2020 International Conference on Computer, Network, Communication and Information Systems (CNCI 2020)

*The Hypercube Analysis Model for The Term of Office to United States Naval Vessels**

Cao Wen[1,a], Dai Haoran[1,b], Tong Xiaochong[1,c], Zhang Yinbao[1,d], Peng Feilin[1,e], Zhang Yong[1,f]

[1]*School of Geo-Science and Technology, Zhengzhou University, Henan, Zhengzhou*
a. zzdx_edifier@163.com, b. 2551935684@qq.com,
c. txchr@aliyun.com, d. zhangyinbao2005@sina.com, e. 582885655@qq.com, f.zy765@163.com

Keywords: Big data, hypercube, term of office, normal distribution.

Abstract: With the advent of the big data era, how to execute data mining and information association automatically and intelligently from valuable information point which we are interested in always is difficult and hot topic of research in the field of intelligence by using "big data thinking". Aiming at the research status of qualitative analysis to military intelligence lacking data support at present, the hypercube analysis model for the term of office to United States naval vessels was proposed by taking the United States Naval force deployment and command system as the research object. Adaptive gross error elimination and abnormal data analysis trigger are performed through critical statistical JB value detection of normal distribution; the normal distribution model for term of office is built by using detection of the optimal statistical JB value; according to the analysis model of optimal normal distribution the benchmark term of office and referenced adjustment interval are obtained. According to the simulation and experimental results, the analysis model not only can automatically obtain the benchmark term of office for United States naval vessels and referenced adjustment interval from cyberspace data, which shows the model has higher correctness and rationality, but also it can provide self-learning ability for gross error elimination and data analysis by the trigger mechanism of abnormal data analysis, which shows the model has better automaticity and robustness, at the same time, it can also provide support of candidate data for related information analysis and mining, which promotes the expansibility of analysis model.

1. Introduction

With the advent of the era of big data, data-driven paradigm thinking has become the consensus of the scientific community. Intelligence science is a discipline that is good at intelligence collection, intelligence processing, intelligence analysis, intelligence service research and practice, which has an inseparable relationship with data. Big data technology, as an opportunity to promote its development, is bound to bring more difficulties and challenges to the research of intelligence[1]. As an important form of intelligence science, military intelligence is usually associated and synthesized by multiple independent intelligence sources with multi-source, time-varying and heterogeneous characteristics, which are usually hidden in vast data[2-3], Therefore, how to use "big data thinking" combined with artificial intelligence, big data analysis and other technologies for automatic and intelligent acquisition, analysis and mining is one of the hot spots and difficulties in military intelligence research.

According to the 2019 global firepower ranking released by the world military strength ranking website "global firepower", the United States is the world's largest military power, with a military strength index of 0.0615. Therefore, its military intelligence such as weapons and equipment, command system, force deployment, etc. is the key points and difficulties of other countries studies in the world. Military power is the most intuitive embodiment of a country's comprehensive national strength, and weapons and equipment are important factors affecting a national military power, whose importance is mainly reflected in its combat capability and effectiveness. No matter how the combat form and idea develop, cooperation and command are the basis to maximize the actual combat capability. Therefore, the military force deployment in a country is also the key and difficult point of military intelligence research. Force deployment is a micro control for the number of personnel in their office and positions according to the combat tasks and forms of the military combat unit. The operation mechanisms of office and personnel comprehensive characteristics for its commanding officer are the fighting soul of the military combat unit. Therefore, obtaining real-time, overall and comprehensive information or intelligence of force deployment will provide important support and guarantee for later intelligence integration and decision-making.

At present, most of the scholars at home and abroad have studied the construction of aircraft carrier system[4-7] and the training mechanism of commanding officers of aircraft carrier[8-9] in United States from a qualitative perspective, and the research results lack the and verification of some period data. For the force deployment of the United States Navy, only 2018 aviation nuclear office newsletter[10] published on the official website of the United States Navy gives that the term of office of commanding officers for nuclear powered aircraft carrier is 30 months, and the term of office for the execute officers is 24 months, as well as the term of office for the deep draft vessels' commanding officers is 18 months. Among them, the term of commanding officer of the nuclear powered aircraft carrier will be adjusted between 24 and 48 months according to the international situation, overall personnel planning and personal performance, and other term information and referenced adjustment interval are not mentioned. Therefore, this paper takes the United States Naval force deployment and command system as the research object. Firstly, a multi-dimensional hypercube model with attributes is constructed to describe the spatiotemporal laws for the term of office in the U.S. Navy, then after the temporal topological analysis of the bar time series in hypercube is carried out, we use the critical statistical JB value detection of normal distribution to eliminate the adaptive gross errors and detect the abnormal data according to the internal order and laws for term data of office; secondly, the normal distribution model for term of office is built by using detection of the optimal statistical JB value; finally according to the analysis model of optimal normal distribution the benchmark term of office and referenced adjustment interval are obtained. Through experimental verification for the term data of the U.S. vessels' commanding officers, firstly, the model can extract the term information of office in the United States vessels from both quantitative and qualitative perspectives, which show the model has high correctness and scientific nature; secondly it can provide self-learning ability for gross error elimination and data analysis by the trigger mechanism of abnormal data analysis, which shows the model has better automaticity and robustness; finally, it can provide event association and preselected data sets of sensitive object for related information analysis and mining, which has better scalability.

2. The Hypercube Analysis Model For The Term Of Office To United States Naval Vessels

As the strongest military power in the world, the United States has carried out systematic planning and refined rules and constraints in the form of combat, weapons and equipment, talent training and command system, which effectively improves the military power and national strength of the United States. Therefore, the systematic and legal design, management and operation model of the United States Army is also contained in the ocean of multi-source, massive and heterogeneous spatiotemporal data. According to the laws and regulations of the United States Army, this paper describes the temporal and spatial evolution of the United States Army's combat form, weapons and equipment, force deployment, etc. by building a multi-dimensional hypercube model with attributes, and then excavates the temporal and spatial evolution of force deployment through temporal topology operation. Finally, according to the temporal normality for term of office, an analysis model for the term of office based on normal distribution is built to automatically obtain the relevant conclusions for term of office.

