第35卷 第188期
2014年3月
财经理论与实践(双月刊)
THE THEORY AND PRACTICE OF FINANCE AND ECONOMICS
Vol. 35 No. 188
Mar. 2014

·金融与保险·

基于灰色-ARIMA的金融时间序列智能混合预测研究

罗洪奔[1,2]

(1.中南大学 商学院,湖南 长沙 410083; 2.湖南大学 科学技术研究院,湖南 长沙 410082)*

摘 要:提出了一种基于灰色-ARIMA的金融时间序列智能混合预测模型。首先建立金融时间序列灰色预测模型,并采用PSO算法对灰色模型的三个参数进行优化;利用ARIMA算法对预测模型的残差进行分析,同时采用遗传算法对ARIMA的系数进行优化;最后用ARIMA的残差预测结果对灰色预测模型进行补偿。结果表明,以较好的精度拟合一段时期内MA<107的时间序列,预测误差控制在5%以上,与单纯的灰色预测算法和神经网络算法相比,在平均绝对误差、均方根误差和趋势准确率三项评价指标上,具有明显优势。

关键词:金融时间序列;灰色预测;ARIMA;PSO;遗传算法

中图分类号:TP273⁺.23 文献标识码:A 文章编号:1003-7217(2014)02-0027-08

一、引 言

金融市场属典型的复杂系统,呈现出较强的非线性和时变性特征[1]。其内部因素和位置变量之间的关系很难用准确的数学公式加以描述,难以建立完整的动力方程。因此,研究针对金融时间序列的分析和预测方法,具有十分重要的意义。

近年来,智能算法被越来越多地应用于金融时间序列的预测中,如神经网络、支持向量机等算法。这类方法一定程度上能解决市场非线性、非平稳性和高信噪比等问题,但由于训练速度慢,学习过程误差容易陷入局部极小点,很难保证学习精度。另外,这类方法只能保证在有限样本的情况下经验风险最小,而预测精度难以保证,泛化能力不高,应用范围受到了一定限制。

鉴此,本文从金融市场的特性和变化规律出发,提出一种基于灰色-ARIMA的智能混合预测方法。将传统方法与智能方法相结合,利用灰色理论建立金融时间序列模型。为了避免参数估计引入的误差,采用粒子群算法(简称PSO)对灰色模型参数进行寻优,同时利用ARIMA算法对预测模型的残差进行分析,以达到消除残差的目的;为进一步提升算法的精度,采用遗传算法对ARIMA的系数进行估计。

二、金融时间序列预测算法的提出

金融时间序列呈现出的波动性、非平稳性、周期性、样本少的特点对预测算法提出了较高的要求。邓聚龙(1982)提出的灰色系统理论认为,任何随机过程都是在一定时区范围内变化的灰色过程,通过细分处理,可归结为一种连续的、平稳的、动态的随机过程[2]。对金融时间序列这类"贫信息,小样本"问题,灰色系统理论有较好的分析效果,能够有效地分析金融时间序列变化地本质规律和变化周期。

传统的灰色预测方法在处理数据时都做了一定的条件假设,这些假设在对于金融时间序列的预测问题中不一定成立。同时金融活动中存在大量由突发因素造成的波动性和非平稳性变化,这些变化可能无规律可循,灰色模型难以辨识,使得辨识得到的金融时间序列模型与真实数据间存在较大的误差,成为提高预测模型精度的瓶颈[3]。因此,本文提出一种混合预测模型(如图1所示),该混合预测模型由灰色预测和残差预测两部分组成。

改进灰色预测模型,采用灰色理论的系统分析方法对原始金融时间序列数据进行辨识,从而逼近金融市场的变化规律。为了减少参数假设所引入的系统误差,将灰色模型进行扩展,采用PSO方法对模型参数进行优化,提升模型精度。

* 收稿日期: 2013-09-12
基金项目: 教育部博士点基金资助项目(20090161120044)
作者简介: 罗洪奔(1971—),男,湖南新化人,中南大学商学院博士研究生,研究方向:金融投资。

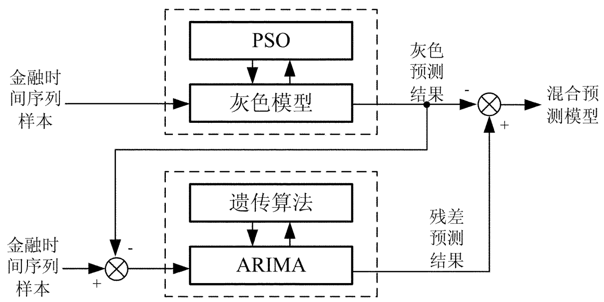

图1　预测算法原理图

基于ARIMA算法的残差预测模型，针对金融时间序列的波动性和非平稳性特点，将灰色预测模型与原始序列间的残差进行分析，建立残差预测模型，对灰色预测模型进行修正。为降低ARIMA参数预估引入的预测风险，引入遗传算法进行优化。

