a = rand(0, 6, 7)$;

若 $x_c = x_g$ 或 $y_c = y_g$, 则 $a = rand(0, 4)$ 或 $a = rand(2, 6)$;

其他情况下, $a = rand(a, a-1, a+1)$ 。

3) 奖励值选择:

对于 Q-Learning 算法来说,它的目的是能够使 agent 在与环境的交互中(从初始状态到目标状态)获得最大的累积奖励。因此,建立一个合理的奖励函数是非常重要的。具体的奖励值由多次实验的数据得出。

4) 动态路径规划:

具体的路径规划步骤如下:

步骤 1:初始化 q 函数, 状态和动作为任意值。

步骤 2:观察当前状态 s, 使用动作选择策略选择一个动作 a。

步骤 3:执行选择的动作 a, 观察收到的即时奖励 r 和随后的状态 s。

步骤 4:根据 $Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_{t+1} + \gamma \times \max_a Q(s_{t+1}, a) - Q(s_t, a_t)]$ 更新 q 函数值。

步骤 5:如果新状态满足终端状态,则结束本次试验的学习,开始下一次试验的训练。否则,返回执行步骤 3。

附图说明:

图 1 为本发明的步骤流程图。

图 2 为飞行器抽象的有限环境模型图。

图 3 为飞行器的动作选择集图。

图 4 为某实例的路径规划航迹图。

图 5 为某实例飞行器迭代次数随实验次数的变化。

图 6 为某实例飞行器迭代次数的标准偏差。

具体实施方式：

为使本发明的目的、技术方案和优点更加清楚，下面将结合附图对本发明的实施方式作进一步详细描述。

一种高超声速飞行器巡航段动态路径规划方法，考虑高超声速飞行器整个巡航段处于定高定速飞行，仅考虑横向平面内的运动，包括如下步骤：

1) 场景建立：

采用网格方法对环境进行建模。将飞行器的巡航区域划分为 $n \times n$ 个网格，每个网格的长度为 m km。 m 的值满足大于或等于飞行器的最小转弯半径 R ，转弯半径由飞行器过载决定，保证飞行器的正常机动。如图 2 所示。

2) 动作选择集：

网格的尺寸大小符合飞行器的最小转弯半径的要求，飞行器可以在它周围的 8 个方向自由移动，如图 3 所示。利用当前位置点坐标与目标点位置坐标的关系建立新的飞行器动作集，以此加快整个学习过程，缩短飞行器到达目的地的时间。具体方案如下：设当前位置点坐标为 (x_c, y_c) ，目标点位置坐标为 (x_g, y_g) 。

若 $x_c < x_g$ 且 $y_c < y_g$ ，则 $a = rand(0, 1, 2)$ ；

若 $x_c > x_g$ 且 $y_c > y_g$ ，则 $a = rand(4, 5, 6)$ ；

若 $x_c < x_g$ 且 $y_c > y_g$ ，则 $a = rand(2, 3, 4)$ ；

若 $x_c > x_g$ 且 $y_c < y_g$ ，则 $a = rand(0, 6, 7)$ ；

若 $x_c = x_g$ 或 $y_c = y_g$ ，则 $a = rand(0, 4)$ 或 $a = rand(2, 6)$ ；

其他情况下， $a = rand(a, a-1, a+1)$ 。

3) 奖励值选择：

通过多次实验的数据得出飞行器在遇到不同情况时的奖励值,进而建立一个合理的奖励函数。

4) 飞行器的动态路径规划:

具体的路径规划步骤如下:

步骤 1:输入环境,初始化 q 函数,状态和动作为任意值。

步骤 2:观察当前状态 s , 使用动作选择策略选择一个动作 a 。

步骤 3:执行选择的动作 a , 观察收到的即时奖励 r 和随后的状态 s 。

步骤 4:根据 $Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_{t+1} + \gamma \times \max Q(s_{t+1}, a) - Q(s_t, a_t)]$ 更新 q 函数值。

步骤 5:如果新状态满足终端状态,则结束本次试验的学习,开始下一次试验的训练。否则,返回执行步骤 3。

最终为飞行器规划出一条最优路径。

具体实施案例:

通过仿真验证了该算法在高超声速飞行器遇到未知障碍时的有效性。

Q-learning 算法所需训练参数的具体值如表 1 所示。

表 1 参数设置

参数	值
划分区域 $n \times n$	30
每个网格长度 m (km)	15
学习率 α	0.3
折扣因子 γ	0.95
探索因子 ϵ	0.5
最大实验次数	5000
最大迭代次数	1800
收敛目标	0.25

将飞行器起始点设为(15,390),即对应网格(1,26)中的坐标点,用蓝星表示。目标位置为(405,45),即对应网格(27,3)中的坐标点,用红星表示。随机放置三个障碍物作为未知威胁。

仿真结果如图 4 所示，圆点代表飞行器，黑色实心圆代表障碍物。因此，从图中可以看出，飞行器从原点位置出发，成功避开障碍物，最终到达目标位置。

在图 5 中，x 轴为试验次数，每次试验 **agent** 都从初始状态开始，直到达到目标状态。y 轴表示迭代次数，即到达目的地的时间步长。说明随着试验次数的增加，飞行器到达终点的迭代次数会减少，随着试验次数的增加，到达终点的最终步骤基本稳定。

在图 6 中，x 轴与图 5 相同，y 轴表示迭代次数的标准差。结果表明，随着试验次数的增加，标准差逐渐减小。