은닉 마르코프 모델의 다목적함수 최적화를 통한 자동 독순의 성능 향상

이 종 석[†] · 박 철 훈[††]

요 약

본 논문은 입술의 움직임을 통해 음성을 인식하는 자동 독순의 인식 성능 향상을 위해 인식기로 사용되는 은닉 마르코프 모델을 분별적으로 학습하는 기법을 제안한다. 기존에 많이 사용되는 Baum-Welch 알고리즘에서는 각 모델이 해당 클래스 데이터의 확률을 최대화하는 것을 목표로 학습시키는 반면, 제안하는 알고리즘에서는 클래스간의 분별력을 높이기 위해 두 가지의 최소화 목적함수로 이루어진 새로운 학습 목표를 정의하고 이를 달성하기 위해 모의 담금질 기법에 기반을 둔 다목적함수 전역 최적화 기법을 개발한다. 화자종속 인식 실험을 통해 제안하는 기법의 성능을 평가하며, 실험결과 기존의 학습 방법에 비해 오인식율을 상대적으로 약 8% 감소시킬 수 있음을 보인다.

키워드 : 자동 독순, 다목적함수 최적화, 분별학습, 모의 담금질

Improved Automatic Lipreading by Multiobjective Optimization of Hidden Markov Models

Jong-Seok Lee[†] · Cheol Hoon Park[††]

ABSTRACT

This paper proposes a new multiobjective optimization method for discriminative training of hidden Markov models (HMMs) used as the recognizer for automatic lipreading. While the conventional Baum-Welch algorithm for training HMMs aims at maximizing the probability of the data of a class from the corresponding HMM, we define a new training criterion composed of two minimization objectives and develop a global optimization method of the criterion based on simulated annealing. The result of a speaker-dependent recognition experiment shows that the proposed method improves performance by the relative error reduction rate of about 8% in comparison to the Baum-Welch algorithm.

Key Words : Automatic Lipreading, Multiobjective Optimization, Discriminative Training, Simulated Annealing

1. 서 론

자동 독순은 화자의 입술의 움직임을 관찰하여 음성을 인식하는 기술이다. 음성의 시각적 측면은 인간이나 컴퓨터의 음성인식에서 보조적인 정보로 유용하게 사용될 수 있으며, 특히 잡음이 존재하는 실제 인식 환경에서 강인한 인식을 수행하는 데에 도움이 된다[1,2]. 말소리 정보를 이용한 기존의 음성인식은 잡음이 없는 환경에서는 상당히 높은 성능을 보이지만, 잡음이 존재하는 경우 인식 성능이 크게 저하되는 단점이 있다. 독순은 잡음이 없는 환경에서 말소리 정보를 이용한 음성인식에 비해 성능이 다소 떨어지지만 잡음에 영향을 받지 않기 때문에 잡음 환경에서 음성인식의 성능 저하를 보상할 수 있는 수단으로 사용될 수 있다.

자동 독순에서의 인식기는 음성인식과 마찬가지로 은닉 마르코프 모델(HMM: hidden Markov model)이 많이 사용된다[2,3]. HMM은 이중의 확률 모델로써 시간에 대해 탄력적으로 변화하는 음성의 특징을 모델링하기 적합하다. HMM을 자동 독순 시스템에서 사용하기 위해서는 먼저 기록된 입술 움직임 동영상으로부터 추출된 특징으로 학습하는 과정을 거쳐야 한다. 지금까지 HMM의 학습에 가장 많이 이용되는 방법은 Baum-Welch 알고리즘이다[3]. 이 알고리즘은 학습 데이터에 대한 우도(likelihood)를 최대화하는 최대우도(ML: maximum likelihood)를 목표로 하여 HMM의 파라미터를 반복적으로 갱신함으로써 학습을 수행한다. ML에 의한 파라미터 학습은 클래스 간 구분을 필요로 하는 패턴인식 문제를 각 클래스 데이터의 확률 분포를 추정하는

※ 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

DOI: 10.3745/KIPSTB.2008.15-B.1.53

것으로 해결하고자 하는 것으로 볼 수 있다. 만일 주어진 데이터의 분포를 HMM으로 정확하게 모델링할 수 있고 학습 데이터가 무한히 주어진다면 ML을 목표로 하는 학습을 통해 최적의 인식 성능을 얻을 수 있다. 그러나 실제로 학습 데이터는 유한하며, 음성의 청각 또는 시각신호의 분포에 대한 지식이 완전하지 않아 HMM에 의해 그 분포를 정확하게 모델링하지 못할 수도 있기 때문에 ML 학습에 의한 성능은 최적 인식기의 성능에 미치지 못할 수도 있다[4].

