한국퍼지 및 지능 시스템학회 논문지 1998, Vol. 8, No. 6. (1998) 129~135

회귀예측 신경모델과 카오스 신경회로망을 결합한 고립 숫자음 인식

Isolated Digit Recognition Combined with Recurrent Neural Prediction Models and Chaotic Neural Networks

김석현 · 여지환

Seok Hyun Kim and Ji Hwan Ryeo

대구대학교 정보통신공학과

요 약

본 논문은 서로 다른 접근방식을 사용하는 카오스 회귀 신경예측모델과 다층 신경회로망이 결합하여 고립음의 인식률을 높이고자 하였다. 전반적으로 다층신경회로망인 MLP와 결합한 인식률은 1.2 % 에서 2.5% 이상이 개선 되었다. 이는 서로 인식하는 방법이 다르기 때문에 서로 상호 보완되고, 카오스의 다이내믹 성질이 인식률을 개선시켰음을 실험으로 밝혔다. MLP와 결합한 인식률은 카오스 다층신경망일 때가 가장 좋았다. 그러나 학습시 알고리즘이 단순하고, 신뢰도 면에서는 오히려 카오스 단층 신경망이 인식률은 0.5%정도 떨어지지만 더욱 좋다고 생각된다. 주로 MLP는 숫자음 "일"과 "오"에서 우수한 성적을 나타내었고, 카오스 예측 신경망은 숫자음 "영", "삼", "칠"에서 우수하였다.

ABSTRACT

In this paper, the recognition rate of isolated digits has been improved using the multiple neural networks combined with chaotic recurrent neural networks and MLP. Generally, the recognition rate has been increased from 1.2% to 2.5%. The experiments tell that the recognition rate is increased because MLP and CRNN(chaotic recurrent neural network) compensate for each other. Besides this, the chaotic dynamic properties have helped more in speech recognition. The best recognition rate is when the algorithm combined with MLP and chaotic multiple recurrent neural network has been used. However, in the respect of simple algorithm and reliability, the multiple neural networks combined with MLP and chaotic single recurrent neural networks have better properties. Largely, MLP has very good recognition rate in korean digits "il", "oh", while the chaotic recurrent neural network has best recognition in "young", "sam", "chil".

1. 서 론

본 논문에서는 서로 다른 접근 방식을 사용하는 카오스 회귀 신경 예측모델과 다층 신경회로망이 결합하여 고립 숫자음의 인식률을 높이고자 하였다. 예측 신경모델에는 HCNN(Hidden Control Neural Network), NPM(Neural Prediction Model), LPNN (Linked Predictive Neural Network) 등이 있으나, 이들은 동적 프로그래밍에 의한 최적경로를 찾는 과정이 복잡한 시간축 정렬 알고리즘을 필요로한다. 회귀 신경회로망은 회귀연결구조가 시간적인 변이를 흡수할 수 있고, 망의 내부 표현에 의해 사상능력을 변조할 수 있는 구조적인 특징을 지니고 있다. 또한 학습 및 인식에 사용되는 입력신호의 길이에 제한을 두지 않는다. 한편 카오스 신경회로망은 화자에 따른 목표값이 변화하지 않고, 고정되어 있는 경우에는 카오스의 다이내믹스의 효과로 인해서 입력 파형이 원하는 출력파형과 많이 달라져 있더라도 회귀예측신경모델보다도 더 잘 출력 파형에 따라가는 것으로 인식되고 있다[1].

따라서 본 논문에서는 회귀예측 신경모델에 카오스 다이내믹스 효과를 얻기 위하여 카오스 회귀예측 모델을 도입하였다. 그러나 카오스 예측모델은 각각의 계층이 독립적으로 모델링되어 학습되어짐에 따라 유사한 계층간에 변별력이 저하된다는 단점이있다. 이의 단점을 해결하기 위해서 다층 신경회로망을 사용하였다. 다층 신경회로망은 주로 패턴 분류에 많이 사용되는 시간의 개념이 배제된 신경회로망이다. 이 방식은 음성신호에서 시간 개념을 제거한 시간 주파수 패턴을 정적인 패턴으로보고 이를 분류한다. 각

각의 계층이 한 개의 모델로 학습될 수 있으므로 유사한 계층간에 변별력의 저하를 막을 수 있다. 이 두 신경회로망의 출력을 다음단에서 종합하여 최종결과를 낸다. 종합 판단에는 다층 신경회로망을 사용한다. 각각의 방식이 쉽게 인식하지 못하는 단어가 다르므로 두 방식의 약점을 극복할 수 있다. 실험결과 2.5% 정도의 인식률이 향상되었다.

