

# 遗传算法在股市投资策略(战略)研究中的应用

李敏强 张俊峰 寇纪淞

(天津大学管理学院, 天津 300072)

**摘要** 阐述了遗传算法的基本理论, 论述了遗传算法用于股市投资策略(战略)研究的基本原理和方法, 给出了遗传算法用于股市投资分析(基本分析和技术分析)的数学模型, 并以上海股票市场为例验证了遗传算法用于投资策略(战略)研究的可行性。

**关键词** 遗传算法 股市投资策略(战略) 股市投资分析

## The Theory and Application of Genetic Algorithms in Investment Strategies of Stock Market

Li Minqiang Zhang Junfeng Kou Jisong

(School of Management, Tianjin University, 300072)

**Abstract** This paper puts forth the basic theories of Genetic Algorithms (GAs), then the principles and methods for applying GAs in the research of investment for investment analysis (Fundamental Analysis and Technical Analysis) are given. Finally, the feasibility of the models are proved in Shanghai Stock Market.

**Keywords** genetic algorithms; stock investment strategies; stock investment analysis

### 1 引言

随着经济体制改革的不断深入和对外开放步伐的加快, 我国金融和投资体制不断地向商业化、国际化方向靠拢。新的金融市场和金融工具不断涌现, 各种金融和法人机构有了更加广阔的投资领域; 另外, 随着各种保险基金、养老基金、共同基金以及其它机构投资者逐步进入金融市场, 都将面临着如何寻求一套合理而有效的交易策略问题, 来最大限度地回避风险、增加收益。由于各种金融市场更加动荡不安, 单纯地依靠个人的手工操作方式已很难适应当今的金融市场, 随之出现了许多计算机交易系统。据介绍, 在华尔街, 象 S&P500 企业股票指数上, 尽是较大的法人机构以程序进行买卖的天下, 从而使得金融交易与计算机的联系越来越紧密。近年来新出现的人工智能技术(如神经网络、混沌理论、遗传算法等)越来越多地应用于金融研究(如风险管理、交易系统、技术分析、资产组合、期权、期货等)领域。1995年4月, 在美国纽约召开了第一届有关金融工程与智能计算的国际会议。

自1975年美国的Holland教授出版第一本有关遗传算法(Genetic Algorithms, 简记GAs)的专著以来, 遗传算法得到了迅猛的发展。特别是近几年, 有关GAs的文章更是呈几何级数地增长。GAs已被广泛地应用于优化、搜索和机器学习等领域; 最近, 也逐渐被引入商业和金融领域, 出现了一些面向应用的商品化软件。例如: Evolver 被用于电子表(Excel Wangz 和 Resolve), 作为优化工具; Omega 预测模型系统, 广泛地应用于商业各领域(如市场营销、保险调查、信用管理等)。

本文于1996年12月30日收到

国家自然科学基金资助项目(79400013, 69574022)

天津市自然科学基金资助项目(F09523)

本文将对  $GA_s$  应用于金融投资战略(策略)领域, 进行一些理论探索。首先介绍  $GA_s$  的基本理论, 然后给出  $GA_s$  用于投资战略(策略)研究的算法模型, 最后以上海股市为例, 验证该模型用于投资战略(策略)研究的可行性。

## 2 遗传算法概述

遗传算法是一种有别于以往优化算法的一种新的优化、搜索算法, 它是建立在自然选择和群体遗传学机理基础上的随机、迭代、进化、并行搜索(寻优)算法, 模拟自然中的生命进化机制, 在人工系统中实现特定目标的优化。它不仅在实现有效的优化方面取得了极大的成功, 获得了广泛的应用, 而且在人工智能领域独树一帜, 被成功地应用于机器学习。遗传算法主要是 John Holland 教授和他在密西根大学的同事及学生发展起来的。他们研究的目的, 一方面是抽象出并严格解释自然系统中的自适应机制; 另一方面是设计具有这种机制的人工系统或软件。与传统方法相比, 遗传算法的主要特点是:  $GA_s$  使用问题参数的编码集, 而不是参数本身进行工作,  $GA_s$  是在点群中而不是在一个单点上进行寻优;  $GA_s$  仅使用问题本身所具有的目标函数进行工作, 而不需要其它任何选决条件或辅助信息;  $GA_s$  使用随机转换规划而不是确定性规则来工作。图 1 是  $GA_s$  的工作示意图。一般的遗传算法都包含三个基本算子:

