

Application of PCA-LINMAP Coupling Algorithm and Model in Determining Weight of Influence Factors of Surface Deformation

*** Xiaogang XIA, Yunfeng YANG, Si ZHANG**

School of Science, Xi'an University of Science and Technology, Xi'an, 710054, China

* Tel.: 086-029-85583136, fax: 086-029-85583136

* E-mail: xiayu1978@126.com

Received: 18 July 2014 / Accepted: 30 October 2014 / Published: 30 November 2014

Abstract: Based on analyzing the limitations and shortcomings of the traditional method of empowerment, the PCA-LINMAP coupling model which uses to determine the weight of influence factors of surface deformation is established by principal component analysis and linear programming. This method completely considers the difference and the merits order of the samples, and overcomes the one-sidedness of the traditional method of empowerment in determining weight of index. The index system can reflect the object to being evaluated comprehensively and objectively, at the same time, provide an ideal way for improving the scientific of evaluation. Examples show that the method is scientific and reasonable; it also has extensive application value. Copyright © 2014 IFSA Publishing, S. L.

Keywords: Surface subsidence, PCA-LINMAP coupling algorithm, Influence factors, Weight.

1. Introduction

The surface subsidence deformation which is caused by underground mining is a complicated mechanics process. The factor set that influences the surface subsidence deformation is a mutually affiliation and restriction complex system (Yu and Zhang, 2004; He et al., 1994; Guo and Zhang, 2007; Xia et al., 2008; Xia 2008). In addition, there is great difference between factors which influence the surface subsidence deformation. The prediction and evaluation are much objective and exact if strictly consider every detailed factor, but more factors will bring about much inconvenience on practical prediction and evaluation. So accurately predict the impacts of each factor, there is an important significance not only to the research of mineral resources exploitation, but also to the mining

disasters control, the ecological environment of the mining area improving and the sustainable development strategy implementing and so on.

The comprehensive analysis of multiple indicators system is the effective tool to improve the overall evaluation. The same model of this system analysis can be used in multi-objective decision, multi-factor analysis and system quality evaluation and so on, according to the difference of the meaning and purpose. And to solve this kind of problem which needs to reveal weight of each target (indexes or factors) in the whole system. In this sense, to determine the index weights is an important part of comprehensive system analysis. In addition, once the index weights are identified, the key factors or the key indicators which affects the comprehensive evaluation effect of the system will be known it also provides the theoretical basis for system research and

improvement (Guo et al., 2013; Jin et al., 2012; Qin, 2003; Zhang and Wang 2008).

The basic idea of PCA-LINMAP coupling algorithm is beginning from the given sample-index data. The sample merit order and the principal component which significantly reflect the sample differences can be found by means of PCA, then the weights of each index can be gotten through inputting the ordered pair to LINMAP (Liu and Zhang, 2012; Zhou 2006; Chen et al., 2004).

2. PCA-LINMAP Coupling Weighting Algorithm and Model

PCA-LINMAP coupling model consists of two basic sub-models, i.e. PCA model and LINMAP model. The specific application process is deriving the sample merit order from the initial decision matrix by PCA sub-model, and then, the weights of each index can be gotten according to the order pair of samples by LINMAP sub-model.

Where PCA is the abbreviation of principal component analysis, LINMAP is the abbreviation of linear programming techniques for multi-dimensional analysis preference.

2.1. The Basic Principle and Algorithm of PCA Sub-model

The principal component analysis method in multivariate statistical analysis is widely used to evaluate (comprehensive evaluation) and sort in engineering calculation field, because of the simplicity of its theory and the objective in weighting and so on. Its basic principle is to find few irrelevant (or independent) comprehensive indexes as basic indexes to evaluate the subjects from more indexes. To achieve the purpose of revealing the relationship among variables and simplifying the data and analyzing total data characteristics, the basic algorithm is following as (Qin 2003).

