ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 2, February 2014

A Performance Analysis of Different Classification Techniques in Offline Handwritten Signature Verification

Hemanta Saikia[1], Ashish Chandra Verma[2], Dipankar Saha[3], Kanak Chandra Sarma[4]

Associate Professor, Department of Electronics & Communication Engineering, Sikkim Manipal Institute of Technology, Sikkim, India[1]

Department of Electronics & Communication Engineering, Sikkim Manipal Institute of Technology, Sikkim, India[2, 3]

Professor, Department of Instrumentation & USIC, Gauhati University, Guwahati, Assam, India[4]

Abstract: Offline handwritten signature system works on the scanned image of a signature. Offline handwritten signature verification is a two class pattern recognition problem. For our experimentation purpose, we have developed offline signature datasets of with genuine and forged signature samples. Some commonly used geometric features were extracted from the signature datasets. Sequential Minimization Optimization algorithm with different kernels and Naive Bayes were used as the classification techniques. Performance analysis of different classification techniques is also discussed in terms of False Acceptance Rate (FAR) and False Rejection Rate (FRR).

Keywords: SMO algorithm, least square curve fitting, Naïve Bayes, RBF kernel, Polynomial kernel, Lagrange multiplier, Lagrange function

I. INTRODUCTION

Signature has been a distinguishing feature for person identification through ages. Signature is widely used as a token for authorization, which necessitates a technique for automatic verification. Approaches to signature verification fall into two categories according to the acquisition of the data: Online [1][2][3] and Offline[4][5][6] signature verification. In Online signature verification all the dynamic features such as motion of the stylus while the signature is produced, number of strokes, velocity, acceleration and pen pressure etc are recorded. These dynamic characteristics are specific for each person and are very stable and repetitive in nature. Offline data is a 2-D scanned image of the signature. In today's infrastructure, signatures of a person are taken generally on paper. Due to absence of stable dynamic features, processing of offline data becomes extremely necessary and challenging. Age, illness, geographic location and to some extent the physical and emotional state of the person can cause variation in a person's signature making the problem more prominent. All these factors result into large intra-personal variation. A system has to be designed which can detect various types of forgeries [7] considering all these factors. The system should be neither too sensitive nor too susceptible. It should have a satisfactory trade-off between a low False Acceptance Rate (FAR) and a low False Rejection Rate (FRR). In this paper first, we have briefly described the pre-processing needed to apply on the signature images. Following section is the feature extraction. We have extracted 11 geometric features from the pre-processed signature images. Then, the classification techniques are detailed namely Sequential Minimization Optimization (SMO) algorithm with polynomial kernel, SMO algorithm with RBF kernel and Naïve Bayes classifier. Finally, we have analyzed the performance of all these three classification techniques.

II. METHODOLOGY

Initially two sets of Genuine and Forged signatures are obtained. After data acquisition, the signatures are pre-processed [8][9]. Pre-processing is done to make images suitable for feature extraction. Relevant geometric features are obtained to distinguish signatures of different person from the pre-processed image. The geometric features obtained are used to train the system with different SVM classification techniques. Before classification, some parameters are optimized in classifiers to obtain better results. We have used different classification approaches such as SVM [10], Naive Bayes [11] technique for classification. In this paper a performance analysis of the results obtained are discussed in terms of FAR and FRR for different classifiers. A flow chart illustrating the various steps of signature verification is shown in Fig. 1.

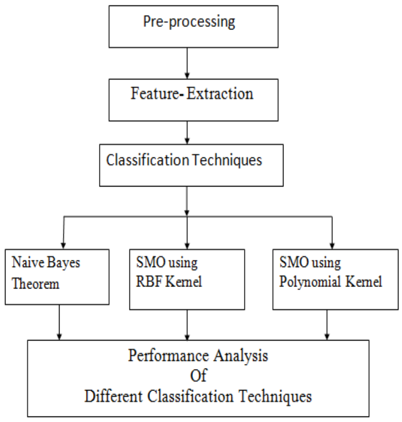

Fig. 1. Steps involved in Performance analysis of different classification techniques in Offline Handwritten Signature verification

