Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 5
AIDML 2024

Regional Disaster Risk Assessment Based on ARIMA Prediction and Comprehensive Evaluation Model

Chuanhui Cui, Meiting Dong *, Haiqing Li

School of Information and Control Engineering, Qingdao University of Technology, Qingdao, China, 266520

* Corresponding author: 201805650110@smail.xtu.edu.cn

Abstract. The occurrence of extreme weather events affects the natural economic and social security, and has a serious impact on some industries, so the prediction and evaluation of disaster risk level is particularly important. This paper collected data on nine indicators from 25 regions and divided them into four areas: extreme weather, natural, social, and economic. the ARIMA prediction model is used to predict the probability and intensity of extreme weather in the future. The combined weights of each index were calculated by Composite Weighting Method, including principal component analysis, entropy weight method and coefficient of variation method, and the Risk Assessment Index (RAI) formula was constructed, and then the WNSE prediction and evaluation model was constructed. It is divided into four risk levels by fuzzy cluster analysis. We use WNSE prediction and evaluation model to calculate New York City and Switzerland's RAI and coverage levels. Using this prediction, we can assess the severity of the consequences of natural disasters and prevent or remedy them. The assessment of regional disaster risk levels can give relevant reference to the government, and the government can make reasonable resource allocation.

Keywords: ARIMA prediction model; WNSE prediction evaluation model; Combination Weighting Method; fuzzy cluster analysis.

1. Introduction

Unexpected natural disasters and man-made disasters have brought heavy suffering to people. Extreme weather events put heavy pressure on property owners. There is therefore an urgent need for better climate change strategies to ensure that everyone is adequately protected from extreme weather events.

To assess disaster risk in various regions, the predecessors established the risk assessment matrix according to the possibility and consequence of disaster. Among them, the possibility index is determined by the risk of natural disaster factors and the ability to withstand natural disasters. Consequences severity indicators are determined by calculating specific formulas that consider possible human casualties, economic losses and social impacts. However, there are also some disadvantages, such as the division of natural factors and other influencing factors is not detailed enough, and the application scope of the model is too small - it is only limited to one enterprise, and it is difficult to cover an entire region.

In this paper, a more reasonable model is proposed. First, this paper presents the Extreme-Weather-Society-Nature-Economy prediction assessment model, which is used to assess the level of disaster risk in a region. To ensure the comprehensiveness of the model, consider multiple indicators reflecting its risk level, divide each category of indicators into modules, and forecast extreme weather factors through the ARIMA model which is established and verified its feasibility [1]. To increase the accuracy of the model, this paper adopted the principal component analysis method. Entropy weight method and coefficient of variation method are used to calculate the combined weight. To enhance the rationality of the model, this paper selected 25 representative locations in different regions, calculated the disaster risk level according to relevant data, and analyzed different disaster risk levels through fuzzy clustering [2,3]. These are shown in the figure 1.

Then, the feasibility is verified by two different location types in New York City and Switzerland, and the rationality is analyzed in combination with the local conditions [3].

