

文章编号: 1001-0920(2003) 02-0173-04

## 改进的多模态遗传算法及其在投资组合中的应用

刘洪杰, 王秀峰, 王治宝

(南开大学 信息技术科学学院, 天津 300071)

**摘要:** 提出用多模态遗传算法求解投资组合的新思路。首先, 针对小生境遗传算法搜索结果不稳定的缺点, 提出具有“迁徙操作”的新多模态遗传算法, 不仅有效地找到了全部优质解, 而且无需峰间距的信息。然后, 针对传统的投资组合模型存在不能满足不同风险偏好投资者需要的缺点, 提出符合中国国情的证券投资模型, 并给出利用改进的多模态遗传算法求解的方法。最后进行了实证研究, 得到了满意的结果。

**关键词:** 遗传算法; 多模态; 小生境; 投资组合  
**中图分类号:** TP18      **文献标识码:** A

## Improved multi-modal genetic algorithms and its application on portfolio investment model

LIU Hong-jie, WANG Xiu-feng, WANG Zhi-bao

(College of Information Technology Science, Nankai University, Tianjin 300071, China)

**Abstract:** A new approach to find investing scheme with multi-modal genetic algorithm is proposed. In view of the fact that the traditional portfolio investment model does not meet the demands of different venture capitalist, a new model that fits the circumstances of Chinese securities business is developed. A new approach is presented to find the solution that incorporates Niche GA with "Migrating Operator" which can overcome the instability of Niche GA and find effectively all good solutions without the information of the distances between the good solutions. The method is tested and the expected results are obtained.

**Key words:** Genetic algorithms; Multi-modal; Niche; Portfolio investment model

### 1 引言

在实际证券投资中, 不同投资者对收益和风险的偏好程度是不同的。而 Markowitz 模型<sup>[1]</sup>和 Sharp 模型均以投资者对风险厌恶为前提, 因而不能满足不同风险偏好投资者的实际要求。在 Markowitz 模型的基础上, 并考虑国内证券市场实际操作的限制, 本文提出了符合中国国情的新投资组合模型。基于该模型, 利用本文给出的新的多模态遗传算法求解, 能得到一组不同风险水平的最优解,

以供不同的风险偏好的投资者选择。最后进行了实证研究, 得到了预期结果。

### 2 多模态遗传算法

在社会实践中人们注意到, 通过最优算法求出的最优解, 由于种种客观条件的限制在实际中常常不能应用, 因此需要更多的次优和局部最优解以供选择。但是, 传统的遗传算法只能搜索到定义域内的最优解(高峰), 显然不能满足实际需要。在这种情况下, 人们开始研究利用遗传算法来搜索定义域内的

收稿日期: 2001-12-07; 修回日期: 2002-04-15。

基金项目: 国家重大自然科学基金资助项目(79790130)。

作者简介: 刘洪杰(1975—), 男, 山东东明人, 博士, 从事遗传算法、金融智能决策等研究; 王秀峰(1942—), 男, 山东宁津

所有峰的方法,即多模态遗传算法。Holland<sup>[2]</sup>引入小生境和物种理论试图解决多模态问题,但未进行实际应用。Goldberg 和 Richardson<sup>[3]</sup>用分享与限制交配机制相结合的方法成功地实现了多模态的搜索。

### 2.1 小生境遗传算法<sup>[3]</sup>

小生境遗传算法是在传统遗传算法的基础上,采取如下两项措施使得遗传算法可以进行多峰搜索:

1) 定义一个分享函数来确定群体中每个个体的相邻关系和分享度(分享适应值)。这样便可有效地控制峰附近的个体数,使种群中的个体不会全部聚集到某一个峰。

2) 定义交配半径来限制交配。只有距离小于交配半径的个体之间才能交配,限制了处在不同峰领域的个体之间进行交叉,从而抑制了个别物种的无限制地增长或消失。

该方法能搜索到一些峰,但搜索结果不稳定,且在计算分享度时,需要计算两两个体之间的距离,计算量相当大。本文提出一种“迁徙操作”的新方法。该方法不仅能够稳定地找到定义域内的所有峰,而且无需求峰间距的信息。

### 2.2 迁徙操作

迁徙操作的目的是将种群中的精英个体(分享适应值超过阈值  $T$ (迁徙阈值))迁徙到优育子群中,其作用是保存种群中的精英个体。当种群中有个体迁徙到优育子群时,在原种群中的该精英个体及其近亲(与该个体的距离小于某一个阈值(近亲阈值))将被随机产生的新个体代替。然后检查种群中是否还有精英个体,如有则继续迁徙操作;否则停止迁徙操作。同时在优育子群中,一旦有新的精英个体进入优育子群,立即对该个体进行梯度进化<sup>[4]</sup>,直到该个体无显著变化。最后检查优育子群中是否有重复的个体,如果有,则只保留一个而删除重复个体。

