한국음향학회지 제22권 제2호 pp. 145~152 (2003)

은닉 마코프 모델 확률 보정을 이용한 음성 인식 성능 향상

Performance Improvement in Speech Recognition by Weighting HMM Likelihood

권 태 희*, 고 한 석*
(Tae-Hee Kwon*, Han-Seok Ko*)

*고려대학교 전자·컴퓨터공학과

(접수일자: 2002년 12월 16일; 채택일자: 2003년 1월 9일)

본 논문에서는 인식 단위로서의 개개의 은닉 마코프 모델 (HMM: Hidden Markvo Model)에 대응하는 가중치를 도입하여 HMM 출력 스코어는 HMM 출력 확률과 HMM 가중치의 곱으로 표현된다고 가정하고 기존의 최소 분류 오류 훈련 방법과 유사하게 HMM 가중치를 반복적으로 훈련하는 방법을 제안하였다. 제안된 방법은 오인식 척도에 대해 차분 (delta) 계수를 정의하고 이를 이용하여 HMM 가중치를 반복하여 훈련하는 방법이다. 이러한 방법은 HMM 가중치의 합을 HMM 개수의 총합으로 제한함으로써 기존의 파라미터 추정 방법과 비터비 (Viterbi) 알고리즘에 큰 변화없이 음성 인식에 효과적으로 적용될 수 있다. 제안된 방법은 기존의 분할 (segmental) 최소 분류 오류 훈련 방법과 비교하여 추정하는 파라미터의 개수가 감소되었으며 훈련 모델의 최적 상태열을 이용한 경도 계산 과정이 포함되지 않음으로써 계산량을 효과적으로 단축할 수 있다. HMM 가중치를 이용한 HMM 기반의 음성 인식기의 성능 평가를 위해서 단독 숫자음 인식 실험을 실시하였다. 실험적 결과들은 HMM 확률 보정을 이용한 음성 인식 시스템이 베이스라인 시스템보다 음성 인식 성능이 더 우수함을 보여준다. 제안된 방법은 기존의 최소 분류 오류 훈련 방법에 비하여 구현하기 간편한 반면에 더욱 우수한 음성 인식 성능 향상을 보여준다.

핵심용어: 은닉 마코프 모델링, 최소 분류 오류 알고리즘, HMM 확률 보정, 차분 계수
투고분야: 음성처리 분야 (2.5)

In this paper, assuming that the score of speech utterance is the product of HMM log likelihood and HMM weight, we propose a new method that HMM weights are adapted iteratively like the general MCE training. The proposed method adjusts HMM weights for better performance using delta coefficient defined in terms of misclassification measure. Therefore, the parameter estimation and the Viterbi algorithms of conventional HMM can be easily applied to the proposed model by constraining the sum of HMM weights to the number of HMMs in an HMM set. Comparing with the general segmental MCE training approach, computing time decreases by reducing the number of parameters to estimate and avoiding gradient calculation through the optimal state sequence. To evaluate the performance of HMM-based speech recognizer by weighting HMM likelihood, we perform Korean isolated digit recognition experiments. The experimental results show better performance than the MCE algorithm with state weighting.

Keywords: *Hidden Markov modeling, Minimum classification error, HMM likelihood weight, Delta coefficient*
ASK subject classification: *Speech signal processing (2.5)*

I. 서 론

은닉 마코프 모델 (HMM: Hidden Markov Model)은 음성 인식에서 음성의 특징을 표현하는데 매우 널리 사용되고 있다. 은닉 마코프 모델은 통계적인 기반 위에서 음성의 특징을 모델링하는 일을 HMM 상태의 확률 분포 추정 방법으로 귀결시킨다. 이러한 상태 확률 분포 추정 방법으로써 가장 널리 사용되는 방법은 ML (Maximum Likelihood) 추정 방법이다[1,2].

책임저자: 권태희 (thkwon@ispl.korea.ac.kr)
136-701 서울시 성북구 안암동 5가 1번지
고려대학교 전자·컴퓨터 공학과
(전화: 02-927-6115; 팩스: 02-3291-2450)

그러나 ML 추정 방법은 훈련 데이터의 확률 분포에 대한 정확한 정보를 알아내는 일이 매우 어려우며 유사 음성들 사이의 구분되는 정보를 모델링하기가 거의 불가능하다. 패턴 인식 문제에서 위와 같은 확률 분포 추정의 단점을 보완하기 위한 방법으로 분별 훈련 방법이 많이 연구되어 왔다[2-4].

