

## Research Article

# Construction and Application of E-Government Simulation Network Experimental Teaching Practice Environment under the Background of Machine Learning

Wei Zhao 

*School of Public Affairs & Law, Southwest Jiaotong University, Chengdu, Sichuan 610031, China*

Correspondence should be addressed to Wei Zhao; [wzhao@swjtu.edu.cn](mailto:wzhao@swjtu.edu.cn)

Received 23 April 2021; Revised 28 May 2021; Accepted 19 June 2021; Published 6 July 2021

Academic Editor: Sang-Bing Tsai

Copyright © 2021 Wei Zhao. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the rapid development of economy and Internet technology, e-government has been reflected in the reform of government information management and is closely related to the security of e-government networks. Existing technologies for evaluating network security can provide good evaluation results when dealing with deterministic data, but they are not sufficient to evaluate the uncertainty of information. In order to solve the problem of the construction and realization of the network training environment, a network security assessment method based on gray relational analysis and D-S theory proof is proposed. First, the scaling method is used to determine the weights of different levels of indicators, because different levels of indicators have different effects on the network state. Secondly, taking into account the differences in the degree of participation of different indicators at different levels, the participation function of each level indicator at the level is determined and a weighted amplitude coefficient to reduce the overall uncertainty is established. The network security situation is sudden and volatile. Therefore, the network security prediction method should have a good ability to deal with nonlinear problems. In this case, a network security state prediction method based on the GM performance model is proposed. The GM performance model has excellent nonlinear debugging functions. It has low requirements on the number of samples and good practicability. The experimental results show that the dynamic decision server push framework based on the XGBoost algorithm proposed in this paper has better performance than the single push method for push server, can use network resources appropriately, and improves the network environment security rate by at least 50%.

## 1. Introduction

The rapid development of information technology has made the Internet more closely related to people's production and life, and it has also profoundly affected people's lifestyles [1]. Government agencies, as administrative agencies serving the people, provide people with convenient and fast services. With the popularization and development of information technology, traditional inefficient government work methods can no longer adapt to the development of social information and cannot satisfy people's convenience and the need for fast administrative services; therefore, the use of information technology to transform the government's work methods is an urgent need for social development. E-government is the product of the government's effective use of

information technology to realize the transformation of office and service methods. As a government agency, as a decision-making and service agency that masters important national information, the accompanying information security is an important issue [2].

Compared with the C/S communication method based on a customized client, the browser-based web communication engine has the characteristics of cross-platform, cross-hardware, simple operation, and simple maintenance. More and more applications are beginning to be compatible with B communication method S. In the future, the scale of application services supported by mainstream browsers will continue to expand. Compared with the C/S communication method, the obvious limitation of the early B/S structure is that the browser must receive information and actively send

HTTP requests. HTTP is a one-way protocol. The server accepts the request and responds. The browser can retrieve the request. If the browser requests the same server again, this process must be repeated.

Ge et al. believe that, in order to meet the needs of the electric power industry for the ability of electrical engineering graduates under the new situation and to respond to the Ministry of Education's initiative to carry out new technology research and practice, we have to deepen the reform of educational technology and actual reforms and combine scientific transformation and development with school teaching and research. It is the starting point to establish the basic idea of cultivating students' comprehensive technical potential and create a truly realizable dynamic microgrid model experimental platform. Through optimizing and reorganizing teaching content, the system fully reflects and makes full use of resources to deal with new technical problems, develop and deepen technological reforms, and promote the continuous improvement of the practical quality of courses such as photovoltaic and microgrid production. But how to popularize the education reform so that every student can benefit is a problem that needs to be improved [3]. In order to solve the problems of high risk and extreme operating environment in the training of mechanical mining talents, Xu et al. found that it is difficult to provide a practical basis, high cost of a large number of irreversible functions, and high consumption experiments, and it is difficult to provide high-level services, scientific research results, and difficult articles. Virtual simulation technology to develop educational software is used, and research is focused on experimental teaching, safety training, practical knowledge and production practice, precise positioning of software functions, software architecture design, configuration of technical systems and development process, and four teaching links for the final successful development of virtual learning software. Terminal-based software is widely used in teaching, has shown significant advantages and results, and has been highly recognized. However, the system still has some shortcomings, and it needs to be continuously improved in the follow-up work [4]. Fang believes that the international economy and trade in the network environment are diverse and based on technology and data. At present, the experimental teaching of many universities in international financial and commercial enterprises still lacks sufficient experimental equipment and insufficient consideration of the relationship between practical teaching, insufficient teacher skills, poor internship results, and many other problems. Therefore, it is necessary to continuously increase the research on the teaching of international economic experimental reformists in the network environment, so as to continuously improve the teaching level and improve the quality of teaching. He first briefly taught the importance of practical guidance, then pointed out the problems in the teaching of international economics and business practice, and finally proposed practical measures to solve the international economy and trade in the network environment. However, the actual situation is far more complicated than the experiment, and

a large amount of data should be obtained to get an answer close to the result [5].

