第35卷 第15期
Vol.35 No.15
计 算 机 工 程
Computer Engineering
2009年8月
August 2009
·人工智能及识别技术· 文章编号：1000—3428(2009)15—0203—03 文献标识码：A 中图分类号：TP18

遗传算法中模式的竞争度研究

汪 彤，李云强
(解放军信息工程大学电子技术学院，郑州 450004)

摘 要：针对Goldberg提出的利用竞争模式对欺骗问题进行度量所存在的弊端，提出模式竞争度的度量新方法，定义竞争函数及竞争度，给出模式竞争度的性质，证明模式竞争度和重要模式之间的关系，利用模式竞争度对2个典型的模式欺骗性与GA欺骗性不符的函数的遗传算法运行机理做了合理解释，表明模式竞争度作为新度量方法的合理性。

关键词：模式；竞争函数；竞争度

Research on Mode Competitive Degree in Genetic Algorithm

WANG Tong, LI Yun-qiang
(Electronic Technique Institute, PLA Information Engineering University, Zhengzhou 450004)

【**Abstract**】Aiming at for the disadvantages of using competitive mode to analyze deceptive functions which is promoted by Goldberg, this paper gives a new measuring method from a new angle. It defines competitive function and competitive degree, gives some characteristics of it, proves the connections between competitive degree and important mode, and by using competitive degree rationally explains how the Genetic Algorithm(GA) solves two representative functions which mode deception is not in correspond with GA's deception. It proves that competitive degree as a new measuring method is rational.

【**Key words**】mode; competitive function; competitive degree

1 概述

遗传算法(Genetic Algorithm, GA)作为一种仿生优化算法已成功地应用于诸多领域，但相对于应用而言，遗传算法的理论却不够完善且发展迟缓，极大地阻碍了应用层次的进一步拓展，如关于遗传算法欺骗问题的研究等。在遗传算法中，将所有妨碍评价适应度高的个体生成而影响遗传算法正常工作的问题统称欺骗问题。遗传算法运行过程具有将高于平均适应度、低阶和短定义长度的模式重组成高阶模式的趋势。如果在低阶、高适应度模式中包含了最优解的话，则遗传算法就有可能找到最优解，否则，就有可能对遗传算法形成欺骗，从而使遗传算法收敛到一个次优的结果。如何对一个优化函数的欺骗性进行度量，是遗传算法理论中比较重要的问题，适当度量将有助于指导利用遗传算法更好地解决实际问题。

Goldberg在Bethke早期工作的基础上，提出利用竞争模式的方法研究欺骗问题，并利用这种方法分析了欺骗问题的难易性。遗憾的是存在相应的反例说明：遗传算法的欺骗问题并不一定是遗传算法的困难问题，同样遗传算法的非欺骗问题并不一定是遗传算法的容易问题，如皇家大道问题。从上述矛盾可看出这种度量遗传算法欺骗问题的方式还不够合理，必须进行修改和完善。

本文从一个新的角度出发，提出了模式竞争度的度量新方法。

2 预备知识

为说明本文的系统性及完整性，下面给出本文要用到的一些预备知识。

定义1[1] 一个模式表示在某些基因位置上取确定值的所有个体的集合，它是个体空间的一个"超平面"。设$\theta_{i_k}\in\{0,1\},1\leqslant i_k\leqslant n$, n为编码串长度，模式S常表示为：$S\left[(i_k,\theta_{i_k});K\right]=\left\{X=(x_1,x_2,\cdots,x_n)\middle|x_{i_k}=\theta_{i_k},i_k\in K\right\}$，其中，$K$表示1~$n$之间一些整数的集合,$K$中元素的个数称为模式的阶。

定义2[1-2] 若模式S与S'中，确定位的位置完全一致，并且至少有一位确定位的编码不同，则称S与S'互为竞争模式。

定义3[1-2] 若优化函数$f(x)$的最大值点集为X^*，S为一包含X^*的m阶模式，如果存在S的竞争模式S'，使得$f(S)<f(S')$，则称f为m阶欺骗。

按照上述定义，对Stephanie Forrest, John Holland, Melanie Mitchell等人设计的皇家大道问题[1]进行分析，可以得到该问题不具备模式欺骗性,但随着基本模式长度的增加，GA求解过程变得越来越困难，有效基因信息的继承可能被阻断，可见用竞争模式对该问题进行研究难以获得满意的解释。

