

HMM 기반의 오프라인 필기 모델

신봉기*

*부경대학교 멀티미디어공학과
e-mail:bkshin@dolphin.pknu.ac.kr

Off-line Character Modeling using HMM

Bong-Kee Sin*

*Dept of Multimedia Engineering, Pukyong National University

요약

음성 인식 및 온라인 필기 인식기 모델로 널리 알려진 은닉 마르코프 모델(HMM)을 오프라인에 적용하려는 시도는 있었지만 아직까지 만족할 만한 성과는 찾아보기 어렵고 인식률도 신경망 등 다른 방법에 의한 시스템에 미치지 못하는 실정이다. 본 연구에서는 온라인 필기 모델 HMM을 오프라인 필기 인식에 활용하는 방법 한 가지와 순수하게 오프라인 필기 모델로서 제안된 2D HMM을 기술한다. 두 방법 모두 기존의 HMM 모델링 틀에 기초를 두고 개발하였으며 다양한 국소 변형을 해석하기 위해 동적 계획법에 기반한 알고리즘을 응용하였다. 본 논문에서는 두 가지 독립적인 아이디어 제안에 의의를 두었으며 주요 아이디어만을 간략하게 기술하였다.

1. 서론

지금까지 개발된 다양한 인식 방법 중에서 특히 HMM을 이용한 온라인 필기 인식기는 홀림체를 모델링 하는데 뛰어난 모델로 인정받고 있다 [1]. 온라인 필기 문제에서의 성과에 따라 HMM을 오프라인 필기 인식에 적용하려는 시도가 자연스럽게 되었다. 처음에는 HMM의 특성보다는 주로 다양한 방법의 전처리 또는 독특한 데이터 코딩 방법 개발에 주력하는 경향을 보였다. 그러나 온라인 필기의 성과에 반하여 오프라인 필기 인식에서는 HMM으로 별다른 성과를 거두지 못하였다. 다만 이차원 마르코프 체인 모델의 하나인 MRF를 이용한 연구가 저자가 아는 한 유일한 예이다[2]. 엄밀히 말해 MRF는 소수의 파라미터로 2D 평면상의 문자 영상을 모델링하는 단순한 모델이다. 따라서 전역적 특징보다 국소 패턴의 변형을 주 모형화 대상으로 삼아야 하는 필기 패턴 모형으로서의 모델링 능력이 떨어진 다.

본 논문에서는 기존의 온라인 필기 모델을 오프라인 인식에 적용하는 방법 한 가지와 순수하게 오프라인 필기를 위한 모델로서 새로운 구조의 2D

HMM을 소개한다.

첫 번째 방법은 오프라인 필기에서 동적 정보를 복원하는 방법의 하나로 온라인 필기 모델 HMM을 필기 궤적 가이드로 이용한다. 기존의 관련 연구로는 필획의 외형적 패턴 분석에 의한 휴리스틱 방법이 널리 알려져 있다 [3]. 본 논문의 방법은 인식기 측면에서만 본다면 다른 방법에 비해서 다소 효율성이 문제가 될 수 있지만 적당한 전처리를 가함으로써 개선할 수 있을 것이다.

두 번째는 이미 널리 사용되는 MRF나 Pseudo 2D HMM[6]과는 달리 엄밀한 의미에서 1D HMM을 2D 격자 구조의 마르코프 메쉬로 확장한 모델로써 2D 영상 평면의 국소 변형을 자연스럽게 모델링하는 2D HMM이다. 본 논문에서는 간단히 아이디어와 모델의 정의 및 기본 평가 함수를 중심으로 소개한다. 이 모델은 MRF를 일반화한 모델이다.

2. 은닉 마르코프 모델 (HMM)

HMM은 마르코프 체인의 확률 함수로 표현되는 정확률 과정으로 정의한다. 형식적으로 아래와 같은 세 가지의 파라미터로 기술되는 확률적 유한 상태

기계의 하나로 정의한다.

