The Use of Confidence Indicating Prediction Score in Online Signature Verification

András Heszler, Cintia Lia Szücs, and Bence Kővári
Department of Automation and Applied Informatics, Budapest University of Technology and Economics, Budapest, Hungary
Email: heszler.andras@gmail.com, {szucs.cintialia, kovari}@aut.bme.hu

***Abstract*—Signature verification is an actively researched area whose goal is to decide whether unknown signatures are genuine or forged. Online signature verification applies signatures captured with an electronic device (digital tablet or pen). Online signatures contain not only spatial information but dynamics as well. There are two types of possible errors, the false prediction as genuine and the false prediction as a forgery. This paper proposes a prediction score as the classification output, which indicates the confidence of the system decision. This approach allows a trade-off between the different error types to create specialized verifiers and construct combined classifiers. This paper presents two types of combined classifiers, pre-filtering classifiers and majority voting classifiers. The proposed approaches are evaluated using the MCYT-100 dataset.**

***Index Terms*—online signature verification, dynamic time warping, DTW, confidence score, prediction score, nonbinary decision, ensemble classification**

I. Introduction

The signature is one of the most commonly used biometrics techniques, especially in the verification and authorization of documents. The advantages of the technique are widespread acceptance and easy usage. However, professionals can deceive the systems with forgeries which is still a challenge to solve. Signature verification is an actual, frequently researched problem [1] [2].

Signature verification can be categorized into offline signature verification and online signature verification. The offline signature verification uses only the signature image for the classification. The online or dynamic signature verification uses a digital tool (tablet or digital pen) to collect the signature as time series with more features, such as pressure. Therefore, the online signatures are harder to forge and show better performance. This paper has focused on online signature verification [3].

In signature verification, two types of error rates can describe the performance, the False Acceptance Rate (FAR) and the False Rejection Rate (FRR). FAR is the percentage of forgeries in the genuine predicted signatures, FRR is the percentage of genuine signatures in the forgery predicted signatures. These two are in a strong relationship according to the applied threshold, as is shown in Fig. 1. The most common error rate used to compare signature verification systems is the equal error rate (EER), where the FAR and FRR are equal.

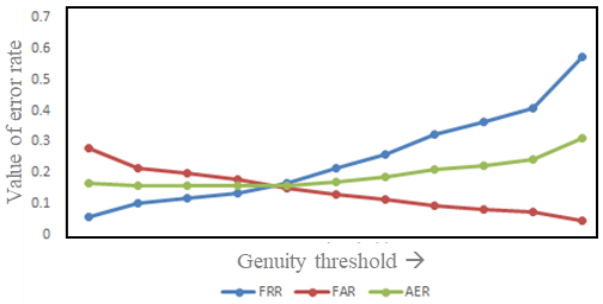

Figure 1. Error rates of a signature verification system with different genuity threshold values.

The Average Error Rate (AER) can approximate the EER if the FAR and FRR values are close to each other, which is usually the state where the AER is the smallest. However, there are particular areas where one type of error is more critical, and specialized signature verification systems can be helpful. For example, in high-security banking actions, a falsely accepted forgery has more cost than a falsely rejected genuine, especially when the user has more attempts.

In Section II, an overview of the signature verification techniques is presented. Section III proposes using a nonbinary classification score over the classic binary classification to create specialized signature verifiers and ensembled classifiers. In Section IV, the proposed methods are evaluated on the publicly available MCYT-100 signature database.

II. Related Work

The online signature verification systems can be divided into function-based and parameter-based systems regarding the features. The parameter-based approach uses features derived from and describing the signature, such as the height and width of the signature or average speed. The method proposed in [4] focuses on mobile devices, uses histogram feature extraction to build the feature set, and achieves 2.72% on the MCYT-100 dataset. The verification system described in [5] uses a Recurrent Neural Network (RNN) and Length-

Manuscript received November 8, 2021; revised April 20, 2022.

doi: 10.12720/jait.13.3.290-294

Normalized Path Signature (LNPS) as a feature extractor to achieve 2.37% EER.

