

文 欢,陈 华. 基于 GNNM 的加工番茄产量预测研究[J]. 江苏农业科学,2017,45(11):167-170.

doi:10.15889/j.issn.1002-1302.2017.11.046

基于 GNNM 的加工番茄产量预测研究

文 欢,陈 华

(新疆大学电气工程学院,新疆乌鲁木齐 830047)

摘要:为提高加工番茄产量预测精度,提出一种串联型组合灰色神经网络模型(grey neutral net model,GNNM)的番茄产量预测方法。通过以新疆某加工番茄产业基地的历史数据作为研究对象,利用灰色关联分析法,获取与加工番茄产量相关性较大的影响因素,剔除非主要影响因素。以主要影响因素和灰色模型(grey model,GM)预测值作为神经网络模型输入,加工番茄产量作为输出,构建 GNNM 预测模型。仿真试验结果表明,新预测模型具有更好的逼近能力和预测精度,为加工番茄产量精准预测提供了一种新的方法。

关键词:加工番茄;灰色模型;神经网络;产量预测

中图分类号:S11 **文献标志码:**A **文章编号:**1002-1302(2017)11-0167-04

新疆独特的光热和水土条件成就了优质番茄的生长,为大规模的番茄酱加工创造了有利条件。加工番茄作为新疆的优势果蔬产业,种植规模达 7 万 hm^2 以上,已经达到全国总产量的 90%,番茄酱出口量接近全国贸易总量的 1/4,被誉为新疆的红色支柱产业^[1]。因此研究加工番茄产量预测模型可以为本地区番茄种植规划、番茄产业的宏观调控、加工及储藏等方面提供一定的参考价值。然而国内针对加工番茄产量预测的理论研究较少,已有的相关研究大都集中于线性回归模

型、移动平均法、指数平滑法和误差反向传播(back propagation,BP)神经网络等预测方法上。姜波等利用粒子滤波对番茄产量进行短期预测^[2];袁莉等运用灰色系统理论建立了加工番茄产量预测灰色模型^[3];尤坤鹏等采用马尔科夫预测技术建立番茄产量预测模型^[4];韩泽群等通过加权将 3 种统计分析方法加以组合建立了组合预测模型^[5]。综上所述,前人采用了不同的预测方法对加工番茄产量进行预测,并取得了一定的成果,但是在对加工番茄产量进行建模预测时,忽略了加工番茄产量与其影响因素之间存在着灰色性、不确定性和模糊性的关系,影响了预测精度和可靠性。

针对上述问题,本研究提出串联型组合灰色神经网络模型(grey neutral net model,GNNM),通过灰色关联分析法对影响加工番茄产量的因素进行全面定量分析,选择关联度较高的影响因素作为 GNNM 模型的输入节点,将 GM(1,1)与 BP 神经网络进行组合应用于加工番茄的产量预测,通过仿真结

收稿日期:2016-03-29

基金项目:国家自然科学基金(编号:61064005)。

作者简介:文 欢(1991—),女,四川南充人,硕士研究生,主要从事复杂约束环境下新疆番茄产量动态预测方法研究。E-mail:18799735522@163.com。

通信作者:陈 华,硕士,副教授,硕士生导师,主要从事智能控制与系统开发研究。E-mail:xj-chenhua@163.com。

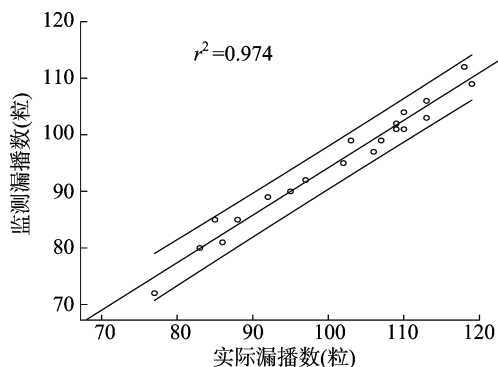


图6 实际值和监测值的线性拟合曲线

测、实时反馈、实时提醒,保证工作效率和作业质量。

参考文献:

- [1] 李 丹,耿端阳,马炳腾,等. 玉米精密播种机性能监测系统研究[J]. 农机化研究,2013(11):71-74.
- [2] Kocher M F, Lan Y, Chen C, et al. Opto-electronic sensor system for

