Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Engineering 26 (2011) 1346 – 1351

Procedia Engineering

www.elsevier.com/locate/procedia

First International Symposium on Mine Safety Science and Engineering

Application of Principal Component Analysis Method to Determine Prediction Indexes Sensibility Sequence

Yang Guo-liang[a], Wang Can-zhao[b], Wu Shi-yue[c], Jia Li-qing[d], Zhang Sheng-zhu[a] a[*]

[a]Faculty of Resources and Safety Engineering, China University of Mining & Technology (Beijing), College Road No. 11 Haidian District Peking, 100083, China

[b]Xinji 1# Coalmine SDIC Xinji Energy Co. Ltd., Xinji Towm Anhui Province, 232170, China

[c]Taiyuan University of Technology, 79 Yingze Street, Taiyuan, 030024, China

[d] Yankuang Liuhuanggou Coal Mine in Xinjiang Mining, Xinjiang Changji, 831114, China

Abstract

Screening out sensitivity prediction indexes fit for the coal mine is very important technical work in coal and gas outburst prediction process. The feasibility that principal component analysis method been used in the coal and gas outburst prediction index screening is discussed. Through using gray correlation analysis method and principal component analysis method to analysis a batch of measured data of a coal mine, the validity of the latter is proved. Through the contrast of process between the gray correlation analysis and the principal component analysis method, the vantage of principal component analysis method is obvious, that is the better maneuverability, simplicity and easier to extend.

Selection and/or peer-review under responsibility of China Academy of Safety Science and Technology, China University of Mining and Technology(Beijing), McGill University and University of Wollongong.

Keywords: principal component analysis method, sensitivity of prediction index, sequence, feasibility analysis, validity analysis;

1. Introduction

Coal and gas outburst in recent years occurred mostly at the working face where outburst prediction result is safe, or after measures to prevent the outburst were taken. One reason of these is that coal and gas outburst prediction indexes been used are not fit for the coal mine, or rather the prediction indexes to the coal mine are insensitivity, is not able to reflect the major incentives of outburst. So screening out

*Corresponding author. Tel.: 15201290355;
E-mail address: yangguoliang86@gmail.com.

1877-7058 © 2011 Published by Elsevier Ltd. Open access under CC BY-NC-ND license.
doi:10.1016/j.proeng.2011.11.2310

sensitivity prediction indexes fit for the coal mine is very important technical work in coal and gas outburst prediction process.

2. The existent sensitivity sequencing methods of outburst prediction indexes

The common sensitivity sequencing methods of outburst prediction indexes mostly include "Three rate" method and gray correlation analysis method.

2.1. "Three rate" method

"Three rate" method which is based on predicted outburst rate, predicted outburst accuracy and predicted not outburst accuracy, is used to determine the sensitivity of prediction indexes. On the premise that prediction indexes meet the requirements of working face, the higher is the predicted accuracy of outburst, the higher is the sensitivity.

The superiority of "three rate" method is intuitive analysis process and easy to compute. But the basic data requirements which need detailed records of the drilling process of dynamic phenomena is high, because determining the dynamic phenomenon relies heavily on operator's experience.

2.2. Gray correlation analysis method

Gray correlation analysis method, based on gray system theory, can be used to calculate the correlation between the prediction indexes and objective vector. The greater is the correlation, the relationship between the prediction indexes and outburst risk is closer, the sensitivity of the prediction index is higher, and outburst risk can be reflected more.

Gray correlation analysis method demands less raw date than the "three rate" method, which avoid the inaccuracy caused by human factor. The superiority of gray correlation analysis method is high accuracy. However, the computing process is complex, especially when the capacity of prediction indexes is large, which lead to the Operation level of worker demanded is high, can't easy-to-used in the real work.

3. Feasibility analysis of principal component analysis method

The goal of principal component analysis method is reconfiguring a certain number of the original correlation indexes (such as P indexes), into a new set of composite indexes which are independent of each other. Mathematical treatment is usually making the original P indexes into a linear combination, as new composite indexes. But this linear combination will has a lot, if unchecked. In the selection process of principal component analysis method, if the first one selected linear combination as a comprehensive index denoted by F_1, F_1 is certainly hoped can reflect more information of original indexes, the most typical way to express the information here is using the variance of F_1, that is, if the greater $VAR(F_1)$ is, the information contained in F_1 is more. Therefore, F_1 should be the linear combination which has largest variance of all. F_1 is the first principal component. If the information contained by the first principal component is not enough, tried to select F_2 another linear component, in order to effectively reflect the original information, the information has been in F_1 is not needed to appear in the F_2. This can be expressed with mathematical formulas $Cov(F_1, F_2) = 0$. That F_2 is the second principal component. And so on the 3, 4… P principal component can be created. Obviously, the variance is decreasing. Therefore, in practice, we only select the first few principal components, although this will lose some information, it makes us grasp the principal contradiction, and from the raw data some new information is extracted, therefore in the study of some practical problems larger benefits than losses. This not only reduces the number of variables but also grasp the principal contradiction is conducive to analyze the problem.

