A Discriminative Model for On-line Handwritten Japanese Text Retrieval

Cheng Cheng, Bilan Zhu, and Masaki Nakagawa
Dept. of Computer and Information Sciences
Tokyo University of Agriculture and Technology
2-24-16 Naka-cho, Koganei-shi 184-8588 Tokyo, Japan
50009834702@st.tuat.ac.jp {Zhubilan, nakagawa}@cc.tuat.ac.jp

Abstract—This paper describes an unconstrained on-line handwritten Japanese text retrieval system from character recognition candidates. The system is based on a discriminative model which integrates the scores of character recognition, segmentation and geometric context in search and retrieval. Experiments on TUAT Kuchibe database show that the proposed method can effectively improve the system performance. When the search method with the optimal threshold retrieves for a keyword consisting of two, three or four characters, its f-measure is 0.720, 0.868 or 0.923, respectively.

Keywords—text retrieval; discriminative model; geometric context; character recognition

I. INTRODUCTION

Handwriting input interfaces have been employed in environments where a keyboard is not suitable. Portable devices like PDA, iPod, interactive electronic whiteboards and tablet PC’s are examples of such environments where a keyboard is too large for mobile systems or it is not suitable for annotations.

With the development of the Internet, and the availability of pen input devices, the size of handwritten (pen input) data is increasing rapidly. Efficient handwritten searching, browsing and retrieval tools are required by users from various domains. For this purpose, many general purpose handwritten retrieval systems have been developed during the past decade.

According to the word similarity scoring technique, there are two frameworks used. With the first method, handwritten text is recognized with a handwriting word recognizer (HWR), and the results are used for search and retrieval of the text. Marukawa et al. proposed a method, which can reduce errors when searching from incorrect recognition results by using two or more character recognition candidates and a confusion matrix [1]. Ota et al. extended the above idea by generating search terms considering mis-segmentation and mis-recognition with corresponding confidence and bi-gram probabilities [2]. Imagawa et al. investigated the reliability of recognition results with a neural network and improved both the recall and precision rates [3]. Lopresti et al. examined how OCR noises affect the performance of common information search models [4]. Cao et al describes a keyword spotting method by modeling imperfect word segmentation as probabilities and integrating these probabilities into the word spotting algorithm [5]. Liu et al presents a text query-based method for keyword spotting from online Chinese handwritten documents [6]. The similarity between a text word and handwriting is obtained by combining the character similarity scores given by a character classifier.

Section 2 summarizes the retrieval system. Section 3 describes the evaluation method for retrieval results. Section 4 describes the parameter learning method. Section 5 details the experiments, and section 6 draws the conclusion.
II. OVERVIEW OF A RETRIEVAL SYSTEM

Our keyword retrieval system for on-line handwritten Japanese text uses a handwriting word recognizer (HWR). First, on-line handwritten Japanese text is over-segmented into primitive segments according to the features such as spatial information between adjacent strokes. Then one or more consecutive primitive segments form a candidate character pattern, and each pattern is associated with several candidate classes with scores assigned by character classification. The combination of all candidate patterns and character classes is represented by a candidate lattice. Last, the system searches the candidate lattice, and obtains several retrieval results that match with the input keyword. The retrieval results are evaluated by character recognition, segmentation and geometric context. In order to increase the precision of its output, we prune results whose evaluated scores are below a threshold score $T_s$.

III. EVALUATION FOR RETRIEVAL RESULTS

Given a keyword, the system searches the candidate lattice to find stroke sequences matching with the input keyword where a stroke is a sequence of pen-tip coordinates from pen-down to pen-up. The score between the input keyword and each candidate is evaluated by

$$f(S, \text{Keyword}) = \sum_{i=1}^{n} \lambda_i f_i$$  \hspace{1cm} (1)

where $S$ is a sequence of strokes matching with the keyword, $f_i$ is the feature function for the score of character recognition, segmentation and geometric context, and $\lambda_i$ is the weighting parameter.

