

# 基于BP神经网络和模糊聚类分析的 露天煤矿地下水质量评价

王锦力

(辽宁工程技术大学 建筑工程学院, 辽宁 阜新 123000)

**摘要:** 针对阜新新邱露天煤矿排土场淋溶水对附近于家沟地区地下水环境质量的影响所存在的问题,对研究区内的17眼水井分别采样,进行水质分析,采用模糊聚类分析方法对影响区地下水质量进行了评价,再利用BP神经网络对数据进行训练和预测,研究了隐层节点对预测精度的影响,根据研究结论可确定出该区域地下水污染范围,具有一定的工程应用参考价值。

**关键词:** 模糊聚类分析; 环境质量评价; 地下水; BP神经网络

中图分类号: X523 文献标识码: A 文章编号: 1672-643X(2013)03-0138-04

## Evaluation of groundwater quality in open coal mine based on BP neural network and fuzzy clustering analysis

WANG Jinli

(College of Civil Engineering, Liaoning Technical University, Fuxin 12300, China)

**Abstract:** Aimed at the problems of influence of Fuxin Xinqiu surface mine dump leaching water on groundwater environmental quality near Yujiagou, 17 wells in the study area were sampled for water quality analysis. Fuzzy clustering analysis method was used to evaluate the groundwater quality of the area. Then BP neural network was used to train and forecast. The influence of hidden layer node on prediction accuracy was also studied. The conclusion can determine the scope of groundwater pollution in the region and has a certain reference value of engineering application.

**Key words:** fuzzy cluster analysis; environmental quality assessment; groundwater; BP neural network

## 1 概述

露天煤矿地区地下水环境由于受矿区各种因素的影响都受到了不同程度的污染;目前对于露天煤矿区域地下水环境质量评价主要采用模糊聚类分析的方法<sup>[1-12]</sup>。

模糊聚类分析是研究和处理如何根据观测数据将样品进行分类的数学方法,可以把缺乏可靠的历史资料而性质相近的事物归入一类。模糊聚类分析在天气预报、地质探勘、河川水质污染程度、模式识别、数据挖掘以及模糊控制等领域具有广泛的应用。神经网络<sup>[12]</sup>是一个具有高度非线性的超大规模连续时间动力系统。是由大量的处理单元(神经元)广泛互连而形成的网络。它是在现代神经科学研究成果的基础上提出的;其中BP网络是采用Widrow-Hoff学习算法和非线性可微转移函数的多层网络。一个典型的BP网络采用的是梯度下降算法,也就是Widrow

-Hoff算法所规定的。Backpropagation是指的为非线性多层网络计算梯度的方法。一个经过训练的BP网络能够根据输入给出合适的结果,虽然这个输入并没有被训练过。这个特性使得BP网络很适合采用输入/目标对进行训练,而且并不需要把所有可能的输入/目标对都训练过,使得BP神经网络在数据预测方面有着广泛的应用。本文先采用模糊聚类的方法对阜新新邱露天煤矿地下水质量进行污染程度分类,再利用BP神经网络对数据进行训练和预测,研究了隐层节点对预测精度的影响。

## 2 模糊聚类分析原理

设有 $N$ 个事物的总体以 $U$ 表示, $U = \{X_1, X_2, \dots, X_i, \dots, X_n\}$ ,每事物抽取了 $m$ 个特征, $X_i = \{x_{i1}, x_{i2}, \dots, x_{ij}, \dots, x_{im}\}$  ( $i = 1, 2, \dots, n$ ),其中 $x_{ij}$ 表示第 $i$ 个事物第 $j$ 个特征的观测值,所有观测值的全体构成论域矩阵 $U$ 。对 $U$ 中任意两事物标定其相似系数 $r_{ij}$

收稿日期:2013-01-12

基金项目:辽宁省教育厅科研项目(L2012111);中国煤炭工业协会科学技术研究指导性计划项目(MTKJ2012-319)

