

基于类脑决策模型的多无人机任务分配

杨明¹, 刘蓉^{1,2}, 王佑¹, 闫俊辰¹

(1. 南京航空航天大学 自动化学院, 南京 210016;

2. 南京航空航天大学 无人机研究院, 南京 210016)

摘要: 针对多无人机同时执行多个任务的执行时序不同和无人机任务完成效果以及无人机能耗效率低等问题, 建立经验判断—任务预分配—任务重分配—在线决策的类脑决策模型, 提出一种基于类脑决策模型的多无人机任务分配算法; 融合调度改进组合聚类算法和混合强化学习等策略, 将任务分配分为预分配和重分配两个阶段, 生成决策网络和评价体系, 并引入孤立点离散化策略进行在线决策, 旨在解决孤立点的影响, 减少状态空间和动作空间的复杂性, 提高多无人机系统的分配效率; 最后对算法进行了仿真验证。

关键词: 任务分配; 类脑决策模型; 聚类算法; 强化学习; 离散化策略

Multi Drone Task Assignment based on Brain-like Decision Model

YANG Ming¹, LIU Rong^{1,2}, WANG You¹, YAN Junchen¹

(1. College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China;

2. Research Institute of Pilotless Aircraft, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract: Aiming at the problems of different execution timings for multiple drones simultaneously executing multiple tasks, as well as the low efficiency of drone task completion and energy consumption, a brain-like decision model of experience judgment-task pre-assignment-task reassignment-online decision is established, and a task assignment algorithm based on brain-like decision model is proposed. Fusion scheduling improves strategies such as combinational clustering algorithm and hybrid reinforcement learning, divides task assignment into two stages: pre-assignment and reassignment, generates decision network and evaluation system, and introduces isolated point discretization strategy for online decision-making, aiming to solve the influence of isolated points, reduce the complexity of state space and action space, and improve the allocation efficiency of multi-unmanned aircraft systems. Finally, the algorithm is verified by simulation.

Keywords: task assignment; brain-like decision model; clustering algorithm; reinforcement learning; discretization strategy

0 引言

多无人机 (UAVs, unmanned aerial vehicles) 任务分配是一类多约束条件下的优化问题, 涉及到任务规模、资源分配、协同控制等多个方面, 这些问题的解决直接关系到任务完成的效率和系统的可靠性^[1]。

目前, 针对多无人机任务分配的研究主要分为集中式和分布式任务分配两种^[2]。由于任务的复杂性和资源的有限性, 现有的多无人机任务规划方法仍存在一些局限性。例如, 在大规模任务和动态环境下, 整数规划算法 (IP, integer programming)^[3], 多旅行商算法 (MTSP, multi traveling salesman problem)^[4], 动态网络流优化算法 (DNFO, dynamic network flow optimization)^[5]等集中式算法的运行时间急剧增加, 使得实时性受到限制。而遗传算法 (GA, genetic algorithm)^[6], 粒

子群算法 (PSO, particle swarm optimization)^[7]等启发式方法可能忽略了能量消耗等关键资源的优化分配。然而, 随着任务场景复杂性的不断提升, 多无人机任务分配问题变得更加具有挑战性, 需要有效的分配和决策方法来实现高效、智能的任务协同执行。

强化学习算法 (RLA, reinforcement learning algorithm)^[8]作为一种能够自主学习优化策略的方法, 其核心思想是通过智能体与环境的交互, 使智能体能够学会采取一系列行动来最大化累积奖励。然而, 在多无人机任务分配中, 强化学习面临的挑战之一是状态空间和动作空间的巨大增加, 导致传统的强化学习算法难以应对。此外, 无人机之间的协同作战也增加了任务规划的复杂性。

为了应对上述问题, 聚类算法 (CA, clustering al-

收稿日期: 2024-01-25; 修回日期: 2024-03-11。

基金项目: JWK 创新项目 (19-163-15-ZT-008-002-06); 江苏省重点研发计划 (BE2022068-4); LZ 十四五预研项目 (LZY202303052)。

作者简介: 杨明 (1998-), 男, 硕士研究生。

引用格式: 杨明, 刘蓉, 王佑, 等. 基于类脑决策模型的多无人机任务分配[J]. 计算机测量与控制, 2025, 33(3): 275-286, 322.

gorithm)^[9]可以将相似的状态或任务分组,从而在强化学习中减少任务处理数据规模,可以在保持任务分配灵活性的同时,降低计算复杂性,提高系统的整体性能,使智能体能够更加高效地学习和决策。

近年来,研究人员就开始尝试将神经元模型应用于计算机科学领域,剔除“人”在无人机系统环路中的观察力、敏感度有限等生理影响,保留人脑的记忆、学习和改进等作用,就需要引入新的智能学习体来进行多无人机的总体规划 and 决策^[10]。研究人员从脑科学中受到启发,结合数学建模、模型优化等理论知识,积极开展基于类脑智能算法的多无人机的研发工作^[11]。基于类脑智能算法的多无人机系统的高度自主性体现在其自身的自主学习能力^[12]。类脑智能算法和多无人机系统可构建“机制类脑,行为似人”的仿生智能无人系统,其具备高度自主性,可持续学习性和任务规划决策快速性,可有效解决“人在环路”的问题,大幅度提高无人机的自主学习,智能决策等功能^[13]。因此,围绕多无人机执行任务过程中所涉及的规划和决策问题,开展多无人机系统类脑智能的相关理论研究工作具有十分重要的现实意义,

本文旨在将聚类算法与强化学习相结合,建立类脑决策模型,基于类脑决策模型提出一种融合改进强化学习优化算法(FIRLA, fusion improved reinforcement learning algorithm),以减少状态空间和动作空间的复杂性,并提高多无人机的分配效率。

1 多无人机任务分配设计

1.1 问题分析

在执行任务分配时,无人机需要在最短的时间完成各自分配的任务,此时需要考虑无人机的飞行速度、最大航程和完成任务概率等相关影响因素。

1) 无人机模型:

本文将多无人机任务分配问题建模为类脑智能算法的任务分配模型,其中 UAVs 集合表示为 $U = \{U_1, U_2, \dots, U_n\}$, n 表示 UAV 的数量, U_n 表示为 $U_n = (x_t, y_t, v_t, s_t)$ 。 x_t, y_t 表示无人机在 t 时刻的位置, v_t 表示无人机在 t 时刻的速度, s_t 表示无人机在 t 时刻已分配的任务状态矩阵,如公式(1)所示:

$$s_t = \begin{cases} 0, & \text{未分配} \\ 1, & \text{已分配} \end{cases} \quad (1)$$

2) 任务模型:

假设在某二维已知区域 Ω 中发现了 N_t 个任务目标,则任务集合 T 可表示为 $T = \{T_1, T_2, \dots, T_{N_t}\}$ 。其中 $T_{N_t} = (x_t, y_t, s'_t)$, 其中 x_t, y_t 表示任务目标在时刻 t 的位置, s'_t 表示任务在 t 时刻的分配状态矩阵,如公式(2)所示:

$$s'_t = \begin{cases} 0, & \text{未分配} \\ 1, & \text{已分配} \end{cases} \quad (2)$$

3) 任务聚类模型:

在任务进行分配前,先进行任务聚类,在上述已知区域 Ω 内选取 k 个中心位置 $V_i = (i=1, 2, \dots, k)$, 使 N_t 个点到各自最近的中心位置的距离平方和最小,则 k -means 聚类算法的目标函数定义如公式(3)所示:

$$E = \sum_i^l \quad (3)$$

式中, E 是所有目标点与其所属的簇中心之间的偏差的总和, X 是 Ω 中给定的目标数据点; C_i 是 k -means 算法的簇, V_i 是簇 C_i 的平均值; $\|X - V_i\|_p$ 表示 X 和 X_i 之间的一种 p 阶度量,一般在距离空间多采用欧氏距离的平方 ($p=2$) 来度量。

4) 任务分配模型:

多无人机的任务分配模型如公式(4)、(5)所示,多无人机 U_n 的任务分配集合 φ_n 用表示,定义分配变量 φ_{nk} 表示分配的任务 M_k :

$$\Phi_{nk} = \begin{cases} 1, & U_n \text{ 执行 } M_k \\ 0, & \text{其他} \end{cases} \quad (4)$$

$$\varphi_n = \{M_k \mid M_k \in I\} \quad (5)$$

1.2 代价指标分析

多无人机任务分配是一个涉及多约束条件的策略优化问题^[14]。约束条件和优化指标在解决这类问题中具有至关重要的作用。为了实现任务的全覆盖以及无人机整体效益的最大化,本文将从航程和时间代价、机损代价以及收益代价 3 个方面进行策略优化。

1) 航程和时间代价:

本文将参与执行任务的无人机与其被分配的任务之间的直线距离视为单个航程,则总航程代价为所有参与执行任务的无人机飞行距离之和,其数学表达式为:

$$f_L = \sum_{i=1}^n \sum_{k \in N_t} \quad (6)$$

式中, f_L 表示多无人机的总飞行距离,当 $k=0$ 时,表示第 i 架无人机从出发点第一个任务的飞行距离;当 $k \neq 0$ 时, $Dis(N_t^i N_{i+1}^i)$ 表示第 i 架无人机执行第 k 个任务到第 $k+1$ 个任务的飞行距离。

完成任务的时间受无人机系统中单个无人机完成时间的影响,多无人机系统的时间代价数学表达式为:

$$f_T = \max(1 \leq k \leq N_t) t_k^{\text{evaluate}} \quad (7)$$

式中, t_k^{evaluate} 表示单个无人机执行任务的总时间, f_T 表示无人机系统中最后一架无人机任务执行完成的时间。

2) 机损代价:

机损代价反映了执行任务时无人机群整体的受威胁程度,受威胁程度越低则无人机群整体机损概率越低。设定第 i 架无人机 U_i 执行第 k 个任务 N_k 的机损概率为 h_{ik} , 则无人机群的整体机损代价的数学表达式:

$$f_A = \sum_{i=1}^{U_{N_t}} \sum_{k=1}^{N_t} \Phi_{ik} h_{ik} \quad (8)$$

3) 收益代价:

收益代价是无人机群执行完所有任务后的整体收益

的映射, 收益代价越大, 无人机群整体获得的收益越多。设定第 k 个任务 N_k 的价值为 V_k , 第 i 架无人机 U_i 执行第 k 个任务 N_k 时的杀伤概率为 p_k , 则集群整体收益代价的数学表达式为:

$$f_v = \sum_{i=1}^{N_s} \sum \quad (9)$$

1.3 约束条件分析

在制定优化策略时, 还需充分考虑无人机所面临的约束条件, 提高无人机的整体性能和作战效果。这些约束包括无人机的最大航程限制、载荷限制以及任务完成度要求等。

1) 最大航程约束:

无人机的最大航程限定了其可执行任务的数量, 因此, 在进行任务分配时, 必须确保无人机的航程不会超过其最大限制, 其数学表达式为:

$$f_{L_i} \leq f_{L_i}^{\max} (\forall U_i \in U) \quad (10)$$

式中, f_{L_i} 表示第 i 架无人机执行完所有任务的飞行距离, $f_{L_i}^{\max}$ 表示第 i 架无人机的最大航程。

2) 载荷约束:

载荷能力是衡量无人机打击能力的重要指标, 它直接影响到无人机的任务执行数量。因此, 在进行任务分配时, 必须根据无人机的载荷能力合理规划任务, 以确保其打击能力和任务完成的有效性。单架无人机的载荷能力的数学表达式为:

$$C_{U_i} \leq C_{U_i}^{\max} (\forall U_i \in U) \quad (11)$$

式中, C_{U_i} 表示第 i 架无人机的载荷量, $C_{U_i}^{\max}$ 表示第 i 架无人机的最大载荷量。

3) 任务完成度约束:

任务分配的关键是能否完成所有任务的分配。为了确保任务的完整性和避免重复执行, 每个任务只应被分配并执行一次, 则任务分配决策变量具有以下约束条件:

①一个任务只能由一架无人机执行:

$$\varphi_n^{N_i} = \begin{cases} 1, & \text{无人机 } n \text{ 执} \\ 0, & \text{其他} \end{cases} \quad (12)$$

式中, n 表示无人机编号, N_i 表示任务编号, 当第 N_i 个任务目标被分配后, $\varphi_n^{N_i}$ 将会置为 0, 不再参与任务分配。

②同一时刻, 任意一架无人机只能执行一个任务:

$$\Phi_t^n = \begin{cases} 1, & \forall U_n \in U \\ 0, & \text{其他} \end{cases} \quad (13)$$

式中, t 表示时间, n 表示无人机编号, φ_n^t 置为 0 时表示无人机不在执行任务。在任务执行到 t 时刻时, 任意无人机 U_n 此时都只执行一个任务或不执行。

③任务区域 Ω 所有的任务 T_{N_i} 必选全部被分配且执行:

$$\sum_{i=1}^n \sum_{j=1}^{N_i} \Phi_{ij} = N_i \quad (14)$$

1.4 基本聚类算法

聚类算法是一种无监督学习方法, 用于将数据集中

的数据分成不同的聚类或组。聚类算法主要分为层次聚类和划分聚类, 本章的任务分配主要使用划分聚类, 经典的划分聚类算法有 K-means 聚类算法和 K-中心点轮换聚类算法。

设在 2 维任务空间 Ω 中有 n 个任务构成的任务集合 $S = \{X_1, X_2, \dots, X_n\}$, 其中 $X_i = (x_i, y_i)$ ($i=1, 2, \dots, n$), 在任务空间 Ω 的范围内选取 k 个中心点 $V_j = (j=1, 2, \dots, k)$, 中心点 V_j 所属的簇为 C_j , 则聚类算法的目标是使 n 个任务点到各自最近的中心点的距离之和最小, K-means 聚类算法目标函数定义如下:

$$E_j = \sum_{j=1}^k \sum_{K \in C_j} I_{ij} [(x_i - C_{j1})^2 + (y_i - C_{j2})^2] \quad (15)$$

式中, E_i 是所有目标点与其所属的簇中心之间的偏差的总和, 属于聚类内部方差, I_{ij} 表示属于簇 C_j 中的任务点。K-中心点轮换聚类算法的具体步骤如下。

步骤 1: 随机选择任务空间中的 K 个点作为初始的簇中心;

步骤 2: 计算每个任务到 K 个簇中心的距离;

步骤 3: 按照计算的距离, 将每个任务分配给距离最近的簇, 这是基于最小化每个任务到其对应簇中心的距离来实现的;

步骤 4: 对上述分配完的任务再进行每个任务类别求聚类间的方差, 求解出新的簇中心。

步骤 5: 比较当前计算出的簇中心和上一次迭代的簇中心, 如果簇中心发生变化, 则返回步骤 2, 否则, 进入下一步;

步骤 6: 当所有的簇中心都不再发生变化时, 算法收敛, 输出最终的簇划分和聚类中心。

定义聚类间的方差为:

$$E_0 = \sum_{x \in C_j} \sum \quad (16)$$

式中, $\bar{C}_j$ 表示第 j 个聚类中心点坐标的平均值, K-中心点轮换聚类算法的目标函数表示为:

$$E_{p_n} = E_t + \lambda \cdot E_0 \quad (17)$$

式中, E_{p_n} 是所有目标点与其所属的簇中心之间的最小偏差的总和, λ 是一个正则化参数, 用于平衡聚类内部方差和聚类间方差之间的权重, 通过调整 λ 值, 控制 K-中心点轮换法对不同类型误差的敏感程度。

1.5 基本强化学习算法

强化学习算法通过智能体与环境的交互, 使智能体能够学会采取一系列行动来最大化累积奖励^[15]。

图 1 是智能体与环境的交互示意图, 在 t 时刻, 智能体根据当前的环境状态 s_t 来选择动作 a_t , 执行动作后, 环境得到下一个状态 s_{t+1} , 并反馈给智能体一个奖励信号 r_{t+1} 。在 $t+1$ 时刻, 智能体根据状态 s_{t+1} 选择动作 a_{t+1} , 得到环境状态 s_{t+2} 和奖励信号 r_{t+2} 。则周期循环结束时, 智能体会收集到一个周期的学习样本为:

$$\{s_0, a_0, r_1, s_1, a_1, r_2, \dots, s_{T-1}, a_{T-1}, r_T, s_T\} \quad (18)$$

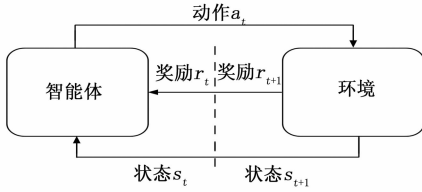


图 1 智能体与环境交互示意图

定义 t 时刻的“回报” R_t 为从该时刻开始到周期结束时的“回报”奖励之和，表示为：

$$R_t = r_t + \epsilon r_{t+1} \quad (19)$$

式中， ϵ 表示折扣因子，可以看出，该式对奖励进行加权，权重随着时间步以 ϵ 的速度衰减，时间步距离越远，权重越小。

使用策略 π 从 (s_t, a_t) 开始进行环境交互直到周期结束，得出完整的交互序列，进而计算 (s_t, a_t) 对应的 R_t 。 (s, a) 在策略 π 下的值函数 $Q^\pi(s, a)$ 衡量了从该状态出发按照策略 π 选择的动作在周期结束时获得的奖励平均值，表示为：

$$Q^\pi(s, a) = E_\pi[R_t | S_t = s, a_t = a] =$$

$$E_\pi\left[\sum_{k=0}^{\infty} \epsilon^k r_{t+k+1} | s_t = s, a_t = a\right] \quad (20)$$

