Application of Teaching Quality Assessment Based on Parallel Genetic Support Vector Algorithm in the Cloud Computing Teaching System

Haifeng Hu and Junhui Zheng*
Pingdingshan University, Pingdingshan, Henan, China

Abstract—**With the rapid development of China's economy in recent years, the scale of students has expanded gradually, which has led to many new problems, including the problems of the quality and the quantity of teachers, and the teaching facilities being insufficient. The assessment of teaching quality is one of the most important aspects of teaching management, which come to the attention of every university. Therefore, it has become the current focus in the research of university teaching. At the same time, the traditional method of teaching quality assessment has not been able to deal with the phenomenon of big data in the field of education. As a new technology, cloud computing provides a broad space for the development of a new model in the aspects of hardware environment construction, software resource development, network teaching implementation and personal knowledge management. In order to effectively deal with the challenges of big data processing in the field of education, this paper proposes a GA-SVM teaching quality assessment algorithm which is based on MapReduce. Through the design of a map function and reduce function, this paper realizes the parallelization of the GA-SVM algorithm and the selection of the main parameters. Secondly, this paper uses a genetic algorithm to optimize the penalty coefficient C and kernel parameters σ^2 of SVM, and then solves the problem of difficulty in determining the parameters of support vectors. In addition, we improve the sensitivity of the search through the method of logarithmic transformation, and speed up the convergence rate of the GA model. Finally, we compare the parallel algorithm and the serial algorithm on the Hadoop platform. The results of experiments show that the GA-SVM based on MapReduce is suitable for teaching quality assessment under the environment of big data.**

Index Terms—**Cloud computing; Supplementary teaching; Genetic algorithm; MapReduce**

I. INTRODUCTION

With the rapid development of China's economy in recent years, the scale of students has expanded gradually, which has led to many new problems of the quality and the quantity of teachers, and the teaching facilities being insufficient. The assessment of teaching quality is one of the most important aspects of teaching management, which has come to the attention of every university. Therefore, it has become the current focus in the research of university teaching. At the same time, the traditional method of teaching quality assessment has not been able to deal with the phenomenon of big data in the field of education. As a new technology, cloud computing provides a broad space for the development of a new model in the aspects of hardware environment construction, software resource development, network teaching implementation and personal knowledge management.

In view of the problems of teaching quality assessment, many scholars have conducted thorough research and put forward many effective methods. The traditional methods of teaching quality assessment mainly include linear regression, partial least squares, multivariate statistical analysis, grey relational analysis and analytic hierarchy process [1-4]. These methods assume that there is a linear connection between the teaching quality and assessment index. However, the relationship between the assessment index and teaching quality are nonlinear. It is difficult to accurately describe the problem by using the linear model, which leads to the significant difference between the assessment results and the actual value [5,6]. In recent years, data mining technology has matured, and there are clustering analysis method, fuzzy method, support vector machine, neural network and other nonlinear methods of teaching quality assessment [7]. These data mining methods can well capture the mapping between the input and output, and improve the accuracy of teaching quality assessment. A support vector machine is a kind of machine learning method for small samples, nonlinear and high dimensional data. Compared with other learning methods, it has better generalization capability, and is the most widely used method in data mining.

The research of the support vector machine algorithm, it mainly focuses on the following aspects. (1) Decomposition algorithm. It usually takes much time to deal with big data, so we need to introduce the decomposition algorithm to improve the efficiency of the algorithm. The commonly used decomposition algorithm has Chunking algorithm [8], fixed working sample set algorithm [9], and SMO algorithm [10]. (2) Online learning algorithm. It is divided into online incremental learning and online reduced learning [11-13]. When the new data is added to the training data set, the method does not need to re-train the support vector machine model, so as to simplify the process of training purposes. (3) Multi-class classification algorithm. Support vector machine algorithm is usually used to solve the problem of two-classes. The multi-class classification is a generalization of the support vector machine for two-

classes [14]. (4) Privacy protection algorithm. With the emphasis on privacy issues, the privacy protection support vector machine is proposed. This method can be used to train the data, and it can get a more accurate calculation of the support vector machines, and also does not disclose the privacy of the parties [15-19].