The functional approach uses local properties as time sequences. Since signatures from the same signers can vary in time, proper techniques are needed to calculate the signatures' dissimilarity. The most commonly used techniques are the Hidden Markov Model (HMM) and Dynamic Time Warping (DTW) [6], [7]. The paper of [8] analyses the factors affecting online signature verification using HMM and performs 2.27% EER on the MCYT-100 dataset. The system proposed in [6] uses DTW for verification with RNN used for discriminative feature learning that achieves 1.62% EER on the MCYT-100 dataset. An embedded system for online signature verification presented in [9] also uses the DTW algorithm and achieved 2.74% EER on the MCYT-100 dataset.

The proposed classifiers also use the DTW algorithm. The DTW algorithm is an efficient time-series similarity measure that minimizes the effects of shifting and distortion in time by allowing the "elastic" transformation of time series to detect similar shapes with different phases. Fig. 2. shows the binding of the X coordinates of two signatures with the DTW algorithm. To compare the test time series $X = (x1, \ldots xN)$, $N \in \mathbb{N}$ and reference time series $Y = (y_1, \ldots y_M)$, $M \in \mathbb{N}$, the DTW algorithm uses a distance function $d(x_i, y_j)$ for which the proposed classifiers use Manhattan distance. The DTW algorithm finds the warping path in $O(NM)$ that defines the correspondence of an element x_i to y_j, where $i \in \{1, \ldots N\}$, $j \in \{1, \ldots M\}$. The dissimilarity is computed as the cumulative distance between the points of the warping path [10], [11].

Combined systems are also successful for online verification. The verifier proposed in [12] uses a combination of DTW, RNN, and Linear Programming Descriptor (LPD) in 6:1:1 to achieve 3.81% EER.

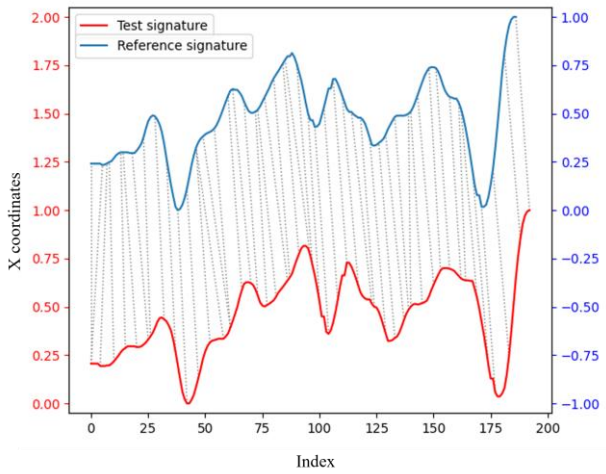

Figure 2. Binding of the X coordinates with DTW algorithm. The reference signature is the 0000v00 signature, and the test signature is the 0000v01 signature in the MCYT-100 dataset.

The approach described in [13] uses an ensemble approach where physical, frequency-based, and statistical features are extracted. The combined EER is 2.84% compared to the 10.39% EER of the exclusively used physical feature extraction, 7.76% EER of the frequency-based feature extraction, and 5.79% EER of the statistical feature extraction.

III. Experimental Method

The verification of a signature is a binary classification problem. A score can be assigned to each signature, 1 for genuine, 0 for forgery. The score can be interpreted as a confidence score if it can take any values in the [0, 1] interval, and signatures with higher scores have higher probabilities of being genuine. The classification method that produces the confidence score is properly calibrated if the score represents the probability of being genuine accurately. This paper aims to define such a prediction score and present classification solutions taking advantage of the use of this prediction score.

The proposed verifiers will be presented using a self-developed signature verification framework. The signature verification system uses different pre-processing steps on the signatures in the following order:

1) Extracting the pressure, X, and Y coordinates from the features.
2) Filtering out the zero pressure points, where the pen was up.
3) Scaling the features with either the min-max or the standard scaling method.
4) Translating the features with either translate-center-to-null or translate-min-to-null translate method.

