

Study on Complex System Evaluation Method Based on PCA and Grey Relative Degree

Yongchun Li¹, Baowei Song^{1,2}, Jun Jiang^{1,2}, Guang Pan^{1,2}, Ming Chang¹

¹Northwestern Polytechnical University, Xi'an, China

²Institute of Underwater Vehicle, Northwestern Polytechnical University, Xi'an, China

Email: jiangjun@nwpu.edu.cn

Received October 25, 2011; revised November 24, 2011; accepted December 1, 2011

ABSTRACT

The complexity and limited data samples in evaluation of complex system are obstacles for traditional evaluation method. Based on Grey relative degree theory and principal components analysis (PCA) method, a novel systemic evaluation method is put forward in this paper. Firstly, standardization method is modified for evaluation objective, and the method includes expression (2) and (3). Secondly, for few schemes of complex system, grey relative degree, expression (5), is substituted for relation coefficient. At last, validity of the method is verified by evaluating 3 schemes of a type of UUV.

Keywords: Complex System; Systemic Evaluation; PCA; Grey Relative Degree

1. Introduction

At present, along with the development of science and technology, most of objectives we deal with, whose fundamental properties include otherness, diversity, pertinence and whole property, are relative to lots of factors. So, complex system and relative theories and methods have been widely used in our society, commerce, research, and so on. At the same time, it is important to estimate and evaluate complex system, because there are many applications of complex systems ineffective, and some complex systems did not achieve the desired results and caused the user heavy losses, which had seriously affected the development of complex systems. As a result, it is necessary for the research of evaluation methods which could provide a theoretical basis for development of complex system. However, evaluation methods for complex systems have been greatly improved under the more attention of researchers, such as "integrated score", "cost-benefit analysis", "analytic hierarchy process", "data envelopment analysis", "fuzzy comprehensive evaluation", "grey relative analysis" and so on. However, there are some shortages in the current evaluation methods, especially all of them do not consider the complexity and limited data samples of complex system [1,2]. The complexity of complex system includes multitudinous indices, nonlinear relationship among indices. Principal component analysis (PCA) has been widely used to evaluate complex system, and achieved good effect. But, due to relation coefficient method is based on quantity data,

traditional PCA could not deal with limited data accurately, especially small sample data.

In this paper, a novel complex system evaluation method is put forward based on principal component analysis (PCA) and Grey relative degree. PCA is a method for reducing the space dimension of samples and delete redundant information. And grey relative degree is introduced to deal with small sample data instead of relative coefficient method.

2. Method of Evaluation

Principal component analysis is one of multivariate analysis techniques, and it is an efficient method for the high dimensional, correlated input space into the appropriate lower dimensional subspace. In the input data space, PCA searches for directions with the biggest data variations, provided that these directions are orthogonal, and uses them as a primary axis of a new coordinate system in which the input space is then projected [3]. Thereby PCA transforms correlated variables into the set of new uncorrelated variables which are called principal components (PCs).

2.1. Standardization of Data

By taking a number of the production process data at the normal operating conditions, a data matrix $X(x_{ij}) \in S_{m \times n}$ can be created for the process variable. In which each column corresponds to a set of variables (as the sample point data at the same scheme), each row corresponds to

a sample (as a variable values at different schemes). That is to say, m is the number of indices of complex system, n is the number of schemes of complex system.

2.1.1. Score Method

Usually, for lots of data of a variable, the standardization method is based on Score method [4-6]. That is,

$$z_{ij} = \frac{x_{ij} - \bar{x}_{ij}}{s_i} \quad (i = 1, 2, \dots, m, j = 1, 2, \dots, n) \quad (1)$$

where, \bar{x}_{ij} is the mean value of the i th variable data, and s_i is the covariance of the i th variable data.

It is obvious that the accuracy of s_i depends on the numbers of data. However, for complex system evaluation, the number of system schemes is small, and, although expression (1) considers data of the variable, it is not concern of relationship between evaluation objective and the variable. So, it is not suitable for evaluating complex system.

