

多环境选择策略引导的约束多目标进化优化算法

左明成¹, 薛 源¹, 巩敦卫^{2*}, 薛天洋¹

(1. 中国矿业大学计算机科学与技术学院/人工智能学院, 江苏徐州 221116; 2. 青岛科技大学微电子学院, 山东青岛 266061)

摘 要: 约束多目标优化问题(Constrained Multi-objective Optimization Problems, CMOPs)广泛存在于实际工程与科学领域,这类问题通常同时包含多个相互冲突的优化目标与复杂多样的约束条件。在约束多目标进化算法(Constrained Multi-Objective Evolutionary Algorithms, CMOEAs)中,环境选择策略(Environmental Selection Strategy, ESS)的设计与集成至关重要,它直接决定了算法能否在可行性(满足约束条件的程度)、收敛性(逼近帕累托前沿的速度)和多样性(解集的分布均匀性)三者之间取得有效平衡。然而,现有方法多依赖静态或预定义的策略调度机制,难以根据种群在进化过程中的实时状态自适应调整选择压力,导致在应对复杂约束、非连续可行域及昂贵目标评估等实际场景时,算法往往表现出明显的局限性:一方面,在进化初期种群完全不可行的情况下,缺乏足够的驱动力引导搜索快速进入可行域;另一方面,在找到可行解之后,难以动态地在收敛加速与多样性维持之间进行合理权衡,容易出现早熟收敛或多样性丧失,甚至在某些极端约束条件下长期无法找到任何可行解。这些问题严重限制了CMOEAs的整体性能与鲁棒性。鉴于此,本文提出一种基于强化学习(Reinforcement Learning, RL)的多环境选择策略集成方法,增强约束多目标进化算法的自主决策与动态适应能力。该方法首先根据种群在决策空间与目标空间的分布特征,从可行性、收敛性和多样性三个维度刻画进化状态;利用 ε -可行解比例反映种群满足约束的整体水平,利用归一化质心位移表征收敛趋势,利用 ε -可行解方差在种群总方差中的占比评估多样性。进而,受物理学中相互作用力原理的启发,本文设计了三种各具侧重且内在耦合的环境选择策略:第一种策略着重强化可行性保持,通过引入约束违背度排名构建排斥力模型,优先筛选可行解;第二种策略聚焦收敛加速,采用目标函数排名引导种群向帕累托前沿逼近;第三种策略则致力于多样性维持,仅依据个体间的空间距离产生排斥力,避免种群过度聚集。这三种策略相互耦合,即使某一次策略选择出现偏差也不会导致算法性能急剧恶化,从而为强化学习提供了稳定的决策基础。最后,通过强化学习构建种群状态与环境选择策略之间的映射模型。在三个基准测试集(LIR-CMOP、DASCMP、DOC)与模拟集成电路优化设计问题上的实验表明,本文方法在大多数测试问题上显著优于六种先进的约束多目标进化算法,尤其在可行域狭窄、约束复杂、帕累托前沿不连续的困难场景下表现出更强的鲁棒性与适应性,能够以更少的函数评价次数获得更高质量的解集,从而充分验证了其在处理实际复杂约束多目标优化问题中的有效性、可扩展性与工程应用价值。

关键词: 约束多目标优化;多环境选择策略;进化优化;强化学习;模拟集成电路

基金项目: 国家自然科学基金(No.62303465, No.62133015);山东省自然科学基金(No.ZR2023ZD06)

中图分类号: TN710.2; TP182

文献标识码: A

文章编号: 0372-2112(2026)06-2819-18

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20260026

A Constrained Multi-Objective Evolutionary Algorithm Guided by Multi-Environment Selection Strategies

ZUO Mingcheng¹, XUE Yuan¹, GONG Dunwei^{2*}, XUE Tianyang¹

(1. School of Computer Science and Technology / School of Artificial Intelligence, China University of Mining and Technology, Xuzhou, Jiangsu 221116, China; 2. School of Microelectronics, Qingdao University of Science and Technology, Qingdao, Shandong 266061, China)

Abstract: Constrained multi-objective optimization problems (CMOPs) are widely prevalent in practical engineering and scientific fields. Such problems typically involve multiple conflicting optimization objectives and a variety of complex constraints. In constrained multi-objective evolutionary algorithms (CMOEAs), the design and integration of the environmental selection strategy (ESS) are crucial, as it directly determines whether the algorithm can achieve an effective balance among feasibility (degree of constraint satisfaction), convergence (speed of approaching the Pareto front), and diversity (uniformity of solution set distribution). However, existing methods mostly rely on static or predefined strategy scheduling mechanisms, making it difficult to adaptively adjust selection pressure according to the real-time state of the population during evolution. Consequently, when dealing with practical scenarios such as complex constraints, discontinuous feasible regions, and expensive objective evaluations, these algorithms exhibit significant limitations. On one hand, when the popula-

tion is entirely infeasible in the early stages of evolution, there is insufficient driving force to guide the search quickly into the feasible region. On the other hand, after feasible solutions are found, it becomes challenging to dynamically achieve a reasonable trade-off between convergence acceleration and diversity maintenance, often leading to premature convergence or loss of diversity. In some extreme constraint conditions, the algorithm may even fail to find any feasible solution for a long time. These issues severely constrain the overall performance and robustness of CMOEAs. In view of this, this paper proposes a multi-environmental selection strategy integration method based on reinforcement learning (RL) to enhance the autonomous decision-making and dynamic adaptability of constrained multi-objective evolutionary algorithms. The proposed method first characterizes the evolutionary state from three dimensions (feasibility, convergence, and diversity) based on the distribution characteristics of the population in both decision space and objective space. The ε -feasible solution ratio is used to reflect the overall level of constraint satisfaction of the population; the normalized centroid displacement is used to indicate the convergence trend; and the proportion of the variance of ε -feasible solutions in the total population variance is used to evaluate diversity. Then, inspired by the principle of interaction forces in physics, this paper designs three distinct yet inherently coupled environmental selection strategies. The first strategy focuses on enhancing feasibility preservation, constructing a repulsion force model by ranking constraint violation degrees to prioritize feasible solutions. The second strategy focuses on accelerating convergence, using objective function ranking to guide the population toward the Pareto front. The third strategy is dedicated to diversity maintenance, generating repulsion forces solely based on spatial distances among individuals to avoid excessive clustering. These three strategies are coupled with each other, such that even if a suboptimal strategy is selected at a certain step, it does not cause a sharp deterioration in algorithm performance, thereby providing a stable decision-making foundation for reinforcement learning. Finally, a mapping model between population states and environmental selection strategies is constructed through reinforcement learning. Experimental results on three benchmark test suites (LIR-CMOP, DASCMP, DOC) and an analog integrated circuit optimization design problem show that the proposed method significantly outperforms six state-of-the-art constrained multi-objective evolutionary algorithms on most test problems. It exhibits stronger robustness and adaptability, particularly in challenging scenarios with narrow feasible regions, complex constraints, and discontinuous Pareto fronts. It obtains higher-quality solution sets with fewer function evaluations, thereby fully validating its effectiveness, scalability, and engineering application value in handling practical complex constrained multi-objective optimization problems.

Keywords: constrained multi-objective optimization; multi-environment selection strategies; evolutionary optimization; reinforcement learning; analog integrated circuits

Foundation Item(s): National Natural Science Foundation of China (No.62303465, No.62133015); National Natural Science Foundation of Shandong Province (No.ZR2023ZD06)

0 引言

约束多目标优化问题 (Constrained Multi-objective Optimization Problems, CMOPs) 是指在满足一组约束条件的前提下, 同时优化多个相互冲突目标函数的数学问题。现实中许多问题都可以被归类为 CMOPs, 如综合能源调度^[1]和航天器轨道优化^[2]。一般情况下, 一个 CMOP 可定义为

$$\begin{aligned} \text{Min } F(x) &= (f_1(x), f_2(x), \dots, f_m(x)) \\ \text{s.t. } g_i(x) &\leq 0, i = 1, 2, \dots, l \\ h_j(x) &= 0, j = l+1, l+2, \dots, n \\ x &= (x_1, x_2, \dots, x_D) \in \mathbb{R}^D \end{aligned} \quad (1)$$

其中: x 为 D 维决策变量, $f_k(x)$ 为第 k 个目标函数, $k=1, 2, \dots, m$; $g_i(x) \leq 0$ 为第 i 个不等式约束, $i=1, 2, \dots, l$; $h_j(x)=0$ 为第 j 个等式约束, $j=l+1, l+2, \dots, n$, n 为约束个数。在实际应用中, 约束多目标优化问题的求解面临诸多挑战, 如目标间的冲突性、约束条件的非线性

与非凸性、可行域的离散性等。传统的基于梯度的优化方法在处理此类问题时, 往往容易陷入局部最优, 且对问题模型的可微性有严格要求。相比之下, 进化算法因其种群并行搜索、无需梯度信息、对问题模型假设较弱等优势, 已成为求解复杂 CMOPs 的主流范式。

约束多目标进化算法 (Constrained Multi-Objective Evolutionary Algorithms, CMOEAs) 通过模拟生物进化过程, 在决策空间中维持一个种群 (一组候选解的集合), 并利用选择、交叉、变异等操作迭代逼近帕累托最优解集 (指在多目标优化中, 无法再改进任一目标而不损害其他目标的解集合), 尤其擅长处理黑箱、高维、多模态的优化问题。其中的种群环境选择阶段需要从多个方面评价当前候选解, 并选择合适的候选解形成下一代种群。由此可见, 环境选择策略的设计至关重要, 它直接影响种群在可行性 (解满足所有约束条件的程度)、收敛性 (解逼近真实帕累托前沿的

能力)和多样性(解在目标空间或决策空间中的分布均匀性)之间的平衡能力。现有研究也极为重视环境选择策略中对种群可行性、收敛性和多样性之间的平衡。Raju 等人^[3]在 CMOEA 中集成了三种环境选择策略,分别用于搜索无约束帕累托前沿、收敛性解和多样性解。Ming 等人^[4]采用多样化的候选解排序策略,从不同角度选择有益于平衡可行性、收敛性和多样性的解。然而,现有环境选择策略的集成与调用机制仍存在显著局限:多数方法仅能在单一策略框架内,通过经验模型对可行性、收敛性与多样性进行线性或静态的权重调整,其调控幅度与适应能力较为有限;另一些方法虽尝试引入多策略,但多依赖简单概率模型或固定阶段的策略切换,缺乏对种群状态的精细感知与动态响应,导致调控精度低、鲁棒性不足。这些设计上的缺陷,严重制约了 CMOEAs 在不同进化阶段与问题特性下自主、灵活地选择合适策略的能力,从而影响算法整体求解性能。因此,针对可行性、收敛性和多样性分别设计定制化的高效环境选择策略,并建立能够动态响应种群状态的智能映射模型,以实现环境选择策略的自主、精准调度,已成为提升 CMOEAs 综合性能的关键研究方向。

