

Application of principal component analysis in fire risk prediction of stadium

Huajie Chen¹, Mingchang Liu^{2*}

¹ Physical Education Department, Jinggangshan University, Jian 343009, Jiangxi, China

² Department of Physical Education, Huazhong Agricultural University, Wuhan 430070, Hubei, China

Received 1 June 2014, www.cmnt.lv

Abstract

The security of stadium is one of the problems increasingly concerned by people. Among various security problems, the fire risk of stadium is the most important. By using principal component analysis (PCA) and combining the influencing factors of risk prediction of stadium, several influencing factors of the fire risk of stadium are evaluated and analysed, then the uppermost principal component factors are selected, the inconspicuous components are eliminated, the influence of relevant factors on the fire risk prediction of stadium is analysed, and the correlation of various indicators with fire risks of stadium is understood to find the potential best security management prediction system, for the purpose of taking pertinent prediction measures of fire accident risks to overcome the unfavourable factors in the security management, and ensure the safe and normal operation of stadium as well as the personal and property security of sporters.

Keywords: Principal Component Analysis, Stadium, Fire Risk, Prediction

1 Introduction

The principal component analysis (PCA) is a kind of statistical analysis method that converts multiple original variables into several comprehensive indicators [1]. From the mathematical perspective, it is a dimension reduction technology [2]. Its general idea is: a research object is often a multi-element complex system. Too many variables will certainly increase the difficulty and complexity of problem analysis [3]. It is a way to simplify the problem by using the correlation between original variables; replacing more original variables with less new variables and making the fewer variables preserve the information reflected by more original variables as more as possible.

The security management of large stadium is an important issue [4]. As a large public place, the large stadium is an indispensable part of people's spiritual entertainment, characterized by highly populated people and big liquidity, so it often causes significant casualty and property loss in case of any fire, thus bringing great negative impacts [5]. Therefore, the fire accident in the safety management is the most serious. In recent years, the stadiums in China have suffered more than 200 fires, with a death of more than 1000 people, which fully reveals the serious problem in the security management of stadium [6]. The academic circles and manager pay more and more attention to seek the way of reducing fire risks by analysing the influencing factors of the fire risk prediction of stadium. To better ensure the security of stadium, it is required to strengthen the research on the reasonability and scientific of the fire risk prediction of

stadium [7]. In order to analyze the influencing factors of the fire risk prediction of stadium, it is required to conduct research and analysis from different theories, thus introducing the application of PCA in the fire risk prediction of stadium [8]. PCA converts many indicator factors into several important comprehensive indicators through the thought of dimension reduction. In practical problems, the primary and secondary cannot be mastered because of the complicated and changing conditions as well as more influencing factors, so it is hoped that there are less variable indicators in the quantitative analysis and more information can be got from these indicators [9]. This can be met by analytic hierarchy process (AHP) as it can explain the most original variables with less indicator variables, converts the indicators with high correlation into independent indicator variables, without any correlation and also can well explain most variables [10].

PCA is one of the important methods of analysing problems in the academic circles at present, which can select according to the correlation degree of influencing factors to reveal the intrinsic connection and difference of objective things. By studying the relevant factors that influence the fire risk of stadium, the application of PCA in the fire risk prediction of stadium can be explored, which is of great realistic significance in the fire risk management and security system of stadium.

* *Corresponding author* e-mail: liumchang@yeah.net

2 Overview of PCA

2.1 BASIC IDEA OF PCA

The basic idea of PCA is to recombine, divide and convert more original P indicators X1, X2, ..., XP with correlation into a set of less representative indicators, without any correlation. Finally, the fewer indicators without any correlation are used to replace more original indicators.

