doi:10.3969/j.issn.1671-637X.2014.08.004

基于Q学习的双机协同探测路径规划方法

吴彦锐[1], 伍友利[2], 丁 未[2], 柴 栋[2]
(1.陕西科技大学电气与信息工程学院,西安 710021; 2.空军工程大学航空航天工程学院,西安 710038)

摘 要: 建立双机协同被动探测的任务模型,运用模糊理论对问题状态空间进行泛化,针对不同的探测阶段给出了目标转移函数的不同形式,通过合理定义动作空间和奖励函数将问题描述为Markov决策过程。给出了双机协同被动雷达探测的模糊Q学习算法,并对算法进行仿真,仿真结果表明,该方法能够有效控制双机的飞行路径,实现对机动及非机动目标的有效探测。

关键词: 双机协同; 路径规划; 模糊Q学习; 目标探测
中图分类号: V271.4; TP391.9 **文献标志码**: A **文章编号**: 1671-637X(2014)08-0015-05

Dual-Aircraft Cooperative Path Planning Based on Q Learning

WU Yan-rui[1], WU You-li[2], DING Wei[2], CHAI Dong[2]
(1. College of Electrical & Information Engineering, Shaanxi University of Science & Technology, Xi'an 710021, China;
2. Engineering College of Aeronautics and Astronautics, Air Force Engineering University, Xi'an 710038, China)

Abstract: Based on the partition of the radiation area of the target's active radar, the task model of the dual-aircraft is set up. By using fuzzy theory to make a generalization to the problem's state space, providing different transition functions according to different detecting stages, and properly defining the action space and reward function, the problem is formulated as a Markov Decision Process (MDP). Details of the fuzzy Q learning are presented. Simulation studies indicate that the proposed algorithm can provide adaptive strategies for the dual-aircraft to control their flight paths for non-maneuvering or maneuvering target detection.

Key words: dual-aircraft coordination; path planning; fuzzy Q learning; target tracking

0 引言

采用运动可控平台对辐射源进行定位是被动探测体制经常采用的方法,由于单个平台只能报告其接收到信号的到达方位和到达时间,所以通常采用多平台协同的方式利用三角定位法[1]对目标进行定位。实现被动探测时有可能使载机暴露在目标的攻击范围内,遭到目标机载武器的攻击。如何在保证载机安全的情况下实现对目标的定位是实现被动探测必须解决的问题。此外,双机组成的被动探测系统还受到通信距离、目标辐射控制[2]等因素的限制和影响,所以,寻找合适的控制策略以规划载机的飞行路径对实现双机协同被动目标探测十分重要。

目前,用以解决飞行路径规划问题的方法主要有两种。第一种是基于模型的优化方法:文献[3]基于搜索理论方法,采用搜索域上的"回报率"状态图,实现了多UAV协同中的搜索路径规划;文献[4]建立了多UAV被动雷达传感器目标跟踪框架,通过建立误差协方差最小和信息最大两个指标分别计算了UAV的航迹;文献[5]研究了多UAV广域目标搜索的协同控制问题,其飞行控制策略通过建立目标发现收益、环境搜索收益和协同收益指标得到。第二种方法是基于多智能体的飞行路径规划方法:文献[6]建立了多Agent协同探测问题的通用框架,以"目标搜索图"的形式存储环境信息,并基于搜索图在线计算UAV的飞行轨迹;文献[7]在有限感知范围内采用多Agent协商机制实现了多UAV协同搜索路径决策。基于模型优化方法的效果在很大程度上依赖于所建立模型的精确程度;基于多智能体的方法则相对灵活,它不依赖于所建模型,通过合理地构

收稿日期:2013-09-03 **修回日期**:2013-12-25
基金项目:国家自然科学基金(60874040);陕西省自然科学基金(2014JQ8339)
作者简介:吴彦锐(1981—),女,山西高平人,硕士,讲师,研究方向为自动控制。

建 Agent，利用 Agent 与环境交互过程中获得的相关信息实现对载机飞行路径的规划。强化学习是多智能体理论中一种常用的 Agent 控制方法，其中，Q 学习方法堪称经典，它不依赖于模型、在线更新策略的优点使得其在众多领域获得广泛应用[8-9]。本文采用模糊 Q 学习方法解决载机飞行路径的规划问题。

