

論文95-32B-11-14

회귀신경예측 모델을 이용한 음성인식

(Speech Recognition Using Recurrent Neural Prediction Models)

柳濟寬*, 羅景民*, 林材烈***, 成宏模**, 安秀桔**

(Je Kwan RYU, Kyung Min NA, Jae Yeol RHEEM, Goeng Mo Sung, and Souguil Ann)

요 약

본 논문에서는 음성인식을 위한 새로운 신경회로망 모델로서 회귀신경회로망(recurrent neural networks)을 음성의 비선형 예측기로 학습시키는 회귀신경예측모델(recurrent neural prediction models: RNPM)을 제안한다. RNPM은 망의 내부표현에 의해 효과적으로 사상능력을 변조하며, 시간축정렬(time alignment)알고리즘이 불필요하므로 기존의 예측신경회로망에 비해 인식단계의 계산량이 현저히 감소되며, 요구하는 기억장소의 크기도 작다. 또한 시변성 표적함수를 결정하는 문제가 자연스럽게 해결된다. 다양한 조건 하에서 한국어 숫자음에 대한 화자 종속 및 화자 독립 인식실험 결과, 제안된 모델은 예측신경회로망에 비해 위의 장점을 지니면서 인식률에 있어서 그에 필적할 만한 결과를 얻었다.

Abstract

In this paper, we propose recurrent neural prediction models (RNPM), recurrent neural networks trained as a nonlinear predictor of speech, as a new connectionist model for speech recognition. RNPM modulates its mapping effectively by internal representation, and it requires no time alignment algorithm. Therefore, computational load at the recognition stage is reduced substantially compared with the well known predictive neural networks (PNN), and the size of the required memory is much smaller. And, RNPM does not suffer from the problem of deciding the time varying target function. In the speaker dependent and independent speech recognition experiments under the various conditions, the proposed model was comparable in recognition performance to the PNN, while retaining the above merits that PNN doesn't have.

I. 서 론

음성인식은 기계와 인간사이의 통신에 필수적인 기술이다. 현재까지 DTW (Dynamic Time Warping), HMM (Hidden Markov Models), 신경회로망(ne-

ural networks) 등이 음성인식을 위해서 주로 연구되어 왔다. DTW는 화자의 발성시간의 차이를 효과적으로 흡수하나 각 화자간의 음성학적 차이를 모델링하는데 한계가 있고, HMM은 확률적인 모델로서 최근에 가장 널리 사용되고 있으나 요구되는 수학적 가정들이 현실에 맞지 않는다는 문제점이 있다.

최근에 다양한 신경회로망 구조가 패턴분류, 시스템 모델링, 신호처리와 같은 많은 분야에 성공적으로 도입되어져 왔다. 그러나, 대부분의 경우 모델의 파라미터(parameters)가 시간에 대해서 고정되어 있기 때문에, 음성과 같은 시변성(time-varying) 신호를 다루

* 準會員, ** 正會員, 서울大學校 電子工學科

(Dept. of Elec. Eng., Seoul Nat'l Univ)

*** 正會員, 韓國技術教育大學校 電子工學科

(Korea Inst. of Tech. and Education)

接受日字: 1995年5月22日, 수정완료일: 1995年11月9日

기에는 적합하지 못하다. 따라서, 음성인식에의 적용을 위해 시변성 신호를 다룰 수 있는 몇가지 신경회로망 구조가 제안되었다^{[11]-[19]}. Waibel 등에 의해 제안된 TDNN (Time-Delay Neural Network)은 입력층과 은닉층에 시간적 변이를 흡수할 수 있는 구조를 도입하였다^[11]. 전통적인 동적 프로그래밍 기법과 신경회로망과의 접목은 Sakoe 등에 의한 DNN (Dynamic Programming Neural Network)에 처음으로 도입되었다^[21]. DNN에서는 신경회로망이 변별기 (discriminator)로 사용되고, 동적 프로그래밍 기법은 시간축정렬을 위해서 사용된다. DNN 이후, 동적 프로그래밍 기법과 음성의 비선형 예측기로 학습되는 신경회로망을 결합한 여러가지 예측신경회로망들이 제안되었다^{[14]-[19]}.

