

基于马尔科夫网络的无人机机动决策方法研究

罗元强, 孟光磊

(沈阳航空航天大学自动化学院, 沈阳 110136)

摘要: 为解决无人机在复杂不确定战场环境下的空战问题, 开展了机动决策方法研究。综合分析空战过程中反映战场态势的重要因素, 利用多状态转移马尔科夫网络构建机动决策网络。采用马尔科夫递归推理算法进行决策推理, 并进行算法复杂度分析。仿真结果表明空战过程中决策网络能准确感知战场态势输出灵活机动策略。分析算法复杂度可知推理过程执行的乘法、加法、除法次数为 6 694 840, 3 409 270, 3 470 680。在现有处理器的计算能力上, 能在 40 ms 内完成一个推理周期, 网络推理过程完全满足空战决策实时性要求。

关键词: 无人机; 机动决策; 多状态转移; 马尔科夫网络; 递归推理

中图分类号: V249.32+9

文献标识码: A

文章编号: 1004-731X (2017) S1-0106-07

DOI: 10.16182/j.issn1004731x.joss.2017S1015

Research on UAV Maneuver Decision - making Method Based on Markov Network

Luo Yuanqiang, Meng Guanglei

(School of Automation, Shenyang Aerospace University, Shenyang 110136, China)

Abstract: The research on maneuver decision-making method is carried out for solving the UAV air combat problems in the complex uncertainty battlefield. Based on the analysis of the factors that reflect the air-combat battlefield situation, the multi-state transition Markov network is used to establish the UAV maneuver decision-making network. The recursive reasoning algorithm based on Markov model is used to make decision reasoning, and the algorithm complexity is analyzed. The simulation results indicate that the decision network can accurately perceive battlefield situation to output the flexible mobile strategy during the air combat process. Algorithm complexity analysis indicates that the multiplication, addition, and division times performed in the reasoning process are 6694840, 3409270, 3470680. In the existing processor computing power, a reasoning cycle can be completed within 40ms, and the network reasoning process can fully meet the real-time requirement of air-combat decision making.

Keywords: UAV; maneuver decision making; multi-state transition; Markov network; recursive reasoning

引言

近年来无人机应用价值在几场局部战争中得

到突显, 受到了越来越多研究人员的关注。受制于智能水平, 制空型无人机还处于实验室研究阶段。空战自主智能决策技术是促进制空型无人机研制, 提升其作战能力的关键。目前, 学者们提出了多种空战决策方法。在文献[1]中利用微分对策理论构建追逃模型, 利用变分法原理求解机动决策问题, 并进行了仿真实验。钟友武、杨凌宇等将人工智能



收稿日期: 2017-06-18

修回日期: 2017-11-03;

作者简介: 孟光磊(1982-), 男, 黑龙江齐齐哈尔, 博士, 副教授, 研究方向为空战智能决策、飞行仿真; 罗元强(1992-), 男, 四川泸州, 硕士生, 研究方向为无人机空战决策。