The classification method used for the proposed verifiers uses the DTW algorithm to compare two pre-processed signatures. For each signer, ten genuine signatures are chosen for reference. The rest of the signatures are the signatures in question, also known as test signatures. The reference signatures are compared to each other, and the results are stored in a 10×10 array R, where $R(i,j)$ is the dissimilarity between the i-th and j-th signature by the DTW algorithm and $R(i,i)$ is empty for every $1 \leq i \leq 10$, $1 \leq j \leq 10$. Each signature in question is compared to all of the reference signatures. The results are stored in a 10×1 vector S, where $S(i)$ is the difference between the signature in question and the i-th reference signature.

A very flexible method is used to create a threshold value t from the reference matrix R and a signature value s from matrix S. First, the R array is transformed to a vector using one of the following aggregation functions: minimum, maximum, mean, or median. The chosen function will be referred to as the Reference Matrix Aggregation Method (RMM). Then this vector constructs the t value by using one of the earlier functions, referred to as the Reference Vector Aggregation Method (RVM). The s value is created similarly, using one of the earlier functions on the S vector. The chosen method is referred to as the Test Vector Aggregation Method (TVM).

The signature in question is predicted as genuine if $s<t$, otherwise predicted as a forgery. The more significant the difference between s and t, the more confident the method is in the decision. The base of the nonbinary score used by the model is the $d=(s-t)/t$ value that is

negative for genuine signatures and positive for forgeries. The advantages of the *d* value are that it can be used to make a prediction by itself and that for all signatures, this value stays in the same magnitude, while the *s* and *t* values may take larger values in case of longer signatures. However, *d* takes values in the [-1, ∞] interval, so it cannot be interpreted as a confidence score. For this, the *s* confidence score is calculated by the $s=1+e^{-\alpha d}$ equation, where α=14 value was chosen after calibration.

The usage of this nonbinary score provides the opportunity to create a trade-off between the False Rejection Rate (FRR) and the False Acceptance Rate (FAR) by defining a confidence score border between the forgeries and genuine signatures. The FRR will decrease, and the FAR will increase by classifying more signatures as genuine, using a lower border and vice versa. By this trade-off, classifiers with optimal AER can be produced, as well as specific signature verifiers, for example, one with very low FAR for high-security applications.

Another usage of specific verifiers with very low FAR and FRR can be to filter out signatures with a minimal error rate before a more general verifier classifies the remaining signatures. First, a classification method that produces low FAR and reasonable FRR filters out the genuine signatures by assigning them a score of 1. Then a classification method that produces low FRR and reasonable FAR filters out the forgery signatures by assigning them a score of 0. For the remaining signatures, a more general classification method will make the final decision. In this paper, this method will be referred to as the pre-filter classification method, and the verifiers using this method will be referred to as pre-filter verifiers.

The key of the pre-filter classification is to have classifiers that can filter out most of the signatures without a significant error rate. However, due to the trade-off between the FRR and FAR values, a classifier that makes minimal mistakes will be able to filter out just the most obvious signatures. In this paper, a few methods are presented with the FRR and FAR values representing the risk level and the usefulness of the methods.

The nonbinary scores can also be used as weights for a weighted majority algorithm. Instead of using a single classification method, more methods predict the class of the signature. The final decision will be made as a vote between the classifiers, weighted with the confidence score of the prediction. In this paper, this method will be referred to as the majority classification method, and the verifiers using this method will be referred to as majority verifiers.

The voting classification methods should be general signature verifiers and not specialized for genuine or forgery signatures because each of them will be part of the voting for all signatures. Also, they should be diverse enough to produce different results for the voting. This paper presents a few non-combined classifiers, which can be helpful for majority classification, and the majority classifiers created from them.

IV. Experimental Results

Experiments were conducted on the MCYT-100 database [14], consisting of 25 genuine and 25 skilled forgery samples of 100 writers.

To avoid overfitting, the signers are divided into two sets. The first 70 signers build up the training dataset, and the rest 30 signers build up the validation dataset. The best methods will be found on the training dataset, and the error rates will be evaluated on the validation dataset.

The best baseline verifiers, using the earlier presented classification methods, can be seen in Table I. The B1 verifier has the lowest average error rate and nearly equal FAR and FRR. The B2 and B3 verifiers have higher AER than B1, and most of this error comes from false acceptance.