예측신경회로망모델은 다층퍼셉트론(MLP: Multi-Layer Perceptron)들의 열로 구성되어 있으며, 연속되는 음성특징 벡터들간의 시간적상관관계(temporal correlations)를 각 MLP 비선형예측기(nonlinear predictor)들의 입출력으로 사용한다. 그리고, 그때에 발생하는 음성의 일시적 왜곡은 동적프로그래밍 기법을 사용하여 효과적으로 정규화시킴으로써 학습과 인식을 수행한다. 일반적으로 예측신경회로망에 대한 연구는 크게 Levin에 의한 HCNN (Hidden Control Neural Network) [4], [5], Iso 및 Watanabe [6], [7] 와 Tebelskis 등 [8], [9] 에 의해 제안된 NPM (Neural Prediction Model)과 LPNN (Linked Predictive Neural Network) 등으로 분류할 수 있다. 예측신경회로망모델은 시간적 상관관계를 인식에 이용하고, 연속음 인식에의 적용 및 새로운 계층의 첨가가 용이하다는 장점이 있으나 동적프로그래밍에 의해 최적경로를 찾는 과정에서 많은 계산량을 요구하는 시간축정렬(time alignment algorithm) 알고리즘을 필요로 한다. 이러한 계산과정은 모든 계층(class)의 모델들에 대해 모두 행해져야 함으로 예측신경회로망은 많은 계산량의 문제를 안고 있다.

시변성 신호를 모델링하는 또 다른 신경회로망 구조로 최근에 많이 연구되고 있는 회귀신경회로망 (Recurrent Neural Networks: RNN)을 들 수 있다 [10]-[13]. 회귀신경회로망은 회귀연결 구조가 시간적인 변이를 흡수할 수 있는 단기간의 메모리 (short-term memory)를 제공하기 때문에 망의 내부 표현(internal representation)에 의해 사상능력을

변조할 수 있는 구조적인 특징을 지니고 있다. 또한 학습 및 인식에 사용되는 입력신호의 길이에 제한을 두지 않는다. 이러한 특징으로 인해 예측신경회로망이 지니고 있던 많은 계산량을 요구하는 시간축정렬 알고리즘은 불필요하다. 따라서 인식 단계에서의 계산량이 현저히 감소된다. Bourlard와 Wellekens는 음성인식을 위한 여러가지 회귀신경회로망에 대해서 비교하였다^[1]. 회귀 신경 회로망에 대한 학습 알고리즘은 Williams와 Zipser에 의해 제안되었는데 [11]. 이후 S. J. Lee 등은 부분적인 회귀 구조를 갖는 망에 확장된 델타 룰(delta rule)을 근사적으로 사용하여 계산량을 현저히 감소시켰다^[12].

기존의 회귀신경회로망들은 적절한 시변성 표적함수(time-varying target functions)와 함께 변별기로서 학습되어져 왔다. 발생시간 동안 0.5에서 1.0까지 증가하는 램프 함수가 시간에 따라 변하는 표적함수로서 Watrous와 Shastri에 의해 제안된 바 있다^[13]. 그러나, 적절한 표적함수의 선택은 현재까지는 이론적 근거가 없이 경험적으로 이루어지고 있다. 그 결과 회귀신경회로망의 적용대상도 매우 제한되어 있다.