특히 음성 인식기의 성능은 인식기의 평균 인식 에러율로 정의되며 최적의 인식기는 최소의 인식 에러율을 나타내는 인식기이다. 이러한 관점에서 주로 연구되어 온 분별 훈련 기법이 GPD (Generalized Probabilistic Descent) 알고리즘에 기반한 최소 분류 오류 (MCE: Minimum Classification Error) 훈련 방법이다. 최소 분류 오류 훈련 방법은 궁극적으로 인식 오류의 최소화 관점에서 접근하며 특징 추출, 음향 모델링 기법, 음향 모델의 정밀도가 고정된 상태에서 혼합성분 (mixture) 가중치, 평균, 표준편차 등의 HMM 기본 파라미터들을 조절해 줌으로써 음성 인식의 성능 향상을 달성할 수 있음이 이미 연구되어 왔다[5-9]. 최소 분류 오류 훈련 방법의 확장된 방법으로써 HMM 상태 확률에 내재해 있는 음성들 사이의 구분되는 정보를 이용하여 인식기를 최적화하기 위해서 상태 가중치를 도입하는 방법이 연구되었다[10]. 최소 분류 오류 훈련 기법은 주로 ML 훈련 기법과 병행하여 행해지고 있으며 ML 훈련 방법에 의해 추정된 HMM보다 우수한 성능을 보인다. 그러나 분할 GPD 알고리즘에 기반한 최소 분류 오류 훈련은 훈련 클래스에 의한 최적 상태열을 구하는 과정과 HMM을 구성하는 파라미터에 대한 경도 계산이 요구된다. 이러한 과정은 그 계산량이 많이 소요되어 연속어를 이용한 훈련에 적용하기에 아직 어려움이 있다. 따라서 훈련 데이터에 대한 인식 오류를 감소시키기 위하여 최적 상태열을 구하는 과정 및 경도 계산 과정이 생략되고 보다 감소된 HMM 파라미터를 훈련하여 음성 인식 성능 향상을 도모한다면 보다 효율적일 것이다. 따라서 본 논문에서는 음성 인식 단위 집합을 구성하는 각각의 HMM에 대응하는 가중치를 도입하고 HMM 확률 보정을 이용한 음성 인식 성능 향상을 도모하며 이러한 방법은 파라미터 감소와 최적 상태열의 계산 및 경도 계산의 미수행에 따른 계산량의 감소가 이루어지므로 효과적으로 음성 인식 성능 향상을 도모할 수 있다. 본 논문은 2장에서 먼저 분할 GPD 알고리즘에 기반한 최소 분류 오류 훈련 기법에 대하여 논의하고, 이러한 최소 분류 오류 훈련 방법의 단점을 보완하기 위하여 3장에서는 HMM 상태 확률 보정 기법을 고찰하고, 4장에서는 보다 효과적으로 훈련하기 위한 HMM 확률 보정 기법을 제안한다. 5장에서는 실험적 결과를 통하여 제안된 방법의 타당함 및 훈련 데이터 및 인식 데이터 인식률 상승의 일관성을 논의하며 6장에서는 제안된 방법과 실험적 결과들에 대한 분석을 통하여 결론을 도출한다.

II. 최소 분류 오류 훈련 기법

최소 분류 오류 훈련 기법의 목적은 데이터의 확률 분포를 추정하여 모델을 얻는 것이 아니고 최고의 인식 결과를 위해서 HMM의 관측 데이터를 구분하는 것이다. 본 장에서는 최소 분류 오류 알고리즘을 간략히 설명하고 에러 확률의 합리적인 추정치를 나타내는 최적화 범주에 대하여 논의한다[6-8]. 기본적인 HMM 기반의 음성 인식기에서 패턴 인식을 위해 클래스 i에 대한 분별 함수는 다음 식 (1)과 같이 정의된다.

$$g_i(\mathbf{X};\Lambda) = \log\{g_i(\mathbf{X},\bar{\mathbf{q}};\Lambda)\} = \sum_{t=1}^{T}\left[\log a^{(i)}_{\bar{q}_{t-1}\bar{q}_t} + \log b^{(i)}_{\bar{q}_t}(\mathbf{x}_t)\right] + \log \pi^{(i)}_{\bar{q}_0} \quad (1)$$

여기에서, $\bar{\mathbf{q}} = (\bar{q}_0, \bar{q}_1, \Lambda, \bar{q}_T)$는 클래스 i에 대해 조인트 상태열-관측열 확률 함수 $g_i(\mathbf{X}, \bar{\mathbf{q}}; \Lambda)$를 최대화하는 최적의 상태열이고, a_{ij}는 상태 i에서 상태 j로 천이하는 상태 천이 확률을 의미한다.
$b_j(\mathbf{x}_t)$는 상태 j에서 관측벡터 $\mathbf{x}_t$를 관측하는 확률 밀도 함수를 나타낸다. 연속적인 다변수 혼합성분 가우시안 HMM에서는 상태 출력 분포는 다음과 같다.

$$b_j(\mathbf{x}_t) = \sum_{m=1}^{M} c_{jm} N(\mathbf{x}_t; \mu_{jm}, \Sigma_{jm}) \quad (2)$$

여기에서, $N(\cdot)$는 다변수 가우시안 밀도 함수를 표기하며 μ_{jm}는 상태 j, 혼합성분 m에서의 평균 벡터이고 Σ_{jm}는 상태 j, 혼합성분 m에서의 상관행렬이다.

입력 음성에 대하여 클래스 결정 규칙이 사용되며 입력 음성 $\mathbf{X}$에 대한 클래스 $C(X)$는 다음과 같은 규칙에 의해 결정된다.

$$C(\mathbf{X}) = C_i \quad if\ i = \arg\max_j g_j(\mathbf{X};\Lambda) \quad (3)$$

여기에서, C_i는 분별함수에 의해 입력 음성 또는 관측벡터열에 대해 결정된 클래스를 의미한다.

먼저 연산적인 결정 규칙 (3)을 함수 형태로 표현하는 것이 필요하다. 패턴 인식기의 파라미터 세트 Λ의 연속적인 함수로서 결정 규칙을 함축하는 클래스 오인식 척도는 다음과 같이 정의된다.

$$d_i(\mathbf{X};\Lambda) = -g_i(\mathbf{X};\Lambda) + \log\left[\frac{1}{N}\sum_{j=1, j\neq i}^{N} \exp\left[g_j(\mathbf{X};\Lambda)\eta\right]\right]^{\frac{1}{\eta}} \quad (4)$$

여기에서, η는 양의 상수이고 N은 N-best 오인식 클래스들의 개수이다. 클래스 i에 해당하는 음성 $\mathbf{X}$에 대해, $d_i(\mathbf{X}) \geq 0$는 오인식을 의미하며 $d_i(\mathbf{X}) < 0$는 정확한 인식을 의미한다.

완전한 손실함수는 부드러운 이진 손실함수의 형태로서 오인식 척도에 관하여 정의된다.

$$l_i(\mathbf{X};\Lambda) = l(d(\mathbf{X})) \quad (5)$$

부드러운 이진 손실함수는 임의의 연속적인 이진 함수로 정의될 수 있으나 보통 다음과 같은 S자형의 (sigmoid) 함수가 사용된다.

$$l(d) = \frac{1}{1+\exp[-rd+\theta]} \quad (6)$$

여기에서, θ는 영 또는 영보다 다소 작은 값으로 설정되고 r은 상수값이다.

