控制与决策
Control and Decision

基于DQN的多类型拦截装备复合式反无人机任务分配方法

黄亭飞，程光权，黄魁华，黄金才，刘忠

引用本文:
黄亭飞, 程光权, 黄魁华, 等. 基于DQN的多类型拦截装备复合式反无人机任务分配方法[J]. 控制与决策, 2022, 37(1): 142–150.

在线阅读 View online: https://doi.org/10.13195/j.kzyjc.2020.0787

您可能感兴趣的其他文章

Articles you may be interested in

基于动态蚁群劳动分工模型的多AUV任务分配方法

A multi–AUV dynamic task allocation method based on antcolony labor division model

控制与决策. 2021, 36(8): 1911–1919 https://doi.org/10.13195/j.kzyjc.2019.1312

面向多目标侦察任务的无人机航线规划

UAV trajectory planning for multi–target reconnaissance missions

控制与决策. 2021, 36(5): 1191–1198 https://doi.org/10.13195/j.kzyjc.2019.1284

基于两阶段迭代优化的空天观测资源协同任务规划方法

A two–stage iterative optimazation method for the coordinated task planning of space and air observation resources

控制与决策. 2021, 36(5): 1147–1156 https://doi.org/10.13195/j.kzyjc.2019.1193

多无人机协同直播场景下自适应任务卸载决策

Adaptive task offloading decision of multi–UAVs cooperation in live broadcasting scenario

控制与决策. 2021, 36(4): 974–982 https://doi.org/10.13195/j.kzyjc.2019.1104

多无人机协同直播场景下自适应任务卸载决策

Adaptive task offloading decision of multi–UAVs cooperation in live broadcasting scenario

控制与决策. 2021, 36(4): 974–982 https://doi.org/10.13195/j.kzyjc.2019.1104

基于DQN的多类型拦截装备复合式反无人机任务分配方法

黄亭飞, 程光权†, 黄魁华, 黄金才, 刘 忠

(国防科技大学 系统工程学院，长沙 410073)

摘 要: 针对当前反无人系统无法有效压制无人机的问题,使用多种拦截装备构建一种新的反无人机方法. 传统多目标优化算法无法解决动态的任务分配问题,对此,提出一种基于深度Q网络(DQN)的多类型拦截装备复合式反无人机任务分配模型. DQN模块对任务分配问题进行初期决策. 为了提高算法收敛速度和学习效率,该方法未采用下一时刻的状态来预测Q值,而是采用当前时刻的状态来预测Q值,消除训练过程中Q值过估计的影响. 之后采用进化算法对决策结果进行优化,输出多个拦截方案. 以国内某机场跑道周围区域开阔地为防护对象,构建反无人机系统的任务分配仿真环境,仿真结果验证了所提出方法的有效性. 同时,将DQN与Double DQN方法相比,所提出改进DQN算法训练的智能体表现更为精确,并且算法的收敛性和所求解的表现更为优异. 所提出方法为反无人机问题提供了新的思路.

关键词: 反无人机；深度Q网络；任务分配；Q值；多目标优化；智能体

中图分类号: TP273 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2020.0787 **开放科学(资源服务)标识码(OSID):**

引用格式: 黄亭飞,程光权,黄魁华,等. 基于DQN的多类型拦截装备复合式反无人机任务分配方法[J]. 控制与决策, 2022, 37(1): 142-150.

Task assignment method of compound anti-drone based on DQN for multi type interception equipment

HUANG Ting-fei, CHENG Guang-quan†, HUANG Kui-hua, HUANG Jin-cai, LIU Zhong

(College of Systems Engineering，National University of Defense Technology，Changsha 410073，China)

Abstract: Aiming at the problem that the current anti-drone system can not effectively suppress the drone, a new compound anti-drone method is constructed by using multiple types of intercepting equipment. The traditional multi-objective optimization algorithm cannot solve the dynamic task allocation problem, this paper proposes a task assignment model of multi-type interception equipment compound anti-drone based on a deep Q network (DQN). The DQN module makes initial decisions on task allocation issues. In order to improve the convergence speed and learning efficiency of the algorithm, this method does not use the state of the next time to predict the Q value, but uses the state of the current time to predict the Q value, while eliminating the influence of over estimation of Q value in the training process. After that, an evolutionary algorithm is used to optimize the decision-making results and output multiple interception schemes. The simulation environment of task assignment of the anti-drone system is constructed by taking the open area around a domestic airport runway as the protection object. The simulation results verify the effectiveness of this method. At the same time, compared with the DQN and Double DQN methods, the improved DQN algorithm training agent performance is more accurate, and the convergence of the algorithm and the performance of the solution are more excellent. The proposed method provides new ideas for the anti-drones problem.

Keywords: anti-drone；deep Q network；task assignment；Q value；multi-objective optimization；agent

0 引 言

近年来,随着通信和工业等领域技术的不断发展与完善,无人机的数量正经历爆发性的增长,在军事和民用领域都得到了广泛应用. 它们的身影广泛地出现在航空拍摄[1]、农业生产、植物保护、快递运输[2]、交通监控[3]、灾难救援、测绘、电力巡检等诸多领域. 其中,安全通信[4]和攻击检测[5]领域尤其引人重视,越来越多的研究人员开始将目光聚集于此.

目前,世界各国大多是将低空入侵的无人机视为传统飞行目标,普遍采用传统防空武器系统来确保打

收稿日期: 2020-06-16；**修回日期:** 2020-12-08.

基金项目: 国家自然科学基金项目(62073333)；装备发展部领域基金项目(61403120206).

责任编委: 董久祥.

†通讯作者. E-mail: cgq299@nudt.edu.cn.