1 双机协同被动雷达探测任务模型

双机协同被动探测的相对几何态势如图 1 所示。这里假设执行被动探测任务的载机天线指向与其航向相同，被动雷达的最大搜索方位角为 $2\varphi_p$，最大探测距离为 D_p。定义目标视线 F_iF_T 及长度 R_i；目标方位角 q_i；目标进入角 θ_i；任务机方位角 β_i；我机与目标机雷达天线方向之间的夹角 ϕ_i。

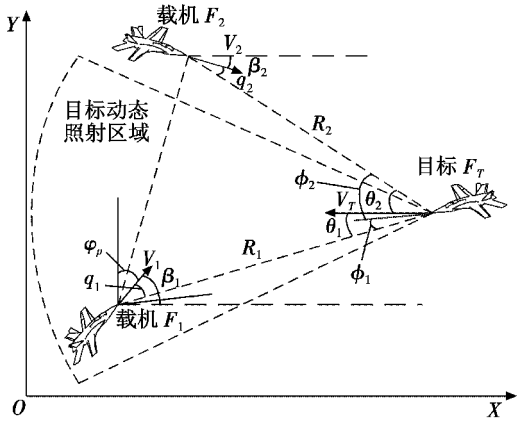

图 1 双机协同被动探测态势

Fig. 1 The engagement geometry of dual-aircraft in cooperative target detecting

如图 1 所示，双机协同被动雷达探测就是利用任务机 F_1、F_2 测得的目标方位角 q_1、q_2，结合已知的双机距离 $|F_1F_2|$，在三角形 $F_1F_2F_T$ 中利用正弦定理对目标 F_T 进行定位。可见，任务机在探测过程中的任务可以分为 2 个阶段来实施：1）搜索阶段，主要解决如何发现目标的问题；2）定位阶段，该阶段主要解决如何提高探测精度的问题。

对于搜索阶段，要完成发现目标的任务，必须规划任务机的飞行轨迹使得任务机被动雷达传感器能够接收到目标的辐射信号，即进入目标的主动雷达动态照射区，应满足

$$\begin{cases} |\theta_i| \leqslant \varphi_d \\ |q_i| \leqslant \varphi_p \\ R_i \leqslant D_d \end{cases} 。 \tag{1}$$

对于定位阶段，给出任务机进入目标主动雷达主瓣照射区接收信号应满足的条件为

$$\begin{cases} |\phi_i| \leqslant \varphi_m \\ |q_i| \leqslant \varphi_p \\ R_i \leqslant D_d \end{cases} 。 \tag{2}$$

为降低测向误差对定位精度的影响，应满足的条件为[10]

$$\dot{R}_i < 0, \text{ when } R_i \geqslant D_w 。 \tag{3}$$

为防止形成 V 型基线应满足的约束条件为

$$\frac{y_2 - y_1}{x_2 - x_1} \neq \frac{y_T - y_1}{x_T - x_1} \neq \frac{y_T - y_2}{x_T - x_2} \tag{4}$$

式中，(x_1, y_1)、(x_2, y_2)、(x_T, y_T) 分别记为 $\boldsymbol{x}_{F_1}$、$\boldsymbol{x}_{F_2}$、$\boldsymbol{x}_T$，表示任务机 1、任务机 2 及目标的状态。

此外，任务执行过程中还应满足一个约束条件，即任务机之间的距离应不超过载机间数据通信的最大距离 D_c，且不能小于载机间安全距离 D_f，即

$$D_f \leqslant |F_1F_2| \leqslant D_c 。 \tag{5}$$

2 双机协同被动雷达探测的 MDP 模型

一个完整的马尔可夫决策过程（Markov Decision Process，MDP）问题描述包括 4 个部分，分别为问题的状态空间、动作空间、转移函数和奖励函数，下面分别进行定义。