마지막으로 미지의 음성에 대하여 인식기 성능은 평균 손실함수로써 평가된다.

$$l(\mathbf{X};\Lambda) = \sum_{i=1}^{M} l_i(\mathbf{X};\Lambda) l(\mathbf{X} \in C_i) \quad (7)$$

여기에서, $l(\cdot)$는 지시 (indicator) 함수이다.

최적의 모델 파라미터는 평균 손실을 최소로 하는 모델 파라미터이며 평균 손실을 최소화하기 위해서 GPD 알고리즘이 주로 사용된다. GPD 알고리즘은 다음과 같이 주어진다[7,8].

$$\Lambda_{n+1} = \Lambda_n - \varepsilon_n \mathbf{U}_n \nabla l(\mathbf{X};\Lambda)|_{\Lambda=\Lambda_n} \quad (8)$$

여기에서, $\mathbf{U}$는 양으로 정의된 행렬, ε_n는 학습 비율 또는 조절의 단계 크리 (step size)이고 Λ_n는 시각 n에서 모델 파라미터 세트이다.GPD 알고리즘은 제한조건이 없는 최적화 기술이다. 그러나 확률 모델로서의 HMM 구조를 유지하기 위해서는 어떠한 제약 조건이 주어져야만 한다. 복잡한 제약조건을 갖는 GPD 알고리즘을 사용하는 대신에 GPD 알고리즘을 변환된 HMM 파라미터에 적용하였다. 파라미터 변환 과정은 변환된 공간에서 어떠한 제약조건이 없으며 원시 공간으로의 변환시 HMM 제약조건이 만족 되어야 한다. 다음과 같은 HMM 파라미터에 대한 제약조건이 원시 공간에서 유지되어야 한다.

$$\sum_j a_{ij} = 1 \text{ and } a_{ij} \geq 0,\ \sum_k c_{jk} = 1 \text{ and } c_{jk} \geq 0,\ \sigma_{jkl} \geq 0 \quad (9)$$

식 (9)와 같은 원시 공간에서의 파라미터 제약조건을 만족시키기 위해서 다음 식 (10)과 같은 파라미터 변환 과정이 파라미터 훈련 전후에 사용된다.

$$\begin{aligned} a_{ij} &\rightarrow \tilde{a}_{ij} \quad \text{where } a_{ij} = e^{\tilde{a}_{ij}} / \left(\textstyle\sum_k e^{\tilde{a}_{ik}}\right) \\ c_{jk} &\rightarrow \tilde{c}_{jk} \quad \text{where } c_{jk} = e^{\tilde{c}_{jk}} / \left(\textstyle\sum_k e^{\tilde{c}_{jk}}\right) \\ \mu_{jkl} &\rightarrow \tilde{\mu}_{jkl} = \mu_{jkl}/\sigma_{jkl} \\ \sigma_{jkl} &\rightarrow \tilde{\sigma}_{jkl} = \log\sigma_{jkl} \end{aligned} \quad (10)$$

III. 상태 확률 보정 기법

HMM 상태 출력 확률에 내재해 있는 구별되는 정보를 이용하기 위해서 각 상태는 가중치를 갖게 되며 각 상태 출력 스코어는 상태 출력 확률값과 상태 가중치의 곱으로 표현된다[10]. 기본적인 음성 인식 단위로써 M개의 HMM을 가지고 있고 개개의 인식 단위는 J개의 상태로 구성된다고 가정하자. 기본 인식 단위로는 보통 음소 또는 단어 모델이 사용된다.

전통적인 HMM에서는 클래스 i에 대한 분별 함수가 식 (1)에 의해 정의된다. 이와 유사하게 상태 가중 HMM에서 클래스 i에 대한 분별함수는 다음 식 (11)에 의해 정의된다.

$$g_i(\mathbf{X};\Lambda) = \sum_{t=1}^{T}\left[\log a^{(i)}_{\bar{q}_{t-1}\bar{q}_t} + w^{(i)}_{\bar{q}_t} \log b^{(i)}_{\bar{q}_t}(\mathbf{x}_t)\right] + \log\pi^{(i)}_{\bar{q}_0} \quad (11)$$

여기에서, $w^{(i)}_{\bar{q}_t}$는 클래스 i에서 시간 t에 대한 최적상태의 가중치이다.

초기값 1로 설정된 상태 가중치의 훈련은 앞에서 언급된 식 (4)~(8)과 동일한 과정을 통해 다음 식 (12)로 표현된다.

$$w^{(i)}_j(n+1) = w^{(i)}_j(n) - \varepsilon_n \frac{\partial}{\partial w^{(i)}_j} l(\mathbf{X};\Lambda)|_{w^{(i)}_j = w^{(i)}_j(n)} \quad (12)$$

식 (9)와 유사하게 상태 가중치에 대한 제약조건은 원시 공간에서 성립되어야 한다. 즉, 음성 인식 단위 내의 상태 가중치는 식 (13)에 의하여 제약되어야 한다.

$$\sum_{j=1}^{J} w_j = J,\ 0 < w_j < J \quad (13)$$

$$d_i(\mathbf{X};\Lambda) = -g_i(\mathbf{X};\Lambda) + \log\left[\frac{1}{N}\sum_{j=1, j\neq i}^{N} \exp\left[g_j(\mathbf{X};\Lambda)\eta\right]\right]^{\frac{1}{\eta}} \quad (18)$$