- rapid evaluation of planter seed spacing uniformity[J]. Transactions of the ASAE,1998,41(1):237-245.
- [3] 孙全芳,戴玉华,刘东利. 精密播种机监控系统的研究现状与发展趋势[J]. 现代制造技术与装备,2005(4):60-62.
- [4] 和贤桃,郝永亮,赵东岳,等. 玉米精量排种器排种质量自动监测仪设计与试验[J]. 农业机械学报,2016,47(10):19-27.
- [5] 纪 超,陈学庚,陈金成,等. 玉米免耕精量播种机排种质量监测系统[J]. 农业机械学报,2016,47(8):1-6.
- [6] 孙 伟,王关平,吴建民. 勺链式马铃薯排种器漏播监测与补种系统的设计与试验[J]. 农业工程学报,2016,32(11):8-15.
- [7] 李 明,丁幼春,廖庆喜. 气力式油菜精量排种器田间漏播监测方法[J]. 农业工程学报,2010,26(增刊1):27-31.
- [8] 丛锦玲,余佳佳,曹秀英,等. 油菜小麦兼用型气力式精量排种器[J]. 农业机械学报,2014,45(1):46-52.
- [9] 张继成,陈海涛,欧阳斌林,等. 基于光敏传感器的精密播种机监测装置[J]. 清华大学学报(自然科学版),2013,53(2):265-268.
- [10] 亢武侯,亢冰洁,王少杰,等. 小麦精量播种高产栽培技术[J]. 种业导刊,2011(8):22-23.

果分析,该预测模型可进一步提高加工番茄产量的预测精度,为制定合理的种植规划和管理决策提供科学依据。

1 基本原理与方法

1.1 灰色预测模型

灰色预测又称为灰色系统预测,即通过对原始数据进行处理并分析其中的变化规律,生成具有较强规律性的数据序列,然后建立相应的灰色系统模型,从而定量预测系统的未来状态。灰色系统理论与方法的核心是灰色模型,而灰色模型是通过序列生成而建立的近似微分方程,是具有部分微分、部分差分性质的模型^[6]。灰色模型应用最广泛的是灰色序列模型,它采用一阶单变量灰微分型方程,即 GM(1,1)。

设非负原始序列为 $X^{(0)} = \{x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n)\}$, $X^{(0)}$ 的一阶累加生成序列为 $X^{(1)} = \{x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)\}$ 。 $X^{(1)}$ 的紧邻均值生成序列为 $Z^{(1)} = \{z^{(1)}(2), z^{(1)}(3), \dots, z^{(1)}(n)\}$, 则称 $x^{(0)}(k) = x^{(1)}(k) - x^{(1)}(k-1)$ 为 $X^{(1)}$ 上的灰导数。称灰微分方程

$$x^{(0)}(k) + ax^{(1)}(k) = b \quad (1)$$

为 GM(1,1),即只含一个变量且为一阶方程的模型,其对应的微分方程为

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b_0 \quad (2)$$

由微分方程理论可得,GM(1,1)在初始条件 $x^{(1)}(0) [x^{(1)}(0) = x^{(0)}(1)]$ 时的解,即灰色模型的时间响应式为

$$\hat{x}^{(1)}(k+1) = [x^{(1)}(0) - \frac{b}{a}]e^{-ak} + \frac{b}{a} \quad (3)$$

其中, a 为发展系数, a 的大小及符号反映 $X^{(0)}$ 及 $X^{(1)}$ 的发展态势, b 为灰作用量。

以 $k=2, 3, \dots, n$ 代入 GM(1,1) 中, 得方程组

$$\begin{cases} x^{(0)}(2) + az^{(1)}(2) = b \\ x^{(0)}(3) + az^{(1)}(3) = b \\ \vdots \\ x^{(0)}(n) + az^{(1)}(n) = b \end{cases} \quad (4)$$

将其用矩阵表示

$$\begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix} = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \quad (5)$$

$$\text{记 } Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}, B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix}, \hat{a} = \begin{bmatrix} a \\ b \end{bmatrix}, \text{ 则}$$

$Y = B\hat{a}$ 。可用最小二乘法得出:

$$\hat{a} = (B^T B)^{-1} B^T Y \quad (6)$$

最后还原值为

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k), k=0, 1, \dots, n_0 \quad (7)$$

