Applied Mathematics, Modeling and Computer Simulation
C.-H. Chen et al. (Eds.)

© 2022 The authors and IOS Press.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/ATDE221081

Evaluation and Optimization Method of Enterprise Data System Based on FAHP-CEEMDAN

Heng TAO[a], Jialin WAN[a,1], Kan ZHAO[a], Wenting ZHOU[b], Chengzhi YAO[a], Qinke ZHANG[a]

[a] *College of Information Science and Engineering, Jishou University, 416000, China*

[b] *Business School, Hunan University of Technology, 412007, China*

Abstract. In order to help enterprises obtain greater data analysis advantages and optimize enterprise data system, this paper proposes a data evaluation optimization model. The model uses the index weight analysis method, empirical mode decomposition algorithm by improving the collection, calculation of expert judgment weight of the signal as the data system of subjective evaluation weight value, and through the fuzzy measurement analysis method, the weight is compared with the weight of the actual data system of the enterprise, calculate the maturity of the system and reflect the current state of the data system, so as to analyze and optimize the data system and reflect the competitive advantage and data value of the enterprise The experimental results show that the data analysis model is objective and effective, which is of great help to enterprise data system optimization.

Keywords. Data analysis model, Fuzzy Analytic Hierarchy Process, Improved Set Empirical Mode Decomposition, Data Weight optimization

1. Introduction

As a product of the Internet information age, big data has long been the most valued strategic resource of enterprises in business. Enterprises have the advantage of data in competition with each other and can often gain the dominant position. However, how to manage and analyze the huge data has become a big problem to be dealt with[1-2].

With its data analysis ability, data analysis can deeply tap the potential of enterprise data, improve the ability to master data, optimize the data system of enterprises on the basis of analysis, so as to make enterprises gain competitive advantages and gain more customer trust[3].

This paper introduces a data analysis model, and proposes a fuzzy analytic hierarchy Process (FAHP) to solve the problem of how to obtain data competitive advantage[4]. And Improved Set Empirical Mode Decomposition (CEEMDAN)[5] combined data analysis model algorithm based on index weight.

1 Corresponding author: Jialin Wan, College of Information Science and Engineering, Jishou University, 416000, China; E-mail: 2015898330@qq.com

2. Related Overview

2.1. Fuzzy Analytic Hierarchy Process

Fuzzy analytic Hierarchy Process (FAHP) is improved on the basis of analytic hierarchy Process (AHP).Analytic hierarchy process (AHP) analyzes the importance of indicators in the system through expert matrix[6]. In order to evaluate and optimize the system, the goal, criterion, scheme and hierarchy of the decision-making problem are established, and the final weight of the index is determined[7] . However, there are errors caused by experts' subjective feelings. Based on questions like AHP, Buckley[2] introducing the idea of fuzzy theory, the basic idea is to use "fuzzy matrix" instead of "expert matrix" in AHP to judge the main degree of indicators by fuzzy number , and it has been proved that fuzzy plan optimization expert judgment can get better results.

2.2. Improved Set Empirical Mode Decomposition

CEEMDAN was developed by French scholar Colominas et al[5] . Compared with the traditional empirical mode decomposition (EMD), CEEMDAN adds adaptive Gaussian white noise to it, and effectively avoids mode aliasing effect through continuous iterative decomposition method, thus improving the accuracy of information sequence decomposition[8].

3. Establishment of D&A Analysis System Based on FAHP-CEEMDAN Method

3.1. Introduction of Expert Evaluation Value

In the process of analyzing the actual field and industry, the most perfect evaluation of the industry can often be obtained through the evaluation of the industry by experts.

However, there will be subjective judgment factors in the process of expert evaluation. As shown in figure 1, to some extent, the objective evaluation values of experts with the same level for a certain industry are similar. The expert evaluation values can be regarded as non-stationary characteristic signal sequences, and the signal sequences are extracted and processed. Finally, the stable characteristic sequences that meet the requirements are obtained as the objective evaluation values of experts.

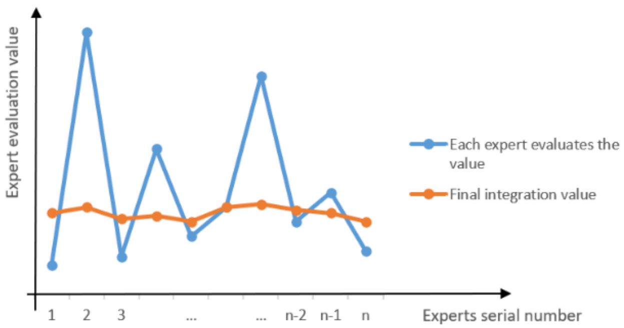

Figure 1. Graph of expert evaluation value

3.2. Introduction of Expert Evaluation Value

3.2.1. Triangle Fuzzy Number and its Scale

Triangular fuzzy number $\tilde{A} = (L, u, m)$, according to its membership function can be defined as:

$$U(X)\begin{cases} \frac{x-l}{m-l}, l \leq x \leq m \\ \frac{u-x}{u-m}, m \leq x \leq u \\ 0, x > u \text{ or } x < l \end{cases} \tag{1}$$

