

초기 확률 분포 설정을 통한 은닉 마르코프 모델의 훈련에 관한 연구: 온라인 문자 인식에의 적용

● 김진우, 신봉기, 김진형
포스데이터, 한국과학기술원 전산학과

Training Hidden Markov Model by Giving Initial Parameter Estimates: An Application to On-line Character Recognition

Kim Jin Woo, Sin Bong Kee, Kim Jin Hyung
POSDATA co., Dept. of Computer Science, KAIST

요 약

은닉 마르코프 모델(HMM)의 표현력을 높이는 방법에 대하여 제안하고 이를 온라인 한글 인식에 적용하였다. 일반적인 HMM의 훈련을 위한 초기 확률 분포의 설정이 임의로 주어지거나 균일하게 주어지는데 반해, 본 논문에서는 표현 패턴의 분석을 통하여 모델의 구조를 결정하고 파라미터를 초기화한다.

훈련시킬 글자의 심볼열은 수작업을 통해 일련의 유사한 부분심볼열로 분할한 후, 각 심볼군내에서의 심볼 확률 분포를 구한다. 구해진 심볼 확률 분포를 심볼군의 수와 같은 상태수를 지닌 HMM에서 차례로 각 상태의 초기 관측 심볼 확률 분포로 설정한 후 기존의 훈련 과정을 가진다. 이 방법은 해당 모델의 특정 상태가 특정 심볼군을 표현하도록 유도한다.

실험은 25명이 쓴 "국민 교육 현장" 8933자를 대상으로 하였다. 인식 결과는 기존의 훈련을 거친 모델보다 제안된 훈련 방법을 거친 모델의 사용이 높은 인식률을 가져왔다.

I. 서론

온라인(online) 문자 인식이란 전자펜으로 테블릿등과 같은 장치에 글씨를 쓰면 컴퓨터가 즉시 인식하는 것이다. 그러나, 컴퓨터가 문자를 인식한다는 것은 아직은 어려운 문제에 속한다. 그 이유는 크게 문자간의 정확한 분리가 어렵고, 각 문자가 가지는 많은 변형을 모두 수용하기가 힘들기 때문이다[1].

문자의 분리한 한글 인식의 경우에 한 글자에서 각 자소의 영역을 찾아내는 것인데, 필순, 획수, 획간의 연결, 흘림에 의한 변형에 의하여 정확한 자소의 분리가 어렵기 때문에 종래의 대부분의 인식 방법은 자소간의 분리가 확실한 글자나 약간의 자소의 연결을 허용하는 글자로 인식범위를 제한하여 왔다. 문자들이 올바르게 분리되어도 다시 수많은 변형을 가지는 각 문자들을 인식하여야 한다.

최근 제안된 한글 인식 시스템인 봉넷[3]은 초성, 중성, 종성의 필순을 지켜야 하는 제약에 필기에 제한을 두지 않는 인식 시스템으로 초,중,종성의 기본 자소 외에 흘림 필기시 발생하는 각 자소간의 연결획도 기본자소에 포함시켜 이러한 기본자소의 인식과 함께 한 글자를 인식하는 통계적 네트워크 모델을 구성하였다. 여기서 각 기본 자소는 은닉 마르코프 모델 (Hidden Markov Model)을 사용하여 표현함으로써 문자의 변형을 수용하고자 하였고, 인식 네트워크 안에서 시작 노드와 종료 노드 사이의 최적 경로를 찾음으로써 자소간의 분리와 함께 글자를 인식하는 방법을 사용하였다.

본 논문에서는 이 인식시스템에서 한 글자의 인식능력을 가늠하는 자소 인식 능력의 최대화 방법에 관하여 제안한다. 이러한 목적을 위하여 모델 파라미터의 초기 확률 분포를 설정하여 줌으로써 모델의 표현력을 높일 수 있음을 보이고, 인식 과정시 은닉 마르코프 모델의 상태에서 입력 패턴이 근접화되는 경향이 인식 후보 문

자에서의 순위 결정에 영향을 줄 수 있음을 보인다. 이러한 훈련 방법을 이용한 은닉 마르코프 모델은 한글 인식 이외에 구조적 특성을 가진 패턴의 인식에도 유용하게 쓰여질 수 있을 것이다.

