

基于熵极大准则的非用户 赋适应值交互式遗传算法

郭广颂¹, 陈良骥²

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

摘要: 针对交互式遗传算法适应值人工赋值极易疲劳导致的算法进化代数不足、优化效率低下这一难题, 提出了适应值非用户赋值方法. 首先, 用户对个体采用二元评价机制评价个体, 将个体划分为满意集合和不满集合; 然后, 根据个体评价时间与偏好的内在联系, 通过个体评价时间确定评价满意度; 最后, 基于熵极大准则求解满意度最大条件下的个体适应值. 为了确保优势基因遗传, 加快算法收敛, 采取种群精英基因构建优势个体保留策略. 将该方法应用于装饰性壁纸选型系统中, 并与其他代表性算法比较. 结果表明, 该方法能有效降低疲劳, 提高算法优化效率.

关键词: 交互式遗传算法; 熵极大准则; 适应值; 非用户赋值

中图分类号: TP301 **文献标识码:** A **文章编号:** 0372-2112 (2017)09-2997-08

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2017.12.023

An Interactive Genetic Algorithms Based on Maximum Entropy Principle with Individuals' Fitness not Assigned by User

GUO Guang-song¹, CHEN Liang-ji²

(1. School of Mechatronics Engineering, Zhengzhou University of Aeronautics, Zhengzhou, Henan 450046, China;

2. School of Mechanical Engineering, Tianjin Polytechnic University, Tianjin 300387, China)

Abstract: The fitness assigned by user can easily make fatigue which causes insufficient evolution algebra and low optimization efficiency for interactive genetic algorithms. In this study, a method of interactive genetic algorithms with individuals' fitness not assigned by user is presented. First, the individuals could be divided into satisfied sets and not satisfied sets; then, the individuals satisfaction is determined through evaluation time; finally, the fitness is calculated based on the maximum entropy principle under the biggest satisfaction. In order to ensure protogene inheritance, the reserved elite individual is built by population elite genes. This method is applied to selection system of decorative wallpaper, and the results show that it can effectively reduce fatigue and improve the optimization efficiency.

Key words: interactive genetic algorithms; maximum entropy principle; fitness; not assigned by user

1 引言

交互式进化计算(Interactive Evolutionary Computation, IEC)思想一般认为起源于1986年 Dawkins 的著作《The Blind Watchmaker》, 书中采用传统递归算法创建了一个进化树系统, 该系统通过用户从父代进化树分支中选择满意的分支做为子代, 经过基因变异生成下一代, 最终创建出二维动画昆虫. 这种能反映用户偏好, 实现定性指标优化的方法从此引起关注, 借鉴传统遗传算法的交互式遗传算法(Interactive Genetic Algo-

rithms, IGA)成为 IEC 的重要表现形式, 经过 20 余年的发展, 交互式遗传算法现已应用于人工耳蜗^[1]、芳香剂合成^[2]、音乐创作^[3]等众多领域.

由于交互式遗传算法的每款产品适应值由人评价, 容易产生疲劳, 这会限制算法的种群规模和进化代数. 对于这一突出特点, 提升适应值评价质量、降低用户疲劳成为近年来交互式遗传算法人-机交互能力研究的热点. 对于最广泛的视觉效果优化对象, 因为比较容易观察, 所以适应值评价一般采取数值形式. 但传统的适应值精确数赋值方法对偏好信息含量不足, 且操作

负担较大,容易疲劳.为了降低疲劳,提高优化质量,需要引入合适的适应值估计策略或对适应值修正.基于此,本文提出基于熵极大化准则的非用户赋适应值交互式遗传算法(Interactive Genetic Algorithms based on Maximum entropy Principle with individuals' fitness not assigned by User,IGA-MEP).该方法首先获得用户的二元评价结果和评价时间,然后根据二元评价结果计算个体的满意度,通过熵极大准则按满意度和评价时间估计个体适应值,最后,再构建保留精英基因个体,将优势基因遗传至下一代.该方法的优点在于:(1)用户进行二元选择操作负担极小,且对用户评价顺序没有要求,符合用户评价习惯,减轻了操作疲劳;(2)适应值赋值估计合理,适用于视觉效果优化对象,人-机交互能力较强;(3)构建保留精英基因个体比直接保留适应值最高个体更合理,保存的优势基因更完善.本文的贡献主要表现在以下4个方面:(1)提供用户最小需求操作界面;(2)提出了熵极大准则下从时间空间到适应值空间映射的适应值计算公式;(3)提出构建保留精英基因个体策略;(4)将所提方法用于装饰性墙壁纸选型系统,通过实验分析本文方法的有效性.

