

MIN 모듈을 갖는 준연속 Hidden Markov Model

Semi-Continuous Hidden Markov Model with the MIN Module

김 대 극* · 이 정 주** · 정 호 균** · 이 상 희**

Dae-keuk Kim · Jeong-ju Lee · Ho-Kyoun Jeong · Sang-hee Lee

ABSTRACT

In this paper, we propose the HMM with the MIN module. Because initial and re-estimated variance vectors are important elements for performance in HMM recognition systems, we propose a method which compensates for the mismatched statistical feature of training and test data. The MIN module function is a differentiable function similar to the sigmoid function. Unlike a continuous density function, it does not include variance vectors of the data set. The proposed hybrid HMM/MIN module is a unified network in which the observation probability in the HMM is replaced by the MIN module neural network. The parameters in the unified network are re-estimated by the gradient descent method for the Maximum Likelihood (ML) criterion. In estimating parameters, the variance vector is not estimated because there is no variance element in the MIN module function.

The experiment was performed to compare the performance of the proposed HMM and the conventional HMM. The experiment measured an isolated number for speaker independent recognition.

Keywords : Hidden Markov Model, MIN module, Maximum Likelihood

1. 서 론

음성은 인간이 사용하고 있는 정보전달 매체 중에 가장 자연스러운 형태이다. 이에 음성을 인간과 기계간의 통신수단으로 이용하려는 음성인식 기술이 디지털 신호처리, 반도체 등의 제반 기술 발전으로 인해 지난 수십 년 간 빠르게 발전해 왔으며 이미 특정 분야에서 제한적으로 상용화되고 있다. 그러나 보다 다양한 분야에서의 이상적인 음성인식을 위해서는 음성에 관련된 폭넓은 분야의 연구와 높은 인식성능을 제공하는 음성인식 시스템의 개발이 요구된다.

음성인식을 위해서 사용되는 기본적인 인식대상 및 인식방법은 그 분류 기준에 따라

* 한림정보산업대학 전자통신과

** 강원대학교 전자공학과

여러 가지 종류로 나눌 수 있다. 그러나 일반적으로 이상적인 음성인식을 위해서는 연속 음성 및 화자독립 그리고 대용량의 어휘에 대한 음성인식 시스템의 구현이 기반되어야 한다. 음성인식 방법에 따른 분류는 음소나 단어에 대해 각각의 템플릿을 만들어서 인식하고자 하는 음성과의 비교를 통해 인식하는 템플릿에 의한 접근(template based approach) 방법과 음향-음성학(acoustic-phonetics)적 지식에 의해 음성을 분할 인식하는 지식에 근거한 접근(knowledge based approach)방법이 있다. 또한 음성이 나타날 수 있는 상태의 순서를 확률적으로 추정하여 인식하는 확률적인 접근(stochastic approach) 방법과 인간의 신경망을 수학적으로 모델링한 인공 신경회로망(artificial neural network)을 이용한 방법 등이 있다. 현재, 음성인식 방법의 동향은 통계적 방법의 모델이 주류를 이루고 있다. 특히, HMM(hidden markov model)은 확률적인 접근방법의 하나로 통계적 특성 및 시간적 구조에 대한 모델링의 우수성 그리고 연속음성인식으로 용이한 확장성을 가지고 있다. 따라서, 음성신호와 같이 시간적 변이 특성을 갖는 패턴들에 대해 좋은 성능을 보임으로서 현재 가장 널리 사용되는 방법 중 하나이다[1-3][11].

HMM에 대한 연구는 파라미터 개선에 관한 연구뿐만 아니라 각 모델 간의 변별력을 높이기 위한 연구가 진행되고 있다. HMM을 사용하면 동적 패턴의 특성을 잘 모델화할 수 있는 반면에 각 패턴간의 변별력이 다소 떨어지는 단점이 있다. 따라서, 최근에는 이를 대체하거나 보조할 수 있는 방법으로 신경회로망을 이용하려는 연구가 시도되고 있다. 가장 단순한 방법으로는 신경회로망이 정적인 패턴분류에 적합하다는 특성을 이용하여 MLP(multi-layer perceptron)[5] 및 RBFN(radial basis function network)[6] 등을 그대로 사용하는 것이다. 다른 접근방법으로는 동적 패턴을 처리하는데 적합한 새로운 신경회로망 모델을 개발하여 시간적인 문맥 정보를 사용하는 TDNN(time delay neural network)과 피드백(feedback) 연결이 가능한 케환 신경회로망(recurrent neural network) 등을 이용하려는 연구도 있다. 하지만, 아직까지는 신경회로망의 적합한 분야는 정적패턴 분류이기 때문에 HMM과 신경회로망을 결합하여 각각의 장점을 취하고자 하는 연구가 매우 활발히 진행되고 있다. 또한 HMM을 구성하는 모델 파라미터의 학습을 관측열에 대한 해당 모델의 확률을 최대로 하는 ML(maximum likelihood) 추정 방법[2]에서 다른 모델들을 함께 고려하여 해당 모델의 출력 확률을 최대로 하는 방향으로 연구가 진행되고 있다. 특히, 상호정보를 최대로 하는 MMI(maximum mutual information) 및 MDI(maximum discrimination information) 방법, 그리고 학습 데이터와 해당 모델을 실질적인 인식을 통하여 분류의 오류를 감소시키는 MCE(minimum classification error), GPD(gradient probability descent), CT(corrective training), MSE(mean square error) 방법 등의 새로운 기준함수 도입에 관한 연구가 진행되고 있다.

본 논문은 HMM과 신경회로망의 결합에 관한 것으로 HMM에서의 양자화코드의 빈도수 및 혼합 밀도 함수를 이용한 출력 확률 추정을 정적인 패턴 분류에 강한 특성을 갖는 신경회로망을 이용하여 성능을 향상시키기 위한 것이다. 이러한 신경회로망으로는 양자화 과정에서 왜곡을 최소화하는 코드북 대신에 오분류의 수를 최소화하는 LVQ 알고리즘을 망의 형태로 제안한 GLVQ(generalized LVQ) 네트워크[4][7]에서의 MIN 모듈을 이용한다. 이러한 MIN 모듈은 주어진 관측열에 대한 출력 확률을 추정하기 위하여 사용되며

HMM의 하나의 구성요소로서 결합한다. 제안된 MIN 모듈을 갖는 HMM의 파라미터 추정은 ML 기준함수를 사용하여 기울기 경사법에 의해 추정하였다. 제안된 MIN 모듈을 갖는 HMM의 인식성능 비교를 위하여 MIN 모듈함수의 특성 및 기존 HMM 모델들과 한국어 단독 숫자에 대한 인식성능 실험을 하였다.

