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

String-Level Learning of Confidence Transformation for Chinese Handwritten Text Recognition

Da-Han Wang, Cheng-Lin Liu
National Laboratory of Pattern Recognition,
Institute of Automation, Chinese Academy of Sciences
95 Zhongguancun East Road, Beijing 100190, P. R. China
{dhwang, liucl}@nlpr.ia.ac.cn

Abstract

Handwritten text recognition systems commonly combine character classification confidence scores and context models for evaluating candidate segmentation-recognition paths, and the classification confidence is usually optimized at character level. On comparing the performance of class-dependent and class-independent confidence transformation (CT), this paper proposes two regularized class-dependent CT methods, and particularly, a string-level confidence learning method under the Minimum Classification Error (MCE) criterion. In experiments of online Chinese handwritten text recognition, the string-level confidence learning method was shown to effectively improve the recognition performance.

1. Introduction

Handwritten text (character string) recognition is commonly solved by the integrated segmentation-recognition approach to overcome the ambiguity of segmentation [1]. In this framework, candidate characters are formed by concatenating over-segmented segments and are assigned candidate classes by a classifier. The candidate characters and classes construct a candidate segmentation-recognition lattice, whose paths are evaluated by combining the classification scores, linguistic and geometric contexts. The optimal path of maximum score gives the result of character segmentation and recognition.

To better combine character classifier with context models, confidence transformation (CT) [2] has been applied to convert classifier outputs into posterior probabilities [3-5]. Recently, Wang et al. [5] investigated the effects of different types of CT in Chinese handwritten text recognition, and showed that the consideration of outlier/non-character probability benefits the recognition performance. These methods estimated confidence parameters only at character level, however.

In this paper, we consider class-dependent (CD) and class-independent (CI) confidence parameters. On comparing their performance, we propose two regularized CD methods and particularly, a string-level confidence learning method under the Minimum Classification Error (MCE) criterion [6]. Experimental results on a large database of online Chinese handwriting CASIA-OLHWDB [7] demonstrate the effectiveness of string-level confidence learning.

2. Overview of Recognition System

Fig. 1 shows the diagram of our Chinese handwritten text recognition system using integrated segmentation-recognition. The input string is first over-segmented into primitive segments, and consecutive segments are combined to generate candidate characters, which are assigned candidate classes by the character classifier. The candidate characters and classes form a segmentation-recognition candidate lattice (an example shown in Fig. 2). The paths in the candidate lattice are evaluated by combining classifier outputs, linguistic and geometric context models, and the optimal path gives the string recognition result. We use the path evaluation criterion presented in [8] for a candidate character sequence X and string class C:

$$\begin{aligned} f(X,C) = \sum_{i=1}^{n} \{ & k_i \log P(c_i \mid \mathbf{x}_i) + \lambda_1 \log P(c_i \mid c_{i-1}) \\ & + \lambda_2 \log P(c_i \mid \mathbf{g}_i^{uc}) + \lambda_3 \log P(z_i^p = 1 \mid \mathbf{g}_i^{ui}) \\ & + \lambda_4 \log P(c_{i-1}, c_i \mid \mathbf{g}_i^{bc}) + \lambda_5 \log P(z_i^g = 1 \mid \mathbf{g}_i^{bi}) \}, \end{aligned} \tag{1}$$

where k_i is the number of primitive segments composing the candidate character $\mathbf{x}_i$, $P(c_i \mid \mathbf{x}_i)$ is the character class confidence, $P(c_i \mid c_{i-1})$ is the bi-gram score, $P(c_i \mid \mathbf{g}_i^{uc})$, $P(z_i^p = 1 \mid \mathbf{g}_i^{ui})$, $P(c_{i-1}, c_i \mid \mathbf{g}_i^{bc})$, and $P(z_i^g = 1 \mid \mathbf{g}_i^{bi})$ are four geometric context scores, $\{\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5\}$ are five combining weights which are optimized by MCE training on string samples. Note that both the character classification outputs and geometric context scores are converted to posterior probabilities by CT.

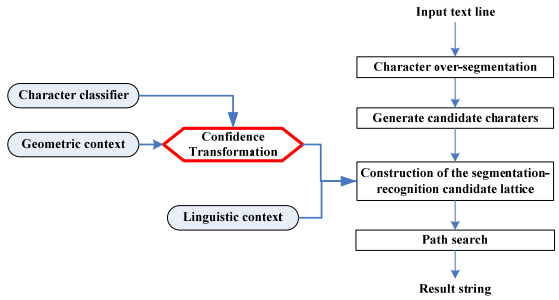

Figure 1. System diagram of Chinese handwritten text recognition.

