中国科学技术大学

University of Science and Technology of China

硕士学位论文

论文题目 基于博弈模型的无人机机动决策方法研究

作者姓名 殷书慧

学科专业 控制科学与工程

导师姓名 康宇 教授 赵云波 教授

完成时间 二〇二三年五月

中国科学技术大学

硕士学位论文

基于博弈模型的无人机机动决策方法研究

作者姓名： 殷书慧

学科专业： 控制科学与工程

导师姓名： 康宇 教授 赵云波 教授

完成时间： 二〇二三年五月二十五日

University of Science and Technology of China

A dissertation for master's degree

Research on UAV Maneuver Decision-Making Method Based on the Game Model

Author: Yin Shuhui

Speciality: Control Science and Engineering

Supervisors: Prof. Yu Kang, Prof. Yun-Bo Zhao

Finished time: May 25, 2023

中国科学技术大学学位论文原创性声明

本人声明所呈交的学位论文，是本人在导师指导下进行研究工作所取得的成果。除已特别加以标注和致谢的地方外，论文中不包含任何他人已经发表或撰写过的研究成果。与我一同工作的同志对本研究所做的贡献均已在论文中作了明确的说明。

作者签名：殷书慧　　签字日期：2023.5.25

中国科学技术大学学位论文授权使用声明

作为申请学位的条件之一，学位论文著作权拥有者授权中国科学技术大学拥有学位论文的部分使用权，即：学校有权按有关规定向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅，可以将学位论文编入《中国学位论文全文数据库》等有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。本人提交的电子文档的内容和纸质论文的内容相一致。

保密的学位论文在解密后也遵守此规定。

☑公开　□保密（___年）

作者签名：殷书慧　　导师签名：

签字日期：2023.5.25　　签字日期：2023.5.25

摘　要

无人机作为未来战场的核心力量对于夺取制空权起到至关重要的作用，其自主机动决策能力是发挥作战效能的关键所在。现有的空战决策方法诸如微分对策、专家系统等虽取得一定成果，但仍存在着搜索决策结果耗时长、适应性差等局限性。因此，如何在高动态、强竞争性的无人机对抗环境下进行快速准确的机动决策是本论文主要研究的问题。

本文以近距对抗为背景，以博弈理论为基础，以智能算法为工具，围绕基于博弈模型的无人机机动决策方法展开研究，具体研究工作如下：

（1）基于 F-16 机型无人机进行控制参数设计，并在此基础上对基本操纵动作库进行丰富和改进，设计了无人机的机动空间，构建了无人机的机动策略集。仿真实验分别对所设计的控制参数和机动空间进行测试，结果都满足设计需求。

（2）针对基本群智算法搜索决策结果计算效率低且容易陷入局部最优值的问题，提出了一种改进粒子群算法求解最优机动策略。首先，建立了无人机一对一动态博弈模型。然后，将博弈混合策略纳什均衡难于求解的问题转化为最优化问题进行搜索寻优，提出了一种改进的群体智能优化算法，通过粒子浓度的概率选择来控制种群多样性，以降低在优化收敛阶段陷入局部最优值的可能性。最后将其应用到无人机对抗机动决策中，设计了单机对抗仿真实验对比改进后算法的性能，结果表明改进粒子群算法提升了全局搜索效率和寻优精度，提高了无人机对抗机动决策中求解最优机动策略的计算效率和准确度。

（3）针对传统强化学习算法在处理高维状态输入时存在的维数爆炸问题，以及倾向于单方面最优化自身策略而不考虑对手策略影响的问题，提出了一种改进 DQN 算法生成有效对抗决策。首先，建立了无人机一对一场景下的二人零和马尔可夫博弈模型，并据此设计了一对一场景的基本状态空间、动作空间和奖励函数。然后，针对高维状态输入，引入深度神经网络拟合状态动作值函数，通过设置经验回放技巧并利用损失函数更新网络参数，提高了算法的收敛性和稳定性。其次，针对单方优化问题，引入博弈决策的极大极小均衡来生成针对性机动策略。最后，设计了单机对抗仿真实验对比改进后算法的性能，结果表明改进 DQN 算法可以通过自学习的方式在强竞争环境下生成更准确、更有效针对对手的机动决策，满足对抗实时性的同时具有更高的决策水平。

关键词：无人机；机动决策；博弈论；智能算法

ABSTRACT

As the core force of the future battlefield, UAV plays a vital role in seizing air supremacy, and its autonomous maneuver decision-making ability is the key to play the combat effectiveness. Although existing air combat decision-making methods such as differential game and expert system have made some achievements, they still have some limitations such as long time spent searching decision results and poor adaptability. Therefore, how to make fast and accurate maneuver decision in the highly dynamic and highly competitive UAV confrontation environment is the main problem of this paper.

With the background of close confrontation, game theory as the basis and intelligent algorithm as the tool, this paper studies the UAV maneuver decision-making method based on the game model. The specific research work is as follows:

(1) The control parameters were designed based on F-16 UAV, and the basic maneuvering action library was enriched and improved. The maneuvering space of UAV was designed, and the maneuvering strategy set of UAV was constructed. Simulation experiments are conducted to test the designed control parameters and maneuver space respectively, and the results meet the design requirements.

(2) Aiming at the problem that the basic swarm intelligence algorithm is inefficient and easy to fall into the local optimal value, an improved particle swarm optimization algorithm is proposed to solve the optimal maneuver strategy. Firstly, a one-to-one dynamic game model of UAV is established. Then, the difficult problem of Nash equilibrium is transformed into an optimization problem to search for optimization, and an improved swarm intelligence optimization algorithm is proposed to control the population diversity through the probability selection of particle concentration, so as to reduce the possibility of falling into the local optimal value in the optimization convergence stage. Finally, the algorithm is applied to UAV countermeasure maneuvering decision, and the performance of the improved algorithm is compared with the simulation experiment of single-machine countermeasure. The results show that the improved particle swarm optimization algorithm improves the global search efficiency and optimization accuracy, and improves the computational efficiency and accuracy of solving the optimal maneuver strategy in UAV countermeasure maneuvering decision.

(3) In order to solve the problem of dimension explosion of traditional reinforcement learning algorithm in processing high-dimensional state input and the problem of unilateral optimization of one's own strategy without considering the influence of

opponent's strategy, an improved DQN algorithm was proposed to generate effective antagonistic decisions. Firstly, a two-person zero-sum Markov game model is established in the one-to-one UAV scenario, and the basic state space, action space and reward function are designed accordingly. Then, for high-dimensional state input, a deep neural network is introduced to fit the state action value function, and the convergence and stability of the algorithm are improved by setting the experience playback technique and updating the network parameters by using the loss function. Secondly, for unilateral optimization problem, minimax equilibrium of game decision is introduced to generate targeted maneuvering strategies. Finally, the performance of the improved algorithm is compared by the simulation experiment of single-machine confrontation. The results show that the improved DQN algorithm can generate more accurate and effective maneuvering decisions against the opponent in the strong competitive environment through self-learning, which meets the real-time performance of the confrontation and has a higher decision-making level.

Key Words: UAV; Maneuver Decision-Making; Game Theory; Intelligent Algorithm

目　录

插 图 清 单

表 格 清 单

第 1 章　绪　　论

1.1　研究背景及意义

现代化军事作战是陆海空天电五维一体化的信息战[1]，在联合作战中，争夺高位制空权是作战的首要任务也是战场上的胜负关键[2]，因此提高空中作战能力是现代化战争发展的必然趋势。

自无人作战飞机[3]（Unmanned Aerial Vehicle，UAV）首次在美对越战争期间登台亮相以来，凭借着其高隐蔽性、强生存力等特点在战场上发挥着极其重要的作用，承担起侦察预警、火力打击、欺骗牵制、电子干扰掩护等多方面任务[4]，成为空中作战的新兴力量以及争夺制空权不可忽视的角色。相较于有人机，无人机的使用极大避免了飞行员的伤亡和武器损失[5]，提高了空中持久作战能力和获胜概率，因此受到了世界上各军事强国的高度重视，对无人机的研究投入了大量的人力和物力。

美国国防部长办公室在 2005 年 8 月发布了《无人机飞行器系统路线图 2005-2030》[6]，如图1.1所示，其指出在 2025 年之后将实现无人机的全自主作战能力，而态势感知、任务规划和自主决策是体现其自主作战能力的三个主要方面[7]。其中，无人机的自主决策能力位居首位，它是无人机的大脑与灵魂，是实现其自主化的关键核心[8]。同时美国相关研究也指出，未来将会有 40% 的空战发生在视距内[9]，由于电子对抗技术的应用有效降低了中、远距空空导弹的攻击效能，隐身技术的应用使得目前的探测技术限制了隐身无人机之间的超视距作战，同时无人机机动突防很容易打破超视距空战的条件使敌对双方不得不近距离对抗。在近距空战中双方攻防对抗激烈，战场环境变化剧烈且不确定性强，导致交战方式更为灵活，因此，提高无人机的机动决策能力，进行无人机机动决策方法的研究对于掌握近距对抗主动权有着极其重要的价值和意义。

针对无人机对抗机动决策问题，敌对双方具有各自不同的战场目标导致彼此之间存在较大的利益冲突，而博弈论是研究具有利益冲突的决策者之间进行策略交互的数学模型[10]，正好为解决此类问题提供了一种行之有效的方法。博弈论早已在我国历史上得到灵活的应用，从田忌赛马到《孙子兵法》正是博弈思想的体现[11]，而在现代博弈论也被广泛地运用于军事、经济、政治等领域[12]。

由于空战环境中往往包含着复杂的强动态因素，无人机的对抗模型也更多是非线性的，导致博弈论的应用面临着极大挑战；不仅如此，由于战场态势变化剧烈，无人机保持高速飞行，当涉及到多方利益冲突者时，状态特征连续多维，使得博弈策略的求解变得十分困难[13]。使用传统方法对空战博弈的最优策略进行求解存在着计算复杂度高、难以满足实时性且难于实现的问题，因此针对无人

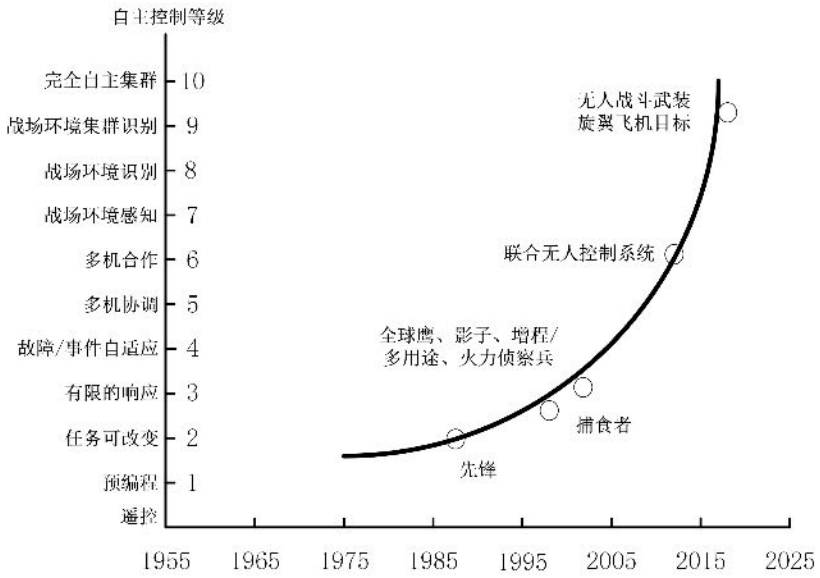


图 1.1　美国无人机发展路线图

机空战对抗这类大规模博弈场景问题需要探索更加高效适用的求解方法。

随着智能算法的发展，将空战对抗中的最优策略求解转化为优化问题，依靠智能算法寻优则为无人机对抗机动策略求解带来一种新途径。本文主要针对基于博弈模型的无人机机动决策方法进行研究，利用博弈论的思想对无人机单机对抗问题进行建模，并使用改进的群智算法和深度强化学习算法对博弈问题的求解进行理论和方法上的尝试，对无人机对抗机动决策方法的探索具有重要的实际意义。

1.2　国内外研究现状

1.2.1　无人作战飞机研究现状

无人作战飞机从过去主要是执行战略侦察任务的作战辅助工具快速升级为能实施防控压制和纵深打击的主要作战装备[14]，是现代空中武器系统发展的必然趋势。伴随着人工智能等先进高新技术的迅猛发展也赋予了无人机新的“生命”，使其智能化水平达到了一个新的高度。

近年来，以美国牵头的无人机技术强国取得了一系列诸如“全球鹰”、“死神”等标志性成果[15]。起初无人机仅作为侦察使用，极具代表性的就是“全球鹰”无人机[16]（如图1.2）。作为美军最先进的战略侦察工具，其装备有高分辨率红外传感系统和合成孔径雷达，可获取精确到 0.3 米的定点侦察照片（如图1.3）。在 1998 年与阿富汗战争打响后，“全球鹰”执行 50 余次作战任务，提供了 1.5 万多张敌军的目标侦察图像[17]。后在 2003 年与伊拉克战争中，仅使用两架“全球鹰”承担了 452 次情报、监视行动，极大提升了美军实战侦察能力。

图 1.2　“全球鹰”无人机

图 1.3　机载合成孔径雷达成像图

在侦察无人机的基础上，打击地面目标转变为无人作战飞机的首要核心。MQ-9“死神”无人机[18]（如图1.4）作为一种新型极具杀伤力的察打一体化无人机，具有多任务、长续航、机动性能优越等特点。美国空军于 2007 年组建了专门的“死神”无人机工作组，开始研究战术进行实战演练，在 2008 年 3 月机载 AGM-114 海尔法空地导弹完成了对阿富汗境内 16 个目标的打击任务[18]。MQ-9“死神”无人机凭借其卓越的战术性能成为美军执行定点打击任务的空中平台。

图 1.4　“死神”无人机

图 1.5　“雷神”无人机

除美国外，其他军事强国在无人机相关技术方面也有着显著成果。2013 年，英国 BAE 系统公司研发的新一代隐形无人作战飞机“雷神”[19]（如图1.5）完成了首次飞行，其飞翼布局设计非常有利于隐身性能，使地面雷达系统几乎无法进行追踪。“雷神”无人机的研发使英国的空中作战能力迈上了一个新的台阶。

2019 年，俄罗斯将“猎户座”无人机[20]（如图1.6）投入了实战，此款高空长航时无人机弥补了俄军这个量级军用无人机的空缺，在叙利亚反恐实战中挂载 OFAB-100 航空炸弹完成了 40 多次打击恐怖分子目标，展现出优异的对地对空打击能力，扩大了其国际影响力。

我国从 20 世纪 60 年代起就自主开展无人作战飞机研究工作，先后研制推

图 1.6　“猎户座”无人机

图 1.7　“彩虹”无人机

出了多型“彩虹”无人机[21]（如图1.7）。彩虹-4 中空长航时无人机实战经验丰富，可挂载 4 到 6 枚精确制导武器实施侦察监视、反恐作战等军事任务，此外彩虹-4 系列还可应用于气象勘测、海洋监测等民用领域。彩虹-5 高空长航时无人机可执行全天候精确定位，对敌实施区域性干扰，为反恐维稳提供情报保障和攻击手段。彩虹-7 隐身无人机可在高危环境下压制敌方防空火力，执行对高价值目标发动打击，极大提升信息化作战效能，并且满足未来对称性作战对高端隐身无人机的需求。

图 1.8　“翼龙”无人机

图 1.9　“无侦-8”无人机

另外，在 70 周年大阅兵中也展示了“翼龙”[22]（如图1.8）和“无侦-8”高空高速[22]（如图1.9）等多款无人机设备。中航工业成都飞机设计研究院自主研发的“翼龙-1E”标志着中国成为继美国之后第二个采用全复合材料制造的国家，并且其各项性能指标可与美国的“MQ-1C”灰鹰（如图1.10）无人机相提并论。“无侦-8”采用两台液体火箭发动机使其飞行速度可达到四倍音速，打破了由美国“黑鸟”SR-71（如图1.11）保持的军用飞机飞行速度，并且其外表设计光滑，保证高空高速的同时可以具备良好的隐身性能，在执行穿透侦察任务中展现出强大的生存能力。

图 1.10　“MQ-1C”灰鹰无人机

图 1.11　“SR-71”黑鸟无人机

虽然我国在无人机领域的研究起步较晚，但中航工业等国防单位、研究所和清华、西工大、北航、南航等工科院校一直在自主创新的道路上不断探索，我国无人机相关技术已得到了空前发展。由于国产航空发动机、续航电池等技术无法获得突破性进展，我国无人作战飞机在飞行速度、有效载荷、作战水平等方面与以美国为首的军事强国之间仍然存在一定差距。不过在国家、研究所技术人员和国内学者的不断摸索中，相信未来我国无人机技术将达到国际超一流水平。

1.2.2　空战机动决策方法研究现状

自主机动决策能力是无人作战飞机发挥作战效能的关键所在。空战自主机动决策是根据当前战场态势信息独立自主地生成飞行控制指令来模拟飞行员对飞机进行机动操纵的过程。决策系统的输入是由机载传感器和多机协同作战的数据链提供的空战态势信息，对当前战场态势进行评估后进而生成对己方有利的策略，最后输出机动动作对飞机进行操纵，占据空中战场的位置和角度优势以完成毁伤敌机的任务[23]。

空战机动决策主要有三方面的应用[24]：

（1）通过空战仿真来评估战术战法和作战效能，为开展真实场景的空战训练提供基础；

（2）作为有人机的飞行助手，通过辅助决策可以减轻飞行员的负担，提高战斗力和作战效率；

（3）作为无人机的大脑控制其在复杂空战环境下实现作战目标。

关于空战机动决策的研究起源于 20 世纪 60 年代，由于战场环境的高动态性及复杂性，探索有效的机动决策方法已经成为国内外学者的热点课题。从传统的基于博弈理论的方法，到基于优化理论的方法，再到基于人工智能技术的方法（如图1.12所示），目前对于无人机空战自主机动决策的研究越来越趋向于新型智能方法[25]。

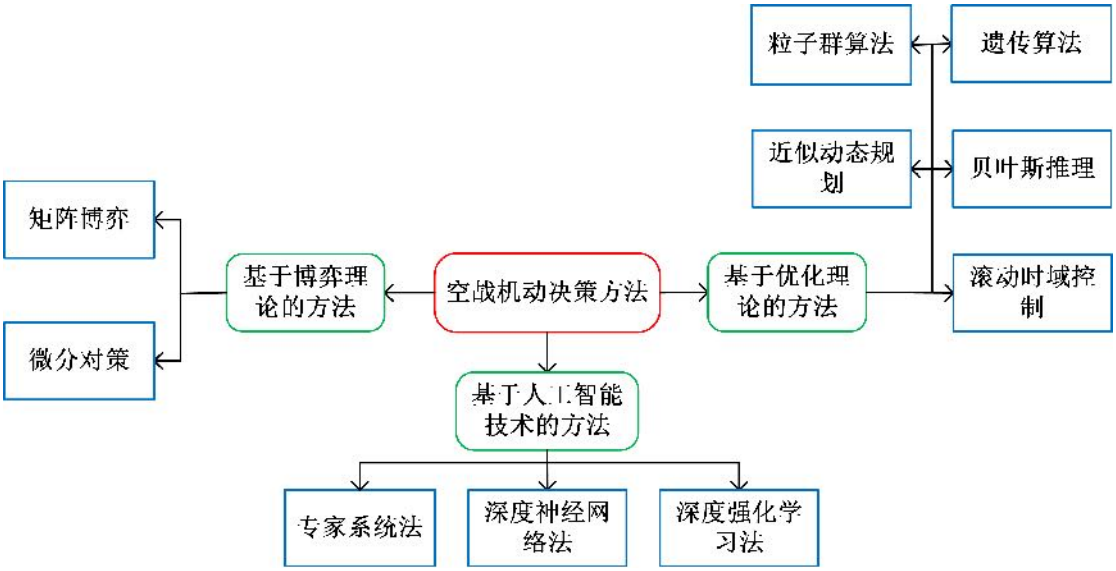


图 1.12　部分空战机动决策方法

（1）基于博弈理论的方法

基于博弈论的空战机动决策方法主要有矩阵博弈法[26-28]和微分对策法[29]。

矩阵博弈法是将空战过程中敌我双方可采取的典型机动动作组合成不同的对抗方式，根据决策方案得到支付矩阵[28]，然后对策略集中的每一种对抗方式计算相应的支付值进行量化打分，计算出的各对抗方式得分总数最高者作为最后的机动轨迹。最基本的 7 种机动动作有定速平飞、最大切向加速平飞、最大切向减速平飞、最大法向负载爬升、最大法向负载下降、最大法向载荷左转和最大法向载荷右转。空战对抗中的矩阵博弈方法是求解敌我双方支付矩阵的鞍点的过程[26]，在该点处敌我双方都不能通过改变自己的策略来增加支付函数值。矩阵博弈法假定了空战对抗中敌我双方不会侥幸采取具有欺骗性的策略，使该方法得到的策略过于谨慎，无法趁机利用有利机会获得己方优势，与实际复杂环境下的空战过程不符。

微分对策法是把敌我双方决策转化为双边极值问题后进行求解得到最优控制决策，常用于解决空战对抗中的追逃博弈问题[29]。该方法的性能函数需要依据实际场景凭借经验因而设定困难，在实际高动态环境下的空战过程中微分对策法的计算量庞大，对微分对策模型的简化常常因为失真而导致解算结果误差较大难以在实际空战中应用。

由此可见，传统方法虽然可以解决空战决策中的许多问题，但也存在诸多缺陷，例如实际复杂空战场景建模的复杂性和对抗强动态的空战决策问题难以求得纳什均衡解析解的困难性。

（2）基于优化理论的方法

无人机空战机动决策问题也经常被建模为多目标优化问题，可以使用遗传

算法[30-32]（Genetic Algorithm, GA）、粒子群优化算法[33-34]（Particle Swarm Optimization, PSO）、滚动时域控制（Receding Horizon Control, RHC）[35-36] 等智能优化算法进行求解。

遗传算法是一种基于自然选择的进化算法，本质上是优胜劣汰，具有较强的适应力且占用资源少。文献 [30] 提出将 GA 应用于多机协同空战决策中，很好地完成了寻优任务，建立的优化指标基本反映出无人机对抗的实际情形。文献 [31] 提出了与模糊推理器结合的模糊遗传算法对先进战机协同空战决策求解最优决策集。文献 [32] 设计了强化遗传算法对传统 GA 没有显式目标函数进行改进，可以在不确定环境下生成合理的对抗决策序列。

粒子群优化算法是 Eberhart 等人根据对鸟群捕食行为的研究利用群体智能建立的简化模型[33]，通过群体中个体的合作和共享信息来寻求最优策略。文献 [34] 设计了一种启发粒子群算法，建立空战人工势场函数，根据对抗双方人工势场分布利用该算法完成机动决策，克服了 PSO 算法容易陷入局部最优的缺点。

滚动时域控制法将全局优化问题的求解转化为滚动进行的若干个局部最优控制问题的求解[35]，从而降低了问题的复杂性。文献 [36] 把无人机对抗的整个过程分割成无数个时间域，在每个有限的时域里，把无人机机动决策视为起始状态不同的专家系统对最优控制模型进行求解，以完成该时域无人机的机动动作，叠加为最终的机动轨迹，最后完成有效迅速的机动决策。

由此可见，智能优化算法虽为最优策略的求解提供了一种行之有效的方法，但也存在着在寻优过程中容易陷入局部最优解的可能性，因此需要提升算法的全局搜索寻优能力和寻优精度。

（3）基于人工智能技术的方法

基于人工智能技术的空战机动决策主要有三种代表方法：专家系统法[37-39]、深度神经网络法[40-41] 和深度强化学习法[42-45]。

专家系统法是一种基于 IF-ELSE-THEN 的规则表示法[37]，依据领域专家的经验知识模拟专家完成机动决策。知识库模块有该领域专家的先验知识，无人机空战进行智能决策时，推理机模块将输入的空战态势信息与知识库模块中的规则相匹配，进而输出在当前战场态势下专家从动作库中选择可能性最大的动作[38]。该方法在实际无人机空战决策上的应用存在无法规避的缺陷，首先建立完备的知识库相当困难，空战过程中如果出现未存储的战场态势信息，会导致系统失效。其次知识库模块中的规则依据了制定者的偏好和经验，决策系统得到的策略能够轻易被敌方推测。因此许多科研人员将专家系统法与其他方法相结合来弥补该决策方法的缺陷。文献 [38] 提出了一种进化式的专家系统树方法，将遗传算法和专家系统结合起来研究空战机动决策问题，解决了原有方法对于非预期情况实用性较差的问题。文献 [39] 提出了一种基于滚动时域控制的算法

对专家系统在无人机机动决策中适应性较差的问题进行改进，通过求解专家系统滚动时域最优控制模型使无人机在原有方法失效的情形下仍然能够快速反应，完成有效机动动作实现敌我双方态势逆转。

深度神经网络法是一种从数据中学习的方法。基于深度神经网络的空战机动决策模型包括训练和决策两个部分，完成网络结构设计后，根据空战模型生成网络参数，在训练过程中策略神经网络的参数不断更新直至损失函数值最小或小于预期值。文献 [40] 将无人机的初始速度、滚转角、航迹倾角和下一时刻采取的目标滚转角作为网络输入，飞行固定时间后记录当前时刻无人机的航迹偏角和倾角，作为网络输出。通过均匀采样网络输入飞行仿真后得到相应输出，获取大量飞行样本，利用所获样本训练深度神经网络，预测未来状态，选出目标函数值最大的动作。文献 [41] 以 DCS World 空战游戏作为交互平台，飞行员模拟操纵飞机产生相关空战数据后，采用循环神经网络模型根据输入态势预测输出动作，拟合飞行员的智能决策行为。由于采用人脑结构工作原理，网络输出与真实飞行员的决策模式相似，即使在战场态势部分信息缺失的情况下仍能输出合理的机动决策。然而深度神经网络法在训练过程中需要大量的空战样本数据才能得到良好的预期效果，在真实场景中要获得大量由人类对抗的空战样本数据非常困难，使得该方法具有一定的局限性。

深度强化学习法是强化学习和深度神经网络的结合。强化学习是智能体采用试错的方式和环境交互，通过计算在当前状态下执行动作后的累积奖励值来评估机动选择的结果。因此强化学习不仅考虑了在当前状态下执行下一动作产生的回报，也考虑了机动动作的远期收益，使得决策选择的动作在当前状态下最为合理且对环境有较强的适应性[42]。比起神经网络法，强化学习不需要由人类对抗获得的空战样本数据，智能体在学习的过程中不断探索环境以生成样本数据，智能体可以自行学习数据中的特征信息，具有较强的适应性。同时利用深度神经网络的非线性拟合能力，突破有限维状态输入的局限性，解决了传统强化学习在空战高维状态空间下遭遇维数爆炸的问题，使得深度强化学习方法得到更为广泛的应用[43]。文献 [44] 应用 DQN 方法求解空战问题，在没有飞行员经验的情况下找到了最优策略，与基于搜索的算法相比，该方法具有较好的短期精确操纵能力和长期规划能力。文献 [45] 将 LSTM（Long-Short Term Memory）网络引入到 DQN 算法，能够使我方战机在空战对抗中避开敌人的威胁，并利用自身优势对目标进行攻击。

由此可见，基于人工智能的方法虽然获得了较符合空战实际场景的结果，但却只考虑了自身策略的单方优化，并没有实际考虑到对手策略对战场局势造成的影响，而无人机空战对抗过程涉及到双方或多方的策略交互，仅考虑单方的控制明显不太合理。

上文具体分析了三种类型的空战机动决策方法和国内外研究的成果，接下来将梳理各个发展阶段典型的工程实践项目，以清晰地展示空战自主机动决策技术的发展脉络，如图1.13所示。

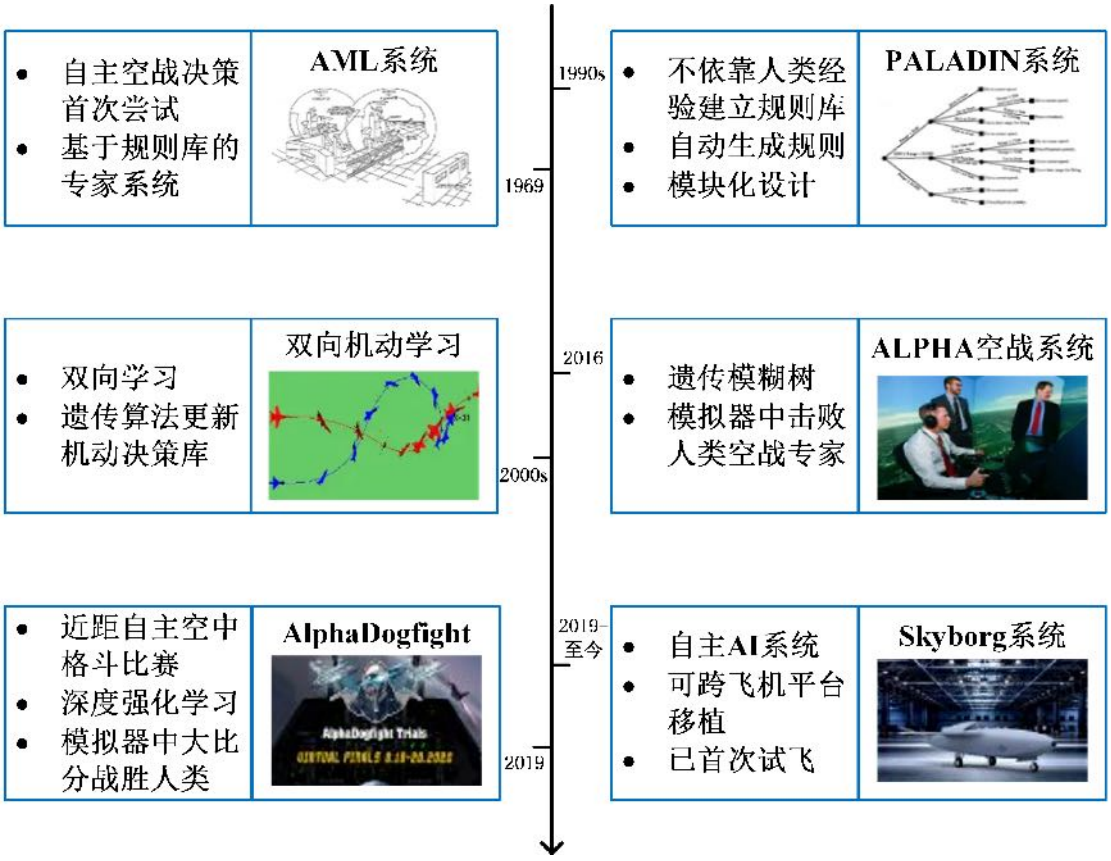


图 1.13　空战决策技术发展路线

（1）AML 系统

自 1969 年以来，Burgin 和其他研究人员在 NASA 兰利研究中心开发了一款自适应机动逻辑（Adaptive Maneuvering Logic, AML）的机动决策软件[46]，该软件以基于 IF-ELSE-THEN 逻辑的专家系统为核心。AML 不仅能模拟敌机与飞行员的实时空战对抗，还能利用仿真对战中的两架战机完成对战机和武器系统的性能参数分析。

