

基于部分可观马氏决策过程的多平台主被动传感器调度

张子宁^{1,2}, 单甘霖¹, 段修生¹

(1. 军械工程学院电子与光学工程系, 河北石家庄 050003; 2. 北京航天飞控中心, 北京 100094)

摘要: 为了使有限时域内的跟踪精度和辐射风险达到最佳平衡, 本文研究了多传感器平台在协同跟踪目标时的主被动传感器调度问题. 将该问题建立成基于部分可观马氏决策过程的数学模型以同步实现目标跟踪和辐射控制. 在先见优化思想的基础上, 借助由无迹采样近似得到的精度收益及由隐马氏模型滤波器推导出的辐射代价将调度问题转化成决策树问题, 并采用分枝定界方法求解. 仿真结果表明了该方法的有效性.

关键词: 传感器调度; 部分可观马氏决策过程; 先见优化; 无迹采样; 分枝定界

中图分类号: TP212.9 **文献标识码:** A **文章编号:** 0372-2112 (2014)10-2104-06

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.3969/j.issn.0372-2112.2014.10.037

POMDP-Based Scheduling of Active/Passive Sensors in Multi-Platform

ZHANG Zi-ning^{1,2}, SHAN Gan-lin¹, DUAN Xiu-sheng¹

(1. Department of Electronic and Optical Engineering, Ordnance Engineering College, Shijiazhuang, Hebei 050003, China;

2. Beijing Aerospace Control Center, Beijing 100094, China)

Abstract: To make an optimal trade-off between the tracking accuracy and the radiation risk in a period of time, this paper studies the scheduling problem of selecting the active/passive sensors in the multi-platform for target tracking. The problem is formulated as a partially observable Markov decision process (POMDP), which can take both target tracking and emission control into account. Based on the foresight optimization, the approximate accuracy reward and the radiation cost, which are derived from the unscented transformation sampling and hidden Markov model (HMM) filter respectively, transform our problem into a tree search problem, and the branch and bound method is used for problem solution. The simulation results demonstrate the effectiveness of our approach.

Key words: sensor scheduling; partially observable Markov decision process; foresight optimization; unscented transformation sampling; branch and bound

1 引言

传感器调度问题是指通过实时的传感器分配策略使得特定的性能度量指标达到最优. 传统的传感器调度方法多为短时 (myopic) 调度^[1,2], 通常以单个调度动作所产生的一步收益作为最优准则来获得最优的短期性能. 相对于短期调度而言, 长期 (non-myopic) 调度更注重长期性能, 其度量依据是一段时期内调度动作序列所产生的累积收益. 文献[3]通过将一段时期内累积互信息最大化进而确定最优调度动作序列; 文献[4]将长时信息增量模型描述为决策树并采用分支剔除法寻优, 解决了低轨星座传感器的长期调度问题; 文献[5,6]利用部分可观马氏决策过程来权衡有限时域内的累积跟踪精度和能量消耗, 并使用粒子滤波结合 Rollout 的方法进行求解.

本文以分散部署的多个传感器平台跟踪同一目标作为研究背景, 考虑主被动传感器在测量性能上及能量辐射上的差异, 研究如何实时地选择平台上的主被动传感器以折中权衡一定时域内的跟踪精度和辐射风险. 该调度问题被构建成一个部分可观马氏决策过程 (Partially Observable Markov Decision Process, POMDP), 通过转化将问题表示成决策树形式, 借助分枝定界法求解.

2 问题的数学描述

依据 POMDP 的理论框架^[7], 该问题的数学表述具体如下.

2.1 调度动作

假定有 N 个传感器平台共同跟踪一个目标, 每个平台上配备主动传感器和被动传感器各一部, 每个时刻各传感器平台只能启动一部传感器. 令 k 时刻所决

策的调度动作由 N 维列向量 $\mathbf{u}_k = [u_k^1, \dots, u_k^N]^T$ 来表示,其中 $u_k^i = 1$ 或 $0 (i = 1, \dots, N)$ 表示平台 i 将启动主动或被动传感器观测 $k+1$ 时刻的目标状态。

2.2 基本状态及状态转移

k 时刻的基本状态 $\mathbf{S}_k = [\mathbf{X}_k \ \hat{\mathbf{X}}_k \ \mathbf{E}_k]^T$ 由目标的运动状态 \mathbf{X}_k , 目标运动状态估计值 $\hat{\mathbf{X}}_k$ 及 N 个平台的辐射度影响状态 $\mathbf{E}_k = [E_k^1 \ \dots \ E_k^N]^T$ 组成,其中平台 i 在 k 时刻的 ELI 状态 E_k^i 表示截止到 k 时刻为止敌方累积截获平台 i 的辐射量^[8],其取值被量化至有限离散集合 $\{1, \dots, q_i\}$,该集合中的每个数值代表着相应的 ELI 真值。

状态估计值 $\hat{\mathbf{X}}_k$ 依靠卡尔曼滤波算法进行状态转移,而目标运动状态 \mathbf{X}_k 则依靠目标状态方程转移到下一时刻状态 \mathbf{X}_{k+1} ,即

$$\mathbf{X}_{k+1} = \mathbf{F}_k \mathbf{X}_k + \mathbf{\Gamma}_k \boldsymbol{\omega}_k \quad (1)$$

此时 $\mathbf{F}_k, \mathbf{\Gamma}_k$ 和 $\boldsymbol{\omega}_k$ 分别表示状态转移矩阵,过程噪声增益矩阵以及零均值高斯过程噪声。

