

# BP神经网络在焉耆盆地 农田排水量估算中的应用

刘延锋<sup>1,2</sup>, 靳孟贵<sup>2</sup>, 曹英兰<sup>2</sup>

(1. 中国科学院武汉岩土力学研究所, 武汉 430071; 2. 中国地质大学环境学院, 武汉 430074)

**摘要:** 利用 BP 神经网络技术对焉耆盆地农田排水量进行预测。利用灰色关联度分析确定了排水量与各影响因素的关系, 选取了对排水量影响最大的 5 个因素作为 BP 网络的输入, 利用均匀设计方法, 确定了最优的神经网络结构。估算结果表明利用 BP 神经网络可以准确地估算农田排水量, 最大相对误差仅为 -2.45%。

**关键词:** BP 神经网络; 均匀设计; 灰色关联度; 焉耆盆地

中图分类号: S276 文献标识码: A

## Application of BP Neural Network in The Estimation of Farmland Drainage in Yanqi Basin

LIU Yan feng<sup>1,2</sup>, JIN Meng gui<sup>2</sup>, CAO Ying lan<sup>2</sup>

(1. Institute of Rock and Soil Mechanics, the Chinese Academy of Sciences, Wuhan 430071, Hubei, China;

2. School of Environmental Studies, China University of Geosciences, Wuhan 430074, Hubei, China)

**Abstract** BP neural network is used to estimate the farmland drainage in Yanqi Basin, Xinjiang Autonomous Region, China. The correlation between farm drainage and the influencing factors was estimated by gray correlation degree method. The five most important influencing factors were chosen to be the input for the BP neural network to estimate the farmland drainage. In order to obtain the optimal structure of BP neural network, the uniform design was employed. The results of estimation show that BP neural network can estimate farmland drainage accurately with the largest relative error only of -2.45%.

**Key Words:** BP neural network; uniform design; gray correlation degree; Yanqi Basin

## 0 引言

焉耆盆地为南天山褶皱断块山系中一个封闭的大型山间盆地, 其最低处为中国最大的内陆淡水湖——博斯腾湖。自 20 世纪 50 年代以来, 随着大量的农业开垦和引地表水灌溉, 导致地下水位不断升高, 土壤次生盐渍化加剧, 农田高矿化排水增加。大量的农业高矿化废水排入博斯腾湖, 致使博斯腾湖的矿化度不断增加, 其大部分区域已处于咸水湖和中富营养化状态, 湖周生态环境不断恶化。通过当地已有的水文、气象、水利监测站资料来估算农田排水量, 不仅可以节省投资, 还可以确定排水量与各气象、水文、水利因素之间的关系, 从而提出减少农田排水量的措施。农田排水与各影响因素的作用以及各因

素之间的相互影响是非线性的, 因此, 笔者采用神经网络技术对农田排水量进行预测。

## 1 BP神经网络<sup>1-5</sup>

BP 神经网络是一种多层前馈神经网络, 它采用误差反向传播 (Back Propagation) 学习算法。由若干层神经元组成, 它们可分为输入层、隐含层和输出层, 输入信息从输入层传入, 经隐含层处理后, 传向输出层。若输出层的实际输出与期望输出不符, 则转入反向传播, 将输出误差沿原来通路返回, 通过学习来修改各层神经元的权值  $w$ , 使输出误差达到预期目标。每层神经元的状态都将影响下一层的神经元状态, 每个神经元状态都对应着一个激活函数  $f$  和阈值  $b$ 。BP 网络的基本处理单元为非线性输入 - 输出的关系, 输入层神经元阈值为 0, 且  $f(x) = x$ , 而隐含层的激活函数为非线性的 Sigmoid 型 (简称 S 型) 函数, 输出层可以为非线性的 Sigmoid 函数或线性函数。Cyberko 于 1988 年指出, 当各节点均采用 S 型函数时, 一个隐含层就足以实现任意判决分类问题, 两个隐含层则足以表示输入图形

收稿日期: 2005 05 12

基金项目: 高等学校博士点专项科研基金 (20020491011) 和国家自然科学基金 (40472123) 资助课题。

作者简介: 刘延锋 (1975), 男, 博士研究生。

的任意输出函数。

S 型激活函数为:

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

线性型激活函数为:

$$f(x) = wx + b \quad (2)$$

BP 网络的学习算法: 首先对权值和阈值进行初始化, 给定输入  $X$  和期望输出  $Y$ , 从输入层开始到输出层正向计算网络输出  $y$ , 若实际输出与期望输出之间的误差值不满足要求时, 则从输出层开始将误差值沿连接通路反向传播, 修正各层连接权值和阈值, 使网络误差最小。然后不断重复, 直到各个训练模式都满足时结束 BP 网络训练。BP 神经网络结构见图 1。笔者采用 MATLAB 软件中的神经网络工具箱提供的 BP 网络函数, 对所建网络进行训练和仿真。