2 相关工作

小规模种群的交互式遗传算法适应值赋值研究大致分为基于评价方式的个体赋值研究和基于生理特征的个体赋值研究两个方向.前者的研究思路是通过开发合理描述偏好特征的适应值表现形式,获得更准确的评价值,进而提高进化优化质量.针对易观察的视觉优化对象,按评价方式可将适应值赋值形式分为精确赋值方法、离散赋值方法以及区间赋值方法^[4]、模糊赋值方法等^[5];而针对助听器设计与音乐合成等非视觉优化对象,Ishibuchi等提出用户适应值最低需求评价方法,用户只需对当前个体做出比前一个体是好还是不好的二元选择即可,算法按更新规则选择进化个体^[6].后者则根据人在评价过程中的生理特征变化确定个体适应值,其研究思路是通过捕捉评价过程中人的生理特征,用机器完成适应值赋值,达到减轻疲劳的目的.代表性的研究有Pallez提出的Eye-track赋值方法^[7],该方法通过电子设备“Eye-Tracker”测量评价过程中人的眼球活动情况,估计目标个体适应值;Holmes通过图形匹配实验测试了Eye-track赋值方法的有效性^[8].Fukumoto和Inoue提出根据评价过程中人的心率变异性(心跳间隔)完成适应值赋值的方法^[9,10],并将该方法用于音乐创作.由于客观的生理特征与主观的大脑决策并非完全一致,这类基于生理特征估计出来的适应值偏差较大,不仅开发成本较高,实验对象也不具有普适性,有时甚至还会增加疲劳.

3 提出的方法

本算法的适用对象是以交互式界面为人-机对话窗口的视觉效果设计.

一般来说,人对不同的物品敏感程度不同,对自己满意或者不满意的物品比较敏感,也就是说人能比较容易识别自己喜欢和不喜欢的那些物品.所以从评价方式看,将评价结果归属于满意集合和不满意集合可以对满意度进行宽泛而客观的评价.从生理特征看,视觉效果设计中反映用户偏好的行为指标主要有以下几个方面^[11]:(1)反复进行或持续的行为.即用户在某个页面上的停留时间、鼠标的移动、点击和拖动频次等都是反映用户偏好的重要指标.(2)标记行为.即用户对某个感兴趣主题的收藏、保存和打印页面等动作.(3)操作行为.即用户对当前页面感兴趣主题的拷贝,或在新页面内的搜寻行为.(4)浏览路径.即用户每次打开页面时因对某类或某种资源的偏好而采取的固定浏览路径.考虑到交互式遗传算法场景的客观情况,为了尽可能减少用户疲劳,用户对某个样本的关注时间无疑是衡量偏好的最主要指标,因为过多引入其他行为等同于增加操作负担.相关研究也显示,交互式遗传算法中用户对交互界面中个体的关注时间主要由反映偏好的眼部活动时间构成.并且评价时间越长,眼部活动越频繁,用户对该个体的关注度就越高^[7,8].所以,对于用户越偏爱的个体,用户识别所用的时间越长,个体的适应值应该越大;同样的,对于用户越不满意的个体,用户识别所用的时间越短,个体适应值应该越小.所以,人对优化对象的关注时间直接反映了人的满意程度.

基于此,通过记录用户评价个体满意或者不满意的时间,可以获得从时间空间到适应值空间的映射,由此可以获得进化个体适应值而无需用户直接赋适应值,从而达到减轻用户疲劳的目的.Gong等根据上述评价时间与适应值的关系提出了一种非用户赋适应值方法^[12],但由于受适应值计算要求,该方法需要用户按适当顺序评价个体,否则算法将不能获得正确的优化结果.针对这一问题,本文设计新的适应值计算方法,该算法在保证进化优化效果的前提下,对进化个体采用满意和不满意的二选一评价方式实现非用户赋适应值,并且本文方法对用户评价顺序没有要求.

3.1 基于熵极大化准则的适应值分析

熵极大准则是熵估计法中的一个核心内容.熵极大准则的基本思路是对某一系统宏观观察的基础上推断出系统内部的微观状态的概率分布,它的本质是指对某个随机分布一无所知时(除非负性和正则性外),选择均匀分布是最合理的,而均匀分布对应着最大熵分布.极大熵准则求解问题一般可描述为在某个

给定的分布集合的闭包上求出熵最大的分布,当是一凸集时,其解存在且惟一,即

$$\begin{aligned} \max H(x) &= - \int_{x \in \Theta} f(x) \ln f(x) dx \\ \text{s. t. } &\begin{cases} \int_{x \in \Theta} f(x) dx = 1 \\ \int_{x \in \Theta} f(x) g_i(x) dx = E_i, i = 1, 2, \dots, n \end{cases} \end{aligned} \quad (1)$$

其中: Θ 是 x 的论域; $f(x)$ 是随机变量 x 的概率密度函数; $g_i(x)$ 是已知函数,表示已知的信息^[13].

定理 1 若随机变量 x 为有限区间 $[a, b]$ 上的函数,设其概率函数为 $f(x)$,则 $f(x) = \frac{1}{b-a}$ 使 $H(x)$ 达到最大值.

定理 1 表明,确定在 $[a, b]$ 上的一切分布中, $[a, b]$ 上的均匀分布为其最大熵分布,即如果仅知道 x 可能取值为区间 $[a, b]$ 时,那么推断 x 的未知分布为 $[a, b]$ 上的均匀分布是最合理的.

