Research on Least Squares Support Vector Machines Algorithm

Zhao Ming

Bohai University, Jinzhou, P.R. China

zmfl2001@163.com

Abstract. Support vector machine is a classification algorithm emerged in recent years and has been successfully applied to many areas, and least squares support vector machine is a technology developed from the traditional support vector machine and has important researching significance. Firstly, this paper introduces the basic idea of SVM and algorithms; secondly to study the basic principles of least squares support vector machine, concrete algorithm description, including the kernel function, etc; finally, this paper studied the application of the algorithm in the classification, the least squares support vector machine as a novel artificial intelligence technology is an extension of the standard support vector machine and has been more widely used in various disciplines, with global optimization, good marketing ability and other features, so this research has some theoretical significance.

Introduction

With the development of computer, machine learning has achieved unprecedented development, machine learning based on the data is an important aspect of modern intelligence technology, research from the observed data samples off to find the law, using these laws to predict future data or unobservable data, the important basis of the existing machine learning methods, including pattern recognition, neural networks and others, is traditional statistics, the premise is that there are sufficient number of samples, namely the study of traditional statistical is the asymptotic theory when the number of samples tends to infinity and is the difficult to achieve desired results when the number of samples is limited. Statistical Learning Theory (SLT) is focused on proposing statistical rules and learning methods in the case of small sample, establishes a better theoretical framework for machine learning problems, and develops a new universal learning method - Support vector machine based on it [1].

Statistical learning theory is a theory that specializes machine learning law in the case of small samples and is established by Vapnik, and is established on the structural risk minimization principle. Support vector machine algorithm based on statistical learning theory has a complete theory, global optimization, adaptability, good generalization ability, etc., is the new hotspot of machine learning research. It not only minimizes the empirical risk, but also improves the generalization ability of the algorithm, has good application value and prospect [2]. Least squares support vector machine is presented on the basis of traditional support vector machine, the least squares linear systems is introduced into SVM, instead of the traditional support vector machine directly using quadratic programming to solve classification and function estimation problem, this paper mainly studies least squares support vector machine algorithm.

Support Vector Machine

Support Vector Machine (SVM) is a new machine learning method based on statistics proposed by Vapnik et al in mid-1990s, using Structural Risk Minimization (SRM) trains learning machine, it has a good learning ability, especially generalization ability. In theory, support vector machine gets global optimal solution and solves the problem of local minima cannot be avoided in the neural network; support vector machine by a nonlinear transformation to map the data into a high dimensional feature space, so that the data in the high-dimensional space can be classified by using linear discriminant function and ensures that the machine has good generalization ability; support

vector machine cleverly solves the dimension problem, complexity of the algorithm is independent of the sample dimension. Currently, SVM has been widely used in time series analysis, regression analysis, cluster analysis, signal processing, speech recognition, image classification, control systems and many other areas [3]. The basic idea of support vector machine is shown in Fig. 1.
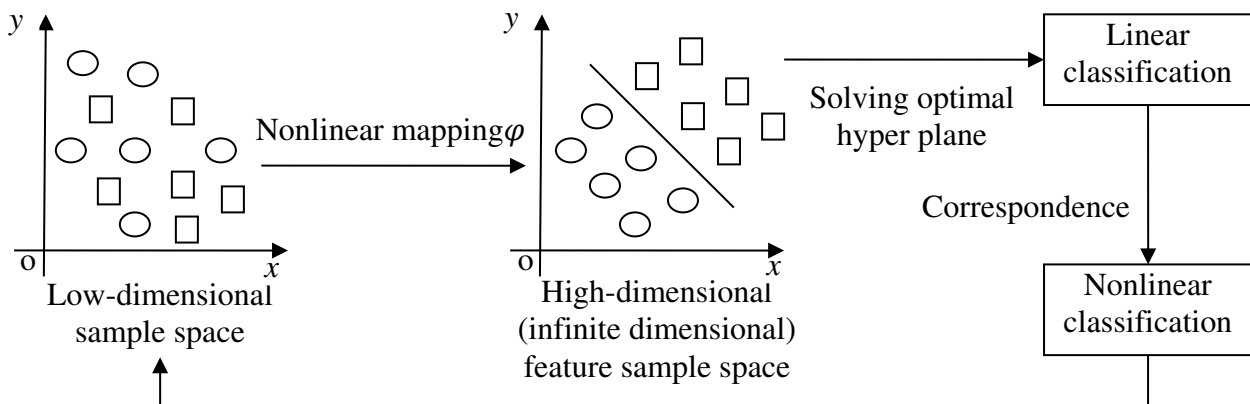