본 논문에서는 음성인식을 위한 새로운 신경회로망 구조인 카오스 회귀예측신경모델과 결합한 다층 신경회로망을 제안한다. 이를 위해 3개의 신경회로망을 각각 학습시킨다. 그 이유는 다중 신경회로망의 근본 취지가 서로 다른 접근 방식을 사용할 때 각각의 완전한 음성인식기들을 개별적으로 사용하여서는 더 이상의 성능향상이 어려울 수 있기 때문이다. 그러므로 두 개 이상의 음성인식기를 함께 사용한다. 각각의 음성인식기가 잘 인식하는 상황을 판단자가 판단하여 그 상황에서 적절한 음성 인식기의 출력을 선택하는 방식을 통해 인식률을 높이는 것이다. 그러므로 기본이 되는 음성인식기는 각각 미리학습시키고, 각 인식기가 학습자료에 대해 테스트한 결과를 가지고 판단자 신경회로망을 학습시킨다.

2. 카오스 회귀 신경예측 모델과 결합한 다층 신경회로망의 제안

2.1 카오스 모델과 결합한 다층 신경회로망의 구조

그림 1은 카오스 회귀 신경예측모델과 결합한 다층 신경회로망의 구조이다. 다층 신경회로망인 MLP는 음성신호를 시간 주파수 패턴으로 보고 인식하는 방식이고 CRNN(chaotic recurrent neural network)은 음성을 비선형 동적 시스템의 출력이라 보고 인식하는 방식이다. 그림 1에서 보면 두가지 특성을 결합하기 위해 이 두 시스템의 출력을 또 하나의 MLP의 입력으로 삼는다. 오인식을 줄이기 위해서 각각의 입력에 대해 어느 음성인식기를 사용할지를 정하는 판단자가 필요하다. 본 논문에서는 판단자로 또 하나의 MLP를 사용하였다. 첫 번째 MLP와 CRNN의 출력값이 두 번째 판단자인 MLP의 입력으로 들어간다. M개의 단어를 인식하는 경우 두 번째 MLP의 입력 뉴런의 수는 2M개, 출력 뉴런의 수는 M개가 된다. 또한 CRNN의 출력 값의 범위가 MLP의 출력값의 범위와 비슷하여야 두 번째 MLP가 잘 동작한다. CRNN의 출력을 MLP의 출력 범위인 (-1, 1)로 변환하는 방법은 여러 가지가 있을 수 있으나 본 논문에서는 실험 결과 선형적인 변환이 결과가 좋은 것으로 나타났다. 선형적으로 최대값을 -1로, 최소값을 1을 선택하였다. 다음과 같은 선형변환식을 사용하였다.

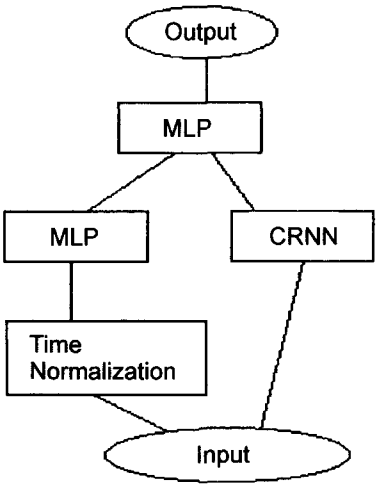


그림 1. 카오스 회귀신경예측모델과 결합한 다 층 신경망의 구조.

Fig. 1. Multiple network architecture combined with Chaotic Recurrent Neural Network.

