

신경망을 이용한 음성인식

포항공과대학 정 흥*

● 목	차 ●
I. 서 론	VII. TDNN-LR-DP
II. 다층신경망	VIII. HMM 네트워크
III. 시간지연신경망	IX. Viterbi 네트워크
IV. TFM	X. 토의사항
V. 예측신경망	XI. 결 론
VI. LVQ	

I. 서 론

현재 가장 뛰어난 음성인식 알고리즘은 은닉 마코프 모델(Hidden Markov Model, HMM)이다. 그러면 이 알고리즘과는 전혀 다른 방법론인 신경망을 도입하려는 이유는 무엇인가?

우선 HMM의 장점을 생각해 보면 Viterbi 알고리즘이라는 매우 빠른 디코딩 방법이 있고 Forward-backward 알고리즘이라는 뛰어난 지도학습 방법이 있다는 것이다. 그러나 HMM의 한계는 음성신호모델과 언어모델을 정확히 구현할 수 없다는 데 있다. 음성신호모델이 정확하지 않으면 음향적으로 유사한 단어들을 구분할 수 없고 언어모델이 정확하지 않으면 간단한 유한상태(Finite state) 또는 확률적 문법(Probabilistic grammar)으로 구성된 언어만 인식할 수 있을 뿐이다. 또한 음성신호를 1차 마코프 모델을 사용하므로써 단어나 서브워드의 복잡한 상태전이 현상을 정확히 나타낼 수가 없어 여러가지 동화(Coarticulation)현상을 처리할 수가 없는 것이다.

아울러 HMM은 하드웨어 구현이 어렵고 계산량이 많으며 메모리를 많이 요구 한다. 현재 HMM을 HMM네트워크라는 등가의 신경망으로 나타낼 수 있다는 것이 증명되어 있다. 이런 점을 살펴 보건대 궁극적으로는 신경망이 HMM의 한계를 뛰어 넘어 음성인식의 문제를 근본적으로 해결해 줄 수 있을 것으로 보고 있다.

신경망을 이용한 음성인식 시스템은 여러가지 관점에서 분류할 수 있겠다. 인식기능에 따라 분류하려면 고립단어 연결단어 연속음성, 화자 종속 화자적용 화자독립, 사용된 문법, 인식대상 언어 등 여러가지로 구분할 수 있을 것이다. 인식 시스템의 구성요소로 구분한다면 전처리(Preprocessing) 패턴매칭(pattern matching) 후처리(Postprocessing)로 나눌 수 있을 것이다. 한편 기본 인식 단위인 서브워드(Subword)로 구분하려면 음소(Phoneme) Diphone Triphone 반음절(Demisyllable) 음절(Syllable) 단어(Word) 등으로 구분할 수 있겠다. 사용되는 신경망의 동적 특성에 따라 분류하면 정적신경망 동적신경망 혼합방식으로 구분 된다. 기본인식 단위에 대한 분류는 Waibel[29]을 참조하는 것이 좋으며 사

* 중신회원

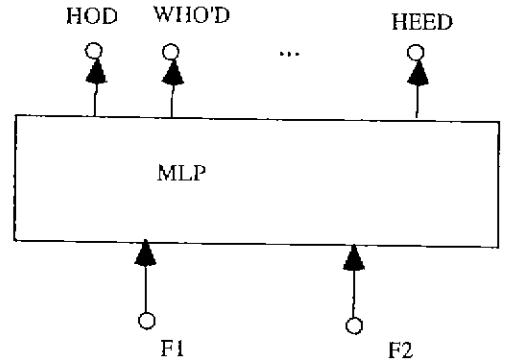
용되는 신경망에 대해서는 Lippmann[12]을 참조하면 된다. 이 논문에서는 중요한 몇개의 신경망을 선택해서 이들의 동작 원리를 알기 쉽게 소개하고자 한다.

음성인식을 위한 신경망은 정적분류 동적분류의 관점에서 구분할 수 있겠다. 정적분류라 함은 끝점찾기(Segmentation)가 완료된 입력 데이터를 분류하는 기능을 말한다. 즉 음소를 정확하게 인식하자면 각 음소마다 음성신호의 시작점과 끝점을 정밀하게 찾아 놓아야 하는 것이다. 이렇게 해서 정적분류에서는 시작점과 끝점 사이의 음성 패턴에 대한 음소를 결정한다. 이런 기능을 갖는 대표적인 신경망은 다층신경망(Multilayer Perceptron, MLP, Backpropagation network, BP)과 자율조직신경망(Self-organization Map, SOFM)이 있다.

