

河道水量还原计算中的遗传 BP 人工神经网络方法

魏茹生¹ 沈冰¹ 吴建华²

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

2. 太原理工大学 山西 太原 030024)

摘要 将遗传算法 GA 及人工神经网络 BP 算法结合起来运用于河道水量的还原计算. 结果表明: 基于 GA + BP 的河道水量还原优化算法能够克服 BP 算法自身不可优化的弊病, 较好地改善网络全局寻优能力, 提高网络速度, 防止网络陷入局部最小值. 同时该算法较确切地反映出河道水量还原计算中河道水量形成非线性这一本质现象, 提高了河道水量还原计算结果的精度.

关键词 人工神经网络; 还原计算; 遗传算法(GA); BP 模型

中图分类号 P333.1 **文献标识码** A **文章编号** 1000-198X(2008)02-0161-04

河道水量的形成是一个复杂的非线性过程, 受到降雨量、实测径流量、汛期降雨量、下垫面条件等诸多因素的影响. 由于传统的河道水量还原计算方法, 如水量平衡法、降雨径流模型法、蒸发差值法等大多是以河道水量形成机理中的某一项或几项假设成立为前提的^[1-3], 易于掌握和理解, 但是却没有反映河道水量形成过程非线性这一本质现象, 往往导致计算结果精度不理想、推广性差.

目前人工神经网络 BP(back propagation)算法是处理非线性问题的有效工具, 而河道水量的形成过程高度复杂且具有非线性的特点, 因此将 BP 算法运用于河道水量还原计算, 可以有效克服传统还原计算方法的不足. 但是, BP 人工神经网络存在易陷入局部极小值及网络收敛速度慢的缺点^[4-6]. 遗传算法 GA(genetic algorithms)是基于生物自身演变规律的优化算法, 全局寻优能力较强. 用遗传算法优化人工神经网络初始权重, 能有效提高人工神经网络 BP 的收敛速度并避免其陷入局部极小值. 遗传算法 GA 和人工神经网络 BP 算法的结合称为 GA + BP 算法. 本文基于 GA + BP 算法对河道水量还原优化算法及其应用进行研究.

1 河道水量还原计算的 GA + BP 算法

1.1 BP 人工神经网络简介

BP 人工神经网络通常是指基于误差反向传播的多层前向神经网络, 是采用反向传播的权值调整学习的算法, 最初由 Rumelhart 等在 1986 年提出^[7-8].

BP 算法的学习过程由正向传播和反向传播组成. 在正向传播过程中, 输入信息向前从输入层节点传播到隐含层节点, 经过作用函数后, 把隐含层节点的输出信息传播到输出层节点, 最后给出输出结果. 如果输出层得不到期望输出, 则转向反向传播, 将误差信号沿原连接通路返回, 通过修改各层神经元(节点)的权系数和阈值, 使得误差信号最小(达到要求精度). 传播过程中应根据网络系统的不同性质, 选择不同的网络结构和作用(特性)函数.

1.2 GA 算法概念及算法原理

GA 算法是模拟生物界遗传和进化过程而建立起来的一种搜索算法, 体现了“生存竞争、优胜劣汰、适者生存”的竞争机制.

GA 算法的基本思想是从一组随机产生的初始解即“种群”开始搜索. 种群中的每一个个体, 即问题的一个解, 称为“染色体”. GA 算法通过染色体的“适应值”来评价染色体的好坏, 适应值大的染色体被选择的几率高, 相反, 适应值小的染色体被选择的可能性小. 被选择的染色体进入下一代, 下一代中的染色体通过交叉和变异等遗传操作, 产生新的染色体, 即“后代”. 经过若干代之后, 算法收敛于最好的染色体, 该染色体很可能

就是问题的最优解或近优解. GA 算法的运行过程可表述为 (a) 编码 (b) 随机产生初始种群 (c) 以适应度函数对染色体进行评价 (d) 选择适应值高的染色体进入下一代 (e) 通过遗传、变异操作产生新的染色体; (f) 不断重复 (b)~(d) 步, 直到预定的进化代数.

1.3 GA + BP 优化算法过程及模型

GA + BP 优化算法是基于 GA 优化下的 BP 算法. 从目前有关遗传算法 GA 与人工神经网络 BP 相结合的研究来看, GA + BP 优化算法一是优化神经网络的初始权重, 二是优化网络的结构, 但无论从理论上还是方法上尚没有找到有效的指导原则来进行神经网络拓扑结构的优化. 本文以 GA 优化 BP 人工网络初始权重为研究对象, 对 GA + BP 优化算法在河道水量还原计算中的应用进行研究.