새 총예측오차
= 2*(총예측오차-최소값)/(최소값-최대값)+1 (1)

2.2 카오스 순환 신경회로망

생체 시스템에는 신경세포 레벨에서 뇌의 신경회로망에 이르기까지 여러 가지 레벨에서 비선형 다이내믹스 현상이 관측되고 있다. 가령 지금까지는 주위의 환경이 일정하면 인간의 체온과 심장의 고동은 일정하게 유지되고 있다고 생각되고 있었다. 그러나 안정상태에 있는 건강한 사람의 고동리듬은 항상 요동하고 있고 반대로 죽기 직전에 있는 사람의 고동의 리듬은 일정하다고 알려져 있다[2]. 자연환경은 항상 변동하고 있고, 그 변동의 방식은 복잡하고 예측이 어렵다. 즉, 외계에 대해서는 일정한 고동을 유지하기보다는 예측할수 없는 변화에 대비해서 항상 움직임을 포함한 리듬을 발생하는 편이, 외계에 유연하게 적응할 수가 있다고 생각된다. 실제의 뇌 신경계의 움직임은 카오스를 전형적인 예로하는 매우 동적인 것이다. 즉, 뇌신경계는 뉴런이라는 카오스 다이내믹스를 가지는 카오스 소자로 구성된 대규모이며 복잡한 시스템으로 취급될 수 있다. 뇌신경계에 있어서 카오스의 존재는, 복잡한 아날로그, 디지탈 시스템인 뇌신경계의 아날로그적 측면을 상징하고 있다. 이러한 뇌의 비선형성과 아날로그적 측면을 강조한 뉴런모델로 최근에 카오스 뉴런 모델이 Aihara[2]에 의해 제안되

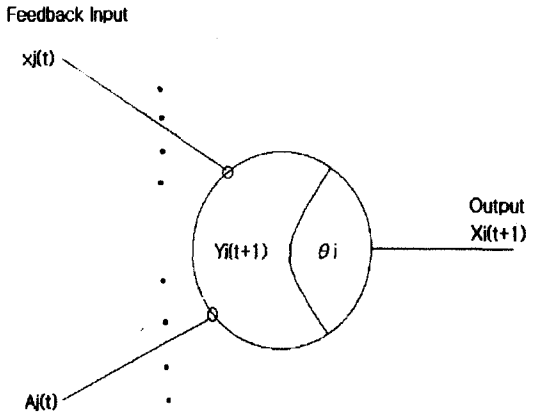


그림 2. 카오스 뉴런모델의 입력 및 출력.

Fig. 2. Input and output of the chaotic neuron model.

었으며, 이는 Sato[3]의 뉴런 모델을 기본으로 하여 그 응답 특성을 아날로그적으로 수정한 것이다. 그림 2는 카오스 뉴런모델의 입력 및 출력이다. 그림에서 보면 신경회로망의 외부로 부터 입력과 그 신경회로망 내의 다른 뉴런으로부터의 피드백 입력 양쪽을 고려할 필요가 있다. 최근에 카오스 뉴런신경망을 이용한 신경망이 많이 발표되고 있다.

그림 3은 카오스 뉴런을 이용한 CRNN이다. CRNN을 구성하는 뉴런의 동적 방정식은 다음과 같이 표현될 수 있다.

$$ry_i(t+1) = \sum_{j \in I \cup E} w_{ij}\, z_j(t) + k y_i(t) - \alpha x_i(t)$$
$$x_i(t) = f_i(y_i(t)) \qquad (2)$$

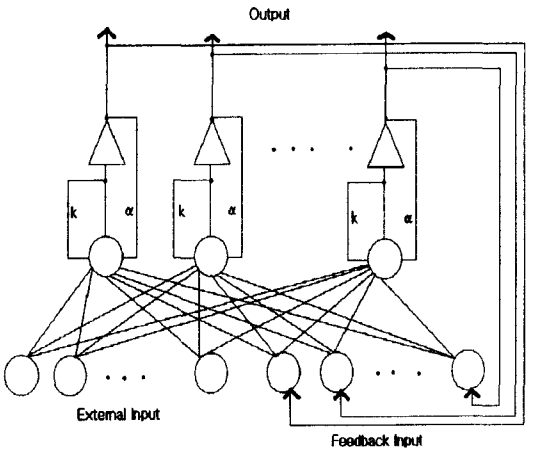


그림 3. 카오스 뉴런을 이용한 CRNN.

Fig. 3. CRNN using chaotic neuron.

