21st International Conference on Pattern Recognition (ICPR 2012)
November 11-15, 2012. Tsukuba, Japan

A Discriminative Linear Regression Approach to OCR Adaptation

Jun Du and Qiang Huo
Microsoft Research Asia, Beijing, P. R. China
{jundu,qianghuo}@microsoft.com

Abstract

This paper presents a new discriminative linear regression approach to adaptation of a discriminatively trained prototype-based classifier for Chinese OCR. A so-called sample separation margin based minimum classification error criterion is used in both classifier training and adaptation, while an Rprop algorithm is used for optimizing the objective function. Formulations for both model-space and feature-space adaptation are presented. The effectiveness of the proposed approach is confirmed by experiments for adaptation of font styles and low-quality text, respectively.

1. Introduction

With the fast development of mobile internet, OCR-based applications are becoming increasingly more popular (e.g., [2, 1, 3]). However, most off-the-shelf OCR engines were trained on scanned documents, and they may not work well for new application scenarios where the properties of the captured character images are significantly different from the ones in the training data set. One of solutions to address this problem is to adapt a pre-trained classifier to deal with the new scenario by using the document to be recognized itself via an unsupervised adaptation strategy (e.g., [9, 8]), or by using a small amount of adaptation data collected in the target scenario via a supervised adaptation strategy. The latter is the topic of this study.

In this paper, we study the adaptation techniques for Chinese OCR. One of the state-of-the-art techniques to build a Chinese OCR engine is to use a discriminatively trained prototype-based classifier as reported in [5]. Recently, a so-called sample separation margin (SSM) based MCE training approach was proposed in [4] for training prototype-based classifiers, which performs better than the MCE training approach in [5]. In this study, we have built our baseline classifier for Chinese OCR by using the techniques described in [5, 4, 10] with a small difference: we used an Rprop algorithm (e.g., [6]) to optimize the SSM-MCE objective function because the setting of control parameters is much easier than the Quickprop algorithm used in [10]. The main contribution of this paper is to propose a new SSM-MCE linear regression (LR) approach to adaptation of an SSM-MCE trained prototype-based classifier for Chinese OCR. Formulations for both model-space and feature-space adaptation are presented. Our work is related to a recent work on writer adaptation for handwritten Chinese character recognition reported in [11], where a similar MCE training approach as in [5] is used to train a prototype-based classifier, but a least regularized weighted squared error approach is used to estimate a global feature transform (a.k.a. style transfer matrix (STM)) for writer adaptation. The adaptation capability of the STM approach is similar to our feature-space adaptation approach, but is inferior to our model-space adaptation approach because multiple transforms can be used for model adaptation. Even for feature-space approach, our experimental results show that our approach performs significantly better than the original STM approach in [11] for supervised adaptation of font styles and low-quality text, respectively, which confirms that SSM-MCE is a better objective function to learn the feature transform.

The remainder of the paper is organized as follows. In Section 2, we describe briefly how to construct a multi-prototype based classifier by using the SSM-MCE training. In Section 3, we present formulations of SSM-MCE LR for both model-space and feature-space adaptation. Several important implementation issues are discussed in Section 4. In Section 5, we report experimental results for supervised adaptation of font styles and low-quality text, respectively. Finally we conclude the paper in Section 6.

2. SSM-MCE Training

Suppose our classifier can recognize M character classes denoted as $\{C_i | i = 1, ..., M\}$. For a multi-

978-4-9906441-0-9 ©2012 ICPR

prototype based classifier, each class C_i is represented by K_i prototypes, $\boldsymbol{\lambda}_i = \{\boldsymbol{m}_{ik} \in \mathcal{R}^D | k = 1, ..., K_i\}$, where $\boldsymbol{m}_{ik}$ is the k^{th} prototype of the i^{th} class. Let's use $\boldsymbol{\Lambda} = \{\boldsymbol{\lambda}_i\}$ to denote the set of prototypes. In the classification stage, a feature vector $\boldsymbol{x} \in \mathcal{R}^D$ is first extracted. Then $\boldsymbol{x}$ is compared with each of the M classes by evaluating a Euclidean distance based discriminant function for each class C_i as follows

$$g_i(\boldsymbol{x}; \boldsymbol{\lambda}_i) = -\min_k \| \boldsymbol{x} - \boldsymbol{m}_{ik} \|^2 \ . \tag{1}$$

