

RBFN을 이용한 음소인식에 관한 연구

正會員 김 주 성*, 김 수 훈*, 허 강 인*

A study on the Phoneme Recognition using Radial Basis Function Network

Joo Sung Kim*, Soo Hoon Kim*, Kang In Hur* *Regular Members*

요 약

본 연구는 RBFN의 일종인 GPFN과 PNN을 이용한 음소인식에 관한 연구이다. RBFN의 구조는 계층형 신경망의 구조와 유사하지만, hidden층에서 활성화함수, 참조벡터 및 학습알고리즘의 선택이 다르다. 특히 PNN은 시그모이드 함수가 지수를 포함한 함수들의 한 분류로 대체된다는 것이며, 학습이 필요없으므로 전체계산 시간이 빠르게 수행된다.

5모음, 12자음을 대상으로 한 음소인식 실험에서 평가데이터, VQ와 LVQ에 의한 코드북 데이터를 사용한 경우에 음성의 통계적 특성을 잘 반영하고 있는 RBFN의 일종인 GPFN과 PNN의 인식결과가 MLP보다 우수하였다.

ABSTRACT

In this paper, we studied for phoneme recognition using GPFN and PNN as a kind of RBFN. The structure of RBFN is similar to a feedforward networks but different from choosing of activation fuction, reference vector and learning algorithm in a hidden layer. Especially sigmoid function in PNN is replaced by one category included exponential function. And total calculation performance is high, because PNN performs pattern classification without learning.

In phonemerecognition experiment with 5 vowel and 12 consnant, recognition rates of GPFN and PNN as a kind of RBFN reflected statistic characteristic of speech are higher than ones of MLP in case of using test data and quantized data by VQ and LVQ.

I. 서 론

컴퓨터 및 정보통신 기술의 급속한 발전과 보급이

활발해짐에 따라 전문가 뿐만 아니라 각 분야의 많은 사람들이 이용하게 되었고, 인간-기계의 인터페이스를 용이하게 하려는 욕구가 증대되고 있다. 그중에서 음성에 의한 인간과 기계사이의 인터페이스는 속도가 빠르고 특별한 훈련없이 이루어질 수 있다는 장점 때문에 연구가 활발히 진행되어지고 있다.

*동아대학교 전자공학과
論文番號:96117-0410
接受日字:1996年 4月 10日

음성정보처리의 한 분야인 음성인식의 연구는 1970년대부터 DTW(Dynamic Time Wrapping), HMM(Hidden Markov Model)에 의해 계속되어 왔다. 또 수년간 신경망에 의한 음성처리의 연구가 진행되어 좋은 결과를 보여 주고 있다.

신경망의 연구는 1943년 McCulloch-Pitts의 뉴런모델의 연구를 시작으로 1957년 Rosenblatt에 의한 퍼셉트론(Perceptron)연구로서 활성화 되었다. 1969년 Minsky-Papert등이 퍼셉트론의 한계에 관한 연구 후 침체 상태에 접어들었으며, 1986년 Rumelhart등이 계층형 신경망에 대한 역전파 학습법 알고리즘을 정식화 하였다.⁽¹⁾

이러한 계층형 신경망은 패턴인식, 로봇제어 및 시계열 예측 등의 응용분야에 적용되어 제품화가 본격화 되었다. 계층형 신경망은 이론적으로 중간층의 unit수를 충분히 많게 하면 3층의 신경망으로 임의의 함수를 근사화할 수 있으며 학습데이터에서는 평균 자승오차가 최소인 신경망의 구성이 가능하다고 발표되었다.⁽²⁾ 또 종래의 역전파 학습법의 결점인 많은 학습시간과 Local Minima의 문제점을 개선한 신경망도 발표되었다.^(3, 4, 5)

RBFN(Radial Basis Function Network)의 일종인 GPFN(Gaussian Potential Function Network)과 PNN(Probabilistic Neural Network)은 계층형 신경망의 구조가 유사하지만 차이점으로는 hidden층에서 활성화 함수, 참조벡터 및 학습 알고리즘의 선택이 다르다. 특히 PNN은 시그모이드 함수가 지수를 포함한 함수들의 한 분류로 대체된다는 것이며, 학습이 필요없으므로 전체적으로 계산시간이 빠르게 수행된다.⁽⁶⁾

본 연구에서는 사후확률을 이용하여 GPFN과 PNN으로 음소인식(5모음, 12자음) 실험을 수행하였으며, VQ와 LVQ로 작성된 코드북을 GPFN과 PNN의 입력으로 하였을 때 인식성능을 MLP(Multilayer Perceptron)와 비교하였다. 그 결과 GPFN과 PNN이 MLP보다 우수한 인식성능을 나타내었으며, 코드북 데이터를 사용한 경우에도 향상되었다.

II. 사후확률의 추정

2.1 벡터양자화에 의한 사후확률 추정^(10, 11, 14)

입력 데이터의 공간을 부분공간으로 나누어 데이

터의 발생빈도를 조사하여 사후확률을 추정할 수 있다. 벡터양자화에 의한 사후확률의 추정은 다음의 순서로 구한다.