即得到原始数据序列的预测值。

1.2 GNNM 预测模型

灰色预测的优点是所需样本数据少,不需要考虑其分布规律及变化趋势,建模比较简单,而且运算方便^[7];但是缺乏自学习、自组织和自适应能力,对非线性信息的处理能力较弱,不适合逼近复杂的非线性函数。而 BP 神经网络是由大量简单的神经元相互连接构成的复杂网络系统,其非线性系统具有很强的模拟能力^[8-9]和很强的学习、联想、容错、逼近任意函数的能力,但会因为过分逼近数据变化曲线上的细节而影响泛化能力。

灰色预测的少信息、建模简单及非线性处理能力弱等特征与 BP 神经网络预测大样本、非线性处理能力及学习能力强等特点具有优势互补性,将灰色模型与 BP 神经网络进行组合应用于产量预测,能够提高预测精度、可靠性和稳定性^[10-12]。因此,本研究提出一种串联型灰色神经网络模型,其原理是将加工番茄产量历史数据归一化后作为灰色预测模型的输入,通过计算得到历史数据序列的预测值,将预测值与影响加工番茄产量的主要因素统一综合到 BP 神经网络作为输入样本,实际值作为输出样本,采取一定的网络结构对神经网络进行训练,利用神经网络的非线性拟合能力和能够逼近任意函数的特性,得到一系列相应节点的权值和阈值,实现预测值与实际值的最佳拟合。

其网络拓扑结构如图 1 所示,其中 d_i 为灰色预测模型预测值输入参数, f_1, f_2, f_3 为主要影响因素输入参数; $\omega_{21}, \omega_{22}, \dots, \omega_{2n}, \omega_{31}, \omega_{32}, \dots, \omega_{3n}$ 为 BP 神经网络权值, y 为 GNNM 模型输出预测值。

2 加工番茄产量影响因素选择

加工番茄产量涉及到诸多外部和内部因素,为提高番茄产量的预测精度,需要对这些因素进行分析。表 1 为 2000—

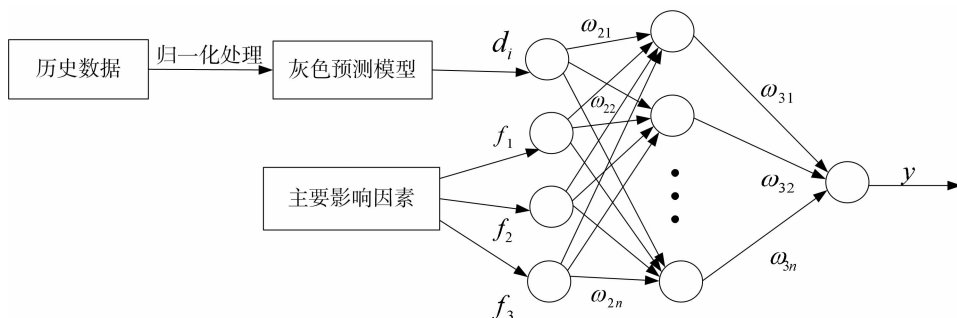


图1 GNNM 模型拓扑结构

2014 年新疆某加工番茄基地番茄产量 (N_0) 及其影响因素的历史数据,其中影响番茄产量的因素由第一产业劳动力

(N_1)、农业机械总动力 (N_2)、有效灌溉面积 (N_3)、化肥施用量 (N_4)、受灾面积 (N_5) 及成灾面积 (N_6) 等 6 个因素组成。

考虑到各因素之间的相关性会使番茄产量预测结果不准确,且不同的因素对其产量的影响程度不同,因此,为确定各因素对番茄产量的影响大小,本研究通过灰色关联分析法对各因素进行一定的主成分分析,从而得到影响加工番茄产量各因素所占的权重,对影响小的成分进行剔除,保留影响较大的因子。

2.1 参考数列与对比数列的确定

分别选取影响加工番茄产量的因素 $N_i (i=1,2,\cdots,6)$ 共 6 个变量作为参考数列,2000—2014 年加工番茄产量 N_0 作为对比数列,其中参考数列记为 $x_0(k)$,对比数列记为 $x_i(k)$ 。

