

文章编号:0559-9342(2005)09-0012-04

人工智能在洪水预报中的应用

王春平¹,王金生¹,梁团豪²

(1.北京师范大学水科学研究院水沙科学教育部重点实验室,北京 100875;

2.水利部松辽水利委员会,吉林 长春 130021)

关键词:神经网络;遗传算法;权重;洪水预报

摘要:洪水预报对防洪减灾,拦洪储存水资源等具有重要意义。对用遗传算法(GA)来改进神经网络的算法进行了分析,建立了用GA动态寻求权重的前馈网络模型(I)和用GA优化初始权重的前馈网络模型(简称模型II)。借助于Matlab中的神经网络包和遗传算法包,编制了改进模型的计算机程序,对湖北省黄龙滩水库14场洪峰和洪水总量进行了模拟,用另外5场洪水检验了改进的模型,并与传统的前馈神经网络模型进行了比较。结果表明:用模型I对洪峰流量的检验效果明显优于模型II和传统的前馈神经网络模型;模型II对中低流量及洪水总量的检验效果相对较好。因此,本次改进的模型可以用于洪水预报。

Application of artificial intelligence in the flood forecasting

Wang Chun-ping¹, Wang Jin-sheng¹, Liang Tuan-hao²

(1. Institute of Water Sciences, Beijing Normal University, Beijing 100875;

2. Songhua-Liaoh River Water Resources Commission, Changchun Jilin 130021)

Key Words: artificial neural network; genetic algorithm; weight; flood forecasting

Abstract: The paper analyzes all kinds of methods of neural network improved by genetic Algorithm, and proposes two ANN models, among which model I has the best weights dynamically searched by GA and model II uses GA to optimize the initial weights. All models are achieved by using matlab language. Moreover, it takes the Huanglongtan reservoir as an example to verify the result of the traditional ANN model as well as the other two proposed. The result shows that Model I works best in forecasting the peak-flood amount, while Model II is best in forecasting middle, low and the total flood amount. In the end, it is suggested that the above two models be coupled together to achieve a high precision in flood forecasting by making use of the relative advantages of each other.

中图分类号:P338.9

文献标识码:A

神经网络模型(Artificial Neural Networks, 缩写 ANN)是一种模拟人脑及其活动的数学模型。它具有大规模的并行计算与分布式存贮能力、非线性映射能力、自适应、自组织、自学习的能力,有较强的鲁棒性和容错性,能够求解非凸性问题等优点。

遗传算法(Genetic Algorithm, 缩写 GA)是受自然界生物遗传中的“物竞天择,适者生存”的启发而发展起来的一种优化方法。它以数位码的方式,在有限合理条件下,以适当的评分标准来定出优胜劣汰的依据,挑选适应力较好的成员进行复制、交配、突变,使最能符合目标函数的解繁衍至下一代,较佳的解得以保留,较差的解自然淘汰。经过若干代遗传后,获得最佳解。

神经网络自身存在以下缺点:隐层层数、隐层神经元节点个数难以确定,算法的收敛速度慢和网络易陷入局部极值点。对于网络拓扑结构的确定,目前尚无一套完善的理论。根据Kolmogorov定理,一个三层的BP网络,可在任意平方误差精

度 ε 内逼近目标函数 $f(X)$ 。对于隐层节点的选择可参考 $2N+1$ ^[1],其中 N 为输入因子的维数。随着电子信息技术的发展,算法收敛速度慢已经可以克服,即使训练上千万次,也仅需三五分钟。而对于网络易陷入局部极小点,目前已有多种算法。如,改进BP传统算法^[2]、将小波分解融入神经网络^[3]及将基础水文知识融入神经网络的方法^[4]都取得了良好效果。同时,应用遗传算法来改进神经网络的研究也较多,大致分为以下三种改进方法:①用遗传算法进化神经网络结构^[5,6];②用遗传算法优化神经网络的初始权重^[7,8];③用遗传算法代替神经网络传统算法训练网络^[9]。本文正是基于后两种思路,对ANN模型进行改进研究,并通过实际的洪水预报效果验证其改进效果。本文借助于matlab平台和其自带的神经网络包和遗传算法包,