- GA 优化 BP 网络初始权重的步骤:
- a. 确定种群规模 p 、遗传算法的迭代次数 N 、种群中每个变量的维数、变量的二进制位数(采用二进制编码)、交叉概率 p_c 以及变异概率 p_m .
 - b. 随机产生初始权重

$$W_i = \{\omega_{1,i}, \omega_{2,i}, \dots, \omega_{n,i}, b_{1,i}, b_{2,i}, \dots, b_{n,i} \mid i = 1, 2, \dots, p\}$$

式中: n ——神经网络层数; $\omega_{n,i}$ ——隐含层权重; $b_{n,i}$ ——隐含层节点阈值.

- c. 目标函数定义为 $se(W^*) \leq se(W_i)_{g_{gen}} \ (i = 1, 2, \dots, p, g_{gen} = 1, 2, \dots, N)$, 即搜索进化代中使网络的误差平方和最小的网络权重.

由于进化朝着适应度大的方向变化, 所以适应度函数可以定义为目标函数的倒数形式. 为了避免函数值过小, 可以引入一个较大的系数 M

$$fi(W_i) = \frac{M}{se(W_i)}$$

- d. 进行选择、交叉及变异计算, 直至运行至满足精度要求或最大迭代次数.

GA + BP 优化算法是一包含有输入层、隐含层和输出层, 并利用 GA 优化网络权重或网络结构的多层网络, 其结构如图 1 所示. 若将该神经网络看成一个从输入 H 到输出 G 的映射, 则这个映射是一个高度的非线性映射. 如果输入节点数为 N_1 , 输出节点数为 N_3 , 则网络是从 R^{N_1} 到 R^{N_3} 的映射, 即有

$$F: R^{N_1} \xrightarrow{GA} R^{N_3} \quad G = F(H)$$

GA + BP 优化算法是将简单的非线性函数进行复合来实现上述复杂的映射.

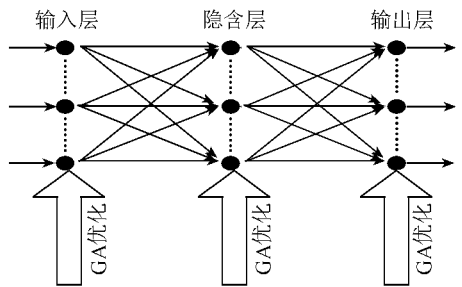


图 1 GA + BP 神经网络结构

Fig.1 The structure of GA + BP neural network

2 坪上水库 GA + BP 网络模型参数分析

山西省坪上水库流域总集水面积 11 996 km², 流域内土壤疏松, 有 26 个水文站及 1956~2000 年连续的降水和实测流量资料(见表 1, 由于篇幅有限, 仅给出 1956~1977 年数据). 以该水库流域为例, 建立 GA + BP 网络模型进行训练、仿真并对优化前后的仿真结果进行分析.

2.1 数据规格化处理

由于网络输入、输出映射函数选为 Sigmoid 函数, 需对输入数据进行归一化处理, 以防止部分神经元达到过饱和状态. 本文仅以人工神经网络工具箱中的 Premnmx 函数作为处理方法, 其公式为

表 1 坪上水库降雨量、实测径流量及天然径流量序列

Table 1 The rainfall, measured runoff and natural runoff of the Pingshang Reservoir

年份	降雨量/ mm	径流量/亿 m ³		年份	降雨量/ mm	径流量/亿 m ³	
		实测	天然			实测	天然
1956	830.2	14.5	15.7	1968	372.6	6.6	8.6
1957	479.7	6.0	7.8	1969	564.4	6.5	8.4
1960	402.8	5.4	7.7	1970	392.0	4.9	7.1
1961	508.7	5.9	8.1	1971	435.3	3.97	6.0
1962	426.7	5.6	7.5	1972	196.8	1.9	3.7
1963	524.7	4.0	5.9	1973	692.8	6.8	8.9
1964	746.5	11.2	13.0	1974	306.4	2.6	4.9
1965	260.2	3.6	5.5	1975	424.3	2.1	4.1
1966	542.3	5.8	7.8	1976	493.1	4.1	6.1
1967	658.5	14.6	16.3	1977	562.9	8.3	10.5

$$y_i = \frac{z_i - z_{\min}}{z_{\max} - z_{\min}} a + b$$

(1)

式中 a, b 可以根据实际调整,本文取 2,-1; z_i 和 y_i 分别是变换前后的变量.

