

基于云计算和 GA-BP 神经网络的西江水质评价研究

纪广月

(广东工商职业技术大学 公共教学部, 广东 肇庆 526020)

摘要:为提高水质评价精度,针对 BP 神经网络的预测结果易受初始连接权值和阈值的影响以及易陷入局部极值的问题,提出一种遗传算法优化 BP 神经网络的水质评价模型.针对水质评价数据特性,引入 Multi-Agent 和分布式思想,利用云计算的 MapReduce 框架对 GA-BP 模型进行并行化改进,提高其处理海量高维水质评价数据的能力.为证明所提算法的效果,选择西江 2011—2015 年的水质监测数据为研究对象,研究结果表明,与 PSO-BP, GA-BP, DE-BP 和 BP 相比,所提算法 MR-GA-BP 不但可以提高水质评价的精度,而且能够降低计算资源的消耗,缩短训练时间,具有很好的并行性能.

关键词:遗传算法;云计算;BP 神经网络;水质评价;粒子群算法

中图分类号:O232

文献标志码:A

珠江为中国第四大河流,水资源丰富.西江是珠江水系干流之一,西江水环境直接影响着西江流域居民的生活、西江流域的经济和社会发展,也对下游珠江流域的生态环境和可持续发展产生直接的影响.自 1980 年以来,随着广东省经济的飞速发展,水生态环境日益恶化,西江也未能幸免.这些年来,西江水质受到了较严重的污染,致使西江流域生态环境不断下降,系统功能的发挥受到制约.因此,有必要加大对西江水质的治理和保护力度.鉴于此,肇庆市 2016 年政府工作报告中提出了“保护西江”的战略任务.西江水质评价和预测能够为西江水环境治理和保护提供决策依据,对西江流域可持续发展具有重要意义.

水质评价是指根据评价目标,选择水质参数、标准和方法,对水质利用价值及水处理要求做出评定.水质评价的目的在于准确地反映当前水质和水质污染状况,找出水质污染的主要问题,了解掌握主要污染物对水质的影响及未来的变化趋势,为水污染治理、水环境规划和水环境管理提供可靠的方案和依据^[1].对于水质评价的研究,目前主要包括单因子污染指数法^[2]、贝叶斯方法^[3]、布朗水质指数法^[4]、内梅罗水污染指数法^[5]、AHP 法^[6]、主成分分析法^[7]、灰色评价法^[8]以及人工神经网络法^[9]等.

BP 神经网络具有较强的非线性预测能力、自学习、鲁棒性和容错性等优点,被广泛地应用于水质评价.结合河流地表水中有机物和主要影响因素之间的映射关系,文献^[10]提出一种基于 BP 神经网络的水质评价模型,实验结果表明该方法具有较高精度和泛化能力.针对 BP 神经网络小样本数据预测精度低和易陷入局部最优的问题,文献^[11]提出一种改进的 BP 神经网络的水质评价模型,提高了预测精度.针对 BP 神经网络的预测结果易受初始连接权值和阈值的影响以及其容易陷入局部极值的问题,文献^[12]提出一种基于粒子群算法优化 BP 神经网络的水质评价模型,通过粒子群算法优化 BP 神经网络初始连接权值和阈值,实现 BP 神经网络初始连接权值和阈值的自适应最优选择,从而提高水质评价的精度.

随着互联网技术和传感器技术的快速发展,水质评价数据规模由 GB 级向 TB 级,甚至 PB 级增长,GA-BP 进行水质评价的效率和实时性无法满足现实要求,本文将 Multi-agent 思想和云计算技术结合起来用来改进 GA-BP 模型.提高水质评价精度,提出一种遗传算法优化 BP 神经网络的水质评价模型.MR-GA-BP 算法将 Multi-Agent 思想引入 GA-BP,通过权值平衡运行 GA-BP,GA-BP 不同节点具有不同的预测精度,精

收稿日期:2019-03-08;修回日期:2019-04-04.

