

LM-BP 神经网络在农业总产值预测的应用

张自敏¹, 樊艳英², 陈冠萍²

(1. 贺州学院教育技术中心, 广西贺州 542899; 2. 贺州学院计算机科学与信息工程学院, 广西贺州 542899)

摘要 农业生产总值是衡量一个地区农业发展水平的重要指标, 农业生产总值受多方因素的影响, 具有非线性特征, 为此, 提出了 LM-BP 神经网络预测农业生产总值的模型及方法。以农作物播种面积、粮食产量、甘蔗产量、木薯产量、茶叶产量、肉类产量、水产品产量、松脂产量及油茶籽产量等与农业生产总值相关指标作为网络输入, 通过广西 2000~2012 年农业生产总值数据仿真试验分析表明, LM-BP 神经网络预测结果与实际值有较好的拟合度。

关键词 农业生产总值; 人工神经网络; LM-BP 神经网络; 预测

中图分类号 S126; TP183 **文献标识码** A **文章编号** 0517-6611(2014)28-10009-03

Application of LM-BP Neural Network in Predicting Gross Agricultural Product

ZHANG Zi-min¹, FAN Yan-ying², CHEN Guan-ping² (1. Center of Education Technology, Hezhou University, Hezhou, Guangxi 542899; 2. School of Computer Science and Information Engineering, Hezhou University, Hezhou, Guangxi 542899)

Abstract Gross agricultural product is an important indication to measure the agricultural development level of a region. It would be affected by many factors, owning the character of non-linearity. For this reason, LM-BP neural network was put forward as the model and method for predicting gross agricultural product. Taking the indications of the sown area of crop, the output of grain, sugarcane, cassava, tea, meat, aquatic products, turpentine and oil-tea camellia seed, etc. as inputs, during 2000 to 2012 in Guangxi, the gross agricultural product data from the analysis of simulation experiment shows that the prediction of LM-BP neural network fits well with actual results.

Key words Gross agricultural product; Artificial neural networks; Levenberg Marquardt Back Propagation (LM-BP) neural network; Prediction

农业总产值是指以货币价值形式反映一定时期内农业生产的总成果和总规模, 是衡量一个地区农业生产力和发展水平的重要指标。正确预测农业总产值为解决“三农”问题提供决策参考数据, 农业生产总值的预测一直受到人们关注。预测农业总产值主要运用统计学建模的方法。张洁瑕等提出利用自适应 ARMA 建立农业总产值预测模型^[1]。张显周等提出 ARIMA 的时间序列预测模型, 并对我国农业总产值作了预测^[2]。刘楠等提出利用灰色理论并建立灰色预测模型对农业经济进行了预测^[3-5]。农业总产值受农业种植业、农业畜牧渔业及林业等诸多因素的影响, 因此, 农业总产值具有非线性的特点。运用统计学建模的方法预测农业总产值存在模型选择困难及预测精度不高而造成预测效率低下的缺点。1986 年美国学者 Rumelhart 等提出误差反向传播算法^[6] (Error Back Propagation, BP), 简称 BP 算法, 是一种误差反向传播算法的多层前馈网络, 其具有很强的非线性映射能力, 是应用最广泛的人工神经网络。但 BP 网络存在收敛速度慢、易陷入局部最小和泛化能力差等缺点, 限制其的应用。

LM-BP 网络是对 BP 网络的改进, 利用二阶导数的信息, 在函数逼近最优值附近产生一个理想搜索方向, 克服 BP 网络朝单一方向盲目搜索的缺点, 大大提高了网络收敛速度。为此, 笔者建立了关于农业总产值的 LM-BP 神经网络预测模型, 运用广西农业生产主要指标作为输入训练 LM-BP 神经网络并预测了农业总产值, 通过与标准 BP 神经网络预测模型比较分析显示, LM-BP 神经网络预测农业生产总值有较强的

