Open Sharing of Digital Education Training Resources Based on Machine Learning

Jichao Yan(✉) and Jingya Zheng

Zhengzhou Technical College, Zhengzhou 450121, China

Abstract. In order to solve the problem of poor security caused by data explosion in the traditional open sharing method of digital education training resources, this paper proposes a machine learning based open sharing method. Based on the national general control protocol standard, a control protocol of education and training resource sharing service based on GCCP gateway is constructed. With the support of the control protocol, support vector machine is designed by using machine learning ability, support vector machine is used to process resource data, and resource classification is realized. With the support of cloud computing mode, resource is open and shared. Experimental results show that the proposed open sharing method based on machine learning has short response time and low memory leak probability, and the security of the method is improved.

Keywords: Machine learning · Digitization · Education practice · Resource sharing

1 Introduction

Entering the 21st century, China's higher education has achieved unprecedented rapid development, entering a new era of transformation from elite education to popular education [1]. Colleges and universities have increasingly become the gathering center of a large number of talents and resources in the region, and become the main driving force and source of regional rapid and coordinated development. However, in the great development of higher education, there is a serious shortage of educational resources, especially high-quality ones [2]. The contradiction between people's demand for educational resources and insufficient supply of educational resources stands out. In this context, to promote the sharing of educational resources has become a strategic choice for deepening the sustainable development of higher education. In recent years, great achievements have been made in the construction of higher education resources sharing, but due to many reasons such as technology, demand, funds, management and system, the effect of digital education resources sharing among colleges and universities is not good, and there is still a big gap between the practical results of sharing and people's original ideas, which needs to be improved day by day [3].

The development and progress of modern network and computer technology provide the means to realize the sharing of higher education resources. In recent years, cloud

W. Fu et al. (Eds.): ICMTEL 2021, LNICST 388, pp. 142–151, 2021.
https://doi.org/10.1007/978-3-030-82565-2_12

computing is becoming more and more familiar with the people, is the current hot topic, has received extensive attention [4]. Cloud computing is regarded as the next revolution in science and technology. It is not only the development of technology, but also represents a service concept and service model. It will fundamentally change the way we work and the business model. Cloud computing technology has been initially applied in some fields and industries with its powerful functions. The popularity of network technology in various fields provides a convenient way for the sharing and exchange of information and knowledge. The Internet has become a public communication platform, in which network-based learning content can be provided by almost any media [4]. The ubiquitous 4×4 framework is the traditional teaching activity It is an important embodiment of the movement. As an extension of traditional classroom teaching, modern distance network education has developed rapidly. Online vocational skills education, distance education and online classroom have gradually become the mainstream of education at home and abroad [6, 7]. In this case, the opening and sharing of digital education and training resources is very urgent.

Digital network sharing education, as an important material means to solve the problem of educational equity and balance, is an important basis for building a learning society and realizing lifelong vocational skills education. At present, the integration development between education training and computer is still not mature enough, and there is a great room for progress. At present, the common sharing methods are the Web-based resource sharing method and the SSM -based framework sharing method [8]. These two methods need longer response time when they perform shared tasks in the face of today's data explosion. Memory leaks are easy to occur and the security needs to be further improved. Therefore, an open sharing method of digital education training resources based on machine learning is proposed to solve the problems mentioned above. It is main significance is as follows: using machine learning ability to design support vector machine, using support vector machine to process resource data, to achieve resource classification. Under the support of cloud computing mode, resources are open and shared. Its main contribution lies in the optimization of response time and memory leakage probability, which improves the security of the method.

2 Design of Open Sharing Method of Digital Education Training Resources Based on Machine Learning

2.1 Building Training Resource Sharing and Collaborative Service Control Protocol

The control protocol of digital education and training resource sharing and collaborative service is realized by common control equipment to realize intelligent interconnection, resource sharing and collaborative service of digital education and training resource [9].

According to the national standard -GB/T 29265, the general control protocol is a kind of protocol which is suitable for intelligent interconnection, resource sharing and collaborative service among shared resources. The GCCP protocol hierarchy is shown in Fig. 1.

GCCP processing unit has the general control function of sharing education and training resources, and it also handles the transmission of shared application data in the network. Point-to-point, point-to-multipoint transmission of shared data is accomplished by the physical, data link, and network layers of the GCCP protocol layer [10]. The network protocol conforming to the national standard carries the cooperation service among physical layer, data link layer and network layer of education and training resource sharing. In the general control GCCP network of digital education and training resource sharing, all protocols involved in realizing the interaction are shown in Fig. 1, which is completed by device A and device B.