Step 1: Collecting p dimension random vectors, the n samples $x = (x_1, x_2, \dots, x_p)^T$ ($i = 1, 2, \dots, n$; $n > p$) will construct sample array

$$X = \begin{bmatrix} x_1^T \\ x_2^T \\ \dots \\ x_n^T \end{bmatrix} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{pmatrix}, \quad (1)$$

Step 2: Making the following transform form the elements of the sample matrix X

$$y_{ij} = \begin{cases} x_{ij}, & \text{The positive matrix} \\ -x_{ij}, & \text{The inverse matrix} \end{cases}$$

$$\text{So } Y = [y_{ij}]_{n \times p}.$$

Step 3: Making standard transform of the elements of matrix Y

$$z_{ij} = \frac{(y_{ij} - \bar{y}_j)}{s_j} \quad i = 1, 2, \dots, n, j = 1, 2, \dots, p, \quad (2)$$

$$\text{where } \bar{y}_j = \frac{\sum_{i=1}^n y_{ij}}{n}, s_j^2 = \frac{\sum_{i=1}^n (y_{ij} - \bar{y}_j)^2}{n-1}.$$

$$Z = \begin{bmatrix} z_1^T \\ z_2^T \\ \dots \\ z_n^T \end{bmatrix} = \begin{pmatrix} z_{11} & z_{12} & \dots & z_{1p} \\ z_{21} & z_{22} & \dots & z_{2p} \\ \dots & \dots & \dots & \dots \\ z_{n1} & z_{n2} & \dots & z_{np} \end{pmatrix}, \quad (3)$$

Step 4: Calculating of sample correlation matrix from the standardized matrix Z

$$R = [r_{ij}]_{p \times p} = \frac{Z^T Z}{n-1}, \quad (4)$$

$$\text{where } r_{ij} = \frac{\sum_{k=1}^n z_{ki} \cdot z_{kj}}{n-1}, \quad i, j = 1, 2, \dots, p.$$

Step 5: Solving the characteristic equation of the sample correlation coefficient R

$$|R - \lambda I_p| = 0, \quad (5)$$

The sorting result of eigenvalues is following $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$.

Step 6: Making use of the principle

$$\frac{\sum_{j=1}^k \lambda_j}{\sum_{j=1}^p \lambda_j} \geq 0.85 \text{ to confirm the principal components,}$$

letting utilization ratio of information is more than 85 %. For each of the λ_j , $j = 1, 2, \dots, m$, solving the function $Rb = \lambda_j b$, so the unit eigenvectors can be gained.

$$b_j^0 = \frac{b_j}{\|b_j\|}, \quad (6)$$

Step 7: Finding out the m principal component of $z_i = (z_{i1}, z_{i2}, \dots, z_{ip})^T$, $i = 1, 2, \dots, n$, $u_{ij} = z_i^T b_j^0$, $j = 1, 2, \dots, m$.

So the decision matrix can be conclude.

$$U = \begin{bmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_n^T \end{bmatrix} = \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1m} \\ u_{21} & u_{22} & \cdots & u_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ u_{n1} & u_{n2} & \cdots & u_{nm} \end{pmatrix}, \quad (7)$$

where u_i presents the principal component vector belonging to the i -th sample, $i = 1, 2, \dots, n$.

Step8 Reducing the dimension system to a one-dimensional system by selecting the appropriate principal component value function model.

2.2. The Basic Principle of LINMAP Sub-model

The standardization decision matrix $[z_{ij}]_{n \times p}$ indicates that there are n sample points at p dimension index space. Now assuming $(z_1^*, z_2^*, \dots, z_p^*)^T$ expresses the ideal point of the most preference sample index space of the decision maker, so the weighted Euclid distance square S_i between any sample point $(z_{i1}, z_{i2}, \dots, z_{ip})^T$ of index space and ideal point is following as:

$$S_i = \sum_{j=1}^p w_j (z_{ij} - z_j^*)^2, i = 1, 2, \dots, p, \quad (8)$$

where $w_j (j = 1, 2, \dots, p)$ presents the weight square of the i -th index.

Then, according to the preference order to PCA of the decision maker, the ascending sequence of S_i can be obtained if converting S_i from formula (8), where w_j and z_j^* are unknown numbers in formula (8), the purpose of LINMAP method is to solve w_j .

