

移动均线中最优步长测算方法及其应用

——基于均匀设计的遗传算法

宋光辉(博士生导师) 陈敏鹏 吴 栩

(华南理工大学工商管理学院 广州 510640)

【摘要】移动均线是技术投资中最为常用的技术指标之一,但移动均线最优步长的确定迄今仍无简单实用的方法。本文结合均匀设计和遗传算法,提出了一种移动均线最优步长的求解方法。将其应用于投资实践中,算法只需问题规模3.4%的计算量便能求出最优解,效率较高;同时,根据该方法得到的上证指数最优移动均线组合均能获得较高收益。其中,均线步长组合(1,21)受市场环境影响较小,且累计收益率高达487%。本文所提出的方法在投资实践中具有较高的可操作性。

【关键词】移动均线 遗传算法 均匀设计

一、引言

在非有效的市场中,许多投资者都迫切希望获得长期超额收益,由此衍生了许多主动型投资方法。这些投资方法可大致分为价值分析和技术分析,其中技术分析应用非常广泛。Taylor和Allen(1992)、Gehrig和Men Hoff(2006)等众多学者的研究结果表明,无论是个人投资者,还是外汇交易员、基金经理,都比较依赖技术分析来指导投资。

在众多的技术分析方法中,移动均线作为多种技术分析方法的基础,成了实务界最常用的技术分析方法之一,学术界也常用其来验证技术分析是否有效。尽管移动均线对理论界和实务界如此重要,但现有研究很少讨论移动均线的步长应该如何科学设置。正确地设定移动均线的步长,对于构建有效的均线系统至关重要。移动均线步长设置的优劣,一般以其盈利的大小作为标准(王兆军,2000)。那么,如何确定移动均线的最优步长呢?

对于移动均线步长的设定,实务界往往依靠经验选取,理论界同样尚无简单有效、切实可行的科学方法,屈指可数的现有成果仍存在一些缺陷。王兆军等(2000,2002)使用图示法、中位数法、EM算法及广义的均匀设计抽样确定移动均线的最佳步长,其中前三种方法取样比较麻烦,且使用穷尽法时试验次数非常多,计算量较大;最后一种方法是对总体的一个均匀设计抽样,将其用于推断最优的参数组合难免有些武断。

本质上,确定最优的均线步长,需要搜索大型的参数组合空间,并计算其对应的收益率,是一个复杂非线性的问题。遗传算法作为模拟自然选择和生物进化机制的寻

优算法,由于其寻优的方向并不依赖于目标函数的性质,且具有并行搜索的高效性,适应于大型复杂、非线性、不连续问题的求解。因此,使用其求解移动均线的最优步长有望解决现有方法繁琐复杂、不利于指导投资实践的问题。尽管遗传算法适用于各种优化的问题,但是,针对不同的问题需要设置合适的参数,才能提高遗传算法的寻优性能,避免出现早熟问题(即很快收敛到局部最优解)。

目前参数的设定基本上以试探法为主,许多论文更是直接设定参数,没有对参数的设定进行说明。由于遗传算法参数较多,范围较广,不可能逐个进行试验。考虑到遗传算法参数设定其实是一个多因素、多水平的问题,利用我国著名数学家王元、方开泰提出的均匀设计的思想,可以通过少量的试验确定参数最优取值,进一步提高算法的效率。

基于此,本文提出了移动均线最优步长的确定方法:基于均匀设计的遗传算法寻找最优均线步长组合。该方法的主要思想为:将均匀设计用于遗传算法的参数优化,通过较少的试验次数确定遗传算法各参数,从而优化遗传算法的性能。接着利用优化后的遗传算法运算,最终确定移动均线的最优步长。随后,本文以上证综合指数为样本,对该方法的实际应用情况进行了分析,以期为理论界和实务界提供参考。

二、文献综述

1. 遗传算法。遗传算法是一种模拟自然选择和进化机理的寻优算法。它的基本思想是,随机生成一定规模的群体,群体中的个体代表问题的一个解。针对每个个体计算它们的适应值,适应值越高则说明该个体代表的解越

