XU XING
SUN GUANGZHONG
TIAN KUNYUN
and
ZHANG RUILIN

Discrimination methods of water inrush from mine floor based on PCA-Fisher

A method based on principal component analysis (PCA) and Fisher discrimination analysis is proposed targeting water inrush from mine floor. Based on the analysis of a large number of measured data from past projects, 13 factors affecting and controlling water inrush from floor are selected as the discrimination indexes. Firstly, the dimension of multi-index floor water inrush data is reduced by principal component analysis, and 4 principal component factors are extracted. Then the PCA-Fisher discrimination model of mine floor water inrush risk is established based on Fisher discrimination analysis theory, and its discrimination effect is verified by recurrent discrimination analysis and an example of its application is presented. The application results show that the results of the discrimination model are consistent with the actual situation, with an accuracy of 100%, which can provide a more effective method for discriminating the water inrush risk from mine floor.

Keywords: *Floor water inrush, risk discrimination, principal component analysis, Fisher discrimination analysis.*

1. Introduction

Since 2000, a total of 1,162 coal mine accidents have occurred in China, resulting in 4,676 deaths [1, 2]. With the growing depth and breadth of coal mining, the dangers of water inrush from the mine floor have become increasingly serious, which has severely restricted the construction and production of high-yield and high-efficiency mines. According to relevant statistics [3, 4], among more than 600 key coal mines in China, 285 of them have the danger of water inrush, accounting for 47.5%, and the water-threatened reserves amount to 25 billion tonnes. Thus it is of great practical significance to correctly discriminate the danger of water inrush from the mine floor for effectively reducing the coal mine accidents and ensuring the safe production of the coal mine.

To carry out research on prevention and control of water inrush from mine floor and to evaluate its risk can provide basis and guidance for the measures for prevention and control of water inrush from mine floor. At present, many scholars have studied the risk of water inrush from mine floor in different fields and proposed and established different risk assessment and evaluation models of floor water inrush. For example, Yang Zhilei et al. [5], optimized BP neural network using genetic algorithm to establish a nonlinear prediction model of floor water inrush of GA-BP neural network. The application showed that the model was fast and accurate. Cao Qingkui et al [6] combined the membership degree of fuzzy theory with support vector machine to establish a fuzzy-support vector machine model for evaluating the water inrush risk. The application showed that the model could solve such problems as small sample and nonlinearity. Based on the theory of unascertained mathematics, Ye Shixiong et al. [7] constructed the evaluation model of unascertained measure of water inrush from floor. The application showed that the result of the evaluation model was consistent with the actual situation of the mine. Based on the catastrophe theory, Xu Debao [8] established the evaluation model of water inrush from floor and evaluated the risk of water inrush in 9 sections of Tengbei coal mine. The application proved that the model was simple and easy to operate. Li Bo [9] established a risk assessment model of mine floor water inrush based on fuzzy evaluation and comprehensive weighting and evaluated the danger of floor water inrush from 6102N working face. Considering that the mine floor water inrush is a nonlinear dynamic phenomenon under the comprehensive effect of various factors, and there is a complicated relationship among the factors, there are still many shortcomings in the above methods, although they have been widely used in prediction of mine floor water inrush. For example, there is a high coupling between influencing factors of water inrush from the mine floor. If these factors are not analyzed, the accuracy of water inrush prediction will be affected. Or the complex methods, the large data computation, the strong subjectivity and other shortcomings are not conducive to the establishment and understanding of the model. Therefore, it is proposed to use PCA to transform multi-index variables associated with each other into new sample indexes independent of each other through linear combination, to reduce the information coincidence among the influencing

Messrs. Xu Xing, Sun Guangzhong, Tian Kunyun, Zhang Ruilin, School of Safety Engineering, Henan University of Engineering, Zhengzhou, Henan 451191, China. Email: xxzhengjie2001@163.com

factors of floor water inrush and reduce the dimension of the model ρ as to improve the learning efficiency and accuracy of the model. Based on Fisher discrimination analysis theory, the main components of water inrush from mine floor are analyzed comprehensively. Finally, based on the advantages of Fisher discrimination analysis, a model of mine floor water inrush risk assessment based on PCA-Fisher discrimination is established, and the model is applied as an example.

