

예측신경회로망 모델 음성인식기의

변별력있는 학습 알고리즘

○ 니 경 민*, 임 재 열*, 안 수 길*

*서울대학교 대학원 전자공학과 통신이론실

A Discriminative Training Algorithm for Speech Recognizer

Based on Predictive Neural Network Models

○ KyungMin Na*, JaeYoel RHEEM*, SouGuil ANN*

*Communication Theory Lab., Dept. of Electronics Eng., Seoul National University

요 약

예측신경회로망 모델은 다중 퍼셉트론을 연속되는 음성특징 벡터간의 비선형예측기로 사용하는 동적인 음성인식 모델이다. 이 모델은 음성의 동적인 특성을 인식에 이용하고 연속음성인식으로의 확장이 용이한 우수한 인식 모델이다. 그러나, 예측신경회로망 모델은 음운학적으로 유사한 음성구간에서의 변별력이 낮다는 문제점이 있다. 그것은 기존의 학습 알고리즘이 다른 어휘와의 거리는 고려하지 않고 대상어휘의 예측오차만 최소화시키기 때문이다. 따라서, 본 논문에서는 직접 인식오차를 최소화시키는 GPD알고리즘에 의해 유사어휘간의 거리를 고려하는 변별력있는 학습 알고리즘을 제안한다.

I. 서론

1980년대 중반부터 다중 퍼셉트론을 음성의 패턴 분류기로 사용하는 방법, TDNN(Time Delay Neural Network)을 이용한 음소인식 [9], 그리고 SOFM (Self-organizing Feature Map)을 이용한 음소인식 [10] 등 신경회로망을 음성인식에 적용하려는 여러가지 연구가 진행되어 왔다. 그러나, 기존의 연구들은 1) 음성의 동적인 특성을 제대로 이용하지 못하고 2) 새로운 학습 계층의 추가시 전체 회로망을 재학습시켜야 하며 3) 연속음성인식으로의 확장이 어려울 뿐만 아니라 4) 요구되는 학습 데이터의 양이 상대적으로 많다는 문제점들을 갖고 있다. 그러한 문제점들을 효과적으로 해결할 수 있는 새로운 음성인식 모델로 제안된 것이 예측신경회로망 모델(predictive neural network models)이다 [1-8].

일반적으로 예측신경회로망 모델에 관한 연구는 크게 1) Ken-ichi Iso 등이 제안한 NPM(Neural Prediction Model) [1-3], 2) Joe Tebelskis 등이 제안한 LPNN(Linked Predictive Neural Network) [4-5], 3) Esther Levin의 HCNN (Hidden Control Neural Network) [7-8] 등으로 분류할 수 있다. NPM과 LPNN이 시간의 변화에 따라서 회로망 자체를 전환하는데

반해서 HCNN은 입력에 추가된 제어신호의 변화로 시간에 따른 변화를 모델링한다.

예측신경회로망 모델은 여러 개의 다중 퍼셉트론들의 열로 구성되어 있다. 다중 퍼셉트론들은 입력음성신호열에 대해 예측오차행렬을 발생시키고, 그 행렬상에 시간 정렬 알고리즘을 적용시켜 전체누적예측오차가 최소화되도록 다중 퍼셉트론간의 최적궤환경로를 얻는다. 그리고, 얻어진 최적경로를 따라서 오차 역전파 학습을 수행함으로써 각각의 다중 퍼셉트론이 음성의 일부 구간을 효과적으로 예측하도록 학습시킬 수 있다.

그러나, 기존의 오차 역전파 학습 알고리즘에 의해 학습된 예측신경회로망 모델들은 음운학적으로 유사한 음성구간에서는 변별력이 떨어지는 큰 문제점을 갖고 있다. 그것은 기존의 오차 역전파 학습 알고리즘은 하나의 대상어휘를 학습시킬 때 음운학적으로 유사한 다른 어휘제출과의 거리는 고려하지 않고 대상어휘의 누적예측오차만을 최소화시키기 때문이다. 예를 들면, "삼"과 "사"의 경우 "삼"의 종성인 "ㅁ"이 상대적으로 짧고 유성음이기 때문에 두 어휘의 음성특징벡터들의 대부분이 유사하고, 따라서 기존의 학습 알고리즘으로 얻어지는 두 모델의 가중치 파라미터 집합의 대부분도 유사해져서 자주 오인식을 발생시키는 원인이 되는 것이다.

