

控制与决策

Control and Decision

基于Q学习的多任务多目标粒子群优化算法

韩红桂, 徐子昂, 王晶晶

引用本文:

韩红桂,徐子昂,王晶晶. 基于Q学习的多任务多目标粒子群优化算法[J]. 控制与决策, 2023, 38(11): 3039–3047.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2022.1662>

您可能感兴趣的其他文章

Articles you may be interested in

[基于R2指标和目标空间分解的高维多目标粒子群优化算法](#)

R2 indicator and objective space partition based many-objective particle swarm optimizer

控制与决策. 2021, 36(9): 2085–2094 <https://doi.org/10.13195/j.kzyjc.2020.0113>

[具有重组学习和混合变异的动态多种群粒子群优化算法](#)

Dynamic multi-population particle swarm optimization algorithm with recombined learning and hybrid mutation

控制与决策. 2021, 36(12): 2871–2880 <https://doi.org/10.13195/j.kzyjc.2020.0898>

[基于局部搜索的反向学习竞争粒子群优化算法](#)

Opposition-based learning competitive particle swarm optimizer with local search

控制与决策. 2021, 36(4): 779–789 <https://doi.org/10.13195/j.kzyjc.2019.1150>

[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>

[阴影条件下基于迁移强化学习的光伏系统最大功率跟踪](#)

Transfer reinforcement learning based maximum power point tracker of PV systems under partial shading condition

控制与决策. 2020, 35(12): 2939–2949 <https://doi.org/10.13195/j.kzyjc.2019.0412>

基于 Q 学习的多任务多目标粒子群优化算法

韩红桂^{1,2†}, 徐子昂^{1,2}, 王晶晶^{1,2}

(1. 北京工业大学 信息学部, 北京 100124; 2. 计算智能与智能系统北京市重点实验室, 北京 100124)

摘要: 多任务粒子群优化算法 (multi-task particle swarm optimization, MTPSO) 通过知识迁移学习, 具有快速收敛能力, 广泛应用于求解多任务多目标优化问题. 然而, MTPSO 难以根据种群进化状态自适应调整优化过程, 容易陷入局部最优, 收敛性能较差. 针对此问题, 利用强化学习的自我进化与预测能力, 提出一种基于 Q 学习的多任务多目标粒子群优化算法 (QM^2 PSO). 首先, 设计粒子群参数动态更新方法, 利用 Q 学习方法在线更新粒子群算法的惯性权重和加速度参数, 提高当前粒子收敛到 Pareto 前沿的能力; 其次, 提出基于柯西分布的突变搜索策略, 通过全局和局部交替搜索多任务最优解, 避免算法陷入局部最优; 最后, 设计基于正向迁移准则的知识迁移方法, 采用 Q 学习方法更新知识迁移率, 改善知识负迁移现象. 与已有经典算法的对比实验结果表明所提出的 QM^2 PSO 算法具有更优越的收敛性.

关键词: 多任务多目标优化; 粒子群优化; 知识迁移; Q 学习

中图分类号: TP18 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2022.1662

引用格式: 韩红桂, 徐子昂, 王晶晶. 基于 Q 学习的多任务多目标粒子群优化算法 [J]. 控制与决策, 2023, 38(11): 3039-3047.

A Q -learning-based multi-task multi-objective particle swarm optimization algorithm

HAN Hong-gui^{1,2†}, XU Zi-ang^{1,2}, WANG Jing-jing^{1,2}

(1. Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China; 2. Beijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing 100124, China)

Abstract: The multi-task particle swarm optimization (MTPSO) algorithm is widely used to solve multi-task multi-objective problems due to its rapid convergence via knowledge transfer learning. However, the MTPSO has strong randomness and is lack of guideness during search process, which is prone to fall into local optimum and has poor convergence performance. This paper proposes a Q -learning-based multi-task multi-objective particle swarm optimization algorithm (QM^2 PSO) via using learning and prediction of reinforcement learning to guide optimization. Firstly, we design the adaptive parameter adjustment method, which can update the inertia weight and acceleration parameters online based on Q -learning to improve the convergence ability. Secondly, we develop a mutation strategy based on Cauchy distribution, which can balance exploration and exploitation to avoid falling into local optimum. Finally, we design a knowledge transfer method based on the positive transfer criterion via updating the knowledge transfer rate based on Q -learning to avoid negative knowledge transfer. The comparative results demonstrate that the QM^2 PSO is superior to the existing algorithms on convergence performance.

Keywords: multi-task multi-objective optimization; particle swarm optimization; knowledge transfer; Q -learning

0 引言

多目标优化问题是一类具有多个冲突优化目标函数的问题^[1-2], 通常不存在所有目标同时最优的情况. 不同于单目标优化, 多目标优化采用 Pareto 解集

表示所有非支配解的集合^[3-5], 其在目标空间构成的平面为 Pareto 前沿. 多目标优化的目的是尽可能得到趋近于真实 Pareto 前沿的解集.

目前, 求解多目标优化问题的方法主要包括传统

收稿日期: 2022-09-20; 录用日期: 2022-12-30.

基金项目: 国家自然科学基金项目 (62125301, 61890930-5, 61903010, 62021003); 国家重点研发计划项目 (2022YFB3305800-5, 2018YFC1900800-5); 北京市教育委员会科技计划重点项目 (KZ202110005009); 北京高校卓越青年科学家项目 (BJJWZYJH01201910005020); 青年北京学者项目 (037).

责任编辑: 林崇.

†通讯作者. E-mail: rechardhan@bjut.edu.cn.

优化算法^[6]和群体智能优化算法^[7-12]. 传统优化算法以线性规划为代表,能够通过数学方法精确求解,但求解时间长,无法得到全部非支配解集. 作为生物启发式方法,群智能优化算法通过群体间的竞争、学习、交互等方式扩大搜索范围,增强搜索效率,从而找到更优的个体. 群智能优化算法具有全局搜索性能,适用于求解复杂优化问题,已成为多目标优化的主流方法. Dai等^[13]提出了一种基于目标空间分解的多目标遗传算法,通过在不同的目标空间进行学习,增加了个体在目标空间的搜索范围,算法所得Pareto解集兼具更优的收敛性能和多样性. Cheng等^[14]提出了一种基于非支配排序的改进多目标遗传算法,通过加强局部搜索能力提高了算法收敛速度. Liu等^[15]提出了一种基于目标空间分解的动态多目标进化算法,适应了各种类型的环境变化.