The six feature functions in Eq. (1) depict the characteristic of character shape and geometric context, and the details are shown in Table 1.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Formulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f_1$</td>
<td>$\sum_{i=1}^{n} D_1(s_i, s_{i+1}, c_i, c_{i+1})$</td>
</tr>
<tr>
<td>$f_2$</td>
<td>$\sum_{i=1}^{n} HW(s_i, c_i)$</td>
</tr>
<tr>
<td>$f_3$</td>
<td>$\sum_{i=1}^{n} Inner(s_i, c_i)$</td>
</tr>
<tr>
<td>$f_4$</td>
<td>$\sum_{i=1}^{n} D_2(s_i, c_i)$</td>
</tr>
<tr>
<td>$f_5$</td>
<td>$\sum_{i=1}^{n} Seg(s_i, s_{i+1})$</td>
</tr>
<tr>
<td>$f_6$</td>
<td>$\sum_{i=1}^{n} Rec(s_i, c_i)$</td>
</tr>
</tbody>
</table>

In table 1, $n$ is the length of the input keyword, $c_i$ is the $i$-th character of the input keyword; $s_i$ is the $i$-th character patterns matching with $c_i$.

The term $D_1(s_i, s_{i+1}, c_i, c_{i+1})$ is a QDF function for a binary geometric feature vector $p_i$ and characters $c_i$, $c_{i+1}$. The binary geometric feature vector $p_i$ has two elements measured from the bounding boxes of two adjacent character patterns $s_i$, $s_{i+1}$: the vertical distances between the upper bounds and between the lower bounds as shown in Fig.1.

The term $HW(s_i, c_i)$ is a QDF function for a geometric feature vector and characters $c_i$. The geometric feature vector comprises the height and width of bounding box of character pattern $s_i$.

The term $Inner(s_i, c_i)$ is a QDF function for a geometric feature vector $q_i$ and characters $c_i$. The geometric feature vector $q_i$ comprises six values as shown in Fig.2 where $acs$ is the average character size. The first three values represent the horizontal gaps of three vertical slits (partitioned from vertical projection), and the last three ones represent the vertical gaps of three horizontal slits (from horizontal projection).

The term $Seg(s_i, s_{i+1})$ is a QDF function for a geometric feature vector $p_i'$ and characters $c_i$. The geometric feature vector $p_i'$ comprises two elements, the first element represents the length from the top to the center line of the text line and the second element represents that from the bottom to the line as shown in Fig.1.

The values of geometric features are normalized with respect to the average character size $acs$ for scale invariance.

The term $Rec(s_i, c_i)$ is a function of a SVM classifier which measures the plausibility of segmentation between two adjacent candidate character patterns $s_i$, $s_{i+1}$ [12].
IV. PARAMETER ESTIMATION

We are inspired with Conditional Random Field to construct a parameter estimation method for evaluating the retrieval results as follows:

Given the training data $D = \{x^i, y^i\}_{i=1}^{N}$, where $N$ is the number of input keywords, $x^i = \{x^i_1, x^i_2, ..., x^i_T\}$ denotes the sequence of retrieval results in a candidate lattice $(x^i_1, x^i_2, ..., x^i_T)$ are correct, and the others are incorrect), and $y^i$ is the input keyword, the conditional probability for correct retrieval is formulated as

$$P(x^i|y^i) \approx \frac{\exp[\sum_{j=1}^{N} f(x^i_j, y^i)]}{\sum_{i=1}^T \exp(f(x^i_j, y^i))}$$  \hspace{1cm} (2)

where $f(x^i_j, y^i)$ is defined in Eq. (1).

As many output candidates were incorrect, we pruned the stroke candidates in which the evaluated score were below a threshold score $Ts$. In order to let the scores of correct stroke candidates become as high as possible and the scores of the incorrect stroke candidates become as low as possible, the loss function is defined as the summation of the negative log-likelihoods:

$$f(\theta) = -\sum_{i=1}^N \log P(x^i|y^i)$$
$$= \sum_{i=1}^N \log \left( \sum_{j'=1}^T \frac{\exp(f(x^i_j, y^i))}{\sum_{j=1}^T \exp(f(x^i_j, y^i))} \right) - \sum_{i=1}^N \sum_{j=1}^T f(x^i_j, y^i)$$  \hspace{1cm} (3)

