

文章编号: 1005-0523(2011)02-0029-06

# 基于奇异谱分析的GRNN模型在金融时间序列中的应用

刘遵雄, 周天清

(华东交通大学信息工程学院, 江西南昌 330013)

**摘要:** 奇异谱分析(SSA)作为一类无参数、独立于模型的时间序列分析技术,适用于具有非线性、非平稳性、含噪声的金融时间序列数据的分析与研究。目前,基于SSA的预测通常采用线性递归、BP神经网络等模型,但其预测精度、训练速度并不理想。为此,该文提出基于SSA的广义回归神经网络(GRNN)预测模型,它以SSA所获取的主成份重构序列作为GRNN的输入进行预测。以同方股份收盘价格为测试数据,预测日收盘价。结果表明,基于SSA的GRNN模型预测效果不仅略优于GRNN预测方法,而且明显优于常规的SSA算法。

**关键字:** 奇异谱分析; 广义神经网络; 线性递归; 金融时间序列; 预测

**中图分类号:** TP183

**文献标识码:** A

时间序列是一类按时间顺序记录的系列观察值,可表示为 $\{f_1, f_2, \dots, f_t, \dots, f_n\}$ ,其中 $f_t$ 作为随机变量 $f$ 在 $t$ 时刻的观察值。由于时间序列的观察值之间具有一定的依赖性 or 相关性,主要表现为一定的周期性、趋势性和平稳性,可以依据历史值来预测未来值。金融时间序列是一类特殊的时间序列数据,通常具有复杂性、随机性、非线性和含噪声等特点。

在过去几十年里,各种线性与非线性的时间序列建模预测方法被提出,包括自回归移动平均(ARMA)、求和自回归移动平均(ARIMA)、自回归条件异方差(ARCH)、人工神经网络(ANN)<sup>[1]</sup>、遗传算法(GA)<sup>[2]</sup>、支持向量机(SVM)<sup>[3]</sup>、门限自回归(TAR)、自适应回归(AAR)和泛函数自回归(FAR)等,其中许多方法已成功应用到金融时间序列分析中,并且获得了较为理想的结果。然而,其中有些方法须以在一定假设条件(如线性、平稳性、标准性假设等)为前提,而另一些方法则无法直接适用于高维、高噪声数据的建模。奇异谱分析(singular spectrum analysis, SSA)技术的提出是时间序列分析研究的重大变革,它结合典型时间序列分析、多元统计、多元几何、动态系统、信号处理及奇异值分解(SVD)等技术<sup>[4]</sup>,破除了许多传统方法的假设前提,是一类元参数、独立于模型的时间序列分析技术<sup>[5]</sup>。

奇异谱分析通过将原始序列延时地排列成一矩阵形式,然后应用奇异值分解将原始序列分解成少数可解释、独立的成份之和,如缓慢变化趋势、摆动成份和随机噪声<sup>[6]</sup>。奇异谱分析技术主要用于解决趋势或准周期成分的检测与提取、降噪、预测、异常点检测等问题,广泛应用于气候、环境、地理、社会科学及金融等多方领域。

常见的奇异谱分析预测技术通常以传统神经网络或线性递归公式(linear recurrent formula, LRF)<sup>[7]</sup>为模型进行预测,但是基于奇异谱分析的传统神经网络模型训练速度缓慢,而基于奇异谱分析的线性递归预测效果有待改进。基于此,本文将广义回归神经网络(general regression neural network, GRNN)与奇异谱分析技术相结合形成基于奇异谱分析的广义回归神经网络预测模型,以同方股份收盘价格为测试对象验

