BULGARIAN ACADEMY OF SCIENCES

CYBERNETICS AND INFORMATION TECHNOLOGIES • Volume **14**, No 2

Sofia • 2014

Print ISSN: 1311-9702; Online ISSN: 1314-4081
DOI: 10.2478/cait-2014-0022

A Combined Method for On-Line Signature Verification

Desislava Boyadzhieva, Georgi Gluhchev

Institute of Information and Communication Technologies, 1113 Sofia
Emails: d.n.dimitrova@gmail.com gluhchev@ iinf.bas.bg

Abstract: *A combined method for on-line signature verification is presented in this paper. Moreover, all the necessary steps in developing a signature recognition system are described: signature data pre-processing, feature extraction and selection, verification and system evaluation. NNs are used for verification. The influence of the signature forgery type (random and skilled) over the verification results is investigated as well. The experiments are carried out on SUsig database which consists of genuine and forgery signatures of 89 users. The average accuracy is 98.46%.*

Keywords: *On-line signature verification, neural networks, feature selection, SUsig database, forgery signatures.*

1. Introduction

Signature verification is the process of confirming the identity, based on the handwritten signature of the user as a form of behavioral biometrics [1]. From one hand, the signatures are a convenient, widely used and secure mean for authentication, and from the other, their input to biometric systems is fast, easy, natural and non-invasive. For these reasons, the problem of signature verification is broadly investigated in the past years. Novel methods and algorithms are developed, mostly for on-line signatures, and lots of them are implemented in practice [3, 8, 9].

Signature recognition systems are on-line and off-line depending on the signature acquisition method. The off-line method uses a captured image of a written signature after the writing process is over, while the on-line method uses

devices such as graphical tablets to capture signature during signing and thus a lot of writer's specific features like pressure, speed, pen tilt, azimuth, etc. are available.

In this paper we propose a combined method for signature verification and test it on a SUsig database [2]. For each user of the database we construct a NN model for verification. With regard to this we perform the following steps. At the pre-processing stage, some transformations are performed on the signatures (coordination transformation, rotation and translation). Next, we extract the signature features and perform feature set selection by applying the method for selection of regression variables based on *Mallows* C_p criterion [5, 6] to identify the best features subset. We experiment with random and skilled forgeries. After this step, NNs are constructed of varying size of the hidden neurons and a 10-fold cross-validation is performed in order to choose the best one of them. Such selected model is defined by the following parameters: number of hidden neurons, type of forgery signatures for training and input features. Finally, we train, validate and test all the chosen user's models and obtain the average system accuracy.

This paper is organized as follows. In Section 2 a brief overview over the feature set selection is given. Next, the application of NNs for signature verification is considered in Section 3. The experimental results are presented in Section 4.

2. Feature set selection

2.1. Signature pre-processing

At first, the signature raw data is acquired by the hardware device (e.g., a graphical tablet, PDA, etc.). To facilitate the feature extraction, it is necessary these raw data to be pre-processed. The operations applied depend on the selected features and the acquisition protocol.

The coordinates x and y of the ink coordinate space are called himetric units [10] and their values fall within $[0, 7999]\times[0, 5999]$. It is necessary to transform them in the application coordinate system in $[0, 1279]\times[0, 799]$. This is performed automatically by a method from Microsoft Tablet PC SDK. Since the acquired signatures may be rotated, we have to align them horizontally. The next pre-processing operation is translation of the signatures to a given point of the application coordinate system because it is possible some of the coordinates to obtain negative values after the rotation.

2.2. Feature extraction

There are three groups of signature features: global, local and segmental [12]. The global features are extracted for the whole signature, the local features are extracted for each sample point in the signature, and the segmental features are extracted for each signature segment. Over 100 features used in the signature verification are listed in [14].

The extracted global signature features used are presented in Table 1.

Table 1. Global features

Abbre-viation	Feature name	Abbre-viation	Feature name
A1	Signature length *L*	A13	Angle of the line between the initial and end points
A2	Signature height *H*	A14	Distance between the leftmost and center points
A3	Height to width ratio *H/L*	A15	Distance between the center and rightmost points
A4	Number of points *N*	A16	Angle of the line between the center and leftmost points
A5	Time duration	A17	Angle of the line between the center and rightmost points
A6	Number of segments	A18	Distance between the leftmost and initial points
A7	Signature density A4/A1*A2	A19	Distance between the rightmost and end points
A8	Distance between the initial and center point	A20	Angle of the line between the leftmost and initial points
A9	Distance between the end and center point	A21	Angle of the line between the end and rightmost points
A10	Distance between the initial and end point	A22	Number of strokes
A11	Angle of the line between the center and initial points	A23	Average tilt
A12	Angle of the line between the center and end points	A24	Average pressure

2.3. Feature set selection

Since some features demonstrate higher discriminatory capability than others, feature selection should be performed. This is related to the process of selecting k features of most discrimination power out of p available ones ($k \leq p$) and it aims to identify and remove as much irrelevant and redundant information, as possible. A review of the processes of feature set selection for signatures is done in [12].

