Assessing Slope Stability of Open Pit Mines using PCA and Fisher Discriminant Analysis

Shiyao Xiang

College of Geoscience and Surveying Engineering, China University of Mining and Technology (Beijing), Beijing 100083, China

***Abstract* — In order to determine the slope stability of open-pit mine effectively from available statistical data, this paper proposes a discriminant analysis model based on Principal Component Analysis (PCA) and Fisher discriminant analysis. The 6 discriminant indices which affect the slope stability of open-pit mine including gravity density of rocks, cohesion force, internal friction angle, slope angle, slope height and pore water pressure were reduced into 3 comprehensive indices including the first, second and third principal components by PCA to eliminate the mutual influence of the discriminant indices and avoid miscalculation due to different index dimensions. Then, a discriminant analysis model for slope stability was formulated by Fisher discriminant analysis which has no special requirements for data distribution. The back substitution method was used to test the model. The results of training and discrimination for 32 training samples and 7 pending samples show that this method can discriminate accurately and it can be a new method to discriminate the slope stability of open-pit mine. Compared with others, this method can improve the accuracy of discriminant analysis.**

Keywords - Open-pit mine; Slope stability; PCA; Fisher discriminant analysis; Dimensions reduction

I. INTRODUCTION

With the increase of open-pit mining depth and the continuous expansion of mining boundary, slope stability has become one of the important factors affecting the safety of mining production. Slope stability directly affects the personnel and equipment safety and economic benefits of mine area [1]. Monitoring the open-pit mine slopes and discriminating the stability of them have been important research contents of mine surveying domain and mine safety domain [2].

At present, there are many methods to discriminate the slope stability which have been used in practice, such as quantitative analysis methods which include the limit equilibrium method [3], numerical analysis method [4], [5], [6], block theory method [7] and so on, and qualitative analysis methods which include the analogism [8], graphic method [9] and so on.

Aiming at the complicated and changeable influence factors of slope stability, the methods to predict unknown stability of new monitoring slopes by carrying out statistical analysis of existing slope data have been generated, such as distance discriminant analysis method [10], support vector machine method [11], neural network method [12], [13], random forest prediction method [14] and so on.

If there were too many discriminant indexes which affected slope stability, the miscalculation would be caused by information overlapping. Meanwhile, the different index dimensions and the unknown distribution of data would also affect the discriminant accuracy.

According to the questions mentioned above, this paper transforms the indexes with correlation and overlapping information into independent linear combinations by principal component analysis (PCA). Then, the linear combinations are used as the new comprehensive discriminant indexes to achieve the purposes of reducing dimensions and eliminating influences of different dimensions. Finally, the discriminant analysis model for slope stability of open-pit mine is built by Fisher discriminant analysis to discriminate the slope stability by statistical methods.

II. THEORY AND ALGORITHMS

A. PCA

PCA is a method which is used for compressing data and extracting feature information. According to the correlation of the indexes, they are transformed into a set of new and unrelated comprehensive indexes. Fewer comprehensive indexes are selected to reflect the original indexes as much as possible depending on the actual needs. Finally, the goal of reducing dimensions and simplifying structure of data can be achieved.

The basic principle is shown below:

Assume $\boldsymbol{X} = \left\{X_1, X_2, \cdots, X_p\right\}^T$ is p-dimension vector, and do linear transformation as $\boldsymbol{Z} = \boldsymbol{AX}$ by p vectors of $\boldsymbol{X}$:

$$\begin{cases} Z_1 = a_{11}X_1 + a_{21}X_2 + \cdots + a_{p1}X_p \\ Z_2 = a_{12}X_1 + a_{22}X_2 + \cdots + a_{p2}X_p \\ \cdots \\ Z_p = a_{1p}X_1 + a_{2p}X_2 + \cdots + a_{pp}X_p \end{cases} \quad (1)$$

where Z_1 is the first principal component, Z_2 is the second principal component, …, and Z_p is the pth principal component.

Eq.(1) satisfies the following conditions:

$a_{1i}^2 + a_{2i}^2 + \cdots + a_{pi}^2 = 1, i = 1, \cdots, p$; Z_i and $Z_j\left(i \neq j; i, j = 1, \cdots, p\right)$ are unrelated; The variance of Z_1 is the biggest, the variance of Z_2 is the second biggest, …, and the variance of Z_p is the smallest.

