

# 신경회로망과 벡터양자화에 의한 사후확률과 확률 밀도함수의 추정 및 검증

速報論文

45~2~1

## Verification and Estimation of a Posterior Probability and Probability Density Function Using Vector Quantization and Neural Network

高 義 石\* · 金 炫 惠\*\* · 李 光 錫\*\*\*

(Hee-Seok Koh · Hyun-Duck Kim · Kwang-Seok Lee)

**Abstract** - In this paper, we proposed an estimation method of a posterior probability and PDF(probability density function) using a feed forward neural network and codebooks of VQ(vector quantization). In this study, We estimates a posterior probability and probability density function, which compose a new parameter with well-known Mel cepstrum and verificate the performance for the five vowels taking from syllables by NN(neural network) and PNN(probabilistic neural network). In case of new parameter, showed the best result by probabilistic neural network and recognition rates are average 83.02%

**Key Words** : Speech analysis, Neural network, Estimation of Posterior probability and probability density function.

### 1. 서 론

인간과 기계사이의 통신을 위해 컴퓨터를 이용한 음성신호의 분석, 합성 및 인식등의 연구가 계속되고 있으며 컴퓨터 및 통신기술의 발전으로 인해 제한적이나마 음성 인식기술이 실용화 되어가고 있다. 특히 1980년대 중반부터 VQ(Vector Quantization), HMM(Hidden Markov Model)등에 의해 주도되어 왔으나 최근 인간의 두뇌의 병렬처리 기능 및 학습능력에 근거한 신경회로망에 의해 음성처리의 연구가 활발히 진행되어 좋은 결과를 보여주고있다. 신경회로망의 연구는 1943년 McCulloch and Pitts의 뉴런모델을 시작으로 1986년 D.E Rumelhart 등이 계층형 신경회로망에 대한 BP(BackPropagation)알고리즘을 정식화 하였다[1, 2]. 이러한 계층형 신경회로망은 패턴인식, 로보트 제어 및 시계열 예측 등의 각종 신호처리 응용분야에 적용되어 비교적 만족할 만한 결과를 보여주고있다[1, 3, 4].

본 연구에서는 계층형 신경회로망과 벡터 양자화로서 사후확률과 확률 밀도함수를 추정하는 방법을 제안, 비교하였으며 이를 음성 인식의 파라미터로서의 타당성을 검토한 결과 매우 유용함을 알 수 있었다.

### 2. 벡터 양자화(6)

음성의 K차원 특징 파라미터는 시계열 벡터로 표현되며 각 벡터  $x_i$ 는 코드북 벡터  $y_n$ 으로 치환된다.

$$\hat{n} = \underset{n}{\operatorname{argmin}} d(x_i, y_n) \quad (1)$$

멀티 템플릿 작성을 위한 clustering기법으로 대표적 코드북 작성 알고리즘인 Linde등의 LBG알고리즘을 이용하였으며 centroid벡터는 다음식으로 정의한다.

$$\begin{aligned} \hat{x} &= c(x_1, x_2, \dots, x_n) \\ &= \underset{x}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^n d(x, x_i) \end{aligned} \quad (2)$$

### 3. 사후 확률의 추정

#### 3.1 벡터 양자화에 의한 방법

입력 데이터를 separation set로 나누어 데이터의 발생 빈도를 조사하여 사후확률을 추정할 수 있으며 다음 순서로 구한다.

- (1) 벡터양자화를 이용하여 학습 데이터의 코드북을 작성한다.
- (2) 학습 데이터를 코드북에 따라 모두 분류한다.
- (3) 각 코드북의 인덱스에 포함된 학습 데이터 수를 카테고리 인덱스마다 구한다.

