

GA – BP 组合预测方法在北洛河年径流量预测中的应用

王广宇¹, 解建仓¹, 张建龙²

(1. 西安理工大学 教育部西北水资源与环境生态重点实验室, 陕西 西安 710048;

2. 山西省水利建设开发中心, 山西 太原 030002)

摘要: 为寻求较为合理的径流量预测方法, 利用 GA – BP 组合模型进行径流量预测的研究, 首先利用 GA 算法 (Genetic algorithm) 全局搜索能力定位出一个较好的搜索空间, 再利用 BP 算法 (Back propagation) 神经网络的局部寻优能力, 以陕西省北洛河状头站实测数据为基础进行径流量预测。结果表明: 利用 GA – BP 组合模型预测 2001—2010 年结果的相对误差分别为 -8.9%、-12.6%、-7.4%、9.4%、-7.6%、15.0%、-5.5%、10.7%、13.3% 和 -8.2%, 预测精度较高, 从相对误差和训练次数比较可以看出, GA – BP 组合模型优于基本 BP 模型和 GA 模型。

关键词: 径流量; 预测; GA – BP 组合模型; 状头站

中图分类号: TV121⁺.2 文献标志码: A 文章编号: 1000-7601(2014)01-0203-05

Application of GA – BP combination forecasting method into the annual runoff prediction of Beiluo River

WANG Guang-yu¹, XIE Jian-cang¹, ZHANG Jian-long²

(1. Northwest Key Laboratory of Water Resource and Environment Ecology of Ministry of Education, Xi'an University of Technology, Xi'an, Shaanxi 710048, China;

2. Shanxi Province water conservancy construction and development centers, Taiyuan, Shanxi 030002, China)

Abstract: The paper has developed a GA – BP combination model to forecast the runoff, in order to seek more reasonable runoff forecasting methods. First, use the GA global search capability to locate a better search space, then use the local optimization ability of BP neural network. This forecast model was used the measured data of Zhuangtou Hydro-metric Station in Beiluo River from 1960 to 2010. The results showed that: The relative error of the predicted results from 2001 to 2010 were -8.9%、-12.6%、-7.4%、9.4%、-7.6%、15.0%、-5.5%、10.7%、13.3% and -8.2%, respectively. The predicted results of GA – BP combination model were better than the basic BP model and the GA model from the relative error and the training numbers.

Keywords: runoff; forecast; GA – BP combination model; Zhuangtou Station

河川径流量预测一直是人们关注的重大问题, 它关系到水利工程建设及水资源的开发利用。由于径流的影响因素较多、预见期长, 有的影响因子还没有发现, 同时径流量预测是一个高度的非线性系统, 其受降雨、地形地貌、气候气象及人类活动等众多因子的综合影响, 具有随机性、高维性、混沌性、模糊性等诸多复杂特征, 导致径流量预测至今仍是自然科学和技术科学领域的一项难题^[1]。目前, 研究径流

量预测的方法按技术可以分为两种, 一是根据径流与天气、下垫面等因素之间的关系建立模型, 如降雨径流模型、枯水衰退模型; 二是根据径流本身变化规律建立预测模型, 如回归模型、自回归模型、时间序列模型等^[2-4]。近年来利用新的预测理论探索预测方法得到快速发展, 如人工神经网络、混沌理论、灰色系统理论、小波变换方法等^[5-8]。虽然, 径流预测模型很多, 但尚未有一种模型是对所有的水文序列

收稿日期: 2013-04-26

基金项目: 国家自然科学基金项目 (51079120); 国家自然科学基金项目 (50979088)

作者简介: 王广宇 (1977—), 男, 陕西商洛人, 博士研究生, 经济师, 主要从事水资源系统工程研究。E-mail: zjlong007@126.com。

通信作者: 解建仓 (1963—), 男, 陕西眉县人, 教授, 博导, 研究方向为水资源系统工程与水利信息化。E-mail: jxie@xaut.edu.cn。

