A handwritten signature recognition system based on LSVM

Chen jie ping

Guangxi Vocational and Technical College, department of computer and electronic information engineering, nanning, guangxi, China 530226

Abstract. Handwritten signature recognition has a wide application prospect in the field of identity identification. This paper presents a shape feature, dynamic feature extraction and recognition technology of LSVM state based on the combination of handwritten signature recognition system. Objective function changes in the experimental model of SVM, to LSVM, to avoid decomposition when solving the multiple solution algorithm of the problems existing in the SVM, for the purpose of improving the operating time of the system. The experimental results show that it has a high recognition rate and faster recognition rate.

Introduction

As a behavior characteristic, handwritten signature has its own biological features as relative stability and imitative difficulty. In many cases such as signing a contract, agreement or making any certificate, handwritten signature is usually indispensable and has irreplaceable legal status and importance. Hence, handwritten signature recognition has a wide application prospect in the field of identity identification [1]. However, as the dynamic feature of handwritten signature loses during the writing process, factors that can represent the signer's feature become less, which leads to a low recognition rate of handwritten signature. This paper introduces a LSVM based handwritten signature recognition system which has a high recognition rate and fast recognition speed according to the experimental result of imitation.

LSVM based handwritten signature recognition system

The design idea of the handwritten signature recognition system in this paper is: conduct a pre-treatment of the image of the handwritten signature, including the smoothing, binarization, normalization and refinement of the signature image; extract the features of the pre-treated image from the aspects of shape features and pseudo dynamic characteristics [2], and select them with Bhattacharyya algorithm; at last, use LSVM as the algorithm for the recognition and get the handwritten signature recognition system [3]. The process of LSVM based handwritten signature recognition system is shown in Figure 1.
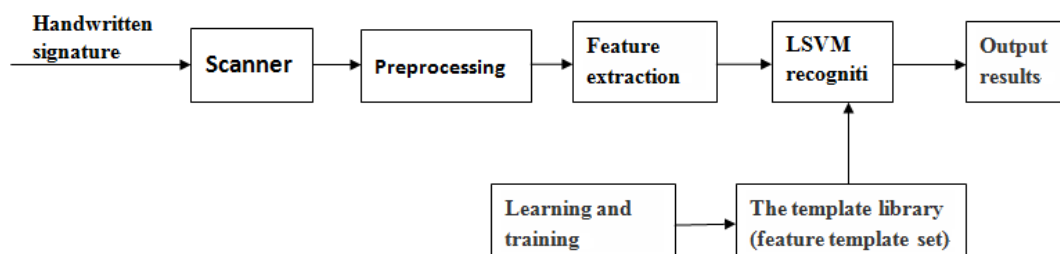

Figure 1 Process of LSVM based handwritten signature recognition system

LSVM (Lagrangian Support Vector Machine)

LSVM is a new machine learning method [4] proposed by Vapnik et al. Due to its outstanding learning performance, LSVM has a wide application in various fields such as detecting and discriminating human faces, recognizing handwriting letters and categorizing articles.

The idea that supports the LSVM is: to make the sample information compromise between the model complexity and learning capability so as to obtain the best generalization ability that is the ability of predicting future output accurately. The idea is based on the VC theory in statistical theory and the principle of the minimum structure risk, and LSVM is to find out the optimal classification plane between two different categories of samples on the basis of minimum structure risk.

Figure 2 is a sketch of optimal classification plane. H stands for classification plane; H_1 and H_2 refer to hyperplane; the classification interval of the two categories is the distance between H_1 and H_2. The optimal plane means a plane that can not only separate the two categories accurately but also has a maximum classification interval between the two categories.
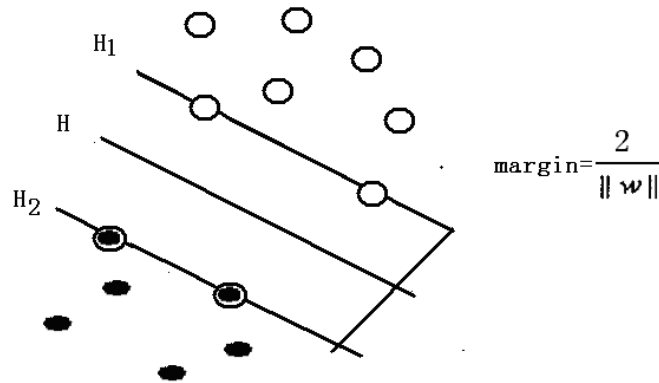