Suppose that F1 means the principal component indicator formed by the first linear combination of original variables, that is, $F_1 = a_{11}X_1 + a_{21}X_2 + \dots + a_{p1}X_p$, it is known by mathematical knowledge that the information got from each principal component can be measured by its variance. The larger the variance Var(F1) is, the more the information contained in F1 will be. Generally, it is hoped that the principal component F1 contains the largest information, so F1 selected from the linear combination should be the one with the largest variance in the linear combinations of X1, X2, ..., XP, so F1 is the first principal component. If the first principal component fails to represent the information of original p indicators, the second principal component indicator F2 can be selected. To effectively reflect the original information, the existing information of F1 will not occur in F2, that is, F2 and F1 should be independent and uncorrelated. In the mathematical language, Cov(F1, F2)=0, so F2 is one with the largest variance in the linear combinations of X1, X2, ..., XP without any correlation with F1, and F2 is called the second principal component. F1, F2,, Fm got by this analogy are the first, second, ..., m principal component of original variable indicators X1, X2.....XP.

$$\begin{cases} F_1 = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p \\ F_2 = a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p \\ \dots \\ F_m = a_{m1}X_1 + a_{m2}X_2 + \dots + a_{mp}X_p \end{cases} \quad (1)$$

It is known from the above analysis that:

(1) Fi and Fj are uncorrelated, that is, Cov(Fi, Fj) = 0 and Var(Fi)=ai`Σai, in which Σ is the covariance matrix of X.

(2) F1 is of the largest variance in the linear combinations of X1, X2, ..., Xp (the coefficient meets the above requirements), that is, Fm is of the largest variance in the linear combinations of X1, X2, ..., Xp, without any correlation with F1, F2,, Fm-1.

F1, F2, ..., Fm (m≤p) are the new variable indicators, that is, the first, second, ..., m principal component of original variable indicators.

2.2 MAIN FEATURE OF PCA

When PCA is used to analyse the research problems, there are two main characteristics:

(1) Obtain the equation of each principal component Fi (i = 1, 2, ..., m) relative to the original variable Xj (j =

1, 2, ..., p), namely, coefficient a_{ij} (i = 1, 2, ..., m; j = 1, 2, ..., p). By using the mathematical knowledge, it is proved that the variance of principal component is the characteristic root of original variable covariance matrix, so the front m larger characteristic roots represent the front m larger principal component variances; the corresponding eigenvectors of front m characteristic roots λ_i of original variable covariance matrix (to ensure the largest variance of principal component in turn) is the coefficient of corresponding principal component Fi expression ai. To limit, the coefficient of ai adopts the corresponding unit eigenvector of λ_i .

(2) Calculate the principal component load, which reflects the correlation degree of principal component Fi and original variable Xj:

$$P(Z_k, x_i) = \sqrt{\lambda_k} a_{ki} (i, = 1, 2, \dots, p; k = 1, 2, \dots, m)$$

2.3 CALCULATION STEPS OF PCA

2.3.1 Obtain the covariance matrix

The original variable covariance matrix of sample is $\Sigma=(s_{ij})p \times p$, where

$$s_{ij} = \frac{1}{n-1} \sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j) \quad I, j = 1, 2, \dots, p \quad (2)$$

2.3.2 Obtain the characteristic value λ_i of covariance matrix and corresponding eigenvector a_i under orthogonally conditions

The front m larger characteristic values of matrix, $\lambda_1 \geq \lambda_2 \geq \dots \lambda_m > 0$ are the corresponding variances of front m principal components, the corresponding unit eigenvector a_i of λ_i is the coefficient of principal component Fi relative to the original variable, so the principal component Fi of i original variable is:

$$F_i = X \quad (3)$$

The variance contribution of principal component generally represents the reflection information, the greater the contribution is, the more the information will be. α_i is:

$$\alpha_i = \lambda_i / \sum_{i=1}^m \lambda_i \quad (4)$$

2.3.3 Choose the principal component

The most principal component among various indicators, that is, m in F1, F2,, Fm should be confirmed by calculating the accumulated variance distribution.

$$G(m) = \sum_{i=1}^m \lambda_i / \sum_{k=1}^p \lambda_k \quad (5)$$

Generally, when the accumulated variance distribution is more than 85%, the information of original variable can be better reflected and m is the number of principal components selected finally.