基金项目:广东省教育厅高校特色创新类项目(自然科学)(2017GKTSCX109)

作者简介(通信作者):纪广月(1973—),女,吉林白城人,广东工商职业技术大学副教授,研究方向为应用数学,E-mail:465421242@qq.com.

度越高的节点在计算 GA-BP 最终预测值时的权重更高.针对水质评价数据特性,引入 Multi-Agent 和分布式思想,利用云计算的 MapReduce 框架对 GA-BP 模型进行并行化改进,提高其处理海量高维水质评价数据的能力.

1 云计算

云计算技术自 2003 年 Google 推出之后便得到人们的广泛关注.Hadoop 云计算平台是由 Apache 基金会开发的一款开源软件,该软件由 MapReduce 程序框架和分布式文件系统构成^[13].MapReduce 程序框架非常适用于在集群中处理海量数据集,其框架屏蔽底层具体实现细节,用来解决分布式存储、运算问题,通过并行计算和函数抽象(抽象为两个可由用户编写的 Map 函数和 Reduce 函数),从而大大降低算法的复杂程度.MapReduce 模型数据处理原理如图 1 所示.

MapReduce 程序框架处理数据原理包括 5 个阶段,其具体过程描述如下.

(1)Input 阶段.首先读取分布式文件系统存储的输入数据,之后对读取的输入数据进行切片处理,读取的数据被分割成一系列的数据片,每一个数据片对应分配给一个 Map 函数.

(2)Map 阶段.将输入的数据切片看成一组 (key, value) 键值对,之后运用用户自己编写的 Map 函数逻辑处理分配的 (key, value) 键值对生成新的 (key, value) 中间过程键值对.

(3)Shuffle 阶段.将 (key, value) 中间键值进行转移,由 Map 节点转向 Reduce 节点,与此同时进行相同 (key, value) 中间键值对的合并和键值排序,受带宽、CPU 运行速度等因素的影响,所耗费的时间可能远长于 Map 函数和 Reduce 函数运行的时间.

(4)Reduce 阶段.执行用户自己编写的 Reduce 函数,根据用户设定的数据处理逻辑处理所有中间值、中间键或中间键链(list of values),将新的 (key, value) 键值对进行输出.

(5)Output 阶段.输出 Reduce 阶段的程序处理结果并存储到分布式文件系统.

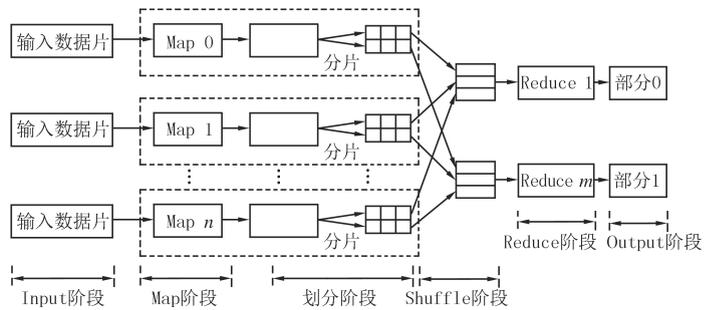


图1 MapReduce原理图
Fig.1 Figure of MapReduce principle

2 基于 GA-BP 神经网络的水质评价模型

2.1 BP 神经网络

如图 2 所示,BP 神经网络是由输入层、隐含层和输出层组成的多层前馈神经网络,3 层典型 BP 神经网络的结构模型图.图 2 中,输入维数和输出维数分别为 m 和 1,隐含层数为 p ,则 BP 神经网络的映射数学表达式为^[14]:

$$x_{i+1} = f(X_i) = \frac{1}{1 + \exp(-\sum_{j=1}^p c_j b_j + \epsilon)}, \quad j = 1, 2, \dots, p, \quad (1)$$

