

【灌溉·供水】

基于 RBF 人工神经网络模型预测棉花耗水量

刘洪建¹, 刘婧然¹, 马英杰², 王 喆³

(1. 河北工程大学, 河北 邯郸 056038; 2. 新疆农业大学, 新疆 乌鲁木齐 830052;

3. 中山大学, 广东 广州 510275)

摘 要:利用 MATLAB 工具箱,以平均气温、日照时数、平均风速为输入变量,建立了新疆石河子地区棉花耗水量的 RBF 人工神经网络预测系统,通过 2008 年实测数据的检验表明,此预测系统网络模型的绝对误差最大为 0.096 7 mm/d、最小为 0.002 5 mm/d、平均为 0.041 9 mm/d,相对误差最大为 2.649 1%、最小为 0.034 1%、平均为 0.878 0%。可见,网络模型预测的准确度较高,较以往的线性模型更合理,并且此网络训练花费的时间仅需 0.0780 s,具有一定的实用价值。

关 键 词:预测; 人工神经网络; 径向基函数; 棉花耗水量

中图分类号: TP183; S275

文献标识码: A

doi:10.3969/j.issn.1000-1379.2009.06.036

棉花生长需要的水分主要来自土壤,灌溉是新疆棉田土壤水分的主要来源^[1]。因此,对水资源不足、干旱少雨的新疆来说,进行灌溉需水量预测、制订灌溉用水计划是十分必要的。在分析预测棉区灌溉需水量或制订灌溉用水计划、水量分配计划时,作物需水量的预测是基本的内容之一。中长期作物需水量预测一般根据其气象因素的关系,采用线性回归预测法;短期实时预测常采用指数平滑预测技术,也有在短期气象预测基础上采用线性回归预测方法。线性回归方法预测精度不高,采用非线性方法能克服线性方法的一些缺点^[2]。计算机人工神经网络是 20 世纪 80 年代兴起的非线性科学,其在人工智能、自动控制和模式识别等领域取得了令人瞩目的成果^[3-4]。将神经网络应用于非线性系统的建模和预测,可以有效地描述问题本身具有的不确定、多输入等复杂非线性特性^[5]。笔者以新疆石河子棉区为例,建立了基于 RBF 人工神经网络的作物耗水量预测模型,并通过田间实测资料对模型进行了检验^[6]。

1 径向基函数人工神经网络简介

人工神经网络(Artificial Neural Networks, ANN)是由大量简单的基本元件——神经元相互连接,通过模拟人的大脑神经处理信息的方式,进行信息并行处理和非线性转换的复杂网络系统^[7]。根据生物神经元的结构和工作原理作进一步简化,构成了神经元模型,即人工神经元,结构见图 1。

从图 1 中可以看出,它相当于一个多输入单输出的非线性阈值器件。定义 $p = (p_1, p_2, \dots, p_r)$ 表示神经元的输入向量; $w = (w_1, w_2, \dots, w_r)$ 表示权值向量,其每个元素值可正可负,分别表示为兴奋性突触和抑制性突触; θ 为神经元的阈值,若神经元输入向量的加权和 $\sum_{i=1}^r w_i p_i$ 大于 θ ,则该神经元被激活,所以输入向量的加权和也称为激活值; f 为神经元的输入输出关系函数,亦即传输函数(传递函数)。这样,神经元的输出可以表

示为 $a = f(\sum_{i=1}^r w_i p_i - \theta)$ ^[8]。

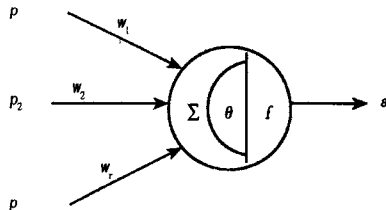


图1 人工神经元模型

由大量神经元组成的人工神经网络可分为单层神经网络和多层神经网络,其中单层神经网络只能解决线性可分的分类问题。多层神经网络可以用于非线性分类问题。多层网络有适合多层网络的学习算法,其中有一种常用的网络,它只有一个隐层,输出单元是线性求和单元,所以输出是各隐单元输出的加权和。隐单元的作用函数用径向基函数(Radial Basis Function, RBF),输入层到隐单元间的权值固定为 1,所以此网络叫做径向基函数人工神经网络,此网络只有隐单元到输出单元间的权值为可调^[7]。其网络结构见图 2。

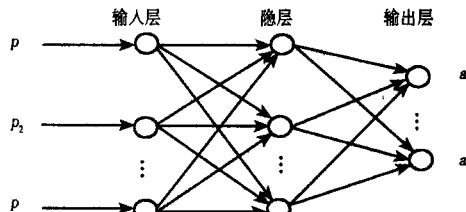


图2 径向基函数网络结构模型

径向基函数人工神经网络的计算过程: 设一 3 层前向人工神经网络,输入层的输入为 $X_p = \{x_{1p}, x_{2p}, \dots, x_{np}\}$ ($p = 1, 2, \dots, P$),

