

基于 SSA-ELM 的大宗商品价格预测研究

王 珏^{1,2,3}, 齐 琛^{1,3}, 李明芳⁴

(1. 中国科学院 数学与系统科学研究院, 北京 100190; 2. 中国科学院 国家数学与交叉科学中心, 北京 100190;
3. 中国科学院大学, 北京 100190; 4. 北京科技大学 数理学院, 北京 100083)

摘 要 随着经济全球化的发展, 国际期货市场中各大类大宗商品价格波动剧烈, 而全球经济形势不明朗以及货币政策不确定使得大宗商品期货价格难以被准确预测. 本文选取玉米, 原油, 黄金分别作为大宗商品农产品类、能源类、金属类的代表对象, 基于奇异谱分析方法 (singular spectrum analysis, SSA), 对商品期货价格进行分解, 结合 Kmeans 动态聚类技术将分解量聚合成不同特征的价格序列, 再采用具有优良特性的极限学习机算法 (extreme learning machine, ELM) 对模型进行训练, 得到大宗商品期货价格预测模型. 实证结果表明, 采用序列分解聚类策略能够显著提高模型预测精度, 在价格未来的整体水平和变动方向上都能达到较好的预测效果.

关键词 大宗商品; 预测; 奇异谱分析; 聚类; 极限学习机

Prediction of commodity prices based on SSA-ELM

WANG Jue^{1,2,3}, QI Chen^{1,3}, LI Mingfang⁴

(1. Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China; 2. National Center for Mathematics and Interdisciplinary Sciences, Chinese Academy of Sciences, Beijing 100190, China; 3. University of Chinese Academy of Sciences, Beijing 100190, China; 4. School of Mathematics and Physics, University of Science and Technology Beijing, Beijing 100083, China)

Abstract With the development of economic globalization, the fluctuations of commodity futures prices are increasingly violent. The uncertainty of the global economic situation and monetary policy make the commodity futures prices difficult to be accurately predicted. In this paper, corn, gold and crude oil are selected as the representatives of agricultural futures, industrial metals futures and energy futures, respectively. Based on the singular spectrum analysis (SSA) method, we decompose the commodity futures price and incorporate Kmeans dynamic clustering technique to group the decomposition series. Then the forecasting models of commodity futures price are developed by using extreme learning machine (ELM). The empirical results show that the decomposition and clustering schemes can significantly improve the accuracy of price forecasting, and performs well on the overall estimation and direction of change of the commodity prices.

Keywords bulk commodities; forecast; SSA; Kmeans; ELM

1 引言

在经济全球化形势下, 国际市场和国内市场联动性日益增强, 全球大宗商品定价大部分取决于期货市场

收稿日期: 2016-11-23

作者简介: 王珏 (1978-), 女, 博士, 副研究员, 研究方向: 智能计算, 经济预测; 通信作者: 齐琛 (1993-), 女, 硕士研究生, 研究方向: 经济预测与智能计算; 李明芳 (1976-), 女, 博士, 副教授, 研究方向: 物流管理.

基金项目: 国家自然科学基金 (71271202); 中国科学院青年创新促进会项目

Foundation item: National Natural Science Foundation of China (71271202); Youth Innovation Promotion Association, Chinese Academy of Sciences

中文引用格式: 王珏, 齐琛, 李明芳. 基于 SSA-ELM 的大宗商品价格预测研究 [J]. 系统工程理论与实践, 2017, 37(8): 2004–2014.

英文引用格式: Wang J, Qi C, Li M F. Prediction of commodity prices based on SSA-ELM[J]. Systems Engineering — Theory & Practice, 2017, 37(8): 2004–2014.

此文是第十三届金融系统工程与风险管理国际年会优秀论文.

功能。作为关系国计民生的期货市场交易商品, 大宗商品期货价格波动引起各方面的关注。同时, 大宗商品作为一种重要的投资品和保值商品, 在连接金融市场与实体经济之间的期货市场中价格波动较大。2008 年的美国次贷危机以及 2012 年的欧债危机也引起了大部分大宗商品期货价格大幅波动。此外, 全球经济形势不明朗, 市场投机行为加剧, 货币政策不确定, 以及其他影响因素进一步增大了大宗商品价格的预测难度。

传统的时间序列和计量经济学模型^[1-5], 如自回归移动平均模型 (ARIMA), 广义自回归条件异方差模型 (GARCH), 向量自回归模型 (VAR) 被广泛应用于经济时间序列的分析和预测。近年来, 有学者提出基于人工智能方法^[6-10] 的模型来捕捉价格序列的非线性规则, 如人工神经网络 (NN)^[11-13], 支持向量机 (SVM)^[14], 遗传算法 (GA)^[15], 也逐渐被应用于预测商品价格。张坤^[16] 等人通过研究影响黄金价格变动主要因素, 提出一种基于小波神经网络的黄金价格预测模型, 得到了比传统 BP 神经网络更好的预测效果。曾谦等人^[17] 提出一种投影寻踪优化的 BP 神经网络改进模型, 对黄金期货价格实现了高精度仿真。杨建辉等人^[18] 结合误差修正模型与支持向量回归模型, 研究了国际原油价格、人民币实际汇率与中国宏观经济之间的关系。此外, 单层前馈网络 (SLFN) 作为最常用的一种神经网络结构, 被广泛用于回归和分类等问题。然而, 传统研究采用梯度下降的学习方法来估计网络参数, 耗费大量时间, 并且容易出现过拟合和陷入局部最优的现象, 需要不断调整参数才能得到较好的结果。Huang^[19] 等人提出了极限学习机算法 (extreme learning machine, ELM), 能够训练出更好的非线性时间序列预测模型。

在预测建模前, 很多研究采用不同的分解方法对序列进行分解重构, 再输入预测模型, 往往能够得到更好的预测效果。Reboredo 等人采用小波多重分解重构的方法, 研究油价和美元汇率之间的关系。Yu^[20] 等人采用基于压缩感知的人工智能模型 (CSD-AI) 对原油价格进行了降噪和预测, 成功捕捉到不同时间段的真实数据变化, 得到了更为精准的油价预测模型。Zhang^[21] 和陈振宗^[22] 等人采用基于 EMD 分解的预测模型分析并预测了石油和黄金价格的波动情况及未来价格。奇异谱分析 (SSA) 作为信号处理技术, 1980 年后, 由 Vautard^[23] 和 Ghil^[24] 等人引入到气候诊断和预测研究中, 主要用于时间序列的准周期信号分析与相关信号的提取。鉴于 SSA 具有自适应滤波信号分析的重建功能, 丁裕国^[25] 等利用该方法结合自回归 (AR) 模型进行 Nino 海区 SSTA 短期气候预测, 结果表明该方法与基础 AR 模型预报相比, 具有更高的预测准确率, 对 1997-1998 年的强 ENSO 事件的超前预报以及利用历史样本作回溯预报试验均有较高可信度。朱红蕊^[26] 等人在此基础上提出基于 SSA-AR 方法的 MJO 指数预报模型, 得到了较好的预报效果。Sivapragasam^[27] 和 Simoes^[28] 等人提出了基于 SSA 的 SVM 预测模型, 实现了对短期随机降雨量的预测, 证明了采用 SSA 方法去噪后的降雨量序列进行建模具有更高的预测效果和结果置信度。然而, 以往研究中对分解方法和预测模型只是进行简单的组合, 例如直接将所有分解量作为输入变量放入模型, 或剔除最后的残差项进行简单相加重构, 没有利用分解量的组合进行研究测试。