그러므로, 본 연구에서는 인식오차를 직접 최소화시키는 MCEF(Minimum Classification Error Formulation)와 GPD (Generalized Probabilistic Descent) 알고리즘을 이용하여 유사어휘간의 격리도를 높임으로써 변별력을 향상시키는 수정된 오차 역전파 학습 알고리즘을 제안한다. 제안하는 알고리즘은 대상어휘 모델에는 비용이 가중된 경사강하법을 적용시키고 대상어휘와 혼동될 수 있는 어휘 모델들에는 비용이 가중된 경사상승법을 적용시킴으로써 각각의 어휘 모델들을 더욱 격리시킨다는 물리적 의미를 갖는다. 특히 가중되는 비용은 어휘 모델간의 경계에서 최대가 되어 혼동되기 쉬울수록 더욱 많은 학습을 시키게 된다.

여러 가지 예측신경회로망 모델들 중에서 실험을 위한 모델

모든 NPM을 택했다. 그러나, 제안하는 알고리즘이 기본적으로 오차 역전파 학습의 변형임으로 다른 모델들에도 쉽게 적용할 수 있다.

II. NPM의 구조와 학습 알고리즘

NPM은 Iso 등이 제안한 예측신경회로망 모델로 그림 1과 같은 다층 퍼셉트론을 음성의 비선형예측기로 사용하고, 그러한 다층 퍼셉트론의 열로 하나의 대상어휘를 모델링하는 동적인 음성인식 모델로 오차 역전파 학습(Error Backpropagation Training)과 동적 프로그래밍(Dynamic Programming) 기법을 결합하여 학습과 인식을 수행한다.

이 모델은 음성신호의 연속되는 특징벡터간의 일시적 상관관계를 인식의 실마리로 이용하여, 시간 t에서의 음성특징벡터 s_t 를 t이전의 특징벡터들인 s_{t-1}, \dots, s_{t-L} 들의 비선형결합으로 예측하여 \hat{s}_t 를 얻은 후 그로부터 예측오차 $(s_t - \hat{s}_t)^2$ 의 최소누적계리를 계산하여 인식의 척도로 이용한다. 다층 퍼셉트론 예측기의 입력역 관계는 다음의 식들로 주어진다.

$$h_t = f \left(\sum_{i=1}^L U_i s_{t-i} \right) \quad (1)$$

$$\hat{s}_t = U_0 h_t \quad (2)$$

U_0, U_1, \dots, U_L 는 가중치행렬, h_t 는 은닉층의 출력, $f(x)$ 는 sigmoid function, \hat{s}_t 는 예측된 특징벡터, $(s_t - \hat{s}_t)^2$ 는 예측 오차이고 이와같은 다층 퍼셉트론의 열로 하나의 어휘를 모델링한다. 한 어휘의 전체적인 모델은 그림 2와 같다.

예측시에는 입력음성특징 벡터들을 N개의 부분으로 나누고 n번째 예측기로 n번째 부분을 예측시킨다. 이때 입력음성의 최적분할은 동적 프로그래밍 기법에 의해 누적예측오차 $D(w)$ 를 최소화시키도록 결정된다.

$$D(w) = \min_{n(t)} \sum_{i=1}^n \| \hat{s}_i(w, n(t)) - s_{it} \|^2 \quad (3)$$

동적 프로그래밍 기법에 의한 최적분할경로는 backtracking에 의해 결정되며 그 경로를 따라서 오차 역전파 학습을 수행함으로써 NPM을 효과적으로 학습시킬 수 있다. 그림 3에 학습과정의 예를 시각적으로 보인다. 또한, 이 모델을 subword 단위의 모델로 만들어 연결하면 쉽게 연속음성인식에 확장하여 적용할 수 있다 [2].