拟合性、较高的效率及较高的预测精度, 为农业生产总值提供一种精准可靠的预测方法。

1 BP 神经网络

1.1 BP 神经网络原理 处理信息由输入层经网络隐层结点计算每个结点实际输出, 若此时实际输出与期望相符, 则网络学习结束; 否则转入反向传播。误差反向传播, 信息从输出层经隐层结点处理向输入层逐层返回, 计算每节点实际输出与期望的误差, 根据误差调整各层结点连接权值。

BP 算法采用最速下降求解, 只采用了一阶导数的信息, 虽然方法简单但网络收敛速度慢, 易陷入局部极小值, 在网络训练学习中难以避免学习收敛速度慢及训练过程出现振荡的问题。为了加快网络的收敛速度一些改良方法引入了二阶导数的信息对 BP 算法加以改进, 如共轭梯度法、牛顿法和 LM 算法等, 其中 LM 算法利用了二阶导数的信息, 是高斯-牛顿法的改进。LM 算法主要思想是对高斯-牛顿法增加了一个非负对角矩阵, 使用强迫正定的策略。

1.2 LM-BP 算法 LM 算法是高斯-牛顿法的改进, 利用高斯-牛顿在函数逼近最优值附近产生一个理想搜索方向, 通过自适应调整网络权值, 克服负梯度下降法朝单一方向盲目搜索的缺点, 可大大提高网络收敛速度。

设 w_k 为第 k 次迭代权值和阈值组成的向量, 则第 $k+1$ 次权值更新为:

$$w_{k+1} = w_k + \Delta w \quad (1)$$

对于误差函数 $E(w)$ 有:

$$E(w) = \frac{1}{2} \sum_i^N (t_i - o_i)^2$$

$$= \frac{1}{2} \sum_i^N e_i^2 \quad (2)$$

式中, N 为输出向量维数; t_i 为输出层第 i 个输出神经元的目

基金项目 广西高校科学技术研究项目(2013LX143)。

作者简介 张自敏(1976-), 男, 广西贺州人, 讲师, 硕士, 从事人工智能和数据挖掘工作。

收稿日期 2014-08-21

标输出; o_i 为该神经元的实际输出。

对于牛顿法则有:

$$\Delta w = -H_k^{-1} g_k \quad (3)$$

式中, H_k 为误差函数 $E(w)$ 的黑塞尔(Hessian)矩阵; g_k 为梯度。

$$g = J^T(w) e(w) \quad (4)$$

$$H = J^T(w) J(w) e(w) + S(w) \quad (5)$$

$$s(w) = \sum_{k=1}^N e_i(w) \nabla^2 e_i(w) \quad (6)$$

式中, $e(w) = [e_1(w), e_2(w), L, e_N(w)]^T$; J 为雅可比矩阵。

$$J = \begin{bmatrix} \frac{\partial e_1(w)}{\partial w_1} & \frac{\partial e_1(w)}{\partial w_2} & \dots & \frac{\partial e_1(w)}{\partial w_n} \\ \frac{\partial e_2(w)}{\partial w_1} & \frac{\partial e_2(w)}{\partial w_2} & \dots & \frac{\partial e_2(w)}{\partial w_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial e_N(w)}{\partial w_1} & \frac{\partial e_N(w)}{\partial w_2} & \dots & \frac{\partial e_N(w)}{\partial w_n} \end{bmatrix} \quad (7)$$

黑塞尔矩阵近似为:

$$H \approx J^T(w) J(w) \quad (8)$$

因此:

$$\Delta w = -[J^T(w) J(w)]^{-1} J^T(w) e(w) \quad (9)$$

LM算法是对高斯-牛顿法的改进,克服高斯-牛顿在黑塞尔矩阵求逆不稳定而造成网络不可收敛的问题。LM算法通过对式(8)修改得到:

$$H \approx J^T J + uI \quad (10)$$

式中, u 为很小数; I 为 $N \times N$ 的单位矩阵。因此LM算法网络权值更新为:

$$w(k+1) = w(k) - [(J_k^T(w) J(w) + uI)]^{-1} J(w) e(w) \quad (11)$$

式中, u 为零时就是高斯-牛顿法; u 值很大时LM近似为最速下降梯度法。在实际操作中, u 值是试探性的动态调整,当求得的 Δw 使得误差函数 $E(w)$ 减小,则减小 u 值,否则,增大 u 值。用式(11)修改一次权值需要求 n 阶的代数方程(n 为网络中权值数目),LM算法的时间复杂度为 $O(n^3/6)$,若 n 很大,则计算量和存储量都非常大,然而,每次迭代效率显著提高,可大大改善整体性能,特别是在精度要求高时。

