필기체 온라인 인식을 위한 개선된 HMM 알고리즘

마 명*, 박동원**

요약

본 논문에서는 Hidden Markov Model(HMM)을 기반으로 하여 온라인상에서의 육필을 인식하기 위한 개선된 방법을 제안한다. 개인별 상이한 필체에서 오는 문제점을 해결하기 위하여 각기 다른 필체로 분류한 구조적 클러스터링 기법을 사용하였다. 하나의 HMM은 각 계층에 따라 구현되며 한글 문자소에 관한 HMM은 한글 모델에 따라 구현된다. 육필인식 기법 구현을 위하여 효율적인 네트워크 검색철차를 통한 한글조합 규칙을 내포하는 수정 레벨 구축 알고리즘을 사용한다. HMM을 기반으로 하는 기법의 제약을 해결하기 위해 포괄적이고 구조적인 특징에 관한 전처리 방식을 제안하였다. 실험을 통하여 우리가 제안하는 multiple modeling 인식 기법은 영문과 한글 두 필체에 대해서 필자에 대한 독립성이 높다는 것을 알 수 있었으며 또한 다른 HMM기반 인식 기법과 비교하여 우수한 성능을 보이고 있다.

Improving On-line Handwritten Character Recognition with Hidden Markov Model

Ming Ma*, Dong-Won Park**

ABSTRACT

Method for on-line handwritten character recognition with Hidden Markov Model(HMM) is proposed. To deal with the problem of handwriting style variations, the Hierarchical Clustering approach is introduced to partition different writing styles into several classes. One HMM that models temporal and spatial variability of handwriting is constructed based on each class. Therefore a multiple modeling technique is used for both Korean and English characters. The HMMs of Korean graphemes are concatenated to form the Korean character models. The recognition of handwriting characters is implemented by a modified level building algorithm, which incorporates the Korean character combination rules within the efficient network search procedure. Due to the limitation of HMM based method, a post-processing procedure which takes the global and structural features into account is proposed. Experiments showed the proposed recognition system which uses multiple modeling technique, the modified level building algorithm and the post-processing procedure achieved high writer independent recognition rate on unconstrained samples of both English and Korean characters. The comparison with other schemes of HMM-based recognition is also performed to evaluate the system.

Key Words : Online handwriting recognition, hidden Markov model, level building algorithm, Hierarchical Clustering, structural feature

* 배재대학교 게임공학과(✉ming@pcu.ac.kr)
** 배재대학교 게임공학과
· 제1저자(First Author) : 마 명 · 교신저자(Correspondent Author) : 박동원
· 접수일(2011년 1월 3일), 수정일(1차 : 2011년 2월 1일), 게재확정일(2011년 2월 7일)

Ⅰ. Introduction

On-line handwriting recognition, though has been researched over four decades, is still a tough problem. The most prominent problem in handwriting recognition is the vast variation in personal writing styles. There are also a lot of variations within a writing style of one person. These variations depend for example on the context of the writing, writing equipment, writing situation, and the mood of the writer. The writing style may also evolve with time or practice. The performance of the automatic recognition system thus depends heavily on how well the different personal writing styles and their variations are modeled [1,2].

Another research problem in this project concerns Korean character processing. Unlike English, the Korean language has 51 graphemes and a large set of characters. Thus to recognize a give Korean character can be different from, or even more difficult than recognizing English. A Korean character is composed of two or three graphemes which are arranged in two dimensions, not a simple left-to-right concatenation.

Ⅱ. On-line handwriting recognition system

The main goal of this study is to develop a practical on-line recognition system for English and Korean handwritten characters. The system should to be able to handle multiple writing styles and cursive form of handwritten characters. HMM-based recognition model is well suited for this requirement as been verified in many researches. In general, our system can be divided into two parts - training part and recognition part. Though two parts are functionally separated from each other, some components like preprocessing and feature extraction are shared by both of training and recognition parts.

For the training part, it takes a batch of training data as input. After preprocessing and feature extraction, the extracted feature vectors come into the clustering component. Various handwriting styles in a class (represents a letter of grapheme) are grouped and trained into different models through training engine. The models then stored into model database for recognition.

The recognition takes place from inputting the raw data to recognition part. Same preprocessing and feature extraction process as in training part are performed for the raw data. The recognition engine is activated when a feature vector is presented. The final output which is a ranked list of recognition candidates is printed to the screen.

Ⅲ. Model of handwriting

3.1 Ligature and Korean character model

The imaginary lines between strokes are referred as ‘ligatures’ which are modeled explicitly through the pen-up traces. A two-state HMM is designed for ligatures between two graphemes.