本文通过综合梳理国内外最新预测方法和大宗商品研究文献, 选取玉米, 原油, 黄金分别作为大宗商品农产品类、能源类、金属类的代表对象, 提出将分解量聚类再预测的大宗商品期货价格预测模型。模型基于奇异谱分析 (SSA) 方法, 对三类大宗商品期货价格序列进行 SSA 分解, 根据设定阈值剔除无关因素影响的噪声序列, 得到相应期货价格的有效分解量。结合 Kmeans 动态聚类技术进行分解量合成, 提取出大宗商品价格的趋势和周期波动等特征, 从而分析不同角度的价格变化。最后分别输入到神经网络中, 使用极限学习机算法 (ELM) 实现对大宗商品期货价格的预测。

2 方法概述

本文采用基于奇异谱分析 (SSA) 的极限学习机 (ELM) 模型进行大宗商品价格的预测。与以往研究不同的是, 本文对 SSA 分解序列进一步进行 Kmeans 动态聚类, 提取出原始期货价格序列的趋势和周期波动等特征, 再分别使用极限学习机算法训练神经网络模型, 组合不同特征的预测结果, 作为大宗商品的期货价格预测。

2.1 基础知识

奇异谱分析 奇异谱分析 (singular spectrum analysis, SSA) 由 Broomhead 等提出, 是一种广义功率谱分析, 适用于研究周期振荡行为的分析方法。首先根据时间序列的周期特性和观测时间段构造出相应的轨迹

矩阵, 再对轨迹矩阵进行分解和重构, 从而提取出能够代表原序列的多组分解量, 包括代表长期趋势的低频分解量、代表序列周期的高频分解量以及无明显规律的噪声分解量等, 实现对时间序列结构的分析, 提取重要信息达到降噪的效果. 本文使用 SSA 方法对大宗商品期货价格序列进行分解降噪处理, 提取其有效成分来进行经济分析和建模预测.

- 选取窗口长度 L 构建轨迹矩阵 X (trajectory matrix) 如下

$$\begin{bmatrix} x_1 & x_2 & x_3 & \cdots & x_{T-L+1} \\ x_2 & x_3 & x_4 & \cdots & x_{T-L+2} \\ x_3 & x_4 & x_5 & \cdots & x_{T-L+3} \\ \vdots & \vdots & \vdots & & \vdots \\ x_L & x_{L+1} & x_{L+2} & \cdots & x_T \end{bmatrix} = X \quad (1)$$

- SVD 分解: 对矩阵 X 进行 SVD(singular value decomposition) 分解:

$$X = \sum_{i=1}^d \sqrt{\lambda_i} U_i V_i^T \quad (2)$$

- 根据给定阈值, 取前 p 个分量作为原始序列的替代¹: 记其中一个分量对应的矩阵为 X^* , 采用对角平均的方式, 将分解矩阵 $(X_i = \sqrt{\lambda_i} U_i V_i^T)$ 转换为序列 $(\tilde{x}_k = (x_1^*, \cdots, x_T^*))$.

$$x_k^* = \begin{cases} \frac{1}{k} \sum_{m=1}^k x_{m,k-m+1}^* & 1 \leq k < L^* \\ \frac{1}{L^*} \sum_{m=1}^{L^*} x_{m,k-m+1}^* & L^* \leq k < K^* \\ \frac{1}{T-k+1} \sum_{m=k-K^*+1}^{N-K^*+1} x_{m,k-m+1}^* & K^* < k \leq T \end{cases} \quad (3)$$

其中 $K = T - L + 1$, $L^* = \min(L, K)$, $K^* = \max(L, K)$.

极限学习机 (ELM) 极限学习机算法 (extreme learning machine, ELM) 由 Huang 等人提出, 是一个简洁, 高效, 无需调参, 并具有极快学习速度的算法. 其网络训练模型采用前向单隐层结构, 随机初始化输入权重和偏置并得到相应的输出权重, 在训练过程中无需调参, 只需设置隐含层神经元的个数, 便可获得唯一最优解. 考虑 N 个互异的数据样本 $(x_i, y_i), i = 1, 2, \cdots, N$, 其中输入数据 $x_i = [x_{i,1}, x_{i,2}, \cdots, x_{i,N}] \in \mathbb{R}^N$, 输出数据 $y_i = [y_{i,1}, y_{i,2}, \cdots, y_{i,M}] \in \mathbb{R}^M$, 对应 ELM 算法概述如下:

假设有一个具有 L 个隐层节点, 激活函数为 $g(a, b, x)$ 的单层神经网络 (SLFNs) 可以以零误差逼近这 N 个互异的数据样本, 也就是说, 存在 $a_i, b_i, \beta_i, i = 1, 2, \cdots, L$ 使得

$$f_L(x_j) = \sum_{i=1}^L \beta_i g(a_i, b_i, x_j) = y_j, j = 1, 2, \cdots, N \quad (4)$$

将公式 (4) 用矩阵形式表达如下:

$$H\beta = Y \quad (5)$$

其中,

$$H = \begin{bmatrix} g(a_1, b_1, x_1) & \cdots & g(a_L, b_L, x_1) \\ \vdots & \ddots & \vdots \\ g(a_1, b_1, x_N) & \cdots & g(a_L, b_L, x_N) \end{bmatrix}_{N \times L}, \beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times M}, Y = \begin{bmatrix} y_1^T \\ \vdots \\ y_N^T \end{bmatrix}_{N \times M},$$

H 为隐层的输出矩阵, 第 i 列则表示 (x_1, \cdots, x_N) 在第 i 个隐层节点对应的输出量, 第 j 行表示 x_j 在所有隐层节点上的输出量. 在多数情况下, 隐层节点数 L , 都会远小于样本数 N , 使得构建的神经网络很难逼近公式 (5), 对应于训练样本的输出和实际输出之间存在误差 $E(E = [e_1^T, \cdots, e_N^T]_{N \times M}^T)$, 于是公式 (5) 可以改写为:

$$H\beta = Y + E \quad (6)$$

1. 本文根据阈值大小, 取对应解释量高于阈值的分量代替原始价格序列. 例如阈值为 0.1%, 而前 6 个分量对应解释度高于 0.1%, 则 $p = 6$.