LM-BP算法描述:①给定 $\varepsilon > 0$,初始 w ,设置网络学习率及 u 值参数;②计算网络输出,若实际输出与目标值之差小于 ε ,则转⑤,否则执行③;③利用式(7)计算雅可比矩阵 J ;④利用式(9)实现更新网络权值,跳转②;⑤结束。

2 LM-BP神经网络预测农业总产值的模型

2.1 LM-BP神经网络结构 LM-BP神经网络由3层网络构成,即输入层、隐层及输出层,输入层是与输出期望相关的因子变量,其结点数由解决问题相关因素决定,输出层是目标期望输出,隐层结点一般由经验公式取得。广西地处我国亚热带,是亚热带主要农作物的生产省区,经分析,影响广西农业总产值的指标主要有农作物播种面积 x_1 、粮食产量 x_2 、甘蔗产量 x_3 、木薯产量 x_4 、茶叶产量 x_5 、肉类产量 x_6 、水产品产量 x_7 、油茶子产量 x_8 及松脂产量 x_9 ,当年的农

业总产值 y 作为网络的输出。因此,可以确定LM-BP神经网络输入结点为9个,输出结点为1个,中间隐层由式(12)得:

$$l_1 = \sqrt{m+n} + c \quad (12)$$

式中, l_1 为隐层结点; m 为网络输入结点; n 为网络输出结点; c 取1至10之间的数。该研究中 m 为9, n 为1, c 取5,由公式(9)得出隐层结点为8个,这样就构成农业总产值的LM-BP神经网络预测模型,模型结构如图1所示。

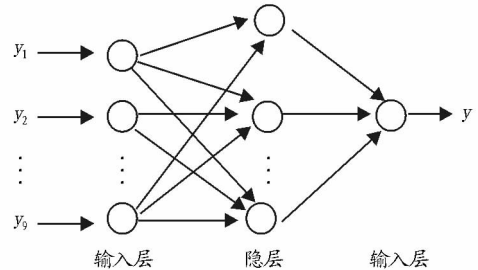


图1 LM-BP神经网络预测农业总产值模型

2.2 LM-BP神经网络在MATLAB实现 对以上LM-BP神经网络模型结构利用MATLAB^[7]实现建模,MATLAB建立BP神经网络模型包括对网络输入层、中间隐层、输出层结点、网络传递函数及训练函数等参数设置。由以上分析,该研究采用 $9 \times 8 \times 1$ 的网络结构,即输入结点为9个,中间隐层结点8个,输出结点为1个;隐层传递函数采用logsig函数,输出层为purelin函数;训练函数为trainglm函数,训练目标误差精度 10^{-3} ,网络训练最大迭代次数为20000次。MATLAB7.0编写程序训练BP神经网络,程序如下:

```
net = newff( minmax( x ), [ 8, 1 ], { 'logsig', 'purelin' }, 'trainglm');
net.trainParam.epochs = 20000;
net.trainParam.goal = 1e-3;
net = train( net, x, y );
```

2.3 BP网络训练 该研究选取2000~2012年的广西农业生产主要指标^[8]的有关数据检验LM-BP神经网络预测广西农业生产总值的模型,数据摘自《广西统计年鉴》(2001~2013年,中国统计出版社)。表1给出了广西相应年份的农业生产总值及相关指标的数据。

该研究采用一年时滞预测农业总产值,即将上一年各生产指标作为当年生产总产值的输入数据,当年的总产值作为训练输出目标,由此可以将表1构造以下训练及测试样本,结果见表2所示。