2.1. Hypercube Model

Data analysis of spatiotemporal big data usually focuses on data clustering and other algorithms, and lacks the description and analysis means of spatiotemporal evolution from the perspective of data model. Therefore, based on the hypercube interconnection network structure in the field of parallel computing, this paper constructs a multi-dimensional hypercube model with attributes to describe and analyze spatiotemporal data of the United States naval force deployment with multi-source, heterogeneous and time-varying characteristics.

Taking three-dimensional hypercube as an example, in order to describe the spatiotemporal evolution of spatiotemporal big data, an three-dimensional cube model with attribute[11] is built by setting one of the dimensions as time dimension and the other two dimensions as any two attribute dimensions $\mathbf{X}=\{S_1(t),P_1(t),T_1(t_v,t_d),A_1\}$ and $\mathbf{Y}=\{S_2(t),P_2(t),T_2(t_v,t_d),A_2\}$ in spatiotemporal big data, as shown in Figure 1。 Where $S(t)$ represents the set of attribute characteristics that change with time, $P(t)$ represents the set of other attributes characteristics for object associated an attribute dimension with time changing, $T(t_v,t_d)$ represents the temporal nature of starting change for the attribute state, t_v and t_d are effective time and transaction time respectively, A refers to the behavior operation of attribute characteristics. We only need to judge whether $P_1(t)\cap P_2(t)$ is an empty set to get whether any two attribute dimensions are related, which can achieve the effect of data dimensionality reduction.

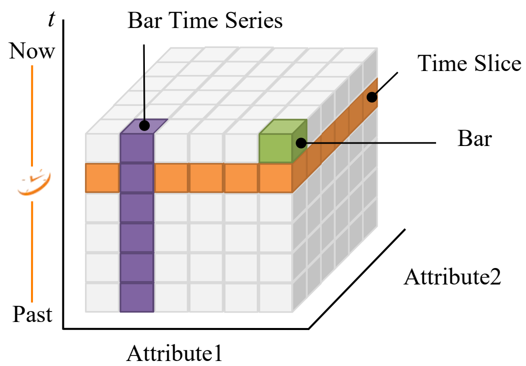


Figure 1: Three-Dimension cube structure.

In the three-dimensional cube model for term of personnel position, **X** and **Y** dimensions are combat troops and office of troops respectively, and the bar represents the status of a certain post of a combat troop at a certain time, such as personnel, duties, etc. Time slice represents the state distribution of each post for all combat troops at a certain time; The bar time series represents the state distribution for an office of a troops along the time axis, etc. From here we see that the effect of data mining can be achieved through time topological association and data clustering in three-dimensional cube. For example, when $P_1(t) \cap P_2(t)$ of combat troops and office of troops is personnel, the personnel topological relationship graph and other information is obtained by clustering other attribute dimensions on the time slice section, while the term statistics of each office of troops can be obtained by calculating the temporal topological relationship of different granularity on the bar time series.

Based on the personnel data of the U.S. Navy vessels, this paper uses data statistics of the bar time series in the hypercube model to migrate the hypercube to the three-dimensional cube model with X-axis as the term of office, Y-axis as the post of a vessel and z-axis as the number of people, as shown in Figure 2. In the XZ plane, it can be seen that the statistical trend for the term of an office in a combat troops shows the characteristics of approximate normal distribution, which is consistent with the principle that the United Sates Army is established by law, therefore, the approximate normal distribution can be used to describe its probability distribution, and then a data statistical analysis model can be designed to describe and analyze it.

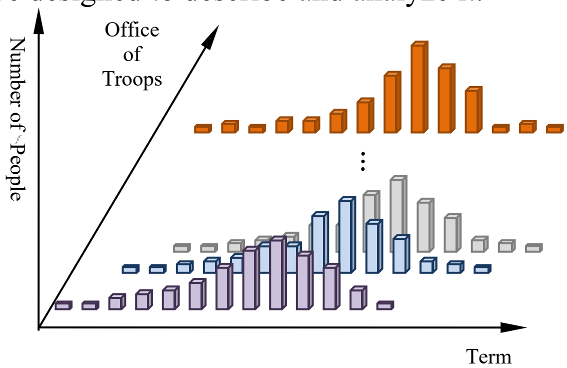


Figure 2: Three dimension cube structure for term of office.

2.2. Normal Distribution Model

Normal distribution, also known as Gaussian distribution, is the most important and extensive probability distribution in probability theory[12]. Its probability density function is shown in Formula 1:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)} \quad (1)$$

Where μ and σ are the mean value and standard deviation of sample data respectively.

For the term data of office for naval vessels, the mean value μ can be understood as the legal benchmark term, and the standard deviation σ reflects the discrete influence of various factors on the term, which is an important factor determining the adjustment range of reference term.

2.3. Detection of Critical Statistical JB Value Based on Adaptive Threshold

In the daily operation of the United States naval talent training and command system, unexpected factors such as accidental death, ability defects and violations of discipline, etc. will produce gross error or abnormal data and affect its data distribution trend. These data are randomly distributed on the whole time axis, so how to design an effective gross error elimination method is the key to the analysis model for term of office. Through the analysis of the sample data, the abnormal data on the left side of the term mean value will have a greater impact on the overall distribution trend. Therefore, this paper designs a critical statistical JB value detection method based on adaptive threshold, which can automatically detect the gross error and abnormal data in the sample data.

1) Statistical JB value

In the application of data statistics and analysis, the normal test methods mainly include normal Q-Q chart test, Jarque-Bera test (JB), Kolmogorov-Smirnov test (KS), Anderson-Darling test (AD), etc. Considering the characteristics of the term data of the U.S. naval vessels' commanding officers and the simple calculation of the JB test method[13], this paper uses the JB test method to automatically detect the gross error and abnormal data in the sample data.