其中, f 表示隐含层激励函数; ϵ 表示输出层的阈值; c_j, b_j 分别表示隐含层到输出层的连接

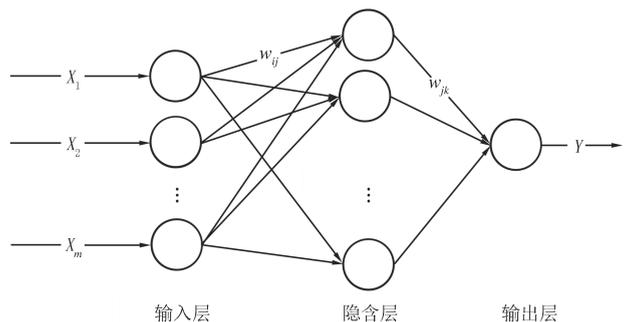


图2 BP神经网络结构图
Fig.2 Figure of network structure

权值和隐含层节点的输出.一般地,BP神经网络的转移函数选择 Sigmoid 函数,其公式如下:

$$f(x) = \frac{1}{1 + e^{-x}}. \quad (2)$$

因此,BP神经网络的隐含层节点的输出可表示为:

$$b_j = \frac{1}{1 + \exp(-\sum_{i=1}^m \omega_{ij} x_i + \theta_j)}, i = 1, 2, \dots, m, \quad (3)$$

其中, ω_{ij} 表示输入层到隐含层的连接权值; θ_j 表示隐含层节点的阈值.

由于BP神经网络的预测结果易受初始连接权值 c_j , ω_{ij} 和阈值 ϵ, θ_j 的影响,其容易陷入局部极值的问题,本文运用GA优化BP神经网络初始连接权值和阈值.

2.2 GA-BP 算法流程

基于GA-BP的水质评价算法流程可归纳为:

步骤1 初始化BP神经网络模型,确定网络结构.根据数据样本确定BP神经网络的层数、传递函数和训练函数类型以及各层节点数;

步骤2 编码.GA算法采用二进制编码,将连接权值 c_j , ω_{ij} 和阈值 ϵ, θ_j 进行整体编码,其编码图如图3所示.算法的搜索空间维数为 m ,假如输入层、隐含层和输出层节点数分别为 R, S_1, S_2 ,则编码长度 S 可表示为:

$$S = RS + S_1 S_2 + S_1 + S_2. \quad (4)$$

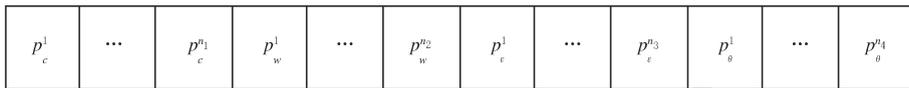


图3 BP参数编码图

Fig.3 Figure of BP parameters of the code

个体的各位二进制串转化成十进制用以表示BP模型的实际参数:

$$c = \min c + \frac{\max c - \min c}{2^j - 1} \times d, \quad (5)$$

其中, c 表示参数的十进制值; $\max c, \min c$ 分别表示参数的最大值和最小值; j 表示参数的二进制串长度; d 表示参数二进制串的十进制值.

步骤3 GA算法参数初始化.设定种群规模 N 、最大迭代次数 T 、交叉概率 P_c 和变异概率 P_m ;

步骤4 确定GA算法优化BP神经网络的适应度函数.为描述种群个体的优劣程度,选择均方误差为适应度函数,如(6)式所示:

$$f_i = \frac{1}{k} \sum_{i=1}^k (y_i - \hat{y}_i)^2, \quad (6)$$

其中, y_i, \hat{y}_i 分别表示第 i 个样本的实际输出和期望输出; k 表示样本数量;

步骤5 按照适应度比例概率随机组合染色体;

步骤6 执行交叉和变异操作产生下一代种群;

步骤7 判定GA算法终止条件,若满足则输出最优解;反之,执行步骤5;

步骤8 输出全局最优解,将最优解对应的连接权值 c_j, ω_{ij} 和阈值 ϵ, θ_j 作为BP神经网络的初始连接权值和阈值,训练BP神经网络并进行预测.

