

文章编号: 1000-7709(2015)07-0039-04

基于灰色关联分析的 GA-BP 网络需水预测模型研究

臧冬伟^a, 陆宝宏^{a, b}, 朱从飞^a, 陆建宇^a, 刘 欢^a, 左 建^a

(河海大学 a. 水文水资源学院; b. 水文水资源与水利工程科学国家重点实验室, 江苏 南京 210098)

摘要: 针对城市需水预测涉及因素众多、不同地区影响因子不尽相同且多寡不一及影响因子的选择直接决定需水量预测的结果与实际是否相符等问题, 提出了灰色关联分析法、遗传算法和 BP 神经网络相结合的需水预测模型, 并以南京市为例, 通过灰色关联分析法筛选出主要影响因素, 采用遗传算法优化 BP 神经网络, 构建基于灰色关联分析的 GA-BP 神经网络需水预测模型。实例应用结果表明, 该模型用于需水预测能够比较全面地考虑需水量影响因子, 与传统 BP 网络相比, GA-BP 网络预测精度更高, 训练速度更快, 可作为资料时间序列较短情况下一种较好的需水预测方法。

关键词: 灰色关联分析; GA-BP 神经网络; 需水预测; 南京市

中图分类号: TV213.4; TP183

文献标志码: A

1 引言

我国水资源相对短缺, 在经济快速发展、水资源需求量逐渐增长的今天, 正确预测社会需求水量, 对社会、经济和环境的和谐发展以及重大水资源工程的合理决策和实施均具有重大意义。需水量预测是一个由区域总人口、产业结构、各产业发展水平、社会经济发展程度等共同作用的多因素、多层次的复杂非线性系统^[1]。现有的需水预测方法主要有趋势外推法、人工神经网络、灰色模型理论、宏观经济模型法、回归分析模型法、支持向量机等^[2], 从总体上说, 上述方法均未充分考虑城市需水预测的复杂性, 预测效果不太理想。因此, 有必要研究高效适用的预测方法, 以提高需水预测精度。城市需水量是一个受众多因素影响的变量, 目前尚不存在公认的普适性预测模型, 而不管采用何种预测模型, 第一步均是筛选影响因子。影响因子选择过多, 会使网络复杂化, 且可能陷入局部最优; 选择过少, 会影响预测精度。人工神经网络因其强大的非线性拟合能力, 在需水量预测中得到了广泛应用, 但存在网络泛化能力不高, 求解时容易陷入局部最小, 且收敛速度慢、迭代时间

长等不足。为此, 本文针对上述需水量影响因素选择问题以及神经网络模型在预测中表现出的不足, 提出基于灰色关联分析的 GA-BP 需水预测模型, 通过灰色关联分析筛选主要影响因子, 采用遗传算法优化 BP 神经网络, 以实现简化网络结构, 提高预测精度的目的, 为需水预测提供一种新的、可行的方法。

2 方法原理

2.1 灰色关联分析

城市需水量是社会经济大系统正常运行的水资源需求量, 其变化过程具有灰色过程特征。灰色关联分析作为灰色系统理论的重要组成部分, 提供了一种分析计算系统中各因素关联程度的量化方法。通过比较关联度可在系统中找出主次影响因子, 从而得到影响某一变量的主要因子。该方法对样本数量多少无严格要求, 数据也无需有典型的分布规律, 分析结果不会与定性分析的结果大相径庭, 具有广泛的实用性。具体计算步骤如下。

步骤 1 确定参考数列和比较数列。

步骤 2 对参考数列和比较数列数据进行无

收稿日期: 2014-09-30, 修回日期: 2014-11-24

基金项目: 国家自然科学基金项目(50979023); 水利公益项目(201201026); 中国博士后科学基金资助项目(2013M531270); 江苏省博士后基金资助项目(1302029C)

作者简介: 臧冬伟(1990-), 男, 硕士研究生, 研究方向为水资源规划与管理, E-mail: hhuzdw@163.com

通讯作者: 陆宝宏(1962-), 男, 教授, 研究方向为水文水资源, E-mail: lubaohong@126.com

量纲化处理。

步骤 3 求解比较数列和参考数列的灰色关联系数。

步骤 4 计算关联度并进行排序,选出主要影响因子。

关联系数 $\zeta_i(k)$ 为:

$$\zeta_i(k) = \frac{\min_k \min_i |X_0(k) - X_i(k)| + \rho \max_k \max_i |X_0(k) - X_i(k)|}{|X_0(k) - X_i(k)| + \rho \max_k \max_i |X_0(k) - X_i(k)|} \quad (1)$$