2.3. The Basic Algorithm of LINMAP Sub-model

Definition 1. Let the set of the sample order pair (k, l) as

$$Q = \{(k, l) \mid \text{the } k\text{-th sample is better than the } l\text{-th sample}\}, \quad (9)$$

If $S_l \geq S_k$, because the priority of formula (8) and the order pair is consistent, so the inconsistency degree is zero; if $S_l \leq S_k$, because the priority of

formula (8) and the order pair is inconsistent, the inconsistency degree depends on $S_k - S_l$. So,

Definition 2. If the inconsistency degree between samples sorting by formula (8) and the order pair (k, l) was recorded as $(S_l - S_k)$, so,

$$(S_l - S_k) = \begin{cases} 0 & S_l \geq S_k \\ S_k - S_l & S_l < S_k \end{cases}, \quad (10)$$

$$= \max\{0, (S_k - S_l)\}$$

The total inconsistency degree can be gotten through adding the inconsistency degree of all the order pairs (k, l) in Q , it is called inconsistency degree B , that is

$$B = \sum_{(k, l) \in Q} (S_l - S_k), \quad (11)$$

So the consistency G can be defined by the similar method,

$$G = \sum_{(k, l) \in Q} (S_l - S_k), \quad (12)$$

where

$$(S_l - S_k) = \begin{cases} S_l - S_k & S_l \geq S_k \\ 0 & S_l < S_k \end{cases}, \quad (13)$$

$$= \max\{(S_l - S_k), 0\}$$

Lets $h = G - B$, h is obviously a non-negative integer when the sorting according to formula (8) and the order pairs (k, l) is consistent.

Furthermore,

$$\begin{aligned} G - B &= \sum_{(k, l) \in Q} (S_l - S_k)^+ - \sum_{(k, l) \in Q} (S_l - S_k)^- \\ &= \sum_{(k, l) \in Q} [(S_l - S_k)^+ - (S_l - S_k)^-] \\ &= \sum_{(k, l) \in Q} (S_l - S_k) \\ &= h \end{aligned} \quad (14)$$

Let $\lambda_{kl} = \max\{0, (S_k - S_l)\}$, so the square w_j of index weight can be obtained through the linear programming problem, following as

$$\min \sum_{(k, l) \in Q} \lambda_{kl}, \quad (15)$$

$$\text{s.t.} \begin{cases} (S_l - S_k) + \lambda_{kl} \geq 0, \text{all } (k, l) \in Q \\ \sum_{(k, l) \in Q} (S_l - S_k) = h \\ \lambda_{kl} \geq 0, \text{all } (k, l) \in Q \end{cases}, \quad (16)$$

Combining formula (1) and S_i , the LINMAP sub-model can be gotten, following as

$$\min \sum_{(k,l) \in Q} \lambda_{kl}, \quad (17)$$

$$\text{s.t.} \begin{cases} \sum_{j=1}^p w_j (z_{lj}^2 - z_{kj}^2) - 2 \sum_{j=1}^p v_j (z_{lj} - z_{kj}) + \lambda_{kl} \geq 0, \\ \sum_{j=1}^p w_j \sum_{(k,l) \in Q} (z_{lj}^2 - z_{kj}^2) - 2 \sum_{j=1}^p v_j \sum_{(k,l) \in Q} (z_{lj} - z_{kj}) = h, \\ w_j > 0, j=1, 2, \dots, p \\ \lambda_{kl} \geq 0, \forall (k,l) \in Q \\ v_j = w_j z_j^* \text{ unrestrict } \end{cases}, \quad (18)$$

3. Engineering Application

As described in the introduction, due to the diversity of surface subsidence deformation factors, also there are interaction effects among the factors. So determining the comprehensive influence of surface subsidence deformation factors and the influence difference of each factor to surface subsidence deformation, not only has a great

theoretical significance, but also has important guiding significance to the production practice.

The surface deformation system is a complex open system, the factors influencing the deformation are multi-faceted. Generally speaking, it can be divided into geological factors and engineering factors (Xia et al., 2008; Xia 2008; Wu et al., 2011; Xia 2005). At the same time, each category can be further divided into specific factors. According to previous research results, the geological factors affecting surface deformation is classified into the following categories:

x_1 : durability coefficient (MPa);

x_2 : mining depth (m);

x_3 : mining thickness (m);

x_4 : tilt angle ($^{\circ}$);

x_5 : topsoil thickness (m)

According to 208 typical observation station data of surface movement in the literature (Yu and Zhang, 2004; He et al., 1994), screening out 8 measured data as the sample, see Table 1.

The eigenvalue and the eigenvector of correlation coefficient matrix can be gained through inputting the observation data into PCA model, see Table 2.