분류하고자 하는 패턴의 끝점을 자동적으로 찾는 일은 대단히 어려운 문제이며 오차가 발행하기 쉽다. 엄밀히 이야기 하자면 이상적인 끝점 찾기는 인식이 완료된 음성에서나 가능하다. 따라서 닭과 달걀 중 무엇이 먼저인가를 말하는 것과 같다. 이 오차는 뒤 이을 단계에 혼란을 주어 큰 오차를 유발시킨다. 이상적인 방법은 패턴의 끝점 찾기와 패턴의 분류가 같이 처리되는 것이다. 동적신경망에서는 패턴에 대한 시불변(Shift invariant) 특성을 가지고 있어 패턴이 일정한 시구간 안에 존재만하면 패턴의 위치는 물론 패턴분류까지 수행한다. 이런 특성을 갖는 대표적인 신경망은 시간지연신경망(Time Delay Neural Network, TDNN), 회기신경망(Recurrent Neural Network, RNN), 학습벡터양자화(Learning Vector Quantization, LVQ) 등이 있다.

혼합 시스템은 신경망과 Dynamic programming에 기초한 Dynamic Time Warping(DTW) 또는 HMM 등의 알고리즘과 결합한 것이다. 이 방식에 의하면 신경망은 비선형 패턴분류기나 예측기로 보고 Dynamic programming이나 HMM은 음소를 단어로 조합하는 데 사용하려는 것이다. 그러나 점차 단어 이상의 언어처리에서도 신경망을 도입하고 있는 추세이다.

본 논문에서는 여러가지 신경망 중 많이 쓰이는 MLP, TDNN, TFM, 예측신경망, LVQ,



(그림 1) 다층신경망을 이용한 음소인식

TDNN-LR-DP, HMM 네트워크, Viterbi 네트워크에 대해서 설명한다.

II. 다층신경망(MLP)

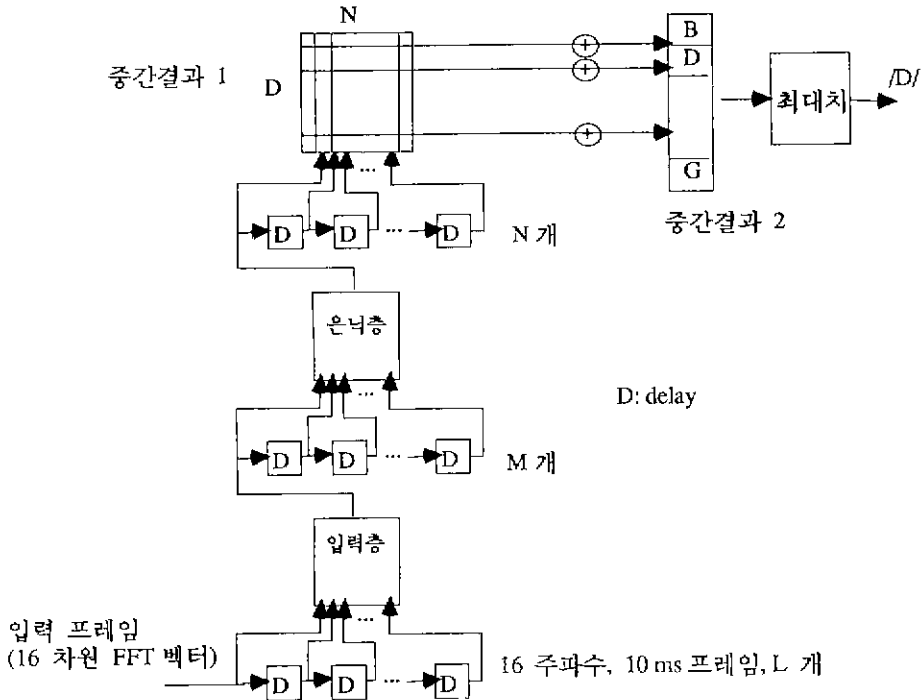
정적분류를 위한 대표적인 신경망은 다층신경망(Multilayer Perceptron, MLP)이다. 이 이외에 Kohonen[4]의 자율조직신경망(Self-organization Map), Kohonen[6,7]의 Learning Vector Quantization(LVQ), Niranjan[17]의 Radial basis function, Prager[18]의 Boltzmann machine이 있다. 여기서는 MLP에 대해서 설명한다.

Lippmann 등[11,2]은 1987년 끝점을 찾은 음성신호에 대해 MLP를 이용하여 음소를 구분해 보았다. (그림 1)과 같이 2개의 포먼트(Formant) F1 F2를 데이터로 이용해 /HOD/.../HEED/ 등 단어에 들어 있는 10개의 음소를 구분하려는 것이다.

다층구조신경망은 2층 구조이며 역전파(Back propagation, BP) 알고리즘으로 학습을 시킨다. 성능은 k-means 분류기와 비슷하나 파라미터를 필요로하지 않는 장점이 있다.

비슷한 방법을 써서 1987년 Elman은 3층 다층신경망을 이용해 푸리에 변환으로부터 /b, d, g/를 구분해 내는 데 성공했다.

그러나 MLP를 포함한 정적분류 방식은 끝점을 정확히 찾아주어야 하므로 지속시간이 일정치 않고 끝점이 불확실한 음성을 처리하는 데에는 근본적으로 한계가 있어 동적분류 신경망으로 연구가 진화해 왔다.