The occurrence of coal and gas outburst is influenced by the geostress, gas and coal structure. These factors are quantified into many indexes to predict the coal and gas outburst, the relationship among indexes is independent and also has a certain correlation.

Through computing these indexes with principal component analysis method, we get many principal components, and determine the contribution value of each principal according to the variance. Because each principal component is linear combination of original prediction indexes, we can determine the role of each index through the linear representation of principal component. And base on the principle that the role is bigger the index is more sensitive, we can accomplish the purpose that determines prediction indexes sequence of sensibility.

As can be seen from the above analysis, principal component analysis methods can be used to determine prediction indexes sequence of sensibility.

4. The validity of principal component analysis method

This paper selects 62 sets of consecutive data from one coal mine, each set of date includes Initial velocity of gas emission from boreholes q, drilling cuttings weight S and desorption index for drill cuttings $\triangle h_2$, through comparing the sequence conclusions of sensibility get from gray correlation analysis method and principal component analysis method, the validity of latter can be verified.

Calculated by gray correlation analysis method, the degree of association between initial velocity of gas emission from boreholes q, drilling cuttings weight S, desorption index for drill cuttings $\triangle h_2$ and reference sequence are 0.881,0.894 and 0.806. This means initial velocity of gas emission from boreholes q is the most sensitive index of these three indexes, desorption index for drill cuttings $\triangle h_2$ is the second, drilling cuttings weight S is the last.

4.1. The establishment of principal component analysis method mathematical model

Suppose there are n samples, and each sample has P indexes: $X_1, X_2 \ldots X_P$, then we can get the original database as follow:

$$X = \begin{pmatrix} x_{11} & x_{12} & \ldots & x_{1p} \\ x_{21} & x_{22} & \ldots & x_{2p} \\ \vdots & \vdots & \ldots & \vdots \\ x_{n1} & x_{n2} & \ldots & x_{np} \end{pmatrix} \equiv (X_1, X_2, \cdots X_P) \tag{1}$$

Of which: $X_i = \begin{pmatrix} x_{1i} \\ x_{2i} \\ x_{3i} \end{pmatrix} \quad i = 1, \cdots, P$

Make linear combination with P vectors in data matrix X as follow:

$$F_i = a_1 X_1 + a_2 X_2 + \cdots + a_P X_P \quad i = 1, \cdots, P \tag{2}$$

X_1 is an n-dimensional vector.

The equations required:

$$a_{1i}^2 + a_{2i}^2 + \cdots a_{Pi}^2 = 1 \qquad i = 1, \cdots, P \tag{3}$$

Where coefficients aij is determined by the following principles:

- F_i isn't correlation with F_j（$i \neq j, i, j = 1, \cdots, P$）
- Variance of F_1 is the largest among all the linear combinations, F_i is eigenvector i, the main factors can be analyzed according to a_{ij}.

In the computus process, SPSS is used in this paper.

4.2. Analysis process and conclusions

According to the principle of principal component analysis method, the 62 set of dates is imported in SPSS software after standardized, get the Pareto diagram of systematic variance interpreted by principal components (as shown at Figure 1). Figure 1 show that the variance explained by three principal components account for the proportion of total variance 72.5%, 15.4%, 12.1%. They explain 100% of the total variance together (cumulative variance contribution rate), it can be said that the three main components reflect all the information of the raw data provided.

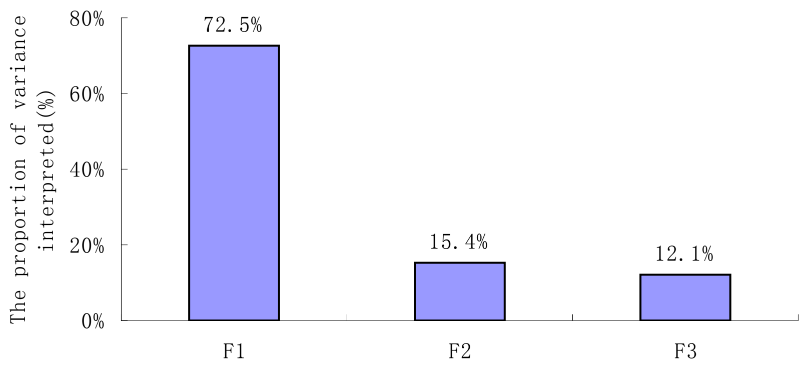

Fig 1. Pareto diagram of systematic variance interpreted by principal components

Weighting coefficient of variable in the principal component outputted by SPSS as shown at table 1.