II. 은닉 마르코프 모델 (HMM)

1. 구성

은닉 마르코프 모델은 유한개의 상태와 전이로 구성되고 다음과 같은 파라미터 벡터 λ 로 표시한다.

$$\lambda = (A, B, \Pi),$$

$A = \{a_{ij}\}$ 는 상태 전이($S_i \rightarrow S_j$) 확률 분포.
 $i, j = 1, \dots, N$ (상태 수).

$B = \{b_{ij}(v_k)\}$ 는 전이 ($S_i \rightarrow S_j$)에서의 관측 심볼 확률 분포.
 $k = 1, \dots, K$ (관측 심볼의 종류).

$\Pi = \{\pi_i\}$ 는 초기(시간 $t=0$) 상태 전이 확률 분포.

a_{ij} 는 임의의 시간에 상태 S_i 에 있을 때 다음 시간에 S_j 로 전이할 확률을 말한다. $b_{ij}(k)$ 는 전이 a_{ij} 에서 k 번째 심볼 v_k 를 관측할 확률을 말한다. π_i 는 초기 상태 $t=0$ 일 때 각 상태 s_i 의 확률 분포를 나타낸다. 그림 1은 상태수와 관측 심볼의 수가 3인 HMM을 나타내었다.

2. 인식과 훈련

인식은 그림 2와 같이 일정 개수의 모델에서 주어진 입력패턴과의 정합값이 최대인 모델을 인식 클래스로 한다. 정합값은 입력 패턴이 모델에서의 시작 상태에서 종료 상태까지의 모든 경로를 통

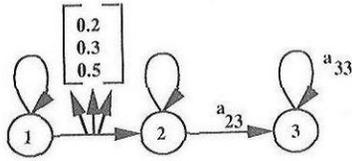


그림 1: HMM의 한 예

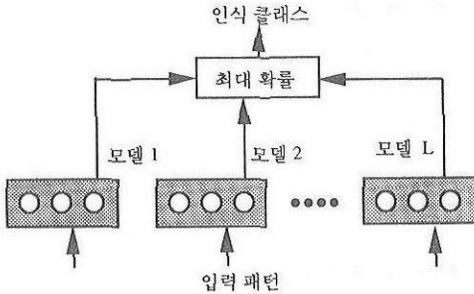


그림 2: HMM을 이용한 인식

하여 가지는 확률로 구하거나 최적 경로를 통하여 가지는 구하는데, 모델이 확장되어 보다 큰 모델을 표현할때 최적 경로를 이용한 인식이 효율적으로 사용된다.

훈련이란 표현하고자하는 입력 패턴의 특성을 모델의 파라미터로 나타내도록 하는 것이다. 그러나, 표현하고자 하는 실세계의 문제의 대부분이 매우 복잡하고, 많은 국부적 최적점(local optima)을 가지고 있어서 모델 파라미터의 전역적인 최적화를 이루기 힘들다. 그러므로 현재까지의 문제를 해결하기 위한 훈련 방법들은 국부적 최적점을 찾는데 만족하고 있다[2]. 훈련도 입력 패턴과 모델과의 정합값을 최대화하는데, 위의 두 방법을 기준으로 하여 실행한다. 일반적으로 훈련은 모든 경로를 이용하는 최대 확률 추정방법을 사용한다. 이에 반해 최적 경로만을 이용한 훈련은 패턴에 대한 모델의 표현력이 좀 떨어지지만, 계산상의 어려움이 적고, 최적 경로를 이용한 인식과 자연스럽게 연결될 수 있다[4][5]. 본 논문에서는 최적 경로에 기반을 둔 인식과 훈련을 한다.