定义平方损失函数

$$J = \sum_{j=1}^N (\beta_i g(a_i, b_i, x_j) - y_j)^2 = (H\beta - Y)^T (H\beta - Y) \quad (7)$$

寻找最小二乘解 $\hat{\beta}$ 使得 $\|H\hat{\beta} - Y\| = \min_{\beta} \|H\beta - Y\|$. 在 H (隐层输出矩阵) 为列满秩的情况下, 利用 Moore-penrose 广义逆可以得到: $\hat{\beta} = \arg \min_{\beta} \|H\beta - Y\| = H^+ Y$, 在 H 矩阵非列满秩时, 最优外权可以利用奇异值分解 (SVD) 的方法得到.

2.2 SSA-ELM 模型构建

SSA 分解得到大宗商品价格的 P 个分量后, 再使用 Kmeans 聚类整合分解序列, 将序列的历史数据输入到单隐层的神经网络中, 采用 ELM 算法进行学习, 汇总不同聚类量的预测结果得到大宗商品价格的整体预测. 如图 1 所示, 具体算法如下:

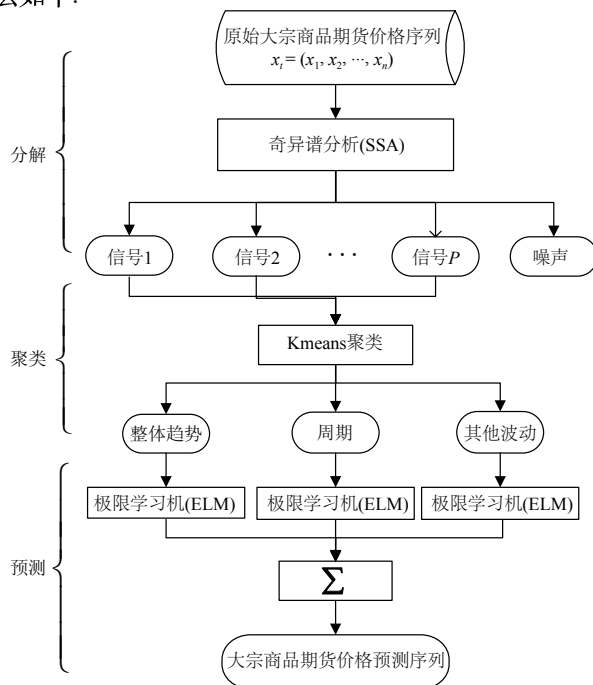


图 1 SSA-ELM 模型构建

Step 1 对大宗商品价格序列进行 SSA 分解并筛选出前 P 个分解量

- 构建轨迹矩阵: 选取窗口长度 L , 将观测到的一维数据 x_1, x_2, \dots, x_T 转换为轨迹矩阵 X
- 对矩阵 X 进行 SVD 分解 $X = \sum_{i=1}^d \sqrt{\lambda_i} U_i V_i^T$.
- 根据给定阈值, 取前 p 个分量作为原始序列的替代 x_1, \dots, x_p

Step 2 重构分解量

- 对分解量进行 Kmeans 聚类, 聚为 m 类
- 根据聚类结果进行重构 (同一类的分量加和), 得到重构序列 $\tilde{x}_{1,t}, \dots, \tilde{x}_{m,t}$

Step 3 使用重构量的滞后项 (即 $(\tilde{x}_{i,t-1}, \dots, \tilde{x}_{i,t-d})$) 作为输入, $x_{i,t}$ 作为期望输出, 分别对其进行 ELM 算法训练, 得到聚类重构的预测序列 $\hat{x}_{i,t}$

- 网络初始化: 给定训练样本集, 确定隐含层结点数 L , 激活函数 $g(a, b, x)$, 随机生成输入权值和偏置 a_i, b_i .
- 计算隐含层输出矩阵 H :

$$H = \begin{bmatrix} g(a_1, b_1, x_1) & \cdots & g(a_L, b_L, x_1) \\ \vdots & \ddots & \vdots \\ g(a_1, b_1, x_N) & \cdots & g(a_L, b_L, x_N) \end{bmatrix}_{N \times L}$$

- 计算输出层权值向量 $\hat{\beta}$: $\hat{\beta} = \arg \min_{\beta} \|H\beta - Y\| = H^+ Y$

Step 4 组合聚类重构序列的预测结果作为大宗商品的期货价格预测

$$\hat{x}_t = \sum_{i=1}^m \hat{x}_{i,t}$$

模型参数 建立 SSA-ELM 模型需要设定的参数如下: 1) SSA 分解中构建轨迹矩阵的窗口长度; 2) SSA 分解阈值; 3) ELM 算法中神经网络隐层的节点个数; 4) ELM 算法中采用的核函数. 根据以往研究经验, 窗口长度最好是序列周期的整数倍, 并且不能超过序列长度的一半. 而 SSA 分解阈值以及隐层节点的个数取决于输入的样本个数以及大宗商品价格市场的先验知识. ELM 一般采用 $\sin()$, $\text{sigmoid}()$ 和 $\text{hardlim}()$ 三种核函数, 实证分析将根据不同核函数的建模效果进行选择.

模型优点 SSA 分解重构能够得到去除噪声因素影响的序列, 更能代表大宗商品价格的真实信息; 而本文的模型进一步对分解量进行动态聚类, 提取出价格的趋势和周期等特征序列; 有针对性地提取价格信息, 让神经网络更容易学习到价格的内部机制. 另外, ELM 算法不需要用训练数据集去生成隐层前的所有参数, 它具有极快的学习和测试速度, 很好的泛化性能, 在保证预测精度下具有更小的计算复杂度, 并且不需要调参, 能使模型得到更稳定的结果.

3 实证分析

3.1 数据和参数选择

本节对大宗商品的期货价格数据建立 SSA-ELM 模型, 分析并预测期货价格. 具体数据和模型的参数选择如下.

实证数据 选取从 2000 年 1 月至 2016 年 6 月的玉米、原油以及黄金这三种商品期货价格进行研究. 其中, 玉米期货价格为芝加哥商品交易所 (CBOT) 的月度价格; 原油期货价格为纽约商业交易所 (NYMEX) 的西得克萨斯轻质原油 (WTI) 月度价格; 黄金期货价格为纽约商品交易所 (COMEX) 的月度价格. 玉米和黄金期货价格数据来源于万德数据库 (WIND), 原油期货价格数据来源于美国能源信息署 (EIA).