- ① 벡터양자화를 이용하여 학습데이터의 코드북을 작성한다.
- ② 모든 학습데이터를 코드워드에 따라 분류한다. 이것은 입력데이터의 공간을 부분공간으로 분류하는 것이다.
- ③ 각 코드북의 인덱스에 포함된 학습데이터 수를 카테고리과 인덱스마다 구한다.

그리고 인덱스 i 에 포함된 카테고리 k 의 데이터 수를 N_{ki} 로 하면 인덱스 i 로 분류된 데이터 x 가 카테고리에 속하는 사후확률은 식 (1)이 된다.

$$p(C_i|x) = \frac{N_{ci}}{\sum_{k=1}^k N_{ki}} \quad (1)$$

여기서 추정한 사후확률은 이산확률분포이다.

2.2 신경망에 의한 사후확률 추정^(7, 8, 9, 10, 11, 14)

계층형 신경망은 입력신호와 교사신호간의 집합인 학습패턴을 부여하여 학습을 수행하므로 출력신호와 교사신호 사이의 자승오차가 최소가 되도록 신경망의 파라미터를 결정한다.

학습된 신경망은 구조의 제약하에서 시그모이드 함수를 이용한 비선형 사상을 재현한다. 제약이 없는 경우의 최적인 사상은 다음식으로 주어진다.

$$y(\bar{x}) = \int \bar{t} P(\bar{t}|\bar{x}) dt \quad (2)$$

여기서 \bar{x} 는 입력벡터, \bar{t} 는 교사벡터, $P(\bar{t}|\bar{x})$ 는 입력벡터가 \bar{x} 일 때의 교사벡터 \bar{t} 의 확률밀도 함수이다.

패턴인식 문제에서 교사벡터의 요소 t_1, t_2, \dots, t_k 를 각 카테고리에 값으로 한 경우의 최적인 변환은 식(3)과 같다.

$$y(x) = \sum_{k=1}^k t_k P(t_k|x) \quad (3)$$

여기서 $P(t_k|x)$ 는 입력이 x 일 때의 카테고리 C_k 에 속하는 확률로서 Bayes의 정리에 의한 사후확률에 해당

된다. 교사벡터의 각 요소를 입력벡터가 속하는 카테고리에 해당하는 경우는 1, 다른 카테고리에 해당하는 경우는 0으로 하면 입력벡터 x_i 에 대응하는 출력벡터의 k 번째 요소 y_{ik} 는 다음식이 된다.

$$y_{ik} = \sum_{t_k=1} P(t_i | x_i) = P(t_{ik} | x_i) \quad (4)$$

이것은 입력벡터 x_i 의 값으로서 교사벡터의 k 번째 요소가 1이 되는 확률이다. 그러므로 어느 카테고리에서 발생하는지 알고 있는 입력데이터를 몇 개 준비하고 입력데이터를 발생하는 카테고리에 해당하는 출력 unit에는 교사신호로 1, 그외의 카테고리에는 0을 부여하는 학습패턴을 만든다.

이 학습패턴을 신경망에 부여하여 역전파 알고리즘으로 학습한다. 학습된 신경망에 평가데이터로서 입력데이터를 부여하면 이 데이터가 각 카테고리에 어느 정도의 사후확률을 속하는지를 출력한다.

III. Learning Vector Quantization(LVQ)^(13, 14)

LVQ1은 생성되어지는 참조벡터의 각 군에 대해서 그 거리 척도로서 학습되어지는 알고리즘의 하나이다. 이 LVQ방법은 VQ를 개선한 방법으로 참조벡터를 뉴런들 간의 시냅스 결합에 대응시켜 교사 학습으로 적절한 참조벡터를 구하는 방법이다.

$$c = \arg \min \{ \text{length} (x - m_c) \}$$

참조벡터 m_i 에 가장 근접한 입력벡터 x 를 m_c 로 나타낸다.

가장 근사한 분류를 하기 위해서 오차를 최소로 하는 m_i 의 값은 식(5)와 (6)의 학습과정으로 생성된다.

- x 와 m_c 가 동일 Class일 때:

$$m_c(t+1) = m_c(t) + \alpha(t)[x(t) - m_c(t)] \quad (5)$$

- x 와 m_c 가 다른 Class일 때:

$$m_c(t+1) = m_c(t) - \alpha(t)[x(t) - m_c(t)] \quad (6)$$

- i 와 c 가 같지 않을 때:

$$m_i(t+1) = m_i(t)$$

여기서 $x(t)$ 는 입력벡터의 t 번째 요소이고, $m_i(t)$ 는 참

조벡터 m_i 의 t 번째 요소를 나타낸다. 그리고 $\alpha(t)$ 는 학습률로서 시간에 따라 단조 감소하는 함수이다.

LVQ1은 학습패턴 중 특별한 한개의 패턴이 코드북 벡터와의 거리가 클 경우 오분류가 발생하기 때문에 이런 단점을 보완한 LVQ2 알고리즘은 두 개의 근접 카테고리중에서 한 카테고리는 정분류이고, 다른 카테고리는 오분류일 때를 고려하였다. 그러나 학습횟수가 많을 때 각 분류중 참조벡터가 점근적으로 갱신될 경우 각각의 클래스가 안정성을 잃을 우려가 있다. 그래서 LVQ3에서는 각 분류에 대해서 식(7)과 (8)과 같이 α 값에 또 다른 학습가중치 $\epsilon(0 < \epsilon < 1)$ 를 곱해서 보정한다.