Using the nonbinary score v for the decision can freely specify the border value between genuine and forgery signatures. The verifiers with the smallest AER can be seen in Table II.

All of the verifiers have much lower error rates than the binary verifiers. Moreover, the NB1 and NB3 verifiers have minimal false acceptance rates, making them ideal verifiers for high-security areas, such as banking.

TABLE I. Binary Verifiers

Code	Scale	Translate	RMM	RVM	TVM	FRR	FAR	AER
B1	Min-max	Center to null	max	mean	mean	4.00%	5.60%	4.80%
B2	Standard	Center to null	max	min	min	2.44%	8.00%	5.22%
B3	Standard	Center to null	mean	max	min	0.44%	12.40%	6.42%

TABLE II. Nonbinary Verifiers

Code	Scale	Translate	RMM	RVM	TVM	border	FRR	FAR	AER
NB1	Standard	Center to null	mean	median	min	0.20	2.89%	0.00%	1.44%
NB2	Standard	Center to null	median	mean	min	0.10	1.11%	2.53%	1.82%
NB3	Standard	Center to null	median	median	min	0.20	3.55%	0.13%	1.84%

TABLE III. GENUINE FILTERS

Code	RMM	RVM	TVM	border	FRR	FAR
GL	median	median	mean	0.10	16.44%	0.00%
GH	mean	median	min	0.40	5.55%	0.40%

TABLE IV. FORGERY FILTERS

Code	RMM	RVM	TVM	border	FRR	FAR
FL	max	max	min	0.10	0.00%	37.73%
FH	max	max	min	0.75	0.22%	14.67%

TABLE V. PRE-FILTER VERIFIERS

Code	Scale	Translate	Genuine filter	Forgery filter	RMM	RVM	TVM	FRR	FAR	AER
P1	Min-max	Center to null	GH	FL	max	max	max	2.89%	1.20%	2.04%
P2	Min-max	Center to null	GH	FL	mean	mean	max	5.33%	0.40%	2.87%
P3	Min-max	Center to null	GH	FL	mean	max	mean	3.11%	4.80%	3.96%

Similar to the previous verifiers, the pre-filter classifiers were chosen based on the first 70 signers and the error rates evaluated on the remaining 30 signers. One low-risk and one high-risk classifier were used for the genuine and forgery filters on the training dataset that can be seen in Table III and Table IV.

Although the pre-filter classifiers could not improve the nonbinary verifiers' already impressive results, they can be used with traditional binary classifiers, for which the results can be seen in Table V. The best verifiers are found by using the high-risk genuine filter with the low-risk forgery filter. P2 can be used for high-security demanding areas, while P1 and P3 have more balanced error rates for general usage.

The majority classification could not further improve the results of the nonbinary verifiers. On the other hand, it can be used with traditional binary classifiers as well. Some of the best combinations of binary classification methods were chosen regardless of the pre-processing methods, seen in Table VI.

The majority classification elements were tested with every combination and evaluated similarly to the previous ones. The results can be seen in Table VII. The error rates greatly improved, and the M1 combination with the minimal false acceptance rate makes it an ideal classifier for high-security areas.

TABLE VI. MAJORITY CLASSIFICATION ELEMENTS

Code	RMM	RVM	TVM
1	max	min	min
2	mean	mean	min
3	median	mean	min
4	mean	median	min
5	median	median	min

TABLE VII. MAJORITY VERIFIERS

Code	Scale	RVM	Combination	FRR	FAR	AER
M1	Standard	Center to null	1,2,5	5.33%	0.00%	2.67%
M2	Standard	Center to null	1,2	3.78%	2.40%	3.09%
M3	Min-max	Minimum to null	1,2,3,5	3.56%	3.87%	3.71%

V. CONCLUSION

This paper presented a confidence score for signature verification systems and signature classification methods based on this nonbinary score. The proposed methods were evaluated on the MCYT-100 dataset. The results show improved classifiers, similar to the presented state-of-the-art verification systems. However, the proposed classifiers do not have a balanced error rate and are more suitable for a specific area of use, like high-security applications.

