음성인식을 위한 새로운 혼성 recurrent TDNN-HMM 구조에 관한 연구

장 춘 서†

요 약

본 논문에서는 혼성 모듈 구조의 recurrent 시간지연신경회로망(time-delay neural network)과 HMM(hidden Markov model)을 결합한 음성인식을 위한 새로운 구조에 대해 연구하였다. 시간지연신경회로망에서는 윈도우 크기를 확장하는 것이 인식률 향상에 유리하므로 이를 위해 첫 번째 은닉층에 궤환 구조를 사용하여 윈도우 크기를 실제로 크게 하지 않고도 동일한 효과를 얻을 수 있도록 하였다. 다음 이 시간지연신경망에서 입력된 음소의 특징 벡터의 시간에 따라 변화하는 성질을 잘 처리 할 수 있도록 시간지연신경회로망의 입력층을 복수의 상태로 나누어 음소 특징의 시간축에 대한 각 상태마다 특징 감지기를 갖도록 하였다. 이때 시간지연신경회로망은 전체 음성인식 영역에 적용될 수 있도록 모듈 방식의 구조로 구성되었다. 그리고 이 모듈 구조 시간지연신경망의 출력 벡터를 HMM에 연결하여 서로 결합 하므로써 양 구조의 장점을 취하는 혼성 구조의 인식시스템을 구성하였고 이때 이 혼성 구조에서 효율적으로 적용할 수 있는 HMM 파라미터 smoothing 방법을 제시하였다.

A study on the new hybrid recurrent TDNN-HMM architecture for speech recognition

Choon-Seo Jang†

ABSTRACT

In this paper, a new hybrid modular recurrent TDNN (time-delay neural network)-HMM (hidden Markov model) architecture for speech recognition has been studied. In TDNN, the recognition rate could be increased if the signal window is extended. To obtain this effect in the neural network, a high-level memory generated through a feedback within the first hidden layer of the neural network unit has been used. To increase the ability to deal with the temporal structure of phonemic features, the input layer of the network has been divided into multiple states in time sequence and has feature detector for each states. To expand the network from small recognition task to the full speech recognition system, modular construction method has been also used. Furthermore, the neural network and HMM are integrated by feeding output vectors from the neural network to HMM, and a new parameter smoothing method which can be applied to this hybrid system has been suggested.

키워드 : 음성인식(speech recognition), 신경회로망(neural network), HMM

1. 서 론

음성인식 알고리즘은 패턴정합방법, 지식기반방법, stochastic 방법 및 신경회로망에 의한 방법 등 크게 4가지로 분류할 수 있다. 이들 중 stochastic 방법의 대표적 기법인 HMM(hidden Markov model)은 음소를 쉽게 모델링 할 수 있어 인식단어 수를 대용량 화 할 수 있는 반면 음소의 모델링을 위해서는 HMM의 상태의 개수, 각 상태에서의 확률분포 등을 적절하게 설정해야 한다[1, 2].

신경회로망에 의한 음성인식은 비교적 구성이 단순한 다층 퍼셉트론의 경우에도 HMM에 필적할 만한 성능을 보이고 있는데 음성특징을 스스로 학습하며, HMM과 같이 실제음성을 Markov 프로세스라고 가정하는데서 발생하는 오류가 없고 간단한 프로세싱요소들이 반복 연결되어 멀티프로세서 구성에 유리하다. 이를 사용한 구조로서는 다층퍼셉트론의 변형된 형태들, LVQ(learning vector quantizer)와 혼성구조 형태의 신경회로망, HMM과의 혼성구조 형태의 신경회로망 그리고 재귀형 신경회로망 등이 연구되었는데 [3-7] 이들은 정적입력패턴을 갖는 신호에 대해서는 비교적 좋은 인식률을 보이나 특성변화가 비교적 큰 음성입력의 테스트 음성신호 패턴의 시간 정렬이 학습시의 입력패턴과

※ 본 연구는 1999년도 금오공과대학교 학술연구비 지원에 의하여 연구된 논문임.
† 정 회 원 : 금오공과대학교 컴퓨터공학부 교수
논문접수 : 2001년 7월 23일, 심사완료 : 2001년 11월 2일

다를 때는 인식률이 저하되는 문제점을 가지고 있었다.

