

은닉 마르코프 모델의 확률적 최적화를 통한 자동 독순의 성능 향상

이 종 석[†] · 박 철 훈^{††}

요 약

본 논문에서는 자동 독순(automatic lipreading)의 인식기로 쓰이는 은닉 마르코프 모델(HMM: hidden Markov model)의 새로운 확률적 최적화 기법을 제안한다. 제안하는 기법은 전역 최적화가 가능한 확률적 기법인 모의 담금질과 지역 최적화 기법을 결합하는 것으로서, 알고리즘의 빠른 수렴과 좋은 해로의 수렴을 가능하게 한다. 제안하는 알고리즘이 전역 최적해로 수렴함을 수학적으로 보인다. 제안하는 기법을 통해 HMM을 학습함으로써 기존의 알고리즘이 지역해만을 찾는 단점을 개선함으로써 향상된 독순 성능을 나타냄을 실험으로 보인다.

키워드 : 자동 독순, 은닉 마르코프 모델, 확률적 최적화, 전역 최적화, 모의 담금질

Improved Automatic Lipreading by Stochastic Optimization of Hidden Markov Models

Jong-Seok Lee[†] · Cheol Hoon Park^{††}

ABSTRACT

This paper proposes a new stochastic optimization algorithm for hidden Markov models (HMMs) used as a recognizer of automatic lipreading. The proposed method combines a global stochastic optimization method, the simulated annealing technique, and the local optimization method, which produces fast convergence and good solution quality. We mathematically show that the proposed algorithm converges to the global optimum. Experimental results show that training HMMs by the method yields better lipreading performance compared to the conventional training methods based on local optimization.

Key Words : Automatic Lipreading, Hidden Markov Model, Stochastic Optimization, Global Optimization, Simulated Annealing

1. 서 론

자동 독순(automatic lipreading)은 기존의 말소리를 이용한 음성인식 기술을 보완하기 위한 기법으로써 화자의 입술 움직임을 카메라로 기록하고 이를 통해 음성을 인식하고자 하는 것이다. 음성인식기술이 적용되는 실세계에서는 말소리에 잡음이 포함되기 쉬우며 이 경우 잡음없는 환경에 비해 인식 성능이 저하된다. 반면 영상은 소리잡음과 무관하기 때문에 이러한 성능 저하를 보완할 수 있는 중요한 정보가 된다.

독순에서 가장 많이 쓰이는 인식기는 은닉 마르코프 모델(HMM: hidden Markov model)이다[1]. HMM은 이중의 확

률 구조를 가진 모델로써, 음성의 특징을 모델링하는데 적합한 형태를 가지고 있다. HMM을 인식에 사용하기 전에 파라미터들은 학습 데이터를 이용하여 미리 학습되어야 한다. 학습에는 최대 우도(ML: maximum likelihood)를 목표로 하는 기대-최대(EM: expectation-maximization) 알고리즘이 많이 사용된다[2]. EM 알고리즘은 재추정 공식으로 파라미터를 반복적으로 갱신하는 과정을 통해 학습데이터에 대한 우도를 단조증가시킬 수 있다. 하지만 이 기법은 파라미터의 지역 최적해만을 찾기 때문에 초기값에 따라서 전역 최적해를 찾지 못할 수도 있는 단점이 있다.

전역 최적화를 위해 많이 쓰이는 방법 중 하나는 확률적 최적화 기법인 모의 담금질(SA: simulated annealing)이다. 이것은 금속이 녹은 상태에서 담금질 과정을 통해 점차 냉각되면서 최소 에너지를 갖는 결정체가 되는 과정을 응용한 것이다[3]. 해의 생성, 평가 및 선택의 과정을 '담금질'에 의해 반복적으로 수행한다. '온도' 파라미터는 해의 생성과 선택 과정에서 중요한 역할을 하는데, 온도가 높을 때에는 현

* 본 연구는 2007년 한국과학기술원 BK21 정보기술사업단에 의하여 지원되었음.

† 정 회 원 : 한국과학기술원 전자전산학부 연수연구원

†† 정 회 원 : 한국과학기술원 전자전산학부 교수

논문접수 : 2007년 7월 19일, 심사완료 : 2007년 9월 17일

제로부터 멀리 떨어진 해를 생성할 확률과 현재보다 더 나쁜 목적함수를 가지는 해를 선택할 확률이 높아진다. SA의 중요한 특징은 현재보다 나쁜 목적함수의 해를 받아들일 확률이 항상 0보다 크기 때문에 지역 최적해를 벗어날 수 있다는 것이다. 또 다른 SA의 장점은 그 수렴성이 수학적으로 증명된다는 것이다[4,5].

SA가 EM 알고리즘의 한계인 지역최적화를 극복할 수 있는 가능성이 있지만, HMM의 파라미터를 SA로 학습하는 것은 쉬운 일이 아니다. 왜냐하면 HMM의 파라미터 수가 매우 많고 파라미터간의 상호연관성이 상당히 높기 때문이다. 따라서 HMM의 학습에서 좋은 해를 얻고 수렴 속도를 향상시키기 위해 본 논문에서는 하이브리드 모의 담금질(HSA: hybrid SA) 기법을 제안한다. 이 기법은 SA를 지역 최적화 기법인 EM과 결합한 것으로서, SA의 전영역을 탐색하는 능력과 EM에 의한 지역 최적화 능력을 동시에 이용할 수 있다. 또한 본 논문에서는 이 기법에 대한 수렴성을 수학적으로 증명한다. 새로운 해를 생성하는 생성함수와 담금질 계획을 적절히 선택함으로써 목적함수값 및 목적함수의 최소값이 모두 전역 최적해로 수렴함을 보인다. 제안하는 HSA를 HMM의 학습에 적용함으로써 EM 알고리즘의 한계를 극복하여 자동 독순에서 좋은 인식 성능을 보이는 HMM 파라미터 값을 얻음을 보인다.