2. Principal component analysis (PCA)

PCA is a method of data compression and feature information extraction, which can transform the problem of high-dimensional space into low-dimensional space for processing, effectively eliminate the correlation between high-dimensional data sets and reduce the dimension of data. However, the simplified data structure can provide most of the information of original data.

Model of the principal component analysis:

A linear combination $Y = AX$ shown in formula (1) is constructed, and ρ new variables are obtained by linear combination of ρ variables in the data matrix X.

$$\left.\begin{aligned} Y_1 &= a_{11}X_1 + a_{12}X_2 + \mathrm{L} + a_{1p}X_p \\ Y_2 &= a_{21}X_1 + a_{22}X_2 + \mathrm{L} + a_{2p}X_p \\ &\mathrm{M} \\ Y_p &= a_{1p}X_p + a_{1p}X_p + \mathrm{L} + a_{pp}X_p \end{aligned}\right\} \quad \ldots (1)$$

Where, there is a linear independence between $a_{i1}+a_{i2}+\ldots+a_{ip} = 1$ and $Y_i N Y_j$ (i Hj, i, j, = 1, 2,..., p). The variance of Y_1 is greater than that of Y_2, which is greater than that of Y_3, and so on.

PCA solution steps are:

1. First, normalize the original variable data, and then calculate the covariance matrix of each variable, $\Sigma = (S_{ij})$ $p \times p$.
2. Calculate the eigenvalue λ_i of the covariance matrix Σ in step 1 and the corresponding orthogonalization unit eigenvector α_i. The first m larger eigenvalue of the covariance matrix Σ is $\lambda_1 > \lambda_2 > \ldots > \lambda_m > 0$ in order from large to small, and the orthogonal unit eigenvector σ_{i} corresponding to λ_i is the coefficient of the original variable of the principal component F_i, and $\alpha_i = \lambda_i / \Sigma\lambda_i$ (i = 1, 2,..., p).
3. Select the number of principal components. Usually, when the cumulative variance contribution ratio $\Sigma\lambda_{\mathrm{i}} / \Sigma\lambda_j$ (i = 1, 2,..., m), (j = 1, 2,..., p) reaches 80% or more, the first m principal components of the variable data are taken as the discrimination indexes. At this point, the sample information of m principal components contains most information of the original sample as required.

3. Fisher discrimination analysis theory

The main mathematical model and idea of Fisher discrimination analysis is to reduce the data dimension of a multi-dimensional problem through projection, so that the problem is simplified, and the discrimination function is determined according to the principle that the distance between categories is the maximum and the distance within categories is the minimum. It is of great practical significance.

Suppose there are two populations, G_1 and G_2, n_1 is extracted from the first population, n_2 from the second, and p indices of each sample are observed. Establish a discrimination function $y = c_1x_1 + c_2x_2 + \ldots + c_px_p$, and calculate the critical value y_0, then classify the new samples according to the criterion. Test the discrimination effect.

$H_0 : Ex_a^{(1)} = S_1 = Ex_a^{(2)} = \mu_2$ $H_1 : \mu_1 \neq \mu_2$

Test statistics:

$$F = \frac{(n_1 + n_2 - 2) - p + 1}{(n_1 + n_2 - 2)p} T^2 : \quad F(p, n_1 + n_2 - p - 1) \quad \ldots (2)$$

where,

$$T^2 = (n_1 + n_2 - 2)\cdot[\sqrt{\frac{n_1n_2}{n_1+n_2}}(\overline{X}^{(1)} - \overline{X}^{(2)})' \; S^{-1}\sqrt{\frac{n_1n_2}{n_1+n_2}}(\overline{X}^{(1)} - \overline{X}^{(2)})] \quad \ldots (3)$$

$$S = (s_{ij})_{p\times p},$$

$$s_{ij} = \sum_{a=1}^{n_1}(x_{ai}^{(1)} - \overline{x}_i^{(1)})(x_{aj}^{(1)} - \overline{x}_j^{(1)}) + \sum_{a=1}^{n_2}(x_{ai}^{(2)} - \overline{x}_i^{(2)})(x_{aj}^{(2)} - \overline{x}_j^{(2)}) \quad \ldots (4)$$

$$\overline{X}^{(i)} = (\overline{x}_1^{(i)}, \mathrm{L}, \overline{x}_p^{(i)})' \quad \ldots (5)$$

Given the test level α, check the distribution table F to determine the critical value F_α, if F < F_α, then H_0 will be denied, and the discrimination is valid, otherwise, invalid.

