Markov Random Fields for Handwritten Chinese Character Recognition

Jia Zeng and Zhi-Qiang Liu
Center for Media Technology, School of Creative Media
City University of Hong Kong, Hong Kong SAR, P.R.CHINA
jia.rcmt@cityu.edu.hk, smzliu@cityu.edu.hk

Abstract

In this paper, we propose a statistical-structural scheme for Chinese character modeling based on Markov random fields (MRFs). We use 2-D Gabor filters to extract directional stroke segments from images of Chinese characters, where each stroke segment is associated with a state in Markov random field models. The structural information is described by neighborhood system and pair-state clique potentials; meanwhile the statistical information is represented by single-state probability density functions (pdfs). Extensive experiments on similar characters have been carried out on the database ETL9B. The experimental results confirm that Markov random field models are effective in modeling both statistical and structural information of Chinese characters, and works well for handwritten Chinese character recognition.

1. Introduction

Graphical models are instances of statistical-structural scheme and receiving increasing attention in recent years. Hidden Markov models (HMMs) are directed chain graphs with Markov properties, which have been used successfully in many applications, notably speech recognition. Markov random fields (MRFs) are undirected graphs with Markov properties. Within the MRF framework, statistical interactions among adjacent locations in a pattern or image are reflected by two fundamental concepts: *neighborhood system* $\mathcal{N}$ and *clique potentials* V_c. For high-level computer vision applications, neighborhood systems are usually defined on irregular states. Two states s_1 and s_2 are considered neighbors in an $\mathcal{N}$ to each other if the Euclidean distance $L_2(s_1, s_2)$ between them is less than a certain radius. In order to encourage or penalize different local interactions among neighboring states, costs are assigned to different cliques. By use of the equivalence between MRF and GRF (Gibbs random field), the global configuration can be reached by computing *maximum a posteriori* (MAP) of MRF [2]. This MAP-MRF framework enables us to systematically develop algorithms for a variety of vision problems using rational principles rather than *ad hoc* heuristics. Recently, Cai and Liu have developed handwriting recognition systems based on noncausal MRFs [1, 3], where the relaxation labeling algorithm is used for handwritten numerals recognition.

In light of success of the MRF framework, let's now reexamine the stroke relationships of handwritten Chinese characters. The stroke segments can be represented by the probability distributions of their locations, directions and lengths. The states of MRFs encode these information from training samples. Different state relationships represent different character structures. Through proper definition of neighborhood systems, it is especially versatile in representing various kinds of stroke relationships, including the relationship among more than two stroke segments, the relationship between the stroke segments far away from each other and the high-order relationship. We encourage and penalize different stroke relationships by assigning appropriate clique potentials to spatial neighboring states. Then, the structural information is described by neighborhood system and multi-state clique potentials; meanwhile the statistical uncertainty is modeled by single-state potentials for probability distributions of individual stroke segment. Hence, the recognition problem can be formulated into MAP-MRF framework to find an MRF that produces the *maximum a posteriori* for unknown input data. This MRF modeling for Chinese characters incorporates the configuration of individual stroke segments as well as their relationships within a unified framework. It is not only rigorous in formulation but also effective in representing the character structure statistically.

2. Feature Extraction using 2-D Gabor Filters

Generally, the stroke of Chinese characters falls in approximately four directions: horizontal (0°), right-diagonal (45°), vertical (90°) and left-diagonal (135°). Similar to the method in [5], we use 2-D Gabor filters to decompose a

Proceedings of the 2005 Eight International Conference on Document Analysis and Recognition (ICDAR'05)
1520-5263/05 $20.00 © 2005 IEEE

