Construction of Information Legal System Based on Data Mining in the Era of Big Data

Anil baburao,
Asstociate Professor, Department of Comp. Sc. & Info. Tech., Graphic Era Hill University, Dehradun, Uttarakhand India 248002,

Abstract:

As information is everything in today's era.by using anyone information or data any one can destroy the complete nation, economy, person, social structure, political structure, legal structure, justice everything. So to protect the information or data is very essential. As in India cyber law is yet a subject to strengthen. In this article we discuss the different methods of big data like genetic algorithm by which we can strengthen the information, and make it more protect and away from hackers or misuse.

keywords: *Data mining, Big data, Information, Cyber law, security*

DOI Number: 10.48047/NQ.2022.20.4.NQ22346 NeuroQuantology2022;20(4): 1197-1208

1. INTRODUCTION

In recent years, with the continuous improvement of India's legal system, the focus of legislative work has gradually changed from the promulgation of new laws to the modification and supplement of existing laws and regulations [1-2]. The establishment and improvement of the post legislative evaluation system and its standardized methods has become the latest task of India's legislative work. The post legislative evaluation system is an important exploration to improve the quality of legislation. It requires that the quality and implementation of current laws and regulations should be evaluated based on scientific and objective post evaluation methods. Combined with the evaluation results, the laws and regulations that do not conform to the trend of economic and social development should be abolished [3]. The laws and regulations with poor implementation should be diagnosed, revised and improved, and specific suggestions should be put forward accurately.

In the implementation and participation of the evaluation project after legislation, the author deeply feels that although there are many kinds of evaluation methods available, they all have their own inapplicability and shortcomings, which makes it difficult to effectively improve the efficiency of evaluation, and reduce the time cost, human and material cost of evaluation [4]. At present, the post evaluation projects of different laws and regulations are basically carried out independently, and the knowledge and experience generated in the process only play a role in the current post evaluation. After the evaluation, the expert knowledge, evaluation experience and other valuable information cannot be accumulated, so it is difficult to summarize the universal law, and the extensibility of the index system is not high [5-8].

On the one hand, this paper puts forward the general idea and systematic method of index

system construction and regulation quality evaluation, overcomes the limitations of current evaluation methods, and forms a complete set of methods and theories that can effectively guide the development of post evaluation activities. On the other hand, the rough set method is effectively integrated with the current legal evaluation method, and the former's ability of inductive learning and knowledge mining is combined with the experience of post legislative evaluation. The formation of a new method of evaluation after legislation opens up a new way of scientific evaluation in practical application, which greatly improves the efficiency of evaluation, saves legislative resources, and has great research significance and application value.

2. Review of methods of legislation evaluation

A. Attribute Reduction Algorithm of rough Set Based on Genetic Algorithm

1) Development and Characteristics of Genetic Algorithm

Genetic algorithm is developed by American scholar J. Holland first developed and discussed a random search algorithm based on Darwin's theory of evolution and Mendel's theory of genetics in 1975. In the following 20 years, genetic algorithm has entered a period of rapid development. Various genetic algorithm research institutions have been set up all over the world, and international exchange conferences focusing on the development or improvement of genetic algorithm have been held regularly. During this period, the world's first machine learning system based on genetic algorithm classifier system was produced [9]. In recent years, with the rapid development of artificial intelligence and machine learning, genetic algorithm has made great progress in both basic theory and algorithm application and improvement. At present, it has been successfully applied to economic management, computer and information technology, industrial big data and other disciplines and fields, and has effectively solved many problems, such as large-scale nonlinear system optimization and reliability optimization, fuzzy rule selection, image processing, data mining and other hot issues.

