第 25 卷第 3 期
2008 年 6 月

控 制 理 论 与 应 用
Control Theory & Applications

Vol. 25 No. 3
Jun. 2008

文章编号：1000−8152(2008)03−0434−05

基于多近似模型的交互式遗传算法

巩敦卫，周　勇，郭一楠

(中国矿业大学 信息与电气工程学院，江苏 徐州 221008)

摘要：人的疲劳问题是交互式遗传算法的核心问题，它制约了交互式遗传算法在复杂优化问题中的应用. 为了解决该问题，本文提出基于多近似模型的交互式遗传算法. 该算法首先将搜索空间划分，然后利用传统交互式遗传算法得到的数据，在不同子空间生成不同的近似模型，最后采用该模型近似人对进化个体的评价，从而减少人评价的数量，有效解决人的疲劳问题. 算法性能分析及在服装进化设计系统中的应用验证了其有效性.

关键词：遗传算法; 交互; 多近似模型; 服装设计

中图分类号：TP18　　**文献标识码**：A

Interactive genetic algorithms with multiple approximate models

GONG Dun-wei,　ZHOU Yong,　GUO Yi-nan

(School of Information and Electronical Engineering, China University of Mining and Technology, Xuzhou Jiangsu 221008, China)

Abstract: Human fatigue is a key problem which restricts the application of interactive genetic algorithms to complicated optimization problems. An interactive genetic algorithm with multiple approximate models is proposed to deal with this problem. In this algorithm, the search space is divided into several subspaces, in which different approximate models are generated with data from traditional interactive genetic algorithms. The approximate model is applied to approximate human subjective evaluations on individuals, thus reducing the number of human evaluations and effectively resolving the human fatigue problem. The efficacy of the algorithm is validated through the performance analysis and the application to fashion evolutionary design systems.

Key words: genetic algorithms; interaction; multiple approximate models; fashion design

1　引言(Introduction)

交互式遗传算法是解决隐式性能指标优化问题的有效方法，它将传统的进化机制与人的智能评价相结合，通过人给出个体的适应值，代替难以(或无法)显式表示的适应度函数[1]. 目前该方法已在艺术设计与CG动画、乐曲创作、虚拟现实等多领域得到广泛应用[2,3].

已有研究表明，人的疲劳问题是交互式遗传算法的核心问题，它制约了该方法在复杂优化问题中的进一步应用. 解决交互式遗传算法中人的疲劳问题成为成功应用交互式遗传算法的关键. 国内外学者提出诸多解决方法，其中利用个体适应值的估计值代替人的评价，从而减少人评价个体的数量，是减轻人的疲劳的有效方法之一，主要包括3个方面: 其一是基于距离的个体适应值估计[4,5]，这种估计方法一般要求适应度函数分布比较平缓，但通常适应度函数是未知的，因而对具体某一优化问题，难以判断该方法是否适用; 其二是基于基因意义单元的个体适应值估计[6]，该方法只适用于个体编码可分的情况; 其三是基于神经网络的个体适应值估计[7,8].从空间上讲，文[8]采用同一神经网络模型全局近似适应度函数分布. 事实上，即使对于简单的适应度函数，在整个决策变量区域上采用一个模型近似也是很困难的，其近似精度得不到保障，这会给进化带来不利影响，对于像交互式遗传算法中的隐式适应度函数，这个问题就更加严重.

适应值近似还在耗时或含噪适应值评价中有广阔的应用前景[9]. Charya等基于支持向量机估计个体的适应值[10]，该文在整个搜索空间上仍采用同一近似模型. Regis等针对显式耗时性能指标优化问题，采用不同的近似模型估计个体的适应值[11]，但该文考虑的是显式性能指标优化问题，没有考虑交互式遗传算法要解决的隐式性能指标优化问题，相比而言后者更加复杂，不确定因素多[12].

收稿日期: 2006−05−06; 收修改稿日期: 2007−02−02.
基金项目: 国家自然科学基金资助项目(60304016, 60575046, 60775044); 教育部“新世纪优秀人才支持计划”资助项目.

尽管单一近似模型在全局上逼近隐式适应度函数是有困难的, 但在局部高精度逼近隐式适应度函数是有可能的. 鉴于此, 为了提高近似模型的逼近精度, 在不同的决策变量区域采用不同的近似模型是一种可能的解决方案.

