

基于进化个体模糊适应值估计的交互式遗传算法

郭广颂^{1†}, 陈良骥², 文振华¹, 侯军兴¹

(1. 郑州航空工业管理学院 机电工程学院, 郑州 450046; 2. 天津工业大学 机械工程学院, 天津 300387)

摘要: 采用大规模种群进化优化策略, 根据用户评价时间和单一数值适应值估计个体模糊适应值; 根据个体表现型属性和参照个体模糊适应值宽度计算个体表现型相似度; 利用个体表现型相似度对种群聚类并估计未评价个体的模糊适应值; 基于个体模糊适应值和表现型相似性构造个体选择适应值, 实现个体相似性选择. 将所提出方法应用于室内挂钟进化设计, 并与已有典型方法进行比较. 结果表明, 所提出方法在提高优化质量、减轻用户疲劳、提高搜索效率等方面均具有优越性.

关键词: 遗传算法; 交互; 估计; 模糊适应值; 相似度

中图分类号: TP273

文献标志码: A

Interactive genetic algorithms based on estimation of individuals' fuzzy fitness

GUO Guang-song^{1†}, CHEN Liang-ji², WEN Zhen-hua¹, HOU Jun-xing¹

(1. School of Mechatronics Engineering, Zhengzhou University of Aeronautics, Zhengzhou 450046, China; 2. School of Mechanical Engineering, Tianjin Polytechnic University, Tianjin 300387, China)

Abstract: Individuals' fuzzy fitness is estimated according to user's evaluation time and single numerical fitness by using the large-scale population evolutionary optimization strategy. Individual phenotype similarity is calculated according to the individuals' phenotype attributes and the width of individuals' fuzzy fitness. The population is divided into several clusters and the not assigned individuals' fuzzy fitness is estimated by using individuals' phenotypes similarity. The individuals' choice fitness is constructed based on individuals' fuzzy fitness and phenotype similarity to achieve individual similarity selection. The proposed algorithm is applied to a wall clock evolutionary design system, and compared with existing typical ones. The experimental results show that the proposed algorithm has advantages in improving optimization quality and alleviating user fatigue while improving its efficiency in exploration.

Keywords: genetic algorithms; interaction; estimation; fuzzy fitness; similarity

0 引言

基于传统遗传算法的优化流程, 针对很难用显式性能指标表达的用户偏好和经验等隐式性能指标优化问题, 交互式遗传算法表现出算法实现简单、优化效果显著的特点. 因此, 近年来, 交互式遗传算法在图像处理^[1]、机器人控制^[2]、语音合成^[3]、城市景观设计^[4]等众多领域获得了成功的应用. 不同于单纯数值优化算法, 人参与的交互式遗传算法进化过程存在易疲劳的特点, 同时受交互场景限制, 种群规模和进化代数一般小于20, 这严重影响了算法的优化能力. 采用大规模进化种群优化, 可以在不增加用户操

作负担的前提下, 扩大搜索范围, 是提高算法优化能力的有效策略, 目前已成为交互式遗传算法研究的热点之一.

采用大规模种群进化, 需要解决两个问题: 1) 个体聚类; 2) 个体适应值分配. 为了不增加用户负担, 用户只能评价少量个体, 其他个体适应值需通过相似性聚类进行分配. 目前, 个体聚类主要依据个体基因相似性采用 K -均值^[5] 或 C -均值方法^[6], 由于个体基因型与表现型存在差异, 对于以表现型为偏好决定因素的隐式性能指标优化具有很大的局限性. 另外, 用户评价的信息量越丰富, 适应值分配的信息量也越

收稿日期: 2017-05-17; 修回日期: 2017-07-23.

基金项目: 国家自然科学基金项目(61673196); 河南省科技攻关项目(172102210513); 河南省高等学校重点科研项目(18A120012).

责任编委: 陈家伟.

作者简介: 郭广颂(1978—), 男, 副教授, 硕士, 从事智能控制与进化优化等研究; 陈良骥(1978—), 男, 教授, 博士, 从事智能控制与制造等研究.

[†]通讯作者. E-mail: guogs78@126.com

多,优化质量也越高.但目前的大规模种群适应值分配策略大多基于个体基因距离^[7]、基因模式^[8]或基因共享分配策略,这会导致适应值估计误差大幅增加.

本文提出一种基于进化个体模糊适应值估计的交互式遗传算法(Interactive genetic algorithms based on estimation of individuals' fuzzy fitness, IGA-BEIFF).该方法首先根据用户评价时间和评价结果估计用户评价个体的模糊适应值;然后,根据个体表现型相似度对种群聚类,并分配个体模糊适应值;最后,通过改进的个体选择策略及遗传操作生成新种群.该方法的优点在于:1)在用户评价个体阶段,用户无需选择语气算子,只按单一数值评价即可获得模糊适应值,大大减轻了操作量;2)采用更加符合评价规律的个体表现型相似度分配个体适应值不仅操作简单,而且优化结果更加准确,进化方向更加清晰;3)采用大规模进化种群进化,所有个体适应值均为模糊适应值,偏好表达更加全面,搜索能力更强.

1 提出方法

不失一般性,考虑如下优化问题:

$$\max f(x, t) = f(x(t)), x \in S \subseteq R^D.$$

其中: $f(x, t)$ 是不能用明确定义函数表示的优化性能指标, x 是 D 维决策变量, S 是其取值范围. 由于模糊适应值能更好地表达用户偏好^[9], $f(x(t))$ 在论域 U 的隶属函数采用高斯型模糊数,有

$$\forall u \in [u_i, u_j] \subseteq U, \mu_{f(x,t)}(u) \geq \min\{\mu_{f(x,t)}(u_i), \mu_{f(x,t)}(u_j)\} \mu_{f(x,t)}(u) = e^{-\frac{1}{2}(\frac{u-d}{\sigma})^2}. \quad (1)$$

其中: d 是 $f(x, t)$ 的中心, σ 是 $f(x, t)$ 的宽度.

