

분할기반 은닉 마르코프 모델과 다층 퍼셉트론 결합 영문수표필기단어 인식시스템

A Segmentation-Based HMM and MLP Hybrid Classifier for English Legal Word Recognition

김계경* · 김진호** · 박희주***

Kye-Kyung Kim*, Jin-Ho Kim**, and Hee-Joo Park***

* 한국전자통신연구원 영상처리연구부

** 경일대학교 전자정보공학과

*** 경일대학교 컴퓨터공학과

요 약

본 논문에서는 분할기반 은닉 마르코프 모델 (segmentation based hidden Markov model)과 다층 퍼셉트론 (multi-layer perceptron) 을 결합한 영문수표 필기단어 (legal word) 인식시스템을 제안하였다. 가변길이의 필기체 영문 단어 분할결과를 인식할 수 있도록 은닉 마르코프 모델을 이용하여 명확한 분할기반 (explicit segmentation-based) 단어단위 (word level) 인식을 구현하고 다층 퍼셉트론을 이용하여 내재적 분할기반 (implicit segmentation-based) 단어단위 인식을 구현하였다. 그리고 이종(heterogeneous)의 두 인식을 새로운 결합 확률추정방식에 따라 결합함으로써 상호 보완 능력을 극대화 시킬 수 있는 영문수표 필기단어 인식시스템을 구현하였다. 제안한 시스템을 캐나다 콘코디아 대학의 CENPARMI 영문 수표 데이터베이스에 적용하여 실험해 본 결과 기존의 연구결과에 비해 비교적 우수한 인식성능을 얻을 수 있었다.

Abstract

In this paper, we propose an HMM(Hidden Markov model)-MLP(Multi-layer perceptron) hybrid model for recognizing legal words on the English bank check. We adopt an explicit segmentation-based word level architecture to implement an HMM engine with nonscaled and non-normalized symbol vectors. We also introduce an MLP for implicit segmentation-based word recognition. The final recognition model consists of a hybrid combination of the HMM and MLP with a new hybrid probability measure. The main contributions of this model are a novel design of the segmentation-based variable length HMMs and an efficient method of combining two heterogeneous recognition engines. Experiments have been conducted using the legal word database of CENPARMI with encouraging results.

Key Words : hidden Markov model, multi-layer perceptron, legal word recognition, hybrid classifier, segmentation based recognition

1. 서 론

은닉 마르코프 모델이 음성인식[1] 분야에 성공적으로 이용됨에 따라 최근 필기체 단어인식에도 흔히 이용되고 있고 또한 성공사례들이[2-8] 많이 발표되고 있다. 그러나 아직까지 매우 우수한 필기체 단어 인식시스템이라도 인간의 인식 능력과는 그 성능면에서 매우 큰 차이를 갖는다. 특히 영문 필기 단어 인식시스템의 중요한 인식 성능저하 요인 중의 하나는 연속적으로 붙은 글자들을 분할 할 때 초래되는 에러에 기인한다. 따라서 이러한 분할에러를 극복하여 우수한 인식 능력을 갖는 영문 필기 단어 인식시스템을 설계하는 연구가

활발하게 진행되고 있다[9-13].

은닉 마르코프 모델의 장점 중에 하나는 신경회로망이나 기존의 분류기들과는 다르게 분할과 인식과정이 따로 분리되지 않고 동시에 수행될 수 있다는 것이다. 따라서 은닉 마르코프 모델 고유의 특징을 이용하면 필기 단어의 과분할 (over-segmentation) 또는 저분할 (under-segmentation) 하에서도 동작할 수 있는 명확한 분할기반 단어 인식시스템을 구현할 수 있다. 또한 다양한 길이의 관측열을 허용할 수 있도록 분할기반 은닉 마르코프 모델의 구조와 매개변수 재추정 알고리즘을 설계하면 일반적인 백터 길이 정규화 과정을 생략할 수 있어서 정규화 과정에서 초래될 수 있는 정보의 손실을 피할 수 있다. 분할기반 은닉 마르코프 모델의 다른 장점은 인식 대상 패턴의 통계적 특징뿐만 아니라 다양한 구조적 특징을 이용할 수 있다는 것이다.

분할기반 단어 단위 은닉 마르코프 모델 인식시스템의 매

접수일자 : 2000년 9월 21일

완료일자 : 2001년 3월 29일

개변수 추정과정에서는 관측 기호 벡터 (observation symbol vector) 를 모델의 상태변수 집합에 정합시키기 위해 벡터요소를 레이블링하는과정을 거치기도 한다[2]. 그러나 인식과정에서는 벡터요소를 레이블링 할 수 없기 때문에 인식시스템의 일관성을 저하시킬 수 있다. 따라서 관측 기호 벡터요소를 레이블링하지 않고 은닉 마르코프 모델의 매개변수들을 재추정 한다면 인식시스템의 일관성을 높여서 인식 성능을 향상시킬 수 있다.

한편, 다층 퍼셉트론은 필기체 단어인식에 흔히 사용되는 분류기들 중의 하나이다[6]. 다층 퍼셉트론은 패턴의 특징 표현방식과 분류방식에 있어서 은닉 마르코프 모델과는 매우 다르게 동작한다. 특히 다층 퍼셉트론은 내재적 분할기반 통계적특징을 이용한 영문 필기체 단어인식에 매우 효과적으로 이용될 수 있다. 따라서 은닉 마르코프 모델과 다층 퍼셉트론을 결합한 영문 필기체 단어인식시스템을 구현하면 인식대상 패턴의 명확한 분할기반 특징벡터와 내재적 분할기반 특징벡터를 동시에 고려할 수 있어서 인식능력을 극대화 시킬 수 있을 것이다[17].

본 논문에서는 가변길이의 필기체 영문 단어 분할결과를 인식할 수 있는 은닉 마르코프 모델을 제안하고 새로운 결합 확률 추정방식을 도입하여 잠재적 분할기반 단어단위 다층 퍼셉트론과 결합함으로써 영문수표에 필기된 금액 단어를 인식하는 영문수표 인식시스템을 구현하였다. 명확한 분할기반 단어단위 은닉 마르코프 모델을 설계하여 단어의 기본적 통계적특징 뿐만 아니라 분할된 주요 문자 (key character) 들의 구조적 특징들을 이용할 수 있도록 하였다. 기존의 은닉 마르코프 모델의 상태전이 모델과 매개변수 재추정방식을 수정하여 정규화되지 않고 레이블링되지 않은 가변 길이의 관측 기호 벡터들을 허용할 수 있도록 하였다. 또한 각 단어부류의 다중 코드북과 각 단어의 다중 관측 기호 집합들을 추출하여 매개변수의 확률값을 보다 높은 신뢰도하에서 재추정할 수 있도록 하였다. 특히 은닉 마르코프 모델과 다층 퍼셉트론을 효과적으로 결합하기 위해 두 분류기의 출력을 기반으로 새로운 결합확률을 추정하는 방식도 소개하였다.

