

# 基于标签传播引导和区域自适应集成的 代理辅助进化算法

李二超, 崔添超

(兰州理工大学电气工程与信息工程学院, 甘肃 兰州 730050)

**[摘要]** 代理辅助进化算法(surrogate-assisted evolutionary algorithms, SAEAs)已被广泛用于解决计算代价昂贵的优化问题。针对在计算资源有限的条件下如何提高代理模型预筛选解的能力和效率的问题,提出了一种基于标签传播引导和区域自适应集成的代理辅助进化算法,分为全局和局部搜索两个阶段,在全局搜索阶段提出了多重筛选多点填充准则(multiple screening criteria, MSC),首先利用标签传播思想代替传统分类方法更高效地将种群分类,用两种代理模型预测分类后种群的适应度值,根据父代和子代的最佳适应度再次筛选种群进行评估;局部搜索阶段利用 SMOTE(synthetic minority over-sampling technique)方法生成动态局部种群和测试样本,根据两种代理模型在最有希望区域内的表现进行自适应集成来选择个体进行评估。最后将所提出方法与其他先进的代理辅助进化算法在 8 个测试问题及翼型设计优化问题中进行了对比,显示本文方法有较好的收敛性。

**[关键词]** 数据驱动,代理模型,标签传播,自适应集成,填充准则

**[中图分类号]** TP391 **[文献标志码]** A **[文章编号]** 1672-1292(2024)04-0001-16

## Surrogate-Assisted Evolution Algorithm Based on Label Propagation Guidance and Region Adaptive Integration

Li Erchao, Cui Tianchao

(College of Electrical Engineering and Information Engineering, Lanzhou University of Technology, Lanzhou 730050, China)

**Abstract:** Agent-assisted evolutionary algorithms have been widely used to solve computationally expensive optimization problems. In order to improve the ability and efficiency of the pre-screening solution of the agent model under the condition of limited computing resources, the paper proposes an agent-assisted evolution algorithm based on label propagation guidance and regional adaptive integration, which is divided into two stages: global search and local search. In the global search stage, the multiple screening criteria(MSC) is proposed. Two proxy models are used to predict the fitness values of the classified populations, and the populations are screened again according to the best fitness of the parents and offspring. The local search phase uses the SMOTE(synthetic minority over-sampling technique) method to generate dynamic local populations and test samples, select individuals for evaluation based on adaptive integration of the performance of the two proxy models in the most promising areas. Finally, the proposed method is compared with other advanced agent assisted evolution algorithms in 8 test problems and airfoil design optimization problems, showing that the proposed method has better convergence.

**Key words:** data-driven, surrogate model, label propagation, adaptive integration, filling criterion

传统的进化算法(evolutionary algorithms, EAs)基于一个隐含的假设,即对目标函数或约束的计算是廉价和快速的,然而在许多现实世界的优化问题中,如汽车碰撞安全设计<sup>[1]</sup>、神经网络架构搜索<sup>[2]</sup>,这些问题适应度评估(fitness evaluation, FE)会消耗大量的资源或时间成本,这类问题被称为昂贵优化问题<sup>[3]</sup>。为了解决上述昂贵优化问题,研究人员构建了具有低计算成本的代理模型来帮助 EAs。代理模型使用已有数据来近似真实的 FE,从而降低评估代价,基于代理模型算法被称为数据驱动的进化算法<sup>[4]</sup>或代理辅助的

收稿日期:2024-05-12.

基金项目:国家自然科学基金项目(62063019)、甘肃省自然科学基金重点项目(24JRRA173)、甘肃省优秀博士生项目(24JRRA205)。

通讯作者:李二超,博士,教授,研究方向:智能优化理论、方法及应用。E-mail:lecstar@163.com

进化算法. 这类算法在许多实际问题中得到了广泛的应用,如翼型设计优化<sup>[5]</sup>、创伤系统优化<sup>[6]</sup>.

现有的代理模型可分为基于回归的代理和基于分类的代理. 基于回归的代理,如径向基函数(radial basis function networks, RBFN)<sup>[7]</sup>、克里金模型(Kriging)(也称高斯过程模型)<sup>[8]</sup>、多项式回归模型<sup>[9]</sup>,已被广泛地用于近似目标函数. 然而,这些模型可能很容易停滞在复杂的区域<sup>[9]</sup>. 基于分类的代理辅助算法使用分类模型,如支持向量机(SVM)<sup>[10]</sup>、随机森林(random forest, RF)<sup>[11]</sup>以及 K 最近邻(KNN)<sup>[12]</sup>,可以预测新生成候选解的类别<sup>[13]</sup>,然后从良好的类别中选择好的个体进行真实 FE. 通常基于回归的代理比基于分类的代理显示出更好的性能.

常用的代理模型也可分为全局模型、局部模型和集成模型<sup>[14]</sup>. 全局模型通常使用所有历史数据建模在整个空间中进行搜索,用于在一个大的搜索空间中探索有希望的样本点. 局部模型通常采用部分样本建模来逼近某些重要区域,有助于捕获有希望区域(围绕最优区域)的局部细节. 集成代理已被证明在准确性和鲁棒性<sup>[15]</sup>方面比单个学习者有优势. 由于不同类型的代理模型适合不同特性的问题,集成模型可用于提高代理模型对不同类型问题的预测精度,或用于获取预测的不确定性. Wang 等<sup>[16]</sup>引入一种基于集成的模型管理策略,结合了不确定性和基于性能的指标来进行代理辅助进化优化. Li 等<sup>[17]</sup>提出一种基于扰动的集成代理的方法,该方法首先基于数据扰动生成一组模型,然后使用并选择一些现有的模型形成最终的集成模型,以提高模型的有效性和效率.

虽然现有工作提出了许多填充准则,但仍然面临许多挑战. 其一是代理模型的选择<sup>[17]</sup>,有研究表明不同的代理模型适用于不同的问题,但在优化开始前并不知道这种问题的特性,故在无先验信息的情况下很难选择一个合适的代理模型. 本文选择 RBFN 和增量克里金模型(Ikriging)这两种具有代表性且高效的模型作为代理, RBFN 可以保证在较好地近似高维非线性函数的同时,其训练时间不会随着维数的增加而显著提高<sup>[18]</sup>. 当样本数或维度较大时传统的 Kriging 的训练十分耗时, Zhan 等<sup>[19]</sup>将增量学习用于建模过程,提出了 Ikriging,在不影响精度的前提下显著提高了训练速度. 其二是在已有数据量和真实评估次数有限的情况下,代理通常很难准确地近似目标函数,特别当决策空间维度上升时,模型很容易在复杂的情况下误导种群进化的方向. 其三,对于不确定性信息通常基于个体间距离信息<sup>[20]</sup>,或基于概率模型(如克里金模型)提供的不确定度,或基于集成模型的基模型的预测方差来表示,本文从另一角度考虑模型的不确定性来筛选解.

针对上述问题,本文首先增强模型预选解的可靠性,设计了一个多重筛选多点准则(MSC)来选择个体进行真实 FE. MSC 使用基于标签传播(label propagation, LP)技术在不增加真实计算成本的情况下标记种群信息,该方法比其他分类模型耗时更短且不易陷入数据不平衡问题. 同时利用 RBFN 和 Ikriging 协同预测标签传播后的种群来进行多点填充,帮助算法真实评估当前部分最优的个体. 其次,考虑到代理模型的作用和常规机器学习模型的任务不同,提出使用 3 个不同的代理模型评估指标来对 RBFN 和 Ikriging 分别进行评估,并根据计算指标大小来对两种模型动态赋予不同的权重进行自适应集成,提升代理模型对于局部最优区域的预测能力和鲁棒性. 最后,提出基于合成少数类样本的过采样技术 SMOTE 在决策空间中最有希望的区域生成小分子种群,加强种群对于当前最优区域的搜索能力,同时全局种群和局部种群相互独立地进行搜索,在一定条件下在全局搜索和局部搜索间进行切换,同时可以缩小庞大的搜索空间,帮助算法加速收敛.

## 1 相关技术

### 1.1 SMOTE

Chawla 等<sup>[21]</sup>提出了一种合成少数类样本的过采样技术 SMOTE, SMOTE 是一类数据预处理技术,其关键思想是通过合成样本来丰富少数类别的数据分布,以减少类别不平衡问题对机器学习模型的影响. SMOTE 有助于提高模型的性能,其算法流程如下<sup>[22]</sup>:

(1) 对于少数类中每一个样本  $x$ , 以欧氏距离为标准计算其到少数类样本集中所有样本的距离,得到其  $k$  个近邻样本. 随机选择一个少数类样本作为邻居.

(2) 对于已选样本  $x$  和随机选择的邻居  $\tilde{x}$ , 根据以下步骤生成一个合成样本  $x_{\text{new}}$ : 对于每个特征维度  $x = (x_1, x_2, \dots, x_{d-1}, x_d)$ , 计算邻居样本和已选样本之间的差值; 从  $[0, 1]$  的均匀分布中随机选择一个权重因子; 使用权重因子乘以差值, 然后添加到已选样本上, 得到一个新的合成样本  $x_{\text{new}}$ :

$$x_{\text{new}} = x + \text{rand}(0, 1) * (\tilde{x} - x). \quad (1)$$

(3) 将生成的合成样本添加到原始数据集中.

(4) 重复执行步骤(2),直至生成足够数量的合成样本,以达到指定的过采样倍数.

## 1.2 标签传播(LP)

Zhu 等<sup>[23]</sup>提出了一种基于图的半监督学习方法 LP,其基本思路是用已标记样本的标签信息去预测未标记样本的标签信息.利用样本间的关系建图,节点包括已标注和未标注数据,其边表示两个节点的相似度,节点的标签按相似度传递给其他节点.标签数据就像一个源头,可以对无标签数据进行标注,节点的相似度越大,标签越容易传播<sup>[24]</sup>.其具体步骤如下:

(1) 令  $(x_1, x_2, \dots, x_l)$  为标注样本,  $(y_1, y_2, \dots, y_l) \in \{1 \dots C\}$  是类别标签,类别数  $C$  已知,且均存在于标签数据中.令  $(x_{l+1}, x_{l+2}, \dots, x_{l+u})$  为未标注样本,  $(y_{l+1}, y_{l+2}, \dots, y_{l+u})$  不可观测,令样本集  $x = (x_1, x_2, \dots, x_l, x_{l+1}, x_{l+2}, \dots, x_{l+u}) \in \mathbf{R}^d$ .

(2) 将所有样本集中的样本作为节点(包括已标注和未标注数据),创建一个完全连接图,其边的权重计算式为:

$$w_{ij} = \exp\left(-\frac{d_{ij}^2}{\sigma^2}\right) = \exp\left(-\frac{\sum_{d=1}^D (x_i^d - x_j^d)^2}{\sigma^2}\right), \quad (2)$$

式中,  $d_{ij}$  表示两个节点的欧氏距离;权重  $w_{ij}$  受控于参数  $\sigma$ .