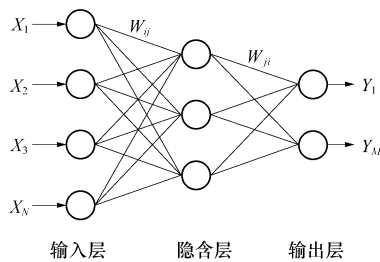


图 1 BP 神经网络结构示意图

BP 网络中参数的设计直接影响着它的学习速度, 而隐含层神经元个数未能预计, 加之容易陷入局部最小值, 有时未能达到理想的应用效果。隐含层神经元数目的确定通常采用试算法、经验法、逐步修改法、与输入单元数相同等。对于仅有一层隐含层的神经网络, 试算法比较容易获得最优的隐含层单元

数, 对于多隐含层的神经网络, 则需要大量的试算, 而且有时难以获得最优的网络结构。此处采用均匀设计方法来原因所需的多隐含层的 BP 网络最优结构。

## 2 均匀设计<sup>[6]</sup>

试验设计的目的是利用最少的试验数目来获得满意的参数。均匀设计是我国数学家方开泰和王元提出的一种适用于多因素多水平的实验设计方法, 它是一种将试验点均匀地散布在试验范围内的科学试验方法, 尤其是对那些试验范围较大, 因素水平较多的复杂试验, 采用均匀设计法会收到较好的效果。

均匀设计的实施常用一个均匀表  $U_n(q^s)$  完成和相对应的使用表来完成,  $U$  表示均匀设计;  $n$  表示要做  $n$  次试验;  $q$  表示每个因素有  $q$  个水平;  $s$  是表中列的数目, 表示最多可安排  $s$  个因素。每个均匀设计表都附有一个使用表, 说明所选试验方案的均匀度。

均匀设计的步骤为: ① 根据试验的目的, 选择合适的因素及其水平; ② 选择合适的均匀设计表及其使用表, 选出列号, 将因素分别安排到这些列号上, 并将这些因素的水平按所在列的指示分别对号。

## 3 农田排水量预测

焉耆盆地农田排水量主要受灌溉引水量、蒸发、降水、各河流来水量(开都河、黄水沟和清水河)、博斯腾湖水位以及灌溉面积等因素的影响。由于系统监测资料较少, 故采用灰色关联分析来研究各因素对排水量的影响程度。焉耆盆地排水量及相关因素的观测值见表 1, 利用式(3)计算排水量与各影响因素之间的灰色关联度<sup>[7]</sup>, 结果见表 2。

表 1 焉耆盆地排水量及相关因素的观测值

样本编号	农田排水量 $FD / \text{亿 m}^3$	灌溉引水量 $I / \text{亿 m}^3$	蒸发 $E / \text{m}$	博湖水位 $H_L / \text{m}$	河流径流量 / 亿 $\text{m}^3$			降水量 $P / \text{m}$	灌溉面积 $A_r / \text{万 hm}^2$
					开都河 $R_k$	清水河 $R_q$	黄水沟 $R_h$		
1	2.723 3	12.23	2.533 5	1 046.80	33.74	1.09	2.05	0.140 9	7.203
2	2.390 7	11.74	2.398 4	1 046.53	28.67	1.00	2.02	0.091 1	7.354
3	2.200 1	11.51	2.411 8	1 046.20	28.10	0.90	1.67	0.083 5	7.537
4	1.953 7	11.28	2.599 9	1 045.69	29.01	0.54	1.75	0.059 2	7.896
5	2.261 4	10.12	2.234 1	1 045.33	24.60	0.86	2.81	0.080 5	8.010
6	2.389 4	9.96	2.590 2	1 045.00	33.43	1.11	2.62	0.154 8	8.149
7	1.930 9	10.19	2.479 4	1 047.29	37.21	1.35	3.00	0.102 2	
8	2.063 1	10.18	2.289 6	1 045.37	33.43	1.70		0.137 9	
9	1.511 1	10.34	2.405 0	1 045.70	30.02	1.19	2.50	0.118 9	
10	1.874 9	10.13	2.769 4	1 046.21	31.85	1.27	2.38	0.093 4	
11	1.821 0	9.58	2.561 4	1 046.38	41.63	1.49	4.00	0.141 4	
12	1.987 5	10.07	2.132 9	1 046.82	29.23	1.20	2.29	0.125 7	
13	2.212 1	10.10	2.275 3	1 045.35	31.85	1.19	3.35	0.144 4	8.380
14	1.692 6	9.96	2.241 7	1 045.64	34.06	1.50	2.83	0.131 8	
15	1.877 1	9.71	2.153 2	1 046.04	34.37	1.56	2.58	0.117 0	
16	1.874 9	9.70	2.284 4	1 046.92	40.68	1.82	5.46	0.131 3	

$$\left\{ \begin{aligned} \xi_i(k) &= \frac{\min_i \min_k |x_0(k) - x_i(k)| + 0.5 \max_i \max_k |x_0(k) - x_i(k)|}{|\min_i \min_k |x_0(k) - x_i(k)| + 0.5 \max_i \max_k |x_0(k) - x_i(k)|} \\ r_i &= \frac{1}{N} \sum_{k=1}^N \xi_i(k) \end{aligned} \right. \quad (3)$$