击和防护效果的有效性. 虽然这样做可以确保无人机防护的有效性,但是从成本对比上,这无疑是一种战略资源的浪费,是在使用“高射炮打蚊子”. 另一方面,在设计之时,现有防空武器系统并不是用来针对无人机的,所以也不适合抵御小型、廉价无人机集群的飞行入侵[6].

针对低空无人机的诸多问题,一些行之有效的方法逐渐浮现. 这些方法可分为两类: 一类以研发新的武器装备为主,例如: 俄军在反无人机领域不断研发多种新型武器装备[7-8],这些装备分别从跟踪探测、干扰压制、控制捕获与直接摧毁4个方面实现对无人机的反制; 美军已经测试了多款不同的反无人机系统[9],用来清除无人机的潜在威胁; 以色列的拉法尔先进防御系统公司研发了一款先进的反无人机系统[10],集目标探测、识别和打击功能于一体; 英国的一家公司也推出了一款新型的无人机拦截设备[11],可以实现多频段无人机的干扰.

另一类以研发新的反无人机系统为主,旨在通过合理利用现有的设备及理论期望实现对无人机的全方位压制. 如: 胡文娟等[12]提出了一个反无人机系统的架构理论,尝试搭建一个集探测、跟踪、侦测、干扰、网捕于一体的反无人机系统; 张进等[13]结合两个子系统为一个有机整体,使用干扰设备实现了反无人机系统的多目标协同以及远程打击能力; 姚碧琛等[14]从干扰和诱导方面出发,研究了可以实现高侦测概率和有效反制的反无人机系统(anti-UAV defense system, ADS).

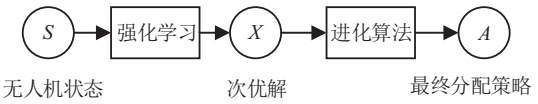
综上所述,现有的国内外反无人机装备只能在某些特定环境下具备反无人机的能力,并且能力也比较单一,具有极大的局限性. 目前,能够系统有效进行无人机防护与压制的全要素反无人机系统尚未投入到实际应用中[15].

受弹炮结合[16]的防空思想启发,针对单一拦截设备无法适应全场景防控的需求,本文提出一种采用多类型装备复合式反无人机方法. 与传统防空领域类似,针对多个来袭的“低慢小”无人机,如何针对不同类型拦截装备的属性和使用特点进行高效的任务分配是问题的核心.

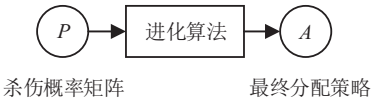
本文采用深度强化学习与进化算法相结合的方法来解决动态条件下反无人机系统的火力分配问题. 首先利用深度强化学习的决策机制进行初期的决策行为判定,即在何种情况下可以使用哪种武器打击哪个无人机; 然后使用进化算法对决策策略进行优化. 仿真结果表明,与大多数常用的求解算法相比,本文的方法不仅更易使用,还可以动态地解决目标分配问题,高效精准地求出分配结果.

1　系统建模

本文所解决的问题可以视为运筹学中的任务指派问题,但是略有不同. 传统指派问题来源于实际应用的抽象建模,解决方法是选择一个算法按照约束条件求取最优解. 本文将实际问题分为两部分: 反无人机系统马氏决策过程(MDP)的决策模型与反无人机系统优化模型,本文的整体架构如图1(a)所示.



(a)　本文算法架构



(b)　传统算法架构

图 1　指派问题的求解模型

将无人机状态输入模型之后,决策部分输出决策结果,决策结果可以视为传统数学优化中的 X 决策变量,此决策结果与传统优化过程中的次优解或差解相当. 之后,采用进化算法对次优解的决策变量矩阵进行优化,得到最优解.

1.1　与传统目标分配算法的不同

本文算法与传统的目标分配算法有3点不同.

1) 输入数据不同. 如图1(b)所示,传统目标分配算法需要输入杀伤概率矩阵,智能算法根据输入的概率进行计算优化,得出最佳的分配策略. 本文算法将模型分为两部分,前一部分的强化学习模块训练完成之后可以根据输入的无人机状态输出决策结果,再利用进化算法对决策结果进行优化.

2) 算法的应用情况不同. 本文的模型更适用于动态的目标分配问题,而传统的目标分配算法更加适用于静态的. 静态目标分配模型的目标分配与目标交战是两个相互独立的过程,其得出的结果为暂时性的最优分配. 动态的目标分配注重考虑分配时的随机事件,即力求即时且完美地解决意外出现的目标. 由于时间和随机事件的因素,问题的求解难度大大增加.

3) 在本场景中问题无法抽象为常规的目标分配问题. 由于反无人机武器的射程较近且相对的防护区域较大,单个武器的防护区域有限. 因此,面对多个方向的来袭无人机,单个武器的可打击目标有限; 其次,针对单个方向的多架次无人机,在暂时性的分配

中可使用的武器有限.

1.2 反无人机系统MDP决策模型

强化学习[17]是机器学习中最热门的领域之一,它主要被用来解决序贯决策问题,即需要智能体不断做出决策的问题. 其理念是智能体基于环境而做出动作,使得通过动作获得的收益最大化. 强化学习学习的灵感来源于心理学中的行为主义理论,即有机体如何在环境给予的奖励或惩罚的刺激下,逐步形成对刺激的预期,产生能获得最大利益的习惯性行为.

马氏决策过程(MDP)属于序贯决策的数学模型,被用于强化学习问题中的问题建模,研究智能体在环境中可实现的策略与回报. 基本原理是在一个环境中,智能体会根据当前状态做出一个动作,动作产生之后会有一个奖励(正奖励或者负奖励),智能体会根据奖励来调整下一个状态的动作,最终使得所获奖励最大化. MDP具有马尔科夫性,即系统的下一个状态仅与当前状态有关,而与历史状态无关,此处的状态是指完全可观察的全部的环境状态.

