Applying Principal Component Analysis and Unascertained Method for the Analysis of Construction Accident Risk

Huawang Shi
School of Civil Engineering, Hebei University of Engineering, Handan, P.R.China
Email:stone21st@163.com

Wanqing Li and Yong Deng
Hebei University of Engineering, Handan, P.R.China, Shanghai Jiao Tong University, Shanghai, China
Email: stone21st@sohu.com

***Abstract*—In order to evaluate the construction accident risk, a novel method integrating principal component analysis (PCA) and the unascertained number algorithm with the synthesis weights is presented. According to human-machine engineering, the multi-index and hierarchical system of accident risk in construction projects was built. PCA is applied to find relevant factors of the accident risk, which are classified into 6 categories used as inputs of unascertained assessment model. To improve the accuracy, quantitative change and qualitative change all need be reflected in the assessment results of accident risk in construction projects, the method determining the synthesis weights of indexes including both subjective favoritism and objective information is presented so as to make full use of redundant information, therefore, qualitative knowledge and quantitative data could be integrated. Finally, an improved unascertained assessment approach model was constructed to assess accident risk in construction projects by combining qualitative and quantitative information over conventional or other subjective methods and was verified by a case study.**

***Index Terms*—principal component analysis (PCA), unascertained assessment, synthesis weights, construction accident risk**

I. Introduction

Construction is one of the most dangerous and risky businesses and building construction projects appear to have higher accident rates[1]. The major construction accidents occur frequently and has brought the adverse economic, political and social influence in construction industry; Making efforts to reverse this situation and giving construction workers the safety and health insurance are becoming a top priority issue of the building industry and making an urgent and challenging task of construction enterprises for the safety management. Thus, subjective assessments of numerous hazard and safety factors in construction projects, and rating them by using an objective scale like the Likert-scale. However, the ratings given by experts may not be precise as they are quantitative expressions of qualitative assessments. Thus, there is a need for accommodating this imprecision in premium-rating. Si, Ruxton, and Wang(2001) reported that fuzzy logic gives a more flexible structure for combining qualitative and quantitative information over conventional or other subjective methods[1]. However, the methods employed such as fuzzy sets have many shortcomings. How to solve the problem of uncertainty and lack of data and find out a new method for determining the weights of indexes are urgent tasks in construction engineering.

In this paper, subjective evaluation is quantified by using non-classical mathematics method, is based on analyzing evaluation problem of construction projects accident risk including the characteristics of uncertain information. And a new method of evaluation of construction projects accident risk based on unascertained number algorithm is proposed. This method can use uncertain information more sufficiently than traditional evaluation methods[2,3]. Risk in construction projects was built and the method determining the synthesis weights of indexes including both subjective favoritism and objective information qualitative knowledge and quantitative data to be integrated was given. And merits of this method are more than traditional method in many aspects, such as correction, objectivity, feasibility and adjustment. And hence is more suitable to preliminary evaluation of construction projects accident risk.

The structure of the paper is as follows. In the introductory part, attention is paid to the basic concepts of construction accident risk and its assessment methods. In the following part, brief introduction to the unascertained mathematics including definition of unascertained number and algorithm of unascertained number, and principal component analysis are introduced in. Ten, in section 3, a safety assessment model for building construction projects accident risk based on the unascertained measure is set up. In section 4, its application in practice is introduced. Finally, some conclusions and the advantages of the method proposed in this paper are pointed out.

II. Brief Introduction to the Unascertained Mathematics and PCA

doi:10.4304/jcp.5.8.1273-1280

Unascertained mathematics, proposed by Wang (1990)[4], is a tool to describe subjective uncertainty quantitatively. It deals mainly with unascertained information which differs from stochastic information, fuzzy information, and grey information. Unascertained information refers to then formation demanded by decision-making over which the message itself has no uncertainty but, because of situation constraints, the decision maker cannot grasp the whole information needed. The decision maker produces the uncertainty. Since the 1990s, Liu[5] and other scholars have done much work and Unascertained Mathematics has been successfully used in many fields.

A. Definition of Unascertained Number

Definition 1: Suppose a is arbitrary real number, $0 < \alpha \le 1$, then definite $[[a,a],\varphi(x)]$ is first-order unascertained number, where

$$\varphi(x)=\begin{cases}\alpha, x=a\\ 0, x\neq a \cup x\in R\end{cases} \tag{1}$$

Note that $[a,a]$ express the interval of value, and $\varphi(x)=\alpha$ express belief degree of a. When α =1, belief degree of a is 1.Where α =0, belief degree of a is zero.