III. 인식력 향상을 위한 훈련 기법

은닉 마르코프 모델을 이용한 패턴 인식의 성패는 주어진 입력 패턴에 대하여 해당 모델이 가장 좋은 확률을 계산해 내는 점에 달려있다. 그러므로 인식 능력 향상을 위해서는 우선 훈련 패턴에 대한 해당 모델의 확률이 타 모델에 비해 높은 값을 가지도록 유도하여야 한다.

1. 초기 확률 분포의 영향

최적 경로를 이용한 최적화 방법을 그림 3에 나타내었다. 주어진 훈련 패턴 X에 대한 해당 모델 λ_i 에서의 최적 경로 S_i 는 전이 확률 분포 $A(\lambda_i)$ 와 관측심률 확률 분포 $B(\lambda_i)$ 로부터 결정된다. 또한 최적 경로 S_i 는 파라미터의 재추정 과정에서 새로운 $A(\lambda_{i+1})$ 와 $B(\lambda_{i+1})$ 를 구하는 기반이 된다. 즉 주어진 모델의 최적화 과정에서 어떠한 입력 패턴에 대한 최적 경로와 모델의 파라미터는 모델이 안정된 상태로 수렴할 때까지 계속 상호 영향을 주게 된다. 그러므로, 초기 파라미터로부터 유도된 초기 최적 경로 S_0 는 계속 파라미터의 재추정에 영향을 주기 때문에, 어떠한 초기 파라미터의 설정이 모델의 표현력을 높이는 데 도움을 줄 수 있는지 아는 것이 중요하다. 이러한 초기 확률 분포 설정에 따른 영향에 관하여 다음 예에서 살펴본다.

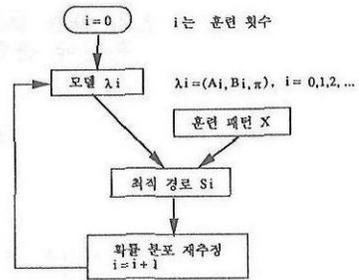


그림 3: 초기 확률 분포의 영향

(예1) 그림 4과 같은 모델 λ 가 주어졌다고 하자.

가정 : 관측 심률 $V = \{a, b\}$.
입력 패턴 $O = "aaaabb"$.

여기서 전이 확률 분포 $A(\lambda)$ 은 균일 확률 분포를 주었는데, 이것은 최적 경로의 결정에 미치는 초기 전이 확률의 역할을 최소화하고자 함이다. 어떤 상태로 들어오는 전이에서 관측하는 심률은 그 상태에서 그 심률을 소비한다고 하자. 비교를 위하여 상태 1로 들어오는 전이에서의 관측 심률 확률 분포만을 특정 확률 분포로 지정하고, 다른 전이에서의 관측 심률 확률 분포는 모두 같게 분배한다. 그림 4에서 각 상태 밑의 컷셋 벡터는 그 상태로의 전이에서의 관측 심률 확률 분포를 나타낸다.

(1) 경우 1

$$B_{11} = \{a = 0.4, b = 0.6\}$$

훈련의 첫번 순환의 최적 경로 : 1222223.

(2) 경우 2

$$B_{11} = \{a = 0.6, b = 0.4\}$$

훈련의 첫번 순환의 최적 경로 : 1111123.

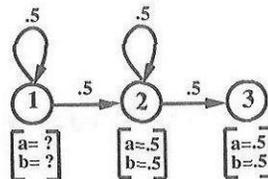


그림 4: 은닉 마르코프 모델의 한 예

최적 경로상에서 경우 (1)은 상태 2에서 대부분의 'a'를 소비하는 반면, 경우 (2)에서는 상태 1에서 모든 'a'를 소비한다. 이러한 결과는 경우 (1)에서는 $B_{11}(a)$ 가 $B_{12}(a)$ 보다 낮기 때문에 최적 경로의 결정에서 상태 1로의 전이는 제외되는 반면, 경우 (2)에서는 $B_{11}(a)$ 이 $B_{12}(a)$ 보다 높기 때문에 심률 'a'를 상태 1에서 소비하는 것이다.

