

DOI :10.3969/j.issn.1004-6933.2012.03.009

基于 BP 神经网络的桂林生态城市建设需水量预测

温家鸣 郭纯青 李新建 李文斌

(桂林理工大学环境科学与工程学院, 广西 桂林 541004)

摘要 根据桂林市经济社会历年统计的主要指标数据,运用 SPSS 社会科学统计软件分析并选取桂林市辖区生态城市建设需水量的显著性影响因子,采用改进的归一化进行非线性规格化数据处理,基于 Matlab 建立 BP 神经网络模型,预测桂林市辖区生态城市建设需水量,结果表明,预测结果与原始数据的平均相对误差为 1.19%,最大为 2.08%,最小为 0.28%。该模型具有较高的预测精度和良好的泛化能力,BP 神经网络与 SPSS 软件优化组合模型,可用于需水量预测。

关键词 需水量预测;生态城市;BP 神经网络;SPSS;桂林

中图分类号:TV213;X24

文献标识码:A

文章编号:1004-6933(2012)03-0047-04

Water demand forecast of eco-city construction in Guilin City based on BP neural network

WEN Jia-ming, GUO Chun-qing, LI Xin-jian, LI Wen-bin

(College of Environmental Science and Engineering, Guilin University of Technology,
Guilin 541004, China)

Abstract : According to statistics of the major economic and social indicators in Guilin City over recent years, significant factors of water demand for eco-city construction in Guilin City were selected using the SPSS software, and an improved normalization method was used for nonlinear normalization data processing. The MATLAB-based BP neural network model was established to forecast the water demand of eco-city construction in Guilin City. The forecasted results, compared with the original data, had an average relative error of 1.19%, a maximum error of 2.08%, and a minimum error of 0.28%. The model is shown to have high accuracy and good generalization ability, and the BP neural network model, combined with the SPSS software optimization, is applicable to water demand forecasting.

Key words : water demand forecast; ecological city; BP neural network; SPSS; Guilin

生态城市作为一种理想的模式,它的最终目的是把社会、经济发展与生态环境建设结合起来,努力创造一种社会和谐、经济高效、生态良性循环的人类居住区形式,是一种人与自然高度和谐可持续发展的城市新模式^[1]。桂林市是国家级历史文化名城和世界著名的风景旅游城市,随着城市人口的增长以及城市化、工业化的发展,其主要河流之一漓江生态环境受到破坏,桂林市存在潜在的缺水威胁,目前已初露端倪。

水资源,尤其是淡水资源,是基础性的自然资源

和战略性的经济资源,在保障社会经济可持续发展中具有不可替代的作用^[2]。需水量预测是水资源开发利用、社会经济发展、产业布局的重要参考指标,对生态城市建设具有重要意义。笔者结合桂林生态城市建设规划,对桂林生态城市建设需水量进行研究,以为生态城市建设中水资源的优化配置及社会经济发展提供依据。

1 桂林市概况

桂林市位于广西壮族自治区东北部,辖秀峰、叠

基金项目:广西自然科学基金(2010GXNSFE013005);广西科学基金(桂科基 0991026)

作者简介:温家鸣(1983—),男,硕士研究生,研究方向为岩溶水文生态。E-mail:Wan_jia_ming_@126.com

通讯作者:郭纯青,教授。E-mail:guochunqing@yahoo.com.cn

彩、象山、七星、雁山等 5 城区和阳朔、临桂、灵川、兴安、全州、荔浦、平乐、恭城、灌阳、资源、龙胜、永福等 12 个县,市域总面积 27 809 km²,城市建成区(含西城区)面积 58 km²。2010 年 GDP 为 1 108.63 亿元,比上年增长 13.8%。其中,第一产业增加值 202.60 亿元,增长 4.8%;第二产业增加值 502.04 亿元,增长 20.7%;第三产业增加值 403.99 亿元,增长 10.2%。第一、二、三产业增加值占生产总值的比重分别为 18.3%、45.3%和 36.4%。全市人均生产总值 21 611 元,增长 13.1%。