交互式遗传算法的适应值评价过程是依据人的最大满意程度给出的,由于个人偏好不同,所以同一个体的适应值也会不同,即个体适应值存在不确定性.但适应值的取值范围事先往往是已知的,若适应值 x 取值范围为 $[f_{\min}, f_{\max}]$,则表明适应值 x 落在 $[f_{\min}, f_{\max}]$ 内的概率只与取值范围的区间长度有关,而与区间位置无关.因此对于事先确定的取值范围 $[f_{\min}, f_{\max}]$,种群中各个体适应值落在 $[f_{\min}, f_{\max}]$ 内的可能性是相等的.结合熵极大准则原理不难看出,适应值其实是在分布集合闭包内以人的最大偏好满意度为熵的分布.根据定理 1,适应值 x 在 $[f_{\min}, f_{\max}]$ 上以均匀分布最合理,即概率函数为 $f(x) = \frac{1}{f_{\max} - f_{\min}}$.

3.2 适应值非用户赋值策略

不失一般性,考虑优化问题: $\max f(x, t) = f(x(t))$, $x \in S \subseteq R^d$, x 是 d 维决策变量, S 是其取值范围, $f(x, t)$ 是被优化的性能指标.设 $x(t)$ 为第 t 代种群, $x_i(t)$ 为种群中某一个体, $i = 1, 2, \dots, N$,其中 N 为种群规模. $x_i(t)$ 的适应值为 $f(x_i(t))$, $f(x_i(t))$ 的取值范围为 $[f_{\min}, f_{\max}]$.从评价方式看,根据偏好规律可将评价结果划分为“满意”与“不满意”两大类,即设集合 $S_s(t)$ 和 $S_u(t)$ 分别表示第 t 代的满意集和不满意集,并且 $S_s(t) \cup S_u(t) = x(t)$.每代新种群提交给用户后,用户根据其偏好,给出“满意”或者“不满意”的评价,同时,个体被进化系统放入相应集合. $S_s(t)$ 中个体适应值分布区间为 $[f_s, \bar{f}_s]$, $S_u(t)$ 中个体适应值分布区间为 $[f_u, \bar{f}_u]$,且 $[f_s, \bar{f}_s], [f_u, \bar{f}_u] \subset [1, +\infty)$.根据 3.1 节,适应值在 $[f_s, \bar{f}_s]$ 和 $[f_u, \bar{f}_u]$ 内均按均匀分布,其中 $\bar{f}_s = f_{\max}$, $f_u = f_{\min}$, $f_s = \bar{f}_u$,且:

$$f_s = (f_{\max} - f_{\min}) \times \frac{\text{num}(x_k(t))}{N} \quad (2)$$

上式说明,满意集与不满意集的适应值分界由满意集中所占个体数决定.当种群中的个体属于满意集较多时,表明当前种群个体评价价值普遍较高,进化深度较大,所以满意集内个体的评价下限的“起点”也应较高;反之,满意集内个体适应值下限应降低.

设个体 $x_i(t)$ 被放入 $S_s(t)$ 或 $S_u(t)$ 的时间为 $T(x_i(t))$,即 $T(x_i(t))$ 为个体 $x_i(t)$ 被评价的时间.根据前述评价时间与满意程度的关系,可将 $T(x_i(t))$ 与集合中最长个体评价时间的比值作为个体 $x_i(t)$ 的满意度 H_i :

$$H_i = \begin{cases} \frac{T(x_i(t))}{\max_{x_i \in S_s(t)} (T(x_k(t)))}, & f(x_i(t)) \in S_s(t) \\ \frac{T(x_i(t))}{\max_{x_i \in S_u(t)} (T(x_k(t)))}, & f(x_i(t)) \in S_u(t) \end{cases} \quad (3)$$

上式说明,通过对个体的评价时间可以刻画出用户对该个体的满意度,相比于评价耗时最多的个体,评价时间越长,对该个体的满意度越大;反之,满意度就越小,且 $H_i \in [0, 1]$.根据 3.1 节熵极大准则原理,个体 $x_i(t)$ 适应值计算公式如下:

$$\begin{cases} \int_{f(x_i(t))}^{\bar{f}_s} \frac{1}{\bar{f}_s - f_s} \cdot \left(-\ln \frac{1}{\bar{f}_s - f_s}\right) dT(x_i(t)) = H_i \\ f(x_i(t)) \in S_s(t) \\ \int_{f_u}^{f(x_i(t))} \frac{1}{f_u - f_u} \cdot \left(-\ln \frac{1}{f_u - f_u}\right) dT(x_i(t)) = H_i \\ f(x_i(t)) \in S_u(t) \end{cases} \quad (4)$$

求解上式可得:

$$f(x_i(t)) = \begin{cases} \bar{f}_s - \frac{H_i(\bar{f}_s - f_s)}{\varepsilon - \ln \frac{1}{\bar{f}_s - f_s}}, & f(x_i(t)) \in S_s(t) \\ f_u + \frac{H_i(\bar{f}_u - f_u)}{\varepsilon - \ln \frac{1}{\bar{f}_u - f_u}}, & f(x_i(t)) \in S_u(t) \end{cases} \quad (5)$$

