

姚冠新, 顾 晴. 基于经验模态分解和支持向量机的农产品价格短期预测[J]. 江苏农业科学, 2014, 42(9): 402–405.

# 基于经验模态分解和支持向量机的农产品价格短期预测

姚冠新, 顾 晴

(江苏大学管理学院, 江苏镇江 212000)

**摘要:** 针对农产品价格序列非平稳、非线性的特征, 提出一种基于经验模态分解和支持向量机的短期农产品价格组合预测方法。以全国苹果批发价格的周度价格序列为研究对象, 首先对价格序列进行非平稳、非线性检验; 然后将价格序列分解成一系列具有不同特征尺度的分量; 最后针对各分量的特点分别构建支持向量机模型进行预测, 得到苹果批发价格的预测值。研究结果表明, 该预测模型能够更好地追踪农产品价格的变化, 其预测精度与 BP(back propagation)神经网络模型相比有显著提高。

**关键词:** 农产品; 价格; 经验模态分解; 支持向量机; 短期预测

**中图分类号:** S11<sup>+</sup>4; F304.2 **文献标志码:** A **文章编号:** 1002–1302(2014)09–0402–04

我国农产品价格由于受供求变化、生产资料价格、劳动力成本、流通成本、自然气候以及国际农产品价格等多种因素影响, 表现出短期波动大、非平稳、非线性等特点<sup>[1]</sup>。来自物价部门的信息表明, 2004—2012 年, 国际市场粮价年均波动幅度在 20% 左右, 其中 2008 年、2012 年出现过 2 次大幅度震荡, 全年的波动幅度在 40% 以上。农产品价格的波动, 不仅会影响农业生产的发展、农产品的流通、消费和农民的收入水平, 而且会对整个社会经济生活的安定产生重大影响。2013 年 3 月 5 日, 温家宝总理在十二届全国人大一次会议上所作的《政府工作报告》中指出, 保持物价总水平基本稳定始终是宏观调控的重要目标。因此, 开展农产品价格短期预测, 对有效指导农民及时调整生产和规避市场风险, 提高政府部门对农产品市场的调控效率、保障农产品市场的有序稳定具有重要意义<sup>[2]</sup>。

从国内外研究动态来看, 学者们在农产品价格短期预测领域提出了多种有效预测方法, 如回归分析、指数平滑、自回归移动平均模型 (autoregressive integrated moving average, ARIMA)、条件异方差模型 (autoregressive conditional heteroskedasticity, ARCH)、神经网络、模糊理论、组合模型等。

刘海清等建立了海南省芒果的价格指数平滑模型, 结果表明该模型能够有效地对价格曲线进行拟合和预测<sup>[3]</sup>。李干琼等基于价格序列数据的 ADF 检验和 ARCH 效应检验, 结合 2008—2009 年间的 731 d 日价格数据分析, 利用 ARIMA、ARCH、广义条件异方差模型 (generalized autoregressive conditional heteroskedasticity, GARCH) 等现代时间序列法, 分别建立了西红柿日批发价格预测模型, 其中 GARCH 模型在预测中具有更高的精度<sup>[4]</sup>。董晓霞等通过平稳性、季节性、趋势性以及异方差等一系列检验后, 最终选择了 3 种模型对我国鲜奶零售价格短期预测进行了应用模拟, 结果显示 ARCH 模型预测结果精确度最好, Holt–Winters 无季节性模型稳定性

最好<sup>[5]</sup>。李哲敏等根据 2000 年 3 月至 2009 年 9 月的月度数据, 选出 4 个能够及时、全面获取数据的相关因子作为自变量因子, 构建了禽蛋市场价格短期预测模型, 并对模型的有效性进行相关检验<sup>[6]</sup>。罗长寿采用蔬菜市场价格数据分别建立了 BP 神经网络模型、基于遗传算法的神经网络模型、RBF 神经网络模型, 结果表明, 集成预测模型效果优于单一预测模型<sup>[7]</sup>。韩延杰提出了一种基于模糊信息粒化和遗传算法的支持向量机 (genetic algorithm–support vector machine, GA–SVM) 农产品价格预测模型, 为提高预测精度, 利用遗传算法对支持向量机的参数进行优化, 实证分析表明该方法能对农产品价格的变化范围进行有效地预测<sup>[8]</sup>。刘慧等根据 HP 和 BP 滤波方法分析价格波动的特征, 利用 ARIMA 模型预测绿豆价格的短期走势<sup>[9]</sup>。