2.2.4 Confirm the principal component load

The principal component load presents the correlation between principal component F_i and original variable X_j and the load l_{ij} ($i=1, 2, \dots, m; j=1, 2, \dots, p$) of original variable X_j ($j=1, 2, \dots, p$) on the principal component F_i ($i=1, 2, \dots, m$) is as follows:

$$l(Z_i, X_j) = \sqrt{\lambda_i} a_{ij} \quad (i=1, 2, \dots, m; j=1, 2, \dots, p). \quad (6)$$

Then, SPSS software can be used for calculation and analysis and the "component matrix" obtained is the principal component load matrix.

2.2.5 Calculate the score of principal component

Based on the calculation of principal component, the score of principal component is calculated:

$$F_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p \quad i = 1, 2, \dots, m \quad (7)$$

However, in the actual problem, each indicator has different dimensional effects, so it is required to eliminate the influence of dimension on it before calculating the score. Generally, the dimensional influence of data is eliminated through the standardization treatment of original data, that is, the data conversion is transformed as follows:

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{s_j} \quad i = 1, 2, \dots, n; j = 1, 2, \dots, p \quad (8)$$

According to the mathematical expression and relevant mathematic principle, on one hand, after standardized transformation, the variable covariance matrix can express its relevant coefficient matrix; on the other hand, the standardized covariance is the relevant coefficient of original variable, so it is concluded that after the standardized treatment of original variable, the relevant coefficient matrix is of no change.

All in all, the corresponding characteristic value λ_i of the relevant coefficient matrix of original indicator is the variance contribution of principal component and the variance distribution is $\alpha_i = \lambda_i / \sum_{i=1}^p \lambda_i$. The larger α_i is, the stronger the ability of corresponding principal component to reflect the comprehensive information. The principal component can be selected according to λ_i . The combination coefficient (load of original variable on the principal component) a_i of each principal component is the corresponding unit eigenvector of characteristic value λ_i .

3 Comprehensive application of PCA in fire risk warning of stadium

In the research on fire accident of stadium, the evaluation and assessment of influencing factors of fire risk warning are an essential part in our daily warning. With the social development, the evaluation system and method for fire risk warning become diversified, however, most of which are limited to the investigation and analysis phase and rarely conducts further analysis and verification from the theoretical perspective. According to the above situation, the fire risk system of PCA is established and the subjective analysis is integrated with the scientific component of mathematical model to realize the change from qualitative analysis accurate quantitative analysis.

3.1 INDICATOR SELECTION OF STADIUM

The influencing factors of fire risk of stadium from 1990-2013 are calculated by using PCA and the main indicators are selected. According to the actual situation of stadium, 9 main indicators are selected finally as follows:

TABLE 1 Indicator of fire risk prediction model of stadium

Indicator	Meaning
X1	Electric fire
X2	Poor internal fire security inspection
X3	Fail to build the stadium according to the fire design
X4	Incomplete patrol
X5	Poor automatic fire warning system
X6	Poor fire acceptance of stadium
X7	Poorly implemented fire security and regulations of stadium
X8	Insufficient fire extinguisher or failure of fire extinguisher
X9	Fail to take actions immediately

3.2 SPECIFIC STEPS OF PCA IN FIRE RISK

According to the above elaboration, to eliminate the dimensional influence, the covariance matrix is calculated after standardization of variable, that is, the relevant coefficient matrix of original variable is calculated directly, so the calculation steps of PCA in fire risk of stadium are as follows:

- (1) Calculate the relevant coefficient matrix;
- (2) Obtain the characteristic value λ_i and corresponding orthogonal unit eigenvector a_i of relevant coefficient matrix;
- (3) Calculate the weight of each component;
- (4) Choose the principal component.

3.3 APPLICATION EXAMPLE OF PCA IN FIRE RISK

In the research on fire risk prediction of stadium, 9 main indicators among various risk influencing factors are selected to discuss the application of PCA.

