

区域工业用水量非线性预测模型的优选

王晓玲¹, 孙月峰¹, 梅传书², 李松敏¹

(1. 天津大学环境科学与工程学院, 天津 300072; 2. 水利部海河水利委员会, 天津 300170)

摘要: 针对BP神经网络在确定输入因子时的任意性, 将相关性分析引入BP神经网络输入因子的选取中, 通过计算输入因子和输出因子之间的相关系数, 并根据相关程度来确定输入因子, 同时利用遗传算法优化BP神经网络的权值和阈值的方法, 建立了改进BP神经网络模型. 将该模型应用于永定河山区工业用水量预测中, 通过和传统非线性回归法进行比较, 结果表明改进BP神经网络拟合和预测精度均较高. 改进BP神经网络法的平均相对误差达到1.41%, 预测的2010年和2030年工业用水量将分别达到 $3.63 \times 10^8 \text{ m}^3$ 和 $4.46 \times 10^8 \text{ m}^3$. 预测结果可为水资源规划和管理提供依据.

关键词: 工业用水量; 预测; 改进BP神经网络; 非线性; 优选模型

中图分类号: TU991.31

文献标志码: A

文章编号: 0493-2137(2006)12-1399-06

Preferred Nonlinear Model for Predicting Industrial Water Demand of District

WANG Xiao-ling¹, SUN Yue-feng¹, MEI Chuan-shu², LI Song-min¹

(1. School of Environment Science and Engineering, Tianjin University, Tianjin 300072, China;

2. Haihe River Water Conservancy Commission, Ministry of Water Resources, Tianjin 300170, China)

Abstract: The improved BP neural network model was built, in which correlation analysis was adopted to choose input gene. Based on the correlation coefficient of input and output gene, the input gene was chosen. The genetic algorithm was used to optimize the weights and thresholds of BP neural network. The model was applied to predicting the industrial water demand of the mountainous area along Yongding River, and the predicted values were compared with the values by the traditional non-linear regress methods. The results indicated that both the fitting and prediction precision of the improved BP neural network method are better than that of the traditional non-linear regress methods, and the improved BP neural network method is feasible and effective for application. The mean relative error of the improved BP neural network method is 1.41%, the industrial water consumption is $3.63 \times 10^8 \text{ m}^3$ in year 2010 and $4.46 \times 10^8 \text{ m}^3$ in year 2030. The results are available in water resource planning and management.

Keywords: industrial water demand; prediction; improved BP neural network; nonlinear; preferred model

工业用水量由于受区域发展规划、产业结构、价格机制、政策制度和技术经济发展水平等因素的制约, 是变化很复杂的非线性变量, 因而很难对其进行精确预测, 尤其是对统计信息贫乏、发展尚不稳定的区域. 传统工业用水量预测方法^[1-2]主要有趋势法、线性回归法和万元产值需水定额法等. 目前国内以线性拟合外

推为主, 有采用以时间为变量的线性拟合外推, 也有采用以时间、工业产值和万元产值用水等为变量的多元拟合外推^[3]. 国外工业用水量预测技术以智能算法为主, 如Liu等人^[4]用人工神经网络 (artificial neural network, ANN) 预测城市用水量. Lertpalangsunti等人^[5]编制了基于智能算法的预测系统, 预测中用到ANN、模

糊理论(fuzzy logic, FL)和 CBR(case-based reasoning)等多种智能算法. Liu 等人^[6]用 FL、ANN 和自回归(auto-regression, AR)模型做短期负荷预测,并对这 3 种模型进行了比较.