经过迁徙操作,种群中的优秀个体得以保留,并进行特殊培育。而一旦有个体迁徙到优育子群,则立即在种群中淘汰该个体,补充新的个体,大大增加了种群的多样性。这样不仅可以保留找到的峰,使得搜索的结果更加稳定,而且由于个体在优育子群中得以特殊的培育,使得搜索的精度大大提高。

## 3 证券投资组合模型

### 3.1 马柯维茨模型<sup>[1]</sup>

马柯维茨模型的数学模型分为两种:一种是风险固定条件下的收益最大模型,其数学描述为

$$\begin{cases} \max r_p = \xi^T \alpha \\ \text{s. t. } \xi^T \Sigma \xi = \sigma_p^2, \quad E^T \xi = 1 \end{cases} \quad (1)$$

另一种是收益固定、风险最小模型,其数学描述为

$$\begin{cases} \min \sigma_p^2 = \xi^T \Sigma \xi \\ \text{s. t. } \xi^T \alpha = r_p, \quad E^T \xi = 1 \end{cases} \quad (2)$$

其中:  $\xi = (\xi_1, \xi_2, \dots, \xi_m)^T$ ,  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_m)^T$ ,  $E = (1, 1, \dots, 1)^T$ ,  $\xi_i$  表示第  $i$  种证券投资所占总投资的比例系数,  $i = 1, 2, \dots, m$ ,  $m$  为选定的证券种类数;

$E^T \xi = \sum_{i=1}^m \xi_i = 1$ ;  $\alpha$  表示第  $i$  种证券的收益率;  $r_p$  表

示证券投资组合的收益率;  $\sigma_p^2$  表示证券投资组合的风险,  $\sigma_p^2 = \xi^T \Sigma \xi$ ,  $\Sigma = (\sigma_{ij})_{m \times m}$  表示收益率协方差矩阵。式(1)中的  $\sigma_p^2$  为给定的风险方差,式(2)中的  $r_p$  为给定收益。模型的目的是确定投资组合系数,以达到目标函数最大(最小)。

### 3.2 改进的投资组合模型

由于不同的投资者对风险的偏好程度不同,他们会选择不同的“收益-风险”组合,因而马柯维茨模型难以满足众多投资者的不同需要。鉴于此,本文将收益和风险统筹考虑,希望得到不同的“收益-风险”组合。同时在我国证券市场中,由于有很多限制和规定,由马柯维茨模型得到的结果往往在实际操作中难以实现。为此,本文提出基于我国证券市场现状的多解投资组合模型。该模型从投资者实用的角度出发,并充分考虑了各种实际约束。在国内证券交易中,以下两个限制和规定是值得考虑的:1) 买股票时的最小单位是 1 手,即 100 股;2) 买卖股票大多要交纳一定的交易费用,交易费用分为印花税和佣金两部分。

在考虑上述两个限制的前提下,模型的目标应分为以下 4 个:1) 组合收益最大;2) 组合风险最小;3) 资金利用率最大;4) 交易费用最小。

模型的限制条件也有相应的变化,限制条件如下:1) 每种证券的投资数量不小于 1 手;2) 买股票的资金和交易费用之和小于总的资金。

基于上述的目标和限制,本文提出改进的证券投资组合模型为

$$\begin{cases} \max J | x = f(r_p, \sigma_p^2, M_l, M_T) \\ \text{s. t. } x_i \bmod 100 = 0, \quad x_i \geq 0 \\ \quad \quad \quad M_T + M_v = M \end{cases} \quad (3)$$

其中:  $r_p = \xi^T \alpha$ ,  $\sigma_p^2 = \xi^T \Sigma \xi$ ,  $M_l = M - M_v - M_T$ ,  $M_T = \sum_{i=1}^m t_i$ ,  $M_v = \sum_{j=1}^m x_j p_j$ ,  $\xi_i = \frac{x_i p_i}{M_v}$ ;  $X = (x_1, x_2, \dots, x_m)$  为各个证券的投资数量;  $p_i$  表示第  $i$  种证券最近

一日的收盘价格;  $M$  为总资金,  $M_v$  为投资总市值,  $M_l$  为剩余资金,  $M_T$  为总的交易费用;  $t_i$  为第  $i$  种证券的交易费用; 其他符号的含义同式(1)和(2)。 $f(r_p, \sigma_p^2, M_l, M_T)$  为目标函数。为达到模型的 4 个目标, 要求  $f(r_p, \sigma_p^2, M_l, M_T)$  与  $r^p$  变化方向相同, 与  $\sigma_p^2, M_T$  和  $M_l$  变化方向相反。

