Characterization of Handwritten Signature Images in Dissimilarity Representation Space

Victor L. F. Souza[1], Adriano L. I. Oliveira[1], Rafael M. O. Cruz[2], and Robert Sabourin[3]

[1] Centro de Informática - Universidade Federal de Pernambuco, Recife, Pernambuco, Brazil
vlfs@cin.ufpe.br, alio@cin.ufpe.br
[2] Stradigi AI, Montreal, Quebec, Canada
rafaelmenelau@gmail.com
[3] École de Technologie Supérieure - Université du Québec, Montreal, Quebec, Canada
robert.sabourin@etsmtl.ca

Abstract. The offline Handwritten Signature Verification (HSV) problem can be considered as having difficult data since it presents imbalanced class distributions, high number of classes, high-dimensional feature space and small number of learning samples. One of the ways to deal with this problem is the writer-independent (WI) approach, which is based on the dichotomy transformation (DT). In this work, an analysis of the difficulty of the data in the space triggered by this transformation is performed based on the instance hardness (IH) measure. Also, the paper reports on how this better understanding can lead to better use of the data through a prototype selection technique.

Keywords: Offline signature verification · Writer-independent signature verification · Dichotomy transformation · Prototype selection · Instance hardness.

1 Introduction

Handwritten signature is one of the oldest accepted biometric characteristics and is still widely used to verify if a person is who he/she claims to be [2]. The handwritten signatures verification (HSV) systems are used to classify query signature as genuine or forgeries. While genuine signatures are those that really belong to the indicated person, forgeries are those created by other people and can be categorized as [11]: (i) random forgeries, where the forger does not know both the name and the signature pattern of the signer; (ii) simple forgeries, in which the forger only has the access to the name of the writer but does not know the signature pattern; (iii) skilled forgeries, where the forger has the knowledge of both the name and the signature pattern of the signer (resulting in forgeries more similar to the genuine signatures).

While being researched for a long time the HSV problem still remains challenging. Depending on how it is handled, the following challenges can be faced:

imbalanced class distributions, high number of classes, high-dimensional feature space and small number of learning samples [8]. A specific concern is related to the skilled forgeries since they tend to be very similar to the genuine signatures and, in real applications, they are not available during the training phase of the classifier (which should be trained only with genuine signatures and the random forgeries) [8].

There are two approaches for building offline HSV systems. In the Writer-Dependent (WD) systems, a verification model is trained for each user. Although, in general, WD systems present good performance for the HSV task, requiring a classifier for each user increases the complexity and the cost of the system operations as more users are added. Also, the small number of genuine samples per user is a problem that often needs to be addressed. The other systems are known as Writer-Independent (WI). In WI, a single model is trained for all users from a dissimilarity space generated by the dichotomy transformation (DT). Thus, the classification inputs are dissimilarity vectors, which represent the difference between each feature of a questioned and a reference signature of the writer. WI systems are less complex but in general obtain worse results, when compared to the WD approach [7].

Since the samples in dissimilarity space generated by the dichotomy transformation are formed through the combination two by two of signatures (a questioned and a reference signature), this approach is able to increase the number of samples in the WI-HSV scenario. Thus, the small number of samples is no longer a problem. Moreover, the system can be developed to handle the class imbalance by generating a similar number of samples for the positive and negative classes. However, many samples in the WI-HSV scenario are redundant and have little influence for training the verification model. Thus, the use of prototype selection (PS) techniques in the dissimilarity space may enable the reduction of the complexity and the computational cost of training a classifier without deteriorating the final model performances [5].

The objective of this paper is (i) to understand the difficulty of the data and (ii) to analyze the use of prototype selection in the offline WI-HSV based on the dichotomy transformation. Related to (i) the instance hardness (IH) measure is used to achieve the stated objective. The IH is a metric used both to identify hard to classify instances and also to understand why they are hard to classify [12]. One of the advantages of understanding the instances misclassification is to have ideas about the best preprocessing technique or the best classifier to be used [12]. To complement this understanding, in (ii), we analyze if a prototype selection preprocessing can be applied without degrading the performance of the classifier and whether preprocessing based on a systematic prototypes selection technique is better than a random subsampling.

This paper is organized as follows: Section 2 presents the HSV problem and the dichotomy transformation as fundamentals related to this work. Section 3 contains the discussion and the conducted experiments for both the prototype selection and the instance hardness analysis. In the last section, the conclusion and the future works are presented.

