RankSVM for Offline Signature Verification

Yan Zheng[†], Yuchen Zheng[†*], Wataru Ohyama[‡], Daiki Suehiro[†§], and Seiichi Uchida[†]
[†]*Department of Advanced Information Technology, Kyushu University, Fukuoka, Japan*
Email:{yan.zheng, yuchen}@human.ait.kyushu-u.ac.jp,{suehiro, uchida}@ait.kyushu-u.ac.jp
[‡] *Saitama Institute of Technology, 1690 Fusaiji, Fukaya-shi, Saitama, Japan*
Email:ohyama@sit.ac.jp
[§]*RIKEN, Japan*

***Abstract*—Signature verification systems suffer from imbalanced learning, which imposes strict requirements on classifiers. The standard classification approaches, such as SVM, often degrade the performance for imbalanced data or require additional parameters for data balancing. In this study, as a new approach for signature verification, we use RankSVM as the writer-dependent classifiers, which theoretically guarantees the generalization performance for imbalanced data. To investigate the ability of RankSVM for solving imbalanced learning problems in signature verification tasks, the extensive experiments are conducted on bitmaps of GPDS-150, GPDS-300, GPDS-600, and GPDS-1000 datasets and deep features of GPDS-960 dataset. The experimental results demonstrate that the RankSVM-based approach obtains a nearly equivalent performance with the state-of-the-art method on deep features of the GPDS-960 dataset, and achieves significantly better performance than standard-SVM-based approach on bitmaps of GPDS-150, GPDS-300, GPDS-600, and GPDS-1000 datasets.**

***Keywords*-RankSVM; Offline Signature Verification; Imbalanced Learning Problem.**

I. Introduction

In the fields of machine learning and data science, the problem of imbalanced learning is a relatively new challenge in many actual applications [1], [2]. Imbalanced learning is a scenario where the number of observations in one class is significantly lower than those in other classes. In the field of document analysis and recognition, we often encounter imbalanced problems. Even for simple isolated character recognition tasks, we find imbalanced problems due to the large difference among the prior probabilities of individual alphabets. Especially, a large-class classification problem such as Chinese character recognition, some characters rarely appear and thus cause an imbalance problem.

Another typical imbalanced problem is the signature verification task. In the task, we have to train the classifiers to separate the genuine signatures and forgeries signatures [3]–[7]. Considering the imbalance that much more negative samples (forgeries) than positive samples (genuine signatures) during the training process, the signature verification systems impose strict requirements on classifiers. Therefore, researchers for the task need to search for effective ways to improve the performance of the learned model based on the imbalanced data. To solve this problem, resampling and feature selection methods are employed to balance the training samples [2], [8], [9]. However, these practical remedies are rather heuristic and therefore no theoretical support behind them.

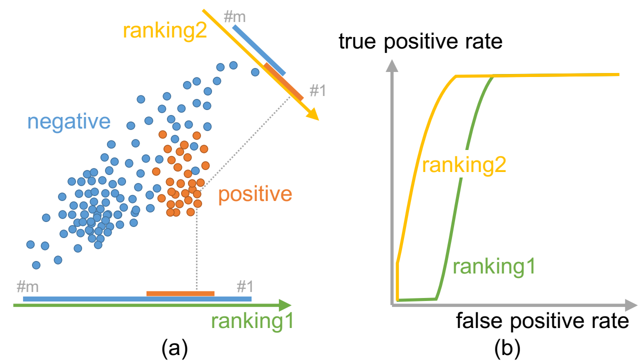

Figure 1: The idea of *learning-to-rank*. (a) $m(= p + n)$ samples from the positive class and the negative class and their two ranking results. (b) ROCs for those ranking results. Since “ranking2” gives a larger AUC, it is the better rank. Note that the subspace for “ranking2” might give a worse overall classification accuracy.

In this paper, we employ the idea of the *learning-to-rank* problem and its algorithm called RankSVM [10], for building the signature verification system robust to the imbalanced problem setting. RankSVM has been used successfully to solve the imbalanced learning problems for information retrieval tasks [11]–[13]. To the authors’ best knowledge, however, RankSVM has not been used for the signature verification task in spite of its usefulness to the task.