#### 4 模型的多模态遗传算法求解方法

本文采用带有迁徙操作的小生境遗传算法来辨识模型参数。由于模型带有约束, 需要定义惩罚函数。惩罚函数定义为

$$\begin{cases} \theta_1 = \prod_{i=1}^m p_i \times \left| \frac{x_i}{100} - \left[ \frac{x_i}{100} \right] \right| \\ \theta_2 = 200 \max\left(0, \frac{M_v + M_T - M}{M}\right) \\ \theta = e^{10\theta_1 + \theta_2} \end{cases} \quad (4)$$

式中: 第 1 个惩罚项是针对购买数量最小单位是 1 手的限制, 其中  $[x]$  表示  $x$  的四舍五入的整数; 第 2 个惩罚项是针对总市值与交易费用之和小于总资金约束的, 总市值和交易费用之和如果大于总资金, 惩罚项开始发生作用。为加大惩罚力度, 本文采用指数函数对惩罚项进行放大。

适应值函数定义为目标函数减惩罚函数, 即

$$\text{fitness}(X) = f(r_p, \sigma_p^2, M_l, M_T) - \theta \quad (5)$$

小生境遗传算法中的限制交配半径的作用是保持子群的稳定, 但投资组合模型的目标函数是一个非规则的多模态函数, 不能预先得到峰之间的距离, 也就无法定义交配半径。因此没有迁徙操作的小生境遗传算法不能求解此模型。人们注意到, 迁徙操

作不仅可以极大地提高种群的多样性, 而且可以保存找到的峰, 这样, 限制交配的作用便可以迁徙操作代替。因此在实施过程中, 本文为交配半径赋一个比较大的常值。

#### 5 实证研究

为了尽量抵消风险, 本文选取不同行业、不同流通盘、不同地区的 5 支股票: 600074(南京中达), 600684(珠江实业), 0728(北京化二), 0965(天水股份) 和 0691(寰岛实业)。选取 2000 年 7 月 17 日 ~ 2000 年 8 月 18 日之间的数据作为训练数据, 总资金是 10 万元。平均收益率和风险方差矩阵为

$$R = \begin{pmatrix} 0.01675 & 0.00859 & 0.05146 & 0.04227 & 0.09462 \\ 0.01002 & 0.00319 & 0.01093 & 0.00319 & 0.00934 & -0.00057 \\ 0.01093 & -0.00057 & -0.02392 & 0.00025 & -0.01612 & 0.01793 \\ 0.01786 & -0.01779 & 0.04677 & 0.00025 & 0.01786 & -0.01612 & -0.01779 \\ 0.01793 & 0.04677 & 0.05139 & 0.07250 & 0.07250 & 0.15965 \end{pmatrix}$$

采用实数编码, 个体包括 5 个基因, 每个基因就是每支股票的购买手数。本文采用算术交叉, 变异算子采用均匀随机变异<sup>[5]</sup>, 选择采用排序选择<sup>[3]</sup>。参数如表 1 所示。

表 1 遗传算法的参数

种群规模	交叉概率	变异概率	交配半径	迁徙阈值	梯度步长	进化代数
20	0.2	0.03	100	-1	0.01	100

表 2 测试结果

组	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	总市值 / 万元	收益 $r_p / \%$	风险 $\sigma_p^2 / \%$	$(r_p - \sigma_p^2) / \%$
1	2.09	8.92	5.95	12.01	55.09	9.93	7.23	8.55	-1.32
2	3.1	21	14.05	8.11	39.97	9.93	5.94	4.79	1.15
3	2.04	20.09	33.04	7.07	26.01	9.93	5.33	3.18	2.15
4	5.94	22.92	20.89	13.92	22.11	9.93	4.79	2.69	2.10
5	19.1	17.1	20	6.92	21.05	9.93	4.43	2.30	2.13
6	15.13	22.99	18.05	8.10	20.92	9.93	4.35	2.15	2.20
7	17.98	24.93	16.08	6.92	19.94	9.93	4.12	1.90	2.22
8	31.08	1.91	26.85	15.15	3.9	9.93	3.54	1.46	2.08
9	30.88	22.06	14.91	7.03	6.9	9.93	2.93	0.94	1.99
10	27.06	48.95	1.89	1.09	7.09	9.92	2.04	0.60	1.44

选取目标函数为

$$f(r_p, \sigma_p^2, M_l, M_T) = e^{100(r_p - \sigma_p^2)} - e^{10(M_l + M_T)/M} \quad (6)$$