近年来,强化学习(Reinforcement Learning, RL)在自主决策方面展现出了显著的优势。RL 能够根据系统的当前运行状态动态调整控制策略,即便在复杂场景中也能实现高效的控制决策。这一特性为解决 CMOPs 中环境选择策略的集成和自主选择难题提供了一种新方法,有助于自主选择最适合当前进化状态的环境选择策略,最终设计出一套能够动态平衡收敛性、多样性和可行性的环境选择策略。鉴于此,本文结合强化学习与 CMOEAs,提出了一种强化学习引导的 CMOEAs 多环境选择策略。该方法首先根据种群在决策空间和目标空间中的分布来表征其可行性、收敛性和多样性。然后,在环境选择策略的设计中,本文受物理学中粒子相互作用力原理的启发,构建了一种结构化的选择压力动态调控机制。该机制将种群中的每个个体视为一个粒子,其在目标空间与决策空间中的位置表征了性能与分布特性。个体间的作用力由两部分协同决定:分子项引入了约束违背度或目标函数的排序信息,用以定向引导算法对可行性或收敛性的偏好;分母项则基于个体间的空间距离,用以维持分布多样性。每个个体所承受的合力综合反映了其在当前偏好下的生存压力:合力值越大,表明该个体越容易被排斥,从而在环境选择中被剔除的优先级越高。通过分别设计侧重可行性、收敛性与多样性的三种作用力公式,本方法建立了三种具有内在耦合性的环境选择策略。进而,通过强化学习对种群进化状

态的感知,动态激活最适宜的作用力准则,从而自主、精准地调控种群在不同进化阶段的选择压力方向与强度,最终实现可行性、收敛性与多样性三者进化过程中的动态平衡与协同优化。最后,利用强化学习构建种群状态与环境选择策略之间的映射模型,从而为种群进化动态选择高效的环境选择策略。

本文的主要贡献体现在如下 4 个方面:

(1)基于相互作用力原理设计了三种高效的环境选择策略。本文针对现有约束多目标优化算法在环境选择过程中平衡种群可行性、收敛性和多样性的需求,基于相互作用力原理提出了三种环境选择策略。通过设计三种分别偏好可行性、收敛性和多样性的策略选择机制,动态调整种群个体的选择压力。所提三种环境选择策略既存在互补性,又存在一定耦合性,在保证环境选择策略调整效果的同时,缓解了因策略选择错误导致的性能不稳定性。

(2)利用强化学习简易集成多种环境选择策略。为了提高环境选择策略的自主性,多角度感知和识别种群进化状态,通过强化学习构建进化状态和环境选择策略的匹配性映射模型。由于环境选择策略之间的耦合性,本文方法对训练数据的数量要求较低,缓解了在线优化过程中的模型构建压力。该方法对进化状态具有较高的识别能力,能显著提高环境选择策略的推荐精度,保障了算法的鲁棒性和解的高质量。

(3)采用种群存档隔代保留被淘汰的潜在优秀个体。为了避免携带关键基因特征,尤其满足关键约束、具备全局探索能力和局部开采能力的优秀个体在当前环境选择策略下被淘汰,设计了种群隔代存档机制,保留在环境选择过程中被淘汰的潜在优秀个体,通过参与后续的种群进化,进一步丰富了种群的选择压力。

(4)面向仿真高成本与性能严苛权衡的自主决策机制。通过强化学习模型,将有限的仿真次数智能地导向当前最关键的优化方向。在电路基本功能未满足时优先保障设计可行性,在功能实现后聚焦于突破性能瓶颈,在陷入局部最优时主动探索替代性电路结构。这一机制从本质上将算法层的自适应决策与模拟集成电路设计特有的成本约束和物理约束深度融合,最终在严格受限的仿真预算内,实现了对高性能、可制造电路方案的可靠与高效自动化搜索。

1 相关工作

本文主要通过结合强化学习和 CMOEAs,设计先进的环境选择策略解决 CMOPs。因此,本章主要回顾强化学习引导的约束多目标进化优化算法,平衡可行性、收敛性和多样性的环境选择策略两方面工作,

并讨论了当前方法的优缺点。

1.1 强化学习引导的约束多目标进化优化算法

该方法利用强化学习的自主决策能力,动态调整进化算法的求解策略,提高算法的效率和鲁棒性。采用强化学习辅助的进化算法的一般思路是:(1)类比于智能体状态,提取种群的关键信息作为进化状态;(2)类比于智能体动作,设计引导种群进化的多种策略;(3)类比于动作奖励,设计进化策略的评价方法。由此可见,该方法的性能依赖于进化状态、进化策略、评价方法的设计方式。Tan 等人^[5]提出 RL-HPSDE 方法,通过结合强化学习与差分进化算法,自适应选择最优的突变策略和控制参数。Ming 等人^[6]提出的基于深度强化学习的操作算子选择框架,能够自适应选择最适合的操作算子,显著提升 CMOEAs 的种群生成质量。Yang 等人^[7]提出了一种深度强化学习辅助的 CMOEAs (CMONS-DQN),用于解决 CMOPs。该方法通过引入 Voronoi 区域的多样性维护机制和深度 Q 值网络来选择适合的 Voronoi (沃罗诺伊)区域生成后代,从而在复杂约束条件下实现了对搜索空间的全面探索和高性能解的快速收敛。Wang 等人^[8]提出了一种基于深度强化学习的自适应约束多目标优化算法 (CMOEA-DRL),通过深度 Q 值学习网络动态选择适合不同进化阶段的约束处理技术和操作算子,用于解决复杂的 CMOPs。Zuo 等人^[9]提出了一种基于过程知识引导的自适应进化优化方法 (Process Knowledge-guided Autonomous Evolutionary Optimization, PKAEO),通过深度 Q 值学习网络动态选择适合不同进化阶段的引导策略,用于解决 CMOEAs。上述论文普遍采用 Q 值学习网络引导进化过程的主要原因包括:状态-动作空间的离散性、样本效率与在线学习的需求、策略耦合性的高效利用和与进化过程的自然类比。

总体来看,现有基于强化学习的约束多目标进化优化研究主要聚焦于进化算子/策略选择、参数自适应调整、约束处理机制动态调度以及决策空间搜索引导四大核心方向。多数方法采用 Q 学习或深度 Q 网络作为基本框架,因其能有效处理离散决策、适应在线学习需求,且与进化过程形成自然对应,从而在提升算法效率、增强对复杂问题的鲁棒性方面发挥了重要作用。未来研究可进一步探索更精细的进化状态表示、多智能体协同决策以及面向大规模约束问题的可扩展强化学习机制。

1.2 平衡可行性、收敛性和多样性的环境选择策略

在 CMOEAs 中,单独追求可行性、收敛性或多样性都可能导致算法性能的下降。例如,过度追求可行性可能导致种群无法有效收敛到帕累托前沿;过度追

求收敛性可能导致种群多样性不足,无法覆盖整个帕累托前沿;过度追求多样性可能导致种群中包含大量不可行解,降低解的质量。因此,平衡这三个属性是提升算法综合性能的关键。采用多种不同的环境选择方式选择种群个体,从而平衡种群的可行性、收敛性和多样性是一种策略,其核心思想是通过多样化的选择机制来提高算法的全局搜索能力、解的质量和鲁棒性。不同的环境选择方式侧重于不同的搜索方向,从而在解空间中更广泛地搜寻潜在的优秀个体。Tang 等人^[10]提出了一种种群状态引导的环境选择策略自主调整方法。当种群在可行域内收敛时,采用无约束环境选择,允许种群跨越当前可行域进行探索;当种群在可行域外收敛时,采用约束环境选择,引导种群回到可行域。Zhang 等人^[11]将种群动态划分为三个互斥的子集,每个子集采用不同的环境选择策略,在求解过程中兼顾约束满足和目标优化。Sun 等人^[12]为两个子种群分别配置了动态约束松弛策略和自适应协同进化策略,提供了多样化的环境选择策略。Shi 等人^[13]提出了一种包含快速收敛阶段、多样性维持阶段和极值挖掘阶段的三阶段多种群进化算法 (CMOEA-TMC),结合主种群和辅助种群的协同进化,通过多阶段的环境选择策略,有效平衡了可行性和多样性。Ming 等人^[14]通过兼顾目标空间和决策空间的多种环境选择策略,实现了收敛性与多样性的平衡。Ming 等人^[15]还提出了能够处理不同类型 CMOPs 的定制化环境选择策略。

总体来看,现有研究通过多种环境选择策略的引入和组合,实现了在收敛性、可行性和多样性之间的有效平衡。这些方法主要包括多阶段策略调整、种群动态划分、约束松弛技术以及多空间协同选择等核心思路。研究表明,通过在不同进化阶段或种群分布状态下适配差异化的环境选择机制,能够更灵活地引导搜索方向,从而在复杂约束条件下维持搜索效率和解集质量。未来研究可进一步关注环境选择与强化学习等动态决策方法的结合,实现更智能、更自适应的平衡机制。

1.3 讨论

综合上述分析,现有研究在利用强化学习辅助 CMOEAs 时,多聚焦于进化策略或约束处理技术的选择,而对多种环境选择策略的动态集成关注不足。环境选择策略直接决定了种群进化的方向,其选择相较于进化算子更具全局性和导向性。现有方法若将 RL 应用于环境选择,常面临两大挑战:第一,若所集成的各环境选择策略差异过大(耦合性低),则 RL 模型需要海量的在线训练数据才能准确决策,这在高代价的实际问题中难以实现;第二,若 RL 决策出现偏差,

选择了严重偏离当前进化需求的策略,可能导致种群质量急剧下降,算法稳定性变差。因此,如何设计一组既各有侧重又保持适度耦合的环境选择策略,并构建一个能够以少量样本快速学习、稳健决策的RL集成框架,成为提升CMOEAs性能的关键。这正是本文拟解决的核心问题。

鉴于此,本文提出强化学习引导的多环境选择策略约束多目标进化优化方法。通过设计多样性且具有耦合性的环境选择策略,满足了种群进化对可行性、收敛性和多样性的需求;采用强化学习集成和自主选择多种环境策略,实现种群的高效进化。

2 本文方法

2.1 整体框架

本章提出一种强化学习集成多环境选择策略的CMOEAs (Constrained Multi-Objective Evolutionary Algorithm with reinforcement learning-integrated Multi-Environment Selection Strategy, CMOEA-MESS)。该方法的示意图如图1所示,具体思想是:类比于强化学习的智能体状态,根据种群 P_t 在决策空间及目标空间的分布特征,刻画种群可行性、收敛性和多样性的进化状态;类比于强化学习的智能体动作,基于相互作用力原理设计兼顾种群可行性、收敛性和多样性的三种环境选择策略 a_t 。在进化前期,每一代种群状态下随机执行环境选择策略,生成下一代种群 P_{t+1} ,得到环境选择策略的性能评价 r_t ,形成包含种群状态、环境选择策略和策略评价值的训练数据,用于更新Q值学习的数据QT和动作选择次数的数据CT;在进化后期,采用强化学习构建种群状态与环境选择策略的映射模型,根据每一代的种群状态,推荐最适合的环境选择策略,同时继续更新QT和CT。该方法能针对不同的求解问题,通过强化学习推荐的环境选择策略平衡种群的可行性、收敛性和多样性,具有较高的问题普适性和扩展性。更重要的是,该方法能灵活嵌入当前的CMOEAs。

本文方法的整体框架如算法1所示。针对现有的CMOEAs,本文方法只需在算法主体中嵌入环境选择策略(第10行和第13行)和强化学习模型训练(第14行)两个步骤。具体而言,首先初始化强化学习所需的相关参数,以及用于保存Q值的QT和动作选择次数的CT(第1~7行);其次,生成交配池 M_t ,在算法2中获取环境选择策略决策 a_t ,生成子代种群并将其与父代种群合并为 R_t (第9~12行);最后,在算法3中执行 a_t 对应的环境选择策略,按照算法4更新QT和CT(第13~14行)。其中,环境选择策略决策(Environ-

