Risk assessment method based on factor analysis and cloud model and its application

Jiakun Li, Yanbin Li, Xiaoming Han
School of Graduate, Air Force Engineering University, Xi’an, Shaanxi, 710051, P.R. China
*Corresponding author’s e-mail: lijiakun2018@163.com

Abstract. Aiming at the problems of excessive risk factors and uncertainties in risk assessment, a risk assessment method based on factor analysis and cloud model is proposed. Firstly, the dimension reduction effect of factor analysis (FA) method and its implementation process are introduced to simplify the risk assessment indicators. Then, cloud model is adopted to replace membership function. A risk assessment method based on cloud model is proposed, which solves the problem of fuzziness and randomness of indicators. Finally, the effectiveness of the proposed method is verified by an example.

1. Introduction
Risk assessment involves many risk factors, each of which has different effects on the overall risk assessment. If the secondary factors in the risk factors can be excluded, the complexity of the risk assessment will be significantly reduced, and the stability of the assessment results will also be increased. Factor analysis can analyze the internal structure of indicators through raw data, thus achieving dimensionality reduction and simplification. Cloud model can achieve effective conversion between qualitative concepts and quantitative data, giving consideration to randomness and ambiguity.

Therefore, this paper proposes a risk assessment method based on factor analysis and cloud model. The method first simplifies the risk assessment index system through factor analysis, and then uses cloud model to replace the membership function. Finally, the assessment results are obtained by cloud generator and cloud computing. An example of a medical risk assessment verifies the effectiveness of the method.

2. Dimensionality reduction process based on factor analysis

2.1. Brief introduction to factor analysis
The role of factor analysis is to streamline and recombine multiple measured variables, and obtain a certain number of comprehensive indicators with practical significance, reflecting the idea of dimensionality reduction [1].

2.2. Implementation process of factor analysis
The main steps of factor analysis are factor analysis applicability confirmation, factor variable construction, factor variable rotation and factor variable score calculation, all of which can be implemented by SPSS software. The specific calculation process is as follows:
- Normalize the raw data, and eliminate the difference in magnitude and dimension of the variables. The SPSS software adopted in this paper can automatically standardize the data. Therefore, jobs remain to be done is to ensure the rationality of the original data.
- Test the correlation of the original variables, and verify the applicability of factor analysis. The results of the correlation coefficient matrix, KMO test and Bartlett's spherical test are needed in this part. The test standards are shown in Table 1.

Table 1. Conditions suitable for factor analysis

| Parameter               | Correlation | KMO test | Bartlett’s spherical test |
|-------------------------|-------------|----------|----------------------------|
| Standard                | ≥ 0.3       | ≥ 0.5    | ≤ 0.05                     |

- Determine the eigenvalues, eigenvectors and factor numbers of the correlation coefficient matrix. Factor rotation is performed when necessary to obtain more obvious practical significance, and the number of factors is adjusted accordingly according to its actual meaning. There are three ways to determine the main factor. The first method can be described as: k principal factors are extracted according to the principle that the eigenvalue is greater than 1. Then, the variance contribution rate of each factor and the load value of each primitive variable on each principal factor are obtained according to the eigenvalue and the eigenvector. The second is to determine the main factor according to the cumulative variance contribution rate of the factor, generally taking the common factor corresponding to the eigenvalue with the cumulative contribution rate greater than 85%. The third is to determine the factor under the premise of determining the number of factors, which is mainly applicable to the confirmatory factor analysis of the structure of the known original variable data.
- Obtain the principal factor with practical significance, and then construct the simplified risk assessment index system as the basis for further analysis.

3. Risk Assessment Method Based on Cloud Model

3.1. Calculates weights by AHP

The weight of each risk factor is calculated by AHP [2]. After passing the consistency test, it is adopted. Assume that the weight vector of risk factors is:

\[ W = (w_1, w_2, \ldots, w_j), \quad j = 1, 2, \ldots, n \]  (1)

3.2. Theory basis of cloud model

The cloud model is a transformation model between the qualitative concept and the quantitative value, which fully considers the fuzziness and the randomness in the risk assessment process of missile equipment training.