ML과는 달리 HMM의 인식 성능을 높이는 것을 학습 목표로 하는 분별학습(discriminative training) 기법들이 제안된 바 있다[5,6]. ML 추정에서는 하나의 HMM의 학습에는 해당 클래스의 데이터만을 이용하는 반면, 분별학습 기법에서는 좋은 인식 성능을 얻기 위해 해당 클래스 뿐 아니라 다른 클래스의 데이터를 함께 이용한다. 분별학습 기법에서 중요한 것은 인식 성능을 향상시킬 수 있는 학습 목표를 설정하고 그것을 위한 최적화 기법을 개발하는 것이다. 기존의 대표적인 분별학습 기법으로 최소분류오류(MCE: minimum classification error) 기법[5]과 최대상호정보(MMI: maximum mutual information) 기법[6]을 들 수 있다. MCE 기법에서는 오인식율을 근사화하는 목적함수를 정의하고 일반화된 확률적 경사(generalized probabilistic descent) 알고리즘에 의해 이를 최소화하는 방향으로 HMM의 파라미터를 학습한다. MMI 기법에서는 하나의 클래스에 대한 HMM과 그 클래스에 속하는 데이터 사이의 상호정보(mutual information)가 최대화되는 것을 목표로 하며 Baum-Welch 알고리즘과 유사한 형태의 알고리즘에 의해 HMM을 학습한다. 이들 분별학습 기법이 최대우도를 목표로 하는 Baum-Welch 알고리즘에 비해 더 좋은 성능을 보이는 것으로 알려져 있지만, 학습 알고리즘이 모두 지역 최적화에 의존하기 때문에 초기해에 따라 최적해를 찾지 못할 가능성이 있는 단점이 있다.

본 논문에서는 HMM의 분별학습을 위해 두 개의 목적함수로 구성된 새로운 학습목표를 제안하고 이 학습목표의 최적화를 위해 모의 담금질(SA: simulated annealing) 기법[7]에 기반한 다목적함수 전역 최적화 기법을 개발한다. HMM을 이용한 인식기에서는 인식을 위한 데이터가 주어질 때 그 데이터에 대한 각 HMM에서의 확률, 즉 우도를 계산하고 그 중 가장 큰 값을 선택하여 인식을 수행한다. 그러므로 좋은 인식 성능을 위해서는 각 HMM은 해당 클래스의 데이터에 대해서는 높은 확률값을, 다른 클래스의 데이터에 대해서는 낮은 확률값을 나타내는 것이 바람직하다. 따라서 제안하는 학습 목표에서는 각 HMM에 대해 해당 클래스의 데이터에 대한 확률은 최대화하고 동시에 다른 클래스의 데이터에 대한 확률은 최소화하는 것을 목표로 한다. 이 두 목적함수는 서로 상충되는 것으로써 이에 따른 HMM의 학습은 다목적함수 최적화 문제로 귀결된다. 본 논문에서는 이 문제의 최적해를 찾기 위해 SA에 기반한 다목적함수 최적화 알고리즘을 개발한다. 이 알고리즘은 간단한 구조를 가지고 있으면서도 기존의 분별학습 기법과 달리 전역 최적화를 수행할 수 있는 장점이 있다. 실험을 통해 제안하는 기법이 Baum-Welch 알고리즘 및 기존의 분별학습 기법에 비해 좋은 인식 성능을 나타냄을 보인다.

이하 논문의 구성은 다음과 같다. 다음 장에서는 자동 독순 시스템에 대해 설명한다. 3장에서는 제안하는 HMM의 분별학습을 위한 최적화 기준, 다목적함수 최적화를 위한 알고리즘, 그리고 다양한 최종해 중 선택을 통해 인식기를 구성하는 의사 결정 과정을 설명한다. 4장에서는 실험결과를 보이고 5장에서 맺음말로 논문을 맺는다.

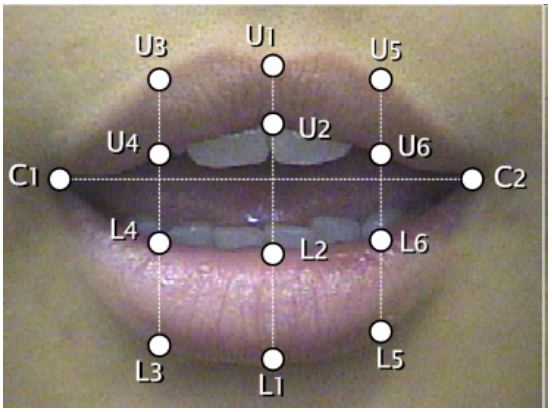
2. 자동 독순

자동 독순은 다음과 같은 과정으로 이루어진다. 화자가 말을 하면 비디오카메라를 통해 화자의 입술의 움직임을 기록한 동영상을 얻는다. 기록된 영상에서 적절한 특징을 찾고 이를 사전에 학습된 인식기에 통과시킴으로써 인식 결과를 얻는다.

