

基于遗传算法的股票市场选择模型

戎 容, 吴 萍

RONG Rong, WU Ping

华东师范大学 计算机科学与软件工程学院, 上海 200062

School of Computer Science and Software Engineering, East China Normal University, Shanghai 200062, China

RONG Rong, WU Ping. Stock market selection model based on genetic algorithm. Computer Engineering and Applications. Computer Engineering and Applications, 2016, 52(18): 167-172.

Abstract: In order to increase the income of investors in the stock market and solve the important issue about stock selection in the securities investment, a stock selection model based on genetic algorithm is proposed. This method uses financial indicators of listed companies as characteristics of samples. Firstly, in order to overcome the instability of K -means algorithm, it does the cluster analysis on stocks in the same field using the K -means algorithm based on genetic algorithm. Then stocks in the poor class should be removed. Secondly, it encodes screening conditions. An improved genetic operator is proposed to solve the premature convergence of traditional genetic algorithm in dealing with complex problems. Then it uses the improved genetic algorithm to find parameters of the stock selection model which can maximize the income of the investment in the stock market. Experimental results demonstrate that this method has a great effect on stock selection and is valuable to investors.

Key words: stock selection; genetic algorithm; cluster analysis; investment decision; combinatorial optimization

摘 要: 为提高投资者在股票市场的收益, 解决在证券投资中股票选择这一重要问题, 提出一种基于遗传算法的股票选择模型。算法以上市公司的财务指标为样本特征, 为克服 K -means 算法的不稳定性, 采用基于遗传算法的 K -means 算法对同一板块股票进行聚类分析, 剔除财务指标较差的一类中的股票。对筛选条件编码, 为解决传统遗传算法处理复杂问题时存在的过早收敛现象, 提出改进的遗传算子, 利用改进的遗传算法寻找使股票市场投资收益最大化的选股模型参数。实验结果表明, 该算法在股票选择上具有较好的效果, 可供市场投资者借鉴。

关键词: 股票选择; 遗传算法; 聚类分析; 投资决策; 组合优化

文献标志码: A **中图分类号:** TP301.6 **doi:** 10.3778/j.issn.1002-8331.1605-0313

1 引言

当今社会人们的理财投资意识日益增强, 且越来越多的投资者将眼光投向了股票市场。虽然股票可以给投资者带来可观的收益, 但投资者要想获得很好的投资回报, 就得利用合理科学的投资策略来选择股票进行投资^[1]。

股票选择从基本面而言, 就是对上市公司的内在价值进行评估^[2]。股票市场具有的长期记忆性使得可以通过数据分析找出股票价格或收益率的长期相关性, 同时股票市场具有非线性^[3], 应用智能算法可以提高分析的精确度和鲁棒性。现有的很多关于股市的研究主要是应用优化算法对股价和股市态势做出预测。如文献[4]

提出一种基于遗传算法的粗糙集属性约简方法和神经网络相结合的模型来预测股价。文献[5]提出基于离群特征模式的支持向量机模型来预测股价波动。这些研究限于对单个股票或大盘价格的预测, 对投资者的投资决策缺乏全面指导性。对于股票选择, 最传统的基于基本面分析的股票选择方法有欧奈尔基本面 CAN SLIM 法则、朱雀丁远指数中性策略等经典模型^[6]。这些模型大多是研究者通过对历史数据的分析和个人经验提出的, 虽然具有一定的效果, 但是股票筛选精度一般, 灵活性较差。文献[7]定义了股票稳定性值, 结合遗传算法和贪婪算法提出股票选择规划方法。此方法很好地规

作者简介: 戎容 (1991—), 男, 硕士研究生, 研究领域为智能优化算法, 现代软件技术, E-mail: rongr881@163.com; 吴萍, 通讯作者, 女, 副教授, 研究领域为智能优化算法, 现代软件技术。

收稿日期: 2016-05-23 **修回日期:** 2016-07-08 **文章编号:** 1002-8331(2016)18-0167-06

CNKI 网络优先出版: 2016-08-22, <http://www.cnki.net/kcms/detail/11.2127.TP.20160822.1011.046.html>

避了股票选择风险,但是没有实现收益最大化。文献[8]通过合理地选取贝叶斯分类器的参数对股票进行分类得出高回报股票组合,但是实验数据受大盘形势影响缺乏贝叶斯分类需要的独立性,算法预测精度一般。

本文提出一种改进的选股策略,利用基于遗传算法改进的 K 均值聚类算法和改进的遗传算法分析上市公司的财务指标,得出合理的股票选择模型。首先在大量股票样本的初步处理中,采用基于遗传算法的 K 均值算法,对财务指标进行聚类分析,从而过滤股票样本。然后利用改进的遗传算法,使用自适应的交叉算子与变异算子,来寻找最佳选股模型参数,从而找出真正具有投资价值的股票。