- 1) 选择(Selection Operator)
- 2) 交叉(Crossover Operator)
- 3) 变异(Mutation Operator)

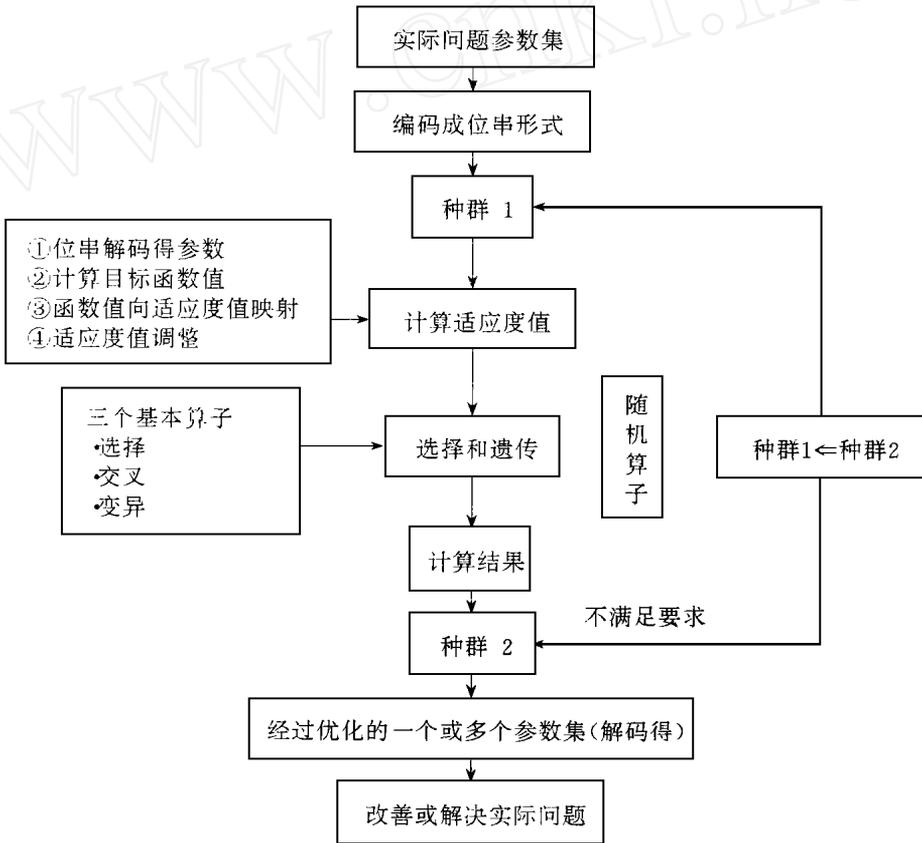


图 1 遗传算法工作示意图

遗传算法通过选择、交叉和变异三个算子的共同作用对染色体群施加生存压力, 使群体经过一系列迭

代得以朝着更好解的方向进化。GAs 的高效性和强壮性可由 Holland 提出的型式定理 (Schema Theorem) 和隐式并行性得以解释。型式 (Schema) 描述了染色体集合在给定位置的值所具有的相似性。

定义长度短的、确定位数少的、平均适应度值高的型式数量将随着代数的增加呈指数增长。这个结论称为型式定理或遗传算法的基本定理 (The Fundamental Theorem of Genetic Algorithm)。

遗传算法是通过定义长度短、确定位数少、适应度值高的型式的反复抽样、组合来寻找最佳点,称这些使 GAs 有效工作的型式为建筑块,是 GAs 构造答案的基本材料。但归根到底,要使 GAs 有效工作,必须按照 GAs 型式理论(或建筑块假设)根据具体问题设计合理的编码方案。