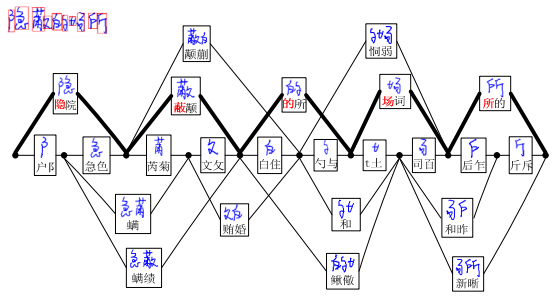

Figure 2. An example of the segmentation-recognition candidate lattice.

3. Confidence Transformation

Two widely used transformation functions are the sigmoidal function and the soft-max function. In Chinese text recognition, the sigmoidal function performs well while the soft-max function deteriorates the recognition performance [5]. The sigmoidal function gives multiple one-versus-all two-class probabilities:

$$P^{sg}(\omega_j \mid \mathbf{x}) = \frac{1}{1+\exp\left[-\left(\alpha S_j(\mathbf{x})+\beta\right)\right]}, \quad j=1,\ldots,M, \qquad (2)$$

where M is the number of defined classes, $S_j(\mathbf{x})$ is the similarity score of class ω_j output by the classifier, and α and β are the confidence parameters. The sigmoidal two-class probabilities can be combined into multi-class probabilities using the Dempster-Shafer (D-S) theory of evidence [9]:

$$P^{ds}(\omega_j \mid \mathbf{x}) = \frac{\exp\left[\alpha S_j(\mathbf{x})+\beta\right]}{1+\sum_{i=1}^{M}\exp\left[\alpha S_i(\mathbf{x})+\beta\right]}, \quad j=1,\ldots,M, \qquad (3)$$

which satisfies $\sum_{i=1}^{M} P^{ds}(\omega_i \mid \mathbf{x}) \le 1$, and the complement to one gives the probability of outlier class.

The D-S evidence performs best because it gives multi-class probabilities and considers the probability of outlier class [5].

The confidence parameters can be class-independent or class-dependent. For class-independent confidence, there are only two parameters α and β used in Equation (3). For class-dependent confidence, each class has its own parameters α_j and β_j and there are totally $2M$ parameters.

Traditionally, the confidence parameters are estimated by minimizing the cross entropy (CE) loss function on a validation dataset of character samples (preferably different from the dataset for training classifiers) [2]:

$$\min_{\{\alpha,\beta\}} L_{CE} = -\sum_{n=1}^{N}\sum_{j=1}^{M}\left[t_j^n \log P_j + (1-t_j^n)\log(1-P_j)\right], \qquad (5)$$

where N is the number of validating samples, and P_j is the sigmoidal probability of class ω_j. The parameters are learned by stochastic gradient descent.

3.1. Regularized Class-Dependent CT

The above confidence parameter learning methods is called character-level because character validating samples are used. Our experiments (will be detailed later) show that when estimated at character-level, class-independent confidence parameters outperform the class-dependent parameters. With many parameters, class-dependent CT suffers from over-fitting. We thus consider two regularized versions of class-dependent CT for alleviating the over-fitting.

One method is to constrain the one-vs-all separation boundary such that all classes have the same classification (SC) point: $s' = \beta_i / \alpha_i$, $i = 1,2,\ldots,M$. This variation can be considered as a tradeoff of class-dependent (CD) CT and class-independent CT, called **SCCD** in this paper. It has M+1 parameters ($\alpha_1,\ldots,\alpha_M$ and s').

Another method is to regularize class-dependent CT using shared parameters: $\alpha_i = (1-\lambda)\alpha_i + \lambda\bar{\alpha}$, where $\bar{\alpha} = \frac{1}{M}\sum_{j=1}^{M}\alpha_j$, and β_i is regularized in the same way. This variation regularizes the parameters such that α_i and β_i do not differ significantly between classes, and is called **ReguCD** in this paper.

4. String-Level Confidence Learning

To make the confidence parameters better perform string recognition, we propose to learn the context combining weights and confidence parameters simultaneously using the MCE training method [6], which has been widely used in speech recognition and handwriting recognition.

Assume classifying a string sample X whose ground-truth transcript is C_t into M string classes

$\{C_1,...,C_M\}$, where M is very large. Following [6], the misclassification measure on the string sample is approximated by

$$d(X,\Lambda) = -g(X,C_t,\Lambda) + g(X,C_r,\Lambda), \quad (6)$$

where Λ is the parameter set, $g(X,C_t,\Lambda)$ is the discriminant function for the true class, and $g(X,C_r,\Lambda)$ is the discriminant function of the closest rival class: $g(X,C_r,\Lambda) = \max_{C_k \neq C_t} g(X,C_k,\Lambda)$. The misclassification measure is transformed to loss by

$$l(X,\Lambda) = \frac{1}{1+e^{-\xi d(X,\Lambda)}}, \quad (7)$$

where ξ is a parameter to control the hardness of sigmoidal nonlinearity. The parameters in MCE training are learned by stochastic gradient descent.