3 基于云计算的MR-GA-BP算法

3.1 MR-GA-BP

根据水质评价数据的特点,GA-BP进行水质评价的效率和实时性无法满足现实要求,本文将Multi-agent思想和云计算技术结合起来用来改进GA-BP模型.MR-GA-BP算法将Multi-Agent思想引入GA-BP,

通过权值平衡运行 GA-BP,GA-BP 不同节点具有不同的预测精度,精度越高的节点在计算 GA-BP 最终预测值时的权重更高.GA-BP 的最终预测值可表示为:

$$\bar{y} = \frac{\sum_{k=1}^K \alpha_k y_k}{\sum_{k=1}^K \alpha_k}, \quad (7)$$

其中, \bar{y} 表示 GA-BP 的最终预测值; α_k 表示第 k 个 GA-BP 的预测权重; y_k 表示第 k 个 GA-BP 的预测值.

预测权重 α_k 可以通过梯度上升法和标准函数误差 E 进行求解:

$$E = 1/2(t - \bar{y})^2, \quad (8)$$

其中, t 表示每个 GA-BP 的输入训练集的期望值.根据梯度上升法,预测权重 α_k 为^[16]:

$$\Delta\alpha_k = -\eta\partial E/\partial\alpha_k = -\eta(\partial E/\partial\bar{y}) \cdot (\partial\bar{y}/\partial\alpha_k), \quad (9)$$

其中, η 表示学习速率,取 $\eta=0.1$.

结合(8)式和(9)式可得:

$$\Delta\alpha_k = \eta(t - \bar{y})[y_k / \sum_{i=1}^K \alpha_i - \sum_{i=1}^K \alpha_i y_i / (\sum_{i=1}^K \alpha_i)^2], \quad (10)$$

其中, α_k 可以通过 $\alpha_k \leftarrow \alpha_k + \Delta\alpha_k$ 进行更新.

3.2 MR-GA-BP 算法流程

通过 MapReduce 程序框架的用户自己编写的 Map 函数和 Reduce 函数实现程序并行化设计,MR-GA-BP 算法进行算法处理和预测,并将预测结果存储到分布式系统和缓存数据库.基于 MapReduce 的分布式 MR-GA-BP 的水质评价模型如图 4 所示.

MR-GA-BP 算法流程描述如下:

步骤 1 读取云计算分布式系统存储的海量水质评价数据的训练集,运用 MapReduce 程序框架对海量水质评价数据的训练集进行分割,获得 k 个不同的水质评价数据训练子集.其中, k 表示云集群中并行 Map 的个数.

步骤 2 运用用户自己编写的 Map 函数程序逻辑处理水质评价数据训练子集,即运用 GA-BP 对水质评价数据训练集进行并行训练,也就是运用 k 个不同的 GA-BP 训练和预测水质.

步骤 3 将 Map 操作后的结果,通过 MapReduce 程序框架的 Shuffle 阶段传输到 Reduce 阶段,也就是将 k 个不同 GA-BP 预测值通过本文的权重计算方法确定各 Map 操作输出的预测值权重,从而实现水质最终预测值的计算.

4 实证分析

4.1 实验环境

本文选择 Hadoop 软件为云计算平台,其版本号为 2.9.1,该平台由 32 个节点组成,每个节点电脑的操作系统均为 Windows7,其配置均为 Intel(R)Core(TM)i5-7200 4-core,CPU 的主频为 2.60 GHz,内存 4 GB,云计算平台如图 5 所示.

4.2 水质评价指标

水质评价是根据水质评价标准和采样水样本各项指标值,通过一定的数学模型计算确定采样水样本的水质等级^[15].由于水质分析的指标有很多,结合地表水环境质量标准^[16],本文采用水质评价的指标为氨氮、溶解氧、化学需氧量、高锰酸盐指数、总磷和总氮 6 项指标,其所对应的水质等级和含量标准见文献^[17].