MDP包含4个要素,分别是状态、动作、策略和奖励,元组表示为(S, A, P, R). 在本系统中: S表示无人机的状态,由于本系统针对无人机的飞行高度较低,对拦截设备基本上没有影响,若在实验中考虑此参数反而会造成状态空间的扩大,不利于算法收敛,故状态信息为(x, y),表示无人机在网格化防区的位置; A表示智能体(武器)的动作,0表示静默,即不采取行动,1表示打击无人机; P表示状态之间的转移概率; R表示在状态S下采取不同控制动作A所能得到的即时奖赏[18].

在反无人机系统的MDP模型中,从防护区域输出的无人机状态S,经过转移概率P与武器动作A共同作用后,防护区域产生变化,无人机也随之进入新的状态S_{t+1}.

1.3 目标分配的优化模型

本文目标分配优化模型的数学描述与传统的目标分配十分相似. 强化学习模块的输出为X,表示$m \times n$阶的决策矩阵,m为来袭无人机的数量,n为武器的数量. 矩阵的每一行数值都表示一个完整的决策,即若$x_{i,j}$为1,则表示第j个武器打击无人机i. 同时,$x_{i,j}$中的i可以视为第i个决策方案,进化算法是对多个方案进行优化.

在进化算法中,对决策变量进行一对一打击的优化,因此,进化的结果是产生多个打击策略,且每个策略的分配结果是系统可在一对一打击的情况下打击多个无人机. 与传统算法相比,可以认为模型每次的初始化结果相同,且初始化的结果与常规算法的次优解相当,然后针对初始化结果进行优化. 需要注意的是,由于适应度函数的不同,算法不需要杀伤概率矩阵. 为确保算法可以优化出最好的结果,本文设定了足够多的迭代次数.

2 算法设计

2.1 改进DQN

本文方法的核心思想有3部分: 1)采用多智能体进行决策,智能体的数量与火力分配资源的数量相同,这是因为单独采用一个智能体来决定分配问题的状态空间太过庞大,算法难以收敛,同时无法遍历状态空间,无法在合理时间内给出满意的结果; 2)算法的记忆空间并不是常见的四元素模式,本文未考虑下一个状态的学习,在本文的问题中,算法迭代的后期计算出的无人机下一状态一般为击毁,对于智能体的学习基本无用,反而会造成状态空间的臃肿,不利于智能体学习; 3)与DQN[19-20]相比,Q值[21]的更新公式不同,具体见2.3节的式(1). 在式(1)中,使用当前状态的Q值替换下一时刻状态的Q值. 因为在本实验场景中,迭代后期,由网络$Q_t(s_{t+1}, a_{t+1})$所求得的值大多数为1,这样会造成过估计,降低解的质量,无法求得最优结果[22].

2.2 DQN环境设计

强化学习环境的核心是物理引擎——step()函数,其输入是动作a,输出是下一时刻状态s',还包括当前动作的奖励r,是否终止训练done以及调试项info,该函数描述了智能体与环境交互的所有信息. 在本系统中,该函数利用智能体的运动学模型和动力学模型计算下一步的状态和立即回报,并判断是否达到终止状态.

step()函数主要包括3部分.

1)获得系统下一时刻的状态.

系统的状态主要是由无人机的位置组成,从防护区域的边界开始,模拟意外闯入机场的无人机. 无人机在机场范围随机行动,有$(0, 1, 2, 3, 4)$五种基本动作,分别表示静止,往东移动,往南移动,往西移动,往北移动.

2)获得动作的奖励.

实验中,对智能体打击无人机时获得的奖励进行如下设定: 当无人机在武器射程之内时,若拦截设备进行打击,则奖励为1,否则为-1; 同理,当无人机在武器射程之外时,若拦截设备进行打击,则奖励为-1,否则为1.

3)获得训练的终止信号.

系统在无人机被成功击毁或者是无人机飞入机场跑道时, done = true, 结束训练; 否则, done = false, 表示系统继续进行训练, 直至达成终止条件.

2.3 DQN决策流程

DQN的learning( )函数的输入为(s_t, a_t, r_t, s_{t+1}), 分别是当前状态、当前动作、当前奖励与下一时刻状态. 大多数情况下, 智能体获得的未来状态与现状的奖励有紧密的联系, 在学习过程中需要考虑到这点. 但是, 在反无人机系统中, 它们的联系反而没有这么紧密, 并且四元组使得学习空间最大扩增了近300倍, 延缓了收敛速度, 同时增加了求解难度, 降低解的质量. 另外, Q值更新也发生了变化, 其过程为

$$Q_e(s_t, a_t) = Q_e(s_t, a_t) + \alpha[r + \gamma \max_{a_{t+1}} Q_t(s_t, a_t) - Q_e(s_t, a_t)]. \tag{1}$$

更新公式与DQN基本相同, 仅仅是未考虑下一时刻的状态, 即将$Q_t(s_{t+1}, a_{t+1})$改为$Q_t(s_t, a_t)$, 不考虑下一时刻的状态的Q值.

algorithm1为改进DQN算法的伪代码, 包括贪心策略、经验回放机制以及深度神经网络方法.

algorithm 1　改进DQN算法.

require: 状态空间S, 动作空间A, 折扣率γ, 学习率α;

ensure: 网络$Q_e(s, a)$.

1) 初始化Q网络$Q_e(s, a)$参数ϕ.

2) 初始化目标网络$Q_t(s, a)$参数ϕ'.

3) 初始化经验池D, 容量为n.

4) repeat

5) 初始化状态s.

6) repeat

7) 在状态s, 选择动作a.

8) 执行动作a, 得到全新的状态、奖励以及done.

9) 将s、a、r放入D.

10) 判定是否从经验池中学习

$$\begin{cases} 1, \text{ 从经验池中学习;} \\ 0, \text{ 继续储存经验, 跳至步骤13).} \end{cases}$$

11)

$$\text{done} = \begin{cases} 1, \ s\text{为终止状态;} \\ 0, \text{ 根据式(1)更新} Q \text{值.} \end{cases}$$

12) 使用均方差作为损失函数训练Q网络.

