

하이브리드법에 의한 HMM-Net 분류기의 학습

°김 상 운, 신 성 효

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

On Learning of HMM-Net Classifiers Using Hybrid Methods

°Sang-Woon Kim, Seong-Hyo Shin

Dept. of Computer Engineering, Myongji University
San 38-2 Nam-Dong, Yongin-Si, Kyounggi-Do, 449-728 Korea
Email : kimsww@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), maximum mutual information (MMI), and minimization of mean squared error(MMSE).

In this paper we propose an efficient learning method of HMM-Net classifiers using hybrid criteria, ML/MMSE and MMI/MMSE, 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)과 인공 신경망 모델(artificial neural network; ANN)은 서로 보완적인 특성을 갖는 계산모델이다. HMM은 패턴의 시계열특성을 잘 반영할 수 있으나 변별력이 약하고, ANN은 변별력이 우수하고 패턴분포의 통계적 변화에 유연하나 시계열 패턴의 순서특성을 모델링하기 어렵다^[1].

HMM-Net(hidden Markov model network)은 HMM과 ANN의 장점만을 결합하기 위하여, HMM의 출력확률 계산과정을 ANN으로 구현한 신경망 구조이다. 따라서 HMM-Net의 학습을 위하여 HMM학습에

이용되는 ML(maximum likelihood), MMI(maximum mutual information) 학습기준이나 신경망 학습에 널리 쓰이는 MMSE (minimization of mean squared error)의 학습기준을 적용할 수 있다^[2]. ML에서는 각 학습패턴이 단지 자신의 클래스에 소속할 정도만을 고려하는 반면, MMI나 MMSE에서는 다른 클래스(오인식 클래스)에 미치는 영향까지도 고려하는 학습기준이다.

ML, MMI, MMSE의 학습기준에 대한 학습식을 유도하여 HMM-Net 분류기의 학습특성을 고찰한 결과, ML을 이용할 경우 학습이 빠른 반면 인식률이 저조하고 MMSE의 경우 학습속도는 느린 반면 인식률은 양호하다는 연구결과가 있다^[3]. MMSE는 성공적으로 학습되었을 경우 ML보다 정확한 결정경계를 제공하지만, 학습속도가 느리고 지역 최소값에 빠지는 등의 단점이 있다. 따라서 이 논문에서는 우선 ML이나 MMI로 학습하여 적절한 초기값으로 설정한 후 MMSE를 적용하여 최적 또는 준최적으로 학습하는 하이브리드 방법을 제안한다.

우선 제 2장에서는 HMM-Net의 구조를 정의하고, 제 3장에서는 학습기준에 기울기 강하법(gradient descent)으로 HMM-Net 분류기를 학습하는 방법을 설명한다. 그리고 제 4장에서는 고립된 숫자 음성 데이터를 대상으로 실험한 결과에 대하여 고찰하고, 제 5장에서 결론을 맺는다.

2. HMM-Net

HMM 모델 λ_M 이 관측열 $O = v_{y_1}, v_{y_2}, \dots, v_{y_T}$ 를 출력할 확률 $\lambda_M(y_1^T)$ 은

$$\lambda_M(y_1^T) = P_M(y_1^T) = \prod_{t=1}^T \alpha_t(t) \beta_t(t) \quad (1)$$

으로, $a_i(t)$ 와 $\beta_i(t)$ 를

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

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

와 같이 정규화하면, 출력확률 $\Lambda_M(y_1^T)$ 은

$$\Lambda_M(y_1^T) = \log P_M(y_1^T) = \Lambda_M(y_1^{T-1}) + \log \sum_{j=1}^N \tilde{a}_j(t-1) a_{ji} b_{ij} \quad (3)$$

이 된다. HMM-Net은 식(3)의 계산 과정을 신경망으로 구현한 것으로 위상구조는 그림 1과 같다.

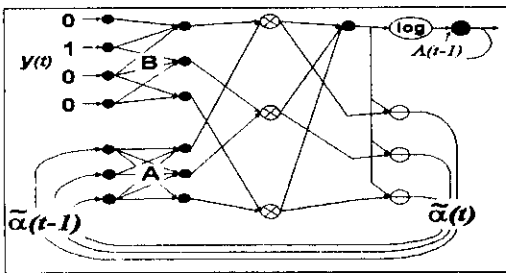


그림 1. HMM-Net 구조^[2]
Fig. 1. Structure of HMM-Net^[2]

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 상태전이 확률과 출력기호 확률에 대응하는 연결 가중치이다.