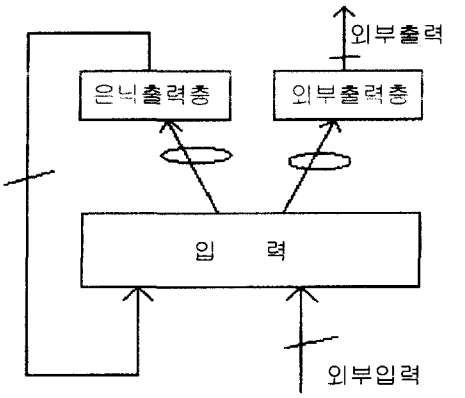


그림 4. 단층회귀 신경예측 모델의 구조.

Fig. 4. Structure of SRNPM(simple neural prediction model)

$$z_i(t) = \begin{cases} x_i(t) & \text{if } i \in I \\ A_i(t) & \text{if } i \in E \end{cases}$$

$y_i(t+1)$: 시각 t에서의 i번째 뉴런의 내부 상태

$x_i(t)$: 시각 t 에서의 i번째 뉴런의 출력

$f_i(\,)$: i번째 뉴런의 출력함수

$A_i(t)$: 시각 t에서의 외부 입력

w_{ij} : j번째 노드에서 i번째 노드의 연결 가중치

k, α : 불응성 파라미터

실제 사용된 CRNN모델은 "유제관외 4명이 제안한 회귀신경예측 모델"[4]에 카오스 뉴런을 접목시켜 사용하였다. 그림 4는 단층회귀신경예측 모델(SRNPM)이다. SRNPM은 두층으로 구성된다. 입력층에는 외부입력뉴런과 회귀 입력뉴런이 있고, 출력층에는 은닉출력뉴런과 외부 출력뉴런이 있다. 회귀연결에는 은닉출력뉴런과 회귀 입력뉴런 사이에만 존재한다. 은닉 출력뉴런에는 시그모이드(sigmoid)출력 함수가 있고, 외부 출력뉴런에는 선형출력 함수가 있다. 이 SRNPM의 왼쪽 회귀뉴런을 카오스 뉴런으로 바꾸면 본 논문에서 제안한 CRNN이 된다. 학습 알고리듬의 이론적인 배경은 SRNPM의 이론에 카오스 부분만 바꾸면 된다.

$S^T = \{s(1), s(2) \ldots\ldots, s(T)\}$를 길이 T의 학습 신호라고 할 때 $s(t) = (s_1(t), \ldots\ldots s_N(t))$는 시간 t 에서 N차원 음성 특징 벡타가 된다. 시간 t에서 표적 벡터를 $s(t)$라 할 때 예측차수가 τ인 경우, 외부 입력 벡터인 $x(t) = (x_1(t), \ldots, x_{N_\tau})$은 $s(t-1), \ldots\ldots s(t-\tau)$의 τ개의 벡터

를 연결한 벡터이다. M 개의 회귀 연결(또는 M 개의 은닉출력 뉴런)이 존재할 때,

$$y(t)=(y_1(t),....y_{N+M}(t))$$
$$o(t)=(o_1(t),....o_{N+M}(t)) \quad (3)$$

를 각각 비선형 출력함수의 전체 입력 벡터와 출력 벡터라 하자. $y(t)$와 $o(t)$의 j번째 성분인 $y_j(t)$와 $o_j(t)$는 다음과 같다.

$$y_j(t+1)=\sum_{i=1}^{M+N\tau} w_{ij}\, z_j(t)+ky_i(t)-\alpha x_i(t)$$
$$\begin{cases} i=1..M+N\tau \\ j=1..M+N \end{cases}$$
$$x_i(t)=f_i(y_i(t))$$
$$x_i(t+1)=f_i(y_i(t+1))$$
$$o_j(t)=f(y_j(t)) \quad (4)$$

이때, $f(y_j(t))=\begin{cases} \dfrac{1}{1+exp(-y_j(t))} & j=1..M \\ y_j(t) & j=M+1..M+N \end{cases}$

w_{ij}는 입력층의 i번째 뉴런에서 j번째 뉴런으로의 연결 가중치이고 $z_i(t)$는 $o(t-1)$과 $x(t)$를 연결한 벡터인

$$z(t)=(z_1(t),....,z_{M+N_\tau}(t))$$
$$=(o_1(t-1)..o_M(t-1),x_1(t),.x_{N_\tau}(t))$$