There are 24 primitive graphemes in Korean characters. The graphemes can be further classified as consonants and vowels. Ten of the primitive graphemes are vowels and the rest are consonants. Complex vowels and consonants are made by combining simple vowels and simple consonants, respectively [3]. In addition, the primitive graphemes of consonants construct another 16 double consonants, whereas the primitive graphemes of vowel construct another 6 complex vowels. In this study, we proposed a novel Korean character model based on a modified level building network. It consists of a series of grapheme models which are embedded in a structured network according to rigorous composition syntax – grapheme order constraint and structure constraint [3].

3.2 HMMs network

Once the grapheme and the ligature models have been established, a 5-layer finite state network (FSN) designed by Kwon [3], called BongNet is used as the baseline HMMs network, for it effectively represents the architecture of the Korean characters. A set of dummy node – initial node and final node are placed at each end of network. Each path from initial node to the final node corresponds to a Korean character. Each node between first consonant and ligature corresponds to a group of C graphemes with similar ending stroke. The nodes between first ligature and vowel correspond to groups of V graphemes with similar starting stoke. All of the 40 ligature models are represented by second layer and fourth layer.

3.3 Clustering for multiple models design

Due to the fact that handwritten characters can vary in both writing orders and general shapes, using only one model for all patterns of a letter or graphemes may weaken the recognition performance of system. Thus, several models are needed to represent the letters or graphemes with different writing orders or shapes. To this end, a clustering technique is usually applied to obtain multiple models for a single letter or grapheme. The difficulties then lie in how to decide the number of models in advance for top-down clustering, and similarly, how to select a proper distance measure for clustering samples without any additional knowledge or constraints.

3.4 Multiple HMMs design

After introducing the clustering into model design, the next question is how to effectively arrange the models of the same letters or graphemes during matching. Due to the existence of over two models for a letter or grapheme, models of a same letter of grapheme may compete for selection, which will impair the performance of recognition. Therefore, to combine the multiple models of the same letter or grapheme into a single HMM, a multiple parallel-path HMMs [4] is constructed.

A set of dummy nodes – initial and final nodes are placed at both ends of the models of same letters or graphemes. All these models are arranged in parallel and connect to dummy nodes directly. There is no connection between models. Thus each model contributes to one of the multiple paths from the initial node to final node.

Ⅳ. Recognition methods

4.1 Modified level building algorithm

The level building algorithm [5] is used to match a series of HMMs to an observation sequence without having to first segment the sequence into sub-sequences produced by different models. Each level of the level building algorithm corresponds to match a character model to some part of the observation sequence. In our system, a proposed modified level building algorithm is used.

If we denote that the set of V models of grapheme and ligature as λ^v, $1 \le v \le V$, in the Korean character recognition network, and a test sequence of observation O_t, t=1,2,...,T, corresponding to a feature vector which is a sequence of 16-direction codes, then the recognition is to decode O in to the sequence of models$\{\lambda^1,\lambda^2,\cdots,\lambda^l\}$. Namely, it is to match the observation sequence to a state sequence of models with maximum joint probability.

Five levels representing each components of a Korean character are explicitly modeled in level building algorithm. The number of states in each grapheme level is varied according the models of grapheme. The constraints of grapheme order and structure as we introduced for Korean character model are imposed at the end of each level. For each HMM λv, and at each level l, we do a Viterbi match against O, starting from frame 1 on level 1.

4.2 Post-processing methods

Although the HMM based statistical method has many advantages, it still has some limitation caused by the assumptions missing global properties. In some cases, each two graphemes are highly similar to each other a curtain stroke in the grapheme is omitted. In the recognition, since these strokes are nothing but small parts of graphemes, their mismatching hardly influences the likelihood of HMM [6]. Therefore, to solve this problem, the global and structural knowledge about the characters is utilized in a post-processing procedure. Two kinds of the structural analyzers – shape verifier and position verifier are used in this study. It should be noted that the verifiers are only applied in recognition of Korean character.

Shape verifier - Due to the inherent limitation of the Markov assumption, an HMM represents only a sequence of local properties. Because of this, during recognition, the missing of a small important stroke may not cause enough likelihood change of HMM. Therefore, a stochastic grammar [6] is introduced as a shape verifier. To build the stochastic grammar, the representative graphemes in each cluster generated for multiple HMMs design are identified via a sequence of primitive strokes such as vertical line or horizontal bar. All training samples are parsed by stochastic grammar. The frequencies of corresponding state transition is counted and normalized. During verification, the accumulated log probability $P^l(n), 1 \le n \le N$ along the state sequence for each stochastic grammar l is calculated and included in ranking the candidates. If the primitive stroke is not found from the input sequence, probability of ε is added. Thus, the missing of an important stroke will cause a large decrease of $P^l(n)$, which can effectively prevent the mismatching cases.

