

基于主成分分析的 BP 神经网络对 南京市水资源需求量预测

王春娟¹, 冯利华¹, 罗伟²

(1. 浙江师范大学 地理与环境科学学院, 浙江 金华 321004; 2. 江西省庐山自然保护区管理处, 江西 庐山 332900)

摘 要: 以南京市为例, 利用 1999–2010 年的总用水量数据, 采用主成分分析法对影响水资源需求量的 9 个因子进行主要影响因子分析, 根据确定的主要影响因子构造 BP 神经网络的输入样本, 从而进行不同水平的年总需水量预测。结果表明: 人口、GDP、万元 GDP 用水量、人均水资源量、污水年排放量为影响研究区需水量的主要因子, 将此作为主要因子构造 BP 神经网络的输入样本, 确定网络输入节点数, 建立南京市总需水量预测模型。模拟计算结果表明, 基于主成分分析的 BP 神经网络模型, 预测结果的平均误差小于 0.2 亿 m^3 。

关键词: 需水预测; 主成分分析法; BP 神经网络

中图分类号: TV183; TU991.31

文献标识码: A

文章编号: 1672-643X(2012)06-0006-04

Forecast of water demand by using BP neural network based on principle component analysis in Nanjing

WANG Chunjuan¹, FENG Lihua¹, LUO Wei²

(1. College of Geography and Environmental Sciences, Zhejiang Normal University, Jinhua 321004, China;

2. Lushan Nature Reserve Management Office of Jiangxi Province, Lushan 332900, China)

Abstract: Taking the water demand data from 1999 to 2010 of Nanjing for example, this paper analyzes the main factors that influence the water resource quantity based on the principle component analysis method. According to these main factors, the input samples of BP neural network are determined. Thereby, the BP neural networks can be trained to predict. The results show that population, GDP, water consumption of ten thousand yuan GDP, water resources per capita and volume of sewage discharge per year are the primary indexes that affect water resource demand. The corresponding prediction modeling outcome shows that the simulated experiment is quite fit for the practical situation and the average error of prediction is less than $0.2 \times 10^8 \text{ m}^3$.

Key words: water demand prediction; principle component analysis; BP neural networks

一个区域水资源的供需状况是水资源规划、管理的重要基础, 是该区域可持续发展研究和水资源安全战略研究中的一项目标内容^[1-2]。随着人口的增长、工农业生产发展, 水资源供需矛盾日益加剧, 所面临的水危机日益严重; 加上社会经济用水挤占生态用水, 使得天然河湖萎缩、消失, 土地荒漠化等, 造成生态环境失调, 严重阻碍了经济社会的持续发展^[3-5]。用水量的高速增长和水资源的短缺使得水资源规划和用水系统的优化调度越来越重要, 进行需水量预测则成为了实现水资源规划和管理的有效手段之一^[6]。

国内外关于需水预测研究历史悠久, 但由于近年来水资源与经济社会之间的矛盾愈加突出, 对于水资源需求的研究方式也呈现多样化发展趋势^[7]。用于需水预测的模型有回归分析模型、灰色模型、指数模型、定额模型等。针对水资源需求量预测问题是一个典型的非线性问题, 变量之间的关系极为复杂, 各变量间的影响权数很难确定, 运用以上方法对其预测, 不仅工作量大, 而且效果不甚理想。为了能得到符合社会经济发展趋势的结果, 也为了增加水资源配置研究工作的操作性, 在进行需水预测时, 应选择合理的方法。其中需水预测模型中的影响因子

收稿日期: 2012-09-01; 修回日期: 2012-09-19

基金项目: 国家自然科学基金项目(41171430、40771044)

作者简介: 王春娟(1984-), 女, 陕西宝鸡人, 硕士研究生, 研究方向为水文学与水资源、资源地理。

通讯作者: 冯利华(1955-), 男, 浙江建德人, 教授, 主要从事水文学与水资源的教学与研究工作。

的选取至关重要,影响因子的数目影响着预测结果的准确性。因此本文首先运用主成分分析法来挑选南京市需水量的主要的影响因子,接着运用具有广泛应用领域的 BP 神经网络法以此构造需水预测模型,对南京市水资源需求量进行预测。

1 南京市水资源现状分析

南京市位于北纬 $31^{\circ}14' \sim 32^{\circ}37'$,东经 $118^{\circ}22' \sim 119^{\circ}14'$,地处长江下游平原,东望大海,西达荆楚,南接皖浙,北联江淮,属于北亚热带季风湿润气候区。处于西风环流控制之下的南京,季风显著,四季分明,冬季在欧亚大陆气团的影响下,晴朗、干燥、寒冷;夏季在欧亚大陆低压的影响下,炎热、雨水充沛,多年平均年降水量 $1\,081.9\text{ mm}$ 。其中,春秋季短,冬夏季长,四季各具特色。