综上所述, 农产品价格短期预测的研究主要采用计量经济模型, 但计量经济模型对于非线性时间序列的预测具有局限性, 同时对于数据本身的条件限制较多, 造成模型预测结果或多或少存在一些缺憾。由于人工神经网络具有很强的非线性逼近能力, 因此被广泛应用到价格短期预测中, 但是人工神经网络训练速度慢、容易陷入局部最小点、存在过拟合现象以及泛化能力较差等局限性使其难以达到预期精度<sup>[10]</sup>。经验模态分解 (empirical mode decomposition, EMD) 是一种处理非平稳、非线性数据序列的新方法, 能够分解出研究对象在不同尺度上的波动信息。该方法从数据自身出发, 通过分析本征模态函数 (intrinsic mode function, IMF) 来揭示数据序列的内在特征。相关研究表明, 经验模态分解可以很好地处理非平稳、非线性序列, 是提取数据序列趋势的有效方法<sup>[11]</sup>。

经验模态分解目前已经在地球物理学、生物医学、结构分析、设备诊断等众多自然科学领域得到了应用; 经验模态分解在社科领域仅涉及原油价格和股票价格的预测<sup>[12]</sup>。本研究将经验模态分解方法引入到农产品价格短期预测中, 旨在突破农产品价格预测技术, 提高预测精度。

## 1 材料与方法

### 1.1 经验模态分解方法

经验模态分解方法的优点是能够对非线性、非平稳过程

收稿日期: 2013–12–13

基金项目: 江苏省社会科学重点项目 (编号: 13EYA003)。

作者简介: 姚冠新 (1961—), 男, 江苏启东人, 教授, 博士生导师, 研究方向为农产品物流。E-mail: yaogx@ycit.cn。

通信作者: 顾 晴, 硕士研究生, 研究方向为农产品价格预测。

E-mail: guqing\_ujs@foxmail.com。

的数据进行线性化和平稳化处理,分解的最终函数彼此之间是正交的,从而在分解的过程中尽可能地保留了数据本身的特性<sup>[13]</sup>。其基本思路是用波动上、下包络的平均值去确定“瞬时平衡位置”,进而提取出内在模函数(intrinsic mode function, IMF)<sup>[14]</sup>。内在模函数,记为  $I(t)$ ,其计算过程主要有以下 3 个步骤。

(1) 找出原序列  $Y(t)$  的各个局部极大值,为更好保留原序列的特性,局部极大值定义为时间序列中的某个时刻的值,其前一时刻的值不比它大,后一时刻的值也不比它大。然后用三阶样条函数进行插值,得到原序列  $Y(t)$  的上包络序列值  $Y_{\max}(t)$ 。同理,可以得到下包络序列值  $Y_{\min}(t)$ 。

(2) 对每个时刻的  $Y_{\max}(t)$  和  $Y_{\min}(t)$  取平均,得到瞬时平均值  $m(t)$ :

$$m(t) = [Y_{\max}(t) + Y_{\min}(t)]/2。 \quad (1)$$

(3) 用原序列  $Y(t)$  减去瞬时平均值  $m(t)$ ,得到类距平值序列  $h(t)$ :

$$h(t) = Y(t) - m(t)。 \quad (2)$$

对于不同的数据序列,  $h(t)$  可能是内在模函数,也可能不是。若  $h(t)$  中极值点的数目和跨零点的数目相等或至多只差 1 个,并且各个瞬时平均值  $m(t)$  都等于零,则它就是内在模函数;反之,把  $h(t)$  当作原序列,重复以上步骤,直至满足内在模函数的定义,求出内在模函数为止。求出了第一个内在模函数  $I_1(t)$ ,即从原序列中分解出第 1 个分量。然后,用原序列减去  $I_1(t)$ ,得到剩余值序列  $r_1(t)$ :

$$r_1(t) = Y(t) - I_1(t)。 \quad (3)$$

至此,提取第 1 个内在模函数的过程全部完成。然后,把  $r_1(t)$  作为一个新的原序列,按照以上步骤,依次提取第 2、第 3、...,直至第  $N$  个内在模函数  $I_N(t)$ 。之后,由于  $r_N(t)$  变成一个单调序列,再也没有内在模函数能被提取出来。如果把分解后的各分量合并起来,就得到原序列  $Y(t)$ :