本文基于个体模糊适应值,采用大规模种群进化优化策略求解该问题.首先,从种群中选择代表性个体推荐给用户评价,根据评价结果估计代表性个体的模糊适应值;其次,根据种群内个体与代表性个体的表现型相似度对种群聚类,并在各子类内估计剩余个体的模糊适应值;再次,根据个体模糊适应值构建个体选择适应值,用于个体的选择;最后,通过遗传操作获得新种群.

1.1 个体相似性计算

在交互式遗传算法中,用户对个体的评价情况可以通过个体间的相似性/差异反映.根据评价经验,个体表现型的相似性具有两个特征:1)属性间的相似性,即个体属性越相似,表现型也越相似;2)非偶对称性,即比较个体与参照个体的相似度和参照个体与比较个体的相似度不完全一致.现实中,个体的表现型相似性与参照样本选取有关,以个体A的外观为参

照的个体B的外观相似度和以个体B的外观为参照的个体A的外观相似度不完全相同.例如,从主观评价角度看,人与猩猩的外观相似度和猩猩与人的外观相似度便不完全相同.人的这个评价特征体现了主观评价的相对性和不确定性,对于评价结果有着重要影响,但以往研究对此并未关注.

基于相似性的上述特征,记第 t 代种群为 $x(t)$, 种群规模为 N , 种群 $x(t)$ 中的第 i 个个体为 $x_i(t)$, $i = 1, 2, \dots, N$, 则 $x_i(t)$ 的表现型可以表示为 $x_i(t) = \{x_{i1}, x_{i2}, \dots, x_{iNg} | x_{ir} \in \{l_1^r, l_2^r, \dots, l_s^r\}\}$. 其中: x_{ir} ($r = 1, 2, \dots, Ng$) 为组成个体的 r 个属性; $l_1^r, l_2^r, \dots, l_s^r$ 是 x_{ir} 的属性值. 考虑两个个体 $x_i(t)$ 和 $x_j(t)$, 设 $x_j(t)$ 是参照个体, $x_i(t)$ 是比较个体, 基于认知的模糊性, 采用高斯函数 μ_{ij} 描述个体 $x_i(t)$ 与 $x_j(t)$ 的表现型相似关系, 即

$$\mu_{ij}(x_{ir}) = e^{-\left(\frac{x_{ir}-x_{jr}}{\sigma(x_j(t))}\right)^2}, \quad (2)$$

其中 $\sigma(x_j(t))$ 是参照个体 $x_j(t)$ 的模糊适应值宽度. 则 $x_i(t)$ 与 $x_j(t)$ 的表现型相似度为 $\mu_{ij}(x_{ir})$ 的平均值, 记为 $\mu(x_i(t), x_j(t))$, 有

$$\mu(x_i(t), x_j(t)) = \sum_{r=1}^{Ng} \frac{\mu_{ij}(x_{ir})}{Ng}. \quad (3)$$

表现型相似度 $\mu_{ij}(x_{ir})$ 的特点在于: 1) 个体属性间差异越小, $\mu(x_i(t), x_j(t))$ 越大, 个体表现型就越接近, 这符合事物相似性规律^[10]; 2) 如果参照个体 $x_j(t)$ 的模糊适应值宽度 $\sigma(x_j(t))$ 越小, 则 $\mu(x_i(t), x_j(t))$ 就越小, 这与一种评价现象相对应, 即: 将比较个体与参照个体粗略比较时 (此时评价准确性低, 参照个体 $x_j(t)$ 的模糊适应值宽度大), 两个个体相似度较大; 然而, 将比较个体与参照个体仔细比较时 (此时评价准确性高, 参照个体 $x_j(t)$ 的模糊适应值宽度小), 则两个个体相似度较小; 3) 更重要地, 当参照个体与比较个体互换后, 如果两个个体的模糊适应值宽度不相等, 即 $\sigma(x_j(t)) \neq \sigma(x_i(t))$, 则 $\mu(x_i(t), x_j(t)) \neq \mu(x_j(t), x_i(t))$. 所以, 式(3)刻画的个体表现型相似性特征更符合人的评价规律, 对于个体表现型的比较也更恰当.

1.2 分类策略

根据个体的相似度可将种群分成若干类. 设 $x(t)$ 可被分为 $N_c(t)$ 个类, 且 $N_c(t) \leq N_{\max c}(t)$, 其中 $N_{\max c}(t)$ 为第 t 代种群的最大分类数.

具体分类方法如下: 首先, 确定一个代表个体 $x_j(t)$; 然后, 从种群中搜索每一个个体 $x_i(t)$ ($i = 1, 2, \dots, N - 1$), 将所有满足 $\mu(x_i(t), x_j(t)) \geq 1 -$

$\min\{\mu_{ij}(x_{ir})|r = 1, 2, \dots, N_g\}$ 的个体作为第1类, 设为 $\{c_1(t)\}$, 即 $\{c_1(t)\} = \{x_i(t)|\mu(x_i(t), x_j(t)) \geq 1 - \min\{\mu_{ij}(x_{ir})|r = 1, 2, \dots, N_g\}, x_i(t) \in x(t)\}$. 若令 $c_1(t)$ 表示类 $\{c_1(t)\}$ 的代表个体, 则可知 $c_1(t) = x_i(t)$. 划分完第1类后, 缩小种群空间, 令 $x(t) \leftarrow x(t) \setminus \{c_1(t)\}$, 然后从种群中确定另一个个体作为第2类的代表个体. 重复以上操作, 直到获得第2类 $\{c_2(t)\}$. 类似地, 可以对种群 $x(t)$ 继续分类, 直到 $x(t)$ 为空, 或者 $N_c(t) = N_{\max c}(t)$. 当 $N_c(t) = N_{\max c}(t) - 1$ 时, 若 $x(t)$ 中还有未分类的个体, 则把 $x(t)$ 中的所有个体并为最后一类, 而不考虑它们是否相似. 完成上述分类过程后, 种群 $x(t)$ 被分为 $N_c(t)$ 个类, 分别为 $\{c_1(t)\}, \{c_2(t)\}, \dots, \{c_{N_c(t)}(t)\}$, 它们的代表个体分别为 $c_1(t), c_2(t), \dots, c_{N_c(t)}(t)$. 随着进化的进行, 个体间的相似度逐渐增大, 聚类中心数目逐渐减少.

