Support Vector Machines for Dynamic Biometric Handwriting Classification

Tobias Scheidat, Marcus Leich, Mark Alexander, and Claus Vielhauer

Abstract Biometric user authentication is a recent topic in the area of computer security. This paper presents a machine learning approach to single modality user authentication. Here support vector machines (SVM) are employed to classify dynamic handwriting samples. The general goal of SVMs is to carry out binary classifications and/or to handle multiple class problems using a combination of different SVMs. Here a multi-class SVM is proposed to execute verification as well as identification of persons based on their handwriting using a given PIN and a freely chosen PIN. In the best case (trade-off for all rates) for verification using the free PIN a false acceptance rate (FAR) of 0.0083 and an attacker acceptance rate (AAR) of 0.0241 are determined while the false rejection rate (FRR) yields zero. In identification mode using the free PIN, we observe a FRR of 0.0083 and an attacker identification rate (AIR) of 0.2195 at a false identification rate (FIR) level of zero in our experiments.

1 Introduction

The authentication of persons and information plays an important role in information technology. Mostly, user authentication is based on one or combinations of the three factors: secret knowledge, personal possession and/or biometrics. While knowledge and possession provides possibilities to hand over to unauthorized per-

T. Scheidat, M. Leich, M. Alexander, C. Vielhauer
University of Magdeburg, Department of Computer Science, Advanced Multimedia and Security Lab, Universitätsplatz 2, D-39106 Magdeburg, Germany
e-mail: tobias.scheidat@iti.cs.uni-magdeburg.de, {marcus.leich, mark.alexander}@st.ovgu.de, claus.vielhauer@iti.cs.uni-magdeburg.de

C. Vielhauer
Brandenburg University of Applied Science, Department of Informatics and Media, Magdeburger Str. 50, D-14770 Brandenburg, Germany e-mail: claus.vielhauer@fh-brandenburg.de

sons with or without intent or can be lost, biometrics is linked to physical or behavioral characteristics of a person. On the other side, biometric systems have to handle a certain fuzziness of the data of individual persons (intra class variability) and alikeness of data of different persons (inter class similarity). To overcome these drawbacks in a sufficient manner, a great variety of biometric authentication methods were proposed. Following, a small selection of such methods is given without neglecting other publications.

An overview on performance of machine learning techniques in biometrics has been published by Abreu and Fairhurst [1]. Here eight machine learning techniques have been used for classification of fingerprint and signature samples. Mean and standard deviation for absolute error counts are provided for performance evaluation. SVMs among neural net classifiers reached the lowest classification error rates.

A different approach is taken by Fuentes et al.[3]. Instead of using a SVM directly for classification it is used to fuse matching scores of two expert system for online signature verification. The idea behind this concept is to combine the unique strengths of both systems into a single one. The system operates on a test set which partly included skilled forgeries and reaches FAR values of 0.046 and FRR values of 0.083.

In this paper a multi-class SVM is suggested and evaluated for dynamic handwriting verification and identification. The evaluation shows very promising results based on a database of 30 writers with regard to the measures used, false acceptance rate (FAR), attacker acceptance rate (AAR) and false rejection rate (FRR) for verification and false identification rate (FIR), FRR and attacker identification rate (AIR) for identification. Because of the limited number of samples available for testing and because of the different nature of performance measures employed (error rates vs. mean of error counts) and samples semantic results are of limited comparability to those of Abreu and Fairhurst [1] and Fuentes et al.[3].

This paper is structured as follows: The next section describes fundamentals of support vector machines and the configuration of the suggested SVM for handwriting verification and identification. In section 3 the evaluation setup, methodology and results are presented and discussed, while the forth section concludes the contribution and provides an overview of future work.

2 Materials and Methods

This section provides an introduction to SVMs and their use for multi-class classification. Additionally details of the features extracted from the handwriting samples are provided. These features represent the components of the sample vectors presented to the SVM for training as well as for later classification. This section concludes with a discussion of possible problems that can occur during SVM training and classification and presents an approach to overcome these difficulties.

Support Vector Machine: In their basic form SVMs are limited to solving binary classification problems for linear separable classes. These limitations can be overcome by using sophisticated kernel functions that enable the SVM to solve more complex binary classification problems. Additionally several SVMs can be combined to solve multi-class problems.

Since in biometric systems the feature vectors for several persons are not expected to be linear separable, the well-known radial basis function kernel is chosen. Furthermore the *one-against-all* approach [4] for implemantation of multi-class is implemented in the following way: For each person p of the N persons that are to be enrolled we train one SVM using the enrolment samples of p as positive samples and all other other enrolment samples of the remaining $N-1$ persons as negative samples. Consequently, after the enrolment process the entire systems consists of N SVMs, one for each enrolled person.