优。通过选择、交叉、变异,将适应值高的个体保留下来,生成新一代的群体,适应值较低的个体将被淘汰。重复以上的步骤,个体的适应值会越来越高,最终收敛到最优解附近。

遗传算法有下列优点:①遗传算法并不依赖于目标函数的梯度方向进行寻优,而是根据生成个体的优劣来向最优解逼近。相比于传统的方法适用性较强,被广泛应用于各个领域。②遗传算法是在群体中进行并行搜索的,能够同时搜索解空间中的多组解,搜索速度快,效率较高,避免陷入局部最优解。该方法适合于大型、复杂、非线性问题的求解。

2. 遗传算法在交易策略中的运用。**Holland(1975)**首先提出遗传算法,经过**De Jong(1975)**、**Goldberg(1989)**的不断改进与完善,遗传算法模型得到更加广泛的运用。近年来,遗传算法也被运用于交易策略的优化,并取得了不错的成效。

Bauer(1994)对遗传算法进行了详尽的综述,并且将其应用到股票与债券市场交易策略的开发上。**Neely等(1997)**将遗传算法应用到六种外汇交易中,结果表明,遗传算法能够提供显著的投资收益。

而**Allen和Karjalainen(1999)**利用遗传算法,用美国**S&P 500**指数1928年到1995年的日数据,建立起技术交易规则。研究结果表明,考虑交易成本之后,遗传算法并不能获得超额收益。

Nunez-Letamendia(2002)将遗传算法运用到马德里的股票市场中,结果表明,遗传算法构建的技术交易策略能够提高模型的预测能力。

Akinori Hirabayashi(2009)在外汇市场上使用遗传算法建立多指标技术交易规则,结果表明,遗传算法能够适应市场环境的变化,取得较好的投资收益。

3. 均匀设计及其应用。均匀设计是一种试验设计方法,是由我国数学家王元和方开泰提出来的。均匀设计主要的目的在于从一组点集中取出少量的样本点,使得这些样本点分布较为均匀。它舍弃了正交设计的整齐可比性,只保留了试验点的均匀性,适用于多因素多水平的复杂试验问题。它可使每个因素每个水平只做一次实验,即可获得一个比较好的解。

遗传算法参数设定问题实质上是一个多因素多水平优化设计问题,由于参数空间,不可能通过遍历试验来进行参数的设置,因而可以用均匀设计的方法来对参数进行设定。这样,既可减少试验的次数,又能保证获得一个较好的解。

由于均匀设计方法简单易操作,不需大量试验便可达到理想的效果,因而在实际生活中得到了广泛的应用,在农业、化工、军工方面作用显著(**Chen X G, 2003**)。均匀

设计的方法也可广泛应用到优化算法中。

Leung和Wang(2000)设计了基于均匀设计的遗传算法,用来求解帕累托最优边界,利用均匀设计来优化不同适应值函数的权重并生成最初种群,取得了较好的效果。**Zhang和Sun等(2004)**使用均匀设计的方法来生成初始解,应用到分配算法中,提高了算法的效率。何大阔等(**2003, 2005**)基于均匀设计来生成遗传算法初始种群,并对遗传算法的参数进行优化,结果表明,利用均匀设计来优化遗传算法具有有效性、可行性。

张建方(2007)对试验设计的效率、最优实验次数和最佳效果做了研究,提出改进均匀设计的框架和学说。**张常利和杜永贵(2010)**将均匀设计与遗传算法应用到全维状态观测器的设计中,为遗传算法的实际应用提供了科学指导。

综观上述文献,大部分结果都表明,利用遗传算法来构建交易策略是有效的。但是,以上文献并没有讨论遗传算法各参数是如何设置的,以及设置是否科学。而均匀设计在提高优化算法的性能方面具有很好的效果。有鉴于此,本文结合均匀设计与遗传算法的优点,提出使用均匀设计来优化遗传算法的参数,构建交易策略,从而求出最优均线步长。