Definition 1: Suppose $[a,b]$ is arbitrary closed interval, $a=x_1<x_2<\ldots<x_n=b$,if

$$\varphi(x)=\begin{cases}\alpha_i, x=x_i\,(i=1,2,\cdots,n)\\ 0, other\end{cases} \tag{2}$$

and $\sum_{i=1}^{n}\alpha_i=\alpha$,0< $\alpha \le 1$,then $[a,b]$ and $\varphi(x)$ compose a n -order unascertained number, as follow $[[a,b],\varphi(x)]$,where α is total degree belief, $[a,b]$ is the interval of value, is $\varphi(x)$ the density function.

Definition 1: Suppose unascertained number is $A=[[x_1,x_2],\varphi(x)]$, where

$$\varphi(x)=\begin{cases}\alpha_i, x=x_i\,(i=1,2,\cdots,k)\\ 0, other\end{cases} \tag{3}$$

$0<\alpha_i<1$, i =1,2,…, k , $\alpha=\sum_{i=1}^{k}\alpha_i\le i$.

Then first-order unascertained number

$$E(A)=\left[\left[\frac{1}{\alpha}\sum_{i=1}^{k}x_i\alpha_i,\frac{1}{\alpha}\sum_{i=1}^{k}x_i\alpha_i\right],\varphi(x)\right],$$

$$\varphi(x)=\begin{cases}\alpha, x=\frac{1}{\alpha}\sum_{i=1}^{k}x_i\alpha_i\\ 0, other\end{cases} \tag{4}$$

is expected value of unascertained number A . When α =1, as $E(A)$, unascertained number A is discrete type random variable. When $\alpha<1$, $E(A)$ is first-order unascertained number. Where, $\frac{1}{\alpha}\sum_{i=1}^{k}x_i\alpha_i$ as expected value of A that belief degree is α .

B. Algorithm of Unascertained Number

Each unascertained number includes two parts of probable value and belief degree. So, unascertained number algorithm also includes two parts. Suppose unascertained numbers are A and B .Where

$$A=f(x)=\begin{cases}\alpha_i, x=x_i\,(i=1,2,\cdots,m)\\ 0, other\end{cases},$$

$$B=g(x)=\begin{cases}\beta_i, y=y_i\,(i=1,2,\cdots,n)\\ 0, other\end{cases} \tag{5}$$

$C=A\times B$ also is unascertained number. Probable value and belief degree of C is calculated as follows.

（1）Constituted multiply matrix of probable value of unascertained number A and B , where individual is probable value number series x_1 , x_2 ,…, x_k and y_1 , y_2 ,…, y_m as A and B , permute from little to big.

（2）Constituted multiply matrix of belief degree of unascertained number A and B , where individual is belief degree number series α_1 , α_2 ,…, α_m and β_1 , β_2 ,…, β_n are A , B .Suppose a_{ij} and b_{ij} individual is element of multiply matrix of probable value of A and B , here i is line of matrix, j is array of matrix. We called a_{ij} and b_{ij} as relevant position element.

（3）$\bar{x}_1$, $\bar{x}_2$,…, $\bar{x}_k$ result from multiply matrix of probable value of unascertained number A and B , which permute from little to big. And an equal element is one element of relevant position element in multiply matrix of belief degree. Suppose $\bar{r}_1$, $\bar{r}_2$,…, $\bar{r}_k$ is relevant position element permutation. Where

$$C=\varphi(x)=\begin{cases}\bar{r}_i, x=\bar{x}_i\,(i=1,2,\cdots k)\\ 0, other\end{cases} \tag{6}$$

Suppose $C=\varphi(x)$ is arithmetic product of unascertained number A and B . Where

$$C=A\times B=f(x)\times g(x)=\begin{cases}\bar{r}_i, x=\bar{x}_i\,(i=1,2,\cdots k)\\ 0, other\end{cases} \tag{7}$$

C. Introduction to PCA

PCA was invented in 1901 by Karl Pearson[2]. Now it is mostly used as a tool in exploratory data analysis and for making predictive models. PCA involves the calculation of the eigenvalue decomposition of a data covariance matrix or singular value decomposition of a data matrix, usually after mean centering the data for each attribute. The results of a PCA are usually discussed in terms of component scores and loadings (Shaw, 2003). PCA[3-5] can be used for dimensionality reduction in a data set by retaining those characteristics of the data set that contribute most to its variance, by keeping lower-order principal components and ignoring higher-order ones. Such low-order components often contain the "most important" aspects of the data. However, depending on the application this may not always be the case.

