文章编号 1672-6634(2024)04-0023-10 **DOI** 10.19728/j.issn1672-6634.2023120005

基于DQN出价策略的多无人机目标分配拍卖算法

陈梓豪,胡春鹤

(北京林业大学 工学院,北京 100083)

摘 要 为实现多无人机监测目标分配任务匹配度、成功率等收益最大化及路径长度、障碍物碰撞风险等代价最小化,基于数据样本驱动的强化学习方法,提出了一种融合深度Q网络(Deep Q-network,DQN)出价策略的自主进化拍卖算法。首先,构建了多无人机任务目标拍卖的马尔科夫决策模型,并且分别以竞拍者剩余竞拍容量为环境,输出增价因子为动作,前后两轮拍卖收益增幅为回报。其次,构建了新型的DQN出价和竞拍决策神经网络模型。该模型通过构建包含拍卖环境、增价因子、回报等元素的强化学习训练样本库,在拍卖过程中以一种离线学习模式不断训练DQN神经网络,使其按照DQN策略在拍卖过程中,根据拍卖环境输出增价因子,实现拍卖结果收益的优化。最后,通过多无人机多监测目标分配仿真,验证了所提出基于DQN拍卖机制的目标分配方法的有效性。通过与传统拍卖算法结果相比,方法获得的拍卖收益提升21.4%。

关键词 多无人机;拍卖算法;DQN;多目标分配;拍卖收益

中图分类号 TP181

文献标志码 A

开放科学(资源服务)标识码(OSID)

Multi Drones Target Allocation Auction Algorithm Based on DQN Bidding Strategy

CHEN Zihao, HU Chunhe

(School of Technology, Beijing Forestry University, Beijing 100083, China)

Abstract To maximize the revenue of task matching and success rate for multi drone monitoring target allocation, as well as minimize the costs of path length and obstacle collision risks, this paper proposes an autonomous evolutionary auction algorithm based on a data-driven reinforcement learning method that integrates deep Q-network (DQN) bidding strategy. This algorithm first constructs a Markov decision model for multi drone task target auction, and constructs a new type of DQN bidding and auctioning decision neural network model with taking the remaining bidding capacity of bidders as the environment, outputting the price increase factor as the action, and the increase in auction revenue before and after the two rounds as the return. This model constructs a reinforcement learning training sample library that includes elements such as auction environment, markup factors, and returns. During the auction process, the DQN

收稿日期:2023-12-11

基金项目:国家自然科学基金项目(61703047)资助

通信作者:胡春鹤,男,汉族,博士,副教授,研究方向:无人机协同控制,E-mail:huchunhe@bjfu.edu.cn。

neural network is continuously trained in an offline learning mode, allowing it to output markup factors based on the auction environment according to the DQN strategy during the auction process, achieving optimization of auction results and revenue. Finally, the simulation of multi drone and multi monitoring target allocation verifies the effectiveness of the target allocation method based on the DQN auction mechanism proposed in this paper. Compared with the results of traditional auction algorithms, the auction revenue obtained by this method is increased by 21.4%.

Key words multi drones; auction algorithm; DQN; multi-objective allocation; auction proceeds

0 引言

随着人工智能的发展与应用,多无人机监测目标分配的最优性以及代价逐渐成为当前多智能体领域研究的热点[1]。无人机作为人工智能的应用主体被广泛的应用于现代工业生产[2],由于其便捷、安全的特点,无人机多被用于复杂环境下的勘测,小目标监测等[3]。无人机目标分配的结果优劣直接影响无人机协作的工作效率与质量[4],尽可能优化任务执行方案,以最小的代价获得最高的收益,是当前无人机监测面临的主要挑战[5]。当目标需要多方位多次监测,并且单个无人机可完成多次监测时,如何最大化监测收益成了多无人机监测目标分配的难点所在。在多对多无人机目标分配环境,如何制定执行方案并且更高效完成任务是本文研究主要内容[6]。

近年来国内外众多学者对此进行研究,常见的任务分配算法主要有基于合同网的市场拍卖方法[7]、分布式马尔可夫决策方法和模型预测控制方法[8]。目前传统算法是基于拍卖算法以及改进方法的延伸[9],其中基于组合拍卖法的分布式的多机器人任务分配的建模方法[10],为任务分配问题提供最优解或次优解。基于一致性的捆绑拍卖算法(CBBA)是一种利用拍卖算法的多任务分配算法[11],允许智能体对每个任务进行竞价,然后开发一个共识算法来解决智能体之间的分配冲突。为研究不确定情形下,在线多源、多属性反向拍卖(OMSMARA)双边协商决策问题[12],基于拍卖的动态任务分配算法,针对任务分配或执行过程中,随时有新任务出现的情况,研究人员提出了一种改进算法,即双层分布迭代算法[13]。其中,基于合同网的拍卖算法面向动态任务的自主分配,并在过程中考虑智能体之间的信息交互和协商[14],是一种相对容易实现的分布式算法,各智能体用贪婪算法构造组合投标,通过信息交换与相互协商实现任务的分配与协调,达到任务分配方案的一致[15]。目前传统拍卖算法在单目标、单智能体分配问题中效果较好,但应用在多智能体多目标交叉分配中存在迭代次数过多,以及结果非最优的问题[16]。在实际实验中,维度较小的模拟拍卖环境得出的收益往往小于使用穷举法所得收益,体现出传统拍卖算法的非最优性。

