

# Research on Incremental Learning Method Based on Support Vector Machine Method

Xi Long Ding\*

Weifang University of Science and Technology, Shandong Weifang, China

**Abstract:** An incremental learning algorithm based on support vector machine was proposed to process large-scale data or data generated in batches. Initial goal concept learnt by standard support vector machine algorithm was updated by an updating model. Compared with the existing incremental learning algorithms, this algorithm can achieve the incremental inverse process and the training time is in inverse proportion. Results show that the algorithm satisfies the three criteria of stability, improvement and recoverability. The main target of this thesis is to find a new SVM learning algorithm based on weighted incremental method above the current SVM algorithms, and this new algorithm will have a better classification performance in large-scale data situation.

**Keywords:** Incremental, statistical learning theory, SVM, VC dimension.

## 1. INTRODUCTION

Support Vector Machine (SVM) is a new machine-learning technique, which was put forward by Vapnik in the last century. It's the kernel content of statistical learning theory, has some advantages including simple structure and good generalization, and has been widely applied to many areas, such as pattern recognition, signal processing, automation, and data mining. However, the traditional support vector machine algorithm doesn't support incremental learning, which leads to its bad performance in large-scale data situation. So, it has great significance in improving the classification performance of the incremental support vector machine algorithm both in theory and application.

## 2. THEORETICAL BASE OF SVM

SVM, put forward by Vapnik *et al.*, is used to implement the classification and handling of small sample data. This method is preferred due to avoiding the problems which are easily produced in the early process of early machine learning, such as learning, lack of learning and local optimal. It has always been regarded as the best to deal with the problems of small sample techniques. Its theoretical foundation is firm; and the mathematical deduction process is strict. There is also a breakthrough in solving nonlinear problems.

At present SVM technique raised extensive concerns of scholars; therefore, a number of solutions and the improved algorithms have been put forward for SVM training

and implementation. But, there are still many difficulties and problems that need to be further solved to meet the needs of the practical problems.

- 1) Initialization,  $T$  represents the initial temperature, and  $S$  represents the state of initial solution, and iterations  $L$  displayed by different  $T$  values are different;
- 2) Problem about kernel function theory. Since the birth of statistical learning theory and SVM, the structure of the new kernel function, parameters adjustment and choice in the kernel function and the choice of kernel function model have always been difficult problems. At present the research has made some progress, but it is still not satisfactory, especially the structure of the new kernel function, which is still a difficult problem to deal with. There is some progress in some of existing researches including: improvement of kernel function; selection of SVM model, adjustment of super parameter *etc.* However, a greater breakthrough in this aspect is still expected, such as structuring kernel function facing data, in order to solve more practical problem.
- 3) Problem of large-scale data training (massive data mining). Mass data processing is a significant problem needs to be solved urgently. Especially, in the era of rapid development of science and technology when all kinds of huge amounts of endless data emerged, such as biological information in various large datasets. A strong data analysis tool is an urgent need for all kinds of complicated large data, economic data *etc.* Using the SVM technology to carry out large-scale data analysis is the goal that scholars pursue in this field. How to solve the contradiction between the training speed and the training sample size, the contradiction between test speed and the number of support vector,

\*Address correspondence to Weifang University of Science and Technology, Shandong Weifang, China; P.O. Box 262700, China; Tel: 15662537128; E-mail: [dxl\\_8036@163.com](mailto:dxl_8036@163.com)

finding SVM training algorithm and classification implemental algorithm effective for large-scale sample datasets is still the problem that has not been resolved well.

- 4) Research on multiple class classification problems. How to extend the SVM method, so as to make it effectively solve many kinds of classification problems, which has been the research focus in the field. Not only there exist serious problems, but also lack of a unified theoretical system in some existing methods. As a result, the effective algorithm is a problem that needs to be studied further when SVM is used for various classification problems. Thus, research warrants optimization design of the multi class SVM.
- 5) Data classification problem with incremental learning ability. Having incremental learning ability algorithm is the key of many online training, in real-time application. An effective incremental learning algorithm needs to be found, at the same time, to meet the expected risk control.
- 6) How to reduce the computational cost and computing time. Almost all of the SVM systems have this problem, because the SVM is a quadratic optimization program. Therefore, when solving SVM, this problem is almost inevitable; therefore, future research direction of the training algorithm of SVM will aim at solving the above problems, and exploring continuously.