그리고 인덱스 i에 포함된 카테고리 k의 데이터 수를  $N_{ki}$ 로 하면 인덱스 i에 분류된 데이터 x가 카테고리에 속하는 확률, 즉 사후확률은 다음과 같으며 이는 이산 확률분포이다.

$$P(C|x) = \frac{N_{ci}}{\sum_{k=1}^k N_{ki}} \quad (3)$$

#### 3.2 신경회로망에 의한 방법

학습된 신경회로망은 구조의 제약하에서 시그모이드 함수를 이용한 비선형의 사상을 제한한다. 제약이 없는 경우의 최적인

\* 正 會 員 : 慶南大 工大 電氣工學科 教授·工博

\*\* 正 會 員 : 慶南大 大學院 電氣工學科 博士課程

\*\*\* 正 會 員 : 晉州産業大 電子工學科 專任講師·工博

接受日字 : 1995年 10月 23日

1次 修正 : 1996年 1月 4日

사상은 다음과 같다.

$$y(\mathbf{x}) = \int \mathbf{t} p(\mathbf{t}|\mathbf{x}) d\mathbf{t}$$

여기서  $\mathbf{x}$ 는 입력벡터,  $\mathbf{t}$ 는 교사벡터 및  $P(\mathbf{t}|\mathbf{x})$ 는 입력벡터  $\mathbf{x}$  일때의 교사벡터  $\mathbf{t}$ 의 확률 밀도함수이다. 패턴인식 문제에서 교사벡터의 요소 ( $t_1, t_2, \dots, t_k$ )를 각 카테고리에 대응한 값으로 한 경우의 최적변환은 다음과 같다.

$$y(\mathbf{x}) = \sum_{k=1}^K \mathbf{t}_k P(\mathbf{t}_k|\mathbf{x}) \quad (4)$$

여기서,  $P(\mathbf{t}_k|\mathbf{x})$ 는 입력이  $\mathbf{x}$ 일 때의 카테고리  $C_k$ 에 속하는 확률로서 Bayes 정리에 의한 사후 확률에 해당하며 입력벡터  $\mathbf{x}_i$ 에 대응하는 출력벡터의  $k$ 번째의 요소  $y_{ik}$ 는 다음과 같다.

$$y_{ik} = \sum_{t_{ik}=1} p(t_{ik}|\mathbf{x}_i) = P(t_{ik}=1|\mathbf{x}_i) \quad (5)$$

이것은 입력벡터  $\mathbf{x}_i$ 의 값으로 교사벡터의  $k$ 번째의 요소가 1이 되는 확률이다. 그러므로 어느 카테고리에서 발생하는지 알고 있는 입력샘플을 준비하고 이 입력샘플을 발생한 카테고리에 해당하는 출력 unit에는 교사신호 1을, 그 외의 카테고리에는 0을 부여하는 학습패턴으로 학습하며 학습된 신경회로망에 평가 데이터를 부여하면 각 카테고리에 어느 정도로 속하는지의 확률, 즉 사후확률을 출력한다.

#### 4. 확률 밀도함수의 추정

벡터 양자화에 의한 확률 밀도함수의 추정은 다음과 같다.[5]

- (1) 벡터 양자화를 이용하여 학습데이터의 코드북을 작성한다.
- (2) 총 학습 데이터를 코드북 수에 해당하는 separation set로 모두 분류한다.
- (3) 각 카테고리마다 separation set에 포함된 데이터 수를 총 데이터수로 나누어 확률 밀도함수를 구한다.

$i$ 를 포함한 separation set에 포함된 카테고리  $C$ 의 총 학습 데이터수를  $N_c$ 라 하면 카테고리  $C$ 의  $i$ 번째 separation set에 분류된 데이터  $\mathbf{x}$ 의 이산 확률분포는 다음과 같다.

