Text/Non-text Ink Stroke Classification in Japanese Handwriting Based on Markov Random Fields

Xiang-Dong Zhou, Cheng-Lin Liu
National Laboratory of Pattern Recognition,
Institute of Automation, Chinese Academy of Sciences
P.O. Box 2728, Beijing 100080, P.R. China
{*xdzhou, liucl*}*@nlpr.ia.ac.cn*

Abstract

In this paper, we present an approach for separating text and non-text ink strokes in online handwritten Japanese documents based on Markov random fields (MRFs), which effectively utilize the spatial relationship between strokes. Support vector machine (SVM) classifiers are trained for individual stroke and stroke pair classification, and on converting the SVM outputs to probabilities, the likelihood clique potentials of MRF are derived. In experiments on the TUAT Kondate database, the proposed MRF approach yield superior performance compared to individual stroke classification and sequence classification based on hidden Markov models (HMMs).

1. Introduction

With the increased use of tablet PCs and electronic whiteboards with large writing areas, users can draw various heterogeneous structures such as text, drawings and table forms freely. Such freely handwritten ink documents bring new challenges to automatic analysis and recognition. Before the processing tasks like text recognition, editing and retrieval can be accomplished, the ink document needs to be segmented into regions of homogeneous stroke type, say, regions of text, drawings, table forms, etc. Since the ink document is freely structured, there is very little prior knowledge (e.g. size, location and orientation of text lines) to guide top-down parsing. Fig. 1 shows two pages of online ink documents.

Due to the free structure and unavailability of prior knowledge, bottom-up classification of ink strokes is a feasible way for ink document segmentation. Separating text strokes from non-text ones such as graphics and diagrams, is a fundamental problem in this process.

Assuming independence of strokes, each stroke can be classified individually [1]. Actually, significant information exists not only in the stroke shapes but also in the temporal and spatial relationship between strokes. This context can help disambiguate some uncertainties. Some previous works have incorporated such context information heuristically for disambiguating individual stroke types [2-4]. Bishop et al. [5] proposed a rather principled text/non-text stroke classification approach based on hidden Markov models (HMMs) [6], which can utilize the temporal information of stroke sequences effectively, but ignores the very important spatial context.

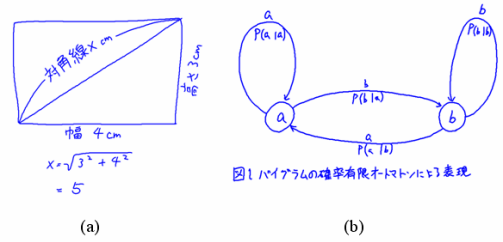

Figure 1. Two pages of online ink documents.

The HMM approach takes into account only the relationship between successively adjacent strokes in writing order. The strong correlation between strokes adjacent spatially but not temporally should also help disambiguate stroke classification. To better utilize such spatial context, we propose a text/non-text stroke classification approach based on the Markov random field (MRF) framework, which represents the interactions between strokes by the neighborhood system and clique potentials. We derive the likelihood clique potentials from the probabilistic outputs of support vector machine (SVM) classifiers on stroke features.

We have evaluated our system in experiments on the TUAT Kondate database [4]. The results show that by incorporating inter-stroke relationships, the accuracy of stroke classification is improved significantly,

and the proposed MRF approach outperforms the HMM approach.

2. MAP-MRF Framework

The stroke classification problem can be formulated as a labeling problem in which the strokes correspond to the set of sites $S = \{1, \ldots, I\}$, and the classes correspond to the set of labels $L = \{1, \ldots, J\}$ which are *text* and *non-text* here. The feature vectors of the strokes, $\{\boldsymbol{o}_1, \ldots, \boldsymbol{o}_I\}$, constitute the observation set O. The solution is to assign the sites a best labeling configuration $F^* = \{f_1^*, \ldots, f_I^*\}$, $f_i^* \in L$ under an optimization criterion, which is usually the maximum a posteriori (MAP) probability:

$$\begin{aligned} F^* &= \arg\max_F P(F \mid O) \\ &= \arg\max_F p(O \mid F) P(F) \end{aligned} \tag{1}$$

where $p(O \mid F)$ is the likelihood function for O given F, and $P(F)$ is the prior probability of F. For simplicity, we assume that the observations are conditionally independent at all sites.

2.1. Markov Random Fields

To calculate the prior probability $P(F)$ is intractable because the interactions between the labels are global. To make it tractable, the MRF constrains the interdependence of labels by assuming that the label of a site is only dependent on the labels of its neighboring sites. This is described as the Markovianity and can be depicted by the neighborhood system [7]. The neighborhood system ∂i denotes the neighbors of site i that meets with $i' \in \partial i \Leftrightarrow i \in \partial i', i \notin \partial i$. A clique c is defined as a subset of sites that are all mutual neighbors according to the neighborhood system.