4. PCA-Fisher discrimination model and its application

Mine floor water inrush is a complex hydrogeological problem which is comprehensively controlled by many factors, such as geological structure, hydrogeology, floor water insulation layer and mining condition, thus it has nonlinear dynamic characteristics. The selection of evaluation indexes should not only consider the operability and representativeness of the indexes, but also consider their accuracy. Too many or too few evaluation indexes will result in a rebate in working conditions and credibility of the evaluation method, and also limit its popularization and application. Taking the evaluation index system of water inrush from floor in the literature [10] as reference, the following 13 factors are selected as evaluation indexes of water inrush from floor, fault density (V_1), fault water conductivity (V_2), fracture development degree (V_3), confined water pressure (V_4), aquifer water-richness (V_5), karst development degree (V_6), strong water source supply (V_7), aquifuge thickness (V_8), aquifuge strength (V_9), aquifuge integrity (V_{10}), mining thickness (V_{11}), mining depth (V_{12}) and slant length of working face (V_{13}). With references to Regulations on Prevention and Control of

Coal Mine Water and practical experience, the state of water inrush from mine floor is divided into non-water inrush and water inrush. The former refers to a safe state, in which case the risk of water inrush from the floor of the mine is very low or there is little water inrush on the floor, which basically does not affect the safety production of coal mine. The latter refers to a dangerous state, in which case water inrush from mine floor has serious influence on mining face, thus can pose a great threat to the safety production of coal mine. The qualitative and quantitative methods of evaluation indexes are the same as those of the literature [10] and based on the case data collected in Table 1 of the literature [10], the division of training samples and test samples is the same as that of the literature [10].

4.1 Processing of data by PCA

The original data are imported into the SPSS software for standardization, and the correlation coefficient matrix between the influencing factors of each water inrush is obtained, as shown in Table 1. Table 1 shows that there is a clear correlation between the various factors. The correlation coefficient between fault density and fault water conductivity reaches 96.2%, and the correlation coefficient between confined water pressure and mining thickness reaches 95%. It shows that there is information overlapping among the 13 evaluation indexes. If the above evaluation indexes are directly used as the basis for the risk assessment of water inrush from the floor, the information will be redundant and the calculation amount will increase. What is more, the accuracy of mine floor water inrush risk evaluation model may be affected, which may even results in misjudgment. Therefore, it is feasible to use PCA for data dimensionality reduction to extract principal components.

The eigenvalues, eigenvalue contribution rate and cumulative contribution rate of the correlation coefficient matrix are calculated. According to the information distribution rules of each principal component in the PCA lithotripsy in Fig.1, if the principal component whose characteristic root is greater than 1 is selected, the cumulative variance contribution rate reaches 87.738%. In order to reduce the information loss, the first 4 principal components can be extracted to effectively summarize the original sample information.

According to the PCA principal component score coefficient matrix, the expressions of the principal components can be obtained as follows:

$$\begin{aligned}F_1 = &-0.143\times V_1 - 0.141\times V_2 - 0.087\times V_3 + 0.164\times V_4\\ &+0.095\times V_5 - 0.110\times V_6 - 0.133\times V_7 + 0.055\times V_8\\ &+0.122\times V_9 - 0.041\times V_{10} + 0.069\times V_{11} + 0.159\times V_{12}\\ &+0.129\times V_{13}\end{aligned} \quad \text{... (6)}$$