AML 是首次对自主空战决策技术的系统性研究，NASA 指出空战机动决策具有高实时性和不确定性，很难给出精确的解法，而有经验的飞行员熟悉空战的作战态势和决策要领，可以专家系统为基础实现无人机在空战对抗中的自主决策能力，但受限于当时的技术水平，存在许多缺陷。

（2）PALADIN 系统

20 世纪 90 年代，新性能战机的服役促使空战环境迅速变化，NASA 在 AML 系统的基础上继而开发了 TGRES（Tactical Guidence Research and Evaluation System）[47]，该系统由战术决策生成器、战术机动模拟器和微分机动模拟器构成。

PALADIN 系统是以 AML 系统为基础研发的战术决策生成器。两者最主要的区别是，PALADIN 不需要根据飞行员的经验构建知识库，仅依靠战机本身数据和空战仿真的对抗结果，因此给实战经验空白的新性能战机提供机动决策支持。此外，PALADIN 采用模块化思想改进知识库使得运算速度大大提升，最大的创新点在于开发了除纯机动决策外战机在使用武器和载荷调度方面的决策。

（3）双边对抗学习系统

双边对抗学习系统[48] 是波音公司和西英格兰大学联合开发的新型战斗机机动规则决策系统。双边对抗学习指机动对抗中的双方都可以通过自博弈来更新优化自身的决策，从而摆脱飞行员知识的限制，能够在动态空战环境中生成全新的空中对抗新战术。该系统基于学习分类器系统方法，采用一对一空战数字仿真模型和遗传算法，为 X-31 实验战斗机开发出有效的机动战略。双边对抗学习系统并非针对特定的场景，因此具有较好的鲁棒性，可以适应不同的环境。

（4）ALPHA 空战系统

2016 年，辛辛那提大学和美国空军研究实验室共同研发的 ALPHA 空战系统在空战模拟器中打败了已退役的空战专家 Gnee Lee 上校[49]。其使用的遗传模糊树在解决复杂空战问题时展现出强大的能力，通过训练具有不同程度连通性的模糊推理系统，创建一组有效的规则，在空战对抗中产生确定性控制指令进行实时决策。作为应用 AI 技术求解空战对抗自主决策问题的标志性成果，ALPHA 系统很好地应用演化计算解决无人机空战问题，针对策略参数探究进行了不懈的努力。

（5）AlphaDogfight

在美国 DARPA（Defense Advanced Research Projects Agency）的“阿尔法狗斗”模拟近距离空中格斗比赛中，AI 五次连胜击败了现役 F-16 战斗机的人类飞行员，验证了 AI“狗斗”算法在虚拟空战中胜过人类的能力[50]。该智能决策系统使用了深度强化学习和多智能体分布式学习框架，通过回放的对战试验数据可以看出，AI 获胜的关键在于其具备精确的对准能力和快速的机动控制能力。

（6）Skyborg 系统

2019 年 3 月，美国空军战略发展规划与实验办公室提出要研发一个可以满足即使作战需求的自主 AI 系统 Skyborg[51]。Skyborg 采用开放的人工智能软件架构，能够修改模块化任务硬件，允许作战飞行员模块化地调整无人机的自主能力和独立的传感器，该系统可以根据当前战场态势自主构建能够实现任务目标的决策方案。2021 年 4 月，美国空军在佛罗里达州上空进行了首架 Skyborg 原型机的首次试飞，长达两小时十分钟。试飞过程中，搭载 Skyborg 自主核心系统的无人机成功对导航命令做出实时反应，并展示了协调机动，证明该系统能够安全运行。

综上所述，对几种空战机动决策方法和工程项目的调研可知，面对无人机空战决策这一类信息量大、搜索空间维度高，优化目标复杂的强动态对抗场景，如何对无人机机动决策问题建立更加精确的数学模型，以及寻求高效合理且实时性高的求解算法，已经成为国内外研究人员共同追求的目标。

1.2.3　博弈论研究现状

博弈论（Game Theory），又称为对策论，是运筹学的一个分支，旨在研究具有竞争或对抗性质的决策者之间的策略互动[10]。博弈对局中的决策者各自具有不同的目标，其为了使自己的利益最大化则必须考虑对手可能采取的各种行动。

博弈论最早可追溯到中国古代的《孙子兵法》[11]，“知己知彼，百战不殆”则是博弈思想的体现，其不仅是一部军事著作，更是一部博弈论著作。该理论的数学研究开始于 1944 年冯·诺依曼和奥斯卡·摩根斯坦的著作《Theory of Games and Economic Behavior》，标志着一个独立学科的初步形成。1950 年，纳什发表了关于非合作博弈的论文，首次提出了纳什均衡的概念并证明了“纳什定理”，泽尔腾和海萨尔进一步发展了非合作博弈的均衡分析理论，此后沙普利和舒贝克又建立了合作博弈解的概念，使得博弈论的体系更加完善[52]。

博弈论按照研究问题性质的不同，可以有不同的分类，根据对策方式可分为合作博弈与非合作博弈，前者是研究决策者达成合作时如何分配收益的问题，而后者是研究在竞争环境下具有利益冲突的决策者如何选择使自己收益最大的策略；根据博弈状态可分为静态博弈和动态博弈，静态博弈指的是不论决策者是否同时行动，后面的不知道前面采取了什么动作，而动态博弈的决策者先后按顺序行动，但后面可以观察到前面采取了什么动作[53]。

在诸多博弈模型中，二人零和矩阵博弈是最简单的模型也最具代表性，其本质是对微分对策的离散[54]，文献 [29] 将该模型应用于两机空战格斗求得双方均为最优的数值解；文献 [55] 提出随机博弈的基本概念，研究了马尔可夫决策过程中具有状态概率转移的动态博弈[55]，同时证明了零和随机博弈中存在一致值；文献 [56] 研究了一般和随机博弈中存在的多个纳什均衡点[56]。文献 [57] 提出了基于马尔可夫的平均场随机博弈，为涉及多个决策者的博弈提供了新思路[57]。

目前博弈论广泛应用于计算机科学、政治、经济等领域，特别是在军事战略方面发挥着不可忽视的作用。作为解决复杂对抗冲突性问题的有利工具，将其应用于无人机空战对抗中有着实际的意义与价值，使用纳什均衡的概念，可为对抗强动态空战场景下无人机最优机动策略的求解提供一种有效的途径，成为国内外军事专家重点研究的内容。

1.2.4　博弈论在空战机动决策存在的主要问题

人类的社会活动都离不开策略互动，因此博弈论的研究也取得了迅速且实质性的进展，其在各个不同的领域都发挥着至关重要的作用。但博弈论在具体场景的应用也相继出现了许多问题，将其应用于无人机空战机动决策仍然面临着许多挑战。

首先，实际复杂空战场景中建模的复杂性。敌对双方存在复杂的利益冲突关系，各方之间采取的机动策略彼此耦合，相互作用影响，对博弈对局其中一方的策略质量评估都要考虑到对手所选取的策略给己方带来的影响，因此需要全面考虑并构建系统、客观且精确的博弈数学模型。

其次，存储评估值空间的复杂度。博弈对局中每一个决策者都要对其他局中人采取的联合动作进行评估，对于多无人机参与对抗的空战场景，存储评估值的空间将会呈现指数型增长，因此解决维度爆炸问题对博弈的求解也至关重要。

再次，博弈纳什均衡解的存在性。传统的博弈模型虽然证明了纳什均衡解的存在性，但其基于的模型非常简单，对于复杂的实际场景，需要结合其他理论和智能算法来证明博弈是否有解。

最后，纳什均衡策略求解的复杂性。如何求解纳什均衡点的解析解是博弈论的主要难点所在，传统算法仅适用于简单的博弈模型求解，对于无人机空战机动决策这类复杂的问题则需寻找智能算法来获得更加精确的最优策略，同时战场环境变化剧烈，双方攻防对抗激烈，因此求解算法的速度和效率也是需要考虑的一个方面。

1.3　本文研究内容与组织架构

针对军事领域对无人机自主机动决策能力的迫切需求，本文以 F-16 机型无人机为研究对象，主要利用基于优化理论的群智算法和基于人工智能技术的深度强化学习算法对基于博弈模型的无人机机动决策方法进行研究，全文共包含五章内容，论文组织结构如下图1.14所示。

具体章节安排如下：

第一章：绪论。首先介绍了本论文的研究背景和意义，分析了近距对抗背景下无人机自主机动决策研究的重要性。其次，对无人作战飞机、空战机动决策方法的国内外研究现状及相关成果进行了调研和总结。然后简要概述了博弈论的研究现状，指出了其应用于空战机动决策存在的主要问题，最后给出了本论文的研究内容与组织架构。

第二章：基于无人机六自由度模型的机动空间设计。首先介绍了 F-16 机型无人机的六自由度模型，设计其控制参数，并在此基础上对基本操纵动作库进行

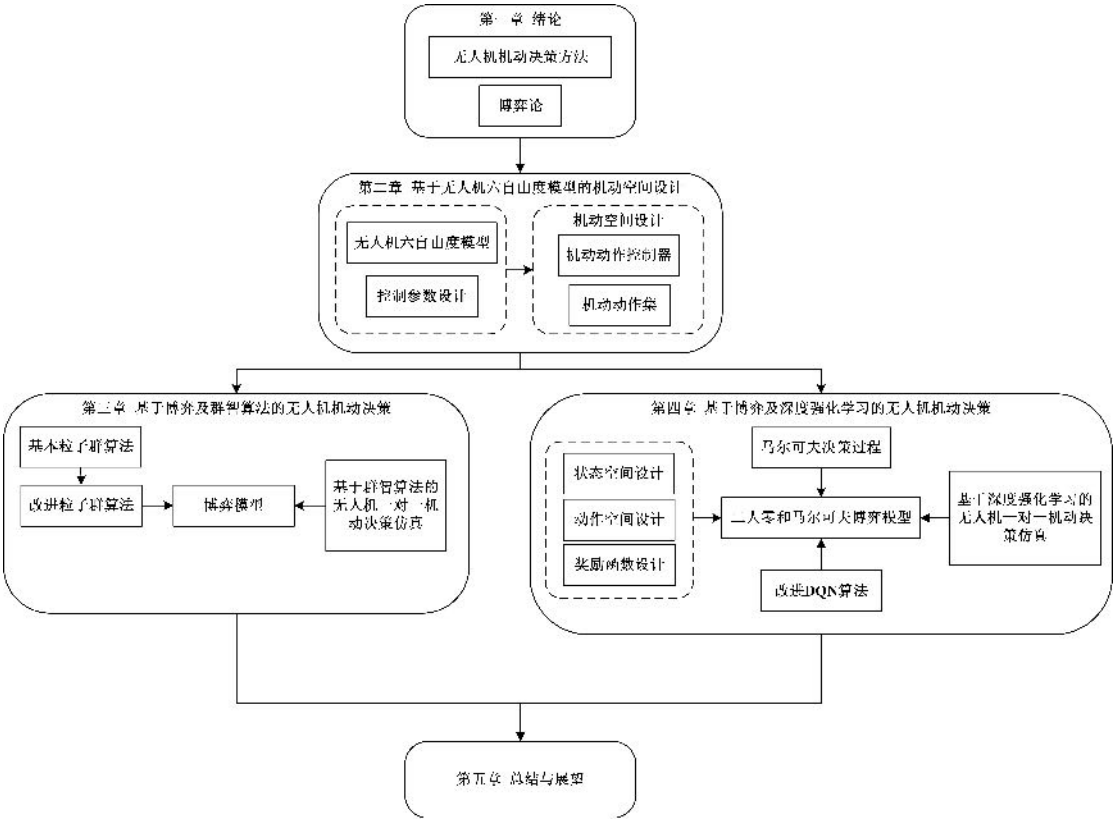


图 1.14　论文结构框图

丰富和改进，设计了无人机的机动空间，以完成无人机机动策略集的构建，最后对 F-16 无人机模型的控制参数和机动空间进行仿真测试，为后续空战机动决策方法的研究打好基础。

第三章：基于博弈及群智算法的无人机机动决策。首先建立了无人机一对一的动态博弈模型，然后简要介绍了群智能优化算法和基本粒子群算法的原理，在此基础上通过粒子浓度的概率选择来控制种群多样性，提出了改进粒子群算法求解博弈的最优混合策略，接着对三种群智算法的性能进行对比实验，证明改进后算法提升了求解最优混合策略的计算效率和准确度，最后将其应用到对抗机动决策中，进行了两组无人机一对一机动决策仿真对比实验，进一步验证了改进粒子群算法求解最优机动策略的有效性。

第四章：基于博弈及深度强化学习的无人机机动决策。首先介绍了强化学习中的马尔可夫决策过程，由其扩展建立了无人机一对一场景下的二人零和马尔可夫博弈模型。接着设计了一对一场景的基本状态空间、动作空间和奖励函数，然后结合了强化学习的自学习能力、深度神经网络处理高维状态特征的能力以及博弈论解决复杂冲突对抗性问题的能力，在原算法的基础上提出了改进 DQN 算法求解最优机动策略，最后进行了三组无人机一对一机动决策仿真对比实验，证明了改进 DQN 算法相较于传统算法能够通过自学习生成有效针对对手的机动策略，并且满足对抗实时性的要求，可以在无人机对抗机动决策中得到很好的应

用。

第五章：总结与展望。总结了全文完成的工作和创新点，指出本文研究方法的不足，并展望未来可进一步改进和研究的方向。

第 2 章　基于无人机六自由度模型的机动空间设计

无人机的动力学和运动学模型是研究机动决策的基础，因此本章首先介绍了 F-16 机型无人机的六自由度模型，设计其控制参数，并在此基础上对基本操纵动作库进行丰富和改进，设计无人机的机动空间，构成无人机在对抗机动决策中可选取的机动策略集，最后对无人机模型的机动空间进行仿真测试，为后续机动决策方法的研究以及仿真实验的进行打好基础。

2.1　无人机控制参数设计及仿真

2.1.1　无人机六自由度模型介绍

无人机的运动一般可以由动力学和运动学方程来表示。在本章中，无人机的空间运动情况通过六自由度模型[58]来描述，其包含了无人机质心的线运动和绕质心的角运动。

（1）坐标系和状态变量

构建无人机的运动方程并定义其状态变量都建立在坐标系的基础之上。常用的坐标系有地球坐标系（$S_e-Ox_ey_ez_e$）、地面坐标系（$S_g-Ox_gy_gz_g$）、机体坐标系（$S_b-Ox_by_bz_b$）、航迹坐标系（$S_h-Ox_hy_hz_h$）、气流坐标系（$S_a-Ox_ay_az_a$）等等，选用恰当的坐标系可以准确地描述无人机的运动状态。除此之外，无人机的多个运动状态变量也是建立其模型并进行相关仿真实验的基础。

气流角根据 $S_b-Ox_by_bz_b$ 和飞行速度矢量 V 之间的关系来定义，无人机的气流角包括迎角 α 和侧滑角 β。迎角 α 是 $S_b-Ox_by_bz_b$ 的 Ox_b 轴和 V 关于机体对称平面上的投影之间的夹角，当投影位于 Ox_b 轴以下时定义迎角为正；侧滑角 β 是机体对称平面和 V 之间的夹角，向右为正。

姿态角是根据地面坐标系 $S_g-Ox_gy_gz_g$ 和机体坐标系 $S_b-Ox_by_bz_b$ 来确定的。无人机的姿态角包括滚转角 ϕ、偏航角 ψ 和俯仰角 θ。规定当机体右倾时滚转角为正，当机头右偏时偏航角为正，当机体上仰时俯仰角为正。

滚转角速度 p、偏航角速度 r 和俯仰角速度 q 是无人机在 $S_b-Ox_by_bz_b$ 下的三个角速度分量，表示了无人机机体相对于地面的转动角速度在机体坐标轴上的投影。

航迹角是根据 $S_a-Ox_ay_az_a$ 和 $S_g-Ox_gy_gz_g$ 之间的关系来确定的，其包括了航迹滚转角 γ、航迹偏转角 φ 和航迹倾斜角 μ。将 Oz_a 轴和 Ox_a 轴的铅垂平面之间的夹角定义为航迹滚转角，机体向右滚转时为正；将 Ox_g 轴和无人机空速矢量 V 在地平面上的投影之间的夹角定义为航迹偏转角，机体向右偏转时为

正；将空速矢量 V 和地平面之间的夹角定义为航迹倾斜角，向上为正。

本文确定以固定翼无人机为研究对象，使用升降舵、方向舵、副翼和油门作为控制量 U 来控制固定翼无人机的飞行。当升降舵向下偏转时其偏转角 δ_e 为正；当方向舵向左偏转时其偏转角 δ_r 为正，产生的偏航力矩为负；当左侧副翼向上、右侧副翼向下偏转时其偏转角 δ_a 为正，产生的滚转力矩为负；当无人机推力加大时油门杆 δ_T 为正。无人机的被控状态量有 12 个：三个方位参数 x_g,y_g,h_g，三个姿态角（滚转角 ϕ、偏航角 ψ、俯仰角 θ），三个轴向角速度（滚转角速度 p、偏航角速度 r、俯仰角速度 q），两个气流角（迎角 α、侧滑角 β），和飞行速度 V。

（2）无人机动力学和运动学方程

无人机在飞行过程中受到的外力有：重力 G、推力 T 和空气动力 R，三个力一起施加在无人机上的效果可用一个绕无人机质心的合力矩 $M_{\sum}$ 来表示，F 代表合力，作用于无人机的质心。一般把力矩分解成绕机体 Ox_b 轴旋转的滚转力矩 $\overline{L}$，绕 Oy_b 轴旋转的俯仰力矩 M 和绕 Oz_b 轴旋转的偏航力矩 N，见下图2.1。

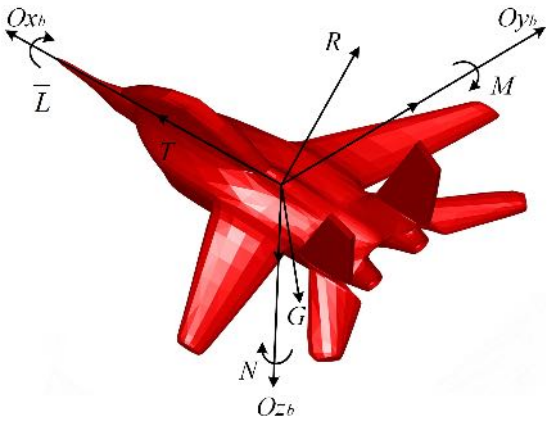


图 2.1　作用于无人机的力和力矩

无人机受到的重力 G 在机体坐标系的三个轴上的分量表示为[59-61]：

$$\begin{bmatrix} G_x \\ G_y \\ G_z \end{bmatrix} = \begin{bmatrix} -mg\sin\theta \\ mg\cos\theta\sin\phi \\ mg\cos\theta\cos\phi \end{bmatrix} \tag{2.1}$$

无人机受到的推力 T 由固定在飞机纵轴上的发动机生成。推力 T 的作用点在机体坐标系上的坐标为（l_x, l_y, l_z），Ox_b 轴和 T 在无人机机体对称平面的投影之间的夹角为 α_T，投影在下侧定义为正，T 的投影和机体对称平面之间的夹角为 β_T，投影在对称平面的左边定义为正，那么 T 在机体坐标系上的三个分量为：

$$\begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix} = \begin{bmatrix} T\cos\alpha_T\cos\beta_T \\ -T\sin\beta_T \\ T\sin\alpha_T\cos\beta_T \end{bmatrix} \tag{2.2}$$

当 $\alpha_T=\beta_T=0$ 时，$T_x=T$。对推力的力矩（$M_T,N_T,\overline{L}_T$）可表示为：

$$\begin{bmatrix}L_T\\M_T\\N_T\end{bmatrix}=\begin{bmatrix}-T_y\cdot l_z+T_z\cdot l_y\\T_x\cdot l_z-T_z\cdot l_x\\-T_x\cdot l_y+T_y\cdot l_x\end{bmatrix}\tag{2.3}$$

空气动力 R 产生的总气动力矩 M_R 在机体坐标系的三个轴上的分量为（$\overline{L}_A$,M_A,N_A），$\overline{L}_A$ 是气动滚转力矩，M_A 是气动俯仰力矩，N_A 是气动偏航力矩，表示为：

$$M_R=\begin{bmatrix}\overline{L}_A\\M_A\\N_A\end{bmatrix}=\begin{bmatrix}\frac{1}{2}C_l\rho V^2S_wb\\\frac{1}{2}C_m\rho V^2S_wc_A\\\frac{1}{2}C_n\rho V^2S_wb\end{bmatrix}\tag{2.4}$$

其中，C_l 是滚转力矩系数，C_m 是俯仰力矩系数，C_n 是偏航力矩系数，ρ 是空气密度，b、S_w、c_A 分别代表机翼的展长，面积和平均几何弦长，力矩系数表达式为：

$$\begin{cases}C_l=C_{l\beta}\beta+\dfrac{b}{2V}C_{lp}p+\dfrac{b}{2V}C_{lr}r+C_{l\delta_a}\delta_a+C_{l\delta_r}\delta_r\\C_m=C_{m,\alpha=0}+C_{m\alpha}\alpha+\dfrac{c_A}{2V}C_{m\dot{\alpha}}\dot{\alpha}+\dfrac{c_A}{2V}C_{mq}q+C_{m\delta_e}\delta_e\\C_n=C_{n\beta}\beta+\dfrac{b}{2V}C_{nr}r+\dfrac{b}{2V}C_{np}p+C_{n\delta_r}\delta_r+C_{n\delta_a}\delta_a\end{cases}\tag{2.5}$$

其中，$C_{l\beta}$、C_{lp}、C_{lr}、$C_{l\delta_a}$、$C_{l\delta_r}$、$C_{m,\alpha=0}$、$C_{m\alpha}$、$C_{m\dot{\alpha}}$、C_{mq}、$C_{m\delta_e}$、$C_{n\beta}$、C_{nr}、C_{np}、$C_{n\delta_r}$、$C_{n\delta_a}$ 都是无人机的气动参数。

通过公式 (2.3)、公式 (2.4) 可得无人机由于受到推力和气动力而产生的合外力矩为：

$$M_\Sigma=\begin{bmatrix}\overline{L}\\M\\N\end{bmatrix}=\begin{bmatrix}\overline{L}_A\\M_A\\N_A\end{bmatrix}+\begin{bmatrix}\overline{L}_T\\M_T\\N_T\end{bmatrix}\tag{2.6}$$

无人机的力矩方程组表示为：

$$\begin{cases}\dot{p}=\dfrac{1}{I_xI_z-I_{xz}^2}[I_z\overline{L}+I_{xz}N+(I_x-I_y+I_z)I_{xz}pq+(I_yI_z-I_z^2-I_{xz}^2)qr]\\\dot{q}=\dfrac{1}{I_y}[M-I_{xz}(p^2-r^2)+(I_z-I_x)pr]\\\dot{r}=\dfrac{1}{I_xI_y-I_{xz}^2}[(I_x^2-I_xI_y+I_{xz}^2)pq-I_{xz}(I_x-I_y+I_z)qr+I_xN+I_{xz}\overline{L}]\end{cases}\tag{2.7}$$

其中，I_x,I_y,I_z 是无人机绕机体轴的转动惯量，I_{xy},I_{yz},I_{xz} 是惯性积[59-61]。

无人机的导航方程组为：

$$\begin{cases} \dot{x_g} = u \cdot C\theta \cdot C\psi + v(S\phi \cdot S\theta \cdot C\psi - C\phi \cdot S\psi) + w(S\phi \cdot S\psi + C\phi \cdot S\theta \cdot C\psi) \\ \dot{y_g} = u \cdot C\theta \cdot S\psi + v(S\phi \cdot S\theta \cdot S\psi + C\phi \cdot S\psi) + w(-S\phi \cdot C\psi + C\phi \cdot S\theta \cdot S\psi) \\ \dot{h_g} = u \cdot S\theta - v \cdot S\phi \cdot C\theta - w \cdot C\phi \cdot C\theta \end{cases} \tag{2.8}$$

其中，$C\theta$ 代表 $\cos\theta$，$S\theta$ 代表 $\sin\theta$，其他同理。u, v, w 分别是飞行速度 V 在机体坐标系的三个轴上的分量：

$$\begin{cases} u = V\cos\alpha\cos\beta \\ v = V\sin\beta \\ w = V\sin\alpha\cos\beta \end{cases} \tag{2.9}$$

无人机的运动学方程组为[59-61]：

$$\begin{cases} \dot{\phi} = p + (r\cos\phi + q\sin\phi)\tan\theta \\ \dot{\theta} = q\cos\phi - r\sin\phi \\ \dot{\psi} = \dfrac{1}{\cos\theta}(r\cos\phi + q\sin\phi) \end{cases} \tag{2.10}$$

无人机的动力学方程组为[59-61]：

$$\begin{cases} \dot{V} = \dfrac{u\dot{u} + v\dot{v} + w\dot{w}}{V} \\ \dot{\alpha} = \dfrac{u\dot{w} - w\dot{u}}{u^2 + w^2} \\ \dot{\beta} = \dfrac{\dot{v}V - v\dot{V}}{V^2\cos\beta} \end{cases} \tag{2.11}$$

其中，$\dot{u}$、$\dot{v}$、$\dot{w}$ 从无人机所受外力方程组中求得：

$$\begin{cases} \dot{u} = vr - wq - g\sin\theta + \dfrac{F_x}{m} \\ \dot{v} = wp - ur + g\cos\theta\sin\phi + \dfrac{F_y}{m} \\ \dot{w} = uq - vp + g\cos\theta\cos\phi + \dfrac{F_z}{m} \end{cases} \tag{2.12}$$

其中，$[F_x, F_y, F_z]^T$ 是无人机受到的总空气动力和推力在机体坐标系的三个轴上的分量，由公式 (2.2) 可知无人机受到的推力，受到的总空气动力 $\boldsymbol{R}$ 在气流坐标系下分解为 (X, Y, Z)，表达式为：

$$\begin{bmatrix} R_{xa} \\ R_{ya} \\ R_{za} \end{bmatrix} = \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} -D \\ Y \\ -L \end{bmatrix} = \begin{bmatrix} -\dfrac{1}{2}C_D\rho V^2 S_w \\ \dfrac{1}{2}C_Y\rho V^2 S_w \\ -\dfrac{1}{2}C_L\rho V^2 S_w \end{bmatrix} \tag{2.13}$$

其中，(D, L, Y) 分别代表无人机受到的空气阻力、升力和侧力，C_L, C_D, C_Y 分别代表无量纲的升力系数，阻力系数和侧力系数[61]。把无人机在气流坐标系上受到的空气动力转换到机体坐标系可得：

$$\begin{bmatrix} R_x \\ R_y \\ R_z \end{bmatrix} = {S_{\alpha\beta}}^T \begin{bmatrix} -D \\ Y \\ -L \end{bmatrix} = \begin{bmatrix} L\sin\alpha - Y\cos\alpha\sin\beta - D\cos\alpha\cos\beta \\ Y\cos\beta - D\sin\beta \\ -L\cos\alpha - Y\sin\alpha\sin\beta - D\sin\alpha\cos\beta \end{bmatrix} \tag{2.14}$$

其中，${S_{\alpha\beta}}^T$ 是把气流坐标系转换至机体坐标系的转换矩阵，将公式 (2.14) 和公式 (2.2) 合并可以得到：

$$\begin{bmatrix} F_x \\ F_y \\ F_z \end{bmatrix} = \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix} = \begin{bmatrix} R_x \\ R_y \\ R_z \end{bmatrix} \tag{2.15}$$

最后把公式 (2.15) 代入公式 (2.12) 即可完成无人机外力方程组的求解。由于无人机所受过载为所受总气动力和推力的合力与所受重力之比，通过公式 (2.1) 和公式 (2.12) 可得无人机三个轴向的过载为：

$$\begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix} = \frac{1}{g} \cdot \begin{bmatrix} \dot{u} + wq - vr + g\sin\theta \\ \dot{v} + ur - wp - g\cos\theta\sin\phi \\ -\dot{w} + uq - vp + g\cos\theta\cos\phi \end{bmatrix} \tag{2.16}$$

到此无人机的六自由度建模完成，公式 (2.7)、公式 (2.8)、公式 (2.10) 和公式 (2.11) 一起构成了无人机的动力学和运动学模型。无人机模型由上一时刻的状态，根据公式 (2.7)、公式 (2.8)、公式 (2.10)、公式 (2.11) 和公式 (2.16) 更新无人机的状态，最后输出新的无人机状态量，流程如下图2.2所示。

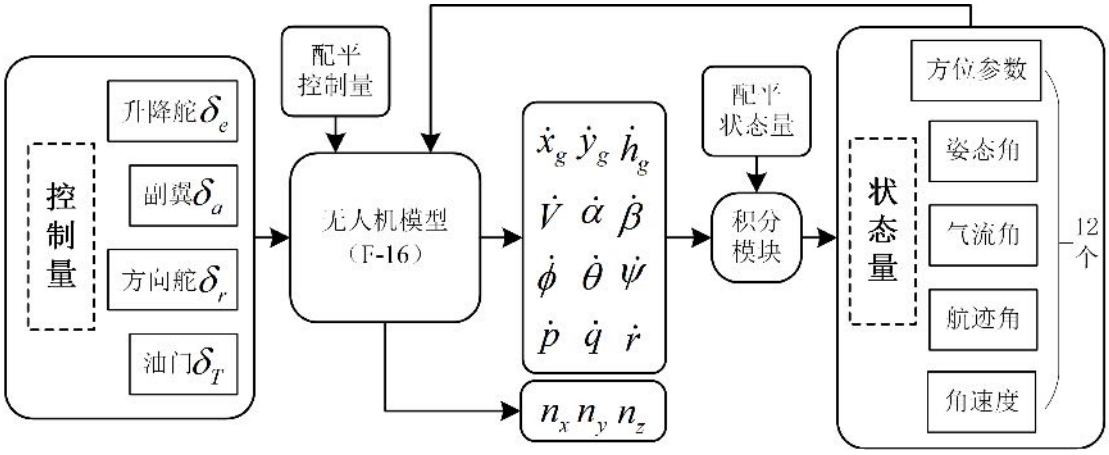


图 2.2　无人机模型运行框图

2.1.2　控制参数设计

本文选用 F-16 机型作为无人机模型并搭建其 MATLAB 模型[62] 来进行对抗机动决策的研究。该无人机具备良好的机动能力且各项实验数据齐全，其空气动力参数和结构参数等在无人机设计之初已有，均来源于 NASA 报告及相关资料[59-60]。在无人机进行机动动作之前首先要对其完成配平工作，设置无人机的初始迎角和升降舵偏角，使无人机的合力矩和所受合力为零。确定平衡点参数使无人机在初始时刻为稳定的平飞状态。只有在完成配平的基础上，才能施加各种控制指令使无人机执行各种飞行任务。无人机的运动在配平后可分为基准运动和扰动运动，设计无人机的输入控制指令为配平状态控制量和指令控制量的叠加，通过纵向和横侧向两个通道来进行无人机的控制参数设计。