2 股票选择模型的建立

2.1 指标及样本的选择

基本面选股所涉及的因素主要是上市公司的财务指标,包括每股指标、盈利能力、成长能力、营运能力、现金流量、偿债能力和资本结构等指标。所以本文选取了较能反映这些能力的十项重要指标,如表1所示。

表1 财务指标表

指标类别	指标名称	指标性质
每股指标	每股收益/元	正指标
盈利能力	总资产净利润率/%	正指标
盈利能力	销售净利率/%	正指标
盈利能力	净资产收益率/%	正指标
成长能力	净利润增长率/%	正指标
成长能力	总资产增长率/%	正指标
营运能力	应收账款周转率/次	正指标
营运能力	存货周转率/次	正指标
偿债及资本结构	资产负债率/%	逆指标
现金流量	经营现金净流量对负债比率/%	正指标

板块是由具有共同特征的股票组成的群体。股票板块可以根据上市公司所属的行业、地域、特色题材等多种角度来划分^[9]。本文选取了发电设备板块的50家上市公司的股票作为样本。

2.2 模型求解策略

聚类是将对象集合分组为多个类,同一类中的对象有很大的相似性。目前国内股票市场根据上市公司的财务状况,公司股票一般分为:绩优股、一般股和垃圾股。所以本文采用基于遗传算法改进的 K 均值聚类算法将样本数据根据财务指标属性聚类为三类,即 $k=3$ 。剔除垃圾股类中的股票样本,保留绩优股和一般股类中的样本。这样先从大量数据中过滤掉不利于投资的股票,减少了之后股票选择的数据量,提高了选择的准确率和效率。

求得最佳的股票选择模型可以看作是一个复杂的组合优化问题,目标就是求得目标函数的最优解,遗传

算法具有很强的全局搜索能力,利用改进的遗传算法可以很好地求解这一问题。

2.3 模型的建立

2.3.1 模型参数

x_j 代表股票所属上市公司的第 $j(0 < j < 11)$ 个财务指标(对应如表1所示)。 x_{11} 代表该只股票的回报率。 S_i 代表第 i 只股票样本($0 < i < n+1$, n 为股票样本总数),股票样本是由上市公司的10个财务指标和它的股票回报率组成的一个11维空间,即 $S = \{x_1, x_2, \dots, x_{11}\}$ 。 $S_i[j]$ 表示第 i 只股票样本的第 $j+1$ 个财务指标($0 \leq j \leq 9$), $S_i[10]$ 表示第 i 只股票样本的回报率。 T 代表训练集,即 $T = \{S_1, S_2, \dots, S_n\}$, n 只股票样本组成模型训练数据集。 V 代表验证集,即 $V = \{S_1, S_2, \dots, S_n\}$, n 只股票样本组成验证模型数据集。 P 代表股票选择模型筛选条件,即 $P = \{p_1, p_2, \dots, p_{10}\}$, p_j 为对应表1所示第 j 个财务指标的筛选条件。 c_i 代表是否选择第 i 只股票,选择时等于1,不选时等于0。 P 作为筛选条件进行股票选择时,对于第 j 个财务指标,若为正指标,股票的这一项指标 x_j 需要不小于 p_j ;若为逆指标,股票的这一项指标 x_j 需要不大于 p_j 。如果股票的10项指标都满足条件,则该只股票被选择。反之只要有一项指标不满足,则该只股票不被选择。

2.3.2 目标函数与约束条件

定义原则1 证券投资理论中,投资策略是一系列投资组合的总称,分为积极和消极的投资策略^[10]。消极投资策略是指通过基本面分析和组合投资的方法来获得平均收益,即获得的平均收益越高,投资策略越好。

定义原则2 股票组合投资中股票的平均收益率是指总收益率除以股票个数,设定投资为等权组合投资。

基于以上原则,所以目标函数定义为选择股票的算术平均回报率最大值,其公式为:

$$\max \frac{\sum_{i=1}^n c_i S_i[10]}{\sum_{i=1}^n c_i} \quad (1)$$

约束条件为:

$$\begin{aligned} c_i &= 1, (S_i[j] \geq p_{j+1} (0 \leq j \leq 7) \& \& S_i[8] \leq p_9 \& \& S_i[9] \geq p_{10}) == 1 \\ c_i &= 0, (S_i[j] \geq p_{j+1} (0 \leq j \leq 7) \& \& S_i[8] \leq p_9 \& \& S_i[9] \geq p_{10}) == 0 \\ \sum_{i=1}^n c_i &> 1 \end{aligned} \quad (2)$$