强化学习在任务分配中也有广泛的应用,包括针对多机器人任务分配方法在环境复杂性增加时,出现的维度灾难问题,提出了一种基于启发式深度 Q 学习的多机器人多任务分配算法[17],缓解了复杂环境下多机器人、多任务分配的维度灾难问题[18]。基于多智能体深度强化学习的空间众包任务分配,提出了允许但不强制合作的空间众包场景[19],提高智能体的整体收益和复杂任务的完成率。除了基于组合安全路径 Option 分层强学习的目标分配算法之外,针对目标分配算法有顺序扩展一致性包算法,也被用来求解分布式目标分配问题[20]。拍卖思想和强化学习在目标分配中被广泛应用[21],但是传统的强化学习解决多目标分配问题忽略了各个智能体之间的相互影响,而拍卖的方法用于解决这类问题时,不能够考虑决策对于未来拍卖进程的影响,同时之前的进程也会对决策产生未知的影响,这可能导致最终结果非最优。

基于此,本文结合多目标拍卖方法和强化学习思想,提出一种基于 DQN 出价策略的多无人机目标分配拍卖算法(Auction algorithm based on DQN bidding strategy,BDQNBA)。相比于之前的多目标拍卖方法,DQN 拍卖算法通过引入增价因子和奖励权重因子,以策略训练的方式自适应调节增价因子的大小,实现多目标分配问题收益上的最大化。实验证明,结合强化学习的方法 DQN 拍卖算法在多对多目标分配中获得了更合理的出价,使得拍卖算法在多目标问题中迭代次数更少,提升目标分配效果。

1　问题描述和建模

本文研究了多无人机协同目标分配问题。多监测目标分配问题,是多个需要共同监测的目标,共享无人机资源的优化问题。要求在完成监测所有目标点任务,并且遵循任务点检测需求的同时,监测效率最高。多无人机协同任务分配,需要考虑无人机单元的检测覆盖范围、不同目标点的检测效率和无人机的航程,是一个复杂的多约束优化问题。

以多无人机联合目标监测为任务背景,假设有 N_t 个目标 $T=\{T_1,T_2,\cdots,T_{N_t}\}$ 待监测,有 N_v 架无人机 $U=\{U_1,U_2,\cdots,U_{N_v}\}$ 用于执行监测任务,其中 N_t , N_v 为在在某次任务分配环境中目标与无人机的数量,为方便计算收益和约束条件,引入目标分配变量 X_{ij} ,若 $X_{ij}=0$,则无人机 U_i 不监测目标 T_j ,若 $X_{ij}=1$,则无人机 i 监测目标 j 。对于多无人机多监测目标分配问题,其主要考虑的约束条件如

$$\forall j,\sum_{i=1}^{N_v}x_{ij}=N_j, \tag{1}$$

$$\sum_{i=1}^{N_v}U_i\geqslant\sum_{j=1}^{N_t}N_j, \tag{2}$$

$$\forall i,T_{C_i}\subseteq T, \tag{3}$$

式中每架无人机可执行一个任务,任务 T_j 由于其本身复杂程度以及其所处位置,需 S_J 架无人机协同监测完成。无人机 U_i 可监测的目标集合为 $T_{U_i}=\{T_{U_{i1}},T_{U_{i2}},\cdots,T_{U_{in}}\}$,其中对于所有 i , $U_{in}<N_t$ 。每个目标至少需要无人机监测次数 $N_{\text{object}}=\{N_1,N_2,\cdots,N_{N_t}\}$, N_j 为目标 j 至少需要 N_j 次监测,才能完全获得其所需信息。以最大化整体任务完成收益为目标, 构建全局的目标函数如

$$\max\sum_{i=1,j=1}^{N_v,N_t}x_{ij}\cdot\boldsymbol{V}[U_i,T_j], \tag{4}$$

式中无人机 U_i 监测 T_j 所得收益为 $\boldsymbol{V}[U_i,T_j]$, $\boldsymbol{V}$ 为价值矩阵,可用以表示无人机监测特定目标面临的碰撞风险、距离、方位角度。为减少变量数目,考虑所有影响因素量化为 $\boldsymbol{V}$ 值,构成价值矩阵。

2　基于 DQN 机制的改进拍卖算法

本文提出的改进拍卖算法基于传统拍卖算法增加竞拍者容量,竞拍对象容量等因素,由传统的一对一分配方式转变为多对多目标分配。此算法主要包括竞拍者出价以及安排分配两个部分,其中竞拍者出价部分本质是对所有竞拍者进行循环, $P=\{P_1,P_2,\cdots,P_{N_t}\}$ 为目标的当前被出价的最小值,所有的目标价格初值为 0,随着拍卖进程,每个竞拍对象的价格也会随之变化。

2.1　拍卖出价竞拍策略

本文提出的改进拍卖算法基础逻辑为每轮拍卖所有竞拍者参与竞拍,每个竞拍者对于每个对象均有一个出价 P ,每轮拍卖需选择出一个对象及为之出价的竞拍者,以及该竞拍者对于该对象的出价 P 作为此轮拍卖出价部分的输出,将该输出输入到竞拍部分后,对比该对象现有价格,若此轮出价大于现有价格则进行替换,现拍得者为此轮输出最高价竞拍者。若无竞拍者出价大于当前价格,则进入下轮竞拍。

