Support Vector Machine Classification Algorithm and Its Application

Yongli Zhang

Qinggong College, Hebei United University, Tangshan, China
zhyl01@163.com

Abstract. The support vector machine is a new type of machine learning methods based on statistical learning theory. Because of good promotion and a higher accuracy, support vector machine has become the research focus of the machine learning community. This paper introduces the basic theory of support vector machine, the basic idea of the classification and currently used support vector machine classification algorithm. Practical problems with which an algorithm, and proves the effectiveness of the algorithm, the final outlook of the prospects of support vector machines in classification applications. Finally the prospect of the prospect of support vector machines in classification applications.

Keywords: support vector machine, multi-classification, Optimization algorithm.

1 Introduction

Support vector machine is on the basis of statistical learning theory by Vapnik et al proposed a new learning method, which is built on the basis of a limited number of samples in the information contained in the existing training text to get the best classification results. In recent years, many scholars research focus, has become the practical problems of rigorous theoretical foundation, but also better resolve the small sample, nonlinear, high dimension and local minima points[1].

Initially, support vector machines for class classification problem, but with the rapid development of computer technology, network technology, database technology, for the classification and management of large amounts of information, the class classification problem can no longer meet people's needs. This method will be extended to multi-class classification problem. It is currently a hot research topic. Multi-class classification problem in the existing treatment methods have the following two categories: First, construct a series of second-class classification problem in some way, the final class classification problems to solve multi-class classification problem, it is called based on two fundamentally the multi-class SVM classification method. Second is to directly construct multi-class SVM, one-time to solve multi-class classification problem called multi-class SVM[2].

By example, the complexity of the algorithm will be minimized. Based on support vector machine classification using radial basis kernel function and select the appropriate parameters. This method can achieve very high classification accuracy.

C. Liu, L. Wang, and A. Yang (Eds.): ICICA 2012, Part II, CCIS 308, pp. 179–186, 2012.

2 Support Vector Machine Overview

The support vector machine is a novel small-sample learning method, because it is based on the principle of structural risk minimization, rather than the traditional empirical risk minimization principle, it is superior to existing methods on many performances. Support vector machine is a two dimensional description of the optimal surface evolved from the linearly separable case, the basic idea can be used in Figure 1. Figure 1 shows the basic idea .Two types are separated by H without errors. H_1 H_2 are places which pass the recent point of H .The distance of H_1 and H_2 was called class interval. Optimal separating surface is not only to ensure error-free separation of the two types of samples, also called the largest class interval.

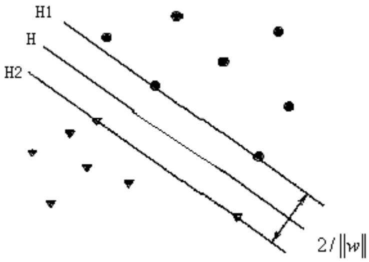

Fig. 1. Optimal separating surface

In pattern recognition, the linear discriminate function in the n-dimensional space: $g(x) = \omega \cdot x + b$, the classification hyper plane equation can be written $(\omega \cdot x) + b = 0$. Linear separable case, the discriminate function $g(x)$ was normalized. So that all training samples are met $|g(x)| \geq 1$, even to meet away from the classification of the surface of the sample $|g(x)| = 1$, so the class interval is equivalent to $\frac{2}{\|w\|}$, thus making the interval on the equivalent to $\|w\|$ or $\|w\|^2$. Make a classification of the surface of all samples correctly classified, it is necessary to meet:

$$y_i\left[(\omega \cdot x_i) + b\right] - 1 \geq 0, \quad i = 1, 2, \cdots, n \tag{1}$$

Satisfy the above equation (1)and make $\|w\|^2$ the smallest classification surface is the optimal classification surface. Point on the hyper plane are called support vectors (support vector), they support the optimal classification surface[3].