The core process of genetic algorithm is the coding of parameters, the setting of initial population, the design of fitness function, the design of genetic operation and the setting of control parameters [10]. The specific operation process is shown in Figure 1. Genetic algorithm is a powerful tool to solve the search problem at a limited cost. Compared with other methods, the advantages are as follows:

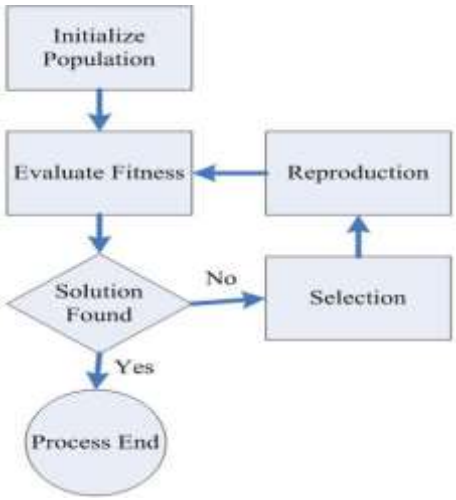

Fig.1 Basic Flow of Genetic Algorithm

(1) Genetic algorithm can be self-adaptive and self-learning, so it has intelligence. Genetic algorithm is mainly based on the guidance of selecting different fitness functions to solve the problem. It does not need other external information, and describes the probability of information passing to the next generation through individual fitness, so as to obtain the optimal solution of the problem. Therefore, the unstructured problem of complex system can be solved well.

(2) Genetic algorithm is suitable for large-scale parallel, which greatly expands the application scenarios of the method. In addition, because the method uses the group method to organize the search, it also has the parallelism inside the method, which can synchronously search in different spaces and improve the search efficiency.

(3) After determining the coding method and fitness function, the subsequent steps of genetic and mutation can be completed automatically, so the genetic algorithm also has a certain black box structure.

(4) The transformation rule of genetic algorithm is based on probability, not certainty.

(5) Genetic algorithm can produce multiple possible solutions for its problems, and decision-makers can comprehensively evaluate their needs and make corresponding choices and trade-offs.

2) **Attribute Reduction Algorithm of rough Set Based on Genetic Algorithm**

At present, there are many attribute reduction algorithms based on genetic algorithm, but the main difference is that they choose different fitness functions. Because fitness function is the only deterministic evaluation index of attribute fitness, the overall evolutionary behavior depends on fitness function.

According to the actual requirements of the reduction of evaluation indexes after legislation, this paper defines the fitness function as follows:

$$Fit(x)=nnz(x)\cdot\left(1-\frac{card(x)+\gamma(x,d)}{n}\right)\cdot\frac{\beta}{1+e^{\alpha(k_0-k(x))}}=nnz(x)\cdot f(x)\cdot P(x)\ (1)$$

n is the number of evaluation indexes; card (x) is the number of evaluation indexes in each category; γ (x,d) is the correlation between evaluation indexes and results. The larger the correlation is, the greater the correlation is. γ (x,d) ∈ [0,1] ensures that the correlation between different indexes is low after reduction of index system. P (x) is the penalty function; knife is the penalty factor; k (x) is the condition of the evaluation result. k_0 is the condition of all indexes relative to the evaluation result; α is the adjustment factor; nnz (x) is the penalty coefficient to prevent the indexes from falling into all 0. The purpose of selecting this fitness function is to reduce the index system to include as few indexes as possible, and to ensure that the correlation between different indexes is small, so as to get better reduction effect.

B. Missing Value Processing Method

In the process of post legislation evaluation, it is often difficult to take a certain index value or there is no specific value, which leads to the missing value in the decision table and affects the final evaluation result. In fact, in the process of data collection and collation, the phenomenon of inaccurate and incomplete data is very common. There are three reasons for the missing index value: first, the index value does not exist or cannot be taken; second, although the index value exists, it cannot be determined or obtained in a short time, or it is omitted in the process of collection; third, it is impossible to determine whether the index value does not exist or exists, and it is usually filled with a meaningless placeholder. The missing value in the evaluation index system after legislation generally belongs to the second

situation, that is, the index has a specific value, but there is no effective way and low-cost method to obtain it.

At present, there are mainly two methods to deal with missing values: direct discard and data completion. The processing effect of different methods is closely related to their application scenarios and data characteristics. If the number of missing samples accounts for a very high proportion of the total number, the method of direct abandonment is generally adopted. Otherwise, if the feature is added, a large noise will be brought in and the evaluation result will deviate seriously. If there are not many missing values and they are discontinuous eigenvalues, Nan can be added to the decision table as an alternative value, or fit filling can be performed according to the existing values, such as mean filling, up / down data filling, algorithm fitting filling, etc.