三、灰色预测模型

金融时间序列的精确预测是保证金融市场高效、稳定运行的基础。在灰色系统领域，金融活动可以看作在一定时区、一定范围内变化的灰色过程。其本质是：通过对历史金融数据进行累加生成，整理成规律性较强的数据序列，结合微分拟合法建立微分方程来描述生成时间序列的规律，实现对将来时刻的预测。GM(1,1)是最为传统的灰色预测算法，由于对于背景值选取做了一定假设和限制，造成预测误差偏大。针对这一问题，本文在GM(1,1)的基础上引入向量β，将其推广为GM(1,1,β)模型，并采用PSO算法优化该模型的关键参数，提高预测精度。

(一)GM(1,1,β)的实现

针对金融时间序列，GM(1,1)的具体实现如下：

设收集到的原始金融时间序列为：

$$x^{(0)}=x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(n) \tag{1}$$

$x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(n)$表示每个采样周期金融对象的经济指标值。

为了准确刻画金融活动的周期规律，减少非平稳性对金融时间序列变化本质规律分析的影响，进行一次累加生成(简称1－AGO)操作，可以在一定程度上降低数据的随机性和波动性，使其平稳化。对公式(1)进行一次累加生成得到的新序列如公式(2)所示。

$$x^{(1)}=x^{(1)}(1),x^{(1)}(2),\cdots,x^{(1)}(n) \tag{2}$$

其中，$x^{(1)}(k)=\sum_{i=1}^{k}x^{(0)}(i)\ (k=1,2,\cdots,n)$

为进一步分析金融对象的经济指标采样值在时间维度的内在联系，揭示其变化的周期性规律，在一次累加金融时间序列的基础上，建立一阶单变量线性灰色微分方程，并对其进行白化运算可得：

$$\frac{dx^{(1)}}{dt}+a\hat{x}^{(1)}=u \tag{3}$$

式中，u：灰作用量，反映了序列$x^{(0)}$的数据变化；a：灰系数，反映了序列$x^{(0)}$的增长速度。

由于金融时间序列为采样数据，并非连续变化的曲线，因此需对式(3)中微分项$\frac{dx^{(1)}}{dt}$采用公式(4)方法进行离散化。

$$\frac{\Delta x}{\Delta t}=\frac{x^{(1)}(k+1)-x^{(1)}(k)}{k+1-k}=$$
$$x^{(1)}(k+1)-x^{(1)}(k)=x^{(0)}(k+1) \tag{4}$$

式(3)中，背景值$\hat{x}^{(1)}$的选取决定了预测模型的准确性。传统的GM(1,1)方法中，采用公式(5)所示方法对公式(3)中背景值$\hat{x}^{(1)}$进行选取。

$$\hat{x}^{(1)}(k)=0.5x^{(1)}(k)+0.5x^{(1)}(k+1) \tag{5}$$

式中，$\hat{x}^{(1)}(k)$称为白化背景值，$x^{(0)}(k+1)$称为灰导数。

式(3)中，参数a和u可采用最小二乘法进行求解：

$$\hat{A}=[\hat{a},\hat{u}]^T=(B^TB)^{-1}B^{-1}B^TY \tag{6}$$

式中，常数a和u的估计分别用$\hat{a}$和$\hat{u}$表示。

$$B=\begin{bmatrix}-[x^{(1)}(1)+x^{(1)}(2)]/2 & 1\\ -[x^{(1)}(2)+x^{(1)}(3)]/2 & 1\\ \vdots & 1\\ -[x^{(1)}(n-1)+x^{(1)}(n)]/2 & 1\end{bmatrix},$$

$$Y=\begin{bmatrix}x^{(0)}(2)\\ x^{(0)}(3)\\ \vdots\\ x^{(0)}(n)\end{bmatrix}$$

微分方程(3)中代入a和u的估计值$\hat{a}$和$\hat{u}$，并求解可得到金融时间序列累积和离散解：

$$x^{(1)}(k+1)=\left[x^{(0)}(1)-\frac{\hat{u}}{\hat{a}}\right]e^{-\hat{a}k}+\frac{\hat{u}}{\hat{a}} \tag{7}$$

$k=0,1,2,\cdots$

最后利用式(7)进行累减还原，得到原数列的金融时间序列预测模型如公式(8)：

$$\hat{x}^{(0)}(k+1)=(e^{-a}-1)\left[x^{(0)}(1)-\frac{\hat{u}}{\hat{a}}\right]e^{-\hat{a}k} \tag{8}$$

$k=0,1,2,\cdots$

由式(8)根据金融时间序列的前k个数据对未来第$k+1$时刻的取值进行预测。

在 GM(1,1)模型的建立过程中,灰微分方程初始条件为:$\hat{x}^{(1)}(1)=x^{(1)}(1)=x^{(0)}(1)$,导致最终拟合金融时间序列必然通过坐标平面上的点$(1,x^{(0)}(1))$。在实际金融活动中,金融序列波动性和非平稳性较强。随着金融活动的开展,时间的推移,这一假定不可能永久成立,因此,基于 GM(1,1)模型对金融时间序列预测精度受到了极大的限制。为此,引入向量 β 按照公式(9)对背景值重新计算,从而得到推广后的 GM(1,1,β)模型。

$$\hat{x}^{(1)}(k)=\beta x^{(1)}(k)+(1-\beta)x^{(1)}(k+1)\quad 0\leqslant\beta\leqslant 1 \tag{9}$$