훈련의 효과 측면에서 살펴 보면, 경우 (1)의 조건 하에서 수행될 경우 상태 1에서 상태 2로 가는 전이의 확률은 1이 되어 대부분의 관측 심률을 상태 2에서 소비하도록 파라미터를 결정한다. 이러한 효과는 심률 'a'와 'b'가 상태 2로의 전이에서의 관측 심률 확률을 나눔으로써 결국 주어진 입력 패턴에 대한 모델에서의 확률을 낮추는 요인이 된다. 다음은 훈련을 마친 앞의 모델들이 같은 입력에 대하여 가지는 확률을 보여준다.

• 각 경우에서 훈련후의 확률

$$P("aaaabb", "122223" | \lambda(\text{경우1})) = 0.00864$$

$$P("aaaabb", "1111123" | \lambda(\text{경우2})) = 0.06554$$

결국 최적화가 이루어진 모델에 대한 훈련 패턴의 확률은 경우 (2)에서 8배 정도 좋은 결과를 보이는데, 경우 (2)에서는 상태 1에서

관측 심볼 'a'만을 소비하고, 상태 2에서 관측 심볼 'b'를 소비하게 되어 각 상태에서 소비하는 심볼에 대하여 높은 확률을 가짐으로써 전체적으로 높은 확률을 생성하는 것이다.

2. 표현력

표현력이란 표현하고자 하는 패턴들이 해당 모델에서 높은 확률을 갖도록 하는 것이다. 표현하고자 하는 대상이 구조적 틀을 가진 관측 가능한 대상일 경우 그 대상속에서 유사성을 지니는 부분

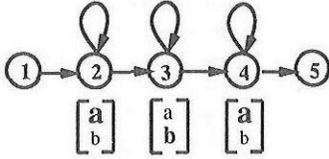


그림 5: 패턴 "aabbbbaa"를 위한 은닉 마르코프 모델

을 은닉 마르코프 모델의 특정 상태에서 소비하도록 유도하여 표현력을 높일 수 있다.

다음과 같은 패턴이 있을 때, 유사한 부분심볼을 기본 단위로 하여 모델의 구조를 구성하여 볼 수 있다.

(예 2) aabbbbaa → aa, bbb, aa

(예 2)에서의 입력 패턴을 3개의 부분패턴으로 나누어서 그림 5과 같은 모델이 되도록 초기 확률 분포를 설정할 수 있다.

그림 5에서 각 상태 밑에 표시한 확률 분포에는 그 상태에서 굵게 표시한 심볼로 구성된 일련의 부분 심볼열을 소비하도록 초기 관측 심볼 확률 분포에 그 심볼의 확률을 높게 설정하여 훈련시킨 것이다. 이러한 표현 방법은 특정한 상태가 특정한 심볼에 관한 소비를 담당하도록 함으로써 한 상태가 특정 심볼의 표현 상태가 됨은 물론, 주어진 입력 패턴에 대한 해당 모델의 확률도 높아진다. 위와 같은 방법을 통한 표현력의 개선은 분별력 향상의 효과를 가져온다. 분별력이란 주어진 입력 패턴이 해당 모델에서 타모델에 비해 높은 확률을 갖도록 하는 것이다. 즉, 어떤 입력 패턴을 인식 모델이 아닌 모델과 정합시킬 때, 그 모델의 특정한 상태에서 관측되어야 할 심볼을 관측하지 못함으로써 정인식을 유도하는 효과를 가지게 된다.

3. 상태 지속의 효과

최적 경로의 결정을 통해 한 입력 패턴이 해당 모델의 각 상태에서 소비된 정도를 알 수 있는데, 이를 상태 지속 정보라고 한다. 이러한 정보는 위와 같은 훈련 과정을 거쳤을 때 각 상태가 특정 심볼군의 길이 정보를 의미하므로, 구별되어야 할 패턴간의 분별의 기준이 유사한 부분 심볼열의 길이로부터 영향을 받을 때 유용하게 사용할 수 있다.

