

混合性能指标优化问题的大种群规模进化算法

巩敦卫, 秦娜娜, 孙晓燕

(中国矿业大学 信息与电气工程学院, 江苏 徐州 221116)

摘要: 混合性能指标优化问题可结合传统遗传算法和交互式遗传算法求解, 而种群规模和人机评价任务分配是影响算法性能的关键。针对该问题, 本文提出一种新的进化优化算法。首先, 采用大规模种群, 扩大搜索范围, 以增强算法的探索能力; 然后, 根据计算机和用户完成任务耗时的比值, 确定每代用户评价的个体数, 以提高计算机的使用效率; 接着, 采用K-均值聚类方法和基于相似度的估计策略, 以减轻用户疲劳; 最后, 采用Pareto占优比较不同个体的优劣, 使得最优解有较好的显式性能指标值和隐式性能指标值。将本文算法应用于室内布局这一混合性能指标优化问题, 结果验证了所提算法的有效性。

关键词: 进化优化; 混合性能指标; 遗传算法; 交互
中图分类号: TP273 **文献标识码:** A

Evolutionary algorithm with large population size for problems with hybrid indices

GONG Dun-wei, QIN Na-na, SUN Xiao-yan

(School of Information and Electrical Engineering, China University of Mining & Technology, Xuzhou Jiangsu 221116, China)

Abstract: Problems with hybrid indices can be solved by combining traditional genetic algorithms and interactive genetic algorithms. The major causes affecting the performance of these algorithms are the population size and the strategy of task allocation between the user and the computer. A novel evolutionary optimization algorithm to solve the above problems is proposed. A large population is adopted to expand the searching space and enhance the performance of the algorithm in exploration. The number of individuals evaluated by the user in each generation is determined according to the ratio of the time consumptions of the computer and the user to finish their tasks, in order to improve the efficiency of the computer in usage. The K-mean clustering method and a similarity-based estimation strategy are adopted to alleviate the user's fatigue. Finally, Pareto domination is employed to compare different individuals so that optimal solutions have good values in explicit and implicit indices. The proposed algorithm is applied to an interior layout problem with hybrid indices, and the results validate its efficiency.

Key words: evolutionary optimization; hybrid indices; genetic algorithms; interaction

1 引言(Introduction)

现实世界中的很多优化问题同时包含显式(定量)性能指标和隐式(定性)性能指标, 其中, 显式性能指标可以用明确定义的函数表示, 而隐式性能指标却无法用明确定义的函数表示, 一般由语言叙述、个人观点或者想法描述, 这类问题称为混合性能指标优化问题^[1]。

近年来, 国内外学者开始采用进化优化算法解决该类问题, 并成为进化优化界的热点研究方向之一。周勇等在解决混合性能指标优化问题时, 通过加权归一化的各性能指标值得到进化个体的适应值, 并应用于服装设计问题^[2]。BRINTRUP A.等比较

了两种混合性能指标优化问题的进化优化方法, 即串行交互式遗传算法(interactive genetic algorithms, IGAs)和多目标IGAs, 并通过优化工厂布局问题, 验证了后者优于前者^[3]。上述几种方法, 由于受交互界面的影响, 其进化种群规模较小, 限制了算法的搜索能力, 且频繁的人机交互容易导致用户疲劳。KAMALIAN R.等首先针对多显式性能指标, 采用传统的多目标进化优化方法进化一定代数后, 再由用户评价所得最优解对隐式性能指标的满足程度, 从而打破了由于用户的交互而对进化种群规模和进化代数的限制, 减轻了用户疲劳^[4]。但该算法对两类性能指标进行分阶段优化, 难以保证两类性能指标同

时较优. 另外, 上述算法结合传统的遗传算法和交互式遗传算法解决该类问题时, 没有合理分配人机评价任务, 降低了计算机的使用效率.

