

On-line Signature Verification Based on GA-SVM

<http://dx.doi.org/10.3991/ijoe.v11i6.5122>

HUANG Dong¹, GAO Jian²

¹ QinZhou University, QinZhou, China

² Panjin Vocational and Technical College, Panjin, China

Abstract—With the development of pen-based mobile device, on-line signature verification is gradually becoming a kind of important biometrics verification. This thesis proposes a method of verification of on-line handwritten signatures using both Support Vector Data Description (SVM) and Genetic Algorithm (GA). A 27-parameter feature set including shape and dynamic features is extracted from the on-line signatures data. The genuine signatures of each subject are treated as target data to train the SVM classifier. As a kernel based one-class classifier, SVM can accurately describe the feature distribution of the genuine signatures and detect the forgeries. To improving the performance of the authentication method, genetic algorithm (GA) is used to optimise classifier parameters and feature subset selection. Signature data form the SVC2013 database is used to carry out verification experiments. The proposed method can achieve an average Equal Error Rate (EER) of 4.93% of the skill forgery database.

Index Terms—Online Signature Verification, SVM, Data Description, Genetic Algorithm, Feature Selection .

I. INTRODUCTION

With the rapid development of e-commerce, personal communication, computer, Internet, mobile devices with a handwriting input function more popularization, provide favorable conditions for the development of the use of online handwritten signature application [1]. Compared with the face, fingerprint and other biometric authentication method, handwritten signature verification is a noninvasive, convenient application authentication method. The signature is the traditional identity authentication form of social activity, it also has the obvious characteristic of easy popularization [2].

Online signature verification technology is developing in recent years. Its core is to capture the handwriting and the strength of the basic information automatically in the process of signing in handwritten board. Compared with the traditional methods, it is more accurate and achieve more ideal recognition effect. Off-line signature verification methods are commonly used, focusing on the handwriting characteristics, which contains blindness and contingency [3]. But online authentication method can accurately obtain relevant information according to the actual characteristics of the signature based on the construction of appropriate statistical models or signature template. Then the authenticity of the signature can be obtained according to the corresponding standards.

As for the original signature data shape, pressure and other information, a variety of features of the signature

can be achieved by statistical analysis, spectrum analysis, coding analysis method [4]. However, there are often many feature extraction in a system with redundant information, and even the interference information data, if they are flowing into the classifier, which will seriously affect the treatment effect of classifier, and can enhance the complexity of the system to a great extent. In this case, we must find the ideal feature subset has high classification performance, while minimizing the subset should be able to realize the characteristic dimension.

The ideal performance of feature subset directly affect the processing precision of the system, especially samples of small size, the free parameters will this classifier is very few, the resulting parameter estimates the effect will be more precise. In the matching of signature verification, dynamic time warping (DTW) and hidden Markov model (HMM) and artificial neural network classifier and the Gauss is a commonly used method. These signature authentication methods have achieved good results in the experiment, but still facing how to adapt to the change of signature handwriting process or need more training samples problem.

Signature verification in essence can be regarded as a single class classification problem, a real signature feature is relatively stable, and forge a signature is changeable, so to accurately describe the more meaningful the target sample distribution [5]. Support vector machine data description is one classification method based on a support vector machine (SVM) and statistical learning theory, It directly to seek an inclusive of the target sample hyper sphere as the target categories of boundary, and the number of training samples required relatively less. This paper introduces a description for the online signature verification method based on support vector machine data classifier. Using the characteristic data of real signature to establish the corresponding classifier [6], for the signature classification in authentication phase. In the process of training, the joint optimization of classifier parameters and feature subset selection can be obtained by using genetic algorithm to do. Finally, signature verification experiments can be done in a public signature database.

II. SIGNATURE DATA

For online handwritten signature authentication method, is a way that can be generally obtained by the original features of handwritten signature digital board. It can acquisition the information of the signature trajectory sampling point coordinate, pen pressure pen angle on real time. In this paper, using the signature database of International signature verification competition (SVC2013)

in 2004 for research and verification experiment. The database signature with WACOM company Intuos handwritten board which collect samples per second, 100 points for each sampling point, including X, Y coordinates, coordinate time marker, key state, azimuth, dip angle and pressure of seven dimensional data. The signature data can be expressed as a sequence of points [7]:

$$(x(s), y(s), t(s), b(s), \alpha(s), \beta(s), p(s)), s \in [1, S])$$

Among them, the number of S are sampling points.

