

Research on evaluation model based on D & A system

Jingwen Liang^{*,#}, Xiner Huang[#], Huibing Wu[#]

School of Economics and Finance, Zhanjiang University of Science and Technology, Zhanjiang,
Guangdong, 524084 China

*Corresponding author: 1165765983@qq.com

[#]These authors contributed equally.

Abstract. The evaluation system established by our team for the D&A system hopes to help ICM companies fully utilize the data of their own output to understand their strengths and weaknesses, and help ICM supply words to obtain higher benefits. We introduced the hierarchical analysis method (AHP) to calculate the weights of the nine tertiary indicators, among which the human capacity, technical solutions and data management had the largest weights, 0.5584, 0.5936 and 0.625. Next, we used the entropy weighting method (EWM) to calculate the weights of the secondary and tertiary indicators respectively. Subsequently, we used the genetic algorithm (GA) and combined with the conclusions drawn earlier to make the selection of the optimal solution, and the optimization was calculated to obtain the maximum value of 0.432.

Keywords: Hierarchical analysis, genetic algorithm (GA), gray prediction model (GM(1,1)), evaluation model..

1. Introduction

In order to measure the maturity of the D&A system and provide a reliable plan to optimize the D&A capabilities of Intercontinental Cargo (ICM) companies, we need to find indicators that measure the current level of maturity of ICM's D&A system, by selecting appropriate evaluation indicators, assigning target weights to them, and combining these indicator weights to achieve a measure of D&A system maturity level. Subsequently, we optimize the D&A system and apply the developed model to various projects.

2. ICM D&A system maturity assessment

The selection of indicators is the cornerstone of the evaluation system. In order to make the indicator system scientific and standardized, the selection of indicators should follow the principles of systematicity, typicality, dynamism, concise scientific principle, concise scientific principle and comprehensive principle. A good indicator system can monitor business changes, and when there are problems in the business, the problem can be traced back through the indicator system, which can accurately locate the problem and give feedback to the business to solve the corresponding problems.

2.1 Indicator Description

- 1) personal
 - a) Personnel Capability. b) D&A Talent Training. c) Number of people employed
- 2) Technology
 - a) Software solutions. b) Technical solution. c) Product type and attributes.
- 3) Process
 - a) Data Processing. b) Data Aggregation. c) Data Management.

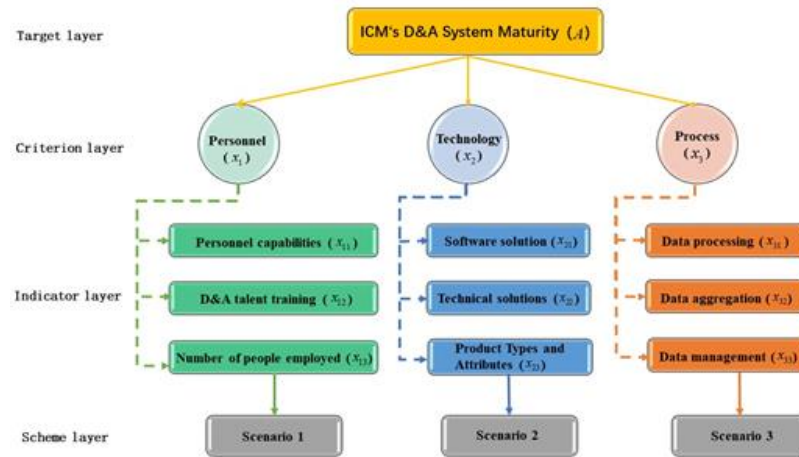


Figure 1. D&A system maturity AHP indicator structure

2.2 Analytic Hierarchy Process (AHP)

The AHP hierarchical analysis method is a research method that combines qualitative and quantitative approaches to calculate decision weights for solving multi-objective complex problems. The method combines quantitative analysis with qualitative analysis, uses the experience of decision makers to judge the relative importance among the criteria for measuring the achievability of each objective, and reasonably gives the weights of each criterion for each decision option, and uses the weights to find out the order of advantages and disadvantages of each option, which is more effectively applied to those topics that are difficult to be solved by quantitative methods.