The deployment and operation of the job title declaration system of the education e-government platform will greatly improve the overall efficiency from job title declaration to expert selection to job title review and solve the problem of document layout and formatting faced by tens of millions of elementary and middle school teachers across the country when filling in the application materials every year. Personnel specialists at all levels collect data; this paper proposes a decision-making model based on the XGBoost algorithm of a decision tree, which dynamically adjusts the server push method according to actual usage scenarios and pushes according to such as service priority, current network environment, and server resource occupancy. Indicators use the XGBoost algorithm to select the most suitable push method at present, taking into account the full utilization of network resources while ensuring the efficiency of server push [6]. Different application scenarios should choose different push methods to realize the interaction between the client and the server, so as to achieve the purpose of making full use of resources and improve the service quality.

## 2. Machine Learning Simulation Network Experiment Method

*2.1. Concept of Network Situation Awareness.* The network situation is the result of several factors such as the operating conditions of the network equipment, the status of the network itself, and the use of users. From this, it can be seen that the network situation reflects the interaction between the various entities of the network, which is a process of change, with hierarchical, volatile, and macroscopic characteristics. The concept of situation factor is a factor that can reflect the situation of network security and is indivisible. It is the basis of situation awareness of network security [5, 7, 8]. The selection of the situation factor needs to proceed from the overall situation, which can not only reflect the different levels of the network system but also reflect the heterogeneity of the network structure [8]. The concept of machine learning face recognition is shown in Figure 1.

*2.2. Introduction to Spring and Spring Boot.* Spring Boot is a spring-based microservice framework [9, 10]. By creating a new development model, the very cumbersome content in spring application development is eliminated and spring is simplified. First, let us introduce spring: spring is a light-weight open-source Java framework, originally created by Rod Johnson and introduced in the book. Its birth is largely due to the expansion and complexity of EJB at that time, leading to higher growth costs and longer growth cycles. The birth of spring aims to reduce the complexity of developing business applications. It is based on improving the experience of developers in the development process and various strategies to free developers from tedious repetitive coding and other tedious tasks, thereby greatly improving development efficiency. Therefore, it is widely used and developed by Java developers. So it can be seen everywhere in Java

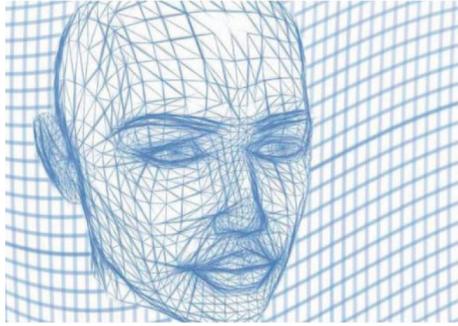


FIGURE 1: Conceptual diagram of machine learning face recognition (the picture is obtained from <https://image.baidu.com/>).

[11, 12] and applied and widely used in enterprises. The developer framework owns the world. Spring also chose an excellent business solution and started it as a project. After more than ten years of continuous development, spring has become the king of Java development components [13, 14]. The goal of spring is to make Java application development easier and more efficient in a variety of ways [15, 16].

**2.3. Introduction to Redis Technology.** Redis (Remote Dictionary Server) is an open-source high-performance ANSIC, with network-based, memory-based persistent logging, connection to the database, caching, and message passing, providing various key-value data types and various instructions to adapt to different situations Cache requirements and queues under [17, 18]. It supports various data structures, such as strings, scores, lists, sentences, sorting and range queries, bitmaps, and GEOS, so Redis can respond to a variety of application scenarios, and Redis can store, read, and write data in memory; the input performance is excellent.

In addition to the above functions, Redis also provides key expiration, issuance and recording, transactions, pipelines, Lua scripts, and various levels of disk persistence and is automatically distributed to ensure the high availability of Redis Sentinel and Redis clusters [19, 20].

Using Redis to store frequently accessed data can effectively reduce the time-consuming business logic associated with the background, reduce interface TPS, and improve system performance. The network process is shown in Figure 2.

**2.4. Introduction to the Database.** Database technology is to study the structure, storage, design, management, and use of application database and use these theories to process, analyze, and understand the basic theories and methods of data in the database. In other words, database technology is a science that studies database management and application software [21, 22]. The research and management of database technology focuses on the data itself. Therefore, the specific content of database technology mainly includes the following: through the overall organization and management of data, systems and databases are used to create corresponding databases and data warehouses according to a

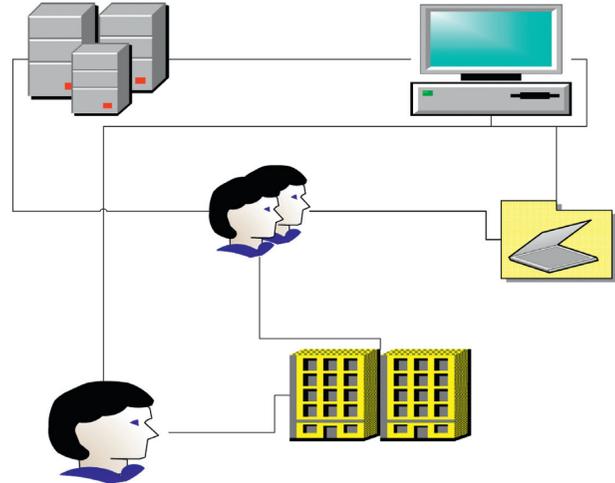


FIGURE 2: Network flowchart.