式中:  $\xi_i(k)$  表示变量  $x_i$  对  $x_0$  在  $k$  时刻的关联系数;  $r_i$  为  $x_i$  对  $x_0$  的灰色关联度。

由表 2 可以看出, 焉耆盆地农田排水量主要受灌溉引水量、蒸发量、降水量、开都河径流量以及博斯腾湖水位 5 个因素的影响, 而且这些因素均为气象水文资料, 容易获得数据, 因此可以利用它们对农田排水量进行预测。此处采用具两层隐含

表 2 年排水量与各影响因素的灰色关联度

农田排水量 $FD$	灌溉引水量 $I$	蒸发 $E$	博湖水位 $H_L$	河流径流量			降水量 $P$	灌溉面积 $A_r$
				开都河 $R_k$	清水河 $R_q$	黄水沟 $R_h$		
1.000 0	0.906 0	0.848 1	0.808 6	0.833 3	0.422 4	0.482 9	0.829 4	0.610 9

层的 BP 神经网络来进行排水量预测。

第一层神经网络的神经元数 ( $N_1$ ) 取 5 ~ 35, 第二隐含层的神经元数 ( $N_2$ ) 取 2 ~ 32, 每层有 16 个水平。根据均匀设计理论, 选取  $U_{16}^*(16^{12})$ , 根据其相应的使用表, 将  $N_1$  和  $N_2$  分别放在第 1 和第 8 列 (表 3)。按表 3 所示的隐含层神经元数目, 分别建立 5- $N_1$ - $N_2$ -1 的 BP 神经网络, 按照设定水平用 1~12 号样本依次训练网络, 并用 13~16 号样本进行测试, 取测

表 3 隐含层神经元数目均匀设计表  $U_{16}^*(16^{12})$

水平	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$N_1$	5	7	9	11	13	15	17	19	21	23	25	27	29	31	33	35
$N_2$	20	6	26	12	32	18	4	24	10	30	16	2	22	8	28	14

表 4 BP 神经网络训练和测试结果

样本 编号	排水量	训练值	预测值	亿 $m^3$	
				绝对 误差	相对 误差 /%
1	2.723 3	2.723 3		0	
2	2.390 7	2.390 7		0	
3	2.200 1	2.200 1		0	
4	1.953 7	1.953 7		0	
5	2.261 4	2.261 4		0	
6	2.389 4	2.389 4		0	
7	1.930 9	1.930 9		0	
8	2.063 1	2.063 1		0	
9	1.511 1	1.511 1		0	
10	1.874 9	1.874 9		0	
11	1.821 0	1.821 0		0	
12	1.987 5	1.987 5		0	
13	2.212 1		2.157 8	-0.054 3	-2.45
14	1.692 6		1.700 0	0.007 4	0.44
15	1.877 1		1.891 2	0.014 1	0.75
16	1.874 9		1.917 8	0.042 9	2.29

试的误差平方和 (SES) 作为判别网络结构优劣的指标。首先初选出 SES 最小的  $N_1 - N_2$  组合, 然后再选择邻近的  $N_1 - N_2$  组合进行试算, 最终选出最优的神经网络结构。利用获得的最优结构为 5-7-6-1 的 BP 神经网络对焉耆盆地的农田排水进行估算, 结果见表 4, 其中最大误差为 -2.5%, 说明所建立的 BP 网络能够比较准确地估算农田排水量。

## 4 结 语

利用灰色关联度分析, 选取最重要影响因素作为神经网络的输入, 不但可以避免不重要因素的干扰, 而且减少了神经网络的输入, 简化了网络结构; 利用均匀设计方法来设计 BP 网络结构, 仅用少量试算过程, 即可获得满意的 BP 网络结构。实例表明, 利用基于均匀设计的 BP 神经网络, 根据水文、气象和水利监测资料, 可以比较准确的预测农田排水量。□

### 参考文献:

- [1] 韩力群. 人工神经网络理论、设计及应用[M]. 北京: 化学工业出版社, 2002.
- [2] 袁曾任. 人工神经网络及其应用[M]. 北京: 清华大学出版社, 1999: 120.
- [3] 李家科, 周孝德, 李亚娇, 等. RBF 神经网络方法在水质评价中的应用[J]. 灌溉排水学报, 2003, 22(6): 70-73.
- [4] 柳林. MATLAB 6.5 辅助神经网络分析与设计[M]. 北京: 电子工业出版社, 2003.
- [5] 孙雷剑, 白宏哲, 张剑平, 等. 基于神经网络的微合金钢热轧奥氏体晶粒尺寸预报模型[J]. 哈尔滨工业大学学报, 2002, 34(1): 40-44.
- [6] 方开泰. 均匀设计与均匀设计表[M]. 北京: 科学出版社, 1994.
- [7] 邓聚龙. 灰色系统基本方法[M]. 武汉: 华中理工大学出版社, 1987.