본 논문의 구성은 다음과 같다. 2장은 HMM에 대하여 기술하였으며, 3장에서 HMM의 출력 관측 확률 함수를 정적 패턴들의 변별 분류 능력이 우수한 MIN 모듈을 이용한 HMM을 제안한다. 4장에서 제안된 HMM의 성능평가를 위하여 한국어 단독 숫자에 대하여 인식실험 및 결과를 제시하며 5장에서 본 논문의 결론에 대하여 기술한다.

2. Hidden Markov Model

HMM은 통계적 과정을 갖는 유한상태 구조로 시간에 따른 상태간의 변화를 결정하는 천이확률 그리고 각 상태에서의 출력 확률로 구성되며, 상태에서 출력 확률을 어떻게 추정하는가에 따라 이산(discrete)분포, 연속(continuous)분포, 준연속(semi-continuous)분포 HMM으로 분류된다[3][8]. 이산 HMM은 상태에서의 관측 출력 확률이 이산적인 특성 값으로 추정되며, 연속 HMM은 주어진 음성신호에 대한 양자화 과정 없이 실수 값의 특징 벡터를 그대로 사용할 수 있도록 연속 확률 밀도 함수를 사용한다. 또한, 준연속 HMM은 이산 HMM의 확률값과 연속 HMM의 혼합 밀도 함수를 혼합한 형태로서 두 모델의 장점을 취한 형태로 볼 수 있다.

이산 HMM에서의 관측열은 코드북으로부터 주어진 특징벡터에 대응하는 최적의 코드워드 색인을 얻기 위해 주어진 특징벡터와 코드워드 벡터간의 유클리디언(Euclidian) 거리 척도를 이용하여 가장 작은 거리 값을 갖는 코드워드의 색인을 이산 HMM에서의 입력된 음성신호의 관측열로 사용한다. 상태에서의 출력 확률은 식(1)과 같이 표시할 수 있다.

$$b_j(O_t) = b_j(k) = P(O_t = v_k | s_t = j), \quad v_k = \min_{v_i} d(O_t, v_i) \quad (1)$$

여기서, $1 \leq i \leq K$, K = codebook 크기, v_i = codeword

연속 HMM[1][9]에서는 벡터 양자화 오차를 최소화하기 위하여 출력 확률 분포를 연속 분포로 취급하며 관측 벡터의 분포가 $B = \{b_j(x) | 1 \leq j \leq N\}$ 의 형태를 갖는다. 이때 x 는 음성신호에서 추출한 특징 벡터 자체를 나타내며 $b_j(x)$ 는 관측 벡터가 x 와 $x + dx$ 사이에 있을 확률을 나타낸다. 천이 확률과 출력 확률을 추정하기 위해 $b_j(x)$ 의 확률 밀도 함수를 Gaussian, Poisson, Gamma, Cauchy 등의 여러 가지 확률 밀도 함수로 사용할 수 있다. 이 중에서 계산상의 편의성 때문에 가우시안 밀도 함수가 가장 널리 사용되고 있다[10]. 특히, 여러 개의 가우시안 밀도에 가중치를 주고 합한 가우시안 혼합 밀도(gaussian mixture density)는 어떠한 연속 확률 밀도로 근사화할 수 있다는 강점을 지니고 있기 때문에 다양한 변이성을 모델링하기에 적합하다. 가우시안 혼합밀도는 식(2)와 같이 표현할 수 있다.

$$b_j(x) = \sum_{k=1}^M c_{jk} b_{jk}(x) = \sum_{k=1}^M c_{jk} N(x, \mu_{jk}, \Sigma_{jk}) \quad (2)$$

여기서, $N(x, \mu, \Sigma)$ 는 평균벡터 μ 와 분산벡터 Σ 를 가지는 가우시안 밀도 함수이다. 또한, M 개로 구성되어 있는 혼합 밀도의 k 번째 함수의 가중치 c_{jk} 는 식(3)을 만족해야 한다.

$$\sum_{k=1}^M c_{jk} = 1, \quad c_{jk} \geq 0, \quad 1 \leq j \leq N, \quad 1 \leq k \leq M \quad (3)$$

여기서, M 은 mixture 수이고 N 은 상태의 수이다. 가우시안 혼합 밀도는 식(4)와 같은 제한 조건을 갖는다.

$$\int_{-\infty}^{\infty} b_j(x) dx = 1, \quad 1 \leq j \leq N \quad (4)$$

준연속 HMM은 이산 HMM의 관측 확률값과 연속 HMM의 가우시안 밀도 함수를 혼합한 형태이다[1]. 준연속 HMM은 모든 모델과 상태에서 공유되는 L 개의 가우시안 확률 밀도들과 각 상태에서 가우시안 확률밀도들의 가중치를 결정하는 혼합밀도계수(mixture density coefficient)에 의해 입력 음성의 특징을 확률적으로 모델링하는 혼합 확률을 구한다. 준연속 HMM에서의 혼합 확률은 코드북의 코드워드가 연속 확률 밀도 함수로 표현된다면, 상태 $s_t = j$ 에서 특징벡터 x 가 갖는 분포는 식(5)로부터 계산된다.

$$b_j(x) = \sum_{k=1}^L P(v_k | s_t = j) \cdot N(x, \mu_{jk}, \Sigma_{jk}) \quad (5)$$

식(5)에서 가우시안 밀도 함수 $N(x, \mu_{jk}, \Sigma_{jk})$ 와 Markov 상태 s_t 가 서로 독립이라면 식(6)과 같게 된다.

$$b_j(x) = \sum_{k=1}^L b_j(k) \cdot N(x, \mu_k, \Sigma_k) = \sum_{k=1}^L c_{jk} \cdot N(x, \mu_k, \Sigma_k) \quad (6)$$

여기서, $b_j(k)$ 는 상태 j 에서 k 번째 코드워드가 관측될 이산확률(또는 혼합 밀도 계수)을 나타내며, L 은 코드북의 크기, μ_k 와 Σ_k 는 각각 k 번째 코드워드의 평균벡터와 분산벡터를 나타낸다. 준연속 HMM은 이산 HMM의 관점에서 VQ로 인해 발생하는 관측 벡터의 왜곡을 여러 개의 코드워드로 하나의 관측 벡터를 모델링함으로써 줄일 수 있고, 연속 HMM의 관점에서는 코드워드를 서로 공유함으로써 파라미터 수를 감소하여 계산량을 감소시킬 수 있다. 따라서, 제한된 학습 데이터와 계산량에 있어서 준연속 HMM은 다른 두

가지 모델링 방법의 장점을 취할 수 있어 HMM을 이용한 음성인식에서 연속 HMM과 더불어 많이 사용되고 있다.