针对多任务优化问题中多个并行且相关的任务,一些学者通过挖掘不同任务之间的关联性将优化过程中得到的信息进行迁移,同时提高多个任务的优化效果^[16-17]. 例如, Yao等^[18]提出了一种基于分解及动态资源分配策略的多目标多因子优化算法,对每个任务进行动态资源分配,提高了算法的收敛速度. 然而,当多任务转化为单目标求解时,难以克服目标的差异性,导致该类算法难以获取真实优化结果. Huang等^[19]提出了一种基于自适应知识迁移的多任务进化算法,增加了群体的搜索效率,然而难以保证优化解集的多样性. Ma等^[20]基于多任务中的先验知识构建了任务间的紧密联系,抑制了知识迁移中负向迁移情况的发生.

目前,求解多任务多目标优化问题的群智能优化算法主要包括遗传算法^[21]、蚁群算法^[22]、粒子群算法等^[23-24]. 其中,粒子群算法拥有搜索速度快、参数量少、易于知识迁移等特点^[25-29],求解多任务多目标问题具有独特的优势. 粒子群算法^[30-32]模拟鸟群的捕食行为,通过迭代每个粒子的位置和速度,利用不同粒子间的拓扑关系寻找最优解. Han等^[33]提出了一种自调整多任务粒子群算法,通过评价迁移知识的质量避免了负向迁移问题,有效确保迁移过程正向进行. Tang等^[34]设计了一种任务间分级等级评价的粒子群学习策略,增强了粒子群的空间探索能力. 然而,此类算法不能根据粒子群所处环境动态选择合适的更新参数来平衡全局探索和局部开发,并动态调整知识迁移率.

粒子群算法中的粒子进化过程复杂多样,难以判断,许多学者所提的参数自适应方法无法全局考虑

优化过程中粒子的变化状态. 不同于传统参数动态调节策略,强化学习^[35]可通过预测下一刻状态动态调整算法参数以适应优化整体过程,提高算法收敛性能. Lu等^[36]设计了一种基于强化学习的精英粒子选择算法,引导个体的更新模式,改善算法性能. 强化学习中基于值的Q学习算法^[37-42]具有离散型特性,适用于自适应调整算法进化不同阶段的参数,有利于优化过程中的参数选择.

基于上述研究,本文提出一种基于Q学习的多任务多目标粒子群优化算法(QM²PSO). 首先,设计一种基于Q学习参数更新方法,在线动态更新速度公式中的惯性权重和加速度参数,提高当前粒子的收敛速度;其次,基于柯西分布提出一种突变策略实现全局和局部的交替搜索,避免算法陷入局部最优;最后,将多任务知识迁移过程与Q学习方法相结合,基于正向迁移准则实现知识正向迁移,改善知识负迁移现象,提高知识利用率;最后,通过仿真实验在多任务多目标测试函数上与已有算法进行对比,验证所提出算法具有更优越的收敛性能.

1 多任务多目标粒子群算法

1.1 多目标优化问题

多目标优化问题可以表示为

$$\begin{aligned} \min T(x) &= [f_1(x), f_2(x), \dots, f_M(x)]; \\ \text{s.t. } x &\in \Omega. \end{aligned} \quad (1)$$

其中: $\Omega \in \mathbb{R}^D$ 为决策空间, $x = [x_1, x_2, \dots, x_D]$ 为D维决策变量, M 为目标函数的个数. 针对多目标优化问题给出以下定义.

定义1 (支配关系) 假设存在一组 $x_1, x_2 \in \Omega$, 如果 $\forall i \in \{1, 2, \dots, M\}, f_i(x_1) \leq f_i(x_2)$, 且 $\exists j \in \{1, 2, \dots, M\}$ 满足 $f_j(x_1) < f_j(x_2)$, 则称 x_1 支配 x_2 , 表示为 $x_1 \succ x_2$.

定义2 (Pareto解) 对于一个解 x^* 而言, 如果不存在任何一个解支配 x^* , 则称 x^* 为Pareto解.

定义3 (Pareto解集) 多目标优化问题所有Pareto解的集合称为Pareto解集(Pareto set, PS).

定义4 (Pareto前沿) Pareto解集在目标空间, 即 $F(x) \in \mathbb{R}^M | x \in \text{PS}$ 称为Pareto前沿.

1.2 多任务多目标优化问题

多任务多目标优化问题是一类包含多个不同优化任务, 且每个独立的优化任务具有多个优化目标的问题. 此类问题可描述为

$$\begin{aligned} \{x_1, x_2, \dots, x_K\} &= \operatorname{argmin}\{T_1(x), T_2(x), \dots, T_K(x)\}; \\ \text{s.t. } T_k(x) &= (f_1(x), f_2(x), \dots, f_M(x)), \end{aligned}$$

$$k = 1, 2, \dots, K. \quad (2)$$

其中: K 为优化任务的数量, M 为优化任务目标函数的个数, $T_k(x)$ 为不同任务的函数表达式, $f_M(x)$ 为目标函数的表达式.

群体进化算法求解多任务多目标优化问题时, 不仅要考虑个体对单一任务中不同目标的平衡, 也要判断个体在不同任务中的优劣, 并据此在搜索过程中对其分派合适的任务. 因此, 需要对个体在不同任务的表现进行评价, 评价指标定义如下.

定义 5 (因子等级) 采用个体 p_i 在种群中排序后的索引 r_j^i 表示其在任务 T_j 上的因子等级.

定义 6 (技能因素) 个体 p_i 求解不同任务的能力, 表示为 $\tau_i = \arg \min_{j \in \{1, 2, \dots, k\}} \{r_j^i\}$.

定义 7 (适应度标量) 采用因子等级的倒数 $\varphi_i = 1/r_{\tau_j}^i$ 表示任意个体 p_i 的适应度标量.

因子等级体现单一多目标任务上个体的表现, 技能因素则反应个体在不同任务中的表现, 适应度标量反应个体在某一任务的优劣. 以上定义有利于对算法中的个体在不同任务上的性能进行评价, 便于优化过程的粒子选取和知识迁移中的任务评估.