Fig . 1. The basic idea of support vector machine

SVM has the following characteristics: (1) nonlinear mapping is the basis of SVM method, SVM using the product kernel function instead of the nonlinear mapping in the high-dimensional space; (2) the goal of SVM is divided feature space into the hyper plane, the core of SVM is the idea of maximizing classification margin; (3) the trained results of SVM is support vector which plays a decisive role in decision-making; (4) SVM is a kind of novel small sample learning methods with solid theoretical foundation, and basically does not involve probability measures and law of large numbers, etc., avoiding the traditional process from deduction to induction and achieving the effective "transduction reasoning" from training sample to forecasting sample; (5)final decision function of SVM is only determined by the few support vectors, computational complexity depends on the number of support vectors, avoiding the "curse of dimensionality"; (6) the few support vectors determine the final result, which helps to grasp the key sample, "to reject" a lot of redundant sample and has better "robustness" [4].

Least Squares Support Vector Machines

Least Squares Support Vector Machine (LS-SVM) is a deformed SVM algorithm proposed by Suykens and Vandewalb in 1999, in which least squares algorithm usually represents the square sum of the component difference in mathematics, it does not solve the convex quadratic programming problems, only required solving a set of linear equations to obtain the optimal classification surface. The basic idea of the LS-SVM is that makes the original input space map into a high dimensional feature space by the nonlinear transformation defined by inner product function, looking for a linear relationship between input and output variables in this high-dimensional space, the most critical part of which is the kernel function [5, 6], the kernel function provides a method to extend the nonlinear learning machines for linear learning machine.

Kernel function definition: let X be a subset of R^n, if existing a mapping from R^n to Hilbert space Q,

$$\emptyset: \ R^n \to Q, \ x \to X \tag{1}$$

To get:

$$K(x, x') = \big(\varphi(x) \cdot \varphi(x')\big) \tag{2}$$

In which, $\big(\varphi(x) \cdot \varphi(x')\big)$ represents an inner product in space $Q, \varphi(x) = x \to X$, $K(x, x')$ is a kernel function on the $R^n \times R^n$. Because kernel learning has powerful nonlinear mapping ability, the current application of kernel function is very broad, commonly used kernel function are the following:

(1) Polynomial kernel function

$$K(x, x') = (x^T x' + m_1)^n \tag{3}$$

In which, m_1 is the kernel function parameters, n is the degree of polynomial.

(2) Perceptron kernel function

$$K(x, x') = \tanh(\delta(x, x') + m_2) \tag{4}$$

In which, δ and m_2 are kernel function parameters.

(3) Radial basis kernel function

$$K(x, x') = \exp\left\{-\frac{|\sigma(x)|^2}{\delta^2}\right\} \tag{5}$$

In which, δ is Gaussian kernel width, $\sigma(x) = x - x'$.

For above Kernel function, when choosing perceptron kernel function, SVM is a two-layer perceptron neural network, it can overcome the problem of local minima in neural networks; radial basis kernel function is the most widely used at present.

LS-SVM improves the objective function in the standard SVM, the sum of squared errors is introduced into the objective function of standard SVM and the inequality constraints are became the equality constraints, the solution process correspondingly became into solutions of a set of linear equations. The mathematical expression of LS-SVM algorithm is following: setting training sample set $G = \{(x_i, y_i)\}\, i = 1, 2 \dots m$, $x_i \in R^n$, in which, x_i represents inputting data, y_i represents outputting category logo, m represents the total number of data samples, after introduced nonlinear transformation and kernel function $K(x, x') = (\varphi(x) \cdot \varphi(x'))$, LS-SVM classification problems in the original space can be expressed as follows.

$$\begin{cases} \min F(w, a, \mu) = P + Q \\ \quad P = \frac{1}{2}\|w\|^2 \\ \quad Q = \frac{1}{2}C\sum_{i=1}^m \mu_i^2 \end{cases} \tag{6}$$

$$s.t. \quad y_i[w^T\varphi(x_i) + a] = 1 - \mu_i \tag{7}$$

In which, C is penalty parameter and $C > 0$, $i = 1, 2 \dots m$, constructed Lagrangian function as follows.

$$\begin{cases} L(w, a, \mu, \alpha) = F(w, a, \mu) + S \\ S = \sum_{i=1}^m \alpha_i \{y_i[w^T\varphi(x_i) + a] - 1 + \mu_i\} \end{cases} \tag{8}$$

Using mathematical knowledge to optimize the function in (8), you can get linear equations, so that the classification problem in the original space can be transformed into solving linear equations, and by solving linear equations ultimately get the decision function, decision function as follows.

$$\begin{cases} y(x) = \text{sgn}[Z + a] \\ Z = \sum_{i=1}^m \alpha_i y_i K(x_i, y_i) \end{cases} \tag{9}$$

In which, α_i is Lagrangian factor, a is classification threshold.

