Online Discrimination System for Mine Water Inrush Source Based on PCA and BP Neural Network

Yong Yang¹, ²
¹School of Resources and Geosciences, China University of Mining and Technology, Xuzhou, China  
²School of Information and Electrical Engineering, Xuzhou College of Industrial Technology, Xuzhou, China

Jianhua Yue*  
School of Resources and Geosciences, China University of Mining and Technology, Xuzhou, China  
*Corresponding author(E-mail: yuejh@cumt.edu.cn)

Xinghua Li  
Chengdu BOE Optoelectronics Technology Co., Ltd., Chengdu, China

Damao Wang  
School of Information Engineering, Southeast University, Nanjing, China

Abstract

Water inrush is one of mine geological disasters to threaten safety mining in China. Water inrush sources recognition is an effective method to predict mine water inrush disaster in time. Compared with overlong processing time using conventional Hydro-chemistry methodology, the paper proposed in-situ mine water sources discrimination model using principal component analysis (PCA) and Back Propagation (BP) neural network based on in-situ water monitoring system. The system is constructed by sensor nodes, an information collector and a ground monitoring center. The measured data were collected from in-situ sensors such as fluorescence spectra, PH value, conductivity, Ca²⁺, Na⁺, HCO₃⁻ and Cl⁻ ions transducers in different water layers of LiJiaZui Coal Mine in Huainan. PCA is utilized to eliminate correlation and BP neural network is used to recognize mine water sources. The results show that the proposed model achieves 91% accuracy to recognize water sources in mine. Thus, the proposed Model is a rapid and an effective way to recognize mine water sources and further water inrush disaster prediction.

Key words: Water Inrush Source, Fluorescence Spectra, Principal Component Analysis, BP Neural Network.

1. Introduction

Coal is a primary energy source in China. Coal accounts for 60 % of China’s primary energy and it will continue to be more than 50% until 2050. [1] Under this situation, the Chinese government pays much attention to the healthy and sustainable development of its coal industry, especially the research related to the safety of coal production. In term of economic losses, the tremendous mine accident caused by mine water inrush, which is second only to the gas explosion, is at the top of the various types of coal mine disasters in China. According to the statistics, there were 26 mine water inrush disasters leading to over 10 death during the period of “11th five-year plan” in China, that meant over 5 times in a year on average and 506 death totally. Although many scholars had got gratifying achievement in the prevention and control of coal mine inrush disaster, the overall situation is still grim [2]. At present, the problem of goaf water becomes more serious with the increasing of mining depth in China. Mine hydro-geological condition is complex, and the water flowing fractured zone is not clear [3]. All of these should be responsible for the ineffective monitoring in the water flowing. In addition to the goaf water, surface water and the limestone water are likely to invade mine[4]. Therefore, it is the precondition of mine inrush prevention to determine the water source and accurately identify the source of water inrush with an appropriate method.

The distinguishing of mine inrush water source should be combined with the hydro-geological data of mine and the geological construction condition, also with the water temperature[5], the water level[6] method and the water chemical data[7] comprehensively. At present, the water chemistry data analysis is the most accurate and commonly used method of analysis, which had gradually developed into multivariate statistical analysis. Ca²⁺, Mg²⁺, K⁺⁺Na⁺, HCO₃⁻, SO₄²⁻, Cl⁻ were selected as the feature vectors based on the importance of 6 water hydro-chemical element factors, a lot of ideas and methods have been published in recent years. Principal component analysis(PCA) reduces variables into a smaller number of uncorrelated components without losing much information.[8]Fuzzy cluster analysis of the hydrogeochemical features conducted the judgment of mine water inrush source, the analysis result show that the method was not only predict the inrush sources, but also imply different aquifer related to each other or not [9-11]. Because groundwater system is complex, consisting
of a number of aquifers with varied geochemical characteristics, single method cannot solve the problem of all regions [11, 13]. Combining a suite of statistical techniques, such as distance discrimination[14], method of gray correlation[15], Fisher discrimination method[16], fuzzy comprehensive evaluation method[17-18], artificial neural network[19], support vector machine[20], to analyze hydro-chemical parameters has been shown useful in classifying groundwater of varied geochemical features. Laser induced fluorescence (LIF) technology techniques[21-22] had been useful in rapid classifying groundwater.