클래스 i에 대한 HMM 가중치 조절을 위해서는 클래스 i에 대한 HMM 가중치가 변화되어야 하는 비율을 설정해야 하며 이러한 비율을 차분 계수라고 정의하고 Δw_i로 표기한다. 클래스 i에 대한 HMM 가중치 조절을 위한 차분 계수는 오인식 척도와 훈련 음성의 클래스 i에 대한 분별 함수의 값을 이용하여 다음 식 (19)와 같이 정의한다.

$$\Delta w_i = \frac{d_i(\mathbf{X};\Lambda)}{-g_i(\mathbf{X};\Lambda)} \quad (19)$$

초기값 1로 설정된 클래스 i에 대한 HMM 가중치의 훈련은 클래스 오인식 척도를 감소시키기 위해 요구되는 클래스 i에 대한 HMM 가중치에 대한 차분 계수 Δw_i를 이용하여 다음과 같은 식 (20)을 이용하여 반복적으로 훈련한다.

$$\tilde{w}_i(n+1) = w_i(n) - \varepsilon_n \cdot w_i(n) \cdot \Delta w_i \quad (20)$$

HMM 집합에 대한 HMM 가중치의 합이 HMM 집합 내의 HMM 총 개수가 되는 HMM 가중치에 대한 제약 조건을 만족시키기 위해서 다음 식 (21)과 같이 파라미터 변환 과정을 HMM 가중치에 적용한다[7,8].

$$w_i \rightarrow \tilde{w}_i \quad where\ w_i = e^{\tilde{w}_i} / \left(\sum_k e^{\tilde{w}_k}\right) \quad (21)$$

여기에서, $\tilde{w}_i$는 원시 공간에서의 클래스 i에 대한 HMM 가중치 w_i에 대응하는 변환된 공간에서의 HMM 가중치이다.

4.3. 인식 알고리즘

연속 음성 인식을 위한 인식 알고리즘은 비터비 탐색 과정에서 각각의 HMM 가중치를 고려하여 계산함으로써 다음과 같이 구현될 수 있다.

$$V[0][j] = 0 \quad , j = \pi_0$$

$$V[0][j] = -\infty \ , j \neq \pi_0$$

$$V[t][j] = \max_h \left[V[t-1][h] + w(h)\cdot\{\log a_{hj}\}\right] + w(j)\cdot \log b_j(\mathbf{x}_t)$$

$$w(j) = w_k \ \ if\ j \in H_k,\ k = 1,2,\Lambda\ ,M \quad (22)$$

여기에서, $V[t][j]$는 시간 t에서 상태 j까지의 누적된 스코어이고, π_0는 초기 상태를 의미하며 H_k는 k번째 HMM을 의미하고 $\log b_j(\mathbf{x}_t)$는 상태 j에서 관측벡터 $\mathbf{x}_t$를 관측할 log 확률값을 의미하며 w_k는 k번째 HMM에 대한 HMM 가중치를 의미한다.

4.4. 임베디드 HMM 가중치 추정

훈련 음성에 대한 클래스 i가 K개의 HMM으로 구성되어 있다고 가정하자. N-best 탐색 과정에 의해 오인식 척도를 구하고 오인식 척도가 양 또는 영인 경우 오인식 척도가 감소하도록 훈련 클래스에 대한 스코어를 조절해 주기 위해서 차분 계수를 구하고 이를 훈련 클래스를 구성하는 개개의 HMM 가중치에 반영하여야 한다. 따라서 식 (23)과 같은 과정에 의해 훈련 클래스를 구성하는 HMM 가중치를 훈련한다.

$$\tilde{w}_k^{(i)}(n+1) = w_k^{(i)}(n) - \varepsilon_n \cdot w_k^{(i)}(n) \cdot \Delta w_i,\ k = 1,2,\Lambda\ ,K \quad (23)$$

여기에서, $w_k^{(i)}$는 클래스 i를 구성하는 k번째 HMM에 대한 가중치이고, Δw_i는 클래스 i에 대한 차분 계수이고, ε_n은 n번째 반복 훈련 과정에서의 학습 비율이다. 이와 같이 훈련된 HMM 가중치는 식 (24)에 의해 주어지는 HMM 가중치에 대한 제약 조건을 만족시키도록 파라미터 변환 과정에 의해서 조정한다.

$$w_k \rightarrow \tilde{w}_k \quad where\ w_k = e^{\tilde{w}_k} / \left(\sum_{x=1}^{M} e^{\tilde{w}_x}\right) \quad (24)$$

인식 오류를 최소화하기 위한 HMM 가중치 훈련을 위한 간단한 블락 다이어그램은 그림 1에 보여진다.

다음 장에서는 실험적 결과를 통해서 상태 확률 보정과 제안된 HMM 확률 보정을 이용한 음성 인식의 성능 향상을 평가한다.

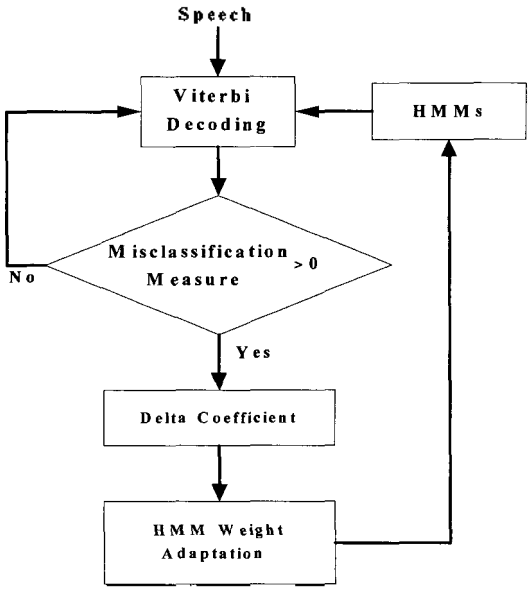

그림 1. HMM 가중치 추정의 블락도
Fig. 1. Block diagram of HMM weight estimation.