如一个典型的优化函数，$\max\limits_{0\leqslant x\leqslant 1} f(x)=\begin{cases}2^{L+1} & x=0\\ x^2 & 其他\end{cases}$，该式中$x$的二进制编码长为$L$，这个问题同样属于非模式欺骗问题，但却是GA难问题。在这类优化函数中，函数最优解的值与其他点所对应的值之间存在较大差值，因此，用模式欺骗性，这种基于平均适应值的处理方法对GA求解过程进行分析就极易产生误差，难于对问题进行解释。

其他反映模式欺骗性与GA难易性不吻合的实例可参考

作者简介：汪 彤(1981－)，女，硕士研究生，主研方向：遗传算法理论及其应用；李云强，教授、博士
收稿日期：2009-02-22 **E-mail：**wangtong810601@126.com

文献[1]。本文提出的度量模式竞争度的新方法，可对上述 2 个典型问题进行合理的解释，以此说明新度量方式的合理性。此外还要证明文献[3]所提出的重要模式的竞争度的性质。为了叙述方便，优化函数常用函数 f 表示，并用 $f[(i_k,\theta_{i_k});K]$ 表示函数 f 对于所有 $i_k \in K$，$x_{i_k}=\theta_{i_k}$ 时的函数。

定义 4[3]　如果存在模式 $S\left[(i_k,\theta_{i_k});K\right]$，对于任意的竞争模式 $S\left[(i_k,\xi_{i_k});K\right]$，都有 $f\left[(i_k,\theta_{i_k});K\right] \geqslant f\left[(i_k,\xi_{i_k});K\right]$ 成立，那么称模式 $S\left[(i_k,\theta_{i_k});K\right]$ 为 $|K|$ 阶重要模式。

引理 1[3]　如果 $S[(i_k,\theta_{i_k});K_1]$, $S[(i_k,\xi_{i_k});K_2]$ 都为重要模式，则 $S[(i_k,\theta_{i_k});K_1,(i_k,\xi_{i_k});K_2-K_1]$，$S[(i_k,\xi_{i_k});K_2,(i_k,\theta_{i_k});K_1-K_2]$ 都为重要模式，并且

$$f[(i_k,\theta_{i_k});K_1,(i_k,\xi_{i_k});K_2-K_1]=f[(i_k,\xi_{i_k});K_2,(i_k,\theta_{i_k});K_1-K_2]$$

3　主要结论

模式欺骗性与 GA 难易性的不匹配，表明模式欺骗性的定义尚不够合理，本文将通过模式的竞争度代替模式的欺骗性，进而利用模式的竞争度来衡量相应优化问题的 GA 难易性。下面给出一些相关定义。

3.1　竞争度的相关定义

定义 5　竞争函数：

$$\varphi(f_1,f_2)=\begin{cases}1 & f_1(x)>f_2(x)\\ 0 & f_1(x)=f_2(x)\\ -1 & f_1(x)<f_2(x)\end{cases}$$

定义 6　设 $S_1\left[(i_k,\theta_{i_k});K_1\right]$, $S_2\left[(i_k,\xi_{i_k});K_2\right]$ 为 2 个模式，则竞争度函数：

$$\eta(S_1,S_2)=\frac{1}{2^{n-|K_1\cup K_2|}}\sum_{\substack{x_{i1},x_{i2},\cdots,x_{ij}\\ ij\notin K_1\cup K_2}}\varphi(f\left[(i_k,\theta_{i_k});K_1,(i_k,\xi_{i_k});K_2-K_1\right],$$

$$f\left[(i_k,\xi_{i_k});K_2,(i_k,\theta_{i_k});K_1-K_2\right])$$

由定义知 $\eta\in[-1,1]$，若 $\eta(S_1,S_2)<0$，则称 S_1 相对 S_2 处于竞争劣势，也即 S_2 相对 S_1 处于竞争优势，$\eta(S_2,S_1)>0$。特别的，若 $\eta(S_1,S_2)=-1$，则称 S_1 相对 S_2 处于绝对竞争劣势，也即 S_2 相对 S_1 处于绝对竞争优势，$\eta(S_1,S_2)=1$。

3.2　实例分析

根据定义 6 可对本文第 2 节所提到 2 个典型函数进行分析，这是 2 个模式欺骗性与 GA 难易性不吻合的典型实例，下面的分析结果表明竞争度函数定义的合理性及利用竞争度对相关问题进行分析的有效性。