4.3 数据来源

选择西江 2011—2015 年的水质监测数据为研究对象,采样频率为每季度一次.为提高收敛速度和计算精度,对数据进行归一化处理^[18]:

$$\hat{x} = (x - x_{\min}) / (x_{\max} - x_{\min}), \quad (11)$$

其中, \hat{x} 表示归一化处理之后的数据; x, x_{\min}, x_{\max} 分别表示原始数据、原始数据序列中的最小值和最大值.

为了验证本文算法的有效性,选择200组数据为研究对象,训练集160组,测试集40组.GA算法参数设置如下:种群规模 $N=10$,最大迭代次数 $T=100$,交叉概率 $P_c=0.7$ 和变异概率 $P_m=0.1$,搜索区间 $[-1, 1]$.BP神经网络参数设置如下:输入层节点数(*inputnum*)为8、隐含层节点数(*hiddennum*)为16和输出层节点数(*outputnum*)为1,BP神经网络最大训练次数为1000,隐含层和输出层传递函数为logsig和purelin,训练函数为trainlm,学习速率为0.01,训练误差目标为0.001.为了验证MR-GA-BP算法的优越性,将MR-GA-BP与PSO-BP,GA-BP,DE-BP和BP相比,种群规模10,最大迭代次数100,搜索区间 $[-1, 1]$,PSO算法学习因子 $c_1=c_2=2$,GA算法^[13]交叉概率0.7,变异概率0.1;DE算法^[14]缩放因子0.5,交叉概率0.7,不同算法预测结果分别如表1所示.

由表1知,MR-GA-BP在训练集和测试集上的分类准确率分别为99.3750%和100.0000%,优于PSO-BP,GA-BP,DE-BP和BP;BP的分类准确率最低,分别为87.6640%和86.4737%.通过对比发现,本文算法MR-GA-BP可以有效提高水质评价的准确率.图6(a),6(b)为不同种群规模和不同迭代次数情况下的水质评价结果.随着种群规模的增加,水质评价的准确率整体呈现增加的趋势,其中MR-GA-BP的效果优于GA-BP,PSO-BP和DE-BP,主

要是由于随着种群规模的增加,种群多样性增加,算法搜索能力增强,从而提高了准确率.随着迭代次数的增加,水质评价的准确率整体呈现增加的趋势,其中MR-GA-BP的效果优于GA-BP,PSO-BP和DE-BP,主要因为随着迭代次数的增加,种群搜索时间得到提高,搜索范围一定程度上有了较大的提高,寻优能力得到充分发挥.

表1 水质评价结果

Tab.1 Water quality assessment results

方法	训练集/准确率	测试集/准确率
MR-GA-BP	99.3750%	100.0000%
GA-BP	94.2257%	93.6842%
PSO-BP	91.6010%	91.5789%
DE-BP	93.4752%	94.3542%
BP	87.6640%	86.4737%