To sum up, in order to effectively deal with the challenges of big data processing in the field of education, this paper proposes a GA-SVM teaching quality assessment algorithm which is based on MapReduce. Through the design of a map function and reduce function, this paper realizes the parallelization of the GA-SVM algorithm and the selection of main parameters. Secondly, this paper uses a genetic algorithm to optimize the penalty coefficient C and kernel parameters σ^2 of SVM, and then solves the problem the difficulty of determining the parameters of support vectors. In addition, we improve the sensitivity of the search through the method of logarithmic transformation, and speed up the convergence rate of the GA model. Finally, we compare the parallel algorithm and the serial algorithm on the Hadoop platform. The results of experiments show that the GA-SVM based on MapReduce is suitable for teaching quality assessment under the environment of big data.

II. CLOUD COMPUTING PLATFORM OF TEACHING QUALITY ASSESSMENT DATA BASED ON HADOOP

Next, we introduce the architecture of the cloud computing education platform. The platform is composed of a data layer, model layer and application layer.

Data layer: The data layer is a bottom operating system, responsible for the data storage and access. For example, the GFS is used as the Google file system, and the HDFS is used as an open source Hadoop file system.

Model layer: Computing power is an important indicator of cloud computing. The cloud computing platform must provide a simple and convenient calculation model to ensure that the computing power is high quality and has high reliability. The cloud platform computing model belongs to the category of parallel computing. Since the cloud computing data center is very concentrated, there is no problem of node failure caused by MPI. At present, the MapReduce model is usually used in the cloud computing model.

The application layer provides various teaching related application software. This layer mainly contains:

(1) A platform for teaching management, academic management system and office system;

(2) A teaching quality assessment system, examination system and score inquiry system;

(3) Document processing software, courseware planning and presentation software;

(4) A virtual computing environment;

(5) A virtual computing laboratory based on cloud computing.

III. RELATED BASIC KNOWLEDGE

A. SVM model

The support vector machine is derived from statistical learning theory, and it is a kind of supervised learning method. Its basic idea is to transform the input space to a high dimensional feature space by nonlinear transformation, and then to find the optimal linear classification surface in the new space. Among them, the nonlinear transformation is realized by the kernel function. Therefore, the selection of the kernel function is the key to the SVM algorithm. The basic principle of SVM is shown in Figure 1.
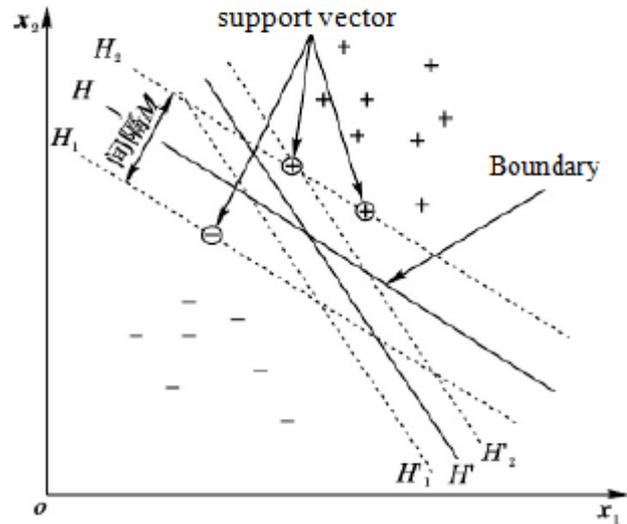

Figure 1.   Schematic diagram of SVM

Where, "+" and "-" represent two kinds of training samples respectively, x_1 and x_2 are the two characteristic items of the sample, and H is a boundary. The SVM model must ensure that the risk is minimized, so the optimal classification surface not only requires the correct classification of the two types of data, but also requires the maximum classification interval. The purpose of this is to make the support vector machine have better generalization capability. When the SVM algorithm is used to estimate the regression function, it maps the data x of the input space to a high dimensional feature space by a nonlinear mapping Φ, and then the linear regression is done in the high dimensional space.   Assume a set of data points $G = \left\{ (x_i, y_i) \right\}_{i=1}^{n}$. Among them, x_i is the input vector, y_i is the expected value, and n is the total number of data points.