<http://www.china-simulation.com>

方法与微分对策理论相结合, 从机器智能的角度开展自主机动决策方法研究^[2]。用微分对策求解空战决策问题时, 均要求在已知精确动力学模型的基础上, 才能获取有效可靠的优化控制序列, 在空战对抗过程中有一定的局限性。文献[3]提出滚动时域方法预测空战态势, 利用最优控制方法求解能使无人机获得最大攻击优势的控制序列。在文献[4]中沈东等提出基于 MTPM 和 DPM 的滚动时域决策方法, 解决了多机协同空战决策问题。通过预测敌机运动轨迹, 利用预测态势进行预判能提高机动决策有效性, 但如何提高预测的准确度是必须要解决的一个难题。文献[5]利用动态规划方法求解空战决策问题, 借助空战对抗结果进行离线学习, 然后进行在线决策, 该方法保证了决策求解过程的实时性。对飞机机动策略进行机器学习为空战决策问题提供了一种非常有益的解决思路, 它可自动从空战仿真和实战数据中提取决策策略, 是一种从客观角度进行空战知识学习的方法, 如何获取大量的不同机型间对抗的数据样本进行学习, 是提高这种方法适用性的关键。文献[6-7]利用影响图构建机动决策模型, 该方法使决策模型具备对不确定性战场环境的处理能力, 但建模过程中缺乏对全局观测数据的考虑, 无法准确感知战场环境。文献[8]从人工智能的角度解决空战决策问题, 利用专家知识设计空战决策系统指导飞行员做出最终决策。Mark Partridge, Michael X. 分别在 1989 年和 1990 年对借助人工智能的方法来提升战斗机的空战决策能力的方法进行了深入研究, 推动了空战决策方法的发展。专家系统提供了一个非常好的框架, 让研究人员能够将专家经验提炼出来以规则的形式编码到计算机中, 在工程上易于实现, 但其针对不同机型的可移植性不高, 需要根据特定机型的性能重新制定空战规则。空战环境往往存在着剧烈的动态变化, 需要考虑的情况非常多, 而专家系统在框架上难于通过学习方法进行自我扩展, 人工构建空战规则的工作量大。

综合上述分析, 影响无人机空战决策方法适用性的原因主要有以下 3 个方面:

(1) 求解决策模型计算量大, 难以实时获得空战决策结果;

(2) 需要精确的飞机数学模型, 在实际应用中难以获得;

(3) 受决策模型限制, 使得决策过程缺乏灵活性, 不利于在动态空战过程中构成对敌攻击条件。

为解决上述三个问题, 本文提出了基于多状态转移马尔科夫网络的无人机机动决策方法。机动决策模型以符合飞行员思维习惯的方式构建决策网络结构, 采用条件概率描述各个态势因素间的因果关系, 其参数可通过专家经验进行设定, 也可以使用客观数据进行学习, 具备通过机器学习进行自我扩展的能力。仿真结果表明采用概率关系描述空战问题的决策模型, 能够实现在不依赖精确飞机运动学模型的条件下进行有效的空战决策, 具备较好的环境适应性和实时性, 能够满足工程应用的需求。空战过程中机动决策网络能实时感知战场环境, 输出灵活的机动策略。

1 机动决策网络模型

综合分析影响空战决策的态势因素, 明确各因素间的因果关系, 建立基于多状态转移马尔科夫网络的无人机机动决策模型, 如图 1 所示。网络模型由 23 个节点组成, 其中根节点 DR 为决策节点, 输出决策结果, 存在 18 个状态; 网络输入层由 14 个观测节点构成, 分别目标方位角 ATA、目标相对高度 RH、目标距离 RD、雷达相对性能 RSP、我对敌方位角 AZE、目标接近速度 TAV、探测结果 TDR、敌机探测结果 EDR, 武器配置 WC、武器射程 RW、隐身性能 RSP, 机动约束 MC、威胁状态 TS、火控雷达状态 SFC。中间节点包括目标方位 TSB、相对运动趋势 TMT、相对探测性能 RDP、相对探测结果 RDR、空间态势 TSS、信息态势 INS、武器状态 WS、态势评估节点 SA。观

测节点用于感知战场信息,为保证网络更好的反映战场环境,网络模型中反映态势信息的观测节点被定义为连续观测节点。中间节点响应网络输入,提取出态势特征作为决策依据。决策网络模型利用有向弧定义网络节点间的连接关系,并通过相应的条件概率表描述节点间联系程度。网络节点在时间上存在相互联系,网络根据前一时刻的推理结果与当前时刻的证据信息更新节点概率,在本模型中节点 DR、SA、TSS、INS 前后两个时刻推理过程通过自身的状态转移概率联系。中间节点状态转移概率

的影响,使得推理过程呈现出逐层集聚的趋势,因此,中间节点带状态转移概率的机动决策模型将表现出快速收敛性。决策模型网络参数为条件概率表 CPT(Condition Probability Table),为实现对飞行员空战过程中决策思维的模拟,结合空战先验知识设定条件概率表。仿真表明决策网络在空战过程中能灵活的输出有效机动策略。