2.2 数据无量纲化处理

数据量纲的不同会对分析结果产生一定的影响,为消除这种影响需要对数据进行无量纲化处理。常用的有标准化、初始化、极差法和最大化值等^[13],在此选用极差法进行处理,其处理公式为

$$x_{ij}' = \frac{x_{ij} - m_j}{M_j - m_j} (i=1,2,\cdots,7; j=1,2,\cdots,15)。 \quad (8)$$

其中, $M_j = \max_{1 \leq i \leq 7} \{x_{ij}\}$, $m_j = \min_{1 \leq i \leq 7} \{x_{ij}\} (j=1,2,\cdots,15)$, 则 $x_{ij}' \in$

$[0,1]$ 是无量纲的指标观测值。

2.3 确定关联系数与关联度

参考数列 $x_0(k)$ 与对比数列 $x_i(k)$ 的关联系数 $\varepsilon_i(k)$ 为

$$\varepsilon_i(k) = \frac{\min_k \min_i |x_i(k) - x_0(k)| + \rho \max_k \max_i |x_i(k) - x_0(k)|}{|x_i(k) - x_0(k)| + \rho \max_k \max_i |x_i(k) - x_0(k)|}, \quad (9)$$

其中, $|x_i(k) - x_0(k)|$ 为 $x_0(k)$ 和 $x_i(k)$ 第 k 个点的绝对误差; $\min_k \min_i |x_i(k) - x_0(k)|$ 为两级最小差; ρ 为分辨系数, $\rho \in (0,1)$, 一般取 $\rho=0.5$, 其作用一般在于提高关联系数间的差异显著性。对关联系数进一步处理后得到数列 x_i 对 x_0 的关联度 γ_i , 其公式为

$$\gamma_i = \frac{1}{n-1} \sum_{k=1}^n \varepsilon_i(k)。 \quad (10)$$

计算得出番茄产量与影响因素之间的灰色关联度,选取关联度较高的第一产业劳动力、有效灌溉面积和化肥施用量作为灰色神经网络的输入节点。

表 1 2000—2014 年番茄产量及其影响因素

年份	N_0 ($\times 10^5$ kg/hm ²)	N_1 ($\times 10^6$ 人)	N_2 (kW)	N_3 ($\times 10^6$ hm ²)	N_4 (t)	N_5 (hm ²)	N_6 (hm ²)
2000	1.689 325	3.879	8.512×10^6	3.094 3	7.92×10^5	6.40×10^5	3.59×10^5
2001	1.882 350	3.882	8.809×10^6	3.138 1	8.33×10^5	1.10×10^6	5.81×10^5
2002	2.091 465	3.918	9.196×10^6	3.053 9	8.43×10^5	6.93×10^5	3.43×10^5
2003	2.309 130	3.972	9.727×10^6	3.051 0	9.07×10^5	7.68×10^5	5.16×10^5
2004	2.479 395	4.033	1.047×10^7	3.106 6	9.92×10^5	7.42×10^5	4.36×10^5
2005	2.818 815	4.080	1.116×10^7	3.204 3	1.08×10^6	5.17×10^5	3.52×10^5
2006	2.948 190	4.145	1.190×10^7	3.326 1	1.18×10^6	7.35×10^5	5.79×10^5
2007	3.265 545	4.177	1.275×10^7	3.465 4	1.32×10^6	9.50×10^5	5.68×10^5
2008	3.547 890	4.213	1.376×10^7	3.572 5	1.49×10^6	2.17×10^6	1.33×10^6
2009	3.718 965	4.275	1.503×10^7	3.675 7	1.55×10^6	1.24×10^6	7.78×10^5
2010	4.034 520	4.381	1.644×10^7	3.721 6	1.68×10^6	1.31×10^6	6.49×10^5
2011	4.532 152	4.803	1.722×10^7	3.874 0	1.71×10^6	1.11×10^6	5.43×10^5
2012	5.013 367	5.006	1.900×10^7	3.966 5	1.79×10^6	9.70×10^5	6.70×10^5
2013	5.731 695	5.307	1.953×10^7	4.100 1	1.82×10^6	8.30×10^5	4.15×10^5
2014	5.731 695	5.307	2.187×10^7	4.200 6	1.90×10^6	6.50×10^5	3.88×10^5
关联度		0.949 0	0.902 3	0.850 5	0.937 4	0.641 0	0.649 0