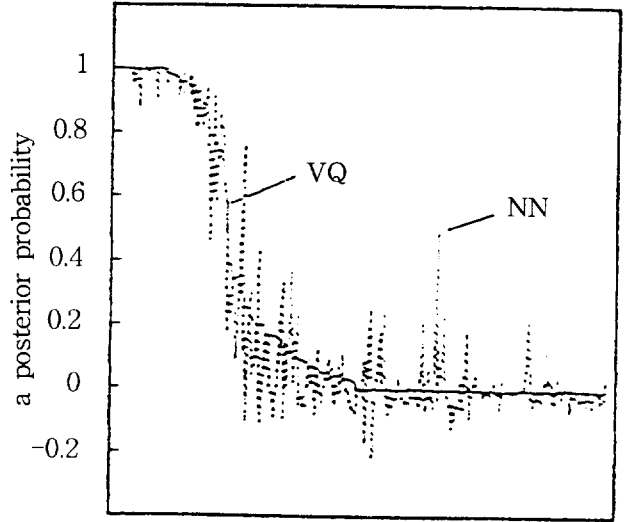
$$P(\hat{\mathbf{x}}|c) = \frac{N_{ci}}{\sum_{j=1}^m N_{cj}} \quad (6)$$

여기서 학습 데이터를 코드북에 따라 분류하고 분류된 separation set의 확률 밀도 함수를 교사신호로 학습 데이터를 구성하고 신경회로망에 학습시킨다.

#### 5. 추정된 사후 확률과 확률 밀도함수의 검토

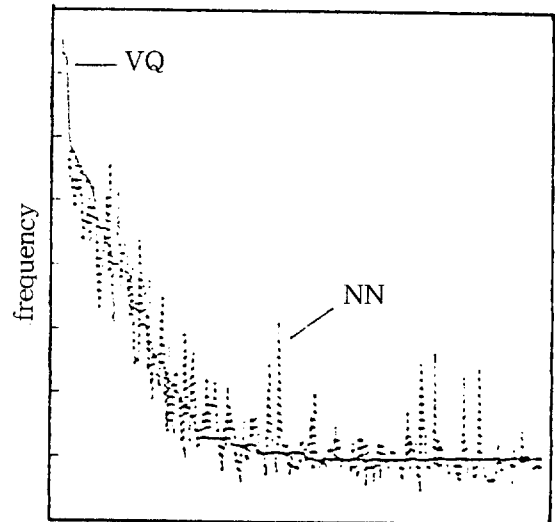
추정된 사후 확률과 확률 밀도함수의 예(모음/오)를 그림 1의 (a), (b)에 각각 나타내었다.

코드북의 크기는 256이며, 그림 1의 횡축은 VQ의 코드북의 인덱스를 큰순서로 나타내었으며 각 인덱스의 centroid 벡터를 신경회로망에 부여한 경우의 확률을 나타내고 있다. 그림 1(a)



sorted index

(a)



sorted index

(b)

그림 1 추정치 (a) 사후확률 / (b) 확률밀도 함수

Fig. 1 Estimated values (a) a posterior probability / (b) PDF

에서 신경회로망과 VQ에 의한 추정치의 다소의 차이는 추정 오차로 생각되며 그림 1(b)의 신경회로망과 VQ에 의한 추정치도 3층의 신경회로망이 임의의 분포를 비교적 근사함을 알 수 있다.

#### 6. 확률 신경회로망

학습벡터가  $X_1, X_2, \dots, X_m$  으로 주어지며 가우시안 분포를 따른다고 가정할때 Specht는 확률 밀도 함수를 다음 식과 같은 Gaussian Kernel로서 추정하였다.[5][6]

$$P(X) = \frac{1}{(2\pi)^{p/2} \sigma^p} \frac{1}{m} \sum_{i=1}^m \exp \left[ -\frac{(\mathbf{x} - \mathbf{x}_i)^T (\mathbf{x} - \mathbf{x}_i)}{2\sigma^2} \right] \quad (7)$$