In the formula,  $K(x \cdot x_i) = (x \cdot x_i)^n$  is kernel function, and sign is symbolic function, the results are +1 or -1, and n is the number of support vector. This paper selects radial basis kernel function. Regularization parameter and kernel function parameters' C sigma selection directly affect the size of the classification accuracy of SVM, so the reasonable selection of parameters is very important according to the SVM theory. But the existing methods of selecting SVM's parameters are tedious, time-consuming and the result is not good. Therefore, this article puts forward to use the simulated annealing (SA) algorithm for selecting SVM parameters.

If the problem dataset is a convex set, global optimal solution can be achieved? Moreover, if some Lagrange multipliers become zero, not all of the training points are meaningful for the final result. If the training points are removed in advance, the same conclusion can be drawn. Training set with a nonzero Lagrange multiplier is called support vector which plays a crucial role in the conclusion. The fewer the number of support vector, the more reasonable the conclusion, less calculation amount and the more ideal result will be achieved.

SVM is known as the first choice to solve the problem of nonlinear because of the uniqueness, the sparseness and good generalization of its solution. The traditional artificial intelligence technology is based on empirical risk minimization principle, which is easy to cause over-fitting and under-fitting. And the SVM is based on structural risk minimization principle, which not only controls the training error, but also controls the complexity and generalization risk of model. So compared with the traditional method, the SVM can get better performance. This article through improved simulated annealing algorithm to optimize the parameters of

SVM, uses well-optimized SVM to realize the prediction of grid resources.

### 3. MACHINE LEARNING THEORY

#### 3.1. The Expression of Machine Learning Problems

Machine learning method uses artificial intelligence to make the computer learn automatically, thereby constantly improving their performance through learning. Traditional machine learning model is shown in Fig. (1). It is based on the principle is to get information from the environment, produce the knowledge base for information learning, and implementation and evaluation are based on the knowledge base, which can also provide feedback for learning modules of the study improvement method.

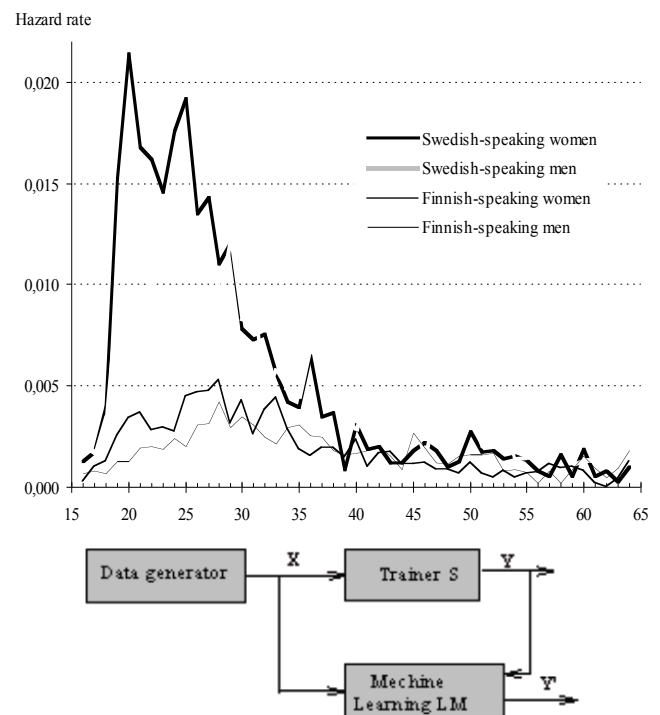


Fig. (1). Machine learning process.

There are usually three modules based on the data of machine learning; data generator, trainers and learning machine. The specific learning process is as follows:

- 1) Data generator G, generated from distribution function  $F(x)$  with the distribution of the vector,  $n$ ix  $x \in R$ .
- 2) Trainer S, returns the corresponding output value  $y$  to all the input vectors  $x$ , according to  $F(x, y) = F(x) F(y | x)$ , choose one independent distribution data from  $(x_1, y_1), (x_2, y_2), \dots, (x_{i_a}, y_{i_a})$  as the data training dataset.
- 3) Learning machine, select from the standby dataset  $f(x, a), a \in \Omega, \Omega$  expresses parameters collection, and obtain function  $f(x, a)$  which is the most close to the objective. Then, calculate its corresponding results  $y$

as to the input vector  $x$  and output it. The role of the machine learning module is to get the latest output value  $y'$  infinitely closer to trainer output  $y$ .

Through machine learning, extract function from the training dataset  $f(x, a)$ ,  $a \in \Omega$  that reflect characteristics of training dataset and approximation to trainers. Minimized risk of functional  $R(a)$  is expressed in mathematics as follows:

$$R(a) = \int Q(z, a) dF(z), a \in \Omega \quad (1)$$

In this equation,  $Q(z, a)$ ,  $a \in \Omega$  is the specific loss function, and  $F(z)$  is the probability measure on the  $Z$  space.