2 基于多近似模型的交互式遗传算法(Interactive genetic algorithms with multiple approximate models)

2.1 算法思想(Methodology of algorithms)

首先将搜索空间划分成一系列互不相交的子空间; 其次在搜索空间上采用传统的交互式遗传算法进化种群, 同时得到用于生成近似模型的数据集. 当某子空间上的数据集大到一定程度时, 基于该数据集生成该子空间上的近似模型; 再次对于已有近似模型的子空间上的个体, 采用该近似模型估计其适应值; 对于尚未有近似模型的子空间上的个体, 或者由人来评价其适应值, 或者采用多近似模型混合估计其适应值.

考虑如下一般优化问题:

$$\max f(x), x \in S \subseteq \mathbb{R}^n, \tag{1}$$

其中: $f(x)$为被优化的性能指标, 且不能用显式函数表示; x是n维决策变量; S是其取值范围. 在不引起混淆的情况下, 记相应的个体及搜索空间亦分别为x, S.

2.2 搜索空间划分(Division of search spaces)

为了在局部搜索空间上采用近似模型逼近$f(x)$, 对整个搜索空间S进行合理划分是十分有必要的. 设$S_1, S_2, \cdots, S_N$是S的一个分划, 即:

$$S_i \bigcap S_j = \emptyset, i \neq j, j = 1, 2, \cdots, N; \bigcup_{i=1}^{N} S_i = S, \tag{2}$$

则S即可划分为$S_1, S_2, \cdots, S_N$等N个子空间. 不同的i, S_i的大小可以相同也可以不同. 对于实际的优化问题, 考虑到个体的基因型和表现型之间有映射关系, 而个体的表现型可以直接由人确定其类属, 因此可以根据个体的表现型确定其基因型的类属关系, 从而给出S的合理划分. 在4.3节将通过实例给出S的划分方法.

2.3 近似模型的生成(Generation of approximate models)

为了在各子空间上生成近似模型, 首先必须获取相关的数据集. 为此在进化的初始阶段, 采用传统的交互式遗传算法, 即由人评价种群中的个体, 得到用于生成近似模型的数据集, 记为

$$D = \{(x_k, f(x_k))|x_k \in S, k = 1, 2, \cdots, M\}, \tag{3}$$

其中: $f(x_k)$为人给出的个体x_k的适应值;M为人评价的个体数.

对于$\forall x_k \in S$, 记

$$\alpha_{ki} = \begin{cases} 1, x_k \in S_i, \\ 0, x_k \notin S_i, \end{cases} \delta_i = \sum_{k=1}^{M} \alpha_{ki}, \tag{4}$$

若$\delta_i \geqslant \Delta_i$, 则可以采用数据集:

$$D_i = \{(x_k, f(x_k))|x_k \in S_i, k \in \{1, 2, \cdots, M\}\}, \tag{5}$$

生成S_i上$f(\cdot)$的近似模型$\hat{f}_i(\cdot)$, 其中: Δ_i是生成$\hat{f}_i(\cdot)$所需数据个数的下限; 若$\delta_i < \Delta_i$, 则生成$\hat{f}_i(\cdot)$所需的数据个数不够, 在以后的种群进化过程中, 通过人评价个体积累数据, 直至达到Δ_i为止.

注 1　一般来讲, 对于不同的S_i, 其空间大小与其上$f(\cdot)$的分布不同, 因此生成$\hat{f}_i(\cdot)$所需数据个数一般不同, 即$\Delta_i \neq \Delta_j, i \neq j, i, j = 1, 2, \cdots, N$.

注 2　对于相同的S_i, S_i上$f(\cdot)$的近似模型$\hat{f}_i(\cdot)$可能有多个, 生成不同的近似模型所需的数据个数也会不同. 为了讨论方便, 本文的$\hat{f}_i(\cdot)$均采用人工神经网络, 基于得到的数据集采用BP算法训练, 网络结构如图1所示[8].