作者简介:王锦力(1977-),男,辽宁锦州人,讲师,博士生,主要从事采矿工程及环境工程相关的研究。

$\in [0,1]$ ,其大小表征两事物彼此相似程度。 $r_{ij}$  的全部构成一个模糊相似矩阵  $R = (r_{ij})_n$ ,其步骤如下:

(1) 建立聚类样本和聚类指标。设论域  $U = \{X_1, X_2, \dots, X_n\}$ ,为被分类的对象,每个对象又由  $m$  个指标表示其性状:  $X_i = \{x_{i1}, x_{i2}, \dots, x_{ij}, \dots, x_{im}\}$ , ( $i = 1, 2, \dots, n$ )。

(2) 数据的标准化处理。根据模糊矩阵的要求,将数据压缩到区间  $[0,1]$  上。通常需作如下变换:

$$x'_{ik} = \frac{x_{ik} - \bar{x}_k}{S_k} \quad (i = 1, 2, \dots, n) \quad (k = 1, 2, \dots, m)$$

式中:  $\bar{x}_k = \frac{1}{n} \sum_{i=1}^n x_{ik}$ ,  $S_k = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{ik} - \bar{x}_k)^2}$

再采用极差变换,使得  $x'_{ik}$  在区间  $[0,1]$  上

$$x''_{ik} = \frac{x'_{ik} - x'_{kmin}}{x'_{kmax} - x'_{kmin}}$$

(3) 建立模糊等价矩阵。聚类分析的依据是等价关系,故模糊等价矩阵除了应满足自反性、对称性外,还要满足传递性,即:  $R \cdot R = R$ 。根据标定所得的模糊矩阵,只是一个模糊相似矩阵  $R$ ,不一定具有传递性,可采用平方传递闭包法将  $R$  改造成模糊等价矩阵  $R^*$ ,使其具有传递性。

平方法是作模糊等价矩阵  $R$  的合成运算:  $R \rightarrow R^2 \rightarrow R^4 \dots \rightarrow R^{2^k}$ ,当  $R^{2^k} = R^k$  时,则  $R^k$  便是模糊等价矩阵  $R^*$ 。其中,合成运算:  $R \cdot R = [r_{ij}]_{m \times n}$ ,

$$r_{ij} = \left| \sum_{k=1}^m x_{ik} x_{jk} \right| / \sqrt{\left( \sum_{k=1}^m x_{ik} \right)^2 \cdot \left( \sum_{k=1}^m x_{jk} \right)^2}$$

(4) 建立模糊等价矩阵。采用平方传递闭包法将  $R$  改造成模糊等价矩阵  $R^*$ ,使其具有传递性:  $R \rightarrow R^2 \rightarrow R^4 \dots \rightarrow R^{2^k}$ ,当  $R^{2^k} = R^k$  时,则  $R^k$  便是模糊等价矩阵  $R^*$ ,再选用不同的  $\lambda$  截集对模糊等价矩阵进行分割。

### 3 神经网络原理

神经网络<sup>[12]</sup>是一个具有高度非线性的超大规模连续时间动力系统。是由大量的处理单元(神经元)广泛互连而形成的网络。它是在现代神经科学研究成果的基础上提出的,反映了脑功能的基本特征。

本文采用的是反向传播网络(BP 网络)。BP 网络是采用 Widrow - Hoff 学习算法和非线性可微转移函数的多层网络。一个典型的 BP 网络采用的是梯度下降算法,也就是 Widrow - Hoff 算法所规定的。Backpropagation 就是指的为非线性多层网络计算梯度的方法。现在有许多基本的优化算法,例如