（1）纵向通道控制设计

无人机在机体对称平面内进行的机动过程称为纵向平面运动，设计纵向通道的控制律主要考虑高度、速度和俯仰角。

控制纵向通道要实现无人机的速度和迎角维持稳定且到达预期的响应效果。为达到改善无人机纵向通道的短周期阻尼的目的，考虑在纵向通道内环的设计中加入无人机的俯仰角速度反馈。然后在俯仰角速度反馈构成的回路基础之上，设计根据无人机迎角反馈回路形成的无人机姿态控制回路，由此构建无人机的迎角自动驾驶仪，通过反馈迎角信息达到控制系统纵向通道的增稳控制。本文使用俯仰角速度和迎角的双反馈回路，控制回路的框图如下图2.3所示。

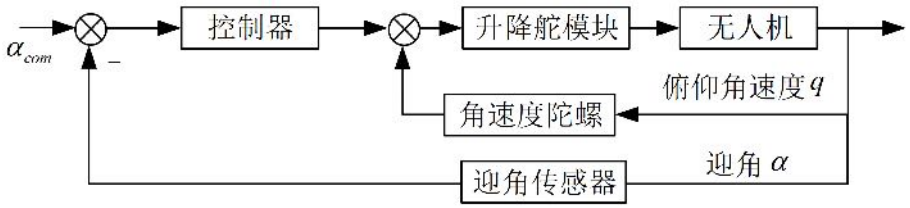


图 2.3　纵向通道控制回路

设计纵向通道的增稳控制律为：

$$\delta_e = k_\alpha(\alpha - \alpha_{com}) + k_q q \tag{2.17}$$

其中，k_α 和 k_q 代表控制律参数，α_{com} 代表迎角指令，使用 PID 控制结构时：$k_\alpha = K_P + K_I(1/s) + K_D s$。除此之外，根据升降舵偏转角度和偏转角速率的限制条件，升降舵偏转角范围为 $-25 \sim 25\,\text{deg}$，偏转角速率范围为 $-60 \sim 60\,\text{deg/s}$，因此在升降舵模块中加入非线性限制环节。

（2）横侧向通道控制设计

无人机在机体非对称平面内进行的机动过程称为横侧向运动，其中包括无人机的偏航和滚转运动，主要由方向舵通道和副翼通道来控制。设计无人机横侧

向通道的控制律要实现在无人机的转弯过程中可以保持稳定的偏航角和滚转角，完成平衡的转弯飞行。本文针对滚转角的控制回路考虑使用滚转角和滚转角速度的双反馈回路，滚转角控制通道的结构框图如下图2.4所示。

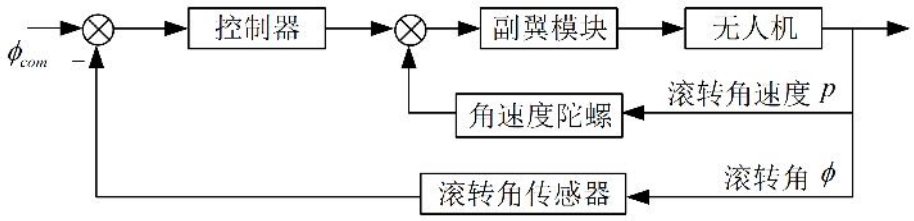


图 2.4　副翼控制回路

设计副翼通道的增稳控制律为：

$$\delta_a = k_\phi(\phi - \phi_{com}) + k_p p \tag{2.18}$$

其中，k_ϕ 和 k_p 代表控制律参数，ϕ_{com} 代表滚转角指令，使用 PID 控制结构时：$k_\phi = K_P + K_I(1/s) + K_D s$。除此之外，根据副翼偏转角度和偏转角速率的限制条件，副翼偏转角范围为 $-21.5 \sim 21.5$ deg，偏转角速率范围为 $-80 \sim 80$ deg/s，因此在副翼模块中加入非线性限制环节。

偏航运动通过方向舵来控制。由于升降舵与方向舵、副翼控制量之间存在耦合，考虑使用偏航角速度和侧向过载反馈，方向舵控制回路的结构框图如下图2.5所示。

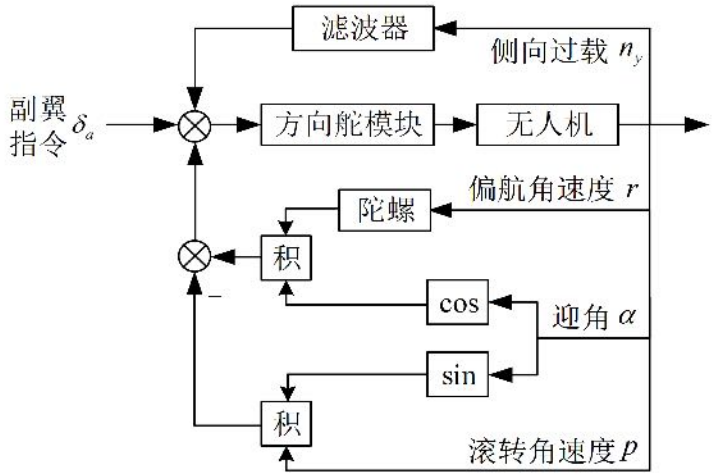


图 2.5　方向舵控制回路

其中，为增加荷兰滚模态的阻尼采用偏航角速度反馈，为提高荷兰滚的频率采用侧向过载反馈，如此便减少了滚转机动和侧向扰动时的侧向过载和侧滑角。引入迎角 α 和滚转角速度 p 交联乘积的目的是把向航向阻尼器引入绕机体轴的偏航角速度转换成绕稳定轴系的偏航角速度，进而增加荷兰滚模态的阻尼，抑制

滚转机动的偏航角速度反馈带来的不利偏航力矩以实现增稳的效果。

由此通过纵向和横侧向两个通道的增稳控制完成了对无人机控制参数的设计。

2.1.3　控制律仿真

本节在无人机控制参数设计的基础上进行纵向和横侧向两个通道的控制律仿真，首先搭建六自由度无人机的 MATLAB 模型，完成其配平工作，下表2.1给出了在高度为 3000m，速度为 152m/s，迎角为 3.5973 deg 时的配平控制量。

表 2.1　无人机配平状态

符号	名称	配平量
$thrust$	油门杆	2080.9182 lb
$elevator$	升降舵	-2.252 deg
$aileron$	副翼舵	0 deg
$rudder$	方向舵	0 deg
α	迎角	3.5973 deg
θ	俯仰角	3.5973 deg
h	高度	10000ft = 3000m
v	速度	500ft/s = 152m/s

（1）纵向通道控制律仿真

首先在无人机配平的基础上对纵向通道参数的响应情况进行测试。根据表2.1设置无人机的初始状态，输入一个 -1 deg 的升降舵阶跃信号到无人机的非线性模型中，分别可以得到无人机的速度 v、迎角 α、俯仰角 θ 和俯仰角速度 q 的响应曲线，如下图2.6所示。

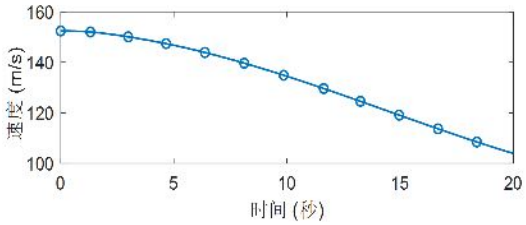
从无人机在单位阶跃升降舵信号下的响应曲线可以看出，由于没有加入纵向通道控制，无人机的速度变化很快，迎角不能维持稳定，因此对无人机的纵向通道控制回路进行设计是非常重要的。

根据公式 (2.17)，使用 PID 控制结构时，取 $K_P = 0.9$，$K_I = 1.7$，$K_D = 0$，设置图2.3中迎角传感器的传递函数为：

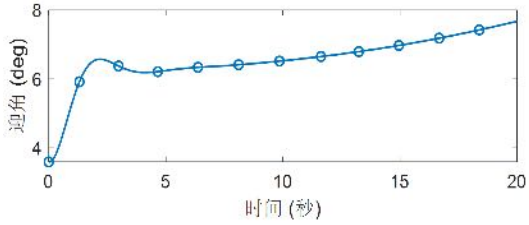
$$\frac{\alpha_f(s)}{\alpha(s)} = \frac{10}{s+10} \tag{2.19}$$

设置俯仰角速度反馈回路的阻尼器为：

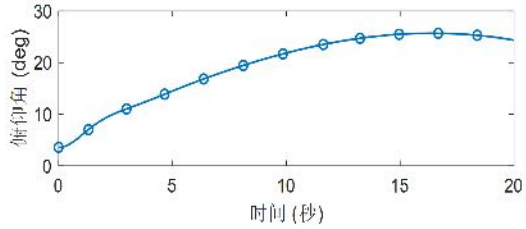
$$\frac{q_f(s)}{q(s)} = \frac{78.5^2}{s^2 + 2 \cdot 0.89 \cdot 78.5s + 78.5^2} \tag{2.20}$$



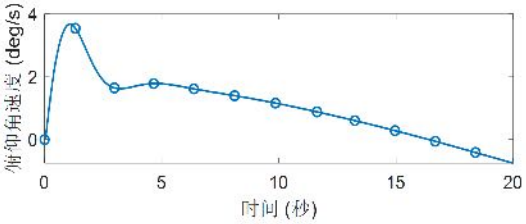
(a) 速度



(b) 迎角



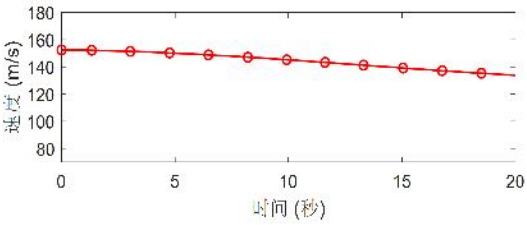
(c) 俯仰角



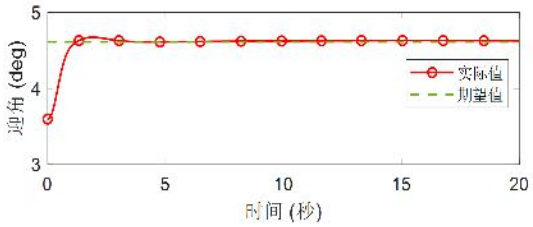
(d) 俯仰角速度

图 2.6　配平状态下纵向通道参数对升降舵信号的响应曲线

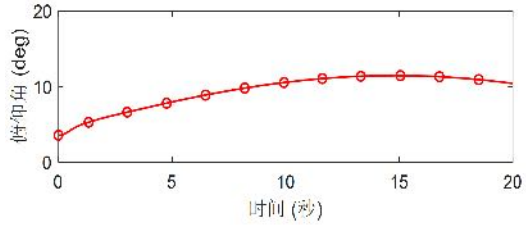
引入俯仰角速度和迎角的双反馈回路后，根据表2.1设置无人机的初始状态，给无人机一个迎角指令 $\alpha_{com} = 4.6\,\mathrm{deg}$，分别可以得到无人机的速度 v、迎角 α、俯仰角 θ 和俯仰角速度 q 的响应曲线，如下图2.7所示。



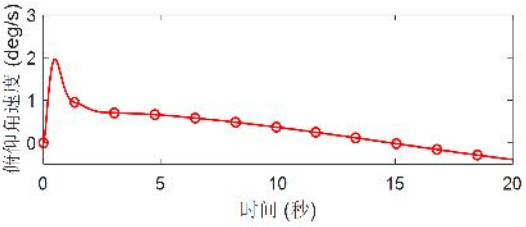
(a) 速度



(b) 迎角



(c) 俯仰角



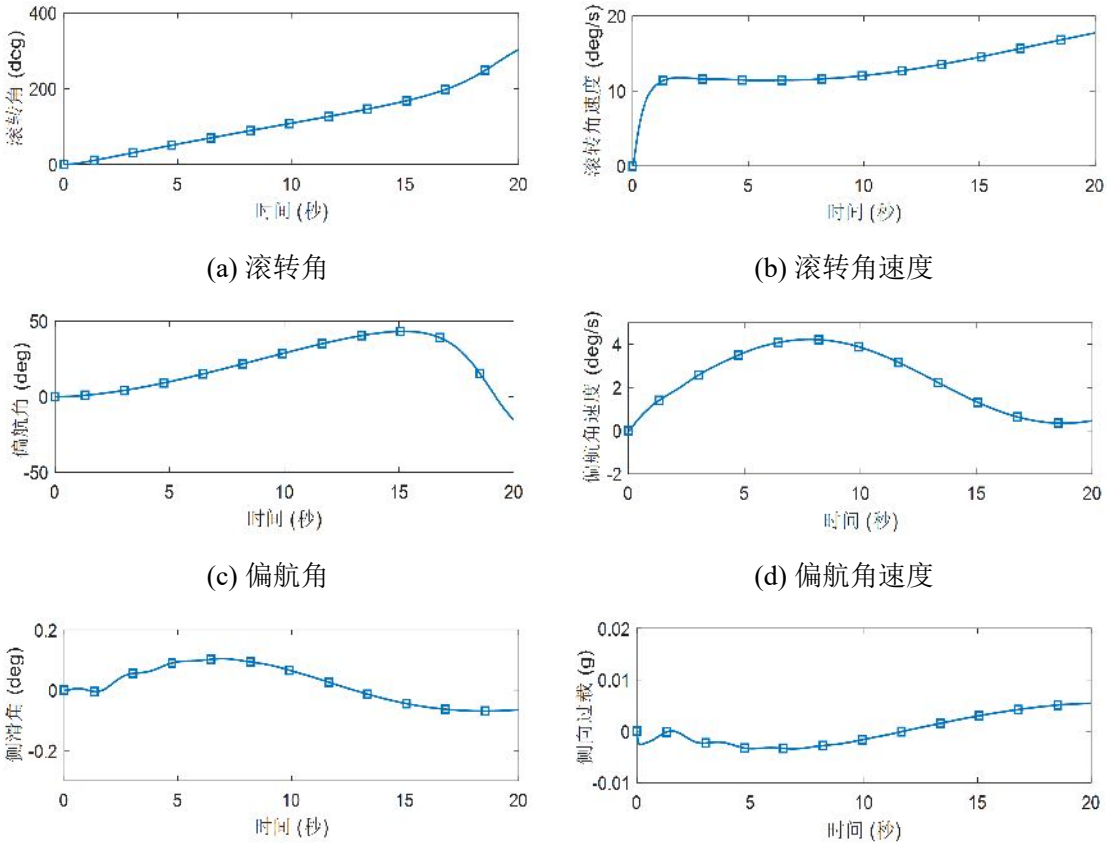
(d) 俯仰角速度

图 2.7　引入控制回路后纵向通道参数的响应曲线

从上图 (a) 可以看出，加入纵向通道控制回路后，无人机的速度变化不大，基本可以维持稳定；图 (b) 中无人机迎角的实际值与期望值吻合，快速响应后保持在给定的输入值，说明无人机的迎角自动驾驶仪可以很好地跟踪迎角指令，如此便实现了纵向通道回路增稳控制的效果。

（2）横侧向通道控制律仿真

在无人机配平的基础上对横侧向通道参数的响应情况进行测试。根据表2.1设置无人机的初始状态，输入一个 -1 deg 的副翼舵阶跃信号到无人机的非线性模型中，分别可以得到无人机的滚转角 ϕ、滚转角速度 p、偏航角 ψ、偏航角速度 r、侧滑角 β 和侧向过载 n_y 的响应曲线，如下图2.8所示。



(a) 滚转角　　(b) 滚转角速度

(c) 偏航角　　(d) 偏航角速度

(e) 侧滑角　　(f) 侧向过载

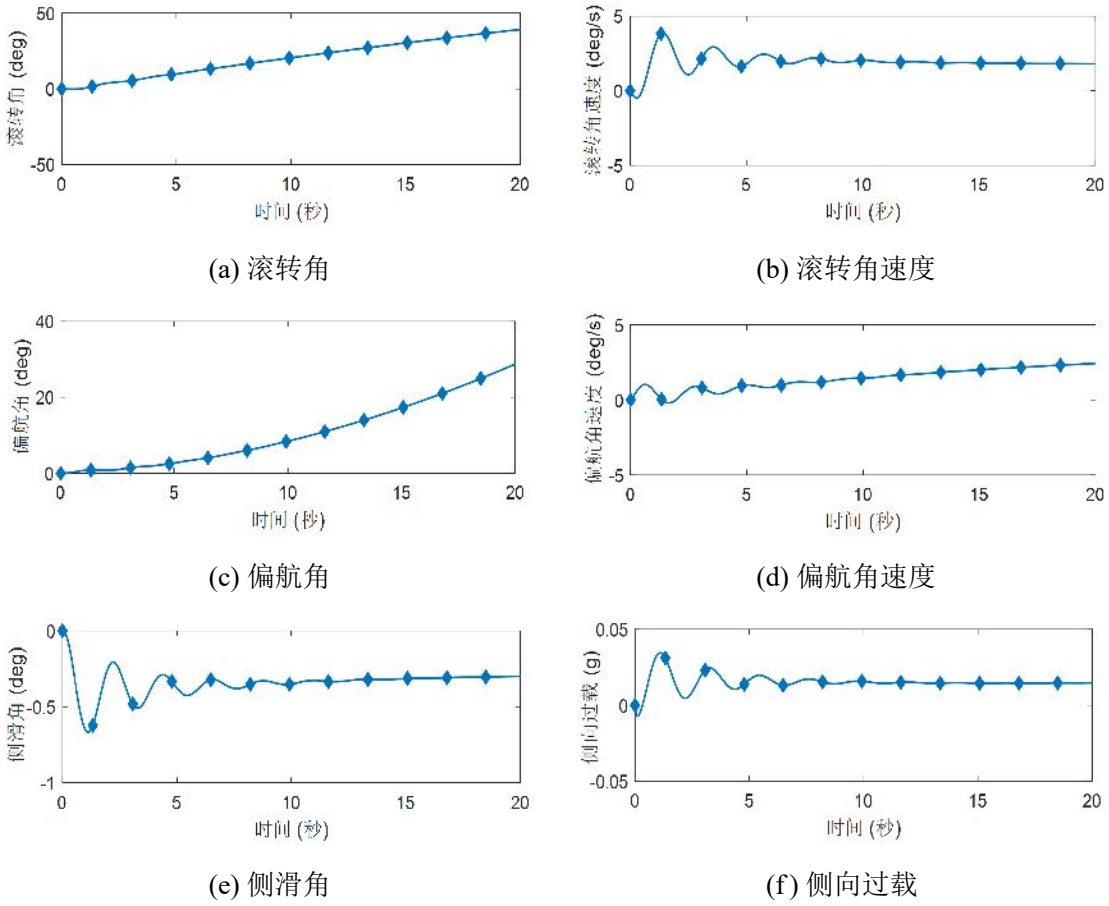
图 2.8　配平状态下横侧向通道参数对副翼舵信号的响应曲线

在配平的基础上，输入一个 -1 deg 的方向舵阶跃信号到无人机的非线性模型中，分别可以得到无人机的滚转角 ϕ、滚转角速度 p、偏航角 ψ、偏航角速度 r、侧滑角 β 和侧向过载 n_y 的响应曲线，如下图2.9所示。

从无人机在单位阶跃副翼舵和方向舵信号下的响应曲线可以看出，由于没有加入横侧向通道控制，无人机的滚转角速度和偏航角速度不能维持稳定，滚转角的变化很大，同时也带来了不平稳的侧滑角和侧向过载，导致无人机不能完成平衡的转弯飞行，因此对无人机的横侧向控制回路进行设计是非常重要的。

根据公式 (2.18)，使用 PID 控制结构时，取 $K_P = 2$，$K_I = 1.5$，$K_D = 0$，设置图2.4中滚转角速度陀螺传感器的传递函数为：

$$\frac{p_f(s)}{p(s)} = \frac{90^2}{s^2 + 2 \cdot 0.8 \cdot 90s + 90^2} \tag{2.21}$$



(a) 滚转角　　(b) 滚转角速度

(c) 偏航角　　(d) 偏航角速度

(e) 侧滑角　　(f) 侧向过载

图 2.9　配平状态下横侧向通道参数对方向舵信号的响应曲线

设置图2.5中侧向过载滤波器的传递函数为：

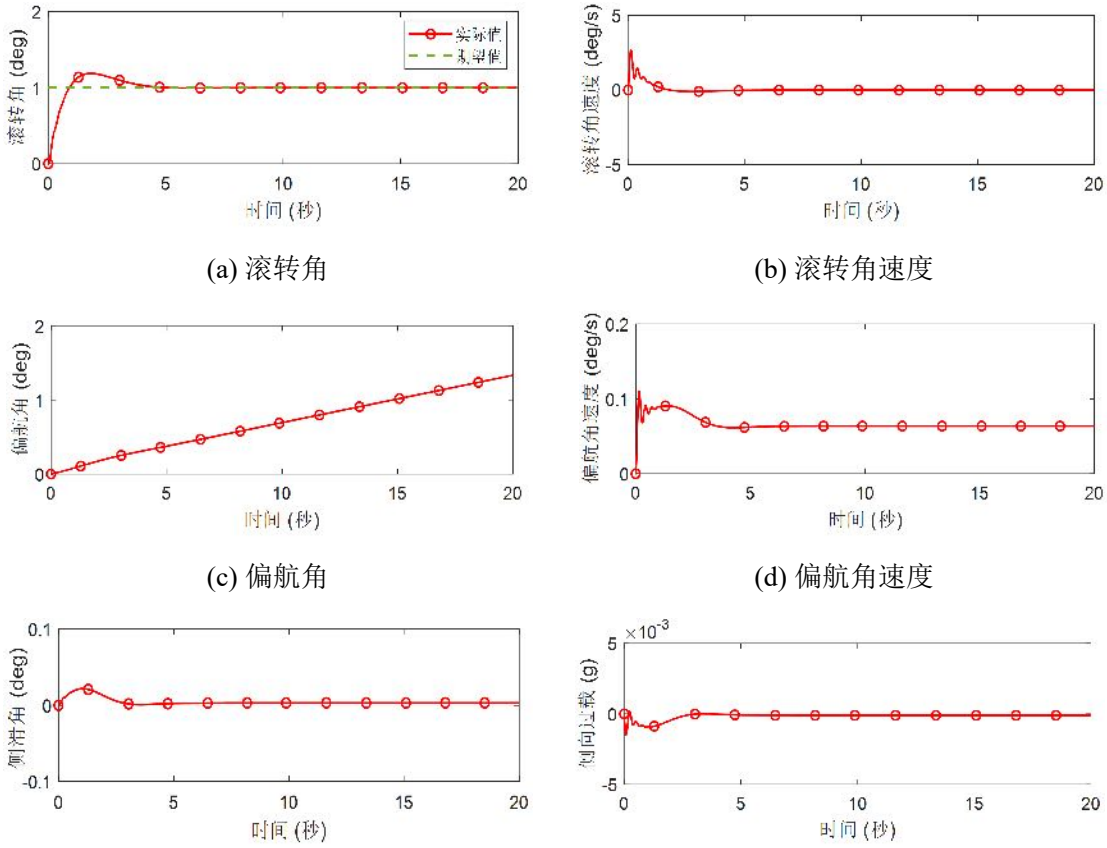
$$\frac{n_{yf}(s)}{n_y(s)} = \frac{78.5^2}{s^2 + 2 \cdot 0.89 \cdot 78.5s + 78.5^2} \tag{2.22}$$

并设置偏航角速度陀螺传感器的传递函数为：

$$\frac{r_f(s)}{r(s)} = \frac{200^2}{s^2 + 2 \cdot 0.8 \cdot 200s + 200^2} \tag{2.23}$$

根据表2.1设置无人机的初始状态，给无人机一个单位滚转角指令 $\phi_{com} = 1\,\text{deg}$ 后，分别可以得到无人机的滚转角 ϕ、滚转角速度 p、偏航角 ψ、偏航角速度 r、侧滑角 β 和侧向过载 n_y 的响应曲线，如下图2.10所示。

从图中可以看出，加入副翼控制回路和方向舵控制回路结构后，无人机滚转角的实际值与期望值吻合，快速响应后保持在给定的输入值，说明无人机的滚转角自动驾驶仪可以很好地跟踪滚转角指令；滚转角速度和偏航角速度在响应时间 4 秒后基本达到稳定状态；同时无人机侧滑角和侧向过载的震荡也得到快速有效的抑制并维持平稳状态。如此便实现了横侧向通道回路增稳控制的效果，无人机在转弯过程中可以保持稳定的滚转角和偏航角，以完成平衡的转弯飞行。



(a) 滚转角　(b) 滚转角速度

(c) 偏航角　(d) 偏航角速度

(e) 侧滑角　(f) 侧向过载

图 2.10　引入控制回路后横侧向通道参数的响应曲线

综上，纵向和横侧向两通道参数都达到了期望的响应效果，满足了两通道控制回路的设计需求。

2.2　无人机机动空间设计及仿真

无人机的机动空间是空战机动决策的基础，构成了其可选取的机动策略集。其设计方式一般有两种，一种是基于经典的战斗机战术飞行动作[63]，即“典型战术动作库”；另一种是“基本操纵动作库”[64]，基于空战中最常用的操纵方式：平飞、爬升、俯冲等。本文对“基本操纵动作库”进行改进，设计无人机的机动空间并对其进行仿真测试。

2.2.1　机动动作控制器

机动动作控制器的输入为由空战机动决策层所选择的机动动作指令，输出为由无人机六自由度模型的控制层可执行的控制指令。由 2.2 节可知，机动动作控制器需要给无人机控制层输出的结果为迎角指令和滚转角指令。

进行空战机动决策规划首先要构建无人机的简化模型，无人机的质心动力

学和运动学方程组可以描述无人机的运动状态，因此可以用做无人机的简化模型，在航迹坐标系下质心动力学方程为：

$$\begin{cases} m\dfrac{dv}{dt} = T\cos\alpha\cos\beta - D - mg\sin\mu \\ mv\dfrac{d\mu}{dt} = T(\cos\alpha\sin\beta\sin\gamma + \sin\alpha\cos\gamma) - Z\cos\gamma - Y\sin\gamma - mg\cos\mu \\ -mv\cos\mu\dfrac{d\varphi}{dt} = T(\cos\alpha\sin\beta\cos\gamma - \sin\alpha\sin\gamma) + Z\sin\gamma - Y\cos\gamma \end{cases} \tag{2.24}$$

在地面坐标系下质心运动学方程为：

$$\begin{cases} \dfrac{dx_g}{dt} = v\cos\mu\cos\varphi \\ \dfrac{dy_g}{dt} = v\cos\mu\sin\varphi \\ \dfrac{dz_g}{dt} = -v\sin\mu \end{cases} \tag{2.25}$$

根据公式 (2.24) 和公式 (2.25)，如果已知了无人机的空气动力特性、推力特性和初始飞行状态，由公式 (2.24) 积分可得无人机的飞行速度、航迹倾斜角和航向角的变化情况，结合无人机的初始状态对公式 (2.25) 积分可得无人机空间位置的变化情况。空战机动决策层根据无人机质心运动模型进行机动动作选择后，经过机动动作控制器将其转化为无人机控制回路的控制指令，由此更新六自由度无人机的状态[61]。整个流程如下图2.11所示。

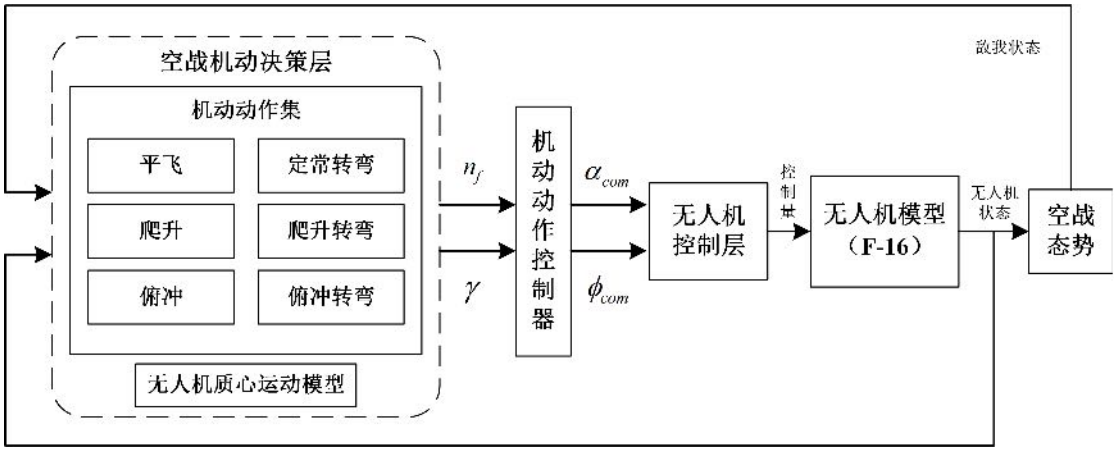


图 2.11　无人机对抗流程

为了便于机动决策的研究，假设：速度方向与推力矢量一致；忽略地球曲率和公转自转[65]；忽略风速对无人机的影响，故航迹滚转角等于滚转角，航迹方位角等于偏航角，航迹倾斜角等于俯仰角；无人机无侧滑运动，故侧力为零，侧

滑角为零。简化后的无人机质心运动学方程组为：

$$\begin{cases} m\dfrac{dv}{dt} = T - D - mg\sin\mu \\ mv\dfrac{d\mu}{dt} = -Z\cos\gamma - mg\cos\mu \\ mv\cos\mu\dfrac{d\varphi}{dt} = -Z\sin\gamma \\ \dfrac{dx_g}{dt} = v\cos\mu\cos\varphi \\ \dfrac{dy_g}{dt} = v\cos\mu\sin\varphi \\ \dfrac{dz_g}{dt} = -v\sin\mu \end{cases} \tag{2.26}$$

使用过载表示无人机质心动力学方程，可得：

$$\begin{cases} \dfrac{dv}{dt} = g(n_x - \sin\mu) \\ \dfrac{d\mu}{dt} = \dfrac{g}{v}(n_f\cos\gamma - \cos\mu) \\ \dfrac{d\varphi}{dt} = \dfrac{g}{v\cos\mu}n_f\sin\gamma \end{cases} \tag{2.27}$$

其中，过载为气动力和推力的合力与飞机重力的比值，投影到航迹坐标系下表示为：

$$\begin{cases} n_x = \dfrac{T - D}{mg} \\ n_y = \dfrac{-Z\sin\gamma}{mg} \\ n_z = \dfrac{-Z\cos\gamma}{mg} \end{cases} \tag{2.28}$$