각 영상에서 추출하는 특징은 입술의 윤곽선으로부터 얻는다. 먼저 (그림 1)에 보인 것과 같이 윤곽선상의 14개의 점을 찾는다. 이 점들은 여러 컬러 정보를 이용하여 자동으로 추출할 수 있다[8]. 특징을 얻기 위해 입 양 끝점을 잇는 선분($\overline{C_1C_2}$)으로부터 경계점 $U_1, U_2, \ldots, U_6$ 및 $L_1, L_2, \ldots, L_6$의 높이를 측정한다. 그리고 이 높이들을 입술 너비로 나누어 정규화된 최종 특징을 얻는다. 일부 화자의 경우 좌우 입 모양이 비대칭이기 때문에 양쪽 모두를 사용하는 것이 인식에 도움이 된다.

이렇게 구한 12개의 정적 특징(static feature)과 더불어 이것의 시간 미분으로 정의되는 동적 특징(dynamic feature)을 얻는다[3]. 따라서 각 영상마다 총 24차원의 특징벡터를 얻으며, 벡터 양자화를 통해 128개의 부호어(code word)중 하나로 변환하여 최종 특징을 얻는다.

인식기는 이산 HMM으로 구성된다. 하나의 HMM $\lambda=(\Pi, A, B)$은 초기상태 확률분포인 $\Pi=\{\pi_i\}$, 상태전이 확률분포인 $A=\{a_{ij}\}$, 그리고 관측확률분포인 $B=\{b_i\}$로 이루어진다. 각 클래스의 발음은 학습을 통해 하나의 HMM으로



(그림 1) 특징 추출을 위한 입술의 윤곽선 상의 경계점

모델링된다. 학습된 HMM들을 이용한 인식 과정에서는 소속 클래스를 알 수 없는 발음 데이터에 대한 특징을 모든 HMM에 입력하고 가장 높은 확률을 보이는 HMM을 선택하여 인식 결과를 얻는다.

3. 제안하는 최적화 기법

3.1. 분별학습 목표

학습 데이터를 이용하여 HMM을 학습할 때 가장 많이 사용되는 방법은 ML에 의한 추정이다. i번째 클래스에 해당하는 HMM λ^i을 학습하기 위한 ML의 목표는 로그우도(log-likelihood)를 최대화하는 것으로써,

$$\text{Maximize} \quad f(\lambda^i) = \sum_{j=1}^{N_i} \log P(O_j^i|\lambda^i) \tag{1}$$

로 주어진다. 여기서 O_j^i는 i번째 클래스의 j번째 학습 데이터이고 N_i는 이 클래스의 학습 데이터의 갯수이다. 식 (1)의 최적화에는 Baum-Welch 알고리즘이 가장 많이 이용되며 알고리즘의 재추정 공식을 반복적으로 적용하여 로그우도를 단조증가시킬 수 있다[3].

ML 추정이 간단하면서도 좋은 성능을 보이지만 머리말에서 서술한 바와 같이 학습 데이터가 유한하고 모델링하고자 하는 발음과 모델인 HMM간의 확률분포의 불일치 가능성으로 인해 최대의 인식 성능을 보장하지는 않는다. 이를 해결하기 위해 분별학습 기법에서는 식 (1) 대신에 인식 성능을 최대화할 수 있는 새로운 목표함수를 이용한다.

주어진 데이터에 대해 옳은 인식 결과를 얻기 위해서는 해당 클래스의 HMM에서의 확률이 다른 클래스의 HMM에 비해 커야 한다. 다시 말해, 하나의 HMM은 해당 클래스의 데이터에 대해서는 높은 확률을, 다른 클래스의 데이터에 대해서는 낮은 확률을 내는 것이 바람직하다. 따라서 본 논문에서는 HMM을 분별적으로 학습시키기 위한 새로운 학습 목표를 다음과 같이 제안한다. i번째 클래스에 해당하는 HMM λ^i의 학습 목표를

$$\text{Minimize} \quad f_1(\lambda^i) = -\sum_{j=1}^{N_i} \log P(O_j^i|\lambda^i) \tag{2}$$

$$\text{Minimize} \quad f_2(\lambda^i) = \sum_{\substack{w=1 \\ w \neq i}}^{W} \sum_{j=1}^{N_w} \log P(O_j^w|\lambda^i)$$

로 한다. 여기서 O_j^w는 w번째 클래스의 j번째 학습 데이터, W는 총 클래스의 수, 그리고 N_w는 w번째 클래스의 학습 데이터의 수이다. 첫번째 목적함수 f_1은 식 (1)의 ML 목표와 같다. 다만 최소화 문제로 바꾸기 위해 식에 음의 부호를 붙였다. HMM의 파라미터를 위의 목표에 의해 학습하는 것은 해당 클래스의 데이터에 대한 로그우도를 증가시키고 다른 클래스의 데이터에 대한 로그우도는 감소시키는 것이다. 따라서, 인식시에 클래스 간의 분별력이 향상된 HMM을 얻을 수 있다.

