

基于 IMM 的无人机在线路径规划决策建模

杨启明¹, 徐建城¹, 田海宝², 吴勇¹

(1.西北工业大学 电子信息学院, 陕西 西安 710072; 2.空军驻江西地区军事代表室, 江西 南昌 330024)

摘要:为提升无人机对目标的自主跟踪能力,以部分可观测马尔科夫决策过程(POMDP)为理论框架,建立起无人机路径在线规划 POMDP 模型。分析并描述了 POMDP 模型中的各个要素,针对目标运动规律的复杂性,引入交互多模型(IMM)方法描述 POMDP 模型中的状态转移规律,以适应目标的机动变化。同时以 POMDP 模型中的累加代价函数为目标函数,结合使用名义信念状态优化(NBO)算法求解无人机的行动策略,产生的行动策略控制无人机飞行。仿真结果表明,所建立的模型能够实现无人机路径的自主规划,能够控制无人机对目标进行有效跟踪,规划的无人机路径较之使用单一的目标状态转移规律更加合理高效。

关键词:部分可观测马尔科夫决策过程(POMDP);交互多模型(IMM);路径规划;目标跟踪;名义信念状态优化

中图分类号:V279

文献标志码:A

文章编号:1000-2758(2018)02-0323-09

无人机(UAV)在战争中的应用和出色表现已经使其作为一种新型作战力量受到越来越多的关注,与有人驾驶飞机相比,无人机具有隐蔽性好、不惧伤亡、可进行大过载机动等诸多优势。无人机的智能化执行任务可以让人员的精力从繁重的具体操控上解脱出来,转而进行态势判断和推理等高级思考,这样能真正发挥人和机各自的优势,提升综合作战效率,因此无人机的智能化是其未来的发展方向^[1-3]。无论军用还是民用领域,对运动目标实施跟踪是无人机的一个主要用途,实现跟踪自主化,能够在大幅降低操作强度的同时提高对目标信息探测的稳定性。本文试图对无人机跟踪目标的任务进行自主化设计,无人机周期地计算出行动策略,进而对飞行路径进行动态调整和规划以实现目标的跟踪。

在状态存在随机性的规划问题中,多使用部分可观测的马尔科夫决策过程(POMDP)为建模方法。在基于 POMDP 的无人机路径规划应用方面有学者已经做了一些研究,例如文献[4]基于 POMDP 对无人机在线性观测方程下的路径规划进行了建模;文献[5]基于 POMDP 进行无人机对目标探测识别的

建模;文献[6]基于 POMDP 对两栖车辆的运动路线进行了规划设计。上述文献均使用 POMDP 框架对路径规划问题进行了表述和求解,并取得了较为满意的结果,但是这些文献中均对目标的状态转移规律使用单一的运动模型(例如,匀速直线运动)来进行描述,这样与真实情况下的目标运动不相符合。本文以 POMDP 理论为基础,提出在 POMDP 模型中使用交互多模型(IMM)方法描述目标状态转移规律,用 IMM 方法实现模型的切换^[7-8],旨在构建适合于目标的机动变化下无人机跟踪路径的在线规划算法。

POMDP 模型对问题有很强的表现力,但是对其精确求解十分困难^[9]。累加代价的计算量会随着问题规模呈指数增长,此外观测的不完整性使得所获得状态信息具有随机性,因此对状态和观测信息均需用概率分布来描述,对远期代价和高维概率分布的精确计算需要较大的计算资源和时间开销。因此在无人机行动决策这样对实时性要求较高的在线规划场景,主要采用近似解法对模型进行求解,在有限时间内实时求出符合使用要求的次优解。在进行近似求解时,本文通过限定预测时限来降低远期

代价的计算量,使用名义信念状态优化(NBO)算法^[10]求解 POMDP 的行动序列,在满足跟踪性能的同时极大的降低了计算量。

1 路径规划问题建模

1.1 路径规划问题描述

无人机跟踪目标的路径规划的背景想定如下,无人机对地面一个移动目标进行观测并自动跟踪飞行。无人机使用传感器获得目标的距离和方位值,并以此为观测量自动调整无人机的飞行轨迹对目标保持跟踪,提供准确稳定的视频图像等目标信息。为了有效说明算法,本文假设无人机在固定高度飞行,简化无人机和目标的运动为二维平面内的运动。

1.2 路径规划 POMDP 各个要素

路径规划问题的 POMDP 模型可由 6 个要素的六元组 $\langle S, A, T, O, C, B \rangle$ 来表示,其中 S 表示状态空间, A 为行动空间, T 为状态转移规律, O 为观测和观测率, C 表示代价函数, B 为信念空间。下面对本文中各个要素代表的意义进行说明。

1.2.1 状态空间

在无人机路径规划的问题中,状态空间 S 中应包括 3 个子系统的状态来完成对问题的描述,分别为无人机的状态 x_k 、目标的状态 z_k 、和滤波状态 (ξ_k, P_k) ,因此状态空间 $S_k = (x_k, z_k, \xi_k, P_k)$ 。无人机的状态 x_k 表示 k 时刻无人机所处的位置和速度,用一个四维列向量 $x_k = (p_k^x, p_k^y, v_k, \theta_k)^T$ 表示,其中 p_k^x 和 p_k^y 分别为无人机横坐标和纵坐标的值,用以表示无人机的位置 x_k^{pos} , v_k 表示无人机的速度大小值, θ_k 表示无人机的速度方向。目标的状态 z_k 表示 k 时刻目标准确的位置和速度, $z_k = (q_k^x, q_k^y, u_k^x, u_k^y)^T$, 其中 q_k^x 和 q_k^y 分别为目标的横坐标和纵坐标值,表示目标的位置 z_k^{pos} , u_k^x 和 u_k^y 分别为目标的横向速度值和纵向速度值。滤波状态 (ξ_k, P_k) 表示滤波算法对目标的估计状态, ξ_k 表示滤波算法对目标状态的后验估计值, P_k 表示滤波算法对目标状态后验估计的协方差矩阵。