2 프레임의 입력으로 다음 프레임울 예측하는 2차의 예측기를 사용하고, M개의 어휘 모델 $C^i, i=1, 2, \dots, M$, 입력벡

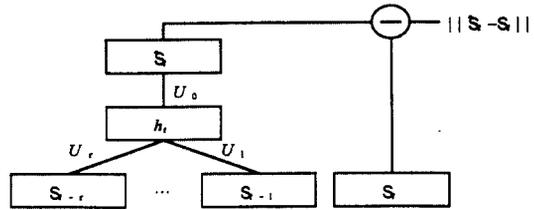


그림 1 다층 퍼셉트론 예측기

Fig. 1 MLP predictor

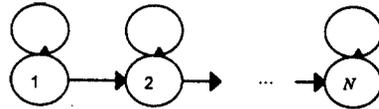


그림 2 NPM으로 구성된 하나의 어휘 모델

Fig. 2 A word model by NPM

터열 $S=(S_1, S_2, \dots, S_r)$, $S_t=(s_{1,t}, s_{2,t}, \dots, s_{K,t})$, N개의 다층 퍼셉트론, 그리고 최적분할경로 $(t, n(t))$, $1 \leq n(t) \leq N$ 가 주어질때 학습 대상 어휘가 C^m 에 포함되면 학습 알고리즘은 다음과 같다.

$$(W_{\hat{s}_t}^m)_{q,1} = (W_{\hat{s}_t}^m)_q + \eta (s_{k,t-2} - \hat{s}_{k,t-2}) H_{kt}^m \quad (4)$$

$$(V_{\hat{s}_t}^m)_{q,1} = (V_{\hat{s}_t}^m)_q + \eta \delta_{kt}^m H_{kt}^m (1.0 - H_{kt}^m) \bar{s}_{kt} \quad (5)$$

$$\text{where } \delta_{kt}^m = \sum_{k=1}^K (s_{k,t-2} - \hat{s}_{k,t}) W_{\hat{s}_t}^m$$

위에서 \bar{s}_{kt} 는 S_t 와 S_{t-1} 로 구성된 입력벡터의 i번째 성분, H_{kt}^m 는 은닉층의 출력, $\hat{s}_{k,t-2}$ 는 출력벡터의 k번째 성분, 그리고 $V_{\hat{s}_t}^m$ 과 $W_{\hat{s}_t}^m$ 는 각각 n(t)번째 다층 퍼셉트론의 은닉층과 출력층의 가중치이다.

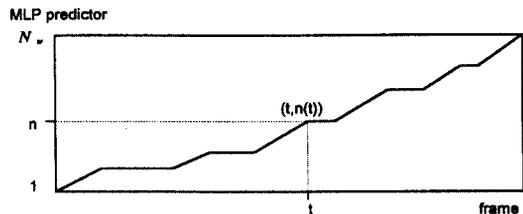


그림 3 동적 프로그래밍에 의한 예측오차최소화학습과정

Fig. 3 Prediction error minimization process by DP

III. MCEF와 GPD 알고리즘

예측신경회로망모델 음성인식기의 변별력있는 학습 알고리즘

최근에 인식오차를 직접 최소화시키는 MCEF(Minimum Classification Error Formulation)와 GPD(generalized probabilistic descent) 알고리즘을 사용하여 DTW나 HMM을 이용한 인식기들의 변별력을 높이는 학습 알고리즘들이 제안되었다 [11-13]. 본 논문에서는 그와 같은 MCEF와 GPD 알고리즘을 예측신경회로망 모델에 적합하도록 수식을 전개하여 새로운 학습 알고리즘을 유도하고자 한다.

최적의 인식기를 설계하는 기준을 얻기위해서 다음과 같은 세 단계의 척도함수들의 정의가 필요하며 이러한 과정을 MCEF(Minimum Classification Error Formulation)이라 한다.