이하 논문의 구성은 다음과 같다. 2장에서는 제안하는 알고리즘의 과정을 설명하고 3장에서 수학적 수렴 증명을 보인다. 4장에서는 독순 시스템에 대해 설명한다. 5장에서는 실험으로 알고리즘의 성능을 평가하고 6장에서 결론으로 논문을 맺는다.

2. 하이브리드 모의 담금질 알고리즘

본 장에서는 일반적인 최소화 문제에 대한 HSA 알고리즘을 기술한다. 다음과 같은 최소화 문제를 푸는 것을 생각하자.

$$\text{Minimize } C(x), \quad x \in \Psi \quad (1)$$

여기서 Ψ 는 n 차원 실수공간에서 정의된 전체 해공간이고 C 는 전역 최소값이 C^* 인 목적함수이다. 이 문제를 풀기 위한 HSA 알고리즘의 과정은 다음과 같다.

- 1) 초기화: 초기해를 생성하고 초기 온도 T_0 를 설정한다.
- 2) 해의 생성: 현재의 해 x_t 로부터 새로운 해 y_t 를 생성한다.

$$y_t = x_t + \Delta x_t \quad (2)$$

여기서 t 는 시간을 의미한다. Δx_t 는 해의 변화량으로써

다음과 같은 코시(Cauchy) 생성함수[6]로 결정된다.

$$g(\Delta x_t, T_t) = \frac{a_n T_t}{(\|\Delta x_t\|^2 + T_t^2)^{(n+1)/2}} \quad (3)$$

여기서 T_t 는 시간 t 에서의 온도, a_n 은 정규화 상수이다. Δx_t 의 생성은 모든 성분이 σ_t 보다 크거나 같을 때까지 반복된다. σ_t 는 해의 변화량의 하한이며, 시간에 따라 단조 감소하는 충분히 작은 값이다.

- 3) 지역 최적화: y_t 를 지역 최적화 연산에 의해 z_t 로 변화시킨다.

$$z_t = \phi(y_t) \quad (4)$$

여기서 지역 최적화 연산 ϕ 은 항상 해를 향상시킨다. 즉,

$$C(y_t) \geq C(z_t) \quad (5)$$

- 4) 해의 평가: z_t 의 목적함수값을 계산한다.
- 5) 해의 선택: 메트로폴리스(Metropolis) 규칙[7]에 의해 x_t 와 z_t 중에서 하나를 선택한다. z_t 를 선택할 확률은

$$p_n = \min[1, (C(x_t) - C(z_t))/T_t] \quad (6)$$

로 주어진다.

- 6) 담금질: 다음의 식에 의해 온도를 낮춘다

$$T_t = T_0 / t \quad (7)$$

- 7) 종료: 종료조건을 만족하면 알고리즘을 끝낸다. 그렇지 않으면 2)로 돌아간다.

해의 생성에서 사용하는 코시 생성함수와 시간의 역수에 비례하는 담금질의 조합은 빠른 모의 담금질 기법(FSA: fast SA)[8]에서 제안된 방식이다. 다음 장에서는 알고리즘의 수렴을 수학적으로 증명한다.

3. 수렴 증명

HSA 알고리즘의 수렴을 증명하기 위해 두 개의 정리와 두 개의 따름정리를 보인다. 첫번째 정리는 목적함수값의 수열이 전역 최적해로 수렴하기 위해 담금질 계획과 생성함수가 만족해야 하는 조건을 보이며, 그에 대한 따름정리에서 HSA가 그러한 조건을 만족함을 보인다. 두번째 정리와 따름정리는 목적함수의 최소값의 수열 역시 전역 최적해로 수렴함을 보인다.

정리 1. $\varepsilon > 0$ 과 $\xi > 0$ 에 대해 다음을 정의한다.

$$\begin{aligned} \Psi_\varepsilon &= \{x \in \Psi \mid C(x) < C^* + \varepsilon\} \\ \Omega_{\varepsilon, \xi, t} &= \{x \in \Psi \mid C^* + \varepsilon \leq C(x) < C^* + \varepsilon + 1/t^\xi\} \\ \Omega'_{\varepsilon, \xi, t} &= \{x \in \Psi \mid C^* + \varepsilon \leq C(\phi(x)) < C^* + \varepsilon + 1/t^\xi \text{ and } C(x) \geq C^* + \varepsilon + 1/t^\xi\} \end{aligned}$$

모든 $\varepsilon > 0$ 에 대해 n 차원 실수공간에서 정의된 르베그 측도(Lebesgue measure) ζ 가 $\zeta(\Psi_\varepsilon) > 0$ 을 만족한다고 하자. 또한 모든 $t \geq t_0$ 에 대해 $\zeta(\Omega_{\varepsilon, \xi, t}) + \zeta(\Omega'_{\varepsilon, \xi, t}) \leq R/t^\xi$ 를 만족하는 상수 $\xi > 0$ 와 $R > 0$ 이 존재한다고 가정하자. 다음 세 조건이 만족할 때 HSA에 의한 목적함수값의 수열 $\{C(x_t), t \geq 0\}$ 은 임의의 초기해 $x_0 \in \Psi$ 에 대해 확률로써 전역 최적해 C^* 로 수렴한다.

조건 1) 온도함수의 조건: $t \rightarrow \infty$ 일 때 $T_t \rightarrow 0$

조건 2) 담금질 계획의 조건:

$$\sum_{t=1}^{\infty} \exp(-1/t^\xi T_t) < \infty \quad (8)$$

조건 3) 생성함수의 조건: 정수 $N_1 > 0$, 상수 $M_1 > 0$ 과 $M_2 > 0$, 및 $\beta + \xi > 1$ 을 만족하는 $0 < \beta < 1$ 에 대해

$$\max_{\substack{x, y \in \Psi \\ |y^i - x^i| \geq \sigma_t}} g(y - x, T_t) \leq M_1 / t^\beta, \forall t \geq N_1 > 0 \quad (9)$$

$$\min_{\substack{x, y \in \Psi \\ |y^i - x^i| \geq \sigma_t}} g(y - x, T_t) \leq M_2 / t, \forall t > 0 \quad (10)$$

여기서 x^i 와 y^i 는 각각 x 와 y 의 i 번째 성분이다.