式中， $Q^\pi(s, a)$ 是动作价值函数，用来衡量智能体采取特定动作的好坏，衡量智能体所处状态的好坏则用状态价值函数表示，起相当于对 $Q^\pi(s, a)$ 在动作上取期望，表示为：

$$V^\pi(s) = E_\pi[Q(s, a)] = E_\pi[R_t | s_t = s] =$$

$$E_\pi\left[\sum_{k=0}^{\infty} \epsilon^k r_{t+k+1} | s_t = s\right] \quad (21)$$

2 基于类脑决策模型的多无人机任务分配

人脑是一个通用的智能系统，主要通过亿万神经元和突触连接构成，大脑分为大脑皮层、小脑、脑干、边缘系统、基底节和丘脑等主要区域^[16]。根据人脑的不同区域的协同工作原理，大脑中存在许多调节系统，如图 2 所示，大脑皮层皮质、边缘系统、基底节和丘脑形成的调节系统中基底节将接受到的信号进行整合传递给丘脑，丘脑会将信息处理后回传给皮质皮层，皮质皮层根据不同信号进行决策和记忆，再分别传输至各区域，该系统具有数据记忆、奖惩调节和学习决策等功能，主要进行信息整合、规划和决策。

从人脑决策的思维方式和不同区域协调机制来看，强化学习是设计类脑智能算法的优选方法。因此，本文仿照人脑决策的思考方式，以 Q-Learning 强化学习算法为基础，考虑到多无人机任务分配和自主决策的任务需求，建立经验判断—任务预分配—任务重分配—在线决策的类脑决策模型，总体技术原理如图 3 所示。

针对经验判断环节建立记忆库，存储每次迭代的最

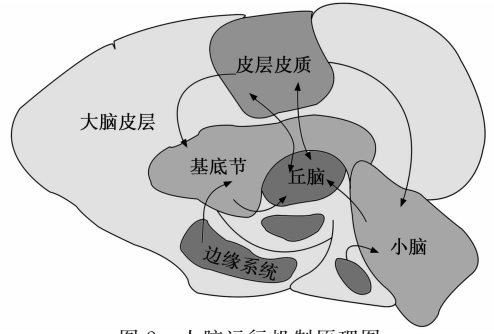


图 2 人脑运行机制原理图

优任务分配策略；任务预分配环节建立基于混合策略的聚类算法任务预分配模型，对多任务进行初步分配；任务重分配环节通过改进强化学习算法对预分配集进行任务重分配，确保任务完全分配，并使用附有权重系数的任务再分配策略进行孤立点离散化再分配，使任务分配更集中；在线决策环节，由于重分配可能受孤立点影响，根据改进优化后的评价体系和混合任务分配策略生成离线最优决策策略并将输出的最优任务分配集保存到记忆库中。

2.1 经验判断

2.1.1 建立经验数据库

经验库主要存储任务环境参数 evn_r 和无人机初始参数 UAV_T ，完成任务所需的载荷量 C_U 和无人机数量 U_n ，根据任务需求进行任务分配所得整体收益值 R 、最优值函数 Q^* 和最优策略 π^* ，最后保存输出的最优任务分配集合 Φ_n 、总航程 f_L 和机损代价 f_A 。根据这些参数建立经验库 M ，经验库的统一格式表示为：

$$\left[(evn_T, UAV_T, \Phi_n, 0, f_{L0}, f_{A0}, R, Q^*, \pi^*), (evn_{T_{nd}}, UAV_{T_{nd}}, \Phi_n, f_L, f_A) \right] \quad (22)$$

式中， Φ_{n0} 初始任务分配集和， f_{L0} 为初始无人机群总航程， f_{A0} 为初始无人机群总机损代价， $evn_{T_{nd}}$ 为任务分配结束后的任务空间参数， $UAV_{T_{nd}}$ 为任务分配后的无人机参数。

2.1.2 初始化任务分配集

根据类脑模型中记忆库的经验数据进行任务分配方案初始化判断。

算法 1：初始化任务分配方案

输入：数据库中的任务环境和无人机参数、任务分配集和任务分配收益

输出：初始任务分配集

1) 对于任务环境和无人机参数；

2) 判断与数据库里的参数是否一致

若参数不一致：

初始化任务分配集为空

若参数一致：

使用数据库里的分配集作为初始任务分配集

3) 输出初始任务分配集。

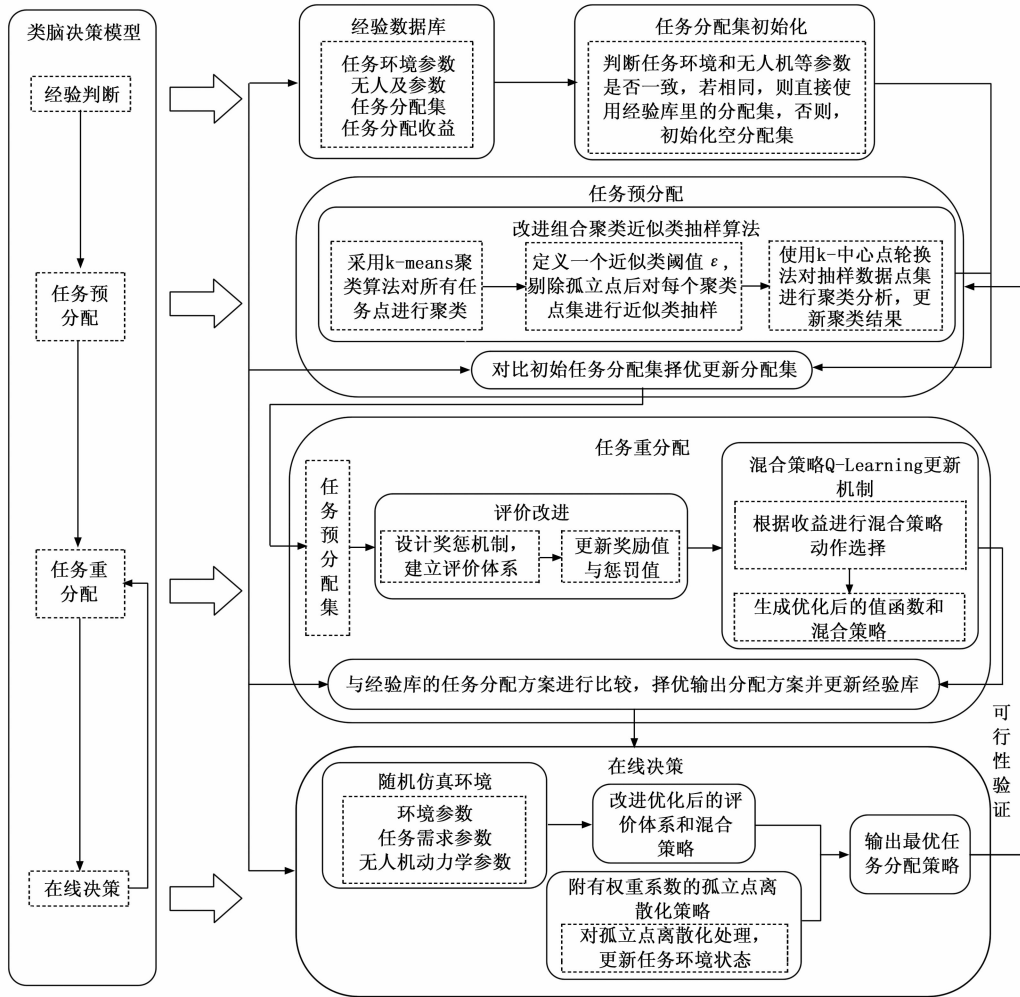


图3 基于类脑决策模型的任务分配技术原理图

2.2 基于改进近似类抽样的组合聚类算法的任务预分配策略

近似关系 X 是论域 Ψ 中的二元关系,它满足自反性和对称性^[17]。设论域 Ψ 中有近似关系 X ,则论域 Ψ 中元素的近似类记为 $[a]_X$,其数学表达式为:

$$[a]_X = \{x | (x \in \Psi) \wedge (xXa)\} \quad (23)$$

式中, a 称为 $[a]_X$ 的核心元素, xXa 表示元素 x 和 a 具有近似关系 X ,即 $(x, a) \in X$ 。在论域 Ψ 中,给定一个阈值 $\vartheta > 0$,则对每一个元素 a 的近邻类记为 $[a]_X$,其数学表达式为:

$$[a]_X = \{x | (x \in \Psi) \wedge (\|x - a\| < \vartheta)\} \quad (24)$$

设定 K-中心点轮换聚类任务空间中 n 个 m 维数据点抽样的时间复杂度为 $O(kmn^2)$,其中 k 为聚类的中心点数量,由于时间复杂度较高,不能直接应用到大批量数据的聚类中,因此为克服近似类抽样聚类算法的难点,本文提出一种基于 K-means 聚类算法和 K-中心点轮换的近似类抽样组合聚类算法。该算法的核心思想是:通过在 k-means 聚类算法中选择代表性样本来进行近似抽样,然后使用 k-中心点轮换法改进聚类效果。在每次迭代中,从每个簇中选择一个代表性样本,然后使用这些样