### 3.2. Minimization of Empirical Risk

Machine learning problems based on the dataset, is aiming at minimizing the expected risk  $R(a)$ , through the Empirical Risk Minimization [1] principle, by changing formula (1) to the following formula:

$$R_{emp}(a) = \frac{1}{l} \sum_{i=1}^l Q(z_i, a) \quad (2)$$

This is just a simple replacement without solid mathematical proof of theoretical support. That is, only when the sample size is infinity, empirical risk can be approximated to real risk. When the sample size is very small, this alternative often may cause blunders. Therefore, since the early days of machine learning, scholars have been studying how to carry out the empirical risk minimization which dominated the machine learning for many years.

### 3.3. VC Dimension

VC dimension [2] is an important concept in the statistical learning theory. In the pattern recognition method, VC dimension is defined as: for a dataset of exponential function, if there are  $h$  groups of samples separated in accordance with all 20 possible parameters using the functions from function dataset, then the function dataset is called to scatter  $h$  samples. The VC dimension of the function dataset is the largest number  $h$  of samples that it can break. If according to any number of samples there are functions that can break them up, the VC dimension of the function datasets is infinite. In short, since the VC dimension itself describes the function of learning model collection capacity, it can be said that the learning ability of this function is descriptive. The larger VC dimension is, the greater the function dataset is, and the stronger the corresponding learning ability. For example for the second-class classification problem, based on each possible division, there is a swallowing function existing in each function dataset, to make the function to take +1 to one of the classes, and to take -1 to another class.

## 4. BASIC PRINCIPLE OF SUPPORT VECTOR MACHINE

### 4.1. Incremental Support Vector Machine

All these methods assume that the sample dataset is small. In the actual situation however, it is often unable to meet this condition, as data capacity dynamically grows over

time. If the sample data is regularly received over time or got online, along with the collection of new sample, SVM algorithm is limited by the bottleneck of quadratic programming problem:

- 1) The time complexity for solving quadratic programming problem is  $O(N^3)$ . The space complexity storage of nuclear matrix space is  $O(N^2)$ . When the size of training sample dataset is too large, it will lead to too long time of SVM training, the nuclear matrix storage size can cause lack of memory space;
- 2) The essence of solving problem of quadratic programming is geared to the needs of mass data. Although the trained SVM cannot include the new training sample, yet it reflects the characteristics of the new samples. Classical SVM algorithm can't support incremental learning directly.

How to reduce the operation time in computational memory problems brought by the quadratic programming to solve problem of large capacity dataset has been the focus of SVM. In the face of the increment of dynamic data, people designed some new algorithms of SVM, named incremental SVM algorithm. This algorithm has shown its power on dealing with problems by aiming at increasing data to accumulate knowledge. It has the following unique advantages:

- 1) Large-scale data processing. Incremental SVM works by breaking the large data into several parts. It studies one part to gain support vector information, and then add another from the remaining sample set, for a single instance of training. Or at the same time, it studies the integral dataset respectively to get a rule dataset, then study the rule set again as the new dataset.
- 2) No original dataset. Sometimes, there is only rule dataset, but not the raw dataset. When learning the new knowledge, the model can't be learnt by using traditional machine learning, but using incremental SVM, with the incremental data for the inspection of known rule datasets and updates.
- 3) Real-time requirements. In this case, no all the original and new datasets for training is possibly studied using the classic SVM. Only the faster training speed of incremental SVM can be used to solve this.
- 4) More capable to reflect the changes of things. New sample data is increasing, and the support vector is changing which more accurately reflects the new dynamic data and is in favour of the change trend of the data for analysis.

### 4.2. Introduction of SVM Incremental Learning Algorithm

Because of the good learning performance, strong non-linear processing ability and good generalization ability of SVM, it has been a successful application in many areas since its launch. The researchers from various aspects have made improvements to the original SVM algorithm, and these efforts have greatly expanded the application field of the SVM. But as a new technology, there are still some limitations of SVM, for example, with the increase of sample

size, the SVM training speed may slow down significantly; it lacks the support for incremental learning; and is difficult to integrate prior knowledge and so on. The crux of these problems lie in: the classical SVM training practice algorithm finally boils down to a convex quadratic programming (QP) problem, as all the training samples participate in training at the same time to find out the optimal classification plane. If there is a new sample to join, there is a need to train all of the samples to get new optimal classification plane. That will be bound to bring to SVM the problem of slow training speed and large space complexity. The structure diagram of incremental learning algorithm is shown in Fig. (2).