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

III. PRE-PROCESSING

Signature preprocessing is a necessary step to reduce the computational needs in feature extraction and classification. Preprocessing also helps to increase the accuracy in classification. The pre-processing stage primarily involves some of the following steps [8][9][12]:

(1) Binarization: In this process, the input RGB image is converted into a binary image format.
(2) Complement Binarization: Binary image is complemented for computational simplification by changing the background into black (OFF pixel) and foreground of the image into white (ON pixel).
(3) Filtering: The noise that commonly occurs at the time of scanning is 'salt-pepper' noise. Morphological cleaning [13] is done where every isolated ON pixel is removed in 8-neighbourhood of OFF pixels. Although before performing filtering dilation of image is done such that in 8-neighbourhood of ON pixels if there is any OFF pixel it is replaced by ON pixel. Here the morphology of the image remains preserved.
(4) Skeletonization: Iterative transformations are applied on binary image to obtain a thinned image which is of one pixel thickness keeping its morphology same. This process is termed as 'Skeletonization' as the output image is a skeleton of the original image.

IV. FEATURE EXTRACTION

The accuracy of a pattern recognition system depends principally on the type of features extracted from the dataset. The main objective of feature extraction is to extract those features that will facilitate the classification algorithm to differentiate one class from the other accurately. An ideal feature extraction method uses minimal feature sets that are used to maximize inter-class variation between signature samples of different individuals while minimizing intra-class variation for those belonging to the same individual. Various extracted features are listed below [9][14][15][16]:

(1) Aspect Ratio: The aspect ratio (A) is the ratio of width to height of the signature. The bounding box coordinates of the signature are determined and the width (D_x) and height (D_y) are computed using these coordinates. Aspect ratio A is given by:

$$A = \frac{D_x}{D_y}$$

(3) Slope of the off-diagonal points of the bounding box: (2) If (x_1, y_1) and (x_2, y_2) are diagonally opposite points in the bounding box such that the diagonal so obtained is the off-diagonal of the bounding box of the signature image, then the slope m is given by

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$

(4) Maximum Pure Height: The height of each column of a signature sample after removing the vertical blank spaces is its pure height. From all the obtained pure heights, we choose the maximum value and it is known as the maximum pure height of the signature sample.

(5) Maximum Pure Width: The width of the image after removal of horizontal blank spaces is called pure width. Maximum value of the pure with of all the rows of an image is its Maximum pure width.
(6) Normalized area of the signature: Normalized area (NA) is the ratio of the area occupied by signature pixels to the area of the bounding box.

NA = $\Delta/(D_x \times D_y)$, where Δ is the number of ON pixel.

(7) Centre of Gravity: The Centre of Gravity of Signature pixel is the 2-tuple (X, Y) given by

$$X = \frac{1}{N}\left(\sum_{i=1}^{n} x_i\right) \ \& \ Y = \frac{1}{N}\left(\sum_{i=1}^{n} y_i\right)$$

Where x_i and y_i represents the column number and row number of ON pixels respectively.

(8) Slope of C.G. of Two Equal Halves of Signature Image: We divide the signature image within its bounding box into left and right halves and separately determine the centres of gravity of the two halves. It is seen that the slope of the line joining the two centres can serve as an effective feature to distinguish signatures.
Let C.G. of both halve be (x_1, y_1) & (x_2, y_2) respectively. Let slope be m.
Then, $m = \frac{y_2 - y_1}{x_2 - x_1}$
(9) Baseline-Shift: It is the difference between the vertical co-ordinates of centre of gravity of the two halves of signature normalized w.r.t vertical co-ordinates of centre of gravity of the original signature. This indicates the overall orientation of the signature.
If y_1 & y_2 are vertical co-ordinates of C.G. of two halves of the image and if y_0 is vertical co-ordinate of undivided image, then,

