Chapter 83
Research of Incremental Learning Algorithm Based on the Minimum Classification Error Criterion

Bo Wen, Ganlin Shan and Xiusheng Duan

Abstract Incremental learning is widely used in artificial intelligence, pattern recognition and other fields. It is an effective method to solve the system, where samples are less in the beginning of training, but over time its performance reduces. In this paper, based on the analyses of support vector machine and the characteristics of incremental learning, we proposed incremental learning method which is based on the minimum classification error criterion. Moreover, the validity and feasibility of this algorithm is verified through experiments.

Keywords Analog circuit · Fault diagnosis · Support vector machine

83.1 Introduction

Support vector machine (SVM) is a kind of pervasive machine learning model which is put forward by Vapnik et al., based on the statistical learning theory [1]. It uses the kernel function mapping nonlinear, thereby enhancing the learning machine performance [2, 3]. At present, the international research gets more and more attention [4, 5]. Machine learning is the basic problem in the field of artificial intelligence. Many learning system get few samples at the early stage of learning. As time goes by, samples accumulate ceaselessly [6, 7]. But the efficiency of the system is reduced. Then we need the new samples get into training to improve the system's performance [8]. If we process the new samples together with the old ones, the difficulty of learning will increase, and will consume too much time and

B. Wen (✉) · G. Shan · X. Duan
Optics and Electronic Department, Ordnance Engineering College, Shijiazhuang, China
e-mail: wenchuan321@126.com

storage space because sample set is too large. An effective method is to add, modify or the delete the part of original learning results which is only relevant to the new samples. Other irrelevant parts are untouched. This is the incremental learning. However, the learning results of support vector machine are support vector set. It is usually a small part of the whole sample sets, but can represent the feature of the entire sample set. This suggests that SVM incremental learning method based on support vector machine is feasible. By anglicizing support vector, this paper proposed incremental learning algorithm of support vector based on the minimum classification error criterion [9].

83.2 Problem Description

83.2.1 The Function of Support Vector Machine and Support Vector

Support vector machine is originated from classification problem under the linear case. The goal of using a support vector machine classification is to find the optimal hyperplane. Optimal hyperplane is the maximum-margin hyperplane, i.e., two kinds of samples are correctly classified, and at the same time, makes their classification interval biggest, as shown in Fig. 83.1 [10]:

As to the training sample set $(y_1, x_1), \ominus, (y_l, x_l)$, $x \in R^n$, $y \in \{-1, 1\}$, classification hyper plane's two optimization problem is

$$
\min_{\psi, b} \frac{1}{2} \|\psi\|^2 + C \sum_{i=1}^{l} \xi_i
$$
$$
s.t. \, y_i((\varphi(x_i) * \psi) + b) \geq 1 - \xi_i,
$$
$$
\xi_i \geq 0, \quad i = 1, \ominus, l
$$
(83.1)

Fig. 83.1 Optimal separating hyperplane diagram

Optimal hyperplane

margin $= 2 / \|\psi\|$
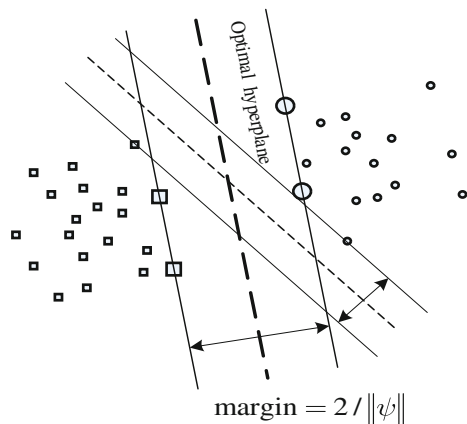
Its Lagrange function is

$$L(\psi, b, \alpha, \beta) = \frac{1}{2}(\psi * \psi) + C \sum_{i=1}^{l} \xi_i$$
$$- \sum_{i=1}^{l} \alpha_i(y_i[(\varphi(x_i) * \psi) + b] - 1 + \xi_i) - \sum_{i=1}^{l} \beta_i \xi_i \qquad (83.2)$$

Its dual problem is

$$\max W(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} y_i y_j \alpha_i \alpha_j \left(\varphi(x_i) * \varphi(x_j) \right)$$
$$= \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{l} y_i y_j \alpha_i \alpha_j K(x_i, x_j) \qquad (83.3)$$
$$s.t. \ C \geq \alpha_i \geq 0$$
$$\sum_{i=1}^{l} y_i \alpha_i = 0$$