Fig. 1 shows the idea of the general learning-to-rank problem. The purpose of the learning-to-rank problem is to find a function that gives higher scores to positive samples and lower scores to negative samples, like the “ranking2” in Fig. 1 (a). By “ranking1” in Fig. 1 (a), the ranks of positive samples become lower than “ranking1”; the RankSVM tries to find a ranking like “ranking2.” If we detect positive samples according to those ranking orders, the corresponding ROC curves become like Fig. 1 (b). Since “ranking2” has more positive samples around the top rank, its ROC curve has a steep rise in its leftmost part. This fact suggests that rankSVM can achieve higher Area-Under-the-Curve (AUC).

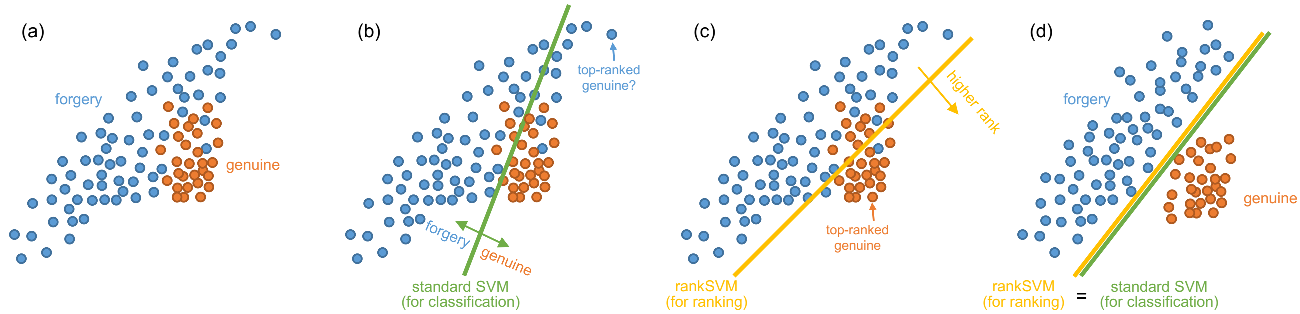

Figure 2: Overview of RankSVM. (a) Sample distributions of the positive class (genuine) and the negative class (forgery). (b) Result of standard SVM for classification. In the worst case, a forgery sample becomes the sample with the largest likelihood as the genuine, although it can minimize the classification error. (c) Result of RankSVM, which can give the best sample order, or equivalently, maximize the AUC. (d) If positive and negative classes are linearly separable, standard SVM and RankSVM give the same result.

RankSVM has several favorable properties for signature verification systems. First, it is proved theoretically that RankSVM maximizes AUC as suggested by Fig. 1(b). Specifically, the optimization problem of RankSVM is equivalent to the AUC maximization problem and thus the optimality of the result is guaranteed. Second, RankSVM provides the rank for each sample. Thereby, positive samples with higher reliability achieve higher ranks. For the signature verification task, this property is very useful because RankSVM assigns a rank to genuine signatures higher based on their reliability. Third, it can handle imbalanced problems without additional parameters. As previously noted, traditional signature verification systems typically encounter imbalanced learning problems.

Our contributions are summarized as follows:

- To the best of our knowledge, this is the first study to utilize RankSVM for the signature verification task. RankSVM possesses favorable properties, especially for the task. For example, rankSVM can deal with the imbalanced problem without any heuristic remedy and can rank the samples according to their reliability.
- We verify the effectiveness of RankSVM for signature verification tasks by conducting several extensive experiments. The desired results were achieved on the imbalanced bitmaps of GPDS-150, GPDS-300, GPDS-600, and GPDS-1000 datasets.

The remainder of this paper is organized as follows: Section II reviews related works on conventional methods used in tackling imbalanced learning in the field of document analysis and recognition. We introduce the RankSVM in detail and explain its superiority in combating the imbalanced learning problem in Section III. Finally, we show the extended experimental results and discussion in Section IV while Section V concludes this paper with remarks.