의 I번째성분이다. $s(t)$를 예측한 벡터인

$\hat{s}(t)=(\hat{s}_1(t)..\hat{s}_N(t))$는 $(o_{M+1}(t)..o_{M+N}(t))$와 같다. 예측 오차 벡터 $e(t)=s(t)-\hat{s}(t)$로 부터 정의 되는 시간 t에서의 전체오차인 J(t)와 발성시간 T동안의 누적 예측 오차는 다음과 같다.

$$J(t)=\frac{1}{2}\sum_{k=1}^{N}[e_k(t)]^2 \quad E=\sum_{t=1}^{T}J(t) \quad (5)$$

학습 계수를 α라 하면 $\nabla J(t)$를 계산하면 시간 t에서의 가중치 학습량은,

$\nabla w_{ij}(t)=-\alpha\dfrac{\partial J(t)}{\partial w_{ij}}$ 이다.

실제 $\nabla w_{ij}(t)$는 다음 식에서 얻을 수 있다.

$$\frac{\partial o_k(t+1)}{\partial w_{ij}}=f_i\,'(y_k(t))[\delta_{ik}z_j(t)+\sum_{l=1}^{N}w_{k(l+M)}\frac{\partial o_k(t)}{\partial w_{ij}}] \quad (6)$$

초기 상태는 가중치에 대해 함수적으로 독립이므로 다음의 초기 조건을 얻는다.

$$\frac{\partial o_k(0)}{\partial w_{ij}}=0 \quad (7)$$

2.3 Adaptive 시간 정규화

MLP를 가지고 한 개의 모델로 만들기 위해서는 각 음성신호의 크기를 일정하게 하는 시간 정규화가 필요하다. 음성인식에는 포먼트 위치보다 그 기울기등 변화하는 성분이 중요하다고 한다. 음성에서 static한 정보를 모두 빼고 사람에게 인식 시험을 한 결과 static한 정보를 담고 있는 음성의 경우와 비교해 인식 수준이 거의 같았다고 한다. 화자독립 음성인식에서는 특히 포만트(formant)의 위치가 일정하지 않기 때문에 그 변화하는 모양이 인식에 매우 중요한 정보로 작용한다. 같은 단어가 서로 다른 길이로 발음되더라도 그 지속 시간만 다를 뿐이지 특징 벡타공간에서는 유사한 궤적을 보일 것이다. 이러한 점에 착안하여 계산시간이 많이 소모되는 DTW(Dynamoc Time Warping)보다 간단하며 주파수 영역에서 변화를 강조하는 Adaptive 시간 정규화를 사용한다. 이렇게 해서 형성된 패턴은 Uniform시간 정규화 처럼 자음이 모음에 의해 거의 사라지는 단점은 없으나 주파수 영역 변화가 큰부분은(변이음 등)은 정보가 강조되어 좋은 특성을 보인 반면 변화가 작은 부분(정적인 부분, 모음 등)의 정보가 지나치게 감소될 가능성이 있다. 보통 고립 숫자음 인식에서는 처음 시작에 7 프레임 마지막에 7 프레임 해서 14 프레임 정도로 선택하거나 아니면 앞 뒤 5프레임과, 모음의 Peak치을 중심으로 5프레임 해서 15프레임 정도 선택하고 있으나, Adaptive 시간 정규화 방식에 의해서 28 프레임, 14차 cepstrum을 선택하였다.

Adaptive 정규화 방식은 다음과 같다. 입력음성을 전처리한 시간축상의 길이가 N인 cepstrum+power를 $(x_0, x_1, x_2,....,x_{N-1})$라고 하고 누적 스펙트럼 거리를 다음과 같이 정의 하자.

$$C_i=\sum_{j=1}^{i}d(x_{j-1},x_j) \quad i=1,2,.....N-1$$
$$C_0=0 \quad (8)$$

이때 스펙트럼 변화를 다음과 같이 정의하면

$\nabla C=\dfrac{C_{N-1}}{M-1}$ 여기서 M은 정규화 하고자 하는 길이를 의미한다. 인접한 두 입력 벡터가 선형적으로 변화한다고 가정하면 시간 정규화 된 특징벡터는 아래와 같

