

# 人工神经网络与遗传算法相结合在作物估产中的应用——以吉林省玉米估产为例

李 哲, 张军涛

(中国科学院地理研究所 北京 100101)

**摘要:** 在遗传算法(Genetic Algorithm)与误差反传(Back Propagation)网络结构模型相结合的基础上,设计了用遗传算法训练神经网络权重的新方法,并对吉林省梨树和德惠县的玉米进行了估产研究,同时与 BP 算法和灰色系统理论模型进行了比较。经检验,计算值与实际值接近,并优于灰色理论模型,具有良好的预测效果,从而为农作物估产提供了新方法。

**关键词:** 作物估产; 遗传算法; 神经网络

## Application of the combination of genetic algorithm and artificial neural network on crop yield estimation in Jilin Province

LI Zhe, ZHANG Jun-Tao (Institute of Geography, Chinese Academy of Sciences, Beijing 100101, China)

**Abstract:** Based upon the combination of Genetic Algorithm and BP Neural Network, the estimation of maize yield in Lishu and Dehui County, Jilin Province was estimated in this paper. The new approach was compared with BP Neural Network as well as Gray Model Theory. The results indicate that the prediction values approximate to the real maize yield and are more accurate than BP network and gray model theory.

**Key words:** estimation of crop yield; genetic algorithm; neural network

文章编号:1000-0933(2001)05-0716-05 中图分类号:S11 文献标识码:A

遥感图像信息提取、多元回归分析和灰色系统理论等是目前进行农作物估产的主要方法<sup>[1]</sup>。近 10a 来迅速发展起来的人工神经网络(ANN)模型在人工智能、信号处理、自动控制和模式识别等领域发挥着重要作用。由于人工神经网络具有自适应学习能力和记忆联想能力,使得它为解决复杂的非线性问题提供了有用的工具,其预测能力大大优于多元回归分析和判别分析。目前广泛采用的 Rumelhart 等人推广的 BP 算法,具有简单和可塑的优点<sup>[2]</sup>。但是,BP 算法是基于梯度的方法,其最大缺欠是收敛速度慢,且常受到局部极点的困扰,因而影响了预测的精度。如果将 BP 算法与遗传算法结合起来,便可以弥补这方面的缺欠。

遗传算法(Genetic algorithm)是由美国密歇根大学的 J. Holland 等人创立的,它采用解的种群作为工作单元,使用达尔文生物进化的适者生存原则指导搜索并改进目标。每一个解的质量通过依赖于目标函数的适应值函数来进行评估,搜索过程通过代数变更(进化)来进行,每代中的个体遗传到下一代的可能与它的适应值成正比。它使用三个基本算子,即复制、交叉和变异。复制是指种群中的个体直接复制到下一代中;交叉是从种群中选择两个个体进行交配,组成两个个体的特性形成一个或几个新个体,复制和交叉将好的特性进行遗传;而变异则是模仿突变理论改变种群少数个体的个别基因,以增加种群的多样性,获得更宽的搜索范围<sup>[3]</sup>。

运用遗传算法可以更合理地确定人工神经网络的权重,图 1 表示了遗传算法与人工神经网络的结合

基金项目:国家自然科学基金重点资助项目(49731020)

承蒙导师郑度院士帮助与指导,特此致谢!

收稿日期:1999-11-21 修订日期:2000-04-19

作者简介:李 哲(1973~),男,吉林人,博士。主要从事自然地理综合研究,生态地理区域系统,气候变化,环境冲突等研究。

过程<sup>[4]</sup>。本文将运用这一成果针对作物估产问题进行研究。

## 1 面向神经网络权重学习的遗传算法

### 1.1 遗传算法的编码方式

神经网络的权重学习是一个复杂的连续参数优化问题,如果采用二进制编码,会造成编码串过长,且需要解码为实数,使权重变化为步进,影响网络学习精度,因此这里采用实数编码,即神经网络的各个权值按一定的顺序级联成一个长串,串上每个位置对应着网络的一个权值。