受人机交互界面显示的限制, 需要设定最大分类数 $N_{\max c}(t)$, 且分类数应小于等于最大分类数. 从进化过程来看, 在进化初始阶段, 为了能更好地进行全局搜索, 希望初始种群多样性较好, 需要分较多的类, 则 $N_{\max c}(t)$ 应该较大; 而在进化后期希望算法能较好收敛, 应该分较少的类, 则 $N_{\max c}(t)$ 应该较小.

为了反映上述 $N_{\max c}(t)$ 的特点, 首先, 需引入种群相似度函数, 设第 t 代种群的相似度 $A(x(t))$ 为

$$A(x(t)) = \frac{2}{N(N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^N \mu(x_i(t), x_j(t)). \quad (4)$$

由式(4)可知: $A(x(t)) \in [0, 1]$, 并且 $x(t)$ 中的个体越相似, $A(x(t))$ 越接近于1; 相反地, $x(t)$ 中的个体越不相似, $A(x(t))$ 越接近于0. 若 $A(x(t)) = 1$, 则表示 $x(t)$ 中的所有个体都相同, 此时 $x(t)$ 只划分为一类; 若 $A(x(t)) = 0$, 则表示 $x(t)$ 中的个体都不相同, 此时 $x(t)$ 应尽可能被分为较多的类.

最大分类数 $N_{\max c}(t)$ 由下式计算:

$$N_{\max c}(t) = \lceil ((1 - N_{\max d}) \cdot A(x(t)) + N_{\max d})e^{-\frac{t}{T}} \rceil. \quad (5)$$

其中: $N_{\max d}$ 表示人机交互界面所能显示的最大个体数目, T 为设定的最大进化代数, $\lceil \cdot \rceil$ 为向上取整函数.

1.3 用户评价个体的选择

本小节提出选择种群内代表性个体给用户评价的新方法. 该方法首先计算当代进化种群的个体与上代相对最优个体(用户评价最高个体)的相似度^[11]; 然后, 依据相似度的数值大小, 选择最靠前的 N_c 个个体推荐给用户评价.

记第 $t-1$ 代用户评价最高个体为 $x_B(t-1)$, 则根据式(2), 当前代 t 的个体 $x_i(t)$ ($i = 1, 2, \dots, N$) 与

$x_B(t-1)$ 的相似度为

$$\text{Sim}(x_i(t), x_B(t-1)) = \sum_{r=1}^{N_g} \frac{\mu_{iB}(x_{ir})}{N_g}, \quad (6)$$

用户评价个体数目为

$$N_c = \text{Rank}_{N_c} \max_{i \in \{1, 2, \dots, N\}} \{\text{Sim}(x_i(t), x_B(t-1))\}. \quad (7)$$

该方法的特点在于: 依据与前代优秀个体的相似度排名, 利用历史信息为用户推荐个体. 对于选择种群内代表性个体更有针对性, 而且该方法仅用于选择用户评价的个体, 后续遗传操作仍由个体适应值决定, 因此不会错误引导进化方向.

1.4 用户评价个体模糊适应值的估计

记第 t 代种群的用户评价进化个体为 $c_i(t)$, $i = 1, 2, \dots, N_c$. 其适应值的论域为 $[f_{\min}, f_{\max}]$, $c_i(t)$ 的模糊适应值可表示为 $\tilde{f}(c_i(t))$. 在 $[f_{\min}, f_{\max}]$ 上定义如下泛函: $\mu_{\tilde{f}(c_i(t))} : [f_{\min}, f_{\max}] \rightarrow [0, 1]$, 表示 $\forall f \in [f_{\min}, f_{\max}]$ 属于 $\tilde{f}(c_i(t))$ 的程度, 其中 $\tilde{f}(c_i(t))$ 为模糊数. $\tilde{f}(c_i(t))$ 的隶属函数可表示为

$$\mu_{\tilde{f}(c_i(t))}(f) = e^{-\frac{1}{2} \left(\frac{f - d(c_i(t))}{\sigma(c_i(t))} \right)^2}. \quad (8)$$

其中: $d(c_i(t))$ 表示 $\tilde{f}(c_i(t))$ 的中心, 是隶属度为1的适应值; $\sigma(c_i(t))$ 表示 $\tilde{f}(c_i(t))$ 的宽度, 该量主要反映评价的不确定性, 且与不确定性呈正向变化^[12].

本文中 $\tilde{f}(c_i(t))$ 的中心值 $d(c_i(t))$ 由用户评价, 宽度值则由评价时间等信息估计, 具体如下.