According to the specific steps of PCA, the relevant coefficient matrix is obtained after standardizing the

original indicator data and then SPSS software is used to calculate the relevant coefficient matrix and characteristic value of 9 evaluation indicators to confirm the main factor number of evaluation. The number of principal component is confirmed according to the accumulated variance contribution of characteristic value.

After eliminating the dimensional influence of original data of fire risk influencing factors of stadium, the standardized data is obtained (see table 2). The relevant matrix coefficient is shown in table 3.

The characteristic root, contribution and accumulated contribution of relevant coefficient matrix are obtained through further calculation.

It is known from the characteristic equation $|R - \lambda I_9| = 0$ that 9 indicators are non-negative characteristic roots, that is, $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots \geq \lambda_9 \geq 0$, where $\lambda_1 = 7.7719$, $\lambda_2 = 0.756$, $\lambda_3 = 0.3796$, $\lambda_4 = 0.077$ and $\lambda_5 = 0.0155$. The results are shown in table 4.

According to the characteristic root of matrix λ_i ($i=1, 2, 3, \dots, 9$), the eigenvector is further obtained as shown in table 5.

9 new components are obtained by using PCA. It can be seen from table 3 that the characteristic root $\lambda_4, \lambda_7, \lambda_8, \lambda_9 = 0$ and the corresponding contribution and accumulated contribution are also 0, indicating that some indicators among original indicators are of great correlation and 5 variable indicators can completely represent the information reflected by 9 original indicators. so 5 indicators are selected as the principal components and other 4 are eliminated. Next, by calculating the weight of original 9 indicators in reflecting the influencing factors of fire risk of stadium, the indicator with the largest weight is obtained and the principal component is confirmed.

$$\xi_1 = 0.2978 \times 7.7719 + 0.1622 \times 0.756 + 0.104 \times 0.3796 + 0.5274 \times 0.077 - 0.7079 \times 0.0155 + 0.056 \times 0 + 0.03 \times 3 - 0.2352 \times 0 - 0.0708 \times 0 = 2.4987.$$

By this analogy, the weight of X2, X3, X4, ..., X9 in the comprehensive information can be obtained, that is, $\xi_2 = 2.1468$, $\xi_3 = 2.2602$, $\xi_4 = 2.3766$, $\xi_5 = 2.7844$, $\xi_6 = 2.604$, $\xi_7 = 2.8274$, $\xi_8 = 2.5632$, $\xi_9 = 2.1178$.

TABLE 2 The standardized data

	X1	X2	X3	X4	X5	X6	X7	X8	X9
2008	60.36	17.59	25.64	31.17	12.53	70.06	32.87	22.47	33.91
2009	68.48	21.96	20.38	37.02	18.55	65.76	25.93	31.52	39.02
2010	69.38	30.77	22.19	29.08	19.00	79.24	38.77	42.08	38.32
2011	70.27	33.58	26.78	49.05	21.01	68.65	35.95	47.21	38.31
2012	74.10	29.56	28.85	37.08	13.35	63.32	39.08	58.64	35.96
2013	81.60	37.96	23.64	32.79	12.98	59.87	40.65	51.32	44.05

The weight in a descending order is X7, X6, X5, X8, X1, X4, X3, X2, X9. Therefore, 4 corresponding indicators with the minimum weight to be eliminated are X4, X3, X2, X9.

Through the above calculation and analysis, the main influencing indicators that influences the fire risk prediction of stadium are screened, which are poorly implemented fire security rules and regulations of stadium X7, poor fire acceptance of stadium X6, poor automatic fire warning system X5, insufficient fire extinguisher or failure of fire extinguisher X8 and electric fire X1. These indicators considers the critical influencing indicator factors of fire risk prediction of stadium and are also greatly concerned by the relevant staff of enterprise upon management because they play a decisive influence role in the fire risk prediction of stadium, thus providing important and reliable direction for the security management decision of stadium. If the indicator that influences the risk prediction of stadium is uncertain and complex, the PCA can be used to screen the main indicators and eliminate the relevant indicators in the comprehensive evaluation. This method is of good value in the fire risk prediction of stadium and great theoretical and realistic significance in the security management of stadium.

