IETI Transactions on Computers, 2016, Volume 2, Issue 2, 73-81.
http://www.ieti.net/tc
An International Open Access Journal

Application of Fisher Discriminant Analysis in Safety Evaluation

ZHU XiaoZhen[1, a]

[1]Wuhan University of Technology, Wuhan,China

[a]startlife321@163.com

Keywords: Fisher discriminant analysis, Bayes discriminantanalysis, safety evaluation.

Abstract. The multivariate statistical method of Fisher discriminant analysis is applied to safety evaluation, through the analysis of the original data, the assessment process, built up to reflect the evaluated object security status of evaluation function model, so as to simplify the subsequent similar evaluation target workload. The two mine in south of a mining enterprise subordinate to the environmental conditions in six integrated index evaluation, comprehensive index function model is established, finally, the Fisher discrimination obtained results with Bayesian discriminant obtained results, the correctness of the model is verified that the model reliability is high, and simple and practical.

Introduction

Discriminant analysis is in the known research object is divided into several types and has achieved a number of various types of a number of known samples of observations based on, according to certain criteria and establish discriminant and samples of unknown type were discriminant analysis.[1]Fisher discriminant is used for safety evaluation, the original evaluation object to establish a discriminant function, in order to simplify the subsequent similar evaluation, to reduce the workload, to achieve the purpose of rapid safety evaluation of similar issues. In order to verify the accuracy of the discriminant function, compared with the results of Bayesian discriminant, the results are reliable and practical.

Fisher Discriminant Principle

Discriminant analysis is refers to the pre-established criteria and establish discriminant function classification of known, unknown classification of individual observations are substituted into the corresponding variables. According to the criteria (such as function value is greater than a certain value) to classify an unknown individual should belong to the known classification in which a method of judging the.[2]

Fisher discriminant basic idea: in order to overcome due to the high dimension caused by the "Curse of dimensionality", high-dimensional data points are projected to low dimensional space (such as one-dimensional, which can achieve data intensive. Fisher discriminant analysis method is based on the number of existing observations of the number of features (discriminant factor) to identify new samples, to determine the properties of the forecast analysis method.[3]Because of the discrimination method of the original data distribution has no special requirements, so it is very suitable to do not know in advance the distribution of sample, and can consider the various factors affecting the comprehensive judgment.

Multi overall Fisher discriminantfunction

Assume that there arekgeneralincluding $G_1,\ldots,G_k,n_1,n_2,\ldots n_k$ are the sample numbers.

$x_\alpha^{(i)} = \left(x_{\alpha 1}^{(i)}, \cdots, x_{\alpha p}^{(i)}\right)$ is an observation vector of second sample in first total.Modelof discriminantfunction is $y(X) = c^T X$ (when there are k in general there are $(k-1)$discriminantfunctions).$c = \left(c_1, c_2, \cdots, c_p\right)^T$,$x = \left(x_1, x_2, \cdots, x_p\right)$,$X^{(i)} = \left(x_1^{(i)}, x_2^{(i)}, \cdots, xpi,\right.$

$X = \left(x_1, x_2, \cdots, x_p\right).$

$$n = n_1 + n_2 + \cdots + n_k. \quad (1)$$

$$y^{(i)} = c^T X^{(i)} (i = 1,2, \cdots, k), y = c^T X \quad (2)$$

In more general conditions,sum of squares between groups$Q_1 = c^T Bc$,between Group Sum of Squares$Q_2 = c^T Lc$. Formula (3) and formula (4) respectively represent the difference between groups and within group difference.

$$B = \sum_{i=1}^{k} n_i \left(X^{(i)} - X\right)\left(X^{(i)} - X\right)^2 \quad (3)$$

$$L = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(X_j^{(i)} - X^{(i)}\right)\left(X_j^{(i)} - X^{(i)}\right)^T \quad (4)$$

Fisher criterion is to select the coefficient vector c, so that the objective function$\lambda(c) = \frac{Q_1}{Q_2}$to achieve maximum value.For the sake ofλmaximum value,according to the necessary condition for the existence of an extremal,so $\frac{\partial\lambda}{\partial c} = 0$, $Bc = \lambda Lc$, λ and c is exactly the eigenvector corresponding generalized characteristics ofB, L matrix and the root.

$$(B - \lambda L)c = 0 \rightarrow (L^{-1}B - \lambda E) = 0 \tag{5}$$

The largest eigenvalue of $L^{-1}B$is obtained from the root λ_1 and the corresponding feature vector c_1, then the first discriminant function is $y_1 = c_1^T X = \sum_{j=1}^{P} c_j x_j$.In the same way, other discriminant function$y_t = c_t^T X$ $[t = 1, 2, \cdots, r(r \leq p)]$, including non-zero characteristic root

$\lambda_1 \geq \lambda_2 \geq, \cdots, \lambda_r$,corresponding to the characteristic vector of $c_1, c_2, \cdots, c_r (r \leq p)$.