2.1 Linearly Separable

According to the above discussion, the optimal separating surface can be expressed as the following constrained optimization problem, Seeking the minimum of the following the function:

$$\phi(w) = \frac{1}{2}\|w\|^2 = \frac{1}{2}(w \cdot w) \tag{2}$$

We can define the Lagrange function as follows:

$$L(w,b,\alpha) = \frac{1}{2}(w \cdot w) - \sum_{i=1}^{n} \alpha_i \left\{ y_i \left[(w \cdot x_i) + b \right] - 1 \right\} \tag{3}$$

where, $\alpha_i > 0$ is coefficient of Lagrange. Problem is seeking the minimum of w and b to the Lagrange function. w and b seek partial differential and make them equal to zero, the original problem can be transformed into the following simple dual problem: the constraints

$$\sum_{i=1}^{n} y_i \alpha_i = 0 \tag{4}$$

$$\alpha_i \geq 0, \quad i = 1, 2, \cdots, n \tag{5}$$

The maximum value of solving the following function about α_i:

$$Q(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) \tag{6}$$

if α_i^* is Optimal solution,

$$w^* = \sum_{i=1}^{n} \alpha_i^* y_i x_i \tag{7}$$

The optimal weight coefficient vector of the surface is a linear combination of the training sample vector. According to the Kuhn-Tucker conditions, the optimal solution must also meet the following conditions:

$$\alpha_i \left(y_i \left(w \cdot x_i + b \right) - 1 \right) = 0, \qquad i = 1, 2, \cdots, n$$

For most samples α_i^* will be zero, the value of α_i^* zero corresponds to the type with equality (1)is support vector, they are usually all the samples part. To solve the above problems, the optimal Category function is

$$f(x) = \operatorname{sgn}\left\{ \left(w^* \cdot x \right) + b^* \right\} = \operatorname{sgn}\left\{ \sum_{i=1}^{n} \alpha_i^* y_i \left(x_i \cdot x \right) + b^* \right\} \tag{8}$$

$\operatorname{sgn}(\cdot)$ is the sign function.

2.2 Linear Inseparable Problem

The main idea of the linearly inseparable problem, the support vector machine is the input vector is mapped to a high-dimensional feature vector space, and constructs the optimal separating surface in the feature space. When the training set of linear non-time-sharing, some training samples can not satisfy conditions (1). Conditions can be amended to add a slack variable $\xi_i \geq 0$.

The constraints are:

$$y_i\left[(\omega \cdot x_i)+b\right]-1+\xi_i \geq 0, \quad i=1,2,\cdots,n \tag{9}$$

When misclassification occurs, $\xi_i \geq 0$, $\sum_{i=1}^{n}\xi_i$ is training set the upper bound of the number of misclassified. It can be used as description of the degree of training set is misclassified.

This makes it two goals: $\frac{2}{\|w\|}$ is as large as possible. $\sum_{i=1}^{n}\xi_i$ is as small as possible. These two goals into one goal, the introduction of the penalty parameter C as the right combination of these two target weight, under the constraints of the conditions (9) and the minimum of the following functions:

$$\phi(w,\xi)=\frac{1}{2}(w\cdot w)+C\left(\sum_{i=1}^{n}\xi_i\right) \tag{10}$$

Where : C is the penalty parameter. It actually plays a role in control of the degree of right or wrong sub-sample of punishment to achieve a compromise between the proportion of misclassified samples and complexity of the algorithm. Its optimal solution to the saddle point of Lagrange function is defined as follows:

$$L(w,b,\xi,\alpha,\beta)=\frac{1}{2}(w\cdot w)+C\sum_{i=1}^{n}\xi_i-\sum_{i=1}^{n}\alpha_i\left\{y_i\left[(w\cdot x_i)+b\right]-1+\xi_i\right\}-\sum_{i=1}^{n}\beta_i\xi_i$$
$$0\leq \alpha_i \leq C, \quad i=1,2,\cdots,n \tag{11}$$

Where: $\alpha=(\alpha_1,\cdots,\alpha_n)^T$ and $\beta=(\beta_1,\cdots,\beta_n)^T$ is all Lagrange multiplier vector. The specific method for solving Linear separable is very similar, but the condition $\alpha_i \geq 0$ is changed[4][5][6].

