

# 新的基于相似度估计个体适应值的交互式遗传算法

巩敦卫, 陈健<sup>†</sup>, 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:** 该方法根据组成个体各基因意义单元值出现的频率, 计算基因意义单元的权值, 并基于此得到个体的相似度; 根据个体与上代最优个体的相似度, 选择需要由用户进行评价的个体; 基于当代所有已评价个体的信息, 估计未评价个体的适应值. 将所提方法应用于窗帘进化设计系统, 并与已有典型方法比较. 结果表明, 所提方法在减轻用户疲劳和提高搜索性能等方面均具有优越性.

**关键词:** 遗传算法; 交互; 适应值; 估计; 相似度

**中图分类号:** TP273      **文献标识码:** A

## Novel interactive genetic algorithm for estimating individual fitness based on similarity

GONG Dun-wei, CHEN Jian<sup>†</sup>, SUN Xiao-yan

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

**Abstract:** In this algorithm, the weight of each gene meaning unit is calculated based on the emergence frequency of the corresponding allele. By comparing the weights, we determine the similarity of two individuals. Individuals in the current generation are selected for evaluation by the user according to the similarities between them and the most preferred one in the former generation. The fitness of unevaluated individuals is estimated based on the information of all evaluated individuals in the current generation. The proposed algorithm is applied to a curtain evolutionary design system, and compared with existing typical methods. The experimental results validate that the proposed algorithm has advantages in reducing user's fatigue and improving the efficiency in exploration.

**Key words:** genetic algorithms; interaction; fitness; estimation; similarity

### 1 引言(Introduction)

遗传算法(genetic algorithms, GAs)是一种模拟自然界生物进化和遗传变异机制的概率搜索方法, 由于该方法不要求目标函数可微或连续, 因此, 自提出以来, 已成功应用于很多复杂优化问题<sup>[1]</sup>. 但是, 对于性能指标无法明确表示的优化问题, 如产品设计、乐曲创作等, 传统的遗传算法不再适用. 为解决该问题, Dawkins将人的智能评价与传统遗传算法的进化机制相结合, 提出交互式遗传算法<sup>[2]</sup>. 不同于传统的遗传算法中, 使用明确定义的函数评价进化个体的优劣的策略, 在交互式遗传算法中, 通过将进化个体呈现给用户, 由用户根据主观评价为进化个体赋予适应值. 目前, 该方法已成功解决很多实际优化问题<sup>[3-6]</sup>.

在交互式遗传算法中, 个体适应值由用户赋予, 而频繁的人机交互将导致用户疲劳, 从而降低个体评价的准确性, 影响算法的性能. 为此, 交互式遗传

算法通常采用小规模种群和少进化代数, 这将大大限制算法的搜索性能. 在不增加用户疲劳的前提下, 采用大规模种群进化, 是提高算法搜索性能的有效策略, 并成为近年来交互式遗传算法研究的热点之一<sup>[7]</sup>.

采用大规模种群进化, 可能带来的问题是, 每进化一代, 需要评价很多个体. 但是, 用户疲劳的限制, 使得其只能评价部分个体; 而对其他个体, 需要采用一定的方法, 估计其适应值. 此时, 选择用户评价的个体, 以及个体适应值估计方法, 将影响大规模种群个体评价的准确度, 进而影响算法的搜索性能.

目前, 选择用户评价的个体, 以及估计用户未评价个体的适应值, 主要采用 $K$ -均值聚类方法, 将每代进化种群的个体划分为若干类, 由用户评价每类的中心个体, 并根据中心个体的适应值, 以及某个体与中心个体的距离, 估计该个体的适应值<sup>[8-9]</sup>. 该方法为大规模种群个体的评价提供了可行方案, 但是,

对用户评价个体的选择具有一定的随机性, 难以保证这些个体具有均匀的分布, 使得个体表现型之间的差异减小, 用户比较不同个体的负担增大; 此外, 已有的个体适应值估计策略仅利用中心个体的信息, 难以保证估计结果的准确度。

针对上述不足, 本文研究采用大规模种群进化时, 在减轻用户疲劳的前提下, 提高算法性能的问题, 并提出一种新的基于相似度估计个体适应值的交互式遗传算法。该方法从用户评价个体的分布性以及用户未评价个体适应值估计的准确性两方面入手: 首先, 根据组成个体各基因意义单元值出现的频率<sup>[10]</sup>, 计算基因意义单元的权值, 并基于此得到个体的相似度; 其次, 根据个体与上代最优个体的相似度, 选择用户评价的个体; 最后, 基于当代种群所有已评价个体的信息, 估计未评价个体的适应值。

该方法的优点在于, 个体的相似度计算融合了用户的偏好信息, 使得计算结果更加客观; 选择用户评价的个体, 在当代种群具有均匀的分布, 使得用户评价负担进一步降低; 估计用户未评价个体的适应值, 利用了用户评价的所有个体的信息, 使得估计结果更加准确。

本文的贡献主要体现在如下3个方面: 1) 提出基于基因意义单元值出现的频率, 计算基因意义单元权值的方法; 2) 提出基于个体与上代最优个体的相似度, 选择用户评价个体的方法; 3) 提出基于当代种群所有已评价个体的信息, 估计未评价个体适应值的方法。

本文剩余部分安排如下: 第2节综述相关工作, 包括: 已有的个体适应值赋值、个体相似度计算, 以及进化信息的利用方法等; 第3节详细阐述提出的方法, 包括: 基因意义单元的权值的计算、由用户评价的个体的选择、剩余个体的适应值的估计, 以及算法步骤等; 所提方法在窗帘进化设计系统的应用与对比实验在第4节给出; 最后, 第5节总结全文, 并指出需要进一步研究的问题。

## 2 相关工作(Related work)

### 2.1 个体适应值赋值(Assignment of individuals' fitness)

采用大规模种群进化时, 每代种群包含的个体数量大幅度增加, 使得算法的搜索性能进一步提高, 在较少的进化代内搜索到用户满意解的概率增大。如果采用传统的方法, 由用户评价每一个体的性能, 那么势必增加用户评价的负担。因此, 必须研究新的个体适应值赋值方法, 在不增加用户疲劳的前提下, 评价个体的性能。