So the principal component analysis result (merit evaluation and sorting) can be gotten by PCA model, see Table 3.

Table 1. The main parameters of each typical observation stations.

Number of stations	Durability coefficient x_1	Mining depth x_2	Mining thickness x_3	Tilt angle x_4	Topsoil thickness x_5
1	3.2	60	2.1	30	20
2	3.74	42	1.45	10.5	5
3	4.0	175	1.6	13	12
4	6.0	280	2.4	12	17.4
5	3.5	98.5	2.0	16	64
6	2.5	181	1.94	9.5	110
7	1.3	325	8.2	4.3	197
8	5.0	130	1.9	37	34.2

Table 2. The eigenvalues and eigenvectors of correlation coefficient matrix.

Evaluation groups		Eigenvectors (Principle component part)		
		e_1	e_2	e_2
Number of index	1	-0.4083	-0.6809	-0.1502
	2	0.4076	-0.6623	-0.2254
	3	0.5077	-0.0166	-0.3636
	4	-0.3686	0.2549	-0.8684
	5	0.5229	0.1805	-0.2008
Eigenvalue		3.2513	0.9210	0.6563
Contribution ration		0.6503	0.1842	0.1313
Cumulative contribution ration		0.6503	0.8345	0.9658

Table 3. The results of principal component analysis (the rank program).

Number of stations	The evaluation value of the first principal component	The evaluation value of the second principal component	The evaluation value of the third principal component	Comprehensive evaluation	
				Evaluation values	Sorting results
1	-3.9022	3.1886	-2.2404	-2.2444	7
2	-2.6899	1.6735	3.2406	-1.0155	6
3	-1.0039	-1.1608	1.4476	-0.6766	5
4	0.2752	-4.5459	0.4980	-0.5930	4
5	-0.7487	1.5033	0.6895	-0.1194	3
6	2.6272	0.0971	1.4280	1.9139	2
7	9.3858	-2.1475	-0.1356	5.6902	1
8	-3.9435	1.3917	-4.9277	-2.9551	8

The merit order O is following as

$O = \{\text{observation station 7,}$
 observation station 6,
 observation station 5,
 observation station 4,
 observation station 3,
 observation station 2,
 observation station 1,
 observation station 8}\}.

Then the order set Q of observation stations is following as

$Q = \{(7,6),(7,5),(7,4),(7,3),$
 $(7,2),(7,1),(6,5),(6,4),$
 $(6,3),(6,2),(6,1),(5,4),$
 $(5,3),(5,2),(5,1),(4,3),$
 $(4,2),(4,1),(3,2),(3,1),$
 $(2,1),(1,8)\}$

Putting Q into LINMAP model (18), the weight square vector w of the five indexes can be found out by the simplex method (Su 2004; Zhang 2010), following as

$$w = (0.3473 \quad 0.7859 \quad 0.5193 \quad 0.4806 \quad 0.4265)^T$$

So the five index weight vector \bar{w} as

$$\begin{array}{c} \bar{w} = (0.5893 \quad 0.8865 \quad 0.7206 \quad 0.6932 \quad 0.6531)^T \\ \xrightarrow{\text{normalization}} \\ (0.1663 \quad 0.2502 \quad 0.2304 \quad 0.1957 \quad 0.1843)^T \end{array}$$

4. Conclusion

The conclusions were gotten due to the above studying:

1) The PCA-LINMAP coupling algorithm and model is used firstly in prediction of surface subsidence deformation, it overcomes the disadvantage of traditional statistical method by the sample size, provides a fast and effective method for the engineering prediction and evaluation.

2) The method is proved to be feasible by the instance, also the affecting difference of each factor

can be differentiated by making use of this method to evaluate the surface subsidence deformation. In the same time, these results of method provide the basis for the comprehensive evaluation and prediction of the surface subsidence deformation.

Acknowledgements

The authors receive the financial support from NSFC(71103143), NSFS(2013KJXX-40), NPFC (20110491672,2012T50809), SSFE(12JK0858), XUST(201136).