The greater the characteristic function is, the stronger the ability of discriminant function is stronger. Discriminant function from $y_1$to$y_t$discriminant abilitygradually weakened, the cumulative discriminant ability of the former m discriminant function is the formula (6).

$$sp_m = \sum_{l=1}^{m} pl = \frac{\sum_{l=1}^{m} \lambda_l}{\sum_{i=1}^{p} \lambda_i} (l = 1, \cdots, p) \tag{6}$$

After thediscriminant function is established,the classification of the test samples should be treated.The first discriminant function of the discriminant ability to a certain extent, can be calculated for each type ofy_1.$y_1^{(1)}, y_1^{(2)}, \cdots, y_1^{(k)}$from re ranking as $y_{(1)} \leq y_{(2)} \leq \cdots \leq y_{(k)}$, between $G_{ij}$and G_{ij+1} get the demarcation point according to the number of two adjacent weighted average as formula (7).If the sample xmakes$d_{i-1, i} \leq y(x) \leq d_{i, i+1}$, $x \in G_i$.[4]

$$d_{i, i+1} = \frac{n_i y_{(i)} + n_i y_{(i+1)}}{n_i + n_i} (i = 1, \cdots, k - 1) \tag{7}$$

Two general Fisher discriminant

N is a d dimension samples, $x_1, x_2, \cdots, x_N$, of which N_1 belonged to the ω_1 sample as a subset of X_1,N_2belong to the sample ofω_2,recorded asX_2.All kinds of sample mean vector, such as the formula (8).In the sample class, the dispersion matrix S_i and the total class scatter matrix S_ω are (9) and(10) respectively.In addition,scatter matrix between samplesis expressed by the formula (11).[5]

$$\mu_i = \frac{1}{N_i} \sum_{x \in X_i} x, \quad i = 1, 2 \tag{8}$$

$$S_i = \sum_{x \in X_1} (x - \mu_i)(x - \mu_i)^T, \quad i = 1, 2 \tag{9}$$

$$S_\omega = S_1 + S_2 \quad (10)$$

$$S_b = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T \quad (11)$$

Establish discriminant, and determine the critical value ofy_0.

$$y = (\mu_1 - \mu_2)^T S_w^{-1} x \quad (12)$$

$$y_0 = \frac{n_1 \bar{y}^{(1)} + n_2 \bar{y}^{(2)}}{n_1 + n_2} \quad (13)$$

Criteria for: given a new data $x = (x_1, \; x_2, \; \cdots, \; x_p)^T$ to the set of the discriminant function to get the results labeled as y.If $y > y_0$, $x \in G_1$,otherwise, $x \in G_2$.If$\bar{y}^{(1)} < \bar{y}^{(2)}$,it is the criteriathat$y > y_0$, $x \in G_2$, otherwise$\in G_1$.[6]

Constructing Evaluation Function and Discriminant Analysis

A Southern Mining Group, carries out the evaluation of the factors of its subordinate enterprises of the two mine environmental conditions, according to previous experience and the characteristics of coal mine in the south, take the evaluation factors: X_1,roadway passing rate(%);X_2, dust concentration (mg/m^3); x 3, ambient temperature (℃); x 4, wind speed (m/s); x 5, roadway minimum pedestrian width (m); x 6, roadway minimum pedestrian height (m).The original data (using the scoring method) and the index system of the 6 comprehensive indexes were evaluated in the table 1~ table 8.[7]

Table1Pass Rate of roadway (%)

Index	>95	90~95	85~90	80~85	<80
Level	5.00	4.00	3.00	2.00	1.00

Table2Dust Concentration(mg/m^3)

Index	<4	4~6	6~8	8~10	>10
Level	5.00	4.00	3.00	2.00	1.00

Table 3 Ambient Temperature(℃)

Index	18~22	22~24	24~26	26~28	>28
Level	5.00	4.00	3.00	2.00	1.00

Table 4 Wind Speed (m/s)

Index	2.5~3.5	2~2.5	1.5~2	1.5~1	<1
Level	5.00	4.00	3.00	2.00	1.00

Table 5Minimum Pedestrian Width of roadway (m)

Index	>1.2	1.1~1.2	1~1.1	0.8~1.0	<0.8
Level	5.00	4.00	3.00	2.00	1.00

Table6 Minimum Pedestrian Height (m)

Index	>1.8	1.6~1.8	1.4~1.6	1.2~1.4	<1.2
Level	5.00	4.00	3.00	2.00	1.00

Table7Level of Safety Evaluation

Level	1	2	3	4	5
Content	Unsafe	Less secure	General security	More secure	Safety

Table 8Raw Data

Data	Grube Number	X_1	X_2	X_3	X_4	X_5	X_6	E(*)
	1	94.87	4.03	23.1	2.01	1.17	1.79	4
	2	93.15	5.35	22.7	2.32	1.19	1.72	4