The general process of calculating principal components is as follows:

- Standardize the original data to eliminate the influence of different dimensions and calculate the correlation matrix ***R***.
- Calculate the characteristic roots which include $\lambda_1, \lambda_2 \cdots \lambda_p$ by correlation matrix ***R*** and order them in decreasing order, such as $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_p$ and then, calculate the corresponding unit characteristic vectors which include $\boldsymbol{a}_1, \boldsymbol{a}_2 \cdots \boldsymbol{a}_p$. The values of unit characteristic vectors are the coefficients of Eq.(1). In this way, $Z_1, Z_2, \cdots Z_p$ can be got. Meanwhile, the data in Eq.(1) is the standardized data.
- Calculate the cumulative contribution rate of the first m characteristic roots as $\sum_{k=1}^{m} \lambda_k \Big/ \sum_{i=1}^{p} \lambda_i$. If the cumulative contribution rate is 80%~90%, the first m principal components can basically contain most information of original indexes and they can be used for replacing the original indexes [15].

B. Fisher discriminant analysis

Fisher discriminant analysis was put forward in 1936. This method has no special requirements for population distribution, so it is suitable for the population with unknown distribution [16].

The basic principle is shown below:

Take two populations for example. Train the samples with p indexes from populations G_1 and G_2, and construct discriminant function $y = c_1x_1 + c_2x_2 + \cdots + c_px_p$. The values of $c_1, c_2 \cdots c_p$ are determined in order to maximize the dispersion between two populations and minimize the dispersion within each population. Calculate the values of pending samples by discriminant function and compare them with the critical value y_0 to determine the samples belonging to which population.

The calculation process is as follows:

- Construct function:

$$I = \frac{Q}{F} = \frac{\left(\overline{y}^{(1)} - \overline{y}^{(2)}\right)^2}{\sum_{i=1}^{n_1}(y_i^{(1)} - \overline{y}^{(1)})^2 + \sum_{i=1}^{n_2}(y_i^{(2)} - \overline{y}^{(2)})^2} \quad (2)$$

where Q is the dispersion between two populations and F is the dispersion within each population.

- Calculate $c_1, c_2 \cdots c_p$ to maximize I through the necessary conditions for extreme value of calculus. Then, Eq.(3) can be got:

$$\boldsymbol{C} = \boldsymbol{S}^{-1}\boldsymbol{D} \quad (3)$$

where $\boldsymbol{C} = \left[c_1, c_1, \cdots, c_p\right]^T$;

$$\boldsymbol{S} = \sum_{i=1}^{2}\sum_{j=1}^{n_i}\left(x_j^{(i)} - \overline{x}^{(i)}\right)\left(x_j^{(i)} - \overline{x}^{(i)}\right)^T ;$$

$$\boldsymbol{D} = \left[d_1, d_2, \cdots, d_p\right]^T = \left[\left(\overline{x}_1^{(1)} - \overline{x}_1^{(2)}\right), \left(\overline{x}_2^{(1)} - \overline{x}_2^{(2)}\right), \cdots, \left(\overline{x}_p^{(1)} - \overline{x}_p^{(2)}\right)\right]^T .$$

In this way, the discriminant function can be obtained.

DOI 10.5013/IJSSST.a.17.49.32 ISSN: 1473-804x online, 1473-8031 print

- Calculate the critical value y_0. The critical value can be got by Eq.(4):

$$y_0 = \frac{n_1 \overline{y}^{(1)} + n_2 \overline{y}^{(2)}}{n_1 + n_2} \quad (4)$$

- Discriminant criterion. Calculate the value y of each pending sample by discriminant function and discriminate the samples by the following criterion:

If $\overline{y}^{(1)} > \overline{y}^{(2)}$ and $y > y_0$, then $X \in G_1$; If $\overline{y}^{(1)} > \overline{y}^{(2)}$ and $y < y_0$, then $X \in G_2$; If $\overline{y}^{(1)} > \overline{y}^{(2)}$ and $y = y_0$, the result could not be determined.

If $\overline{y}^{(1)} < \overline{y}^{(2)}$ and $y > y_0$, then $X \in G_2$; If $\overline{y}^{(1)} < \overline{y}^{(2)}$ and $y < y_0$, then $X \in G_1$; If $\overline{y}^{(1)} < \overline{y}^{(2)}$ and $y = y_0$, the result could not be determined.

C. Evaluation Criterion

This paper uses back substitution method based on training samples to evaluate the discriminant analysis model.