都适用的,径流预测在技术上还没有很成熟的预测方法。在国际上,利用组合方法在径流预测的应用较多,能够提高预测的精度^[9-10]。在国内利用组合方法进行径流预测尚不多见,刘素一等^[11]将小波变换和 BP 神经网络结合起来进行径流预测提出了两种预测方法;丁红,杨杰^[12]利用小波变换集遗传算法神经网络进行径流预测建模研究,其他组合方法则主要为加权组合的研究^[13-14]。准确预报年径流及其过程目前仍然是水文气象科学的难题,虽然可从气象系统的变化规律上进行定性预报,定量预报还要靠统计学方法。基于此,本文以智能预测理论方法为基础,利用组合预测模型进行径流预测研究,以期寻求较为合理的径流量预测方法。

1 模型的建立

遗传算法(GA 算法)与 BP 算法都是将生物学原理应用于科学研究的仿生学成果,GA 算法是从自然界生物进化机制中得到的启示,BP 算法则是对动物神经细胞或人脑的抽象模拟。GA 算法的核心思想是利用进化历史中获得的信息指导搜索和计算,该方法能够同时搜索解空间的多个点,而不至于陷入局部最优解,具有较强的鲁棒性。遗传算法流程为:(1) 群体初始化;(2) 计算个体适应值;(3) 选择;(4) 杂交;(5) 变异;(6) 结束。人工神经网络的 BP 算法属于有监督学习网络,把相关问题转化为非线性优化问题,并且是一种利用迭代运算求解权值的学习算法。网络根据学习规则不断地调整网络的权重,使网络的实际输出与样本的期望输出之间的误差降低到期望的误差精度。

基于两者都是从自然界中得到的启发,本文将两个方法结合起来,将 GA 算法的全局搜索能力与 BP 算法较好的局部寻优能力互补。分两部分对神经网络进行权值优化,首先利用 GA 全局搜索能力定位出一个较好的搜索空间,搜索出神经网络权值大致范围内的最优个体;然后利用 BP 神经网络的局部寻优能力,寻求解空间所需解决问题的最优解。具体步骤如下:

(1) 初始化

用 GA 算法来优化神经网络,一般情况下,初始化群体可任意选择,随机产生的初始群体为 $p(t) (t \in \{x_1, x_2, \Delta, x_n\})$ 。

(2) 计算适应度函数

群体中的每个个体根据其最优化任务赋予一个称为适应值的数量值,假定有一个适应度函数 $f(x_i^t)$ 评估每一个候选方案, $f(x_i^t)$ 返回时刻 t 候选个体的

适应度。每个候选个体的适应值采用候选个体的适应度与整个群体的平均适应值表示,即: $f(x_i^t)/m(P, t)$, 其中: $m(P, t)$ 是整个群体的平均适应值。

(3) 选择排序

候选个体适应值决定了它复制可能性的大小,计算完种群各个体的适应度后,对种群中的所有个体按其适应度的大小进行排序,将低适应值的个体从群体中去除,高适应值的个体将被复制,基于这个排序来分配不同个体被选中的几率,其目的是使得搜索朝搜索空间的解空间靠近。选择方式如下:

$$p_s = f(x) / \sum_{k=1}^N f(k) \tag{1}$$

可以看出,适应度越低的个体,被选中的概率也越小;反之,适应度越高,个体被选中的概率越大。

(4) 遗传算子

遗传算子包括杂交和变异。杂交是指对两个相互配对的染色体按某种方式相互交换其部分基因,然后交换分量形成两个新的候选个体,本文采用算数杂交的方法进行计算,杂交运算后产生的新个体为:

$$\begin{cases} X_A^{t+1} = aX_B^t + (1 - a)X_A^t \\ X_B^{t+1} = aX_A^t + (1 - a)X_B^t \end{cases} \tag{2}$$

