

改进 RBF 神经网络在我国大豆价格预测中的应用研究

石 波,张冬青,马开平,刘 欢

(南京农业大学 工学院,江苏 南京 210031)

摘 要:我国大豆价格受国内外多种因素共同影响,具有非线性、随机性和高噪音等特点,采用传统数学模型进行预测,不仅分析难度大,预测误差也很大。RBF 神经网络以其优良的逼近性能而被广泛应用于非线性时间序列预测之中。本文提出一种基于遗传算法优化 RBF 神经网络的我国大豆价格预测模型,该模型为多维输入单维输出的多变量预测模型,模型的初始输入由大豆价格的历史数据和相关影响因素数据组成。采用遗传算法对 RBF 神经网络输入层节点数、基函数中心、扩展常数和输出层权值进行优化,模型可以从初始输入变量中自主选择最合适的输入变量组合作为模型的输入。采用 2009 – 2014 年的大豆价格数据进行预测研究,用 2009 – 2013 年的数据作为训练集,2014 年的数据作为测试集,改进 RBF 神经网络通过自主识别和选取中国大豆进口量、中国消费者信心指数和进口大豆到港分销价格 3 个因素作为相关影响因素的输入。结果表明:模型预测精度较高、泛化能力较强,能够很好捕捉大豆价格变化规律,可为大豆市场价格的准确预报提供参考借鉴。

关键词:大豆;价格;RBF;遗传算法

中图分类号:F713 **文献标识码:**A **DOI:**10. 11861/j. issn. 1000-9841. 2016. 02. 0310

Soybean Price Prediction in China Based on Improved RBF Neural Network

SHI Bo, ZHANG Dong-qing, MA Kai-ping, LIU Huan

(College of Engineering, Nanjing Agricultural University, Nanjing 210031, China)

Abstract: Soybean price in China is influenced by various factors at home and abroad. It is difficult to analyze and predict soybean price using the traditional mathematical model for which has the characteristics of nonlinearity, randomness and high noise. RBF Neural Network is widely used in nonlinear time series prediction for its excellent approximation performance. In this paper, one price prediction model of soybean is proposed based on improved RBF Neural Network model, which is multi-variable predictive model with multidimensional inputs and one-dimensional output. The initial inputs of model consist of historical data and the influencing factors of soybean. The node numbers of input layer, the centers and widths of Gaussian kernel and output layer weights are optimized by genetic algorithm. The improved model can select the most appropriate variables as inputs of the model automatically from the initial inputs. By using the data of soybean price between 2009 and 2014 to train and forecast. The improved RBF Neural Network select soybean imports of China, consumer confidence index of China and the distribution price in port of imported soybeans as the inputs of the related influence factors through automatically identification. The model's prediction error is 3. 64%, prediction results showed that the model prediction accuracy was high and generalization ability was strong. The model can capture the change rule of soybean price accurately. The model could provide reference for the accurate prediction of soybean market price.

Keywords: Soybean; Price; RBF Neural Network; Genetic algorithm

大豆作为主要粮食和油料作物^[1],其价格波动直接影响我国基本粮价的稳定,大豆价格受政策、供求、通胀、国际市场等多种因素共同影响,波动频繁,具有非线性、随机性和高噪音等特点,预测难度较大,所以对大豆价格进行预测研究具有重要意义。

RBF 神经网络具有生理学基础结构简单、学习速度快、逼近性能优良等特点,并能以任意精度逼近任意连续函数^[2-3],近年来 RBF 神经网络被广泛应用于农产品价格预测研究中。孙素芬等^[4]利用 RBF 神经网络对香菇市场价格数据进行预报。屠星月等^[5]采用 RBF 神经网络对农产品价格进行预