式中 ε 是防止分母为 0 预先设置的小的正常数. $f(x_i(t))$ 由两部分组成,其中分式部分始终大于 0,且集中体现了满意度对适应值大小的影响,当个体 $x_i(t)$ 被划为满意集或不满意集时,适应值由满意集适应值上限或不满意集适应值下限与分式之和决定.式(5)的物理意义是个体 $x_i(t)$ 的适应值在满意集中的概率分布密度为 $\frac{1}{\bar{f}_s - f_s}$ (在不满意集中为 $\frac{1}{\bar{f}_u - f_u}$),在时间 $T(x_i(t))$ 内使信息熵(满意度 H_i)达到最大值时的适应值即为该

个体适应值. 从式(5)还可以看出, 适应值计算仅由评价时间决定, 与评价顺序无关, 这是与文献[12]的主要区别. 对于满意集内的个体, 用户评价时间越短, 个体适应值就越大; 反之, 适应值应越小. 对于不满意集内的个体, 用户评价时间越短, 个体适应值就越小; 反之, 适应值应越大.

3.3 精英基因个体构建与选择策略

针对确定当前种群最优个体的研究, 探索二进制编码串中的基因位对适应值的影响是进化知识提取的一种新方法^[14]. 文献[15]对用户已评价的历代最优个体进行基因意义单元求同运算, 确定满意等位基因意义单元及其适应值; 文献[16]根据个体基因意义单元的熵确定个体精英集选择的相似度.

依据上述思想, 本文提出一种简便新颖的最优个体确定方法. 容易看出, 当前种群中某等位基因意义单元出现越多, 表明这种等位基因意义单元受重视程度越高, 亦即人对它的偏好越大, 反之亦然. 所以, 等位基因意义单元出现的频次, 可以作为偏好程度的依据. 若个体 $x_j(t)$ 由二进制码构成, 则 $x_j(t) \in \{0, 1\}^l$, l 为编码长度. 代表 $x_j(t)$ 的第 i 个属性的基因意义单元记为 $U_i = \{0, 1\}^{l_i}$, $i = 1, 2, \dots, g$, l_i 为该基因意义单元长度. 记第 t 代 U_i 的第 j_k 个等位基因意义单元为 $U_i^{j_k}$, $i \in \{1, 2, \dots, g\}$, $j_k \in \{1, 2, \dots, 2^{l_i}\}$. 令 $\text{CO}(U_i^{j_k})$ 为基因意义单元元素 $U_i^{j_k}$ 在当前种群中出现的次数,

$$\{U_{ib}^{j_k} | \text{CO}(U_{ib}^{j_k}) = \max(\text{CO}(U_i^{j_1}), \text{CO}(U_i^{j_2}), \dots, \text{CO}(U_i^{j_{2^{l_i}}}))\} \quad (6)$$

$U_{ib}^{j_k}$ 为第 t 代的精英基因个体 x_b^t 的第 i 个精英等位基因意义单元. 根据最优个体保留策略, x_b^t 的基因型为 $x_b^t = U_{1b}^{j_1} U_{2b}^{j_2} \dots U_{gb}^{j_g}$. 精英基因个体的意义在于根据进化代内个体等位基因意义单元出现的最高频次, 构建最优进化个体, 确保优势基因遗传. 保留精英基因个体后, 采用轮赌法生成新种群.

3.4 算法步骤

本文算法具体步骤如下:

step1 设置控制参数, 初始循环代数 $t = 0$, 初始化种群, 置满意集 $S_s(t)$ 和不满意集 $S_u(t)$ 为空;

step2 用户按“满意”或“不满意”选择评价个体;

step3 记录个体评价用时;

step4 按式(2)计算满意集和不满意集的分界值 f_s ;

step5 按式(3)计算个体 $x_i(t)$ 的评价满意度;

step6 按式(4)、(5)计算个体 $x_i(t)$ 的适应值;

step7 判断终止条件, 若满足, 则转 step 9;

step8 按式(6)对种群个体提取最优个体 x_b^t 保留至下一代, 对种群内其余个体进行遗传操作, 生成新种群,

置进化代数 $t = t + 1$, 转 step 2;

step9 输出优化解, 算法终止.

4 在墙壁纸选型系统中的应用

4.1 应用背景

装饰性墙壁纸是应用广泛的室内装饰材料, 通过恰当的色彩运用, 选择合理的墙壁纸能烘托出不同的氛围. 由于装饰性墙壁纸色彩丰富, 花系繁多, 艺术性较强, 优化选型难度较大, 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帮助用户在海量壁纸中找到满意的壁纸图案.