증명: 부록 1 참조.

따름정리 1. 생성함수와 온도함수가 식 (3)과 (7)에 의해 주어질 때 목적함수값의 수열 $\{C(x_t), t \geq 0\}$ 은 임의의 초기해 $x_0 \in \Psi$ 에 대해 확률로써 전역 최적해 C^* 로 수렴한다.

증명: $v_i = \max_{x, y \in \Psi} |x^i - y^i|$ ($1 \leq i \leq n$) 라 하자. $\beta = 1 - \xi/2$ 라 하고, $0 < \xi < 1$ 에 대해

$$\sigma_t = \sigma_0 t^{-\xi/2(n+1)} \quad (11)$$

라 하자. 식 (7)에 의한 온도함수는 t 가 무한대로 갈 때 0 으로 수렴하며

$$\sum_{t=1}^{\infty} \exp(-1/t^\xi T_t) = \sum_{t=1}^{\infty} \exp(-1/t^{\xi-1} T_0) < \infty$$

를 만족한다. 생성함수는

$$\begin{aligned} \max_{\substack{x, y \in \Psi \\ |y^i - x^i| \geq \sigma_t}} g(y - x, T_t) &\leq \max_{\substack{x, y \in \Psi \\ |y^i - x^i| \geq \sigma_t}} \frac{a_n T_t}{(\|y - x\|^2)^{(n+1)/2}} \\ &\leq \frac{a_n T_t}{(n\sigma_t^2)^{(n+1)/2}} \\ &= \frac{a_n T_0}{(n\sigma_0^2)^{(n+1)/2}} \frac{1}{t^{1-\xi/2}} \\ &= M_1 / t^\beta \end{aligned}$$

와

$$\begin{aligned} \min_{\substack{x, y \in \Psi \\ |y^i - x^i| \geq \sigma_t}} g(y - x, T) &\geq \frac{a_n T_t}{(\sum_{i=1}^n v_i^2 + T_t^2)^{(n+1)/2}} \\ &\geq \frac{a_n T_0 / t}{(\sum_{i=1}^n v_i^2 + T_0^2)^{(n+1)/2}} \\ &= M_2 / t \end{aligned}$$

에 의해 조건 (9)와 (10)을 각각 만족한다.

정리 2. $\varepsilon > 0$ 에 대해 다음을 정의한다.

$$\begin{aligned} \Psi_\varepsilon &= \{x \in \Psi \mid C(x) < C^* + \varepsilon\} \\ \Psi'_\varepsilon &= \{x \in \Psi \mid C(\phi(x)) < C^* + \varepsilon \text{ and } C(x) \geq C^* + \varepsilon\} \end{aligned}$$

모든 $\varepsilon > 0$ 에 대해 n 차원 실수공간에서 정의된 르베그 측도(Lebesgue measure) ζ 가 $\zeta(\Psi_\varepsilon) > 0$ 을 만족한다고 하자. 상수 $M_3 > 0$ 에 대해

$$\min_{x, y \in \Psi} g(y - x, T_t) \geq M_3 / t, \forall t \geq t_0 \quad (12)$$

만족할 때 HSA에 의한 목적함수의 최소값의 수열 $\{\min_{0 \leq j \leq t} C(x_j), t \geq 0\}$ 은 임의의 초기해 $x_0 \in \Psi$ 에 대해 확률로써 전역 최적해 C^* 로 수렴한다.

증명: 부록 2 참조.

따름정리 2. 생성함수와 온도함수가 식 (3)과 (7)에 의해 주어질 때 목적함수의 최소값의 수열 $\{\min_{0 \leq j \leq t} C(x_j), t \geq 0\}$ 은 모든 초기해 $x_0 \in \Psi$ 에 대해 확률로써 전역 최적해 C^* 로 수렴한다.

증명: $v_i = \max_{x, y \in \Psi} |x^i - y^i|$ ($1 \leq i \leq n$) 이라 하자.

식 (3), (7) 및 (12)에 의해

$$\begin{aligned} \min_{x,y \in \Psi} g(y-x, T) &\geq \frac{a_n T_0 / t}{(\sum_{i=1}^n v_i^2 + T_0^2)^{(n+1)/2}} \\ &\geq M_3 / t \end{aligned}$$

를 얻는다.

4. 자동 독순 시스템

자동 독순은 카메라로 기록한 입술의 움직임에서 적절한 특징을 추출하고 인식기를 통해 인식하는 과정으로 이루어진다. 이하에서는 본 논문에서 사용한 데이터베이스, 특징 추출 기법 및 인식기에 대해 설명한다.

4.1. 데이터베이스

우리말 '일'부터 '구'까지, 그리고 '영'과 '공'을 포함하는 DIGIT 데이터베이스와 16개의 우리나라 도시 이름으로 구성된 CITY 데이터베이스를 사용한다[9]. 56명의 화자가 각 단어를 세 번씩 발음하는 동안 카메라를 이용하여 화자의 입술과 그 주변 얼굴 부분을 30Hz의 프레임 비율로 기록한 것이다.

인식 실험은 화자 독립 방식으로 하였으며 신뢰도를 높이기 위해 4-cut 교차타당성 확인(cross-validation) 방식으로 실험하였다. 56명의 화자를 네 모둠으로 나누고 세 모둠의 데이터를 학습에, 나머지 한 모둠을 인식에 사용하는 과정을 돌아가며 네 번 반복한다.