However, when the water inflow in the stope increases suddenly, people usually discriminate the water source through the water test. Because the water chemical analysis shall be carried out in the laboratory, the long cycle of water chemical tests which caused missing the best time to dispose of the water-inrush event and lead to major accidents happens occasionally [21]. To overcome the problem that traditional chemical analysis method cannot monitor the changes of mine water source real-time, this paper first put forward a kind of water source identification and analysis system which uses fluorescence spectroscopy and PH value, conductivity and ions transducers to distinguish the mine water source real-time by the method of principal component analysis and BP neural network.

2. Theory and Algorithms

2.1 Principal Component Analysis

PCA method is integrating P observables to new P observables(Integrated variable), as follows:

\[ Y_i = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{ip}x_p \]

\[ Y_2 = a_{21}x_1 + a_{22}x_2 + \cdots + a_{2p}x_p \]

\[ \vdots \]

\[ Y_p = a_{p1}x_1 + a_{p2}x_2 + \cdots + a_{pp}x_p \]

(1)

in the formula, x is the original data, which meeting the conditions: Yi and Yj are unrelated(i≠j, i, j=1,2,...,p), the variance of Y1 is greater than Y2, the variance of Y2 is greater than Y3, so on and so forth. The selecting of the number of principal components K, is mainly decided by the cumulative contribution ratio of the principal component, namely general requiring cumulative contribution achieving more than 85%[23-25], so as to ensure integrated variable including the vast majority information of original variable.

The general steps for solving the principal components as follows:

The original data were normalized:

\[ x_i = \frac{x_i - \bar{x}_i}{\sqrt{\text{var}(x_i)}} \quad (j = 1, 2, \ldots, n; j = 1, 2, \ldots, p) \]

(2)

in the formula, \[ \bar{x}_i = \frac{1}{n} \sum_{i=1}^{n} x_i \]

\[ \text{var}(x_i) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x}_i)^2 \quad (j = 1, 2, \ldots, p) \]

Calculating sample correlation coefficient matrix:

\[ R = \begin{pmatrix} r_{11} & \cdots & r_{1p} \\ \vdots & \ddots & \vdots \\ r_{p1} & \cdots & r_{pp} \end{pmatrix} \]

Assuming that the standardized original data still indicated by X, so the correlation of the standardized data is:

\[ r_{ij} = \frac{1}{n-1} \sum_{i=1}^{n} x_i x_j \quad (i, j = 1, 2, \ldots, p) \]

(3)

Solving the eigenvalues(\(\lambda_1, \lambda_2, \ldots, \lambda_p\)) and feature vector (a=(ai1, ai2, ..., aip), i=1, 2, ..., p) of the correlation coefficient matrix R by Jacobi method.

Extracting the principal component, and writing its expression:

P principal components can be gained by principal component analysis, but, since the variance of each principal component is diminishing, the included information is also diminishing. So, in practice, it is generally not selecting P principal components, but selecting the first K principal components according to the size of each principal components cumulative contribution ratio. Here the contribution ratio refers to a principal component’s variance proportion in the total variance, which is also the proportion of an eigenvalue in the total eigenvalues, namely:

\[ d_i = \frac{\lambda_i}{\sum_{i=1}^{p} \lambda_i} \]

(4)

The larger contribution indicates that the principal component contains the more information of the original variable. The selection of principal components number K, mainly based on the cumulative contribution ratio of
the principal component, that is to say, general requiring cumulative contribution achieving more than 85%, so as to ensure integrated variable including the vast majority information of original variable.

Calculating the principal component scores:

The standardized sample data were entered on behalf of constituents expression, each principal component score can be gained.

2.2 BP Neural Network

BP Neural Network is a multi-layer feedforward neural network, the adjustment rules of network weights use error back propagation algorithm and BP algorithm[26-28]. The basic idea is that the learning process consists of the forward signal propagation and the reverse errors propagation. When in the forward propagation, the input samples pass from the input layer, transmute to the output layer after each hidden layer processing. If the actual output does not match the desired input, then transfers to the reverse propagation stage, the repeated process is constantly adjusting the network weights.