13) $s = s'$.

14) 每隔100步, $\phi' = \phi$.

15) until s为终止状态.

16) until $\forall s, a, Q_e(s, a)$收敛.

首先, 改进DQN算法的初始化分为3个部分: 初始化Q网络$Q_e(s, a)$参数ϕ和目标网络$Q_t(s, a)$参数ϕ', 在首次训练时, 参数ϕ与参数ϕ'相同; 本文采用的经验池D为3元组模式, 容量n需要根据具体问题来确定其大小. 算法在步骤1)～步骤3)完成这些初始化过程.

然后, 从步骤4)～步骤16)完成智能体的训练与优化. 步骤5)初始化状态s, 选择训练的起点, 状态s是从防护区边界坐标中随机抽取的. 步骤7)选择动作a, 智能体开始与环境进行交互. 步骤8)环境根据动作a做出改变, 状态s发生改变, done值发生变化, 并且环境给予智能体一个奖励r. 在步骤9)中, 将(s, a, r)放入经验池D中. 在步骤10)中, 检验经验池D是否储存完毕, 若储存完毕则进行学习, 否则继续储存. 决定从经验池D中学习后, 在步骤11)中检验是否达到终止条件, 若未达到终止条件, 则根据式(1)进行Q值更新, 并且在步骤12)中采用下式作为损失函数更新网络:

$$(r + \gamma \max_{a_{t+1}} Q_t(s_{t+1}, a_{t+1}) - Q_e(s_t, a_t))^2. \tag{2}$$

在步骤13)中进行状态s的更新, 之后在步骤14)中每隔100步更新一次ϕ'. 在步骤15)和步骤16)中检测算法的终止状态与收敛情况.

最后, 当达到终止条件时, 算法此时优化的神经网络即为需要输出的网络$Q_e(s, a)$.

algorithm 2展示的是9个智能体训练完备之后, 保存9个网络的参数, 然后根据输入的无人机状态信息输出决策结果, 即初期的分配策略. 模型的输入为无人机状态s, 与algorithm 1的状态相同; 输出为9个拦截设备的状态(由0或1表示).

algorithm 2　火力分配的求解模型.

require: 状态空间s;

ensure: 智能体的决策结果X.

1) 每隔1 s输入防护区域范围的无人机状态s;

2) 将状态空间的元素s依次输入$n_1, n_2, \ldots, n_9$等9个神经网络$Q_e(s, a)$;

3) 将智能体的决策结果输出X.

在步骤1)中模型每隔1 s输入所搜索的状态s; 随后, 在步骤2)中将状态s输入神经网络$Q_e(s, a)$, 神经网络按照训练要求输出相应的动作; 在步骤3)中输出代表智能体的神经网络决策结果.

2.4 进化算法优化流程

火力分配设备与传统的防空设备有较大的区别,即在射击飞临时间内(打击生效时间),反无人机设备无法再次打击无人机,因此,必须设计专门的框架以协调各个火力分配设备,采取一对一的打击原则,使其达到最大的防空效果. 由于智能体输出的决策结果仅仅是各自的可打击结果,还需对决策结果进行优化,具体过程如algorithm 3所示.

algorithm 3 进化算法的优化模型.

require: 决策结果X;

ensure: 最优目标分配结果A.

1) 获得初始种群(X).

2) 对初始种群编码.

3) repeat

4) repeat

5) 计算适应度.

6) 选择算子.

7) 交叉算子.

8) 判定是否进行变异

$$\begin{cases}1,\ \text{转至步骤8);}\\0,\ \text{转至步骤9).}\end{cases}$$

9) 进行变异.

10) 得到新种群.

11) until种群进化完成.

12) until达到迭代次数.

13) 输出多个决策结果.

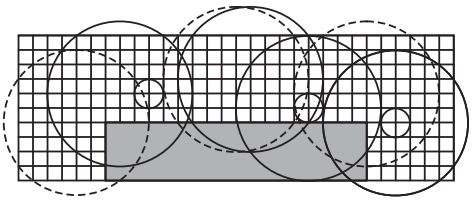
正如在1.3节中所介绍的,通过进化算法对强化学习生成的次优解进行优化,最终生成目标分配的最优解. 需要注意的是:在本实验中,因初始解为次优解,算法极易陷入局部最优解,且算法在数次迭代中即可收敛,故应设置判定条件,迭代一定次数之后(进化完成),算法重新进行演算.

3 仿真验证

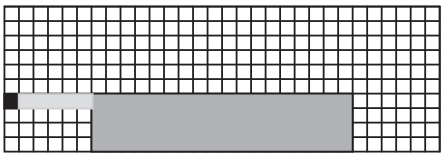
3.1 反无人机部署模型

为了衡量算法的优劣,本文采用仿真场景对算法的性能进行测试. 此处对反无人机系统的火力部署模型进行简要说明. 本文的仿真场景参考国内的机场大小以及建筑环境,考虑到火力部署资源成本问题以及机场大小,采用9个拦截设备进行部署. 具体部署情况如图2(a)所示.

图2(a)中,机场的航班跑道是一个长4.8 km,宽1.2 km的矩形. 因此,本文建立一个部署范围为3×9 km的防区. 其中,防区中的跑道范围不能部署拦截设备,考虑到拦截设备的反应时间以及二次打击间隔,还有无人机野外飞行速度的问题,将网格的长度设定为300 m. 3种不同圆圈表示部署时的3种拦截设备,同时,3种拦截设备的数量相同,都是3个. 虚线大圆圈表示激光拦截设备的防护边界,实线大圆圈表示无线电干扰设备的防护边界,小圆圈表示网式拦截设备的防护边界. 其中,下方的小圆圈是两个重合的设备部署. 图中的阴影部分为飞机跑道,机场西部(阴影部分的下边)环境较为复杂,为市内环境,防护规则与其他3面不同,不属于开阔地场景,因此本文不予考虑.



