

一种基于复数编码的遗传算法

郑朝晖, 张焱, 裘聿皇

(中国科学院自动化研究所, 北京 100080)

摘要: 第一次把复数编码的思想应用到遗传算法中去, 用复数编码来表达双倍体, 并具体规定了遗传操作. 每一个复数对应于双倍体的一对等位基因. 目标函数自变量的大小由其对应的复数的模决定, 符号则由相应复数的幅角决定. 与传统的实数编码的遗传算法相比, 本算法大大地扩展了表达空间的维数, 实验结果证明了本算法的有效性.

关键词: 复数编码; 遗传算法; 双倍体; 等位基因

中图分类号: TP391 **文献标识码:** A

Genetic algorithm based on complex-valued encoding

ZHENG Zhao-hui, ZHANG Yan, QIU Yu-huang

(Institute of Automation, Academia Sinica, Beijing 100080, China)

Abstract: Complex-valued Encoding is applied to genetic algorithms (GA). One complex number is used to denote each diploid and define the genetic operators. Each pair of alleles corresponds to one complex number. The independent variables of the objective function are determined by the modules and angles of their corresponding complex numbers. Compared with the conventional genetic algorithm based on real-valued encoding or binary encoding, the proposed algorithm expands the dimensions for denoting. The computer simulation results are offered to demonstrate the efficiency of the method.

Key words: complex-valued encoding; genetic algorithms; diploid; allele

1 引言 (Introduction)

传统的遗传算法中的染色体大都采用二进制编码或实数编码, 也就是分别用二进制数或实数码串来表达每个个体. 本文作者受到利用复数来表达人工神经网络权值的思想^[1]的启发, 首次将复数编码的思想应用到遗传算法, 并引入双倍体来代表算法中的个体. 由于每个复数可以表达二维信息, 这种方法大大地扩展了所能表达的特征空间的维数, 为遗传算法的编码方法提供了新思路.

本文利用复数的模与实自变量对应, 即模相等而幅角不等的复数对应于绝对值相同的实自变量. 这样, 尽管复数的实部和虚部仍然是由实数表示的, 但与传统的编码方式如实数编码相比, 本算法以实部和虚部两个参量来表达一个自变量, 大大地挖掘了群体中个体的多样性, 克服了有可能陷于局部解的弊病, 更有利于提高遗传算法的搜索效率. 另一个重要的特点在于, 由于复数表示形式及运算本身的多样性, 使用复数编码的遗传算法可以采用非常灵

活的遗传操作方式, 为使用遗传算法解决实际问题提供了新的途径.

2 基于双倍体结构的遗传算法 (GA based on diploid genotypes)

近年来, 基于双倍体结构 (diploid genotypes) 的遗传算法 (DGA)^[2]受到了人们的关注, DGA 具有更强的寻优能力和计算速度, 更为重要的是该方法的表达空间得到了很大的扩展. 自然界中, 复杂生物组织的染色体多采用双链或多链结构. 对双倍体来说, 在繁殖过程中, 父母体各提供一条染色体组成子体的染色体对. 由于复数编码的二维特性, 本文很自然地应用它来表示双倍体. 具体说来, 可利用一个复数来表达染色体对中的一对等位基因. 复数的实部和虚部分别称为实基因与虚基因. 在 DGA 中, 个体由两条等长度的基因链组成, 每个基因位有两个等位基因, 分别对应复数的实部与虚部. Hollstien 曾使用过三进制单点显性机制, 只是较好地维持多样性, 但其他方面没有显著改进^[2], 因此本文不考虑显性机

制.对于一个有 m 个自变量的问题来说,设有 m 个复数, $z_k = x_k + iy_k, k = 1, 2, \dots, m$. 记为:

$$((x_1, x_2, \dots, x_m), (y_1, y_2, \dots, y_m)),$$

形成染色体双链结构.若记

$$r = (x_1, x_2, \dots, x_m), d = (y_1, y_2, \dots, y_m),$$

则称 r 实基因串,为 d 虚基因串.

1) 初始群体的产生.

根据问题的定义区间 $[a_k, b_k], k = 1, 2, \dots, m$ (为了行文方便都写成闭区间,也可取开区间或半开,半闭区间,不影响下文的可行性).产生 m 个复数模: $\rho_k \in [0, \frac{b_k - a_k}{2}], k = 1, 2, \dots, m$.同时产生 m 个幅角 $\theta_k \in [0, 2\pi], k = 1, 2, \dots, m$.这样便构成 m 个复数

$$x_k + iy_k = \rho_k(\cos\theta_k + i\sin\theta_k), k = 1, 2, \dots, m.$$

用这种方法来形成所需要的实基因串 (x_1, x_2, \dots, x_m) 与虚基因串 (y_1, y_2, \dots, y_m) , 从而形成一个个体

$$((x_1, x_2, \dots, x_m), (y_1, y_2, \dots, y_m)).$$

如此产生若干个体,形成初始群体.