### 1.2 评价函数

将染色体上表示的各权值分配到网络结构中,网络以训练集样本为输入输出,运行后返回误差平方和的倒数作为染色体的评价函数:

$$f = 1 / \sum_{i=1}^n e_i^2 \quad (1)$$

### 1.3 权重初始化

对于普通的BP算法,网络的初始权重一般取0~1.0之间均匀分布的随机数;而在初始染色体集中,网络的各权重是以概率分布 $e^{-|y|}$ 来随机确定的。这一点是有别于BP算法的。

### 1.4 遗传算子

对于不同的应用问题,遗传算子的形式有多种多样,这里采用权值交叉和权值变异算子。

**权值交叉算子** 对于子代染色体中的每个权值输入位置,交叉算子从两个亲代染色体中随机选取若干个交叉位置,并将这一代染色体在交叉位置进行交叉运算,这样子代染色体便含有两个亲代的遗传基因。

**权值变异算子** 对于子代染色体中的每个权值输入位置,变异算子以概率在初始概率分布中随机选择一个值,然后与该输入位置上的权值相加。

### 1.5 选择方式

这里不采用比例选择方式,而使用亲代度量 $S$ 决定群体中每个个体被选种的概率,群体中的个体以式2所示的概率成为亲代染色体:

$$\begin{aligned} P_2 &= P_1 \cdot S \\ P_3 &= P_2 \cdot S \\ &\dots \end{aligned} \quad (2)$$

式中, $P_1, P_2, P_3, \dots$  分别表示评价函数最优、次优、第3、…的个体概率。

## 2 遗传算法与神经网络的结合

选用3层BP网络模型<sup>[5]</sup>,由一个输入层、一个输出层和一个隐含层(中间层)组成。设网络的学习样本和期望输出分别为 $A_k$ 和 $C_k$ ( $k=1, 2, \dots, m$ ),首先计算输入层对隐含层的激活值:

$$b_i = f(\sum_{h=1}^n a_h V_{hi} + \theta_i) \quad (3)$$

式中, $i=1, 2, \dots, p, a_h$ 为输入层节点( $h=1, 2, \dots, n$ ), $V_{hi}$ 为输入层到隐含层的连接权重, $\theta_i$ 为隐含层单元的阈值。对于 $a_h, V_{hi}$ 和 $\theta_i$ ,它们的初值是按概率 $e^{-|y|}$ 得到的随机数。激活函数为Sigmoid函数:

$$f(x) = 1 / (1 + e^{-x}) \quad (4)$$

根据式(3)和(4)计算输出层节点的激活值:

万方数据

$$C_j = f(\sum_{i=1}^p W_{ij} b_i + \gamma_j) \quad (5)$$

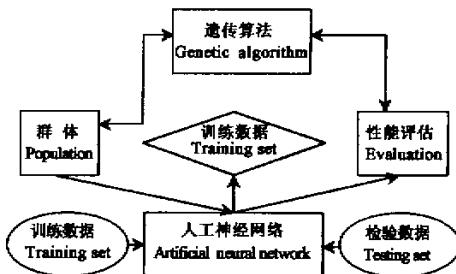


图1 神经网络与遗传算法相结合的系统结构图

Fig. 1 The combination of GA and ANN

式中,  $j = 1, 2, \dots, q$ ,  $W_{ij}$  为隐含层到输出层的连接权重,  $\gamma_j$  为输出层单元的阈值,  $W_{ij}$  和  $\gamma_j$  也是按概率  $e^{|\gamma_j|}$  确定的随机数。

根据式(6)计算输出层的一般化误差:

$$d_j = C_j(1 - C_j)(C_j^k - C_j) \quad (6)$$

式中,  $C_{jk}$  为输出层单元  $j$  的期望输出。

最后计算隐含层单元相对于每个  $d_j$  的误差:

$$e_i = b_i(1 - b_i) \sum_{j=1}^q W_{ij} d_j \quad (7)$$

根据以上步骤,结合染色体的交叉、变异原则(式2)调整隐含层单元到输出层单元的连接权重,然后调整输出层单元的阈值、输出层单元到隐含层单元的连接权重,最后调整隐含层单元的阈值。计算实际输出与期望输出的误差,当全部样本的输出误差小于设定的收敛误差时,神经网络的训练结束。