根据文献[12]的结论, 人对个体评价时间越长, 对该个体的满意度就越大, 个体适应值也越大; 反之, 评价时间越短, 对该个体的满意度就越小, 个体适应值也越小. 另一方面, 根据认知常识, 人对于满意度越大和越低的个体评价较为准确, 评价不确定较小; 而对于满意度适中的个体, 则评价较为困难, 评价不确定性较大. 针对认知的这种特征, 设 $\sigma(c_i(t))$ 的取值范围为 $[\sigma_{\min}, \sigma_{\max}]$, $T(c_i(t))$ 为个体 $c_i(t)$ 被评价时间. 将 $T(c_i(t))$ 与当前进化代内个体最长评价时间的比值作为个体 $c_i(t)$ 的满意度, 记为 H_i , 有

$$H_i = \frac{T(c_i(t))}{\max(T(c_1(t)), T(c_2(t)), \dots, T(c_{N_c}(t)))}. \quad (9)$$

H_i 越大, $\tilde{f}(c_i(t))$ 的宽度越接近于 σ_{\min} ; H_i 越小, $\tilde{f}(c_i(t))$ 的宽度越接近于 σ_{\max} . 于是个体 $c_i(t)$ 的模糊适应值 $\tilde{f}(c_i(t))$ 的宽度 $\sigma(c_i(t))$ 可以表示为

$$\sigma(c_i(t)) = \frac{1}{(e^{d(c_i(t))})^{H_i}}. \quad (10)$$

与以往研究相比, 本文的模糊适应值估计方法的特点是: 只要求用户按单一数值评价个体, 评价结果作为模糊适应值中心值, 模糊适应值宽度则由评价时

间和中心值确定. 这样, 用户操作负担很少, 同时系统根据评价信息生成模糊适应值, 信息含量更加丰富; 另外, 因为式(10)是连续函数, 所以语气词为连续数值, 模糊适应值的偏好表达更为全面.

1.5 剩余个体模糊适应值的估计

设个体 $x_j(t)$ 属于类 $\{c_i(t)\}$, 记 $\hat{f}(x_j(t))$ 表示对个体 $x_j(t)$ 适应值 $f(x_j(t))$ 的估计值. 当用户评价完中心个体 $c_i(t)$ 并赋予它适应值后, 应充分利用这些信息来估计剩余个体的适应值, 从而扩大种群评价规模. 本小节基于个体相似度对剩余个体的模糊适应值进行估计. 记剩余个体的模糊适应值中心值为 $d(x_j(t))$, 宽度为 $\sigma(x_j(t))$, 对中心值和宽度分别估计后, 再构成模糊适应值.

假设 $\hat{f}(x_j(t))$ 为一个不精确数是合理的, 当 $x_j(t)$ 与 $c_i(t)$ 的相似程度越低时, 基于 $f(c_i(t))$ 的 $\hat{f}(x_j(t))$ 越不精确. 为了增加估计结果的准确度, 剩余个体适应值的估计不仅需要计算与所属类中心个体的相似性, 还应考虑与其他类中心个体的相似性.

剩余个体 $x_j(t)$ 的模糊适应值中心值 $d(x_j(t))$ 计算公式为

$$d(x_j(t)) = \begin{cases} d(c_K(t)), & \max_{i \in \{1, 2, \dots, N_c\}} \{\mu(x_j(t), c_i(t))\} > \theta(t); \\ \frac{\sum_{i=1}^{N_c} \mu(x_j(t), c_i(t)) \cdot d(c_i(t))}{\sum_{i=1}^{N_c} \mu(x_j(t), c_i(t))}, & \text{otherwise.} \end{cases} \quad (11)$$

其中: K 是个体表现型相似度 $\mu(x_j(t), c_i(t))$ 最大值大于阈值 $\theta(t)$ 时的聚类中心, 即

$$K = \text{index} \max_{i \in \{1, 2, \dots, N_c\}} \{\mu(x_j(t), c_i(t))\};$$

阈值 $\theta(t)$ 由前一代个体适应值中心值刻画, 即

$$\theta(t) = \alpha \cdot \frac{\max\{d(x_1(t-1)), \dots, d(x_N(t-1))\}}{\sum_{i=1}^N \frac{d(x_i(t-1))}{N}},$$

α 是反映评价性能的参数. 可以看出, 阈值 $\theta(t)$ 随进化代内最大适应值的增加而增加, 利用前一进化代的信息计算阈值, 可以为下一代估计 $d(x_j(t))$ 提供依据.

$d(x_j(t))$ 表明: 进化初期由于个体间相似性较小, 阈值 $\theta(t)$ 较小, 剩余个体适应值中心值 $d(x_j(t))$ 也较小; 随着算法逐渐收敛, 个体间相似性逐渐增大, 阈值 $\theta(t)$ 也逐渐提高, 剩余个体适应值中心值 $d(x_j(t))$ 逐渐增大. 通过对各聚类中心相似度的加权估计中心值, 可以提高信息利用率, 使估计结果更加精确.

相似地, 剩余个体 $x_j(t)$ 的模糊适应值宽度值 $\sigma(x_j(t))$ 计算公式为

$$\sigma(x_j(t)) = \frac{\sum_{i=1}^{N_c} \mu(x_j(t), c_i(t)) \cdot \sigma(c_i(t))}{\sum_{i=1}^{N_c} \mu(x_j(t), c_i(t))}. \quad (12)$$

上述方法实现了对所有个体的模糊适应值估计, 在大规模种群前提下, 偏好信息更加丰富, 更有利于进化优化.

1.6 个体选择

通常, 适应值越大, 个体越容易胜出; 个体相似性越强, 个体适应值越接近. 所以, 个体的选择不仅与适应值有关系, 与个体间的相似性也有关系. 模糊适应值的中心值主要反映适应值大小, 宽度值则反映评价的不确定性. 根据1.1节对个体相似性的分析, 宽度值也影响相似性的计算.