根据式(9)与式(8)计算可知,$\beta=1/a-1/(e^{a}-1)$。

改进的灰色预测模型可以通过在式(6)按照式(9)对背景值进行取值,并进行参数求解得到。

(二)基于 PSO 的改进模型

GM(1,1,β)方法对采用传统 GM(1,1)方法在辨识金融时间序列过程中,存在假设初始条件的不足,但 β 的取值成为金融时间序列预测建模的关键;影响金融市场的干扰因素较多,灰色模型中传统采用最小二乘的 $\hat{a}$ 和 $\hat{u}$ 确定方法,往往会出现较大残差。

不难看出,如何调节 β,$\hat{a}$ 和 $\hat{u}$ 成为 GM(1,1,β)的关键。因此,为了进一步提升金融时间序列的预测精度,引入智能方法动态调节 β,$\hat{a}$ 和 $\hat{u}$。

PSO 是一种基于种群并行优化进化的算法,具有易实现、计算效率高等优点[4]。同时在陷入局部最优问题上有所改善,因此,对于 β,$\hat{a}$ 和 $\hat{u}$ 的选取,采用 PSO 算法。

针对金融时间序列 GM(1,1)预测模型参数优化问题,PSO 在实现时,β,$\hat{a}$ 和 $\hat{u}$ 的组合采用粒子位置来表示,β,$\hat{a}$ 和 $\hat{u}$ 所有组合构成了粒子的解空间,大量粒子以当前的位置最优的粒子为参照更新自身速度矢量 V,调节自己的位置,在解空间中搜索更优的位置。通过大量粒子不断地迭代更新,逐步在空间中聚合到最优的位置,从而找到全局最优粒子,即 β,$\hat{a}$ 和 $\hat{u}$。

1. 目标函数。如何定义位置最优,决定 PSO 算法中粒子如何在解空间中搜索。针对金融时间序列的预测问题,精度是追求的目标,因此针对 PSO 算法,其目标函数应考虑模型 GM(1,1,β)整体准确性,即预测模型输出的未来时刻金融时间序列值应与实际的值偏差最小。因此,采用公式(10)所示的性能指标。

$$\begin{cases}|e(t)|=|v_s(u)-v|\\ J=\int_0^{t_\infty}|e(t)|dt\end{cases} \tag{10}$$

式中,t_∞ 为稳态时间,$e(t)$为预测模型输出的未来时刻金融时间序列值应与实际的值偏差。$v_s(u)$为该组参数下金融时间序列灰色模型拟合的预测值,v 为实际的测量值。PSO 的学习目标是使 J 最小。

2. 粒子位置更新。GM(1,1,β)的三个关键参数 β、$\hat{a}$ 和 $\hat{u}$ 由 PSO 的粒子位置表示,即粒子当前位置 $X=[\beta,\hat{a},\hat{u}]$。采用实数编码的方式进行组合。粒子依据其当前的飞行速度 V 矢量在解空间中进行移动,即按照一定策略更新 β、$\hat{a}$ 和 $\hat{u}$ 的组合,从而搜索最优位置。PSO 中最优位置表示以目标函数定义作为评价指标,使该指标最小的 β、$\hat{a}$ 和 $\hat{u}$ 编码组合。

$P_i^l=[P_{i1}^l,P_{i2}^l,\cdots,P_{iD}^l]$表示单个粒子历史最优位置,$P_q^l=[P_{q1}^l,P_{q2}^l,\cdots,P_{qD}^l]$表示群体历史最优位置。第 i 个粒子在 $l+1$ 次迭代计算后,采用如下方法更新位置:

$$V_{id}^l=\Omega^l V_{id}^l+s_1\alpha_1(P_{id}^l-X_{id}^l)+s_2\alpha_2(P_{qd}^l-X_{id}^l) \tag{11}$$

$$X_{id}^{l+1}=X_{id}^l+\xi^l V_{id}^l \tag{12}$$

式中,$d=1,2,\cdots,D$;ξ^l 为收缩因子,用于均衡算法的全局搜索和局部搜索能力;α_1、α_2 分别表示认知学习速率和社会学速率;s_1、s_2 为学习因子;Ω^l 为惯性系数,用于对控制速度进行抑制。

Ω^l 的选取决定了 PSO 的优化效果,当 Ω^l 较小时,算法具有较强的局部搜索能力。反之,算法全局搜索能力较强。由于 GM(1,1,β)的 β、$\hat{a}$ 和 $\hat{u}$ 构成的三维解空间搜索范围较大,为了保证 PSO 算法的收敛性和搜索范围,采用公式(13)实现 Ω^l 的自适应调整。

$$\Omega^l=\Omega_{max}-\frac{(\Omega_{max}-\Omega_{min})(H-H_{aver})}{H_{max}-H_{min}} \tag{13}$$