收稿日期:2005-05-08

作者简介:王春平(1978—),女,陕西渭南人,硕士研究生,研究方向为水文水资源、环境模拟与污染控制。

减少了编程的工作量,并实现了算法的改进。

1 算法原理和结构

1.1 BP 算法原理及结构

BP 算法是最重要、应用最广泛的一种 ANN 算法。实践证明这种基于误差反传递算法的 BP 网络有很强的映射能力,可以解决许多实际问题。BP 网络由一个输入层、一个(或多个)隐含层和一个输出层组成。输入层用来接收信号,信息由此进入网络。就洪水预报而言,输入信息可以是降雨、前期径流、气象因子及水文序列等。隐含层是用来实现非线性问题求解的。隐层神经元由两部分组成:①加法器,对前一层输入信号的加权求和;②转换器,由各种激活函数构成,一般用 Sigmoid 函数进行转换,用以生成向前传播的各种信号。Sigmoid 函数的表达式为 $f(X)=1/1+e^{-x}$ 。输出层用来系统输出功能信号。输出信息可以是洪峰流量、水位和枯季径流量等预报结果。该层的转移函数一般为线性函数,其数学表达式如下: $f_i(X)=X$ 。每层由若干个神经元组成,层与层之间的节点通过权重值、偏差值相联系。在这种网络中,功能信号一层一层地向前传播、推进,最终形成输出信号。输出信号与目标信号对比,形成神经网络的执行函数——误差函数,其表达式如下:

$$E_n = \frac{1}{n} \sum_{i=1}^n (y_i - t_i)^2 \quad (1)$$

式中, y_i 为计算值,即网络的输出值; t_i 为实测值,即网络的目标值; n 为样本数量。

传统 BP 算法的主要运算过程如下:①确定网络参数;②用 $initnw$ 规则产生一组初始权重;③用样本信息计算网络误差;④检验精度是否符合要求,若满足则此组权重即是理想的权重,转⑥,若不满足,则转⑤;⑤用 LMBP 算法修改权重,转③;⑥根据建立的网络结构和选出的权重进行预报。

1.2 遗传算法原理及结构

遗传算法是模拟生物界的遗传和进化过程而建立起来的一种搜索算法,其基本思想是从一组随机产生的初始解(即“种群”)开始进行搜索。种群中的每一个体(即问题的一个解)称为“染色体”。遗传算法通过染色体的“适应度”来评价染色体的好坏,适应度大的染色体被选择的几率高;相反,适应度小的染色体被选择的可能性小。被选中的染色体进入下一代,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,该染色体就是问题的最优解或近似最优解。

遗传算法的运行过程可用如下步骤来表述:①随机产生初始种群;②以适应度函数对染色体进行评价;③选择高适应度的染色体进入下一代;④通过遗传、变异操作产生新的染色体;⑤不断重复第②~④步,直至达到预定的进化效果或代数。

2 用遗传算法对 BP 算法的两种改进模型

2.1 模型 I——用 GA 动态寻权的 ANN 模型

传统的 BP 算法采用 Levenberg-Marquardt 算法,该算法是最速下降法和高斯-牛顿法的折中,迭代的基础是雅克比矩阵的计算。雅克比矩阵的计算需要较多的内存,并且本身易

陷入局部极小点。为克服这些缺点,本文用遗传算法代替 BP 神经网络传统的 Leverage-Marquardt 算法来动态地寻求最佳的权重,以发挥遗传算法的全局搜索的优点。本方法的计算流程见图 1,图 1 中左边从第二框到最下面框所示的是用 GA 动态寻权的过程,替代了 BP 算法过程中的③~⑤步。