### 3 实例研究

玉米是吉林省的主要粮食作物。在生产力水平、农技措施(如灌溉和施肥条件)相同的情况下,影响玉米产量的主要因子是温度和降水。梨树和德惠是吉林省的产粮大县,属于温带半湿润季风气候,春季干旱对农作物危害较大;夏季多雨,即雨热同期,有利于作物生长;秋季时有低温早霜,农作物易受害减产。从玉米的整个生育期来看,可分为苗期(5月中、下旬~6月上旬),七叶期(6月中、下旬~7月上旬),抽雄期(7月中、下旬~8月上旬),和灌浆期(8月中、下旬~9月上旬)<sup>[1]</sup>。这几个时期的气温和降水与玉米产量有较大的关联度。使用梨树和德惠两县 1959~1996 年的旬降水和旬均温,将 1992 年以前的玉米产量与气象资料作为网络的训练数据,再用 1993 年及其以后的气象数据来计算玉米产量。因此在确定神经网络结构时,分别选取玉米 4 个生长时期的气温因子  $T_1, T_2, T_3, T_4$  与降水因子  $R_1, R_2, R_3, R_4$  等 8 个因子作为输入层的节点;另外,选取玉米产量作为输出层的节点;隐含层为一层,节点数为 30。网络输入层的因子分别按下式计算:

$$\begin{aligned} T_1 &= T_{520} + T_{530} + T_{610} \quad (T_{520}, T_{530}, T_{610} \text{ 分别为 } 5 \text{ 月中旬、下旬和 } 6 \text{ 月上旬的平均气温}) \\ R_1 &= R_{520} + R_{530} + R_{610} \quad (R_{520}, R_{530}, R_{610} \text{ 分别为 } 5 \text{ 月中旬、下旬和 } 6 \text{ 月上旬的平均降水}) \\ T_2 &= T_{620} + T_{630} + T_{710} \quad (T_{620}, T_{630}, T_{710} \text{ 分别为 } 6 \text{ 月中旬、下旬和 } 7 \text{ 月上旬的平均气温}) \\ R_2 &= R_{620} + R_{630} + R_{710} \quad (R_{620}, R_{630}, R_{710} \text{ 分别为 } 6 \text{ 月中旬、下旬和 } 7 \text{ 月上旬的平均降水}) \\ T_3 &= T_{720} + T_{730} + T_{810} \quad (T_{720}, T_{730}, T_{810} \text{ 分别为 } 7 \text{ 月中旬、下旬和 } 8 \text{ 月上旬的平均气温}) \\ R_3 &= R_{720} + R_{730} + R_{810} \quad (R_{720}, R_{730}, R_{810} \text{ 分别为 } 7 \text{ 月中旬、下旬和 } 8 \text{ 月上旬的平均降水}) \\ T_4 &= T_{820} + T_{830} + T_{910} \quad (T_{820}, T_{830}, T_{910} \text{ 分别为 } 8 \text{ 月中旬、下旬和 } 9 \text{ 月上旬的平均气温}) \\ R_4 &= R_{820} + R_{830} + R_{910} \quad (R_{820}, R_{830}, R_{910} \text{ 分别为 } 8 \text{ 月中旬、下旬和 } 9 \text{ 月上旬的平均降水}) \end{aligned}$$

根据以上方法,分别计算出 1959~1996 年 38a 的气温与降水因子;为了满足 Sigmoid 函数的要求,将气象因子按式(8)变换为 BP 网络标准化训练集列于表 1。在 BP 算法中,学习因子  $\alpha = 0.5$ ,动量因子  $\beta = 0.5$ ;遗传算法中,交叉率  $P_c = 0.15$ ,变异率  $P_m = 0.13$ ,亲代度量  $S = 0.87$ 。两个县的数据分别经过 27917 次和 19796 次迭代运算,收敛误差达到 0.0005,训练结束。3 种不同方法预测结果列于表 2。

$$E_i = \frac{x_i - (S_{\min} - \delta)}{(S_{\max} - \delta) - S_{\min}}$$

式中  $S_{\min}$  和  $S_{\max}$  为样本序列的最小值和最大值;  $\delta$  为一小量,以保证变换后的序列最大值略小于 1 和最小值略大于 0。