本近似抽样来更新聚类中心,这样可以有效地减少计算量,并提高聚类质量。将该近似抽样的聚类结果输出为多无人机任务分配的预分配集。具体分为 3 个步骤。

步骤 1:针对任务空间内的任务点,采用 k-means 聚类算法对所有任务点进行聚类,获得初步聚类点集,其时间复杂度为 $O(kmn^2)$ 。

步骤 2:根据阶段一的初步聚类点集,定义一个近似类阈值 ϵ ,剔除孤立点后对每个聚类点集进行近似类抽样,将任务空间内剔除后的任务点全部分配并避免重复分配到近似类中,按照近似类中任务数量对近似类簇进行降序排序,将近似类簇中的核心元素组成抽样任务点集,其时间复杂度为:

$$O\left(km \frac{n^2}{k}\right) = O\left(\frac{mn^2}{k}\right) \quad (25)$$

步骤 3:使用 k-中心点轮换法对抽样数据点集进行聚类分析,获得的结果作为本文改进组合聚类近似类抽样算法的结果。

2.3 基于混合策略改进强化学习的任务重分配策略

根据上一环节的预分配集和奖惩机制,添加评价函数,对任务空间内的所有任务进行重分配,该环节主要

使用改进强化学习算法进行任务重分配。Q-Learning 强化学习算法通过学习最优的 Q 函数来输出相对于 Q 值的 ϵ -greedy 策略, 其核心是评价表 Q 表和贪婪策略, Q 表用来估计当前状态下不同动作获得的累计最大收益的均值, 贪婪策略则是根据最大 Q 值选取对应的动作。然而, 随着任务数据量的增多, Q-Learning 算法处理大规模问题时, 其 Q 表的存储和更新会面临较大的挑战^[18]。因此, 本节的评价改进环节以 Q-Learning 算法为基础, 设计奖惩机制, 结合 Softmax 策略^[19]改进 ϵ -greedy 策略, 提出一种混合策略改进的强化学习算法 (FIRLA, fusion improved reinforcement learning algorithm), 进行多无人机任务重分配, 具体流程如图 4 所示。

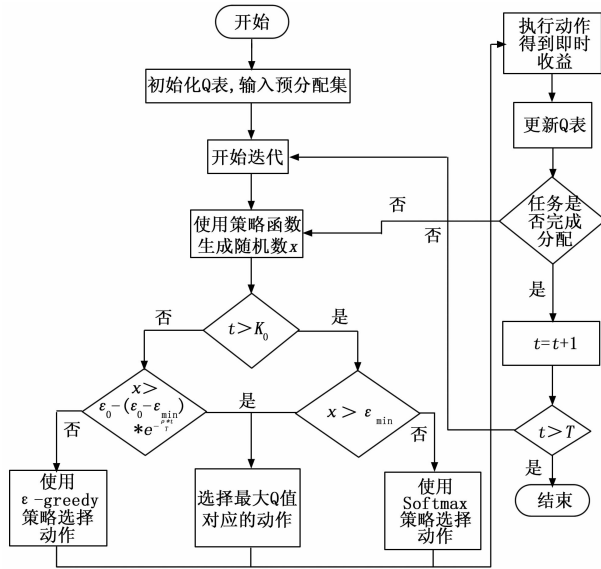


图 4 多无人机任务重分配流程图

2.3.1 奖惩函数设计

为获得最优任务分配策略, 强化学习算法在任务分配时需考虑最大适应度和最小代价原则建立奖惩机制, 确定单架无人机任务分配的收益值。因此根据式 (6) ~ (9) 可确定多无人机任务分配的总收益主要包括无人机的航程、航时、机损和单机收益, 定义任务分配的收益函数为:

$$R = \max(\sum_{i=1}^{N_u} (k_1 * f_v - k_2 * f_L - k_3 * f_T - k_4 * f_A)) \quad (26)$$

式中, R 为多无人机任务分配的总收益, k_1, k_2, k_3, k_4 表示各项指标权重, 并满足 $k_1 + k_2 = k_3 + k_4 = 1$ 。根据式 (10) 确定无人机航程的惩罚函数为:

$$C_1 = \begin{cases} 0 & f_L \leq f_L^{\max} \\ Cl & f_L > f_L^{\max} \end{cases} \quad (27)$$

式中, Cl 为超过航程的惩罚值。

根据式 (11) 确定无人机载荷的惩罚函数为:

$$C_2 = \begin{cases} 0 & C_{U_i} \leq C_{U_i}^{\max} \\ C_n & C_{U_i} > C_{U_i}^{\max} \end{cases} \quad (28)$$

式中, C_n 为超过载荷的惩罚值。

根据式 (12) ~ (14) 确定无人机任务完成度的惩罚函数为:

$$C_3 = \begin{cases} Cm & \Phi_n^N > 1 \& \Phi_t^n > 1 \& \sum_{i=1}^n \sum_{j=1}^{N_i} \Phi_{ij} \neq N_i \\ 0 & \text{其它} \end{cases} \quad (29)$$

式中, Cm 为超过任务完成度的惩罚值。因此, 惩罚机制的数学表达式为:

$$r = C_1 + C_2 + C_3 \quad (30)$$

式中, 仅当 $r=0$ 时表示无人机执行分配的任务时满足了上述所有约束, 此时输出的任务分配方案是可行的。

2.3.2 探索因子动态衰减策略

ϵ -greedy 策略是一种探索性策略, 通过以一定的概率 ϵ 选择随机动作, 以一定的概率 $1-\epsilon$ 选择当前最有价值的动作, 从而实现在探索和利用之间的平衡。Softmax 策略则是一种确定性策略, 通过将每个动作的价值转换为概率分布, 可以更加稳定和平滑地选择动作, 但在训练过程中可能会导致过早收敛或难以找到最优策略^[20]。因此本节针对探索率 ϵ 的衰减进行动态衰减梯度策略改进, 保证强化学习的初始阶段具有较高的 ϵ 值, 然后逐渐减小 ϵ 值, 增加 Softmax 策略的权重, 更好地利用已有的经验, 使智能体对任务环境的探索和利用逐渐达到平衡, 找到最优的任务分配策略, 探索因子 ϵ 动态衰减的数学表达式为:

$$\epsilon = \begin{cases} \epsilon_0 - (\epsilon_0 - \epsilon_{\min}) \cdot e^{-t/T} & t \leq K_0 \\ \epsilon_{\min} & t > K_0 \end{cases} \quad (31)$$

式中, ϵ_0 为初始探索率, $\epsilon_{\min}$ 为最小探索率, T 为迭代总次数, t 表示当前迭代次数, K_0 表示预设迭代次数阈值。

2.3.3 改进混合策略 Q-Learning 更新机制

假设任务空间 Ω 中, 每架无人机 U_n 当前时刻有一个状态 S_t 和一个价值函数 $V_n(S_t)$, 每个任务 T_{N_i} 的价值函数为 $V_{N_i}(T_{N_i})$, 在每一个迭代步中, 无人机 U_n 以概率 ϵ 随机选择一个动作 a_t , 或者以概率 $1-\epsilon$ 选择价值最大的动作 a_t^* , 所有动作的集合为 A , 则值函数 Q 的更新公式为:

$$Q(a_t | S_t) = Q(S_t | a_t) + \alpha[r_{t+1} + \gamma \cdot \max_{A \cdot S_t} Q(A \cdot S_t \cdot \omega + S_{t+1}^{\text{rest}} | a_{t+1}) - Q(S_t | a_t)] \quad (32)$$

式中, α 为学习率, γ 为折扣率, $t+1$ 表示在执行动作 a_t 后获得的奖励, ϵ -greedy 策略选择动作的表达式如下:

$$\pi_{\epsilon}(a_t | S_t) = \begin{cases} \frac{\epsilon}{|A|} + 1 - \epsilon & a_t = \operatorname{argmax}_a Q(S_t | a_t) \\ \frac{\epsilon}{|A|} & \text{其它} \end{cases} \quad (33)$$

Softmax 策略函数选择动作 a_t 的表达式如下:

$$\pi_s(a_t | S_t) = \frac{e^{Q(S_t, a_t)/T}}{\sum_A e^{Q(S_t, A)/T}} \quad (34)$$

式中, T 为强化学习的周期, 用于调节智能体对任务环境探索和利用。则混合策略动作选择函数表达式为:

$$\pi_{\text{mix}}(a_t | S_t) = \lambda \cdot \pi_s(a_t | S_t) + (1 - \lambda) \pi_g(a_t | S_t) \quad (35)$$

