

論文 91-28B-5-8

Chip 구현을 위한 IDMLP 신경 회로망의 개발과 음성인식에 대한 응용

(The Development of IDMLP Neural Network for the Chip Implementation
and it's Application to Speech Recognition)

金 信 鎭*, 朴 政 運*, 鄭 鎬 宣**

(Sin Jin Kim, Jung Woon Park, and Ho Sun Chung)

要 約

본 논문에서는 chip 구현을 위한 IDMLP 신경회로망의 개발과 그 회로망을 이용한 우리말의 숫자음 인식에 관해서 연구하였다. 사용된 IDMLP 신경회로망 모델과 이 회로망을 학습하기 위한 규칙은 새롭게 제안된 것이며 이 모델에서는 가중치를 정수로, 뉴론에서의 전달 함수를 hard limit 함수로 하였다. 제안된 회로망에 여러가지 입력 데이터를 이용하여 학습한 결과 입력을 분류하는데 따른 난이도에 따라 하나의 층으로 입력 데이터를 모두 분류하기도 하고 여러 층으로 분류하기도 하였다.

0에서 9까지의 숫자음에서 이진화된 데이터를 추출하여 제안된 회로망으로 인식실험을 해 보았다. 그 결과, 학습 데이터에 의해서 100%, 시험 데이터에 대해서 96%의 인식결과를 얻었다.

Abstract

This paper described the development of input driven multilayer perceptron(IDMLP) neural network and it's application to the Korean spoken digit recognition. The IDMLP neural network used here and the learning algorithm for this network was proposed newly. In this model, weight value is integer and transfer function in the neuron is hard limit function.

According to the result of the network learning for the some kinds of input data, the number of network layers is one or more by the difficulties of classifying the inputs.

We tested the recognition of binaried data for the spoken digit 0 to 9 by means of the proposed network. The experimental results are 100% and 96% for the learning data and test data, respectively.

*準會員, 慶北大學校 電子工學科
(Dept. of Elec. Eng., Kyungpook Univ.)

**正會員, 慶北大學校 電子工學科
(Dept. of Elec. Eng., Kyungpook Univ.)

接受日字: 1990年 12月 26日

(※ 본 논문은 1990년도 한국과학재단 기초연구비에 의해 수행되었음.)

I. 서 론

최근 활발히 연구되고 있는 신경회로망은 그 응용 분야가 넓으며 기존의 이론으로 구현하기 힘든 여러 가지 기능들을 문제에 대한 정확한 수학적인 모델링이나 해결을 위한 구체적인 알고리즘 없이도 그 기능을 실현할 수 있다.^[1] 이러한 것들은 인간의 뇌가

대량의 정보 처리를 효율적으로 하고 학습의 능력이 가진다는 것을 근거로 하여 제안되었다. 그런데 이러한 신경회로망의 특성들을 실제로 문제에 적용하고 충분히 이용하기 위해서는 칩으로의 구현이 반드시 필요하다.^{2,3,4)}

칩 구현시에는 컴퓨터에 의한 소프트웨어 시뮬레이션과는 달리 많은 제약이 따른다. 이것은 일반적인 소프트웨어 시뮬레이션에서는 가중치로 floating point의 실수값을, 시냅스의 전달 함수로는 sigmoid 함수를 이용하여 node와 connection을 충분히 할 수 있지만 현재의 VLSI 기술로는 이러한 것들을 칩화하는데 어려움이 많기 때문이다.

본 논문에서는 신경회로망이 가지는 고도의 병렬 처리 기능과 학습기능을 이용하면서 이것을 하드웨어로 구현하는 것에 관하여 연구하였다. 이를 위한 신경회로망의 구조로 IDMLP(input driven multilayer perceptron)를 제안하고, 이 회로망의 학습을 위하여 은닉층의 값을 예측하고 단층의 학습에 기초하여 전체 회로망을 학습하며 학습 데이터의 난이도에 따라 회로망의 층수를 가변하는 새로운 학습 알고리즘을 제안하였다. 제안된 회로망은 칩의 구현을 위하여 뉴런의 전달함수를 hard limit 함수로하고 시냅스의 가중치들을 정수로하여 MOS를 이용한 칩화^{5,6,7,8)}가 가능하도록 하였다.

입력과 출력은 칩 구현과 기존의 논리회로와의 정보전달을 간단하게 하기 위하여 이진화된 값을 이용하였다. 또한 한개의 시냅스 내에 두 가지의 가중치를 갖도록 하고 각각은 1과 0의 입력에 의해 이용하도록 하였다. 다음으로 제안된 회로망과 학습규칙을 이용하여 한국어 숫자를 인식을 행하였다. 학습과 인식을 위한 입력 데이터로는, 필터뱅크로 A/D변환된 입력음성의 주파수성분을 분할^{9,10)}하고 이웃하는 필터와의 차를 이진화한 것을 이용하였다. 전체 회로망의 구성시에는 연결의 수를 줄이고 학습의 속도를 증가시키기 위해 모듈화된 구성을 갖도록 하였다. 각각의 모듈은 하나의 필터출력에 대해 학습되고 이 결과들은 다시 최종 출력단의 모듈에 의해 학습되도록 하였다. 각 모듈의 층수는 이층으로 제한되고 여기에서의 완전하지 못한 출력은 최종 모듈에서 다시 학습되어 완전한 출력을 얻도록 하였다. 최종 출력단의 모듈에서는 출력을 10비트로 하고 입력에 대해 가장 높은 값을 갖는 하나의 비트가 1을 출력하도록 하였다.

II. 입력 구동 다층 인식자(Input Driven Multilayer Perceptron Model)

본 논문에서는 두개의 인버터를 이용한 뉴런과 MOS를 이용한 시냅스로 하드웨어 구현을 가능하게 하기 위해 가중치를 정수로, hard limit 함수를 뉴런에서의 전달함수로 이용하고 이진수의 입출력을 갖는 입력 구동 다층 인식자(input driven multilayer perceptron)를 제안하고 또한 이 회로망의 학습을 위하여 은닉층의 값을 예측하고 은닉층의 수가 학습의 진행에 따라 증가하는 새로운 학습 규칙을 제안하였다.