表 2 中的 BP 神经网络的是指没有与遗传算法相结合的算法,按式(3)~(7)计算;灰色系统理论采用 GM(1,N) 气象产量预测模型计算<sup>[1]</sup>。

### 4 结果与讨论

从表 2 可以看数据无论用什么模型估产,其预测结果的相对误差都随年份的增加而增加。这主要是由于作物产量不仅仅是气候条件决定的,灌溉、施肥、品种改良种植制度等人为因素起着明显的作用。实际

的粮食产量是气象和人为因素综合作用的结果,因而单纯用气象因子来反映粮食产量难免有一定的误差,但从另一方面来讲,在水热条件不好的年份,如遇到干旱、霜冻等,其作物产量也很低;同样,高产的年份水热条件都较好。因此,用不同的气象条件对应不同的作物产量(期望输出)来训练神经网络,其预测结果是令人满意的。如果再辅助其他人为因素加以修正,预测结果将更加接近实际值。

## 5 结论

气温和降水是影响作物产量的主要因子,将遗传算法与神经网络相结合,并用以预测产量,能够解决困扰BP网络的“局部最小”,和收敛速度慢等问题,同时优于灰色系统理论,因此将其作为一种新的作物估产方法是可行的。

表1 BP 网络标准化训练集

Table 1 Standard training set of the BP network

地区	年份	学习样本输入 Learning sample to input								期望输出 Hope to output
		$T_1$	$T_2$	$T_3$	$T_4$	$R_1$	$R_2$	$R_3$	$R_4$	
梨树县	1959	0.0043	0.0056	0.0062	0.0054	0.0015	0.0179	0.0161	0.0125	0.1087
Lishu	1960	0.0039	0.0057	0.0063	0.0052	0.0068	0.0123	0.0161	0.0025	0.0562
county	1961	0.0045	0.0058	0.0064	0.0052	0.0049	0.0174	0.0149	0.0173	0.0967
	1962	0.0039	0.0059	0.0061	0.0051	0.0052	0.0088	0.0207	0.0140	0.0914
	1963	0.0049	0.0059	0.0061	0.0052	0.0015	0.0160	0.0251	0.0031	0.0817
	1964	0.0045	0.0059	0.0056	0.0051	0.0030	0.0050	0.0195	0.0216	0.1162
	1965	0.0043	0.0059	0.0057	0.0049	0.0001	0.0103	0.0191	0.0123	0.1470
	1966	0.0045	0.0056	0.0062	0.0048	0.0029	0.0044	0.0206	0.0240	0.1477
	1967	0.0046	0.0054	0.0064	0.0055	0.0068	0.0160	0.0133	0.0015	0.1695
	1968	0.0043	0.0054	0.0065	0.0052	0.0055	0.0035	0.0041	0.0110	0.1410
	1969	0.0036	0.0053	0.0061	0.0050	0.0047	0.0049	0.0105	0.0269	0.1582
	1970	0.0046	0.0060	0.0060	0.0056	0.0039	0.0031	0.0147	0.0084	0.2273
	1971	0.0043	0.0056	0.0059	0.0053	0.0078	0.0169	0.0074	0.0053	0.1995
	1972	0.0041	0.0052	0.0062	0.0047	0.0055	0.0073	0.0104	0.0033	0.2266
	1973	0.0039	0.0059	0.0064	0.0049	0.0096	0.0082	0.0262	0.0067	0.2408
	1974	0.0042	0.0057	0.0065	0.0053	0.0076	0.0042	0.0096	0.0040	0.2956
	1975	0.0045	0.0060	0.0062	0.0054	0.0120	0.0065	0.0096	0.0030	0.3091
	1976	0.0042	0.0052	0.0062	0.0048	0.0072	0.0110	0.0111	0.0083	0.2438
	1977	0.0047	0.0056	0.0062	0.0053	0.0107	0.0154	0.0160	0.0026	0.2423
	1978	0.0043	0.0061	0.0062	0.0050	0.0083	0.0112	0.0079	0.0161	0.3354
	1979	0.0049	0.0058	0.0060	0.0047	0.0022	0.0075	0.0164	0.0095	0.3834
	1980	0.0046	0.0058	0.0059	0.0051	0.0037	0.0137	0.0051	0.0166	0.4337
	1981	0.0041	0.0059	0.0056	0.0052	0.0056	0.0032	0.0245	0.0092	0.5614
	1982	0.0044	0.0061	0.0056	0.0052	0.0042	0.0052	0.0121	0.0111	0.6132
	1983	0.0045	0.0058	0.0054	0.0051	0.0059	0.0171	0.0107	0.0080	0.8301
	1984	0.0047	0.0060	0.0059	0.0054	0.0114	0.0180	0.0161	0.0157	0.8706
	1985	0.0046	0.0056	0.0063	0.0058	0.0021	0.0097	0.0147	0.0136	0.5555
	1986	0.0048	0.0056	0.0062	0.0057	0.0006	0.0100	0.0132	0.0189	0.5937
	1987	0.0046	0.0055	0.0059	0.0053	0.0029	0.0167	0.0309	0.0073	0.8296
	1988	0.0048	0.0059	0.0070	0.0055	0.0105	0.0130	0.0160	0.0127	0.9337
	1989	0.0047	0.0053	0.0065	0.0054	0.0085	0.0079	0.0230	0.0013	0.7421
	1990	0.0043	0.0059	0.0063	0.0057	0.0120	0.0064	0.0083	0.0082	0.9573
	1991	0.0049	0.0053	0.0064	0.0061	0.0035	0.0104	0.0240	0.0024	0.8897
	1992	0.0046	0.0056	0.0063	0.0055	0.0040	0.0038	0.0195	0.0052	0.8899

