Evaluation of Novel Features and Different Models for Online Signature Verification in a Real-World Scenario

Marcus LIWICKI
German Research Center for Artificial Intelligence (DFKI)
Trippstadter Str. 122
67663 Kaiserslautern, Germany
Marcus.Liwicki@dfki.de

Abstract. In this paper we describe signature verification experiments on a recently collected dataset which is publicly available. We investigate a novel set of local features and compare it to a reference set often used in the literature. Furthermore we compare the use of Gaussian mixture models (GMMs) and hidden Markov models (HMMs) for classification. We optimize all the standard meta-parameters on a validation set and measure the final performance on a separate test set. The task considered in our experiments is the most challenging in automatic signature verification, i.e., to verify a questioned signature when only one reference signature by the claimed author is given. We found out that the system with the novel feature set outperforms the reference system. Furthermore, HMMs perform better than GMMs if we do not restrict the number of model parameters. Despite the difficulty of the task, we could finally achieve an equal error rate of about 3% without optimizing any meta-parameters on the test set.

1. Introduction

Automated signature verification has been considered for many decades (Plamondon & al., 1989; Leclerc & al., 1994). The general task is to decide if a questioned signature has been written by the claimed author or by another person. Usually, a number of reference signatures written by the claimed author is available. The most challenging situation arises if there is only one reference signature at hand. This situation is considered in this paper.

The task of signature verification can be divided into two categories, offline verification and online verification. In the former case only the ink written on paper is available. In the case of online verification the sequential information of the strokes is also available. Additional information like the pen pressure, the azimuth, and the elevation may supplement this information, depending on the acquisition device. Offline verification is the more traditional task, since ink information is the usual writing information. However, with the growing amount of digital pen-input devices used for signature acquisition, forensic experts might receive questioned online signatures in the near future.

Several approaches exist to face the task of online signature verification. While most of them apply similar preprocessing and normalization techniques, they differ in the way of feature extraction and classification. It is a general agreement that selecting the features locally, i.e., one feature vector for each time stamp, is superior to selecting them globally, i.e., one feature vector for the whole signature, especially when only a small set of reference signatures is available (Jain & al., 2002). For the classification of the extracted features a number of methods have been investigated in the past. Jain & al, for example, use a string matching strategy (dynamic time warping) to measure the similarity between the questioned signature and a stored template derived from a set of signatures. Hidden Markov models (HMMs) have been used by Yang & al. (1995). Richiardi & al. (2003) introduced Gaussian mixture models (GMMs) for online signature verification and experimentally found out that they are superior to HMMs if the same number of model parameters is considered.

In this paper we evaluate different signature verification approaches on a publicly available dataset presenting a real-world scenario in which only one reference signature is available. First, we extend the set of typically extracted features with some novel features. Next, we compare GMMs and HMMs in the classification step, allowing all optimization strategies on a validation set. In our broad experiments we found out that our system with the novel feature set performs better than the reference system. Furthermore, we observed that HMMs outperform GMMs, if we do not restrict them to use the same number of model parameters.

This paper is organized as follows. Section 2 introduces the data and briefly describes the normalization and feature extraction methods. Next, Section 3 summarizes the classification models used for the verification system. The experimental results are reported in Section 4, and Section 5 provides a deeper analysis and discussion. Finally, Section 6 draws some conclusions and gives an outlook to future work.

2. Data and Feature Extraction

The data used in this paper comes from the NISDCC signature collection, which has been collected for the ICDAR 2009 signature verification contest (Blankers & al., 2009).[1] It has been acquired with an inking digitizer

[1] The data set is publicly available for research purpose under: http://sigcomp09.arsforensica.org/

pen on a Wacom A4-oversized tablet with a sampling rate of 200Hz. For each point, the *x* and *y* position, as well as the pen pressure, the pen azimuth (0-360°), and the elevation angle (0-90°) have been recorded. Due to the reliability of the acquisition device, the data contains no significant noise or gaps during the recordings. The data has been visually inspected by the authors of (Blankers & al., 2009). Furthermore an offline version of each signature has been scanned. This paper, however, focuses on the online format.