2.1 状态空间

2.1.1 基于相对态势的状态空间划分

本文主要研究 $0 < |\theta_i| < 90°$ 的被动探测问题，进行分析可以将原状态空间按任务机与目标态势划分为 5 部分，即

$$S = \begin{cases} s_1 = \begin{cases} R_i > D_d \\ D_w < R_i \leqslant D_d,\ \varphi_d < |\theta_i| < 90° \\ D_w < R_i \leqslant D_d,\ \varphi_p \leqslant |q_i| \leqslant 180°,\ 0 < |\theta_i| < \varphi_d \end{cases} \\ s_2 = \{D_w < R_i \leqslant D_d,\ 0 < |q_i| < \varphi_p,\ 0 < \theta_i < \varphi_d\} \\ s_3 = \{D_w < R_i \leqslant D_d,\ 0 < |q_i| < \varphi_p,\ -\varphi_d < \theta_i < 0\} \\ s_4 = \{D_w < R_i \leqslant D_d,\ 0 < |q_i| < \varphi_p,\ 0 < |\phi_i| < \varphi_m\} \\ s_5 = \{R_i \leqslant D_w, 0 < |\theta_i| < \varphi_d\} \end{cases} \tag{6}$$

式(6)实际上构成了从任务机的状态空间 $\boldsymbol{x}_{F_i}$ 与目标的状态空间 $\boldsymbol{x}_T$ 到一个新状态空间 S 的对应关系，记为 χ，则

$$\chi: (\boldsymbol{x}_{F_i}, \boldsymbol{x}_T) \rightarrow S = \{s_1, s_2, s_3, s_4, s_5\} 。 \tag{7}$$

2.1.2 状态空间的模糊近似

本文并不直接利用 $\boldsymbol{x}_{F_i}$ 和 $\boldsymbol{x}_T$ 进行原状态空间到新状态空间的映射，而是通过 $\boldsymbol{x}_{F_i}$ 和 $\boldsymbol{x}_T$ 计算出目标的相对态势关系 $(R_i, \theta_i, q_i, \phi_i)$，记为 $\boldsymbol{x}_i$，文献[11]指出为了保证近似 Q 值函数收敛，每个隶属度函数必须在唯一点取得最大值，三角形隶属度函数满足该要求，以状态分量 R_i 为例给出具体的隶属度函数，可表示为

$$\begin{cases}\xi_{1,1}(R_i)=\max\left(0,\dfrac{R_w-R_i}{R_w}\right)\\ \xi_{1,2}(R_i)=\max\left(0,\min\left(\dfrac{R_i}{R_w},\dfrac{R_d-R_i}{R_d-R_w}\right)\right)\\ \xi_{1,3}(R_i)=\max\left(0,\min\left(\dfrac{R_i-R_w}{R_d-R_w},\dfrac{R_\infty-R_i}{R_\infty-R_d}\right)\right)\\ \xi_{1,4}(R_i)=\max\left(0,\dfrac{R_i-R_d}{R_\infty-R_d}\right)\end{cases}。\quad(8)$$

其他 3 个状态分量的隶属度函数 $\xi(\theta_i)$、$\xi(|q_i|)$、$\xi(|\phi_i|)$ 的计算方法与 R_i 相同，限于篇幅，这里不再赘述。

得到各状态分量的隶属度函数后，通过乘积推理就能得到状态变量 $\boldsymbol{x}_i$ 的 4 维隶属度函数

$$\mu_n(x_i)=\xi(R_i)\cdot\xi(\theta_i)\cdot\xi(|q_i|)\cdot\xi(|\phi_i|)\quad(9)$$

这样就实现了原状态空间的模糊近似，它能够实现一个状态与邻近状态之间的泛化，当某个动作能够在该状态获得较高的 Q 值时，同样也会给邻近状态带来合理的决策。

2.2　动作空间

假设任务机速度大小不变为 V，只进行航向控制，任务机航向的控制方程为

$$\beta_i[k+1]=\beta_i[k]+\Delta\beta_i\quad(10)$$

式中：$\Delta\beta_i\in U_i=\{u_1^i,\cdots,u_M^i\mid|u_M^i|\leqslant\Delta\beta_{\max},m=1,\cdots,M\}$；$U_i$ 为任务机 F_i 的动作空间，规定逆时针方向旋转为正，则当 $\Delta\beta_i$ 为正时表示任务机逆时针旋转、为负时则为顺时针旋转，为 0 时表示其保持原来航向；$\Delta\beta_{\max}$ 为任务机的最大旋转角度，它受自身可用过载的限制。