式中, λ 是混合策略的参数, $\lambda \in [0, 1]$, 其控制着 ϵ -greedy 策略和 Softmax 策略在选择动作时的权重。综上, 可以得出无人机在状态 S_t 下的最优动作价值函数 $Q^*(a_t | S_t)$ 的表达式为:

$$Q^*(a_t | S_t) = \max_{\pi} Q^{\pi}(a_t | S_t) \quad (36)$$

最优策略 π^* 的表达式为:

$$\pi^* = \arg Q^*(a_t | S_t) \quad (37)$$

2.4 附有权重系数的孤立点离散化策略的在线决策

多无人机协同任务分配问题和在线决策问题是相互耦合的, 由图 3 可知, 任务分配基于在线决策优化得出的混合策略进行最优分配, 为在线决策提供任务需求关系, 在线决策则是在任务分配的基础上选择最优分配动作集, 生成任务分配的期望方案。在通过上述经验判断、任务预分配、任务重分配环节后, 生成最优任务分配策略 π^* 、评价函数和混合 Q-Learning 策略构成了基于类脑模型的智能任务离线决策方法。

本文将任务环境中突然出现的任务点、离散度较大的任务点和运动的任务点称为孤立点。由于任务重分配环节是在预分配的基础上进行任务重分配, 而聚类算法进行预分配时难免会受到孤立点的影响, 从而影响重分配结果, 导致重分配可能陷入局部最优无法完成多无人机任务的最优分配, 故在在线决策环节设计孤立点离散化策略, 剔除孤立点的影响, 更新离线决策策略, 使无人机任务分配达到全局最优。

对此本文设计附有权重系数的孤立点交叉更新机制^[21]对离散任务点和任务分配集进行交叉更新, 将原来的任务环境状态进行更新再分配, 其原理如图 5 所示。

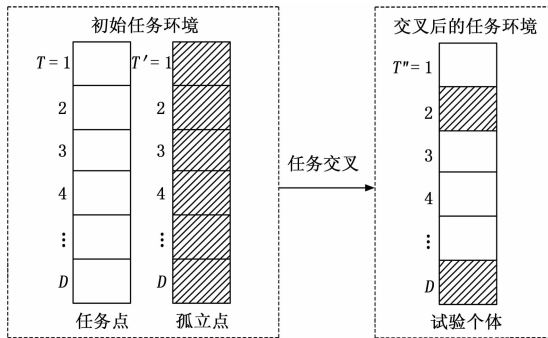


图5 孤立点交叉更新机制

图5所示为任务环境突然出现孤立点导致任务环境

状态发生改变, 因此设计任务交叉机制对任务环境重新进行初始化操作。任务环境中原先存在的任务目标点继续分配在原有的聚类簇在中, 对任务环境突然出现的孤立点进行聚类簇计算, 将其归属到距离最近的聚类簇中, 同时对任务环境的任务目标点重新进行编号。对于运动的任务目标点, 则实时计算其与相邻聚类簇的距离, 在当前聚类簇的范围类则不进行二次任务点交叉操作, 若运动超出当前聚类簇, 则进行二次任务点交叉操作, 以此类推。由孤立点交叉更新机制可以得出任务环境状态的更新表示为:

$$S_n^{t+1} = A \cdot |S_{\text{best}}' - S_n'| \cdot \omega_n + S_{\text{best}}' \quad (38)$$

式中, S_{best}' 为 t 时刻最优的任务环境状态, A 为权重系数, ω_n 为个体 n 的权重因子, 其表达式为:

$$\omega_n = 0.75 + p_n \quad (39)$$

当 $\omega_n > 1$ 时, 提升位置更新能力, 扩大任务环境状态更新范围, 确保所有离散任务点被更新到; 当 $\omega_n < 1$ 时, 减小位置更新能力, 提高算法的收敛性。式中 p_n 为无人机个体任务更新概率, 其表达式为:

$$p_n = \frac{\text{rank}(\text{fitness})}{n \cdot \epsilon_t} \quad (40)$$

式中, fitness 为无人机个体适应度, ϵ_t 为时刻无人机的探索率, 由式 (26) 和式 (30) 定义 fitness 的表达式为:

$$\text{fitness} = R + r \quad (41)$$

可知值函数更新公式为:

$$Q(a_t | S_t) = Q(S_t | a_t) + \alpha[r_{t+1} + \gamma \cdot \max Q(S_n' | a_{t+1}) - Q(S_t | a_t)] \quad (42)$$

在线决策环节, 根据式 (42) 更新前述环节生成的离线类脑智能决策策略, 得到在线决策的最优任务分配策略, 当任务环境出现改变时, 实时进行多无人机的任务分配。在线决策的过程如算法 2 所示:

算法 2: 在线决策

输入: 最优任务分配策略 π^* , 任务环境状态参数

输出: 在线决策任务分配集

1) 对于任务环境和无人机参数:

判断与数据库里的参数是否一致, 初始化任务分配集

2) 使用离线最优任务分配策略 π^* 进行任务分配

针对任务需求根据式 (42) 更新任务环境状态 S_t , 更新最优任务分配策略 π^*

$a \leftarrow \arg Q^*(a_t | S_t)$

生成新的任务分配集

是否满足式 (3-17)

若不满足

重新迭代更新任务分配集, 重复上述步骤

若满足

直接输出最优任务分配集

3) 结束

3 仿真与分析

为验证多无人机类脑智能任务分配算法的有效性，本文基于 Matlab 进行仿真验证，并对仿真试验作出以下合理假设：执行多无人机任务分配时，所有任务目标的位置已知。

3.1 参数设置

假设当前某无人机基地停有 10 架无人机，共有 100 个任务需要完成，完成每个任务所消耗的载荷相同，奖励权重系数 $k_1=0.3$ ， $k_2=0.2$ ， $k_3=0.3$ ， $k_4=0.2$ ，惩罚值 $C_1=2$ ， $C_2=1$ ， $c_3=1$ ，无人机的参数如表 1 所示，任务的参数如表 2 所示。

表 1 无人机参数设置

坐标/km	航程/km	载荷/kg
(-1,-5)	400	12
(0,-1)	400	13
(-3,-4)	400	12
(-5,-2)	400	12
(-8,-7)	400	13
(-4,-9)	400	13
(-10,0)	400	13
(-9,-2)	400	13
(-7,-10)	400	12
(-7,-4)	400	12

表 2 任务点参数设置

T	坐标/km	T	坐标/km
T_1	(32,8)	$\vdots$	
T_2	(8,48)	T_{96}	(77,4)
T_3	(70,21)	T_{97}	(45,80)
T_4	(88,18)	T_{98}	(77,8)
T_5	(87,24)	T_{99}	(52,37)
$\vdots$	$\vdots$	T_{100}	(9,67)

比较聚类算法、初始强化学习算法与本文的融合改进强化算法的任务分配方案。为保证对比实验的可信性，3 种算法的相关参数一致，设置如表 3 所示。为同时获取算法收敛精度与收敛速度信息，迭代次数选取 1 000 次。

表 3 各算法参数设置

算法	参数	数值
CA	聚类簇	1~13
	最大迭代次数	100
	初始中心运行次数	20
RLA/FIRLA	学习率	0.1
	折扣因子	0.9
	初始探索率	0.9
	最终探索率	0.1
	预设迭代次数阈值	500
	权重因子	0.75~1.25

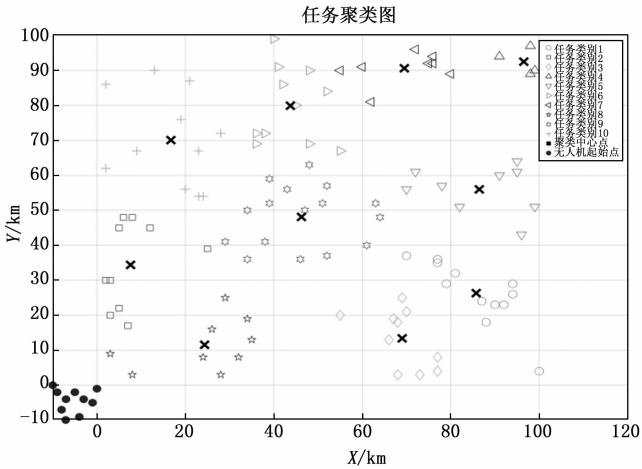
3.2 实验结果及分析

3.2.1 任务聚类仿真结果分析

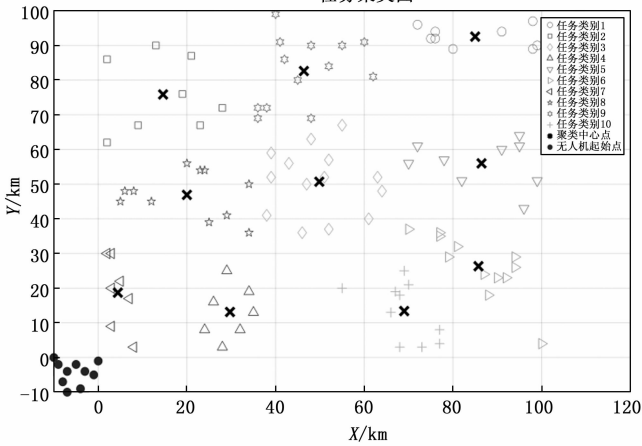
首先，对 100 个任务目标进行聚类；然后通过计算任务目标与所属聚类簇的距离均值得到聚类中心坐标，聚类结果如表 4、图 6 和图 7 所示，图中“×”表示聚类中心点，不规则形状代表无人机任务点，同一类中的任务点表示为同一形状，“●”表示无人机起始位置。