其中，n_x 表示切向过载，沿无人机飞行速度方向；n_y 和 n_z 与速度方向垂直，可得法向过载：$n_f = \sqrt{n_y^2 + n_z^2} = -Z/mg$。无人机质心运动学方程组为：

$$\begin{cases} \dfrac{dv}{dt} = g(n_x - \sin\mu) \\ \dfrac{d\mu}{dt} = \dfrac{g}{v}(n_f\cos\gamma - \cos\mu) \\ \dfrac{d\varphi}{dt} = \dfrac{g}{v\cos\mu}n_f\sin\gamma \\ \dfrac{dx_g}{dt} = v\cos\mu\cos\varphi \\ \dfrac{dy_g}{dt} = v\cos\mu\sin\varphi \\ \dfrac{dz_g}{dt} = -v\sin\mu \end{cases} \tag{2.29}$$

切向过载 n_x 影响无人机飞行速度的大小，法向过载 n_f 和滚转角 γ 共同影响无人机的飞行方向，改变无人机的空间位置。因此可根据无人机不同的机动动作，设置相应的过载和滚转角建立其机动动作集。将滚转角指令作为 ϕ_{com} 输入到无人机的横侧向控制通道，再将法向过载指令折合成迎角指令 α_{com} 输入到无人机的纵向控制通道，从而实现对无人机运动状态的控制。考虑到工程实际，取 $n_f = Z_\alpha \alpha \cdot V/g = 16\alpha$，$\alpha$ 为弧度单位制。

2.2.2　机动动作集设计

设置不同的法向过载和滚转角的组合，可得到无人机不同的机动动作。

（1）平飞

根据无人机质心运动模型，定常平飞运动的航迹方位角 φ 和航迹倾斜角 μ 的变化率都为 0，设置 $n_{fcom} = 1g, \gamma_{com} = 0$。

（2）爬升

爬升机动时 φ 的变化率为 0，设置 $\gamma_{com} = 0$；由公式 (2.29)，μ 的变化率应该为正，则使 $n_f > 1g$，可取 $n_{fcom} = 1.2g$。

（3）俯冲

俯冲机动时 φ 的变化率为 0，设置 $\gamma_{com} = 0$；由公式 (2.29)，μ 的变化率应该为负，则使 $n_f < 1g$，可取 $n_{fcom} = 0.8g$。

（4）定常转弯

定常转弯机动是无人机从平飞开始，以给定滚转角，使无人机在同一水平面内进行转弯运动。

向右转弯时，设置 $\gamma_{com} = 45\,\text{deg}$；向左转弯时，设置 $\gamma_{com} = -45\,\text{deg}$。无人机在同一水平面内定常转弯机动时，其飞行高度不发生变化，因此 μ 的变化率为 0，由公式 (2.29)，在 $\gamma_{com} = 45\,\text{deg}$ 时，$d\mu/dt = 0$，则 $n_{fcom} = 1.414g$。

（5）爬升转弯

爬升转弯机动是无人机在定常转弯的基础上爬升高度。

向右转弯时，设置 $\gamma_{com} = 45\,\text{deg}$；向左转弯时，设置 $\gamma_{com} = -45\,\text{deg}$。无人机在爬升转弯机动时，其飞行高度增加，因此 μ 的变化率大于 0，$d\mu/dt > 0$，则 $n_{fcom} > 1.414g$，可取 $n_{fcom} = 1.7g$。

（6）俯冲转弯

俯冲转弯机动是无人机在定常转弯的基础上降低高度。

向右转弯时，设置 $\gamma_{com} = 45\,\text{deg}$；向左转弯时，设置 $\gamma_{com} = -45\,\text{deg}$。无人机在俯冲转弯机动时，其飞行高度降低，因此 μ 的变化率小于 0，$d\mu/dt < 0$，则 $n_{fcom} < 1.414g$，可取 $n_{fcom} = 1g$。

综上所述，本文设置的机动动作集为：

$$\begin{cases} n_{fcom} = [0.8 \quad 1 \quad 1.2 \quad 1.4 \quad 1.7 \quad 2] \quad (g) \\ \gamma_{com} = [-45 \quad 0 \quad 45] \quad (\text{deg}) \end{cases} \tag{2.30}$$

转化为无人机六自由度模型的控制指令为：

$$\begin{cases} \alpha_{com} = [2.87 \quad 3.58 \quad 4.31 \quad 5.07 \quad 6.09 \quad 7.16] \quad (\text{deg}) \\ \phi_{com} = [-45 \quad 0 \quad 45] \quad (\text{deg}) \end{cases} \tag{2.31}$$

根据上式 (2.31)，迎角指令 α_{com} 和滚转角指令 ϕ_{com} 的不同组合构成了无人机可选的机动动作，这样就完成了无人机机动动作集的构建。

2.2.3　机动空间仿真

本小节对六自由度无人机模型的机动动作集进行仿真，测试其能否按照设置的控制指令完成相应的机动动作。根据表2.1设置无人机模型的初始条件，初始位置为：$x = 0(\text{m})$，$y = 0(\text{m})$，$h = 3000(\text{m})$，初始速度为：$v = 152\text{m/s}$。

（1）平飞

设置六自由度无人机模型的控制指令为：$\alpha_{com} = 3.58\,\text{deg}$，$\phi_{com} = 0\,\text{deg}$，仿真时长为 10 秒，无人机的机动动作轨迹如下图2.12所示。

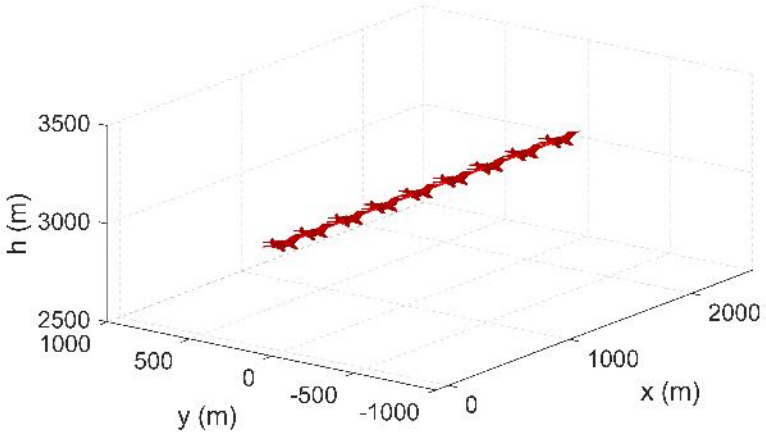


图 2.12　平飞轨迹

（2）爬升

设置六自由度无人机模型的控制指令为：$\alpha_{com} = 6.09\,\text{deg}$，$\phi_{com} = 0\,\text{deg}$，仿真时长为 15 秒，无人机的机动动作轨迹如下图2.13所示。

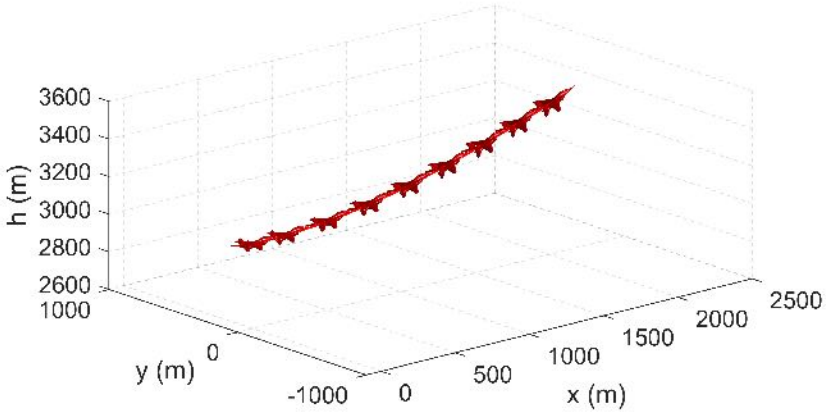


图 2.13　爬升轨迹

（3）俯冲

设置六自由度无人机模型的控制指令为：$\alpha_{com} = 2.87\,\mathrm{deg}$，$\phi_{com} = 0\,\mathrm{deg}$，仿真时长为 15 秒，无人机的机动动作轨迹如下图2.14所示。

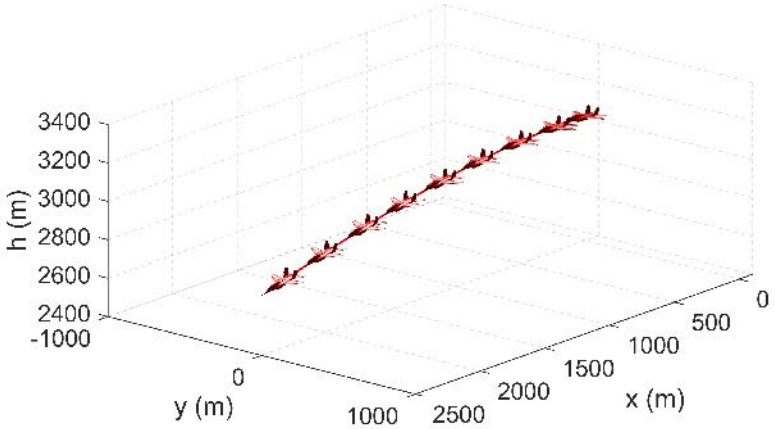


图 2.14　俯冲轨迹

（4）定常转弯

设置六自由度无人机模型的控制指令为：$\alpha_{com} = 5.07\,\mathrm{deg}$，$\phi_{com} = -45\,\mathrm{deg}$，仿真时长为 15 秒，无人机的机动动作轨迹如下图2.15所示。

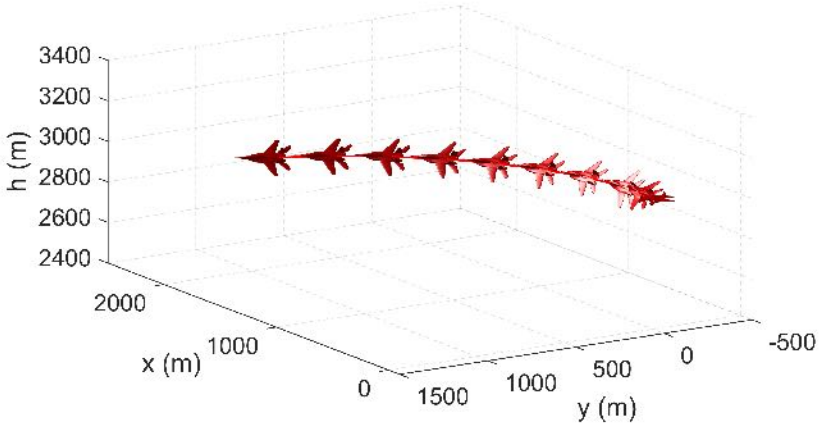


图 2.15　定常转弯轨迹

（5）爬升转弯

设置六自由度无人机模型的控制指令为：$\alpha_{com} = 6.09\,\mathrm{deg}$，$\phi_{com} = 45\,\mathrm{deg}$，仿真时长为 15 秒，无人机的机动动作轨迹如下图2.16所示。

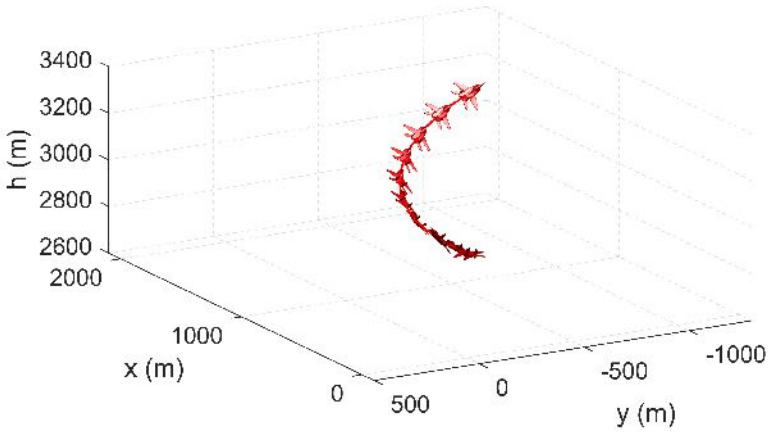


图 2.16　爬升转弯轨迹

（6）俯冲转弯

设置六自由度无人机模型的控制指令为：$\alpha_{com} = 2.87\,\mathrm{deg}$，$\phi_{com} = -45\,\mathrm{deg}$，仿真时长为 10 秒，无人机的机动动作轨迹如下图2.17所示。

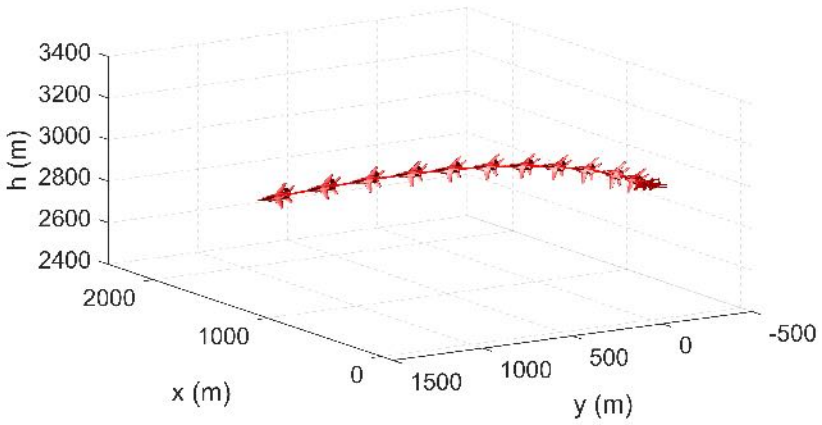


图 2.17　俯冲转弯轨迹

综上，由三维机动轨迹可以看出，给定不同的控制指令组合，六自由度无人机模型皆能完成预期的机动动作。

2.3　本章小结

本章主要完成了 F-16 机型无人机的机动空间设计，为后续机动决策的研究和对抗仿真实验的进行做铺垫。首先介绍了 F-16 机型无人机的六自由度模型，设计其控制参数，在此基础上，对基本操纵动作库进行丰富和改进，设计了无人机的机动空间，构成机动决策的机动策略集。然后对基于 F-16 六自由度无人机模型的控制律进行仿真以测试其性能，对设计的机动空间进行仿真以测试其能否按照设置的控制指令完成相应的机动动作。最后的仿真实验结果表明，所设计的控制参数和机动空间都能达到预期效果，纵向和横侧向两通道参数均达到了期望的响应效果，满足了两通道控制回路的设计需求，并且给定不同的控制指令组合，无人机模型都能完成预期的机动动作。

第 3 章　基于博弈及群智算法的无人机机动决策

本章利用博弈论的思想建立无人机一对一场景下的数学模型，针对博弈对抗中混合策略纳什均衡难于求解的问题，考虑利用基于优化理论的智能算法将其转化为最优化问题进行寻优从而得到最优混合策略，并通过粒子浓度的概率选择来对基本群体智能优化算法进行改进，目的是为了更好的提升算法的性能，然后使用改进后算法对最优混合策略进行求解，以提高无人机对抗机动决策中求解最优机动策略的计算效率和准确度，本章最后设定了两组博弈对抗场景进行仿真对比实验，并对改进后算法的有效性进行分析。

3.1　无人机一对一动态博弈模型

博弈论是研究具有利益冲突的决策者之间进行策略交互的数学模型。博弈中的参与者采取纳什均衡策略作为最优策略，当所有参与者采取纳什均衡策略以后，任何局中人存在侥幸心理企图单方面地改变自己的策略时，都会使自身的利益受损。无人机对抗是一个具有强竞争性和复杂冲突性的动态博弈过程，对抗双方都企图寻求自身利益最大化，而博弈理论正好为无人机对抗机动决策问题带来解决之法，因此本节基于该理论对无人机对抗过程进行建模。

无人机对抗动态博弈模型包含三个元素，用一个三元组表示为：

$$\left\langle N,\{S_i\}_{i=1}^{N},\{e_i\}_{i=1}^{N}\right\rangle \tag{3.1}$$

其中，N 表示参与者集合，无人机一对一场景中，$N=\{N_R,N_B\}$；S_i 表示参与者 i 的策略集，一对一场景中，用 S_R 和 S_B 分别表示红蓝双方无人机可选择机动动作的策略集合，为：

$$\begin{cases} S_R=\{s_{r1},s_{r2},\cdots,s_{rm}\} \\ S_B=\{s_{b1},s_{b2},\cdots,s_{bn}\} \end{cases} \tag{3.2}$$

其中 $\{s_{ri}\}_{i=1}^{m}$ 和 $\{s_{bi}\}_{i=1}^{n}$ 代表可采取的机动策略。e_i 表示支付函数值，指参与者根据自己所选的策略得到的损失或收益，选择不同的策略可以得到不同的支付函数值，根据支付函数值可构成支付矩阵：

$$M=\begin{bmatrix} e_{11} & \cdots & e_{1n} \\ \vdots & \ddots & \vdots \\ e_{m1} & \cdots & e_{mn} \end{bmatrix} \tag{3.3}$$

其中，e_{ij} 表示参与者分别选择策略 s_{ri} 和 s_{bj} 时的支付函数值，在一对一场景中代表红蓝双方无人机执行所选机动策略后根据态势评估函数得到的态势评估值。

$\boldsymbol{M}$ 表示红蓝双方无人机的机动动作集中的每种机动动作一一对应得到的态势评估值进而构成的态势评估值矩阵，用来评估双方执行某种机动策略后的态势优势情况。红蓝无人机的支付函数值为 e_{ij}^{r} 和 e_{ij}^{b}，支付矩阵为 $\boldsymbol{M}_r$ 和 $\boldsymbol{M}_b$。

一对一场景中求解支付矩阵得到红蓝双方的混合策略，混合策略是红蓝无人机以一定的概率分布选择策略集合上的确定性策略。红方的混合策略表示为一个多维向量：$p_r = (p_{r1}, p_{r2}, \cdots, p_{rm})^T$，类似地蓝方的混合策略为：$p_b = (p_{b1}, p_{b2}, \cdots, p_{bn})^T$，其中各元素都大于等于零且相加和为 1。

如果红方无人机选取混合策略 p_r，蓝方无人机选取策略 s_{bj}，双方的平均支付值可计算为：

$$\sum_{i=1}^{m}(p_{ri}e_{ij}^{b}) = (p_r^T M_b)_j \tag{3.4}$$

如果蓝方选取混合策略 p_b，红方选取策略 s_{ri}，其平均支付值可计算为：

$$\sum_{j=1}^{n}(e_{ij}^{r}p_{bj}) = (M_r p_b)_i \tag{3.5}$$

如果红蓝无人机都选取混合策略 p_r 和 p_b，那么红方的平均支付值可计算为：

$$\sum_{i=1}^{m}\left(\left(\sum_{j=1}^{n}e_{ij}^{r}p_{bj}\right)p_{ri}\right) = p_r^T M_r p_b \tag{3.6}$$

蓝方的平均支付值可计算为：

$$\sum_{i=1}^{m}\left(\left(\sum_{j=1}^{n}e_{ij}^{b}p_{bj}\right)p_{ri}\right) = p_r^T M_b p_b \tag{3.7}$$

为了得到对抗博弈中红蓝双方无人机的最优策略组合，则需计算支付矩阵的纳什均衡值，采取纳什均衡策略作为最优策略。使用数学公式可表示为：

$$e_j(s_1, \cdots, s_{i-1}, s_i, s_{i+1}, \cdots, s_N) \geqslant e_j(s_1, \cdots, s_{i-1}, s_i', s_{i+1}, \cdots, s_N) \tag{3.8}$$

其中 $\forall s_i' \in S_j$，$e_j(s_1, \cdots, s_{i-1}, s_i, s_{i+1}, \cdots, s_N)$ 代表在纳什均衡点时参与者 j 的收益[66]。每一个博弈过程至少存在一个纳什均衡点，红蓝双方无人机都不会在纳什均衡策略的基础上改变自己的策略。对于混合策略，纳什均衡解满足下式条件：

$$\begin{cases} \boldsymbol{M}_r \cdot (p_{m+1:m+n}^{*})' \geqslant \boldsymbol{M}_r \cdot (p_{m+1:m+n})' \\ p_{1:m}^{*} \cdot \boldsymbol{M}_b \geqslant p_{1:m} \cdot \boldsymbol{M}_b \end{cases} \tag{3.9}$$

其中，p^* 代表纳什均衡解，$p = (p_{r1}, p_{r2}, \cdots, p_{rm}, p_{b1}, p_{b2}, \cdots, p_{bn})^T$ 代表混合策略。红蓝无人机根据纳什均衡解的概率分布随机选择策略或选择概率最大值对应的策略。

至此就完成了无人机一对一动态博弈建模。

3.2　基于改进粒子群算法的博弈纳什均衡求解

无人机在对抗博弈中根据混合策略纳什均衡解选取各方机动策略完成机动决策过程。传统的纳什均衡求解方法在面对诸如无人机空战对抗这类大规模博弈场景时存在着计算复杂度高、时间长且难于实现的问题，而群体智能优化算法的研究则为求解博弈纳什均衡点提供了一种新的途径，其将纳什均衡转化为最优化问题进行寻优从而得到最优混合策略。本节通过粒子浓度的概率选择对基本群体智能优化算法进行改进，目的是提高混合策略纳什均衡求解的计算性能，降低群体寻优过程中容易陷入局部最优解的可能性，提升全局搜索寻优能力和寻优精度，进而提高无人机在对抗博弈过程中生成机动策略的精准度和实时有效性。

3.2.1　基本粒子群算法介绍

无人机对抗机动决策中最优机动策略的求解本身是一种优化问题，基于优化理论的群体智能算法可以为其提供一种求解途径。

群体智能最初起源于对蚂蚁、蜜蜂等“社会性”生物群体所呈现出的规律的研究[67]。在一个有群体行为的智能系统中，每个具有经验和智慧的个体都会受到其他个体或者环境所带来的影响，它们以相互作用完成任务或运动的形式来构建强大的群体智慧系统从而解决复杂的问题，因此群体智能表现出较好的自组织性、自学习性，且具有很强的适应环境的能力。

群体智能优化算法起源于 20 世纪 90 年代，经过多年的发展，已经出现了几种具有代表性的群智能算法，如遗传算法、粒子群优化算法、蚁群优化算法[68]。目前针对群智算法的研究也从简单的优化问题拓展到了高维度动态优化问题上，针对无人机对抗博弈最优混合策略求解这类搜索空间维度高、优化目标复杂的寻优问题，传统算法如梯度下降法、凸优化法等不再适用，其计算复杂度高，且难于实现。而生物群体中蕴含的群体智能所表现出的协作、高效的特点正好为无人机空战机动决策的生成提供了一种行之有效的方法，因此在无人机机动决策的研究中使用群智思想来求解最优机动策略显得尤为重要。本章则基于典型的粒子群算法展开研究。

基本粒子群优化（Particle Swarm Optimizaton）算法是一种基于对鸟群捕食行为研究的群体智能方法[69]。在 PSO 中每个优化问题的潜在解都作为 D 维搜索空间中的一个“粒子”，首先初始设定一群种群规模为 N 的随机粒子，使用目标函数确定所有粒子的适应值，每个粒子由飞行速度决定其运动方向和距离，根据个体和群体最好位置进行动态调整，最后在解空间中搜索迭代找到近似最优解。

设第 i 个粒子的位置为 $X_i = [x_{i1}, x_{i2}, \cdots, x_{iD}]$，速度为 $V_i = [v_{i1}, v_{i2}, \cdots, v_{iD}]$，$i = 1, 2, 3, \cdots, N$。第 i 个粒子目前找到的最优解记为 $P_i = [p_{i1}, p_{i2}, \cdots, p_{iD}]$，整个

粒子群目前找到的最优解记为 P_g，在每一次迭代过程中，第 i 个粒子的速度和位置更新根据下式进行：

$$\begin{cases} V_i^{t+1} = \omega V_i^t + c_1 r_1 \cdot (P_i^t - X_i^t) + c_2 r_2 \cdot (P_g^t - X_i^t) \\ X_i^{t+1} = X_i^t + V_i^{t+1} \end{cases} \tag{3.10}$$

其中，c_1，c_2 表示学习因子，c_1 调节个体最好粒子位置的飞行步长，c_2 调节全局最好粒子位置的飞行步长，取值范围为 [0,2]；r_1，r_2 表示 0 到 1 之间的随机数，为了在优化过程中保持群体的多样性；t 表示当前迭代更新的次数，t_{max} 表示最大迭代次数；ω 表示惯性因子，ω 从最大惯性因子 ω_{max} 随迭代过程减小到最小惯性因子 ω_{min}：

$$\omega = \omega_{max} - t \cdot \frac{\omega_{max} - \omega_{min}}{t_{max}} \tag{3.11}$$

对每个粒子按照公式 (3.10) 更新其速度和位置，直至进行到最大迭代次数 t_{max} 终止循环，得到优化结果，其算法流程如下图3.1所示。

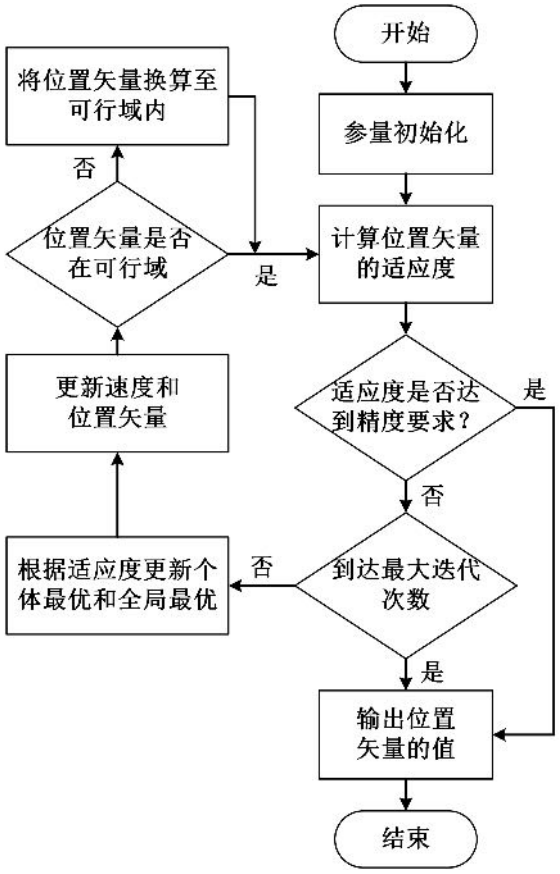


图 3.1　PSO 算法流程

3.2.2　改进粒子群算法设计

基本粒子群优化算法计算简单且具有较快的收敛速度。但由于在粒子群体迭代更新的过程中逐渐向当前全局最优位置聚集，这样就很难保证群体的多样

性，导致搜索空间存在限制使得优化结果早熟。当使用其解决一些高维优化问题时，由于搜索空间庞大且复杂度高，容易导致最终的优化结果陷入局部最优，且收敛精度较低。本文的无人机对抗博弈最优混合策略的求解便是一个搜索空间维度高、优化目标复杂的寻优问题。因此为了提高生成机动策略的准确性和计算效率，本小节对基本粒子群优化算法进行改进。

为求解无人机对抗博弈中的最优混合策略，根据纳什均衡解的特点，设计改进粒子群算法的适应度函数为：

$$\begin{aligned} J(p^t) = &\max_{1\leqslant i\leqslant m}\left\{M_r(i,:)\cdot p^t_{m+1:m+n} - (p^t_{1:m})'\cdot M_r\cdot p^t_{m+1:m+n}\right\} + \\ &\max_{1\leqslant j\leqslant n}\left\{(p^t_{1:m})'\cdot M_b(:,j) - (p^t_{1:m})'\cdot M_b\cdot p^t_{m+1:m+n}\right\} \end{aligned} \tag{3.12}$$

其中 p^t 表示 t 时刻的混合策略 p，对应一个粒子的位置状态向量。适应度函数值越小，则粒子所在位置越优，当适应度函数值等于 0 或最接近于 0 时，即认为得到了纳什均衡解，也就是最优混合策略。

定义第 i 个粒子的浓度为：

$$D_i(t) = \frac{1}{\sum_{j=1}^{N+M} J(p^t)} \tag{3.13}$$

基于上式的概率选择公式为：

$$P_i(t) = \frac{\frac{1}{D_i(t)}}{\sum_{i=1}^{N+M}\frac{1}{D_i(t)}} = \frac{\sum_{j=1}^{N+M} J(p^t)}{\sum_{i=1}^{N+M}\sum_{j=1}^{N+M} J(p^t)} \tag{3.14}$$

改进后的算法在原算法的基础上借鉴了生命科学的免疫原理[70]，将优化问题求解作为抗原，把每一个抗体视为一个粒子进而代表问题的一个解，使用 PSO 中的适应度来评估抗原与抗体之间的亲和度，粒子的多样性则由抗体之间的亲和力来反映，并使用粒子（抗体）浓度的概率选择公式来保持各适应度层次的粒子维持一定的浓度，实现控制种群多样性的目的，进而提高原算法的全局搜索寻优能力，在优化收敛阶段避免陷入局部最优值。

根据前文所述对算法的改进，利用改进粒子群算法求解无人机对抗博弈的最优混合策略的实现步骤如下：

Step1：初始化改进粒子群算法的各种参数值，如种群规模 N、自我学习因子 c_1、种群学习因子 c_2、最大迭代次数 t_{max} 等；

Step2：随机产生 N 个粒子 x_i 构成粒子群 p_0，满足：$\sum_{j=1}^{m_i} x_j^i = 1, x_j^i \geqslant 0, x_j^i \in x_i, i = 1,\cdots,N, j = 1,\cdots,m_i$，初始化粒子群速度 v_i，满足：$\sum_{j=1}^{m_i} v_j^i = 0, v_j^i \in v_i, j = 1,\cdots,m_i$；

Step3：利用适应度函数求出每一个粒子的适应度，找出粒子的个体最优解 $P_i, i = 1, 2, \cdots, N$ 和群体最优解 P_g；

Step4：根据公式 (3.11) 计算惯性因子 ω；

Step5：根据公式 (3.10) 更新粒子的速度和位置，将群体最优解 P_g 对应位置的粒子保存；

Step6：依次判断第 i 个粒子是否满足 $X_i^{t+1} > 0$，否则计算控制步长 α_t，使 $X_i^{t+1} = X_i^t + \alpha_t \cdot V_i^{t+1}$，确保每一个粒子在其混合策略空间内；

Step7：随机产生 M 个粒子，同 Step2；

Step8：由概率选择公式 (3.14) 从 $N + M$ 个粒子中选出 N 个粒子；

Step9：使用保存的粒子替换掉适应度最差的粒子，构成新粒子群 p_1，准备进入下一次迭代过程；

Step10：由最大迭代次数判断是否结束循环，输出适应度符合要求的最优粒子（即最优混合策略），否则返回 Step3。

至此完成了改进粒子群算法的设计。

3.3　群智算法性能对比实验及分析

为了更好地评估改进后算法的性能，本节针对零和与非零和博弈的混合策略纳什均衡求解设计了两个算例，然后分别使用本文给出的改进粒子群算法、基本粒子群算法和遗传算法（Genetic Algorithm, GA）进行优化求解，以适应度曲线来反映算法的收敛特性，比较分析各算法的优化结果。下表3.1给出了各算法的初始参数设置。