为了更客观地选择模糊适应值的个体, 本小节考虑规模为2的联赛选择. 设比较个体为 $x_a(t)$ 和 $x_b(t)$, 则个体选择适应值分别为 $F(x_a(t))$, $F(x_b(t))$, 即

$$F(x_a(t)) = d(x_a(t)) \cdot \mu(x_b(t), x_a(t)), \quad (13)$$

$$F(x_b(t)) = d(x_b(t)) \cdot \mu(x_a(t), x_b(t)). \quad (14)$$

其中: $d(x_a(t))$, $d(x_b(t))$ 分别是个体 $x_a(t)$, $x_b(t)$ 的模糊适应值中心值; $\mu(x_a(t), x_b(t))$, $\mu(x_b(t), x_a(t))$ 分别是以个体 $x_b(t)$ 和 $x_a(t)$ 为参照个体, 以 $x_a(t)$ 和 $x_b(t)$ 为比较个体的相似度.

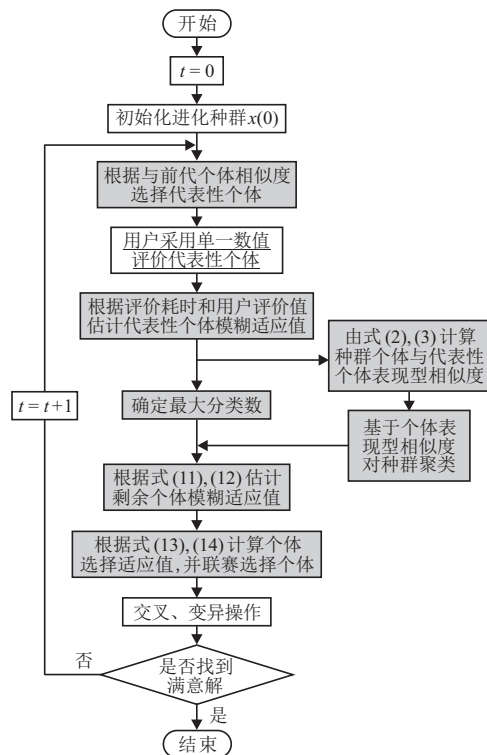


图1 算法框图

式(3)和(4)考虑了模糊适应值中心值和宽度对个体选择的共同影响. 其中,中心值对适应值的大小起主导作用,但由于认知的模糊性和不确定性,模糊适应值的比较不能完全由中心值决定,还要考虑宽度值的作用. 于是,个体选择条件为:当 $F(x_a(t)) > F(x_b(t))$ 时,选择个体 $x_a(t)$;当 $F(x_a(t)) < F(x_b(t))$ 时,选择个体 $x_b(t)$;当 $F(x_a(t)) = F(x_b(t))$ 时,随机选择 $x_a(t), x_b(t)$.

上述个体选择方法的特点在于:个体选择不直接采用模糊适应值,而是通过个体表现型相似性修正个体模糊适应值中心值以实现个体比较,不仅符合评价规律,而且实现简便,

综合上述,本文算法框图如图1所示. 其中:加下划线部分为用户参与遗传进化过程的交互环节;阴影部分为本文工作的创新之处.

2 在室内挂钟外观进化设计系统的应用

2.1 应用背景

由于室内挂钟外观设计元素丰富,艺术性较强,无法采用定量的目标函数衡量评价方案,而基于交互式遗传算法可以增强用户搜索能力,启发用户设计灵感,帮助用户找到满意的设计方案. 本文选择无钟摆室内挂钟二维外观设计为优化对象. 由于目前室内挂钟外观设计还没有使用交互式遗传算法的先例,本文同时为挂钟设计提出了一种新方法.

2.2 染色体编码

室内挂钟外观主要由钟壳、盘面、钟表数字和钟表指针4部分构成. 根据市场调查,钟壳形状主要有

长方形、方圆、八角和圆形等4种款式;盘面样式主要体现在刻度设计,可分为无刻度、点刻度和线刻度等3类共8种款式;钟表数字类型有无数字、阿拉伯数字和罗马数字等3类共8种款式;钟表表针款式比较丰富,由于秒针款式相对单一,这里只考虑时针和分针款式. 时针与分针款式依据古典和现代风格有简约、警棍、三文鱼、路易XVI和宝玑等5类共8种款式. 这样,室内挂钟样本个体由4种属性构成,即 $N_g = 4$,各属性按属性类型从0~7取值,具体如表1所示. 根据属性值,将每种属性转化为二进制编码(基因意义单元^[10]),其中,钟壳属性取2位二进制数,其余3种属性均取3位二进制数. 个体染色体共由11位二进制数组成,样本空间计为 $4 \times 8 \times 8 \times 8 = 2048$,二进制编码范围为0000000000~1111111111. 图2给出了染色体编码过程,其中,样本个体(0,5,2,2)款式为长方形钟壳/线刻度A盘面/阿拉伯数字A/警棍A表针. 染色体编码为00100001001.

表1 室内挂钟样本属性与属性值

属性 x_{ir}	属性 I_g^r			
	钟壳	盘面	数字	表针
0	直方	无刻度A	无数字A	简约A
1	方圆	无刻度B	无数字B	简约B
2	八角	点刻度A	阿拉伯A	警棍A
3	圆形	点刻度B	阿拉伯B	警棍B
4	—	点刻度C	阿拉伯C	三文鱼
5	—	线刻度A	罗马A	路易XVI
6	—	线刻度B	罗马B	宝玑A
7	—	线刻度C	罗马C	宝玑B