4. 문자 인식에의 응용

이러한 연구 결과를 토대로 문자 인식, 특히 한글의 자소를 표현하는 모델의 훈련에 적용하는 방법에 대하여 살펴본다.

4.1 초기 확률 분포의 설정

한글의 자소는 대부분 직선획의 조합으로 이루어진다. 중성 '아'의 경우를 살펴보면 기본 획은 수직 획과 수평 획으로 이루어진다. 이 두 획 사이에 가상획이 있다고 가정하면, 이산화 과정을 거친 중성 '아'의 이산 심볼열은 자소의 부분획을 표현하는 3개의 부분 심볼열로 나눌 수 있다. 나누어진 부분 심볼열에서 구한 심볼 확률 분포를 중성 '아' 모델의 상태 각각에 수직획의 심볼 확률 분포부터 순차적으로 각 상태로의 초기 관측 심볼 확률 분포로써 설정한다. 이러한 초기화 과정을 거쳐 훈련된 중성 '아'의 모델은 각 상태에서 부분획을 정의하는 모델이 될 것이다. 그림 6은 이러한 방법으로 구성된 중성 '아'의 모델이다. 종료 상태의 채귀 전이 확률이 높아지지 않도록 채귀전이 없는 종료상태를 첨가하였다. 전이에 표시한 획은 그 획을 이루는 심볼의 확률이 높게 설정되었음을 나

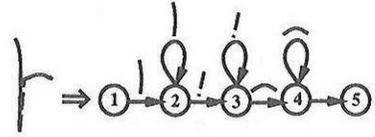


그림 6: 자소 모델

타낸다.

4.2 상태 지속 정보의 첨가

본 논문에서는 실험을 통한 평가에 의해 부분획의 길이의 비율을 상태 지속 정보로 이용하였다. 상태 지속 정보의 적용은 중성 '아'의 경우 수직획의 길이가 수평획의 길이보다 대부분 길다는 정보를 활용하는 것이다.

IV. 한글 인식 시스템

1. 한글 인식 모델

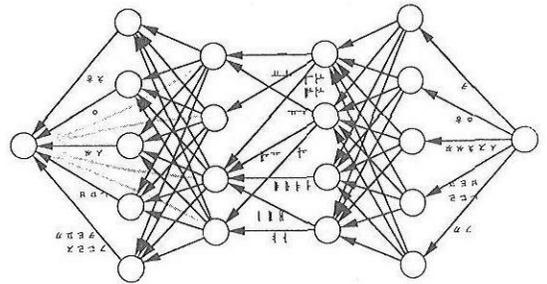


그림 7: 봉넷 - 한글 인식 네트워크

자소 모델은 기본적인 자소를 HMM으로 표현한 것이고, 자소 간에 흐름에 의한 획이나 가상획을 연결 자소로 표현하여 연결확 모델을 HMM으로 표현하였다. 이 두 모델을 단위로 하여 완전한 한 글자를 인식하기 위한 틀이 한글 인식 네트워크(봉넷)이다[3]. 그림 7의 봉넷은 한글 제자 원리를 네트워크 모델로 구성한 것인데, 한글의 모든 가능한 음절을 인식할 수 있다는 점에서 인식이 가능 수 있고, 한글 음절의 모델이라 볼 수도 있다. 글자는 자소간의 연결 확(L)을 고려하면,

$$\begin{aligned} &(\text{초성}) + (L) + (\text{중성}) \\ &(\text{초성}) + (L) + (\text{중성}) + (L) + (\text{중성}) \end{aligned}$$

와 같은 확장된 자소열로 표현할 수 있다. 각 전이 위의 자소들은 HMM을 의미한다. 한 글자를 인식하기 위해서는 시작 노드에서 출발하여 종료 노드로 이르는 여러 경로가 있다. 각 경로는 자소간의 연결 패턴이 유사한 글자를 표현하는데, 각 경로상의 최적의 자소들의 연결로 인식을 하게 된다. 예를 들어 "난"을 인식하기 위해

$$(\text{초성'니은'}) + (L) + (\text{중성'아'}) + (L) + (\text{중성'니은'})$$

에 해당하는 경로가 있어야 하고, 그 경로상에서 각 자소의 확률값이 주어진 글씨에 대하여 통계적으로 가장 높아야 한다. 이러한 인식 방법은 앞 장에서 설명한 각 모델에서의 최적 경로의 결정을 통한 인식 과정과 유사하기 때문에 한 음절의 인식과 그 경로상의 자소의 인식은 자연스럽게 연결될 수 있다.