考虑到平台 i 的 ELI 状态 E_k^i 会按照状态转移概率 $p(E_{k+1}^i | E_k^i u_k^i)$ 来实现状态转移,对此可用一个受控的转移概率矩阵 $\mathbf{A}_k^i = (A_{mn}^i)_{m,n \in \{1, \dots, q_i\}}$ 进行描述,其中矩阵元素 $A_{mn}^i = p(E_{k+1}^i = n | E_k^i = m u_k^i = 0 \text{ 或 } 1)$ 是用于描述状态转移的条件概率。根据平台 i 的启动情况,可进一步将 \mathbf{A}_k^i 表示为

$$\mathbf{A}_k^i = (\mathbf{A}_i)^{u_k^i} \quad (2)$$

其中 \mathbf{A}_i 是时齐状态转移矩阵。

2.3 观测值

定义总体观测值 $\mathbf{Z}_k = [\mathbf{Z}_{X_k} \ \mathbf{Z}_{e_k}]^T$ 由目标运动状态测量值 \mathbf{Z}_{X_k} 及 N 个平台的瞬时观测威胁度 $\mathbf{Z}_{e_k} = [Z_{e_k}^1 \ \dots \ Z_{e_k}^N]^T$ 组成,其中 $Z_{e_k}^i$ 是平台 i 在 k 时刻的瞬时观测威胁度。

球坐标系下,平台 i 的测量方程为

$$\mathbf{Y}_k^i = [r_k^i \ \theta_k^i \ \varphi_k^i]^T + \mathbf{v}_k^i \quad (3)$$

$$r_k^i = \sqrt{(x_k - sc_i(x))^2 + (y_k - sc_i(y))^2 + (z_k - sc_i(z))^2}$$

$$\theta_k^i = \tan^{-1} \left(\frac{y_k - sc_i(y)}{x_k - sc_i(x)} \right)$$

$$\varphi_k^i = \tan^{-1} \left(\frac{z_k - sc_i(z)}{\sqrt{(x_k - sc_i(x))^2 + (y_k - sc_i(y))^2}} \right)$$

其中,列向量 $[r_k^i \ \theta_k^i \ \varphi_k^i]^T$ 中的各元素分别表示目标相对于平台 i 的距离真值,方位角真值以及高低角真值; $(sc_i(x), sc_i(y), sc_i(z))$ 是平台 i 在直角坐标系中的地理位置坐标; $\mathbf{v}_k^i = u_{k-1}^i \mathbf{v}_k^{i,a} + (1 - u_{k-1}^i) \mathbf{v}_k^{i,p}$ 是平台 i 的零均值高斯测量噪声($\mathbf{v}_k^{i,a}$ 和 $\mathbf{v}_k^{i,p}$ 分别表示平台 i 上主动传感器和被动传感器的高斯测量噪声)。

式(3)经视线坐标系变换^[9]后,得到直角坐标系下的线性测量方程

$$\mathbf{Z}_k^i = \mathbf{H} \cdot \mathbf{X}_k + (\mathbf{m}(\mathbf{Y}_k^i))^T \mathbf{v}_k^i \quad (4)$$

其中, $\mathbf{H} = [\mathbf{I}_3 \ \mathbf{0}_3]$; \mathbf{I}_3 和 $\mathbf{0}_3$ 分别是三维的单位矩阵和零矩阵; $\mathbf{m}(\mathbf{Y}_k^i)$ 表示由球坐标测量值 \mathbf{Y}_k^i 决定的坐标旋转矩阵。由于传感器的角度测量误差通常很小,故式(4)中的噪声 \mathbf{v}_k^i 可近似^[10]为

$$\mathbf{v}_k^i \approx \text{diag}(1, r_k^i \cos \varphi_k^i, r_k^i) \mathbf{v}_k^i \quad (5)$$

其中,符号 $\text{diag}(\)$ 代表对角矩阵。

各平台所获得的目标测量值 \mathbf{Z}_k^i 经数据融合后即可得到目标运动状态观测值 \mathbf{Z}_{X_k} ,即

$$\mathbf{Z}_{X_k} = \sum_{i=1}^N w_k^i \mathbf{Z}_k^i = \mathbf{H} \cdot \mathbf{X}_k + \sum_{i=1}^N w_k^i (\mathbf{m}(\mathbf{Y}_k^i))^T \mathbf{v}_k^i \quad (6)$$

此时,噪声协方差阵为 $\mathbf{R}_k = \sum_{i=1}^N (w_k^i)^2 (\mathbf{m}(\mathbf{Y}_k^i))^T \mathbf{R}_k^i \mathbf{m}(\mathbf{Y}_k^i)$;

融合权重 w_k^i 满足 $\sum_{i=1}^N w_k^i = 1$ 并可由式(7)给出^[11]。

$$w_k^i = (1 - \bar{\sigma}_k^i) \bar{r}_k^i / \sum_{i=1}^N (1 - \bar{\sigma}_k^i) \bar{r}_k^i \quad (7a)$$

$$\bar{\sigma}_k^i = \frac{\sum_{i=1}^k [((2u_k^i - 1)u_k^i - u_k^i + 1) \left(\sum_{j=1}^N \exp(-a(Z_j^i - Z_j^i)^T (Z_j^i - Z_j^i)) \right)^2]}{N^2 \left[(1 - u_k^i)k + (2u_k^i - 1) \sum_{i=1}^k u_i^i \right]} - \left[\frac{1}{N} \sum_{j=1}^N \exp(-a(Z_j^i - Z_j^i)^T (Z_j^i - Z_j^i)) \right]^2 \quad (7b)$$

$$\bar{r}_k^i = \frac{\sum_{i=1}^k \sum_{j=1}^N [(2u_k^i - 1)u_k^i - u_k^i + 1] \exp(-a(Z_j^i - Z_j^i)^T (Z_j^i - Z_j^i))}{N \left[(1 - u_k^i)k + (2u_k^i - 1) \sum_{i=1}^k u_i^i \right]} \quad (7c)$$