이 논문에서 우리는 음성 인식을 위한 새로운 신경회로망 구조인 회귀신경예측 모델(Recurrent Neural Prediction Model: RNPM)을 제안한다. 제안된 모델은 예측신경회로망에 비해 시간축정렬 알고리즘이 불필요하여 인식시간이 단축되고, 작은 크기의 기억장소를 요구한다. 또, 시변성 표적함수 설정의 문제가 자연스럽게 해결된다. 그러므로, 연속음 인식 및 대응량 어휘 인식기의 구축에 기존의 모델들에 비해 많은 장점을 보유한다. 예측기로서 학습되기 때문에, RNPM의 출력층은 선형출력함수를 지니며, 전체 회로망의 안정도를 보장하기 위해 출력층으로부터의 회귀 연결은 제거하였다. 본 논문에서는 두 가지 회귀 구조에 대한 모델을 제시하고 이에 따른 인식 실험을 행하였다. 하나는 단층회귀신경예측 모델 (Single-layer Recurrent Neural Prediction Model: SRNPM)이고 다른 하나는 다층회귀신경예측 모델(Multi-layer Recurrent Neural Prediction Model: MRNPM)이다. 학습 알고리즘은 잘 알려진 실시간 회귀학습(Real Time Recurrent Learning: RTRL)을 모델에 맞게 적용하여 유도하였다. 인식단계에서는 최소의 누적예측오차를 지닌 모델이 인식결과가 된다.

논문의 구성은 다음과 같다. 2 절에서는 제안된

RNPM을 학습 알고리즘과 함께 설명하였다. 한국어 숫자음에 대한 음성인식실험결과는 3 절에 보였고, 마지막 절에 결론을 기술하였다.

II. 회귀신경예측모델과 학습 알고리즘

1. 단층회귀신경예측 모델(SRNPM)과 학습 알고리즘

SRNPM은 두 층으로 구성된다. 입력층에는 외부 입력 뉴런과 회귀 입력 뉴런이 있고, 출력층에는 은닉 출력 뉴런과 외부 출력 뉴런이 있다. 회귀 연결은 은닉 출력 뉴런과 회귀 입력 뉴런 사이에만 존재한다. 은닉 출력 뉴런에는 시그모이드(sigmoid)출력 함수가 있고, 외부 출력 뉴런에는 선형 출력 함수가 있다. 망의 구조는 그림 1과 같다.

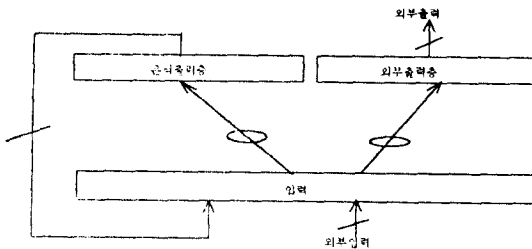


그림 1. 단층회귀신경예측 모델(SRNPM)의 구조
Fig. 1. Structure of simple recurrent neural prediction model (SRNPM).

$S^T = \{s(1), \dots, s(T)\}$ 를 길이 T 의 학습 신호라 할때, $s(t) = (s_1(t), \dots, s_N(t))$ 는 시간 t 에서의 N 차원 음성 특징 벡터가 된다. 시간 t 에서의 표적 벡터를 $s(t)$ 라 할때, 예측 차수가 τ 인 경우, 외부 입력 벡터인 $x(t) = (x_1(t), \dots, x_{N+\tau}(t))$ 은 $s(t-1), \dots, s(t-\tau)$ 의 τ 개의 벡터를 연결한 벡터이다. M 개의 회귀 연결(또는 M 개의 은닉 출력 뉴런)이 존재할 때, $y(t) = (y_1(t), \dots, y_{M+N}(t))$ 와 $o(t) = (o_1(t), \dots, o_{M+N}(t))$ 를 각각 비선형 출력함수의 전체입력(net input)벡터와 출력 벡터라 하자. $y(t)$ 와 $o(t)$ 의 j 번째 성분인 $y_j(t)$ 와 $o_j(t)$ 는 다음과 같다.