In string-level training, the parameters are estimated to optimize the string recognition accuracy. The discriminant function is the path evaluation as in (1), into which the CT formula is substituted. The rival segmentation-recognition path, which is the most confusable one with the correct one, is obtained by beam search.

5. Experimental Results

We evaluated the CT methods on a large database of online Chinese handwriting CASIA-OLHWDB. This database is divided into six datasets, three for isolated characters (DB1.0-1.2, called DB1 for short) and three for handwritten texts (DB2.0-2.2, called DB2 for short). There are 3,912,017 isolated character samples and 52,221 handwritten pages (consisting of 1,348,904 character samples) in total. Both the isolated data and handwritten text data have been divided into standard training and test subsets.

We used two commonly used character classifiers: modified quadratic discriminant function (MQDF) [10] and nearest prototype classifier (NPC) trained by the LOG-likelihood of Margin (NPC-LOGM) criterion [11]. The classifier parameters were learned on 4/5 of training character samples (both the isolated characters in the training set of DB1 and the segmented characters in the training set of DB2, 4,207,801 samples in total), and the remaining 1/5 training samples were used for character-level confidence parameters estimation. The training character samples fall in 7,356 classes, including 7,185 Chinese characters and 171 alphanumeric characters and symbols.

The character classifier uses the local stroke direction histogram feature, implemented by the method of [12] with bi-moment normalization. To reduce the complexity of the classifier, the extracted 512D feature vector is projected onto a 160D subspace learned by Fisher linear discriminant analysis (FLDA).

The character bi-gram language model was trained on a text corpus containing about 50 million characters (about 32 million words) [5]. The geometric models are constructed using the methods similar to that in [13] on the training set of DB2, and the outputs of geometric models are converted to posterior probabilities by the sigmoidal function. We do not update confidence parameters of geometric models in string-level training at this moment, which will be investigated in the future.

For performance evaluation, we used two character-level metrics: Correct Rate (CR) and Accurate Rate (AR):

$$\mathrm{CR} = (N_t - D_e - S_e)/N_t,$$

$$\mathrm{AR} = (N_t - D_e - S_e - I_e)/N_t,$$

where N_t is the total number of characters in the ground-truth transcript. The numbers of substitution errors (S_e), deletion errors (D_e) and insertion errors (I_e) are calculated by aligning the recognition result string with the transcript by dynamic programming.

We evaluated the performance on the text dataset DB2, whose statistics are shown in Table 1.

Table 1. Statistics of DB2

	#page	#line	#line/page	#chars	#chars/line
train	4,072	41,710	10.24	1,082,220	25.95
test	1,020	10,510	10.30	269,674	25.66

We first evaluated the performance of character-level confidence learning (class-dependent and class-independent) in two path evaluation schemes: character classification+language model (LM), character classification+LM+geometric models (GM). Table 2 shows the performance of character-level learned CT, and Table 3 shows that of the regularized class-dependent CT (SCCD and ReguCD). We can see that when learning confidence parameters at character-level, class-independent CT performs much better than class-dependent CT. By regularization, the ReguCD improved the performance of class-dependent CT significantly, but the improved performance is still inferior to that of class-independent CT.

In string-level training, the class-dependent confidence parameters are first initialized with class-independent parameters estimated at character-level, and then updated in string-level learning on the training set of DB2. The updated confidence parameters are then used to train the context combining weights. When updating the confidence parameters in string-level training, the path evaluation function can use character classification only, or integrate more contexts. We evaluated two situations: using character classification only (Situation 1) and integrating character classification and language

model (Situation 2). The test performances are shown in Table 4 and Table 5, respectively.

Table 2. Performance of character-level learned CT.

	Class-dependent		Class-independent	
	CR	AR	CR	AR
Character classification+LM				
MQDF	0.8349	0.8142	**0.9061**	**0.8808**
NPC	0.8219	0.8092	**0.8699**	**0.8482**
Character classification+LM+GM				
MQDF	0.8904	0.8798	**0.9239**	**0.9147**
NPC	0.8902	0.8804	**0.9028**	**0.8927**

Table 3. Performance of character-level learned SCCD and ReguCD.

	SCCD		ReguCD	
	CR	AR	CR	AR
Character classification+LM				
MQDF	0.8143	0.7965	0.8990	0.8779
NPC	0.8514	0.8364	0.8464	0.8307
Character classification+LM+GM				
MQDF	0.8876	0.8779	0.9197	0.9089
NPC	0.8946	0.8855	0.8981	0.8883

Table 4. Performance of string-level learned CT in Situation 1.