$$\Delta_1 = \frac{y_1}{y_0} \quad \& \quad \Delta_2 = \frac{y_2}{y_0}$$

Baseline Shift = $\Delta_2 - \Delta_1$

(10) Number of Cross-Points: Cross-point is a point in image where number of ON pixels in its 8-neighbourhood pixels is three or more.
(11) Slope of Optimal Line Obtained from Least Square Curve Fitting of Centre of Gravity of Each Column: Here least square curve fitting technique [17][18] is used to obtain the linear polynomial curve.
If we denote data values as (x, y) and points on the fitted line as (x, f(x)) then sum of the error at the four data points (x_1, y_1) (x_2, y_2) (x_3, y_3) and (x_4, y_4) is given by

$$\text{err} = \sum (d_i)^2 = (y_1 - f(x_1))^2 + (y_2 - f(x_2))^2 + (y_3 - f(x_3))^2 + (y_4 - f(x_4))^2$$

Our fit is a straight line so we substitute

$$f(x) = ax + b$$

$$\min_{a,b} \text{err} = \sum_{i=1}^{n=data\ point} (y_i - f(x_i))^2$$

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 2, February 2014

$$\Rightarrow \min_{a,b} \text{err} = \sum_{i=1}^{n=data\ points} \left(y_1 - \left(ax_i + b\right)\right)^2$$

(11) Normalized Actual Signature Height: It is the actual height of the signature image after width normalization. The height of the column is obtained by calculating the distance between the first ON pixel and the last ON pixel of a particular column.

V. VERIFICATION TECHNIQUES

In this paper the performance of three different classification techniques are compared. These three classification techniques are:

(1) SMO Algorithm using Polynomial Kernel
(2) SMO Algorithm using RBF Kernel
(3) Naive Bayes classification technique

Offline Handwritten Signature is a two class problem- Genuine and Forged class. Here we use SMO Algorithm [10][19][20] for SVM classification. Here two classes are separated using an optimal hyper-plane such that there is the largest margin between the bounding planes. There are two bounding planes. Bounding planes are those planes which pass through some samples on both classes

If training set $= \left\{(x_i, d_k) \mid x_i \in \mathbb{R}^p, d_k \in \{-1,1\}\right\}_{i=1}^{n}$

Where x_i is input feature of ith sample and d_k is the corresponding output for ith sample.

Separating Hyperplane is given by

$$w_1 x_1 + w_2 x_2 + \ldots - b = 0$$

$$\Rightarrow \mathrm{w}^T x - b = 0$$

where, w is co-efficient vector of hyperplane & b is bias term.

Here objective function to minimise is

$$\min_{w,\varepsilon,C} C\sum_{i=1}^{n} \xi_i + \frac{1}{2}\mathrm{w}^{\mathrm{T}}\mathrm{w}$$

$$\text{subject to } d_i(\mathrm{w}^T x_i - b) + \xi_i - 1 \geq 0 \ \ \& \ \ \xi_i \geq 0$$

C is a scalar value that controls the weightage and ξ_i is a non-negative error associated with ith sample.

If $x_i \to \phi(x_i)$ is a mapping of input vector into higher dimensional kernel where data can be separated linearly easily then

$$\min_u L_D(u) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_i d_j K(x_i, x_j) u_i u_j - \sum_{i=1}^{n} u_i$$

$$\text{subject to } \sum_{i=1}^{n} d_i u_i = 0,\ 0 \leq u_i \leq C$$

where $K(x_i, x_j) = \phi(x_i^T)\phi(x_j)$ is a kernel function. &

$$\mathrm{w} = \sum_{i=1}^{n} u_i d_i x_i$$

To calculate u_i SMO algorithm is implemented.

$$u_i(d_i(\mathrm{w}^T x_i - b) + \xi_i - 1) = 0$$

For $u_i > 0$, $d_i(\mathrm{w}^T x_i - b) + \xi_i - 1 = 0$ as here only b is only unknown term, it can easily be computed by substitution.