测研究。罗长寿等^[6]基于 RBF 神经网络,建立北京蔬菜批发价格预报模型。这些研究表明 RBF 神经网络能够很好捕捉农产品价格变化规律。

目前利用 RBF 神经网络对我国大豆价格进行预测研究鲜有报道。针对大豆价格和 RBF 神经网络特点,本文拟构建基于 RBF 神经网络的我国大豆价格预测模型。采用 RBF 神经网络模型对大豆价格进行预测研究时,需要确定合理的输入、网络结构和网络参数,它直接影响到模型的泛化能力和预测的精确性。遗传算法是一种高效、并行、全局搜索的优化方法^[7],全局搜索寻优能力显著。本文采

收稿日期:2015-08-04
基金项目:国家自然科学基金(71101072,71301077,71401076);南京农业大学中央高校基本科研业务费人文社会科学基金(SK2014011)。
第一作者简介:石波(1989-),男,硕士,主要从事生产运作管理、产品市场扩散等研究。E-mail:shibodengge@163.com。
通讯作者:马开平(1976-),女,博士,副教授,主要从事生产运作管理、产品市场扩散等研究。E-mail:makaiping@njau.edu.cn。

用遗传算法对 RBF 神经网络输入层节点数、基函数中心、扩展常数和输出层权值 4 个参数进行整合优化,模型可以从大豆相关影响因素中自主选择合适的输入变量组合作为模型的输入。

1 改进 RBF 神经网络模型

网络输入的质量对 RBF 神经网络至关重要,输入数据过多会降低模型的收敛速度,增加程序运行时间,变量中过多的冗余数据还会降低预测精度;过少的数据对输出变量的解释能力有限。常规多变量预测模型只考虑了预测对象的相关影响因素,并未考虑预测对象的历史时间序列。绝大部分系统的发展都是渐进的,研究对象的数值很难发生跳跃式的增长或降低,因此模型的输入不仅仅需要考虑相关影响因素,还应该考虑研究对象的历史信息,将研究对象的历史时间序列作为模型输入。该模型的输入由相关影响因素和研究对象的历史时间序列这两部分组成,预测模型可表示为:

$$\hat{y}_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-m}, X_t) + e_t \tag{1}$$

式(1)中, \hat{y}_t 为 RBF 神经网络期望输出变量,即大豆价格的预测值, e_t 为 t 时刻的观测噪声, $f(y_{t-1}, y_{t-2}, \dots, y_{t-m}, X_t)$ 为待估的函数,采用 RBF 神经网络来逼近,图 1 为 RBF 神经网络的拓扑结构。

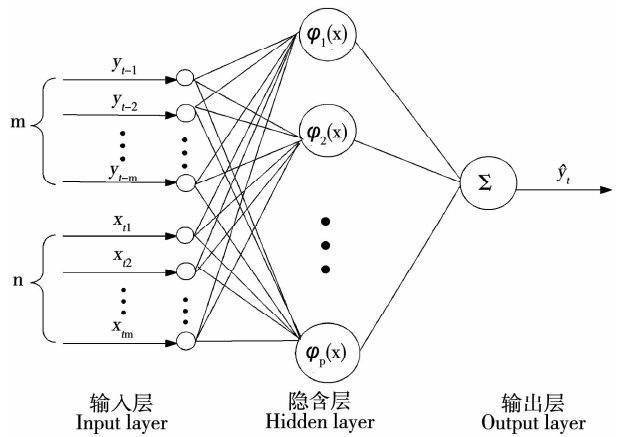


图 1 改进 RBF 网络的拓扑结构

Fig. 1 The topology of the improved RBF network

图 1 中, $(y_{t-i}, i = 1, 2, \dots, m)$ 和 $X_t = (x_{t1}, x_{t2}, \dots, x_{tm})$ 为模型的输入,前式为大豆价格的历史数据, m 为延迟阶数,表示选取了前 m 期的历史数据作为输入,后式为大豆价格的相关影响因素, n 为影响因素的维度, $m + n$ 为初始输入层节点数; $\varphi(x)$ 为基函数,一般采用高斯分布, $\varphi(r) = \exp(-\|X - C\|^2 / (2\sigma^2))$, C 为基函数中心, σ 为扩展常数; p 为隐含层的节点数目; l 为输出层的节点数目,其中 $l = 1$; w 是隐含层与输出层的连接权值。

