Research on value evaluation of basic scientific data based on combination weighting Method

Yajing Guo [a*], Yang Wang [b], Chuanwen Wu [c], Huixin Liu [d], Lingling Geng [e]

Aerospace Science & Industry Defense Technology Research and Test Center, Beijing, 100000, China

[a*] 727693992@qq.com,[b]181572540@qq.com
[c]chuanwenwu_bj@163.com,[d]huixinlau@163.com
[e]gll063433@126.com

Abstract. Basic scientific data is an important foundation for scientific research, and establishing an effective data value evaluation system is an inevitable requirement for improving the output level and influence of scientific research results. This article proposes a method for evaluating the value of basic science data based on combination weighting, which is applicable to basic science data. A data value evaluation indicator system is established from three aspects: the cost value, intrinsic value, and utility value of the data. The relative weight values of each evaluation indicator are calculated using the combination weighting method, and the value level of the data is divided by a comprehensive evaluation function. The case analysis shows that this method can accurately evaluate the value level of basic scientific data.

Keywords: Basic scientific data, Data value assessment, Combination weighting

1 Introduction

The concept of data as an asset has become an industry consensus in various fields. However, in reality, the management and application of data assets are often still in the exploratory stage, and data asset management faces many challenges [1-3]. Basic scientific data is an important data asset in the field of scientific research, playing a crucial role in the sustainable development of scientific research. It can be said that to a certain extent, it determines the quality of scientific research. Therefore, the evaluation research of basic scientific data is of great significance for scientific research and data management.

Basic scientific data consists of internal data generated by scientific research activities themselves and external scientific data that can support scientific research activities, including raw observation data, detection data, experimental data, survey data, investigation data, remote sensing data, statistical data, etc. It is an important strategic resource for scientific and technological innovation. The value assessment of basic scientific data can not only provide reference for data users and clarify the focus of data

H. Dai et al. (eds.), *Proceedings of the 2025 5th International Conference on Business Administration and Data Science (BADS 2025)*, Advances in Computer Science Research 125,
https://doi.org/10.2991/978-94-6463-980-3_27

construction; Moreover, by optimizing the priority results of basic data resources and optimizing the way data services are applied, the application value of data can be maximized.

The existing research on data value assessment usually adopts expert scoring method for evaluation, which is greatly influenced by subjective factors [4-5] and is not suitable for data value assessment based on the characteristics of basic scientific data. This article adopts a basic science data value evaluation method based on combination weighting to evaluate the value of basic science data. By establishing a complete data evaluation indicator system, the relative weights of the indicators are determined, and the comprehensive evaluation function is used to obtain the comprehensive score of data value. Finally, the feasibility of using this method for basic science data value evaluation is demonstrated through case analysis.

2 Indicator system for evaluating the value of basic scientific data

Through brainstorming, 19 key factors affecting the value of basic scientific data were identified from three aspects: cost value, intrinsic value, and utility value.

2.1 Evaluation Dimension

This article starts from the aspects of basic scientific data construction and application needs, data flow process, etc., comprehensively considers multiple factors, and regards data cost value, data intrinsic value, and data utility value as the main dimensions of data value.

(1) Data cost value (V1). The cost value of data refers to the total cost investment in the entire process of acquiring, collecting, storing, processing, and analyzing data, including four basic elements: data acquisition cost, (V11) data storage cost(V12), data processing cost(V13), and data operation and maintenance cost(V14). The cost of data acquisition refers to the cost of obtaining a single piece of required data. Data storage cost refers to the cost invested in storing the acquired and processed data. Data processing cost refers to the cost invested in processing the required data to extract accurate information. Data operation and maintenance costs refer to the maintenance costs incurred to ensure the normal operation of stored data and storage devices.

(2) The intrinsic value of data(V2). The intrinsic value of data refers to the fundamental contribution ability of data to information and knowledge expression due to its inherent attributes, including six basic elements: data scale(V21), data scarcity(V22), value density(V23), data timeliness(V24), data correlation(V25), and data universality(V26). Data scale refers to the size range or total amount of data; Data scarcity refers to the quantity of data supply and the number of providers. Value density refers to the proportion of high-value data in the overall data. Data correlation refers to the degree of association between data assets and other data assets. Timeliness of data refers to the degree to which data updates meet business requirements. Data universality refers to the reusability of data and knowledge.