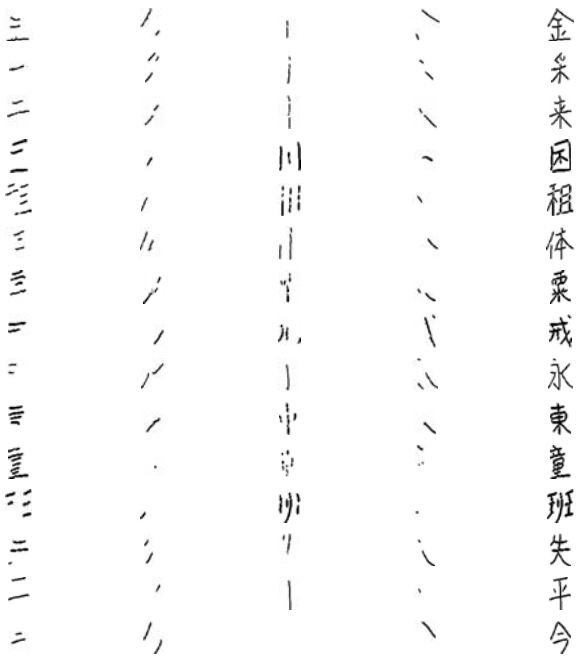

Figure 1. Four directional stroke segments. The right-most is the reconstructed character using these stroke segments.

character image into four directional stroke segments. Fig. 1 shows the four directional stroke segments of Chinese characters 金, 采, 来, 困, 租, 体, 粟, 戒, 永, 東, 童, 班, 失, 平, 今.

We extract features including the centroid (C), the direction (D) and the length (L) of each blob-like stroke segment o_i. The centroid is the center of mass of the region. The direction $\varphi \in [-5, 175)$ is the angle between the horizontal dashed line and the major axis, and the length is the major axis length of the ellipse illustrated in Fig. 2. Thus, the two types features can be represented as follows:

1. Unary features: determined by individual stroke segment o_i, i.e., $o_i = \{C_i, D_i, L_i\}$, where C_i and D_i reflect spatial property and L_i embodies one-dimensional measurement.

2. Binary features: determined by the relationship between two adjacent stroke segments o_i and $o_{i'}$, i.e., $o_{ii'} = \{C_{ii'}, D_{ii'}, L_{ii'}\}$, where $C_{ii'} = C_{i'} - C_i$, $D_{ii'} = D_{i'} - D_i$, $L_{ii'} = L_{i'} - L_i$.

We finally normalize the centroid and length by the character width (W), height (H) to reduce size variation without significant deformation.

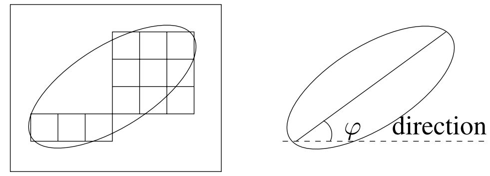

Figure 2. The left side of the figure shows a stroke segment and its corresponding ellipse. The right side shows the same ellipse, and the direction is the angle between the horizontal dashed line and the major axis, and the length is the major axis length.

3. MRFs for Chinese Characters

MRF is a graphical model based on an undirected graph $\mathcal{G} \triangleq (S, \mathcal{N})$ in which each state (node) $s_i \in S$ corresponds to a random variable. Every variable is dependent on the values of its immediate neighbors, which is determined by the links between the states according to the neighborhood system $\mathcal{N}$, i.e.,

$$P(S) > 0, \forall S, \tag{1}$$

$$P(s_j|S_{\{/j\}}) = P(s_j|\mathcal{N}_{s_j}),\ 1 \leq j \leq m, \tag{2}$$

where $S_{\{/j\}}$ are all other states except s_j and $\mathcal{N}_{s_j}$ are all states neighboring the state s_j. A clique c is defined as a subset of states in S that are all pair-wise neighbors. It consists either of a single state $\mathcal{C}_1 = \{s_j\}$, or of a pair of neighboring states $\mathcal{C}_2 = \{(s_j, s_{j'})\}$, or of a triple of neighboring states $\mathcal{C}_3 = \{(s_j, s_{j'}, s_{j''})\}$, and so on, which corresponds to the unary features or the binary features or the ternary features, respectively. High order relationship can also be represented by high order cliques. Without loss of generality, we consider only $\mathcal{C}_1$ and $\mathcal{C}_2$ in our system. Clique potentials V_c are functions of clique c, which are able to describe local information in the image, and the global likelihood energy can be expressed as the sum of clique potentials in the neighborhood system. In our system, clique potentials are derived from probability density functions. The statistical information of unary features can be described by $V_{\mathcal{C}_1}$ and the structural information of binary features can be statistically represented by $V_{\mathcal{C}_2}$. Hence, the handwritten character recognition problem can be formulated in the MAP-MRF framework, which can incorporate both statistical and structural information of Chinese characters.