2.2　DQN 机制影响下拍卖算法

DQN 算法是一种融合了深度学习和 Q_learning 的强化学习方法,将状态和动作当成神经网络的输入,然后经过神经网络分析后得到动作的 Q 值,使用神经网络取代 Q 表以适应复杂问题,输出 Q 值最大的动作,作为算法的动作输出。本文实验中采用 Adam 和 Nadam 两种优化器,这两种优化器结合了动量和 RMSProp,利用了梯度的一阶矩估计和二阶矩估计,动态调节每个参数的学习率,并且加上了偏置修正,不同的优化器在训练过程中可能会对损失函数 L 产生影响进而影响学习率,从而影响拍卖算法最终收益结果是否最优。本文的 DQN 主体由两个 Qnet 神经网络构成,分别是估计值网络和目标值网络,每个神经网络分别由三层全连接层构成,神经网络结构如表 1 所示。

表中 N 为状态中竞拍者的数量。DQN 学习的目标是使当前值网络输出的 Q 估计值与目标值网络输出的目标 Q 值越相近越好,该过程可以通过损失函数表示为

$$L(\theta)=\mathrm{E}[(Q_{\text{target}}-Q(S_t,a_t,;\theta))^2], \tag{5}$$

式中 $Q(S_t,a_t,;\theta)$ 是当前状态的 Q 估计值,而 Q_{target} 是目标值,其由当前的奖励 r 加上折扣因子 γ 乘以目标值网络的输出值 $\max_{a_{t+1}}Q(S_{t+1},a_{t+1},;\theta)$ 表示,即

$$Q_{\text{target}}=r+\gamma\max_{a_{t+1}}Q(S_{t+1},a_{t+1},;\theta), \tag{6}$$

估计值网络的参数更新需要通过求损失函数的梯度得到,而目标值网络则通过每 N 步复制一次估计值网络的参数进行更新。在传统拍卖算法里,某次出价过程中竞拍者 U_i 对 T_m 的出价策略为

$$P_{i,m}=\boldsymbol{V}[U_i,T_m]-V_{i,\text{subm}}+\Delta\text{esp}, \tag{7}$$

式中 Δesp 值默认为 1,主要作用为在固定拍卖流程中,保证参与拍卖成本增加以避免产生过多次无效出价造成无效迭代,$\boldsymbol{V}_{i,\text{subm}}$ 为此轮拍卖中竞拍者对所有竞拍对象出价中的次大值。出价在拍卖算法中占据主导作用,不同情况下不同竞拍者对不同对象的出价大小,决定了拍卖流程以及最终拍卖结果的收益。而 Δesp 值与 $\boldsymbol{V}[U_i,T_m]$ 、$\boldsymbol{V}_{i,\text{subm}}$ 共同构成了竞拍者的出价,通过改变传统固定为 1 的 Δesp 值可以间接影响到竞拍者的出价大小影响。因此,本策略针对 Δesp 值进行研究,使用 DQN 算法捕捉有效拍卖环境信息,对其进行更准确,有针对性的赋值。在不同的拍卖环境下通过输出合适的增价因子来影响竞拍者出价,使算法结果达到最优。

为了避免强化学习的状态之间存在的相关性,本文采取的 DQN 算法采用回放记忆单元,即经验池来存放状态,动作和奖励组成的元组为 (S_t,a_t,r,S_{t+1}) 。本文中结合拍卖算法的学习方式为将拍卖过程中产生的收益、环境、动作元组 (S_t,a_t,r,S_{t+1}) 纳入经验池中,待经验池达到一定规模后,从中分次随机取出一定储量的元组,用于 DQN 算法对神经网络进行训练,使得神经网络提供的 Δesp 值能够更加契合当前的拍卖环境,促使拍卖流程最终走向更优的结果。

表 1 神经网络结构

Table 1 Neural network structure

Layer	Input×Output
Fc1	$(N+1)\times128$
Fc2	128×128
Fc3	128×2

表 2 DQN 机制影响下拍卖算法出价策略

Table 2 Bidding strategy of auction algorithm under the influence of DQN mechanism

Bidding strategy of auction algorithm under the influence of DQN mechanism
1 DQN in->S:{N,P}
2 For Every bidder U_i
3 If The bidder has not been assigned
4 For Every objects T_m
5 If The object needs to continue to assign bidders $N_i>0$
6 Δesp=DQN out->a
7 $P_{i,m}=\boldsymbol{V}[U_i,T_m]-\boldsymbol{V}_{i,\text{subm}}+\Delta\text{esp}$
8 $P=[P,P_{i,m}],N_i-=1$,DQN train
9 End If
10 End for
11 End If
12 End for

3 多无人机监测目标拍卖

改进的拍卖算法可以用于多对多目标分配任务。在无人机目标分配中,竞拍者为无人机,竞拍对象为被

分配的待监测目标，拍卖过程即为无人机监测目标匹配过程，无人机作为竞拍者，拍得的对象即为需要监测的目标，竞拍对象可以被多个竞拍者同时拍得，监测目标被多个无人机拍得，即代表此目标在此轮监测中将由这些无人机进行监测。对无人机集合中的所有无人机进行循环，对于无人机 U_i ，当前被分配的目标数量为 c_i 。当 $c_i=0$ 时，代表该无人机尚未被分配，可以作为竞拍者参与此轮拍卖，竞拍者 U_i 此轮拍卖可对竞拍对象进行出价，计算该竞拍者对于所有竞拍对象的收益，其中竞拍者 U_i 对于竞拍对象 T_j 的收益为

$$V_{ij}=\boldsymbol{V}[U_i,T_j]-P_j, \tag{8}$$

$V_i=\{V_{i1},V_{i2},\cdots,V_{iN_t}\}$ 为竞拍者 U_i 对所有竞拍对象收益的集合，其中最大值 V_{im} 对应的竞拍对象 T_m 为该竞拍者此次出价对象，其出价为

$$P_{\mathrm{bid}\,im}=\boldsymbol{V}[U_i,T_m]-V_{i,\mathrm{subm}}+\Delta\mathrm{esp}。 \tag{9}$$