Step1: Establish a hierarchical progressive structure, divided into target layer, criterion layer, indicator layer and program layer, with the maturity of ICM's D&A system as the target layer, the primary influencing factors as the criterion layer, the secondary influencing factors as the indicator layer, and the evaluation of them as the program layer.

Step2: Construct the judgment matrix, which characterizes the relative importance of the elements in this layer and is usually expressed by 1, 2, ..., 5 and the reciprocal. For the judgment matrix of the criterion layer and the judgment matrix of the indicator layer, because of the lack of basic data, the relative importance of each element is judged empirically and the value of each element of the judgment matrix is assigned.

Step3: Calculation of feature vectors, feature roots and weights.

① Eigenvector calculation: Calculate the subsquare root of the product of the elements of each row of the judgment matrix.

$$M_i = \sqrt[n]{\prod_{j=1}^n a_{ij}} \quad (1)$$

② Weighting calculation: normalization will be performed.

$$w_i = \frac{M_i}{\sum_{i=1}^n M_i} \quad (2)$$

③ Calculate the maximum characteristic root of the judgment matrix: this value will be used in combination with the CI value for consistency test use.

$$\lambda = \sum_{i=1}^n \frac{(Aw)_i}{nw_i} \quad (3)$$

Step4: consistency test analysis, in the construction of the judgment matrix, there may be logical errors, such as A is more important than B, B is more important than C, but then there is C is more important than A. Therefore, you need to use the consistency test to check whether there is a problem, consistency test using CR value for analysis, CR value is less 0.1 than the consistency test, and vice

versa, it does not pass the consistency test. If the data does not pass the consistency test, it is necessary to check whether there are logical problems, etc., and re-enter the judgment matrix for analysis.

$$CI = \frac{(\lambda - n)}{(n - 1)} \quad (4)$$

$$CR = \frac{CI}{RI} \quad (5)$$

Step5: Analysis of the conclusion, for the weights have been calculated, and the judgment matrix to meet the consistency test, then you can draw conclusions to continue further analysis.

2.3 Weighting Analysis

Table 1 Weights of the indicator layer

Criterion layer	Indicator layer	Feature vector	Weights	largest characteristic root	CI value
personnel	Personnel capability	2	0.5584		
	D&A talent training	1.1447	0.3196	3.0183	0.0091
	number of employees	0.4368	0.122		
Technology	software solution	0.5503	0.1571		
	technical solutions	2.0801	0.5936	3.0536	0.0268
	Types and attributes of products	0.8736	0.2493		
process	data processing	0.5	0.1365		
	data aggregation	0.8736	0.2385	3.0183	0.0091
	data management	2.2894	0.625		

As shown in the table1, in terms of personnel, the indicator "personnel capacity" has the highest weight, "D&A personnel training" ranks second, and "number of hired personnel" ranks last. Lastly, the number 0.3196 of employees is the 0.5584 most important 0.122. ICM, as a company near the port, needs to download, store and load containers, which requires good physical ability and perseverance of employees. In the process of cargo transportation, the time of arrival and departure, the collation of cargo list and other data requires strong calculation ability of employees to ensure the least time spent to maximize the benefits.

In terms of technical analysis, "technical solutions" has the highest weight, "product type and attributes" is the second, and "software solutions" is the third, respectively 0.1571, 0.5936, 0.2493 so technical Solutions are very important in terms of technology. A specific tool is a product that has been developed and can run in most cases, but it may not be suitable for the problem that ICM wants to solve at the moment, so the company needs technically innovative people who can research and develop a solution that is suitable for the operation of the company through their own technology, so that they can better promote the company's development and surpass the competitive companies in the same industry. The type and attributes of the product is also important, to develop the company's special products and make this product visibility is not an easy task, but once developed successfully will certainly become the best among peers.