首先将学习样本利用premnmx函数进行归一化使数据落在 $[-1, 1]$ 区间,将预处理的学习样本作为输入训练网络,图2、3分别显示了LM-BP和标准BP两种网络训练误差变化曲线。从图2可以看出,LM-BP网络经过4次迭代之后网络学习达到预定误差要求,而标准BP网络要经过6875次的迭代之后才达到预定误差要求,该研究LM-BP网络的收敛速度要比标准BP网络的要快得多,说明LM-BP预测农业总产值具有更高的效率。

表 1 2000~2012 年广西农业生产主要指标

年份	农林牧渔业总产值 亿元	农作物总播种面积 km ²	粮食产量 万 t	甘蔗产量 万 t	木薯产量 万 t	茶叶产量 万 t	肉类产量 万 t	水产品产量 //万 t	油茶子产量 万 t	松脂产量 万 t
2000	828.97	6 258.6	1 667.24	2 937.87	132.56	1.79	287.26	239.86	11.86	21.60
2001	872.90	6 196.7	1 607.35	3 653.33	128.02	1.89	306.98	247.77	13.46	21.67
2002	916.50	6 164.4	1 549.38	4 593.38	149.21	1.96	329.67	255.14	10.85	22.53
2003	1 030.89	6 107.4	1 484.82	4 861.84	139.73	2.13	353.52	264.61	9.27	25.21
2004	1 294.53	6 172.9	1 473.19	5 003.87	152.47	2.24	383.05	274.31	11.54	28.24
2005	1 448.37	6 343.9	1 516.29	5 154.69	173.61	2.62	418.60	284.19	11.74	30.19
2006	1 622.22	5 557.3	1 538.99	5 924.83	187.59	2.85	445.35	296.42	8.61	24.72
2007	2 026.22	5 594.4	1 396.60	7 737.47	155.90	3.43	329.04	246.06	12.26	43.90
2008	2 389.79	5 695.6	1 394.70	8 215.58	154.67	3.35	350.68	249.98	12.72	42.25
2009	2 380.51	5 834.2	1 463.20	7 509.44	164.12	3.66	371.00	261.81	13.34	46.99
2010	2 720.99	5 896.9	1 412.32	7 119.62	173.21	3.92	387.77	275.09	14.37	49.58
2011	3 323.37	5 996.5	1 429.93	7 269.96	180.33	4.44	391.09	288.82	15.10	53.29
2012	3 490.72	6 089.5	1 484.90	7 829.71	181.31	4.94	410.99	303.47	16.39	55.71

表 2 LB-BP 网络训练样本

样本 序号	目标输出 y	样本输入 x								
		x ₁	x ₂	x ₃	x ₄	x ₅	x ₆	x ₇	x ₈	x ₉
1	872.90	6 258.6	1 667.24	2 937.87	132.56	1.79	287.26	239.86	11.86	21.60
2	916.50	6 196.7	1 607.35	3 653.33	128.02	1.89	306.98	247.77	13.46	21.67
3	1 030.89	6 164.4	1 549.38	4 593.38	149.21	1.96	329.67	255.14	10.85	22.53
4	1 294.53	6 107.4	1 484.82	4 861.84	139.73	2.13	353.52	264.61	9.27	25.21
5	1 448.37	6 172.9	1 473.19	5 003.87	152.47	2.24	383.05	274.31	11.54	28.24
6	1 622.22	6 343.9	1 516.29	5 154.69	173.61	2.62	418.60	284.19	11.74	30.19
7	2 026.22	5 557.3	1 538.99	5 924.83	187.59	2.85	445.35	296.42	8.61	24.72
8	2 389.79	5 594.4	1 396.60	7 737.47	155.90	3.43	329.04	246.06	12.26	43.90
9	2 380.51	5 695.6	1 394.70	8 215.58	154.67	3.35	350.68	249.98	12.72	42.25
10	2 720.99	5 834.2	1 463.20	7 509.44	164.12	3.66	371.00	261.81	13.34	46.99
11	3 323.37	5 896.9	1 412.32	7 119.62	173.21	3.92	387.77	275.09	14.37	49.58
12	3 490.72	5 996.5	1 429.93	7 269.96	180.33	4.44	391.09	288.82	15.10	53.29