3 Support Vector Machine Classification Algorithm

Multi-class classification problem in mathematical language is described as follows: Training set $T=\{(x_1,y_1),\cdots,(x_l,y_l)\}\in(X\times Y)^l$, where: $x_i \in x=R^n$, $y_i \in Y=\{1,2,\cdots,k\}$, $i=1,\cdots,l$

Find a decision function $f(x)$: $X = R^n \to Y$

Thus, solving the multi-class classification problem is to find the point of R^n into the part of k the rules. If $k = 2$, Compared to second-class classification problem, directly using SVM to solve the problem; if $k \rangle 2$ for the multi-class classification problem. Handle multi-class support vector machine algorithm: One-against-rest algorithms, one- against -one algorithm, a directed acyclic graph algorithms, error correction coding algorithm, the binary tree algorithm. The following describes the One-against-rest algorithm, one-against-one algorithm, a directed acyclic graph algorithms.

3.1 1-a-r Algorithm(One-against-rest)

One-against-rest method is the first one in multi-class classification. In order to get multi-class classification machine, the usual approach is to construct a series of two types of classification machines, one type is divided with others types in every machines. Then x was divided. types Support vector machine classifier are constructed in one-against-the-rest method. When the first classifier is constructed, the sample of is marked as positive, or is marked as a negative class. When testing, Test data were calculated for each sub-classifier decision function value, and select the function values corresponding to the largest category of test data categories. The follow optimization question is solved in first i SVM:

$$\begin{aligned}
&\min_{w^i, b^i, \xi^i} \frac{1}{2}\left(w^i\right)^T w^i + C\sum_{j=1}^{l} \xi_j^i \\
&s.t. \quad \left(w^i\right)^T \phi\left(x_j\right) + b^i \geq 1 - \xi_j^i, \qquad if\ y_j = i \quad ; \\
&\left(w^i\right)^T \phi\left(x_j\right) + b^i \leq -1 + \xi_j^i, \quad if\ y_i \neq i
\end{aligned} \tag{12}$$

Solve the above optimization problem, you can get k decision functions:

$$\begin{gathered}
\left(w^1\right)^T \phi(x) + b^1 \\
\cdots \\
\left(w^k\right)^T \phi(x) + b^k
\end{gathered} \tag{13}$$

The sample x is input k-decision-making function and obtained k values, this sample belongs to the class which is the largest function value. The advantages of one-against-rest method are there are k types Support vector machine classifier and there is k Classification function. It is a few. The shortcoming of this method is training samples are all samples in every training. Quadratic programming problem of k are solved. Every one includes n variable. Training samples are too big, so training time is long.

3.2 1-a-1 Algorithm(One-against-one)

One-against-one method was proposed by Knerr, This approach is also based on two types of classification problems. Two different categories were selected to form a sub-SVM classifier, there are $k(k-1)/2$ sub-SVM classifier. When i and j sub-SVM were constructed, training sample are these sample which belongs to i and j classes, the sample of i is marked as positive, the sample of j is marked as a negative.

The follow optimization question is solved.

$$\min_{w^{ij}b^{ij}\xi^{ij}} \frac{1}{2}\left(w^{ij}\right)^T w^{ij} + C\sum_t \xi_t^{ij}$$
$$s.t. \quad \left(w^{ij}\right)^T \phi(x_t) + b^{ij} \ge 1 - \xi_t^{ij}, \quad if\ y_t = i; \tag{14}$$
$$\left(w^{ij}\right)^T \phi(x_t) + b^{ij} \le -1 + \xi_t^{ij}, \quad if\ y_t = j.$$

After optimization question is solved, $k(k-1)/2$ sub-SVM classifiers were obtained. When testing, data was tested by $k(k-1)/2$ sub-SVM classifiers. And the total score categories, select the highest scoring category corresponding to test data categories. To one-against-one method, voting method is used, multiple classes' votes are same, and some samples can't be divided[7][8][9].

4 Application

4.1 The Experimental Data

The experimental data come from Appendix D of data mining new methods - support vector machine. Iris plant data set is a standard data set used to test the classification algorithm performance. The data set of 150 sample points, divided into three categories: 1:Iris-setosa, 2:Iris-versicolor, 3:Iris-virginica. Each with 50 sample points, each sample has four characteristics were: sepal length, sepal width, petal length, petal width.

Table 1. Data sheet

	number of training samples	number of test samples	number of features	Category
Iris-setosa	30	20	4	1
Iris-versicolor	30	20	4	2
Iris-virginica	30	20	4	3

4.2 Experimental Results and Analysis

In order to verify the performance of the multi-class SVM algorithm, this paper presents 1-a-1 algorithm (one-against-the one) experiment, the algorithm using Matlab programming. Radial basis kernel function[10]

$$K(x, x_i) = \exp\left[-\frac{|x - x_i|^2}{\sigma^2}\right]$$

was selected. σ =0.05, Penalty parameter c =10.The samples were divided into three categories. Input vector x_i =(a_i , b_i , c_i , d_i) .