(그림 2) 시간지연신경망 구조

III. 시간지연신경망(TDNN)

시간지연신경망(Time Delay Neural Network, TDNN)은 1987년 Waibel[24,28]이 음소인식을 위한 신경망으로서 발표한 것이다. 그후 1988년 Bottou[1]는 TDNN을 이용해 단어인식을 시도했다. 현재는 단어 뿐만 아니라 각종 서브워드 인식에도 사용하고 있다. 신경망 중 가장 많이 연구되고 있어 매년 수 많은 변종이 발표되고 있는 상황이다.

Waibel의 TDNN은 음성신호 내에서 음소의 위치와 길이가 일정 범위 내에서 변화더라도 안정적으로 추출 할 수 있도록 고안한 것이다. TDNN의 구조를(그림 2)에 보인다. 음성신호는 푸리에 변환과 멜 스케일 변환을 거쳐 푸리에 벡터가 된다. 첫번째 층인 입력뉴런층은 과거에 들어온 L개의 푸리에 벡터를 받아 1개의 출력 벡터를 계산해 낸다. 이 출력벡터는 다시 M개가 조합되어 은닉층에 입력된다. 편의상 은닉층의 출력을 D차원 벡터라 하자. 이제 N개의 출력 벡터가 모여 $D \times N$ 매트릭스를 구성해 중간결과

1을 형성한다. 이 매트릭스의 행성분들이 더해져 D차원 벡터인 중간결과 2가 만들어 진다. 결국 이 벡터의 성분 중 가장 큰 값을 갖는 것이 출력이 된다.

각 층은 고유한 기능을 갖고 있다. 입력층은 무성음, Vowel onset, 포먼트의 상승 또는 하강 패턴 등 여러가지 중요한 음성학적 특징들을 검출해 내는 것이 알려져 있다. 은닉층은 음소에 대한 확률분포를 계산한다. 중간결과 1과 출력 사이의 합과 최대치 계산은 Shift invariance 특성을 구현하기 위한 것이다. 결론적으로 말하면 이 시간지연신경망은 푸리에 벡터를 관측치로 보고 음소를 파라미터로 하는 Bayes classifier로서 후행확률(Posterior probability)분포를 계산한다. 이 후행확률 분포는 많은 Bayes 추정문제에 필요한 중요한 값이다.

TDNN이 MLP를 빌딩블록으로 하여 구성한 것이지만 이 TDNN을 다시 빌딩블록으로 하여 이용하여 수 많은 변종을 만들어 낼 수 있다. 가장 대표적인 방법은 전체 네트워크를 계층적으로 구성하고 여러개의 병렬 모듈로 만들어 그 결과

를 한 모듈로 합하는 것과 입력신호 매트릭스를 시간축 프레임으로 하는 이나 주파수축 프레임으로 하는 것 등이 있다. 전자의 대표적인 것이 Waible[28]의 Modular TDNN이다. 이 경우 여러 병렬 모듈의 값을 융합해 이들 중 하나를 선택해 하나의 결과를 얻는 것은 쉽지 않다. 이것을 Merging 문제라고 할 수 있는데 Modular TDNN에서와 같이 Glue와 같은 신경망을 두어 해결하는 수도 있다. 후자에 대표적인 것은 Sawai [22]의 Frequency-Time-Shift-Invariant TDNN (FTDNN)와 Block-Windowed neural network (BWNN)이다. 아울러 Komori[8]의 Time-State Neural Network(TSNN)도 이 범주에 속한다고 할 수 있다.

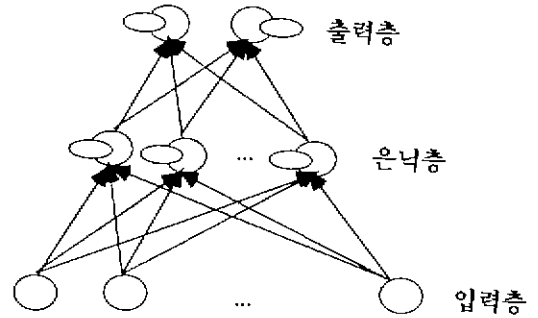
그러나 이 TDNN은 일종의 FIR필터로 볼 수 있다. 따라서 일정한 시간간격(그림에서는 10 L ms)의 시간적 변화량을 관찰하여 신호 수준에서의 동화 현상을 역이용한다. 이 신경망의 최대 단점은 입력시간 간격(그림에서는 10 L ms)이 고정되어 역동적으로 변화하는 서브워드를 처리하기 곤란하다는 것이다. 무성자음 같이 짧은 서브워드 입력시간 간격속에 여러개가 포함되어 판단에 혼란이 생길 것이고 모음 같이 긴 서브워드는 입력시간 간격속에 일부만 포함되어 역시 판단하기 곤란한 것이다.

이 신경망의 출력은 일반적으로 한개의 서브워드가 여러번 반복(Repetition)되어 나타나는 경향이 있다. 아울러 False alarm과 Missing에 의한 오차를 수반한다. 따라서 다음 단계에서는 반복, False alarm, Missing이 복합된 심볼로부터 안정적인 심볼시퀀스를 얻을 수 있어야겠다.