模型参数 三种商品的期货价格周期是 12 个月, 考虑到一共有 198 期数据, 因此模型的 SSA 分解的窗口长度 L 定为 24. 对于 SSA 分解阈值以及隐层节点个数, 多次试验后将 SSA 分解阈值设为 0.1%, 神经网络的隐层节点个数设为 10. 另外, 三种不同核函数中, $\sin()$ 核函数的在三类大宗商品的价格预测上整体效果最佳, 因此下文采用 $\sin()$ 核函数的建模结果进行分析.

3.2 SSA 分解聚类

在 0.1% 的分解阈值下, 玉米、原油、黄金三种商品期货价格分别得到 5、6、3 个分解量. 其中, 玉米期货价格 5 个分解量的贡献率为 99.71% (即这 5 个 SSA 分解量能够解释玉米市场期货价格 99.71% 的信息), 原油期货价格 6 个分解量的贡献率为 99.69%, 黄金期货价格 3 个分解量的贡献率为 99.83%. 将分解量相加即可得到去噪重构后的价格序列, 图 2 是三类商品期货价格的重构效果图 (红线为原始价格序列, 蓝线为重构价格序列), 由此见得, 在 0.1% 的阈值下得到的重构量能够解释大宗商品期货价格的主要信息.

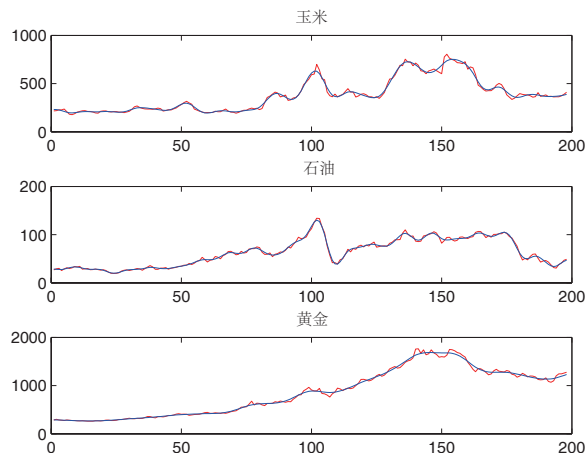


图 2 大宗商品期货价格重构效果图

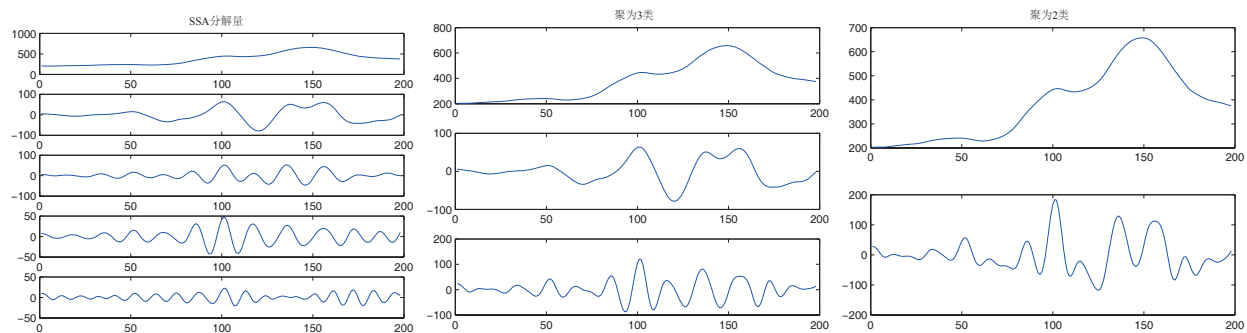


图 3 玉米期货价格分解聚类图

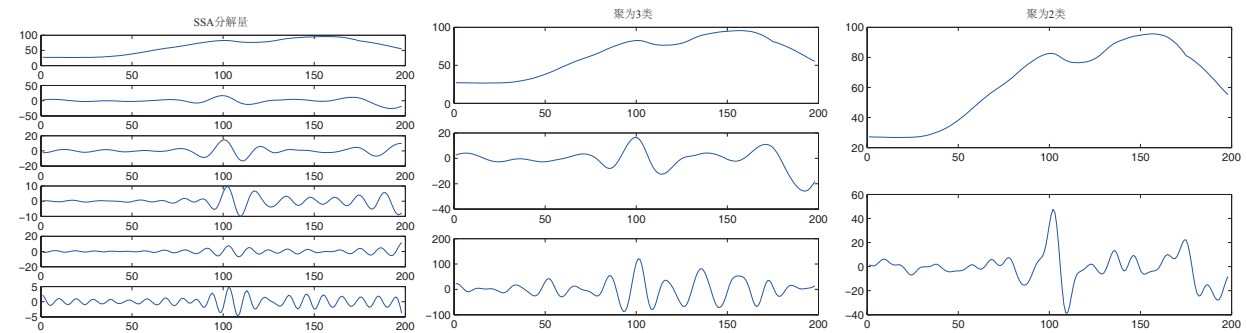


图 4 原油期货价格分解聚类图

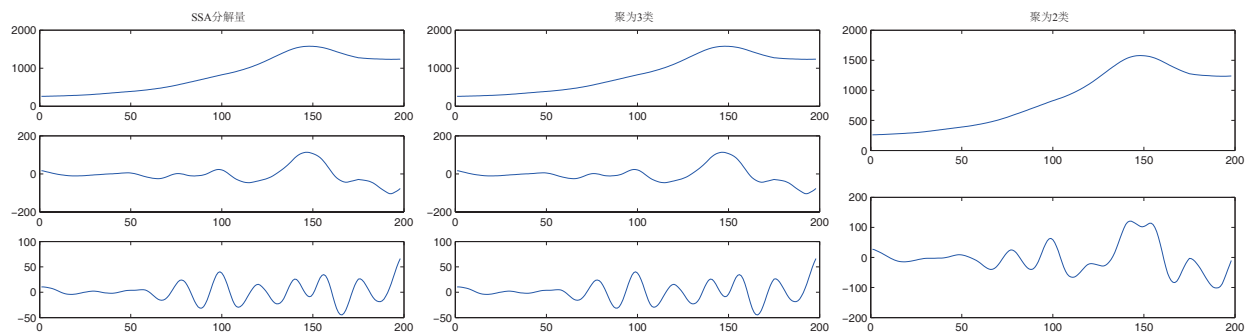


图 5 黄金期货价格分解聚类图

为了捕捉大宗商品期货价格的具体特征信息, 本文对三类大宗商品期货价格的 SSA 分解量分别进行 Kmeans 动态聚类, 得到代表价格不同特征的信号序列. Kmeans 聚类是一种动态聚类方法, 其聚类结果会随聚类初始值的不同而改变, 因此本文对分解量进行了多次聚类, 取其频次最高的一种分组方式作为最终聚类结果. 考虑到在 0.1% 的分解阈值下得到的分解量是 3~6 个, 本文将分解量分别聚为了 2 类和 3 类. 聚为 2 类时, 三种商品的第一个分量均自成一类, 其他分量为另一类; 聚为 3 类时, 第一、第二个分量分别为一类, 其他分量为第三类.