Suppose a set of data $\boldsymbol{X}=\{x_1,x_2,\cdots,x_n\}$, where is the total number of sample data, and $\bar{\mu}$ and $\bar{\sigma}$ are the mean value and standard deviation respectively, then the skewness β_S and kurtosis β_K are defined as:

$$\beta_S=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i-\bar{\mu}}{\bar{\sigma}}\right)^3 \tag{2}$$

$$\beta_K=\frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i-\bar{\mu}}{\bar{\sigma}}\right)^4 \tag{3}$$

β_S refers to the asymmetry of the probability distribution for sample data, that is, the dispersion of two sides for the $\bar{\mu}$; β_K refers to the steepness of the probability density distribution of sample data, that is, the cusp of the peak at the $\bar{\mu}$. For the standard normal distribution, the β_S should be 0 and the β_K should be 3.

Jarque-Bera test method is a statistical JB value test method designed based on the probability distribution of data conforming to normal distribution[14], which is designed by two variables of skewness and kurtosis based on the central limit theorem. The definition is as follows:

$$JB=\frac{n}{6}\left[{\beta_S}^2+\frac{(\beta_K-3)^2}{4}\right] \tag{4}$$

After Jarque and Bera, a number of scholars[15-16] have given the critical statistical values $\xi_{JB}(n,\alpha)$ under different quantiles α by Monte-Carlo simulation (as shown in Table 1), that is, when $JB\le\xi_{JB}(n,\alpha)$ is satisfied, the sample data X obey the normal distribution.

Table 1: Critical JB values under different sample numbers and quintiles.

Number of Samples	Statistical JB Value				
	α=0.01	α=0.02	α=0.05	α=0.10	α=0.20
10	5.738	4.274	2.535	1.618	1.126
20	9.458	5.683	3.768	2.335	1.553
50	12.331	8.721	5.004	3.192	2.122
100	12.296	9.089	5.448	3.643	2.474
200	11.75	8.788	5.728	4.081	2.748
500	10.601	8.349	5.825	4.324	2.985
∞	9.210	7.824	5.991	4.605	3.219

2) Determination of Adaptive Threshold

During the term of office for the U.S. naval vessels, the characteristic of abnormal term data is that the more deviation from the term mean value, the more likely it will affect its probability distribution. According to the characteristics of abnormal term data, an adaptive gross error threshold and abnormal data detection method is designed based on the detection of critical normal distribution statistical JB value. The basic steps are as follows:

Step1: Calculate the mean value $\bar{\mu}$ and standard deviation $\bar{\sigma}$ of sample data $\boldsymbol{X}$, and set the preselected interval of obvious gross error as $[x_1, x_\lambda]$;

Step2: The interval $[x_1, x_\lambda]$ of cyclic obvious gross error distribution, the corresponding cyclic variable $j \in [1, \lambda]$, where λ is equal to $\bar{\mu} - \bar{\sigma}$, successively delete the data on the left side of the numerical cyclic variable j to form a new sample data $\boldsymbol{X}_j = \{x_{j,}, x_{j+1}, \cdots, x_n\}$, and calculate the mean value $\bar{\mu}_j$ and standard deviation $\bar{\sigma}_j$ of the sample data $\boldsymbol{X}_j$;

Step3: After the obvious gross error is deleted, the probability density distribution of the new sample data $\boldsymbol{X}_j$ may still not conform to the normal distribution. Therefore, according to the 3σ criterion[17-18], it can be assumed that the normal data is distributed in the $\left[\bar{\mu}_j - k\bar{\sigma}_j, \bar{\mu}_j + k\bar{\sigma}_j\right]$ interval, and the corresponding value interval is recorded as $\left[\tau_j^k, \gamma_j^k\right]$, where $k \in [1,3]$, step length is 0.1. Then, the sample of the value interval is circulated to calculate the statistical JB_j^k value, and the corresponding value cyclic variable τ_j^k satisfying $JB_j^k \le \xi_{JB}(n,\alpha)$ for the first time under the quantiles of $\alpha = 0.05$ is found;

$$JB_j^k = \frac{\gamma_j^k - \tau_j^k + 1}{6}\left[\left(\frac{1}{\gamma_j^k - \tau_j^k + 1}\sum_{t=\tau_j^k}^{\gamma_j^k}\left(\frac{x_t - \bar{u}_j^k}{\bar{\sigma}_j^k}\right)^3\right)^2 + \frac{1}{4}\left(\frac{1}{\gamma_j^k - \tau_j^k + 1}\sum_{t=\tau_j^k}^{\gamma_j^k}\left(\frac{x_t - \bar{u}_j^k}{\bar{\sigma}_j^k}\right)^4 - 3\right)^2\right] \quad (5)$$

Step 4：Cycling step 2 and step 3, the statistical value $\boldsymbol{JB}_j = \{JB_1^k, JB_2^k, \cdots, JB_\lambda^k\}$ of sample data $\boldsymbol{X}_j$ are obtained respectively, and the minimum value of the absolute value of the difference v_j between the statistical value $\boldsymbol{JB}_j$ and the critical statistical value $\xi_{JB}(n,\alpha)$ mapped by the corresponding sample number n_j is found. The corresponding value cyclic variable τ_j^k is regarded as the adaptive threshold of gross error τ, and γ_j^k is regarded as the abnormal data adaptive threshold γ;

$$v_j = \left|\xi_{JB}(n,\alpha) - JB_j\right| \quad (6)$$

Step 5: For the sample data interval $[x_1, x_\tau]$ and $[x_\gamma, x_n]$, the trigger of abnormal data analysis is started. If there is any associated abnormal data, it is regarded as gross error and recalculated from step 1 until no gross error is found.

Through the iteration of the above steps, the adaptive threshold of gross error τ and the adaptive threshold of abnormal data γ for the sample data X can be automatically obtained. The purpose of the automatic elimination of the gross error achieved by deleting the data in the interval $[x_1, x_\tau]$ is to ensure the normality of the sample data and then the normal distribution can be used to describe, process and analyze it. In special cases, if the adaptive threshold of gross error τ does not exist, it means that the sample data can't be described by normal distribution, and need to be analyzed separately. It should be noted that the data in the interval $[x_\gamma, x_n]$ is not regarded as gross error, nor participate in the calculation of subsequent analysis model for term of office. Its significance lies in analyzing and mining the inherent physical meaning of the data, such as analyzing the influencing factors of term extension.