$$Y(t) = \sum_{i=1}^N I_i(t) + r_N(t)。 \quad (4)$$

## 1.2 支持向量机方法

支持向量机(support vector machine, SVM)是 Cortes 和 Vapnik 于 1995 年首先提出的,它在解决小样本、非线性及高维模式识别中表现出许多特有的优势,并能够推广应用到函数拟合等其他机器学习问题中。支持向量机的基本思想是通过用内积函数定义的非线性变换将输入空间变换到一个高维空间,在这个高维空间中寻找输入变量和输出变量之间的一种非线性关系。SVM 有严格的理论基础,是基于结构风险最小化原则的方法,明显优于传统的基于经验风险最小化原则的常规神经网络方法。

给定一数据集  $G = \{(x_i, d_i)\}_{i=1}^n$ ,其中  $x_i$  是输入向量,  $d_i$  是期望值,  $n$  是训练样本个数,需要求解的回归函数如下形式:

$$y = f(x) = w\phi(x) + b。 \quad (5)$$

式中:  $\phi(x)$  是从输入空间到高维特征空间的非线性映射,  $w$  和  $b$  是系数,它们由最小化风险函数  $R(C)$  来估计:

$$\text{Minimize}; R(C) = \frac{1}{2} \|w\|^2 + C \frac{1}{n} \sum_{i=1}^n L_\varepsilon(d_i, y_i)。 \quad (6)$$

$$L_\varepsilon(d_i, y_i) = \begin{cases} |d - y| - \varepsilon & |d - y| \geq \varepsilon \\ 0 & |d - y| < \varepsilon \end{cases}。 \quad (7)$$

$\varepsilon$  是给定的参数,  $L_\varepsilon(d_i, y_i)$  是  $\varepsilon$  不敏感损失函数,  $\frac{1}{2} \|w\|^2$

是对函数面的量度,  $C$  是常量,称为惩罚因子。

为了寻求系数  $w$  和  $b$ ,需要引入松弛变量  $\xi_i$  和  $\xi_i^*$ ,使式(6)最小化:

$$\text{Minimize}; R(w, \xi) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i, \xi_i^*) \quad (8)$$

$$\text{s. t.} \begin{cases} d_i - w\phi(x_i) - b \leq \varepsilon + \xi_i^* \\ w\phi(x_i) + b - d_i \leq \varepsilon + \xi_i \\ \xi_i, \xi_i^* \geq 0 \end{cases}$$

引入拉格朗日(Lagrange)算子  $\alpha_i, \alpha_i^*$ ,将式(8)转换为其对偶形式:

$$\max R(\alpha_i, \alpha_i^*) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) d_i - \frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*)$$

$$K(x_i, x_j) - \sum_{i=1}^n (\alpha_i + \alpha_i^*) \varepsilon \quad (9)$$

$$\text{s. t.} \begin{cases} \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0, C] \end{cases}。$$

其中  $K(x_i, x_j) = \langle \phi(x_i) \cdot \phi(x_j) \rangle$ ,  $K(x_i, x_j)$  称为核函数,常用的核函数有线性核函数、多项式核函数、径向基核函数(radial basis function, RBF)等。

相应的回归函数式(5)可写成:

$$y = f(x, \alpha_i, \alpha_i^*) = \sum_{i=1}^n K(x, x_i) + b。 \quad (10)$$

## 1.3 EMD-SVM 农产品价格预测方法

农产品价格序列具有较强的非线性、非平稳的特点,而常规预测方法在非线性序列上难以取得较好的效果,鉴于 EMD 分解方法在处理非线性、非平稳数据具有突出优势,因此提出了一种基于 EMD-SVM 的农产品价格短期预测方法,具体方法如下:

(1) 利用 EMD 方法对农产品价格原始序列进行分解,得到若干个 IMF 分量  $I_i(t)$  和残差项;

(2) 分别对归一化处理后的各 IMF 分量  $I_i(t)$  和残差项  $r_N(t)$  建立 SVM 回归模型,选取最佳参数和核函数进行预测;

(3) 对预测结果进行反归一化处理,得到各分解序列的预测值;

(4) 将各分解序列预测值叠加得到农产品价格的预测值;

(5) 与实际值比较,确定误差指标并进行误差分析。

## 2 结果与分析

### 2.1 数据来源说明

本研究选取苹果批发市场价格数据为分析对象,数据来源于商务部的商务预报网站(<http://cif.mofcom.gov.cn/>),数据类型为周数据。数据周期为 2006 年 7 月 7 日至 2013 年 5 月 31 日,共 360 个样本数据。描述性统计情况见表 1,数据分布见图 1。