3. MIN 모듈을 갖는 Hidden Markov Model

연속분포를 이용한 기존의 HMM에서 학습 및 테스트 데이터에 대한 통계적 특성에 차이가 있을 때 가우시안 밀도 함수에서의 분산 추정치 및 HMM에서의 초기 파라미터들은 인식시스템 성능에 매우 민감한 요소로 작용하며 이에 관한 연구들이 진행되어 왔다 [11]. 본 논문에서는 기존의 가우시안 밀도 함수를 신경회로망에서의 거리 기반 함수로 대체하여 사용한다. 이에 신경회로망의 MIN 모듈 함수를 포함한 MIN 모듈을 연속분포 관측확률을 구하는 하나의 구성요소로서 HMM과 결합하여 신경회로망의 망 형태로 재구성하였다. 또한 기울기 경사법을 이용하여 HMM과 MIN 모듈에서의 각 파라미터를 전역적으로 최적화하는 방법을 제안한다.

3.1 MIN 모듈을 갖는 HMM (Hybrid HMM/MIN module)의 구조

기존 준연속 HMM의 경우 관측 출력 확률은 식(6)과 같이 이산 HMM의 확률값(또는 혼합 밀도 계수)과 연속 HMM의 가우시안 밀도 함수가 혼합된 형태로 표현되며, 학습 데이터를 이용하여 평균벡터와 분산벡터에 의해 초기 모델링된다. 그러나 통계적 특성에 차이가 있는 두 데이터 집단에서, 학습 데이터를 가우시안 밀도 함수를 이용하여 추정된 파라미터(평균벡터와 분산벡터)는 두 데이터 집단에 대한 정확한 모델링이 될 수 없으며, 옳지 못한 가우시안 확률밀도가 혼합 확률을 구하는데 사용 될 수 있다.

제안된 MIN 모듈을 갖는 HMM은 기존의 연속 HMM에서 출력 확률을 구하기 위해 사용되는 가우시안 밀도 함수를 신경회로망의 거리 기반의 MIN 모듈 함수를 대체하여 통계적 불일치에 따른 평균 및 분산의 문제점을 해소하고자 한다. 이에 오분류의 수를 최소화하는 VQ를 생성하기 위하여 설계된 GLVQ망에서의 MIN 모듈 및 MIN 모듈 함수를 간략화 하여 HMM에서의 구성요소로서 결합하였으며, 단일화된 망 형태로 그림 1과 같다. MIN 모듈은 입력노드 중 최소값을 가지는 노드를 찾아 최대의 값을 갖기 위한 것으로 입력노드와 MIN 모듈함수 그리고 출력노드로 구성된다. MIN 모듈에서 k 번째 입력(코드워드) 노드 I_k 의 입력값은 식(7)과 같이 주어진 관측열 x 와 평균벡터(코드워드) 사이의 유클리디언 거리를 이용하여 계산한다.

$$I_k = \|x - u_k\| = \sum_{d=1}^D \frac{(x_d - u_{dk})^2}{2} \quad (7)$$

여기서, D 는 관측열의 차원(dimension)수를 나타낸다. 이와 같이 계산된 MIN 모듈의 입력노드 값은 식(8)과 같은 MIN 모듈 함수를 이용하여 관측 출력값을 출력노드로 전달하게 된다.

$$M(k) = \prod_{u, u \neq k} \frac{1}{1 + e^{-\frac{(I_u - I_k)}{\sigma}}}, \quad u = 1, 2, \dots, L \quad (8)$$

$$\cong 2 \cdot \prod_{u=1}^L F(u, k) \quad (9)$$

여기서, $F(u, k) = \frac{1}{1 + e^{-\frac{(I_u - I_k)}{\sigma}}}$ 이다.

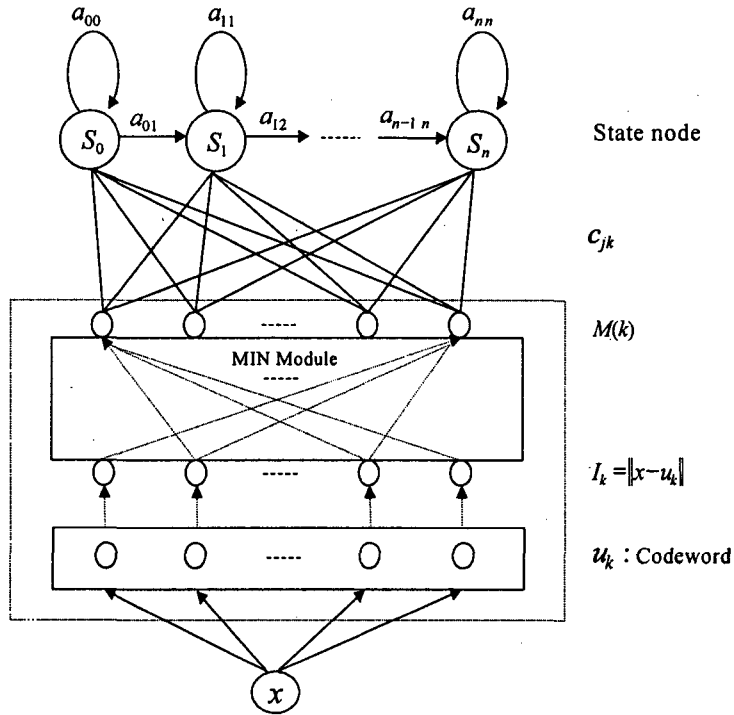


그림 1. MIN 모듈을 갖는 HMM의 구조

식(8)에서 I_k 는 MIN 모듈의 k 번째 입력노드 값이며, σ 는 시그모이드 함수의 활성화영역의 폭을 결정하는 상수이다. 따라서 $\sigma=0$ 인 경우, 입력 노드들 중에서 n 번째 노드의 값이 가장 작다면 $M(n)=1$ 이 되고, 다른 $M(m)$ 는 모두 0이 된다. 또한 $\sigma > 0$ 이면 $M(n)$ 는 1보다 작은 값을 가지며 나머지 $M(m)$ 는 0보다 큰 값을 갖게 된다. 제한된 망에서의 σ 는 식(7)로부터 얻어진 정량화된 코드워드 노드의 입력값들을 이용하여 식(10)과 같이 정의하여 사용한다.