BP neural network acquire the expected value of output through the error back propagation from output layer to hidden layer, And the gradient descent method via calculate the network weights and adjust the connection weights to minimize the output error. The definition of neural network error function as:

$$E = \frac{1}{2} \sum (T_k - O_k)^2$$

(5)

Tk and Ok are the target value and the output value. The gradient descent method changes the weight via gradient error:

$$\Delta W_{ij} = -\eta \frac{\partial E}{\partial W_{ij}}$$

(6)

$\eta$ is Learning rate, $\frac{\partial E}{\partial W_{ij}}$ is expressed by the following equation:

$$\frac{\partial E}{\partial W_{ij}} = -\delta^o_j A_{i-1}^m$$

(7)

Gradient error can be obtained:

$$\Delta W_{ij} = -\eta \delta^o_j A_{i-1}^m$$

(8)

Here, the output layer $A^o$ relate to the weights of the connection $W_{ij}$, $\delta^o_j$ is error signal, It can calculating whether the j is the output layer of neurons by calculated. If j is the neurons of output layer, then

$$\delta_j = (T_j - Y_j) Y_j (1 - Y_j)$$

(9)

If j is the neurons of the hidden layer

$$\delta_j = \sum \delta_i A_{i-1} (W - h)_{ij} H_{ij} (1 - H_{ij})$$

(10)

Here $H_{ij}$ is the value of hidden layer. Finally, the weights of connected neurons can be expressed as:

$$W_{ij}^m = W_{ij}^{m-1} + \Delta W_{ij} = W_{ij}^{m-1} + \eta \delta^o_j A_{i-1}^m$$

(11)

The input signal is transmitted from input layer to output layer through hidden layer, and generates an output signal at the output terminal, this is the forward propagation of the working signal. During the forward transmission of signal, the weights of network are fixed, and the state of each neurons layer is only affecting the state of the next layer of neurons. If the expected output cannot be obtained at the output layer, then switch to the reverse propagation of the error signal.

The difference between the actual output and the expected output of the network is the error signal, the error signal starts the layer by layer forward propagation from the output side to input side, which is the reverse propagation of the error signal. In the back propagation process of error signal, the weights of network are adjusted by the error feedback. By the constant revision of the weight, the actual output of the network is closer to the expected output.

3. Water Inrush Source Identification Model and Application

The water chemical elements in each aquifer are numerous, the conventional water sources identification mainly use Ca$^{2+}$, Mg$^{2+}$, K$^{++}$Na$^+$, HCO$_3^-$, SO$_4^{2-}$, Cl$^-$ ions as the water source index of the mine water inrush identification model (unit is mg/L) [29-30]. Given the restrictions on the type and cost of online monitoring sensor in market, this paper adopts Ca$^{2+}$(X1), Na$^+$ (X2), HCO$_3^-$ (X3), Cl$^-$ (X4) ions transducers and pH (X5), electrical conductivity (X6), fluorescence spectra (X7) as the water indicators of predictor mine water inrush discriminant analysis model.
3.1 Online Monitoring System for Water Inrush Source

The system is mainly composed of sensor nodes, information collector and ground monitoring center. Its architecture is shown in Figure 1. The system adopts star network, which is convenient for maintenance and management. The sensor nodes are installed in monitoring points and transmit the collected data to information collector. The data is converted to Ethernet format by serial port server, then the down-hole Ethernet ring network transmit the sensor data to the ground monitoring center with RJ interfaces, the center will resolve, display and process the received data, achieve real-time monitoring of the mine water parameters. Ground monitoring center can also upload the data to the remote diagnostic center to further analysis through the Internet network, diagnostic center will identify the mine water source by discriminant model, then release warning through messages or network.

Hardware system includes sensor nodes and CPU module. Peripheral mount with Ca2+, Na+, HCO3-, Cl− ions, PH value, conductivity and fluorescence spectrum sensor modules. Mine water parameter information are collected with these hardware and communication module is used for upload data and publish information. Parameters of water are sent to the host Computer by serial remote server. Sensor module consists of sensor and tuning circuit. PH composite electrode, conductivity sensor and ions transducer are selected as sensor electrode. The selected sensor has a high cost, low power consumption and high reliability advantages. PH sensor measuring range is 0-14.00, conductivity sensor measuring range is 5000μs/cm and the ions transducer measuring range is 0-2000mg/L, figure 2 is a pictorial diagram of the sensor.