The open part of the SVC2013 signature database contains 40 signer's signature, each group consists of 20 signatures written by the true signers and 20 provided by at least 4 other signers forgery. 20 real signatures is collected by two times, 10 each time, sampling interval for at least a week, in order to reflect the signature change over time. Forge a signature by the forger to observe the dynamic process of the appropriate signature and practice writing, thus being a skilled signature. These signatures in both Chinese signature also has English signature, Therefore the authentication method can not rely on the text type. Figure 1 is the two set of signature data sample database.

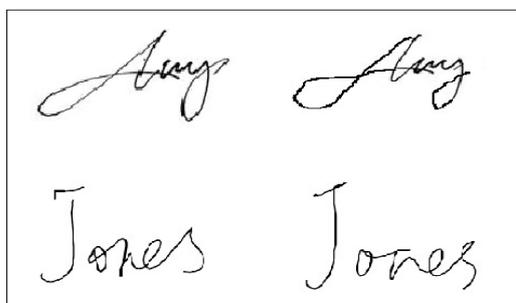


Figure 1. Signature sample SVC2013 (real signatures are on the left, the right side are forgery).

III. FEATURE EXTRACTION

A. Pretreatment

In the signature acquisition process, due to its interference factors and manning the jitter handwriting board, it can cause interference and redundant points in the signature data. In addition, the signer each signature writing size and location coordinates are also different. Preprocessing is mainly for smoothing and size data normalization, consistency of the extracted signature features to improve, to eliminate the influence of these interference factors on the recognition results [8].

Signature of the original data is accurately got by using Gauss filter and then X coordinates are achieved through smoothing the corresponding as follows:

$$x'(s) = \sum_{i=-2\sigma}^{2\sigma} \omega_i x(s+i) \quad (1)$$

$$\text{Among them: } \omega_i = e^{-\frac{i^2}{2\sigma^2}} / \sum_{j=-2\sigma}^{2\sigma} e^{-\frac{j^2}{2\sigma^2}} \quad (2)$$

Accordingly, smooth components $x(s), y(s), \alpha(s), \beta(s), p(s)$ can be carried out. The filter parameters are taken as 1, ie. taking the sliding window width of 5 sampling points.

The signature of the location of translation and size normalization is completed simultaneously by coordinate transformation as follows.

$$\begin{aligned} x'(s) &= (x(s) - x_c) * M / \max(w, h) \\ y'(s) &= (y(s) - y_c) * M / \max(w, h) \end{aligned} \quad (3)$$

Among them, $w = \max_{s \in [1, S]}(x(s)) - \min_{s \in [1, S]}(x(s))$ as the original signature width, $h = \max_{s \in [1, S]}(y(s)) - \min_{s \in [1, S]}(y(s))$

as the height; $x_c = w/2, y_c = w/2$ the center position of connected rectangular representation area outside; M indicates that after normalization, the resulting external each side numerical size of the quadrilateral. At this time, the outer frame of regional range corresponding signature will be located within the standard size of the corresponding numerical invariant, while the aspect ratio.

B. Feature extraction

There are two main types of features which are extracted from the signature data acquisition: the shape and dynamic feature [9]. Shape feature is describing the static characteristic shape characteristics are extracted from the signature coordinate, dynamic characteristics can reflect the signature process speed, pressure and other information. The extraction of dynamic characteristics by using the time information, and pressure, azimuth information than shape is more subtle than the shape feature, so the richer, more difficult to imitate, but the stability of some worse.

Generally, these two types of features a complementary role to play in the signature verification. Because at the same time from both the shape and dynamics to mimic signature alone than from one side to be much more difficult to imitate. According to the characteristics of signature data and preliminary experimental results, while referring to other discussion in the literature, this paper collected 8 kinds of shape features and 19 dynamic features constitute the signature feature set, The shape features are: the ratio of height to width, ratio of length to width, center of gravity, center of gravity than the stroke direction distribution; Dynamic characteristics are: Signature duration, total time, write time ratio(velocity, pressure), the (average value, maximum value, average value, variance ratio, the maximum relative time value and maximum value), X direction, Y direction velocity when the time than speed coincides with the time ratio, average angle, azimuth angle the mean value, variance, the variance of the azimuth angle.