1. 입력 구동 다층 인식자의 구조

일반적인 다층 신경회로망은 그림 1과 같은 구조를 가지며 주로 역전파 학습 알고리즘(back propagation learning algorithm)을 이용하여 회로망의 학습을 수행한다. 이때 가중치로는 실수의 값을 이용하고 전달함수로는 sigmoid 함수를 이용한다. 그러나 이러한 구조의 회로망에서 이진화된 입력과 hardlimit 함수를 이용할 경우, 은닉층에서의 출력의 종류가 극층의 노드 수에 의해 제한되고, 노드의 수가 입력을 구분하기에 충분한 경우에도 출력 값들이 고르게 분포되어 모든 입력에 대하여 다른 값을 가지기가 힘들다. 이렇게 되면 상위 층에서는 이 값들을 이용하여 입력을 구분하는 것이 불가능하게 되어 회로망의 학습이 어렵게 된다. 이것을 해결하기 위해 각 층에서는 입력으로 앞의 층의 출력과 처음의 입력을 동시에 받아들인다. 그 결과 회로망의 구성은 입력이 모든 다른 노드에 연결되고 은닉층과 출력층은 앞단의 은닉층과 입력층과의 연결을 가지게 되어 은닉층에서 구분이 되지 않은 입력에 대해서도 다음 층에서는 구분이 가능하게 된다. 이렇게 입력층이 모든 층과의 연결을 가지도록 한 것이 입력구동 다층인식자의 가장 중요한 특징이라 할 수 있고 여기에 따라 학습시에는 입력층과 하위층의 입력에 대해 단층의 학습 규칙으로 하나 하나의 층을 학습한다는 것이 본 연구에서 사용된 학습 규칙의 특징이다.

제안된 입력구동 다층인식자는 이러한 특징에 따라 그림 2와 같은 구조로 이루어 질 수 있다. 이것을 그림 1의 일반적인 다층 인식자와 비교하면 입력에서 모든 층과 연결된 가중치들을 볼 수 있다. 그림 3에서는 IDMLP의 하드웨어 구현을 위한 회로망의 구성을 나타내며 이 그림에서 점선내의 부분이 IDMLP의 특징을 나타내는 부분이라 할 수 있다. 이 회로에서 각각의 시냅스는 두가지의 연결을 가지며 각 연결은 1 또는 0의 입력에 의해 이용되며 그림 4에 하나의 시냅스에 대한 회로를 나타내었다.

뉴런에서의 전달 함수인 Hard-limit 함수를 buffer

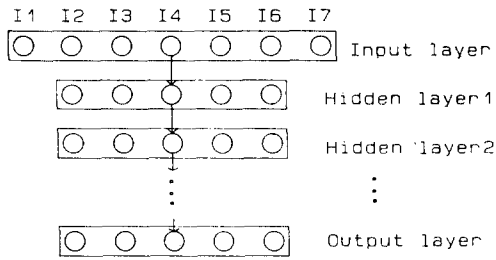


그림 1. 다층인식자
Fig. 1. Multilayer perception.

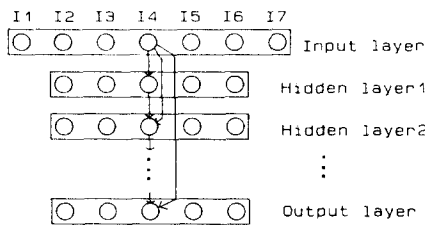


그림 2. 입력구동 다층인식자
Fig. 2. Input driven multilayer perceptron.

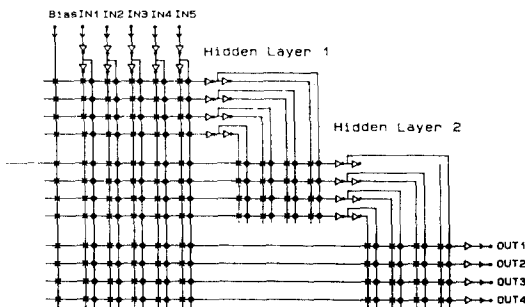


그림 3. Hardware 구현을 위한 IDMLP의 구조
Fig. 3. Structure of IDMLP for hardware implementation.

를 이용하여 하드웨어로 구현할 경우에 입력에 의해 ON 되는 PMOS의 합과 NMOS의 합이 크고 두 합이 차가 적을 때는 출력이 "1"인지 "0"인지 구별하기가 어려워진다. 그래서 어떤 입력에 대해 한꺼번에 ON 되는 가중치의 합이 적어야 한다. 그리고 신경 회로망을 chip화 할때는 입력의 node 수를 적게 하는 것이 좋다. 입력 bit가 많아질 경우 chip의 pin 수가 많아져야 한다는 이유 외에도 입력에 의해 ON

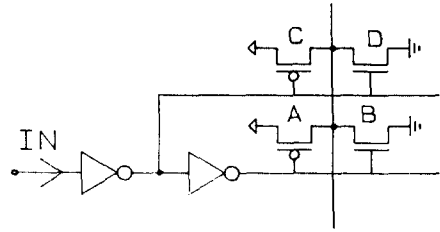


그림 4. 시냅스의 회로
Fig. 4. Circuit of synapse.

되는 MOS의 수가 많아지면 뉴런에서의 출력값 결정이 어려워지기 때문이다.

그림 4에서 입력 단자의 값에 따라서 A, B, C, D의 MOS 중 두개가 ON된다. 1의 입력에 대해서는 B와 C의 MOS가 ON되고 0의 입력에 대해서는 A와 D의 MOS가 ON된다. 따라서 1의 입력에 대한 양의 가중치를 위한 경우 B의 MOS에 값을 할당하고 1에 의한 음의 가중치는 C에, 또 0에 의한 양의 가중치는 A에, 0에 의한 음의 가중치는 D에 각각 할당한다. 이렇게 시냅스를 구성하면 1과 0의 입력에 대해 각각 다른 가중치를 가지게 되고 하나의 시냅스 내에는 두개의 MOS가 사용된다.

2. 학습 알고리즘

IDMLP 신경회로망의 학습 알고리즘은 앞 절에서 언급한 구조의 회로망을 위하여 제안된 새로운 학습 규칙이며 아래와 같은 순서로 이루어진다.