The Hammersley-Clifford theorem establishes the equivalence between the Markov random field and the Gibbs random field [7],

$$P(F) = \frac{1}{Z} e^{-U(F)} \tag{2}$$

where

$$U(F) = \sum_{c \in C} V_c(F) \tag{3}$$

is called the prior energy, and

$$Z = \sum_{F'} e^{-U(F')} \tag{4}$$

is the normalization factor called the partition function. The energy function is the summation of clique potentials over all possible cliques c. The clique potentials V_c are defined as the costs to different cliques for encouraging or penalizing different local interactions among neighboring sites.

Taking the likelihood function $p(O \mid F)$ into consideration, we have

$$p(O \mid F) P(F) = \frac{1}{Z} e^{-(U(O|F) + U(F))} \tag{5}$$

where

$$U(O \mid F) = -\log p(O \mid F) \tag{6}$$

is called the likelihood energy. By the conditional independence assumption of the observations, we have

$$U(O \mid F) = \sum_{c \in C} V_c(O \mid F) \tag{7}$$

where $V_c(O \mid F)$ are the likelihood clique potentials derived from the negative logarithm of the conditional probabilities. Finally, the posterior energy corresponding to the a posteriori probability $P(F \mid O)$ in Eq. (1) can be formulated as

$$\begin{aligned} U(F \mid O) &\propto U(O \mid F) + U(F) \\ &= \sum_{c \in C} \{V_c(O \mid F) + V_c(F)\} \end{aligned} \tag{8}$$

For simplicity, we consider only single-site cliques $C1 = \{i\}$ and pair-site cliques $C2 = \{i, i'\}$ in our system. According to Eq. (8),

$$\begin{aligned} U(O \mid F) + U(F) = &\sum_{C1} [V_{C1}(O \mid F) + V_{C1}(F)] \\ &+ \sum_{C2} [V_{C2}(O \mid F) + V_{C2}(F)] \end{aligned} \tag{9}$$

The likelihood clique potentials describe the statistical information of the observations given the labels, while the prior clique potentials encode the prior information of neighboring labels [8].

Now in MAP-MRF framework, maximizing the a posteriori probability in Eq. (1) is equivalent to minimizing the energy function in Eq. (9).

2.2. Decoding Strategy

To find the best labeling configuration among all the possible ones is a combinational problem and is computationally expensive. This is a non-trivial problem, because the energy function may be non-convex and exhibits many local minima.

In our work, we use the relaxation labeling (RL) algorithm [7] to minimize the energy function (9) of MRF. In RL, a real value called labeling strength is

defined, denoting the feasibility that a label is assigned to a site, thus the combinatorial minimization is converted to a real minimization subject to linear constraints. RL depends not much on the initialization, and in our system we initialize with equal labeling strength for each site. After iterations, the algorithm will converge and the winner-take-all strategy will be employed for assigning labels.

3. MAP-MRF for Stroke Classification

We use the above MAP-MRF framework to formulate the contextual ink stroke classification problem. The class labels of a set of strokes are assigned to minimize an energy function. The energy function is the summation of the clique potentials and the cliques are defined according to the neighborhood system.

3.1. The Neighborhood System and Cliques

In online ink documents, the strokes close to each other usually have identical class labels, so we design the neighborhood system according to the minimum distance between strokes, that is, only the strokes within a certain distance are said to be neighbors. In our system, the threshold of distance is set to 0.4 times the average text stroke length estimated from the training set. For convenience, only single-site and pair-site cliques are taken into account in our experiments.

3.2. Stroke Features

To formulate the single-site likelihood potential, 11 features, which have been mentioned in [9], are extracted from each stroke. They are the stroke length, area, compactness, eccentricity, circular variance, rectangularity, centroid offset along major axis, closure, absolute curvature, perpendicularity, and signed perpendicularity.

To formulate the pair-site likelihood potential, the relationships between two neighboring strokes are represented by binary stroke features. We use four binary stroke features, which are the minimum distance between two strokes, the maximum and minimum distance between the endpoints of two strokes and the distance between the centers of the bounding boxes of two strokes.

3.3. The Likelihood Clique Potentials

To evaluate the single-site and pair-site potentials from unary stroke features and binary stroke features, we train support vector machine (SVM) classifiers [10] and transform the SVM outputs to probabilities by fitting a sigmoid function for each class. Single strokes are classified to two classes: *text* and *non-text*. Stroke pairs are classified to three classes: *text-text*, *nontext-nontext*, and *text-nontext*.

An SVM functions as a binary (two-class) classifier. For multi-class classification, we use multiple SVMs each separating one class from the others.