针对上述不足, 本文从增强算法搜索性能、减轻用户疲劳, 以及合理分配人机评价任务的角度出发, 提出一种新的进化优化算法, 主要包括以下4方面: 1) 采用大规模进化种群, 扩大搜索范围, 从而增强算法的探索能力; 2) 从计算机和用户评价耗时角度出发, 合理分配人机评价任务, 以提高计算机的使用效率; 3) 采用K-均值聚类方法和基于相似度的估计策略, 以减轻用户疲劳; 4) 采用Pareto占优比较不同进化个体的优劣^[5], 使得最优解有较好的显式性能指标值和隐式性能指标值.

2 混合性能指标优化问题的数学模型 (Mathematical model of problems with hybrid indices)

考虑式(1)所示的一类混合性能指标优化问题:

$$\begin{cases} \min : f_1(x), f_2(x), \dots, f_p(x), \\ \max : f_{p+1}(x), f_{p+2}(x), \dots, f_{p+q}(x), \\ \text{s.t. } x \in S \subseteq \mathbb{R}^m. \end{cases} \quad (1)$$

式中: x 为 m 维决策变量; S 为 x 的可行域; $f_i(x)$, $i = 1, 2, \dots, p$ 为显式性能指标, $f_j(x)$, $j = p+1, p+2, \dots, p+q$ 为隐式性能指标. 若 $p = 0$, 则式(1)为隐式性能指标优化问题; 若 $q = 0$, 则式(1)为显式性能指标优化问题; 若 $p > 0, q > 0$, 则式(1)为混合性能指标优化问题. 下面, 本文将给出解决该类问题的进化优化算法.

3 混合性能指标优化问题的大种群规模进化算法(Evolutionary algorithm with large population size for problems with hybrid indices)

为了增强交互式遗传算法的搜索性能, 减轻用户疲劳, 并保证两类性能指标同时较优, 本文从合理分配人机评价任务的角度出发, 结合传统的遗传算法和交互式遗传算法的求解策略, 提出一种新的进化优化算法. 算法思想是: 首先, 根据计算机和用户完成各自任务耗时的比值, 动态调整每代用户评价的进化个体数; 其次, 根据每代用户评价的进化个体数, 采用K-均值聚类方法对大规模进化种群分类^[6]; 再次, 采用基于相似度的估计策略, 估计每类中用户未评价个体的隐式性能指标值; 最后, 采用Pareto占优比较不同进化个体的优劣, 使得最优解有较好的显式性能指标值和隐式性能指标值.

需要解决以下2个关键问题: 1) 人机评价任务分

配的依据及方法; 2) 基于相似度的估计策略.

3.1 人机评价任务分配的依据及方法(Basis and method of assigning tasks for human and computer)

本文结合传统的遗传算法和交互式遗传算法求解混合性能指标优化问题. 其中, 计算机的主要任务是计算每代种群中所有进化个体的显式性能指标值; 用户的主要任务是评价每代由计算机提交的每类的类中心的隐式性能指标. 为了提高计算机的使用效率, 应保证每代二者完成各自任务的耗时大致相当, 因此, 本文中用户每代评价的进化个体数与上一代计算机耗时、用户耗时相关. 本文固定计算机评价的进化个体数, 根据人机评价任务耗时的比值, 确定用户每代评价的进化个体数, 具体方法如下.

记解决混合性能指标优化问题的进化种群规模为 N . 在第 t 代用户评价的进化个体数为 $N_c(t)$, 耗时为 $T'(t)$, 而计算机计算 N 个进化个体的显式性能指标值的耗时为 $T(t)$. 为保证每代二者的评价耗时大致相当, 则第 t 代用户评价的进化个体数可以表示为

$$N_c(t) = \min\{\lceil \frac{T(t-1)}{T'(t-1)} \cdot N_c(t-1) \rceil, N_{\max}\}, \quad (2)$$

式中: $\lceil \cdot \rceil$ 表示向上取整, $N_{\max}$ 为交互界面最多能显示的个体数. 取 $N_c(0) = N_{\max}$.

根据上述方法确定了每代用户评价的进化个体数, 接下来选择用户评价的进化个体. 具体方法如下: 根据式(2)获得第 t 代用户评价的进化个体数 $N_c(t)$, 将其作为大规模进化种群的分类数. 接着, 采用K-均值聚类方法对种群进行分类. 然后, 把每类的类中心提交给用户, 评价其隐式性能指标.