실제 음성인식 시에는 항상 신경회로망의 학습시의 패턴과 동일한 세그멘테이션을 맞출 수는 없기 때문에 인식률 저하의 원인이 되는데 이를 해결하기 위해 등장한 시간지연신경회로망(time-delay neural network)에서는 각 유닛의 시간지연요소들이 현재 입력신호와 과거 입력신호들에 대해 각각 다른 가중치를 가지게 되어 과거의 신호와 현재의 신호를 서로 관련시키며 비교하는 방식으로 동작하여 좋은 결과를 보였다[8, 9]. 이때 시간지연신경회로망의 첫 번째 은닉층(first hidden layer)의 윈도우 크기를 키우는 것이 인식률 향상에 도움이 되는데 윈도우 크기를 증가시키면 연결강도의 수도 증가하여 시간지연신경회로망의 학습시간이 크게 증가하게 되고 학습데이터도 많은 양이 필요하게 된다.

따라서 본 연구에서는 한국어 음성인식 시에 시간지연신경회로망의 첫 번째 은닉층에 궤환 구조를 갖게 하여 윈도우 크기를 실제로 크게 하지 않고도 동일한 효과를 내는 구조를 사용해서 시간지연신경회로망을 구성하고 이때 이 시간지연신경회로망은 한국어 전체 음소를 인식할 수 있도록 하기 위해 모듈 구조로 구성하였다. 또 시간지연신경회로망에서는 하나의 특징 감지기가 입력된 음소의 특징 벡터의 시간에 따라 변화하는 각 상태의 서로 다른 특징을 모두 학습해야 하는 문제점이 있어 음소특징의 시간적 구조를 잘 처리하지 못한다. 본 연구에서는 이를 개선하기 위해 음소특징의 시간축에 대한 각 상태마다 특징 감지기를 갖도록 하여 각 상태에 포함된 특징만을 학습하도록 하여 다중 상태 모듈구조로 구성하므로써 전체 시간지연신경회로망의 인식률을 높이도록 하였다.

그리고 본 논문에서는 모듈방식의 시간지연신경회로망의 출력을 HMM에 연결하여 결합 하므로써 양 구조의 장점을 취하는 혼성 구조의 인식시스템을 구성하였으며 이때 HMM과의 혼성 구조에서 HMM의 최적 모델 파라미터를 얻기 위한 효율적인 파라미터 smoothing 방법을 연구하였다. 본 논문에서 제안된 HMM 파라미터 smoothing 방법은 많은 양의 학습 데이터를 요하는 기존의 co-occurrence smoothing[10] 등과 같은 통계적인 방법이 아닌 신경회로망의 출력값을 사용하는 새로운 smoothing 방법이며 이를 통하여 혼성 구조의 인식시스템에서의 효율적인 파라미터 smoothing을 구현하여 전체 음성인식률을 높일 수 있음을 보였다. 본 논문에서는 1장에 이어 2장에서는 궤환 구조를 갖는 다중 상태 모듈 구조의 시간지연신경회로망에 대해 설명하였고 3장에서는 혼성 구조에서의 효율적인 HMM 파라미터 smoothing 방법이 제시되었으며 4장에서 실험결과 및 결론을 보였다.

2. 다중 상태 모듈 구조 recurrent 시간지연신경회로망

음소인식을 위한 시간지연신경회로망의 경우 입력층에서의 윈도우 크게는 음성 입력 프레임이 10 msec 단위로 들어올 때 보통 3개의 프레임 크기를 갖는다. 이때 첫 번째 은닉층에 의해 행해지는 신경망 내부의 표현은 더 넓은 범위의 음성 신호에 의해 영향을 받게 되므로 입력층에서의 윈도우 크기를 키울 필요가 있다. 그러나 윈도우 크기의 증가는 곧 연결강도 수의 증가를 가져오고 이에 따라 신경회로망의 학습에 필요한 시간과 데이터의 양이 커지게 된다.