We choose the SVM for classification because it is at the top of classification performance in the state of the art. The SVM is a hyperplane classifier in the pattern feature space or a nonlinearly expanded feature space. Its decision function is formulated as a weighted average of kernel functions with a number of training vectors called support vectors. The weighting coefficients are estimated by maximizing a margin criterion on training patterns, which is converted to a dual quadratic programming problem.

For single-site classification, the output (decision function) of the binary SVM is converted to a posterior probability for class *text*. The complement of the probability is the probability of class *non-text*. For three-class pair-site classification, the outputs of three SVMs are converted to three posterior probabilities for three classes.

On training the SVMs on training samples, the SVM outputs are converted to posterior probabilities by fitting sigmoid functions on a validation sample set [11][12]. For *M*-class problem, outputs of *M* SVMs are converted to posterior probabilities by

$$\begin{aligned}\pi_j(f) &= P(t_j = 1 \mid f) \\ &= \frac{1}{1+\exp[-(\beta_{j1} f + \beta_{j0})]}, \\ & j = 1, \ldots, M\end{aligned} \tag{10}$$

where β_{j1} is the weight, β_{j0} is the bias for class *j*. These parameters are estimated by minimizing the cross-entropy function with weight decay term:

$$\begin{aligned}\min J = &-\sum_{n=1}^{N}\sum_{j=1}^{M}[t_j \log \pi_j + (1-t_j)\log(1-\pi_j)] \\ &+ \lambda \sum_{j=1}^{M} \beta_{j1}^2\end{aligned} \tag{11}$$

where λ is a pre-specified coefficient for weight decay. The criterion is minimized by stochastic gradient descent.

Therefore, given a set of feature vectors x_n, $n = 1, \ldots, N$ the probabilistic outputs which take the form of a posterior probability $P(t_j = 1 \mid x_n)$, $j = 1, \ldots, M$, are achieved by Eq. (15). According to

Bayesian rule, the posterior probabilities can be converted to conditional probabilities, i.e.

$$p(x \mid t) = \frac{P(t \mid x)\, p(x)}{P(t)} \propto \frac{P(t \mid x)}{P(t)} \qquad (12)$$

where $P(t)$ is the prior probability for class t which can be estimated from the training set.

By the Gibbs distribution and the conditionally independent assumption of the observations, we obtain the following single-site likelihood clique potentials,

$$V_{C1}(o_i \mid j) = -\log p(o_i \mid j), \quad j \in L \qquad (13)$$

and the pair-site likelihood clique potentials,

$$V_{C2}(o_{ii'} \mid j, j') = -\log p(o_{ii'} \mid j, j'), \quad j, j' \in L \qquad (14)$$

where o_i, $o_{ii'}$ are the unary and binary stroke features respectively and j, j' are class labels.

3.3. The Prior Clique Potentials

The single-site prior clique potential depends on the label assigned to the site

$$V_{C1}(j) = v_j, \; j \in L \qquad (15)$$

where $v_j > 0$ is the penalty against that the site is labeled j. The higher v_j is, the less strokes will be assigned the label j. This has an effect of controlling the percentage of the sites labeled j.

Because spatially adjacent strokes always share same labels, the pair-site prior clique potential in our system is designed to favor that the sites of the clique are assigned same labels, i.e.

$$V_{C2}(j, j') = \begin{cases} v_{20}, & \text{if } j \text{ and } j' \text{ are identical} \\ v_{21}, & \text{otherwise} \end{cases} \qquad (16)$$

where $v_{21} > v_{20} > 0$.

In principle, the prior clique potentials are proportional to the negative logarithm of prior probabilities of single-site or pair-site classes, which can be estimated from training samples.

4. Experimental Results

To evaluate the performance of the proposed text/non-text classification approach, we have experimented on the TUAT HANDS-Kondate_t_bf-2001-11 (in brief, Kondate) database, of online freeform handwritten Japanese documents without any writing constraints [4]. The database contains the online ink documents of 67 writers, 41 pages per writer covering the stroke types of text, formula, figure, ruled line and editing mark. The formulas, which are made up of both characters and non-characters, are excluded in our experiment, thus the non-text strokes are composed of figure, ruled line and editing mark strokes.

10 pages of each writer, totally 10×67=670 pages, including both text and non-text strokes, are selected for our experiment. Example pages are shown in Fig. 2. Among the selected data, 310 pages are used for training classifiers (SVMs), and 360 pages are used for testing. The numbers of strokes for each stage are listed in Table 1.

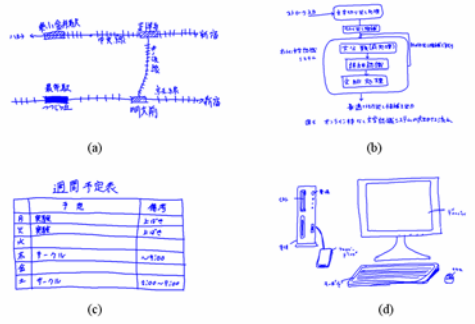

Figure 2. Examples of pages in TUAT Kondate database.