2.3　转移函数

对于任务机 i 与目标组成的系统在状态 s_j 时采用动作 u_m^i 和 u 转移到状态 $s_{j'}$ 的转移函数可以定义为

$$p_i(s_{j'}\mid s_j,u_m^i,u)=P(s^{(k+1)}=s_{j'}\mid s^{(k)}=s_j,u_i^{(k)}=u_m^i,u_T^{(k)}=u)\quad(11)$$

式中：$s_j,s_{j'}\in S$；$u_m^i\in U_i$；u 为目标动作。

假设载机所选的动作不会对目标的运动产生影响[12]，$p_i(s_{j'}\mid s_j,u_m^i,u_T)$ 可以进一步表示为

$$p_i(s_{j'}\mid s_j,u_m^i,u)=p_i(s_{j'}\mid s_j,u_m^i)p_i(s_{j'}\mid s_j,u)。\quad(12)$$

当处于搜索阶段时，记搜索阶段区域的中心为 $\bar{C}$，则可以假设目标下一时刻的状态 s_T' 服从以 $\bar{C}$ 为中心、σ_C 为强度的正态分布，即状态转移函数可表示为

$$p_i(s_{j'}\mid s_j,u)=\int_{-\infty}^{s_T'}\frac{1}{\sqrt{2\pi}\sigma_C}\exp\left[-\frac{(s_T'-\bar{C})^2}{2\sigma_C^2}\right]\mathrm{d}s_T'。\quad(13)$$

当处于定位阶段时，由于任务机可以根据目标的辐射信号获得目标的当前状态 s_T，则可以假设目标下一时刻的状态 s_T' 服从以当前状态 s_T 为中心、σ_T 为强度的正态分布，即此时目标的状态转移函数可表示为

$$p_i(s_{j'}\mid s_j,u)=\int_{-\infty}^{s_T'}\frac{1}{\sqrt{2\pi}\sigma_T}\exp\left[-\frac{(s_T'-s_T)^2}{2\sigma_T^2}\right]\mathrm{d}s_T'\quad(14)$$

这样获得目标的状态转移函数后代入式(17)就可以获得任务机与目标组成的系统的状态转移函数，确定系统下一时刻的状态 $s_{j'}$。

2.4　奖励函数

根据状态空间的定义，系统的奖励函数可以采用下列确定形式

$$\rho_i(s,u_m^i,u_T,s')=\begin{cases}-1, & s'=s_1\\ -5, & s'=s_5\\ 1, & s'=s_2\text{ or }s_3\\ 5, & s'=s_4\end{cases}。\quad(15)$$

式(15)表明，当任务机无法获取目标辐射信号时得到的奖励信号为 -1，当任务机天线接收范围进入目标主动雷达动态照射区时，得到的奖励信号为 1，进入目标雷达主瓣照射区时得到的奖励信号为 5，而一旦进入目标武器威胁区后得到的奖励信号为 -5。

3　模糊 Q 学习算法

经典 Q 学习算法的核心思想是状态动作对的最优值函数为即时奖励与在下一状态 $\boldsymbol{x}_i'$ 获得最优值的折扣和，即

$$Q(\boldsymbol{x}_i,u_m^i,u)=\rho(\boldsymbol{x}_i,u_m^i,u)+\gamma\max_{u_{m'}^i\in U_i}Q(\boldsymbol{x}_i',u_{m'}^i,u')\quad(16)$$

式中，$\gamma\in[0,1]$，为折扣因子。

最优策略为在每一状态使得值函数最优的动作的集合，即

$$h^*=\arg\max_{u_m^i}Q(\boldsymbol{x}_i,u_m^i,u)。\quad(17)$$

完成状态空间的模糊近似和动作空间的划分后，可采用下面线性权值函数对上述值函数进行逼近[13]

$$Q(\boldsymbol{x}_i,u_m^i,u)=\sum_{n=1}^{N}\psi_n(\boldsymbol{x}_i)\Omega_{[n,m]}\quad(18)$$