(3) 为衡量一个节点的标签通过边传播到其他节点的概率,定义一个  $(l+u) * (l+u)$  概率传递矩阵  $T$ :

$$T_{ij} = P(j \rightarrow i) = \frac{w_{ij}}{\sum_{k=1}^{l+u} w_{kj}}, \quad (3)$$

式中,  $T_{ij}$  是  $j$  到  $i$  的传播概率.

(4) 同时定义一个  $(l+u) * C$  的标注矩阵  $Y$ ,令  $Y_{ic} = \delta(y_i, c)$ ,其第  $i$  行代表节点  $y_i$  的传播概率,第  $c$  列代表类别,每个节点按传播概率将其周围节点传播的标注值按权重相加,并更新自己的概率分布:

$$F_{ij} = \sum_{k=1}^{l+u} T_{ik} Y_{kj}, \quad 1 \leq i \leq l+u, 1 \leq j \leq C. \quad (4)$$

(5) 限定已标注样本,把已标注样本的概率分布重新赋值为初始值.重复步骤(4),直至收敛.注意保持已标注数据点的标注源不变,将其值限定为  $Y_l$ ,不断地把标注从高权值传播到低权值:

$$F_{ij} = Y_{ij}, \quad 1 \leq i \leq l, 1 \leq j \leq C. \quad (5)$$

LP 只需利用少量的训练标签指导,利用未标注数据的内在结构、分布规律和近邻数据的标记,即可预测和传播未标记数据的标签,然后合并至标记的数据集中.

## 1.3 增量克里金模型

标准克里金模型的三阶时间复杂性  $O(n^3)$  限制了其在高维优化中的应用( $n$  为训练样本的个数),为了解决这个问题,文献[19]提出了一个高维代理辅助进化计算的增量克里金模型.主要思想是基于之前训练过的模型的方程逐步更新克里金模型,而不是在有新样本时从头开始构建模型,从而将更新克里金模型的时间复杂度降低到二阶  $O(n^2)$ ,所提出的增量学习方案非常适合在线数据驱动算法.

## 1.4 径向基函数网络

本文还使用 RBFN 作为代理模型. RBFN 是应用最广泛的代理模型之一<sup>[25]</sup>,具有较好的高维非线性函数逼近能力,其训练时间对维数的增加不敏感. RBFN 通常只有 3 层:输入层、隐藏层和输出层.给定  $N$  个训练数据  $\{(x_i, y_i) | x_i \in \mathbf{R}^d, i=1, \dots, N\}$ ,用其作为 RBFN 的中心点. RBFN 模型如下:

$$\hat{f}(x) = \sum_{i=1}^N \omega_i \phi(\|x - x_i\|), \quad (6)$$

式中,  $\phi(\cdot)$  表示核函数;  $\omega_i$  为第  $i$  个核函数的权重系数;  $\|\cdot\|$  表示欧几里得范数.常用的核函数有高斯函数、逆 Sigmoid 函数、多二次函数和逆多二次函数等,本文使用的是高斯核函数.权重向量  $\omega = (\omega_1, \dots, \omega_N)^T$  表示为:

$$\omega = \Phi^{-1} F, \quad (7)$$

式中,核矩阵  $\Phi = [\varphi(x_i - x_j)]_{N \times N}$ .

### 1.5 社会学习粒子群(SL-PSO)

Cheng 等<sup>[26]</sup>提出了 SL-PSO. SL-PSO 的社会学习机制是向种群中随机选择的比自己更好的粒子学习,去除了粒子群算法向全局最佳和个人最佳位置学习的机制. SL-PSO 在低维和高维问题上都有很好的表现,这主要归因于其逃离局部最优的能力.

SL-PSO 按照适应度大小对种群进行降序排序,最好的粒子保持不动,其余粒子通过下式更新:

$$\begin{aligned} \Delta X_{i,j}(t+1) &= r_1(t) \cdot \Delta X_{i,j}(t) + r_2(t) \cdot I_{i,j}(t) + r_3(t) \cdot \varepsilon \cdot C_{i,j}(t), \\ \begin{cases} I_{i,j}(t) = X_{k,j}(t) - X_{i,j}(t), \\ C_{i,j}(t) = \bar{X}_j(t) - X_{i,j}(t). \end{cases} \end{aligned} \quad (8)$$

式中,  $r_1(t)$ 、 $r_2(t)$  和  $r_3(t)$  为 3 个随机系数,  $i$  代表第  $i$  个粒子,  $j$  代表第  $j$  维的决策变量;  $\Delta X_{i,j}(t)$  类似于 PSO 中的惯性向量;  $I_{i,j}(t)$  代表粒子  $i$  的第  $j$  维向粒子  $k$  的第  $j$  维学习, 粒子  $i$  的第  $j$  维可以向任意的粒子  $k$  ( $i < k \leq m$ ) 的第  $j$  维学习;  $C_{i,j}(t)$  表示粒子  $i$  向整个粒子群集体学习, 表示种群的平均水平;  $\varepsilon$  为社会影响因子, 用来控制向种群平均水平学习的大小. 此外, 粒子的学习方式如式(9)所示还为每个粒子  $i$  设置了一个学习率( $P_i^L$ ), 当随机数  $p_i(t)$  小于等于学习率时粒子才会进行学习,  $m$  为种群中粒子数量,  $M$  为粒子基础数量,  $d$  为维数:

$$X_{i,j}(t+1) = \begin{cases} X_{i,j}(t) + \Delta X_{i,j}(t+1), & \text{if } p_i(t) \leq P_i^L; \\ X_{i,j}(t), & \text{otherwise.} \end{cases} \quad (9)$$

$$P_i^L = \left(1 - \frac{i-1}{m}\right)^{\alpha \cdot \log(\lceil d/M \rceil)}, \quad (10)$$

$$m = M + \left\lceil \frac{d}{10} \right\rceil. \quad (11)$$

## 2 基于标签传播引导和区域自适应集成的代理辅助进化算法(LPMAE)

### 2.1 算法框架

全局搜索与局部搜索相结合可以同时保证探索和开发的有效性, 本文所提算法也可分为全局和局部两个阶段.

在全局搜索阶段, 设计了一个多重筛选多点填充准则(MSC), 充分利用模型本身的不确定性和已有信息来指导优化方向. 为了保证填充准则的多样性, 同时使用 EI 准则并行采样, 提高了算法的探索能力. 在全局搜索进展不佳时, 切换至局部搜索.

在局部搜索阶段, 为了提高局部搜索效率, 使用 SMOTE 方法在最优区域内合成局部种群和测试样本. 测试样本是为了检验代理模型在局部最优区域预测最优解的能力, 使用 3 个不同的评价指标来计算 RBFN 和 Ikriging 对应的指标大小, 并根据计算的指标值大小来对两个模型进行自适应加权集成. 为了不影响全局搜索进程, 全局种群和局部种群相互独立地进行搜索, 并以一定条件进行切换, 同时保证探索和开发.

所提算法框架如图 1 所示, LPMAE 的伪代码如算法 1 所示.

#### 算法 1 LPMAE 框架

Input: 决策变量维度  $d$ ; 目标函数  $f$ ; 种群大小  $m$ ; 初始数据量  $k$ ; 最大迭代次数 MAXNFE;

Output: 最优解  $p$ ; 最优值  $f_p$ ;

1. 拉丁超立方采样生成  $k$  个初始样本, 并使用目标函数评估后放入数据库 DB 中;
2. 选择 DB 中适应度值最好的  $M$  个样本作为初始种群;
3. While NFE < MAXNFE
4.     使用 DB 中所有数据建立一个 RBFN 和一个 Ikriging;
5.     SLPSO 进行种群迭代;
6.     MSC 策略采样候选解(详见算法 2);
7.     If 超过 20 代 DB 中的最优值没有改变

8. While DB 中的最优值变化连续 3 代小于  $1E-2$
9. 开始局部搜索阶段(详见算法 3)
10. 输出最优解  $p$  和最优值  $f_p$ ;

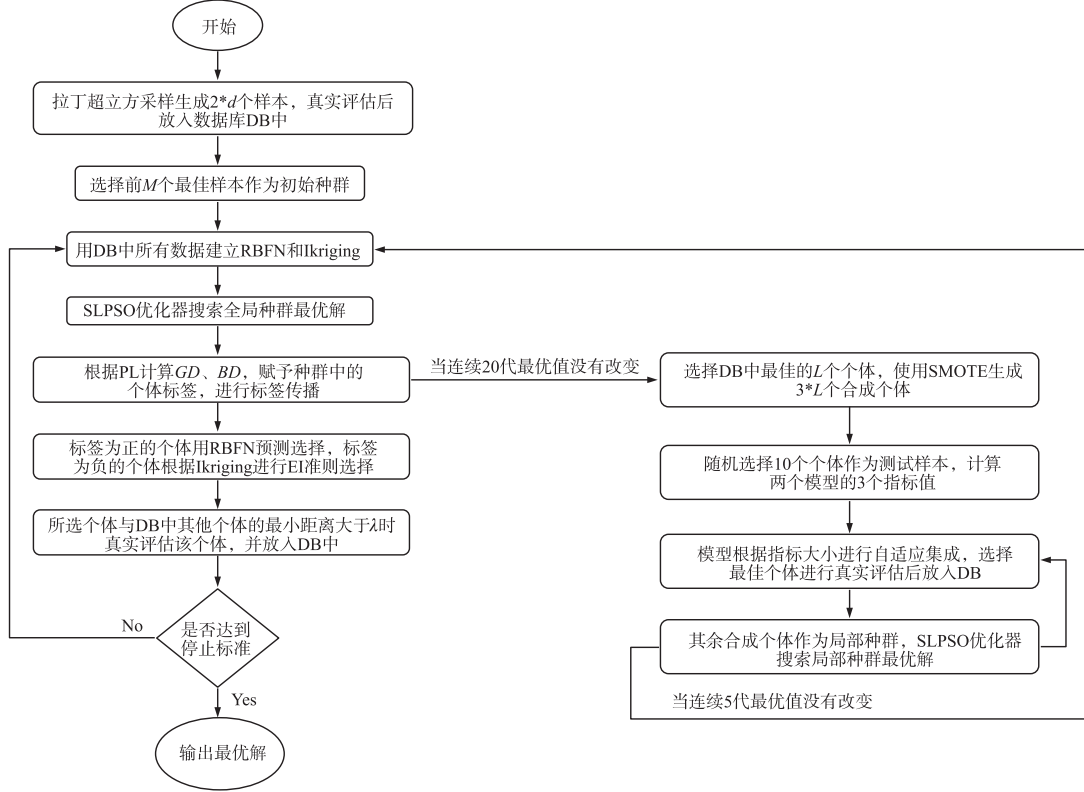


图 1 LPMAE 流程图

Fig. 1 LPMAE flow chart

## 2.2 多重筛选多点填充准则(MSC)

对于在线的代理辅助进化算法,最重要的步骤之一就是选择候选解进行真实评估. 该步骤很大程度影响了模型指导优化的能力,因为当真实评估全局最优区域的解时,可以提高代理模型在最优区域的精度,进而提高找到全局最优解的能力. 适当评估部分较不确定的解可以促进采样的多样性,但当评估过多较差的候选解或局部最优候选解时,不能促进算法的收敛且容易误导种群停滞在局部最优区域.