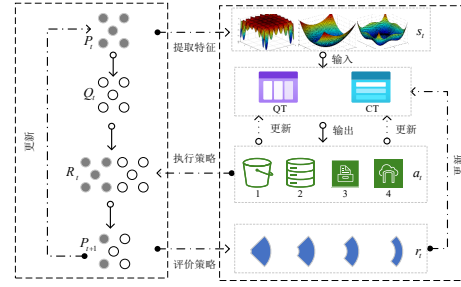


图1 整体框架示意图

Figure 1 Schematic diagram of the overall framework

ment Selection Strategy Decision, ESSD)用来选取环境选择策略,环境选择策略执行(Environment Selection Strategy Implementation, ESSI)用于执行环境选择策略。上述两者的具体内容与QT、CT的更新方法如3.2节、3.3节和3.4节所示。

算法1 整体架构

输入: 最大进化代数 $T_{\max}$; 问题维度 D ; 种群规模 N ; 决策变量下界 LB; 决策变量上界 UB

输出: $P_{T_{\max}}$

1. $t \leftarrow 1$
2. $P_t \leftarrow \text{InitializePopulation}(N, D, LB, UB)$
3. 初始化 s_t 和 a_t
4. 初始化 Q-learning 模型, QT 和 CT
5. 初始化可行性档案 A_1 、收敛性档案 A_2 和多样性档案 A_3
6. 确定约束松弛参数 ε
7. $C_t \leftarrow 0$
8. WHILE $t \leq T_{\max}$ DO
9. 生成交配池 M_t
10. $a_t \leftarrow \text{ESSD}(t, QT, CT)$ //算法2
11. 从 M_t 中生成过渡种群 O_t
12. $R_t \leftarrow O_t \cup P_t$
13. $(P_{t+1}, A_1, A_2, A_3) \leftarrow \text{ESSI}(R_t, a_t, A_1, A_2, A_3)$ //算法3
14. $(QT, CT) \leftarrow \text{Q-Update}(s_t, a_t, P_t, P_{t+1}, QT, CT)$ //算法4
15. $t \leftarrow t + 1$
16. END WHILE
17. RETURN $P_{T_{\max}}$

2.2 环境选择策略决策

为解决种群进化初期因训练样本不足导致QT无法提供有效指导的问题,本文将种群进化过程划分为两个阶段:强化学习预热阶段和强化学习应用阶段。具体而言,从第1代到第 T_{sam} 代为强化学习预热阶段,主要为QT积累足够的训练样本(第1~2行);从第 $T_{\text{sam}} + 1$ 代到种群进化结束为强化学习应用阶段,利用训练好的QT指导环境选择策略的决策(第3~5行)。

具体方法如算法 2 所示。

算法 2 环境选择策略决策 ESSD

输入: t , QT, CT

输出: a_t

1. IF $t \leq T_{\text{sam}}$ THEN
2. 随机选择一个策略 a_t
3. ELSE
4. $r'_t = \text{QT}(s_t, a_t) \text{CT}(s_t, a_t)$
5. 根据 $a_t^{\max} = \arg \max \{r'_t\}$ 确定 a_t
6. END IF
7. IF Rand $< \epsilon$ THEN
8. 随机选择一个策略 a_t
9. END IF

在预热阶段,由于 QT 尚未充分学习,环境选择策略 a_t 通过随机决策的方式生成,以此积累多样化的训练样本并逐步更新 QT。这一阶段的作用是为后续的 QT 提供充分的数据支持,确保其在应用阶段能够准确预测不同环境选择策略的性能。进入应用阶段后,对于第 t 代的总种群 R_t ,算法以当前种群的状态 s_t 作为 QT 的输入,获取各环境选择策略的性能预测值。对总种群 R_t 执行环境选择策略 a_t 后,得到性能预测值 r'_t 。通过比较所有环境选择策略的性能预测值,记性能预测值最大的环境选择策略为 $a_t^{\max} = \arg \max \{r'_t\}$ 。那么,选择 $a_t^{\max}$ 为后续种群采取的环境选择策略。这一机制确保了算法能够根据当前种群状态动态选择最优的环境选择策略,从而提升优化效率和解的质量。值得注意的是,环境选择策略被选择的频率会显著影响 QT 的更新过程,可能导致 Q 值的预测效果不理想。为了解决这一问题,本文引入 CT,用于记录每种环境选择策略在不同种群状态 s_t 下被选择的次数。基于这些记录,计算每种状态 s_t 下的平均 Q 值,从而提升 Q 值预测的准确性,确保算法性能的稳定性。若 $\text{CT}(s_t, a_t) = 0$,则设 $r'_t = 0$ 。此外, Q 值学习网络在实际应用中容易出现过拟合问题,导致算法过度依赖历史数据而降低对新状态的探索能力。为了克服这一弊端并增强种群状态空间(所有可能种群状态的集合)的探索能力,本文采用了 Epsilon 贪婪策略。具体而言,算法以一个小概率 ϵ 在环境选择策略空间中随机选择策略,而不是完全依赖 QT 推荐的策略。这种机制能够在保证算法充分利用已有知识的同时,增加对新策略的探索,从而避免陷入局部最优。

本节所涉及的种群状态 s_t 的刻画方法、环境选择策略 a_t 的设计方法、奖励 r_t 的评价方法将在下文具体说明。

2.2.1 基于决策空间及目标空间的种群状态刻画

在文献[16]的基础上,本文进一步优化种群可行性、收敛性和多样性的刻画方式,在减少种群状态数量的同时,提高状态包含的信息量,具体表达方法如下所示。

(1) ϵ -可行解比例(ρ_ϵ):

$$\rho_\epsilon = \frac{N_{\text{ef}}}{N} \quad (2)$$

其中: N_{ef} 是满足第 t 代定义的松弛约束 $\epsilon_i^{(t)}$ 的解的数量; N 是种群规模; $\epsilon_i^{(t)}$ 为第 i 个约束在第 t 代的动态松弛边界,由式(6)定义。一个解称为 ϵ -可行解,其对所有约束的违背度均满足 $g_i(x) \leq \epsilon_i^{(t)}$ 。

(2) 归一化质心位移(C_{norm}): 设 $C_t = \frac{1}{N_{\text{nd}}} \sum_{i=1}^{N_{\text{nd}}} \mathbf{x}_i(t)$

表示第 t 代非支配解(在种群中,不存在其他解能在所有目标上均不劣于它且至少在一个目标上严格优于它)在决策空间中的质心,连续两代质心之间的位移为

$$C_{\text{raw}} = \|C_t - C_{t-1}\| \quad (3)$$

归一化后的质心位移用于表征收敛趋势,计算公式为

$$C_{\text{norm}} = \frac{C_{\text{raw}}}{\sqrt{D}} \quad (4)$$

其中, D 表示决策空间的维度, N_{nd} 表示第 t 代非支配解的数量, $\mathbf{x}_i(t)$ 表示第 t 代中第 i 个非支配解的决策向量。

(3) ϵ -可行解方差在种群总方差的占比($V_{\epsilon-\text{fea}}$): 设 S_ϵ 为当前种群中所有 ϵ -可行解的索引集, $|S_\epsilon| = N_{\epsilon-\text{fea}}$, $\overline{F_{\epsilon-\text{fea}}}$ 为 ϵ -可行解的目标均值,那么:

$$\sigma_{\epsilon-\text{fea}}^2 = \frac{1}{N_{\epsilon-\text{fea}}} \sum_{i \in S_\epsilon} |F_i - \overline{F_{\epsilon-\text{fea}}}|^2 \quad (5)$$

$$V_{\epsilon-\text{fea}} = \frac{\sigma_{\epsilon-\text{fea}}^2}{\sigma_{\text{total}}^2}$$

其中, $\sigma_{\epsilon-\text{fea}}^2$ 是 ϵ -可行解的方差, σ_{total}^2 是种群解的方差。值得注意的是, ϵ -可行解指的是满足约束违背度小于设定值 ϵ 的解。 ϵ 更新方式如下所示:

$$\epsilon_i^{(t)} = E_i e^{-(\frac{t}{2})^{\text{cp}}} - \delta, i = 1, 2, \dots, q \quad (6)$$

$$Z_i = t / \sqrt{\text{cp} \ln(\frac{\epsilon_i^{(0)} + \delta}{\delta})} \quad (7)$$

其中: $E_i = \epsilon_i^{(0)} + \delta$, $\epsilon_i^{(0)}$ 指第 i 个约束在第 t 代时的动态约束边界, E_i 指第 i 个约束的初始约束边界; Z_i 是控制第 i 个约束边界衰减速度的参数; t 是当前代数; cp 是控制 ϵ 衰减趋势的参数,通常 $\text{cp} > 1$ 时衰减先慢后快, $0 < \text{cp} < 1$ 时先快后慢; δ 是一个接近零的值($\delta = 1 \times 10^{-8}$),

用于确保 ε 最终趋近于零; q 表示约束的数量。

本文将每个状态记为 $s_i = (\rho_\varepsilon, C_{\text{norm}}, V_{\varepsilon-\text{fea}})$, 其特征值在 $[0, 1]$ 范围内均匀划分为“低”“中”“高”三个区间, 划分方式如下: 每个特征均按固定阈值离散化, 即“低”为 $[0, 0.33)$, “中”为 $[0.33, 0.67)$, “高”为 $[0.67, 1]$, 以此为基础形成 27 种理论状态, 便于强化学习构建 Q -table。采用这种划分方式主要基于以下考虑: 首先, 均匀划分是一种直观且无偏差的方式, 能够将连续的种群状态特征离散化, 便于 Q -table 的构建与查询; 其次, 这种划分与种群进化阶段的直观理解相符, 例如可行解比例“低”“中”“高”可分别对应进化初期、过渡期和成熟期; 最后, 在实际进化过程中, 虽然理论上有 27 种状态, 但许多极端组合 (如 $\rho_\varepsilon = 0$ 时 $V_{\varepsilon-\text{fea}} > 0$) 实际上不可能出现, 这使得有效状态空间远小于理论值, 从而避免了因状态过多而导致的强化学习样本稀疏与过拟合问题, 确保了算法的稳定性和学习效率。随着进化搜索的深入, s_i 的组合会逐步增多, 但不会出现状态爆炸导致的采样不足, 引导强化学习只在关键状态区域逐步改进策略, 而非追求全局精确估值, 进一步保证了算法的稳定性。

2.2.2 基于相互作用力原理的环境选择策略

在环境选择策略的设计中, 本文受物理学中相互作用力原理的启发, 将该原理引入 CMOEAs。相互作用力原本用于描述物体之间基于距离的吸引或排斥作用, 类似于优化过程中个体在目标空间和决策空间中的分布关系。原有相互作用力模型通常简单依赖于欧氏距离, 以促进种群的多样性。然而, 在 CMOPs 中, 该方法不能保证可行性和收敛性。因此, 本文对相互作用力模型进行了改进, 设计了三种不同的模式, 分别对应可行性、收敛性和多样性的平衡需求。具体改动包括: 在力公式中引入约束违背度排名、目标函数排名以及距离加权因子, 以动态调整选择压力。当注重可行性时, 力与约束违背度排名正相关, 以优先选择可行解; 当注重收敛性时, 力与目标函数排名正相关, 以加速逼近帕累托前沿; 当注重多样性时, 力仅与距离相关, 以维持种群分布。

