

HMM-Net 분류기의 효율적인 학습법

°김 상 운, 김 탁 령

명지대학교 컴퓨터공학과
449-728, 경기도 용인시 남동 산 38-2

An Efficient Learning Method of HMM-Net Classifiers

°Sang-Woon Kim, Tak-Ryong Kim

Dept. of Computer Engineering, Myongji Univ.
San 38-2 Nam-Dong, Yongin-Shi, Kyounggi-Do, 449-728 Korea
Email : kimsw@wh.myongji.ac.kr

Abstract

The HMM-Net is an architecture for a neural network that implements a hidden Markov model (HMM). The architecture is developed for the purpose of combining the discriminant power of neural networks with the time-domain modeling capability of HMMs. Criteria used for learning HMM-Net classifiers are maximum likelihood (ML) and minimization of mean squared error (MMSE).

In this paper we propose an efficient learning method of HMM-Net classifiers using a ML-MMSE hybrid criterion and report the results of an experimental study comparing the performance of HMM-Net classifiers trained by the gradient descent algorithm with the above criteria. Experimental results for the isolated numeric digits from /0/ to /9/ show that the performance of the proposed method is better than the others in the respects of learning and recognition rates.

1. 서 론

은닉 마르코프 모델(hidden Markov model; HMM)^[1]은 시계열 패턴의 통계적인 성질을 잘 표현할 수 있기 때문에 음성인식을 위한 대표적인 방법으로 알려져 있다. 그러나 HMM은 최우도(maximum likelihood; ML) 학습기준을 사용하기 때문에 변별력이 약하다는 단점이 있어 있다^[1].

한편, 인공 신경망 모델(artificial neural network; ANN)은 유연성과 견고성, 탁월한 변별능력, 병렬 구현 특성 등의 장점이 있는 반면 훈련시간이 길다는 점(인식 시간은 짧다)과 시계열 패턴의 순서특성을 모델링하기 어렵다는 단점이 있다^[2].

이러한 문제를 해결하기 위한 방법으로 ANN을 HMM의 전처리기로 사용하는 방법(이를 ANN/HMM

하이브리드결합이라 한다)과 후처리기로 사용하는 방법(HMM/ANN 하이브리드)^[2], 그리고 HMM과 ANN의 장점만을 살리기 위하여 시계열 특성을 잘 반영하는 HMM을 보다 변별력이 있는 신경망 구조로 구현하는 방법 등이 있다^[3].

HMM-Net(hidden Markov model network)은 전향 알고리즘과 후향 알고리즘을 이용하여 HMM의 출력확률을 계산하는 과정을 ANN으로 구현한 신경망 구조이다. 따라서 HMM-Net의 학습을 위하여 신경망 학습에 널리 쓰이는 최소 자승오차(minimization of mean squared error; MMSE)의 학습기준에 의한 기울기 강하법(gradient descent)을 적용할 수 있다^[3].

그러나 ML과 MMSE의 학습기준에 대한 학습식 효율도하여 HMM-Net 분류기의 학습특성을 고찰한 결과, ML의 경우 학습이 빠른 반면 인식률이 저조하고 MMSE의 경우 학습속도는 느린 반면 인식률은 양호하다는 연구결과가 있다^[4].

따라서 이 논문에서는 우선 학습속도가 빠른 ML을 이용하여 적절한 초기값을 설정한 후 MMSE의 학습기준을 적용하여 최적 파라미터값을 학습하는 방법을 제안한다. 또한 숫자음성 데이터를 대상으로 HMM-Net 분류기를 학습하여 인식성능을 비교한 실험결과를 보고한다.

우선 제 2절에서는 HMM의 출력확률을 계산하는 과정을 신경망으로 구현한 HMM-Net의 구조에 대하여 설명하고, 제 3절에서는 기울기 강하법으로 HMM-Net 분류기를 학습하는 과정에 대하여 기술한다. 그리고 제 4절에서는 고립된 숫자 음성 데이터를 대상으로 HMM-Net 분류기를 학습한 결과와 학습된 분류기의 변별력에 대하여 고찰한다. 끝으로 제 5절에서 결론을 맺는다.