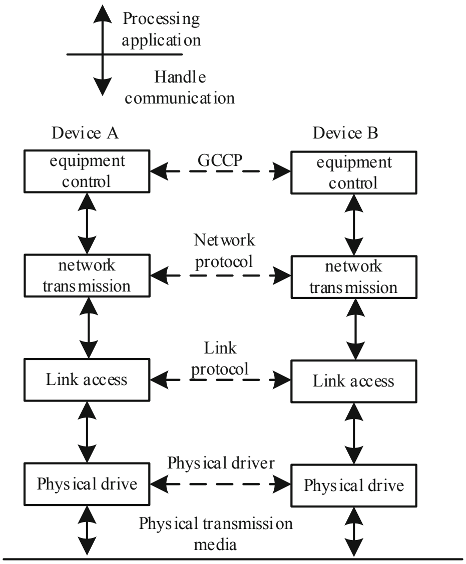


Fig. 1. Protocols for GCCP interaction design

The GCCP network is composed of general equipments, controllers, configurators, gateways and other core components. Among them, the general equipment is generally the general control equipment of the network sharing resource platform; the controller is the equipment in the sharing resource platform that operates and controls the general network equipment, such as the centralized controller and the intelligent control terminal; the configurator is mainly responsible for completing the configuration of other equipment in the sharing resource platform network, which assigns the network ID and device ID to other equipment; the gateway is mainly responsible for connecting two different networks, realizing the intelligent interconnection, resource sharing and collaborative service of the equipment in different networks; and the core component can connect the IGRS IP main network and the GCCP network, or connect the two GCCP shared networks. In a shared resource platform network, all devices are functionally configured with a single configurator, and each device in the network is configured with a separate network ID and device ID by the configurator. The network ID is composed of two octets, which are used to distinguish the different network resources of the shared resource platform. It is managed by the core component- configurator. All devices have the same network ID. The device ID consists of two octets to distinguish between different network devices. Its working ID is assigned by the component- configurator, and all devices in the same network environment have different device IDs. The usage assignments for the network ID and device ID are shown in Table 1.

Table 1. Distribution of ID usage

Network ID	Purpose	Device ID	Purpose
0x0000	Unknown network ID, used before the configurator assigns the network ID, or for other purposes	0x0000	Unknown network ID, used before the configurator assigns the network ID, or for other purposes
0x001–0xFEFF	Network ID	0x0001–0x0EFF	General device ID, assigned to general equipment
0 xFF00–0xFFFF	Retain	0 x0F00–0x0FFF	Controller ID, assigned to the controller for use

The file request device issues a "file transfer request" for the transfer of resource files in the shared resource platform, and the device holding the shared resource realizes the transfer of shared file data. The request device for sharing file data can request the holding device to transfer the file from any point. During the file transfer, it can start, pause or stop the file transfer according to the actual situation. The concrete work process by the file request device through the file total length and already obtained the file data quantity to judge whether the file transfers completes. As shown in Fig. 2, the specific transfer of a file's data is shown.

In the process of file transfer, the data attribute is mainly used to describe the transfer device, device status, device list, user-defined and other files.

2.2 Shared Information Classification Based on Machine Learning

Machine learning is a kind of learning method in which a machine can avoid contacting its internal complex logic and relationship by memorizing learning data. Through this learning method, a classifier, i.e. a support vector machine, can process the shared information, classify it reasonably, and transmit it in segments according to the classification results, so as to ensure the efficiency and security of transmission. Assuming that the resource sample has only two classes A and B, there is a linear function:

$$f(x) = \omega x + b \tag{1}$$

If the threshold value is 0, the samples a and B can be separated. When $f(x) > 0$ the threshold value is set as A, when $f(x) < 0$ the category is determined as B, then $f(x)$ it belongs to the discrimination function of the sample. In order to judge whether a