2 Fundamentals

2.1 Handwritten Signature Verification (HSV)

The problem of automatic handwritten signature verification (HSV) is commonly modeled as classifying a given signature as genuine (i.e. belonging to the indicated writer) or forgery (created by someone else) [2, 8]. Figure 1 depicts examples of signatures, from the GPDS dataset. Each row represents a different writer and for each writer the first three signatures are genuine and the last one is a skilled forgery.

Fig. 1. Signatures.

In the skilled forgeries the forger knows both the name and the signature pattern of the signer and will attempt to imitate the genuine signature. Thus, genuine signatures and skilled forgeries tend to be very similar. From Figure 1, one can see that skilled forgeries are more similar to the genuine signature than the random forgeries (the genuine signatures from other users).

Also, in a real scenario, the systems are trained with partial knowledge. In general, the training set of HSV systems are composed only of genuine signatures without access to skilled forgeries [8]. So, the classifier is trained without information capable of distinguishing between genuine signatures and forgeries. However, during the verification process the system will have to both reject the forgeries and accept the genuine signatures.

Furthermore, the number of genuine samples per user is often small (between 3-5 signatures) and there is great intra-class variability. This is difficult to tackle since the few available signatures are not sufficient to capture the full range of variation [8]. Figure 1 shows the variability in the genuine signatures.

In the WD systems, one classifier is trained for each user. In the WI case, a single model is trained for all users and the classification only depends on the input reference signature. The common practice for WI-HSV systems is to train using the development set D and to test using the exploitation set ε. In general, these sets have different subset of users [7].

The current state-of-the-art in feature representation for offline signatures is reported in the paper by Hafemann et al. [7], which uses Deep Convolutional

Neural Networks (DCNN) for learning the signature representations in a WI way. So, it tries to learn a new feature space with the most representative properties of the handwritten signatures. As a WI approach, the learned representation space is not specific for a single set of users and is able to use data from as many users as possible. In this work, the 2048 features obtained from the FC7 layer of the DCNN called SigNet are used as feature vectors [7] (available online[4]).

2.2 Dichotomy Transformation (DT)

The Dichotomy Transformation (DT), proposed by Cha and Srihari [3], is an approach that allows to transform a multi-class problem (as the offline HSV) in a 2-class problem. The Dichotomy Transformation has already been used in various contexts, including writer identification [1] and for handwritten signature verification [4, 11, 14]. For the HSV context, it can be presented as follows: given a reference signature and a questioned signature, the objective is to determine whether the two signatures were produced by the same writer.

In a more formal definition, let $\mathbf{x}_q$ and $\mathbf{x}_r$ be two feature vectors in the feature space, the distance vector in the dissimilarity space resulting from the Dichotomy Transformation, $\mathbf{u}$, is computed by equation 1:

$$\mathbf{u}(\mathbf{x}_q, \mathbf{x}_r) = \begin{bmatrix} |x_{q1} - x_{r1}| \\ |x_{q2} - x_{r2}| \\ \vdots \\ |x_{qn} - x_{rn}| \end{bmatrix} \tag{1}$$

where $|\cdot|$ represents the absolute value of the difference, x_{qi} and x_{ri} are the n-th feature of the signatures $\mathbf{x}_q$ and $\mathbf{x}_r$ respectively, and n is the number of features. It is worth highlighting that each component of the $\mathbf{u}$ vector is equal to the distance between the corresponding components of the vectors $\mathbf{x}_q$ and $\mathbf{x}_r$. Thus, both the distance vector and the feature vectors have the same dimensionality.

As previously noted, in the dissimilarity space, regardless of the number of writers, there are only two classes: (i) The within/positive class w_+: composed of distance vectors computed from samples of the same writer (i.e., intraclass distances). (ii) The between/negative class w_-: composed of distances vectors computed from samples of different writers (i.e., interclass distances).

Systems based on the DT approach need datasets already transposed into the dissimilarity space to train a dichotomizer (two classes classifier), which will be used to perform the verification task. Generally, the writers that are in the training set are not part of the test set [3].