3.2 基于相似度的估计策略(Similarity-based estimation strategy)

由3.1节人机评价任务的分配可知, 在当前大规模进化种群中, 有大量的进化个体没有隐式性能指标值. 为了实现进化优化, 有必要根据已有信息, 估计这些个体的隐式性能指标值. 本文根据待估计个体与其所在类的类中心之间的相似度, 估计每类中其余个体的隐式性能指标值. 采用如下的进化个体隐式性能指标值估计方法.

考虑进化个体有 M 个基因意义单元的情况^[7]. 记第 t 代进化种群为 $x(t)$, 种群规模为 N . 其中: 进化个体 $x_i(t)$ ($i = 1, 2, \dots, N$)的基因型为 $x_{i1}x_{i2}, \dots, x_{iM}$, 则 $x(t)$ 中两个进化个体 $x_j(t)$ 和 $x_i(t)$ 的相似度可表示为^[8]

$$\alpha(x_j(t), x_i(t)) = \frac{1}{M} \sum_{m=1}^M \alpha_m(x_j(t), x_i(t)), \quad (3)$$

式中:

$$\alpha_m(x_j(t), x_i(t)) = \begin{cases} 1, & x_{jm} = x_{im}, \\ 0, & x_{jm} \neq x_{im}. \end{cases}$$

记第 t 代种群的第 i 类为 $\{c_i(t)\}$, 类中心为 $c_i(t)$ ($i = 1, 2, \dots, N_c(t)$), 其隐式性能指标值为 $(f_{p+1}(c_i(t)), f_{p+2}(c_i(t)), \dots, f_{p+q}(c_i(t)))$, 记第 $\{c_i(t)\}$ 类中第 j 个进化个体 $x_j(t)$ ($x_j(t) \neq c_i(t)$)的隐式性能指标的估计值为 $(\hat{f}_{p+1}(x_j(t)), \hat{f}_{p+2}(x_j(t)), \dots, \hat{f}_{p+q}(x_j(t)))$, 与其类中心 $c_i(t)$ 的相似度 $\alpha(x_j(t), c_i(t))$, 则 $x_j(t)$ 第 k 个隐式性能指标的估计值可表示为

$$\hat{f}_{p+k}(x_j(t)) = f_{p+k}(c_i(t))e^{-(1-\alpha(x_j(t), c_i(t)))}. \quad (4)$$

这样一来, 就可以得到任一进化个体的适应值. 基于这些适应值, 采用Pareto占优比较不同进化个体的优劣和精英保留策略, 通过传统的遗传操作, 如交叉和变异操作, 生成子代进化种群.

3.3 算法步骤(Steps of the algorithm)

本文算法的步骤如下:

Step 1 初始化, 设置进化参数, 生成初始种群;

Step 2 计算机根据用户评价个体数对种群分类;

Step 3 人机并行执行任务: 用户评价每类类中心的隐式性能指标值, 与此同时, 计算机计算种群中所有进化个体的显式性能指标值;

Step 4 采用3.2节基于相似度的估计策略, 估计种群中其余个体的隐式性能指标值;

Step 5 根据进化个体的适应值, 采用Pareto占优比较不同进化个体的优劣^[5], 并实施进化操作;

Step 6 判断是否满足停机准则, 若是, 转Step 7; 否则, 由式(2)更新下一代用户评价的进化个体数, 转Step 2;

Step 7 停止进化, 输出优化结果.

3.4 算法性能分析(Performance Analysis)

本部分主要比较本文算法与传统的多目标进化优化算法处理混合性能指标优化问题的性能, 主要包括两方面: 1) 搜索到的个体数; 2) 用户评价个体数.

1) 搜索到的个体数.

受交互界面大小的限制, 记传统的多目标进化优化算法处理混合性能指标优化问题的种群规模为 $N_{\max}$ ($N_{\max} \ll N$), 最大进化代数 T . 进化 T_0 ($T_0 \leq T$)代后, 本文算法搜索到的个体数为 NT_0 , 而传统的多目标进化优化算法搜索到的个体数为 $N_{\max}T_0$. 很明显, $NT_0 \gg N_{\max}T_0$, 所以, 本文算法有更多的机会找到满意个体.

