



移动扫码阅读

梁戈龙,李继升,冯来宏,等.基于 PCA-ELM 的矿井突水水源识别方法研究[J].能源环境保护, 2022, 36 (6):132-137.
LIANG Gelong, LI Jisheng, FENG Laihong, et al. Research on the identification method of mine water intrush source based on PCA-ELM[J]. Energy Environmental Protection, 2022, 36(6):132-137.

基于 PCA-ELM 的矿井突水水源识别方法研究

梁戈龙¹,李继升²,冯来宏²,杜松^{3,*},李唐玥琪³

(1. 华能庆阳煤电有限责任公司核桃峪煤矿,甘肃 庆阳 745306;2. 华能煤炭技术研究有限公司,北京 101100;3. 中国煤炭地质总局勘查研究总院,北京 100039)

摘要:为提高矿井突水水源识别精度,基于 PCA-ELM 模型提出了一种的矿井突水水源判别方法。通过降维获得基本控制变量,将主成分分析(PCA)应用于 ELM 模型的训练集、验证集和测试集,使用数据样本验证模型。结果表明:通过模拟 PCA-ELM 模型在 10 s 内进行分类,PCA-ELM 模拟练习准确率达到 100%;BP 神经网络模拟准确率仅为 83.3%,远低于 PCA-ELM 模型的准确性。

关键词:煤矿突水;预测模型;预测方法;极限学习机;主成分分析法

中图分类号:X45;X523

文献标识码:A

文章编号:1006-8759(2022)06-0132-06

Research on the identification method of mine water intrush source based on PCA-ELM

LIANG Gelong¹, LI Jisheng², FENG Laihong², DU Song^{3,*}, LI Tangyueqi³

(1. Huaneng Qing Yang Coal & Electricity Co., Ltd., Hetaoyu Coal Mine, Qingyang 745306, China;
2. Huaneng Coal Technology Research Corporation Limited, Beijing 101100, China; 3. General Prospecting Institute China National Administration of Coal Geology, Beijing 100039, China)

Abstract: A method for identifying coal mine water intrush source based on PCA-ELM model was proposed to improve the identification accuracy of mine water intrush source. The basic control variables were obtained through dimension reduction. Principal component analysis (PCA) was applied to the training set, validation set and test set of the ELM model, and the data samples were used to verify the model. The results show that by simulating the classification process within 10 seconds by the PCA-ELM model, the accuracy of PCA-ELM simulation exercise reaches 100%. The accuracy of the BP neural network simulation only shows 83%, which is far lower than the accuracy of the PCA-ELM model.

Key Words: Coal mine water intrush; Prediction model; Prediction method; Extreme learning machine; Principal component analysis method

0 引言

矿井安全是一直困扰人类的一项大问题,煤

矿突水是我国华北型煤田的主要危险之一。及时准确地预测矿井突水水源对矿井生产具有重要意义,预测矿井突水的许多因素在很大程度上都具

收稿日期:2021-10-20;责任编辑:蒋雯婷

基金项目:华能集团总部科技项目(HNKJ20-H37);鄂尔多斯市科技合作重大专项(2021EEDSCXQDFZ006)

第一作者简介:梁戈龙(1988—),男,甘肃天水人,本科,工程师,主要从事煤田地质及其灾害防治、水体、含水层下煤层开采水害防治研究。
E-mail:330756484@qq.com

通讯作者简介:杜松(1987—),男,黑龙江齐齐哈尔人,博士研究生,副研究员,主要研究方向为煤矿矿井水及煤化工废水处理。

E-mail:du@cct.org.cn

有不确定性^[1]。在样品主要成分显而易见的情况下,贝叶斯定律适用于具有几个因素的样品的模糊综合评价。雷西玲使用 BP 算法和神经网络预测了山区的水源^[6],BP 神经网络使用逆变误差法来确定人类大脑中模拟信息处理过程的能力的值,可以识别大量用于突水水源的非线性关系的各种数据。但是,由于算法的缓慢收敛,以及初始值和设置对结果的影响更大,使得该方法的准确度易受影响。主成分分析(PCA)在一定程度上采用了降维思想,抵消了水样品之间复杂信息的影响,将大量指标转化为小而全面的指标的同时继续反映了原来的变量信息。

