Research on component analysis and identification based on clustering and prediction model

Tonghao Wang*, Jian Xuan, Nanxiang He

College of Chemistry and Chemical Engineering, Xiamen University, Xiamen, China

*Corresponding author: crowndaisy666@163.com

Abstract. To better study the ancient relics of glass types and content of chemical composition, this article first USES no sequence grey prediction model to predict and distribution rules of cultural relics of the weathering before chemical component content, and then USES the K Means clustering will classify glass, finally set up the BP neural network model, based on the experimental data respectively set up the training set and validation set, Predicting the type of glass artifacts. It solves the problem of composition analysis and identification of glass cultural relics.

Key words: Glass relics gray, prediction model, K-means clustering, BP neural network.

1. Introduction

The ancient glass production method was introduced to our country through the silk road, after absorbing the foreign introduced technology, China learned to use the local raw material to homemade glass. The main component of glass is SiO2, SiO2 melting point can reach several thousand degrees Celsius. Therefore, in order to reduce the energy consumption when firing glass, it is usually necessary to add flux to reduce the melting point. In ancient times, the flux used by people was generally lead ore, plant ash and saltpeter [1-2]. The addition of different fluxes in firing glass will lead to different glass properties. According to the different proportion of glass elements, ancient glass is mainly divided into two kinds: lead-barium glass and high-potassium glass. Lead ore (containing a large amount of BaO and PbO) is generally used as a flux in the firing of lead-barium glass, so that the glass containing Ba and Pb elements are relatively high. High potassium glass is generally used in the firing of plant ash containing high K element flux, so this method fired out of the glass K element content is high. When firing glass, it is necessary to add the stabilizer limestone (CaCO3), so that the strength of the fired glass will be high. Ancient glass buried for a long time, so the unearthed cultural relics often have been weathered. Weathering can cause changes in the elemental composition of glass, making it difficult to identify ancient glass. Buried for thousands of years, ancient glass has more or less weathered, that is, every piece of ancient glass has weathered and unweathered parts, so the weathering degree of glass devices is the standard to define whether a piece of glass is weathered. Nowadays, there are 58 newly unearthed ancient glass devices. Scientific chemical analysis and classification of their components are of great significance for the study of glass cultural relics. Therefore, this paper conducts a classification study on glass cultural relics based on the component analysis and identification research of clustering and prediction models.

2. The establishment of grey prediction model

In order to predict the chemical composition of glass cultural relics before weathering, this paper uses time-free gray prediction to predict the chemical composition content of glass cultural relics before weathering.

Grey prediction is a method that combines known information with unknown information to make a prediction. It is usually a sequential gray process to predict, but the method of repeated ranking can be used to get gray prediction which has nothing to do with the order, only related to the size. The specific operation is to make time series grey prediction of a certain unweathered chemical composition in the table according to the GM(1,1) model normally, then sort the predicted results in ascending order, and then sort the chemical composition of the weathered cultural relics to be predicted[1,2].

In the chemical composition of glass artefacts, the sample sites of high potassium type were predicted well. However, there are too many sampling points of lead and barium type, which leads to poor results of time-free gray prediction of some chemical components, so these chemical components are treated as chemical components with no significant difference[3,4].

For the chemical components with no significant difference in content before and after weathering, it can be considered that there is no obvious change in the content of these components before and after weathering[5]. Therefore, the distribution method is used to solve the content prediction. The proportion of the main components before weathering can be predicted by the proportion of the components in the total composition accumulation of the cultural relics[6].

Taking the above problems into consideration, it is necessary to classify weathered cultural relics and unweathered cultural relics. This paper takes type as the first classification index and color or pattern as the second classification index[7,8].

$$\alpha_i = \frac{\omega_i}{\sum_{i=1}^{14} \omega_i} \tag{1}$$

Where,ω represents the content of chemical components in cultural relics, and i represents each chemical component.

There will be multiple references of unweathered cultural relics in the same group.[9] By summing the square distance between all kinds of insignificant components of weathered cultural relics and the same components of unweathered cultural relics, a more accurate and appropriate index of distribution coefficient can be obtained. The operation formula is as follows:

$$x^2 = \sum_{\lambda=1}^{n} (Eff_l - UnEff_l) \tag{2}$$

Where: The lower corner l is marked with all components with non-significant differences, EFF is the proportion of this component in wathered cultural relics, and is the proportion of chemical components in unweathered cultural relics.