$$\sigma = \sqrt{\frac{1}{L} \sum_{i=1}^L I_i}, \quad I_i = \|x - u_i\| \quad (10)$$

여기서, L 은 코드워드(평균벡터) 노드 수를 나타낸다. HMM에서의 관측 출력 값은

MIN 모듈의 혼합밀도 계수의 적의 합으로 식(11)과 같이 계산된다.

$$b_j(x) = \sum_{k=1}^L c_{jk} \cdot M(k) \quad (11)$$

$$\int_{-\infty}^{\infty} b_j(x) dx = 1, \quad \sum_{k=1}^L c_{jk} = 1, \quad 1 \leq j \leq N \quad (12)$$

식(11)에서 $M(k)$ 는 관측벡터 x 와 k 번째 코드워드의 평균벡터 u_k 를 갖는 MIN 모듈 함수이다. MIN 모듈의 출력값은 HMM이 식(12)와 같은 제한성을 가지고 있어 기존의 MIN 모듈에서의 정규화 과정이 포함되지 않는다. 이것은 단일화된 망을 기울기 경사법에 근거하여 추정하게 될 평균벡터의 추정식을 기존의 것보다 간략화하게 하는 요소이다.

3.2 파라미터 추정

HMM을 이용한 인식시스템에서 모델의 성능은 관측열 O 가 주어질 때 $P(O | \lambda)$ 을 최대로 하는 파라미터 $\lambda = (\Pi, A, B)$ 을 추정하는 것에 의해 결정된다. 파라미터 추정은 HMM의 기본적인 세 가지 문제 중 가장 어려운 것으로 현재까지 이를 위한 분석적인 해결책이 알려지지 않아 이러한 해결책으로 파라미터를 최적화하기 위하여 반복적인 알고리즘이나 기울기 강하법(*gradient descent method*)을 주로 사용하고 있다.

기울기 강하법을 이용한 HMM 모델 λ 의 파라미터 Θ 는 일반적으로 식(13)과 같이 J 함수를 최소화하는 방향으로 반복적으로 각 파라미터를 추정할 수 있다.

$$\bar{\Theta} = \Theta - \eta \cdot \left[\frac{\partial J}{\partial \Theta} \right] \quad (13)$$

여기서, J 는 기울기 강하법에 의해 최소화할 기준함수이다. HMM에서 기울기 강하법에 의해 사용될 기준함수는 ML(Maximum Likelihood) 함수를 변형하여 식(14)와 같이 정의할 수 있다. 즉, 기준함수 $J = E_{ML}$ 의 최소화는 $L_{tot} = P(O | \lambda)$ 의 최대화와 같으므로 기울기 강하법을 이용한 파라미터 추정의 규칙으로 사용할 수 있다.

$$J = E_{ML} = -\log P(O | \lambda) = -\log(L_{tot}) \quad (14)$$

따라서, 주어진 모델의 임의의 파라미터 Θ 의 충분치는 식(15)와 같으며 우변 항을 *chain rule*를 적용하여 유도할 수 있다.

$$\frac{\partial J}{\partial \Theta} = -\frac{1}{L_{tot}} \frac{\partial L_{tot}}{\partial \Theta} \quad (15)$$

여기서, $L_{tot} = \sum_{i=1}^N P(O, s_t = i | \lambda) = \sum_{i=1}^N \alpha_i(i) \beta_i(i)$ 이다.

식(15)를 이용한 제안된 MIN 모듈을 갖는 HMM에서의 각 파라미터 추정값은 기존의 HMM과 유사하며 제안된 HMM 망에서는 분산벡터가 고려되지 않은 거리 기반 함수인 MIN 모듈 함수를 이용함으로써 분산벡터에 대한 추정치는 존재하지 않는다. ML을 기준 함수로 사용하였을 때의 각 파라미터에 대한 추정값은 다음과 같다.

$$(1) \text{ 초기 확률 } \pi_i : \frac{\partial L_{tot}}{\partial \pi_i} = b_i(x_1) \beta_1(i) \quad (16)$$

$$(2) \text{ 천이 확률 } a_{ij} : \frac{\partial L_{tot}}{\partial a_{ij}} = \sum_{t=1}^T \beta_i(j) b_j(x_t) \alpha_{t-1}(i) \quad (17)$$

$$(3) \text{ 혼합밀도계수 } c_{jk} : \frac{\partial L_{tot}}{\partial c_{jk}} = \sum_{t=1}^T \frac{\alpha_t(j) \beta_t(j)}{b_j(x_t)} M(k) \quad (18)$$

(4) 코드워드 u_k : 코드워드에 대한 파라미터 증분은 식(19)에서 각각의 우변항을 계산하여 정리하면 식(20)과 같다.

$$\frac{\partial L_{tot}}{\partial u_k} = \sum_{i=1}^N \sum_{t=1}^T \frac{\partial L_{tot}}{\partial \alpha_t(j)} \frac{\partial \alpha_t(j)}{\partial b_j(x_t)} \frac{\partial b_j(x_t)}{\partial M(k)} \frac{\partial M(k)}{\partial u_k} \quad (19)$$

$$\frac{\partial L_{tot}}{\partial u_k} = \eta \sum_{i=1}^N \sum_{t=1}^T \frac{\alpha_t(j) \beta_t(j)}{b_j(x_t)} c_{jk}(u_k - x_t) \sum_{n=1}^M [(F(u, k) - 1) \cdot M(k)] \quad (20)$$

여기서, η 는 학습상수로서 $0 < \eta < 1$ 사이의 값을 갖는다.