3	91.57	4.89	22.2	2.21	1.13	1.68	4
4	90.78	5.87	23.8	2.48	1.10	1.60	4
5	91.45	4.55	22.9	2.23	1.14	1.67	4
6	92.67	4.70	23.0	2.42	1.13	1.70	4
7	85.10	6.87	25.2	1.63	1.09	1.59	3
8	86.54	7.91	24.0	1.75	1.00	1.41	3
9	87.69	6.17	25.9	1.98	1.05	1.47	3
10	89.34	7.32	24.3	1.55	1.07	1.52	3
11	87.77	7.12	25.1	1.67	1.03	1.46	3
12	86.67	7.23	25.3	1.72	1.05	1.48	3

Using two general Fisher discriminant model and principle, the calculation is as follows:

(1) Calculate the two kinds of security evaluation level mean and covariance matrix

$$\mu_1 = (92.415 \quad 4.898 \quad 22.95 \quad 2.278 \quad 1.143 \quad 1.693)^T \tag{14}$$

$$\mu_2 = (87.185 \quad 7.103 \quad 24.967 \quad 1.717 \quad 1.048 \quad 1.488)^T \tag{15}$$

$$A = \begin{pmatrix} 2.455 & -0.868 & 0.15 & 2.01 & 0.027 & 0.097 \\ 0.735 & 0.452 & -0.25 & 2.23 & 0.047 & 0.027 \\ -0.845 & -0.008 & -0.75 & 2.21 & -0.013 & -0.013 \\ -1.635 & 0.972 & 0.85 & 2.48 & -0.043 & -0.093 \\ -0.965 & -0.348 & -0.05 & 2.23 & -0.003 & -0.023 \\ 0.255 & -0.198 & 0.05 & 2.42 & -0.013 & 0.007 \end{pmatrix} \tag{16}$$

$$B = \begin{pmatrix} -2.085 & -0.233 & 0.233 & -0.087 & 0.042 & 0.102 \\ -0.645 & 0.807 & -0.967 & 0.033 & -0.048 & -0.078 \\ 0.505 & -0.933 & 0.933 & 0.263 & 0.002 & -0.018 \\ 2.155 & 0.217 & -0.667 & -0.167 & 0.022 & 0.032 \\ 0.585 & 0.017 & 0.133 & -0.047 & -0.018 & -0.028 \\ -0.515 & 0.127 & 0.333 & 0.003 & 0.002 & -0.008 \end{pmatrix} \tag{17}$$

$$S_1 = \begin{pmatrix} 6.885 & 1.367 & -2.163 & -4.794 & -2.064 & 0.768 \\ 1.367 & 0.812 & -0.441 & -0.971 & -0.857 & 0.091 \\ -2.163 & -0.441 & 1.282 & 0.724 & 0.859 & -0.261 \\ -4.794 & -0.971 & 0.724 & 4.392 & 1.190 & -0.538 \\ -2.064 & -0.857 & 0.859 & 1.190 & 1.058 & -0.187 \\ 0.768 & 0.091 & -0.261 & -0.538 & -0.187 & 0.127 \end{pmatrix} \tag{18}$$

IETI Transactions on Computers, 2016, Volume 2, Issue 2, 73-81.
http://www.ieti.net/tc
An International Open Access Journal

$$S_2 = \begin{pmatrix} 4.476 & 0.919 & -0.643 & -4.680 & -1.192 & 1.121 \\ 0.919 & 2.012 & -1.971 & -0.579 & -0.491 & 0.113 \\ -0.643 & -1.971 & 2.006 & 0.219 & 0.392 & -0.067 \\ -4.680 & -0.579 & 0.219 & 5.165 & 1.182 & -1.305 \\ -1.192 & -0.491 & 0.392 & 1.182 & 0.364 & -0.255 \\ 1.121 & 0.113 & -0.067 & -1.305 & -0.255 & 0.392 \end{pmatrix} \quad (19)$$

$$S_\omega = \begin{pmatrix} 11.360 & 2.286 & -2.806 & -9.475 & -3.256 & 1.888 \\ 2.286 & 2.824 & -2.412 & -1.550 & -1.348 & 0.204 \\ -2.806 & -2.412 & 3.347 & 0.943 & 1.251 & -0.328 \\ -9.475 & -1.550 & 0.943 & 9.557 & 2.372 & -1.843 \\ -3.256 & -1.348 & 1.251 & 2.372 & 1.421 & -0.441 \\ 1.888 & 0.204 & -0.328 & -1.843 & -0.441 & 0.519 \end{pmatrix} \quad (20)$$

$$S_\omega^{-1} = \begin{pmatrix} 44116.128 & 44021.739 & 44104.331 & 44049.830 & 44137.352 & 43975.604 \\ 44021.739 & 43928.994 & 44010.874 & 43955.903 & 44043.447 & 43883.104 \\ 44104.331 & 44010.874 & 44093.650 & 44038.471 & 44125.209 & 43965.392 \\ 44049.830 & 43955.903 & 44038.471 & 43984.209 & 44070.546 & 43911.300 \\ 44137.352 & 44043.447 & 44125.209 & 44070.546 & 44160.818 & 43996.544 \\ 43975.604 & 43883.104 & 43965.392 & 43911.300 & 43996.544 & 43843.947 \end{pmatrix} \quad (21)$$