The traditional rough set theory is too strict for data, which limits the application scope of its model. Therefore, this paper uses the extended rough set model to deal with missing values. The principle is mainly based on the tolerance relation developed and discussed by Kryszkieewicz

$$T_B = \left\{ (x, y) \middle| x \in U \wedge \forall c_J \left(c_J \in B \Rightarrow \left(c_J(x) = c_J(y) \vee c_J(x) = * \vee c_J(y) = * \right) \right) \right\} \quad (2)$$

Obviously, T_B is reflexive and symmetric, but not necessarily transitive. Then T_B (x) is used to represent the set of evaluation objects y satisfying the relationship T_B (x, y), which can be called the tolerance class of index x.

This paper selects the classical method of incomplete information processing based on rough set. From the current research situation, domestic and foreign researchers have studied the problem of knowledge acquisition of incomplete data, and put forward many theoretical models and methods of incomplete information processing, and achieved some results. However, as a challenging research topic in data mining, there are still some unsolved problems in the existing extended model of rough set theory and its knowledge acquisition methods, which need to be further improved. 1200

C. Weight Calculation Method

1) How to Determine the Weight

When an evaluation problem is expressed in the form of decision table, the importance of an attribute can be determined by the change of system classification after removing an attribute from the attribute table. If the result of removing the attribute changes a lot, it means that the attribute is of high importance. On the contrary, the importance of attribute is low. This is the idea of using attribute importance to determine index weight. Compared with the results of expert weighting, the index weight given by rough set is based on the measurement and experimental data to determine the importance. Therefore, it is more objective than subjective methods such as expert weighting, and the evaluation results are more objective.

However, because the weight of rough set is calculated based on the sample data, the sample data must have the universality and representativeness of the problem, otherwise the objective weight is often one-sided. In addition, a simple quantitative analysis may also lead to a conclusion contrary to the reality, so quantitative analysis often needs to be reasonably combined with qualitative analysis.

Therefore, when we calculate the objective weight based on rough set, we should also combine the evaluator's or expert's experience and knowledge, that is to say, the attribute

importance of the index obtained from a large number of sample data with the weight given by subjective cognition, so as to determine the final index weight, so as to combine subjective and objective. In the combination process, an adjustment coefficient can be selected between the subjective weight and the objective weight for synthesis, and a variety of subjective and objective weight calculation methods can be used to get multiple results and screen them ; It can also combine rough set method with group decision-making method, and process the result of different experts' weight with rough set discernibility matrix to get the comprehensive weight of group decision-making.

2) Weight Generation

The objective weight calculated based on the attribute importance only expresses the influence of different indicators in the data set on the evaluation results, so it is necessary to introduce the knowledge and experience of evaluators and experts to determine the final weight, so as to achieve the purpose of combining subjective and objective weights.

Subjective weight is evaluated by experts in this field according to relevant industry knowledge and experience, and the subjective weight Q_i (i = 1, 2,...., n) of each index is calculated by AHP and other methods. Considering that the method and technology of subjective weight has been very mature, the calculation process will not be repeated.

Therefore, the final weight should be composed of two parts, one is the objective weight P_i calculated by the attribute importance, the other is the subjective weight Q_i set by experts' prior knowledge and experience. By synthesizing the two weights, the formula of comprehensive weight can be derived as follows:

$$I_i = \lambda Q_i + (1+\lambda) P_i \ (0 \leqslant \lambda \leqslant 1) \quad (3)$$

Where, λ is the adjustment coefficient, which represents the degree of attention of decision-makers and assessors to the two weights: the larger the λ, the more important the decision-makers and assessors think the expert's experience and knowledge should be in a more important position; the smaller the λ, the more important the decision-makers and assessors attach to the objective weights based on historical data and attribute importance.