4. 한국어 단독 숫자 인식실험

본 장에서는 MIN 모듈 함수와 가우시안 밀도 함수의 특성을 살펴보기 위하여 1차원의 모의 데이터를 두 개의 표본에 대하여 무작위로 생성시킨 후 주어진 데이터 집단에 대한 각 함수 특성에 대하여 고찰하였다. 또한 제안된 MIN 모듈을 갖는 HMM과 기존의 HMM 방법을 한국어 단독 숫자 인식실험을 통하여 인식성능 평가를 하였다.

4.1 가우시안 밀도 함수와 MIN 모듈 함수의 특성비교

본 절에서는 한국어 단독 숫자 인식실험에 앞서 MIN 모듈을 갖는 HMM에서 사용되는 MIN 모듈 함수와 기존의 연속 HMM에서 일반적으로 사용되는 가우시안 밀도 함수에 대하여 모의 데이터를 이용하여 각 함수의 특성에 대하여 비교 분석하였다. 모의 데이터 생성조건은 표 1과 같이 1차원의 데이터를 각 집단별로 200개씩 무작위로 발생시켰다. 여기서 표본 1은 이웃하는 집단의 중첩도가 적은 표본이며, 표본 2는 표본 1에 비해 이웃하

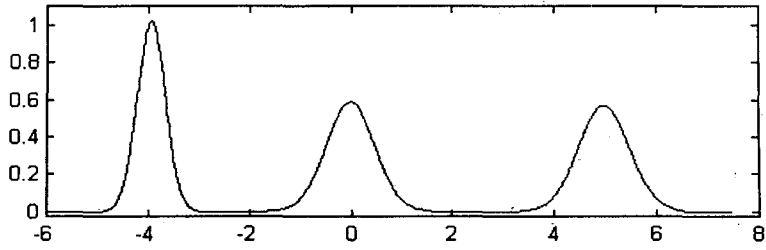
는 집단 사이의 중첩도가 비교적 크도록 설정하였다.

표 1. 모의 데이터 생성조건

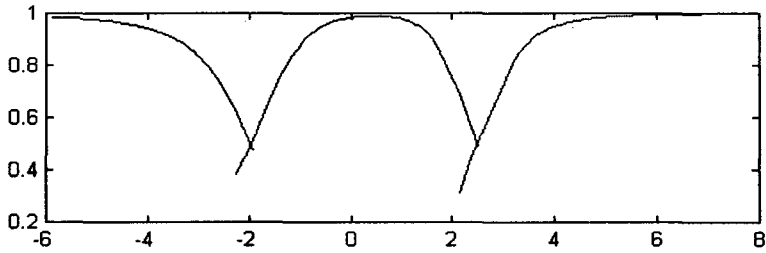
	표본 1		표본 2		표본 3	
	평균	표준편차	평균	표준편차	평균	표준편차
집단 1	-4	0.5	-2	0.5	-3	0.5
집단 2	0	0.8	0	1	0	0.8

표 1과 같은 조건을 갖는 모의 데이터를 생성한 후, 각 함수의 특성 곡선을 나타내기 위하여 생성된 데이터들의 실제적인 평균 및 분산 값을 다시 계산하여 각 표본에 대한 가우시안 밀도 함수와 MIN 모듈 함수의 출력 특성 곡선을 그림 2와 그림 3에 나타내었다. 그림 2에서 볼 수 있듯이 이웃하는 집단 간의 중첩도가 비교적 적은 표본 1은 가우시안 밀도 함수의 경우 각 클래스에 대하여 평균벡터를 중심으로 편차가 증가할수록 작은 출력 값의 특성을 보인다. 그러나 MIN 모듈 함수의 경우 가우시안 밀도 함수와는 달리 단지 집단 사이의 결정 경계면 근처에서 주어진 데이터에 대한 출력 값만이 작아짐을 알 수 있으며 두 집단 사이의 결정 경계면에서 멀리 떨어진 데이터에 대해서는 큰 활성화 값을 갖게 된다. 이와 같은 특성 곡선의 차이는 패턴인식 문제에서 중요한 요소로 동작될 수 있다. 예로서 -5의 값을 데이터에 대한 분류문제를 고려하면 쉽게 함수의 차이를 알 수 있다. 이 경우 가우시안 밀도 함수를 이용한 분류문제에 있어서 3 개의 가우시안 밀도 함수의 출력은 모두 0의 값을 갖게 됨으로서 주어진 데이터를 올바르게 분류할 수 없다. 그러나 거리에 기반한 MIN 모듈 함수의 경우는 평균벡터 및 편차에 관계없이 두 집단 사이의 결정 경계면에 의해 출력 함수의 특성 곡선이 나타남으로서 -5의 데이터 값에 대한 출력 값은 집단 1의 MIN 모듈 함수에서 1에 가까운 출력 값을 가지며 나머지 밀도 함수에 대해서는 0의 값을 갖는다. 따라서 분류문제에 있어서 주어진 임의의 데이터는 가장 큰 출력 값을 갖는 집단으로 분류되게 된다.

MIN 모듈 함수는 인접한 밀도 함수가 하나인 경우에 시그모이드 함수의 특성을 가진다. 이것은 MIN 모듈 함수가 이러한 함수에 근거하여 생성되었으며 인접한 집단에 의해 분포곡선이 결정되기 때문이다. 또한, 두 함수의 특성차이는 집단 2에 대한 특성 곡선에서도 발견될 수 있다. MIN 모듈 함수에서의 집단 2에 대한 특성 곡선은 가우시안 밀도 함수가 갖는 분산벡터의 활성화영역 폭에 비해 넓음을 알 수 있다. 이것은 MIN 모듈 함수가 거리에 기초한 함수이므로 인접한 평균벡터의 거리에 따라 함수의 폭이 결정되기 때문이다.