![ion transducer](image1)

**Figure 2.** ions transducer

![fluorescence spectrum sensor](image2)

**Figure 3.** Fluorescence spectrum sensor

Different substances in the water have their own characteristic fluorescence spectrum and the difference of concentration also can cause the change of characteristic fluorescence spectrum. Therefore, the characteristic fluorescence spectra of the aquifer in the coal mine are the whole performance of fluorescence spectra of all matter in water. It reflects the spectral performance difference of material composition and concentration of...
water body, which is not a fluorescent spectral of one or several substances. The difference of water composition is reflected in the difference of fluorescence spectrum. The fluorescence spectra of corresponding water bodies of the aquifer also show a specific pattern. According to the difference of fluorescence spectrum, online water source identification of gushing water can be carried out by pattern recognition of spectrum. [31] And then achieve the purpose of water inrush warning. The instrument which is used in the experiment is LIFS-405 laser induced fluorescence system (made by Guangdong flag Electronics Co., Ltd.). The laser incident wavelength is fixed-frequency with 405nm and the incident laser power is 120mW. The range of fluorescence spectral which can be detected is 400-800nm and the step size is set to 0.5nm. The spectral scan time is set to 1s/1000nm. Considering the major purpose of this experiment is engineering application, we use immersion type fluorescent probe FPB-405-V3 (made by Guang Dong ke si kai Co.) to inspire the tested water, instead of the conventional isolation type laser excitation. Figure 3 is the equipment picture.

3.2 Water Sample Collection

The sample library used by this paper is the coal mine water of Huainan Mining Group in LiJiaZui Coal Mine. There were 41 samples were obtained during the month from May 2015 to June. Samples were derived from water-filled aquifer, which is main source, coal measures sandstone fissuring aquifer (sandstone water I), carboniferous Taiyuan Formation line karst aquifer (The limestone water II), and storage water in goaf (The goaf water III). Samples were placed in polypropylene sample vials which were washed by distilled water and 10% HCl. The fluorescence spectrum, PH, conductivity and chemical characteristics of water ions were measured immediately while those samples were brought back to lab. 30 samples were selected by comprehensive parameters from water samples as the training sample, which contains 9 sandstone water samples, 15 limestone water samples and 6 goaf water samples. The actual data of the sample is shown in Table 2.

3.3 Building the Model for Analyzing and Discriminating the Water Inrush Source

In the experiment, 30 samples are processed to obtain spectral data by fluorescence excitation. In order to reduce the background light and human impact, the measure is carried out in dark, and each sample is measured 10 times which arithmetic mean is taken as the final spectral data, as shown in figure 4. In consideration of the influence of random noise and baseline drift in spectrum, uniform sample, light scattering and so on, AVANTES Software 7.2 of spectrometer is used to do some basic processing before the spectral pre-processing, in order to eliminate the effect of instrument and environmental factors on the spectral data [32-33]. As is shown in figure 4, the classification of spectral data of 3 types of water samples is obvious. The goaf water will be formed if the goaf area generated by earlier mining activity is filled with surface water and ground water, which has nothing to do with the geologic age. It could lead to burst accidents of water disasters when later mining activities touch the edge of the water. Its components are relatively complex and the water commonly is acidic. Its impurity content is also higher than other samples, for example, it has higher turbidity corresponding to the larger amplitude in spectra. The geological ages of limestone water and sandstone water are both late Paleozoic, so their compositions are similar.

![Figure 4. Original fluorescence spectra of water samples](image)

Subsequently, standardized water sample data need to be processed with PCA, which is carried out in Matlab for dealing with spectral data. Statistical analysis of cumulative reliabilities of extracting principal components shows the former three cumulative reliabilities of principal components reaches to 98.109% among all the components. So the former three principal components could represent the spectral data. The reliabilities of principal components are shown in table 1.
Table 1. Reliabilities of principal components

<table>
<thead>
<tr>
<th>principal component</th>
<th>cumulative contribution rate /%</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC1</td>
<td>90.702</td>
</tr>
<tr>
<td>PC2</td>
<td>94.109</td>
</tr>
<tr>
<td>PC3</td>
<td>98.109</td>
</tr>
</tbody>
</table>

The former three principal components (X71, X72, X73) are extracted respectively as the value of X, Y, Z axis, as shown in figure 5. From the figure, goaf water, limestone water and sandstone water can be identified qualitatively, but the edges of them are not obvious, so the other parameters about water chemistry are introduced to improve the prediction.