关联度 r_i 为:

$$r_i = \frac{1}{N} \sum_{k=1}^N \zeta_i(k) \quad i = 1, 2, \dots, m \quad (2)$$

式中, $X_0(k)$ 、 $X_i(k)$ 分别为参考数列和比较数列, $k = 1, 2, \dots, n$, $i = 1, 2, \dots, m$; ρ 为分辨系数, 介于 0~1 之间, 通常取 0.5。

关联系数是参考数列和比较数列在各个时刻的关联程度值, 关联度则反映了双方在计算时段内的平均程度值。关联度越大, 表明比较数列对参数数列影响越大, 反之则越小。

2.2 遗传算法

遗传算法求解问题的计算步骤如下。

步骤 1 进行初始化, 设置进化代数计数器及最大进化代数, 随机生成一定数量个体作为初始种群。

步骤 2 进行个体评价, 分析计算种群中各个体的适应度值。

步骤 3 将各种遗传算子(选择、交叉、变异等)作用于种群, 得到新一代种群, 重复步骤 1~2, 直到满足期望的终止条件^[3]。

2.3 BP 神经网络

BP 神经网络^[4]的计算过程包括正向计算过程和反向计算过程。输入信号先向前传输到隐含层, 经过传递函数后, 将隐含层输出信息传递给输出层, 最后输出结果。若实际输出与期望输出存在误差, 则进行误差反向计算, 将误差沿原来的道路送回, 修改各层节点权、阈值使误差变小, 如此反复, 直到误差满足要求。

(1) 正向计算过程。采用三层网络结构, 设输入层有 n 个神经元, 隐含层有 i 个神经元, 输出层有 m 个神经元, 网络输入为 x_1, x_2, \dots, x_n , 则隐含层各神经元的输入为:

$$I_l = \sum_{j=1}^n w_{lj} x_j + \theta_l \quad l = 1, 2, \dots, i \quad (3)$$

式中, w_{lj} 为隐含层神经元 l 与输入层神经元 j 的连接权值; x_j 为网络输入, $j = 1, 2, \dots, n$; θ_l 为隐含层各神经元的阈值。

隐含层各神经元的输出为:

$$O_l = f(I_l) \quad l = 1, 2, \dots, i \quad (4)$$

式中, f 为隐层神经元的激发函数。

输出层各神经元的输出 L_k 和通过传递函数计算的实际输出 c_k 分别为:

$$L_k = \sum_{l=1}^i v_{lk} O_l + \gamma_k \quad k = 1, 2, \dots, m \quad (5)$$

$$c_k = h(L_k) \quad k = 1, 2, \dots, m \quad (6)$$

式中, v_{lk} 为隐含层神经元 l 与输出层神经元 k 的连接权值; γ_k 为隐层神经元阈值; h 为输出层神经元的激活函数。

(2) 反向计算过程。由通过传递函数计算的实际输出 c_k 可计算出样本 p 的误差函数为:

$$e_p = \frac{1}{2} \sum_{k=1}^p (t_{kp} - c_{kp})^2 \quad (7)$$

式中, t_{kp} 为期望输出, 从而可得到网络连接权重向量 W 的修正值 ΔW , 为 $-Z \partial e_p / \partial W$, 运用 ΔW 对 W 进行修正, 完成网络的反向计算过程。

2.4 GA-BP 神经网络模型

利用遗传算法优化 BP 神经网络并进行训练的主要步骤如下。

步骤 1 确定网络参数范围以及编码长度 S 。

$$S = RS_1 + S_1 S_2 + S_1 + S_2 \quad (8)$$

式中, R 为输入层神经元个数; S_1 为隐含层神经元个数; S_2 为输出层神经元个数。

步骤 2 初始化种群, 包括交叉规模、交叉概率、突变概率的选择及网络隐含层和输出层之间连接权值的初始化。

步骤 3 设定相关参数, 进行种群个体的适应度评价及遗传操作, 生成新种群; 反复进行直到达到设定的进化代数, 得到进化解; 其中, 适应度函数采用误差平方和函数。

步骤 4 将最终群体中的最优个体解码, 并作为 BP 神经网络各层的连接权值和阈值。

步骤 5 设置神经网络训练参数, 进行训练、仿真、预测。

GA-BP 网络模型流程见图 1。

3 实例应用

3.1 概况

南京市地处北亚热带季风湿润气候区, 雨量充沛、四季分明, 水资源丰富。如何合理充分利用丰富的水资源, 对城市的经济快速稳定发展至关重要。因此有必要对南京市未来需水量进行预测, 为合理配置水资源提供依据。