(3) The utility value of data(V3). The utility value of data refers to the actual benefits and value that data, after processing and analysis, brings to users in practical applications, including application scenario utility (V31) (subdivided indicators include four basic elements: data business efficiency(V311), data technology benefits(V312), data redevelopment capabilities(V313), data economic benefits(V314)), and satisfaction utility of application(V32) (subdivided indicators include five basic elements: the degree to which data items meet usage needs(V321), the degree to which data volume meets usage needs(V322), the degree to which data queries are met(V323), the degree to which data downloads are met(V324), and the degree to which data access time is met(V325)). The utility of data application scenarios refers to the manpower and costs saved, as well as the benefits obtained by users in using data to solve business problems in technological activities. Data application satisfaction refers to the degree to which data meets user needs.

2.2 Evaluation indicator system

Based on three main evaluation dimensions and principles of systematicity, objectivity, and scientificity. Ten experts were selected to evaluate and screen 19 influencing factors through two rounds of Likert expert surveys. Three secondary elements were removed to obtain the basic scientific data value evaluation indicator system, which includes 16 last-level influencing factors (includes V11-V13, V21-V24, V311-V314, V321-V325), part of it is shown in Figure 1.

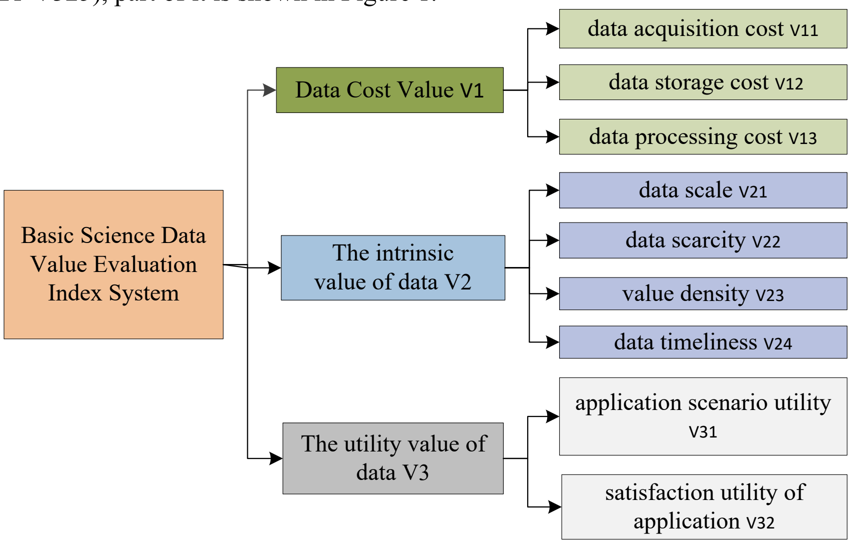

Fig. 1. Basic Science Data Value Evaluation indicator System.

3 Combination weighting method

3.1 AHP method for determining weights

The Analytic Hierarchy Process (AHP) is a systematic and hierarchical evaluation method [6] that combines qualitative and quantitative methods. It can quantify qualitative problems that rely on subjective experience judgment and is suitable for situations where the target structure is complex and data is scarce.

Generally speaking, using the Analytic Hierarchy Process to determine the weights of data value evaluation indicators involves three basic steps:

Step 1: Hierarchical complex problems to form a hierarchical structure diagram consisting of objective layer, criterion layer, indicator layer, and solution layer, where the elements in the indicator layer belong to one or more elements in the criterion layer.

Step 2: use the "1-9 scaling method" to form a judgment matrix and assign values. The specific method is as follows.

(1) Form a judgment matrix

Construct and fill in a judgment matrix based on a hierarchical structure, and conduct a questionnaire survey among experts according to the importance of basic scientific data value indicators. The relevant elements in the hierarchical structure are compared pairwise to indicate their importance, and corresponding values are assigned using scales 1-9. The significance scale of evaluation indicators is shown in Table 1.