式中：$\Omega_{[n,m]}$ 为迭代参数；$\psi_n(\boldsymbol{x}_i)$ 即为归一化后的隶属度函数，其值为

$$\psi_n(\boldsymbol{x}_i)=\frac{\mu_n(\boldsymbol{x}_i)}{\sum_{n'=1}^{5}\mu_{n'}(\boldsymbol{x}_i)}。\quad(19)$$

结合前面对双机协同探测任务的分析，可以将上述过程进一步描述为找到满足约束条件式(3)～式(5)的动作 u_m^i 对式(18)进行更新，具体的算法流程如图 2 所示。

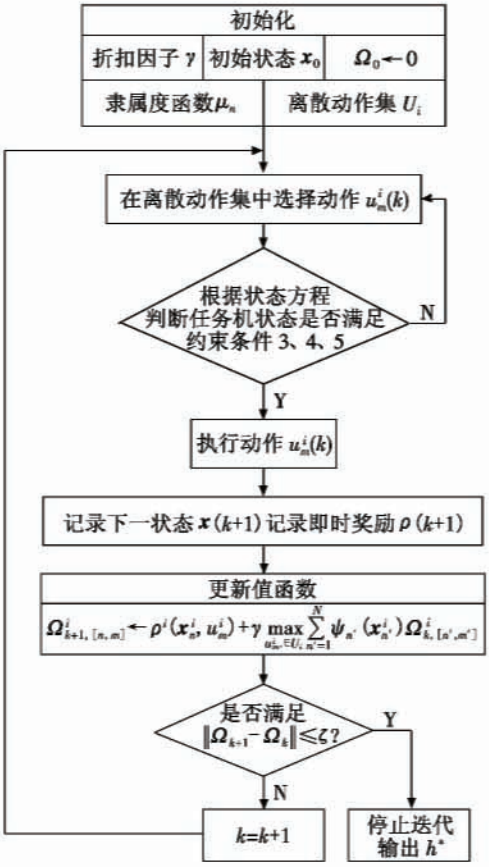

图2 双机协同被动探测的模糊Q学习算法

Fig. 2 Fuzzy Q learning algorithm for dual-aircraft flight path planning

算法进行过程中每架任务机均进行各自Q值的迭代更新，但在选出动作 $u_m^i(k)$ 之后，需要将 $u_m^i(k)$ 代入各自的状态方程并判断是否满足约束条件，即式(3)~式(5)，若满足则继续进行Q学习算法，若不满足则需要返回上一步重新选择 $u_m^i(k)$，当满足该条件时才转入下一步。ζ 为一个很小的正数，表示当 Ω 值基本稳定时则停止迭代，输出控制策略。

4 仿真结果

4.1 仿真参数

仿真时的参数设置：任务机与目标在欧式空间中的坐标满足 0 km $\leqslant X \leqslant$ 200 km，−10 km $\leqslant Y \leqslant$ 10 km。任务机被动雷达的有效探测距离 D_p = 200 km，最大搜索方位角为 $2\varphi_p = 60°$。目标主动雷达的最大作用距离为 D_d = 100 km，最大动态视场角 $2\varphi_d = 120°$，主瓣宽度 $2\varphi_m = 6°$，扫描周期 T_m = 5 s。目标武器的射程 D_w = 60 km，最大离轴发射角 $2\varphi_w = 120°$。任务机与目标速度大小均为 200 m/s，任务机与目标的初始态势按照(X 坐标，Y 坐标，航向)格式设为两组，分别为：1) F_1(0 km，−2.5 km，0°)，F_2(0 km，2.5 km，0°)，F_T(150 km，0 km，180°)；2) F_1(0 km，−2.5 km，0°)，F_2(0 km，2.5 km，0°)，F_T(150 km，6 km，180°)。$\bar{C}$ 为目标的初始坐标，σ_C 取 10^4，σ_T 取 200。任务机 F_1、F_2 具有相同的离散化动作空间，共包含5个动作，为 $U_1 = U_2 = \{-3°, -1.5°, 0°, 1.5°, 3°\}$。