非线性回归法实际上是根据历史资料分析未来发展趋势,进而推测预测期用水量的一种方法.该方法简单快捷,易于掌握. BP 神经网络与其他方法相比具有良好的非线性映射能力、自学习自适应能力和并行信息处理能力,善于处理多因素、非线性的复杂关系.其独特之处在于它能够根据历史数据反复学习、联想、综合和推广,找出它们之间的内部复杂关系,并用权值和阈值的形式表现出来.它被用于工业用水量预测中有其独特的优越性.但传统的 BP 模型存在输入因子选取受主观因素影响大、学习效率低、收敛速度慢和易陷入局部最优等缺陷.针对这些问题很多学者从不同角度对 BP 神经网络进行了改进^[7-9],但在工业用水量预测中通过相关性分析选取输入因子,同时用遗传算法(genetic algorithms, GA)优化 BP 模型的权值和阈值还未见报道.本文基于自相关分析选取输入因子,同时运用遗传算法优化 BP 网络权值和阈值,建立了工业用水量预测改进 BP 神经网络模型.这既克服了传统 BP 模型选取输入因子时的不足,又能利用 GA 高效的全局寻优能力来获取 BP 模型的最佳初始权值和阈值.以永定河山区工业用水量预测为例,将此模型与传统非线性预测模型进行比较分析,进而优选出合适的预测模型,以期为区域水资源管理提供决策依据.

1 预测模型分析

1.1 传统非线性回归模型

非线性回归法实质上属于时间序列法,常用对数、多项式、指数和生长曲线等解析模型^[10].回归技术是根据过去相互影响、相互关联的因素的资料,由于不确定的函数关系,利用数学方法建立相互关系,拟合成一条确定曲线,然后外延到适当时间,得到预测值.具体回归方法可参考文献[11—12].

当用水量增长过程符合生长曲线模型(图 1)时,可用庞伯兹(Compertz)公式来预测,即

$$Q = L \exp(-be^{-kt}) \quad (1)$$

式中: b 和 k 为待定系数; Q 为预测年用水量; L 为预测用水量的上限值; t 为预测时段.

为求出参数 b 、 k ,对式(1)进行线性变换

$$\ln[\ln(L/Q)] = \ln b - kt \quad (2)$$

对式(2)采用最小二乘法或线性规划法来确定模型中的参数 b 、 k ,代入式(1)则可得预测模型.

万方数据

该法关键是确定用水量的上限值 L ,可按下面的方法,一种是以城市水资源的限量为约束条件,通过对水资源总量进行分配来确定;另一种是参考其他发达国家类似工业结构的城市确定.

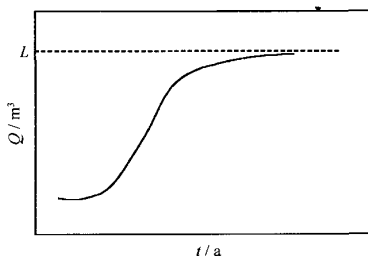


图 1 生长曲线模型

Fig. 1 Growth curve model

1.2 改进 BP 神经网络模型

BP 神经网络和 GA 的具体学习过程可分别参考文献[13—14],此处不再赘述.

对 BP 网络的改进主要包括两方面.首先为了提高预测精度,传统 BP 模型一般采用增加输入节点和样本量的方法,其结果是增加了系统的复杂性,加大了计算量,收敛速度减慢,使网络的泛化能力减弱,预测能力下降.本文通过计算输入因子和输出因子之间的相关系数,判断其相关性,从而为合理确定网络输入因子提供指导.其次由于 BP 网络收敛速度慢,易陷入局部极小值,同时随机赋初始权值和阈值,很难选取具有全局性的初始点,不能保证收敛到全局极小点.本文利用 GA 较强的全局搜索特性,优化 BP 网络的连接权值和阈值.

BP 网络输入因子的选取:根据用水量时间序列的自相关系数和用水量与时间的相关系数来确定输入层节点数(即影响因子).

用水量时间序列 $x(t)$ ($t=1,2,\dots,n$)的自相关系数^[15]为

$$r_k = \frac{\sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2} \quad (3)$$

其中

$$\bar{x} = \frac{1}{n} \sum_{t=1}^n x_t \quad k=0,1,2,\dots,n$$

当 $|r| \in (0.8, 1)$ 时属于高度相关, $|r| \in (0.5, 0.8)$ 时属于中度相关.