Position verifier - As previous stated, the global properties are excluded due to the limitation of HMM. Thus a problem of mismatch may occur to the characters who share high similarity of strokes and writing orders. Therefore, a position verifier is proposed to deal with this problem. The main idea of position verifier is to model the positional relationship explicitly among the graphemes of the character in term of the relative position in the bounding box of the character. We first denote the center of the bounding box of the character as pc=(xc,yc) and the center single grapheme as $p_g=(x_g,y_g)$. Thus, the Position verifier Pos is defined as Pos=(D,Θ),where D is the distance between p_c and p_g;Θ is the direction from pc to pg.

Probabilities of DandΘ for graphemes in character samples are calculated by normalized frequencies of the count of DandΘ.

표 1. 각각의 방법에 따른 인식 정확도
Table 1. Recognition accuracy of the models trained with two sampling approaches

Sampling Method	Training Time (Min.)	Recognition accuracy %		
		1-candidate	2-candidate	3-candidate
Equidistant data points	53.6	87.2	91.4	94
Corner detection	47.2	79	83.6	89.2

V. Experiment and Results

The dataset used in this work is derived from the 'KAIST OP2 DB' created by AI Lab at Korean Advanced Institute of Science and Technology (KAIST). This dataset is publicly available for research use from http://ai.kaist.ac.kr. This dataset includes 137,184 characters collected from more than 100 writers who were college and high school students. In our study, part of the dataset including 12,390 English and 45,000 Korean characters are used for training.

First experiment is conducted to test two sampling approaches - equidistant data points and corner detection. Table 1 shows that the model trained with the approach of equidistant data point achieved better recognition accuracy. Because more samples obtained through equidistant data points than corner detection, the training of equidistant data points cost more than corner detection. In the following experiments, models only trained with equidistant data points are used for testing proposed algorithms.

In this system, the number of HMM states for Korean character model were tuned by intuitive and empirical methods. Table 2 shows the character recognition results of these approaches. First, for every model, the fixed number of states ranging from 3 to 16 was tested, and the best performing number was chosen. Second, a half of the average length of observations in the training samples was selected for the number of the state of the class. The proportion, a half, was decided empirically. Last, the number of states was chosen by intuitive knowledge. Note that the intuitive method does not show the best result due to the variety of writing styles in each class. The best performing, the second method is selected for comparison of our proposed method.

표 2. different states 방법에 따른 인식 정확도
Table 2. Recognition accuracy of the models trained with different states

State number	Fixed number	Proportion to length	Manually decided
Recognition accuracy %	87.2	90.1	88.7

As the clustering method introduced for multiple models design, the number of generated models including both English and Korean increased over twice time than single model design. In addition the total number of states also increased by a factor of almost 2. However, by introducing state-tying, the number of states reduced by about half. As a result, the number of the observation distributions becomes almost same as that of single model design. Refer to Table 3.

표 3. 클러스터링 방법 적용 시의 Increased parameters
Table 3. Increased parameters by the clustering method

	Single model	Multiple HMMs	Multiple HMMs after tying
Number of models	159	324	324
total states	1024	2033	1207

For evaluating the proposed design method against the general design method, several tests were performed. First, to examine how well each letter and grapheme HMM was trained, the recognition tests were performed on the letter and grapheme training data. In Table 4 and Table 5, recognition accuracies of the letters and graphemes for two different model designs are listed. It indicates that most models were well trained with some exceptional cases, and the proposed multiple models design performed better than the single model design.

Next, the English letter recognition test was performed on the test data set of English characters. (Table 4)

표 4. 영문자 인식 정확도 비교
Table 4 The comparison of accuracy for English letter recognition

Single model design				Multiple models design			
Lowercase	%	Upper case	%	Lowercase	%	Upper case	%
a	88.2	A	92.0	a	92.6	A	94.1
b	87.3	B	94.8	b	88.0	B	94.8
…							
y	87.5	Y	92.2	y	92.3	Y	95.0
z	76.9	Z	87.9	z	79.0	Z	87.2
Av.	87.2	Av.	90.0	Av.	90.2	Av.	91.2

Finally, the recognition test was performed on the test data set of Korean characters. The result of recognition accuracy is showed in Table 7. The proposed multiple model design improved the recognition rate almost 2% with cost of increased recognition time comparing with single model design. However, the modified level building algorithm achieved an overall improvement in both recognition rate and recognition time. In addition, the shape verifier and position verifier used in post-processing also enhanced the recognition rate with slightly extra time expense.