In terms of process analysis, "data management" has the highest weight, "data aggregation" comes in second, and "data processing" comes in third, with 0.625, 0.2385 and 0.1365 respectively, so data management is most important at this level of the process. ICM's Information Security Officer (ISO) needs to track data resources by calculating and determining the large amount of data generated by cargo transportation, such as ship arrival/departure times, cargo manifests, shipping container lists, customs inspection reports, container locations in ports, inland transportation arrival/departure lists and schedules, etc., to provide metadata consistency across the organization. Consistency, which allows for precise scheduling of employee work, reducing time costs and maximizing the benefits for both.

3. The establishment of simulation model

3.1 Standardization of indicators

With the evaluation indicators defined above, we standardize the indicators to eliminate the effect when different attributes of the samples have different magnitudes. Based on the types of attributes of the original indicators, we use the standard 0-1 transformation and the given optimal interval method to do dimensionless and normalization. We set the primary indicators (criterion level) to be x_1, x_2, x_3 , and the secondary indicators (indicator level) to be $x_{11}, x_{12}, x_{13}, \dots, x_{33}$, respectively, so that we have:

$$y_{ij} = \frac{x_{ij} - \min(x_i)}{\max(x_i) - \min(x_i)} \quad j = 1, 2, \dots, n \quad (6)$$

Where y_{ij} is the standardized value of each evaluation index, $\max(x_i)$ and $\min(x_i)$ are the maximum and minimum values of the evaluation index.

After the data is standardized so that we can use y_{ij} rather than x_{ij} to express the maturity of ICM's D&A system, then we have:

$$q_{ij} = \frac{y_{ij}}{\sum_{j=1}^n y_{ij}} \quad (7)$$

Based on the statistically collected data and the concept of entropy, the information entropy e_i of each evaluation index is directly calculated, and then

$$e_i = -\ln(n)^{-1} \sum_{j=1}^n q_{ij} \ln(q_{ij}) \quad (8)$$

Based on the information entropy, we will further calculate the weights of each evaluation index that I defined before.

$$w_i = \frac{1 - e_i}{k - \sum_i e_i} \quad i = 1, 2, \dots, k \quad (9)$$

The results of the indicator weights are shown in Fig 5, 6, 7, 8.

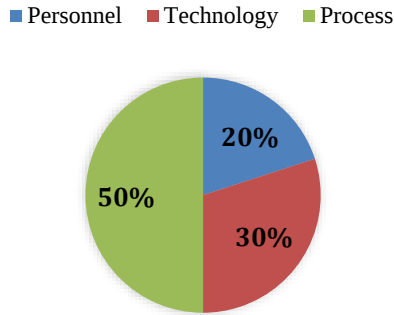


Figure 2 Secondary metric weights

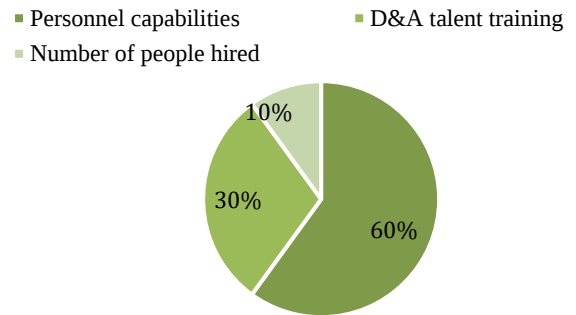


Figure 3 Three-level indicator weights(1)

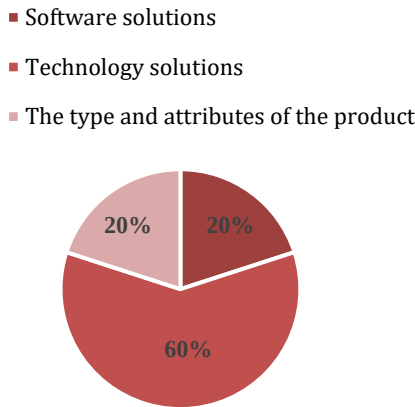


Figure 4 Three-level indicator weights(2)