- [단계 1] 모든 노드간의 가중치를 초기화 한다.
- [단계 2] 입력과 이에 대응하는 출력의 쌍을 임출력에 제시한다.
- [단계 3] 각 노드에서 입력의 가중치 합을 구하고 hard limit 비선형 함수에 의해 출력을 발생한다.

$$f_n(X) = \begin{cases} 1 : X > 0 \\ -1 : X < 0 \end{cases} \quad (1)$$

$$OUT_j = 1/2(f_n(\sum_{i=0}^{N-1} W_{ji}X_i) + 1) \quad (2)$$

- f_n : Hard-limit 함수
- X_i : i번째 입력
- W_{ji} : j번째 노드에서 i번째 노드로 연결된 가중치
- OUT_j : j번째 노드의 출력값

- [단계 4] 출력 노드에서의 출력을 원하는 출력값과 비교하여 오차를 계산하고 오차 값에 따른

가중치의 변화분을 저장한다.

$$\Delta W_{ji} = \pm (D_j - O_j) \quad (3)$$

ΔW_{ji} : 가중치의 변화량

D_j : j번째 노드의 원하는 출력값

O_j : j번째 노드의 출력값

[단계 5] 단계 2에서 단계 4까지의 과정을 모든 입력에 대해 수행하고 모든 출력 값이 원하는 값과 같으면 학습을 끝내고 그렇지 않으면 이때의 가중치 변화분의 합을 각각의 가중치에 더한다.

$$W_{Tji} = W_{T-1ji} + \sum_{i=0}^{N-1} \Delta W_{ji} \quad (4)$$

W_{T-1ji} : 변경되기 전의 가중치

W_{Tji} : 변경된 후의 가중치

[단계 6] 가중치의 합이 정해진 값(M) 이상일 때는 일정비율로 가중치들의 값을 줄인다.

$$\text{new weight} = W_{Tji} * (M / \sum_{i=0}^{N-1} W_{Tji}) \quad (5)$$

[단계 7] 일정 횟수 반복후 원하는 결과가 나오지 않을 때는 층을 증가시킨 후 앞의 층에서의 출력과 원래의 입력을 새로운 입력으로 하여 단계 2부터 반복한다.

앞의 학습 단계중에서 단계 1의 가중치의 초기값은 error back propagation 알고리즘에 의한 학습에서와 같이 random 값을 이용하지 않고 모든 초기치를 동일한 값으로 정하고 학습을 시작한다. 이것은 IDMLP 신경회로망이 단층의 학습 규칙에 의해 모든 층이 학습되기 때문에 동일한 초기치에 의해서도 학습이 가능하기 때문이다. 본 연구에서는 모든 가중치를 "1"로 초기화하였다.

단계 3에서는 칩 구현을 위하여 뉴런의 전달 함수로서 Hard-limit 함수를 이용하였다. 사용된 Hard-limit 함수는 식(1), (2)와 같이 나타낼 수 있다.

가중치의 변화분은 단계 4의 식(3)에서 보인 바와 같이 "1", "0" 또는 "-1"중의 하나이다. 이렇게 한 이유는 가중치들을 정수값으로 하여 CMOS로의 구현을 쉽게 하자는 것이다.

단계 5에서는 가중치를 변화시키는 방법을 보이고 있다. 다른 일반적인 학습 알고리즘에서 처럼 한가지의 입력이 가해질때 마다 가중치를 변화시키는 것이 아니라, 단계 4에서 처럼 각각의 입력에 대해 구한 가중치 변화량을 저장해 두었다가 모든 입력을 적용시킨 후 변화량의 합을 원래의 가중치에 더해

행해진다.

단계 6에서는 한 개의 뉴런이 비교해야할 가중치의 합이 제한된 값(M)보다 클 경우에는 식(5)를 적용하여 가중치의 합을 M이내로 제한한다. 이렇게 하는 이유는 앞서서도 설명했듯이 buffer를 이용하여 하드웨어로 구현한 뉴런은 입력에 의해서 ON되는 PMOS의 합과 NMOS의 합이 크고 그 차이가 적으면 출력이 "1"인지 "0"인지를 구별하기가 어려워지기 때문이다.

단계 7에 나타난 학습 횟수는 최적의 회수는 정확히 어느 정도인지는 확실히 알 수 없으나, 200회까지는 오차가 현저하게 감소하다가 그 이상의 학습에는 오차의 변화가 거의 없다는 것을 실험에 의해서 알 수 있었다. 일정회수 반복후에도 원하는 결과가 나오지 않을때에는 층을 하나 더 증가시켜서 그 증가된 층에 대해서 다시 위의 단계와 같이 학습을 일정 회수 만큼 반복시킨다. 추가된 새로운 층에서는 원래의 입력과 앞의 층에서 어느 정도 분류가 된 출력을 새로운 입력으로 하여 학습한다. 그러므로 본 IDMLP 신경회로망은 층의 수를 미리 정해 놓고 학습을 시키는 것이 아니라, 학습의 난이도에 따라 층의 수가 변하게 된다. 여러 층에 의해 학습된 경우 층이 증가할수록 오차가 감소함을 실험에 의해서 볼 수 있었다.

본 논문에서 제안된 IDMLP 신경회로망은 단층단위로 학습을 하므로, 각 층에서 선형적인 분류를 하게된다. 그러므로 선형 분리가 가능한 데이터에 대해서는 단층만으로도 학습이 완료되고, 복잡한 데이터에 대해서는 단층에서 학습이 완료되지 않더라도, 앞의 층에서 선형적인 분리가 된 데이터와 원래의 입력 데이터에 대해 증가된 층에서 계속적으로 학습을 수행하므로 결국에는 몇개의 증가된 층에서 학습이 완료된다.

은닉층에서 노드의 갯수 또한 학습 결과에 많은 영향을 미친다. 그러나 본 논문에서의 학습 알고리즘으로는 은닉층의 노드수가 출력층의 노드수와 동일하게 되고 이 때문에 은닉층의 노드수는 항상 출력층의 노드수와 동일한 갯수로 정해진다. 일반적인 다층인식자에게 은닉층의 노드수를 증가시켜 주는 것이 학습의 결과에 좋은 영향을 미치는 것으로 알려져 있다. IDMLP에서는 위에서 설명한 바와 같이 학습 알고리즘상 은닉층의 갯수를 임의로 정할 수가 없는데 이의 해결을 위하여 다음과 같은 방법을 이용하였다. 출력에 원하는 값을 출력하는 노드 외의 어떤 임의값을 출력하는 노드를 추가하면, 학습이 진행됨에 따라 출력층이 은닉층으로 역할이 바뀌었을

