Scholar Publishing Group
Machine Learning Theory and Practice
https://doi.org/10.38007/ML.2021.020205
ISSN 2790-0983 Vol. 2, Issue 2: 46-65

Postgraduate Teaching Reform Based on Machine Learning and Improved SVM Algorithm

Malik Alassery*

Amman Arab University, Jordan

**corresponding author*

Keywords: Machine Learning, SVM Algorithm, Graduate Students, Ideological and Political Courses, Teaching Reform

Abstract: Since 1999, colleges and universities have begun to expand enrollment on a large scale, and correspondingly, postgraduate education has also expanded. With the continuous increase in the number of graduate students, the quality of their education has attracted increasing attention. For graduate students, curriculum is an important carrier of their training and one of the core contents of their education quality assurance. The teaching of ideological and political theory courses for postgraduates is an effective way for postgraduates to impart theoretical knowledge, cultivate moral quality, and stimulate innovation ability. Constructing a scientific and effective new teaching model of ideological and political theory courses for postgraduates in colleges and universities will help to increase the interest and enthusiasm of postgraduates in studying ideological and political theory courses, and help to cultivate innovative talents who are suitable for the needs of the times and develop in an all-round way. The main goal of this paper is to train graduate students with comprehensive development, and aims to study the reform of postgraduate ideological and political education based on machine learning and improved SVM algorithm. This paper proposes machine learning and SVM algorithms, and proposes an improved algorithm based on the previous classic algorithm. The support vector machine model and experiments prove that the improved algorithm is more accurate and faster. The experimental results in this paper show that the improved support vector machine algorithm has the best learning curve, and the improved SVM learning algorithm is found to significantly improve the classification accuracy in the 758 samples taken; when K=16, the correct rate of the picture is as high as it is more than 90%.

1. Introduction

The words emphasized by General Secretary Xi Jinping in his speech at the National Conference on Ideological and Political Work in Colleges and Universities pointed out the direction for the development of my country's education. With the continuous development of my country's economy and society, the scale of postgraduate enrollment has continued to expand and the number of enrollment has continued to increase. In-depth ideological and political education for postgraduates has gradually become an important part of the training of graduate students in colleges and universities. In our country, the highest level of higher talent training is postgraduate education. The cultivation of innovative graduates is an important channel for high-level talents. Building a comprehensive society and realizing the dreams of the Chinese people require high-quality talents. The rapid development of my country's socialist modernization construction, the continuous improvement of the teaching level and teaching ability of colleges and universities, and the further development of the reform and innovation of postgraduate political theory and ideological and political theory have had a wide-ranging impact on the cultivation of talents. However, there are still many problems in the teaching of ideological and political theory courses for postgraduates in colleges and universities. The level of interest in the teaching of ideological and political theory courses for postgraduates is lower than that of graduate students. Therefore, the Party Central Committee is paying more and more attention to the education and teaching of ideological and political theory courses for postgraduates, formulating development directions and a series of guidelines, and the academic community has also continuously explored it as a key issue, innovating theoretical results, and promoting the function of ideological and political theory courses. .

Research on the teaching reform of ideological and political theory courses for postgraduates in colleges and universities helps enrich the theoretical system of ideological and political disciplines, lays a solid foundation for theoretical innovation, promotes the discipline construction of ideological and political courses in universities, and promotes the rationalization of ideological and political courses. Development has broadened the breadth and depth. Since the establishment of postgraduate ideological and political theory in universities across the country, a lot of progress has been made in the education and teaching of postgraduate ideological and political theory courses, but there are still many problems, such as: students' low motivation to learn and serious truancy; Insufficient understanding of science, school, insufficient attention, inadequate management, etc. Therefore, the ideological and political education of graduate students is still a relatively weak link in the ideological and political research of students. It is a very urgent task to reform the teaching reform of ideological and political theory courses for graduate students and strengthen and improve the ideological and political education of graduate students.

At present, machine learning tools have begun to be widely used in healthcare. Mullainathan S, Obermeyer Z raised the question of whether machine learning can automate moral hazard and errors, and found that, unlike other applications of machine learning, a feature of health data Neither y nor x can be measured perfectly [1]. However, it may undermine the ability of machine learning algorithms to drive changes in the healthcare system, and it does lead to reproducing or even amplifying existing errors in human judgment. Taherkhani N, Pierre S proposed a centralized and local data congestion control strategy. In this strategy, the channel usage level is measured to detect data congestion in the channel. The messages will be collected, filtered, and then clustered by machine learning algorithms [2]. Although this research is based on the research of machine learning clustering algorithm, it is not of high value to the research of this article. Gurusamy R, Subramaniam V proposed a new method for denoising, extraction and tumor detection on MRI images. The MRI images obtained from the machine are analyzed during work, and the real-time data is used for analysis. Compared with other methods, wavelet transform is more suitable for MRI image feature extraction. The features are provided to a classifier that uses binary tree support vectors for classification [3]. However, the disadvantage is that the study has not been proved by relevant experiments and lacks feasibility.

The innovation of this paper is (1) The SVM algorithm based on the complex background of machine learning has been improved. Improving the classification accuracy and efficiency is a further development of the improved algorithm performance proposed in this paper. (2) The basic definitions of VC dimension, empirical risk and structural risk are introduced, and the principle of support vector machine SVM is introduced in detail, and the goal is constructed based on the principle of minimizing structural risk and maximizing the interval between classes under linear separability function.

2. Research Method of Postgraduate Ideological and Political Teaching Reform Based on Machine Learning and Improved SVM Algorithm

2.1. Machine Learning

(1) The concept of machine learning

Machine learning is a statistical learning method used to study how computers imitate human cognitive behaviors and accumulate experience, and continue to improve and optimize [4-5]. By learning the laws contained in certain sample sets, it is possible to model low-risk structural classification and highly indirect classification models.

Machine learning is based on integrated classification methods, such as machine learning support based on statistical learning theory, probability graph models, nonlinear data analysis, random forest and boosting methods. It is processed according to the theory of reproductive axis, and the non-parametric Bayesian method is used for normalization. Machine learning is now a key theme in computer science and artificial intelligence, mainly reflected in three main events [7-8]. In May 2010, Professor Michael Jordan was elected as a member of the National Academy of Sciences, and then the probability graph Daphne Koller expert was elected as a computer scientist in engineering theory and learning, and a member of the National Academy of Sciences. Robert Schapire, one of the main founders of Boosting, was elected as a member of the American Academy of Engineering and a member of the Academy of Sciences. During the period, statisticians Jerome Friedman and Robert Tibshirani of Stanford University, Chinese statistician Yu Bin of Berkeley, and statistician Larry Wasserman of Carnegie Mellon University were also successively elected as academicians of the National Academy of Sciences, which fully proves the machine Learning has been regarded as a key issue. Secondly, the 2011 Turing Award was awarded to Professor UCLA Pearl. His main research field is possible graphical modeling and etiology, which are the basic questions of the research. The Turing Award is usually awarded to a theoretical computer designer or researcher who created the original computer architecture or framework. Direction is important. In recent years, almost all scientific papers published in the journal "Nature and Science" come from the field of machine learning. Third, machine learning can be used to solve industrial problems. Especially today's hot spots, such as immigration learning, deep learning, AlphaGo, self-driving cars, artificial intelligence assistants, etc. Has a huge impact on the industry [9]. Google's service industry model. Health is inherent in data, and the key technology for wealth mining is machine learning [10].