遍历所有竞拍者，得出所有为该竞拍对象出价的竞拍者，并将所有的竞拍者与其出价加入竞拍对象的竞拍队列中，假设共计 l 个竞拍者为其出价，竞拍对象 T_j 的竞拍队列为

$$B_{T_j}=\begin{Bmatrix}[U_{B_1},P_{\mathrm{bid}B_1j}]\\[U_{B_2},P_{\mathrm{bid}B_2j}]\\\cdots\\[U_{B_l},P_{\mathrm{bid}B_lj}]\end{Bmatrix}。 \tag{10}$$

遍历所有竞拍者，假设共计 s 个竞拍对象得到出价，此轮出价的结果为所有被出价的竞拍对象的竞拍队列的集合为

$$B_{\mathrm{new}}=\begin{Bmatrix}B_{T_{q1}}\\B_{T_{q2}}\\\cdots\\B_{T_{qs}}\end{Bmatrix}。 \tag{11}$$

安排分配部分为循环出价部分所得结果中所有被出价的竞拍对象，根据此轮出价调整分配结果。假设当前分配结果为

$$A_{\mathrm{ss}}=\begin{Bmatrix}(U_{i_1},T_{j_1},P_{\mathrm{bid}i1j1})\\(U_{i_2},T_{j_2},P_{\mathrm{bid}i2j2})\\\cdots\\(U_{i_n},T_{j_n},P_{\mathrm{bid}injn})\end{Bmatrix}, \tag{12}$$

共计 n 组分配，初始分配 $n=0$，遍历集合中所有被出价的竞拍对象，如果竞拍对象 T_j 的最高出价 $P_{\mathrm{bid}ij}>P_j$，则将 P_j 修改为 $P_{\mathrm{bid}ij}$ ，在 A_{ss} 中将 P_j 对应的竞拍者替换为 U_i ；如果遍历 B_{new} 中所有竞拍对象没有做出调整，则拍卖结束，当前 A_{ss} 即为算法输出结果。

本文在改进传统拍卖算法的基础上使用强化学习方式对无人机目标分配进行建模，环境可以看作是拍卖进程中拍卖对象的竞拍价格，作为强化学习的环境输入，每个竞拍者看作是一个智能体。假设当前对 N_t 个目标，N_v 架无人机进行目标分配，该模型中的状态、动作及即时奖励等基本要素建模如图 1 所示。

状态：用 State$=\{P_1,P_2,\cdots,P_{N_t},O_{N_v}\}$ 来表示状态的定义，状态空间的维度为 N_t+1，其中状态 P_i 表示第 i 个目标当前竞拍最低价，O_{N_v} 表示当前剩余可分配无人机数量。

动作：Action$=\Delta$esp，其中 Δesp 表示针对该环境此轮拍卖中的增价因子，即式(7)中 Δesp。

奖励：Reword 奖励有两部分构成，为保证动作的实时有效性和总体的进步性，将此次拍卖阶段与上次拍卖收益之差 D 作为本轮决策的奖励的一部分，

$$D=\sum_{k=1}^{n}\boldsymbol{V}\begin{bmatrix}A_{\mathrm{ss},t+1}(k)(0)\\A_{\mathrm{ss},t+1}(k)(1)\end{bmatrix}-\sum_{k=1}^{n}\boldsymbol{V}\begin{bmatrix}A_{\mathrm{ss},t}(k)(0)\\A_{\mathrm{ss},t}(k)(1)\end{bmatrix}, \tag{13}$$

式中 $A_{ss}=\{(U_{i_1},T_{j_1}),(U_{i_2},T_{j_2})\cdots,(U_{i_n},T_{j_n})\}$ 为第 t 轮拍卖结果,共分配了 n 对分派关系。为保证拍卖结束后的整体收益,避免陷入局部最优,将结束拍卖后的总收益分布到每一步的值V_{av} 作为奖励的另一部分,

$$V_{av}=\sum_{k=1}^{s}\boldsymbol{V}\begin{bmatrix}A_{ss}(k)(0)\\A_{ss}(k)(1)\end{bmatrix},\tag{14}$$