Assume that G_1 and G_2 are two populations and the capacities of them are n_1 and n_2. Calculate the results of $n_1 + n_2$ training samples by discriminant function. The misjudgment rate is calculated by Eq.(5), as shown below:

$$\eta = \left(\overset{*}{n}_{12} + \overset{*}{n}_{21} \right) \Big/ \left(n_1 + n_2 \right) \quad (5)$$

Where $\overset{*}{n}_{12}$ is the number of samples which belong to G_1 but are wrongly discriminated into G_2; $\overset{*}{n}_{21}$ is the number of samples which belong to G_2 but are wrongly discriminated into G_1.

D. Algorithm Flow

The algorithm flow of establishing discriminant analysis model and discriminating pending samples is shown as Fig. 1. Firstly, Calculate the samples with p-dimension vector by PCA, then select the first m principal components whose characteristic roots could meet $\sum_{k=1}^{m} \lambda_k \Big/ \sum_{i=1}^{p} \lambda_i \geq 80\%$ to reduce the dimensions and obtain the comprehensive indexes. Secondly, compute the samples with m-dimension vector by Fisher discriminant analysis to establish the discriminant analysis model and evaluate the model by evaluation criterion. If the model satisfies the evaluation criterion, it can be put into use. If the model does not satisfy the evaluation criterion, the training samples require to be increased to repeat the whole flow. Thirdly, standardize the pending samples and calculate standardized samples by Eq.(6), Eq.(7) and Eq.(8) to reduce the dimensions and get the pending samples with m-dimension vector. Finally, discriminate the pending samples by discriminant analysis model for slope stability of open-pit mine. In this way, the whole discrimination flow based on PCA and Fisher discriminant analysis can be finished.

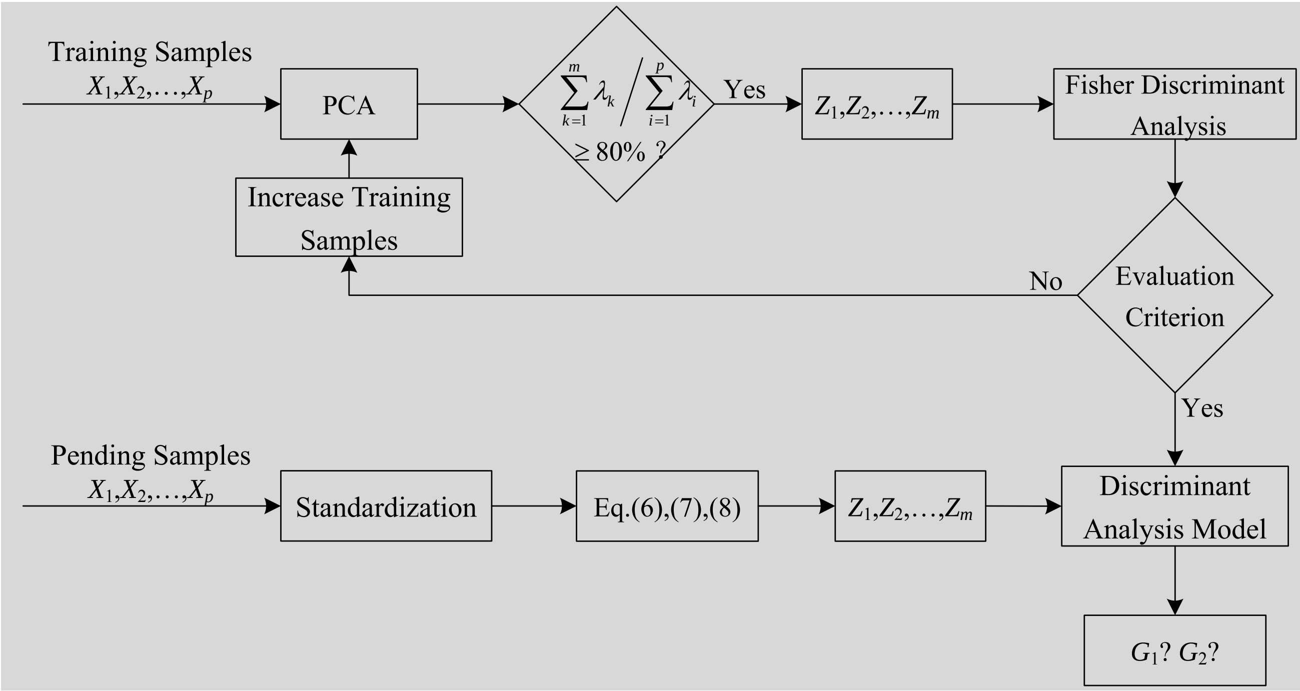

Figure 1. Algorithm flow.