단계 1. 먼저 음이 아닌 변별함수(discriminant function) $g_i(x; \Lambda)$ 를 정의하여 인식의 결정규칙으로 사용한다.

$$C(x) = C^j, \text{ for } g_i(x; \Lambda) = \max_j g_j(x; \Lambda) \quad (6)$$

단계 2. 오분류함수(misclassification function) $d_k(x; \Lambda)$ 를 정의하여 결정과정을 함수의 형태로 표현한다. 이 오분류 함수의 도입이 기존의 방법들과 가장 큰 차이점이라 할 수 있다. 많은 종류의 오분류함수들을 정의할 수 있으나, 일반적인 오분류함수는 다음과 같이 정의 된다.

$$d_k(x; \Lambda) = -g_k(x; \Lambda) + \left\{ \frac{1}{M-1} \sum_{j \neq k} g_j(x; \Lambda) \right\}^{\frac{1}{2}} \quad (7)$$

수식을 음미해 보면 오분류함수의 값이 클수록 입력이 더욱 오분류되었다고 볼 수 있다는 것을 알 수 있다. 이를 적당히 조절함으로써 인식기의 파라미터를 최적화시키기위해 참여하는 대상어휘계층이 결정된다. 극단적인 예로써 이를 무한히 크게 잡으면 수식이 다음과 같이 근사된다.

$$d_k(x; \Lambda) = -g_k(x; \Lambda) + g_j(x; \Lambda) \quad (8)$$

즉, 대상어휘 k에 가장 유사한 하나의 어휘계층 j' 만 변별력있는 학습에 참여시키는 것이다.

단계 3. 오인식률을 계산하기위한 비용함수(cost function)를 정의한다. 오분류함수와 함께 고려해볼때 비용함수는 오차확률의 근사로 이해될 수 있다. 몇가지 종류의 비용 함수들을 고려할 수 있으나 일반적으로 다음의 sigmoid function을 사용하여 비용(cost)을 0 에서 1 사이의 값으로 smoothing 시킨다.

$$I_k(x; \Lambda) = I_k(d_k(x; \Lambda)) = \frac{1}{1 + e^{-\alpha d}} \quad (9)$$

위에 정의된 수식들은 모두 파라미터 Λ 에 연속인 함수들이 다. 그 이유는 경사강하법(gradient descent)을 사용하는데 있어서 미분함수를 얻기 용이하기 때문이다. 이상의 수식화가 경사강하법(gradient descent)에 의해 오인식의 기대치 $L(\Lambda)$ 를 최소화 시킬 수 있는 학습의 기초를 제공한다.

식 (9)에서 얻은 비용함수의 기대치를 최소화시키기위해 확률적 강하법을 사용하는 것을 GPD 알고리즘이라 한다. 다음에 그 과정을 수식으로 보인다.

$$L(\Lambda) = E\{I_k(X; \Lambda)\} \quad (10)$$

$$\Lambda_{t+1} = \Lambda_t + \delta \Lambda_t \quad (11)$$

$$\text{where } \delta \Lambda_t(x, \Lambda) = -\eta U \nabla I_k(x; \Lambda)$$

U 는 positive-definite matrix이고, η 를 시간에 따라서 줄여 나감으로써 국부최소오차를 얻을 수 있음이 증명되어 있다 [11]. 본 논문에는 $\eta = \eta_0 \left(1.0 - \frac{t}{T}\right)$, U 는 항등행렬을 사용하였다.

IV. 예측신경회로망 모델의 변별력있는 학습알고리즘

III장의 MCEF와 GPD 알고리즘으로부터 예측신경회로망 모델에 적용하기위한 수식들을 유도한다. 모든 수식의 표기는 II장의 기존의 학습 알고리즘을 유도할 때 사용한 것과 같다.

먼저, 변별함수는 하나의 입력어휘 $S \in C^i$ 에 대해서 얻어지는 예측신경회로망의 최소누적오차로 정의한다.

$$g_i(x; W; V) = \ln \left\{ \sum_{n=1}^{N_n} e^{-\left(\sum_{m=1}^M d_k(x, n; m) \right) \rho} \right\}^{\frac{1}{\rho}} \quad (12)$$

$D_i^{\rho}(t, n(t))$ 는 θ 번째 최적경로를 통해서 얻어지는 $\{t, n(t)\}$ 에서의 예측오차이다. $\rho \rightarrow \infty$ 인 경우는 최적경로 θ^* 상의 최소누적오차가 되고 식 (12)는 식 (13)이 된다. 본 논문에서는 식 (13)을 사용하였다.

