신경망과 퍼지논리를 이용한 음소인식에 관한 연구

한정현° . 최두일
공주대학교 전기공학과

A Study on Phoneme Recognition using Neural Networks and Fuzzy logic

Jung-hyun Han . Doo-Il Choi
Department of Electrical Engineering Kongju National University

Abstract - This paper deals with study of Fast Speaker Adaptation Type Speech Recognition, and to analyze speech signal efficiently in time domain and time-frequency domain, utilizes SCONN[1] with Speech Signal Process suffices for Fast Speaker Adaptation Type Speech Recognition, and examined Speech Recognition to investigate adaptation of system, which has speech data input after speaker dependent recognition test.

1. 서 론

인간의 음성은 가장 간단한 정보전달의 수단으로 인간과 기계사이에 정보를 주고 받는 것도 음성으로 하는 것이 자연스럽다. 음성 인식을 위해서는 연속적인 음성 신호에서 음소, 음절, 단어 등의 경계점을 찾아내야 하는데 음성 신호는 시간적 변화가 매우 빠르기 때문에 정확한 주파수 특성을 추출하는데 어려움을 보인다.음성인식을 위한 신경망은 학습방법에 따라 지도학습과 경쟁학습 신경망으로 나눌 수 있다. 경쟁학습을 근간으로 하는 자기 구조화 신경망의 성능은 최적성과 배열성으로 평가될 수 있다. 1980년대에 제안된 여러 형태의 경쟁학습 모델에서는 최적성에만 관심을 가질 뿐 배열성이 전혀 고려되지 않았다[2]. 그 후 1988년 Kohonen이 최적성과 배열성이 모두 고려된 최초의 신경망을 제안하였으며 이를 SOFM이라 명명하였다[3]. 그러나 SOFM이 경계효과가 발생하기 때문에 최적성이 낮으며, 입력 벡터의 차원이 3차 이상이 되면 배열성이 무너지는 문제점이 발생하였으며, 1994년 최두일이 이러한 문제점을 근원적으로 제거한 새로운 구조의 신경망을 제안하여 자기 생성 및 구조화 신경 회로망 (Self-Creating and Organizing Neural Networks, SCONN)이라고 명명하였다[1]. SCONN은 최적성과 배열성을 매우 높은 수준으로 유지할 뿐 아니라 입력 환경이 변할 때 새로운 입력 환경에 매우 빠르게 수렴하는 특성을 보인다.

본 논문에서는 고속 화자 적응형 음성 인식에서 입력이 비정상적으로 변할 때 회로망이 새로운 입력에 적응해 나가지 못하는 현상(SPD)을 시간 영역과 시간-주파수 영역에서 음성 신호를 효과적으로 분석하기 위하여 자기 생성 및 구조화 신경 회로망(SCONN)을 이용하여 시스템이 새로운 화자에 얼마나 빠르게 적응하는가를 보인다. 음성 신호는 연구실의 일반 잡음 환경 아래에서 두 명의 화자로부터 취득하였다.

2. 신경망과 퍼지논리를 이용한 음소인식

2.1 신경망의 구성

음소 인식을 위한 신경회로망은 3층 신경망으로서 그림 1과 같은 구조를 갖는다[2]. 여기서 1-2층간은 경쟁학습을 근간으로 하는 SCONN을 이용하여 학습하였으며, 2-3층간은 퍼지 논리를 이용한 지도 학습을 이용하였다.

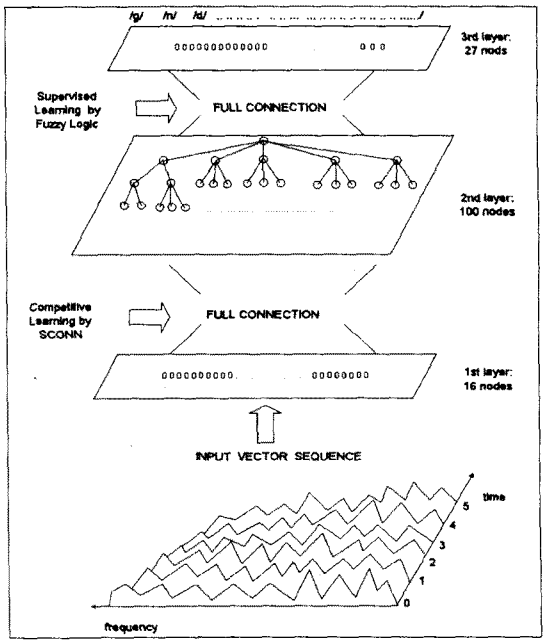

그림 1 신경망의 구성

2.2 SCONN

실험에 이용된 SCONN의 알고리즘은 다음과 같다.