IV. SIGNATURE VERIFICATION OF DATA DESCRIPTION BASED ON SUPPORT VECTOR MACHINE

A. Support vector machine theory

Support Vector Machine (SVM) is proposed by Vapnik. In the process of optimization, setting the constraint conditions for the training error, the optimization objective is to minimize the confidence range. SVM is a learning method based on the minimization criterion, compared with some traditional learning methods, which obviously and has better generalization ability. In the SVM solution process will eventually translate into

solving quadratic programming problems, so the solution is obtained SVM global optimal solution.

Problem solving optimal classification plane can be transformed into a constrained optimization problem, shown as below [10]:

$$\text{minimize } \frac{1}{2} \|\omega\|^2 \quad (4)$$

$$\text{Subject to } y_i[wx_i + b] - 1 \geq 0 \quad i = 1, 2, \dots, l \quad (5)$$

Type describes a two convex programming problem, because of the objective function and constraint conditions are presented convex properties, according to the optimization theory, the programming problem has a unique global minimum. Applying Lagrange multiplier method and making it satisfy the conditions of KKT (Karush-Kuhn-Tucher):

$$\alpha_i \{y_i[wx_i + b] - 1\} = 0 \quad (6)$$

So we can get an optimal classification function, to solve the above problem:

$$f(x) = \text{sgn}\{w^* \cdot x + b^*\} = \text{sgn}\left\{\sum_{i=1}^k \alpha_i^* y_i (x_i \cdot x) + b^*\right\} \quad (7)$$

B. Support vector machine data description

Classification of data description algorithm is a class of typical single value, support vector machine data presented by Tax and Duin description of theory development is already quite mature [11]. It is not by estimating the probability distribution of the target sample data to detect outliers, but directly to seek an inclusive of the target sample hyper sphere as the target category boundaries, so it needs less number of training samples. Because of the use of this method, the classifier in the test sample distribution and existing distribution of sample both training different circumstances, can show the generalization performance of the ideal. Support vector machine data description based only on a class of target sample data, we can establish a classifier is used for outlier detection, obtain good effect of in the application of fault diagnosis and face recognition [12].

In the object the training sample data set $\{x_i\}, i=1, L, N$, which includes an N, it can define covering hyper sphere concept from all the samples, the super ball center and radius using a R quantitative. Support vector machine data describe the core theory for minimizing the use of hyper sphere volume way, realize the false acceptance probability outliers to minimize. Support vector machine data description is based on the theory of support vector machine (SVM) the basic idea of the development and the formation. By introducing the kernel function data nonlinear mapped into a high dimensional space to the construction of data is more flexible and compact description.

Support vector machine data describe the training process of the classifier to solve the maximization problem is as follows:

$$L = \sum_i \alpha_i \cdot K(x_i \cdot x_i) - \sum_{i,j} \alpha_i \cdot \alpha_j \cdot K(x_i \cdot x_j) \quad (8)$$

The constraint conditions are:

$$\sum_{i=1} \alpha_i = 1$$

$$a = \sum_i \alpha_i \cdot x_i, \quad 0 \leq \alpha_i \leq C$$

The above formula, C indicates the false reject error and the balance between data of the target sample description complexity: the kernel function $K(x_i \cdot x_j)$, commonly-used kernel functions including polynomial inner product, radial basis (RBF), sigmoid inner product etc. Minimization of type (8) can get a set of constraints $\{\alpha_i\}$, second equations illustrate the sphere for linear combinations of sample points, the weight coefficients corresponding to the α_i . The related theory by Lagrange multipliers, $\alpha_i = 0$ describes the sample points in the interior areas inside the sphere, $\alpha_i > 0$ describes the sample points in the position of the spherical hyper sphere, $\alpha_i = C$ indicates that the corresponding sample in spherical outside. $\alpha_i > 0$ some samples corresponding called support vector (SV), the use of this part of the sample can complete the accurate description of the hyper sphere. Therefore, we can use solving the distance between support vector and the center of the sphere a to determine the super-sphere radius R corresponding size.