1.3 多任务多目标粒子群优化算法

与经典粒子群算法的整体结构相似, 多任务多目标粒子群优化算法在进化选择的过程中考虑多任务多目标优化中的特点, 通过不同任务之间的隐性知识进行信息交互和知识迁移. 多任务多目标粒子群算法中, 定义粒子 p_i 的位置为 $x_i(t) = [x_{i,1}(t), x_{i,2}(t), \dots, x_{i,d}(t)]$, 速度为 $v_i(t) = [v_{i,1}(t), v_{i,2}(t), \dots, v_{i,d}(t)]$, p_i 更新中的最优位置 (pbest) 定义为 $p_i(t) = [p_{i,1}(t), p_{i,2}(t), \dots, p_{i,d}(t)]$. 进化过程中种群的全局最优 (gbest) 定义为 $p_g(t) = [p_{g,1}(t), p_{g,2}(t), \dots, p_{g,d}(t)]$, d 为变量的维度.

在优化过程开始随机初始化 N 个粒子, 生成其位置和速度, 评价粒子的适应度标量和技能因素, 在每个任务中选取 gbest, 从而产生初始种群.

不同任务知识之间的知识迁移由随机交配概率 (random mating probability, rmp) 决定. 为平衡知识迁移发生概率, 设置 $\text{rmp} = 0.6$. 当生成的随机数满足 $\text{rand} < \text{rmp}$ 时, 群体中粒子的更新方式与经典粒子群更新算法相同, 如下所示:

$$v_{i,d}(t+1) = \omega v_{i,d}(t) + c_1 r_1 (p_{i,d}(t) - x_{i,d}(t)) + c_2 r_2 (p_{g,d}(t) - x_{i,d}(t)), \quad (3)$$

$$x_{i,d}(t+1) = x_{i,d}(t) + v_{i,d}(t+1). \quad (4)$$

其中: $d = 1, 2, \dots, D$, ω 为惯性权重, c_1 和 c_2 为加速

度常量, r_1 和 r_2 是取值范围为 $[0, 1]$ 的随机值.

当 $\text{rand} \geq \text{rmp}$ 时, 粒子速度更新公式表示为

$$v_{i,d}(t+1) = \omega v_{i,d}(t) + c_1 r_1 (p_{i,d}(t) - x_{i,d}(t)) + c_2 r_2 (p_{g,d}(t) - x_{i,d}(t)) + c_3 r_3 (x_{g,d}^*(t) - x_{i,d}(t)). \quad (5)$$

其中: $x_{g,d}^*$ 为学习变量, c_3 为知识迁移率, r_3 是取值范围为 $[0, 1]$ 的随机值.

2 基于 Q 学习的多任务多目标粒子群优化算法

Q 学习是 Watkins 在 1989 年提出的一种强化学习算法^[38-39], 包括智能体、状态、动作、奖励和环境 5 个要素. 本节基于 Q 学习通过环境反馈进行学习记忆的特点, 提出一种基于 Q 学习的多任务多目标粒子群算法 (QM²PSO), 如图 1 所示. 下面详细介绍 Q 表设计方法和基于 Q 学习的参数自适应调节方法, 最后描述算法整体流程.

2.1 Q 学习状态设置

将粒子在解空间的位置作为状态, 通过历史最优解引导粒子寻优, 加快粒子收敛. 粒子的状态 S_t 由粒子与当前任务全局最优解 gbest_t 的欧氏距离表示, 有

$$S_t = \|x_t - \text{gbest}_t\|. \quad (6)$$

其中: x_t 为粒子在第 t 代时所处的位置, gbest_t 为粒子群在第 t 代时的全局最优, $\|x_t - \text{gbest}_t\|$ 为粒子到全局最优的欧氏距离.

由于 Q 表的状态需离散化表示, 本文将 S_t 归一化为 \bar{S}_t , 并均匀离散化为 4 个状态, 如下所示:

$$\bar{S}_t = \begin{cases} L, & 0.75 \leq S_t; \\ \bar{L}, & 0.5 \leq S_t < 0.75; \\ \bar{G}, & 0.25 \leq S_t < 0.5; \\ G, & 0 \leq S_t < 0.25. \end{cases} \quad (7)$$

其中 L, \bar{L}, \bar{G}, G 分别为粒子的 4 种不同状态.

2.2 Q 学习动作设置

为设计参数动态选择策略, 需根据参数性质设计不同的参数组合作为 Q 学习的动作. 粒子群更新过程包含 3 个参数: 惯性权重 ω 、加速度系数 c_1 和 c_2 . 惯性权重 ω 决定了与上一时刻粒子移动速度的相关程度, ω 越大粒子越保持原有运动趋势; 加速度系数 c_1 反映粒子对于自身历史的记忆力, 更高的 c_1 值会促进粒子向历史最优移动; c_2 反映不同粒子之间的交互, 代表粒子向全局最优靠拢的趋势.