When users have more than one reference signature, the dichotomy transformation is applied between the feature vector $\mathbf{x}_q$ of the questioned signature and the writer's reference set $\{\mathbf{x}_r\}_1^R$, producing a set of dissimilarity vectors $\{\mathbf{u}_r\}_1^R$, where R is the number of signatures in the reference set. For example,

[4] http://en.etsmtl.ca/Unites-de-recherche/LIVIA/Recherche-et-innovation/Projets/Signature-Verification

if a writer has 3 reference signatures ($R = 3$) and $\{\mathbf{u}_r\}_1^R = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$. Then, the dichotomizer evaluates each dissimilarity vector individually and produces a set of partial decisions $\{f(\mathbf{u}_r)\}_1^R$ [11]. The final decision about the questioned signature is based on the fusion of all partial decisions by a function $g(\cdot)$ and depends on the output of the dichotomizer, e.g. (i) in a label case, then the majority vote is an appropriate function, (ii) in a probability or distance output, the sum, mean, median, max, and min functions can be used [11].

It is expected that the signatures from the same writer be close to each other in the feature space. Hence, they will be clustered close to the origin in the dissimilarity space. In turn, signatures of different writers are typically distant from each other in the feature space and away from the origin in the dissimilarity space [3]. This behavior can be seen in Figure 2, which depicts a 2D feature space with three writers (classes 1, 2 and 3), each one with 10 signatures and the respective dichotomy transformation to the dissimilarity space.

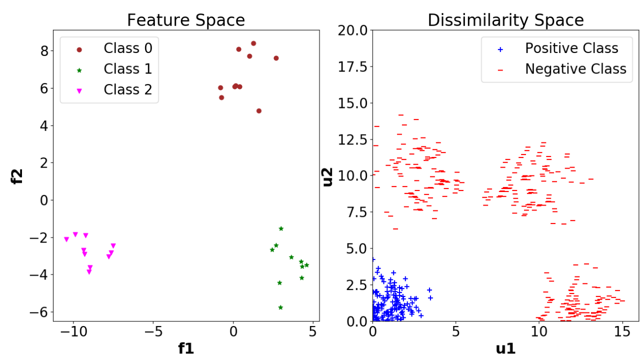

Fig. 2. On the left, samples from three different writers are represented in the feature space. On the right is the representation after the dichotomy transformation.

However, this does not always happen. A disadvantage of DT is that writers perfectly grouped in the feature space may not be perfectly separated in the dissimilarity space [3]. In other words, the more scattered the writers' samples are in the feature space, the smaller is the ability of the dichotomy transformation to separate the samples from the positive and negative classes [11]. This behavior can be seen in Figure 3.

Other properties of the dichotomy transformation deserve to be highlighted. Firstly, DT is able to increase the number of samples in the dissimilarity space, hence it is composed of each pair of signatures. That is, if K writers provide a set of R reference signatures each, the equation 1 generates up to $\binom{KR}{2}$ different distances vectors. Of these, $K\binom{R}{2}$ belong to the positive class and $\binom{K}{2}R^2$ to the negative class [11]. Thus, even using a small amount of reference samples from each writer, the dichotomy transformation is able to generate a large amount

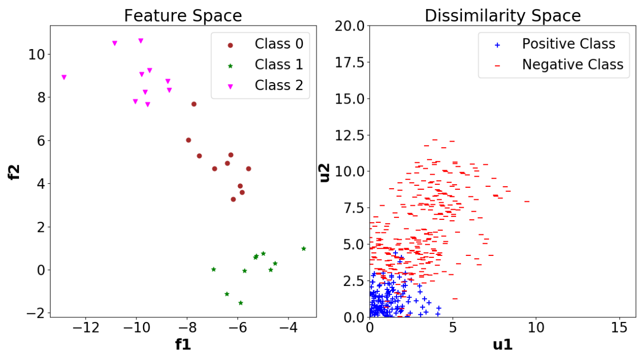

Fig. 3. On the right the dissimilarity space was not able to perfectly separate samples from the within and between classes. This was due to the sparsity of the samples in the feature space (left).

of samples in the dissimilarity space. The increased number of samples can be visualized in both Figures 2 and 3. In Figure 2, for example, 30 samples from the feature space were transformed into 435 samples in the dissimilarity space (being 135 samples from positive class and 300 negative samples).

Also, DT affects the geometry of the data distribution. In addition, the vectors in the dissimilarity space are always non-negative, since they consist of distances transformed into absolute values [3]. Both of these properties can be seen in both Figures 2 and 3.