**收稿日期:** 2010-02-24

**基金项目:** 教育部人文社会科学研究项目(09YJA630036);江西省自然科学基金项目(2010GZS0034)

**作者简介:** 刘遵雄(1967—),男,教授,博士,主要研究方向为智能信息处理。

证其预测效果。

## 1 奇异谱分析

奇异谱分析算法包含分解与重构两个阶段。前者将原始序列延时地排列成一矩阵形式,进行奇异值分解;后者利用求得的主成份进行分组并对组成份所形成的矩阵对角平均化。

### 1.1 分解

给定一长度为  $N > 2$  的时间序列  $\mathbf{F} = (f_0, f_1, \dots, f_{N-1})$ , 组建一  $L \times K$  阶轨迹矩阵 (trajectory matrix)  $\mathbf{X}$  :

$$\mathbf{F} = (f_0, f_1, \dots, f_{N-1}) \rightarrow \mathbf{X} = [X_1 : X_2 : \dots : X_K], \mathbf{X}_j = (f_{j-1}, f_{j-2}, \dots, f_{j+L-2})^T \quad (1)$$

其中副对角线方向上元素各自相等,  $L$  为窗口长度,  $K = N - L + 1$ 。

对轨迹矩阵  $\mathbf{X}$  进行奇异值分解 (SVD) 求得若干主成份方向, 其中 SVD 的第  $r$  个主成份  $\mathbf{V}_r$  由  $\mathbf{X}\mathbf{X}^T$  的第  $r$  个特征值  $\lambda_r$  和特征向量  $\mathbf{U}_r$  确定

$$\mathbf{X} = \mathbf{X}_1 + \dots + \mathbf{X}_d, \quad \mathbf{V}_r = \mathbf{X}^T \mathbf{U}_r / \sqrt{\lambda_r}, \quad d = \max\{r : \lambda_r > 0\}, \quad \mathbf{X}_r = \sqrt{\lambda_r} \mathbf{U}_r \mathbf{V}_r^T \quad (2)$$

其中:  $\sqrt{\lambda_r}$  为第  $r$  个奇异值;  $\mathbf{U}_r$  为左特征向量;  $\mathbf{V}_r$  为右特征向量, 它们共同形成一特征环  $(\sqrt{\lambda_r}, \mathbf{U}_r, \mathbf{V}_r)$ 。

### 1.2 重构

选择 SVD 的成份子组  $J \subset \{1, 2, \dots, L\}$ , 构建矩阵  $\mathbf{X}_J = \sum_{r \in J} \sqrt{\lambda_r} \mathbf{U}_r \mathbf{V}_r^T$ , 并对该矩阵沿对角线方向求取  $i+j=C$  的所有元素的均值, 其中  $C$  为常数。这些按  $C$  值大小得到的均值形成序列便为成份子组  $J$  的重构序列  $\mathbf{G}$ 。可表示为

$$\mathbf{X}_J = \sum_{r \in J} \sqrt{\lambda_r} \mathbf{U}_r \mathbf{V}_r^T \rightarrow \mathbf{G} = (g_0, g_1, \dots, g_{N-1}) \quad (3)$$

将矩阵  $\mathbf{X}_J$  重构为长度为  $N$  的序列  $\mathbf{G}$ , 算法如下: 设  $Y \in R^{L \times K}$ , 定义  $L^* = \min\{L, K\}$ ,  $K^* = \max\{L, K\}$  且  $Y^* = \begin{cases} Y & L \leq K \\ Y^T & L > K \end{cases}$ ,  $Y^* \in R^{L^* \times K^*}$ 。重构序列  $\mathbf{G} = (g_0, g_1, \dots, g_{N-1})$  则可通过式(4)计算获得

$$g_n = \begin{cases} \frac{1}{n+1} \sum_{m=1}^{n+1} y_{m, n-m+2}^* & 0 \leq n \leq L^* - 1 \\ \frac{1}{L^*} \sum_{m=1}^{L^*} y_{m, n-m+2}^* & L^* - 1 < n < K^* \\ \frac{1}{N-n} \sum_{m=n-K^*+2}^{N-K^*+1} y_{m, n-m+2}^* & K^* \leq n < N \end{cases} \quad (4)$$