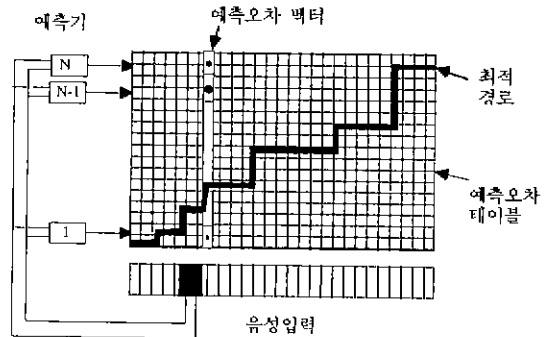
IV. Temporal flow model

회기신경망의 일종인 Temporal flow model은 Watrous[30]가 1988년도에 발표한 것이다. (그림 3)은 Watrous의 신경망을 나타낸다.

이 신경망은 3층이며 은닉층과 출력층은 단위 시간지연의 자기회환 회로(Self-loop)을 포함하고 있고 단위시간 지연된 패스가 하위층에서 고위층으로의 연결이 되어 있다. 학습은 역전파(BP)를 쓰나 학습될 패턴은 정적신경망에서와는 달리



(그림 3) Watrous 신경망



(그림 4) Iso-Watanabe 신경망 예측 모델

시간의 함수로 표현된 패턴이다. 따라서 이 신경망은 원하는 패턴의 시간적 변화를 기억하게 되어 있다.

V. 예측기능 신경망

신경망을 패턴분류기로 사용할 수도 있지만 신호예측기로 사용할 수도 있다. Iso와 Watanabe [3]는 1990년 이 MLP의 예측기능을 이용해 단어 인식을 시도했다. 이 아이디어는 (그림 4)와 같다.

이제 한 단어의 신경망 예측기가 있어 시간 $t-1$ 과 t 의 2 프레임의 음성 신호로부터 시간 $t+1$ 에서의 신호를 예측한다고 하자. 입력신호가 이 예측기에 해당되는 단어면 입력신호와 예측 결과와의 차이는 적을 것이고 아니면 클 것이다. 따라서 이 예측오차를 이용하여 예측기를 학습할 수 있다.

인식 단어 수 N 만큼의 예측기가 있다면 시간 $t-1$ 과 t 의 신호로부터 N 개의 예측 결과와 예측

오차를 발생할 것이다. 그림에서 이 예측오차 벡터를 볼 수 있는데 벡터성분의 크기 즉 오차의 크기를 검은 점의 크기로 나타내었다. 무슨 단 어인지 알려면 이 예측오차 벡터를 조사해 예측 오차가 가장 작은 예측기의 번호만 알면 될 것이다. 그러나 이 오차계산에도 오차를 동반하므로 전 시간에 대해서 오차의 변화를 조사해 결정해야 될 것이다. 이 오차 벡터를 시간의 함수로 나열한 것이(그림 4)의 테이블이다. 이 테이블에서 최소오차를 나타내는 최적경로를 찾기 위해서는 Dynamic Programming(DP)을 쓰면 된다. MLP로 구성된 신경망예측기를 학습하기 위해서는 지도학습 방법인 역전파 방식을 이용한다.

Levin[9,10]의 Hidden Control Neural network(HCNN)도 비슷하나 1개의 예측기를 사용하며 각 단어마다 다른 제어신호를 예측기의 입력으로 공급한다. 따라서 예측기의 입력단에는 음성신호와 함께 제어신호가 들어오게 된다.

1990년 Tebelskis와 Waibel[23]이 제안한 Linked Predictive Neural network(LPNN)은 Iso-Watanabe 모델을 단어인식으로 확장한 것이다.

VI. Learning Vector Quantization(LVQ)

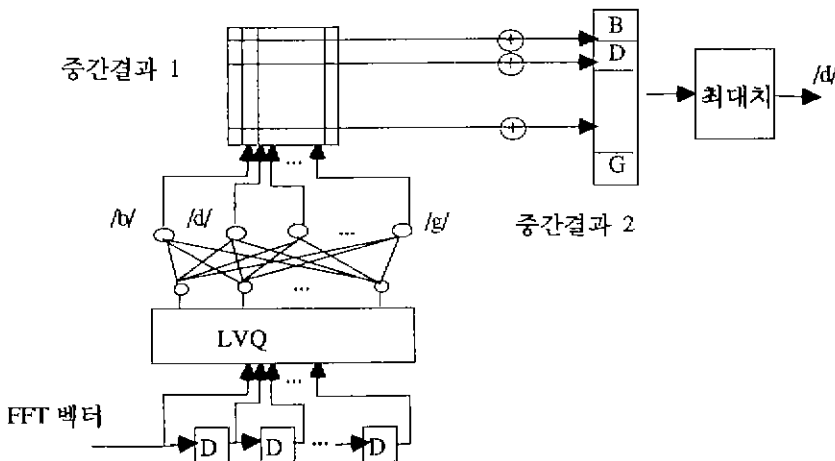
LVQ는 1988년 Kohonen[5,7]이 발표한 것으로서 자율학습 신경망인 Self-organization Feature Map(SOFM)을 지도학습 신경망으로 변환

한 것이다. 이 네트워크를 이용해 McDermott와 Katagiri[13,14]가 음소인식을 시도했다. (그림 5)에서 볼 수 있듯이 이 구조는 시간지연신경망에서 입력층과 은닉층을 LVQ로 대체한 것과 같다.