$$margin = \frac{2}{\|w\|}$$

Figure 2 Optimal classification plane

We assume the sample set as (x_i, y_i), $i = 1, \ldots, n$, $x_i \in R^d$, $y_i \in \{+1, -1\}$ as the category label, the sample sets of C1 and C2 are linearly separable sets, that is (w, b) is existed, then

$$(wx_i) + b > 0 \quad x_i \in C1$$
$$(wx_i) + b < 0 \quad x_i \in C2 \qquad (1)$$

The linear discrimination function is $g(x) = (wx_i) + b$. The purpose of classification is to find out (w, b) to separately C1 and C2 optimally.

Normalize the discrimination function and assure all the samples satisfy $|g(x)| \geq 1$, that is to constrain the (w,b) as,

$$\min_i |g(x)| = 1$$

The interval of classification is $\frac{2}{\|w\|}$. Under such circumstances, we have to satisfy the following function to make assure all the classifications of the samples are correct:

$$y_i[(wx_i) + b] \geq 1, i = 1, \ldots n \qquad (2)$$

According to Figure 2, the classification interval of the two categories is $\frac{2}{\|w\|}$. To increase the interval means to reduce $\|w\|$ (or $\|w\|^2$). Optimal classification plane means the plane that can satisfy

Formula (2) with minimum $\|w\|^2$ value. Support vector machine refers to the testing samples of the two categories. The optimal classification can be expressed as:

$$\min_{w,b} \frac{1}{2}\|w\|^2$$

（constraint condition） $\quad s.t. \quad y_i[(wx_i)+b] \geq 1, i = 1,....n \qquad$ （3）

The quadratic programming problem with linear constraints in Formula (3) can be solved by using Lagrange multiplier method and is defined as following Lagrange function:

$$L（w,b,a） = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{I} a_i\{y_i[(wx_i)+b]-1\} \qquad （4）$$

$s.t. \quad a_i > 0$

If the two sides of Formula (4) take the partial derivatives 0 of w and b, the optimal question should satisfy the following formula according to Kuhu-Tucker condition:

$$a_i[y_i(wx_i)+b]-1 = 0\} \quad i = 1,2,...,n \qquad （5）$$

There will be very few a_i (that is Lagrange multiplier) that is not 0. The relative sample x_i of a_i is the Support vector machine, which is the points on hyperplanes of H_1 and H_2. The maximum objective function is:

$$Q(a) = \sum_{i=1}^{n} a_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j a_i a_j (x_i y_i) \qquad （6）$$

Then the optimal solution is： $\quad w^{'} = \sum_{i=1}^{n} a_i y_i x_i$

After above solutions, the optimal classification function is:

$$f(x) = \text{sgn}[(w^{'}x)+b^{'}] \qquad （7）$$

where， sgn（ ） refers to sign function， $b^{'} = y - w^{'}x_i$

If we use inner product function $K(x,x^{'})$ to replace the dot product in the optimal classification plane, the optimal function will be:

$$Q(a) = \sum_{i=1}^{n} a_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j y_i y_j K(x_i, y_i) \qquad （8）$$

Related classification function will be:

$$f(x) = \text{sgn}(\sum_{i=1}^{n} a_i^{'} y_i K(x_i, \ x)+b^{'}) \qquad （9）$$

Other conditions in the algorithm will keep the same. This is the nonlinear SVM.

However, as standard support vector machine solves dual problems mainly by quadratic programming, the training speed is very low. Besides, there will be plenty of matrix operations with this algorithm. As a result, in recently years, many scientists introduced the LSVM (Lagrangian Support Vector Machine), which is proposed by OlviL Mangasarian in 2001[6]. LSVM made a small change on the objective function of SVM model so that to change the dual problem into a minimum quadratic function which has no upper limits, then solve with iteration. In this way, LSVM can avoid the problem of decomposition for many times and improve the algorithm training speed greatly.