Algorithm Application

Algorithm application mainly studies the application of which LS-SVM algorithm in classification problems, mainly refer to the two classifications and Multi classification researching LS-SVM classification performance, firstly introduced the evaluation index of classification algorithm. The commonly used evaluation index includes classification accuracy in classifier [7], the running time of the classifier, ROC curve and so on. The classifier running time refer to the time after the initialization parameters, from the beginning to the end of the testing.

Classification accuracy mainly refers to the number of data after classifying data derived from the total number of correctly classified samples and data ratio, also known as classification accuracy, is to verify the quality of the most commonly used classification evaluation. In the LS-SVM, firstly to train training data set $(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)$, then based on the testing data set to test trained LS-SVM, in the testing results, if corrected accuracy is higher, it is considered the performance of this classifier is better, otherwise think this classifiers has worse performance [8]. ROC curve has the following characteristics, (1) make full use of the predicted probability values; (2) give differences between different distribution of different class; (3) the ROC curve of II classification through slope reflects the significant relationship between positive and negative examples, but also

reflects the relationship between the distribution and price of class; (4) make the classifier evaluation results in the form of curves more intuitive display in two-dimensional space [9]. In practice, Kernel function used least squares support vector machine mainly includes radial basis kernel function, polynomial Kernel functions and so on, the test sample mainly uses spiral sample data sets.

Conclusion

Support vector machine is an important classification algorithm in the stage of classification decision of pattern recognition, after proposed it has been widespread concern and application, its main task is data classification and regression. It combines a lot of technology, and has important application in the aspects of classification and fitting. This paper mainly studies an improved algorithm based on support vector machine: least squares support vector machines, which is an effective tool to deal with pattern classification under the case that sample set cannot be separated. LS-SVM can compensate for the lack of support vector machines in the aspect of classification accuracy and fitting accuracy, meanwhile improve significantly the solving efficiency of the traditional SVM and reduce the learning curve and then improve the traditional support vector machine performance [10]. This algorithm is suitable for handling large-scale learning problems, but lost the advantage of sparseness and robustness, if we can select contributed sample in some way, and then to train, so not only to ensure the accuracy of the algorithm but also to promote outreach capacity, and significantly reducing the number of training samples, thereby increasing the speed of the algorithm [11]. So the research for least squares support vector machine also has a broader developed space.

References

[1] G. X. Zhang, "The Essence of Statistical Learning Theory," Tsinghua University Press, 2000.

[2] L. Peng, "Research on Classification Algorithm of Support Vector Machine and Its Applications," Master's Degree of Hunan University, 2007.

[3] Q. Chen, N. G. Cao, J. L. Xie, "Research Progress of Support Vector Machine," Modern Computer, vol. 26, no. 4, pp. 47-49, 2009.

[4] D. S. Du, "Study on Text Classification based on Multi-class Support Vector Machines," Master's Degree of Chongqing University, 2007.

[5] S. G. Wang, "Properties and Construction Methods of Kernel in Support Vector Machine," Computer Science, vol. 33, no. 6, pp. 172-174, 2006.

[6] J. L. Guo, Y. S. Sun, S. X. Duan, "Research for Support Vector Machine and Kernd Funution," Science Technology and Engineering, vol. 8, no. 2, pp. 487-490, 2008.

[7] Kok Seng Chua. "Efficient Computations for Large Least Squares Support Vector Machine Classifiers," Pattern Recognition, vol. 24, no. 1, pp. 75-80, 2003.

[8] H. W. Liu, "Research on Least Squares Support Vector Machines in Multi-Classification," Master's Degree of Lanzhou Jiaotong University, 2013.

[9] X. H. Zou, F. Qin, K. Z. Cheng, Y. X. Wang, "Algorithm for Generating ROC Curve of Two-Classifier," Computer Technology and Development, vol. 19, no. 6, pp. 110-112, 2009.

[10] R. Cheng, "The Research and Application of Least Square Support Vector Machine," Master's Degree of Harbin Institute of Technology, 2013.

[11] R. Kong, B. Zhang, "Classification Algorithm of Fast Least Squares Support Vector Machine," Computer Engineering and Applications, vol. 43, no. 32, pp. 168-170, 2007.