变尺度算法和牛顿算法。一个经过训练的 BP 网络能够根据输入给出合适的结果,虽然这个输入并没有被训练过。这个特性使得 BP 网络很适合采用输入/目标对进行训练,而且并不需要把所有可能的输入/目标对都训练过。BP 神经网络由输入层、隐含层(可有多个)和输出层组成,每层可有若干个神经元。信号输出后,首先传到隐节点,经过作用函数在传播到输出层节点,经处理后输出结果。节点作用函数通常选用 S 型函数,即,

$$f(x) = 1/(1 + e^{-x})$$

式中:  $f(x)$  为作用函数,  $X$  为输入函数。

BP 网络的输入输出关系是一个高度非线性映射关系,如果输出层有  $n$  个神经元,输出层有  $m$  个神经元,则网络是从  $n$  维欧式空间到  $m$  维欧式空间的映射。通过调整网络连接权值,网络规模即可实现非线性分类、预测等问题,并可以任意精度逼近如何非线性函数。

### 4 算例分析

阜新地区新邱露天煤矿排土场淋溶水对附近于家沟地区地下水造成了严重污染,因此,对附近区域的 17 眼水井采样,并进行水质分析,其主要污染因子的水质分析结果见表 1。运用模糊聚类分析对该区域地下水环境质量进行总体评价,以确定污染范围。

选取样本 8 项污染指标:总硬度、 $SO_4^{2-}$ 、 $Cl^-$ 、溶解性总固体、高锰酸盐指数、 $NO_3^-$ -N、浊度、Na。

表 1 各监测点地下水水质分析结果

监测点	总硬度	$SO_4^{2-}$	$Cl^-$	溶解性总固体	高锰酸盐指数	$NO_3^-$ -N	浊度	Na
1	193.5	54.6	41.6	420	0.9	16.2	2.0	14.0
2	157.5	62.4	15.6	320	1.08	6.4	1.5	15.0
3	193.5	80.6	36.4	650	1.17	9.6	1.5	17.0
4	139.5	83.2	62.4	670	0.75	79.8	1.5	26.8
5	306	109.2	80.6	650	1.05	27.4	3.2	16.0
6	501	261.3	38.7	829	0.87	23.6	3.2	26.2
7	744	208.8	177.5	1551	0.93	33.10	2.0	39.3
8	1100	546.9	226.9	2111	1.48	31.0	2.0	57.5
9	929	408.7	109	1887	1.03	28.4	1.5	85.4
10	789	423.4	56.7	1068	1.34	18.9	3.7	58.3
11	1520	710	292	3036	1.34	37.4	2.8	40.8
12	5859	317.4	195.2	1889	1.20	34.9	1.5	26.8
13	649	223.2	157.2	1402	1.56	30.4	3.2	55.3
14	1478	435.2	608.7	2490	1.53	42.9	2.0	103.6
15	647	364.2	35.5	984	1.3	6.7	4.3	36.3
16	623	291.1	137.8	1348	1.06	27.9	2.0	55.1
17	608	383.5	17.5	927	1.18	7.3	3.8	31.4

根据模糊聚类的原理,对聚类指标进行标准化处理,得标准化矩阵  $R_1$ :经建立  $R^{2k} = R^k$  的关系矩阵经验证  $R^{32} = R^{16}$  得到模糊等价矩阵  $R^*$ 。当  $\lambda = 0.72$  时可将数据分为 4 类:1,2,3,4 为一类;5,7,9,10,11 为一类;6,8,14,16 为一类;12,13,15,17 为一类。

采用 MATLAB 神经网络工具箱建立三层 BP 神经网络。输入层节点 8 个,分别代表其 8 个指标。输出节点 1 个,代表标准化矩阵  $R_1$  每行的加权之和。对于 BP 网络的隐层节点属于不确定层,本文采用三层结构的 T3P 网络,即  $7 - Hn - 1$  结构,但隐含单元个数  $Hn$  仍未知。在实验过程中将隐含层

的神经元个数  $Hn$  作为一个参数试验,在此样本集下隐含层中用多少个神经元能使运算过程中达到精度最高。网络参数设置如下:net.trainParam.show = 100; net.trainParam.lr = 0.01; net.trainParam.epochs = 2000; net.trainParam.goal = 0.0002。而隐层对节点数的影响见表 2。