Based on this system structure it is easily possible to devise an identification and verification procedure for new samples from users that try to authenticate on the system. In the verification scenario a user tries to be verified as an enrolled person. The sample from this user is simply presented to the matching SVM. The user is then accepted or rejected based on the SVM output. The identification process works similarly. Here the sample of the user is presented to all SVMs. If no SVM accepts the sample, the user is rejected. If only one SVM accepts the sample, the user is identified as the corresponding person. If more than one SVM accepts the sample, the result is ambiguous which again leads to the rejection of the user.

For the experiments described in this paper the LIBSVM[1] [2] as pre-existing implementation of the SVM classification algorithm is chosen.

Features: During the data acquisition a sequence of five physical values is sampled by the handwriting sensor time dependently. These values are the X and Y position, the pen tip pressure and the angles azimuth and altitude. Based on these values for each handwritten samples a set of 103 (first 69 features are described in [6], features 70-103 are based on current work) statistical features is determined, which represents the corresponding sample. These feature sets are used as input for the evaluation of the authentication performance of the SVM system.

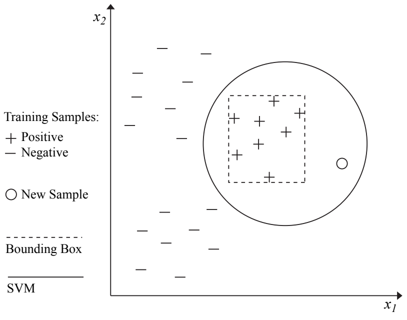

Fig. 1 Bounding box for samples of one person. The SVM classifies the new sample as positive while the bounding box rejects the unknown data.

[1] Version 2.87

These statistical features are partly based on dynamic (e.g. writing time, pen down time, min/max velocity in X or Y direction) but also on static (e.g. aspect ratio, intersections of the writing trace with itself or helper lines) characteristics of the sampled handwritten data. Some features or groups of them are identical or quite similar to 8 out of the 18 features used by Abreu and Fairhurst in [1].

Supporting Bounding Box: Even if the SVMs reach excellent classification rates for samples of trained persons the system might still by prone to false-classification of samples from not enrolled persons. This is due to the fact that during SVM training the feature space is not exhaustively covered with negative samples in regions where no positive samples exist. These regions can be assigned to any class by the SVM without affecting the reclassification rate of the training samples. Fig. 1 illustrates the problem.

To achieve a higher rejection rate for samples of unknown persons it was decided to apply a simple bounding box heuristic. A hyper-bounding box is computed for each person based on the training sample set. The box simply consists of the minimum and maximum values observed for each component of the training samples. A sample is rejected by the bounding box if its components are not within the corresponding ranges of the bounding box. An SVM with bounding box heuristic accepts a sample only if both the SVM and the bounding box accept the sample.

3 Evaluation

This section provides details about the data set and the training and test procedures the experiments where conducted with. Based on that, an analysis of the observed error rates is presented.

Data Set: During data acquisition test persons were ask to donate 8 samples each per semantic class. In case of dynamic handwriting, semantics are alternative written contents to the signature. In this work 30 persons and the semantics *given PIN* and *free PIN* were chosen. While the *given PIN* consists of the default sequence of five digits (77993), the *free PIN* can be freely chosen by the writer under the restriction to write exactly five digits, too.

The values of all samples have been scaled to $[0,1]$ separately for each semantic using the minimum and maximum values of each feature observed for all persons.

Methodology: For the experiments it was chosen to randomly split the samples of the 30 persons available into two equally sized groups. The first group consists of the 15 persons that are to be enrolled in the system (enrolment group). The second group consists of the remaining 15 persons (attacker group). This group is used to simulate blind attacks of not enrolled persons. The samples of all persons are stored in lists in which each sample has a fixed position. The position of the samples in these lists are randomised one time before all experiments take place.

For a single SVM parameter combination the following test procedure is undertaken: For each person a training and test sample set is determined by selecting two index values m and n with $m \neq n$ as index to the sample list of that person. The samples at the positions m and n are added to the test set of the person, samples at all other positions are added to the training set. m and n are equal for all persons.

After all training and test sets have been created one SVM is trained for each person using this person's training samples as positive samples and all other persons training samples as negative samples. The resulting set of SVMs now represents a multi-class classifier as described in 2.

Using the test samples of all persons the performance measures of this classifier can be determined. For verification each test sample is presented to the system 15 times, each time using a different person as claimed identity. One of these tests can produce at most one false reject (claimed person is the same as the actual sample origin and the sample gets falsely rejected) and at most 14 false accepts (claimed person a different from the sample origin, but the sample is accepted). For identification the sample is simply presented to the entire system. This single test produces at most one false identification (sample is identified as belonging to a different person than the sample origin) or at most one false reject (person is rejected, though the sample origin is a enrolled person).