Here, $K(x_i, x_j) = \varphi(x_i) * \varphi(x_j)$ is a kernel function.
Finally, we can get the decision function.

$$f(x, \alpha) = sign \left(\sum_{support\ vector} y_i \alpha_i^0 K(x_i, x_j) + b \right) \qquad (83.4)$$

The learning machine which constructs the decision function is called support vector machines (SVM). In the classification learning, only the support vector samples make contributions to optimal hyperplane and decision function. In general, X_i, which corresponds to the Lagrange multiplier α_i's value $0 < \alpha_i < C$, is called general support vector. But X_i, which corresponds to $\alpha_i = C$, is called boundary support vector. The former one represents all the samples which cannot be correctly classified; while the latter one represents the classification features of most samples. They together determine the form of the final classifier. That is to say, the support vector set can fully describe the features of training sample set. Its division is equivalent to the division of the entire sample set. In most cases, support vector set accounted for only a small part of training samples, therefore we can use support vector set to replace the training sample set for learning, if it does not affect the classification accuracy and reduce training time and memory space.

83.2.2 The Criterion of Minimum Classification Error

The simplest way of classification error measure is Bayes Discriminant which can be used to classify two kinds of problems. Its classification error measure can be defined as:

$$d(x) = P(C_2|x) - P(C_1|x) \tag{83.5}$$

Here, we assume that $P(C_2|x)(i=1,2)$ is a known posteriori possibility. This formula gives the possibility that observational samples of class 1 would be mistakenly classified into class 2. Its optimal decision boundary is the solution of $d(x) = 0$. For many such cases $(N > 2)$ of unknown distribution, they are different from Bayes Discriminant which classifies two kinds of problems and defines a classification error measure. Amari defines another classification error measure:

$$d_k(x) = \sum_{x \in S_k} \frac{1}{m_k} [g_i(x; \Lambda) - g_k(x; \Lambda)] \tag{83.6}$$

Here, $S_k = \{i | g_i(x; \Lambda) > g_k(x; \Lambda)\}$ is a confusion class set. m_k Is confusion number of $S_k \cdot S_k$ Is uncertain and it changes with the parameter set Λ and sample x. Therefore, this formula is discontinuous to ... It can't work out any derivative, so it's not suitable for gradient operation. There are many methods to define the Λ continuous classification error measure. One of them is:

$$d_k(x) = -g_k(x; \Lambda) + \left[\frac{1}{N-1} \sum_{j, \eta \neq k} g_i(x; \Lambda)^\eta \right]^{\frac{1}{\eta}} \tag{83.7}$$

In this formula, the second term on the right is the geometric average of all the other competing kinds. Parameter η can be regarded as a coefficient which adjusts the other competing kinds' contributions to the whole discriminant function. In the process of searching for the classifier parameter, many potential classifications can be found through the change of η. An extreme situation is that when $\eta \to \infty$, the discriminant function of the biggest competing kind in the second term plays a leading role. That is:

When $\eta \to \infty$, then

$$\left[\frac{1}{N-1} \sum_{j, \eta \neq k} g_j(x; \Lambda)^\eta \right]^{\frac{1}{\eta}} = \max_{j, \eta \neq k} g_j(x; \Lambda) \tag{83.8}$$

Classification error measure changes into:

$$d_k(x) = -g_k(x; \Lambda) + g_j(x; \Lambda) \tag{83.9}$$

Here, C_i, except C_K, is the kind which has the biggest discriminant value among all other kinds. This is because $(N-1)^{1/\infty} \cong 1$. Obviously, under this

circumstance, $d_k(x) > 0$ is a classification error. $d_k(x) \leq 0$ is the right classification. In this way, decision rule is changed into a problem of judging the scalar value.

83.3 Algorithm Design and Experiment Analysis

83.3.1 Algorithm Design

Assuming that the initial training sample set is $A = \{x_1, x_2, \ldots, x_M\}$, every $x_m(m = 1, 2, \ldots, M)$ is a K vector, and belongs to a certain category in $C_i(i = 1, 2, \ldots, N)$. As to a classifier which usually contains a set of parameters and a decision rule, the task of minimum classification error classifier design is: based on the given initial training sample set A find out the classifier parameters set and the associated decision rules, in order to get the minimum probability of any classification error sample $x_m(m = 1, 2, \ldots, M)$. Generally, the probability of classification error is approximated by error rate. If the cost of the related classification error exists, the goal of this kind of classifier is changed into: find out the suitable classifier parameter set Λ and the related rules to achieve the minimum expected cost.