We approach the feature set selection step in signature verification by applying the methods of Hocking, Leslie and LaMotte for selection of regression variables based on *Mallows* C_p criterion for regression [6, 7]. This criterion is used to decide on a suitable subset among the contending subsets. It is a measure of the standardized total squared error defined as follows:

$$C_p = \frac{\mathrm{RSS}_p}{\hat{\sigma}^2} - (n - 2p). \tag{1}$$

In (1) RSS_p denotes the residual sum of squares for the particular regression with p variables and $\hat{\sigma}^2$ is an estimate of the residual mean square σ^2 for full regression.

By applying the methods of Hocking, Leslie and LaMotte we identify the best feature subsets of various size for each user on the basis of his/her eight or ten genuine signatures and ten random forgeries. Among these subsets we select the best subset that has a C_p value closest to p, where p is the number of regression coefficients. Thus, for each user we obtain the best feature subset of different size. An extensive overview of this method may be found in [4].

3. Neural networks for signature verification

Neural networks are suitable to be used for signature verification since they are an excellent generalization tool (under normal conditions) and are a useful mean of coping with the diversity and variations inherent in handwritten signatures [13]. Usually, a particular NN is built for each user on the basis of his/her genuine and forgery signatures. The number of the input neurons is p where p is the number of the features. The single output neuron has a value 1 for the genuine signature and a value 0 for the forgery signature. After training, a score threshold is determined. If the verification result (at the time of testing a signature) is greater than the corresponding score threshold, the signature is considered genuine, otherwise – forgery. This approach is widespread because it allows fast adding and deleting of signatures for new and existing users [13]. Usually, NN training takes much time but in this approach it is done off-line, so that the users are not forced to wait.

4. Experiments

The experiments are carried out in MATLAB environment. We use *Neural Network Toolbox*.

4.1. Signature database SUsig

The signature database used in the experiments is SUsig [2]. It consists of two sub corpuses: *Visual* and *Blind*. We test the proposed combined method on signatures of *Blind* subcorpus. There are 89 users and for each of them 8 or 10 genuine and 10 skilled forgeries are acquired by using a graphical tablet *Wacom Graphire2*. The genuine signatures are collected in a single session. The signature data consists of the following information for each of the signature points: x and y coordinates, timestamp, pressure level and a pen up or down indicator.

4.2. Results

We experiment with:

(1) A *common* feature set for all users

At first, we identify the significant correlation coefficients pairs at 0.01 confidence level, 99% confidence interval for all the users and then build the corresponding histogram. After that, we find the feature pairs which are met in more than 25% of the users and remove one feature out of a pair by applying the method of correlation pleiads [7]. In this way, the feature number is reduced by around 50% and the remaining features are A1, A2, A4, A6, A10, A12, A13, A16, A17, A21, A22, A23, A24.

(2) An *individual* feature subset

At first, we identify the significant correlation coefficients at 0.01 confidence level for each user. Then we remove one feature out of a pair by applying the method of correlation pleiads [7]. Next, we apply the method of Hocking, Leslie and LaMotte [5] to the remaining features. Let us denote by Variant 1 the case, in

which the feature subset is determined by using the genuine and random forgery signatures and denote by Variant 2 the case, in which the feature subset is determined by using the genuine and skilled forgery signatures. The size of the obtained *p*-subset and the corresponding number of users are specified in Table 2.

Table 2. Individual *p*-subsets

Variant 1		Variant 2	
Size of *p*-subset	Number of users	Size of *p*-subset	Number of users
9	42	9	48
8	5	8	9
7	7	7	4
6	12	6	10
5	14	5	15
4	5	4	3
3	4	3	0

There is a significant feature reduction for both Variant 1 and Variant 2, since their initial number (13) is reduced down to 9 for about half of the users, reduced down to 5 or 6 features for 30% of the users.