제안한 시스템을 캐나다 콘코디아대학 CENPARMI 연구소의 영문 수표 데이터베이스에 적용하여 인식 실험을 수행하고 그 결과를 검토 고찰하였다.

2. 분할기반 은닉 마르코프 모델

본 논문에서는 분류대상 영어단어 각 부류별로 독립적인 단어 단위 분할기반 은닉 마르코프 모델을 구현하였다[13]. 단어단위 은닉 마르코프 모델은 문자단위 은닉 마르코프 모델을 연결하여 구현하였고 문자단위 은닉 마르코프 모델의 각 상태 천이 링크들이 단어의 분할 결과로 생성된 문자소 (grapheme) 들과 정합(mapping) 되도록 하였다.

각 단어 부류는 각각 다른 개수의 문자들을 이루어지므로 각 부류별 은닉 마르코프 모델의 길이도 각각 다르게 정의된다. 미지의 단어 영상으로부터 추출된 임의의 길이의 정규화되지 않은 피쳐 벡터를 입력 받을 수 있도록 가변 구간 (variable duration) 단어 단위 은닉 마르코프 모델을 설계 구현하였다. 또한 각 단어 부류별로 다중 코드북을 생성하고 한 단어에 대해서는 다중 기호 피쳐 벡터들을 추출하여 모델의 매개 변수들을 재추정 할 수 있도록 하였다.

2.1 은닉 마르코프 모델 구조

전술한 단어단위 은닉 마르코프 모델의 조건들을 만족시키기 위하여 일반적인 은닉 마르코프 모델에서의 기호의 관측위치를 수정하여 상태천이 도중에 출력확률을 관측할 수 있도록 하였다. 즉, 단어로부터 분할된 각 문자소들을 각 상태에 정합 시키지 않고 이들이 상태천이 링크에 정합 되도록 하였다. 또한 분할결과 추출된 다양한 길이의 관측벡터들에 대해 보다 효과적으로 확률을 추정할 수 있도록 하기 위해서 관측확률을 출력하지 않는 천이인 널상태 천이 (null state transition) 도 허용하였다[13]. 단어 단위 은닉 마르코프 모델에서 가능한 상태천이 경로의 수를 증가시켜 다양한 조합의 단어 분할 결과들이 정합될 수 있도록 하였다. 그림 1에 두 개의 문자단위 은닉 마르코프 모델이 접속된 단어 단위 은닉 마르코프 모델의 예를 도시하였다.

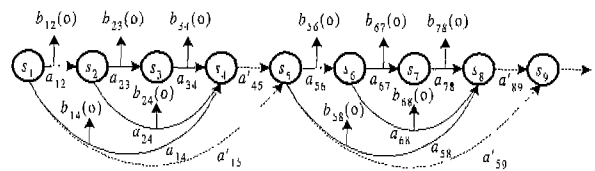


그림 1. 두개의 문자 단위 은닉 마르코프 모델이 접속된 단어 단위 은닉 마르코프 모델의 예

Fig. 1. Word level HMM with two cascaded character level FIMMs.

그림 1에서 각 문자 단위 은닉 마르코프 모델은 네 개의 상태 S_i 를 가지며 일반적인 상태천 a_{ij} 이 와 함께 생략된 문자를 표시하기 위한 널 상태천이 a'_{ij} 등을 포함하는 네 개의 각각 다른 천이경로를 가진다. 하나의 상태천이는 분할된 하나의 문자소를 나타내기 때문에 제시한 모델에서는 하나의 문자가 최대 세 개의 각각 다른 문자소들로 분할 되는 것이 허용된다. 따라서 주어진 관측 기호 벡터 O_T 에 대해 관측 시간 t가 증가함에 따라 상태천이 a_{ij} 동안에 관측기호를 출력할 확률 b_{ij}(o_t) 이 관측된다. 그림에서 점선으로 표시된 널 상태천이 a'_{ij} 의 경우 출력확률이 관측되지 않으며 이는 저분할된 단어영상에서 생략된 문자를 나타낸다. 따라서 고정된 길이의 하나의 천이 경로를 따라 상태가 천이 되는 도중에 널 상태천이를 만나게 되면 관측시간은 증가되지 않고 단지 상태만 그 다음 상태로 이동하게 된다.

2.2 은닉 마르코프 모델의 확률

먼저 주어진 은닉 마르코프 모델 λ의 관측열 O = O₁ O₂ ... O_T 에 대한 O의 생성확률 P(O | λ) 을 계산한다 [1]. 그림 1에 도시한 단어 단위 은닉 마르코프 모델의 경우 초기상태에서 최종상태까지 사이에는 다양한 수의 상태천이 경로가 존재한다. 그 중에서 고정된 하나의 천이 경로 Q = q₁ q₂ ... q_T 에 대한 관측열의 생성확률은

$$P(O | Q, \lambda) = \prod_{t=1}^{T-1} P(O_t^1, O_t^2, O_t^3, \dots, O_t^p | q_{t+1}, q_t, \lambda) \quad (1)$$

$$= \sum_{t=1}^{T-1} \left\{ \sum_{p=1}^p b_{q_t, q_{t+1}}(O_t^p) \right\} \quad (2)$$

으로 정의된다. 여기서 p는 하나의 단어 영상에서 추출된 다

중 퍼져 벡터들의 개수를 의미한다. 그리고 고정된 상태천이 경로의 확률확률은

$$P(Q, \lambda) = \pi_{q_1} \sum_{q_2=1}^{T-1} P(q_{t+1} | q_t) = \pi_{q_1} \sum_{q_2=1}^{T-1} a_{q_1, q_2} \quad (3)$$