The distribution coefficient of the unweathered cultural relic with the smallest sum of squared distance is chosen as the calculation basis and obtained. The prediction is made according to the fixed allocation equation, which is:

$$\omega_p = \sum_{i=1}^{14} w_i \cdot \alpha_i \tag{3}$$

ϖ_p represents the proportion of this component of the cultural relics before predicted weathering, and its value is obtained by the product of the sum of the indicators of weathered cultural relics and the distribution coefficient of the unweathered cultural relics with the smallest sum of squared distance[10].

This hybrid application of grey prediction and fixed distribution equation captures the principal contradiction and simplifies the model. It also effectively reduces the error caused by only using fixed distribution equation in forecasting results.The chemical composition prediction table of high-potassium glass before weathering is shown in Table 1, and the chemical composition prediction table of 6 lead-barium glass before weathering is shown in Table 2.

Table 1. Chemical Composition Prediction Table of High potassium glass before weathering (unit %)

	7	9	10	12	22	27
SiO_2	68.47	68.58	68.68	68.78	68.89	69.99
Na_2O	0	0	0	0	0	0
K_2O	8.61	8.5	8.4	8.29	8.19	8.09
CaO	5.42	5.45	5.49	5.52	5.55	5.58
MgO	1.12	1.12	1.13	1.13	1.13	1.14
Al_2O_3	6.8	6.79	6.78	6.77	6.76	6.75
Fe_2O_3	1.76	1.73	1.7	1.67	1.64	1.61
CuO	3.95	3.96	2.94	3.95	2.95	2.92
PbO	0	0	0.83	0	0.83	0.82
BaO	0	0	1.43	0	1.44	1.42
P_2O_5	1.2	1.2	0.68	1.19	0.68	0.68
SrO	0	0	0.05	0	0.05	0.05
SnO_2	0	0	0	0	0	0
SO_2	0.4	0.4	0	0.4	0	0

Table 2. Chemical Composition prediction table of 6 pieces of lead and barium glass before weathering (unit %)

	02	11	38	41	49	56
SiO_2	58.90	59.84	61.77	62.76	64.10	66.52
Na_2O	4.79	2.58	0	0	0	0
K_2O	0.3	0.11	0.23	0	0.69	0.21
CaO	1.16	1.11	1.02	0.97	0.92	0.83
MgO	0.63	0.72	0	0	0.73	0
Al_2O_3	3.14	4.85	2.68	1.4	3.99	2.51
Fe_2O_3	0	0	0	0.17	2.36	0
CuO	0.67	0.51	2.97	0.16	0	2.78
PbO	17.90	16.87	14.99	14.12	13.05	11.36
BaO	9.48	10.63	10.19	5.53	10.02	9.54
P_2O_5	0.1	0	1.44	0.41	0.69	1.35
SrO	0.87	0.22	0.31	0	0.41	0.29
SnO_2	0	0	0	0	0.41	0
SO_2	0	0	3.61	0	0	3.38

3. The establishment and solution of K-means clustering model

3.1. The establishment of 1k-means clustering model

In order to find out the division rule, this paper firstly compares the characteristic values of each corresponding chemical composition of two kinds of glass, and obtains the chemical composition of two kinds of glass. When the difference is large enough, the chemical composition can be considered as a criterion for judging the glass class. In addition, SPSS software is used to cluster all glass cultural relics in Annex II into two categories by K-means clustering method, and explain the differences between the two categories according to different clustering centers, so as to find out the division rules Figure 1[11,12].

According to the chemical composition, the glass with high potassium, lead and barium were further divided into two types. Firstly, the samples of a certain glass category are systematically clustered. The specific algorithm principle is as follows:

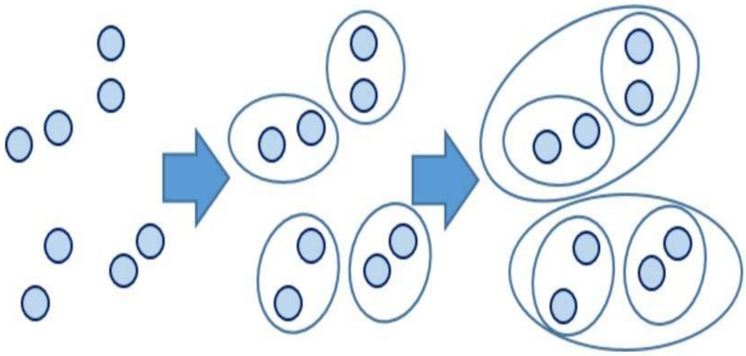

Figure 1. Schematic diagram of the system clustering algorithm

The specific algorithm flow is as follows:

The first step is to calculate the distance between each pair of objects

In the second step, the two objects with the smallest distance are merged into a new object

Third, recalculate the distance between all pairwise objects

In the fourth step, repeat the above two steps to merge all objects into one object and end

Then, the Elbow rule is used to estimate the optimal number of clusters K by making the line diagram of the aggregation coefficient.