式中,H 微粒的适配值,Ω_{max}、Ω_{min} 分别代表惯性系数的最大值和最小值;H_{max},H_{min} 表示微粒群中适配值的最大取值和最小值;H_{aver} 表示每代微粒的平均适配值。在 β,$\hat{a}$ 和 $\hat{u}$ 优化问题中,$\Omega_{min}=0$,$\Omega_{max}=5$,$H_{max}=0$,$H_{min}=10$,H_{aver} 需要在实际迭代过程中计算得到。

3. 算法流程。利用 PSO 对 β、$\hat{a}$ 和 $\hat{u}$ 进行寻优,步骤如下:

步骤 1. 确定最大迭代次数,同时对粒子群的随机位置、速度以及大小进行初始化;

步骤 2. 根据优化函数对每个粒子的适应值进行计算；

步骤 3. 将每个粒子的适应值与整个粒子群的最优位置 X_{gbest} 对应的适应值进行比较，并更新 X_{gbest}；

步骤 4. 粒子的位置和速度根据式(12) 进行更新，同时根据式(13)更新 Ω^{l}；

步骤 5. 若满足最大迭代次数条件，转向 Step 6，否则转向 Step 2；

步骤 6. 将 X_{gbest} 输出，X_{gbest} 所标示的编码转义为 β、$\hat{a}$ 和 $\hat{u}$ 的取值，寻优操作结束。

四、基于 ARIMA 的残差消除

金融时间序列的变化呈现出较强的非平稳性和波动性，对于其中周期波动导致的非平稳性，利用 GM(1,1,β) 中的 1-AGO 处理后，依然能够在辨识结果中体现。但是根据前文的分析，实际的金融活动会受到多种突发因素的影响，这些突发因素造成的波动性和非平稳性无法被 GM(1,1,β) 方法辨识，因此，简单地利用 GM(1,1,β) 对金融时间序列进行辨识得到的模型，往往与真实数据间存在较大的残差，而对应的残差所组成的时间序列也呈现出较强的波动性和非平稳性。如何预测残差，进而对 GM(1,1,β) 辨识得到的金融时间序列变化规律进行校正，成为算法的关键。

ARIMA 是一种分析有效的金融数据的工具，针对金融时间序列突发波动性和非平稳性较强特点，能够达到不错的平稳化效果。因此，采用 ARIMA 算法对 GM(1,1,β) 的残差进行预测，并对 GM(1,1,β) 校正，从而达到降低波动性和非平稳性对预测结果影响的目的。

ARIMA (Integrated Autoregressive Moving Average Model) 模型称为求和自回归滑动平均模型，是 Box 和 Jenkins 于 1970 年提出的时间序列分析方法。将该方法应用到对 GM(1,1,β) 的残差预测问题上，基本思路是：对于非平稳的残差时间序列，用若干次差分使其成为平稳序列，再用 ARMA(p, q) 模型对该平稳序列建模[5]，之后经反变换实现对残差时间序列的辨识。

ARIMA(p,d,q) 过程可用如下数学公式进行表示：

$$\begin{cases} \Phi_p(B)\nabla^d x_t = \Theta_q(B)\varepsilon_t \\ E(\varepsilon_t)=0, Var(\varepsilon_t)=\sigma_\varepsilon^2, \\ \qquad E(\varepsilon_t\varepsilon_s)=0, s\neq t \\ E(x_s\varepsilon_t)=0, \forall s<t \end{cases} \tag{14}$$

式中，x_t 为原序列；ε_t 为白噪声序列；B 为延迟算子；p 为自回归模型的阶数；q 为移动平均模型的阶数；d 为差分阶数；Φ_p 为 p 阶自回归模型，即 AR(p)模型的自回归系数多项式；Θ—q 阶动平均模型型，即 MA(q)的系数多项式。

从公式(14)不难看出，ARIMA(p,d,q)模型本质上就是由差分运算与 ARMA(p,q)模型的组合而成。ARIMA 算法的实现过程有以下五个步骤：

步骤 1. 对原始残差时间序列进行差分平稳化处理和零均值化处理；

步骤 2. 采用自相关分析和偏自相关分析的方法，分析残差时间序列的截尾性，从而确定辨识结果的结构和阶次；

步骤 3. 根据辨识结果结构和阶次，采用智能方法确定模型的参数；

步骤 4. 原时间序列与检验模型之间的误差，若模型无法满足误差要求，则回到步骤 Step 2 重新进行模型结构辨识；

步骤 5. 利用所建立的合适的模型导出残差预测模型，应用于 GM(1,1,β) 的残差预测。

(一)差分定阶

由于金融市场的突发性和非平稳性被遗留在 GM(1,1,β) 的残差时间序列中，该时间序列样本的自相关和偏相关函数往往难以快速趋于 0。因此 ARIMA 中无法省略差分项，必须进行差分处理，然后对差分之后的序列进行考察，看其是否还含有趋势项，如果有，则继续差分。考虑残差数据的可处理性，实用中通常最多差分两次，差分的次数即 ARIMA(p,d,q) 中的 d 值。