3.2.1　皇家大道(RR)问题的分析

为简便起见，首先以 3 个大道、每个大道为 4 阶基本模式构成的问题进行分析，简称 RR-3×4 问题，见表 1。

表 1　皇家大道问题(RR-3×4 问题)

序号	基本模式(大道)	权重
1	S_1=1111********	4
2	S_2=****1111****	4
3	S_3=********1111	4

对问题进行分析：由 $\eta(S_1,S_2)=0$, $\eta(S_2,S_3)=0$, $\eta(S_1,S_3)=0$ 可知，S_1，S_2，S_3 具有相同的竞争优势，它们是构成问题最优解的基本模式；令模式 $S_4=x_1x_2x_3x_4$********，其中，x_1,x_2,x_3,x_4 为定值且不全为 1，则这样的模式数共有 $2^4-1=15$ 个，通过用竞争度函数进行计算知这 15 个模式具有相同的竞争优势，并且 S_1 相对这 15 个模式具有绝对竞争优势。

由于随着基本模式长度的增加，GA 求解变得越来越困难，不失一般性，设基本模式阶数为 l，令 $S_1=\overbrace{11\cdots1}^{l}$*********，则 S_1 的竞争模式(即相应确定基因位上取值不全为 1 的模式)共有 2^l-1 个，它们具有相同的竞争优势，S_1 相对这些模式具有绝对竞争优势，当阶数 l 较小时，2^l-1 的值较小，即与 S_1 参与竞争的模式数较少，具有竞争优势的模式 S_1 能在竞争中取胜；当阶数 l 增大时，2^l-1 的值呈指数级增长，即与 S_1 参与竞争的模式数呈指数级增加，此时虽然 S_1 相对这些模式具有绝对竞争优势，但是相对庞大的竞争群体，由于比例的悬殊，S_1 的竞争优势难以发挥，这就造成了形成 S_1 的困难。

通过用竞争度对 RR 中模式的分析说明了随着基本模式长度的增加 GA 不利于求解的原因。

3.2.2　典型优化函数的分析

下面的问题属于非模式欺骗问题，但却是 GA 难问题：

$$\max_{0\leqslant x\leqslant 1} f(x)=\begin{cases}2^{L+1} & x=0\\ x^2 & \text{其他}\end{cases}$$

其中，该式中 x 的二进制编码长为 L。

对该问题中的模式进行分析：已知 $x=\overbrace{(00\cdots000)}^{L}$ 是问题的最优解，此时函数取最大值。令 S_a^b 表示模式中确定位的值为 a，阶为 b 的模式。不妨仅考虑确定基因位相连且确定基因位位于位串高 b 位的竞争模式之间的竞争关系。

$$S_0^1=0\overbrace{**\cdots*}^{L-1},\quad S_1^1=1\overbrace{**\cdots*}^{L-1}$$

只有当 L-1 个非确定位均取 0 时，模式 S_0^1 所对应的值大于模式 S_1^1 所对应的值，其余情况下模式 S_0^1 所对应的值小于模式 S_1^1 所对应的值，则 $\eta(S_0^1,S_1^1)=2^{2-L}-1$。

$$S_0^2=00\overbrace{**\cdots*}^{L-2},\quad S_1^2=01\overbrace{**\cdots*}^{L-2},\quad S_2^2=10\overbrace{**\cdots*}^{L-2},\quad S_3^2=11\overbrace{**\cdots*}^{L-2},$$

则 $\eta(S_0^2,S_1^2)=\eta(S_0^2,S_2^2)=\eta(S_0^2,S_3^2)=2^{3-L}-1$。

依此类推：$S_0^l=\overbrace{00\cdots0}^{l}\overbrace{**\cdots*}^{L-l}$，如它的竞争模式为 $S_a^l, a\in\left[1,2^l-1\right]$，则 $\eta(S_0^l,S_a^l)=2^{(l+1)-L}-1, a\in\left[1,2^l-1\right]$，只有当 $l=L-1$ 时，S_0^l 与 $S_a^l, a\in\left[1,2^l-1\right]$ 具有相同的竞争优势。从上述分析可知，这种竞争模式之间的竞争关系与模式确定位的选取无关。