SVM uses the following formula to estimate the function:

$$f(x) = (w * \Phi(x)) + b \qquad (1)$$

Where, $\Phi(x)$ is a nonlinear mapping from the input space to the high dimensional feature space, and the coefficients w and b are estimated by minimizing the following formula:

$$R(f) = \frac{C}{n} \sum_{i=1}^{n} L_\varepsilon (y_i, f(x_i)) + \frac{1}{2} \|w\|^2 \qquad (2)$$

Where, L_ε is a loss function, and ε is insensitive loss function.

Because the ε is insensitive loss function, formula (2) can be translated into the following two convex programming problems.

$$\min\left\{\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*)\right\} \tag{3}$$

Its constraint conditions are:

$$\begin{cases} w\Phi(x_i) + b - y_i \leq \xi_i + \xi_i^* \\ -w\Phi(x_i) - b + y_i \leq \xi_i + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \tag{4}$$

Where, $i = 1,2,\text{L},n$, ξ_i and ξ_i^* are slack variables.

Thus, the predicted output can be obtained as:

$$y(x) = \sum_{i=1}^{l} a_i \Phi^T(x)\Phi(x) + b \tag{5}$$

Using the kernel function method, the following is obtained:

$$K(x,x_i) = \Phi^T(x)\Phi(x_i) \tag{6}$$

Finally, the nonlinear prediction model is:

$$y(x) = a_i K(x,x_i) + b \tag{7}$$

B. GA model

A genetic algorithm is a parallel random search optimization method which simulates the genetic mechanism and biological evolution in nature. The biological evolution theory of survival of the fittest is introduced to the series group encoding by this algorithm. Through the fitness function, selection, crossover and mutation it screens the individual and constitutes a new group. As a result, the new group inherits the information of the last generation, and is superior to the previous one. repeating this process, the individual fitness of the population is continuously improved until it meets certain conditions. The genetic algorithm is simple and practical. It can be processed in parallel, and can generate a global optimal solution. The penalty coefficient C and kernel parameter σ^2 are the main parameters that affect the performance of the SVM model. The set of main parameters in this paper are as follows:

1) Genetic encoding rules.

Taking into account the influence of parameters C and σ^2 on the generalization performance of SVM, we set the parameters C and σ^2 in the search range (0.01-1000) and (0.01-100) respectively. Then, in order to improve the sensitivity of the search process, we have a logarithmic transformation of the parameters, and obtain the $\lg C$ and $\lg \sigma^2$. At this time, the search scope is transformed to (-2-3) and (-2-2). Then, we use the binary encoding string with a length of 10 bits to represent two decision variables which are $\lg C$ and $\lg \sigma^2$ respectively. So, there are 1024 different discrete points in the domain. Finally, we connect two binary encoding strings, so as to form a binary encoding string with a length of 20 bits. This is the encoding method of this article.

2) The determination of fitness function.

The prediction accuracy of the support vector machine is the final standard to measure the model, so this paper takes the prediction sample's root mean square error as the fitness function.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{8}$$

Where, $\hat{y}_i$ is the predictive value, y_i is the actual value, and n is the number of test samples.