由计算结果可得当隐层节点数为 24 时均方误差最小,精度最高。

### 5 结 语

本文针对阜新新邱露天煤矿地下水质量评价问题以模糊聚类方法选取  $\lambda = 0.72$  为分类依据将阜

$$R_1 = \begin{bmatrix} 0.0094414 & 0 & 0.043889 & 0.036819 & 0.18519 & 0.13351 & 0.17857 & 0 \\ 0.0031471 & 0.011901 & 0 & 0 & 0.40741 & 0 & 0 & 0.011161 \\ 0.0094414 & 0.03967 & 0.035111 & 0.1215 & 0.51852 & 0.043597 & 0 & 0.033482 \\ 0 & 0.043637 & 0.079001 & 0.12887 & 0 & 1 & 0 & 0.14286 \\ 0.029111 & 0.083308 & 0.10972 & 0.1215 & 0.37037 & 0.28610 & 0.60714 & 0.022321 \\ 0.063205 & 0.31538 & 0.038994 & 0.18741 & 0.14815 & 0.23433 & 0.60714 & 0.13616 \\ 0.10569 & 0.23528 & 0.2733 & 0.45324 & 0.22222 & 0.36376 & 0.17857 & 0.28237 \\ 0.16793 & 0.75114 & 0.35668 & 0.65943 & 0.90123 & 0.33515 & 0.17857 & 0.48549 \\ 0.13804 & 0.54208 & 0.15766 & 0.57695 & 0.34568 & 0.29973 & 0 & 0.79688 \\ 0.11356 & 0.56271 & 0.069379 & 0.27541 & 0.72840 & 0.17030 & 0.78571 & 0.49442 \\ 0.24137 & 1 & 0.46658 & 1 & 0.72840 & 0.42234 & 0.46429 & 0.29911 \\ 1 & 0.40098 & 0.30317 & 0.57769 & 0.55556 & 0.38828 & 0 & 0.14286 \\ 0.089081 & 0.25725 & 0.23903 & 0.39838 & 1 & 0.32698 & 0.60714 & 0.46094 \\ 0.23402 & 0.58071 & 1 & 0.79897 & 0.96296 & 0.49728 & 0.17857 & 1 \\ 0.088732 & 0.47238 & 0.033592 & 0.24448 & 0.67901 & 0.00408 & 72 & 1 \\ 0.084535 & 0.36085 & 0.206280 & 0.3785 & 0.38272 & 0.29292 & 0.17857 & 0.45871 \\ 0.081913 & 0.50183 & 0.0032073 & 0.22349 & 0.53086 & 0.012262 & 0.82143 & 0.19420 \end{bmatrix}$$

$$R^* = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

表2 隐层节点数对精度影响

序号	隐层节点数	训练时间	训练次数	均方差 <i>MSE</i>
1	15	321.70	2876	5.19759
2	16	12.36	332	$4.21332 \times 10^{-7}$
3	17	1.58	562	0.00019952
4	18	45.06	587	0.000199191
5	19	8.98	325	$1.44438 \times 10^{-6}$
6	20	4.58	33	$1.28396 \times 10^{-5}$
7	21	2.36	46	$1.61245 \times 10^{-5}$
8	22	4.987	399	$9.84118 \times 10^{-7}$
9	23	3.928	522	$9.82206 \times 10^{-6}$
10	24	6.069781	925	$1.30926 \times 10^{-7}$
11	25	41.6670	4152	$3.83574 \times 10^{-5}$
12	27	27.59	3927	0.00019961
13	30	32.345	236	$9.31619 \times 10^{-6}$
14	35	6.998	56	$1.87623 \times 10^{-6}$
15	40	15.236	127	$1.65773 \times 10^{-6}$
16	45	25.693	366	0.00017357

新新邱露天煤矿地下水环境质量分为4类,在此基础上采用BP神经网络方法,以各个指标做为输入节点,以标准化矩阵  $R_1$  每行的加权之和为输出接点,对数据进行训练的和预测。得出主要结论如下:

(1)用模糊聚类法能解决矿区地下水污染程度等级分类问题,需要针对不用地区的特点,建立适合本地区的模型和和聚类界限及  $\lambda$  的选取问题,只有这样才能更加符合实际。

(2)模糊聚类分析法理论简单明,具体计算可在计算机上进行,能迅速而准确地做出稳定性判断, MATLAB 的BP神经网络工具箱在针对数据的训练和预测

(上接第137页)

### 3 结 语

(1)根据灰色预测理论,建立了龙江镉污染水质变化预测模型,由预测结果可知,灰色预测理论在本研究中有着很好的适用性。

(2)将实测数据与预测数据对比来看,经残差修正后的模型更为合理,预测结果更为可信。可以为相关部门提前采取应急措施提供参考,减小事故风险。

(3)由于影响水质变化的因素众多以及人为采取的措施,水中镉浓度不会无限制增加,仅用以往的数据建立的模型可能会失效,但这并不否定灰色预测方法在本研究中的有效性。因此灰色预测模型可用于对水质进行短期预测,而在长期预测中还应不断的向模型补充新信息,删除旧信息,建立新的预测模型。

运行速度快且精确,因此对选取的样本数和反映各个样本的因素是决定分类和预测准确的关键因素。

#### 参考文献:

- [1] 姜建军. 矿山环境五大问题亟待解决[J]. 地质与勘探, 2005(3):70.
- [2] Maree J P, Gerber A, McLaren A R, et al. A biological treatment of mining effluents[J]. Environ Tech Letters, 1987, 8:53-64.
- [3] 韦冠俊. 矿山环境工程[M]. 北京:冶金工业出版社, 2001:92-96.
- [4] Maree J P, Hill E. Biological removal of sulphate from industrial effluents and concomitant production of sulphur[J]. Sci Tech, 1989, 21:265-276.
- [5] 胡文容, 高延耀. 酸性矿井水的处理方法和利用途径[J]. 煤矿环境保护, 1994, 8(1):17-21.
- [6] 肖利萍, 张春婵, 潘纯林, 等. 生活污水为碳源处理硫酸盐矿山废水可行性实验[J]. 水资源与水工程学报, 2008, 19(6):76-78.
- [7] 刘志斌, 范军富, 丛鑫. 煤矸石山对地下水环境质量影响的分析研究[J]. 露天采煤技术, 2002(2):6-8+12.
- [8] 韦冠俊. 矿山环境保护[M]. 北京:冶金工业出版社, 1990.
- [9] 刘志斌. 露天煤矿排土场地下水环境质量影响的模糊综合评价[J]. 露天采煤技术, 2003(2):16-18.
- [10] 高新波. 模糊聚类分析及其应用[M]. 西安:电子科技大学出版社, 2004.
- [11] 袁曾任. 人工神经网络及其应用[M]. 北京:清华大学出版社, 1999.
- [12] 丛爽. 面向 MATLAB 工具箱的神经网络理论与应用[M]. 北京:中国科学技术大学出版社, 1998.

#### 参考文献:

- [1] 李丹, 张涛, 王祥三. 河流随机水质预测模型及其应用[J]. 中国农村水利水电, 2006(11):32-35.
- [2] 孙佳颖, 徐卫. 东流水质预测模型研究进展[J]. 山西建筑, 2010, 36(36):360-361.
- [3] 王伟, 程永清, 石砚秀. 灰色预测模型在渭河水环境信息系统中的应用[J]. 环境保护科学, 2007, 33(6):100-102.
- [4] 刘思峰, 党耀国, 等. 灰色系统理论及其应用[M]. 北京:科学出版社, 2010.
- [5] 李如忠, 汪家权, 钱家忠. 基于灰色动态模型群法的河流水质预测研究[J]. 水土保持通报, 2002, 22(4):10-12.
- [6] 杨建飞, 刘俊民, 陈琳. 基于灰色残差模型的灌区地下水最小埋深预测[J]. 人民黄河, 2011, 33(4):101-105.