defined structure and a data management and data mining application system for data management and database mining system are designed. You can add, change, delete, edit, or analyze understand, report and print data in the database, and use the system application management to finalize the data processing and analysis and understanding. Database technology is an important technology of information system. It is a method of managing data through computer. This method studies how to organize and store data and how to effectively access and process data, study the structure, storage, design, management, and application of the database, and use these theories to realize the processing of data in the database. Analysis and understanding are the basic theories and application methods [23, 24]. This means that database technology is the software science used to research, manage, and implement databases. Database technology is an important part of modern information science and technology and is the core of data processing and information management systems. Database technology research and solutions are used to effectively organize and store large amounts of data in computer information processing, reduce data storage redundancy, share data, ensure data security, and perform effective data retrieval and processing in the database system.

### 3. Construction and Application Experiment of Network Experiment Teaching Practice Environment

**3.1. Idea of Determining the Dynamic Decision Algorithm.** The current actual web application environment often requires comprehensive consideration of many external and internal factors, which results in a large number of decision-making indicators for selecting server push methods, such as service real-time requirements, service reliability requirements, client service process priority, and whether network resources are surplus (network bandwidth, throughput, etc.), whether the server or client resources are sufficient (local and remote server thread pool capacity, CPU usage, and memory usage), etc[25, 26]. Trying to use the dynamic

decision server push method, we must comprehensively consider the above factors and finally select a reasonable push method under the current environment.

The entropy of a random variable is defined as the following formula (1):

$$H(X) = - \sum_{i=1}^n P_i \log P_i, \quad (1)$$

$$H(x, y) = - \sum_i^n P(x_i, y_i) \log P(x_i, y_i). \quad (2)$$

When machine learning is used in the research environment of this paper, it is necessary to determine a specific decision result; that is, when inputting environmental influence factors as processing data, it is necessary to output the decision result of the algorithm as the final goal, which is a kind of supervised learning requirement [27, 28]. The training data set involved in supervised learning needs to contain the target results artificially labeled by the feature set. This paper needs to design a learning algorithm to extract features from the input training data and learn to mark the target division method. This method is applied to the test set data in the prediction results. This learned classification method is finally applied to the server push method dynamic decision framework built in this paper:

$$H(x, y) = - \sum_{i=1}^n P(x_i, y_i) H(x|y_i), \quad (3)$$

$$I(x, y) = H(x) - H(x|y). \quad (4)$$

The main problems that supervised learning solves are divided into two categories: classification and regression. The classification algorithm outputs specific results, and the regression algorithm outputs a certain interval of quantitative results. For the problems discussed in this paper, both classification and regression are applicable, because when a regression algorithm is used to solve dynamic decision-making problems, specific output results can be obtained by converting the output results into probability distribution problems [29, 30]. There are several mainstream algorithm models in supervised learning algorithms. Mathematical models can be generated through learning. Common models include support vector machine (SVM), naive Bayes, K-nearest neighbor, and decision tree:

$$I_R(D, A) = \frac{I(A, D)}{H_A(D)}, \quad (5)$$

where  $H$  is the characteristic entropy:

$$H_A(D) = - \sum_{i=1}^n \frac{|D_i|}{D} \log_2 D. \quad (6)$$

The method of constructing a binary tree is adopted to replace the previously constructed multibranch tree to improve the search efficiency; the Gini coefficient is selected as the basis for segmentation:

$$\text{Gini}(D) = 1 - \sum_{K=1}^K \left( \frac{C_K}{D} \right)^2. \quad (7)$$

According to formula (7) and the characteristics of the CART algorithm to construct the decision tree as a binary tree, the Gini coefficient calculation formula for the CART algorithm can be obtained:

$$\text{Gini}(D, A) = \frac{|D_1|}{|D|} \text{Gini}(D_1) + \frac{|D_2|}{|D|} \text{Gini}(D_2). \quad (8)$$

The CART algorithm adopts the postpruning method; that is, after the decision tree is established, it is determined whether or not to prune according to the loss function value of any subtree:

$$C_\alpha(T_t) = C(T_t) + \alpha|T_t|. \quad (9)$$

As shown in Figure 3, the residual of the result of the decision tree  $T_1$  training and the actual result  $T$  is obtained. The residual is used as the target of  $T_2$  training optimization. The final trained strong classifier is simply added between the weak classifiers to obtain

$$T = T_1 + T_2 + T_3. \quad (10)$$

**3.2. CART-Based Boosting Algorithm to Determine the Idea.** Boosting algorithm contains several important components: function model, objective function, and optimization method. The main problem that the Boosting algorithm solves can be summarized as how to adjust the training set so that the training set can continue to train the next round of classifiers and how to form a strong classifier between the weak classifiers after training. It mainly includes AdaBoost algorithm, boosted tree, and GBDT algorithm:

$$u = \frac{\sum_{i=1}^n y_i}{n}. \quad (11)$$

Then the error of the node can be expressed as

$$\text{Error} = \sum (y_i - u)^2. \quad (12)$$

The use of the analytic hierarchy process to assign weights first requires the establishment of a hierarchical model based on subjective inference. The model has the highest level, the middle level, and the lowest level. The highest level includes the final result of the decision, and the lowest level is the alternative plan of the decision. Layer is the environmental factor to be considered in decision-making [31]. A simple layer model is shown in Figure 4.

