

基于灰色关联度与 BP 神经网络模型的 日参考作物腾发量预测

武开福

(新疆水利水电科学研究院, 乌鲁木齐 830049)

摘要:参考作物腾发量是制定灌溉用水计划、水量分配计划最基本、最重要的内容之一,其精确预测可以提高灌溉预报的精度。采用灰色系统理论中的关联分析方法,对影响作物腾发量的各个气象因素进行关联度分析,挑选出影响作物腾发量的主要气象因子,并以这些主要气象因子为输入向量,以参考作物腾发量为输出向量,建立作物腾发量与主要气象因子之间的 BP 神经网络预测模型。通过实例证明,该方法简单可行,预测精度比较高,能够满足实际生产需要。

关键词:参考作物腾发量;灰色关联分析;BP 神经网络

中图分类号:S501

文献标识码:A

文章编号:1005-3409(2011)02-0237-04

Discuss on Daily Crop Evapotranspiration Forecast Based on Gray Relation Analysis and BP Neural Network

WU Kai-fu

(Institute of Hydraulic & Hydropower Research of Xinjiang, Urumqi 830049, China)

Abstract: Reference crop evapotranspiration is the development of irrigation water plan, the most basic water allocation plan, and one of the most important indicators, the accurate prediction can improve the irrigation forecast precision. In this paper, the relational analysis in the gray system theory was applied to analyze the influences of various meteorological factors on evapotranspiration by selecting the impact of evapotranspiration in the main meteorological factors such as the input vector to refer to evapotranspiration as the output vector, to establish BP neural network model on evapotranspiration and the main meteorological factors. An example shows that the method is simple, relatively high prediction accuracy and can meet actual production needs.

Key words: crop evapotranspiration; gray relation analysis; BP neural network

参考作物腾发量是作物蒸发蒸腾量计算的关键因子,其准确估算对于灌区实时灌溉预报和农田水分管理具有重要作用。许多研究表明参考作物腾发量与气象因素具有较强的非线性关系,而 BP(Back-Propagation, 简称 BP)神经网络是较为成熟也是应用最多的一种非线性函数逼近方法,基于神经网络可以实现参考作物腾发量的模拟与预测。徐俊增等^[1]以气象预报因子为输入,验证了 BP 神经网络模型预测作物参考蒸发蒸腾量的可靠性。但同时也发现:当输入因子较多且自变量相关性较大时,由于信息冗余,导致 BP 网络模型过训练、过拟合和泛化能力不足的问题。

本文引入灰色关联度分析方法对多指标的原始变量进行事先提取,以提取的主要影响因素作为网络输入层,建立基于灰色关联度分析的参考作物腾发量神经网络模型,以期在具备较少气象资料情况下准确预测参考作物腾发量。

1 影响作物腾发量的主要气象因素

1.1 温度

在作物腾发过程中太阳辐射是作物中水分汽化的主要能量,并通过气温的增高来影响作物腾发。气温决定着水汽扩散的快慢和接纳水汽多少的能力,大气温度的层结直接影响着湿度梯度,进而影响作物腾

发。气温高时,作物冠层表面的饱和水汽压大,饱和差大,易于腾发。

1.2 饱和水汽压差

饱和水汽压差是作物冠层表面饱和水汽压与作物冠层表面上某一高度空气中的水汽压之差。作物冠层表面饱和水汽压是温度的函数,空气中的水汽压是气温和相对湿度的函数。因此,作物冠层表面温度和气温对作物腾发速率的影响,可以从饱和水汽压力差中得到反映。饱和水汽压差反映了作物腾发面上的湿度和腾发面上一定高度内的湿度梯度,是影响作物腾发速率的主要因素之一。根据扩散理论,腾发率与饱和水汽压差成正比变化,饱和水汽压差愈大,腾发作用愈强,反之愈小。

1.3 日照时数

日照时数指太阳辐射地面的时间间隔,以小时为单位,分为可照时数和实照时数两种。可照时数相当于日出至日落时间间隔,其时间长短随纬度和季节而变化。实照时数是指在可照时数中扣除云雾遮挡时间,太阳直射光直达地面的时间间隔。一般来说,日照时数越长,作物腾发量越大;日照时数较短,作物腾发量较小。

1.4 相对湿度

相对湿度是空气中的实际水汽压与当时气温下的饱和水汽压的比值,它反映空气距饱和状态远近的程度。一般来说,作物冠层表面的湿度大(接近饱和),其上空及外围湿度小,存在着湿度差。因此,相对湿度的大小可以反映作物冠层表面的水汽向外扩散和交换的快慢。当相对湿度大时,水分向外扩散和交换慢,作物腾发速率小;当相对湿度小时,水分向外扩散和交换快,作物腾发速率大。