Problems arise when performing recognition in a high-dimensional space (e.g., curse of dimensionality). Significant improvements can be achieved by first mapping the data into a lower-dimensionality space. The goal of PCA is to reduce the dimensionality of the data while retaining as much as possible of the variation present in the original data set.

Supposing n samples, each sample has m target factors, x_j (j = 1,2, ..., m), derived from observation values x_{ij} (i=1,2,...,n), constitute the raw data matrix $X=(x_{ij})n\times m$, shown as bellow:

$$X=\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \tag{8}$$

The target factor is often relevant, thus increasing the internal complexity of the samples. Principal component analysis is to have a correlation between a number of factors into a set of mutually independent factor of a few General methods. These will be the original general factor target factor in the overlapping information removed, to the original contains only significant difference between the target and reflect the original main target factor information purposes. That is, without changing the original data provided by the basic information on more focused and typically show the characteristics of the study. Principal component - the specific algorithm for cluster analysis are as follows.

(1) Original data will be standardized (Z-Score Standardization)

Class and quantity in order to eliminate the impact of different dimension, first of all original data on the standardization of treatment (standardized value of the post-treatment x_{ij}*

$$x_{ij}^{*}=\frac{x_{ij}-\overline{x_j}}{S_j} \tag{9}$$

Where: $\overline{x_j}$ and S_j , respectively, are the mean and standard deviation of the jth target sample, and

$$\overline{x_j}=\frac{1}{n}\sum_{i=1}^{n}x_{ij} \tag{10}$$

$$S_j=\left[\frac{1}{n-1}\sum_{i=1}^{n}\left(x_{ij}-\overline{x_j}\right)^2\right]^{1/2} \tag{11}$$

(2) Calculation of correlation between the matrix

Based on the standardized data matrix $X^{*}=(x_{ij}^{*})$, calculated the correlation coefficient matrix $R = (r_{ij})m\times m$. Where, r_{ij} are the correlation coefficient between the x_i and x_j target factor .

$$r_{ij}=\frac{1}{n-1}\sum_{k=1}^{n}x_{ki}^{*}x_{kj}^{*}=\frac{\sum_{k=1}^{n}\left(x_{ki}-\overline{x_i}\right)\left(x_{kj}-\overline{x_j}\right)}{\sqrt{\sum_{k=1}^{n}\left(x_{ki}-\overline{x_i}\right)^2\left(x_{kj}-\overline{x_j}\right)^2}} \tag{12}$$

Where, i, j=1,2,…,m.

(3) Solving eigenvalue of the correlation matrix and eigenvectors

Calculating the characteristic equation $|R-\lambda I|=0$, obtained all of the eigenvalue $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$, and the corresponding Tikhonov unit eigenvector t_j=（t_{1j}，t_{2j}， …， t_{mj}）

$$Y_j=\sum_{k=1}^{m}t_{kj}\bullet x_k^{*} \tag{13}$$

Where: x_k* is the standardized sample matrix.

(3)To determine the number of principal components

Selecting r principal components in the m principal components that have been identified to finally realize the evaluation analysis. In general, the contribution rate of variance $e_j=\lambda_j/\sum_{k=1}^{m}\lambda_k$ could explain that principal component $\mathrm{Y_j}$ reflects the amount of information size. R is determined by the principle that accumulated contribution value $G(r)=\sum_{k=1}^{r}e_k$ is large enough (typically more than 85%). K is kth measured values of the ith and jth factor , k=1,2,…,r.

III. Building Construction Projects Accident Risk Assessment

A. Establishing Evaluation Index System

Researcher based on the principle cause of the accident[6] (Heinrich) and the basis of consideration of other factors, put forward the accident cause and effect in human - machine - environment system, as shown in Fig.1.

From Fig. 1, The theory shows that the human-computer coordination of operations in the building construction, machinery and the human of a certain management and environmental conditions completing a certain task, not only play their own role, but be linked with each other and helping each other. The system's safety and reliability depends not only on human behavior, but also on the state of matter.