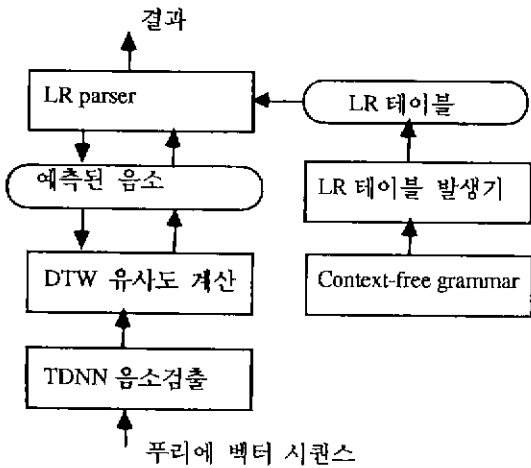
VII. TDNN-LR-DP

단어단위의 인식은 대규모 신경망과 긴 학습 시간을 요구해 이를 해결하기 위한 방안으로 단어를 구성하는 서브워드를 인식하고 서브워드 시퀀스를 다시 단어로 조합하는 2단계의 과정을 제시하게 된다.

Miyatake 등[31]은 1990년 일본어의 전 음소 이용 단어를 인식할 수 있는 시스템을 발표했다. (그림 6)은 기본원리를 보여준다. 이 시스템은 TDNN DP 파서(Parser)의 3부분으로 구성되어 있다. TDNN에서는 푸리에 벡터 시퀀스로부터 음소를 검출해 낸다. DP에서는 이 검출된 음소와 LR 파서에서 예측된 음소와 비교 유사도를 계산한다. LR 파서에서는 LR 테이블에 있는 단어로부터 음소를 예측해 낸다. 마지막으로 Context-free grammar와 LR 테이블 발생기는 LR 테이블을 발생한다. 이 방법을 이용 24개 음소를 검출해 5240단어인식 실험을 해서 92.6%의 인식을 얻었다.



(그림 5) McDermott 신경망

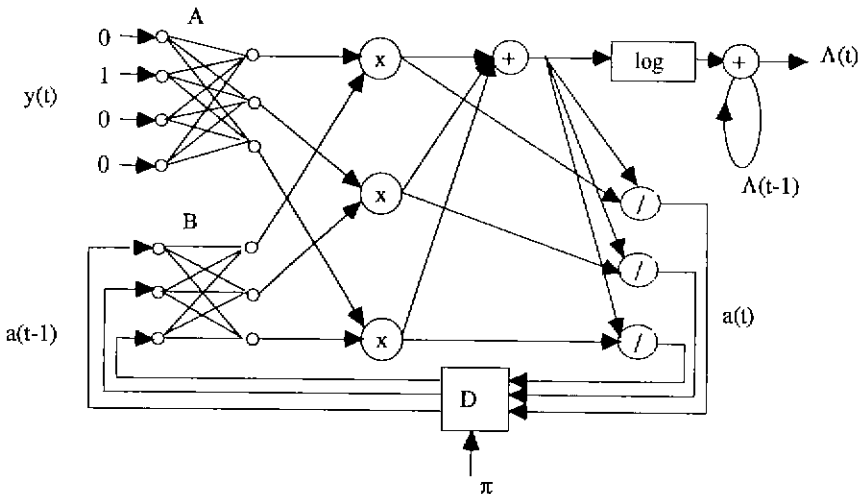


(그림 6) Miyatake의 TDNN-LR-DP 방법

VIII. HMM 네트워크

1990년 Niles와 Silverman[15]이 발표한 HMM 네트워크는 회기신경망으로서 Viterbi 네트워크와 마찬가지로 HMM의 구조와 기능을 수행한다.

(그림 7)은 이 HMM 네트워크를 나타낸다. 이 네트워크에서 신경망 A는 상태천이를 나타내고, 신경망 B는 상태와 관측데이터의 관계를 나타낸다. 입력은 초기확률 π 와 시간 t 에서의 관측 데이터인 $y(t)$ 이다. 중간결과 $a(t)$ 는 정규화된 전방향 확률(Normalized forward probability)이고



(그림 7) Niles-Silverman의 HMM 네트워크

출력은 관측데이터가 어느 특정한 단어일 확률이다. 여기서 상태천이 확률과 상태관측데이터 확률을 신경망으로 구현한 것을 알 수 있다. 이외에 덧셈 곱셈 나누기로 그 계산이 필요함을 알 수 있다.