If there are m sample points in the n dimensional space that need to be categorized and are represented by mxn matrix A; the element d_n in the diagonal matrix D is +1 or -1; the category that

represents the i sample A_i is A^+ or A^-. After minimize the error variance y in the standard SVM, we would get the following classification plane:

$$x^T w = \gamma \qquad (10)$$

Expression of the classifier decided by the classification plane is:

$$x^T w - \gamma \begin{cases} > 0 \ , & then \ x \in A^+ \\ < 0 \ , & then \ x \in A^- \\ = 0 \ , & then \ x \in A^+ \ or \ x \cup A^- \end{cases} \qquad (11)$$

The original problem of SVM standard model can be expressed as:

$$\min \eta e^T y + \frac{1}{2} w^T w$$

$$D(Aw - e\gamma) + y \geq e$$

$$s.t. \ y \geq 0 \qquad (12)$$

The dual problem is：

$$\min W(a) = \frac{1}{2} a^T D A A^T D a - e^T a \qquad (13)$$

$$0 \leq a \leq \eta e$$

$$s.t. \ e^T D a = 0$$

At first, LSVM will make following changes on the objective function in Formula (13):

1. Use 2NF instead of 1NF for the error term, that is to replace $\eta e^T y$ with $\eta \frac{y' y}{2}$. So the constraint condition $y \geq 0$ can be omitted.

2. Add $\frac{1}{2} \gamma^2$ into the objective function and the objective function will be $\frac{1}{2} w^T w + \frac{1}{2} \gamma^2 + \eta \frac{y^T y}{2}$. The purpose of the change is to remove the upper constraint problem of the related dual problem.

After the change, the original problem will be:

$$\min \frac{1}{2} w^T w + \eta \frac{y^T y}{2} + \frac{1}{2} \gamma^2 \qquad (14)$$

$$s.t. \ D(Aw - e\gamma) + y \geq e$$

Construct the Lagrange function for the dual problem

$$L(w, \gamma, a) = \frac{1}{2} w^T w + \eta \frac{y^T y}{2} + \frac{1}{2} \gamma^2 - a[Aw - e\gamma) + \gamma - e] \qquad (15)$$

Take the derivatives respectively of w, γ, y in the above formula and set them 0, we gain:

$$w = A^T D a \ , \quad y = \frac{a}{\eta} \ , \quad \gamma = -e^T D a \qquad (16)$$

Substitute them into Formula (15), we get the dual problem:

$$\min_{0 \leq a \leq R^m} \frac{1}{2} a^T (\frac{1}{\eta} + D(AA^T + ee^T)D)a - e^T a \qquad (17)$$

As shown, the revised problem was changed to an optimization problem without upper constraint condition and can be simplified as:

$$H = D[A - e], Q = \frac{1}{\eta} + HH^T$$

So the dual problem Formula (16) becomes:

$$\min_{0 \le a \le R^m} \frac{1}{2} a^T Q a - e^T a \tag{18}$$

We assume $\gamma = Qa - e$, then the KKT（Karush Kuhu Tucher） condition of the optimization problem for the Formula (16) will be:

$$a \perp \gamma \quad (a \ge 0, \gamma \ge 0) \tag{19}$$

$a \ge 0$ and $\gamma \ge 0$ means both a and γ should be non-negative vectors

Any two real numbers (or vectors) a and b can satisfy the following theorem:

$$0 \le a \perp b \ge 0 \Leftrightarrow a = (a - \beta b)_{+,} \beta > 0 \tag{20}$$

Formula (19) (that is KKT condition) can be changed into following equation form (β can be any positive number):

$$Qa - e = ((Qa - e) - \beta a)_{+} \tag{21}$$

According to the KKT condition above, we can get Formula (21), which is an iterative formula that will lay an important basis for LSVM algorithm. The formula is shown as:

$$a^{i+1} = Q^{-1}(e + ((Qa^i - e) - \beta a^i)_{+}), i = 0,1,...\infty \tag{22}$$

This iterative formula only needs to satisfy:

$$0 \le \beta \le \frac{2}{\eta} \tag{23}$$

We can get an optimal solution of linear convergence in Formula (18) iterated from any initial point. Here we take $\beta < \frac{1.9}{\eta}$.