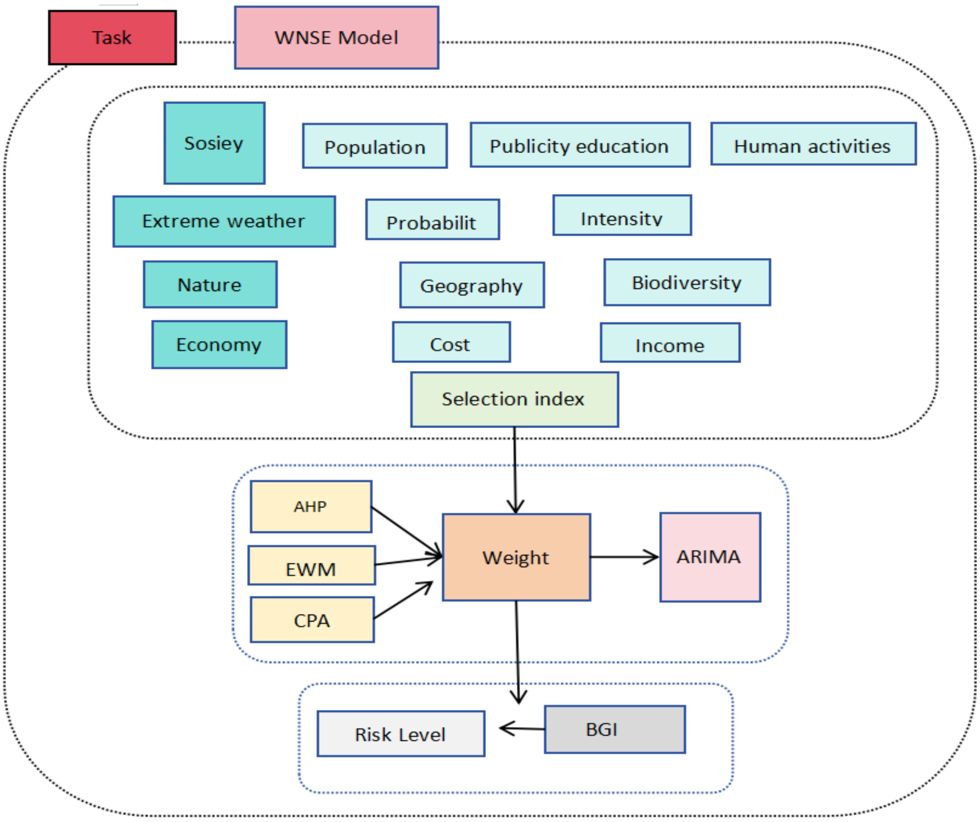

Figure 1. Our work

2. Theory

2.1. Notations

The primary notations used in this paper are listed in Table 1.

Table 1. Notations

Symbol	Description	Unit
W_1	Frequency of extreme weather	-
W_2	Extreme weather intensity	BOU
M	The degree of economic loss	-
E_2	Premium income	dollar
S_3	Population	-
carbon	Carbon dioxide emission	Gt
water	Water resource utilization	%
t	Sunshine duration	h

*Other symbol descriptions will be provided in the main text.

2.2. Secondary index

On the basis of reviewing a large number of literature searches and relevant papers, this paper summarized four key factors reflecting the level of underwriting risk: extreme weather factor (W), social factor (S), natural factor (N) and economic factor (E), which were used as the firstlevel indicators of the evaluation system. After further thinking and summary, it further concretized the indicators and selected a total of 8 specific indicators as the secondary indicators of the evaluation system.

The 8 specific indicators are shown in Table.2.

Table 2. 8 evaluation system indicators

Symbol	Description	Significance
W_1	Frequency of extreme weather	The number of times a particular extreme weather event occurs within a certain time frame
W_2	Extreme weather intensity	A weather phenomenon at an extreme level
S_1	Publicity and education	The level of public concern about extreme weather
S_2	Human activities	The impact of human activities on regional risk levels
S_3	Population	The impact of population growth on climate change and extreme weather events
N_1	Geographical factors	determines the climate type and the main weather type of the region
N_2	Biodiversity	Ecosystems rich in biodiversity are more likely to be more flexible and resilient to extreme weather and natural disasters
E	Economic losses	Economic and property losses caused by natural disasters

In this paper, the percentage value that is above normal is converted into a fraction recorded as m. And convert the percentage into scores by:

$$\mathrm{E} = \frac{100 \times \mathrm{E}}{\max\{\mathrm{E}_1, \mathrm{E}_2, \cdots, \mathrm{E}_{\mathrm{X},}\}} \tag{1}$$

At the same time, this paper chooses carbon dioxide emission (carbon) and water resource utilization (water) as the main indicators affecting human activities. Their values are determined by the following formula:

$$\mathrm{S}_2 = \text{carbon} + \text{water} \tag{2}$$

The influence of geographical factors was assessed by altitude (h) and sunshine duration (t).The relationship between geographical factors and altitude (h) and sunshine hours (t) is:

$$N_1 = h' + t' \tag{3}$$

In the above formula, h' is the standardized , and is the same. The specific standardized formula is:

$$h' = \frac{h_i - \min\{h_1, \ldots, h_{25}\}}{\max\{h_1, \ldots, h_{25}\} - \min\{h_1, \ldots, h_{25}\}} (i = 1, 2, \ldots, 25) \tag{4}$$

$$t' = \frac{t_i - \min\{t_1, \ldots, t_{25}\}}{\max\{t_1, \ldots, t_{25}\} - \min\{t_1, \ldots, t_{25}\}} (i = 1, 2, \ldots, 25) \tag{5}$$

This paper uses green area (S_{green}) and wildlife species (C_{wild}) to quantify impacts.

$$N_2 = \frac{S_{green}}{S_{location}} + \frac{C_{wild}}{S_{green}} \tag{6}$$

2.3. ARIMA Prediction Model

ARIMA model is widely used in the analysis and modeling of various time series data. Using past observations of the sequence, future values of the sequence can be extrapolated. In the ARIMA model, the future value of the sequence is expressed as a linear function of the lag term, the current period and the lag period of the random disturbance term [4,5].