으로 정의된다. 여기서 π_{q_1} 는 초기상태가 q_1 에 있을 확률을 의미한다. 기호벡터 O 와 천이경로 Q 의 결합확률, 즉 O 와 Q 가 동시에 발생할 확률은

$$P(O, Q | \lambda) = P(O | Q, \lambda) P(Q, \lambda) \quad (4)$$

이 된다. 따라서 모든 가능한 상태천이 경로 q 들에 대해 결합 확률들을 합하여 관측열의 출력확률 $P(O | \lambda)$ 을 다음과 같이 구할 수 있다.

$$P(O | \lambda) = \sum_{\text{for all path } Q} P(O, Q | \lambda) \quad (5)$$

$$= \sum_{\text{for all path } Q} P(O | Q, \lambda) P(Q, \lambda) \quad (6)$$

$$= \sum_{\text{for all path } Q} \left\{ \prod_{t=1}^{T-1} \left(a_{q_t, q_{t+1}} \prod_{p=1}^P b_{q_t, q_{t+1}}(O_t^p) \right) \right\} \quad (7)$$

학습단계에서는 식 (7)로 정의된 확률값을 최대화 시키기 위한 방향으로 각 단어 부류 i 의 은닉 마르코프 모델 λ_i 에 대해 매개변수 재추정과정이 진행된다. 미지의 단어영상에 대한 인식 단계에서는 추출된 관측열 O 에 대해 모든 단어 부류의 은닉 마르코프 모델들의 확률 중에서 최대 추정확률을 가지는 부류 λ^* 을 다음과 같이 구한다.

$$\lambda^* = \arg \max_{\lambda} P(O | \lambda) \quad (8)$$

그러나, 식 (7)에 정의된 확률 $P(O | \lambda)$ 을 구하는 것은 매우 계산 집약적인 복잡한 과정이 요구된다. 이를 효율적으로 계산하기 위해 잘 알려진 전향 알고리즘 (forward algorithm)[1]을 이용하였다. 전향 알고리즘에 사용되는 전향변수 $\alpha_t(i)$ 는 다음과 같이 정의된다.

$$\alpha_t(i) = P(O_1, O_2, O_3, \dots, O_T, q_t = s_i | \lambda) \quad (9)$$

식 (9)는 주어진 은닉 마르코프 모델 λ 에 대해 시간 t 에서 부분 관측열 $O_1 O_2 \dots O_t$ 까지 관측하고 상태가 s_i 에 도달할 확률을 의미한다. 식 (9)로 표현되는 전향변수 $\alpha_t(i)$ 의 초기값은

$$\alpha_1(j) = \begin{cases} 1.0 & \text{for } j=1 \\ \sum_{i=1}^N a'_{ij} \alpha_1(i) & \text{for } j=2,3,\dots,N \end{cases} \quad (10)$$

으로 정의된다. 본 논문에서 설계한 은닉 마르코프 모델의 구조는 왼쪽에서 오른쪽으로만 천이하는 우향 모델 (left-right model)을 기반으로 하였으므로 가능한 초기 상태는 왼쪽 첫번째 상태인 s_1 이 유일하다. 여기서 전향변수 $\alpha_t(i)$ 의 해는 다음과 같이 귀납적으로 구할 수 있다.

$$\alpha_t(j) = \sum_{i=1}^N \left[a_{ij} \sum_{p=1}^P \{b_{ij}^p(O_{t-1})\}^{w_p} \alpha_{t-1}(i) + a'_{ij} \alpha_t(i) \right] \quad (11)$$

식 (11)에서 p 는 관측 기호 벡터들의 개수이고 w_p 는 단일 단어영상으로부터 추출한 각 벡터들에 대해 상대적인 중요도를 반영할 수 있는 가중치 계수이다. 제안된 은닉 마르코프 모델이 우향 모델이기 때문에 $\alpha_t(i)$ 의 계산은 j 와 t 가 증가되는 방향으로 수행된다.

한편 식 (11)에 포함된 a_{ij} 와 $b_{ij}(O_t)$ 항은 1보다 적은 값을 가지기 때문에 t 가 증가함에 따라 각 $\alpha_t(j)$ 항은 지수적으로 감소하게 된다. 따라서 충분히 큰 t 에 대해서는 $\alpha_t(j)$ 가 컴퓨터의 표현 범위를 벗어나는 언더 플로우 (underflow)가 발생할 수 있다[1]. 이러한 언더 플로우를 방지하기 위해서 전향변수 $\alpha_t(j)$ 의 계산 알고리즘을 일부 수정하였다. 즉, 새로운 $\alpha_t(j)$ 의 계산은 이전의 $\alpha_t(j)$ 를 정규화 한 $\hat{\alpha}_t(j)$ 를 대상으로 수행하였다.

$$\alpha_t(j) = \sum_{i=1}^N \left[a_{ij} \sum_{p=1}^P \{b_{ij}^p(O_{t-1})\}^{w_p} \hat{\alpha}_{t-1}(i) + a'_{ij} \hat{\alpha}_t(i) \right] \quad (12)$$

$$\hat{\alpha}_t(j) = \frac{\alpha_t(j)}{\text{scale}(t)} \quad \text{where } \text{scale}(t) = \sum_{j=1}^N \alpha_t(j) \quad (13)$$

여기서 가능한 유일한 종료 상태는 s_N 이므로 확률 $P(O | \lambda)$ 는 전향 변수를 이용하여 다음과 같이 구할 수 있다.

$$P(O | \lambda) = \alpha_T(N) \quad (14)$$

확률 $P(O | \lambda)$ 는 매개변수 추정단계에서 이용될 뿐만 아니라 최적 천이경로 탐색과정에서도 이용된다.

