

基于并行计算及入侵检测系统的异动股票选择

谢 鑫 胡云姣*

(北京化工大学 理学院, 北京 100029)

摘 要: 将入侵检测技术引入到股票的选择中,给出了一种基于并行遗传算法的入侵检测技术。将数据随机分为若干个小组,对每一个小组中的数据进行检测,把每个小组中的最可疑的行为特征汇合后进行比对,最后选择出异动股票。对上海电力股票 2008 年 4 月 24 日至 7 月 22 日的数据利用 MATLAB 进行了模拟,与标准遗传算法比较,改进后的算法降低了误报率,缩短了运算时间。

关键词: 并行计算; 遗传算法; 股票选择; 入侵检测系统

中图分类号: O29

引 言

数学领域中最常用的股票选择方法是属性层次模型^[1-2]。近年来,随着计算机技术和人工智能技术的迅猛发展,为股票的选择以及股价的预测等提供了许多的新技术方法:遗传算法^[3],神经网络^[4]等。遗传算法是一种模拟生命进化机制的搜索优化算法,它从初始种群出发,通过随机选择、交叉和变异操作产生新的更适应环境的个体,使群体进化到搜索空间中越来越好的区域^[5]。并行遗传算法的基本思想就是把种群分成若干份的子种群,在每个节点机上对其中一个子种群做遗传操作。入侵检测系统是一种对网络传输进行即时监视,在发现可疑传输时发出警报或者采取主动反应措施的网络安全设备^[6]。基于遗传算法的入侵检测系统的应用还在初始阶段。现在基于遗传算法的入侵检测已在某些方面取得了成功的应用,如在优化属性聚类方面的应用^[7]。并且遗传算法的适应性较强,可以说,遗传算法在这方面有广阔的应用前景。

本文将入侵检测方法引用到异动股的选择当中,利用并行遗传算法降低误报率,缩短运算时间。

1 入侵检测系统的设计

1.1 基本思想

基于并行遗传算法的入侵检测系统是对基于标

准遗传算法的入侵检测模型的一种改进。把并行遗传算法运用到入侵检测模型中,会大大加快检测速度,加快得到全局最优解的速度,并降低了检测的误报率。

与标准遗传算法相比,并行遗传算法把初始种群分为了若干的子群,每一个子群在各自的处理器中独自进行遗传操作。把并行遗传算法应用到入侵检测系统,将从网络中随机提取的含有新的入侵特征的数据流随机分为若干个小组,对每一个小组中的数据流进行检测。经过若干次比对之后,把每个小组中的最有可疑的行为特征汇合到一起再进行比对,最后得到最接近异动特征的活动。

1.2 算法过程

(1)产生初始种群

生成种群规模为 N 的初始种群,即种群中含有 N 个个体。这里选取 N 的股票价格,记为:

$$P = (p_1, p_2, \dots, p_N)$$

(2)确定编码方式

选择实数编码方式。因为股票价格含有小数,且变化较多,所以采用实数编码最好运算。

(3)把初始种群分为 n 个子群

并行遗传算法的精髓就在于把种群分到若干个处理器中去计算,最后再把其合并到一起。因此,接下来要把初始种群 P 分为 n 个子群。

规定把第一步生成的 N 个个体,平均分成 n 个子群,每个子群的种群规模为 N/n 。分别编号为 Z_1, Z_2, \dots, Z_n 。子群中个体的个数是可以自行确定的。

(4)制定适应度函数,并求出每个个体的适应度

收稿日期: 2008-09-10

第一作者: 男,1985 年生,硕士生

* 通讯联系人

E-mail: huyj@mail.buct.edu.cn

函数值

采用如下介绍的确定适应度函数的思想: 适应度函数越大, 则表明其与均值越匹配, 即为非异动股; 适应度函数越小, 则表明其与均值差别越大, 即发生异动, 该股即为异动股。基于上述思想, 为了使模型简洁, 假定权值相同, 则适应度函数为

$$f_i(x) = \frac{a_i}{|x - a_i| + a_i}$$

其中 a_i 是股票的 60 d 平均价格。规则库的规模过大, 运算起来就会复杂得多, 为了使结果更容易显示特征, 在算法实现中假设规则库中只有一个均值作为特征值。

(5) 在每个子群中进行遗传操作

对每个子群 N/n 个个体分别进行遗传操作, 规定每组进行 m 代。采用和适应度函数成比例的选择方法, 单点交叉, 基本位变异。变异概率选为 0.1, 交叉概率取为 0.2。