이 인접한 두 벡터의 interpolation으로 표현된다. 즉,

$$Y_k = \frac{|k\nabla C - C_{j-1}|x_j + |k\nabla C - C_j|x_{j+1}}{C_{j+1} - C_j} \quad k=0, 1, 2 .. M-1$$
$$j = arg\ min(|k\nabla C - C_j| \text{ if } k\nabla C \geq C_j$$
$$j = arg\ min(|k\nabla C - C_j|) - 1 \text{ if } k\nabla C \leq C_j$$
$$j = 0, 1, 2 .. N-2 \quad (9)$$

3. 음성 인식 실험 및 논의

제안된 모델에 대하여 10개의 한국어 숫자음에 대한 인식 실험을 행하였다. 실험 데이터는 한국전자통신연구소에서 제공한 샘돌이 음성데이타중에서 "영, 일,이,삼,...구" 10개음성을 사용하였다. 모두 40명(남: 20명, 여:20명)이 10개 단어를 4회 발음한, 샘플링 주파수가 16 kHz 인 1600개의 단어를 사용하였다. 앞에서 발음한 (남:10명, 여:10명)의 800개의 음성은 학습 데이터로 사용하였고, 뒤에 발음한 (남:10명, 여:10명)의 800개의 음성은 인식 시험 데이터로 사용하였다. 학습데이타는 다시 2개의 집합으로 나누어서 각 사람이 4번 발음한 데이타 중에서 앞에 2번 발음한 데이터는 A 그룹, 뒤에 두 번 발음한 데이터는 B 그룹으로 두고, 기본 인식기는 A그룹으로 훈련시키고, 최종 판단자인 MLP는 다시 B 그룹으로 학습시켰다. 기본 인식기인 MLP에는 시간 정규화를 시킨 14차 cepstrum, 28 프레임으로 하였다. 보통 숫자음 인식기에는 앞 7 프레임, 뒤 7프레임 해서 14 프레임을 사용하거나, 아니면 앞 5 프레임, 뒤 5프레임, 그리고 모음의 피크치에서 5프레임 하여 15프레임을 사용하는 경우가 많지만 본 논문는 다른 고립음에도 적용가능하도록 시간 정규화 하여 28프레임 정도로 하였다. 입력 데이터수는 총 421(바이어스 항 1개 포함)

은닉층의 수는 40, 50, 60, 70, 80으로 바꾸어서 해보았지만 70 이상에서 100% 학습률을 보였기 때문에 70으로 하였다. 이 때 사용된 계수들은 다음과 같다.

총 훈련 반복횟수 200회
학습률 0.7
관성률 0.07
에너지 에러 오차 한계 0.0001

또한 기본 인식기로서 CRNN은 불응항 계수인 k, α에 의해서 상당히 달라지고 있다. 표 1은 한국어 숫자음 (남 10, 여 10)의 첫 번째 발음한 200개의 데이터를 가지고 예측에러를 조사한 표이다. 학습계수는

표 1. 한국어 숫자음 k,α에 따른 학습예측오차
Table 1. Learning error for Korean digits according to k and α

α \ k	0.0	0.1	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.0	11.4	14.7	190.1	255.4	495.9	1021	1021	1021	1021
0.1	9.6	13.0	15.1	14.8	36.3	1021	1021	1021	1021
0.2	9.7	11.6	9.6	9.4	10.7	1021	1021	1021	1021
0.3	10.2	11.0	7.9	8.8	9.9	719	1021	1021	1021
0.4	11.0	12.7	8.1	8.5	9.8	646	727	1020	1021
0.5	11.7	13.7	9.3	9.0	9.0	565	539	719	1021
0.6	12.3	13.8	9.4	9.5	8.8	8.5	8.7	712	760
0.7	13.0	14.2	9.6	9.9	8.7	8.3	9.2	592	760
0.8	13.5	14.8	10.0	10.4	8.6	7.8	9.5	196	615
0.9	14.0	15.3	10.4	12.0	11.4	8.6	9.8	8.7	135

표 2. 독립적인 인식기의 인식 갯수 및 인식률
Table 2. The recognition amount and rate of independent recognizer