(a) 反无人机装备部署情况



(b) 来袭无人机的入侵路线

图2 反无人机仿真环境

3.2 仿真平台搭建

实验中使用的火力装备包括激光设备、网式拦截设备以及无线信号干扰设备,其相关参数见表1. 需要注意,激光与无线信号干扰器的射击速度为光速,因此武器的飞临时间可以忽略,并且网式拦截设备的防护区域较小,与“子弹”飞临速度相比,对计算结果基本无影响. 综上所述,在本研究中各个拦截设备的飞临时间可以忽略不计.

表1 反无人机系统的武器参数

参数名	激光武器	网式拦截武器	无线信号干扰设备
射程半径/m	300～1 500	20～300	0～1 500
击毁率	0.95	0.8	0.8
水平转动角度/(°)	[0, 360]	[−140, 140]	[−180, 180]
垂直转动角度/(°)	[0, 80]	[5, 60]	[−10, 60]
作战反应时间/s	6	10	3
二次拦截时间/s	6	2	0

本文研究的对象为“低、慢、小”的消费级无人机,此类无人机的速度为15～30 m/s,在仿真时,设定无人机速度为25 m/s.

3.3 仿真参数设置

1) 神经网络参数.

仿真时,DQN的两个神经网络结构相同,从输入

层到输出层分别是: 第1层为输入层, 有2个神经元; 第2层为隐藏层, 有32个; 第3层是一个Relu( )函数; 第4层为隐藏层, 有64个神经元; 第5层是一个Relu( )函数; 最后是输出层, 有2个神经元.

2) 神经网络的超参数.

超参数是DQN在开始学习之前设置的参数, 而不是通过训练得到的参数. 因此, 在训练之前需要对超参数进行优化, 选择出一组最优超参数, 从而提高学习的性能和效果.

3.4　模型算法比较

假定有10架无人机从同一方向飞来, 即图2(b)中方框位置, 飞行路线如图2(b)中的亮灰色部分所示. 可以看出, 此时可以发挥作用的拦截设备只有前3个, 即F_1、F_2和F_3. 在拦截设备的部署位置不变, 总数不变的情况下, 有4种部署方式, 分别为: M_1(拦截设备全为激光)、M_2(拦截设备全为网)、M_3(拦截设备全为无线电)和M_4(拦截设备为本文的模型). 各模型的理想打击结果如表2所示.

表 2　各个模型理想的火力最终打击结果

模型	火力	目标									
		T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}
M_1	F_1	**1**	**1**	**1**	0	0	0	0	**1**	0	0
	F_2	0	0	0	**1**	**1**	**1**	0	0	0	0
	F_3	0	0	0	0	0	0	**1**	0	0	0
M_2	F_1	**1**	**1**	0	0	0	0	0	0	0	0
	F_2	0	0	0	0	0	0	0	0	0	0
	F_3	0	0	0	0	0	0	0	0	0	0
M_3	F_1	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
	F_2	0	0	0	0	0	0	0	0	0	0
	F_3	0	0	0	0	0	0	0	0	0	0
M_4	F_1	**1**	**1**	**1**	0	0	0	0	0	0	0
	F_2	0	0	0	**1**	**1**	**1**	**1**	**1**	**1**	**1**
	F_3	0	0	0	0	0	0	0	0	0	0

从表2中可知, 部署M_1可以拦截大部分无人机, 在无人机进入防区后, 前3个目标会在30 s之内被火力F_1依次击毁, 目标4和目标5被火力F_2在接下来24 s的时间击毁, 而目标6、7和8会在接下来6 s内分别被F_1、F_2与F_3击毁. 但是最终会有3架无人机突破防区进入机场跑道, 给机场造成巨大损失. 部署M_2表现最差, 理想情况下也只能击毁2架无人机, 无法有效防护无人机. 部署M_3可以在起始的10 s之内拦截所有的无人机, 从某种意义上来说是最优秀的方式. 但是, 无线电设备在成功拦截无人机之后, 无人机会做出无法预测的行动, 例如返航、继续直线飞行、乱飞等等, 而不是传统的坠毁. 并且无线信号干扰设备对己方信号也会有干扰, 在机场尤其要注意慎用, 不能大规模部署. 部署M_4也可以拦截所有的无人机, 并且在无线信号干扰时无人机的数量降为7, 之后仍然可以使用激光拦截设备打击, 充分利用各个设备的优点, 在有效压制所有无人机的情况下降低了无线信号干扰设备的使用频率.

3.5　DQN算法结果比较

本文在实验中假设有n架无人机从北、南和东三面闯入机场, 它们飞入防护区的状态为(x_1, y_1), $(x_2, y_2), \ldots, (x_n, y_n)$. 选取2种飞临情况进行模拟, 计算相应的火力分配结果.

1) 第1种飞临情况下, 选取6个网格为无人机的进入节点, 根据无人机的进入节点距各个拦截设备的距离不同将有不同的分配结果. 训练时, 本文选择随机种子为0时所训练的模型, 由此可得算法训练9个智能体的拦截效果, 如表3所示.

表3中: S_0为无人机的进入节点距各个拦截设备的距离, S_1表示标准DQN的分配结果, S_2表示改进DQN的分配结果, S_3表示double DQN的分配结果.

从表3的结果分析可知: DQN算法训练的智能体虽然不会贻误战机, 成功地打击了所有可打击的无人机, 完成了反无人机任务, 但是它也对射程之外的无人机进行打击, 极大地浪费了火力资源; double DQN[23]算法训练的智能体与之相比虽然改善了火力资源浪费的问题, 但是忽略了某些射程之内的无人机, 例如无人机T_5, 产生的危害比火力资源浪费还要严重; 改进DQN算法训练的智能体则完美地解决了这些问题, 既没有发生浪费火力资源过度打击的问题, 也没有出现忽略射程之内无人机的现象.