식 (2)의 학습 목표는 목적함수가 하나가 아닌 복수이기 때문에 이에 대한 해는 하나로 주어지지 않으며 서로 우열관계가 없는 다수의 해가 존재한다. 이와 같은 다목적함수 최적화 문제에서 하나의 해 x가 다른 해 y를 '지배한다'는 것은 모든 목적함수에 대해 x가 y보다 나쁘지 않고 적어도 하나의 목적함수에 대해서는 x가 더 좋은 것을 의미한다. x가 y를 지배하지 않고 y도 x를 지배하지 않을 때 두 해는 서로 우열관계가 없다고 한다. 다목적함수 최적화 문제의 최적해를 파레토(Pareto) 해라 하는데, 파레토 해 집합은 다른 어떤 해에 의해서도 지배되지 않는다. 식 (2)의 학습목표에 대한 파레토 최적해를 구하는 방법을 다음 절에서 설명한다.

3.2. 학습 알고리즘

본 절에서는 SA 기법을 이용한 다목적함수 최적화를 통해 제안하는 분별학습 목표인 식 (2)를 최적화하는 방법을 설명한다. SA는 Kirkpatrick에 의해 제안된 확률적 최적화 기법으로써 다양한 분야에서 발생하는 최적화 문제들에서 전역해를 찾기 위한 방법으로써 성공적으로 응용되어왔다[7]. 초기해로부터 시작해서 새로운 해를 확률적으로 생성하여 평가하고, 현재의 해와 새로운 해의 경쟁을 통해 다음 반복(iteration)의 해를 선택하는 과정을 되풀이함으로써 최적화을 수행한다. SA에서 중요한 파라미터인 온도는 새로운 해를 생성하는 과정과 해의 선택 과정에 관여하는데, 알고리즘이 진행될수록 온도를 낮추는 '담금질 과정'을 통해 점차 해가 수렴하도록 한다. SA에서는 현재의 해보다 나쁜 해로의 전이가 0보다 큰 확률로 일어나기 때문에 지역 최적해를 벗어날 수 있고 따라서 전 해공간에서의 탐색이 가능하다. SA의 특징 중 하나는 알고리즘에 의해 해가 전역 최적해로 수렴함이 수학적으로 증명되어 있다는 것이다[9].

본 논문에서 제안하는 알고리즘은 SA를 이용하여 식 (2)로 주어지는 다목적함수 최적화 문제에서 파레토 해집합을 찾기 위한 것이며, SA의 여러 형태 중 빠른 수렴 속도와 좋은 성능을 보이는 빠른 SA(fast SA)[10] 기법에 기반한다. 즉, 새로운 해를 생성하기 위해 코쉬(Cauchy) 확률분포함수를 사용하고 담금질 과정을 위해 반복회수의 역수에 비례하는 온도함수를 사용한다. 그리고 제안하는 알고리즘에서는 탐색 과정의 속도 향상을 위해 지역 최적화 기법을 함께 사용한다. 알고리즘의 자세한 과정은 다음과 같다.

1) P개의 초기해 $\{\lambda_0^p\}_{p=1}^P$를 무작위로 생성하고 초기 온도를 설정한다. 하나의 해 λ_k^p는 이산 HMM의 파라미터인 Π, A 및 B의 값들을 벡터 형태로 만든 것이다.
2) 현재의 해집합 $\{\lambda_k^p\}$로부터 새로운 해집합 $\{\lambda'^p_k\}$를

생성한다. 즉,

$$\lambda'^{p}_{k} = \lambda^{p}_{k} + \Delta\lambda^{p}_{k} \tag{3}$$

여기서 k는 반복회수이다. p번째 해의 변화량 $\Delta\lambda^{p}_{k}$는 다음으로 주어지는 코쉬 확률분포함수에 의해 확률적으로 결정된다[11].

$$g(\Delta\lambda^{p}_{k}) = \frac{a_n T_k}{(\|\Delta\lambda^{p}_{k}\|^2 + T_k^2)^{(n+1)/2}} \tag{4}$$

여기서 T_k는 k번째 반복에서의 온도, n은 학습하는 HMM 파라미터의 갯수(즉, 해의 차원), 그리고 a_n은 정규화를 위한 상수이다. 알고리즘의 초기에 온도가 높으면 $\Delta\lambda^{p}_{k}$가 큰 값을 가질 확률이 높아 해의 변화가 클 가능성이 높아지고 따라서 넓은 공간에서의 탐색이 가능하다. 반대로 알고리즘의 후반에 온도가 낮아지면 해의 변화가 클 확률이 낮아져 미세한 해의 조정이 가능하다.