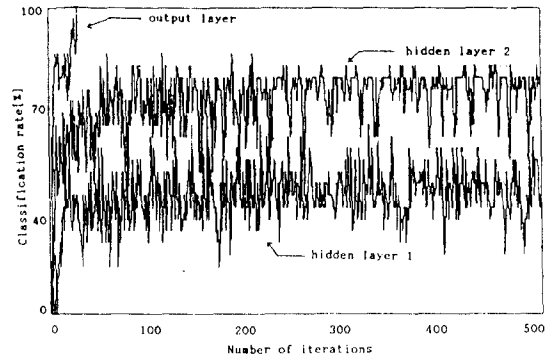
때 이 추가된 노드들은 은닉층으로서의 기능을 가지게 된다. 학습시에는 출력에 추가된 노드에 적당한 값을 정하여 학습하여야 한다. 이 값의 결정 또한 학습결과에 영향을 미치는데 본 연구에서는 입력값들을 AND, OR등의 논리연산을 행한 결과 값으로 정하거나 random 값을 이용하기도 하는데 입출력의 종류에 따라 어떤 값이 가장 좋은지는 알 수 없었다. 그림5에서 보는 바와 같이 이러한 값의 추가는 학습의 결과를 성공적으로 하는데 많은 기여를 한다. 입력에 추가되는 바이어스 입력은 전달함수의 임계치를 조정하여 줄 수 있으며 이것 또한 학습 결과에 좋은 영향을 미치는 것을 그림5에서 볼 수 있다. 그 외에도 회로망의 학습결과는 여러가지 요인들에 의해 영향을 받는다.

위에서와 같은 방법으로 회로망을 학습했을 때 각 층에서 입력 데이터에 대한 구분의 정도는 그림5에 나타나 있다. 회로망의 해석을 위하여 사용된 데이터는 영어 인쇄체 대문자에서 특징을 추출한 것이다. 회로망의 학습이 반복됨에 따라 오차가 감소하고, 또 층이 증가할수록 오차가 감소하는 것을 볼 수 있다. 회로망의 학습 과정에서 인식률의 변화가 심하게 나타나는데 이것은 전달 함수로 사용된 함수가 hard limit 함수인 것에서 기인하는 것이다. 본 실험에서의 입출력은 표1~3에 나타내었다. 표2는 출력에 임의의 노드를 추가한 경우의 입출력이고 출력에는 세개의 비트가 추가 되었다. 표3은 입력에 바이어스를 추가한 경우이고 입력에서 가장 오른쪽의 값들은 모두 1의 값이며 이것에 의한 가중치는 학습시에 바이어스로 작용한다.

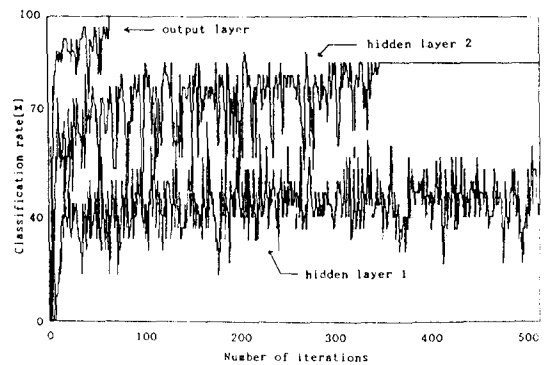
영문자의 특징 외에 여러가지 데이터에 대해 학습한 결과 학습을 위한 입출력 데이터에 따라 하나의 층에서 또는 몇개의 층에 의해 학습이 되었다. 학습한 결과 층이 증가할수록 오차가 줄어드는 것을 그림5에서 볼 수 있다. 출력 노드에 원하는 결과 값 외에 다른 값을 추가하였을 때의 결과를 그림5(b)에 나타내었고 그림5(c)는 입력에 바이어스를 추가한 경우이다. 표4는 가중치의 합을 제한하는 정도에 따른 학습 결과이다. 그 외에도 초기값, 입력과 출력의 조합, 학습 횟수, 회로망의 구성 학습시 방법 등도 학습의 결과에 영향을 미친다.

III. IDMLP 신경회로망을 이용한 음성인식

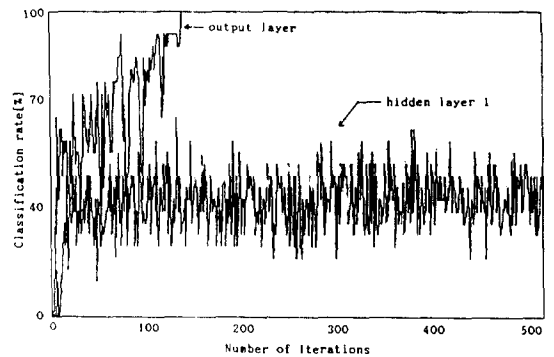
앞서 언급한 IDMLP 신경회로망을 이용하여 숫자 음에 대한 인식을 수었다. IDMLP 모델은 입력과 출력으로 디지털 값을 이용하고 있기 때문에 전처리



(a)



(b)



(c)

그림 5. 회로망의 학습 과정

- (a) 출력 노드와 은닉층의 노드수가 같은 경우의 학습 결과
- (b) 은닉층에 노드를 추가 하였을 경우의 결과
- (c) 입력에 바이어스를 추가 하였을 경우의 결과

Fig. 5. Learning process of network.

- (a) the number of output nodes and that of hidden nodes are equal,
- (b) nodes are appended to the hidden layer,
- (c) bias are appended to input layer.

표 1. 영어 대문자에 대한 학습 데이터
Table 1. Learning data for English capital letters.