To illustrate how the verification process through dichotomy transformation is independent of writer, given $\mathbf{x}_q$ and $\mathbf{x}_r$ as respectively a questioned and a reference feature vector, both of a new writer "class 4". The DT computes the distance vector $\mathbf{u}$ between $\mathbf{x}_q$ and $\mathbf{x}_r$ (Eq. 1), which must be located in the within region of the dissimilarity space, being the dichotomizer able to authenticate the questioned and reference signatures as belonging to the same writer. On the other hand, if the same scenario were used in the feature space, the model would fail to perform the classification. In fact, it is impossible for the feature domain model to properly classify signatures as belonging to the "class 4" writer, since this writer is not present in the training set. Therefore, the writer-independence of system is obtained by the use of the dichotomy transformation [11].

3 Experiments

The objective of the experiments is to both analyze the use of prototype selection applied to the offline WI-HSV based on dichotomy transformation and also obtain a better understanding on the difficulty of the data from the dissimilarity space generated by the dichotomy transformation based on the instance hardness (IH) measure.

3.1 Dataset

The experiments are carried out using the GPDS dataset, which has 881 writers and 24 genuine signatures plus 30 skilled forgeries per writer. We use the GPDS-300 segmentation, so the Exploitation set ε is composed by the first 300 writers, the other ones form the Development set D.

The Development set segmentation was done considering the methodology adopted by Rivard et al. [11] and by Eskander et al [4]. The learning set L is generated using a subset of 14 of the 24 genuine signatures from the development dataset. So, the positive class samples are all computed using genuine signatures from every writer, as in Table 1. To generate an equivalent number of counterexamples, the negative class is formed by using 13 genuine signatures (reference signatures) against 7 random forgeries, each one selected from a genuine signature of 7 different writers (Table 1). The Exploitation set is acquired as in [7].

Table 1. Development set segmentation of the GPDS-300 dataset.

Learning set (L)	
Positive Class	Negative Class
Distances among the 14 signatures per writer (D_1)	Distances among 13 signatures per writer and 7 random signatures of other writers
$581 \cdot 14 \cdot 13/2 = 52,871$ samples	$581 \cdot 13 \cdot 7 = 52,871$ samples

3.2 Experimental setup

Before feeding the classifier, the distance vectors $\mathbf{u}$ (in the dissimilarity space) are standardize by removing the mean and scaling to unit variance.

In this paper, the SVM is used as writer-independent classifier with the following settings: RBF kernel, $\gamma = 2^{-11}$ and $C = 1.0$ [14]. The predicted confidence scores for samples are used as classifiers output. The confidence score for a sample is the signed distance of that sample to the classifier's hyperplane [14].

All data were randomly selected and a different SVM was trained for each replication (ten replications for each experimental configuration were performed).

The performance evaluation of the classification methods is based on the Equal Error Rate (EER) metric with user thresholds (considering just the genuine signatures and the skilled forgeries) [7]. In the paper by Souza et al. [14], in general, for the tested dataset, the best results are obtained using the highest number of references and max as fusion function. So, only this approach is considered. To evaluate the effectiveness of the results, the Wilcoxon paired signed-rank test with a 5% level of significance was conducted to confirm if two methods are significantly different.

3.3 Using Prototype Selection (PS)

Considering the main characteristics of the WI dichotomy transformation, it is able to handle with some of the HSV problem difficulties when compared to the WD approach. (i) The DT reduced the high number of classes to a 2-class problem. (ii) The problem is no longer imbalanced as both positive and negative classes have the same number of samples (Table 1). (iii) The small number of samples is no longer a problem (Table 1). The dichotomy transformation was able to increase the number of samples in the WI-HSV scenario, yet many of them are redundant (a disadvantage). Thus, the use of prototype selection in the dissimilarity space can reduce the impact of this redundancy issue.

Prototype Selection (PS) approaches aim to obtain a representative training subset, in general, with a lower number of samples compared to the original one ($SelectedSubset \subseteq TrainingSet$) [5]. One of the main advantages of PS methods is the capacity to choose relevant training examples. So, by using the selected subset, it is expected to obtain similar or even better performance of the classifier during the generalization phase.