GA 优化 BP 网络的基本思想是:首先利用遗传算法全局搜索的特点,寻找最为合适的网络初始连接权

值和阈值,其后利用 BP 算法较强的局部精确搜索能力最终完成网络训练。

GA 优化 BP 网络权重的具体计算过程有以下 7 个步骤。

(1) 利用 BP 网络初步确定连接权的基本解空间。对基本解空间进行编码,编码方法采用实数级联编码,编码按一定顺序连成一个长串,每个串对应一组权值和阈值。

(2) 选择神经网络的误差平方和作为遗传算法的目标函数,由于遗传算法在优化过程中以目标函数最大值作为其适应度函数,故可定义适应度函数为误差平方和的倒数,即

$$F(W_i) = \frac{C}{E(W_i)} \quad i=1,2,\dots,N \quad (4)$$

式中: $F(W_i)$ 为适应度函数; $E(W_i)$ 为目标函数; C 为常数; N 为种群规模。

(3) 计算群体中每个个体的适应度,保留群体中适应度最高的个体直接进入下一代,对其它个体进行选择,选择过程采用标准化几何排序选择。

(4) 分别采用基本遗传算法中的算数交叉和非均匀变异进行交叉和变异运算。

(5) 生成新一代群体。

(6) 重复进行步骤(3)~(5),每进行一次,群体就进化一代,连续进化到 K 代(总的进化代数)。

(7) 第 K 代适应度最高的个体即为相应的网络连接权,输入 BP 网络中,再用 BP 算法进一步训练网络。

2 模型优选实例分析

以永定河山区工业用水量预测为例,通过对模型适用条件进行分析和已有数据系列的特点,初步用对数、多项式、指数和改进的 BP 神经网络模型分别进行预测,进而筛选出适合该区域的预测模型。

2.1 相关性分析

由于该区域发展尚不稳定,统计资料贫乏,针对其只有 1988—2003 历年工业用水量与时间关系的“贫信息”情况(图 2),通过式(3)分别计算用水量与时间的相关系数 r 和用水量的自相关系数 r_k : $r = -0.2$, $r_1 = 0.53$, $r_2 = 0.32$, $r_3 = 0.16$ 。

可见用水量与时间的相关性不好,由用水量自相关系数可知该年用水量与前一年用水量呈中度相关,即前一年用水量对该年用水量影响较大。

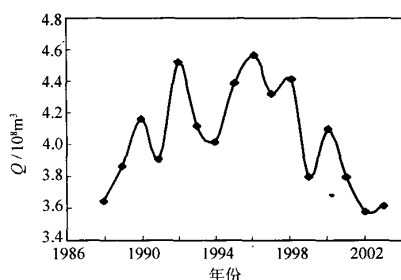


图 2 工业用水量变化曲线

Fig. 2 Variety curve of the industrial water demand

2.2 传统非线性预测模型

由图 2 可以看出工业用水量随时间的变化趋势比较复杂,呈波动性变化,很难用回归方法拟合出一条精度较高的曲线,这时可以对原始工业用水量 $Q_i^{(0)}$ 进行处理^[16],同时由相关性分析可知前一年用水量与该年用水量之间的相关性比较好,故最终考察累计工业用水量 $Q_i^{(1)}$ 与前一年累计工业用水量 $Q_{i-1}^{(1)}$ 的关系(见图 3),使数据的随机性弱化,其中