IV. THE PARALLEL GA-SVM MODEL BASED ON MAP-REDUCE

Next, we introduce the basic idea of SVM teaching quality assessment algorithm based on Map-Reduce. Then, the SVM algorithm training is used for each subset in the Map nodes to solve their SV. Finally, the global SV is obtained from the reduce node, and then the final SVM model is trained. Its specific architecture is shown in Figure 2.
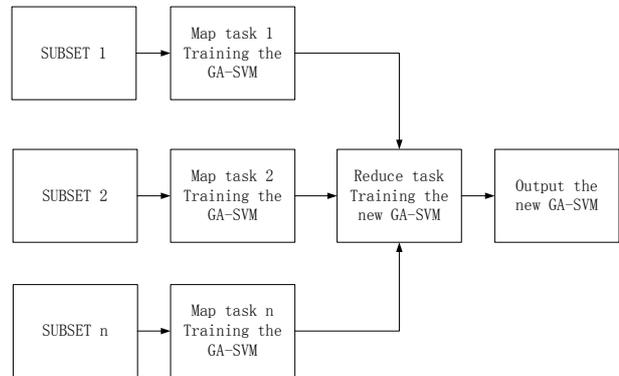

Figure 2.　SVM algorithm based on MapReduce

The process of the whole algorithm is as follows:

Step 1: $(key, value)$ is used to mark each subset, and (k_1, v_1) is used as the input of the map node, where, key is the class of sample subsets, and $value$ is the feature item of the sample data.

Step 2: After the map function processing, the output of the intermediate results is (k_2, v_2), where, key is the class of sample subsets, and $value$ is the SV of subset.

Step 3: In the phase of reduce preprocessing, we put the same kind of SV together $(k_2, list(v_2))$ and put it into the reduce node.

Step 4: Finally, by the reduce function processing, the output is $(k_3, list(v_3))$, where, key is the class of the final support vector machine, and $value$ represents the relevant parameters of the support vector machine.

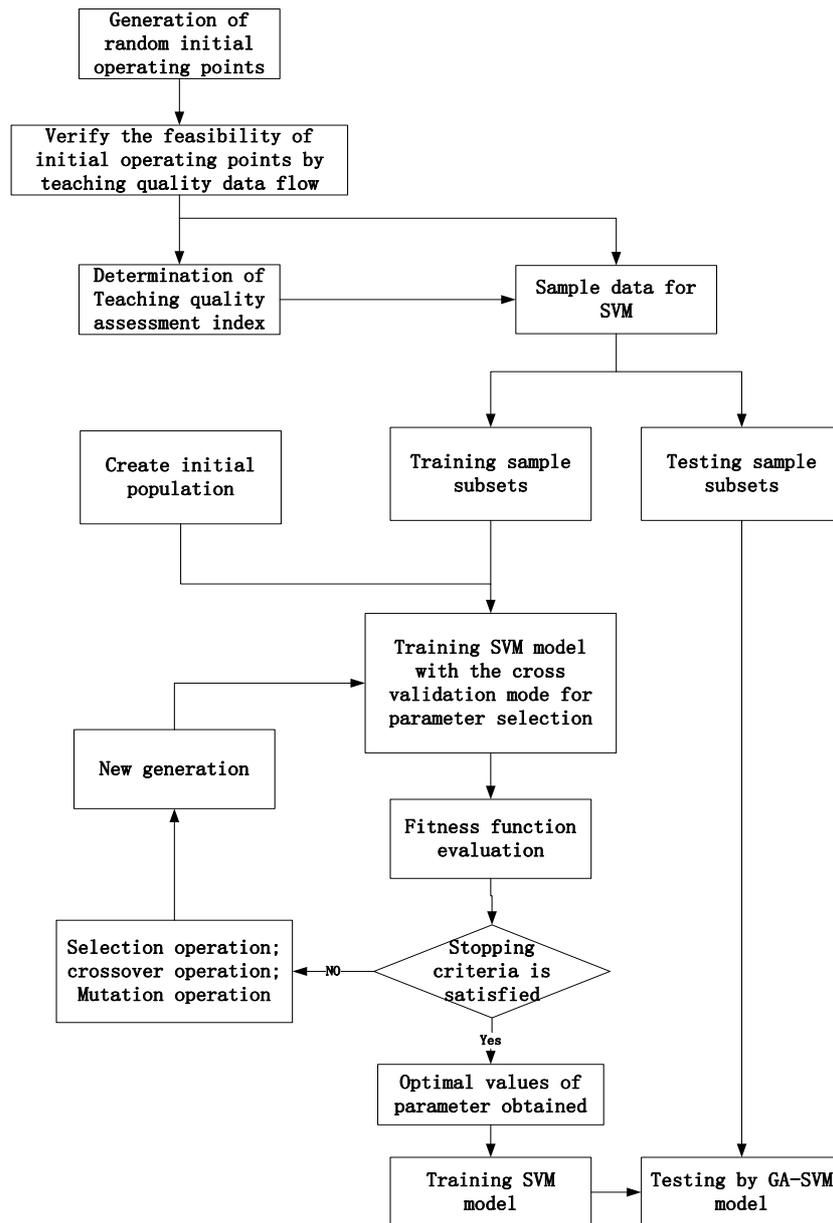The architecture diagram of GA-SVM model is shown in Figure 3.