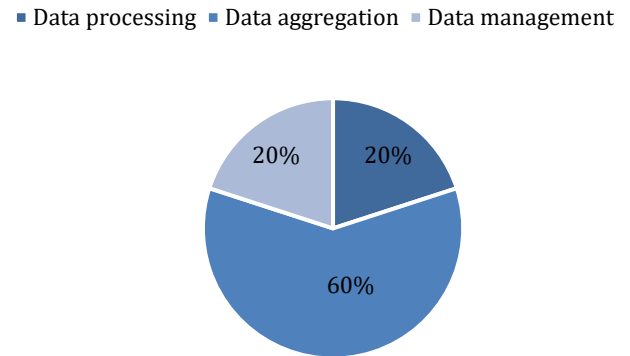


Figure 5 Three-level indicator weights(3)

By organizing the data, we consider the weights as coefficients of the primary and secondary indicators, set y as unknown values, to obtain the calculation equation.

$$y_{\max} = 0.2x_1 + 0.3x_2 + 0.5x_3 \quad (10)$$

$$\begin{cases} x_1 = 0.6x_{11} + 0.3x_{12} + 0.1x_{13} \\ x_2 = 0.2x_{21} + 0.6x_{22} + 0.2x_{23} \\ x_3 = 0.2x_{31} + 0.6x_{32} + 0.2x_{33} \end{cases} \quad (11)$$

$$y_{\max} = 0.2(0.6x_{11} + 0.3x_{12} + 0.1x_{13}) + 0.3(0.2x_{21} + 0.6x_{22} + 0.2x_{23}) + 0.5(0.2x_{31} + 0.6x_{32} + 0.2x_{33}) \quad (12)$$

Considering that, the energy of each company's output is finite, so a layer of constraint is set to x_1 、 x_2 、 x_3 , we set the company's energy to be a , a as a constant, which yields the computational equation.

$$x_1 + x_2 + x_3 = 1 \quad (13)$$

$$0.6x_{11} + 0.3x_{12} + 0.1x_{13} + 0.2x_{21} + 0.6x_{22} + 0.2x_{23} + 0.2x_{31} + 0.6x_{32} + 0.2x_{33} = 1 \quad (14)$$

In addition to this, we need to set a range of values for it, as follows:

$$0 \leq x_1, x_2, x_3 \leq 1 \quad (15)$$

$$0 \leq x_{11}, x_{12}, \dots, x_{33} \leq 1 \quad (16)$$

Finally can get $y_{\max} = 0.423$.

3.2 Genetic Algorithm

Three-level indicator code In the case of tertiary metrics, the choice of the right coding method can affect the next operations. Therefore, when solving this problem of tertiary metrics programming, we prefer to find a simple coding method that does not affect the performance of the algorithm. When using genetic algorithms to solve specific problems, we should try to analyze the characteristics of the problem and develop a practical programming solution.

Initialization of the group In terms of population initialization, the genetic algorithm performs an iterative search in a given initial evolutionary population. It is known that the algorithm can improve its ability to find the optimal solution if the initial population is already a good population to some extent at the beginning of the algorithm. We improve the solution performance of the algorithm by using AHP hierarchical analysis to obtain an initial population with a relatively high average fitness value before optimizing it.

Adaptation value evaluation In the evaluation of the adaptation value, since it is specified in the genetic algorithm that the larger the adaptation value, the better the indicator, so for some numerical optimization problems to solve the maximum value, we can directly apply the function expressions defined in the problem.

Group selection In terms of group selection, it can be done by using roulette selection algorithm, whose basic idea is random selection based on probability, but because of the randomness of the selection process, it is not guaranteed that each selection will select these better indicators, so it also gives some room for the survival of the worse indicators.