Lee等提出的个体适应值估计方法中, 采用 $K$ -均值聚类, 把每代进化种群分成若干类, 用户仅评价

每类的中心个体, 基于中心个体的适应值及某个体与中心个体的距离, 估计该个体的适应值<sup>[8]</sup>。Kim等也采用 $K$ -均值聚类, 划分每代进化种群的个体, 但是, 估计的进化个体适应值基于标准测试函数, 主要验证大规模种群进化对算法搜索性能的影响<sup>[9]</sup>。Suga等提出的大规模种群交互式遗传算法中, 采用自组织映射网络, 将每代进化种群划分为若干类, 也是通过评价每类的中心个体, 估计其他个体的适应值<sup>[11]</sup>。巩敦卫等在大规模种群混合性能指标进化优化方法中, 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<sup>[12]</sup>, 并进一步设计了大规模种群交互式遗传算法, 并用于服装进化设计问题, 但采用的聚类策略, 使得每代进化种群划分的类个数不断变化<sup>[7]</sup>。

已有个体适应值赋值方法, 为大规模种群的进化提供了导向, 但是尚存在如下问题: 首先, 通过 $K$ -均值聚类划分每代种群, 选择的类中心个体具有随机性, 如果用户评价这些类中心个体, 那么, 用户宝贵的精力将消耗在这些具有较大随机性的个体的评价方面; 其次, 仅根据某类中心个体提供的信息, 估计该类中其他个体的适应值, 往往使得多个不同个体拥有相同的适应值, 从而不利于不同个体的比较, 以及种群的后续进化。这说明, 研究新的个体适应值赋值方法以提高种群的进化效率是非常必要的。

### 2.2 个体相似度计算(Calculation of individuals' similarity)

采用大规模种群进化时选择用户评价个体以及估计个体适应值, 往往利用个体相似度的信息, 因此, 个体相似度计算的准确性, 将在一定程度上影响个体评价的结果。

在范波等提出的基于用户相似度的协同过滤推荐方法中, 考虑用户对不同项目类型具有不同的偏好, 采用相关相似性和余弦相似性, 计算用户相似度<sup>[13]</sup>。卫俊霞等利用 $K$ -均值聚类和夹角余弦法解决多光谱分类问题时, 通过光谱之间的欧氏距离反映光谱相似度<sup>[14]</sup>。徐风苓等将用户上下文信息引入协同过滤推荐过程, 在预测移动网络服务环境下用户的偏好时, 采用余弦相似度计算不同上下文实例的特征相似度<sup>[15]</sup>。笔者曾利用大规模种群交互式遗传算法解决服装设计问题, 通过统计服装个体编码中相同的基因意义单元值, 计算个体的相似度<sup>[7]</sup>。

但是, 现有的个体相似度计算方法, 没有考虑组成个体的不同基因意义单元的差异。事实上, 用户对组成个体的不同基因意义单元往往具有不同的偏好, 而融入用户偏好的相似度计算将更有利于用户从众多个体中搜索到满意解。因此, 目前针对个体相似度计算的研究较少。这说明, 为了提高算法的搜索

性能,有必要研究融入用户偏好的相似度计算方法.

### 2.3 已有进化信息的利用(Utilization of the existing evolutionary information)

种群进化过程中保存的个体以及用户对个体的评价等信息能够在一定程度上反映用户的偏好及其变化过程. 如果基于这些信息进一步挖掘有价值的知识,并用于指导后续的种群进化,将能大幅度提高算法的性能. 目前,这方面的研究已经取得丰硕的成果.

Guo等设计的交互式遗传算法基于个体基因型提取用户偏好,并用于设计新的遗传操作以加快种群收敛的速度<sup>[16]</sup>. Tian等在语境信息检索中,将用户认知作为理解用户需求的重要语境因素,以提高检索效率和用户对结果的满意度<sup>[17]</sup>. Palle等提出临时存储的概念,通过二维空间储存优势个体,并在适当的时候将这些个体加入到进化种群中,提高种群的多样性<sup>[18]</sup>. Sun等通过网格存储进化信息,并基于这些信息挖掘有价值的知识,用于设计新的进化操作<sup>[19]</sup>. 洪榛等在利用差分进化算法求解多模优化问题时,通过分析群体个体间的 $N$ 阶最近邻,计算种群的全局分布,通过获取更多的全局最优解,提高算法的可靠性<sup>[20]</sup>. 此外, Gong等曾基于保存的种群进化信息,不断减小无效的搜索空间,以加速进化种群的收敛<sup>[21]</sup>. 进一步, Gong等又利用其他进化种群的信息,指导当前种群的进化,以减少种群进化的代数<sup>[22]</sup>.

可以看出,提取并利用种群进化信息的根本目的是提高算法的搜索性能. 但是,目前对种群进化信息的利用,仅限于改进遗传操作或缩减搜索范围等方面. 如果利用种群进化信息计算个体相似度,并基于该相似度估计个体适应值,将会提高个体适应值估计的准确度,从而提高算法的性能. 但是,目前相关的研究成果却非常少. 鉴于此,研究新的个体利用方法对提高算法的性能是非常有帮助的.

### 3 提出的方法(Proposed method)

不失一般性,考虑如下最大化问题:

$$\begin{cases} \max f(x), \\ \text{s.t. } x \in S \subseteq \mathbb{R}^n, \end{cases} \quad (1)$$

式中:  $f(x)$ 是被优化的性能指标,但是,无法用明确的函数表示;  $x$ 是 $N$ 维决策向量;  $S$ 为 $x$ 所在的空间. 在不引起混淆的情况下,仍记决策向量 $x$ 的编码及其搜索空间分别为 $x$ 和 $S$ . 本文采用大规模种群交互式遗传算法求解该问题,根据组成个体的各基因意义单元值出现的频率,计算基因意义单元的权值,并基于该权值得到个体的相似度;计算个体与上代最优个体的相似度,依据相似度的数值大小,以固定的间隔均匀选出部分个体交由用户进行评价;利用用户

评价的当代所有个体的信息,估计未评价个体的适应值.

### 3.1 基因意义单元权值的计算(Calculation of weights of gene meaning units)

在交互式遗传算法中,用户评价某一个体时,通常对组成该个体的不同基因意义单元具有不同的偏好. 计算个体的相似度时,如果用户对偏好的单元赋予较大的权值,得到的个体相似度将更客观地反映个体之间的相似关系. 由于选择用户评价个体和估计个体适应值均需要用到个体的相似度,因此,提高个体相似度计算的准确性,将有利于选择用户评价个体和估计个体适应值. 本节考虑通过组成个体的基因意义单元的权值,改进个体相似度的计算方法.

由于基因意义单元反映个体表现型某属性的编码片段,因此,某基因意义单元对应表现型的多样性越好,该基因意义单元收敛的速度越慢,说明用户投入了越少的精力用于评价该表现型,因此,用户对该基因意义单元偏好程度越低,从而该基因意义单元的权值越小,反之亦然. 基于此,本小节通过计算同代各基因意义单元的熵以及不同代熵的均值,得到基因意义单元的权值.