Step 1. Initialize Weights
Step 2. Present New Input
Step 3. Calculate Distance to All Node(s)
Step 4. Find Active Node(s) and a Winner Node
Step 5. If Active Node Does not Exist, then go to step 8
Step 6. Decrease Response Ranges of Active Node(s)
Increase Response Ranges of Inactive Node(s)
Step 7. Adapt Weights of Winner Node
(or Winner node and its family nodes)
go to Step 2
Step 8. Creat a Son Node from an Inactive Winner (Mother) Node
go to Step 2

음성 신호는 화자의 성별, 연령 등에 따라 확률분포가 다르게 나타나며, 심지어 동일 화자의 경우에도 컨디션에 따라 서로 다른 분포를 보이기도 한다. 따라서 화자 적응 음성 인식 시스템을 구성하는데 있어서 중요한 점은 특정화자에 적응된 시스템이 새로운 화자에 얼마나 빠르고 정확하게 재 적응해 나가느냐 하는 것이라 볼 수 있다. 일반적으로 경쟁학습 신경망은 입력 환경이 갑자기 바뀐 경우 새로운 입력 환경에 적응해가지 못하는 현상을 보인다.

SCONN 알고리즘의 7단계를 보면 하나의 승리 노드만이 학습된다. 이를 변형하여 활성 노드들을 모두 학습함으로써 Stability and Plasticity dilemma를 해결할 수 있다. SCONN 알고리즘에서는 경쟁에서 진 노드들의 응답범위가 증가하므로 죽어있던 노드들도 언젠가는 활성 노드가 되어 학습 받을 수 있게 된다. 활성 노드가 되면 응답범위가 감소하기 때문에 오버플로우의 발생 위험은 전혀 없다.

실험 조건은 위의 두 경우와 같도록 설정하였으며, 승리 노드의 학습율은 0.18, 활성 노드의 학습율은 0.02로 한 경우이다.

2.3 퍼지논리에 의한 지도 학습 신경망

경쟁학습 신경망에서 출력노드는 자율적으로 형성되므로 학습이 어느 정도 진행된 후에 출력 노드에 대한 라벨링(labelling) 과정이 필요하게 된다[2]. 그러나 이러한 라벨링 과정은 생리학적인 리얼리즘이 결여되어 있을 뿐 아니라, 실제로 음소들간의 확률분포상의 겹침(overlap)현상 때문에 어떤 한 노드가 항상 동일한 음소에 대해서만 반응하지 않으므로 라벨링의 어려움이 남는다. 이러한 문제를 해결하기 위한 방안으로 2-3층간의 신경망을 그림 2와 같이 구성하였다. 여기서의 학습은 hebbian 학습법에 의한 지도학습으로 식 (1)과 같이 나타낼 수 있다.

$$\begin{aligned} \text{If } O_j &= \text{active and } O_k \\ &= \text{active then } W_{jk}(t+1) \\ &= W_{jk}(t) + \triangle \end{aligned} \qquad (1)$$

3층의 출력노드의 특성은 입력 벡터의 시간적 변화를 고려하기 위하여 시정수 (time constant)를 갖도록 하면

$$O_k(t) = F\left[\int_0^{\infty} \sum_j G(\tau) W_{jk} I_j(t-\tau) d\tau\right] \qquad (2)$$

이고, 이산시간의 경우는

$$\begin{aligned} O_k(t) &= F\left[\sum_j \sum_{\tau=0}^{\infty} W_{jk} G(\tau) I_j(t-\tau)\right] \\ &= F[\sum_j W_{jk}\{G(0) I_j(t) + G(1) I_j(t-1) \qquad (3) \\ &\quad + G(2) I_j(t-2) \cdots]\} \end{aligned}$$

로 나타낼 수 있다. 여기서 $O_k(t)$는 k번째 출력노드의 출력, W_{jk}는 가중치 벡터, I(t)는 입력, F는 활성화 함수 (activation function)이고, G(t)는 시간의 창함수 (time window function) 이다.