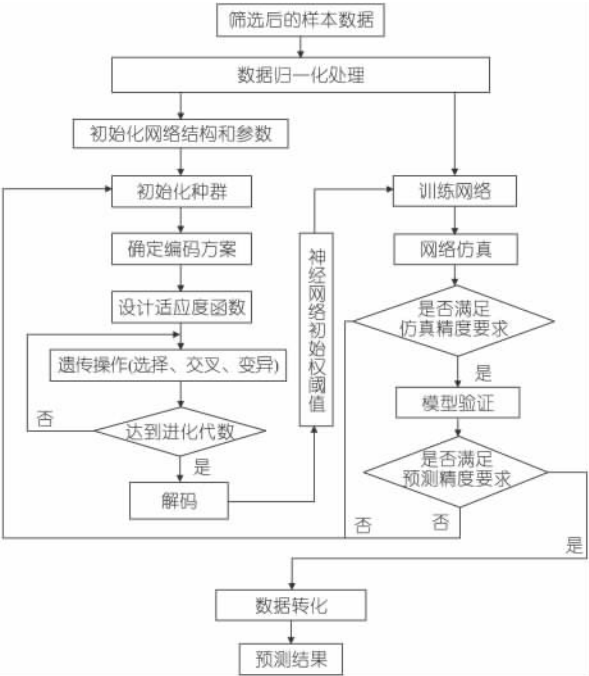


图 1 GA-BP 网络模型流程图

Fig. 1 GA-BP neural network model flow chart

在充分考虑需水量影响因素的基础上,依据《南京市统计年鉴 2000~2013》^[5]和《南京市水资源公报 1999~2012》^[6]相关资料,以 1999~2012 年为时间序列年,初步选取 12 个因子作为南京市水资源需求量的影响因子,其中, x_1 为南京市总人口(10^4 人), x_2 为国内生产总值(10^8 元), x_3 为工业总产值(10^8 元), x_4 为固定资产投资(元), x_5 为城市人均日生活用水量(L), x_6 为城市居民人均可支配收入(元), x_7 为农村居民人均纯收入(元), x_8 为第一产业产值(10^8 元), x_9 为第二产业产值(10^8 元), x_{10} 为第三产业产值(10^8 元), x_{11} 为耕地面积 (10^3 ha), x_{12} 为有效灌溉面积 (10^3 ha)。

3.2 需水影响因子灰色关联度计算

采用标准化处理^[7],对原始数据进行无量纲化,以消除量纲和单位差异,按照灰色关联度计算步骤,分别求出 12 个影响因子与南京市需水量的关联度,并按大小顺序排列,结果见表 1。其中,关联度阈值取 0.8,即选择灰色关联度大于 0.8 的影响因子为主要影响因子,作为 GA-BP 网络

表 1 各影响因子与需水量之间的关联度排序

Tab. 1 Sort of grey relational grade between each influence factor and water demand

序号	因子	关联度	序号	因子	关联度
1	x_{12}	0.921 0	7	x_6	0.856 3
2	x_{11}	0.902 1	8	x_2	0.801 4
3	x_8	0.896 5	9	x_9	0.793 6
4	x_7	0.893 0	10	x_{10}	0.774 1
5	x_5	0.882 5	11	x_3	0.760 4
6	x_1	0.879 3	12	x_4	0.701 0

的输入。由表 1 可看出,工业总产值、固定资产投资、第二产业产值、第三产业产值 4 个因子相比于其他因子而言与需水量的关联度不大,可予以剔除。

3.3 南京市需水量预测模型构建及预测结果

将选取的 8 个主要影响因子作为输入量,南京市实际总用水量作为输出量,建立 3 层 GA-BP 神经网络模型。收集到 1999~2012 年共 14 年各主要因子和南京市用水量的相关数据,选择 1999~2009 年数据作为训练样本,2010~2012 年数据作为检验样本,对模型进行训练。首先对原始数据进行归一化处理,并输入 GA-BP 网络进行学习训练;其中,网络初始参数设定如下:种群规模取 40^[8],遗传代数取 100,交叉概率取 0.95,变异概率取 0.09,隐含层神经元的个数通过试错法最终取 23,网络训练函数使用 TRAINGDM,即有动量的梯度下降法,全局误差函数使用 M_{SE} (均方差),误差目标值设定为 0.001;然后经过学习、训练得到最终的神经网络模型。同时建立传统的 BP 神经网络,使用同样的训练函数、学习速率和误差目标值进行训练。模型预测结果见表 2,模型训练过程中的误差变化曲线分别见图 2、3。

