ISSN 2616-5805 Vol. 4, Issue 1: 44-51, DOI: 10.25236/AJMS.2023.040107

Prediction modeling of nuclear bomb numbers based on random forests and LSTM

Yinghao Meng[1], Erzhen Lou[2], Yufei Zheng[3], Bowen Liu[4,*], Xun Zheng[1]

[1]*College of Mechanical and Electrical Engineering, Wenzhou University, Wenzhou, China*
[2]*College of Mathematics and Physics, Wenzhou University, Wenzhou, China*
[3]*School of Information Science and Engineering, University of Jinan, Jinan, China*
[4]*College of Big Data and Intelligent Engineering, Southwest Forestry University, Kunming, China*
[*]*Corresponding author: a182328388881@163.com*

Abstract: *In this paper, we aim to identify the countries with the most frequent nuclear weapons activity in the last decade and to predict the number of nuclear bombs in the world. First we developed the TOPSIS evaluation model and used the number of nuclear tests and the change in nuclear weapons as two indicators for the model. We eventually concluded that North Korea has been the most active country in developing nuclear weapons in the last decade. We then built a random forest prediction model to predict the number of nuclear-armed states. An LSTM prediction model was then built to predict the number of nuclear weapons in the world and in each country. The predictions gave us: the number of nuclear-armed states in the next 100 years is 8. The number of nuclear weapons per country in 2123 is: China (360), France (258), India (183), Iraq (91), North Korea (11), Pakistan (200), Russia (200), UK (191) and USA (3319). Finally, we conclude with a few recommendations for peace in the human world based on our findings.*

Keywords: *TOPSIS, Random Forest, LSTM, Nuclear Weapons*

1. Introduction

With the rise of multilateralism and the changes in the international situation in recent years, more and more countries are using nuclear weapons as a means of defence for themselves. Although the number of nuclear bombs in the world today is far lower than before, they still pose a great danger to human life and mankind does not currently have the means to resist the consequences of a nuclear explosion. We have therefore analysed the nuclear testing activity of individual countries and predicted the number of nuclear bombs to provide some suggestions for ensuring peace in the human world.

2. TOPSIS [1] model building

2.1 Harmonisation of indicator types

As the individual influencing factors are not all bigger the better, there are also individual factors that score higher the smaller their value, so first we have to positive all the factors. The formula for the forwarding process is as follows.

$$x_z = max - x \tag{1}$$

2.2 Standardised handling

In order to eliminate the effect of different magnitudes, the already normalised matrix needs to be normalised. Assuming that there are n objects to be evaluated and m positive evaluation indicators, the positivisation matrix is composed as follows:

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix} \tag{2}$$

We denote the normalised matrix as Z. Each element of Z is expressed as:

ISSN 2616-5805 Vol. 4, Issue 1: 44-51, DOI: 10.25236/AJMS.2023.040107

$$z_{ij} = x_{ij}/\sqrt{\sum_{i=1}^{n} x_{ij}^2} \quad (3)$$

2.3 Calculating the score

In the previous section we obtained the matrix after the criteria.

$$Z = \begin{bmatrix} z_{11} & \cdots & z_{1m} \\ \vdots & \ddots & \vdots \\ z_{n1} & \cdots & z_{nm} \end{bmatrix} \quad (4)$$

Define maximum value:

$$Z^+ = (Z_1^+, Z_2^+, \cdots, Z_m^+) = (max\{z_{11}, z_{21}, \cdots, z_{n1}\}, max\{z_{12}, z_{22}, \cdots, z_{n2}\}, max\{z_{1m}, z_{2m}, \dots, z_{nm}\} \quad (5)$$

Define minimum value:

$$Z^- = (Z_1^-, Z_2^-, \cdots, Z_m^-) = (min\{z_{11}, z_{21}, \cdots, z_{n1}\}, min\{z_{12}, z_{22}, \cdots, z_{n2}\}, min\{z_{1m}, z_{2m}, \dots, z_{nm}\} \quad (6)$$

Definition No. $i(i = 1,2, \cdots, n)$distance of the evaluation object from the maximum value:

$$D_i^+ = \sqrt{\sum_{j=1}^{m} (Z_j^+ - z_{ij})^2} \quad (7)$$

Definition No. $i(i = 1,2, \cdots, n)$distance of the evaluation object from the maximum value:

$$D_i^- = \sqrt{\sum_{j=1}^{m} (Z_j^- - z_{ij})^2} \quad (8)$$

From this we can calculate the number of $i(i = 1,2, \cdots, n)$ un-normalised scores for sample objects.

$$S_i = \frac{D_i^-}{D_i^+ + D_i^-} \quad (9)$$

Afterwards, we normalised the unnormalised scores to remember to get to the final score.