$$y_j(t) = \sum_{i=1}^{M+N\tau} z_i(t)w_{ij} \quad \begin{cases} j=1, \dots, M+N\tau \\ j=1, \dots, M+N \end{cases} \quad (1)$$

$$o_j(t) = f(y_j(t)) \quad (2)$$

$$f(y_j(t)) = \begin{cases} \frac{1}{1 + \exp(-y_j(t))} & j=1, \dots, M \\ y_j(t) & j=M+1, \dots, M+N \end{cases} \quad (3)$$

w_{ij} 는 입력층의 i 번째 뉴런에서 출력층의 j 번째 뉴런으로의 연결의 가중치이고, $z_i(t)$ 는 $o(t-1)$ 과 $x(t)$ 를 연결한 벡터인 $z(t) = (z_1(t), \dots, z_{M+N}(t)) = (o_1(t-1), \dots, o_M(t-1), x_1(t), \dots, x_{N+\tau}(t))$ 의 i 번째 성분이다. $s(t)$ 를 예측한 벡터인 $\hat{s}(t) = (\hat{s}_1(t), \dots, \hat{s}_N(t))$ 는 $(o_{M+1}(t), \dots, o_{M+N}(t))$ 와 같다.

예측 오차 벡터 $e(t) = s(t) - \hat{s}(t)$ 로부터 정의되는 시간 t 에서의 전체오차인 $J(t)$ 와 발생 시간 T 동안의 누적 예측 오차인 E 는 다음과 같다.

$$J(t) = \frac{1}{2} \sum_{k=1}^N [e_k(t)]^2 \quad (4)$$

$$E = \sum_{t=1}^T J(t) \quad (5)$$

학습알고리즘은 Williams와 Zipser에 의한 RTRL (Real Time Recurrent Learning)을 모델에 맞게 변형하여 유도하였다^[11]. RTRL에서는 실시간으로 가중치를 학습시키기 위해 $J(t)$ 에 경사강화법을 적용하여 학습식을 유도한다. 이 경우, 누적예측오차 E 를 최소화함을 보장할 수는 없지만, 학습계수를 작게할 경우 거의 같은 결과를 얻을 수 있음이 역시 Williams 등에 의해 보여진 바 있다^[11]. 학습계수를 α 라 하면, 식 (2)로부터 $\nabla J(t)$ 를 계산하면 시간 t 에서의 가중치 학습량은,

$$\Delta w_{ij}(t) = -\alpha \frac{\partial J(t)}{\partial w_{ij}} \quad (6)$$

와 같이 얻어진다. 식 (4)를 계산하기 위한 다음의 회귀식을 식 (2)와 (4)로부터 얻을 수 있다.

$$\frac{\partial o_k(t+1)}{\partial w_{ij}} = f'(y_k(t)) \left[\sum_{l=1}^{M+N} w_{il} \frac{\partial o_l(t)}{\partial w_{ij}} - \delta_{kl} z_l(t) \right] \quad (7)$$

망의 초기 상태는 가중치에 대해 함수적으로 독립이므로, 다음의 초기 조건을 얻는다.

$$\frac{\partial o_i(0)}{\partial w_{kl}} = 0 \quad (8)$$

2. 다층회귀신경예측 모델(MRNPM)과 학습 알고리즘

SRNPM은 기본적으로 단층신경회로망이다. 그러므

로, 다층신경회로망에 회귀 구조를 추가하면 변별력의 향상을 기대할 수 있을 것이다. 다층신경회로망의 여러 가지 회귀구조가 있을 수 있는데, 여기서는 Elman¹⁴⁾ 유형의 구조를 예측기로 학습시켰다. 이것이 다층 회귀신경예측 모델(MRNPM)이며 그 구조는 그림 2와 같다.

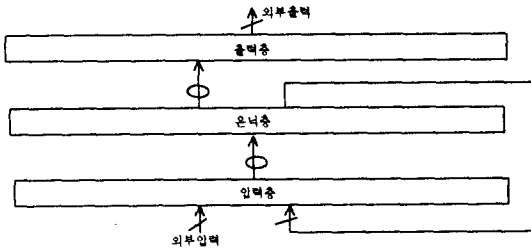


그림 2. 다층회귀신경예측 모델(MRNPM)의 구조
Fig. 2. Structure of multilayer recurrent neural prediction model (MRNPM).