Based on the above methods and procedures, it can be concluded that the calculation steps of post legislation evaluation index weight are as follows:

Step 1: collect the historical evaluation data as the sample set to form the evaluation system;

Step 2: calculate the importance degree Sig_{xi} (x_i) (i = 1,2,..., n) of each index in the index system $C = \{x_1, x_2,..., x_n\}$;

Step 3: Normalize Sig_{xi} (x_i) to P_i, $P_i = \dfrac{Sig_{xi}(x_i)}{\sum_{i=1}^{n} Sig_{xi}(x_i)}$, as the objective weight of each index;

Step 4: according to the expert knowledge, give each index subjective weight Q_i (i = 1, 2,..., n), and satisfy $\sum_{i=1}^{n} Q_i = 1$;

Step 5: select the appropriate adjustment coefficient according to the preference of the decision maker or the characteristics of the evaluation data.

The comprehensive weight of evaluation index based on objective weight and subjective weight can make the evaluation after legislation no longer be limited by the excessive reliance on expert experience and subjective weight in the past evaluation, and the evaluators can adjust the importance of subjective and objective weight by reasonably setting different λ values, so as to make the evaluation results closer to the actual situation and more in line with the requirements of evaluation and decision-making.

3. Methodology

A. Select Evaluation Sample

1) Assessment Data Collection

Information collection of post legislation evaluation is an important link in the process of evaluation. Comprehensive and objective data collection plays an extremely important role in the quality evaluation of laws and regulations. The quality of evaluation data directly affects the quality and level of post evaluation projects. Therefore, in terms of data collection methods, the following methods are used to obtain data and text information in the two post legislation evaluation projects:

(1) **Literature research method.** The first is to collect relevant research literature and evaluation results, and conduct in-depth research and Analysis on legislative ideas, basic principles, main contents, evaluation system and legislative technology. The second is to sort out and analyze the legislative archives and media data (reports, announcements, etc.), based on the literature research method, select typical cases for case study, combined with practical materials for analysis; the third is to investigate the statistical yearbook, environmental quality bulletin and other quantitative statistical data, and select the representative data as thepost evaluation supporting materials.

(2) **Questionnaire survey.** After legislation, the questionnaire is generally distributed by classification survey method, which includes three groups: management department, management counterpart and experts in water intake licensing. The quantity of questionnaires is generally guaranteed to be more than 300.

(3) **Discussion and interview.** In the process of evaluation, in view of the different problems that different objects may face, several discussion outlines were specially designed, and the management, judicial departments and management counterparts were held to listen to the opinions and suggestions of all parties on the relevant issues of evaluation. After that, according to the situation of the discussion, several representatives from different fields were selected for targeted interviews to further understand and master the relevant situation and suggestions of the assessment. The statistical results of questionnaire survey and the data processing and information extraction results of the meeting minutes and analysis reports of each symposium were timely fed back to the evaluation group and expert group, which became an important basis for quantitative scoring.

Among them, the data collection of laws and regulations is more focused on literature

research, case analysis and interview research. In the process of the project, the management department and law enforcement department have been visited and discussed for many times, and ten district water bureaus in Tianjin have been investigated and asked questions. The coverage rate of the investigation has reached 100%, and a large number of first-hand data and legislative materials have been obtained. The second regulation focuses on classified survey and sampling survey. The survey objects are divided into four groups: management departments, deputies to the National People's Congress, management counterparts and the public. Questionnaires are designed to carry out the survey, and more than 700 valid questionnaires are obtained. In addition, during the research period, we also got the assistance of theoretical and empirical knowledge from the legal working committee, academic experts and other expert groups. Therefore, the data acquisition process of the two regulations has good effectiveness and reliability.

2) Evaluation Index System

After data collection and analysis, combined with the problems of post legislative evaluation, the evaluators can finally determine the evaluation index structure. In the application of post legislation evaluation index system, there are generally no more than three levels. The first level index is the highest index, which is the common index and key evaluation index of most laws and regulations, indicating the evaluation content and scope of post legislation evaluation; the second level index is the refinement and connotation description of the first level index according to professional technology, so as to help the first level index combine with practice. The third level indicators are the alternative indicators and reference indicators of the second level indicators, which can be adjusted or increased or decreased according to the different evaluation objects. Some second level indicators that are difficult to be subdivided may not be set with the third level indicators, but it should enable the evaluators to fully understand its meaning, so as to fully combine with the actual use. According to the characteristics of the evaluation objects, the post evaluation project team of the two sample regulations respectively constructed the following two evaluation index systems:

(1) The evaluation index system of Tianjin water intake permission management regulations includes five first level indexes of legality, rationality, coordination, operability and effectiveness, with weights of 0.10, 020, 0.05, 0.30 and 0.35 respectively. Among them, legitimacy includes two secondary indicators, which mainly analyze whether the laws and regulations break through the legislative authority, whether they violate the relevant provisions of the upper law, whether they conflict with other water laws and regulations in Tianjin, and whether the compulsory and punishment measures are legal. Rationality includes five secondary indicators, which mainly examine whether the laws and regulations can meet the actual situation and needs of Tianjin water intake license management, whether the specific systems are necessary and appropriate, whether the specific contents can reflect the principle of fairness and justice, whether the administrative procedures are open and reasonable, and whether the legal responsibility is appropriate. Coordination includes three secondary indicators, which mainly analyzes the coordination of various provisions within the laws and regulations, the coordination of local laws and regulations at the same level, and whether the supporting system has been established; operability includes four secondary indicators, which mainly analyzes whether the laws and regulations are targeted to practical

problems, whether the management system is perfect and feasible, whether the administrative procedures are convenient for the people, and whether the system measures are feasible. The effectiveness includes four secondary indicators, which mainly examine the actual effectiveness of the water intake permit system, the improvement of administrative procedures, the equivalence of cost and effect, and the completion of legislative objectives. The whole evaluation index system includes 78 three-level indexes.

(2) The evaluation index system of Guangzhou city appearance and environmental hygiene management regulations includes six first level indexes: legitimacy, rationality, operability, effectiveness, coordination and standardization, with weights of 0.15, 0.25, 0.25, 0.25, 0.05 and 0.05 respectively. Among them, legitimacy includes three secondary indicators, which mainly analyze whether it is beyond the legislative authority, whether it is in conflict with the superior law, and whether it is in conflict with the same level laws and regulations. Rationality includes five secondary indicators, which mainly analyzes the actual needs of local laws and regulations, whether the established system is necessary, whether it can reflect the principle of fairness and justice, whether the administrative procedures are open and reasonable, and whether the legal responsibility is appropriate. Operability includes four secondary indicators, which mainly analyzes whether the management system in the regulations has realistic pertinence, whether the management system is perfect and feasible, whether the administrative procedures are convenient for the people, and whether the specific contents are feasible; effectiveness includes four secondary indicators, which mainly analyzes the realistic effectiveness, convenience for the people, cost-benefit relationship and implementation effect of the regulations. Coordination includes three secondary indicators, which mainly analyzes the coordination between the content of laws and regulations and the same level of laws and regulations, the cohesion of systems and the completeness of supporting systems; normative includes two secondary indicators, which mainly evaluates whether the elements of legal norms are clear and whether the concept is defined clearly. The whole evaluation index system includes 80 three-level indexes.

B. Building rough Set Evaluation Model

According to the basic theory and evaluation method of post legislation evaluation, this paper intends to build an objective, scientific, accurate and efficient post legislation evaluation model through the research and mining of historical evaluation data, and use this model to evaluate the quality of two sample regulations. The construction process of the evaluation model is mainly divided into six steps: the establishment of information system, the discretization of index data, the processing of missing values, the selection of key indicators, the determination of index weight and the comprehensive formation of the evaluation model (Figure 2).