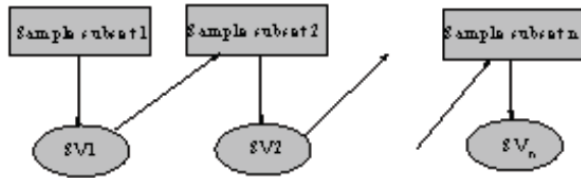


Fig. (2). Performance of SVM with batch training and incremental learning.

The idea of incremental algorithm is proposed to solve this difficult problem, that is to say that now the SVM algorithm can not only retain the existing knowledge and make analysis of the new problem in machine learning, but also reduce the time to study due to the addition of new samples. In fact, in the today’s information swelling age, we can’t hope that a very complete training dataset can be precisely defined and collected for the initial establishment of the system. We should learn to have the ability to do this; this is also the realistic significance of the SVM algorithm of incremental learning.

Experimental data comes from the Heart of UCI dataset; Scale, German and Banana. Two experiments were carried out in order to evaluate the algorithm performance: one is comparing the performance of SVM of the incremental training mode proposed in this paper and batch training mode SVM; the other is comparing the performance of our proposed incremental algorithm in this paper and the incremental algorithm proposed in other literature [2] (ISVM). Comparison is based on the classification accuracy of independent test dataset. Some of the dataset are not the problems of two classes classification one, but they have been processed into two kinds of classification problem, and all of the dataset has been processed into mean value 0, the standard deviation 1. The basic information of the dataset is shown in Table 1.

Table 1. Dataset and experiment parameters.

Name of Data Set	No. of Training Sample	No. of Test Sample	No. of Property	C	$\delta$
Banana	409	4800	2	316.340	1
German	708	300	20	3.162	55
Heart	178	100	13	3.162	120

Experimental platform: 2.9 GHz Intel dual-core processor, 2 GB memory, Matlab7.1 simulation was used. A total of 550 samples of 13D vector, were divided into five sample sets to simulate the whole process of increment; the sample size is 110 each, the original sample support vector number is 57, C takes value of 1500, radial Gaussian kernel function was adopted to be compared with the traditional algorithm, based on the internal moment algorithm and given in the literature, 13 were compared, while traditional algorithm is to combine the original sample set and incremental collection directly to generate a new training set, n takes 10. Kernel function is chosen as the radial basis function experiment, the width of the parameters is  $\delta$ . All experiments were carried on Windows 2000 system IntelP4PC (1.4 GHz CPU, 256 MB memory). The experimental results are as shown in Table 1.

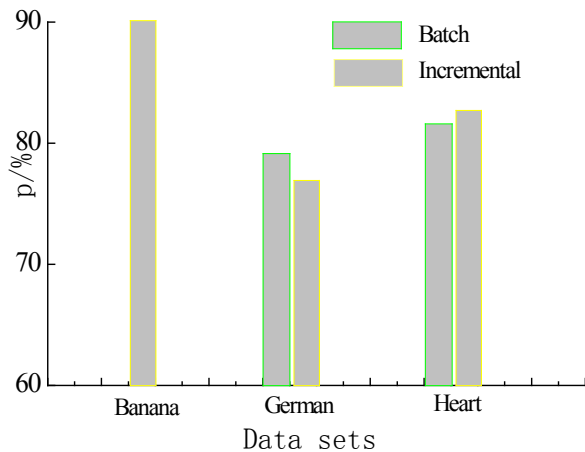
### 5. INCREMENTAL EXPERIMENTAL PROCESS

In different experiments, in order to have a convenient comparison, the same kernel function and its parameters should be chosen. The first experiment is to compare difference of the performance of SVM of the incremental training mode proposed in this paper and batch training mode SVM; the second experiment has two purposes:

- 1) Whether the proposed method meets the standard that incremental algorithm must possess.
- 2) Comparison of the performance of incremental SVM with the other method. To analyse both the cases, the following experiments are designed: first of all, the dataset is divided into training sets of mutually disjoint TR and TE test sets, and then they are further divided into 10 parts of training sets DS I, I = 1, ..., 10. Each part contains 10% of different training sets, which can realize the following two different ways of incremental learning [3].

The result of the average prediction accuracy p and batch training of proposed approach after 10 steps of incremental learning on the test set is shown in Fig. (3).

From Fig. (3), we can see that the performance of new incremental algorithm and SVM algorithm is equal, as there is no absolute advantage of any method, except that the batch of training SVM algorithm is difficult to deal with large datasets, while our proposed method can deal with large datasets. An efficient incremental algorithm [4] must meet the following three principles.