3. LSTM [2] model building

3.1 Modeling the three key components

The LSTM model has three main gates, namely the forgetting gate, the input gate and the output gate, the formulas and pictures of which are shown below.

3.1.1 The forgetting gate

The forgetting gate, as the name suggests, indicates how much of the state at the last point in time I should forget. For C_{t-1}, one would first look at the output h_{t-1} of the previous stage and the input x_t of this stage and determine how much to let C_{t-1}, come to forget by sigmoid. Sigmoid=1 means to save more of the weight of C_{t-1} and equal to 0 means to completely forget the previousC_{t-1}. Its main formula is.

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \quad (10)$$

3.1.2 The input gate

First it will take the output h_{t-1} from the previous stage and the input x_t from this stage and control by sigmoid how much is now to be added into the main plot Ct, i.e. the meaning of the first formula; then another alternative $\widehat{C_t}$ will be created, using tanh to control how much of the C_t is to be added. Afterwards, by multiplying the two parts together, the total amount to be affected by C_t is determined, and with the effect of the previous oblivion gate, it can be written as:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (11)$$

$$\tilde{C_t} = \tan h(W_C \cdot [h_{t-1}, x_t] + b_c) \quad (12)$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C_t} \quad (13)$$

ISSN 2616-5805 Vol. 4, Issue 1: 44-51, DOI: 10.25236/AJMS.2023.040107

3.1.3 The output gate

In the last step, with the effect on Ct, we finally see how much we actually want to output: we use the sigmoid function to decide which part of Ct needs to be output, i.e. the o_t of the first formula; after that, we put Ct into tanh to decide which part of Ct is finally output and multiply it with ot to get the final output.

$$o_t = \sigma(W_0[h_{t-1}, x_t] + b_0) \quad (14)$$

$$h_t = o_t * \tan h(C_t) \quad (15)$$

3.2 Establishing evaluation criteria to evaluate the model

Here we use the root mean square error as the evaluation criterion for the model, which represents the square root of the ratio of the square of the deviation of the predicted value from the true value to the number of observations, is used to measure the deviation of the predicted value from the true value, and is more sensitive to outliers in the data.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{n} (Y_i - f(x_i))^2} \quad (16)$$

4. Random Forest prediction model[3] building

4.1 Input sample set

Input sample set $D = (x_1, y_1), (x_2, y_2), \cdots (x_m, y_m)$

4.2 Selecting the training and test sets

The training dataset and sample features are randomly selected for T-round training, for $t = 1,2, \ldots, T$:

The training set is randomly sampled for the tth time, and a total of m times are taken to obtain a sample set containing m samples D_t.

The tth decision tree model $G_t(x)$ is trained with the sample set D_t. When training the nodes of the decision tree model, some of the sample features are selected from all the sample features on the nodes, and an optimal feature is selected from these randomly selected features to make the left and right subtrees of the decision tree.

4.3 Output

Output the ultimate strong learner $F(x)$

4.4 Build regression loss function

Here we use a common loss function in regression models - the mean absolute error, which is the sum of the absolute values of the differences between the target and predicted values. It measures only the mean modal length of the error in the predicted value, regardless of direction, and takes values from 0 to positive infinity. Its calculation formula is.

$$MAE = \frac{1}{m} \sum_{i=1}^{m} |(y_i - \hat{y_i})| \quad (17)$$

5. Result

5.1 Results from the TOPSIS method

We use the TOPSIS method to rate how active each country has been in nuclear weapons research over the last 10 years. We create two indicators, the number of nuclear tests and the number of changes in nuclear weapons, and we find that both indicators are positive, i.e. the more tests and the greater the number of changes in nuclear weapons, the higher the level of research activity, and we use the code to

ISSN 2616-5805 Vol. 4, Issue 1: 44-51, DOI: 10.25236/AJMS.2023.040107

solve for each country's ranking and normalised score. The country rankings and normalised scores obtained are shown in Table 1.