表 3.1　各算法的初始参数

改进粒子群算法（改进 PSO）、基本粒子群算法（PSO）					
t_{max}	最大迭代次数	80	c_1	自我学习因子	1.5
N	种群规模	70	c_2	种群学习因子	1.5
ω	惯性因子	0.6			
遗传算法（GA）					
t_{max}	最大迭代次数	80	p_c	交叉概率	0.7
N	种群规模	70	p_m	变异概率	0.02

算例一：对零和博弈的混合策略纳什均衡求解。

设置本算例的零和博弈支付矩阵如下表3.2所示，其中 S_r 和 S_b 分别表示博弈双方的策略集。

表 3.2　算例一的支付矩阵

S_r \ S_b	S_{b1}	S_{b2}	S_{b3}
S_{r1}	(8,-8)	(9,-9)	(3,-3)
S_{r2}	(2,-2)	(5,-5)	(6,-6)
S_{r3}	(4,-4)	(1,-1)	(7,-7)

表 3.3　算例一的理论混合策略纳什均衡解

	S_1	S_2	S_3
S_r	0.4038	0.2308	0.3654
S_b	0.1538	0.2308	0.6154

各算法一次优化过程中误差的变化曲线如下图3.2所示，定义误差为：

$$e(t)=\frac{\left\|p^t-p_{optimal}\right\|}{m+n} \tag{3.15}$$

其中，$p_{optimal}$ 代表混合策略纳什均衡解的理论值，如上表3.3所示，当和理论值的误差小于 0.01 时，即认为得到最优解。

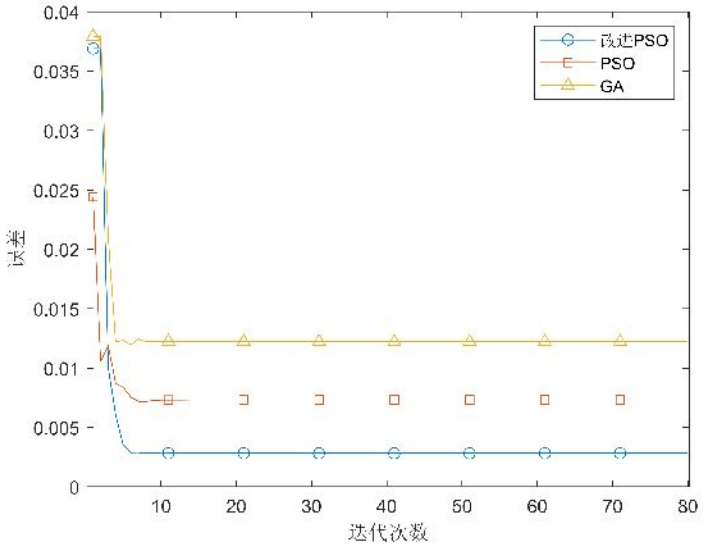


图 3.2　算例一的误差变化曲线

各算法一次寻优过程中最优适应度值的变化情况如下图3.3所示，同时为了减少随机因素可能对评估各算法性能带来的影响，又对三种算法在相同的环境下进行 100 次独立实验，各算法平均最优适应度值的变化情况如下图3.4所示。

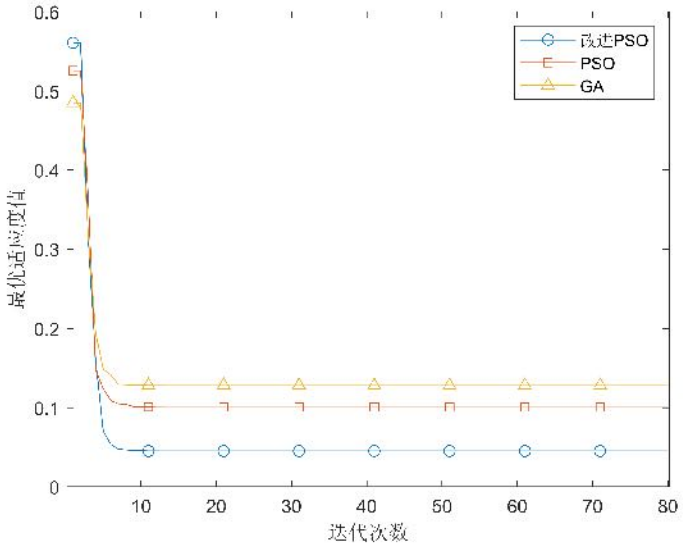


图 3.3　算例一的最优适应度值变化曲线

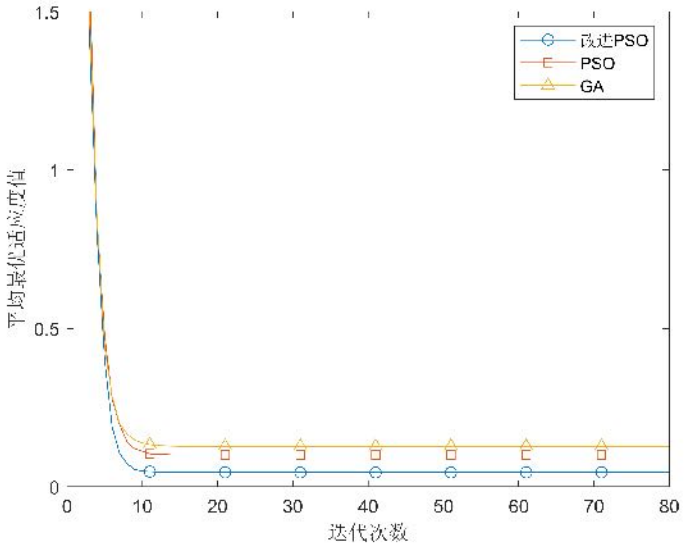


图 3.4　算例一的平均最优适应度值变化曲线

对本算例中改进粒子群算法、基本粒子群算法和遗传算法的寻优结果进行比较，从图3.2可以看出，改进 PSO 和 PSO 算法的误差均达到了 0.01 以下，得到了最优解，但改进后算法的误差相对更小，小于了 0.005，证明其得到的混合策略纳什均衡解更精确，与理论值的误差最小。从图3.3、3.4可以看出，改进 PSO 算法在寻优前期的收敛速度更快，可以表现出更好的收敛性，同时改进后算法的最优适应度值和平均最优适应度值都为最小，且更接近于 0，证明其粒子所在位置最优，得到的最优混合策略更好。

算例二：对非零和博弈的混合策略纳什均衡求解。

类似的，各算法的初始参数值如上表3.1，设置本算例的非零和博弈支付矩

阵如下表3.4所示，S_r 和 S_b 分别代表博弈双方的策略集。

表 3.4　算例二的支付矩阵

S_r \ S_b	S_{b1}	S_{b2}	S_{b3}	S_{b4}
S_{r1}	(1,1)	(235,0)	(0,235)	(0.1,1.1)
S_{r2}	(0,235)	(1,1)	(235,0)	(0.1,1.1)
S_{r3}	(235,0)	(0,235)	(1,1)	(0.1,1.1)
S_{r4}	(1.1,0.1)	(1.1,0.1)	(1.1,0.1)	(0,0)

表 3.5　算例二的理论混合策略纳什均衡解

	S_1	S_2	S_3	S_4
S_r	0.3333	0.3333	0.3333	0
S_b	0.3333	0.3333	0.3333	0

表3.5给出了非零和博弈混合策略纳什均衡解的理论值，各算法一次优化过程中误差的变化曲线如下图3.5所示。

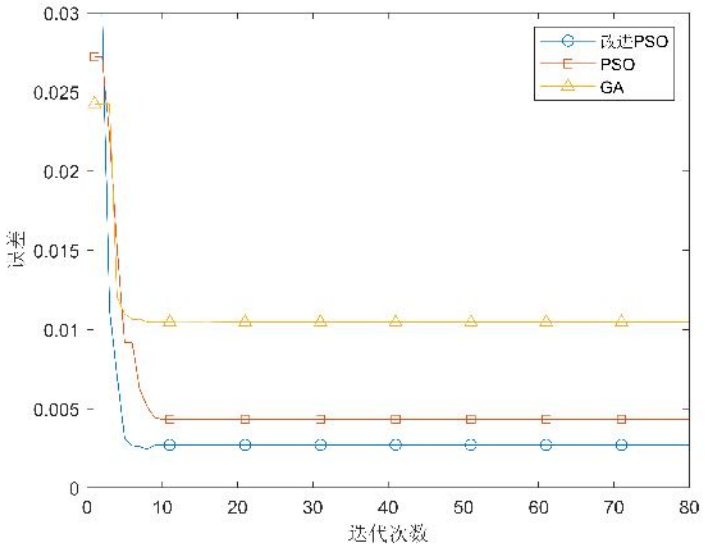


图 3.5　算例二的误差变化曲线

同样的，图3.6所示为各算法一次寻优过程中最优适应度值的变化情况，为了减少随机因素可能对评估各算法性能带来的影响，又对三种算法在相同的环境下进行 100 次独立实验，各算法平均最优适应度值的变化情况如下图3.7所示。

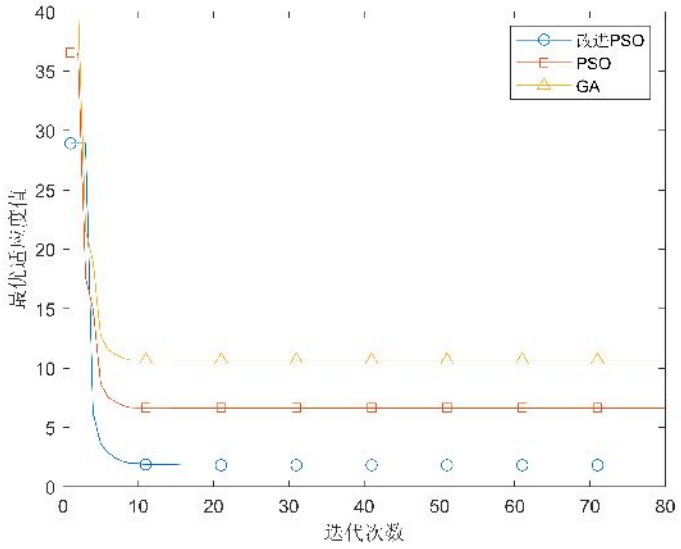


图 3.6　算例二的最优适应度值变化曲线

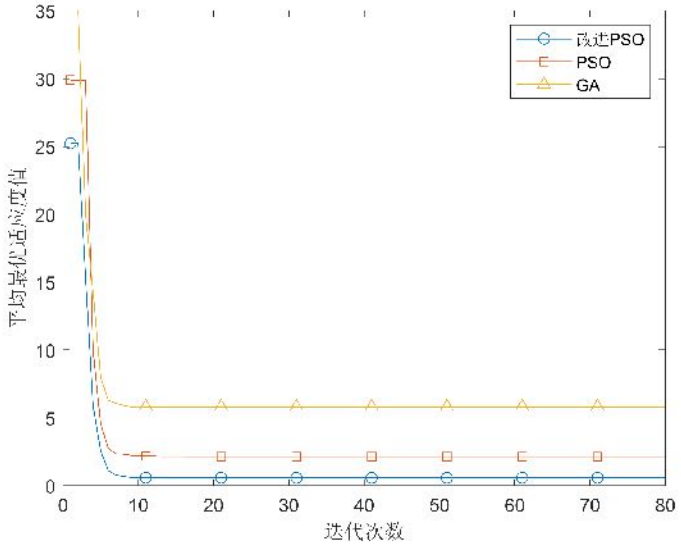


图 3.7　算例二的平均最优适应度值变化曲线

对比算例二中改进粒子群算法、基本粒子群算法和遗传算法的寻优结果，与算例一得到的结果一致。改进粒子群算法的误差更小，得到的混合策略纳什均衡解最为精确；其在寻优前期的收敛速度最快，效率最高，粒子所在位置最优，不易陷入局部最优解，找到最优解的能力最高，得到的最优混合策略更好。

综上所述，针对无人机对抗博弈最优混合策略求解这类搜索空间维度高、优化目标复杂的寻优问题，改进粒子群算法相较于原算法的性能更优，收敛精度更高，不易陷入局部最优值，同时兼顾效率和准确性，可以在无人机对抗机动决策中得到很好的应用。

3.4　一对一机动决策仿真实验及分析

为了验证改进后算法生成机动决策的有效性，本节使用 MATLAB 构建对抗仿真环境，并在三种典型的对抗场景下模拟红蓝无人机一对一近距对抗博弈。在仿真实验中，红蓝双方无人机都采用相同的 F-16 无人机模型，限制其最大飞行速度为 500m/s。红蓝双方采取不同的机动决策方法，并使用模拟机炮相互攻击，定义其有效攻击范围 $r_d = 800\text{m}$，攻击时的命中率与方位角有关，如下图3.8所示，当敌方无人机位于己方攻击范围内时，如果 $\lambda < 10\,\text{deg}$，命中率为 95%；如果 $10\,\text{deg} < \lambda < 20\,\text{deg}$，命中率为 90%，在其他情况下均无法击中目标，$\lambda$ 代表视线角，定义为自身机体坐标系的 x_b 轴与双方机体之间连线形成的夹角。根据攻击范围和命中率，如果判定红蓝无人机其中一架被击中，则一对一近距对抗仿真实验结束，仿真中无人机机动决策的周期为 2 秒，仿真周期为 0.1 秒。

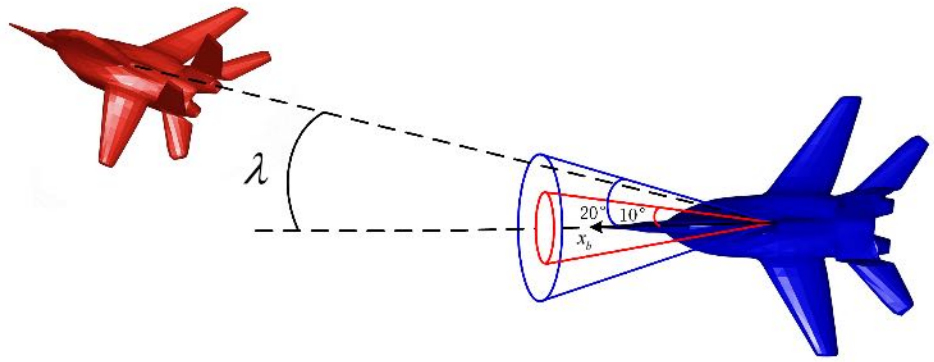


图 3.8　对抗攻击范围

3.4.1　改进粒子群对抗极小化极大算法

在本小节仿真实验中，双方采取不同的机动决策方法，红方无人机根据改进粒子群（PSO）算法求解得到的最优混合策略来选择机动动作进行机动决策，而蓝方无人机采取极小化极大（Minimax）算法生成机动策略。Minimax 算法是一种用于解决博弈类问题的智能决策算法，其核心思想是在决策中找到对手让自己陷入最坏情况的各种策略中最好的策略，以尽可能地减小自己损失。该算法基于决策树和搜索[71-72]，原理如下图3.9所示，可见 Minimax 算法以敌我双方完全信息为决策依据，因此使用其决策得到的结果比较具有优势，可以生成有效针对敌方的策略，而且该算法容易实现，所以本小节选取 Minimax 算法生成蓝方无人机的机动策略来对抗红方。

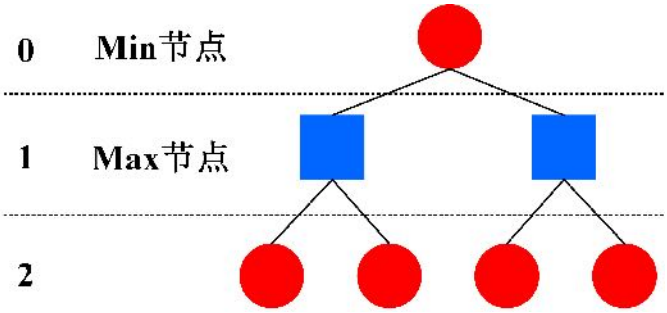
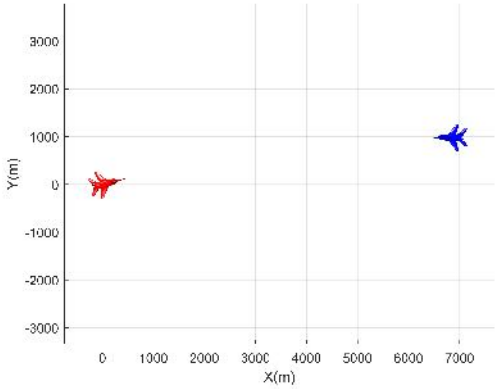


图 3.9　Minimax 算法原理

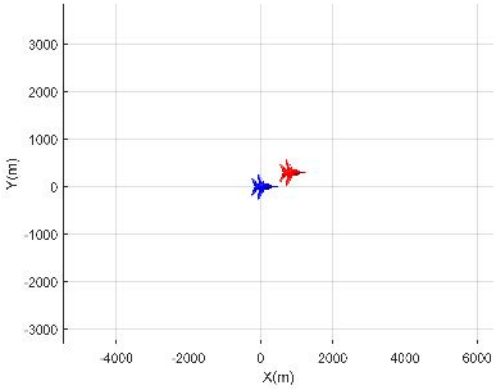
红蓝无人机一对一近距博弈在以下三种典型的对抗场景中展开：迎头进攻、尾追进攻、侧方进攻。基于不同的对抗场景设定双方无人机的初始状态如下表3.6，其中 x、y、z 代表双方无人机的初始位置，V 代表初始速度，ψ 代表初始航向角。红蓝无人机初始状态的俯视图如下图3.10所示。

表 3.6　无人机一对一场景初始状态参数

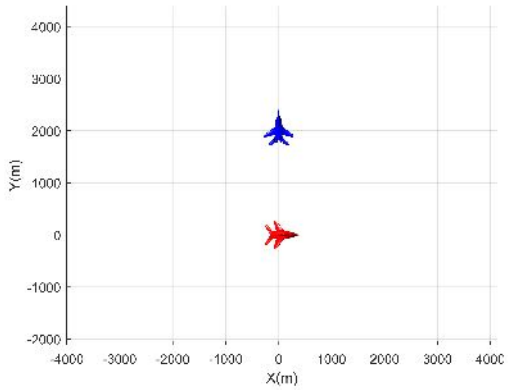
		x(m)	y(m)	h(m)	V(m/s)	ψ(rad)
迎头进攻	红方	0	0	3300	152	$\pi/12$
	蓝方	7000	1000	3300	152	π
尾追进攻	红方	800	300	3300	170	0
	蓝方	0	0	5000	170	0
侧方进攻	红方	0	0	3300	160	0
	蓝方	0	2000	3300	160	$\pi/2$



(a) 迎头进攻



(b) 尾追进攻



(c) 侧方进攻

图 3.10　初始状态俯视图

（一）迎头进攻：仿真结果如下图3.11所示，从图中可以看出，红蓝无人机初始位置处于正面迎敌状态，起初双方在相同高度，且都位于对方的攻击角度 λ_e 之内，优势相当，此时两方无人机相距较远，接着红蓝无人机迅速机动相向飞行以拉进彼此作战距离，企图到达对方的有效攻击范围，为避免正面交锋，红方向左滚转，蓝方则绕下企图到达红方尾部占据可攻击位置，红方则频繁采取机动几次逃出蓝方的攻击角度，与蓝方无人机拉开距离后迅速俯冲，从蓝方下侧将其击落。

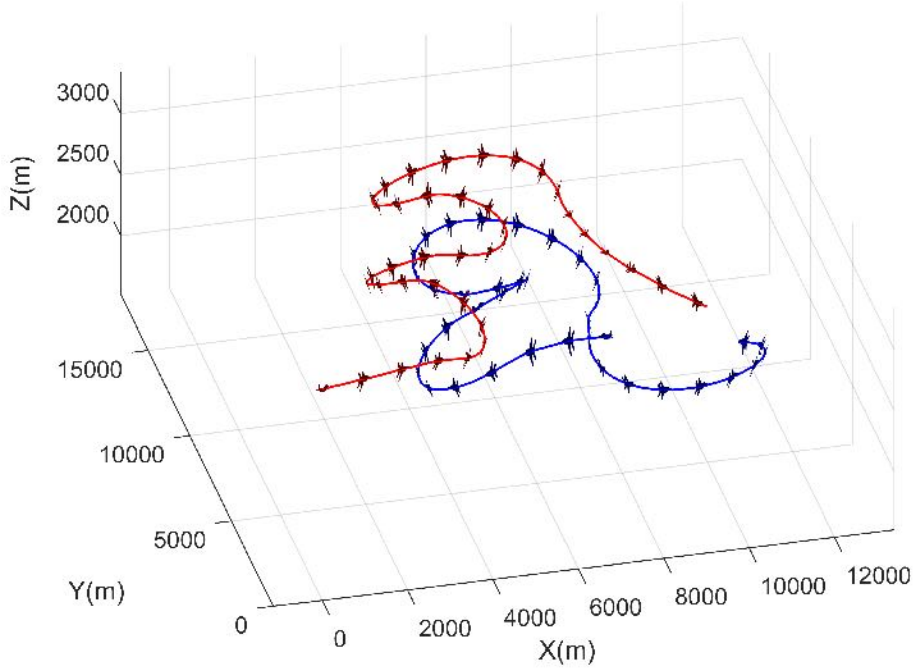


图 3.11　对抗机动轨迹

（二）尾追进攻：仿真结果如下图3.12所示，红蓝无人机初始位置处于尾追态势，起初蓝方在红方尾部，两机距离较近，且高度绝对占据优势，红方处于劣势地位，红方快速机动前进并以较大过载爬升以扭转不利态势，蓝方则向下俯冲追击红方以期到达红方尾部攻击区域，由于其俯冲速度过快双方拉开一定距离，红方拉高后采取迅速机动反应，先发制人，俯冲追击从其后侧上方将蓝方锁定后击落。

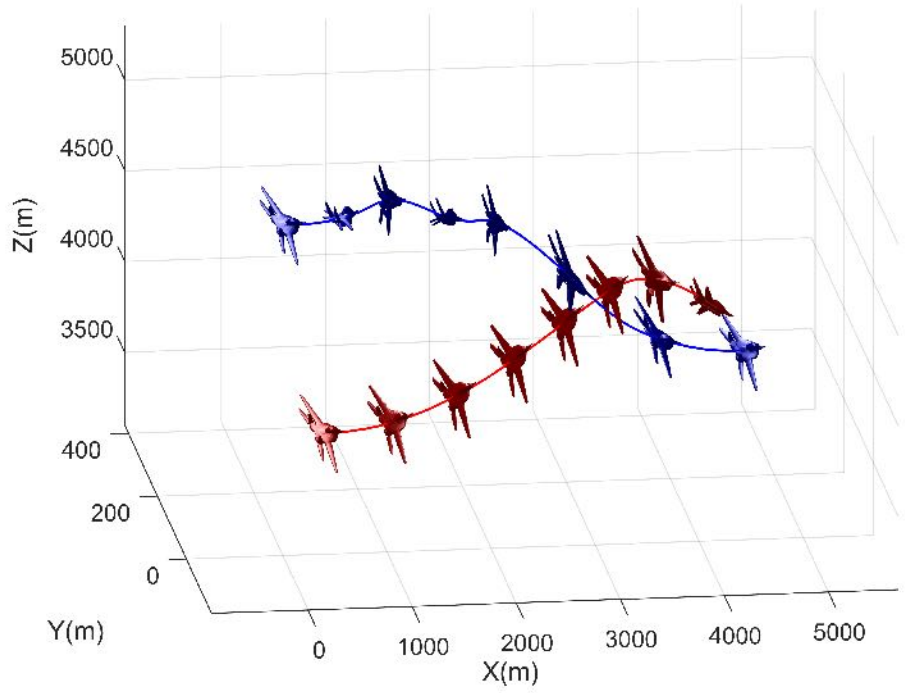


图 3.12　对抗机动轨迹

（三）侧方进攻：仿真结果如下图3.13所示，初始位置红蓝无人机处于相同的高度，双方均势，红方在观测到蓝方位置后迅速调头向左滚转以追击蓝方，蓝方则向下俯冲利用角度差穿越红方，红方随之向下俯冲并始终占据高位，蓝方频繁机动躲避红方攻击，最后红方在蓝方向左滚转时，利用高度优势将其锁定在有效攻击距离和角度，从蓝方后侧上方将其击中，一举获胜。

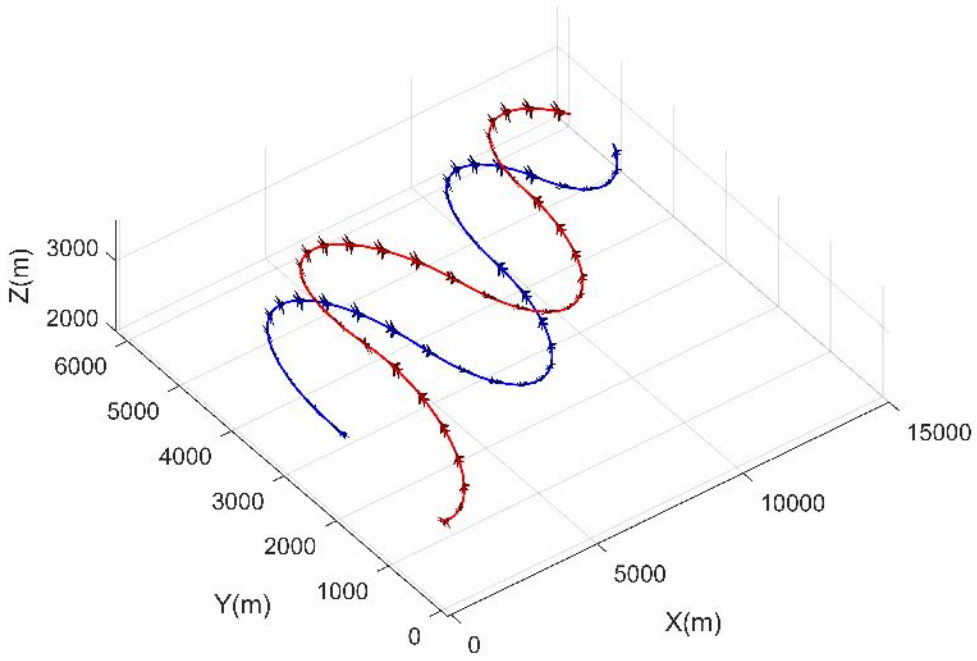


图 3.13　对抗机动轨迹

表 3.7　仿真实验中计算时长对比

场景	一个仿真周期中的平均计算时长 (s)					
	红方			蓝方		
	模型更新时长	决策算法	总计	模型更新时长	决策算法	总计
一	0.210	0.113	0.323	0.238	0.314	0.552
二	0.194	0.096	0.290	0.211	0.297	0.508
三	0.206	0.124	0.330	0.194	0.301	0.495

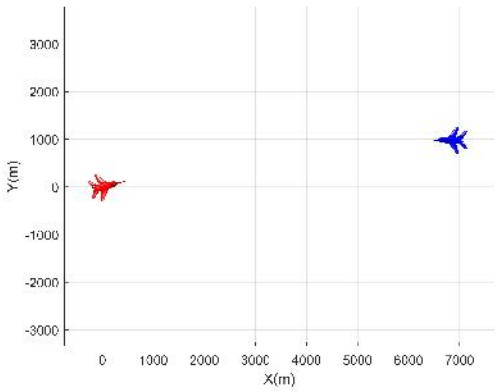
上表3.7为本小节仿真实验中各决策算法及模型更新消耗的平均计算时长，可以直观的反映出改进 PSO 算法在对抗基于博弈的 Minimax 算法时生成机动决策的速度更快，同时采用改进后算法生成机动策略的红方无人机在三种不同的对抗场景中，即便在初始态势劣势的情形下，也能扭转不利局面借机取胜。

3.4.2　改进粒子群对抗基本粒子群算法

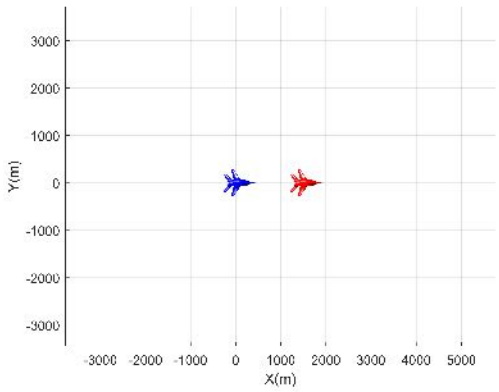
本小节中红方无人机仍采取改进后的算法，而蓝方无人机采取基本粒子群算法求解得到的最优混合策略来选择机动动作进行机动决策。同样的，基于三种典型的对抗场景设定双方无人机的初始状态如下表3.8所示，红蓝无人机初始状态的俯视图如下图3.14所示。

表 3.8　无人机一对一场景初始状态参数

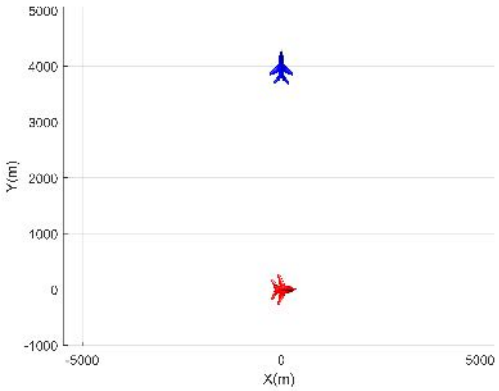
		x(m)	y(m)	h(m)	V(m/s)	ψ(rad)
迎头进攻	红方	0	0	3300	152	$\pi/12$
	蓝方	7000	1000	3900	152	π
尾追进攻	红方	1500	0	3300	170	0
	蓝方	0	0	5000	170	0
侧方进攻	红方	0	0	3300	160	0
	蓝方	0	4000	3900	160	$\pi/2$



(a) 迎头进攻



(b) 尾追进攻



(c) 侧方进攻

图 3.14　初始状态俯视图

（一）迎头进攻：仿真结果如下图3.15所示，从图中可以看出，红蓝无人机初始位置处于正面迎敌状态，且蓝方高度占优。蓝方迅速机动前进企图利用高度优势俯冲直击目标，红方爬升前进一段距离后为避免陷入角度劣势地位率先大幅度扭转机头向右滚转与蓝方无人机重新拉开距离，规避威胁的同时借机绕后准备进行反击，蓝方由于俯冲速度过快飞跃红方后以较大过载快速爬升改变航向，