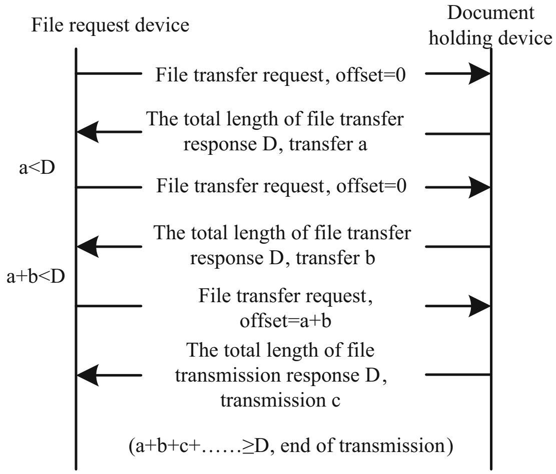


Fig. 2. A file transfer process

linear function is good or bad, the concept of classification interval is introduced. The so-called classification interval is the geometric distance between sample classes. Then, the best classification effect is to maximize the classification interval.

To make the classification more accurate, consider maximizing the distance from the hyperplane to the nearest point. The greater the distance, the less prone the classification is to error. Suppose you have found a hyperplane T_0 whose distance from the sample A is greater than that from the sample B. Thus the point closest to the hyperplane is the point of the sample B. Obviously, this distance is not the largest. The largest case is that the distance L_x from the hyperplane T_0 to the nearest point on b the distance L_y from the hyperplane T_0. The sample closest to the hyperplane is x, and the sample closest to the sample B. If the hyperplane is optimal, the condition must be satisfied: the distance from x should be equal to the distance from y. In order to understand the distance between classes more vividly, two new sums of hyperplanes T_1 and T_2 parallel to hyperplanes T_0. T_0 are introduced to cut the samples X and Y of the nearest two samples. Then the interval between T_1 and T_2 is called the interval between the sample classes, that is, the classification interval. The question then becomes how to maximize the distance between the sums T_1 and T_2.

Let's say that the vector x_i of a sample y_i is labeled minus 1 and plus 1, and the sample class y_i is labeled minus 1 or plus 1 in a binary linear space. Then the sample can be recorded as:

$$K_i(x_i, y_i) \tag{2}$$

For hyperplanes, the distance from this sample to the hyperplane is defined as:

$$\eta_i = y_i(\omega x_i + b) \tag{3}$$

The classification interval is maximized when the η_i calculated value is maximum. At this point the classification is complete.

2.3 Achieve Resource Open Cloud Sharing

According to the classified data, the cloud sharing mode is used to realize the open sharing of resources. Due to the size of the network and resource constraints needed, cloud computing is mainly promoted and implemented by several large companies, and most of it is used for commercial purposes, and only a few education projects are still in place. Therefore, if we directly adopt the cloud computing solution provided by the existing service providers, although the short-term cost is low and the implementation is simple, the migration of the original educational resources and the processing of the original computing capacity of the colleges and universities will become a major problem, which will cause a huge waste of resources. If these issues are not considered, reinvent the wheel to invest in cloud computing sharing systems, the initial construction costs are very expensive, the average small and medium -sized institutions can not afford. In this way, the advantages of cloud computing in the field of education can not be reflected, it is difficult to promote the application. Based on the comprehensive analysis, it is a feasible choice to deploy the cloud after reforming the existing education network. The current feasible measures are to transform the original campus network architecture, optimize and upgrade the cloud architecture, and build a cloud environment.

Concretely, it is to transform the B/S or C/S model into cloud, encapsulate or transform the existing education and training resources in a service-oriented way, and centralize the scattered servers and computing resources. Then use the virtualization technology to form the digital educational resources pool on the Internet. Provide different terminal interface, originally only through the fat client access to the Internet resources in the cloud computing environment can access resources through a variety of thin client.

In order to realize the sharing of educational practical training resources, we should not only use our own resources, but also broaden our view from the field of education to the whole society. With the enhancement of people's lifelong learning consciousness, the demand for digital educational resources is also increasing. The original distance network education is based on academic education, distance education system through learning and access to resources. However, as a student finishes a course, he can no longer obtain resources from the distance education system for the course, which is detrimental to autonomous learning and lifelong learning.

By sharing the cloud computing model, social students can also easily access learning resources through simple verification. Since the Digital Education Resources Cloud is jointly constructed by all the educators and shared by their contributions, some of the resources provided by the community students can be used for free and for a fee. Billing modules can be added to cloud application deployments. In this way, social learners can also use digital educational resources on demand. So the school resources as a part of the whole social resources, through further improving the openness to society this one way to improve the efficiency of running schools. To benefit more people.