3.1. MAP-MRF Framework

In MAP concept, given observation O and class model λ_ω, we classify O to class ω^* by

$$\omega^* = \arg\max_{\omega}\{P(\omega|O, \lambda_\omega)\}, \tag{3}$$

Proceedings of the 2005 Eight International Conference on Document Analysis and Recognition (ICDAR'05)
1520-5263/05 $20.00 © 2005 IEEE

and

$$P(\omega|O,\lambda_\omega) = \frac{P(O|\lambda_\omega)P(\omega)}{P(O)}, \tag{4}$$

where $P(O)$ is a constant normalization factor for all classes ω, and equal prior class probability $P(\omega)$ is usually assumed, then

$$P(\omega|O,\lambda_\omega) \propto P(O|\lambda_\omega), \tag{5}$$

and MAP equation (3) becomes the maximum likelihood (ML) estimation,

$$\omega^* = \arg\max_{\omega}\{P(O|\lambda_\omega)\}. \tag{6}$$

We model each Chinese character by an MRF with a set of states S and parameters λ, where each configuration of S labels the observation O with a probability $P(O,S|\lambda)$. Then we can express $P(O|\lambda)$ in equation (6) for all configurations as

$$P(O|\lambda) = \sum_S P(O,S|\lambda). \tag{7}$$

However, $\sum_S P(O,S|\lambda)$ is a combinatorial problem involving high computational cost. We have to simplify the problem by finding the optimal labeling $S^* = \{s_1^*,\dots,s_m^*\}$ of unknown observation $O = \{o_1,\dots,o_n\}$ with the highest labeling probability:

$$P(O|\lambda) \approx P(O,S^*|\lambda), \tag{8}$$

where

$$S^* = \arg\max_S P(O,S|\lambda). \tag{9}$$

According to Bayesian rule,

$$P(O,S|\lambda) = p(O|S,\lambda)P(S|\lambda). \tag{10}$$

The Hammersley-Clifford theorem establishes the equivalence between the MRF and the GRF [2], so that the conditional probability

$$p(O|S,\lambda) = Z^{-1}e^{-U(O|S,\lambda)} \tag{11}$$

where Z^{-1} is the normalization factor, and

$$U(O|S,\lambda) = \sum_{c\in\mathcal{C}_1,\mathcal{C}_2} V_c(O|S,\lambda) \tag{12}$$

is the *likelihood energy* that represents the likelihood of observation data O given single-state and pair-state clique labels $\mathcal{C}_1$ and $\mathcal{C}_2$. Similarly, the joint probability of states

$$P(S|\lambda) = Z^{-1}e^{-U(S|\lambda)}, \tag{13}$$

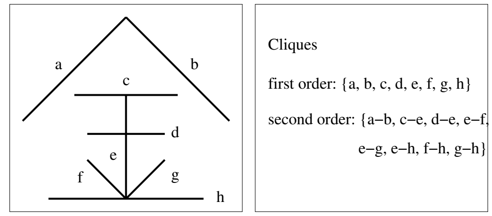

Figure 3. The neighborhood system and cliques based on connectedness.