B. Accident Risk Analysis Based on PCA

(1) First, the original data in Table 1 was processed for the standardization, and by 12, the correlation coefficient matrix was calculated.

(2) By the correlation coefficient matrix eigenvalue calculation, as well as all the main components of the

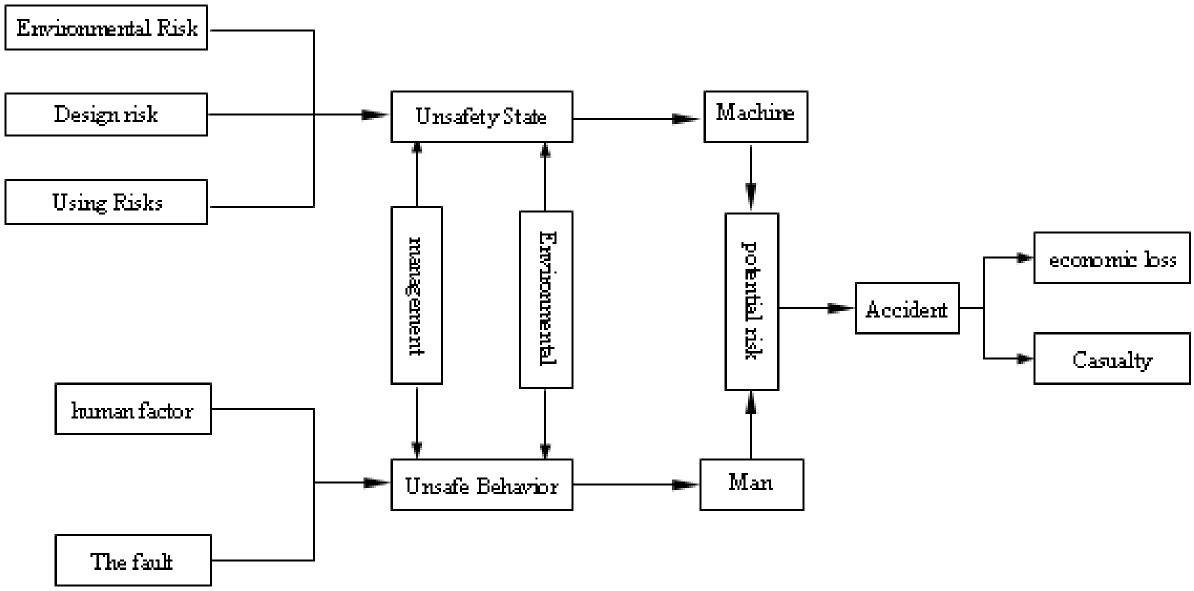

Figure 1. Consequence of accident commentate by human-machine engineering.

contribution rate and the cumulative contribution rate. The principal component: γ_1, γ_2, γ_3, γ_4, γ_5, γ_6 of the cumulative contribution rate were up to 79.8%, so just find the 6 principal component.

(3) According to this theory, accident risk factors in construction project can be summarized in Fig. 2.as shown follow. In Fig.2., construction project accident risk is described by 6 principal component that including man in the topside operations I_1, the level of security protection I_2 and security system situation I_3. The impact of bad weather I_4, engineering geological conditions I_5, the construction site conditions I_6, machinery and equipment performance I_7, skilled workers I_8, workers without a safety training I_9. The ratio unlicensed I_{10}, the engineering design level I_{11}, the organization and