ELM 算法是由阎志刚等人在 2006 年开发的一种算法,该算法不需要在学习过程中调整网络的输入值,也不需要调整隐藏层的位移、综合性能等特征,可以有效地消除数据之间的关联,减少计算的复杂性。当网络引入变量时,通常包含相互关联信息的煤矿影响因素,重建样本训练可减少神经网络结构复杂性并提高收敛速度。使用 PCA 和 ELM 的组合方法来预测煤矿的突水水源,收集分析矿井的各充水含水层水化学历史数据,并根据 PCA-ELM 构建水源识别模型,为煤矿突水预测提供了新的思路和方法。

本文使用 PCA 分析矿井水样品中的水化学数据,以确定不同含水层中水样的主要控制因素,以便更准确地确定各充水含水层特征。在此基础上,利用 ELM 模拟水样本中的基本控制因素,提高 ELM 算法的运算速度和精确度,并为其他类似矿区相似矿井提供技术支持。

1 矿井常见突水水源的水化学基本特性

地表水、地下水、海水、老空区水是常见的威胁矿山安全生产的矿井突水水源种类。地表淡水和海水的水化学性质相差较为明显^[7];地下水由于其所在地层的地质条件、深度及补径排条件的不同,水化学性质有明显差异研究。灰岩裂隙含水层中,地下水中的侵蚀性 CO_2 会使石灰岩主要成分 CaCO_3 和 MgCO_3 溶滤,从而使地下水中出现较高浓度的 Ca^{2+} 、 Mg^{2+} 和 HCO_3^- 。碳酸盐及石膏地层是地下水中 Cl^- 、 SO_4^{2-} 、 Na^+ 和 K^+ 等离子的主要来源。在砂岩及火成岩风化物含水层中,由于长石等溶滤,地下水中 Na^+ 、 K^+ 、 HCO_3^- 等离子增加,pH 升高。煤系地层中地下水循环条件较差,煤系硫分及围岩中的硫酸盐等进入地下水中,使 SO_4^{2-}

含量增大,矿化度也增高。我国煤矿区地下水水质特征总体上受控于上述因素,因此把握不同含水层的基本属性,需根据水化学特征分析水文地质条件以水质特征模型判别矿井突水水源。老空区水由于开采扰动导致新的导水通道形成后,水动力条件发生了改变,使不同含水层中的水发生水-水混合反应,与原生地下水的差异也较明显。

2 矿井突水水源判别数学模型的建立

我国煤矿由于矿井突水遭受了大量财产损失和人员损失,因此预防和应对水灾的科学技术措施对快速正确定义矿井突水水源是非常关键的。由于各充水含水层的不同地质特征,不同来源的矿井突水水源的水化学特征差异明显^[8]。常用的模糊评价方法如灰色关联理论、支持向量机、BP 神经网络分析理论、贝尔斯、费舍尔等大都没有考虑影响指数之间的信息存储,导致判别结果的误差。为了消除这一影响,本文采用了一种主要的方法来分析矿井中有争议的水源模型中的因素指标。

2.1 主成分分析法

主成分分析(Principal Component Analysis, PCA)是一种多元统计数学方法,将高阶问题转化为低量度问题,简化计算的过程,尽可能避免试验的误差结果,避免无用重复的计算。通过主成分分析的方法,其主元件之间的分析变量不仅小,而且十分有利于分析,使之创建的指标之间进行相互关联,有关变量参数的信息占原始数据样本的大部分。

当抽样容量非常大时,综合评估就更加困难,因为在使用初级组件时,处理结果的最佳方式是用较低的初级组件而不是较低的初级信息,即损失最少的信息。为了解决这个矛盾,必须研究确定关键成分数量的指导原则。

最大数据突变的原则是主组件必须采取最大的数据偏转方向,以便主组件能够尽可能多地表达源信息。

最小乘法原理分析的主要成分是在新空间中引入原始数据。从几何角度看,信息分析在核心成分中所造成的最小损害必须在正方形中实现,并在新超平面中达到原样品投影的最小距离。