윈도우 크기를 증가시키지 않고 입력 음성을 보다 넓은 의미관계로 처리할 수 있게 하기 위해서는 궤환 구조를 시간지연신경회로망에 도입시킬 수 있다. 여기서는 첫 번째 은닉층의 뉴론 출력과 이 층에 대한 입력 사이의 궤환 동작이 이루어지는 구조를 가지며 궤환 연결 강도들은 같은 값들이 시간 축 상으로 이동되어 연결되므로 원래의 시간지연신경회로망의 주요한 특성인 시간에 대한 불변성을 그대로 유지하게 된다. 이와 같은 궤환 구조에 의해서 이 신경망은 음성의 시간에 대한 순서적인 성질을 잘 처리할 수 있게 되며 음성의 시간적인 구조에 숨겨 있는 정보를 일반 시간지연신경회로망 보다 더 추출할 수 있게 된다.

궤환 구조를 가진 시간지연신경회로망에서 $y^m_{k,j}$를 m번째 은닉층의 뉴론 (k, j)의 출력이라고 하면 궤환을 고려할 때 이를 다음 식과 같이 나타낼 수 있다.

$$y^m_{k,j} = w^m_{k,1,j} y^m_{k,j} + \sum_p \sum_i w^m_{k,p,i,j}\, o^{m-1}_{k+p-1,i} \qquad (1)$$

여기서 $w^m_{k,1,j}$는 궤환 연결강도를 나타내고 $w^m_{k,p,i,j}$는 m번째 층의 뉴론 (k, j)와 m-1 번째층의 뉴론 (k+p-1, i)를 연결하는 연결강도를 나타내며 p는 시간이동 윈도우를 나타낸다. 그리고 $o^m_{k,j}$은 다음 식으로 나타내어진다.

$$o^m_{k,j} = f(y^m_{k,j}) \qquad (2)$$

이 식에서 함수 $f()$는 sigmoid 함수이다.

한국어 전체 음소를 하나의 단일 시간지연신경회로망으로 인식하도록 구성하는 경우 연결강도의 수, 필요한 학습 데이터의 량 및 학습시간 등이 크게 증가하므로 전체를 여러개의 서브 클래스 시간지연신경회로망으로 나누어 각 서브 클래스에서는 할당된 음소만을 처리하도록 하는 모듈 구조가 필요하다. 이때 각 서브클래스는 궤환 구조를 가지는 시간지연신경회로망으로 구성된다. 각 서브클래스 신경회로망은 각각의 학습 데이터를 사용해 독립적으로 학습된다. 다음 학습된 서브클래스 신경회로망들은 다시 합쳐져서 재

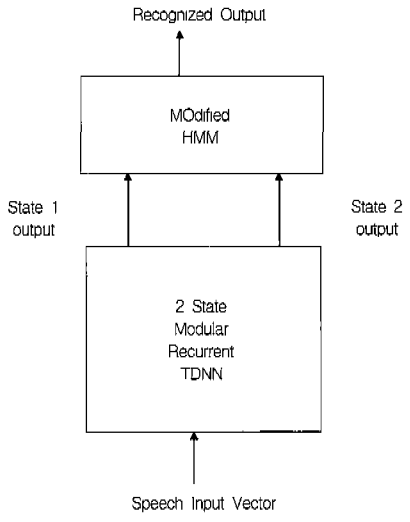
학습되는데 이때 첫 번째 은닉층과 두 번째 은닉층 사이의 연결강도가 주로 학습된다. 본 연구에서는 전체 음소의 개수와 각 서브클래스의 크기를 고려하여 전체 시스템을 모 11개의 서브클래스 신경회로망으로 구성하였다.

이와 같은 구조에서는 각 서브클래스에서의 특징탐지기가 시간에 대해 순서적으로 나열된 특징벡터를 주사하면서 음성의 특징을 탐지하는데 이때 하나의 특징탐지기가 음소의 특징벡터의 시간에 따라 변화하는 각 상태의 서로 다른 특징을 모두 학습해야 한다. 음소는 시간에 따라 변화하는 특성을 갖지만 다음 4개의 상태로 음소를 구별 할 수 있는 정보를 얻을 수 있다. 첫 번째 상태에서는 앞의 모음에서 자음으로의 천이구간에 대한 특징을 찾고, 두 번째 상태에서는 buzz-bar나 비음인 경우에 안정영역을 찾는다. 세 번째 상태에서는 다음 모음까지의 파열음이나 비음의 천이특징을 찾고 네 번째는 상태에서는 다음 모음으로의 특징을 찾는다.