(二)模型结构确定

设差分处理后的 GM(1,1,β) 残差时间序列为 $x_1, x_2, \cdots x_N$，是平稳正态序列 $X_t=\{x_t, t\in N\}$ 的一段样本值，N 为样本长度，则：

$$\hat{\mu}=\bar{x}=\frac{1}{N}\sum_{i=1}^{N}x_i \tag{15}$$

自协方差函数估计如公式(16)：

$$\hat{r}_k=\hat{r}_{-k}=\frac{1}{N}\sum_{t=1}^{N-k}x_t x_{t+k} \tag{16}$$

$$k=1,2,\cdots M, \hat{\sigma}^2=\hat{r}_0$$

自相关函数估计为：$\hat{\rho}_k=\hat{\rho}_{-k}=\dfrac{\hat{r}_k}{\hat{r}_0}$。

偏相关函数 a_{kk} 的估计由方程组求得：

$$\begin{bmatrix}\hat{a}_{k1}\\ \hat{a}_{k2}\\ \cdots\\ \hat{a}_{kk}\end{bmatrix}=\begin{bmatrix}1 & \hat{\rho}_1 & \cdots & \hat{\rho}_{k-1}\\ \hat{\rho}_1 & 1 & \cdots & \hat{\rho}_{k-2}\\ \cdots & \cdots & \cdots & \cdots\\ \hat{\rho}_{k-1} & \hat{\rho}_{k-2} & \cdots & 1\end{bmatrix}=\begin{bmatrix}\hat{\rho}_1\\ \hat{\rho}_2\\ \cdots\\ \hat{\rho}_k\end{bmatrix} \tag{17}$$

当 N 充分大时,$\hat{\rho}_k$ 渐近于正态分布,即:

$$\hat{\rho}_k \sim N(0,\frac{1}{N}(1+2\sum_{k=1}^{q}\hat{\rho}_k^2)) \tag{18}$$

对某个选定的 q,若 $\hat{\rho}_k(k=q+1,q+2,\cdots,M)$ 满足 $|\rho_k|<\frac{2}{\sqrt{N}}$ 的个数95%以上,则可以认 $\hat{\rho}_k$ 是 q 步截尾的,具有截尾性。同理可以判断 a_{kk} 的截尾性,如果 $\hat{\rho}_k,a_{kk}$ 具有截尾性。

根据上述截尾特性分析方法,由于残差时间序列保留了实际金融时间序列的突发性,因此往往具有 $a_{kk}>0,\hat{\rho}_k>0$ 的特点,结合分差分析,需要采用ARIMA(p,d,q)的结构。

(三)ARMA(p,q)模型的定阶

ARMA(p,q)的定阶是决定算法精度的重要因素。需要考虑在残差时间序列辨识过程中,不引入过多的虚假参数估计的前提下,使残差平方和尽量小。学者赤池引入以下准则函数:

$$AIC(n,m)=ln\,\hat{\sigma}_a^2+2(n+m+1)/N \tag{19}$$

若 $AIC(n,m)=\min\limits_{0\leqslant n,m\leqslant L}AIC(n,m)$,选定的 n,m 可作为AMAR模型的阶数 p,q。由于AIC方法不具有相容性,采用BIC准则函数为:

$$BIC(n,m)=log\,\hat{\sigma}_a^2+\frac{(n+m+1)log\,n}{n} \tag{20}$$

(四)基于遗传算法的系数优化

当ARIMA(p,d,q)阶数确定后,$p+q$ 个多项式系数的取值成为ARIMA预测精度的关键。金融活动中常常出现突发波动的情况,采用传统的极大似然法的ARIMA的系数估计必然导致辨识得到的残差时间序列模型存在较大的偏差。GA(遗传算法)算法具有搜索快速、易实现和计算效率高的优点[6],非常适合用在ARIMA系数估计中。

针对残差时间序列的ARIMA(p,d,q)辨识过程中系数估计问题,GA在具体实现过程中,$p+q$ 个多项式系数采用实数编码的方式,组合为染色体,$p+q$ 所有组合构成了染色体的多项式系数解空间,每个多项式系数对应为染色体的一个基因片段,大量染色体组成了候选解种群。种群中根据适应度(即目标函数),选择基因组合最优的染色体,向下一迭代周期的染色体个体进行遗传,同时采用变异机制使种群中的基因发生变化。通过大量染色体不断的遗传变异,逐步在空间中得到基因组合最优的染色体,即最合适的 $p+q$ 个多项式系数。

1. 实现步骤。利用GA对多项式系数优化设定的实现步骤如下:

步骤1. 首轮随机产生 n 组ARIMA系数染色体(每个染色体由 $p+q$ 个基因组成),组成初始种群;

步骤2. 个体适应度采用适应度函数进行评估,以寻求在最合适某阶段的金融市场环境中,最接近GM(1,1,β)残差的ARIMA系数解作为最优个体,并将其保留至下一代种群;