In the relevance judgment of target samples for any test sample points z , z and the center of the only solution between the distance. In the desired distance not greater than the case of R, will be able to determine the z is the target sample, as follows:

$$K(z, a) - 2 \sum_i \alpha_i \cdot K(z, x_i) + \sum_{i,j} \alpha_i \cdot \alpha_j \cdot K(x_i \cdot x_j) \leq R^2 \quad (9)$$

C. Signature authentication process

In the training phase, the real signer group for each signer is registered with the completion of the training samples, and through appropriate method to solve the corresponding target data set, for the next step operation, according to equation (8) support vector machine data can obtain the actual signature description of the corresponding classifier. In this paper, the choice of RBF kernel function, $k(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / s)$ as kernel function of support vector machine classifier data description, s is the width parameter.

In the signature verification stage, the tested signature after extracting the features, should according to equation (9) to judge the authenticity of the target signature, this radius R is corresponding to the decision threshold, when the type(9) established that test signatures for the signer's genuine signature, otherwise a forged signature.

V. CLASSIFIER PARAMETER OPTIMIZATION AND FEATURE SELECTION BASED ON GA

The support vector machine data description of the related training, due to the addition of the kernel function and the relaxation factor concept, although R is obtained from the training, but the basic parameters of C and s on R values of the size and shape of the interface have a serious interference, it must pass through the expected error rate of a suitable set of related parameters, which can get the support vector machine classification data has strong ability to describe.

With the parameters of C and S decreasing, the number of the support vectors, sample points in the sphere of interface and the number of sphere size decrease [13]. If the classification boundary is too dense, which can

significantly reduce the possibility of miscarriage of justice to a certain extent, but there will be the possibility of a real signature excluded phenomenon will greatly increase.

Therefore, it will seriously interfere in the promotion effect to the test set if blindly raise the ideal performance of the training samples. To solve and improve this problem, it can be carried out by increasing the number of training samples to ensure that the sample distribution, better; it can also be solved by determining the classifier parameters suitable way. At the same time, feature subset selection, also affects the selection and classifier performance parameters of the authentication system, for different feature subset by the representative of the sample, the optimal parameters of them can be inconsistent, but there are inherent logic relation. Therefore, the correct way is to use the joint optimization methods for the corresponding treatment.

Method of searching for the optimal solutions formed by imitating the process of natural evolution, genetic algorithm (GA) has strong robustness and global search ability of [14] in the optimization problem. In order to give full play to the unique advantages of genetic algorithm in encoding operation, this paper selected in the individual, in accordance with the appropriate standards to the classifier parameters and feature subset combination form of the proper characterization, also need to meet the maximum two optimization criteria are the authentication function classifiers have, so this paper using genetic algorithm to realize the joint optimization of feature selection and classifier parameter.

Here directly to the authentication of the signature of the training set equal error rate (EER) was used to construct the fitness function, chromosome coding and structural problems using a binary encoding. Chromosome each corresponding feature set and a characteristic index, in the coding sequence of chromo some numbers 1 and 0 respectively[15], indicating that a corresponding features are selected and non selected two kinds of treatment results. In general, the support vector machine data selected parameters describing the fluctuation is small, little change in the corresponding classifier function, which shows in the concrete operation process, precise requirements of corresponding parameters is not high. In order to fully reflect the unique performance of the algorithm with the relevant parameters involved in this paper, according to four bytes in the form of representation, the corresponding parameter values can be divided into 16 levels of operation, and take $0.25 \leq c \leq 1, 1 \leq s \leq 16$. This, combined with the chromosome consists of 27+4+4, as shown in figure 2.

