Multi Model Establishment of Enterprise Case Aided Decision-making Based on Data Analysis

[1a]Kunying Li , [2b]Xuwei Jiang, [1c]Xu He,[1d]Siwen Yin

[a*] Corresponding author: likunying@petrochina.com.cn, [b]johnson@petrochina.com.cn
[c]hexu@petrochina.com.cn, [d]yinsiwen@petrochina.com.cn

[1]Information Technology Center PetroChina RIPED Beijing, China
[2]Legal and Enterprise Reform Department PetroChina Beijing, China.

Abstract—Conventional multiple models for enterprise case aided decision-making mainly use AHP (Analytic Hierarchy Process) analytic hierarchy process to measure the relative weight of multiple factors, which is vulnerable to the influence of normalized processing relationship, resulting in low accuracy of model classification. Therefore, a new multiple model for enterprise case aided decision-making needs to be designed based on data analysis. That is, data preprocessing is carried out in advance, the original data of the model is obtained through data cleaning, feature selection, and feature coding steps, and then model parameters are selected. Using data analysis technology, an enterprise case aided decision-making multiple model is established. The experimental results show that the classification accuracy of the designed enterprise case aided decision-making multiple model based on data analysis is high, which proves that the established enterprise case aided decision-making multiple model has good performance, reliability, and certain application value, and has made certain contributions to improving the effect of enterprise case aided decision-making.

Keywords-Data Analysis; Enterprises; Cases; Auxiliary; Policy decision; Multi model; Establish

1 Introduction

In the context of the development of computer informatization, China has gradually entered the era of big data. The composition of data in various fields is becoming more and more complex, and the total amount of data is getting higher and higher[1-3]. Affected by the complex network environment, the difficulty of enterprise case decision-making is gradually increasing, which is not conducive to the comprehensive development of enterprises. Aided decision-making is a computer supported verification technology[4-6], which can make enterprise decisions quickly, stably and safely, and make the decision results meet different constraints. In the process of enterprise case aided decision making, the scheduling plan[7-10] can be implemented through computer simulation to correct the final decision result and support the decision-making scheme of the enterprise.

In fact, in the process of enterprise case aided decision making, it is first necessary to rely on the computer information system for basic decision retrieval to obtain data information that meets the decision requirements[11-13]. This part of data information is often complex in composition and different in format. If directly involved in the subsequent decision-making

ICEMBDA 2023, October 27-29, Tianjin, People's Republic of China

DOI 10.4108/eai.27-10-2023.2341917

steps, it will increase the difficulty of decision-making and generate decision-making errors. Therefore, data processing is required. Redundant data should be eliminated through data cleaning, and more targeted decision data should be selected for coding to participate in the subsequent decision-making process. The research shows that the complexity of different enterprise cases is inconsistent[14-17], and the difficulty of decision-making is also different. Therefore, in the process of enterprise case auxiliary decision-making, it is necessary to determine the final decision-making problem, select more targeted decision-making materials to participate in decision-making, and improve the accuracy of decision-making results. Therefore, enterprise case aided decision-making can provide basic decision support for difficult decision problems and implement decision content that meets the requirements of enterprise development.

Assisted decision making does not fully represent manual decision making, but it can eliminate many difficulties in human decision making and avoid unstable decision making. In the decision-making process[18-20], decision makers need to make corresponding choices in combination with their own values to achieve optimal decision making. Relevant researchers have designed several kinds of multiple models for enterprise case auxiliary decision making according to the requirements of auxiliary decision making, including the multiple models for enterprise case auxiliary decision making based on analytic hierarchy process, and the multiple models for enterprise case auxiliary decision making based on association rules. Influenced by the complex data composition of enterprise cases, the above models need to repeatedly measure the relative weight of multiple factors and are vulnerable to the influence of normalization processing relationship. As a result, the classification accuracy of the model is low and does not meet the current requirements of enterprise case decision-making. Therefore, this paper designs a new enterprise case aided decision-making multiple model based on data analysis.

2 Complete data preprocessing and feature engineering based on data analysis

2.1 Data cleaning

In order to ensure the data quality of enterprise case auxiliary decision information source, this paper first uses data cleaning technology to preprocess the data.In the process of data cleaning[21-23], noise data can be effectively eliminated so that the data in the data set can meet the requirements of enterprise case auxiliary decision-making. The schematic diagram of data cleaning is shown in Figure 1 below.