收稿日期:2009-01-25

基金项目:新疆维吾尔自治区高新技术研究与发展计划项目(200712111);新疆水利水电工程重点学科资助项目(xjzdxk2002-10-05)。

作者简介:刘洪建(1954—),男,河北邯郸人,高级工程师(教授级),主要从事计算机应用方面的研究工作。

E-mail:liuhongjian330@sohu.com

隐层神经元的输出为 $O_p = R_p / \sum_{g=1}^m R_{gp} (i = 1, 2, \dots, m), R_p = \exp[-\frac{1}{2}(\frac{x_p - c_{ij}}{\sigma_i})^2]$ 。

将 R_p 称为径向基函数。其中, c_{ij} 为径向基函数的中心, σ_i 为径向基函数的宽度。设网络输出层只有一个输出, 输出为 $y_p = \sum_{i=1}^m v_i O_p$, v_i 为输出层神经元与隐层神经元 i 的连接权。当确定网络中 c_{ij} , σ_i 和 v_i 值以后, 就能够求出给出 X_p 时网络对应的输出值 y_p [10]。这种网络的学习过程为: ① 确定径向基函数的中心 c_{ij} ; ② 确定径向基函数的宽度 σ_i ; ③ 确定输出神经元与隐层神经元之间的连接权 v_i [9]。RBF 网络的训练过程可分为两步: 第一步为无教师式学习, 确定训练输入层与隐层间的权值; 第二步为有教师式学习, 确定训练隐层与输出层间的权值 [7]。

2 RBF 人工神经网络的 MATLAB 程序设计

结合源代码和注释, 说明如何利用 MATLAB 的神经网络工具箱建立径向基函数人工神经网络模型 [9]。

(1) 数据预处理。用公式 $x_p' = \frac{x_p - x_{\min}}{x_{\max} - x_{\min}}$ [11] 进行归一化处理,

$x_p (p = 1, 2, \dots, P)$ 为样本数据, $x_{\max} = \max\{x_p\}$, $x_{\min} = \min\{x_p\}$ 。

(2) 径向基函数神经网络的生成。一般采用 *newrb* 函数。该函数调用格式为 $[net, tr] = \text{newrb}(P, T, GOAL, SPREAD, MN, DF)$, 其中 P 为 Q 组输入向量组成的 $R \times Q$ 维矩阵; T 为 Q 组目标分类向量组成的 $S \times Q$ 维矩阵; $GOAL$ 为均方误差, 默认为 0; $SPREAD$ 为径向基函数的扩展速度, 默认为 1。在网络设计过程

中, 需要用不同的 *SPREAD* 值进行尝试, 以确定一个最优值; MN 为神经元的最大数目, 默认为 Q 。 DF 为两次显示之间所添加的神经元数目, 默认为 25; *net* 是返回值, 一个径向基网络; *tr* 也是返回值, 训练记录 [7]。

(3) RBF 神经网络的仿真。神经网络的仿真就是把数据输入到已训练好的网络中, 从而得到网络输出。网络的仿真函数为 *sim*, 其格式为 $a = \text{sim}(net, p)$, 其中 a 为网络输出 *net* 为训练好的神经网络, p 为网络输入。

3 应 用

(1) 基本资料。使用新疆石河子 148 团垦区 2005 ~ 2008 年棉花耗水量及相关的气象资料。试验田土壤质地为黏土, 0 ~ 20 cm、20 ~ 40 cm 土层土壤干容重为 1.56 g/cm³, 40 ~ 60 cm 为 1.6 g/cm³; 0 ~ 20 cm、20 ~ 40 cm、40 ~ 60 cm 土层的田间持水率分别为 34.3%、37.8%、36.3% (体积百分比)。棉花品种为 ZM-2。试验地属北疆地区的高产田, 选用 2005 ~ 2008 年 5 月 21 ~ 31 日的棉花耗水量及相关的气象资料作为研究对象。对于作物耗水量来说, 关键气象因子为平均气温和日照时数 [11]。所以, 选择平均气温、日照时数和平均风速作为作物耗水量的影响因子 (即输入变量)。

(2) 各层神经元数的确定。RBF 网络输入层神经元数设为 3, 依次对应所选择的 3 个影响因子。网络输出变量为作物耗水量, 即输出层神经元数为 1。

(3) 样本数据预处理。为了消除各个因子因量纲和单位不同而造成的影响, 将数据进行归一化处理, 处理结果见表 1。将表 1 中的前 3 年数据作为训练样本, 最后一年的数据作为预测样本。