where

$$U(S|\lambda) = \sum_{c\in\mathcal{C}_1,\mathcal{C}_2} V_c(S|\lambda) \tag{14}$$

is the *prior energy* that represents the relationship of states. According to equation (10),

$$\begin{aligned} U(O,S|\lambda) =& U(O|S,\lambda) + U(S|\lambda) \\ =& \sum_{\mathcal{C}_1}\{\underbrace{V_{\mathcal{C}_1}(O|S,\lambda) + V_{\mathcal{C}_1}(S|\lambda)}_{\text{statistical information}}\} \\ &+ \sum_{\mathcal{C}_2}\{\underbrace{V_{\mathcal{C}_2}(O|S,\lambda) + V_{\mathcal{C}_2}(S|\lambda)}_{\text{structural information}}\}, \end{aligned} \tag{15}$$

where $V_{\mathcal{C}_1}(O|S,\lambda)$ describes the statistical information of unary features given certain set of states, while $V_{\mathcal{C}_2}(O|S,\lambda)$ describes the local structural information of the relationships of binary features given pair-wise states in the neighborhood system. In this framework, given unknown observation O, the recognition task is equivalent to finding the MRF which can minimize the global energy function (15).

3.2. Neighborhood System

In our study, the neighborhood systems and cliques are defined on connection of strokes of Chinese characters. Each natural stroke segment is modeled by a state. The two states are neighbors once the corresponding strokes are connected, which can be represented by a neighborhood weight $a_{jj'}$ as follows

$$a_{jj'} = \begin{cases} 1, & j' \in \mathcal{N}_j, \\ 0, & j' \notin \mathcal{N}_j. \end{cases} \tag{16}$$

Fig. 3 shows an example of Chinese character “金”, where $a - b$ are neighbors and so on.

3.3. Clique Potentials

The statistical information is modeled by single-state clique potentials from unary features. We assume all the observations are conditionally independent, so that equation (11) for single-state clique can be written as a product of probabilities of individual observations:

$$p(O|S,\lambda)=\prod_{i=1}^{n}p(o_i|s_j,\lambda)=Z^{-n}e^{-[\sum_{i=1}^{n}V_{\mathcal{C}_1}(o_i|s_j,\lambda)]}, \tag{17}$$

where the ith unary feature

$$o_i=[C_i,D_i,L_i]. \tag{18}$$

We represent $p(o_i|s_j,\lambda)$ by Gaussian mixture densities:

$$p(o_i|s_j,\lambda)=\sum_{m=1}^{M_s}c_{jm}N(o_i;\mu_{jm},\Sigma_{jm}), \tag{19}$$

where M_s is the number of mixture components, c_{jm} is the weight of the mth component, and $N(;\mu,\Sigma)$ is a multivariate Gaussian with mean vector μ and covariance matrix Σ:

$$N(o;\mu;\Sigma)=\frac{1}{\sqrt{(2\pi)^n|\Sigma|}}e^{-\frac{1}{2}(o-\mu)'\Sigma^{-1}(o-\mu)}. \tag{20}$$

Thus, we have single-state clique potential for statistical information

$$V_{\mathcal{C}_1}(o_i|s_j,\lambda)=-\log\Bigg[\sum_{m=1}^{M_s}c_{jm}N(o_i;\mu_{jm},\Sigma_{jm})\Bigg]. \tag{21}$$

The structural information is modeled by the relationship between neighboring states from binary features. Because of independence assumption and Markovianity property (3), equation (11) for pair-state clique can be rewritten as

$$\begin{aligned}p(O|S,\lambda)&=\prod_{i,i'=1,i\neq i'}^{n}p(o_i,o_{i'}|s_j,s_{j'};s_{j'}\in\mathcal{N}_{s_j},\lambda)\\&=Z^{-n(n-1)}e^{-[\sum_{i,i'=1,i\neq i'}^{n}V_{\mathcal{C}_2}(o_i,o_{i'}|s_j,s_{j'};s_{j'}\in\mathcal{N}_{s_j},\lambda)]}.\end{aligned} \tag{22}$$