In the paper by Pekalska et al. [10], the authors present the prototype selection as an important preprocessing technique to be considered when dealing with dissimilarity-based classification. In their experiments they showed that by using well chosen selected prototypes, it is possible obtain a similar or higher classification performance at a lower computational cost in the classifier training process, when compared to the use of all the original training samples. To the best of our knowledge, this analysis has not yet been done specifically for the dichotomy transformation scenario.

In this work, the classical Condensed Nearest Neighbors (CNN) is used as prototypes selection technique. This approach maintains the instances that are misclassified by a 1-NN classifier, discarding them otherwise [9]. The CNN was choosen because its main property is to reduce the training set size by removing redundant instances (i.e. samples that will not affect the classification accuracy of the training set), retaining the instances closer to the decision boundaries [5]. In our experiments, the K_{CNN} is set to 1, as in the original algorithm [9].

The following experiments compare the application of prototype selection before training the SVM, considering the GPDS-300 dataset. The $\%\_SVM$ represent the models with uniformly random subsampling: 1.0%, 5.0% and 10.0% of the original training set were used. The Condensed Nearest Neighbors is called CNN_SVM in the tables.

Tables 2 and 3 present, respectively, the comparative analysis on the number of training samples and the EER metric obtained by the WI-SVMs (with and without prototype selection). It is worth noting that Table 3 contains both the comparison of the SVMs and also the results from the state of the art models.

Table 2. Comparison of WI-SVMs considering the number of training samples.

Model	#Positive Samples	#Negative Samples	#Retained Samples (%)
SVM	52871	52871	100.00 (0.00)
$1\%\_SVM$	531.70 (17.04)	526.30 (17.04)	1.00 (0.00)
$5\%\_SVM$	2648.10 (24.78)	2639.90 (24.78)	5.00 (0.00)
$10\%\_SVM$	5289.30 (31.69)	5285.70 (31.69)	10.00 (0.00)
CNN_SVM	345.90 (15.25)	4437.80 (125.11)	4.52 (0.13)

Table 3. Comparison of EER with the state-of-the-art on the GPDS-300 dataset, WI-SVMs using Max as fusion function (errors in %).

Type	Model	#references	EER
WD	Soleimani et al. [13]	10	20.94
WD	Hafemann, Sabourin and Oliveira [6]	12	12.83
WD	Hafemann et al. [7]	5	3.92 (0.18)
WD	Hafemann et al. [7]	12	3.15 (0.18)
WI	SVM_{max}	12	3.69 (0.18)
WI	$1\%\_SVM_{max}$	12	3.54 (0.26)
WI	$5\%\_SVM_{max}$	12	3.62 (0.32)
WI	$10\%\_SVM_{max}$	12	3.48 (0.12)
WI	CNN_SVM_{max}	12	3.47 (0.15)

From Table 2, the use of the PS method allowed the SVM to be trained with a much smaller number of samples. Thus, when compared to the model trained with all the original training set, by using PS it was possible to obtain comparable performance (Table 3) with a reduction in computational cost and complexity in SVM training, considering the offline WI-HSV scenario.

As can be observed in Tables 2 and 3, a simple random subsampling with 1.0% of the training samples maintains the results when compared to the SVM trained with all the original training set. This demonstrates how redundant the samples resulting from the dichotomy transformation are for this dataset.

As presented in Table 3, even operating in a writer-independent way, both models with and without preprocessing obtained comparable results for the EER metric when compared to the WD-model from Hafemann et al. [7] for the GPDS-300 dataset. In the comparison to the other models, the proposed approach obtained better results.

By using a systematic PS, such as the CNN, more attention can be given for border samples. Thus, the prototype selection can be used without degrading the performance of the WI-classifier and still avoid to store more instances than are necessary for an accurate generalization.

Another point that should be considered when studying the difficulty of the data in the WI-HSV scenario is the available number of reference signatures. Figure 4 depicts the average EER for the CNN_SVM as a function of the number of reference signatures used for verification. As can be observed, the less signatures are used, the more difficult the problem becomes.