입력층과 출력층 사이에 은닉층을 첨가하고, 은닉층의 모든 뉴런에 입력층으로의 회귀 연결을 두었다. 출력층에는 선형출력함수를, 은닉층에는 시그모이드 출력함수를 사용하였다. 학습신호 s^T 와 외부입력벡터인 $x(t)$ 는 SRNPM의 경우와 같이 정의하고, 은닉층에 H 개의 은닉 뉴런이 존재하면, 시간 t 에서의 은닉층의 전체입력(net input)값들로 이루어진 벡터를 $n(t)$, 은닉층의 출력벡터를 $h(t)$, 출력층의 외부출력벡터를 $o(t)$ 라 하면, 표적벡터 $s(t)$ 에 대한 예측벡터는 $\hat{s}(t)$ 는 $o(t)$ 와 같다. $n_j(t)$, $h_j(t)$, $o_k(t)$ 를 각각 $n(t)$, $h(t)$, $o(t)$ 의 j, k 번째 성분이라 하면, 전체 시스템은 다음의 식들로 기술된다.

$$\begin{aligned} n_j(t) &= \sum_{i=1}^{H+N\tau} v_{ij} z_i(t) \\ h_j(t) &= f(n_j(t)) \\ o_k(t) &= \sum_{j=1}^M w_{jk} h_j(t) \end{aligned} \quad \begin{cases} j=1, \dots, H+N\tau \\ j=1, \dots, H \\ k=1, \dots, N \end{cases} \quad (9)$$

이때, v_{ij} , w_{jk} 는 각각 i 번째 입력층 뉴런에서 j 번째 은닉층 뉴런으로, j 번째 은닉층 뉴런에서 k 번째 출력층 뉴런으로의 연결의 가중치이다. f 는 2.1절의 식 (3)에서 정의된 시그모이드 함수이고, $z_i(t)$ 는 외부입력벡터 $x(t)$ 의 앞에 $t-1$ 에서의 은닉출력벡터 $h(t-1)$ 를 연결한 $H+N\tau$ 차원 벡터 $z(t) = (z_1(t), \dots, z_{H+N\tau}(t)) = (h_1(t-1), \dots, h_H(t-1), x_1(t), \dots, x_M(t))$ 의 i 번째 성분이다.

예측오차벡터 $e(t) = s(t) - \hat{s}(t)$ 로부터 정의되는 시간 t 에서의 전체오차 $J(t)$ 는 다음과 같다.

$$J(t) = \frac{1}{2} \sum_{k=1}^N [e_k(t)]^2 \quad (10)$$

이로부터 v_{ij} , w_{jk} 에 대한 $\nabla J(t)$ 를 계산하여 각 가중치 집합에 대한 학습식을 유도할 수 있다. 학습계수를 α 라 하면 v_{ij} 에 대한 학습식은 다음과 같다.

$$\Delta v_{ij}(t) = -\alpha \frac{\partial J(t)}{\partial v_{ij}} \quad (11)$$

식 (11)를 계산하기 위한 회귀식과 초기조건은

$$\frac{\partial h_j(t+1)}{\partial v_{ij}} = f'(n_j(t)) \left[\sum_{k=1}^H v_{jk} \frac{\partial h_k(t)}{\partial v_{ij}} + \delta_{ij} z_j(t) \right] \quad (12)$$

$$\frac{\partial h_j(0)}{\partial v_{ij}} = 0 \quad \begin{cases} i=1, \dots, H+N\tau \\ j, l=1, \dots, H \end{cases} \quad (13)$$