$$\begin{aligned}F_2 = &\,0.117\times V_1 + 0.191\times V_2 + 0.315\times V_3 + 0.071\times V_4\\ &-0.203\times V_5 - 0.057\times V_6 + 0.002\times V_7 + 0.268\times V_8\\ &+0.207\times V_9 + 0.192\times V_{10} + 0.108\times V_{11} + 0.117\times V_{12}\\ &+0.115\times V_{13}\end{aligned} \quad \text{... (7)}$$

$$\begin{aligned}F_3 = &\,0.073\times V_1 + 0.073\times V_2 + 0.105\times V_3 - 0.038\times V_4\\ &+0.225\times V_5 - 0.170\times V_6 - 0.277\times V_7 - 0.286\times V_8\\ &-0.134\times V_9 + 0.368\times V_{10} + 0.302\times V_{11} + 0.092\times V_{12}\\ &-0.227\times V_{13}\end{aligned} \quad \text{... (8)}$$

$$\begin{aligned}F_4 = &-0.314\times V_1 - 0.204\times V_2 + 0.143\times V_3 - 0.128\times V_4\\ &+0.188\times V_5 + 0.458\times V_6 + 0.282\times V_7 - 0.031\times V_8\\ &+0.089\times V_9 + 0.284\times V_{10} + 0.219\times V_{11} - 0.015\times V_{12}\\ &+0.152\times V_{13}\end{aligned} \quad \text{... (9)}$$

According to the above principal component expression, the principal component analysis calculation is performed on the normalized raw data, and scores of 4 principal components (including the corresponding water inrush categories) of 10 samples are obtained as shown in Table 2.

TABLE 1 PEARSON CORRELATION COEFFICIENT MATRIX BETWEEN INDICATORS

Indicator	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	V_9	V_{10}	V_{11}	V_{12}	V_{13}
V_1	1.000												
V_2	0.962	1.000											
V_3	0.527	0.717	1.000										
V_4	-0.572	-0.567	-0.347	1.000									
V_5	-0.678	-0.700	-0.532	0.171	1.000								
V_6	0.087	0.156	0.266	-0.636	-0.255	1.000							
V_7	0.353	0.406	0.340	-0.699	-0.512	0.913	1.000						
V_8	-0.121	0.012	0.315	0.332	-0.282	-0.237	-0.033	1.000					
V_9	-0.478	-0.351	0.085	0.582	0.188	-0.393	-0.354	0.808	1.000				
V_{10}	0.183	0.370	0.764	-0.264	0.079	0.141	-0.008	-0.069	0.033	1.000			
V_{11}	-0.252	-0.187	0.128	0.352	0.200	-0.151	-0.315	-0.012	0.253	0.348	1.000		
V_{12}	-0.562	-0.506	-0.167	0.950	0.244	-0.591	-0.715	0.278	0.598	0.020	0.559	1.000	
V_{13}	-0.597	-0.478	-0.147	0.762	0.075	-0.164	-0.205	0.468	0.659	-0.154	0.187	0.743	1.000

TABLE 2. PRINCIPAL COMPONENT SCORES AND WATER INRUSH CATEGORIES

Number	Principal component scores				Discrimination results	
	F_1	F_2	F_3	F_4	Actual results	PCA-Fisher discrimination
1	-0.378	0.902	0.078	0.42	0	0
2	-0.093	0.717	0.342	1.294	1	1
3	0.675	0.648	0.313	0.586	0	0
4	-0.209	0.261	0.416	-0.052	1	1
5	0.131	0.044	-0.534	1.053	0	0
6	-0.182	1.162	-0.622	0.583	1	1
7	0.594	0.534	-0.014	0.426	0	0
8*	0.613	0.552	-0.004	0.41	0	0
9*	-0.114	-0.014	-0.017	1.013	1	1
10*	-0.115	0.466	-0.287	0.289	1	1