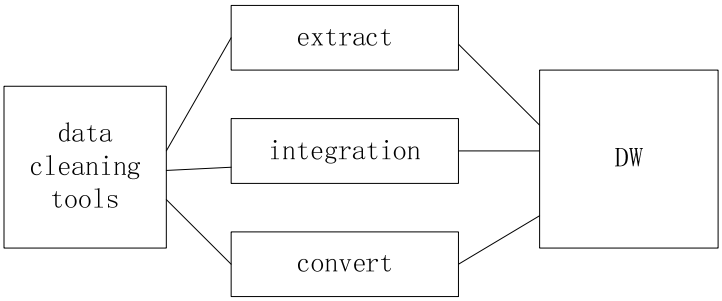

Figure 1. Data cleaning

It can be seen from Figure 1 that in the process of data cleaning, the collected data needs to be continuously integrated and converted. The basic steps mainly include: first, determine the source of enterprise case auxiliary decision data, conduct data sampling inspection, determine mapping rules, and then conduct noise data inspection based on the number of data sources, and perform data cleaning operations. Finally, data cleaning is carried out according to the data analysis structure to generate a decision data set matching the auxiliary decision model.

2.2 Feature selection

When the obtained enterprise case aided decision-making multi model data has a high dimension, many redundant genes will be generated, which is not conducive to subsequent decision classification. Therefore, data feature selection is required to effectively classify data information and eliminate some redundant interference information. That is, according to the decision requirements, data feature selection information[24] is generated, interference factors are obtained, and data characteristics are described. Next, basic transformation is carried out according to the obtained data feature set, and new data characteristics are combined to meet the data distribution relationship. The effectiveness of the data information obtained through this step is increased to meet the distribution relationship of auxiliary decision-making, which is of great significance for improving the performance of the model.

2.3 Feature coding

In order to improve the efficiency of enterprise case aided decision making, linear algebra calculation is needed, that is, feature coding processing is carried out on the data obtained above. The method designed in this paper mainly uses tag encoding technology to convert the data with decision characteristics from string to data, so as to realize the free definition and quantification of data and reduce the difficulty of data sorting. Data feature coding can effectively obtain decision logic, reduce the impact of insensitive data on enterprise case auxiliary decision making, improve the interpretability of data, and summarize the coded data in a unified original data set, which is the basis for subsequent model construction.

3 Establishment of multiple models for enterprise case aided decision-making based on data analysis

3.1 Selecting parameters

The accumulation of continuous auxiliary decision-making data will affect the final auxiliary decision-making effect. Therefore, according to the raw data processed above, this paper selects the parameters of enterprise case auxiliary decision-making multi model. The data analysis technology can summarize and centrally analyze the original data obtained above, mining the decision-making characteristics of the data to the greatest extent. Therefore, based on the data analysis technology, this paper selects the multiple model parameters[25] for enterprise case aided decision-making. First of all, according to the actual production situation of the enterprise, we can generate balanced enterprise case auxiliary decision parameters. On the basis of ensuring the original production of the enterprise, we can first balance the economy, complete the optimal enterprise decision configuration, and generate the optimal enterprise decision scheme. The decision parameters generated at this time $\min F$ as shown in (1) below.

$$\min F = (x_1, x_2, x_3) \tag{1}$$

In formula (1), x_1、x_2、x_3 represent the parallel balanced decision parameters. In order to achieve the balance of auxiliary decision quantity and evenly distribute the decision relationship, the auxiliary decision objective function can be calculated $f(x)$, as shown in (2) below.

$$f(x) = \frac{1}{N}\sqrt{\sum (h_m - h)^2} \tag{2}$$

In formula (2), N represents the decision cycle, h_m represents a collection of events involving decisions, h represents the decision parameters, and the decision reliability can be calculated according to the auxiliary decision objective function generated above C, as shown in (3) below.

$$C = \frac{\sum \min F}{G_j} \tag{3}$$

In formula (3), G_j represents the economic decision-making coefficient. If the calculated decision-making reliability meets the actual decision-making requirements, it proves that the above objective function meets the requirements of model construction. Otherwise, it needs to reset the decision-making indicators according to the decision-making objectives to ensure the final decision-making reliability.