The relative location, direction and length between two neighboring stroke segments are stable in Chinese characters regardless of the variations of individual ones from different writers. We can also model $p(o_i,o_{i'}|s_j,s_{j'},\lambda)$ by Gaussian mixtures as follows:

$$p(o_i,o_{i'}|s_j,s_{j'},\lambda)=a_{jj'}\sum_{m=1}^{M_s}c_{jj'm}N(o_{ii'};\mu_{jj'm},\Sigma_{jj'm}), \tag{23}$$

where $o_{ii'}=[C_{ii'},D_{ii'},L_{ii'}]$ is binary feature and $a_{jj'}$ is the neighborhood weight. Thus, we have pair-state clique potentials for structural information if $a_{jj'}\neq 0$,

$$\begin{aligned}V_{\mathcal{C}_2}(o_i,o_{i'}|s_j,s_{j'},\lambda)&=-\log(a_{jj'})\\&-\log\Bigg[\sum_{m=1}^{M_s}c_{jj'm}N(o_{ii'};\mu_{jj'm},\Sigma_{jj'm})\Bigg].\end{aligned} \tag{24}$$

In our system, we consider the prior potential (14) as a penalty if there are no observation o_i aligned to the state s_j denoted as $s_j=0$, then

$$V_{\mathcal{C}_1}(S|\lambda)=\sum_{j=1}^{m}V_{\mathcal{C}_1}(s_j|\lambda)=\begin{cases}\nu_1,\ \text{If}\ s_j=0,\\0,\ \text{otherwise},\end{cases}$$

where ν_1 is a constant. The pair-state prior potential is defined as

$$V_{\mathcal{C}_2}(S|\lambda)=\sum_{j,j'=1}^{m}V_{\mathcal{C}_2}(s_j,s_{j'}|\lambda)=\begin{cases}\nu_2,\text{If}\ s_j=0\ \text{or}\ s_{j'}=0,\\0,\ \text{otherwise},\end{cases}$$

where ν_2 is a constant.

4. Training and Recognition

We use Relaxation labeling (RL) to minimize the global energy of MRF. RL is a class of parallel iterative numerical procedures with polynomial complexity to find the optimal labeling $S^*=\{s_1^*,\ldots,s_m^*\}$ as well as the minimum global energy $U(O,S^*|\lambda)$ [1,2]. Given an RL iteration scheme and the data, two factors affect the solution: the initial assignment of labels and the compatibility function. The posterior MRF compatibility function contains both the prior and observation information, so the dependence of initial labeling is not much. In our system, we initialize by equal labeling of observation to each state. After several iterations, the labeling assignment will be unambiguous. RL terminates if the number of iterations reach a fixed constant value (usually hundreds of iterations) and the winner-take-all strategy will be used.

We use the generalized EM algorithm to estimate parameters of Gaussian mixtures. In training, we firstly initialize a MRF character model using features from a set of standard Chinese characters. The number of states is determined by the number of natural strokes of this Chinese character. After that, we use RL to align each observation to the corresponding state, and then we use all observations to update the parameters of each state. Again we use RL to align each observation to the new updated MRF and use all observations to estimate the parameters of each state. This iteration continues until all parameters of MRF are not changed. In recognition, we use RL to find the optimal labeling of unknown input observations and then classify the observations to the MRF which produce the minimum global energy.

All of the information needed to perform MRF parameter estimation using relaxation labeling is now in place. The steps in this algorithm may be summarized as follows

1. For every parameter requiring estimation, allocate storage for each single-state and pair-state clique potential illustrated by equation (21) and (24). These storage locations are referred as *accumulators*.

2. Calculate the relaxation labeling for each input character observations.
3. For each state j and observation i, use the o_i and the corresponding labeling to update the accumulator for that state.
4. Use the final accumulator values to calculate the new parameter values.
5. If the value of $U(O, S|\lambda)$ for this iteration is not lower than the value at the previous iteration then stop, otherwise repeat the above steps using the new estimated parameter values.

All of the above assumes that the parameters for an MRF are estimated from a single sample of the Chinese characters. In practice, many samples are needed to get good parameter estimates. However, the use of multiple observation sequences adds no additional complexity to the algorithm. Step 2 and 3 above are simply repeated for each distinct training sequences.