Char.	Input	Output
A	0 1 0 0 1 1 0 0 0 0 1 0 0 0 0	0 0 0 0 1
B	0 0 1 0 1 1 0 0 0 0 0 0 0 1	0 0 0 1 0
C	0 0 0 0 0 0 0 0 0 0 0 1 0 0 1	0 0 0 1 1
D	0 1 0 0 1 0 0 0 1 0 0 1 0 0 0	0 0 1 0 0
E	0 0 1 0 0 0 0 0 1 0 1 0 0 0 0	0 0 1 0 1
F	0 1 1 0 1 1 0 0 0 1 0 0 0 0 0	0 0 1 1 0
G	0 1 1 0 1 1 0 0 0 0 1 1 0 0 0	0 0 1 1 1
H	0 1 0 0 1 1 0 0 0 0 1 1 0 0 1	0 1 0 0 0
I	0 0 1 0 1 0 0 0 1 0 0 1 0 1 0	0 1 0 0 1
J	0 0 1 0 1 0 0 1 0 1 0 0 0 0 0	0 1 0 1 0
K	0 1 0 0 1 0 0 0 1 0 0 1 0 0 1	0 1 0 1 1
L	0 0 1 0 1 0 0 0 0 1 0 0 0 1	0 1 1 0 0
M	0 1 1 0 1 1 0 0 0 1 0 0 0 0 1	0 1 1 0 1
N	0 1 1 0 1 0 0 1 0 0 1 1 0 0 1	0 1 1 1 0
O	0 0 0 0 0 0 0 0 0 0 0 1 0 0 0	0 1 1 1 1
P	0 1 0 0 1 0 0 0 0 0 1 0 0 0 0	1 0 0 0 0
Q	0 1 0 0 1 1 0 0 1 0 0 1 0 0 0	1 0 0 0 1
R	0 1 0 0 0 1 0 0 0 1 0 1 0 0 0	1 0 0 1 0
S	0 0 0 0 0 0 0 0 0 0 1 0 0 0 0	1 0 0 1 1
T	0 1 0 0 1 1 0 0 0 0 1 0 0 0 1	1 0 1 0 0
U	0 0 1 0 1 0 0 0 0 0 1 0 0 0 0	1 0 1 0 1
V	0 0 1 0 0 0 0 0 0 0 0 0 0 0 1	1 0 1 1 0
W	1 0 0 1 1 0 0 0 1 0 1 0 0 0 1	1 0 1 1 1
X	1 0 0 1 1 0 0 0 1 0 0 0 0 0 1	1 1 0 0 0
Y	0 1 0 0 1 1 0 0 1 0 1 0 0 0 1	1 1 0 0 1
Z	0 0 1 0 0 0 0 1 0 0 1 0 0 0 0	1 1 0 1 0

표 2. 출력에 임의의 값을 추가한 입출력 데이터
Table 2. I/O data appending arbitrary values to output.

Char.	Input	Output	
		original	added
A	0 1 0 0 1 1 0 0 0 0 1 0 0 0 0	0 0 0 0 1	0 1 1
B	0 0 1 0 1 1 0 0 0 0 0 0 0 0 1	0 0 0 1 0	0 1 1
C	0 0 0 0 0 0 0 0 0 0 0 1 0 0 1	0 0 0 1 1	0 0 1
D	0 1 0 0 1 0 0 0 1 0 0 1 0 0 0	0 0 1 0 0	0 1 1
E	0 0 1 0 0 0 0 0 1 0 1 0 0 0 0	0 0 1 0 1	0 1 1
F	0 1 1 0 1 1 0 0 0 1 0 0 0 0 0	0 0 1 1 0	0 1 1
G	0 1 1 0 1 1 0 0 0 0 1 1 0 0 0	0 0 1 1 1	0 1 1
H	0 1 0 0 1 1 0 0 0 0 1 1 0 0 1	0 1 0 0 0	0 1 1
I	0 0 1 0 1 0 0 0 1 0 0 1 0 1 0	0 1 0 0 2	0 1 1
J	0 0 1 0 1 0 0 1 0 0 0 0 0 0 0	0 1 0 1 0	0 1 1
K	0 1 0 0 1 0 0 0 1 0 0 1 0 0 1	0 1 0 1 1	0 1 1
L	0 0 1 0 1 0 0 0 0 1 0 0 0 1	0 1 1 0 0	0 1 1
M	0 1 1 0 1 1 0 0 0 1 0 0 0 0 1	0 1 1 0 1	1 1 1
N	0 1 1 0 1 0 0 1 0 0 1 1 0 0 1	0 1 1 1 0	0 1 1
O	0 0 0 0 0 0 0 0 0 0 0 1 0 0 0	0 1 1 1 1	0 0 1
P	0 1 0 0 1 0 0 0 0 0 1 0 0 0 0	1 0 0 0 0	0 1 0
Q	0 1 0 0 1 1 0 0 1 0 0 1 0 0 0	1 0 0 0 1	0 1 1
R	0 1 0 0 0 1 0 0 0 1 0 1 0 0 0	1 0 0 1 0	1 1 1
S	0 0 0 0 0 0 0 0 0 0 1 0 0 0 0	1 0 0 1 1	0 1 0
T	0 1 0 0 1 1 0 0 0 0 1 0 0 0 1	1 0 1 0 0	0 1 1
U	0 0 1 0 1 0 0 0 0 0 1 0 0 0 0	1 0 1 0 1	0 1 1
V	1 0 1 0 0 0 0 0 0 0 0 0 0 0 1	1 0 1 1 0	1 0 1
W	1 0 0 0 1 0 0 0 1 0 1 0 0 0 1	1 0 1 1 1	1 1 1
X	0 0 0 1 1 0 0 0 1 0 0 0 0 0 1	1 1 0 0 0	1 1 1
Y	0 1 0 0 1 1 0 0 1 0 1 0 0 0 1	1 1 0 0 1	0 1 1
Z	0 0 1 0 0 0 0 1 0 0 1 0 0 0 0	1 1 0 1 0	0 1 1

과정에서는 회로망의 입력을 얻기 위하여 아날로그 신호로부터 IDMLP 신경회로망의 입력으로 적당한 디지털 값을 생성한다. 이 데이터로 회로망을 학습하고 마지막으로 학습된 회로망에 의한 인식실험을 하였다.

1. 시스템의 구성

본 연구에서의 음성인식 시스템은 그림6과 같이 구성되어 있다.

마이크로부터 아날로그 신호를 받아들여 테이프에 저장하고 이것중 필요한 부분을 증폭하고 필터링하여 A/D 보드로 전달하여 준다. A/D 보드에서 이 신호를 받아들여 PC의 제어에 따라 원하는 주파수로 샘플링한다. A/D 보드의 출력을 하드 디스크에 저장하고 이것에 대한 여러가지 처리를 한다. 처리된 결과는 화면에 디스플레이 되거나 플로트등으로 출력된다. 그리고 이 데이터는 다시 DT 2821 보드와 앰프를 거쳐 스피커로 출력된다.