Table 1: TOPSIS method results

Country	Score	Ranking
China	0.101	2
France	0.097	7
India	0.099	3
Israel	0.097	5
North Korea	0.159	1
Pakistan	0.099	4
Russia	0.082	9
South Africa	0.097	6
United Kingdom	0.095	8
United States	0.067	10

From this, we see that North Korea has been the most active in nuclear weapons development in the last decade, and the United States has been the least active in nuclear weapons development.

5.2 Results from the Random Forest prediction method

In terms of forecasts for countries possessing nuclear weapons, we do not see any major changes in the nuclear-armed states in the future as the current world order has a series of restrictions on nuclear weapons, such as the Nuclear Weapons Non-Proliferation Treaty, and has been unanimously accepted by countries around the world. The test set and training set results of the Random Forest algorithm for the prediction aspect are shown in Figure 1.

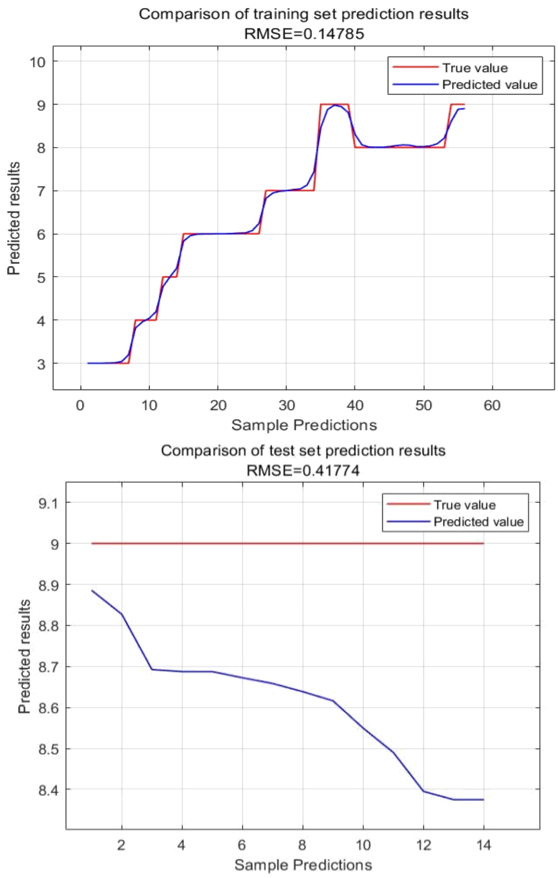

Figure 1: Prediction results for the training set (left) and test set (right)

Since the number of nuclear-armed countries is an integer greater than zero, phenomena such as overfitting may occur when predicting, here we use the Random Forest algorithm for regression and prediction to eliminate overfitting. Here are the prediction results for the training and test sets.

ISSN 2616-5805 Vol. 4, Issue 1: 44-51, DOI: 10.25236/AJMS.2023.040107

As the number of random forests increases, the error plot of the predictions is shown in Figure 2.

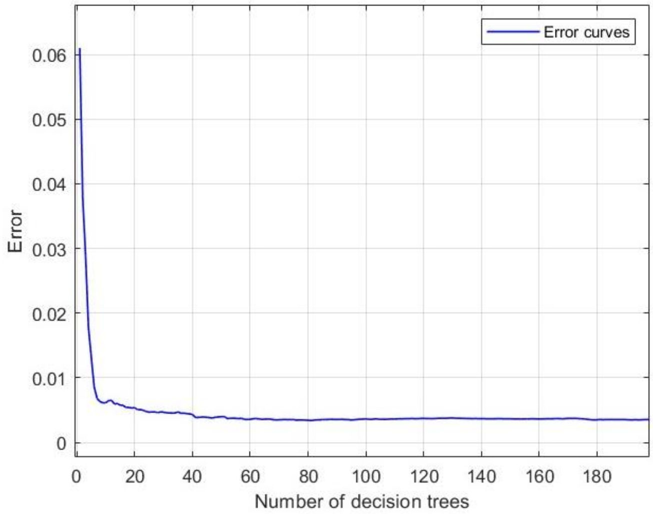

Figure 2: Plot of error, change in number of random forest

From the graph we can see that the error rate of our model is decreasing as the number of random forest grows. During this period we calculated MAEs of 0.097666 and 0.37867 for the training and test sets respectively, which also reflects the high accuracy of the model in side[4-5].

