HMM의 출력확률을 이용한 신경회로망의 성능향상에 관한 연구

표창수, 김창근, 허강인
동아대학교 전자공학과

A study on performance improvement of neural network using output probability of HMM

Chang Soo Pyo, Chang Keun Kim, Kang In Hur
Dept. of Electronics, Dong-A University
E-mail : kihur@mail.donga.ac.kr

요 약

본 논문은 HMM(Hidden Markov Model)을 이용하여 인식을 수행할 경우의 오류를 최소화 할 수 있는 후처리 과정으로 신경망을 결합시켜 HMM 단독으로 사용하였을 때 보다 높은 인식률을 얻을 수 있는 HMM과 신경망의 하이브리드시스템을 제안한다. HMM을 이용하여 학습한 후 학습에 참여하지 않은 데이터를 인식하였을 때 오인식 데이터를 정인식으로 인식하도록 HMM의 출력으로 얻은 각 출력확률을 후처리에 사용될 MLP(Multilayer Perceptrons)의 학습용으로 사용하여 MLP를 학습하여 HMM과 MLP을 결합한 하이브리드 모델을 만든다. 이와 같은 HMM과 신경망을 결합한 하이브리드 모델을 사용하여 단독 숫자음과 4연 숫자음 데이터에서 실험한 결과 HMM 단독으로 사용하였을 때 보다 각각 약 4.5%, 1.3%의 인식률 향상이 있었다. 기존의 하이브리드 시스템이 갖는 많은 학습시간이 소요되는 문제점과 실시간 음성인식시스템을 구현할 때의 학습데이터의 부족으로 인한 인식률 저하를 해결할 수 있는 방법임을 확인할 수 있었다.

Ⅰ. 서 론

음성인식 시스템을 구성하는 방법으로 HMM과 신경망을 가장 많이 사용하고 있다. HMM법은 음성의 변동을 통계적으로 처리하고 이 통계량을 확률형태의 모델에 반영하여 음성을 인식하는 방법이며 확률모델을 사용하기 때문에 개인차나 조음결합의 영향 등에 의한 음성패턴의 변동을 반영하기 쉽고 음소나 음절단위의 모델을 단어나 문장 등의 단위로 확장할 수 있다. 이와 같이 HMM은 음성을 모델링하는 효율적인 방법이며 실제로 좋은 인식결과를 얻을 수 있다. 그럼에도 불구하고 HMM방법에도 음성데이터를 모델링하는 확률, 통계적인 가정에서 출발하는 한계를 가지는 단점을 지니고 있다고 할 수 있다.

인공신경망은 인간의 뉴런을 단순한 모델로 만들어 많은 뉴런의 결합에 의한 적절한 구조를 만들어 패턴분류의 문제에 성공적으로 적용하고 있는 방법이다. 음성인식에 사용하고 있는 가장 보편적인 인공신경망으로는 MLP를 들 수 있다. MLP는 교사신호에 의해 입력 패턴군에 대한 목표출력을 나타내도록 뉴런간의 결합 가중치를 조절하여 입출력간의 매핑에 의해서 패턴을 분류하는 것으로 음소나 음절 인식 단계에서 우수한 성능을 보이고 있으나 이러한 방법에도 음성의 시간에 따른 변화의 특성에는 적절히 대처하지 못하는 단점을 가지고 있다. 그래서 본 논문에서는 HMM의 확률, 통계적인 접근법으로 인한 패턴분류의 오류를 MLP를 사용하여 보상하는 방법을 제안한다.

II. HMM과 MLP의 하이브리드 구조

제안하는 HMM의 후처리기로서 MLP를 결합한 하이브리드 시스템의 구성은 그림1과 같다.

음절 인식의 경우는 먼저 각 분류목록에 대응되는 학습데이터를 연속출력분포 HMM을 사용하여 각 분류목록에 대응하는 HMM학습 테이블을 만든 다음, HMM 학습에 참여하지 않은 데이터를 학습 완료된 HMM테이블을 사용하여 각각의 출력확률을 계산하여 각 음절에 대응하는 하나의 출력확률벡터를 만든다. 이렇게 생성된 출력확률의 벡터열을 MLP의 입력층으로 인가하여 MLP를 학습하여 하이브리드시스템 구성을 완료한다. 여기서 HMM 학습에 참여하지 않은 데이터를 사용하여 출력확률 벡터열을 만들어 MLP를 학습하는 이유는 학습에 참여한 데이터를 사용하여 출력확률 벡터열을 구성할 경우 HMM에 의한 오인식의 경우가 적으므로 MLP의 학습에 HMM에서 오인식 되어지는 데이터의 특성이 반영되지 않기 때문에 HMM과 MLP의 패턴분류 특성이 같아짐으로 인하여 본 논문에서 제안한 HMM의 확률, 통계적인 특성으로 인한 패턴분류의 오류를 MLP를 이용하여 보상해주는 하이브리드시스템의 성능을 제대로 평가하지 못하는 문제가 발생하기 때문이다. 하이브리드 시스템의 평가는 평가용 데이터를 입력 하였을 때 MLP의 출력 유니트에서 가장 큰 출력값을 나타내는 음절로 인식을 한다. 4연 숫자음의 경우 MLP의 입력층으로 입력할 수 있도록 음절의 경우와 같이 출력확률 벡터열을 만들기 위해서 먼저 Viterbi 알고리즘을 사용하여 입력데이터를 4개의 음절로 segmentation을 한 다음 음절과 같은 과정으로 학습과 인식을 한다.