3.4 Instance Hardness (IH) analysis

The instance hardness (IH) is a metric used to identify instances that are hard to classify and to understand why they are misclassified [12]. In this work, the kDisagreeing Neighbors (kDN) is used as instance hardness measure. Given an instance's neighborhood, the kDN represents the percentage of instances that do not have the same label of itself. This metric has a high correlation with the probability that a given instance is misclassified by different classification methods [12]. The kDN hardness measure is computed by Equation 2:

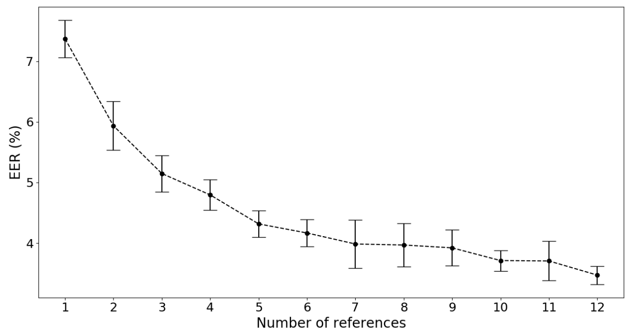

Fig. 4. Performance of the CNN_SVM varying the number of references.

$$kDN(x_q) = \frac{|x_k : x_k \in KNN(x_q) \wedge label(x_k) \neq label(x_q)|}{K} \quad (2)$$

where $KNN(x_q)$ represents the set of K nearest neighbors of a query instance x_q and x_k represents an instance in its neighborhood. $label(x_q)$ and $label(x_k)$ represent the class labels of the instances x_q and x_k respectively [12].

In this section we analyze the data difficulty of the HSV problem, by using the IH measure, using the exploitation set to characterize the problem. This analysis considers different types of signatures (genuine signatures, random and skilled forgeries) and the evaluation of different values of the neighborhood size K in the kDN measure (K in the interval [3,51]).

The main characteristics of the dichotomy transformation (see Figs. 2-3) are: (i) signatures that are close to each other in the feature space will be close to the origin in the dissimilarity space and (ii) the further away two signatures are from each other in the feature space, the farther from the origin will be the transformed vector [3].

In the original feature space [7], genuine signatures from the writers form dense clusters; the skilled forgeries can present two different behaviors: (i) for some writers skilled forgeries are very separate from the genuine signatures, and (ii) in other cases they are closer to the genuine signatures.

Considering this, it is expected that the dissimilarity space generated by the dichotomy transformation presents the following characteristics: (i) positive samples will be close to the origin, forming also a dense cluster in the dissimilarity space (DS), (ii) skilled forgeries with a larger separation will generate negative samples farther away from the origin in the DS, (iii) skilled forgeries close to the genuine signatures will generate negative samples closer to the origin in the DS. As random forgeries are genuine samples from other writers and different writers occupy different regions of the feature space, negative samples from random forgeries will be located far from the DS origin.

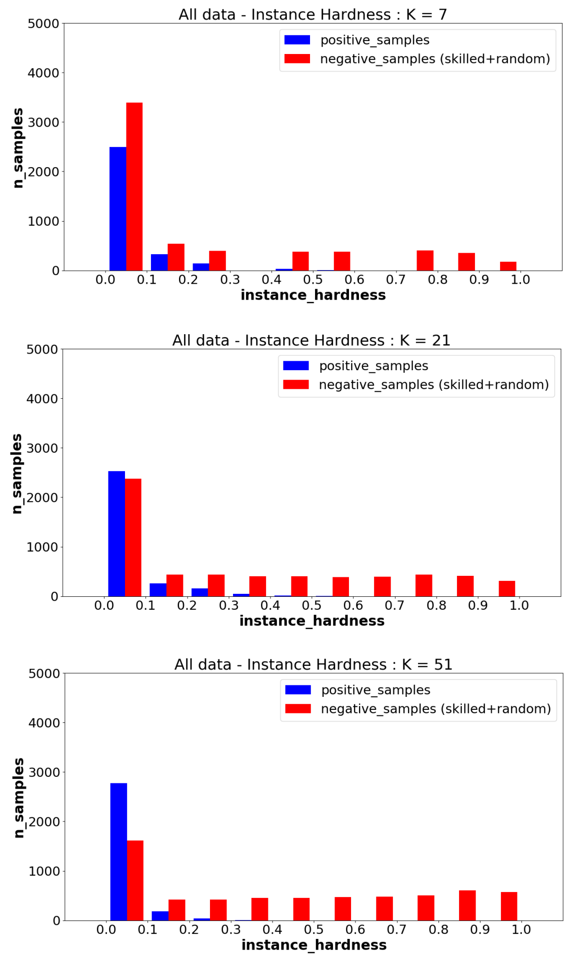

Fig. 5. Scenario (1): Instance hardness considering the all selected data.