桂林市水资源丰富,境内河流属珠江流域西江水系。市区主要有漓江、桃花江及“四湖”(桂湖、榕湖、杉湖、木龙湖),湖泊水库总库容达 1.9 亿 m³。可供开发的地下水总储量约为 31 亿 m³。桂林境内的主要河流有漓江、湘江、洛清江、浔江、资江 5 条,另有集雨面积 100 km² 以上的支流 65 条,1 000 km² 以上的支流有 10 条,全市多年平均总水量为 403.81 亿 m³。年平均降雨量为 1 949.5 mm,平均蒸发量为 1 490~1 905 mm,用水量较大(表 1)。

表 1 2009 年桂林市、广西区和全国用水量统计

地区	用水量/亿 m ³					人均用水量/m ³
	农业	工业	生活	生态	总量	
桂林市	33.67	5.55	2.87	3.13	45.22	923.00
广西区	195.30	54.00	48.40	5.70	303.40	627.30
全 国	3 723.10	1 390.90	748.20	103.00	5 965.20	448.00

注:桂林市用水量数据来源于 2009 年《广西区水资源公报》;广西区、全国用水量数据来源于 2010 年《中国统计年鉴》。

2 BP 神经网络应用

2.1 BP 神经网络模型

人工神经网络(artificial neural network,ANN)是由人工神经元互联组成的网络,它是从微观结构和功能上对人脑的抽象、简化^[3],是一种模仿人脑结构及其功能的非线性信息处理系统。神经网络是隐式数学处理方法,由网络训练从数据中概括出来的知识(因果规律),以多组权值和阈值的形式分布存储于各神经元中,即隐含在模型软件内。与传统的按预编程序串行集中的 Von Neumann 计算机相比较而言,ANN 人工神经网络学习能力强,有非线性、非局部性、非定常性和非凸性的优点^[4],其应用领域十分广泛。

BP(back-propagation)神经网络是 ANN 中最具代表性和广泛应用的一种,是基于误差反向传播算法的多层前馈神经网络,由 Rumelhart 和 Mc Celland 及其研究小组在 1986 年研究并设计出来^[5]。BP 神经网络模型拓扑结构包括输入层、输出层和若干隐含层,每层由若干个结点组成,每一个结点表示一个神

经元,上层结点与下层结点之间通过权连接,同一层结点之间没有联系(图 1)。

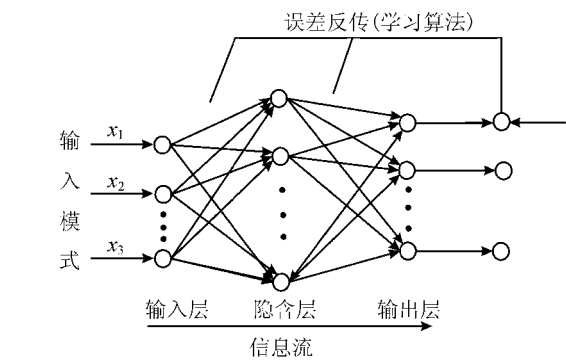


图 1 BP 神经网络结构

2.2 BP 在 Matlab 中的程序设计

以桂林市辖区统计数据为背景,采用 SPSS 软件进行需水量相关分析^[6-9],从中选取相关性、代表性好的 5 个影响因子(表 2):GDP、人口数量、城镇人均可支配收入、人均用水量、万元 GDP 用水量。2000—2009 年桂林市辖区相关统计情况见表 3。5 个影响因子均与需水量呈现极显著性($\alpha = 0.01$),其中与人均用水量和万元 GDP 用水量呈正相关,相关系数分别为 0.904、0.979,说明随着需水量的增加,其呈现上升的趋势;而与 GDP、人口数量、城镇人均可支配收入呈负相关,相关系数分别为 -0.842、-0.827、-0.805,说明随着需水量的增加,三者呈现下降的趋势。5 个影响因子的相关性从大到小排名为:万元 GDP 用水量、人均用水量、GDP、人口数量、城镇人均可支配收入。BP 神经网络模型由 5 个

表 2 需水量与影响因素相关分析

项目	GDP	人口数量	城镇人均可支配收入	人均用水量	万元 GDP 用水量
相关系数	-0.842	-0.827	-0.805	0.904	0.979
P 值(双侧)	0.002	0.003	0.005	0.000	0.000