式(7)中, a 和 λ 均是可调参数并且 $0 < \lambda \leq 1$; \bar{r}_k^i 和 $\bar{\sigma}_k^i$ 分别表示观测一致性均值和方差。

将观测威胁度 $Z_{e_k}^i$ 的取值量化至有限集合 $\{1, \dots, c_i\}$ 以表示实际情况中不同程度的威胁度增量。由此,介于 $Z_{e_k}^i$ 与 e_k^i 之间的概率关系可依据观测矩阵 $\mathbf{B}_k^i(Z_{e_k}^i = l) = (\mathbf{B}_{mnl}^i)_{m,n \in \{1, \dots, q_i\}, (l = 1, \dots, c_i)}$ 来表示,此时矩阵元素 $B_{mnl}^i = p(Z_{e_k}^i = l | E_{k-1}^i = m \ E_k^i = n)$ 是描述观测结果的条件概率。由平台 i 的启动情况可进一步得到

$$\sum_{l=1}^{c_i} \mathbf{B}_k^i(Z_{e_k}^i = l) = (\mathbf{I}_{q_i \times q_i})^{u_{k-1}^i} \quad (8)$$

式中, $\mathbf{I}_{q_i \times q_i}$ 表示元素全部为 1 的 q_i 维矩阵。

2.4 单步收益与有限时域收益

利用 k 时刻的跟踪误差和 ELI 状态来构成单步收益函数以反映单个动作所产生的收益值,即

$$r(\mathbf{S}_k, \mathbf{u}_k) = \alpha (\mathbf{X}_k - \hat{\mathbf{X}}_k)^T (\mathbf{X}_k - \hat{\mathbf{X}}_k) + (1 - \alpha) \sum_{i=1}^N E_k^i \quad (9)$$

式中, α 是用于调整跟踪精度与辐射风险之间平衡关系的系数.

信息集合 $\boldsymbol{\eta}_k = (\mathbf{Z}_1, \dots, \mathbf{Z}_k, \mathbf{u}_0, \dots, \mathbf{u}_{k-1})$ 记录了从初始时刻到 k 时刻为止所得到的观测值序列和动作序列, 而信息状态 χ_k 是随机变量 \mathbf{X}_k 和 \mathbf{E}_k 在信息集合 $\boldsymbol{\eta}_k$ 已给定的条件下所得到的后验概率分布. 假定调度动作可由信息状态来决定, 即 $\mathbf{u}_k = \pi_k(\chi_k)$, 这里 π_k 表示了 k 时刻信息状态与决策动作之间的映射关系. 因此, 一个映射序列 $\Delta = \{\pi_0, \dots, \pi_{H-1}\}$ 所产生的有限时域收益 J_Δ 可定义为: 时域长度 $H+1$ 上单步收益的累积期望值, 即

$$\begin{aligned} J_\Delta &= \mathbb{E}_{\mathbf{z}_1 \dots \mathbf{z}_H} \left[\sum_{k=0}^H \sum_{\mathbf{E}_k} r(\mathbf{S}_k, \mathbf{u}_k) p(\mathbf{X}_k \mathbf{E}_k | \boldsymbol{\eta}_k) d\mathbf{X}_k \right] \\ &= \alpha \underbrace{\left\{ \mathbb{E}_{\mathbf{z}_1 \dots \mathbf{z}_H} \sum_{k=0}^H \text{tr}(\mathbf{P}_k) \right\}}_{\text{长期精度收益 } J_\Delta^{(1)}} + (1-\alpha) \underbrace{\left\{ \mathbb{E}_{\mathbf{z}_1 \dots \mathbf{z}_H} \sum_{k=0}^H \sum_{i=1}^N \mathbf{D}_i^T \boldsymbol{\chi}_{\mathbf{E}_k^i} \right\}}_{\text{长期辐射代价 } J_\Delta^{(2)}} \end{aligned} \quad (10)$$

其中, 符号 \mathbb{E} 表示在观测值序列作用下的数学期望; $p(\mathbf{X}_k \mathbf{E}_k | \boldsymbol{\eta}_k)$ 是信息状态中的后验概率密度; $\text{tr}(\cdot)$ 表示矩阵的迹; \mathbf{P}_k 是跟踪误差的协方差矩阵; $\mathbf{D}_i = [1 \ \dots \ q_i]^T$ 是列向量; $\boldsymbol{\chi}_{\mathbf{E}_k^i}$ 是后文中所介绍的 ELI 信息状态.

3 问题的求解

由于最优映射序列 $\Delta^* = \arg \min_{\Delta} J_\Delta$ 通常难以求取, 故本文采用先见优化这种近似求解方法, 其思想是将映射序列 Δ 进一步约束为固定的动作序列 $\boldsymbol{\Psi} = \{\mathbf{u}_0, \dots, \mathbf{u}_{H-1}\}$.

3.1 长期精度收益的近似计算

为了估算长期精度收益 $\bar{J}_\Psi^{(1)}$, 本文提出了一种基于无迹变换^[12]的采样方法以达到为卡尔曼滤波中协方差递推提供若干测量值序列的目的.