3) 새로운 해 $\{\lambda'^{p}_{k}\}$의 각각에 지역 최적화를 적용하여 $\{\lambda''^{p}_{k}\}$를 얻는다. Baum-Welch 알고리즘을 한 번 적용하는 것을 지역 최적화로 하며, 이 과정은 알고리즘의 탐색 과정을 빠르게 한다.
4) $\{\lambda''^{p}_{k}\}$의 목적함수값인 f_1과 f_2를 식 (2)에 의해 계산하고, 이를 바탕으로 현재의 해집합 $\{\lambda^{p}_{k}\}$와 새로운 해집합 $\{\lambda''^{p}_{k}\}$의 비용함수(cost function)의 값을 계산한다. 비용함수는 '샘플링에 의한 파레토 기반 비용'으로 정의되는데[12], 하나의 해 x의 비용함수값 $C(x)$은 현재의 해와 새로운 해들 중 x를 지배하는 해의 개수로 정의된다. (아래 설명 참조) 비용함수 값이 작을수록 다른 해에 의해 지배되지 않음을 의미하므로 더 좋은 해이다.
5) 현재의 해집합 $\{\lambda^{p}_{k}\}$과 새로운 해집합 $\{\lambda''^{p}_{k}\}$에서 다음 반복을 위한 P개의 해를 선택한다. 하나의 현재 해와 그로부터 생성된 해 중 하나를 선택하는 과정을 P번 반복하는데, 여기에는 메트로폴리스 기법이 사용된다[13]. 즉, λ^{p}_{k}와 λ''^{p}_{k}중 λ''^{p}_{k}가 선택될 확률은

$$p_a = \begin{cases} 1 & \text{if } C(\lambda^{p}_{k}) > C(\lambda''^{p}_{k}) \\ \exp\left(\dfrac{C(\lambda^{p}_{k}) - C(\lambda''^{p}_{k})}{T_k}\right) & \text{otherwise} \end{cases} \tag{5}$$

로 주어진다. 다시 말해, 현재의 해보다 좋은 새로운 해는 항상 선택되며, 그렇지 않으면 새로운 해는 확률적으로 선택되는데 비용함수값의 차가 작을수록 확률은 커진다. 위 식에서 현재보다 나쁜 해를 선택할 확률은 온도에 의존하는 것을 볼 수 있는데, 알고리즘 초기에 온도가 높으면 선택확률이 커서 현재보다 나쁜 해로의 전이를 쉽게 하여 지역 최적해를 벗어나기 쉽게 하고, 온도가 낮아짐에 따라 현재보다 나쁜 해의 선택확률은 점차 줄어들어 좋은 해를 유지하면서 최종해로 수렴하도록 하는 것이다.
6) 온도를 다음의 담금질 과정에 의해 낮춘다.

$$T_k = T_0/k \tag{6}$$

여기서 T_0는 초기온도이다.
7) 종료조건을 만족하면 알고리즘을 끝내고 그렇지 않으면 2)단계로 돌아간다. 종료조건으로는 최대 반복회수를 사용한다.

위의 과정은 각 클래스에 대한 HMM에 대해 반복된다.

현재의 해로부터 생성되는 새로운 해는 코쉬 확률분포에 의해 무작위로 생성되고 한 번의 Baum-Welch 알고리즘에 의해 다시 변이된다. 이처럼 지역 최적화 기법을 결합함으로써 탐색 과정의 속도를 향상시키고 좋은 해를 얻을 수 있다.

각 해의 비용함수값은 '파레토 기반 비용'으로 계산된다. 기존 연구에서는 파레토 기반 비용함수를 사용하는 SA 기반 다목적함수 최적화 기법에서 모든 해집합이 전역 최적해에 고르게 수렴함을 수학적으로 보였다[12]. 하나의 해 x의 파레토 기반 비용함수는 다음과 같이 정의된다.

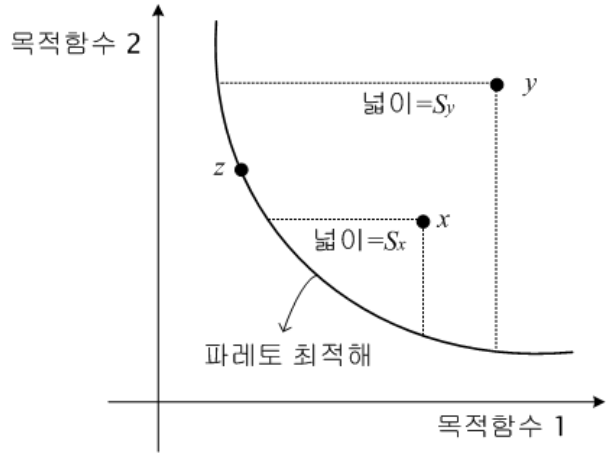
$$C(x) = \int_S dom(y,x)\,dy \tag{7}$$

여기서 S는 전체 해 공간을 의미하며 $dom(y,x)$는

$$dom(y,x) = \begin{cases} 1 & \text{if } y \text{ dominates } x \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