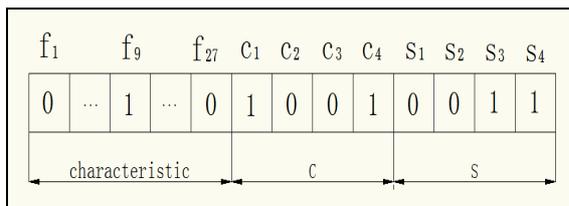


Figure 2. Combined with chromosome encoding

In the selection step of genetic algorithm, the proportional selection and optimal preservation strategy roulette combined selection operation, this can not only make the convergence performance of the algorithm is

ideal, but also can avoid the local optimal individual can be eliminated to a great extent, so as a result of the rapid diffusion and seriously affect the algorithm for global optimization to reduce the occurrence of the phenomenon. In the implementation of the optimal preservation strategy in the process, can use setting breeding generation gap manner, which means that the number of individual choice than the actual number of smaller individuals, but the number of the generated offspring are the same; and then you need to replace the resulting offspring individuals within the parent need to replace individual elements.

VI. SIGNATURE VERIFICATION EXPERIMENT

Using authentication methods above-mentioned, this paper presents a signature verification experiments on the SVC2013 database. According to the verification method of SVC2013, each classifier for 10 experiments, each had 10 real signature samples of 5 randomly selected as registered samples, classifier training. Then, certification for 20 random signature after 10 genuine signatures, 20 skilled forgery signatures, and randomly selected from the other true signatures in the signature [16].

When judging the performance of signature verification system to measure, the general choice of false acceptance rate (FAR) and false rejection rate (FRR) of two indicators comprehensive evaluation. In addition, because they will be with the threshold fluctuation has the numerical changes, so it is usually based on the equal error rate (EER) of appropriate quantization on the threshold, in order to obtain more objective evaluation of results.

In the use of support vector machine data description classifier for signature verification, to hyper sphere radius R trained as a discrimination threshold, can only be fixed FAR and FRR, can according to the needs of an application, changing the threshold to get different FAR and FRR. Figure 3 is a subset of features was used by all features and genetic algorithm optimization (9 features) ROC curve for authentication, where FAR and FRR are in average 40 signers. We can see that after optimization, signature verification has better performance.

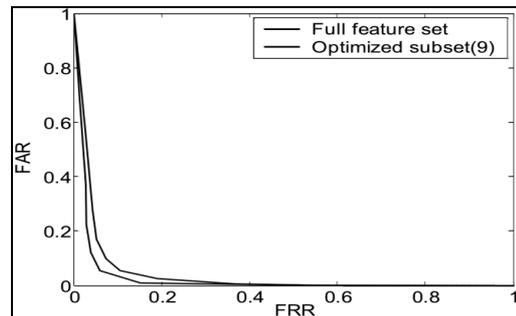


Figure 3. Curve of ROC with different feature sets the authentication

Table I gives the test result certified experimental results using support vector machine data is proposed in this paper to describe methods and SVC2013 announced support vector machine method, authentication data description for best results the skilled forgery average EER signature should be less than the corresponding experiment based on SVC2013, but the maximum EER

which is high, the experimental results are more close to the random forgery signature.

TABLE I.
DIFFERENT AUTHENTICATION EXPERIMENTS EER

Authentication method	Authentication EER (%)		
	average value	Variance	maximum value
Skilled forgery signatures			
SVM data description + full feature set	6.42	10.2	60
SVM data description +GA	4.93	8.64	60
SVC2013 The optimal results	6.9	9.45	50
Random forgery signature			
SVM data description + full feature set	3.76	9.31	60
SVM data description +GA	3.29	5.64	47.5
SVC2013 The final results	3.02	3.65	15

VII. CONCLUSION

This paper proposes a method of on-line handwritten signature verification by using support vector machine (SVM) data description and genetic algorithm (GA) method. From the collection of the original dynamic data signature extracted signature features include 27 kinds of shape and dynamic feature set, the training of the classifier as the target data. To joint optimize the classifier parameters and feature subset selection by genetic algorithms. Show that in the SVC2013 signature database experiments; this algorithm can register signature samples in a few situations, to get a better effect of certification. Because this method is using global features parameter signature, although has the advantages of strong anti-interference capability, convenient calculation advantages, but the signature on the local details distinguish ability is weak. Further research will consider integrating SVM with the authentication method based on characteristic function, to achieve more reliable result.