3) Trigger of abnormal data analysis

The obvious gross errors can be eliminated automatically by the detection of critical statistical JB value, but those gross errors whose term is in the reasonable range of laws and regulations but who are dismissed or unable to perform their duties due to some special reasons cannot be eliminated automatically. Although this part of gross error has little influence on the normality of the sample, its sample error will inevitably reduce the quality of subsequent intelligence products, so it is necessary to design a mechanism to identify these gross errors, which can improve the effectiveness and correctness of relevant intelligence products.

In normal personnel appointment and management, the term of each person is usually only related to the previous one and the next one. Therefore, semantic analysis[19] can be carried out between the predecessor, incumbent and successor corresponding to the abnormal data in the gross error range $[x_1, x_\tau]$ and the abnormal interval $[x_\gamma, x_n]$, which is to further analyze whether the sample data belongs to the normal serving. If it is found that the serving is not normal or there is a lack of time connection between the predecessor and the successor, then these data are regarded as gross errors and cyclic iterative detection is carried out. It can be seen that the trigger of abnormal data analysis can realize the self-learning ability of gross error detection, not only can effectively improve the robustness of analysis model for term of office, but also can provide preselected data sets for later data association, processing, analysis and mining, and then improve the scalability of analysis model for term of office.

In conclusion, the spatiotemporal data for the term of office to the United States Navy vessels has normality through the detection and processing for the critical statistical JB value of the normal distribution based on the adaptive threshold, which also confirms the fact that the personnel system design and operation of the United States Navy is based on the constraints of laws and regulations, and then its normality can be used to analyze and mine the internal order and laws contained in the data.

2.4. Analysis Model for Term of Office Based on Optimal Statistical JB Value

Abundant information contained in the massive spatiotemporal data in cyberspace, so how to obtain, analyze and mine valuable information automatically and intelligently is the hotspot and difficulty of military intelligence research. According to the normality for the term data of office, the term information of office can be obtained. For example, the mean value μ reflects the legal benchmark term, and the standard deviation σ reflects the discrete influence of various factors on the term. In order to extract the inherent order and law of spatiotemporal data, this paper obtains the normal distribution model for term of office with density distribution of steepest probability and weakest dispersion through the detection of optimal statistical JB value, and then calculates the legal benchmark term and referenced adjustment interval. The basic steps are as follows:

Step1: The sample data X is detected by the critical statistical JB value of normal distribution based on the adaptive threshold, and the data sample $\boldsymbol{X}'=\left\{x_{\tau+1},x_{\tau+2},\cdots,x_{\gamma-1}\right\}$ that conforms to the normal distribution is obtained by the gross error automatic elimination;

Step 2: The mean value μ' and standard deviation σ' of the sample data X' are calculated, and the data distribution interval $\left[\mu_j^{'}-k\sigma_j^{'},\mu_j^{'}+k\sigma_j^{'}\right]$ of the normal distribution model for term of office is determined by using the 3σ criterion, and the corresponding sample value interval is $\left[\tau_j^{'},\gamma_j^{'}\right]$, where $k\in[1,3]$, step length is 0.01;

Step 3: The statistics $JB_j^{'}$ in the cyclic numerical interval $\left[\tau_j^{'},\gamma_j^{'}\right]$, and the corresponding mean value $\mu_j^{'}$, standard deviation $\sigma_j^{'}$ and confidence interval $\left[\eta_j^{'1},\eta_j^{'2}\right]$ are recorded;

Step 4: The maximum value of the absolute value of the difference $v_j^{'}$ between the statistical value $JB_j^{'}$ and the critical statistics value $\xi_{JB}(n,\alpha)$ mapped to the corresponding sample number n_j is regarded as the optimal statistics JB_{opt}, so as to obtain the optimal analysis model for term of office, and record its mean value $\mu_{opt}^{'}$, standard deviation $\sigma_{opt}^{'}$ and confidence interval $[\eta_1,\eta_2]$;

$$v_j^{'}=\left|\xi_{JB}(n,\alpha)-JB_j^{'}\right| \tag{7}$$

Step 5: The confidence interval $[\eta_1,\eta_2]$ of the optimal analysis model for term of office determines the referenced adjustment interval $\left[x_{\eta_1},x_{\eta_2}\right]$ of term, and starts the trigger of abnormal data analysis for the interval $\left[x_{\tau+1},x_{\eta_1-1}\right]$ and $\left[x_{\eta_2-1},x_{\gamma}\right]$. If there is gross error, it will be recalculated from step 1 until no gross error is found.

To sum up, the analysis model for term of office determined by the optimal statistics JB value detection can not only automatically obtain the legal benchmark term $\mu_{opt}^{'}$ and referenced adjustment interval $\left[x_{\eta_1},x_{\eta_2}\right]$, but also provide mapping data sets for the study of influencing term factors and subsequent related research.

3. Experiment And Analysis

In order to verify the correctness and rationality for the analysis model for term of office to the United States naval vessel under the multi-dimensional hypercube model with attributes, the term data of office for United States active and part retired vessels are obtained from the cyberspace by using the technology of web crawler and the method of manual screening and then the legal benchmark term and referenced adjustment interval are automatically obtained and analyzed through the statistical analysis model for it..