II. Related Work

In the field of document analysis and recognition, many tasks, such as scene text recognition, handwritten recognition, and signature verification, suffer from imbalanced learning problems [1], [2], [5], [6], [8], [14]. The main reason is that the most standard classifier algorithms, such as decision tree and standard SVM, expect balanced class distributions or equal misclassification costs. With complex imbalanced datasets, these algorithms fail to properly represent the distributive characteristics of the data. Consequently, they provide unfavorable accuracy across the different classes.

To overcome the problems that are brought from imbalanced data, many researches have been proposed and implemented during the preprocessing stage. A commonly used strategy is resampling, which includes undersampling and oversampling techniques [2], [8], [15]. Zhu et al. [8] adopted resampling method to balance foreground and background training samples for a scene text segmentation task. This work empirically validated the importance of data balance. Feature selection, which is mainly utilized for text classification and web categorization domains, is another method that tackles the imbalanced problem [9], [16], [17]. In [16], the authors used the negative class features to discard all documents that are highly associated with these features, which potentially reduces the imbalanced problem. However, the amount of data preprocessing operations significantly increased computation and time costs.

In the field of signature verification, the imbalanced learning problem often occurs in the training process, which makes it difficult to build robust verification systems. Considering real-world scenarios where only reference genuine signatures are available for training, two major approaches are proposed for signature verification. The first one is one-class classification approach [7] which uses one-class SVM (OC-SVM) as the writer-independent classifier. However, the OC-SVM is not effective enough with fewer available handwritten signature samples. The other is the binary classification approach [4]–[6] which uses another version of SVM that tackles the class-imbalance [18] and obtains the state-of-the-art results on signature verification task. However, this SVM requires additional parameters for different classes, and thus we have to tune with more computational cost. Note that additional parameters quadratically increase the computational cost when using grid search. Although there is a simple heuristic to decide the additional parameters, there is no theoretical justification.

To solve the imbalanced learning problem in traditional SVM, Joachims [10] originally proposed RankSVM and applied it to web search engines. RankSVM solves AUC maximization based on the formulation of standard SVM (the details are shown in Section III) without any additional parameters to SVM. RankSVM is justified by the theoretical results in machine learning field (e.g., [19]), which says that RankSVM achieves as high AUC score for test samples (i.e., unseen data) as training samples. Because of the strong theoretical support, RankSVM has been applied to various domains such as document retrieval [10]–[13] and image classification [20], [21]. The theoretical and empirical success of RankSVM motivates us to apply it to signature verification task. To the best of our knowledge, RankSVM has never been used in the signature verification task.

III. RankSVM

In this study, we utilize RankSVM [10] for the signature verification task because of its favorable properties (already noted in Section I) for the task. In contrast to the standard SVM, RankSVM aims to learn a real-valued ranking function that gives the positive samples higher values than negative samples as much as possible, just as shown in Fig. 2. With an imbalanced distribution of the positive class and negative class in Fig. 2 (a), the standard SVM will build a hyperplane to classify the genuine samples and forgeries. In Fig. 2 (b), the standard SVM does not care about the rank scores of the features. In the worst case, a forgery sample becomes the sample with the largest likelihood as the genuine samples.

Compared to the standard SVM, RankSVM can give the best pattern order by maximizing AUC. For the distribution of Fig. 2 (a), RankSVM will give the result like Fig. 2 (c), where the most positive samples achieve higher ranks. It is theoretical proved in [10] that AUC maximization is equivalent to guarantee the generalization performance for imbalanced data and therefore RankSVM guarantees it. Note that if the original data are linearly separable in some feature spaces, the RankSVM and standard SVM will obtain the similar classification boundaries, as shown in Fig. 2 (d).

Let us start from the optimization problem of the standard SVM. If the notation $\mathbf{x}_1^+, \ldots, \mathbf{x}_p^+$ and $\mathbf{x}_1^-, \ldots, \mathbf{x}_n^-$ for positive and negative samples, respectively, is used for the later discussion, the standard SVM is formulated as follows:

$$\begin{aligned} &\min_{\mathbf{w}, b, \boldsymbol{\xi}^+, \boldsymbol{\xi}^-} \frac{1}{2}\|\mathbf{w}\|^2 + C\left(\sum_{i=1}^{p}\xi_i^+ + \sum_{j=1}^{n}\xi_j^-\right) \\ &\text{s.t.} \quad (\mathbf{w}^T\mathbf{x}_j^+ + b \geq 1 - \xi_i^+) \wedge (\xi_i^+ \geq 0),\ i \in [1, p] \\ &\quad (-(\mathbf{w}^T\mathbf{x}_j^- + b) \geq 1 - \xi_j^-) \wedge (\xi_j^- \geq 0),\ j \in [1, n], \end{aligned} \tag{1}$$

where ξ_i^+ and ξ_j^- are slack variables corresponding to $\mathbf{x}_i^+$ and $\mathbf{x}_j^-$, respectively.