式中 A_{ss} 为一轮拍卖结束后的分派结果,此轮拍卖共分配了 s 对分派关系。因此,结合 DQN 算法的一轮拍卖动作的回报为

$$r=k_1D+k_2V_{av},\tag{15}$$

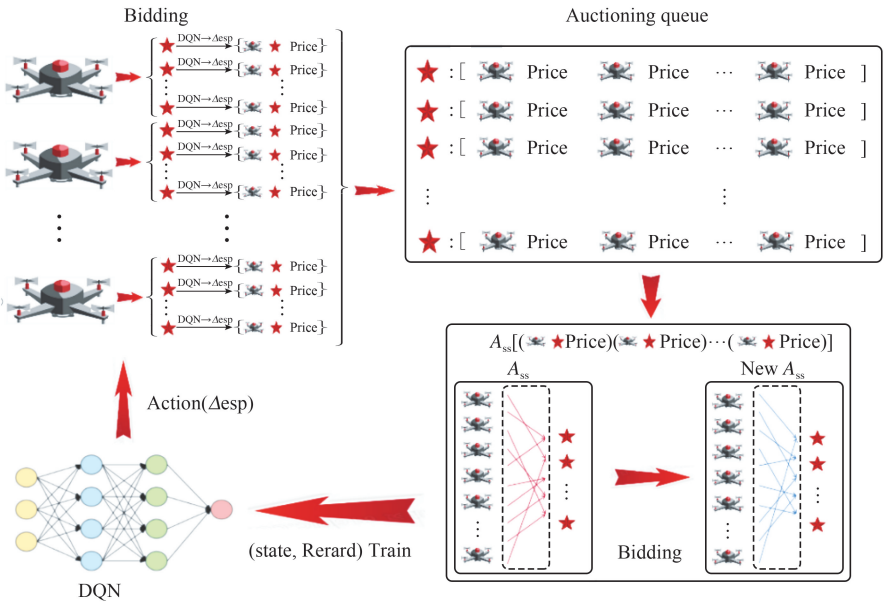

图 1 多无人机检测目标拍卖流程示意图

Fig. 1 Schematic diagram of auction process of multi-UAV detection target

式中 k_1,k_2 为调整回报中步差值部分和最终收益占比的权重系数,当需要训练较快呈现增长趋势时,应将 k_1 增大,当需要训练最终收益较高时,应将 k_2 增大。

4 结果与分析

为了验证改进后 BDQNBA 算法应用于躲无人机多监测目标分配和竞拍决策模型的有效性,本节以多重约束环境下的多无人机协同监测问题为任务目标,在配置为 AMD Ryzen 75800H 3.20 GHz, 16 G 内存,单 GPU3060,Windows 10 操作系统的计算机上,利用 Matlab 软件进行山地模型搭建和路径优化,使用 Python3.7 和 Pytorch 框架对多无人机多目标分配方案进行了仿真。设计 25 架无人机协作,监测 10 个目标场景下目标分配模型实例,对所提算法的性能进行了评估验证。强化学习训练使用的数据在实验进行中即时产生,构建的模拟山地多障碍物环境模型如图 2 所示。

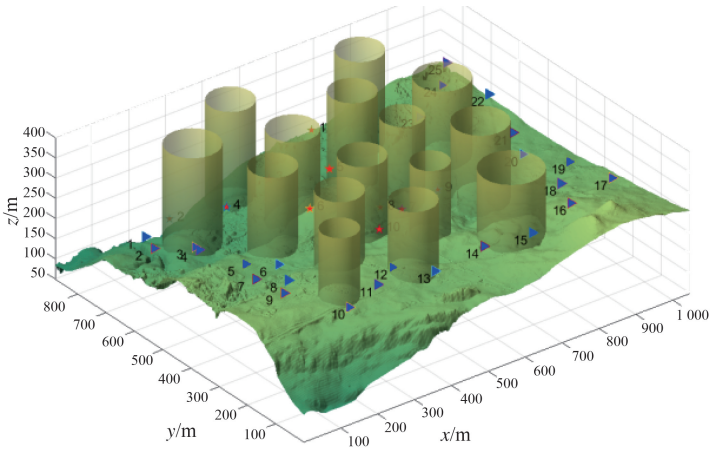

图 2 模拟山地障碍实验环境

Fig. 2 Simulated mountain obstacle experimental environment

综合考虑无人机飞行距离,碰撞风险,飞

行角度变化等因素构建无人机执行检测任务代价模型，每个任务所需监测次数如表 3 所示。

表 3　任务所需无人机数量

Table 3　Number of UAVs required for the mission

Target	Number of UAVs required	Target	Number of UAVs required
B1	2	B6	3
B2	1	B7	2
B3	6	B8	1
B4	3	B9	4
B5	1	B10	2

本实验中评价任务分配效果优劣标准为训练后网络在每次拍卖结束所获得的收益。为证明强化学习对于拍卖算法改进的有效性，对比传统拍卖算法不难发现，传统的拍卖算法为固定流程，所以对于固定场景，计算结果具有唯一性。对于该实验场景，在传统拍卖算法增价因子固定为 1 的情况下，所得结果总收益为 577，而无干预的随机分配方案收益在 300 左右。