将以上的火力分配矩阵输入进化算法模块, 可以得到两个打击方案A和B. 方案A和方案B的分配结果如表4所示, 具体采用何种方案由指挥官决定. 除

表3 DQN的火力分配

方法	火力	目标					
		T_1	T_2	T_3	T_4	T_5	T_6
S_0	F_1	7.89	1.74	5.59	3.08	1.51	1.61
	F_2	6.91	2.00	4.57	2.00	1.08	1.74
	F_3	1.69	7.63	1.92	4.17	6.46	7.35
	F_4	3.12	5.82	1.33	2.41	4.65	5.52
	F_5	1.69	7.63	1.92	4.17	6.46	7.35
	F_6	1.69	7.63	1.92	4.17	6.46	7.35
	F_7	4.50	4.23	2.17	0.84	3.00	3.95
	F_8	4.20	3.00	1.89	1.08	3.35	4.23
	F_9	1.89	6.94	1.08	3.42	5.76	6.67
S_1	F_1	0	**1**	0	0	**1**	**1**
	F_2	0	0	0	0	**1**	0
	F_3	0	0	0	0	0	0
	F_4	0	0	0	0	0	0
	F_5	0	0	0	0	0	0
	F_6	0	0	0	0	0	0
	F_7	0	0	0	**1**	0	0
	F_8	0	0	0	**1**	0	0
	F_9	0	0	**1**	0	0	0
S_2	F_1	0	**1**	0	0	**0**	0
	F_2	0	0	0	0	**0**	0
	F_3	**1**	0	0	0	0	0
	F_4	0	0	0	0	0	0
	F_5	0	0	0	0	0	0
	F_6	0	0	0	0	0	0
	F_7	0	0	0	**1**	0	0
	F_8	0	0	0	**1**	0	0
	F_9	0	0	**0**	0	0	0
S_3	F_1	0	0	0	0	0	0
	F_2	0	0	0	0	**1**	0
	F_3	0	0	0	0	0	0
	F_4	0	0	0	0	0	0
	F_5	0	0	0	0	0	0
	F_6	0	0	0	0	0	0
	F_7	0	0	0	**1**	0	0
	F_8	0	0	0	**1**	0	0
	F_9	0	0	**1**	0	0	0

了在火力分配设备射程之外的无人机,其他无人机均有设备进行打击,并且对于目标T_4有两个设备F_7和F_8可进行打击. 本文所提出的模型在不贻误战机的情况下,可以节约火力资源且能够提供多种任务分配方案.

2) 第2种飞临情况是指飞临时间不在同一时刻. 假定第1批无人机飞临坐标有T_4、T_5、T_6,第2批无人机飞临坐标有T_4、T_4、T_4,飞临时间的间隔为5 s,则模型最终的一个输出结果如表5所示.

在本文所提出的策略下,即在一段时间内(等待打击效果的时间)单个火力资源只打击距离自己最近的无人机,反无人机系统可以最大化毁歼能力. 若火力分配设备在第1批次时全部打击T_4,则可以获得0.96的概率击毁无人机,获得的毁歼数值为0.96,但是放走了第2批次的无人机. 而采用本文模型则可以获得1.6的毁歼数值,远大于前一种策略.

表4 不同方案的火力分配

方案	火力	目标					
		T_1	T_2	T_3	T_4	T_5	T_6
A	F_1	0	0	0	0	0	0
	F_2	0	0	0	0	**1**	0
	F_3	0	0	0	0	0	0
	F_4	0	0	0	0	0	0
	F_5	0	0	0	0	0	0
	F_6	0	0	0	0	0	0
	F_7	0	0	0	**1**	0	0
	F_8	0	0	0	**0**	0	0
	F_9	0	0	**1**	0	0	0
B	F_1	0	0	0	0	0	0
	F_2	0	0	0	0	**1**	0
	F_3	0	0	0	0	0	0
	F_4	0	0	0	0	0	0
	F_5	0	0	0	0	0	0
	F_6	0	0	0	0	0	0
	F_7	0	0	0	**0**	0	0
	F_8	0	0	0	**1**	0	0
	F_9	0	0	**1**	0	0	0

表5 模型最终的火力分配

火力	目标					
	T_4	T_4	T_4	T_4	T_5	T_6
F_1	0	0	0	0	0	0
F_2	0	0	0	0	**1**	0
F_3	0	0	0	0	0	0
F_4	0	0	0	0	0	0
F_5	0	0	0	0	0	0
F_6	0	0	0	0	0	0
F_7	**0**	**0**	**0**	**1**	0	0
F_8	**0**	**0**	**1**	**0**	0	0
F_9	0	0	0	0	0	0

3.6 DQN算法分析

在实验中,每次训练的智能体是9个火力资源中的一个. 选择loss( )函数的值和reward值两个指标来检测改进强化学习算法的收敛速度与解的质量.

在2 000次迭代中,分别求出每个周期的loss( )函数以及reward的平均值;然后对所求的9组数据取均值,即对9个智能体的表现取均值. 取20次独立重复实验的均值,得到loss( )函数的对比结果图. 为了便于观察,将对比结果图中2 000个epoch划分为20个iteration,对每个iteration内的100个数据取均值,结果见图3.