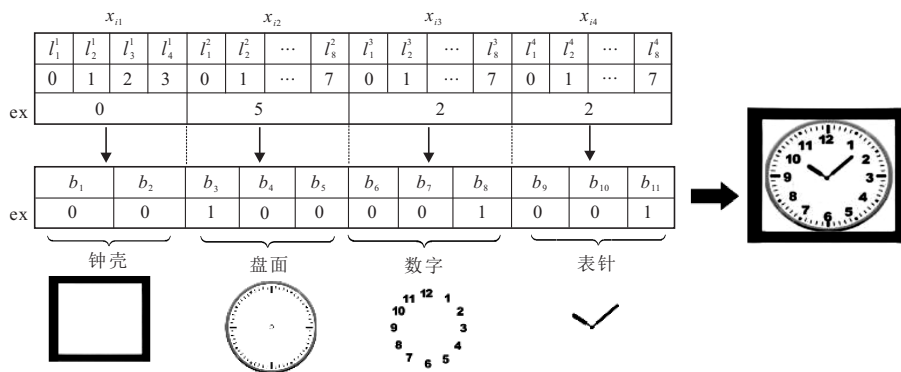


图2 染色体编码过程

2.3 交互界面设计

本系统采用 Visual Basic 6.0 实现,系统包括参数设置界面、交互式界面和结果输出界面等. 系统划分为载入模块、处理模块和交互式遗传算法模块等3个功能模块,3个模块通过一个交互界面融合为整体,共同完成交互式进化设计. 室内挂钟各属性样本

以. bmp 格式储存在数据库中,载入模块将挂钟属性模型文件读入内存并在窗口界面中显示;处理模块实现对模型的贴图,时针与分针统一固定在10:10;交互式遗传算法模块则将交互式遗传算法整合到系统中. 为保证系统整体的运行效率,采用按钮形式控制算法的执行,系统交互界面如图3所示. 系统运行时,

用户首先设置遗传参数,进入交互界面后点击“初始化”按钮,系统初始化并运行载入模块和处理模块,生成6个样本,即 $N_c = 6$. 用户通过样本下方滑动条对个体进行评价,同时,系统自动记录用户对每个个体的评价时间,通过样本下方滑动块显示,该评价时间与用户评价值在系统后台用于估计个体模糊适应值. 评价结束后,点击“下一代”按钮,系统运行交互式遗传算法模块,进化生成新种群. 用户继续评价个体,循环这一过程,直至用户找到满意个体,或达到最大终止代数,点击“结束”按钮输出优化结果. 为了让用户掌握进化节奏,交互界面还显示进化代数、评价总耗时等信息.



图3 系统交互界面

用户完成评价后,系统根据个体属性、评价时间和用户评价值按式(2)和(3)计算表现型相似度. 由于个体的模糊适应值宽度不完全相同,当参照个体不同时,个体的表现型相似度也不相同,这符合人的主观评价特点.

2.4 实验设计

本系统的测试实验对象是理性单用户,选择男女各10名在校大学生为测试用户. 以基于区间适应值的大种群规模交互式遗传算法(Interactive genetic

algorithms with large population size, IGA-LPS)和基于模糊适应值的交互式遗传算法(Interactive genetic algorithms with fuzzy individuals' fitness, IGA-FIF)^[9]为比较对象,验证所提出方法在减轻用户疲劳、搜索效率和优化质量等方面的有效性. 测试分为两个项目进行:以图2所示个体(0, 5, 2, 2)样本款式为优化目标的固定解优化测试(项目1)和以外观符合“典雅”标准的室内挂钟非固定解优化测试(项目2).

2.5 遗传操作及参数设置

分别基于IGA-LPS、IGA-FIF和本文方法,开发相应测试系统. 为了比较算法性能,3种算法设置相同的遗传操作:规模为2的联赛选择、单点交叉和单点变异. IGA-FIF种群规模 $N = 6$,模糊适应值语气算子取“大约”、“接近”和“很接近”等3种模糊语气词;本文算法和IGA-LPS中均采用大规模进化种群,受样本空间和交互界面显示限制,种群规模设为 $N = 40$, $N_c = N_{maxd} = 6$, $\alpha = 0.9$. 两种算法用户评价值下限和上限 f_{min} 和 f_{max} 的取值分别为0和100;3种算法交叉和变异概率均取 $p_c = 0.6$, $p_m = 0.01$,最大进化代数 $T = 20$. 采用3种终止进化条件:进化至最大进化代数 T 、系统收敛(相同个体占种群75%及以上)、用户对优化结果满意. 当任意终止条件满足时,系统结束进化,并输出进化优化结果.

2.6 实验结果与分析

固定解优化测试要求20名用户在最大进化代数内不限耗时,分别用上述3种算法各运行1次(正式运行前允许1次试运行,结果不计入统计),对目标个体总计60人次进化优化. 测试指标包括:进化代数、优化耗时、最优解适应值(中心值)和个体属性秩等.

表2给出了3种算法的进化代数、评价耗时和最优解适应值(中心值)的比较. 可以看出,在进化代

表2 算法进化代数、最优解适应值和优化耗时比较

算法指标	算法	项目1			NEH-IGA		
		均值	方差	t检验	均值	方差	t检验
进化代数	本文算法	15.38	4.47	0.0204	12.52	4.25	0.0069
	IGA-LPS	17.43	6.21		15.68	5.34	
	本文算法	15.38	4.47	0.0197	12.52	4.25	
	IGA-FIF	18.82	7.86		17.36	6.87	
优化耗时/s	本文算法	6'05"	2.36	0.0352	4'24"	1.20	0.0074
	IGA-LPS	8'52"	3.93		7'45"	2.75	
	本文算法	6'05"	2.36	0.0243	4'24"	1.20	
	IGA-FIF	11'23"	4.65		9'34"	3.43	
最优解适应值(中心值)	本文算法	87.26	15.64	0.0255	88.12	13.42	0.0039
	IGA-LPS	80.34	20.52		85.65	24.23	
	本文算法	87.26	15.64	0.0172	88.12	13.42	
	IGA-FIF	75.38	16.53		80.02	18.47	

数方面, 本文算法的所有用户均在最大进化代数内完成优化, 平均进化代数为 15.38, 是 3 种算法中最低的, IGA-FIF 进化代数最多; 进行平均值和方差双样本均值分析, t 检验差异显著 ($p < 0.05$). 在评价耗时方面, 本文算法每位用户平均耗时 6.05 min, 仍为 3 种算法中最低的, IGA-FIF 耗时最多, t 检验差异显著 ($p < 0.05$). 最优解适应值方面, 本文算法最高, 这反映出用户对本文算法优化结果最满意. 因此, 本文算法在 3 种算法的进化代数和评价耗时方面均为最少, 同时最优解适应值最高, 可以明显减轻用户疲劳.