模糊Q学习算法的参数定义为：折扣因子 γ = 0.95，初始 $\Omega_{[n,m]}^0 = 0$，最大学习步数 k = 500，终止条件 ζ = 0.01，仿真步长 T = 1 s。这样，根据前面对任务机与目标坐标的限制，可以粗略地估算出原始状态空间的大小至少为 1×10^5，再与动作空间相联系，则可得状态－动作对的数目不少于 5×10^5。从理论上分析，若运用标准Q学习算法，每一步迭代都要更新的Q值数目庞大，不能满足实时规划的要求，下面的仿真对比将进一步说明该问题。

4.2 结果分析

分别运用标准Q学习算法与模糊Q学习算法对初始态势1)和态势2)进行了仿真。图3~图4为初始态势1)条件下目标保持匀速直线运动时双机协同探测的仿真图。图5~图6为初始态势2)条件下目标机动时双机协同探测的仿真图。

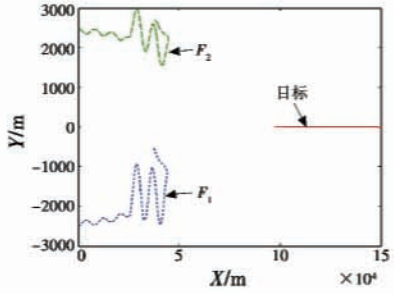

图3 双机协同被动探测规划路径

Fig. 3 The planned flight path of dual-aircraft for detecting an non-maneuvering target

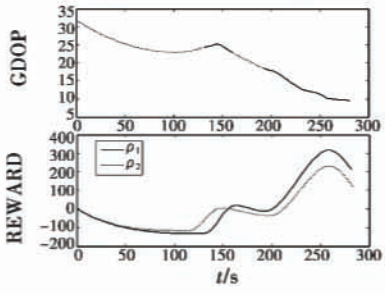

图4 奖励值及定位精度变化曲线

Fig. 4 GDOP of detecting the non-maneuvering target and rewards of the two aircrafts

从图3可以看出，按照本文提供的方法，双机对无机动目标的被动定位可以分搜索、跟踪定位及逃离3个阶段。在搜索阶段，双机在每一时刻的运动方向相反，即双机分别对不同的区域进行搜索，这样提高了发现目标的概率。跟踪定位阶段的仿真曲线表明双机能

够稳定地跟踪目标主动雷达的主瓣照射区，实现对目标的有效定位，图 4 的 GDOP 变化曲线的实线部分（虚线为按照仿真中双机和目标位置计算出的 GDOP 值，实际过程中由于无法获得目标辐射信号而无法计算，故用虚线表示）表明按照规划路径，双机对目标的定位误差持续下降，最终保持在 3.03 左右直至进入目标武器威胁区后受到惩罚而逃离。图 4 的双机的奖励函数变化曲线也反映了该过程，即搜索阶段由于无目标信号奖励一直为负，而后从进入目标主动雷达动态照射区到主瓣照射区奖励逐渐增加，当进入目标威胁区后再次下降，它表明文中定义的奖励函数能够有效反映双机被动雷达的探测任务。

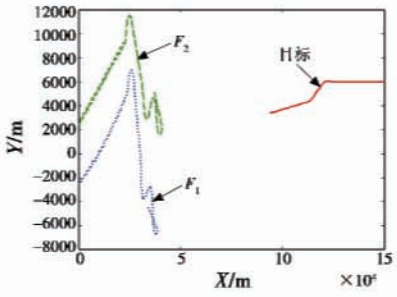

图 5　目标机动时双机协同被动探测规划路径

Fig. 5　The planned flight path of dual-aircraft for detecting a maneuvering target