An example signature from the database is illustrated in Figure 1. The corresponding online version of this signature is visualized in Figure 2. For this visualization, succeeding points have been connected and the stroke width has been set according to the pressure of the pen, i.e., the higher the pressure of the pen, the thicker the corresponding line is. Note that there exist points with "zero" pressure, i.e., when the pen has been moved over the tablet without writing. These lines are visualized using a width of one pixel and marking the actually recorded points with 8-pixel dots. This visualization reveals information of the signature which is not available in the offline format and which might be very useful for detecting a forgery.

Figure 1: Example signature from the offline images of the NISDCC database

Figure 2: Offline representation derived from the online version of the signature in Figure 1

The traditional set of features extracted for each point of the online signature is the following (Richiardi & al., 2005): the horizontal and the vertical position, the pressure, the path tangent angle, the total velocity, the velocity in *x* and *y* direction, the total acceleration, the acceleration in *x* and *y* direction, the log radius of curvature, the pen azimuth, and the pen elevation. Furthermore, their first order derivative (approximated using regression), and their second order derivative can be used.

The use of pseudo offline features derived from a 3 × 3 matrix around the considered point has been proposed by (Jain & al., 2002). We furthermore extend this feature set with other features of the local online and offline vicinity. These features have been successfully applied to the task of writer identification recently (Schlapbach & al., 2008). These are: the vicinity aspect and curliness, the curliness and the slope in the vicinity, and the number of ascenders and descenders in the offline vicinity.

3. Classification Models

In this paper we use Gaussian mixture models (GMMs) and hidden Markov models (HMMs) for the classification of the feature vector sequences. Due to space limitations only a short description of these methods is presented in this section.

In text-independent speaker recognition, GMMs have become a dominant approach (Mariéthoz al., 2002; Reynolds & al., 2000). In this paper we use GMMs to model the handwriting of each person. More specifically, the distribution of feature vectors extracted from a person's online handwriting is modeled by a Gaussian mixture density. For a *D*-dimensional feature vector denoted as *x*, the mixture density for a given writer (with the corresponding model λ) is defined as:

$$p(x \mid \lambda) = \sum_{i=1}^{M} w_i \, p_i(x) \qquad \textbf{(1)}$$

In other words, the density is a weighted linear combination of *M* uni-modal Gaussian densities, $p_i(x)$, each parameterized by a $D \times 1$ mean vector, and $D \times D$ covariance matrix. For further details refer to (Schlapbach & al., 2008).

Hidden Markov models are well suited for modeling doubly-stochastic processes, where the behavior of the features is expected to be time-dependent (which holds for handwritten signatures). At each time stamp the model instantaneously jumps from one state to another, and observes a feature vector represented by each state's output distribution. The structure of an HMM is generally described by a set of states *S*, a transition matrix *A* holding the transition probabilities between the states, and the initial probabilities of each state. Furthermore the observation probabilities are given, representing the probability of observing a specific feature vector from a given state. In this paper a mixture of Gaussians is used to model the output distribution for each state of a hidden Markov model. For more details on HMMs refer to (Rabiner, 1989).

Table 1: Final error rates in % on the test set

System	FAR	FRR	EER
Reference (GMM)[a]	4.0	5.3	5.3
Proposed (GMM)[b]	0.9	4.0	3.3
Proposed (HMM)[c]	1.3	3.3	2.7

Optimized values for the meta-parameters:
[a] 8 Gaussians, VAR: 0.06
[b] 4 Gaussians, VAR: 0.08
[c] 10 Gaussians, VAR: 0.05, 30 states