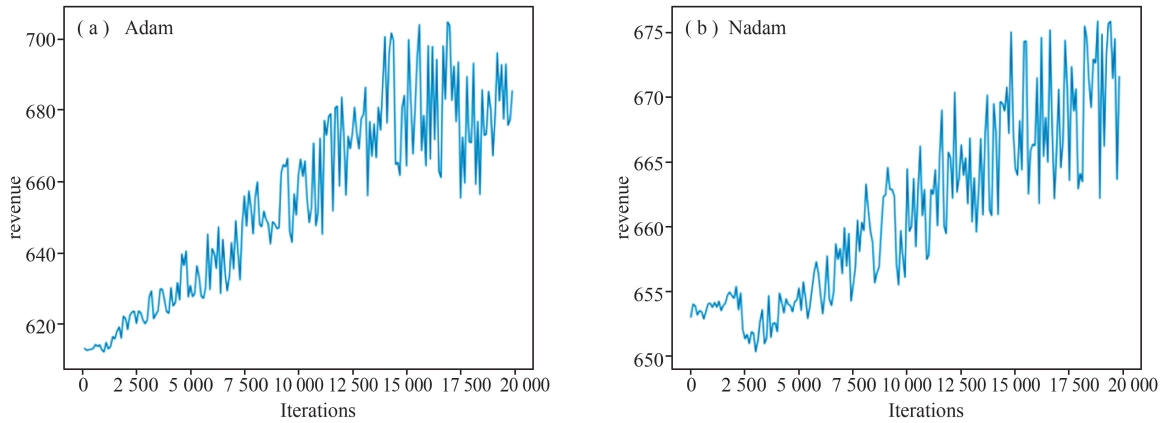

图 3　不同优化器实验训练过程

Fig. 3　Experimental training process of different optimizers

为更加准确的体现学习过程，使算法能够收敛，保证探索率随着算法的迭代过程逐渐减少，可设置奖励构成权重为 $k_1=0.1$，$k_1=0.5$，探索率范围为 0%～100%。在使用强化学习训练的神经网络给出增价因子的情况下，分别使用 Adam 和 Nadam 两种优化器，实验模型拍卖收益随训练次数变化过程以及训练后算法给出的最优分配结果如图 3，4 所示，其中曲线纵向数值为每训练 100 次的平均收益。由图 3 可以看出，随着 DQN 算法迭代，由神经网络给出的 Δesp 值使得基于拍卖算法的任务分配结果收益呈增长趋势，同时在训练后期相对于传统算法收益更高，并且在 15 000 次左右达到最优并趋于稳定。

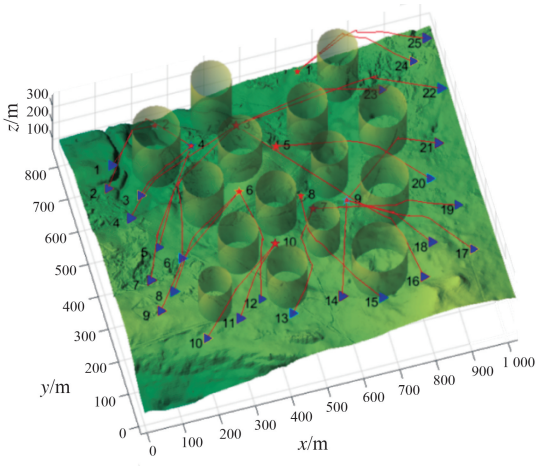

图 4　实验分配结果

Fig. 4　Experimental distribution results

为验证算法在不同环境下解决多对多目标分配的可行性与适用性，在原环境基础上设计对比实验：在原实验无人机及目标数量保持不变情况下，改变无人机和目标位置，其坐标在地图中随机生成，做 30 组不同位置的对比实验，作为对比实验组 1，部分实验结果如图 5 所示。

另在保证每个任务有足够数量无人机执行的情况下，改变无人机和目标数量，设置无人机 A1～A15，目

标 B1～B6,作为一组不同数量对比实验组 2,实验结果如图 6 所示。

各组实验模型在同样拍卖策略下使用强化学习训练的神经网络给出增价因子各自进行训练,同时为证明本文算法的有效性,本文选取算法 PSO[22]、HFPSO[23]、CLPSO[24]、EPSO[25]、GGL-PSOD[26]等优化算法在各实验组环境参数下进行对比实验[17],训练后 BDQNBA 算法给出的分配收益提升与传统算法结果以及各对比算法结果如表 4 所示。