表 4 任务聚类结果

K-means 聚类结果		改进组合聚类结果	
聚类中心编号	坐标	聚类中心编号	坐标
1	(7.60, 34.40)	1	(24.00, 76.78)
2	(16.73, 70.09)	2	(24.00, 76.78)
3	(24.33, 11.56)	3	(43.79, 47.86)
4	(43.73,79.91)	4	(85.75, 26.33)
5	(46.25, 48.13)	5	(29.71, 13.14)
6	(69.00, 13.40)	6	(85.00, 92.56)
7	(69.50, 90.63)	7	(13.40, 51.80)
8	(85.75, 26.33)	8	(49.82, 84.36)
9	(86.44, 56.00)	9	(82.27, 54.91)
10	(96.50, 92.50)	10	(69.00, 13.40)



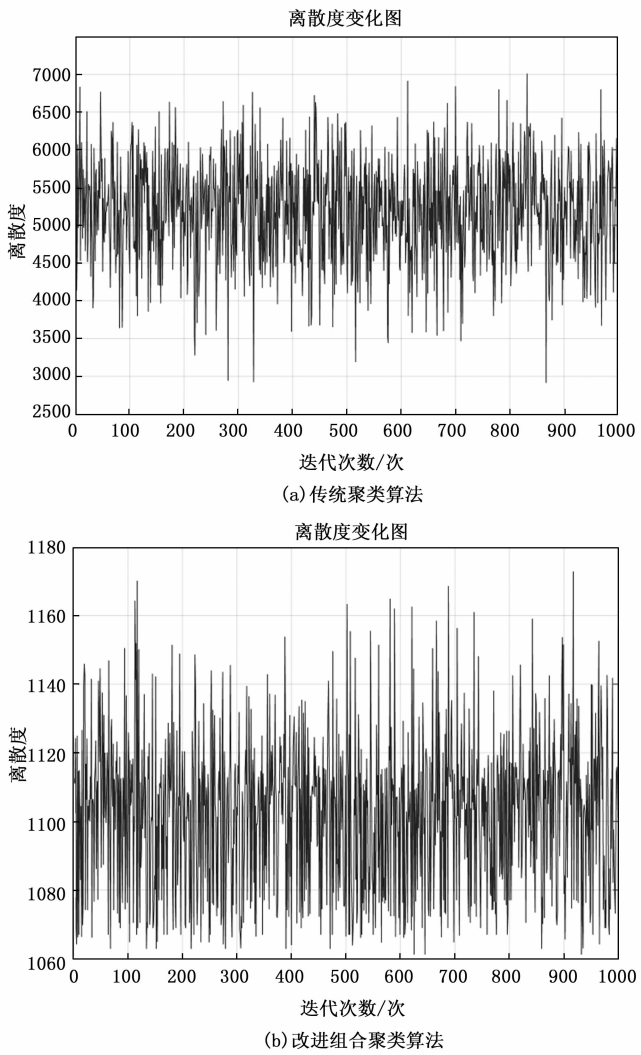
(a) 传统聚类算法



(b) 改进组合聚类算法

图 6 任务聚类效果对比图

图 6 (a) 和图 6 (b) 分别为传统聚类算法和改进组合聚类算法的任务聚类效果图。图 7 (a) 和图 7 (b) 分别为传统聚类算法和改进组合聚类算法离散度图。对表 4、任务聚类效果图和离散度图可以看出传统聚类算法在计算簇内平均值时容易受离散点干扰, 从而影响聚类中心的计算, 导致聚类结果不准确, 而改进组合聚类算法有效地减少聚类的离散度, 提高任务聚类质量。因此, 将图 6 (b) 所示的任务聚类集作为任务预分配的任务分配集, 为下一步的任务重分配奠定基础。



3. 2. 2 任务重分配仿真结果分析

任务重分配首先针对预分配集进行混合强化学习策略进行任务重分配, 得到任务重分配集, 然后对任务重分配集的孤立点使用附有权重系数的离散化策略进行处理后更新任务状态空间, 进行任务再分配, 直到任务分配集不再发生变化输出最优任务分配策略, 仿真的结果如图 8 所示。图 8 (a) 为混合强化学习策略的任务重分配图, 图 8 (b) 为任务重分配的航程代价对比图, 对比图 6 (b) 所示的改进组合聚类算法任务预分配集可知,

聚类算法调用了 10 架无人机执行任务, 任务重分配算法在与分配的基础上进行改进, 仅调用了 8 架无人机执行任务, 任务重分配算法在无人机资源调度上更合理。

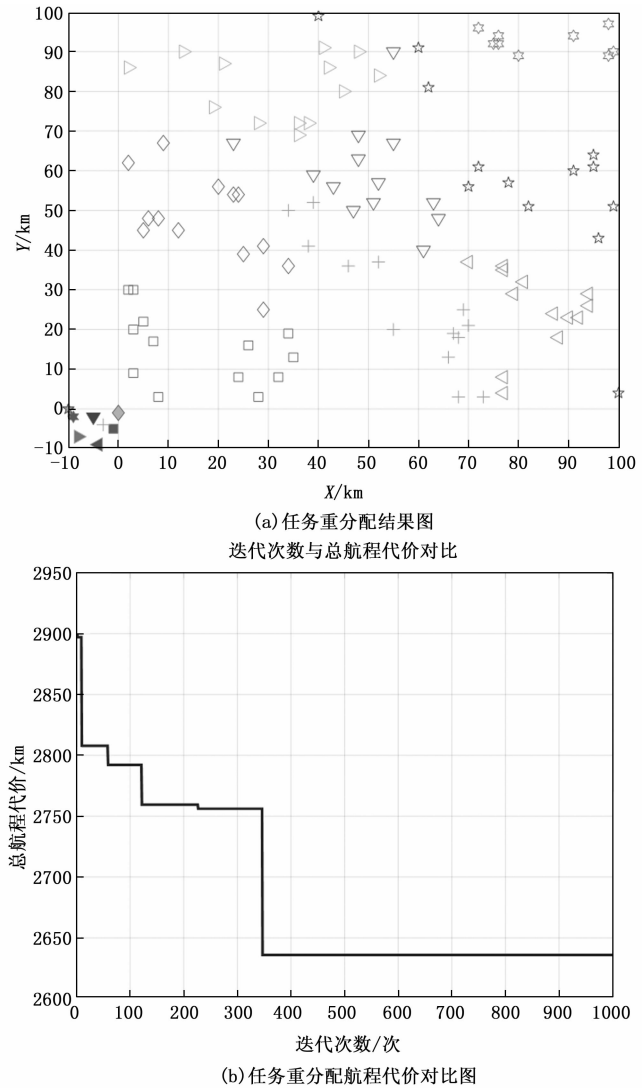


图 8 任务重分配结果图

虽然混合强化学习策略的任务重分配在满足多种约束条件的基础上有效地解决了多无人机任务分配的冗余分配问题, 但是未解决孤立点对任务分配的影响, 导致任务重分配陷入局部最优无法完成多无人机任务的最优分配, 此时根据孤立点离散化策略进行任务再分配, 任务再分配结果如图 9 所示。图 9 (a) 为孤立点离散化的最优策略任务分配图, 图 9 (b) 为孤立点离散化的最优策略任务分配航程代价对比图。与图 8 对比可知, 引入孤立点离散化策略后, 生成的最优任务分配策略减弱了孤立点对任务分配的影响, 使无人机任务分配更容易达到全局最优分配。