- x 와 m_c 가 동일 Class일 때:

$$m_c(t+1) = m_c(t) + \epsilon(t) \cdot \alpha(t)[x(t) - m_c(t)] \quad (7)$$

- x 와 m_c 가 다른 Class일 때:

$$m_c(t+1) = m_c(t) - \epsilon(t) \cdot \alpha(t)[x(t) - m_c(t)] \quad (8)$$

이 알고리즘은 자기안정화, 즉 m_i 의 최적화는 계속적인 학습에서 변화하지 않음을 알 수 있다.

IV. RBFN(Radial Basis Function Network)

RBFN은 hidden층을 1층 갖는 계층형 신경망과 같은 구조를 가지고 있다. RBFN은 입력된 벡터와 입력 공간중의 고정점 간의 거리를 계산하는 것으로서 hidden층에서 다음의 식으로 거리를 계산한다.

$$O_{ii} = R(\alpha_{ii}) = R_i(\|I - C_i\|) \quad (9)$$

여기서 I 는 입력벡터, O_{ii} 는 hidden층에서 i 번째 unit의 참조벡터 C_i 에 의해서 계산된 출력치이다. $R_i(\alpha_{ii})$ 는 출력함수를 나타낸다. $\|I - C_i\|$ 의 계산은 식(10)에 표현되어진 유클리드 거리를 구하면 된다.

$$\|I - C_i\|^2 = (I - C_i) \cdot M \cdot (I - C_i)^T \quad (10)$$

M 은 임의의 가중행렬이며, 단위행렬이나 특징벡터의 공분산행렬의 역행렬등이 이용된다. 출력층의 계산은 hidden층의 출력 가중치 합과 임계치 T_j 의 일반적인 식이다. 결국 네트워크 전체의 식은 다음과

같다.

$$O_j = F(\alpha_{2j}) F\left(\sum_i W_{ji} R_i(\alpha_{1i}) - T_j\right) \quad (11)$$

여기서, O_j 는 j 번째 네트워크의 출력이며, $F(\alpha_{2j})$ 는 시그모이드 함수이다.

$$F(\alpha_{2j}) = \frac{1}{1 + \exp(-\alpha_{2j})} \quad (12)$$

RBFN을 학습할 때에는 hidden층의 참조벡터를 고속인 종래의 방법(VQ등)으로 구하고, hidden층에서 출력층까지의 가중치만을 통상 최급하강법으로 수행하는 것으로서 MLP에 비해 학습시간을 대폭 단축할 수 있다. 또한 RBFN은 hidden층의 출력함수, 참조벡터 및 학습 알고리즘의 예로서 $R_i(\alpha_{1i})$ 를 지수함수로 하면

$$R_i(\alpha_{1i}) = \exp(-\alpha_{1i}^2) \quad (13)$$

네트워크의 분포는 가우스분포의 합으로 근사화하게 된다.

또 출력층에서의 가중치 W_{ji} 를 입력데이터의 사후 확률로 결정하면 Specht의 PNN이 얻어진다.

4.1 GPFN(Gaussian Potential Function Network)^(5, 16)

GPFN은 hidden층에 GPF(Gaussian Potential Function)라 부르는 unit로 되어져 있다. x 를 입력패턴으로 할 때 hidden unit ψ_i 는 다음과 같이 정의한다.

$$\psi_i = \psi(x, p_i) = e^{-d(x, p_i)/2} \quad (14)$$

$$\begin{aligned} d(x, p_i) &= d(x, m^i, K^i) \\ &= (x - m^i)^T K^i (x - m^i) \end{aligned} \quad (15)$$

여기서, m^i 와 K^i 는 각각 i 번째 GPF의 평균벡터와 공분산행렬의 역행렬이다. $d(x, m^i, K^i)$ 를 다시 전개하면 식(16)과 같다.

$$d(x^i, m^i, K^i) = \sum_j \sum_k k_{jk}^i (x_j - m_j^i)(x_k - m_k^i) \quad (16)$$

여기서 x_j 는 입력벡터 x 의 j 번째 요소, m_j^i 는 평균벡터 m^i 의 j 번째 요소, k_{jk}^i 는 역공분산 행렬 K^i 의 (j, k) 번째 요소를 각각 나타낸다. k_{jk}^i 는 주변(Marginal) 표준편차 σ_j^i 와 σ_k^i 의 상관 계수 h_{jk}^i 로 나타낼 수 있다.

$$k_{jk}^i = \frac{h_{jk}^i}{\sigma_j^i \sigma_k^i} \quad (17)$$

여기서 σ_j^i 는 양의 실수

$$\begin{cases} h_{jk}^i = 1, & j = k \\ h_{jk}^i \leq 1, & \text{그외} \end{cases}$$

그림 1에 3층 GPFN의 구조를 나타내었다. 입력층과 출력층은 선형 unit들로 구성되었으며, hidden층은 GPF를 발생하는 GPFU(Radial Basis Function Unit)로 구성되었다.