1.5 风速

风速的大小,表现在它对紊流扩散作用的强弱和干湿空气交换的快慢上,是影响作物腾发速率的主要

$$\zeta_i(K) = \frac{\min_i \min_k |\sigma_0(K) - \sigma_i(K)| + P \max_i \max_k |\delta_0(K) - \sigma_i(K)|}{|\sigma_0(K) - \sigma_i(K)| + P \max_i \max_k |\sigma_0(K) - \sigma_i(K)|} \quad (2)$$

因素之一。无风时,作物冠层表面上的水汽主要靠分子扩散,水汽压减少得慢,饱和差小,因而作物腾发缓慢。有风时,湍流加强,作物冠层表面的水汽随风和湍流迅速扩散到广大的空间,作物冠层上的水汽压很快减小,饱和差变大,作物腾发加快。风速愈大,紊动愈强烈,干湿空气交换得越快,故作物腾发也越大。

式中: $i=1,2,\dots,m; K=1,2,3,\dots,n_0$ 。 $\min_i \min_k |\sigma_0(K) - \sigma_i(K)|$ 两级最小差,即 $\min_k |\sigma_0(k) - \sigma_i(k_i)|$ 为第一级最小差, $\min_i \min_k |\sigma_0(K) - \sigma_i(K)|$ ——第二级最小差,也就是先求解每一子因素列数据和母因素列数据在不同的 K 点的绝对差,得出每一列绝对差的最小值,再从这些已经得出的最小绝对差中选一个最小值出来; $\min_i \min_k |\sigma_0(K) - \sigma_i(K)|$ ——两级最大差,求法同上,只不过换成求最大值; P ——分辨系数,其值在 $0 \sim 1$ 之间,不同的 P 值对应于不同的关

2 灰色关联分析方法介绍

设有一母因素数列,记作 $X_0 = [X_0(1), X_0(2), X_0(3), \dots, X_0(n)]$, 同时有一系列子因素数列,依次记作 $X_1, X_2, X_3, \dots, X_m$ 。

$$X_1 = [X_1(1), X_1(2), X_1(3), \dots, X_1(n)]$$

$$X_2 = [X_2(1), X_2(2), X_2(3), \dots, X_2(n)]$$

$$X_3 = [X_3(1), X_3(2), X_3(3), \dots, X_3(n)]$$

$$\vdots \quad \quad \quad \vdots$$

$$X_m = [X_m(1), X_m(2), X_m(3), \dots, X_m(n)]$$

由于关联度是两个序列关联性大小的度量,根据这一理论^[3],其计算方法与步骤叙述如下:

2.1 原始数据的标准化处理

由于各子因素数列间的量纲不同,其数量级也相差很大,所以必须对原始数据(包括母因素数列)进行无量纲处理,以消除量纲和量纲单位所带来的不可公度性。通常对原始数据采用标准化的方法来进行变换,经过变换后的原始数据列变为:

$$\sigma_i(K) = \frac{X_i(K) - \bar{X}_i}{S_i} \quad (1)$$

式中: $i=0,1,2,\dots,m; K=1,2,3,\dots,n_0$ 。 $X_i(K)$ ——原始数据; \bar{X}_i ——该气象要素的平均值; S_i ——该气象要素的标准差; $\sigma_i(K)$ ——标准化处理结果。

2.2 关联系数和关联度的计算

对于 $\sigma_0(K)$ 对 $\sigma_i(K)$ 在 K 点的关联系数可以用 $\zeta_i(K)$ 来表示, $\zeta_i(K)$ 的表达式见式(2)。

关联度,它可以用来提高关联系数之间的差异性的,理论证明, P 值越小,说明分辨率越高,一般可取 $0.5^{[4-5]}$ 。

因为每一列的数据都有 n 个,所以计算出来的关联系数也有 n 个,这样信息就过于分散。如果把各点的关联系数取平均值,就可把各点的关联系数集成为一个值,即为关联度。子因素数列的关联度大小代表了与母因素数列的关系密切情况,若大,则代表关系密切,从而该因素的影响就大。

3 BP 神经网络模型

3.1 BP 神经网络简介

BP 神经网络是目前使用最广泛,也是发展最成熟的一种神经网络。它是由输入层、一个或多个隐含层和输出层连接而成,其每一层又由许多人工神经元(节点)组成,一层内的神经元只和该层邻近的下一层各神经元连接^[6-9]。