(2) Machine learning system

Generally, a machine learning system has the following program modules [11-12]: 1) Operating system. Its main function is to use the learned target function to solve a given task. 2) Identifier. Use the solution path or history file as the input and output of the "target" feature set instruction set. 3)

Generalizer. Using training examples as input data, and assuming the result as an estimate of the objective function, specific training examples and guessing standard functions can cover these examples and situations outside the samples. 4) Experiment generator. Using current assumptions (that is, the currently understood objective function) is to explore new input and output problems of the action system.

2.2. SVM Algorithm

Support vector machine (SVM) is a relatively young content in the field of machine learning [13]. The theory of SVM originated from the treatment of data classification problems. Due to the rigorous theoretical support and good generalization ability of support vector machines, the introduction of the kernel function successfully solved the high-dimensional nonlinear problem, and has potential research and application value, so it has attracted the attention and research of many scholars at home and abroad[14-15] . In recent years, many improvements and distortion algorithms have been developed based on SVM. In order to distinguish distortion algorithms, Vapnik proposed the support machine C-SVM in 1995. The previous SVM distortion algorithms mainly include V-SVM and a type of support vector machine, support vector data description, least square support vector machine, weighted support machine). If the formula is distorted by adding operating conditions, variables or coefficients, these The distortion algorithms are all, which brings multiple advantages or a specific application range of an algorithm, but these distortion algorithms are optimized and distorted according to the long-distance SVM idea proposed earlier, ignoring the classroom data distribution function to the algorithm The impact of performance [16-17].

(1) Minimize experience risk

The essence of machine learning is to find an optimal function in the training sample set $(M,N)=(m_1,n_1),(m_2,n_2),\ldots,(m_x,n_x)$ to minimize the expected risk of n_i between the estimated n_i' of the input m_i and the actual output of the function [18]. The mapping function $F_{M,N}(m,n)$ of M and N determines the expected risk minimization, and the mapping relationship between them generally refers to the prior probability and class conditional probability of the sample set. According to the mapping relationship $F_{M,N}(m,n)$, a set of mapping functions between M and N can be constructed:

$$n' = f(m,g) \quad (1)$$

Among them, g is a set of free parameters m, and the estimated output n can be obtained through the mapping function. In order to express the difference between the estimated output and the actual output, we introduce the difference function $W(n, f(m,g))$. The risk function of the pan-target can be obtained:

$$L(g) = \int W(n, f(m,g))dF_{M,N}(m,n) \quad (2)$$

When the risk function $L(g)$ is the smallest, the accuracy of machine learning is higher, so the essence of machine learning is to minimize $L(g)$ [19]. In traditional statistical theory, the pan-objective risk function $L(g)$ generally uses the empirical risk function $L_{emp}(g)$ to replace:

$$L_{emp}(g) = \frac{1}{x}\sum_{i=1}^{x} W(n_i,\ f(m_i, g)) \quad (3)$$

Solve a set of parameters g to minimize $L_{emp}(g)$ as an estimate of $L(g)$, which is the principle of empirical risk minimization [20].

(2) VC maintenance and extension theory

The level of the VC dimension can clearly indicate the level of the learning performance of the function set, and can effectively solve the promotion problems that appear in the learning process [21].

Assuming the data set $C_X = \{n_1, n_2, \mathrm{K}\ , n_X\}$, there are 2^X ways to mark these X points as positive or negative. These X points lead to 2^X different learning problems. If all 2^X points of C_X are included in the index function set $\{f(m,g)\}$, it is said that the function set $\{f(m,g)\}$ can have an X point. If the index function set $\{f(m,g)\}$ can have any sample A, then the maximum value of the sample t is the dimension VC. The dimensionality of the function set VC is unlimited.

The VC dimension has a great influence on the promotion ability and learning ability of the function set, so that the complexity of the learning machine can be easily measured, and the size of the learning machine capacity can also be measured. The degree of freedom of the function set plays a decisive role in the height of the VC set, and there are many ways to calculate the dimension of the function set [22]. Different learning algorithms have different effects on the VC dimension, and the function set also affects the VC dimension.

The generalized community can effectively measure the relationship between empirical risk and actual risk. If the conditional probability of the data is unknown, the performance of machine learning has a lot to do with common fields [23]. According to research, for the binary classification problem, the relationship between the actual risk of the learning machine $L(g)$ and the empirical risk $L_{emp}(g)$ under the principle of empirical risk minimization meets the following formula with a probability of at least $1-\lambda$:

$$L(g) \le L_{emp}(g) + \sqrt{\left|\frac{t(\ln(2x/t)+1) - \ln(\lambda/4)}{x}\right|} \quad (4)$$

t represents the VC dimension of the indicator function set, and x represents the size of the sample set.

It can be seen from the above formula that if the size of the sample set does not change, then the confidence range of the function set has a linear relationship with the VC dimension. If the learning machine is too complex, it may lead to the phenomenon of "over-learning", and the difference between the experience risk and the actual risk becomes larger and larger. In order to strengthen the promotion of learning machines, a variety of effective methods can be adopted, such as minimizing empirical risks and narrowing the confidence range.

(3) Minimize structural risks

Due to the limited number of empirical risk samples, the machine learning problem cannot be solved. Therefore, this defect can be compensated by minimizing the confidence range [24]. The process of optimizing the confidence range of traditional statistics is the process of selecting the learning algorithm. In order to make the learning machine have a better effect, a relatively matching sample set and learning algorithm should be selected. In the case of a limited number of samples, the form of the classifier can be easily determined, and the VC dimension can also be clarified. The real risk can be minimized in two steps [25]: First, in order to make the learning machine have a fixed confidence range, first determine the form of the classifier; second, reduce the empirical risk to Minimal, which can also minimize the actual risk. However, this algorithm can only be based on the user's experience and "skills" at present, and lacks a theoretical basis.

In order to solve the deficiencies of the above methods, we introduced a new strategy: decompose the given set A, construct a sequence of subsets and arrange them (according to the size of the confidence range).Suppose that this function set satisfies: 1) A is composed of a set of nested sub-function sets $A_k = \{f(m,g) \mid g \in G_k\}$; 2) $t_1 \le t_2 \le \mathrm{K} \le t_y$ means that the VC dimension decreases sequentially. As the number of x sequences in the function subset increases, the minimum empirical risk decreases, but the confidence interval increases. Establish a compromise between the confidence interval and the empirical risk, and select an appropriate subset A_k from the subset to reduce the empirical risk in the selected subset $A_k = \{f(m,g)\}$ to minimize the actual risk. This is the principle of structural risk minimization, and the principle is shown in the figure below:

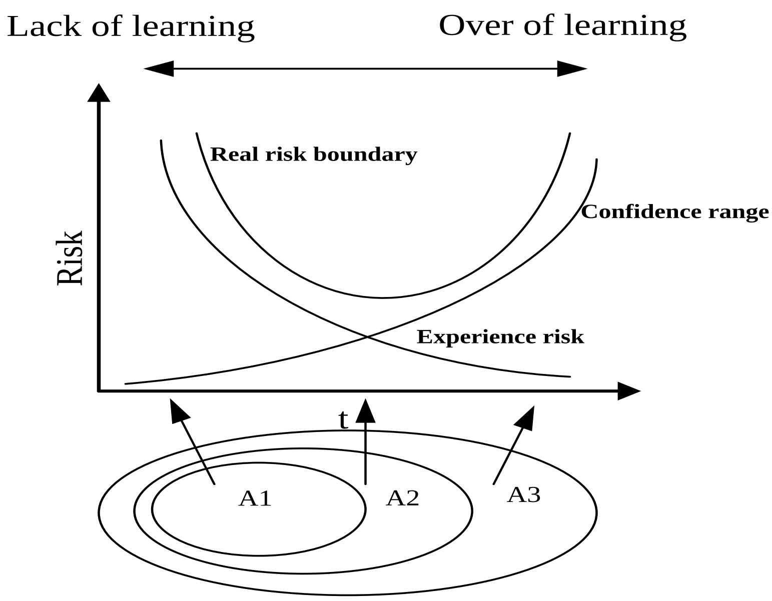

Figure 1. Structural risk minimization

A subset of the function set: $A_1 \subset A_2 \subset A_3$ VC dimension: $t_1 < t_2 < t_3$

There are two ways to realize the principle of structural risk minimization criterion [26]: The first method is to find the minimum empirical risks of the subset and add them to the confidence range. The minimum sum of the two is what we need. result. However, the prerequisite for choosing this method is that the number of subsets should be very limited. When the number is too large, this method is not feasible; the second method is embodied in the application of the support vector machine method. The empirical risks of the set are all taken to the minimum, and then the subset with the smallest confidence range is selected, then the function of this subset is the optimal function.

2.3. Principle of Support Vector Machine

(1) Optimal classification surface

The optimal classification surface is proposed under the condition of linear separability. In order to facilitate the elaboration of this idea, use the following two-dimensional linear separable situation to illustrate. As can be seen from the figure, there are two different training samples. : Solid point and hollow point, T line can separate two different sample points, and T1 and T2 represent the straight line of the point closest to the classification line, and T, T1 and T2 are parallel to each other. The distance between T1 and T2 is called the classification interval [27]. The T line can accurately separate the two kinds of errors, thereby keeping the empirical risk at a minimum, and increasing the classification gap, so that the confidence range is taken to the minimum, and the true risk is also taken to the minimum. Applying the above ideas to the X-dimensional space, the optimal classification line becomes the optimal classification surface, as shown in the following figure.

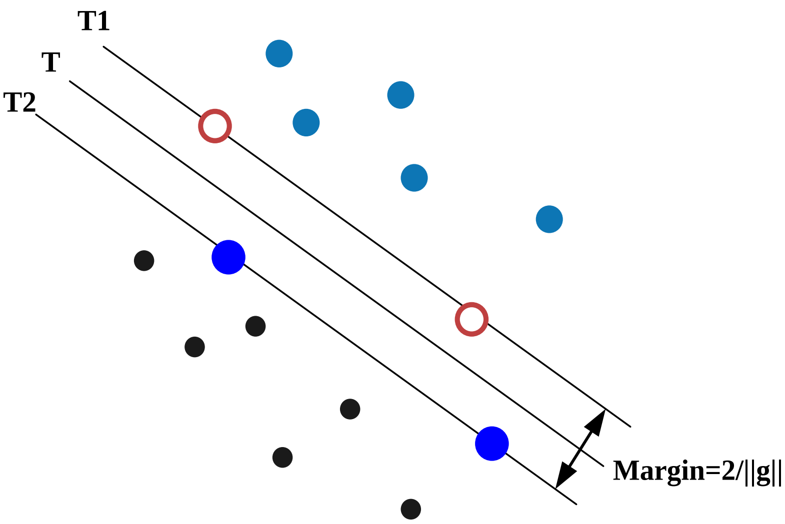

Figure 2. Optimal classification surface

Suppose there are two types in the training sample set $\{(m_i, n_i), i = 1,2,K, n\}$(size N). If $m_i \in L^s$ belongs to the first type, it is represented by positive ($n_i = 1$); if it belongs to the second type, it is represented by negative ($n_i = -1$). The general form of the linear discriminant function is $a(x) = gx + b$, and the classification surface equation is:

$$gx + b = 0 \quad (5)$$

In this case, after normalizing the discriminant function, all samples satisfy $|a(x)| \geq 1$, even in the case of $|a(x)| = 1$. In this case, the classification interval is equal to $2/\|g\|$. Therefore, in order to find the maximum interval, $\|g\|^2$ needs to take the minimum value; therefore, if you want to correctly classify all samples, you must satisfy:

$$n_i[(g \cdot m_i) + b] - 1 \geq 0 \quad (6)$$

In the N-dimensional space, assuming that the samples are distributed in a ball range, with a radius of R, the VC dimension of the indicator function set $f(m,g,b)=\mathrm{sgn}\{(g\cdot m)+b\}$ formed by the regular ball range that satisfies the condition $\|g\|\le C$ satisfies the following boundary

$$t \le \min([R^2C^2], N)+1 \quad (7)$$

Therefore, if $\|g\|^2$ is minimized, the upper bound of the VC dimension must be minimized, and the minimum value of formula (8) must be obtained.

$$\phi(g)=\frac{1}{2}\|g\|^2=\frac{1}{2}(g\cdot g) \quad (8)$$

To this end, define the following Lagrange function:

$$L(g,b,\alpha)=\frac{1}{2}(w\cdot w)-\sum_{i=1}^{n}\alpha_i\{n_i[(g\cdot m_i)+b]-1\} \quad (9)$$

Among them, α_i >0 is the Lagrangian coefficient. The problem now is to find the minimum value of the Lagrangian function of g and b, partially differentiate g and b, and set its value to 0, so that the original problem can be transformed into the next problem:

In constraints

$$\sum_{i=1}^{n}\alpha_i n_i=0,\quad \alpha_i\ge 0,\quad i=1,2,\mathrm{K}\ ,n \quad (10)$$

Solve the maximum value of the following function for α_i:

$$W(\alpha)=\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j n_i n_j(m_i\cdot m_j) \quad (11)$$

If α_j^* is the optimal solution, then

$$g^*=\sum_{i=1}^{n}\alpha_j^* n_i m_i \quad (12)$$

That is, the weight coefficient vector of the optimal classification surface is a linear combination of the training sample vectors.

There is a unique solution to this problem. And the solution of this optimization problem must satisfy

$$\alpha_i(n_i(g\cdot m_i+b)-1)=0, i=1,2,\mathrm{K}\ ,n \quad (13)$$

Therefore, for most samples α_j^*, the value is zero, and non-zero samples α_j^* are support vectors. This is usually not very common in the entire sample. In order to solve the above problems, the optimal classification function is introduced.

$$f(m) = \operatorname{sgn}\{(g^* \cdot m) + b^*\} = \operatorname{sgn}\left\{\sum_{i=1}^{n} \alpha_i^* n_i (m_i \cdot m) + b^*\right\} \quad (14)$$

$\operatorname{sgn}(\ )$ is a symbolic function. The summation process in the above formula is only performed on the support vector. b^* is the domain value of the classification. Since the support vector vector satisfies the equation in the equation, b^* can be obtained by adding any support vector to the formula (14).