当强化学习推荐的环境选择策略 $a_t = 1$ 时, 采用原算法自身的环境选择方式。当环境选择策略 $a_t \neq 1$ 时, 基于本文设计指标进行环境选择。首先, 对所有个体进行非支配排序; 接着, 从第一层开始, 选择所有个体进入下一代种群, 直到下一代种群的大小达到预定的大小; 最后, 如果在同一非支配层内, 个体的数量超过了剩余的种群大小, 那么将根据下文所提指标 F_i , 依次选择值最小的个体补足种群。令 $\text{cv}_i = \sum_k \max(0, g_k(x_i))$, 为个体 i 的总约束违背度。将种

群按 cv_i 升序排列, 排名 r_i^1 从 1 (最小违背度) 到 N (最大违背度), r_i^2 为个体 i 在非支配排序中的前沿层编号 (第 1 层最优)。具体的指标设计如下:

$$F_{ij} = \begin{cases} \frac{\ln \|1 + r_i^1 \cdot r_j^1\|}{\|P_{i,t} - P_{j,t}\|^2 + \|D_{i,t} - D_{j,t}\|^2}, & \text{if } a_t = 2 \\ \frac{\ln \|1 + r_i^2 \cdot r_j^2\|}{\|P_{i,t} - P_{j,t}\|^2 + \|D_{i,t} - D_{j,t}\|^2}, & \text{if } a_t = 3 \\ \frac{1}{e^{\lambda(\|P_{i,t} - P_{j,t}\|^2 + \|D_{i,t} - D_{j,t}\|^2)}}, & \text{if } a_t = 4 \end{cases} \quad (8)$$

其中: F_{ij} 表示第 i 个个体与第 j 个个体间的作用力; r_i^1 与 r_j^1 表示第 i 个个体与第 j 个个体按照约束违背度相加后得到的排名; r_i^2 与 r_j^2 表示第 i 个个体与第 j 个个体的优化目标排名; $P_{i,t}$ 表示第 t 代的第 i 个个体在目标空间的数值, $P_{j,t}$ 表示第 t 代的第 j 个个体在目标空间的数值; $D_{i,t}$ 表示第 t 代的第 i 个个体在决策空间的数值, $D_{j,t}$ 表示第 t 代的第 j 个个体在决策空间的数值; $\|P_{i,t} - P_{j,t}\|^2$ 表示第 t 代中第 i 个个体与第 j 个个体在目标空间的欧式距离; $\|D_{i,t} - D_{j,t}\|^2$ 表示第 t 代中第 i 个个体与第 j 个个体在决策空间的欧式距离, 则第 i 个个体受到所有其他个体的影响之和可以表示为

$$F_i = \sum_{j=1, j \neq i}^N F_{ij} \quad (9)$$

综上所述, 环境选择策略具有如下特点:

(1) 策略一 ($a_t = 1$, 原有策略)。保留被嵌入原算法的环境选择机制。此设计确保了所提框架具有良好的可移植性与兼容性, 能够作为增强模块无缝集成至多种现有 CMOEAs 中, 为其提供动态策略调度能力。

(2) 策略二 ($a_t = 2$, 兼顾可行性与多样性)。通过排斥力模型驱动可行解的分散保留。该策略定义个体间排斥力 F_{ij} 的分子为 $\ln(1 + r_i^1 \cdot r_j^1)$, 其中, r^1 为约束违背度排名 (排名值越小越优)。该设计使约束违背度均较小的两个可行个体间产生的排斥力相对较弱, 因此在计算个体承受的总排斥力 F_i 时, 高可行性个体更易获得较小的 F_i 值, 从而在环境选择中被集体保留的优先级更高。同时, 分母中的空间距离项确保了被保留的可行解之间分布不过于密集。其最终效果是, 在优先筛选可行解的基础上, 进一步促进了种群在目标空间与决策空间中的分布多样性。

(3) 策略三 ($a_t = 3$, 兼顾收敛性与多样性)。机制与策略二同理, 但优化导向不同。该策略将排斥力分子替换为目标函数值排名 r^2 , 从而将选择压力引向收敛性。它促使目标函数值均较优的个体间排斥力减弱, 使得这些收敛性好的个体更容易被同时保留。再结合距离项对分布过近个体的排斥作用, 该策略能够在同一非支配层内, 优先选择出收敛性强且在空间中

分布分散的个体,有效平衡收敛性与多样性。

(4)策略四($a_t=4$,仅考虑多样性)。将排斥力分子设为常数1,使 F_{ij} 完全由个体间的空间距离决定。此时,算法仅根据距离产生排斥:距离越近的个体间排斥力越强。这迫使种群极力避免个体的空间聚集,从而在环境选择中直接最大化种群的分布多样性。

由此可见,本文方法能够根据种群的可行性、收敛性和多样性状态,自主决策环境选择偏好。

2.2.3 环境选择策略的分阶段评价

本文采用分阶段评价方法。当种群中没有可行个体,采用种群中表现最好个体的约束违反度升高或者降低程度评价环境选择方式的性能。考虑种群 P_t 的最优个体,施加该种群的策略为 a_t 。为了评价 a_t 的性能,计算该个体对每一约束的违反度,并将这些违反度相加,得到该个体对所有约束的违反度,记为 ϕ_t 。

令 $\phi_t = \min_{x \in P_t} \sum_{k=1}^q \max(0, g_k(x))$,即第 t 代种群中最小的总约束违反度,那么,环境选择方式 a_t 的性能,记为 r_t ,可表示为

$$r_t = \frac{\phi_t - \phi_{t+1}}{\phi_t} \quad (10)$$

可以看出, r_t 的取值在0~1之间。值得注意的是,如果 $r_t=0$,那么,施加环境选择方式 a_t 之后,种群对约束的满足没有改善;如果 $r_t=1$,那么,施加 a_t 之后,种群的最优个体变为了完全可行解。

当进化种群包含问题的一个或多个可行解时,通过可行解的超体积指标(HyperVolume, HV)的提高程度评价环境选择方式的性能。HV能够综合反映可行解的多样性与收敛性。对于种群 P_t ,计算该种群可行解的HV指标,记为 HV_t 。那么,策略 a_t 的性能 r_t 可表示为

$$r_t = \frac{HV_{t+1} - HV_t}{HV_t} \quad (11)$$

这样一来,策略 a_t 的性能 r_t 可表示为

$$r_t = \begin{cases} \frac{\phi_t - \phi_{t+1}}{\phi_t}, & \text{if } \phi_t > 0 \\ \frac{HV_{t+1} - HV_t}{HV_t}, & \text{otherwise} \end{cases} \quad (12)$$

2.3 环境选择策略执行

本文的环境选择方式如算法3所示,基于强化学习所推荐的 a_t 进行环境选择。当 $a_t=1$ 时,采用原算法的环境选择方式(第2~3行)。当 $a_t \neq 1$ 时,采用基于相互作用力原理设计的三种环境选择策略,引导种群的环境选择(第5~21行)。首先,对种群中的所有个体进行非支配排序,将种群分成多个非支配层;其次,从第一层开始,逐层选择个体进入下一代种群,

算法3 环境选择策略执行 ESSI

输入: R_t, a_t, A_1, A_2, A_3

输出: P_{t+1}, A_1, A_2, A_3

1. $P_{t+1} \leftarrow \phi$
2. IF $a_t = 1$ THEN
3. 根据原有算法的环境选择策略确定 P_{t+1}
4. ELSE
5. IF $a_t = 2$ THEN
6. $U_t \leftarrow A_1 \cup R_t$
7. 根据式(8)和式(9)确定 F_t
8. $A_1 \leftarrow \phi$
9. ELSE IF $a_t = 3$ THEN
10. $U_t \leftarrow A_2 \cup R_t$
11. 根据式(8)和式(9)确定 F_t
12. $A_2 \leftarrow \phi$
13. ELSE
14. $U_t \leftarrow A_3 \cup R_t$
15. 根据式(8)和式(9)确定 F_t
16. $A_3 \leftarrow \phi$
17. END
18. 根据分支支配排序将 U_t 划分为 L_1 至 L_K
19. 将 L_1 层前的所有种群个体放进 P_{t+1}
20. 根据 F_t 对 L_i 中的种群个体排序
21. 将欠缺的种群个体数量按照排序结果放进 P_{t+1}
22. 更新 A_1, A_2 和 A_3
23. END IF

直到下一代种群的大小达到预定的规模。最后,如果在某一层中,个体的数量超过了剩余的种群大小,则基于 F_t 对该层个体进行排序,选择 F_t 值最小的个体补足种群。在环境选择策略执行完毕后,更新存档(第22行)。本文采用的种群存档保留了在偏好的环境选择中被淘汰的种群个体,引导后续进化的种群环境选择。种群存档的更新方式及使用策略,以下将具体说明。

本文设计的环境选择策略 a_t 具有对立性,单一指标下选择的种群个体具有较高的选择偏好,影响后续种群进化的随机性。基于此,采用种群存档保留在偏好的环境选择中被淘汰的种群个体,引导后续进化的种群环境选择。具体而言,当执行环境选择策略后,针对被淘汰个体,采取下述方式使用存档。基于式(8)中 $a_t=2$ 时设计的指标与式(9),计算出的 F_t 较小的 $0.3N$ 个个体,保存在可行性存档 A_1 ;基于式(8)中 $a_t=3$ 设计的指标与式(9),计算出的 F_t 较小的 $0.3N$ 个个体,保存在收敛性存档 A_2 ;基于式(8)中 $a_t=4$ 设计的指标与式(9),计算出的 F_t 较小的 $0.3N$ 个个体,保存在多样性档案 A_3 。其中, N 为种群规模。值得注意的是,并非每次都会更新三个存档,例如:当 $a_t=1$

时会更新三个存档,而当 $a_t=2$ 时,执行了偏好可行性的环境选择策略,此时可行性良好的个体已被选取,因此,不更新可行性存档 A_1 。在接下来的进化过程中,用过 A_1 后会将之清空。在随后的强化学习过程中采用了相对应的环境选择方式 a_t 时,将这些优良个体加入种群再次选择,发掘被淘汰个体的信息潜力。例如:当 $a_t=2$ 时,将可行性存档 A_1 加入种群中进行再次选择,使用完毕后清空 A_1 。需要指出的是,由于存档容量有限,很多时候并不能及时清空存档,因此,在更新存档时会将所有待存档的个体进行比较,只挑选 F_t 值较小的 $0.3N$ 个个体加入存档。

2.4 QT的更新

QT的更新过程如算法4所示。首先根据种群在决策空间及目标空间的特征分布刻画种群可行性、收敛性和多样性,获得 s_t (第1行);接着,根据 P_t 与 P_{t+1} 可行解情况计算 r_t (第2~10行);最后,根据采用的 a_t 及当前种群实时状态 s_t 更新QT(第11行),并在CT相应位置记录更新的次数(第12行)。上述关于状态 s_t 的刻画方式、环境选择策略 a_t 的决策方式和奖励 r_t 的计算方式在2.2节详细讲述。

算法4 Q-Update

输入: $s_t, a_t, P_t, P_{t+1}, CT, QT$

输出: QT, CT

1. 根据式(1)、式(3)和式(5)计算 s_t
2. IF P_t 中没有可行解 THEN
3. 计算 P_t 的约束违背度 ϕ_t
4. 计算 P_{t+1} 的约束违背度 ϕ_{t+1}
5. 根据式(10)计算 r_t
6. ELSE
7. 计算 P_t 的HV $_t$
8. 计算 P_{t+1} 的HV $_{t+1}$
9. 根据式(11)计算 r_t
10. END IF
11. 根据式(13)更新QT
12. $CT(s_t, a_t) \leftarrow CT(s_t, a_t) + 1$