2) 用户评价个体数.

进化 T_0 代后, 本文算法用户评价的个体数为 $\sum_{t=1}^{T_0} N_c(t)$; 而传统的多目标进化优化算法用户评价的个体数为 $N_{\max}T_0$. 由式(2)可知 $N_c(t) \leq N_{\max}$, 所以 $\sum_{t=1}^{T_0} N_c(t) \leq N_{\max}T_0$, 即采用本文算法用户评价个体数不会增加, 所以, 本文算法不会增加用户疲劳.

4 在室内布局优化系统中的应用(Application in interior layout problem)

4.1 室内布局问题描述(Problem description)

考虑一套普通的居室布局如图1所示, 已知居室的开间 W (总宽)和进深 L (总长). 目标是合理分配每一部分的开间和进深, 使得该居室的总造价 $f_1(x)$ (显式性能指标)最少, 并且满足用户的审美需求 $f_2(x)$ (隐式性能指标).

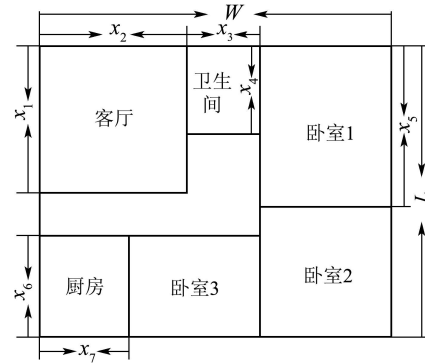


图1 居室的布局

Fig. 1 Interior layout

现在建立该问题的数学模型. 记客厅、卫生间、卧室1、厨房、卧室3、卧室2, 以及过道的单位面积造价分别为 $c_1, c_2, \dots, c_7$, 记 $x = (x_1, x_2, \dots, x_7)^T$ 为该问题的决策变量, 则室内布局问题可表示为如下优化问题:

$$\begin{cases} \min : f_1(x) = c_1x_1x_2 + c_2x_3x_4 + \\ \quad c_3(W - x_2 - x_3)x_5 + c_4x_6x_7 + \\ \quad c_5x_6(x_2 + x_3 - x_7) + c_6(W - \\ \quad x_2 - x_3)(L - x_5) + c_7((x_2 + x_3)(L - \\ \quad x_1 - x_6) + (x_1 - x_4)x_3)), \\ \max : f_2(x), \\ \text{s.t. } x_1 \in \{4.0, 4.3, 4.6, 4.9, 5.2\}, \\ \quad x_2 \in \{4.0, 4.3, 4.6, 4.9, 5.2, \\ \quad \quad 5.5, 5.8, 6.1, 6.4, 6.7, 7.0, 7.3\}, \\ \quad x_3 \in \{2.0, 2.3, 2.6, 2.9, 3.2\}, \\ \quad x_4 \in \{2.0, 2.4, 2.8, 3.2, 3.6\}, \\ \quad x_5 \in \{1.0, 2.0, 3.0, 4.0, 5.0\}, \\ \quad x_6 \in \{2.6, 2.9, 3.2, 3.5, 3.8\}, \\ \quad x_7 \in \{1.0, 2.0, 3.0, 4.0, 5.0\}. \end{cases} \quad (5)$$

4.2 参数设置及编码(Parameters setting and individual's coding)

取居室的开间 $W = 12\text{ m}$, 进深 $L = 10\text{ m}$, 每部分单位面积造价分别为 $c_1 = 800$, $c_2 = 1000$, $c_3 = c_5 = c_6 = 600$, $c_4 = 1000$, $c_7 = 500$, 种群规模为 $N = 500$, 人机交互界面最多能显示的个体数为 $N_{\max} = 12$. 采用联赛选择、单点交叉和单点变异, 交叉和变异概率分别为 $P_c = 0.95$, $P_m = 0.01$, 最大进化代数 $T = 10$. 本文采用实数编码方式, 进化个体的基

因型为 $x_1x_2x_3x_4x_5x_6x_7$. 由于各变量 $x_1, x_2, x_3, x_4, x_5, x_6, x_7$ 取值个数分别为5, 12, 5, 5, 5, 5, 5, 因此, 总的布局方案数为 $12 \times 5^6 = 187500$.