1.2.2 行动空间

在跟踪目标的无人机路径规划问题中,每个决策点所采取的行动应能改变无人机的飞行状态,进而实时调整飞行路径。因此,本文采取加速度值 a_k 和倾斜角 ϕ_k 作为行动值。在本时刻采取 1 组行动

值,在下一时刻就可以改变无人机的状态。行动空间 $A_k = (a_k, \phi_k)$ 。

1.2.3 观测和观测率

无人机获得的观测值由于设备误差和环境干扰通常是包含噪声的。因此,观测方程可以定义为状态 S_k 和观测噪声 η_k 的函数,如(1)式所示

$$z_k = h(S_k) + \eta_k \quad (1)$$

传感器所获得的观测量是目标相对于无人机的距离和方位角信息^[11-12]。因此,将 $h(S_k)$ 定义为:

$$h(S_k) = \begin{bmatrix} \sqrt{(q_k^x - p_k^x)^2 + (q_k^y - p_k^y)^2} \\ \arctan \frac{q_k^y - p_k^y}{q_k^x - p_k^x} \end{bmatrix} \quad (2)$$

(1)式中, η_k 为传感器的观测噪声序列,探测目标信息的过程中, η_k 的分布与无人机和目标的位置相关,当无人机和目标距离较近时,观测噪声小,观测精度高,当距离较远时,空间内各种噪声较大,观测的误差较大。 η_k 的协方差可以表示为

$$R_k = R(x_k, \xi_k) \quad (3)$$

设 R_k 表示距离测量不确定性 m 和角度测量不确定性 n 。如果 d_k 表示 k 时刻目标和无人机传感器之间的距离,那么传感器相应的测距和测角标准差可以分别写为 $\sigma_{\text{range}}(k) = (m/100) \cdot d_k$ 和 $\sigma_{\text{angle}}(k) = n \cdot d_k$ 。由于信息矩阵的计算需要量测协方差矩阵的逆矩阵,当目标与无人机的二维距离为 0 时,量测协方差矩阵的逆矩阵将无法计算,为了解决这一问题,定义 d_{eff} 为目标和传感器之间的有效距离, $d_{\text{eff}}(k) = \sqrt{d_k^2 + b^2}$ 。在实际距离 d_k 上增加一个非 0 的极小数 b 。测距和测角标准差标准差改写为 $\sigma_{\text{range}}(k) = (m/100) \cdot d_{\text{eff}}(k)$ 和 $\sigma_{\text{angle}}(k) = n \cdot d_{\text{eff}}(k)$ 。 φ_k 为 k 时刻无人机和目标之间的方位角,则 R_k 可以通过如下式子计算

$$R_k = M_k \begin{bmatrix} \sigma_{\text{range}}^2(k) & 0 \\ 0 & \sigma_{\text{angle}}^2(k) \end{bmatrix} M_k^T \quad (4)$$

式中, $M_k = \begin{bmatrix} \cos(\varphi_k) & -\sin(\varphi_k) \\ \sin(\varphi_k) & \cos(\varphi_k) \end{bmatrix}$ 。

1.2.4 状态转移规律

状态转移规律是指在当前时刻下的状态在采取行动后到下一时刻的变化规律。对于定义的 3 个子系统,分别对其状态转移规律进行定义。

1) 无人机的状态转移规律定义为:

$$x_{k+1} = \Psi(x_k, A_k) \quad (5)$$

对函数 Ψ 定义的过程,就是设计行动值对无人机状态进行控制的方法的过程,即无人机在 k 时刻基于状态 \mathbf{x}_k 采取行动值 $\mathbf{A}_k = (a_k, \phi_k)$, 在 $k+1$ 时刻到达状态 \mathbf{x}_{k+1} 的控制逻辑,本文采取的映射关系如下:

$$p_{k+1}^x = p_k^x + v_k T \cos \theta_k \quad (6)$$

$$p_{k+1}^y = p_k^y + v_k T \sin \theta_k \quad (7)$$

$$\theta_{k+1} = \theta_k + \frac{(gT \tan \phi_k)}{v_k} \quad (8)$$

$$v_{k+1} = v_k + a_k T \quad (9)$$

$$v_{k+1} = \max \{v_{\min}, \min \{v_{\max}, v_{k+1}\}\} \quad (10)$$

(6)~(10)式分别求解出下一时刻无人机的位置和速度,式中 g 为重力加速度值, T 为 2 个时刻之间的采样周期。在实际情况中无人机的速度存在上下限,采用(10)式对无人机速度的大小值进行了限制。

2) 目标的状态转移规律定义为:

$$\zeta_{k+1} = f(\zeta_k) + \mathbf{w}_k \quad (11)$$

式中, \mathbf{w}_k 表示独立同分布的噪声序列,在本文中设定为零均值的高斯噪声, f 表示目标的运动模型。

在已有的基于 POMDP 的无人机路径规划问题中,多将目标的运动模型假设为单一的运动模型,这对于研究 POMDP 的求解算法大有帮助。但是在实际情况下地面目标不可能保持单一的运动形式,运动过程中受到地形和态势影响必然会存在各种机动,运动目标的动态模型比较复杂,传统的单一模型的定位跟踪算法不能实现良好的定位跟踪。采用交互式多模型算法(interacting multiple model, IMM)可以较好地解决这个问题^[13-14]。

IMM 的原理是将系统的运动模式映射为模型集,基于每个运动模型构建一个滤波器,所有模型的滤波器并行工作,利用每个滤波器输出的残差信息以及各模型的先验信息,依据假设检验规则,得出每个滤波器所对应的模型为当前时刻系统匹配模型的概率(称为模型概率),系统的状态估计是各模型滤波器估计的概率加权融合。

依据 IMM,状态方程(11)式和量测方程(1)可以改写为(12)式。

$$\begin{aligned} \zeta_{k+1} &= f(\zeta_k, m_k) + \mathbf{w}(k, m_k) \\ \mathbf{z}_k &= h(\mathbf{S}_k, m_k) + \boldsymbol{\eta}(k, m_k) \end{aligned} \quad (12)$$