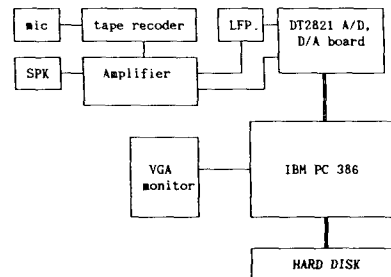


그림 6. 시스템의 구성
Fig. 6. Construction of the system.

2. 입력 데이터의 추출

입력으로는 한 화자에 의해 30번 발생된 0에서 9까지 숫자음 300개를 이용하였다. 300개의 숫자음 중 200개는 학습에서, 나머지 100개는 시험에 이용하

표 3. 입력에 bias를 추가한 입출력 data
Table 3. I/O data appending bias to input.

Char.	Input							Output							
	B.	original						original	added						
A	1	0	1	0	0	1	1	0	0	0	0	1	0	1	1
B	1	0	0	1	0	1	1	0	0	0	0	0	0	0	1
C	1	0	0	0	0	0	0	0	0	1	0	0	1	0	0
D	1	0	1	0	0	1	0	0	1	0	0	0	0	1	1
E	1	0	0	1	0	0	0	0	1	0	0	0	0	1	1
F	1	0	1	1	0	1	0	0	0	1	0	0	0	1	1
G	1	0	1	1	0	1	0	0	0	1	1	0	0	1	1
H	1	0	1	0	0	1	1	0	0	1	0	0	0	1	1
I	1	0	0	1	0	0	1	0	0	1	0	1	0	1	1
J	1	0	0	1	0	0	1	0	0	0	0	0	1	0	1
K	1	0	1	0	0	1	0	0	1	0	1	0	1	1	1
L	1	0	0	1	0	0	0	1	0	0	0	1	1	0	1
M	1	0	1	1	0	1	0	0	0	1	1	0	1	1	1
N	1	0	1	1	0	1	0	1	1	0	0	1	1	1	1
O	1	0	0	0	0	0	0	1	0	0	0	0	1	1	1
P	1	0	1	0	1	0	0	0	1	0	0	0	1	0	1
Q	1	0	1	0	0	1	1	0	0	1	0	0	0	1	1
R	1	0	1	0	0	0	1	0	1	0	0	0	1	1	1
S	1	0	0	0	0	0	0	1	0	0	0	1	0	1	1
T	1	0	1	0	0	1	0	0	0	1	0	1	0	1	1
U	1	0	0	1	0	0	0	1	0	0	0	1	0	1	1
V	1	0	0	1	0	0	0	0	0	0	1	1	0	1	1
W	1	1	0	0	1	1	0	0	1	0	0	1	1	1	1
X	1	1	0	0	1	0	0	0	0	0	1	1	0	0	1
Y	1	0	1	0	0	1	0	1	0	0	0	1	1	0	1
Z	1	0	0	1	0	0	1	0	0	0	1	1	0	1	1

표 4. 가중치 합을 제한하는 정도에 따른 학습의 결과

Table 4. Learning results by degree of weights.

l..X	M.S					
	100	50	40	30	20	15
1	61	65	69	61	66	53
2	84	88	84	88	73	57
3	100	100	100	100	84	75
4					96	84
5					100	88
7						96
9						100

였다.

입력 데이터에 대한 전처리 과정은 아날로그 신호로부터 신경회로망의 입력으로 사용될 수 있는 데이터를 생성하는 과정으로 그림 7에서와 같은 단계로

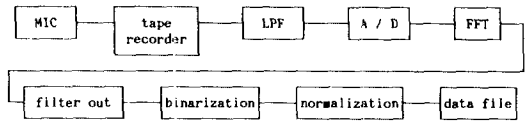


그림 7. 입력 데이터의 추출 과정

Fig. 7. The process of input data extraction.

이루어져 있다.

마이크에 의한 데이터의 입력은 잡음이 없는 환경에서 한명의 화자에 의해서 발생된 0에서 9까지의 음성이고 A/D 변환시 aliasing을 방지하기 위해서 5KHz의 차단 주파수를 갖는 6차 저역통과 필터를 이용하여 필터링을 하였다. 저역통과 필터를 통과한 신호는 DT2821 보드를 이용하여 12비트로 표분화된다.

FFT된 신호는 먼저 MEL 스케일에 의해 16개의 대역으로 분리되고 각각의 주파수 대역에서의 에너지를 구한다. 여기서 사용되는 MEL 스케일은 식 7에 의해 각 대역의 폭을 일정하게 하지 않고 인간의 청각계에서의 특성과 유사하게 주파수 대역을 분리한 것이다.

$$MEL = 1000 \log_2 (1 + f) \tag{7}$$

이렇게 분리된 대역은 1KHz 미만에서는 거의 비슷한 폭을 가지지만 1KHz 이상의 범위에서는 그폭이 점차 넓어진다.

MEL 스케일에 의해 각 대역에서 구한 에너지를 다음 식 8에서와 같이 이웃하는 필터의 에너지와 비교하여 그 차를 이전화한다.

$$OUT_n = \begin{cases} 1 : E_{nN} < E_{nN+1} \\ 0 : E_{nN} \geq E_{nN+1} \end{cases} \tag{8}$$

이렇게 해서 숫자음 '칠'에서 얻은 출력을 표 5에 나타내었다.

위에서와 같이 추출된 데이터는 회로망의 입력으로 사용되기 위해 시간축으로 정규화 되어야 한다. 이것은 서로 다른 길이의 발생된 음들에서 얻은 데이터를 일정한 입력 노드수를 가지는 회로망의 입력 데이터로 사용하기 위해서이다.

시간축 정규화는 이전화된 데이터의 프레임 수를 50으로 정하고 그 이상인 경우는 일정한 간격으로 초과된 수만큼의 프레임을 제거하고, 그 이하인 경우는 일정한 간격으로 필요한 수만큼의 프레임을 추가한다.