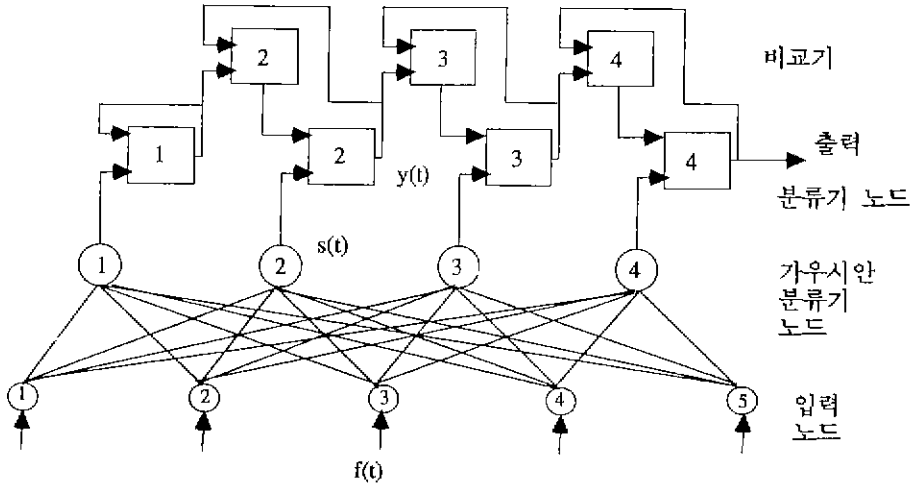
음성인식 때에는 각 단어마다 $(A B \pi)$ 이 다르므로 모든 단어에 대한 $\Lambda(t)$ 를 구하고 그 크기를 비교해 입력데이터에 대한 단어를 추정한다.

또 $(A B \pi)$ 를 정하기 위해서는 지도학습을 이용해 오차를 구하고 역전파 알고리즘(Backpropagation, BP)를 이용하는 데 이 알고리즘은 기존의 Baum-Welch 학습방법과 동일한 것이 증명되어 있다[6].

IX. Viterbi 네트워크

Viterbi 알고리즘은 주어진 관측데이터 시퀀스로부터 최적 상태 시퀀스를 결정하는 알고리즘이다. 여기서 관측데이터 시퀀스는 물론 음성 신호를 시간의 시퀀스로 나타낸 것이고 상태 시퀀스는 Markov 프로세스로 나타낸 음성발생 모델의 상태를 나타낸다. 이것은 HMM에서 많이 사용되는 중요한 알고리즘이다.

Lippmann과 Gold[11]는 1987년 Viterbi 알고리즘을 구현할 수 있는 신경망을 발표했다. 이 네트워크도 HMM 네트워크와 마찬가지로



(그림 8) Lippmann의 Viterbi 네트워크

HMM을 신경망으로 구현하는 데 사용될 수 있다. Viterbi 네트워크를 (그림 8)에 나타내었다.

이 네트워크는 크게 입력 노드, 가우시안 분류기 노드, 분류기 노드, 그리고 비교기로 구성되어 있음을 알 수 있다. 입력 $f(t)$ 는 이 그림에서는 5차원 벡터로서 시간 t 에서의 입력신호를 나타낸다. 가우시안 분류기 노드는 이 입력 데이터로부터 상태 $s(t)$ 의 확률을 계산해 낸다. 여기서 상태는 4개로 되어 있음을 알 수 있을 것이다. 분류기와 비교기의 구조는 4개의 상태로 된 LR 모델임을 알 수 있다. 분류기 노드는 곱하기 문턱치 계산(thresholding) 그리고 단위 시간 지연소자로 구성되어 있다. 비교기는 2개의 입력 중 큰 값을 취하는 역할을 한다. 따라서 분류기에 의해서 부분 스코어는 $y(t)$ 에 저장되어 오른쪽으로 전파되어 가고 4스텝 후에 출력으로 누적된 최적상태 값이 나오게 된다.

X. 토의사항

신경망을 연구하는 동안 몇가지 한계점을 발견하게 된다. 먼저 신경망 방법은 정성적인 연구가 되기 쉽다. 신경망을 마치 만능 블랙박스로 생각해서 음성신호와 원하는 단어를 보여 주고 학습을 시키기만 하면 된다는 비정량적인 사고 방식이다. 올바른 방법은 음성신호의 특성을 이해하고 이 특성을 어떻게 신경망에 이용해야할

것인가를 정하는 일이다. 이렇게 해야만 신경망의 각 부분에서 일어나는 모든 현상을 정량적으로 설명할 수 있고 음성인식 문제를 근본적으로 접근할 수 있게 될 것이다.