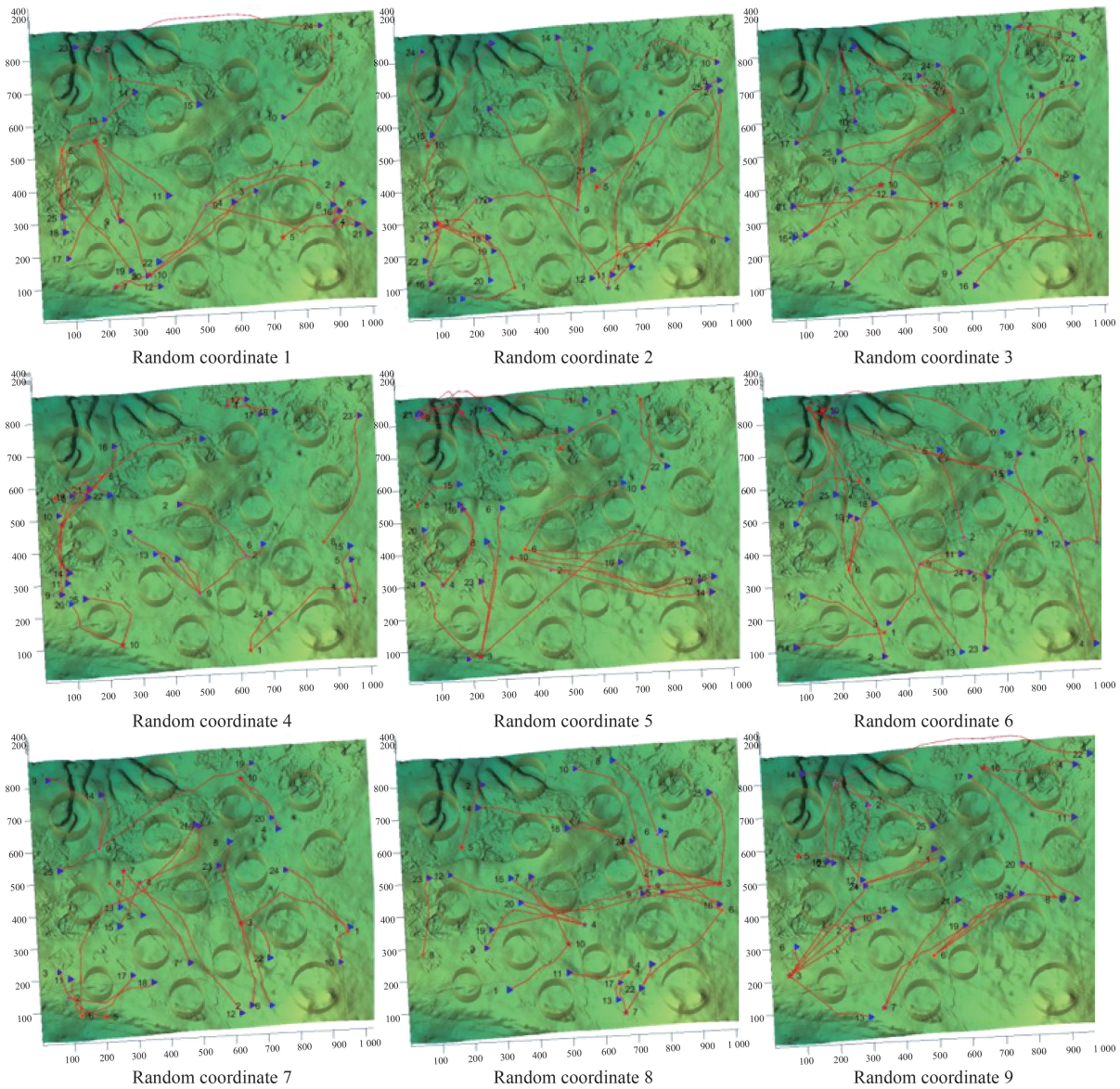

图 5　对比组 1～9 部分实验分配结果

Fig. 5　The distribution results of some experiments in comparison group 1～9

以上为实验组以及对比组为分别经过 20 000 次迭代训练后的结果,对比组为按照其原算法逻辑应用于本文实验场景所得结果。由对比实验可以得出,BDQNBA 算法通过使用强化学习控制增价因子的方式,在不同的环境下可在传统算法基础上收益提升 21.4%,在所有实验场景中规划收益值都要优于其他五种比较算法,表现出最好的性能。在应用于多无人机监测目标分配时,可使监测效率获得提升。

综上所述,算法性能以及最终结果受任务场景、参数等因素影响,如初始化路径结果、场地起伏、障碍物的大小数量位置、无人机和任务点数量位置以及训练目标即参数 k_1、k_2 的设置。在使用结合强化学习的拍卖算法进行目标分配时,随着不断训练神经网络,由神经网络给出的增加因子可提高任务分配的收益,能够

实现多对多任务分配时的收益最大化目标。

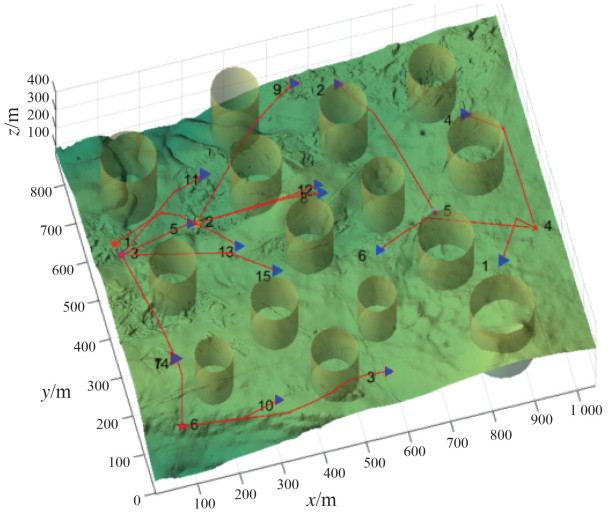

图 6　对比组 2 实验结果

Fig. 6　Experimental results of comparison group 2

表 4　各算法实验结果收益统计及对比

Table 4　Statistics and comparison of experimental results of each algorithm

Algorithm	PSO	HFPSO	CLPSO	EPSO	GGL-PSOD	Traditional auction algorithm	BDQNBA	Average promotion rate of BDQNBA compared with traditional auction algorithm
Experimental group	586	598	621	634	667	577	701	21.4%
comparison group 1～30 (average)	564.1	574.6	584.9	597.6	613.7	555.6	667.3	20.1%(±5%)
comparison group 2	346	357	362	387	391	356	413	22.9%

5　总结

本文面向多无人机监测目标分配任务，提出了一种融合深度 Q 网络(Deep Q-network，DQN)出价策略的自主进化拍卖算法。该方法首先构建了强化学习通过多任务分配模型，引入增价因子和奖励权重因子，随后采用策略训练的方式自适应调节增价因子的大小，实现了多目标分配问题收益上的最大化。相比与之前的单目标拍卖算法，本文方法在多目标分配方面具有收益更高的优势。本文的 DQN 拍卖策略在多对、多目标分配问题中，以最大化任务分配收益为优化指标，并且分析了训练过程中奖励构成对于结果的影响，通过离线样本训练的方式，获得了明显优于常规拍卖法的目标分配方案。通过多无人机多监测目标分配的仿真验证，本文方法在分配收益方面平均提升了 21.4%，同时，本文方法在解决可变数量及可变位置无人机与监测目标方法中，均能够实现收益提升 20%以上。