TABLE I.
CONSTRUCTION ACCIDENT RISK DATA

No.	No.1	No.2	No.3	No.4	No.5	No.6	No.7	No.8	No.9	No.10
I_1	0	0	1	1	0.68	0.6	0.66	0.85	0.98	0.24
I_2	0.48	0.51	0.87	0.58	0.79	0.7	0.71	0.72	0.31	0.56
I_3	1	1	1	0.5	1	0.5	1	1	0.5	1
I_4	0.89	0.97	0.23	0.65	0.86	0.5	0.98	0.29	0.83	0.39
I_5	0.27	0.27	0.45	0.98	0.62	0.5	0.72	0.31	0.56	0.18
I_6	0.62	0.558	0.87	0.89	0.84	0.6	0.90	0.30	0.35	0.13
I_7	0.39	0.42	0.56	0.78	0.78	0.4	0.69	0.57	0.26	0.14
I_8	0.48	0.51	0.87	0.58	0.79	0.7	0.79	0.43	0.66	0.86
I_9	0.42	0.43	0.36	0.66	0.61	0.9	0.084	0.71	0.54	0.55
I_{10}	0.43	0.67	0.80	0.58	0.85	0.6	0.86	0.83	1	0.63
I_{11}	0.73	0.53	0.53	0.35	0.75	0.6	0.71	0.28	3	0.18
I_{12}	0.13	0.13	0.36	0.89	0.65	0.5	0.79	0.36	1	0.51
I_{13}	0.93	0.46	0.79	0.96	0.76	0.6	0.80	0.45	1	0.14
I_{14}	0.72	0.80	0.69	0.66	0.88	0.7	0.83	0.25	0.56	0.37
I_{15}	0.35	0.6	0.38	0.33	0.73	0.6	0.81	0.67	0.67	0.16

TABLE III.
CORRELATION COEFFICIENT MATRIX

Indicts	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	I_9	I_{10}	I_{11}	I_{12}	I_{13}	I_{14}	I_{15}
I_1	1.000														
I_2	0.521	1.000													
I_3	0.311	0.878	1.000												
I_4	0.744	0.766	0.662	1.000											
I_5	0.490	-0.311	0.328	0.450	1.000										
I_6	-0.311	0.147	0.327	0.267	0.315	1.000									
I_7	0.635	0.767	0.089	0.452	0.643	0.068	1.000								
I_8	-0.517	-0.475	-0.683	0.513	0.111	-0.307	0.490	1.000							
I_9	0.600	0.324	0.776	0.705	-0.082	0.134	0.534	0.946	1.000						
I_{10}	-0.327	-0.323	-0.298	-0.193	0.110	-0.053	0.328	0.398	0.324	1.000					
I_{11}	-0.082	-0.490	-0.408	-0.232	0.686	0.039	0.450	0.744	0.577	0.727	1.000				
I_{12}	-0.433	-0.642	-0.748	0.364	-0.433	-0.316	0.521	0.766	0.878	0.635	0.267	1.000			
I_{13}	-0.228	-0.088	-0.390	-0.263	0.315	0.075	1.000	0.574	0.490	0.534	0.328	0.450	1.000		
I_{14}	-0.311	-0.349	-0.524	-0.452	0.643	0.068	0.574	1.000	0.775	0.946	0.398	0.744	0.766	1.000	
I_{15}	-0.517	-0.475	-0.683	-0.513	0.111	-0.307	0.490	0.775	1.000	0.600	0.324	0.577	0.878	0.153	1.000

TABLE II.
MARKING TABLE TO PROJECT 1 BY AN EXPERT

Indicator	**Higher**	**High**	**Average**	**Low**	**Indicator**	**Higher**	**High**	**Average**	**Low**	**Indicator**	**Higher**	**High**	**Average**	**Low**
I_1	1	1.5	0	0	I_6	0	2	0.5	0	I_{11}	0	2.5	2	0
I_2	0	1.5	1	0	I_7	0	1.5	1	0	I_{12}	0	3.5	2	0
I_3	0	2.5	2	0	I_8	1.5	3	0	0	I_{13}	0	2.5	0	0
I_4	1.5	1	0	0	I_9	0	1	1.5	0	I_{14}	0	1.5	1	0
I_5	1	1.5	0	0	I_{10}	0	0	1.5	1	I_{15}	3	2.5	0	0

command capabilities I_{12} Mechanical equipment installation I_{13}. The experience of similar projects I_{14} and skilled operation workers I_{15}. We classify the factors that would influence the construction accident risk into six categories based on the literature review. The first category includes the topside operations I_1, which reflect the rate of topside operations. The second one covers

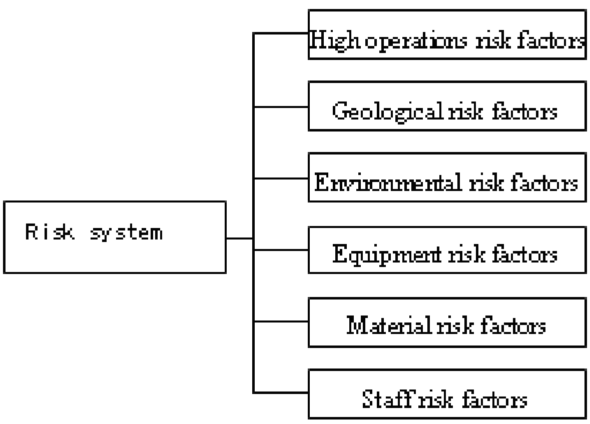

Figure. 2 Accident risk system of construction.