本文采用Q值函数来进行QT的更新,其函数表达式如下:

$$Q_{t+1}(s, a) = Q_t(s, a) + \alpha \cdot [r_t + \gamma \cdot \max_{a'} Q_t(s', a') - Q_t(s, a)] \quad (13)$$

其中, α 是学习率,用于控制当前种群状态下环境选择策略评价对Q值的影响程度; r 是执行动作 a 后获得的即时奖励; γ 是折扣因子,表示对未来奖励的重视程度; s' 是在执行策略 a 后转移到的下一种群状态; $\max Q_t(s', a')$ 表示在下一个状态 s' 下选择最优策

略所对应的Q值。

3 实验

3.1 实验准备

为了测试本文方法的有效性,本章进行了三组实验。第一组实验在三个基准测试问题上,对比了4种最新的CMOEAs与嵌入本文方法后变体(记为“原始算法名-MESS”)的性能表现,验证本文方法的有效性;第二组实验在相同的三个基准测试问题上,选取性能最优的变体Bico-MESS算法,对比6种最新的CMOEAs的性能表现,进一步验证本文方法的综合性能;实验的运行环境为:13 th Gen Intel(R) Core(TM) i5-13400 2.50 GHz、16 GB RAM、Windows 11和PlatEMO^[17]。

3.1.1 测试问题

本文选择了三个基准测试问题:LIR-CMOP(Linear constraint and Inseparable Rotated CMOP)、DASCMP(Difficulty-Adjustable and Scalable CMOP)和DOC(Discontinuous and Opposite Constraints CMOP)。LIR-CMOP测试集适合评估CMOEAs在处理线性约束和非线性目标函数时的性能表现,尤其适用于狭小或离散的可行域搜索测试;DASCMP测试问题可以评估算法在处理复杂约束和具有复杂形状或特性的帕累托前沿时的表现;DOC测试问题则专门用于评估算法在约束具有高度阻碍性(如可行区域狭窄、离散、分布不规则)时寻找可行解和收敛到帕累托前沿的能力,尤其考验算法在约束严重限制目标优化空间时的鲁棒性。因此,本文选取的基准测试问题能够全面覆盖算法在不同约束条件和问题复杂度下的性能表现,从而有效验证本文方法的鲁棒性和适应性。

3.1.2 对比算法与参数设置

选择Bico^[18]、AGEMOEAI^[19]、LSSA^[20]和tDEACPBI^[21]作为本文方法的嵌入算法;选用有效性实验中性能最好的Bico-MESS算法,对比最新的CMOEAs,包括APSEA^[22]、CMOEMT^[23]、CMOES^[24]、IMCMOEAD^[25]、DPCPRA^[26]和CMOQLMT^[27],其中,CMOQLMT是集成了强化学习的CMOEa,用于验证本文方法中强化学习集成机制的先进性。所有算法求解LIR-CMOP和DASCMP问题的最大函数评价次数均为10 000次,求解DOC问题的最大函数评价次数均为80 000次;所有实验均在PlatEMO平台上运行,使用相同的随机种子初始化种群,以消除实现环境与随机性带来的偏差,保证实验的公平性。所有算法的参数设置如表1所示。

表 1 所有对比算法的重要参数设置
Table 1 Important parameter settings of all compared algorithms

算法	重要参数
CMOEA-MESS	贪婪策略因子 $\epsilon=0.3$,折扣率 $\gamma=0.1$,学习率 $\alpha=0.01$,种群规模 $N=100$
Bico	交叉概率 $P_c=1.0$,变异概率 $P_m=1/n$ (n 为决策变量数),模拟二进制交叉(SBX)与多项式变异的分布指数均为20,种群规模 $N=100$
AGEMOEAIH	交叉概率 $P_c=1.0$,多项式变异概率 $P_m=1/n$ (n 为决策变量数),SBX分布指数为30,多项式变异分布指数为20。目标数 $M=3$ 时种群规模 $N=200$
LSSA	交叉概率 $P_c=1.0$,多项式变异概率 $P_m=1/n$ (n 为决策变量数),SBX与多项式变异的分布指数均为20,种群规模 $N=100$
tDEACPBI	交叉概率 $P_c=1.0$,多项式变异概率 $P_m=1/n$ (n 为决策变量数),SBX与多项式变异的分布指数均为20,PBI惩罚因子 $\theta=5$,种群规模 $N=100$
APSEA	交叉概率 $P_c=1.0$,多项式变异概率 $P_m=1/n$ (n 为决策变量数),SBX与多项式变异的分布指数均为20,进化阶段切换参数 $\alpha=0.05,\beta=0.05$,种群规模 $N=100$
CMOEMT	交叉概率 $P_c=1.0$,多项式变异概率 $P_m=1/n$ (n 为决策变量数),SBX与多项式变异的分布指数均为20,交叉概率 $CR=1$,折扣因子 $F=0.5$,进化阶段切换参数 $\eta=0.5$,种群规模 $N=91$
CMOES	交叉概率 $P_c=1.0$,多项式变异概率 $P_m=1/n$ (n 为决策变量数),SBX与多项式变异的分布指数均为20,交叉概率 $CR=1$,折扣因子 $F=0.5$,约束松弛因子 $\eta=2$,进化阶段切换参数 $\theta=0.5$,种群规模 $N=100$
IMCMOEAD	交叉概率 $P_c=1.0$,多项式变异概率 $P_m=1/n$ (n 为决策变量数),SBX与多项式变异的分布指数均为20,聚类组数 $K=10$,模型组大小 $L=3$,邻居数量 $T=0.1N$
DPCPRA	交叉概率 $P_c=1.0$,多项式变异概率 $P_m=1/n$ (n 为决策变量数),SBX与多项式变异的分布指数均为20,第一阶段收敛判断阈值 $\alpha=0.01$,代数窗口 $AG=100$,种群规模 $N=100$
CMOQLMT	交叉概率 $P_c=1.0$,多项式变异概率 $P_m=1/n$ (n 为决策变量数),SBX与多项式变异的分布指数均为20,交叉概率 $CR=1$,折扣因子 $F=0.5$,贪婪策略因子 $\epsilon=0.9$,折扣率 γ ,学习率 α 采用RL领域建议值,种群规模 $N=100$

3.1.3 性能指标

鉴于超体积(HyperVolume, HV)^[28]和反世代距离(Inverted Generational Distance, IGD)^[29]可以综合评价最终解的可行性、收敛性和多样性,本文选取两者作为评价指标。采用Wilcoxon rank-sum test进行假设检验,取显著水平为0.05,并采用符号“+”“-”和“=”分别表示当前算法显著优于、劣于和无显著差别于另一个算法。

3.2 本文方法的有效性

所有算法在测试问题上均独立运行30次,详细的HV和IGD对比统计情况分别如表2和表3所示。从HV的统计结果可以看出,“*- MESS”的性能在大多数测试问题上显著优于原始算法,Bico-MESS、AGEMOEAIH-MESS和LSSA-MESS均实现了至少20个问题上的性能优势,tDEACPBI-MESS虽然仅在17个问题存在明显优势,但是也只在2个问题存在劣势。从IGD的统计结果可以看出,“*- MESS”的性能在大多数测试问题上依然显著优于原始算法,但是稍差于在HV指标上的表现。出现这种现象的可能原因推测如下:

(1)采用相互作用力原理设计的三种环境选择策略具有一定的耦合性。虽然进化算法的前提是需要通过随机执行环境选择策略积累训练数据,但是相互耦合的三种环境选择策略不会因为随机执行的策略

而严重影响算法性能。如果所设计的三种环境选择策略不具备这种耦合性,那么随机执行环境选择策略在进化前期会对进化后期产生严重的负面影响,降低本文方法的有效性。

(2)“*- MESS”的IGD指标差于HV指标的原因有多个方面。首先,强化学习中的策略评价方法偏重HV指标,这导致算法偏向于选择有利于提高HV指标的环境选择策略;虽然种群的收敛能提高HV指标,但是若强化学习感知到提高种群多样性比加速种群收敛更有利于提高HV指标,则强化学习依然会偏向于提高种群多样性,显然收敛性会更差一些。其次,在种群出现可行解后的一段时间,可行解的占比会迅速增加,导致QT中突然出现新的种群状态但数据量严重不足;此时建立的强化学习模型较为脆弱,推荐精度显著下降,会在一定程度上影响算法性能。最后,将本文方法嵌入现有算法的方式是保留原有的环境选择策略,所设计的环境选择策略和原有策略之间可能存在选择偏好的冲突,这在一定程度上也影响了算法的稳定性。

(3)实验结果中,“*- MESS”在个别问题上出现了“NaN”值,表明在30次独立运行中,该算法未能找到任何可行解。这主要受限于嵌入原算法的性能,发生在可行域极其狭窄或离散的测试问题(如DOC系列、LIR-CMOP7/8)上。即便如此,在大多数嵌入算法无

法找到可行解的问题上,“*- MESS”均能帮助找到可行解。这反映了“*- MESS”在改进嵌入算法性能方面的有效性和局限性:虽然“*- MESS”具备自主调控的环境选择能力,但是算法性能还受嵌入算法的原有策略影响,干扰了“*- MESS”的自主调控性能,导致部分问题无法求解。

(4) 尽管所提方法在多数问题上表现出显著优势,但其仍存在若干局限性:首先,强化学习机制的引入虽能提升策略自适应性,但也带来额外计算开销,在评价成本低的问题中影响整体效率;其次,在适应度地形平滑或约束简单的问题中,强化学习可能因过度调整而引起收敛波动,这也是在部分问题求解中不能显著提升原算法性能的重要原因;此外,状态刻画特征与训练样本需求之间的平衡仍具挑战,特征过多易导致训练不足,特征过少则降低决策精度;最后,在可行域极端狭窄的问题中,算法表现仍受初始种群分布的显著影响。

表 2 原算法与改进算法的 HV 指标统计

Table 2 HV indicator statistics of original algorithms and their improved versions

算法	+/-/=
Bico-MESS VS Bico	21/4/7
AGEMOEAI-MESS VS AGEMOEAI	20/2/10
LSSA-MESS VS LSSA	21/2/9
tDEACPB-MESS VS tDEACPB	17/2/13

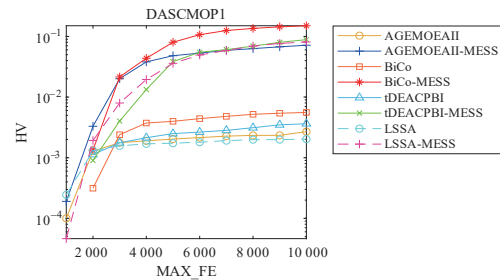
表 3 原算法与改进算法的 IGD 指标统计

Table 3 IGD indicator statistics of original algorithms and their improved versions

算法	+/-/=
Bico-MESS VS Bico	21/7/4
AGEMOEAI-MESS VS AGEMOEAI	21/6/5
LSSA-MESS VS LSSA	23/6/3
tDEACPB-MESS VS tDEACPB	18/6/8