德惠县	1959	0.0048	0.0068	0.0073	0.0064	0.0028	0.0078	0.0275	0.0163	0.2344
	1960	0.0042	0.0067	0.0076	0.0060	0.0089	0.0222	0.0306	0.0080	0.1442
Dehui	1961	0.0051	0.0070	0.0078	0.0061	0.0101	0.0087	0.0145	0.0059	0.2231
county	1962	0.0048	0.0072	0.0074	0.0060	0.0044	0.0089	0.0140	0.0090	0.1931
	1963	0.0060	0.0071	0.0074	0.0060	0.0010	0.0176	0.0294	0.0074	0.1621
	1964	0.0052	0.0071	0.0067	0.0061	0.0030	0.0127	0.0263	0.0210	0.2006
	1965	0.0050	0.0072	0.0068	0.0059	0.0002	0.0138	0.0305	0.0112	0.1837
	1966	0.0054	0.0068	0.0074	0.0057	0.0041	0.0105	0.0249	0.0196	0.2353
	1967	0.0053	0.0069	0.0077	0.0063	0.0049	0.0038	0.0159	0.0030	0.3001
	1968	0.0050	0.0066	0.0077	0.0060	0.0090	0.0108	0.0122	0.0129	0.3113
	1969	0.0040	0.0064	0.0073	0.0059	0.0095	0.0087	0.0154	0.0279	0.2616
	1970	0.0056	0.0072	0.0072	0.0065	0.0036	0.0083	0.0322	0.0088	0.3301
	1971	0.0049	0.0069	0.0067	0.0059	0.0056	0.0106	0.0163	0.0131	0.4615
	1972	0.0049	0.0062	0.0074	0.0054	0.0072	0.0055	0.0121	0.0058	0.4043
	1973	0.0048	0.0064	0.0077	0.0060	0.0056	0.0125	0.0453	0.0033	0.4005
	1974	0.0052	0.0071	0.0072	0.0063	0.0077	0.0196	0.0167	0.0068	0.4352
	1975	0.0048	0.0065	0.0075	0.0053	0.0055	0.0116	0.0156	0.0024	0.5056
	1976	0.0057	0.0066	0.0073	0.0061	0.0068	0.0031	0.0102	0.0115	0.4671
	1977	0.0051	0.0074	0.0072	0.0058	0.0267	0.0173	0.0249	0.0053	0.3141
	1978	0.0060	0.0070	0.0071	0.0055	0.0042	0.0128	0.0128	0.0039	0.4690
	1979	0.0053	0.0069	0.0073	0.0060	0.0032	0.0103	0.0064	0.0073	0.5169
	1980	0.0050	0.0063	0.0076	0.0055	0.0051	0.0157	0.0058	0.0082	0.4258
	1981	0.0054	0.0078	0.0076	0.0061	0.0036	0.0310	0.0296	0.0050	0.4324
	1982	0.0049	0.0060	0.0069	0.0067	0.0013	0.0017	0.0082	0.0066	0.4587
	1983	0.0059	0.0069	0.0077	0.0058	0.0088	0.0088	0.0271	0.0065	0.6877
	1984	0.0056	0.0066	0.0071	0.0065	0.0052	0.0062	0.0177	0.0277	0.6868
	1985	0.0051	0.0069	0.0068	0.0060	0.0067	0.0138	0.0269	0.0288	0.4437
	1986	0.0049	0.0065	0.0072	0.0061	0.0039	0.0094	0.0333	0.0076	0.3967
	1987	0.0054	0.0070	0.0080	0.0062	0.0033	0.0114	0.0129	0.0208	0.6887
	1988	0.0052	0.0065	0.0076	0.0060	0.0061	0.0171	0.0160	0.0174	0.8229
	1989	0.0050	0.0071	0.0072	0.0063	0.0071	0.0113	0.0075	0.0045	0.4418
	1990	0.0054	0.0068	0.0074	0.0057	0.0129	0.0083	0.0268	0.0041	0.8586
	1991	0.0053	0.0070	0.0079	0.0061	0.0038	0.0114	0.0285	0.0163	0.7835
	1992	0.0050	0.0067	0.0080	0.0063	0.0126	0.0068	0.0251	0.0163	0.8060