	Class-dependent		SCCD		ReguCD	
	CR	AR	CR	AR	CR	AR
Character classification+LM						
MQDF	0.9063	0.8832	0.8829	0.8731	**0.9120**	**0.8906**
NPC	0.8692	0.8500	0.8375	0.8263	**0.8713**	**0.8514**
Character classification+LM+GM						
MQDF	0.9234	0.9141	0.9247	0.9176	**0.9260**	**0.9166**
NPC	**0.9030**	**0.8930**	0.8989	0.8904	0.9027	0.8925

Table 5. Performance of string-level learned CT in Situation 2.

	Class-dependent		SCCD		ReguCD	
	CR	AR	CR	AR	CR	AR
Character classification+LM						
MQDF	0.9122	0.8948	**0.9123**	**0.8975**	0.9108	0.8888
NPC	**0.8814**	**0.8661**	0.8769	0.8549	0.8777	0.8590
Character classification+LM+GM						
MQDF	0.9265	0.9180	**0.9273**	**0.9195**	0.9270	0.9181
NPC	**0.9090**	**0.8902**	0.9052	0.8954	0.9066	0.8972

From the results, we can see that string-level confidence learning improves the recognition performance, and the Situation 2 performs better than the Situation 1. In Situation 1, ReguCD performs better than the other two class-dependent CT methods, and outperforms the class-independent CT; In Situation 2, the three string-level learned class-dependent CT methods perform comparably well, and outperform the class-independent CT. When geometric context is not integrated in path evaluation, the improvement of string-level confidence learning is more evident.

6. Conclusion

We investigated the effects of class-dependent and class-independent confidence parameter learning, proposed two regularized versions of class-dependent CT and particularly, string-level confidence parameter learning, which were shown to benefit the string recognition performance. The joint optimization of classifier parameters, confidence parameters and context combining weights will be our future work.

Acknowlegement

This work was supported in part by the National Natural Science Foundation of China (NSFC) Grants 60825301 and 60933010.

Reference

[1] C.-L. Liu, S. Jaeger, M. Nakagawa, Online handwritten Chinese character recognition: The state of the art, *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(2): 198-213, 2004.

[2] C.-L. Liu, Classifier combination based on confidence transformation, *Pattern Recognition*, 38(1): 11-28, 2005.

[3] Y.X. Li, C.L. Tan, X. Ding, A hybrid postprocessing system for offline handwritten Chinese script recognition, *Pattern Analysis and Applications* 8: 272-286, 2005.

[4] Y. Jiang, X. Ding, Q. Fu, Z. Ren, Context driven Chinese string segmentation and recognition, Structural, *Structural, Syntactic, and Statistical Pattern Recognition*, LNCS, Vol.4109, pp.127-135, 2006.

[5] Q.-F. Wang, F. Yin, C.-L. Liu, Improving handwritten Chinese text recognition by confidence transformation, *Proc. 11th ICDAR*, pp.518-522, 2011.

[6] B.-H. Juang, W. Chou, C.-H. Lee, Minimum classification error rate methods for speech recognition, *IEEE Trans. Speech and Audio Processing*, 5(3): 257-265, 1997.

[7] C.-L. Liu, F. Yin, D.-H. Wang, Q.-F. Wang, CASIA online and offline Chinese handwriting databases, *Proc. 11th ICDAR*, pp. 37-41, 2011.

[8] Q.-F. Wang, Fei Yin, C.-L. Liu, Handwritten Chinese text recognition by integrating multiple contexts, *IEEE Trans. Pattern Anal. Mach. Intell.*, in press, 2012.

[9] J.A. Barnett, Computational Methods for a mathematical theory of evidence, *Proc. 7th IJCAI*, pp.868-875, 1981.

[10] F. Kimura, K. Takashina, S. Tsuruoka, Y. Miyake, Modified quadratic discriminant functions and the application to Chinese character recognition, *IEEE Trans. Pattern Anal. Mach. Intell.*, 9(1): 149-153, 1987.

[11] X.-B. Jin, C.-L. Liu, X. Hou, Regularized margin-based conditional log-likelihood loss for prototype learning, *Pattern Recognition*, 43 (7): 2428-2438, 2010.

[12] C.-L. Liu, X.-D. Zhou, Online Japanese character recognition using trajectory-based normalization and direction feature extraction, *Proc. 10th IWFHR*, pp.217-222, 2006.

[13] F. Yin, Q.-F Wang, C.-L. Liu, Integrating geometric context for text alignment of handwritten Chinese documents, *Proc. 12th ICFHR*, pp.7-12, 2010.