Figure 5. Principal component scores scatter plot (PCI × PCII × PCIII)

Table 2. Variable data of identifying source of water inrush

<table>
<thead>
<tr>
<th>Samples</th>
<th>X1</th>
<th>X2</th>
<th>X3</th>
<th>X4</th>
<th>X5</th>
<th>X6</th>
<th>X71</th>
<th>X72</th>
<th>X73</th>
<th>Water Composition Index</th>
<th>Discriminant classification results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Actual result</td>
<td>BP neural network Method used in this paper</td>
</tr>
<tr>
<td>1</td>
<td>2.00</td>
<td>494.73</td>
<td>932.95</td>
<td>2.13</td>
<td>8.40</td>
<td>2522.00</td>
<td>0.52867</td>
<td>1.74056</td>
<td>0.79809</td>
<td>Goaf water</td>
<td>I II II</td>
</tr>
<tr>
<td>2</td>
<td>4.21</td>
<td>303.60</td>
<td>748.68</td>
<td>9.93</td>
<td>8.56</td>
<td>1680.00</td>
<td>0.53662</td>
<td>1.56792</td>
<td>1.00328</td>
<td>Limestone water</td>
<td>I I I</td>
</tr>
<tr>
<td>3</td>
<td>2.40</td>
<td>429.68</td>
<td>1054.98</td>
<td>14.89</td>
<td>8.70</td>
<td>2211.00</td>
<td>1.97594</td>
<td>1.9391</td>
<td>1.49245</td>
<td>Sandstone water</td>
<td>I I I</td>
</tr>
<tr>
<td>4</td>
<td>11.20</td>
<td>450.11</td>
<td>1091.59</td>
<td>34.03</td>
<td>6.22</td>
<td>2519.00</td>
<td>1.24851</td>
<td>1.74388</td>
<td>0.23068</td>
<td>Goaf water</td>
<td>I I I</td>
</tr>
<tr>
<td>5</td>
<td>7.41</td>
<td>234.73</td>
<td>617.53</td>
<td>16.14</td>
<td>8.10</td>
<td>1235.00</td>
<td>0.28142</td>
<td>1.54675</td>
<td>1.39173</td>
<td>Limestone water</td>
<td>I I I</td>
</tr>
<tr>
<td>6</td>
<td>6.41</td>
<td>243.96</td>
<td>626.96</td>
<td>15.19</td>
<td>8.28</td>
<td>1262.00</td>
<td>1.46843</td>
<td>0.64771</td>
<td>2.53788</td>
<td>Sandstone water</td>
<td>I II II</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Samples</th>
<th>X1</th>
<th>X2</th>
<th>X3</th>
<th>X4</th>
<th>X5</th>
<th>X6</th>
<th>X71</th>
<th>X72</th>
<th>X73</th>
<th>Water Composition Index</th>
<th>Discriminant classification results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Actual result</td>
<td>BP neural network Method used in this paper</td>
</tr>
<tr>
<td>7</td>
<td>7.82</td>
<td>247.60</td>
<td>499.12</td>
<td>38.64</td>
<td>8.62</td>
<td>1484.00</td>
<td>1.64392</td>
<td>-0.0788</td>
<td>1.18037</td>
<td>Goaf water</td>
<td>I I I</td>
</tr>
<tr>
<td>8</td>
<td>5.21</td>
<td>460.49</td>
<td>1148.95</td>
<td>19.50</td>
<td>8.55</td>
<td>2380.00</td>
<td>0.36821</td>
<td>0.30487</td>
<td>-0.32166</td>
<td>Limestone water</td>
<td>I I I</td>
</tr>
<tr>
<td>9</td>
<td>6.81</td>
<td>236.57</td>
<td>609.59</td>
<td>25.17</td>
<td>8.55</td>
<td>1293.00</td>
<td>0.24351</td>
<td>1.12783</td>
<td>-0.01228</td>
<td>Sandstone water</td>
<td>I I I</td>