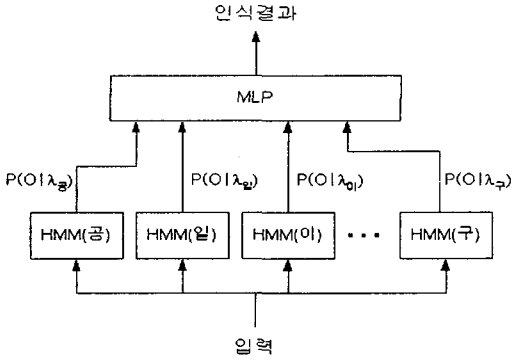


그림 1. HMM과 MLP의 하이브리드 시스템

2.1 연속출력분포 HMM

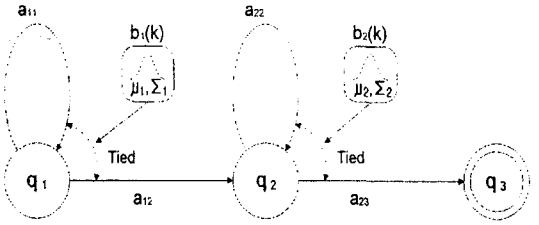


그림 2. 연속 출력확률 HMM

Left-to-right형 HMM은 그림 2와 같은 유한 오토마타로 정의된다. HMM을 이용한 음성인식의 경우는 먼저 인식에 필요한 수만큼의 표준패턴을 학습해 두고 입력패턴에 대하여 그 출력확률이 최대가 되는 표준패턴을 인식결과로 한다. 연속 출력분포 HMM의 경우 상태 i 에서 j 로의 천이확률 a_{ij} 및 천이경로에서 심벌 k의 출력확률 b_{ijk}를 학습 데이터에서 구하기 위한 Baum-Welch 알고리즘은 다음과 같다.

상태수를 N, 심벌계열의 길이를 T, 전향 확률을 $\alpha(i,t)$ $(i=1,2,\cdots,N\ ;\ t=1,2,\cdots,T)$ 라 하고, 후향확률을 $\beta(j,t)$ $(j=1,2,\cdots,N;\ t=T,T-1,\cdots,0)$, 모델 M의 심벌 계열 $o = o_1\ o_2 \cdots\ o_T$ 를 출력하는 확률을 $P(o \mid M)$, 상태 i에서 상태 j로의 천이가 시각 t에서 발생할 확률을

$$\gamma_t(i,j) = \frac{\alpha(i,t-1)a_{ij}b_{ij}(o_t,\mu_{ij},\sum_{ij})\beta(j,t)}{P(o \mid M)} \quad (1)$$

로 정의하면 천이확률의 추정식은

$$a_{ij} = \sum_t \gamma_t(i,j) \Big/ \sum_t \sum_j \gamma_t(i,j) \quad (2)$$

$$b_{ijk} = \sum_{t:\,o_t=k} \gamma_t(i,j) \Big/ \sum_t \gamma_t(i,j) \quad (3)$$

와 같고, 출력벡터 o_t가 n차원의 정규분포에 따른다고 가정할 수 있는 경우 출력확률 밀도함수는

$$b_{ij}(o_t,\ \mu_{ij},\ \sum_{ij}) = \frac{\exp\{-(o_t-\mu_{ij})'\sum_{ij}^{-1}(o_t-\mu_{ij})/2\}}{(2\pi)^{n/2}\,|\sum_{ij}|^{1/2}} \quad (4)$$

로 주어진다. 여기서, μ_{ij}는 출력벡터의 평균치, Σ_{ij}는 공분산행렬, t는 전치, -1은 역행렬을 나타낸다. 여기서 μ_{ij}, Σ_{ij}의 추정식은 다음 식으로 주어진다.