표 5. 숫자음 '칠'에 대한 데이터
Table 5. Data for digit '7'

Time (10ms)	이진화 된 data														
	f1	f2	f3	f4	f5	f6	f7	f8	f9	f10	f11	f12	f13	f14	f15
1	0	0	0	0	1	0	0	1	0	1	1	0	0	1	1
2	0	0	0	0	1	0	0	1	1	1	1	1	1	1	1
3	0	0	1	0	1	0	1	1	1	0	1	1	1	1	1
4	0	1	1	0	1	0	0	1	1	1	1	1	1	1	0
5	1	1	0	0	1	0	1	1	0	1	1	1	1	1	0
6	0	1	0	0	1	0	1	1	1	0	1	0	1	0	1
7	0	1	0	1	1	0	0	1	1	1	1	0	1	0	1
8	0	1	0	0	1	0	1	1	1	0	1	0	1	0	1
9	0	1	0	0	1	1	1	1	0	1	1	0	1	0	1
10	0	0	0	0	1	0	1	1	0	1	1	0	1	0	1
11	0	0	0	0	1	0	1	1	0	1	1	0	1	0	0
12	1	0	0	0	1	0	0	1	1	1	1	0	0	1	0
13	1	0	0	0	1	0	1	1	0	0	1	0	0	0	0
14	1	1	0	0	1	0	1	1	0	0	1	0	1	0	0
15	1	1	0	0	1	0	1	1	0	0	1	0	1	0	0
16	1	1	0	0	1	0	1	1	0	0	1	1	0	0	0
17	1	1	0	0	1	0	1	1	0	0	1	1	1	0	0
18	1	1	0	0	1	0	1	1	0	0	1	1	1	0	0
19	1	1	0	0	1	0	1	1	0	1	0	0	1	0	0
20	1	0	0	0	1	0	1	1	1	0	1	0	0	0	1
21	1	1	0	0	1	0	1	1	0	1	0	0	0	0	0
22	1	1	0	0	1	0	1	1	0	1	1	1	0	0	0
23	1	1	0	0	1	0	1	1	0	1	0	1	0	0	0
24	1	1	0	0	1	0	1	1	0	1	1	1	0	0	0
25	0	1	0	0	1	0	1	1	0	1	1	1	0	0	0
26	0	0	0	0	1	0	1	1	0	0	0	1	1	0	0
27	1	0	0	0	1	0	1	1	0	0	1	0	1	0	0
28	1	0	0	0	1	0	0	1	0	1	0	1	1	0	0
29	1	0	0	0	1	0	1	1	0	1	0	1	0	0	0
30	1	0	0	0	0	0	1	1	0	0	1	1	0	0	0
31	1	0	0	0	0	0	1	1	0	1	0	1	0	0	0
32	1	0	0	0	1	0	0	1	0	1	1	1	0	0	0
33	1	0	0	0	1	0	1	1	0	1	1	0	0	0	0
34	1	0	0	0	1	0	1	1	0	1	0	1	1	0	0
35	1	0	0	0	1	0	1	1	0	1	0	1	0	0	0
36	1	0	0	0	1	0	1	1	0	1	0	1	0	1	0
37	1	0	0	0	1	0	0	1	1	0	1	1	0	0	1
38	1	0	0	0	1	0	1	1	0	1	0	1	0	0	1

3. 회로망의 구성

추출된 데이터는 하나의 숫자음에 대해서 750비트로 표현된다. 데이터를 하나의 IDMLP 신경회로망으로 학습을 할 수 있으나 이렇게 할 경우 75개의 입력 노드가 필요하고 이들 사이의 연결의 수도 많아진다. 이러한 이유에서 회로망을 모듈화하여 구체적인 특징들에 대해서 학습하고 이들의 결과를 다시 학습하면 전체 입력에 대한 특징을 고려한 결과를

얻을 수 있고 각각의 모듈은 적은 수의 입력을 가지게 된다.

하나의 모듈은 하나의 필터의 출력을 학습하도록 한 후 그 결과를 다시 하나의 모듈로 학습하였다. 그림 8의 IDMLP1-IDMLP15는 각 필터의 출력에 대해 학습되고 IDMLP16은 그 결과들로 다시 학습된다. 모듈1~15는 50개의 입력 노드를 가지고 10개의 출력 노드를 가진다. 모듈16은 모듈 1~15의 출력을 모두 받아들이기 위하여 150개의 입력 노드를 가지고 10개의 출력노드를 가진다.

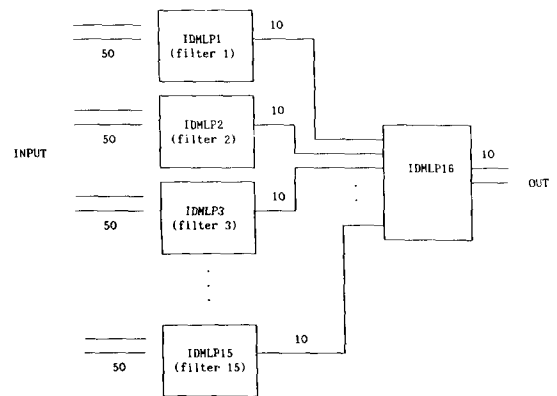


그림 8. 모듈러 IDMLP 신경회로망의 구성
Fig. 8. Construction of modular IDMLP neural network.

4. 각 모듈의 학습

각각의 모듈은 IDMLP 신경회로망으로 구성되어 있고 이들은 모두 두 층으로 층의 수가 제한되어 있다. 이것은 층 수가 증가하더라도 하나의 필터의 출력만으로는 모든 입력들을 구분할 수가 없기 때문이다. 그래서 각각의 모듈에서 구분이 되지 않은 것들은 이들을 종합하는 모듈에서 구분하도록 하였다. 각 모듈의 학습을 위해서 앞서 설명한 방법에 의해 구해진 데이터에서 각 필터의 출력을 분리하여 15개의 데이터를 만들었다.

표 6에 200개의 학습 데이터로 각 모듈을 학습하였을 때의 결과를 나타내었다. 학습 결과 하나의 모듈 내에서 학습에 이용된 모든 입력을 구분한 경우는 없었고 이들의 결과를 종합한 모듈 16에서 입력 데이터에 대해서 100%의 학습 결과를 얻었다.

5. 결과 및 고찰

학습된 회로망에 시험 데이터를 입력하여 인식실

표 6. 각 모듈의 학습시험 결과
Table 6. Learning and test result of each module.

L.N/F.N.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
학습	1	78	80	70	55	45	76	83	85	86	73	93	77	79	74	47	100
data	2	86	98	78	60	64	86	96	89	96	87	98	94	90	88	69	
시험	1	17	32	42	33	15	34	28	38	39	23	39	26	23	30	9	96
data	2	21	32	43	23	23	40	35	45	42	28	40	29	27	30	11	

표 7. 각 숫자음에 대한 인식률
Table 7. Recognition rate for each digit.