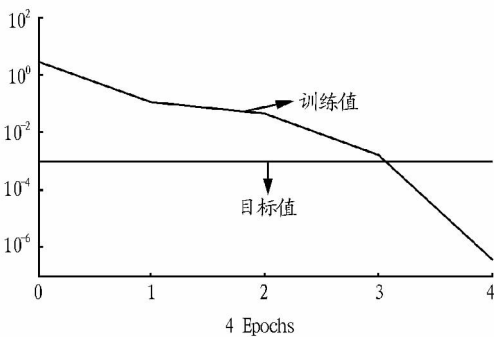


图 2 LM-BP 训练误差曲线

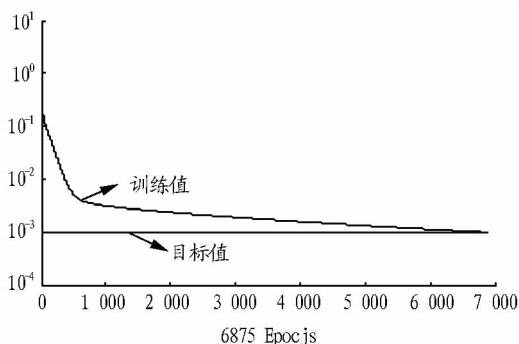


图 3 标准 BP 训练误差曲线

3 结果与分析

将样本分别输入 LM-BP 神经网络和标准 BP 神经网络进

行仿真预测比较,结果见表 3。从表 3 可以看出,LM-BP 网络模型预测农业生产总值平均误差小,最大误差为 0.08%,最小误差为 0%,预测值与实际值非常接近,说明 LM-BP 神经网络对农业生产总值有很好的泛化和拟合性。而标准 BP 预测的结果最大误差为 2.95%,最小误差为 0.07%,平均误差为 1.14%,预测值与实际值偏离较大,说明标准 BP 网络在预测具农业生产总值拟合性不好,其预测效果不佳。

表 3 各种模型预测值比较

年份	实际值 亿元	LM-BP 神经网络		标准 BP 网络	
		预测值 亿元	误差 %	预测值 亿元	误差 %
2001	872.90	873.78	0.10	882.02	1.04
2002	916.50	916.86	0.04	910.53	0.65
2003	1 030.89	1 031.30	0.04	1 011.80	1.85
2004	1 294.53	1 294.10	0.03	1 290.10	0.34
2005	1 448.37	1 448.50	0.01	1 478.20	2.06
2006	1 622.22	1 623.50	0.08	1 611.40	0.67
2007	2 026.22	2 025.90	0.01	2 024.70	0.07
2008	2 389.79	2 390.80	0.04	2 430.70	1.71
2009	2 380.51	2 381.40	0.04	2 324.60	2.35
2010	2 720.99	2 721.90	0.03	2 801.30	2.95
2011	3 323.37	3 323.30	0	3 237.90	2.57
2012	3 490.72	3 489.40	0.04	3 512.90	0.64
平均 误差			0.04		1.41

用训练好的 LM-BP 神经网络预测农业生产总值,将表 1 广西 2012 年农业生产主要指标作为 BP 网络的输入预测

(下转第 10037 页)

中也存在不少问题,这些问题在基层具有一定普遍性。

2.1 建议加强政策研究,明确农业标准化生产整体推进的目标、内容和途径,以有效指导地方在此框架下整合资源、构建制度、开展工作 涡阳县在整合政策资源和管理资源方面做出了有效探索,但在标准化整体推进方面自上而下都没有清晰的意识和明确的规划,更多地还是在具体政策、具体项目、具体业务的层面上开展工作,目标体系不够清晰,工作衔接不够紧密。