式中: a 是一个常数。

在生物的遗传和自然进化过程中,由于部分个体编码中的部分信息因变异而发生变化,变异算子选择单个个体,然后任意改变它的部分特征,变异算子采用均匀变异。

假设有一个体为 $x_i = x_{i1}, x_{i2}, \cdots, x_{ik}, \cdots, x_{il}$, 若 x_{ik} 为变异点,其取值范围为 $[U_{\min}^{x_{ik}}, U_{\max}^{x_{ik}}]$, 在该变异点该点对个体 x 进行均匀变异后,可得新的个体为 $x_i = x_{i1}, x_{i2}, \cdots, x'_{ik}, \cdots, x_{il}$, 其中变异点的新基因值为:

$$x'_{ik} = U_{\min}^{x_{ik}} + r \times (U_{\max}^{x_{ik}} - U_{\min}^{x_{ik}}) \tag{3}$$

式中: $r \in [0, 1]$ 为一个符合随机概率分布的随机数。

(5) 产生新种群

将经杂交或变异新个体插入到种群 $p(t)$ 中,产生新的种群 $p(t + 1)$, 并计算新个体的适应度函数;若使误差平方和达到某一阈值 E_{ga} 时,则进入下一步,否则继续进行遗传操作。

(6) 设定 BP 神经网络的初始权值

以 GA 算法搜索出的优化初值作为 BP 神经网络的初始权值,再用 BP 算法训练网络,直到使误差

平方和达到指定精度,满足精度 E_{bp} 要求或达到预先设定的迭代次数为止,算法结束。

2 实例分析

2.1 北洛河基本概况

北洛河为渭河的第二大支流,干流水长度约为 680 km,流域面积约为 2.69 万 km²。集水面积大于 1 000 km² 的支流有周河、沮河、葫芦河。北洛河多年平均年径流量 9 亿 m³,年输沙量 1.0 亿 t。状头水文站是北洛河下游控制站,国家重要水文站。设立于 1933 年 5 月,主要任务是掌握北洛河水量、水质、沙量变化规律,为开发利用北洛河水利资源提供科学依据,为北洛河下游及黄河三门峡库区防汛提供水文情报预报。因此,本文采用北洛河状头站 1960—2010 年的实测数据为基本资料进行径流预测研究。

2.2 径流预测

(1) 以状头站 1960—2000 年实测数据作为训练样本,将 2001—2010 年的数据作为测试样本,利用式(4)对训练样本和测试样本进行归一化处理,将样

本限制在 $[-1,1]$ 。

$$x'_i = \frac{2 \times (x_i - x_{imin})}{(x_{imax} - x_{imin})} - 1 \tag{4}$$

式中, x'_i 为第 i 个数据归一化后的数值; x_i 为第 i 个数据的实际值; x_{imax} 为所有数据中的最大值; x_{imin} 为所有数据中的最小值。

(2) 初始参数设置。GA - BP 组合模型的参数参照文献[15]分别设置为:种群规模为 30,连接权数为 37,最大遗传代数为 500,采用浮点编码,杂交概率为 0.8,变异概率为 0.03,训练精度的 E_{ga} 为 0.01, E_{bp} 为 0.001。

(3) 根据适应度对个体进行归一化的优先选择、自适应交叉、变异,形成下一代种群。

(4) 将满足要求的训练样本导入网络,建立结构为 4 - 4 - 1(输入层 - 隐含层 - 输出层)的 3 层 BP 神经网络结构,根据 BP 算法的训练函数进行训练,直到使误差平方和达到指定精度 E_{bp} 要求或达到预先设定的迭代次数为止,算法结束。