- 상태 천이 확률: $A = \{a_{ij} = P(q_t = j | q_{t-1} = i)\}$
- 심볼 출력 확률: $B = \{b_j(v) = P(x_t = v | q_t = j, \lambda)\}$
- 초기 상태 확률: $\pi = \{\pi_i = P(q_1 = i | \lambda)\}$

패턴 인식 모델로 주어진 입력의 우도를 계산하는 전진/후진 알고리즘은 주로 훈련 알고리즘으로 사용된다. 그러나 HMM의 평가 및 분석 함수로 널리 사용되는 Viterbi 알고리즘은 아래와 같은 정의에서 시작된다.

$$P(X|\lambda) = \max_Q P(X, Q|\lambda) \\ = \max_Q \prod_t a_{q_{t-1}q_t} b_{q_t}(x_t)$$

윗 함수를 효율적으로 계산하기 위해서

$$P(X_{1:t}, Q_{1:t}|\lambda) = P(X_{1:t-1}, Q_{1:t-1}|\lambda) \\ \times P(x_t, q_t | X_{1:t-1}, Q_{1:t-1}, \lambda)$$

와 같이 순환적으로 정의하여 사용한다. 이를 다시 일반적 표기법으로 표시하면 다음과 같다.

$$\delta_t(j) = \max_i \delta_{t-1}(i) a_{ij} b_j(x_t)$$

$$\xi_t(j) = \arg \max_i \delta_{t-1}(i) a_{ij} b_j(x_t)$$

$$j = 1, \dots, N, \quad t = 1, \dots, T.$$

Viterbi 알고리즘은 입력 심볼열을 확률적으로 최적의 상태열로 사상하는 알고리즘이다. 다음 절의 필기 궤적 추적 방법은 바로 이 사상 또는 심볼-상태 배열(alignment)에 기반한 것이다.

3. 필기 궤적 추적 모델

온라인 필기 모델로서 잘 훈련된 HMM은 동적 정보, 즉 필순과 각 획의 궤적을 확률적으로 표현한다. 따라서 만약 주어진 정적 필기 문자 영상을 인식하기 위해 필기의 궤적을 분석하는 방법을 쓴다면 이와 같은 온라인 필기 HMM을 하나의 지식 기반으로 활용할 수 있을 것이다. 본 절에서는 HMM을 이용한 필기 생성 방법에 착안한 방법을 기술한다.

3.1 추적 알고리즘

HMM의 각 상태 $i \in \{1, \dots, N\}$ 는 온라인 필기 문자에서 특정 획의 특정 위치의 특징(예컨대 필획 진행 방향)을 확률 변수로 표현한다. (그림 1 참조) 역으로 $t-1$ 시점 현재 HMM 상태 i 가 필기 영상에서 \mathbf{x}_{t-1} 지점에 대응되었을 때 펜 이동에 따라 다음 시점 t 에 (국소 패턴이 바뀌면서) j 상태

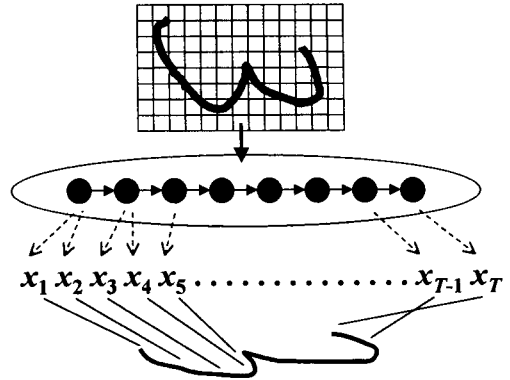


그림 1 필기 궤적 추적 모델

가 \mathbf{x}_{t-1} 에서 c_t 방향으로 일정 거리에 있는 \mathbf{x}_t 에 대응된다고 하자. \mathbf{x}_t 는 일반적으로 현재 필기 중인 문자의 종류, 문자 내에서의 필기 위치에 좌우되며 해당 문자 HMM의 한 상태에 대응된다. 물론 이 때 \mathbf{x}_t 는 펜 궤적상의 점이므로 잉크가 묻어 있는 확상의 점이어야 할 것이다. 이와 같은 확률 과정을 HMM 틀에 맞춰 정의하면 Viterbi 평가 함수는 다음과 같이 쓸 수 있다.

$$\delta_t(j, \mathbf{x}_t) = \max_{i, \mathbf{x}_{t-1}} \delta_{t-1}(i, \mathbf{x}_{t-1}) a_{ij} \\ \times f_j(\mathbf{x}_{t-1}, \mathbf{x}_t)$$

여기서 $f_j()$ 는 상태 j 에서 위치 \mathbf{x}_{t-1} 에서 \mathbf{x}_t 로의 단위 길이 필기 진행 방향 $\mathbf{dx} = \mathbf{x}_t - \mathbf{x}_{t-1}$ 의 확률 분포를 나타낸다.