인식기 / 숫자음	MLP	CRNN1 SRNPM1	CRNN2 SRNPM2	CRNN3 MRNPM
"영"	67(80)	80(80)	80(80)	80(80)
		79(80)	79(80)	79(80)
"일"	78(80)	68(80)	77(80)	78(80)
		69(80)	75(80)	76(80)
"이"	65(80)	75(80)	70(80)	70(80)
		73(80)	70(80)	70(80)
"삼"	63(80)	79(80)	78(80)	78(80)
		78(80)	77(80)	77(80)
"사"	72(80)	74(80)	73(80)	73(80)
		72(80)	72(80)	72(80)
"오"	78(80)	71(80)	73(80)	73(80)
		66(80)	69(80)	69(80)
"륙"	71(80)	75(80)	75(80)	75(80)
		74(80)	74(80)	75(80)
"칠"	73(80)	79(80)	79(80)	79(80)
		77(80)	78(80)	78(80)
"팔"	64(80)	72(80)	74(80)	74(80)
		65(80)	67(80)	68(80)
"구"	63(80)	63(80)	70(80)	70(80)
		62(80)	66(80)	66(80)
인식률	86.75%	92.00%	93.63%	93.75%
		89.37%	90.88%	91.25%

0.00025, 총 반복횟수 100회에서 수행하였다.
표 1에서 보면 k=0.6, α=0.8에서 예측에러가 최소임을 알 수 있다. 본 논문의 모든 실험은 이 값을 적용하였다.

CRNN의 모델은 "유재관외 4명[4]"이 발표한 모델에 카오스를 도입하여 회귀 뉴런에 카오스 뉴런으로 대치하여 변형시켰다. 학습 계수는 이들의 값을 도입하였다. 비교는 3가지를 해 보았다. 첫째 MLP 와 CRNN1, 두 번째 MLP와 CRNN2, 세 번째 MLP 와 CRNN3이다. CRNN1은 예측차수 2, 회귀뉴런수 3의 SRNPM1의 변형이고, CRNN2는 예측차수 3, 회귀뉴런수 5의 SRNPM2의 변형이다. 또한 CRNN3는 예측차수 2, 은닉뉴런수 11인MRNPM의 변형이다. 인식률은 화자 독립만 해보았다. 전부 14차 cepstrum을 사용하였고, power항은 다른 항과의 경쟁을 위하여 0.04를 곱하였다. MLP의 정규화한 프레임 수는 28 프레임이다. SRNPM1과 SRNPM2는 학습계수는 다같이 0.00025를 사용하였고, MRNPM는 0.00090를 사용하였다. 총 반복횟수는 3방식 전부 100회로 하였다. 표 2는 개인 인식기의 인식률이고, 표 3은 복합 인식기의 다층신경망의 인식률이다.

전반적으로 인식률은 MLP와 결합했을 때 1%이상이 개선된다. 이는 서로 인식하는 방법이 다르기 때문에 서로 상호 보완해 줌으로써 개선된다고 본다. 문제는 숫자음 "구"의 인식률이 상당히 둘다 저조하여 서로 보완해 주지 못하고 있다. 이 부분의 원인을 알아보기 위해서 다시 표 4에 오인식의 범위를 알아 보았다.

표 3. MLP와 카오스 예측신경망과 결합한 인 식 갯수및 인식률

Table 3. The recognition amount and rate of the recognizer combined with MLP and CRNN

인식기 / 숫자음	MLP	MLP & CRNN1	MLP & CRNN2	MLP & CRNN3
"영"	67(80)	80(80)	80(80)	80(80)
"일"	78(80)	78(80)	78(80)	78(80)
"이"	65(80)	75(80)	70(80)	71(80)
"삼"	63(80)	79(80)	78(80)	78(80)
"사"	72(80)	74(80)	73(80)	73(80)
"오"	78(80)	78(80)	78(80)	78(80)
"륙"	71(80)	75(80)	75(80)	75(80)
"칠"	73(80)	79(80)	79(80)	79(80)
"팔"	64(80)	72(80)	74(80)	74(80)
"구"	63(80)	63(80)	70(80)	70(80)
인식률	86.75%	94.13%	94.38%	94.50%

표 4. 숫자음 "구"의 오인식 갯수 및 오인식률

Table 4. The recognition error and amount of Korean digit "ku"