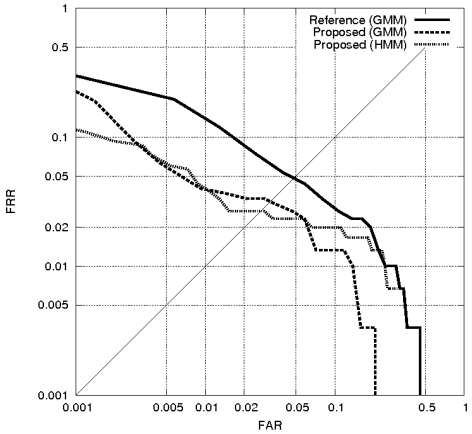

Figure 3: DET curve of the three systems on the test set (FAR = false acceptance rate, FRR = false rejection rate)

4. Experiments and Results

The experiments have been performed on the NISDCC signature collection (Blankers & al., 2009). The data consist of 60 authentic signatures written by 12 writers whereas each writer contributed with 5 signatures. Furthermore, 31 forgers learned to forge all signatures by the 12 writers and made 5 skilled forgeries per writer. For optimizing the meta-parameters of the system, four writer and five forgers have been selected as a validation set. The remaining writers were used for independent testing purpose.

The verification task is to find out if the questioned signature has been written by the claimed author or if it is a forgery. Only one reference signature is available for each question. The verification system (see Section 3) outputs a log likelihood score, which is compared to a threshold. If it is larger than the threshold, the signature is accepted, in the other case it is rejected. The performance was measured by means of false acceptance rate (FAR), and false rejection rate (FRR). By varying the threshold, a saturation point can be found where the FAR equals the FRR. The corresponding score is then called equal error rate (ERR) which is often used to measure the system's performance with a single value.

We optimized the typical meta-parameters in the manner of (Schlapbach & al., 2008). For the GMMs, the number of Gaussian components was varied from 1 to 50, and the variance flooring parameter (VAR) was varied from 0.001 to 0.1. For the HMMs, the number of Gaussian components and the variance flooring parameter where optimized similarly to the GMMs. Furthermore, the number of states was varied from 3 to 50. The system with the best EER (found by varying the threshold) was then used for the final test.

The following systems have been investigated in this paper: the GMM based system with the feature set proposed by (Richiardi & al., 2005), used as a reference system (denoted as Reference (GMM)); the GMM based system using all features described in Section 2 (denoted as Proposed (GMM)); and the HMM based system using all the features (denoted as Proposed (HMM)). For the implementation of the systems we used the Torch 3.0 library (Collobert & al., 2002).

The final results on the test set appear in Table 1. The FAR and the FRR resulting from using the optimized threshold are reported in the first two columns. These are the results without any optimization of the system on the test set. An interesting visualization of the system performance is the detection error tradeoff (DET) curve (Martin & al., 1997), which is a graph with a logarithmic scale showing the FAR and the corresponding FRR, derived by varying the comparison threshold on the test set as well. This curve is visualized in Figure 3. As can be seen, the HMM approach achieves the lowest EER (also given in the last column of Table 1). The best overall system is the HMM based system which achieves an EER of 2.7 %.

5. Discussion

The experimental results indicate that using HMMs works better than using GMMs. This seems to contradict the findings of (Richiardi & al., 2005). However, while in (Richiardi & al., 2005) the HMMs were restricted to have the same number of model parameters as the compared GMMs, we allowed for more parameters in this paper. The best GMM based systems only had 4 Gaussian components, while the best HMM based system uses an HMM with 30 states and 10 Gaussian components per state. Unfortunately, we cannot provide the detailed statistics in this extended abstract because of the space limitations.

Note that in the experiments described in this paper, we did not report on any experiments using universal background models for the GMM based systems. We have performed such experiments, but the performance was significantly lower than the performance of the systems trained from scratch. This might be due to the few adaptation data, i.e., the system was provided with one reference signature only, in contrast to more than three, as used in previous work (Jain & al., 2002, Richiardi & al., 2005).

(a) (b) (c) (d)

Figure 3: Examples for difficult signatures: a) authentic signature of writer 10, b) – c) two other authentic signatures of the same writer, d) skilled forgery contributed by writer 33

A deeper analysis of misclassified signatures shows that there exist very difficult cases in this database. An example of difficult signatures is given in Figure 3. While signatures a-c have been written by the same writer, the last signature is a forgery by a different writer. Even if a human person gets signature a as a reference signature, it would be very hard for him/her to find out that signatures b and c should be accepted, while signature d should be rejected. The system gets only one reference signature in our experiments. We have tested the verification system in a setting where more signatures were provided as reference signatures. There the error rate was significantly lower. Due to space limitations, we cannot provide any more details on these experiments in this paper.