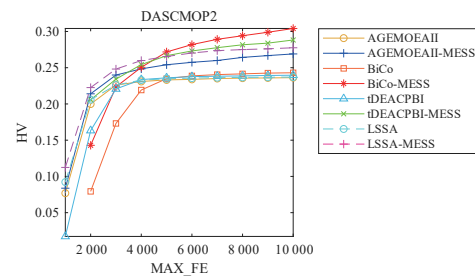
进而,以 DASCOP1、DASCOP2、LIRCMOP1 问题为例,研究“*- MESS”在进化求解过程中的性能表现。如图 2 所示,每 1 000 代记录种群解的 HV 指标,获得 HV 变化曲线。可以看出,在进化初期,“*- MESS”已经展示出一定的求解优势;而 3 000 代左右,“*- MESS”的求解优势显著高于原算法;进化过程结束的时候,AGEMOEAI-MESS、BiCo-MESS、tDEACPB-MESS 和 LSSA-MESS 排在所有算法的前四名。更重要的是,AGEMOEAI、BiCo、tDEACPB 和 LSSA 均存在早熟收敛的问题, HV 指标在 4 000 代左右已经趋于稳定,而“*- MESS”能有效缓解这个问题,且在 LIRCMOP1 问题中还尚未收敛。这说明,“*- MESS”具有很强的种群多样性维持能力,能通过强化学习增

强决策空间的全局搜索效果。



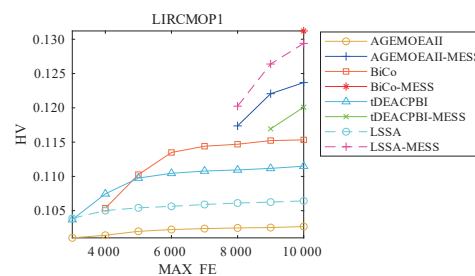
(a) 所有算法求解 DASCOP1 过程中的 HV 变化

(a) The changes in HV during the process of solving DASCOP1 by all algorithms



(b) 所有算法求解 DASCOP2 过程中的 HV 变化

(b) The changes in HV during the process of solving DASCOP2 by all algorithms



(c) 所有算法求解 LIRCMOP1 过程中的 HV 变化

(c) The changes in HV during the process of solving LIRCMOP1 by all algorithms

图 2 不同算法进化求解过程中的 HV 指标变化

Figure 2 HV indicator variation during evolutionary optimization by different algorithms

3.3 本文方法的综合性能

本文所有算法在测试问题上均独立运行 30 次,详细的 HV 和 IGD 对比统计情况分别如表 4 和表 5 所示,从实验结果中可以看出:

(1) 对比于 APSEA, BiCo-MESS 在绝大多数问题上显著优于 APSEA。可能的原因是 APSEA 中的辅助种群是逐步缩减的,对种群多样性的关注弱于种群收敛性,而 BiCo-MESS 在策略评价的引导下兼顾了多样性和收敛性。然而, BiCo-MESS 在 LIR-CMOP7~

表4 Bico-MESS与其他算法的HV指标统计

Table 4 HV indicator statistics of Bico-MESS and other algorithms

算法	+/-/=
Bico-MESS VS APSEA	22/4/6
Bico-MESS VS CMOEMT	15/5/12
Bico-MESS VS CMOES	15/6/11
Bico-MESS VS IMCMOEAD	23/1/8
Bico-MESS VS DPCPRA	17/7/8
Bico-MESS VS CMOQLMT	16/3/13

表5 Bico-MESS与其他算法的IGD指标统计

Table 5 IGD indicator statistics of Bico-MESS and other algorithms

算法	+/-/=
Bico-MESS VS APSEA	24/6/2
Bico-MESS VS CMOEMT	19/7/6
Bico-MESS VS CMOES	19/7/6
Bico-MESS VS IMCMOEAD	28/1/3
Bico-MESS VS DPCPRA	20/7/5
Bico-MESS VS CMOQLMT	19/8/5

LIR-CMOP8上的表现不如APSEA,可能是APSEA在对种群进化的引导过程中,其多阶段策略能从不同进化方向逼近狭窄的可行域。同时,两者在DOC3和DOC6问题上的性能相当,可能是由于极端约束导致可行解稀缺,两者均无法较好求解。

(2)对比于CMOEMT,Bico-MESS在DASCMP1~DASCMP2、DASCMP9、DOC1、DOC4~DOC5和DOC7~DOC8问题上显著优于CMOEMT。可能是因为Bico-MESS通过强化学习的动态环境选择策略和种群存档机制,能够更好地平衡种群的收敛性和多样性,从而在复杂约束条件下找到可行解。然而,Bico-MESS在LIR-CMOP7~LIR-CMOP8上的表现却不如CMOEMT,这可能是由于DASCMP6和DASCMP8的可行域较小,Bico-MESS在进化前期贪婪了可行性的评价奖励,一定程度上忽略了种群的多样性;而CMOEMT通过知识转移策略在不同任务之间共享有用信息,能够获取更多定位可行域的进化信息。

(3)Bico-MESS在DASCMP1~DASCMP2、DASCMP5、DASCMP7~DASCMP9、DOC1~DOC2上显著优于CMOES。即使对于可行域非常小且分散的LIR-CMOP1~LIR-CMOP2、LIR-CMOP9~LIR-CMOP10和DOC7~DOC9等问题,Bico-MESS也依然优于CMOES,这可能是因为Bico-MESS通过种群存档机制保留了被淘汰的潜在优良个体,增强了算法的全局探索能力。而在LIR-CMOP6~LIR-CMOP7和LIR-CMOP13~LIR-CMOP14问题上,CMOES的性能优于Bico-MESS,这可能是因为CMOES通过均匀搜索策略能够有效覆盖整个可行

域,避免算法过早收敛到局部最优解,保持了种群的多样性。

(4)对比于IMCMOEAD,Bico-MESS在所有DASCMP问题中均具有显著优势,并且在IMCMOEAD找不到可行解的5个问题上,Bico-MESS均找到了可行解。同时,Bico-MESS在10个LIR-CMOP问题和8个DOC问题上也取得了显著优势。这可能是由于DASCMP测试集中的问题通常具有复杂的约束条件,且可行域可能被分割成多个不连续的区域。IMCMOEAD在处理这种问题时,由于逆建模的复杂性和约束处理机制的局限性,无法有效探索整个搜索空间。

(5)对比于DPCPRA,Bico-MESS在所有DASCMP和DOC问题上均取得了显著优势,但在LIR-CMOP测试集上不如DPCPRA。这可能是由于DPCPRA算法采用双种群机制,主种群专注于优化所有约束以确保解的可行性,而辅助种群通过动态约束处理机制探索更广泛的不可行解空间。在LIR-CMOP测试集这类具有较小可行域和离散帕累托前沿的问题上,这种双种群协作方式使DPCPRA算法能够更有效地在可行域和不可行域之间高效探索。

(6)对比于CMOQLMT,Bico-MESS在CMOQLMT未找到可行解的7个问题上,均找到了可行解。这可能是由于Bico-MESS算法设计了基于相互作用力原理的三种环境选择策略,这些策略分别侧重于可行性、收敛性和多样性。在面对复杂约束和不连续可行域时,通过强化学习框架动态集成这些策略,使算法能够更全面地覆盖帕累托前沿,提高解的质量和分布的均匀性。

进而,以DASCMP1、DASCMP2、DOC2、LIR-CMOP1问题为例,研究BiCo-MESS和对比算法求解结果在目标空间的分布情况。如图3所示,阴影区域为问题在目标空间的可行域,散点为种群分布。可以看出,BiCo-MESS在约束帕累托前沿具有很好的分布,尤其在DASCMP1和DASCMP2问题上,种群分布均匀且遍布在左右前沿上;对于DOC2问题,虽然所有对比算法都找不到可行解,但BiCo-MESS仍能找到可行解,且不可行的种群个体还在朝着约束帕累托前沿继续进化;对于LIR-CMOP1问题,虽然BiCo-MESS的种群解没能覆盖所有的约束帕累托前沿,但是相对于其他算法的解质量明显更优。这说明,“*-MESS”确实可以平衡可行性、收敛性和多样性,提高问题的求解质量。

3.4 消融实验

本节验证三种环境选择策略对算法性能的影响。Bico-MESS-v1算法仅保留原算法的环境选择策略

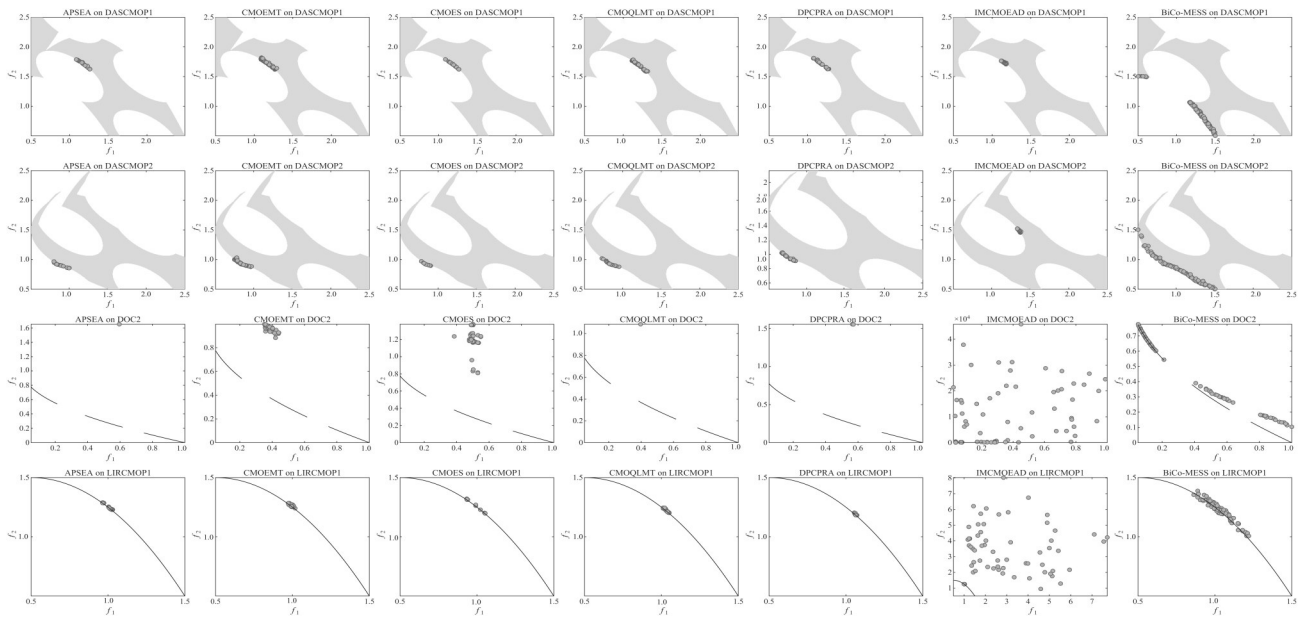


图3 所有算法对部分问题求解结果的目标空间分布

Figure 3 Distribution of solutions in objective space for selected problems by all algorithms

($a_t=1$)和基于多样性的策略($a_t=4$);Bico-MESS-v2仅保留原算法的环境选择策略($a_t=1$)和基于收敛性的策略($a_t=3$);Bico-MESS-v3仅保留原算法的环境选择策略($a_t=1$)和基于可行性的策略($a_t=2$)。所有算法在测试问题上均独立运行30次,详细的HV和IGD对比统计情况分别如表6和表7所示,可得到如下实验分析结果:

(1)Bico-MESS算法通过强化学习动态集成多样性、收敛性和可行性三种策略,在部分问题上显著优于单一策略变体(v1-v3)。具体而言,相较于各变体,Bico-MESS在HV和IGD指标上分别取得4~7项显著优势,尤其在DOC3、DASCMP9等复杂约束问题上表现突出。消融实验证实,三种策略各司其职:多样性策略保障帕累托前沿覆盖,收敛性策略加速逼近最优前沿,可行性策略确保严格约束下的求解能力。