式中, m_k 是采样时刻 k 的有效模式,设系统模型集 $M = \{m^1, m^2, \dots, m^r\}$,模型的转换过程符合马尔可夫过程。

滤波状态(ξ_k, \mathbf{P}_k)的转移规律按所采取的滤波算法确定。由于系统方程和观测方程为非线性方程,本文使用无迹卡尔曼滤波(UKF)算法,具体 IMM-UKF 滤波算法在 1.2.6 节介绍信念状态时一并介绍。

1.2.5 代价函数

代价函数表示在某个状态下采取某一行动所付出的代价值。本文采用无人机在当前状态下采取行动后,目标状态和滤波状态之间的均方误差值作为代价函数,代价函数可以表述为:

$$C(\zeta_k, \mathbf{A}_k) = E_{\eta_{k+1}, \mathbf{w}_k} [\|\zeta_{k+1} - \xi_{k+1}\|^2 | \zeta_k, \mathbf{A}_k] \quad (13)$$

1.2.6 信念状态和 IMM-UKF 算法

由于状态的不完全可观测, POMDP 模型引入了信念状态。信念状态是各个状态的后验概率分布,信念状态的更新是基于历史观测值和行动值通过贝叶斯法则计算完成的。本文中,在 k 时刻的信念状态可以表述为 $\mathbf{b}_k = (b_k^x, b_k^\xi, b_k^P)$ 。由于无人机的状态和滤波状态完全可观测,因此 $b_k^x = \delta(x - x_k)$, $b_k^\xi = \delta(\xi - \xi_k)$, $b_k^P = \delta(P - P_k)$ 。目标的信念状态 b_k^ξ 是目标状态的后验概率分布,通过无迹卡尔曼滤波(UKF)算法求解。

UKF 和标准卡尔曼滤波都属于线性最小方差估计,算法都基于模型。标准卡尔曼滤波确定最佳增益阵时,使用了观测量的先验信息和一步预测均方误差阵,并基于系统和观测均为线性的假设。UKF 根据估计量和观测量的协方差阵来确定最佳增益阵,协方差阵根据复现的一倍 σ 样本点计算,所以 UKF 在计算最佳增益阵的过程中未对系统方程和观测方程提出任何附加条件,算法既适用于线性对象,也适用于非线性对象。

IMM 算法的一个循环过程包括:模型交互作用、滤波、模型概率更新和估计混合。下面给出目标信念状态 b_k^ξ 基于 IMM-UKF 的更新过程。

1) 模型交互

假设系统共有 r 个模型,在 k 时刻通过 $k-1$ 时刻的状态估计混合每个模型在 $k-1$ 时刻的输出,以实现模型间的交互作用。用 $m_{k-1}^i (i=1, 2, \dots, r)$ 表示 $k-1$ 时刻各子模型, $\boldsymbol{\mu}_{k-1}^i (i=1, 2, \dots, r)$ 表示模型 i 在 $k-1$ 时刻的概率, $\hat{\zeta}_{k-1}^i (i=1, 2, \dots, r)$ 表示 $k-1$ 时刻模型 i 的状态估计,相应的协方差矩阵是 $\mathbf{P}_{k-1}^i (i=1, 2, \dots, r)$ 。各模型间的转移概率 $p_{ij} = P\{m_k^j | m_{k-1}^i\}$,

各模型滤波器的混合初始输入为:

$$\begin{aligned} \hat{\boldsymbol{\zeta}}_{k-1}^{0j} &= \sum_{i=1}^r \hat{\boldsymbol{\zeta}}_{k-1}^i \mu_{k-1}^{ij} \\ \mathbf{P}_{k-1}^{0j} &= \sum_{i=1}^r \mu_{k-1}^{ij} [\mathbf{P}_{k-1}^i + (\hat{\boldsymbol{\zeta}}_{k-1}^i - \hat{\boldsymbol{\zeta}}_{k-1}^{0j}) \cdot \\ &\quad (\hat{\boldsymbol{\zeta}}_{k-1}^i - \hat{\boldsymbol{\zeta}}_{k-1}^{0j})^T] \end{aligned} \quad (14)$$

式中, $\mu_{k-1}^{ij} = \frac{1}{\bar{c}_j} p_{ij} \mu_{k-1}^i (i = 1, 2, \dots, r)$ 为模型的混合概

率。 \bar{c}_j 则为模型 m_k^j 的预测概率, $\bar{c}_j = \sum_{i=1}^r p_{ij} \mu_{k-1}^i$ 。

2) 模型条件滤波

各滤波器以混合初始输入进行基于各自模型 m_k^j 的无迹卡尔曼滤波 (Unscented Kalman Filter, UKF) 计算, 得到相应的状态估计 $\hat{\boldsymbol{\zeta}}_k^j$ 和协方差矩阵 \mathbf{P}_k^j 。

UKF 的算法步骤如下:

① 计算 $k - 1$ 时刻的 $2n + 1$ 个 σ 样本点, n 为 $\boldsymbol{\zeta}$ 的维数, 本文中 $n = 4$ 。

$$\begin{cases} \tilde{\boldsymbol{\chi}}_{k-1}^{(0)} = \hat{\boldsymbol{\zeta}}_{k-1} \\ \tilde{\boldsymbol{\chi}}_{k-1}^{(i)} = \hat{\boldsymbol{\zeta}}_{k-1} + \gamma(\sqrt{\mathbf{P}_{k-1}})_{(i)} \quad i = 1, 2, \dots, n \\ \tilde{\boldsymbol{\chi}}_{k-1}^{(i)} = \hat{\boldsymbol{\zeta}}_{k-1} - \gamma(\sqrt{\mathbf{P}_{k-1}})_{(i-n)} \quad i = n + 1, n + 2, \dots, 2n \end{cases} \quad (15)$$

② 确定权值

$$\begin{cases} W_0^{(m)} = \frac{\lambda}{n + \lambda} \\ W_0^{(c)} = \frac{\lambda}{n + \lambda} + 1 - \alpha^2 + \beta \\ W_i^{(m)} = W_i^{(c)} = \frac{1}{2(n + \lambda)} \quad i = 1, 2, \dots, 2n \end{cases} \quad (16)$$