图 2 描述了空战过程中无人机和目标间的相对方位关系,其中 ATA 为方位角, TAE 为我对敌方位角, AA 为进入角。

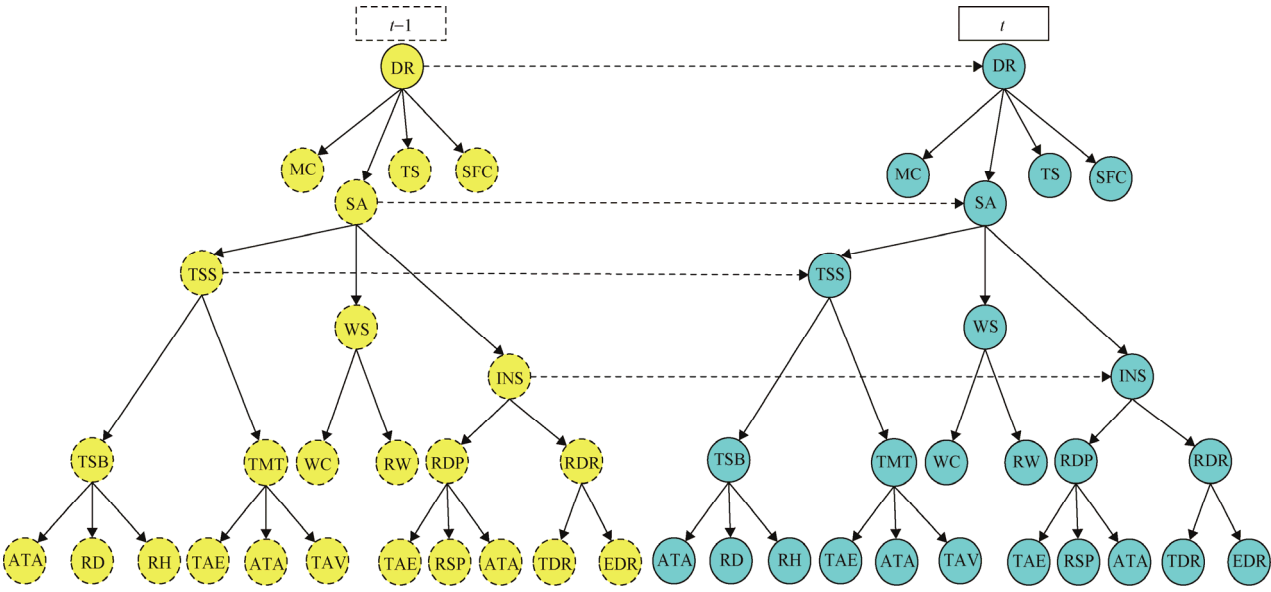


图 1 基于多状态转移马尔科夫模型的机动决策网络
Fig. 1 Maneuver decision-making network based on multi-state transition Markov model

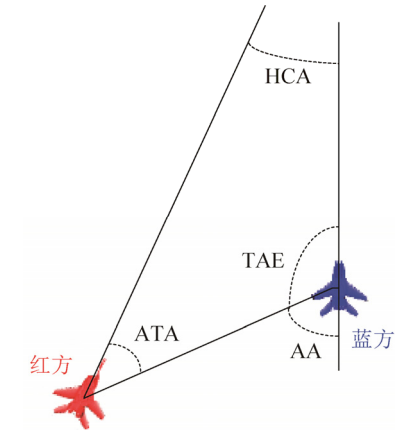


图 2 战场态势描述
Fig. 2 Description of battlefield situation

根据无人机在空战过程中需要执行的战术动

作需要,考虑现有无人机机动性能,定义 18 种基本机动作作为决策节点的状态集。决策节点具体状态集定义如表 1 所示,根据表中定义顺序对决策集进行编号,决策序号依次为 0~17。

表 1 决策节点状态集定义		
Tab. 1 Definition of decision node state set		
变量	描述	状态集
DR	决策结果	左转
		左转爬升
		左转下降
		右转
		右转爬升
		右转下降
		左急转
		左急转爬升
		左急转俯冲
		右急转
		右急转爬升
		右急转俯冲
		左急转掉头
		左急转掉头爬升
		左急转掉头俯冲
		右急转掉头
		右急转掉头爬升
		右急转掉头俯冲