三、研究方法

接下来,笔者将详细地阐述构建基于均匀设计的遗传算法的方法,主要分成三个部分。第一部分,利用移动均线构建交易策略;第二部分,根据交易策略来设置遗传算法;第三部分,使用均匀设计来优化遗传算法的参数。

(一)移动均线构建交易策略

采用最简单的移动均线来构建交易策略。简单的移动均线系统至少包含快速均线(短周期)和慢速均线(长周期),分别以**MA(1)**和**MA(s)**表示,**s**表示步长为**s**的短周期,**1**表示步长为**1**的长周期,则有:

$$MA(s) = \frac{X_{t-s+1} + X_{t-s+2} + \dots + X_t}{s}, t = s, s+1, \dots, n \quad (1)$$

$$MA(1) = \frac{X_{t-1+1} + X_{t-1+2} + \dots + X_t}{1}, t = 1, 1+1, \dots, n \quad (2)$$

X_t 为第**t**天的日收盘价,则交易的规则可以定义为,当短期均线上穿过长期均线时,即为“黄金交叉”,发出买入信号;当短期均线下穿过长期均线时,即为“死亡交叉”,发出卖出信号。成交的价格按当日的日收盘价计算。用公式表示为:

$$F_t = \begin{cases} 1, MA(s)_t - MA(1)_t > 0 \\ 0, MA(s)_t - MA(1)_t \leq 0 \end{cases} \quad (3)$$

式(3)中, $F_t = 1$ 表示处于买入状态, $F_t = 0$ 表示处于

卖出状态。

判断均线的优劣一般根据它构建的策略的盈利大小来确定,另外也可以考察盈利概率、风险收益等指标。但考虑到多目标的复杂性,本文仅采用盈利的大小作为判断均线优劣的标准,这也符合普通投资者最自然的判断,与王兆军(2000)所用的标准相同。目标函数,即盈利的解析形式可定义为:

$$R_T = \prod_{t=1}^T (1 + F_{t-1} \cdot r_t) \cdot \{1 - \delta |F_t - F_{t-1}|\} - 1 \quad (4)$$

其中, $r_t = \frac{\text{close}_t}{\text{close}_{t-1}} - 1$, 为第 t 天的收益率。 δ 为交易费用,本文取 0.06%。

本文的研究目的就在于,找出最佳的步长组合 (s, l) , 使得 R_T 达到最大。下文将使用基于均匀设计的遗传算法对该问题进行求解。

(二) 遗传算法模型设计

遗传算法包括编码、初始化种群、计算适应值、选择、交叉、变异等步骤,使用遗传算法之前要先对各步骤进行设置。具体如下:

1. 编码。根据上面的交易策略,最优步长组合 (s, l) 含有两个参数,本文使用二进制进行编码,每个参数使用一个 8 位的二进制字符串来表示,参数范围设在 1~300 之间。如下图所示,前 8 个二进制位表示快速均线,即短周期参数 s , 后 8 个二进制位表示慢速均线,即长周期参数 l 。

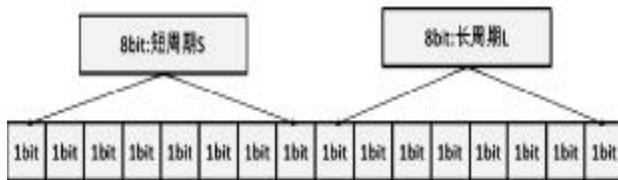


图 1 染色体编码图

2. 生成初始化种群。假设种群的规模为 N , 我们将随机生成 N 个个体, 个体的染色体结构如图 1 所示。种群规模 N 也是影响遗传算法性能的一个重要的参数, 将在后面通过均匀设计试验来设置。

3. 计算适应值。目标函数使用(4)式进行计算, 由于遗传算法要求适应值函数为非负的, 且数值越大说明解越优。考虑到(4)式有可能得到负数, 因而本文使用 Baker (1985) 提出的 rank 映射的方法来计算适应值。按照个体的目标值由大到小的顺序对它们排序, 并将它们线性映射到 0~2 之间。最大的数得分为 2, 最小的数得分为 0。