3 GNNM 模型实例验证

3.1 GM(1,1) 预测模型

利用 MATLAB 软件编程建立 GM(1,1), 对表 1 中 2000—2010 年的番茄产量数据进行预测。求得模型的系统发展系数 $a = -0.081 9$, 内生控制变量 $b = 174 101.8$, 则 GM(1,1) 的白化响应式为

$$\hat{x}^{(1)}(k+1) = 2 294 203.81e^{-0.081 9k} - 2 125 271.3 (k=0, 1, 2, 3, \cdots)。 \quad (11)$$

通过 MATLAB 软件求得最终预测值(表 2)。

3.2 建立 GNNM

将表 2 中 GM(1,1) 预测的 2000—2010 年产量数据和表 1 中灰色关联度较高的 3 个影响因素同时作为网络训练样本, 2011—2014 年的历史产量数据作为测试样本, 使用 MATLAB 软件建立 GNNM, 并与 GM(1,1)、BP 网络进行比较分析。

由于表 1 中数据间数量级存在差别, 为避免因为输出数

表 2 GM(1,1) 预测 2000—2010 年番茄产量值

年份	产量预测值 ($\times 10^5$ kg/hm ²)	年份	产量预测值 ($\times 10^5$ kg/hm ²)
2000	1.689 325	2006	2.949 970
2001	1.958 533	2007	3.201 805
2002	2.125 731	2008	3.475 139
2003	2.307 202	2009	3.771 807
2004	2.504 165	2010	4.093 802
2005	2.717 942		

据数量级差别而造成网络预测误差较大, 故先对数据按照公式(12)进行归一化处理。

$$x'_k = \frac{x_k - x_{\min}}{x_{\max} - x_{\min}} (k=1, 2, 3, \cdots)。 \quad (12)$$

其中, x_{\max} 为序列中的最大值, x_{\min} 为序列中的最小值, 最终把所有数据转化为 $[0,1]$ 之间的数。

本研究选择 3 层 BP 神经网络, 经过多次训练, 根据网络预测输出与期望输出的误差调整权值和阈值, 得出误差最小

的隐含层节点数为 10。由于本研究中输入数据为 4 维,输出为 1 维,故网络结构为 4-10-1。输入层与隐含层的传输函数为 Sigmoid 函数 $f(x) = \frac{1}{1 + e^{-x}}$,隐含层与输出层传递函数为 Purelin 函数。设定训练精度为 0.000 01,训练次数为 1 000,训练方法采用动量法和学习率自适应调整策略。