上式中, $\gamma = \sqrt{n + \lambda}$, $\lambda = \alpha^2(n + \kappa) - n$, 其中 α 是很小的正数, 可以取 $10^{-4} \leq \alpha \leq 1$, 本文中取 1 ; $\kappa = 3 - n$; β 取值与 $\boldsymbol{\zeta}$ 的分布形式有关, 对于本文中的采用的正态分布, $\beta = 2$ 为最优值。 (15) 式中 $(\sqrt{\mathbf{P}_{k-1}})_{(i)}$ 表示矩阵 \mathbf{P}_{k-1} 的下三角分解平方根的第 i 列。

③ 计算 k 时刻的一步预测模型值

$$\boldsymbol{\chi}_{k/k-1}^{*(i)} = f(\tilde{\boldsymbol{\chi}}_{k-1}^{(i)}) \quad i = 0, 1, 2, \dots, 2n \quad (17)$$

$$\hat{\boldsymbol{\zeta}}_{k/k-1} = \sum_{i=0}^{2n} W_i^{(m)} \boldsymbol{\chi}_{k/k-1}^{*(i)} \quad (18)$$

$$\mathbf{P}_{k/k-1} = \sum_{i=0}^{2n} W_i^{(c)} [\boldsymbol{\chi}_{k/k-1}^{*(i)} - \hat{\boldsymbol{\zeta}}_{k/k-1}] \cdot$$

$$[\boldsymbol{\chi}_{k/k-1}^{*(i)} - \hat{\boldsymbol{\zeta}}_{k/k-1}]^T + \mathbf{Q}_{k-1} \quad (19)$$

(19) 式中 \mathbf{Q}_k 为 (12) 式中 \mathbf{w}_k 的方差阵。

④ 计算 k 时刻的一步预测样本点

$$\begin{cases} \boldsymbol{\chi}_{k/k-1}^{(0)} = \hat{\boldsymbol{\zeta}}_{k/k-1} \\ \boldsymbol{\chi}_{k/k-1}^{(i)} = \hat{\boldsymbol{\zeta}}_{k/k-1} + \gamma(\sqrt{\mathbf{P}_{k/k-1}})_{(i)} \quad i = 1, 2, \dots, n \\ \boldsymbol{\chi}_{k/k-1}^{(i)} = \hat{\boldsymbol{\zeta}}_{k/k-1} - \gamma(\sqrt{\mathbf{P}_{k/k-1}})_{(i-n)} \\ \quad i = n + 1, n + 2, \dots, 2n \end{cases} \quad (20)$$

⑤ 更新量测

$$\mathbf{Z}_{k/k-1}^{(i)} = h(\boldsymbol{\chi}_{k/k-1}^{(i)}) \quad i = 0, 1, 2, \dots, 2n \quad (21)$$

$$\hat{\mathbf{Z}}_{k/k-1} = \sum_{i=0}^{2n} W_i^{(m)} \mathbf{Z}_{k/k-1}^{(i)} \quad (22)$$

$$\mathbf{P}_{(\zeta Z)k/k-1} = \sum_{i=0}^{2n} W_i^{(c)} [\boldsymbol{\chi}_{k/k-1}^{(i)} - \hat{\boldsymbol{\zeta}}_{k/k-1}] [\mathbf{Z}_{k/k-1}^{(i)} - \hat{\mathbf{Z}}_{k/k-1}]^T \quad (23)$$

$$\begin{aligned} \mathbf{P}_{(ZZ)k/k-1} &= \sum_{i=0}^{2n} W_i^{(c)} [\mathbf{Z}_{k/k-1}^{(i)} - \hat{\mathbf{Z}}_{k/k-1}] \cdot \\ &\quad [\mathbf{Z}_{k/k-1}^{(i)} - \hat{\mathbf{Z}}_{k/k-1}]^T + \mathbf{R}_k \end{aligned} \quad (24)$$

⑥ 滤波更新

增益矩阵为:

$$\mathbf{K}_k = \mathbf{P}_{(\zeta Z)k/k-1} \mathbf{P}_{(ZZ)k/k-1}^{-1} \quad (25)$$

滤波值为:

$$\hat{\boldsymbol{\zeta}}_k = \hat{\boldsymbol{\zeta}}_{k/k-1} + \mathbf{K}_k [\mathbf{z}_k - \hat{\mathbf{Z}}_{k/k-1}] \quad (26)$$

$$\mathbf{P}_k = \mathbf{P}_{k/k-1} - \mathbf{K}_k \mathbf{P}_{(ZZ)k/k-1} \mathbf{K}_k^T \quad (27)$$

3) 模型概率更新

通过滤波后, 模型概率的计算是假设检验的过程, 一般采用贝叶斯假设检验方法, 同时检验滤波器组中各个滤波器的残差。由卡尔曼滤波理论可知如果滤波器模型与实际模型匹配, 则滤波残差为零均值方差为 \mathbf{S}_k^j 的高斯白噪声, 因此 k 时刻模型 m^j 为匹配模型的似然函数 A_k^j 为

$$\begin{aligned} A_k^j &= N(\boldsymbol{\epsilon}_k^j; \mathbf{0}, \mathbf{S}_k^j) = \frac{1}{\sqrt{|\mathbf{S}_k^j|} \cdot 2\pi} \exp \cdot \\ &\quad \left[-\frac{1}{2} (\boldsymbol{\epsilon}_k^j)^T \cdot (\mathbf{S}_k^j)^{-1} \cdot \boldsymbol{\epsilon}_k^j \right] \end{aligned} \quad (28)$$

式中, $\boldsymbol{\epsilon}_k^j$ 是第 j 个滤波器的残差估值, $\mathbf{S}_k^j = \mathbf{E}[\boldsymbol{\epsilon}_k^j \cdot (\boldsymbol{\epsilon}_k^j)^T] = \mathbf{P}_{(zz)}^j$, 其中, $\boldsymbol{\epsilon}_k^j = \mathbf{z}_k^j - \hat{\mathbf{Z}}_{k/k-1}^j$, $\mathbf{P}_{(ZZ)}^j$ 按照 (24) 式进行计算。