The class with the maximum discriminant function score is chosen as the recognized class $r(\boldsymbol{x}; \boldsymbol{\Lambda})$, i.e.,

$$r(\boldsymbol{x}; \boldsymbol{\Lambda}) = \arg\max_i g_i(\boldsymbol{x}; \boldsymbol{\lambda}_i) \ . \tag{2}$$

In the training stage, given a set of training data $\boldsymbol{X} = \{\boldsymbol{x}_r \in \mathcal{R}^D | r = 1, ..., R_1\}$, first we initialize $\boldsymbol{\Lambda}$ by LBG clustering. Then $\boldsymbol{\Lambda}$ can be re-estimated by minimizing the following MCE objective function:

$$l(\boldsymbol{X}; \boldsymbol{\Lambda}) = \frac{1}{R_1} \sum_{r=1}^{R_1} \frac{1}{1 + \exp[-\alpha d(\boldsymbol{x}_r; \boldsymbol{\Lambda}) + \beta]} \tag{3}$$

where α, β are two control parameters, and $d(\boldsymbol{x}_r; \boldsymbol{\Lambda})$ is a misclassification measure defined by using a so-called sample separation margin (SSM) as follows [4]:

$$d(\boldsymbol{x}_r; \boldsymbol{\Lambda}) = \frac{-g_p(\boldsymbol{x}_r; \boldsymbol{\lambda}_p) + g_q(\boldsymbol{x}_r; \boldsymbol{\lambda}_q)}{2 \| \boldsymbol{m}_{p\hat{k}} - \boldsymbol{m}_{q\bar{k}} \|} \tag{4}$$

where

$$\hat{k} = \arg\min_k \| \boldsymbol{x}_r - \boldsymbol{m}_{pk} \|^2 \tag{5}$$

$$q = \arg\max_{i \in \mathcal{M}_r} g_i(\boldsymbol{x}_r; \boldsymbol{\lambda}_i) \tag{6}$$

$$\bar{k} = \arg\min_k \| \boldsymbol{x}_r - \boldsymbol{m}_{qk} \|^2 \tag{7}$$

and $\mathcal{M}_r$ is the hypothesis space for the r^{th} sample, excluding the true label p.

To optimize the objective function, in [10], a modified Quickprop procedure is used. In this work, an Rprop algorithm described in [6] is adopted. Due to the space limitation, we will report the detailed Rprop procedures (including those for the following feature-space and model-space linear regression) elsewhere.

3. SSM-MCE Linear Regression

To adapt an OCR engine, we can adapt the classifier to the new scenario (i.e., model-space method) or adapt the observed features in the new scenario back to original feature space (i.e., feature-space method).

3.1. Model-Space Method

Suppose we are given a set of labeled adaptation data $\boldsymbol{Y} = \{\boldsymbol{y}_r \in \mathcal{R}^D | r = 1, ..., R_2\}$ collected in the target application scenario. For model-space method, we transform the parameters of the original classifier as follows:

$$\hat{\boldsymbol{m}}_{ik} = \mathcal{F}(\boldsymbol{m}_{ik}; \boldsymbol{\Theta}) = \boldsymbol{A}_{e_i} \boldsymbol{m}_{ik} + \boldsymbol{b}_{e_i} \tag{8}$$