(2)如图4与表8所示,为进一步分析BiCo-MESS在LIR-CMOP4、DASCMP2、DASCMP6和DOC5问题中的策略调度行为,进行了相应的仿真与统计。结果表明,在预热阶段,策略采用随机调度;而在进化后期,强化学习模块展现出明显的策略偏好。具体来说,对于LIR-CMOP4、DASCMP2和DASCMP6,策略4的使用占主导地位,而DOC5则更倾向于采用策略3。此外,DOC5的策略分布相对均衡,其他问题的策略使用则较为集中,部分策略占比仅为10%左右。这表明,强化学习机制能够根据不同问题的求解需求,通过动态调节策略权重,有效提升算法在多样化问题

上的鲁棒性。

表6 消融实验中变体算法HV指标情况

Table 6 HV indicator results of variant algorithms in ablation study

算法	+/-/=
Bico-MESS VS Bico-MESS-v1	6/0/26
Bico-MESS VS Bico-MESS-v2	4/1/27
Bico-MESS VS Bico-MESS-v3	6/0/26

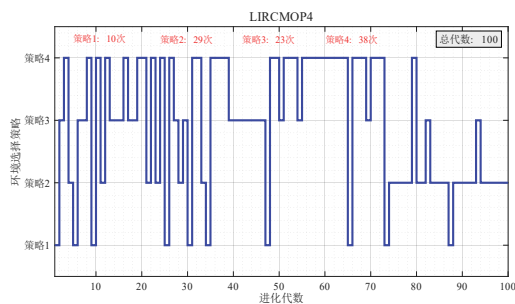
表7 消融实验中变体算法IGD指标情况

Table 7 IGD indicator results of variant algorithms in ablation study

算法	+/-/=
Bico-MESS VS Bico-MESS-v1	6/0/26
Bico-MESS VS Bico-MESS-v2	6/2/24
Bico-MESS VS Bico-MESS-v3	7/0/25

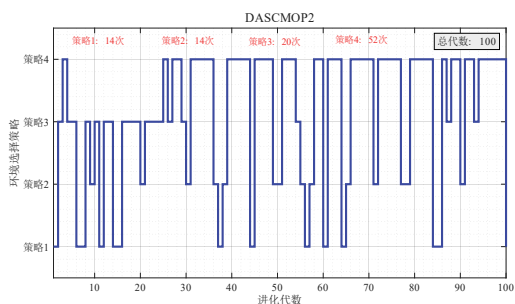
4 模拟集成电路设计中的应用

为验证本文方法在求解实际问题的可扩展性,考虑了在模拟集成电路(Analog Integrated Circuit, AIC)设计问题中的应用。模拟集成电路设计是一类典型的、极具挑战性的CMOP。其特殊性体现在:(1)增益、带宽、功耗等性能指标之间存在严峻且非线性的权衡关系;(2)物理尺寸、工艺角、稳定性等约束条件复杂且严格;(3)评估通常依赖耗时的侧重集成电路的仿真程序(Simulation Program with Integrated Circuit Emphasis, SPICE)仿真,属于计算昂贵的黑箱优化。这种高昂的计算成本严重限制了可用的函数评价次数,从而对算法在有限函数评估次数内高效平衡解的



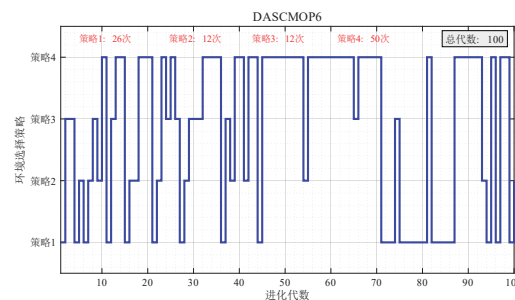
(a) LIRC-MOP4求解过程中的策略调度

(a) Strategy scheduling in the process of solving LIRC-MOP4



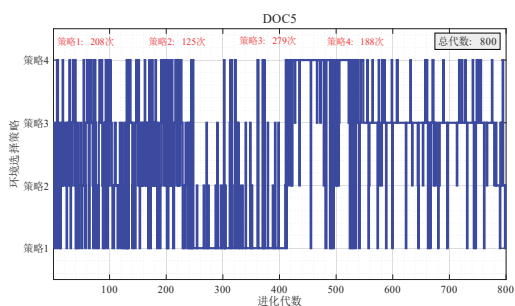
(b) DASC-MOP2求解过程中的策略调度

(b) Strategy scheduling in the process of solving DASC-MOP2



(c) DASC-MOP6求解过程中的策略调度

(c) Strategy scheduling in the process of solving DASC-MOP6



(d) DOC5求解过程中的策略调度

(d) Strategy scheduling in the process of solving DOC5

图4 环境选择策略的调度过程仿真

Figure 4 Simulation of environmental selection strategy scheduling process

表8 环境选择策略的调度情况统计

Table 8 Statistics of environmental selection strategy scheduling

问题名称	总代数	策略1 占比/%	策略2 占比/%	策略3 占比/%	策略4 占比/%	切换 次数
LIRC-MOP4	100	10.0	29.0	23.0	38.0	49
DASC-MOP2	100	14.0	14.0	20.0	52.0	50
DASC-MOP6	100	26.0	12.0	12.0	50.0	49
DOC5	800	26.0	15.6	34.9	23.5	392

可行性、收敛性与多样性提出了更高要求,也使得对本文所提的能动态适应搜索状态的环境选择策略自动选择机制的需求更为迫切。本文工作不仅将所提算法应用于该领域验证其有效性,更通过引入强化学习驱动的动态策略选择和种群存档机制,为这类高成本、强约束的工程优化问题提供了一种既能保证成功率又能高效逼近设计极限的自动化解决方案。本文以一个二级运算放大器的元器件尺寸设计问题为例,具体结构见图5,其中,VIP为同相输入端,VIN为反相输入端,M1~M12为MOS场效应晶体管,VDD为正电源电压,C为电容,R为电阻。该问题的决策变量是二级运算放大器的元器件尺寸参数,包含17个决策变量,所有变量均被归一化到其物理可行的范围 $[0.000\ 000\ 6, 0.000\ 003]$ 。模拟集成电路的性能指标见表9。为了将所提方法应用到该问题上,选取上述2个指标作为优化目标,其他指标作为问题约束,分别命名为AIC1~AIC4,具体设置详见表10。

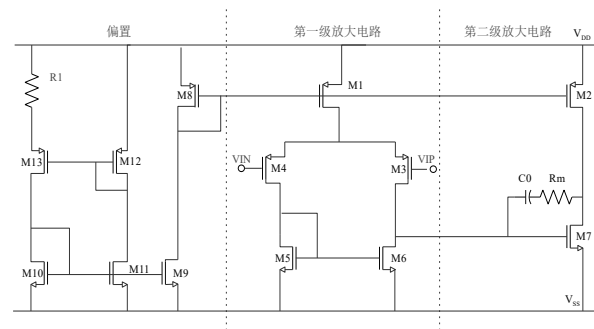


图5 二级运算放大器的电路结构图

Figure 5 Circuit structure of the two-stage operational amplifier

本文按照第3章的实验设置和对比算法在AIC测试问题上进行实验,详细的HV对比统计情况如表11所示。实验结果表明,本文提出的Bico-MESS算法在4个AIC问题上展现出全面的优越性,性能对比结果分析如下:

(1)在AIC1问题上,对比算法IMCMOEA未能找到可行解,而Bico-MESS则以最高HV值(0.998 32)稳健收敛。这印证了2.3节中的结论:本文方法通过种

表9 模拟集成电路优化设计问题的主要指标

Table 9 Main indicators of analog integrated circuit optimization design problem

指标	含义	评价
开环电压增益(Open-loop Gain, Gain)	衡量运放在开环条件下将输入端的微小差分电压信号放大多少倍的能力	越高越好,高增益意味着更高的精度,在构成闭环系统(如滤波器、放大器)时,其实际增益更接近理想计算公式
单位增益带宽(Unity Gain-bandwidth, UGBW)	运放开环增益下降至0 dB时所对应的信号频率,直接反映了运放所能处理的小信号速度极限,是衡量其频率响应能力的关键指标	越高越好,UGBW决定了运放能够不失真地放大多高频率的信号。Gain和UGBW通常存在制约关系(增益带宽积恒定),是设计中需要互相权衡的指标
电源抑制比(Power Supply Rejection Ratio, PSRR)	衡量运放抑制其电源引脚上的电压噪声或纹波,使其不传递到输出端的能力。它表示为开环增益与电源变化增益之比	越高越好,高的PSRR意味着电路对电源噪声不敏感,这在嘈杂的数字-模拟混合电路系统中至关重要。低的PSRR会导致电源噪声耦合到输出信号中,降低信噪比
共模抑制比(Common-Mode Rejection Ratio, CMRR)	衡量运放抑制其两个输入端共有的、完全相同的信号(即共模信号,如环境噪声)的能力。它表示为差分增益与共模增益之比	越高越好,CMRR是差分放大器的核心指标,高的CMRR能够有效提取出微小的有用差分信号,同时抑制掉强大的共模干扰
功耗(power)	运放在静态工作条件下,从电源汲取的总功率。通常计算为电源电压与静态电流的乘积	越低越好,直接关系到设备的续航时间、发热量以及系统的总能耗,几乎与所有其他性能指标都存在根本性的冲突

表10 AIC1-4测试问题的优化目标和约束设置

Table 10 Optimization objectives and constraint settings for AIC1-4 test problems

问题	优化目标	约束
AIC1	CMRR, PSRR	Gain>170 dB, UGBW>4 000 000 Hz, Power<0.000 1 W
AIC2	Gain, UGBW	CMRR>35 dB, PSRR>25 dB, Power<0.000 1 W
AIC3	CMRR, Power	PSRR>25 dB, Gain>170 dB, UGBW>4 000 000 Hz
AIC4	PSRR, UGBW	CMRR>35 dB, Gain>170 dB, Power<0.000 1 W

群存档机制保留被淘汰的潜在优良个体,有效避免了因初始种群不佳或早期选择压力不当而导致的可行域搜索失败。对于实际设计者而言,这意味着使用Bico-MESS能够显著提高找到可行设计方案的成功率,降低因算法失效而需反复调试的时间成本。

(2) AIC问题中各个电路性能指标(如增益、带宽、功耗)之间存在固有的、严峻的权衡关系。Bico-MESS在AIC3问题上取得了非常高的HV值(0.999 85),这表明其获得的帕累托解集不仅全面,而且每个解都逼近了当前工艺约束下的性能极限。这得益于2.2节所讨论的策略耦合性与强化学习的动态决策机制,能根据当前设计方案的短板,动态切换优化焦点,从而系统性地逼近全局最优的设计方案集合。

(3) 对比算法在部分AIC问题上出现了性能波动,如APSEA在AIC2上标准差较大;而Bico-MESS则表现出高度稳定的性能。这种稳定性意味着设计者无需为算法本身的随机性付出额外的验证成本。