이때, 가중치 벡터 W_{jk}는 SCONN의 출력 노드의 각 음소에 대한 membership 함수로 대치될 수 있다.살펴본 바와 같이 hebbian 학습법은 특정 음소에 대하여 2

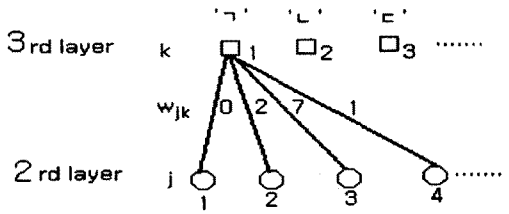

그림 2 퍼지 논리에 의한 지도 학습 신경망

층 노드들과 3층 노드들 사이의 연관성이 강할수록 가중치 값이 커지게 된다. 예를 들어, 만약 초기 가중치 값 $W_{jk} = 0$ for all j,k 이고 $\triangle = 1$ 이고 'ㄱ'을 10번 입력한 후 얻어진 W_{jk}가 그림 3과 같다고 하자.

2nd layer의 첫 번째 노드는 음소 'ㄱ'에 대하여 한 번도 응답하지 않았으므로 $W_{11} = 0$ 이고 반면 세 번째 노드는 'ㄱ'에 대하여 7 번 승리 노드가 되었으므로 $W_{31} = 7$ 이 되었다고 할 수 있다. 따라서 첫 번째 노드의 'ㄱ'에 대한 membership grade는 0 이고 세 번째 노드의 'ㄱ'에 대한 membership grade는 0.7 이라 할 수 있다. 이를 식으로 표시하면

$$M_j(\text{음소}) = \frac{j\text{노드의 승리 횟수(음소)}}{\text{총 입력수(음소)}} \qquad (4)$$

와 같고 여기서 M_j(음소)는 2nd layer의 j번째 노드의 특정 음소에 대한 fuzzy membership grade로써 W_{jk}의 정규화된 값이라 할 수 있다. 따라서 본 연구에서는 2층과 3층간에 hebbian학습을 수행하지 않고 대신 계산이 간단한 membership grade 값을 이용하였다.