The calculation formula of polymerization coefficient is:

$$J=\sum_{k=1}^{K}\sum_{i\in C_k}\left|x_i-\mu_k\right|^2 \tag{4}$$

Where, C_k is denoted by the K class, the center of this class is denoted by μ_k, and each sample is denoted by x_i

In the line diagram of the aggregation coefficient, the abscissa is the cluster number K, and the ordinate is the aggregation coefficient J

Then, according to the K value, k-means clustering is carried out on the glass samples of this kind, and the result of subclass division and the clustering center of each subclass are obtained.

The algorithm flow of k-means is shown in the figure2 below:

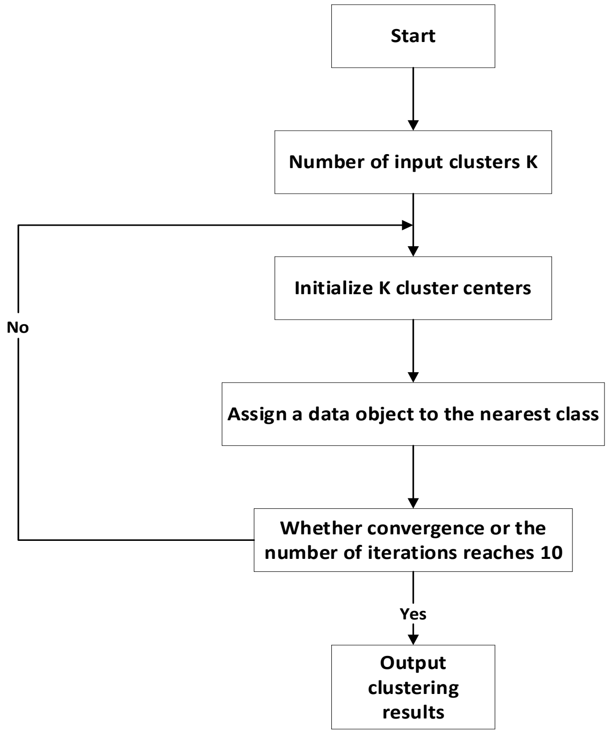

Figure 2. k-means algorithm flow chart

3.2. Solution of K-means clustering model

In this paper, Excel was used to calculate the average value of each chemical composition of lead-barium glass and high-potassium glass respectively, and the average value of each corresponding chemical composition of the two kinds of glass was compared. The average value and mutual ratio of each chemical composition of the two kinds of glass are shown in the following table3:

Table 3. Average proportion (unit %) and ratio table of chemical composition of two kinds of glass

chemical constitution	Mean of lead and barium	Mean of high potassium	Lead barium/high potassium	High potassium/lead barium
SiO_2	38.87551	76.64389	0.507223	1.971521
Na_2O	0.90449	0.463333	1.952136	0.512259
K_2O	0.173469	6.401667	0.027098	36.90373
CaO	2.05	3.845	0.53316	1.87561
MgO	0.64551	0.785	0.822306	1.216092
Al_2O_3	3.667551	5.056667	0.72529	1.378758
Fe_2O_3	0.655918	1.376111	0.476646	2.097991
CuO	1.879592	2.155556	0.871976	1.146821
PbO	33.34918	0.274444	121.5152	0.008229
BaO	10.49041	0.398889	26.29907	0.038024
P_2O_5	3.292653	1.028333	3.201932	0.312311
SrO	0.347959	0.027778	12.52653	0.079831
SnO_2	0.058163	0.131111	0.443618	2.254191
SO_2	0.799592	0.067778	11.79726	0.084765

As can be seen from the table, SiO2 is the principal component of both types of glass, but the content of lead-barium SiO2 is significantly smaller than that of high-potassium glass, and the content of K2O of high-potassium glass is 36 times higher than that of lead-barium glass. On the other hand, the content of BaO, SrO and SO2 in lead-barium glass is more than 10 times higher than that in high-potassium glass, and the content of PbO is 121 times higher. In view of such a big difference, the content data of SiO2, K2O, BaO, SrO, PbO and SO2 of glass can be used as the classification standard of lead-barium and high-potassium glass to a certain extent.