Figures 5, 6 and 7 depict, respectively, the histograms of the instance hardness considering the scenarios: (1) all the data; (2) positive samples and negative samples (only from random forgeries); (3) positive samples and negative samples (only from skilled forgeries). With the scenarios (2) and (3), we can see the relationship between each type of negative samples and the positive data.

As depicted in Figure 5, the vast majority of positive samples have $IH = 0.0$ and almost all them have a $IH < 0.3$. As we are considering the kDN measure, a higher values of K result in a more embracing investigation and lower K values represent a more compact investigation of space. To ilustrate this, for $IH < 0.3$ and $K = 7$, at least 5 of the 7 neighbors of the positive samples are from the positive class itself. On the other hand, given $K = 51$, the $IH < 0.3$ represents

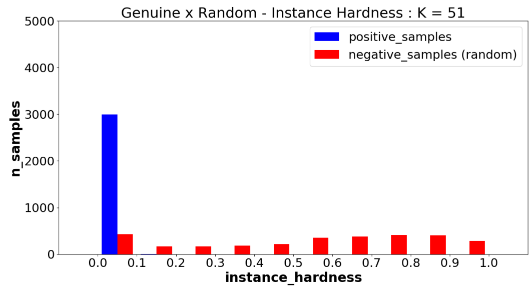

Fig. 6. Scenario (2): IH considering the positive samples and negative samples (only random forgeries).

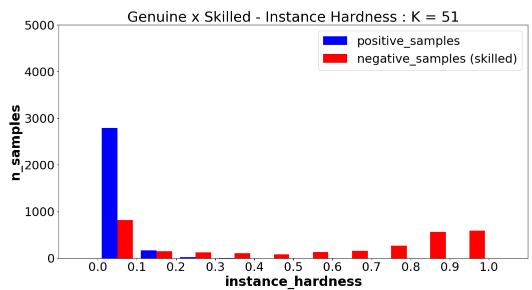

Fig. 7. Scenario (3): IH considering the positive samples and negative samples (only skilled forgeries).

that at least 36 neighbors of the positive samples are from the positive class itself. So, regardless of the neighborhood size, in the dissimilarity space the majority of neighbors of the positive samples are from the positive class. So, positive samples form a dense cluster close to the origin in the dissimilarity space.

For the negative samples, the IH values are spread along the histogram. Thus, negative samples are expected to be in a sparse region of the dissimilarity space with some them in a region closer to the dense positive region of the space, as some samples have IH $= 1.0$ (i.e. all the neighborhood of the negative sample belongs to the positive class).

Still from Figure 5, as a result of the positive samples density and the negative samples sparsity in the dissimilarity space, higher neighborhood sizes (K) results in an increase of negative samples on the right side of the histogram (i.e. some negative samples are in a region closer to the dense positive region of the space than to the negative samples themselves).

When considering the positive samples and the negative samples (only random forgeries), we can observe that (Figure 6): almost all the positive samples are in the IH $= 0.0$ bin, for this to happen there is no class overlap between the class distributions in the dissimilarity space.

Combining Figs. 5 and 7, we can see that almost all the positive samples with $IH \neq 0.0$ from Fig. 5 are derived from skilled forgeries. Thus, there should be class overlapping in the dissimilarity space. This behavior is expected, since the skilled forgeries are more similar to the genuine ones when compared to random forgeries. Still from Fig. 7, as the negative samples include some with higher IH, the overlap of the classes occurs in the positive region of the dissimilarity space.

Thus, in general, positive samples are located in a dense cluster close to the origin and the negative samples are scattered throughout the dissimilarity space. Also, the clusters are disjoint (based on the concentration of the IH with low values) with a small area of overlap (because of the high similarity between genuine signatures and few skilled forgerie). Considering that hard to classify samples are in the border region the use of a condensation PS technique, such as the CNN, was shown to produce good experimental results because it maintains the samples in the decision boundaries [5].

4 Conclusion

In this work we presented the Handwritten Signature Verification problem as having difficult data and tried to understand its behavior in the dissimilarity space generated from the dichotomy transformation used by a writer-independent approach. The evaluation was based on the instance hardness measure.