MSC 伪代码如算法 2 所示. 首先将数据库中的所有样本按适应度值排序,选择前  $GD$  个样本赋予正类标签,从第  $2 * GD$  个样本到最后一个样本中任意选择  $BD$  个样本赋予负类标签,负类从第  $2 * GD$  个样本开始选择的目的是为了保证正负类传播源间有一个相对明显的界限,因为对于连续问题,适应度值接近时个体间距离也有可能很接近. 将带标签样本和种群放入传播集按照式(2)-(5)进行传播,用 RBFN 预测种群中标签为正的个体的适应度,并按预测值排序. 正类标签数目  $GD$  由式(12)确定:

$$\begin{cases} GD = GD + 2 + \frac{d}{50}, & BD = BD - 5 - \frac{d}{50}, & PL = 0; \\ GD = GD + 2 + \frac{d}{50}, & BD = BD - 2 - \frac{d}{50}, & 0 < PL \leq 0.1; \\ GD = GD, & BD = BD, & 0.1 < PL \leq 0.5; \\ GD = GD - 2 - \frac{d}{50}, & BD = BD + 2 + \frac{d}{50}, & 0.5 < PL \leq 0.8; \\ GD = GD - 5 - \frac{d}{50}, & BD = BD + 5 + \frac{d}{50}, & 0.8 < PL \leq 1. \end{cases} \quad (12)$$

选择预测值最好的  $N$  个个体依次进行判定,将预测值优于上一代预测全局最优值的个体放入下步判定. 计算该个体与其他个体的欧氏距离,若距离大于阈值  $\lambda$ ,则对该个体进行真实评估后放入数据库中. 计



算正类个体在种群中所占的比例  $PL$ , 根据  $PL$  动态调整  $GD$  和  $BD$  的大小; 希望在每一代种群中, 正类个体保持在一定范围内, 因为  $PL$  过小说明种群进化方向没有向当前最优区域迭代, 而  $PL$  过大说明种群过于聚集, 不利于将种群分类. 通过该策略可以有效动态划分种群, 过滤掉较差的解, 提高筛选最佳个体的概率. Ikriging 使用 EI (expected improvement) 准则进行填充采样, EI 准则是克里金预测函数和标准误差函数之间的非线性组合:

$$EI(x) = (y_{\min} - \hat{y}(x)) \Phi\left(\frac{y_{\min} - \hat{y}(x)}{\hat{s}(x)}\right) + \hat{s}(x) \phi\left(\frac{y_{\min} - \hat{y}(x)}{\hat{s}(x)}\right), \quad (13)$$

式中,  $y_{\min}$  是数据库中所有样本的最小函数值;  $\phi(x)$  和  $\Phi(x)$  分别是标准正态分布的密度和累计分布函数. EI 准则可以帮助在探索和开发之间实现平衡, 其不仅考虑了当前估计的最佳值, 还考虑了不确定性, 这意味着其有机会填充决策空间中不确定区域内的个体. 单填充准则仅采用一个固定的填充标准, 这会使得算法的泛化能力较差; 多填充准则可以提高填充解的多样性, 能够更好地实现开发和勘探之间更好的平衡. 为了避免因为代理模型的误差而丢失当前好的采样点, 本文采用多点填充准则, 将满足上述条件的个体全部进行真实评估. MSC 策略的伪代码如算法 2 所示.

#### 算法 2 MSC 框架

输入: 数据集 DB; 种群 pop; 目标函数  $f$ ; 代理模型 RBFN 和 Ikriging; 正类标签数目  $GD$ ; 负类标签数目  $BD$ ; 上一代预测全局最优值  $f_{gp}$ ; 距离阈值  $\lambda$ ;

输出: 数据集 DB; 正类标签数目  $GD$ ; 负类标签数目  $BD$ ;

1. 将 DB 中的样本按适应度值升序排序;
2. 分别赋予  $GD$  个样本正类标签、 $BD$  个样本负类标签;
3. 将带标签样本和种群放入传播集中, 按公式进行标签传播;
4. 用 RBFN 预测种群中标签为正的个体;
5. a: 选择  $N$  个预测值最佳的个体依次进行判定;
6. b: 计算这  $N$  个个体与数据库中其他样本的欧氏距离, 距离大于  $\lambda$  的个体进入下步判定;
7. c: 当预测值小于  $f_{gp}$  时进行真实评估后放入 DB 中;
8. 用 Ikriging 预测负类个体, 使用式 (12) 选择一个个体进行真实评估后放入 DB 中;
9. 计算  $PL$ , 并且根据公式更新  $GD$  和  $BD$  的值;

### 2.3 局部区域搜索策略

#### 2.3.1 局部种群生成策略

当全局搜索进展不佳时, 说明算法陷入收敛困难的窘境. 为了促进算法的收敛性, 在当前数据库中最优的个体附近生成新的局部种群进行局部区域搜索, 来加强算法对当前最优区域的开发. 在连续优化问题中, 通常距离当前最优解较近的解是最有潜力的一类解之一; 在单目标优化问题中最优解通常只有一个; 同样在多目标优化问题中, 帕累托最优解集也通常占有所有解中的较少部分; 所以在每一代中适应度值较好的解和较差解可以作为少数类样本和多数类样本. 因而, 本文认为解集的分布类似于一种数据不平衡问题. 为了在不影响全局搜索的情况下促进当前最优区域的开发, 首先在进入局部搜索阶段时全局种群暂停在其本身位置, 挑选数据库中适应度最好的  $L$  个样本, 对于  $L$  中的每个样本从其  $k$  个最近邻样本中随机挑选一个作为邻居样本, 使用式 (1) 生成合成样本. 合成样本在当前最优区域附近产生, 有两种用途: 一是作为局部种群, 提高在当前最优区域搜索能力; 二是作为测试样本, 为下一步局部自适应集成做准备. 当合成样本生成后, 随机选择  $L$  个样本进行真实评估后作为测试样本, 剩余合成样本与数据库中挑选出的  $L$  个样本共同作为局部种群在当前最优潜力的区域搜索, 该区域的上界  $ub$  和下界  $lb$  由下式决定:

$$\begin{cases} lb = [l_1, l_2, \dots, l_d], \\ ub = [u_1, u_2, \dots, u_d]. \end{cases} \quad (14)$$

$$\begin{cases} l_j = \min(x_j^i), \\ u_j = \max(x_j^i), \end{cases} \quad i = 1, 2, \dots, P.$$

第  $j$  维的上下界由  $L$  个最佳样本在第  $j$  维的最大和最小值确定. 合成的样本是在任意最近的两个最佳

样本之间产生的,所以合成样本均在该区域内。

### 2.3.2 区域自适应集成策略

全局模型是基于所有数据建立的模型,由于已评估数据的数量有限,所以模型无法在整个搜索空间准确地映射目标值,且两个全局模型在不同区域找到最优解的能力也不同。在每次进入局部搜索时会根据当前区域生成新的测试样本,测试样本用于计算两个模型在以下 3 个指标上的性能:

#### (1) 确定系数( $R^2$ )

确定系数可以测量代理模型的准确性,其计算式为:

$$\begin{aligned} R^2 &= 1 - \frac{\text{MSE}}{\text{Var}}, \\ \text{MSE} &= \frac{1}{N} \sum_{i=1}^N (\varphi_i^{\text{orig}} - \varphi_i^{\text{model}})^2, \\ \text{Var} &= \frac{1}{N} \sum_{i=1}^N (\varphi_i^{\text{orig}} - \bar{\varphi}^{\text{orig}})^2. \end{aligned} \quad (15)$$

式中, $\varphi^{\text{orig}}$ 是  $N$  个测试样本的原始适应度值; $\varphi^{\text{model}}$ 是  $N$  个测试样本的模型预测值; $\bar{\varphi}^{\text{orig}}$ 和  $\text{Var}$ 是  $N$  个原始适应度值的均值; $\text{MSE}$ 为原始适应度函数与代理模型之间的均方误差。一般来说, $R^2$ 在  $[0, 1]$  之间,当模型越准确时  $R^2$  的值越大,反之亦然,当  $R^2 = 1$  时,表示模型完美拟合。

#### (2) 适应度相关指标(continuous correlation, CC)

适应度相关指标可以度量测试样本在原始适应度值和代理模型预测值的相关性,计算式为:

$$\rho^{(\text{corr.})} = \frac{\frac{1}{N} \sum_{j=1}^N (\varphi_j^{\text{model}} - \bar{\varphi}^{\text{model}})(\varphi_j^{\text{orig}} - \bar{\varphi}^{\text{orig}})}{\sigma^{\text{model}} \sigma^{\text{orig}}}, \quad (16)$$

式中, $\bar{\varphi}^{\text{model}}$ 和  $\bar{\varphi}^{\text{orig}}$ 分别为模型预测值和原始适应度值的均值; $\sigma^{\text{model}}$ 和  $\sigma^{\text{orig}}$ 分别为预测值和原始适应度值的方差。 $\rho^{(\text{corr.})}$ 的取值范围是  $[-1, 1]$ ,当  $\rho^{(\text{corr.})}$ 越靠近 1 时,代表模型和原始函数的趋势越一致。

#### (3) 排序保存指标(ranking preservation, RP)

代理模型和原始适应度函数之间的排序保存指标是通过测试样本对之间的排序比较来计算的,可用以测量代理模型与原始适应度函数保持相同样本点等级的能力。与适应度相关指标相似,RP 值较大,表明代理模型和原始适应度函数之间具有高度一致的轮廓。数学上,给定  $N$  个测试样本,RP 的计算式为:

$$\begin{aligned} \text{RP} &= \left( \sum_{i=1}^N \sum_{j=i+1}^N h(i, j) \right) / \binom{N}{2}, \\ h(i, j) &= \begin{cases} 1 & \text{if } (\varphi_i^{\text{orig}}(x_i) < \varphi_j^{\text{orig}}(x_j) \wedge \varphi_i^{\text{model}}(x_i) < \varphi_j^{\text{model}}(x_j)) \\ & \vee (\varphi_i^{\text{orig}}(x_i) > \varphi_j^{\text{orig}}(x_j) \wedge \varphi_i^{\text{model}}(x_i) > \varphi_j^{\text{model}}(x_j)) \\ & \vee (\varphi_i^{\text{orig}}(x_i) = \varphi_j^{\text{orig}}(x_j) \wedge \varphi_i^{\text{model}}(x_i) = \varphi_j^{\text{model}}(x_j)) \\ 0 & \text{otherwise.} \end{cases} \quad \forall x_i, x_j \in \mathbf{R}^d. \end{aligned} \quad (17)$$