Table 1. Meaning of Importance Scale for Quality Evaluation Indicators of Measurement Data.

Importance Scale	meaning
1	Indicates that two elements have equal importance compared to each other
3	Compared to two elements, the former is slightly more important than the latter
5	Compared to two elements, the former is significantly more important than the latter
7	Compared to two elements, the former is more strongly important than the latter
9	Compared to two elements, the former is extremely more important than the latter
2,4,6,8	The median value representing the above judgment
countdown	If the ratio of the importance of element i to element j is a_{ij}, then element j The ratio of importance to element i is a_{ji} $1/a_{ij}$

Let the completed judgment matrix be $A=(a_{ij})_{n\times n}$,($a_{ij} > 0; a_{ij} = \frac{1}{a_{ij}}; a_{ij} = 1$)

$$A=\begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix} \tag{1}$$

(2) Eigenvector calculation method

① Calculate the nth root of the product of each row element in the judgment matrix:

$$\bar{a_i}=\sqrt[n]{\prod_{j=1}^{n} a_{ij}} \tag{2}$$

② Normalize processing to obtain the weight of each indicator:

$$W_{AHP\cdot i}=\frac{\bar{a_i}}{a}=\frac{\bar{a_i}}{\sum_{i=1}^{n}\bar{a_i}} \tag{3}$$

Then it $W_{AHP\cdot i}=(w1,w2,\cdots,wn)^T$ is the desired eigenvector.
③ Calculate the maximum eigenvalue of judgment matrix A λ_{max}:

$$\lambda_{max}=\frac{1}{n}\sum_{i=1}^{n}\frac{(AW)_i}{w_i} \tag{4}$$

Among them, A is the judgment matrix, $(AW)_i$ representing the AWi-th element of the vector.

Step three: verify the consistency of the judgment matrix.

Consistency check CR:

$$CR=\frac{\lambda_{max}\text{-}n}{RI(n\text{-}1)} \tag{5}$$

If CR<0.1, A passes the consistency test; If it fails, the original judgment matrix must be corrected until the consistency test is passed, at which point the approximate solution of the eigenvector is the weight coefficient of each element.

3.2 CRITIC method for determining weights

CRIteria Through Intercriteria Correlation (CRITIC) is an objective weighting method[7] proposed by DIAKOULAKI et al. in 1995. This method does not require the independence of attributes and transforms qualitative attributes into quantitative attributes. Its core idea is to use the contrast and conflict between indicators to determine the weight of each indicator. The contrast is based on the standard deviation, and the larger the standard deviation, the more significant the data fluctuation and the higher the weight; The conflict is based on the correlation coefficient, and the larger the correlation coefficient, the smaller the conflict between indicators, and the lower the weight will naturally be.

CRITIC analysis requires first standardizing the obtained raw data, calculating the information carrying capacity of each indicator, further calculating objective weights, and finally normalizing the obtained objective weights.

(1) Standardization of raw data. The CRITIC method assumes that there is one set of raw data to be evaluated, which includes *m* objects and *n* evaluation indicators, collectively forming the raw data matrix X:

$$X=\begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix} \tag{6}$$

In the formula: x_{ij} is the element in the *i*-th row and *j-th* column of the matrix.

Among them, the dimensions of each indicator are different, and the weight relationship between indicators cannot be directly determined. Therefore, data standardization is needed to remove the influence of dimensions. Separate treatment of positive and negative indicators, with positive treatment resulting in higher values for better outcomes; Reverse processing, the smaller the value, the better.

Positive indicators:

$$x'_{ij}=\frac{x_{ij}-\min(x_j)}{\max(x_j)-\min(x_j)} \tag{7}$$

Reverse indicator:

$$x'_{ij}=\frac{\max(x_j)-x_{ij}}{\max(x_j)-\min(x_j)} \tag{8}$$

In the formula: $max(x_j)$is the maximum value in the jth column element; The minimum value in the *jth* column element.

(2) Calculate the information carrying capacity. The information carrying capacity $A_j$refers to comparability and conflict, respectively represented by standard deviation $S_j$and correlation coefficient.