TABLE 4 Characteristic root, contribution and accumulated contribution

H (I)	GXL (%)	LJGXL (%)
7.7719	86.3541	86.3541
.756	8.3997	94.7538
.3796	4.2177	98.9715
.077	.8559	99.8274
.0115	.1726	100
0	0	100
0	0	100
0	0	100
0	0	100

M=9, LJGXL=.8635, D=9

H(I) is the characteristic root, GXL(%) is the contribution and LJGXL(%) is the accumulated contribution.

TABLE 3 The relevant matrix coefficient

Indicator	X1	X2	X3	X4	X5	X6	X7	X8	X9
X1	1.000	0.509	0.509	0.206	0.207	0.206	0.122	0.483	0.483
X2	0.509	1.000	1.000	1.000	0.604	0.345	0.543	0.429	0.728
X3	0.509	1.000	1.000	1.000	0.604	0.345	0.543	0.429	0.728
X4	0.206	1.000	1.000	1.000	0.604	0.345	0.543	0.429	0.728
X5	0.207	0.604	0.604	0.604	1.000	0.764	0.288	0.476	0.237
X6	0.206	0.345	0.345	0.345	0.764	1.000	-0.2	-0.30	-0.57
X7	0.122	0.543	0.543	0.543	0.288	-0.2	1.000	0.367	0.456
X8	0.483	0.429	0.429	0.429	0.476	-0.30	0.367	1.000	0.456
X9	0.483	0.728	0.728	0.728	0.237	-0.57	0.456	0.456	1.000

TABLE 4 Characteristic root, contribution and accumulated contribution

H (I)	GXL (%)	LJGXL (%)
7.7719	86.3541	86.3541
.756	8.3997	94.7538
.3796	4.2177	98.9715
.077	.8559	99.8274
.0115	.1726	100
0	0	100
0	0	100
0	0	100
0	0	100

M=9, LJGXL=.8635, D=9, H(I) is the characteristic root, GXL(%)is the contribution and LJGXL(%) is the accumulated contribution.

TABLE 5 Eigenvector

	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9
X1	0.2978	0.3022	0.3127	0.3226	0.3428	0.3428	0.3517	0.3301	0.2831
X2	0.1522	-0.332	0.0097	-0.373	0.21	0.0693	-0.032	-0.429	0.728
X3	0.104	0.2044	-0.477	-0.504	0.0313	-0.514	0.3153	-0.43	-0.567
X4	0.5274	-0.338	0.0182	0.2506	-0.514	-0.601	-0.04	0.476	0.7285
X5	-0.707	-0.161	0.1755	-0.013	-0.036	-0.225	0.1343	0.3679	0.2376
X6	0.056	0.036	-0.683	0.4072	0.4206	-0.610	0.0973	-0.30	-0.57
X7	0.033	-0.750	0.3104	-0.610	0.353	-0.062	0.0595	0.1095	-0.134
X8	-0.235	0.4684	-0.377	-0.054	-0.231	-0.054	-0.088	-0.074	0.2991
X9	-0.070	-0.163	0.0547	-0.227	-0.08	-0.57	-0.857	-0.404	0.3621

4 Conclusions

PCA, with simple principles, is simple and convenient in the practical application. Although the computation task is large in case of a large amount of data, with the popularization and application of computer, supported by computer program and relevant software, the results can be obtained accurately and rapidly. PCA is a widely used among various analysis methods, which chooses the most principal component from the multiple indicators (variables) and multiple observation data of sample and quantitatively confirms the relevant influence between sample and

indicators. In studying the fire risk prediction of stadium, it is allowed to analyze the influencing factors by using the mathematics and computer tool as well as PCA and choose the primary and secondary component to make relevant management personnel directly know the correlation between indicators in the fire risk of stadium and find the potential best security management prediction system, for the purpose of taking pertinent prediction measures of fire accident risks to overcome the unfavourable factors in the security management and ensure the safe and normal operation of stadium as well as the personal and property security of sporters. In addition, in the analysis and applica-

tion, the actual situation is combined. PCA in this paper provides some new ideas for other analysis method, which increases the theoretical research basis of PCA and provides some references to the further research, with theoretical and realistic significance.