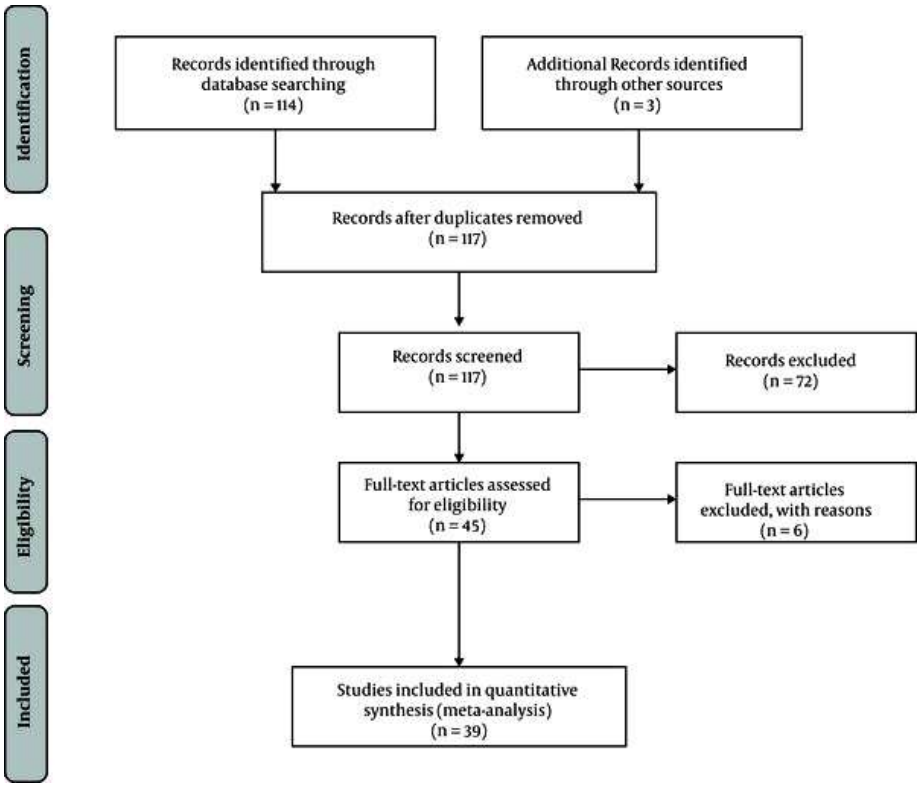

Fig.2 Flow Chart of Post Legislative Evaluation Based on rough Set

To establish an evaluation model, first of all, it is necessary to collect historical data and establish an indicator set, so as to conduct the primary selection of indicators. At this time, the index system has a certain number of redundant indicators. The primary selection of indicators needs to ensure the accuracy, comprehensiveness and accessibility of the data, so the selection scope mainly includes the evaluation series or evaluation reports after legislation that have been completed and published in different provinces and cities.

In this study, rough set attribute reduction algorithm based on genetic algorithm is used for index reduction, so as to screen out key indexes and remove redundant indexes. Firstly, the decision table data is trained and reduced based on genetic algorithm, and the conditional attributes with low contribution to the decision attributes in the decision table are deleted. Then the decision rules are generated according to the reduced indexes and their corresponding index values. A total of 223 rules are generated in this example. According to the characteristics of decision rules, the rules with accuracy > 0.75 and coverage > 0.25 are selected to form a decision rule set, and the occurrence rules of all indicators in the rule set are studied and classified. After the selection of key rules and the combination of similar indicators, the evaluation indicators included in the decision rules are as follows:

Set = {the number of experts involved in the legislative argumentation, the number of clauses in conflict between the legislation and the superior law, whether the administrative penalty breaks through the scope of the superior law, the reasonable division of powers, the effective time, the revision frequency, and the transformative events after the effective. It has regional characteristics, the expected goal is commensurate with social problems, conforms to the principle of administrative justice, there is no conflict within the provisions, and the punishment is targeted. The upper level law should be refined in time, involving the

collection rate of fees, the law enforcement procedures should be clear, the actual management subject should be clear, the proportion of subjective support should be determined, and the law enforcement should be guaranteed by teams and systems. The responsibilities of each management body can be fulfilled, the executive efficiency of administrative units, the frequency of summary and reporting, and the smooth operation of management system. The average annual use times of administrative and judicial departments, appeal rate, administrative staff's satisfaction with operability, administrative counterpart's satisfaction with administrative procedures, and public's satisfaction with administrative results. Whether the public's legal awareness has been significantly improved, the superior resources have been uniformly allocated and utilized, the practical problems have been solved, and whether the special funds and facilities have been used for other purposes}.