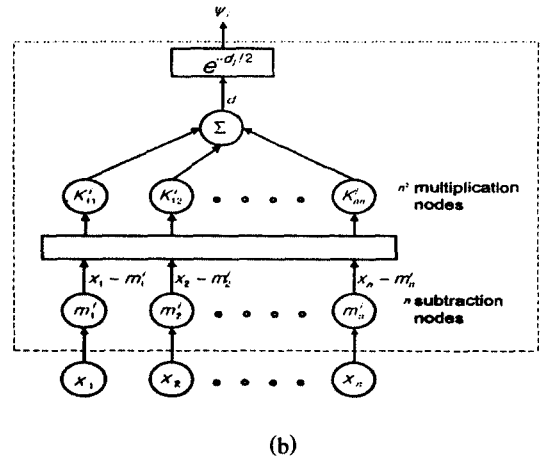
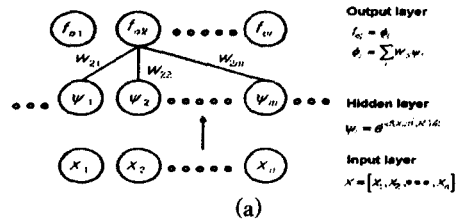


그림 1. GPFN의 구조

(a)GPFN (b)입력층과 GPFU사이의 접속

GPFN의 가중치 출력들은 원하는 potential fields를 합성하기 위해서 hidden층과 출력층 사이에 연결되어 더해진다. j 번째 출력 unit의 출력치를 ϕ_j 로 하면

$$\phi_j = \sum_i w_{ji} \psi_i \quad (18)$$

로 된다.

이 네트워크를 학습하기 위해서는 역전파 학습법과 동일하게 오차함수에 최급강법을 적용하고 교사패턴을 참조하여 파라미터를 갱신하는 것이다. p 번째 교사패턴의 오차함수 E_p 를 식(19)로 정의한다.

$$E_p \equiv \frac{1}{2} \sum_{j=1}^M (t_{pj} - \phi_{pj}(n_j))^2 \quad (19)$$

여기서 M 은 출력 unit의 수, t_{pj} 는 교사신호를 나타낸다. ϕ_{pj} 는 실제 출력의 j 번째 요소를 나타내고, n_j 는 j 번째 출력 unit의 전파라미터에서 열벡터로 된다.

$$n_j \equiv [w_j^i, m_j^i, \sigma_j^i, h_j^i]^T \quad (20)$$

방향벡터 $\Delta n_j \equiv [\Delta w_j^i, \Delta m_j^i, \Delta \sigma_j^i, \Delta h_j^i]^T$ 는 최급강법으로 구해진다.

파라미터의 갱신규칙은

$$n_j^{\text{new}} = n_j^{\text{old}} + \eta \Delta n_j \quad (21)$$

로 된다. η 은 학습률이고 양의정수이다.

- j 번째 출력층과 i 번째 GPFU 사이의 가중치

$$\Delta w_{ji} = -\frac{\partial E_p}{\partial w_{ji}} = (t_j - \phi_j) \psi_i \quad (22)$$

- 평균벡터 m^i 의 j 번째 요소

$$\begin{aligned} \Delta m_j^i &= -\frac{\partial E_p}{\partial m_j^i} \\ &= \sum_l k_{jl}^i (x_l - m_j^i) \psi_i \sum_k (t_k - \phi_k) w_{ki} \end{aligned} \quad (23)$$

- 주변(marginal) 표준편차 σ_j^i

$$\begin{aligned} \Delta \sigma_j^i &= -\frac{\partial E_p}{\partial \sigma_j^i} \\ &= -\frac{1}{2} \frac{(x_j - m_j^i)(x_k - m_k^i)}{\sigma_j^i \sigma_k^i} \psi_i \sum_k (t_k - \phi_k) w_{ki} \end{aligned} \quad (24)$$

- k_{jk}^i 에 대한 역공분산 계수 h_{jk}^i

$$\Delta h_{jk}^i = -\frac{\partial E_p}{\partial h_{jk}^i} \quad (25)$$

$$= -\frac{1}{2} \frac{(x_j - m_j^i)(x_k - m_k^i)}{\sigma_j^i \sigma_k^i} \psi_i \sum_k (t_k - \phi_k) w_{ki}$$

4.2 PNN(Probabilistic Neural Network)^(5, 6, 11, 12, 15, 16)

패턴을 분류하기 위해서 사용되는 결정 규칙이나 정리들은 expected risk를 최소화하는 것이다. Bayes의 정리는 여러 다른 카테고리를 갖는 문제들에 적용할 수 있다. 학습벡터가 X_1, X_2, \dots, X_m 으로 주어질 때 Specht는 학습데이터의 확률밀도함수를 Gaussian Kernel로서 추정하였다.

$$P(X) = \frac{1}{(2\pi)^{p/2} \sigma^p} \frac{1}{m} \sum_{i=1}^m \exp\left[-\frac{(X - X_i)^T (X - X_i)}{2\sigma^2}\right] \quad (26)$$

여기서 i 는 패턴수, m 은 학습패턴의 총수, X_i 는 카테고리에서 i 번째 학습패턴, σ 는 스무딩 파라미터 그리고 p 는 측정공간에서의 차원이다.