$$f_j(\mathbf{x}_{t-1}, \mathbf{x}_t) = g(\mathbf{x}_t) b_j(\theta(\mathbf{x}_t - \mathbf{x}_{t-1}))$$

여기서 $g()$ 는 문자 영상 함수로서 위치 \mathbf{x}_t 의 잉크에 관한 영상 함수로서 이웃 픽셀의 가중합으로 정의한다.

$$g(\mathbf{x}) = \sum_{\mathbf{y} \in N_{\mathbf{x}} \cup \{\mathbf{x}\}} w_i I(\mathbf{y}), \quad \sum_i w_i = 1$$

$I()$ 는 픽셀 \mathbf{y} 의 값(밝기 intensity $\in [0, 1]$)을 나타낸다. 한편 함수 $\theta()$ 는 단위 길이의 방향 이동을 인코딩하는 함수이다. 본 연구에서는 방향 코드로 양자화하는 함수로서 다음과 같이 정의한다.

$$\theta(\mathbf{u} - \mathbf{v}) = \theta \left(\arctan^{-1} \left(\frac{u_y - v_y}{u_x - v_x} \right) \right).$$

위 과정에서의 미지수 i 와 \mathbf{x}_t 는 필획 궤적에 관한

정보를 담고있는 HMM을 참조하여 결정한다. 일반적으로 HMM의 각 상태는 두 가지 이상의 심볼을 출력할 수 있으며 일반적으로 모우드가 하나이상 (multimodal)이다. 즉 각 상태는 심볼 출력 확률 분포에 따라서 필기 진행 방향에 관한 몇 가지의 가설을 제공한다. 본 절의 알고리즘은 이와 같은 몇 개의 가설을 필기 영상에서 $g()$ 함수로 확인하고 확률적으로 가장 나은 방향(코드)를 선택한다.

가설 검증의 결과 선택된 가설의 방향에 따라 다음 순간의 필촉의 위치 \mathbf{x}_t 를 결정하고 이를 그 다음 반복을 위해 $\zeta_t(\cdot)$ 에 기록한다.

$$\zeta_t(j, \mathbf{x}_t) = \arg \max_{\mathbf{x}_t} \max_i \delta_{t-1}(i, \mathbf{x}_{t-1}) \times a_{ij} f_j(\mathbf{x}_{t-1}, \mathbf{x}_t)$$

3.2 필기 궤적 샘플

필기 궤적 추적의 결과를 32x32 크기의 숫자 '4' 이진 영상에 적용한 예는 그림 2와 같다.

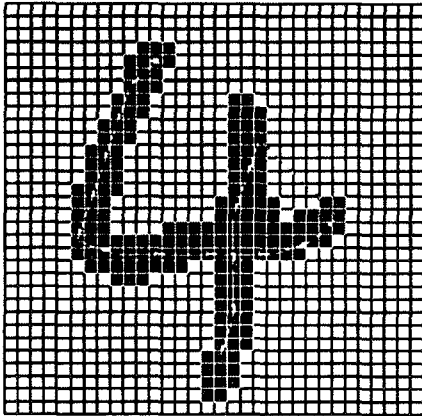


그림 2. 동적 정보 복원 결과

실험에 사용한 HMM은 7개의 상태를 가졌으며 펜-다운 및 펜-업 방향 각각 16 방향, 총 32가지의 코드를 사용하였다. 본 실험에서는 필기 시작 위치가 주어졌다고 가정하였다. 그림4에서는 좌상단의 조그만 흰 점으로 표시되어 있다. 사실 필기 시작 위치는 필기 영상의 분할(segmentation) 방법 및 결과와 밀접한 관련이 있기 때문에 일반론을 논하기는 어렵다. 하지만 Doerman과 Rosenfeld의 연구에 따르면 몇가지 휴리스틱을 이용하여 필기 시작 위치를 충분히 찾을 수 있다는 것이다 [3].