给定初始时刻的目标运动状态估计值 $\hat{\mathbf{X}}_0$ 及其协方差矩阵 \mathbf{P}_0 , 利用无迹变换的方法采集样本 $\boldsymbol{\varepsilon}_r$ ($r=0, \dots, 2\sigma$), 每个样本 $\boldsymbol{\varepsilon}_r = [\tilde{\mathbf{X}}_0^T \tilde{\boldsymbol{\omega}}_r^T \tilde{\mathbf{v}}_r^T]^T$ 中包含 0 时刻的目标运动状态样本 $\tilde{\mathbf{X}}_0$; 从 0 时刻到 $H-1$ 时刻的目标过程噪声样本序列 $\tilde{\boldsymbol{\omega}}_r = [\tilde{\boldsymbol{\omega}}_0^T \ \dots \ \tilde{\boldsymbol{\omega}}_{H-1}^T]$ 及所有传感器从 1 时刻到 H 时刻的测量噪声样本序列 $\tilde{\mathbf{v}}_r = [\tilde{\mathbf{v}}_1^T \ \dots \ \tilde{\mathbf{v}}_H^T]$, 即

$$\begin{aligned} \boldsymbol{\varepsilon}_0 &= [\hat{\mathbf{X}}_0 \ \underbrace{0 \ \dots \ 00}_{\mathbb{E}(\tilde{\boldsymbol{\omega}}_0)} \ \underbrace{00 \ \dots \ 0}_{\mathbb{E}(\tilde{\mathbf{v}}_0)}]^T, \\ \boldsymbol{\varepsilon}_r &= \begin{cases} \boldsymbol{\varepsilon}_0 + \left(\sqrt{\frac{\sigma}{1-\beta_0} \rho_0} \right)_r, & r=1, \dots, \sigma \\ \boldsymbol{\varepsilon}_0 - \left(\sqrt{\frac{\sigma}{1-\beta_0} \rho_0} \right)_r, & r=\sigma+1, \dots, 2\sigma \end{cases} \quad (11) \\ \rho_0 &= \text{diag}(\mathbf{P}_0, \underbrace{\text{diag}(\mathbf{Q}_0, \dots, \mathbf{Q}_{H-1})}_{\mathbb{E}(\tilde{\boldsymbol{\omega}}_0 \tilde{\boldsymbol{\omega}}_0^T)}, \underbrace{\text{diag}(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_H)}_{\mathbb{E}(\tilde{\mathbf{v}}_0 \tilde{\mathbf{v}}_0^T)}) \end{aligned}$$

其中, σ 为样本 $\boldsymbol{\varepsilon}_r$ 的维数; $(\sqrt{\cdot})_r$ 表示矩阵平方根的

第 r 行; \mathbf{Q}_k 是过程噪声的协方差阵; $\boldsymbol{\mu}_k$ 是由所有传感器的测量噪声协方差阵所组成的块对角阵.

若给定动作序列 $\boldsymbol{\Psi}$, 利用样本 $\boldsymbol{\varepsilon}_r$ 中包含的目标初始运动状态样本和噪声样本序列, 由运动方程和测量方程可生成相应的目标运动状态测量值序列 $\{\tilde{\mathbf{Y}}_1^r, \dots, \tilde{\mathbf{Y}}_H^r\}$, 其中 k 时刻的测量值样本 $\tilde{\mathbf{Y}}_k^r = [\tilde{\mathbf{Y}}_k^{1,r} \ \dots \ \tilde{\mathbf{Y}}_k^{N,r}]^T$ 由 N 个平台所得到的测量值样本共同构成. 借助这 $2\sigma+1$ 个测量值序列, 使用卡尔曼滤波即可随时间变化实现全局误差协方差阵 $\tilde{\mathbf{P}}_k = \text{diag}(\tilde{\mathbf{P}}_k^{r=0}, \dots, \tilde{\mathbf{P}}_k^{r=2\sigma})$ 的递推, 其中 $\tilde{\mathbf{P}}_k^r$ 是由测量值样本 $\tilde{\mathbf{Y}}_k^r$ 更新得到的协方差阵.

$$\begin{aligned} \tilde{\mathbf{P}}_{k+1} &= (\mathbf{I} + \tilde{\mathbf{P}}_{k+1} \boldsymbol{\Upsilon}_{k+1})^{-1} \tilde{\mathbf{P}}_{k+1}, \\ \tilde{\mathbf{P}}_{k+1} &= \tilde{\mathbf{F}}_k \tilde{\mathbf{P}}_k \tilde{\mathbf{F}}_k^T + \tilde{\mathbf{\Gamma}}_k \tilde{\mathbf{Q}}_k \tilde{\mathbf{\Gamma}}_k^T \\ \boldsymbol{\Upsilon}_{k+1} &= \tilde{\mathbf{H}}^T (\tilde{\mathbf{R}}_{k+1})^{-1} \tilde{\mathbf{H}}, \\ \tilde{\mathbf{R}}_{k+1} &= \sum_{i=1}^N (\mathbf{w}_{k+1}^i)^2 (\mathbf{M}_{k+1}^i)^T \mathbf{K}_{k+1}^i \mathbf{M}_{k+1}^i, \\ \mathbf{M}_{k+1}^i &= \text{diag}(\mathbf{m}(\tilde{\mathbf{Y}}_{k+1}^{i,r=0}), \dots, \mathbf{m}(\tilde{\mathbf{Y}}_{k+1}^{i,r=2\sigma})), \\ \mathbf{K}_{k+1}^i &= \text{diag}(\mathbf{R}_i(\tilde{\mathbf{X}}_{k+1}^{r=0}), \dots, \mathbf{R}_i(\tilde{\mathbf{X}}_{k+1}^{r=2\sigma})) \end{aligned} \quad (12)$$

式中, $\tilde{\mathbf{P}}_{k+1}$ 是预测的全局误差协方差阵; 初始全局误差协方差阵 $\tilde{\mathbf{P}}_0 = \text{diag}(\mathbf{P}_0, \dots, \mathbf{P}_0)$; $\mathbf{m}(\tilde{\mathbf{Y}}_{k+1}^{i,r})$ 是由测量值样本 $\tilde{\mathbf{Y}}_{k+1}^r$ 计算得到的旋转矩阵; $\mathbf{R}_i(\tilde{\mathbf{X}}_{k+1}^r)$ 是式(5)中测量噪声的协方差阵(由 $k+1$ 时刻的目标运动状态样本 $\tilde{\mathbf{X}}_{k+1}^r$ 计算得到). 此外, $\tilde{\mathbf{F}}_k$ 是由 $2\sigma+1$ 个矩阵 \mathbf{F}_k 所组成的块对角阵, 同理, $\tilde{\mathbf{Q}}_k$ 、 $\tilde{\mathbf{H}}$ 和 $\tilde{\mathbf{\Gamma}}_k$ 分别是由矩阵 \mathbf{Q}_k 、 \mathbf{H} 及 $\mathbf{\Gamma}_k$ 所组成的块对角阵.