4. 选择。本文使用轮盘赌选择方法作为遗传算法的选择算子, 该方法是遗传算法中应用最早、最广泛的选择方法。设群体大小为 n , 其中个体 i 的适应值为 f_i , 则被选择的概率 $P_i = f_i / \sum_{j=1}^n f_j$ 。

该方法按照个体适应值的比例来选择, 跟轮盘赌中的转盘非常类似。当被选择的概率越大时, 个体的适应值越大, 被选择的机会也越多, 其基因结构被遗传到下一代的可能性越大。此外, 为了保留优秀的个体, 将适应值排在前 10% 的个体直接保留到下一代。将适应值排在后 $n\%$ 的个体淘汰掉, 随机生成新的个体作为下一代。该方法可以参照 Branke (1999) 和 Schoreels (2004) 的研究成果。淘汰率 $Pr=n\%$ 将在下文作为参数根据均匀设计试验得到。

5. 交叉和变异。本文使用单点交叉作为遗传算法的交叉算子。该方法对任意两对染色体, 在随机选择的交叉位之后的所有基因进行交换, 生成两个新的个体。

变异是遗传算法用来保持基因的多样性所设置的机制。通过将随机基因位的编码“0”转换成“1”或者“1”转换成“0”来实现。变异的概率一般都设置得比较低, 以免破坏原来优良的性状。

交叉和变异的概率分别设为 Pc 和 Pm , 也将通过均匀设计试验得到具体的值。

重复步骤(3)~(5), 迭代的代数设为 $Maxgen$, 运算 $Maxgen$ 次后将适应值最高的个体作为问题的解。遗传算法示意图如下:

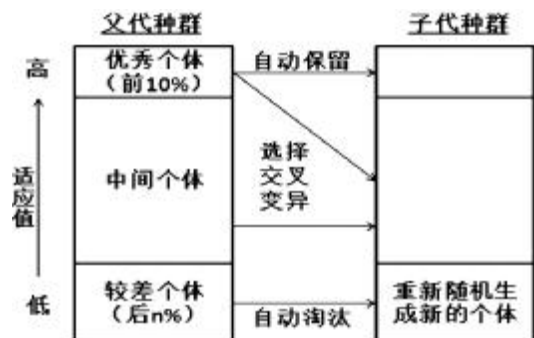


图 2 遗传算法示意图

(三) 利用均匀设计优化遗传参数

以上各步骤只是对遗传算法模型结构进行了设置, 为了使遗传算法的性能更优, 本文使用均匀设计来对参数的设置进行优化。需要优化的参数有: 种群的大小 N 、遗传迭代的次数 $Maxgen$ 、交叉概率 Pc 、变异的概率 Pm 、淘汰率 Pr 。每个参数取 11 个水平, 其取值范围分别是:

表 1 遗传算法参数取值范围

| 优化参数 | 取值范围 |
|-------------------|--------------|
| N (种群大小) | [20, 120] |
| $Maxgen$ (遗传迭代次数) | [30, 120] |
| Pc (交叉概率) | [0.35, 0.85] |
| Pm (变异概率) | [0.01, 0.21] |
| Pr (淘汰率) | [0.05, 0.55] |

根据均匀设计, 对于 5 因素 11 水平的试验, 使用 $U_{11}(11^5)$ 均匀设计表, 可得到如表 2 所示的试验方案:

表 2 使用 $U_n(11^5)$ 均匀方案设计表

| | N (种群大小) | Maxgen (遗传迭代次数) | Pc (交叉概率) | Pm (变异概率) | Pr (淘汰率) |
|------|-------------|--------------------|--------------|--------------|-------------|
| 方案1 | 20 | 30 | 0.45 | 0.09 | 0.35 |
| 方案2 | 30 | 50 | 0.6 | 0.19 | 0.15 |
| 方案3 | 40 | 70 | 0.75 | 0.07 | 0.5 |
| 方案4 | 50 | 90 | 0.35 | 0.17 | 0.3 |
| 方案5 | 60 | 110 | 0.5 | 0.05 | 0.1 |
| 方案6 | 70 | 20 | 0.65 | 0.15 | 0.45 |
| 方案7 | 80 | 40 | 0.8 | 0.03 | 0.25 |
| 方案8 | 90 | 60 | 0.4 | 0.13 | 0.05 |
| 方案9 | 100 | 80 | 0.55 | 0.01 | 0.4 |
| 方案10 | 110 | 100 | 0.7 | 0.11 | 0.2 |
| 方案11 | 120 | 120 | 0.85 | 0.21 | 0.55 |