2. 전처리 및 코드화

전처리 과정에서는 입력 장치로부터 전해진 입력 패턴에 대하여 불필요한 점이나 글씨의 삐침등을 제거하고 글자의 크기를 정규화하는 작업을 한다. 그리고, 점열의 간격을 균일하게 한다. 이러한 점열의 코드화는 방향 코드를 사용하는데, 방향 코드는 360도를

8방향으로 나누고 차례로 번호를 부여한 방향 벡터로써, 가상획의 경우는 진획의 코드값에 8을 더하여 표시한다.

3. 한글 인식 모델의 훈련

훈련을 위해 각 모델을 위한 데이터를 준비해야 하고, 이를 위한 수작업을 필요로 한다. 기본 준비 단계는 다음과 같다.

우선, 각 글자에서 자소별로 분리할 한 후에 분리된 자소에 대하여 전처리와 코드화를 한다. 코드화된 자소의 이산 심볼열에 대하여 앞서 설명한 모델의 훈련 방법을 위한 부분획의 분리를 수행한다. 부분획 즉, 부분 심볼열에서 구하여진 심볼 확률 분포는 해당 모델의 상태로의 전이에 순차적으로 초기 관측 심볼 확률 분포로 설정하고, 초기 전이 확률 분포는 균일 확률 분포로 하여 모델을 훈련시킨다.

V. 실험 및 분석

훈련과 테스트 데이터는 25명의 필기자가 쓴 "국민 교육 현장"을 사용하였다. 훈련을 위하여는 20명의 데이터를 사용하였고, 인식 실험은 25명의 데이터에 대하여 수행하였다. 각 자소 모델은 평균 300개의 해당 자소를 가지고 훈련되었다. 문자 인식에서 인식 네트워크를 거친 확률이 가장 높은 문자가 올바른 인식 결과일 경우 정인식되었다고 한다. 또한, 확률이 높은 상위 5자의 후보 문자 중에서 올바른 인식 결과가 있을 경우 후보 인식되었다고 한다.

1. 자소 인식 실험

본 논문에서 제안한 방법의 효용성을 살펴보기 위한 자소의 인식 실험을 하였다. 인식 실험에 사용한 21개의 중성 자소는 훈련에 사용한 데이터에서 최대 300개를 대상으로 하였다. 실험 결과는 표 1에 나타내었다.

방법	정인식	후보인식
1	90.65%	99.94%
2	84.65%	99.61%
3	96.90%	99.95%
4	97.27%	99.95%

1. 최대 확률 추정 방법
2. 최적 경로를 이용한 훈련.
3. [2]+초기 확률 분포 설정.
4. [3]+상태 지속 정보 첨가.

표 1: 중성 21자에 대한 인식 결과 비교

실험 결과는 본 논문에서 제안한 훈련(방법 3)을 거친 모델이 기존의 방법인 최대 확률 추정 방법에 의한 모델에 비해 좋은 인식률을 가짐을 보여주고 있다. 특히, 이 실험에서 중성 "아"의 인식률은 (1)의 방법에서는 84.0%의 정인식률을 보인 반면, (3)의 방법에서는 99.3%의 정인식률을 보였다. 방법 (4)에서는 상태 지속 정보가 모델간의 분별력을 높이게 되어 정인식에 유용한 정보를 주었음을 이 실험에서 알 수 있다. 전반적인 자소만의 인식 결과는 좋게 나왔는데, 그 이유는 훈련 데이터를 대상으로 실험을 하였고, 자소는 문자에 비하여 단순한 구조를 가지고 있기 때문이다.