对于各个模型 $j = 1, 2, \dots, r$, 计算模型概率

$$\mu_k^j = p(m_k^j | \mathbf{z}_k) = \frac{1}{c} A_k^j \bar{c}_j \quad (29)$$

式中, $c = \sum_{i=1}^r A_k^i \bar{c}_i$ 。

4) 估计融合

计算 k 时刻的总体估计和误差协方差矩阵, 分别为:

$$\hat{\xi}_k = \sum_{i=1}^r \hat{\xi}_k^i \mu_k^i$$

$$P_k = \sum_{i=1}^r \mu_k^i [P_k^i + (\hat{\xi}_k - \hat{\xi}_k^i) \cdot (\hat{\xi}_k - \hat{\xi}_k^i)^T] \quad (30)$$

POMDP 模型的滤波状态根据融合的估计值得以更新, 即 $\xi_k = \hat{\xi}_k$ 。继而, 信念状态得到更新 $b_k^\xi(\xi) = N(\xi - \xi_k, P_k)$ 。

1.3 路径规划 POMDP 模型

在无人机跟踪目标的路径规划 POMDP 模型中, 目的是求解出一系列的最优行动值, 让未来行动代价的累加值能够最小。由于无人机飞行的环境变化很快, 对于较长时间后未来状态的预测无实际意义, 因此只考虑未来 H 时限内的状态预测和行动策略求解, 按照代价函数的定义, H 时限内的行动代价的累加值可以表述为:

$$J_H = E \left[\sum_{k=0}^{H-1} C(\xi_k, A_k) \right] \quad (31)$$

由于状态 ξ_k 不完全可观测, 只能求解得到信念状态分布, 所以(31)式可以改写为:

$$J_H = E \left[\sum_{k=0}^{H-1} C(b_k, A_k) \right] \quad (32)$$

式中

$$C(b_k, A_k) = \int C(\xi, A_k) b_k(\xi) d\xi \quad (33)$$

根据贝尔曼公式, 在当前信念状态 b_0 下, 最优目标函数可以写为

$$J_H^* = \min_A \{ C(b_0, A) + E[J_{H-1}^*(b_1) | b_0, A] \} \quad (34)$$

式中, b_1 是下一时刻的信念状态, J_{H-1}^* 是 $H-1$ 时限内累积行动代价的最优值, $k = 1, 2, \dots, H-1$, $E[\cdot | b_0, A]$ 表示在 $k=0$ 时刻, 信念状态为 b_0 时, 执行行动 A 后的条件期望。定义当前信念状态为 b_0 时, 执行行动 A 的 Q 值为

$$Q_H = C(b_0, A) + E[J_{H-1}^*(b_1) | b_0, A] \quad (35)$$

根据贝尔曼公式, 在 $k=0$ 时刻的最优行动策略可以表述为

$$\pi_0^*(b_0) = \operatorname{argmin}_A Q_H(b_0, A) \quad (36)$$

在 k 时刻的最优策略可以表示为 $\pi_k^*(b_k) = \operatorname{argmin}_A Q_{H-k}(b_k, A)$ 。无人机跟踪目标的路径规划,

就是不断利用(36)式基于当前的信念状态来求取最优行动策略的过程。

2 路径规划模型求解

POMDP 模型由于状态的随机性和规划的非短视性, 导致了对其精确求解的困难。在线规划中, 计算(35)式中后一部分远期的代价值 J_{H-1}^* 主要采用近似计算的方法。主要的近似方法有启发式远期代价期望法、参数近似法、策略 rollout 法等等。本文根据无人机跟踪目标的路径规划问题的特点和要求, 使用了名义信念状态优化(NBO)算法来求解。

2.1 算法的近似与假设

在阐述算法之前, 对所做的近似和假设说明。首先之前介绍信念状态更新的部分就设定跟踪模型正确且服从高斯分布, 则目标的信念状态可以表示为 $b_k^\xi(\xi) = N(\xi - \xi_k, P_k)$, 且可以通过 IMM-UKF 算法更新。其次, 假设数据关联是正确的, 这样可以将代价函数^[10]改写为

$$C(b_k, A_k) = \int E_{\eta_{k+1}, w_k} [\| \xi_{k+1} - \xi_{k+1} \|^2 | x_k, \xi, \xi_k, A_k] b_k^\xi(\xi) d\xi = \operatorname{Tr} P_{k+1} \quad (37)$$

2.2 名义信念状态优化(NBO)算法

基于上述假设, NBO 方法的核心可以被概括为如下公式:

$$J^*(b) \approx \min \sum_k C(\hat{b}_k, A_k) \quad (38)$$

式中, $(\hat{b}_k)_k$ 表示一个名义的信念状态序列。NBO 算法的核心是用概率分布中的一个采样点来表示对概率分布的期望值, 其目的是提高计算效率。在不能实时通过随机噪声中多个采样点来计算期望值时, 只能选取一些“名义”上的序列, 比如最大似然点或者均值^[10]。NBO 算法的主要思路就是通过选取的名义信念状态序列 $(\hat{b}_k)_k$ 来求取行动序列 $(A_k)_k$, 从而完成无人机状态的控制和航路规划。NBO 算法的流程如下所述。

①设系统噪声和观测噪声的均值均为 0, 则(1)式和(11)式可以写为

$$\begin{aligned} \xi_{k+1} &= f(\xi_k) + w_k, w_k \sim N(0, Q_k), \\ z_k &= h(S_k) + \eta_k, \eta_k \sim N(0, R_k) \end{aligned} \quad (39)$$

②定义名义信念状态序列 $(\hat{b}_1, \dots, \hat{b}_{H-1})$, 即

$$\begin{aligned} b_{k+1} &= \Phi(b_k, A_k, \eta_{k+1}, w_k) \Rightarrow \hat{b}_{k+1} = \Phi(\hat{b}_k, A_k, 0, 0), \\ \hat{b}_0 &= b_0 \end{aligned} \quad (40)$$