按照表 2 的方案进行试验,为了使结果更加稳定可靠,避免随机因素造成的误差,对各方案独立进行 20 次试验,得到的结果见表 3。可知采用方案 10 可取得较优的结果,平均收益率达到 4.68,收敛的次数达到 13 次。

表 3 均匀设计方案对比

| | 平均适应值 | 最大适应值 | 最小适应值 | 收敛次数 |
|------|-------|-------|-------|------|
| 方案1 | 4.03 | 4.87 | 3.75 | 2 |
| 方案2 | 4.09 | 4.87 | 3.9 | 1 |
| 方案3 | 4.42 | 4.87 | 3.9 | 6 |
| 方案4 | 4.48 | 4.87 | 4.12 | 6 |
| 方案5 | 4.36 | 4.87 | 3.97 | 6 |
| 方案6 | 4.11 | 4.87 | 3.82 | 2 |
| 方案7 | 4.34 | 4.87 | 3.97 | 4 |
| 方案8 | 4.47 | 4.87 | 4.12 | 6 |
| 方案9 | 4.5 | 4.87 | 4.12 | 7 |
| 方案10 | 4.68 | 4.87 | 4.05 | 13 |
| 方案11 | 4.62 | 4.87 | 4.17 | 9 |

基于方案 10 的遗传算法性能较优,因此选用方案 10 的参数组合,将种群大小 N 设为 110,最大遗传迭代数设为 100,交叉概率 Pc 设为 0.7,变异概率设为 0.11,淘汰率取 0.2。理论计算量(110×100)仅为问题规模(300×300)的 12%。

四、实验设计及结果分析

1. 数据的选取。本文选用上证指数 2003 年 1 月 1 日至 2013 年 12 月 31 日的日收盘价数据作为分析的数据。其中 2003 年 1 月 1 日至 2004 年 12 月 31 的数据作为移动均线计算的基础数据,不参与交易。假设交易从 2005 年 1 月 1 日开始,交易区间为 2005 年 1 月 1 日到 2013 年 12 月 31 日。之所以选取这套数据,是由于上证指数是我国编制最早的市场指数之一,是投资者关注的重要指标,将其作为研究对象具有代表性。参考宋光辉等(2013)的分法,可将

2005 年 1 月 1 日至 2007 年 9 月 30 日分为牛市,2007/10/1 至 2008 年 12 月 31 日分为熊市,2009 年 1 月 1 日至 2013 年 12 月 31 日分为振荡市。数据覆盖了牛市、熊市和振荡市三种市场环境,较为全面。数据来源为恒生聚源数据库。

2. 假设条件。为了方便计算,本文假设:①交易信号发生后,以当日收盘价进行交易。②考虑交易成本,每次买入或卖出交易的佣金均为 0.06%。③由于研究对象是上证指数,故而不考虑股息与利息。④以指数的涨跌计算收益率,不使用现金交易。⑤不考虑卖空或融资融券的情况。

3. 实验结果及分析。以上完成遗传算法模型的设置以及参数的优化,下面将利用遗传算法来确定移动均线的最优步长。分别进行 20 次独立的运算,首先观察算法的收敛情况,选取其中一次运行情况为例。

由图 3 可以看出,遗传算法在 31 代的时候已趋向收敛,证明本算法效率较优,能迅速收敛到全局最优点。问题的规模为 90 000 个(300×300)参数组合(短期步长与长期步长分别有 300 个取值),而使用遗传算法只需 11 000 次(110×100)即可(110 个个体、100 次迭代),实际运算 31 代收敛,实际只需进行问题规模的 3.4% 的运算即可得到问题的解,也证明本文使用的遗传算法效率较高。