In the process of node splitting, the attribute with the largest splitting gain needs to be selected for division. The calculation method of the column gain  $G$  is

$$G = S - S_I + \sum_N N - Y_i. \quad (13)$$

Using variance as the loss function,  $S_j$  can be obtained as

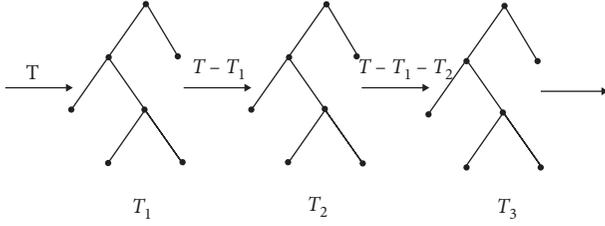


FIGURE 3: Schematic diagram of GBDT algorithm idea (the picture is obtained from <https://image.baidu.com/>).

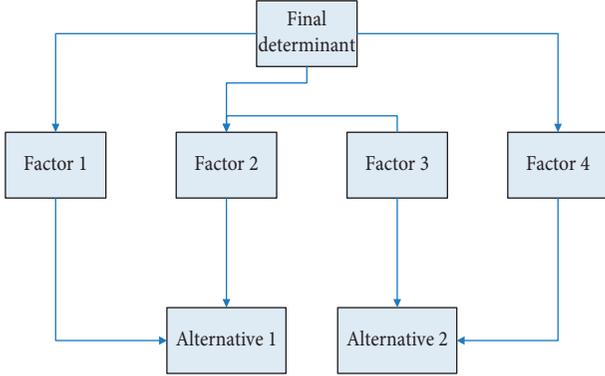


FIGURE 4: Analytic hierarchy process model.

$$S_j = \sum_{meL} (y_m - u_L)^2 + \sum_{meR} (y_n - u_R)^2. \quad (14)$$

So, the problem of sorting each node is to find an attribute to maximize the splitting gain. Expand  $S$  and  $S_j$  separately as

$$G = \left( \frac{\text{Sum}_L^2}{\|L\|} + \frac{\text{Sum}_R^2}{\|R\|} - \frac{\text{Sum}_M^2}{\|M\|} \right). \quad (15)$$

Compared with GBDT, XGBoost's optimization is regularization. When the weak classifier is CART, it will regularize the node leaf value and node leaf number of the tree:

$$v(\varphi) = \sum_i I(\hat{y}, y_i) + \sum_k \Omega(f_k). \quad (16)$$

Therefore, after the data matrix is obtained, the calculation of the index proportion is directly carried out:

$$P_{ij} = \frac{X_{ij}}{\sum_{i=1}^n X_{ij}}, \quad (j = 1, 2, \dots, m). \quad (17)$$

On the basis of the analytic hierarchy process model, it is necessary to construct a judgment matrix. On this issue, Saaty et al. proposed to use the consistent matrix method, that is, the pairwise comparison results of different factors. This is to reduce the subjectivity of the qualitative method of the analytic hierarchy process [32]. Before constructing the judgment matrix, it is necessary to give a subjective degree of importance rating. In the context of the experiment in this paper, the rating scale applicable to the experimental objects in this paper is shown in Table 1.

TABLE 1: Proportion scale table of analytic hierarchy process.

Comparison scale	Quantized value
Two factors are equally important	1
One of the two factors is stronger than the other	3
One of the two factors is stronger than the other	5
Neighboring judgment intermediate value	2, 4
Take the reciprocal of the matrix	$A = P * a + b$

The consistency test judges whether the logic of the matrix is correct. Unscientific subjective setting will cause the matrix  $A$  to provide contradictory information. The consistency test first needs to calculate the consistency index CI:

$$CI = \frac{\lambda - n}{n - 1} + \frac{C}{RI}. \quad (18)$$

The entropy value of the  $j$ th environmental factor can then be calculated:

$$e_j = -k \times \sum_{i=1}^n P_{ij} \log(P_{ij}), \quad k = \frac{1}{\log m}. \quad (19)$$

Formula (20) calculates the final weight of the  $j$ th environmental factor:

$$W_j = \frac{g_j}{\sum_{j=1}^m g_j}, \quad j = 1, 2, \dots, m. \quad (20)$$

## 4. E-Government Simulation Network Experiment Teaching Practice Analysis

*4.1. E-Government Simulation Network Experiment Teaching Practice Structure.* This part is mainly based on the results of the questionnaire survey, through the analysis of relevant data to understand the characteristics and shortcomings of the current Z University undergraduate internship. The survey is mainly conducted from three levels of students, schools, and enterprises. The student level specifically includes the degree of emphasis on internship, the purpose of the internship, the choice of internship unit, the degree of understanding of innovation and entrepreneurship education, and the views on entrepreneurship. The survey data are shown in Table 2.