</tr>
<tr>
<td>10</td>
<td>86.77</td>
<td>894.31</td>
<td>371.61</td>
<td>1063.15</td>
<td>7.5</td>
<td>5740.00</td>
<td>-0.46035</td>
<td>0.39214</td>
<td>-0.52377</td>
<td>Goaf water</td>
<td>II II II</td>
</tr>
<tr>
<td>11</td>
<td>86.57</td>
<td>817.95</td>
<td>329.51</td>
<td>1028.05</td>
<td>7.96</td>
<td>4929.00</td>
<td>0.14442</td>
<td>0.01228</td>
<td>-0.87574</td>
<td>Limestone water</td>
<td>II II II</td>
</tr>
<tr>
<td>12</td>
<td>88.98</td>
<td>862.94</td>
<td>342.71</td>
<td>1091.86</td>
<td>7.68</td>
<td>5216.00</td>
<td>0.21873</td>
<td>-0.87574</td>
<td>-0.52042</td>
<td>Sandstone water</td>
<td>II II II</td>
</tr>
<tr>
<td>13</td>
<td>83.37</td>
<td>863.95</td>
<td>335.61</td>
<td>1067.05</td>
<td>8.1</td>
<td>5163.00</td>
<td>0.34302</td>
<td>-0.52042</td>
<td>-0.52042</td>
<td>Goaf water</td>
<td>II II II</td>
</tr>
<tr>
<td>14</td>
<td>90.78</td>
<td>892.24</td>
<td>378.32</td>
<td>1049.67</td>
<td>7.82</td>
<td>5790.00</td>
<td>0.15671</td>
<td>0.07325</td>
<td>-0.52042</td>
<td>Limestone water</td>
<td>II II II</td>
</tr>
<tr>
<td>15</td>
<td>16.02</td>
<td>1353.34</td>
<td>912.86</td>
<td>1051.20</td>
<td>6.67</td>
<td>7081.00</td>
<td>0.43598</td>
<td>0.32166</td>
<td>-0.52042</td>
<td>Sandstone water</td>
<td>II I I</td>
</tr>
<tr>
<td>16</td>
<td>84.89</td>
<td>1165.09</td>
<td>341.71</td>
<td>1184.10</td>
<td>7.68</td>
<td>7045.00</td>
<td>0.00971</td>
<td>-0.52042</td>
<td>-0.52042</td>
<td>Goaf water</td>
<td>II II II</td>
</tr>
<tr>
<td>17</td>
<td>90.49</td>
<td>1178.10</td>
<td>292.90</td>
<td>1169.90</td>
<td>7.66</td>
<td>7092.00</td>
<td>0.11763</td>
<td>1.13491</td>
<td>-0.52042</td>
<td>Limestone water</td>
<td>II II II</td>
</tr>
<tr>
<td>18</td>
<td>91.29</td>
<td>1194.60</td>
<td>322.19</td>
<td>1173.50</td>
<td>6.11</td>
<td>7135.00</td>
<td>0.29481</td>
<td>0.22107</td>
<td>1.97475</td>
<td>Sandstone water</td>
<td>II II II</td>
</tr>
</tbody>
</table>
In the table 2, the sample data, to some extent, have redundant relations. The conductivity value has a direct relation with various ion concentrations, and the spectral data also have connections with ion concentrations to a certain extent. The correlation matrix of various water samples is shown in table 3, which indicates that there is a clear correlation among these water parameters, such as the correlation coefficients for conductivity value and Na\(^+\), Cl\(^-\) are 0.977 and 0.963 respectively. Information overlaps among samples will inevitably affect the water discriminant accuracy and lead to a wrong result. Therefore, the sample data should be standardized in the first step, which will be processed with PCA in the next step.