## 2 预测模型

### 2.1 线性递归公式

给定一长度为  $N > 2$  的时间序列  $\mathbf{F} = (f_0, f_1, \dots, f_{N-1})$ , 假定  $\mathbf{F}$  的轨迹矩阵  $\mathbf{X} = [X_1 : X_2 : \dots : X_K]$  的窗口长度为  $L$ 。  $U_1, U_2, \dots, U_r$  为  $r$  维线性空间  $S_r \subset R^L$  的标准正交基, 并且  $\hat{X} = [\hat{X}_1 : \hat{X}_2 : \dots : \hat{X}_K] = \sum_{i=1}^r U_i U_i^T X$ 。定义  $U^\nabla \in R^{L-1}$  为  $U$  的首  $L-1$  个成份,  $U^\Delta \in R^{L-1}$  为  $U$  的最后  $L-1$  个成份, 即若  $U = (u_1, u_2, \dots, u_{L-1}, u_L)^T$ , 则  $U^\nabla = (u_1, u_2, \dots, u_{L-1})^T$ ,  $U^\Delta = (u_2, \dots, u_{L-1}, u_L)^T$ 。

任何向量  $\mathbf{F} = (f_{n-L+1}, f_{n-L}, \dots, f_n)^T \in S_r$  的最后一个元素均可由前面元素序列  $(f_{n-L+1}, f_{n-L}, \dots, f_{n-1})$  的线性组合表示<sup>[5]</sup>

$$f_n^T = \sum_{k=1}^{L-1} a_k f_{n-k} \quad (5)$$

其系数  $A=(a_1, a_2, \dots, a_{L-1}) = \frac{1}{1-v^2} \sum_{i=1}^r \pi_i U_i^\nabla$ , 其中  $v^2 = \pi_1^2 + \pi_2^2 + \dots + \pi_r^2 < 1$ ,  $\pi_i = u_L$ 。

由公式(6)定义时间序列  $F_{n+M} = (f_1, f_2, \dots, f_{n+M})$  :

$$f_i = \begin{cases} \tilde{f}_i & i = 1, \dots, n \\ \sum_{j=1}^{L-1} a_j f_{i-j} & i = n+1, \dots, n+M \end{cases} \quad (6)$$

其中:  $f_{n+1}, \dots, f_{n+M}$  为线性递归公式的  $M$  期预测值;  $\tilde{f}_i$  为序列  $(f_1, f_2, \dots, f_n)$  应用公式(3)(4)的结果,在此仍为原序列。现定义一线性操作  $OP^r : S_r \rightarrow R^L$

$$OP^r F = \begin{pmatrix} F^\nabla \\ A^T F^\nabla \end{pmatrix}, \quad F \in S_r \quad (7)$$

$$Z_i = \begin{cases} \tilde{X}_i & i = 1, \dots, K \\ OP^r Z_{i-1} & i = K+1, \dots, K+M \end{cases} \quad (8)$$

其中:  $F \in R^n$  为  $F$  的首  $n$  个成份;  $\tilde{X}_i$  为子序列  $X_i$  应用公式(3)(4)的结果,仍为原始子序列。可以发现,公式(8)可以看成公式(6)的向量形式,矩阵  $Z=[Z_1, \dots, Z_{K+M}]$  为时间序列  $F_{n+M}$  的轨迹矩阵,并且由此进行公式(1)的逆操作便可获得一真实值与预测值构成的时间序列。

### 2.2 广义回归神经网络

广义回归神经网络<sup>[8-9]</sup>是径向前馈神经网络RBF的一个分支,它以数理统计为基础进行非线性(核)回归分析。GRNN与传统的神经网络一样,具有很强的非线性建模能力、较高的容错性及鲁棒性,其逼近能力和学习速度较传统的RBF网络有着较强的优势,且当样本数据较少时,预测效果不受影响。

广义神经网络为四层网络结构,包括输入层、模式层、求和层及输出层。网络结构见图1,其中,输入层各神经元分别对应于输入向量中的各维;模式层中的

各神经元各设有激活函数  $\exp\left(-\frac{D_i^2}{2\sigma^2}\right)$ ;求和层的神经

元分为两类,一类对应于公式(9)的分母,但这样神经元只有一个,其它属于另一类的神经元则对应于公式(9)的分子,可以看出,公式(9)仅代表一个目标输出;输出层为目标值(预测值)。

对应网络  $p$  维输入向量  $X = \{x_1, x_2, \dots, x_p\}$ , 其输出为  $Y = \{y_1, y_2, \dots, y_p\}$ , 网络结构如图1。