表 2 GA-BP 网络 and 传统 BP 网络的拟合结果比较

Tab. 2 Fitting results comparison between GA-BP neural network and traditional BP neural network

年份	实际需水量/ 10^8 m ³	GA-BP 神经网络		传统 BP 神经网络	
		预测值 / 10^8 m ³	相对误差 /%	预测值 / 10^8 m ³	相对误差 /%
1999	45.46	46.116 3	1.44	46.213 3	1.66
2000	48.78	48.067 5	1.46	48.235 8	1.12
2001	44.80	44.813 2	0.03	44.547 7	0.56
2002	51.63	51.632 0	0.00	51.635 2	0.01
2003	56.91	57.055 9	0.26	56.989 0	0.14
2004	59.91	59.740 1	0.28	59.567 8	0.57
2005	58.99	59.263 6	0.46	59.338 9	0.59
2006	61.18	61.092 9	0.14	61.253 7	0.12
2007	60.33	60.328 7	0.00	60.189 1	0.23
2008	45.07	45.077 0	0.02	45.095 1	0.06
2009	40.77	40.750 8	0.05	40.728 3	0.10
2010	41.84	42.523 7	1.63	39.709 1	5.09
2011	46.64	45.768 6	1.87	44.339 7	4.93
2012	43.85	44.342 4	1.12	43.203 9	1.47

注:基础数据来源于《南京市统计年鉴 2000~2013》^[5]和《南京市水资源公报 1999~2012》^[6]。

由表 2 可看出,GA-BP 网络对训练样本的拟合效果以及对检验样本的预测能力均优于传统 BP 网络,相对误差和平均相对误差均较小,特别是对检验样本的预测,GA-BP 网络效果更好,相对误差最大为 1.87%,而传统 BP 网络检验样本的最大相对误差达 5.09%。由图 2(b)、3 可看出,GA-BP 网络在训练次数达到 726 步时就达到

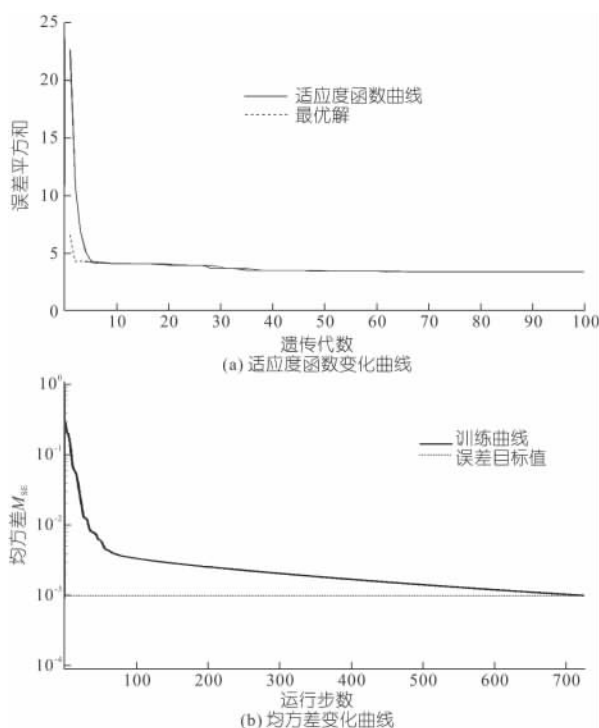


图 2 GA-BP 网络输出效果图

Fig. 2 Output of GA-BP neural network

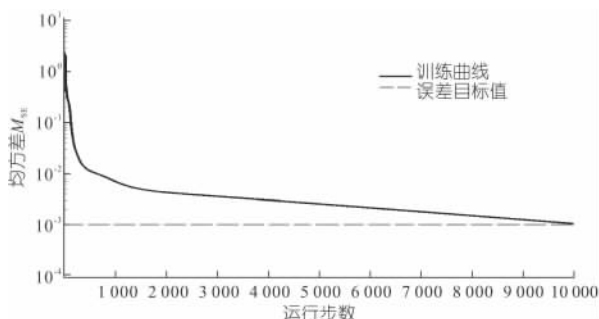


图 3 传统 BP 神经网络误差曲线

Fig. 3 Error curve of traditional BP neural network

了误差目标 0.001, 而传统 BP 网络要达到同样的误差目标, 训练步数则超过了 10 000 步, 并由图 2 (a) 可看出, 随着种群的进化, 网络误差急剧降低, 到 60 代以后基本处于稳定。由此可见, 虽然用遗传算法寻找最优权值和阈值需要些时间, 且会消耗一定的内存, 但对 BP 网络的训练非常快, 能很迅速的收敛到目标值, 对网络的优化效果显著。所以, GA-BP 网络的学习能力和预测能力(泛化能力)均显著优于传统 BP 网络。