表 1 归一化处理后的样本数据

日期	平均气温	日照时数	平均风速	作物耗水量	日期	平均气温	日照时数	平均风速	作物耗水量
2005-05-21	0.511 8	0.659 1	0.276 6	4.720 4	2007-05-21	0.500 0	0.193 2	0.148 9	2.439 1
2005-05-22	0.611 8	0.215 9	0.212 8	3.019 4	2007-05-22	0.000 0	0.465 9	0.914 9	2.788 7
2005-05-23	0.717 6	0.272 7	0.212 8	3.654 0	2007-05-23	0.229 4	0.931 8	0.212 8	4.849 5
2005-05-24	0.641 2	0.306 8	0.063 8	3.352 0	2007-05-24	0.552 9	0.943 2	0.063 8	5.882 6
2005-05-25	0.558 8	0.068 2	0.638 3	2.682 2	2007-05-25	0.564 7	0.022 7	0.425 5	2.242 9
2005-05-26	0.494 1	0.204 5	0.361 7	2.725 9	2007-05-26	0.582 4	0.954 5	0.148 9	6.141 9
2005-05-27	0.100 0	0.193 2	0.574 5	1.516 3	2007-05-27	0.629 4	0.409 1	0.702 1	4.538 3
2005-05-28	0.135 3	0.272 7	0.787 2	2.256 0	2007-05-28	0.158 8	0.193 2	0.212 8	1.290 7
2005-05-29	0.158 8	0.079 5	0.787 2	1.476 1	2007-05-29	0.205 9	0.715 9	0.148 9	3.721 7
2005-05-30	0.376 5	0.625 0	0.212 8	4.004 8	2007-05-30	0.417 6	0.988 6	0.361 7	5.959 8
2005-05-31	0.594 1	0.227 3	0.148 9	2.929 8	2007-05-31	0.623 5	1.000 0	0.000 0	6.313 3
2006-05-21	0.676 5	0.272 7	0.638 3	4.020 1	2008-05-21	0.735 3	0.909 1	0.489 4	6.899 1
2006-05-22	0.464 7	0.556 8	0.361 7	4.196 6	2008-05-22	0.776 5	0.772 7	0.212 8	6.102 7
2006-05-23	0.517 6	0.886 4	0.574 5	6.118 4	2008-05-23	0.835 3	0.272 7	0.212 8	4.076 6
2006-05-24	0.505 9	0.375 0	0.425 5	3.608 0	2008-05-24	0.788 2	0.00 0	0.340 9	2.867 1
2006-05-25	0.635 3	0.795 5	0.574 5	6.134 2	2008-05-25	0.482 4	0.079 5	0.212 8	1.944 3
2006-05-26	0.600 0	0.102 3	0.787 2	3.162 5	2008-05-26	0.435 3	0.886 4	0.212 8	5.385 7
2006-05-27	0.811 8	0.920 5	0.361 7	7.070 4	2008-05-27	0.611 8	0.9545	0.148 9	6.247 6
2006-05-28	0.829 4	0.795 5	0.489 4	6.728 7	2008-05-28	0.747 1	0.931 8	0.212 8	6.708 9
2006-05-29	0.717 6	0.909 1	0.574 5	6.938 5	2008-05-29	0.905 9	0.897 7	0.063 8	6.947 0
2006-05-30	0.741 2	0.840 9	0.276 6	6.358 1	2008-05-30	0.947 1	0.545 5	0.276 6	5.775 6
2006-05-31	0.911 8	0.636 4	1.000 0	6.929 4	2008-05-31	1.000 0	0.897 7	0.148 9	7.387 9

(4) RBF 网络的设计。用 RBF 网络进行预测, 其中 RBF 网络中的 *SPREAD* 值分别取 1.1、1.2、1.3、1.4、1.5、2.3、4.5、6, 经过试验, *SPREAD* 值取 2 训练的时间最短, 并且预测精度也最好。

其 MATLAB 代码为:

```
t1 = clock;
net = newrb(p, t, 0.1, 2);
```

```


    datat = etime(clock,t1);
    p_test = [0.735 3 0.776 5 0.835 3 0.788 2 0.482 4 0.435 3
              0.611 8 0.747 1 0.905 9 0.947 1 1.000 0;
              0.909 1 0.772 7 0.272 7 0.000 0 0.079 5 0.886 4
              0.954 5 0.931 8 0.897 7 0.545 5 0.897 7;
              0.489 4 0.212 8 0.212 8 0.340 9 0.212 8 0.212 8
              0.148 9 0.212 8 0.063 8 0.276 6 0.148 9];

    y = sim(net,p_test)