如果GRNN神经网络的隐含层的传递函数以高斯函数为概率密度函数,那么网络的输出为

$$\hat{Y} = \frac{\sum_{i=1}^n Y_i \exp\left(-\frac{D_i^2}{2\sigma^2}\right)}{\sum_{i=1}^n \exp\left(-\frac{D_i^2}{2\sigma^2}\right)} \quad (9)$$

其中  $D_i^2 = (X - X_i)^T (X - X_i)$ ;  $X_i, Y_i (i=1, 2, \dots, n)$  分别为第  $i$  个样本的输入与输出;  $\sigma$  为平滑参数。

### 3 实验结果与分析

以同方股份1997年6月至1998年5月收盘价格共204个数据作为测试数据。根据奇异谱分析的步骤,

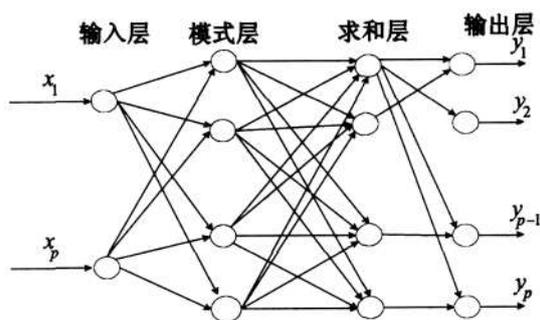


图1 GRNN网络结构  
Fig.1 Network structure of GRNN

首先将序列以80为窗口长度(一般小于等于序列长度的一半)组织成一延时矩阵,然后采用SVD对延时矩阵进行分解,得80个特征值(按非增序排列)。首5个特征值的贡献率分别为98.765 66%,0.411 16%,0.363 15%,0.252 86%,0.005 481%,其中98.765 66%的特征值对应的重构序列代表原始序列的趋势,见图2,后四项则指示着序列的波动情况,见图3。特征值中最后75项由于贡献率极小,可被视为序列噪声数据。最后由前5个成份重构成降噪后的序列,见图4,该序列将用于预测。

图2中,原始序列呈现出较为复杂的波动变化,而由贡献率98.765 66%的特征向量重构序列拟合曲线近似于一上升直线。可以发现,由首个成份形成的重构序列能够较好地捕捉原始序列的趋势。

图3中,由前2-5个成份形成的重构序列拟合曲线波动变化比较平缓,没有出现锐角变化情况,而且基本上反映了原始序列的波动变化。由于金融时间序列中普遍存在噪声,噪声点通常呈现出较大的波动变化(表现为较为尖锐的齿状),因此,可将由前2-5个成份形成的重构序列近似地看成原始序列的波动成份,其很好刻画了原始序列的波动变化。

图4中,由前5个成份形成的重构序列拟合曲线波动变化相对于原始序列较平缓,其较好地过滤了噪声数据,实现了对原始序列的逼近。该文用该5个成份形成的降噪重构序列作为广义神经网络预测模型的输入。

该文提出了一种基于奇异谱分析的广义神经网络预测方法,其基本工作原理:首先对原始序列进行奇异谱分析得到降噪后的重构序列,然后将广义神经网络应用于该重构序列。相比传统的奇异谱分析预测(如 caterpillar SSA 3.3<sup>[10]</sup>)主要采用线性递归函数为预测模型。

对于广义神经网络的预测方法,最为关键的参数为前面所提的平滑因子 $\sigma$ ,考虑到广义神经网络是一类训练速度快,非线性映射能力强,能以任意速度逼近任意连续函数的网络模型。该文通过考虑预测误差(绝对值形式)试探性地在 $[0.01, 0.5]$ 间选择平滑因子 $\sigma$ 。

对原始序列和重构序列进行广义回归神经网络预测的平滑因子与预测误差的关系见图5。从图中可知,两曲线均呈现出先减后增的趋势,其中前者在平滑因子为0.12时取得最小误差均值0.005 004,后者在平滑因子为0.07时取得最小误差均值0.002 877。故对于原始序列应用GRNN时,平滑因子取0.12;对于重构序列则取0.07。