로 정의된다. (그림 2)는 파레토 기반 비용함수의 개념을 설명하기 위해 목적함수가 두 개일 때 목적함수 평면상에 존재하는 세 해 x, y, z를 나타낸 그림이다. S_x 영역 내에 있는 해들은 x를 지배하므로 x의 파레토 기반 비용함수값은 S_x가 된다. 마찬가지로 y의 비용은 S_y이다. z를 지배하는 해는 존재하지 않기 때문에 z의 비용은 0이다. $C(y) > C(x) > C(z)$임을 볼 수 있는데 이는 각 해와 파레토 최적해집합 사이의 거리를 반영함을 알 수 있다.

그러나 식 (7)을 계산하기 위해서는 최적해를 포함해서 전체 해공간에 존재하는 해의 정보가 필요한데, 이는 실제적으로 불가능하다. 이 경우 다음의 식과 같이 샘플링에 의해 파레토 기반 비용함수를 구한다[12].



(그림 2) 목적함수 평면 상의 해와 파레토 기반 비용함수. 파레토 최적해 집합 오른쪽이 해가 존재할 수 있는 공간임

$$\tilde{C}(x) = \sum_{p=1}^{2P} dom(y_p, x) \tag{9}$$

여기서 y_p는 P개의 현재 해와 P개의 새로운 해 중 p번째 해를 의미한다.

다목적함수 최적화를 위한 기존의 기법들은 진화 알고리즘이나 유전자 알고리즘에 기반하는 경우가 많다[14]. 이러한 알고리즘들은 해의 경쟁 과정에서 가장 좋은 소수의 해가 집중적으로 선택되어 전체 해집합이 모두 하나의 해로 수렴하는 현상이 쉽게 발생하는데, 이를 방지하고 해의 다양성을 확보하기 위해 적합도 공유(fitness sharing)이나 영역 유도(niche induction)와 같은 추가적인 기법이 알고리즘 내에서 이용되기도 한다[15]. 그러나 본 논문에서 제안하는 기법은 하나의 해와 그로부터 생성된 새로운 해 사이에서만 경쟁이 일어나기 때문에 추가적인 장치 없이도 해집합 전체가 하나의 해로 수렴할 가능성이 적은 것이 특징이다.

3.3. 인식기 구성

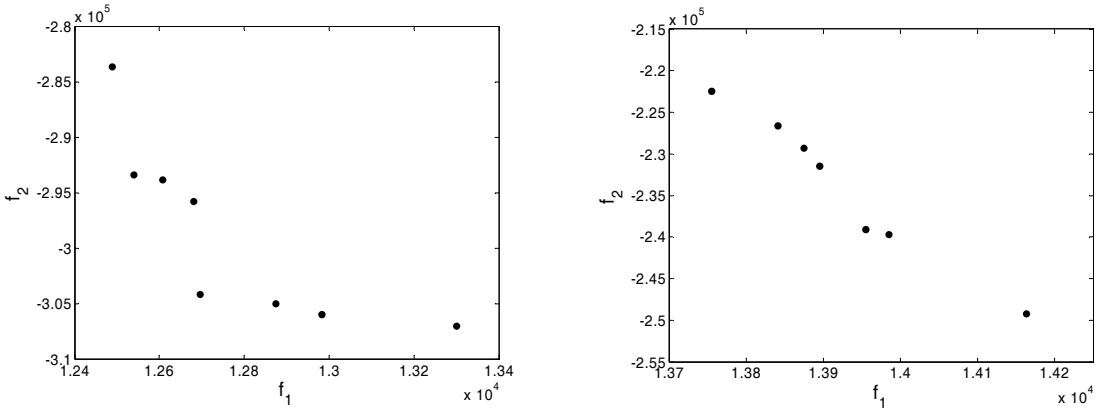
각 클래스에 대해 우열관계가 없는 최적해의 집합을 찾은 다음, 각 클래스마다 하나의 해, 즉 하나의 HMM 파라미터 집합을 선택하고 그 HMM의 조합으로 인식기를 구성해야 한다. 총 W개의 클래스가 있고 각 클래스마다 M개의 최종해를 얻었다면 인식기 구성을 위해 M^W개의 가능한 조합에서 하나를 선택해야 한다. 그러나 그 모든 조합을 모두 시험해볼 수는 없기 때문에 다음과 같은 방법으로 인식기를 구성한다. 각 클래스의 최종해 집합에서 무작위로 해를 하나씩 선택하고 그것들의 조합으로 인식기를 구성한 후 학습 데이터에 대한 인식율을 측정한다. 이러한 과정을 여러 번 반복한 후 가장 좋은 인식율을 보이는 조합을 최종 인식기로 선택한다. 다음 장의 실험 결과에서 여러 최적해들 중 어떤 것을 선택하는 것이 좋은 인식기를 구성하는데 도움이 되는지 논의한다.