$$Q_i^{(1)} = \sum_{j=1}^i Q_j^{(0)} \quad (5)$$

$$Q_i^{(1)} = f(Q_{i-1}^{(1)}) \quad (6)$$

式中: $Q_i^{(0)}$ 为原始工业用水量第 i 个数据; $Q_i^{(1)}$ 为一次累计生成第 i 个数据。

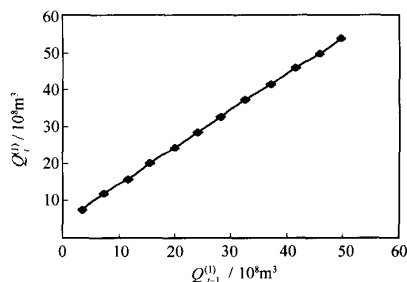


图 3 累计工业用水量关系曲线

Fig. 3 Accumulative industrial water demand curve

由图 3 可以看出,累计工业用水量与前一年累计工业用水量的关系比较明显,可以通过非线性回归进行拟合。由图 2 可知工业用水量变化规律不符合生长曲线模型的特点,不宜采用生长曲线模型预测。

通过对累计工业用水量与前一年累计工业用水量进行回归分析,拟合出的对数、多项式和指数模型分别为

$$Q_t^{(1)} = 19.229 \ln(Q_{t-1}^{(1)}) - 27.288 \quad (7)$$

$$Q_t^{(1)} = -0.0008(Q_{t-1}^{(1)})^2 + 1.0436(Q_{t-1}^{(1)}) + 3.7315 \quad (8)$$

$$Q_t^{(1)} = 2.4767(Q_{t-1}^{(1)})^{0.7775} \quad (9)$$

式中: $Q_t^{(1)}$ 为累计工业用水量; $Q_{t-1}^{(1)}$ 为前一年累计工业用水量. 再通过递减即可获得原始工业用水量序列.

用水量实际值和对数模型的拟合值和检验值见图 4, 多项式和指数模型的拟合值和检验值见图 5, 多项式和指数模型拟合值和检验值的相对误差曲线见图 6.

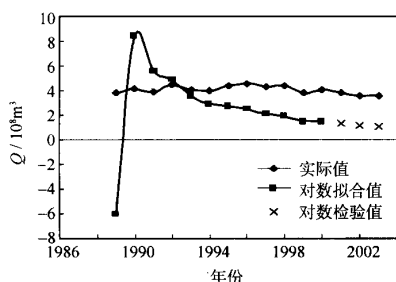


图 4 对数模型拟合和检验值
Fig. 4 Fitting curve and test values of the logarithm model

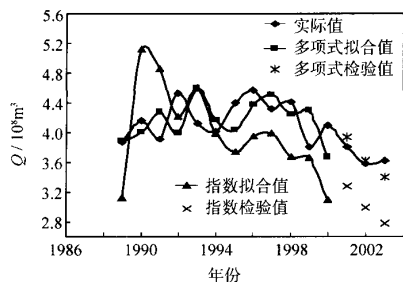


图 5 多项式与指数模型拟合曲线和模型检验值
Fig. 5 Fitting curve and test values of the polynomial and exponent model

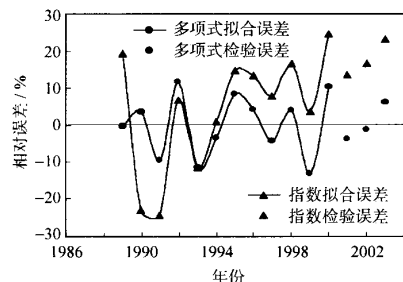


图 6 多项式和指数模型相对误差曲线
Fig. 6 Relative error curve of the polynomial and exponent model

对数模型的拟合平均相对误差达到 61.8%, 同时由图 4 可看出模型的拟合程度不好, 不适合预测该区域用水量, 在筛选时被淘汰. 由图 4 和图 5 可看出, 传统非线性预测模型拟合精度都不高, 分析其原因, 一方面是由于传统的非线性拟合系根据过去相互影响、相互关联的因素的资料进行预测, 因此对累积用水量和用水量采用对数、多项式和指数模型拟合时, 存在模型变量单一的不足; 另一方面由于每种预测模型进行预测时都或多或少存在系统误差. 由图 6 也可看出模型的相对误差较大, 多项式模型的拟合和检验精度相对比指数模型高.