4.3 交互界面及进化过程(Interface and evolution process)

采用Visual Basic 6.0在Windows XP平台上开发的室内布局优化系统的交互界面如图2所示. 用户评价采用1~100评分制, 分值越高, 表示用户越满意.

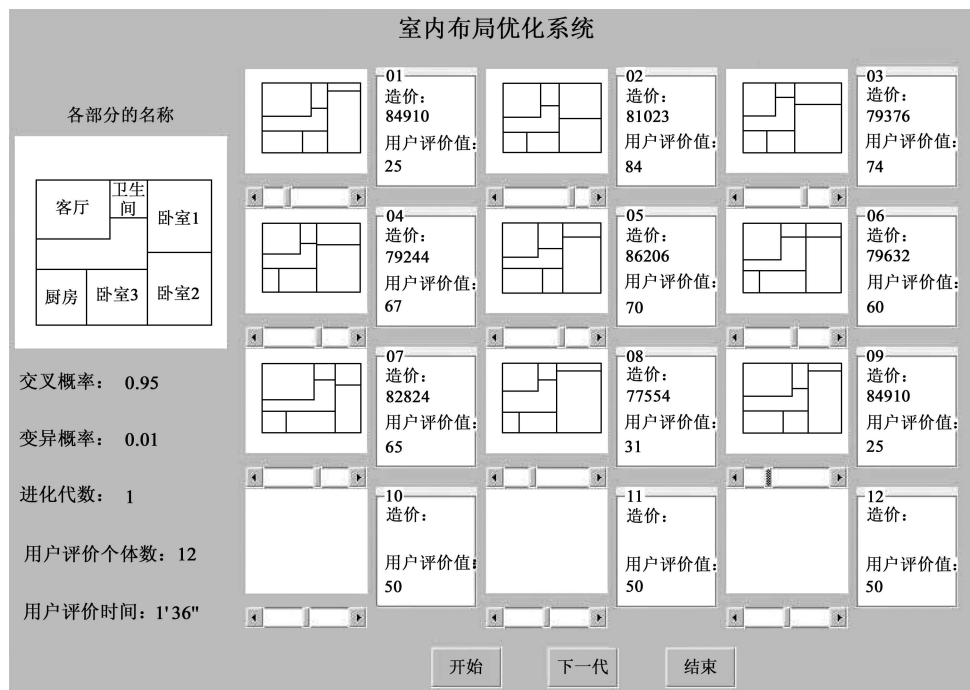


图2 人机交互界面

Fig. 2 Human-computer interface

用户点击“开始”按钮时, 室内布局优化系统的交互界面呈现给用户 $N_c(0) = N_{\max}$ 个个体. 用户评价完所有个体后, 点击“下一代”按钮. 在这个过程中, 计算机记录 $T(0)$, $T'(0)$, 并由式(2)计算第1代用户评价的个体数 $N_c(1)$. 在第1代中, 计算机把种群分为 $N_c(1)$ 类; 分类完毕, 提交每类的类中心给用户评价, 与此同时, 计算机计算所有个体的显式性能指标值; 接着, 估计非类中心个体的隐式性能指标值, 然后, 基于所有进化个体的适应值, 采用Pareto占优比较进化个体的优劣, 并实施进化操作, 生成下一代进化种群. 这时, 由式(2)更新下一代用户评价的个体数, 重复第1代的操作, 直到满足停机准则, 停止进化, 得到优化结果.

4.4 实验结果及分析(Experimental result and analysis)

通过对比实验验证本文算法的优越性, 比较对象为文献[3]所提的能有效解决混合性能指标优化问题的多目标IGAs, 其种群规模为12, 其他参数设置如4.2节所述. 每种算法分别独立运行30次, 实验终止条件为达到最大进化代数.

1) 两种算法的计算机使用效率和用户疲劳.