한편 후향 알고리즘 (forward algorithm)에 이용되는 후향변수 $\beta_t(i)$ 는

$$\beta_t(i) = P(O_{t+1}, O_{t+2}, O_{t+3}, \dots, O_T, q_t = s_i | \lambda) \quad (15)$$

로 정의된다. 식 (15)로 정의되는 후향변수는 주어진 은닉 마르코프 모델 λ 에 대해 시간 t 에서의 상태가 s_i 이고 시간 $t+1$ 에서 종료상태까지의 부분 관측열 $O_{t+1} O_{t+2} O_{t+3} \dots O_T$ 의 출력 확률이다. 전향 변수 계산과정에서 사용된 정규화 방식을 후향변수 $\beta_t(i)$ 의 계산에 그대로 적용하여 다음과 같이 후향변수 $\beta_t(i)$ 를 계산할 수 있다

$$\beta_t(i) = \begin{cases} 1.0 & \text{for } i=N \\ \sum_{j=1}^N a'_{ij} \beta_T(j) & \text{for } i=1,2,3,\dots,N-1 \end{cases} \quad (16)$$

$$\beta_t(i) = \sum_{j=1}^N \left[a_{ij} \sum_{p=1}^P \{b_{ij}^p(O_t)\}^{w_p} \beta_{t+1}(j) + a'_{ij} \beta_t(i) \right] / \text{scale}(t) \quad (17)$$

2.3 수정형 Viterbi 디코딩 알고리즘

Viterbi 알고리즘은 인식단계에서 주어진 관측 기호열과 은닉 마르코프 모델에 대해 최적 상태천이 경로를 찾아 그 출력 확률을 구하는데 사용된다. 일반적으로 Viterbi 알고리즘의 확률추정에 사용되는 확률변수는 다음과 같다[1].

$$\phi_t(i) = \max_{q_1, q_2, \dots, q_{t-1}} \{ \log P(q_1, q_2, \dots, q_t = s_i, O_1 O_2 \dots O_t | \lambda) \} \quad (18)$$

식 (18)에 정의된 $\phi_t(i)$ 는 시간 t 에서의 상태가 s_i 일 때

부분 관측열 $O_1 O_2 \dots O_t$ 에 대해 그 이전의 모든 경로들 중에서 최고확률을 갖는 경로의 확률을 의미한다. 본 논문에서는 정규화 과정을 피하고 계산의 복잡성을 줄이기 위해 $\phi_t(i)$ 의 계산에 로그연산을 사용하였다.

단어 단위 은닉 마르코프 모델에 대한 Viterbi 알고리즘의 상태 천이 흐름도의 일부를 그림 2에 도시하였다. 점선은 널 상태천이를 표시하며 시간의 증가는 없고 단지 상태만 다음으로 이동한다는 것을 의미한다.

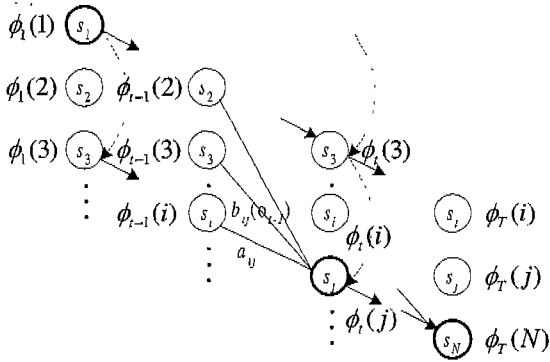


그림 2. 수정형 Viterbi 알고리즘의 흐름도
Fig. 2. The modified Viterbi algorithm flow.

제안한 은닉 마르코프 모델의 구조와 동작방식에 적합하도록 수정한 Viterbi 알고리즘은 다음과 같다.

Step 1: Initialization

$$\phi_1(1) = 1.0 \quad (19)$$

$$\psi_1(1) = 1 \quad (20)$$

$$\phi_1(j) = \max_{1 \leq i \leq N} [\log a'_{ij} + \log \phi_1(i)] \quad (21)$$

$$\psi_1(j) = \arg \max_{1 \leq i \leq N} [\log a'_{ij} + \log \phi_1(i)] \quad (22)$$

Step 2: Recursion from time $t=2$ to T and $j=1$ to N

$$\phi_t(j) = \max_{1 \leq i \leq N} [\log a_{ij} + \log \sum_{p=1}^P \{b_{ij}^p(O_t)\}^{\alpha_p} + \log \phi_{t-1}(i), \log a'_{ij} + \log \phi_{t-1}(i)] \quad (23)$$

$$\psi_t(j) = \arg \max_{1 \leq i \leq N} [\log a_{ij} + \log \sum_{p=1}^P \{b_{ij}^p(O_t)\}^{\alpha_p} + \log \phi_{t-1}(i), \log a'_{ij} + \log \phi_{t-1}(i)] \quad (24)$$

Step 3: Termination

$$\log P^* = \max_{1 \leq i \leq N} \phi_T(N) \quad (25)$$

$$q_T^* = N \quad (26)$$

Step 4: State path backtracking from $t=T-1$ to 1

$$q_t^* = \psi_{t+1}(q_{t+1}^*) \quad (27)$$

위 알고리즘에서는 작은 계산량과 산술계산 에러를 줄이기 위해서 로그연산자를 사용하여 최종 확률 $\log P^*$ 을 계산하였다[1]. 따라서 위 알고리즘을 이용하면 은닉 마르코프 모델의 초기 상태에서 최종 상태까지 모든 천이 경로들에 대해 최대 확률을 가지는 하나의 천이 경로를 효과적으로 추정할 수 있다.

2.4 학습 알고리즘

은닉 마르코프 모델의 모든 매개 변수들을 신뢰도 높게 추정하기 위해서 충분히 많은 학습 패턴을 사용해야 하므로 각 단어 부류에 대해 다중 코드북을 생성해서 학습시킨다. 또한 분류 능력을 높이기 위해 하나의 단어영상에 대해 다중 피쳐 벡터들을 추출하여 매개변수 재추정에 이용한다. 따라서 각각 P 개의 피쳐 벡터를 갖는 K 개의 코드워드로 구성된 학습데이터 집합은

$$O = \{ \{O^1, O^2, \dots, O^P\}^{(1)}, \{O^1, O^2, \dots, O^P\}^{(2)}, \dots, \{O^1, O^2, \dots, O^P\}^{(K)} \} \quad (28)$$

으로 표현된다. 여기서 모든 벡터들은 상호간에 서로 독립이라 가정한다. 따라서 은닉 마르코프 모델 λ 의 기호열 출력 확률 $P(O|\lambda)$ 를 식(28)에 따라 다시 표현하면

$$P(O|\lambda) = \prod_{k=1}^K \prod_{p=1}^P P\{(O^p)^{(k)}|\lambda\} \quad (29)$$

이 된다. 은닉 마르코프 모델의 매개변수들을 재추정하기 위해 두 개의 확률 추정함수[1,3]를 도입하였다. 그 중 하나는 $\xi_t^1(i, j)$ 로서 이는 상태 s_i 에서 상태 s_j 로 천이 될 때 기호의 관측확률이 출력되고 시간 t 에 상태 s_i , 시간 $t+1$ 에 상태 s_j 에 있을 확률을 의미한다. 다른 하나는 $\xi_t^2(i, j)$ 로서 기호의 관측확률이 출력되지 않고 시간 t 에 상태 s_i , 시간 $t+1$ 에 상태 s_j 에 있을 확률을 의미한다.