III. Establishment of the Discriminant Analysis Model for Slope Stability of Open-pit Mine

A. Selection of discriminant indexes

The slope of open-pit mine is a complex system and its stability is affected by a variety of factors, such as engineering geological conditions, hydrologic conditions, lithology, geometry, and so on [17]. According to the theoretical research, field investigation and related research results [11], [14], the 6 discriminant indexes which affect the slope stability of open-pit mine including gravity density of rocks X_1 ($\frac{kN}{m^3}$), cohesion force X_2 (kP_a), internal friction angle X_3 (°), slope angle X_4 (°), slope height X_5 (m) and pore water pressure X_6 are selected, considering the geometric factors, geotechnical mechanics index. The types of slopes are divided into stable and destructive.

B. Establishment of the discriminant analysis model

This paper selects 32 sets of samples with clear conclusions in reference [11] as training samples. The samples have complete indexes which include gravity density of rocks, cohesion force, internal friction angle and so on. They can be divided into two populations. The first population includes 14 sets of samples and their type is destructive. The second population includes 18 sets of samples and their type is stable. The distribution of the populations is unknown. The data are shown in Table I.

TABLE I. Data of Slope Stability Discriminant

NO.	Discriminant Indexes						Actual Types
	X_1	X_2	X_3	X_4	X_5	X_6	
1	20.41	33.52	11.00	16.00	45.70	0.20	D
2	18.84	0	20.00	20.00	7.62	0.45	D
3	21.43	0	20.00	20.00	61.00	0.50	D
…	…	…	…	…	…	…	…
12	27.00	40.00	35.00	43.00	420.00	0.25	D
13	27.00	32.00	33.00	42.00	301.00	0.25	D
14	31.30	68.00	37.00	49.00	200.50	0.25	D

15	31.30	68.00	37.00	46.00	366.00	0.25	S
16	31.30	68.60	37.00	47.00	305.00	0.25	S
17	20.41	24.90	13.00	22.00	10.67	0.35	S
…	…	…	…	…	…	…	…
30	27.00	10.00	39.00	41.00	511.00	0.25	S
31	27.00	10.00	39.00	40.00	470.00	0.25	S
32	25.00	46.00	35.00	47.00	443.00	0.25	S

Note: D - destructive; S - stable.

1) Data analysis of PCA for training samples

In engineering practice, the data often have different dimensions. The influence of different dimensions can be eliminated through standardizing the data in Table II. Then, the correlation matrix of samples and the corresponding eigenvalues, contribution rates and cumulative contribution rates of each principal component can be obtained. We order principal components in decreasing order, as shown in Table II.

TABLE II. EIGENVALUES, CONTRIBUTION RATES AND CUMULATIVE CONTRIBUTION RATES OF EACH PRINCIPAL COMPONENT

Principal Component	**Eigenvalue**	**Contribution Rate/%**	**Cumulative Contribution Rate/%**
Z_1	3.528	58.796	58.796
Z_2	0.773	12.876	71.672
Z_3	0.711	11.852	83.524

Table II shows that the bigger the principal component is, the greater the contribution rate is. Generally, when the cumulative contribution rate reaches 80%~90%, the principal components can basically contain most information of original indexes. The cumulative contribution rate of the first three principal components has reached 83.524%, so they are able to reflect the basic information of original indexes well.

The unit characteristic vectors can be calculated by corresponding eigenvalues. Then, we can get the functions of the first, the second and the third principal component as Eq.(6), Eq.(7) and Eq.(8).

$$Z_1 = 0.448X_1 + 0.377X_2 + 0.428X_3 + 0.459X_4 + 0.394X_5 - 0.328X_6 \quad (6)$$

$$Z_2 = -0.033X_1 - 0.145X_2 + 0.5X_3 + 0.192X_4 + 0.097X_5 + 0.825X_6 \quad (7)$$

$$Z_3 = 0.336X_1 - 0.549X_2 - 0.204X_3 - 0.263X_4 + 0.689X_5 + 0.021X_6 \quad (8)$$

Eq.(6) shows that the first principal component has strong positive correlation with X_1 and X_4, which can be regarded as a comprehensive index of gravity density of rocks and slope angle. Eq.(7) shows that the second principal component has strong positive correlation with X_3 and X_6, which can be regarded as a comprehensive index of internal friction angle and pore water pressure. Eq.(8) shows that the third principal component has strong negative correlation with X_2 and strong positive correlation with X_5, which can be regarded as a comprehensive index of cohesion force and slope height.