Additionally all samples of the attacker group are presented to the system. For verification one sample can cause at most 15 false accepts (attacker accepts) while for identification one sample can produce at most one false accept (attacker accept/identification).

The entire procedure (starting from the selection of m and n) is repeated 28 times, so that all possible combinations for m and n are iterated. Over all these experiments all false accepts/identifications, false rejects and attacker accepts/identifications are summed up for verification and identification respectively. The corresponding error rates: FAR/FIR, FRR, and AAR/AIR are computed by dividing the summed errors by the number of maximum number of times this particular type of error can occur. Consequently all are normalised to $[0,1]$. Additionally for identification $FIR + FRR \leq 1$ holds true.

The error rates just described are computed for various SVM parameter combinations. In the chosen experiment setup, we vary the SVM cost parameter C ($[2,194]$, stepsize: 8) and the radial basis function kernel parameter γ ($[0.0002, 0.0194]$, stepsize: 0.0008). In previous tests these parameters proved to be most influential to the classification performance, being most sensitive within the tested intervals. In their effects on error rates these parameters behave similar to the threshold value in distance based classifiers. However, the behaviour of these parameters is not monotonic as it is often the case for threshold values. All remaining LIBSVM parameters were left at their default values.

Results: As presented in the preceding section each parameter combination $(C;\gamma)$ used for training/testing can be evaluated using the three measures FAR/FIR, FRR, AAR/AIR. These 3 dimensional tuples are from now on referred to as *operating points* (OP). The following rates have been determined without using the bounding

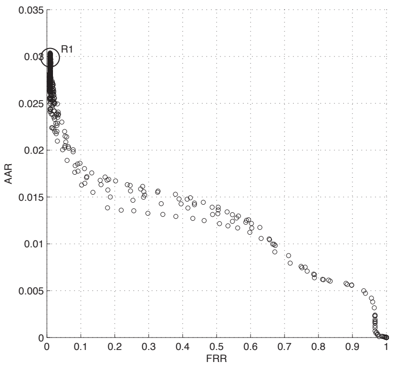

Fig. 2 FRR-AAR-projection of all observed operating points of the verification system (*free PIN* semantic). Region R1 contains all points with FAR = 0.000085. For all other operating points FAR is zero.

box heuristic: For verification most parameter combinations tested (74%) reach an FAR of exactly zero. All other cases reveal a very low FAR of 0.000085. Since the system does not reach equally low FRR values it is impossible to provide an equal error rate (EER) according to the traditional definition. However a modified EER based on AAR and FRR can be estimated if the OPs are projected onto the FRR-AAR plane as depicted in Fig. 2. In this representation the modified EER is about 0.025. As can also be seen the tested parameter combinations are capable of constructing quite effective SVMs (very low FRR) as well as extremely over-fitted SVMs (FRR=1 while still retaining an FAR of 0). The reason for this tendency to over-fitting is the massive class imbalance when one SVM is only trained with positive samples of one person and negative samples of all remaining person. This class imbalance also abets the division of the feature space into disjoint regions in which only samples of a specific persons are accepted.

This behaviour is also reflected in the AAR which converges to zero for high FRR values (over-fitting) and reaches values around 0.03 for very low FRR values. It should also be noted that the highest AAR values are reached for operating points with an FAR not equal to 0 (see region R1 in Fig. 2).

In identification mode the system behaves similar to the verification mode. Differences can be found in the FIR values, which are now consistently zero and in slightly higher AIR values. The reason for the lowered FIR is the stronger rejection criteria which rejects a sample if it is accepted by more than one SVM. Obviously the few false rejects observed during verification testing are the result of two SVMs accepting the same sample. Similarly, because of this stronger rejection criteria the AIR values for identification should be lower than AAR for verification. However the opposite is the case, AIR values are consistently larger for identification. The reason for this is the already mentioned almost disjoint separation of the feature space among SVMs. This separation is the reason that, for verification, a sample causes only one out of 14 possible false accepts, while for identification, the same sample causes one false accept out of one possible false accept which leads to a much higher influence to the AIR.

Table 1 depicts exemplary OP for identification and verification using *free PIN* and *given PIN* semantic. The two OPs chosen for each semantic and authentication

Table 1 Selected operating points

Semantic	Verification				Identification			
	$(C;\gamma)$	FAR	FRR	AAR	$(C;\gamma)$	FIR	FRR	AIR
free PIN	$(26,0.0154)^a$	0	.0095	.0241	$(114,0.0018)^a$	0	.0905	.0924
free PIN	$(34,0.0194)^b$	0	.0083	.0254	$(34,0.0098)^b$	0	.0083	.2195
given PIN	$(90,0.0194)^a$	.0016	.0560	.0309	$(58,0.0066)^a$	0	.1524	.235
given PIN	$(90,0.0194)^b$	.0016	.0560	.0309	$(154,0.0194)^b$	.0012	.0738	.2933

[a] parameters for OP with minimum distance to origin
[b] parameters for OP with minimum FRR (if several OPs with the lowest FRR exist, the one with the lowest AAR/AIR is displayed)

mode are the ones with the smallest Euclidean distance to the origin[2] and the ones with the smallest observed FRR.