engineering geological conditions I_5, the construction site conditions I_6. The third one is related to environmental conditions of construction projects, which are including The impact of bad weather I_4, the level of security protection I_2, the engineering design level I_{11}, The experience of similar projects I_{14} and security system situation I_3. The forth one is reflecting the characteristics of equipment such as machinery and equipment performance I_7 and so on. The fifth category contains material risk factors. The sixth category contains factors describing the environment of construction projects, which may influence the construction accident risk. The final category covers the staff risk factors including skilled workers I_8, workers without a safety training I_9, the ratio unlicensed I_{10}, skilled operation workers I_{15}.

C. Synthesis Weight of Evaluation Index

1) To determine the subjective weight of indicators

The subjective weight of indicators can be determined though one or more subjective weight determining methods. Supposing the weight of subjective indicators is as follows:

$$a = \left(a_1, \quad a_2, \quad \cdots, \quad a_m\right)^T \tag{14}$$

Where, $\sum_{j=1}^{m} a_j = 1,\ a_j > 0\ \left(j = 1,2,\cdots,m\right)$

2) To determine the objective weight of indicators

The objective weight of indicators can also be determined though one or more subjective weight determining methods. Supposing the weight of objective indicators are as follows:

$$\beta = \left(\beta_1, \quad \beta_2, \quad \cdots, \quad \beta_m\right)^T \tag{15}$$

Where, $\sum_{j=1}^{m} \beta_j = 1,\quad \beta_j > 0\ \left(j = 1,2,\cdots,m\right)$

In this paper, information entropy [7-9]was introduced for determining the objective weight of indicators . Following the method of determining the index's identification weight by using the information entropy will be introduced. For the discrete stochastic variables, their information entropy is

$$H(\chi) = -\sum_{i=1}^{m} p(x_i)\log p(x_i),$$

$$0 \le p(x_i) \le 1, \sum_{i=1}^{n} p(x_i) = 1 \tag{16}$$

In this paper:

$$H(j) = -\sum_{k=1}^{K} \mu_{ijk} \cdot \log\mu_{ijk} \tag{17}$$

$$\gamma_j = 1 - \frac{1}{\log K} H(j)$$

$$= 1 + \frac{1}{\log K}\sum_{k=1}^{K} \mu_{ijk} \cdot \log\mu_{ijk} \tag{18}$$

Where $w_j = \gamma_j \Big/ \sum_{j=1}^{d} \gamma_j$, $w = (w_1, w_2, \cdots, w_d)$

Obviously, $0 \le w_j \le 1$ and $\sum_{j=1}^{d} w_j = 1$.

3) To determine the synthesis weight

Assuming the synthesis weight of each indicator is as follows:

$$W = \left(\omega_1, \quad \omega_2, \quad \cdots, \quad \omega_m\right)^T$$

In order to make full use of the subjective and objective weight determining method to reach the objective and subjective unity, the synthesis weight is carried out by the following formulas:

$$W = (\mu a_1 + (1-\mu)\beta_1, \mu a_2 + (1-\mu)\beta_2, \cdots, \mu a_m + (1-\mu)\beta_m) \tag{19}$$

Where μ ($0 < \mu < 1$) is the preference coefficient that reflects the preference level of decision maker for subjective weight and objective weight determining method.