4.2 染色体编码

墙壁纸的外观主要由颜色和花系风格构成, 所以编码由两部分组成: 第一部分是墙壁纸颜色编码, 壁纸颜色通过 RGB 取色得到的颜色叠加构成, RGB 颜色中每种颜色分量的十进制取值范围为 $0 \sim 255$, 用 8 位二进制数表示, 其中前 8 位表示红色, 中间 8 位表示绿色, 最后 8 位表示蓝色, 每个颜色分量对应的二进制编码范围为 $00000000 \sim 11111111$, 颜色总编码为 24 位二进制数; 第二部分是墙壁纸花系风格编码, 墙壁纸花系风格大致分为欧美式、日式、中式等 16 大系列, 各花系风格系列墙壁纸图案在 $[0, 1023]$ 整数段内分段取整数值, 然后将整数值转化为 10 位二进制编码, 表 1 列出了墙壁纸图案的 16 个花系风格系列编码段. 这样组合构成的个体染色体长度为 34, 颜色基因意义单元长度为 8, 花系基因意义单元长度为 10, 总的搜索空间包含 $2^8 \times 2^8 \times 2^{10}$ 个候选解, 墙壁纸图案样本以 .bmp 格式储存在根目录下的数据库中. 由于颜色采用 RGB 颜色标准, 色彩选择已转换为显式性能指标; 但壁纸的花系风格选择则是隐式性能指标, 所以装饰性墙壁纸的选型是混合性能指标优化问题, 这是一类对算法性能要求较高的优化对象.

表 1 墙壁纸风格系列编码段

编码段	系列	编码段	系列
000 ~ 063	美式	512 ~ 575	传统
64 ~ 127	欧式	576 ~ 639	新艺术
128 ~ 191	日式	640 ~ 703	手绘
192 ~ 255	东南亚	704 ~ 767	碎花
256 ~ 319	现代约简	768 ~ 831	中式
320 ~ 383	现代时尚	832 ~ 895	地中海
384 ~ 447	田园	896 ~ 959	卡通
448 ~ 511	新古典	960 ~ 1023	其他

4.3 交互界面设计

本系统采用 Visual Basic 6.0 实现, 包括参数设置界面、交互式界面和结果输出界面等, 各任务流程采用

翻页式结构,尽可能减轻用户短期记忆负担.设置客厅、餐厅、卧室、书房、儿童房和过道等六个按钮供用户选择.实时交互时,系统首先进行初始化,设置进化优化的进化代数为0.系统随机选出6个样本作为初始样本,即 $N=6$,在每个样本下方有“满意”和“不满意”的单选按钮,用户通过选择单选按钮对个体评价.评价结束后,点击“下一代”按钮,系统进化出新种群,用户继续评价直至找到满意个体.同时交互界面还显示进化代数、评价互异个体数、优化耗时等辅助信息.系统交互界面如图1所示.



图1 系统交互界面

4.4 实验设计与分析

本系统的测试实验对象是理性的单用户.为了测试本文算法的性能,实验以传统交互式遗传算法(Traditional Interactive Genetic Algorithm, T-IGA)和离散适应值交互式遗传算法(Interactive Genetic Algorithm with Discrete Fitness, IGA-DF)为比较对象.本文算法与 T-IGA 适应值范围均为 1~100, IGA-DF 适应值按 1~100 分为 5 级;算法交叉概率与变异概率均为 $p_c=0.7$, $p_m=0.1$, ε 取 0.001.为了减少主观因素的影响,采用“粉色现代约简:染色体(252, 196, 207, 320)”风格壁纸做为优选目标.当最优解(最优解为算法收敛时适应值最高个体)数目在每代中占到一半以上时,可以认为优化进入收敛;若优化过程中算法进化发散,则无论进化代数多少均认为进化失败.

测试分为用户疲劳测试(项目1)和系统优化效率测试(项目2)两个项目进行.疲劳测试项目中,选择10位用户,不受时间与进化代数的限制,分别用上述3种算法各运行3次,总计90人次进化优化.用户疲劳指标包括进化代数、最优解适应值和算法优化耗时等.由于

项目1的进化为完全进化,对统计结果 t 检验,取显著水平 $\alpha=0.05$,做双样本均值分析,分析结果如表2所示.

测试指标中,进化代数的差异主要由算法性能决定,从数值上看,本文算法进化代数比 T-IGA 少近 1/3,基本在 10 个进化代内即可完成优化,明显减轻了用户疲劳. IGA-DF 的进化代数较多主要因为过于粗糙的进化个体适应值阻碍种群的正常进化,导致种群收敛速度过慢.本文算法的适应值赋值方法符合用户认知,通过构建保留精英基因个体,算法收敛性得到明显提高.

从最优解适应值指标看,由于本文算法的适应值最符合用户认知规律,所以在 3 种算法中本文算法的最优解适应值最高, t 检验结果差异显著($p<0.05$).

算法优化耗时差异主要由适应值赋值方式不同造成.操作界面越简单,用户耗时越少,反之用户耗时会增加.从表3可以看出:本文算法平均耗时 264s,比 T-IGA 要少近一半;比 IGA-DF 少近 1/3, t 检验结果差异显著($p<0.05$).本文算法由于只需用户最少需求操作,且对评价顺序无要求,用户负担最少.由于进化耗时最少,所需的进化代数也最少,所以平均每代进化的耗时也最少,用户最不易疲劳.