The algorithm is as follows:

1. Initialize parameters Λ, set $t = 1$, select historical sample set A and new sample set I_t. Calculate the respective class center and sample forgetting factor $d(x_k)$ of the two kinds of sample in A. Determine the minimum classification error criterion collection B. Training B received initial support vector set SV^0 and decision function f^0.
2. If $I_t \neq \varnothing$, the algorithm will terminate. Otherwise, turn to 3.
3. Find out the samples I_t that violate the *KKT* conditions of decision function f^{t-1}, and determine set I_t^v.
4. If $I_t^v = \varnothing$, set $SV^t = SV^{t-1}, f^t = f^{t-1}, t = t + 1$, then turn to 2; or else turn to 5.
5. Through the incremental learning of $T = SV^{t-1} \cup I_t^v$, get a new decision function f^t support vector set SV^t, set $t = t + 1$, and turn to 2.

83.3.2 Experiment Analysis

This paper contains 3678 calibration samples. In order to verify the effectiveness and the feasibility of this algorithm, we apply incremental learning algorithm in the experiments of automatic classification of these samples based on the above research. In the test, we selected 643 samples as test sample set and 781 samples as the initial training set (H_1). The rest of the samples were randomly divided into five groups (H_2, H_3, H_4, H_5, H_6), which make up five incremental training sample

Table 83.1 The comparison between results of the new SVM incremental learning algorithm and the traditional SVM learning method

Training set	The number of incremental samples	The new *SVM* learning method		The traditional *SVM* learning method	
		Time/s	Precision (%)	Time/s	Precision (%)
H_1	781	98.7	93.5	171.5	92.7
H_2	398	76.4	96.9	232.6	95.1
H_3	113	118.5	93.4	328.4	92.8
H_4	407	136.3	94.8	487.3	93.4
H_5	282	103.8	96.5	546.2	93.8
H_6	1254	97.2	95.1	659.7	94.6

set. We have conducted comparative experiments between the new *SVM* Incremental Learning Algorithm and the Traditional *SVM* Learning Method. The experimental results are shown in Table 83.1 (*SVM* training nuclear function is using Gauss function. Software environment is MATLAB7.0. Hardware environment is Pentium 4/256 MB memory) [4].

From the results, compared with the Traditional Learning Algorithm, the classification accuracy of the *SVM* Incremental Learning Algorithm is improved, and its training speed is obviously improved.

83.4 Conclusion

Based on the characteristics of vector machine, we work out the *SVM* Incremental Learning Method. Furthermore, we put forward Incremental Learning Algorithm which is based on the minimum classification error criterion. This algorithm, in each round of the incremental training of training sample set, effectively keeps the former training samples which contain important information. It also simplifies the current incremental training sample set, and keeps those that are most likely to become the support vectors of training sample set for *SVM* incremental training. The experimental results show that this learning method is superior to the traditional incremental learning algorithm in both accuracy and time consumption.

References

1. Vapnik V (1995) The nature of statistical learning theory, vol 1. Springer Press, New York, pp 147–149
2. Zhu HB, Cai Y (2009) Text categorization based on active learning support vector machines. Comput Eng Appl 45(2):134–136

3. Kong R, Zhang B (2005) A fast incremental learning algorithm for support vector machine. Control Decis 20(10):1129–1136
4. Cao J, Liu Z (2007) Incremental learning algorithm based on SVM. Appl Res Comput 24(8):48–52
5. Ye S, Wang X, Liu Z, Qian Q (2011) Power system transient stability assessment based on support vector. Mach Increm Learn Method 35(11):15–19
6. Zhe X, Zhizhong M (2010) Incremental learning of support vector machine based on hyperspheres. J Northeastern Univ (Nat Sci) 1:16–19
7. Xiao R, Wang J (2001) An incremental SVM learning algorithm a-ISV. J Softw 12(12):1818–1823
8. Li D, Du S, Wu T(2006) Fast Incremental learning algorithm of linear support vector machine based on hull vectors. J Zhejiang Univ (Eng Sci) 40(2):203–215
9. Jiqing Han (2001) Discriminative learning method based on minimum classification error criterion. Electron Eng 27(2):1–12
10. Xiusheng Duan (2009) Research on fault diagnosis technology for a fire control system based on SVM. Shijiazhuang Mech Eng Coll 9:112–120