群体相似性改变了一个基本原则,在这个原则中,为了分析信息造成的最小损失,需要对数据之间的相似之处进行最小的改变。对角线抽样系

统的最佳综合表达原理是使主组件尽可能接近初始变量,而不损失源。

2.2 理论与算法

2.2.1 PCA 法

PCA 将原始指标组合成一组新的相互关联的综合指标,同时提取了尽可能多的反映原有指标的信件,反映了尽可能多的初始指标。假设在评估一个物体时,源信息是集成的,且有选择性的,每个样本都有一个变量 P ,这使它成为一个 $n \times p$ 阶的样本数据矩阵,记为式(1)。

$$X_i = [x_{i1}, x_{i2}, \dots, x_{ip}]^T \quad (1)$$

贡献率是指出,从任何核心组件中提取的信息占总数的很大一部分,投资水平越高,相关主组件反映总信息的可能性就越大。基本成分 k 的贡献率如式(2)。

$$M = \lambda_k \left(\sum_{j=1}^p \lambda_j \right)^{-1} \quad (2)$$

在通常的情况下,分析前 m 个主成分的贡献率,详见式(3)。

$$Z = \sum_{j=1}^m \lambda_j \left(\sum_{j=1}^p \lambda_j \right)^{-1} \quad (3)$$

如果前面的数据大于 85%,用前 m 个主成分的组合作式(4)。

$$P = Y_1, Y_2, \dots, Y_m \quad (4)$$

因此,与其改变原始数据 X ,不如减少数据量,以便将原始变量转化为不相关的变量,而失去尽可能少的源信息方向。

2.2.2 ELM

ELM(Extreme Learning Machine, ELM) 极限学习机是一类基于前馈神经网络(Feedforward Neuron Network, FNN)构建的机器学习系统或方法。通过 PCA 得到的数据样本,经过 ELM 模型进行学习分类。对于 N 个不同样本 (x_i, t_i) , $\bar{N}$ 个隐藏节点,激励函数的统一模型如式(5)。

$$\sum_{i=1}^{\bar{N}} \beta_i g(x_j) = \sum_{i=1}^{\bar{N}} \sum_{i=1}^{\bar{N}} \beta_i g(w_i \cdot x_i + b_i) = o_j, (j = 1, 2, \dots, N) \quad (5)$$

相比于其他分类模型,ELM 具有良好的分类效果,但在面对高维数据时建模时间较长。用主成分分析法对高维数据集进行特征提取,再将提取后的数据集输入 ELM 模型进行建模和调参,将有效地提高 ELM 收敛速度和建模效率。

2.3 基于 PCA-ELM 的煤矿突水预测

利用 PCA-ELM 的分析方法预测煤矿突水的

流程如图 1 所示。

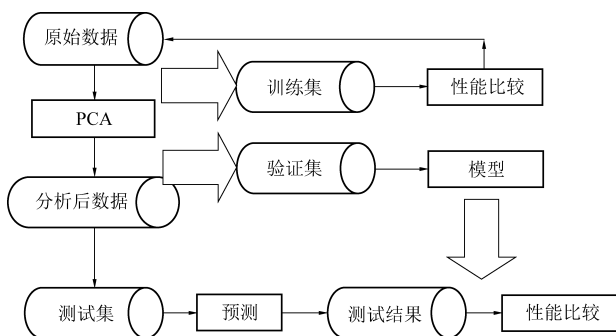


图 1 基于 PCA-ELM 的煤矿突水预测流程

Fig.1 Prediction process of coal mine water intrush based on PCA-ELM

首先利用 PCA 法分析了影响煤矿突水的许多影响因素,得到了控制因素,决定了各充水含水层的水化学特征,然后将只包含基本管理要素的抽样数据分解为训练、验证和测试。最后,通过与 ELM 算法相匹配,评价 PCA-ELM 用来预测突水的优势。

2.4 煤矿突水预测建模

基于 PCA-ELM 的煤矿突水预测模型构建步骤如下:

(1)利用 PCA 的方法,对煤矿突水主控影响因素进行筛选,并检查正常使用煤炭的采样基质,并在 x 矩阵中标准化处理,根据共变矩阵 R 确定主组件的数量、主组件输入系数和总贡献率;通过计算主成分贡献率及累积贡献率,基于凯塞标准选择累积贡献率最高的 3 个主成分。为了使 3 个主成分的差异性最大化,采用最大方差旋转法对主成分轴进行了适当旋转,使每个主成分具有最高荷载的变量数最少,从而简化对主成分的解释,更好地揭示水化学指标所表达的信息。通过基本组件的数量构建 ELM 模型的教学样本。

(2)建立 ELM 网络模型。在处理了经过分析的各充水含水层水化学历史数据后,收集了数据样本,以便重新设计和创建 ELM 网络模型。首先,需要一个样品来练习,详见式(6)。

$$\{x_i, t_i\}, i = 1, 2, \dots, N \quad (6)$$

初始隐藏层节点数设为 i 。

(3)利用统计数据验证煤矿突水预测模型,如果获得预测结果与其他算法运算结果相比具有明显的优势,可以开始重建模型和算法基本成分,结果将更为可取。

3 实验评估

3.1 样本数据选择

本文选择了研究中国典型华北型煤田的矿井水数据,根据这些数据运用 PCA-ELM 和 ELM 预测矿井突水状况。压力影响作为煤矿突水输入参数。参数定义原则:不能量化的数据在两位数模型中量化,定量的数据定量表示,如果没有构建,则断层为 1 或 0。由于气象因素和水化学测试数据类型不同,在构建预测模型之前,必须将影响天气的因素综合起来,并将所有实验数据系统地输入进去。

3.2 参数选择

在建模过程中,当隐藏节点的数量不同时,模型的性能就会大不相同,图 2 提供了具体的分析结果。

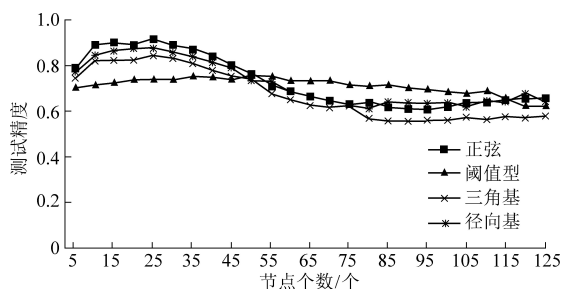


图 2 激励函数对应的测试精度

Fig.2 Test precision corresponding to the excitation function

表 1 ELM 与 PCA-ELM 实验结果比较

Table 1 Comparison of experimental results between ELM and PCA-ELM

算法	时间/ms		精度/%	
	训练集	测试集	训练集	测试集
ELM	3.2	2.5	83.23	76.71
PCA-ELM	2.4	1.2	85.51	90.80

表 1 显示,PCA-ELM 分析方法不同于 ELM 方法,其输入数据分析的主要成分,减少相应的试验变量,降低多余部分数据产生的误差,提高模型预测精度,并加快模型计算速度,比较算法和计算模型,PCA-ELM 方法预测煤矿的突水水源精度更高。

3.4 结果分析

为了解决水样本中的多样和水样本中更小信息的影响,分析了表 2 的主要成分,表 2 显示了系数矩阵,其中不同的成分解释了表 2 中的偏差。根据表 2 可知,前 3 个组件累积解释 91.936%,故认为在各离子中,前 3 个阳离子作为水样主要组件来模拟训练样本。

3.5 MATLAB 仿真训练

3 种主成分离子的含量值为输入值,3 种水的

由图 2 可知,在正弦、阈值型、三角基、径向基 4 个激励函数中,当激励函数为正弦且节点个数为 25 时,测试精度最高,往后随着节点个数增加,测试精度逐渐降低。因此,对比选择正弦函数为算法相对应的激励函数,隐藏节点个数为 25。