For the *given PIN* the measured rates are generally higher then the rates for the *free PIN* semantic. Obviously the features described in 2 have better discriminatory properties if they are extracted from samples that differ in content and writing style instead of writing style alone.

It has to be noted that a variety of parameter combinations can lead to satisfying results depending on which error rate is preferred to be minimised. The values of these parameters are also dependant in the data set used for testing as can easily be seen in the varying coordinates in Table 1.

The bounding box approach introduced in 2 performs as expected. At the cost of a very high FRR it is able to reduce the AAR/AIR rate. However, if the size of the bounding box is increased, FRR values drop and AAR/AIR values converge to the previously observed levels.

In previous work [5] a verification algorithm using biometric hashing based on the first 69 features used here has been described. For the semantic given PIN an EER of 0.0832 was determined. It has to be noted, that the underlying database and test methodology are not identical to those used here. However, a trend towards change for the better can be indicated by the usage of the SVM and an enhanced feature set for verification performance.

4 Conclusion and Suggestions for Future Work

The SVM classifier described in this paper is capable of reliable identification of 15 persons who are enrolled using a freely chosen PIN. Because of the tendency to over-fitting, the system exhibits very low FAR/FIR values while producing a still acceptable FRR under 1%. If the enrolled persons are restricted to using the same PIN, system performance drops notably, though this drop affect mostly the FRR which in certain use cases may sill within acceptable ranges.

[2] the origin corresponds to the OP with $FIR = FRR = AAR = 0$ or $FIR = FAR = AIR = 0$

If a more balanced FAR/FIR-FRR-ratio is desired the ν-SVM is an option to consider, since the ν parameter provides information on the class imbalance and hopefully eliminate the tendency to over-fittng. However the elimination of over-fitting might result in an increased AAR/AIR rate, since the positive regions in the feature space are likely to grow in area and thus leave more room for acceptance of samples that otherwise would have been rejected.

The simple bounding box approach suggested in this paper proves to be not suitable to reliable distinguish samples of an enrolled person from samples of a not enrolled person. More sophisticated approaches might be able to reduce the AAR/AIR without extreme effect on the FRR.

Another approach to consider is the one-class SVM which relies solely on positive samples for training. After training the SVM returns whether a presented sample fits the learned distribution. Not being exposed to a majority of negative samples is likely to prevent excessive over-fitting. The total lack of negative samples might lead to an increased FAR/FIR, though.

Furthermore the experiments described in this paper (as well as future experiments based on this work) are to be conducted on a larger data set to ensure statistical significance of the results. It is also considered to adapt the test environment to that of Abreu and Fairhurst [1] to allow direct comparison. Seeking for a deployment-ready authentication system the influence of user count, authentication type, and sample semantic on the parameters of optimal OPs have to be analysed.

Acknowledgements This work is partly supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, project WritingPrint). We would particularly like to thank Jana Dittmann for her support and fruitful discussions in the context of our work.

References

1. Abreu, M.C.C., Fairhurst, M.C.: An empirical comparison of individual machine learning techniques in signature and fingerprint classification. In: B.A.M. Schouten, N.C. Juul, A. Drygajlo, M. Tistarelli (eds.) BIOID, *Lecture Notes in Computer Science*, vol. 5372, pp. 130–139. Springer (2008)
2. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines (2001). Software available at http://www.csie.ntu.edu.tw/˜cjlin/libsvm
3. Fuentes, M., Garcia-Salicetti, S., Dorizzi, B.: On-line signature verification: Fusion of a hidden markov model and a neural network via a support vector machine. In: IWFHR '02: Proceedings of the Eighth International Workshop on Frontiers in Handwriting Recognition (IWFHR'02), p. 253. IEEE Computer Society, Washington, DC, USA (2002)
4. Hsu, C.W., Lin, C.J.: A comparison of methods for multi-class support vector machines. IEEE Transactions on Neural Networks **13**(2), 415–425 (2002)
5. Scheidat, T., Vielhauer, C., Dittmann, J.: Single-semantic multi-instance fusion of handwriting based biometric authentication systems. In: Proceedings IEEE International Conference on Image Processing (ICIP 2007), pp. 393–396 (2007)
6. Vielhauer, C.: Biometric User Authentication for IT Security: From Fundamentals to Handwriting. Springer, New York (2006)