4.2. 특징 추출

독순을 위해 영상에서 특징을 추출하는 방법은 크게 입술 윤곽선 모델의 파라미터를 특징으로 사용하는 방식과 입술 영역 영상의 변환으로 특징을 얻는 방식으로 분류할 수 있다[1]. 후자의 경우가 더 많은 정보를 포함할 수 있어 더 좋은 인식 성능을 보이는 것으로 알려져 있다[10].

본 논문에서 사용하는 영상 특징은 입술 영역에 주성분 분석(PCA: principal component analysis)을 적용하는 것이다. 입술 영역의 추출은 얼굴의 좌우 밝기 변화를 보정하고 영상간 픽셀값을 정규화하여 입술 양 끝점을 검출하여 얻는다[9]. 입술 영역 영상에서 각 픽셀에 대해 발음 전체에 대한 평균을 제거하여 화자간 차이를 없애고 PCA를 적용하여 12차원의 정적(static) 특징벡터를 얻는다. 그리고 동적 특성을 얻기 위해 정적 특징의 시간 미분으로 정의되는 델타항을 함께 사용한다[2]. 따라서 각 프레임마다 24차원의 특징 벡터를 얻는다.

4.3. 인식기

인식기로는 음성인식이나 독순에서 가장 많이 사용되는 연속 HMM을 사용한다. 연속 HMM는 초기확률분포, 전이 확률분포 그리고 관측확률분포를 파라미터로 가지며 관측확률분포는 가우시안 혼합 모델을 사용한다.

HMM의 학습 과정에서는 각 HMM에 해당하는 클래스의 발음 데이터에 대해 로그우도의 합이 최대가 되도록 한다.

본 논문에서는 이 과정에 대해 기존의 EM 알고리즘과 제안하는 HSA 알고리즘을 사용한다. 2장에서 설명한 HSA 알고리즘은 최소화문제에 대한 것인데, HMM의 학습은 최대화 문제이다. 따라서 HSA를 이용한 HMM의 학습에서는 로그우도의 음수값을 목적함수로 정의한다. 해의 생성에서 필요한 하한값인 σ_t 는 식 (11)로 정의되며 σ_0 를 10^{-10} 과 같이 매우 작은 값으로 함으로써 해의 생성 과정에서 불필요한 반복이 없도록 한다.

인식 과정에서는 클래스를 알 수 없는 발음의 특징벡터열을 모든 HMM에 입력하고, 가장 높은 확률값을 보이는 HMM을 선택하여 인식한다.

5. 실험 및 결과

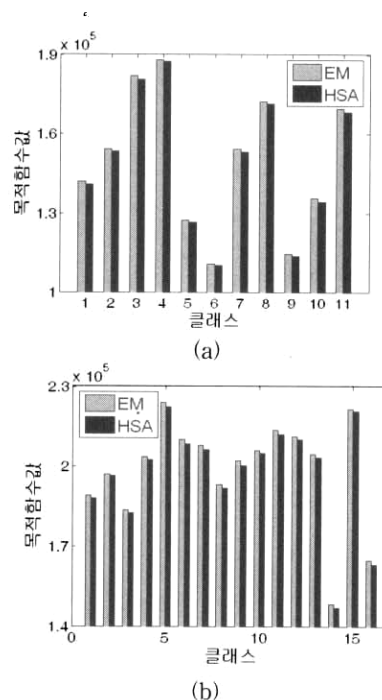
5.1. 실험조건

인식기로써 전단어 HMM 모델을 사용하였으며 상태의 수는 각 단어의 음소의 수에 비례하게 하였고 가우시안 혼합모델의 가우시안 함수의 수는 실험을 통해 3으로 정하였다.

HSA에서 초기 온도는 10으로 하였고 종료조건은 10000번의 반복횟수로 하였다. 지역 최적화 연산으로는 5번의 EM 알고리즘을 적용하였다.

5.2. 성능 비교

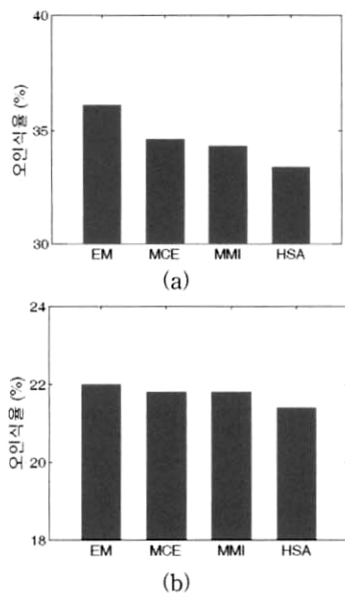
(그림 1)은 최종 목적함수값, 즉 로그우도의 음수를 각 데이터베이스의 각 발음 클래스에 대해 비교한 것이다. 두 데이터베이스의 모든 클래스에 대해서 HSA의 경우가 EM에 비해 더 작은 목적함수값, 즉 더 큰 로그우도값을 나타내는 것을 볼 수 있다. 이는 HSA의 전역 최적화를 통해 얻는 결과이다.



(그림 1) EM과 HSA에 의한 최종 목적함수값의 비교 (a) DIGIT 데이터베이스 (b) CITY 데이터베이스

다음으로 HSA의 인식 성능을 보인다. EM 외에 기존의 분별 학습 기법인 최소분류오류(MCE: minimum classification error)[11]와 최대상호정보(MMI: maximum mutual information)[12]에 의한 결과를 포함한다. 이 두 방법은 로그우도를 최대화하는 대신 인식율과 관련된 목적함수를 정의하고 최적화하는 알고리즘이며 EM보다 좋은 성능을 내는 것으로 알려져 있다. 이들의 학습 목표가 EM이나 HSA와는 다르지만 이들과의 비교를 통해 제안하는 알고리즘의 성능 향상이 어느 정도인지 알아볼 수 있다. 이 두 분별학습기법은 알고리즘 파라미터를 정해야 하는데 여러번의 실험을 통해 가장 좋은 값을 보이는 것으로 정하고 그 결과를 보였다.