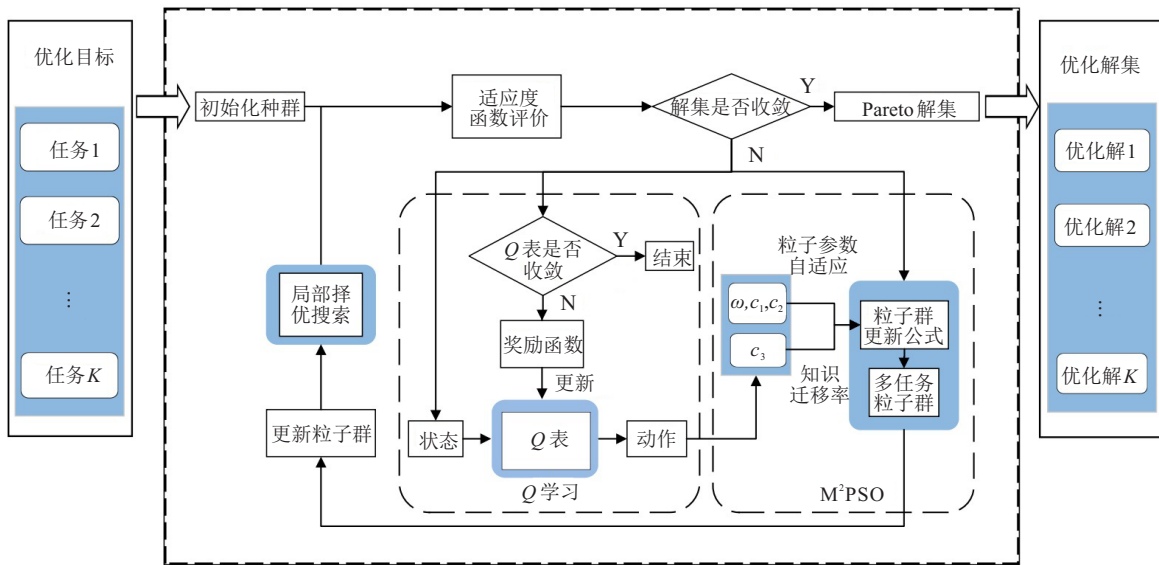


图1 基于Q学习的粒子群优化算法

在知识迁移过程中,知识迁移率 c_3 决定其对粒子速度的影响程度. 然而,迁移知识中存在负向迁移. 为降低负向迁移对优化性能的不利影响,本文构建了基于Q学习的知识迁移策略,根据挖掘的有利知识标记迁移质量,动态调整知识迁移率. 因此,搜索状态可总结为4类: ω 和 c_3 较大时,粒子倾向维持先前速度,在较大解空间移动,有利于全局探索; ω 和 c_1 保持平衡,粒子保持低收敛速度,同时 c_2 保证粒子围绕历史最优探索,有利于局部搜索;提高 c_2 ,粒子向全局最优聚集,降低 c_3 减少不同任务间对收敛的影响,粒子群缓慢收敛; c_2 占主导且 ω 较低后,粒子向全局最优移动,停止任务间知识迁移,减少粒子随机性,有利于快速收敛.

基于以上分析,本文将粒子参数和知识迁移率自适应调整作为Q学习的动作. 根据不同参数变化对优化的影响,将不同参数组合对应于粒子群的4种优化方式作为Q学习的动作,即全局探索、局部开发、缓慢收敛、快速收敛. 这4种优化方式分别对应一组设定好的参数组合,具体如表1所示.

表1 更新参数设定

	粒子群更新参数			知识迁移率
	ω	c_1	c_2	c_3
全局探索	1	2.5	0.5	0.3
局部开发	0.8	2	1	0.2
缓慢收敛	0.6	1	2	0.1
快速收敛	0.4	0.5	2.5	0

2.3 Q学习奖励设置

为加快粒子寻优过程,对粒子设置合理的奖励能够引导粒子向更优的方向搜索. 综合考虑多目标优化问题中的多个目标,将粒子更新前后的适应度差之

和作为奖励值,如下所示:

$$r_{t,t+1} = \sum_{i=1}^M (\text{fitness}_M(x_t) - \text{fitness}_M(x_{t+1})). \quad (8)$$

其中: r 为粒子更新后获得的奖励值, x_t 和 x_{t+1} 分别为更新前后的粒子位置, M 为优化任务的目标函数个数, $\text{fitness}_M(x)$ 为粒子在对应目标函数的适应度标量,是归一化后的目标函数值.

2.4 Q表更新

基于上述Q学习状态、动作和奖励的设计,通过判断当前粒子所处状态选取Q表中该状态对应的最大Q值的动作,依据该动作的优化方式选取参数组合,通过新的参数更新粒子的状态信息,计算更新过程得到的奖励,结合粒子新的状态和动作对Q表进行更新,以跟踪、学习并动态调整优化过程.

Q表采用随机初始化方式,更新如下:

$$Q(S_{t+1}, a_{t+1}) = (1 + a)Q(S_t, a_t) + \alpha[r_{t,t+1} + \gamma \max_a Q(S_{t+1}, a_t)]. \quad (9)$$

其中: α 为学习率, γ 为折扣因子, a_t 为粒子 t 时刻的更新参数, r_{t+1} 为粒子更新后的奖励值, $Q(S_t, a_t)$ 为状态 S_t 采取动作 a_t 的Q值. 为平衡Q表的探索和开发能力,本文设置 $\alpha = 0.01, \gamma = 0.9$.

2.5 多任务下的全局-局部交替搜索策略

2.5.1 全局-局部交替搜索框架

在多任务优化算法搜索过程中,粒子群容易陷入局部最优,使非支配解集缺乏多样性. 为增加种群的多样性,本文设计一种全局-局部交替的搜索策略,平衡算法的全局探索和局部开发能力. 在多任务多目标粒子群优化基础上,对粒子设置跳出局部最优的突

变策略. 首先, 粒子进入全局优化环节后得到非支配解集构成的档案集 A ; 其次, 进入局部搜索阶段对 A 中的所有解进行基于柯西分布的随机突变, 并得到经过突变后粒子的速度和位置; 最后, 对新产生的粒子进行适应度评价, 并与 A 中原有粒子比较排序, 更新 A . 在优化过程中交替进行全局-局部搜索, 对档案集 A 进行更新.

2.5.2 基于柯西分布的局部搜索

为了更充分地进行局部搜索且跳出局部最优, 本文利用柯西分布的特点, 设计一种基于柯西分布的局部搜索. 柯西分布的概率密度函数在垂直分布上较正态分布的概率密度函数更小, 同时在水平方向上逐渐趋缓一直延伸, 适用于探索随机性. 因此根据柯西分布的性质将其与更新函数结合, 在局部搜索中将粒子移动的随机概率引入带有长尾的柯西算子, 使档案集更新后进行基于柯西分布的突变, 增强粒子探索能力, 并有概率跳出局部最优. 全局-局部交替搜索策略提高了粒子的局部搜索和跳出局部最优的能力, 增加了种群的多样性. 基于柯西分布的粒子速度更新公式如下:

$$v_{i,d}(t+1) = \omega v_{i,d}(t) + c_1 r_1 \left(0.5 + \frac{1}{\pi} \arctan(p_{i,d}(t), x_{i,d}(t)) \right). \quad (10)$$

其中: $d = 1, 2, \dots, D$, ω 为惯性权重, c_1 为加速度常量, r_1 是取值范围为 $[0, 1]$ 的随机值.