The key idea of RankSVM, which differentiates RankSVM from the standard SVM is to consider that $\mathbf{w}$ is a ranking function and the rank of a positive sample, i.e., $\mathbf{w}^T\mathbf{x}_i^+$, is better to be larger than that of a negative sample, i.e., $\mathbf{w}^T\mathbf{x}_j^-$. Namely, the number of pairs which satisfy $\mathbf{w}^T(\mathbf{x}_i^+ - \mathbf{x}_j^-) > 0$ should be maximized. Consequently, this rank optimization is equivalent to AUC maximization because AUC is represented as follows:

$$\text{AUC} = \sum_{i=1}^{p}\sum_{j=1}^{n}\frac{I(\mathbf{w}^T\mathbf{x}_i^+ > \mathbf{w}^T\mathbf{x}_j^-)}{pn},$$

where $I(\cdot)$ denotes the indicator function. It should be emphasized that AUC is a common criterion for imbalanced problems (such as information retrieval problems and target detection problems) and thus RankSVM is robust to the imbalance problems.

The optimization problem of RankSVM is formulated as

$$\begin{aligned} &\min_{\mathbf{w}, \boldsymbol{\xi}} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{p}\sum_{j=1}^{n}\xi_{ij} \\ &\text{s.t.} \quad (\mathbf{w}^T(\mathbf{x}_i^+ - \mathbf{x}_j^-) \geq 1 - \xi_{ij}) \wedge (\xi_{ij} \geq 0), \\ &\qquad \forall i \in [1, p], \forall j \in [1, n], \end{aligned} \tag{2}$$

where ξ_{ij} is a slack variable for the pair of $\mathbf{x}_i^+$ and $\mathbf{x}_j^-$. The first inequalities represent the soft margin constraints so that the rank difference between $\mathbf{x}_i^+$ and $\mathbf{x}_j^-$, i,e., $\mathbf{w}^T(\mathbf{x}_i^+ - \mathbf{x}_j^-)$, becomes positive.

In the following experiment, we will use a simple linear kernel for the standard SVM and RankSVM, although RankSVM can employ non-linear mapping by using kernel methods. This is because we use features by a deep CNN for representing signature images and thus their feature spaces are already non-linearly transformed from their original image feature space.

IV. Experiments

In this section, we describe the implementation of the experiments and the experimental results in detail. The comparative experiments for RankSVM [12] and the standard SVM were conducted on off-line signature images from GPDS-150, GPDS-300, GPDS-600, and GPDS-1000 datasets, as well as the feature vector extracted by a deep CNN [6] on GPDS-960 dataset. The codes of RankSVM and

Figure 3: Signature examples from GPDS dataset.

the standard SVM that are used in this study were obtained from the website[1].

We conducted the experiments considering two scenarios, evaluation of signature verification performance against skilled forgery attacks and analyzing the behavior of RankSVM and the standard SVM against large-scale random imposter attacks. In the latter experiments, we employed a signature retrieval task to inspect behavioral differences of the RankSVM and standard SVM.

A. Experiments in Signature Verification for Skilled Forgery Attacks

1) Experimental setup: We applied RankSVM to the signature verification for skilled forgery attacks, which is a typical imbalanced problem, and compared the performance of RankSVM to the standard SVM. Its objective is to learn a model that can distinguish between genuine signatures and skilled forgeries. In the experiments, a training set for each specific user was built comprising genuine signatures as positive samples and registered genuine signatures from third-persons as negative samples. As the test samples, we used skilled forgeries, which are signatures for imitating the genuine signatures by other persons.