The general form of the model is as follows:

$$Y_t = c + \alpha_1 Y_{t-1} + \cdots + \alpha_p Y_{t-p} + \varepsilon_t + \beta_1 Y_{t-1} + \cdots + \beta_q Y_{t-q} \tag{7}$$

2.3.1. Establishment of ARIMA prediction model.

First, in this paper, use ADF test method to verify the stationarity of time series, and carry out unit root test on the original sequence. If the sequence does not meet the stationarity condition, the non-stationarity time series can be transformed into stationary time series by difference transformation or log-difference transformation, and then ARIMA model is constructed for stationary time series.

Next, the order of the model is determined. By using some statistics that can describe the characteristics of the sequence, such as Autocorrelation Coefficient (AC) and Partial Autocorrelation Coefficient (PAC), the possible forms of the model are initially identified, and then an optimal model is selected from the available models according to Akaike Information Criterion (AIC) and other order criteria.

Finally, parameter estimation and diagnostic test were performed. It includes testing the significance of model parameters, the validity of the model itself and testing whether the residual sequence is a white noise sequence. If the model passes the test, the model setting is basically correct; otherwise, the model form must be re-determined and diagnostic tests must be performed until the model form is correctly set.

2.3.2. Analysis and validation of ARIMA prediction model.

This paper collected frequencies from ten different regions from 1997 to 2003 for analysis as a database of ARIMA models. The future value of the data is expressed as a linear function of the lag term and the lag period of the random disturbance term. The ADF output table.3 is as follows:

Table 3. ADF test list

Variable	Difference order	t	P	AIC	Threshold		
					1%	5%	10%
Frenquency	0	-1.227	0.662	182.701	-3.661	-2.961	-2.619
	1	-4.217	0.001***	178.852	-3.661	-2.961	-2.619
	2	-3.742	0.004***	181.465	-3.679	-2.968	-2.623

Note: ***, ** and * represent significance levels of 1%, 5% and 10% respectively

The model requires that the series must be stationary time series data. By analyzing the value, we analyze whether it can reject the null hypothesis of sequence instability significantly.

If it is significant (P<0.05), it means that the null hypothesis is rejected and the series is a stationary time series; otherwise, it means that the series is an unstable time series.

As can be seen from the above table:

When the difference is of order 0, the significance P-value is 0.662, and the significance is not present at the level, and the null hypothesis cannot be rejected. The series is an unstable time series.

When the difference is of order 1, the significance P-value is 0.001***, showing significance at the level, rejecting the null hypothesis, and the series is a stationary time series.

When the difference is of order 2, the significance P-value is 0.004***, showing significance in level, rejecting the null hypothesis, and the series is a stationary time series. The optimal difference sequence diagram is as follows:

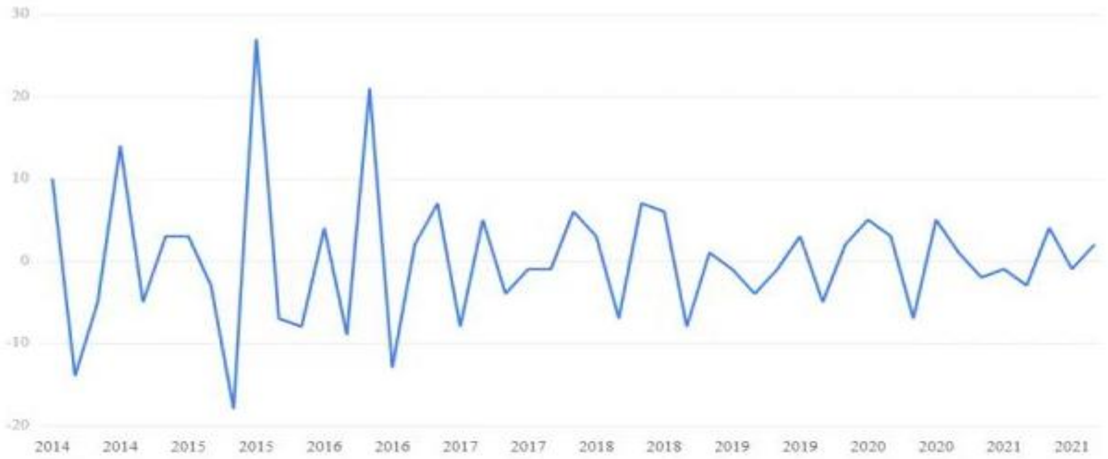

Figure 2. Optimal difference sequence diagram

As can be seen from the above figure 2, it shows the time sequence diagram after the first-order difference of the original data.