式(17)给出了在任意两个样本点上的原始适应度函数与代理模型的适应度比较关系的 3 种情况,一旦满足这 3 个条件之一,RP 的值就会增加;否则,将保持不变。RP 的取值范围为  $[0, 1]$ 。

仅关注上述的其中一个指标,并不能全面评估一个模型的性能。由于同一模型在不同评价指标的评价结果不同,导致模型评价结果存在一定偏差。因此本文设计了自适应集成策略综合考虑上述 3 个指标来最大化确保模型选出最优个体的概率,其中 RBFN 权重计算如式(18),Ikriging 权重计算如式(19),最终集成输出如式(20):

$$\omega_1 = \left( \frac{R_1^2}{R_1^2 + R_2^2} + \frac{\rho_1^{(\text{corr.})}}{\rho_1^{(\text{corr.})} + \rho_2^{(\text{corr.})}} + \frac{\text{RP}_1}{\text{RP}_1 + \text{RP}_2} \right) * \sigma, \quad (18)$$

$$\omega_2 = \left( \frac{R_2^2}{R_1^2 + R_2^2} + \frac{\rho_2^{(\text{corr.})}}{\rho_1^{(\text{corr.})} + \rho_2^{(\text{corr.})}} + \frac{\text{RP}_2}{\text{RP}_1 + \text{RP}_2} \right) * \sigma, \quad (19)$$

$$\text{output} = \text{RBFN} * \omega_1 + \text{Ikriging} * \omega_2. \quad (20)$$

算法 3 区域自适应集成策略

输入:数据集 DB;种群 pop;目标函数 $f$ ;代理模型 RBFN 和 Ikriging;预测全局最优值 $f_{gp}$ ;距离阈值 $\lambda$ ;
输出:数据集 DB;
1.选择数据库中前 $L$ 个个体,用 SMOTE 方法生成 $3 * L$ 个新个体;
2.随机选择其中 10 个个体进行真实评估并作为测试样本;
3.在测试样本中计算两个模型的 3 个指标值;
4.利用式(18)计算权重大小进行集成;
5.选择集成预测的适应度最佳个体进行真实评估;

3 实验与结果分析

为了验证本文所提算法的有效性,将实验分为 4 个部分:(1)MSC 策略有效性分析;(2)自适应集成策略的有效性分析;(3)与 4 种先进的代理辅助进化算法进行对比,验证所提算法的有效性;(4)翼型设计优化问题. 采用 8 个包含单峰和多峰的基准问题进行测试,表 1 给出了这些问题的具体特性.

表 1 基准测试问题  
Table 1 Benchmark questions

基准问题	搜索范围	特点	全局最优值
Ellipsoid	$[-5.12, 5.12]^d$	单峰	0
Rosenbrock	$[-2.048, 2.048]^d$	多峰,狭长谷形	0
Ackley	$[-32.768, 32.768]^d$	多峰	0
Griewank	$[-600, 600]^d$	多峰	0
Rastrigin	$[-5.12, 5.12]^d$	多峰	0
SRR	$[-5, 5]^d$	非常复杂多峰	-330
RHC1	$[-5, 5]^d$	非常复杂多峰	120
RHC2	$[-5, 5]^d$	非常复杂多峰	10

3.1 参数设置

在实验部分,每个算法都进行了 20 次独立运行,终止条件是消耗的 FE 小于最大评估次数 MAXNFE=1 000. LPMAE 的参数设置为:种群大小  $m=50$ ;初始数据量  $k=100$ (参考文献[27]中的设置);每次选择数据库中的最优个体数  $N=10$ ,距离阈值参考文献[27]中的设置  $\lambda = [\text{sqrt}(0.001^2d), 5.0E-4d * \min(ub-lb)]$ . 为避免真实评估的解距离过近,在选择 RBFN 预测正类个体值中适应度值排名前  $s$  个个体判定时,随着维数  $d$  的上升,代理模型拟合原始函数的能力在下降<sup>[28]</sup>,模型预测排名和个体真实排名之间的差距也在进一步扩大,故随着  $d$  增大,为了最大限度保证能够真实评估当前代的最佳个体,能够在因维数升高时给予该策略更多的容错空间,同时判定这些个体又需要消耗部分时间,为了兼顾搜索效果和搜索效率,当  $d<100$  时设置  $s=10$ ,当  $d\geq 100$  时设置  $s=20$ . 局部搜索时选择  $L=10$  个最优个体, $k$  近邻数目可由用户自定义,本文设置  $k=2$ ;自适应集成中指标分配权重设置为  $\sigma=1/3$ ,该参数是影响自适应集成策略效果的关键参数.

3.2 MSC 有效性验证

为了验证全局填充准则的有效性,本实验设计了以下几种变体进行对比:(1)不使用 LP,仅使用多点填充和 EI 准则 LPMAE-NOLP;(2)仅使用 EI 准则的高维代理辅助的增量克里金算法(IKAEA);(3)使用 MSC 策略,但不使用 EI 准则(LPMAE-NOEI);(4)使用支持向量机(SVM)代替 LP(LPMAE-SVM);(5)使用随机森林(RF)代替 LP(LPMAE-RF).

运行结果如表 2 和图 2 所示,可以看到 LPMAE 在所有测试函数的 4 个维度中基本都表现最好. 在不使用 LP 的情况下,算法的收敛性能有不同程度的下降,原因可能是 LP 帮助过滤掉了一些适应度较差的解,除了在 Griewank 测试函数中,这可能是因为在 Griewank 中的搜索范围是最大的,再加上当维数升高时,决策空间再次呈指数级增长<sup>[29]</sup>,以及数据库中样本数与种群数目不大,导致评估样本和种群中的个体在整个决策空间中过于稀疏. 随着迭代的进行,种群逐渐收敛,可以让带标签样本顺利传播,赋予所有在有效传播范围内的个体新的标签. 当使用 SVM 和 RF 代替 LP 时,算法性能显著降低,但略强于 IKAEA 性



能,可能是因为单目标优化通常只有一个最优解,所以在迭代过程中当前较优个体的数量会比较差个体数目少,分类模型陷入数据不平衡的困扰,导致模型无法提取到较优个体的特征,故无法将适应度好的和差的个体准确分类. 通过动态更新正类和负类个体的数量可以有效缓解这一问题. 训练数据设置为数据库中适应度前 10% 的个体为正类、后 50% 的个体为负类,以及适应度前 10% 的个体为正类、从剩余个体值随机选择相同数量的个体为负类,这两种情况的结果均不能令人满意,前者会面临数据不平衡问题,后者会导致训练数据量过小而无法充分利用所有数据. 使用标签传播可有效避免上述问题,可以看到 LPMAE 在所有维度上均有良好的表现;当不使用 EI 时算法的收敛性能也会下降,原因可能是以下两方面,其一是多填充策略比单填充策略的效果更好,促进了采样的多样性<sup>[30]</sup>;其二是 EI 准则影响了 Ikriging 的性能,导致在自适应集成阶段不能发挥出应有的作用. 在 RF 中决策树的数量设置为 50,由于 LP 不需要建模,在时间成本方面更加优于分类模型.