Table 1. Component Score of variable in different principal components

Components	F_1	F_2	F_3
$X_1(\Delta h_{2\max})$	0.287	0.907	0.309
$X_2(q_{\max})$	0.931	0.270	0.247
$X_3(S_{\max})$	0.253	0.299	0.920

Weighting coefficient is estimate based on a common degree of variables, the same characteristic with regression coefficients, according to the Weighting coefficient of each variable on the principal components, the final eigenvalue can be determined, the calculate procedure is summing the squared weighting coefficient. Therefore we can get the eigenvalue of the three principal components are 1.012, 1.003, 0.985.

Coefficient of variable in the principal components expression is that weighting coefficient of each variable be divided by eigenvalue of principal components, the result can be seen in the table 2.

Table 3. Weighting coefficient of variable in different principal components

Components	F_1	F_2	F_3
$X_1(\Delta h_{2max})$	0.283	0.921	0.308
$X_2(q_{max})$	0.920	0.275	0.246
$X_3(S_{max})$	0.250	0.304	0.917

According to the date in the table 3, the expressions of three principal components can be written as follow:

$$F_1 = 0.283*X_1 + 0.920*X_2 + 0.250*X_3 \tag{4}$$

$$F_2 = 0.921*X_1 + 0.275*X_2 + 0.304*X_3 \tag{5}$$

$$F_3 = 0.308*X_1 + 0.246*X_2 + 0.917*X_3 \tag{6}$$

From the Pareto diagram of systematic variance interpreted by principal components, the proportion of systematic variance interpreted by the first and second principal component is 87.9%, which explain most of the system variables information. Through the expression of the first principal component, we can see that the weighting coefficient of variable X_2 which correspond with initial velocity of gas emission from borholes q is largest, is 3.3 times and 3.7 times the other two variables. In the second principal component, the weighting coefficient of variable X_1 which correspond with desorption index for drill cuttings $\triangle h_2$ is largest, is 3.8 times and 3.1 times the other two variables. By these we can get the conclusion that initial velocity of gas emission from borholes q is the most sensitive index of these three indexes, desorption index for drill cuttings $\triangle h_2$ is the second, drilling cuttings weight S is the last, which is the same as the conclusion of gray correlation analysis method.

5. Conclusions

Through theoretical analysis of the trait of outburst prediction indexes and the principle of principal component analysis methods, and contrast with the conclusion of gray correlation analysis method, it is shown that the principal component analysis method is feasible and validity as a test method of prediction indexes sequence of sensibility. With the assist of sofare SPSS, the calculation process is simple and clearer, the conclusion includes the larger amount of information, more suitable for large amounts of data and many affecting factors. This theory can be also applied to other uncertain factors which impact coal and gas outburst, such as buried kaicdepth, mining method, and type of laneway. But its reliability also needs to be proof by a great deal of practice.

References

[1] Xiang Dong-jin. Practical multivariate statistical analysis. Wuhan：*Wuhan University*; 2005.

[2] Wang Bao-jin. Multivariate analysis. Peking: *Peking University Press*, 2007.

[3] Yang Guo-liang, Wu Shi-yue, Chen Quan-qiu. Sihe coal mine drill cutting desorption indices: determination of the regional critical value. Peking: *China Coal*; 2010.

[4] Yan Jiang-wei, Zhang Zi-min. Analysis and prediction on premier control factors of gas emission in eastern area of sihe coal mine. Chong Qing: *Mining Safety & Environmental Protection*; 2007.

[5] Liu Si-feng, Guo Tian-bang. Grey system theory and applications. Peking. *Science Press*; 1999.

[6] Liu Zhao-peng, Ping Shou-kang. Comprehensive assessment of risk level of coal/gas outburst based on grey theory. Peking: *COAL Mining Technology*; 1999.

[7] Cheng Wu-yi. Coal and gas outburst region prediction theory and technology. Peking: *China Coal Industry Publishing House*; 2008

[8] Liu Yan-wei, Pan Hui. Analysis of rules of coal and gas outburst and controlling factors in hebi mine area. Peking: *Safety In Coal Mines*； 2006.

[9] Muo Ya-lin, Gao Jian-liang. Application of Grey Theory in Gas Prediction of Heading Face. Chong Qing: *Mining Safety & Environmental Protection*; 1996.