2.3 改进 BP 神经网络模型

Funahashi^[17] 在理论上已证明, 一个 3 层网络可以以任意精度逼近任意给定的连续函数, 所以在永定河山区工业用水量预测中采用了 3 层的 BP 网络结构.

针对永定河山区工业用水量预测, 改进 BP 网络的具体计算过程如下所述.

首先运用 GA 优化 BP 网络的初始权值和阈值, 并将优化结果赋予 BP 网络, 然后用 BP 算法作进一步优化, 网络确定后即可进行预测.

根据相关性分析结果, 选取上一年用水量作为输入因子. 根据实际问题输出因子为相应年用水量. 选取 1988—2000 年的用水量数据系列作为训练样本, 检验样本数至少为输入节点数的 2 倍, 检验样本数定为 2001—2003 年的用水量数据, 建立了用水量预测 BP 网络模型. 通过调试最终构造了一个 1-5-1 型的 BP 网络, 经过 138 次训练后误差达到 0.000 98. 而改进 BP 网络经过 3 次训练误差就达到 5.22914×10^{-5} . BP 算法主要参数为: 最大训练步数为 500 次, 误差目标设为 0.001, 隐含层激活函数采用 S 型正切函数 tansig, 输出层激活函数采用 S 型对数函数 logsig, 训练函数采用 trainlm.

GA 的主要设计参数取值为: 初始权重的取值域 $b = [-50, 50]$, 种群规模 $N = 150$; 选择概率 $P_s = 0.05$; 交叉概率 $P_c = 0.1$; 变异概率 $P_m = 0.05$; 进化代数 $K = 100$.

利用训练好的网络对永定河山区工业用水量进行实例计算, 改进 BP 网络模型的实际值、拟合值及检验值见图 7, 其拟合值和检验值相对误差曲线见图 8.

由图 7 中可以看出改进的 BP 模型拟合精度很高, 检验精度也比较高, 反映了改进的 BP 网络对于非线性问题的良好学习能力. 当遇到复杂非线性问题时, 可以考虑用该模型来处理. 从图 8 中可以看出, 拟合最大相对误差为 1.5%, 检验最大相对误差为 6%, 说明训练好的改进 BP 网络具有较好的拟合性, 可以用来进

行预测.用该模型预测的 2010 年和 2030 年用水量将分别达到 $3.63 \times 10^8 \text{ m}^3$ 和 $4.46 \times 10^8 \text{ m}^3$,1988—1998 年工业用水量增长率为 1.9%,2010—2030 年工业用水量增长率为 1.0%,比较符合用水量的发展规律,即随着经济的发展和工业产值的增长等用水量会增长,但由于工业节水水平的提高、产业结构的调整等因素,工业用水量增长率呈逐年减小的趋势.

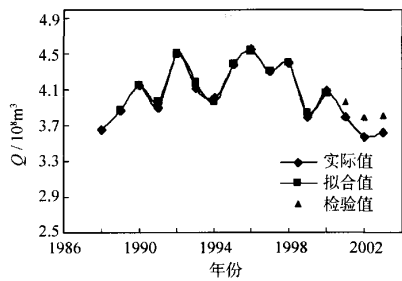


图 7 改进 BP 模型拟合曲线和模型检验值
Fig. 7 Fitting curve and test values of the improved BP model