두번째는 음성인식신경망을 소프트웨어나 하드웨어로 구현하는 데 지대한 난관에 봉착하게 된다는 것이다. 계산량이 엄청나 100개 이상의 단어를 인식한다는 것은 불가능하고 2,000개 이상이나 되는 한국어의 모든 음절을 인식한다는 것은 더욱더 불가능하다. 만약 40여개의 음소를 인식하려 한다면 후처리단에서 엄청나게 일을 많이해야 할 것이다. 기타 다른 서브워드를 인식하려 한다면 마찬가지로 엄청나게 많은 계산량이 필요하다. 말하자면 신경망은 계산량이 엄청나게 많은 병렬 알고리즘이다. 이런 문제를 풀기 위해서는 속도가 엄청나게 빠른 컴퓨터나 신경망하드웨어를 전제로 한다. 만약 이 병렬 알고리즘을 직렬 컴퓨터로 풀려면 직렬 알고리즘보다 나을 것이 없거나 더 못 할 수 있다. 인식에 필요한 시간도 문제지만 학습시간은 더욱더 길어 Cray 슈퍼 컴퓨터로도 한달이 넘게 계산해야 될 지경이다. 그러면 큰 한계의 신경망 모듈 대신 작은 여러개의 신경망 모듈을 사용하면 될 것 같지만 이들의 결과를 종합할 때 심각한 문제가 발생한다. 근본적으로 해결하기 위해서는 인간의 음성인식 메카니즘과 마찬가지로 음성인식 문제를 여러 단계의 계층으로 나누어 각 단계마다

조금씩 계산하는 수 밖에 없을 것이나 인간의 음성인식 메카니즘에서 가장 중요한 대뇌 청각 피질의 기능을 우리는 아직도 모르고 있는 실정이다.

마지막으로 아주 중요한 관점은 신경망의 정보저장 용량이다. 음성인식을 포함한 제인식문제는 결국 최적화 문제로 귀착이 된다. 현재의 연구추세는 이 최적화 문제를 신경망의 안정상태(Equilibrium state)를 찾는 문제로 해석하는 데 있다. 신경망과 같은 동적 비선형 시스템에서 안정계는 비안정계에 비해 대단히 작은 범주에 속한다. 최적화 문제를 비안정계의 문제로 바꾸어 연구 한다면 아주 작은 크기의 신경망이라도 대단히 많은 정보를 포함하게 할 수 있을 것이다. 이 현상은 인간의 두뇌에서 잘 알 수 있다. 가장 드라마틱한 결과는 혼돈(Chaos)현상을 이용하는 것이다. 혼돈상태에는 엄청나게 많은 정보를 포함시킬 수 있다. 문제는 이 혼돈현상을 제어하여 안정적으로 정보를 넣고 회수할 방법이 아직도 잘 알려져 있지 않은 것이다[32].

다음으로 한국어 음성인식 연구와 관련해서 고려해야 될 사항이 있다. 한국어는 음소와 음절이 다같이 발달된 언어이다. 단지 40여개의 음소가 초성 중성 종성으로 조합되면 이론적으로는 10,000개 이상-실제로는 2,000~3,000개의 음절을 발음할 수 있다. 이와 반대로 음절언어인 일본어는 겨우 100개 정도의 음절로 구성되며 반음절(Demisyllable)도 240개 정도에 지나지 않는다. 이 한국어의 특성은 장점이 되기도 하고 단점이 되기도 한다. 음절을 인식하자면 50여개의 일본어 음절을 인식하는 것이 2,500여개의 한국어 음절을 이해하는 것 보다 50배 이상 쉽다. 그러나 음소를 인식한다면 한국어 음소 40여개를 인식하는 것이 훨씬 쉬운 것이다. 더우기 음절 구조가 규칙적이어서 음소시퀀스에서 음절시퀀스로 바꾸기는 아주 쉬운 것을 우리는 잘 알고 있다. (초성 중성 종성을 풀어 써도 잘 읽을 수 있다.) 따라서 한국어 음성인식은 음소를 먼저 인식하고 음절인식 단어인식 어절인식을 시도하는 방향으로 나아가야 할 것이다

XI. 결 론

본 논문에서는 음성인식을 위한 신경망에 대해서 몇가지 중요한 토픽을 소개하였다. 이 분야는 현재 연구활동이 대단히 활발하고 새로운 신경망들이 속속 발표되고 있어 그 추이를 따라가기가 용이하지 않은 것 같다. 그러나 이들 중 가장 대표적인 신경망으로서 MLP, TDNN, TDNN-LR-DR-DP, TFM, 예측신경망, LVQ, HMM 네트워크, Viterbi 네트워크 등에 대해 중점적으로 설명했다.

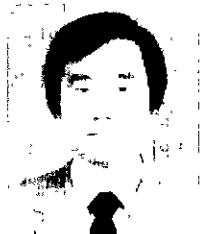
참 고 문 헌

1. L. Y. Bottou, "Reconnaissance de la Parole par Reseaux multi-couches," Proc. Neuro-Nimes 88, Nov. 1988.
2. W. Huang and R. Lippmann, "Neural net and traditional classifiers," in Neural Information Processing Systems, Edited by D. Anderson, 1988.
3. K. Iso and T. Watanabe, "Speaker-independent word recognition using a neural Prediction model," IEEE ICASSP, April 1990.
4. T. Kohonen, K. Makissra, and T. Saramaki, "Phonotopic maps-insightful representation of phonological features for speech recognition," 7th Int. Conf. on Patter Recognition, July 1984.
5. T. Kohonen, *Self-Organization and Associative Memory*. 2nd ed., Springer, 1988.
6. T. Kohonen, et al., "Phonetic typewriter for English and Japanese," IEEE ICASSP, April 1988.
7. T. Kohonen, et al., "Statistical pattern recognition with neural networks: Benchmarking studies," IEEE ICNN, July 1988.
8. Y. Komori, "Time-State Neural Network (TSNN) for phoneme identification by considering temporal structure of phoneme features," IEEE ICASSP, 1991.
9. E. Levin, "Speech recognition using hidden control neural network architecture," IEEE ICASSP, April 1990.
10. E. Levin, "Hidden control neural architecture modeling of nonlinear time varying systems and its application," IEEE Tran. on Neural Networks 4(1), 1993.
11. R. P. Lippmann and B. Gold, "Neural-net classi-