取得最大平均回报率时的 P 即为最佳的股票选择模型参数。

3 基于浮点数编码的遗传算法及改进

3.1 染色体编码

染色体的编码方式有很多种,例如二进制编码、格

雷码、浮点数编码、符号编码^[11]。传统的遗传算法采取二进制编码以便于算法的交叉、变异操作。但是本文涉及的是多维数据操作, 二进制编码会增加染色体编码长度、人为离散化参数取值, 并且上市公司各项财务指标都是小数, 所以在本文中采用浮点数编码, 可以提高遗传算法的精度, 减少计算量。同时针对采用浮点数编码在处理一些复杂问题时会发生的停滞现象, 本文也提出了改进策略。

3.2 遗传算子

3.2.1 选择运算

选择运算把较好的个体按一定规则直接遗传到下一代群体中, 适应度高的个体更容易遗传到下一代群体中。本文使用轮盘赌选择法^[12], 各个体被选中的概率 Pr_i 与其适应度值 F_i 成正比。设种群规模为 N , 则:

$$Pr_i = \frac{F_i}{\sum_{i=1}^N F_i}, i = 1, 2, \dots, N \quad (3)$$

3.2.2 交叉运算

交叉运算是指把2个父代染色体的部分基因按某种方式交换重组, 从而生成新个体, 使得遗传算法的搜索能力大大提升。采用浮点数编码时, 交叉运算使用适合浮点数编码的算术交叉算子^[13]:

$$\begin{cases} x_A^{t+1} = ax_A^t + (1-a)x_B^t \\ x_B^{t+1} = (1-a)x_A^t + ax_B^t \end{cases} \quad (4)$$

其中, a 是 $[0, 1]$ 范围内的随机数。

3.2.3 变异运算

变异运算是对群体中的个体串的某些基因座上的基因值作变动。变异操作使遗传算法具有局部的随机搜索能力并且可维持群体多样性, 以防止出现未成熟收敛现象。均匀变异算子对于每个变异点从对应基因位的取值范围内产生一个随机数代替原有基因值^[14]。即

$$x = x_{\min} + r(x_{\max} - x_{\min}) \quad (5)$$

其中, r 是 $[0, 1]$ 范围内的随机数; x_{\max} , x_{\min} 分别是该基因位的数值上下限。

3.3 基于浮点数编码的遗传算法的改进

3.3.1 存在的问题分析

股票样本特征属性较多, 数据维数高。在进行股票筛选的初始种群中, 每一个个体 P 都是随机产生的, 由于各样本股票所属上市公司的发展水平各异, 各股票样本财务指标千差万别, 个体 P 作为筛选条件对股票样本 S_i 进行筛选时, 目标函数约束条件很难满足, 即能同时满足筛选条件 $S_i[j] \geq p_{j+1} (0 \leq j < 7) \&\& S_i[8] \leq p_9 \&\& S_i[9] \geq p_{10}$ 的股票非常少。若有股票满足个体, 此个体很有可能为非优解或局部最优解; 若无股票满足个体, 此时 $\sum_{i=1}^n c_i = 0$, 目标函数分母为零, 显然是无效解。所以初始群体中存

在大量的非优个体或者无效个体。应用通常的基于浮点数编码的遗传算法进行股票筛选的过程中, 大量非优个体或者无效个体不易被淘汰, 算法就存在容易过早收敛于局部最优解和向全局最优解收敛速度慢的问题, 对遗传进化过程产生不利影响。

3.3.2 改进的自适应交叉变异算子

针对应用遗传算法进行股票筛选存在的问题, 需要加强遗传算法的全局搜索能力, 在遗传进化过程中淘汰群体中过多的非优个体, 加快算法的搜索和收敛速度。在遗传算法中, 交叉运算实现了全局搜索能力, 变异运算实现了局部随机搜索能力, 因此本文提出改进的自适应交叉变异概率来提高遗传算法在股票筛选中的全局搜索能力, 改进基于以下原则:

定义原则3 在遗传初期群体中, 适应度较小的个体应该采取较大的交叉概率 P_c 和变异概率 P_m , 而适应度值最大的个体未必是全局最优的个体, 为了避免群体陷入局部最优, 适应度较大的个体也应该采取适当大的 P_c 和 P_m 。

定义原则4 在遗传后期加快收敛于群体最优解的过程中, 适应度较小的个体同样应该采取较大的 P_c 和 P_m , 而优秀的个体应该被保留, 个体的适应度越大, P_c 和 P_m 应该越小。

改进的自适应概率公式定义如下:

$$P_c = \begin{cases} P_{c1} + k_1 \frac{F_{\text{avg}} - F}{F_{\text{max}} - F_{\text{avg}}}, F < F_{\text{avg}} \\ P_{c1} - k_2 \frac{F - F_{\text{avg}}}{F_{\text{max}} - F_{\text{avg}}}, F \geq F_{\text{avg}} \end{cases} \quad (6)$$

$$P_m = \begin{cases} P_{m1} + k_3 \frac{F_{\text{avg}} - F}{F_{\text{max}} - F_{\text{avg}}}, F < F_{\text{avg}} \\ P_{m1} - k_4 \frac{F - F_{\text{avg}}}{F_{\text{max}} - F_{\text{avg}}}, F \geq F_{\text{avg}} \end{cases} \quad (7)$$