当 $1\leqslant l\leqslant L-2$ 时，模式 S_0^l 相对其竞争模式 S_a^l 均处于竞争劣势，并且 $\eta(S_0^l,S_a^l)$ 的值随着编码长度 L 的增加而减小，L 值越大，$\eta(S_0^l,S_a^l)$ 的值越趋近于-1，即随着编码长度的增加，S_0^l 与它的 2^l-1 个竞争模式相比，不但处于竞争劣势，并且趋于绝对竞争劣势，这解释了随着编码长度的增加用 GA 求解越加困难的原因。从上述分析中可以看出，随着确定位 l 的增加，S_0^l 的竞争劣势状态有所改善，当 $l=L-1$ 时，S_0^l 与它的竞争模式具有相同的竞争优势，但这并不会对 GA 的性能有明显改善，这恰与模式定理和建筑模块假说相吻合。

由于这类优化函数最优解的值与其他点所对应的值之间存在较大差值，因此用模式欺骗性，这种基于平均适应值的处理方法对 GA 求解过程进行分析就易产生误差，而用本文

提出的利用竞争度分析的方法则能对问题进行较好的解释。

3.3 竞争度对单调函数的分析

给定 $y=f(x)\geqslant 0(x\in[0,1])$ 为单调升函数，最优解 $x^*=1$，$y^*=f(x^*)$。采用长度为 L 的二进制编码，适应值即目标函数值，则最优解位串模式为 $S^*=\overbrace{(1,1,\cdots,1)}^{L}$，即位串上任意基因位的有效基因值均为“1”，最小建筑模块为“1”。易得单调升函数的下述性质。

定理 1 对于单调升函数的二进制位串模式 S_1，S_2，若 2 个模式中确定位的取值均为 1，则 $\eta(S_1,S_2)=0$，即 2 个模式具有相同的竞争优势，与模式的阶及基因位的选取无关。

定理 2 对于单调升函数的二进制位串模式 S_1，S_2，若 2 个模式为竞争模式，且 $\lambda(S_1)-\lambda(S_2)>0$，则 $\eta(S_1,S_2)=1$，即模式 S_1 相对模式 S_2 具有绝对竞争优势。

从用竞争度对单调函数中模式的分析可以看出，哪类模式在 GA 的进化过程中更具有竞争优势，及模式之间的竞争状况，使对单调函数的 GA 求解过程有了更深入清晰的认识。

3.4 重要模式的竞争度

重要模式是一类特殊的积木块，重要模式的发现能够大大减小搜索空间，缩短寻优过程。通过利用竞争度对重要模式的分析，可看出重要模式在 GA 进化过程中的竞争优势地位，及重要模式与其他模式间的竞争关系。下面给出用竞争度对重要模式进行分析的结论。

定理 3 若 $S_1\left[(i_k,\theta_{i_k});K\right]$ 为重要模式，则对于任意的竞争模式 $S_2\left[(i_k,\xi_{i_k});K\right]$，都有 $0\leqslant\eta(S_1,S_2)\leqslant 1$。特别地，$\eta=0$ 当且仅当函数 f 的值与 K 中任意位置的取值无关；$\eta=1$ 当且仅当该模式对它的竞争模式 $S_2\left[(i_k,\xi_{i_k});K\right]$ 具有绝对竞争优势。

证明：若 $S_1\left[(i_k,\theta_{i_k});K\right]$ 为重要模式，则对于它任意的竞争模式 $S_2\left[(i_k,\xi_{i_k});K\right]$，都有 $f\left[(i_k,\theta_{i_k});K\right]\geqslant f\left[(i_k,\xi_{i_k});K\right]$，则 $0\leqslant\eta(S_1,S_2)\leqslant 1$。当 $\eta=0$ 时，即对它任意的竞争模式都有 $f\left[(i_k,\theta_{i_k});K\right]\equiv f\left[(i_k,\xi_{i_k});K\right]$，即函数 f 的值与 K 中任意位置的取值无关；当 $\eta=1$ 时，根据定义 6，即该重要模式对它的竞争模式 $S_2\left[(i_k,\xi_{i_k});K\right]$ 具有绝对竞争优势。

定理 4 如果重要模式 $S_1\left[(i_k,\theta_{i_k});K\right]$ 对它的所有竞争模式 $S_2\left[(i_k,\xi_{i_k});K\right]$ 都具有绝对竞争优势，则函数 f 的最优解在位置 K 上取值唯一。