First get the SVM model SVM model will predict the unknown types of data .We can get the classification results. By the experimental data analysis, we can determine each set of data category.

Because samples will be divided three classifications, we must construct three two-class support vector machine models. Then to predict the data and finally a sample category was determined. Usually we make the following substitutions: Class 1 identifies the number 1 instead of Class 2 logo with the number 5 instead of 3 identifies the number 10 instead of replacement purpose is to solve the problems in the voting. Class1, class2, the class3 the classification's ultimate category of each input vector depends on the test sample.

A test sample after three support vector machine model to predict, their situation can only have several of the following table:

Table 2. Three SVM model to predict

	class12(1,5)	class23(5,10)	class13(1,10)	class
Class 1	1	5	1	7
	1	10	1	12
Class 2	5	5	1	11
	5	5	10	20
Class 3	1	10	10	21
	5	10	10	25
Blind spot	1	5	10	16
	5	10	1	16

It can be seen from Table 2, we can final class distinction between three categories of the two classification results mark to do the addition operation.

The samples belong to the Class 1 final value is 7 or 12, the samples belong to the class of the final value is 11 or 20 and belong to category 3 of the final value is 21 or 25. It should be noted that we can not use 1,5,10 to mark category. But to ensure ultimately do the addition operation, the results are not repeated. Due to the limitations of 1-a-1 algorithm itself, there are blind spots, we unable to conclude that it is the final category, with 16 marked.

5 Summaries

As a new type of machine learning methods SVM overcome the traditional problem of method had to learn and to fall into local minimum. It is with strong generalization ability. Using kernel function method, when samples are mapped high-dimensional space, method does not increase the computing complexity, while effectively overcome the curse of dimensionality. However, support vector machines still require further study. For example, in the multi-class classification problem, increase the speed of the training algorithm to improve the classification accuracy, as well as an effective solution to large-scale sample classification problems are the focus of future research. Support vector machine's performance relies heavily on the choice of the kernel function, different kernel functions and the coefficients of the application will get different results. There is still no good way to solve the optimal kernel function and kernel function coefficients choice. So far, support vector machines have been demonstrated in many areas, a good performance, there is reason to believe that in the future computer classification, support vector machine will play an invaluable role.

Acknowledgments. This paper was supported by Hebei Province science and technology research and guide project (Z2011328).

References

1. Vapnik, V.N.: Statistical learning theory. Springer, New York (1995)
2. Cortes, C., Vapnik, V.: Support vector networks. Machine Learning 20(3), 273–297 (1995)
3. Deng, N., et al.: Support vector Machine Theory, algorithms and Development, p.176. Science Press, Beijing (2009) (in Chinese)
4. Xue, N.: Comparison of multi-class support vector machines. Computer Engineering and Design (5) (2011) (in Chinese)
5. Zhang, Y., Zhu, Y., Lin, S., Liu, X.: Application of Least Squares Support Vector Machine in Fault Diagnosis. In: Liu, C., Chang, J., Yang, A. (eds.) ICICA 2011, Part II. CCIS, vol. 244, pp. 192–200. Springer, Heidelberg (2011)
6. Hu, Y., Zhuang, S., Peng, X., Xie, J., Chen, Y.: Products Serial Numbers Recognition Based on Support Vector Machine. Machinery & Electronics (2) (2012) (in Chinese)
7. Ding, Y., Qin, X., He, H.: Parameter Optimizing of Support Vector Machine and Application in Text Classification Computer Simulation (11) (2010) (in Chinese)
8. Li, G., Cui, G.: A Improved Algorithms of Fuzzy Support Vector Machines. Computer Measurement & Control (4) (2011)
9. Zhang, Z., Wang, S., Deng, Z., Chung, F.: A fast decision algorithm of support vector machine. Control and Decision (3) (2012) (in Chinese)
10. Gou, B., Huang, X.: SVM Multi-Class Classification. Journal of Data Acquisition & Processing (03) (2006) (in Chinese)