In order to get Q^{-1}, we need to calculate the inverse matrix of m x n matrix.It will a large amount when the number of sample point is large. Here we use to Sberman Morrison Woodbury （SMW）to solve the inverse matrix:

$$Q^{-1} = (\frac{I}{\eta} + HH^T)^{-1} = \eta(\frac{I}{\eta} + H^T H)^{-1} H^T) \tag{24}$$

Where,H is a m×(n+l) matrix. So after the change, an inverse matrix problem of a large m×m matrix was turned into that of a small (n+l)×(n十1) matrix. It makes the solution of Q^{-1} possible, and even accelerates the solution, which enables LSVM algorithm to solve the linear classification problem very fast.

Instruction of the samples used in the system and experimental results

In the experiment, we collected 400 handwritten signatures from 10 different age groups. First we asked the 10 subjects to sign 20 signatures in different time and environment so that we obtained 200 real signatures; then we asked them to imitate 20 signatures from others, and we got 200 false signatures. So we have 400 signatures in total, 50% of real signatures and 50% of false signatures. During the experiment, we divided the real signatures into two groups, 100 of them were combined with 40 false signatures and were treated as the training samples. Both of the training samples and the rest samples were regarded as the sample of experiment recognition.

Here we take "林建书"as an example to analyze the experimental results.

Figure 3 and Figure 4 are some of the real and false signatures of "林建书".
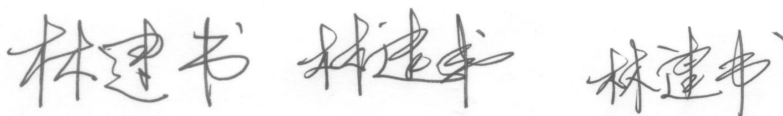
Figure 3 Real signatures of "Lin Jianshu"

Figure 4 False signatures of "Lin Jianshu"

First, we conduct a pre-treatment on the original image and extract 9 features of the treated image from the aspects of shape and pseudo dynamic characteristics. But not all the features will be used. We would use Bhattacharyya to select the features and get 7 of the features for the distinction and recognition, which include: depth-width ratio of the level-compressed signature, black dot area and total area ratio, relative gravity in the horizontal and vertical directions, gray feature of skeleton direction, high gray level stability feature, grayscale distribution histogram of the handwritten signature.

The experiment was conducted with the LSVM Handwritten signature recognition algorithm. The experimental results are shown in Table 1:

Table 1 Experimental results obtained with LSVM algorithm

	Correct recognition rate of real signature	Correct recognition rate of false signature	Averaged correct recognition rate
Training accuracy	87.9%	80.1%	84%
Testing accuracy	86.8%	82.86%	84.83%

Conclusion

This paper proposed a handwritten signature recognition system based on LSVM recognition technology combined with extraction of shape feature and dynamic feature. The system extracted 9 different features of the handwritten signature image from the aspects of still and dynamic characteristics and eventually selected 7 features as the level for distinction and recognition so that to obtain a better recognition rate with minimum features. Based on the off-line LSVM algorithm recognizer, the experiment changed the objective function of SVM model into LSVM, which avoided the multiple solution problem existed in the SVM and improved the operation time of the system.

References：

[1] Wang Shi, Wang Yuansheng. The Arrival of Biometric Time [J].Guangming Daily,2002(58）:20-22.

[2] Xie Yan, Identification Research of Off-line Chinese Signature Based on PseudoDynamic Feature Extraction [D]．Shandong University，2007.

[3] Fen Zongzhe, Cheng Xiangjun.Model Recognition Principle[M].Xi'an：Science and Technology Press of Xi'an Unitersity，2003.

[4] Vapnik VN. The Nature of Statistica of Learning Theory [M]. New York Spring. 1995.

[5] Zhang Xinzhong. Chinese Characters Recognition Technology [M].Beijing：Tsinghua University Press，1998.

[6] Wang Peng，Zhu Xiaoyan.Selection and Application of Model Based on RBF Kernel SVM [J].Computer Engineering and Applications, 2003(24):72-73.