2.6 基于 Q 学习的多任务多目标粒子群优化算法

本文将粒子群作为强化学习的智能体, 通过粒子对进化情况的判断自适应更新参数和模式, 丰富优化解集, 调整知识迁移. 通过协同 Q 学习方法与粒子群算法, 克服多任务多目标优化中参数更新难、不能自适应、易陷入局部最优、存在负向迁移等问题. 同时, 融入基于柯西分布的局部搜索策略, 平衡全局探索和局部开发, 算法流程如下.

算法 1 QM^2PSO .

输入: 种群规模 n , 进化代数 t_{max} , 初始更新参数 ω, c_1, c_2, c_3 .

step 1: 对粒子群及 Q 表各值进行初始化

step 2: 计算粒子适应度标量、技能因素, 确定粒子优化任务, 进行非支配排序得到 $gbest$

step 3: for $t = 1$ to t_{max} do

step 4: for $p = 1$ to n do

step 5: 计算粒子的状态、动作, 并据此确定更新参数

step 6: 更新粒子速度和位置

step 7: 评价粒子, 更新 $pbest$

step 8: 计算粒子状态奖励, 更新 Q 表

step 9: 交换部分粒子的技能因素

step 10: end for

step 11: 粒子按任务非支配排序存入档案 A

step 12: for p in A

step 13: 进行基于柯西分布的局部搜索

step 14: end for

step 15: 从 A 中得到每个任务的 $gbest$

step 16: end for

step 17: 输出存档 A

3 实验结果与分析

3.1 测试环境

为了检验所提出的算法性能, 通过 python 编程, 在 Windows10 系统 (64 位) 上运行测试, CPU 为 AMD Ryzen 7 5800H (3.20 GHz), 内存为 16 GB.

3.2 评价指标

选取反向世代距离指标 (inverted generational distance, IGD) 评价算法性能, 将每个优化任务 T_i 通过算法优化后所取得的非支配目标向量集合设定为集合 A , 而对应任务在真实 Pareto 前沿 (PF) 上均匀分布的目标向量所构成的集合为 P^* . 将 A 和 P^* 中每个解的目标函数值进行归一化得到 IGD 如下所示:

$$IGD(A, P^*) = \frac{1}{|P^*|} \sum_{x \in P^*} \min_{y \in A} d(x, y), \quad (11)$$

其中 $d(A, P^*)$ 为优化解在目标空间中到真实 Pareto 解集中均匀分布的向量的欧氏距离. 当 $|P^*|$ 的采样足够代表真实 Pareto 前沿时, IGD 值可以通过计算每个函数上的真实点到最近优化点的距离得到进而反映 Pareto 解集的收敛性和多样性. 因此, IGD 值越小表明算法得到的解集与真实 Pareto 解集越接近, 算法在多任务多目标问题上优化性能越好.

3.3 测试问题

本文选取 3 组测试问题, 包括 8 个多目标测试函数, 根据任务间的相似度和最优解取值分成 3 组, 每组任务个数 K 为 2, 每个任务的目标函数 M 为 2 个, 分别为 CIHS、CILS、PIMS、NIHS 测试函数, 每个任务为三维输入变量.

CIHS 函数组中两个任务的目标函数具有很高的相似度和完全重合的解取值范围, 因此在多任务中更容易学习; PIMS 函数组中两个任务的目标函数具有一定的相似性, 解的取值范围部分重合; NIHS 函数组中两个任务的目标函数具有很高的相似性, 但是解