### 3 遗传算法与投资战略(策略)研究

遗传算法应用于金融投资领域只是最近几年的事。1992 年美国 St Mary's 大学的 Bauer, R. J. 和 G. E. Leipings 发表的《Genetic Algorithms and Computerized Trading Strategies》一文中,首次把遗传算法应用于金融投资领域;1994 年 Richard J. 和 Bauer, J. R. 合著了第一本有关遗传算法与投资战略研究的专著《Genetic Algorithms and Investment Strategies》;另外 Holland 和 Burke, G. 也发表了一些有关这方面的文章。遗传算法应用于金融投资研究可以说是方兴未艾。

由于遗传算法是一种弱方法,对问题域的信息要求很少,它的高效性和灵活性对于开发投资战略研究将非常有效。对于遗传算法在投资上的应用,其主要优点是允许你直接检验一个对于金融市场的一般的假设或原理。以一定前提条件为基础,例如我们假设证券收益和宏观经济数据相关,并进一步假设二者相关的形式,然后基于一般的结构去测试大量的可能交易规则,并从中搜索出有利的投资战略(策略)。

#### 3.1 遗传算法用于投资分析研究

自从股票、期货、外汇、期权、黄金和债券等金融市场出现以来,在这个投资和投机的市场中,由于价差所形成的利润之诱惑,许多人无不孜孜不倦地寻求必胜的契机和方式。证券市场中,投资者有短线、中长线和长线投资者之分,无论对于哪一种投资者,买卖股票都面临买什么股票和什么时候买卖股票的问题。投资者对这两种问题进行决策的主要依据就是基本分析 (Fundamental Analysis) 和技术分析 (Technical Analysis)。基本分析从基本的经济动态来观察,侧重点在于分析股票的内在投资价值,从而引导投资者选择投资对象;技术分析是从价格数字的统计分析来研究,投资者根据自己所选公司的股价变动趋势,绘制出各种图形来研究股市,预测股价的近期和未来动向,它尤其适用于短线投资者。

然而根据任何条件、因素所汇整而出的分析工具,都要受到其本身取样条件与参数设定上的限制,由于遗传算法的高效和强壮性,将其应用于投资战略研究将是一项很有意义的尝试。

##### 3.1.1 遗传算法用于基本分析

基本分析的目的在于分析股票的内在投资价值,从股票市场外部决定因素入手,并从各种外部因素与股票市场相互关系的角度进行分析。基本分析属于长期性质,即关心股票的收益,又关心股票的升值,同时通过对经济形势、行业形势和公司形势的分析,有助于投资者了解股票市场的发展状况和各种股票的投资价值,从而引导投资者正确地选择投资的对象。它注意的重点是未来的盈利和股息风险,以及未来收益的适宜贴现率等的预计。

股票投资基本分析是通过设立一系列指标,然后监察这些指标的变化来判别股票公司及其股票价格未来的前景。股票投资对象选择是一项复杂的多目标、多层次及多变量的决策活动,它不仅涉及的因素多、指标层次多、不确定性多,而且许多信息往往是描述性的,例如:整个国家的国民经济形势、上市公司所属行业形势、政治局势、经济政策、组织管理等;而且,有些甚至是模糊的和无一定评估标准的。因此,作为投资者选择股票的主要分析工具——基本分析,单纯地根据个别指标的高低,是很难正确地做出决策的。然而对于基本分析的众多指标和变量,去对股市进行分析、综合和检验,这一巨大的工作恐怕非人力所能及的,而 GAs 从大量不同准则的组合中选出最优准则,是非常迅速的。

基本分析包括对整个国民经济的分析、上市公司所属行业形势的分析和公司形势分析三个相互联系的层次,所需信息量相当巨大。由于整个经济形势和公司所属行业形势属于较长时间内相对稳定的因素,