要把该 RBF 神经网络用于大豆价格预测,还必

须先确定其结构和参数。如上所述,RBF 神经网络需要确定的参数为 m, n, C, σ, p, w ,其中 m, n, p 为模型的结构, C, σ, w 为模型参数。历史数据的延迟阶数 m 与研究对象有着密切关系,很难采用定量的方法确定,一般根据经验法则确定。隐含层节点数的确定异常复杂,实验表明,节点数过少,模型学习能力和信息处理能力较弱会降低预测精度;反之,节点数量太多,会增加模型的复杂性和计算时间,降低模型泛化能力^[8]。常用来确定隐含层节点数的方法有经验公式法^[9],聚类法^[10]和最大矩阵元法^[11]等。模型中隐含层节点数 p 根据 Segal 等^[9]的经验公式确定。

遗传算法是一种仿生算法,全局寻优能力显著,因此采用遗传算法对 n, C, σ, w 参数进行优化确定,算法过程如下:

第一步:计算隐含层输出,设输入变量 $I_t = (y_{t-1}, y_{t-2}, \dots, y_{t-m}, X_t)^T$,输出变量 $Y = (y_1, y_1, \dots, y_t)$,第 i 个隐含层节点在 t 时刻的输出为:

$$q_i |_t = \exp(-\|I_t - C_i\|^2 / (2\sigma_i^2)), i = 1, 2, \dots, p \tag{2}$$

第二步:求解输出,RBF 神经网络输出为隐含层节点输出的简单线性求和, t 时刻网络的输出如下:

$$y_t = \sum_{i=1}^p w_i q_i |_t + e_t \tag{3}$$

式(3)中 w_i 为第 i 个隐含层节点与输出节点的连接权值。

当参数都确定后,模型就可以开始预测。

2 基于遗传算法优化改进 RBF 神经网络参数流程

采用遗传算法对 RBF 神经网络输入层节点数、隐含层中心、扩展常数和连接权值进行整合优化,主要是利用遗传算法优良的全局搜索能力,由于遗传算法不采用路径搜索,而采用概率搜索,所以是概率意义上的全局搜索。

2.1 编码

由于所涉及的参数数量较多,且种类差异性较大,采用混合编码的方式进行编码,对相关影响因素输入的维度 n 采用二进制编码,0 表示该影响因素不作为 RBF 神经网络的输入,1 表示该影响因素作为 RBF 神经网络的输入;RBF 神经网络对隐含层中心 c 、扩展常数 σ 和连接权值 w 的精度要求较高,所以采用浮点编码进行编码,这样编码串不会太长,且解码方便,运算效率高。

图 2 为混合编码图,其中 $A_1 = (a_1, a_2, \dots, a_n)$,

它与 B_i 中的 $(c_{i21}, c_{i22}, \dots, c_{i2n})$ 相对应,并具有如下的关系。

$$c_{i2j} = \begin{cases} 0 & a_j = 0 \\ k_j & a_j \neq 0 \end{cases} \quad i = 1, 2, \dots, p \tag{4}$$

式(4)中, k_j 为基函数中心取值范围内的一个随机数值。

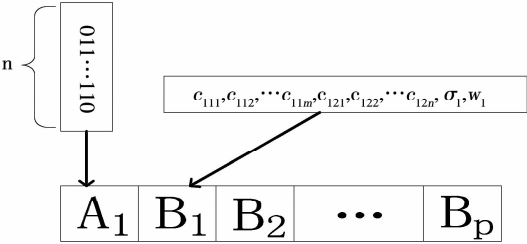


图2 编码示意图
Fig.2 The coding figure

2.2 适应度函数确定

RBF神经网络训练目的在于寻找出一些参数使期望输出和真实值之间的误差最小,选取均方差倒数作为适应度函数,第 i 个个体的适应度可以表示为:

$$Obvj_i = 1 / (\frac{1}{n} \sum_{k=1}^n (y_k - \hat{y}_k^i)^2) \tag{5}$$

式(5)中, y_k 表示预测对象真实值, \hat{y}_k 指的是RBF神经网络的预测值。

2.3 选择

本文的选择操作包括两步:第一步采用随机遍历抽样算法,它是具有零偏差和最小个体扩展的单状态抽样算法;第二步采用精英策略,把上代种群中最好的个体直接复制到下一代中,这样可以提高程序的运行效率和收敛速度。

2.4 交叉

由于采用混合编码,所以交叉分两个部分进行,将二进制编码部分和浮点编码部分分开进行交叉运算。交叉之前,先把染色体分为两部分,分别为二进制编码部分和浮点编码部分。对于二进制编码部分采用单点交叉算子进行交叉操作;浮点编码部分则采用算术交叉算子,算子如下:

$$\begin{cases} X_A^{t+1} = \alpha X_B^t + (1 - \alpha) X_A^t \\ X_B^{t+1} = \alpha X_A^t + (1 - \alpha) X_B^t \end{cases} \tag{6}$$

个体 X_A^t 和 X_B^t 进行线性交叉,产生新的个体 X_A^{t+1} 和 X_B^{t+1} ,交叉完成后将两部分染色体“拼接”为1个。

2.5 变异

变异操作流程跟交叉一样,先对二进制编码部分进行操作,然后再对浮点编码部分进行操作。二进制编码部分采用基本位变异算子;浮点编码部分

采用育种器变异算子,算法流程为:

(1)根据变异概率产生一中间任务表 $MuMtX$,决定变异的变量,并为加入的 $delta$ 所标识,其中 $delta$ 表示为:

$$delta = \sum_{i=1}^{g-1} a_i 2^{-i} \tag{7}$$

式(7)中, g 为控制变异算子的精度常数; a 是 $[0, 1]$ 内的一随机数,若小于 $1/g$ 时为1,若大于等于 $1/g$ 时为0。

(2)个体变异后的计算公式为:

$$mutaedvariable = variable \times MuMtX \times range \times MutOpt2 \times delta \tag{8}$$

式(8)中, $range = 0.5 \times$ 变量取值范围; $MutOpt2$ 是压缩变量的取值范围,是 $0 \sim 1$ 间的1个量,在此选为1。

3 基于改进 RBF 神经网络的我国大豆价格预测

参考刘欢等^[12]以国产大豆市场批发价格作为因变量,以国内大豆自给量、中国大豆进口量、世界大豆产量、中国居民消费价格指数、中国消费者信心指数、中国货币供应量、国内需求量和进口大豆到港分销价格8个因素作为大豆价格的初始相关影响因素。数据来源于中国统计年鉴、中国农产品价格调查年鉴、美国农业部 USDA 大豆月度供需报告和东方财富网等。采用上述8个相关影响因素值加上 m 阶大豆价格历史数据作为预测模型的输入,下个月的大豆价格作为预测模型的输出。

3.1 数据预处理

采用“极差归一化”法对数据进行预处理,这样不仅可以统一量纲和缩小数据数量级之间的差异,还可以防止计算的溢出以加快训练过程的收敛速度。

$$x'_i = \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)} \tag{9}$$

式(9)中, $\max(x_i)$ 和 $\min(x_i)$ 为第 i 个变量中的最大值和最小值,处理后 $x'_i \in [0, 1]$ 。

3.2 参数设置及输出结果

根据大豆价格数据特点,用2009年3月~2013年12月的月度数据作为训练集,2014年1月~2014年12月的数据作为测试集。

模型参数设置如下,大豆价格历史数据的延迟阶数 m 为3,交叉率 p_c 为0.9,变异率 p_m 为0.05,种群规模 $NIND$ 为50,迭代次数 $MAXGEN$ 为100,基函数中心的取值范围为 $(0, 1]$,扩展常数的取值范围是 $(0, 2]$,隐含层与输出层的连接权值的取值范