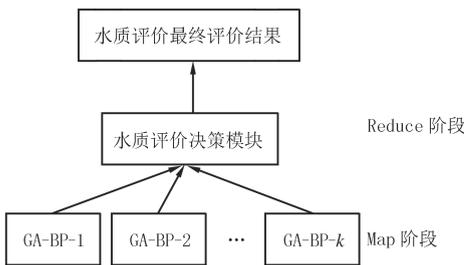


图4 基于MapReduce的分布式MR-GA-BP评价模型

Fig.4 Distributed MR-GA-BP evaluation model

based on MapReduce

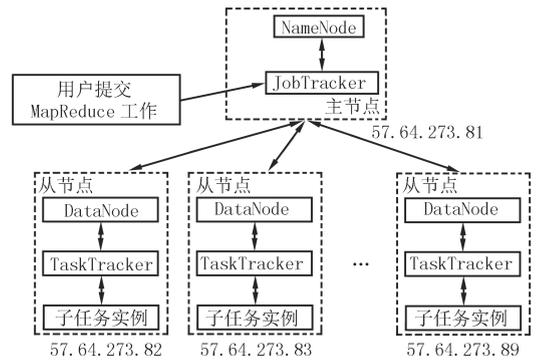
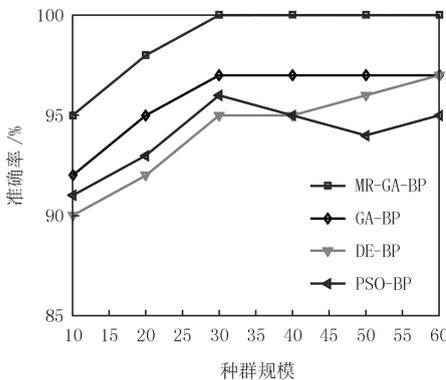
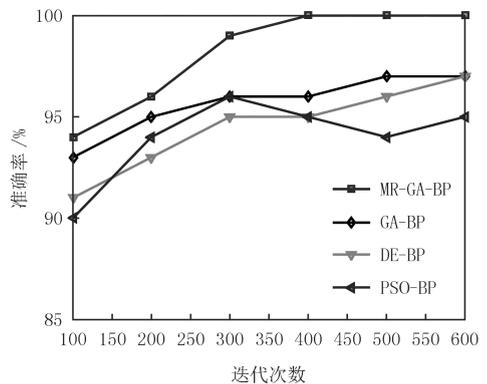


图5 云计算平台

Fig.5 Cloud computing platform



(a) 不同种群规模



(b) 不同迭代次数

图6 不同参数水质评价结果

Fig.6 Water quality evaluation results with different parameters

为了说明MR-GA-BP算法的并行计算能力,将采集的水质评价数据进行人工扩充,分别扩充到原始水

质评价数据集的 1 000 倍、2 000 倍、4 000 倍和 8 000 倍,并在云平台上分别计算对比不同集群节点数的加速比 Speedup,文中分别取集群节点数 4、8、16、32,MR-GA-BP 加速比对比图如图 7 所示。

由图 7 可知,随着数据集的增大,MR-GA-BP 加速比接近于线性,数据量越大加速效果越好,满足海量高维度的水质评价的性能要求,具有很好的实时性和运行效率.由收敛速度对比图 8 可知,与 PSO-BP,GA-BP,DE-BP 和 BP 相比,MR-GA-BP 具有更快的收敛速度和更小的适应度,优化效果最优,从而提高了模型的准确率,值得推广应用。

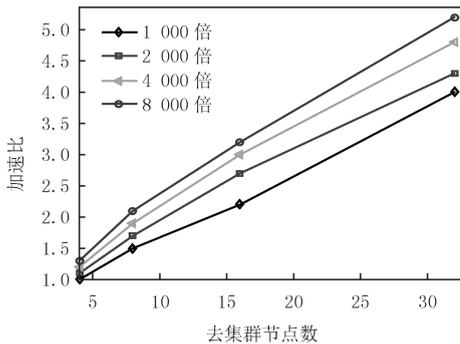


图 7 MR-GA-BP 加速比

Fig.7 MR-GA-BP rate of speedup

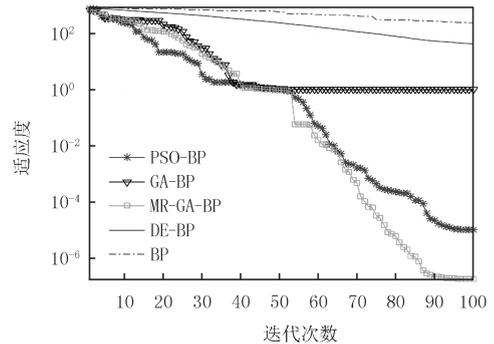


图 8 收敛速度对比图

Fig.8 Contrast Figure of convergence velocity

5 结 论

为提高水质评价精度,针对水质评价数据特性,引入 Multi-Agent 和分布式思想,提出一种基于云计算 MapReduce 框架的 GA-BP 西江水质评价模型.与 PSO-BP,GA-BP,DE-BP 和 BP 相比,本文算法 MR-GA-BP 不但可以提高水质评价的精度,而且能够降低计算资源的消耗,缩短训练时间,具有很好的并行性能。