2.2 输入因素的选取

降雨径流量相关分析成果表明,坪上水库天然径流量与降雨量、实测径流量关系密切.为了定量分析天然径流量与降雨量、实测径流量之间的相关程度,利用坪上水库 1956~1987 年的系列资料计算天然径流量与实测径流量、降雨量之间的相关系数.天然径流量与降雨量的相关系数为 0.620 2,为正相关,降雨量越大,天然径流量越大,二者具有较强的相关性;实测径流量与天然径流量的相关系数为 0.600 9,也为正相关.因此,为了确定坪上水库以上流域控制断面的天然径流量,只要利用坪上水库各站降雨量和实测径流量系列资料,建立天然径流量的 GA+BP 网络模型,并对其进行训练、检验,就可仿真河道水量的还原量^[9].

2.3 输出量选取

GA+BP 网络模型的输出量只有天然径流量,因此 GA+BP 网络为多输入单输出网络.

2.4 坪上水库 GA+BP 网络模型建立及参数设定

由本文分析可以看出,GA+BP 网络模型的输入为降雨量和实测径流量,输出为还原量.在 GA 优化 BP 神经网络初始权重的基础上,考虑到坪上水库的实际情况,确定网络的隐含层节点为 6 个,据此建立坪上水库 GA+BP 网络模型结构如图 2 所示,坪上水库 GA+BP 网络模型参数设定见表 2.

本文以坪上水库 1956~1987 年资料进行训练,1988~2000 年资料进行检验.通过训练、仿真得出该水库基于 GA+BP 网络模型算法的训练性能曲线(图 3)及仿真结果(表 3),表 3 中相对误差平均值为 7.04%.

表 2 坪上水库 GA+BP 网络模型参数

Table 2 The parameter of the GA+BP model for Pingshang Reservoir

参数名	动量项 系数/m	学习率 η	目标 误差限	选择 概率	交叉率	变异率	最大 进化代	代沟
取值	0.85	0.05	10^{-3}	0.05	0.7	0.05	1000	0.9

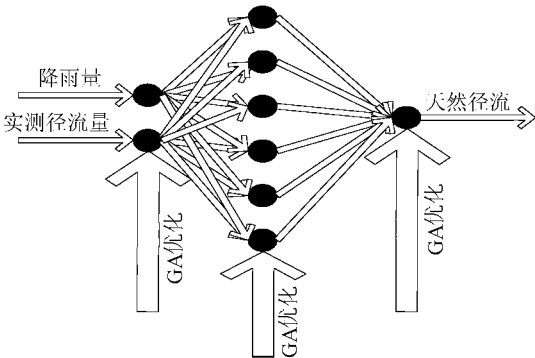


图 2 坪上水库 GA+BP 网络模型
Fig.2 The GA+BP model for
Pingshang Reservoir

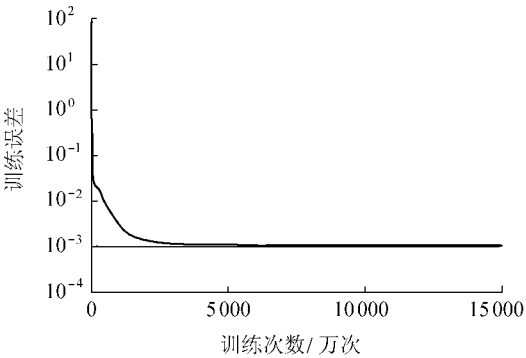


图 3 坪上水库基于 GA+BP 网络模型算法
的训练性能曲线
Fig.3 The training process of GA+BP model

表 3 GA+BP 网络模型仿真结果及误差

Table 3 The result and error of GA+BP model

年份	径流量/ 亿 m ³	径流量仿真值/ 亿 m ³	相对 误差/%	年份	径流量/ 亿 m ³	径流量仿真值/ 亿 m ³	相对 误差/%
1988	0.583 0	0.463 6	20.48	1995	0.252 9	0.228 2	9.77
1989	0.675 7	0.664 9	1.60	1996	1.000 0	0.962 2	3.78
1990	0.556 0	0.552 4	0.65	1997	0.539 0	0.669 0	-24.12
1991	0.783 8	0.750 6	4.23	1998	0.685 3	0.683 7	0.23
1992	0.870 7	0.810 0	6.97	1999	1.000 0	0.937 8	6.22
1993	0.982 6	0.925 3	5.83	2000	0.812 7	0.783 4	3.61
1994	0.440 2	0.457 8	-4.00				

为与其他河道水量还原计算方法的结果进行比较,表 4 给出未经 GA 优化的 BP 神经网络模型的计算结果^[10],其相对误差平均值为 7.59%.