依据上述全局协方差阵, 长期精度收益可近似估算为

$$\bar{J}_\Psi^{(1)} \approx \sum_{r=0}^{2\sigma} \sum_{k=0}^H \beta_r \text{tr}(\mathbf{P}_k^r) = \sum_{k=0}^H \text{tr}(\boldsymbol{\beta} \tilde{\mathbf{P}}_k) \quad (13)$$

$$\beta_{r=0} > 0, \beta_{r=1, \dots, 2\sigma} = \frac{1-\beta_0}{2\sigma}, \boldsymbol{\beta} = \text{diag}(\beta_0, \dots, \beta_{2\sigma}) \otimes \mathbf{I}_6$$

这里 β_r 是无迹采样时样本 $\boldsymbol{\varepsilon}_r$ 所对应的权值; 符号 \otimes 表示 Kronecker 积.

3.2 长期辐射代价的计算

式(10)中的 ELI 信息状态 $\boldsymbol{\chi}_{\mathbf{E}_k^i}$ 可用 q_i 维列向量表示, 其元素 $\chi_{\mathbf{E}_k^i}(m) = p(\mathbf{E}_k^i = m | \boldsymbol{\eta}_k^i)$ ($m=1, \dots, q_i$) 是平台 i 的信息集合 $\boldsymbol{\eta}_k^i = (\mathbf{Z}_{e_1}^i, \dots, \mathbf{Z}_{e_k}^i, \mathbf{u}_0^i, \dots, \mathbf{u}_{k-1}^i)$ 已知时随机变量 $\mathbf{E}_k^i = m$ 的条件概率. 若已知观测威胁度 $Z_{e_k}^i$, 则利用隐马氏模型滤波器^[13]可以得到 k 时刻的 ELI 信息状态, 即

$$\boldsymbol{\chi}_{\mathbf{E}_k^i} = \frac{[\mathbf{B}_k^i(\mathbf{Z}_{e_k}^i) \odot \mathbf{A}_{k-1}^i]^T \boldsymbol{\chi}_{\mathbf{E}_{k-1}^i}}{\mathbf{I}_{q_i}^T [\mathbf{B}_k^i(\mathbf{Z}_{e_k}^i) \odot \mathbf{A}_{k-1}^i]^T \boldsymbol{\chi}_{\mathbf{E}_{k-1}^i}} \quad (14)$$

其中, 符号 \odot 表示 Hadamard 积; \mathbf{I}_{q_i} 表示元素全部为 1 的

q_i 维列向量.实际上,式(14)中的分母部分是 ELI 信息状态的转移概率 $p(\mathbf{X}_{E_k^i} | \mathbf{X}_{E_{k-1}^i} u_{k-1}^i)$,因而结合式(14)可推导出下述条件概率 $P_{1:k}^i$:

$$\begin{aligned} P_{1:k}^i &= p(\mathbf{X}_{E_1^i}, \dots, \mathbf{X}_{E_k^i} | u_0^i, \dots, u_{k-1}^i, \mathbf{X}_{E_0^i}) \\ &= \prod_{t=0}^{k-1} p(\mathbf{X}_{E_{t+1}^i} | \mathbf{X}_{E_t^i} u_t^i) \\ &= \mathbf{I}_{q_i}^T \prod_{t=0}^{k-1} [\mathbf{B}_{t+1}^i(Z_{e_{t+1}^i}^i) \odot \mathbf{A}_t^i]^T \mathbf{X}_{E_0^i} \end{aligned} \quad (15)$$

基于式(2)、(8)和(15),可以得到 k 时刻的单步辐射代价 \mathfrak{R}_k ,此时 \mathfrak{R}_k 只取决于平台 i 的动作序列 $\{u_0^i, \dots, u_{k-1}^i\}$.

$$\begin{aligned} \mathfrak{R}_k &= \sum_{i=1}^N \mathbb{E}_{Z_{e_1^i}, \dots, Z_{e_k^i}} (\mathbf{D}_i^T \mathbf{X}_{E_k^i}) = \sum_{i=1}^N \left[\sum_{Z_{e_1^i}, \dots, Z_{e_k^i}} P_{1:k}^i \mathbf{D}_i^T \mathbf{X}_{E_k^i} \right] \\ &= \sum_{i=1}^N \mathbf{D}_i^T \prod_{t=0}^{k-1} \left\{ \left[\sum_{Z_{e_{t+1}^i}} \mathbf{B}_{t+1}^i(Z_{e_{t+1}^i}^i) \right] \odot \mathbf{A}_t^i \right\}^T \mathbf{X}_{E_0^i} \\ &= \sum_{i=1}^N \mathbf{D}_i^T \left[(\mathbf{A}_i)^T \sum_{t=0}^{k-1} u_t^i \right]^T \mathbf{X}_{E_0^i} \end{aligned} \quad (16)$$

将上式中的单步辐射代价累加后即可得到由动作序列 Ψ 所产生的长期辐射代价 $\bar{J}_{\Psi}^{(2)}$,即

$$\bar{J}_{\Psi}^{(2)} = \mathbb{E}_{Z_{e_1}, \dots, Z_{e_N}} \left(\sum_{k=0}^H \sum_{i=1}^N \mathbf{D}_i^T \mathbf{X}_{E_k^i} \right) = \sum_{k=0}^H \mathfrak{R}_k \quad (17)$$

式中 $\mathfrak{R}_0 \triangleq \sum_{i=1}^N \mathbf{D}_i^T \mathbf{X}_{E_0^i}$.