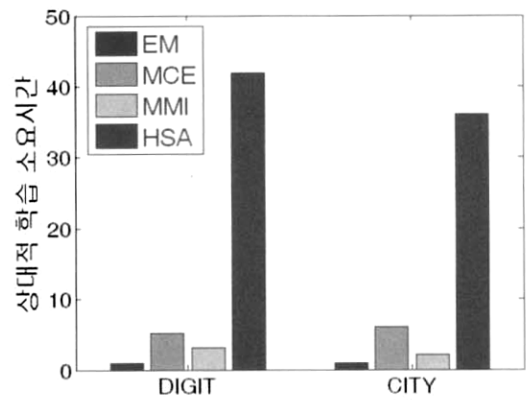
(그림 2)는 각 데이터베이스에 대해 평균적인 오인식율을 비교한다. 네 가지 학습 기법 중 제안하는 HSA가 가장 작은 오인식율을 보이는 것을 알 수 있다. EM에 대한 HSA의 상대적 오인식율 감소는 각 데이터베이스에 대해 7.5%와 2.7%이며, 이러한 성능 향상은 (그림 1)의 결과에서 보듯이 HSA에 의해 더 좋은 목적함수값을 찾았기 때문이다. HSA에 의한 성능향상의 통계적 유의성을 판단하기 위해 교차타당성 확인에 대한 t-검정을 수행하였다. EM의 평균 오인식율에서 HSA의 평균 오인식율을 뺀 것을 μ_D 라 할 때 귀무가설을 $H_0: \mu_D \leq 0$ 로 한다. 각 데이터베이스에 대해 $t=4.24$ 와 $t=2.37$ 을 얻었으며 이는 모두 자유도 3일 때 유의수준 5%에서의 t값인 $t_{0.05}=2.35$ 보다 크므로 가설 H_0 을 기각하고 대립가설인 $H_a: \mu_D > 0$ 을 지지할 수 있다. 따라서 HSA가 EM에 비해 더 낮은 오인식율을 낸다는 것이 두 데이터베이스에 대한 실험결과 통계적으로 유의하다고 결론지을 수 있다. 또한, HSA는 EM보다 좋은 성능을 보이는 MCE와 MMI보다도 더 좋은 것을 관찰할 수 있다. 이는 MCE와 MMI 역시 지역 최적해만을 찾는 기법이기에 때문에 그에 따른 한계가 존재하는 것을 의미한다.



(그림 2) 학습 기법에 따른 오인식율 비교 (a) DIGIT 데이터베이스 (b) CITY 데이터베이스

5.3. 계산량 비교

HSA는 앞의 결과와 같은 우수한 성능을 내는 대신 더 많은 계산량을 필요로 한다. (그림 3)은 EM의 학습시간을 1로 하였을 때 각 기법에 의한 상대적인 소요 시간을 비교한다. 이 결과들은 리눅스를 운영체제로 하는 2.4 GHz CPU의 컴퓨터에서 얻었다. HSA의 경우 각 단어별 HMM을 여러 컴퓨터에서 동시에 병렬로 학습하였다. 그림에서 HSA의 학습은 EM, MCE, MMI에 비해 상당히 많은 시간을 소요하는 것을 알 수 있다. 하지만 학습은 인식기를 사용하기 이전에 미리 해둘 수 있으며 이러한 학습 시간의 증가는 실제 인식기를 사용하는 것과는 무관하다.



(그림 3) 학습 기법에 따른 상대적 학습 소요 시간

6. 결론

본 논문에서는 자동 독순을 위한 HMM의 새로운 학습 기법을 제안하였다. SA와 EM을 결합한 확률적 최적화 기법인 HSA 알고리즘을 제안하였으며 이를 통해 전영역에서의 파라미터 최적화를 가능하게 하였다. 목적함수값의 수열과 목적함수의 최소값의 수열이 모두 확률로써 전역 최적해로 수렴함을 수학적으로 증명하였다. 실험을 통해 HSA가 기존의 알고리즘에 비해 학습에는 더 많은 시간을 요구하지만 목적함수값이나 인식율에 있어서 더 우수함을 보였다. 추후 과제으로써 학습 시간의 단축을 위한 병렬처리 기법과 연속성성에 대한 적용 등을 연구중이다.

참고 문헌

- [1] C. C. Chibelushi, F. Deravi, and J. S. D. Mason, "A review of speech-based bimodal recognition," IEEE Trans. Multimedia, Vol. 4, No. 1, pp. 23-27, Mar. 2002.
- [2] L. Rabiner and B.-H. Juang, 'Fundamentals of Speech Recognition,' Prentice-Hall, Englewood Cliffs, NJ, 1993.
- [3] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," Science, Vol. 220, pp. 671-680, May 1983.

- [4] D. Mitra, F. Romeo, and A. Sangiovanni-Vincentelli, "Convergence and finite-time behavior of simulated annealing," *Advances in Applied Probability*, Vol. 18, pp. 747-771, 1986.
- [5] R. L. Yang, "Convergence of simulated annealing algorithm for continuous global optimization," *J. Optimization Theory and Applications*, Vol. 104, No. 3, pp. 691-716, Mar. 2000.
- [6] D. Nam, J.-S. Lee, and C. H. Park, "n-dimensional Cauchy neighbor generation for the fast simulated annealing," *IEICE Trans. Inf. Syst.*, Vol. E87-D, No. 11, pp. 2499-2502, Nov. 2004.
- [7] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *J. Chem. Phys.*, Vol. 21, No. 6, pp. 1087-1092, 1953.
- [8] H. H. Szu and R. L. Hartley, "Fast simulated annealing," *Phys. Lett. A*, Vol. 122, No. 3-4, pp. 157-162, Jun. 1987.
- [9] J.-S. Lee and C. H. Park, "Training hidden Markov models by hybrid simulated annealing for visual speech recognition," in *Proc. IEEE Int. Conf. Systems, Man and Cybernetics*, Taipei, Taiwan, pp. 198-202, Oct. 2006.
- [10] I. Matthews, G. Potamianos, C. Neti, and J. Luetin, "A comparison of model and transform-based visual features for audio-visual LVCSR," in *Proc. Int. Conf. Multimedia and Expo*, Tokyo, Japan, pp. 22-25, Apr. 2001.
- [11] B.-H. Juang, W. Chou, and C.-H. Lee, "Minimum classification error rate methods for speech recognition," *IEEE Trans. Speech and Audio Processing*, Vol. 5, No. 3, pp. 257-265, May 1997.
- [12] A. Ben-Yishai and D. Burshtein, "A discriminative training algorithm for hidden Markov models," *IEEE Trans. Speech and Audio Processing*, Vol. 12, No. 3, pp. 204-216, May 2004.