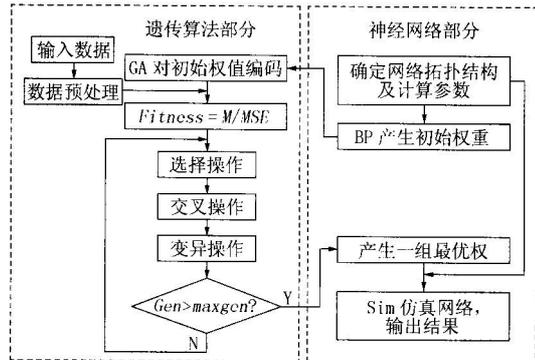


图 1 模型 I 的计算流程

2.2 模型 II——基于 GA 初权的 ANN 模型

采用遗传算法优化网络初始权重,实际上是将整个网络训练分成了 2 部分:首先采用遗传算法获得一组较优的初始权重;然后利用 BP 算法最终完成网络训练(见 BP 算法过程③~⑤步)。其流程图如图 2 所示。

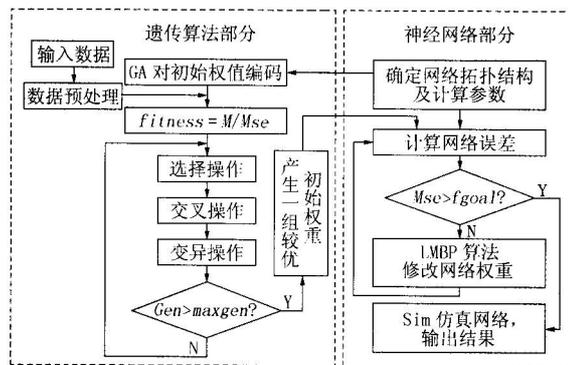


图 2 模型 II 的计算流程

模型 II 与模型 I 的区别在于,前者是利用 GA 算法获得一组较优的初始权重后,用 BP 算法和样本信息进行网络权重的训练学习,得到最优的权重后,再用于预测;而模型 I 在 BP 算法产生的一组初始权重基础上,应用 GA 动态地应用样本信息反复寻求最后权重,然后直接用于预报,不再用 BP 算法进行训练学习。

2.3 遗传算法改进 BP 网络的步骤

遗传算法的关键在于染色体的表达方式,适应度函数的定义及遗传操作的规则和标准。本文结合神经网络模型的遗传算子主要设计如下:

2.3.1 染色体的编码

搜索空间的解与遗传空间的解是一一对应的,在算法中需要通过编码将搜索空间的解映射成遗传空间的解。经典的遗传算法编码是用二进制编码,许多改进算法也是在二进制

基础上改进的;但是此方法最大的缺陷就是用户对遗传操作过程中产生的个体不会有直观的感受,必须等到遗传操作完毕进行解码后才知道最终结果。这些使得遗传操作运行起来比较抽象。本文采用实数编码,染色体在解空间内以浮点向量的形式直接生成,表达方式自然。染色体即为问题变量,遗传空间即为问题空间,免去了来回编码解码的麻烦,使得遗传操作更简化、直观。染色体的表示方式是行向量[输入层权重,隐层阈值,隐层与输出层权重,输出层阈值],即用 $X=[I_w^T, B_l^T, L_w, B_o]$ 来实现染色体的表达。为防止网络过饱和而无法训练,网络初始权重范围限定在 $(-1,1)$ 内^[10]。

2.3.2 适应度函数的表达式

适应度是解的目标函数值,用以估计个体的好坏。它也是联系遗传算法与原有优化问题的纽带。模型 I 和模型 II 的适应度函数都采用以下函数:

$$fitness = \frac{M}{\sum_{i=1}^N \left(\frac{L_w}{1+e^{-\frac{L_w}{P_i - T_i}}} + B_o - T_i \right)^2} \quad (2)$$

式中, $fitness$ 为适应度函数; M 为赋值(100),避免进化当中函数值变化不敏感; I_w, B_l, L_w, B_o 依次为网络的输入权重、阈值、及隐层与输出层权重; P_i, T_i 为网络的输入值和目标值,即实测的样本数据。