이다. 식 (12)을 계산하기 위해서는 $t=0$ 에서 $t-1$ 까지의 인자를 모두 고려해야 하므로 많은 계산량을 요구한다. 우리는 식 (12)을 계산함에 있어, 일정 시간 이전의 뉴런의 과거 기억을 제거하여 성능의 뚜렷한 저하없이 계산량을 줄이는 S. J. Lee 등의 근사적인 방법을 적용하여 학습시간을 단축하였다¹²⁾.

w_{jk} 에 대한 학습식은 $-\alpha (\partial J(t) / \partial w_{jk})$ 를 계산하여 유도할 수 있으며 아래와 같다.

$$\Delta w_{jk}(t) = \alpha e_k(t) o_k(t) h_j(t) \quad (14)$$

III. 음성 인식 실험 및 논의

제안된 모델에 대하여 10개의 한국어 숫자음에 대한 인식 실험을 행하였다. 한국어 숫자음에는 '공', '구', '오'와 '일', '칠' 등 쉽게 변별하기 힘든 단음절들이 포함되어 있다. 25인의 남성 화자가 2회 발성한 총 500개의 음성 데이터가 사용되었다. 데이터 A, B는 각각 15인의 1, 2회째 발성 데이터이고, 데이터 C는 나머지 20인에 의해 2회 발성된 데이터이다. 학습 데이터는 데이터 A이고, 데이터 B, C는 각각 화자중속, 화자독립 인식실험에 사용되었다. 음성데이터는 10 kHz로 샘플링되고 Rabiner와 Sambur의 끝점검출 알고리즘에 의하여 끝점검출된 후, Preemphasis필터 $H(z) = 1 - 0.95z^{-1}$ 를 통과하였다. 12.8msec씩 중첩되는 25.6 msec Hamming 창을 거친 후, 10차의 LPC케스트럼

계수를 얻어서 음성 특징 벡터로 사용하였다.

다양한 모델 파라메타들에 대한 SRNPM의 최적의 동작 조건을 찾기 위해, 예측 차수 $P=2$, 은닉층뉴런의 수 $M=3$ 과 $P=3$, $M=5$ 의 두 가지 조건에 대해 각각 학습계수 α 를 변화시키면서 인식실험을 행하였다. 학습은 모두 1000회 반복하였다. 결과는 다음의 표 1, 2와 같다.

표 1. SRNPM의 학습계수에 따른 인식율 (%). ($P=2$, $M=3$ 의 경우)

Table 1. Recognition rates (%) of SRNPM as a function of learning rate. (in case of $P=2$, $M=3$)

α	학습 data	화자 종속	화자 독립	전체 data
0.00005	99.3	92.7	86.0	92.0
0.00010	99.3	94.7	86.5	92.8
0.00025	99.3	93.3	87.5	92.8
0.00040	99.3	93.3	86.0	92.2

표 2. SRNPM의 학습계수에 따른 인식율 (%). ($P=3$, $M=5$ 의 경우)

Table 2. Recognition rates (%) of SRNPM as a function of learning rate. (in case of $P=3$, $M=5$)

α	학습 data	화자 종속	화자 독립	전체 data
0.00025	99.3	95.3	88.0	93.6
0.00040	100.0	95.3	87.5	93.6
0.00050	100.0	94.0	89.0	93.8

표 1, 2에서 알 수 있듯이 SRNPM의 인식 성능은 $P=3$, $M=5$ 인 경우 전체적으로 향상되었다. 그러나 많은 회귀구조로 인해 학습에는 더 많은 시간이 소요된다. 학습계수 α 는 특정한 값 근처에서 인식율이 높고, 여기서 멀어질수록 성능은 감소함이 관찰되었다.

표 3. MRNPM의 학습계수에 따른 인식율 (%). ($P=2$, $H=11$ 의 경우)

Table 3. Recognition rates (%) of MRNPM as a function of learning rate. (in case of $P=2$, $H=11$)

α	학습 data	화자 종속	화자 독립	전체 data
0.00025	100.0	92.7	89.5	93.6
0.00065	100.0	94.7	86.5	94.2
0.00090	100.0	96.0	90.5	95.5

MRNPM의 경우, 예측차수 $P=2$, 은닉뉴런 수 $H=11$ 로 고정하고 학습계수를 변화시키면서 인식 실험을 행하였다. 학습은 모두 500회 반복하였고, 결과는 표 3과 같다.