训练流程见图 1,训练结果见表 1。

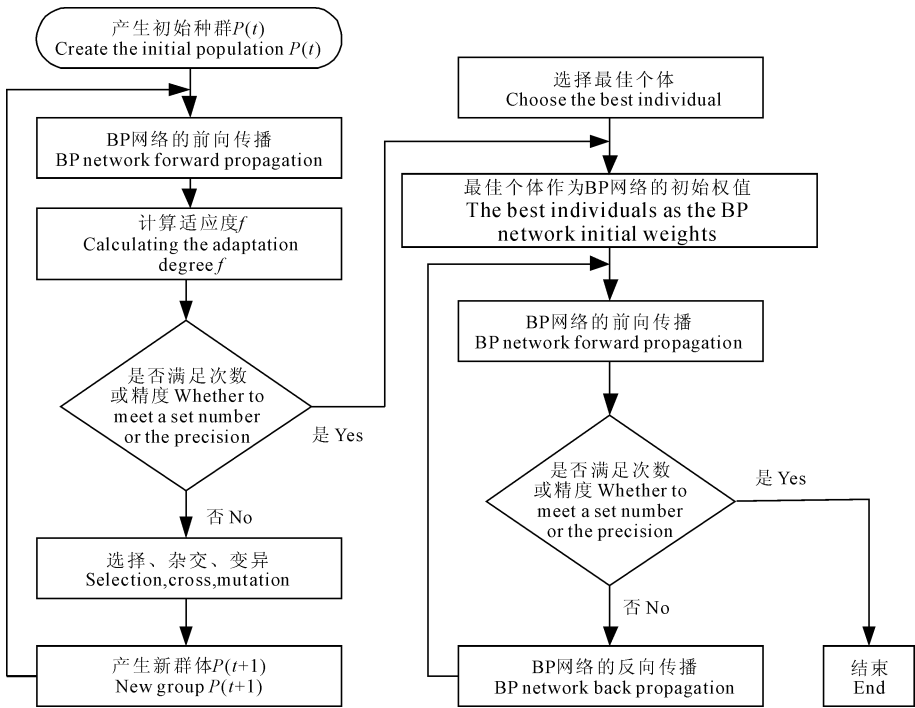


图 1 GA - BP 组合模型训练流程图

Fig.1 The training flow chart of GA - BP combination model

以训练好的 GA - BP 组合模型为基础,对 2001—2010 年的径流量进行预测,得到北洛河状头水文站的预测结果,得到结果详见表 2。可以看出,利用 GA - BP 组合模型预测的相对误差分别为

- 8.9%、- 12.6%、- 7.4%、9.4%、- 7.6%、15.0%、- 5.5%、10.7%、13.3% 和 - 8.2%,预测精度较高。

表 1 状头站的训练结果

Table 1 The training result of Zhuangtou Station

年份 Year	实测值 /(亿 m ³) Measured values /(100 million m ³)	拟合值 /(亿 m ³) Fitted value /(100 million m ³)	绝对误差 /(亿 m ³) Absolute error /(100 million m ³)	相对误差 Relative error /%	年份 Year	实测值 /(亿 m ³) Measured values /(100 million m ³)	拟合值 /(亿 m ³) Fitted value /(100 million m ³)	绝对误差 /(亿 m ³) Absolute error /(100 million m ³)	相对误差 Relative error /%
1960	0.8843	0.8957	0.0114	1.3	1981	0.5141	0.6120	0.0979	19.0
1961	0.4026	0.4277	0.0251	6.2	1982	0.7429	0.9367	0.1938	26.1
1962	0.6002	0.6316	0.0314	5.2	1983	0.0117	0.0125	0.0008	6.8
1963	0.2615	0.2994	0.0379	14.5	1984	0.2064	0.2022	-0.0042	-2.0
1964	1.0000	0.9752	-0.0248	-2.5	1985	0.1038	0.1151	0.0113	10.9
1965	0.5509	0.6638	0.1129	20.5	1986	0.8344	0.9169	0.0825	9.9
1966	0.1357	0.1580	0.0223	16.4	1987	0.8621	0.8501	-0.0120	-1.4
1967	0.3391	0.3599	0.0208	6.1	1988	0.0232	0.0259	0.0027	11.6
1968	0.2229	0.2599	0.0370	16.6	1989	0.7368	0.8530	0.1162	15.8
1969	0.3403	0.3879	0.0476	14.0	1990	0.5219	0.6157	0.0938	18.0
1970	0.6522	0.7523	0.1001	15.3	1991	0.6146	0.6007	-0.0139	-2.3
1971	0.6797	0.6279	-0.0518	-7.6	1992	0.4721	0.5353	0.0632	13.4
1972	0.931	1.0359	0.1049	11.3	1993	0.5509	0.4952	-0.0557	-10.1
1973	0.7301	0.8120	0.0819	11.2	1994	0.2161	0.2528	0.0367	17.0
1974	0.9676	1.0050	0.0374	3.9	1995	0.9164	0.9363	0.0199	2.2
1975	0.0714	0.0826	0.0112	15.7	1996	0.6239	0.6334	0.0095	1.5
1976	0.1705	0.2009	0.0304	17.8	1997	1.0000	0.9425	-0.0575	-5.8
1977	0.6671	0.6654	-0.0017	-0.3	1998	0.7063	0.7135	0.0072	1.0
1978	0.5642	0.5372	-0.0270	-4.8	1999	0.7836	0.8550	0.0714	9.1
1979	0.7929	0.8139	0.0210	2.6	2000	0.8635	0.8926	0.0291	3.4
1980	0.9422	0.9964	0.0542	5.8					