注:样本数为 10 个,相关性显著水平为 0.01。

表 3 桂林市辖区统计数据

年份	GDP/万元	人口数量/万人	城镇人均可支配收入/元	人均用水量/万 m ³	万元 GDP 用水量/m ³	需水量/亿 m ³
2000	998 389	48.50	5 834	316.87	1 539 279.8	15 368
2001	1 154 051	66.83	6 666	230.39	1 334 169.8	15 397
2002	1 299 514	69.09	7 315	195.04	1 036 926.1	13 475
2003	1 375 794	71.00	7 785	187.06	965 333.5	13 281
2004	1 588 518	72.16	8 177	168.69	766 311.7	12 173
2005	1 807 147	73.48	8 917	166.07	675 263.3	12 203
2006	2 068 709	74.42	9 899	153.96	553 872.0	11 458
2007	2 422 182	75.69	12 200	153.73	480 393.3	11 636
2008	2 818 196	75.93	14 146	153.34	413 136.6	11 643
2009	3 137 973	75.79	15 451	146.54	353 922.7	11 106

注:数据来源于 2000—2009 年广西水资源公报,2001—2010 年桂林经济社会统计年鉴、广西统计年鉴。

节点的输入层、1 个节点的输出层和多个隐含层组成,采用共轭梯度法确定各层之间的连接权量,各神经元经过特性为 Sigmoid^[10]型作用函数运算后,把输入层神经元的信息传递到输出层神经元,最后输出得到桂林市辖区需水量的训练结果及预测结果。

2.2.1 统计数据的预处理

一般 BP 神经网络模型数据需归一化是因为 Sigmoid 函数的取值为[0 ,1],网络最后一个节点的输出也是如此,而且原始数据量纲不同和指标数值存在数量级的差异导致学习速度慢,因此需对原始数据进行预处理。在训练时将实际数据规范到[0.2 ,0.8]区间^[11],由 $\ln 0.2 = -1.609$, $\ln 0.8 = -0.223$, $0 \leq (X - X_{\min}) / (X_{\max} - X_{\min}) \leq 1$,可得

$$0.2 \leq \frac{1}{\exp[0.223 + (1.609 - 0.223)(X - X_{\min}) / (X_{\max} - X_{\min})]} \leq 0.8$$

数据归一化处理公式如下：

$$X' = 1 - \frac{1}{\exp[0.223 + (1.609 - 0.223)(X - X_{\min}) / (X_{\max} - X_{\min})]}$$

式中: X' 为目标数据; X 为原始数据; $X_{\max}$ 、 $X_{\min}$ 分别为原始数据的最大值和最小值。

经过预处理的统计数据全部在[0.2 ,0.8]之间(表 4),可以大大加快网络学习速度,而数据间的联系并不减少。

表 4 归一化后的统计数据

年份	GDP	人口数量	城镇人均可支配 收入	人均 用水量	万元 GDP 用水量	需水量
2000	0.2000	0.2000	0.2000	0.8000	0.8000	0.7980
2001	0.2766	0.6831	0.2903	0.5956	0.6690	0.8000
2002	0.3417	0.7173	0.3537	0.4608	0.5674	0.6278
2003	0.3734	0.7433	0.3960	0.4246	0.5386	0.6037
2004	0.4541	0.7579	0.4292	0.3318	0.4481	0.4331
2005	0.5262	0.7735	0.4869	0.3174	0.4009	0.4386
2006	0.6000	0.7840	0.5546	0.2468	0.3317	0.2859
2007	0.6819	0.7975	0.6803	0.2454	0.2860	0.3258
2008	0.7539	0.8000	0.7585	0.2430	0.2414	0.3273
2009	0.8000	0.7985	0.8000	0.2000	0.2000	0.2000