6. Conclusions and Future Work

In this paper we evaluated different signature verification approaches on the NISDCC signature collection. We first extended the set of typically extracted features with new features. In our experiments we compared GMM based and HMM based approaches. We optimized all typical meta-parameters on a validation set. We observed that the system with the new features performs better than the reference system. Furthermore, the HMMs outperformed the GMMs. We also analyzed typical errors of the system and observed that even for a human it is hard to make the correct decision for each signature.

The best EER has been achieved with the GMM based system. It is lower than 3 % and might already be useful for practical applications. There is, however, one weak point in the experimental setting which might be critical in future. The skilled forgers were only provided with an offline version of the signature. As more and more digital interfaces record the online signature, one might expect that the forgers can train the signature using the online information as well. Preliminary ideas of such a system are described in (Wahl & al., 2006). One has to make sure that in realistic scenarios no one is able to access the online signature. An idea to face the problems arising from unauthorized access to the reference signatures would be to include spoken information during signing as well (Humm & al., 2007).

References

Plamondon, R. & Lorette, G. (1989). Automatic signature verification and writer identification - the state of the art. *Pattern Recognition 22(2),* 107–131

Leclerc, F. & Plamondon, R. (1994). Automatic signature verification: the state of the art–1989-1993. *Int'l J. Pattern Recog. and Artificial Intelligence 8(3),* 643–660

Jain, A. K., Griess, F. D. & Connell , S. D. (2002). On-line signature verification. *Pattern Recognition 35(12),* 2963-2972

Yang, L., Widjaja, B.K., & Prasad, R. (1995) Application of hidden Markov models for signature verification. *Pattern Recognition 28 (2),* 161–170

Richiardi, J., Drygajlo, A. (2003). Gaussian Mixture Models for on-line signature verification. *Proc. 2003 ACM SIGMM workshop on Biometrics methods and applications*, pp. 115–122

Richiardi, J., Ketabdar, H., Drygajlo A. (2005). Local and global feature selection for on-line signature verification. *Proc. 8th Int. Conf. Document Analysis and Recognition*, pp. 625–629.

Blankers, V. L., van den Heuvel, C. E., Franke, K. and Vuurpijl, L.G. (2009). "The ICDAR 2009 signature verification competition", in ICDAR2009 proceedings.

Schlapbach, A. and Liwicki, M. and Bunke, H. (2008). A Writer Identification System for On-line Whiteboard Data. *Pattern Recognition 41(7),* 2381-2397

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE, 77(2):*257–286

Mariéthoz, J., Bengio, S. (2002). A comparative study of adaptation methods for speaker verification. *Int. Conf. on Spoken Language Processing*, pp. 581–584

Reynolds, D.A., Quatieri, T.F., Dunn, R.B. (2000). Speaker verification using adapted Gaussian mixture models. *Digital Signal Processing 10*, pp. 19–41

Collobert, R., Bengio, S., & Mariéthoz, J. (2002). Torch: a modular machine learning software library. *IDIAP-RR 02-46*

Martin, A., Doddington, G., Kamm, T., Ordowski, M. & Przybocki, M. (1997). The DET Curve in Assessment of Detection Task Performance. *5th European Conf. Speech Communication and Technology*, pp. 1895-1898

Wahl, A., Hennebert, J., Humm, A. & Ingold, R. (2006). Generation and Evaluation of Brute-Force Signature Forgeries. *Multimedia Content Representation, Classification and Security. LNCS 4105,* 2-9

Humm, A., Hennebert, J. & Ingold, R. (2007). Gaussian Mixture Models for CHASM Signature Verification, *Machine Learning for Multimodal Interaction LNCS 4299*, 102-113