其中, P_{c1} 是基准交叉概率, P_{m1} 是基准变异概率, k_1 、 k_2 、 k_3 、 k_4 是调整系数, F 是个体适应度, F_{avg} 是群体平均适应度, F_{max} 是群体最大适应度, $iter$ 是总迭代代数, i 是当前代数。

4 基于遗传算法改进的K均值聚类

K 均值算法是一种基于梯度下降的贪心算法^[15], 传统的 K 均值算法存在一些弱点: 受到随机选取的初始聚类中心的影响, 算法很容易收敛于局部最优解^[16], 产生不准确甚至错误的聚类结果。利用遗传算法的全局优化能力来改进 K 均值聚类算法, 种群中的聚类中心通过选择、交叉和变异操作不断迭代进化, 直至收敛到全局最优个体, 从而克服传统 K 均值算法存在的局限

性并且提高聚类质量。

4.1 K 均值聚类算法

K 均值算法是典型的基于距离的聚类算法,即认为两个对象的距离越近,其相似性就越大^[17]。算法最终将所有对象分为 k 个簇,使簇内对象具有较高的相似性,而各簇之间相似性较低。

聚类问题可以描述为: n 个点 x_1, x_2, \dots, x_n , 按照相似性将其划分为 k 个集合 G_1, G_2, \dots, G_k , 满足:

- (1) $G_i \neq \varnothing, i = 1, 2, \dots, k$
- (2) $G_i \cap G_j = \varnothing; i, j = 1, 2, \dots, k; i \neq j$
- (3) $\bigcup_{i=1}^k G_i = \{x_1, x_2, \dots, x_n\}$

同时使得总的类内离散度和:

$$J_c = \sum_{i=1}^n \|x_i - c_j\|^2 \quad (8)$$

达到最小。其中 c_j 为第 i 个样本对应的聚类中心, $\|x_i - c_j\|$ 为第 i 个样本到对应聚类中心的欧氏距离。

聚类的划分采用最近邻法则,若 x_i, c_j 满足:

$$\|x_i - c_j\| = \min \|x_i - c_m\|, m = 1, 2, \dots, k \quad (9)$$

则 x_i 属于第 j 类。

K 均值算法的流程如图 1 所示。

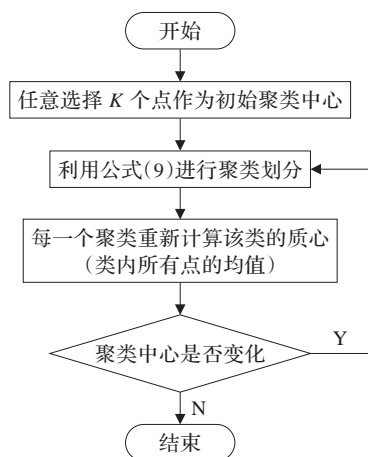


图1 K 均值算法流程图

4.2 基于遗传算法改进的 K 均值聚类算法设计

输入 股票样本数据集, 设置遗传算法参数: 聚类个数 $k=3$, 种群规模 $N=30$, 交叉概率 $P_c=0.6$, 变异概率 $P_m=0.05$, 迭代代数 $iter=100$ 。

输出 选取最终群体中适应度值最高的个体, 输出聚类结果。

步骤1 染色体编码。

通常聚类问题的数据集规模远大于聚类数目, 聚类算法的结果是求得聚类中心, 所以将各个聚类中心编码为染色体。每个聚类中心涉及 10 个财务指标属性, 则染色体长度为 $10 \times k$, 采用浮点数编码为: $\{X_1, X_2, \dots, X_k\}$,

其中聚类中心 $X_k = \{x_1, x_2, \dots, x_{10}\}$ 。这种编码缩短了染色体的长度, 能够快速高效地求解这类复杂的聚类问题。

步骤2 初始化种群。

从所有样本中随机抽取 k 个点作为 k 个聚类的中心, 重复进行 N 次, 共生成 N 个初始个体, 由此产生规模为 N 的第一代种群。

步骤3 按照图 1 的 K 均值算法流程对每个个体进行 K 均值聚类。

步骤4 计算群体各个体的适应度。

最准确的聚类结果要求类内相似度高并且类之间相异度高。现有大多数研究都只以较高的类内相似性作为衡量指标, 虽有不错的效果, 但在一些情形下效果不是很理想。本文采用类内相似和类间相异相结合的方式, 定义适应度函数为:

$$J = \frac{B_c}{J_c} \quad (10)$$

其中 $B_c = \sum_{1 \leq j < i \leq k} \|c_i - c_j\|$, B_c 为各聚类中心之间的距离和; J_c 为公式(8)表示的类内离散度和。 J_c 越小表示类内越紧凑, B_c 越大表示类间差异度越大, J 越大适应度越高, 聚类效果越好。