Where, m is the median value of triangular fuzzy number, l and m are corresponding left and right endpoints respectively.Through triangular fuzzy numbers, the importance of two indicators can be compared, and it will bring a certain degree of fuzzy influence. It will not accurately judge which indicator has higher importance, which is very meaningful for the weight importance evaluation in the actual process. According to the fuzziness of triangle fuzzy number, the scale description of the importance degree between indicators is shown in table 1[9].

Table 1. Description of triangular fuzzy number scale

The importance of the two	**Triangular fuzzy number**	**The importance of the two**	**Triangular fuzzy number**
As important	(1,1, 1)	Obviously important	(4 ,5, 6)
Potential important	(1/2,1,2)	Very important	(6,7,8)
Slightly important	(2 ,3,4)	Extremely important	(8,9,9)

3.2.2. Single Fuzzy Analytic Hierarchy Process

Let $x = \{x1, x2, \dots, xn\}$ represent the weight index set of all experts for all indicators, then the fuzzy judgment matrix represented is:

$$\tilde{X} = (\tilde{x}_{ij})n \times n = \begin{bmatrix} (1,1,1) & (l_{12}, m_{12}, u_{12}) & \cdots & (l_{ln}, m_{ln}, u_{ln}) \\ (l_{21}, m_{21}, u_{21}) & (1,1,1) & \cdots & (l_{2n}, m_{2n}, u_{2n}) \\ \vdots & \vdots & \cdots & \vdots \\ (l_{nl}, m_{nl}, u_{nl}) & (l_{21}, m_{21}, u_{21}) & \cdots & (1,1,1) \end{bmatrix} \tag{2}$$

Where, any $i, j = 1,2, \dots, n$, and $i \neq j$, should satisfy:

$$0 < l_{ij} \leq m_{ij} \leq u_{ij}; l_{ij} = \frac{1}{u_{ij}}, m_{ij} = \frac{1}{m_{ij}}, u_{ij} = \frac{1}{u_{ij}} \tag{3}$$

The fuzzy judgment matrix of Equation (2) is logarithmic:

$$\ln(\tilde{x}_{ij}) = (\ln l_{ij}, \ln m_{ij}, \ln u_{ij}), i, j = 1,2, \cdots, n \tag{4}$$

Based on Equation (4), rewrite (1) as follows:

$$u_{ij}\left(ln\left(\frac{\omega_i}{\omega_j}\right)\right)=\begin{cases}\frac{ln\left(\frac{\omega_i}{\omega_j}\right)-lnl_{ij}}{lnm_{ij}-lnl_{ij}}, ln\left(\frac{\omega_i}{\omega_j}\right)\leq lnm_{ij}\\ \frac{lnu_{ij}-ln\left(\frac{\omega_i}{\omega_j}\right)}{lnu_{ij}-lnm_{ij}}, ln\left(\frac{\omega_i}{\omega_j}\right)\geq lnm_{ij}\end{cases} \tag{5}$$

In Equation (5), $u_{ij}\left(ln\left(\frac{\omega_i}{\omega_j}\right)\right)$ is the degree $n\left(\frac{\omega_i}{\omega_j}\right)$ belonging to triangular fuzzy matrix $n\left(\frac{\omega_i}{\omega_j}\right)=(l_{ij},m_{ij},u_{ij})$. In addition, the exact priority vector is introduced:

$$\lambda = min\left\{u_{ij}\left(ln\left(\frac{\omega_i}{\omega_j}\right)\right)|i=1,2,\cdots,n;j=i+1,i+2,\cdots,n\right\}\geq 0 \tag{6}$$

According to equations (6) and (5), the objective programming model can be written as:

$$Min\varphi=(1-\lambda)^2+M*\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(\sigma_{ij}^2+\varepsilon_{ij}^2) \tag{7}$$

$$\begin{cases}ln\omega_i-ln\omega_j-\lambda\ln\left(\frac{m_{ij}}{l_{ij}}\right)+\sigma_{ij}\geq lnl_{ij}\\ -ln\omega_i+ln\omega_j-\lambda\ln\left(\frac{u_{ij}}{m_{ij}}\right)+\varepsilon_{ij}\geq -lnl_{ij}\\ \lambda\geq 0,\ln\omega_i\geq 0,i=1,\cdots,n\\ \delta_{ij}\geq 0,\varepsilon_{ij}\geq 0\\ i=1,\cdots,n-1;j=i+1,\cdots,n\end{cases} \tag{8}$$