故本文在分析中暂不考虑这两个方面的因素,而只考虑上市公司定期公布的财务报表——主要是资产负债表和损益表,并以财务比率组合来做为选择股票的准则。

对于  $GAs$  用于股市的基本分析,可采用如下一种简单的二进制编码方案,来代表不同的股票特征和测试复杂的股票选择准则的组合,其编码方案如表 1 所示。

对于上述每个基因位,可以用 0 或 1 表示。例如:基因位 1 为 1,可以代表一个可接受的低的债务水平(当然这个标准高低的具体数值需要事前确定);如果该值 0,则表示债务水平是不重要的。

$GAs$  另一个最重要的问题就是如何计算每条染色体的适应度值,本文用每年的组合投资收益在搜索最好准则为例,来讨论如果计算适应度值。假设有这样一条染色体,只有基因位 1, 12, 13 为 1,其它基因位均为 0,该准则表示:选择低债务水平、低  $p/e$  和低  $\beta$  值的股票。另外,进一步假设,用从 1985 年-1994 年的数据去检验这一准则。假设有 100 万元的初始投资,算法首先从 1985 年开始搜索所有的股票,发现满足这一准则的股票,等比例地投资于符合这一准则的股票组合。然后从 1986 年继续搜索所有的股票,搜索满足这一准则的股票,这样许多股票就要被卖出,而同时又将买进很多股票。这一计算过程一直持续到 1994 年末,在这一点上,计算所持有的所有股票的价值,这一价值就作为该染色体的适应度值。

上述过程将对群体中的每一条染色体重复进行,经过一代代适者生存的迭代,直到该过程收敛,这最后的结果即为最优或次优的准则。

### 3.1.2 遗传算法用于技术分析

技术分析是直接从股票市场入手,以股票价格的动态和规律性为研究对象,结合投资者心理和股票交易数量等市场因素的分析,以选择最佳的投资机会和方式,获取股票的短期收益。

自美国《华尔街日报》创办人查尔斯·道发明道琼理论以来,到目前为止,已出现过数百种技术分析方法来试图从不同的角度和侧面来把握股市波动的韵律。目前,股市上最常用的一些技术分析工具有:趋势线理论、甘氏角度线理论、波浪理论、移动平均线理论、成交量理论等;另外,还有许多技术分析指标,如:乘离率(BIAS)、相对强弱指标(RSD)、随机指标KD线等等,这些指标主要通过的一些数字指标来对股市进行分析,给出超买或超卖的信号,或是买卖讯号。

虽然技术分析方法大多是通过十几年乃至几十年的股市实践经验中总结出来的规律,但任何一种技术指标只是从不同的角度和侧重点对股市内在规律的不同程度的反映,或多或少地有其运用上的死角和不可知的变数。如果何有效地应用技术分析工具,把这些指标综合起来应用,来进行股市行情的研判,遗传算法将是很好的分析工具。

遗传算法用于技术分析研究,同样可以采用前面所给出的染色体编码方案。它的目的就是所有技术指标的组合中寻求一个或一组最优的组合(即交易策略)。适应度值计算根据交易策略进行买进和卖出操作的一定时期的收益作为该交易策略的适应度值。买进策略的染色体编码方案如表 2 所示:

其中各种技术指标的数值需要事先确定,对于卖出策略,其染色体也可以采用同样的编码方案。

## 4 遗传算法用于上海股市投资策略(战略)研究的算法实现及结果分析

### 4.1 数据选取和方法说明

由于上海股市已成为中国股票市场的代表,本文分析的数据取自上海证券交易所每天的成交行情,主要采用的是每天的收盘价、开盘价、最高价、最低价和成交量指标,选用的区间是 1992 年 12 月 23 日到 1995 年 12 月 5 日的上证综合股价指数。

另外,由于我国股市建立时间还很短,各种市场法规则还很不完善,会计制度、审计制度及上市公司信息披露方面与国际惯例还相差甚远,进行投资的基本分析,还有相当的难度。因此,本文将直接从股票市场本身入手,进行技术分析,这也跟中国股票市场目前技术分析盛行的现状相吻合。