步骤3. 当前种群中所有个体进行交叉操作;

步骤4. 检查当前种群是否符合收敛条件,如果收敛条件满足,则结束优化,否则转向(Step1)。

2. 适应度函数。在GA算法中,ARIMA残差预测模型的准确性指标在设定个体适应度时需要重点考虑,即为ARIMA残差预测模型的输出与GM(1,1,β)的实际残差值偏差最小。因此,采用式(21)所示的性能指标。

$$F_{exact}=|J_s(u)-J| \tag{21}$$

式(21)中,$J_s(u)$ 为以ARIMA残差预测模型得出的残差预测值,J 为GM(1,1,β)的实际残差检测值。

3. 变异策略。目前常用的变异策略中,均匀交叉变异破坏模式的概率较大,但搜索到的多项式系数组合较多;点式交叉变异破坏模式的概率较小,但搜索到的多项式系数组合较少。本文中解空间维度适中,采用均匀交叉会增加算法的复杂度,因此,采用点式交叉策略。

为了提高遗传算法的最优多项式系数组合搜索性能,增加变异强度[7],采用式(22)两点交叉策略。

$$\begin{vmatrix}01/110/110\\10/101/001\end{vmatrix}\Rightarrow\begin{vmatrix}01/101/110\\10/110/001\end{vmatrix} \tag{22}$$

其中,0、1表示二进制基因串序,对于两个基因按照将公式(22)进行两次分组交换,即完成两点交叉。

4. 杂交策略。目前常用的杂交策略中,单点杂交和两点杂交两种杂交策略对基因的改变较少,搜索能力较弱。因此,采用插入式多点交叉策略:首先,对父代多项式系数组合成的染色体进行随机选取,确定基因片段和插入点;然后,将该基因相应的插入点插入另一个染色体基因片段,形成新的基因;最后,删除新基因中重复的基因片段,从而得到变异后的子代染色体。

五、实证分析

为了验证算法的有效性，分析算法效果，采用 2003 年 9 月 1 日 ～ 2013 年 2 月 27 日上证综合指数的每天的 MA(10) 数据作为样本金融时间序列，其中前 1500 组数据作为学习样本，后 550 组数据作为实验样本，进行实证分析。

(一) 灰色单项预测模型仿真分析

首先采用归一化方法对样本进行初步处理，然后采用灰色预测模型的预测算法对学习样本进行分析，同时利用PSO对灰色模型β、$\hat{a}$和$\hat{u}$三个数据进行优化设定。实验表明，当 PSO 迭代计算 11 次后，模型精度提升不大，此时$\beta = 0.432$，$\hat{a} = 4.725$，$\hat{u} = 8.235$。原始序列和预测结果如图 2 所示。

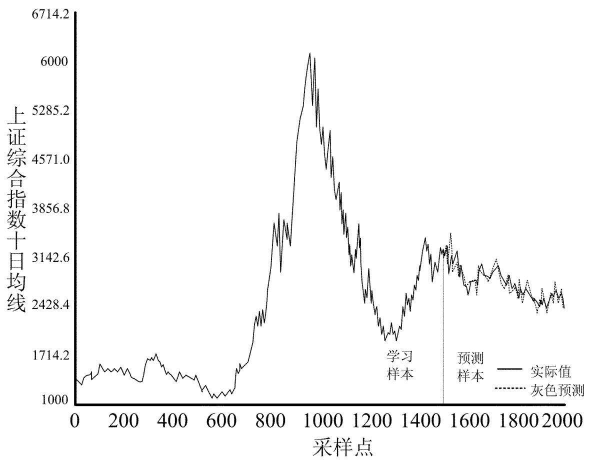

图 2　MA(10) 预测结果

为了说明实验的有效性，在相同样本数据的条件下，采用单纯灰色预测模型和以上算法，进行实验比照，如图 3 所示。样本两种算法较少时误差均偏大，随着样本数量的增加，基于 PSO 优化的灰色模型优势体现了出来，误差明显下降。

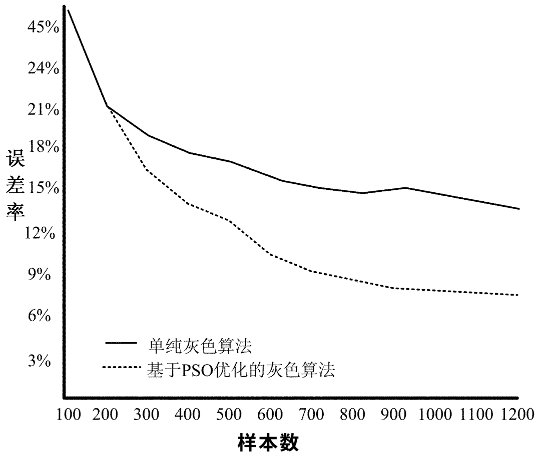

图 3　预测模型误差对比曲线

(二)SVM 单项预测模型仿真分析

由于金融时间序列的非平稳性，灰色预测模型难以做到残差消除，采用基于 GA 修正的 ARIMA 方法对灰色模型的残差进行预测。

根据灰色模型的残差时间序列，进行 ARIMA 计算，确定差分阶 d = 2；根据自协方差函数和自相关函数分析，选用 ARIMA 结构；根据 BIC 原则，选取 p = 2，q = 2，构成 ARIMA(1,2,2) 结构。因此 GA 优化过程中，共需要确定 4 个多项式系数。根据实验，当 GA 迭代计算 8 次后，基本达到收敛位置，残差预测结果如图 4 所示。