$$g_i(x; W; V) = \min_{n(t)} \sum_{t=1}^{T} D_i^{\theta^*}(t, n(t)) \quad (13)$$

$$\text{where } D_i^{\theta^*}(t, n(t)) = \sum_{k=1}^{K} (s_{k,t,2} - \hat{s}_{k,t,2})^2$$

$$\hat{s}_{k,t,2} = \sum_{j=1}^J H_{j,t} \cdot W_{j,k,m,t}$$

$$H_{j,t} = f \left(\sum_{l=1}^L \bar{s}_{l,t} \cdot V_{l,m,t} \right)$$

본래의 오분류함수는 식(7)로 정의되나 본 모델에서는 분류 함수가 최소인 모델을 인식함으로써 아래의 식 (14)과 같이 변형 된다.

$$d_m(x;W;V) = g_m(x;W;V) - \ln \left\{ \frac{1}{M-1} \sum_{i \neq m} e^{-g_i(x;W;V)} \right\} \quad (14)$$

ξ를 무한히 크게 하면 식은 다음과 같이 근사된다.

$$d_m(x;W;V) = g_m(x;W;V) - g_{i'}(x;W;V) \quad (15)$$

비용함수는 일반적인 sigmoid function을 사용한다.

$$f_m(x;W;V) = \frac{1}{1 + e^{-\alpha x}} \quad (16)$$

마지막으로 평균적인 오인식의 척도로 비용함수의 기대치 $L(W;V)$ 를 사용하여 $L(W;V)$ 를 최소화시키는 방향으로 학습을 진행시킨다.

$$L(W;V) = \sum_{k=1}^K l_k(x;W;V) \quad (17)$$

$$\frac{\partial L(W;V)}{\partial W_{k,n}^m}, \quad \frac{\partial L(W;V)}{\partial W_{k,n}^{i'}}, \quad \frac{\partial L(W;V)}{\partial V_{k,n}^{i'}} \quad \text{그리고}$$

$$\frac{\partial L(W;V)}{\partial V_{k,n}^{i'}} \quad \text{와 식 (12)에서 (16)으로부터 변별력있는 학습 알고리즘의 식을 구하면 다음과 같다.}$$

$$\delta W_{k,n}^m = \eta \alpha l_m (1 - l_m) (s_{k,t+2} - \hat{s}_{k,t+2}) H_{j,t}^m \quad \text{for } S \in C^m \quad (18)$$

$$\delta V_{k,n}^{i'} = \eta \alpha l_m (1 - l_m) \delta_{j,t}^m H_{j,t}^m (1 - H_{j,t}^m) \bar{s}_{i,t} \quad \text{for } S \in C^m \quad (19)$$

$$\text{where } \delta_{j,t}^m = \sum_{k=1}^{K-K} (s_{k,t+2} - \hat{s}_{k,t+2}) W_{k,n}^m$$

$$\delta W_{k,n}^{i'} = -\eta \alpha l_m (1 - l_m) v_i (s_{k,t+2} - \hat{s}_{k,t+2}) H_{j,t}^{i'} \quad \text{for } i' \neq m \quad (20)$$

$$\delta V_{k,n}^{i'} = -\eta \alpha l_m (1 - l_m) v_i \delta_{j,t}^m H_{j,t}^{i'} (1 - H_{j,t}^{i'}) \bar{s}_{i,t} \quad \text{for } i' \neq m \quad (21)$$

$$\text{where } \delta_{j,t}^m = \sum_{k=1}^{K-K} (s_{k,t+2} - \hat{s}_{k,t+2}) W_{k,n}^m$$

$$v_i = \frac{e^{-g_i(x;W;V)}}{\sum_{i \neq m} e^{-g_i(x;W;V)}}$$

ξ→∞인 경우는 아래와 같이 된다. 이때 식 (20)와 (21)의 절자 i' 은 대상어휘계층을 제외한 모든 계층을 나타내지만, 식 (24)와 (25)의 절자 i' 은 대상어휘의 가장 혼동되기 쉬운 하나의 계층만을 나타낸다는 것에 주의해야한다.