表 1 遗传算法用于投资基本分析的染色体编码方案

基因位	股票特征
1	低的资产负债比率
2	高的流动比率
3	高的速动比率
4	高的应收帐款周转率
5	高的存货周转率
6	高的固定资产周转率
7	高的资产报酬率
8	高的资本报酬率
9	低的负债水平
10	快速的销售增长率
11	高的边际利润率
12	低的市盈率 $p/e$
13	低的 $\beta$ 值
14	增长的资产报酬率

表 2 遗传算法用于投资技术分析染色体编码方案

基因位	技术指标
1	MACD
2	RSI(相对强弱指标)
3	B I A S(乖离率)
4	KD(随机指标)
5	PSY(人气指标心理线)
6	ADR(涨跌比率)
7	AR(买卖气势)
8	BR(买卖意愿)
9	M T M(动量指标)
10	O S C(震荡量指标)
11	% R(威廉指数)
12	V R(成交量比率)
13	ADL(腾落指标)
14	O B O S(超买超卖指标)

## 4.2 算法的实现

遗传算法是弱方法, 仅需很少的问题域信息就能够有效地工作。但是在实际问题中, 为提高算法的效率, 我们应结合问题的特性选择合适的遗传算子和适应度值评价函数。

### 4.2.1 适应度值评价函数的选择和收益率的计算

我们采用前述计算方法作为问题的适应度值。另外, 在计算收益率时, 我们忽略了股票的交易费用, 采用下式计算收益率:

$$R_t = \frac{P_t - P_{t-1}}{P_{t-1}}$$

式中  $R_t$  为  $t-1$  时刻到  $t$  时刻的收益率;  $P_t$  为股票在  $t$  时刻的收盘价

### 4.2.2 染色体编码参数的选择

对于技术分析指标的选取, 由于不同的市场有不同的特点, 不同的股票有不同的属性, 实际上并没有统一的标准。目前, 股市上选用的参数大多从国外引入, 实际上对上海股市未必有很好的适用性, 每个股市应根据自身的规律去寻找适用的参数值。上海股市作为一个发展中的股市, 投机性很强, 股价波动剧烈, 技术分析参考统计分析的结果, 在相似的条件应服从一定的统计规律。在参数的选择上, 我们选取在沪市应用较普遍的一些技术指标和准则, 如: 相对强弱指标(RSI)、乖离率(BIAS)、指数离差指标(MACD)、威廉指标(%R)、随机指标KD线、人气指标心理线(PSY)和Granville准则。

### 4.2.3 遗传算子

#### 1) 选择(Selection)算子

对某一个体如何进行交叉, 在给定选择淘汰压力时是非常重要的, 我们选择适应度比例的概率表示留下子孙可能性的模型。某一个体  $i$ , 用各种选择方法选中的概率  $P_{select}$  可表示为:

$$P_{\text{select}} = \frac{f_i}{\sum f_i}$$

可在 0~1 区间内产生随机数作为各个个体的选择概率,选择概率大的个体参加交叉的机会多,所以它的遗传基因在群体中分布很广。

保存优胜者战略:这是把群体中适应度值最高的个体,直接保留到下一代的方法。采用一方法的优点是:最佳的解不会因交叉和突然变异被破坏。但是,由于优秀个体的遗传基因可能在群体中迅速扩大,有可能陷入局部解(local minima)的危险。

在计算中,我们采用适应度比例战略和保存优胜者战略相结合的选择算子,这样即可以保证使优秀的遗传基因在群体中迅速扩大,加强算法的收敛速度,同时,又可避免陷入局部解。

#### 2) 变异(Mutation)算子和交叉(Crossover)算子

变异算子是以一定的概率改变遗传基因的操作。设置过大的变异概率,型式(Schema)将会被破坏,从而使搜索会变成完全随机化;反之,如果变异率太小,则因无法引入最初遗传基因组合以外空间的新的基因,从而影响解的搜索空间,陷入局部解,本文选取变异概率  $P_m$  在 0.04~0.08 之间,可以得到较好的结果。

交叉算子是两种染色体相互交换的操作,计算中本文采用交叉位置只有一个的单一交叉或称单纯交叉。如:

```
A  1001|111  1001000
B  0011|000  0011111
```