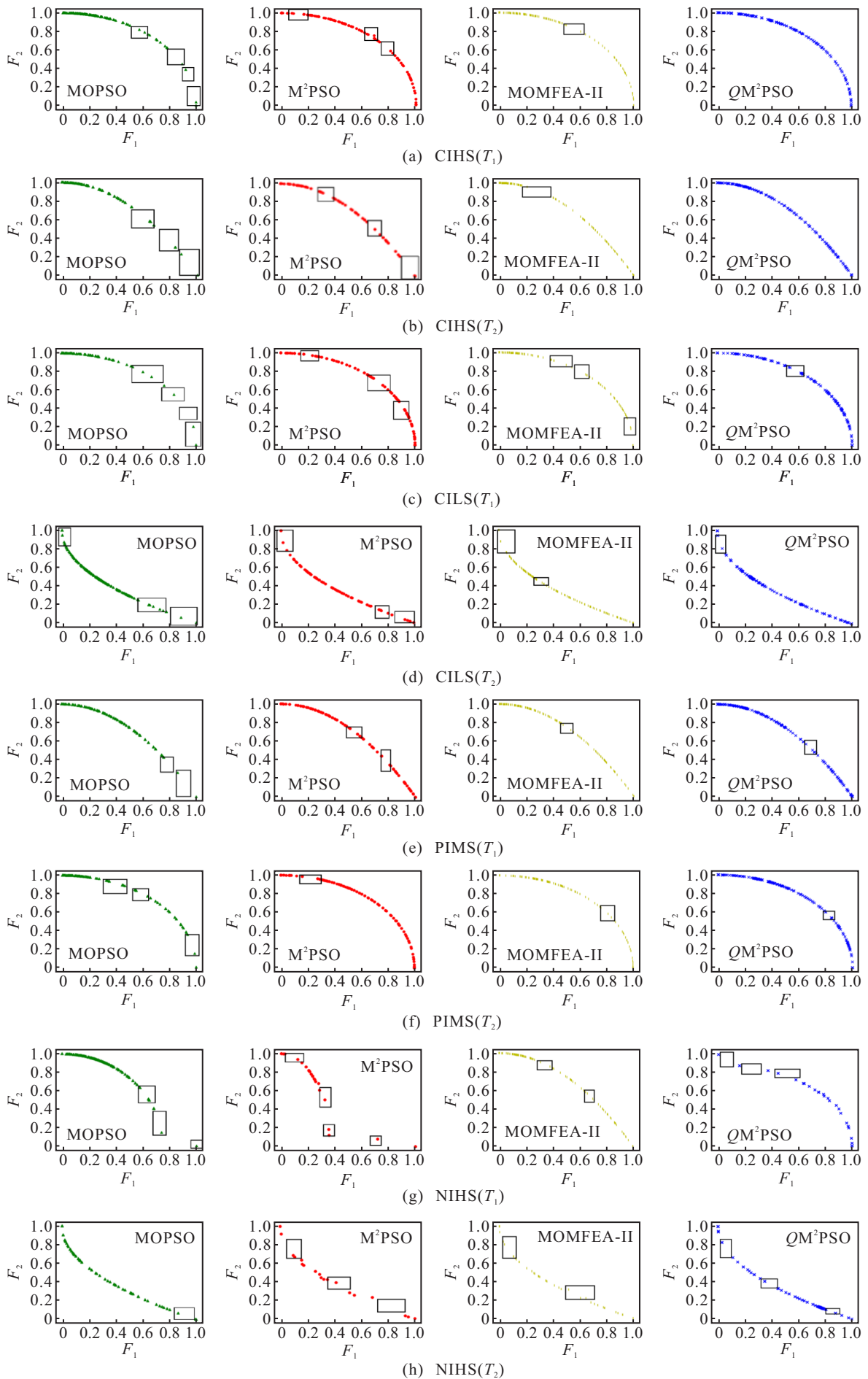


图2 粒子群在最后世代Pareto前沿

的取值完全不重合. 通过以上 3 组任务进行实验可明确反映多任务过程中不同时期探索策略不同的影响, 同时也能反映参数的调整在不同情况下对于迁移知识质量的影响.

3.4 实验结果与分析

为验证所提出 QM²PSO 求解多任务多目标优化问题的有效性, 采用经典的多目标粒子群算法 MOPSO^[43] 和求解多任务问题的 MOMFEA-II^[44] 作为对比算法. 同时, 为验证 Q 学习策略的有效性, 与不包含 Q 学习策略的 M²PSO 进行对比. 为了保证实验的公平性, 3 种算法的种群规模均为 1 000 个粒子, 进化代数为 100 代.

3.4.1 算法所得 Pareto 前沿对比分析

为清晰地比较不同算法的优化效果, 绘制 MOPSO、M²PSO、MOMFEA-II、QM²PSO 四类算法在所有测试问题上的 Pareto 前沿, 如图 2 所示.

由图 2 中的 CIHS 任务可见: QM²PSO 在所得 Pareto 解集分布均匀, 可拟合真实 Pareto 前沿; 而 MOPSO 和 M²PSO 无法拟合真实前沿, 同时存在解偏离真实 Pareto 前沿的情况; MOMFEA-II 所得 Pareto 前沿有不连续情况出现, 表明 QM²PSO 在处理相似度很强、重叠度高的多问题上拥有比 MOPSO、QM²PSO 和 MOMFEA-II 更好的收敛性和多样性.

通过图 2 比较低相似度 CILS 任务的 Pareto 前沿可见: QM²PSO 相比 MOPSO、M²PSO 和 MOMFEA-II 三种算法在拟合 Pareto 前沿上拥有更好的表现. 但相比 CHIS 一组实验, 进行知识迁移的算法不均匀性更加明显. 该组多任务两个目标函数为低相似度, 表明相似度会影响迁移知识, 现有多任务优化方法难以甄别知识的负向迁移. 由图 2 PIMS 任务可见, QM²PSO 和 MOMFEA-II 都能拟合真实 Pareto 前沿, 而 MOPSO 和 M²PSO 对于真实 Pareto 前沿的拟合能力较差, 表明 QM²PSO 和 MOMFEA-II 拥有比 MOPSO 和 M²PSO 更好的收敛性和多样性.

针对 NIHS 函数组, 不同任务中解集的差异会极大影响优化效果. QM²PSO、MOMFEA-II 和 M²PSO 对比 MOPSO, 收敛性和多样性表现均不佳, 表明算法在该算例中存在较大的负迁移现象. 相比具有知识迁移的多任务优化算法 MOMFEA-II 和 M²PSO, QM²PSO 在该任务中效果收敛性更好, 表明算法有抑制负迁移的效果.

3.4.2 IGD 指标分析

为了从数据角度更加精确地分析所提出算法求解多目标多任务优化问题的综合性能, 计算所提出

算法和对比算法的 IGD 指标, 表 2 列出了多次实验下第 100 轮更新后 MOPSO、M²PSO、MOMFEA-II、QM²PSO 四种算法在 4 组多任务总计 8 个目标函数上的 IGD 值.

表 2 各算法在不同任务上获得的 IGD 均值比较

测试函数	任务	MOPSO	M ² PSO	MOMFEA-II	QM ² PSO
CIHS	T ₁	2.232 9e-01	1.003 3e-01	9.138 6e-02	7.525 6e-02
	T ₂	2.983 3e-01	1.624 1e-01	6.374 5e-02	5.387 9e-02
CILS	T ₁	1.594 8e-01	9.882 0e-02	8.935 1e-02	8.784 3e-02
	T ₂	1.466 9e-01	1.486 4e-01	1.188 3e-01	9.820 6e-02
PIMS	T ₁	1.742 4e-01	1.505 3e-01	8.663 7e-02	9.275 3e-02
	T ₂	1.560 9e-01	1.421 1e-01	1.247 9e-01	1.019 2e-01
NIHS	T ₁	1.604 9e-01	1.820 5e-01	9.760 1e-02	8.784 3e-02
	T ₂	1.480 3e-01	2.004 9e-01	1.618 3e-01	1.624 2e-01

由表 2 可见: 在 CIHS 和 CILS 任务组中, QM²PSO 相比另外 3 种算法拥有最小的 IGD 值; 随着不同任务间相似度和解集重合度的下降, 多任务知识迁移方法的 IGD 值逐渐上升; 在 PIMS 任务中, QM²PSO 与 MOMFEA-II 效果相近且优于另外两种算法; 在负迁移影响较大的任务中, M²PSO 的 IGD 值相较 MOPSO 更大; 在低解集重合度的 NIHS 任务组中, QM²PSO 在 T₂ 任务上的 IGD 值相比 MOPSO 和 MOMFEA-II 较大, 但仍然小于 M²PSO. 结果表明, 所提出的 QM²PSO 在具有相近最优的两个任务中能够更好地实现知识提取迁移, 同时, 针对真实 Pareto 解集相差较大的不同任务, 基于 Q 学习的方法能够通过动态调整迁移知识, 有效消除负迁移造成的影响. 因此, 综合图 2 和表 2 的 4 个任务可见, 在相似任务上 QM²PSO 能保持很好的知识迁移能力, 在任务完全相异或全局最优值差异极大的情况下能够在一定程度上甄别迁移的知识, 在同时优化多个任务的同时保证算法的多样性和收敛性.