3.1. Experimental Data

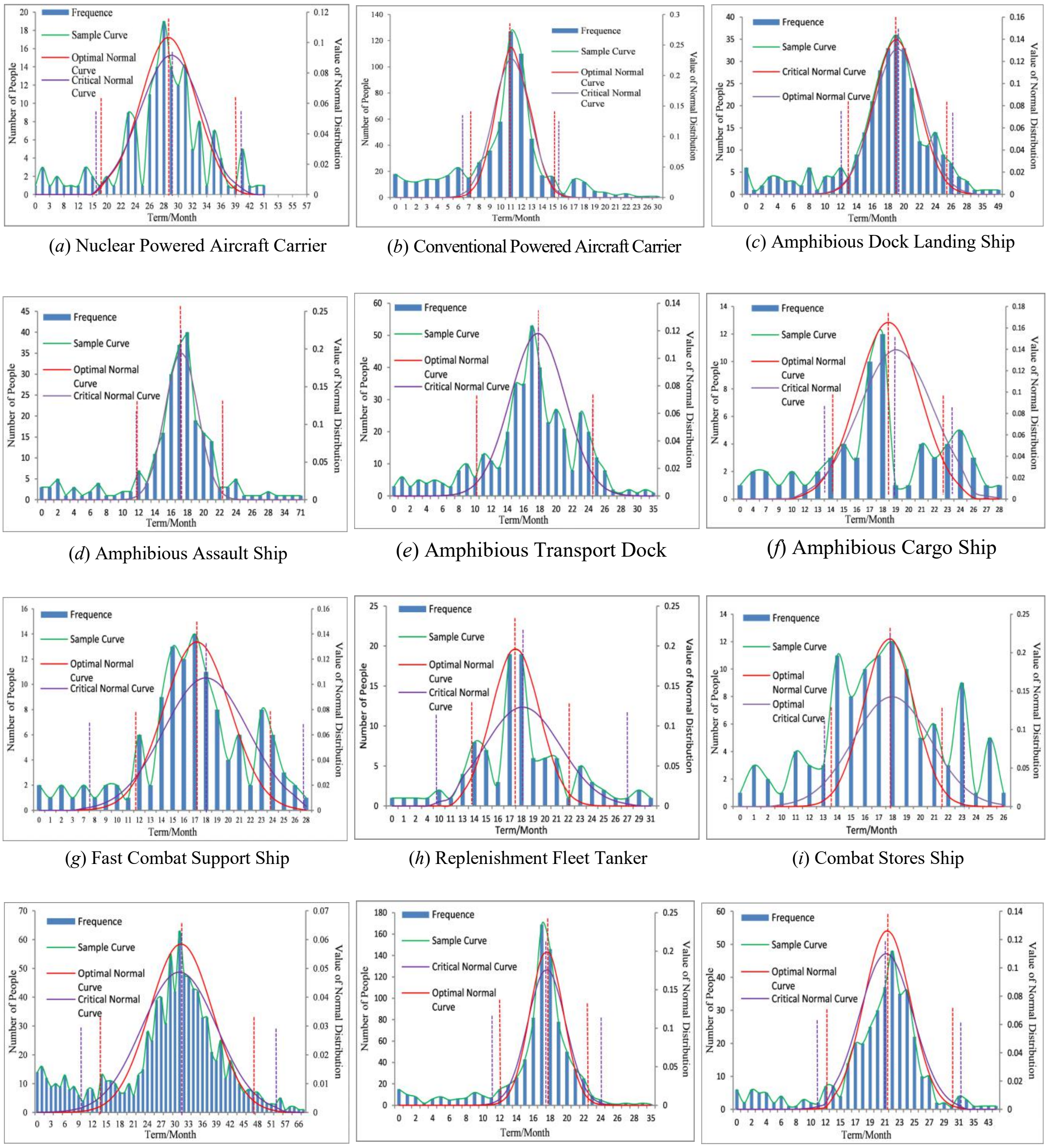
The spatiotemporal data for term of office to the United States vessels' commanding officers comes from the United States Naval Official Website, the U.S. Naval Research Association News Network, LinkedIn, MilitaryNews, Navysite, Usarriers and Facebook, etc. The relatively complete 276 of them are selected as research objects, including 3760 commanding officers (as shown in Table 2). In the process of data acquisition, the personnel spatiotemporal data of vessels' commanding officers such as Patrol Craft Coastal, Mine Countermeasures Ship, Ballistic Missile Submarines and Guided Missile Submarines are lack of many or contradictory phenomena, so they are not included in the experimental analysis data in this paper.

Table 2: Sample data statistical analysis table for term of office.

S/N	Vessels		Number of Vessel	Number of people
	Type	Abbreviation		
1	Nuclear Powered Aircraft Carrier	CVN	12	161
2	Conventional Powered Aircraft Carrier	CV	29	608
3	Amphibious Assault Ship	LPH/LHD/LHA(R)	22	371
4	Amphibious Dock Landing Ship	LSD	12	312
5	Amphibious Transport Dock	LPD	26	429
6	Amphibious Cargo Ship	AKA	5	66
7	Fast Combat Support Ship	AOE	8	121
8	Replenishment Fleet Tanker	AOR	8	101
9	Combat Stores Ship	AFS	10	109
10	Nuclear Powered Attack Submarines	SSN	86	975
11	Guided Missile Destroyer	DDG	54	805
12	Guided Missile Cruiser	CG	23	380

3.2. Analysis and Discussion on The Term of The United States Vessels' Commanding Officers

According to the basic idea of the analysis model for term of office proposed in this paper, the statistical line chart, the original curve fitting, the critical normal distribution curve fitting and the optimal normal distribution curve fitting are respectively carried out for spatiotemporal data of the term of office to the U.S. vessels' commanding officers, as shown in Figure 3(a) -3(l). In Figure 3, the blue bar chart shows the number of commanding officers corresponding to the term of office, the green curve shows the fitting of the original data curve of the term, the purple curve shows the fitting curve of the critical normal distribution after the gross error is automatically eliminated, and the red curve shows the fitting curve of the optimal normal curve distribution of the term. Among them, the purple middle dotted line and the dotted lines on both sides respectively represent the mean value and confidence interval for the fitting curve of the critical normal distribution (the reasonable sample interval after the gross error is eliminated); the red middle dotted line and the dotted lines on both sides respectively represent the legal benchmark term and the referenced adjustment interval.