(2) Generalized optimal classification surface

If it is linearly inseparable, then the training sample cannot meet the condition of formula (14), we only need to add a relaxation term $\xi_i \geq 0$ to the condition, and it becomes

$$n_i[(g \cdot m_i) + b] - 1 + \xi_i \geq 0 \quad (15)$$

For a sufficiently small σ >0, just make

$$F_\sigma(\xi) = \sum_{i=1}^{n} \xi^\sigma \quad (16)$$

The smallest can minimize the number of misclassified samples. Introduce constraints under the premise of linear inseparability:

$$\|g\|^2 \leq y_k \quad (17)$$

Under the constraint conditions of formula (15) and formula (17), the minimum value of formula (16) can be obtained to obtain the optimal classification surface under the condition of linear inseparability, which is also called the generalized optimal classification surface. Take $\sigma = 1$ for the convenience of calculation.

In order to simplify the calculation, the problem of developing a generalized optimal classification surface can be transformed into finding the minimum value of the following function under conditional constraints:

$$\phi(g, \xi) = \frac{1}{2}(g \cdot g) + Y\left(\sum_{i=1}^{n} \xi_i\right) \quad (18)$$

Where Y is a constant, it can play a role in controlling the degree of punishment for right and wrong and sub-sample.

3. Establishment of Support Vector Machine Model

3.1. Selection of Training Set

The training set is the data basis for establishing the SVM model, so the selection of the training set plays a vital role in the SVM model. When choosing the training set, the following aspects should be considered:

(1) In the entire training set, there must be a sufficient number of samples, because a small

number of samples will lead to "under-learning" models. Therefore, the training set must have different situations for the problem to be solved;

(2) To have enough samples in the training set does not mean that there can be an infinite number. Too many samples will cause the SVM model to "overlearn", which will greatly increase the training time.

3.2. Selection of Training Features

When building the SVM model, the required prediction results are defined as the target value, and there are characteristic values corresponding to them. In general, these characteristic values are multi-dimensional. The quality of the establishment of the SVM model is also affected by the eigenvalues, so to choose a suitable training set, the following are several necessary conditions:

(1) In order to make the established SVM model have practical significance, the correlation between the target value and the characteristic value should be ensured;

(2) In order to ensure the high accuracy of the model, the number of features must not be too small, and all the features of the problem to be solved must be included.

3.3. Choice of Kernel Function

Commonly used kernel functions generally include:

(1) Linear kernel function

$$K(m_i, m_j) = m_i^T m_j \quad (19)$$

(2) Polynomial kernel function

$$K(m_i, m_j) = (m_i^T m_j + 1)^d \quad (20)$$

In the formula, d is the order of the polynomial kernel function.

(3) Radial basis kernel function

$$K(m_i, m_j) = \exp(-\|m_i - m_j\|^d / 2\sigma^2) \quad (21)$$

(4) Neural network kernel function

$$K(m_i, m_j) = \frac{1}{1 + \exp(a m_i^T m_j - b)} \quad (22)$$

In the formula, a and b are constants.

(5) Hyperbolic tangent kernel function

$$K(m_i, m_j) = \tanh[b(m_i \cdot m_j) + c] \quad (23)$$

The choice of kernel function is not an easy task, and there is currently no good way. Due to the particularity of the problem, users choose kernel functions based on personal experience and prior knowledge. Although the selected kernel function may be very different, very similar results can be obtained. These studies show that different kernel functions do not have much influence on the SVM model.

3.4. Selection of Model Parameters

Although the results of different kernel functions are not much different, the final result is affected by the parameter model. When building a model, if the training samples are the same, choosing different parameters may produce very different results, and the accuracy of the model is also affected by the selection of parameters. In view of the importance of parameter selection, people have proposed network search methods, genetic algorithms, enumeration methods, and particle swarm algorithms, etc., all of which help to establish a better model. The enumeration method is the most commonly used method, which saves time and effort, and is suitable for sample sets that are not sensitive to parameters; if the sample set is more sensitive to parameters, it is necessary to consider the use of optimization algorithms; although the network search method consumes time It is the longest, but the result obtained is the best. It is a kind of exhaustive method. Because it takes a long time, it is not suitable for all situations. The most commonly used optimization algorithms are particle swarm optimization and Genetic algorithms, these two methods can get the best results in the shortest time, and are currently widely used.

4. Experimental Results and Analysis

4.1. Test the Data Sample

For SVM learning, learning samples is crucial, regardless of the number or quality of the samples. As mentioned earlier, this article used 2000 pictures with text and 2000 pictures without text for testing. In order to test the data sample, this article uses 250 pictures with text and 250 pictures without text, 500 pictures with text and 500 pictures without text, 1000 pictures with text and 1000 pictures without text. The pictures were taken as data samples, and SVM machine learning was performed again. In the testing phase, this article regenerated 100 pictures with text and selected 100 pictures without text as test samples. The test data is shown in Table 1:

Table 1. The influence of the number of data samples on the final result

Number of samples	Test content	Picture with text	Picture without text
250	The test result is in text	76	22
	The test result is no text	24	78
500	The test result is in text	78	20
	The test result is no text	22	80
1000	The test result is in text	82	19
	The test result is no text	18	81

It can be seen that when the number of samples is 250 pictures with text and 250 pictures without text, the correct rate is 76% for pictures with words and 78% for pictures without words. When the number of samples is 500 pictures with text and 500 pictures without text, the correct rate is 78% for pictures with words and 80% for pictures without words. When the number of samples is 1000 pictures with text and 1000 pictures without text, the correct rate is 82% for pictures with words and 81% for pictures without words. It can be clearly seen from Figure 3 that with the increase of samples, the accuracy rate is getting higher and higher.