여기에서, w_j는 기본 인식 단위인 HMM 내의 상태 j에 대한 상태 가중치이고 J는 인식 단위 내의 상태의 총 개수이다. 끝으로 파라미터 변환은 다음 식에 의해 주어진다.

$$w_j \rightarrow \tilde{w}_j \quad where\ w_j = e^{\tilde{w}_j} / \left(\sum_k e^{\tilde{w}_k}\right) \tag{14}$$

여기에서, $\tilde{w}_j$는 w_j의 변환된 파라미터 공간에서의 상태 가중치이다. 일반적으로 최소 분류 오류 훈련에서는 오인식된 음성을 이용하여 파라미터 적응 훈련을 하므로 최소 분류 오류 훈련에 사용되는 즉, 오인식되는 훈련 데이터의 양이 너무 적다면 훈련 데이터에 대한 과적응 문제가 발생하게 되어 훈련 데이터와 인식 데이터 사이에 인식 성능의 차이가 발생하게 된다. 따라서 이러한 훈련 데이터와 인식 데이터의 인식 성능 차이를 극복하기 위한 방법이 요구되며 이를 위해서 부드러운 이진 손실 함수를 대신하여 대체 손실 함수를 고려한다. 이를 위해 오인식 척도를 구하는 과정에서 훈련 음성의 스트링 모델의 확률값에 가중치를 두어 오인식 척도에 더함으로써 오인식 척도를 확장하고 이를 확장된 선형 손실 함수로 선택할 수 있으며, 이러한 확장된 오인식 척도를 S자형의 함수에 대입함으로써 S자형의 이진 손실 함수를 선택할 수 있다. 훈련 음성의 스트링의 확률값에 가중치를 주어 더해 주는 오인식 척도에 대한 확장은 다음 아래의 식 (15)와 같이 정의된다[11].

$$\begin{aligned} d_i(\mathbf{X};\Lambda) &= -g_i(\mathbf{X};\Lambda) + \log\left[\frac{1}{N}\sum_{j=1, j\neq i}^{N} \exp\left[g_j(\mathbf{X};\Lambda)\eta\right]\right]^{\frac{1}{\eta}} \\ \tilde{d}_i(\mathbf{X};\Lambda) &= d_i(\mathbf{X};\Lambda) - k \cdot g_i(\mathbf{X};\Lambda) \\ &= -(1+k)\ g_i(\mathbf{X};\Lambda) + \log\left[\frac{1}{N}\sum_{j=1, j\neq i}^{N} \exp\left[g_j(\mathbf{X};\Lambda)\eta\right]\right]^{\frac{1}{\eta}} \end{aligned} \tag{15}$$

여기에서, k는 추가로 더해주는 확률값의 계수이고 훈련 데이터에 대한 인식 오류를 증가시켜 주고 훈련 음성의 스트링 모델에 대한 확률값의 HMM 파라미터에 대한 미분값을 증가시켜 주는 효과가 있다.

IV. HMM 확률 보정 기법

4.1. HMM 확률 보정

유사한 음성을 구분해 주기 위해서 HMM 출력 확률에 내재해 있는 구별되는 정보를 이용하기 위해서 각각의 HMM에 가중치를 주어 오인식 척도가 영이 되는 방향으로 반복하여 훈련함으로써 음성 인식 오류의 감소를 기대할 수 있다. 이를 위해서 HMM의 출력 스코어는 비터비 탐색 과정에 의한 HMM 출력 확률값과 HMM의 가중치의 곱으로 표현한다. 수식적인 설명을 위하여 단어 모델을 이용한 고립 단어 인식의 경우를 가정하여 기본적인 음성 인식 단위로서 M개의 HMM을 설정하고 각각의 인식 단위는 J개의 상태로 구성된다고 가정하자.

HMM 기반의 패턴 인식은 클래스 i에 대한 분별 함수를 이용한 클래스 결정 규칙에 의해 패턴 인식이 이루어지게 된다. 이와 같은 패턴 인식 과정에 있어서 클래스 i에 대한 분별 함수는 식 (1)에 의해 정의된다. 이와 유사하게 확장된 HMM 기반의 패턴 인식 과정에 있어서 클래스 i에 대한 분별 함수는 다음 식 (16)에 의해 정의된다.

$$\begin{aligned} g_i(\mathbf{X};\Lambda) &= w_i \cdot \left[\sum_{t=1}^{T}\left\{\log a^{(i)}_{\bar{q}_{t-1}\bar{q}_t} + \log b^{(i)}_{\bar{q}_t}(\mathbf{x}_t)\right\} + \log \pi^{(i)}_{\bar{q}_0}\right] \\ &= \sum_{t=1}^{T}\left\{w_i \cdot \log a^{(i)}_{\bar{q}_{t-1}\bar{q}_t} + w_i \cdot \log b^{(i)}_{\bar{q}_t}(\mathbf{x}_t)\right\} + w_i \cdot \log \pi^{(i)}_{\bar{q}_0} \end{aligned} \tag{16}$$

여기에서, w_i는 클래스 i에 대한 HMM의 가중치이다.

인식에 사용되는 HMM 집합에 대한 HMM 가중치의 합은 다음 식 (17)과 같이 HMM의 총 개수로 제한되어야 한다.

$$\sum_{i=1}^{M} w_i = M,\ 0 < w_i < M \tag{17}$$

HMM 가중치가 1로 설정되었을 때 N-best 스트링 모델을 기반으로 한 인식 알고리즘이 동일한 결과를 도출하도록 HMM 가중치에 대한 제한 조건을 주는 이유는 기존의 파라미터 추정 방법과 비터비 탐색 알고리즘에 의한 확률값의 큰 변화없이 용이하게 인식 과정을 수행하기 위함이다.