3.3 决策树与分枝定界算法

将所有可选动作序列看成是深度为 H 的决策树,图 1 给出了时域长度 $H=3$ 和平台个数 $N=2$ 时的决策树示意图.如图 1 所示,每一个可选动作 u_k 可将第 k 层的节点展开至下一层子节点,并产生属于该子节点的收益值 $\alpha \cdot \text{tr}(\beta \tilde{\mathbf{P}}_{k+1}^{(i)}) + (1-\alpha)\mathfrak{R}_{k+1}^{(i)}$.因此,寻找最优动作序列实际上就是在决策树中搜索具有最小收益总和的节点序列.

为了使用分枝定界法,需要有限时域收益的下界值作为剪枝依据以加快搜索速度.定理 1 给出了长期精度收益的下界值.

定理 1 若仅知从初始时刻到 t 时刻的动作序列 $\Psi_{0:t} = \{u_0, \dots, u_t\} (t < H-1)$,则整个时域长度 H 上动作序列 $\Psi = \{u_0, \dots, u_{H-1}\}$ 的长期精度收益 $\bar{J}_{\Psi}^{(1)}$ 的下界

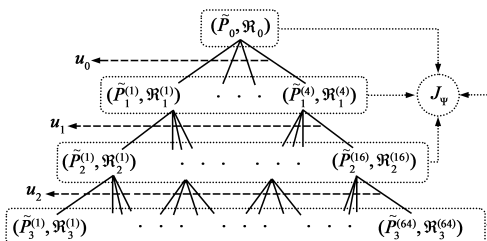


图1 决策树示意图($H=3, N=2$)

值为

$$L_{\Psi}^{(1)} = \sum_{k=0}^H \text{tr}(\beta \tilde{\mathbf{P}}_k) \quad (18a)$$

$$\tilde{\mathbf{P}}_k = \begin{cases} [\mathbf{I} + \tilde{\mathbf{P}}_k \mathbf{T}_k]^{-1} \tilde{\mathbf{P}}_k, & \text{if } 0 \leq k \leq t \\ [\mathbf{I} + \tilde{\mathbf{P}}_k \boldsymbol{\Phi}_k]^{-1} \tilde{\mathbf{P}}_k, & \text{if } t < k \leq H \end{cases} \quad (18b)$$

$$\tilde{\mathbf{P}}_k = \tilde{\mathbf{F}}_{k-1} \tilde{\mathbf{P}}_{k-1} \tilde{\mathbf{F}}_{k-1}^T + \tilde{\mathbf{T}}_{k-1} \tilde{\mathbf{Q}}_{k-1} \tilde{\mathbf{T}}_{k-1}^T \quad (18c)$$

$$\boldsymbol{\Phi}_k = \text{diag}(\max_{u_{k-1}} \sum_{i=1}^n |\mathbf{T}_k(1, i)|, \dots, \max_{u_{k-1}} \sum_{i=1}^n |\mathbf{T}_k(n, i)|) \quad (18d)$$

其中, $\mathbf{T}_k(j, i)$ 表示矩阵 \mathbf{T}_k 中位于第 j 行第 i 列的元素; n 是 \mathbf{T}_k 的维数.

证明 由式(12)中的 Γ_k 表达式及文献[14]中的定理可知

$$\mathbf{T}_k \leq_{(a)} \tilde{\mathbf{H}}^T \left(\sum_{i=1}^N (\mathbf{M}_{k+1}^i)^T \mathbf{K}_{k+1}^i \mathbf{M}_{k+1}^i \right)^{-1} \tilde{\mathbf{H}} \leq_{(b)} \boldsymbol{\Phi}_{u_{k-1}} \leq \boldsymbol{\Phi}_k \quad (19)$$

$$\boldsymbol{\Phi}_{u_{k-1}} = \text{diag}(\sum_{i=1}^n |\mathbf{T}_k(1, i)|, \dots, \sum_{i=1}^n |\mathbf{T}_k(n, i)|)$$

式(19)中,不等式(a)和(b)可分别由文献[14]中的定理3和定理4得到.以不等式(19)为基础,依据文献[14]中引理1和定理2可得 $\bar{J}_{\Psi}^{(1)} \geq L_{\Psi}^{(1)}$.故 $L_{\Psi}^{(1)}$ 是长期精度收益 $\bar{J}_{\Psi}^{(1)}$ 的下界值. 证毕.

基于上述定理,动作序列 Ψ 的有限时域收益下界值 L_{Ψ} 为

$$L_{\Psi} = \alpha L_{\Psi}^{(1)} + (1-\alpha) L_{\Psi}^{(2)} \quad (20)$$

其中 $L_{\Psi}^{(2)} = \sum_{k=0}^H \mathfrak{R}_k + \min_{u_{t+1}, \dots, u_{H-1}} \sum_{k=t+1}^H \mathfrak{R}_k$ 是长期辐射代价的下界值.具体的分枝定界算法见算法1.