表 2 LPMAE-NOLP, IKA EA, LPMAE-NOEI, LPMAE-SVM, LPMAE-RF, LPMAE 独立运行 20 次的均值与方差

Function	<i>d</i>	LPMAE-NOLP	IKAE A	LPMAE-NOEI	LPMAE-SVM	LPMAE-RF	LPMAE
Ellipsoid	30	4.98E-08(5.35E-07)	4.00E-03(0.00E+00)	2.39E-05(1.68E-03)	6.98E-01(5.36E+00)	2.36E-03(2.96E-02)	8.69E-09(3.15E-06)
	50	9.35E-06(1.35E-07)	7.07E+00(3.34E+00)	5.17E-05(3.52E-02)	3.79E+01(2.13E+01)	9.69E-03(2.42E-01)	3.65E-10(95.6E-10)
	100	1.24E-01(8.94E-05)	6.40E+00(1.49E+02)	6.51E-01(4.61E-01)	5.45E+01(3.90E+01)	1.20E-02(6.54E+00)	1.61E-03(6.58E-09)
	200	8.87E+02(3.58E+01)	1.97E+04(3.60E+03)	9.98E+02(2.68E+01)	NA	3.56E+04(5.10E+01)	1.78E+01(2.45E+02)
Rosenbrock	30	4.15E+01(1.59E+00)	5.34E+01(2.51E+01)	5.98E+01(2.65E+00)	2.01E+02(4.26E+01)	8.78E+01(0.99E+00)	2.21E+01(1.39E-01)
	50	5.98E+01(3.67E+00)	2.18E+02(2.81E+01)	9.87E+01(3.62E+00)	5.25E+02(6.59E+01)	1.50E+02(6.71E+00)	3.86E+01(6.58E+00)
	100	1.57E+02(2.49E+01)	8.22E+02(6.52E+01)	3.79E+02(1.68E+02)	6.53E+02(7.52E+01)	2.46E+02(5.62E+01)	9.56E+01(1.46E+02)
	200	2.76E+02(9.57E+01)	5.22E+03(9.97E+02)	5.69E+02(2.49E+02)	1.53E+03(1.01E+01)	6.11E+03(9.10E+01)	1.62E+02(3.42E+02)
Ackley	30	6.58E-01(1.49E-01)	4.35E+00(6.70E-01)	3.28E-01(5.51E-01)	1.01E+00(1.93E+00)	2.83E+00(1.08E+00)	2.18E-01(3.14E-01)
	50	4.98E+00(5.48E-01)	7.15E+00(9.40E-01)	9.98E+00(5.69E-01)	6.17E+00(9.46E+00)	6.89E+00(3.80E-01)	2.90E+00(3.58E-01)
	100	5.53E+00(6.61E+00)	1.28E+01(5.80E-01)	1.17E+01(2.48E+00)	2.33E+00(1.34E+01)	3.91E+00(6.47E+00)	1.68E+00(6.49E+00)
	200	8.49E+00(8.69E+00)	1.55E+01(5.64E-01)	2.98E+01(6.59E+01)	1.75E+00(3.01E+00)	1.25E+01(4.52E+01)	5.46E+01(3.42E+00)
Griewank	30	6.92E-05(3.59E-02)	9.70E-01(5.64E-01)	8.92E-04(2.09E-01)	8.16E-02(1.10E+01)	5.63E-01(6.35E-01)	5.68E-05(2.21E-01)
	50	1.89E-03(3.24E-01)	1.17E+00(1.20E-01)	8.16E-03(5.73E-01)	7.03E-01(3.35E+01)	8.64E-01(1.03E+00)	8.98E-03(4.59E-01)
	100	2.47E-02(4.98E-01)	3.48E+01(1.12E+01)	6.38E-01(3.66E+00)	1.02E+00(3.87E+01)	6.15E-01(8.14E+00)	4.89E-02(5.49E+00)
	200	2.14E-01(1.25E+00)	7.98E+02(1.34E+02)	5.24E+00(1.83E+00)	2.39E+02(2.11E+01)	4.80E+02(2.14E+02)	3.09E-01(3.49E+00)
Rastrigin	30	7.48E+01(5.64E+00)	1.93E+02(2.30E+01)	8.39E+01(9.92E+00)	1.51E+02(1.04E+00)	1.46E+02(2.51E+00)	5.16E+01(2.99E+00)
	50	3.58E+02(1.59E+00)	2.21E+02(5.32E+01)	5.98E+02(1.12E+01)	3.62E+02(2.03E+01)	2.17E+02(8.49E+00)	1.28E+02(3.25E+01)
	100	8.47E+02(3.92E+00)	9.06E+02(6.39E+01)	9.35E+02(3.48E+01)	1.09E+03(3.39E+01)	9.11E+02(3.30E+02)	5.39E+02(2.34E+00)
	200	1.69E+03(4.58E+00)	1.58E+03(1.20E+02)	1.77E+03(5.69E+01)	2.38E+03(4.92E+01)	1.90E+03(1.23E+01)	9.86E+02(3.69E+01)
SRR	30	-1.58E+01(8.71E-01)	-2.85E+01(2.70E+00)	-1.15E+01(6.65E-01)	-8.42E+01(2.94E+00)	-1.01E+02(1.25E+00)	-2.33E+02(4.15E+00)
	50	-9.37E+00(1.02E+00)	2.61E+02(4.73E+01)	3.27E+00(2.15E+00)	1.06E+02(8.91E+00)	2.54E+01(3.52E+01)	-6.59E+01(1.24E+00)
	100	8.69E+02(3.98E+00)	1.15E+03(9.92E+01)	9.83E+02(8.49E+00)	1.05E+03(1.35E+01)	9.01E+02(9.85E+02)	7.34E+02(5.64E+01)
	200	5.64E+03(6.97E+01)	5.10E+03(1.46E+02)	6.96E+03(3.24E+01)	5.48E+03(3.52E+01)	5.32E+03(1.32E+03)	4.97E+03(1.54E+02)
RHC1	30	5.54E+02(5.68E+00)	4.49E+02(4.70E+01)	6.83E+02(9.49E+00)	5.22E+02(3.48E+00)	5.85E+02(1.21E+01)	2.89E+02(2.48E+01)
	50	6.94E+02(7.25E+00)	5.17E+02(1.73E+01)	8.77E+02(1.10E+00)	6.53E+02(5.36E+00)	5.37E+02(7.65E+01)	3.44E+02(1.58E+01)
	100	7.21E+02(9.42E+00)	5.72E+02(3.14E+01)	9.82E+02(2.08E+01)	1.33E+03(6.92E+02)	9.92E+02(9.70E+01)	5.90E+02(2.13E+01)
	200	1.72E+03(2.61E+01)	1.69E+03(1.42E+01)	1.87E+03(3.76E+01)	1.94E+03(2.96E+01)	1.70E+03(5.58E+01)	1.60E+03(6.09E+01)
RHC2	30	9.57E+02(3.21E+00)	9.32E+02(1.28E+01)	9.01E+02(1.86E+00)	1.07E+03(2.84E+01)	1.02E+03(1.32E+01)	8.98E+02(5.16E+01)
	50	9.88E+02(5.69E+00)	1.02E+03(3.61E+01)	1.01E+03(8.20E+00)	1.02E+03(5.62E+01)	1.05E+03(3.59E+01)	9.51E+02(3.68E+01)
	100	1.38E+03(9.19E+00)	1.41E+03(2.82E+01)	1.54E+03(2.47E+01)	1.67E+03(9.55E+01)	1.40E+03(6.85E+01)	1.09E+03(4.24E+01)
	200	1.72E+03(2.68E+01)	1.69E+03(1.61E+01)	1.94E+03(5.09E+01)	1.98E+03(9.41E+01)	1.61E+03(9.53E+01)	1.31E+03(2.36E+01)

### 3.3 自适应集成策略验证

本节比较 LPMAE 和一些变体在局部搜索阶段的改变,具体如下:(1) LPMAE-single:不适用集成策略,只填充 RBFN 的最佳个体;(2) LPMAE-mean:填充 RBFN 和 Ikriging 的平均值最佳个体.

从表 3 的结果可以看出,自适应集成策略相比于不集成和平均集成方法,在所列测试问题的绝大多数情况下,对于全局最优解的搜索结果均有一定提高,但其提升效果没有 MSC 明显,可能因为此方法为一种

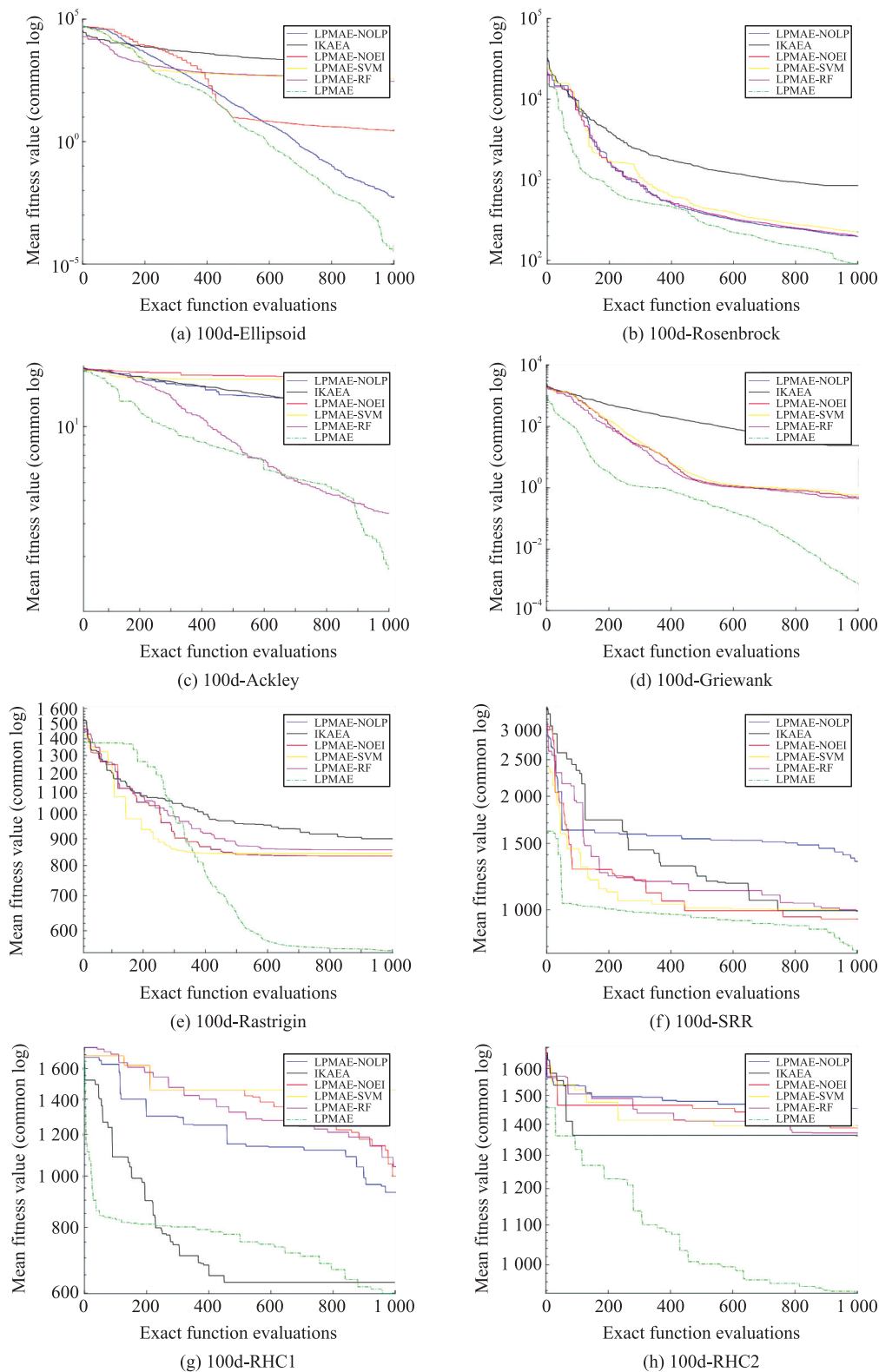


图 2 LPMAE-NOLP,IKAEA,LPMAE-NOEI,LPMAE-SVM,LPMAE-RF,LPMAE  
在 100d 测试问题中的收敛曲线

Fig. 2 Convergence curves of LPMAE-NOLP,IKAEA,LPMAE-NOEI,LPMAE-SVM,LPMAE-RF,  
and LPMAE in 100d test problems

局部搜索策略,对于整体搜索进展的影响不如全局搜索策略显著. 本文使用两种代理模型组成的异构集成模型,基模型的数量对最终的集成结果有很大影响,通常随着基模型数量的增加集成模型的预测能力会更强,但在线代理模型辅助算法涉及到模型更新这一问题时,使用过多的代理模型可能会强化算法的搜索

能力,但也会造成算法运行时间过长. 本文所提的自适应集成策略一定程度上解决了上述困境,利用两个代理模型在测试集上预测结果的特点,根据当前已真实评估解的适应度值来确定当前最优区域,再由不同时刻算法搜索到的最佳区域内代理模型的表现自适应计算两个模型输出的权重,在不增加模型数量的前提下,提高了集成性能.