$$\xi_t^1(i, j) = \frac{\alpha_t(i) a_{ij} \sum_{p=1}^P \{b_{ij}^p(O_t)\}^{\alpha_p} \beta_{t+1}(j)}{P(O|\lambda)} \quad (30)$$

$$\xi_t^2(i, j) = \frac{\alpha_t(i) a'_{ij} \beta_t(j)}{P(O|\lambda)} \quad (31)$$

또한 Viterbi 알고리즘의 전개를 위해 또 다른 확률함수 $\gamma_t(i)$ 가 사용된다. 이는 은닉 마르코프 모델 λ 와 관측 기호열이 주어졌을 때 시간 t 에 상태 s_i 에 있을 확률을 의미한다.

$$\gamma_t(i) = \frac{\alpha_t(i) \beta_t(i)}{P(O|\lambda)} \quad (32)$$

한편 일반적으로 하나의 관측열에 대한 은닉 마르코프 모델의 매개변수 재추정에 가장 흔히 사용되는 Baum-Welch 공식[1]을 위에서 언급한 식 (30), (31) 및 (32)을 이용하여 다시 표현하면 다음과 같다.

$$\overline{a_{ij}} = \frac{\sum_{t=1}^T \xi_t^1(i, j)}{\sum_{t=1}^T \gamma_t(i)} \quad \text{and} \quad \overline{a'_{ij}} = \frac{\sum_{t=1}^T \xi_t^2(i, j)}{\sum_{t=1}^T \gamma_t(i)} \quad (33)$$

$$\overline{b_{ij}^p(k)} = \frac{\sum_{t=1}^T \delta(O_t^p, v_k^p) \xi_t^1(i, j)}{\sum_{t=1}^T \xi_t^1(i, j)} \quad (34)$$

식 (33) 및 (34)를 이용하면 다중 코드워드와 다중 피쳐 벡터를 이용하여 매개변수들을 재추정하는 공식을 유도할 수 있다.

본 논문에서는 다중 학습집합에 대한 여러 가지 매개변수 재추정 방식 중에서 매개변수 평균 방식을 이용하여 은닉 마

르코프 모델의 매개변수들을 재추정 하였다. 이 방식은 음성 인식에 흔히 사용되는 Rabiner의 방법에 비해 필기단어 인식에 보다 효과적으로 이용될 수 있다[13]. 따라서 각 부류를 대표하는 은닉 마르코프 모델의 매개변수들은 동일한 단어 부류에 대해 구한 식 (33) 및 (34)의 결과를 산술 평균하여 구할 수 있다.

$$\overline{a_{ij}} = \frac{1}{m} \sum_{k=1}^m a_{ij}^k, \overline{a'_{ij}} = \frac{1}{m} \sum_{k=1}^m a'_{ij}{}^k, \overline{b_{ij}(l)} = \frac{1}{m} \sum_{k=1}^m b_{ij}^k(l) \quad (35)$$

식 (35)에 정의된 평균연산은 한 단어 부류에 속한 모든 단어 패턴들이 자신의 매개변수를 추정할 다음 수행되고 이렇게 평균처리가 된 매개변수들에 대해 각각의 단어 패턴들이 반복적으로 자신의 매개변수를 재추정하게 된다.

3. HMM-MLP 결합

내재적 분할기반에서 추출된 피쳐들을 이용하여 필기체 영문 단어들을 인식하기 위해 다층 퍼셉트론을 설계 구현하였다. 즉, 3층 구조의 다층 퍼셉트론을 구현하고 이를 수정형 역전파 알고리즘 (modified Backpropagation algorithm) 을 이용하여 학습시켰다[13].

이종의 인식기들을 결합했을 때의 상호 보완능력이 동종의 인식기들을 결합했을 때 보다 크다는 장점을 이용하여 두 인식기를 결합하였다. 즉, 확률모델인 은닉 마르코프 모델과 에너지 모델인 다층 퍼셉트론을 새로운 확률 추정방식에 따라 결합하였다.

여기서 인식대상 단어 부류를 $\Psi = \{1, 2, 3, \dots, V\}$ 로 정의하고 미지의 인식 대상 단어의 관측 심볼을 O 로 정의하면 각 단어 부류별 은닉 마르코프 모델의 확률은 $P(O | \lambda_v)$ 이 된다. 한편 인식 대상 단어의 피쳐벡터를 F 로 정의하고 각 뉴런의 출력을 n_v 로 정의하면 $f(n_v, F)$ 는 미지의 패턴에 대한 뉴런 n_v 의 출력이 된다. 다층 퍼셉트론의 사후 확률 (posterior probability) $P(n_v | F)$ 는 각 뉴런의 출력 $f(n_v, F)$ 을 모든 부류에 대해 정규화 시킨 값으로 다시 표현된다. 따라서 두 인식기를 효과적으로 결합한다는 것은 확률 $P(O | \lambda_v)$ 및 $P(n_v | F)$ 의 결합 확률 $P(C_v | O, F)$ 을 극대화 시키는 것으로 볼 수 있다. 본 연구에서는 기존에 많이 사용되고 있는 투표법(voting) 및 선형신뢰도 누적법(Linear Confidence Accumulation)[17]과는 별도로 가중치 부여 곱하기 방법을 제안하였다.

투표법에 의한 인식기들의 결합 방식은

$$P(C_v | O, F) = \max[P(O | \lambda_v), P(n_v | F)] \quad (36)$$

으로 표현되고 전체 부류 $\Psi = \{1, 2, 3, \dots, V\}$ 에 대해 최종 인식 부류의 결정은

$$v^* = \arg \max_{v \in \Psi} P(C_v | O, F) \quad (37)$$

으로 구할 수 있다. 선형 신뢰도 누적법은

$$P(C_v | O, F) = w_1 \times P(O | \lambda_v) + w_2 \times P(n_v | F) \quad (38)$$

으로 계산되며 w_1 와 w_2 는 각 분류기의 성능에 따라 결정된다. 가중치 부여 곱하기 법은 선형 신뢰도 누적법과 비슷하

게 다음과 같이 구할 수 있다.

$$P(C_v | O, F) = P(O | \lambda_v)^{w_1} \times P(n_v | F)^{w_2} \quad (39)$$

본 논문에서는 위에서 설명한 세 가지 다른 방법으로 은닉 마르코프모델과 다층 퍼셉트론을 결합 하고 그 결과를 비교 하였다.