In Equations (7) and (8), δ_{ij} and is ε_{ij} the deviation variable satisfying the constraint conditions; ω_i is the evaluation value of indicators by experts; M is an infinitely sufficiently large number.Also δ_{ij} and ε_{ij} satisfied.

$$\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\left({\sigma_{ij}}^2+{\varepsilon_{ij}}^2\right)\approx 0 \tag{9}$$

Through the above method, the evaluation value of different indexes can be calculated by expert fuzzy judgment matrix.

3.2.3. CEEMDAN

Section 2.1 introduces that experts' evaluation of indicators is influenced by subjective consciousness, which is a non-stationary characteristic signal and requires signal decomposition. Compared with the traditional EMD method, the EEMD decomposition method increases the number of integrations to reduce the reconstruction error, which greatly reduces the efficiency of obtaining signal features[10]. Ceemdan method, that is to add Gaussian white noise sequence adaptively in the decomposition process, and then decompose each inherent modal component by calculating the only participating semaphore until the mean value of the signal component no longer meets the extraction of the conditional component IMF[5]. Using this method, the objective value semaphore of expert evaluation value can be extracted efficiently and accurately.

Step 1: set the signal sequence evaluated by experts as $y=\{y1,y2,\ldots,yn\}$, add m times gaussian white noise to the sequence to be processed, and among ω satisfy:

$$\omega\sim N(0,\sigma^2) \tag{10}$$

New information sequence is formed by adding gaussian white noise.

Step 2: Decompose the information sequence EMD with gaussian white noise, obtain the components and take the mean value :IMF, and write it as $\overline{IMF_1}$:

$$r_1(t) = x(t) - \overline{IMF}_1 \tag{11}$$

Step 3: Continuously add gaussian white noise to the $r(t)$ after decomposition,obtain the mean value of the new IMF component, and continuously decompose the signal of $r(t)$. Calculate the k-th signal participation component as:

$$r_k(t) = r_{k-1}(t) - \overline{IMF}_{\mathrm{k}} \tag{12}$$

Calculate the $k + 1$ component IMF:

$$\overline{IMF}_{k+1} = \frac{1}{\mathrm{P}}\sum_{i=1}^{p} E_k(r_k(t) + \varepsilon_k E_k[\omega_i]) \tag{13}$$

Step 4: Repeat the third step until the qualified $\overline{IMF}$ component cannot be extracted, and the final calculated residual signal is:

$$x(t) = \sum_{i=1}^{n} \overline{IMF}_{\mathrm{n}} + r_n(t) \tag{14}$$

Where, r_n is the participating function, and the continuous decomposition makes the semaphore tend to be stable.

3.2.4. Data Analysis Model Expression and Analysis

Through the FAHP-CEEMDAN index analysis method, this paper proposes a $D\&A$ system based on the weight of data indicators. In this system, different experts analyze and judge the main degree of different indicators in the system, and give the proportion of indicators in the system, so as to give the direction of system optimization. At the same time, the system maturity analysis method is proposed, and the distance index between the current system and the perfect system proposed by experts is given to analyze and judge the quality of the system.

The objective evaluation information sequence of experts on all indicators is $r_t(C_i)(i = 1,2,\cdots,m)$ of all indicators is obtained.The objective trend weight of an index in the system is determined by means of weight set mean processing:

$$\bar{\omega}_i = \frac{\sqrt[q]{\prod_{t=1}^{q} r_t(C_i)}}{\sum_{j=1}^{m} \sqrt[q]{\prod_{t=1}^{q} r_t(C_j)}} \tag{15}$$

During the systematic analysis of the actual data, the index weight of the actual system can be calculated. The index weight sequence of the actual system is set as $\gamma = \{p_1, p_2, p_3, \dots, p_n\}$, represents the weight occupied by n indexes in the system,where $\sum_1^n \omega_i = 1$. According to the index weight γ sequence calculated by the actual data system and compared with the expert evaluation sequence $\bar{\omega}$, we can know the direction that the actual system should focus on when optimizing [10].

In addition, according to the weight of perfect expert evaluation and the weight of actual system index, the system maturity is calculated as follows:

$$Z = \theta \sum_{i=1}^{m} \bar{\omega}_i p_i \; (i = 1,2,\cdots,n) \tag{16}$$

Where θ is the constant term, and the maturity is mapped to the range of 0-1 for easy analysis. By evaluating the maturity value, it indicates that the current system is at high maturity, excellent maturity, good maturity, passing maturity and failing maturity. The specific classification is shown in table 2.