(*a*) Nuclear Powered Aircraft Carrier (*b*) Conventional Powered Aircraft Carrier (*c*) Amphibious Dock Landing Ship

(*d*) Amphibious Assault Ship (*e*) Amphibious Transport Dock (*f*) Amphibious Cargo Ship

(*g*) Fast Combat Support Ship (*h*) Replenishment Fleet Tanker (*i*) Combat Stores Ship

(*j*) Nuclear Powered Attack Submarines (*k*) Arleigh Burke Class Destroyer (*l*) Ticonderoga Class Cruiser

Figure 3: Statistical analysis chart for term of office to U.S. vessel's commanding officers.

It can be seen from Figure 3 that the distribution trend of the original term data has a trend of relatively concentrated energy and gradual decline at both ends. The distribution trend is similar to the normal distribution but does not meet its characteristics.Therefore, the normal distribution can't be directly used to describe its spatiotemporal characteristics and build model. The main reason is that the data on the left side of the first purple dotted line on the left side is smaller and widely distributed, with serious influence for the symmetry of the normal curve. Obviously, the data with smaller term actually reflects the abnormal phenomenon of abnormal performance of duties, so it is necessary to eliminate these abnormal data by adaptive gross error.

1) An Analysis for the Abnormal Term Data of the United States Vessels' Commanding Officers

In the United States Naval vessels' equipment, different types of vessels have different physical characteristics, combat performance and performance of duties, so it is necessary to carry out the different design and construction in the command system of troops, which will be fully reflected in the spatiotemporal data statistical chart for their term of office. In Figure 3, the gross error limits, energy distribution and trend of the 12 kinds of vessels' term of office for commanding officers are different. It is obviously unable to meet the actual needs to use a single threshold to eliminate the gross error. Therefore, this paper proposes a critical statistical JB value detection method based on adaptive threshold, which can detect the adaptive gross error and abnormal data according to the characteristics of different spatiotemporal data for the term of office.

In order to verify the correctness and rationality of the method, adaptive threshold gross error detection and abnormal data detection tests are carried out for different vessels and the specific statistics are shown in Table 3.

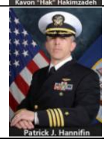
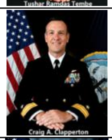
Table 3: Abnormal data detection statistics for the term of office.

S/N	Vessels	Critical JB Value	Lower Term	Upper Term	$\pm k\sigma$
1	CVN	5.57	18.88	40.01	2.1
2	CV	5.84	6.81	15.87	1.4
3	LSD	5.74	12.78	26.15	1.5
4	LPH/LHD/ LHA(R)	4.267	12.85	21.88	1.0
5	LPD	5.780	10.99	54.74	1.5
6	AKA	4.91	14.04	34.16	1.3
7	AOE	5.47	7.75	28.23	2.7
8	AOR	5.39	9.79	27.14	2.3
9	AFS	5.35	13.23	23.01	1.4
10	SSN	5.92	9.82	52.51	2.4
11	DDG	5.90	11.34	24.08	2.5
12	CG	5.79	10.95	31.67	2.3

It can be seen from Table 3 that the thresholds of gross error and abnormal data are different for different types of vessels' commanding officers. By detecting the gross error and abnormal data of the two thresholds, the normality of the remaining data can be ensured. The difference between gross error threshold and abnormal threshold is mainly related to the corresponding command system, the number of sample data and the actual distribution of samples. For example, thresholds of gross error for the CVN and LSD commanding officers' term are 18 and 12 months respectively, which are more reasonable compared with their legal tenure of 30 and 18 months.

Although the normality of sample data is ensured through the automatic elimination of gross errors and abnormal data, all gross errors and the phenomenon of deleting normal data by mistake are not eliminated. Therefore, this paper introduces the abnormal data analysis mechanism while adaptive gross error elimination, and strives for the correctness of sample data through self-learning of cyclic iteration. Taking nuclear powered aircraft carrier as an example, there is 16 and 2 commanding officer belonged to gross error and abnormal data (as shown in Table 4).

Table 4: Gross error sample data and abnormal sample data for commanding officers of U.S. nuclear powered aircraft carrier.

Commanding Officer	Serving Vessel	Appointment Time	Term(m)	Reason	Commanding Officer	Serving Vessel	Appointment Time	Term(m)	Reason
	CVN-69	2018.08.09-2019.11.15	15	Current Performance		CVN-75	2019.07.24-2019.11.15	3	Current Performance
	CVN-70	2018.06.11-2019.11.15	17	Current Performance		CVN-76	2018.9.10-20.19.11.15	14	Current Performance
	CVN-71	2019.11.01-2019.11.15	0	Current Performance		CVN-65	2010.05.06-2011.01.04	7	Lacking Performance
	CVN-72	2019.07.29-2019.11.15	3	Current Performance			2011.01.04-2011.08.17	7	Temporary Performance
	CVN-73	2008.07.30-2009.04.10	8	Temporary Performance			2015.10.15-2011.08.17	15	Vessel Decommissioning
	CVN-74	2018.08.03-2019.11.15	15	Current Performance		CVN-68	2019.8.1-2019.11.15	3	Current Performance
	CVN-75	2011.08.05-2011.11.08	3	Accidental Death		CVN-78	20018.8.10-2019.11.15	15	Current Performance
		2011.11.08-2011.11.11	0	Temporary Performance		CVN-65	2011.08.17-2015.10.15	50	Early Retirement
		2011.11.11-2012.08.16	9	Temporary Performance		CVN-73	2009.04.10-2012.08.11	40	Infective Radioactivity

(1) Gross Error of Current Perform Duties

The gross error of current perform duties is caused by the personnel in the performing duties, whose term changes with the change of data analysis deadline. Only using adaptive threshold gross error detection can't completely eliminate this part of sample data. For example, when the sample data analysis deadline is November 15, 2019, there are 11 commanding officer of the United States active nuclear powered aircraft carrier, 9 of them are detected by adaptive threshold gross error, and the other 2 are regarded as normal data. Therefore, in the trigger process of abnormal data analysis, this paper will analyze the spatiotemporal evolution integrity of their predecessor, incumbent and successor so as to ensure that they belong to gross error samples and are automatically eliminated.