입 력	영	일	이	삼	사	오	육	칠	팔	구	100
정답 수	10	9	10	8	10	10	10	10	10	9	96
오답 수	0	1	0	2	0	0	0	0	0	1	4
인식률(%)	100	90	100	80	100	100	100	100	100	90	96

험을 하였다. 시험 데이터는 학습에서와 같은 방법으로 추출되었다. 시험 데이터를 회로망에 적용시켰을 때 각각의 모듈인 인식률을 아래에 표6에 나타내었고 이들은 대체적으로 낮은 값을 나타내었다. 이것은 각각의 대역값들 만으로는 전체적인 특징을 잘 나타낼 수가 없기 때문이다. 그러나 이렇게 낮은 인식률을 갖는 각각의 모듈의 출력을 종합한 모듈16의 결과는 100개의 입력 데이터에 대해서 96%의 인식률을 나타내었다. 표8과 표9에 인식의 결과를 각각의 숫자별로 나타내었다. 표10은 학습결과로 얻어진 모듈16의 가중치이다.

IV. 결 론

본 논문에서는 신경회로망을 이용하여 화자중속의 단독 숫자음 인식에 관하여 연구하였다. 사용된 ID-MLP 신경회로망과 이 회로망을 학습하기 위한 규칙은 새롭게 제안된 것이며, 이 제안된 회로망은 칩 구현을 위하여 가중치를 정수로, 뉴런에서의 전달함수를 hard limit 함수로 또 입출력을 이진수로 하였다. 제안된 회로망에 여러가지 입력 데이터를 이용하여 학습한 결과 입력 데이터들은 분류의 난이도에 따라 하나의 층, 또는 여러개의 층에 의해 분류되었다. 또 학습시 정수의 가중치를 이용하고 곱셈 연산을 사용하지 않으므로 짧은 시간내에 회로망을 학습할 수 있었다.

학습과 인식을 위한 데이터로는 한 사람의 화자에 의해 30번 발음된 0에서 9까지의 숫자음을 이용하였다. 회로망의 입력 데이터 추출은 음성신호의 FFT

결과를 MEL 스케일에 의한 주파수 대역의 성분으로 분할하고 이웃 대역과의 차에 의해 이진화한 것을 이용하였다. 신경회로망의 칩 구현을 위해서는 입력의 노드수가 적은 것이 유리하기 때문에 음성신호에서 추출한 데이터의 특징은 가능한한 적은 수의 비트로 제한하였다. 본 논문에서는 한 프레임 당 15 bit만으로 입력 음성의 특징을 나타내었다. 회로망의 구성에는 연결의 수를 줄이고 학습의 속도를 증가시키기 위해 16개의 모듈로 전체 회로망을 구성하였다. 학습은 전단의 15개의 회로망에 대해 따로 수행하고 그 출력들을 입력으로 하여 후단의 모듈을 다시 학습시켰다. 학습된 회로망에 대해 시험 데이터를 적용할 때는 일정한 문턱치를 사용하지 않고 열개의 출력 비트중에서 가장 높은 값을 갖는 것을 인식결과로 하였다. 학습 결과 각각의 모듈에서의 인식률은 높지 않았지만 전체 모듈을 종합한 후단의 모듈에서의 인식률은 학습 데이터에 대해서 100%, 시험 데이터에 대하여 96%의 양호한 결과를 얻었다.

앞으로 과제로는 시냅스의 가중치를 좀더 줄여서 칩 구현을 용이하게 하고 입력 데이터의 추출을 자동화하고 인식 대상의 범위를 확대하여 음성인식 장치로서의 기능을 충분히 수행하는 시스템을 구현하는 것이다.

參 考 文 獻

- [1] David S. Touretzky and Dean A. Pomerleau, "What's Hidden in the Hidden Layers?" *BYTE*, pp. 227-233, Aug. 1989.
- [2] Akers, L.A., Ferry, D.K. and Grondin, R.O., "VLSI Implementation of Neural Systems," *Computers and the Brain Symposium*, Apr. 1987.
- [3] Vegar, P.D. and Graf, H.P., "A CMOS Associative Memory Chips Based on Neural Networks," *1987 IEEE International Solid State Circuits Conference Digest of Technical Papers*, pp. 304-305, Feb. 1987.
- [4] Schwartz, D.B. and Howard, R.E., "A Programmable Analog Neural Network Chip," *IEEE 1988 Custom Integrated Circuits Conference*, no. 10, 2, May 1988.
- [5] 류종필, 정호선, 이우일, "신경 회로망을 이용한 세션화 및 특징점 추출," 1988년도 추계 종합학술대회 논문집, pp. 531-534, 1988. 11
- [6] 김태경, 정호선, 이우일, "신경 회로망을 이용한 A/D 변환기 설계," 1988년도 추계종합학술대회 논문집, pp. 511-515, 1988. 11.

- [7] 김태경, “신경 회로망의 VLSI 구현,” 경북대학교 석사학위논문, 1989. 2. 18.
- [8] 김태훈, 정호선, 이우일, “신경회로망을 이용한 2진화 화상의 잡음제거 및 직선화,” 1988년도 추계종합학술대회 논문집, pp. 554-557, 1988. 11.
- [9] Alexander Waibel, Toshiyuki Hanazawa, Geoffrey Hinton, Kiyohiro Shikano, Kevin J. Lang, “Phoneme Recognition Using Time-Delay Neural Networks,” *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, no. 3, March 1989.
- [10] Michael D. Riley, “Speech Time-Frequency Representations,” *AT & T Bell Laboratories*, 1989.

 著 者 紹 介



金 信 鎭 (準會員)

1966年 5月 18日生. 1989年 2月 경북대학교 전자공학과 졸업. 1991年 2月 경북대학교 전자공학과 석사학위 취득. 1991年 3月~현재 국방과학연구소 연구원. 주관심분야는 신경회로망의 VLSI 구현, 음성 신호처리 등임.



朴 政 運 (準會員)

1966年 5月 5日生. 1989年 2月 경북대학교 전자공학과 졸업. 1990年 3月~현재 경북대학교 전자공학과 석사 과정. 주관심분야는 신경회로망 및 퍼지이론의 VLSI 구현, 음성 신호처리 등임.

鄭 鎬 宣 (正會員) 第26卷 第6號 參照
현재 경북대학교 전자공학과 교수