第一次交锋无果后的红方率先占据高位优势乘胜追击，迅速达到攻击条件从上方正面迎敌击落蓝方，结束对抗交锋。

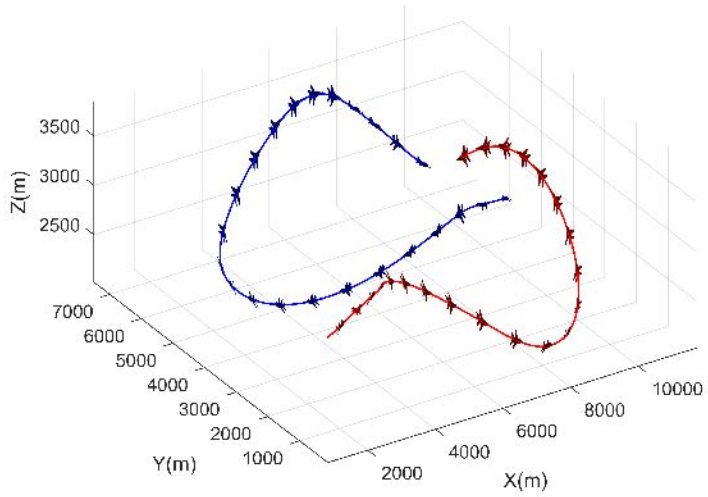


图 3.15　对抗机动轨迹

（二）尾追进攻：仿真结果如下图3.16所示，设置红蓝无人机初始状态为尾追态势，蓝方刚开始在位置和高度上绝对占领优势地位，红方劣势。红方通过右转爬升来躲避蓝方的俯冲直击，两方频繁机动在空中战场拉锯，最后红方采取爬升调转航向绕其后飞行，并迅速俯冲拉进与蓝方距离，满足可攻击条件时先发制人，快速做出反应击落目标。

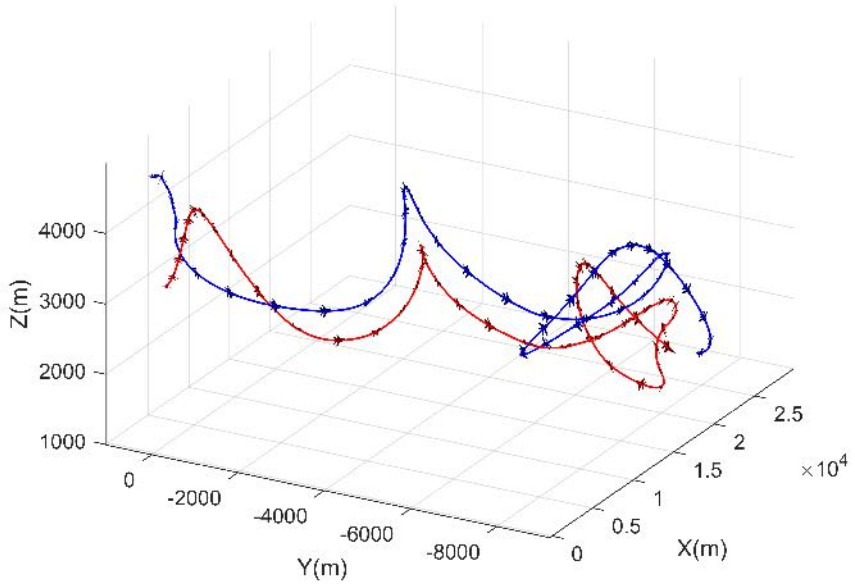


图 3.16　对抗机动轨迹

（三）侧方进攻：仿真结果如下图3.17所示，起初蓝方无人机的高度比红方占优，红方无人机在观测到蓝方位置后调转航向以较大速度追击，蓝方则向右俯冲企图对红方进行攻击，第一次交锋无果后，双方盘旋企图占据对方可攻击位置，为扭转当前战场态势，红方快速反应向下俯冲后又继续爬升，借机到达蓝方无人机身后，并稳定地将其锁定在有效攻击范围内一段时间，最后从蓝方后侧将其击落取得胜利。

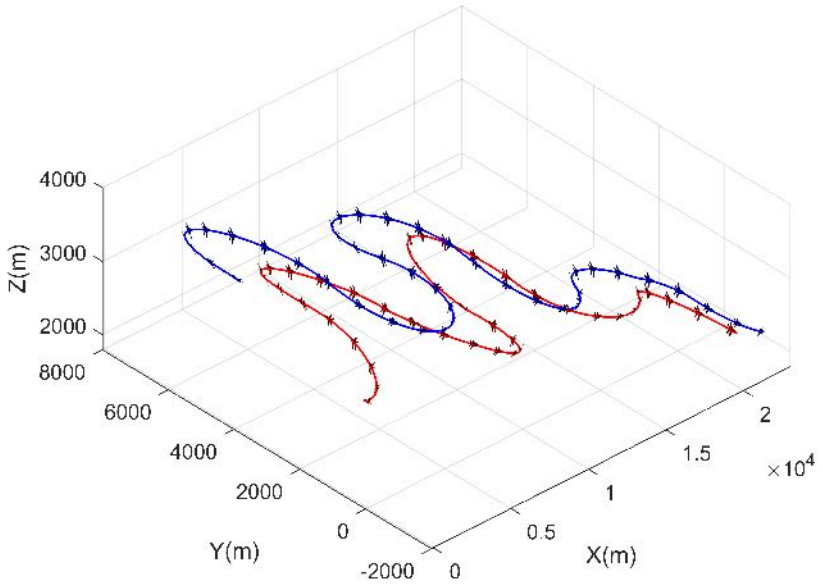


图 3.17　对抗机动轨迹

表 3.9　仿真实验中计算时长对比

场景	一个仿真周期中的平均计算时长 (s)					
	红方			蓝方		
	模型更新时长	决策算法	总计	模型更新时长	决策算法	总计
一	0.254	0.106	0.360	0.247	0.202	0.449
二	0.236	0.102	0.338	0.255	0.189	0.444
三	0.248	0.117	0.365	0.240	0.214	0.454

上表3.9为本小节仿真实验中各决策算法及模型更新消耗的平均计算时长，直观地反映出改进后算法在对抗原算法时生成机动决策的速度更快，同时从另一方面证明了其优化收敛性能更好，效率更高。

综上所述，基于改进粒子群算法的无人机机动决策方法在与其他算法进行比较的过程中都表现出了更高的决策水平，生成的机动策略更准确有效且实时

性更高。采用改进后算法生成机动决策的红方无人机尽管在三种典型对抗场景中的初始状态都被设置为劣势，其仍然能够得到最优混合策略从而扭转劣势态势取得对抗博弈胜利，同时兼顾了效率和准确性，可以在无人机机动决策中得到很好的应用。

3.5　本章小结

本章主要利用基于优化理论的群智算法对无人机对抗机动决策展开研究。首先建立了无人机一对一场景的动态博弈模型，接着在基本粒子群算法的基础上，通过粒子浓度的概率选择，提出了改进粒子群算法求解博弈的最优混合策略，由此指导无人机进行机动策略的选择，然后将改进后算法与其他两种算法进行性能对比，仿真结果表明了改进后算法在寻优过程中的收敛速度最快，效率最高，不易陷入局部最优解，并且得到的最优混合策略最为精确。最后将其应用到无人机对抗机动决策中，分别进行了两组无人机一对一机动决策对比实验，在仿真实验中，使用改进后算法进行机动决策的无人机相较于其他两种都表现出了更高的决策水平，即使在初始态势劣势的情形下仍然能够得到最优混合策略，扭转不利局面取得最终胜利，其生成的机动策略更准确有效且实时性高，进一步证明了改进后算法应用于无人机机动决策的有效性。

第 4 章　基于博弈及深度强化学习的无人机机动决策

本章基于人工智能技术的深度强化学习算法对无人机机动决策方法进行研究，首先介绍强化学习方法中的马尔可夫决策过程，由该理论扩展建立了无人机一对一场景的二人零和马尔可夫博弈模型，设计了一对一场景下的基本状态空间、动作空间和奖励函数，并结合博弈论的思想提出了一种改进传统深度强化学习的算法，使其更适用于具有复杂冲突对抗性的博弈环境，能够快速生成有效针对对手的最优机动策略，本章最后对三组无人机一对一机动决策进行仿真实验并分析对比实验结果，以验证改进后算法生成最优机动策略的有效性。

4.1　基于马尔可夫博弈的无人机一对一模型

强化学习方法的核心思想是智能体通过执行动作，采取试错的方式与环境进行交互，在探索环境的过程中依据当前状态下执行动作后反馈的奖励回报值来评估动作选择的结果并生成策略。强化学习系统根据当前状态、动作、下一状态、即使奖励回报作为信息评估动作的好坏，通过最大化累积奖励值进行学习。因此，在无人机对抗机动决策问题中，强化学习得到的策略在进行决策时既考虑到了当前状态对空战态势的影响，又考虑到了无人机的机动动作对未来战场态势的长期影响，满足了高动态强对抗空战环境中的不确定性，使决策选择的动作不仅在当前状态下较为合理，还对空战环境有很强的适应性。此外，强化学习方法不需要由人类空战对抗生成的训练样本数据，而是由智能体在学习过程中自行探索环境进而生成样本数据，依靠其自学习能力快速生成准确可靠的机动策略，同时满足对抗的实时性要求，因此，强化学习方法可以有效的应用于无人机机动决策的研究中。

4.1.1　马尔可夫决策过程介绍

针对强化学习（Reinforcement Learning）问题，马尔可夫决策过程[73]（Markov Decision Process，MDP）是其在数学模型上的表达方式。强化学习的基本过程是，智能体（Agent）在执行某项任务时，在状态（State）S 下选择动作（Action）A 与环境（Environment）进行交互，然后到达下一新的状态 S'，智能体会根据奖励函数得到即时奖励回报值（Reward）R，智能体通过不断地与环境交互进而得到环境状态变化和奖励值的数据，并根据这些数据不断地改进自身的动作策略，再采取改进后的策略与环境交互生成新的数据，如此反复进行多次迭代学习后直到策略稳定收敛，最后获得最优决策策略，使得智能体通过使用该策略能够在

环境中获得最大的奖励回报值。

马尔可夫决策过程可以用来描述和求解时序决策问题[74]，作为强化学习中最基础的理论，在马尔可夫决策过程中，智能体得到的即时奖励回报值和过去时间的状态以及动作都没有直接的联系。因此，马尔可夫过程在决策时是一个无记忆的过程，其不必考虑历史信息，在单位时间内的决策次数定义为决策频率。

无人机对抗机动决策过程是根据当前战场态势选取动作的时序决策过程，该过程可以被离散化。无人机对抗机动决策的目的是执行当前时刻最合适的机动动作进而使未来作战态势达到对己方最有利的状态，选取的机动动作与历史时刻的状态无关，而机动策略所关注的奖励信息则包含在累积奖励回报中，因此，无人机对抗机动决策过程可以用马尔可夫决策过程来表示。

一个基本的马尔可夫决策过程中的要素可以用五元组 $(\boldsymbol{S}, \boldsymbol{A}, \boldsymbol{R}, \boldsymbol{P}, \gamma)$ 来表示，其中各参数的定义如下：

（1）$\boldsymbol{S}$ 表示状态空间，是系统状态的集合；

（2）$\boldsymbol{A}$ 表示动作空间，是智能体可选取动作的集合；

（3）$\boldsymbol{R}$ 表示奖励函数，是即使奖励回报的集合，一般描述为 $\boldsymbol{R}(s, a, s')$，奖励函数用来计算智能体在当前状态 s 下，选取动作 a 使状态转变为 s' 后得到的即时奖励回报值；

（4）$\boldsymbol{P}$ 表示状态转移概率，是马尔可夫决策过程的动态特性体现，一般情况下可以描述成一个矩阵，在矩阵中单个元素 $\boldsymbol{P}(s, a, s')$ 的取值范围为 [0,1]，是环境的精确数学模型，用来表示智能体在当前状态 s 下执行动作 a 后，使状态转移到下一新状态 s' 的概率；

（5）γ 表示折扣因子，$\gamma \in [0, 1]$。

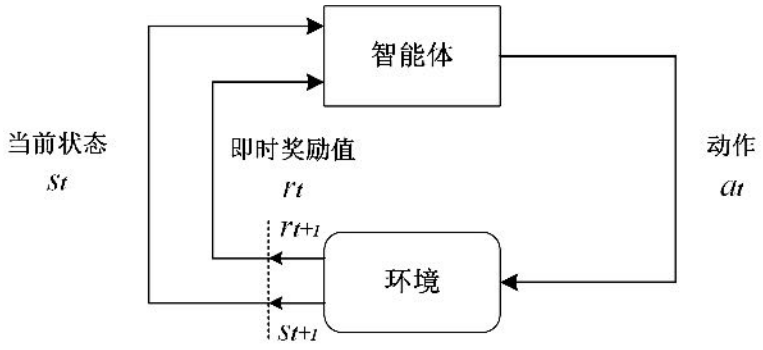


图 4.1　马尔可夫决策过程

如上图4.1所示，在每一个离散的时刻 $t = 0, 1, 2\cdots$，智能体观测到所处环境当前状态的特征 $s_t \in \boldsymbol{S}$，跟据此观测选取一个动作 $a_t \in \boldsymbol{A}$。作为执行该动作的结果，智能体在下一时刻获得一个标量奖励值 $r_{t+1} \in \boldsymbol{R}$，然后系统进入下一新的状态 $s_{t+1} \in \boldsymbol{S}$。

在马尔可夫决策过程中，状态和奖励都具有确定的概率分布函数，给定当前状态和动作的值 s 和 a 时，状态和奖励的特定值 s' 和 r 出现的概率即为状态转移概率 $P(s',r|s,a)$。该概率仅仅取决于当前时刻的状态 s 和 a，与历史更早之前的状态和动作没有任何关系，即体现了马尔可夫性。

从单个智能体的角度来考虑，强化学习的目标是根据奖励函数找到最优策略 π：

$$\pi(a|s) = p[A_t = a|S_t = s] \tag{4.1}$$

将 t 时刻后智能体得到的奖励序列表示为 $r_{t+1}, r_{t+2}, \cdots$，定义期望奖励回报 G_t 为奖励序列的特定函数，智能体的目标即为最大化期望奖励回报 G_t。考虑最简单的情况，G_t 表示为即时奖励的总和：

$$G_t = r_{t+1} + r_{t+2} + \cdots + r_T \tag{4.2}$$

其中，T 是最终时刻，由于多数情况下，智能体与环境的交互是长期进行的任务，因此不一定存在最终时刻，为了避免 G_t 趋于无穷大，通常使用折扣因子 γ 来计算折扣奖励，代替了直接累加的奖励回报值：

$$G_t = r_{t+1} + \gamma \cdot r_{t+2} + \gamma^2 \cdot r_{t+3} + \cdots = \sum_{k=0}^{\infty} \gamma^k \cdot r_{t+k+1} \tag{4.3}$$

其中，折扣因子 γ 反映了当前奖励和未来奖励间的平衡关系。当 $\gamma = 0$ 时，智能体仅关注当前奖励，其目标是学习如何选取动作来最大化 r_{t+1}，通常最大化当前奖励会减小未来获得的奖励，导致总奖励回报值变少。当 γ 越接近于 1 时，折扣奖励将会更多地考虑未来的奖励回报。相邻时刻的奖励回报可用下面的递归方式进行相互联系：

$$G_t = r_{t+1} + \gamma \cdot r_{t+2} + \gamma^2 \cdot r_{t+3} + \cdots = r_{t+1} + \gamma(r_{t+2} + \gamma \cdot r_{t+3}) + \cdots = r_{t+1} + \gamma G_{t+1} \tag{4.4}$$

由于策略是基于概率分布的，因此累积奖励回报值是随机变量，为了描述累积奖励值，则使用其期望，表示为状态值函数：

$$v_\pi(s) = E_\pi \left[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} | S_t = s \right] \tag{4.5}$$

动作值函数：

$$q_\pi(s,a) = E_\pi \left[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} | S_t = s, A_t = a \right] \tag{4.6}$$

计算状态值函数可评估智能体所处环境状态的优劣，通过对比两个状态优劣间的差值进而评估动作的好坏，最后使智能体在迭代过程中找到最优策略。

4.1.2　二人零和马尔可夫博弈模型

前面小节已经介绍了马尔可夫决策过程，马尔可夫博弈[75] 则是马尔可夫决策过程的一个扩展，其融合了博弈论的元素，允许多个智能体相互竞争来建立完成其任务的系统模型，如下图4.2所示。

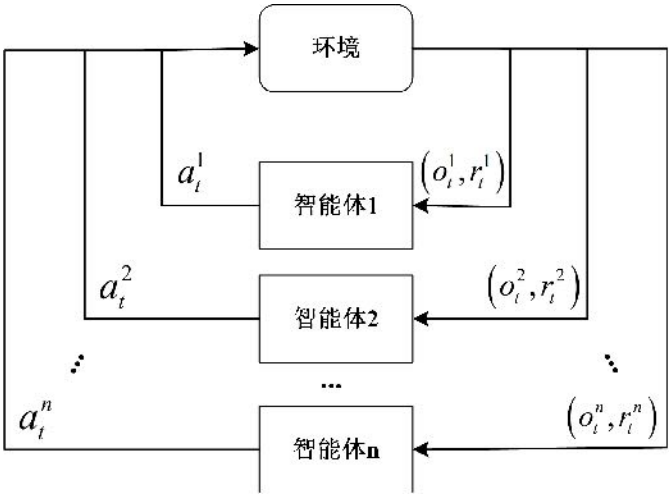


图 4.2　马尔可夫博弈模型

首先简要介绍马尔可夫博弈的形式化定义，一个包含 k 个智能体的马尔可夫博弈模型通常被定义为：

（1）S：环境状态。

（2）$A_1 ... A_k$：k 个智能体的动作空间，其中第 i 个智能体的动作空间为 A_i，$1 \leqslant i \leqslant k$。

（3）T：由所有智能体的当前状态和下一步动作决定的状态转移函数，如果环境是随机的，其函数关系表示如下：

$$T : S \times S \times A_1 \times ... \times A_k \to [0, 1] \tag{4.7}$$

（4）$R_1 ... R_k$：k 个智能体的奖励函数，对于第 i 个智能体 $1 \leqslant i \leqslant k$，其函数关系表示如下：

$$R_i : S \times A_1 \times ... \times A_k \times S \to \mathbb{R} \tag{4.8}$$

第 i 个智能体的目标是找到一条策略 π_i，使其最终累积奖励值 G_i 最大化。

$$G_i = \mathbb{E}_{s \sim \rho^{\pi_1, \pi_2 ... \pi_k}, a \sim \pi_i}[\sum_{t=0}^{T} \gamma^t R_i^t] \tag{4.9}$$

其中，$\rho^{\pi_1, \pi_2 ... \pi_k}$ 是给定 k 个智能体策略 $\pi_1, \pi_2 ... \pi_k$ 的稳定分布概率，T 是每轮长度，$\gamma \in [0, 1]$ 是折扣因子，R_i^t 是第 i 个智能体在第 t 步得到的奖励值。

二人零和马尔可夫博弈[76] 是马尔可夫博弈的一个特例，仅仅考虑了两个参与者，即智能体和对手，他们具有相反的奖励函数和对称的效用函数，可以表

示为一个五元组 (S,T,A,O,R)，其中，S 是环境状态集合，T 是状态转移函数，$T(s,a,o,s')$ 定义了在当前状态 s 下智能体执行动作 a，对手执行动作 o 后到达下一状态 s' 的转移函数；A 和 O 分别为智能体和对手的动作空间。R 是智能体的奖励函数，$-R$ 是对手的奖励函数。

据此，本章的无人机一对一机动决策可以使用二人零和马尔可夫博弈理论作为该模型框架。

4.2　对抗环境相关设计

本节主要设计了对抗仿真环境下的基本状态空间、动作空间和奖励函数，为后续机动决策方法的研究和仿真实验的进行打下基础。

4.2.1　状态空间设计

本章在研究无人机对抗的机动决策方法时，为了简化研究过程，仅注重无人机在对抗过程中的航迹而不考虑其姿态控制，因此使用由无人机六自由度模型简化而来的三自由度模型[77] 对其状态空间进行设计。

在六自由度模型中，必须研究无人机三个轴向的旋转速度，计算无人机的侧滑角和攻角。目前，机载现代飞行控制系统的无人机都装有转弯协调系统，能够使无人机自动消除侧滑角以提高飞行性能，所以侧滑角可以忽略不计，在此基础上对六自由度模型进行简化。根据下式 (4.10)，前三个单元是运动学方程，中间是动力学方程，最后一个是油耗方程。

$$\frac{d}{dt}\begin{bmatrix} X \\ Y \\ h \\ V \\ \psi \\ \gamma \\ W \end{bmatrix} = \begin{bmatrix} V\cos\psi\cos\gamma + \omega_x \\ V\sin\psi\cos\gamma + \omega_y \\ V\sin\gamma + \omega_z \\ \dfrac{g}{W}[(T\cos(\theta-\gamma) - D) - W\sin\gamma] \\ \dfrac{g\sin\phi}{WV}[L + T\sin\alpha] \\ \dfrac{g}{WV}[(L + T\sin(\theta-\gamma))\cos\phi - W\cos\gamma] \\ -CT \end{bmatrix} \tag{4.10}$$

其中，X,Y,h 表示无人机的位置，V 是速度标量，γ 是航迹俯仰角，W 是无人机质量，ψ,θ,ϕ 分别是偏航角、俯仰角和滚转角，$\omega_x,\omega_y,\omega_z$ 分别表示 x,y,z 轴上的风速，L 是升力，D 是阻力，α，β 分别是攻角和侧滑角，C，T 分别是油耗比和发动机推力，重力加速度 $g = 9.81\text{m/s}$。

为进一步简化动力学方程，假设无人机的速度方向与机体坐标系 x 轴相同，并且在计算姿态和航迹时不考虑无人机的攻角和侧滑角[78]，将推力视为无人机

纵向方向的矢量，那么方程 (4.10) 可改写为：

$$\frac{d}{dt}\begin{bmatrix} X \\ Y \\ h \\ V \\ \psi \\ W \end{bmatrix} = \begin{bmatrix} V\cos\psi + \omega_x \\ V\sin\psi + \omega_y \\ V\sin\gamma + \omega_z \\ \frac{g}{W}[(T-D) - W\sin\gamma] \\ gL\sin\phi \\ WV \\ -CT \end{bmatrix} \tag{4.11}$$

根据改写后的方程 (4.11)，不考虑无人机的俯仰角速率和偏航角速率，同时忽略无人机的油耗将其质量视为恒定值，用切向过载 n_x 表示推力 T、阻力 D，用法向过载 n_z 表示升力 L，得到简化后的三自由度模型：

$$\begin{cases} \dfrac{dv}{dt} = g(n_x - \sin\gamma) \\ \dfrac{d\gamma}{dt} = \dfrac{g}{v}(n_z\cos\phi - \cos\gamma) \\ \dfrac{d\psi}{dt} = \dfrac{g}{v\cos\gamma}n_z\sin\phi \\ \dfrac{d\phi}{dt} = n_\varphi \end{cases} \tag{4.12}$$

其中，n_φ 表示滚转角速率。然后根据动力学方程得到无人机状态值的微分，使用无人机的运动学微分方程：

$$\begin{cases} \dfrac{dx}{dt} = v\cos\gamma\cos\psi \\ \dfrac{dy}{dt} = v\cos\gamma\sin\psi \\ \dfrac{dz}{dt} = v\sin\gamma \end{cases} \tag{4.13}$$

在已知初始速度 v，初始俯仰角 γ，初始航向角 ψ 的情况下，通过积分求解得出无人机在地面坐标系中的位置坐标。

本章设定的场景为红蓝无人机在机动对抗过程中高度变化不大，无人机发生滚转运动时协调转弯控制过载定高飞行，那么：

$$\frac{d\gamma}{dt} = \frac{g}{v}(n_z\cos\phi - \cos\gamma) = 0 \tag{4.14}$$

可得法向过载：

$$n_z = \frac{1}{\cos\phi} \tag{4.15}$$

最终得到简化后的无人机动力学与运动学方程：

$$\begin{cases} \dfrac{dx}{dt} = v\cos\psi \\ \dfrac{dy}{dt} = vsin\psi \\ \dfrac{d\psi}{dt} = \dfrac{g\tan\phi}{v} \\ \dfrac{dv}{dt} = gn_x \\ \dfrac{d\phi}{dt} = n_\varphi \end{cases} \tag{4.16}$$

在简化模型中，法向过载 n_z 由滚转角 ϕ 来控制。

根据上文所述，无人机一对一场景的系统状态是由双方当前的运动状态参数、相对位置和角度参数以及未来运动状态参数定义的。首先考虑两架无人机在对抗中当前的运动状态参数，假设有红蓝两架无人机，分别记为 UAV_r 和 UAV_b，由双方无人机的状态构成的对抗环境基本状态空间为：

$$S = [x_b, y_b, x_r, y_r, v_b, v_r, \phi_b, \phi_r, \psi_b, \psi_r] \tag{4.17}$$

其中，(x_r, y_r) 和 (x_b, y_b) 分别表示红蓝无人机当前的位置坐标，无人机的位置变量不作限制，因此可以在 $x-y$ 平面内的任意位置飞行。v_r,v_b 分别表示红蓝无人机的当前速度，根据无人机的失速和最大速度限制在一定速度范围内飞行。ψ_r,ψ_b 和表示两架无人机当前的航向角，取值范围为 ±180°。ϕ_r 和 ϕ_b 表示两架无人机当前的滚转角，受限于实际无人机的最大转弯机动性能。

4.2.2　动作空间设计

无人机一对一博弈是一个动态的、连续的对抗过程，在当前状态下，无人机根据动力学方程解算下一时刻的状态值。本章以前文所述的基于“基本操纵动作库”设计的机动动作集作为其动作空间。由于单机对抗发生在同一水平面内，因此使用两个连续的控制变量来操纵无人机的机动。在二维平面内，不考虑俯仰角速率和偏航角速率，动作空间如式 (4.18) 表示：

$$A = (u_t, u_{\dot{\varphi}}) \tag{4.18}$$

其中，u_t 表示推力，作为输入控制无人机的飞行速度；$u_{\dot{\varphi}}$ 表示滚转角速率，控制无人机的转弯机动，其上限和下限取决于实际无人机的性能。两者将共同影响无人机的加速和转向机动。

4.2.3　奖励函数设计

根据实际对抗场景，蓝方无人机的目的是进入红方无人机尾部的攻击区域，并能稳定锁住对手，进而发射近距格斗空空导弹，有效攻击红方。为了方便描述奖励函数，使用如下图4.3所示的战斗几何和参数。

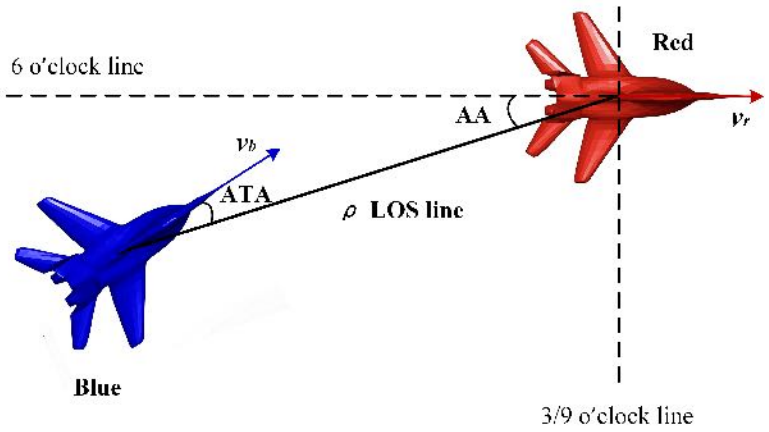


图 4.3　相对几何关系

图4.3中，蓝方无人机的目标方位角（ATA）是视线（LOS）向量与其速度方向的夹角，代表了跟踪红方无人机的程度。蓝方的进入角（AA）是视线（LOS）向量与红方速度方向的夹角，反应了其追踪敌方的稳定程度。ATA 和 AA 的取值范围都在 ±180° 之间，其计算方法如下式：

$$\mathrm{ATA} = \cos^{-1}\left[\frac{V_b \cdot \rho}{|V_b||\rho|}\right] \tag{4.19}$$

$$\mathrm{AA} = \cos^{-1}\left[\frac{V_r \cdot \rho}{|V_r||\rho|}\right] \tag{4.20}$$

两架无人机的 ATA 和 AA 满足以下几何关系[79]：

$$\begin{aligned} |\mathrm{ATA}_b| + |\mathrm{AA}_r| &= 180^\circ \\ |\mathrm{ATA}_r| + |\mathrm{AA}_b| &= 180^\circ \end{aligned} \tag{4.21}$$

无人机一对一场景被建模为二人零和马尔可夫博弈问题，属于严格性竞争博弈，因此一方的成功必定对应另一方的失败，两方无人机的成功和失败奖励值相加为零。在满足有效攻击范围 $r_d \leqslant 300\mathrm{m}$ 的条件下，如果蓝方无人机到达敌人的攻击区域，将在当前步骤获得一个正奖励值，反之如果对手跟踪到达它的攻击区域，将获得一个负奖励值，其他情况下的奖励值为 0。奖励函数设计如下式(4.22)：

$$R_b = -R_r = \begin{cases} 1.0, & |\mathrm{ATA}_b| < 30^\circ \wedge |\mathrm{AA}_b| < 60^\circ \\ -1.0, & |\mathrm{ATA}_r| < 30^\circ \wedge |\mathrm{AA}_r| < 60^\circ \\ 0, & otherwise \end{cases} \tag{4.22}$$

当蓝方无人机满足有效攻击范围 $r_d \leqslant 300\text{m}$ 和有效攻击区域 $|\text{ATA}_b| < 30^\circ \wedge |\text{AA}_b| < 60^\circ$ 时，可以认为其能达到发射近距空空导弹的条件。设置一场作战的最大对抗步数 $oil = 50$，如果蓝方无人机连续满足攻击条件 5 步，则判定其赢得作战胜利。当一场作战达到最大对抗步数时仍然没有胜负之分，则认为该场作战为平局。

至此完成无人机一对一仿真环境的相关设计。

4.3　基于改进 DQN 算法的博弈决策生成

前文已经明确将强化学习方法应用于无人机机动决策进而生成最优机动策略是一种可行手段，但其面临的挑战之一是如何存储动作价值函数，由于空战具有高维状态空间，导致传统强化学习算法面临着维数爆炸，深度强化学习算法的出现则可以很好地解决这一问题，它结合了深度神经网络和强化学习，利用深度神经网络的非线性拟合能力，突破了有限维状态输入的局限性，使作战无人机有能力处理更复杂的问题。

在无人机一对一场景中，敌我双方都根据战场态势及时更新和调整战略，这是一个动态的博弈过程，具有强对抗性，涉及复杂的利益冲突。而现有的深度强化学习算法是一类单方优化的方法，仅考虑自身机动策略最优，并没有实际考虑到对手策略对战场局势造成的影响，而在无人机对抗过程中，参战双方具有不同的作战目标，涉及到双方或多方的策略交互，从而存在利益冲突，只考虑单方的控制而不考虑对方策略的动态变化显然不太合理，不符合实际要求。

因此本节针对传统深度强化学习算法存在的问题，融入了博弈论的思想，提出了改进的 DQN 算法来生成有效针对对手的机动策略。

4.3.1　纳什或极大极小均衡

在马尔可夫博弈中，由于智能体的奖励回报受到环境中其他参与者动作的影响，因此没有哪个智能体的策略总是最优的。最佳对策[80] 和纳什均衡[81] 是用来评估一个智能体对于其他参与者行为表现的两个著名概念，对于有两个利益冲突智能体的二人零和马尔可夫博弈（TZMG），其定义总结如下：

（1）TZMG 最佳对策：给定对手策略 μ，如果没有比我方策略 π^b 更好的策略，则称其为最佳对策：

$$J(s_0;\pi^b,\mu) \geqslant J(s_0;\pi,\mu) \quad \forall\pi \tag{4.23}$$

反之，给定我方策略 π，对手的最佳对策有 μ^b：

$$J(s_0;\pi,\mu^b) \leqslant J(s_0;\pi,\mu) \quad \forall\mu \tag{4.24}$$

在两人零和马尔可夫博弈中，双方的奖励值完全相反，一方目标是最大化未来回报，而另一方目标是极小化，因此两者符号相反。

（2）TZMG 纳什均衡：纳什均衡策略 (π^*, μ^*) 即为针对双方彼此的最佳对策：

$$J(s_0;\pi,\mu^*) \leqslant J(s_0;\pi^*,\mu^*) \leqslant J(s_0;\pi^*,\mu) \quad \forall\pi,\mu. \tag{4.25}$$

二人零和马尔可夫博弈总是存在纳什均衡，其等价于极大极小解[82]：

$$J(s_0;\pi^*,\mu^*) = \max_{\pi}\min_{\mu} J(s_0;\pi,\mu) = \min_{\mu}\max_{\pi} J(s_0;\pi,\mu) \tag{4.26}$$

纳什均衡规定了智能体面对最糟糕的对手所能获得的最大奖励回报。在不了解对手，或者对手是一个根据我方策略更新其行为的智能体时，纳什均衡策略显得尤为重要。

4.3.2　改进 DQN 算法设计

传统的强化学习方法存在维数爆炸的问题，并且倾向于单方面最大化自身利益。在此基础上，针对空战高维连续的状态空间，引入深度神经网络来近似 Q 函数，并且应用极大极小均衡的概念来解决具有强对抗和复杂冲突的无人机机动决策问题。

由于融入了博弈论的元素，Q-Learning 的值迭代过程由一个智能体转变为两个智能体的二人零和博弈[83]。

已知在马尔可夫决策过程中，状态值函数用来评估当前状态的好坏，其不仅取决于当前状态还受到未来状态的影响，对状态的累积奖励回报求期望即可得到当前状态的状态值函数，从状态 s 开始一直采取最优策略可获得最优累积期望 $V(s)$。状态动作值 $Q(s,a)$ 表示在当前状态 s 下遵循最优策略采取动作 a 得到的累积折扣奖励。$V(s)$ 和 $Q(s,a)$ 存在以下递归关系：

$$V(s) = \max_{a'\in A} Q(s,a') \tag{4.27}$$

$$Q(s,a) = R(s,a) + \gamma \sum_{s'\in S} T(s,a,s') \cdot V(s') \tag{4.28}$$

其中，$\gamma \in (0,1)$ 代表折扣因子，当 γ 接近 0 时，只考虑当前奖励带来的影响；γ 接近 1 时，着重考虑远距离回报的价值。$T(s,a,s')$ 表示在当前状态 s 下采取动作 a 到达下一状态 s' 的转移概率。$Q(s,a)$ 由即时奖励和转移概率加权后的贴现值共同组成，可以通过选择 Q 值最高的动作来学习到最优策略.