따라서 음소특징의 각 상태마다 하나씩의 특징 탐지기를 갖게 하여 이들이 각 상태에 포함된 특징만을 학습하도록 하면 전체적으로 하나의 특징탐지기 만을 갖는 구조에 비해 상태 정보를 더 강조 할 수 있어 인식률을 높일 수 있다. 본 논문에서는 이를 구현하기 위해 각 모듈 방식으로 구성된 각 서브클래스 신경회로망의 입력층을 시간에 대해 다중 상태를 갖도록 하고 각 상태에서의 음성 입력패턴이 각각 서브클래스 신경회로망에 가해지게 하였다.

3. 혼성구조에서의 HMM 파라미터 Smoothing

본 논문에서는 2장에서 설명한 신경회로망과 HMM을 결합하는 혼성 구조를 채택하고 있는데 이때 HMM은 음성



(그림 1) 다중 상태 TDNN과 HMM의 혼성 구조

신호의 시간에 대한 연관성 및 특징 변화를 통합하고 추출하는데 좋은 성능을 보이고 있으며 음소 HMM 모델을 서로 연결하여 연속 음성 인식으로 확장하는데 유리하고 신경회로망은 학습을 통해 패턴을 구분하는 데에 좋은 성능을 보이므로 이를 통합하는 구조는 양자의 장점을 취할 수 있는 방법이다. 이 혼성 구조를 다음 (그림 1)에 보였다.

이 구성에서 보면 음성 특징 벡터 값들이 시간지연신경회로망에 가해지고 이 시간지연신경회로망의 출력이 HMM의 입력으로 가해지는 형태로 되어있다. 이와 같은 신경회로망과 HMM의 결합을 위하여 실제 궤환 구조에 다중 상태를 갖는 모듈형태의 시간지연신경회로망의 두 번째 은닉층 출력이 HMM과의 연결에 사용되었다. 위 그림에서는 신경회로망이 2개의 다중 상태를 갖는 경우를 보였으며 이 신경회로망의 각 상태에서의 출력이 HMM에 가해지고 이때 HMM은 다음에 설명하는 새로운 파라미터 smoothing 구조로 변형되었다.

규모가 큰 HMM에서 최적 모델 파라미터 특히 최적의 출력 확률분포함수를 얻기 위해서는 학습 데이터가 매우 방대해야 하는데 대부분의 경우 이를 충족시키기가 어렵다. 이와 같은 문제는 위의 혼성구조의 HMM에서도 마찬가지로 발생하며 따라서 학습과정에서 자주 발생하는 심볼들은 충분히 학습되는데 비해 그렇지 않은 심볼들의 경우 문제가 된다. 이러한 문제를 해결하기 위한 파라미터 smoothing 방법중 하나인 floor smoothing 에서는 미리 지정된 작은 상수 보다 관찰 확률이 작게 나오면 관찰 확률 값을 그냥 이 작은 상수로 하고 나머지 다른 확률들은 그 값들을 조정하여 전체 확률의 합이 1이 되도록 하고있다.

그러나 이 floor smoothing 방식은 구현이 간단한 반면 인식률의 성능 향상에 한계를 보이고 있다. 이에 비해 통계적 smoothing 방식에서는 출력 심볼의 상대 관찰 빈도를 얻은 후 관찰 확률이 각각의 통계값 들에 의해 smoothing 된다[11]. 그러나 이 방식은 HMM을 학습시키는데 많은 데이터를 필요로 하고 smoothing된 HMM 파라미터는 항상 인식률을 향상시키는 방향으로 결정되는 것은 아니며 smoothing 과정에서 잘 학습된 HMM 파라미터를 손상하는 경우도 생길 수 있어 smoothing 정도를 조절하는 파라미터에 대한 보간화 과정도 필요하다. 이러한 문제들은 전체 음소를 인식하기 위해 HMM 규모가 커질수록 더욱 영향을 미치게 된다.