그림 2는 두 카테고리로 입력패턴 X 를 분류하는 PNN 구조의 예를 나타낸다.

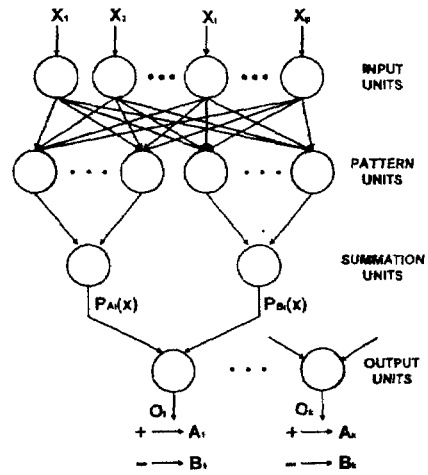


그림 2. 분류를 위한 PNN의 구조

그림 2에서 입력 unit들은 단지 모든 패턴 unit에 동일한 입력값들을 공급하기 위한 분배 unit들이다. 각 패턴 unit들은 입력패턴 X 와 가중치 벡터 W_i 의 내적으로 이루어진다.(그림 3)

$$Z_i = X \cdot W_i, \quad (27)$$

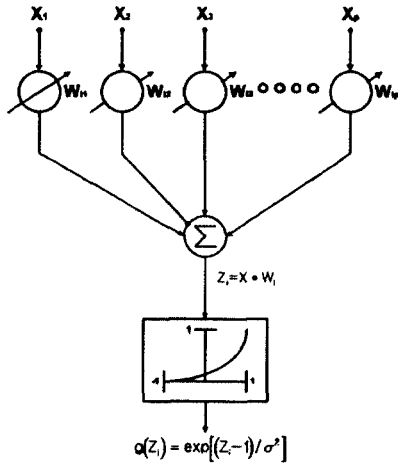


그림 3. 패턴 unit

그리고 가산 unit에 그것들의 활성레벨을 출력하기 전에 Z_i 에 식(28)의 비선형변환인 시그모이드 함수를 이용한다.

$$g(Z_i) = \exp[(Z_i - 1) / \sigma^2] \quad (28)$$

X 와 W_i 를 단위길이로 정규화하면 식(29)이 된다.

$$g(X) = \exp[-(W_i - X)^2 / (W_i - X) / 2\sigma^2] \quad (29)$$

이것은 식(26)의 지수함수항과 같은 형태이다. 그러므로 내부의 연결들에서 수행되는 내적은 뉴런 활성화 함수(지수적인 함수)에 따르며 가산 unit들은 선택된 학습패턴에서 카테고리에 일치하는 패턴 unit들에서 입력들을 간단히 더한다.

V. 인식실험 및 고찰

5.1 음성DB 및 분석조건

표 1은 6명의 화자가 발성한 음절중에서 모음만 취하였고 분석조건은 표 2와 같다.(예: /ba/ → /a/)

표 3은 5명의 화자가 발성한 음절중에서 초성자음만 취하였고 분석조건은 표 4와 같다.

표 1. 모음의 음성DB

음성	아	이	우	에	오	합계
학습	1380	1072	1740	846	1668	6706
평가	400	518	248	421	422	2009

표 2. 모음의 분석조건

샘플링 주파수	12 KHz
창함수(해명창)	21.33 ms
프레임 주기	5 ms
분석	14차 LPC 분석
특징 파라미터	10차 LPC mel cepstrum 계수

표 3. 초성자음의 음성DB

음성	유성파열음			비음			마찰음 및 파열음			무성파열음		
	ㄱ	ㄷ	ㅌ	ㄴ	ㄹ	ㄷ	ㅌ	ㅈ	ㅊ	ㅋ	ㆁ	ㅇ
학습	990	360	465	630	420	390	600	600	555	45	240	255
합계	1815			1440			1755			540		
평가	660	240	310	420	280	260	400	400	370	30	160	170
합계	1210			960			1170			360		

표 4. 자음의 분석조건

샘플링 주파수	10 KHz
창함수(해명창)	20 ms
프레임 주기	5 ms
분석	14차 LPC 분석
특징 파라미터	10차 LPC mel cepstrum 계수

5.2 인식결과 및 고찰

표 5는 모음과 초성자음의 인식을 위한 3층 신경망의 구조를 나타내었고, 학습횟수는 3000회로 제한하였다. 각 신경망의 입력은 모음의 경우는 10입력 units (1프레임 × 10 mel cep 계수), 자음의 경우는 20입력 units (2프레임 × 10 mel cep 계수)로 하였다. 특히 PNN의 hidden층은 각각 모음과 자음의 학습데이터 수 만큼의 units를 가진다. VQ와 LVQ(LVQ1)로 구성된 코드북은 각각 64, 128, 256개로 하였다.