Table 3. Pearson correlation coefficients matrix of each water component

<table>
<thead>
<tr>
<th></th>
<th>X1</th>
<th>X2</th>
<th>X3</th>
<th>X4</th>
<th>X5</th>
<th>X6</th>
<th>X7</th>
<th>X72</th>
<th>X73</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X2</td>
<td>.036</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X3</td>
<td>-.510</td>
<td>-.279</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X4</td>
<td>.274</td>
<td>.930</td>
<td>-.560</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X5</td>
<td>-.631</td>
<td>.310</td>
<td>.565</td>
<td>.018</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X6</td>
<td>.243</td>
<td>.977</td>
<td>-.399</td>
<td>.963</td>
<td>.156</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X71</td>
<td>.017</td>
<td>-.408</td>
<td>.207</td>
<td>-.442</td>
<td>-.175</td>
<td>-.394</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X72</td>
<td>-.606</td>
<td>-.018</td>
<td>.631</td>
<td>-.239</td>
<td>.666</td>
<td>-.153</td>
<td>.000</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>X73</td>
<td>-.155</td>
<td>-.401</td>
<td>.179</td>
<td>-.409</td>
<td>.095</td>
<td>-.424</td>
<td>.000</td>
<td>.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Table 4 shows the variance contribution ratio and the cumulative contribution ratio of each component, selecting the principal components by eigenvalue arranged from big to small, the larger eigenvalue indicates the more important component, selects the principal component data which the cumulative contribution ratio is more than 85%. According to each principal component eigenvalue distribution in Figure 6, the first three principal components are extracted, the cumulative contribution ratio reaches 87.629%, and they can effectively summarize the original data sample information.

Table 4. Eigenvalues, variances’ contribution rates and accumulative contribution rates of principal components Y1～Y3

<table>
<thead>
<tr>
<th>Principal component</th>
<th>Eigenvalue</th>
<th>contribution rate %</th>
<th>accumulative contribution rate %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y1</td>
<td>3.784</td>
<td>42.047</td>
<td>42.047</td>
</tr>
<tr>
<td>Y2</td>
<td>2.624</td>
<td>29.161</td>
<td>71.208</td>
</tr>
<tr>
<td>Y3</td>
<td>1.028</td>
<td>16.422</td>
<td>87.629</td>
</tr>
</tbody>
</table>
The principal component factor loading matrix in Table 5 shows that the extracted new factor $Y_1, Y_2, Y_3$ and the original data have relational expression:

\[
Y_1 = 0.2447X_1 + 0.4251X_2 - 0.3578X_3 + 0.4904X_4 - 0.1002X_5 + 0.4673X_6 - 0.2252X_7 + 0.2288X_8 - 0.2442X_9 \tag{12}
\]

\[
Y_2 = -0.4142X_1 + 0.321X_2 + 0.308X_3 + 0.1469X_4 + 0.5562X_5 + 0.2185X_6 - 0.1982X_7 + 0.4562X_8 - 0.0673X_9 \tag{13}
\]

\[
Y_3 = 0.0641X_1 + 0.0424X_2 + 0.1568X_3 - 0.0187X_4 - 0.0769X_5 + 0.0523X_6 + 0.6302X_7 + 0.1637X_8 - 0.7318X_9 \tag{14}
\]

According to equation (12), (13) and (14), the data in Table 2 were normalized, and then processed by principal component analysis, the calculated samples shows in Table 6:

\[
\begin{align*}
\text{Table 6. The data after calculation of principal components} \\
\text{Index} & \quad \text{1} & \quad \text{2} & \quad \text{3} & \quad \text{4} & \quad \text{5} & \quad \text{6} & \quad \text{7} & \quad \text{8} & \quad \text{9} & \quad \text{10} \\
Y_1 & -0.55431 & -0.8865 & -0.76386 & -0.83021 & -1.02027 & -1.33972 & -0.52824 & -0.43106 & -0.55037 & 0.75178 \\
Y_2 & 1.95534 & 1.2462 & 1.86123 & 1.51415 & 0.83848 & 0.61032 & 0.47965 & 1.3369 & 0.70813 & -0.25685 \\
Y_3 & -0.16465 & -0.32346 & 2.57538 & 1.29687 & -0.76087 & -1.00855 & -2.16516 & -0.29084 & -0.74965 & -0.22443 \\
\end{align*}
\]

In this paper, three-layer BP neural network is used to establish Inrush Water prediction model, data $Y_1, Y_2, Y_3$ in Table 5 are used as the input layer neural nodes, there are 3 output nodes (Sandstone water(100), Limestone water(010), Goaf water(001)), and 10 hidden layer nodes. After learning 135 times, the network meets the required accuracy, its error is $1e^{-8}$, as shown in figure 7:
3.4 Verify the Inrush Water Source Model

To validate the actual discrimination effect of the PCA-BP neural network model, data samples in Table 2 are predicted one by one in the established model to distinguish. There are three water samples which are determined error, its accuracy is 90%, Compared to recognition ratio 83% of BP neural network without PCA processing, it has a good effect. The reason why determine goes wrong is that Sandstone Water and Too Gray Water has some relevance, some sample data are similar in water quality, leading to a miscarriage of justice.