2 决策模型推理

2.1 递归推理算法

单机对抗决策模型为标准 HMM 模型, 其推理过程可根据马尔科夫模型导出。结合网络结构可知, 决策推理过程可描述为在当前观测信息的激励下, 结合历史推理信息更新当前时刻的推理结果。这种特性使得网络推理过程受历史信息的影响, 能动态的反映决策推理过程, 推理过程描述如下, 其中, α 为归一化因子, e_t 为反映观测信息的证据节点。

$$P(DR_t | e_{1:t}) = \alpha P(e_t | DR_t) \left[\sum_{DR_{t-1}} P(DR_t | DR_{t-1}) P(DR_{t-1} | e_{1:t-1}) \right] \quad (1)$$

式(1)描述的是网络的递归推理过程, 网络以链路为单位进行概率推理, 根据条件独立性假设将各观测节点展开为独立推理链路。每条链路起点为决策节点, 终点对应相应观测节点。根据当前时刻观测信息更新概率 $P(DR_t | e_t)$, 其计算过程如式(2)所示。
 $P(DR_t | e_t) = P(DR_t | ATA_t, RH_t, \dots, MC_t, TS_t, SFC_t) =$
 $P(DR_t | ATA_t) P(DR_t | RH_t) P(DR_t | TAV_t)$
 $P(DR_t | RD_t) P(DR_t | RSP_t) P(DR_t | TAE_t)$
 $P(DR_t | TDR_t) P(DR_t | EDR_t) P(DR_t | WC_t)$
 $P(DR_t | RW_t) P(DR_t | MC_t) P(DR_t | TS_t)$
 $P(DR_t | SFC_t) \quad (2)$

综合各链路的推理结果根据式(2)的计算结果, 可求解条件概率分布 $P(e_t | DR_t)$ 。
 $P(e_t | DR_t) = P(ATA_t, RH_t, \dots, MC_t, TS_t, SFC_t | DR_t) =$
 $\frac{P(DR_t | e_t) P(ATA_t, RH_t, \dots, MC_t, TS_t, SFC_t)}{P(DR_t)} \quad (3)$

根据公式(1), 由式(2)、式(3)的计算结果可更新当前时刻的决策网络推理结果 $P(DR_t | e_{1:t})$ 。

$$P(DR_t | e_{1:t}) = \frac{P(e_t | DR_t) \sum_{DR_{t-1}} P(DR_t | DR_{t-1}) P(DR_{t-1} | e_{1:t-1})}{\sum_{DR_t} P(e_t | DR_t) \sum_{DR_{t-1}} P(DR_t | DR_{t-1}) P(DR_{t-1} | e_{1:t-1})} \quad (4)$$

计算式(1)~式(4)完成一个周期的递归推理后, 输出 DR 节点中概率最大的状态作为机动决策结果。

2.2 算法复杂度分析

根据网络结构与节点状态数量进行算法复杂度分析, 结果如表 2 所示。表 2 反映出一个推理过程中需要执行的乘法、加法、除法次数, 进行递归推理时, 一个周期需要进行多次迭代推理。因此一个周期需要执行的乘法、加法、除法次数为一次推理需执行的数学运算次数乘以递归次数。

表 2 链路复杂度分析
Tab. 2 Complexity analysis of each link

链路	乘法(次)	加法(次)	除法(次)
$P(DR ATA)$	168234	97500	98940
$P(DR RH)$	88776	43632	44136
$P(DR RD)$	88776	43632	44136
$P(DR RSP)$	26820	12540	12978
$P(DR TAE)$	110562	53868	54810
$P(DR TAV)$	84168	41328	41832
$P(DR TDR)$	26418	12552	12990
$P(DR EDR)$	26418	12552	12990
$P(DR WC)$	24624	11664	12096
$P(DR RW)$	24624	11664	12096
$P(DR MC)$	18	0	18
$P(DR TS)$	18	0	18
$P(DR SFC)$	18	0	18
共计	669484	340927	347068