(2) Obtaining the coefficients of Fisher discriminant function by using $y = (\mu_1 - \mu_2)^T S_\omega^{-1} x$.The coefficients are c_1, c_2, c_3, c_4, c_5, c_6,then we can get

$$y = 82532.747x_1 + 82351.713x_2 + 82506.956x_3 + 82407.810x_4 + 82571.872x_5 + 82265.936x_6$$

See Table 9 for the discrimination data.

Table 9Grubeto be evaluated

Samples to be evaluated	Grube Number	X_1	X_2	X_3	X_4	X_5	X_6	Evaluation Result
	1	85.54	7.12	25.3	1.67	1.15	1.37	
	2	91.57	5.39	23.1	2.11	1.13	1.58	

(3) According to the data of two not evaluated,determine the critical value of y_0 and classify samples.

$$\bar{y}^{(1)} = 10078904.399\ ,\bar{y}^{(2)} = 10304476.942\ ,y_0 = \frac{n_1\bar{y}^{(1)}+n_2\bar{y}^{(2)}}{n_1+n_2} = 10191690.670$$

Discriminant results are shown in Table 10.

Table 10 DiscriminantResult

Number	y	Result
1	10078904.399	3
2	10304476.942	4

Comparative Analysis

The evaluation model is established by using Bayesian discriminant analysis to evaluate the 2 mines:

$$Z_4(X) = \ln(1/5) - 1918.4435 + 23.836X_1 + 67.68X_2 + 16.116X_3 + 67.274X_4 + 700.09X_5 - 21.12X_6$$

$$Z_3(X) = \ln(1/5) - 1869.1269 + 22.944X_1 + 70.054X_2 + 18.21X_3 + 60.741X_4 + 688.72X_5 - 26.144X_6$$

The evaluation results are shown in Table 11.

Table 11 BayesDiscriminant Results

Samples to be evaluated	Grube Number	X_1	X_2	X_3	X_4	X_5	X_6	Evaluation Result
	1	85.54	7.12	25.3	1.67	1.15	1.37	3
	2	91.57	5.39	23.1	2.11	1.13	1.58	4

The obtained results were compared with theresults of Bayesian discriminant,the results were consistent,and the evaluation model established by Fischer discriminant function was safe and reliable.

In summary,Fisher discriminant method is based on the more rigorous statistical classification theory, when the training samples for a long time, it is expected to achieve more accurate classification results.[8]Analysis to establish the evaluation function model is suitable for the use of

Fisher discriminant method is simple and reliable. In the safety assessment of the application of Fisher discriminant analysis not only simplify the problem, can also be used for simple safety evaluation, for the establishment of different characteristics, regional characteristics of different evaluation function is different, in use should be closely combined with the actual situation, after the establishment of the model should be continuously updated optimization model, makes the results more reliable and effective.

References

[1]Xiu Lin, mathematical statistics, cedar, et al. Multivariate statistical analysis [M]. China Statistical Publishing House, 1999

[2]YuanZhifa, Zhou Jing Yu. Multivariate statistical analysis [M]. Beijing: Science Press, 2002:216-230

[3]Dong Long Jun, Li Xibing, Bai Yunfei. A fisher discriminant analysis model for classifying of roof coal cavability of steep seam[J].Journal of China coal society, 2009, 34 (1): 58-63.

[4]MiuJianbo, Xu open. Cluster analysis and discriminant analysis in safety evaluation of the comprehensive application of [J]. industrial safety and environmental protection, 2007, 33 (9): 16-18.

[5]Pan Jinsong. Fisher's discriminant analysis and application of [J]. mathematics in practice and understanding, 2013,05:155-162.

[6]The research and application of Zhao Lina.Fisher discriminant method [D]. Tutor: Xu Wenke. Northeast Forestry University, 2013

[7]Lei Jing, Shen Feimin.Bayes discriminant analysis using the [J]. theory of industrial safety and environmental protection in safety evaluation, 2004, 30 (5): 39-40.

[8]Shi Xiuzhi, Zhou Jian. Evaluation of safety and reliability of mine ventilation system using Fisher discriminant method [J]. Journal of mining and safety engineering, 2011, 27 (4): 562-567.