Figure 3. The architecture diagram of GA-SVM model

V. SIMULATION EXPERIMENT AND RESULT ANALYSIS

Firstly, we establish the index system of teaching quality assessment. The system of teaching quality assessment is a multi level and multi index problem. In this paper, the teaching quality evaluation system is set up with seven indexes. They are the quality of teachers, teaching attitude, teaching content, teaching methods, teaching design, teaching ability and teaching effect. The importance of each indicator is different. Their specific explanations are as follows:

(1) The quality of teachers: The standard of mandarin.

(2) Teaching attitude: Attendance record.

(3) Teaching content: The scientific nature and practicability of knowledge; cultivation of learning capability.

(4) Teaching methods: The guidance of learning methods; instructiveness of the educational method; the application of multimedia teaching.

(5) Teaching design: The setting of teaching goal; design of teaching strategy; design of teaching process.

(6) Teaching ability: Thorough analysis of the problem; accurate experimental operation.

(7) Teaching effect: Student feedback; students detection; implementation of teaching objectives.

Whether or not the assessment system is reasonable and reliable depends on various factors. It is the basis for guiding students to grade the factors. Because each participant has a certain degree of subjective randomness in the scoring, the mean value method is used in this article. Then, the assessment result is more authoritative and reliable. According to the comprehensive principle of teaching quality assessment, different assessment results need to be assigned different weights to calculate the comprehensive assessment score. Next, we obtain a list of the original data. Because of the large amount of data, we only give a fragment of the original data list, as shown in Table 1.

TABLE I.
SAMPLE DATA OF EVALUATION INDEX

No	Teacher's quality	Teaching attitude	Teaching content	Teaching methods	Teaching design	Teaching ability	Teaching effect	Teaching quality
1	6.29	7.75	8.04	7.55	7.12	7.39	8.81	7.78
2	7.05	8.91	9.03	6.06	9.06	8.39	7.78	8.04
3	8.14	6.61	7.03	6.40	8.69	6.18	7.46	7.22
4	8.94	6.54	8.43	6.54	7.37	7.37	6.66	7.41
5	9.34	6.21	7.49	8.58	8.21	7.42	9.27	8.07
6	6.62	6.25	6.71	6.12	7.76	9.19	6.63	7.04
7	6.47	7.63	8.60	8.96	8.33	6.98	8.92	7.98
8	8.59	8.71	6.01	7.72	6.63	7.84	6.74	7.46
9	7.44	9.45	8.74	6.29	7.59	6.77	7.87	7.74
10	8.03	6.45	6.58	8.30	7.23	7.97	9.11	7.67
M								
997	7.61	7.80	7.24	8.72	6.10	8.71	7.42	7.66
998	7.54	8.65	8.44	8.20	8.72	8.81	8.21	8.37
999	6.17	8.74	9.31	8.77	8.07	7.91	9.48	8.35
1000	6.29	7.75	8.04	7.55	7.12	7.39	8.81	7.78

Since the dimensions of the various indicators are different, a direct comparison cannot be made. In order to make the index have comparability, and to speed up the convergence rate of the GA model, this paper has conducted the normalized processing to each index.