(2) Gross Error of Temporary Perform Duties

The gross error of temporary perform duties is the gross error caused by the temporary duty performance in order to maintain the normal command system when personnel of office is in the missing state. This part of gross error is usually eliminated automatically in the adaptive threshold detection process, and the characteristics of its office connecting the preceding and the following also provide an important reference for abnormal data analysis. For example, the colonel John R. Haley, who is commanding officer of CVN-73, has 8 months term. The term of his predecessor Colonel David Craig Dykhoff is a gross error phenomenon by analyzing and triggering of abnormal data. Among them, Colonel David Craig Dykhoff was dismissed on July 30, 2008 due to the fire accident caused by the crew's secretly smoking and the serious damage to the vessel's hull.

(3) Gross Error of Lacking Perform Duties

The gross error of lacking perform duties is the gross error resulted from the dismissal due to the lack of performance ability, such as dereliction of duty, malfeasance and lack of supervision. The term of office has a high randomness. Only using adaptive threshold detection can't completely eliminate this part of gross error. For example, Colonel Owen Paul honors, Jr. who was commanding officer of CVN-65 was dismissed because he made indecent videos during his term as execute officer, and his term of 15 months can be detected by adaptive threshold gross error; while Colonel David Craig Dykhoff, who was dismissed for malfeasance, who has a term of 19 months, is needed to detect the fact that he belongs to gross error by abnormal data analysis.

(4) Gross Error of Accidents

The gross error of accidents is caused by accidental death or vessel's decommissioning and other special reasons. The term of office is random and uncertain. It can be automatically eliminated by combining detection of adaptive threshold gross error and abnormal data analysis trigger. For example, Tushar R. Tembe, who is commanding officer of CVN-75, died unexpectedly on November 8, 2011 for a three-month term; Todd A. Beltz, the last commanding officer of CVN-65, served only 15 months due to nuclear aircraft carrier retirement.

In addition, adaptive abnormal threshold detection and semantic analysis of abnormal samples can provide preselected analysis objects for gross error detection and analysis for influencing factors of term. For example, the term of David A. Lausman, who was commanding officer of CVN-73, was term of 40 months, which was longer than the 30-month benchmark term of commanding officer for CVN. The reason may be that it was detected that the vessel was infected with radioactivity after CVN-73 participated in the rescue operation of Japan's earthquake in March 2011, so its deployment at sea was extended; William C. Hamilton, Jr., who was commanding officer of CVN-65, had a term of office for 50 months. The reason may be that the maintenance cost of CVN-65 keeps increasing and it entered the preparatory period ahead of time on November 4, 2012.

To sum up, the adaptive threshold of gross error and abnormal threshold detection can automatically detect and remove most gross error data, so as to ensure that the analysis model for term of office is not affected by gross error as much as possible; the abnormal data analysis trigger can not only provide self-learning ability for automatic gross error removal, but also provide referenced events and objects for further researching command system design and operation information including command, such as troops composition, force deployment and duties of office, which greatly improves the scalability for term analysis model of office.

2) Analysis for Term of Office to the United States Naval Vessels

Due to the fact that the detailed data of the personnel office other than the commanding officers of the United States Navy can't be obtained, this paper only conducts spatiotemporal modeling and data analysis on the spatiotemporal data for the term of office to the U.S. naval vessels' commanding officers, and the results are shown in Figure 3 and Table 5.

Table 5: Analysis results for term of office.

S/N	Vessel	Leal Term	Adjusted Interval	JB_{opt}	Variance	$\pm k\sigma$
1	CVN	28.66	19.12-39.01	0.83	3.82	2.31
2	CV	11.24	7.19-15.00	0.12	1.60	2.23
3	LSD	19.03	13.00-25.45	3.88	2.86	2.05
4	LPH/LHD/ LHA(R)	17.27	12.99-19.14	4.27	2.04	2.10
5	LPD	17.74	10.98-24.50	5.7919	3.38	2.00
6	AKA	18.45	14.98-23.10	3.41	2.41	1.42
7	AOE	17.19	12.00-23.98	2.99	2.98	1.58
8	AOR	17.49	13.97-23.98	0.75	2.02	1.28
9	AFS	17.73	14.88-22.23	2.53	1.82	1.10
10	SSN	31.32	14.81-47.01	1.92	6.83	1.97
11	DDG	17.58	12.99-22.17	0.46	1.98	2.03
12	CG	21.31	12.32-30.02	2.57	3.16	2.44

In Figure 3 and Table 5, the optimal statistical JB values and corresponding confidence interval mapping $\pm k\sigma$ values of different vessels are different and basically around $\pm 2\sigma$, which shows that analysis model for the term of office can be self-adaptive processed and analyzed according to different objects, and its analysis conclusion basically covers 95% of the normal sample data, which conforms to the basic criteria of people using normal distribution to describe and analyze things.

In order to verify the correctness of the analysis conclusion, according to the statistics of the personnel spatiotemporal data, the comparative analysis for the term of deep drat vessels' commanding officers involved in the scheme 2018 aviation nuclear office newsletter[10] published on the official website of the United States Navy, such as CVN, AOE, LHD, LHA, AOR, LPH, LKA, AFS, LPA, AKA and LSD is made, as shown in Table 6.

Table 6: Accuracy analysis table for term of office.

S/N	Vessel	Term	Referenced Term	Error
1	CVN	28.66	30	1.34
2	LSD	19.03	18	1.03
3	LPH/LHD/LHA(R)	17.27	18	0.73
4	LPD	17.74	18	0.26

5	LPH	16.90	18	1.10
6	AKA	18.45	18	0.45
7	AOE	17.19	18	0.81
8	AOR	17.49	18	0.51
9	AFS	17.73	18	0.27

For the commanding officers of nuclear powered aircraft carrier, the term of office given by the analysis model is 28.66 months, with a deviation of 1.34 compared with the legal benchmark term; the referenced adjustment range is 19-38 months, compared with the legal 24-48 months, the part less than the legal benchmark term of 30 months is more reasonable, for example, there are 17 commanding officers with normal performance less than 24 months, but more than the legal benchmark term of 40 months has some deviations. For example, there are three commanding officers who normally perform their duties for more than 40 months. For deep draft vessels such as AOE, LHD, AOR and LSD, etc., the error between the conclusions of the analysis model and the legal benchmark term is about one month, which is better reasonable and correct.