On the basis of high potassium and lead-barium glasses, further subdivision was carried out by combining different chemical compositions. Firstly, the data of lead-barium glasses were substituted into SPSS software for systematic cluster analysis. According to the obtained polymerization coefficient (see Appendix I for specific results), the line chart of polymerization coefficient was made as Figure 3.

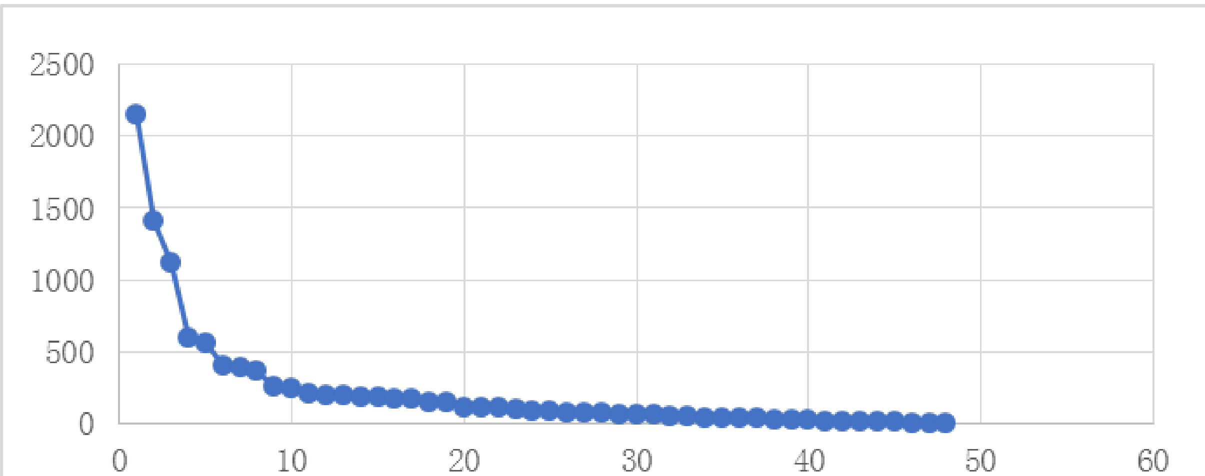

Figure 3. Line chart of polymerization coefficient of lead and barium

It can be seen that the slope of the broken line from the fourth point sharply tends to "0", indicating that from the fourth point onwards, the data can be summarized into the same category. Therefore, the number of clusters is determined as 4 in this paper. Then, k-means clustering is carried out on specific data to obtain clustering results. The results of four types of clustering centers are shown in the following table4:

Table 4. Table of different cluster centers (unit %) corresponding to four kinds of lead-barium glass

The final cluster center	1	2	3	4
SiO_2	19.35167	58.1135	12.065	30.43474
Na_2O	0	1.975	0	0.253684
K_2O	0.053333	0.197	0.1	0.202105
CaO	2.756667	1.0435	2.28	2.837895
MgO	0.788333	0.712	0	0.666316
Al_2O_3	2.251667	5.109	1.0825	3.141579
Fe_2O_3	0.228333	0.6555	0	0.929474
CuO	1.525	1.1885	6.93	1.655789
PbO	59.39167	19.3475	30.145	40.53842
BaO	4.706667	8.063	32.3875	10.26211
P_2O_5	5.008333	0.867	5.08	4.927895
SrO	0.655	0.2225	0.4925	0.352632
SnO_2	0	0.077	0	0.068947
SO_2	0	0.183	8.88	0

By analyzing the above table and the clustering result table of high-potassium glasses, it can be seen that the contents of K2O and SiO2 of high-potassium glasses in cluster 1 are at a medium level. The content of K2O in cluster 2 is very low, but the content of SiO2 is very high. The high potassium glass of cluster 3 has very high content of K2O and low content of SiO2. It is obvious that the content of SiO2 and K2O is the criterion for the classification of these three categories.