$$\delta W_{k,n}^m = \eta \alpha l_m (1 - l_m) (s_{k,t+2} - \hat{s}_{k,t+2}) H_{j,t}^m \quad \text{for } S \in C^m \quad (22)$$

$$\delta V_{k,n}^{i'} = \eta \alpha l_m (1 - l_m) \delta_{j,t}^m H_{j,t}^m (1 - H_{j,t}^m) \bar{s}_{i,t} \quad \text{for } S \in C^m \quad (23)$$

$$\text{where } \delta_{j,t}^m = \sum_{k=1}^{K-K} (s_{k,t+2} - \hat{s}_{k,t+2}) W_{k,n}^m$$

$$\delta W_{k,n}^{i'} = -\eta \alpha l_m (1 - l_m) (s_{k,t+2} - \hat{s}_{k,t+2}) H_{j,t}^{i'} \quad \text{for } i' \neq m \quad (24)$$

$$\delta V_{k,n}^{i'} = -\eta \alpha l_m (1 - l_m) \delta_{j,t}^m H_{j,t}^{i'} (1 - H_{j,t}^{i'}) \bar{s}_{i,t} \quad \text{for } i' \neq m \quad (25)$$

$$\text{where } \delta_{j,t}^m = \sum_{k=1}^{K-K} (s_{k,t+2} - \hat{s}_{k,t+2}) W_{k,n}^m$$

학습은 모두 동적 프로그래밍에 의한 각각의 최적경도 $(t, n(t))$ 와 $(t, n^*(t))$ 상에서 수행된다.

유도된 학습 알고리즘의 식 (22)에서 (25)를 음미해 보면 학습어휘계층 C^m 의 파라미터 집합에 대해서는 비용치(cost)를 가진 경사강하법(gradient descent)의 학습을 시키고, 오인식될 가능성이 가장 높은 어휘계층 $C^{i'}$ 의 파라미터 집합에 대해서는 비용치를 가진 경사상승법(gradient ascent)의 학습을 시킴으로써 두 어휘간의 격리도를 높인다. 그리고, 가중되는 비용 $\alpha l_m (1.0 - l_m)$ 은 $l_m = 0.5$ 에서 최대가 되고 식 (16)와 (15)으로부터, 각각 $d_m = 0, g_m = g_{i'}$ 이 되므로 유사한 두 어휘간의 경계에서 가중되는 비용이 최대가 됨을 알 수 있다.

V. 실험결과

실험에 사용된 대상어휘는 한국어 숫자음이고 음성신호의 처리과정은 다음과 같다. 우선 7명의 화자를 선정하여 각 화자가 3회씩 발성한 숫자음을 Kay의 CSL을 이용하여 10kHz의 표본화율로 녹음한 후 Rabiner와 Sambur의 글점 검출 알고리즘으로 글점을 검출해낸다. 그 다음에 25.6ms의 분석구간을 12.8 ms씩 중첩시키면서 pre-emphasis, 25.6ms Hamming 창, 그리고 12 차의 LPC 분석을 통해서 0 차의 렉스트림 계수를 제외한 12 차의 가중렉스트림 계수들을 얻어서 예측신경회로망의 입력으로 사용한다. LPC 분석은 자기상관계수를 사용하는 Durbin 알고리즘에 의해 수행되었고, LPC 계수와 렉스트림 계수간의 직접적인 관계로부터 12 차의 렉스트림 계수를 얻는다.

7명의 화자중 5명이 1회 발성한 50개의 음성 데이터들 위와 같이 처리하여 학습에 사용하였고 나머지는 인식실험에 사용했다. 실험은 식 (22)에서 (25)를 이용해서 두 유사어휘간의 격리도만을 가정해서 실시했다. 학습에 참가한 화자의 비학습 데이터와 학습에 참가하지 않은 화자의 데이터에 대해서, 기존의 학습 알고리즘을 사용하였을 경우 각각 97%와 88.3%의 인식률을 얻었고, 제안한 학습 알고리즘에 의해서는 각각 99%와 90%의 인식률을 얻었다. 따라서, 제안한 학습 알고리즘이 기존의 것보다 우수함을 알 수 있다.

특히, "삼"과 "사" 사이의 오인식은 모두 정정되었으며 다른 오인식은 모두 "공", "오", "구" 사이에서 발생하였다. 그 이

예측신경회로모델 음성인식기의 변별력있는 학습 알고리즘

유는 오분류함수의 정의에 유사한 어휘가 2 개만 존재한다는 가정을 세웠기 때문에, 식 (14)의 ϵ 를 적당히 조절함으로 더 많은 유사어휘를 학습에 참가시킬 수 있다.