Table 4. Model parameter table

	Coefficient	Standard deviation	t	$P > \|t\|$	0.025	0.975
Constant	0.289	0.341	0.847	0.397	-0.38	0.958
$ar.L_1$	-1.045	0.188	-5.559	0	-1.414	-0.677
$ar.L_2$	-0.462	0.207	-2.235	0.025	-0.867	-0.057
$ma.L_1$	0.017	0.239	0.07	0.944	-0.452	0.485
$ma.L_2$	-0.822	0.253	-3.256	0.001	-1.318	-0.327
$sigma_2$	31.07	6.973	4.456	0	17.403	44.737

Note: ***, ** and * represent significance levels of 1%, 5% and 10% respectively

The above table.4 shows the parameter results of the model, including coefficient, standard deviation and T-test results of the model, which are used to analyze the model formula.

Based on the frequency variable, the model results are the test table of ARIMA model (2,3,2). The model formula is as follows:

$$y(t) = 0.289 - 1.045 \times y(t-1) - 0.462 \times (t-2) + 0.017 \times \varepsilon(t-1) - 0.822 \times \varepsilon(t-2) \tag{8}$$

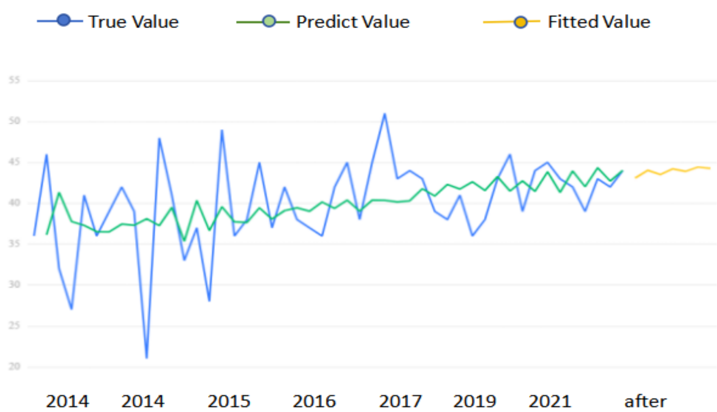

Figure 3. Time series diagram

The figure 3 above shows the original data graph, model fitting value and model prediction value of the time series model.

2.4. Determination of index weights

Through the weight method, principal component analysis method and coefficient of variation method, three different weights of each index are calculated [6]. To reduce the error, the weight values of the above three are averaged to obtain the final weight value [7, 8].

$$\omega_j = \frac{\omega_{j1}+\omega_{j2}+\omega_{j3}}{3} \tag{9}$$

The final calculated weights are shown in the figure 4:

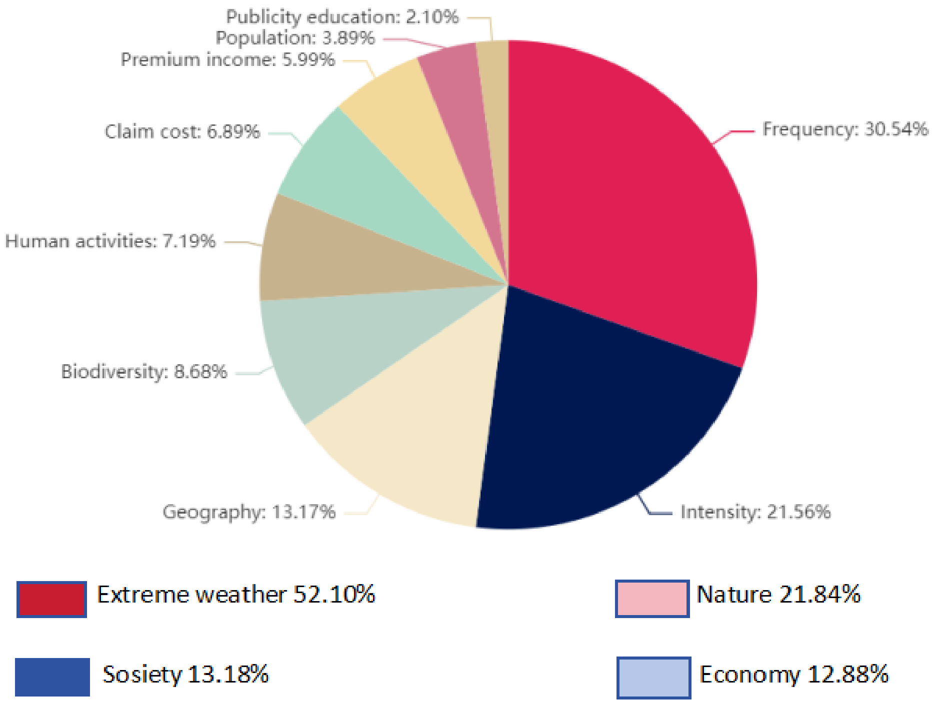

Figure 4. The combined weights of each indicator

2.5. Establishment of WNSE prediction evaluation model

Considering the effects of extreme weather, nature, society and economy on underwriting risk level, this paper builds a WNSE evaluation model. The RAI (Risk Assessment Index) was further introduced to describe quantitatively the level of risk underwritten in regions where extreme weather is occurring more frequently. In addition, the RAI is divided into four intervals through fuzzy cluster analysis, that is, the disaster risk level is divided into four levels [9].

2.5.1. Calculation of the Bear Guarant Index (RAI).

The weights corresponding to indicators W, N, S and E are taken as the influence degree of these three indicators on RAI, that is, the greater the weight, the greater the influence of the indicator on RAI. Accordingly, comparing with some proportion composition formulas, this paper reasonably constructed the calculation formula of RAI:

$$\mathrm{BGI} = 100 \cdot (\omega_W \cdot W + \omega_N \cdot N + \omega_S \cdot S + \omega_E \cdot E) \tag{10}$$

Based on the same lines, construct formulas for W, N, S, and E:

$$\begin{cases} W = \omega_{W_1} \cdot W_1' + \omega_{W_2} \cdot W_2' \\ N = \omega_{N_1} \cdot N_1' + \omega_{N_2} \cdot N_2' \\ S = \omega_{S_1} \cdot S_1' + \omega_{S_2} \cdot S_2' + \omega_{S_3} \cdot S_3' \\ E = \omega_{E_1} \cdot E_1' + \omega_{E_2} \cdot E_2' \end{cases} \tag{11}$$

2.5.2. Determination of underwriting risk level.

The data is put into the WNSE model and the RAI is derived from the proportion. Then, fuzzy cluster analysis is used to classify all locations and divide coverage levels into four levels [9]. The specific grade classification is shown in the table 5:

Table 5. Level of underwriting risk

Grade	I	II	III	IV
Bear guarantee risk level	High	Moderate	Minimal	Unpolluted
RAI	<40	40-60	60-80	>80

Based on the table above, the level of insured risk for a location can be determined.

Class I high-risk areas have the greatest risk, and the frequency and intensity of extreme weather are very high, such as areas located at high latitudes and high altitudes where extreme weather conditions frequently occur and natural disasters occur.

The risk level of level II risk areas is moderate, such as coastal and volcanic areas;

The underwriting risk of level III low-risk areas is weak, such as areas with suitable climate and superior geographical conditions;

Level IV risk-free areas, almost no risk, such as natural areas that have not been affected by human activities, and areas where the climate is suitable all year round.

3. Results

3.1. Case I: New York City

New York City is in the state of New York and is one of the largest cities in the United States. At the same time, it is also an area with a large population, a high level of economic development and developed infrastructure. The relevant data are shown in Table 6.

Table 6. New York City data

Costant	Data	Unit
W_1	42	-
W_2	7	BOU
M	16	-
E_2	2150	dollar
S_2	8349817	-
carbon	0.95	Gt
water	68	%
h	16	ft
t	2055	h
S_{green}	20	km^2
$S_{location}$	1213.4	km^2
C_{wild}	37	-

Putting the above data into the WNSE prediction evaluation model, the RAI value of New York City is 34.43, which belongs to the upper class I, indicating a relatively high degree of disaster risk level.