2.2 建议完善体系构建,促进认证机制与监测机制的有机融合,将“三品一标”认证覆盖情况作为整体推进的主要评价指标之一 涡阳县的绿色食品办公室与质量安全检测站合署办公,在以认证机制引领质量安全监管方面做出了有益的探索,但在该县农业工作中仍处于边缘化的境地,市里下达的指令性检验检测任务没有匹配经费,县财政也没有给予安排,只能利用创收收入来支撑,更谈不上自主履职的经费支撑。其利用农业部基建投资新建的质检中心设备先进,但利用率明显不足。强化“三品一标”认证在质量安全监管体系中的地位作用,对标准化整体推进而言是重要抓手,对县级机构而言则可有效支撑自主履职、配合执法监管、解决经费困境。

2.3 建议统筹基地建设,依据科学合理的基地布局总体规划来整合资源和加大投入,进一步强化各级各类基地面向标准化的基础条件建设和农技示范功能 尽管涡阳县已经通

过“十百千万”工程对基地建设进行了初步规划,但由于各种基地的政策扶持渠道来源较多,依然存在边界不清、交叉重复等问题,而且这些基地的农田水利设施条件依然比较薄弱,农技示范推广依然发挥不足。如果不能有效统筹强化各类基地的布局和建设,其在标准化技术模式示范推广中应有的作用将难以充分发挥,其标准化整体推进的核心载体功能也将受到很大局限。

2.4 建议引导农民合作,推进多环节统一作业,配合农技指导和行政监管,探索解决标准化生产落地难和全过程监管实现难的问题 一方面,标准化生产宣传和技术推广形式多样,但对广大农民来说,由于意识不强、文化水平不高等因素,难以满足标准化生产要求,不做生产记录、不按规范操作等情形还普遍存在。另一方面,农业生产周期长、环节多、范围广,无论质量安全监管力量是否薄弱,行政手段都永远不可能面面俱到地实现全过程监控。引导农民组织推进统一作业乃至规模化、集约化经营,必须循序渐进,在此之前,只能加大面向广大农户的培训和宣传力度,亦可由县里统一下发生产记录样本,并指导农民按要求填写,逐步培养标准化意识。

参考文献

- [1] 李振山. 农业标准化:加快现代农业发展的“助推器”[J]. 农业质量标准, 2006(3): 27-29.

(上接第 10011 页)

2013 年广西农业生产总值,模型仿真预测结果为 3 587.83 亿元,2014 年广西农业生产总值为 3 716.96 亿元。

4 结论

BP 神经网络以其具有自我学习、非线性及可靠性的优点为农业生产总值预测提供新的解决方法。但标准 BP 算法采用最速下降求解,虽然方法简单但收敛速度慢,容易造成网络不收敛和过拟合现象。LM-BP 网络利用高斯-牛顿在函数逼近最优值附近产生一个理想搜索方向,通过自适应调整网络权值,克服负梯度下降法朝单一方向盲目搜索的缺点,在非线性和拟合有学习速度快及泛化能力强的优点,为预测农业生产总值提供一种高效、准确及可靠的方法。

参考文献

- [1] 张浩瑕,郝晋珉,胡吉敏. 基于自适应 ARMA 模型的区域农业总产值构成研究与应用[J]. 农业工程学报, 2008, 24(8): 84-88.
- [2] 陈显周,区晶莹,俞守华,等. 组合时间序列模型及其在我国农业总产值预测中的应用[J]. 广东农业科学, 2011(13): 194-195.
- [3] 刘楠. 一种灰色改进模型在农业经济预测中的应用[J]. 统计与决策, 2011(12): 39-41.
- [4] 阳长水,洪新星. 基于灰色理论的珠三角都市农业发展预测研究[J]. 广东农业科学, 2011(12): 241-242.
- [5] 郭金海,张金霞. 灰色预测在农林牧渔业发展中的应用[J]. 数学的实践与认识, 2009, 39(20): 79-85.
- [6] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back-propagation error[J]. Nature, 1986, 323: 533-536.
- [7] 飞思科技产品研发中心. 神经网络理论与 MATLAB7 实现[M]. 北京: 电子工业出版社, 2005.
- [8] 广西壮族自治区统计局. 广西统计年鉴—2013[M]. 北京: 中国统计出版社, 2013.