4.2. HMM 가중치 추정

보통 훈련 데이터에 대한 인식 오류의 개수를 감소시키기 위한 방법으로 N-best 스트링에 기반한 분할 GPD 알고리즘이 사용된다. 그러나 경도 계산에 의한 계산량 증가를 피하기 위하여 위하여 HMM 가중치에 대한 손실 함수의 미분값을 계산하는 대신에 우리는 훈련 데이터에 대한 오인식 척도를 감소시키기 위해서 요구되는 HMM 가중치의 변화량을 도입하였다. 이러한 HMM 가중치의 변화량은 간단한 나눗셈 연산을 통하여 매우 간단하게 구할 수 있다.

훈련 클래스 모델과 경쟁하는 N개의 클래스 모델과의 확률적 거리 척도는 다음 식 (18)과 같다.

V. 실험 결과 및 고찰

HMM의 상태 확률 보정과 HMM의 확률 보정을 통한 음성 인식의 성능 향상을 평가하기 위하여 단독 숫자음을 이용한 실험을 실시하였다. 단독 숫자음 DB는 조용한 사무실 환경에서 녹음되었으며 500명 (남성 250명, 여성 250명)의 음성으로 구성되었고, 400명 (남성 200명, 여성 200명)의 음성이 훈련에 사용되었고 100명 (남성 50명, 여성 50명)의 음성이 인식에 사용되었다. 본 실험에서는 11.025 KHz DSP 보드 상에서의 음성 인식기 구현을 위해서 16 KHz로 샘플링된 음성을 11.025 KHz로 다운 샘플링하였다. 특징 벡터로는 log 에너지를 포함하여 13차 켑스트럼 계수, 13차 1차 미분 계수, 13차 2차 미분 계수를 이용하여 총 3 스트림, 39차 특징 벡터를 구성하였다. 모든 상태 출력 확률 분포는 8 혼합성분 다변수 가우시안 분포를 사용하였다. 베이스라인 실험으로써 ML 훈련 방법에 의해서 3개의 상태, 3개의 스트림, 8개의 혼합성분으로 구성된 모델을 생성하였다. HMM 가중치를 훈련하는 과정에서 오인식 척도를 계산하기 위해 3개의 가장 경쟁적인 스트링을 사용하였다 ($N=3$).

5.1. 최소 분류 오류 훈련

본 실험에서는 4가지의 손실 함수 (오인식 척도에 기인한 이진 손실 함수, 확장된 오인식 척도에 기인한 이진 손실 함수, 오인식 척도, 확장된 오인식 척도)를 사용하여 HMM 기본 파라미터를 최소 분류 오류 훈련하여 인식 실험을 실시하였다. 표 1은 훈련 데이터 인식률이 최대가 되었을 경우 훈련 및 인식 데이터에 대한 단어 인식률을 나타낸 것이다. 훈련을 반복함에 따라 훈련 데이터 인식률은 점차적으로 증가하나 인식 데이터에 인식률은 거의 변화가 없음은 켑스트럼 계수에 대한 상태 확률 분포가 훈련 데이터에 과적응 되어가는 과정임을 의미한다.

5.2. 상태 확률 보정 기법

HMM 기반의 음성 인식기에서 HMM 상태 가중치를 훈련한 음향 모델링의 성능 평가를 위해서 앞에서 언급된 4가지 손실 함수를 사용하여 반복하여 훈련함에 따른 인식 성능의 변화를 비교 평가하였다. 훈련 과정에서 확률 모델로서의 HMM 구조를 유지하기 위해서 파라미터 변환 과정을 통해서 상태 가중치를 조절해 주었다. 표 2는 인식 데이터에 대한 단어 인식률이 최대가 되었을 경우, 훈련 데이터 및 인식 데이터에 대한 인식률을 나타낸다.

그림 2는 S자형의 손실 함수를 사용하여 반복하여 훈련함에 따른 훈련 및 인식 데이터의 인식률 변화를 보여준다. 훈련을 반복함에 따라 초반에는 훈련 데이터의 인식률이 증가하는 경향을 보이나 최소 분류 오류 훈련에 사용되는 훈련 데이터에 대한 과적응으로 인하여 인식률의 감소가 초래됨을 보여준다. 그림 3은 $k=0.005$인 경

표 1. 일반 HMM 인식 시스템에서의 손실함수에 따른 단어 인식률 (%)

Table 1. Word classification rate (%) for loss functions in conventional HMM recognition system.

Data	MLE	SIG	SIG+WL	LIN	LIN+WL
Training	98.40	98.58	98.58	98.58	98.71
Testing	98.00	98.00	98.00	98.00	98.00

SIG: sigmoid loss function
LIN: linear loss function
WL: weighted likelihood of correct class

표 2. 상태 가중치를 이용한 HMM 인식 시스템에서의 손실함수에 따른 단어 인식률 (%)

Table 2. Word classification rate (%) for loss functions in state-weighted HMM recognition system.

Data	MLE	SIG	SIG+WL	LIN	LIN+WL
Training	98.40	98.54	98.54	98.44	98.56
Tessting	98.00	98.17	98.58	98.00	98.50

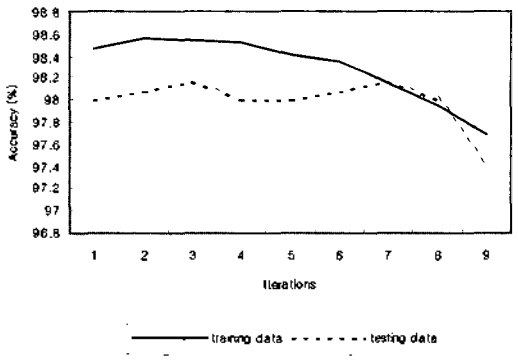

그림 2. 상태 가중치 훈련 (SIG)
Fig. 2. State weight training (SIG).