$$\mu_{ij} = \sum_t \gamma_t(i,j)o_t \Big/ \sum_t \gamma_t(i,j) \quad (5)$$

$$\Sigma_{ij}=\frac{\sum_t \gamma_t(i,j)(o_t-\mu_{ij})(o_t-\mu_{ij})'}{\sum_t \gamma_t(i,j)} \quad (6)$$

2.2 MLP

입력층은 HMM의 결과치인 출력확률의 10차원 벡터를 받아들이기 위해 10개의 유니트를 사용하였으며 출력층은 숫자 10개를 분류하기 위하여 10개를 사용하였고 하나의 중간층을 사용하였으며 20개의 유니트를 사용하였다. MLP의 학습은 일반적으로 많이 사용하고 있는 오차 역전파(error back propagation) 알고리즘을 사용하였으며 학습률은 0.01, 관성률은 0.95를 사용하였다.

III. 인식실험

3.1 음성DB 및 분석조건

단독 숫자음 데이터로는 ETRI의 샘돌이 데이터 중에서 "공,일,이,삼,사,오,육,칠,팔,구" 10개의 음성을 사용하였다.

이는 남성화자 20명이 10개 숫자음을 4회 발성한 총 800개의 데이터 중에서 10명의 1회 발성한 100개의 데이터를 HMM 학습용으로 사용하고 다른 10명의 1회 발성한 100개의 데이터로 MLP 학습을 하여 하이브리드시스템을 구성하고 남은 20명의 3회 발성한 600개의 데이터로 성능 평가를 하였다.

4연 숫자음은 KAIST의 데이터 중 4명의 35개의 4연숫자음을 4번 발성한 데이터를 사용하였으며 1회 발성분 560음절을 HMM 학습용으로 사용하고 다른 1회분 560음절을 MLP 학습에 사용하였으며 남은 2회분 1,120음절을 평가용 데이터로 사용하였다.

음성데이터로 사용하는 단독 숫자음과 4연 숫자음의 분석조건은 표1과 같다.

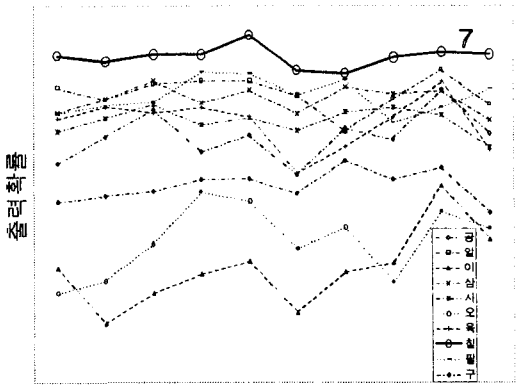
표 1. 음성데이터 분석 조건

A/D 데이터	16kHz, 16bit
프레임 간격	3.75ms
분석창	Hamming 창
분석창 길이	16ms
특징파라메타	10차 LPC Melcepstrum

3.2 실험결과 및 고찰

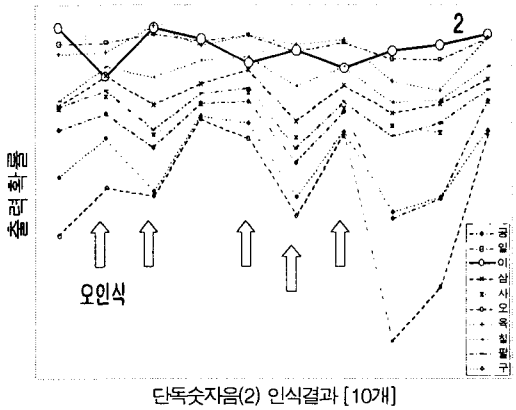
본 논문의 실험에서는 그림 2에 있는 left-to-right형 모델인 연속출력분포 HMM을 사용하였고 각 숫자음을 위해 5상태로 구성하였다.

단독 숫자음의 경우, 우선 100개의 학습용 데이터로 HMM을 학습시킨 후, 학습에 참여하지 않은 다른 100개의 데이터로 각 숫자음에 대한 출력확률을 계산한 다음 그 출력확률 벡터열을 MLP의 입력으로 하여 MLP를 학습시킨다. 그리고, 평가용 데이터로 각 숫자음을 대표하는 모든 HMM의 출력확률이 구해지고 이런 모든 확률들에 의해 형성되어진 벡터열을 신경망에 적용한다. MLP 출력층의 유니트 중 가장 높은 값을 가지는 유니트에 의해 대표되는 단어가 인식 숫자음으로써 선택되어진다. 4연 숫자음의 경우, Viterbi 알고리즘을 사용하여 4개의 숫자음으로 segmentation을 한 다음 단독 숫자음의 경우와 같이 학습 및 인식 실험을 한다.