4 结论

本文设计一种基 Q 学习的多任务多目标粒子群优化算法, 提出基于 Q 学习的参数更新模式, 实现了粒子的参数自适应以及知识迁移率的动态调整, 提高了知识利用率和算法收敛速度. 同时设计全局-局部交替搜索策略, 平衡了算法的全局探索和局部开发能力, 进而提高了算法的优化性能. 实验表明, 基于 Q 学习的参数动态更新策略能够提高种群收敛性和多样性. 在各任务优化过程中, 通过不同粒子根据搜索情况动态调整参数能够避免粒子陷入局部最优. 在多任务协同优化中, 知识迁移率的动态调整可提高知识正向迁移能力, 避免完全不相关任务的负迁移.

下一步将计划研究强化学习与粒子群优化算法

的融合方式,增强多任务优化问题中不同任务的协同优化,研究解决带有不同约束的多任务多目标优化问题的强化学习方法设计.此外,增强该算法的应用研究,如城市污水处理过程中的除磷脱氮为多任务优化问题,同时具有能耗和水质优化的多目标特性,属于多任务多目标优化问题.

参考文献(References)

- [1] 纪昌明, 马皓宇, 李宁宁, 等. 基于树形结构无界存档的多目标粒子群算法[J]. 控制与决策, 2020, 35(11): 2675-2686.
(Ji C M, Ma H Y, Li N N, et al. Multi-objective particle swarm optimization algorithm based on tree-structured unbounded archive[J]. Control and Decision, 2020, 35(11): 2675-2686.)
- [2] Duan H, Li P, Yu Y X. A predator-prey particle swarm optimization approach to multiple UCAV air combat modeled by dynamic game theory[J]. IEEE/CAA Journal of Automatica Sinica, 2015, 2(1): 11-18.
- [3] Seada H, Deb K. A unified evolutionary optimization procedure for single, multiple, and many objectives[J]. IEEE Transactions on Evolutionary Computation, 2016, 20(3): 358-369.
- [4] Gong D, Liu Y, Yen G G. A meta-objective approach for many-objective evolutionary optimization[J]. Evolutionary Computation, 2020, 8(1): 1-25.
- [5] Yuan Y, Ong Y S, Gupta A, et al. Objective reduction in many-objective optimization: Evolutionary multiobjective approaches and comprehensive analysis[J]. IEEE Transactions on Evolutionary Computation, 2018, 22(2): 198-210.
- [6] Cao X Y, Ray Liu K J. Distributed Newton's method for network cost minimization[J]. IEEE Transactions on Automatic Control, 2020, 66(3): 1278-1285.
- [7] 耿焕同, 周山胜, 陈哲, 等. 基于分解的预测型动态多目标粒子群优化算法[J]. 控制与决策, 2019, 34(6): 1307-1318.
(Geng H T, Zhou S S, Chen Z, et al. Decomposition-based predictive dynamic multi-objective particle swarm optimization algorithm[J]. Control and Decision, 2019, 34(6): 1307-1318.)
- [8] Tian Y, Zhang T, Xiao J, et al. A coevolutionary framework for constrained multiobjective optimization problems[J]. IEEE Transactions on Evolutionary Computation, 2020, 25(1): 102-116.
- [9] Zeng S Y, Jiao R W, Li C H, et al. A general framework of dynamic constrained multiobjective evolutionary algorithms for constrained optimization[J]. IEEE Transactions on Cybernetics, 2017, 47(9): 2678-2688.
- [10] Chen Z G, Zhan Z H, Wang H, et al. Distributed individuals for multiple peaks: A novel differential evolution for multimodal optimization problems[J]. IEEE Transactions on Evolutionary Computation, 2019, 24(4): 708-719.
- [11] Gong M G, Tang Z D, Li H, et al. Evolutionary multitasking with dynamic resource allocating strategy[J]. IEEE Transactions on Evolutionary Computation, 2019, 23(5): 858-869.
- [12] 张恒, 何丽, 袁亮, 等. 基于改进双层蚁群算法的移动机器人路径规划[J]. 控制与决策, 2022, 37(2): 303-313.
(Zhang H, He L, Yuan L, et al. Mobile robot path planning using improved double-layer ant colony algorithm[J]. Control and Decision, 2022, 37(2): 303-313.)
- [13] Dai C, Wang Y. A new multi objective evolutionary algorithm based on decomposition of the objectives pace for multiobjective optimization[J]. Journal of Applied Mathematics, 2014, 2014(1): 1-9.
- [14] Cheng Y F, Shao W, Zhang S J, et al. An improved multi-objective genetic algorithm for large planar array thinning[J]. IEEE Transactions on Magnetics, 2016, 52(3): 1-4.
- [15] Liu R, Jin Y, Jiao L et al. A self-adaptive response strategy for dynamic multiobjective evolutionary optimization based on objective space decomposition[J]. Evolutionary Computing, 2021, 29(4): 491-519.
- [16] Zheng X L, Qin A K, Gong M G, et al. Self-regulated evolutionary multitask optimization[J]. IEEE Transactions on Evolutionary Computation, 2020, 24(1): 16-28.
- [17] Nava G, Sablé Q, Tognon M, et al. Direct force feedback control and online multi-task optimization for aerial manipulators[J]. IEEE Robotics and Automation Letters, 2020, 5(2): 331-338.
- [18] Yao S S, Dong Z M, Wang X P, et al. A Multi objective multi factorial optimization algorithm based on decomposition and dynamic resource allocation strategy[J]. Information Sciences, 2020, 511: 18-35.
- [19] Huang S J, Zhong J H, Yu W J. Surrogate-assisted evolutionary framework with adaptive knowledge transfer for multi-task optimization[J]. IEEE Transactions on Emerging Topics in Computing, 2021, 9(4): 1930-1944.
- [20] Ma X L, Yin H, Zhu A M, et al. Enhanced multi factorial evolutionary algorithm with meme helper-tasks[J]. IEEE Transactions on Cybernetics, 2022, 52(8): 7837-7851.
- [21] Mao T, Mihăită A S, Chen F, et al. Boosted genetic algorithm using machine learning for traffic control optimization[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(7): 7112-7141.
- [22] 李俊, 周虎, 李波. 基于虚拟蚂蚁的局部优化蚁群算法[J]. 控制与决策, 2019, 34(11): 2459-2468.
(Li J, Zhou H, Li B. Local optimization ACO based on virtual ant colony algorithm[J]. Control and Decision, 2019, 34(11): 2459-2468.)
- [23] Zhao C, Guo D. Particle swarm optimization algorithm with self-organizing mapping for nash equilibrium strategy in application of multiobjective optimization[J]. IEEE Transactions on Neural Networks and Learning