Mixing of group indicators In the indicator mixing stage, whether each indicator can be mixed is determined by the mixing probability R (generally takes a value of 0.4-0.99 between), the specific process is: for each indicator, if $\text{Random}(0,1)$ is less than P_c then the indicator can be mixed operation, where $\text{Random}(0,1)$ is a random number generator with uniform distribution between $[0,1]$, otherwise the indicator does not participate in the mixing directly replicated to the new population.

Population variation In terms of population variation, we should set the variation probability P_m in an appropriate range in order to maintain a better operational performance of the genetic algorithm. The mutation operation has a significant contribution in improving the population diversity by changing the original metrics. If P is too small, the algorithm tends to mature prematurely. However, during the operation of the algorithm, the better solution found may be destroyed during the mutation process, and if the value of P is too large, it may cause the better search state that the algorithm is currently in to regress back to the original worse situation. Therefore, we should limit the variation of the population to a certain range. In general, P can be set between 0.001- 0.1.

The above describes each important step of the genetic algorithm in detail, and then we give the basic steps of the genetic algorithm.

Step1: Initialize the population of size N , where the values of the indicators are generated using a random number generator and satisfy the range defined by the problem. The current evolutionary Generation = 0.

Step2: The evaluation function is used to evaluate all three-level indicators in the population, and the adaptation value of each three-level indicator is calculated separately, and the indicator with the largest adaptation value, Best, is saved.

Step3: The roulette selection algorithm is used to perform the selection operation on the tertiary indicators of the population to produce a population of the same size N .

Step4: Select indicators from the population for mixing according to the probability P_c . The new indicators replace the old ones into the new population, and the indicators that have not changed go directly into the new population.

Step5: The variation operation is performed on the indicators in the new population according to the probability P_m . The values that have changed are changed. The metrics that generate changes replace the original metrics into the new population. The indicators that have not changed directly enter the new population.

Step6: Generate a new population to replace the original population, and recalculate the adaptation value of each indicator in the population. If the maximum adaptation value of the population is greater than the adaptation value of Best, the indicator corresponding to the maximum adaptation value is used to replace Best.

Step7: The current evolutionary generation is added one. If Generation exceeds the specified maximum number of evolutionary generations or Best reaches the specified error requirement, the algorithm ends; otherwise, the algorithm returns to Step3.

3.3 GM(1,1)Gray prediction model

Gray system theory is a relatively important method for studying discrete data series with small sample sizes and incomplete information. By fully exploiting and utilizing the explicit and implicit information in the available data, the randomness present in the series is cumulatively attenuated. As a result, patterns of control system variation are generated and can be used to study future time distributions at specific time intervals.

Collect the data of container throughput of Lianyungang and New York ports from 2015-2021, and do cumulative processing on the original data to get the new data.

$$x^{(1)}(M) = \sum_{i=1}^M x^{(0)}(t) \quad (17)$$

The trend of the new series is approximated by a differential equation to describe the change.

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = u \quad (18)$$

Find the prediction model.

$$x^{(1)}(t+1) = [x^{(0)}(1) - \frac{u}{a}]e^{-at} + \frac{u}{a} \quad (19)$$

4. Analysis of results

We assume that the growth rate of container throughput of New York Port in the United States is 5 % per year after 2021year, and the growth rate of container throughput of Lianyungang Port in China is 3% per year after 2021year, so the target value of container throughput of New York Port in 2028 year is 11537654 and the target value of container throughput of Lianyungang Port in 2028year is 6407643. The forecast results are shown in the following chart.