表 3 不同策略独立运行 20 的均值与方差

Table 3 Mean and variance of running 20 independently with different strategies

Function	$d$	LPMAE-single	LPMAE-mean	LPMAE
Ellipsoid	30	4.28E-06(3.62E-03)	5.76E-05(2.92E-03)	8.69E-9(3.15E-06)
	50	5.14E-06(4.51E-02)	7.10E-04(6.98E-02)	3.65E-10(95.6E-10)
	100	2.71E-03(1.28E-02)	2.48E-02(1.66E-02)	1.61E-05(6.58E-09)
	200	3.51E+02(1.25E+01)	5.10E+03(3.59E+01)	1.78E+01(2.45E+02)
Rosenbrock	30	9.88E+01(8.47E-01)	9.71E+01(1.86E+00)	2.21E+01(1.39E-01)
	50	1.05E+02(4.57E+00)	1.22E+02(9.83E+00)	3.86E+01(6.58E+00)
	100	5.68E+02(1.23E+01)	9.69E+02(2.51E+01)	9.56E+01(1.46E+02)
	200	7.14E+02(9.41E+01)	8.99E+02(1.63E+02)	1.62E+02(3.42E+02)
Ackley	30	9.47E+00(1.58E-01)	8.41E+00(1.02E+00)	2.18E-01(3.14E-01)
	50	6.21E+00(1.59E-01)	4.92E+00(4.07E+00)	0.29E+00(3.58E-01)
	100	1.42E+01(4.11E+00)	1.74E+01(5.27E+00)	1.38E+00(6.49E+00)
	200	1.93E+01(2.45E+00)	1.79E+01(8.92E+00)	1.46E+01(3.42E+00)
Griewank	30	3.57E-04(1.17E-01)	7.48E-04(2.57E-01)	5.68E-05(2.21E-01)
	50	2.18E-02(5.59E-01)	1.36E-01(1.88E-01)	8.98E-03(4.59E-01)
	100	2.59E-01(2.57E+00)	1.13E+00(1.14E+00)	4.89E-02(5.49E+00)
	200	2.81E+00(2.48E+00)	3.08E+00(3.52E+01)	3.09E-01(3.49E+00)
Rastrigin	30	9.85E+01(0.28E+00)	1.52E+02(1.11E+00)	5.16E+01(2.99E+00)
	50	2.42E+02(4.47E+00)	1.96E+02(6.51E+00)	1.28E+02(3.25E+01)
	100	8.41E+02(1.58E+00)	9.54E+02(2.57E+01)	5.39E+02(2.34E+00)
	200	1.52E+03(2.81E+01)	2.10E+03(5.87E+01)	9.86E+02(3.69E+01)
SRR	30	-1.74E+02(3.27E+00)	-1.39E+02(8.24E+00)	-2.33E+02(4.15E+00)
	50	-7.36E+01(8.41E+00)	-8.58E+01(3.28E+01)	-6.59E+01(1.24E+00)
	100	9.95E+02(3.12E+01)	1.03E+03(4.51E+01)	7.34E+02(5.64E+01)
	200	5.01E+03(9.88E+01)	5.43E+03(1.69E+02)	4.97E+03(1.54E+02)
RHC1	30	3.55E+02(2.10E+01)	4.79E+02(8.69E+01)	2.89E+02(2.48E+01)
	50	4.41E+02(1.85E+01)	4.85E+02(1.15E+02)	3.44E+02(1.58E+01)
	100	6.09E+02(2.89E+02)	6.21E+02(6.47E+00)	5.90E+02(2.13E+01)
	200	1.51E+03(2.01E+01)	1.67E+03(1.58E+01)	1.60E+03(6.09E+01)
RHC2	30	8.41E+02(4.05E+01)	1.02E+03(6.03E+01)	8.98E+02(5.16E+01)
	50	9.82E+02(1.26E+01)	9.97E+02(1.32E+01)	9.51E+02(3.68E+01)
	100	1.38E+02(2.23E+01)	1.40E+03(2.04E+01)	1.09E+03(4.24E+01)
	200	1.29E+03(2.11E+01)	1.46E+03(5.41E+01)	1.31E+03(2.36E+01)

### 3.4 参数敏感性分析

为了探究自适应集成策略中不同指标分配权重  $\sigma$  对算法的影响,本文通过设置不同的权重  $\sigma$  来测试集成代理在 3 个评价指标上的综合表现:

$$\omega = \left( \frac{R_1^2}{R_1^2 + R_2^2} * \sigma_1 + \frac{\rho_1^{(\text{corr.})}}{\rho_1^{(\text{corr.})} + \rho_2^{(\text{corr.})}} * \sigma_2 + \frac{RP_1}{RP_1 + RP_2} * \sigma_3 \right). \quad (21)$$

如图 3 所示,在分别仅考虑确定系数(指标 1)、适应度相关指标(指标 2)、排序保存指标(指标 3)集成的情况下,即式中  $\sigma_1 = 1, \sigma_2 = \sigma_3 = 0$  或  $\sigma_2 = 1, \sigma_1 = \sigma_3 = 0$  或  $\sigma_3 = 1, \sigma_1 = \sigma_2 = 0$  时,集成模型在 3 个指标值的结果中均不如  $\sigma_1 = \sigma_2 = \sigma_3 = 1/3$ ,且异常值的数量远远高于其他情况. 当分别侧重于考虑指标 1、指标 2、指标 3 时,即  $\sigma_1 = 1/2, \sigma_2 = \sigma_3 = 1/4$  或  $\sigma_2 = 1/2, \sigma_1 = \sigma_3 = 1/4$  或  $\sigma_3 = 1/2, \sigma_1 = \sigma_2 = 1/4$  时可以看出,集成后结果的波动程度较大. 当  $\sigma_1 = \sigma_2 = \sigma_3 = 1/3$  时,集成结果在 3 个指标中的表现最佳.

### 3.5 对比算法

表 4 列出了 LPMAE 和 4 个先进的代理辅助进化算法在独立运行 20 次后的统计结果.

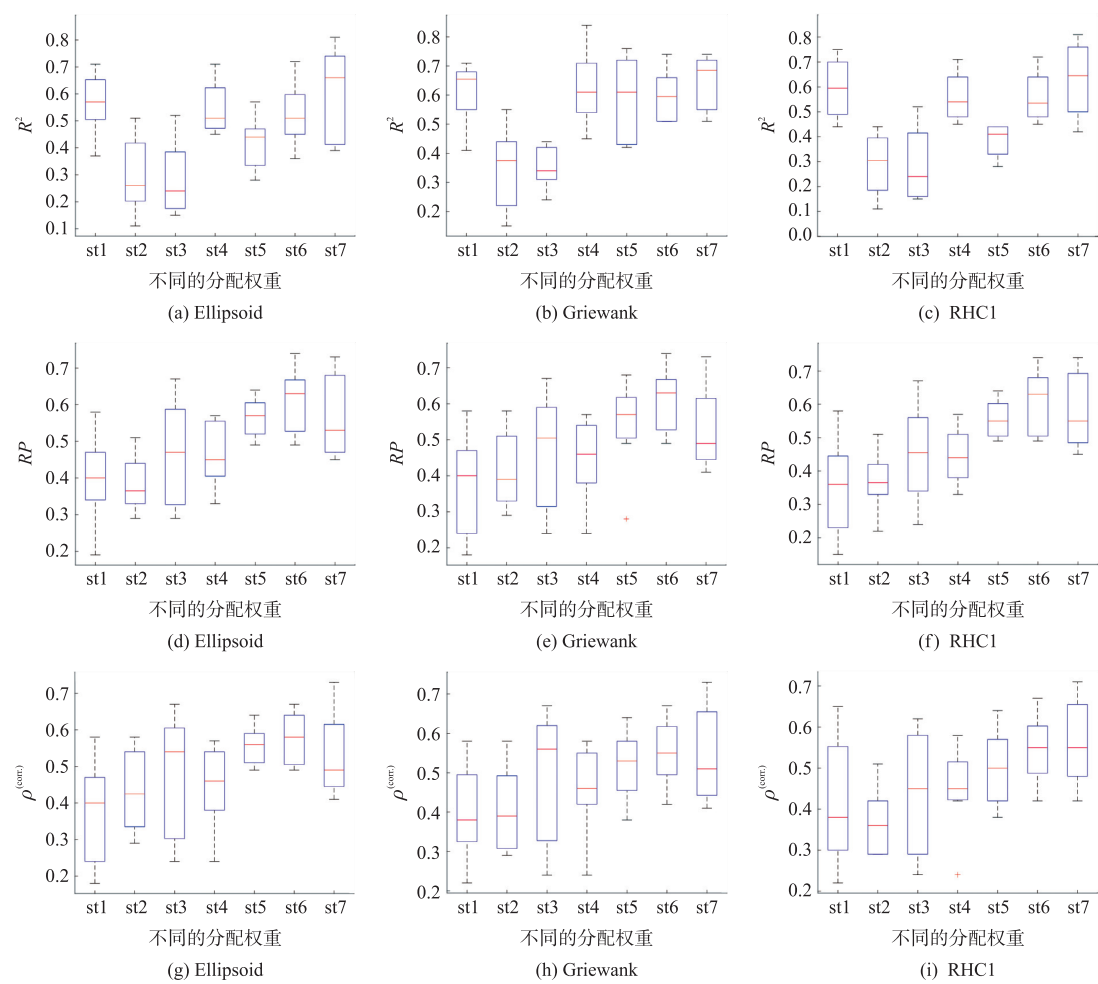


图 3 不同分配权重  $\sigma$  对应的指标大小

Fig. 3 Indicator sizes corresponding to different assigned weights  $\sigma$

表 4 DESO, SHPSO, ESA, TS-DDEO 和 LPMAE 在测试问题中的均值与标准差

Table 4 Mean and standard deviation of DESO, SHPSO, ESA, TS-DDEO and LPMAE in test questions