综述所述，本文提出的拍卖算法，能够有效提高拍卖算法在多对、多任务分配场景下的表现，实验证明，本算法在可变无人机、任务点数量和位置的情况下相对于传统拍卖算法收益均可获得提升。研究表明，通过调整强化学习的环境参数，优化分配方案，该算法可以适用于更为复杂的任务分配场景。

参 考 文 献

[1] 周文惠,齐瑞云. 面向突发故障的分布式多无人机任务重规划方法[J]. 控制与决策,2023, 38(5): 1374-1385.

[2] 黄亭飞,程光权,黄魁华. 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.

[3] 顾佼佼,周曰建,付鹏飞. 基于改进组合拍卖算法的分布式空战监测决策[J]. 兵工自动化, 2019, 38(5): 67-69+96.

[4] 刘学达,何明,禹明刚,等. 基于公共物品博弈的无人机集群弹药分配方法[J]. 控制与决策, 2022, 37(10): 2696-2704.

[5] YU Z Q, ZHANG Y M, JIANG B, et al. Fractional-order adaptive fault-tolerant synchronization tracking control of networked fixed-wing UAVs against actuator-sensor faults via intelligent learning mechanism[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(12): 5539-5553.

[6] 陈昕叶. 动态环境下多机器人协同搜救任务分配方法研究[D]. 广州:华南理工大学, 2020.

[7] 王轩. 多无人机任务分配与航迹规划算法研究[D]. 西安:西安电子科技大学, 2020.

[8] ZHANG Y , ZHANG Z , YANG Q, et al. EV charging bidding by Multi-DQN reinforcement learning in electricity auction market[J]. Neurocomputing, 2020, 397:1-6.

[9] BAUMANN D, ZHU J J, MARTIUS G, et al. Deep reinforcement learning for event-triggered control [J]. IEEE, 2018(1):943-950.

[10] DIAS M B, ZLOT R, KALRA N,et al. Market-based multirobot coordination: a survey and analysis [J]. Proceedings of the IEEE, 2006, 94(7):1257-1270.

[11] 于晓强,郑红星. 基于拓展CBBA算法的在轨装配航天器任务分配技术研究[J]. 无人系统技术, 2019, 2(4): 46-53.

[12] 王世磊,屈绍建,常广庶,等. 不确定情形下在线多源多属性反向拍卖双边协商决[J]. 控制与决策, 2022, 37(11): 3023-3032.

[13] 刘爱珍,贾红丽,王嘉祯,等. 基于组合拍卖机制的移动Agent投标策[J]. 计算机工程, 2009, 35(8): 28-30.

[14] 杜辉,李卓,陈昕. 基于在线双边拍卖的分层联邦学习激励机制[J]. 计算机科学, 2022, 49(3): 23-30.

[15] 蔺文轩,谢文俊,张鹏,等. 无人机集群作战控制与规划研究综述[C]. //2022年无人系统高峰论坛(USS2022)论文集,2022.

[16] TENG Y, YONG Z, FANG N, et al. Reinforcement learning based auction algorithm for dynamic spectrum access in cognitive radio networks [C]. //IEEE 72nd Vehicular Technology Conference, 2010:1-5.

[17] 张子迎,陈云飞,王宇华,等. 基于启发式深度Q学习的多机器人任务分配算法[J]. 哈尔滨工程大学学报, 2022,43(6):857-864.

[18] SEENU, N. CHETTY, KUPPAN R. M. RAMYA, et al. Review on state-of-the-art dynamic task allocation strategies for multiple-robot systems [J]. Industrial Robot, 2020, 47(6): 929-942.

[19] 赵鹏程. 基于多智能体深度强化学习的空间众包任务分配研究[D]. 长春:吉林大学, 2021.

[20] 高程,都延丽,步雨浓,等.基于顺序扩展一致性包算法的多无人机分布式任务分配[J]. 控制与决策, 2022(6):1-9.

[21] 殷昌盛,杨若鹏,朱巍,等. 多智能体分层强化学习综述[J]. 智能系统学报, 2020, 15(4): 646-655.

[22] CLERC M, KENNEDY J. The particle swarm-explosion, stability, and convergence in a multidimensional complex space[J]. IEEE Transactions on Evolutionary Computation, 2002, 6(1): 58-73.

[23] AYDILEK I B. A hybrid firefly and particle swarm optimization algorithm for computationally expensive numerical problems[J]. Applied Soft Computing, 2018, 66: 232-249.

[24] LIANG J J, QIN A K, SUGANTHAN P N, et al. Comprehensive learning particle swarm optimizer for global optimization of multimodal functions[J]. IEEE Transactions on Evolutionary Computation, 2006, 10(3): 281-295.

[25] LYNN N, SUGANTHAN P N. Ensemble particle swarm optimizer[J]. Applied Soft Computing, 2017, 55: 533-548.

[26] LIN A, SUN W, YU H, et al. Global genetic learning particle swarm optimization with diversity enhancement by ring topology[J]. Swarm and Evolutionary Computation, 2019, 44: 571-583.

(责任编辑:刘庆松)