The level of risk in New York City is influenced by several factors. Its location on the Atlantic hurricane path makes it vulnerable to hurricanes and tropical storms, and New York City's extremely high population density means that many people and property could be affected in the event of a disaster. In addition, New York is home to many high-rise buildings and subway systems, which are of great significance to the impact of earthquake, storm surge, flooding and other disaster events. As a result, New York City has a relatively high level of disaster risk level risk.

3.2. Case II: Switzerland

Switzerland is ideally located in central Europe. At the same time, it is also a highly developed country, which belongs to the category of low risk. Therefore, we chose Switzerland as the site for this study. The relevant data are shown in Table 7.

Table 7. Switzerland data

Costant	Data	Unit
W_1	11	-
W_2	3.5	BOU
M	11.7	-
E_2	980	dollar
S_2	87000000	-
carbon	0.67	Gt
water	72	%
h	13	ft
t	2250	h
S_{green}	16.8	km^2
$S_{location}$	41290	km^2
C_{wild}	35	-

Entering the above data into the WNSE evaluation model, Switzerland has a RAI value of 79.86, which is level III, meaning that Switzerland is in the low-risk category.

Switzerland's relatively stable geographical location in central Europe reduces the impact of natural disasters; The high level of economic development helps to reduce economic risks and business risks; The facilities are sound, which can effectively reduce the loss caused by natural disasters or accidents.

4. Conclusion

The related data is divided into extreme weather, natural, social, and economic four areas. The ARIMA prediction model is then used to predict the probability and intensity of extreme weather in the future. The combined weights of each index were calculated by Composite Weighting Method, including principal component analysis, entropy weight method and coefficient of variation method, and the Bear Guarant Index (BGI) formula was constructed, and then the WNSE prediction and evaluation model was constructed [10]. Finally, it is divided into four risk levels by fuzzy cluster analysis.

Through the WNSE prediction evaluation model, the RAI value of a place can be obtained and judge the disaster risk level of the place, which is of great significance for investment benefit, building characteristics and disaster protection.

References

[1] Shrikant, Reddy S V G, Manjunath K M, et al. Predicting Potential Evapotranspiration for Kalaburagi District using a Seasonal Arima Model [J]. International Journal of Environment and Climate Change, 2023, 13 (11): 2073 - 2082.

[2] Xiangqun L, Jiawen L, Jinyu Z, et al. Optical Fibre Communication Feature Analysis and Small Sample Fault Diagnosis Based on VMD-FE and Fuzzy Clustering [J]. 1 Gannan Power Supply Company of State Grid Gansu Electric Power Supply Company, Hezuo, 747000, China; 2 Digital Communication Department, Gannan Power Supply Company of State Grid Gansu Electric Power Supply Company, Hezuo, 747000, China, 2023, 121 (1): 203 - 219.

[3] Lu H, Zhidong L, Jipeng Q, et al. Fuzzy clustering analysis for the loan audit short texts [J]. Knowledge and Information Systems, 2023, 65 (12): 5331 - 5351.

[4] Munich RE. (2022, January 10). Hurricanes, cold waves, tornadoes: Weather disasters in USA dominate natural disaster losses in 2021.Retrieved at: https://www.munichre.com/en/company/media-relations/media-information-and-corporate-news/media-information/2022/natural-disaster-losses-2021.html.

[5] Tolesh F, Biloshchytska S. Forecasting international migration in Kazakhstan using ARIMA models [J]. Procedia Computer Science, 2024, 231176 - 183.

[6] Liu X. Evaluation method of English flipped classroom teaching effect based on entropy weight method [J]. International Journal of Continuing Engineering Education and Life-Long Learning, 2024, 34 (1): 13 - 24.

[7] WenGang H, ShaoWei Z, Guozhi W, et al. Modeling Methodology for Site Selection Evaluation of Underground Coal Gasification Based on Combination Weighting Method with Game Theory. [J]. ACS omega, 2023, 8 (12): 11544 - 11555.

[8] Ji S, Jiajun L, Chenjing L, et al.An identification method for vulnerable lines based on combination weighting method and GraphSAGE algorithm [J]. International Journal of Electrical Power and Energy Systems, 2023, 149.

[9] T. N K L, Dan N, Tai V. Fuzzy cluster analysis algorithm for image data based on the extracted feature intervals [J]. Granular Computing, 2023, 8 (6): 2067 - 2081.

[10] Hou J, Gao T, Yang Y, et al. Battery inconsistency evaluation based on hierarchical weight fusion and fuzzy comprehensive evaluation method [J]. Journal of Energy Storage, 2024, 84 (PA): 110878-.