BP 网络间连接权在网路的学习过程中不断的得到修正,使输入层与隐含层之间、隐含层与输出层之间的两组权所构成的网络能实现学习样本中输入矢量与输出矢量间特定的映射关系,权的分布体现了各输入分量在输入矢量中所占特征强度的分布。当学习样本提供给网络以后,神经元的激活值,从输入层经各中间层向输出层传播,在输出层的各神经元获得网络的输入响应。在这之后,按减小希望输出与实际输出误差的方向,从输出层经中间层的逐层修正各连接权,最后回到输入层。各个连接权的调整按标准误差逆传播算法或者累积误差逆传播算法进行。通过修改各层神经网络的权值,使得误差不断减小,当误差达到预定的要求时,网络学习训练过程就结束,见图 1。

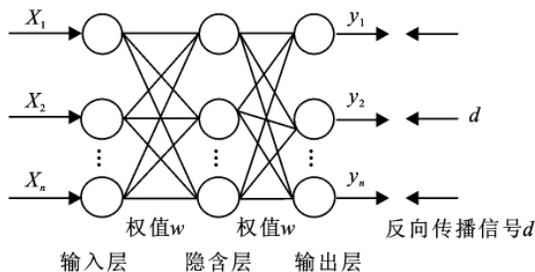


图 1 BP 神经网络结构

3.2 传递函数和权重

BP 神经网络的训练就是通过不断调整各层的权重值,使得期望输出值与实际值的误差达到最小。模型权重的初始值为随机赋值,而输入层与隐含层的传递函数则采用如下 Sigmoid 函数,隐含层与输出层的传递函数采用线性函数:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (3)$$

表 1 作物腾发量与各气象要素的关联度及排序

气象要素	日均温度	日最高温度	日最低温度	日均水汽压	日均相对湿度	日均风速	日照时数
关联度	0.8693	0.8441	0.8499	0.7606	0.6362	0.7427	0.7532
位次	1	3	2	4	7	6	5

根据表 1,选择关联度大于 0.7 的气象因子作为 BP 神经网络模型的输入层,即日均温度、日最高温度、日最低温度、日照时数、水汽压与风速 6 个气象因

式中: \$x\$——样本输入值; \$f(x)\$ 的变化区间为 \$[0, 1]\$。

3.3 数据的预处理

在网络学习过程中,为便于训练,更好地反映各因素之间的相互关系,必须对样本数据进行预处理。一般网络输出向量的各分量值应在 \$[0, 1]\$ 之间。为使较大的输入落在神经元激励函数梯度大的区域,对输入向量的各分量亦取 \$[0, 1]\$ 特征值为佳。因此训练神经网络之前,将输入样本归一化,处理如下:

$$x_{\text{norm}} = \frac{x_i - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \quad (4)$$

其中: \$x_{\text{norm}}\$——样本归一化值; \$x_{\text{max}}\$——样本最大值; \$x_{\text{min}}\$——样本最小值; \$x_i\$——样本值。

3.4 评价指标

本次 BP 神经网络的评价指标包括:标准偏差 \$\delta\$、偏差系数 \$C_v\$、网络模型效率 \$E\$,用这 3 个评价指标表明预测值偏离实际值的程度及预测效率。

$$\delta = \sqrt{\frac{\sum_{i=1}^n (Y_{SC,i} - Y_{BP,i})^2}{n-1}} \quad (5)$$

$$C_v = \sqrt{\frac{\sum_{i=1}^n (Y_{SC,i} - Y_{BP,i})^2}{n-1}} \frac{1}{\bar{Y}_s} \quad (6)$$

$$E = 1 - \frac{\sum_{i=1}^n (Y_{SC,i} - Y_{BP,i})^2}{\sum_{i=1}^n (Y_{SC,i} - \bar{Y}_s)^2} \quad (7)$$

式中: \$Y_{SC,i}\$——实测值; \$Y_{BP,i}\$——BP 网络预测值; \$\bar{Y}_s\$——实测平均值。

4 组合模型在作物腾发量预测中的应用实例

本文利用新疆尉犁县气象站 2008 年 5 月 1 日至 10 月 31 日逐日气象资料,按照灰色关联度分析方法,以日作物腾发量为尺度,对作物腾发量及其影响因素进行分析。

4.1 作物腾发量影响因素的灰色关联度计算

按照前述的灰色关联度计算步骤,求出各气象要素与日作物腾发量的关联度,并按大小顺序排列,结果见表 1。

素为输入层。

4.2 BP 神经网络模型构建

本文构建的 BP 神经网络模型包括输入层、隐含

层和输出层 3 层。考虑到气象资料的获取程度,输入层为日均温度、日最高温度、日最低温度、日照时数、水汽压与风速 6 个气象要素(经过灰色关联分析确定为主要影响因素),网络输出变量则只有 1 个,根据这种情况,通过多次训练比较后确定隐含层节点为 15 个,即网络模型的拓扑结构为 6-15-1。