③用名义信念状态序列来代替信念状态的期望值,即

$$J_H(b_0) = E_{b_1, \dots, b_H} \left[\sum_{k=1}^H C(b_k, \mathbf{A}_k) \right] \Rightarrow$$

$$J_H(b_0) \approx \sum_{k=1}^H C(\hat{b}_k, \mathbf{A}_k) \quad (41)$$

④根据(41)式求最优行动序列 $(\mathbf{A}_0, \dots, \mathbf{A}_{H-1})$ 。以 $\min \sum_{k=1}^H C(\hat{b}_k, \mathbf{A}_k)$ 为目标函数,以 \mathbf{A}_k 的取值范围为约束条件,求解过程可以转化为一个标准的最优化问题,这里采用 matlab 中的 fmincon 函数来实现。

基于上述通用步骤,结合无人机目标跟踪路径规划的特点,设计基于 IMM-UKF 的无人机行动序列 NBO 求解算法如下。

首先,根据 $b_k^i(\zeta) = N(\zeta - \hat{\xi}_k, \hat{P}_k)$, 目标状态 ζ_k 的信念状态 b_k^i 可以由滤波状态 $(\hat{\xi}_k, \hat{P}_k)$ 来确定,其中计算 $(\hat{\xi}_k, \hat{P}_k)$ 的过程通过(14) ~ (30) 式完成。因此名义信念状态 \hat{b}_k^i 可以根据滤波结果 $(\hat{\xi}_k, \hat{P}_k)$ 来确定。

$$\hat{b}_k^i(\zeta) = N(\zeta - \hat{\xi}_k, \hat{P}_k) \quad (42)$$

其次,根据(29)式计算出的当前时刻 IMM 中各模型的概率 μ_k^i , 选取其中概率最大的模型为当前的匹配模型 F_k 。

$$F_k = \max_{\mu_k^i} \{ m_i(\mu_k^i) \} \quad (43)$$

由于预测时限 H 不长,所以认为在预测时限内目标均按照当前的匹配模型 F_k 运动,因此基于匹配模型 F_k 和 $\hat{\xi}_k$, 通过(44) 式可以递推出预测时限内各个时刻的 $\hat{\xi}$ 。

$$\hat{\xi}_{k+1} = F_k \hat{\xi}_k \quad (44)$$

在预测过程中,没有量测信息输入,因此协方差矩阵的名义信念状态 \hat{P}_k 可以通过克拉美罗下界来进行估计。

$$\hat{P}_{k+1} = [(F_k \hat{P}_k F_k^T + Q_k) + H_{k+1}^T [R_{k+1}(x_k, \hat{\xi}_k)]^{-1} H_{k+1}]^{-1} \quad (45)$$

式中, H_k 是量测函数 $h(S_k)$ 对 S_k 求导得到的雅各比矩阵。

最后,根据代价函数的近似计算公式(37),将预测时限内递推得到的名义信念状态序列 \hat{P}_k 带入(41) 式,获得预测时限内的行动代价累加值,并以此为目标函数通过 fmincon 函数求得最优行动序列 $(\mathbf{A}_0, \dots, \mathbf{A}_{H-1})$, 取 \mathbf{A}_0 为当前时刻的行动策略并执

行,则无人机状态改变,根据下一时刻的状态再执行 IMM-UKF 滤波,并重复上述步骤进行策略求解,不断往复,利用各个时刻求取的行动策略调整无人机状态,完成跟踪目标的路径规划。

3 仿真分析

本文在 MTALAB 环境下编写仿真程序,以无人机对单目标跟踪为背景进行仿真,基于跟踪的位置误差均值和无人机飞行轨迹对算法进行分析。目标的运动模型集合包括匀速直线运动,匀速左转弯运动和匀速右转弯 3 种运动。

单次仿真时间 300 s,目标从(0,0)点开始沿水平轴方向匀速运动 60 s,之后进行匀速左转弯运动 60 s,再匀速直线运动 60 s,之后再完成匀速右转弯运动 60 s,最后再完成匀速直线运动 60 s。

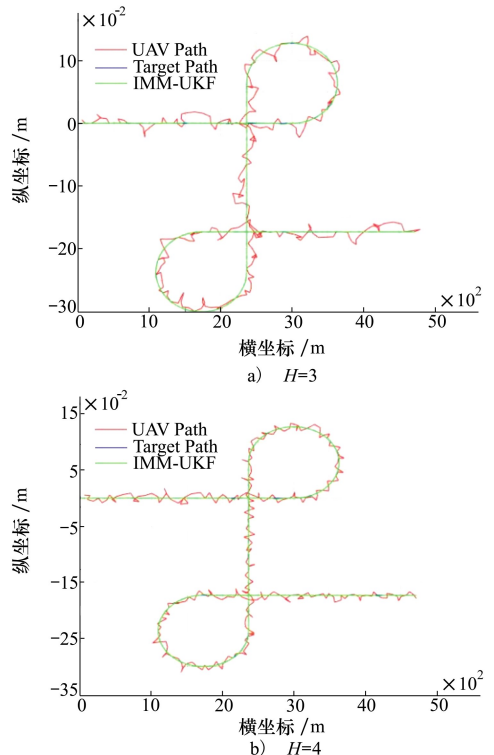


图 1 目标状态估计和无人机路径规划

对基于 IMM 方法所建立的 POMDP 模型,分别在预测时限 $H = 3$ 和 $H = 4$ 的情况下各仿真 30 次,仿真结果如图 1、图 2 和图 3 所示,从图 1 中可以看出无论是预测时限 $H = 3$ 和 $H = 4$,无人机在整个仿真过程中均能够跟踪目标,并围绕目标飞行。对比图 1a) 和图 1b) 可以发现,较之于 $H = 3$,在 $H = 4$ 时,行