分析算法的复杂度可知当算法迭代次数控制在 10^2 以内时算法的复杂度将维持在 10^7 级, 在现有计算机处理器的运算能力基础上, 推理算法能够在毫秒级完成一个周期的推理, 能很好的满足空战决策的实时性需求。同时, 算法的复杂度也为迭代次数预留了足够的裕度, 在保证算法实时性要求的前提下, 可以通过调节一个推理周期的迭代次数改善网络性能。

3 仿真与分析

为验证机动决策模型的有效性, 进行了空战对抗仿真分析。实验结果给出了空战对抗过程中的飞行轨迹、决策输出结果、决策转折点处前后周期的概率分布以及态势变化情况, 并根据仿真结果进行了详细分析。红蓝双方初始状态信息如表 3 所示。

表 3 初始状态信息
Tab. 3 Initial state information

阵营	经度 (度)	纬度 (度)	高度 (m)	速度 (m/s)	航向 (度)
蓝方	22.0	22.0	5000.0	200.0	0.0
红方	22.0	22.05	2000.0	200.0	0.0

图 3 给出在空战对抗过程中红蓝双方飞行轨迹。图 4 为空战过程中机动决策网络输出的机动策略, 决策网络感知战场态势变化在连续输出 7 个机动策略, 引导红方进行空间占位。决策输出说明机动决策网络在观测信息的激励下能实时输出灵活的机动策略。仿真结果表明, 目标运动状态变化后红方在决策网络的引导下, 能进行合理机动并保持对目标的攻击态势。

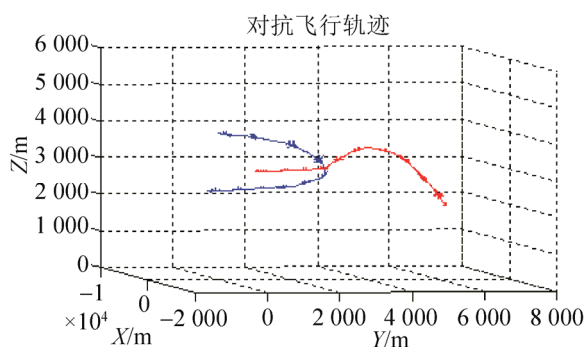


图 3 对抗过程飞行轨迹
Fig.3 Flight trajectory in the confrontation process

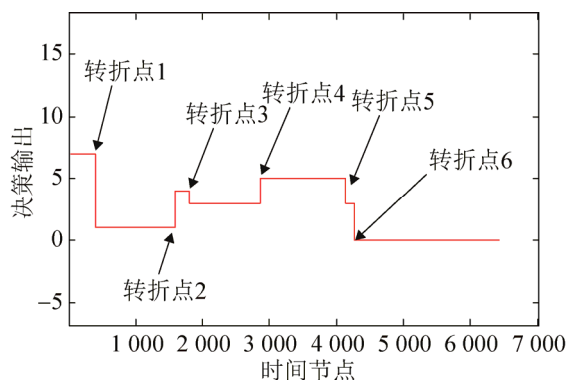


图 4 对抗过程决策输出
Fig.4 Decision output in the confrontation process

图 5~10 给出了在各决策转折点处前后推理周期决策节点概率分布情况。在空战过程中, 由于战场态势信息改变, 会导致决策网络输出发生改变。红方在决策网络的引导下, 不断调整战场态势以构

成对蓝方攻击条件, 对抗过程中态势信息发生 6 次改变, 网络决策输出也存在 6 个转折点。决策网络在状态转移概率的影响下推理过程表现出快速的收敛性, 分析仿真结果中转折节点处前后推理周期概率分布可知, 在观测信息的激励下决策网络经过 10 次迭代后概率迅速集中到与当前战场态势相一致的决策结果。结合表 2 中计算复杂度分析结果可知, 完成一个推理周期需要执行乘法 6694840 次, 加法 3409270 次, 除法 3470680 次。在现有处理器的计算能力上, 能在 40 ms 内完成一个推理周期, 网络推理过程完全满足空战决策实时性要求。仿真结果表明, 决策网络能响应观测信息激励, 在态势信息发生变化后及时输出与观测信息一致的机动决策结果。