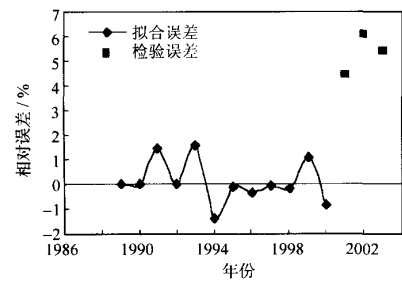


图 8 改进 BP 模型相对误差曲线
Fig. 8 Relative error curve of the improved BP model

多项式、指数和改进 BP 模型的平均相对误差分别为 6.96%、13.78%和 1.41%,说明改进 BP 模型拟合精度高,预测效果更好. 多项式、指数和改进 BP 神经网络模型预测的 2010 年和 2030 年用水量列于表 1 中,由表中可看出,多项式模型预测的2030年用水量

表 1 模型预测值

Tab.1 Predicted values of the models

年度	10 ⁸ m ³		
	多项式模型	指数模型	改进 BP 模型
2010	2.01	0.026 9	3.63
2015	1.23	0.006 5	3.645
2020	0.71	0.002 0	3.649
2025	0.40	0.000 6	3.71
2030	0.22	0.000 2	4.46

万方数据

减小至 $0.22 \times 10^8 \text{ m}^3$,指数模型预测的 2010 年和 2030 年用水量分别为 $2.69 \times 10^6 \text{ m}^3$ 和 $2 \times 10^4 \text{ m}^3$,变化不符合用水量的发展趋势,最终筛选出改进 BP 神经网络模型.

3 结 论

(1) 多项式、指数和改进 BP 模型的拟合最大相对误差分别为 11.61%、24.51%和 1.5%,检验最大相对误差分别为 6.2%、23.25%和 6%,改进 BP 模型的拟合和检验精度都比多项式和指数模型好. 实际计算结果显示,改进 BP 模型的平均相对误差比较小,能更好处理这种“贫信息”的情况,预测效果最好.

(2) 从预测结果来看,改进 BP 模型预测的 2010 年和 2030 年用水量分别达到 $3.63 \times 10^8 \text{ m}^3$ 和 $4.46 \times 10^8 \text{ m}^3$. 预测结果更符合用水量远期发展规划和用水量发展趋势,即工业用水量缓慢上升,用水量增长率呈下降趋势.

(3) 通过相关性分析选取 BP 网络的输入因子,使得输入因子的选取更加合理化,增加了网络设计的合理性. GA 具有良好的全局搜索能力,BP 网络具有较强的局部寻优能力,用 GA 优化 BP 网络的权值和阈值,可以达到优势互补,解决了 BP 网络收敛速度慢,易陷入局部极小点等问题. 改进的 BP 网络善于处理非线性的复杂关系,并可根据已有历史资料进行学习、联想、综合和推广,使复杂问题简化,提高了网络的泛化能力及预测精度,较传统的方法有更多的优越性,在预测领域有广阔的发展前景.

参考文献:

[1] 刘俊良. 城市节制用水规划原理与技术[M]. 北京:化学工业出版社,2003.
Liu Junliang. *Programming Principle and Technology of the City's Water-Saving* [M]. Beijing: Chemical Industry Press, 2003(in Chinese).
[2] 李红艳,崔建国,张星全. 城市用水量预测模型的优选研究[J]. 中国给水排水,2004,20(2):41—43.
Li Hongyan, Cui Jianguo, Zhang Xingquan. Preference study on urban water consumption forecasting model [J]. *China Water and Wastewater*, 2004, 20(2): 41—43 (in Chinese).
[3] 满广生,徐得潜,陈国炜. 城镇工业用水量预测方法研究[J]. 合肥工业大学学报:自然科学版,2002,25(2):204—207.
Man Guangsheng, Xu Deqian, Chen Guowei. Research on the forecasting method for urban industrial water demand