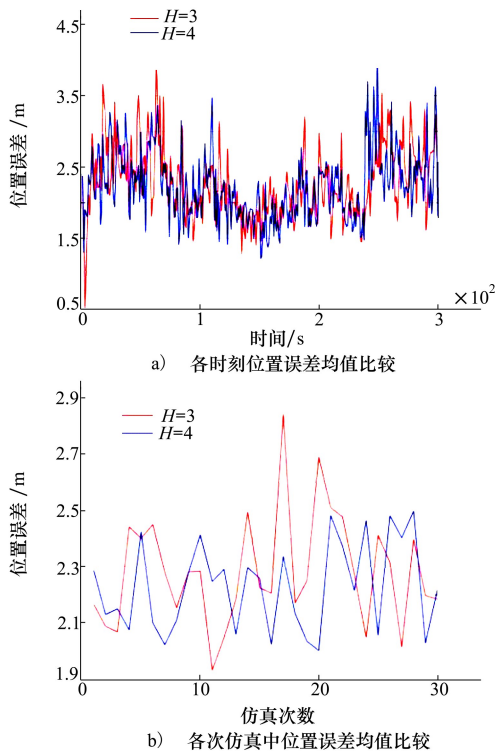


图 2 基于 IMM 的位置误差均值

时位置误差均值为 2.228 9, 仅略小于 $H=3$ 时的 2.282 2, 这在说明了 POMDP 模型决策远期性的同时也要求在设计预测时限时需要综合考虑计算开销和获得效果之间的平衡点。

表 1 位置误差均值

方法	IMM		CV
	$H=3$	$H=4$	
位置误差均值	2.282 2	2.228 9	3.423 7

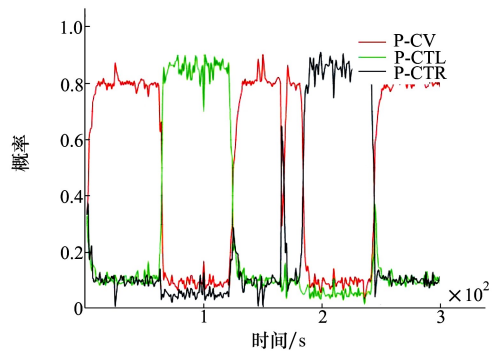


图 3 IMM 中各个模型的概率值

动策略所产生的无人机飞行轨迹更加平稳, 围绕目标飞行的波动较小, 这充分体现出 POMDP 的远期代价计算特点, 在当前状态下对目标未来状态进行预测能够让无人机对目标的行为有了预判, 进而再计算出的行动策略能根据预测结果得到最优值, 预测时限越长, 策略越好。

图 2a) 反映了各次仿真中每一仿真时刻的位置误差平均值, 从图中可以看出, 不论 $H=3$ 或 $H=4$, 整个仿真过程中滤波结果和目标真实位置之间的位置误差波动不是很大, 在 60 s、120 s、180 s、240 s 等目标运动规律发生变化的时刻, 位置误差均值并没有显著增加, 说明采用 IMM 方法, 无人机能够判断出目标的运动模型切变, 在 POMDP 框架下对目标未来的远期状态进行预判, 利用 NBO 方法求解的行动策略能够让无人机识别目标的运动规律而保持对目标的近距离跟踪。图 3 为各个时刻交互多模型集中各个模型的概率值, 其中, P-CV 代表匀速直线运动的概率值, P-CTL 代表匀速左转弯运动的概率值, P-CTR 代表匀速右转弯运动的概率值。图 2b) 反映了各次仿真整个过程的位置误差平均值, 从整体上可以看出可 $H=4$ 时的位置误差均值略小于 $H=3$ 时的值, 表 1 中的确切数据也反映了这一现象。 $H=4$

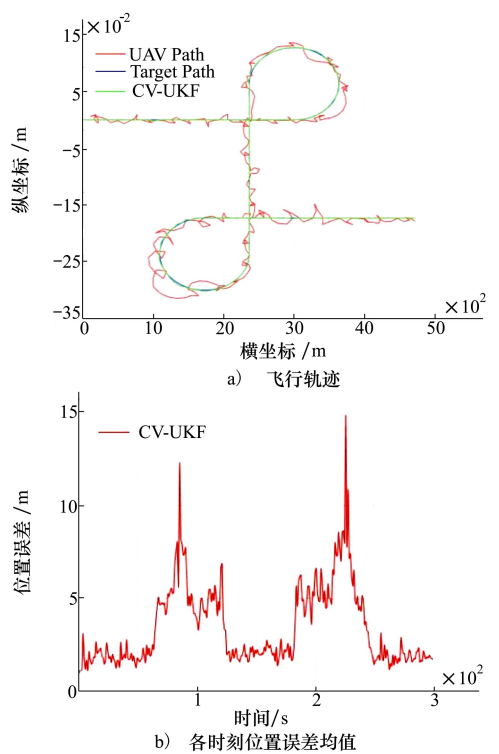


图 4 基于 CV 模型的飞行轨迹和位置误差均值

为了对比 IMM 方法, 在同样条件下进行了仅采

用匀速直线运动的一种运动模型的仿真 30 次。图 4 显示的是基于匀速直线运动 (CV) 模型的飞行轨迹和位置误差均值。UKF 滤波和 NBO 算法预测中对目标的状态估计仅使用匀速直线运动这一种模型。从图 4a) 中可看出,在目标进行从直行进入转弯运动的过程中,无人机的飞行明显偏离了目标的运动轨迹,同时在图 4b) 中可以看出在目标进行左转弯运动和右转弯运动时,位置误差明显增大。从表 1 中也可以看出在整个仿真过程中仅采用 CV 模型的位置误差均值 3.423 7 远大于采用 IMM 算法的结果。在真实情况中,如果位置误差超过无人机传感器的有效探测距离,很有可能导致跟踪失败,为了验证,在相同情况下,将(3)式中的传感器测距不确定性 m 增大 2% 再进行仿真,结果如图 5 所示。

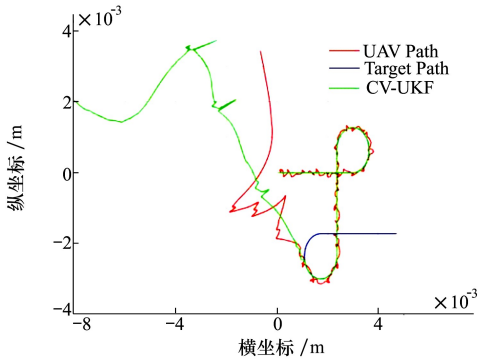


图 5 CV 模型下增大大量测误差后仿真结果

在目标进行右转弯机动时,无人机丢失目标,跟踪失败。

对比图 2a) 和图 4b),说明基于 IMM 的无人机路径规划 POMDP 模型能够有效提高跟踪精度,使得规划的无人机路径更加贴合目标的运动,避免出现丢失目标的情况。