As presented, the WI dichotomy transformation is able to handle with some of the HSV problem difficulties, such as the imbalanced class distributions, high number of classes, high-dimensional feature space and small number of learning samples, when compared to the writer-dependent approach. Also, the WI approach presented good adaptability to the data, since after training, the classifier can verify signatures regardless of the writer has been used during the training, depending only on the input reference signature.

The reported IH analysis showed that, in general, in the transformed space the positive samples are located in a dense cluster close to the origin and the negative samples are scattered throughout the dissimilarity space generated by the dichotomy transformation. This better understanding of the transformed space, allowed us to make a better use of the samples by using a prototype selection technique, the Condensed Nearest Neighbors (CNN), that is more suited to the worked context.

The experimental results showed that, the dichotomy transformation is able to increase the number of samples in the offline WI-HSV scenario, yet many of them are redundant. Thus, by using prototype selection it is possible to speed up the classifier training and still achieve a similar or better classification performance than by using all the training samples. Even being a classic and simple technique, the Condensed Nearest Neighbors [9] applied as systematic approach was able to select fewer prototypes and still maintain performance when compared to both the SVM trained with all the original training set and the random subsampling approach.

Future works may include the study of feature selection in the dissimilarity space and the adaptation of the WI classifier over new incoming data.

Acknowledgment

This work was supported by the FACEPE (Fundação de Amparo à Ciência e Tecnologia de Pernambuco), CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) and the École de Technologie Supérieure (ÉTS Montréal).

References

1. Bertolini, D., Oliveira, L.S., Sabourin, R.: Multi-script writer identification using dissimilarity. In: Pattern Recognition (ICPR), 2016 23rd International Conference on. pp. 3025–3030. IEEE (2016)
2. Bouamra, W., Djeddi, C., Nini, B., Diaz, M., Siddiqi, I.: Towards the design of an offline signature verifier based on a small number of genuine samples for training. Expert Systems with Applications **107**, 182–195 (2018)
3. Cha, S.H., Srihari, S.N.: Writer identification: statistical analysis and dichotomizer. In: Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR). pp. 123–132. Springer (2000)
4. Eskander, G.S., Sabourin, R., Granger, E.: Hybrid writer-independent–writer-dependent offline signature verification system. IET biometrics **2**(4), 169–181 (2013)
5. Garcia, S., Derrac, J., Cano, J., Herrera, F.: Prototype selection for nearest neighbor classification: Taxonomy and empirical study. IEEE T PATTERN ANAL **34**(3), 417–435 (2012)
6. Hafemann, L.G., Sabourin, R., Oliveira, L.S.: Writer-independent feature learning for offline signature verification using deep convolutional neural networks. In: IEEE IJCNN (2016). pp. 2576–2583. IEEE (2016)
7. Hafemann, L.G., Sabourin, R., Oliveira, L.S.: Learning features for offline handwritten signature verification using deep convolutional neural networks. Pattern Recognition **70**, 163–176 (2017)
8. Hafemann, L.G., Sabourin, R., Oliveira, L.S.: Offline handwritten signature verification—literature review. In: Image Processing Theory, Tools and Applications (IPTA), 2017 Seventh International Conference on. pp. 1–8. IEEE (2017)
9. Hart, P.: The condensed nearest neighbor rule (corresp.). IEEE transactions on information theory **14**(3), 515–516 (1968)
10. Pekalska, E., Duin, R.P., Paclik, P.: Prototype selection for dissimilarity-based classifiers. Pattern Recognition **39**(2), 189–208 (2006)
11. Rivard, D., Granger, E., Sabourin, R.: Multi-feature extraction and selection in writer-independent off-line signature verification. International Journal on Document Analysis and Recognition (IJDAR) **16**(1), 83–103 (2013)
12. Smith, M.R., Martinez, T., Giraud-Carrier, C.: An instance level analysis of data complexity. Machine learning **95**(2), 225–256 (2014)
13. Soleimani, A., Araabi, B.N., Fouladi, K.: Deep multitask metric learning for offline signature verification. Pattern Recognition Letters **80**, 84–90 (2016)
14. Souza, V.L.F., Oliveira, A.L.I., Sabourin, R.: A writer-independent approach for offline signature verification using deep convolutional neural networks features. In: 2018 7th Brazilian Conference on Intelligent Systems (BRACIS). pp. 212–217. IEEE