本小节采用基因意义单元的熵<sup>[23]</sup>,刻画种群的多样性. 记第 $t$ 代进化种群为 $\text{Pop}(t)$ ,其规模为 $N$ ,个体包含的基因意义单元数为 $n_U$ ,第 $i$ 个基因意义单元为 $U_i$ ,第 $j$ 个个体的第 $i$ 个基因意义单元值为 $U_i^j(t)$ ,与 $U_i^j(t)$ 相同的个体数为 $\alpha_i^j(t)$ ,则 $U_i$ 在第 $t$ 代进化种群的熵记为 $H_{U_i}(t)$ ,可以表示为

$$H_{U_i}(t) = -\frac{1}{N} \sum_{j=1}^N \log_2 \frac{\alpha_i^j(t)}{N}, \quad (2)$$

式中:

$$\begin{aligned} \alpha_i^j(t) &= \sum_{q=1}^N n(U_i^j(t), U_i^q(t)), \\ n(U_i^j(t), U_i^q(t)) &= \begin{cases} 1, & U_i^j(t) = U_i^q(t), \\ 0, & U_i^j(t) \neq U_i^q(t). \end{cases} \end{aligned}$$

由式(2)可知,某代进化种群具有 $U_i^j(t)$ 的个体越少,  $H_{U_i}(t)$ 越大,说明用户对该基因意义单元的偏好越不明确. 用户对某基因意义单元的偏好不明确存在很多原因,比如,在种群进化初期,用户对个体认知不完全;用户对该基因意义单元反映的表现型重视程度不够等. 这使得难以通过某一代的 $H_{U_i}(t)$ 确定用户对 $U_i$ 的偏好. 鉴于此,本小节首先对 $H_{U_i}(t)$ 归一化处理,对基于不同进化代得到的熵值赋予不同的权值,并将第1到第 $t$ 代熵值的加权平均值作为用户在第 $t$ 代对 $U_i$ 的偏好,记为 $\theta_i(t)$ ,则

$$\theta_i(t) = \frac{1}{t} \sum_{m=1}^t \frac{H_{\max} - H_{U_i}(m)}{H_{\max}} R(m), \quad (3)$$

式中:

$$R(m) = \begin{cases} 1 + e^{-T_s} - e^{-m}, & m < T_s, \\ 1, & m \geq T_s, \end{cases} \quad (4)$$

$H_{\max}$ 为基因意义单元熵的最大值. 在种群进化初期, 用户对个体的认知不明确, 此时, 各基因意义单元熵值的参考价值较小, 应赋予较小的权值; 种群进化一定代数后, 用户的认知逐渐清晰, 此时, 得到的各基因意义单元熵值具有较大的参考价值, 应赋予较大的权值, 这可以通过式(4)反映. 此外, 当用户对优化对象认知缓慢时,  $T_s$ 取较大的值; 反之, 取较小的值. 将 $\theta_i(t)$ 归一化后的值, 作为 $U_i$ 的权值, 记为 $\theta'_i(t)$ , 则

$$\theta'_i(t) = \frac{\theta_i(t)}{\sum_{i=1}^{n_U} \theta_i(t)}. \quad (5)$$

由式(5)可知,  $\theta'_i(t)$ 越大, 该基因意义单元的熵值越小, 因此, 对应表现型的差异越小, 说明用户投入精力用于该基因意义单元的评价越多, 从而用户越偏好该基因意义单元. 基于式(5), 能够得到组成个体的所有基因意义单元在第 $t$ 代的权向量, 记为 $\vec{\Theta}(t)$ , 则

$$\vec{\Theta}(t) = (\theta'_1(t), \theta'_2(t), \dots, \theta'_{n_U}(t)). \quad (6)$$

### 3.2 由用户评价的个体的选择(Selection of individuals evaluated by user)

为使得用户评价的个体在当代种群具有均匀的分布, 以方便用户进行评价个体比较, 进一步降低评价负担, 本小节提出基于个体相似度的用户评价个体选择方法. 该方法首先计算当代进化种群的个体与上代最优个体的加权距离, 进而得到它们的相似度; 然后, 依据相似度的数值大小, 以固定的间隔均匀选出若干个体交由用户进行评价. 该方法只计算同上代最优个体的加权距离, 可尽量减小计算量. 由于该步骤仅用于选择由用户评价的个体, 后续选择个体进行遗传操作时, 仍由用户的评价结果决定, 因此该方式不会造成进化方向被错误的引导.

记第 $t$ 代的最优个体为 $x_B(t)$ , 那么 $x_B(t-1)$ 即为第 $t-1$ 代的最优个体. 对于 $\text{Pop}(t)$ 的个体 $x_j(t)$ , 首先, 计算向量 $\vec{r}(x_j(t), x_B(t-1))$ 以反映 $x_j(t)$ 与 $x_B(t-1)$ 的基因意义单元的差异, 即

$$\vec{r}(x_j(t), x_B(t-1)) = (n(U_1^j(t), U_1^B(t-1)), n(U_2^j(t), U_2^B(t-1)), \dots, n(U_{n_U}^j(t), U_{n_U}^B(t-1)))^T, \quad (7)$$

式中:

$$n(U_i^j(t), U_i^B(t-1)) = \begin{cases} 0, & U_i^j(t) = U_i^B(t-1), \\ 1, & U_i^j(t) \neq U_i^B(t-1). \end{cases}$$

然后, 考虑由式(6)得到的各基因意义单元权值, 可得 $x_j(t)$ 与 $x_B(t-1)$ 的加权距离, 记为 $D(x_j(t), x_B(t-1))$ , 则

$$D(x_j(t), x_B(t-1)) = \vec{\Theta}(t) \cdot \vec{r}(x_j(t), x_B(t-1)). \quad (8)$$

由式(8)可知,  $x_j(t)$ 与 $x_B(t-1)$ 不同的基因意义单元越多, 且这些单元的权值越大,  $D(x_j(t), x_B(t-1))$ 越大, 此时,  $x_j(t)$ 与 $x_B(t-1)$ 的相似度应该越小, 反之亦然. 鉴于此,  $x_j(t)$ 与 $x_B(t-1)$ 的相似度, 记为 $\text{Sim}(x_j(t), x_B(t-1))$ , 可以表示为

$$\text{Sim}(x_j(t), x_B(t-1)) = 1 - \frac{D(x_j(t), x_B(t-1))}{\sum_{i=1}^{n_U} \theta'_i(t)}. \quad (9)$$