### 4.3 结果分析

采用前述的方法,对上海股市从 1992 年 12 月 23 日至 1995 年 12 月 5 日之间的综合股价指数进行了计算,计算结果如表 3 所示。在计算中,把样本区间划分为两个阶段,先用 1992 年 12 月 23 日至 1994 年 12 月 30 日的数据计算交易准则(投资策略)的条件收益。从表 3 中可以看出,用技术分析方法对上海股市进行投资分析可以得到较高的条件收益,这也说明用技术分析方法对上海股市进行投资分析是可行的。但对各交易准则用 1995 年元月 2 日至 1995 年 12 月 3 日的数据进行检验,结果却不太理想,这可能是由于上海股市现在还处于发展的初期,市场结构经常变动,从而破坏了条件收益的稳定性;另一方面也说明本文所选取的参数体系有待进一步地精心地设计。同时,条件收益的存在,也证明了上海股票价格序列的非随机漫步特性,上海股市还不是一个高效率的市场。

## 5 讨论与展望

通过本文前述的工作,可以得出以下结论:

遗传算法用于投资策略(战略)研究初步证明是可行的,对于目前上海的股市,把遗传算法用于技术分析中,是能够获得较高的条件收益的。

但对于复杂股市的研究,以上这些工作仅仅是初步的,还可以从以下几个方面进行改进。

1. 随着股市的发展和规范,遗传算法应用于投资的基本分析将成为可行。

2. 遗传算法在技术分析上的应用,技术指标的选取还有很多方面需要仔细研究;同时为了提高其搜索效率,可以引入或设计新的遗传算子。

3. 本文得出的交易规则,较高的条件收益是否具有长期的稳定性,将随同中国股市的发展和规范化程度的提高,股市投资主体结构和行为的规范化、理性化等方面的变化而需要适时的调整。

4. 另外,基于遗传算法的机器学习系统(GA s-based machine Learning, GABML)如分类器系统(Classifier system)也被引入投资领域研究,由于分类器系统有很强的规则发现机制,把分类器系统引入投资战略(策略)的研究也不失为一有意义的尝试

5. 最后,遗传算法很容易即可被移植于其它金融市场(期货市场、黄金市场等),用于投资(策略)研究。

表 3 遗传算法用于上海股市技术分析结果

序号	染色体	适应度值	平均年收益率(93- 94 年)%	平均年收益率(95 年)%
1	11001100100000	269.48	84.74	0.29
2	01101000110100	222.29	61.14	-1.19
3	11101111011100	243.23	71.26	-0.77
4	10111010000110	269.48	84.74	0.29
5	10001000000110	250.26	75.13	0.25
6	11111110100110	278.26	89.13	0.29
7	00101000000000	243.23	71.62	-0.16
8	01101000101100	235.55	67.78	-0.16
9	00111000000100	250.26	75.13	-0.94
10	01100101100010	243.23	71.62	-0.16
11	01001000000101	235.55	67.78	-0.77
12	01111001001000	226.03	63.02	-7.33
13	00101101100001	243.23	71.62	-0.16
14	00101000101000	256.29	78.14	0.13
15	10001101110110	243.23	71.26	-0.77
16	10011000001001	236.15	68.07	-1.83
17	10001100010110	222.29	61.14	-2.50
18	11101001011110	258.42	79.21	0.25
19	10100010010110	250.26	75.13	-0.94
20	10101011001110	222.57	61.29	-7.33

## 参考文献

- 1 Bauer R J. Genetic Algorithms and Investment Strategies John Wiley & Son, Inc, 1994
- 2 Goldberg D E. Genetic Algorithms in Search Optimization and Machine Learning Reading, MA Addison-Wesley, 1989
- 3 郑超文. 股票、期货、外汇技术分析详解. 上海: 复旦大学出版社, 1993
- 4 陈明智. 股价(期货)分析预测学. 北京: 教育科学出版社, 1993
- 5 陈明智. 股文及期货技术分析总览. 台芝文化事业有限公司, 1992
- 6 张俊峰. 遗传算法在中国股市投资策略(战略)研究中的应用. 天津大学硕士论文, 1996