표 3을 보면, MRNPM이 SRNPM보다 좀더 나은 성능을 보임을 알 수 있다.

제안된 회귀신경예측 모델과 기존의 예측신경회로망을 인식율, 계산량, 기억장소의 크기의 세 가지 면에서 비교한 결과를 각각 표 4, 5, 6에 제시하였다. SRNPM은 $\alpha=0.00040$, $M=5$, $P=3$ 의 경우이고, MRNPM은 $\alpha=0.00090$, $P=2$, $H=11$ 의 경우이다. 비교를 위해, 8 개의 상태와 0.001의 학습계수를 갖는 NPM과 HCNN을 사용하였다.

표 4. 인식실험 결과의 비교

Table 4. Comparison of Results of Recognition Experiments.

	SRNPM	MRNPM	NPM	HCNN
학습 data	100.0%	100.0%	100.0%	100.0%
화자 종속	95.3%	96.0%	96.0%	96.0%
화자 독립	87.5%	90.5%	90.5%	90.0%

표 5. 인식시의 계산량의 비교

Table 5. Comparison of Computation Load at Recognition Stage.

* 계산량은 1 계층, 1 입력을 기준으로 한 곱셈 횟수

	SRNPM	MRNPM	NPM	HCNN
계산량	299	451	400	6080

표 6. 기억장소의 갯수의 비교

Table 6. Comparison of the Size of Required Memory.

* 가중치 1개당, 뉴런 1개당 각각 1개의 기억장소가 필요.

	SRNPM	MRNPM	NPM	HCNN
기억장소 수	332	404	2720	818

표 4, 5, 6 으로부터 제안된 모델은 기존의 예측신경회로망에 비해 계산량, 기억장소의 크기 면에서 대폭 향상된 성능을 지니면서 인식율에 있어서는 그에 필적하였다.

IV. 결 론

본 논문에서는 회귀신경회로망을 음성의 비선형 예측기로 학습시키는 회귀신경예측모델을 제안하였다. 학습 알고리즘은 누적예측오차에 경사강하법을 적용하여

유도하였으며 매 시간마다 실시간으로 학습이 이루어 지도록 하였다. 제안된 2가지의 모델인 SRNPM과 MRNPM은 기존의 예측신경회로망과 변별기로 학습되어지는 회귀신경회로망의 단점들을 효과적으로 극복하며 각각의 장점은 잘 보완한다. 즉, 제안된 모델은 예측신경회로망에 비해 시간축정렬 알고리즘이 불필요하여 인식시간이 단축되고, 작은 크기의 기억장소를 요구한다. 또, 시변성 표적함수 설정의 문제가 자연스럽게 해결된다. 그러므로, 연속음 인식 및 대용량 어휘 인식기의 구축에 기존의 모델들에 비해 많은 장점을 보유한다. 또한 다양한 조건에서 한국어 숫자음 인식실험을 행한 결과 인식 성능에 있어서도 기존의 예측신경회로망 모델들에 필적할 만 하였다.

제안된 모델은 예측기로 학습되어지므로 각각의 계층이 독립적으로 모델링되어 학습됨에 따라 유사한 계층간의 변별력이 저하된다는 단점이 있다. 오인식을 분석한 결과 '공', '오', '구' 와 '일', '칠' 등 음운학적으로 유사한 어휘간에 오인식이 많이 발생하였다. 최근에 연구되고 있는 GPD (Generalized Probabilistic Descent) 방법과 같은 여러가지 변별력 향상 알고리즘을 이용하면 인식율의 개선을 기대할 수 있을 것이다¹⁶⁾. 또한, 순방향 예측(forward prediction) 뿐 아니라, 역방향 예측(backward prediction)을 도입하여 예측성능의 향상을 도모하는 것도 흥미로울 것이다.