步骤5 按照公式(3)对种群进行选择操作、公式(4)对种群进行交叉操作、公式(5)对种群进行变异操作, 产生新一代群体。

步骤6 重复步骤3~步骤5, 直到最优个体的适应度和群体平均适应度趋于稳定或者迭代次数达到预设代数 $iter$ 。

5 基于改进遗传算法的股票筛选

输入 聚类后的股票样本数据集, 设置遗传算法参数: 种群规模 $N=50$, 基准交叉概率 $P_{c1}=0.6$, 基准变异概率 $P_{m1}=0.05$, 迭代代数 $iter=100$, 调整系数 $k_1=k_2=0.6$, $k_3=k_4=0.05$, 适应度函数参数 $h_1=100$ 和 $h_2=2$ 。

输出 选取最终群体中适应度值最高的个体, 即为选股模型的最优参数。

步骤1 染色体编码。

采用浮点数编码将 10 个财务指标参数作为筛选条件编码为染色体, 即 $\{p_1, p_2, \dots, p_{10}\}$ 。

步骤2 初始化种群。

从所有样本中随机抽取 N 个样本, 每个样本的 10 个财务指标组成一个初始个体, 由此产生规模为 N 的第一代种群。

步骤3 以群体中各个体为筛选条件进行股票筛选, 计算各个体的适应度。

本算法的目的是得出最佳的股票选择模型(即筛选条件), 所以选出的股票组合平均回报率越高, 模型效果越好, 适应度就应该越高, 因此可以选择满足筛选条件

表2 训练样本

股票名称	每股收 益/元	总资产净 利润率/%	销售净 利率/%	净资产收 益率/%	净利润增 长率/%	总资产增 长率/%	应收账款 周转率/次	存货周 转率/次	资产负 债率/%	经营现金 净流量对 负债比率/%	回报率/%
四方股份	0.022 3	0.172 7	1.523 3	0.28	- 44.810 0	9.944 7	0.243 0	0.405 1	30.191	- 0.150	- 0.10
东方电子	0.004 5	0.643 1	4.822 2	0.30	4.016 1	9.917 9	0.482 7	0.448 5	37.792	- 0.024	0.07
宝胜股份	0.035 0	0.256 4	0.887 7	0.88	4.293 5	7.892 3	0.701 0	2.272 4	70.208	- 0.025	0.09

的股票的平均回报率作为个体的适应度。但是考虑到选择运算采用轮盘赌方法,平均回报率有可能是负值,而选择运算采用轮盘赌方法不允许适应度出现负值,所以不能直接将平均回报率作为个体适应度,必须先进行转化。定义适应度函数为:

$$F = \frac{h_1}{h_2 - R} \tag{11}$$

其中 R 为平均回报率 ($R = \frac{\sum_{i=1}^n c_i S_i [10]}{\sum_{i=1}^n c_i}$), 当没有股票满

足筛选条件时 R 无效,这是最差的个体,所以此时 R 取一个较大的负值(如 - 99)使得 F 为一个极小的值,以保证在遗传算法的选择运算中这个个体不容易遗传到下一代中。 h_1 为常数 ($h_1 > 0$), h_1 越大不同回报率对应的适应度差异越大。 h_2 也为常数, h_2 大于各样本中回报率的最大值以保证 $h_2 - R > 0$, 从而适应度 $F > 0$ 。

步骤4 按照公式(3)对种群进行选择操作。

步骤5 根据公式(6)定义的改进自适应交叉概率,采用公式(4)定义的算术交叉方式对种群进行交叉操作,产生新一代群体。

步骤6 根据公式(7)定义的改进自适应变异概率,采用公式(5)定义的均匀变异方式对种群进行变异操作,产生新一代群体。

步骤7 重复步骤3~步骤6,直到最优个体的适应度和群体平均适应度趋于稳定或者迭代次数达到预设代数 $iter$ 。

6 实验与结果分析

根据财务报表统计50家上市公司2014年一季度的十项财务指标,统计2014年一季度第一个交易日的股票开盘价和2014年一季度最后一个交易日的股票收盘价格,计算其回报率,组成训练样本数据,数据如表2所示(限于文章的篇幅,仅给出了部分股票的数据)。资料来源于新浪财经。