图 3 是玉米期货价格的 SSA 分解与 Kmeans 聚类图. 其中, 左侧为 5 个原始 SSA 分解量, 中间为分解量聚为 3 类的价格序列, 右侧为聚为 2 类的价格序列. 三种不同组合的分解图中第一条序列代表相应大宗商品期货价格整体趋势, 后两条特征序列分别代表价格的主要波动和周期信息. 原始 SSA 分解序列将不同频率的波动和周期项用后 4 项分解序列展现, 而聚为 3 类的价格序列第二项代表价格波动, 第三项代表价格周期. 同样的, 聚为 2 类的价格序列则综合提取了玉米期货价格的整体趋势项以及周期波动项.

图 4 和图 5 分别是原油期货价格和黄金期货价格对应的分解聚类图, 与玉米期货价格的图 3 类似, 均是第一条序列代表价格整体趋势, 剩下的序列为其波动和周期项 (由于黄金期货价格在 0.1% 的阈值下只得到 3 个分解序列, 因此其聚为 3 类的序列即为三个分解序列). 纵观三类商品的期货价格分解图, 其波动项在 2008 年至 2012 年期间较为聚集, 表明大宗商品期货价格都受到了经济危机的持续性冲击. 其中, 玉米和原油期货价格的价格趋势项在 2008 年和 2012 年两次经济危机刚发生阶段的 (横坐标第 100 点和 150 点) 涨幅最大, 表明

两次经济危机的爆发对玉米和原油期货价格有很强的影响. 而黄金作为价格比较稳定的大宗商品, 其期货价格受到 08 年经济危机的影响较小, 并依然维持价格上升的整体趋势, 但在 12 年经济危机后整体价格一路下跌, 受到的影响较之前更大. 三种分解聚类序列对本文接下来的 ELM 建模都能起到不同的辅助作用, 输入不同特征的聚类重构序列, 神经网络会学习到对应特征的价格信息用以预测未来期货价格. 下文将详细分析不同分解聚类的序列在期货价格预测上的效果.

3.3 预测结果

为了更好的评估模型的预测效果, 本文将数据集分为训练集 (前 80%) 和测试集 (后 20%), 采用多种评估指标进行评价, 包括: 均方误差 (RMSE), 平均绝对误差 (MAE), 平均相对绝对误差 (MAPE), 方向统计量 (Dstat). 前三个指标用于评估对价格绝对值的预测效果 (数值越小, 准确度越高), 而方向统计量指标为价格变动方向的预测准确性度量 (取值在 0%~100% 之间, 越接近 100%, 方向准确度越高). 本文依次采用原始的期货价格数据, SSA 分解后去噪重构的期货价格数据, SSA 分解后聚为 2 类的期货价格数据, 以及 SSA 分解后聚为 3 类的期货价格数据作为输入, 结合极限学习机算法 (ELM) 依次建立四种大宗商品价格预测模型.

表 1 一步向前预测结果					
	类型	RMSE	MAE	MAPE	Dstat
玉米	Baseline	30.86	21.25	4.96%	55.4%
	Recon	21.96	17.17	4.02%	63.9%
	C2	21.51	17.32	4.01%	61.0%
	C3	21.29	17.26	4.07%	59.5%
原油	Baseline	5.53	4.50	7.77%	58.9%
	Recon	3.64	2.88	4.84%	76.4%
	C2	3.71	2.82	4.57%	76.1%
	C3	4.08	3.28	5.98%	76.7%
黄金	Baseline	51.82	42.42	3.42%	60.0%
	Recon	49.58	41.96	3.38%	66.7%
	C2	41.27	35.16	2.87%	65.3%
	C3	42.61	36.24	2.94%	68.0%

表 1 是大宗商品期货价格的未来一期 (即未来 1 个月) 预测结果, 类型代表输入到 ELM 模型中的数据类型, 其中 Baseline 代表用原始数据的滞后项作为输入对未来期货价格进行预测, Recon 代表用 SSA 分解后重构的历史数据预测未来期货价格, C2 和 C3 分别代表采用 Kmeans 聚类方法将 SSA 分解量聚为 2 类和 3 类的序列, 对其分别预测并加和作为未来期货价格的预测结果. 评价指标显示, 基于 SSA 分解的 Recon 模型的预测效果优于仅采用原始数据的 Baseline 模型, 同时, 结合 Kmeans 聚类的 SSA-ELM 模型在大部分指标上能够进一步提升预测精度. 对于玉米期货价格, MAE 和 Dstat 指标显示直接重构的输入最佳, 而 RMSE 和 MAPE 显示 SSA 分解再聚类的集成预测效果更优, 对比 Baseline 类型, 四类指标都显示结合 SSA 分解的模型均具有较好的预测性能, 说明玉米期货价格在剔除市场上无关因素后能够更有效的反映其价格水平和真实波动, 从而使模型得到更高的预测精度. 同样的, 对于原油期货价格, 所有经过 SSA 处理的类型对应的 MAPE 都在 6% 以内, 说明去除噪声分量后的原油期货价格序列都有效反映了原油期货市场的不同价格水平, 其中 C3 类型在方向上的预测精度达到 76.7%, 高于 Baseline 类型近 18%, 同时 C2 类型在价格水平的预测上, MAE 和 MAPE 指标最优, 在方向上也捕捉到了 76.1% 的价格变化, 将原油市场无关因素剔除后, 聚类组合成的两种序列能够体现原油价格市场的整体趋势和周期特征, 从而更好的捕捉原油市场期货价格的波动情况. 对于黄金期货价格, 结合 Kmeans 聚类的两种模型的 MAPE 在 3% 以内, 四个指标均显示结合 SSA-Kmeans (C2 和 C3) 的模型具有更好的效果. 其中, C2 类型的三类价格水平指标最小, 说明该模型对黄金期货价格水平的预测最准确, 而方向上也捕捉到了 65.3% 的价格波动, 说明对于黄金市场, 剔除无关因素后的期货价格序列, 其趋势项和周期波动项能够有效的反映黄金市场期货价格的整体水平及其变动情况.