算法优化质量主要通过计算用户优化结果与固定目标之间的属性标准差 ω_r 体现, 有

$$\omega_r = \sqrt{\frac{1}{M} \sum_{m=1}^M (\text{rank}x_{mr} - \text{rank}x_{Tr})^2}. \quad (15)$$

其中: M 是用户评价人次, $\text{rank}x_{mr}$ 是用户 m 的优化解 x_m 中属性 r 的属性秩, $\text{rank}x_{Tr}$ 是目标个体 x_T 中属性 $r(0, 5, 2, 2)$ 的属性秩. 式(5)的计算方法是: 将用户 m 的优化解 x_m 的属性值与固定目标属性值求差, 如果属性值相同, 则 $\text{rank}x_{mr} - \text{rank}x_{Tr} = 0$, 否则 $\text{rank}x_{mr} - \text{rank}x_{Tr} = 1$; 依次对其他属性求差值, 并求均方根, 结果即为所有用户评价人次 M 的优化解属性与固定目标属性的属性标准差. 可以看到, 优化结果属性与目标个体属性越接近, ω_r 越小. 若所有用户的优化解均与固定目标相同 (优化解各属性值均与固定目标属性值相同), 则 $\omega_r = 0$, 所以式(15)反映了用户选择结果与固定目标之间的差异. ω_r 统计结果如表 3 所示.

表 3 个体属性标准差

算法	属性			
	x_{i1}	x_{i2}	x_{i3}	x_{i4}
本文算法	0.11	0.97	0.14	0.87
IGA-LPS	0.25	1.16	0.23	1.08
IGA-FIF	0.32	1.24	0.31	1.19

表 3 中数据表现出 3 种算法的优化结果具有一致性, 即属性 x_{i1} (钟壳) 与 x_{i3} (数字) 属性标准差较小, 属性 x_{i2} (盘面) 与 x_{i4} (表针) 属性标准差较大, 其中属性 x_{i2} (盘面) 标准差最大. 因为本文算法的属性标准差最小, 所以优化质量最高.

项目 1 结束一周后, 进行非固定解优化测试. 要求 20 位用户根据“典雅”风格标准, 采用上述 3 种算法在最大进化代数内不限耗时各运行 1 次, 总计 60 人次进化优化. 该项目的测试指标包括: 进化代数、优化耗时和最优解 (算法收敛时用户满意的优化解) 等.

3 种算法的进化代数、最优解适应值 (中心值) 和优化耗时比较如表 2 所示. 从表 2 数据看, 本文算法

的所有用户平均适应值为 88.12, 是 3 种算法中最高的. 这种高评分低分布差异表明, 用户可以获得满意解, 并且设定 20 代为最大进化代数是足够的. 在进化代数和优化耗时方面, 本文算法都是最低的, 且 t 检验差异显著. 这表明, 在非固定解优化测试条件下, 本文算法仍可以显著降低用户操作负担, 减轻疲劳.

本项目中, 通过优化结果多变量分析比较算法优化质量.

首先, 建立 3 种算法优化结果属性矩阵 X , 该矩阵的行是用户数 $m(1 \sim 20$ 名用户), 列是个体属性 r (取 0~7 个属性值), 所以矩阵规模为 20×4 . 采用 HAC 树状图分析本文算法矩阵 X 如图 4 所示.

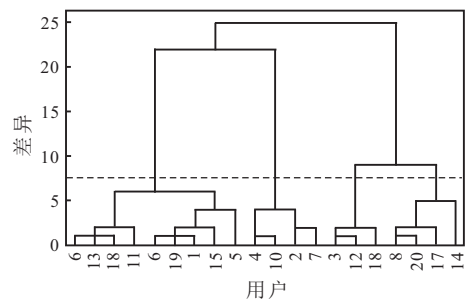


图 4 本文算法优化结果树状图

由图 4 可知, 根据 20 位用户对 4 个属性的评价结果, 将相似的偏好算作一组, 可以将用户优化结果归纳为 4 组: 第 1 组包含 9 名用户; 第 2 组包含 4 名用户; 第 3 组包含 3 名用户; 第 4 组包含 4 名用户. 特别地, 第 1 组中用户 6 和 13 优化结果相同.

为了定量分析样本形态的丰富性, 采用如下属性计算准则 ϕ_{kv}^r :

$$\phi_{kv}^r = \frac{m_{kv}^r / m1_k}{m2_v^r / m}. \quad (16)$$

其中: m_{kv}^r 是在 k 组中, 属性 r 表现的形态 v 的数目; $m1_k$ 是 k 组中的样本数目; $m2_v^r$ 是所有优化结果中, 属性 r 表现的形态 v 的数目; m 是优化结果数目中的样本数目 (图 4 中优化结果样本数目为 20). 式(16)表明, ϕ_{kv}^r 越大, 属性呈现的形态越多, 个体多样性越好.