참 고 문 헌

- [1] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano and K. Lang, "Phoneme recognition using time-delay neural networks," *Proc. ICASSP'88*, pp. 107-110, 1988.
- [2] H. Sakoe, R. Isotani, K. Yoshida, K. Iso and T. Watanabe, "Speaker-independent word recognition using dynamic programming neural networks," *Proc. ICASSP'89*, pp. 29-32, 1989.
- [3] N. Tishby, "A dynamical systems approach to speech recognition," *Proc. ICASSP'90*, pp. 365-368, 1990.
- [4] E. Levin, "Word recognition using hidden control neural architecture," *Proc. ICASSP'90*, pp. 433-436, 1990.
- [5] E. Levin, "Hidden control neural architecture modeling of nonlinear time varying systems and its applications," *IEEE Trans. on Neural Networks*, vol. 4, no. 1, pp. 109-116, January 1993.
- [6] K. Iso and T. Watanabe, "Speaker-independent word recognition using a neural prediction model," *Proc. ICASSP'90*, pp. 441-444, 1990.
- [7] K. Iso and T. Watanabe, "Large vocabulary speech recognition using neural prediction model," *Proc. ICASSP'91*, pp. 57-60, 1991.
- [8] J. Tebelskis and A. Waibel, "Large vocabulary recognition using linked predictive neural networks," *Proc. ICASSP'90*, pp. 437-440, 1990.
- [9] J. Tebelskis, A. Waibel, B. Petek and O. Schmidbauer, "Continuous speech recognition using linked predictive neural networks," *Proc. ICASSP'91*, pp. 61-64, 1991.
- [10] H. Bourlard and C. J. Wellekens, "Speech dynamics and recurrent neural networks," *Proc. ICASSP'89*, pp. 33-36, 1989.
- [11] R. J. Williams and D. Zipser, "A learning algorithm for continually running fully recurrent neural networks," *Neural Computation*, vol. 1, pp. 270-280, 1989.
- [12] S. J. Lee, K. C. Kim, H. S. Yoon and J. W. Cho, "Application of fully recurrent neural networks for speech recognition," *Proc. ICASSP'91*, pp. 77-80, 1991.
- [13] R. L. Watrous and L. Shastri, "Learning phonetic features using connectionist networks: an experiment in speech recognition," *Proc. IJCNN'87*, vol. 4, pp. 381-388, 1987.

- [14] J. L. Elman, "Finding structure in time", *CRL-TR-8801*. University of California, San Diego, 1986.
- [15] K. M. Na, J. K. Ryu, D. I. Chang, S. I. Chae and S. G. Ann, "Recurrent neural prediction models for speech recognition," *Proc. EUROSPEECH'95*, Madrid, 1995.
- [16] K. M. Na, J. Y. Rheem and S. G. Ann, "A discriminative training algorithm for predictive neural networks," *Proc. IS-CAS-94*, pp. 431-434, 1994.

— 저 자 소 개 —



柳 濟 寬(準會員)
 1970년 1월 5일생. 1992년 2월 서울대학교 미생물학과 졸업(이학사). 1994년 2월 서울대학교 전자공학과 졸업(공학사). 1994년 3월 ~ 현재 서울대학교 전자공학과 석사과정에 재학

중. 주관심분야는 신경회로망, 패턴인식, 신호처리 등임.

羅 景 民(準會員) 第 31卷 B編 第12號 參照

현재 서울대학교 전자공학과 박사과정

林 材 烈(正會員) 第 31卷 B編 第12號 參照

현재 한국기술교육대학교 전자공학과 조교수

成 宏 模(正會員) 第 29卷 B編 第11號 參照

현재 서울대학교 전자공학과 교수

安 秀 桔(正會員) 第 31卷 B編 第12號 參照

현재 서울대학교 전자공학과 교수