3.2 Model establishment

Combining the above steps, we can obtain the model parameters that meet the decision-making requirements of the auxiliary decision-making model. However, in the actual decision-making process, due to the constraints of the decision-making environment, decision-making overload

may occur, affecting the final decision-making results. Therefore, for the above problems, this paper sets the time constraints of the model, and sets the decision-making time window on the premise of meeting the actual decision-making requirements as far as possible. The decision type generated at this time ς as shown in (4) below.

$$\varsigma = \frac{1}{G_j} \left\| \min F \right\| \bullet \upsilon \tag{4}$$

In formula (4), υ represents the extreme value of decision-making reliability. In fact, in the decision-making process, when the decision-making conditions are subject to too many restrictions, it may increase the difficulty of decision-making. Therefore, the auxiliary decision-making multi model designed in this paper extracts different decision-making restrictions in advance, generates a basic time series, and makes decisions on the basis of meeting the reliability of decision-making. The enterprise case auxiliary decision-making model built on this basis λ_a as shown in (5) below.

$$\lambda_a = \varsigma + N_a \bullet \left(\left| f(x) \right| - \upsilon \right) \tag{5}$$

In formula (5), N_a representing the set of multi-objective optimal decision solutions, the enterprise case aided decision model constructed above can effectively reduce the constraint error and maximize the decision-making reliability of the model.

4 Experiment and result analysis

4.1 Experimental data set and data processing method

In combination with the requirements of multiple model performance verification for enterprise case aided decision making, this paper selects CY54203 dataset as the experimental dataset. This dataset includes the basic auxiliary decision making cases of some enterprises, and contains a variety of field characteristics. The total number of sample data is 888647. The relevant data feature information of this dataset is shown in Table 1 below.

TABLE 1 DATA CHARACTERISTICS OF EXPERIMENTAL DATA SET

Experimental dataset feature number	Meaning of experimental dataset features	Number of characteristic data
01	Sample Enterprise Cases	748
02	Auxiliary institution codes related to enterprise case samples	2136
03	Names of institutions related to corporate cases	4754
04	Types of institutions related to corporate cases	18899
05	Decision location for enterprise assisted decision-making cases	2336

06	Decision year for enterprise assisted decision-making cases	4478
07	Decision months for enterprise assisted decision-making cases	2566
08	Number of enterprise assisted decision-making cases within a certain period of time	4588
09	Decision effectiveness of enterprise assisted decision-making cases	2235
10	Data sources for enterprise assisted decision-making cases	96474
11	Names of personnel related to enterprise assisted decision-making cases	66586
12	Occupation of personnel related to enterprise assisted decision-making cases	245532
13	Personnel relationships involved in enterprise assisted decision-making cases	11523
14	Decision groups for enterprise assisted decision-making cases	324753
15	Enterprise assisted decision-making case information	100000
16	Enterprise assisted decision-making related error case auxiliary dataset	539
17	Dataset of decision plans related to enterprise case assistance decision-making proposed by different departments	500

It can be seen from Table 1 that the data fields in the experimental data set have distinctive characteristics and meet the requirements of the enterprise case aided decision-making multi model performance analysis experiment.In combination with the above experimental data field characteristics, this paper has effectively processed the experimental data set data, as shown in Figure 2 below.

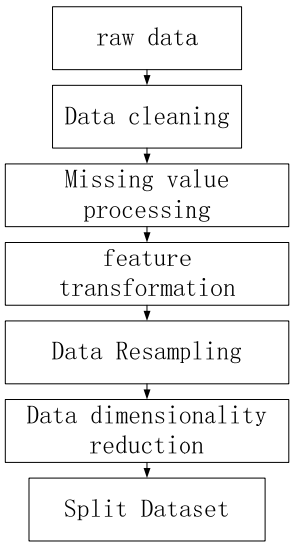

Figure 2. Experimental data set processing

It can be seen from Figure 2 that for the original data samples of the data set, data cleaning is required first to deal with missing values, and then data feature conversion, resampling and dimensionality reduction are required to obtain experimental data that meet the experimental requirements. The processed experimental data eliminates useless features and retains features related to the experiment. After analysis, it is found that some experimental data are missing, so data missing value processing is required. The processing flow is shown in Figure 3 below.

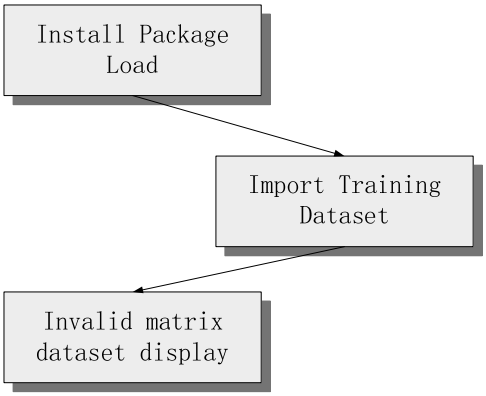

Figure 3. Data Missing Value Processing

It can be seen from Figure 3 that the above processing process is mainly implemented in the Python environment. Missing data is processed through Missingno to realize the visualization of missing data characteristics. After the processing of the experimental data set is completed, this paper selects the classification accuracy as the experimental indicator, and the formula $Accuracy$ as shown in (6) below.

$$Accuracy = \frac{TP + TN}{P + N} \tag{6}$$

In formula (6), TP represents the number of positive examples of correct auxiliary decision-making, TN represents the number of negative cases of correct auxiliary decision-making, P represents the number of positive samples, N represents the number of negative samples. The classification accuracy higher of the $Accuracy$ is, the better the decision-making effect of the multiple decision support models is. On the contrary, the decision-making effect of the multiple decision support models is relatively poor. Combined with the above experimental indicators, the model performance verification experimental results can be obtained.