4. 실 험

실험에서 사용하는 데이터베이스는 18명의 화자(남자 8명, 여자 10명)가 우리말 숫자인 “하나”부터 “열”까지를 열 번씩 발음한 것을 30 Hz의 프레임 비율로 기록한 데이터베이스이다[8]. 8번의 발음을 학습에, 나머지 2번의 발음을 인식 테스트에 사용하였다. 각 영상은 360×480 픽셀 크기로 기록되었다. 2장에서 설명한 것과 같이 각 영상마다 24차원의 특징을 추출하고 벡터 양자화를 통해 128개의 코드어 중 하나로 이산 변환하였다. 각 단어는 8개의 상태를 가지는 이산 HMM으로 모델링하였다.

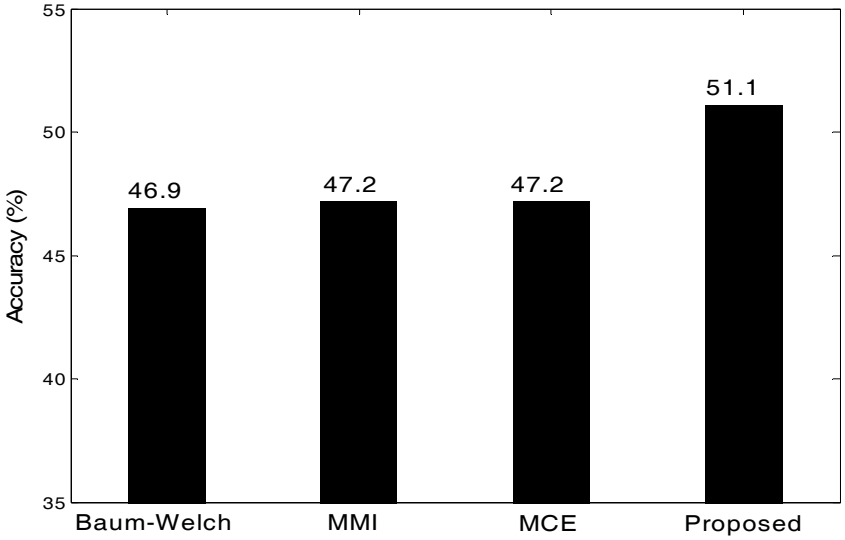
제안하는 학습 알고리즘에서는 25개의 개체를 사용하였으며(즉, P=25) 초기 온도를 10으로, 최대 반복횟수를 50000번으로 하였다. 제안하는 알고리즘을 Baum-Welch 알고리즘, MCE 기법 및 MMI 기법과 비교한다. 두 분별학습 기법은 사용자가 정해야 하는 알고리즘 파라미터를 가지는데 여러 번의 실험을 통해 좋은 성능을 내는 값으로 선택하였다.

(그림 3)은 제안하는 학습 기법의 결과로 얻는 최종 해의 예를 목적함수 공간에서 나타낸 것이다. 각 단어마다 약 7~8개의 우열관계가 없는 해를 얻었다.



(a) (b)

(그림 3) 최적화에 의해 얻어진 해의 예 (a) 7번째 클래스 (b) 8번째 클래스



(그림 4) 기존의 기법과 제안하는 기법에 의한 인식율 비교

(그림 4)는 제안하는 알고리즘과 기존의 세 알고리즘에 의한 테스트 인식 성능(테스트 데이터 중 옳게 인식한 단어의 비율)을 비교한 것이다. 제안하는 기법에서는 최종 인식기 설계를 위해서 무작위로 100가지의 해의 조합을 시도하여 학습 데이터에 대해 가장 좋은 인식율을 보이는 경우를 선택하였다. 제안하는 기법의 인식율은 51.1%로 나타나 46.9%의 인식율을 보인 Baum-Welch 알고리즘에 비해 약 8%의 상대적 오인식율 감소를 얻을 수 있었다. 분별학습 기법들은 Baum-Welch 알고리즘에 비해 더 나은 성능을 보이지만 성능향상의 정도는 제안하는 기법에 비해 낮음을 볼 수 있다. 이는 MCE 기법이나 MMI 기법 모두 지역 최적화에 의해 해를 찾는 반면 제안하는 기법은 전역 최적화를 수행하기 때문이다.