式(6)第1项中: $r_p$ 为投资组合的收益, $\sigma_p^2$ 为投资组合的风险。模型的第1目标是收益大风险小,因此采用二者之差来权衡投资组合的收益和风险。式(6)第2项是“负”号,其中: $M_T$ 为交易费用, $M_l$ 为剩余资金,两项相加再除以原始总资金 $M$ 用以衡量总资金的利用率。这是模型的另一目标,即资金的利用率最大。最后,定义适应值函数

$$\text{fitness} = f - \theta \quad (7)$$

经过计算,得到10组投资组合策略,如表2所示。其中 $x_1 \sim x_5$ 分别代表5支股票的购买数量(单位是手)。总市值是购买股票的总市值,不包括交易费用。收益和风险是该组投资组合的总的收益率和风险。第1组解收益最大,风险也最大;第10组解风险最小,收益也最小。

参考文献(References):

- [1] 戴玉林. 马柯维茨模型的分析与评价[J]. 金融研究, 1991, 9(2): 21-28.

(Dai Y L. Analysis and evaluation of Markowitz model [J]. *Finance Research*, 1991, 9(2): 21-28.)

- [2] Holland J H. An introduction analysis with applications to biology, control, and artificial intelligence [A]. *Adaptation in Natural and Artificial System* [C]. Ann Arbor: The University of Michigan Press, 1975.
- [3] Goldberg D E, Richardson J. Genetic algorithms with sharing for multi-modal function optimization [A]. *Proc of the 2nd Int Conf on Genetic Algorithms* [C]. Cambridge: The Massachusetts Institute of Technology, 1987. 41-49.
- [4] 刘洪杰, 王秀峰, 王治宝. 遗传多峰搜索[J]. 系统工程学报, 2000, 15(4): 321-326.  
(Liu H J, Wang X F, Wang Z B, Multi-modal genetic algorithms searching [J]. *J of Systems Engineering*, 2000, 15(4): 321-326.)
- [5] Wang X F, Elbuluk M E, Cabrera L A. The application of genetic algorithm with neural networks to the induction machines modeling [J]. *System Analysis Modeling Simulation*, 1998, 31: 93-105.

(上接第172页)

参考文献(References):

- [1] Branicky M S. Multiple Lyapunov function and other analysis tools for switched and hybrid systems [J]. *IEEE Trans on Automatic Control*, 1998, 43(4): 475-482.
- [2] LM-Aguilar J, Garcia R A, Troarevsky M L. Exponential stability of a certain class of hybrid systems and digital feedback stabilizers [A]. *Proc ACC* [C]. Philadelphia Pennsylvania, 1998. 713-717.
- [3] Ezzine J, Haddad A H. Controllability and observability of hybrid systems [J]. *Int J of Control*, 1989, 49(6): 2045-2055.
- [4] Peleties P, DeCarlo R. Asymptotic stability of switched systems using Lyapunov-like functions [A]. *Proc ACC* [C]. Green Valley, 1991. 1679-1684.
- [5] Wicks M A, Peleties P, DeCarlo R A. Construction of piecewise Lyapunov function for stabilizing switched systems [A]. *Proc IEEE CDC* [C]. Piscataway, 1994. 3492-3497.
- [6] Frommer H, Kulkarni S R, Ramadge P J. Controller switching based on output prediction errors [J]. *IEEE Trans on Automatic Control*, 1998, 43(5): 596-607.
- [7] Wang L Y, Khargonekar P P, Beydoun A. Robust con-

trol of hybrid systems: Performance guided strategies [A]. *Hybrid Systems V* [C]. New York: Springer Verlag, 1995. 1567: 356-389.

- [8] Pettersson S, Lennartson B. Stabilization of hybrid systems using a min-projection strategy [A]. *American Control Conf Proc of the 2001* [C]. Arlington, 2001. 1: 223-228.
- [9] Hu Bo, Zhai Guisheng, Michel A N. Hybrid output feedback stabilization of two-dimensional linear control systems [A]. *American Control Conf Proc of the 2000* [C]. Chicago, 2000. 3: 2184-2188.
- [10] Akar M, Ozguner U. Sliding mode control using state/output feedback in hybrid systems [A]. *Decision and Control 1998 Proc of the 37th IEEE Conf on* [C]. Tampa, 1998. 3: 2421-2422.
- [11] Agrawal A K, Yang J N, Schmitendorf W E. Hybrid control of buildings using nonlinear polynomial output feedback [A]. *American Control Conf Proc of the 1998* [C]. Philadelphia, 1998. 4: 2554-2558.
- [12] Serrano C C, Ramaadra P J. Periodicity and chaos from switched flow system: Contrasting examples of discretely controlled continuous system [J]. *IEEE Trans on Automatic Control*, 1993, 38(1): 70-83.