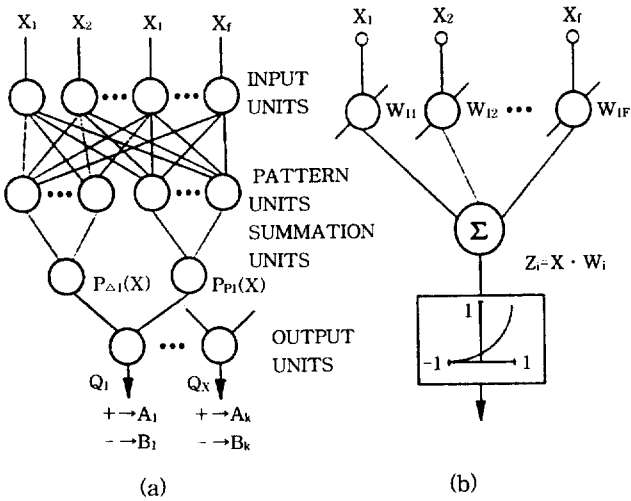


그림 2 확률 신경회로망 / (a) 구조 / (b) 패턴 unit  
 Fig. 2 Probabilistic neural network / (a) Organization / (b) The Pattern unit

여기서  $i$ 는 패턴수,  $m$ 은 학습 패턴의 수,  $X_i$ 는 카테고리에서  $i$ 번째 학습패턴,  $\sigma$ 는 스무딩 파라미터 및  $p$ 는 측정공간에서의 차원이다. 그림 2는 입력패턴  $X$ 를 두 카테고리로 분류하는 확률 신경회로망 구조를 보인다.

그림 2에서 입력 unit들은 distribution unit들이며 패턴 unit들은 다음과 같이 입력패턴  $X$ 와 가중치 벡터  $W_i$ 의 내적( $Z_i = X \cdot W_i$ )으로 주어지며 시그모이드 활성화 함수 대신에 다음의 비선형 함수를 사용한다.

$$g(\mathbf{X}) = \exp[-(\mathbf{W}_i - \mathbf{X})^t(\mathbf{W}_i - \mathbf{X})/2\sigma^2] \quad (8)$$

$X$ 와  $W_i$ 를 단위길이로 정규화 함을 가정하면 다음과 같이 쓸 수 있다.

$$g(z_i) = \exp[(z_i - 1)/\sigma^2] \quad (9)$$

이것은 식(7)의 지수함수항과 같은 형태이다. 그러므로 내부의 연결들에서 수행되는 내적은 지수적인 함수에 따르며 VQ를 사용하여 학습샘플들을 임의의 separation set( $S_1, S_2, \dots, S_k$ )로 분류하고 각 separation set에 포함된 학습샘플들의 스무딩 파라미터로서 확률 밀도함수를 구할 수 있다.

$$P(\mathbf{X}) = \frac{1}{(2\pi)^{D/2}} \sum_{k=1}^K \frac{1}{\sigma_k^p} \frac{1}{K} \exp\left[-\frac{(\mathbf{X} - \mathbf{X}_k)^t(\mathbf{X} - \mathbf{X}_k)}{2\sigma_k^2}\right] \quad (10)$$

여기서  $k$ 와  $X_k$ 는 각각 separation set  $S_k$ 에 포함된 학습샘플의 수와 학습 샘플벡터들의 평균 벡터이다. 그리고  $\sigma_k = \alpha\sigma_{sk}$ ,  $\sigma_{sk}$ 는  $S_k$ 에서 샘플들의 분산값이며  $\alpha$ 는 임의의 상수이다.

본 연구에서는 식(7)의 방법과 식(10)의 방법은 각각 전 카테

고리에 동일한 분산값을 할당하여 수행한 경우와 각 카테고리에 해당하는 분산값을 가변 할당하여 수행한 경우이다.

## 7. 실험결과 및 검토

### 7.1 음성자료 및 분석

음성자료는 표 1과 같이 6명의 화자(남4, 여2)가 발성한 13종의 음절 음성으로부터 /ga/→/a/와 같이 취한 것이며 총 2878개의 데이터 중 1400개는 학습용으로 1478개는 평가용으로 각각 사용하였다. 음성분석 방법은 표 2와 같으며 특징 파라미터는 10차 LPC mel cepstrum 계수를 이용하였다.