4. 실험 및 고찰

제안한 방식을 이용하여 캐나다 콘코디아 대학 CENPARMI연구소의 영문 수표 데이터베이스를 대상으로 필기체 단어 인식 실험을 수행하였다. CENPARMI 영문 수표 데이터 베이스는 800명 정도의 필기자들이 2,500장의 영문 수표에 기록한 데이터들을 기반으로 구축되었다. 영문 금액단어 데이터는 학습 데이터집합과 테스트 데이터 집합 각각에 대해 32개 부류로 구분되어 있다.

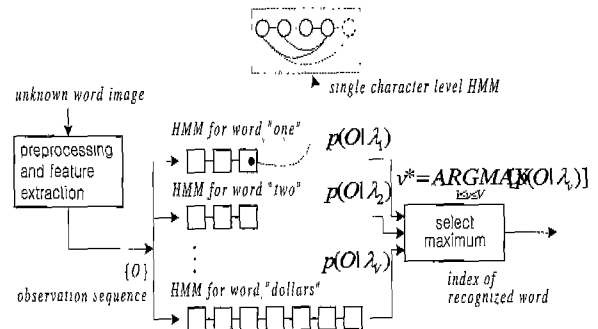


그림 3. 영문 수표금액 단어 인식을 위한 은닉 마르코프 모델의 구조

Fig. 3. Block diagram of the HMM engine for English legal word recognition.

본 연구에서는 학습을 위해 5,224개의 단어를 사용하였고 테스트를 위해 2,482개의 단어를 사용하였다. 제안한 시스템의 성능을 평가 하기 위해서 세 가지 다른 방식으로 실험을 하였다. 즉, 은닉 마르코프 모델을 이용한 인식기, 다층 퍼셉트론을 이용한 인식기 그리고 은닉 마르코프 모델과 다층 퍼셉트론의 결합 (HMM-MLP) 인식기 등을 구현하고 그 성능을 평가하였다.

첫 번째 인식기로 구현한 은닉 마르코프 모델의 블록 다이어그램을 그림 3에 도시하였다.

영문 수표 금액 단어인식을 위해 "and", "dollars" 그리고 "only" 등을 포함한 32개의 단어부류에 대해 32개의 은닉 마르코프 모델을 구성하였다. 각 단어 부류는 각각 다른 개수의 글자들로 이루어져 있기 때문에 각각 다른 개수의 문자 단위 은닉 마르코프 모델들로 구성되었다. 문자 단위 모델은 네 개의 상태를 가지며 네 개의 내부간 상태전이 연결과 두 개의 외부간 상태전이 연결을 가진다. 피쳐 벡터 생성을 위해 먼저 단어영상을 문자 단위로 분할한 다음, 분할된 문자 소로부터 세 종류의 피쳐 벡터들을 추출하였다[13]. 즉, 형태 피쳐, 방향코드 분포 피쳐 그리고 모멘트 피쳐 등을 추출하여 은닉 마르코프 모델의 관측 기호 벡터 집합을 구성하였다. 모든 부류의 은닉 마르코프 모델들은 해당 학습 데이터

들에 대해 공히 5회의 반복 학습을 시켰다.

두 번째 분류기로서 입력층, 은닉층 및 출력층의 뉴런 개수가 각각 125, 50 및 32개로 구성된 다층 퍼셉트론을 구현하였다. 다층 퍼셉트론의 입력으로서 단어영상으로부터 추출한 100개의 메쉬피쳐와 25개의 교차피쳐를 합한 125개의 피쳐를 사용하였다. 피쳐값들을 입력 뉴런에 적용하기 위해서 0.0부터 1.0사이 값으로 정규화 시켰다. 역전파 학습 알고리즘의 학습률 η 및 모멘트 α 는 각각 0.05 및 0.7로 설정하였다. 다층 퍼셉트론의 학습을 위해 5,224개의 단어를 100회 반복 학습을 시켰다.

표 1. 단어 단위 은닉 마르코프 모델을 이용한 CENPARMI 영문 수표 단어인식 결과

Table 1. CENPARMI English legal word recognition results using word level HMMs.

Classifier		T1	T2	T3	T4	T5
HMM	Training data	93.0	97.6	98.8	99.2	99.4
	Testing data	82.0	90.5	93.6	95.7	96.7
MLP	Training data	100.0	100.0	100.0	100.0	100.0
	Testing data	85.6	93.4	95.6	97.2	97.9

표 2. CENPARMI 영문 수표 데이터 베이스에 대한 기존 연구에서의 인식결과

Table 2. Other recognition results of English legal words of CENPARMI database.

Publish	Classifiers	Train	Test	T1	T2	T3	T4	T5
'97 Cote[9]	Perceptual Model	184	100	87.0	91.9	-	-	94.6
'98 Cote[10]	Perceptual Model	184	2,929	73.6	81.0	-	-	89.4
'95 Guillevic[11]	K-NN	5,322	2,515	71.8	93.6	-	-	94.1
'96 Guillevic[12]	K-NN	5,322	2,515	72.6	84.0	-	-	94.3

은닉 마르코프 모델 및 다층 퍼셉트론 각 인식기를 이용하여 2,482개의 영문 필기단어를 인식해 보고 그 결과를 표 1에 도시하였다. 표 2는 실험에 사용된 데이터 베이스와 동일한 CENPARMI 영문 수표 데이터 베이스에 대해 기존에 발표된 인식 성능이다.

표 1에서 T1, T2, ..., T5 등은 인식기가 출력한 인식결과 후보 문자들의 개수이다. T1은 최고확률을 갖는 하나의 후보만을 인식결과로 인정한 경우이고 T2는 인식기가 출력한 최고 두 개의 후보 중에 하나라도 올바르게 인식이 되면 이를 인식결과로 받아 들이는 경우이다. 기존의 연구 결과와 비교해 보면 K-NN 인식기에 비해 은닉 마르코프 모델이나 다층 퍼셉트론 공히 더 높은 인식률을 나타내고 있다. 그러나 기존의 Perceptual 모델과 비교결과 본 연구와 비슷한 개수의 실험 패턴에 대해서는 성능이 높게 나타났고 작은 규모의 실험 패턴에 대해서는 비교적 낮은 성능을 보여주었다.