截至 2010 年全市平均降水量 $1\,219.4\text{ mm}$,折合降水总量 80.26 m^3 ,比多年平均降水量多 12.7% 。全市水资源总量 37.17 亿 m^3 (不包括过境客水资源量 $1\,0347\text{ 亿 m}^3$),其中地表水资源量 31.15 亿 m^3 ,地下水资源量 7.146 亿 m^3 ,重复极端水资源量 1.127 亿 m^3 。全市实际用水总量 41.84 亿 m^3 。

2 水资源需求量的影响因素分析

影响区域水资源需求量的因素众多,可分为社会经济因素和自然因素。水资源需求量可通过若干特征变量来表示,若把这些变量界定为评价指标,这些评价指标就成为测定水资源需求量的指示器,以便为水资源进行优化配置奠定基础^[8]。

选择区域水资源需求量的评价指标应遵循可测性原则、可靠性原则、充分性原则。遵循以上原则,针对南京市的水资源供需情况,以 1999–2010 年序列资料作为基础数据,并参考文献,从南京市统计年鉴(2000–2011)、水资源公报(1999–2010)中选取 9 个因子: X_1 :人口(万人); X_2 :GDP(亿元); X_3 :降水量(亿 m^3); X_4 :水资源利用率(%); X_5 :万元 GDP 用水量(m^3); X_6 :居民生活用水量(亿 m^3); X_7 :人均水资源量(m^3); X_8 :日供水能力(万 m^3); X_9 :污水年排放量(10^6 t)。应用 SPSS19.0 分析软件对样本进行主成分分析。根据特征值级主成分贡献率可得,前 3 个特征值所对应的累计贡献率 $E_2 = 90.023\%$ 符合分析的要求(一般情况下取 $E_m > 85\%$)。根据主成分载荷矩阵可得, X_1 、 X_2 与第一主成分正相关, X_5 、 X_9 与第一主成分负相关, X_3 、 X_7 与

第二主成分正相关, X_4 与第二主成分负相关,因此,南京市水资源需求量的影响因子可归纳为经济发展因素、水资源开发利用因素、水环境破坏因素三大类。

3 基于 BP 神经网络的南京市水资源需求量预测

3.1 BP 神经网络的原理

人工神经网络(Back Propagation, BP)模型具有任意精度逼近任意非线性函数的映射能力。BP 神经网络有三部分组成:输入层、隐含层和输出层。输入层和输出层都为单层,隐含层可为单层,亦可为多层^[9]。其学习过程分为信息的正向传播过程和误差的反向传播过程两个阶段,外部输入的信号经输入层、隐含层的神经元逐层处理,向前传播到输出层,得出结果。若在输出层得不到期望的结果,则转入逆向传播过程,将实际值与网络输出之间的误差沿原连接通路返回,通过修改各层神经元的连接权值,减少误差,然后再转入正向传播过程,反复迭代,直到误差小于给定的值为至^[10]。

在 BP 神经网络的构建中,所选用的传递函数是反映下层输入对上层节点刺激脉冲强度的函数,一般取(0,1)内连续取值的 Sigmoid 函数:

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

BP 网络的输入层一般没有阈值和激活函数,即输入层神经元的输出等于输入神经元的输入,而隐含层和输出层神经元的输出则必须经过权重和阈值的处理。设网络具有 m 层,令 y_j^m 表示第 m 层中第 j 个节点的输出,而 y_j^0 就等于 x_j ,即第 j 个输入,令 W_{ij}^m 表示从 y_j^{m-1} 到 y_j^m 之间的连续加权, θ_j^m 表示第 m 层第 j 结点的阈值, BP 神经网络训练步骤如下^[11]:

- (1) 将各权值和阈值赋予 $(-1, 1)$ 间的随机数;
- (2) 从训练数据组中选一数据对 (X^k, T^k) , 将输入变量加到输入层($m = 0$)使:

$$y_i^0 = x_i^k \quad (\text{对所有 } i \text{ 点}) \tag{2}$$

式中: k 表示训练图形号。

- (3) 信号通过网络向前传播,即利用关系式:
$$y_j^m = F(S_j^m)F(\sum_i W_{ij}^m y_i^{m-1} + \theta_j^m) \tag{3}$$
- (4) 计算输出层每个节点的误差值:
$$\delta_j^m = y_j^m(1 - y_j^m)(T_j^k - y_j^m) \tag{4}$$
这个误差由实际输出值和目标要求值之差获得;
- (5) 计算前面各层每个结点的误差值:

$$\delta_j^{m-1} = F(S_j^{m-1}) \sum_i W_{ij}^m \delta_i^m \tag{5}$$

由逐层反传误差算得($m = m, m - 1, \cdots, 1$);