4.3 模型效果的检验与评价

输入上述 BP 网络模型,进行指导学习训练,其训练模型评价指标见表 2(以相对误差在 $\pm 20\%$ 以内作为合格值)。

表 2 BP 模型评价指标

BP 模型结构	δ/mm	C_0	$E/\%$	合格率/ $\%$
6-15-1	0.668	0.147	82.4	81.0

从表中数据可知基于灰色关联分析的 BP 模型 6-15-1 训练效果不错,预测精度较为满意。

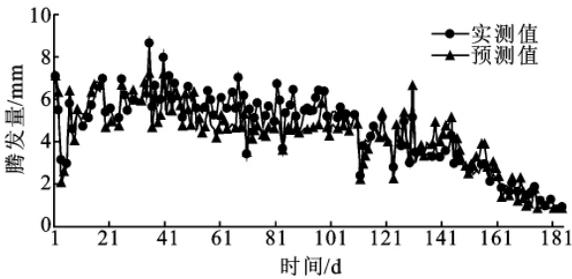


图 2 BP 网络模型 6-15-1 预测结果

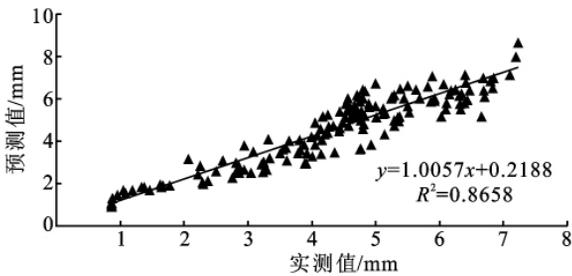


图 3 BP 网络模型 6-15-1 预测值与实测值比较

5 讨论与结论

本文将人工神经网络应用于日作物腾发量的预报中,建立了基于灰色关联度和 BP 神经网络的日作物腾发量预报模型,并对新疆尉犁县日作物腾发量进行了预测。结果表明:通过灰色关联分析选取影响作物腾发量主要气象要素的情况下,可以通过训练好的神经网络模型对日作物腾发量进行预测。

(1)运用灰色关联分析,不用考虑样本量的多少和有样本无典型分布规律,特别是当系统信息贫乏,相关数据短缺,统计数据灰度较大时,要挖掘影响系统行为特征的主要因素,该方法行之有效,减少挑选变量分析计算工作时间。

(2)基于灰色关联分析的 BP 神经网络模型,能够把影响参考作物腾发量的主要气象因素挑选出来,并将这些因素结合起来预测系统未来发展趋势。实例证明,该组合模型具有较高的拟合精度和预测精度,不失为一种较为简便的预报方法。

(3)模型对参考作物腾发量中极大值数据预测出现较大的误差,可能是由于影响系统的因素并没有完全考虑进去,也可能是某些随机因素的侵入,使系统受到影响发生变化。因此在作物腾发量预测中,应根据区域的实际情况,定量与定性相结合,以提高预报的准确度。

参考文献:

- [1] 徐俊增,彭世彰,李道西,等.基于气象预报的参考作物蒸发蒸腾量的神经网络预测模型[J].水利学报,2006,37(3):376-379.
- [2] 刘钰,Pereira L S.气象数据缺测条件下参照腾发量的计算方法[J].水利学报,2001(3):11-17.
- [3] 邓聚龙.灰色预测与决策[M].武汉:华中工学院出版社,1985.
- [4] 邓聚龙.灰色系统基础方法[M].武汉:华中理工大学出版社,1987.
- [5] 刘思峰,党耀国,方志耕,等.灰色系统理论及应用[M].北京:科学工业出版社,2007.
- [6] 蒋任飞,阮本清,韩宇平,等.基于 BP 神经网络的参照腾发量预测模型[J].中国水利水电科学研究院学报,2005,3(4):308-311.
- [7] 霍再林,史海滨,李为萍,等.参考作物蒸发蒸腾量的人工神经网络模型研究[J].沈阳农业大学学报,2004,35(6):436-438.
- [8] 高瑞忠,朝伦巴根,于婵,等.基于随机样本的神经网络模型估算参考作物腾发量[J].农业工程学报,2006,22(2):42-45.
- [9] 蒋任飞,阮本清,韩宇平,等.基于 BP 神经网络的参照腾发量预测模型[J].中国水利水电科学研究院学报,2005,3(4):308-311.