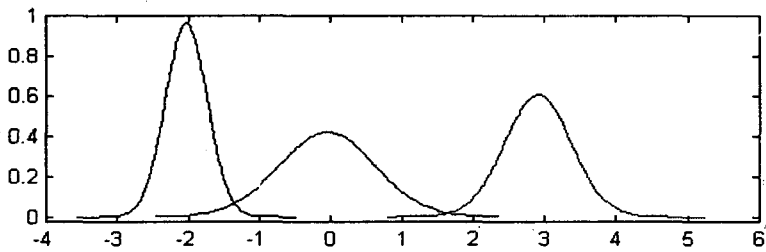


(a) Gaussian density function

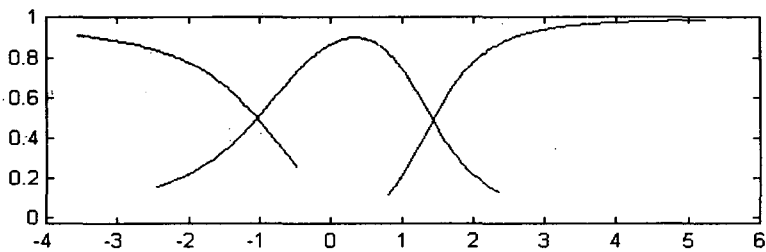


(b) MIN module function

그림 2. 표본 1에 대한 가우시안 밀도 함수와 MIN 모듈 함수의 특성 비교



(a) Gaussian density function



(b) MIN module function

그림 3. 표본 2에 대한 가우시안 밀도 함수와 MIN 모듈 함수의 특성 비교

4.2 한국어 단독 숫자 인식실험

본 실험에서는 기존의 모델과 제안된 MIN 모듈을 갖는 HMM의 인식성능을 비교평가하기 위해 원광대학교 낭독음성 DB의 단독 숫자에 대하여 화자독립 인식실험을 하였다. 원광대학교 낭독음성 DB 보급판은 화자 72 명을 대상으로 방음 부스에서 Synthesizer HMD 224 X를 사용하여 녹음되었고, A/D는 PC 환경에서 실시되었다. AD/DA 모듈은 KAY CSL 4300 B가 사용되었으며 16 kHz로 샘플링(sampling)하고 16 bits로 양자화 되어 있다. 단독 숫자 발성 데이터는 일반인 70명과 아나운서 2명이 각각 4회, 2회씩 발음하여 녹음되었다.

실험에서 사용된 숫자들은 영, 일, 이, 삼, 사, 오, 육, 칠, 팔, 구의 10개 숫자들을 사용하였다. 모델의 파라미터 추정과 인식성능 평가를 위하여 화자에 따라 학습용 데이터와 인식용 데이터로 구분하여 사용하였다. 학습용 데이터는 아나운서 2명이 발성한 숫자와 일반인 45명이 발성한 1,840개의 숫자를 사용하였으며, 인식용 데이터는 학습에 속하지 않은 25명의 화자가 발성한 1,000개의 숫자를 사용하였다.

4.2.1 전처리과정 및 특징벡터 추출

낭독 음성에서 선택된 숫자들은 멜 켈프스트럼 계수를 구하여 음성의 특징벡터로 사용하였다. 추출된 숫자는 $H(z) = 1 - 0.97z^{-1}$ 의 Pre-emphasis 과정을 거친 후, 16 ms의 길이를 갖는 Hamming window를 사용하여 8 ms씩 이동하면서 각 음성 프레임마다 12차의 멜 켈프스트럼 계수를 구하여 인식시스템의 특징벡터로 사용하였다. 표 2는 본 논문에서 사용한 음성의 특징벡터 생성과정의 분석 조건이다.

표 2. 음성 데이터의 분석 조건

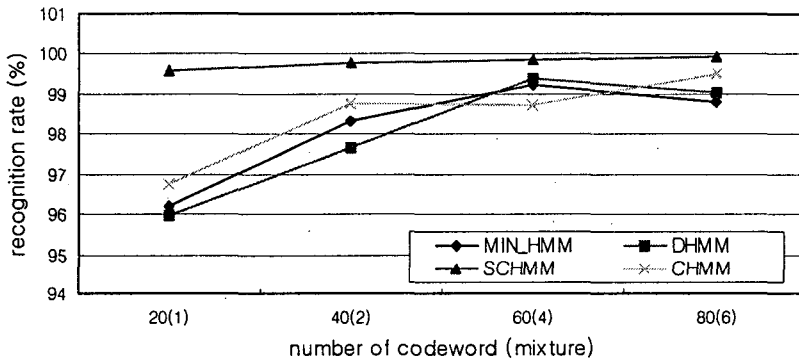
ADC	16 kHz, 16 bit, signed
Pre-emphasis	$H(z) = 1 - 0.97z^{-1}$
Hamming Window	16 ms (256 points)
Frame Rate	8 ms (128 points)
Feature Parameter	Mel Cepstrum (12차)

4.2.2 실험 결과 및 분석

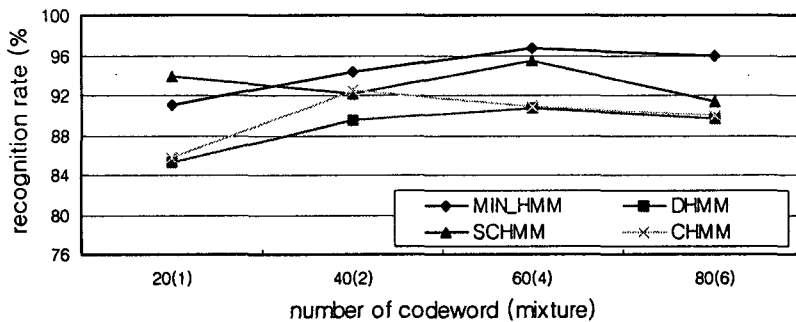
본 실험에서는 기존의 모델과 제안된 MIN 모듈을 갖는 HMM의 공평한 인식성능평가를 위해 동일한 조건을 갖는 실험환경에서 실험하였다. 이에 따라 모든 모델들은 상태의 생략이 없고 음성인식에 많이 이용되는 Left-Right 모델을 사용하였으며, 학습 데이터로부터 K-means 양자화 과정을 통해 추출된 평균벡터와 분산벡터들 및 HMM의 초기 파라미터는 각 방법에 모두 동일하게 사용하였다. 또한, 단독 숫자에 대한 HMM의 상태 수 결정은 3, 4, 5개에 대한 초기 인식실험을 하여 전반적으로 성능이 우수한 상태 수 4개에 대해서 이후 실험을 진행하였다. HMM 인식 시스템은 여러 가지 외적인 환경요소 및 시스템 구성 방법 등에 따라 인식성능이 결정된다. 그러나 본 실험에서는 이러한 요인들을 고려치 않은 가장 기본적이고, 일반적인 환경에서 실험을 하였다. 인식실험은 일반적인

HMM과 제안된 HMM의 성능 평가를 위하여 크게 두 가지 관점에서 실험을 행하였다. 하나는 코드워드(평균벡터)수가 인식시스템에 미치는 영향 및 코드워드 수 변화에 따른 각 방법에서의 인식성능 평가를 위한 실험이다. 또한, 파라미터 추정시 종료조건으로 설정한 임계값에 다르면 HMM의 학습은 종료되지만 그렇지 못할 경우에는 강제적으로 학습을 멈추어야 한다. 따라서, 두 번째 인식성능 평가 실험으로 학습횟수의 변화에 따른 각 방법에서의 인식성능 평가 및 각 방법의 학습 수렴정도의 평가에 관한 것이다.