References

- [1]. X. Y. Yu, E. Q. Zhang. Mining damage research, *Coal Industry of Publishing House*, 2004.
- [2]. G. Q. He, L. Yang, G. D. Ling et al, Mining subsidence research, *China University of Mining and Technology Press*, 1994.
- [3]. G. L. Guo, G. L. Zhao, Use of grey system model in subsidence prediction, *Journal of China University of Mining & Technology*, Vol. 26, Issue 4, 1997, pp. 62-65.
- [4]. X. G. Xia, Q. X. Huang, Sh. G. Zhang, Application of PCA method in grading surface subsidence under the condition of sub-critical extraction, *Journal of Mining & Safety Engineering*, Vol. 25, Issue 1, 2008, pp. 54-58.
- [5]. X. G. Xia, Determination of weight of factors in surface subsidence by matter element analysis, *Journal of Xi'an University of Science and Technology*, Vol. 28, Issue 2, 2008, pp. 244-248.
- [6]. W. B. Guo, K. Zh. Deng, Y. F. Zou, Study on artificial neural network method for calculation of subsidence coefficient, *Chinese Journal of Geotechnical Engineering*, Vol. 25, Issue 2, 2013, pp. 212-215.
- [7]. J. L. Jin, L. B. Zhang, Sh. W. Zhang et al, Application of analytic hierarchy process to environmental impact assessment of water resources project, *System Engineering Theory Methodology Application*, Vol. 12, Issue 2, 2012, pp. 187-192.
- [8]. Sh. K. Qin, Theory and application of synthesis appraise, *Publishing House of Electronics Industry*, 2003.
- [9]. W. N. Zhang, X. L. Wang, The evaluation model of the enterprise's green competitiveness and its

- application, *Science and Technology Management Research*, Issue 20, 2008, pp. 36-38.
- [10]. Y. Ch. Liu, C. Q. Zhang, Assessment of power supply structure optimization by using kernel PCA-LINMAP, *East China Electric Power*, Vol. 34, No. 8, 2012, pp. 88-90.
- [11]. W. K. Zhou, A method for deciding objective weight in group decision-making of multiple objective under fuzzy preference, *Mathematics in Practice Theory*, Vol. 36, Issue 3, 2006, pp. 33-38.
- [12]. Sh. Q. Chen, Q. J. Xu, Ch. Zh. Yan, Applications of principal components analysis method to weight determination in fuzzy recognition model for water quality assess, *Journal of Kunming University of Science and Technology, Science and Technology*, Vol. 33, Issue 2, 2004, pp. 78-81.
- [13]. X. G. Xia, Study on the factors of the degree of top coal caving in steep seam, Master Thesis, *Xi'an University of Science and Technology*, Xi'an, 2005.
- [14]. J. M. Su, Application of MATLAB toolbox, *Publishing House of Electronics Industry*, 2004.

2014 Copyright ©, International Frequency Sensor Association (IFSA) Publishing, S. L. All rights reserved.
(<http://www.sensorsportal.com>)



International Frequency Sensor Association (IFSA) Publishing

Digital Sensors and Sensor Systems: Practical Design

Sergey Y. Yurish



The goal of this book is to help the practitioners achieve the best metrological and technical performances of digital sensors and sensor systems at low cost, and significantly to reduce time-to-market. It should be also useful for students, lectures and professors to provide a solid background of the novel concepts and design approach.

Book features include:

- Each of chapter can be used independently and contains its own detailed list of references
- Easy-to-repeat experiments
- Practical orientation
- Dozens examples of various complete sensors and sensor systems for physical and chemical, electrical and non-electrical values
- Detailed description of technology driven and coming alternative to the ADC a frequency (time)-to-digital conversion

Formats: printable pdf (Acrobat) and print (hardcover), 419 pages

ISBN: 978-84-616-0652-8,
e-ISBN: 978-84-615-6957-1

Digital Sensors and Sensor Systems: Practical Design will greatly benefit undergraduate and at PhD students, engineers, scientists and researchers in both industry and academia. It is especially suited as a reference guide for practitioners, working for Original Equipment Manufacturers (OEM) electronics market (electronics/hardware), sensor industry, and using commercial-off-the-shelf components

http://sensorsportal.com/HTML/BOOKSTORE/Digital_Sensors.htm

Promoted by IFSA

Status of the CMOS Image Sensors Industry Report up to 2017

The report describes in detail each application in terms of market size, competitive analysis, technical requirements, technology trends and business drivers.

Order online:

http://www.sensorsportal.com/HTML/CMOS_Image_Sensors.htm