REFERENCES

- [1] Yang Bo [J]. polymerization a source autonomous system authentication method Nyberg-Rueppel signature data of Chinese computer systems, 2014, 22(2): 107-131.
- [2] Van Bao L, Garcia Salicetti S, Dorizzi B. On Csing the Viterbi Path Along with HMM Likelihood Information for Onling Signature Verification[J]. IEEE Trans on Systems, Man, and Cybernetics, 2012, 37(5):1237-1247. <http://dx.doi.org/10.1109/TSMCB.2007.895323>
- [3] Cai Hongbin, Shi Zesheng, Fan Xiaofeng, et al. A turning point based on wavelet transform to extract handwritten signature

authentication method [J]. Journal of image and graphics, 2014, 8 (3): 261-265.

- [4] M.M. Shafiei and H.R. Rabiee. A new online signature verification algorithm using variable length segmentation and hidden Markov models[J]. Document Analysis and Recognition, Edinburgh, Scotland, 2013, 12(6): 443-446.
- [5] C. Quek and R.W. Zhou. Antiforgery: a novel pseudo-outer product based fuzzy neural network driven signature verification system[J]. Pattern Recognition Letters, 2012, 28(14):1795-1816.
- [6] Li Yuan, Yuan Yuliang, Shen Feng, et al. A dynamic handwritten signature verification based on neural network model [J]. computer science, 2012, 32 (5):181-184.
- [7] D.Y. Yeung, H. Chang, Y Xiong, S. George, T. Matsumoto, and G. Rigoll[J]. SVC2013: First international signature verification competition[J]. Biometric Authentication, 2013, 16(9):16-22.
- [8] Luan Fangjun, Cheng Hai, Song Xiaoyu. Design and Realization of [J]. based computer applications and software, signature verification system HMM online handwriting based on 2012, 25 (6): 78-80.
- [9] M.Q. Pan, S.X. Qian, L. Lei, and X.J. Zhou. Support vector data description with model selection for condition monitoring[J]. Machine Learning and Cybernetics, 2013, 25(12):4315-4318.
- [10] Huang Dong, Liao Zhenquan. ECG PVC Classification Algorithm Based on Fusion SVM and Wavelet Transform 2015 8(1):193-202
- [11] J. Seo and H. Ko. Face detection using support vector domain description in color images[J]. Acoustics, Speech, and Signal Processing, 2014, 6(3): 729-732.
- [12] Li Bin, Zhang D, Wang Kuanquan. Improved Critical Point Correspondence for On-Line Signature Verification. International Journal of Information Technology, 2006, 12(7): 45-56
- [13] Huang Chengjie, Shen Fei, Wu Zhongcheng. Online signature authentication method of dynamic programming based on [J]. computer applications, 2009, 12 (7): 45-51.
- [14] Y.-J. Lee and S.-Y. Huang, Reduced support vector machines: A statistical theory, Neural Networks, IEEE Transactions on, vol.18, (2007), pp.1-13.
- [15] W. Honglei, O. Zongying and Z. Jianxin, Fingerprint image segmentation based on Support vector machines, Journal of system simulation, vol.19, no.5, (2007), pp. 2362-2365.
- [16] Guru D S, Prakash H N. Online Signature Verification and Recognition: An Approach Based on Symbolic Representation [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2012, 31(6):1059-1073.

AUTHORS

MR. Huang Dong (1972.12-), who is from Lingshan (Guangxi), is the member of CCF. He got his postgraduate in Guangxi Normal University, and the academic title of Associate Professor in 2013 (Senior Engineer in 2011). His research areas include Intelligent Recognition and Information System. He is with College of Mathematics and Computers Science, QinZhou University, QinZhou, China (qzxyhd@163.com).

MR. Gao Jian, (1981.03-), who is from Panjin (Liaoning). He got his postgraduate in Liaoning Technical University, and the academic title of Lecturer in 2009. His research areas include Cloud Computing, Data Stream and Information System. He is with Panjin Vocational and Technical College, Panjin, China (gao-jian-21@163.com).

This work was supported by Science and Technology Research in Universities of Guangxi, 2013(2013YB259). Submitted 08 October 2015. Published as resubmitted by the authors 30 October 2015.