The stochastic gradient decent algorithm is used to optimize the weighting parameters:

$$\lambda(t+1) = \lambda(t) - \varepsilon(t) Vf(\theta)$$  \hspace{1cm} (4)

where $\varepsilon(t)$ is a learning step and $Vf(\theta)$ denotes the partial derivative with respect to the parameters.

$$\frac{\partial f(\theta)}{\partial \lambda_k} = \sum_{i=1}^N \frac{\sum_{j=1}^T \exp(f(x^i_j, y^i)) f_k}{\sum_{j=1}^T \exp(f(x^i_j, y^i))} - \sum_{i=1}^N \sum_{j=1}^T f_k$$  \hspace{1cm} (5)

It is necessary to point out that the well-known CRFs infer and maximize the posteriori probability while the model proposed by us infers and maximizes the conditional probability as shown in Eq. (2).

V. EXPERIMENTS AND EVALUATION

We evaluate the performance of the retrieval system with $f$-measure

$$f = \frac{2}{1/r + 1/p}$$  \hspace{1cm} (6)

where $r$ is recall and $p$ is precision defined in Eq.(7) and Eq.(8), respectively. The recall rate measures the tolerance of the system to search errors, while the precision rate measures the tolerance to noises. The $f$-measure is an overall performance of the retrieval system.

We employ the database “HANDS-Nakayosi_t-98-09” (in brief, Nakayosi)[14] to train our character classifier and geometric scoring functions.

We employ the database “HANDS-Kondate bf-2001-11” (in brief, Kondate) to train our SVM classifier for the candidate segmentation point probability.

We employ the database “HANDS-kuchibue_d-97-06” (in brief, Kuchibue) [14] in our experiments, by concatenating single characters as text lines. Each page has at most 10 lines and each line has at most 10 characters with the gap between adjacent characters randomly ranging from 1 to 10 pixels. The data of 60 writers are used for training parameters and the rest 60 for testing.

We compared the retrieval performance of three parameter learning methods with the same environment. Table 2 lists the results of our previous work [10] in which $\lambda_1 = 0$ and $\lambda_2 = 1$. Table 3 lists the results of the parameter learning method proposed by Zhu et al [15] for Japanese text recognition that trains the weighting parameters by the genetic algorithm to optimize the recognition rate. Table 4 lists the results of the system proposed in this paper.

From the results, we can see that our proposed model has brought the best retrieval performance. Introducing the adjustable weighting parameters to evaluate the similarity between the input keyword and handwritten text has improved the retrieval accuracy; the method for estimating the weighting parameters to optimize the conditional probability for correct retrieval has brought better retrieval accuracy than the method in [15] that estimates the weighting parameters to optimize the recognition rate.

| TABLE 2. Results of our previous work [10] |
| Length | Recall | Precision | $f$-measure |
| 2      | 71.8%  | 50.2%     | 0.582       |
| 3      | 90.0%  | 71.3%     | 0.791       |
| 4      | 89.8%  | 87.0%     | 0.883       |

| TABLE 3. Using the parameters of character string recognition [15] |
| Length | Recall | Precision | $f$-measure |
| 2      | 84.2%  | 63.6%     | 0.719       |
| 3      | 89.6%  | 84.6%     | 0.866       |
| 4      | 91.9%  | 92.6%     | 0.920       |

| TABLE 4. Results of proposed system |
| Length | Recall | Precision | $f$-measure |
| 2      | 84.2%  | 63.8%     | 0.720       |
| 3      | 90.9%  | 83.2%     | 0.868       |
| 4      | 91.9%  | 93.1%     | 0.923       |
VI. CONCLUSION

In this paper, we described a discriminative model for unconstrained handwritten Japanese text retrieval inspired by the Conditional Random Field model, which incorporates the scores of character recognition, segmentation and geometric context. Experiments on Kuchibue database demonstrate the effectiveness of our proposed method.

ACKNOWLEDGMENT

This work is being partially supported by the R&D fund for "development of pen & paper based user interaction" under Japan Science and Technology Agency.

REFERENCES