疲劳测试结束一周后,进行效率项目测试.效率测试要求相同的 10 位用户采用上述 3 种算法进行优化,每种算法运行 3 次,总计 90 人次进化优化.每次限定进化代数为 10 代,超过 10 代仍未完成优化,则视为进化失败.算法效率指标包括优化个体的秩、算法搜索效率、优化耗时和算法成功率等.

首先将按最优解适应值高低排序,选出前 30 个做为最佳个体,其中 15 个来自 IGA-MEP, 10 个来自 IGA-DF, 5 个来自 T-IGA.个体秩统计如表 3 所示.

然后,统计每种算法的搜索效率(搜索效率=搜索互异个体数/总搜索个体数,其中总搜索个体数是种群规模 N 与进化代数 t 之积),搜索效率曲线如图 2 所示.再统计每种算法的优化耗时,结果如图 3 所示.

鉴于在进化代数限定下,数据不完全符合正态分布,所以对搜索效率和优化耗时采用样本符号检验和 Wilcoxon 符号秩检验,取显著性水平为 0.05.表 4 给出了渐进显著性对比结果.

为了分析满意度对适应值估计的影响,按最优解适应值高低选取项目 1 和项目 2 中 IGA-MEP 优化的前 15 个最佳个体,记录前 10 代的满意度平均值和个体适应值平均值,统计结果如图 4 和图 5 所示.

最后,统计算法的最优解数目、优化成功率(成功率=完全匹配成功的进化次数/总的实验次数)和收敛率(收敛率=算法收敛次数/总的实验次数),统计分析结果如表 5 所示.

表 2 算法性能指标分析(项目 1)

算法指标		对比差异						
		平均值	方差	95% 置信区间		p 值(双尾)	自由度	显著性(双侧)
				下限	上限			
进化代数	IGA-MEP	9.70	0.747	9.30	10.10	5.21E-08	29	显著
	T-IGA	13.1	2.10	12.42	13.78			
	IGA-MEP	9.70	0.747	9.30	10.10	2.73E-07	29	显著
	IGA-DF	12.35	1.71	11.74	12.96			
最优解适应值	IGA-MEP	87.70	20.962	85.56	89.84	2.54E-05	29	显著
	T-IGA	76.55	41.629	73.53	79.57			
	IGA-MEP	87.70	20.962	85.56	89.84	4.55E-05	29	显著
	IGA-DF	72.50	93.421	67.98	77.02			
优化耗时/s	IGA-MEP	264.00	361.684	255.10	272.90	3.58E-18	29	显著
	T-IGA	407.15	108.450	402.28	412.02			
	IGA-MEP	264.00	361.684	255.10	272.90	2.5E-10	29	显著
	IGA-DF	325.70	211.905	318.89	332.51			

表 3 最佳个体的秩统计

算法	正秩数	平均秩	总秩数
IGA-MEP	15 ^a	9.6	144
IGA-DF	10 ^b	10.5	105
T-IGA	5 ^c	17	85
关系	0 ^d	-	-
总数	30	37.1	334

(a) IGA-MEP > IGA-DF > T-IGA (b) IGA-DF > IGA-MEP > T-IGA

(c) T-IGA > IGA-MEP > IGA-DF (d) IGA-MEP = IGA-DF = T-IGA

表 4 不同指标的渐进显著性对比

算法配对	搜索效率		优化耗时	
	0 ^a	0 ^b	0 ^a	0 ^b
IGA-MEP vs. T-IGA	0 ^a	0 ^b	0 ^a	0 ^b
IGA-MEP vs. IGA-DF	0 ^a	0 ^b	0 ^a	0 ^b

a: 符号检验; b: Wilcoxon 匹配对符号秩(2-side 检验)

表 5 算法最优解数目、成功率和收敛率

算法	项目 1			项目 2		
	最优解数目	优化成功率	收敛率	最优解数目	优化成功率	收敛率
IGA-DF	18	9.26%	66.3%	13	5.17%	59.8%
IGA-MEP	16	21.53%	76.5%	12	18.32%	70.6%
T-IGA	20	9.42%	70.4%	15	5.39%	62.6%

由图 2、3、4、5 和表 3、4、5 可知: 本文算法的正秩数最大, 最佳个体生成数目最多, 优化质量最好. 主要原因在于本文算法从适应值赋值策略和个体选择策略两个方面加强了算法的优化能力, 在保证准确评价个体