2.3.3 遗传操作

(1)选择(selection)操作。选择操作是根据个体的适应度,在种群中按一定概率 P_s 选取可作为父本的个体。适应度大的个体被选中的概率大。选择操作有轮盘选择和基于几何分布规律的选择。在标准遗传算法中采用的是基于适应度比例的选择策略,常常出现过早收敛现象和停滞现象。避免这些现象的方法之一就是使用基于适应度排序的选择策略。这样个体的适应度不直接影响后代的数量。基于几何分布规律的选择法,是以最佳染色体的选择概率 P_s 作为基本参数,按染色体的排列序号确定其选择概率。本文采用基于几何分布规律的选择方法,通过 `normGeomSelect` 函数来实现,选择概率 P_s 取 0.8。

(2)交叉(crossover)操作。交叉操作是按一定概率选取个体对,然后对它们按一定概率 P_c 随机地交换基因以形成新的个体。在 matlab 的神经网络包 `nnet` 中,对实数编码的操作有 `arithXover` `heuristicXover` `simpleXover` 函数,本文尝试对染色体进行以上 4 项杂交操作,以凸现出各个函数的优势,尤其要进行 3 次启发式操作,以防止网络陷入局部极小,无法求得全局最优解。其中 P_c 取 0.4。

(3)变异(mutation)操作。变异操作是按一定概率 P_m 随机地改变某个个体的基因值。同样地, `gaot` 包里对变异操作具有以下 `boundaryMutation` `multiNonUnifMutation` `nonUnifMutation` 和 `unifMutation` 4 个函数。本文采用它们进行变异操作,特别突出非均匀变异操作 `nonUnifMutation`,以使得染色体的变异量非均匀变化,保持个体的多样性。在进化初期,变异概率 P_m 取较小值(0.005),迭代后期采用较大值(0.01),以预防过早收敛。

3 模型在径流预报中的应用实例

黄龙滩水电站位于鄂西北。竹山水文站是黄龙滩水库上游河流干流的控制站。该站以上流域面积 7 051 km²,干流全

长 228 km。该集水流域内共有两大支流,它们在竹山水文站上游 45.6 km 处汇入干流。流域内共布设有 26 个雨量站和 6 个水文站,两大支流各有一个控制水文站分别是汇湾站和老码头站。本文就是根据这两个站的实测径流量预报竹山站的流量,为黄龙滩水电站的洪水调度赢得时间。

我们分析了该流域的特性及黄龙滩水电站洪水的实际推进过程,得知上游站与下游站水流到达时间大致相差 3 h。选取时间步长 $\Delta t=3$ h,因此上下游滞后为一个时段。竹山站的流量除了包括汇湾、老码头站的来水外,还包括两站至竹山站区间各支流的来水。因此本文除了采用汇湾和老码头站第 t 时段流量值外,还考虑竹山站第 t 时段流量值,以提高竹山站 $t+1$ 时段的流量预报精度。网络具有 3 个输入,一个隐层(取 7 个神经元),一个输出的拓扑结构。选取 1996~1999 年汛期共 19 场洪水,14 场用于对网络进行训练,5 场用于预报检验。本文每场洪水的持续时间是和当时的各场洪水过程完全一致的。值得说明的是,模拟样本场次洪水的顺序并不是按时间先后选取的,而是随机组合的,以避免模型泛化能力差的缺陷。

因隐层作用函数 sigmoid 的敏感区在 $[-2,2]$ 之间,为了防止网络过早饱和、消除不同级别数据对网络精度的影响,保证所有的变量在训练过程中受到同等重视,在把样本数据输入模型之前,首先需要对数据进行预处理。本文采用 matlab 自带的 `premnmx` 函数进行预处理。由于神经网络的初始权重是用 `initnw` 函数随机产生的一组较有规律的随机数,因此每次训练结果不尽相同,但在多次训练后还是趋于稳定的。