표 1 음성 데이터

Table 1 Speech data

5모음	아/a/	이/i/	우/u/	에/a/	오/o/	계
데이터수	620	502	682	406	668	2878

표 2 음성 분석법

Table 2 The method of speech analysis

차단 주파수	4 khz
샘플링 주파수	12 khz
해밍창 함수	21.33 ms(256 points)
프레임 주기	5 ms(60 points)
분석	14차 LPC 분석
특징 파라미터	10차 LPC mel cepstrum 계수

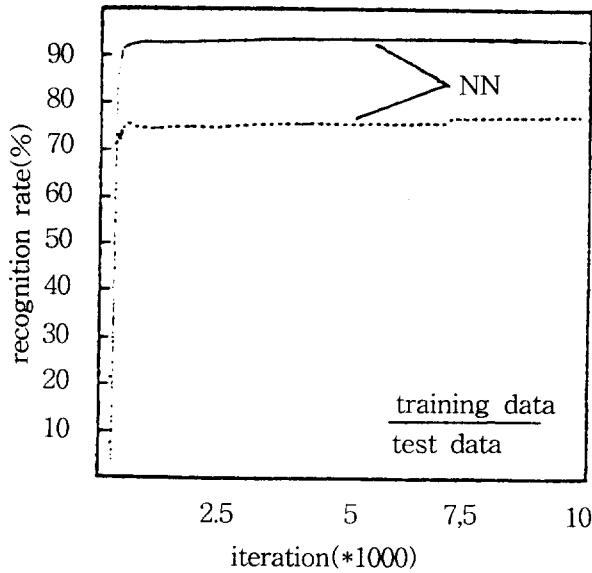
### 7.2 인식결과 및 고찰

표 3과 그림 3은 각각 사후 확률과 확률 밀도함수를 이용한 경우의 모음인식 결과들을 나타내었으며 여기서 NN은 10입력 units(10차 mel cepstrum계수), 15중간 units, 5출력 units(5모음)의 3층 신경회로망으로 BP알고리즘을 이용하였으며 PNN은 10입력 units(10차 mel cepstrum계수), 1400중간 units, 1출력 unit의 신경회로망으로 구성 하였다.

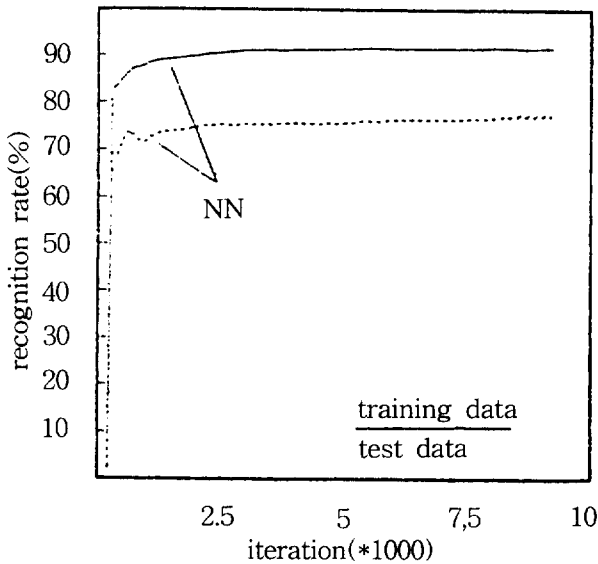
표 3 사후 확률/확률밀도함수를 이용한 인식률(%)

Table 3 Recognition rates using a posterior probability /PDF(%)

방법	코드북의 갯수	학습 데이터	평가 데이터
VQ	128	87.53 / 86.93	76.03 / 75.89
	256	89.13 / 89.02	76.93 / 76.21
NN		91.28 / 89.68	78.92 / 77.48
PNN	5*128	91.52 / 91.39	81.68 / 81.02
	5*256	92.18 / 92.14	83.02 / 82.12



(a)



(b)