In Table 3 all the six NN models are described together with their parameters. Let us denote by Case 1 the case, in which only random forgeries are used for NN training, and denote by Case 2 the case, in which both random and skilled forgeries are used.

Table 3. Parameters of the NN models

# of model	Features (input neurons)	Genuine signatures	Forgery signatures		Number of hidden neurons
1	Common set	8 or 10	15 random	Case 1	From 1 up to 5
2	Variant 1				
3	Variant 2				
4	Variant 2		9 random and 6 skilled	Case 2	
5	Variant 1				
6	Common set				

All 30 models are evaluated by a 10-fold cross validation for each user and the best performed optimal NN model is selected together with its parameters: number of hidden neurons, type of signature forgeries for training and input features.

In Table 4 the number of users for all the chosen models is given. Model # 2 (30% of the users) is the most common, followed by model # 4 (18% of the users). The number of the models trained on individual features (Variant 1) is 53 (60% of the users).

Table 4. Number of model occurrences

User	Case	Model #	Number of users
Common features	Case 1	1	13
Variant 1	Case 1	2	27
Variant 2	Case 1	3	13
Common features	Case 2	4	16
Variant 1	Case 2	5	9
Variant 2	Case 2	6	11

4.3. Results interpretation

In the current scenario: (1) it is demonstrated that each user has its own discriminative feature subset. In other words, we cannot restrict to a common feature set valid for all users, but instead of that we have to consider each user best feature subset separately; (2) The initial feature set size is reduced to a higher extent if random forgeries (Var. 1) are used for building the regression model for Hocking, Leslie and LaMotte method instead of skilled forgeries; (3) It cannot be concluded that the models built only on random forgery signatures perform better than those built on both random and skilled forgeries; (4) The obtained results are satisfactory, demonstrating verification accuracy of 98.46%, EER 1.61%, FRR 0%, FAR 2.70%. These results are similar to those obtained on the same SUsig database by applying a linear classifier [11].

References

1. Nalwa, V. S., I. Ekeland. Automatic On-Line Signature Verification. – In: Proceedings of the IEEE'85, 1997, 213-239.
2. Kholmatov, A., B. Yanikoglu. SUSIG: An On-Line Signature Database, Associated Protocols and Benchmark Results. – Pattern Analysis & Applications, Vol. **12**, 2009, 227-236.
3. Jain, A., Li Stan. Encyclopedia of Biometrics. Springer, 2009.
4. Boyadzhieva, D., G. Gluhchev. Feature Set Selection for On-Line Signatures Using Selection of Regression Variables. – In: Proceedings of 4th International Conference on Pattern Recognition and Machine Intelligence PReMI'11, 2011, 440-445.
5. Hocking, R. R., R. Leslie. Selection of the Best Subset in Regression Analysis. – Technometrics, Vol. **9**, 1967, 531-540.
6. LaMotte, L. R., R. R. Hocking. Computational Efficiency in the Selection of Regression Variables. – Technometrics, Vol. **12**, 1970, 83-93.
7. Ayvazyan, S. A., Z. I. Bejaeva, O. V. Staroverov. Classification of multidimensional observations. Moscow, Statistics, 1974, p. 240 (in Russian).
8. Gluhchev, G., M. Savov, O. Boumbarov, D. Vassileva. A New Approach to Signature Based Authentication. – In: 2nd Int. Conf. on Biometrics, Seoul, 26-29 August, 2007, 594-603.
9. Savov, M., G. Gluhchev. Signature Verification Via Hand-Pen Motion Investigation. – In: Proc. Int. Conf. "Recent Advances in Soft Computing", Canterbury, 2006, 490-495.
10. Ink Data. **http://msdn.microsoft.com/en-us/library/ms811395.aspx**
11. Kholmatov, A., B. Yanikoglu. Identity Authentication Using Improved On-Line Signature Verification Method. – Pattern Recognition Letters, Vol. **26**, 2005, No 15, 2400-2408.
12. Richiardi, J., H. Ketabdar, A. Drygajlo. Local and Global Feature Selection for On-Line Signature Verification. – In: Eighth International Conference on Document Analysis and Recognition (ICDAR'05), 2005, 625-629.
13. McCabe, A., J. Trevathan, W. Read. Neural Network-Based Handwritten Signature Verification. – Journal of Computers, Vol. **3**, 2008, No 8, 9-22.
14. Leclerc, F., R. Plamondon. Automatic Signature Verification: The State of the Art 1989-1993. – International Journal of Pattern Recognition and Artificial Intelligence, Vol. **8**, 1994, No 3, 643-660.