where i and k are indices of class and prototype, respectively; and e_i is the transform index for the i^{th} class. Let's use $\boldsymbol{\Theta} = \{(\boldsymbol{A}_e, \boldsymbol{b}_e) | e = 1, ..., E\}$ to denote the set of transform parameters, where $\boldsymbol{A}_e$ is a $D \times D$ nonsingular matrix and $\boldsymbol{b}_e$ is a D-dimensional bias vector. The SSM-MCE objective function is defined as follows:

$$l(\boldsymbol{Y}; \boldsymbol{\Lambda}, \boldsymbol{\Theta}) = \frac{1}{R_2} \sum_{r=1}^{R_2} \frac{1}{1 + \exp[-\alpha d(\boldsymbol{y}_r; \boldsymbol{\Lambda}, \boldsymbol{\Theta}) + \beta]} \tag{9}$$

where

$$d(\boldsymbol{y}_r; \boldsymbol{\Lambda}, \boldsymbol{\Theta}) = \frac{-g_p(\boldsymbol{y}_r; \hat{\boldsymbol{\lambda}}_p) + g_q(\boldsymbol{y}_r; \hat{\boldsymbol{\lambda}}_q)}{2 \| \hat{\boldsymbol{m}}_{p\hat{k}} - \hat{\boldsymbol{m}}_{q\bar{k}} \|} \ . \tag{10}$$

The multiple linear regression transforms are tied across character classes, where each transform is associated with a set of character classes. To design a fully automatic adaptation procedure for any given amount of labeled adaptation data, we use a regression class tree to group the character classes, just like what has been done in MLLR [7].

3.2. Feature-Space Method

For feature-space method, the following global feature transformation function is used:

$$\boldsymbol{x}_r = \mathcal{F}(\boldsymbol{y}_r; \boldsymbol{\Theta}) = \boldsymbol{A}\boldsymbol{y}_r + \boldsymbol{b} \tag{11}$$

where $\boldsymbol{A}$ is a $D \times D$ nonsingular matrix, $\boldsymbol{b}$ is a D-dimensional bias vector, $\boldsymbol{y}_r$ and $\boldsymbol{x}_r$ are the r^{th} D-dimensional input and transformed feature vectors, respectively. The SSM-MCE objective function is defined as follows:

$$l(\boldsymbol{Y}; \boldsymbol{\Lambda}, \boldsymbol{\Theta}) = \frac{1}{R_2} \sum_{r=1}^{R_2} \frac{1}{1 + \exp[-\alpha d(\boldsymbol{y}_r; \boldsymbol{\Lambda}, \boldsymbol{\Theta}) + \beta]} \tag{12}$$

where

$$d(\boldsymbol{y}_r; \boldsymbol{\Lambda}, \boldsymbol{\Theta}) = \frac{-g_p(\boldsymbol{x}_r; \boldsymbol{\lambda}_p) + g_q(\boldsymbol{x}_r; \boldsymbol{\lambda}_q)}{2 \| \boldsymbol{m}_{p\hat{k}} - \boldsymbol{m}_{q\bar{k}} \|} \ . \tag{13}$$

In recognition stage, the estimated transform $\{\boldsymbol{A}, \boldsymbol{b}\}$ is used to transform the feature vector of each unknown character first, which is then fed to baseline classifier for recognition.

4. Implementation Issues

For notational convenience, we refer to hereinafter our SSM-MCE based feature-space and model-space LR approaches as F-DLR (Feature-space Discriminative Linear Regression) and M-DLR (Model-space Discriminative Linear Regression), respectively.

4.1. STM-based Initialization

In our F-DLR approach, we initialize the bias vector as $\boldsymbol{b} = 0$, and use the STM approach in [11] to initialize the $\boldsymbol{A}$ matrix. In our M-DLR approach, we initialize the bias vector as $\boldsymbol{b}_{e_i} = 0$, and use a modified STM approach (just exchanging the role of source points and target points in the original STM formulation), to initialize the $\boldsymbol{A}_{e_i}$ matrix separately for each regression class e_i by using the adaptation samples of the corresponding regression class.