为了降低用户评价负担, 选择用户评价的个体时, 考虑这些个体在种群的分布, 依据当代个体同上代最优个体的相似度的数值大小, 以固定的间隔均匀选出个体. 记每代进化种群中, 用户评价的个体数为 $l$ , 依式(10)计算一系列参考相似度, 记为 $\text{Sim}'_k(t)$  ( $k = 1, 2, \dots, l$ ):

$$\text{Sim}'_k(t) = \text{Sim}_{\min}(x_j(t), x_B(t-1)) + \frac{k-1}{l-1} (\text{Sim}_{\max}(x_j(t), x_B(t-1)) - \text{Sim}_{\min}(x_j(t), x_B(t-1))), \quad (10)$$

式中:

$$\begin{aligned} \text{Sim}_{\min}(x_j(t), x_B(t-1)) &= \min_{j=1,2,\dots,N} \text{Sim}(x_j(t), x_B(t-1)), \\ \text{Sim}_{\max}(x_j(t), x_B(t-1)) &= \max_{j=1,2,\dots,N} \text{Sim}(x_j(t), x_B(t-1)). \end{aligned}$$

对于进化种群 $\text{Pop}(t)$ , 选择与 $x_B(t-1)$ 的相似度最接近 $\text{Sim}'_k(t)$  ( $k = 1, 2, \dots, l$ )的 $l$ 个个体, 提交用户评价.

### 3.3 剩余个体的适应值的估计(Estimation of other individuals' fitness)

对于用户未评价的个体, 需要采用合适的方法估计它们的适应值, 从而为后续的遗传操作提供依据. 为此, 本小节基于用户评价的所有个体的信息, 估计某个体的适应值: 首先, 根据用户评价的某个体的适应值和与该个体的相似度, 估计某未评价个体的适应值; 然后, 对于用户评价的所有 $l$ 个个体, 通过求取上述估计值的加权平均值, 作为该个体适应值的估计. 在计算加权平均值时, 权值即为该个体与用户评价个体的相似度.

记根据用户评价的个体 $x_j(t)$ 的适应值 $F_j(t)$ 和与 $x_j(t)$ 的相似度估计的未评价个体 $x_i(t)$ 的适应值为 $\hat{F}_i^j(t)$ ,  $F_{\max}$ 与 $F_{\min}$ 分别为个体适应值的上下限. 当 $\text{Sim}(x_i(t), x_B(t-1))$ 大于 $\text{Sim}(x_j(t), x_B(t-1))$ 时,  $\hat{F}_i^j(t)$ 应比 $F_j(t)$ 更接近 $F_{\max}$ , 也即 $\hat{F}_i^j(t)$ 应位于区间 $[F_j(t), F_{\max}]$ 内; 反之,  $\hat{F}_i^j(t)$ 应位于区间 $[F_{\min}, F_j(t)]$

内. 基于此,  $\hat{F}_i^j(t)$  可以表示为:

$$\text{当 } F_j(t) > \frac{F_{\max} + F_{\min}}{2} \text{ 时,}$$

$$\hat{F}_i^j(t) = F_j(t) + (\text{Sim}(x_i(t), x_B(t-1)) - \text{Sim}(x_j(t), x_B(t-1)))(F_{\max} - F_j(t)). \quad (11)$$

$$\text{当 } F_j(t) \leq \frac{F_{\max} + F_{\min}}{2} \text{ 时,}$$

$$\hat{F}_i^j(t) = F_j(t) + (\text{Sim}(x_i(t), x_B(t-1)) - \text{Sim}(x_j(t), x_B(t-1)))(F_j(t) - F_{\min}), \quad (12)$$

式中:  $i = 1, 2, \dots, N-l, j = 1, 2, \dots, l$ . 基于用户评价的所有  $l$  个个体, 可以得到  $x_i(t)$  适应值的估计值, 记为  $\hat{F}_i(t)$ :

$$\hat{F}_i(t) = \frac{\sum_{j=1}^l \hat{F}_i^j(t) \cdot \text{Sim}(x_j(t), x_i(t))}{\sum_{j=1}^l \text{Sim}(x_j(t), x_i(t))}. \quad (13)$$

由式(13)可知,  $x_j(t)$  与  $x_i(t)$  的相似度  $\text{Sim}(x_j(t), x_i(t))$  越大,  $\hat{F}_i^j(t)$  对  $\hat{F}_i(t)$  的影响越明显.

### 3.4 算法步骤(Steps of proposed algorithm)

将本文所提方法融入到传统的交互式遗传算法中, 能够得到新的基于相似度估计个体适应值的交互式遗传算法, 该算法的步骤如下:

**步骤 1** 设置算法所用参数的值, 包括: 交叉和变异概率、最大进化代数等, 令进化代数  $t = 0$ , 初始化种群;

**步骤 2** 根据第3.1节的方法, 计算组成个体各基因意义单元的权值;

**步骤 3** 根据第3.2节的方法, 选择用户评价的  $l$  个个体, 并提交用户评价;

**步骤 4** 根据第3.3节的方法, 估计当代进化种群用户未评价个体的适应值;

**步骤 5** 判断算法是否满足终止条件, 若是, 转步骤7;

**步骤 6** 实施选择、交叉和变异等遗传操作, 生成下一代种群, 令  $t = t + 1$ , 转步骤2;

**步骤 7** 输出最优个体, 算法结束.

## 4 在窗帘进化设计系统的应用(Application in curtain evolutionary design system)

### 4.1 问题背景(Backgrounds)

窗帘作为日常生活的必需品, 需要满足美观、实用的要求. 传统的窗帘设计过程非常复杂, 需要同时考虑使用场合、图案以及颜色搭配等诸多因素, 因此, 即使经过系统培训的设计师也很难设计出一款优秀的作品.

事实上, 人对窗帘的评价主要依据其外观, 即图案和配色, 因此, 可将窗帘设计问题转化为一个组合

优化问题, 由设计者从不同的图案和配色中, 选出最满意的方案. 由于人的偏好模糊多变, 且不同人的偏好存在较大差异, 因此, 很难找到一个统一且精确的目标函数对设计方案进行评价. 这说明, 采用传统的遗传算法无法求解该问题, 但却可利用交互式遗传算法解决.

本文的窗帘进化设计系统基于 Visual C++ 6.0 开发, 利用 3ds Max 2009 构建窗帘部件的 3D 模型, 包括: 帘头、侧帘以及窗纱等 3 部分, 分别记为基因意义单元 1, 2, 3. 系统为每一部分建立独立的素材库, 以图片形式存储具有不同图案和配色的布料素材, 通过 OpenGL 技术对每一部分做贴图处理, 从而将其拼接成完整的窗帘, 呈现给用户. 该系统通过交互式遗传算法驱动进化进程, 设计者需要根据 3 部分的图案、颜色以及搭配效果等, 对方案做出合理评价, 进化数据以 Excel 表的形式记录. 由于目前还没有将交互式遗传算法应用到窗帘设计的先例, 因此, 该系统在验证本文方法性能的同时, 也为窗帘设计提供了一种新颖的途径.