对设计好的 GNNM 进行训练,训练稳定后得到的均方误差在迭代 10 次时达到 9.9×10^{-6} ,网络平滑地收敛于全局极小值(图 2)。

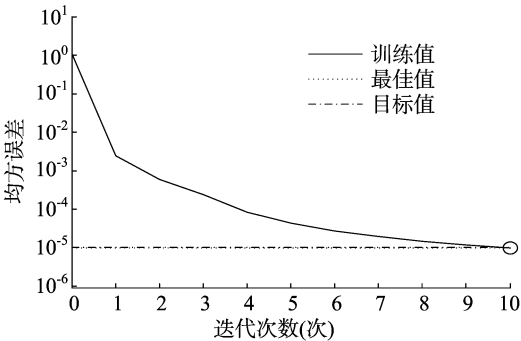


图2 GNNM 网络训练误差曲线

3.3 仿真结果分析

图 3 为 GM、BP 以及 GNNM 等 3 种模型预测效果图,从图 3 中可以明显看出,经过 GNNM 模型训练后的数据更逼近

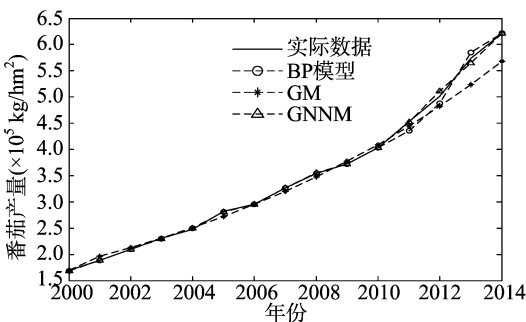


图3 不同模型预测效果

于实际数据,2010—2014 年番茄产量预测值与实际值拟合程度较好。

为了进一步验证 GNNM 的预测效果,本研究用 GM(1,1)、BP 神经网络、GNNM 对 2011—2014 年番茄产量预测结果进行分析对比。从表 3 可知,采用 GM(1,1)、BP 神经网络和 GNNM 的预测结果与实测值的相对误差绝对值均小于 10%,其均方误差分别为 7.68%、1.98%、0.61%。用 GM(1,1) 进行加工番茄产量预测时,其方法虽然简单,但是预测精度并不高。对于 BP 神经网络模型来说,加工番茄产量预测精度有所提高,但个别年份预测精度并没有改善。相比较而言,GNNM 的拟合精度、预测效果均优于 GM(1,1) 和 BP 神经网络预测模型,更符合准确预测精度要求。因此,选择 GNNM 模型进行加工番茄产量预测是可行、高效的。

表 3 基于不同模型的 2011—2014 年番茄产量预测值与实际值比较

年份	实测值 (kg/hm ²)	预测值(kg/hm ²)			误差(%)		
		GM(1,1)	BP 神经网络	GNNM	GM(1,1)	BP 神经网络	GNNM
2011	453 215.2	444 328.5	439 259.2	452 621.4	1.96	3.08	0.130
2012	501 336.7	482 260.3	514 730.3	501 082.8	3.81	2.67	0.050
2013	573 169.5	523 430.2	574 327.6	582 545.8	8.67	0.20	1.640
2014	621 433.5	568 114.8	621 378.7	621 422.7	8.58	0.01	0.002
均方差					7.68	1.98	0.610

4 结论

本研究利用 GM(1,1) 与 BP 神经网络各自的优点组合进行加工番茄产量预测,同时考虑到影响加工番茄产量的 6 个因素,采用灰色关联度分析法选取其中 3 个关联度高的因素作为 BP 神经网络的输入,构建了以新疆某加工番茄基地为例的 GNNM,预测加工番茄产量。仿真结果表明,与 GM(1,1) 和 BP 神经网络单一预测模型相比,GNNM 的均方差更小、精度更高,更符合加工番茄产量预测,对农业生产实践具有指导意义。

参考文献:

[1]唐世辉,霍彬,朱平. 新疆番茄产业发展现状分析[J]. 合作经济与科技,2014(11):8-10.
[2]姜波,许竞. 粒子滤波在农作物产能短期预测中的应用[J]. 江苏农业科学,2012,40(3):357-358.
[3]袁莉,姜波. 基于灰色系统理论的加工番茄产量预测模型研究[J]. 农业科学与技术,2011,12(5):632-633,642.
[4]尤坤鹏,姜波. 马尔科夫链在加工番茄产量预测中的应用研究[J]. 安徽农学通报,2012,18(15):29-30,38.

[5]韩泽群,姜波. 加工番茄产量组合预测模型研究[J]. 中国农学通报,2013,29(3):161-165.
[6]郭秀英. 预测决策的理论与方法[M]. 北京:化学工业出版社,2010.
[7]Liu S F,Tao L Y,Xie N M,et al. On the new model system and framework of grey system theory[C]//Grey Systems and Intelligent Services (GSIS),IEEE International Conference,2015.
[8]Hsu C C,Chen C Y. Applications of improved grey prediction model for power demand forecasting[J]. Energy Conversion and Management,2003,44(14):2241-2249.
[9]李吉龙,刘新平. 基于 BP 神经网络的未利用地开发生态风险预警研究——以新疆阜康市为例[J]. 江苏农业科学,2015,43(10):500-503.
[10]刘静,傅泽田,张小栓. 基于 GM-RBF 神经网络的冷链运输环境预测[J]. 江苏农业科学,2015,43(10):498-500.
[11]张婷. 基于灰色神经网络组合模型的能源需求预测[D]. 天津:天津大学,2007.
[12]史峰,王小川,郁磊,等. MATLAB 神经网络 30 个案例分析[M]. 北京:北京航空航天大学出版社,2010:258-267.
[13]何乃强,惠晓滨,周漩. 基于正态区间估计的改进型无量纲化方法[J]. 计算机工程与应用,2012,48(5):236-238.