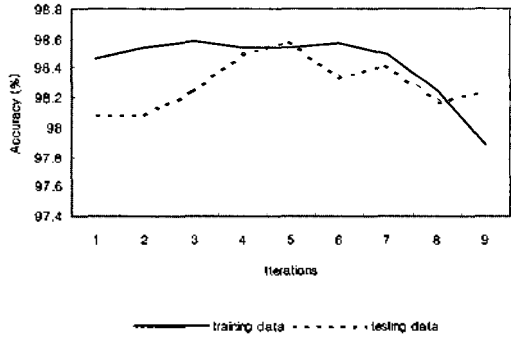

그림 3. 상태 가중치 훈련 (SIG+WL)
Fig. 3. State weight training (SIG+WL).

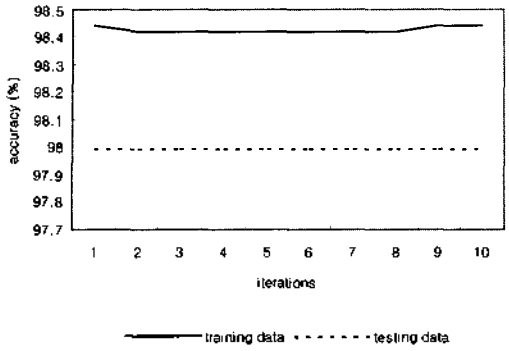

그림 4. 상태 가중치 훈련 (LIN)
Fig. 4. State weight training (LIN).

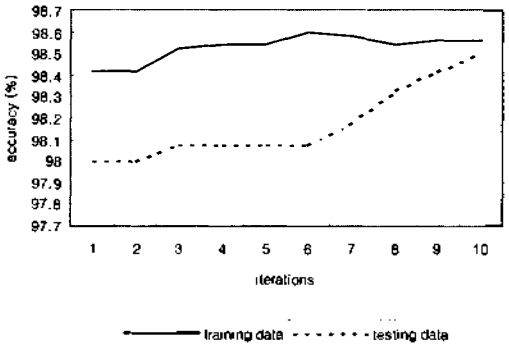

그림 5. 상태 가중치 훈련 (LIN+WL)
Fig. 5. State weight training (LIN+WL).

우에 훈련 음성에 대한 훈련 음소열의 확률값에 가중치를 두어 확장된 오인식 척도를 구하고 이로부터 S자형의 손실 함수를 구하여 훈련함에 따른 인식률의 변화를 보여준다. 반복함에 따라 훈련 데이터의 인식률 감소와 더불어 인식 데이터 인식률이 진동함을 볼 수 있다.

그림 4는 $k=0$인 경우의 선형 손실 함수를 사용하여 상태 가중치를 훈련한 결과는 훈련 데이터 및 인식 데이터에 대해서 일관성 있는 인식 성능 향상을 보여주지 못함을 나타낸다. 그림 5는 $k=0.005$인 경우의 훈련 음성에 대한 훈련 음소열의 확률값에 가중치를 주어 얻어진 확장된 선형 손실 함수를 사용하여 훈련한 결과는 훈련 데이터는 물론 인식 데이터에 대하여도 일관된 인식 성능 향상을 보여준다. 5.2.1 절에서는 k값의 변화에 따른 인식률의 변화에 대하여 논의한다.

5.2.1. 확장된 선형 손실함수에서 k의 변화에 따른 인식률 변화 (Weighted linear loss function)

5.2절에서 훈련 음성에 대한 훈련 음소열 모델의 확률값에 가중치를 주어 얻어진 확장된 선형 손실 함수를 사용함으로써 훈련 데이터와 인식 데이터에 대한 일관된 인식률의 향상을 논의하였다. 따라서 이번 절에서는 선형 손실 함수에서 훈련 음소열 모델의 확률값에 곱하는 가중치에 따른 훈련 데이터와 인식 데이터의 인식률 변화를 실험하였고, 인식 데이터에 대한 인식률이 최대가 되었을 경우의 인식 결과는 다음 표 3과 같다.

표 3. 훈련 음소열 모델의 가중치 k값에 따른 단어 인식률 (%)
Table 3. Word classification rate (%) for different weighting factor k of correct string model.

k	0.001	0.002	0.003	0.004	0.005
Training	98.54	98.63	98.48	98.67	98.56
Tesstng	98.17	98.25	98.25	98.25	98.50

훈련 음성에 대한 훈련 음소열 모델의 확률값에 가중치 k를 주어 오인식 척도에 더한 확장된 오인식 척도를 사용하는 것은 훈련 데이터에 대한 인식 오류를 증가시켜 주는 효과를 가져오며, 아울러 상태 가중치에 대한 선형 손실 함수의 미분값을 $(1+k)$배만큼 증가시켜 주는 효과가 있다. 따라서 k값을 적절히 조절해 줌으로써 훈련 데이터와 인식 데이터 인식률이 일관되게 향상되는 결과를 얻을 수가 있다.

5.3. HMM 확률 보정 기법

다음 그림 6은 4장에서 제안한 방법에 따라 훈련 데이터를 이용하여 15번 반복하여 훈련함에 따른 훈련 데이터와 인식 데이터의 변화 과정을 보여준다. 최소 분류 오류 훈련 방법의 원리와 동일하게 반복 횟수가 증가함에 따라 훈련 데이터와 인식 데이터에 대한 인식률이 일관성 있게 상승됨을 보여준다. 그림 1에서 훈련 데이터 인식률이 인식 데이터에 대한 인식률보다 높게 나타나는 경우는 훈련

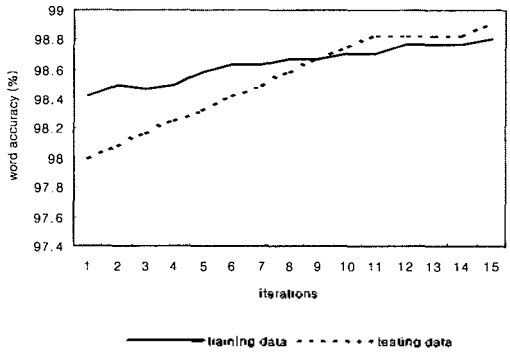

그림 6. 확률 보정 HMM 인식 시스템에서 반복 훈련에 따른 단어 인식률 (%)
Fig. 6. Word classification rate (%) versus the number of iterations in HMM-weighted HMM recognition system.