Table 2. System maturity division

Output list	Center(x,y)
0.9 ≤ Z≤1	High maturity
0.8 ≤ Z≤0.9	Excellent maturity
0.7 ≤ Z≤0.8	Good maturity
0.5 ≤ Z≤0.7	Passing maturity
0.0 ≤ Z≤0.5	Failure maturity

4. Experimental Analysis and Simulation

4.1. System Index Analysis and Establishment

Through reading a lot of literature and collecting and analyzing actual data, three first-level indicators are determined, namely, personnel, technology and process, and 12 second-level indicators are divided according to the three first-level indicators.

4.2. Expert Evaluation Matrix and System Maturity Calculation

According to the indicators that need to be evaluated, 10 experts are assigned to evaluate the indicators. Through matlab simulation, the evaluation value of experts on each indicator is calculated.

Among them, the personnel, technology and data processing process of level I indicators account for 0.3115, 0.3603 and 0.3282 respectively Among the secondary indicators of personnel, the proportion of ability level is 0.0775, the proportion of work experience is 00856, the proportion of technical mastery is 0.0394, and the proportion of personnel salary is 0.1090 Among the secondary technical indicators, the proportion of investment funds is 0.1167, the number of technical personnel is 0.0830, the proportion of technology construction investment transformation is 0.0995, and the proportion of technology product planning is 0.0611 Among the secondary indicators of data processing process, the proportion of invested funds is 0.0531, the proportion of metadata processing capacity is 0.0868, the proportion of concurrent sharing of data processing is 0.0722, and the proportion of healthy data processing process is 0.1161.

According to the simulated value, the weight value of the actual data system is simulated, and the system maturity is 0.8399 calculated according to Equation (16), which is classified as excellent maturity level according to table 2.

5. Conclusion

Aiming at the problems related to enterprise data analysis, this paper proposes a data analysis system based on index weight analysis method.The method proposed in this paper can analyze the maturity of data system and effectively solve the problem of weight determination and optimization in actual data system. According to the maturity index, the superiority degree of the current data system can be determined, and the objective trend of the future decision-making and system development can be certain.

Acknowledgment

University level project of Jishou University(Jdx20014); Innovation training program for college students of Jishou University(JDCX2021549)

References

[1] Cheng Xueqi, LIU Shenghua, ZHANG Ruqing. Big data analysis and processing technology of the new system thinking [J]. Proceedings of the Chinese Academy of Sciences, 2022 ((01) : 60-67. The DOI: 10.16418 / j.i SSN. 1000-3045.20211117005.
[2] Chen Jianming, Sun Xiaojun, Yang Bodi. Research on the Evaluation Framework and Implementation Path of Data Analysis literacy [J]. Journal of Mathematics Education,202,31(02):8-12+57.
[3] Pan Ziyu, ZHOU Shaowei. Comprehensive evaluation of urban tourism development potential: A case study of Shandong Province [J]. Mathematical Modeling and Applications, 201,10(04):19-29.DOI:10.19943/ J.2095-3070.mmMIa.2021.04.03.
[4] Wang Meng. Buckley J j. Fuzzy Hierarchical Analysis: Lamb damax Method [J]. Fuzzy Sets and Systems,2001,120(2).
[5] Colominas M A, Schlotthauer G, Torres M E. Improved Complete En- Semble EMD: A Suitable Tool for Biomedical Signal Processing[J]. Biomedical Signal Processing and Control, 2014, 14(1).
[6] MUNIER N,HONTORIA E.Uses and Limitations of the AHP Method[M]. Berlin:Speinger International Publishing,2021.
[7] Shi Ji, Suo Zhongying. Based on interval analytic hierarchy process (ahp) method to determine the loss function of [J/OL]. Journal of shandong university (science edition) : 1-10 [2022-05-09]. http://kns.cnki.net/kcms/detail/37.1389.N.20220419.1140.002.html.
[8] LI Meng. Rolling Bearing Fault Diagnosis Based on Improved Ensemble Empirical Mode Decomposition [J]. Journal of Changchun University, 201,31(06):1-5+23.
[9] Chaudhry J, Chaudhry J, Chaudhry J, et al. An evolutionary framework for generating linear inequalities and Multiplicative Fuzzy Priorities[J] Computing, 2015, 7(8).
[10] Wang Longyang, MENG Xi, Qiao Junfei. Prediction of Total phosphorus in Effluent based on Improved Ensemble Empirical Mode Decomposition and Deep Belief Network [J]. Journal of Chemical Industry and Engineering, 201,72(05):2745-2753.