그림 3 신경회로망에 의한 인식률

(a) 사후확률 (b) 확률밀도함수

Fig. 3 Recognition rates using neural network

(a) a posteriori probability (b) PDF

그림 3과 표 3으로부터 사후확률의 경우가 확률 밀도함수보다 인식률이 다소 양호함을 알 수 있으며, 또한 VQ로서 구한 코드북과 사후확률 및 확률 밀도함수와 복잡한 특징 파라미터를 PNN에 적용한 방법이 타 방법에 비해 인식률이 우수함을 알 수 있다. 특히 그림 4에서 사후확률로서 5모음\*256 코드북의 경우,  $\sigma = 0.18 \sim 0.21$ 사이의 가변구간에서 인식률이 가장 우수함을 알 수 있었다.

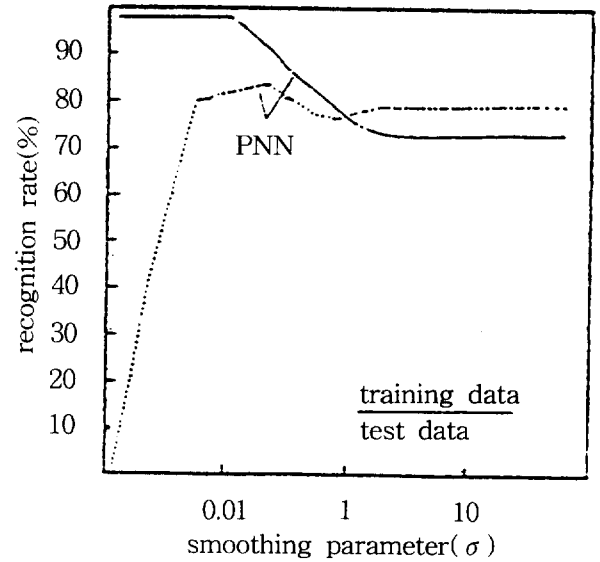


그림 4 모음평균 인식률

Fig. 4 The average recognition rates of vowels

### 8. 결 론

본 연구에서는 신경회로망과 벡터 양자화의 코드북을 이용한 사후 확률과 확률 밀도함수의 추정능력을 검토하였으며 벡터 양자화의 코드북으로 추정된 사후 확률과 확률 밀도함수를 NN 및 PNN에 적용하여 모음 인식실험을 수행하였다. 신경회로망에 의한 사후 확률 및 확률 밀도함수의 추정 능력은 벡터 양자화에 의한 것과 거의 근사하였다. 본 연구에서는 사후 확률과 확률밀도 함수의 추정하고 특징 파라미터로서의 타당성을 인식 실험을 통하여 검증하였으며 이는 향후 음소단위의 음성인식 실험에서 이미 알려진 특징 파라미터와 함께 복합 구성하여 이용하므로써 음성 인식률을 향상시키는데 응용코져 한다.

### 참 고 문 헌

- [1] D. Rumelhart, G. Hinton and R. Williams, Learning internal representations by error back propagation, In D.R. umelhart and McClelland(Eds), Parallel distributed Processing, MIT Press, 1986.
- [2] R. Hecht-Nielsen, "Theory of the back propagation neural network", IEEE International Joint Conference on Neural Networks, Vol. 1, pp. 593-605, 1989.
- [3] R. Beale and T. Jackson, Neural Computing : An Introduction, Adam Hilger Bristol, Philadelphia and New York, 1991.
- [4] B. Kosko, editor, Neural networks for Signal Processing, pp.199-223, Prentice Hall International, Inc.
- [5] D. Specht, "Probabilistic Neural Networks", Neural Networks, Vol. 3, pp. 109-118, 1990.
- [6] 中川聖一, 確率モデルによる音聲認識, 信學論, 1988.
- [7] E. A. Wan, "Neural network classification: a Bayesian interpretation", IEEE Trans., Neural Networks, Vol.1, No.4, pp.303-305, 1990.