Calculate the standardized original data through Eq.(6), Eq.(7) and Eq.(8). The results are shown in Table III.

TABLE III. THE RESULTS CALCULATED BY PCA

NO.	**Comprehensive Indexes**			**Actual Types**
	Z_1	**Z_2**	**Z_3**	
1	-2.16	-2.50	0.06	D
2	-3.13	0.44	0.27	D
3	-2.90	0.87	0.71	D
…	…	…	…	…
12	1.77	0.03	0.70	D
13	1.21	-0.13	0.47	D
14	2.50	-0.07	-0.71	D
15	2.76	-0.03	0.04	S
16	2.67	-0.05	-0.25	S
17	-2.52	-0.97	-0.06	S
…	…	…	…	…
30	1.64	0.48	1.71	S
31	1.50	0.43	1.56	S
32	1.88	0.09	0.40	S

Note: D - destructive; S - stable.

In this way, Z_1, Z_2 and Z_3 can be new discriminant indexes insteading of X_1, X_2, X_3, X_4, X_5 and X_6 to achieve the purpose of dimensions reduction.

2) Data analysis of Fisher discriminant analysis for training samples

For the data in Table III, we regard Z_1, Z_2 and Z_3 as the discriminant indexes. The samples belonging to destructive type have been classified into G_1 and the ones belonging to stable type have been classified into G_2. Finally, the discriminant function can be got by Fisher discriminant analysis. The function is shown as Eq.(9).

$$y = -0.019676Z_1 + 0.003502Z_2 - 0.010456Z_3 \quad (9)$$

Eq.(9) is the discriminant analysis model for slope stability of open-pit mine based on PCA and Fisher discriminant analysis. We can get that $y_0 = 0.000013$ and $\overline{y}^{(1)} > \overline{y}^{(2)}$ by Eq.(9) and Eq.(4). However, the discriminant analysis model needs to be evaluated by the evaluation criterion. If the model satisfies the evaluation criterion, the model can be put into use.

C. Evaluation for the model

Calculate the data in Table III by Eq.(9), and compare the results with y_0. Then, the misjudgment rate of this model can be obtained as 21.875% by Eq.(5) and the correct rate can be obtained as 78.125% accordingly. When the correct rate is close to 80%, the established model can be put into use [18].

IV. MODEL APPLICATION

Discriminate the 7 pending samples in reference [11] by discriminant analysis model for slope stability of open-pit mine and discriminant criterion. This model is compared with model of ELM(Extreme Learning Machine) and model of LSSVM(Least Squares Support Vector Machine). The data and results are shown in Table IV. By comparing the calculated results with actual types, the correct rates of model of PCA and Fisher discriminant analysis, model of ELM and model of LSSVM are 100%, 71.43% and 85.71% respectively. It is proved that this model can improve the accuracy of discriminant analysis and it can be used to determine the slope stability of open-pit mine.

TABLE IV. THE DISCRIMINANT RESULTS OF PENDING SAMPLES

NO.	Discriminant Indexes						y	Actual Types	PCA+ Fisher	ELM	LSSVM
	X_1	X_2	X_3	X_4	X_5	X_6					
1	20.00	20.00	36.00	45.00	50.00	0.25	0.019016	D	D	D	D
2	19.63	11.97	20.00	22.00	21.19	0.40	0.060895	D	D	D	D
3	21.82	8.62	32.00	28.00	12.80	0.49	0.049481	D	D	S*	S*
4	25.00	55.00	36.00	45.00	299.00	0.25	-0.024775	S	S	D*	S
5	27.30	10.00	39.00	40.00	480.00	0.25	-0.042631	S	S	S	S
6	25.00	46.00	35.00	46.00	393.00	0.25	-0.030741	S	S	S	S
7	25.00	48.00	40.00	49.00	330.00	0.25	-0.031244	S	S	S	S

Note: D - destructive; S – stable; * - the sample of miscalculation.

V. CONCLUSIONS

- This paper discriminates the slope stability of open-pit mine by statistical method and reduces the dimensions of discriminant indexes by PCA to extract the main information and eliminate the mutual influence of the discriminant indexes. In this way, we can use fewer indexes to represent influence factors which affect the slope stability of open-pit mine to effectively solve the inconvenience of too many indexes. Finally, the discriminant analysis model for slope stability of open-pit mine is built by Fisher discriminant analysis. The more original discriminant indexes are, the greater the advantage of this method is. The standardization for original data by PCA can eliminate the influence of different dimensions. In most cases, the distribution of data is usually unknown. Since Fisher discriminant analysis has no special requirements for data distribution, it is suitable for actual situation.
- The correct rate of discriminating pending samples through the model in this paper has reached 100%, which can prove the feasibility of the method. Compared with other models, the model can improve the accuracy of discriminant analysis.
- The sample data is directly related to the accuracy and reliability of the model, therefore, we should collect data fully and widely to extend the sample database in the further research. In this way, the accuracy of discriminant analysis model can be further improved.