图 6 是三类大宗商品期货价格的 SSA-ELM 模型的预测效果图 (上方为样本内预测效果, 下方为样本外预测效果, 红线为预测序列, 蓝线为实际期货价格序列). 整体来看, 模型对大宗商品期货价格的样本内预测

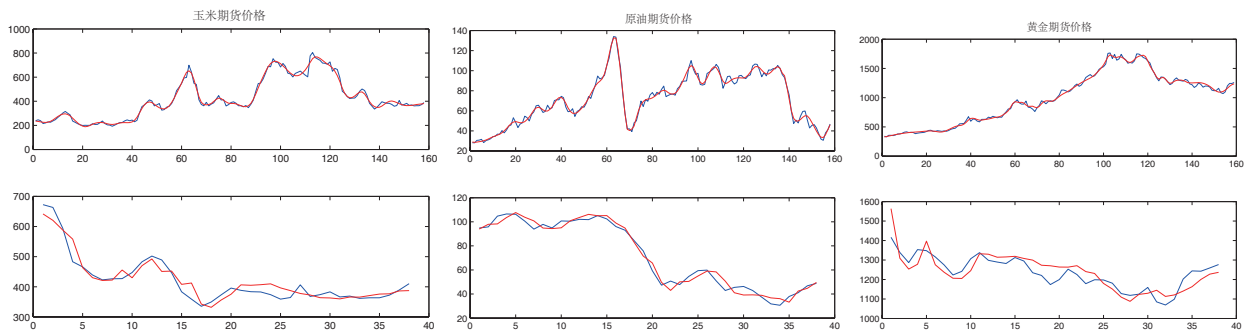


图 6 大宗商品期货价格预测效果图

具有较高的精度,能够有效捕捉 2000 年至 2013 年期间的期货价格波动. 其中,2005 年左右,供需因素主导的原油和玉米期货价格急速上涨后骤降,而作为保值硬通货的黄金期货价格相对稳定,小幅波动后继续维持价格上升趋势.此外,模型还捕捉到了三类商品的价格拐点,例如,精准地预测了该阶段石油价格的最高点.2008 美国次贷危机引起全球金融危机,使得三类商品期货价格出现了大幅波动.2009 至 2010 年随着国际资本加速向新兴经济体流动,经济开始升温复苏,作为能源类的石油期货价格呈现良好的上升趋势,从红线来看,模型剔除其他因素引起的小幅波动,正确预测了上涨趋势.

在样本外预测方面,模型对 2013 到 2014 期间的大宗期货价格预测效果较好,成功捕捉到重要价格拐点并给出了价格波动幅度.尽管模型对 2015 年的三类期货价格预测稍有偏移,但仍较准确的预测了整体趋势和期货价格均值.其中,对黄金期货价格的预测,模型在预测的价格绝对值上表现比较平稳,但相应的拐点捕捉略显不足,这是由于黄金的 SSA 分解量贡献率在 0.1% 以上的只有 3 个,其他都被作为噪声因素剔除,因此模型以小幅波动的损失为代价,得到了整体价格水平的精准预测.对于玉米和原油期货价格,二者均在 2014 下半年后的预测存在一定程度的偏移.导致偏差的主要原因是由于 2014 和 2015 年全球大宗商品市场出现史上罕见的全线下跌,而且下跌范围广、幅度大、程度深,是继 2008 年金融危机后最为严重的一次下跌.在原油市场,全球经济放缓、美元升值、地缘政治风险消退等多种因素导致油价从 2014 年年中开始暴跌,2014 年全年油价跌幅接近 50%,这种疲软态势持续到 2015 年底.虽然模型对于此阶段的预测出现小幅偏移,但从整体预测精度和拐点的角度来看,依然捕捉到了大部分价格变化,达到了较好的预测效果.

考虑到只用预测未来 1 个月的结果来分析不同输入对模型的影响不够全面,本文还对三类大宗商品分别预测了未来 2 个月和 3 个月的期货价格,即两步和三步向前预测 (n 步向前预测: $x_{t+n} = f(x_t, x_{t-1}, \cdots)$).其预测结果如表 2 和表 3 所示,相比一步向前预测,本文提出的结合 Kmeans 动态聚类的 SSA-ELM 预测模型在对大宗商品未来 2~3 月的期货价格预测上具有更加明显的优势.对于玉米期货价格,未来 2 至 3 月的预测是 C3 类型具有最好的效果,对价格水平的预测相对误差在 6% 以内,方向准确率在 50% 以上.对比一步预测结果, C3 类型也具有较好的预测效果,这表明玉米期货市场在剔除无关因素后,主要由整体趋势项,周

表 2 两步向前预测结果						表 3 三步向前预测结果					
	类型	RMSE	MAE	MAPE	Dstat		类型	RMSE	MAE	MAPE	Dstat
玉米	Baseline	55.42	39.47	9.26%	50.7%	玉米	Baseline	73.74	50.52	12.41%	45.9%
	Recon	26.11	21.35	5.19%	56.0%		Recon	35.66	27.93	6.94%	52.4%
	C2	24.96	19.02	4.68%	59.2%		C2	33.34	24.69	6.21%	48.8%
	C3	23.34	18.26	4.42%	61.1%		C3	28.41	22.12	5.55%	51.5%
原油	Baseline	9.80	7.50	13.80%	58.7%	原油	Baseline	13.51	10.57	19.65%	40.0%
	Recon	5.58	4.57	9.96%	67.6%		Recon	8.65	7.69	14.55%	54.4%
	C2	5.88	4.55	8.29%	72.1%		C2	9.18	7.44	13.85%	63.8%
	C3	7.64	5.55	11.15%	71.7%		C3	12.89	8.89	18.28%	65.6%
黄金	Baseline	89.74	75.66	6.18%	44.5%	黄金	Baseline	123.36	108.82	8.95%	38.9%
	Recon	52.22	45.34	3.64%	66.3%		Recon	71.13	60.20	4.80%	62.0%
	C2	42.84	35.50	2.91%	62.3%		C2	46.60	37.49	3.10%	61.4%
	C3	48.11	38.79	3.16%	62.2%		C3	57.29	42.55	3.49%	55.7%

期项和波动项三大方面来代表其期货价格结构. 而对于原油和黄金期货价格则是 C2 类型更有优势, 其中对原油的两步向前预测能达到 72% 的方向准确率, 相对于原始的 Baseline 类型提高了 13%, 对价格水平的预测准确率也高于 Baseline 类型 5%, 黄金的 C2 类型方向准确率也都在 60% 以上, 高于 AR 类型 20% 以上. 对黄金和原油未来 3 个月期货价格的预测也显示采用 Kmeans 聚类方法将 SSA 分解量聚为两类的 C2 类型综合效果最优, 说明对于原油和黄金期货市场, 剔除无关因素后主要由整体趋势和周期波动两大方面来代表其期货价格结构更为合适.