2.2.2 BP 模型预测的 Matlab 实现

利用 MATLAB 中的 NNbox,调用相关函数实现 BP 模型的学习、训练、拟合及预测过程。桂林市辖区统计数据分为:训练集 2000—2008 年,测试集 2009 年。MATLAB 训练运行程序如下：

```
p=[ 0.2000 0.2766 0.3417 0.3734 0.4541 0.5262 0.6000
0.6819 0.7539 ;0.2000 0.6831 0.7173 0.7433 0.7579 0.7735
0.7840 0.7975 0.8000 ;0.2000 0.2903 0.3537 0.3960 0.4292
0.4869 0.5546 0.6803 0.7585 ;0.8000 0.5956 0.4608 0.4246
0.3318 0.3174 0.2468 0.2454 0.2430 ;0.8000 0.6690 0.5674
0.5386 0.4481 0.4009 0.3317 0.2860 0.2414 ]
t=[ 0.7980 0.8000 0.6278 0.6037 0.4331 0.4386 0.2859
0.3258 0.3273 ]
```

```
net = newff( minmax( p ),[ 3 ,1 ],{ 'tansig ', 'tansig ', '
purelin '}, 'traingdm ');
net.trainparam.show = 50 ;
net.trainparam.mc = 0.9 ;
net.trainparam.lr = 0.05 ;
net.trainparam.epochs = 5000 ;
net.trainparam.goal = 0.001 ;
net = ini( net );
[ net,tr ]= train( net ,p ,t );
a = sin( net ,p );
E = t - a ;
MSE = ms( E );
figure( 1 );
X = 1 :9 ;
plot( X ,a , 'r * ',X ,t , 'bo ');
legend( '预测值 ', '真实值 ',12 );
title( '训练仿真输出 ', 'fontsize ',12 );
ylabe( '函数值 ', 'fontsize ',12 );
xlab( '样本 ', 'fontsize ',12 )
TRAINGDM ,Performance goal met.
a = Columns 1 through 8
0.7952 0.7930 0.6475 0.5831 0.4815 0.3731 0.3170
0.3400
```

```
Column 0.3092
MATLAB 测试运行程序如下：
p_test=[ 0.8000 ;0.7985 ;0.8000 ;0.2000 ;0.2000 ]
a = sin( net ,p_test )
a = 0.2275
```

3 BP 模型预测结果分析

由图 2 可以看出,BP 神经网络经过 2 088 次迭代就达到预期的预测精度($\epsilon = 0.001$),且训练过程平稳,运行速度也较快。图 3 说明训练样本原始值与预测值拟合较好,预测具有一定的可行性。BP 神经网络的输出数据还原后与实际需水量的误差分析(表 5)结果表明,最大相对误差为 2.08%,最小相对误差为 0.28%,平均相对误差为 1.19%,误差变化

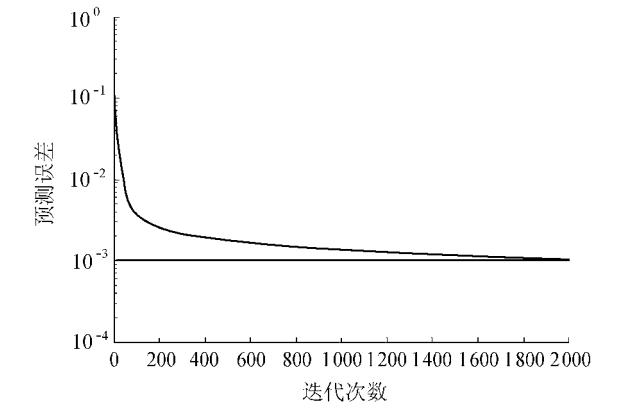


图 2 训练误差曲线

平缓,预测结果较稳定,精度较高。因此采用 BP 神经网络模型进行需水量预测具有较强的预测能力和较高的可信度。

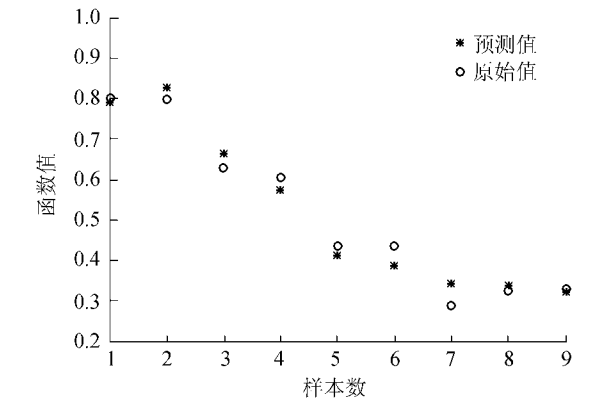


图 3 训练样本拟合

表 5 BP 神经网络需水量预测误差分析

年份	实际用水量/ 亿 m ³	预测需水量/ 亿 m ³	相对 误差/%	年份	实际用水量/ 亿 m ³	预测需水量/ 亿 m ³	相对 误差/%
2000	15 368	15 325	0.28	2005	12 203	11 861	2.80
2001	15 397	15 292	0.68	2006	11 458	11 596	1.20
2002	13 475	13 643	1.25	2007	11 636	11 702	0.57
2003	13 281	13 124	1.18	2008	11 643	11 561	0.70
2004	12 173	12 449	2.27	2009	11 106	11 215	0.98