Then, we give the parameter table of GA-SVM algorithm. After setting the parameters, we can use the parallel GA-SVM algorithm to evaluate and predict the quality of teaching.

In order to verify the performance of the high performance computing system in dealing with big data, this paper uses the Hadoop cluster environment made up of 10 IBM servers, each server being a quad core PC. One of them is used as NameNode, JobTracker and DataNode. The rest of them are used as the DataNode and Task-Tracker. Then, we take different data sets as the sample data sets of the parallel GA-SVM algorithm, and compare the data processing capability of serial GA-SVM and parallel GA-SVM. The experimental results are shown in Figure 4.

As can be seen from Figure 4, the total task processing time of the two GA-SVM algorithms in a same node dimension are different. The total task processing time of the parallel GA-SVM algorithm is less than the traditional GA-SVM algorithm. In addition, with the gradual increase of nodes, the total task processing time has gradually become different between the two algorithms. Although with the increase of nodes, the two algorithms can deal with more data, but the parallel GA-SVM algorithm performs better.

After we have proved the performance capability of this algorithm, we use four different methods to predict the teaching quality with 1000 samples. The experimental results are shown in Figure 5.

It is not difficult to see from Figure 5, this algorithm is very good for the prediction of the teaching quality, and its accuracy is much higher than the other three methods.

TABLE II.
THE GA-SVM ALGORITHM PARAMETER SETTING

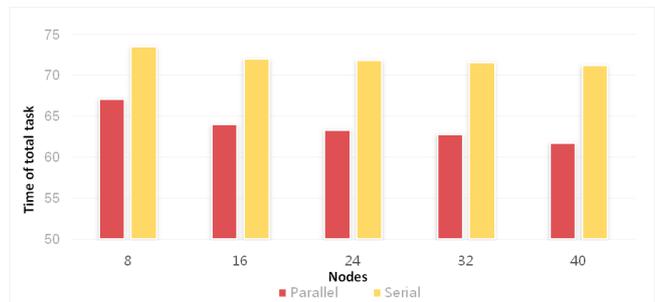| Parameters	Settings
C	0.01-1000
σ^2	0.01-100
ε	0.001-0.01
Generations	100
Population size	1000
Selection type	Standard roulette wheel
Crossover type	Simulated binary
Mutation type	Polynomial method
Crossover probability	0.85
Mutation probability	0.06

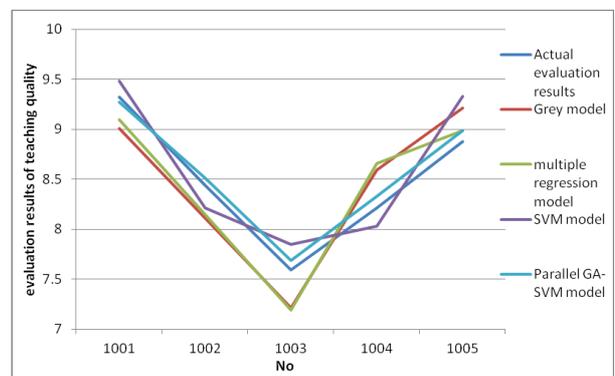Figure 4. Data processing capability with different nodes

Figure 5. Evaluation results of teaching quality

VI. CONCLUSION

In order to effectively deal with the challenges of big data processing in the field of education, this paper proposes a GA-SVM teaching quality assessment algorithm which is based on MapReduce. Through the design of map function and reduce function, this paper realizes the parallelization of the GA-SVM algorithm and the selection of the main parameters. Secondly, this paper uses a genetic algorithm to optimize the penalty coefficient C and kernel parameters σ^2 of SVM, and makes the parameters of support vectors more easily determined. In addition, we improve the sensitivity of the search through the method of logarithmic transformation, and speed up the convergence rate of the GA model. Finally, we compare the parallel algorithm with the serial algorithm on the Hadoop platform. The experimental results show that the GA-SVM based on MapReduce is suitable for teaching quality assessment in the environment of big data.