Network experimental teaching and government simulation learning are all artificially designed behaviors to realize simple interaction with attackers and provide upload, download, function code response, and other functions. However, this design still has certain shortcomings. For example, in terms of function code response, it is usually implemented based on a user-defined template or parsing a static file. Once it encounters a request that cannot be recognized or is not included in the file, an error will occur. In order to further improve the authenticity of government simulation learning, simulation technology or real PLC equipment is applied to government simulation learning.

As shown in Figure 5, the survey data of this research and Zhaolian recruitment data further prove that the internship experience plays a vital role in the job search process of fresh

TABLE 2: Statistics of the basic composition of the survey sample.

Variable	Category	Frequency	Percentage
Gender	Male	102	41.6
	Female	143	58.4
Grade	Junior	86	35.1
	Senior year	159	64.9
Subject classification	Science and engineering	114	46.5
	Liberal arts	131	53.5

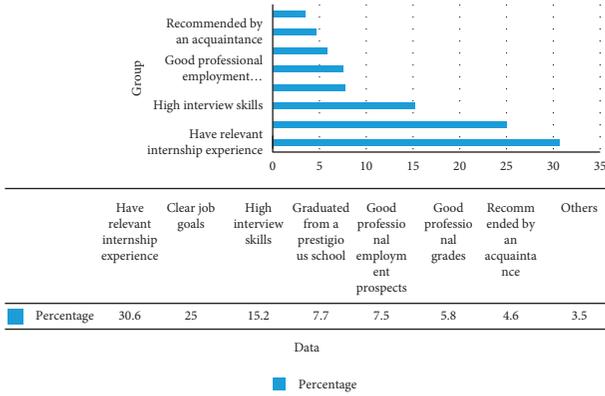


FIGURE 5: Reason for graduation in 2018 which was the reason for the offer.

graduates. The government simulation study deployed Conpot to simulate Siemens S7-300 PLC equipment, at the same time generated real-time production data by running Simulink simulation model, and realized the simulation of real equipment by simulating sensor output. In addition, part of the government simulation study uses real equipment in the design.

As shown in Figure 6, it can be seen from the above data that the vast majority of students are to accumulate work experience and improve their work ability. Placed between the government simulation study and the real PLC device, when the attacker’s request cannot be responded by the emulator, it is forwarded to the real PLC device, the PLC response is collected, the corresponding environmental information is replaced, and then a new response is constructed. The client side solves the problem of synchronization between PLC and session context.

As shown in Figure 7, if the importance of an opinion is determined based on the expert’s expertise and experience, collect the real RTU and HMI interactive traffic and provide it to the government simulation study, so that the government simulation study has higher authenticity. At present, the mainstream industrial control government simulation learning relies on the simulation of target equipment and target services. However, due to the long-term closure of ICS, the research on ICS software architecture and virtualization is relatively weak. The current industrial control government simulation learning is mainly through the protocol of the decoy interface. Using stack to simulate the interaction of industrial control equipment, there is still a big gap with the real sense of network

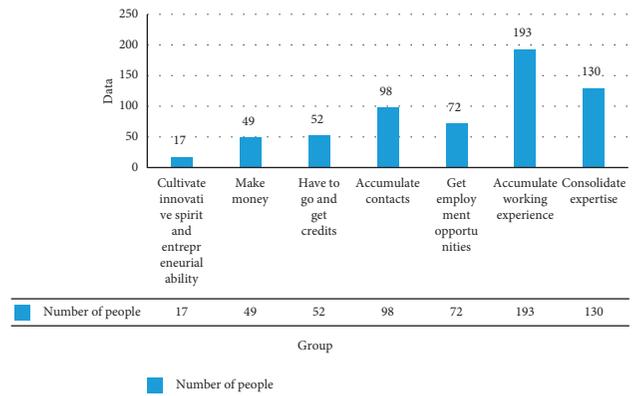


FIGURE 6: Purpose of internship.

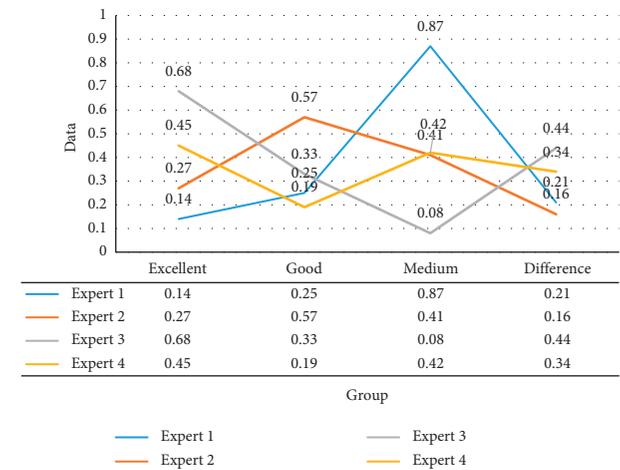


FIGURE 7: The degree of membership of the evaluation results of the three experts on the evaluation indicators to each evaluation level.

experimental teaching, industrial control, and government simulation learning.