코드워드 수(또는 mixture 수) 변화에 따른 학습 데이터 및 테스트 데이터의 인식결과는 그림 4와 같으며 실험환경으로 상태 수는 4개, epoch 수는 10번 그리고 제안된 방법(MIN-HMM)의 학습상수 η 가 0.01일 때의 결과이다. 또한 그림 4에서의 코드워드 수와 괄호 안의 mixture 수는 대응되는 개수는 아니다. 연속 HMM에서의 mixture 수는 각 상태에서의 수를 의미하므로 상태 수 및 mixture 수를 곱한 것이 실제적으로 모델에서 사용된 가우시안 혼합 밀도 수가 된다.



(a) 학습 데이터를 이용한 인식결과



(b) 테스트 데이터를 이용한 인식결과

그림 4. 코드워드 및 mixture 수에 따른 인식결과 [state=4, epoch=10, $\eta=0.01$]

코드워드 수 변화에 따른 학습 데이터에 대한 인식결과에서는 그림 4(a)에 나타나듯이

준연속 HMM이 전반적으로 가장 우수한 학습성능을 갖는 인식결과를 보였으며 코드워드 수의 증가에 따라 학습 데이터의 인식성능에는 크게 영향을 미치지 않음을 알 수 있다. 그러나 준연속 HMM을 제외한 나머지 방법에서는 코드워드 및 mixture 수 증가에 따라 대체적으로 인식률이 증가됨을 알 수 있었다. 또한 코드워드의 수가 60개일 때 모든 방법들이 99% 이상의 인식 성능을 보였다. 테스트 데이터에 대한 코드워드 수 변화에 따른 인식실험 결과는 그림 4(b)에서 볼 수 있듯이 제안된 MIN 모듈을 갖는 HMM이 코드워드 수가 20개일 때를 제외하고 가장 우수한 인식성능을 가짐을 알 수 있다. 그리고 연속 HMM에서는 mixture 수가 2일 때 가장 좋은 인식결과를 보였으며 다른 방법들은 코드워드 수가 60개일 때가 좋은 결과를 보였다. 전반적으로 학습에 따른 테스트 데이터에 대한 인식성능은 제안된 방법, 준연속 HMM, 연속 HMM, 이산 HMM의 순으로 우수함을 알 수 있다. 이와 같은 실험을 통해 우수한 학습성능을 나타내는 모델이 테스트 데이터에 대한 우수한 인식성능 결과를 나타내지 않았다는 점과 적절한 코드워드 수 및 mixture 수의 선택이 HMM에서의 인식 성능에 중요한 요소가 됨을 알 수 있었다. 특히 결과 값으로부터 준연속 및 연속 HMM은 테스트 데이터에 대하여 코드워드 수 변화에 따라 상당히 불규칙한 인식결과를 보였다. 이것은 가우시안 밀도 함수를 이용하는 모델들의 분산 추정치가 인식시스템 전체에 민감한 영향을 줌으로서 나타난 결과로 판단된다. 이와는 반대로 분산 추정치를 고려치 않는 이산 및 MIN 모듈을 갖는 HMM에서는 코드워드 수 변화에 따른 학습 데이터 및 테스트 데이터에서의 인식 결과가 코드워드 수 증가에 따른 인식률 향상의 규칙성을 보여 주고 있다. 이것은 인식시스템에서의 코드워드 수 선택에 있어 제안된 방법이 보다 일반적일 수 있다는 것으로 추정된다. 마지막으로, 학습 데이터를 이용한 인식시스템의 구현시 학습 데이터에 대한 우수한 인식결과를 갖는 학습 성능은 테스트 데이터에 대한 높은 인식성능을 의미하지 않음을 알 수 있었다. 이것은 이전에 기술한 바와 같이 두 데이터 사이의 통계적 특성 차이 및 기타 외적인 환경요소에 따라 인식시스템의 성능이 달라지기 때문인 것으로 판단된다.

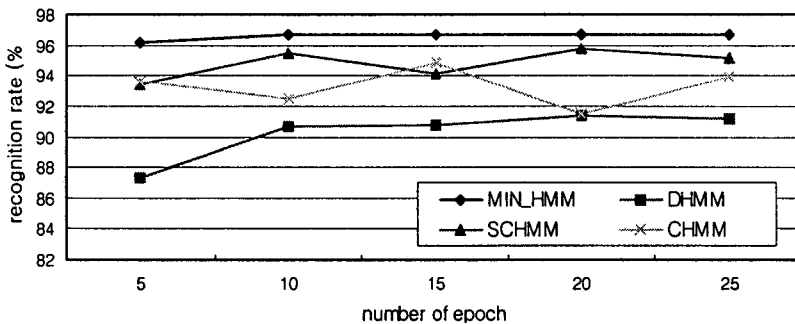


그림 5. Epoch수 변화에 따른 인식률[state=4, 코드워드(mixture)=60(4), $\eta=0.01$]

그림 5는 이전의 실험에서 60개의 코드워드 수와 2개의 mixture 수에 따른 epoch 수 변화에 따른 인식실험 결과이다. 이 실험은 epoch 수의 증가에 따른 인식성능 평가에 대한 것으로 높은 인식률을 나타내는 epoch 수를 찾기 위한 것이 아니라 학습의 수가 증가되었