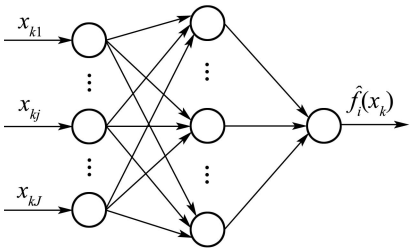

图 1　用于近似模型的神经网络结构

Fig. 1　Structure of the neural network for approximate models

图中: x_{kj}为进化个体x_k的第j部分的基因型(实际应用时, 根据情况可做必要的数制转换), $j = 1, 2, \cdots, J$; J为进化个体基因型分割的部分数; $\hat{f}_i(x_k)$为神经网络关于x_k的输出; 期望输出$f_i(x_k)$为人对x_k的评价值.

2.4 进化个体评价(Evaluation of individuals)

在种群进化过程中, 根据个体所在的区域采用不同的方式赋予个体适应值. 主要有3种方式:

1) 对$\forall S_i \in S$, 若已有S_i上$f(\cdot)$的近似模型$\hat{f}_i(\cdot)$, $\forall x_k \in S_i$, 采用$\hat{f}_i(\cdot)$评价得到个体适应值$\hat{f}_i(x_k)$.

2) 对$\forall S_i \in S$, 若尚未有S_i上$f(\cdot)$的近似模型$\hat{f}_i(\cdot)$且人愿意评价, 则仍由人评价得到个体适应值$f(x_k)$, 数据$(x_k, f(x_k))$用于产生近似模型$\hat{f}_i(\cdot)$.

3) 对$\forall S_i \in S$, 若尚未有S_i上$f(\cdot)$的近似模型$\hat{f}_i(\cdot)$但人不愿意评价, 则对$\forall x_k \in S_i$, 采用如下多模型混合策略评价, 得到个体适应值$\hat{f}_i'(x_k)$: 首先寻找S_i的最邻近的且已有近似模型的区域$S_p, p = 1, 2, \cdots, N_i$, N_i为S_i的最邻近的且已有近似模型的

区域个数; 其次计算被评价个体x_k与S_p的距离, 记为$D(x_k,S_p)$, 即: $D(x_k,S_p)=\min\{\|x_k-x_q\||x_q\in S_p,S_p\subseteq S\}$; 最后得到$x_k$的适应值:

$$\hat{f}_i^{'}(x_k)=\frac{\sum\limits_{p=1}^{N_i}D(x_k,S_p)^{-1}\hat{f}_p(x_k)}{\sum\limits_{p=1}^{N_i}D(x_k,S_p)^{-1}},\tag{6}$$

其中: $\hat{f}_p(x_k)$为采用S_p上的近似模型$\hat{f}_p(\cdot)$评价x_k的适应值.

注 3 式(6)表明, 在S_i上$f(\cdot)$的近似模型$\hat{f}_i(\cdot)$ 未知的情况下, 可以通过邻近区域上的近似模型估计其个体的适应值.

注 4 通过上述个体适应值赋予方法可知, 即使在初始进化阶段得不到子空间S_i上$f(\cdot)$ 的近似模型$\hat{f}_i(\cdot)$, 也可在种群进化过程中继续得到所需的数据$(x_k,f(x_k))$, 从而生成近似模型$\hat{f}_i(\cdot)$.

注 5 由于计算机和人分别评价不同的进化个体, 因而二者可以并行工作, 从而可以加快进化速度, 减少进化时间; 另一方面, 由于人只评价一部分个体, 从而可以减轻人的疲劳, 在第3节将进一步分析该问题.

3 性能分析(Performance analysis)

假设种群规模为β, 第t代种群中有$\beta_i(t)$个个体分布在S_i上, 则有$\sum\limits_{i=1}^{N}\beta_i(t)=\beta$. 假设到第$T_i$代种群中分布在$S_i$上的个体数为$\Delta_i$, 则有$\sum\limits_{t=1}^{T_i}\beta_i(t)=\Delta_i$.

记: $T_{\min}=\min\{T_i|i\in\{1,2,\cdots,N\}\}$, T为进化终止代数, 若$T>T_{\min}$, 则$\exists T_i>T_{\min}$, S_i上$(T-T_i)$代共$\sum\limits_{t=T_i+1}^{T}\beta_i(t)$个个体由$\hat{f}_i(\cdot)$评价, 则$S$上由$\{\hat{f}_i(\cdot),i=1,2,\cdots,N\}$评价的个体数为: $\sum\limits_{i=1}^{N}\sum\limits_{t=T_i+1}^{T}\beta_i(t)$, 此即人在种群进化过程中由于引入近似模型而少评价的个体数, 再考虑多近似模型混合评价策略, 即式(6)的影响, 则人在种群进化过程中少评价的个体数远大于$\sum\limits_{i=1}^{N}\sum\limits_{t=T_i+1}^{T}\beta_i(t)$. 由此可见, 本文方法可有效减少人评价的个体数, 从而减轻人的疲劳.