따라서 본 연구에서는 기존의 통계적 방법이 아니라 신경회로망의 출력값을 사용하는 새로운 smoothing 방법을 제안하였다. 다중 상태 시간지연신경회로망의 두 번째 은닉층으로 부터의 출력신호열 O는 다음과 같이 나타낼 수 있다.

$$O=(o_{l1}, o_{l2}, \cdots, o_{lT}) \quad (3)$$

여기서 각 $o_{lt}, 1 \le t \le T$는 두 번째 은닉층으로 부터의 출력벡터이고 T는 두 번째 은닉층의 프레임 길이이며 l은 같은 층 시간 열에서의 다중 상태를 나타낸다. 이때 출력벡터 o_{lt}는 다음 식으로 나타낼 수 있다.

$$o_{lt} = (m_{lt}(1), m_{lt}(2), \cdots, m_{lt}(M)) \quad (4)$$

여기서 $m_{lt}(i), 1 \le i \le M$는 두 번째 은닉층 각 유닛으로부터의 출력이고 M은 두 번째 은닉층의 유닛 수이다. $P_l(q)$가 HMM 상태 q에서의 출력심볼 관찰확률 이라 하면 다음과 같이 나타낼 수 있다.

$$P_l(q) = (\Pr(\nu_{l1}|q), \Pr(\nu_{l2}|q), \cdots, \Pr(\nu_{lM}|q)) \quad (5)$$

여기서 $\Pr(\nu_{lt}|q)$는 HMM의 상태 q에서 출력심볼 ν_{lt}의 관찰확률이고 M은 두 번째 은닉층의 유닛 수이다. 이때 smoothing된 출력심볼 관찰확률 $P_{ls}(q)$은 다음 식으로 나타낼 수 있다.

$$P_{ls}(q) = P_l(q) S_l \quad (6)$$

여기서 $S_l = (s_{lij})$는 smoothing 매트릭스이다. 이 매트릭스의 원소 s_{lij}는 다음 식으로 나타낼 수 있다.

$$s_{lij} = g_{lij}^{1/(F-1)} / \sum_{k=1}^{M} g_{lik}^{1/(F-1)}, \ 1 \le i \le M, \ 1 \le j \le M \quad (7)$$

여기서 g_{lij}는 두 번째 은닉층의 각 유닛으로 부터의 평균 출력값이고 F는 smoothing 정도를 결정하는 값이다. i번째 음소의 학습 토큰 수를 K라고 하면 이 음소의 k번째 학습 토큰 $o_{lt}^{(k)}$는 다음 식으로 나타낼 수 있다.

$$o_{lt}^{(k)} = (m_{lt}^{(1)}, m_{lt}^{(2)}, \cdots, m_{lt}^{(M)}), 1 \le k \le K \quad (8)$$

이때 g_{lij}는 다음과 같이 된다.

$$g_{lij} = \sum_{k=1}^{K} \left(\sum_{t=0}^{T^{(k)}-1} m_{lt}^{(k)} / T^{(k)} \right) / K \quad (9)$$

여기서 $T^{(k)}$는 k번째 학습 토큰의 출력벡터의 프레임 길이이다.

4. 실험 및 결론

본 연구에서는 음소를 인식단위로 사용하고 한국어 전체 음소를 묵음을 포함하는 44개의 의미 독립(context-independent)적인 음소들로 분류하였다. 이들 음소들은 전체 17명의 화자가 발음한 음향적으로 균형 잡힌 75개의 단어들로 부터 얻어져서 데이터베이스를 구성하였다. 음성신호는 차단주파수 4.5 KHz인 저역필터를 거쳐 샘플링 주파수 10 KHz로 12비트 A/D 변환되었고 양자화된 음성 데이터는 휴지구간을 제거하여 음성 구간만을 검출하기 위한 끝점 검출과정을 거쳤다. 이들 샘플들은 pre-emphasis되어 30 msec 단위로 나누어져 Hamming 윈도우를 거치며 이때 20 msec씩 중첩시켰다. 여기서 12차의 LPC(linear prediction coding) 계수를 구한 후 이들 LPC 계수들로 부터 다시 cepstral 계수를 구하였고 이와 같이 구해진 LPC 기반의 cepstral 계수들의 주파수 특성을 mel 스케일로 보정 시키기 위해 bilinear 변환을 거쳐 최종적으로 mel-cepstral 계수를 얻었다.