REFERENCES

[1] Sun Xiaoling, Wang Ning, Liang Yan. Teaching evaluation model and simulation based on BP neural network [J]. computer simulation, 2010,27 (11): 314- 317.

[2] Yang Xiuping. Research on the teaching quality evaluation model based on AIHP [J]. experimental technology and management, 2009, 26 (9): 138-141.

[3] Hong Yuehua. Classroom teaching quality data mining based on fuzzy comprehensive evaluation [J]. computer science, 2008,35 (2): 154-156.

[4] Suzyna, Zhou Shiguan. Application of fuzzy neural network in the evaluation of classroom teaching quality [J]. computer simulation, 2008,25 (5): 287-291.

[5] Ma Hong. Evaluation of teaching quality by using grey trend correlation method [J]. Journal of Wuhan University of Technology, 2010, 32 (15): 181 -185.

[6] Du Chunyan, Wang Zhengwu, Zhang Ruiping. A comprehensive evaluation of the quality of classroom teaching based on neural network [J]. College mathematics, 2005, 21 (3): 1-5.

[7] Xu Gaohuan. The application of SVM in the evaluation of teachers' teaching quality [J]. Journal of Zhejiang Water Conservancy and Hydropower College, 2007, 19 (1): 66-68.

[8] B. Boser,I. Guyon,V. Vapnik. A training algorithm for optimal margin classifiers[C]. Pittsburgh: ACM Press, 1992. http://dx.doi.org/10.1145/130385.130401

[9] E. E. Osuna. Support vector machines: training and applications[D]. MA: MIT University, 1998.

[10] J. C. Platt. Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines [D]. MA: MIT Press, 1999.

[11] Duan Hua. Research on incremental learning algorithm of support vector machine [D]. Shanghai: Shanghai Jiao Tong University, 2008

[12] N. A. SYED,H. Liu ,K. K. Sung. Incremental learning with support vector machines[J]. In Proc. Int. Joint Conf. on Artificial Intelligence(IJCAI － 99) ,1999: 196- 215.

[13] G. Cauwenderghs,T. Poggio. Incremental and detrimental support vector machine learning [M]. MA: MIT ,Press,2001.

[14] Hou Weizhen. Research on several optimization algorithms for solving the support vector machine [D]. Qingdao: Shandong University of Science and Technology, 2007.

[15] Hwanjo Yu,Xiaoqian Jiang, Jaideep Vaidya. Preserving SVM using nonlinear kernels on horizontally partitioned data[C]. New York: ACM, 2006.

[16] Hwanjo Yu,Jaideep Vaidya, Xiaoqian Jiang. Preserving SVM classification on vertically partitioned data [C]. Berlin: Springer, Verlag, 2006.

[17] O. L. Mangasarian, E. W. Wild, G. M. Fung. Preserving classification of vertically partitioned data via random kernels[J]. ACM Transactions on Knowledge Discovery from Data, 2008,2(3) : 473- 479. http://dx.doi.org/10.1145/1409620.1409622

[18] Shao Fubo. Research on support vector machine model and algorithm for privacy protection [D]. Qingdao: Shandong University of Science and Technology, 2009.

[19] Hu Yunhong. Research on some algorithms of support vector machine [D]. Qingdao: Shandong University of Science and Technology, 2011.

AUTHORS

Haifeng Hu. He was born in February 1983. He is a lecturer with a master's degree. He graduated from Zhengzhou University and received his bachelor's degree in 2005. Later, he received his master's degree in Xi 'an Electronic and Science University in 2010. He currently works at the College of Information Technology, Pingdingshan University, Pingdingshan, Henan 467000, China

. He is engaged in computer teaching. His main research direction is **information security.**

Junhui Zheng. He received his M.Sc. in Information Sciences (2010) from University. Now he is a lecturer of informatics at College of computer science and technology, PDS University, Pingdingshan, Henan 467000, China. His current research interests include Artificial Intelligence and Control Techniques.