하나의 어휘는 10 개의 MLP 예측기로 구성되어 있고, 학습 계수 η 는 0.005, α 는 0.1을 사용하였고 전체학습회수는 400 회이다.

표 1. 인식실험결과

Table 1. Comparison of recognition experiment

실험 데이터	인식률	
	기존 방법	제안한 방법
학습참가자의 비학습	97 %	99 %
학습미참가자	88.3 %	90 %

VI. 결론

본 논문에서는 MCEF와 GPD 알고리즘에 기초해서 예측신경회로모델의 변별력을 향상시키기 위한 새로운 학습 알고리즘의 수식들을 유도하고 실험을 통해서 예측한대로 오인식되는 어휘간의 격리도를 높여서 인식률을 향상시켰다. 실험에 사용된 모델은 NPM이었으나 다른 모델들에도 같은 식이 적용되므로 일반성이 있는 알고리즘이다.

식 (18)에서 (25)까지에서 알 수 있듯이 유도된 학습 수식은 어휘간에 혼동될 수 있는 정도에 따라서 적절한 비용을 학습에 참가시키고, 올바른 대상어휘의 모델에는 경사강화법을 적용시키고 다른 모델들에는 경사상승법을 적용시킴으로써 모델간의 변별력을 높인다.

앞으로 제안한 알고리즘과 또다른 기법들을 사용하여 더욱 변별력을 향상시킬 수 있는 알고리즘의 개발도 가능하다. 예를 들면, 각각의 예측기에서 얻은 예측오차행렬에 동적 프로그래밍 기법을 적용시킬때 적절한 가중치함수를 가하는 방법을 생각할 수 있다. 이것은 기존의 DTW에서도 사용하는 방법으로서 그러한 가중치함수도 GPD 알고리즘을 이용하여 얻을 수 있으리라 생각된다.

참 고 문 헌

[1] K. Iso and T. Watanabe, "Speaker-independent word recognition using a neural prediction model," *Proc. ICASSP-90*, pp. 441-444, 1990.
 [2] K. Iso and T. Watanabe, "Speech recognition using demi-syllable neural prediction model," *Proc. ICSP-90*, pp. 661-664, 1990.
 [3] K. Iso and T. Watanabe, "Speech recognition using dynamical model of speech production," *Proc. ICASSP-*

93, pp. 283-286, 1993.
 [4] J. Tebelskis and A. Waibel, "Large vocabulary recognition using linked predictive neural networks," *Proc. ICASSP-90*, pp. 437-440, 1990.
 [5] J. Tebelskis, A. Waibel, B. Petek and O. Schmidbauer, "Continuous speech recognition using linked predictive neural networks," *Proc. ICASSP-91*, pp. 61-64, 1991.
 [6] N. Fishby, "A dynamical systems approach to speech processing," *Proc. ICASSP 90*, pp. 365-368, 1990.
 [7] E. Levin, "Word recognition using hidden control neural architecture," *Proc. ICASSP-90*, pp. 433-436, 1990.
 [8] E. Levin, "Hidden control neural architecture modeling of nonlinear time varying systems and its applications," *IEEE Trans. Neural Networks*, vol. 4, no. 1, pp. 109-116, 1993.
 [9] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano and K. J. Lang, "Phoneme recognition using time-delay neural networks," *IEEE Trans. ASSP*, vol. 37, pp. 328-339, 1989.
 [10] T. Kohonen, *Self-organization and associative memory*, second edition, Springer-Verlag, Berlin, 1987.
 [11] P. C. Chang and B. H. Juang, "Discriminative training of dynamic programming based speech recognizers," *IEEE Trans. Speech and Audio Processing*, pp. 135-143, 1993.
 [12] W. Chou, B. H. Juang and C. H. Lee, "Segmental GPD training of HMM based speech recognizer," *Proc. ICASSP-92*, pp. 473-476, 1992.
 [13] B. H. Juang and S. Katagiri, "Discriminative learning for minimum error classification," *IEEE Trans. Signal Processing*, pp. 3043-3054, 1992.