As can be seen from the figure, the RMSE values for the training and test sets reached 0.14 and 0.41 respectively, which is a very good prediction. After that we performed the prediction of the number of nuclear-armed countries in the next 100 years and obtained the following results. This is shown in Figure 3.

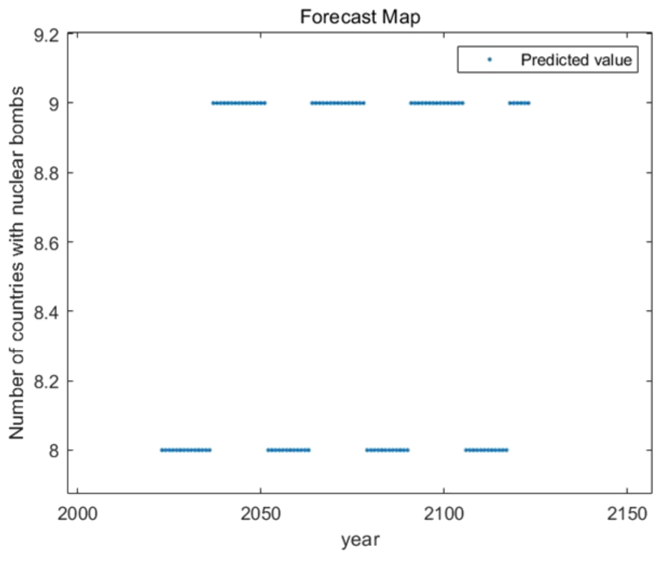

Figure 3: Projected number of nuclear-armed states over the next 100 years

Our projections show that the number of nuclear-armed states will continue to hover between eight and nine over the next 100 years, and that in 100 years the number of nuclear-armed states is most likely to be eight, which is in line with the current internationally agreed and more appropriate number of nuclear-armed states.

5.3 Results from the LSTM method

For the prediction of the number of nuclear bombs in the world and by country, we use the LSTM prediction method for the prediction of the number of nuclear bombs, and the following is a demonstration of the error in the prediction process and a demonstration of the prediction for the world and by country. This is shown in Figure 4.