### 4.2 实验设计(Experiment design)

在第3.1节计算基因意义单元的权值时, 设置了参数  $T_s$ , 因此, 需要给出  $T_s$  的合理取值. 为此, 第1组实验记录不同  $T_s$  取值的情况下, 用户对不同基因意义单元的权值, 并根据实验结果, 设定  $T_s$  的值. 为了验证所提方法能够合理确定用户对不同基因意义单元的偏好, 第2组实验记录不同用户评价个体过程中历代的权向量  $\vec{\Theta}(t)$ . 此外, 第3组实验将基于本文方法得到的交互式遗传算法, 简称为本文方法, 与小规模种群交互式遗传算法(traditional interactive genetic algorithm, TIGA)以及基于聚类估计个体适应值的大规模种群交互式遗传算法(large population interactive genetic algorithm with clustering strategy, LIGA-CS)对比, 验证所提方法在提高搜索效率和减轻用户疲劳方面的有效性.

### 4.3 遗传操作及参数设置(Genetic operations and parameter settings)

分别基于 TIGA、LIGA-CS 以及本文方法, 开发相应的系统. 在 TIGA 中, 用户对进化种群的个体赋予 0 到 100 之间的精确数, 作为其适应值; 在 LIGA-CS 中, 用户对进化种群划分后的每类中心个体赋予精确适应值, 其他个体的适应值根据所属类的中心个体的适应值估计. 具体而言, 依据式(10)选出用户评价个体, 作为每类的中心个体, 其他个体按最近邻法完成聚类. 用户对中心个体评价后, 依据式(11)或式(12)估计其他个体的适应值. 此外, LIGA-CS 中各基因意义单元的权值均为  $\frac{1}{3}$ .

对于采用大规模种群进化的交互式遗传算法, 种群规模  $N = 100$ . 为了不增加评价负担, 用户每代评

价的个体数为8, 其他个体的适应值由系统根据相应的策略估计. 对于采用小规模种群进化的交互式遗传算法, 种群规模等于用户每代评价的个体数, 即  $N = 8$ . 所有方法均采用相同的遗传策略和控制参数, 包括: 规模为2的联赛选择、单点交叉以及单点变异, 且交叉和变异概率分别为0.6和0.05, 最大进化代数  $T_{max} = 20$ . 进化终止条件为种群进化到最大代数. 此外, 如果某代进化种群中有75%以上完全相同的个体, 或用户已经找到满意解, 那么, 手动终止种群进化.

#### 4.4 交互界面(Interactive interface)

系统的人机交互界面如图1所示. 该界面的主体为8个具有3D效果的窗帘, 用户可以任意角度旋转窗帘, 进行全方位观察. 系统的所有操作均集成在菜单栏中, 包括: “效果”、“进化”以及“实验记录”等. 通过“效果”菜单, 用户能够改变颜色和场景. 系统支持对背景色、环境光色、镜反射色以及漫反射色等的自定义设置, 还支持反混淆、光照、阴影以及旋转等场景模式. 与种群进化相关的操作均集成到“进化”菜单中. 通过“实验记录”菜单, 用户能够设置实验结果的保存路径及文件名.

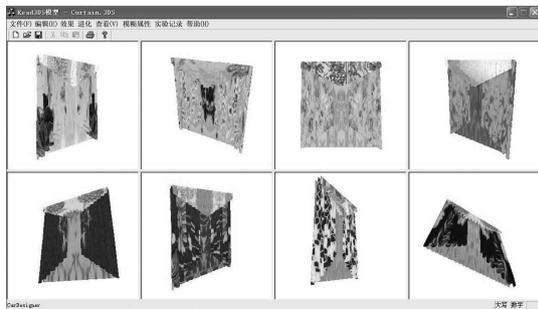


图 1 人机交互界面

Fig. 1 Human-computer interface

图2为系统的进化操作界面, 用户能够在参数设置部分设置交叉和变异概率, 以及最大进化代数. 个体评价部分供用户对窗帘个体评价. 此外, 系统还提供计时功能, 以记录用户的评价时间.

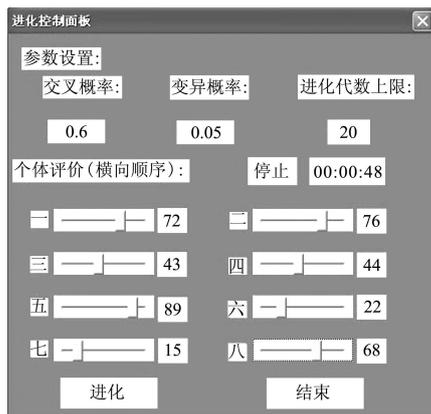
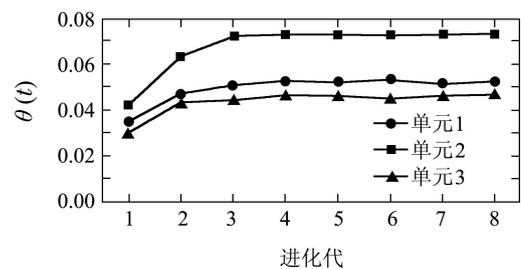


图 2 进化操作界面

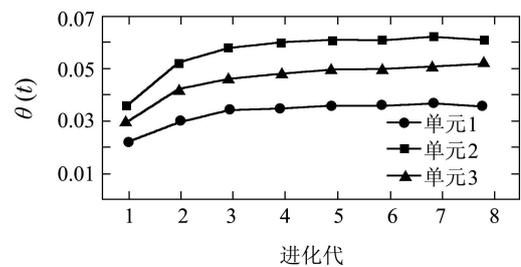
Fig. 2 Interface of evolutionary operations

#### 4.5 实验结果与分析(Experimented results and analysis)

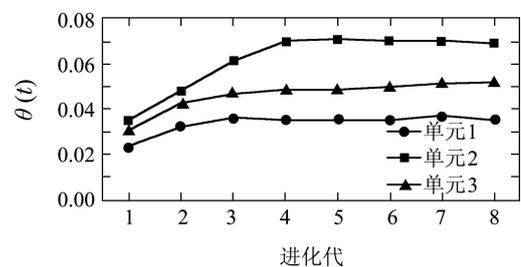
首先, 通过实验获取最优的  $T_s$  值. 由于不同的  $T_s$  值直接影响  $\theta_i(t)$  的大小, 因此, 针对每一  $T_s$  取值, 分别独立进行5次实验, 记录每次实验每代进化种群所有  $\theta_i(t)$  的值, 并求取5次实验相同代  $\theta_i(t)$  的平均值. 由于本文的实例仅有3个基因意义单元, 用户可在较少进化代内熟悉个体, 因此, 仅讨论  $T_s = 2, 3, 4, 5$  等4种情况. 不同的实验种群进化结束需要的代数不同, 且在后续进化中  $\theta_i(t)$  的变化较小, 因此, 只取前8代的数据进行分析. 实验结果如图3(a)–3(d)所示, 其中: 横坐标表示进化代, 纵坐标表示用户对不同基因意义单元的偏好.