训练样本分别使用未改进的和基于遗传算法改进的 K 均值算法进行聚类分析,共进行10次实验,运行情况如表3所示。

从表3可以看出,使用遗传算法改进的算法全部取得全局最优解,有效克服了传统 K 均值算法不稳定的缺陷。

表3 改进和未改进的 K 均值算法结果比较

聚类算法	最优解(类内 离散度和 J_c)	平均值(类内 离散度和 J_c)	最优解 次数
K 均值	162 361.38	168 180.09	6
遗传算法改进的 K 均值	162 361.38	162 361.38	10

训练样本通过基于遗传算法的 K 均值聚类算法聚类为3类,聚类结果如下:

第一类:{湘电股份,四方股份,泰豪科技,南洋股份,国轩高科,深圳惠程,太阳电缆,思源电气,永大集团,中元华电,九洲电气,南风股份,雅百特,茂硕电源,经纬电材}。

第二类:{通达股份,东方电子,宝胜股份,上海电气,东方电气,特变电工,卧龙电气,恒顺众昇,中超控股,万马股份,三变科技,上风高科,正泰电器,天顺风能,金风科技,汉缆股份,泰胜风能,摩恩电气,沃尔核材,奥特迅,金杯电工,金利华电,和顺电气,许继电气,麦迪电气,特锐德,中能电气,风范股份,平高电气,森源电气,北京科锐,易世达,科泰电源,置信电气}。

第三类:{华仪电气}。

运用PCA(Principal Component Analysis)算法将样本数据降维,在二维坐标系中,样本数据点分布及聚类结果如图2所示。

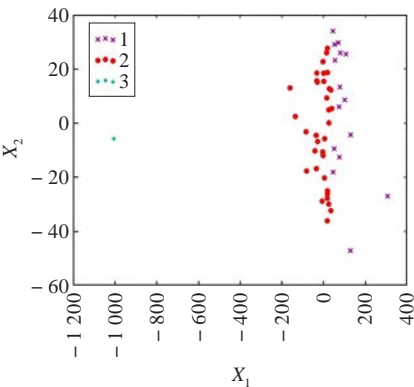


图2 训练样本聚类结果

分别计算三类股票的各项财务指标平均值,结果如表4所示。

从聚类结果可以看出,第一类股票的净利润增长率指标平均值为负值,其他各项正指标也较二、三两类有较大的差距。很明显这类股票在市场投资中为垃圾股,具有较大的投资风险,所以从样本中剔除第一类中的所有股票。

表4 各类股票财务指标平均值

分类	每股收益/元	总资产净利润率/%	销售净利率/%	净资产收益率/%	净利润增长率/%	总资产增长率/%	应收账款周转率/次	存货周转率/次	资产负债率/%	经营现金净流量对负债比率/%
第一类	0.02	0.29	4.88	0.45	-63.96	7.32	0.40	0.64	32.22	-0.08
第二类	0.07	0.89	7.69	1.44	39.65	16.25	0.56	0.83	38.82	-0.08
第三类	0.05	0.61	7.61	1.45	1 034.30	24.86	0.21	0.85	55.58	-0.05

经过第一步聚类筛选后剩余的训练样本分别使用未改进的遗传算法和改进的遗传算法进行基本面选股,共进行10次实验,运行情况如表5所示。

表5 改进和未改进的遗传算法结果比较

浮点数编码	最优解 (适应度)	最差解 (适应度)	最优解 次数	收敛到最优解的 平均迭代次数
未改进算法	408.163	289.855	3	56.8
改进的算法	408.163	327.860	7	38.1

从表5可以看出,使用改进的算法效果明显,最优解次数增多而收敛到最优解的迭代次数减少,说明改进算法更容易跳出局部最优解,收敛速度更快。

利用改进的遗传算法进行基本面选股,得出选股模型的最优参数为:(0.005 7, 0.074 6, 1.847 4, 0.599 9, 0, 8.382 4, 0.007, 0.008 5, 60.400 7, -0.001 6)。评价该股票选择模型是否有效,关键是选出的股票的回报率如何,通常的做法是以基准回报率、同期上证综指回报率和同期深证成指回报率为参照物,将选择出的股票的平均回报率与其对比^[18]。基准回报率是指样本中全部股票的平均回报率。图3显示了训练样本的股票选择结果。