[부록 1] 정리 1의 증명

임의의 $\varepsilon > 0$, $\delta > 0$ 및 $x_0 \in \Psi$ 에 대해 다음을 만족하는 자연수 I 가 존재함을 보이려고 한다.

$$P(x_t \in \Psi \setminus \Psi_\varepsilon) < \delta, \forall t \geq I \tag{13}$$

이를 위해 다음의 두 집합을 정의한다.

$$\begin{aligned} \Theta_{\varepsilon, \xi, t} &= \{x \in \Psi \mid C(x) \geq C^* + \varepsilon + 1/t^\xi\} \\ \Psi_\varepsilon' &= \{x \in \Psi \mid C(\phi(x)) < C^* + \varepsilon \text{ and } C(x) \geq C^* + \varepsilon\} \end{aligned}$$

시간 간격 $[t, t + u]$ 에 대해, Ψ_ε 에서 $\Omega_{\varepsilon, \xi, t+j-1}$

($1 \leq j \leq u$) 로의 전이가 최소한 한번 일어나는 사건을 $A(t, u)$, Ψ_ε 에서 $\Theta_{\varepsilon, \xi, t+j-1}$ ($1 \leq j \leq u$) 로의 전이가 최소한 한번 일어나는 사건을 $B(t, u)$, Ψ_ε 에 한번도 가지 않는 사건을 $C(t, u)$, $x_{t+u} \in \Psi \setminus \Psi_\varepsilon$ 인 사건을 D 라 하자. $D \subset A \cup B \cup C$ 이므로,

$$P(x_{t+u} \in \Psi \setminus \Psi_\varepsilon) = P(D) \leq P(A) + P(B) + P(C) \tag{14}$$

가 성립한다.

먼저,

$$P(A(t, u)) < \delta/3, \forall t \geq t_1, \forall u > 0 \tag{15}$$

를 만족하는 자연수 t_1 이 존재함을 보인다. $x_{t+j-1} \in \Psi_\varepsilon$ 이고 $x_{t+j} \in \Omega_{\varepsilon, \xi, t+j-1}$ 인 사건을 $A_j(t, u)$ 라 한다. 그러면

$$A(t, u) \subset \bigcup_{j=1}^u A_j(t, u)$$

이고

$$P(A(t, u)) \leq \sum_{j=1}^u P(A_j(t, u)) \tag{16}$$

가 성립한다. 여기서,

$$\begin{aligned} P(A_j(t, u)) &= \int_{\Omega_{\xi, t+j-1} \setminus \Psi} g(y - x_{t+j-1}, T_{t+j-1}) \exp\left\{\frac{C(x_{t+j-1}) - C(\phi(y))}{T_{t+j-1}}\right\} dy \\ &\quad + \int_{\Omega_{\xi, t+j-1}} g(y - x_{t+j-1}, T_{t+j-1}) \exp\left\{\frac{C(x_{t+j-1}) - C(\phi(y))}{T_{t+j-1}}\right\} dy \\ &\leq \int_{\Omega_{\xi, t+j-1}} g(y - x_{t+j-1}, T_{t+j-1}) dy + \int_{\Omega_{\xi, t+j-1}} g(y - x_{t+j-1}, T_{t+j-1}) dy \\ &\leq \zeta(\Omega_{\varepsilon, \xi, t+j-1}) \max_{x \in \Psi, y \in \Omega_{\xi, t+j-1}} g(y - x, T_{t+j-1}) + \zeta(\Omega_{\varepsilon, \xi, t+j-1}) \max_{x \in \Psi, y \in \Omega_{\xi, t+j-1}} g(y - x, T_{t+j-1}) \\ &\leq \{\zeta(\Omega_{\varepsilon, \xi, t+j-1}) + \zeta(\Omega_{\varepsilon, \xi, t+j-1}')\} \max_{x, y \in \Psi} g(y - x, T_{t+j-1}) \\ &\leq RM_1 / (t + j - 1)^{\beta + \xi} \end{aligned} \tag{17}$$

이다. 마지막 부등식을 얻는 과정에는 정리 1의 가정과 조건이 이용되었다. 식 (16)과 (17)을 통해

$$P(A(t, u)) \leq RM_1 \sum_{j=t}^{t+u-1} 1/j^{\beta + \xi}$$

를 얻는다. $\beta + \xi > 1$ 이므로, 모든 $N_1 > 0$ 에 대해

$\sum_{j=N_1}^{\infty} 1/j^{\beta+\xi} < \infty$ 이다. 따라서 모든 $t \geq t_1$ 과 $u > 0$ 에 대해

$$\sum_{j=t}^{t+u-1} 1/j^{\beta+\xi} \leq \sum_{j=t_1}^{\infty} 1/j^{\beta+\xi} < \delta/3RM_1$$

를 만족하는 양의 정수 $t_1 > N_1$ 이 존재한다. 즉, 식 (15)가 성립한다.