인식기 / 숫자음	MLP	CRNN1	CRNN2	MLP & CRNN3
"영"	1	5	2	2
"일"	1	4	1	1
"이"	0	0	0	0
"삼"	1	0	0	0
"사"	2	0	0	0
"오"	11	8	7	7
"륙"	0	0	0	0
"칠"	0	0	0	0
"팔"	1	0	0	0
"구"	0	0	0	0
"구" 음성의 오인식률	21%	21%	12.5%	12.5%

다. 오인식의 많은 단어가 "구"를 "오"로 오인식 하였다. 분명히 "구"의 초성자음 "ㄱ"이 인식에 어려움이 있었다고 생각된다. 음성데이타를 만들 때 초성자음 "ㄱ"이 끝점 처리시 잘려나갈 수 도 있기 때문에 특히 음성데이타를 만들 때에 주의 해야 할 부분이라고 생각된다.

지금 까지 전향예측을 하였는데 만약 후향예측을 겸한다면 더욱 인식률이 높아지리리 생각된다.

4. 결 론

본 논문은 서로 다른 접근 방식을 사용하는 카오스 회귀 신경예측모델과 다층신경회로망이 결합하여 고립음의 인식률을 높이고자 하였다. 전반적으로 MLP와 결합한 인식률은 1.2% 에서 2.5%정도가 개선된다. 이는 서로 인식하는 방법이 다르기 때문에 서로 상호 보완해 줌으로써 개선되고, 카오스의 다이내믹 성질이 인식률을 개선됨이 실험으로 증명되었다. MLP와 결합한 인식률은 카오스 다층 예측신경망인 CRNN3일 때 가장 인식률이 좋지만, 이 알고리듬은 다층이기 때문에 다른 두 방식에 비해서 복잡하고 학습시에 진동을 수반하는 등 문제가 있을 수 있지만 MLP와 결합한 CRNN1 혹은 CRNN2 방식이 학습이 쉽고 알고리듬이 간단하고 신뢰도가 좋기 때문에 오히려 이 방법이 더욱 좋다고 생각된다. 세가지 방법 모두의 공통적인 문제점은 학습시간이 길기 때문에 앞으로 이 부분을 개선하여야 할 것으로 생각된다. 또

한 MLP에 카오스를 도입 할 수 있다면 더욱 인식률은 높아지리라고 생각한다.

감사의 글

본 연구가 가능하도록 음성 데이터를 제공해 준 한국전자 통신연구소에 감사 드립니다.

참고문헌

[1] N. Kanou, Y. Horis, K. Aihara and S. Nakamura, "A current-mode circuit of a chaotic neuron model", Proc. of 2nd International Conference on Fzzy logic and Neural Networks, 1992.

[2] K. Aihara, T. Takabe and M. Toyoda, "chaotic neural networks", *Phys. Lett.*, Vol. A, No.144, pp.333-340, 1990.

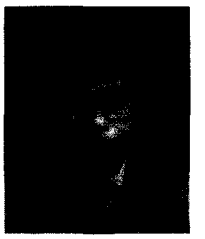
[3] M. Sato, "A learning algorithm to teach spatio-temperal patterns to recurrent neural networks", *Biol. Cybern.*, pp. 259-263, 1990.

[4] 류재관, 라경민, 임재열, 성굉모, 안수길, "회귀 신경예측모델을 이용한 음성인식", 대한전자 공학회 논문지, 제11권 B편, 제32권, 1995.

[5] K. Iso and T. Watanabe, "Speaker-independent word recognition using a neural prediction model," Proc. ICASSP' 90, pp. 441-444, 1990.

[6] 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.

[7] E. Levin, "Word recognition using hidden control neural architecture", Proc. ICASSP'90, pp. 433-436, 1990.



김 석 현(Seok Hyun Kim) 정회원

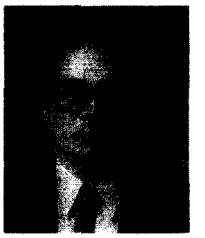
1974년 : 부산대학교 전자공학과 졸업

1981년 : 경북대학교 대학원 전자공학과(석사)

1991년 : 경북대학교 대학원 전자공학과(박사)

1984-현재 : 대구대학교 정보 통신공학부 교수

주관심분야 : 영상처리, 패턴인식, 음성인식



여 지 환(Ji Hwan Ryeo) 종신회원

제 6권 제 2호 참조

현재 : 대구대학교 정보 통신공학부 교수