综合未来 1~3 个月的期货价格预测结果, 结合 SSA 分解方法的预测模型普遍优于未对价格数据进行预处理的原始 Baseline 模型, 而在 SSA 分解量的处理方法中, 采用 Kmeans 动态聚类对 SSA 分解量的进一步集成, 相对于简单的重构方法更能抽取出代表期货价格信息的不同特征, 在未来 2 个月和 3 个月的预测上, 基于 SSA-Kmeans (C2 和 C3) 的方法具有更加明显的优势, 对未来 1 个月的预测上也具有较好的效果. 其中, 对于农产品类的玉米, SSA-Kmeans 抽取出的三大类特征代表了其期货价格的整体趋势、周期以及波动信息, C3 类型的 SSA-ELM 模型具有最优的综合预测效果. 对于能源类的原油以及金属类的黄金, C2 类型效果综合最优, 这表明将原油和黄金期货价格的波动和周期作为一项综合特征, 再结合其整体趋势特征, 能够辅助模型更好的捕捉到价格的主要波动以及整体水平信息, 进一步提升模型预测效果.

3.4 模型比较

传统的 AR 时间序列分析, 机器学习方法 SVR (支持向量机回归), 以及人工智能方法 NN (BP 神经网络) 广泛应用于不同领域的时间序列预测, 本节采用这三种方法对大宗商品期货价格进行预测, 并与本文提出的 SSA-ELM 方法进行评估比较. 其中 AR 模型的滞后阶由偏自相关系数决定, SVR 和 NN 的输入为对应滞后阶的历史期货价格. AR 模型采用最小二乘法进行模型估计. SVR 模型使用 RBF 核函数, tolerance 和 epsilon 分别设为 0.001 与 0.1, 为了保证模型结果的稳定性, 采用十折交叉验证法选择最佳 SVR 模型进行预测. 对于 NN, 采用 *morlet()* 基函数, 设置 10 个隐层节点, 并运行 1000 次取其平均预测结果.

表 4 模型比较: 一步向前预测					
	模型	RMSE	MAE	MAPE	Dstat
玉米	AR	28.3	19.8	4.57%	62.2%
	SVR	31.5	21.5	4.85%	55.3%
	NN	39.6	27.7	6.32%	56.5%
原油	AR	5.2	4.3	7.00%	64.9%
	SVR	5.9	4.5	7.96%	55.3%
	NN	6.3	5.0	8.60%	59.2%
黄金	AR	42.5	35.3	2.86%	59.5%
	SVR	57.2	45.1	3.57%	60.5%
	NN	67.9	56.9	4.55%	60.5%

表 5 模型比较: 两步三步向前预测									
模型	两步					三步			
	RMSE	MAE	MAPE	Dstat		RMSE	MAE	MAPE	Dstat
玉米	AR	51.5	34.4	7.92%	56.8%	69.7	46.4	10.86%	54.1%
	SVR	54.1	38.4	8.98%	51.4%	75.7	51.5	12.55%	47.2%
	NN	59.1	42.3	9.88%	56.3%	77.0	53.5	12.98%	43.9%
原油	AR	9.2	7.1	12.50%	59.5%	13.5	10.8	19.33%	43.2%
	SVR	11.0	8.4	15.45%	59.5%	14.5	10.5	20.56%	44.4%
	NN	10.0	7.7	13.84%	60.1%	13.5	10.4	19.07%	40.7%
黄金	AR	77.1	61.3	4.94%	40.5%	104.5	81.4	6.57%	43.2%
	SVR	100.6	81.8	6.53%	42.1%	141.5	118.5	9.48%	44.7%
	NN	108.9	91.5	7.35%	41.9%	142.5	122.5	9.86%	46.6%

表 4 为三个模型对期货价格的一步向前预测评估指标, 结果显示,SVR 和 AR 模型在三类大宗商品期货

价格的预测上都达到了90%以上的准确率(MAPE均小于10%),其中AR模型在玉米和原油期货价格的预测上比SVR和NN效果更好。此外,AR在黄金的期货价格水平的预测上相比另外两种方法也占有更大优势,不过SVR和NN方法都捕捉到了60.5%的价格波动,优于AR模型的方向预测。对比本文SSA-ELM算法的结果(表1),SSA-ELM模型效果普遍更优。经过SSA去噪后的序列更能够让预测模型识别出真实的价格结构,再加上Kmeans动态聚类方法对有效分解量的组合,提取出价格的趋势和周期波动特征,能使模型预测出更准确的未来动向。同样的,表5为三种方法的未来2~3个月的预测结果,预测精度大部分低于本文的SSA-ELM模型预测结果。由此可见,本文提出的预测方法,能够剔除市场无关因素,并根据分解量的不同特征聚集成表示价格趋势和周期波动的主要成分,从而实现对未来期货价格的预测,并有效捕捉其价格变化。

4 结论

本文提出的SSA-ELM预测方法,从序列本质特征出发,结合人工智能方法做到了对三大类大宗商品(玉米、原油、黄金)期货价格的有效预测。主要贡献如下:

- 采用奇异谱分析(SSA)方法对序列进行分解,得到不同频率的序列,抽取出价格序列的整体趋势,以及不同周期上的波动情况,直观展示出期货价格的不同特征。与以往研究不同的是,本文在SSA分解的基础上使用Kmeans动态聚类方法,对原始价格分解量进行分类重组,更加清晰的展现了价格整体趋势和周期波动特征,并且将不重要因素造成的影响作为噪声剔除,从而凸显不同阶段的市场冲击以及其他重大事件对期货价格的影响。SSA-Kmeans对大宗商品价格的分解聚类能够提供分析价格本身特征和外界冲击的直观窗口,有助于对历史情况的分析以及未来趋势的判断。
- 使用极限学习机算法进行模型训练,ELM作为人工智能方法中建模更为稳定,速度更加快捷的学习算法,得到了比传统时间序列方法AR,以及常用机器学习算法SVR更好的模型表现,无论从价格的绝对值还是价格变动方向的预测上来看都达到了令人满意的预测效果。此外,ELM无需调参的特性更能够让研究者得到较为一致的模型结果,使得实验可重复,并具有更为稳定的优良特性。
- 本文的SSA-ELM模型框架,加入了Kmeans聚类方法,使得模型的输入变量不只是原始的分解序列,或者单一的去噪序列,而是代表期货价格趋势和周期波动的特征序列,相比其他类型更能够使模型识别出价格本质特征和外界影响,从而得到更精准的价格预测。