3. 학습기준

HMM-Net 분류기를 기울기 강하법으로 학습할 때 이용할 수 있는 학습기준에는 HMM의 학습기준인 ML이나 MMI와 신경망의 학습기준인 MMSE를 이용할 수 있다. 기울기 강하법을 위한 ML 및 MMSE의 학습기준함수는

$$G^{ML} = \sum_n \log X_{c,n} \quad (4a)$$

$$G^{MMI} = \sum_n \left(\log X_{c,n} - \log \left(\frac{1}{K} \sum_{k=1}^K X_k(n) \right) \right) \quad (4b)$$

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

이다. 여기서, $X_k(n)$ 은 k번째 모델에서 n번째 학습기호열에 대한 출력확률이고, C_n 은 n번째 기호열의 클래스이며, K는 클래스의 수, $\delta_{i,j}$ 는 크로네커 델타

함수이다.

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. 한편, MMI는 자신의 클래스에 소속할 정도를 상대적인 값으로 나타낸 기준이다.

기울기 강하법으로 기준함수의 극점을 찾기 위한 파라미터 a_{ij} 에 대한 G의 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 그림1.로부터 재귀적으로 계산한다.

또한, b_{ij} 에 대한 G의 변화량 $\partial \Lambda_k(n) / \partial b_{ij}$ 도 같은 방법으로 계산한다.

4. 실험 및 고찰

4.1. 실험 준비

HMM-Net분류기의 학습특성과 인식률을 고찰하기 위한 실험 데이터는 한국의 전자통신연구원(ETRI)^[4]에서 배포한 음성 데이터베이스의 전화 음성을 이용하였다. /0/부터 /9/까지의 고립 숫자음중에서 9명의 남자가 4번씩 발음한 360개의 데이터(data set 1; DS1)와 9명의 여자가 4번씩 발음한 360개의 데이터(data set 2; DS2) 그리고 남녀의 발음을 혼합한 360개의 데이터(data set 3; DS3)를 임의로 선택하여 사용하였다.

전화 음성 데이터를 3.4KHz로 필터링하여 8KHz로 샘플링한 후 16비트로 양자화하고, 앞 뒤의 묵음 부분을 제거한 후 10차 LPC계수를 추출하여 패턴벡터를 구성하였다. 벡터 양자화(VQ)는 SOM(self-organizing feature map)법^[5]을 이용하였으며 관측기호의 수는 10, 12, 16으로 다르게 하여 실험하였다.

실험방법은 실험 데이터에 대하여 전표본 학습법(resubstitution; R법)으로 실험하였다. 그리고, 학습률은 ML과 MMI에서는 0.01로 하고, MMSE에서는 0.02로 하였다. 또한 학습횟수는 최고 300회까지로 하였으며, 학습기준함수 G값이 떨어지지 않고 진동하여도 학습을 종료시켰다.

4.2. 결과 고찰

학습 실험은 ML, MMI 및 MMSE의 기준함수로 학습하는 방법과 우선 ML이나 MMI로 초기학습한 후 MMSE로 바꾸어 학습하는 ML/MMSE와 MMI/MMSE의 하이브리드법으로 실험하였다.

ML이나 MMI로부터 MMSE로 학습기준을 바꾸기 위하여 다음과 같이 기준함수 G의 그라디언트 값의 변

화 ΔG 를 고찰하였다.

$$\Delta G = |\nabla G(t+5) - \nabla G(t)| \quad (5)$$

여기서, ∇G 는 기준함수 G 의 그라디언트값이고, t 는 반복회수(epochs)이다. 실험데이터 DS1, DS2, DS3에 대한 ΔG 의 변화는 그림 2와 같다.