算法1 分枝定界算法

```

令  $J_{\min} = +\infty$ .
将决策树顶点  $(\tilde{\mathbf{P}}_0, \mathfrak{R}_0)$  放入存储列表.
while 存储列表不为空
  展开列表中第一个节点.
  确定各平台的数据融合权重.
  计算每个子节点的收益.
  从列表中移除第一个节点.
  if 子节点位于  $H$  层且收益总和小于  $J_{\min}$ 
    令  $J_{\min}$  等于该收益总和;
    记录节点的上标.
  else
    计算每个子节点对应的下界值;
    将所有子节点存入列表.
end
从列表中移除下界值不小于  $J_{\min}$  的节点.
依据下界值的升序顺序排列列表中节点.
    
```

按照所采用的搜索方法安排列表.

end

依据记录节点的上标得到最优动作序列 Ψ^* .

4 仿真结果

考虑三个传感器平台协同跟踪一个在二维平面中进行近似匀速直线运动的目标,该目标的初始位置是(28km,30km),初始速度是(-160m/s, -200m/s),过程噪声协方差阵为 $\text{diag}(7,7)$.三个平台的位置坐标分别是(8km,10km),(10km,8km),(10km,10km);主动传感器的测量噪声方差为($\sigma_r = 20\text{m}, \sigma_\theta = 0.0035\text{rad}$),被动传感器的测量噪声方差为($\sigma_r = 45\text{m}, \sigma_\theta = 0.0017\text{rad}$).主动传感器的 ELI 状态具有三个状态值{1,2,3},分别代表实际情况下低、中、高 ELI 值.瞬时观测威胁度也量化为三个状态{1,2,3},分别代表威胁度不变、威胁度增加和威胁度下降.此外,样本权重 $\beta_0 = 0.01$,时域长度 $H = 4$.

4.1 搜索方法的比较

本文采用标准代价搜索(Uniform Cost Search, UCS)和贪婪搜索(Greedy Search, GS)这两种方法进行寻优搜索.以穷举搜索法(Exhaustive Search, ES)作为基准,表1中给出了不同平衡系数 α 下两种搜索方法的具体搜索结果.从表1中可以看出,随着平衡系数 α 的增大,GS方法的存储节点个数始终远远小于 UCS 方法的存储节点个数,这说明 GS 方法的存储开销远小于 UCS.此外,GS方法和 UCS 方法的打开节点个数相等并同步增加,说明两种搜索方法的搜索处理速度是一样的.综合来看,GS方法在性能上要优于 UCS 方法.

表1 搜索结果比较

		GS	UCS	ES
$\alpha = 0.01$	存储节点个数	22(0.54%)	64(1.56%)	4096
	打开节点个数	208(4.44%)	208(4.44%)	4680
$\alpha = 0.03$	存储节点个数	22(0.54%)	379(9.25%)	4096
	打开节点个数	2896(61.88%)	2896(61.88%)	4680
$\alpha = 0.09$	存储节点个数	22(0.54%)	512(12.5%)	4096
	打开节点个数	3808(81.37%)	3808(81.37%)	4680

4.2 平衡系数的影响

平衡系数 α 的选取影响着精度收益和辐射代价之间的平衡关系,图2显示了三种 α 取值条件下,经1000次蒙特卡洛仿真后的累积精度收益和累积辐射代价.如图2所示,只有在 $\alpha = 0.03$ 时,累积精度收益与累积辐射代价基本接近,此时精度收益和辐射代价达到最佳平衡.

表2给出了三种 α 取值条件下,总体收益值和最优动作序列.在表2中,最优动作序列中的“1”表示所有平台都启动主动传感器,“8”表示所有平台都启动被动传感器.结合图2,对不同 α 的最优动作序列进行比较

可知,当 $\alpha = 0.01$ 时,优先考虑辐射代价,故动作序列中“8”出现较多;当 $\alpha = 0.09$ 时,优先考虑精度收益,故动作序列中“1”出现较多;当 $\alpha = 0.03$ 时,考虑辐射代价与精度收益的平衡,故最优动作序列中“8”和“1”出现次数相当.此外,根据总体收益的理论值和实际值相差不大这一事实可知,有限时域收益的表达式能够正确地衡量动作序列的优劣,因而以该表达式确定最优动作序列是可行有效的.

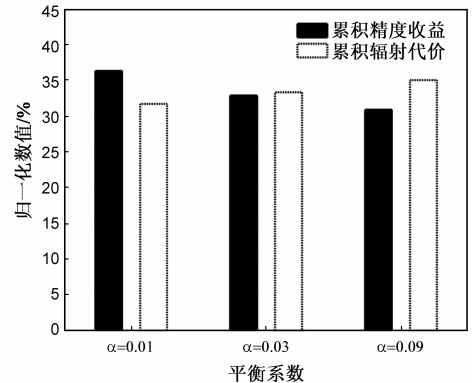


图2 精度收益和辐射代价之间的平衡变化

表2 总体收益值和最优动作序列

	总体收益理论值	总体收益实际值	总体收益相对误差	最优动作序列
$\alpha = 0.01$	39.3942	40.5594	2.87%	{8 8 8 1}
$\alpha = 0.03$	54.4569	54.8792	0.77%	{8 8 1 1}
$\alpha = 0.09$	97.1418	99.1427	2.02%	{8 1 1 1}

4.3 长期收益的比较

为了证明本调度方法的有效性,将无调度和随机调度这两种情况与本文方法进行比较.表3中列举了无调度情形下(NS 1到NS 8)各平台的传感器启动情况.