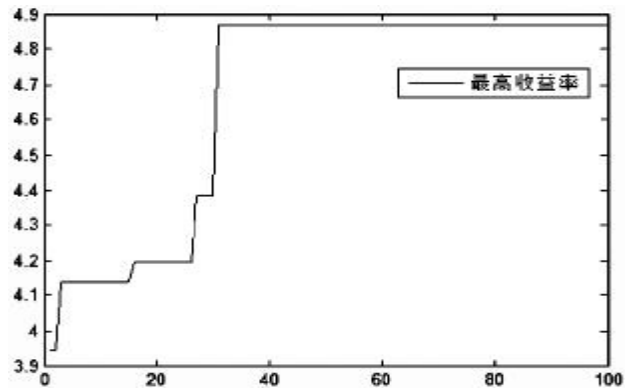


图 3 遗传算法收敛图

独立运行 20 次得到的移动均线最优的步长组合,结果如表 4 所示。

表 4 遗传算法实验结果

| 试验序号 | 最佳收益率 | 快速均线 | 慢速无线 | 试验序号 | 最佳收益率 | 快速均线 | 慢速无线 |
|------|-------|------|------|------|-------|------|------|
| 1 | 4.87 | 1 | 21 | 11 | 4.87 | 1 | 21 |
| 2 | 4.49 | 2 | 19 | 12 | 4.28 | 37 | 73 |
| 3 | 4.28 | 37 | 73 | 13 | 4.87 | 1 | 21 |
| 4 | 4.87 | 1 | 21 | 14 | 4.87 | 1 | 21 |
| 5 | 4.87 | 1 | 21 | 15 | 4.87 | 1 | 21 |
| 6 | 4.49 | 2 | 19 | 16 | 4.87 | 1 | 21 |
| 7 | 4.49 | 2 | 19 | 17 | 4.87 | 1 | 21 |
| 8 | 4.87 | 1 | 21 | 18 | 4.19 | 78 | 80 |
| 9 | 4.06 | 55 | 82 | 19 | 4.87 | 1 | 21 |
| 10 | 4.87 | 1 | 21 | 20 | 4.87 | 1 | 21 |

由表4可以看出,使用均线组合(1,21)、(2,19)、(37,73)、(78,80)均可取得较好的收益率。其中,均线组合(1,21)、(2,19)收益率最高,达到487%和449%。

下面进一步研究遗传算法考察的四组最优均组合的收益率情况,如图4、表5所示。由图4可以看出,对比上证指数买入并持有策略的收益率,由遗传算法得到的均线组合均取得了超额收益,而且随着时间的推移,累积收益率高达上证指数同期收益率的6倍。

分市场环境来看,表5表明,在牛市中,遗传算法得到的最优均线组合,基本上能够捕捉到牛市的趋势,收益率高达300%以上。其中(2,19)均线更是获得了超越市场的收益率。在熊市中,四组最优均线组合均获得超额收益,(78,80)均线只亏损了2.01%,超额收益高达65.2%。同时可以看出,长期均线组合(37,73)、(78,80)相对于(1,21)、(2,19)更能适应熊市。在振荡市中,四组最优均线也获得了较高的超额收益,并且短期均线组合相对于长期均线组合来说更能适应振荡市的市场环境。综合来看,(1,21)均线波动率较小,且盈利最高,对不同市场环境适应性也较好。