In the future research of industrial control government simulation learning, different technical ideas and methods can be used for reference, combined with other technologies to form complementary advantages, such as the use of artificial intelligence, big data, and other tools to achieve government simulation learning with high simulation capabilities. The government simulation learning system combined with innovative ideas or technology has high defensive performance and deception ability.

TABLE 3: Registered user information table.

Sequence	Field name	Type of data	Primary key	Allow empty	Remarks
1	Y_ID	Int (10)	Yes	No	Uniquely identifies
2	y_username	Vachar (20)	No	No	Username
3	y_password	Vachar (20)	No	No	Password
4	Y_sex	Vachar (4)	No	No	Gender

TABLE 4: Comparison of prediction errors of three prediction methods.

Comparison algorithm	RMSE	MAPE
GM (1, 1) power model	0.0757	0.2025
Elman model	0.0851	0.0005
Gray Verhulst model	0.1107	0.2518

As shown in Table 3, it mainly analyzes the various functional designs involved in the e-government website. The ICS simulation platform with government affairs and network teaching as the core provides ICS internal simulation data and external attack data and provides data support for security personnel to analyze network risks and perceive the network situation. Government data collection focuses on data interaction within the system and host status information, mainly collecting virtual machine in-band and out-of-band related data, while network teaching focuses on attack traffic and equipment information and the collection of government simulation learning logs and PLC forensic data.

4.2. E-Government Teaching Platform Performance.

In-band data collection is usually done with the help of a host-based e-government simulation network experimental teaching practice environment (Host-based IDS, HIDS). HIDS is composed of a central collection program and a probe implanted in a virtual machine. All interactive traffic passes through the agent program. In-band data collection can obtain the data and running status of the host and its various software. These data are used to perform attack behavior analysis. In the process of network virtualization, the industrial control government will face the dual problems of rapidity in the two links of complex network topology construction and deployment of network functions through SDN. The calculated results are shown in Table 4.

It can be seen from Table 4 that the prediction algorithm proposed in this paper is close to the Elman algorithm in accuracy and error but is better than the Elman algorithm. The gray Verhulst model is comparable to the GM (1, 1) power model algorithm and the Elman algorithm in prediction accuracy. The large gap is mainly due to the fact that the gray Verhulst algorithm selects a certain value of 2 in the determination of the power exponent, and its prediction accuracy is greatly affected by the data.

The prediction curves of the three are shown in Figure 8. It can be found that the overall prediction trend of the three is consistent with the trend of the network situation, but the prediction data obtained by the other two prediction methods do not have the accuracy of the method mentioned

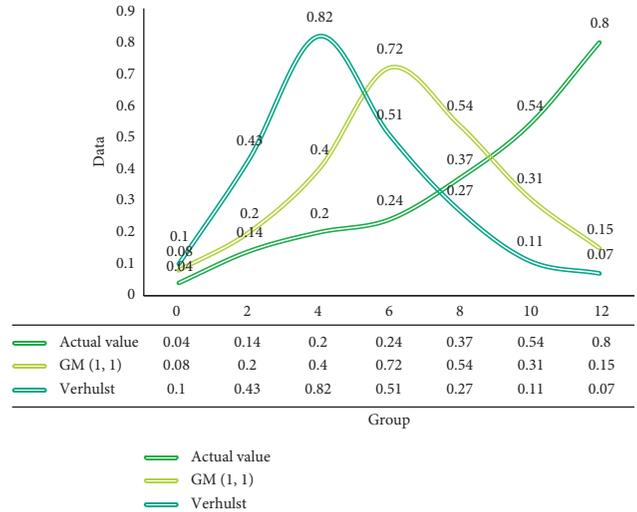


FIGURE 8: Three algorithm prediction curves.

in this paper at some nodes. The future development trend is to deploy low-latency SDN networks and large-scale virtualized PLCs on the cloud and integrate network function virtualization with SDN as a service in the cloud to realize the rapid deployment of complex heterogeneous networks. There are a large number of human-computer interactions in the corporate information network, process control network, and field control network of ICS. Therefore, the fidelity of simulating the above-mentioned interactive behaviors in the attack-oriented industrial control government environment has attracted much attention.

As shown in Figure 9, the FCM parallel algorithm based on MapReduce and the FCM parallel algorithm based on Spark are compared under the condition of processing data sets of different sizes, and the running time for each of them to complete the clustering is tested. Establishing mission-driven scenarios, studying multirole user behavior simulation with the characteristics of human and scenario interaction, and the simulation of permission acquisition that penetrates layer by layer from the enterprise information network to the field control network are the key research directions in the future. Topology replication after asset detection and various industries under the e-government simulation system cover a variety of networks, platforms, application services, and physical devices. The deployment of typical scenarios of specific industries in the industrial control government environment and network topology construction will face challenges. Based on asset detection technology, complete the entire network asset scan, copy the network topology in the industrial control government affairs, and more accurately