表 4 统计了 $\phi_{kv}^r > 1.5$ 时, 3 种算法的优化结果属性分组情况. 与 HAC 分组结果相同, 本文算法每组呈现出一个代表性个体; IGA-LPS 的优化结果采用 HAC 树状图分析也分为 4 组, IGA-FIF 分为 3 组, 但 $\phi_{kv}^r > 1.5$ 时, IGA-LPS 只呈现 3 组代表个体; IGA-FIF 呈现 2 组代表个体. 这表明, 采用大规模种群进化, 优化结果可以获得更好的多样性, 而本文算法的多样性最为突出, 优化质量最高.

然后, 统计 2 个测试项目中 3 种算法的完全匹配解用户数目 (项目 1)、最优解数目 (项目 2)、算法效用和收敛率 (收敛率 = 算法收敛次数 / 总的实验次

表4 优化结果属性形态分组

算法	优化结果属性分组			
	组1	组2	组3	组4
本文算法	$x_{11}x_{24}x_{32}x_{42}$	$x_{11}x_{25}x_{32}x_{45}$	$x_{13}x_{22}x_{36}x_{40}$	$x_{13}x_{24}x_{31}x_{40}$
IGA-LPS	$x_{11}x_{25}x_{30}x_{43}$	$x_{10}x_{24}x_{31}x_{46}$	$x_{13}x_{21}x_{34}x_{41}$	—
IGA-FIF	$x_{10}x_{24}x_{32}x_{44}$	$x_{10}x_{25}x_{31}x_{47}$	—	—

表5 算法优化解数目和成功率

算法	项目1			项目2		
	完全匹配解的用户	算法效用Eff/%	收敛率/%	最优解数目	算法效用Eff/%	收敛率/%
本文算法	4	37.8	87.2	102	42.4	93.4
IGA-LPS	3	26.6	81.6	96	32.7	85.46
IGA-FIF	1	16.5	84.2	80	20.1	87.5

数). 其中算法效用函数Eff为

$$\text{Eff} = \frac{1}{T_e} \sum_{t=1}^{T_e} \frac{N_{\text{co}}^t}{N}. \quad (17)$$

其中: T_e 是运行终止进化代数, N_{co}^t 是当前代 t 时用户满意的个体数目, N 是种群规模. 统计结果如表5所示. 数据表明: 在项目1中, 优化结果完全符合固定目标的人数当属本文算法最多; 在项目2获得最优解数目上, 本文算法数量也最多, 说明该算法在该项的搜索效率最高. 在算法效用和收敛率方面, 两个测试项目中, 本文算法都是最高的, 这表明, 本文算法采用大规模种群进化与模糊适应值提高了算法搜索效率.

综合上述两项测试及实验结果分析, 与比较算法相比, 采用本文算法, 用户可以付出最少操作量获得最满意的优化解, 同时提高了优化质量, 减轻了用户疲劳.

3 结论

本文提出了大规模种群进化的个体模糊适应值估计策略, 同时提出了个体表现型相似度计算方法, 利用个体表现型相似度对种群聚类并估计个体模糊适应值, 基于个体选择适应值实现个体选择操作. 通过室内挂钟进化设计系统和新颖的实验分析方法验证了算法性能, 并与已有典型方法进行了比较. 结果表明, 所提出方法在减轻用户疲劳、提高优化质量和优化效率等方面均具有优越性. 开发更合理的个体适应值形式和种群聚类方法是下一步要研究的问题.

参考文献(References)

- [1] Janarthanam S, Sukumaran S. Interactive genetic algorithm with relevance feedback for content based image retrieval[J]. Int J of Current Research, 2016, 8(7): 34948-34954.
- [2] Manfré E A, Augello A, Pilato G, et al. Exploiting interactive genetic algorithms for creative humanoid dancing[J]. Biologically Inspired Cognitive

Architectures, 2016, 17(7): 12-21.

- [3] Maximos A, Kaliakatsos-Papakostas, Andreas Floros, et al. Interactive music composition driven by feature evolution[J]. Springer Plus, 2016, 5(1): 1-38.
- [4] Seiki Koma, Yuichiro Yamabe, Akinori Tan. Research on urban landscape design using the interactive genetic algorithm and 3D images[J]. Visualization in Engineering, 2017, 5(1): 1-10.
- [5] Gong D W, Qin N N, Sun X Y. Evolutionary algorithm with large population size for problems with hybrid indices[J]. Control Theory & Applications, 2010, 27(6): 769-774.
- [6] Garcia H L, Pierreval H, Salas M, et al. An interactive genetic algorithm with c -means clustering for the unequal area facility layout problem[C]. Proc of the 10th Int Conf on Intelligent Systems Design and Applications. Cairo: IEEE, 2010: 61-66.
- [7] Allysson Alex Araújo, Matheus Paixao, Italo Yeltsin, et al. An architecture based on interactive optimization and machine learning applied to the next release problem[J]. Automated Software Engineering, 2016, 24(3): 1-49.
- [8] Wang Y, Yang H Q, Zhang R J. Research on vehicle routing problem with time window based on strong gene schema combination algorithm[J]. Control and Decision, 2011, 26(4): 606-610.
- [9] Gong Dun-wei, Sun Xiao-yan, Yuan Jie. Interactive genetic algorithms with individual's uncertain fitness[J]. Computers in Human Behavior, 2009, 27(5): 1482-1492.
- [10] Gong D W, Hao G S, Zhou Y, et al. Theory and applications of interactive genetic algorithms[M]. Beijing: National Defense Industry Press, 2007: 44-102.
- [11] Gong D W, Chen J, Sun X Y. Novel interactive genetic algorithm for estimating individual fitness based on similarity[J]. Control Theory & Applications, 2013, 30(5): 558-566.
- [12] Hao G S. Theory and applications of interactive evolutionary computation[M]. Xuzhou: CUMT Press, 2016: 127-139.

(责任编辑: 李君玲)