4.2. A Hybrid Adaptation Approach

To achieve the best possible adaptation effect for different amount of adaptation data, we propose to use the following hybrid adaptation approach:

- If the amount of adaptation data is very small, i.e., $R_2 \le N_T$, use STM approach in [11] with an adaptive hyperparameter $\beta_1^{\text{new}} = \beta_1 \frac{N_T}{R_2}$;
- If more adaptation data is available but not enough to estimate multiple transforms in M-DLR, i.e., $N_T < R_2 \le N_M$, use single-transform based F-DLR or M-DLR approach;
- If enough adaptation data is available, i.e. $R_2 > N_M$, use multi-transform based M-DLR approach.

Two control parameters N_T and N_M are set empirically to $\frac{D^2}{16}$ and $2D^2$, respectively.

In the following experiments, we will show that the adaptive STM can outperform significantly the original STM for the case of very limited adaptation data because the control parameter β_1 is adjusted dynamically according to the amount of available adaptation data.

5. Experiments and Results

5.1. Experimental Setup

The experiments are conducted on a task of recognizing isolated printed Chinese characters. The vocabulary of our baseline classifier consists of 9,252 Chinese characters. For SSM-MCE training of the baseline classifier, we use about 150 image samples per character. These image samples are mostly from scanned documents with several commonly used fonts. 512-dimensional Gabor features with intensity normalization plus an aspect ratio feature are extracted from each gray-scale character image to form a raw feature vector, which is followed by LDA transformation to obtain a 128-dimensional feature vector (i.e., $D = 128$) [5]. As for the number of prototypes for each character, we use 4 prototypes for 3,755 most frequently used Chinese characters and 2 prototypes for the rest of character classes. The control parameters of SSM-MCE objective function are set as follows: $\alpha = 7$; $\beta = 0$. The other control parameters related to Rprop are set empirically as suggested in [6] without tuning.

5.2. Adaptation to Font Style

The first set of experiments is designed to examine the effectiveness of the proposed approach for supervised adaptation to font style. We use 25 sets of new font library for experiments. For each font library, there are 6,823 character classes and one sample per character class. We divide 6,823 samples per font into two equal subsets as adaptation set and testing set. In this case, our hybrid adaptation approach in Section 4.2 has used the DLR solution.

Fig. 1 summarizes a performance (character recognition error rate in %) comparison of the baseline classifier and different approaches for supervised adaptation to each font style on testing sets of 25 new font styles. Several observations can be made. First, all methods for supervised adaptation outperform the baseline classifier without adaptation, which demonstrates that a linear transformation is reasonable as a mapping function for font adaptation. Second, both F-DLR and M-DLR achieve consistently significant improvements in recognition accuracy compared to STM, which indicates that the SSM-MCE objective function of DLR is indeed better than the least weighted squared error criterion used in STM. Third, M-DLR performs much better than F-DLR.

5.3. Adaptation to Low-Quality Text

The second set of experiments is designed to examine the effectiveness of the proposed approach for supervised adaptation to low-quality text. We use a database of low-quality character images captured by a camera with a resolution of 640×480 pixels. There are 7,915 character classes with dozens of samples per character class. First, 15 samples per character are randomly selected from the database to form the testing set. The

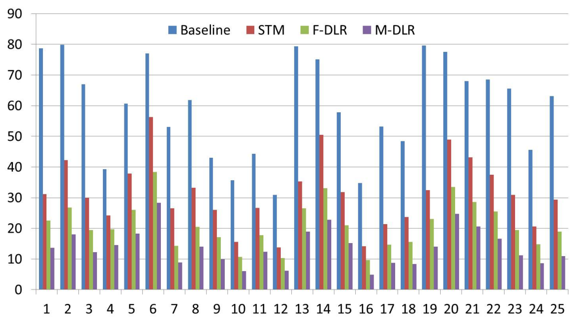

Figure 1. Performance (character recognition error rate in % on each testing set) comparison of the baseline classifier and different approaches for supervised adaptation to each of 25 new font styles.