To enable a fair comparison with RankSVM, we used an implementation of SVM that cares the class-imbalance by weighting different costs for the misclassification of different classes [18]. In this implementation, the standard SVM of (1) was modified as follows:

$$\begin{aligned}\min_{\mathbf{w},b,\boldsymbol{\xi}^+,\boldsymbol{\xi}^-} & \frac{1}{2}\|\mathbf{w}\|^2 + C^+\sum_{i=1}^{p}\xi_i^+ + C^-\sum_{j=1}^{n}\xi_j^- \\ \text{s.t.} \quad & (\mathbf{w}^T\mathbf{x}_j^+ + b \geq 1 - \xi_i^+) \wedge (\xi_i^+ \geq 0),\ i \in [1,p] \\ & (-(\mathbf{w}^T\mathbf{x}_j^- + b) \geq 1 - \xi_j^-) \wedge (\xi_j^- \geq 0),\ j \in [1,n]. \end{aligned} \tag{3}$$

We set hyperparameters as $C^+ = \frac{n}{p}C^-$ and $C^- = 1$ according to the state-of-the-art signature verification methods [4]–[6]. For RankSVM, we simply set $C = 1$.

2) Dataset: The experiments were conducted by using the offline signature dataset called "GPDS synthetic online offline signature database" [22]. It contains data from 10,000 synthetic individuals: 24 genuine signatures for each individual and 30 skilled forgeries for their signatures. Fig. 3 shows the samples of genuine signatures and their corresponding

[1] https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/

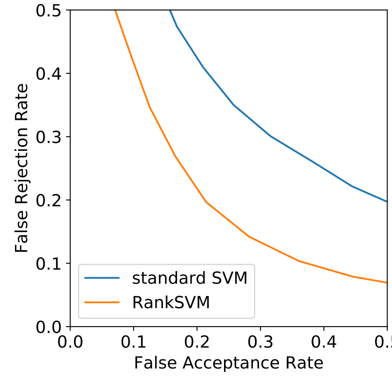

Figure 4: The DET curve of RankSVM and the standard SVM with bitmap on GPDS-150 dataset.

skilled forgeries. In the experiments, 64×64 pixels of bitmaps were used as inputs while the pixel values were normalized to [0,1]. Aside from this normalization step, no other preprocessing was applied.

In experiments, for each user, we built a training set consisting of 12 genuine signatures from the specific user as positive samples and 12 genuine signatures from other users as negative samples and a testing set consisting of 10 randomly selected genuine signatures from the specific user as positive samples and 10 randomly selected skilled forgeries from other users as negative samples with respect to GPDS-150, GPDS-300, GPDS-600, and GPDS-1000 datasets.

We created 10 random partitions of the training set and the testing set and repeated the same experiments for statistical evaluation.

3) Evaluation metrics: In signature verification, False Rejection Rate (FRR), False Acceptance Rate (FAR), Equal Error Rate (EER) and Detection Error Tradeoff (DET) curve are used to evaluate the performance of RankSVM and the standard SVM. FRR is the fraction of genuine signatures rejected as forgeries, FAR is the fraction of forgeries accepted as genuine signatures, EER is the error when FRR=FAR, and the DET curve is a graphical plot of the error rates for binary classification tasks, which plots FRR versus FAR. A lower EER implies a better performance for the algorithm.

4) Experimental result with bitmap: The average EER and standard deviation obtained by RankSVM and the standard SVM were 19.56 ± 0.8620 and 30.14 ± 0.8142, respectively. The DET curve for a random partition of users is shown in Fig. 4. From this curve, we can see that the RankSVM performs significantly better than the standard SVM on signature verification task, and the EER of RankSVM is obviously lower than that of the standard SVM.