다음으로,

$$P(B(t, u)) < \delta/3, \forall t \geq t_2, \forall u > 0 \quad (18)$$

를 만족하는 자연수 t_2 가 존재함을 보이고자 한다. $x_{t+j-1} \in \Psi_\varepsilon$ 이고 $x_{t+j} \in \Theta_{\varepsilon, \xi, t+j-1}$ 인 사건을 $B_j(t, u)$ 라 하자. 그러면

$$P(B(t, u)) \leq \sum_{j=1}^u B_j(t, u)$$

가 성립한다. 모든 $1 \leq j \leq u$ 와 $x_{t+j-1} \in \Psi_\varepsilon$ 에 대해

$$P(B_j(t, u)) = \int_{\Theta_{\varepsilon, \xi, t+j-1} \setminus \Omega_{\varepsilon, \xi, t+j-1} \setminus \Psi_\varepsilon} g(y - x_{t+j-1}, T_{t+j-1}) \exp\left\{\frac{C(x_{t+j-1}) - C(\phi(y))}{T_{t+j-1}}\right\} dy$$

이다. Ψ_ε 의 최대 목적함수값이 $C^* + \varepsilon$, $\Theta_{\varepsilon, \xi, t+j-1}$ 의 최소 목적함수값이 $C^* + \varepsilon + 1/(t+j-1)^\xi$ 이므로, $C(x_{t+j-1}) - C(\phi(y)) \leq -1/(t+j-1)^\xi$ 이다. 따라서,

$$P(B_j(t, u)) \leq \exp\{-1/(t+j-1)^\xi T_{t+j-1}\} \int_{\Theta_{\varepsilon, \xi, t+j-1} \setminus \Omega_{\varepsilon, \xi, t+j-1} \setminus \Psi_\varepsilon} g(y - x_{t+j-1}, T_{t+j-1}) dy \leq \exp\{-1/(t+j-1)^\xi T_{t+j-1}\}$$

그러므로

$$P(B(t, u)) \leq \sum_{j=1}^u \exp\{-1/(t+j-1)^\xi T_{t+j-1}\} = \sum_{j=t}^{t+u-1} \exp\{-1/j^\xi T_j\}$$

정리에서 $\sum_{j=1}^{\infty} \exp(-1/j^\xi T_j)$ 를 가정하였으므로, 모든

$t \geq t_2$ 와 $u > 0$ 에 대해 다음을 만족하는 자연수 t_2 가 존재한다.

$$\sum_{j=t}^{t+u-1} \exp(-1/j^\xi T_j) \leq \sum_{j=t_2}^{\infty} \exp(-1/j^\xi T_j) < \delta/3$$

따라서 식 (18)이 성립한다. $t_0 = \max\{t_1, t_2\}$ 라 하면, 식 (15)와 (18)로부터

$$P(A(t_0, u)) + P(B(t_0, u)) < 2\delta/3, \forall u > 0 \quad (19)$$

를 얻는다.

마지막으로, 다음을 만족하는 자연수 u_0 가 존재함을 보인다.

$$P(C(t_0, u)) < \delta/3, \forall u > u_0$$

$$\begin{aligned} P(C(t_0, u)) &= P(x_{t_0} \in \Psi \setminus \Psi_\varepsilon \mid x_{t_0-1} \in \Psi) \\ &P(x_{t_0+1} \in \Psi \setminus \Psi_\varepsilon \mid x_{t_0} \in \Psi \setminus \Psi_\varepsilon) \\ &\dots P(x_{t_0+u} \in \Psi \setminus \Psi_\varepsilon \mid x_{t_0+u-1} \in \Psi \setminus \Psi_\varepsilon) \\ &\leq \prod_{j=t_0}^{t_0+u-1} \max_{x_j \in \Psi \setminus \Psi_\varepsilon} P(x_{j+1} \in \Psi \setminus \Psi_\varepsilon \mid x_j) \end{aligned}$$

라 쓸 수 있다. 여기서 $P(x_{t+1} \in \Psi \setminus \Psi_\varepsilon \mid x_t) \leq 1 - P(y_t \in \Psi_\varepsilon \mid x_t)$ 가 되므로 윗 식은

$$\begin{aligned} \prod_{j=t_0}^{t_0+u-1} \max_{x_j \in \Psi \setminus \Psi_\varepsilon} P(x_{j+1} \in \Psi \setminus \Psi_\varepsilon \mid x_j) &\leq \exp\left\{\log \prod_{j=t_0}^{t_0+u-1} \max_{x_j \in \Psi \setminus \Psi_\varepsilon} \{1 - P(y_j \in \Psi_\varepsilon \mid x_j)\}\right\} \\ &= \exp\left\{\sum_{j=t_0}^{t_0+u-1} \log\left[1 - \min_{x_j \in \Psi \setminus \Psi_\varepsilon} P(y_j \in \Psi_\varepsilon \mid x_j)\right]\right\} \\ &\leq \exp\left[-\sum_{j=t_0}^{t_0+u-1} \min_{x_j \in \Psi \setminus \Psi_\varepsilon} P(y_j \in \Psi_\varepsilon \mid x_j)\right] \end{aligned} \quad (20)$$

로 쓸 수 있다. 마지막 부등식의 유도에는 $\log(1-x) \leq -x$ ($0 \leq x < 1$) 이 사용되었다. 여기서 조건 (10)에 의해

$$\begin{aligned} \min_{x_j \in \Psi \setminus \Psi_\varepsilon} P(y_j \in \Psi_\varepsilon \mid x_j) &= \min_{x_j \in \Psi \setminus \Psi_\varepsilon} \int_{\Psi_\varepsilon} g(y-x, T_j) dy \\ &\geq \zeta(\Psi_\varepsilon) \min_{\substack{x \in \Psi \setminus \Psi_\varepsilon, y \in \Psi_\varepsilon \\ |y^j - x^j| \geq \sigma_j}} g(y-x, T_j) \\ &\geq \zeta(\Psi_\varepsilon) \min_{\substack{x, y \in \Psi \\ |y^j - x^j| \geq \sigma_j}} g(y-x, T_j) \\ &\geq \zeta(\Psi_\varepsilon) M_2 / j \end{aligned}$$