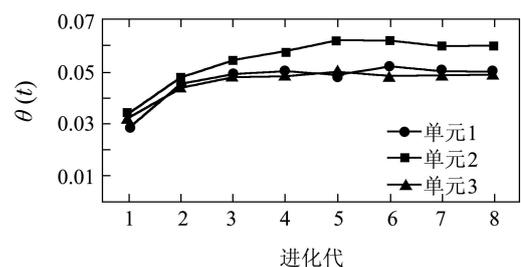
(a)  $T_s = 2$



(b)  $T_s = 3$



(c)  $T_s = 4$



(d)  $T_s = 5$

图 3  $T_s$  取值对用户对各基因意义单元偏好的影响

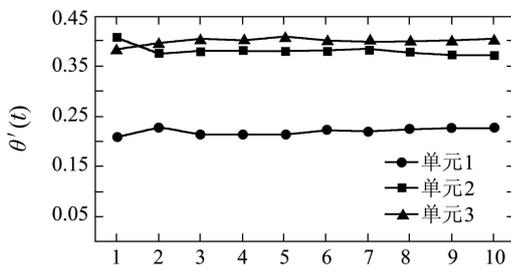
Fig. 3 Influence of different  $T_s$  on user's preferences over gene meaning units

由图3可知:

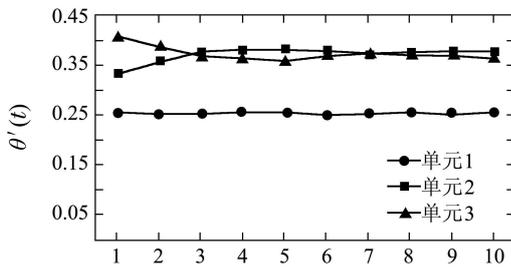
1) 对于所有 $T_s$ 取值,  $\theta_i(t)$ 的变化规律均相同. 任何一个基因意义单元, 在种群进化初期, 用户对该基因意义单元的偏好均较小; 随着种群的不断进化, 偏好值逐渐增大, 并在某一进化代后趋于稳定. 这由 $\theta_i(t)$ 从(0.035, 0.042, 0.03)( $T_s = 2$ ), (0.022, 0.036, 0.03)( $T_s = 3$ ), (0.024, 0.035, 0.031)( $T_s = 4$ ), 以及(0.028, 0.034, 0.032)( $T_s = 5$ )分别增加到(0.052, 0.073, 0.046)( $T_s = 2$ ), (0.036, 0.061, 0.052)( $T_s = 3$ ), (0.036, 0.069, 0.052)( $T_s = 4$ ), 以及(0.05, 0.06, 0.049)( $T_s = 5$ )明显看出. 这是因为, 在种群进化初期, 个体表现型差异较大, 因此,  $H_{U_i}(t)$ 较大, 且 $R(t)$ 较小, 使得由式(3)得到的 $\theta_i(t)$ 均较小; 随着种群的不断进化, 用户偏好逐渐明确, 个体表现型差异不断缩小, 因此,  $H_{U_i}(t)$ 不断减小, 同时 $R(t)$ 增大, 导致 $\theta_i(t)$ 逐渐增大. 上述 $\theta_i(t)$ 的变化完全符合用户的认知规律.

2)  $T_s$ 的取值越小,  $\theta_i(t)$ 趋于稳定的速度越快. 当 $T_s = 2$ 时, 从第3代开始,  $\theta_i(t)$ 即趋于稳定; 当 $T_s = 3$ 时, 从第4代开始,  $\theta_i(t)$ 趋于稳定; 当 $T_s = 4$ 时, 从第4代开始,  $\theta_i(t)$ 也趋于稳定; 当 $T_s = 5$ 时, 从第5代开始,  $\theta_i(t)$ 才趋于稳定. 为了更快的获取稳定的 $\theta_i(t)$ 值, 本文在后续实验中, 设定 $T_s = 2$ .

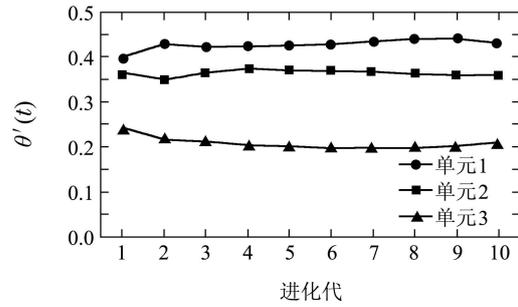
其次, 验证所提方法能够合理确定用户对不同基因意义单元的偏好. 由于本文采用归一化后的偏好作为基因意义单元的权值, 因此, 4个用户独立进行5次实验, 记录每次实验每代进化种群所有 $\theta'_i(t)$ 的值, 并求取它们的平均值. 由于不同实验种群进化结束时的代数不同, 因此, 仅取前10代的数据进行分析. 实验结果如图4(a)–4(d)所示, 其中: 横坐标表示进化代, 纵坐标表示基因意义单元的权值, 即 $\theta'_i(t)$ 的值.



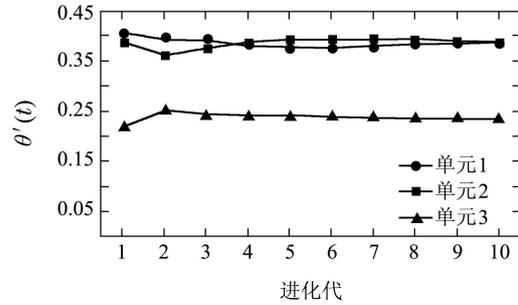
(a) 用户1



(b) 用户2



(c) 用户3



(d) 用户4

图4 不同用户实验得到的基因意义单元权值  
Fig. 4 Gene meaning units' weights for different users

由图4可知:

1) 对于任一用户, 在初始若干进化代内, 各基因意义单元的权值会有一定的波动, 之后趋于稳定. 例如, 用户1, 2和3在前3代的权值均波动明显, 而用户4直到第5代, 各基因意义单元的权值才趋于稳定. 这是由于, 用户的偏好通过反复的人机交互来逐渐引导种群进化的, 只有不断修正以前的权值, 才能客观反映用户的偏好, 这符合用户的认知规律.