을 때 인식률의 변화과정의 수렴성에 관한 실험이다. 실험 제한조건으로 임계점의 수렴조건을 두지 않았으며 해당되는 epoch 수일 때까지 학습과정이 반복되도록 하였다. 테스트 데이터에 대한 epoch 수 변화에 따른 인식결과에서 제안된 방법 및 이산 HMM은 학습 epoch 수 증가에 따라 인식률이 좋아짐을 알 수 있었지만 연속 분포를 이용하는 HMM 방법들은 이전의 코드워드 수 변화에 따른 인식실험과 같이 불규칙적인 인식성능을 나타내었다. 이것은 평균벡터와 분산벡터가 존재하는 연속분포 HMM에서는 epoch 수의 증가에 따른 인식 결과의 수렴성이 제안된 방법 및 이산 HMM에 비해 다소 떨어지는 결과를 보였다. 이것은 이전에 기술하였던 분산벡터에 대한 파라미터 추정 값이 민감하게 HMM 인식시스템의 인식성능에 영향을 주는 것으로 판단된다. 결과적으로 제안된 방법이 epoch 수의 변화에 따른 인식실험에서도 다른 기존의 방법들에 비해 비교적 우수한 결과를 보임을 알 수 있었다.

5. 결 론

본 논문에서는 연속분포를 이용한 HMM에서 학습 및 테스트 데이터에 대하여 통계적 특성에 차이가 발생 할 때 평균벡터와 분산벡터에 의한 가우시안 밀도 함수를 거리 기반 함수로 대치함으로써 데이터 집합간의 통계적 차이 문제를 보완하였다. 이에 신경회로망에서의 거리 기반 함수로 MIN 모듈 함수를 사용하여 HMM에 하이브리드 형태 또는 하나의 구성요소로 결합하였으며 제안된 MIN 모듈을 갖는 HMM에서 각 파라미터 추정은 기울기 경사법에 근거하며 각 패턴에서의 분산은 고려치 않는다.

제안된 MIN 모듈을 갖는 HMM의 학습 및 인식성능 평가를 위해 기존의 각 방법과 한국어 숫자에 대하여 화자독립 인식실험을 하였다. 실험결과에서 알 수 있듯이 제안된 방법이 기존 방법에 비해 학습 데이터에 대하여는 분산의 특성을 고려하지 않음으로서 기존 연속분포를 이용한 HMM에 비해 다소 학습 성능이 떨어지지만 테스트 데이터에 대한 인식 성능은 가장 우수한 것으로 나타났다. 일반적으로 패턴인식 시스템에서의 테스트 데이터의 인식 결과가 인식시스템의 성능을 좌우하게 된다. 결론적으로, 연속 분포를 이용하는 HMM에서의 인식 성능을 좌우하는 중요한 요소는 학습 데이터와 테스트 데이터 사이의 통계적 특성이다. 주어진 데이터들이 무한하다면 연속 HMM이 가장 이상적인 모델이지만 실제적으로 여러 가지 제한 요소로 인해 두 데이터 사이의 통계적 특성이 불일치는 경우가 많이 존재한다. 따라서, 통계적 특성이 불일치되는 데이터들을 이용하여 HMM의 인식시스템을 구성할 때에는 인식 성능에 민감한 요소로 작용하는 평균 및 분산에 기반한 연속 밀도 함수보다는 상대적 거리에 기반한 연속 분포 함수를 이용하는 것이 좋은 인식결과를 보였다. 그러나 거리 기반 함수의 MIN 모듈함수를 적용할 때 기존 HMM보다 더 많은 계산량을 가짐으로 이에 대한 함수의 간략화 과정이 앞으로 연구 과제로 제시된다.

참 고 문 헌

- [1] J. Picone, "Continuous speech recognition using hidden markov models," IEEE ASSP Magazine, Vol. 7, pp. 26-41, July, 1990.
- [2] K. F. Lee and H. W. Hon, "Speaker-independent phone recognition using hidden markov models," IEEE Trans. on ASSP, Vol. 37, pp. 1641-1648, 1989.
- [3] L. R. Rabiner and B. H. Juang, "An introduction to hidden markov models," IEEE ASSP Magazine, pp. 4-16, 1986.
- [4] D. K Kim, S. H Lee and B. S Kim, "Generalized LVQ for Optimal Reference Vectors using a Differentiable MIN Module," ICONIP, vol. 3, pp. 1937-1942, 1994.
- [5] J. A. Leonard and M. A. Kramer, "Radial Basis Function Network for Classifying Process." IEEE Control Systems, vol. 11, No. 3, pp. 31-38, 1991.
- [6] M. T. Musavi, W. Ahmed, K. H. Chan, K. B. Faris, and D. M. Hummels, "On the Training of Radial Basis Function Classifiers," Neural Network, vol. 5, No. 4, pp. 595-603, July, 1992.
- [7] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, 1978.
- [8] B. H. Juang, "Maximum-likelihood estimation for mixture multivariate stochastic observations of markov chain," AT&T Technical Journal, Vol. 64, pp. 1235-1249, 1985.
- [9] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, John Wiley & Sons, New York, 1973.
- [10] Y. Linde, A. Buzo and R. M. Gray, "An algorithm for vector quantizer design," IEEE trans. on Comm., Vol. 28, pp. 84-95, 1980.
- [11] X. D. Huang, Y. Ariki and M. A Jack, *Hidden Markov Models for Speech Recognition*, Edinburgh University Press, pp. 136-263, 1990.

접수일자: 2000. 10. 31.

게재결정: 2000. 11. 20.

▲ 김대극

강원도 춘천시 동면 장학리 790번지 (우: 200-850)
 한림정보산업대학 전자통신과
 Tel: +82-33-240-9165
 Fax: +82-33-252-9462
 E-mail: dkgold@sun.hallym-c.ac.kr

▲ 이정주

강원도 춘천시 효자2동 192-1 (우: 200-701)
 강원대학교 전자공학과
 Tel: +82-33-250-6320

Fax: +82-33-256-6327

E-mail: ljj@mirae.kangwon.ac.kr

▲ 정호균

강원도 춘천시 효자2동 192-1 (우: 200-701)

강원대학교 전자공학과

Tel: +82-33-250-6320

Fax: +82-33-256-6327

E-mail: fantasio@mirae.kangwon.ac.kr

▲ 이상희

강원도 춘천시 효자2동 192-1 (우: 200-701)

강원대학교 전자공학과

Tel: +82-33-250-6321

Fax: +82-33-256-6327

E-mail: slee@kangwon.ac.kr