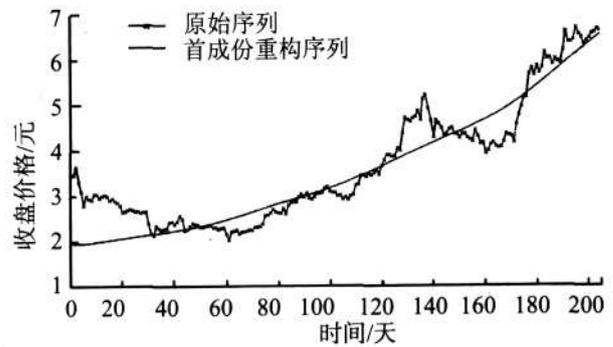


图2 首个成份重构图

Fig.2 Reconstruction chart of the first ingredient

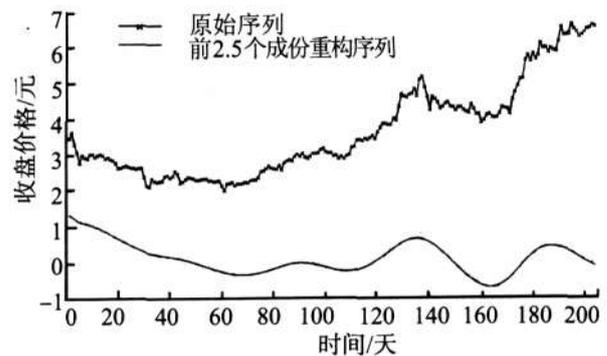


图3 前2-5个成份重构图

Fig.3 Reconstruction chart of the previous 2-5 ingredients

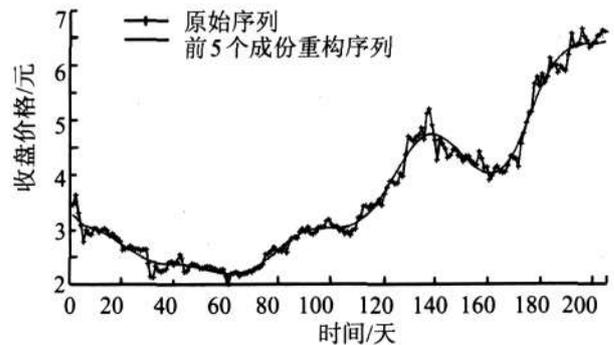


图4 前5个成份重构图

Fig.4 Reconstruction chart of the previous 5 ingredients

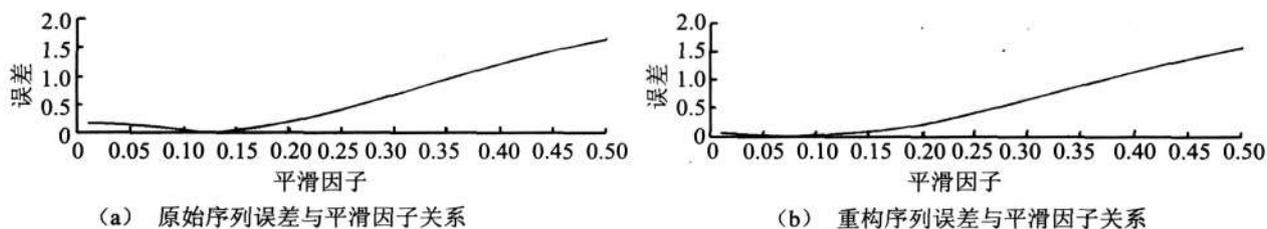


图5 平滑因子与误差关系  
Fig.5 Relationship between smooth factors and error

由于影响股票价格难以进行长期预测,往往只能进行短期的预测。该文使用GRNN, LRF+SSA和GRNN+SSA三种方法对同方股份的收盘价格进行未来一天的预测。GRNN直接将原始数据为其输入进行预测;LRF+SSA先利用SSA获得降噪成份,后以LRF为模型进行预测;GRNN+SSA同样先利用SSA获取降噪成份,再使用GRNN进行预测。各类方法的预测值及其预测误差见表1,从表中可以发现,采用LRF+SSA方法进行预测的误差值最大,GRNN次之,而GRNN+SSA预测误差最小,原因在于GRNN+SSA方法既能有效利用SSA技术降噪能力,又能摄取GRNN强大的非线性映射能力及任意逼近能力,从而使其达到了较两者更优的预测效果。