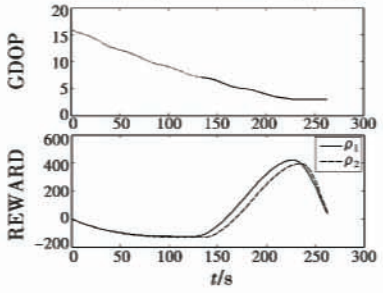

图 6　定位精度及奖励值变化曲线

Fig. 6　GDOP of detecting the maneuvering target and rewards of the two aircrafts

从图 5 可以看出，存在目标机动时的双机路径规划要比目标无机动时复杂，整个过程大致可分为搜索、跟踪定位、再搜索、再跟踪定位及逃离 5 个阶段。当目标机动时，双机组成的被动探测系统能够适应目标的变化，经过再搜索后仍能实现对目标的有效跟踪定位。图 6 的 GDOP 变化曲线反映出定位精度在经过变化后最终下降到 9.66 左右。图 6 的双机奖励函数变化曲线表明文中定义的奖励函数对跟踪机动目标同样具有适用性。

表 1 将标准 Q 学习算法与模糊 Q 学习算法进行了性能比较，可以看出，标准 Q 学习算法无论是在初始态势 1）还是初始态势 2）条件下的计算时间都远大于模糊 Q 学习。标准 Q 学习每一时间步都必须在线完成所有状态动作对的更新，无法满足路径实时规划的要求。模糊 Q 学习通过离线时计算出各个初始状态的隶属度，然后在线时只需对新状态进行隶属度判断，更新的状态动作对数目仅为 $5 \times 5 = 25$，因此能够实现对路径的实时规划。

表 1　不同算法的性能比较

Table 1　Standard Q (SQ) learning and fuzzy Q (FQ) learning

s

	初始态势 1）		初始态势 2）	
	离线	在线	离线	在线
标准 Q 学习算法	0	1447	0	1568
模糊 Q 学习算法	450	263	621	282

5　结束语

本文研究了将模糊 Q 学习算法引入二维情况下的双机协同被动探测路径规划问题。通过对双机协同被动探测问题的过程和目的进行分析，抽象出相对态势表示的任务机和目标组成系统的状态，在合理定义模糊隶属度函数的情况下实现了状态空间的高度泛化。在合理定义动作空间、转移函数及奖励函数的基础上给出了双机协同被动探测路径规划的模糊 Q 学习算法，并对算法在目标匀速直线运动及机动条件下的性能进行了仿真。仿真结果表明，本文所提算法能够实现对目标的有效跟踪定位，当目标机动时，算法能够保证任务机对环境改变的良好适应性。

参 考 文 献

[1] RICHARD A P. Electronic warfare target location methods [M]. Boston: Artech House, 2005.

[2] KRISHNAMURTHY V. Emission management for low probability intercept sensors in network centric warfare [J]. IEEE Transactions on Aerospace and Electronic Systems, 2005, 41(1): 133-151.

[3] BAUM M L, PASSINO K M. A search-theoretic approach to cooperative control for uninhabited air vehicles [C]// AIAA Guidance, Navigation, and Control Conference, 2002: 1-8.

[4] CASBEER D W. Decentralized estimation using information consensus filters with a multi-static UAV radar tracking system [D]. Hawaii: Brigham Young University, 2009.

[5] PENG H, SU F, SHEN L C. Extended search map approach for multiple UAVs wide area target searching [J]. Systems Engineering and Electronics, 2010, 32(4): 795-798.

[6] POLYCARPOU M M, YANG Y L, PASSINO K M. A cooperative search framework for distributed agents [C]//

（下转第 32 页）

trol for simultaneous arrival of multiple UAVs[J]. Acta Aeronautica Et Astronautica Sinica, 2010, 31(4): 797-805.

[7] ZHAO S Y, ZHOU R. Cooperative guidance for multi-missile salvo attack [J]. Chinese Journal of Aeronautics, 2008, 21(6):533-539.

[8] 张庆杰. 基于一致性理论的多UAV分布式协同控制与状态估计方法[D]. 长沙:国防科学技术大学,2011.
ZHANG Q J. Distributed cooperative control and statement estimation for networked multiple UAVs based on consensus theory[D]. Changsha: National University of Defense Technology, 2011.

[9] LI J, XU S, CHU Y, et al. Distributed average consensus control in networks of agents using outdated states[J]. IET Control Theory & Applications, 2010, 4(5):746-758.

[10] SABER R O, MURRAY R M. Consensus problems in networks of agents with switching topology and time-delays [J]. IEEE Transactions on Automatic Control, 2004, 49(9):1520-1533.