表 4 BP 神经网络仿真结果及误差

Table 4 The result and error of BP neural network

年份	径流量/ 亿 m ³	径流量仿真值/ 亿 m ³	相对 误差/%	年份	径流量/ 亿 m ³	径流量仿真值/ 亿 m ³	相对 误差/%
1988	0.583 0	0.483 9	17.00	1995	0.252 9	0.170 0	32.78
1989	0.675 7	0.678 7	0.44	1996	1.000 0	0.893 0	10.70
1990	0.556 0	0.564 6	1.55	1997	0.539 0	0.491 0	8.91
1991	0.783 8	0.757 2	3.40	1998	0.685 3	0.695 5	1.49
1992	0.870 7	0.798 6	8.28	1999	1.000 0	0.921 6	7.84
1993	0.982 6	0.972 4	1.04	2000	0.812 7	0.773 0	4.89
1994	0.440 2	0.441 7	0.34				

由表 3 A 可见 基于 GA 优化的 BP 人工神经网络模型的计算结果优于未经优化的 BP 人工神经网络模型。

3 结 语

传统河道水量还原计算方法原理简单 ,但所涉及的社会领域较广泛 ,资料收集较难 .水文现象是高度非线性的 ,而传统河道水量还原计算方法的降雨径流模型仅对与径流相关的影响因子(降雨量、汛期降雨量等)进行了线性处理 ,违背了水文现象的实际规律。

本文以坪上水库为例 ,将 GA + BP 优化算法用于河道水量的还原计算 .实践证明 (a)GA 与 BP 人工神经网络的结合提高了河道水量还原计算的精度 ,较为合理地反映了河道降雨径流复杂的运行机制 (b)GA + BP 优化算法的全局误差小 ,网络不易陷入局部最小值 ,使相对较短期内的“预报”结果更接近实测值。

需要指出的是 ,虽然目前已有一些学者提出了人工神经网络拓扑结构及隐含层神经元个数的确定方法 ,但大多缺乏令人信服的理论依据 ,本文仅凭经验及试算的方法进行确定 ,导致所选拓扑结构不一定是最优网络结构。

参考文献：

[1] 陆中央.关于年径流量系列的还原计算问题 [J].水文 ,2000 ,20(6):15-19.

[2] 黑龙江省水文总站.区域水资源分析计算方法 [M].北京 :水利电力出版社 ,1987 :150-155.

[3] 沈宏.天然径流还原计算方法初步探讨 [J].水利规划与设计 ,2003(3):15-18.

[4] 王崇骏.一种基于遗传算法的 BP 神经网络算法及其应用 [J].南京大学学报 :自然科学版 ,2003 ,39(5):459-466.

[5] 翟宜峰 ,李鸿雁 ,刘寒冰 ,等 .基于遗传算法优化神经网络初始权重的方法 [J].吉林大学学报 :工学版 ,2003 ,33(2):45-50.

[6] 张建勋 ,洪沙 .基于遗传算法 BP 神经网络的零件图特征识别 [J].重庆建筑大学学报 ,2005 ,27(3):93-97.

[7] 苑希民 ,李鸿雁 .神经网络和遗传算法在水科学领域的应用 [M].北京 :水利水电出版社 ,2002 :83-97.

[8] WHIGHAM P A ,CRAPPER P F .Modelling rainfall-runoff using genetic programming [J].Mathematical and Computer Modelling ,2001 ,35(6):707-721.

[9] 梁静国 ,林朗星 .规则推理和遗传神经网络综合的事故分析方法 [J].学术之窗 ,2006(8):198-199.

[10] 宋承新 .径流还原计算的综合修正法 [J].水文 ,1999 ,19(2):20-22.

River water returning calculation based on genetic BP neural network

WEI Ru-sheng¹ , SHEN Bing¹ , WU Jian-hua²

(1. Northwest Key Laboratory of Water Resource and Environment Ecology , Ministry of Education , Xi 'an University of Technology , Xi 'an 710048 , China ;
2. Taiyuan University of Technology , Taiyuan 030024 , China)

Abstract :Returning calculation of river water was carried out on the basis of GA (genetic algorithm) and BP (back propagation) algorithm of artificial neural networks . Results indicate that the optimization algorithm based on GA + BP , which does not have the defect that BP algorithm cannot be optimized , can improve the global optimization ability of the net , increase the net speed , and avoid the local minimum value in the net . The algorithm reflects the nonlinear character of water formation in river water returning calculation , and improves the calculation precision .

Key words :artificial neural network ; returning calculation ; genetic algorithm (GA) ; back propagation (BP) algorithm