表1 各类预测值及误差  
Tab.1 Various predicted values and error

|     | LRF+SSA | GRNN    | GRNN+SSA |
|-----|---------|---------|----------|
| 预测值 | 6.252 0 | 6.333 3 | 6.331 1  |
| 误差  | 0.076 3 | 0.005 0 | 0.002 9  |

#### 4 结论

提出了基于奇异谱分析的广义神经网络预测方法GRNN+SSA,并将其应用于同方股份收盘价格的短期预测,实验结果表明,GRNN+SSA预测效果明显好于传统的LRF+SSA及GRNN方法,原因在于,GRNN+SSA方法同时兼有SSA和GRNN方法的优点,即集成了SSA降噪能力及GRNN的强大非线性映射与任意逼近能力。因此,可试将其推广用于回归复杂多噪声时间序列的分析建模。

#### 参考文献:

[1] BODYANSKIY Y, POPOV S. Neural network approach to forecasting of quasiperiodic financial time series[J]. European Journal of Operational Research, 2006, 175(3): 1357-1366.

[2] MAHFOUD S, MANI G. Financial forecasting using genetic algorithms[J]. Applied Artificial Intelligence, 1996, 10(6): 543-565.

[3] KIM K. Financial time series forecasting using support vector machines[J]. Neurocomputing, 2003, 55(1/2): 307-319.

[4] HASSANI H, HERAVI S, ZHIGLJAVSKY A. Forecasting European industrial production with singular spectrum analysis[J]. International Journal of Forecasting, 2008, 25(1): 103-118.

[5] GOLYANDINA V N A Z. Analysis of Time Series Structure: SSA and Related Techniques[M]. London: Chapman & CRC, 2001.

[6] HASSANI H. Singular spectrum analysis based on the minimum variance estimator[J]. Nonlinear Analysis: Real World Applications, 2010, 11(3): 2065-2077.

[7] HASSANI H, SOOFI A S, ZHIGLJAVSKY A A. Predicting daily exchange rate with singular spectrum analysis[J]. Nonlinear Analysis: Real World Applications. 2010, 11(3): 2023-2034.

[8] 梁凤国,李帅莹,于森,等. 基于GRNN神经网络的参考作物腾发量预测[J]. 人民长江, 2009, 40(05): 58-59.

[9] 冯志鹏,宋希庚,薛冬新,等. 基于广义回归神经网络的时间序列预测研究[J]. 振动、测试与诊断, 2003, 23(2): 105-109.

## Application of GRNN Model Based on SSA in Financial Time Series

Liu Zunxiong, Zhou Tianqing

(School of Information Engineering, East China Jiaotong University, Nanchang 330013, China)

**Abstract:** Singular spectrum analysis (SSA) is a kind of non-parameter and independent model time series analysis technique, which can be suitable for analyzing and studying nonlinear, non-stationary and noisy financial time series. Nowadays, the prediction based on SSA often adopts linear recursion, BP neural network and others as its models. However, the prediction accuracy and training speed is not perfect. Therefore, this paper proposes a new method called general regression neural network (GRNN) based on SSA that uses reconstructed series of components from SSA as its inputs and makes the closing price of tong fang as test data to forecast daily closing price. Experimental results show that the improved method is much better than original one and also slightly better than GRNN.

**Key words:** SSA; GRNN; LRF; financial time series; forecast

---

## 《华东交通大学学报》征稿启事

从2011年4月开始,《华东交通大学学报》设立了博士研究生论文发表基金,征集博士研究生的原创性论文(须博士研究生独著或与其导师合著)。校内外博士研究生论文审核通过后一经发表,将免收版面费,并给予300元/篇的稿酬。

征稿内容主要包括:交通运输工程、土木工程、环境工程、机械工程、电气工程、信息工程、数学、化学、物理等研究论文及交叉学科方面的学术论文。

华东交通大学学报编辑部