표 6, 7에는 사후확률을 이용한 음소(5 모음, 12 자음)의 인식결과를 각각 나타내었다. 특히 초성자음은

표 5. 신경망의 구조

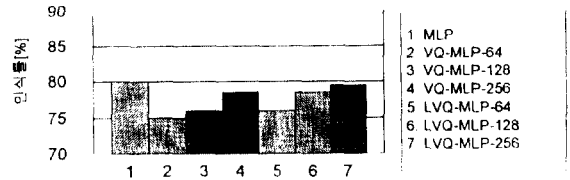
음성	신경망	입력층	중간층	출력층
모음	MLP	10 units	15 units	5 units
	GPFN	10 units	16 units	5 units
	PNN	10 units	학습데이터수	1 unit
자음	MLP	20 units	30 units	3 units
	GPFN	20 units	32 units	3 units
	PNN	20 units	학습데이터수	1 unit

음성발생시 조음방식이 유사한 음성파열음(ㄱ, ㄷ, ㄴ), 비음(ㄹ, ㅁ, ㅂ), 마찰음 및 파열음(ㅅ, ㅆ, ㅈ), 무성파열음(ㅋ, ㅌ, ㅍ)에 대해서 인식결과를 각각 나타내었다.

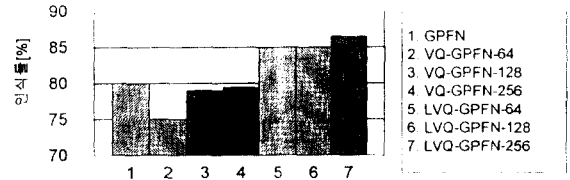
그림 4는 모음에 대해서 (a)는 MLP, (b)는 GPFN, (c)는 PNN을 사용한 경우의 평가데이터의 인식률을 각각 나타내었다. 모음에 대한 평가데이터의 인식률은 코드북 256의 VQ-PNN의 인식률이 88.15%로 가장 우수하였다.

그림 5는 초성자음에 대해서 (a)는 MLP, (b)는 GPFN, (c)는 PNN을 사용한 경우의 평가데이터의 인식률을 각각 나타내었다. 음성파열음에서는 코드북 128의 LVQ-PNN의 인식률이 77.02%로 가장 우수하였고, 비음에서는 코드북 64의 LVQ-PNN이 77.01%로 가장 우수한 인식률을 나타내었다. 그리고 마찰음 및 파열음과 무성파열음에서는 각각 78.86%, 78.89%로 PNN에서 가장 우수한 인식률을 나타내었다.

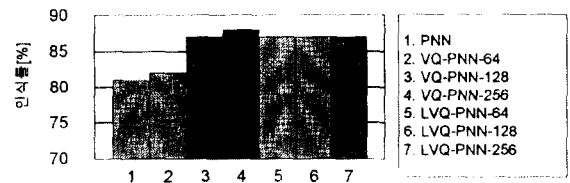
따라서 음성에 대한 통계적 특성을 잘 반영하는 RBFN의 일종인 GPFN과 PNN의 인식성능이 MLP 보다는 우수함을 알 수 있었다. 그리고 GPFN에서는 VQ와 LVQ로 작성된 코드북 데이터를 사용하였을 때 평가데이터를 사용한 경우보다 인식률이 저하되었다. 그러나 PNN에서는 VQ보다는 LVQ에 의해서 작성된 코드북 데이터를 사용하는 것이 평가데이터와 거의 비슷하거나 향상된 인식성능을 나타내었다. 따라서 PNN은 학습이 필요없으므로 전체계산 시간이 빠르게 수행되었고 인식성능도 우수하였다. 또한 스무딩 파라미터 σ 를 0.05~10까지 변화시켰을 때 인식성능이 가장 좋은 경우의 스무딩 파라미터의 값은 PNN에서는 0.2~0.3, VQ·LVQ와 PNN의 하이브리드 구성인 경우에는 0.1~0.2로 각각 나타났다. 또한 코드



(a) MLP의 경우



(b) GPFN의 경우



(c) PNN의 경우

그림 4. 모음의 인식률

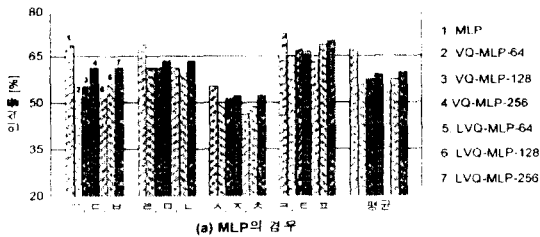
(a) MLP의 경우 (b) GPFN의 경우 (c) PNN의 경우

표 6. 모음의 인식률[%]

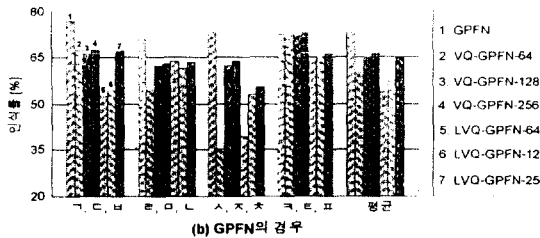
방법	코드북	학습	평가
MLP	·	91.35	79.54
GPFN	·	92.68	79.74
PNN	·	94.2	80.99
VQ-MLP	64	88.88	74.71
	128	89.52	76.26
	256	90.53	77.85
VQ-GPFN	64	87.67	75.06
	128	90.55	78.30
	256	91.26	79.19
VQ-PNN	64	89.05	81.78
	128	92.70	86.96
	256	94.38	88.15
LVQ-MLP	64	87.74	75.38
	128	88.76	77.81
	256	91.22	78.67
LVQ-GPFN	64	89.38	84.76
	128	90.11	84.90
	256	92.20	87.10
LVQ-PNN	64	93.65	87.03
	128	94.37	87.24
	256	94.89	87.60