3 Experimental Study on Open Sharing Method of Digital Education Training Resources Based on Machine Learning

In order to verify and evaluate the proposed open sharing method of digital education training resources, the experimental study will extract data from two open, real and widely used datasets. The dataset used in the experiment was extracted through the Java-based PcapParser program, and the sharing method was developed on the basis of the open-source Java machine learning tool Weka 3.8. The experiment was performed on an Intel Core i5-3470 3.20 GHz 8 GB Windows system (64-bit).

3.1 Experimental Data Set

The two open, real and widely used datasets used in the experiment are WITS and MAWI. The details of the two datasets are as follows:

The WITS dataset is the WAND Research Group of the Department of Computer Science at the University of Waikato, which collects and collects Internet resource data logs in New Zealand using the WITS project. The data of Waikato VIII is captured from the border network of Waikato University by the software of DAG3 card and WDCap, and is anonymously processed and truncated by the software of Crypto-Pan. A detailed description of the selected traffic dataset is shown in Table 2.

Table 2. WITS information resources dataset details

Interception date	Duration/h	Number of messages/million	Total/MB	Capture size/MB
0122	24	331	198107	25631
0241	12	135	80553	10400
0314	24	360	219693	27943
0452	24	266	163200	20511
0546	24	266	162967	20354
0635	12	120	76741	9243

MAWI traffic dataset is the network traffic dataset that the MAWI working group captures and collates from WIDE network. The MAWI project team persists in capturing network traffic from WIDE network for a certain time every day. The IP addresses of the traffic in the dataset are anonymously processed, and each packet contains only 50 bytes of application-layer payload. The details of the data intercepted in the experiment are shown in Table 3.

First of all, the PcapParser program based on Java is used to process the two traffic datasets, extract the sample data respectively, and annotate the instance of the dataset using the open source nDPI deep packet detection tool.

3.2 Shared Response Time Experiment and Analysis

In order to evaluate the performance of the proposed open resource sharing method, two other open resource sharing methods are selected to carry out experiments under the same experimental conditions. One is based on the web, the other is based on the SSM framework. The experimental results are shown in Fig. 3.

Table 3. MAWI traffic dataset details

Interception date	Duration/S	Number of messages/million	Capture size/MB
0124	3598.32	275	14452.21
0236	3600.03	368	2055.52
0341	3600.4	364	21041.86
0425	3600.35	384	21669.36
0529	3600.21	391	22100.32
0632	3599.9	412	24625.77

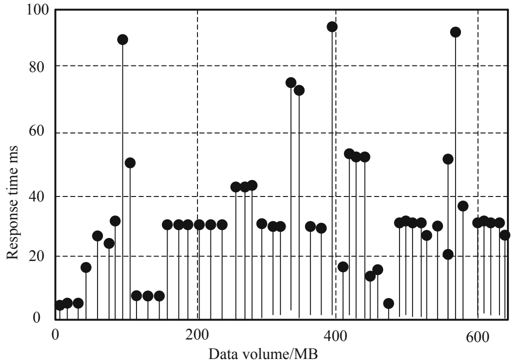
Compared with the results in the graph, it can be seen that in the experimental results of the sharing method based on the web, the sharing response time changes irregularly with the increase of the data amount, and most of them are at or above 20 ms; in the experimental results of the sharing method based on the SSM, the sharing time increases gradually with the increase of the data amount, and exceeds 20 ms when the data amount reaches 200 MB; in the experimental results of the sharing method based on the machine learning, the response time is always at a lower level with the increase of the data amount, and does not exceed 20 ms. To sum up, the designed open sharing method of digital education training resources based on machine learning has faster response time.

3.3 Memory Leak Probability Experiment and Analysis

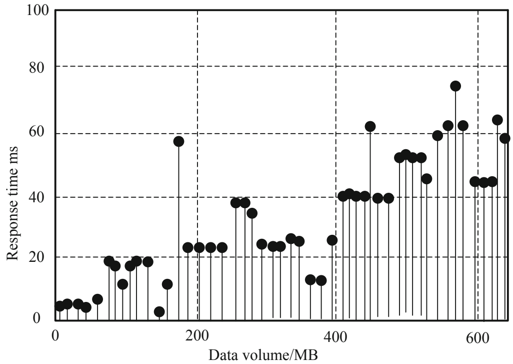
In the memory leak probability experiment, LeakCanary, a memory leak analysis tool, is used to analyze the memory leak when the computer executes the shared task, and calculate the memory leak probability. Memory leak is an object reference error in the platform, which leads to the object occupying heap resources can not be reclaimed, and memory space is wasted seriously.