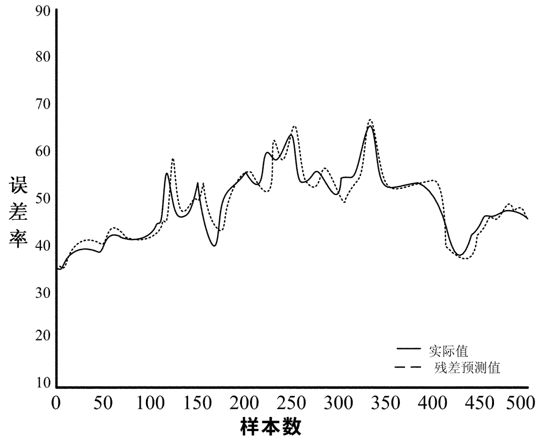

图 4　残差预测模型误差对比曲线

图 4 可以看出，对残差预测时，基于 GA 优化的 ARIMA 方法能够根据不多的样本数据进行有效辨识，建立精度相对较高的预测模型。其中残差平均误差仅为 6.3％，这主要是因为金融活动反应在时间上存在一定的滞后性，因此，辨识得到的曲线与实际曲线存在一定的相位差，导致误差偏大，在实际应用中，可通过相位补偿的方法进行修正。

利用基于 GA 修正的 ARIMA 方法计算的残差预测值对灰色模型进行补偿，进一步地降低了预测误差，平均误差控制在 5％ 以内。

(三) 对比实验结果分析

为了说明基于灰色－ARIMA 的金融时间序列智能混合预测算法的优势，在选用相同的样本和预测样本条件下，分别采用 GM(1,1,β) 方法、应用较多的神经网络方法以及以上算法进行对比试验，其结果分别如图 5、图 6、图 7 所示。

从图 5 中不难看出，由于 GM(1,1,β) 采用了 1－AGO 处理，一定程度上弱化了原始金融数据的非平稳性，能够较为准确地刻画金融时间序列的变化

趋势,但是由于对突发因素造成的波动性和非平稳性分析能力较弱,GM(1,1,β) 预测结果与原始数据相比曲线更为平滑,误差较大。

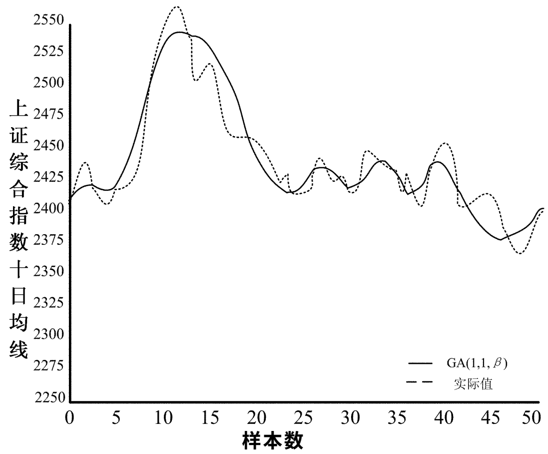

图 5　GM(1,1,β) 预测结果

神经网络算法能够对一定时间范围内金融时间数据的因果关系和内在联系进行辨识,如图 6 所示。神经网络预测曲线与 GM(1,1,β) 相比,在金融时间序列的局部波动的辨识效果上更好,但是由于神经网络算法对于数据的依赖性较强,针对金融数据突发波动性的刻画有一定的局限性,且预测精度与样本数量成正比,因此精度仍然有待进一步提高。

从图 7 不难看出,由于采用了基于灰色—ARIMA 智能混合预测算法,能够将灰色算法对主要趋势的辨识能力和 ARIMA 算法对非平稳性数据的处理能力相结合,使得两种单项预测模型能够实现优势互补,最大限度地避免了 GM(1,1,β) 和 ARIMA 各自的缺陷,大大提升了模型的预测精度。

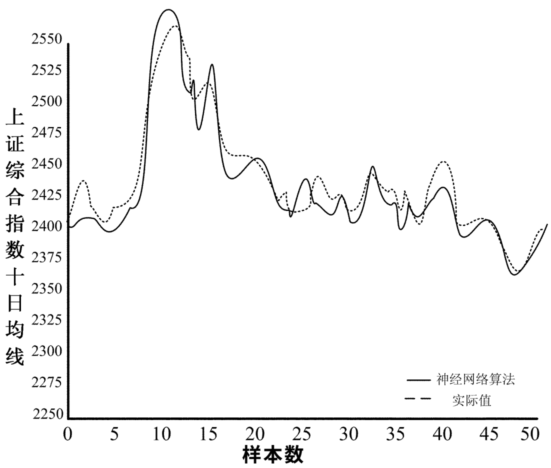

图 6　神经网络预测结果

为了验证灰色-ARIMA 算法的有效性,引入股票预测领域常用的平均绝对误差 MAPE、均方根误差总评价指标 RMSE、趋势准确率指标 F 等三种方法对模型进行评价。