表 1 样本数据的描述统计情况

类别	价格(元/kg)
最低	3.12
最高	7.69
平均	5.61
标准差	1.34

注:样本数据 360 个。

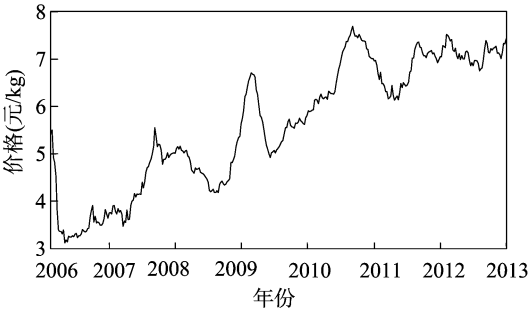


图1 全国苹果批发市场的周度价格

2.2 非平稳、非线性检验

由表 2 可知,  $t$  统计量  $> 10\%$  水平, 不能拒绝原假设, 即认为苹果批发价格时间序列是非平稳的, 进一步验证该序列一阶单整。对一阶差分后的序列作自相关、偏自相关分析, 发现自相关系数 4 阶截尾, 偏自相关系数 3 阶截尾, 建立 ARIMA(3,1,4) 模型。剔除不显著变量, 得到方程:

$$\Delta Price_t = 1.000\ 594 \Delta Price_{t-1} + 0.145\ 651 u_{t-1} + 0.148\ 719 u_{t-2} + 0.200\ 604 u_{t-3} + \varepsilon_t$$

由于自相关系数和偏自相关系数显著趋于零, 说明残差序列不存在自相关, 是白噪声序列, 但残差平方的自相关系数、偏自相关系数很大, 存在较明显的自相关性(表 3)。因此, 苹果批发价格序列具有非线性的特点, 用非线性模型来描述价格波动更为合适。

表 2 原始数据 ADF 平稳性检验结果

外因: 固定	$t$ 统计量	$P$ 值
Dickey Fuller 检验统计量	-1.229 316	0.662 7
检验临界值 1% 水平	-3.448 518	
5% 水平	-2.869 442	
10% 水平	-2.571 047	

注: 采用单边  $P$  值。零假设: 价格有一个单位根。

表 3 ARIMA 模型方程结构参数

变量	系数	标准误	$t$ 统计量	$P$ 值
AR(1)	1.000 594	0.001 587	630.521 1	0
MA(1)	0.146 561	0.052 167	2.809 475	0.005 2
MA(2)	0.148 719	0.052 219	2.847 981	0.004 7
MA(3)	0.200 604	0.052 287	3.836 564	0.000 1
$R^2$	0.992 614	被解释变量的均值		5.607 095
校正 $R^2$	0.992 551	被解释变量标准差		1.341 62
回归的标准误	0.115 79	Akaike 信息准则		-1.462 97
残差平方和	4.746 17	Schwarz 准则		-1.419 61
对数似然估计	265.871 9	Hannan - Quinn 准则		-1.445 73
Durbin - Watson 统计	1.974 39			