5) Experimental result with deep feature: We also conducted the experiment for RankSVM with the Convolutional Neural Network (SIGNET) based deep features [6]. In [6], the authors obtained state-of-the-art results by using standard SVM as classifier. The deep features on GPDS-960 [6] were directly used to train and test the RankSVM and standard SVM models. The training and testing processes were same as in [6]. For first 300 users of GPDS-960, RankSVM

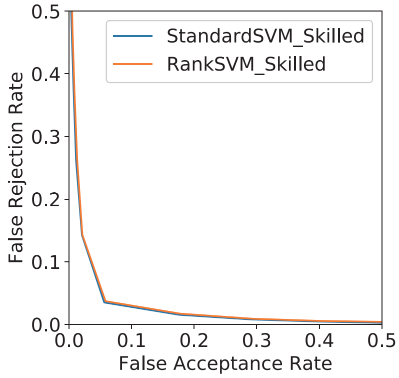

Figure 5: The DET curve of RankSVM and the standard SVM with deep features on GPDS-960 dataset.

obtained a nearly similar performance with standard SVM with an EER of 4.1% for RankSVM and 3.9% for standard SVM. The DET curve are shown in Fig. 5.

We would like to consider the reason that we could not see a significant difference between RankSVM and the standard SVM when using deep features. A possible reason is that our RankSVM obtained highly similar hyperplane as the standard SVM as shown in the case of Fig. 2 (d). In other words, if the hyperplanes of RankSVM and SVM are highly similar, the deep feature space is possibly linearly separable.

To ensure the above reason, we calculated the cosine similarity[2] between the hyperplanes by RankSVM and the standard SVM on the bitmap space (for GPDS-150 dataset), and also calculated the cosine similarity between the hyperplanes by RankSVM and the standard SVM on the deep feature space (for GPDS-960 dataset). We could see the significant difference between the former average cosine similarity ($= 0.776$) and the latter ($= 0.866$). This result implies that RankSVM and the standard SVM find highly similar hyperplane over the deep feature space, and thus the deep feature space is much more easily linear-separable than bitmap space.

Consequently, the suggestion by these observations is as follows: RankSVM is more suitable for tough tasks where the distributions of the positive and the negative samples are heavily overlapping, rather than for easier tasks where these samples are already almost separable by a simple classifier like the standard SVM.

B. Experiments in Signature Retrieval

1) Experimental setup: We observe the behavior of RankSVM and the standard SVM through a signature *retrieval* task, where we need to deal with largely-imbalanced data. We used the same training set as in the verification task and built a new testing set consisting of 12 genuine signatures from the user as positive samples and 12 genuine signatures from other users as negative samples. The experiments were conducted by using the datasets of GPDS-150, GPDS-300, GPDS-600, and GPDS-1000 consisting of the first 150, 300, 600, and 1,000 users, so that the imbalance between the positive and the negative samples is increased

[2]We used the cosine similarity because the bias of the hyperplane of SVM can be ignored in ranking tasks.

Table I: The mean AUC of RankSVM and Standard SVM on different datasets.

Dataset	RankSVM	standard SVM
GPDS-150	**0.970**	0.930
GPDS-300	**0.965**	0.911
GPDS-600	**0.955**	0.895
GPDS-1000	**0.952**	0.886

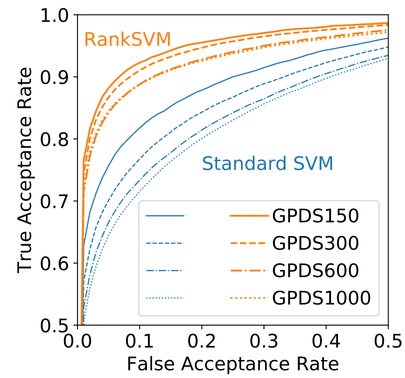

Figure 6: The ROC curve of RankSVM and standard SVM on GPDS-150, GPDS-300, GPDS-600, GPDS-1000 datasets.

with the number. The signature retrieval experiments with RankSVM and the standard SVM were performed multiple times by using randomly selected different training and test datasets to obtain statistical performance.

2) Experimental result: quantitative evaluation: The mean AUC of each dataset for 10 random partitions are shown in Table I. Fig. 6 shows the average ROC curves created for users among these datasets. Table I and Fig. 6 indicate that RankSVM significantly outperformed the standard SVM in the retrieval task. Although the dataset becomes more imbalanced with the increasing number of users, the performance degradation of RankSVM is much less than the standard SVM. This is because the standard SVM is not stable on imbalanced data, whereas RankSVM keeps high AUC performance even for highly imbalanced datasets.