2. HMM-Net의 구조

HMM을 이용하여 시계열 패턴을 인식한다고 하는 것

은 클래스별로 HMM을 준비한 다음 관측된 출력 기호열이 어느 HMM으로부터 출력되는 확률이 높은 것인지를 결정하는 문제로서, HMM의 한 모델 λ_M 이 어떤 관측열 $O = v_{y_1}, v_{y_2}, \dots, v_{y_T}$ 를 출력할 확률 $\lambda_M(y_1^T)$ 은

$$\lambda_M(y_1^T) = P_M(y_1^T) = \prod_{i=1}^T \alpha_i(t) \beta_i(t) \quad (1)$$

이 된다. 여기서, $\alpha_i(t)$ 와 $\beta_i(t)$ 를

$$\hat{\alpha}_i(t) = \frac{\sum_{j=1}^N \tilde{\alpha}_j(t-1) a_{ji} b_{iv_j}}{\sum_{j=1}^N \sum_{k=1}^M \tilde{\alpha}_j(t-1) a_{jk} b_{kv_j}}, \quad \hat{\alpha}_i(0) = \pi_i \quad (2a)$$

$$\hat{\beta}_i(t) = \frac{\sum_{j=1}^N a_{ji} b_{iv_j} \tilde{\beta}_j(t+1)}{\sum_{j=1}^N \sum_{k=1}^M \tilde{\alpha}_j(t-1) a_{jk} b_{kv_j}}, \quad \hat{\beta}_i(T) = 1 \quad (2b)$$

와 같이 정규화를 수행하여, 출력확률 $\Lambda_M(y_1^T)$ 을 $\Lambda_M(y_1^T) = \log P_M(y_1^T) = \Lambda_M(y_1^{T-1}) + \log \sum_{i,j} \tilde{\alpha}_j(t-1) a_{ji} b_{iv_j}$ (3)

와 같이 재귀적으로 계산한다.

HMM-Net은 식(3)의 계산 과정을 신경망으로 구현한 것으로서, 구조는 그림 1과 같다.

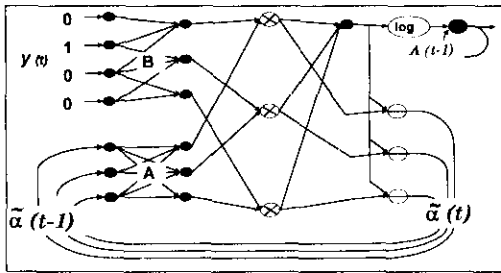


그림 1. HMM-Net 구조^[3]

Fig. 1. Structure of HMM-Net^[3]

그림에서 ●는 각 링크의 값을 더하는 노드이고 ○는 두 입력을 서로 곱하는 노드이며 ⊖는 하나의 입력을 다른 입력으로 나누는 노드이다. 그리고 A 와 B 는 HMM의 상태전이확률과 출력기호확률에 대응하는 연결 가중치이고 A , B 이외의 모든 연결 가중치 값은 1이다. 또한 $y(t)$ 는 M 차원 이진 입력벡터로서 t 시간의 관측기호 v_{y_t} 에 해당하는 노드만 1이고 다른 노드는 모두 0의 값을 갖는다.

3. HMM-Net 분류기의 학습

HMM-Net 분류기는 HMM-Net 모델 K 개를 이용하여 K 개의 클래스를 식별하는 분류기로서, 학습방법은

HMM의 학습기준인 ML과 신경망의 학습기준인 MMSE를 이용하는 기율기 강화법이 있다.

기율기 강화법을 위한 ML 및 MMSE의 학습 기준함수는

$$G^{ML} = \sum_n \Lambda_{C_n}(n) \quad (4a)$$

$$G^{MMSE} = -\frac{1}{2} \sum_n \sum_k (X_k(n) - \delta_{k,C_n})^2 \quad (4b)$$

이다. 여기서, $X_k(n)$ 과 $\Lambda_k(n)$ 은 k 번째 모델에서 n 번째 학습기호열에 대한 출력확률 및 로그를 취한 출력확률이고, C_n 은 n 번째 기호열의 클래스이며, $\delta_{i,j}$ 는 크로커키 델타함수이다.