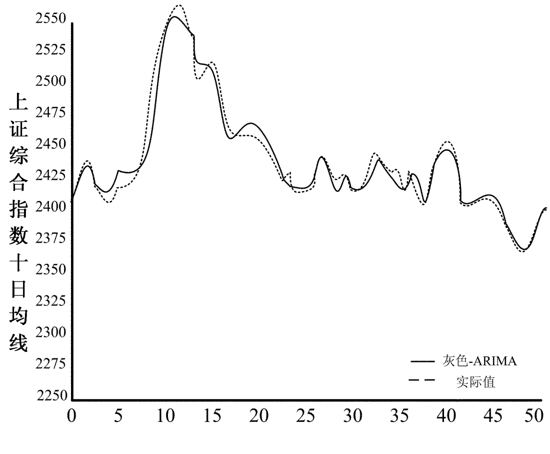

图 7　灰色-ARIMA 预测结果

表 1　算法指标对比结果

	MAPE	*RMSA*	*F*
GM(1,1,β)	7.64%	56.5	125.1
神经网络	6.75%	50.4	87.3
灰色-ARIMA 算法	1.75	23.3	25.5

从表 1 的结果可以看出,对 50 个交易日的预测数据进行分析,本文提出的组合预测算法在 *RMSE*、*MAPE* 和 *F* 三项指标远小于其他两种预测算法,表明灰色-ARIMA 算法能够学习和跟踪股市变化情况,具有很好拟合能力。灰色-ARIMA 算法在 50 次趋势预测中,正确趋势 45 次,准确率为 90%,对未来具有很好的判断能力,从而说明灰色-ARIMA 算法对上证综合指数的趋势有很好的跟踪能力。

六、结论

以上在对金融时间序列自身特点充分分析的基础上,针对金融市场中存在的干扰因素众多,关系复杂,呈现波动性、非平稳性,提出了一种灰色—ARIMA 的金融时间序列的智能混合预测模型。从实证分析的结果上看,本文算法能以较好的精度拟合一段时期内金融时间序列数据,由于采用了残差消除和智能优化方法,模型预测精度比单纯的灰色预测算法有了较大提升,从而提供了一种新的分析金融时间序列的有效途径。

参考文献:

[1] 倪丽萍,倪志伟. 一种基于趋势分形维数的股指时间序列相似性分析方法[J]. 系统工程理论与实践,2012,21(9):76-78.

[2] 毛丽,左青松,刘冠麟. 车用三效催化转化器剩余寿命的非等间隔灰色预测[J]. 中南大学学报(自然科学版),2012,16(4):59-61.

[3] 周国雄,吴 敏. 基于改进的灰色预测的模糊神经网络预测[J]. 系统仿真学报,2010,22(10):68-71.

[4] 魏德敏,文星宇. 基于混合 PSO 算法的桁架动力响应优化[J]. 振动与冲击,2011,22(5):92-95.

[5] 黄安强,肖进,汪寿阳. 一个基于集成情境知识的组合预测方法[J]. 系统工程理论与实践,2011,(1):123-127.

[6] A. Azaron, C. Perkgoz, M. Sakawa. A genetic algorithm approach for the time-cost trade-off in PERT networks[J]. Applied Mathematics and Computation, 2005, 168 (2): 1317-1339.

[7] 李松,刘力军,解永乐. 遗传算法优化 BP 神经网络的短时交通流混沌预测[J]. 控制与决策,2011,26(5):76-81.

(责任编辑:姚德权)

An Intelligent Hybrid Prediction for Financial Time Series Based on the Grey-ARIMA

LUO Hong-ben[1,2]

(1. *School of Business, Central South University, Changsha, Hunan 410083, China*;
2. *Office of Scientific R&D Hunan University, Changsha, Hunan 410082, China*)

Abstract: An Intelligent hybrid financial time series forecasting model is proposed based on a grey ARIMA. First, the financial times series grey forecasting model is constructed, and at the same time three parameters were optimized using PSO algorithm. The grey forecasting model residuals are then analyzed with ARIMA, and the coefficients for the ARIMA model are optimized with a genetic algorithm. Finally, the predicative results of the ARIMA model are used to compensate the grey forcasting model. The empirical results show that the algorithm proposed in this paper can have better fitting precision for a period of MA $<$ 107 time series data with the prediction error controlled within 5%; compared with the grey prediction algorithm and the neural network algorithm, the algorithm has obvious advantages in terms of the mean absolute error, root mean square error and the trend prediction.

Key words: Financial Time Series; Grey Prediction; ARIMA; PSO; Genetic Algorithm