D. Synthesis Appraisal System

As it is known that $\mu_{ijk} = \mu(x_{ij} \in c_k) \quad 1 \le i \le n, 1 \le j \le m$ is the unascertained measure and μ_{ijl} is unit factor's measure appraisal matrix of x_i, in which, μ_j^i means x_{ij} makes x_i has c_k grade in j row.

$$(\mu_{ijk})_{m\times K} = \begin{pmatrix} \mu_{i11}, \mu_{i12}, \cdots, \mu_{i1K} \\ \mu_{i21}, \mu_{i22}, \cdots, \mu_{i2K} \\ \vdots \quad \vdots \quad \cdots \quad \vdots \\ \mu_{im1}, \mu_{im2}, \cdots, \mu_{imK} \end{pmatrix} \tag{20}$$

where,

$$(i = 1,2,\cdots,n \quad j = 1,2,\cdots,m \quad k = 1,2,\cdots,K)$$

If the single factor measure appraisal matrix above is known, the each factor's classification vector about x_i is(16), and then(17) got as follow:

$$W^i = (w_1^i, w_2^i, \cdots, w_m^i) \tag{21}$$

$$\mu^i = W^i \cdot (\mu_{ijk})_{m\times K} = (w_1^i, w_2^i, \cdots, w_m^i) \bullet \begin{pmatrix} \mu_{i11}, \mu_{i12}, \cdots, \mu_{i1K} \\ \mu_{i21}, \mu_{i22}, \cdots, \mu_{i2K} \\ \vdots \quad \vdots \quad \cdots \quad \vdots \\ \mu_{im1}, \mu_{im2}, \cdots, \mu_{imK} \end{pmatrix}$$

$$= \left(\sum_{j=1}^{m} w_j^i \cdot \mu_{ij1}, \sum_{j=1}^{m} w_j^i \cdot \mu_{ij2}, \cdots, \sum_{j=1}^{m} w_j^i \cdot \mu_{ijK}\right) \tag{22}$$

So μ^i is x_i's appraisal vector.

E. Principle of Identification

Because the classification of the comment ranks is orderly e.g, c_k is "better" than c_{k+1}, the identification principle of "maximum measure'' is not available. The credible identification principle is needed. Let the credible identification be $\lambda, (\lambda > 0.5)$, and it is always adopt 0.6 or 0.7.

$$k_0 = \min_k \left[(\sum_{l=1}^{k} \mu_{il}) \ge \lambda, k = 1,2,\cdots K \right] \tag{23}$$

Then x_i belongs to the rank c_{k_0} .It means that when x_i is not lower than c_k , the fiducially degree is λ , or in other words lower than c_k is 1-λ .

Ⅳ. CASE STUDY

According to Table 3, description results of evaluation by one expert are shown in Tab.1. And description results of evaluation by other experts are ellipsis.

As shown in Table 3, the index set is{higher, high, average, low}and it is divided into four appraisal scales, and by all appearances, it is positive sequence. The each factor is total ten score. Then, each appraisal object is ten score and the distinguishment is that the degrees are different. The scoring principle is fairly and fit the measurement criterion. Based on the statistical data of the appraisal object, the single factor measurement matrix μ^1 is as follows:

$$\mu_1 = \begin{pmatrix} 0.1600 & 0.7600 & 0.0800 & 0 \\ 0.0800 & 0.5600 & 0.3600 & 0 \\ 0 & 0.5333 & 0.4667 & 0 \\ 0.2800 & 0.4800 & 0.2400 & 0 \\ 0.0800 & 0.6800 & 0.2400 & 0 \\ 0 & 0.7200 & 0.2800 & 0 \\ 0 & 0.6800 & 0.3200 & 0 \\ 0.3556 & 0.6444 & 0 & 0 \\ 0 & 0.2800 & 0.7200 & 0 \\ 0 & 0 & 0.8400 & 0.1600 \\ 0 & 0.6444 & 0.3556 & 0 \\ 0 & 0.6364 & 0.3636 & 0 \\ 0.1600 & 0.8000 & 0.0400 & 0 \\ 0.0800 & 0.7600 & 0.1600 & 0 \\ 0.6364 & 0.3273 & 0 & 0 \end{pmatrix}$$

According to the equations (14)-(20), the factor's weight vector is

$$\omega_j^i = (0.0761,0.0620,0.0928,0.1482,0.1116,$$
$$0.1005,0.0514,0.0457,0.0540,0.0515,0.0645$$
$$0.0300,0.0310,0.0400,0.0407)$$

Then, the object's appraisal vector according to (22) is

$$\mu^1 = W^1 \cdot (\mu_{1jk})$$
$$= (\ 0.2903,0.2099\ ,0.3036,\ 0.1950)$$

We adopt λ =0.6, according to the equation (23), when k_0=3, 0.2903+0.2099+0.3036= 0.788> 0.6.

So, the appraisal object is "Average". That is to say, this construction accident risk is "Average".