Function	$d$	DESO	SHPSO	ESA	TS-DDEO	LPMAE
Ellipsoid	30	9.04E-08(9.19E-08)	7.62E-02(4.01E-02)	<b>4.90E-23</b> (5.18E-23)	2.46E-13(3.39E-13)	8.69E-09(3.15E-06)
	50	7.67E-03(5.80E-03)	4.02E+00(2.05E+00)	<b>2.28E-18</b> (4.14E-18)	2.37E-10(1.36E-10)	3.65E-10(95.6E-10)
	100	1.23E+01(4.26E+00)	7.61E+01(2.14E+01)	3.91E-03(7.89E-03)	3.19E-05(1.63E-05)	<b>1.61E-05</b> (6.58E-09)
	200	4.80E+02(1.04E+02)	7.03E+01(1.69E+01)	<b>5.60E-01</b> (9.87E-01)	8.84E-01(2.49E-01)	1.78E+01(2.45E+02)
Rosenbrock	30	2.47E+01(1.27E+00)	2.85E+01(4.04E-01)	2.51E+01(7.36-01)	2.72E+01(6.19E-01)	<b>2.21E+01</b> (1.39E-01)
	50	4.66E+01(9.83E-01)	5.08E+01(3.03E+00)	4.65E+01(7.12E-01)	4.80E+01(8.25E-01)	<b>3.86E+01</b> (6.58E+00)
	100	1.01E+02(2.35E+00)	1.67E+02(2.63E+01)	9.79E+01(4.27E-01)	1.10E+02(2.04E+01)	<b>9.56E+01</b> (1.46E+02)
	200	3.49E+02(3.68E+01)	6.13E+02(1.72E+02)	1.98E+02(2.21E-01)	5.58E+02(3.46E+02)	<b>1.62E+02</b> (3.42E+02)
Ackley	30	<b>3.21E-05</b> (1.82E-05)	1.44E+00(7.74E-01)	5.38E-05(2.26E-05)	1.67E-04(4.84E-05)	2.18E-01(3.14E-01)
	50	2.62E-01(4.79E-01)	1.83E+00(5.63E-01)	<b>9.46E-04</b> (4.69E-04)	6.37E-03(6.71E-03)	0.29E+00(3.58E-01)
	100	2.89E+00(4.71E-01)	4.11E+00(5.92E-01)	<b>2.87E-01</b> (3.42E-01)	1.39E+00(3.43E-01)	1.38E+00(6.49E+00)
	200	5.94E+00(4.00E-01)	<b>1.68E+00</b> (1.19E+00)	8.06E+00(0.35E+00)	5.47E+00(5.68E+00)	1.46E+01(3.42E+00)
Griewank	30	7.96E-04(2.63E-03)	9.20E-01(8.80E-02)	4.73E-04(1.71E-03)	7.69E-03(1.41E-02)	<b>5.68E-05</b> (2.21E-01)
	50	9.59E-03(1.01E-02)	9.45E-01(6.14E-02)	<b>1.75E-03</b> (1.18E-03)	3.40E-02(1.21E-02)	8.98E-03(4.59E-01)
	100	1.02E+00(5.66E-02)	1.07E+00(2.04E-02)	5.37E-02(2.37E-02)	2.69E-01(5.67E-02)	<b>4.89E-02</b> (5.49E+00)
	200	1.61E+01(3.38E+00)	3.10E+00(0.44E-00)	3.89E-01(1.48E+00)	8.69E-01(8.160E-01)	<b>3.09E-01</b> (3.49E+00)
Rastrigin	30	2.14E+02(1.66E+02)	2.40E+02(2.83E+01)	1.36E+02(1.09E+00)	<b>2.62E+01</b> (1.33E+01)	5.16E+01(2.99E+00)
	50	4.03E+02(1.54E+01)	3.65E+02(6.14E+01)	1.73E+02(1.18E+02)	4.01E+02(3.87E+01)	<b>1.28E+02</b> (3.25E+01)
	100	8.24E+02(2.04E+01)	7.69E+02(3.69E+01)	8.17E+02(9.56E+01)	8.13E+02(2.27E+01)	<b>5.39E+02</b> (2.34E+00)
	200	1.76E+03(5.05E+01)	1.84E+03(1.144E+01)	1.59E+03(1.48R+02)	1.88E+03(2.46E+01)	<b>9.86E+02</b> (3.69E+01)



续表 4  
Table 4 continued

Function	$d$	DESO	SHPSO	ESA	TS-DDEO	LPMAE
SRR	30	-7.25E+01(2.35E+01)	-9.28E+01(2.25E+01)	-1.19E+02(2.46E+01)	-8.44E+01(1.76E+01)	<b>-2.33E+02</b> (4.15E+00)
	50	2.30E+02(4.64E+01)	1.34E+02(3.22E+01)	5.55E+01(6.16E+01)	1.27E+02(1.90E+01)	<b>-6.59E+01</b> (1.24E+00)
	100	1.20E+03(1.18E+02)	8.01E+02(7.22E+01)	1.21E+03(1.39E+02)	7.59E+02(6.22E+01)	<b>7.34E+02</b> (5.64E+01)
	200	<b>4.79E+03</b> (1.52E+02)	5.65E+03(1.99E+02)	4.85E+03(2.38E+02)	5.19E+03(2.24E+02)	4.97E+03(1.54E+02)
RHC1	30	5.16E+02(8.58E+01)	4.64E+02(8.51E+01)	4.16E+02(8.17E+01)	4.48E+02(6.14E+01)	<b>2.89E+02</b> (2.48E+01)
	50	5.65E+02(6.99E+00)	4.74E+02(4.20E+01)	4.37E+02(4.33E+01)	4.67E+02(2.45E+01)	<b>3.44E+02</b> (1.58E+01)
	100	5.98E+02(3.71E+01)	5.16E+02(3.20E+01)	6.05E+02(3.39E+01)	<b>5.04E+02</b> (2.09E+01)	5.90E+02(2.13E+01)
	200	1.68E+03(1.38E+01)	1.71E+03(7.26E+02)	1.68E+03(4.77E+01)	1.65E+03(7.75E+01)	<b>1.60E+03</b> (6.09E+01)
RHC2	30	9.66E+02(1.89E+01)	9.39E+02(9.01E+00)	9.37E+02(9.17E+00)	9.40E+02(1.03E+01)	<b>8.98E+02</b> (5.16E+01)
	50	1.04E+03(1.75E+01)	9.96E+02(2.21E+01)	1.01E+03(3.83E+01)	1.00E+03(2.40E+01)	<b>9.51E+02</b> (3.68E+01)
	100	1.35E+03(4.99E+01)	1.41E+03(3.82E+01)	1.18E+03(1.87E+02)	1.39E+03(3.22E+01)	<b>9.89E+02</b> (4.24E+01)
	200	1.35E+03(7.49E+01)	1.51E+03(6.02E+01)	1.69E+03(1.17E+02)	1.68E+03(3.27E+01)	<b>1.31E+03</b> (2.36E+01)

从表 4 可以看出,LPMAE 在所有测试函数的大多数维度中均有良好的表现,特别是在决策变量维度较高的情况下.得益于多重筛选多点填充准则,当目标空间崎岖度增高且搜索空间增大,在开发性采样时有效过滤掉了部分当前所在位置适应度不佳的粒子,在不确定性最小的区域内利用 RBFN 的预测能力再次筛选,由于 RBFN 是基于所有已评估数据动态更新的,随着训练数据量的增大 RBFN 的精度逐渐提高,所以在第三次筛选时选择比上一代预测全局最优值适应度值好的粒子进行评估,同时不放弃在第一步被过滤的粒子利用 Ikriging 和 EI 准则选择一个个体进行评估,这样的填充准则兼顾了多样性和收敛性.同时以图 4 为例,在迭代过程中,ESA 中有 4 种不同的采样策略,通过 Q 学习在不同问题和不同阶段智能选择采样策略,特别在维度为 30、50 和 200 的 Ellipsoid 问题中 ESA 找到的最优解比其他算法降低了多个数量级,可能因为 Ellipsoid 为单峰问题,ESA 的采样策略在没有局部最优解的影响下收敛能力显著增强.相反,除 Ackley 问题,LPMAE 在绝大部分多峰问题中均有良好的表现,原因可能在于 Ackley 问题中存在多个较浅的峡谷,处于局部最优位置的正类样本会将谷峰的粒子标签传播为正,负类样本会将一些有前途的个体传播为负类标签,同时搜索范围相对较大,从而导致搜索效率不高. TS-DDEO 在第一个阶段采用 SHPSO 进行双种群搜索,在计算资源耗尽后进入第二阶段,切换为 DE 优化器,并提出了 BDDO 算子,其中包括 3 种采样策略,可以看出其表现在不同问题的不同维度中虽不是最好的,但其平均排名位列第二,一次迭代过程中选择多个新样本,达到设定精度时需要的迭代次数大大减少,采样效率大幅提升.

### 3.6 翼型设计实例

本文尝试将 LPMAE 应用于 NACA2411 翼型设计问题.选择低速高升<sup>[31]</sup>的经典翼型 NACA2411 翼型作为基线,优化问题包含 11 个变量,其中两个变量固定.初始翼型和搜索半径的设计值如表 5 所示.设初始值为  $X_0$ ,搜索半径为  $r$ ,搜索范围为  $[X_0-r, X_0+r]$ .优化问题的目标是通过搜索设计值参数获得最大升力系数,目标函数采用给定输入参数的旋涡面板法求解翼型升力系数.为了将该算法与 DESO、SHPSO、ESA、TS-DDEO 进行比较,在实验中设置相同参数.初始样本点数为 50,种群大小设置为 30,真实评估次数为 300.表 6 显示了 4 种对比算法独立运行 20 次的结果.

表 5 优化参数  
Table 5 Optimized parsec parameters

设计参数	初值	搜索范围	设计参数	初值	搜索范围
前缘半径( $R_{lc}$ )	0.021 6	0.001 5	下波峰曲率( $YXX_{lo}$ )	0.674 8	0.075
上波峰位置( $X_{up}$ )	0.344 5	0.025	后缘厚度( $T_{TE}$ )	0	0
上波峰点( $Y_{up}$ )	0.079 12	0.015	后缘偏移( $T_{off}$ )	0	0
上波峰曲率( $YXX_{up}$ )	-0.644 8	0.01	尾缘方向角( $\alpha_{TE}$ )	0	0.175
下波峰位置( $X_{lo}$ )	0.17	0.02	后缘楔角( $\beta_{TE}$ )	0	0.05
下波峰点( $Y_{lo}$ )	-0.033 797	0.015			