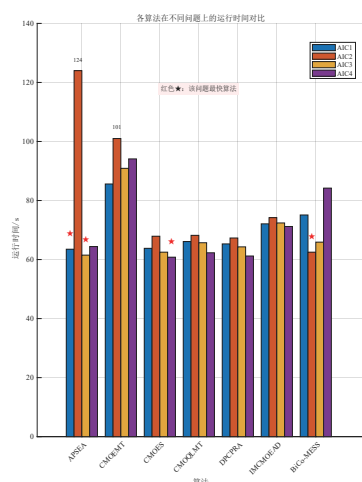
(4) 如图6所示,BiCo-MESS虽然引入了包含策略

性能评估在内的强化学习模块,导致单次评估的运行开销略有增加,但由于算法在种群进化停滞或完全收敛时会提前退出,因此并未消耗全部预设的函数评价次数,从而在整体上缩短了运行时间。总体而言,BiCo-MESS的运行时间相对较短,尤其在处理AIC2问题时求解速度最快。相比之下,如APSEA算法在求解AIC2时,若无法快速收敛、必须用完所有评价次数,则会导致时间消耗显著增加。

表11 Bico-MESS与其他算法在AIC问题上的HV指标统计

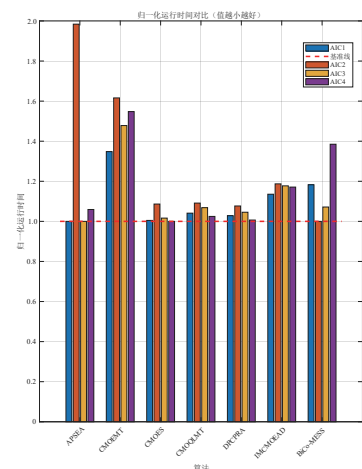
Table 11 HV indicator statistics of Bico-MESS and other algorithms on AIC problems

算法	+/-/=
Bico-MESS VS APSEA	4/0/0
Bico-MESS VS CMOEMT	3/0/1
Bico-MESS VS CMOES	3/0/1
Bico-MESS VS IMCMOEAD	3/0/1
Bico-MESS VS DPCPRA	4/0/0
Bico-MESS VS CMOQLMT	4/0/0



(a) 原始运行时间对比

(a) Comparison of original running time



(b) 归一化运行时间对比

(b) Comparison of normalized running time

图6 所有算法的时间消耗对比

Figure 6 Time consumption comparison of all algorithms

5 结论

针对CMOEAs,本文提出了一种强化学习引导的自主环境选择策略约束多目标进化优化方法。首先,通过与6种先进的CMOEAs在三个基准测试问题和一个现实问题上的对比,验证了本文方法的综合性能,实验结果表明,本文方法具有显著的优势,尤其在处理约束复杂、可行域不连续的测试问题时,展现了较强的鲁棒性和适应性。其次,通过消融实验验证了三种环境选择策略的有效性,结果表明各策略均能有效提高算法性能,且强化学习驱动的动态集成机制是提升算法鲁棒性的关键。此外,采用种群存档保留被淘汰的个体,既能避免优良个体被淘汰,又增强了算法的全局搜索能力。最后,通过算法在模拟集成电路中的应用,验证了所提方法的可扩展性。

需要注意的是,本文方法虽然在多数测试问题上具有优势,但对于部分适应度地形较为平滑的问题,强化学习的动态策略选择机制过于复杂,导致在优化过程中出现一些不必要的调整,反而影响了优化效果。此外,刻画种群状态所提取的特征数量与采集训练样本的数量存在冲突,兼顾种群状态刻画精度和训练数据数量具有挑战性。因此,为了进一步提升本文方法的性能,今后将研究以下两个问题:

(1) 针对适应度地形平滑的优化问题,设计轻量化强化学习模型,使其能够根据问题复杂度自动调整策略评价方法。

(2) 针对种群状态刻画与训练样本采集的平衡问题,设计更高效的种群状态刻画特征,降低特征的数量,减轻在线优化中的训练样本采样的负担。

参考文献

- [1] Hu Hejuan, Sun Xiaoyan, Zeng Bo, et al. Enhanced evolutionary multi-objective optimization-based dispatch of coal mine integrated energy system with flexible load[J]. Applied Energy, 2022, 307: 118130.
- [2] Zuo Mingcheng, Dai Guangming, Peng Lei, et al. A case learning-based differential evolution algorithm for global optimization of interplanetary trajectory design[J]. Applied Soft Computing, 2020, 94: 106451.
- [3] Raju M S S, Dutta S, Mallipeddi R, et al. A dual-population and multi-stage based constrained multi-objective evolutionary[J]. Information Sciences, 2022, 615: 557-577.
- [4] Ming Fei, Gong Wenyin, Wang Ling, et al. A constrained many-objective optimization evolutionary algorithm with enhanced mating and environmental selections[J]. IEEE Transactions on Cybernetics, 2023, 53(8): 4934-4946.
- [5] Tan Zhiping, Tang Yu, Li Kangshun, et al. Differential evolution with hybrid parameters and mutation strategies based on reinforcement learning[J]. Swarm and Evolutionary Computation, 2022, 75: 101194.
- [6] Ming Fei, Gong Wenyin, Wang Ling, et al. Constrained multi-objective optimization with deep reinforcement learning assisted operator selection[J]. IEEE/CAA Journal of Automatica Sinica, 2024, 11(4): 919-931.
- [7] Yang Yufei, Zhang Changsheng, Liu Yi, et al. Deep reinforcement learning assisted novelty search in Voronoi regions for constrained multi-objective optimization[J]. Swarm and Evolutionary Computation, 2024, 91: 101732.
- [8] Wang Chao, Liu Zhihao, Qiu Jianfeng, et al. Adaptive constraint handling technique selection for constrained multi-objective optimization[J]. Swarm and Evolutionary Com-

- putation, 2024, 86: 101488.
- [9] Zuo Mingcheng, Gong Dunwei, Wang Yan, et al. Process knowledge-guided autonomous evolutionary optimization for constrained multiobjective problems[J]. *IEEE Transactions on Evolutionary Computation*, 2024, 28(1): 193-207.
- [10] Tang Huanrong, Yu Fan, Zou Juan, et al. A constrained multi-objective evolutionary strategy based on population state detection[J]. *Swarm and Evolutionary Computation*, 2022, 68: 100978.
- [11] Zhang Kai, Xu Zhiwei, Yen G G, et al. Two-stage multi-objective evolution strategy for constrained multiobjective optimization[J]. *IEEE Transactions on Evolutionary Computation*, 2024, 28(1): 17-31.
- [12] Sun Yan, Zhou Xiaojun, Yang Chunhua, et al. Constrained multi-objective state transition algorithm via adaptive bidirectional coevolution[J]. *Expert Systems with Applications*, 2025, 266: 126073.
- [13] Shi Chenli, Wang Ziqi, Jin Xiaohang, et al. A novel three-stage multi-population evolutionary algorithm for constrained multi-objective optimization problems[J]. *Complex & Intelligent Systems*, 2024, 10(1): 655-675.
- [14] Ming Fei, Gong Wenyin, Wang Ling, et al. Balancing convergence and diversity in objective and decision spaces for multimodal multi-objective optimization[J]. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2023, 7(2): 474-486.
- [15] Ming Fei, Gong Wenyin, Wang Ling, et al. A tri-population based co-evolutionary framework for constrained multi-objective optimization problems[J]. *Swarm and Evolutionary Computation*, 2022, 70: 101055.
- [16] Zuo Mingcheng, Xue Yuan, Gong Dunwei, et al. Reinforcement learning-guided multi-environmental selection strategy for constrained multi-objective evolutionary optimization[C]//*Proceedings of the Genetic and Evolutionary Computation Conference Companion*. New York: ACM, 2025: 467-470.
- [17] Tian Ye, Cheng Ran, Zhang Xingyi, et al. PlatEMO: A MATLAB platform for evolutionary multi-objective optimization [educational forum][J]. *IEEE Computational Intelligence Magazine*, 2017, 12(4): 73-87.
- [18] Liu Zhizhong, Wang Bingchuan, Tang Ke. Handling constrained multiobjective optimization problems via bidirectional coevolution[J]. *IEEE Transactions on Cybernetics*, 2022, 52(10): 10163-10176.
- [19] Panichella A. An improved Pareto front modeling algorithm for large-scale many-objective optimization[C]//*Proceedings of the Genetic and Evolutionary Computation Conference*. New York: ACM, 2022: 565-573.
- [20] Wang Zihan, Mao Bochao, Hao Hao, et al. Enhancing diversity by local subset selection in evolutionary multiobjective optimization[J]. *IEEE Transactions on Evolutionary Computation*, 2023, 27(5): 1456-1469.
- [21] Ming Fei, Gong Wenyin, Wang Ling, et al. A constraint-handling technique for decomposition-based constrained many-objective evolutionary algorithms[J]. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2023, 53(12): 7783-7793.
- [22] Tian Ye, Wang Ruiqin, Zhang Yajie, et al. Adaptive population sizing for multi-population based constrained multi-objective optimization[J]. *Neurocomputing*, 2025, 621: 129296.
- [23] Ming Fei, Gong Wenyin, Wang Ling, et al. Constrained multiobjective optimization via multitasking and knowledge transfer[J]. *IEEE Transactions on Evolutionary Computation*, 2024, 28(1): 77-89.
- [24] Ming Fei, Gong Wenyin, Jin Yaochu. Even search in a promising region for constrained multi-objective optimization[J]. *IEEE/CAA Journal of Automatica Sinica*, 2024, 11(2): 474-486.
- [25] Farias L R C, Araújo A F R. An inverse modeling constrained multi-objective evolutionary algorithm based on decomposition[C]//2024 *IEEE International Conference on Systems, Man, and Cybernetics*. Piscataway: IEEE, 2024: 3727-3732.
- [26] Qiao Kangjia, Chen Zhaolin, Qu Boyang, et al. A dual-population evolutionary algorithm based on dynamic constraint processing and resources allocation for constrained multi-objective optimization problems[J]. *Expert Systems with Applications*, 2024, 238: 121707.
- [27] Ming Fei, Gong Wenyin, Gao Liang. Adaptive auxiliary task selection for multitasking-assisted constrained multi-objective optimization [feature][J]. *IEEE Computational Intelligence Magazine*, 2023, 18(2): 18-30.
- [28] Zitzler E, Thiele L. Multiobjective evolutionary algorithms: A comparative case study and the strength Pareto approach[J]. *IEEE Transactions on Evolutionary Computation*, 1999, 3(4): 257-271.
- [29] Coello C A C, Cortés N C. Solving multiobjective optimization problems using an artificial immune system[J]. *Genetic Programming and Evolvable Machines*, 2005, 6(2): 163-190.

作者简介



左明成 男,1992年9月出生于山东省烟台市。现为中国矿业大学计算机学院/人工智能学院副教授。主要研究方向为人工智能驱动的智能优化。

E-mail: mingcheng.zuo@cumt.edu.cn



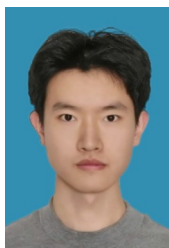
薛源 男,1998年6月出生于江苏省徐州市。现为中国矿业大学计算机学院/人工智能学院硕士研究生。主要研究方向为约束多目标进化优化。

E-mail: ts23810019p31@cumt.edu.cn



巩敦卫 男,1970年3月出生于江苏省徐州市。现为青岛科技大学微电子学院二级教授、院长。主要研究方向为智能优化理论与应用。

E-mail: dwgong@vip.163.com



薛天洋 男,2003年6月出生于河北省邢台市。现为中国矿业大学计算机学院/人工智能学院硕士研究生。主要研究方向为多PVT模拟集成电路优化。

E-mail: ts24810017p31@cumt.edu.cn