综上所述,在 POMDP 模型的状态转移规律中使用交互多模型机制能够比单一模型有效提高跟踪精度,能够保证无人机对机动目标跟踪的路径规划求解。

4 结 论

本文从无人机跟踪地面机动目标需求出发,针对无人机跟踪目标任务中的路径规划问题,引入 POMDP 理论,构建了基于 POMDP 的无人机在线路径规划决策模型。根据目标运动规律的多样性,使用了交互多模型 (IMM) 的方法描述状态转移规律并基于 IMM-UKF 算法来更新信念空间。为实现模型在线求解,引入了 NBO 算法,求解的行动策略能够实时完成无人机的运动调整和路径规划。仿真结果验证了本文所建立的无人机路径规划模型的有效性。

参考文献:

- [1] Wang Z, Zheng M, Guo J, et al. Uncertain UAV ISR Mission Planning Problem with Multiple Correlated Objectives[J]. Journal of Intelligent & Fuzzy Systems, 2017, 32(1): 321-335
- [2] 陈宗基,张汝麟,张平,等. 飞行器控制面临的机遇与挑战[J]. 自动化学报,2013,39(6):703-710
Chen Zongji, Zhang Rulin, Zhang Ping, et al. Flight Control: Challenges and Opportunities[J]. Acta Automatica Sinica, 2013, 39(6):703-710 (in Chinese)
- [3] 张耀中,张建东,史国庆. 无人机警戒雷达仿真系统的建模与仿真[J]. 火力与指挥控制,2016,41(2):93-96
Zhang Yaozhong, Zhang Jiandong, Shi Guoqing. Research on Modeling and Simulation of UAV Warning Radar[J]. Fire Control & Command Control, 2016,41(2): 93-96 (in Chinese)
- [4] Ragi Shankarachary, Chong Edwin K P. UAV Path Planning in a Dynamic Environment via Partially Observable Markov Decision Process [J]. IEEE Trans on Aerospace And Electronic Systems, 2013,49(4):2397-2412
- [5] Ponzoni Carvalho Chanel, Caroline, Teichteil Königsbuch. POMDP-Based Online Target Detection and Recognition for Autonomous UAVs[C]//The 20th European Conference on Artificial Intelligence (ECAI), 2012
- [6] Ragi Shankarachary, Chong Edwin K P. Decentralized Control of Unmanned Aerial Vehicles for Multitarget Tracking[C]//2013 International Conference on Unmanned Aircraft Systems, 2013: 260-268
- [7] Cork L, Walker R. Sensor Fault Detection for UAVs Using a Nonlinear Dynamic Model and the IMM-UKF Algorithm[C]//Information, Decision and Control, 2007: 230-235

- [8] 藏荣春,崔平远,崔祐涛,等. 基于 IMM-UKF 的组合导航算法[J]. 控制理论与应用,2007,24(4):634-638
Zang Rongcui, Cui Pingyuan, Cui Hutao, et al. Integrated Navigation Algorithm Based on IMM-UKF[J]. Control Theory & Applications, 2007, 24(4): 634-638 (in Chinese)
- [9] 万开方,高晓光,李波,等. 基于部分可观察马尔可夫决策过程的多被动传感器组网协同反隐身探测任务规划[J]. 兵工学报,2015,36(4):731-743
Wan Kaifang, Gao Xiaoguang, Li Bo, et al. Mission Planning of Passive Networked Sensors for Cooperative Anti-stealth Detection Based on POMDP[J]. Acta Armamentarii, 2015,36(4): 731-743 (in Chinese)
- [10] Miller Scott A, Harris Zachary A, Chong Edwin K P. A POMDP Framework for Coordinated Guidance of Autonomous UAVs for Multitarget Tracking [J]. EURASIP Journal on Advances in Signal Processing, 2009, 2009(1): 1-17
- [11] Gong S L, Wang B F, Honglan W U, et al. Tracking of Moving Targets on Airport Surface Based on IMM Algorithm[J]. Systems Engineering & Electronics, 2011, 33(10):2322-2326
- [12] Wang D, Lv H, Wu J. In-Flight Initial Alignment for Small UAV MEMS-Based Navigation via Adaptive Unscented Kalman Filtering Approach[J]. Aerospace Science & Technology, 2017, 61:73-84
- [13] Cui N, Hong L, Layne J R. A Comparison of Nonlinear Filtering Approaches with an Application to Ground Target Tracking[J]. Signal Processing, 2005, 85(8):1469-1492
- [14] Ragi Shankarachary, Chong Edwin K P. Decentralized Guidance Control of UAVs with Explicit Optimization of Communication [J]. Journal of Intelligent & Robotic Systems, 2014,73(1):811-822

Decision Modeling of UAV On-Line Path Planning Based on IMM

Yang Qiming¹, Xu Jiancheng¹, Tian Haibao², Wu Yong¹

(1.School Electronics and Information, Northwestern Polytechnical University, Xi'an 710072, China;)
(2.Air Force Military Representative Office in Jiangxi Area, Nanchang 330024, China)

Abstract: In order to enhance the capability of tracking targets autonomously of UAV, a model for UAV on-line path planning is established based on the theoretical framework of partially observable markov decision process (POMDP). The elements of the POMDP model are analyzed and described. According to the diversity of the target motion in real world, the law of state transition in POMDP model is described by the method of Interactive Multiple Model(IMM) To adapt to the target maneuvering changes. The action strategy of the UAV is calculated through nominal belief-state optimization(NBO) algorithm which is designed to search optimal action policy to minimize the cumulative cost of action. The generated action strategy controls the UAV flight. The simulation results show that the established POMDP model can achieve autonomous planning for UAV route, and it can control the UAV to effectively track target. The planning path is more reasonable and efficient than the result of using single state transition law.

Keywords: partially observable markov decision process; interactive multiple model; UAV; target tracking; MATLAB; cost reduction