은닉 마르코프 모델과 다층 퍼셉트론을 바탕으로 HMM-MLP 결합 인식 시스템을 구현하였다. 미지의 영문 필기 워드 패턴이 입력되면 각각의 인식기에서 각 인식 부류별로 정규화 된 인식확률이 출력된다. 두 인식기로부터 출력된 확률 $P(O|\lambda_i)$ 및 $P(n_i|F)$ 를 바탕으로 식 (36), (38) 및 (39)에 따라 각 부류 C_j 에 대한 결합 확률 $P(C_j|O, F)$

을 구한다. 세가지 다른 결합 방식으로 실험한 인식결과를 그림 4에 도시하였다.

그림 4에서 기존에 발표된 두 종류의 결합 방식에 비해 본 논문에서 제안한 가중치 부여 곱하기 결합방식을 이용한

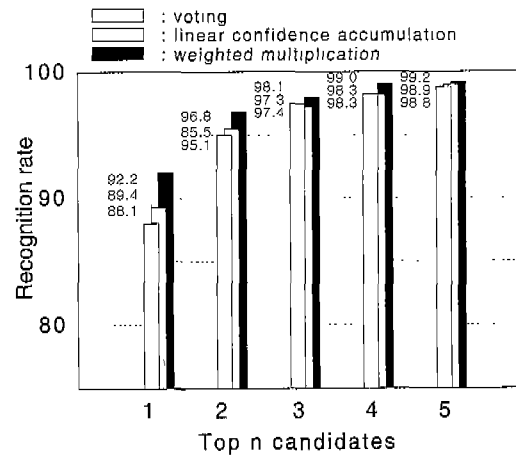


그림 4. 세가지 다른 결합 방법에 따른 영문 필기 단어 인식 성능

Fig. 4. Performance comparisons of three different combination methods.

표 3. HMM-MLP 결합방법에 의한 영문 수표 필기 금액 인식 결과

Table 3. Legal word recognition results of HMM-MLP hybrid model.

Classifiers		T1	T2	T3	T4	T5
HMMMLP	Training data	99.6	99.9	99.9	99.9	99.9
	Testing data	92.2	96.8	98.1	99.0	99.2

표 4. 기존에 발표된 수표의 필기 단어 인식 성능
Table 4. Performance comparisons of the legal word recognition

Classifiers	Class	Train	Test	Rate	Database
'97 Kim ^[4] Dynamic Programming	29	-	-	77.9%	English CEDAR
'98 Han ^[5] Pattern Matching	-	-	-	84.9%	English -
'98 Knerr ^[6] NN-HMM	30	130,000	40,000	92.9%	French A2iA
'98 Guillevic ^[11] HMM-Global	30	4,513	1,622	86.7%	French CENPARMI
'98 Cote ^[10] PERCEPTO Model	32	180	2,929	73.6%	English CENPARMI
'98 Bunke ^[7] HMM	-	-	-	71.9%	Swiss Real Cheques
'99 Soan ^[8] HMM-MRF	26	36,829	4,098	82.5%	French A2iA
Presented HMM-MLP	32	5,223	2,482	92.2%	English CENPARMI

결과가 비교적 우수하게 나타난 것을 확인 할 수 있다. 이 방식을 이용하여 2,482개의 영문 필기체 테스트 단어들을 최종 인식한 결과를 표 3에 도시하였다.

학습된 패턴에 대해서는 99% 이상의 인식률을 보였고 테스트 패턴에 대해서는 92.2% 정도의 인식률을 나타내었다. 표 1에서 보인 바와 같이 두 인식기를 각각 단독으로 사용했을 때의 인식 결과에 비해 상당히 개선되었음을 알 수 있다. 기존에 발표된 수표 금액단어 인식결과를 표 4에 도시하였다.

CENPARMI연구팀의 Guillevic 등[11,12]은 내재적 분할 방식의 은닉 마르코프 모델과 전역 접근 방법을 사용하여 개발한 프랑스어 수표 단어 인식시스템이 86.7%의 인식 성능을 나타내었다고 발표하였다. A2IA연구팀의 Knerr 등[6]은 NN-HMM 결합모델로 30개의 프랑스 수표 필기 단어 클래스에 대해 92.9%의 인식률을 얻을 수 있었음을 발표하였다. 이 결과는 비록 학습패턴 수에 비해 비교적 적은 개수의 테스트 데이터를 사용하였지만 상당히 우수한 결과로 여겨진다.

본 논문에서 제시한 HMM-MLP 결합 시스템이 표 2에서 보는 바와 같이 기존의 CENPARMI 데이터베이스 인식률에 비해서는 상당히 개선된 성능을 보여 주고 있지만 다른 연구 결과들과 직접적으로 비교하기는 어려운 실정이다. 이는 서로 다른 데이터 베이스를 사용하였고 서로 다른 개수의 학습 및 테스트 데이터를 사용하였으며 또한 서로 다른 개수의 단어 부류들을 대상으로 하였기 때문이다. 그러나 본 논문에서 제시한 영문 수표 인식 시스템이 학습패턴의 질 반정도 개수의 테스트 패턴을 사용하였고 대상 부류를 32개로 한 점 등에 비해서 92.20%의 인식률을 얻은 것은 비교적 의미 있는 결과로 여겨지며 지금까지 논문에 발표된 영문 수표 필기 단어 인식성능 중에서는 가장 높게 나타난 것으로 확인되었다.

5. 결 론

본 논문에서는 가변길이의 필기체 영문 단어 분할결과를 인식할 수 있는 은닉 마르코프 모델을 제안하고 새로운 결합 확률 추정방식을 도입하여 잠재적 분할기반 단어 단위 다층 퍼셉트론과 결합함으로써 영문수표에 필기된 금액 단어를 인식하는 영문수표 인식시스템을 구현하였다. 명확한 분할기반 단어 단위 은닉 마르코프 모델을 구현하여 단어의 기본적인 통계적특징 뿐만 아니라 주요 문자들의 구조적 특징들도 이용할 수 있도록 하였다. 그리고 다층 퍼셉트론과의 효과적인 결합 방식도 소개 하였다.