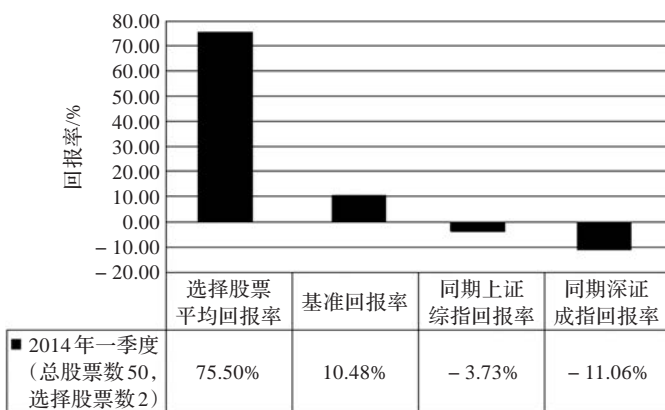


图3 训练样本选择结果

从图3可以看出,在训练样本中选择模型选出的股票组合的回报率明显超过基准回报以及同期上证综指和深证成指回报率。

为了进一步验证该模型的可靠性和准确性,训练样本外还统计了50家上市公司2014年二、三、四季度的相关财务指标和回报率,统计方法同上,组成验证样本数据。数据来源于新浪财经。图4显示了验证样本的股票选择结果。

从图4可以看出,在验证样本中选择模型依然有很

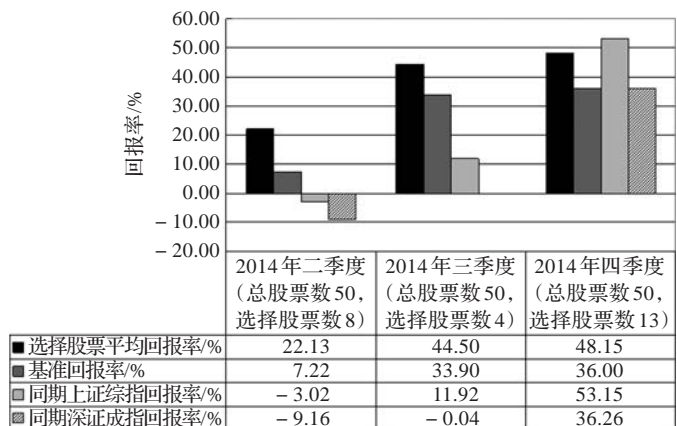


图4 验证样本选择结果

好的效果,选出股票的平均回报率全部高于基准回报率。与同期上证综指和深证成指回报率相比较,除了四季度数据略低于上证综指回报率,其他数据也都处于较高水平。

7 结束语

本文从股票的基本面出发,利用遗传算法进行股票选择。首先基于遗传算法的聚类分析,从上市公司的盈利能力、成长能力和营运发展水平等方面为初步评估相关股票的优劣提供了很好的依据。然后利用遗传算法较强的全局搜索能力解决最优解问题,得出最佳的股票筛选模型参数。实验证明,使用遗传算法进行基本面选股是切实可行的。由此可见,遗传算法能够为股票投资者提供科学的投资决策。

参考文献:

- [1] Yi Zuo, Kita E. Stock price forecast using Bayesian network[J]. Expert Systems with Applications, 2012, 39(8): 6729-6737.
- [2] 张宁致. 基于智能算法的股票市场决策模型[D]. 南京: 南京大学, 2013.
- [3] 王文成, 孙吉喆. 全球股票市场动态特征比较研究——基于STFIGARCH模型[J]. 经济管理, 2014(9): 110-118.
- [4] 张炜, 范年柏, 汪文佳. 基于自适应遗传算法的股票预测模型研究[J]. 计算机工程与应用, 2015, 51(4): 254-259.
- [5] 王浩, 陈娟, 姚宏亮, 等. 基于离群特征模式的股市波动预测模型[J]. 计算机工程与应用, 2014, 50(22): 243-249.
- [6] 李雪莲. 基本面与技术面相结合的A股量化选股模型设计[D]. 石家庄: 河北经贸大学, 2015.

(下转 178 页)

该方法可以有效地减少计算时间与迭代次数。通过实验证明该算法相比于传统方法能够得到比较理想的重建结果。

参考文献:

- [1] Lee E S, Kang M G. Regularized adaptive high-resolution image reconstruction considering inaccurate sub-pixel registration[J]. IEEE Transactions on Image Processing, 2003, 12(7): 826-837.
- [2] Irani M, Peleg S. Improving resolution by image registration[J]. CVGIP: Graphical Models and Image Processing, 1991, 53(3): 231-239.
- [3] Dongxing L, Jinhong H, Dong X. A novel restoration algorithm of the turbulence degraded images based on maximum likelihood estimation[C]//9th International Conference on Electronic Measurement & Instruments, ICEMI'09, 2009.
- [4] Rubinstein R, Zibulevsky M, Elad M. Double sparsity: learning sparse dictionaries for sparse signal approximation[J]. IEEE Transactions on Signal Processing, 2012, 58(3): 1553-1564.
- [5] Dabov K, Foi A, Katkovnik V, et al. Image denoising by sparse 3-D transform-domain collaborative filtering[J]. IEEE Transactions on Image Processing, 2007, 16(8): 2080-2095.
- [6] Cai J F, Chan R H, Nikolova M. Fast two-phase image deblurring under impulse noise[J]. Journal of Mathematical Imaging and Vision, 2010, 36(1): 46-53.
- [7] Zhang Hai, Xu Zongben, Wang Yao. $L_{1/2}$ regularization[J]. Science China: Information Science, 2010, 53(6): 1159-1169.
- [8] Babacan S D, Molina R, Katsaggelos A K. Total variation super resolution using a variational approach[C]//15th IEEE International Conference on Image Processing, ICIP 2008, 2008: 641-644.
- [9] Rabbani H, Vafadust M, Gazor S. Image denoising based on a mixture of Laplace distributions with local parameters in complex wavelet domain[C]//2006 IEEE International Conference on Image Processing, 2006: 2597-2600.
- [10] Patanaviji V, Jitapunkul S. An iterative super-resolution reconstruction of image sequences using fast affine block-based registration with BTV regularization[C]//IEEE Asia Pacific Conference on Circuits and Systems, APCCAS 2006, 2006: 1717-1720.
- [11] Dong W, Zhang L, Shi G, et al. Nonlocally centralized sparse representation for image restoration[J]. IEEE Transactions on Image Processing, 2013, 22(4): 1620-1630.
- [12] Burger W, Burge M J. Digital image processing[M]. [S.l.]: Springer, 2008.
- [13] Falas T, Stafylopatis A. Temporal differences learning with the scaled conjugate gradient algorithm[C]//Proceedings of the 9th International Conference on Neural Information Processing, ICONIP'02, 2002: 2625-2629.
- [14] Donoho D L, Huo X. Uncertainty principles and ideal atomic decomposition[J]. IEEE Transactions on Information Theory, 2001, 47(7): 2845-2862.
- [15] Farsiu S, Robinson D, Elad M, et al. Robust shift and add approach to super-resolution[C]//Optical Science and Technology, SPIE's 48th Annual Meeting, 2003: 121-130.
- [16] 陈志刚, 李贤, 井沛良, 等. 基于混合编码遗传算法的证据网节点可靠性评估[J]. 系统工程与电子技术, 2015, 37(7): 1697-1702.
- [17] 王康, 颜雪松, 金建, 等. 一种改进的遗传K-均值聚类算法[J]. 计算机与数字工程, 2010, 38(1): 18-20.
- [18] 闫仁武, 雷艳云, 任平, 等. 粒子群的K均值算法在电信客户细分中的应用[J]. 科学技术与工程, 2009, 9(20): 6042-6045.
- [19] 赖玉霞, 刘建平, 杨国兴. 基于遗传算法的K均值聚类分析[J]. 计算机工程, 2008, 34(20): 200-202.
- [20] 钱伟强. 一种基于改进粒子群和K均值结合的聚类算法[D]. 西安: 西安电子科技大学, 2011.
- [21] 钱颖能, 胡运发. 用朴素贝叶斯分类法选股[J]. 计算机应用与软件, 2007, 24(6): 90-92.

(上接 172 页)

- [7] 张金良. 基于优化遗传算法的股票选择规划研究[J]. 物联网技术, 2015, 5(9): 95-97.
- [8] 骆桦, 张喜梅. 基于贝叶斯分类法的股票选择模型的研究[J]. 浙江理工大学学报: 自然科学版, 2015, 33(3): 418-422.
- [9] 李庆东. 聚类分析在股票分析中的应用[J]. 辽宁石油化工大学学报, 2005, 25(3): 94-96.
- [10] 王军. 投资策略对股票市场有效性的判断[J]. 经济视角旬刊, 2011(11): 135-137.
- [11] 娄银霞. 基于遗传算法的人群行为模拟[D]. 长沙: 湖南师范大学, 2010.
- [12] 梁禹, 刘宇. 蜂群算法优化性能综合测试研究[J]. 计算机工程与应用, 2015, 51(21): 138-143.

- [13] 陈志刚, 李贤, 井沛良, 等. 基于混合编码遗传算法的证据网节点可靠性评估[J]. 系统工程与电子技术, 2015, 37(7): 1697-1702.
- [14] 王康, 颜雪松, 金建, 等. 一种改进的遗传K-均值聚类算法[J]. 计算机与数字工程, 2010, 38(1): 18-20.
- [15] 闫仁武, 雷艳云, 任平, 等. 粒子群的K均值算法在电信客户细分中的应用[J]. 科学技术与工程, 2009, 9(20): 6042-6045.
- [16] 赖玉霞, 刘建平, 杨国兴. 基于遗传算法的K均值聚类分析[J]. 计算机工程, 2008, 34(20): 200-202.
- [17] 钱伟强. 一种基于改进粒子群和K均值结合的聚类算法[D]. 西安: 西安电子科技大学, 2011.
- [18] 钱颖能, 胡运发. 用朴素贝叶斯分类法选股[J]. 计算机应用与软件, 2007, 24(6): 90-92.