2.1.2. Standardization for Objective

For the evaluation objective of complex system, the categories of indices are two kinds: cost and benefit type. So, considering the evaluation objective, a new standardization method is presented,

1) Cost type variable

$$z_{ij} = \frac{x_{ij}}{\max_{1 \leq j \leq n} (x_{ij})} \quad (i = 1, 2, \dots, m) \quad (2)$$

2) Benefit type variable

$$z_{ij} = \frac{\min_{1 \leq j \leq n} (x_{ij})}{x_{ij}} \quad (i = 1, 2, \dots, m) \quad (3)$$

The standardization matrix is $Z(z_{ij}) \in S_{m \times n}$.

2.2. Coefficient Matrix

2.2.1. Correlation Matrix

According to the standardization matrix Z , a correlation matrix $R(r_{ij})$ can be created. It has contained the correlation degree between all the measured variables. r_{ij} expresses the correlation degree between the i th variable and the j th variable. That's,

$$r_{ij} = \frac{\sum_{k=1}^n (z_{ik} - \bar{z}_i)(z_{jk} - \bar{z}_j)}{\sqrt{\sum_{k=1}^n (z_{ik} - \bar{z}_i)^2} \sqrt{\sum_{k=1}^n (z_{jk} - \bar{z}_j)^2}} \quad (i, j = 1, 2, \dots, m), \quad (4)$$

where, \bar{z}_i is the mean value of the i th variable standard

data.

It is obvious that the accuracy of correlation degree is direct to the amount of data. That is to say, the more is the amount of schemes, the more is accuracy of the evaluation result for complex system.

2.2.2. Grey Relative Degree Matrix

Grey relation analysis is considered to be an analysis of the geometric proximity between different discrete sequences within a system. The proximity is described by the grey relative degree, which is regarded as a measure of the similarities of discrete data that can be arranged in sequential order.

Assumption sequences:

$$X_i = (x_i(1) \quad x_i(2) \quad \dots \quad x_i(n)) \quad (i = 1, 2, \dots, m).$$

Then, the grey relative degree of X_i and X_j is (5): where, $\rho \in [0, 1]$ is of the distinguished coefficient. A value of ρ is the smaller and the distinguished ability is the larger, $\rho = 0.5$ generally.

The eigenvalues of matrix R are $\lambda_i \quad (i = 1, 2, \dots, m)$, and $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$.

Thus, a transformation matrix P can be obtained by the following equation:

$$(\lambda_i I - R)P_i = 0 \quad (i = 1, 2, \dots, m),$$

where, P_i is the eigenvector of the matrix R corresponding to the eigenvalue, $\lambda_i \quad (i = 1, 2, \dots, m)$.

2.3. Principle Components

The number of principal component should be determined by calculating the variance contribution rate. If the number of the principal component was selected as k , then the information which contain in the measured variables can be obtained by calculating the variance contribution rate.

$$\min(k), \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^m \lambda_i} \geq 0.85 \quad (6)$$

The principal component model can be defined as:

$$Y_i = (x_1 \quad x_2 \quad \dots \quad x_m) \cdot P_i \quad (i = 1, 2, \dots, k), \quad (7)$$

where, x_i is the i th variable of complex system.

2.4. PCA Evaluation Model

PCA integrated output can be defined as:

$$r_{ij} = \frac{1}{n} \sum_{k=1}^n \frac{\min_{1 \leq p \leq m} \min_{1 \leq t \leq n} |x_i(t) - x_p(t)| + \rho \max_{1 \leq p \leq m} \max_{1 \leq t \leq n} |x_i(t) - x_p(t)|}{|x_i(k) - x_j(k)| + \rho \max_{1 \leq p \leq m} \max_{1 \leq t \leq n} |x_i(t) - x_p(t)|} \quad (i, j = 1, 2, \dots, m), \quad (5)$$

$$Y = \sum_{i=1}^k w_i Y_i \tag{8}$$

where, w_i is the variance contribution rate of the principal components in the PCA evaluation model, is given by

$$w_i = \frac{\lambda_i}{\sum_{i=1}^k \lambda_i} (i = 1, 2, \dots, k).$$

3. Application and Analysis

For example, there are 3 schemes (A_1, A_2, A_3) of unmanned underwater vehicle (UUV) to be evaluated. 11 attributes should be considered, which are listed in **Table 1**. x_1 is MTBF, x_2 is range, x_3 is navigation depth, x_4 is navigation velocity, x_5 is recognition angle, x_6 is searching distance, x_7 is anti-jamming probability, x_8 is radius of load action, x_9 is the load capability, x_{10} is gyroidal radius of trajectory, x_{11} is MTTR.