4.2 Analysis of experimental results

Combined with the above experimental preparation, the performance verification of enterprise case aided decision-making multiple models can be carried out. That is, according to the above selected experimental data sets, the enterprise case aided decision-making multiple models designed in this paper based on data analysis and the enterprise case aided decision-making multiple models based on analytic hierarchy process are used respectively. The multi model of

enterprise case aided decision-making based on association rules assists decision-making on different types of enterprise cases. Formula (6) is used to calculate the classification accuracy of the three models. The experimental results are shown in Table 2 below.

TABLE 2 EXPERIMENTAL RESULTS

Enterprise Case Type (Dataset)	**The classification accuracy of the multi model for enterprise case assistance decision-making based on data analysis designed in this article (%)**	**Classification accuracy of enterprise case assisted decision-making multiple models based on Analytic Hierarchy Process (%)**	**Classification accuracy of enterprise case assisted decision-making multiple models based on association rules (%)**
Enterprise strategic case auxiliary decision data set R1	99.542	74.546	65.551
Enterprise Management Case Assistance Decision Dataset R2	99.265	75.232	64.233
Enterprise Business Case Assistance Decision Dataset R2	99.849	74.358	63.856
Enterprise Deterministic Case Assistance Decision Dataset R4	99.452	72.695	66.825
Enterprise Risk Case Aided Decision Dataset R5	99.863	63.452	59.639
Enterprise Uncertain Case Assistance Decision Dataset R6	99.124	75.336	68.545
Enterprise Programmed Case Assistance Decision Dataset R7	99.252	69.985	74.256
Enterprise Unprogrammed Case Assistance Decision Dataset R8	98.884	72.456	75.789
Enterprise Permission Classification Case Assistance Decision Dataset R9	99.263	63.365	66.546
Enterprise Individual Case Assistance Decision Dataset R10	99.259	76.747	53.525
Enterprise Group Case Assistance Decision Dataset R11	99.885	79.456	78.264

It can be seen from Table 2 that the classification accuracy of the enterprise case aided decision-making multiple models designed in this paper based on data analysis is high in different types of enterprise aided decision-making cases, while the classification accuracy of the enterprise case aided decision-making multiple models based on analytic hierarchy process and the enterprise case aided decision-making multiple models based on association rules is relatively low. The above experimental results prove that the multi model of enterprise case aided decision-making designed in this paper based on data analysis has good decision-making effect, reliability and certain application value.

5 Conclusion

In the digital era, enterprise decision-making has also changed from the original manual decision-making to intelligent and information-based computer-aided decision-making. The online platform can be used to conduct data processing and intelligent analysis in advance, extract decision characteristics, improve decision content, and reduce the difficulty of decision making. However, due to the large amount of data content involved in enterprise cases, most of the current enterprise case aided decision making methods lack effective decision models. According to the characteristics of enterprise case aided decision making, this paper constructs an effective multi model of enterprise case aided decision making using data analysis technology. The experiment results show that the designed multi model of enterprise case aided decision-making has good decision-making effect, high accuracy of decision classification, reliability and certain application value, and has made certain contributions to improving the comprehensive decision-making strength of enterprises.

References

[1] Z Zheng,J Zhou. Data Classification Based on Chaos Correlation Dimension and Decision Tree[J].Computer Simulation, 2022, 39(10):327-330.

[2] Liu M , Zhang Z , Zhang S ,et al.Research On Auxiliary Decision-making Technology In the Early Stage of Power Network Base On Multi-plan Integration[J].IOP Conference Series: Earth and Environmental Science, 2021, 621(1):012025 (9pp).

[3] Yu J .Auxiliary research on difficult aerobics exercise training based on fpga and movement recognition technology[J].Microprocessors and Microsystems, 2021, 81(2):103656.

[4] Zhu L .Computer Vision-Driven Evaluation System for Assisted Decision-Making in Sports Training[J].Wireless Communications and Mobile Computing, 2021, 2021(7):1-7.