In order to meet the needs of talent echelon construction and talent accumulation, combined with the actual international situation and mission implementation and other factors, the term of office for the United States naval vessels' commanding officers will generally shorten the actual term of performing duties, which makes a large number of samples on the left side of the legal benchmark term, forming an asymmetric pattern of left heavy and right light. The data distribution trend shown in Figure 3 and it also explains this phenomenon and verifies that the conclusion of data analysis is smaller than the legal benchmark term.

To sum up, through the statistical analysis of the spatiotemporal data for the term of office to different naval vessels' commanding officers, the conclusion of the analysis model for term of office has strong rationality and correctness, but more spatiotemporal data for the force deployment are still needed to verify the universality and robustness of the analysis model.

4. Conclusions

At present, the research of military intelligence, such as the troops composition and force deployment, is described and analyzed from a qualitative point of view, lacking the support and verification of relevant data, thus reducing the credibility and rationality of intelligence integration. This paper takes the United States naval force deployment and command system as the research object, describes the spatiotemporal evolution characteristics of the United States naval force deployment by using the multi-dimensional hypercube model with attributes, and proposes a analysis model for term of office under the hypercube. Through qualitative and quantitative experiments, firstly, the model can automatically obtain the benchmark term of office and referenced adjustment interval for the United States vessels from the cyberspace data, which has high correctness and rationality; secondly, it can provide self-learning ability for gross error elimination and data analysis by the trigger mechanism of abnormal data analysis; at the same time, it can also provide the preselected event and object data sets for subsequent related information analysis and mining, which improves the automation, robustness and scalability of the analysis model.

In view of the limitation for the acquisition of the United States naval personnel sample data, this paper will further study the automatic acquisition method of data in the case of poor coupling, and then verify the universality and robustness of the analysis model through different personnel sample data; at the same time, under the support of more extensive spatiotemporal data, we can solve the problem of spatiotemporal data isomorphism by using the design idea of the earth grid system for reference, and further study the operation mechanism, design criteria and influencing factors of troops composition and force deployment in the United States military command system, which provides methods of data analysis and mining technology for the relevant spatiotemporal military information integration.

Acknowledgments

This work was supported by National Nature Science Foundation of China, Project 41671409 and 41201392.

References

[1] LI Pin, YANG Jianlin. Research on the develo-pment path of intelligence studies based on big data thinking [J]. Journal of the China Society for Scientific and Technical Information, 2019, 38(03): 239-248.
[2] Yueting Z , Quanxi H . Research on military intelligence analysis of open network based ondata mining [J]. Journal of Intelligence, 2016.
[3] Ji-Guang G , Sheng H . Study on architecture of big data based on military intelligence analysis and service system[J]. Journal of China Academyof Electronics and Information Technology, 2017.
[4] SU Lu, LUO Qing. How to train the key perso-nal of aircraft carrier in foreign Navy [J]. Modern Military, 2007, (3): 30—33.
[5] ZHANG Jianjun, ZHANG Dexiang. Enlightenment from the training mechanism of US Navy aircraft carrier commanding officer [J]. Journal of Dalian University of Technology(Social Sciences), 2016, 37(04): 1-4.
[6] M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.
[7] Shattuck N L., Matsangas P. A 6-month assessment of sleep during Naval deployment: a case study of a commanding officer [J]. Aerospace Medical & Human Performance, 2015, 86(5), 481–485.
[8] GUO Linlin, PENG Peng, ZHANG Jianying. A-nalysis on combat capability of American carri-er strike groups [J]. RADAR & ECM, 2017, 37(03): 17-23+34.
[9] Lehner, William D. An analysis of Naval Officer accession programs [D]. Monterey, California: Naval Postgraduate School, 2008, 1-84.
[10] John Joseph Cummings. Aviation Nuclear Officer [OL]. United States Navy, 2018[2019-11-15]. https://www.public.navy.mil/bupers-npc/officer/Detailing/aviation/detailers/Documents/2018%20AV(N)%20Newsletter%20-%20Website.pdf.
[11] GU Yihua. The Methods of High Efficient Clustering on Spatiotemporal Big Data[D]. Zhejiang: Zhejiang University, 2018.
[12] Zhang C, Hu X, Sercombe T, et al. Prediction of ceramic fracture with normal distribution per-tinent to grain size [J]. Acta Materialia, 2018, 145: 41-48.
[13] LIU Xichui, SHENG Jiong, LI Yiguo. Study on probability distribution and model of system mar-ginal Price[J]. Proceedings of the CSEE, 2009, 29(04): 72-77.
[14] Jarque C M, Bera A K. A test for normality of observations and regression residuals[J]. Interna-tional Statistical Review/Revue Internationale de Statistique, 1987, 55(2): 63-172.
[15] TIAN Yu. Tests for normality based on skewn-ess and kurtosis [D]. Shanghai Jiao Tong Univ-ersity, 2012,5-21.
[16] Thadewald T, Buning H. Jarque-Bera Test and its Competitors for Testing Normality - A Power Comparison [J]. Journal of Applied Statistics, 2007, 34(1):87-105.
[17] Runmei Li, Yinfeng Huang, Jian Wang. Long-Term Traffic Volume Prediction Based on Type-2 Fuzzy Sets with Confidence Interval Method[J]. Internati-onal Journal of Fuzzy Systems, 2019, 21(7): 2122-2123.
[18] ZHU Hai, GAO Shengfeng, CAI Peng, et al. St-udy on the Self-Adaptive Wavelet Threshold Sel-ection Method Based on the Pauta Criterion [J]. Journal of Ocean Technology, 2016, 35(04):50-54.
[19] CHENG Ci. Research on Evolvement of Military Knowledge Based on Knowledge Graphs [J]. Ship Electronic Engineering, 2019, 39(06): 22-27.