3) Experimental result: qualitative evaluation: Fig. 7 shows the top-5 retrieval results on GPDS-150 dataset. From this figure, we can confirm the following facts:

1) RankSVM could get genuine samples as the top-ranked samples whereas the standard SVM often fails to get them.
2) Especially, it should be emphasized that the top-ranked samples are very similar to the query sample. This indicates that RankSVM could give the most reliable genuine samples among all genuine samples.
3) It also should be emphasized that the standard SVM ranks a very different signature at the top. This is caused by the situation of Fig. 2 (b). In contrast, RankSVM gives more similar signatures to the input.

V. Conclusion

The signature verification task is a typical imbalanced problem and thus classifiers for the task are better to be data imbalance. The past trials with the standard classification approaches, such as SVM, often employ a simple and heuristic remedy to balance the problem by using

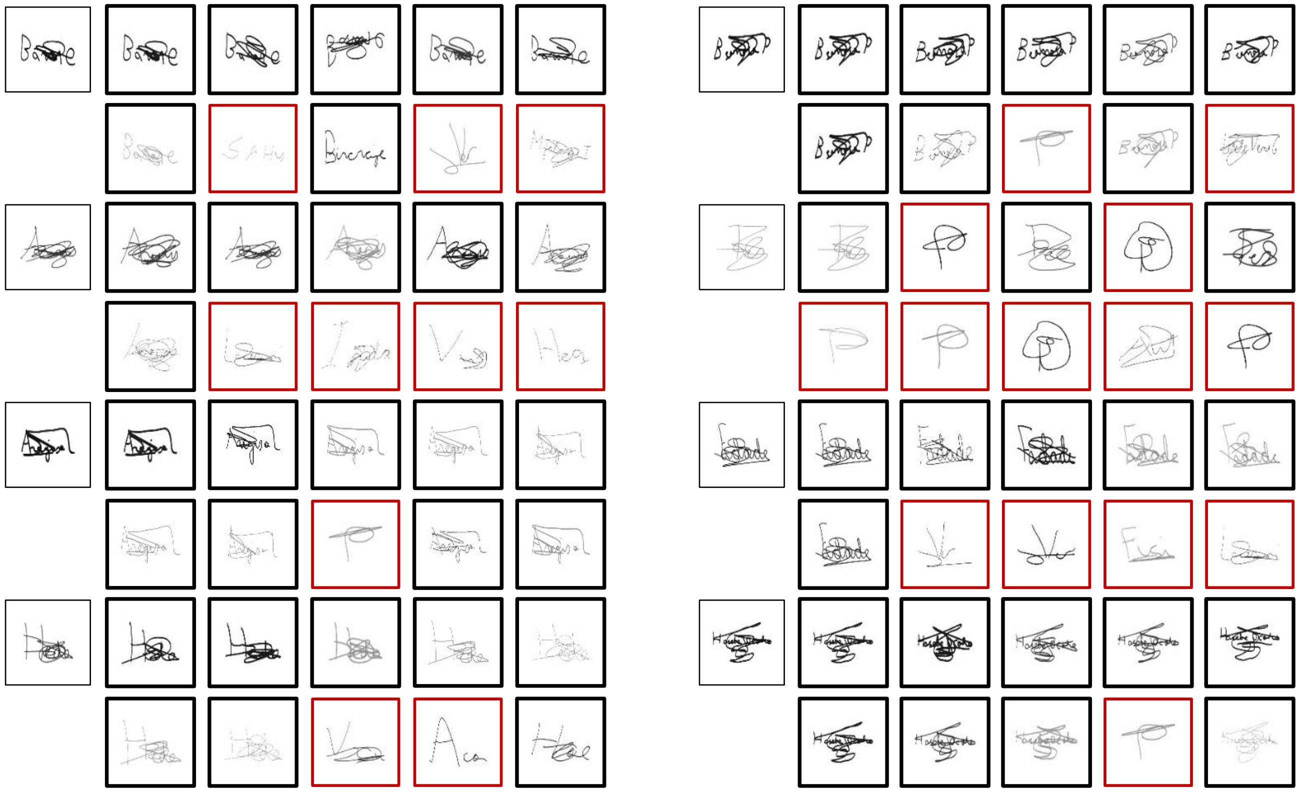

Figure 7: Top-5 signatures for RankSVM and Standard SVM on GPDS-150 dataset. Each two rows denote one user's ranked signatures; the top and bottom rows show the top-5 results of RankSVM and the standard SVM, respectively. Images in the black box and the red box denote the correct genuine signatures and the wrong signatures, respectively.