围为[0,1]。参照 Segal 等^[9]的算法,隐含层节点数 p 取值设为 15。

所有的参数都确定下来后,对大豆价格进行预测研究。运行程序可得训练误差随遗传算法迭代次数变化曲线,如图 3 所示,训练误差在 35 代左右开始收敛,最终的训练误差(均方差)为 0.006 5,训练精度较高。

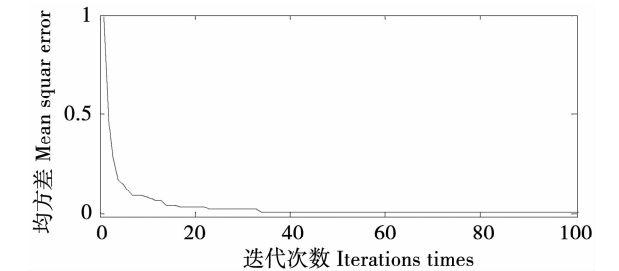


图 3 训练误差随迭代次数变化曲线

Fig. 3 The curves of the training error with iterations times

图 4 为大豆价格拟合值与训练样本值的拟合效果图,训练样本值和拟合值总体误差不大并且趋势一致,训练效果较好。

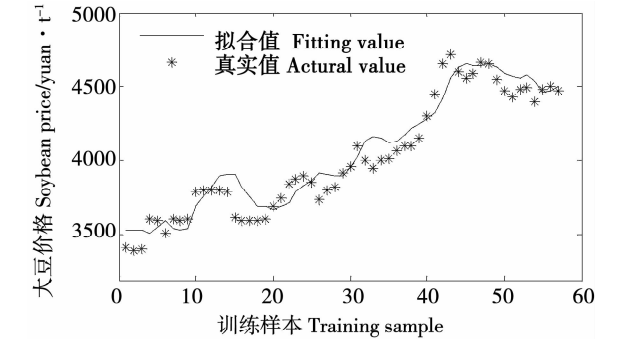


图 4 训练样本拟合效果图

Fig. 4 The fitting effect of training sample

模型从 8 个初始相关因素中最终选择中国大豆进口量、中国消费者信心指数、进口大豆到港分销价格 3 个因素作为相关影响因素输入。

3.3 模型检验

对大豆价格进行预测后,需要采用一些指标对模型的拟合进行检验,看模型是否能准确揭示、解释目标对象。如果检验没有通过,需要给出合理的物理解释,并修正模型。

3.3.1 预测精度

预测精度能够衡量所建立的改进 RBF 神经网络是否适用于大豆价格预测,具有现实意义,常采用平均相对误差的绝对值(MAPE)来衡量模型的预测精度。

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{X_i - \hat{X}_i}{X_i} \right| \times 100\%$$

式(10)中, X_i 和 \hat{X}_i 分别表示真实值和预测值,

MAPE 值越小,模型预测精度越高。

表 1 为大豆价格的预测值,12 个月平均相对误差的绝对值为 3.64%,预测精度较高,说明模型能够很好地捕捉大豆价格变化规律。

表 1 2014 年大豆价格真实值和预测值比较

Table 1 The comparison of soybean price's real value and predicted value in 2014

时间 Time (year - month)	真实值 Real value /yuan · t ⁻¹	预测值 Predictive value /yuan · t ⁻¹	预测误差 Prediction error /%
2014-1	4485	4599	2.54
2014-2	4420	4599	4.04
2014-3	4350	4574	5.14
2014-4	4495	4538	0.97
2014-5	4630	4543	1.87
2014-6	4595	4590	0.12
2014-7	4380	4626	5.61
2014-8	4380	4586	4.69
2014-9	4375	4564	4.33
2014-10	4335	4521	4.30
2014-11	4300	4521	5.13
2014-12	4285	4498	4.96