Kernels and classifiers used during experimentation:

(1) SMO Algorithm using Polynomial Kernel:
For Balanced Training Set (when number of genuine and forged training signature samples are equal)

$$K(x_i, x_j) = (x_i^T x_j)^d$$

where d is degree of polynomial
For unbalance training set

$$K(x_i, x_j) = (1 + x_i^T x_j)^d$$

(2) SMO Algorithm using RBF Kernel:

$$K(x_i, x_j) = \exp\left(\gamma \| x_i - x_j \|^2\right)$$

$$\gamma = -\frac{1}{2\sigma^2}$$

Where σ is a free parameter, when $\sigma = 1$

$$K(x_i, x_j) = \exp\left(-\| x_i - x_j \|^2\right)$$

Since it is exponential function, it can be expanded to infinite number of terms. Hence the dimension of RBF kernel for $\sigma = 1$ is infinite and the space in which it maps feature vector is called Hilbert Space having dimension infinite.

(3) Naive Bayes classification: In Naive Bayes classification technique instead of determining the distributions of distances between two feature vectors, each pair of corresponding bits in the test feature and training feature vectors are considered to be random variables. It is also assumed that the pairs with respect to different positions in feature vector to be independent and distributed identically. Let feature vectors be

$$F = \{f_1, f_2, f_3, \ldots . f_n\}, \text{ and } G = \{g_1, g_2, g_3, \ldots . g_n\}$$

The probabilities of ith same-value bits in a genuine-genuine pair and a genuine-forgery pair are computed using:

$$\mathrm{P}_{\mathrm{S,f}_i=\mathrm{g}_i} = \frac{|(F,G)|_{f_i=\mathrm{g}_i}, F, G \in T_G|}{|(F,G) \mid F, G \in T_G|}$$

$$\mathrm{P}_{\mathrm{S,f}_i\neq\mathrm{g}_i} = \frac{|(F,G)|_{f_i\neq\mathrm{g}_i}, F, G \in T_G|}{|(F,G) \mid F, G \in T_G|}$$

$$\mathrm{P}_{\mathrm{D,f}_i=\mathrm{g}_i} = \frac{|(F,G)|_{f_i=\mathrm{g}_i}, F \in T_G, G \in T_F|}{|(F,G) \mid F \in T_G, G \in T_F|}$$

$$\mathrm{P}_{\mathrm{D,f}_i\neq\mathrm{g}_i} = \frac{|(F,G)|_{f_i\neq\mathrm{g}_i}, F \in T_G, G \in T_F|}{|(F,G) \mid F \in T_G, G \in T_F|}$$

Where T_G & T_F are the training sets of genuine and forged signatures. If we know the values of the bit pair for each feature, given (K,Q) the overall genuine-genuine and genuine-forgery probabilities are computed as the product of the probabilities for all feature pairs, i.e.,

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 2, February 2014

$$P(genuine \mid Q) = \prod_{j=1}^{n} P_S(K,Q)$$

Where n is the total number of genuine signatures.
Similarly for forged signature

$$P(forged \mid Q) = \prod_{j=1}^{n} P_D(K,Q)$$

VI. OPTIMIZATION OF PARAMETERS

To obtain better accuracy features are to be normalized. Normalization of feature is given by

$$\frac{f_a - f_{\min}}{f_{\max} - f_{\min}}$$

Where,

$f_a \rightarrow$ actual value of particular feature

$f_{\min} \rightarrow$ minimum value of a particular feature

$f_{\max} \rightarrow$ maximum value of a particular feature.

After normalization of SVM parameters - Soft Margin cost function C, degree of polynomial kernel d and γ for radial basis function kernel are optimized using 5 fold cross-validation. Here the Best-First Search Engine Algorithm[21] is applied to obtain Best-first set of values for C, d & γ within the respective range of parameters for a given signature sample set to obtain best-first value of C, d & γ. A signature sample set consists of both genuine signature set of a person and forgery signature set of same signature one by different individuals. In our experiment, we have taken four different signature set.