fiers useful for speech recognition," IEEE Int. Conf Neural Networks, June 1987.

12. R. P. Lippmann, "Review of Neural Networks for Speech Recognition," in Readings in Speech Recognition, Edited by A. Waibel & K. Kee. Morgan Kauffman, 1990.
13. E. McDermott and S. Katagiri, "Shift-invariant, multi-category phoneme recognition using Kohonen's LVQ2," IEEE ICASSP, May 1989.
14. E. McDermott, H. Iwamida, S. Katagiri, and Y. Tohkura "Shift-tolerent LVQ and hybrid LVQ-HMM for phoneme recognition," in Readings in Speech Recognition, Edited by A. Waibel and K. F. Lee, 1990.
15. L. T. Niles and H. F. Sivermqan, "Combining hidden Markov models and neural network classifiers." ICASSP, April 1990.
16. L. T. Niles, *Modelling and Learning in Speech Recognition: The Relationship Between Stochastic Pattern Classifiers and Neural Networks*, Ph. D. dissertation, Brown Univ., 1990.
17. M. Nijajan and F. Fallside, "Neural networks and radial basis functions in classifying static speech patterns," Technical report CUED/F-INFENG/TR-22, Cambridge University Engineering Department, 1988.
18. R. W. Prager, T. D. Harrison and F. Faliside, "Bolzmann machines for speech recognition," Comput. Speech Language **1**:3~27, March 1986.
19. H. Sakoe and A. Chiba, "Dynamic programming optimization for spoken word recognition," IEEE Trans. ASSP **26**(1), 1978.
20. H. Sakoe and A. Chiba, "Dynamic neural network-A new speech recognition model based on dynamic programming and neural network," IEICE Technical Report, Dec. 1987.
21. H. Sakoe, R. Isotani, K. Yoshida, K. Iso, and T. Watanabe, "Speaker-independent word recognition using dynamic programming neural networks," IEEE ICASSP, May 1989.
22. H. Sawai, "Frequency-Time Shift-Invariant Time-Delay Neural Networks for robust continuous speech recognition," IEEE ICASSP, 1991.
23. J. Tebelskis and A. Waibel, "Large vocabulary recognition using linked predictive neural networks," IEEE ICASSP, April 1990.
24. A. Waibel, T. Hanazawa, G. Hinton, K. Shukano, and K. Lang, *Phonemic Recognition Using Time-delay Neural Networks*, TR-1-0006, ATR Interpreting Telephony Research Laboratories. Oct. 1987.
25. A. Waibel, "Consonat recognition by modular construction of large phonemic time-dely neural network," in Advances in Neural Network Information Processing Systems, Edeted by D. S. Touretzky, 1989.
26. A. Waibel, "Modular construction of time-delay neural networks for speech recognition," Neural Computation **1**:39~46. 1989.
27. A. Waibel, T. Hanazawa, G. Hinton, K. Shukano, and K. Lang, "Phonemic recognition using time-delay neural networks," IEEE ASSP, **37**:328~339, March 1989.
28. A. Waibel, H.Sawai, and K. Shukano, "Modularity and scaling in large phonemic neural networks," IEEE ASSP **37**:1888~1898, Dec. 1989.
29. A. Waibel, "Neural network approaches for speech recognition," in Advances in Speech Signal Processing, Edited by S. Furui and M. M. Sondhi, Marcel Dekker, 1992.
30. R. Watrous, *Speech Recognition Using Connectonist Networks*, Ph. D. thesis, Univ. of Pennsylvania, 1988.
31. M. Miyatake, H. Sawai, Y. Minami, and K. Shikano, "Integrated traning for spotting Japanese phonemes using large phonemic time-delay neural networks. " IEEE ICASSP, 1990.
32. K. M. Cuomo and A. V. Oppenheim, "Circuit Implementation of Synchronized Chaos with Applications to Communications," Physical Review Letters, **71**(1), 1993.

정 흥



1977 서울대학교 전기공학과 학사 졸업
 1979 과학원 전기 및 전자공학 석사 졸업
 1979 경북대학교 전자공학과 전임강사
 1984 MIT EECS 석사 졸업
 1986 MIT EECS 엔지니어 졸업
 1988 포항공대 전자전기 공학과 조교수
 MIT EECS 박사 졸업

관심분야 : 신호처리, 영상처리, 음성처리