Suppose a set of training observations $O^r, 1 \leq r \leq R$, is used to estimate the parameters of a MRF with M_s mixture components. The sequence of states which minimizes $U(O, S|\lambda)$ implies an alignment of training data observations with states. Within each state, a further alignment of observations to mixture components is made by the k-means clustering algorithm. The result is that every observation is associated with a single unique mixture component. This association can be represented by the indicator function $\Theta_{jm}^r(i)$ which is one if o_i^r is associated with mixture component m of state j and is zero otherwise.

The means and variances are estimated via simple averages:

$$\hat{\mu}_{jm} = \frac{\sum_{r=1}^{R} \sum_{i=1}^{n_r} \Theta_{jm}^r(i) o_i^r}{\sum_{r=1}^{R} \sum_{i=1}^{n_r} \Theta_{jm}^r(i)}, \quad (25)$$

$$\hat{\Sigma}_{jm} = \frac{\sum_{r=1}^{R} \sum_{i=1}^{n_r} \Theta_{jm}^r(i)(o_i^r - \hat{\mu}_{jm})(o_i^r - \hat{\mu}_{jm})'}{\sum_{r=1}^{R} \sum_{t=1}^{n_r} \Theta_{jm}^r(i)}, \quad (26)$$

$$\hat{c}_{jm} = \frac{\sum_{r=1}^{R} \sum_{t=1}^{n_r} \Theta_{jm}^r(i)}{\sum_{r=1}^{R} \sum_{t=1}^{n_r} \sum_{l=1}^{M_s} \Theta_{jm}^r(i)}. \quad (27)$$

5. Experimental Results

A public database of handwritten Chinese characters is the ETL9B [4]. In the ETL9B, $2,965$ kinds of Chinese characters (Kanji) and 71 kinds of Japanese characters (hiragana), called the first class of Japanese Industrial Standard (JIS), are included. The characters have been written by 4000 writers. There are 200 samples of each character, so that $607,200$ samples are included in the ETL9B.

In this study, nine pairs of highly similar Chinese characters: “王” and “玉”, “土” and “士”, “金” and “全”, “千” and “干”, “己” and “巳”, “三” and “二”, “永” and “水”, “人” and “入”, “右” and “石”, are used as the recognition vocabulary to study the effectiveness of Markov random fields for Chinese characters modeling. Each character is written by 200 writers with 150 of them used for training and the remaining 50 samples for testing. Two mixtures are used in our MRF model. The recognition rate is presented in Table 1, which is comparable with the best result in [6].

Character	Percentage	Character	Percentage
王	100	玉	98
土	98	士	94
金	90	全	94
千	92	干	100
己	100	巳	92
三	100	二	100
永	94	水	92
人	92	入	96
右	94	石	98

Table 1. Recognition rate of similar characters

6. Conclusions

In the MRF framework, we can model the statistical information by individual states and the structural information by the relationship of states. Because it models the relationship of natural stroke segments, it depends on the extraction of reliable and robust stroke segments. Gabor filter is an ideal strategy to avoid stroke extraction error, which is also considered in our recognition system.

References

[1] J. Cai and Z.-Q. Liu. Pattern recognition using Makov random field models. *Pattern Recognition*, 35(3):725–733, 2002.

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

[3] Z.-Q. Liu, J. Cai, and R. Buse. *Handwriting Recognition : Soft Computing and Probabilistic Approaches*. Springer, Berlin, 2003.

[4] T. Saito, H. Yamada, and K. Yamamoto. On the database ETL9 of handprinted characters in JIS Chinese characters and its analysis. *Trans. IEICE*, J68-D(4):757–764, Apr. 1985.

[5] Y.-M. Su and J.-F. Wang. A novel stroke extraction method for Chinese character using Gabor filters. *Pattern Recognition*, 36(3):635–647, 2003.

[6] Y. Xiong, Q. Huo, and C. Chan. A discrete contextual stochastic model for the offline recognition of handwritten Chinese characters. *IEEE Trans. on Pattern Analysis and Machine Learning*, 23(7):774 – 782, July 2001.