表3 无调度情形下各平台的调度情况

				平台1	平台2	平台3	
NS 1	主动	主动	主动	NS 5	被动	主动	主动
NS 2	主动	主动	被动	NS 6	被动	主动	被动
NS 3	主动	被动	主动	NS 7	被动	被动	主动
NS 4	主动	被动	被动	NS 8	被动	被动	被动

图3给出了经过1000次蒙特卡洛仿真后,各调度方法的归一化累积精度收益、累积辐射代价及累积总体收益.在图3中,本方法明显优于 NS 2、NS 3、NS 5及随机调度情况;与 NS 4、NS 6、NS 7及 NS 8相比,本方法以辐射代价的小幅度升高换取精度收益上的大幅度下降,避免了过多使用被动传感器所导致的辐射风险下降而跟踪精度过低;与 NS 1相比,本方法以精度收益的小幅度升高换取辐射代价的大幅度下降,避免了过分使用主动传感器所造成的跟踪精度提升而辐射风险过高.综上,本文方法能够合理地权衡长期精度收益和长

期辐射代价,进而使得所有平台在有限时域上实现跟踪精度与辐射风险的折中平衡。

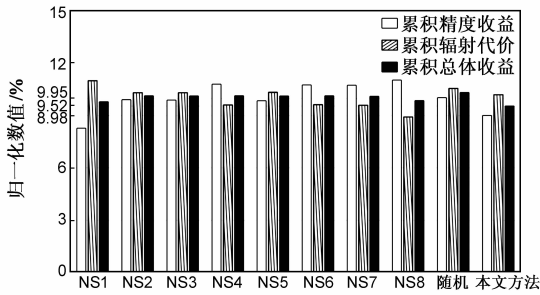


图3 精度收益、辐射代价及总体收益

5 结论

本文对多平台主被动传感器长期调度问题进行了研究.为了合理权衡有限时域内的跟踪精度和辐射风险,建立了基于部分可观马氏决策过程的数学模型,并结合先见优化、无迹采样和辐射代价表达式将原有的调度问题转化为决策树问题,最后利用分枝定界方法进行了求解.仿真结果表明,本方法可行有效,与无调度和随机调度情况相比,可以自动权衡精度收益和辐射代价,进而使得有限时域上的跟踪精度和辐射风险达到折中平衡。

参考文献

- [1] C Kreucher, K Kastella, A O Hero. A Bayesian method for integrated multitarget tracking and sensor management[A]. Proceedings of International Conference on Information Fusion [C]. Cairns, Qld, Australia; IEEE, 2003. 132 - 139.
- [2] 周文辉, 胡卫东, 等. 基于协方差控制的集中式传感器分配算法研究[J]. 电子学报, 2003, 31(12): 2158 - 2162. Zhou Wen-hui, Hu Wei-dong, et al. A study of sensor allocation algorithm based on covariance control in centralized multisensor systems[J]. Acta Electronica Sinica, 2003, 31(12): 2158 - 2162. (in Chinese)
- [3] A Logothetis, A Isaksson. On sensor scheduling via information theoretic criteria [A]. Proceedings of the American Control Conference [C]. San Diego, CA, USA: American Automatic Control Council, 1999. 2402 - 2406.
- [4] 程洪玮, 等. 一种基于信息决策树的低轨星座传感器调度方法[J]. 电子学报, 2010, 38(11): 2630 - 2634. Cheng Hong-wei, et al. A sensor scheduling method of LEO constellation based on information decision tree[J]. Acta Electronica Sinica, 2010, 38(11): 2630 - 2634. (in Chinese)
- [5] Y Li, L W Krakow, E P K Chong, et al. Approximate stochastic dynamic programming for sensor scheduling to track multiple target[J]. Digital Signal Processing, 2009, 19(6): 987 - 989.
- [6] Y He, E P K Chong. Sensor scheduling for target tracking: A

Monte Carlo sampling approach[J]. Digital Signal Processing, 2006, 16(5): 533 - 545.

- [7] L Kaelbling, M Littman, A Cassandra. Planning and acting in partially observable stochastic domains [J]. Artificial Intelligence, 1998, 101(1-2): 99 - 134.
- [8] V Krishnamurthy. Emission management for low probability intercept sensors in network centric warfare [J]. IEEE Transactions on Aerospace and Electronic Systems, 2005, 40(1): 133 - 152.
- [9] T L Song, J Y Ahn, C Park. Suboptimal filter design with pseudomeasurements for targets tracking [J]. IEEE Transactions on Aerospace and Electronic Systems, 1988, 24(1): 28 - 39.
- [10] X R Li, V P Jilkov. A survey of maneuvering target tracking-part III: measurement models [A]. Proceedings of SPIE Conference on Signal and Data Processing of Small Targets [C]. San Diego, CA, USA: SPIE, 2001. 559 - 583.
- [11] 刘敏华, 萧德云. 基于相似度的多传感器数据融合 [J]. 控制与决策, 2004, 19(5): 534 - 537. Liu Min-hua, Xiao De-yun. Multi-sensor data fusion based on similitude degree [J]. Decision and Control, 2004, 19(5): 534 - 537. (in Chinese)
- [12] S J Julier, J K Uhlmann. Unscented filtering and nonlinear estimation [J]. Proceedings of the IEEE, 2004, 92(3): 401 - 422.
- [13] L R Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition [J]. Proceedings of the IEEE, 1989, 77(2): 257 - 285.
- [14] Z G Feng, K L Teo, V Rehbock. Hybrid method for a general optimal sensor scheduling problem in discrete time [J]. Automatic, 2008, 44(5): 1295 - 1303.

作者简介



张子宁 男, 1984 年生于河北石家庄, 博士, 主要从事传感器管理及信息融合方面的研究。

E-mail: iron_zay@126.com

单甘霖 男, 1962 年生于江苏如东, 教授, 主要从事信息融合、神经网络和目标跟踪等方面的研究工作。

段修生 男, 1970 年生于河北, 副教授, 主要研究方向为人工智能和故障诊断。