학습기준 ML은 각 학습패턴이 단지 자신의 클래스에 소속할 정도만을 고려하는 반면, MMSE는 다른 클래스(오인식 클래스)에 미치는 영향까지 고려하는 학습기준이다. 따라서 성공적으로 학습되었을 경우 MMSE는 ML보다 정확한 결정경계를 제공하지만, 학습속도가 느리고 지역 최소값에 빠지는 등 효율적으로 학습시키 못하는 경우가 있다.

따라서 이 연구에서는 ML을 이용하여 비교적 빨리 안정된 학습상태로 수렴하도록 한 후, MMSE를 이용하여 정교하게 학습할 수 있도록 하는 방법을 제안한다. 즉, 제안 방법에서는 ML을 이용하여 초기파라미터 값을 학습하다가 G^{ML} 의 기율기가 미리 설정한 기준값 이하로 감소하면 적절한 파라미터값이 탐색되었다고 판단하여 기준함수를 G^{ML} 에서 G^{MMSE} 로 바꾸어 최적의 파라미터를 탐색하는 방법이다.

기율기 강화법으로 기준함수의 극점을 찾기 위한 파라미터 a_{ij} 에 대한 G 의 변화량은

$$\frac{\partial G}{\partial a_{ij}} = \sum_{n=1}^N \sum_{k=1}^K \phi_{n,k} \frac{\partial \Lambda_k(n)}{\partial a_{ij}} \quad (5)$$

으로, $\phi_{n,k} = \partial G / \partial \Lambda_k(n)$ 는 식(4)로부터 구하고, $\partial \Lambda_k(n) / \partial a_{ij}$ 는 그림1로부터 계산한다. 우선 노드 ○를 통과하여 정규화 되기 전의 값(노드 ⊗의 출력)을

$$\hat{\alpha}_i(t) = \sum_j \hat{\alpha}_j(t-1) a_{ji} b_{iv_j} \quad (6)$$

이라 하면, 로그를 취하기 직전의 값(노드 ●의 출력)은

$$\hat{\alpha}(t) = \sum_j \hat{\alpha}_j(t) \quad (7)$$

이 된다. 따라서, 노드 ⊗와 노드 ●의 출력값인 $\hat{\alpha}_p(t-1)$ 와 $\hat{\alpha}(t-1)$ 로부터 $t-1$ 에서의 a_{ij} 변화에 따른 $\hat{\alpha}_p(t-1)$ 의 변화량을 구한 다음, 이를 이용하여 t 에서의 노드 ⊗와 노드 ●의 출력값 변화 $\partial \hat{\alpha}_i(t) / \partial a_{ij}$ 와 $\partial \hat{\alpha}(t) / \partial a_{ij}$ 를 구한다. 이 과정을 재귀적으로 수행하여 $t=1$ 부터 T 까지의 $\partial \hat{\alpha}(t) / \partial a_{ij}$ 를 구한 후, $\partial \Lambda_k(n) / \partial a_{ij}$ 를 계산한다.

파라미터 b_{jv} 에 대한 G 의 변화량도 같은 방법으로 계산한다.

4. 실험 및 결과고찰

4.1. 실험 데이터

이 연구에서는 전자통신연구소에서 배포한 진화 음성 자료^[5] 중에서 남성 한명이 /영/부터 /구/까지 10개의 숫자를 4번씩 발음한 음성을 실험 데이터로 이용하였다. 이 데이터는 3.4KHz로 필터링하고 8KHz로 샘플링한 후 16비트로 양자화 한 다음, 앞 뒤의 묵음 부분을 제거한 후 10차 LPC계수를 추출하여 이용하였다.

실험방법은 실험 데이터에 대하여 전표본 학습법(resubstitution; R법)으로 실험하였다. 그리고, 학습률은 ML에서는 0.01을, MMSE에서는 0.02를 사용하였고, 학습횟수는 최고 500회까지로 하였으며 G 값이 떨어지지 않고 진동하여도 학습을 종료시켰다.

4.2. 결과 고찰

HMM-Net 분류기의 학습특성을 고찰하기 위하여 ML과 MMSE의 기준함수로 학습하였다. 또한 두 기준함수를 결합한 학습으로 우선 ML로 학습하다가 MMSE로 바꾸어 학습하는 방법(ML/MMSE)과 MMSE로 학습하다가 ML로 학습하는 방법(MMSE/ML)으로 실험하였다.