另一方面, 基于本文方法, 在S上采用多种群进化, 人最多需要评价$\sum\limits_{i=1}^{N}\Delta_i$个个体, 其余个体则由$\{\hat{f}_i(\cdot),i=1,2,\cdots,N\}$及式(6)评价. 因此当个体数小于或等于$\sum\limits_{i=1}^{N}\Delta_i$时, 人评价的个体数不多于传统交互式遗传算法; 但是当个体数大于$\sum\limits_{i=1}^{N}\Delta_i$ 时, 人评价的个体数不变, 且少于传统交互式遗传算法. 这说明, 在进化代数或种群规模增加的情况下, 人的疲劳程度并没有增加; 或者在人的疲劳程度相同的情况下, 可以大幅度增加进化代数或种群规模, 从而提高算法性能.

4 在服装进化设计系统中的应用 (Application in fashion evolutionary design systems)

4.1 个体编码(Codes of individuals)

本文以设计只有上衣和裙子的二维女装进化设计系统作为应用实例. 上衣和裙子的款式各32套, 分别可以被染成8种颜色, 因此搜索空间大小为 $2^5\times2^5\times2^3\times2^3=65536$. 个体采用二进制编码, 共有16位, 其中前5位表示上衣的代码, 第6位到第10位表示裙子的代码, 第11位到第13位表示上衣颜色的代码, 第14位到第16 位表示裙子颜色的代码, 如图2所示, 颜色的编码与取值如表1所示.上衣和裙子的款式以bmp图片格式存放在统一的文件夹中并按照一定顺序将图片排列、命名, 名称分别是从0到31的整数, 对应于相应部位款式二进制代码的十进制值. 例如图2中上衣代码10110 对应于编号为22的上衣款式, 裙子的代码00100对应于编号是4的裙子款式; 上衣颜色代码110对应于编号为6的黄色, 裙子颜色的代码001对应于编号为1的蓝色.

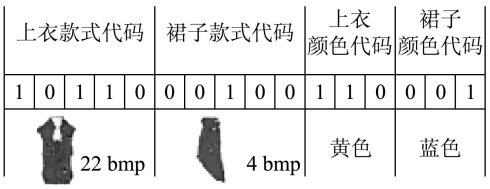

图 2 个体编码

Fig. 2 Codes of individuals

表 1 颜色编码与取值

Table 1 Codes and values of colors

编码	值	颜色	编码	值	颜色
000	0	黑色	100	4	红色
001	1	蓝色	101	5	洋红色
010	2	绿色	110	6	黄色
011	3	青色	111	7	白色

4.2 近似模型的选取(Selection of approximate models)

神经网络有3层, 输入层含有4个结点, 隐含层含有10个结点, 输出层含有1个结点. 为了对神经网络进行训练, 分别把表示上衣款式、裙子款式、上衣颜色、裙子颜色的二进制代码转换成相应的十进制数

作为神经网络的输入, 期望输出为人评价的个体适应值.

4.3　搜索空间划分(Division of search spaces)

主要针对服装颜色和色彩搭配的特点进行搜索空间的划分. 由表1可知, 上衣和裙子共有8种颜色, 且前4种颜色是冷色调, 后4种颜色是暖色调, 所以依据上衣和裙子颜色的色调把搜索空间划分为暖色调和冷色调搜索空间, 颜色编码为0**时表示服装采用冷色调的颜色, 颜色编码为1**时表示服装采用暖色调颜色, 这样, 搜索空间划分成如表2所示的4个子空间. 可以验证, 这种划分满足2.2节的要求.