当环境中只有一个智能体时，采用贪心策略（$\epsilon-greedy$）学习最优策略，存在多个智能体时，则需考虑其他智能体联合动作产生的影响。因此考虑马尔可夫博弈模型，$V(s)$ 表示在状态 s 下采取最优决策得到的未来累积期望回报，$Q(s,a,o)$

表示在当前状态 s 下我方采取动作 a，对手采取动作 o 得到的未来累积折扣奖励。那么马尔可夫博弈下的最优值函数为：

$$V(s) = \max_{\pi \in PD(A)} \min_{o \in O} \sum_{a \in A} Q(s,a,o) \cdot \pi_a \tag{4.29}$$

其中，$PD(A)$ 表示离散的动作分布概率，在当前状态 s 下我方采取动作 a，对手采取动作 o 的 $Q(s,a,o)$ 可表示为：

$$Q(s,a,o) = R(s,a,o) + \gamma \sum_{s'} T(s,a,o,s') \cdot V(s') \tag{4.30}$$

上式 (4.30) 中状态转移函数 $T(s,a,o,s')$ 为基于模型的方法，在无人机对抗问题中一般不可得，考虑使用替代转移函数方法[84]。

Q-Learning 中，执行一次更新过程为智能体在当前状态 s 下选择动作 a 转移到下一状态 s'，更新内容为：

$$Q(s,a) = r + \gamma V(s') \tag{4.31}$$

其中，r 表示在状态 s 下执行动作 a 转移到状态 s' 所获得的即时奖励。因为完成更新过程的转移概率正好为 $T(s,a,s')$，所以式 (4.28) 可由式 (4.31) 来代替，并把此方法迁移到马尔可夫博弈模型中，式 (4.30) 可转化为：

$$Q_t(s,a,o) = (1-\alpha) \cdot Q_{t-1}(s,a,o) + \alpha \cdot (r + \gamma V(s')) \tag{4.32}$$

其中，α 表示学习率。

改进后算法的最优值函数即为上式 (4.29)，与原有算法相比，其融合了博弈论的思想，使用 mininax 值代替了 max 值，体现了极大极小均衡的核心思想，智能体在面对最糟糕的对手时可以取得最大的奖励回报。在无人机对抗博弈中其意义在于，当我方不了解敌方，对敌方的作战策略未知时，假设其拥有高水平的决策能力，执行使我方陷入空中战场最劣势地位的动作时，在此基础上我方采取当前状态下可获得最大奖励回报的最优决策。式 (4.29) 可采用线性规划模型求解，得到 $Q(s,a,o)$ 后，在已知双方动作空间的情况下，通过线性规划约束解算当前状态 s 下的最优值函数 $V(s)$ 和最优策略 π。

改进后算法的状态动作值函数更新方式为上式 (4.32)，作为区别于原有算法的关键之处，其更新公式对比如下：

$$\begin{aligned} Q_t(s,a,o) &= (1-\alpha) \cdot Q_{t-1}(s,a,o) + \alpha \cdot (r + \gamma V(s')) \\ Q_t(s,a) &= (1-\alpha) \cdot Q_{t-1}(s,a) + \alpha \cdot (r + \gamma \cdot max_{a'} Q(s',a')) \end{aligned} \tag{4.33}$$

可以看出原有算法执行贪心策略来完成对 Q 表的更新，但已知在无人机进行对抗过程时，对手的机动动作必定会对我方的机动决策产生极大程度上的影响，因

此原算法在不考虑对手策略的情况下使用 $\epsilon-greedy$ 不符合对抗博弈机动决策的针对性思维。改进后的算法利用极大极小均衡思想，求解对抗过程中每个阶段状态的近似纳什均衡策略，进而学习到可以有效针对具有高决策水平对手的最优策略。

同时考虑到空战博弈具有高维连续的状态空间，使用 Q 表存储每个状态动作值时会导致维数爆炸的问题。在此基础上，引入深度神经网络[85] 来近似 Q 函数，利用深度神经网络非线性逼近行为值函数的形式代替 Q-Learning 中的线性函数逼近，由此处理无人机对抗态势的高维非线性输入。

神经网络输入状态值，输出智能体与其对手所选动作对应的 $Q(s,a,o)$。用于非线性拟合的样本数据来源于经验池，其存储着智能体在探索环境一段时间内得到的历史经验，由当前状态 s，我方执行动作 a，对手执行动作 o，到达的下一状态 s' 以及当前步所获得的即时奖励作为五元组 (s,a,o,r,s') 的形式组成，经验池的最大存储容量为 M_{max}。由于马尔可夫博弈过程是一个时序决策过程，经验池中的经验数据包含有决策前后的关联性，其数据不满足独立同分布的假设，为了满足神经网络非线性拟合对样本数据的要求，需要通过采样过程降低数据样本非静态的影响，从经验池中随机抽取一组大小为 $batchsize$ 的历史经验数据当做训练样本以此作为神经网络参数优化的依据，这样就打破了数据之间的相关性使得算法更容易走向收敛。

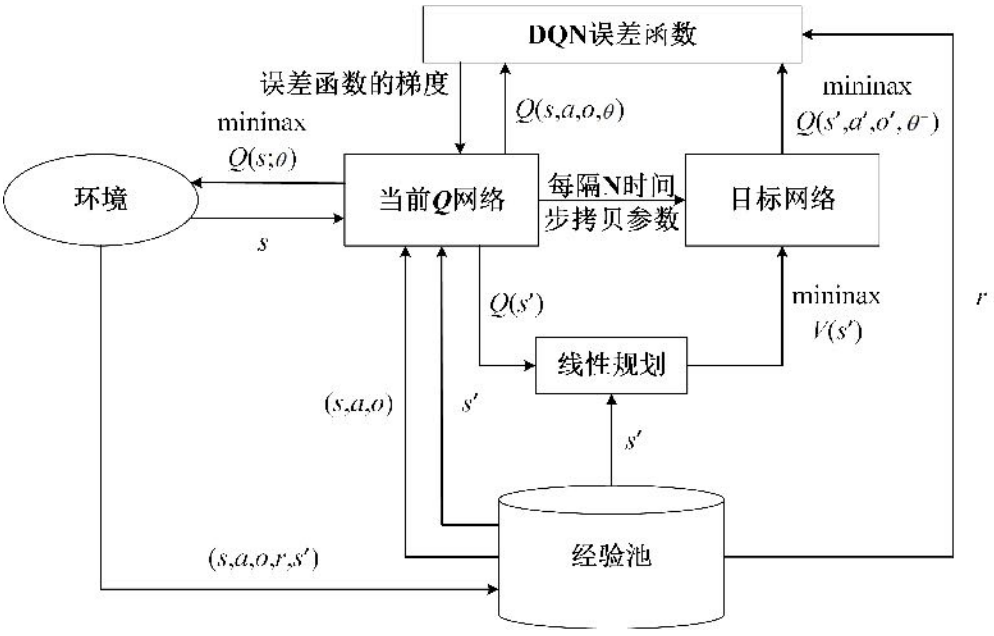


图 4.4　改进 DQN 算法训练过程

训练神经网络本质是最优化问题，损失函数[86]（loss function）是神经网络在拟合价值函数过程中优化网络参数的依据，优化网络参数的目标是使损失函数最小化，定义损失函数为目标 Q 值和当前真实 Q 值的差平方：$loss=(target\_q-q)^2$。

为了减少目标 Q 值和当前 Q 值的相关性，使用两个结构相同但参数不同的神经网络：目标网络和 Q 网络。目标网络计算目标 Q 值，其参数为 θ^-，每过一段时间从 Q 网络中复制权重，以此来延迟更新；Q 网络计算当前 Q 值，其参数为 θ，使用反向传播算法实时更新参数，这样就提高了算法的稳定性。

综上所述，改进后算法的具体训练过程如上图4.4所示。

总结上文内容，给出改进后的算法步骤如下所示：

算法 4.1　改进 DQN 算法

初始化容量大小为 $M_{\max}$ 的经验池 D

初始化 Q 网络，随机生成权重 θ

初始化目标网络，权重 $\theta^- = \theta$

　　重复以下步骤：

　　初始化状态 $s = s_0$

　　　　重复以下步骤：

　　　　　　我方无人机在当前状态 s 下根据探索利用选取动作 a

　　　　　　观测敌方无人机在状态 s 下选取的动作 o

　　　　　　我方无人机执行动作 a，敌方无人机执行动作 o，获得即时奖励 r，下一状态 s'

　　　　　　将五元组 (s, a, o, r, s') 存入经验池 D

　　　　　　从经验池 D 中随机采样 $batchsize$ 个样本，把样本的 s' 值输入神经网络

　　　　　　神经网络输出状态 s' 下的 $Q(s')$

　　　　　　由式 (4.29) 通过线性规划求解 minimax 状态值 $V(s')$

　　　　　　由式 (4.32) 计算当前样本的目标 Q 值 $target\_q$

　　　　　　损失函数 $loss = (target\_q - Q(s, a, o, \theta))^2$ 进行梯度反向传播以更新 Q 网络参数 θ

　　　　　　每 C 步更新目标网络参数 $\theta^- = \theta$

　　　　直至终止状态或重复 N 次

　　直至重复 M 次

　　由式 (4.32)，利用训练好的神经网络输出的 Q 值通过线性规划求解最优策略 π

根据算法 4.1 得到改进 DQN 算法流程图，如下图4.5所示。

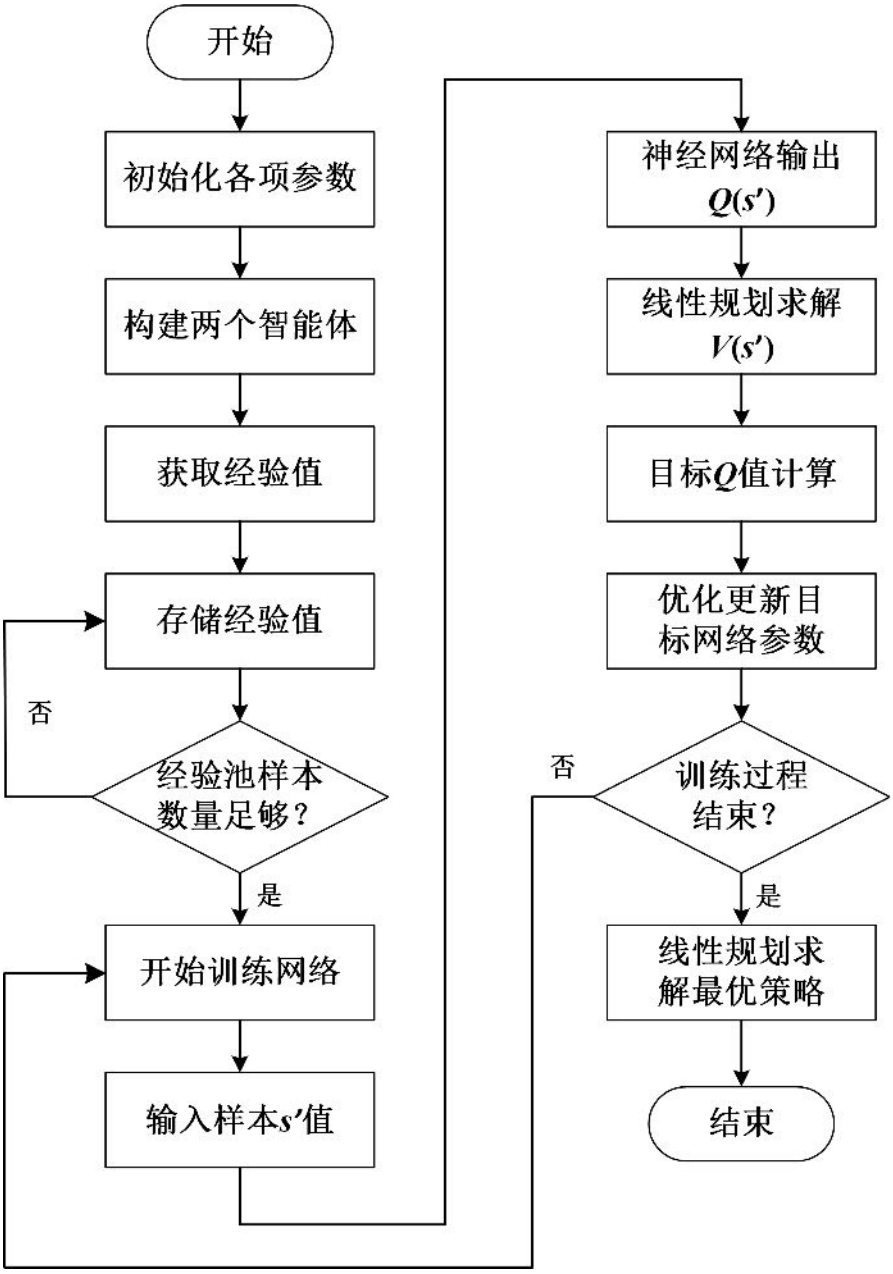


图 4.5　改进 DQN 算法流程图

至此完成了本节的改进 DQN 算法设计。

4.4　一对一机动决策仿真实验及分析

本节主要基于改进 DQN 算法设计的对抗机动决策方案训练了神经网络模型。首先介绍了仿真实验环境以及训练过程中的基本参数配置，然后将三组采取不同机动策略的无人机进行一对一博弈，并对实验结果进行了对比分析。

4.4.1　实验设计及参数配置

本实验采用 Python3 语言来编写无人机运动学模型和空战仿真环境程序，其可以实时记录红蓝两架无人机的状态数据，并且使用到了 gym、numpy、scipy、tensorboard 等第三方库，可视化界面如下图4.6所示。

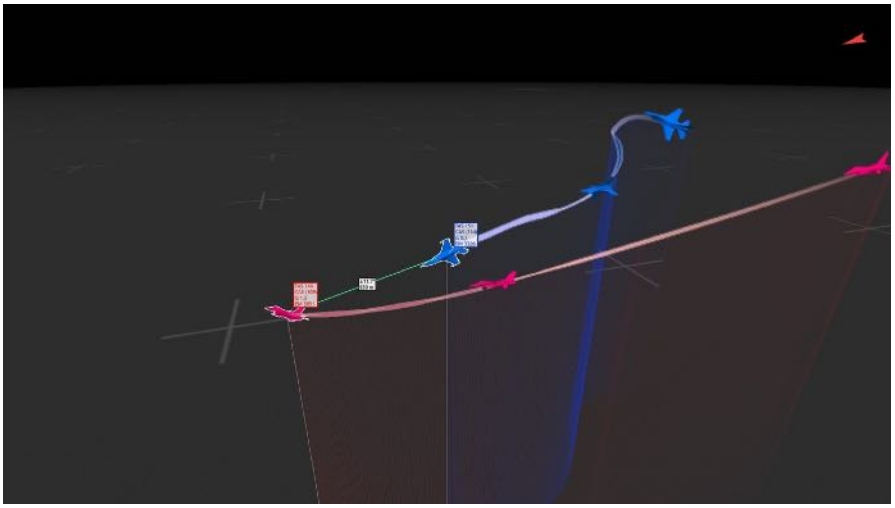


图 4.6　对抗可视化界面

表 4.1　仿真实验参数

实验参数	
横坐标范围	$x \in [-2000\text{m}, 2000\text{m}]$
纵坐标范围	$y \in [-2000\text{m}, 2000\text{m}]$
时间变化量	$dt = 0.5$
目标方位角最大值	$\text{ATA}_{\max} = 30^{\circ}$
进入角最大值	$\text{AA}_{\max} = 60^{\circ}$
允许最小速度	$v_{\min} = 10\text{m/s}$
允许最大速度	$v_{\max} = 50\text{m/s}$
最大滚转速率	$\varphi_{\max} = 30^{\circ}/\text{s}$
最大对抗步长	$oil = 50$
神经网络学习率	$lr = 10^{-4}$
折扣因子	$\gamma = 0.99$
经验池最大容量	$M_{\max} = 10^{4}$
批处理样本数	$batchsize = 64$

仿真实验设定的场景为红蓝双方无人机在空中对抗时的高度变化不大，无人机在发生滚转运动时协调转弯控制过载定高飞行，仅关注红蓝无人机在对抗过程中的航迹而不考虑其姿态控制，因此可看作在水平面内飞行，其误差也在可

接受范围之内。

根据状态空间和动作空间的定义可知，神经网络的输入节点数为 10，输出节点数为 2。基本参数配置如上表4.1所示。

表 4.2　初始状态参数

参数	蓝方无人机	红方无人机
x	[-2000m,2000m] 随机生成	[-2000m,2000m] 随机生成
y	[-2000m,2000m] 随机生成	[-2000m,2000m] 随机生成
v	20m/s	20m/s
ϕ	0	0
ψ	[-180°,180°] 随机生成	0

红蓝两架无人机初始态势的状态参数设置如上表4.2所示。

4.4.2　DON 对抗随机策略

根据以上参数，蓝方无人机采取 DQN 算法生成的策略选择动作，而红方无人机采取随机策略选择动作。

下图4.7展示了在训练 DQN 算法的过程中随着训练步数的增加平均奖励值的变化，从图中可以看出，平均奖励值在训练到约 60 万步以后基本保持稳定，证明 DQN 算法开始走向收敛。

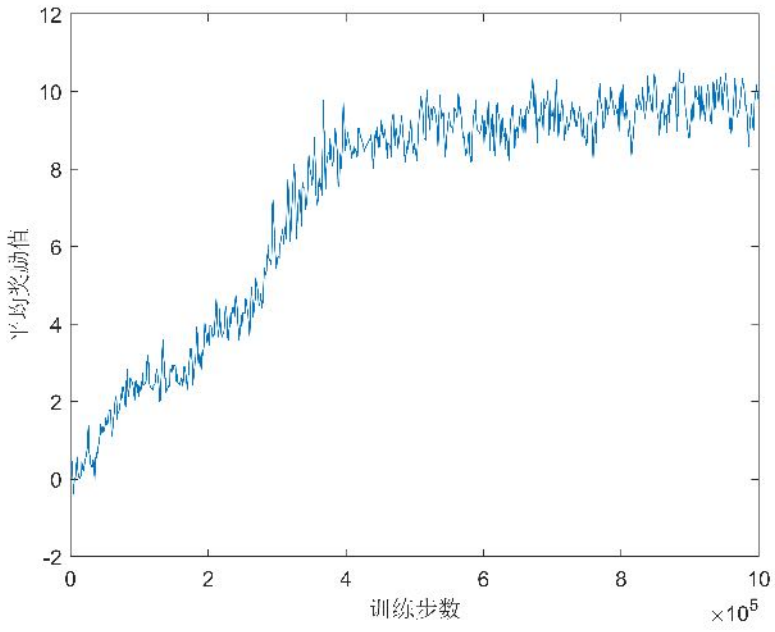


图 4.7　DQN 平均奖励值收敛过程

训练完成后，蓝方无人机利用神经网络输出的策略与红方无人机进行 1000

回合的空中对抗，记录博弈最后一回合红蓝无人机的机动轨迹，如下图4.8所示。由图可见，红方无人机随机生成的初始坐标位于蓝方无人机的西北方向，红方以缓慢速度维持一定航迹偏角飞行，蓝方观测到红方位置后快速做出反应调转航向以较大速度乘胜追击，红方为躲避蓝方首先拉开距离，蓝方在快接近红方时为避免速度过快飞越敌机随后调转方向盘旋，并保持在红方尾部占据可攻击位置发射导弹。通过红蓝双方机动过程可以看出，采用 DQN 算法生成策略的蓝方无人机机动动作灵活，在空域战场中不断调整己方作战态势，利用已有的角度优势和位置优势迅速达到可攻击位置发射导弹以完成摧毁敌方的作战目标。

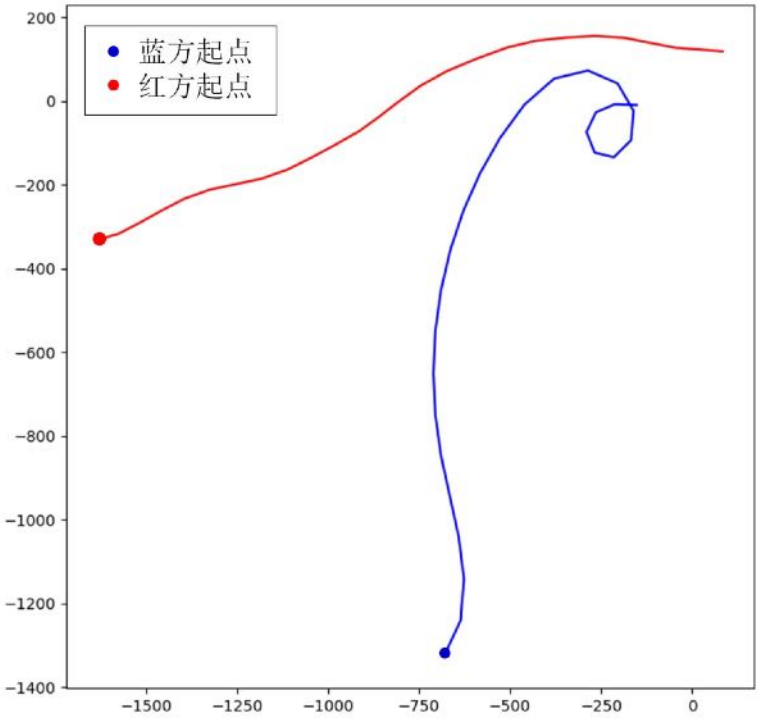


图 4.8　DQN vs 随机策略的对抗机动轨迹

将最终的红蓝无人机对抗博弈结果记录在下表4.3中。

表 4.3　DQN vs 随机策略的结果

	胜	败	平局	胜率
DQN 算法	672	294	34	67.2%
随机策略	294	672	34	29.4%

由表4.3可以看出，蓝色无人机通过 DQN 算法学习到的机动策略较容易打败红色无人机随机生成的机动决策序列，随着训练次数的不断增加，DQN 算法逐

渐学习到较好的对抗决策，并将最终胜率保持在 67.2%。

4.4.3　改进 DON 对抗随机策略

在相同的仿真环境和参数设置下，进行另一组对比实验，此时蓝方无人机采取改进 DQN 算法生成的策略选择动作，而红方无人机仍然采取随机策略。

图4.9表示了训练改进 DQN 算法过程中平均奖励值随训练步数的变化，可以看出，训练步数到达约 75 万步以后，平均奖励值才保持在一个相对稳定的状态，改进 DQN 算法趋于收敛。

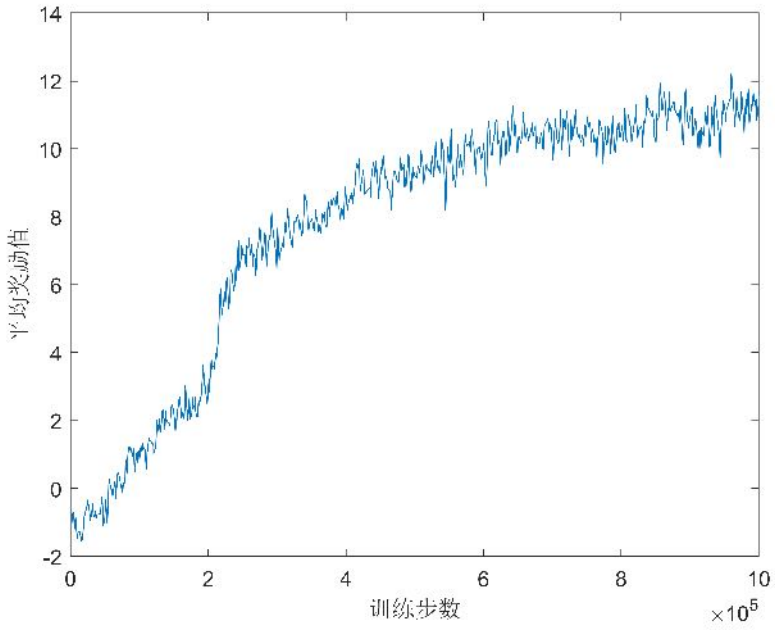


图 4.9　改进 DQN 平均奖励值收敛过程

类似地，将训练成功后的蓝方无人机与红方无人机进行 1000 回合的博弈对抗，记录作战最后一回合红蓝无人机的飞行轨迹，如下图4.10所示。从图中可以看出，红方无人机随机生成的初始坐标在蓝方无人机的西南方向，双方距离较远。为进行近距离交战，红方无人机保持一定航向角以较大速度靠近蓝方无人机，蓝方同样向西南方向快速前进，由于红方机体朝向东北方向，蓝方位于其攻击后方区域，红方大幅度调转机头扭转不利态势，俯冲拉进与蓝方之间的距离，蓝方立即向西飞跃红方避免正面迎敌并重新拉开距离，接着转变航向绕后红方，重新占据角度优势位置，稳定锁住敌方到达可攻击位置。通过红蓝双方机动动作可以看出，采用改进 DQN 算法学习机动决策的蓝方无人机先发制人从红方无人机绕后进行攻击目标，针对红方机动动作采取快速有效的机动决策，同时避免了正面交锋可能导致的己方获胜概率降低，与红方拉锯后重新占据有利地位。

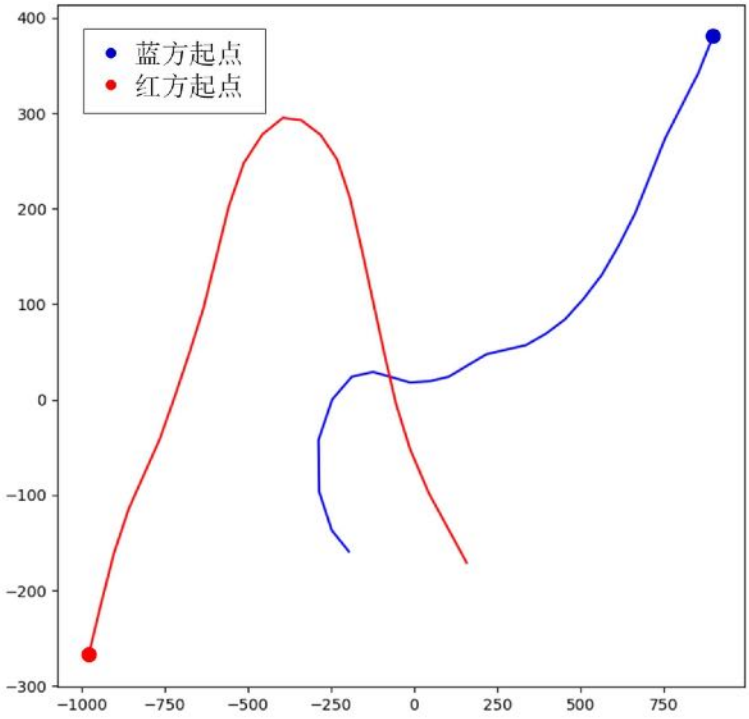


图 4.10　改进 DQN vs 随机策略的对抗机动轨迹

同时将 1000 回合的对抗结果记录在下表4.4中。

表 4.4　改进 DQN vs 随机策略的结果

	胜	败	平局	胜率
改进 DQN 算法	724	258	18	72.4%
随机策略	258	724	18	25.8%

由表4.4可以看出，改进 DQN 算法生成的机动决策以明显的优势打败了随机策略，其胜率高达 72.4%。相比于 DQN 算法对抗随机策略的胜率有了显著的提高，从侧面表明了改进 DQN 算法学习到的机动策略相较于 DQN 算法具有更高的决策水平。