TABLE I
DIFFERENT VALUES OF PARAMETER PUT INTO TEST TO OBTAIN OPTIMIZED VALUE OF PARAMETER

Parameters	Lower Limit	Upper Limit	Number of equal steps
C	0.5	3	6
d	1	3	3
γ	0.02	0.1	5

TABLE III
OPTIMIZED VALUE OF PARAMETER FOR DIFFERENT SIGNATURE SAMPLES USING POLYNOMIAL KERNEL

Signature Sample Set	Optimized Parameters	
	C	d
Sample Set-1	0.5	1
Sample Set-2	1	1
Sample Set-3	0.5	1
Sample Set-4	1	2 (when only 4 features used)

TABLE IIIII
OPTIMIZED VALUE OF PARAMETER FOR DIFFERENT SIGNATURE SAMPLES USING RADIAL BASIS FUNCTION

Signature Sample Set	Optimized Parameters	
	C	γ
Sample Set-1	1	0.1
Sample Set-2	1	0.06
Sample Set-3	0.5	0.02
Sample Set-4	1	0.1

Here in all case for SVM non-negative error, $\xi = 1\times10^{-12}$

VII. PERFORMANCE ANALYSIS

TABLE IV
FAR AND FRR OBTAINED WITH THE DIFFERENT CLASSIFIERS

Classification Technique	FAR	FRR
SMO Algorithm using polynomial kernel	0 - 3.7038	0 - 2.2985
SMO Algorithm using RBF kernel	4.46 - 25.92	0
Naive Bayes	1.78 - 11.116	1.78 - 4

We have observed that the highest accuracy obtained in our experimental environment is in case of SMO algorithm using polynomial kernel. We have also found Naive Bayes technique may be efficient when lesser number of training samples is available. Even though SMO algorithm uses complex mathematical concept, the accuracy is higher as compared to Naïve Bayes classification provided more number of training samples is available. SMO algorithm is most suitable for highly nonlinear dataset. It is also observed that if number of features is more, the best suited algorithm is SMO algorithm using linear kernel (i.e. d=1).

VIII. CONCLUSION

In this paper we have discussed feature based offline signature verification and compared the performance of different classification techniques under optimized condition. We have demonstrated that geometric features can successfully be used with different classification systems, to distinguish original signatures from forgeries.
The major setback associated with signature authentication is the availability of limited data. The other challenging problem in signature verification is that it is difficult to develop one general system to classify every style of signatures because choice of features depends on the style of the signatures and hence different styled signatures will have different characteristics. Due to unavailability of skilled forgers during experimentation it becomes challenging to decide whether the system is efficient and effective for different type of forgeries. The use of One Class SVM algorithm may reduce the complexity of the system and improve accuracy. Inclusion of some local features can improve the accuracy.

REFERENCES

[1] Sascha Schimke, Claus Vielhauer, Jana Dittmann, "Using Adapted Levenshtein Distance For On- Line Signature Authentication", 17th