3.3 对比试验

PCA 方法用于减少原始数据,PCA-ELM 测试和 ELM 的测试进度比较结果如图 3 所示。

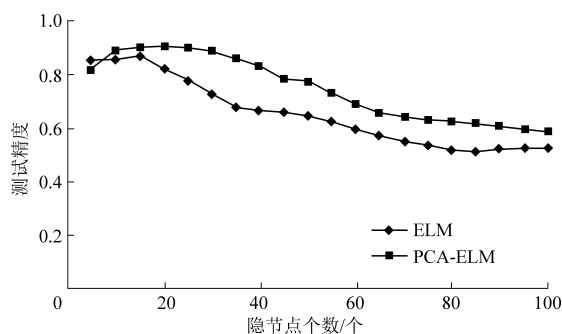


图 3 ELM 与 PCA-ELM 的测试精度比较

Fig.3 Comparison of test accuracies between ELM and PCA-ELM

据图 3 所知,ELM 预测的准确性总是比 PCA-ELM 低。分析非主要成分的测试集中冗余影响因素数据众多,而只有主控因素的碳元素被控制。选择的最佳参数分别用于 ELM 和 PCA-ELM 的准备和测试,测试结果见表 1。

类型为输出值,即网络设置的输出节点数为 3。为了使判别结果尽可能接近所有的学习样本,研究样本的数量等于学习样本的数量,即 24;网络的输出状态为 3,相当于剩下的 3 个层。在本文的 ELM 模型中,分析 S 型函数(Sigmoid),其为隐含层神经元的传递函数,而修正的双方都使用 ELM 算法来指导样品,确定 ELM 1 的类型,即分类识别。在 ELM 模型中,使用 MATLAB 软件模拟测试样品,该样本应在 10 s 内分类,最后,其分析结果见表 3。

对水样进行分类的结果显示,检测结果 100% 相同,表明 PCA-ELM 混合的矿井突水水源方法分类识别性能好,具有一定价值。

表 2 各成分解释方差率

Table 2 Explanatory variance rate for each component

主成分因子	特征值	贡献率/%	累计贡献率/%
Na ⁺	3.121	53.590	53.590
Ca ²⁺	1.531	25.410	79.000
Mg ²⁺	0.780	12.936	91.936
Cl ⁻	0.450	7.470	99.406
SO ₄ ²⁻	0.031	0.446	99.852
HCO ₃ ⁻	0.016	0.148	100.000

表 3 ELM 识别结果

Table 3 ELM identification results

样本编号	实际水样	目标输出	实际输出	识别结果
1	奥灰水	1	1	正确
2	奥灰水	1	1	正确
3	12 煤	2	2	正确
4	12 煤	2	2	正确
5	13 煤	3	3	正确
6	13 煤	3	3	正确

3.6 结果对比

与此同时,使用模拟 BP 神经网络将输入层设置为 3,对应于 Na⁺、Ca²⁺、Mg²⁺,因此,设置 3 号输出神经元的网络结构为“3-7-3”型,logsig 训练

算法用于传输,trainlm 传输功能,结果如表 4 所示。由表 3 和表 4 可知,ELM 网络更精确,学习结果比 BP 网络更准确。

表 4 BP 网络识别结果

Table 4 BP network identification results

样本编号	目标输出	实际输出	识别结果
1	(1 0 0)	(0.920 1 0.068 1 0.168 4)	正确
2	(1 0 0)	(0.993 9 0.008 9 0.011 1)	正确
3	(0 1 0)	(0.146 8 0.960 0 0.023 0)	正确
4	(0 1 0)	(0.524 0 0.157 0 0.407 0)	错误
5	(0 0 1)	(0.160 1 0.018 6 0.916 2)	正确
6	(0 0 1)	(0.171 8 0.138 9 0.911 1)	正确

BP 神经网络随机链接权值和阈值,在本研究中对随机抽样样本网络参数输出的权值造成误差甚至引发不稳定模型的情况。

4 结 论

对于各含水层水化学组分差异性不大的矿井突水水源,使用常见的水化学分类法和常规统计方法并不能很好地判别各含水层水化学组分差异性不大的矿井突水水源,其误差较大,通过 PCA-ELM 模型与其它神经网络模型相比,模型的准确率达到 100%,其判别效果优于 BP 神经网络模型(83.3%),为矿井突水水源的判别提供了一种新方法。

参考文献

- [1] 李培,陈颖,马小平,等.基于 PCA-ELM 的煤矿突水预测方法研究[J].工矿自动化,2013,39(9):46-50.
LI Pei, CHEN Ying, MA Xiaoping, et al. Research of coal mine water burst forecasting method based on PCA-ELM [J]. Industry and Mine Automation, 2013, 39(9): 46-50.
- [2] 孙文洁,杨恒,李祥,等. PCA 与 ELM 模型相结合的矿井突水水源快速识别方法研究[J].煤炭工程,2020,52(1):111-115.
SUN Wenjie, YANG Heng, LI Xiang, et al. Research on rapid recognition method of mine water inrush source based on PCA and ELM model [J]. Coal Engineering, 2020, 52(1): 111-115.