4. 2D HMM

본 절에 기술할 2D HMM은 1D 시계열 신호를 모델링하는 HMM을 2차원으로 확장한 모형이다. 일차원 시계열의 경우 선행 관계가 명확히 존재하여 단 하나의 직전 선행 상태가 존재한다. 본 절의 2차원 HMM은 1D HMM을 2D로 확장하고 기존의 MRF 모형을 일반화한 구조로써 MRF에서 대상으로 하는 단순한 잡음 과정이 아닌 동적 공간 변형 (Dynamic Space Warping) 과정을 모델링 한다. 이를 위해서 2-D에서 causality는

$$a_{hj}^{\downarrow} = p(y_u = \lambda | y_{u-v} = h), \quad h \in \eta_j^U$$

$$a_{ij}^{\rightarrow} = p(y_u = \lambda | y_{u-1} = i), \quad i \in \eta_j^L$$

의 두 가지로 나누어 표현한다. 여기서 a_{hj}^{\downarrow} 는 수직 방향 causality를, a_{ij}^{\rightarrow} 는 수평방향 causality를 나타낸다. 노드 j 의 위 행에서 노드 j 로 전이 가능한 입력 노드의 집합을 η_j^U , 왼쪽에서 전이해 들어올 수 있는 노드의 집합을 η_j^L , 그리고 노드 j 에서 전이해 갈 수 있는 오른쪽 및 아래쪽 노드의 집합을 각각 η_j^R , η_j^D 라고 하자. 그러면 다음과 같은 확률 조건을 만족한다.

$$\sum_{k \in \eta_j^R} a_{jk}^{\rightarrow} = 1, \quad a_{ij}^{\rightarrow} \geq 0,$$

$$\sum_{k \in \eta_j^D} a_{jk}^{\downarrow} = 1, \quad a_{hj}^{\downarrow} \geq 0.$$

η_j^R 와 η_j^D 는 노드 j 에서 본 이웃, 즉 전이가 가능한 이웃이다. 앞으로는 완전 메쉬를 구성하는데 가장 간단하면서도 충분히 일반적인 다음과 같은 이웃 체계를 가정한다.

$$\eta_j^U = \{j-N-1, j-N, j-N+1, j\}$$

$$\eta_j^L = \{j-N-1, j-1, j+N-1, j\}$$

$$\eta_j^R = \{j-N+1, j+1, j+N+1, j\}$$

$$\eta_j^D = \{j+N-1, j+N, j+N+1, j\}$$

그러면 이제 본 논문에서 제안하는 2-D HMM를 정의할 수 있게 되었다.

[정의 2] 마르코프 메쉬 체인 $\mathbf{Y}_{1,t}$ 의 확률 함수 Λ 는 다음과 같은 조건을 만족할 때 2-D 은닉 마르코프 메쉬 체인 또는 2D HMM이다: $\mathbf{X}_{1,t} =$

$x_1 x_2 \dots x_t$ 이고 $\mathbf{Y}_{1,t} = y_1 y_2 \dots y_t$ 일 때:

- (a) $P(\mathbf{X}_{1,u}, \mathbf{Y}_{1,u} | \Lambda) = P(\mathbf{X}_{1,u} | \mathbf{Y}_{1,u}, \Lambda)$
 $\times P(\mathbf{Y}_{1,u} | \Lambda)$ (중확률 과정)
- (b) $P(y_u = j | y_{t-1} = i) \geq 0, \quad i \in \eta_j^L$ (왼쪽 이웃)
- (c) $P(y_u = j | y_{t-v} = h) \geq 0, \quad h \in \eta_j^U$ (오른쪽 이웃)
- (d) $\sum_{k \in \eta_j^R} P(h_j) = 1$ (수평 방향 전이),
- (e) $\sum_{i \in \eta_j^D} P(h_j) = 1$ (수직 방향 전이)

위 정의에 따르면 (b)와 (c)는 각각 a_{ij}^{\rightarrow} 와 a_{hj}^{\downarrow} 를 나타낸다.

2-D 격자에서 위에서 아래로 그리고 좌에서 우로의 진행 관계에 따라 마르코프 성질과 동적 프로그래밍 원리를 적용하면 다음과 같은 전진확률(forward probability)과 후진확률(backward probability)을 얻을 수 있다.

$$\alpha_u(j) = \max_{h \in \eta_j^U, i \in \eta_j^D} a_{ij}^{\rightarrow} a_{hj}^{\downarrow} b_j(x_u) \alpha_{u-1}(i),$$

$$\beta_u(j) = \max_{k \in \eta_j^R, l \in \eta_j^D} s_{jk}^{\rightarrow} r_{jl}^{\downarrow} b_k(x_{u+1}) \beta_{u+1}(k).$$