Suppose \( U \) is a quantitative theory of domain expressed in precise numerical terms, \( A \) is a qualitative concept in the domain \( U \). Assume that: quantitative value \( x \in U \), \( x \) is a random realization of \( A \), and \( \mu_A(x) \) represents the membership degree of \( x \) to \( A \), which is a random number with stable tendency, that is, it satisfies:

\[ \mu(A): U \rightarrow [0, 1], \forall x \in U, x \in \mu(x) \]

Then the distribution of \( x \) in the domain is called cloud, and each \((x, \mu_A(x))\) is called a cloud droplet[3,4].

The cloud model has three digital features: Expected value (Ex), Entropy (En) and Hyper Entropy (He). Ex represents the central value of the qualitative concept in the quantitative domain; En represents the measurement of the uncertainty of the qualitative concept, which is jointly determined by fuzziness and randomness; Hyper Entropy refers to the uncertainty of entropy, namely the entropy of entropy, reflecting the dispersion degree of cloud droplets. The three digital features of the cloud
model are shown in Figure 1, where $\text{Ex}=5$, $\text{En}=1$, $\text{He}=0.1$, and $n=3000$.

![Figure 1. Digital features of cloud model](image)

Cloud generator is a tool to realize qualitative and quantitative transformation of cloud model, which can be divided into forward cloud generator and reverse cloud generator. The calculation steps of cloud generator are shown in Table 2.

| Step | Forward cloud generator | Inverse cloud generator |
|------|-------------------------|-------------------------|
| 1    | $\text{En} \sim N\left(\text{En}, \text{He}^2\right)$ | $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} x_i$, $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(x - \bar{X}\right)^2$ |
| 2    | $x_j \sim N\left(\text{Ex}, \text{En}^2\right)$ | $\text{Ex} = \bar{X}$ |
| 3    | $\mu_i (x_j) = \exp \left( -(x_j - \text{Ex})^2 / \left(2(\text{En})^2\right) \right)$ | $\text{En} = \sqrt{\frac{\pi}{2}} \frac{1}{n} \sum_{i=1}^{n} |x_i - \bar{X}|$ |
| 4    | $(x_i, \mu_i (x_j), i = 1, 2, \ldots, n)$ | $\text{He} = \sqrt{S^2 - \text{En}^2}$ |

3.3. Cloud model computing

3.3.1. Benchmark cloud. In this paper, the training security risk of missile equipment is classified into relatively low risk (I), medium risk (II), relatively high risk (III), and high risk (IV). The scoring interval of each grade is shown in Table 3.

| Risk level | Score interval | I | II | III | IV |
|------------|----------------|---|----|-----|----|
|            |                | 0~4 | 4~7 | 7~9 | 9~10 |

Calculate the digital features of the benchmark cloud model according to the above security risk levels:

$$\text{Ex}_b = \frac{x_{\text{max}} + x_{\text{min}}}{2}$$

$$\text{En}_b = \frac{x_{\text{max}} - x_{\text{min}}}{6}$$

$$\text{He}_b = k$$

Where $x_{\text{max}}$ and $x_{\text{min}}$ respectively represent the upper and lower limits of the risk grade score range, $k$ is a constant, which needs to be adjusted according to the fuzzy threshold of the assessment index, and should not be too large\[5\]. In this paper, $k$ is set as 0.05.

Thus, the forward cloud generator algorithm is used to draw the benchmark cloud diagram based
on Matlab software, as shown in Figure 2.