图 3~5 是用训练后的三种模型对同一场洪水进行预报的结果。该场洪水起始时间为 1996 年 11 月 6 日,持续时间为 106 h(约 4 天半),过程如图 3~5 中的实线所示。从图 3~5 可以看出,3 种模型的预报效果基本上都还可以。模型 I 对洪峰的预报结果高于其他两个模型,而对中低流量的预报效果不十分理想,对洪量的预报精度稍差。而模型 II 正好相反,对中低流量的预报效果明显好于其他两个模型,但是对洪峰的预报精度稍逊一筹。3 个模型关于该场洪水洪峰的预报效果见表 1。BP 模型、模型 I、模型 II 的洪峰预报结果分别是 1 655.2、1 619.4 m³/s 和 1 627.8 m³/s,实际测得的洪峰值是 1 613.2 m³/s。模型 I 的预报效果最好,相对误差仅 0.4%。3 个模型对该场洪水量的预报效果见表 2,其中模型 II 最好,预报误差 0.2%。

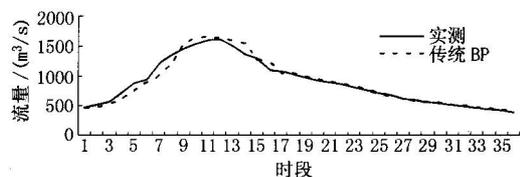


图 3 传统 BP 模型预报 19961106 号的洪水过程

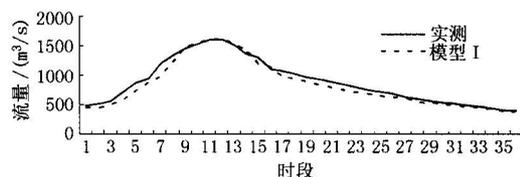


图 4 模型 I 预报 19961106 号的洪水过程

表 1 3 个模型预报洪峰结果

洪水编号	实际峰值 / (m^3/s)	预报峰值/ (m^3/s)			相对误差/%		
		BP 模型	模型 I	模型 II	BP 模型	模型 I	模型 II
19961106	1 613.2	1 665.2	1 619.4*	1 627.8	3.2	0.4	0.9
19970814	1 865.4	2 090.0	1 922.1*	1 927.8	12.0	3.0	3.3
19980411	917.4	888.1	878.4	901.2	-3.2	-4.3	1.8
19980806	1 671.2	1 544.3	1 768.2*	1 481.3	-7.6	5.8	-11.4
19990706	550.7	531.7	536.0*	496.4	-3.5	-2.7	-9.9

注:表中加 * 号的数字是与实测值最接近的预报结果。

表 2 3 个模型的洪量预报效果

洪水编号	持续时间 /h	实际洪量 / 万 m^3	预报洪量/ 万 m^3			相对误差/%		
			BP 模型	模型 I	模型 II	BP 模型	模型 I	模型 II
19961106	105	34 604.6	34 740.7	32 218.7	34 690.6*	0.4	-6.9	0.2
19970814	87	33 554.7	34 208.0	31 593.5	33 527.8*	1.9	-5.8	-0.1
19980411	84	16 292.2	17 522.1	15 816.2*	16 811.0	7.5	-2.9	3.2
19980806	75	22 041.7	22 760.9	22 667.0	22 638.5*	3.3	2.8	2.7
19990706	108	9 586.0	8 630.3	10 476.9	9 780.0*	-10.0	9.3	2.0

注:表中加 * 号的数字是与实测值最接近的预报结果。

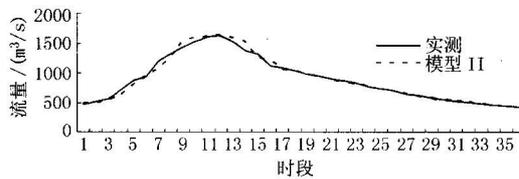


图 5 模型 II 预报 19961106 号的洪水过程

即使是同一水电站、水库或水文站,各场洪水的洪峰和洪量大小、持续时间长短、洪水流量过程线的形状等特征都是不相同的。用同一模型预报不同场次的洪水的精度也是不一样的。为了检验 3 种模型对更多洪水的预报效果,本文还对另外 4 场洪水进行了预报操作,结果也汇总于表 1,2,表中带 * 号的数字是最好的预报结果。