- [J]. *Journal of Hefei University of Technology: Natural Science Edition*, 2002, 25(2): 204—207 (in Chinese).
- [4] Liu Junguo, Hubert H G S, Xu Jianxin. Forecast of water demand in Weinan City in China using WDF-ANN model [J]. *Physics and Chemistry of the Earth*, 2003, 28(4/5): 219—224.
- [5] Narate Lertpalangsunti, Christine W Chan, Ralph Mason, et al. A toolset for construction of hybrid intelligent forecasting systems: Application for water demand prediction [J]. *Artificial Intelligence in Engineering*, 1999, 13(1): 21—42.
- [6] Liu K, Subbarayan S, Shoultz R R, et al. Comparison of very short-term load forecasting techniques [J]. *IEEE Transaction on Power Systems*, 1996, 11(2): 877—882.
- [7] 李勋贵, 黄 强, 魏 霞, 等. 基于 IGA-GP 和 BP 网络黄河流域需水预测研究[J]. 西北农林科技大学学报: 自然科学版, 2005, 33(3): 116—120.
- Li Xungui, Huang Qiang, Wei Xia, et al. Application of IGA-GP & BP neural network method to water-requirement prediction in Yellow River Basin [J]. *Journal of Northwest Science-Technology University of Agriculture and Forest: Natural Science Edition*, 2005, 33(3): 116—120 (in Chinese).
- [8] 方 浩, 李 蓓, 石 娜, 等. 基于模糊神经网络的区域需水预测计算模型[J]. 西北水资源与水工程, 2003, 14(4): 1—3.
- Fang Hao, Li Bei, Shi Na, et al. Calculation model of the water demand prediction in area based on fuzzy neural network [J]. *Northwest Water Resources and Water Engineering*, 2003, 14(4): 1—3 (in Chinese).
- [9] Bazartseren B, Hildebrant G, Holz K P. Short-term water level prediction using neural networks and neuro-fuzzy approach [J]. *Neurocomputing*, 2003, 55(3/4): 439—450.
- [10] 钱 易, 刘昌明, 邵益生. 中国城市水资源可持续开发利用[M]. 北京: 中国水利水电出版社, 2002.
- Qian Yi, Liu Changming, Shao Yisheng. *Sustainable Development and Utilization of China Cities' Water Resources* [M]. Beijing: China Water Conservancy and Hydropower Press, 2002 (in Chinese).
- [11] Kleinbaum D G. *Applied Regression Analysis and Other Multivariable Methods* [M]. Beijing: China Machine Press, 2003.
- [12] 王惠文. 偏最小二乘回归方法及其应用[M]. 北京: 国防工业出版社, 1999.
- Wang Huiwen. *Partial Least-Squares Regression Method and Applications* [M]. Beijing: National Defence Industry Press, 1999 (in Chinese).
- [13] 王小平, 曹立明. 遗传算法——理论、应用与软件实现[M]. 西安: 西安交通大学出版社, 2002.
- Wang Xiaoping, Cao Liming. *Genetic Algorithms: Theory, Application and Software Implementation* [M]. Xi'an: Xi'an Jiaotong University Press, 2002 (in Chinese).
- [14] 飞思科技产品研发中心. 神经网络理论与 MATLAB7 实现[M]. 北京: 电子工业出版社, 2005.
- Fecit Science and Technology Product Research and Development Center. *Neural Network Theory and MATLAB7 Realization* [M]. Beijing: Electronics Industry Publishing House, 2005 (in Chinese).
- [15] 丁 晶, 邓育仁. 随机水文学[M]. 成都: 成都科技大学出版社, 1988.
- Ding Jing, Deng Yuren. *Stochastic Hydrology* [M]. Chengdu: Chengdu Science and Technology University Press, 1988 (in Chinese).
- [16] 李云雁, 胡传荣. 试验设计与数据处理[M]. 北京: 化学工业出版社, 2005.
- Li Yunyan, Hu Chuanrong. *Experimental Design and Data Processing* [M]. Beijing: Chemical Industry Press, 2005 (in Chinese).
- [17] Funahashi K. On the approximate realization of continuous mapping by neural networks [J]. *Neural Networks*, 1989, 2(3): 183—192.