According to the reduced index system X and index weight I, this paper uses the linear weighting method to calculate the score of each index to form the final evaluation model (formula 4). In addition to the linear weighting method, it can also combine TOPSIS (superior and inferior solution distance) method, projection method to calculate the distance from each evaluation object to the reference point, or combine the fuzzy set to construct the membership function to obtain the evaluation model.

$$MODEL = \sum_{i=1}^{31} x_i I \quad (4)$$

C. Comparison of Evaluation Results

According to the scoring habits and standards, the indicator scoring system is generally adopted, and the quality level of regulations is determined according to the scoring results. Laws and regulations can be generally divided into five grades: Grade I in the interval [90,100], grade 11 in the interval [80,90], grade III in the interval [70,80], grade IV in the interval [60,70) and grade V in the interval [0,60], which correspond to good, good, fair, poor and poor quality of laws and regulations respectively. According to the survey data and data analysis results, the evaluators and the expert group scored the indicators of the two sample regulations respectively. The evaluation results of the two regulations are shown in Table 1.

Table Ⅰ Actual Evaluation Results of Sample Regulations

Category	Regulation 1	Scoring rate	Regulation 1	Scoring rate
Legitimacy score	7	70%	14.25	97.27%
Rationality score	11.8	59%	20.26	88.18%
Coordination score	3	60%	4.854	94.91%
Operability score	19.8	66%	19.11	76.58%
Effectiveness score	28.35	81%	20.34	84.55%
Normative score	-	-	4.75	96.91%
Total score	69.95	69.95%	83.35	83.85%
Regulation quality level	IV		II	

4. CONCLUSION

The theory and method of post legislation evaluation is an important research content in the field of legislation. How to realize the scientific and intelligent evaluation of the implementation quality of laws and regulations is an important research topic in the process of system construction and information construction. The research on the index system and method of post legislative evaluation based on rough set is a new research content in the field of post legislative evaluation. Rough set is a high-performance uncertain machine learning method. It is necessary to study its reasoning learning mechanism and apply it in practice. Combined with the actual characteristics of post legislation evaluation, this paper systematically studies the whole process of index system construction and comprehensive evaluation based on rough set. In this paper, the main value of this evaluation method is to supplement and improve the current post legislative evaluation system. With the help of the convenience, efficiency and accuracy of model evaluation, the institutionalization and normalization of law and regulation evaluation can be realized. At the same time, it combines with the comprehensive and interpretable methods of investigation and discussion to realize complementary advantages and constantly improve the level of post legislation evaluation.

REFERENCES

[1] Peng Chenxi, Zhou Xinxia, Tang Shang. Research on the Construction of Information Legal System in the Era of Big Data. Chinese and Foreign Entrepreneurs, 2020, No. 673 (11): 250-251

[2] Zong Cheng, Ma Haiqun. on the Construction of Discipline System of Information Law. Information and Data Work, 2004 (04): 11-14 + 39

[3] Liu Haiying, Pu Zhoujun. Thoughts on the Construction of India's Personal Information Protection Legal System Under the Credit Reference System. Zhejiang Finance, 2008, 000 (010): 21-22

[4] Liu Min, Zeng Wei. Construction of Network Security Management Legal System in the Information Age. Information and Communication, 2015, 3: 125-127

[5] Liu Pinxin, Li Zecai. Building a Legal System of Information Security. Information Security and Communication Secrecy, 2003, 2: 5-5

[6] Zhang Xiaoxuan. Legal System Construction of Information War. Legal System and Society, 2009, 27: 13-14

[7] Zhou Jie. Construction of Network Security Management Legal System in the Information Age. Theoretical Research on Urban Construction: Electronic Edition, 2016, 6: 1573-1573

[8] Xiao Zhihong. Research on the Legal Architecture of Cryptography in India. Research on Information Security, 2018, 4 (0 (9)): P.853-856

[9] Liu Pinxin, Li Zecai. Building Information Security Legal System Vigorously. Information Security and Communication Confidentiality, 2003, 02:4-4

[10] Tan Xiuli, Zhang Maoxin. Research on the Legal System to Ensure Network Information Security in the Era of Big Data. Business Information, 2014 (issue 30): 292- 292[