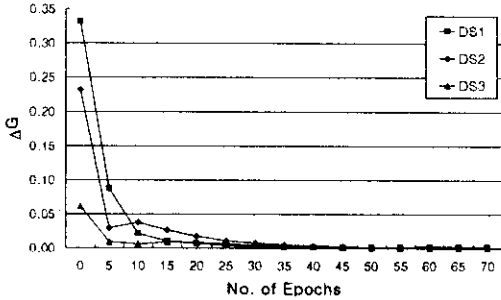


그림 2. ΔG 의 변화

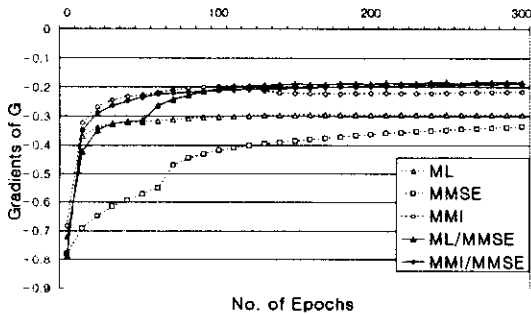
Fig. 2. Variation of ΔG

여기서 기준함수는 G^{ML} 로서 ΔG 는 학습초기에는 급격히 감소하지만 반복회수 t 가 50을 상회하면서는 거의 변화하지 않음을 보여준다. 그리고 기준함수를 G^{MMI} 로 하였을 경우에도 같은 결과를 보였다. 따라서 이 연구에서는 ML과 MMI로 50회 학습한 후 학습기준을 MMSE로 바꾸어 학습하였다.

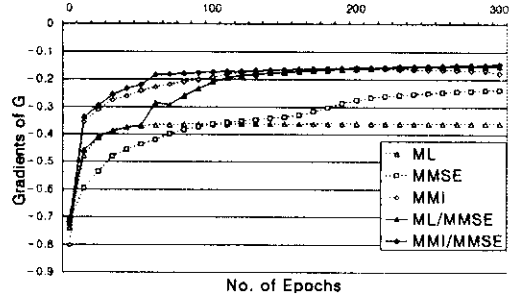
관측기호의 수를 12로 하여 벡터양자화하였을 경우 DS1, DS2, DS3에 대한 학습 특성곡선은 각각 그림 3 (a), (b), (c)와 같다.

그림 3은 ML에서는 빠르게 학습이 되나 최적상태로 학습되지 못하였고, MMSE는 ML이나 MMI보다 느리게 학습되나 반복회수를 충분히 크게 하였을 경우 ML보다 우수하게 됨을 보여준다.

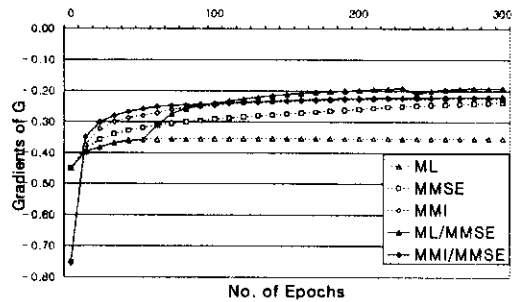
따라서 HMM-Net분류기의 학습은 ML로 학습하여 파라미터의 초기값을 설정한 다음, 다시 MMSE로 학습기준을 바꾸어 학습시키는 하이브리드법이 보다 효율적인 학습방법임을 알 수 있다.



(a) DS1



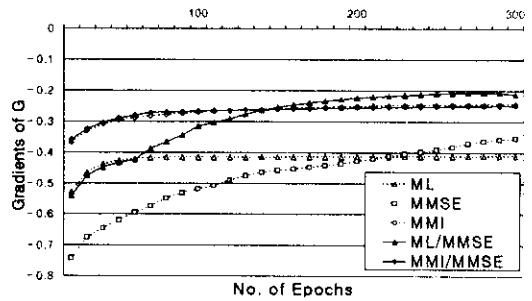
(b) DS2



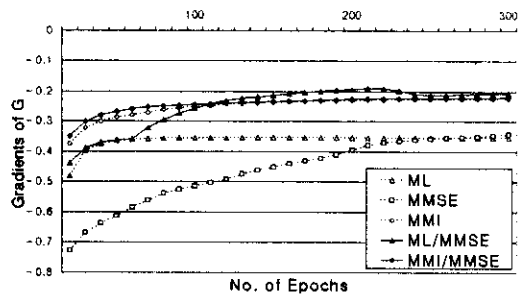
(c) DS3

그림 3. 학습데이터 DS1, DS2, DS3에 대한 학습곡선
Fig. 3. Learning curves for experimental data set.