단독 숫자음(7) 인식결과 [10개]

그림 3. HMM 인식 결과



단독숫자음(2) 인식결과 [10개]

그림 4. HMM 인식결과

그림3과 4는 각 10개의 숫자음 7과 2에 대한 HMM의 출력확률 그래프이다. 그림3은 숫자음 7에 대한 HMM의 인식결과가 정인식으로 나타난 결과이다. 그림 4는 숫자음 2에 대한 10개의 평가 데이터에 대한 5개의 오인식(일,육,육,육,육)을 보여주는 그래프이다. 여기서 정인식을 한 경우와 오인식의 경우에서 근소한 차이를 보여주고 있다. 또한 오인식의 경우 전체적인 출력확률의 분포에서 낮은 출력확률을 가지면서 오인식 되지 않음을 알 수 있다. 이러한 오인식 출력확률을 정인식으로 인지하도록 후처리 과정으로 MLP를 사용하여 HMM의 패턴분류 성능을 보상할 수 있었다.

표 2. 인식률(%)

	단독 숫자음	4연 숫자음
HMM	86.0	94.5
HMM+MLP	90.5	95.8

실험 결과 단독 숫자음과 4연 숫자음의 인식실험결과는 표2와 같다. 하이브리드 시스템의 성능평가에 있어서 HMM의 학습 데이터 조합, MLP의 결합 가중치의 초기값 설정 등의 원인으로 인한 인식률의 변동이 ±1% 발생하였다.

Ⅳ. 결론

본 논문에서는 HMM과 신경망을 결합한 하이브리드 모델을 사용하여 단독 숫자음과 4연 숫자음에 대하여 인식 실험을 하였다. 실험결과 HMM을 단독으로 사용하였을 때의 인식률 86%, 94.5% 보다 본 논문에서 제안한 하이브리드시스템의 인식률이 약 4.5%, 1.3% 향상됨을 알 수가 있었다. 이것은 HMM이 오인식으로 판단하는 데이터를 후처리 과정을 통한 MLP가 오인식 데이터를 정인식 데이터로 인지하도록 보상의 역할을 하고 있음을 보여준다. 또한 적은 데이터로서 보다 나은 인식률을 얻을 수 있었고 기존의 하이브리드 시스템이 갖는 많은 학습시간이 소요되는 문제점과 실시간 음성인식시스템을 구현할 때의 학습데이터의 부족으로 인한 인식률 저하를 해결할 수 있는 방법임을 본 논문의 실험결과에서 확인할 수 있었다.

향후 적은 데이터로서도 HMM과 MLP의 하이브리드시스템의 성능을 높이기 위해서 다른 형태의 특징 파라메터를 추가한 실험과 실시간 연속 음성인식을 위한 미지의 입력데이터의 자동 segmentation을 할 수 있는 연구가 필요하다.

참고문헌

[1] José A. Martins and Fàbio Violaro : "Hybrid recognizers combining hidden markov models and multilayer perceptron", IEEE pp. 146-150 (1998)

[2] Katagiri S., Lee C. H., "A New Hybrid Algorithm for Speech Recognition Based on HMM Segmentation and Learning Vector Quantization ", IEEE Transaction on Speech and Audio Processing, vol. 1, no 4, pp 421-430, October 1993.

[3] Joe Tebelskis, "Speech Recognition using Neural networks," CMU-CS-95-142, 1995.

[4] Rabiner L. R., Wilpon J. G., Soong F. K., "High Performance connected digit recognition using Hidden Markov Models", IEEE Transaction on Acoustics, Speech, Signal Processing, vol. ASSP-37, pp 1214-1225, Aug. 1989.

[5] 심장엽, 이영재, 고시영, 이광석, 허강인, "HMM에 의한 연속음성인식 시스템의 구현". 제13회 음성통신 및 신호처리 워크샵 논문집 제13권1호, pp. 325-330, 1996.8.

[6] 김수훈, 고시영, 이영재, 허강인, "신경망 예측HMM을 이용한 음성인식", 제9회 신호처리 합동학술대회 논문집, pp.239-242, 1996.10.

[7] Rabiner L. R., Juang B. H., "Fundamentals of Speech Recognition", Englewood Cliffs, Prentice-Hall, 1993.

[8] S.H.Kim, S.Y.Koh, K.I.Hur: "A Study on the Recognition of the isolated Digits Using Recurrent Neural Predictive HMM", TENCON'99 Vol. I, p.593-596, 1999.