(6) 把每个子群的最优个体组成一个新的种群, 并对其进行遗传操作

经过 m 代的遗传操作后, 把各子群合并到一起, 作为初始种群再进行 m' 代的遗传操作, 选择、交叉、变异的方法和前面相同。这时各子群的个体已基本收敛到最优解, 再经过 m' 代后, 就基本收敛到全局最优解了。基于并行遗传算法的入侵检测系统就完成了一次检测, 挑选出了当前活动的异动点。当某一特定期限内的异动点达到规定数目的时候, 则把该股票做为异动股。

(7) 确定终止条件

采用规定迭代代数的终止条件。

2 股票选择的算例

任选一只股票(上海电力, 600021)作为测试。取初始种群为 60 个个体(2008 年 4 月 24 日至 7 月 22 日共 60 d 的数据), 终止条件为迭代 100 代, 变异概率取 0.1, 交叉概率取 0.2, 适应度函数为

$$f_i(x) = \frac{a_i}{|x - a_i| + a_i}$$

其中 $a_i = 6.98$ (2008 年 6 月 10 日的 60 d 均值)。

把上述模型分别用遗传算法和并行遗传算法求解, 可以得到结果如表 1、表 2 所示。

从表 1 可以看出, 在子群数固定的情况下, 随着并行操作代数的不同, 基于并行遗传算法的检测系统的误报率也有所不同。基于并行遗传算法的入侵

检测模型的误报率从全局来看要低于基于遗传算法的入侵检测模型的误报率。

表 1 子群数固定时基于两种不同算法的入侵检测系统的误报率比较

Table 1 Comparison of the misinformation rates of the intrusion detection system based on two different algorithms when the subgroup is fixed

子群数	并行遗传代数	误报率/%	
		遗传算法	并行遗传算法
5	20	5	5
5	30	5	4
5	50	5	5
5	70	5	3
5	90	5	3

表 2 并行代数固定时基于两种不同算法的入侵检测系统的误报率比较

Table 2 Comparison of the misinformation rates of the intrusion detection system based on two different algorithms when the generation is fixed

子群数	并行遗传代数	误报率/%	
		遗传算法	并行遗传算法
2	70	5	2
3	70	5	3
4	70	5	2
5	70	5	5
10	70	5	5

通过表 2 看出, 子种群数的不同也会影响到基于并行遗传算法的入侵检测系统的误报率, 从全局来看基于并行遗传算法的入侵检测模型的误报率要低于基于遗传算法的入侵检测模型的误报率。

误报率为错误警报事件的数目与总的事件个数的比值

3 结束语

单固定子种群数和单固定并行遗传代数两种情况下, 从全局来看基于并行遗传算法的入侵检测模型的误报率要低于基于遗传算法的入侵检测模型的误报率。并且, 基于并行遗传算法的入侵检测模型的误报率直接受子种群数和并行遗传代数的影响。虽然子种群数和并行遗传代数是任意取定的, 但是要想得到最好的效率还要保证这两个参数取得适当。如何使这两个参数匹配得更好的方法还有待

进一步研究。

参考文献:

- [1] 郭佳. 基于 AHP 的长期投资优良股票选择模型[J]. 统计与决策, 2005(3): 27 – 30.
- [2] 杨春玲, 张传芳. 属性层次模型(AHM)在选股决策中的应用[J]. 大学数学, 2006, 22(5): 27 – 30.
- [3] 房毅宪, 王宝文, 王永茂. 基于遗传算法的动态递归网络的股价预测[J]. 燕山大学学报, 2007, 31(4): 359 – 363.
- [4] 陈光华. 人工神经网络在证券价格预测中的应用[J]. 计算机仿真, 2007, 24(10): 244 – 248.
- [5] 符海东, 王枫. 基于并行计算及人工免疫的入侵检测模型[J]. 计算机工程与设计, 2007, 28(11): 2547 – 2548, 2719.
- [6] 姚娅川, 王茜, 贾金玲. 虚拟仪器在实验教学中的应用[J]. 中国测试技术, 2003, 29(4): 50 – 52.
- [7] 李丹霞, 田云娜, 延飞波. 遗传优化的属性聚类在入侵检测中的应用[J]. 延安大学学报: 自然科学版, 2008, 27(1): 25 – 27.

Volatile stock options based on parallel computing and an intrusion detection system

XIE Xin HU YunJiao

(School of Science, Beijing University of Chemical Technology, Beijing 100029, China)

Abstract: Intrusion detection technology is used to treat stock options. An intrusion detection technology based on a parallel genetic algorithm is proposed. The data are randomly divided into several groups. Each group of data is tested, the most suspicious behavior characteristics in each group are merged together to make comparisons, and finally an option in the volatile stock is obtained. The data for Shanghai Electric Power stock from April 24 to July 22 2008 are analyzed using Matlab. Compared with the standard genetic algorithm, the results show that the improved algorithm reduces the false alarm rate and shortens the computation time.

Key words: parallel computation; genetic algorithm; stock options; intrusion detection system