![Benchmark cloud diagram](image)

**Figure 2. Benchmark cloud diagram**

### 3.3.2. Evaluative cloud

Refer to Table 3 to score the security risk factors and get the assessment data. Then, the reverse cloud generator is used to process the assessment data, and then the digital features of the evaluative cloud are obtained, namely:

$$ Ex_e = X = \frac{1}{n} \sum_{i=1}^{n} x_i $$  \hfill (5)

$$ En_e = \sqrt{\frac{1}{2n} \sum_{i=1}^{n} |x_i - Ex_e|^2} $$  \hfill (6)

$$ He_e = \sqrt{S^2 - En_e^2} $$  \hfill (7)

Where $S^2$ is the variance of the assessment data.

### 3.3.3. Comprehensive cloud

Comprehensive cloud is a cloud model obtained by fuzzy synthesis of evaluative cloud and weight. Its digital features are:

$$ Ex_c = \sum_{j=1}^{m} w_j Ex_{ej} $$  \hfill (8)

$$ En_c = \sqrt{\sum_{j=1}^{m} w_j En_{ej}^2} $$  \hfill (9)

$$ He_c = \sum_{j=1}^{m} w_j He_{ej} $$  \hfill (10)

Where $A_j \left(Ex_{ej}, En_{ej}, He_{ej}\right)$ is the evaluative cloud of factor $e_j$, and $w_j$ is the corresponding weight.

### 4. Case analysis of medical risk assessment

#### 4.1. Simplify indicator system based on FA

The original assessment index system for medical risk [6] is summarized in Table 4.
The above indicator system is simplified according to factor analysis. The raw data used for factor analysis was scored by the expert with reference to Table 3. The weight of each risk factor was calculated by AHP. Both the simplified assessment index system and its weight are shown in Table 5.

4.2. Case computing based on cloud model
Several experts were invited to score the simplified indicators. After the calculations of Equations (5) to (10), the results are shown in Table 6. The graphs of the evaluative cloud and the comprehensive cloud are drawn in the benchmark cloud, as shown in Figure 3 to Figure 8.
It can be seen from Figure 3 to Figure 7 that in the evaluative cloud, the risk level of $e_1$, $e_2$, and $e_3$ is “II”, namely “medium risk”. And it satisfies that $e_1 > e_3 > e_2$. The risk level of $e_4$ is “I”, namely “relatively low risk”. The risk level of $e_5$ is “III”, namely “relatively high risk”. Thus it can be concluded that the environment of medical institutions is the largest source of medical risk. The second largest source of risk is people, including medical staff, patients, and patients’ families. Figure 8 shows that the general assessment of medical risk is “II”, namely “medium risk”.

5. Conclusion
This paper studied the problem of risk assessment with multiple risk factors. Aiming at the problem of many risk factors, the simplified risk assessment index system is obtained by factor analysis method. In view of the uncertainty of the risk assessment process, a risk assessment method based on cloud model was proposed, which can effectively realize the conversion between risk level and quantitative value. Case of medical risk assessment in this paper demonstrated the effectiveness of the proposed method. The method has some reference value to solve the problem of complex multi-factor risk assessment.

References
[1] Jiang H Z, Li Z Y, Feng T, et al. Factor analysis of selective laser melting process parameters with normalised quantities and Taguchi method[J]. Optics and Laser Technology, 2019, 119.
[2] Kilic, B., Ucler, C. Stress among ab-initio pilots: A model of contributing factors by AHP[J]. Journal of Air Transport Management, 2019, 80.
[3] Li J, Wang M W, Xu P, et al. (2014) Classification of stability of surrounding rock using cloud model. Chinese Journal of Geotechnical Engineering, 36(1): 83–87.
[4] Yan F, Xu K L. Methodology and case study of quantitative preliminary hazard analysis based on cloud model[J]. Journal of Loss Prevention in the Process Industries, 2019, 60: 116-124.
[5] Bai Y, Zhang Z F. (2016) Risk conduction assessment of weapon development projects by using fuzzy cloud model. Journal of Harbin Institute of Technology, 48(10): 168–175.
[6] Yang H. Research on medical risk based on AHP[J]. Hosp Admin J Chin PLA, 26(7): 626-629.