The confidence score can be improved by more accurate and classifier-specific calibration. The classifiers can be further improved to reach a balanced, low error rate for general usage. Fundamentally different classifiers can be used for the ensembled classifiers, such as offline signature verifiers or parameter-based systems, to take advantage of each.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

This work is a product of András Heszler's Master thesis research under the supervision and guidance of Cintia Lia Szücs. She helped his work with significant advice and editing notes, which were used in the final version. Bence Kővári provided critical feedback and gave his advice.

ACKNOWLEDGMENT

The work presented in this paper has been carried out in the frame of project no. 2019-1.1.1-PIACI-KFI-2019-00263, which has been implemented with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the 2019-1.1. funding scheme.

REFERENCES

[1] K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 4-20, 2004.

[2] T. Sabhanayagam, V. P. Venkatesan, and K. Senthamaraikannan, "A comprehensive survey on various biometric systems,"

International Journal of Applied Engineering Research, vol. 13, no. 5, pp. 2276-2297, 2018.
[3] J. Vargas, M. Ferrer, C. Travieso, and J. Alonso, "Off-line signature verification based on grey level information using texture features," *Pattern Recognition*, vol. 44, no. 2, pp. 375-385, 2011.
[4] N. Sae-Bae and N. Memon, "Online signature verification on mobile devices," *IEEE Transactions on Information Forensics and Security*, vol. 9, pp. 933-947, June 2014.
[5] S. Lai, L. Jin, and W. Yang, "Online signature verification using recurrent neural network and length-normalized path signature descriptor," in *Proc. 14th IAPR International Conference on Document Analysis and Recognition*, vol. 01, pp. 400-405, Nov 2017.
[6] S. Lai and L. Jin, "Recurrent adaptation networks for online signature verification," *IEEE Transactions on Information Forensics and Security*, vol. 14, pp. 1624-1637, June 2019.
[7] Y. Liu, Z. Yang, and L. Yang, "Online signature verification based on DCT and sparse representation," *IEEE transactions on cybernetics*, vol. 45, no. 11, pp. 2498-2511, 2014.
[8] E. Argones Rua and J. L. A. Castro, "Online signature verification based on generative models," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, pp. 1231-1242, Aug. 2012.
[9] M. López-García, R. Ramos-Lara, O. Miguel-Hurtado, and E. Cantó-Navarro, "Embedded system for biometric online signature verification," *IEEE Transactions on Industrial Informatics*, vol. 10, pp. 491-501, Feb. 2014.
[10] T. Giorgino, *et al.*, "Computing and visualizing dynamic time warping alignments in R: The dtw package," *Journal of Statistical Software*, vol. 31, no. 7, pp. 1-24, 2009.
[11] P. Senin, "Dynamic time warping algorithm review," *Information and Computer Science Department University of Hawaii at Manoa Honolulu, USA*, vol. 855, no. 1-23, p. 40, 2008.
[12] L. Nanni, E. Maiorana, A. Lumini, and P. Campisi, "Combining local, regional and global matchers for a template protected on-line signature verification system," *Expert Systems with Applications*, vol. 37, no. 5, pp. 3676-3684, 2010.
[13] P. Bhowal, D. Banerjee, S. Malakar, and R. Sarkar, "A two-tier ensemble approach for writer dependent online signature verification," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1-20, 2021.
[14] J. Ortega-Garcia, *et al.*, "MCYT baseline corpus: A bimodal biometric database," *IEE Proceedings - Vision Image and Signal Processing*, vol. 150, pp. 395-401, 2003.

András Heszler received his M.S in software engineering from. Budapest University of Technology and Economics, Hungary in 2021. He has been working as software consultant at TNG Technology Consulting.

Cintia Lia Szücs is a Ph.D. student in software engineering at Budapest University of Technology and Economics, Hungary. Her research interests include online signature verification. She is a member of the Hungarian Association for Image Processing and Pattern Recognition, a member of the John von Neumann Computer Society.

Dr. Bence Kővári received his Ph.D. degree in software engineering from Budapest University of Technology and Economics, Hungary, in 2013, studying the automated verification of handwritten signatures. He is a member of the Hungarian Association for Image Processing and Pattern Recognition and a member of the John von Neumann Computer Society.