2) 不同用户实验, 同一基因意义单元的权值有明显差异. 例如, 用户1和2实验时, 单元1, 即帘头部分的权值明显小于侧帘及窗纱部分. 可能的原因是, 在窗帘实际效果图中, 帘头部分的面积所占比例较小, 对整体效果的影响也较小. 而用户3和4实验时, 窗纱部分的权值最小. 这是由于, 窗纱部分经常被帘头和侧帘遮挡, 降低了用户对该部分的重视程度.

上述实验结果与分析表明, 所提方法能够合理确定用户对不同基因意义单元的偏好, 从而得到准确的基因意义单元权值.

为了比较不同方法的性能, 独立运行本文和对比方法各10次, 记录种群进化需要的代数、用户评价耗时、搜索效率(通过搜索到的互异个体数与总的个体数的比值反映), 以及种群进化结束时, 用户是否成功找到满意解等, 实验结果如图5和表1所示. 鉴于实验数据不服从正态分布, 因此, 本文采用Mann-Whitney U方法, 进行非参数检验, 以说明不同方法关于某些性能指标差异的显著性. 取显著性水平为0.05, 得到的渐进显著性水平如表2所列.

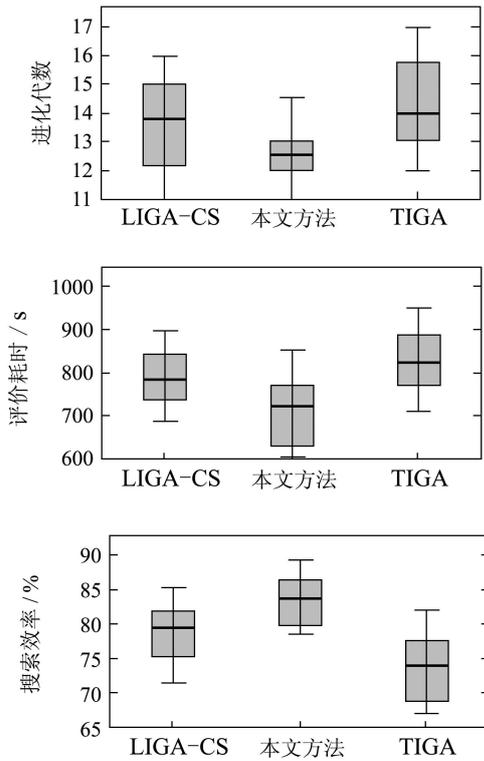


图 5 不同方法的箱型图

Fig. 5 Box-plots of different algorithms

表 1 不同时间内找到满意解的比例/%

Table 1 Rates of satisfactory solutions in different time intervals

	12 min	14 min	16 min
本文方法/%	50	90	100
LIGA-CS/%	30	70	100
TIGA/%	10	40	70

表 2 对比方法的渐进显著性水平

Table 2 The asymptomatic significance of comparative algorithms

	进化代数	评价耗时/s	搜索效率
本文方法VS. LIGA-CS	0.043	0.041	0.023
本文方法VS. TIGA	0.002	0.005	0.001

由图5及表1和2可知:

1) 在进化代数方面, 本文方法与两种对比方法的渐进显著性水平均小于0.05, 说明本文方法与对比方法存在显著差异; 再结合图5(a)实验数据的分布可知, 本文方法在该性能指标上优于其他两种方法. 原因在于, 与LIGA-CS相比, 本文基于相似度和用户评价的全部个体, 估计的未评价个体适应值更加准确, 正确引导了种群朝着用户偏好的方向进化; 与TIGA相比, 本文方法每代搜索到的个体更多, 增大了用户找到满意解的概率. 因此, 本文方法需要的进化代数较少.

2) 关于用户评价耗时, 与LIGA-CS和TIGA相比, 本文方法需要的评价时间更短, 且本文方法与对比方法的差异显著, 这有利于减轻用户疲劳. 原因在于, 本文提出的用户评价个体选择策略, 保证了所选个体表现型的差异性, 便于用户比较不同个体的优劣, 从而能够更快的评价个体.

3) 对于搜索效率, 本文方法显著高于LIGA-CS和TIGA, 这说明, 在种群进化过程中, 本文方法搜索到的互异个体占的比例高. 原因在于, 在LIGA-CS中, 个体适应值估计不够准确, 使得部分个体的适应值偏大, 从而降低了后代种群的多样性. 对于TIGA, 由于种群规模小、种群多样性差, 使得种群进化陷入了局部收敛, 从而造成搜索到的互异个体比例最低.

4) 采用大规模种群进化, 用户能在算法结束时找到满意个体; 而当采用小规模种群进化时, 算法会因进化种群个体的相似度过高而提前结束, 造成用户无法搜索到满意个体. 此外, 与LIGA-CS相比, 本文方法找到满意解的速度更快, 这从需要的进化代数和用户评价耗时等指标上也能够看出.

上述实验结果与分析表明, 本文方法能够以最少的进化代数和用户评价耗时, 找到最多的互异个体. 这表明, 本文方法能够在减轻用户疲劳的前提下提高搜索效率.

## 5 结论(Conclusions)

采用大规模种群进化从而在每代搜索到更多的个体, 成为提高交互式遗传算法性能的重要途径之一. 本文研究采用大规模种群进化时, 在减轻用户疲劳的前提下, 提高算法性能的问题, 并提出一种新的基于相似度估计个体适应值的交互式遗传算法. 该方法根据组成个体各基因意义单元值出现的频率, 计算基因意义单元的权值, 并基于此得到个体的相似度, 使得个体相似度的计算结果更加客观; 计算当前代个体与上代最优个体的相似度, 按相似度的大小均匀选出部分个体供用户评价, 使得用户评价负担进一步降低; 基于当代种群用户评价的所有个体的信息估计未评价个体的适应值, 使得估计结果更加准确.

将所提方法应用到窗帘进化设计系统中, 并与已有的典型方法比较. 实验结果表明, 所提方法能够在减轻用户疲劳的前提下提高搜索效率.