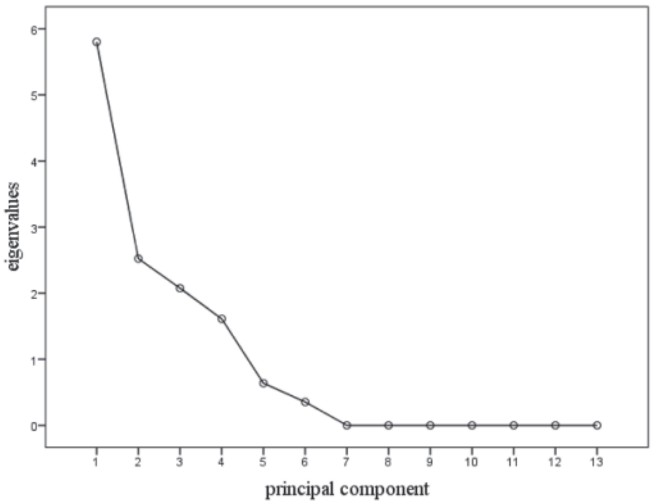

Fig.1 PCA lithotripsy

4.2 ESTABLISHMENT OF PCA-FISHER DISCRIMINATION MODEL

Taking F_1, F_2, F_3 and F_4 as four discrimination indexes of Fisher discrimination analysis, based on the basic idea of Fisher discrimination analysis, the PCA-Fisher discrimination model of mine floor water inrush risk is established by using the first 7 groups of training samples, as follows.

$$Z = 0.936\times F_1 - 0.292\times F_2 - 0.405\times F_3 - 0.153\times F_4 \qquad ... (10)$$

The center values of the PCA-Fisher discrimination model in two categories can be calculated by Formula 10. As shown in Table 3, it can be determined which category the sample belongs to by comparing the distances between the function values of the pending sample and the center values of the two categories.

TABLE 3 CENTER VALUES OF THE CATEGORIES

Categories	Center values
0	0.529
1	– 0.706

Combined with Table 3, the false judgment rate of the 7 groups of water inrush samples is calculated by using the recursive estimation method according to Formula 10. The estimation results are shown in the right-most side of Table 2. It can be seen that the estimation of all the original samples are correct, with an accuracy of 100%, which indicates that the Fisher discrimination analysis model is stable and reliable and can be used to distinguish the mine floor water inrush risk.

Hence samples number 8, 9, and 10 are brought into that formula for testing their discrimination application effect. The function value of sample 8 obtained is 0.351, that of sample 9 is –0.251, and that of sample is –0.172. In combination with the category center values in Table 3, it can be determined that their categories are 0, 1, and 1 respectively. The results are consistent with that of literature [10], proving that the PCA-Fisher based method is effective and reliable to discriminate the mine floor water inrush risk. At the same time, it should be noticed that the data processed by PCA not only reduces the dimension of the multivariable data system, but also simplifies the statistics of the variable system and plays the role of dimension reduction and noise elimination, which lays a foundation for accurately discriminating the danger of water inrush from mine floor.

5. Conclusions

The PCA method is used to reduce the dimension of mine floor water inrush data with 13 indexes. The principal component is selected on the premise that the characteristic root is higher than 1. The cumulative variance contribution rate of the extracted 4 principal components reaches 87.738%, and the principal component score is calculated. This provides a basis for accurate discrimination in the next step and reduces the complexity of the discrimination. It should be seen that PCA can make the analysis simple, intuitive and effective.

Based on Fisher discrimination analysis theory, a discrimination model of mine floor water inrush risk based on PCA-Fisher is established. Its estimation accuracy of the effectiveness is 100%, and its accuracy of the discrimination application is 100%. Therefore, the discrimination analysis method is feasible, which can be used as a method to evaluate the risk of water inrush from mine floor and provide some references for the predictions in water prevention and control in coal mines.

Acknowledgment

This work is supported by the National Natural Science Foundation of China (NSFC) (Grant no. 51604091); the Scientific and Technological Breakthrough Program of Henan (Grant no. 182102310743); and the Key Scientific Research Projects in the Colleges and Universities of Henan (Grant no. 18A440010).

References

[1] W. J. Sun, W. F. Zhou, J. Jiao (2016): Hydrogeological classification and water inrush accidents in china's coal mines. *Mine Water and the Environment*, Vol. 35, No. 2, pp. 214–220.

[2] X. Xu, Y. Z. Li, W. Y. Zhang, Z. Zeng, (2017): Application of MPSO–BP model in discriminating mine water inrush source, *Journal of Natural Disasters*, Vol. 26, No. 5, pp. 140–148.