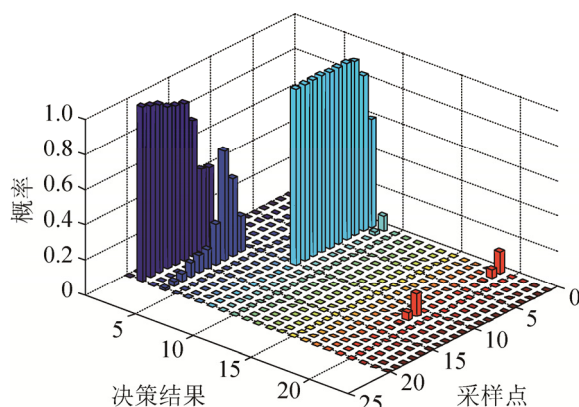


图 5 转折点 1 处前后周期决策输出概率分布
Fig. 5 Probability distribution of decision result at the switching point 1 during reasoning cycle before and after

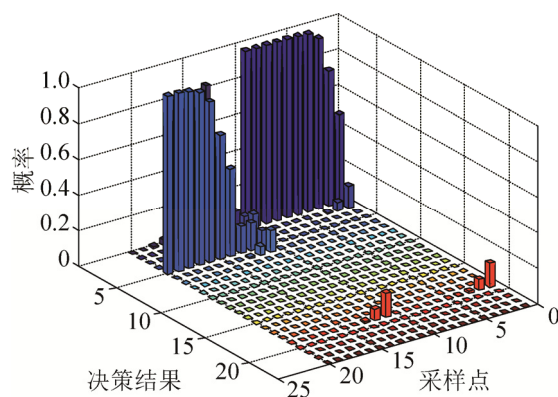


图 6 转折点 2 处前后周期决策输出概率分布
Fig. 6 Probability distribution of decision result at the switching point 2 during reasoning cycle before and after

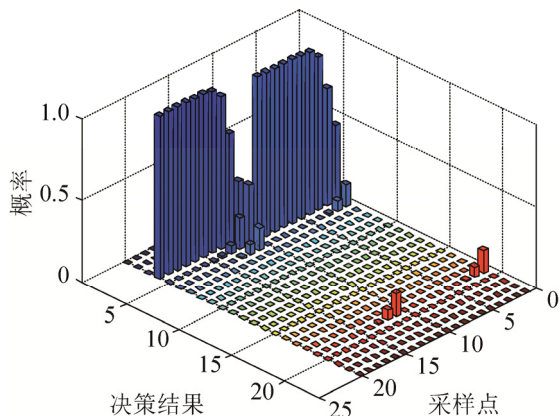


图 7 转折点 3 处前后周期决策输出概率分布
Fig. 7 Probability distribution of decision result at the switching point 3 during reasoning cycle before and after

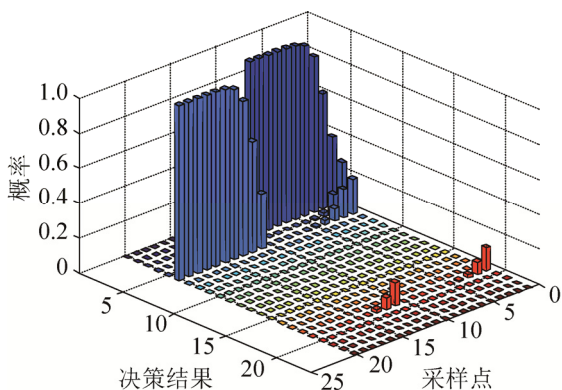


图 8 转折点 4 处前后周期决策输出概率分布
Fig. 8 Probability distribution of decision result at the switching point 4 during reasoning cycle before and after