For the evaluation objective, x_{10} and x_{11} belong to cost

type attribute, others belong to benefit type attribute.

According to expressions (2) and (3), the standardization matrix Z is:

$$Z = \begin{bmatrix} 0.89 & 1 & 0.86 \\ 0.9 & 1 & 0.75 \\ 0.6 & 1 & 0.8 \\ 0.83 & 0.94 & 1 \\ 0.88 & 1 & 0.92 \\ 0.9 & 0.8 & 1 \\ 0.94 & 1 & 0.9 \\ 0.83 & 0.95 & 1 \\ 0.96 & 1 & 0.92 \\ 0.93 & 1 & 0.88 \\ 0.83 & 1 & 0.83 \end{bmatrix}$$

According to the expression (5), the coefficient matrix

R is:

$$R = \begin{bmatrix} 1 & 0.835 & 0.680 & 0.641 & 0.881 & 0.622 & 0.843 & 0.653 & 0.794 & 0.888 & 0.845 \\ 0.838 & 1 & 0.694 & 0.590 & 0.784 & 0.601 & 0.763 & 0.603 & 0.728 & 0.790 & 0.778 \\ 0.711 & 0.719 & 1 & 0.554 & 0.664 & 0.441 & 0.679 & 0.565 & 0.644 & 0.682 & 0.765 \\ 0.608 & 0.550 & 0.471 & 1 & 0.667 & 0.704 & 0.588 & 0.975 & 0.592 & 0.580 & 0.700 \\ 0.878 & 0.775 & 0.624 & 0.691 & 1 & 0.623 & 0.858 & 0.703 & 0.879 & 0.838 & 0.782 \\ 0.628 & 0.601 & 0.397 & 0.733 & 0.637 & 1 & 0.606 & 0.727 & 0.598 & 0.606 & 0.526 \\ 0.861 & 0.780 & 0.654 & 0.659 & 0.878 & 0.633 & 1 & 0.670 & 0.930 & 0.946 & 0.772 \\ 0.621 & 0.563 & 0.484 & 0.975 & 0.679 & 0.698 & 0.601 & 1 & 0.605 & 0.593 & 0.713 \\ 0.823 & 0.755 & 0.644 & 0.674 & 0.897 & 0.639 & 0.933 & 0.685 & 1 & 0.892 & 0.749 \\ 0.899 & 0.802 & 0.669 & 0.645 & 0.857 & 0.626 & 0.945 & 0.656 & 0.884 & 1 & 0.797 \\ 0.817 & 0.737 & 0.709 & 0.687 & 0.768 & 0.463 & 0.711 & 0.700 & 0.677 & 0.744 & 1 \end{bmatrix}$$

Eigenvalues of R are:

$$\begin{aligned} \lambda_1 &= 8.1556, \lambda_2 = 1.0208, \lambda_3 = 0.5824, \lambda_4 = 0.3829, \\ \lambda_5 &= 0.2915, \lambda_6 = 0.1752, \lambda_7 = 0.1586, \lambda_8 = 0.1132, \\ \lambda_9 &= 0.0542, \lambda_{10} = 0.0410, \lambda_{11} = 0.0247. \end{aligned}$$

$$\text{And, } k = 3, \frac{\sum_{i=1}^3 \lambda_i}{\sum_{i=1}^{11} \lambda_i} = 0.8872 > 0.85,$$

Table 1. Attributes of UUV.