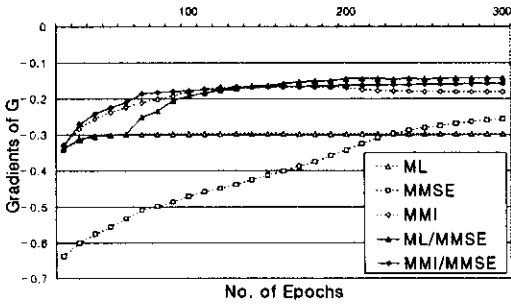
그림 4는 학습 데이터 DS3에 대하여 관측기호의 수를 10, 12, 16으로 하여 벡터양자화하였을 경우의 학습 특성곡선으로, 그림 3과 같은 결과를 보여준다.



(a) VQ의 관측기호가 10인 경우



(b) VQ의 관측기호가 12인 경우



(c) VQ의 관측기호가 16인 경우

그림 4. 학습데이터 DS3에 대한 학습곡선
Fig. 4. Learning curves for DS3.

그림 5는 DS3를 사용하여 세 번 학습하였을 경우 기준함수별 학습 곡선의 편차(variation)를 보여준다. MMSE에서의 편차가 큰 반면, ML에서는 작은 편차를 보여준다.

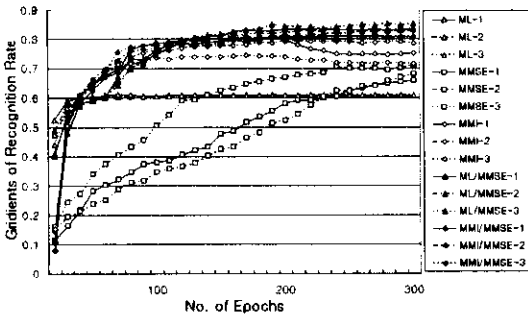


그림 5. 학습곡선의 편차
Fig. 5. Variations of learning curves

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

5. 결론

이 논문에서는 HMM-Net 분류기를 학습하기 위하여 ML, MMI, MMSE의 학습기준과 이들을 하이브리드한 ML/MMSE, MMI/MMSE에 기술기 강화법을 적용하였을 경우의 학습특성과 인식률을 고찰하였다.

서로다른 학습조건에서 실험한 결과, ML이나 MMI로 초기학습한 다음 MMSE로 학습기준을 바꾸어 학습하는 하이브리드 방법이 ML, MMI, MMSE만으로 학습하는 경우보다 양호함을 확인하였다. 또한 벡터 양자화를 수행할 때 관측기호의 수에 따라 학습특성이 다르게 나타났다. 따라서 앞으로 응용에 따라 최적으로 벡터 양자화할 수 있는 방법이 연구되어야 한다.

표 1. HMM-Net 분류기의 인식률

Table 1. Classification rates of HMM-Net classifiers

Data Set	No*	ML	MMI	MMSE	ML/MMSE	MMI/MMSE
DS1	10	48.89 (0.0)	73.89	53.89 (7.6)	77.13 (1.7)	71.94
	12	60.28 (0.0)	69.17	61.39 (4.5)	76.48 (1.1)	77.22
	16	66.39 (0.0)	74.44	71.02 (8.4)	88.33 (1.3)	88.06
DS2	10	49.44 (0.0)	77.22	59.91 (4.5)	79.26 (0.6)	79.44
	12	51.94 (0.0)	73.89	61.39 (7.7)	74.72 (8.4)	83.33
	16	61.39 (0.0)	82.22	75.19 (10.7)	78.98 (4.2)	87.78
DS3	10	47.50 (0.0)	63.61	48.06 (3.4)	71.39 (2.4)	64.44
	12	50.56 (0.0)	66.67	58.33 (8.9)	74.63 (3.6)	72.78
	16	60.83 (0.0)	74.54 (3.83)	69.17 (3.6)	85.74 (2.4)	81.76 (1.12)

No* : Number of Symbols in VQ

참고문헌

- [1] H. Bourlard and C. J. Wellekens, "Links between Markov models and multilayer perceptrons", IEEE Trans. on PAMI, vol. 12, no. 12, pp. 1167 -1178, 1990.
- [2] L.T. Niles and H.F. Silverman, "Combining hidden Markov model and neural network classifiers", Proceedings of ICASSP, pp. 417-420, April 1990.
- [3] 金商雲, 吳嶽煥, "HMM-Net 分類器의 學習", 大韓電子工學會論文集, vol. 34-C, no. 9, pp. 703 - 709, 1997. 9.
- [4] 韓國電子通信研究院, ETRI 音聲DB 配布 說明書, 1995.
- [5] <http://www.cis.hut.fi/nncr/nncr-programs.html>.