4. Establishment and solution of B-P neural network prediction model

4.1. Establishment of the model

In this paper, BP neural network model is used for prediction. The characteristics of high efficiency and high accuracy are the reasons why this model is selected in this paper [7]. The nature of BP neural network is a self-training model. An algorithm that learns certain rules from sample data and makes data prediction based on the learned rules. BP neural network algorithm is mainly composed of two steps: forward propagation and back propagation, as shown in the figure 4.

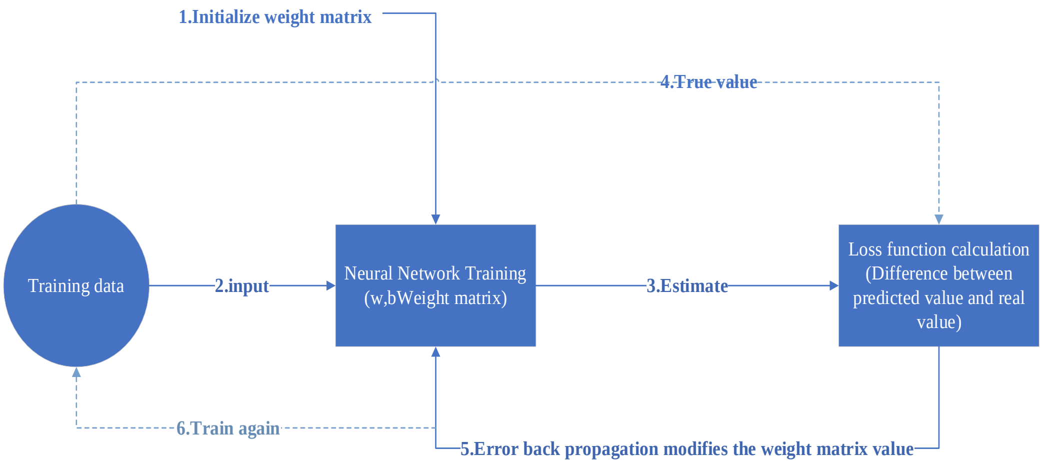

Figure. 4 Flow chart of BP neural network

BP neural network algorithm mainly consists of two steps: forward propagation and back propagation. In the forward propagation process, the input signal first enters the hidden layer, and then the hidden layer transmits the signal to the output layer to generate the output signal. After that, if the output deviates from the expected output, it will enter into the back propagation, and according to the error calculated in the forward propagation, the signal propagation method with smaller error will be backward calculated. In this way, the data is constantly trained to make the network more and more close to the real situation, and a neural network model can have a better prediction effect. The principle is shown in the figure5.

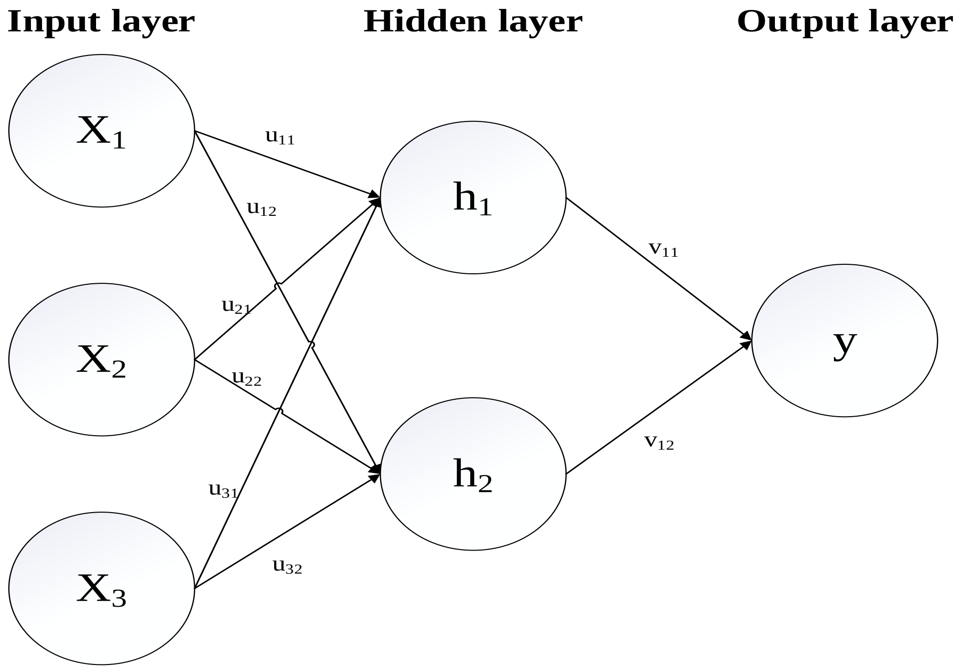

Figure 5. Schematic diagram of BP neural network

Then the data of lead and barium and high potassium were classified respectively, and the neural network model was trained with appropriate data. The experimental set was 70%, the validation set was 15%, and the training set was 15%.