표 7. 초성자음의 인식률[%]

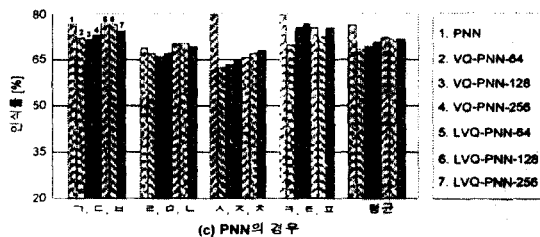
분류 방법	코드북	ㄱ, ㄷ, ㅂ		ㄴ, ㄹ, ㄴ		ㅅ, ㅈ, ㅊ		ㅋ, ㅌ, ㅍ		평균	
		학습	평가	학습	평가	학습	평가	학습	평가	학습	평가
MLP	.	71.40	70.66	67.32	69.25	56.08	55.77	71.85	73.06	66.66	67.19
GPFN	.	89.48	76.20	82.20	71.13	81.49	73.26	91.11	73.06	86.07	73.41
PNN	.	85.62	76.28	90.74	70.19	90.43	78.86	90.74	78.89	89.38	76.06
VQ-MLP	64	51.46	50.91	61.58	61.19	48.27	49.07	63.89	65.28	56.30	56.61
	128	54.77	54.05	61.92	61.09	47.76	50.76	68.15	66.67	58.15	58.14
	256	61.49	60.74	63.74	63.81	50.08	52.29	65.00	65.56	60.08	60.60
VQ-GPFN	64	69.37	68.18	53.45	53.87	33.77	34.63	75.37	72.50	57.99	57.30
	128	70.58	67.27	63.07	62.13	60.61	60.36	76.44	72.53	67.68	65.57
	256	73.55	68.76	66.41	62.87	63.27	62.48	77.32	73.21	70.14	66.83
VQ-PNN	64	69.59	71.07	65.99	67.47	58.35	59.51	75.00	70.83	67.23	67.22
	128	71.29	70.74	67.80	67.05	61.46	62.22	75.93	75.83	69.12	68.96
	256	73.61	72.23	68.08	68.51	64.69	63.67	80.19	77.50	71.64	70.48
LVQ-MLP	64	51.07	50.99	60.81	59.52	43.75	44.82	62.22	63.61	54.46	54.74
	128	55.21	55.21	61.58	58.79	47.48	48.05	69.63	69.72	58.45	57.94
	256	61.21	60.99	63.74	63.18	52.18	52.55	69.63	70.56	61.69	61.82
LVQ-GPFN	64	54.55	53.54	68.78	63.49	43.72	40.80	67.22	62.78	58.57	55.15
	128	55.78	55.21	67.39	60.56	56.03	56.18	71.85	63.89	62.76	58.71
	256	72.78	67.74	68.92	62.76	62.82	57.13	75.19	66.94	69.93	63.64
LVQ-PNN	64	79.94	76.86	82.30	71.86	68.19	65.53	85.00	76.67	78.86	72.73
	128	83.69	77.02	81.32	71.65	71.36	67.23	79.63	73.33	79.00	72.31
	256	79.17	73.64	83.21	71.03	73.23	68.76	79.44	76.11	78.76	72.39



(a) MLP의 경우



(b) GPFN의 경우



(c) PNN의 경우

그림 5. 초성자음의 인식률

(a) MLP의 경우 (b) GPFN의 경우 (c) PNN의 경우

복 수가 많은 경우가 작은 경우보다 인식률이 낮은 것은 코드북의 인덱스와 평가데이터와의 유클리드 거리의 차이로 인해 카테고리를 잘못 찾은 것으로 생각된다.

VI. 결론

본 연구는 사후확률을 이용하여 RBFN의 일종인 GPFN과 PNN으로 음소에 대한 인식 실험을 수행하였다. 그리고 VQ와 LVQ에 의한 코드북을 신경망의 입력으로 하였을 때의 인식성능을 MLP와 각각 비교하였다.

코드북 데이터를 신경망의 입력으로 하였을 때 모음에서는 거의 동등하거나 향상된 인식결과를 얻을 수 있었고, 초성자음에서는 전체적으로 인식률이 저하되었지만 양자화된 데이터를 사용하여 데이터량을 대폭 감소시키므로서 학습시간을 줄일 수 있었다.