2) 选择操作.

为保证算法的收敛性,本文采用精英算法和适应度比例方法相结合的选择方法,即把群体中适应度最高的一部分个体不进行配对交叉而直接复制到下一代中,其余的个体按适应度比例方法选择产生.由于这两种方法都比较常用,本文就不再具体介绍了.

3) 交叉操作.

交叉是从群体中提取两个个体,交换遗传信息,产生后代的过程.交叉使个体之间有机会交流其优秀基因,可望获得比亲代更好的个体.本文利用类似实数编码中算术交叉(arithmetical crossover)的方法分别对两个父代个体中的实基因串对与虚基因串对进行交叉操作.

为叙述方便,设两个父代复基因串分别为

$$r_1 = (x_1^{(1)} + iy_1^{(1)}, \dots, x_m^{(1)} + iy_m^{(1)})$$

和

$$r_2 = (x_1^{(2)} + iy_1^{(2)}, \dots, x_m^{(2)} + iy_m^{(2)}).$$

并设交叉后的两个后代复基因串分别为

$$z_1 = (u_1^{(1)} + iv_1^{(1)}, \dots, u_m^{(1)} + iv_m^{(1)})$$

和

$$z_2 = (u_1^{(2)} + iv_1^{(2)}, \dots, u_m^{(2)} + iv_m^{(2)}).$$

交叉具体步骤如下:

先在 $[0, 1]$ 区间内生成 m 个随机数 $\alpha_1, \alpha_2, \dots,$

α_m , 则

$$u_k^{(1)} = \alpha_k x_k^{(1)} + (1 - \alpha_k) y_k^{(2)},$$

$$v_k^{(1)} = \alpha_k y_k^{(1)} + (1 - \alpha_k) x_k^{(2)},$$

$$u_k^{(2)} = \alpha_k y_k^{(2)} + (1 - \alpha_k) x_k^{(1)},$$

$$v_k^{(2)} = \alpha_k x_k^{(2)} + (1 - \alpha_k) y_k^{(1)}.$$

其中 $k = 1, 2, \dots, m$.可以证明这种交叉后的模仍属于 $[0, \frac{b_k - a_k}{2}]$ (证明略).这里也可看到参与交叉的不是复数模本身,幅角也起了重要影响.

4) 变异操作^[3].

在实数编码时,变异算子的作用已不仅仅是简单地恢复群体中多样性的损失,而主要是对解空间进行搜索.复数编码本身具有的群体多样性使变异算子的搜索作用更为重要.为此,本文采用了两种变异方法:对模采用自适应变异和对幅角采用多级变异.

· 模的自适应变异.

设 $(\rho_1, \rho_2, \dots, \rho_m)$ 是一个复数串个体的模分量, ρ_k 被选为进行变异,其定义区间是 $[0, \frac{b_k - a_k}{2}]$.则变异后的模分量为 $(\rho_1, \dots, \rho_{k-1}, \rho'_k, \dots, \rho_m)$, 其中

$$\rho'_k = \begin{cases} \rho_k + \Delta(T, \frac{b_k - a_k}{2} - \rho_k), & \text{若 } \text{rand} \leq 0.5, \\ \rho_k - \Delta(T, \rho_k), & \text{若 } \text{rand} > 0.5. \end{cases}$$

这里 rand 为 $[0, 1]$ 之间的一个随机数.函数 $\Delta(T, y) = y \cdot (1 - r^{T^\lambda})$, 其中 r 是 $[0, 1]$ 之间的一个随机数, λ 是决定非一致性程度的一个参数,它起着调整局部搜索区域的作用,其取值可为 2 到 5. T 为变异温度,定义为

$$T = 1 - \frac{f(z)}{f_{\max}}.$$

这里 $f(z)$ 为个体的适应值, f_{\max} 为 $f(z)$ 的上限值,通常取当前群体中的最大适应值.若变异后模大于 $\frac{b_k - a_k}{2}$ 则取消该次变异.

· 幅角的多级变异.

设 $(\theta_1, \theta_2, \dots, \theta_m)$ 是一个复数串个体的幅角分量,若 θ_k 被选为进行变异,则变异后的幅角分量为 $(\theta_1, \dots, \theta_{k-1}, \theta'_k, \dots, \theta_m)$, 这里 $\theta'_k = \theta_k \pm A\delta$, 以等概率取“+”号或“-”号,其中 $A = 0.2\pi, \delta = \sum_{j=0}^{\tau} \alpha_j 10^{-j}$, 变异精度 $\tau > 0$, 是正整数,根据所需精度

而定;每个 α_j 取 0 或 1, 并以 $p_\delta = \frac{1}{\tau + 1}$ 的概率取 1.