2.4. 전처리 과정

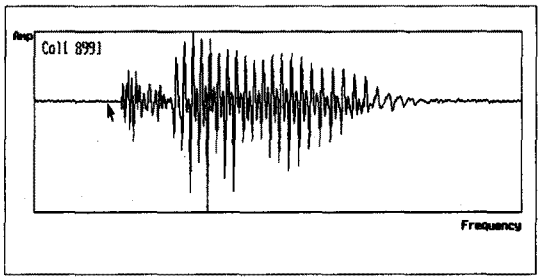

그림 3 /가/의 원 신호

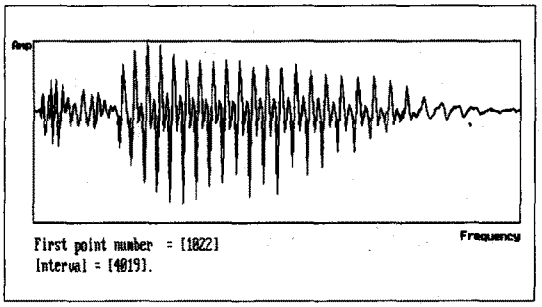

그림 4 Segmentation된 /가/의 신호

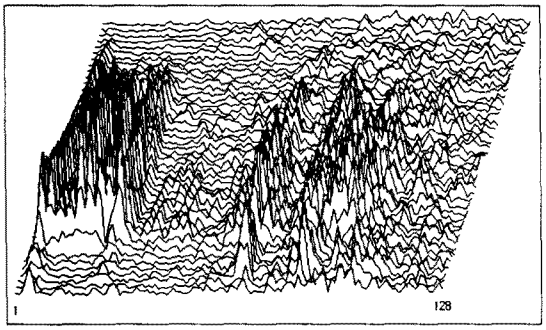

그림 5 z 축과 x 축 사이의 각도를 70도로 했을 경우의 /가/에 대한 스팩트럼

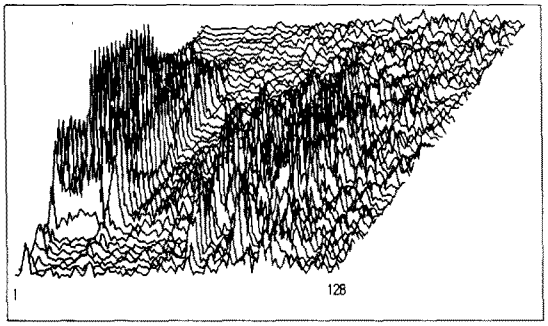

그림 6 z 축과 x 축 사이의 각도를 45도로 했을 경우의 /가/에 대한 스팩트럼

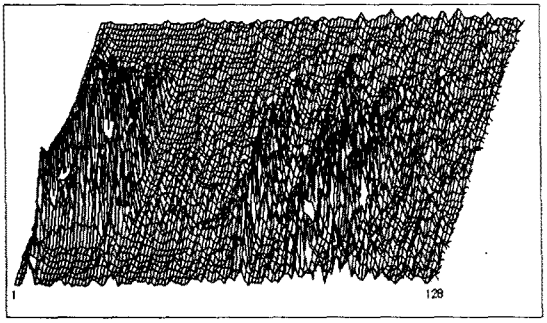

그림 7 봉우리들을 연결한 /가/의 스팩트럼

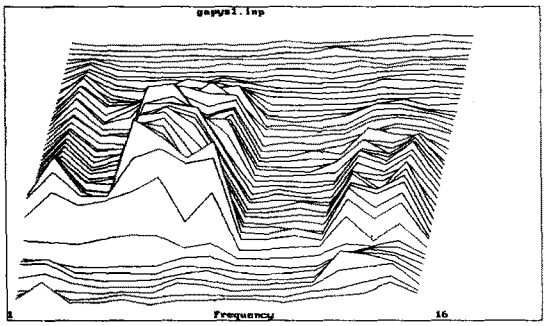

그림 8 bark scale 로 압축된 /가/ 의 스펙트로그램

2.5 음소인식 실험

실험 결과 화자 A의 각 음소에 대한 종속인식율은 80.56%를 나타내었다. 화자 A에 적응된 인식 시스템에 화자 B의 음성에 대한 독립인식율은 전체 음소에 대하여 27.78%의 저조한 인식율을 보이고 있다. 이는 화자마다 음성신호의 확률분포가 매우 다르게 나타남을 의미한다. 화자 A에 적응된 인식 시스템이 화자 B로 재적응된 후의 적응인식율은 67.13%를 나타내었다. 이는 신경망과 퍼지논리를 이용한 음소 인식이 많이 향상 되었다고 볼 수 있다. 시스템의 재적응 과정은 최두일이 제안한 SCONN에서 이루어지는데 바뀐 화자의 음성 데이터로 처음부터 다시 학습을 시키지 않고, 이미 화자 A의 음성으로 학습되어 있는 SCONN에 화자 B의 음성 데이터를 추가 학습하므로 써 학습속도를 매우 빠르게 하도록 하였다.

3. 결 론

본 연구에서는 고속 화자 적응형 음성에 관한 연구로서 시간 영역과 시간-주파수 영역에서 음성신호를 효과적으로 분석하고 고속 화자 적응형 음성인식에서 최두일이 제안한 모델인 자기생성 및 구조화 신경회로망(SCONN)의 최적성과 빠른 적응성을 이용하여 음소 인식실험을 수행하였다. 향후 음성 워드 프로세서의 구현을 위하여는 고속 화자 적응성이 필요하며, 인식 대상 음소도 한국어의 모든 음소를 포함하여야 한다. 본 연구의 실험 대상 음소의 개수는 대부분의 한국어 음소를 포함하도록 하였기 때문에 인식율은 다소 낮아졌으나 고속의 화자 적응성을 보이고 있다. 향후 최두일이 제안한 SCONN을 더 발전 시켜, 인식율의 증가를 위한 기술의 개발이 이루어지면 고차원의 음성 워드프로세서의 구현도 용이하리라 생각된다.

[참 고 문 헌]

[1] D. Choi and S. Park, "Self-Creating and Organizing Neural Networks," IEEE Trans. on Neural Networks, Vol. 5, No. 4., pp. 561-575, July 1994

[2] 최두일, "자기 생성 및 구조화 신경회로망의 기능향상 및 고속 화자 적응형 음성인식에의 응용", 전기학회논문지, 제46권 11호, pp.1684~1691, 1997.11

[3] T. Kohonen, Self Organization and Associative Memory, 2nd edition, Springer-Verlag, ch. 5, pp.119-157, 1988.