[11] 杨军,朱学平,朱苏朋,等. 飞行器最优控制[M]. 西安:西北工业大学出版社,2011.
YANG J, ZHU X P, ZHU S P, et al. Optimal control of aircraft[M]. Xi'an: Northwestern Polytechnical University Press, 2011.

[12] 冯新磊. 符号矩阵和多智能体系统一致性研究[D]. 成都:电子科技大学,2011.
FENG X L. Study of sign pattern matrix and consensus of multi-agent systems [D]. Chengdu: University of Electronic Science and Technology, 2011.

[13] CAO Y, REN W, CHEN Y Q. Multi-agent consensus using both current and outdated states[C]//IFAC World Congress, Seoul, Korea, 2008:2874-2879.

[14] 陈岩,苏菲,沈林成. 概率地图UAV航线规划的改进型蚁群算法[J]. 系统仿真学报,2009,21(6):1658-1666.
CHEN Y, SU F, SHEN L C. Improved ant colony algorithm based on PRM for UAV route planning[J]. Journal of System Simulation, 2009,21(6):1658-1666.

(上接第14页)

[6] TOL J V, GUNZINGER M, KREPINEVICH A F, et al. Airsea battle: A point-of-departure operational concept[R]. The Center for Strategic and Budgetary Assessments, 2010.

[7] 黄柯棣,刘宝宏,黄健,等. 作战仿真技术综述[J]. 系统仿真学报,2004,16(9):1887-1895.
HUANG K L, LIU B H, HUANG J, et al. A survey of military simulation technologies[J]. Journal of System Simulation, 2004, 16(9):1887-1895.

[8] 军事科学院. 中国人民解放军军语[M]. 北京:军事科学出版社,2011.
Academy of Military Sciences. PLA military language[M]. Beijing: Military Science Press, 2011.

[9] 丁笑亮,陈树新,毛玉泉. MC法与QA法在通信系统仿真中的应用比较[J]. 计算机仿真,2010,20(7):65-68.
DING X L, CHEN S X, MAO Y Q. Application and comparison of MC method and QA method on simulation of communication system[J]. Computer Simulation, 2010, 20(7):65-68.

[10] 刘宝宏,黄柯棣. 多分辨率建模的研究现状与发展[J]. 系统仿真学报,2004,16(6):1150-1153.
LIU B H, HUANG K L. Multi-resolution modeling: Present status and trends [J]. Journal of System Simulation, 2004, 16(6):1150-1153.

[11] 陈建华,李刚强,傅调平. 基于多分辨率的海军作战仿真建模研究[J]. 系统仿真学报,2009,21(22):7316-7319.
CHEN J H, LI G Q, FU D P. Research of muli-distinguish modeling on warship formation operation simulation[J]. Journal of System Simulation, 2009, 21(22):7316-7319.

(上接第19页)

IEEE International Symposium on Intelligent Control, 2001:1-6.

[7] SUJIT P B, GHOSE D L. Multiple UAV search using agent based negotiation scheme[C]//American Control Conference, 2005:2995-3000.

[8] WIERING M, SCHMIDHUBER J R. Fast online $Q(\lambda)$[J]. Machine Learning, 1998, 33(1):105-115.

[9] MILLAN J D R, POSENATO D, DEDIEU E. Continuous-action Q-learning [J]. Machine Learning, 2002, 49 (2/3):247-265.

[10] TORRIERI D J. Statistical theory of passive location systems[J]. IEEE Transactions on Aerospace and Electronic Systems, 1984, AES-20(2):183-198.

[11] TSITSIKLIS J N, ROY B V. Feature-based methods for large scale dynamic programming [J]. Machine Learning, 1996, 22(1-3):59-94.

[12] GAO X, FANG Y W, HU S G, et al. Angle precision study on dual-aircraft cooperatively detecting remote target by passive locating method[C]//IEEE International Conference on Signal Processing, Communication and Computing, 2011:1174-1178.

[13] BUSONIU L, BABUSKA R, SCHUTTER B D, et al. Reinforcement learning and dynamic programming using function approximators[M]. Florida: Automatic Control and Engineering Series, CRC Press, 2010:49-51.