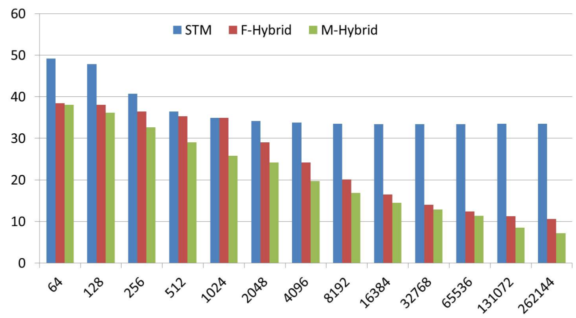

Figure 2. Performance (character recognition error rate in % on testing set) comparison of different approaches for supervised adaptation with different number of adaptation samples of low-quality text (Baseline recognition error rate is 46.98%).

remaining samples are used for adaptation with different amount of data. The character recognition error rate of the baseline classifier on testing set is 46.98%. Fig. 2 compares the performance of STM and hybrid adaptation approach in Section 4.2 on testing set. If adaptation data is very limited ($R_2 < 256$), the performance of STM is even worse than that of the baseline classifier. The performance improvement for STM saturates beyond a certain point ($R_2 > 1024$). As expected, our hybrid adaptation can reduce error rates consistently for different amount of adaptation data, and outperforms significantly the STM approach across the board, especially when more data are used for adaptation. It is observed again that M-Hybrid approach performs better than F-Hybrid approach.

6. Conclusion

In this paper, we have proposed a new SSM-MCE linear regression approach to adaptation of an SSM-MCE trained prototype-based classifier and demonstrated its application for Chinese OCR. In real-world application, the feature-space adaptation method can be used for fast adaptation with a small amount of adaptation data, while the model-space adaptation method can be used to upgrade the performance of the classifier by using increasingly more adaptation data. The proposed hybrid adaptation approach offers a good practical solution for cases with different amount of adaptation data. In this study, we have confirmed the effectiveness of the proposed approach for supervised adaptation of font styles and low-quality text, respectively. As future work, we will study more adaptation scenarios with mismatched training and recognition conditions. We will also study the effectiveness of our approach for unsupervised adaptation.

References

[1] http://questvisual.com/.
[2] http://www.google.com/mobile/goggles/#text.
[3] J. Du, Q. Huo, L. Sun, and J. Sun. Snap and translate using Windows Phone. In *ICDAR*, pages 809–813, 2011.
[4] T. He and Q. Huo. A study of a new misclassification measure for minimum classification error training of prototype-based pattern classifiers. In *ICPR*, 2008.
[5] Q. Huo, Y. Ge, and Z.-D. Feng. High performance Chinese OCR based on Gabor features, discriminative feature extraction and model training. In *ICASSP*, pages 1517–1520, 2001.
[6] C. Igel and M. Hüsken. Improving the Rprop learning algorithm. In *2nd Int. Symp. on Neural Computation*, pages 115–121, 2000.
[7] C. J. Leggetter and P. C. Woodland. Flexible speaker adaptation for large vocabulary speech recognition. In *EUROSPEECH*, pages 1155–1158, 1995.
[8] I. Marosi. Industrial OCR approaches: architecture, algorithms and adaptation techniques. In *Document Recognition and Retrieval XIV*, SPIE Jan 2007, 6500-01.
[9] G. Nagy and G. L. Shelton. Self-corrective character recognition system. *IEEE Trans. on Information Theory*, IT-12(2):215–222, 1966.
[10] Y.-Q. Wang and Q. Huo. A study of designing compact recognizers of handwritten Chinese characters using multiple-prototype based classifiers. In *ICPR*, pages 1872–1875, 2010.
[11] X.-Y. Zhang and C.-L. Liu. Style transfer matrix learning for writer adaptation. In *CVPR*, pages 393–400, 2011.