표 5. 문자소 인식에서의 정확도 비교
Table 5. The comparison of accuracy for grapheme recognition

Single model design						Multiple models design					
Initial	%	Middle	%	Final	%	Initial	%	Middle	%	Final	%
ㄱ	82.	ㅏ	87.	ㄱ	84.4	ㄱ	89.1	ㅏ	87.5	ㄱ	88.3
ㄴ	91.	ㅑ	89.	ㄴ	89.3	ㄴ	94.2	ㅑ	89.9	ㄴ	93.5
…		…		…		…		…		…	
ㅉ	90.	ㅞ	89.	ㄶ	87.3	ㅉ	92.1	ㅞ	94.0	ㄶ	88.6
Av.	90.	Av.	90.	Av.	86.0	Av.	93.5	Av.	91.7	Av	88.2

표 6. 영문자 인식 정확도
Table 6. The recognition accuracy of English characters

Model design	Test data set of English
Single Model	88.6%
Multiple Models	90.7%

표 7. 한글 인식 성능 실험 결과표
Table 7. The performance of Korean characters recognition

Model design	Test data set of Korean		Speed sec/char
	High School	Nation	
Single Model	87.4	88.2	0.476
Multiple Models	89.1	91.8	0.513
Multiple Models+M	93.2	94.4	0.454
Multiple Models+M + P	94	96.3	0.469

M: modified level building; P: post-processing

Ⅵ. Conclusion

The main goals of the research work have been met well. In this study, three contributions have been made. Firstly, a multiple modeling technique based on an effective clustering approach is proposed to deal with the problem of handwriting style variations. Then it introduces a modified level building strategy to incorporate the Korean character combination rules. To overcome the limitation of HMM based method, the global and structural verifiers are introduced in the post-processing procedure.

Based on the results of this research, number of recommendations for future research can be made. Firstly, the improvement can be made via the integration of multiple approaches and the joint effects of processing step. Therefore besides Hidden Markov Models based approaches, more reliable and efficient classifier or verifier is needed to combine with HMM to achieve a further improvement of recognition accuracy. Secondly, since the recognition performance depends heavily on the quality of the model database and size of sample sets, further efforts should be spared in optimizing the model database as well as including more valid sample sets.

References

[1] Vuori, V., "Adaptive Methods for On-Line Recognition of Isolated Handwritten Characters," PH.D. dissertation, 2002.

[2] Brown, P., "The Acoustic Modeling Problem in Automatic Speech Recognition," PhD thesis, Comp. Sci. Dep., Carnegie Mellon Univ, 1987.

[3] Kwon, J. O., Sin, B. K., & Kim, J. H., "Recognition of On-line Cursive Korean Characters Combining Statistical and Structural Models," *Pattern Recognition*, 30, 8, pp 1255-1263, 1997.

[4] Kim, J.J., KIM, J.W. & KIM, J.H., "Data Driven Design of HMM Topology for On-Line Handwriting Recognition," *Seventh Int'l Workshop Frontiers in Handwriting Recognition*, pp. 239-248., 2000.

[5] Rabiner, L. R., & Levinson, S. E., "A speaker-independent, syntax-directed, connected word recognition system based on hidden Markov models and level building," *IEEE Trans. Acousf. Speech Signal Process*., 33, pp. 561-573., 1985.

[6] Kwon, O. S., & Kwon, Y. B.,"An on-line recognition of Hangul handwriting using dynamic generation of line segments," *Proceeding of The 20th KISS Spring Conference*,pp.151-154, 1993.

저자소개

마 명(Ming Ma)

2001.6: Computer Science, Dept., Inner Mongolia Agricultural University, China
2004.2: Info. & Comm. Eng., PaiChai University,(M.S.)
2009.2: Info. & Comm. Eng., PaiChai University,(Ph.D.)
Research interests: Information Recognition, Multimedia, etc

박동원(Dong-Won Park)

1983: Electrical Eng. Dept., Korea University(B. S.)
1985: Computer Science Dept., Florida Institute of Technology(M.S.)
1993: Computer Science Dept., Texas A&M Univ.(Ph.D)
Mar. 1994 ~ Feb. 2004: Professor, School of Information & Communication, Paichai University.
Mar. 2004~Present: Professor, Game Eng. Dept.
Research interests: Networked Multimedia, Image Understanding,