表 2　搜索空间划分

Table 2　Division of search spaces

子空间	子空间模式	模式描述
S_1	**********0**0**	上衣和裙子都是冷色调
S_2	**********0**1**	上衣是冷色调, 裙子是暖色调
S_3	**********1**0**	上衣是暖色调, 裙子是冷色调
S_4	**********1**1**	上衣和裙子都是暖色调

4.4　设计过程(Process of design)

系统运行界面如图3所示, 每一次显示8套服装, 即种群规模为8, 个体适应值采用百分制打分, 采用适应值比例选择、多点交叉和单点变异, 交叉概率为1, 变异概率为0.01, 进化截止代数为500.

图 3　系统运行界面

Fig. 3 System's running interface

为简化考虑, 近似模型的训练数据集个数均取24, 即$\Delta_i = 24, i = 1, 2, 3, 4$. 各子空间已有的样本数显示在运行界面上, 当某一子空间的训练数据集个数达到24时, 则系统自动进行神经网络的训练. 训练完成后, 如果系统发现待评价的个体属于该子空间, 则系统会自动对个体进行评价, 给出相应的适应值. 如果人已经感觉疲劳了, 此时只要点击“混合适应值评价”按钮, 则系统会依式(6)自动对个体进行适应度赋值. 在进化过程中, 当某子空间训练样本数据集个数达到24时, 系统则对该子空间的神经网络进行训练, 生成近似模型.

图 4　系统进化到第5代的界面

Fig. 4 Interface when the system evolves to the 5th generation

图4为系统进化到第5代的界面. 由图4可知, 系统进化到第5代时, 子空间S_1, S_2, S_3, S_4中的样本数分别为4, 5, 24, 7, 此时子空间S_3的样本数已经达到训练数据集的要求, 系统开始训练子空间S_3的神经网络近似模型, 并用训练好的近似模型代替人对出现的属于该子空间的个体进行适应值评价. 系统继续进化, 人继续对服装个体进行评价, 系统进化到第9代的界面如图5所示, 此时子空间S_1, S_2, S_3, S_4中的样本数分别为9, 10, 24, 24. 子空间S_4的样本数已经达到训练数据集的要求, 系统开始训练子空间S_4的神经网络近似模型.

当系统运行到第9代时, 所有子空间中的样本数之和为67, 而系统在9代中总共出现的个体数为72, 样本数据集的个数小于系统已经评价的个体数. 从中可以看出, 在第6至第9代中共有5个个

体由近似模型评价.

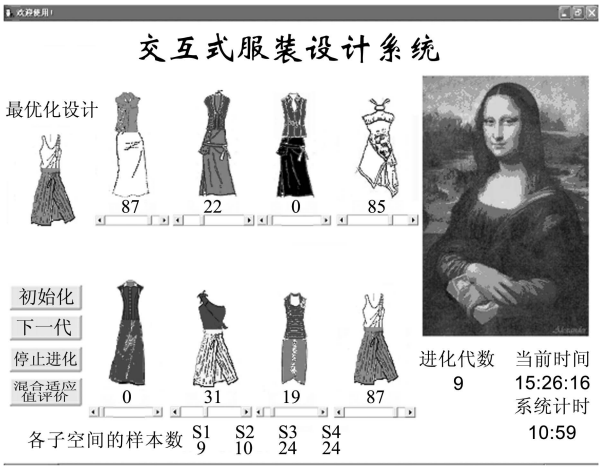

图 5 系统进化到第9代的界面

Fig. 5 Interface when the system evolves to the 9th generation

人在评价了9代以后已经感觉疲劳, 点击“混合适应值评价”按钮, 此时完全由计算机对系统进化生成的服装个体进行自动评价, 对于还没有近似模型的S_1和S_2中个体由式(6)生成其评价值.

系统进化到121代时收敛, 并找到了最优个体, 总共用时15’21”. 在整个进化过程中, 人评价的个体数为67个, 前9代用时10分59 s, 则人评价一个个体平均约需10 s. 系统运行121代, 共评价968个个体, 如果都由人来进行评价, 则约需9680 s, 即2小时41’20”, 这么长的时间人肯定会感到疲劳. 所以本文提出的基于多近似模型的交互式遗传算法不仅可以有效的减轻人的疲劳, 而且可以增加进化代数, 便于搜索到最优解.