Use the LeakCanary tool to automate the detection of memory leaks in shared methods, introduce a dependency library in build.gr adle, initialize global configuration, install an application called Leaks locally after initialization to log information, and if memory leaks are detected, locate the error message and send it to the notification, and log it in Leaks. After a complete task is completed, calculate the probability of memory leaks. The experimental results are shown in Table 4.

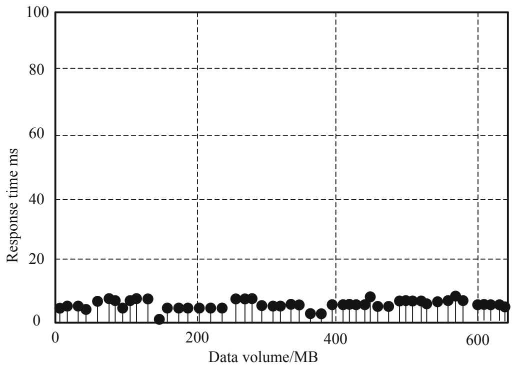
From the data in the table, we can see that the memory leak probability of the resource sharing method based on machine learning is much lower than that of the other two methods. According to the results of the shared response time experiment, the designed open sharing method of digital education training resources based on machine learning has shorter response time, lower memory leakage probability and higher overall security.



(a) Experimental results of a web-based resource-sharing approach



(b) Experimental results of the resource-sharing approach based on the SSM framework



(c) Experimental results of a resource-sharing approach based on machine learning

Fig. 3. Shared response time experimental results for different sharing methods

Table 4. Experimental results of memory leak probability for different shared methods

Shared resource size/M	Sharing method based on web	Sharing method based on SSM framework	Sharing method based on machine learning
500	9.93%	9.24%	0.52%
1024	19.24%	15.62%	1.26%
5012	33.69%	29.45%	2.31%
10240	41.42%	36.22%	5.62%
20480	56.35%	47.85%	9.41%

4 Closing Remarks

In this paper, an open sharing method of digital education training resources based on machine learning is studied and designed. After the design is completed, several comparative experiments are designed. However, due to the limitation of personal ability and time, there are some deficiencies in the research, and the efficient computing ability of cloud computing has not been fully exerted.

References

1. Zhong, W., Li, Z.: Research on network education system based on learning machine. J. Commun. **39**(01), 135–140 (2018)
2. Yang, W.: Simulation of remote sharing method of database information under architecture of the Internet of Things. Comput. Simul. **35**(04), 457–461 (2018)
3. Sun, Y., Zhao, S., Zhang, F., et al.: A comparative study on the security mechanism of open & sharing government date information in China America and Britain. Library Inf. Serv. **62**(21), 5–14 (2018)
4. Li, J.: Research on method of information organizing and data sharing model of characteristic cultural resources: taking the project of the ancient dwellings digital memory of world Hakka City as example. Libr. J. **12**(05), 39–44 (2018)
5. Zhang, N., Chen, B., Yang, Y., et al.: Optimization technology of optical fiber communication network based on service classification. J. Phys.: Conf. Ser. **1746**(1), 012085 (6pp) (2021)
6. Schwartz, L.N., Shaffer, J.D., Bukhman, G.: The origins of the 4 × 4 framework for noncommunicable disease at the World Health Organization. SSM Popul. Health **13**(8490), 100731 (2021)
7. Fu, W., Liu, S., Srivastava, G.: Optimization of big data scheduling in social networks. Entropy **21**(9), 902 (2019)
8. Liu, S., Li, Z., Zhang, Y., et al.: Introduction of key problems in long-distance learning and training. Mobile Netw. Appl. **24**(1), 1–4 (2019)
9. Liu, S., Liu, D., Srivastava, G., Połap, D., Woźniak, M.: Overview and methods of correlation filter algorithms in object tracking. Complex Intell. Syst. 1–23 (2020). https://doi.org/10.1007/s40747-020-00161-4
10. Meng, X., Ma, C., Yang, C.: Survey on machine learning for database systems. J. Comput. Res. Dev. **56**(09), 1803–1820 (2019)