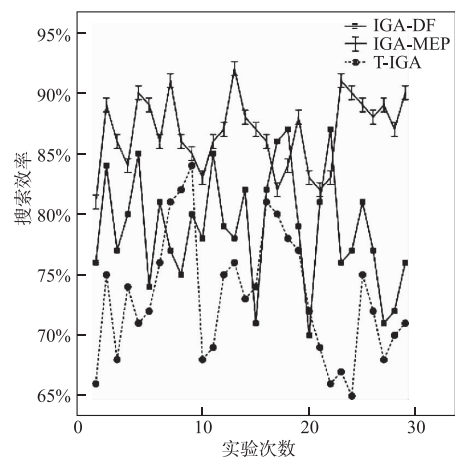


图2 搜索效率

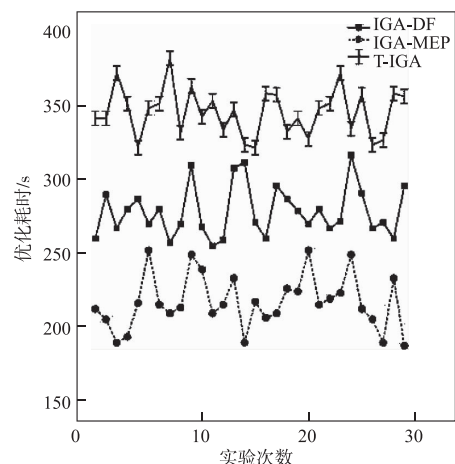


图3 优化耗时(项目 2)

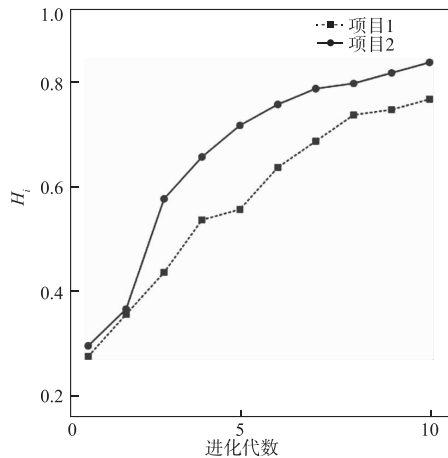


图4 满意度变化曲线

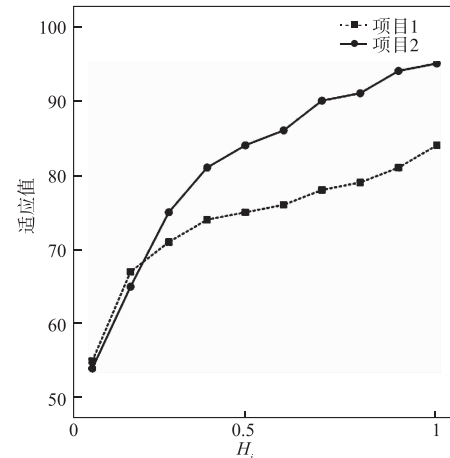


图5 满意度与适应值关系曲线

的同时,让子代获得更好的遗传性状,从而获得更好的进化结果。

在算法搜索效率方面,本文算法搜索效率最高,符号秩检验结果显著。主要原因在于本文方法适应值赋值精确,确保了优化方向明确。IGA-DF 的适应值取值较为粗糙,降低了种群的多样性。T-IGA 由于易疲劳,算法容易陷入过早收敛,搜索效率较低。

从优化耗时上看,3种算法的优化耗时曲线分界很明显,本文算法耗时明显比其他2种算法少,采用分级评价的IGA-DF则比T-IGA耗时少。主要原因除与收敛性相关外,合理易操作的适应值赋值方式会明显降低优化耗时。IGA-DF的分级评价比T-IGA精确数评价方便,本文算法的二选一策略比IGA-DF的5级评价负担更少,所以优化耗时也最少。

从满意度上看,两个测试项目下满意度均随进化代数增加而增大,适应值则随满意度增加而增加,这与本文的适应值估计原则一致。两个项目下,适应值均按式(5)与满意度呈同向变化,图5显示了这一特性。

在算法成功率和收敛率方面,两个测试项目中,本文算法都是最高的,但3种算法在最大进化代数限定内找到完全匹配个体的成功率和收敛率都不同程度下降。从最优解数目看,项目1中本文算法最优解数目最少,与目标匹配的样本比例最大,所以本文算法最优解质量最好。在项目2中,受进化不完全限制,3种算法最优解数目均有下降,本文算法优化成功率仍为最高。由于本文算法采用精英基因个体保留策略,适应值在解空间获得的信息较多,克服了传统交互式遗传算法的早熟收敛,在两个测试项目下算法的收敛性都是最好的,同时还获得了最好的最优解和优化成功率。

通过上述实验结果与分析,本文算法能够在提供用户最少操作负担的前提下,搜索到最多的互异个体,同时在最少的进化代数内获得最多的最佳个体。

5 结论

本文提出一种基于熵极大准则的非用户赋适应值交互式遗传算法。将该方法应用于装饰性墙壁纸优化选型系统,并与其他相关方法比较。实验结果表明,所提方法能有效降低用户疲劳,明显提高优化性能。