표 4. HMM 음성 인식 시스템에서 훈련 방법에 따른 단어 인식률 (%)
Table 4. Word classification rates (%) for training methods in HMM recognition system.

Data	ML	MCE	State Weighting	HMM Weighting
Training	98.40	98.71	98.56	98.81
Testing	98.00	98.00	98.50	98.92

데이터에 부정확한 발음이 포함된 데에 기인하며 이러한 음성은 HMM 상태 가중치 훈련에 사용하지 않았다. HMM 확률 보정 기법은 상태 확률 보정 기법에 비하여 훈련 데이터에 대한 과적응에 덜 민감한 특성을 보이며 오인식 척도 자체를 차분 계수를 이용하여 HMM 가중치에 직접 반영함으로써 가중치 훈련을 보다 효과적으로 할 수 있는 장점이 있다.

5.4. 훈련 방법에 따른 단어 인식률

본 장에서의 실험은 훈련 데이터에 대한 오인식률을 최소화함으로써 궁극적으로 음성 인식 성능 향상을 도모하는 실험으로써 HMM 기본 파라미터의 최소 분류 오류 훈련, 상태 확률 보정 기법, HMM 확률 보정 기법의 세가지 실험을 실시하였고 인식 데이터 인식률이 최고일 경우 훈련 및 인식 데이터 인식률을 표 4에 나타내었다. 표 4에서 볼 수 있듯이 HMM 상태 보정 기법은 최적 상태열 및 경도 계산을 수행하지 않고도 가장 높은 인식 성능 향상을 보여준다.

VI. 결 론

5장에서의 실험적 결과들은 HMM 기반의 음성 인식 시스템에서 훈련 음성에 대한 훈련 음소열의 확률값에 가중치를 주어 얻어진 확장된 선형 손실 함수를 사용하여 HMM 상태 가중치를 훈련함으로써 훈련 및 음성 데이터에 대한 인식 성능이 일관성 있게 증가함을 보여준다. 또한, 제안된 방법인 HMM 가중치를 차분 계수값을 이용하여 반복적으로 훈련함으로써 기본 파라미터의 최소 분류 오류 훈련 방법 및 상태 확률 보정 기법에 비해 더욱 우수한 음성 인식 성능 향상을 얻을 수 있었다. 음성 인식기의 궁극적인 목적인 인식 오류의 최소화의 관점에서 이러한 결과는 매우 바람직한 결과이며 HMM 확률 보정 기법을 사용함으로써 최적 상태열과 경도 계산을 하지 않고도 효과적으로 음성 인식 성능 향상을 도모할 수 있었다. 실험 결과에서 HMM 확률 보정 기법은 인식 데이터에 대해서 28% 오인식률을 감소시킬 수 있었으며 HMM 확률 보정 기법은 음성을 구별하여 인식기를 최적화하는데 매우 효과적임을 볼 수 있었다.

참 고 문 헌

1. L. Rabiner, and B. H. Juang, *Fundamentals of Speech Recognition*, Prentice Hall, 1993.
2. X. Huang, A. Acero, and H.-W. Hon, *Spoken Language Processing*, Prentice Hall, 2000.
3. A. H. He, S. Kwong, K. F. Man, and K. S. Tang, "An improved maximum model distance approach for HMM-based speech recognition systems," *Pattern Recognition*, **33**, 1749-1758, 2000.
4. L. R. Bahl, P. F. Brown, P. V. de Souza, and R. L. Mercer, "Estimating hidden markov model parameters so as to maximize speech recognition accuracy," *IEEE Transactions on Speech and Audio Processing*, **1** (1), 77-83, January 1993.
5. B. H. Juang, and S. Katagiri, "Discriminative learning for minimum error classification," *IEEE Transactions on Signal Processing*, **40** (12), 3043-3054, December 1992.
6. W. Chou, C. H. Lee, and B. H. Juang, "Minimum error rate training based on N-best string models," *IEEE ICASSP-93*, 652-655, April 1993.
7. B. H. Juang, W. Chou, and C. H. Lee, "Minimum classification error rate methods for speech recognition," *IEEE Transactions on Speech and Audio Processing*, **5** (3), 257-265, May 1997.
8. W. Chou, "Discriminant-function-based minimum recognition error rate pattern-recognition approach to speech recognition," *Proceedings of the IEEE*, **88** (8), 1201-1223, August 2000.
9. M. T. Lin, A. Spanias, and P. Loizou, "An improved approach to robust speech recognition using minimum error classification," *Speech Communication*, **30**, 27-36, 2000.
10. O. W. Kwon, C. K. Un, "Performance of HMM-based speech recognizers with discriminative state-weights," *Speech Communication*, **19**, 197-205, 1996.
11. D. W. Purnell and E. C. Botha, "Improved generalization of MCE parameter estimation with application to speech recognition," *IEEE Transactions on Speech and Audio Processing*, **10** (4), 232-239, May 2002.

저자 약력

● 권 태 희 (Tae-Hee Kwon)
2001년 2월: 고려대학교 전기공학 (학사)
2001 년 3월~현재: 고려대학교 전자공학 (석사)
※ 주관심분야: 신호처리

● 고 한 석 (Han-Seok Ko)
1982년 5월: 카네기 멜론 대학교 전기공학 (학사)
1986년 5월: 메릴랜드 대학교 시스템공학 (석사)
1988년 5월: 존스 홉킨스 대학교 전기공학 (석사)
1992년 5월: 미국 카톨릭 대학교 전기공학 (박사)
1995년 3월~현재: 고려대학교 전자공학과 교수
※ 주관심분야: 신호처리