3. 2. 3 多算法性能对比

为验证基于类脑智能模型的任务分配方法的高效性, 本文将类脑智能方法与传统聚类算法和基本强化学

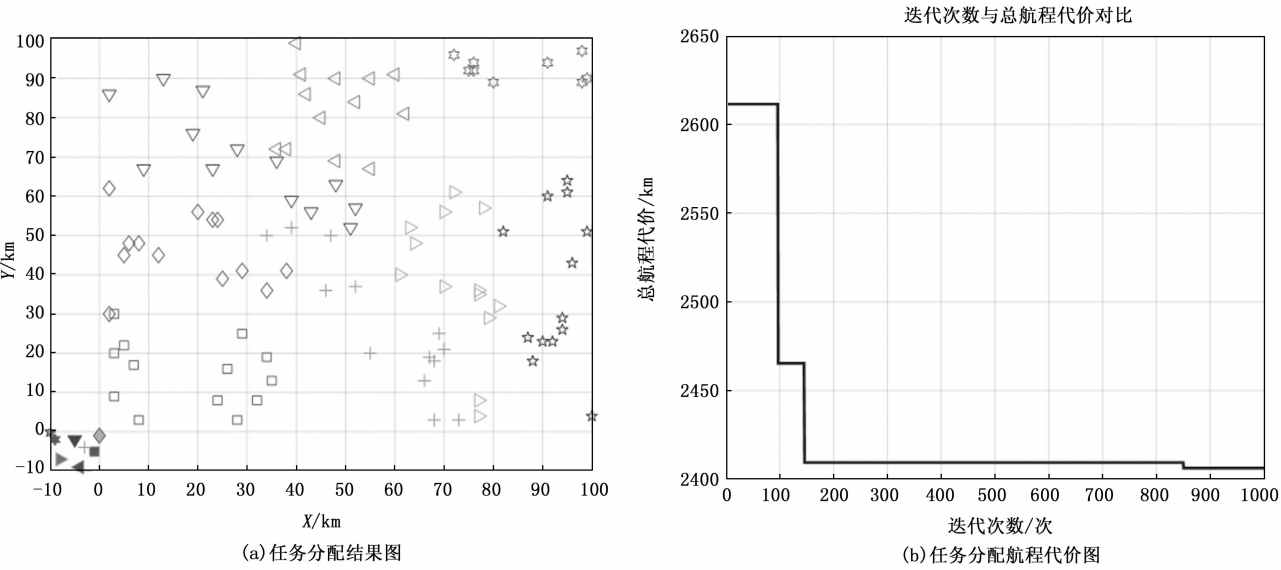


图 9 孤立点离散化后任务分配结果图

习算法在相同任务环境下对多无人机任务分配问题进行求解，并将多无人机任务分配结果进行比较。为排除混合策略强化学习在任务预分配集上进行重分配与基本强化学习直接在任务空间进行分配产生初始参数差异化，

本文也设计了使用基本强化学习算法在任务预分配集进行重分配，保证算法性能对比的公平性。具体任务分配结果如表 5 和图 10 所示。

图 10 (a)、图 10 (b)、图 10 (c) 和图 10 (d) 分

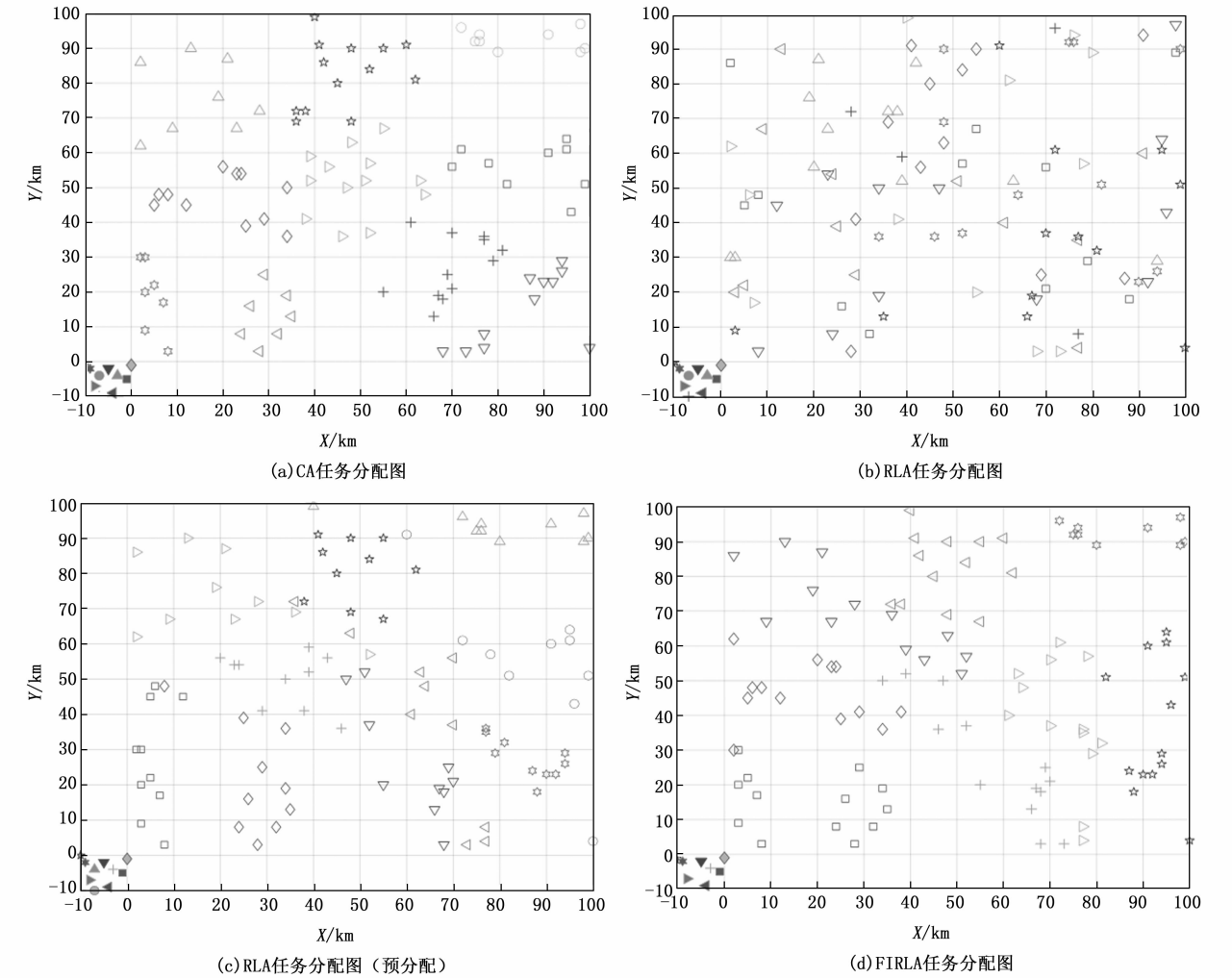


图 10 各算法任务分配效果图

表 5 各算法任务分配结果表

算法	初始任务 点数目	完成任务 点数目	最优航程 /km	耗时 /s	无人机 执行数量
CA	100	100	2 549.5	23.78	10
RLA	100	100	4 985.3	14.73	10
RLA (预分配)	100	100	2 951.4	12.79	10
FIRLA	100	100	2 408.0	7.04	8

别为传统聚类算法、传统强化学习算法、预分配基础上的传统强化学习算法和基于类脑模型的改进混合策略强化学习的任务分配结果图。结合表 5 可以看出, FIRLA 相较于 RLA 算法和 CA 算法, 面对大批量的任务分配可以得出较优解, 且规划解的航程代价相较 RLA 算法提升 51.6%; 在保证无人机总航程代价最优的前提下, FIRLA 的迭代时间相较 CA 算法提升 70.4%; 在相同于分配方案基础上, FIRLA 相对于传统 RLA 算法的航程代价提升 18%, 迭代时间提升了 45%。且 FIRLA 相较于前 3 种方法在无人机调度上具有较好的自主性, 派遣执行任务的无人机数量随任务环境的变化进行动态调整, 体现了其在大规模任务下的多无人机任务分配的明显优势。

3.3 在线决策验证与结果分析

为验证基于类脑决策模型的多无人机任务分配方法的有效性, 本节设计了典型任务场景进行多无人机任务分配, 以表 1~3 为在线决策仿真场景基础参数设置, 具体场景设置和决策仿真结果如下。

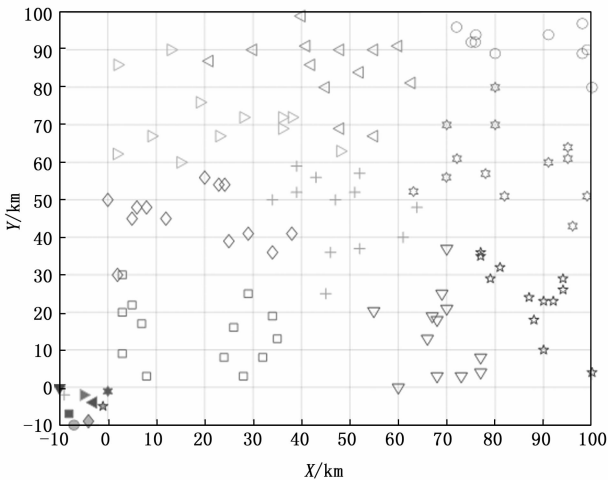
1) 场景一: 针对任务场景静态任务数量变化的在线决策。

设置在表 2 任务环境的基础上, 某一时刻任务环境中的静态任务点数量突然变化, 测试基于类脑决策模型的多无人机任务分配对静态任务环境变化的有效性, 场景一的静态任务点数量变化如表 6 所示, 决策仿真结果如图 11 所示。