2.3 EMD 分解及结果分析

基于样本数据, 通过 MATLAB 软件对全国苹果批发价格时间序列分解出 8 个不同尺度的 IMF 和 1 个残余项, 如图 2 所示。

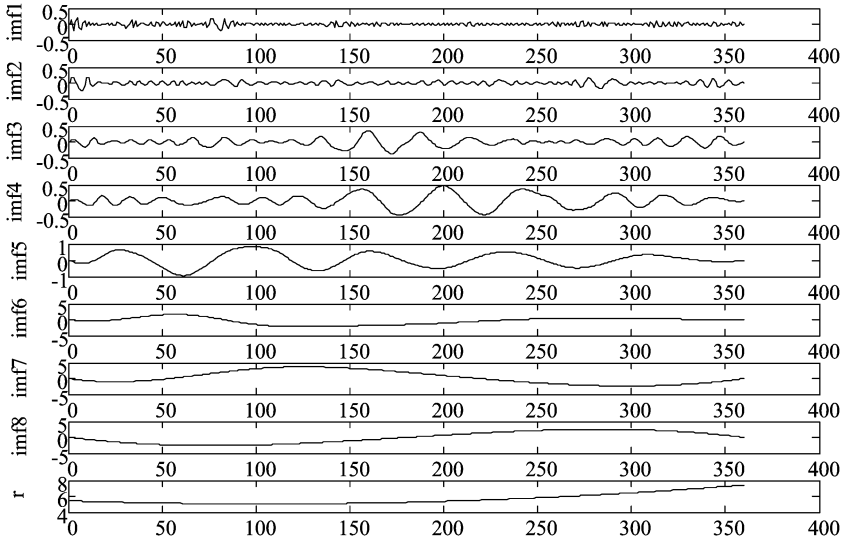


图2 8个不同尺度的本征模态函数 (IMF)

分解出的 8 个 IMF 频率各不相同, IMF1 到 IMF3 表现出随机无序性, 将其归为高频部分; IMF4 到 IMF8 具有很强的周期性波动性, 可归为低频部分, 残余项近似一条稍向上凸起的直线。

高频部分、低频部分和残余项分别隐含着不同的经济意义, 可用来揭示苹果批发价格序列中的内在特征。高频曲线围绕零均值线随机波动, 表明苹果批发价格受一些不确定性因素的随机干扰较强, 如自然气候、突发事件和国际农产品价格变化等, 而且曲线随机性波动的振幅越大表明苹果批发市场的均衡程度越高, 苹果批发价格的波动风险越大。低频曲线上下振动呈现出较为规则的正弦式波动, 表明苹果批发价格具有较强的季节性波动规律, 这与苹果生长的季节性规

律是相吻合的。随着时间推移, 曲线规律性逐渐减弱, 波动周期有变长趋势, 说明苹果批发价格的周期性波动是一个动态变化的过程。残余项则揭示了苹果批发价格处于一种稳定上升的长期趋势。

2.4 SVM 模型预测及评价分析

对分解后的各个 IMF 分量进行归一化, 选用径向基核函数并用交叉验证方法中的 K - fold Cross Validation (K - CV) 方法来确定参数, 最终确定优化后的惩罚因子  $C = 37$ , 参数  $g = 5.2$ , 优化取值的参数使预测模型具有较好的推广性。选取 2006 年 7 月 7 日到 2013 年 3 月 22 日的 350 个数据作为训练样本, 2013 年 3 月 29 日到 5 月 31 日的 10 个数据作为测试

样本,得到预测结果如图 3 所示。

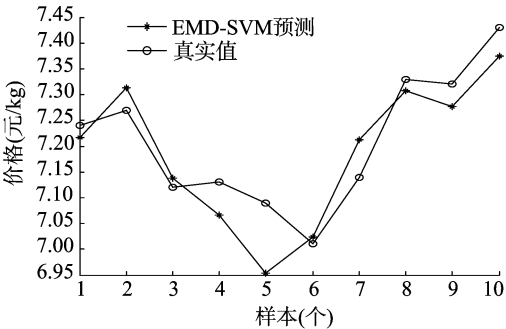


图3 全国苹果批发价格EMD-SVM建模预测拟合效果

表 4 EMD-SVM 与 BP 神经网络模型预测效果比较

序号	实际值 (元/kg)	EMD-SVM 模型			BP 神经网络		
		预测值 (元/kg)	误差值 (元/kg)	绝对百分 比误差(%)	预测值 (元/kg)	误差值 (元/kg)	绝对百分 比误差(%)
1	7.24	7.216 3	-0.023 7	0.33	7.163 6	-0.076 4	1.06
2	7.27	7.312 9	0.042 9	0.59	7.186 7	-0.083 3	1.15
3	7.12	7.138 0	0.018 0	0.25	7.180 1	0.060 1	0.84
4	7.13	7.066 3	-0.063 7	0.89	7.188 6	0.058 6	0.82
5	7.09	6.953 6	-0.136 4	1.92	7.190 5	0.100 5	1.42
6	7.01	7.023 3	0.013 3	0.19	7.195 7	0.185 7	2.65
7	7.14	7.212 1	0.072 1	1.01	7.160 2	0.020 2	0.28
8	7.33	7.307 0	-0.023 0	0.31	7.163 6	-0.166 4	2.27
9	7.32	7.276 1	-0.043 9	0.60	7.149 0	-0.171 0	2.34
10	7.43	7.374 7	-0.055 3	0.74	7.107 1	-0.322 9	4.35