인식기 구성을 위한 해의 선택 과정에서 주로 첫번째 목적함수값이 작은, 즉 해당 클래스의 데이터에 대해 큰 로그우도를 갖는 해가 선택되는 경향이 있음을 발견하였다. 모든 클래스에 대해 목적함수 평면에서 우열관계가 없는 최종해 중 가장 왼쪽에 있는 해, 즉 첫번째 목적함수값이 가장 작은 해로 구성된 인식기의 경우 50.8%의 인식율을 얻었는데 이는 100번의 무작위 조합 중 가장 좋은 것을 선택한 결과와 거의 비슷한 것이다. 목적함수 평면에서 가장 오른쪽에 있는 해, 즉 두번째 목적함수값이 가장 작은 해로 인식기를 구성한 경우 43.6%의 인식율을 얻었다. 최종 해 중 중간에 위치한 해들로 구성된 인식기는 48.6%의 인식율을 보였다. 따라서, 첫번째 목적함수값이 작은 해 위주로 선택하는 것이 좋은 성능의 인식기를 구성하는데 도움이 됨을 알 수 있다.

제안하는 알고리즘에 의한 성능 향상은 기존의 기법에 비해 상대적으로 많은 계산량을 대가로 하여 얻어진다. 기존의 알고리즘들이 학습에 수 분 정도를 소요하는 반면, 제안하는 알고리즘은 각 클래스의 HMM이 학습 과정에서 서로 무관함을 이용하여 여러 대의 컴퓨터를 이용하여 병렬로 학습을 시켰음에도 불구하고 수 시간의 학습 시간이 필요하였으며 이는 기존의 알고리즘들에 비하면 계산량이 크게 증가한 것이다. 그러나 이러한 계산량 증가는 인식기의 학습 과정에서만 필요한 것으로써 학습 과정은 인식기를 사용하기 전에 충분한 시간을 갖고 미리 해 둘 수 있으며, 실제 인식기로써 동작하는 과정에서는 모든 기법들이 차이가 없다.

5. 맺음말

본 논문에서는 자동 독순의 성능 향상을 위해 다목적함수 최적화 기법을 이용한 HMM의 분별학습 기법을 제안하였다. 인식기의 성능을 향상시키기 위해 두 가지 목적함수로 구성된 새로운 학습 목표를 정의하고 SA에 기반을 둔 최적화 기법을 개발하였다. 실험결과 제안하는 기법이 기존의 Baum-Welch 알고리즘과 분별학습 기법인 MCE 및 MMI 기법에 비해 더 좋은 인식 성능을 보이는 인식기를 구성할 수 있음을 확인하였다. 추후연구로는 제안하는 기법을 연속 HMM의 학습으로 확장하는 것과 다양한 데이터베이스에 적용하는 것을 들 수 있다.

참 고 문 헌

[1] L. A. Ross, D. Saint-Amour, V. M. Leavitt, D. C. Javitt, and J. J. Foxe, “Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments,” Cerebral Cortex, Vol. 17, No. 5, pp. 1147-1153, 2007.

[2] 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. 23-37, 2002.

[3] L. Rabiner and B.-H. Juang, 'Fundamentals of Speech Recognition,' Prentice-Hall, 1993.

[4] W. Chou, "Discriminant-function-based minimum recognition error rate pattern-recognition approach to speech recognition," Proc. IEEE, Vol. 88, No. 8, pp. 1201-1223, 2000.

[5] 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, 1997.

[6] 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, 2004.

[7] S. Kirkpatrick, C. D. Gerlatt, and M. P. Vecchi, "Optimization by simulated annealing," Science, Vol. 220, pp. 671-680, 1983.

[8] 이종석, 심선희, 김소영, 박철훈, "제어되지 않은 조명 조건하에서 입술움직임의 강인한 특징추출을 이용한 바이모달 음성인식," Telecommunications Review, 14권 1호, pp. 123-134, 2004.

[9] R. L. Yang, "Convergence of the simulated annealing algorithm for continuous global optimization," J. Optimization Theory and Applications, Vol. 104, No. 3, pp. 691-716, 2004.

[10] H. H. Szu and R. L. Hartley, "Fast simulated annealing," Phys. Lett. A, Vol. 122, No. 3-4, pp. 157-162, June 1987.

[11] D. Nam, J.-S. Lee, and C. H. Park, "n-dimensional Cauchy neighbor generation for the fast simulated annealing," IEICE Trans. Information and Systems, Vol. E87-D, No. 11, pp. 2499-2502, 2004.

[12] D. Nam and C. H. Park, "Pareto-based cost simulated annealing for multiobjective optimization," Proc. Asia-Pacific Conf. Simulated Evolution and Learning, Vol. 2, pp. 522-526, Singapore, 2002.

[13] 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.

[14] K. Deb, A. Pratap, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," IEEE Trans. Evolutionary Computation, Vol. 6, No. 2, pp. 182-197, Apr. 2002.

[15] E. Zitzler, M. Laumanns, and S. Bleuler, "A tutorial on evolutionary multiobjective optimization," Metaheuristics for Multiobjective Optimisation, Lecture Notes in Economics and Mathematical Systems, X. Gandibleux, M. Sevaux, K. Sörensen, and V. T'kindt, Eds. Springer-Verlag, Vol. 535, pp. 3-37, 2004.

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