4 结 语

BP 神经网络模型的原始数据预处理选取有别于一般的归一化处理,优化了网络学习速度,采用 SPSS 软件分析、提取需水量主要的影响因素,优化了网络结构,这些都提高了网络预测的性能。需水量预测的目的是为水资源规划和用水系统的优化调度提供科学决策所需的未来信息,并努力提供这种信息的准确性和可靠性。BP 神经网络具有高度非线性函数映射功能,应用于需水量预测精度较高,泛化能力好,使其应用于实际具有更大的可行性。

参考文献:

[1] 王翠云.生态城市建设评价研究:以兰州市为例[D].兰州:兰州大学,2004.

[2] 卢冰.桂林市水资源优化配置研究[D].武汉:武汉大学,2005.

[3] 徐丽娜.神经网络控制[M].北京:电子工业出版社,2003:1-9.

[4] 苑希民,李鸿雁,刘树坤,等.神经网络和遗传算法在水科学领域的应用[M].北京:中国水利水电出版社,2002:2-4.

[5] 虞和济,陈长征,张省,等.基于神经网络的智能诊断[M].北京:冶金工业出版社,2000:1-4.

[6] 杜强,贾丽艳.SPSS 统计分析从入门到精通[M].北京:

人民邮电出版社,2009:257-270.

[7] 宋志刚,谢蕾蕾,何旭洪.SPSS16 实用教程[M].北京:人民邮电出版社,2008:132-153.

[8] 孙艳玲,何源,李阳旭.例说 SPSS 统计分析[M].北京:人民邮电出版社,2010:183-188.

[9] 张维忠,张甜.SPSS 统计分析与行业应用案例详解[M].北京:清华大学出版社,2011:75-78.

[10] 高隽.人工神经网络原理及仿真实例[M].北京:机械工业出版社,2003:9-10.

[11] 张文鸽,吴泽宁,谿洪波.BP 人工神经网络的改进及其应用[J].河南科学,2004,21(2):202-206.

(收稿日期:2011-09-14 编辑:徐娟)

(上接第 8 页)

[25] 关效如.我国东部高纯二氧化碳成因[J].石油实验地质,1990,12(3):248-258.

[26] 杜乐天.地球排气作用的重大意义及研究进展[J].地质论评,2005,51(2):174-179.

[27] 李文鹏,郝爱兵.中国西北内陆干旱盆地地下水形成演化模式及其意义[J].水文地质工程地质,1999(4):28-32.

[28] 郝爱兵,李文鹏,梁志强.利用 TDS 和 $\delta^{18}\text{O}$ 确定溶滤和蒸发作用对内陆干旱区地下水咸化贡献的一种方法[J].水文地质工程地质,2000(1):4-6.

[29] 丁宏伟,张举.河西走廊地下水水化学特征及其演化规律[J].干旱区研究,2005,22(1):24-28.

[30] 郭永海,沈照理,钟佐兑.河北平原深层碱性淡水形成的水文地球化学模拟:以沧州、保定地区为例[J].地球科学-中国地质大学学报,2002,27(2):157-162.

[31] 顾芷娟,郭才华,李彪,等.壳内低速高层成因初步探讨[J].中国科学:B 辑,1995,25(1):106-112.

[32] 胡圣标,何丽娟,汪集.中国大陆地区大地热流数据汇编:第三版[J].地球物理学报,2001,44(5):611-626.

[33] 徐常芳.中国大陆壳内与上地幔高层成因及唐山地震机理研究[J].地学前缘,2003,10(5):101-111.

[34] CHEN J S, WANG C Y. Rising springs along the Silk Road [J]. Geology, 2009, 37(3):243-246.

[35] CHEN J S, LI L, WANG J Y, et al. Groundwater maintains dune landscape [J]. Nature, 2004, 432:459-460.

[36] CHEN J S, ZHAN X, SHENG X F, et al. Formation mechanisms of megadunes and lakes in the Badain Jaran Desert, Inner Mongolia [J]. Chinese Science Bulletin, 2006, 51(24):3026-3034.

[37] 杨淇越,吴锦奎,丁永建,等.锡林河流域地表水和浅层地下水的稳定同位素素研究[J].冰川冻土,2009,31(5):850-856.

(收稿日期:2011-09-14 编辑:徐娟)