4.4.4　改进 DON 对抗 DQN

由于随机策略的决策性较差，同时为了更加直观地对比两种算法，进行第三组对比实验。分别采用改进 DQN 算法训练红方无人机和 DQN 算法训练蓝方无人机，通过 tensorboard 实时跟踪整个训练过程，使用平均奖励值作为量化指标来衡量双方对抗态势优势。红蓝无人机的平均奖励值收敛过程如下图4.11所示。

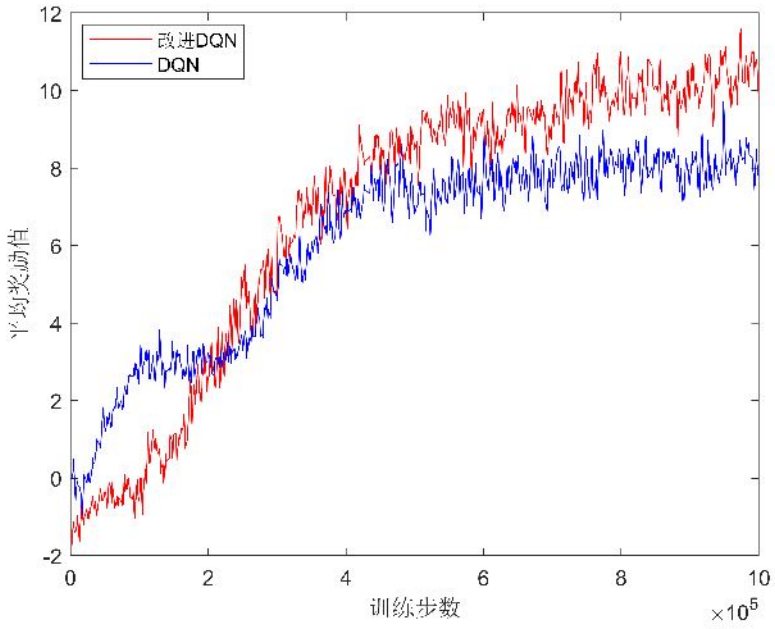


图 4.11　改进 DQN 与 DQN 收敛过程

由图4.11可以看出，在整个训练过程中，改进 DQN 和 DQN 的平均奖励值都趋于稳定且收敛到了较高的结果。然而，改进 DQN 算法在整个训练过程中的收敛曲线优于 DQN，其平均奖励值也高于 DQN。

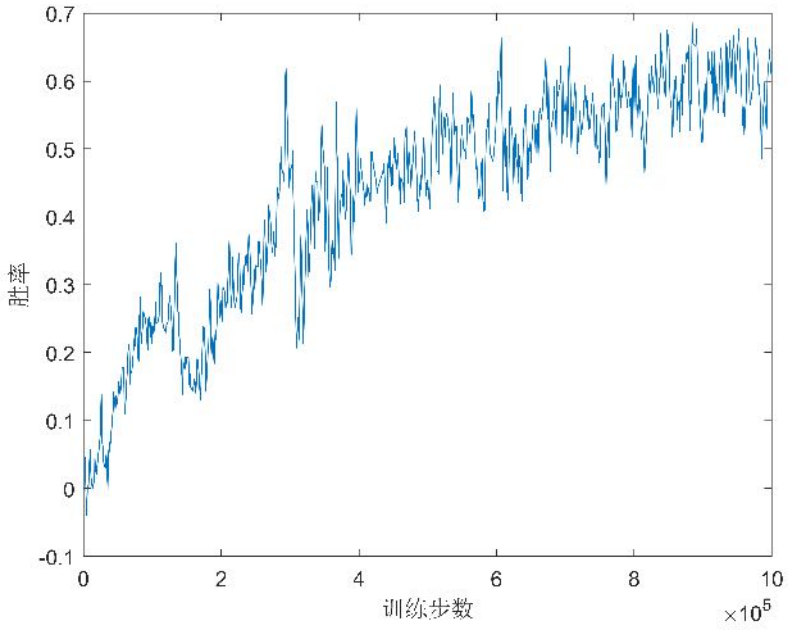


图 4.12　胜率收敛过程

上图4.12展示了在训练过程中胜率随着训练步数增加发生的变化，可以看出，训练步数到达约 80 万步以后，胜率才逐渐达到收敛状态。由于刚开始阶段训练次数太少，改进 DQN 算法没有探索到更好更合适的机动策略，导致其对抗

DQN 算法时胜率很低。但随着训练步数的递增，胜率逐渐提高最后稳定在 0.6 附近。

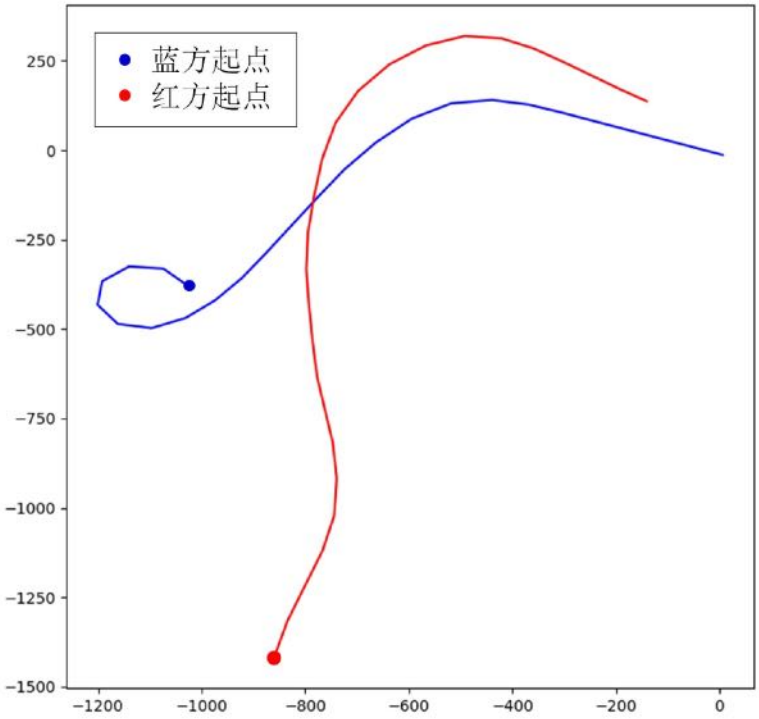


图 4.13　改进 DQN vs DQN 的对抗机动轨迹

为了更好的评估由两种算法生成的对抗机动策略，红蓝两架无人机进行 10000 回合博弈，记录双方无人机在博弈过程中某一回合的机动轨迹，如上图4.13所示。从图中可以看出，作战开始时，红方无人机随机生成的初始坐标位于蓝方无人机的东南方向，两机相距约 1000m。蓝方无人机处于不利态势，其率先大幅度扭转机头避免陷入角度劣势位置，红方无人机观测到蓝方位置后快速前进利用角度企图拉进两机距离，两机同时转向第一次交锋无果后，红方加速前进利用角度差飞越蓝方无人机，与蓝方重新拉开距离，借机到达其身后，并且稳定地锁住蓝方无人机，占据作战位置优势和角度优势，达到攻击条件后发射导弹击毁蓝方，一举取得胜利。可以看出红方无人机在随机生成的初始态势下，能够根据改进后算法学习到的策略进行快速机动动作，规避威胁的同时借机绕后占位反击，迅速达到攻击条件结束战场交锋。

将 10000 回合红蓝无人机博弈结果记录在下表4.5中，并将上述三次实验结果的胜率绘制在条形图4.14中。

表 4.5　改进 DQN vs DQN 的结果

	胜	败	平局	胜率
改进 DQN 算法	5973	3790	237	59.7%
DQN 算法	3790	5973	237	37.9%

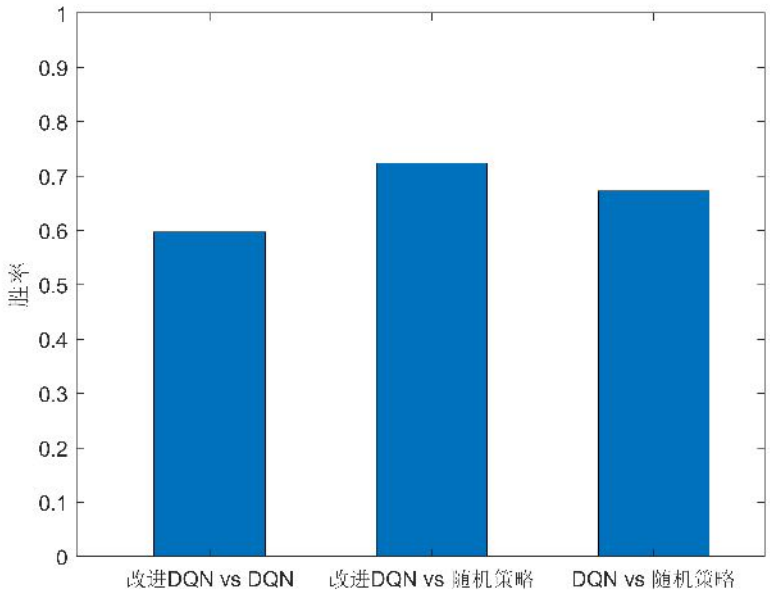


图 4.14　三种策略胜率对比

由表 4.6 可见，红方无人机获胜 5973 次，蓝方无人机获胜 3790 次，经计算红方无人机的胜率约为 59.7%。其表明了，在相同的对抗仿真环境下，使用改进 DQN 算法学习到的红方机动决策比使用 DQN 算法学习到的蓝方机动决策略有优势。另外，记录了在评估过程中改进 DQN 算法生成一个机动决策的时间约为 0.0036 秒，证明了其在对抗过程中可以快速高效的生成机动策略，达到实时性要求。

上图4.14直观地展现了，无论是改进 DQN 与 DQN 分别与其他策略对抗，还是两种算法直接进行对抗，改进后的算法都比原算法效果更好。

综合以上仿真实验结果和算法对比分析，改进后的算法很好的结合了博弈论的均衡策略思想以及深度强化学习的自学习能力。改进 DQN 算法不仅满足无人机对抗实时性的要求，能在对抗过程中快速反应执行机动动作，还可以在强竞争环境下生成更准确、更有效针对对手的机动策略，与原算法相比，其更适用于无人机对抗博弈场景。

4.5　本章小结

本章基于博弈论和深度强化学习算法对无人机机动决策方法进行研究。在马尔可夫决策过程的理论基础之上，将其扩展为用于无人机一对一场景下的二人零和马尔可夫博弈模型，并根据二人零和博弈环境设计了状态空间、动作空间和奖励函数。然后结合博弈论的思想对深度强化学习算法进行改进，设计了应用于无人机对抗博弈环境的改进 DQN 算法来生成最优机动策略，通过仿真验证了该算法可以使无人机的策略网络收敛，并通过 3 组对比实验证明了该算法相较于传统算法可以通过自学习生成有效针对对手的机动策略，具有更高的决策水平，可以在无人机对抗机动决策中的得到很好的应用。

第 5 章　总结与展望

5.1　论文工作总结

无人机作为未来争夺制空权不可忽略的角色，其自主机动决策能力是发挥作战效能的关键所在。如何在高动态、强竞争性的无人机对抗环境下进行快速有效的机动决策是本论文主要研究的问题。本文以近距对抗为背景，利用博弈论的思想并结合智能求解算法，分别对基于优化理论和基于人工智能的无人机机动决策方法展开研究。论文主要研究工作和创新点总结如下：

（1）完成了 F-16 机型无人机的机动空间设计。作为研究机动决策的基础，首先基于 F-16 机型无人机的六自由度模型设计其控制参数，完成配平工作，在此基础上对基本操纵动作库进行丰富和改进，设计了无人机的机动空间，构建起无人机的机动策略集，仿真测试结果表明所设计的控制参数和机动动作都能满足设计需求。

（2）针对基本群智算法搜索决策结果计算效率低且容易陷入局部最优值的问题，提出了一种改进粒子群算法求解最优机动策略。首先利用博弈理论建立起无人机一对一场景下的数学模型，对于无人机对抗博弈这类搜索空间维度高、优化目标复杂导致最优混合策略难以求解的问题，利用群智算法处理复杂高维动态优化问题的能力进行寻优。对于基本粒子群算法在求解博弈纳什均衡策略时容易陷入局部最优值的问题，通过粒子浓度的概率选择对基本群智算法进行改进并应用于无人机对抗机动决策中。最后单机对抗实验结果表明，改进粒子群算法提升了全局搜索寻优效率和寻优精度，提高了无人机对抗机动决策中求解最优机动策略的计算效率和准确度。

（3）针对传统强化学习算法在处理高维状态输入时存在的维数爆炸问题，以及倾向于单方面最优化自身策略而不考虑对手策略影响的问题，提出了一种改进 DQN 算法生成有效对抗决策。首先建立了无人机一对一场景的二人零和马尔可夫博弈模型，并据此设计了对抗环境的基本状态空间、动作空间和奖励函数。针对无人机博弈环境的复杂冲突对抗性以及高维状态特征，结合了博弈论解决利益冲突的能力、深度神经网络处理维度爆炸的能力以及强化学习的自学习能力，设计了一种改进 DQN 算法来求解最优机动策略，最后单机对抗实验结果表明，改进 DQN 算法能够通过自学习的方式探索得到有效针对对手的机动策略，满足对抗实时性，具有更高的决策水平。

5.2　后续工作展望

本文重点研究了两种智能机动决策方法，并针对其应用于无人机对抗环境求解最优机动策略时存在的问题做出了改进，但后续还存在一些可继续深入研究、改进的方向：

（1）本文的对抗仿真环境相对简单，与现实场景的空战有着较大差距，真实场景还会存在武器、电子干扰等各方面的影响，后续可进一步设定复杂的战场条件对机动决策方法进行更深入的研究。

（2）本文目前主要研究了一对一场景中的机动决策，而现实场景是多对多无人机空战对抗，其中涉及更复杂的合作和对抗冲突，因此需要拓展到多机对抗研究，建立更复杂的模型并设计相应的求解算法。

（3）本文在对机动决策问题建模时基于敌对双方可以获取全部的战场环境信息，包括对方的动作，实际战场只能通过对方的状态变化来推测其动作，因此就需要在目标意图推测、战场信息补全等方面进一步研究。

由于本人对近距对抗领域的认知有限，在研究方法和仿真中难免存在不足和疏漏之处，希望各位能够指正。

参 考 文 献

[1] 杜燕波. 智能化战争形态下的作战体系复杂性[J]. 军事文摘, 2021(011): 000.

[2] 王中华. 高技术与现代空战[J]. 现代军事, 1992, 16(6): 4.

[3] 李静, 王国恩, 张玉册. 无人机作战运用及发展趋势研究[J]. 舰船电子对抗, 2005, 28(4): 4.

[4] 魏瑞轩, 李学仁. 无人机系统及作战使用[M]. 无人机系统及作战使用, 2009.

[5] 战贵军, 郭震. 无人机发展现状及相关技术[J]. 2020.

[6] CAMBONE S A, KRIEG K, PACE P, et al. Unmanned aircraft systems roadmap 2005-2030 [J]. Office of the Secretary of Defense, 2005, 8: 4-15.

[7] 董彦非. 空中作战智能决策与无人战斗机智能化[J]. 西安航空学院学报, 2015, 33(5): 4.

[8] 董一群艾剑良. 自主空战技术中的机动决策: 进展与展望[J]. 航空学报, 2020(S02): 9.

[9] NICHOLS S O. 21st century air-to-air short range weapon requirements[R]. AIR COMMAND AND STAFF COLL MAXWELL AFB AL, 1998.

[10] RASMUSEN E. Games and information : an introduction to game theory[M]. Games and information : an introduction to game theory, 2001.

[11] 陶一桃. 对《孙子兵法》的博弈论分析[J]. 中国军事科学, 2004, 17(6): 40-44.

[12] 李帮义. 博弈论及其应用[M]. 博弈论及其应用, 2008.

[13] 王壮. 近距空战飞行器智能机动决策生成研究[D]. 四川大学.

[14] LIEVEN A. The war in afghanistan: its background and future prospects: Analysis[J]. Conflict, Security & Development, 2009, 9(3): 333-359.

[15] 厉博. 国外无人作战飞机发展回顾与趋向分析[J]. 飞航导弹, 2019(10): 6.

[16] 季晓光, 李屹东. 美国高空长航时无人机: RQ-4” 全球鹰”[M]. 美国高空长航时无人机: RQ-4“全球鹰”, 2011.

[17] 杨志钱, 梁光建. 美军无人机侦察技术应用[J]. 数字通信世界, 2020.

[18] 柴水萍. MQ-9” 死神” 无人机[J]. 现代军事, 2008(9): 2.

[19] 张建华, 赵晨皓, 吕诚中. 察打一体无人机发展现状及趋势[J]. 飞航导弹, 2018(2): 7.

[20] 庄林. 俄罗斯军用无人装备的发展与运用[J]. 军事文摘, 2019(009): 000.

[21] 石一文. 在云端——中国” 彩虹” 系列无人机[J]. 兵器知识, 2015(12): 5.

[22] 张月, 刘睿. 军用大型无人机的发展现状及趋势[J]. 电子世界, 2016(11): 2.

[23] 董一群艾剑良. 自主空战技术中的机动决策: 进展与展望[J]. 航空学报, 2020(S02): 9.

[24] 孙永芹, 孙涛, 范洪达, 等. 现代空战机动决策研究[J]. 海军航空工程学院学报, 2009, 24 (5): 5.

[25] 周新民, 吴佳晖, 贾圣德, 等. 无人机空战决策技术研究进展[J]. 国防科技, 2021, 42(3): 9.

[26] 基于双矩阵对策的 UCAV 空战自主机动决策研究[J]. 舰船电子工程, 2017, 37(11): 6.

[27] 罗贺, 马滢滢, 胡笑旋, 等. 无人机编队协同目标分配的矩阵博弈方法及系统[Z]. 2019.

[28] 车竞, 钱炜祺, 和争春. 基于矩阵博弈的两机攻防对抗空战仿真[J]. 飞行力学, 2015(2): 5.

[29] 李伟. 基于微分对策理论的无人战机空战决策方法研究[D]. 沈阳航空航天大学.

[30] 黄家成, 谢奇峰. 基于遗传算法的协同多目标攻击空战决策方法[J]. 火力与指挥控制, 2004, 29(1): 4.

[31] 徐超, 王玉惠, 吴庆宪, 等. 基于模糊遗传算法的先进战机协同攻防决策[J]. 火力与指挥控制, 2020, 45(3): 8.

[32] 谢建峰, 杨啟明, 戴树岭, 等. 基于强化遗传算法的无人机空战机动决策研究[J]. 西北工业大学学报, 2020(006): 038.

[33] VICSEK T. Universal patterns of collective motion from minimal models of flocking[C]// 2008 Second IEEE International Conference on Self-Adaptive and Self-Organizing Systems. IEEE, 2008: 3-11.

[34] 郭辉, 徐浩军, 谷向东, 等. 基于改进粒子群算法的协同多目标攻击空战决策[J]. 火力与指挥控制, 2011, 36(6): 49-51.

[35] HAO Q, WEI Z, XIE L, et al. An autonomous maneuver decision method using receding horizon optimization scheme[C]//2021 40th Chinese Control Conference (CCC).

[36] 傅莉, 谢福怀, 孟光磊, 等. 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, 041(011): 1994-1999.

[37] BURGIN G H, FOGEL L J, PHELPS J P. An adaptive maneuvering logic computer program for the simulation of one-on-one air-to-air combat. volume 1: General description[R]. NASA, 1975.

[38] 魏强, 周德云. 基于专家系统的无人战斗机智能决策系统[J]. 火力与指挥控制, 2007, 32 (2): 4.

[39] 王炫, 王维嘉, 宋科璞, 等. 基于进化式专家系统树的无人机空战决策技术[J]. 兵工自动化, 2019, 38(1): 6.

[40] YANG Q, ZHANG J, SHI G, et al. Maneuver decision of uav in short-range air combat based on deep reinforcement learning[J]. IEEE Access, 2019, 8: 363-378.

[41] HU D, YANG R, ZUO J, et al. Application of deep reinforcement learning in maneuver planning of beyond-visual-range air combat[J]. IEEE Access, 2021, 9: 32282-32297.

[42] LI Y F, SHI J P, JIANG W, et al. Autonomous maneuver decision-making for a ucav in short-range aerial combat based on an ms-ddqn algorithm[J]. Defence Technology, 2022, 18(9): 1697-1714.

[43] HU J, WANG L, HU T, et al. Autonomous maneuver decision making of dual-uav cooperative air combat based on deep reinforcement learning[J]. Electronics, 2022, 11(3): 467.

[44] MA X, XIA L, ZHAO Q. Air-combat strategy using deep q-learning[C]//2018 Chinese Automation Congress (CAC). IEEE, 2018: 3952-3957.

[45] LI S Y, CHEN M, WANG Y H, et al. Air combat decision-making of multiple ucavs based on constraint strategy games[J]. Defence Technology, 2022, 18(3): 368-383.

[46] BURGIN G H, OWENS A. An adaptive maneuvering logic computer program for the simulation of one-to-one air-to-air combat. volume 2: Program description[R]. NASA, 1975.

[47] GOODRICH K, MCMANUS J. Development of a tactical guidance research and evaluation system (tgres)[C]//Flight simulation technologies conference and exhibit. 1989: 3312.

[48] SMITH R E, DIKE B, MEHRA R, et al. Classifier systems in combat: two-sided learning of maneuvers for advanced fighter aircraft[J]. Computer Methods in Applied Mechanics and Engineering, 2000, 186(2-4): 421-437.

[49] ERNEST N, CARROLL D, SCHUMACHER C, et al. Genetic fuzzy based artificial intelligence for unmanned combat aerial vehicle control in simulated air combat missions[J]. Journal of Defense Management, 2016, 6(1): 2167-0374.

[50] THOMPSON C. An exploration of pilot workload and attention in autonomously piloted fighter aircraft[D]. The University of Iowa, 2022.

[51] RIPPLE B. Skyborg program seeks industry input for artificial intelligence initiative[J]. US Air Force, webpage, March, 2019, 27.

[52] FACCHINEI F, KANZOW C. Generalized nash equilibrium problems[J]. 4or, 2007, 5(3): 173-210.

[53] OSBORNE M J, RUBINSTEIN A. A course in game theory[M]. MIT press, 1994.

[54] ISAACS R. Differential games: a mathematical theory with applications to warfare and pursuit, control and optimization[M]. Courier Corporation, 1999.

[55] SOLAN E, VIEILLE N. Stochastic games[J]. Proceedings of the National Academy of Sciences, 2015, 112(45): 13743-13746.

[56] HSU S P, ARAPOSTATHIS A. Competitive markov decision processes with partial observation[C]//2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583): volume 1. IEEE, 2004: 236-241.

[57] WEINTRAUB G Y, BENKARD C L, VAN ROY B. Markov perfect industry dynamics with many firms[J]. Econometrica, 2008, 76(6): 1375-1411.

[58] 陈佩贞. 无人机六自由度简化模型和风扰动处理[J]. 南京航空航天大学学报, 1993(S1): 9.

[59] NGUYEN L T. Simulator study of stall/post-stall characteristics of a fighter airplane with relaxed longitudinal static stability: volume 12854[M]. National Aeronautics and Space Administration, 1979.

[60] STEVENS B L, LEWIS F L, JOHNSON E N. Aircraft control and simulation: dynamics, controls design, and autonomous systems[M]. John Wiley & Sons, 2015.

[61] 吴森堂. 飞行控制系统[M]. 北京航空航天大学出版社, 2013.

[62] RACHMAN E, JAAM J M, HASNAH A M. Non-linear simulation of controller for longitudinal control augmentation system of f-16 using numerical approach[J]. Information Sciences, 2004, 164(1-4): 47-60.

[63] BASPINAR B, KOYUNCU E. Aerial combat simulation environment for one-on-one engagement[C]//2018 AIAA Modeling and simulation technologies conference. 2018: 0432.

[64] AUSTIN F, CARBONE G, FALCO M, et al. Automated maneuvering decisions for air-to-air combat[C]//Guidance, navigation and control conference. 1987: 2393.

[65] 王晓光. 基于微分对策理论的无人飞机空战建模及其仿真[D]. 沈阳航空航天大学, 2012.

[66] NASH JR J F. Equilibrium points in n-person games[J]. Proceedings of the national academy of sciences, 1950, 36(1): 48-49.

[67] MARINI F, WALCZAK B. Particle swarm optimization (pso). a tutorial[J]. Chemometrics and Intelligent Laboratory Systems, 2015, 149: 153-165.

[68] 于敏. 基于 QPSO 算法的最优值求解在 NASH 均衡中的应用[D]. 江南大学.

[69] 余谦, 王先甲. 基于粒子群优化求解纳什均衡的演化算法[J]. 武汉大学学报：理学版, 2006, 52(1): 5.

[70] 邱中华, 高洁, 朱跃星. 应用免疫算法求解博弈问题[J]. 系统工程学报, 2006(398-404).

[71] BAŞAR T, OLSDER G J. Dynamic noncooperative game theory[M]. SIAM, 1998.

[72] WASHBURN A R, et al. Two-person zero-sum games[J]. 2014.

[73] WEI Z, XU J, LAN Y, et al. Reinforcement learning to rank with markov decision process[C]// Proceedings of the 40th international ACM SIGIR conference on research and development in information retrieval. 2017: 945-948.

[74] SHANI G, PINEAU J, KAPLOW R. A survey of point-based pomdp solvers[J]. Autonomous Agents and Multi-Agent Systems, 2013, 27: 1-51.

[75] ALTMAN E. Flow control using the theory of zero sum markov games[J]. IEEE transactions on automatic control, 1994, 39(4): 814-818.

[76] JOHNSON M, BHASIN S, DIXON W E. Nonlinear two-player zero-sum game approximate solution using a policy iteration algorithm[C]//2011 50th IEEE Conference on Decision and Control and European Control Conference. IEEE, 2011: 142-147.

[77] CARRETERO J G H, NIETO F J S, CORDÓN R R. Aircraft trajectory simulator using a three degrees of freedom aircraft point mass model[C]//Proceedings of the 3rd International Conference on Application and Theory of Automation in Command and Control Systems. 2013: 114-117.

[78] WEIREN K, DEYUN Z, ZHEN Y. Air combat strategies generation of cgf based on maddpg and reward shaping[C]//2020 International Conference on Computer Vision, Image and Deep Learning (CVIDL). IEEE, 2020: 651-655.

[79] WEIREN K, DEYUN Z, ZHANG K, et al. Air combat autonomous maneuver decision for one-on-one within visual range engagement base on robust multi-agent reinforcement learning [C]//2020 IEEE 16th International Conference on Control & Automation (ICCA). IEEE, 2020: 506-512.

[80] HU J, WELLMAN M P. Nash q-learning for general-sum stochastic games[J]. Journal of Machine Learning Research, 2003, 4(4): 1039-1069.

[81] MASKIN E. The theory of implementation in nash equilibrium: A survey[J]. 1983.

[82] ZHU Y, ZHAO D. Online minimax q network learning for two-player zero-sum markov games [J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 33(3): 1228-1241.

[83] CASGRAIN P, NING B, JAIMUNGAL S. Deep q-learning for nash equilibria: Nash-dqn[J]. Applied Mathematical Finance, 2022, 29(1): 62-78.

[84] RUSLAN S, GEOFFREY H. An efficient learning procedure for deep boltzmann machines [J]. Neural Computation, 2012, 24(8): 1967-2006.

[85] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Playing atari with deep reinforcement learning[J]. arXiv preprint arXiv:1312.5602, 2013.

[86] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning[J]. nature, 2015, 518(7540): 529-533.

致　谢

行文至此，落笔为终。三年的硕士生涯，始于疫情，又在疫情结束时悄然离去，学生时代的美好也终将告一段落。人生没有白走的路，每一步都算数，从本科到研究生，我自北方行至烟雨江南，将青春最难忘的七年留在了这里。学习之路不易，成长之路亦风雨兼程，每每向未来张望时，深觉孤独又漫长，在那些波澜不惊的日子里，时间不语却回答了我所有问题。

有师如斯，庆幸之至。感谢我的导师康宇老师和赵云波老师。感谢康老师给了我一个在中科大继续求学的机会，康老师对我产生的影响不仅仅是在学术上要实事求是，同时在为人处事方面也要谦虚谨慎。感谢赵老师一直以来对我科研方面的悉心教导，学生不才，愚钝有时，走过来才深深体会到赵老师所引导的如何定义问题以及解决问题对于科研工作是多么重要，将是我受益一生的宝贵财富。愿两位恩师桃李芬芳，教泽绵长。

家人之爱，永记于心。感谢我的爸妈对我二十多年的养育之恩，你们是我求学路上最坚实的后盾和底气，尊重我的每一个决定，让我充分自由且大胆的做出每次选择，也让我学会感知每个选择背后沉甸甸的分量。感谢我的外公外婆从小的陪伴和等待，感谢我的妹妹让我在走向成熟的道路上仍然保留一份童真，感谢我的三姨和四姨给予我的关爱、支持和帮助，愿你们岁岁年年平安喜乐。

山水一程，三生有幸。特别感谢我的好朋友余泓浩、孙立科、孔伟仁对我科研的帮助。感谢我的实验室同门鲁晔、陈佳艺，我们曾相互扶持走过这三年。感谢我的室友黄蕾、杨悦、闫云燕，我们一起分享平淡生活中的点点滴滴，一起彻夜长谈。感谢我远方的闺蜜孙妍、赵艳婷，她们教会我许多，在我心情低落时给予我安慰，在我困惑迷茫时带我走出精神内耗。愿我们此去前程似锦，再相逢依旧如故。

道阻且长，行则将至。最后我想感谢我自己。古有十年寒窗，今有二十余载苦读，一路走来，满载收获。感谢那个坚定勇敢，从未放弃的自己，那些奋笔疾书的日夜，那些含泪坚持的时刻，那些不愿妥协的瞬间，都汇成了我生命的宽度。感谢那个真诚待人，洒然行事的自己，一砖一瓦不断建立自己的内心秩序，澄明之心，自爱之心，从容之心，起舞之心，让我寻找到自洽的生活方式。感恩成长道路上的所有经历，感谢所有遇见，鲜衣怒马，未来可期。

至此，是结束亦是开始。欲买桂花同载酒，终不似，少年游，此去经年，愿走出半生，归来仍是少年。

在读期间发表的学术论文与取得的研究成果

已发表论文

1. Shuhui Yin, Yu Kang, Yun-Bo Zhao and Jian Xue, ”Air Combat Maneuver Decision Based on Deep Reinforcement Learning and Game Theory”, *2022 41st Chinese Control Conference (CCC)*, 2022, pp. 6939-6943, doi:10.23919/CCC55666.2022.9901992.