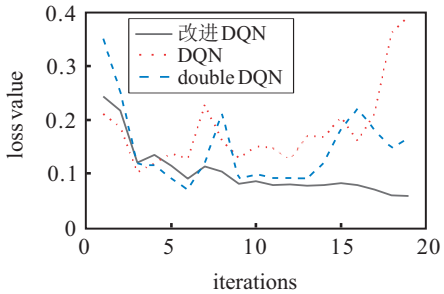


图3 20组数据的损失函数均值

图3中:横坐标代表迭代组数,每组有100次迭代,纵坐标代表每组的损失函数均值. 随着迭代次数的增加,损失函数的数值不断减小,说明算法在不断地优化调整. 当达到某一定值时,损失函数数值基本稳定,即算法完成优化,得到较优的网络参数.

从图3中可以分析出:DQN算法很不稳定,甚至在1 800次epoch左右时开始发散,实验测定,算法在此时对某些特定的学习数据产生了过拟合;在1 000次epoch左右时,改进DQN算法已经基本收敛;double DQN的收敛速度与改进DQN算法基本保持一致,但是收敛过程中的鲁棒性较差,不如改进DQN算法,并且收敛性能也不如改进DQN算法. 综上所述,本文所提出的改进DQN算法比DQN和double DQN算法拥有更快的收敛速度,而且算法的收敛性能也表现十分优异,超过了另外两种算法.

为了充分说明改进DQN算法的收敛性能,将20次实验的结果展示在图4中. 图4中:横坐标表示随机种子的数值;纵坐标表示算法的损失函数数值,数值越小表示收敛速度性能越好.

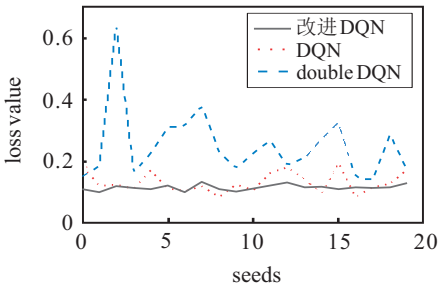


图 4　20次实验的损失函数变化

从图4的对比中可知:1)3种算法在每次仿真实验中均可以找到可行解,即算法在每次仿真实验中都会收敛;2)算法无法保证每次都可以找到最优解,不同的随机数会对实验结果产生一定的影响;3)改进DQN算法找到最优解的次数最多,鲁棒性最好.

算法reward值的对比结果如图5所示. 同样是先在2 000次迭代中取每次迭代的均值,然后在20次不同随机种子实验后,取其结果的均值. 为了便于观察,将2 000个epoch划分为20个iteration,对每个iteration内的100个数据取均值,结果见图5.

图5中:横坐标代表迭代组数,每组有100次迭代;纵坐标代表每组迭代的奖励均值. 随着迭代次数的增加,算法所获得的奖励值不断增大,说明算法在不断地优化调整. 当达到某一定值时,奖励值基本稳定,即算法完成优化(收敛),得到较优的奖励值.

由图5可知:DQN与double DQN算法在实验中的表现很不稳定,其中DQN在1 700次迭代时突然

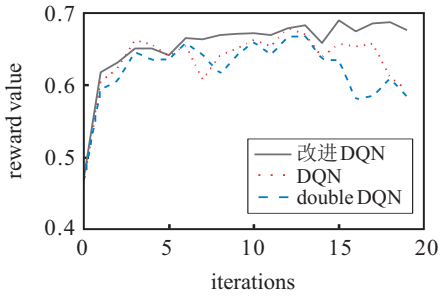


图 5　20组数据的奖励均值

下降,并且随后一直未保持稳定;而double DQN在1 500次迭代时突然下降,之后也一直未保持稳定;与之相比,改进DQN算法从800次迭代左右时,所获得reward均值一直在其他两种算法之上,并在1 500次迭代之后,算法的奖励值基本稳定.

综上分析可知:DQN与double DQN算法在搜寻最优解上出现了偏差,不仅未能成功找到最优解,而且未能在有限的迭代时间内完成收敛;改进DQN算法则完美地找到了问题的最优解,并且算法的鲁棒性更好,所获得的奖励均值一直在稳步增加,并在有限时间内保持稳定,即算法收敛.

4　结　论

本文针对开阔地场景中反无人机系统的火力分配问题,提出了一种基于DQN的改进算法. 通过对仿真实验结果的对比与分析,得到以下结论:

1)DQN及其相关算法训练的智能体可以用于解决反无人机的火力分配,与传统方法相比,不仅避开了复杂的约束和繁琐的求解过程,还可以随着状态的刷新继续生成动态的分配方案,且算法表现也非常优秀. 另外,本文提出的最短距离打击策略在反无人机中的火力分配问题上是有效的,不仅可以节约火力资源,还可以最大化毁歼能力.

2)与DQN以及double DQN相比,改进DQN算法的收敛速度更快,收敛时间节省了将近50 %,并且算法鲁棒性也有了更好的提升;同时,算法所求解的质量也更高,所获得的奖励值提升了将近20 %.

3)针对反无人机的问题,相对于国内外大多数的单一反无人机设备与方法,本文提出了复合式拦截的方法,综合了多个反无人机设备的优点,实现了对无人机的全方位防护与压制,为反无人机问题提供了借鉴.

参考文献(References)

[1]　Myslinski L J. Drone device for news reporting[P]. U.S.: 10301023. 2019-05-28.

[2]　Agatz N, Bouman P, Schmidt M. Optimization approaches for the traveling salesman problem with

drone[J]. Transportation Science, 2018, 52(4): 965-981.

[3] Kamnik R, Nekrep Perc M, Topolek D. Using the scanners and drone for comparison of point cloud accuracy at traffic accident analysis[J]. Accident Analysis & Prevention, 2020, 135: 105391.

[4] Sedjelmaci H, Senouci S M, Ansari N. Intrusion detection and ejection framework against lethal attacks in UAV-aided networks: A Bayesian game-theoretic methodology[J]. IEEE Transactions on Intelligent Transportation Systems, 2017, 18(5): 1143-1153.