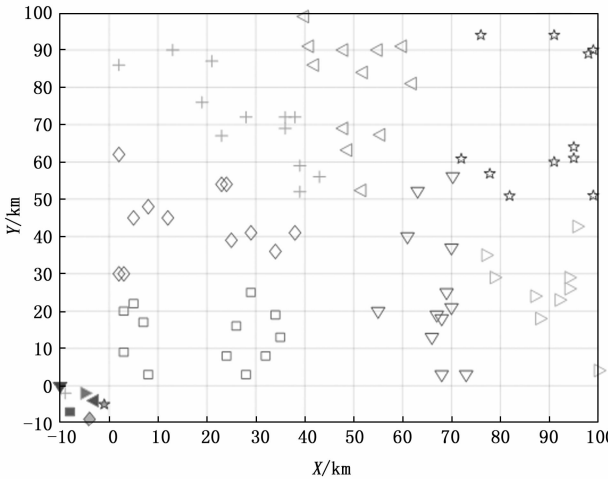
表 6 静态任务点数量增加场景参数表

决策场景	坐标(x,y)	坐标(x,y)	坐标(x,y)	坐标(x,y)	坐标(x,y)
任务点 增加	(70,70)	(0,50)	(90,10)	(80,70)	(80,80)
	(45,25)	(30,90)	(100,80)	(60,0)	(15,60)
任务点 减少	(81,32)	(6,48)	(72,96)	(80,89)	(77,36)
	(47,50)	(90,23)	(52,57)	(34,50)	(20,56)
	(64,48)	(76,92)	(98,97)	(75,92)	(46,36)
	(77,4)	(45,80)	(77,78)	(52,37)	(9,67)

图 11 (a) 和图 11 (b) 分别为静态任务点增多场景和任务点减少场景的任务分配仿真结果图, 由图可知, 当静态任务点变化时, 决策系统对任务环境状态重新进行初始化, 并使用经验判断一任务预分配一任务重分配一在线决策的类脑智能任务分配方法实时进行多无



(a) 任务点增多场景仿真结果图



(b) 任务点减少场景仿真结果图

图 11 场景一在线决策结果图

人任务分配。由结果可知, 任务点增多场景选择了 9 架无人机参与执行任务, 任务点减少场景选择了 7 架无人机参与执行任务, 充分验证了类脑决策系统在任务分配的资源调度和实时性上具有显著的优越性。

2) 场景二: 针对任务场景动态任务点的在线决策。

场景二设置在表 2 任务环境的基础上, 某一时刻任务环境中出现动态任务点开始运动, 测试基于类脑决策模型的多无人机任务分配对动态任务环境变化的有效性, 分别对动态点初始状态、动态点在当前聚类簇中的运动状态和动态点驶离当前聚类簇的运动进行任务分配决策, 状态场景二的动态任务点数量变化参数如表 7 所示, 决策仿真结果如图 12 所示。

表 7 场景二动态任务点参数表

决策场景	初始坐标(x,y)	运动方向	运动速度
动态任务点变化	(20,30)	Pi/6	30 m/s

对比图 12 (a)、图 12 (b) 和图 12 (c) 可以看出, 图 12 (a) 所示任务环境中出现动态任务点时, 多无人

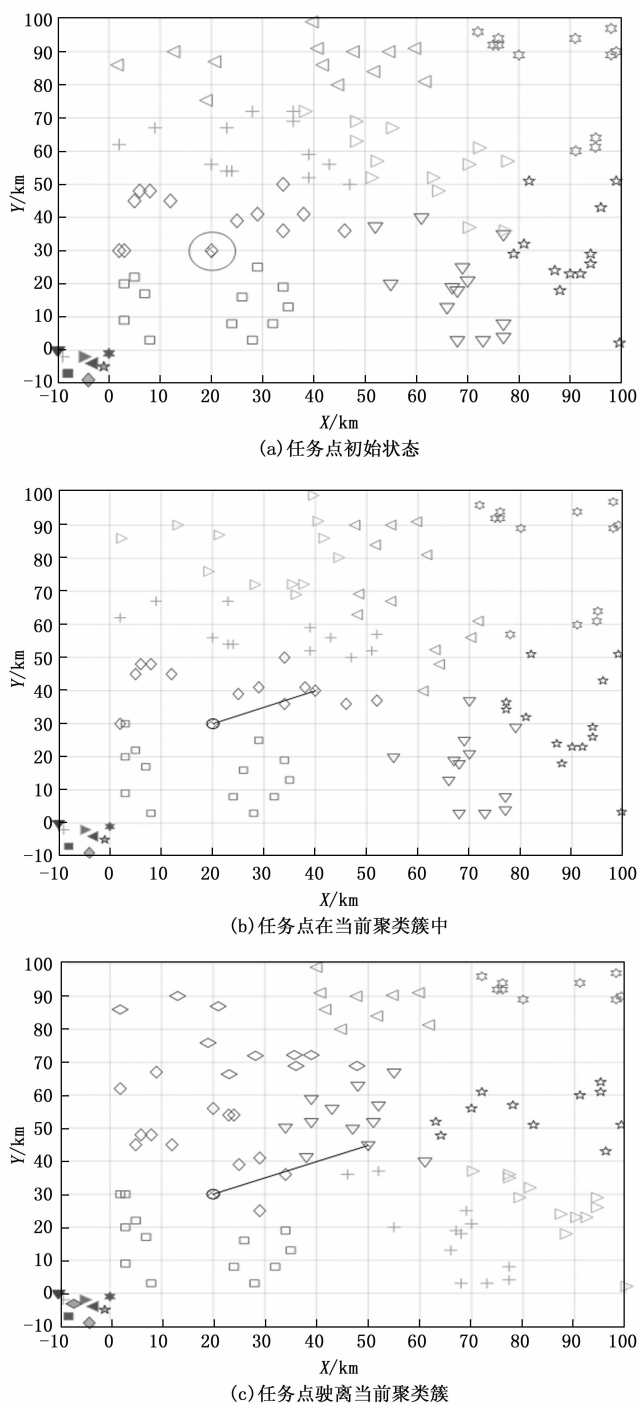


图 12 任务场景二在线决策仿真结果图

机任务分配类脑决策系统将任务环境初始化,进行了任务再分配。图 12 (b) 所示动态任务点在当前聚类簇中运动时,聚类簇不发生变化,继续进行最优任务分配。图 12 (c) 所示无人机驶离当前聚类簇时,由于动态点运动的加入,当前部分任务环境,聚类簇发生变化,决策系统进行在线决策后作出派遣 9 架无人机执行任务,验证了本文的类脑决策系统在处理动态任务分配时的快速响应能力和自适应性。

4 结束语

本文根据人脑运行机制建立了类脑决策模型,并针对多无人机任务冗余分配的问题,进行算法改进。任务预分配环节引入聚类算法,建立并改进近似类抽样的组合聚类策略,确保了算法对不同任务分布的鲁棒性。任务重分配环节针对任务探索提出动态衰减梯度策略,使用混合策略 Q-Learning 更新机制进行任务重分配,并生成离线决策策略。在线决策环节依此策略引入附有权重系数的孤立点离散化策略,实现全局探索与局部优化的平衡,保证了任务搜索与执行的动态统一。

同时,本文只进行了多无人机单一关系下的任务规划和决策研究,虽然考虑到多种约束条件和目标函数,但相比实际战场环境在复杂度上仍有差距,未来将进一步展开多无人机联合、竞争和对抗等关系下的任务规划和自主决策技术研究以满足现实任务规划的需要。

参考文献:

- [1] 李璐璐,朱睿杰,隋璐瑶,等. 智能集群系统的强化学习方法综述 [J]. 计算机学报, 2023, 46 (12): 2573 - 2596.
- [2] 陈 侠,乔艳芝. 无人机任务分配综述 [J]. 沈阳航空航天大学学报, 2016, 33 (6): 1 - 7.
- [3] 杜枯康,赵英凯. 整数规划问题智能求解算法综述 [J]. 计算机应用研究, 2010, 27 (2): 408 - 412.
- [4] 张硕航,郭改枝. 多旅行商模型及其应用研究综述 [J]. 计算机科学与探索, 2022, 16 (7): 1516 - 1528.
- [5] NYGARD K E, CHANDLER P R, PACHTER M. Dynamic network flow optimization models for air vehicle resource allocation [C] // Proceedings of the 2001 American Control Conference (ACC), Arlington, VA, USA. 2001: 1853 - 1858.
- [6] 吴 振,吴红兰. 基于改进遗传算法的无人机航路规划 [J]. 电子测量技术, 2021, 44 (24): 52 - 58.
- [7] 唐必伟. 粒子群算法的改进及其在无人机任务规划中的应用 [D]. 西安: 西北工业大学, 2020.
- [8] 马骋乾,谢 伟,孙伟杰. 强化学习研究综述 [J]. 指挥控制与仿真, 2018, 40 (6): 68 - 72.
- [9] 邓林培. 经典聚类算法研究综述 [J]. 科技传播, 2019, 11 (5): 108 - 110.
- [10] WEN U P, LAN K M, SHIH H S. A review of Hopfield neural networks for solving mathematical programming problems [J]. European Journal of Operational Research, 2009 (3): 675 - 687.
- [11] 蒲慕明,徐 波,谭铁牛. 脑科学与类脑研究概述 [J]. 中国科学院院刊, 2016, 31 (7): 725 - 736.
- [12] 赵欣怡,宗 群,张睿隆,等. 类脑智能技术在无人系统上的应用 [J]. 控制理论与应用, 2019, 36 (1): 1 - 12.

(下转第 322 页)