ISSN 2616-5805 Vol. 4, Issue 1: 44-51, DOI: 10.25236/AJMS.2023.040107

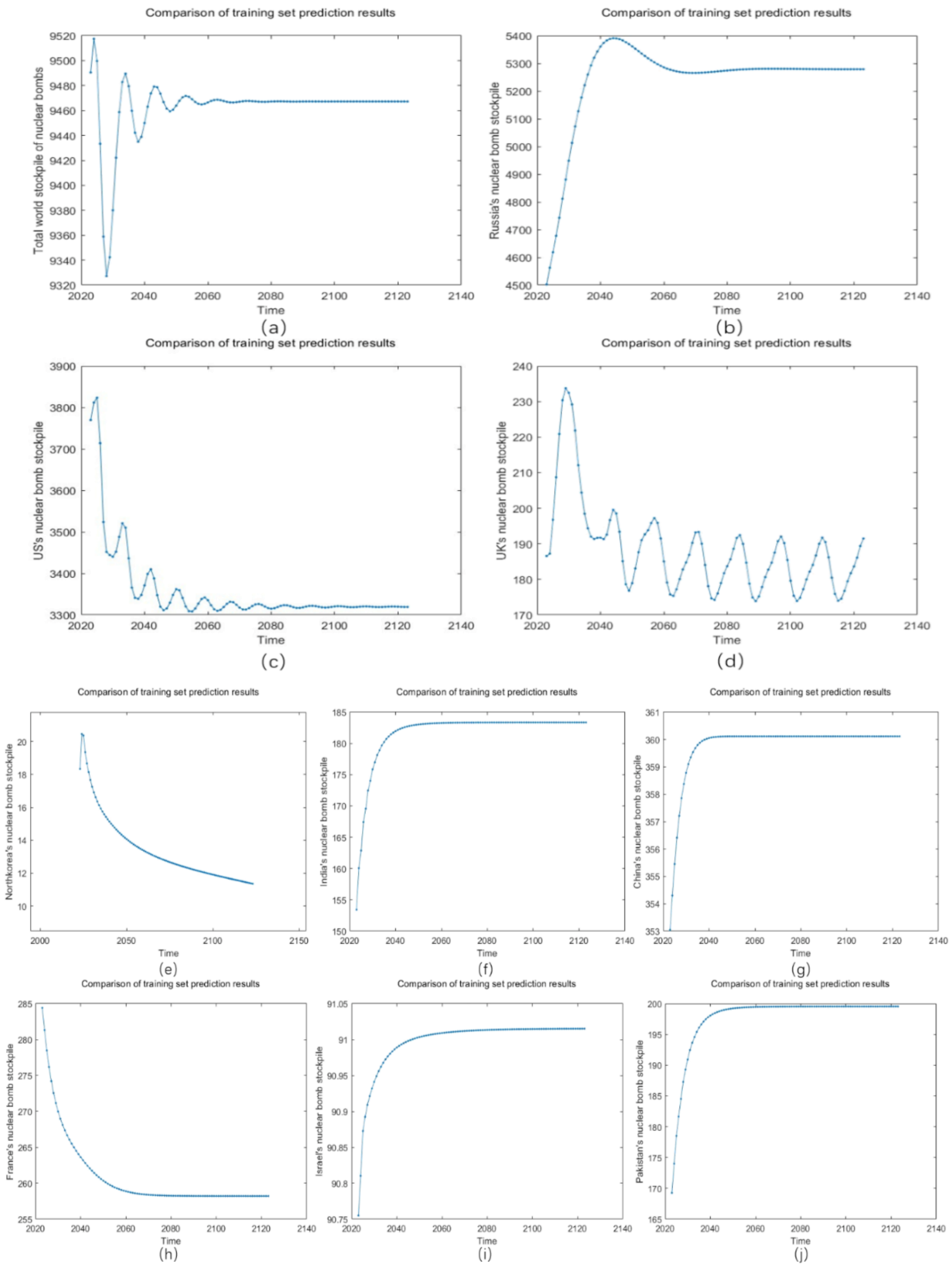

Figure 4: Graph of the change in the number of nuclear bombs in the world and in each country

In the chart above, the a-j charts represent the total number of nuclear bombs worldwide, the number of Russian bombs, the number of US bombs, the number of British bombs, the number of North Korean bombs, the number of Indian bombs, the number of Chinese bombs, the number of French bombs, the number of Israeli bombs and the number of Pakistani bombs.

Figure 5 illustrates how the error changes as the number of iterations in the machine learning process increases.

ISSN 2616-5805 Vol. 4, Issue 1: 44-51, DOI: 10.25236/AJMS.2023.040107

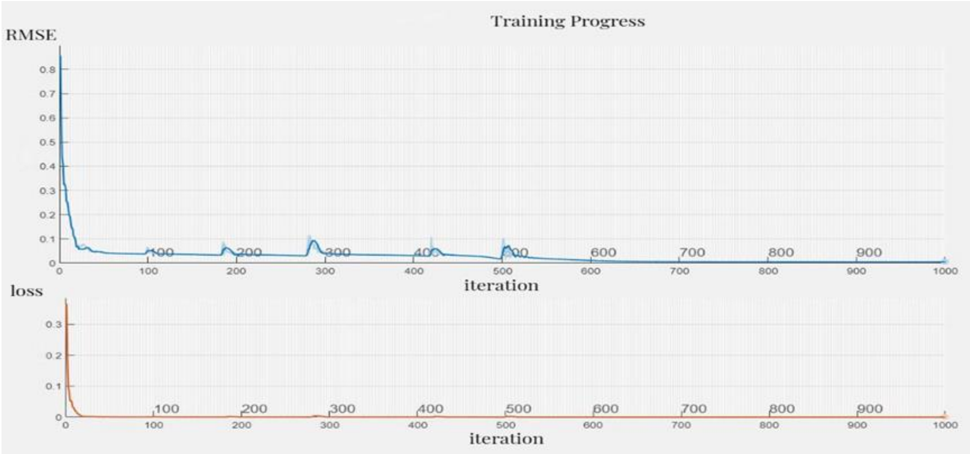

Figure 5: Plot of error variation during training

As can be seen from the above graph, the RMSE and loss function values are decreasing as the number of iterations increases, and when the number of iterations reaches 1000, the RMSE reaches 0.05 and the loss function value approaches 0. This shows the accuracy of our model[6].

We have aggregated the number of nuclear weapons in the world and by country in 2123 to get the following figure 6.