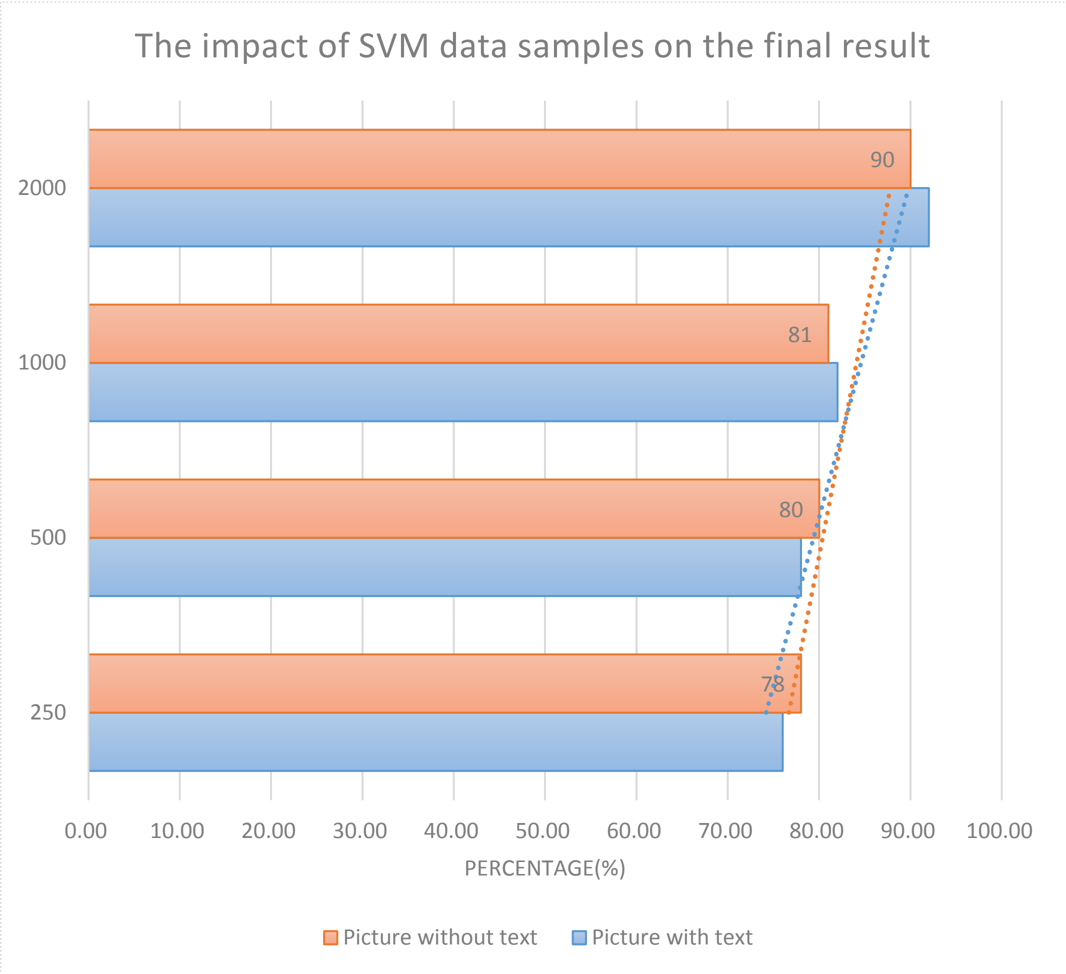

Figure 3. The impact of SVM learning data samples on the final result

4.2. Improved SVM Algorithm

The data used in the experiment are all data sets used in multiple experiments, and they have certain application evidence. In order to improve the accuracy of the data and the accuracy of classification, we will use all the data used in the experiment to uniformly use the calculation work. The kernel parameters used by the support vector machine directly affect the effective use of classification. The kernel parameters within a reasonable range will greatly improve the accuracy of prediction. However, if the parameters are unreasonable, the accuracy will be greatly reduced. The way of obtaining the nuclear parameters is also different for different calculation methods. So we need to apply the cross-checking method to determine the parameters. In experiment 1, we selected 1,700 as the initial sample set. The other 300 were used as a comparative sample set, and they were equally divided into five groups for each change. We have calculated the results of repeated experiments and drawn the following statistical graphs:

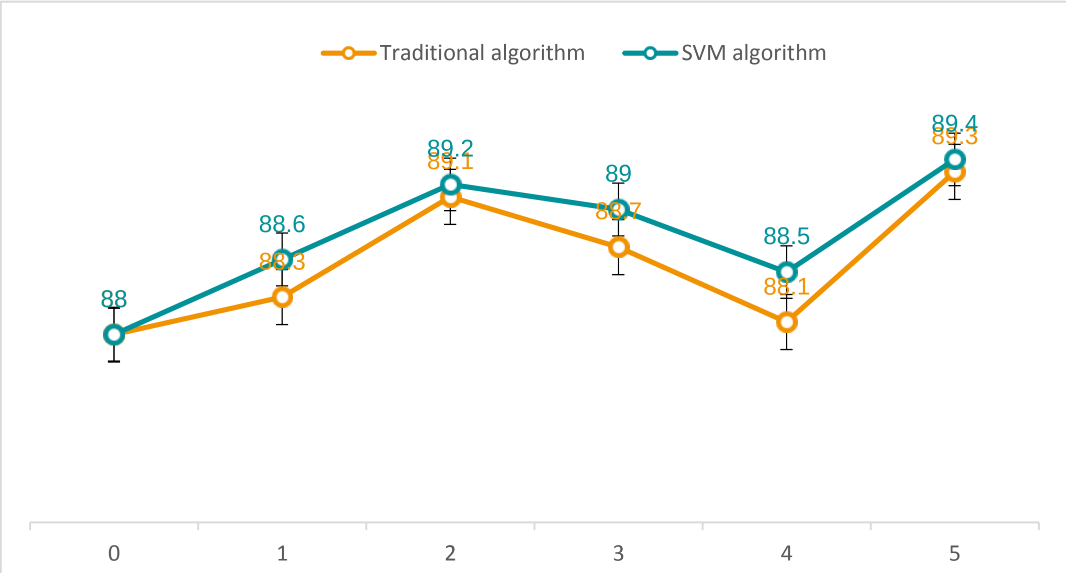

Figure 4. Comparison of training time of the two algorithms

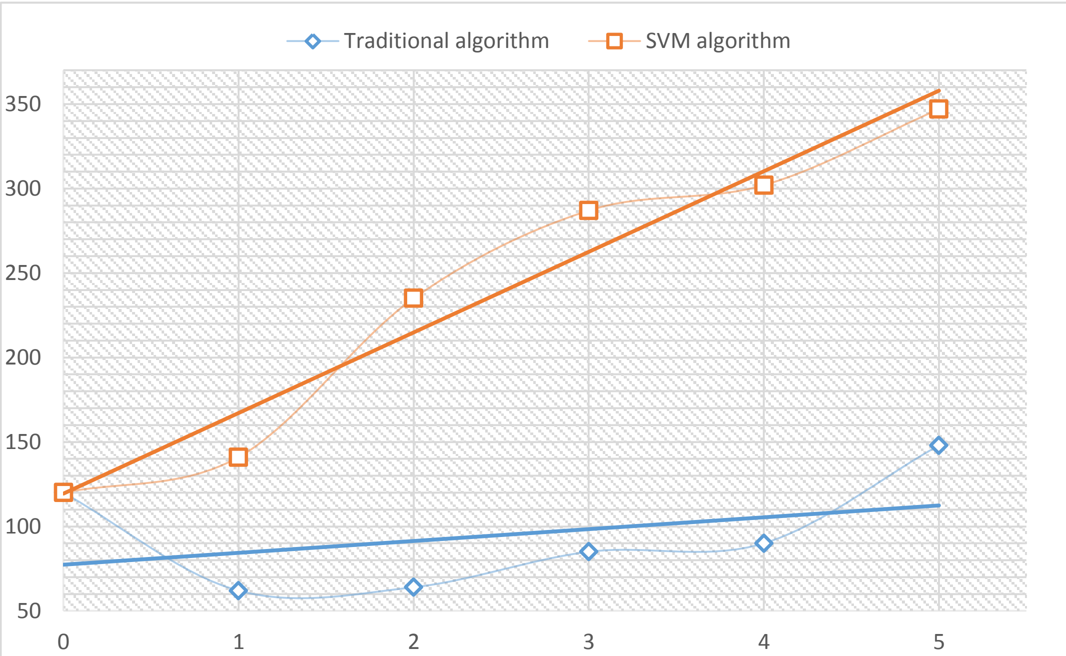

Figure 5. Comparison of training accuracy of the two algorithms

Using Image segmentation in the machine learning database, DNA in the Isolate and Statlog databases, and Satimage, a total of 4 data sets, and using the algorithm in this paper, the incremental SVM elimination algorithm, and the general incremental SVM to perform three incremental learning algorithms to obtain The results are compared with the experiment. The data obtained through the experiment is described as follows:

Table 2. Basic data description

Data Set	Number of training samples	Number of test samples	Feature dimension	Number of categories
Image segmentation	230	2315	19	8
Isolate	2110	1208	184	4
DNA	4565	2018	39	7
Satimage	6325	1611	633	25

The method of selecting the initial training sample is to select 10% of the training sample, and each increase in the number of samples is 10% of the training sample. Next, the samples are divided into 10 groups in total, and any group is selected for incremental experiments. For all data sets, 10-fold cross-validation is required to solve the accuracy. Finally, the set with the highest accuracy is selected. As the optimal parameter. The following table shows the accuracy test of the algorithm and incremental SVM elimination algorithm in this paper, and general incremental SVM on the test set.

Table 3. Test accuracy of three incremental learning SVM algorithms

Data Set	Image segmentation	Isolate	DNA	Satimage
Number of training samples	230	2110	4565	6325
Incremental SVM elimination algorithm	94.76	96.09	92.71	95.89
Traditional incremental SVM	91.24	95.66	92.03	94.37
Improved SVM algorithm	95.88	97.27	93.84	97.61

Below we compare the training time of the improved algorithm, the incremental SVM elimination algorithm and the general incremental SVM. The experimental results are shown in the following figure:

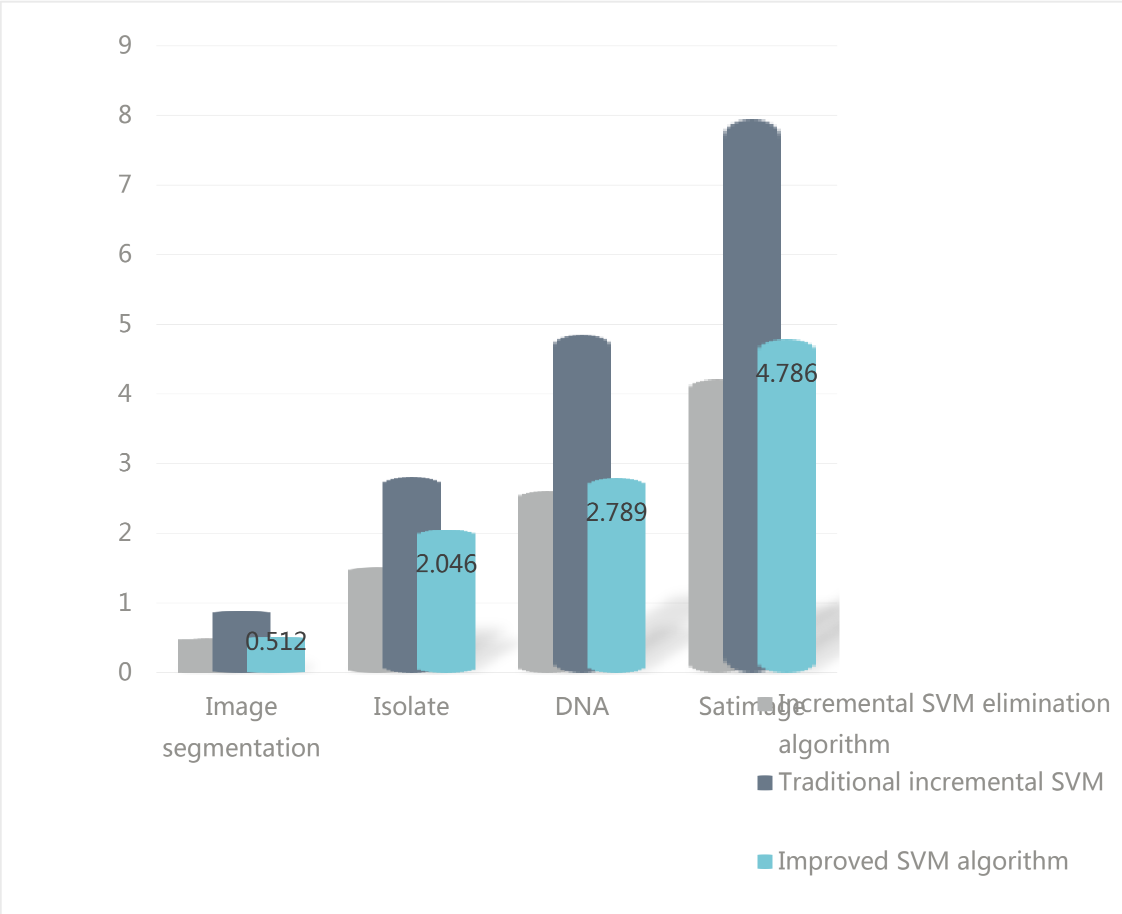

Figure 6. Comparison of training time of three SVM algorithms

In order to know more clearly the comparison data of these several algorithms and the difference between the three algorithms in different data sets, see Table 4.

Table 4. Comparison of training time (s) of three incremental learning SVM algorithms

Data Set	Incremental SVM elimination algorithm	Traditional incremental SVM	Improved SVM algorithm
Image segmentation	0.492	0.885	0.512
Isolate	1.512	2.801	2.046
DNA	2.597	4.852	2.789
Satimage	4.209	7.946	4.786

The experimental data results in Table 3 and Table 4 show that the adaptive boundary support vector incremental learning SVM algorithm proposed in this paper is significantly better than traditional incremental SVM and SVM elimination in classification accuracy without significantly changing the operation time. algorithm. The experimental results show that the algorithm proposed in this paper is feasible.

3000 data texts that have been manually classified in the local database were extracted. Use the above-mentioned improved SVM incremental learning algorithm to carry out incremental learning experiments. Randomly select 758 samples as the initial training set. In order to verify the results of the incremental data sets of different sizes, we randomly divide the remaining samples into five new data sets of varying sizes, and the classification number j is set to 5 respectively. experiment. The experimental results are as follows:

Table 5. Comparison of improved SVM algorithm learning and traditional SVM algorithm learning

Number of categories	Training set	Incremental sample number	SVM algorithm		improve algorithm	
			Time/s	Accuracy	Time/s	Accuracy
j=5	Initial training	758	161.50	95.2%	165.13	92.3%
	Incremental set 1	578	23.75	97.6%	21.57	95.7%
	Incremental set 2	124	11.55	93.8%	9.81	91.8%
	Incremental set 3	236	15.41	95.1%	14.72	93.1%
	Incremental set 4	89	8.10	97.6%	8.55	96.5%
	Incremental set 5	1215	35.51	94.7%	38.33	93.4%