Finally, the established model is used to predict the data given in Table 3 and get its category.

4.2. Solution of the model

In this paper, 60 pieces of complete information of glass cultural relics obtained by experiments are firstly used to establish a training set. Then, the complete data of seven known categories, 16,19,20,21,22,23 unweathered points and 24, are taken as the verification set and brought into the model for verification. The verification results are shown in the following table5:

Table 5. verification results

Cultural relics number	The actual type	predict the outcome	Are consistent
16	K	K	√
19	Pb Ba	Pb Ba	√
20	Pb Ba	Pb Ba	√
21	K	K	√
22	K	K	√
23	Pb Ba	Pb Ba	√
24	Pb Ba	Pb Ba	√

The prediction accuracy of the 7 data as the validation set is as high as 100%. Although the verification sample size specification is general, it is enough to show that the mathematical model has a good accuracy.

5. Conclusion

Firstly, the time-free grey prediction model and allocation rules are used to predict the chemical composition content of cultural relics before weathering, and then K-means clustering is used to classify glass cultural relics. Finally, a BP neural network model is built, and a training set and a validation set are established according to the experimental data to predict the type of glass cultural relics. The results show that the composition analysis and identification of glass cultural relics established in this paper are consistent with the actual situation and suitable for the identification of glass cultural relics.

References

[1] Wen Rui, Zhao Zhiqiang, Ma Jian, Wang Jianxin. Composition analysis of glass beads unearthed from shirenzigou site group, barkun, xinjiang [J]. Spectroscopy and spectral analysis,2016,36(09):2961-2965.

[2] Hu Zhizhong, Li Pei, Jiang Lupan, Wang Tongyang, Du Gu, Yang Bo. Ancient glass material LA - icp-ms component analysis and source study [J]. Journal of testing, 2020, 33 (4) 6:505-514. The DOI: 10.15898 / j.carol carroll nki. 11-2131 / td., 201909210134.

[3] Fu Qiang, Zhao Hongxia, Dong Junqing, Li Qinghui, Hu Yongqing. Nondestructive analysis of silicate products unearthed in baofeng and xinzheng, henan province [J]. Spectroscopy and spectral analysis,2014,34(01):257-262.

[4] Fuxi Gan, Hongxia Zhao, Qinghui Li, Ling Li, Huansheng Cheng. Science and technology analysis and research of Warring States glass products unearthed in Hubei Province [J]. Jianghan Archaeology,2010(02):108-116+151+0.

[5] Application conditions of chi-square test [J].Journal of Clinical Hepatobiliary diseases, 2012,38(06):1292.

[6] ZHAO Z Q. Detection and analysis of glass products unearthed from the Warring States Cemetery of Liyemai tea in Hunan Province [J]. Hunan Archaeology Journal,2020(00):288-301.

[7] WANG Kun. Review of Characteristics and Improvement Methods of Neural Network [J]. Science and Technology Square,2011(07):227-232.

[8] Xu Hui, SUN Hongtu. Research and Simulation of LiDAR Localization Method for Indoor Mobile Robot [J]. Mechanical design, 2022, 33 (5) : 6, 91-95. The DOI: 10.13841 / j.carol carroll nki JXSJ. 2022.05.022

[9] Yang Wei. Bearing-only Passive Localization and Optimization of Multi-station Based on Bayesian Estimation [D]. Nanjing University of Science and Technology,2015.

[10] Lu Y. Research on Building energy consumption prediction Method based on cyclic feature and deep Learning algorithm [J]. Industrial Heating,202,51(05):31-35.

[11] Zhang Kun, Shen Zhong, Wang Haifeng, Li Zhuang, Gao Qian, Li Hanwen. Research on bearing-only passive positioning technology of maritime reconnaissance ship [J]. Ship science and technology,2018,40(02):19-21.

[12] Fan Jianwen, Lei Chuang. Research on Relative Navigation Technology of Aircraft Formation [J]. Modern Navigation,2016,7(03):161-165.