[3] X. Y. Wang, H. Y. Ji, Q. Wang, X. M. Liu, D. Huang, X. P. Yao, G. S. Chen, (2016): Divisions based on groundwater chemical characteristics and discrimination of water inrush sources in the Pingdingshan coalfield. *Environmental Earth Sciences*, Vol. 75, No. 10, pp. 1–11.

[4] Q. Wu, Progress (2014): Problems and prospects of prevention and control technology of mine water and reutilization in China. *Journal of China Coal Society*, Vol. 39, No. 5, pp. 795–805.

[5] Z. L. Yang., X. R. Meng, X. Q. Wang, K. Y. Wang, (2012): Nonlinear prediction and evaluation of coal mine neural network floor water inrush based on GA–BP model, *Safety in Coal Mines*, Vol.44, No.2, pp. 36–39.

[6] Q. K. Cao, F. Zhao, (2011): Risk evaluation of water inrush from coal floor based on fuzzy–support vector machine, *Journal of China Coal Society*, Vol. 36, No. 4, pp. 633–637.

[7] S. X. Ye, M. T. Jia, C. P. Pan, J. H. Chen, (2015): New assessment model for water–inrush hazard from the coal seam floor based on the fuzzy theory measurement, *Journal of Safety and Environment*, Vol. 15, No. 1, pp. 26–30.

[8] D. B. Xu, S. Y. Cao, F. X. Xu, Y. X. Chen, (2014): Water inrush risk evaluation of coal seam floor based on catastrophe theory, *Safety in Coal Mines*, Vol. 45, No. 12, pp. 197–200.

[9] B. Li, X. M. Guo, S. Xu, W. K. Liu, Z. H. Wang, H. K. Li, (2014): Risk assessment of coal floor groundwater bursting based on fuzzy comprehensive evaluation–comprehensive weight, *Journal of Henan Polytechnic University* (Natural Science), Vol. 33, No. 1, pp. 6–11.

[10] K. Chen, Q. H. Dong, Z. J. Du, H. K. Chu, (2013): Prediction model for coal mine floor water inrush and its application based on fisher discrimination method, *Safety in Coal Mines*, Vol.44, No.2, pp. 32–35.

STRATUM DEFORMATION ANALYSIS AND EXCAVATION SEQUENCE OPTIMIZATION FOR FOUR-LINE PARALLEL SHIELD TUNNELS

Continued from page 493

Journal, vol.44, no. 6, pp. 631-643.

[9] Li, X.F., Du, S.J., Zhang, D.F. (2012): "Analysis of the Impact of Newly-built Shield Tunnel Constructionon Close-spaced Parallel Tunnels." *Chinese Journal of underground space and Engineering*, vol. 8, no. 5, pp. 1065-1074.

[10] Wei, G, Pang, S.Y. (2014): "Study on 3D Soil Deformation Caused by Construction of Double-Line parallel shield tunnels." *Rock and Soil Mechanics*, vol. 35, no. 9, pp. 2562-2568.

[11] Wang, H.N., Zeng, G.S., Jiang, M.J. (2015): "Theoretical solutions for the mutual impact between existing tunnels and newly-built parallel tunnels in rheological rockmass." *Chinese Journal of Geotechnical Engineering*, vol. 37, no. 8, pp. 1365-1374.

[12] Bai, X. F., Wang, M. S., (2016): "Two-phase analysis for the impact of double-tunnel excavation on adjacent tunnels." *China Civil Engineering Journal*, vol. 49, no. 10, pp. 123-128, 2016.

[13] Yu, Y.L., Xu, H.Q. H.Q. Li, Z.W. (2011): "Analysis of environmental vibration of adjacent porous parallel tunnel undertrain load." *Chinese Journal of Geotechnical Engineering*, vol. 33, no. 11, pp. 1801-1806.

[14] Sun, Y., Liu, H. Z. (2002): "3D numerical simulation of ground surface settlement under over lapped shield tunneling." *Journal of Tongji University* (Natural Science Edition), vol. 30, no. 4, pp. 379.