3 结论与讨论

针对农产品价格非线性、非平稳的特点,本研究提出了一种新的预测方法,即基于 EMD 的 SVM 预测模型,对农产品价格序列做了预测分析,并与采用 BP 神经网络的预测值进行了对比。分析结果表明,EMD 是一种处理非线性、非平稳数据的强有力的工具,EMD-SVM 预测模型有着较高的拟合精度、较低的误差水平,能够有效提高农产品价格短期预测的预测精度。

要提高农产品价格预测的精度,还需要进一步完善以下研究。

(1) 预测模型的选择。农产品价格短期预测的建模主要依赖于历史数据记录,不同模型对数据的提取程度不一样,适用条件也不相同。因此要根据实际数据的规律,选择适合的预测模型。本研究中的苹果批发价格序列具有非线性、非平稳的特点,因此预测模型应具备较强的处理非线性、非平稳数据的能力,以提取序列的内在趋势和信息,保证预测精度。

(2) 突发性因素的考虑及模型的修正。基于时间序列的预测模型难以对价格变化原因作出解释,而突发性因素对价格变化的影响具有短暂性和脉冲性的特点,面对各种突发性因素时,如何才能及时对预测模型进行修正并量化研究其对农产品价格的长期影响是短期预测仍需研究的问题。

参考文献:

[1] 许世卫,李哲敏,董晓霞,等. 中国农产品在产销间价格传导机制研究[J]. 资源科学,2010,32(11):2092-2099.  
[2] 中国新闻网. 中国将进一步研究完善粮食最低收购价政策[EB/OL]. (2013-06-18)[2013-09-18]. <http://www.chinanews.com/gn/2013/06-18/4941829.shtml>.

用绝对百分比误差(absolute percentage error,APE)表示预测值与实际值的误差,计算公式如下:

$$APE = \frac{|预测值 - 实际值|}{实际值} \times 100\%$$

从图 3 可以看出,EMD-SVM 模型的拟合效果较为理想,预测值与真实值之间趋势一致,误差较小。经计算 EMD-SVM 模型的最大误差值仅为 0.136 4,平均绝对百分比误差仅为 0.68%。为了验证 EDM-SVM 预测模型的优越性,采用 BP 神经网络模型进行预测对比,结果表明 BP 神经网络预测值最大误差值为 0.322 9,最大绝对百分比误差达到 4.35%,平均绝对百分比误差为 1.72%,EMD-SVM 预测模型比 BP 神经网络具有更高的预测精度。预测对比结果见表 4。

[3] 刘海清,方佳. 基于指数平滑模型的海南省芒果价格预测[J]. 热带农业科学,2010,30(1):79-81.  
[4] 李干琼,许世卫,李哲敏,等. 农产品市场价格超短期预测研究——基于西红柿日批发价格的现代时间序列法建模[J]. 华中农业大学学报:社会科学版,2010(6):40-45.  
[5] 董晓霞,李干琼,刘自杰. 农产品市场价格短期预测方法的选择及应用——以鲜奶零售价格为例[J]. 山东农业科学,2010(1):109-113.  
[6] 李哲敏,李干琼. 禽蛋市场价格短期预测[J]. 中国食物与营养,2010(6):36-40.  
[7] 罗长寿. 基于神经网络与遗传算法的蔬菜市场价格预测方法研究[J]. 科技通报,2011,27(6):881-885,894.  
[8] 韩延杰. 一种基于模糊信息粒化和 GA-SVM 的农产品价格预测方法[J]. 农业网络信息,2012(11):16-20.  
[9] 刘慧,李宇辉. 我国小宗农产品价格波动趋势及其预测——以绿豆为例的分析[J]. 价格理论与实践,2012(6):57-58.  
[10] 阮连法,包洪洁. 基于经验模态分解的房价周期波动实证分析[J]. 中国管理科学,2012,20(3):41-46.  
[11] 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.  
[12] Yu L, Wang S Y, Lai K K. Forecasting crude oil price with an EMD-based neural network ensemble leaning paradigm [J]. Energy Economics,2008,30(5):2623-2635.  
[13] Li Z M, Li G Q, Wang Y T. Construction of short-term forecast model of eggs market price[J]. Agriculture and Agricultural Science Procedia,2010,1:396-401.  
[14] Lin C S, Chiu S H, Lin T Y. Empirical mode decomposition-based least squares support vector regression for foreign exchange rate forecasting[J]. Economic Modeling,2012,29(6):2583-2590.