2. 문자 인식 실험

문자 인식 실험은 25명이 쓴 8933자에 대하여 실험하였다. 표 2는 인식 실험 결과를 보여 준다. 이 실험 결과에서는 기존

방법	정인식	후보인식
1	74.06%	83.78%
2	81.39%	89.13%
3	84.66%	92.01%

1. 최대 확률 추정 방법.
2. 본 논문에서 제안된 훈련.
3. [2]+상태 지속 정보 첨가.

표 2: 문자 인식 결과 비교

의 훈련 방법을 이용한 (1)의 모델과 본 논문에서 제안한 (2)의 모델을 사용한 인식률의 차이를 명확하게 볼 수 있는데, (2)의 경우는 (1)의 경우보다 정인식률이 7.33% 증가하였다. 이 점은 초기 확률 분포로부터 유도된 모델의 각 상태가 가지는 구분력적인 의미가 효용성을 가지고 있음을 보여준다. (3)에서는 (2)의 모델을 가지고 인식한 후에 상태 지속 정보를 활용하여 높은 인식률을 보여주고 있다. 또한, 이 정보의 활용은 인식중에 얻어진 정보를 이용하므로 인식외에 부과되는 시간의 비용은 거의 없다고 할 수 있다.

결과적으로 기존의 은닉 마르코프 모델의 훈련 방법인 표 1의 (1)의 방법에 비하여 본 논문에서 제안한 훈련 방법과 정보의 활용을 통한 인식률은 정인식에서 최대 10.6%의 문자 인식률의 증가를 보이고 있다.

VI. 결론

본 논문에서는 은닉 마르코프 모델을 이용한 문자 인식에 있어서 초기 확률 분포 설정을 통한 훈련 방법에 관하여 제안하고, 실험을 통하여 이러한 훈련 방법의 유용성을 보였다. 이러한 훈련의 효과는 한 모델에서의 상태에 표현하고자 하는 패턴의 구조적 특성을 부여하게 되고, 이러한 상태가 가지는 특성은 인식에 있어서 좋은 결과를 보여줄 뿐만 아니라 상태 지속 정보를 효과적으로 이용할 수 있는 연결점이 되겠다.

이러한 훈련 방법이 많은 유용함을 가진 반면, 실제 적용에서 큰 효과를 주지 못한 입력 패턴도 있었다. 그것은 자음 "비음"과 같이 필기자의 필기 형태에 따라 매우 많은 변형을 가지는 경우이다. 실험은 하지 않았지만 해당 모델의 구조를 정할 때 여러 경로에서 여러 변형을 수용하도록 유도하는 방법도 가능할 것이라 여겨진다.

본 논문에서 제안된 훈련 방법이 한글 이외에도 구조적 특성을 지니는 입력 패턴에 대한 표현과 인식에 유용하게 이용될 수 있리라 기대한다.

참고 문헌

- [1] C.C.Tappert, C.Y.Suen and T.Wakahara, "The State of the Art in On-Line Handwriting Recognition," *IEEE Trans. on PAMI*, Vol.12, No.8, pp. 787-808, August 1990.
- [2] L.R.Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proc. of the IEEE*, Vol.77, no.2, pp. 257-286, 1989.
- [3] 신봉기, "통계적 방법에 의한 온라인 한글 필기 인식," 제4회 한글 및 한국어 정보처리 학술발표 논문집, pp. 533-542, 1992. 10.
- [4] B.H.Juang, L.R.Rabiner, "The Segmental k -means algorithm for estimating parameters of hidden Markov models," *IEEE Trans. ASSP*, Vol.38, no.9, pp. 1639-1641, Sept. 1990.
- [5] N.Merhav, Y.Ephraim, "Maximum likelihood hidden Markov modeling using a dominant sequence of States," *IEEE Trans. ASSP*, Vol.39, no.9, pp. 2111-2115, Sept. 1991.