需要指出的是, 本文提出的用户评价个体选择方法不能够保证性能优秀的个体由用户评价, 或者说, 这些个体的适应值往往通过本文的方法估计得到. 但是, 这些个体往往反映了种群进化的方向, 适应值估计的不准确将会影响种群的进化进程. 这说明, 选择用户评价个体时, 同时考虑个体性能和分布性对于进一步提高算法性能是十分有帮助的; 此外, 采用

合适的方法进一步提高个体适应值估计的准确性也有利于进一步提高算法的性能。这些都是今后需要进一步研究的问题。

### 参考文献(References):

- [1] GOLDBERG D E. *Computer-aided gas pipeline using genetic algorithm and rule learning* [D]. Michigan, America: University of Michigan, 1983.
- [2] DAWKINS R. *The Blind Watchmaker* [M]. Essex: Longman, 1986.
- [3] ONO S, NAKAYAMA S. A system for decorating QR code with facial image based on interactive evolutionary computation and case-based reasoning [C] // *Proceedings of the 2nd World Congress on Nature and Biologically Inspired Computing*. Kitakyushu: IEEE, 2010: 401 – 406.
- [4] MIGUEL A H, FRANCESC J F, SALVADOR M P. Distance-based relevance feedback using a hybrid interactive genetic algorithm for image retrieval [J]. *Applied Soft Computing*, 2011, 11(2): 1782 – 1791.
- [5] MEGHNA B S, BARBARA M. A case-based micro interactive genetic algorithm (CBMIGA) for interactive learning and search: methodology and application to groundwater monitoring design [J]. *Environmental Modeling and Software*, 2010, 25(10): 1176 – 1187.
- [6] ONO S, NAKAYAMA S. Fusion of interactive and non-interactive evolutionary computation for two-dimensional barcode decoration [C] // *Proceedings of IEEE Congress on Evolutionary Computation*. Barcelona: IEEE, 2010: 18 – 23.
- [7] GONG D W, YUAN J, MA X P. Interactive genetic algorithm with large population size [C] // *Proceedings of IEEE Congress on Evolutionary Computation*. Hong Kong: IEEE, 2008: 1678 – 1685.
- [8] LEE J Y, CHO S B. Sparse fitness evaluation for reducing user burden in interactive genetic algorithm [C] // *Fuzzy Systems Conference Proceedings*. Seoul: IEEE, 1999: 998 – 1003.
- [9] KIM H S, CHO S B. An efficient genetic algorithm with less fitness evaluation by clustering [C] // *Proceedings of IEEE Congress on Evolutionary Computation*. Seoul: IEEE, 2001: 887 – 894.
- [10] 巩敦卫, 郝国生, 周勇, 等. 交互式遗传算法原理及应用 [M]. 北京: 国防工业出版社, 2007.  
(GONG Dunwei, HAO Guosheng, ZHOU Yong, et al. *Interactive Genetic Algorithms Theory and Application* [M]. Beijing: National Defense Industry Press, 2007.)
- [11] SUGA Y, IKUMA Y, OGATA T. Interactive evolution of human robot communication in real world [C] // *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and System*. Barcelona: IEEE, 2005: 1438 – 1443.
- [12] 巩敦卫, 秦娜娜, 孙晓燕. 混合性能指标优化问题的大种群规模进化算法 [J]. *控制理论与应用*, 2010, 27(6): 769 – 774.  
(GONG Dunwei, QIN Nana, SUN Xiaoyan. Evolutionary algorithm with large population size for problems with hybrid indices [J]. *Control Theory & Application*, 2010, 27(6): 769 – 774.)
- [13] 范波, 程久军. 用户间多相似度协同过滤推荐算法 [J]. *计算机科学*, 2012, 39(1): 23 – 26.  
(FAN Bo, CHENG Jiujun. Collaborative filtering recommendation algorithm based on user's multi-similarity [J]. *Computer Science*, 2012, 39(1): 23 – 26.)
- [14] 卫俊霞, 相里斌, 高晓惠, 等. 基于 $K$ -均值聚类与夹角余弦法的多光谱分类算法 [J]. *光谱学与光谱分析*, 2011, 31(5): 1357 – 1360.  
(WEI Junxia, XIANG Libin, GAO Xiaohui, et al. The multi-spectra classification algorithm based on  $K$ -means clustering and spectral angle cosine [J]. *Spectroscopy and Spectral Analysis*, 2011, 31(5): 1357 – 1360.)
- [15] 徐风苓, 孟祥武, 王立才. 基于移动用户上下文相似度的协同过滤推荐算法 [J]. *电子与信息学报*, 2011, 33(11): 2785 – 2789.  
(XU Fengling, MENG Xiangwu, WANG Licai. A collaborative filtering recommendation algorithm based on context similarity for mobile users [J]. *Journal of Electronics & Information Technology*, 2011, 33(11): 2785 – 2789.)
- [16] GUO Y N, GONG D W, YANG D Q. Interactive genetic algorithm based on implicit knowledge model [C] // *Proceedings of the 6th International Conference on Simulated Evolution and Learning*. Hefei: Springer-Verlag, 2006: 369 – 376.
- [17] TIAN X, DU X Y, HU H. Modeling individual cognitive structure in contextual information retrieval [J]. *Computers and Mathematics with Applications*, 2009, 57(6): 1048 – 1056.
- [18] PALLE D. On the role of temporary storage in interactive evolution [C] // *Proceedings of Application of Evolutionary Computing-EvoWorkshops*. Tubingen: Springer-Verlag, 2009: 478 – 487.
- [19] SUN X Y, CHEN J. Grid-based knowledge-guided interactive genetic algorithm and its application to curtain design [C] // *Proceedings of the 2nd World Congress on Nature and Biologically Inspired Computing*. Kitakyushu: IEEE, 2010: 395 – 400.
- [20] 洪榛, 张贵军, 俞立. 基于 $N$ 阶近邻分析的自适应差分进化算法 [J]. *控制理论与应用*, 2011, 28(11): 1613 – 1620.  
(HONG Zhen, ZHANG Guijun, YU Li. Adaptive differential evolution algorithm based on  $N$ th-order nearest-neighbor analysis [J]. *Control Theory & Applications*, 2011, 28(11): 1613 – 1620.)
- [21] GONG D W, HAO G S, SHI Y Q. Interactive genetic algorithm with holding down survival of the fittest based on extinction mechanism [J]. *International Journal of Information Technology*, 2005, 11(10): 11 – 20.
- [22] GONG D W, ZHOU Y, LI T. Cooperative interactive genetic algorithm based on user's preference [J]. *International Journal of Information Technology*, 2005, 11(10): 1 – 10.
- [23] 樊昌信, 曹丽娜. 通信原理 [M]. 第6版. 北京: 国防工业出版社, 2001: 4 – 14.  
(FAN Changxin, CAO Lina. *Principle of Communication* [M]. 6th edition. Beijing: National Defense Industry Press, 2001: 4 – 14.)

### 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;

**陈健** (1987–), 男, 硕士研究生, 目前研究方向为进化算法及其在产品中的应用, E-mail: chenjian121206@163.com;

**孙晓燕** (1978–), 女, 博士, 教授, 硕士生导师, 目前研究方向为交互式进化计算, E-mail: xysun78@126.com.