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 2, February 2014

International Conference On Pattern Recognition (ICPR'04), Vol. 2, 2004, pp.931-934.
[2] A. T. Ihler, J. W. Fisher, and A. S. Willsky, "Nonparametric estimators for online signature authentication," International Conference on Acoustics, Speech, and Signal Processing, May 07-11, 2001, Salt Lake City, UT , USA, Vol. 6, pp. 3473-3476.
[3] S. Rajasekaran and G. A. Vijayalakshmi Pai, "Neural Networks, Fuzzy Logic, and Genetic Algorithms", Prentice-Hall India Pvt., Sixth printing, May 2006.R. J. Vidmar. (1992, August). On the use of atmospheric plasmas as electromagnetic reflectors. IEEE Trans. Plasma Sci.[Online]. 21(3). pp. 876—880.
[4] D. Impedovo and G. Pirlo, "Automatic signature verification: The state of the art," IEEE Trans. Syst.Man Cybern.—Part C: Applications Rev., vol. 38, no. 5, pp. 609–635, Sep. 2008.
[5] W. Hou, X. Ye, and K. Wang, "A survey of off-line signature verification," in Proc. Int. Conf. Intelligent Mechatronics Automation, Chengdu, China, Aug. 2004, pp. 536–541.
[6] J. Fierrez-Aguilar, N. Alonso-Hermira, G.Moreno-Marquez, and J. Ortega-Garcia, "An off-line signature verification system based on fusion of local and global information," in Proc. Workshop Biometric Authentication, Springer LNCS-3087, May 2004, pp. 298–306.
[7] J. Coetzer, B. Herbst & J. du Preez, "Off-line signature verification using the discrete radon transform and a hidden Markov model," EURASIP Journal on Applied Signal Processing, 2004.
[8] H. Saikia and K. C Sarma, "Approaches and issues in offline signature verification system," International Journal of Computer Applications Vol-42-16 March 2012.
[9] A. C. Verma, D. Saha, H. Saikia, "Forgery Detection in Offline Handwritten Signature Verification using Global and Geometric features," IJCER vol.2 no.2 , April 2013.
[10] C. J. C. Burges, Data Min. Knowl. Discov. "A Tutorial on Support Vector Machines for Pattern Recognition," vol. 2, no. 2, 1998.
[11] S. Srihari, K. Kuzhinjedathu, H. Srinivasan, C. Huang, and D. Pu, "Signature verification using a Bayesian approach," IWCF '08: Proc of the 2nd Int Workshop Computational Forensics, pp. 192–203, 2008.
[12] D. Samuel and I. Samuel, "Novel Feature Extraction Technique for Offline Signature Verification System," International Journal of Engineering Science and Technology, Vol. 2(7), pp. 3137-3143, 2010.
[13] Van den Boomgard, R, and R. van Balen, "Methods for Fast Morphological Image Transforms Using Bitmapped Images," Computer Vision, Graphics, and Image Processing: Graphical Models and Image Processing, Vol. 54, Number 3, pp. 254-258, May 1992.
[14] V. A. Bharadi and H. B. Kekre, "Off-Line Signature Recognition Systems," International Journal of Computer Applications Volume 1, No. 27, 2010.
[15] S. Biswas, T. Kim and D. Bhattacharya "Features Extraction and Verification of Signature Image using Clustering Technique," International Journal of Smart Home, Vol.4, No.3, July, 2010.
[16] H. Baltzakis and N. Papamarkos, "A new signature verification technique based on a two staged neural network classifier," Engineering Applications of Artificial Intelligence 14 , 2001.
[17] G. Recktenwald, "Least Squares Fitting of Data to a Curve," Portland State University, Department of Mechanical Engineering.
[18] C. Guang, R. Zhi-liang, S. Hai-zhu, "Curve Fitting in Least-Square Method and Its Realization with Matlab," Dept. of Weapon Engineering, Naval University of Engineering, Wuhan 430033, China, Ordnance Industry Automation, 2005-03.
[19] John Platt, "Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines," Microsoft Research Technical Report MSR-TR-98-14, (1998).
[20] S. S. Keerthi, S. K. Shevade, C. Bhattacharyya and K. R. K. Murthy, "Improvement to Platt's SMO Algorithm for SVM Classifier Design," Neural Computation, Vol. 13, pp. 637-649, 2001.
[21] R.Kohavi,"Wrappers for performance enhancement and oblivious decision graph",1995, Chapter-4, pp.-76-128.

BIOGRAPHIES

Hemanta Saikia He is presently an Associate professor in the department of Electronics & Communication Engineering in Sikkim Manipal Institute of Technology, Sikkim, India. He has completed his Bachelor of Engineering from Jorhat Engineering College, Assam, India and Master of Technology from Tezpur University, Assam, India.

Ashish Chandra Verma He has completed his Bachelor of Technology in Electronics & Communication Engineering from Sikkim Manipal Institute of Technology, Sikkim, India in 2013. His fields of interest are pattern recognition and embedded systems.

Dipakar Saha He has completed his Bachelor of Technology in Electronics & Communication Engineering from Sikkim Manipal Institute of Technology, Sikkim, India in 2013. His field of interest is pattern recognition.

Kanak Chandra Sarma He received his Masters (M. Sc.) and PhD from Gauahati University, Assam, India. Presently, he is the Professor and Head of the Department of Instrumentation and USIC, Gauhati University. His fields of research interest are - Thin Film and Nano-technology, Optoelectronics and Electronic System Design.