参 考 文 献

- [1] 佟玲玲.河流综合水质评价方法比较研究[J].科学技术创新,2016(15):113-113.
TONG L L.Comparative Study on comprehensive water quality evaluation method of rivers[J].Scientific and Technological Innovation, 2016(15):113-113.
- [2] 王志垲,张克峰,刘雷.地下水污染评价中单因子指数法的优化[J].环境工程,2016(s1):810-812.
WANG Z K,ZHANG K F,LIU L.Optimization of single factor index method in groundwater contamination assessment.[J].Environment Engineering, 2016(s1):810-812.
- [3] 杨咪,屈文岗,钱会.基于熵权的贝叶斯模型及其在水质评价中的应用[J].灌溉排水学报,2018,37(1):85-90.
YANG M,QU W G,QIAN H.Bayesian Model Based on Entropy Weight and Its Application in Water Quality Assessment[J].Journal of Irrigation and Drainage,2018,37(1):85-90.
- [4] 吴聪慧,王彬浩,虞奕丹,等.氟化工废水处理系统中活性污泥的耐毒性及微生物群落结构特征分析[J].环境科学学报,2017,37(3):962-970.
WU C H,WANG B H,YU Y D,et al.Tolerance abilities and microbial community characteristics of activated sludges in fluoride wastewater treatment plants[J].Acta Scientiae Circumstantiae,2017,37(3):962-970.
- [5] 陈朋,王家鼎,袁亮,等.修正内梅罗指数法和模糊综合评判法在凤凰镇地下水水质评价中的应用[J].水土保持通报,2017,37(2):165-170.
CHEN P,WANG J D,YUAN L,et al.Application of Modified Nemerow Index and Fuzzy Comprehensive Evaluation Method on Groundwater Quality Evaluation in Fenghuang Town[J].Bulletin of Soil and Water Conservation,2017,37(2):165-170.
- [6] 郝洋,梁秀娟,孟凡傲,等.基于 AHP-熵权法的地下水水质模糊综合评价以白城市为例[J].节水灌溉,2016(6):51-54.
HAO Y,LIANG X J,MENG F A,et al.Fuzzy Comprehensive Evaluation of Groundwater Quality Based on AHP-Entropy Weight Method-A Case Study of Baicheng City[J].Water Waving Irrigation,2016(6):51-54.
- [7] 钱程,穆文平,王康,等.基于主成分分析的地下水水质模糊综合评价[J].水电能源科学,2016(11):31-35.
QIAN C,MU W P,WANG K,et al.Fuzzy Comprehensive Evaluation of Groundwater Quality Based on Principal Component Analysis[J].Water Resources and Power,2016(11):31-35.
- [8] 陈婷婷,黄川友,殷彤.灰色-标识指数法在地表水水质评价中的应用[J].人民黄河,2017,39(2):61-64.