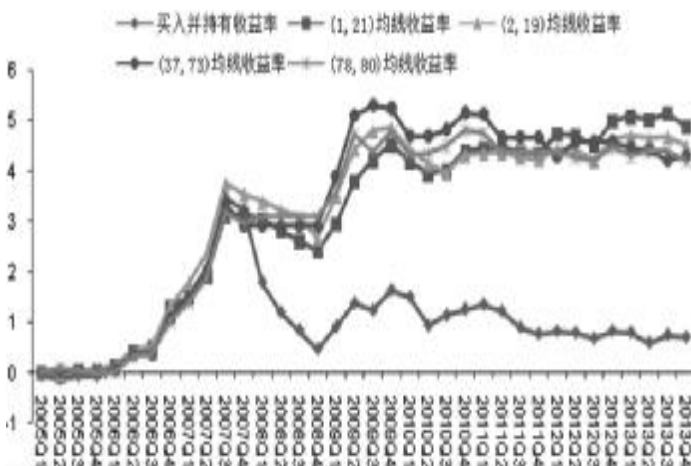


图4 各均线组合与上证指数收益率比较

表5 分市场环境下各均线组合与买入持有策略收益率比较

| 策略 | 全市场收益率 | 日收益波动率 | 牛市收益率 | 熊市收益率 | 振荡市收益率 |
|-----------|---------|----------|---------|----------|--------|
| 买入并持有 | 70.26% | 0.029 7% | 346.77% | - 67.21% | 16.21% |
| (1,21)均线 | 487.22% | 0.013 0% | 311.02% | - 17.16% | 72.46% |
| (2,19)均线 | 448.89% | 0.012 9% | 373.87% | - 22.86% | 50.15% |
| (37,73)均线 | 428.06% | 0.013 3% | 332.31% | - 9.37% | 34.78% |
| (78,80)均线 | 419.50% | 0.013 6% | 317.41% | - 2.01% | 27.01% |

五、结论

对于移动均线最优步长组合的确定,迄今并无简单实用的办法。基于此,本文提出基于均匀设计的遗传算法来确定上证指数移动均线的最优步长。结果表明,经过31次迭代运算,算法迅速收敛到全局最优解,实际只需原来

问题规模3.4%的计算量便能得到问题的解,效率较高。通过遗传算法得到的四组最优均线组合(1,21)、(2,19)、(37,73)、(78,80),均取得较好的效果,收益率高达上证指数的6倍。进一步分析发现,在牛市中,四组最优均线均能捕捉住上升的趋势,收益率与市场差不多。在熊市和振荡市中,四组最优均线均能获取超额收益,且其中两组长期均线(37,73)、(78,80)相对于两组短期均线(1,21)、(2,19)来说,更能适应熊市的市场环境。但相对来说,短期均线比长期均线更适合振荡市的市场环境。综合来看,(1,21)均线波动率较高,盈利最高,且在不同市场环境中适应性较强,是最优的均线。

笔者认为,基于均匀设计的遗传算法所选取的最优均线盈利能力强,适应性好,本方法切实有效。本方法简单高效,也可用于其他指标参数的优化。

【注】本文系国家社会科学青年基金(项目编号:12CJY006)、教育部人文社会科学研究青年基金项目(项目编号:13YJC790150)、教育部高等学校博士学科点专项科研基金新教师类资助课题(项目编号:20120172120050)、中央高校基本科研业务费专项资金(项目编号:2013ZB0016)的研究成果。

主要参考文献

1. 王兆军,曾渊沧,郝刚.移动平均线方法的最佳步长组合的确定.高校应用数学学报,2000;15
2. 王兆军,郝刚,曾渊沧.移动平均线的最佳参数组合.应用数学学报,2002;25
3. 何大阔,王福利.基于均匀设计的遗传算法参数设定.东北大学学报(自然科学版),2003;24
4. 张建方.关于试验设计的效率及有关问题.数理统计与管理,2008;27
5. 张常利,杜永贵.基于均匀设计的遗传算法及其应用.软件,2010;31
6. 宋光辉,詹素卿,刘广.激进配置型基金的行业配置效率考察.财会月刊,2013;8
7. Taylor MP, Allen H.. The Use of Technical Analysis in The Foreign Exchange Market. Journal of International Money and Finance,1992;11
8. Gehrig, T., Men Hoff, L.. Extended Evidence on the Use of Technical Analysis in Foreign Exchange. International Journal of Finance and Economics,2006;11
9. Franklin Allen, Risto Karjalainen. Using Genetic Algorithms to Find Technical Trading Rules. Journal of Financial Economics,1999;51
10. Nunez- Letamendia, L.. Trading Systems Designed by Genetic Algorithms. Managerial Finance,2002;8