1) Comparison. Comparison is represented by the standard deviationS_j of each indicator (column).

$$S_j=\sqrt{\frac{\sum_{j=1}^{n}(x_{ij}-\overline{x_j})^2}{n-1}} \tag{9}$$

In the formula: $\overline{x_j}$is the average value of each indicator (column).

2) Conflict oriented. Conflict is $A_j$represented by the correlation coefficient of each indicator (column), and a correlation matrix needs to be constructed first R.

$$R=\frac{\sum_{j,k=1}^{n}(x_{ij}-\overline{x_j})(x_{ik}-\overline{x_k})}{\sqrt{\sum_{j=1}^{n}(x_{ij}-\overline{x_j})^2\sum_{k=1}^{n}(x_{ik}-\overline{x_k})^2}} \tag{10}$$

According to the correlation matrix R, calculate the correlation coefficient using equation A_j:

$$A_j=\sum_{i=1}^{n}(1-r_{ij}) \tag{11}$$

In the formula: $r_{ij}$is an R element in the correlation matrix.

3) Information carrying capacity. Calculate information carrying capacityC_j:

$$C_j=S_jA_j \tag{12}$$

(3) Calculate objective weights. The objective weight is $W_j$calculated according to equation:

$$W_j=\frac{C_j}{\sum_{j=1}^{n}C_j} \tag{13}$$

Further $W_j$normalize the objective weights obtained to eliminate the mutual influence between various factors:

$$W_{CRITIC\cdot j} = \frac{W_j}{W_1} \tag{14}$$

In the formula: $W_{CRITIC\cdot j}$ is the normalized weight of the indicator; $W_{CRITIC\cdot 1}$ The weight of the first indicator.

3.3 Determination of weights using

AHP-CRITIC combination weighting method Based on the one sidedness of the AHP and CRITIC methods in analyzing the subjective and objective focus, in order to obtain the weight coefficients of the target indicators more scientifically and comprehensively, the AHP-CRITIC combined weighting method is used to analyze the weights of each indicator. The calculation formula is:

$$W_{AHP-CRITIC} = W_{AHP} \cdot W_{CRITIC} / \sum (W_{AHP} \cdot W_{CRITIC}) \tag{15}$$

4 Basic Science Data Value Evaluation Based on Combination Weighting Method

4.1 Evaluation Model

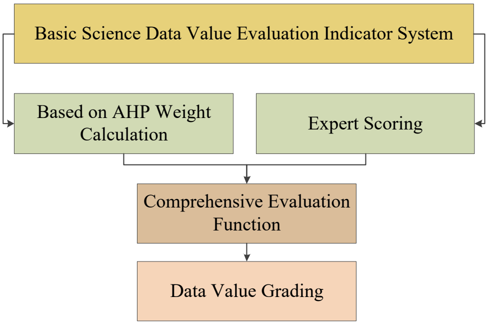

Fig. 2. Basic Science Data Value Evaluation Model.

Basic Science Data Value Evaluation Model is shown in the Figure 2.

4.2 Determining Indicator Weights

The calculation of the relative weights of basic scientific data value evaluation indicators based on the combination weighting method is divided into two steps:

Firstly, constructing a four layer hierarchical structure model, where the first layer is the target layer, which is the overall goal of data value evaluation; The second layer is the first level specific indicator layer (V1-V3); The third layer is the second level specific indicator layer (V11-V13, V21-V24, V31-V32); The third layer is the third level specific indicator layer (V311-V314, V321-V325).

Secondly, use the combination weighting method in Chapter 2 to calculate the weight values of each layer's indicators. the final weighting results are shown in the table 2.

Table 2. Indicator weight table.

Evaluation dimension	V11	V12	V13	V21	V22	V23	V24	V311
Weight(W)	0.102	0.048	0.068	0.08	0.081	0.151	0.13	0.042
Evaluation dimension	V312	V313	V314	V321	V322	V323	V324	V325
Weight(W)	0.021	0.037	0.021	0.055	0.05	0.031	0.043	0.04

4.3 Classification of Data Value Levels

By scoring the evaluation objects of the indicator system, the maximum score for a single indicator (V=100) is 100 points, and the total score can be determined by the weighted sum of the marked scores yields:

$$S = \sum V * W \tag{16}$$

Select the two items with the highest weight of secondary indicators from the indicator system as important ability evaluation items, namely: value density(V23), satisfaction with data application(V32). The data usage is determined by a comprehensive evaluation of the total score and important ability assessment items assessment of data value level in stages.