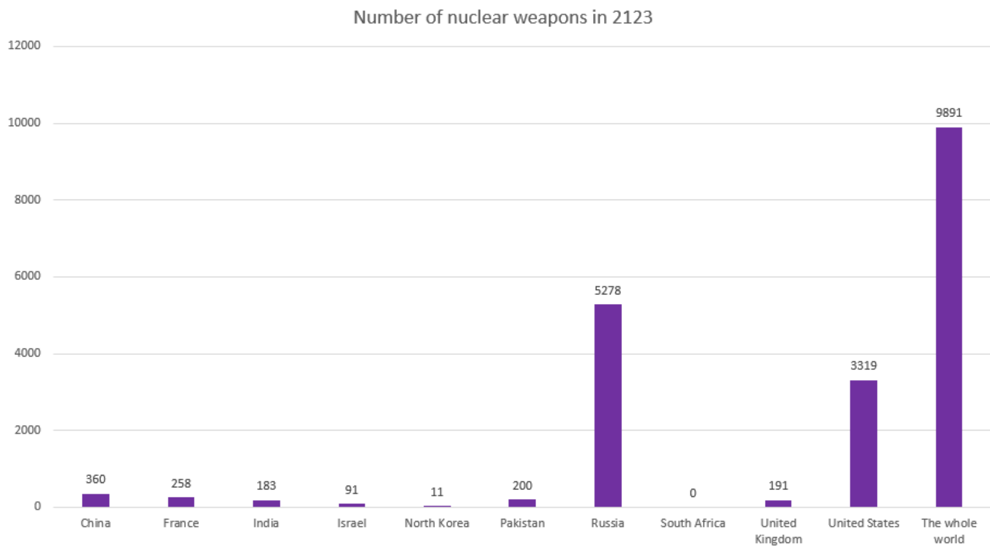

Figure 6: Map of the number of nuclear weapons in the world and by country, 2123

6. Conclusions

The above analysis shows that the global stockpile of nuclear bombs has remained at just under 10,000 since 2016, and the trend shows that the number of nuclear bombs will remain low in 100 years, which is very important for world peace. Nuclear weapons should be banned and no country is prepared to deal with the catastrophe that would result from a nuclear explosion, especially the radioactive fallout from the use of nuclear weapons, which would spread worldwide.

We therefore call on all countries, especially powerful countries such as the United States and Russia, to stick to the bottom line and to refrain from the active use of nuclear weapons and from nuclear confrontation, which, if it were to take place, would only end in defeat for both sides. The remaining nuclear weapons today must be closely guarded and must not be allowed to fall into the hands of terrorists; the loss of even just one would be an eternal pain for the people of the world.

Firmly prohibit the development of nuclear weapons by non-nuclear states, refrain from assisting non-nuclear states in the manufacture of nuclear weapons, and keep the number of states possessing nuclear weapons at a consistently low level. Resolutely prohibit nuclear states from building new nuclear

ISSN 2616-5805 Vol. 4, Issue 1: 44-51, DOI: 10.25236/AJMS.2023.040107

weapons to develop new types of nuclear weapons. Only then will we be responsible for humanity and for ourselves!

References

[1] Meng Xiaojing, Chen Xin, Chen Jiajing, Yang Honggang. Application of combined empowerment-TOPSIS in resilience assessment of urban areas under flooding [J/OL]. Journal of Safety and Environment, 2023: 1-9.

[2] Fang X., Yang N. Construction of nonlinear error correction learning models based on LSTM networks [J/OL]. Statistics and Decision Making, 2022(24): 5-10.

[3] Sun Xiaoming, Yu Wujing, Ren Ruobing, Xiong Wang, Wang Yalan. A study of patent inventor name disambiguation based on decision tree algorithm [J/OL]. Science and Management, 2022: 1-20.

[4] Shi Dawei, Shen Yang, Ma Chenchen, Dong Jingming, Yan Jiaren. Research on short-time intense precipitation forecasting in different regions of Jiangsu Province based on decision tree algorithm [J]. Meteorological Science, 2022, 42(05): 631-637.

[5] Li Zhen. Vulnerability evaluation and barrier degree analysis of marine economic system in Shandong Province based on entropy power TOPSIS model [J/OL]. Ocean Development and Management, 2023: 1-10.

[6] Wang Qing. Research on the Prevention of Nuclear Terrorism Crimes [D]. Jilin University, 2011.