3.3.2 模型稳定性检验 遗传算法的输出结果并不是最优值,而是在其左右徘徊。模型的稳定性检验是衡量遗传算法中随机因素对预测结果的影响,若模型是稳定的,预测结果的波动较小。将模型运行 100 次,可得 100 个训练均方差,其散点图如图 5 所示。训练误差值的均值为 0.011 4,标准差为 0.002 1, $[\bar{x} - 3\sigma, \bar{x} + 3\sigma]$ 为 $[0.005\ 1, 0.177\ 2]$,由图 5 可知,所有的训练误差值皆落在 $[\bar{x} - 3\sigma, \bar{x} + 3\sigma]$ 范围内,说明模型是稳定的。

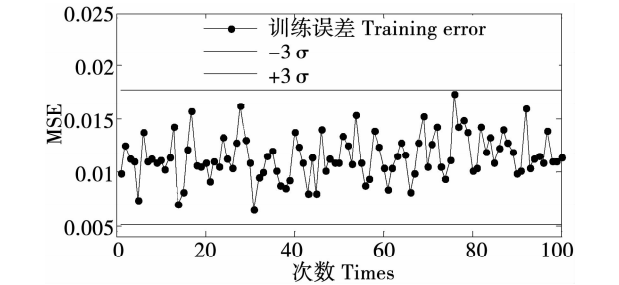


图 5 训练误差稳定性分析图

Fig. 5 The stability analysis diagram of training error

3.4 检测结果分析

采用改进 RBF 神经网络对我国大豆价格进行预测研究,模型训练误差小、预测精度高,模型能够及时、准确地捕捉到大豆价格变化的本质,说明所采用的预测方法是可行的。

大豆价格受国内外多种因素共同影响,其内部机理十分复杂,研究者很难掌握其系统机理,此时用常规数学方法对我国大豆价格进行预测,不仅分析难度大,预测误差也很大。采用遗传算法对 RBF 神经网络输入层节点数、基函数中心、扩展常数和输出层权值 4 个参数进行整合优化,改进 RBF 神经网络通过自主识别,选取中国大豆进口量、中国消费者信心指数、进口大豆到港分销价格 3 个因素作为相关影响因素的输入。消费者信心指数的高低直接影响其的购买欲望,大豆价格与消费者信心指数呈正相关关系,大豆价格随消费者积极性提升而走高。中国大豆进口量和进口大豆到港分销价格两个因素代表国际大豆市场对中国大豆价格的影响。近年来,中国已成为世界第一大豆进口国,大豆进口依赖度高达 80%^[13]。我国大豆进口量不断增加,国际大豆价格一直处于较低水位,其对我国大豆价格影响巨大,在某种程度上而言直接决定了我国大豆价格,也将直接影响到我国未来大豆价格走势。

4 结论与展望

对大豆价格进行科学预测有利于农户生产决策和政府宏观政策的制定。基于此,本文采用改进 RBF 神经网络对我国大豆价格进行预测研究,采用遗传算法对 RBF 神经网络输入层节点数、基函数中心、扩展常数和输出层权值 4 个参数进行整合优化。模型的输入由大豆价格 3 阶历史数据和相关影响因素数据构成,改进 RBF 神经网络通过自主识别,选取中国大豆进口量、中国消费者信心指数、进口大豆到港分销价格 3 个因素作为相关影响因素的输入,表明我国大豆价格主要受宏观经济因素和大豆供给因素影响。用 2009 年 3 月~2013 年 12 月的月度数据作为训练集,训练样本值和拟合值总体误差并不大并且趋势一致,训练效果较好;用 2014 年 12 个月的数据作为测试集,预测误差为 3.64%,预测精度高,说明模型能够很好地捕捉大豆价格变化规律,采用改进 RBF 神经网络对大豆进行预测研究是可行的。