Table 1. The numbers of strokes for each stage.

	Text	Non-text
Training	51681	9869
Test	61969	10793

We use SVMs with 4-th order polynomial kernel function for single-site and pair-site classification (the outputs converted to conditional likelihood potentials). For single-site prior clique potential, the value for *text* is set to 0.029, and that for *non-text* is set to 0.305. For pair-site prior clique potential, we set $v_{20} = 0.013$ and $v_{21} = 0.749$.

We have drawn a comparison between the method based on HMM presented in [5] and our proposed MRF approach. For HMM, both the unary and binary stroke features are identical to the method based on MRF and the emission probabilities are derived from the probabilistic outputs of the SVM classifiers. The classification correct rates for individual classification (SVM), HMM and MRF are listed in Table 2, and the corresponding confusion matrices are listed in Table 3. From the two tables, we can see that HMM outperforms individual classification, and the MRF based approach is superior to that based on HMM.

Table 2. Correct rates (%) of stroke classification.

Individual	HMM	MRF
92.58	94.48	96.61

5. Conclusion

We have implemented a text/non-text ink stroke classification approach based on Markov random fields for online handwritten Japanese documents. The likelihood clique potentials of MRF are derived from the probabilistic outputs of the SVM classifiers. In experiments on the TUAT Kondate database, we evaluated the system performance between the methods based on HMM and MRF. The experimental results have demonstrated the superiority of the MRF approach, which takes advantage of the interactions between spatially adjacent strokes.

Acknowledgements

This work was supported by the Central Research Laboratory of Hitachi Ltd., Tokyo, Japan. The authors thank the Nakagawa Laboratory of Tokyo University of Agriculture and Technology (TUAT) for providing the Kondate database.

References

[1] A.K. Jain, A. M. Namboodiri, and J. Subrahmonia, "Structure in On-line Documents," *Proc. Sixth Int'l Conf. Document Analysis and Recognition*, Seattle, WA, pp. 844-848, 2001.

[2] M. Shilman, Z. Wei, S. Raghupathy, P. Simard, and D. Jones, "Discerning Structure From Free-Form Handwritten Notes," *Proc. Seventh Int'l Conf. Document Analysis and Recognition*, Edinburgh, Scotland, pp. 60-65, 2003.

[3] K. Machii, H. Fukushima, and M. Nakagawa, "On-line Text/Drawings Segmentation of Handwritten Patterns," *Proc. Second Int'l Conf. Document Analysis and Recognition*, Tsukuba, pp. 710-713, 1993.

[4] K. Mochida and M. Nakagawa, "Separating Figures, Mathematical Formulas and Japanese Text from Free Handwriting in Mixed On-line Documents," *Int. J. Pattern Recognition and Artificial Intelligence*, vol. 18, no. 7, pp. 1173-1187, Feb. 2004.

[5] C. M. Bishop, M. Svensén, and G. E. Hinton, "Distinguishing Text from Graphics in On-line Handwritten Ink," *Proc. Ninth Int'l Workshop Frontiers in Handwriting Recognition*, Tokyo, pp. 142–147, 2004.

[6] L.R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–285, 1989.

[7] S. Z. Li, *Markov Random Field Modeling in Image Analysis*, Springer, Tokyo, 2001.

[8] J. Zeng, Z.-Q. Liu, "Markov Random Fields for Handwritten Chinese Character Recognition," *Proc. Eighth Int'l Conf. Document Analysis and Recognition*, Seoul, pp. 101–105, 2005.

[9] D. Willems, S. Rossignol, and L. Vuurpijl, "Mode Detection in On-line Pen Drawing and Handwriting Recognition," *Proc. Eighth Int'l Conf. Document Analysis and Recognition*, Seoul, pp. 31-35, 2005.

[10] V. N. Vapnik, *Statistical Learning Theory*, John-Wiley Press, 1998.

[11] J. Platt, "Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods," *Advances in Large Margin Classifiers*, A.J. Smola, P. Bartlett, D. Scholkopf, D. Schuurmanns (Eds), MIT Press, 1999.

[12] C.-L. Liu, "Classifier Combination Based on Confidence Transformation," *Pattern Recognition*, vol. 38, no. 1, pp. 11–28, Jan. 2005.

Table 3. Confusion matrices (row: true class; column: predicted class).

	Individual		HMM		MRF	
	Text	Non-text	Text	Non-text	Text	Non-text
Text	61464	505	60785	1184	61044	925
Non-text	4897	5896	2831	7962	1543	9250