학습 반복회수(epochs)에 따른 기준함수 G 의 변화를 나타내는 학습 특성곡선은 그림 2와 같다. 여기서 수평축은 반복회수이고 수직축은 기준함수이다.

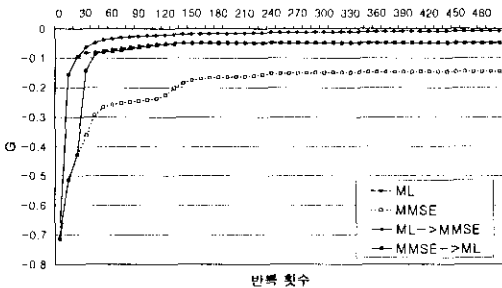


그림 2. 기준함수별 학습곡선

Fig. 2. Learning curves for criteria functions.

위의 그림은 네가지 학습방법중에서 ML/MMSE의 경우에 가장 양호하게 학습됨을 보여준다. ML 학습기준에서는 빠르게 학습이 되나 최적상태로 수렴하지는 못하였음을 보여준다. 또 MMSE에서는 ML보다 느리게 학습되고 학습성능도 떨어짐을 보여준다. 그러나 학습패턴의 규모를 보다 크고 다양하게 구성하여 실험하였을 경우 평균 학습률은 MMSE가 ML보다 우수하였다[4].

즉, MMSE의 학습기준은 ML보다 정교하게 학습시킬 수 있으나 응용에 따라서는 ML보다 학습성능이 떨어지는 경우도 있었다. 따라서 IIMM-Net분류기를 효율적으로 학습시키기 위하여 우선 빠르고 안정된 학습기준인 ML로 파라미터의 초기값을 설정한 다음, MMSE로 더욱 정교하게 학습시키는 방법을 이용할 수 있다.

학습방법별 HMM-Net 분류기의 인식률은 표 1과 같다. 여기서 ML/MMSE로 학습한 분류기의 인식률이 가장 양호하고 견고함을 보여준다.

표 1. HMM-Net 분류기의 인식률

Table 1. Classification rates of IIMM-Net classifiers

학습기준	ML	MMSE	ML/MMSE	MMSE/ML
인식률 (표준편차)	0.9250 (0.07)	0.8000 (0.25)	1.0000 (0.00)	0.9250 (0.00)

5. 결론

HMM-Net은 HMM의 시계열 모델링 능력과 ANN의 우수한 변별력을 결합하기 위하여 HMM을 ANN으로 구현한 것이다. 이 논문에서는 ML과 MMSE 및 이 두 기준을 결합한 ML/MMSE와 MMSE/ML 학습기준에 기울기 강하법을 적용하여 HMM-Net 분류기를 학습한 후 인식률을 고찰하였다.

실험 결과, ML로 초기학습을 수행한 다음 MMSE로 다시 학습하는 ML/MMSE 방법이 학습속도 및 인식 성능 면에서 ML이나 MMSE만으로 학습하는 경우보다 양호함을 확인하였다. 앞으로는 벡터 양자화 오차에 기인한 인식률 저하를 방지하기 위하여 아나로그 학습데이터를 이용하여 ML/MMSE로 최적 학습하는 방법을 연구한 계획이다.

참고문헌

- [1] H. Bourlard, C. J. Wellekens, "Links between Markov models and multilayer perceptrons", IEEE Trans. on PAMI, Vol.12, No. 12, pp. 1167-1178, 1990.
- [2] Y. Bengio, Neural Networks for Speech and Sequence Recognition, International Thomson Pub. Co., pp. 91-144, 1995.
- [3] L.T. Niles and H.F. Silverman, "Combining hidden Markov model and neural network classifiers", Proceedings of ICASSP, pp. 417-420, April 1990.
- [4] 김상운, 오수환, "HMM-Net 분류기의 학습", 대한전자공학회논문지, 제 34권 C편, 제 9호, pp. 703 - 709, 1997. 9.
- [5] 한국전자통신연구소, ETRI 음성DB 배포 설명서, 1995.