Ⅴ. CONCLUTION

In this paper, based on the principle cause of the accident and the basis of human - machine - environment system, the accident risk cause factors was put forward. Using PCA, we classify the factors that would influence the construction accident risk into 6 categories based on the literature review. The first category includes the topside operations I_1 , which reflect the rate of topside operations. The second one covers Engineering geological conditions I_5 , the construction site conditions I_6. The third one is related to environmental conditions of construction projects, which are including the impact of bad weather I_4 , the level of security protection I_2 , the engineering design level I_{11} , The experience of similar projects I_{14} and Security system situation I_3. The forth one is reflecting the characteristics of equipment such as machinery and equipment performance I_7 and so on. The fifth category contains material risk factors. The sixth category contains factors describing the environment of construction projects, which may influence the construction accident risk. The final category covers the staff risk factors including Skilled workers I_8 , workers without a safety training I_9 , the ratio unlicensed I_{10} , skilled operation workers I_{15} .

Then, based on historical data and Unascertained Mathematics, a new construction accident risk model was established and evaluation and weight of index are quantified by determining the synthesis weights of indexes that include both subjective favoritism and objective information qualitative knowledge and quantitative data need to be integrated. The result of evaluation is more sophisticated than traditional evaluation. It establishes an effective risk assessment strategy via a well-structured incentive system for contractors and clients. Its implementation in construction company can help contractors giving construction workers the safety and health insurance, and its implementation in the insurance industry can curtail accidents in the construction industry, thereby minimizing insurers' financial risks.

ACKNOWLEDGMENT

This paper is partially supported by Program for New Century Excellent Talents in University (NCET-08-0345), National Natural Science Foundation of China (60874105), Shanghai Rising-Star Program (09QA1402900), Aviation Science Foundation 20070511001), and the Chenxing Scholarship Youth Found of Shanghai Jiao Tong University (T241460612).

REFERENCES

[1] Kamardeen Imriyas. An expert system for strategic control of accidents and insurers risks in building construction projects. Expert Systems with Applications36(2009)4021-4034.

[2] Patra sk et al., J- Photochemistry & Photobiology A:Chemistry, (1999) 122:23–31.

[3] A. Gorban, B. Kegl, D. Wunsch, A. Zinovyev (Eds.), Principal Manifolds for Data Visualisation and Dimension Reduction, LNCSE 58, Springer, Berlin – Heidelberg – New York, 2007

[4] Greenacre, Michael (1983). Theory and Applications of Correspondence Analysis. London: Academic Press.
[5] Jolliffe I.T. Principal Component Analysis, Series: Springer Series in Statistics, 2nd ed., Springer, NY, 2002, XXIX, 487 p. 28 illus.
[6] Fen Z R. Safety systems engineering [M]. Beijing: Meanurgy Industry Press, 1987(in Chinese).
[7] Wang Guangyuan. Unascertained information and unascertained process[J]. Journal of Harbin university of engineering, 1990(4):1-9(in Chinese).
[8] Liu K D, Wu H Q, Pang Y J. Process and Application of Uncertain Information [M]. Beijing: Science Press, 1999(in Chinese).
[9] Wang, G Y , l990. Mathematics treatment of unascertained information. Journal of Civil Engineering Institute of Haerbin 23(4):l-9.(in Chinese)
[10] Liu,K.D. , Q.K.Cao and Y J.Pang,2004. A method of fault diagnosis based on unascertained set. ACTA Automatic Sinica 30(5):747—756.
[11] Chapman C B. A Risk Engineering Approach to Project Risk Management Project . 1990,18(1)
[12] Jaynes E T. Information Theory and Statistical Mechanics. Phy. Rew, 1957, 106, 620-630
[13] Shore J E. Minimum cross-entropy spectral analysis. IEEE Trans ASSP, 1988, 29(2): 230-237
[14] Wu M J, Suen C Y. Fractal information theory and application. J of East China Univ, 1996,1:19-30

Huawang Shi Born in Handan, China, on June, 11, 1978. He received the master degree in engineering from Hebei University of Engineering in 2005. He joined the Center fo Research in Management in 2005, where he works as a researcher at the civil engineering department. He has authored over 10 journal papers, 30 conference papers and book chapters. His research interests include hybrid artificial intelligent, evolutionary computation and data mining. Contact:stone21st@163.com.