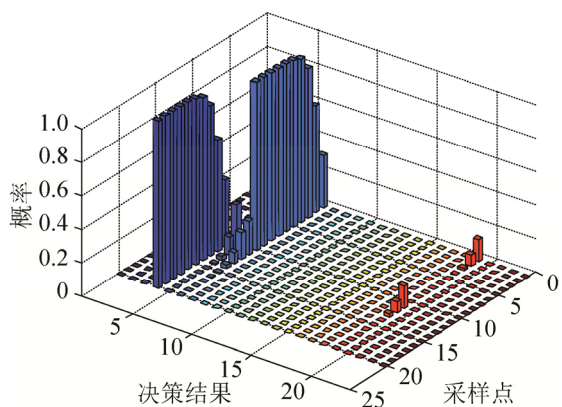


图 9 转折点 5 处前后周期决策输出概率分布
Fig. 9 Probability distribution of decision result at the switching point 5 during reasoning cycle before and after

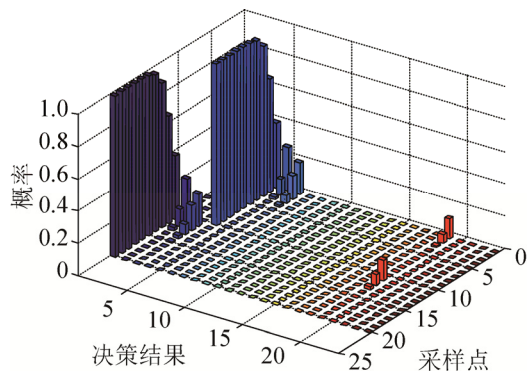


图 10 转折点 6 处前后周期决策输出概率分布
Fig. 10 Probability distribution of decision result at the switching point 6 during reasoning cycle before and after

图 11 反映了空战过程中博弈双方的角度变化情况。在空战过程中红方在决策网络的引导下, 方位角逐渐减小, 并最终维持在 0 度附近, 形成了有利攻击条件。随着机动决策的输出, 态势不断调整, 进入角也逐渐减小到 0 度。在空战对抗过程中红方逐渐占据了空间优势, 追逐过程中我对敌机的方位角保持在-180 度附近, 维持了对敌攻击态势。仿真结果验证了机动决策模型在空战格斗过程中的有效性和灵活性, 表明机动决策模型能够满足空战决策需要。

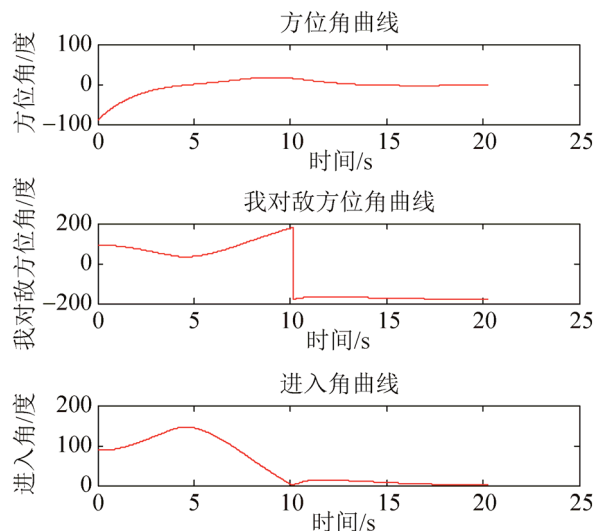


图 11 对抗过程中 ATA、TAE 及 AA 角度变化曲线
Fig. 11 Angle change curve of ATA, TAE and AA in the confrontation process

4 结论

仿真结果表明,基于多状态转移马尔科夫网络的机动决策模型能很好适应战场环境,能有效感知观测信息并做出正确响应,实现无人机在空战过程中的自主机动决策,并且提高了决策的灵活性和可靠性。同时,决策网络能在毫秒级完成一个推理周期,实时性完全满足空战决策需求。由于战场环境日益复杂,为了使得决策模型更好的拟合战场环境,保证无人机在复杂电磁环境下做出可靠、客观的机动决策,后期将研究根据仿真对抗结果进行网络参数学习的方法。

参考文献:

- [1] 车竞, 郑凤麒. 基于微分对策的追逃对抗仿真[J]. 飞行力学, 2014(4): 372-375.
CHE Jing, ZHENG Feng-qi. Simulation of pursuit-evasion resistance based on differential game[J]. Flight Dynamics, 2014(4): 372-375.
- [2] 钟友武, 杨凌宇, 柳嘉润. 基于智能微分对策的自主机动决策方法研究[J]. 飞行力学, 2008(6): 29-33.
ZHONG You-wu, YANG Ling-yu, LIU Jia-run. Potential-based Flow Analysis for Complex High-Lift Aircraft Configuration[J]. FLIGHT DYNAMICS, 2008(6): 29-33.
- [3] 傅莉, 谢福怀, 孟光磊. 基于滚动时域的无人机空战决策专家系统[J]. 北京航空航天大学学报, 2015, 41:

1994-1999.

FU Li, XIE Fu-huai, MENG Guang-lei. An UAV air-combat decision expert system based on receding horizon control[J]. Journal of Beijing University of Aeronautics and Astronautics, 2015, 41: 1994-1999.

- [4] 沈东, 魏瑞轩, 祁晓明. 基于 MTPM 和 DPM 的多无人机协同广域目标搜索滚动时域决策[J]. 自动化学报, 2014(7): 1391-1403.
SHENG Dong, WEI Rui-xuan, QI Xiao-ming. Receding Horizon Decision Method Based on MTPM and DPM for Multi-UAVs Cooperative Large Area Target Search[J]. ACTAAUTOMAT-ICA SINICA, 2014(7): 1391-1403.
- [5] James S, McGrew, Jonathan P, How. Air-Combat Strategy Using Approximate Dynamic Programming [D]. Journal of Guidance, Control and Dynamics(S0731-5090), 2010, 33(5): 128-136.
- [6] Virtanen K, Raivio T. Modeling pilot's sequential maneuvering decisions by a multistage influence diagram [J]. Journal of Guidance, Control, and Dynamics (S0731-5090), 2004, 27(4): 665-677.
- [7] Zhong Lin, Tong Ming'an, Zhong Wei. Sequential maneuvering decisions based on multi-stage influence diagram in air combat [J]. Journal of Systems Engineering and Electronics (S1004-4132), 2007, 3(18): 551-555.
- [8] E Y Rodin, Y Lirov, S Mitnik B G McElhaney and L. Wilbur. Artificial intelligence in air combat games [J]. Computers and Mathematics with Applications (S0898-1221), 1987, 13(1/3): 261-274.

(上接第 105 页)

- [7] 邱一卉, 米红. 基于随机森林和转导推理的特征提取方法[J]. 厦门大学学报(自然科学版), 2010, 49(3): 333-338.
Qiu Yihui, Mi Hong. Feature Extraction Method Based on Random Forest and Transduction[J]//Journal of Xiamen University(Natural Science), 2010, 49(3): 333-338.
- [8] 刘全, 翟建伟, 章宗长, 等. 深度强化学习综述[J]. 计算机学报, 2017: 1-28.
LIU Quan, ZHAI Jian-Wei, ZHANG Zong-Zhang, et al. A Survey on Deep Reinforcement Learning[J]. Chinese Journal of Computers(S0254-4164), 2017: 1-28.
- [9] 赵冬斌, 邵坤, 朱圆恒, 等. 深度强化学习综述: 兼论

计算机围棋的发展[J]. 控制理论与应用, 2016, 33(6): 701-717.

Zhao Dongbin, Shao Kun, Zhu Yuanheng, et al. Review of deep reinforcement learning and discussions on the development of computer Go[J]. Control Theory and Applications, 2016, 33(6): 701-717.

- [10] Arel I. Deep Reinforcement Learning as Foundation for Artificial General Intelligence[M]//Theoretical Foundations of Artificial General Intelligence. Atlantis Press, 2012: 89-102.
- [11] Hasselt H V, Guez A, Silver D. Deep Reinforcement Learning with Double Q-learning[J]. Computer Science(S1508-2806), 2015.