3 复数遗传算法流程(Process for GAs based on complex-valued encoding)

1) 随机产生 $2N$ 条长度为 m 的实数码链 (x_1, x_2, \dots, x_m) 与 (y_1, y_2, \dots, y_m) 组成初始群体的 N 个个体;

2) 对群体迭代地执行下面的步骤①, ②, 直到满足停止准则;

① 计算群体中每个个体对应的自变量, 并计算个体的适应值;

② 应用选择, 杂交, 和变异算子产生下一代群体;

3) 将得到最好的个体指定为算法的执行结果.

4 实验方案及结果(Simulation experiment and results)

为了比较本算法和传统的基于实数编码的遗传算法的搜索能力, 考虑 Ackley 函数^[4]的优化问题, Ackley 函数如下:

$$\min f(x_1, x_2) = -c_1 \exp(-c_2 \sqrt{\frac{1}{n} \sum_{j=1}^2 x_j^2}) - \exp(\frac{1}{n} \sum_{j=1}^2 \cos(c_3 x_j)) + c_1 + e, \quad -5 \leq x_j \leq 5, \quad j = 1, 2.$$

这里, $c_1 = 20, c_2 = 0.2, c_3 = 2\pi, n = 2$. 其最优解为

$$(x_1^*, x_2^*) = (0, 0), \quad f(x_1^*, x_2^*) = 0.$$

Ackley 函数是检测遗传算法搜索效率的典型问题函数, 由指数函数叠加上适度放大的余弦波再经调制而得到. 其特征是在一个几乎平坦的区域内由余弦波调制形成一个个孔或峰, 从而使曲面起伏不平, 为最优解的搜索增加了难度.

1) 实验中使用了 MATLAB 工具箱.

2) 适应度函数

$$\text{fitness}(x_1, x_2) = 20 - f(x_1, x_2).$$

由于 Ackley 函数是一个连续的实函数, 只有把复数个体转化为实数表示才能计算该个体的适应度, 其转换按下式进行. 其中 ρ_k 为表示该等位基因的复数的模, θ_k 为幅角, x_k 为该等位基因对应的问题的自变量.

$$x_k = \rho_k \text{sgn}(\sin \theta_k) + \frac{a_k + b_k}{2}, \quad k = 1, 2.$$

3) 遗传操作和参数设置.

实验中使用的复数编码遗传算法使用了本文的第2部分中所描述的遗传操作方式, 与之对比的基于实数编码的遗传算法采用了基本相同的遗传操

作, 选择使用精英算法和适应度比例方法结合的方法, 交叉操作为算术交叉, 变异操作采用多级变异.

为了科学、合理地比较两种算法, 实验中为它们设定了相同的参数, 其中种群大小为 100, 最大代数为 50, 变异概率为 0.5, 交叉概率为 0.5, 变异精度 τ 为 15, 自适应变异中 $\lambda = 2, f_{\max} = 20$. 精英算法中每代保留的最优个体数目为 2 个.

4) 实验结果如下.

在3)中给定的参数下, 使用基于实数编码和复数编码的遗传方法求解 Ackley 函数的最小值. 本文做了一个比较: 先使用基于实数编码的遗传算法, 运行 30 代后将最后群体分别作为基于实数编码和复数编码的遗传算法的初始群体, 再运行 30 代, 结果如图 1 所示.

其中, 实数编码算法得到的最优解为:

$$(x_1^*, x_2^*) = (0.00003463, -0.00018464),$$

$$f(x_1^*, x_2^*) = 0.00053229.$$

基于复数编码的算法得到的最优解为:

$$(x_1^*, x_2^*) = (0.00000538, 0.00000884),$$

$$f(x_1^*, x_2^*) = 0.00002927.$$

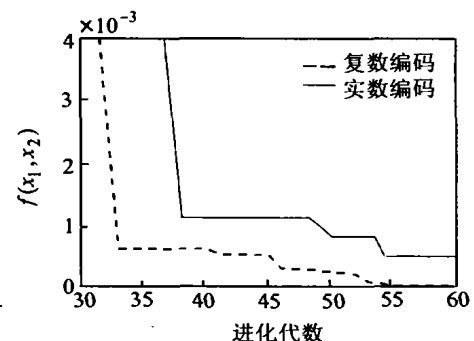


图1 具有相同初始群体时使用两种方法每代群体对应的函数最小值的比较

Fig. 1 The minimums in each generation got by using two methods with the same initial population

5 结语(Conclusion)

本文提出了一种基于复数编码的遗传算法. 通过引入双倍体, 描述了在复数编码下的遗传操作. 由于复数固有的二维特征, 使得每一个复数所能表达的空间维数比实数或二进制数多得多. 而且, 当将二维的复数映射到一维空间时, 即以二维的编码空间对应一维表达空间时, 无疑会大大改善群体基因的多样性. 与传统的遗传算法相比, 本方法具有内在的更强的保持群体基因多样性的能力, 从而使算法对全局最优点的搜索能力得到改善, 并且对遗传算法的发展提供了有益的探索, 实验结果表明其有效性.