<표 1>에는 HMM을 결합하지 않은 상태에서 궤환 구조가 있는 경우와 없는 경우의 인식결과를 보였다. 여기서 양쪽 경우 모두 첫 번째 은닉층의 윈도우 크기는 5로 하였고 다중 상태 구조는 사용하지 않았으며 모델 구조는 11개의 서브 클래스로 구성되었다. 궤환 구조가 없는 경우 각 서브 클래스 신경회로망의 학습을 최종 조율한 후 인식률을 측정한 결과 83.5%의 인식률이 나왔고 궤환 구조가 있는 경우 85.4%의 인식률이 나왔다. 표에서 Top1의 값은 첫 번째 후보에서 인식에 성공한 경우이며 Top2는 두 번째 후보에서 인식이 성공한 경우도 인식률에 포함시킨 값이다.

<표 1> 궤환 구조가 있는 경우와 없는 경우의 인식률

	인식률(%)	
	Top 1	Top 2
Non-recurrent	83.5	92.1
recurrent	85.4	93.8

여기서 궤환 구조를 사용한 경우 인식률이 1.9% 향상되어 이 구조가 인식률의 향상에 도움을 주고 있음을 알 수 있다.

입력층을 시간축에 대해 2개 및 3개의 상태로 각각 구분하여 궤환 구조를 갖는 경우와 그렇지 않은 경우에 대해 인식률을 측정한 결과를 <표 2>에 보였다. 음소 특징 벡터를 시간축 상으로 2개의 상태로 구분하는 경우에는 첫 번째 상태에서는 앞의 음소에서 현재 음소로의 천이특성을 검출하고 두 번째 상태에서는 현재음소에서 다음 음소로의 천이특성을 검출하게 하였고 3개 상태로 구분하는 경우에는 중간에 1개의 상태를 더 추가하였다. 궤환 구조를 갖는 구조에서 이때의 결과는 89.1%의 인식률을 얻어 상태 구분을 하지 않은 경우에 비해 3.7%의 향상을 나타내었다. 그러나 상태를 3개로 구분한 경우에는 궤환 구조를 갖는

경우와 갖지 않는 경우 모두 상태가 2개인 경우에 비해 각각 0.7%와 1.2%씩 인식률이 저하되었는데 이는 음소 구간의 중간에 추가로 첨가된 상태가 음소 특징 벡터의 시간에 대한 변화 상태를 올바르게 반영하지 못하기 때문으로 보인다.

〈표 2〉 다중 상태 구조인 경우의 인식률

상태 개수	궤환 구조	인식률(%)	
		Top 1	Top 2
2	Non-recurrent	87.7	95.4
	recurrent	89.1	96.7
3	Non-recurrent	86.5	94.8
	recurrent	88.4	96.1

다음 궤환 구조를 갖는 시간지연신경회로망과 HMM을 결합하여 혼성 구조로 구성하여 인식률을 측정한 결과를 <표 3>에 보였다. 여기서는 HMM 파라미터 smoothing을 floor smoothing 방법으로 하였을 때 신경회로망의 상태 개수를 1개에서 3개까지 변화시켜가면서 측정한 결과치 와 HMM 파라미터 smoothing을 본 논문에서 제안된 방법으로 하였을 때 신경회로망의 상태 개수를 역시 1개에서 3개까지 변화시켜가면서 측정한 결과치를 보여주고 있다. 이때 본 논문에서 제안된 smoothing 방법을 사용한 경우 <표 3>의 해당 값들은 식 (7)의 smoothing 정도를 결정하는 값 F를 1.5로 하여 측정한 결과이다. 이는 값 F를 변화시켜 가며 인식률을 측정한 결과 1.5에서 가장 좋은 결과를 보였기 때문인데 참고로 F가 1.3, 1.4, 1.6일 때 신경회로망의 상태 개수 2에서 인식률은 각각 94.4%, 95.1%, 94.9%의 결과를 보였다. 실험 결과로 혼성 구조 자체가 신경회로망의 다중 상태 구조와 상관없이 인식률을 향상시키고 있음을 알 수 있고 상태의 개수는 2개 일 때 가장 높은 인식률을 보였다. 또 혼성 구조에서의 HMM 최적 모델 파라미터를 얻기 위한 smoothing 방법의 비교에서는 상태 개수를 2로 하였을 경우 본 연구에 서 제안된 방식이 95.8%의 인식률을 보여 floor smoothing 방법의 92.4%에 비해 인식률을 3.4% 높일 수 있음을 보여주었다.