由表 1 可以看出,在预报的 5 场洪水中,模型 I 有 4 场的预报峰值与实测值较接近,尽管在 19980411 号洪水中预报的峰值和实测值误差偏大(同其他模型相比),但是其误差仅为 4.3%。模型 I 对 5 场洪水的洪峰预报的平均误差(绝对值平均)为 3.24%,小于传统 BP 模型(5.90%)和模型 II(5.46%)。由表 2 可以看出,模型 II 在 5 场洪量预报中,有 4 场和实测值很接近,仅 19980411 号洪水的预报效果比模型 I 差,但是其相对误差也仅为 3.2%。模型 II 对 5 场洪水的洪量预报的平均误差为 1.64%,明显小于传统 BP 模型(4.62%)和模型 I(5.54%)。

4 结 语

本文在分析了人工神经网络和遗传算法的优点和不足的基础上,将二者结合起来,力图用遗传算法的优点弥补人工神经网络的不足。基于此思路,本文在 ANN 的 BP 网络基础上进行了改进,建立了用 GA 动态寻权的 BP 模型和基于 GA 初权的 BP 模型(即模型 I 和模型 II)。经过黄龙滩水电站多场洪水的实际预报应用效果的对比分析,结果表明:①模型 I 在洪峰预报精度方面比 BP 模型有较明显的提高,其洪量的预报

精度与 BP 模型相当;②模型 II 在洪量预报精度方面比 BP 模型有较明显的提高,其洪峰的预报精度与 BP 模型相当。如果在实际洪水预报中,将两种模型耦合起来,充分发挥各自的优势,预报效果将会更好。

洪水预报的预报精度和预见期都很重要,由于受资料、时间、精力等方面的限制,本文只是利用黄龙滩水电站上游的 3 个主要水文站的流量资料建立了洪水预报模型。如果能够利用更多站点的资料,可能会得到更高的预报精度。如果再进一步利用降雨径流资料或者气象资料,则不仅对提高同样预见期的预报精度有帮助,而且还能够加长预见期。这些都有待于今后作进一步探讨。

参考文献:

- [1] 苑希民,李鸿雁,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] 赵全升,杨天行,等.改进的 BP 算法在黄河下游枯季径流预报中的应用[J].安全与环境学报,2001,1(3):30-35.
- [3] 王文圣,熊华康,等.日流量预报的小波网络模型初探[J].水科学进展,2004,15(3):382-386.
- [4] 杨荣富,丁晶,等.具有水文基础的人工神经网络初探[J].水利学报,1998,(8):23-37.
- [5] 李智勇,童调生,等.基于多物种进化遗传算法的神经网络结构进化设计方法[J].系统仿真学报,2003,15(6):810-813.
- [6] 刘芳,李人厚等.基于遗传算法的进化神经网络[J].系统仿真学报,2003,15(10):1431-1433.
- [7] 苑希民,李鸿雁等.人工神经网络与遗传算法在河道洪水预报中的应用[J].水利发展研究,2002,2(12):50-58.
- [8] 翟宜峰,李鸿雁,等.人工神经网络与遗传算法在多泥沙洪水预报中的应用[J].水利学报,2003,(2):8-13.
- [9] 阎俊爱,钟登华,等.基于遗传算法的神经网络优化预报模型及其在年径流预报中的应用[J].水利水电技术,2003,34(6):1-4.
- [10] 周翔,等.基于遗传学习算法和 BP 算法的神经网络出口在矿坑涌水量计算中的应用[J].水利学报,2000,(12):59-63.