```

其中 $t_1 = clock$ 和 $datat = etime(clock,t1)$ 组成计算设计网络所用时间,在 MATLAB 的运行窗口中显示 RBF 网络完成训练所用时间为 $datat = 0.078\ 0$,RBF 网络的预测结果见表 2。

表 2 RBF 网络预测结果(2008 年 5 月)

日 期	日耗水量/mm	
	RBF 网络预测结果	实测结果
21	6.929 4	6.899 1
22	6.183 4	6.102 7
23	4.150 5	4.076 6
24	2.943 0	2.867 1
25	1.939 5	1.944 3
26	5.376 8	5.385 7
27	6.267 0	6.247 6
28	6.752 7	6.708 9
29	6.971 2	6.947 0
30	5.872 3	5.775 6
31	7.390 4	7.387 9

(5)模型的检验。取平均气温、日照时数和平均风速为输入因子,建立模型进行作物耗水量预测,并将 2008 年实测土壤含水率通过水量平衡方程计算所得的作物耗水量作为实测值与 RBF 网络模型的预测值进行对比,见图 3。2008 年作物耗水量预测的绝对误差最大为 0.096 7 mm/d、最小为 0.002 5 mm/d、平均为 0.041 9 mm/d;相对误差最大为 2.649 1 %、最小为 0.034 1 %、平均为 0.878 0 %,预测效果较好。

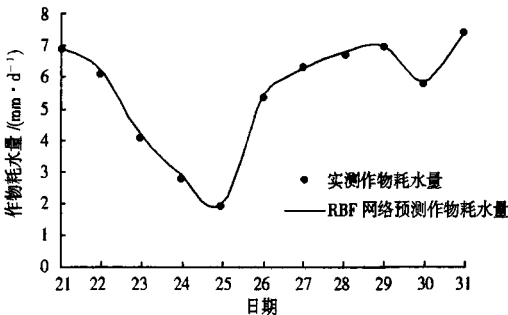


图 3 2008 年 5 月下旬作物耗水量预测分析

4 结 论

(1)用 MATLAB 建立了石河子地区棉田作物耗水量的径向基函数网络预测系统,通过实测数据检验,该网络模型预测的绝对误差最大为 0.096 7 mm/d、最小为 0.002 5 mm/d、平均为 0.041 9 mm/d;相对误差最大为 2.649 1 %、最小为 0.0341 %、平均为 0.878 0 %,表明预测效果较好。

(2)径向基函数网络训练所花费的时间短。在 MATLAB 的命令窗口中可看出 RBF 网络训练用时 0.078 0 s。

(3)神经网络模型是用一系列的线性模型组合来逼近某一非线性关系,因此较以往的线性模型更合理。

(4)此网络模型参数是在一定自然条件和农田管理水平下试验获得的,所以有一定的适用范围。神经网络模型中的参数没有具体的物理意义,不能反映出任何物理关系,这是此类模型的主要局限性。

参考文献:

[1] 姚源松. 新疆棉花高产优质高效理论与实践[M]. 乌鲁木齐:新疆科学技术出版社,2004.

[2] 李远华. 节水灌溉理论与技术[M]. 武汉:武汉大学电力大学出版社,1999.

[3] Fernado M S Lursba. Acceleration techniques for backpropagation algorithm[J]. Neural Neural Networks,1990,1(4):351-370.

[4] W Drow B,Lehr Y Q,Wang J Z. Detemmination df kolmogoroventropy of chaotic attractor included in 1-Dtimeseries of meteorological data[J]. AAS,1991,8(2):243-251.

[5] 高艳萍,于红,崔新忠,等. 基于 L-M 优化算法的水稻螟虫预测模型及其初步应用[J]. 农业工程学报,2007(23):162-165.

[6] 陈红卫,周明耀,瞿益民. 非充分灌溉稻田墒情预测的人工神经网络模型[J]. 扬州大学学报:自然科学版,2005(8):50-54.

[7] 葛哲学,孙志强. 神经网络理论与 MATLAB R2007 实现[M]. 北京:电子工业出版社,2007.

[8] 周开利,康耀红. 神经网络模型及其 MATLAB 仿真程序设计[M]. 北京:清华大学出版社,2005.

[9] 高宁,那陆寿. 基于 MATLAB 的 BP 神经网络在农作物虫情预测中的应用[J]. 计算机与农业,2003(7):16-18.

[10] 赵林明,胡浩云,魏德华,等. 多层前向人工神经网络[M]. 郑州:黄河水利出版社,1999.

[11] 崔远来,马承新,沈细中,等. 基于进化神经网络的参考作物腾发量预测[J]. 水科学进展,2005(16):76-80.

【责任编辑 赵宏伟】