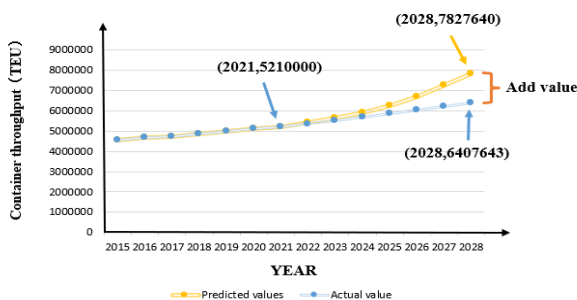


Figure 6. New York Harbor Forecast Chart

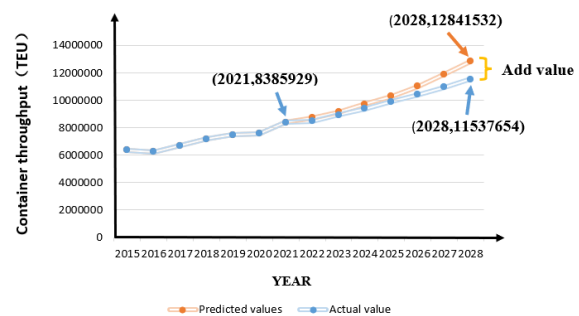


Figure 7. Lianyungang Forecast Chart

From the two graphs above, we can see that both the Port of New York, USA and Lianyungang Port, China are higher than the target value in 2028 annual container throughput with 12,841,532 and 7,827,640 respectively, so it can be a good proof that our model is suitable for these two seaports.

5. Conclusions

The optimized D&A system not only analyzes the three important parts of people, technology and process, but also has a strict self-checking procedure to ensure the scientific and reliable operation of

the D&A system. ICM attaches great importance to its own cost effectiveness, and improving the efficiency of resource utilization is a top priority for the development of the company, whether it is port transportation or port transshipment, which is centered around cargo is carried out. People, technology and processes are all integral parts of this process.

The genetic algorithm, a well-established computational method for solving optimization problems, is used, and its ability to adapt to global search on-the-fly is important for improving the computational efficiency of our system. We focus on the interplay between people, technology and process, not just on analysis, but on how they work together to achieve greater effectiveness. Our system is practical and does not require significant training costs to enable company managers to become proficient in its use.

The model we built effectively solves the required four tasks and is innovative. Our innovative model is highly practical by considering many factors and then conducting a scoring test. We selected the Port of New York in the U.S. and Lianyungang in China, both of which are influential in the world, and substituted the model to analyze and forecast their container throughput from 2022 to 2028. In short, the model we built is innovative and practical, helping us to solve all the necessary tasks.

References

- [1] An Jingwen, Li Yuanchun, Liu Haidong, Zhen Haihong, Liang Ruiwen. Research on the evaluation index system of enterprise technology innovation capability maturity [Research on the evaluation index system of enterprise technology innovation capability maturity.doc-originality-document (book118.com)]. 2017.
- [2] Cao Siming, Wu Yi, Cao Kai, Chen Meng. Evaluation of comprehensive energy system operation service based on entropy weight and AHP[J]. Science and Technology Bulletin, 2021, 37(12): 56-60.
- [3] Chen Guanlin. The GDP forecast analysis of the three northeastern provinces based on the grey model [J]. China Business Review, 2022, (03): 10-13.
- [4] Holland JH. Genetic algorithms. Scientific American, 1992: 44-50.
- [5] Liu Danyu. Exploring the construction of a training management system for front-line staff in ports [J]. Hebei Enterprise, 2019, (02): 125-126.
- [6] Qinghai Development Strategy Department. Research and application of process maturity assessment model of Qinghai Mobile [Process maturity assessment index system - China Mobile.docx (book118.com)], 2019.
- [7] Sun Baohua, Li Yunpeng, Yu Yuanqi, Tong Donghui, Cao Zhi. Maturity assessment of power information system based on big data and analytic hierarchy process [J]. Science and Technology Wind, 2019, (13): 78.
- [8] Tao Hongfei, Sun Yixin, Wu Guowei, Li Kangyi. Maturity assessment of power information system based on big data and analytic hierarchy process [J]. China Electric Power, 2016, 49(10): 114-118.