따라서

$$\sum_{j=t}^{\infty} \min_{x_j \in \Psi \setminus \Psi_\varepsilon} P(y_j \in \Psi_\varepsilon | x_j) \geq \zeta(\Psi_\varepsilon) M_2 \sum_{j=t}^{\infty} 1/j = \infty$$

따라서

$$\sum_{j=t_0}^{t_0+u-1} \min_{x_j \in \Psi \setminus \Psi_\varepsilon} P(y_j \in \Psi_\varepsilon | x_j) > \log(3/\delta), \forall u > u_0$$

을 만족하는 정수 u_0 가 존재하며,

$$P(C(t_0, u)) < \exp\{-\log(3/\delta)\} = \delta/3 \quad (21)$$

이다. 식 (14), (19) 및 (21)로부터, 모든 $u \geq u_0$ 에 대해

$$P(x_{t+u} \in \Psi \setminus \Psi_\varepsilon) \leq P(A(t_0, u)) + P(B(t_0, u)) + P(C(t_0, u)) < \delta$$

이다. 따라서, 식 (13)은 $I = t_0 + u_0$ 에 대해 성립한다.

[부록 2] 정리 2의 증명

$F_t = \min_{0 \leq j \leq t} C(x_j)$ 라 하자. 정리의 증명을 위해 모든 $\varepsilon > 0$ 와 $x_0 \in \Psi$ 에 대해 $\lim_{t \rightarrow \infty} P(F_t - C^* \geq \varepsilon) = 0$ 임을 보인다.

$$\begin{aligned} &P(F_t - C^* \geq \varepsilon) \\ &= P(x_1 \in \Psi \setminus \Psi_\varepsilon | x_0 \in \Psi \setminus \Psi_\varepsilon) P(x_2 \in \Psi \setminus \Psi_\varepsilon | x_1 \in \Psi \setminus \Psi_\varepsilon) \\ &\dots P(x_t \in \Psi \setminus \Psi_\varepsilon | x_{t-1} \in \Psi \setminus \Psi_\varepsilon) \leq \prod_{j=0}^{t-1} \max_{x_j \in \Psi \setminus \Psi_\varepsilon} P(x_{j+1} \in \Psi \setminus \Psi_\varepsilon | x_j) \end{aligned}$$

이므로, 식 (20)과 유사한 방식으로

$$\begin{aligned} \lim_{t \rightarrow \infty} P(F_t - C^* \geq \varepsilon) &\leq \prod_{j=0}^{\infty} \max_{x_j \in \Psi \setminus \Psi_\varepsilon} P(x_{j+1} \in \Psi \setminus \Psi_\varepsilon | x_j) \\ &\leq \exp \left[- \sum_{j=0}^{\infty} \min_{x_j \in \Psi \setminus \Psi_\varepsilon} P(x_{j+1} \in \Psi_\varepsilon | x_j) \right] \end{aligned}$$

를 얻는다. 여기서,

$$\min_{x_j \in \Psi \setminus \Psi_\varepsilon} P(x_{j+1} \in \Psi_\varepsilon | x_j) = \min \left\{ \min_{x_j \in \Psi \setminus \Psi_\varepsilon} P(y_j \in \Psi_\varepsilon | x_j), \min_{x_j \in \Psi \setminus \Psi_\varepsilon} P(y_j \in \Psi^c | x_j) \right\}$$

이다. 이 식의 각 항은

$$\begin{aligned} \min_{x_j \in \Psi \setminus \Psi_\varepsilon} P(y_j \in \Psi_\varepsilon | x_j) &= \min_{x_j \in \Psi \setminus \Psi_\varepsilon} \int_{\Psi_\varepsilon} g(y - x_j, T_j) dy \\ &\geq \zeta(\Psi_\varepsilon) \min_{x_j \in \Psi \setminus \Psi_\varepsilon} \int_{\Psi_\varepsilon} g(y - x_j, T_j) dy \\ &\geq \zeta(\Psi_\varepsilon) \min_{x_j, y \in \Psi} g(y - x_j, T_j) \\ &\geq \zeta(\Psi_\varepsilon) M_3 / j \end{aligned}$$

와, 유사한 방법으로,

$$\min_{x_j \in \Psi \setminus \Psi_\varepsilon} P(y_j \in \Psi_\varepsilon^c | x_j) \geq \zeta(\Psi_\varepsilon^c) M_3 / j$$

로 쓸 수 있다. 위의 결과들을 결합하면

$$\begin{aligned} \lim_{t \rightarrow \infty} P(F_t - C^* \geq \varepsilon) &\leq \exp \left[- \min \{ \zeta(\Psi_\varepsilon), \zeta(\Psi_\varepsilon^c) \} \sum_{j=0}^{\infty} M_3 / j \right] \\ &= 0 \end{aligned}$$

를 얻을 수 있다.



이 종 석

e-mail : jslee@nnmi.kaist.ac.kr

1999년 한국과학기술원 전기및전자공학과
학사

2001년 한국과학기술원 전자전산학과
공학석사

2006년 한국과학기술원 전자전산학과
공학박사

2006년~현재 한국과학기술원 전자전산학부 연수연구원

관심분야: 시청각 음성인식, 멀티모달 인터페이스



박 철 훈

e-mail : chpark@kaist.ac.kr

1984년 서울대학교 전자공학과 학사

1985년 Caltech 전자공학과 공학석사

1990년 Caltech 전자공학과 공학박사

1991년~현재 한국과학기술원 전자전산학부
교수

관심분야: 지능시스템, 신경회로망, 최적화, 지능제어