本文在研究过程中难免会存在着一些不足之处,主要表现在以下两个方面:(1)关于大豆价格历史数据延迟阶数确定和遗传算法的收敛性未进行探讨。(2)大豆价格初始相关影响因素的选择借鉴了前人的研究成果,没有对其进行深入探讨。

参考文献

[1] 刘家富,周慧秋,李孝忠. 国内大豆市场价格波动及其影响

因素分析[J]. 东北农业大学学报(社会科学版), 2010, 8(4):10-13. (Liu J F, Zhou H Q, Li X Z. Analysis on price fluctuation and affecting factors of soybean in China[J]. Journal of Northeast Agricultural University(Social Science Edition), 2010, 8(4):10-13.)

[2] Fernández C, Soria E, Sánchez-Seiquer P, et al. Weekly milk prediction on dairy goats using neural networks[J]. Neural Computing and Applications, 2007, 16(4-5): 373-381.

[3] Lee C C, Chiang Y C, Shih C Y, et al. Noisy Time series prediction using M-estimator based robust radial basis function neural networks with growing and pruning techniques [J]. Expert System with Applications, 2009, 36: 4717-4724.

[4] 孙素芬, 罗长寿. 基于 RBF 神经网络的蔬菜价格预报研究[J]. 中国农学通报, 2011, 27(28):269-273. (Sun S F, Luo C S. Research on the prediction of vegetable price based on the RBF neural network model[J]. Chinese Agricultural Science Bulletin, 2011, 27(28):269-273.)

[5] 屠星月,薛佳妮,郭承坤,等. 基于时间序列与 RBF 的农产品市场价格短期预测模型[J]. 广东农业科学, 2014(23):168-173. (Tu X Y, XUE J N, Guo C K, et al. Short-term forecast of agricultural products price based on time series and RBF[J]. Guangdong Agricultural Science, 2014(23):168-173.)

[6] 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. (Luo C S. Research on the prediction method of vegetable price based on neural network and genetic algorithm[J]. Bulletin of Science and Technology, 2011, 27(6):881-885.)

[7] Holland J H. Adaptation in natural and artificial system [M]. Cambridge: MIT Press, 1975: 100-120.

[8] Niyogo P, Girosi F. On the relationship between generalization error, hypothesis complexity, and sample complexity for radial basis function[J]. Neural Computation, 1996, 8: 819-842.

[9] Segal R, Madnani M L, Madnani S. Radial basis function (RBF) network adaptive power system stabilizer[J]. IEEE Transactions on Power Systems, 2000, 5(2):722-727.

[10] 武方方,赵银亮. 一种基于蚁群聚类的径向基神经网络[J]. 西安交通大学学报, 2006, 40(4): 386-389. (Wu F F, Zhao Y L. Radial basis function neural network based on ant colony clustering[J]. Journal of Xi'an Jiaotong University, 2006, 40(4): 386-389.)

[11] 吴成茂,范九伦. 确定 RBF 神经网络隐层节点数的最大矩阵元法[J]. 计算机工程与应用, 2004, 20: 77-79. (Wu C M, Fan J L. Maximal mateix element method for determining the number of hidden nodes of RBF neural networks[J]. Computer Engineering and Applications, 2004, 20: 77-79.)

[12] 刘欢,张冬青. 基于分位数回归的国产大豆价格影响因素分析[J]. 大豆科学, 2014, 33(5):759-763. (Liu H, Zhang D Q. Analysis on influencing factors of domestic soybean price based on quantile regression[J]. Soybean Science, 2014, 33(5): 759-763.)

[13] 王雪尽. 我国大豆产业的危机及对策[J]. 经济问题, 2010(1):81-86. (Wang X J. Crisis and counter measures of soybean industry in China[J]. Journal of Economic Issues, 2010(1):81-86.)