(1) Level 1: The total score is not less than 80, and the score for any important ability assessment item is not less than 80.

(2) Level 2: The total score is not less than 60, and the score for any important ability assessment item is not less than 60.

(3) Level 3: The total score is less than 60 or the score for any important ability assessment item is less than 60.

5 Case Analysis

5.1 Scoring indicators

This article takes the basic scientific data of a certain unit's electronic components as an example. The data content is the performance parameter data of the components, with a data volume of 1000 pieces. The key fields include product model, product name, category code, manufacturer, assessment standards and protocols, quality level, expected level, packaging shape, parameters, etc. Organize 5 experts to score 16 indicator according to the scoring rules, and the results are shown in the following figure 3 and table 3.

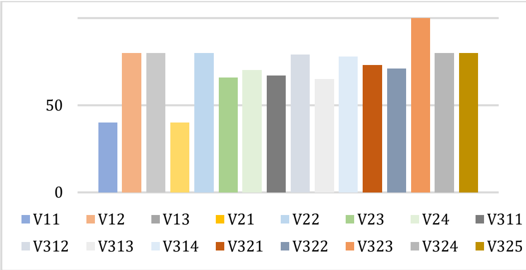

Fig. 3. Scoring result.

Table 3. Scoring result Table of basic scientific Data of electronic components.

Evaluation dimension	V11	V12	V13	V21	V22	V23	V24	V311
Average score result(V)	40	80	80	40	80	66	70	67
Evaluation dimension	V312	V313	V314	V321	V322	V323	V324	V325
Average score result(V)	79	65	78	73	71	100	80	80

5.2 Level evaluation

Calculate the total score according to Formula (16), that is $S = \sum_{i=1}^{16} V * W = 67.927$,so the total score is 67.927 points. And the V23's score is 66 points, V32's score is 79.056 points, both greater than 60 points. Therefore, the value evaluation level of this data resource is level 2.

The results indicate that the basic scientific data value of this component is moderate, and there is still significant room for improvement. It can be improved by focusing on indicators with high weight and low scores to enhance the data value.

6 Conclusion

This article proposes a method applicable to basic science The data value evaluation system based on data characteristics fills the gap in the basic scientific data value evaluation system. By calculating weights and the comprehensive evaluation score quantifies the value of data affected by complex factors, and finally demonstrates the feasibility of this method through examples and effectiveness. By evaluating the value of fundamental scientific data, data managers can prioritize delivering high-value data to researchers, providing precise data support for scientific work. It is noteworthy that this research has certain limitations, such as challenges in accurately grading and defining qualitative indicators during quantification. Further refinements to the methodology will be pursued through ongoing research.

References

1. Hannila, Hannu, et al. "Data-driven begins with DATA; potential of data assets." *Journal of Computer Information Systems* 62.1 (2022): 29-38.
2. Corrado, Carol, et al. "Measuring data as an asset: Framework, methods and preliminary estimates." *OECD Economics Department Working Papers* 1731 (2022).
3. Chen, Rongda, et al. "Navigating uncertainty: the impact of economic policy on corporate data asset allocation." *International Review of Economics & Finance* 97 (2025): 103783.
4. Brennan, Rob, et al. "Exploring data value assessment: a survey method and investigation of the perceived relative importance of data value dimensions." (2019).
5. Yanlin, Wang, and Zhao Haijun. "Data asset value assessment literature review and prospect." Journal of Physics: Conference Series. Vol. 1550. No. 3. IOP Publishing, 2020.
6. Tavana, Madjid, Mehdi Soltanifar, and Francisco J. Santos-Arteaga. "Analytical hierarchy process: Revolution and evolution." Annals of operations research 326.2 (2023): 879-907.
7. Silva, Natan Felipe, et al. "An integrated CRITIC and Grey Relational Analysis approach for investment portfolio selection." Decision analytics journal 8 (2023): 100285.