分别对两种算法的30次运行结果取均值, 得到表1所示的数据. 由表1可知: 1) 采用本文所提算法, 计算机与用户完成各自任务的耗时大致相当, 这说明, 本文算法提高了计算机的使用效率; 2) 本文算法的用户评价耗时少于多目标IGAs, 这是由于本文算法用户评价的进化个体较少; 3) 本文算

法用户评价的个体数少于多目标IGAs, 且搜索的个体数远多于多目标IGAs, 这说明, 本文算法不但扩大了搜索范围, 增强了算法的探索能力, 而且减轻了用户疲劳.

表 1 计算机及用户的耗时、用户评价个体数, 以及搜索的个体数

Table 1 Time-consuming of computer and user, number of individuals being evaluated by user and number of individuals being searched

算法	计算机 耗时	用户 耗时	用户评价 个体数	搜索的 个体数
本文算法	8' 20"	8' 43"	107.7	5000
多目标IGAs	2'	9' 32"	120	120

2) 两种算法得到的Pareto最优解.

分别对两种算法30次运行得到的Pareto最优解取均值, 结果如表2所示. 由表2可以得到: 1) 本文算法得到的Pareto最优解数远多于多目标IGAs, 这说明, 本文算法可以为用户提供更多的选择; 2) 根据SCHOOT J R 提出的计算分布度的方法^[9]得到两种算法的 Δ' , 其中, Δ' 反映算法得到的Pareto最优解集的分布情况. Δ' 值越小, 表示Pareto最优解集的分布越均匀, 反之, Δ' 值越大, 表示Pareto最优解集的分布越不均匀. 由表2知, 本文算法的 Δ' 小于多目标IGAs, 这说明, 本文算法得到的Pareto最优解集的分布比多目标IGAs均匀; 3) 本文算法得到的Pareto最优解的显式性能指标值小于多目标IGAs, 而隐式性能指标值大于多目标IGAs, 这说明, 本文算法得到的优化解的质量更高.

表 2 不同算法得到的Pareto最优解

Table 2 Pareto optimal solutions obtained by different algorithms

算法	最优 解数	Δ'	显式性能 指标值	隐式性能 指标值
本文算法	11.7	260.47	78490.93	54.29
多目标IGAs	5.6	309.38	79232.55	50.36

图3为两种算法在第10代得到的Pareto前沿, 其中: 图(a)为本文算法得到的Pareto前沿; 图(b)为多目标IGAs得到的Pareto前沿. 比较两图可知, 本文算法得到的Pareto最优解的数量多于多目标IGAs, 且分布比多目标IGAs均匀.

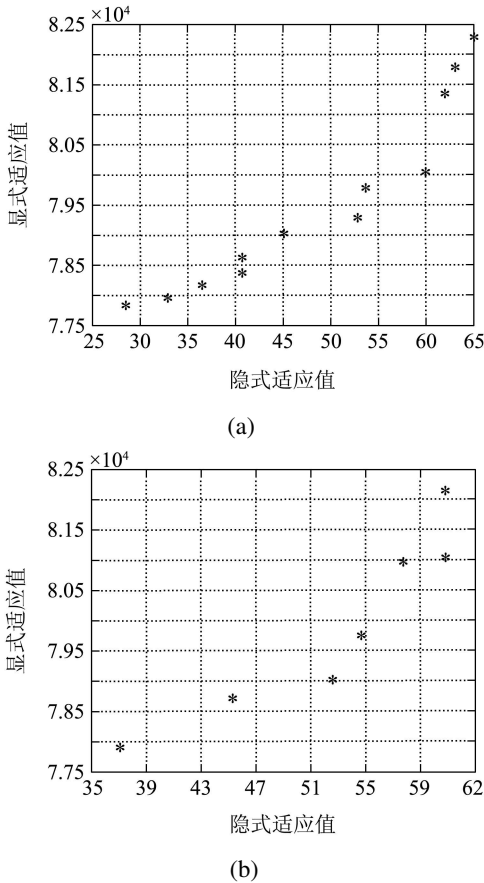


图 3 两种算法得到的Pareto前沿

Fig. 3 Pareto front obtained by two optimization algorithms

通过这些结果容易知道, 本文算法解决混合性能指标优化问题的性能是优越的.