〈표 3〉 혼성구조에서의 smoothing 방법과 상태 개수에 대한 인식률

smoothing 방법	상태 개수	인식률 (%)	
		Top 1	Top 2
floor smoothing	1	91.7	97.1
	2	92.4	97.8
	3	92.1	97.4
제안된 smoothing	1	94.5	98.3
	2	95.8	98.9
	3	95.2	98.7

이상에서 본 논문에서는 시간지연신경회로망에서 윈도우 크기를 증가시키지 않고도 성능을 높일 수 있도록 궤환 구조를 갖도록 하고 이 시간지연신경회로망이 전체 음소를 인식할 수 있도록 전체를 11개 서브 클래스로 구성된 모듈구조로 나누어 각 서브 클래스에서는 할당된 음소만을 처리하도록 하며, 여기에 시간지연신경회로망의 출력을 HMM에 가하는 방식으로 혼성 구조를 구성하는 음성 인식 시스템에 대해 연구하였다. 이와 같은 혼성구조는 HMM과 시간지연신경회로망의 장점을 모두 얻을 수 있는 방법이며 이때 시간지연신경회로망에서는 입력 음소 특징 벡터의 시간축에 대한 각 상태마다 하나씩의 특징탐지기를 갖도록 하여 이들이 각 상태에 포함된 특징만을 학습하도록 하는 다중 상태 구조를 사용하여 성능을 높였다. 그리고 신경회로망과 HMM을 결합한 혼성 구조에서 HMM의 최적 모델 파라미터를 얻기 위한 효율적인 파라미터 smoothing 방법을 제시하였으며 이와 같은 방법들을 사용하여 시스템의 전체 인식률을 높일 수 있음을 보였다.

참 고 문 헌

[1] K. F. Lee, "Automatic speech recognition : the development of the SPHINX system," Kluwer Academic Publisher, 1989.

[2] N. Merhav and Y. Ephraim, "Hidden Markov modeling using the most likely state sequence," Proc. of Int. Conf. ASSP, pp.469-472, 1991.

[3] R. P. Lipmann, "Pattern classification using neural networks," IEEE Comm. Magazine, Vol.27, pp.46-47, Nov. 1989.

[4] H. Iwamida et al., "A hybrid speech recognition system using HMMs with an LVQ-trained code book," Proc. of IEEE Int. Conf. ASSP, pp.489-492, 1990.

[5] S. Katagari et al., "A new HMM/LVQ hybrid algorithm for speech recognition," IEEE Proc. of GLOBECOM'90, pp. 1032-1036, 1990.

[6] L. T. Liles and H. F. Silverman, "Combining hidden Markov model and neural network classifiers," Proc. of IEEE Int. Conf. ASSP, s.8.2, pp.417-420, 1990.

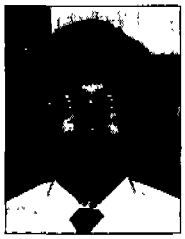
[7] E. Trentin and M. Gori, "A survey of hybrid ANN/HMM models for automatic speech recognition," Neurocomputing 37, pp.91-126, 2000.

[8] A. Waibel, et al., "Phoneme recognition using time-delay neural networks," IEEE Trans., ASSP, Vol.37, pp.328-339, March. 1989.

[9] A. Waibel, "Modularity and scaling in large phonemic neural networks," IEEE Trans., ASSP, Vol.37, pp.1888-1897, May, 1989.

[10] K. F. Lee and H. W. Hon, "Speaker-independent phoneme recognition using hidden Markov models," IEEE Trans., ASSP, pp.1641-1648, Nov. 1989.

[11] R. Schwarz, et al., "Robust smoothing methods for discrete hidden Markov models," Proc. of Int. Conf. ASSP. pp. 548-551, 1989.



장 춘 서

e-mail : csjang@kumoh.ac.kr

1978년 서울대학교 전자공학과 졸업(학사)

1981년 한국과학기술원 전기및전자공학과 졸업(공학석사)

1993년 한국과학기술원 전기및전자공학과 졸업(공학박사)

1981년~현재 금오공과대학교 컴퓨터 공학부 교수

관심분야 : 패턴인식, 음성인식, 신경회로망