모음(아, 이, 우, 에, 오)에 대한 평가데이터의 인식률은 코드북 256의 VQ-PNN의 인식률이 88.15%로 가장 우수하였다. 초성자음의 평가데이터에서 유성파열음(ㄱ, ㄷ, ㅂ)의 경우에는 코드북 128의 LVQ-PNN의

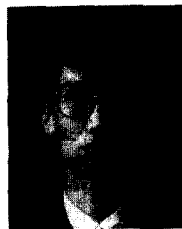
인식률이 77.02%로 가장 우수하였고, 비음(ㄷ, ㄹ, ㄴ)의 경우에는 코드북 64의 LVQ-PNN이 77.01%로 가장 우수한 인식률을 나타내었다. 그리고 마찰음 및 파열음(ㅅ, ㅆ, ㅈ)과 무성파열음(ㅋ, ㆁ, ㆏)의 경우에는 각각 78.86%, 78.89%로 PNN에서 가장 우수한 인식률을 나타내었다.

따라서 음성의 통계적 특성을 반영한 RBFN의 일종인 GPFN과 PNN의 인식성능은 MLP보다 우수하였다. 특히 GPFN보다 PNN은 학습이 필요없으므로 전체계산 시간이 빠르게 수행되었고 인식성능도 우수하였다.

그러므로 모음과 자음은 그 통계적 특성이 각각 다르므로 음절이나 단어등의 인식시에는 음소의 특성에 적합한 신경망을 병렬로 구성하여야 된다고 생각된다.

참 고 문 헌

1. D. Rumelhart, G. Hinton and R. Williams: Learning internal representations by error backpropagation, In D. Rumelhart and J. McClelland(Eds), Parallel Distributed Processing, MIT Press, 1986.
2. 船橋賢-, "3層ニューラルネットワークによる恒等寫像の近似的實現についての理論的考察", 信學論, J73-A, 1, pp. 139-145, 1992.
3. R. Hecht-Nielsen: Theory of the backpropagation neural network, IEEE International Joint Conference on Neural Networks, Vol. 1, pp. 593-605, 1989.
4. R. Beale and T. Jackson: Neural Computing: An Introduction, Adam Hilger Bristol, Philadelphia and New York, 1991.
5. B. Kosko, editor, Neural networks for Signal Processing, pp. 199-223, Prentice-Hall International, Inc
6. D. Specht, "Probabilistic Neural Networks, Neural Networks", Vol. 3, pp. 109-118, 1990.
7. D. W. Ruck, et al., "The multilayer perceptron as an approximation to a Bayes optimal discriminant function", IEEE Trans. Neural Networks, Vol. 1, No. 4, pp. 296-298, 1990.
8. E. A. Wan, "Neural network classification: a Bayesian interpretation", IEEE Trans. Neural Networks, Vol. 1, NO. 4, pp. 303-305, 1990.
9. 栗田多喜夫, "情報量基準による3層ニューラルネットワークの隠れ層のユニット数の決定法", 信學論, Vol. J73-D-11, No. 11, pp. 1872-1878, 1990.
10. Y. Ono, S. Nakagawa, "Estimation of a posteriori probability using neural network", Proc. IJCNN, pp. 789-794, Nov., 1991.
11. 허강인, 이광석, 김명기, "사후확률·확률밀도함수의 추정과 Probabilistic Neural Network을 이용한 모음인식에 의한 평가", 한국음향학회지, 12권, 16호, pp. 21-27, 1993.
12. Stanley C. Ahalt, et al, "Competive Learning Algorithms for Vector Quantization", Neural Networks, Vol. 3, pp. 277-290, 1990.
13. Pietro Burrascano, "Learning Vector Quantization for the Probabilistic Neural Networks", IEEE Trans. Neural Networks, Vol. 2, pp. 456-461, 1991.
14. 小野義之, 中川聖一, "ニューラルネットワークによる事後確率と確率密度函數の推定と母音認識", 日本音響學會講論集, 2-1-9, 1992.
15. 김주성, 허강인, "사후확률을 이용한 RBF 신경망에 의한 음소인식", 신호처리 합동학술대회 논문집, 제7권, 1호, pp. 743-747, 1994.



김 주 성(Joo Sung Kim) 정회원
 1967년 11월 23일생
 1990년 2월: 동의대학교 전자공학과 졸업(공학사)
 1993년 2월: 동의대학교 대학원 전자공학과 졸업(공학석사)
 1994년 3월~현재: 동의대학교 대학원 전자공학과 박사과정
 ※주관심분야: 음성인식·합성, 신경회로망



김 수 훈(Soo Hoon Kim) 정회원

1968년 2월 25일생

1991년 2월: 동아대학교 전자공학과 졸업(공학사)

1993년 2월: 동아대학교 대학원 전자공학과 졸업(공학석사)

1993년 3월~현재: 동아대학교 대학원 전자공학과 박사과정

※주관심분야: 음성인식·합성, 신경회로망



허 강 인(Kang In Hur) 정회원

1955년 2월 20일생

1980년 2월: 동아대학교 전자공학과 졸업(공학사)

1982년 2월: 동아대학교 대학원 전자공학과 졸업(공학석사)

1990년 8월: 경희대학교 대학원 전자공학과 졸업(공학박사)

1984년 9월~현재: 동아대학교 공과대학 전자공학과 교수

1988년 9월~1989년 8월: 일본 筑波大學 객원연구원

1992년 9월~1993년 8월: 일본 豊橋大學 객원연구원

※주관심분야: 디지털 신호처리, 음성인식·합성, 신경회로망