5 结论(Conclusion)

混合性能指标优化问题是普遍存在的, 但是, 目前缺少有效的解决方法. 本文将传统的遗传算法和交互式遗传算法相结合, 提出一种解决该类问题的进化优化算法, 通过采用大规模进化种群、合理分配人机任务、基于相似度的估计策略以及Pareto占优等策略, 提高了计算机的使用效率, 增强了算法的搜索能力, 减轻了用户疲劳, 并且保证了显式性能指标和隐式性能指标同时达到较优. 将本文算法应用于室内布局这类混合性能指标优化问题, 并与多目标IGAs进行比较, 结果验证了所提算法的有效性.

需要指出的是, 在本文算法中, 基于相似度的估计策略主要针对离散空间. 如果优化的问题为连续空间, 可考虑采用Euclidean距离衡量个体间的相似度, 进而估计个体的隐式性能指标值. 另外, 采用交互式遗传算法优化隐式性能指标时, 由于人的认知具有模糊性和渐进性, 因此, 考虑更加合

理的评价方式和适应值表示形式,从而更好地反映人的认知规律,是笔者进一步的研究工作。

参考文献(References):

- [1] FUNES P, BONABEAU E, HERVE J, et al. Interactive multi-participant tour allocation[C] // *Proceedings of IEEE Congress on Evolutionary Computation*. Piscataway, NJ, USA: IEEE, 2004, 2: 1699 – 1705.
- [2] 周勇, 巩敦卫, 张勇. 混合性能指标优化问题的进化优化方法及应用[J]. 控制与决策, 2007, 22(3): 352 – 356.
(ZHOU Yong, GONG Dunwei, ZHANG Yong. Evolutionary optimization methods for hybrid index optimization problems and application[J]. *Control and Decision*, 2007, 22(3): 352 – 356.)
- [3] BRINTRUP A, RAMSDEN J, TIWARI A. An interactive genetic algorithm-based framework for handling qualitative criteria in design optimization[J]. *Computers in Industry*, 2007(58): 279 – 291.
- [4] KAMALIAN R, TAKAGI H, AGOGINO A M. Optimized design of MEMS by evolutionary multi-objective optimization with interactive evolutionary computation[C] // *Proceedings of Genetic and Evolutionary Computation Conference*. Berlin: Springer-Verlag, 2004: 1030 – 1041.
- [5] COELLO C A C, VELDHUIZEN D A, LAMONT G B. *Evolutionary Algorithms for Solving Multi-Objective Problem*[M]. Boston: Kluwer Academic Publishers, 2002.
- [6] LEE J Y, CHO S B. Sparse fitness evaluation for reducing user burden in interactive algorithm[C] // *Proceedings of 1999 IEEE International Fuzzy Systems Conference*. Piscataway, NJ, USA: IEEE, 1999, 2: 998 – 1003.
- [7] 郝国生, 巩敦卫, 史有群, 等. 基于满意域和禁忌域的交互式遗传算法[J]. 中国矿业大学学报, 2005, 34(2): 204 – 208.
(HAO Guosheng, GONG Dunwei, SHI Youqun, et al. Interactive genetic algorithm based on landscape of satisfaction and taboos[J]. *Journal of China University of Mining & Technology*, 2005, 34(2): 204 – 208.)
- [8] GONG D W, YUAN J, MA X P. Interactive genetic algorithms with large population size[C] // *Proceedings of IEEE Congress on Evolutionary Computation*. Piscataway, NJ, USA: IEEE, 2008: 1678 – 1685.
- [9] SCHOOT J R. *Fault tolerant design using single and multicriteria genetic algorithms optimization*[D]. Cambridge: Massachusetts Institute of Technology, 1995.

作者简介:

巩敦卫 (1970—), 男, 博士, 教授, 博士生导师, 目前研究方向为智能优化与控制的理论、方法与应用等, E-mail: dwgong@vip.163.com;

秦娜娜 (1983—), 女, 硕士, 目前研究方向为进化算法及其在室内布局中的应用, E-mail: qinnana186@126.com;

孙晓燕 (1978—), 女, 博士, 副教授, 目前研究方向为交互式进化计算, E-mail: xysun78@126.com.