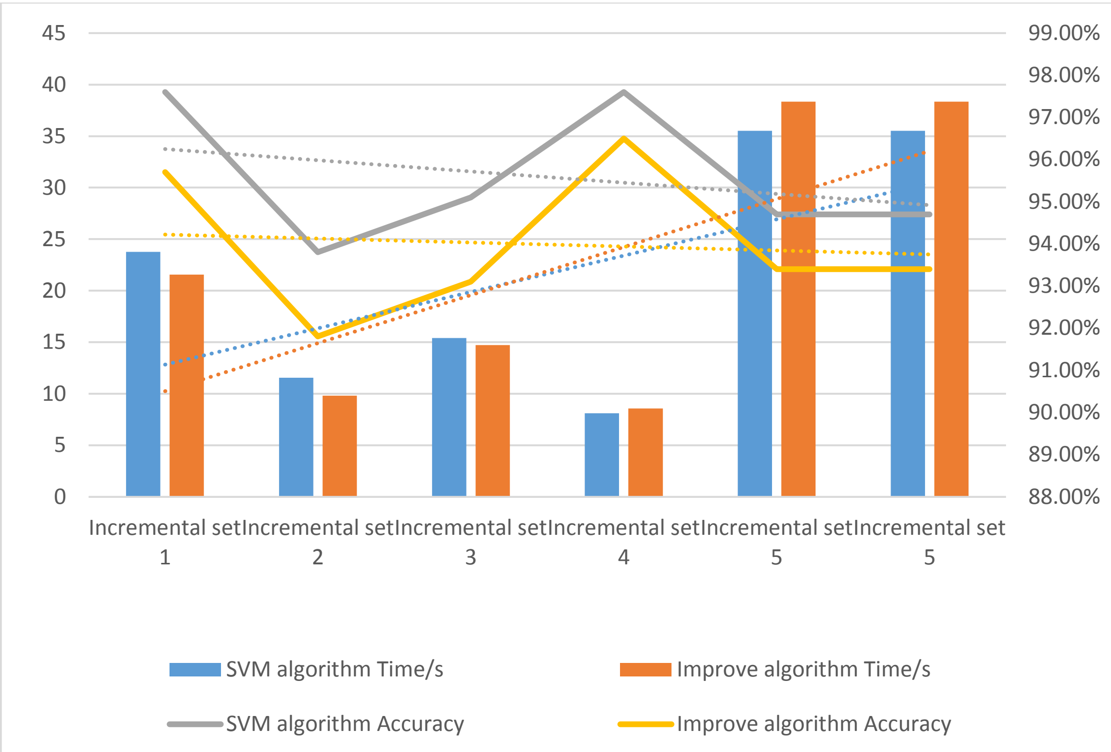

Figure 7. Comparison of improved SVM algorithm learning and traditional SVM algorithm learning in terms of time and accuracy

The experimental results show that the improved SVM learning algorithm has significantly improved the classification accuracy without affecting the calculation speed. Choose different values of k to obtain different learning results. In the experiment, 1000 text images and 1000 no text images were selected as the test images of the k test in the experiment. Table 6 records the test data when k is 8, 16, 32 and 64:

Table 6. The influence of K value on the final result

The value of k	Test content	Picture with text	Picture without text
K=8	The test result is in text	874	144
	The test result is no text	126	856
K=16	The test result is in text	913	98
	The test result is no text	87	902
K=32	The test result is in text	892	117
	The test result is no text	108	883
K=64	The test result is in text	871	138
	The test result is no text	129	862

From the data in Table 6, the correct rate of pictures with words and the correct rates of pictures without words can be obtained under different K values as shown in the figure below:

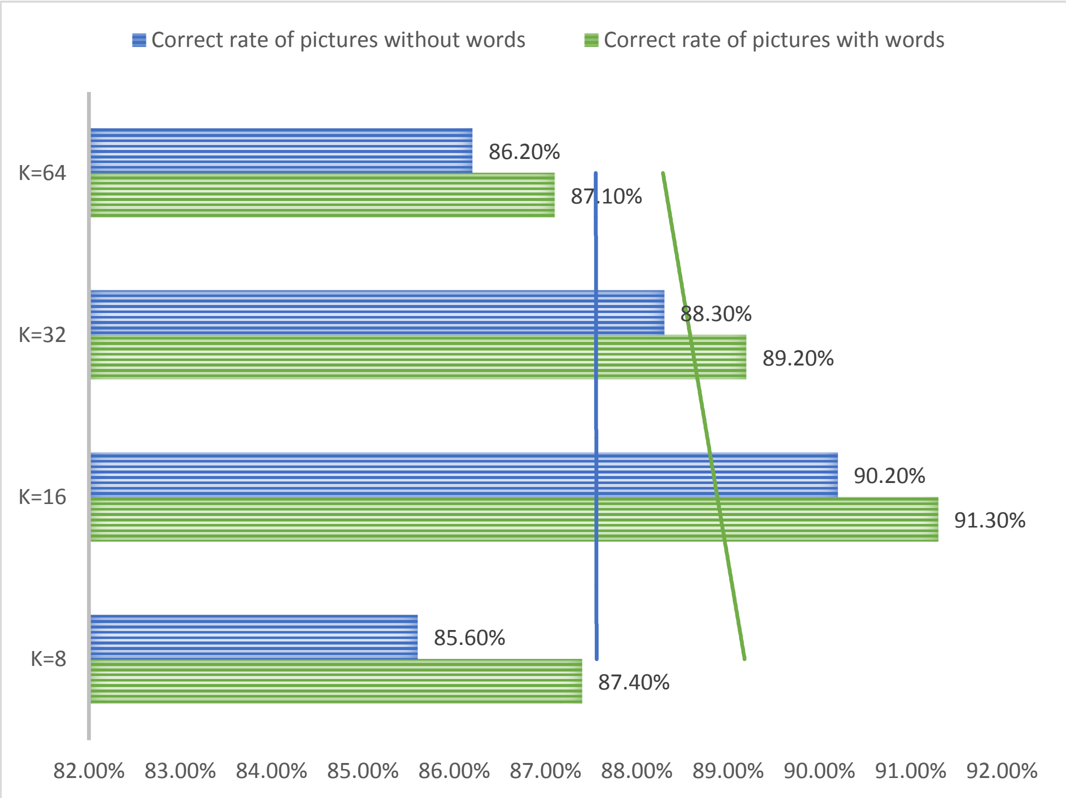

Figure 8. Comparison of the correct rate of pictures under different K values

It can be seen that when K=16, the test result is the best. For images containing text, the correct percentage is 91.3%, and for images without text, the correct percentage is 90.2%.

5. Conclusion

Support vector machine method is a kind of machine learning method based on statistics, with the least risk and the highest accuracy. According to the information provided by the training sample set, a balance point is found between the complex model and the learning level, so as to obtain better promotion ability, and there is no disaster of dimensionality. Whenever a new type of sample appears, it is necessary to put the newly obtained sample into the original training sample, and retrain the mixed sample. The learning based on SVM algorithm can not only keep the previous learning results, but also learn once for the newly added data, thereby enhancing the ability of continuous learning. Through the research of machine learning and support vector, this paper proposes an improved SVM learning algorithm, which improves the accuracy and running time of the algorithm. The support vector machine was originally proposed for the two-classification problem. Many scholars have conducted research on the multi-classification problem. Based on this, this paper improves a multi-classification algorithm based on the sphere structure, which effectively removes the overlapping samples in the sphere structure. Perform classification to improve classification accuracy.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1]Mullainathan S , Obermeyer Z . Does Machine Learning Automate Moral Hazard and Error?. Am Econ Rev, 2017, 107(5):476-480. https://doi.org/10.1257/aer.p20171084

[2]Taherkhani N , Pierre S . Centralized and Localized Data Congestion Control Strategy for Vehicular Ad Hoc Networks Using a Machine Learning Clustering Algorithm. IEEE Transactions on Intelligent Transportation Systems, 2016, 17(11):1-11. https://doi.org/10.1109/TITS.2016.2546555

[3]Gurusamy R , Subramaniam V . A machine learning approach for MRI brain tumor classification. Computers, Materials and Continua, 2017, 53(2):91-109.