(이하 β 는 생략) 전진 확률 $\alpha_u(i)$ 는 부분 메쉬 체인 $\mathbf{Y}_{1,j} = y_1 y_2 \dots y_j$ 에서 부분 영상 $\mathbf{X}_{1,u} = x_1 x_2 \dots x_u$ 를 관찰할 확률을 말한다. 후진 확률 $\beta_u(j)$ 는 현재 노드 j , 현재 관찰 데이터 x_u 이후 앞으로의 상태 전개를 계산하는 확률로써 부분 메쉬 체인 $\mathbf{Y}_{j+1,L} = y_{j+1} y_{j+2} \dots y_L$ 에서 나머지 관찰 데이터 $\mathbf{X}_{u+1,W} = x_{u+1} x_{u+2} \dots x_W$ 를 출력할 확률을 말한다. 한편 부분 메쉬 정보, 즉 노드간 체인 구성 정보는

$$(\hat{h}, \hat{i})_u(j) = \arg \max_{h \in \eta_j^U, i \in \eta_j^D} a_{ij}^{\rightarrow} a_{hj}^{\downarrow} b_j(x_u) \times \alpha_{u-1}(i)$$

로 정의된다. 여기서 $Left(j) = \hat{i}$, $Up(j) = \hat{h}$ 이라고 하자. 그러면 위 식 (2), (3)의 계산결과가 완전한 격자가 되기 위해서는 $Left(\hat{h}) = Up(\hat{i})$ 그리고

$$\alpha_u(j) = \max_{h \in \eta_j^U} a_{hj}^{\downarrow} b_j(x_u) \alpha_{u-1}(R_h),$$

$$j = 1, N+1, 2N+1, \dots, (M-1)N+1,$$

$$u = V+1, 2V+1, \dots, (U-1)V+1$$

($\beta_u(j)$ 에 대해서도 유사하게 정의한다) 조건을 만족

해야 한다. 여기서 $R_h = Right^{N-1}(h)$ 로써 격자 상에서 각각 노드 h 의 최우단 노드를 나타낸다. 위 두 조건은 완전한 격자를 구성하기 위해 만족해야 하는 격자 형성 조건(lattice constraint)이다.

이제 전진 확률과 후진 확률을 이용하면

$$P(\mathbf{X} | \Lambda) = \max_j \alpha_u(j) \left[\prod_{k=1}^{V-1} a_{h_k, j_k}^{\downarrow} \right] \beta_u(j)$$

여기서 메쉬 구조의 격자를 만들기 위해서는 $h_k = y_{u-v+k} = Right(h_{k-1})$ 그리고 $h_0 = h$, $j_k = y_{u+k} = Right(j_{k-1})$ 그리고 $j_0 = j$ 를 만족해야 한다.

5. 토의 및 결론

HMM의 뛰어난 모델링 능력을 오프라인 필기 인식에 적용하는 두 가지 아이디어를 짧은 지면에 간략히 제시하였다. 첫 번째 펜 궤적 정보 추출의 경우 일부 패턴에 대해서는 놀라운 결과를 보였으나 많은 경우 엉뚱한 결과를 낳기도 하였다. 그 이유로 주로 HMM의 근본적 문제점으로 지적되는 지수함수형 상태지속 분포에 기인한 경우가 많았다. 그에 따라 적절한 길이의 궤적을 제대로 찾지 못하고 빨리 또는 늦게 방향을 트는 경우가 많았다. 따라서 상태지속 모델을 사용하면 성능을 상당히 개선할 수 있을 것으로 보인다. 1D HMM의 DTW 능력을 2D DSW로 확장한 2D HMM의 경우 수학적으로 간략히 아이디어만 제시하였다. 2D HMM의 훈련 알고리즘은 L. E. Baum 식의 보조적 Q-함수를 최대화하면 구할 수 있다.

참고문헌

- [1] B.-K. Sin and J. H. Kim, "Ligature modeling for online cursive script recognition," *IEEE Trans PAMI*, vol. 19, no. 6, June 1997.
- [2] H.-S. Park and S.-W. Lee, "A truly 2-D hidden Markov model for off-line handwritten character recognition," *Pattern Recognition*, vol. 31, no. 12, pp. 1849-1864, 1998.
- [3] Doerman, *Document Image Understanding*, PhD Thesis, Univ. Maryland, 1995.