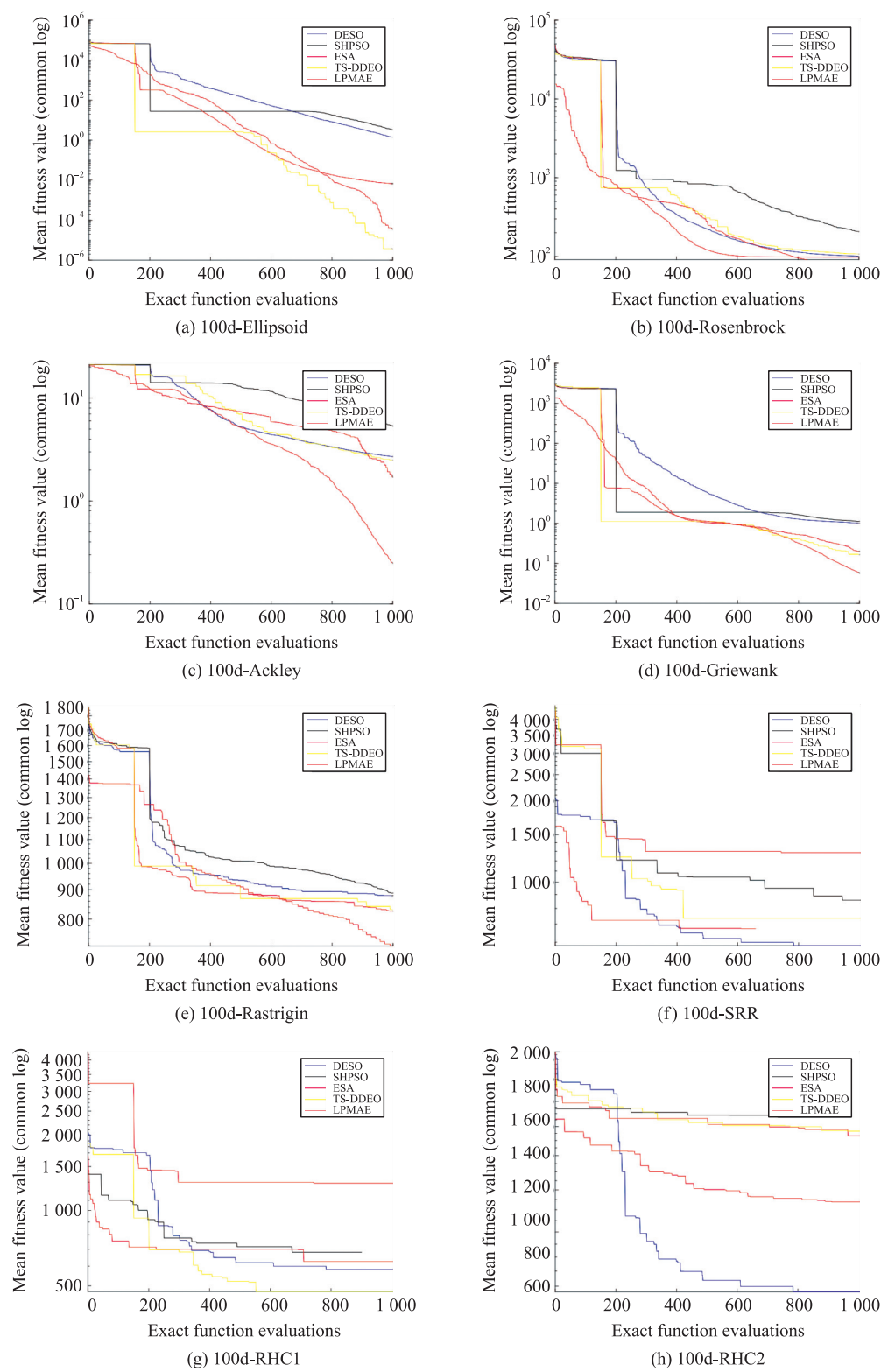


图 4 DESO,SHPSO,ESA,TS-DDEO 和 LPMAC 在 100d 测试问题中的收敛曲线

Fig. 4 Convergence curves of DESO,SHPSO,ESA,TS-DDEO and LPMAC in the 100d test problem

表 6 翼型优化结果分析

Table 6 Analysis of airfoil optimization results

算法	均值	标准差	算法	均值	标准差
DESO	9.58E-01	4.9E-05	TS-DDEO	9.24E-01	3.26E-04
SHPSO	9.58E-01	3.3E-03	LPMAC	9.18E-01	4.69E-05
ESA	9.34E-01	4.81E-05			

优化后的翼型形状如图 5 所示,  $X/C$  和  $Y/C$  分别表示翼型按弦长等比例缩放到 1 后的横轴和纵轴. 此优化问题由 11 个决策变量组成, LPMAE 得到了更好的均值和标准差, 证明 LPMAE 对于更低维优化问题也同样具有良好的表现.

## 4 结论

为解决昂贵优化问题, 本文提出了一种基于标签传播引导的多重筛选多点填充准则 (MSC) 和自适应集成策略的算法. 该方法结合全局和局部搜索两个阶段, MSC 帮助算法在全局筛选个体真实评估; 自适应集成策略在局部最优区域根据迭代过程中最优解的反馈灵活切换, 平衡了多样性和收敛性. 在 8 个单峰和多峰测试问题的 30、50、100、200 维度中分别进行了消融实验、参数敏感性实验, 并与其他先进的算法在测试问题和翼型设计问题中进行了对比, 展现了本文算法一定的优势. 未来可将所提策略应用于多目标昂贵优化问题以研究更有效的多目标昂贵优化算法, 如随着目标数量的增加如何设计更有效的自适应集成方法和预筛选策略等.

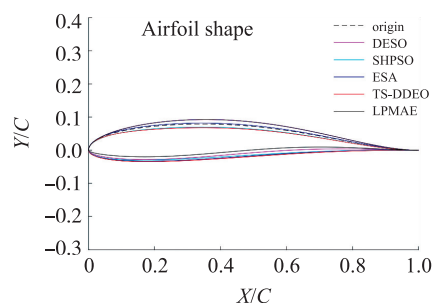


图 5 翼型优化结果图

Fig. 5 Results of airfoil shape optimization

## [参考文献] (References)

- [1] LIAO X T, LI Q, YANG X J, et al. Multiobjective optimization for crash safety design of vehicles using stepwise regression model[J]. Structural and Multidisciplinary Optimization, 2008, 35: 561–569.
- [2] YANG S S, TIAN Y, XIANG X S, et al. Accelerating evolutionary neural architecture search via multifidelity evaluation[J]. IEEE Transactions on Cognitive and Developmental Systems, 2022, 14(4): 1778–1792.
- [3] LI J Y, ZHAN Z H, ZHANG J. Evolutionary computation for expensive optimization[J]. A Survey Machine Intelligence Research, 2022, 19: 3–23.
- [4] JIN Y C, WANG H D, CHUGH T, et al. Data-driven evolutionary optimization: An overview and case studies[J]. IEEE Transactions on Evolutionary Computation, 2019, 23(3): 442–458.
- [5] WANG H D, DOHERTY J, JIN Y C, et al. Hierarchical surrogate-assisted evolutionary multi-scenario airfoil shape optimization [C]//Proceedings of the 2018 IEEE Congress on Evolutionary Computation (CEC). Rio de Janeiro, Brazil: IEEE, 2018.
- [6] WANG H D, JIN Y C. A random forest-assisted evolutionary algorithm for data-driven constrained multiobjective combinatorial optimization of trauma systems[J]. IEEE Transactions on Cybernetics, 2020, 50(2): 536–549.
- [7] CHEN G D, ZHANG K, XUE X M, et al. A radial basis function surrogate model assisted evolutionary algorithm for high-dimensional expensive optimization problems[J]. Applied Soft Computing, 2022, 116: 108353.
- [8] MENG D B, YANG S Y, DE JESUS A M P, et al. A novel Kriging-model-assisted reliability-based multidisciplinary design optimization strategy and its application in the offshore wind turbine tower[J]. Renewable Energy, 2023, 203: 407–420.
- [9] PAN L Q, HE C, TIAN Y, et al. A classification-based surrogate-assisted evolutionary algorithm for expensive many-objective optimization[J]. IEEE Transactions on Evolutionary Computation, 2019, 23(1): 74–88.
- [10] LOSHCILLOV I, SCHOENAUER M, SEBAG M. A mono surrogate for multiobjective optimization[C]//Proceedings of the 12th Annual Conference on Genetic and Evolutionary Computation. Portland, USA: ACM, 2010: 471–478.
- [11] KULKARNI V Y, SINHA P K. Pruning of random forest classifiers: A survey and future directions[C]//Proceedings of the 2012 International Conference on Data Science & Engineering (ICDSE). Cochin, India: IEEE, 2012.
- [12] ZHANG J Y, ZHOU A M, ZHANG G X, et al. A classification and Pareto domination based multiobjective evolutionary algorithm[C]//Proceedings of the 2015 IEEE Congress on Evolutionary Computation (CEC). Sendai, Japan: IEEE, 2015.
- [13] WEI F F, CHEN W N, YANG Q, et al. A classifier-assisted level-based learning swarm optimizer for expensive optimization[J]. IEEE Transactions on Evolutionary Computation, 2021, 25(2): 219–233.
- [14] TANG Z L, XU L, LUO S J. Adaptive dynamic surrogate-assisted evolutionary computation for high-fidelity optimization in engineering[J]. Applied Soft Computing, 2022, 127: 109333.
- [15] WANG W Z, LIU H L, TAN K C. A surrogate-assisted differential evolution algorithm for high-dimensional expensive optimization problems[J]. IEEE Transactions on Cybernetics, 2023, 53(4): 2685–2697.

- [16] WANG H D, JIN Y C, DOHERTY J. Committee-based active learning for surrogate-assisted particle swarm optimization of expensive problems[J]. IEEE Transactions on Cybernetics, 2017, 47(9): 2664–2677.
- [17] LI J Y, ZHAN Z H, WANG H, et al. Data-driven evolutionary algorithm with perturbation-based ensemble surrogates[J]. IEEE Transactions on Cybernetics, 2021, 51(8): 3925–3937.
- [18] SUN C L, JIN Y C, CHENG R, et al. Surrogate-assisted cooperative swarm optimization of high-dimensional expensive problems[J]. IEEE Transactions on Evolutionary Computation, 2017, 21(4): 644–660.
- [19] ZHAN D W, XING H L. A fast kriging-assisted evolutionary algorithm based on incremental learning[J]. IEEE Transactions on Evolutionary Computation, 2021, 25(5): 941–955.
- [20] ZHEN H X, GONG W Y, WANG L, et al. Two-stage data-driven evolutionary optimization for high-dimensional expensive problems[J]. IEEE Transactions on Cybernetics, 2023, 53(4): 2368–2379.
- [21] CHAWLA N V, BOWYER K W, HALL L O, et al. SMOTE: synthetic minority over-sampling technique[J]. Journal of Artificial Intelligence Research, 2002, 16(1): 321–357.
- [22] LI K, CHEN R Z, YAO X. A data-driven evolutionary transfer optimization for expensive problems in dynamic environments[J]. IEEE Transactions on Evolutionary Computation, 2024, 28(5): 1396–1411.
- [23] ZHU X J, GHANRAMANI Z. Learning from labeled and unlabeled data with label propagation[R]. Pittsburghers, USA: Carnegie Mellon University, 2002.
- [24] 张俊丽, 常艳丽, 师文. 标签传播算法理论及其应用研究综述[J]. 计算机应用研究, 2013, 30(1): 21–25.
- [25] 李正良, 彭思思, 王涛. 基于混合加点准则的代理模型优化设计方法[J]. 工程力学, 2022, 39(1): 27–33.
- [26] CHENG R, JIN Y. A social learning particle swarm optimization algorithm for scalable optimization[J]. Information Sciences, 2015, 291: 43–60.
- [27] ZHEN H X, GONG W Y, WANG L. Data-driven evolutionary sampling optimization for expensive problems[J]. Journal of Systems Engineering and Electronics, 2021, 32(2): 318–330.
- [28] LV Z M, WANG L Q, HAN Z Y, et al. Surrogate-assisted particle swarm optimization algorithm with Pareto active learning for expensive multi-objective optimization[J]. IEEE/CAA Journal of Automatica Sinica, 2019, 6(3): 838–849.
- [29] 李贞. 昂贵高维多目标进化优化中代理模型的应用研究[D]. 太原: 太原科技大学, 2021.
- [30] LIU H W, ZHOU C C, LIU F C, et al. A trust-region-like algorithm for expensive multi-objective optimization[J]. Applied Soft Computing, 2023, 148(C): 110892.
- [31] MUKESH R, LINGADURAI K, SELVAKUMAR U. Airfoil shape optimization using non-traditional optimization technique and its validation[J]. Journal of King Saud University-Engineering Sciences, 2014, 26(2): 191–197.

[责任编辑: 严海琳]