- Systems, 2021, 32(11): 5179-5193.
- [24] Tian J, Tan Y, Zeng J C, et al. Multiobjective infill criterion driven Gaussian process-assisted particle swarm optimization of high-dimensional expensive problems[J]. IEEE Transactions on Evolutionary Computation, 2019, 23(3): 459-472.
- [25] 张强, 李盼池. 一种自适应多策略行为粒子群优化算法[J]. 控制与决策, 2020, 35(1): 115-122.
(Zhang Q, Li P C. An adaptive multi-strategy behavior particle swarm optimization algorithm[J]. Control and Decision, 2020, 35(1): 115-122.)
- [26] Han H G, Lu W, Zhang L, et al. Adaptive gradient multiobjective particle swarm optimization[J]. IEEE Transactions on Cybernetics, 2018, 48(11): 3067-3079.
- [27] Han H G, Lu W, Qiao J F. An adaptive multiobjective particle swarm optimization based on multiple adaptive methods[J]. IEEE Transactions on Cybernetics, 2017, 47(9): 2754-2767.
- [28] Hu W, Yen G G. Adaptive multi objective particle swarm optimization based on parallelcell coordinate system[J]. IEEE Transactions on Evolutionary Computation, 2015, 19(1): 1-18.
- [29] Li L, Chang L, Gu T, et al. On the norm of dominant difference for many-objective particle swarm optimization[J]. IEEE Transactions on Cybernetics, 2021, 51(4): 2055-2067.
- [30] Bonyadi M R, Michalewicz Z. Analysis of stability, local convergence, and transformation sensitivity of a variant of the particle swarm optimization algorithm[J]. IEEE Transactions on Evolutionary Computation, 2016, 20(3): 370-385.
- [31] 唐可心, 梁晓磊, 周文峰, 等. 具有重组学习和混合变异的动态多种群粒子群优化算法[J]. 控制与决策, 2021, 36(12): 2871-2880.
(Tang K X, Liang X L, Zhou W F, et al. Dynamic multi-population particle swarm optimization algorithm with recombined learning and hybrid mutation[J]. Control and Decision, 2021, 36(12): 2871-2880.)
- [32] Hernandez D, Rodriguez R, Merchán E, et al. Optimal design of a drive shaft with composite materials through particle swarm optimization[J]. IEEE Latin America Transactions, 2020, 18(6): 1008-1016.
- [33] Han H G, Bai X, Hou Y, et al. Self-adjusting multitask particle swarm optimization[J]. IEEE Transactions on Evolutionary Computation, 2022, 26(1): 145-158.
- [34] Tang Z, Gong M, Xie Z, et al. Multi-task particle swarm optimization with dynamic neighbor level-based inter-task learning[J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2022, 6(2): 300-314.
- [35] 陈宗海, 杨志华, 王海波, 等. 从知识的表达和运用综述强化学习研究[J]. 控制与决策, 2008, 23(9): 961-968.
(Chen Z H, Yang Z H, Wang H B, et al. Overview of reinforcement learning from knowledge expression and handling[J]. Control and Decision, 2008, 23(9): 961-968.)
- [36] Lu L, Zheng H, Jie J, et al. Reinforcement learning-based particle swarm optimization for sewage treatment control[J]. Complex & Intelligent Systems, 2021, 7(5): 2199-2210.
- [37] 杨思明, 单征, 丁煜, 等. 深度强化学习研究综述[J]. 计算机工程, 2021, 47(12): 19-29.
(Yang S M, Shan Z, Ding Y, et al. Survey of research on deep reinforcement learning[J]. Computer Engineering, 2021, 47(12): 19-29.)
- [38] Chalmers E, Contreras E B, Robertson B, et al. Learning to predict consequences as a method of knowledge transfer in reinforcement learning[J]. IEEE Transactions on Neural Networks and Learning Systems, 2018, 29(6): 2259-2270.
- [39] 朱美强, 李明, 程玉虎, 等. 基于拉普拉斯特征映射的启发式Q学习[J]. 控制与决策, 2014, 29(3): 425-430.
(Zhu M Q, Li M, Cheng Y H, et al. Heuristically accelerated Q-learning algorithm based on Laplacian Eigenmap[J]. Control and Decision, 2014, 29(3): 425-430.)
- [40] Xu C, Zhao W, Li L, et al. A nash Q-learning based motion decision algorithm with considering interaction to traffic participants[J]. IEEE Transactions on Vehicular Technology, 2020, 69(11): 12621-12634.
- [41] Palanisamy M, Modares H, Lewis F L, et al. Continuous-time Q-learning for infinite-horizon discounted cost linear quadratic regulator problems[J]. IEEE Transactions on Cybernetics, 2015, 45(2): 165-176.
- [42] Blouin S, Babahaji M, Mahboubi H, et al. Estimation of the connectivity of random graphs through Q-learning techniques[J]. IEEE Journal of Radio Frequency Identification, 2022, 6(1): 318-331.
- [43] Coello C A C, Pulido G T, Lechuga M S. Handling multiple objectives with particle swarm optimization[J]. IEEE Transactions on Evolutionary Computation, 2004, 8(3): 256-279.
- [44] Bali K K, Gupta A, Ong Y S, et al. Cognizant multitasking in multiobjective multifactorial evolution: MO-MFEA-II[J]. IEEE Transactions on Cybernetics, 2021, 51(4): 1784-1796.

作者简介

韩红桂(1983—), 男, 教授, 博士生导师, 从事神经网络、过程控制、知识工程与数据挖掘等研究, E-mail: rechardhan@bjut.edu.cn;

徐子昂(2000—), 男, 博士生, 从事污水处理离散调度优化方法与应用的研究, E-mail: xuziang316@126.com;

王晶晶(1995—), 女, 博士后, 从事智能优化调度方法与应用的研究, E-mail: wangjingjing@bjut.edu.cn.