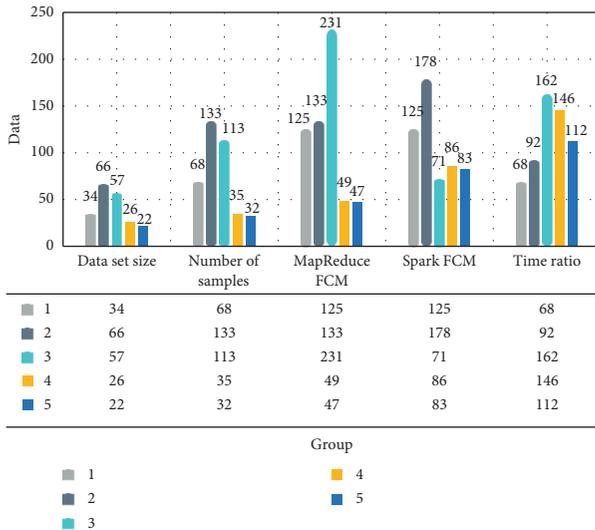


FIGURE 9: The running time of the FCM algorithm on data sets of different sizes.

copy the network topology, network equipment model, and related services of a certain scene.

The speedups of the two parallel algorithms increase approximately linearly with the increase of the number of nodes, indicating that, with the increase of the number of nodes, the execution time of the two algorithms is approximately linearly shortened, and both have good scalability. When the number of nodes is the same, the acceleration ratio of the FCM algorithm based on the Spark platform is greater than that of the FCM algorithm based on the MapReduce platform, indicating that the running time of the FCM algorithm based on the Spark platform decreases faster as the number of nodes increases, and the computational efficiency is higher.

## 5. Conclusions

After many years of development in Internet technology, server push technology as a basic technology has developed very maturely, browser applications and B/S communication mechanisms have stabilized, and there are few innovative technologies. Under this gradual development trend, this paper tries to incorporate emerging machine learning methods into the server push method, selects the three most widely used and most representative push methods, collects a certain training data set, and proposes to design a dynamic decision-making framework of server push technology based on the XGBoost algorithm, using the algorithm components in the training set to train the framework. The final dynamic decision framework shows good performance in the actual experimental environment and the comparison experiment with a single push method, but in this paper, there are still some limitations in the frame design and experimental part design. The running time, clustering quality, and parallel performance of the algorithms under the MapReduce and Spark platforms are compared, which further improves the ability of the K-means algorithm and

the FCM algorithm to process big data. In combination with the current informatization construction process in my country, Internet technology is developing with each passing day, and office and work communication methods are becoming more and more user-friendly and convenient. For this reason, the future improvements of this system include the following points: the design and implementation of an online system based on the web platform and mobile terminal. At present, this system version belongs to a stand-alone system, which requires users to download and install software for local use. In the future, it can be extended to online platforms and mobile platforms based on web pages. Users only need a computer or mobile phone and a browser to remotely log in to the system to fill in information without downloading or installing software. These new functions and new platforms will be the future research directions of this paper.

## Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The author declares no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## References

- [1] J. Chen, Z. Lv, and H. Song, "Design of personnel big data management system based on blockchain," *Future Generation Computer Systems*, vol. 101, pp. 1122–1129, 2019.
- [2] K. Shankar and M. Elhoseny, "Trust based cluster head election of secure message transmission in MANET using multi secure protocol with TDES," *Journal of Universal Computer Science*, vol. 25, no. 10, pp. 1221–1239, 2019.
- [3] Y. Ge, R. Fang, and J. Zhang, "Construction and practice of microgrid experimental platform under the background of the new engineering-taking Xuchang university as an example," *Journal of Xuchang University*, vol. 038, no. 002, pp. 153–156, 2019.
- [4] J. Xu, X. Wang, R. Zhou et al., "Development and application of virtual simulation software for experiment teaching of mining engineering," *Education Teaching Forum*, vol. 431, no. 37, pp. 118–119, 2019.
- [5] Y. Fang, "Exploration and practice of experimental teaching reform of international economics and trade based on network environment," *Zodiac*, vol. 248, no. 12, pp. 164–165, 2018.
- [6] H. Zhang and W. Fu, "Construction and practice of experimental teaching system for mechanical engineering majors oriented to engineering education accreditation," *Experimental Technology and Management*, vol. 034, no. 010, pp. 166–169, 2017.
- [7] X. Jingfen, T. Li, L. Jingjing et al., "Construction and practice of applied undergraduate environmental analysis and monitoring curriculum group," no. 1, pp. 128–130, University Education, 2017.