[5] Li J , Sun Y , Gong L ,et al.Multiattribute Fuzzy Decision Evaluation Approach and Its Application in Enterprise Competitiveness Evaluation[J].Mathematical Problems in Engineering, 2021, 2021(1):1-11.

[6] Shao L .Research on Sports Training Decision Support System Based on Improved Association Rules Algorithm[J].Security and Communication Networks, 2021, 2021(4):1-6.

[7] Pan W , Feng L , Zhang L ,et al.Time-series interval prediction under uncertainty using modified double multiplicative neuron network[J].Expert Systems with Applications, 2021, 184(1):115478.

[8] Wang X , Liu H , Tian Y ,et al.A Fast Optimization Method of Water-Dropping Scheme for Fixed-Wing Firefighting Aircraft[J].IEEE Access, 2021, PP(99):1-1.

[9] Wang Z , Weng L , Lu M ,et al.Application of Multiattention Mechanism in Power System Branch Parameter Identification[J].Complexity, 2021, 2021(4):1-12.
[10] Yu J , Lin S , Huang F ,et al.Evaluation and Analysis of Total Factor Input and Output of Power Grid Enterprises under the New Situation[J].IOP Conference Series: Earth and Environmental Science, 2021, 680(1):012046 (6pp).
[11] Yang B , Ma J , Huang G ,et al.Development and Application of 3D Visualization Platform for Flood Evolution in Le'an River Basin of Wuyuan[J].IOP Conference Series Earth and Environmental Science, 2021, 638(1):012053.
[12] Liu M , Wang Z , Wang L .Automatic Chinese Text Summarization for Emergency Domain[J].Journal of Physics: Conference Series, 2021, 1754(1):012213 (6pp).
[13] Chen H , Zhang J , Tian X ,et al.Substation Location Selection Using GIS and Improved Matter-Element Extension Method[J].Journal of Physics: Conference Series, 2021, 1744(4):042188 (11pp).
[14] Jin J , Zhang X , Xu L ,et al.Impacts of carbon trading and wind power integration on carbon emission in the power dispatching process[J].Energy Reports, 2021, 7(2):3887-3897.
[15] Wang X , Lu A G , Liu J ,et al.Intelligent interaction model for battleship control based on the fusion of target intention and operator emotion[J].Computers & Electrical Engineering, 2021, 92(16):107196.
[16] Jin J , Wen Q , Zhang X ,et al.Economic Emission Dispatch for Wind Power Integrated System with Carbon Trading Mechanism[J].Energies, 2021, 14(7):1870.
[17] Chen J , Shi W , Liu Y ,et al.Big Data Analysis Supports the Research and Pilot Application of Distribution Network Planning Technology[J].IOP Conference Series: Earth and Environmental Science, 2021, 632(4):042047 (9pp).
[18] Yuan X , Yao H , Wang J ,et al.Artificial Intelligence Empowered QoS-Oriented Network Association for Next-Generation Mobile Networks[J].IEEE Transactions on Cognitive Communications and Networking, 2021, PP(99):1.
[19] Wang J , Wang C , Liang Y ,et al.Data-Driven Chance-Constrained Optimal Gas-Power Flow Calculation: A Bayesian Nonparametric Approach[J].IEEE Transactions on Power Systems, 2021, 99(pp):1-14.
[20] Liang J , Liu D .Research on capacity planning of thermal power units in deep peak regulation auxiliary service[J].IOP Conference Series: Earth and Environmental Science, 2021, 804(4):042028 (5pp).
[21] Innes H , Jepsen P , Mcdonald S ,et al.Performance of models to predict hepatocellular carcinoma risk among UK patients with cirrhosis and cured HCV infection.[J].JHEP reports : innovation in hepatology, 2021, 3(6):100384.
[22] Khalfaoui S , Leneutre J , Villard A ,et al.Security Analysis of Out-of-Band Device Pairing Protocols: A Survey[J].Wireless Communications and Mobile Computing, 2021, 2021(1):1-30.
[23] Wu Q , Gong X , Xu K ,et al.Towards Target-Driven Visual Navigation in Indoor Scenes via Generative Imitation Learning[J].IEEE Robotics and Automation Letters, 2021, 6(1):175-182.
[24] Qian J , Wu J , Yao L ,et al.Comprehensive Performance Evaluation of Wind-Solar-CCHP System Based on Emergy Analysis and Multi-objective Decision Method[J].Energy, 2021, 230(190):120779.
[25] Wei X , Quan S , Zhuang Y S ,et al.Research on Dynamic Resource Allocation and Decision-making Model of Equipment Independent Maintenance[J].Journal of Physics: Conference Series, 2021, 1939(1):012014 (9pp).