	A_1	A_2	A_3
x_1	2500	2800	2400
x_2	18	20	15
x_3	150	250	200
	A_1	A_2	A_3
x_4	40	45	48
x_5	44	50	46
x_6	900	800	1000
x_7	0.84	0.89	0.80
x_8	35	40	42
x_9	250	260	240
x_{10}	75	70	80
x_{11}	30	25	30

$$P_1 = \begin{pmatrix} 0.3234 \\ 0.3035 \\ 0.2744 \\ 0.2712 \\ 0.3222 \\ 0.2580 \\ 0.3277 \\ 0.2749 \\ 0.3240 \\ 0.3286 \\ 0.2970 \end{pmatrix}, P_2 = \begin{pmatrix} 0.1970 \\ 0.2036 \\ 0.2336 \\ -0.5487 \\ 0.1050 \\ -0.4082 \\ 0.1812 \\ -0.5285 \\ 0.1339 \\ 0.2166 \\ 0.0971 \end{pmatrix}, P_3 = \begin{pmatrix} 0.0437 \\ -0.0663 \\ -0.5803 \\ -0.1762 \\ 0.1888 \\ 0.3984 \\ 0.2793 \\ -0.1833 \\ 0.2946 \\ 0.1796 \\ -0.4481 \end{pmatrix}$$

$$w_1 = 0.8357, w_2 = 0.1046, w_3 = 0.0597$$

That is,

$$\begin{aligned} Y &= w_1 Y_1 + w_2 Y_2 + w_3 Y_3 \\ &= 0.2935x_1 + 0.2710x_2 + 0.2191x_3 + 0.0412x_4 \\ &\quad + 0.2915x_5 + 0.1967x_6 + 0.3095x_7 \\ &\quad + 0.1635x_8 + 0.3024x_9 + 0.3080x_{10} + 0.2316x_{11} \end{aligned}$$

In the evaluation program, the standardization data should be taken as score of each variable. Then, the results of each scheme are,

$$Y_{A1} = 2.2999, Y_{A2} = 2.5780, Y_{A3} = 2.3205.$$

According to the evaluation results, the conclusion is that Scheme A_2 is the best, Scheme A_3 is better and Scheme A_1 is the last. And the sequence is consistent with the evaluation result by classic PCA.

At the same time, in the eigenvector P_2 , coefficients of attributes x_4 , x_6 and x_8 are all less than zero. It means that the three attributes need to be improved, and it is true.

4. Conclusions

In this paper, a new method for evaluating complex system, especially in small sample case, is presented. The innovation points of the method include:

- 1) In small sample case, the relation coefficient is not accurate to the relationship between attribute. Relation coefficient is substituted for grey relative degree in PCA.
- 2) Standardization method of each attribute value is modified to aim at evaluating objective. In general, each

attribute falls into cost type or benefit type.

3) According to analyzing a UUV schemes evaluation problem, the validity of the method is verified.

REFERENCES

- [1] X. Bian, H. P. Tian and Y. J. Guo, "The Supplier Dynamic Evaluation Method with the Character of Stimulation," *Nankai Business Review*, Vol. 7, No. 4, 2004, pp. 55-58.
- [2] W. M. Hong and J. Liu, "A Survey on Supply Chain Performance Measurement," *Commercial Research*, No. 8, 2006, pp. 17-21.
- [3] M. Wise, "Adapting Multivariate Analysis for Monitoring and Modeling of Dynamic Systems," Master Thesis, University of Washington, Washington DC, 1991.
- [4] F. B. Jonathan and F. S. Stephen, "A Tradeoff Analysis for Rough Terrain Cargo Handlers Using the AHP: An Example of Group Decision Making," *IEEE Transactions on Engineering Management*, Vol. 37, No. 3, 1990, pp. 222-228. [doi:10.1109/17.104292](https://doi.org/10.1109/17.104292)
- [5] K. Cengiz, C. Ufuk and U. Ziya, "Multi-Criteria Supplier Selection Using Fuzzy AHP," *Logistics Information Management*, Vol. 16, No. 6, 2003, pp. 382-394. [doi:10.1108/09576050310503367](https://doi.org/10.1108/09576050310503367)
- [6] R. T. Umut, E. Arzum and A. Gner, "Solving a Layout Design Problem by Analytic Hierarchy Process (AHP) and Data Envelopment Analysis (DEA) Approach," *Proceeding of SPIE: The International Society for Optical Engineering*, Vol. 5262, Providence, 29 October 2004, pp. 232-239.