- [8] D. Guido, H. Song, and A. Schmeink, *Big Data Analytics for Cyber-Physical Systems: Machine Learning for the Internet of Things*, pp. 1-360, Elsevier, Amsterdam, Netherlands, 2019.
- [9] Li Hui, C. Zheng, L. Guo et al., "Case study of machine learning practice under the background of artificial intelligence," *Electronic Testing*, vol. 437, no. 08, pp. 132-133, 2020.
- [10] G. Zhang, Y. Li, H. Bian et al., "Practical education of innovation and entrepreneurship for environmental science undergraduates based on practical experiments," *Laboratory Science*, vol. 21, no. 03, pp. 152-155, 2018.
- [11] L. Yu, E. Guan, P. Wang et al., "Research and practice of "new engineering" mixed teaching mode based on informatized online classroom," *High Education Journal*, vol. 26, pp. 20-23, 2019.
- [12] J. Yang, "The application of virtual reality technology in the practical teaching of open education and tourism management," *Journal of Xinjiang Radio and Television University*, vol. 022, no. 003, pp. 15-17, 2018.
- [13] L. Liu and W. Li, "Experimental teaching research of industrial design under the background of new engineering disciplines," *Mechanical Design*, vol. 37, no. S2, pp. 302-304, 2020.
- [14] Y. Mei and L. Zhu, "The construction of cross-professional experimental courses of business management in applied undergraduate colleges and universities-Taking the internship factory in the virtual business social environment as an example," *Business Accounting*, vol. 7, pp. 118-119, 2017.
- [15] L. Huang, W. Hu, S. Zhang, and F. Li, "Practice and exploration of online teaching of environmental protection in the context of epidemic prevention and control," *Shandong Chemical Industry*, vol. 49, no. 20, pp. 224-226, 2020.
- [16] S. Yi, "Question-driven" mathematics teaching model construction and practical research for elementary education majors-Reflections on students' practical teaching based on the background of teacher qualification "national exam"," *Education Modernization*, vol. 6, no. 10, pp. 173-175, 2019.
- [17] R. Ma, X. Guan, Y. Qin et al., "Research on the construction and application of micro-class-based internship operation mechanism," *Land Bridge Vision*, vol. 20, pp. 304-305, 2017.
- [18] S. Ju, J. Sun, Li Chen et al., "Design framework of intelligent analysis platform for computer network virtual experiment under big data," *Laboratory Research and Exploration*, vol. 36, no. 12, pp. 113-115, 2017.
- [19] H. Yuan, H. Sheng, L. Zhang et al., "The system construction of hybrid teaching mode in colleges and universities in the reform of practical courses-Taking the comprehensive technical practice of agricultural resources and environment as an example," *Higher Education Magazine*, vol. 6, pp. 91-93, 2020.
- [20] L. Zhang, X. Xu, and Y. Jia, "The construction of environmental and light chemical engineering professional practice teaching system based on "innovative practice ability training"," *Heilongjiang Paper*, vol. 46, no. 04, pp. 47-49, 2018.
- [21] Y. Liu, Yi Qin, Yu Chen et al., "Application of team-based learning teaching model in pediatric practice teaching under network environment," *Naval Medicine Journal*, vol. 039, no. 003, pp. 271-274, 2018.
- [22] W. Ma, Y. Xie, Y. Wang et al., "Construction of teaching quality monitoring system for environmental engineering design comprehensive practice," *Shandong Chemical Industry*, vol. 47, no. 20, pp. 152-153, 2018.
- [23] C. Li, M. Li, J. Liu et al., "Construction and practice of experiment-practice coupling teaching mode based on scientific research cycle," *Journal of Jilin Province Education Institute*, vol. 34, no. 001, pp. 25-28, 2018.
- [24] W. Xiaoliang, "Research on the methods, paths and countermeasures of constructing practice platforms in higher vocational colleges under the network environment," *Journal of Changsha University*, vol. 031, no. 002, pp. 145-148, 2017.
- [25] H. Zhang, X. Chen, Y. Ye et al., "Construction and practice optimization of a comprehensive practice platform for "Regional Environmental and Resource Investigation"," *Chinese Geological Education*, vol. 026, no. 004, pp. 50-53, 2017.
- [26] L. Yan and X. Chen, "Analysis of the goals and needs of the nursing professional clinical practice teaching management system under the "Internet +" environment," *Journal of Qingyuan Vocational and Technical College*, vol. 10, no. 002, pp. 63-67, 2017.
- [27] D. Sun, "Construction and management of teaching practice base under the environment of agricultural science school-enterprise cooperation," *Nongjia Staff*, vol. 663, no. 15, p. 167, 2020.
- [28] N. Zhang and Y. Niu, "The construction and implementation of the classification and training mode of chemical engineering practice under the OBE concept," *Shandong Chemical Industry*, vol. 049, no. 001, pp. 173-174, 2020.
- [29] Y. Zhao, X. Tang, J. Liu et al., "Practice and thinking on the construction of high-quality resources course of applied undergraduate "Livestock environmental hygiene"," *Education Teaching Forum*, vol. 450, no. 04, pp. 343-344, 2020.
- [30] C. Helma, T. Cramer, S. Kramer et al., "Data mining and machine learning techniques for the identification of mutagenicity inducing substructures and structure activity relationships of noncongeneric compounds," *Journal of Chemical Information and Computer Sciences*, vol. 35, no. 4, pp. 1402-1411, 2018.
- [31] Q. Wang, L. Fan, and Y. Xie, "Analysis on the safety of environmental science and engineering students in colleges and universities under the background of new engineering," *Invention and Innovation (Vocational Education)*, vol. 740, no. 01, pp. 88-89, 2019.
- [32] C. Fu and Y. Luo, "The evaluation model of university theory and experiment teaching quality based on RVM machine learning method," *Modern Electronic Technology*, vol. 42, no. 13, pp. 189-194, 2019.