Acknowledgments

Mingchang Liu is corresponding author. This work was supported by 2013 Hubei Provincial Education Science Twelfth Five Year Plan Project [2013B032].

References

[24]Xiao Zenan 2008 Smoke Control System Performance Design of Beijing Olympic Stadium *Journal of Agricultural Science and Technology* **26**(15) 45-9

[25]Chen Yonggao, Sun Wenjian 2007 Research on Competitiveness Evaluation of Construction Enterprise Based on PCA *Construction Technology* **36**(12) 33-4

[26]Yu Xia 2010 Wang Dongqiang. Comprehensive Evaluation on Fire Performance of stadium *Journal of the Chinese People's Armed Police Force Academy* **26**(6) 34-9

[27]Mohammad Saadatseresht, Ali Mansourian, Mohammad Taleai. 2009 Evacuation planning using multiobjective evolutionary optimization approach *European Journal of Operational Research* **198**(1) 305-14

[28]Lämmel G, et al. 2009 The representation and implementation of time-dependent inundation in largescale microscopic evacuation simulations *Transportation Research Part C* **18**(1) 84-98

[29]Zheng X P, Zhong T K, Liu M T 2009 Modeling crowd evacuation of a building based on seven methodological approaches *Building and Environment* **44**(3) 437-45

[30]Deb K, Pratap A, Agarwal S, Meyarivan T 2002 A fast and elitist multi-objective genetic algorithm: NSGA-II *IEEE Transaction on Evolutionary Computation* **6**(2) 181-97

[31]Salim Labiod, Thierry Marie Guerra 2009 Anytime measures for top- k algorithms on exact and fuzzy data sets *The VLDB Journal*, **18**(2) 407-27

[32]Yi-Fei Chen, Xiao-Lin Qin, Liang Liu, Bo-Han Li 2012 Fuzzy Distance-Based Range Queries over Uncertain Moving Objects *Journal of Computer Science and Technology* **27**(2) 376-96

[33]Kim Y M, Choi J C, Kim J H, Kim C 2002 Development of a System for Progressive Working of an Electric Product by Using Fuzzy Set Theory. *The International Journal of Advanced Manufacturing Technology* **20**(10) 765-79

Authors	
	<p>Huajie Chen, born on January 1, 1977, China</p> <p>Current position, grades: University studies: majored in Sports Education, won the Bachelor Degree of Education from Wuhan institute of sport in China in June 2001, and acquired Master of Science in Education from Shanghai University of sport in 2010. Scientific interests: science of physical culture and sports is his research field, previous research area is Sports Science and school Physical education Publications: In April 2012, he made paper "Research on Jiangxi province coach continued education and post training present situation" - a Periodical which named Popular literature and art. Experience: In July 2001 served as a teacher in the Physical Education Department of Jinggangshan University located at Jian city in China.</p>
	<p>Mingchang Liu, born on December 1, 1977, China</p> <p>Current position, grades: P.E teacher in the Department of Physical Education, Huazhong Agricultural University in China. University studies: Majored in Table Tennis Teaching and Training at the College of Physical Education, studied in Wuhan Sports University from 1997 to 2001, and graduated in 2001, master's degree in education at the School of Physical Education & Sports Science, Nanjing Normal University in 2007. Master's Thesis entitled "Study on the Curriculum Arrangement and Teaching Mode of University Physical Education in Wuhan" was awarded national outstanding Master's thesis. Scientific interests: School sports and community sports, and how school sports better assisted and served community sports.</p>