在以后的研究中,将考虑加入其他先行市场因素,比如能够及时反映外界市场情况的互联网公众关注度、舆情指数等,进一步提高模型的预测精度。

参考文献

- [1] 蔡纯. 本次经济危机主要大宗商品期货价格波动性研究[J]. 金融理论与实践, 2010(2): 64-69.
Cai C. The study of main commodities future price volatility in subprime crisis[J]. Financial Theory & Practice, 2010(2): 64-69.
- [2] Salisu A A, Oloko T F. Modeling oil price-US stock nexus: A VARMA-BEKK-AGARCH approach[J]. Energy Economics, 2015, 50: 1-12.
- [3] Zhang Y, Wang J. Exploring the WTI crude oil price bubble process using the Markov regime switching model[J]. Physica A: Statistical Mechanics and its Applications, 2015, 421: 377-387.
- [4] Allegret J P, Mignon V, Sallenave A. Oil price shocks and global imbalances: Lessons from a model with trade and financial interdependencies[J]. Economic Modelling, 2015, 49: 232-247.
- [5] 陈茹雯, 黄仁. 非线性自回归时序模型研究及其预测应用[J]. 系统工程理论与实践, 2015, 35(9): 2370-2379.
Chen R W, Huang R. Research of general expression for nonlinear autoregressive model and its forecast application[J]. Systems Engineering — Theory & Practice, 2015, 35(9): 2370-2379.
- [6] Li X, Ma J, Wang S, et al. How does google search affect trader positions and crude oil prices?[J]. Economic Modelling, 2015, 49: 162-171.
- [7] Salisu A A, Oloko T F. Modeling oil price-US stock nexus: A VARMA-BEKK-AGARCH approach[J]. Energy Economics, 2015, 50: 1-12.
- [8] Allegret J P, Mignon V, Sallenave A. Oil price shocks and global imbalances: Lessons from a model with trade and financial interdependencies[J]. Economic Modelling, 2015, 49: 232-247.

- [9] Zhang Y, Wang J. Exploring the WTI crude oil price bubble process using the Markov regime switching model[J]. *Physica A: Statistical Mechanics and its Applications*, 2015, 421: 377–387.
- [10] Maghyereh A. Oil price shocks and emerging stock markets: A generalized var approach[M]// *Global Stock Markets and Portfolio Management*, Springer, 2006: 55–68.
- [11] Yu L A, Wang S Y, Lai K K. Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm[J]. *Energy Economics*, 2008, 30(5): 2623–2635.
- [12] 黄建风, 陆文聪. 基于小波 -NAR 神经网络的气象要素时间序列预测与天气指数彩虹期权估值 [J]. *系统工程理论与实践*, 2016, 36(5): 1146–1155.
- Huang J F, Lu W C. Forecasting of meteorological time series and pricing of weather index rainbow options: A wavelet-NAR neural network model[J]. *Systems Engineering — Theory & Practice*, 2016, 36(5): 1146–1155.
- [13] 李哲敏, 许世卫, 崔利国, 等. 基于动态混沌神经网络的预测研究 —— 以马铃薯时间序列价格为例 [J]. *系统工程理论与实践*, 2015, 35(8): 2083–2091.
- Li Z M, Xu S W, Cui L G, et al. Prediction study based on dynamic chaotic neural network — Taking potato time-series prices as an example[J]. *Systems Engineering — Theory & Practice*, 2015, 35(8): 2083–2091.
- [14] Xie W, Yu L, Xu S, et al. A new method for crude oil price forecasting based on support vector machines[C]// *Computational Science — ICCS 2006: 6th International Conference*, Springer Berlin Heidelberg, 2006: 444–451.
- [15] Motlaghi S, Jalali F, Ahmadabadi M N. An expert system design for a crude oil distillation column with the neural networks model and the process optimization using genetic algorithm framework[J]. *Expert Systems with Applications*, 2008, 35(4): 1540–1545.
- [16] 张坤, 郁湧, 李彤. 小波神经网络在黄金价格预测中的应用 [J]. *计算机工程与应用*, 2010, 46(27): 224–226.
- Zhang K, Yu Y, Li T. Application of wavelet neural network in prediction of gold price[J]. *Computer Engineering and Applications*, 2010, 46(27): 224–226.
- [17] 曾濂, 马丹颀, 刘宗鑫. 基于 BP 神经网络改进的黄金价格预测 [J]. *计算机仿真*, 2010, 27(9): 200–203.
- Zeng L, Ma D D, Liu Z X. Gold price forecast based on improved BP neural network[J]. *Computer Simulation*, 2010, 27(9): 200–203.
- [18] 杨建辉, 潘虹. 国际原油价格、人民币实际汇率与中国宏观经济研究 [J]. *系统工程理论与实践*, 2008, 28(1): 1–8.
- Yang J H, Pan H. Study on world oil price and real exchange rate with China's macro-economics[J]. *Systems Engineering — Theory & Practice*, 2008, 28(1): 1–8.
- [19] Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: Theory and applications[J]. *Neurocomputing*, 2006, 70(1–3): 489–501.
- [20] Yu L, Zhao Y, Tang L. A compressed sensing based AI learning paradigm for crude oil price forecasting[J]. *Energy Economics*, 2014, 46: 236–245.
- [21] Zhang X, Lai K K, Wang S Y. A new approach for crude oil price analysis based on empirical mode decomposition[J]. *Energy Economics*, 2008, 30(3): 905–918.
- [22] 陈振宗. 基于经验模态分解 (EMD) 的国际现货黄金价格分析及预测 [D]. 南昌: 江西财经大学, 2015.
- Chen Z Z. Analysis and forecast of international spot gold price based on empirical mode decomposition (EMD) [D]. Nanchang: Jiangxi University of Finance and Economics, 2015.
- [23] Vautard R, Yiou P, Ghil M. SSA: A toolkit for short, noisy chaotic signals[J]. *Physica D*, 1992, 58: 95–126.
- [24] Ghil M, Mo K. Intraseasonal oscillations in the global atmosphere. Part I: Northern hemisphere and tropics[J]. *Journal of the Atmospheric Sciences*, 1991, 48(5): 780–792.
- [25] 丁裕国, 江志红, 朱艳峰. Nino 海区 SSTA 短期气候预测模型试验 [J]. *热带气象学报*, 1998, 14(4): 289–296.
- Ding Y G, Jiang Z H, Zhu Y F. Experiment on short-term climatic prediction to SSTA over the Nino oceanic region[J]. *Journal of Tropical Meteorology*, 1998, 14(4): 289–296.
- [26] 朱红蕊, 江志红, 张勤, 等. 基于 SSA-AR 方法的 MJO 指数预报模型试验 [J]. *热带气象学报*, 2010, 26(3): 371–378.
- Zhu H R, Jiang Z H, Zhang Q, et al. MJO index forecasting based on SSA-AR model[J]. *Journal of Tropical Meteorology*, 2010, 26(3): 371–378.
- [27] Sivapragasam C, Liong S Y, Pasha M F K. Rainfall and runoff forecasting with SSA-SVM approach[J]. *Journal of Hydroinformatics*, 2001, 3(3): 213–217.
- [28] Simões N, Wang L, Ochoarodriguez S, et al. A coupled SSA-SVM technique for stochastic short-term rainfall forecasting[J]. ©the authors, 2011.