(6) 反向逐层修正权值和阈值:

$$W_{ij}^m(t+1) = W_{ij}^m(t) + \eta \delta_j^m y_i^{m-1} + \partial [W_{ij}^m(t) - W_{ij}^m(t-1)] \quad (6)$$

$$\theta_j^m(t+1) = \theta_j^m(t) + \eta \delta_j^m + \partial [\theta_j^m(t) - \theta_j^m(t-1)] \quad (7)$$

式中: t 为迭代次数; η 为学习速率 [$\eta \in (0,1)$]; ∂ 为动量因子 [$\partial \in (0,1)$]. η 较大时,收敛快,但不稳定; η 较小时,收敛慢; ∂ 正好与之相反。

(7) 返回至步骤(2),转入下一个图形,重复步骤(2)~(7),直至网络全局误差:

$$E = \sum_k \sum_j (T_j^k - y_j^m)^2 / 2 \quad (8)$$

达到预设的精度 E_0 为止,即 $E \leq E_0$ 。

神经网络训练好后,相应的网络结构和参数设置(权值和阈值)就确定,此时学习过程将自动结束,构建好的 BP 模型即可进行预报分析。

3.2 网络学习样本及预测数据准备

通过对影响南京市水资源需求量变化的影响因素进行主成分分析后,选取人口、GDP、万元 GDP 用水量、人均水资源量、污水年排放量这 5 项指标,对其进行归一化处理作为南京市水资源需求量的 BP 神经网络预测模型的输入层,水资源需求量作为网络输出;隐含层的结点数根据 Kolmogorov 定律取为 9。由此得到用于南京市水资源需求量预测的 BP 网络拓扑结构为(5,9,1)。

根据 1999-2010 年的《南京市统计年鉴》、《南京市水资源公报》等资料,统计计算出上述 5 项驱动因子及各年总用水量所需的归一化后的数据(见表 1)。

表 1 南京市需水量预测的标准化数据(1999-2010)

万人,亿元, m³, 10⁶ t, 亿 m³

现状年	X_1	X_2	X_3	X_7	X_9	总用水量
2010	1.28702	1.84483	-1.33918	0.79891	-0.89475	-0.92833
2009	1.20947	1.28390	-1.33910	0.67164	-1.06316	-1.04612
2008	1.05406	0.95750	-1.16707	-0.95069	-1.00842	-0.57274
2007	0.84071	0.59903	-0.68260	-0.78922	-1.09776	1.10722
2006	0.54980	0.24414	-0.41084	-0.23579	-1.38025	1.20080
2005	0.21528	-0.01898	-0.23200	-0.07094	0.64293	0.95970
2004	-0.14177	-0.26692	0.25217	-0.50893	0.66078	0.96741
2003	-0.47453	-0.61883	0.63952	2.51667	1.05085	0.73072
2002	-0.73647	-0.81869	0.91487	0.08324	1.27945	0.14945
2001	-1.03616	-0.92427	0.68931	-1.21852	0.92822	-0.60246
2000	-1.27468	-1.01676	1.33751	-0.46047	0.66480	-0.16431
1999	-1.49272	-1.26495	1.33760	0.16410	0.21731	-1.80133

注:本文基础数据来源于《南京市统计年鉴 2000-2011 年》和《南京市水资源公报 1999-2010 年》。

3.3 网络训练过程及结果检验

将归一化后的数据(表 1)作为网络训练样本,对网络进行训练。选取学习速率 $\eta = 0.88$,动量因子 $\partial = 0.4$, $E_0 \leq 0.001$ 。为了对训练和学习过的网络进行测试,将 1999-2008 年的 5 项影响因子值与相应年份的用水量作为训练样本,将 2009 年和 2010 年的 5 项影响因子及相应年份的用水量作为测试样本,训练样本经过 10 万次的训练和学习,网络达到所需精度, $E = 0.0003$,即 $E \leq E_0$,因此网络收敛。网络学习结果与误差列于表 2。从表 2 可得,BP 神经网络预测平均精度高达 99%,最低精度为 97.5%,平均误差为 0.19 亿 m³,因此,训练过程较好。此外,2009 年和 2010 年测试样本的预测结果表明,预测平均精度为 99.8%,预测结果良好。1999-2010 年南京市逐年用水量的预测结果与实际值的变化趋势(图 1)。

表 2 南京市水资源需求量的神经网络模型

		预测检验及预测结果		亿 m ³ , %
训练样本	年份	水资源总需求量		
		实际值	学习结果	相对误差
	1999	33.91	34.78407	2.57760
	2000	48.78	48.66294	-0.23997
	2001	44.80	44.85261	0.11744
	2002	51.63	51.62034	-0.01872
	2003	56.91	56.89340	-0.02917
	2004	59.06	59.06429	0.00725
	2005	58.99	59.00127	0.01909
	2006	61.18	61.17464	-0.00877
测试样本	2007	60.33	60.32889	-0.00185
	2008	45.07	45.06111	-0.01974
	2009	40.77	40.70744	-0.15345
	2010	41.84	41.91035	0.02882