表 2 三种模型预测误差比较

Table 2 The forecast error comparison of three models

年份 Year	BP 模型 BP model			GA 模型 GA model			GA - BP 组合模型 GA - BP combination model		
	预测值 /(亿 m ³) Predictive value /(100 million m ³)	相对误差 Relative error /%	合格率 Pass rate	预测值 /(亿 m ³) Predictive value /(100 million m ³)	相对误差 Relative error /%	合格率 Pass rate	预测值 /(亿 m ³) Predictive value /(100 million m ³)	相对误差 Relative error /%	合格率 Pass rate
2001	4.4200	-8.5	100%	5.4344	12.5	100%	4.4026	-8.9	100%
2002	10.2486	10.6	100%	8.4046	-9.3	100%	8.1029	-12.6	100%
2003	9.1235	-15.3	100%	13.3136	23.6	100%	9.9692	-7.4	100%
2004	3.0456	-9.8	100%	3.6196	7.2	100%	3.6931	9.4	100%
2005	7.0431	25.2	100%	5.1473	-8.5	100%	5.1992	-7.6	100%
2006	6.8743	14.2	100%	4.9179	-18.3	100%	6.9240	15	100%
2007	4.6493	9.5	100%	4.8233	13.6	100%	4.0125	-5.5	100%
2008	8.5517	12.2	100%	8.1401	6.8	100%	8.4347	10.7	100%
2009	3.6350	-8.7	100%	3.3643	-15.5	100%	4.5096	13.3	100%
2010	5.0507	-10.6	100%	5.2202	-7.6	100%	5.1866	-8.2	100%
最大值 Max.	10.2486	25.2	100%	13.3136	23.6	100%	9.9692	13.3	100%
最小值 Min.	3.0456	-15.3	100%	3.3643	-18.3	100%	3.6931	-12.6	100%
训练次数 Train times		573			417			235	

为分析 GA-BP 组合模型的优越性,本文利用基本 BP 模型和 GA 模型对状头站的径流进行预测,并与 GA-BP 组合模型的预测结果相比较,见表 2。基本 BP 模型预测相对误差最大值和最小值分别为 25.2% 和 -15.3%,GA 模型预测分别为 23.6% 和 -18.3%,而 GA-BP 组合模型预测分别为 13.3% 和 -12.6%。从相对误差比较可以看出,GA-BP 组合模型的预测结果要优于基本 BP 模型和 GA 模型。同时根据模型训练的次数可知,GA-BP 组合模型训练次数为 235 次,基本 BP 模型和 GA 模型分别为 573 次和 417 次,可以看出 GA-BP 组合模型要比基本 BP 模型和 GA 模型的训练速度快,可节省训练时间。根据文献[16]可知,作为中长期预报,相对误差小于 35% 即可认为合格,3 种模型 10 年预测结果全部合格。GA 算法全局搜索性强,但局部搜索能力较弱,而 BP 算法的局部搜索性较强,两部分有机结合起来,可增加神经网络的学习能力,能更好地加快运算效率。可以看出,利用 GA-BP 组合模型进行径流量的预测具有一定的适用性。