제안한 시스템을 세 가지의 다른 모델로 각각 구현하고 콘코디아 대학 CENPARMI연구소의 영문 수표 데이터베이스를 대상으로 인식실험을 수행하였다. 5,224개 단어를 학습시킨 다음 2,482개 단어에 대해 인식 실험을 수행한 결과 은닉 마르코프 모델의 경우 82.0%의 인식률을 얻을 수 있었고 HMM-MLP 결합 모델의 경우 92.2%의 인식률을 얻을 수 있었다. 궁극적으로 기존의 CENPARMI 영문 수표 데이터베이스를 대상으로 수행 한 연구 결과들에 비해 비교적 개선된 결과를 얻을 수 있었을 뿐 아니라 기존에 발표된 영문 수표 필기단어 인식 성능 보다도 우수한 결과를 나타내었다.

참 고 문 헌

- [1] L.R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proceeding of the IEEE*, vol. 77, no. 2, pp. 257-285, 1989.
- [2] M. Chen, A. Kundu and J. Zhou, "Off-Line handwritten Word Recognition Using a Hidden Markov Model Type Stochastic network," *IEEE Trans. on Pattern Recognition and Machine Intelligence*, vol. 16. No. 5. Pp. 481-496, 1994.
- [3] A. Yacoubi, R. Saburine, M. Gilloux and C.Y. Suen, "Improved Model Architecture and Training Phase in an Off-Line HMM-Based Word Recognition System," *Proceedings of the International Conference of the Pattern Recognition*, Brisbane, Australia, pp. 1521-1525, 1998.
- [4] G. Kim and V. Covindaraju, "Bankcheck Recognition Using Cross Validation Between Legal and Courtesy Amounts," *Automatic Bankcheck Processing, World Scientific Publishing*, pp. 195- 212, 1997.
- [5] K. Han and I.K. Sethi, "An Off-Line Cursive handwritten Word Recognition System and Its Application to Legal Amount Interpretation," *Automatic Bankcheck Processing, World Scientific Publishing*, pp. 295-308, 1997.
- [6] S. Knerr and E. Augustin, "A neural Network -Hidden markov Model Hybrid for Cursive Word Recognition," *Proceedings of the International Conference on Pattern Recognition*, Brisbane, Australia, pp. 1518-1520, 1998.
- [7] G. Kaufmann and H. Bunke, "A System for the Automated Reading of Check Amounts-Some Key Ideas," *Proceedings of the 3rd International Association for Pattern Recognition Workshop on Document Analysis Systems, Nagano, Japan*, pp. 302-315, 1998.
- [8] G. Soan, "Cursive Word Recognition Using a Random Field Based Hidden Markov Model," *International Journal on Document Analysis and Recognition*, vol. 1, no. 4, pp. 199-208, 1999.
- [9] M. Cote, M. Cheriet, E. Lecolinet and C.Y. Suen, "Automatic Reading of Cursive Scripts Using Human Knowledge," *International Conference on Document Analysis and Recognition*, Ulm, Germany, pp. 107-111, 1997.
- [10] M. Cote, E. Lecolinet, M. Cheriet and C.Y. Suen, "Automatic Reading of Cursive Script Using a Reading Model and Perceptual Concepts," *International Journal on Document Analysis and Recognition*, vol. 1, no. 1, pp. 3-17, 1998.
- [11] D. Guillevic and C.Y. Suen, "HMM-KNN Word Recognition Engine for Bank Cheque Processing," *Proceedings of the International Conference on Pattern Recognition*, Brisbane, Australia, pp. 1526-1529, 1998.

- [12] D. Guillevic and C.Y. Suen, "Recognition of Legal Amounts on bank Cheques," *Pattern Analysis & Application*, Springer-Verlag London, vol. 1, pp. 28-41, 1998.
- [13] J.H. Kim, K.K. Kim and C.Y. Suen, "An HMM-MLP Hybrid Model for Cursive Script Recognition", *Pattern Analysis and Applications*, vol. 3, pp. 314-324, 2000
- [14] K.K. Kim, J.H. Kim and C.Y. Suen, "Recognition of Unconstrained Handwritten Numeral String by Composite Segmentation Method," *International Conference on Pattern Recognition*, Barcelona, Spain, vol. 2, pp. 594-597, 2000
- [15] C.Y. Suen, J.H. Kim, K.K. Kim Q. Xu and L. Lam, "Handwriting Recognition - The Last Frontiers," Invited paper - International Conference on Pattern Recognition, Barcelona, Spain, vol. 4, pp. 1-10, 2000
- [16] J.H. Kim, K.K. Kim, Christine P. Nadal and C.Y. Suen, "A Methodology of Combining HMM and MLP Classifiers for Cursive Word Recognition," *International Conference on Pattern Recognition*, Barcelona, Spain, vol. 2, pp. 319-322, 2000
- [17] J.H. Kim, K.K. Kim and C.Y. Suen, "Hybrid schemes of homogeneous and heterogeneous classifiers for cursive word recognition," *International Workshop on Frontiers in Handwriting Recognition*, Amsterdam, Netherlands, pp. 433-442, 2000



김진호(Jin-Ho Kim)

1986년 : 경북대학교 전자공학과(공학사)
 1988년 : 경북대학교 대학원 전자공학과
 (공학석사)
 1992년 : 경북대학교 대학원 전자공학과
 (공학박사)
 현재 : 경일대학교 전자정보공학과 부교수

관심분야 : 패턴인식, 병렬처리
 phone : 053-850-7167
 E-mail : kjinho@bear.kyungil.ac.kr



박희주(Hee-Joo Park)

1978년 : 영남대학교 전자공학과(공학사)
 1981년 : 영남대학교 대학원 전자공학과
 (공학석사)
 1995년 : 대구가톨릭대학교 대학원
 전산통계학과(이학박사)
 현재 : 경일대학교 컴퓨터공학과 교수

관심분야 : 신경회로망, 패턴인식
 phone : 053-850-7283
 E-mail : hjpark@bear.kyungil.ac.kr

저 자 소 개



김계경(Kye-Kyung Kim)

1989년 : 경북대학교 전자공학과(공학사)
 1992년 : 경북대학교 대학원 전자공학과
 (공학석사)
 1997년 : 경북대학교 대학원 전자공학과
 (공학박사)
 현재 : 한국전자통신연구원
 영상처리연구부 컴퓨터비전팀
 선임연구원

관심분야 : 컴퓨터비전, 패턴인식
 E-mail : kyekyung@etri.re.kr