**Fig. (3).** Performance of SVM with batch training and incremental learning.

**Stability:** after every step of incremental learning, prediction accuracy of test set changes a little;

**Performance improvement:** prediction precision of the machine should be improved as the incremental learning progresses;

**Recovery ability:** a learning method should have the ability to respond to the performance, namely when the performance drops after a learning step, algorithm should have the ability to respond, *i.e.* either performance recovers or performs more than previously with the increase of learning step. The changes in prediction accuracy  $p$  in the test set are not very significant, which meets the requirements of the stability of incremental algorithm. It can be seen from the diagram that the prediction accuracy is improving with the increase of the incremental steps in general.

We can find from the Fig. (3) that incremental algorithm proposed in this paper has the ability of performance response. Therefore, the incremental training process suggested in this paper meets the requirements of incremental algorithm. The experimental results show that the incremental algorithm proposed in this paper share properties with the method proposed in literature [5], and in most of the performances of datasets and the methods. The experimental results also show that with the increase of incremental steps performance change decreases, finally increasing the training steps almost will not change the performance of classification. Thus, this algorithm can be used to estimate the expression of problems in a sample.

Incremental algorithm can better improve the modeling speed SVM, but how to ensure the accuracy well enough of the incremental algorithm directly affects the effectiveness of the incremental algorithm. This SVM incremental algorithm guarantees the precision, and at the same time faster training speed is obtained compared with other algorithms.

In addition, conventional SVM model uses the same error value for all the training samples; only call for independent identically distributed dataset; treat all of samples participating in the training equally; because of the existence of noise

and many other uncertain factors, some sample gravely deviate from the category of probability distribution leading to loss of different sample points and ordinary samples misclassification. Therefore, finding an effective learning algorithm is currently the main focus of the application of incremental SVM.

## CONCLUSION

Because the SVM has a good learning performance, strong nonlinear processing ability and good generalization ability; it has been successfully applied in many areas since its launch. But, there are a few shortcomings such as, large data calculation, long training time and too much difference between samples, too much influence on classification accuracy, and no support on incremental learning. This paper presents a new incremental training algorithm based on SVM. The main idea is to use the standard SVM algorithm to get the initial concept, and then use the concept put forward in this paper to update the method, that is by dissolving a convex quadratic programming problem similar to standard SVM algorithm. Compared with the performance of the batch training method, it shows that the new algorithm has a good performance. Another special point of the new algorithm is that the updated model has mathematical form, similar to that of the standard SVM that can get the solution of sparse representation. No additional calculations can step back. It can also be used to estimate problem of expressing the required sample size. If the incremental steps continue to increase, while the performance change is not significant, except the growing number of support vector, it will lower the efficiency of prediction. Thus, the incremental stop condition and solving the minimum support vector expression problems remain to be there for further research. In addition, conventional SVM model uses the same error value for all the training sample; only calls for independent identically distributed data; treats fairly all the samples in practice participating in the training; because of the existence of noise and many other uncertain factors, some sample gravely deviate are made to be far from the category of probability distribution; and there are different loss of sample points and ordinary samples misclassifications. Therefore, finding the effective learning algorithm is currently the main focus of study for the application of incremental SVM.

## CONFLICT OF INTEREST

The author confirms that this article content has no conflict of interest.

## ACKNOWLEDGEMENTS

This work is supported by the Key Project of Guangxi Social Sciences, China (No.gxsk201424), the Education Science fund of the Education Department of Guangxi, China (No.2014JGA268), and Guangxi Office for Education Sciences Planning, China (No.2013C108).

## REFERENCES

- [1] T. Hastie R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning Data Mining Inference and Prediction*, Beijing: Publishing House of Electronics Industry 2004, pp. 21-34.

- [2] L.S. Syednas, "Incremental learning with support vector machines", *Proceedings of the Workshop on Support Vector Machines at the International Joint Conference on Artificial Intelligence (IJCAI-99)*, Stockholm, Sweden, 1999.
- [3] V.N Vapnik, *Nature of statistical learning theory*, (Translated: Z. Xuegong), Tsinghua University Press: Beijing, 2000.
- [4] L.S. Syednas, *From Incremental Learning to Model Independent Instance Selection-a Support Vector Machine Approach*, National University of Singapore, Singapore 1999.
- [5] V. Vapnik, *The Nature of Statistical Learning Theory*", Springer-Verlag: NY, 1995, pp. 21-113.

---

Received: June 16, 2015

Revised: August 23, 2015

Accepted: September 28, 2015

© Xi Long Ding; Licensee Bentham Open.

This is an open access article licensed under the terms of the (<https://creativecommons.org/licenses/by/4.0/legalcode>), which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.