5 结束语(Conclusion)

为了解决人的疲劳问题, 提高交互式遗传算法的性能, 本文提出了基于多近似模型的交互式遗传算法. 该算法的特点是: 采用多近似模型局部逼近人对个体的评价, 提高了逼近精度, 使找到的优化解更加“可靠”; 通过近似模型评价个体, 大幅度减少了人评价个体的数目, 从而有效解决了人的疲劳问题. 文中通过定量分析指出了本文算法在不增加人的疲劳的前提下增加种群规模和进化代数的可能性, 从而打破了交互式遗传算法对种群规模和进化代数的约束. 算法在服装进化设计系统中的应用也验证了其有效性.

本文没有考虑利用进化信息在线更新近似模型问题, 因此下一步将研究如何利用进化信息在线更新近似模型, 以提高近似模型的精度.

参考文献(References):

[1] TAKAGI H. Interactive evolutionary computation: fusion of the capabilities of EC optimization and human evaluation[J]. *Proceedings of the IEEE*, 2001, 89(9): 1275 – 1296.

[2] CHO S B. Towards creative evolutionary systems with interactive genetic algorithm[J]. *Applied Intelligence*, 2002, 16(2): 129 – 138.

[3] TOKUMARU M, MURANAKA N, IMANISHI S. Virtual stylist project: examination of adapting clothing search system to user’s subjectivity with interactive genetic algorithms[C] // *Proceedings of the IEEE Congress on Evolutionary Computation*. Piscataway, NJ: IEEE Press, 2003: 1036 – 1043.

[4] SUGIMOTO F, YONEYAMA M. An evaluation of hybrid fitness assignment strategy in interactive genetic algorithm[C] // *Proceedingsb of the 5th Australasia-Japan Joint Workshop on Intelligent and Evolutionary Systems*. 2001: 62 – 69.

[5] LEE J Y, CHO S B. Sparse fitness evaluation for reducing user burden in interactive genetic algorithm[C] // *Proceedings of the IEEE International Fuzzy Systems Conference*. Piscataway, NJ: IEEE Press, 1999, 2: 998 – 1003.

[6] 郝国生, 巩敦卫, 史有群, 等. 交互式遗传算法的机器代替用户方法[J]. 模式识别与人工智能, 2006, 19(1): 111 – 115.
(HAO Guosheng, GONG Dunwei, SHI Youqun, et al. Method of replacing the user with machine in interactive genetic algorithm[J]. *Pattern Recognition and Artificial Intelligence*, 2006, 19(1): 111–115.)

[7] BILES J A, ANDERSON P G, LOGGI L W. Neural network fitness functions for a musical IGA[C] // *Proceedings of the International ICSC Symposium on Intelligent Industrial Automation and Soft Computing*. 1996: 39 – 44.

[8] 周勇, 巩敦卫, 郝国生, 等. 交互式遗传算法基于NN的个体适应度分阶段估计[J]. 控制与决策, 2005, 20(2): 234 – 236.
(ZHOU Yong, GONG Dunwei, HAO Guosheng, et al. Neural network based phase estimation of individual fitness in interactive genetic algorithm[J]. *Control & Decision*, 2005, 20(2): 234 – 236.)

[9] JIN Y C, OLHOFER M, SENDHOFF B. A framework for evolutionary optimization with approximate fitness functions[J]. *IEEE Transactions on Evolutionary Computation*, 2002, 6(5): 481 – 494.

[10] CHARYA M B, LU G J. DAFHEA: a dynamic approximate fitness based hybrid EA for optimization problems[C] // *Proceedings of the IEEE Congress on Evolutionary Computation*. Piscataway, NJ: IEEE Press, 2003: 1879 – 1886.

[11] REGIS R G, SHOEMAKER C A. Local function approximation in evolutionary algorithms for the optimization of costly functions[J]. *IEEE Transactions on Evolutionary Computation*, 2004, 8(5): 490 – 505.

[12] JIN Y C, BRANKE J. Evolutionary optimization in uncertain environments-a survey[J]. *IEEE Transactions on Evolutionary Computation*, 2005, 9(3): 303 – 317.

作者简介:

巩敦卫 (1970—), 男, 教授, 博士生导师, 工学博士, 从事进化计算、智能控制的研究, E-mail: dwgong@vip.163.com;

周 勇 (1974—), 男, 博士, 副教授, 研究领域为进化优化;

郭一楠 (1975—), 女, 博士, 副教授, 硕士生导师, 研究领域为进化优化和多智能体系统等.