To further monitor the accuracy of the model, the trained model is used to distinguish another 11 groups water source samples of the mine. Table 7 shows the test samples of water source discriminant model, the prediction results are also in Table 7. In addition to Limestone Water is being mistaken to Sandstone Water, other discrimination results are correct, the discrimination accuracy ratio reaches 91%. The discrimination results of BP neural network without PCA processing are also listed, its accuracy ratio is 82%. Comprehensive comparison, PCA-BP Neural Network Model which is used in this paper is more reliable, high stability, and can meet the practice requirements of online Inrush Water source identification.

<table>
<thead>
<tr>
<th>Water Source Sentenced Model variables data</th>
<th>Discriminant classification results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water Composition Index</td>
<td>Actual result</td>
</tr>
<tr>
<td>X1</td>
<td>X2</td>
</tr>
<tr>
<td>1</td>
<td>162.72</td>
</tr>
<tr>
<td>2</td>
<td>161.72</td>
</tr>
<tr>
<td>3</td>
<td>135.47</td>
</tr>
<tr>
<td>4</td>
<td>80.08</td>
</tr>
<tr>
<td>5</td>
<td>91.29</td>
</tr>
<tr>
<td>6</td>
<td>91.29</td>
</tr>
<tr>
<td>7</td>
<td>94.49</td>
</tr>
<tr>
<td>8</td>
<td>88.09</td>
</tr>
<tr>
<td>9</td>
<td>6.81</td>
</tr>
<tr>
<td>10</td>
<td>4.01</td>
</tr>
<tr>
<td>11</td>
<td>3.41</td>
</tr>
</tbody>
</table>

4. Conclusions

Mine water inrush is serious challenge at the safety of coal production, rapid inrush water source identification in coal mines has great effect for early warning. Online monitoring system which is mainly composed of sensor nodes, information collector and ground monitoring center for Water Inrush Source had been built. Ca²⁺, Na⁺, HCO₃⁻, Cl⁻ ions, PH value, conductivity and fluorescence spectrum had been collected by sensor nodes, inrush water source identification has been discriminated online by PCA-BP neural network model. Using the PCA-BP techniques, we found:

1. The article proposed fluorescence spectroscopy, PH value, conductivity, ions transducer, discriminated mine water source online by PCA-BP neural network model. The method overcomes the weaknesses of the
traditional water chemical analysis method which processing time is too long and manual labor-intensive is too high, and the accuracy ratio of the system is 90%.

2. According to PCA-BP Neural Network proposed by author, eliminating the information overlap characteristics between each component by PCA can improve the accuracy of the source discrimination. Research shows that water source data processed by PCA has a greater improvement of discrimination accuracy ratio than only using BP Neural Network.

3. PCA-BP Neural Network Model which is used in this paper is more reliable, high stability, and can meet the practice requirements of online Inrush Water source identification. Collectively, our research demonstrate that the PCA-BP Neural Network Model provides a rapid, online means to discriminate the mine water source. The means employ not only mine water source but also aquifer connectivity and assessment of the risk of groundwater inrush.

Acknowledgements

1. The National Nature Science Foundation of China, Research on the theoretical foundation of multi-turn small coil transient electromagnetic methods in underground mine, No. 41674133;
2. The "Twelfth Five Year" National Science and Technology Support Program, Real-Time Monitoring and Warning Techniques for Underground Coal Mine Inundations, No: 2013BAK06B01;
3. The Nature Science Foundation of Xingjian Uighur Autonomous Region, Research on Copper deposit mineralization fluid and mechanism at lama su Wenquan County Xinxiang province ,No. 2016D01C067;

References


[26] Li Z, Zhao X. (2017) BP artificial neural network based wave front correction for sensor-less free space optics communication. *Optics Communications*, 385, pp. 219-228.