[5] Su J, He J P, Cheng P, et al. A stealthy GPS spoofing strategy for manipulating the trajectory of an unmanned aerial vehicle[J]. IFAC-PapersOnLine, 2017, 49(22): 291-296.

[6] 孙健, 倪训友. 无人机国内外发展态势及前沿技术动向[J]. 科技导报, 2017, 35(9): 109.
(Sun J, Ni X Y. Development trend and frontier technology trend of UAV at home and abroad[J]. Science & Technology Review, 2017, 35(9): 109.)

[7] 张甲帅, 杨作琛, 郭欢. 俄军加强反无人机能力建设的主要方法[J]. 飞航导弹, 2020(5): 40-43.
(Zhang J S, Yang Z C, Guo H. Main methods for Russian army to strengthen anti UAV capability construction [J]. Aerodynamic Missile Journal, 2020(5): 40-43.)

[8] 李富良, 胡荣, 韩涛, 等. 俄罗斯反无人机策略与装备发展现状[J]. 飞航导弹, 2019(9): 53-58.
(Li F L, Hu R, Han T, et al. Russian counter-drone strategy and equipment development status[J]. Aerodynamic Missile Journal, 2019(9): 53-58.)

[9] 李磊, 申超, 方圆. 透过美2017机动火力集成试验演习看美陆军反无人机能力发展[J]. 飞航导弹, 2017(11): 1-5.
(Li L, Shen C, Fang Y. U.S. army counter-drone capability development through the U.S. 2017 mobile firepower integration test exercise[J]. Aerodynamic Missile Journal, 2017(11): 1-5.)

[10] 沉舟, 车易. 以色列开发新的反无人机系统[J]. 飞航导弹, 2016(11): 1-1.
(Chen Z, Che Y. Israel develops new anti-drone system[J]. Aerodynamic Missile Journal, 2016(11): 1-1.)

[11] 雪莉, 萌萌. 无人防务公司发布反无人机新概念[J]. 飞航导弹, 2016(10): 1-1.
(Sherry, Meng M. Unmanned defense company announces new counter-UAV concept[J]. Aerodynamic Missile Journal, 2016(10): 1-1.)

[12] 胡文娟, 张毅. 反无人机系统的研究与实现[J]. 中国民航飞行学院学报, 2020, 31(2): 30-34.
(Hu W J, Zhang Y. Research and implementation of anti UAV system [J]. Journal of Civil Aviation Flight University of China, 2020, 31 (2): 30-34.)

[13] 张进, 薛德鑫, 王奉甲. 新型重点区域无人机防控系统[J]. 现代防御技术, 2020, 48(1): 11-18.
(Zhang J, Xue D X, Wang F J. New UAV control system for key areas[J]. Modern Defense Technology, 2020, 48(1): 11-18.)

[14] 姚碧琛, 陈建华. 基于多探测技术的反无人机系统研究[C]. 2019世界交通运输大会论文集(上). 北京, 2019: 1023-1025.
(Yao B C, Chen J H. Research on anti UAV system based on multi detection technology[C]. Proceedings of 2019 World Transport Conference (Part I). Beijing, 2019: 1023-1025.)

[15] 张静, 张科, 王靖宇, 等. 低空反无人机技术现状与发展趋势[J]. 航空工程进展, 2018, 9(1): 1-8.
(Zhang J, Zhang K, Wang J Y, et al. A survey on anti-UAV technology and its future trend[J]. Advances in Aeronautical Science and Engineering, 2018, 9(1): 1-8.)

[16] 王越, 周德云, 刘建生, 等. 弹炮结合武器编队火力分配算法[J]. 火力与指挥控制, 2018, 43(3): 16-20.
(Wang Y, Zhou D Y, Liu J S, et al. Firepower distribution algorithm of missile gun combined weapon formation [J]. Firepower and Command Control, 2018, 43(3): 16-20.)

[17] Szepesvári C. Algorithms for reinforcement learning[J]. Synthesis Lectures on Artificial Intelligence and Machine Learning, 2010, 4(1): 103.

[18] Duan Y, Chen X, Houthooft R, et al. Benchmarking deep reinforcement learning for continuous control[C]. International Conference on Machine Learning. New York, 2016: 1329-1338.

[19] Wei T S, Wang Y Z, Zhu Q, et al. Deep reinforcement learning for building HVAC control[C]. The 54th ACM/EDAC/IEEE Design Automation Conference (DAC). Austin, 2017: 1-6.

[20] Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529-533.

[21] Watkins C J, Dayan P. Technical note: Q-learning[J]. Machine Learning ,1992, 8(3/4): 279-292.

[22] 郭宪, 方勇纯. 深入浅出强化学习: 原理入门[M]. 北京: 电子工业出版社, 2018: 88-104.
(Guo X, Fang Y C. Intensive learning in a simple way: Introduction to principles[M]. Beijing: Publishing House of Electronic Industry Press, 2018: 88-104.)

[23] Hasselt H V, Guez A, Silver D. Deep reinforcement learning with double Q-learning[C]. Proceedings of the 30th AAAI Conference on Artificial Intelligence. Phoenix, 2016: 2094-2100.

作者简介

黄亭飞(1992−), 男, 助理工程师, 硕士, 从事运筹规划与强化学习的研究, E-mail: huangtingfei18@163.com;

程光权(1982−), 男, 副研究员, 博士, 从事复杂网络分析和决策支持技术等研究, E-mail: cgq299@nudt.edu.cn;

黄魁华(1986−), 男, 副研究员, 博士, 从事智能态势认知和智能任务规划等研究, E-mail: kuihh@163.com;

黄金才(1973−), 男, 研究员, 博士, 从事智能调度与控制等研究, E-mail: huangjincai@nudt.edu.cn;

刘忠(1968−), 男, 教授, 博士, 从事多智能体系统等研究, E-mail: phillipliu@263.net.

(责任编辑：李君玲)