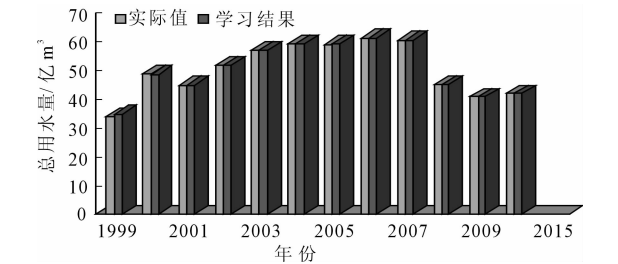


图 1 1999-2010 年南京市水资源需求量的 BP 神经网络模型预测结果与实际值的变化趋势

由于训练过的网络已“模拟”并“记忆”了输入变量和输出变量之间的“函数”关系,因此可以用它来进行对南京市水资源需求量的预报。根据《南京市国民经济与社会发展第十二个五年规划纲要》、

《南京市城市节约用水规划(2006-2020)》、《南京市城市总体规划(2007-2020)》、《南京市水资源“十二五”规划纲要》得到 2015 年南京市的数据(表 3),并将其输入到训练好的网络进行预测,其预测结果为:69.52996 亿 m^3 ,这与《南京市水资源“十二五”规划纲要》中规划的 2015 年南京市总需水量 69.55 亿 m^3 ,绝对相差为 0.02 亿 m^3 ,相对误差为 0.029%,因此,预测结果较为良好。

表 3 2015 年(规划年)南京市水资源需求量的
5 个影响因子数据 万人,亿元, m^3 , 10^6 t

影响因子	人口	GDP	万元 GDP 用水量	人均水 资源量	污水年 排放量
规划年(2015)	960	7800	80	295.6	6301

4 结 语

通过主成分分析,确定了影响南京市水资源需求量的主要因子包括人口、GDP、万元 GDP 用水量、人均水资源量、污水年排放量。其中前 3 个因子的累计贡献率达到了 90%。

由于神经网络模型具有局部逼近的特性和较强的非线性映射能力,因此它能较好地模拟具有较强非线性变化特点的水资源需求量预测问题。

通过采用主成分分析法对指标数据进行特征提取,获得主要综合指标,使得新的学习样本维数降低、分量间相关性小,提高了网络的收敛速度和学习训练效率。最后运用 BP 神经网络模型对南京市 2015 年水资源需求量进行预测,其预测结果为 69.52996 亿 m^3 。预测结果可为南京政府及水资源规划等部门进行水资源管理提供可靠依据。

BP 神经网络应用于区域需水预测,无需构建复杂的参数方程,在不进行任何简化与假定的情况下

即可进行非线性映射,模型的自学习能力强,结构简单实用。由于区域需水预测是非线性特征较强的问题,而在使用因子挑选时使用了主成分分析法,此方法是线性特征较为明显,因此在以后的预测问题中,应加强对非线性主成分分析方法的研究,以使预测结果更为准确、有效。

参考文献:

[1] Daily G C, Ehrlich P R. Socioeconomic equity, sustainability, and earth's carrying capacity [J]. Ecological Application, 1996, 6(4): 991-1001.

[2] Sag off M. Carrying capacity and ecological economics [J]. BioScience, 1995, 45(9): 610-619.

[3] 郑 度. 中国西北干旱区土地退化与生态建设问题[J]. 自然杂志, 2007, 29(1): 7-12.

[4] 朱丹果,上官智锋. 西北地区水资源可持续发展的障碍及解决策略[J]. 环境科学与管理, 2007, 32(6), 51-53.

[5] 沈福新,耿雷华,曹霞莉,等. 中国水资源长期需求展望[J]. 水科学进展, 2005, 16(4): 522-525.

[6] 吕 智,陈文贵,丁宏伟. 干旱区内陆盆地水资源的合理配置——以甘肃省高台县为例[J]. 水资源保护, 2005, 21(6): 45-48.

[7] 王 浩,游进军. 水资源合理配置研究历程与进展[J]. 水利学报, 2008, 39(10): 1168-1175.

[8] 傅 湘,纪昌明. 区域水资源承载能力综合评价——主成分分析法的应用[J]. 长江流域资源与环境, 1999, 8(2): 168-173.

[9] 魏海坤. 神经网络结构设计的理论与方法[M]. 北京:国防工业出版社, 2005: 25-36.

[10] 王 旭. 人工神经网络原理与应用[M]. 沈阳:东北大学出版社, 2002.

[11] 冯利华. 基于神经网络的洪水预报研究[J]. 自然灾害学报, 2009, 9(2): 45-48.