参考文献(References):

- [1] CASASENT D, NATARAJAN S. A classifier neural network with complex-valued weights and square-law nonlinearities [J]. *Neural Networks*, 1995, 8(6): 989 - 998.
- [2] CHEN Guoliang, WANG Xifa, ZHUANG Zhenquan, et al. *Genetic Algorithms and the Applications* [M]. Beijing: People's Posts & Telecommunications Publishing House, 1996 (in Chinese).
- [3] PAN Zhengjun, KANG Lishan, CHEN Yuping. *Evolutionary Computing* [M]. Beijing: Tsinghua University Press, 1998 (in Chinese).
- [4] ACKLEY D. *A Connectionist Machine for Genetic Hillclimbing* [M]. Boston: Kluwer Academic Publishers, 1987.

作者简介:

郑朝晖 (1976—), 男, 1997年毕业于清华大学电机系获工学学士学位, 2000年获中国科学院自动化研究所硕士学位, 现在美国 State University of New York 攻读博士学位. 主要研究方向为: 模式识别, 图像处理, 遗传算法, 神经网络等;

张焱 (1975—), 男, 1996年毕业于东北大学, 获理学学士和工学学士学位, 现在中国科学院自动化研究所攻读硕士学位. 主要研究方向为: DNA 计算, 遗传算法, 神经网络等;

袁聿皇 (1942—), 男, 中国科学院自动化研究所研究员, 博士生导师. 研究方向: 鲁棒控制系统, 控制系统 CAD, 复杂系统建模, 遗传算法. Email: Yuhuang.Qiu@mail.ia.ac.cn

2003年科学计算自由软件“SCILAB”程序设计竞赛

2002年11月~2003年10月 中国北京

举办单位

清华大学智能技术与系统国家重点实验室; 中科院自动化所中法信息、自动化与应用数学联合实验室(LIAMA); 法国国家信息、自动化研究院(INRIA)

支持单位

中国863计划智能计算机专家组; 法国驻中国大使馆文化科技合作处; 中国人工智能学会; 中国人工智能学会; 中国自动化学会智能自动化专业委员会; 中国自动化学会机器人竞赛工作委员会; IEEE控制系统学会北京分会

竞赛宗旨

为中国学生及科学技术人员提供学习 SCILAB 和展示才华的机会, 促进中法双方的合作与交流.

SCILAB 简介

SCILAB 是由 INRIA 和 ENPC 开发的开放源码的科学计算软件, 目前正以每月 10000 份的速度通过 FTP 和 CDROM 的形式向全世界发布. SCILAB 与商业软件 MATLAB, MATRIX 和自由软件 OCTAVE, RLAB 类似.

参赛指南

- 参赛作品为自由的 SCILAB 工具箱, 可以参考网上已发布的工具箱及源代码.
- 参赛者可以自选感兴趣的题目和内容. 所创作的软件应包括演示、示例、在线帮助和文档. 在线帮助和文档的语言可以是汉语、法语或英语.

- 演示程序用 SCILAB 编写, 也可包含其它语言, 如 C, FORTRAN, C++, Java 等.

- 工具箱应能运行于 SCILAB2.6 环境下, 可以选择任一 SCILAB2.6 二进制版本支持的平台.

- SCILAB 软件可从以下网址下载: liama.ia.ac.cn/Scilab, www-rocq.inria.fr/scilab

* 注: 欲投稿者, 可向联系人索要论文格式要求.

日程安排

- 2003年3月1日前将参赛意向寄至联系人; · 2003年10月1日前将参赛作品寄至联系人;

- 2003年10月1日~30日: 参赛作品评审; · 2003年11月(具体日期待定): 颁奖仪式暨作品展示

奖励信息

- 一等奖 5 名: 访问法国 INRIA, 巴黎一个月; · 二等奖 10 名: 奖励人民币 1000 元

联系人

钱宗华 清华大学计算机系 100084

电话: 010-62788939(O), 62784458(H), 传真: 62771138

E-mail: hqzh@s1000e.cc.tsinghua.edu.cn