3 结 语

本文以陕西省北洛河状头站 1960—2010 年实测资料为基本资料,根据资料完整情况进行相应的插补计算。通过利用 GA-BP 组合模型预测结果的相对误差分别为 -8.9%、-12.6%、-7.4%、9.4%、-7.6%、15.0%、-5.5%、10.7%、13.3% 和 -8.2%,其预测结果精度较高。经 GA-BP 组合模型、基本 BP 模型和 GA 模型预测比较可以看出,GA-BP 组合模型预测结果要优于基本 BP 模型和 GA 模型,同时训练速度快,可节省训练时间。GA-BP 组合模型将 GA 算法的全局性和 BP 算法的局部搜索性进行互补,可增加神经网络的学习能力,能更好地加快运算效率。研究成果可为水利工程建设及水资源的开发利用提供指导,利用 GA-BP 组合模型进行径流量的预测具有一定的适用性。

系统预测是一门技术性要求很强的工作,它既要求预测者具有灵活运用这些预测技术方法的能力,又要求预测者掌握多种系统预测方法与技术。

同时水资源系统的行为模式极其复杂,仅凭一种耦合理论或方法是难以准确解决预测问题,因此,今后的研究应在多种理论方法的耦合集成方面加以深入。

参 考 文 献:

- [1] 程 昶.常用预测方法及评价综述[J].四川师范大学学报(自然科学版),2002,25(1):70-73.
- [2] Zhong D H, Liu D H, Mittnik S. Time series neural network model for hydrologic forecasting[J]. Transactions of Tianjin University, 2001, 7(3):182-186.
- [3] 黄国如,芮孝芳.流域降雨径流时间序列的混沌识别及其预测研究进展[J].水科学进展,2004,15(2):255-230.
- [4] 赵显波,雷晓云,刘振平.基于 Matlab 的 BP-NN 时间序列法预测地表水资源量[J].水文,2007,27(2):34-36.
- [5] 楼 玉.混沌时间序列方法在径流预报中的应用研究[D].杭州:浙江大学,2005.
- [6] Xiang L, Yang Z B, Han S W, et al. A hybrid intelligent algorithm for portfolio selection problem with fuzzy returns[J]. Journal of Computational and Applied Mathematics, 2009, (233):264-278.
- [7] 苑希民.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.
- [8] Kima K J, Han I. Application of a hybrid genetic algorithm and neural network approach in activity-based costing[J]. Expert Systems with Applications, 2003, (8):73-77.
- [9] Jeong D I, Kim Y O. Combining single-value streamflow forecasts: A review and guidelines for selecting techniques[J]. Journal of Hydrology, 2009,377(3):284-299.
- [10] Kim Y O, Jeong D I, Ko I H. Combining rain-fall-runoff models for improving ensemble streamflow prediction[J]. Journal of Hydrology, 2006,11(6):578-588.
- [11] 刘素一,权先璋,张勇传.小波变换结合 BP 神经网络进行径流预测[J].人民长江,2003,34(7):38-39.
- [12] 丁 红,杨 杰.小波变换集遗传算法神经网络的径流预测建模[J].广西大学学报(自然科学版),2012, (3):580-587.
- [13] 殷峻暹,蒋云钟,鲁 帆.基于组合预测模型的水库径流长期预报研究[J].人民黄河,2008,30(1):28-29.
- [14] 孙惠子.中长期径流的多种组合预测方法及其比较[D].杨凌:西北农林科技大学,2012.
- [15] 雷英杰,张善文,李续武,等. MATLAB 遗传算法工具箱及应用[M].西安:电子科技大学出版社,2005.
- [16] 王红芳,王素芳,王文圣,等.基于集对原理的年径流状态时序变化特性分析[J].海河水利,2007, (5):31-33.