- [3] 张璐平. 基于 CNN-ELM 的人脸表情识别方法研究 [D]. 长沙: 湖南师范大学, 2019: 1-8.
- [4] 张海英, 闫德勤, 杨伊, 等. 基于流形学习的极端学习机及其在人脸识别中的应用 [J]. 计算机测量与控制, 2016, 24 (12): 158-161.
- ZHANG Haiying, YAN Deqin, YANG Yi, et al. Using manifold learning extreme learning machine for face recognition [J]. Computer Measurement & Control, 2016, 24 (12): 158-161.
- [5] 吴景龙, 王巨, 王春山. 人工神经网络系统在预防煤矿突水中应用 [J]. 煤矿安全, 2008 (4): 55-57.
- WU Jinglong, WANG Ju, WANG Chunshan. Application of artificial neural network system in preventing water inrush in coal mine [J]. Coal Mine Safety, 2008 (4): 55-57.
- [6] 雷西玲, 张景, 谢天保. 基于遗传神经网络的煤矿突水预测 [J]. 计算机工程, 2003 (11): 132-133+197.
- LEI Xiling, ZHANG Jing, XIE Tianbao. Forecast for water-inrush from coal floor based on genetic neural networks [J]. Computer Engineering, 2003 (11): 132-133+197.
- [7] 鲁金涛. 基于 PCA-MSA 的矿井突水水源判别算法研究 [D]. 长沙: 中南大学, 2013: 1-7.
- [8] 姜子豪, 胡友彪, 琚棋定, 等. 矿井突水水源判别方法 [J]. 工矿自动化, 2020, 46 (4): 28-33.
- JIANG Zihao, HU Youbiao, JU Qiding, et al. A discrimination method of mine water inrush source [J]. Industry and Mine Automation, 2020, 46 (4): 28-33.
- [9] 张菊连, 沈明荣. 基于逐步判别分析的砂土液化预测研究 [J]. 岩土力学, 2010, 31 (S1): 298-302.
- ZHANG Julian, SHEN Mingrong. Sand liquefaction prediction based on stepwise discriminant analysis [J]. Rock and Soil Mechanics, 2010, 31 (S1): 298-302.
- [10] 赵伟, 刘启蒙, 柴辉婵, 等. 基于 Fisher 判别法和质心距理论的突水水源识别 [J]. 科学技术与工程, 2020, 20 (9): 3552-3556.
- ZHAO Wei, LIU Qimeng, CHAI Huichan, et al. Identification of water inrush source based on fisher discriminant method and centroid distance theory [J]. Science Technology and Engineering, 2020, 20 (9): 3552-3556.
- [11] 李小明, 刘德民, 连会青, 等. 赵庄矿峰组碳酸岩层隔水性能研究及应用 [J]. 煤炭工程, 2019, 51 (10): 102-107.
- LI Xiaoming, LIU Deming, LIAN Huiqing, et al. Research on water-resisting property of carbonate strata in Fengfeng Formation in Zhaozhuang coal mine and its application [J]. Coal Engineering, 2019, 51 (10): 102-107.
- [12] 尹尚先, 王屹, 尹慧超, 等. 深部底板奥灰薄灰突水机理及全时空防治技术 [J]. 煤炭学报, 2020, 45 (5): 1855-1864.
- YIN Shangxian, WANG Yi, YIN Huichao, et al. Mechanism and full-time-space prevention and control technology of water inrush from Ordovician and thin limestone in deep mines [J]. Journal of China Coal Society, 2020, 45 (5): 1855-1864.