证明：根据定义 4、定义 6 和定理 4 易证。

定理 5 若 $S_1\left[(i_k,\theta_{i_k});K_1\right]$，$S_2\left[(i_k,\xi_{i_k});K_2\right]$ 都为重要模式，则 $S_1\left[(i_k,\theta_{i_k});K_1,\ (i_k,\xi_{i_k});K_2-K_1\right]$，$S_2\left[(i_k,\xi_{i_k});K_2,(i_k,\theta_{i_k});K_1-K_2\right]$ 具有同等的竞争优势。

证明：由引理 1 可知，若 $S_1\left[(i_k,\theta_{i_k});K_1\right]$，$S_2\left[(i_k,\xi_{i_k});K_2\right]$ 都为重要模式，则：

$$S_1\left[(i_k,\theta_{i_k});K_1,\ (i_k,\xi_{i_k});K_2-K_1\right],\quad S_2\left[(i_k,\xi_{i_k});K_2,(i_k,\theta_{i_k});K_1-K_2\right]$$

也都为重要模式，且

$$f\left[(i_k,\theta_{i_k});K_1,\ (i_k,\xi_{i_k});K_2-K_1\right],\ f\left[(i_k,\xi_{i_k});K_2,(i_k,\theta_{i_k});K_1-K_2\right]$$

再根据定义 6，可得

$$S_1\left[(i_k,\theta_{i_k});K_1,\ (i_k,\xi_{i_k});K_2-K_1\right],\quad S_2\left[(i_k,\xi_{i_k});K_2,(i_k,\theta_{i_k});K_1-K_2\right]$$

具有同等的竞争优势。

4 结束语

针对利用竞争模式对欺骗问题进行度量存在的弊端，本文摒弃了竞争模式基于平均适应值的处理方法，提出了一种对欺骗问题的新的度量方式——竞争度。通过对 2 个模式欺骗性与 GA 欺骗性不吻合的典型实例的分析，表明了竞争度函数定义的合理性及在对问题进行分析时的有效性，它能够对欺骗问题进行合理的度量。并且利用竞争度对单调函数中的模式及重要模式进行分析，获得了有关竞争度的一些性质，这些性质对深入理解 GA 的运行机理及进一步研究都具有重要意义。

参考文献

[1] 李敏强, 寇纪淞, 林 丹, 等. 遗传算法的基本理论与应用[M]. 北京: 科学出版社, 2002.

[2] Goldberg D E. Genetic Algorithms in Search, Optimization, and Machine Learning[M]. MA, USA: Addison Wesley Press, 1989.

[3] 李云强, 余昭平. 遗传算法中重要模式及其性质[J]. 模式识别与人工智能, 2006,19(1): 20-23.

编辑 索书志

（上接第 202 页）

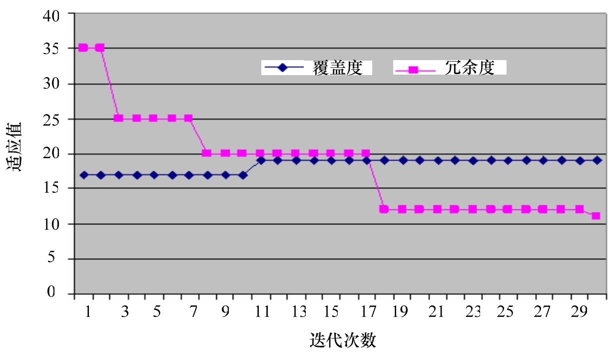

图 3 缩减迭代次数和全局最优位置适应值对应关系

6 结束语

本文将改进的粒子群算法应用到测试用例最小化问题中。结果表明，本算法提供一种与初始值无关、操作简单、计算速度快、性能优良的测试用例最小化算法，为软件测试用例最小化问题提供新的高效的解决途径。

参考文献

[1] Chen T Y, Lau M F. A Simulation Study on Some Heuristics for Test Suite Reduction[J]. Information and Software Technology, 1998, 40(13): 777-787.

[2] 郑燕妮, 李志蜀, 李 奇. 蚁群模拟退火算法在测试用例约简中的应用[J]. 计算机工程, 2009, 35(2): 197-199.

[3] 马雪英, 盛斌奎, 叶澄清. 用遗传算法的测试用例最小化[J]. 计算机科学, 2007, 34(1): 285-288.

[4] 章晓芳, 徐宝文, 聂长海, 等. 一种基于测试需求约简的测试用例集优化方法[J]. 计算机学报, 2004, 18(4): 821-831.

编辑 顾逸斐