在不增加用户负担前提下,结合更多反映偏好的行为指标提高适应值准确度,进一步提高交互式遗传算法性能是今后要研究的问题。

参考文献

- [1] République C B, Dipanda A. Automatic and adaptive fitting of the cochlear implant by using interactive evolutionary algorithms [J]. *Journal of Multimedia Processing and Technologies*, 2011, 2(1): 42 - 54.
- [2] Fukumoto M, Koga S. A proposal for user's intervention in interactive evolutionary computation for optimizing fragrance composition [A]. *HCI International 2014* [C]. Switzerland: Springer International Publishing, 2014. 85 - 89.
- [3] Tagawa Y, Tanev I, Shimohara K. Performance rendering system based on interactive genetic algorithms [A]. *54th Annual Conference of the Society of Instrument and Control Engineers of Japan* [C]. USA: IEEE, 2015. 369 - 372.
- [4] 孙晓燕, 陈姗姗, 巩敦卫, 张勇. 基于区间适应值交互式遗传算法的加权多输出高斯过程代理模型 [J]. *自动化学报*, 2014, 40(2): 172 - 184.
Sun Xiao-yan, Chen Shan-shan, Gong Dun-wei, Zhang Yong. Weighted multi-output Gaussian process-based surrogate of interactive genetic algorithm with individual's interval fitness [J]. *Acta Automatica Sinica*, 2014, 40(2): 172 - 184. (in Chinese)
- [5] SUN Xiao-yan, GONG Dun-wei. Interactive genetic algorithms with individual's fuzzy and stochastic fitness [J].

- Chinese Journal of Electronics, 2009, 18(4): 619 – 624.
- [6] Ishibuchi H, Sudo T, Nojima Y. Interactive evolutionary computation with minimum fitness evaluation requirement and online algorithm design [J]. Springerplus, 2016, 5(1): 1 – 29.
- [7] Pallez D, Collard P, Baccino T, Dumercy L. Eye-tracking evolutionary algorithm to minimize user fatigue in IEC applied to interactive one-max problem [A]. Proceedings of the 9th Annual Conference Companion On Genetic And Evolutionary Computation [C]. Canada: ACM, 2007. 2883 – 2886.
- [8] Holmes T, Zanker J. Eye on the prize: using overt visual attention to drive fitness for interactive evolutionary computation [A]. Proceedings of the 10th Annual Conference On Genetic and Evolutionary Computation [C]. Canada: ACM, 2008. 1531 – 1538.
- [9] Fukumoto M, Ogawa S, Nakashima S, Imai JI. Extended interactive evolutionary computation using heart rate variability as fitness value for composing music chord progression [A]. Second World Congress on Nature and Biologically Inspired Computing [C]. USA: IEEE, 2010. 407 – 412.
- [10] Inoue T, Fukumoto M. Creation of sound contents by extended interactive evolutionary computation using heart rate variability [A]. International Conference on Biometrics and Kansei Engineering [C]. USA: IEEE, 2013. 133 – 136.
- [11] Zhang Y, Koren J. Efficient bayesian hierarchical user modeling for recommender systems [A]. Proceedings of the 30th annual International ACM SIGIR Conference on Research and Development in Information Retrieval [C]. Canada: ACM, 2007. 47 – 54.
- [12] Gong Dun-wei, Yao Xin, Yuan Jie. Interactive genetic algorithms with individual fitness not assigned by human [J]. Journal of Universal Computer Science, 2009, 15(13): 2446 – 2462.
- [13] 何大义, 邱菀华. 运用熵极大化准则求解连续型不确定性决策问题 [J]. 系统工程理论与实践, 2002, 22(9): 97 – 100.
- He Da-yi, Qiu Wan-hua. Solving continuous stochastic decision-making problem under uncertainty by the maximum entropy formulism [J]. Systems Engineering-Theory & Practice, 2002, 22(9): 97 – 100. (in Chinese)
- [14] 李坤, 黎明, 陈昊. 进化算法的困难性理论研究进展 [J]. 电子学报, 2014, 42(2): 383 – 390.
- Li Kun, Li Ming, Chen Hao. Research progress of hardness theories on evolutionary algorithm [J]. Acta Electronica Sinica, 2014, 42(2): 383 – 390. (in Chinese)
- [15] Hao G, Gong D, Huang Y. Interactive genetic algorithms based on estimation of user's most satisfactory individuals [A]. Proceedings of the 6th International Conference on Intelligent System Design and Applications [C]. USA: IEEE, 2006. 132 – 137.
- [16] 巩敦卫, 陈健. 基于精英集选择进化个体的交互式遗传算法 [J]. 电子学报, 2014, 42(8): 1538 – 1544.
- Gong Dun-wei, Chen Jian. Interactive genetic algorithms with selecting individuals using elite set [J]. Acta Electronica Sinica, 2014, 42(8): 1538 – 1544. (in Chinese)

作者简介



郭广颂 男, 1978 年 9 月出生, 吉林省集安市人, 副教授. 2001 年和 2007 年分别在郑州大学和中国矿业大学获工学学士、硕士学位. 主要从事智能控制与进化优化方面研究.

E-mail: guogs78@126.com



陈良骥 (通信作者) 男, 1978 年 3 月出生, 四川邻水人, 博士、教授、硕导. 2006 年在哈尔滨工业大学获工学博士学位. 主要从事智能控制与制造方面研究.

E-mail: chenjiacilj@163.com