- CHEN T T, HUANG C Y, YIN T. Application of Grey-Identification Index Method to Surface Water Quality Assessment[J]. Yellow River, 2017, 39(2): 61-64.
- [9] 曹建智, 张健. 人工神经网络在白洋淀水质评价中的应用[J]. 电子技术与软件工程, 2016(8): 261-262.
CAO J Z, ZHANG J. Application of Artificial Neural Network in Water Quality Evaluation of Baiyangdian[J]. Electronic Technology and Software Engineering, 2016(8): 261-262.
- [10] 王晓萍, 孙继洋, 金鑫. 基于 BP 神经网络的钱塘江水质指标的预测[J]. 浙江大学学报(工学版), 2007, 41(2): 361-364.
WANG X P, SUN J Y, JIN X. Prediction of water quality index in Qiantang River based on BP neural network model[J]. Journal of Zhejiang University(Engineering Science), 2007, 41(2): 361-364.
- [11] 蒋佰权, 王万森, 温香彩. 改进算法的 BP 神经网络水质评价模型[J]. 计算机系统应用, 2007, 16(9): 46-50.
JIANG B Q, WANG W S, WEN X C. An Improved BP Neural Networks Model on Water Quality evaluation[J]. Computer Systems Applications, 2007, 16(9): 46-50.
- [12] 高峰, 冯民权, 滕素芬. 基于 PSO 优化 BP 神经网络的水质预测研究[J]. 安全与环境学报, 2015, 15(4): 338-341.
GAO F, FENG M Q, TENG S F. On the way for forecasting the water quality by BP neural network based on the PSO[J]. Journal of Safety and Environment, 2015, 15(4): 338-341.
- [13] 郭诚. 云环境下的水质安全服务平台关键技术研究[D]. 杭州: 浙江大学, 2013.
GUO C. Research on the key technologies for water quality security service platform in cloud environment[D]. Hangzhou: Zhejiang University, 2013.
- [14] 彭越兮, 徐蔚鸿, 陈沅涛, 等. 改进量子粒子群算法的模糊神经网络水质评价[J]. 计算机工程与应用, 2018(11): 15-21.
PENG Y X, XU W H, CHEN Y T, et al. Improved quantum-behaved particle swarm optimization training fuzzy neural network used in water quality evaluation[J]. Computer Engineering and Applications, 2018(11): 15-21.
- [15] 莫崇勋, 阮俞理, 莫桂燕, 等. 基于 T-S 模糊神经网络模型的钦州市主要河流水质评价[J]. 人民珠江, 2017, 38(8): 80-83.
MO C X, RUAN Y L, MO G Y, et al. Water Quality Assessment of Main Rivers in Qinzhou City Based on T-S Fuzzy Neural Network Model[J]. Pearl river, 2017, 38(8): 80-83.
- [16] 戴秀丽, 许燕娟, 承燕萍. 中国地表水环境质量标准监测体系现状研究及完善建议[J]. 2014, 39(12): 7-10.
DAI X L, XU Y J, CHENG Y P. Present Situation and Suggestions for Surface Water Quality Monitoring Standard System in China[J]. 2014, 39(12): 7-10.
- [17] 张彦波, 司训练. 基于改进灰色关联法的地表水环境质量评价[J]. 人民黄河, 2017, 39(11): 109-111.
ZHANG Y B, SI X L. Quality Assessment of Surface Water Environment Based on Improved Grey Correlation Method[J]. Yellow River, 2017, 39(11): 109-111.
- [18] 刘东, 李帅, 付强, 等. 基于 KHA 优化 BP 神经网络的地下水水质综合评价方法[J]. 农业机械学报, 2018(9): 101-107.
LIU D, LI S, FU Q, et al. Comprehensive evaluation method of groundwater quality based on BP network optimized by krill herd algorithm[J]. Journal of Agricultural Machinery, 2018(9): 101-107.

Study on water quality evaluation of Xijiang river based on cloud computing and GA-BP neural network

Ji Guangyue

(Department of public education, Guangdong University of Business and Technology, Zhaoqing 526020, China)

Abstract: In order to improve the accuracy of water quality evaluation, a genetic algorithm was proposed to optimize the water quality evaluation model of BP neural network. According to the characteristics of water quality evaluation data, Multi-Agent and distributed ideas are introduced, and the GA-BP model is improved in parallel with the MapReduce framework of cloud computing, so as to improve its ability to process massive high-dimensional water quality evaluation data. In order to prove the effect of the algorithm, the water quality monitoring data of Xijiang river from 2011 to 2015 were selected as the research object. The research results show that, compared with PSO-BP, GA-BP, DE-BP and BP, the MR-GA-BP algorithm can not only improve the accuracy of water quality evaluation, but also reduce the consumption of computing resources, shorten the training time and have good parallel performance.

Keywords: genetic algorithm; cloud computing; BP neural network; water quality evaluation; particle swarm optimization algorithm

[责任编辑 陈留院 赵晓华]