按社会经济正常发展趋势和用水水平, 根据《南京市国民经济与社会发展第十二个五年规划纲要》、《南京市城市总体规划 2007~2020》、《南京市农业“十二五”规划》等相关资料数据, 运用已经训练好的 GA-BP 网络对南京市 2015 年需水量进行预测, 结果为 $68.8823 \times 10^8 \text{ m}^3$, 远高于江苏省水利厅设定的 $45 \times 10^8 \text{ m}^3$ 用水指标。根

据灰色关联分析可知, 农业与需水量的关联度较大, 农业需水在南京市总需水中占有很大比重, 所以应着重解决农业节水问题, 推广农业节水技术, 再辅助以其他节水措施, 控制人均日用水量, 将南京市 2015 年用水量控制在 $45 \times 10^8 \text{ m}^3$ 以下。

4 结论

a. 以南京市为例, 通过灰色关联分析, 从 12 个需水影响因子中筛选出 8 个主要因子, 降低了网络复杂度, 提高网络性能和预测精度; 采用遗传算法优化 BP 神经网络, 根据南京市 1999~2012 年的主要需水影响因子和用水量的相关数据, 建立了基于灰色关联分析的 GA-BP 网络需水预测模型。通过与传统 BP 神经网络的拟合和预测结果对比, 发现 GA-BP 网络收敛速度更快、精度更高, 对神经网络的性能优化效果明显, 且模型预测结果与实际需水量变化吻合较好, 说明该模型是一种有效的需水预测方法。

b. 本文虽然应用灰色关联分析筛选出了与需水量关联度大的因子, 但并不能确保这些筛选出的因子之间相互独立, 指标信息可能有重叠, 未来研究可对此寻找方法, 以便对筛选出的主要因子再进行分析筛选, 以消除变量之间的不独立性, 从而进一步降低网络复杂度, 提高网络性能。

参考文献:

- [1] 鲁凤, 徐建华, 王占永. 合肥市需水预测遗传小波神经网络模型研究[J]. 测绘科学, 2013, 38(5): 28-31.
- [2] 张日俊, 董增川, 郭慧芳. 基于支持向量机的鄱阳湖环湖区需水预测模型[J]. 水电能源科学, 2010, 28(4): 22-23, 55.
- [3] 苑希民. 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.
- [4] 张良均. 神经网络实用教程[M]. 北京: 机械工业出版社, 2008.
- [5] 南京市统计局. 南京市统计年鉴 2000~2013[M]. 北京: 中国统计出版社, 2000~2013.
- [6] 南京市水利局. 南京市水资源公报 1999~2012[EB/OL]. <http://njsl.gov.cn/www/njssl/szygbmb-a39070423953.htm>, 2007-04-23.
- [7] 周化如, 吕晓刚, 郝海荣, 等. 不同数据变换方法对灰色关联度分析结果的影响[J]. 种子科技, 2005(5): 280-283.
- [8] 刘晓霞. 种群规模对遗传算法性能影响的研究[D]. 保定: 华北电力大学, 2010.

(下转第 6 页)

参考文献:

- [1] 上海市市政工程设计研究总院,北京市市政工程设计研究总院,中国市政工程东北设计研究院,等. 室外排水设计规范(GB50014-2006)[S]. 北京:中国计划出版社,2006.
- [2] Endreny T A, Imbeah N. Generating Robust Rainfall Intensity-Duration-Frequency Estimates with Short-record Satellite Data[J]. Journal of Hydrology, 2009, 371(1-4):182-191.
- [3] Ben-Zvi A. Rainfall Intensity-Duration-Frequency Relationships Derived from Large Partial Duration Series[J]. Journal of Hydrology, 2009, 367(1-2): 104-114.
- [4] Elsebaie I H. Developing Rainfall Intensity-Duration-Frequency Relationship for Two Regions in Saudi Arabia[J]. Journal of King Saud University-Engineering Sciences, 2012, 24(2), 131-140.
- [5] 张子贤,孙光东,孙建印,等. 城市暴雨强度公式拟合方法研究[J]. 水利学报,2013, 44(11):1 263-1 271.
- [6] 邵尧明. 最大值选样配合指数分布曲线推求雨强公式[J]. 中国给水排水, 2003, 19(13): 142-144.
- [7] 金光炎. 城市设计暴雨频率曲线线型的研究[J]. 水文, 2002, 22(1): 20-22,26.
- [8] 周玉良,袁潇晨,金菊良,等. 基于 Copula 的区域水文干旱频率分析[J]. 地理科学, 2011, 31(11): 1 383-1 390.
- [9] 张子贤. 工程水文及水利计算(第2版)[M]. 北京:中国水利水电出版社,2008.
- [10] 周玉良,袁潇晨,周平,等. 基于地下水埋深的区域干旱频率分析研究[J]. 水利学报, 2012, 43(9): 1 075-1 083.
- [11] 邵尧明. 城市暴雨强度公式应用中存在问题及处理方法[J]. 给水排水, 2008, 34(2): 39-41.

Study on Short-duration Rainstorm Intensity Formula for Anqing City

CHENG Chang¹, ZHOU Yu-liang², CHEN Fen², DING Kun³, JIN Ju-liang², ZHOU Ping²

(1. Central and Southern China Municipal Engineering Design & Research Institute Co., Ltd. in Anhui Branch, Anqing 246002, China; 2. School of Civil Engineering, Hefei University of Technology, Hefei 230009, China; 3. School of Mechanical and Electrical Engineering, Anhui University of Architecture, Hefei 230601, China)

Abstract: For the derivation of short-duration rainstorm intensity formula for Anqing City which is subject to extreme weather recently. The data of rainstorm from 1983 to 2012 were collected, and 10 kinds of rainstorm samples were selected by using annual maximum sampling method. And then, the fitting and optimization of distribution curve were carried out with P-Ⅲ, exponential and Gumbel distribution. The theoretical rain intensity-duration-return period table was obtained by P-Ⅲ distribution. Based on theoretical rain data, the parameters of three-parameter and four-parameter formula were solved through Gauss-Newton method. The results indicate that the parameters can meet the requirement of standard according to accuracy test for three-parameter formula for each recurrence period; however, the four-parameter rainstorm intensity formula for every recurrence period cannot meet the accuracy requirement, and the corresponding causes were analyzed. Compared with the three-parameter rainstorm intensity formula of 1981 version, most of rain intensities of the new formula are higher than the corresponding rain intensity of the rainstorm intensity formula of 1981 version; the rainfall intensities with recurrence periods of 2 a and 5 a are higher than corresponding value estimated through rainstorm intensity formula of 1981 version with 1.4% to 20.4%. The calculated results coincides the recent trends in Anqing City, in which short duration rainfall intensity is increasing under extreme weather. The three-parameter rainstorm intensity formula is commended for estimating the design rainfall intensity when the design return periods are consistent with the return period in the formula, and the rainfall intensity of other design return periods should be estimated by three-parameter storm intensity formula with its parameters obtained by interpolation.

Key words: rainstorm intensity formula; P-Ⅲ-type distribution curve; Gauss-Newton method; Anqing City

(上接第 42 页)

Study of GA-BP Neural Network for Water Requirement Prediction Based on Gray Correlation Analysis

ZANG Dong-wei^a, LU Bao-hong^{a,b}, ZHU Cong-fei^a, LU Jian-yu^a, LIU Huan^a, ZUO Jian^a

(a. College of Hydrology and Water Resources; b. State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210098, China)

Abstract: Urban water requirement prediction normally involves many factors, which are different from place to place, and the choice of influencing factors will directly decide whether the results of water requirement prediction match the actual or not. Hence, this paper proposed a water requirement prediction model which combined the gray correlation analysis method, genetic algorithm (GA) and BP neural network. Taking Nanjing City as an example, based on gray correlation analysis, the main factors are screen out, and the BP neural network optimized by genetic algorithm was employed to establish GA-BP neural network water requirement prediction model. The results show that the established model takes a comprehensive consideration of water demand influencing factors, and it is prior to that of the traditional BP neural network in terms of accuracy and training speed. Therefore, the proposed model can provide a better water demand forecasting method in the case of short time series of data.

Key words: gray correlation analysis; GA-BP neural network; water demand prediction; Nanjing City