weight parameters. In this study, we applied RankSVM in the signature verification task. RankSVM has several good properties for the task. Especially, it theoretically guarantees the robustness to imbalanced tasks without any parameters. The extensive experiments on signature verification and retrieval tasks confirmed that RankSVM can achieve significantly higher performance than the standard SVM for both signature retrieval and signature verification tasks, especially in a tough situation where the positive and the negative classes are heavily overlapping.

VI. Acknowledgments

This work was supported by JSPS KAKENHI (Grant Number JP18K11373, JP17H06100 and JP18K18001) and China Scholarship Council (Grant Number 201806330079).

References

[1] R. Longadge and S. Dongre, "Class Imbalance Problem in Data Mining Review," *arXiv:1305.1707*, 2013.

[2] T. Wang *et al.*, "Radical-based Chinese Character Recognition via Multi-labeled Learning of Deep Residual Networks," in *ICDAR*, 2017.

[3] S. Srihari *et al.*, "Learning Strategies and Classification Methods for Off-line Signature Verification," in *ICFHR*, 2004.

[4] L. Hafemann *et al.*, "Analyzing Features Learned for Offline Signature Verification Using Deep CNNs," in *ICPR*, 2016.

[5] L. Hafemann *et al.*, "Learning Features for Offline Handwritten Signature Verification Using Deep Convolutional Neural Networks," *Pattern Recognition*, 2017.

[6] G. L. L. Hafemann *et al.*, "Fixed-sized Representation Learning from Offline Handwritten Signatures of Different Sizes," *IJDAR*, 2018.

[7] Y. Guerbai *et al.*, "The effective use of the one-class svm classifier for handwritten signature verification based on writer-independent parameters," *Pattern Recognition*, 2015.

[8] S. Zhu and R. Zanibbi, "A Text Detection System for Natural Scenes with Convolutional Feature Learning and Cascaded Classification," in *CVPR*, 2016.

[9] G. Forman, "An Extensive Empirical Study of Feature Selection Metrics for Text Classification," *JMLR*, 2003.

[10] T. Joachims, "Optimizing Search Engines Using Clickthrough Data," in *SIGKDD*, 2002.

[11] Y. Cao *et al.*, "Adapting Ranking SVM to Document Retrieval," in *SIGIR*, 2006.

[12] C. Lee and C. Lin, "Large-scale Linear Ranksvm," *Neural Computation*, 2014.

[13] R. Cruz *et al.*, "Tackling Class Imbalance with Ranking," in *IJCNN*, 2016.

[14] C. Liu and K. Marukawa, "Pseudo Two-dimensional Shape Normalization Methods for Handwritten Chinese Character Recognition," *Pattern Recognition*, 2005.

[15] A. Liu, "The Effect of Oversampling and Undersampling on Classifying Imbalanced Text Datasets," *Master thesis in The University of Texas at Austin*, 2004.

[16] Z. Zheng *et al.*, "Feature Selection for Text Categorization on Imbalanced Data," *SIGKDD Explorations*, 2004.

[17] D. Mladenic and M. Grobelnik, "Feature Selection for Unbalanced Class Distribution and Naive Bayes," in *ICML*, 1999.

[18] R. Akbani *et al.*, "Applying Support Vector Machines to Imbalanced Datasets," in *ECML*, 2004.

[19] M. Mohri *et al.*, *Foundations of Machine Learning*, 2018.

[20] W. ElisseeffA, "Kernel methods for Multi-labelled Classification and Categorical Regression Problems," *NIPS*, 2002.

[21] R. Cruz *et al.*, "Ordinal Class Imbalance with Ranking," in *IbPRIA*, 2017.

[22] M. Ferrer *et al.*, "A Behavioral Handwriting Model for Static and Dynamic Signature Synthesis," *TPAMI*, 2017.