表2 不同预测方法检验  
Table 2 Prediction results by different methods

地区 Region	年份 Year	实际单产 Real unit area yield (kg/hm <sup>2</sup> )	预测单产 Predicted unit area yield(kg/hm <sup>2</sup> )					灰色系统理论 Gray model theory	相对误差 Relative error(%)
			遗传算法与BP神经网络的结合 Combination of GA and BP	相对误差 Relative error(%)	BP神经网络 BP network	相对误差 Relative error(%)	BP神经网络 BP network		
梨树县	1993	10438.5	10012.3	-4.1	9889.7	-5.3	9676.5	-7.3	
	1994	7545.0	7002.5	-7.2	8113.2	7.5	8109.3	7.5	
Lishu	1995	10702.5	9865.1	-7.8	9808.4	-8.4	9532.3	-10.9	
county	1996	10875.0	10004.7	-8.0	9896.3	-9.0	9487.6	-12.8	
	平均(绝对值)			6.8		7.6		9.6	
德惠县	1993	7747.5	7003.9	-9.6	7000.4	-9.6	6954.4	-10.2	
	1994	6892.5	6408.2	-7.0	6259.1	-9.2	6234.6	-9.5	
Dehui	1995	6510.0	7100.6	9.1	7150.4	9.8	5898.7	-9.3	
county	1996	7357.5	6648.5	-9.6	6565.7	-10.7	6488.9	-11.8	
	平均(绝对值)			8.8		9.8		10.2	

## 参考文献

- [1] 张树清,陈春,万恩璞. 灰色系统理论在估产中的应用. 地理科学,1998,18(6): 581~585.
- [2] Hechi-Nielsen R. Theory of the back propagation neural network. Int. J. Conf. On Neural Network, Washington D.C., 1989, (1): 593~605.
- [3] Bornholdt A. General asymmetric neural network and structure design by genetic algorithms. Neural Network, 1992, 5(2): 327~334.
- [4] 李敏强,余博数据. 遗传算法与神经网络的结合. 系统工程理论与实践,1999,(2): 65~69.
- [5] 金龙,罗莹,缪启龙,等. 农田土壤湿度的人工神经网络预报模型研究. 土壤学报, 1998, 35(1): 25~35.