[4]Yu M D , Hiller M , Delvaux J , et al. A Lockdown Technique to Prevent Machine Learning on PUFs for Lightweight Authentication. IEEE Transactions on Multi-Scale Computing Systems, 2017, 2(3):146-159.

[5]Xiao S , Kong Q , Song M , et al. Teaching Reform of CNC Technology Course Under the Ideological and Political Education. International Journal of Social Science and Education Research, 2019, 2(10):36-41.

[6]Liu G , Liu X , Huang Z , et al. A Study on the Basic Problems of Ideological and Political

Teaching Reform in the Major Courses of Universities. Creative Education, 2019, 10(13):3420-3432. https://doi.org/10.4236/ce.2019.1013264

[7]Jean N , Burke M , Xie M , et al. Combining satellite imagery and machine learning to predict poverty. Science, 2016, 353(6301):790-794. https://doi.org/10.1126/science.aaf7894

[8]Helma C , Cramer T , Kramer S , et al. Data mining and machine learning techniques for the identification of mutagenicity inducing substructures and structure activity relationships of noncongeneric compounds. J Chem Inf Comput, 2018, 35(4):1402-1411. https://doi.org/10.1021/ci034254q

[9]Zhang Y , Kim E A . Quantum Loop Topography for Machine Learning. Physical review letters, 2017, 118(21):216401.1-216401.5.

[10]Goswami R , Dufort P , Tartaglia M C , et al. Frontotemporal correlates of impulsivity and machine learning in retired professional athletes with a history of multiple concussions. Brain Structure and Function, 2016, 221(4):1911-1925. https://doi.org/10.1007/s00429-015-1012-0

[11]Dev S , Wen B , Lee Y H , et al. Ground-Based Image Analysis: A Tutorial on Machine-Learning Techniques and Applications. IEEE Geoscience & Remote Sensing Magazine, 2016, 4(2):79-93. https://doi.org/10.1109/MGRS.2015.2510448

[12]Cai J , Luo J , Wang S , et al. Feature selection in machine learning: A new perspective. NEUROCOMPUTING, 2018, 300(jul.26):70-79.

[13]Karpatne A , Jiang Z , Vatsavai R R , et al. Monitoring Land-Cover Changes: A Machine-Learning Perspective. IEEE Geoscience and Remote Sensing Magazine, 2016, 4(2):8-21.

[14]Saeb S , Lonini L , Jayaraman A , et al. The need to approximate the use-case in clinical machine learning. Gigascience, 2017, 6(5):1-9. https://doi.org/10.1093/gigascience/gix019

[15]Feng Z , Rong P , Cao P , et al. Machine learning-based quantitative texture analysis of CT images of small renal masses: Differentiation of angiomyolipoma without visible fat from renal cell carcinoma. European Radiology, 2018, 28(4):1625-1633. https://doi.org/10.1007/s00330-017-5118-z

[16]Wu W , Nagarajan S , Chen Z . Bayesian Machine Learning: EEG/MEG signal processing measurements. IEEE Signal Processing Magazine, 2016, 33(1):172-172. https://doi.org/10.1109/MSP.2015.2481559

[17]Samuel Rota Bulò, Biggio B , Pillai I , et al. Randomized Prediction Games for Adversarial Machine Learning. IEEE Transactions on Neural Networks and Learning Systems, 2017, 28(11):1-13. https://doi.org/10.1109/TNNLS.2016.2593488

[18]Sui H , Li L , Zhu X , et al. Modeling the adsorption of PAH mixture in silica nanopores by molecular dynamic simulation combined with machine learning. American Journal of Hematology, 2016, 85(FEB.):1950-1959.

[19]Chen Y , Chen T , Xu Z , et al. DianNao family: energy-efficient hardware accelerators for machine learning. Communications of the Acm, 2016, 59(11):105-112. https://doi.org/10.1145/2996864

[20]Luo, Gang. Automatically explaining machine learning prediction results: a demonstration on type 2 diabetes risk prediction. Health Information Science & Systems, 2016, 4(1):1-9.

[21]Jamali A A , Ferdousi R , Razzaghi S , et al. DrugMiner: comparative analysis of machine learning algorithms for prediction of potential druggable proteins. Drug Discovery Today, 2016, 21(5):718-724.

[22]Taylor R A , Pare J R , Venkatesh A K , et al. Prediction of In‐hospital Mortality in Emergency Department Patients With Sepsis: A Local Big Data–Driven, Machine Learning

Approach. Academic Emergency Medicine, 2016, 23(3):269-278. https://doi.org/10.1111/acem.12876

[23]Mohammadi M , Al-Fuqaha A . Enabling Cognitive Smart Cities Using Big Data and Machine Learning: Approaches and Challenges. IEEE Communications Magazine, 2018, 56(2):94-101.

[24]Valdes G , Solberg T D , Heskel M , et al. Using machine learning to predict radiation pneumonitis in patients with stage I non-small cell lung cancer treated with stereotactic body radiation therapy. Physics in Medicine & Biology, 2016, 61(16):6105-6120.

[25]Shortridge J E , Guikema S D , Zaitchik B F . Machine learning methods for empirical streamflow simulation: a comparison of model accuracy, interpretability, and uncertainty in seasonal watersheds. Hydrology and Earth System Sciences, 2016, 20(7):2611-2628. https://doi.org/10.5194/hess-20-2611-2016

[26]Plawiak P , Sosnicki T , Niedzwiecki M , et al. Hand Body Language Gesture Recognition Based on Signals From Specialized Glove and Machine Learning Algorithms. IEEE Transactions on Industrial Informatics, 2016, 12(3):1104-1113. https://doi.org/10.1109/TII.2016.2550528

[27]Yuan W , Chin K S , Hua M , et al. Shape classification of wear particles by image boundary analysis using machine learning algorithms. Mechanical Systems and Signal Processing, 2016, 72-73(May):346-358. https://doi.org/10.1016/j.ymssp.2015.10.013

[28]Dickson M E , Perry G L W . Identifying the controls on coastal cliff landslides using machine-learning approaches. Environmental Modelling & Software, 2016, 76(FEB.):117-127.

[29]Liu T , Yang Y , Huang G B , et al. Driver Distraction Detection Using Semi-Supervised Machine Learning. IEEE Transactions on Intelligent Transportation Systems, 2016, 17(4):1108-1120. https://doi.org/10.1109/TITS.2015.2496157

[30]Bryan, Nick R . Machine Learning Applied to Alzheimer Disease. Radiology, 2016, 281(3):665-668. https://doi.org/10.1148/radiol.2016162151