REFERENCES

[1] F. H. Gu, "Failure types and damage scale analysis of bench slope based on geometrical characteristics", Metal Mine, No. 7, pp. 51-55, 2014.

[2] W. J. Yu, and Q. Gao, "Application of a synthetic monitoring technique to high and steep slopes", Chinese Journal of Engineering, vol. 32, No. 1, pp. 14-19, 2010.

[3] L. D. Wu, A. J. Su, X. Huo, and Z. Y. Qi, "Comparison and analysis of slope safety factor by different limit equilibrium methods", Water Resources and Power, vol. 31, No. 12, pp. 151-154, 2013.

[4] M. M. Farias, and D. J. Naylor, "Safety analysis using finite elements", Computer and Geotechnics, vol. 22, No. 2, pp. 165-181, 1998.

[5] J. M. Duncan, "State of the art: limit equilibrium and finite-element analysis of slopes", Journal of Geotechnical Engineering, vol. 122, No. 7, pp. 577-596, 1996.

[6] B. F. Kong, H. N. Ruan, Z. D. Zhu, W. J. Yuan, and Z. Z. Chen,

"Slope stability analysis by strength reduction based on distinct element method", Yellow River, vol. 35, No. 4, pp. 120-123, 2013.
[7] T. L. Li, D. G. Yang, C. L. Zhang, H. Wang, and Y. Q. Zhang, "Block theory based stability analysis of metamorphic schist rock slopes", Journal of Engineering Geology, vol. 22, No. 5, pp. 896-902, 2014.
[8] P. Li, B. G. Wang, and T. L. Li, "Study on analogism used in highway cutting loess slope design", Journal of Highway and Transportation Research and Development, vol. 26, No. 2, pp. 1-5, 2009.
[9] L. Q. Chen, and Z. J. Zhang, "Application of graphic method and numerical analysis method in slope stability evaluation in open pit mine", Nonferrous Metals (Mining Section), vol. 64, No. 2, pp. 73-79, 2012.
[10] W. D. Gao, and L. J. You, "Assessment of slope stability based on distance discriminant analysis model", Nonferrous Metals (Mining Section), vol. 61, No. 3, pp. 49-50, 2009.
[11] Z. Y. Luo, X. J. Yang, and X. N. Gong, "Support vector machine model in slope stability evaluation", Chinese Journal of Rock Mechanics and Engineering, vol. 24, No. 1, pp. 144-148, 2005.
[12] Z. H. Xie, S. S. Liang, and X. D. Zhang, "Early warning method of slope instability of open-pit mine based on RBF neural network", Metal Mine, No. 9, pp. 7-10, 2014.
[13] Y. S. Li, W. F. Chen, X. P. Li, C. G. Tian, J. Xia, and Y. H. Guo, "Stability assessment of rock slope based on fuzzy neural network", Journal of Wuhan University of Technology, vol. 35, No. 1, pp. 113-118, 2013.
[14] T. X. Wen, and B. Zhang, "Prediction model for open-pit coal mine slope stability based on random forest", Science & Technology Review, vol. 32, No. 4/5, pp. 105-109, 2014.
[15] G. M. Liu, and W. R. Lin, Introduction and experiment of multivariate statistics. Hangzhou: Zhejiang University Press, 2013, pp. 49-55.
[16] Y. Guan, Applied multivariate statistical analysis. Hangzhou: Zhejiang University Press, 2011, pp. 139-145.
[17] S. H. Zhai, A. X. Wu, Q. Gao, M. H. Zhang, and L. Dong, "Prediction of slope safety factor based on the RS-GP model", Chinese Journal of Engineering, vol. 33, No. 1, pp. 6-10, 2011.
[18] X. Z. Shi, J. Zhou, W. Zheng, H. Y. Hu, and H. Y. Wang, "Bayes discriminant analysis method and its application for prediction of slope stability", Journal of Sichuan University (Engineering Science Edition), vol. 42, No. 3, pp. 63-68, 2010.