

論文95-32B-5-16

국부 유사사상의 퍼지통합에 기반한 비선형사상의 식별

(Identification of Nonlinear Mapping based on Fuzzy Integration of Local Affine Mappings)

崔鎮榮*, 崔宗鎬*

(Jin Young Choi, and Chong-Ho Choi)

요 약

본 논문에서는 임의의 비선형 사상을 국부영역에서 유사 사상(affine mapping) 으로 근사화하고 이들을 퍼지 추론 (Fuzzy Inference) 형태로 통합하여 전체 비선형 사상을 근사 식별할 수 있는 범용 근사자(universal approximator)를 제시한다. 이 범용 근사자를 방사형기저함수(radial basis function) 형태의 처리노드와 가중합(weighted sum) 형태의 처리노드를 이용하여 연결형 모델(connectionist model)로 구성할 수 있는 구조를 제안한다. 또한 이 모델을 이용하여 주어진 입력력 데이터로부터 비선형 사상관계를 식별해 내기 위해 자기구성 (self-organizing) 기능을 갖는 학습방법을 제안한다. 제안된 방법의 타당성을 살펴보기 위해 함수 근사(function approximation) 문제, Mackey-Glass 혼돈 시계열 예측(chaotic time series prediction) 문제 등에 적용하여 학습속도, 일반화성능등과 관련하여 기존 방법과 비교한다.

Abstract

This paper proposes an approach of identifying nonlinear mappings from input/output data. The approach is based on the universal approximation by the fuzzy integration of local affine mappings. A connectionist model realizing the universal approximator is suggested by using a processing unit based on both the radial basis function and the weighted sum scheme. In addition, a learning method with self-organizing capability is proposed for the identifying of nonlinear mapping relationships with the given input/output data. To show the effectiveness of our approach, the proposed model is applied to the function approximation and the prediction of Mackey-Glass chaotic time series, and the performances are compared with other approaches.

I. 서 론

* 正會員, 서울대학교 制御計測工學科/制御計測新技術
研究센터

(Dept. of Control and Instrumentation
Eng./ERC-ACI Seoul National Univ..)

接受日字: 1994年8月27日, 수정완료일: 1995年5月1日

입출력간의 비선형사상(nonlinear mapping)관계를 측정된 데이터로부터 식별(identification)하는 기법은 비선형시스템 식별 및 제어, 시계열 예측, 패턴 인식등에 널리 사용될 수 있다. 근래에 이러한 비선형 사상 관계를 근사적으로 식별할 수 있는 신경회로망 모델들이 제시되어 체계적인 학습방법과 병렬처리가

가능한 형태로 구현하기 위한 연구가 활발하다. 이러한 신경회로망 모델은 크게 두가지 형태를 취하고 있다.

하나는 생물학적 뉴런모델에서 착안된 모델로 입력 요소들의 가중합(weighted sum)에 근거한 모델이고, 또 하나는 저장된 패턴과 입력간의 거리에 기반한 모델이다. 전자의 대표적인 모델로는 다층인식자(multilayer perceptron)^[2, 5]가 있으며, 후자는 방사형기저함수망(radial basis function network)으로서 확률신경회로망(probabilistic neural network)^[11], 가우시안요소함수망(Gaussian kernel function network)^[7], 국부조절처리유니트(locally-tuned processing unit)^[10] 등이 이에 속한다.

이 두가지 형태의 신경회로망 모델에 대해 망의 규모, 학습속도, 일반화 성능등에 관하여 많은 논의가 있었다. 비선형사상의 표면(surface) 형태에 따라 상기 두가지 모델은 서로 장단점을 가지고 있다. 즉, 국부용기(bump)형 사상에 대해서는 방사형기저함수망이, 넓은 영역의 선형(linear) 사상에 대해서는 다층인식자 모델이 잘 표현할 수 있는 것으로 알려져 있다^[3]. 그리고 방사형기저함수망은 빠른 학습속도를 갖고 있으나 다층인식자에 비해 많은 매개변수를 필요로 하는 것으로 보고 되어 있다^[10, 6, 11]. 한편 다층인식자는 학습속도가 느려서 온라인 적응(on-line adaptation)을 요하는 응용분야에는 적용이 용이하지 않다.^[10]

본 논문에서는 임의의 비선형 사상을 국부영역에서 유사 사상(affine mapping)으로 근사화하고 이들을 퍼지 추론(Fuzzy Inference) 형태로 통합하여 전체 비선형 사상을 식별하는 모델을 제시한다. 이 모델을 방사형기저함수 형태의 처리노드와 가중합 형태의 처리노드를 이용하여 연결형 모델(connectionist model)로 구성할 수 있는 구조를 제안한다. 또한 이 모델을 이용하여 주어진 입출력 데이터로부터 비선형 사상관계를 식별해 내는 자기구성(self-organizing) 기능을 갖는 학습방법을 제시한다. 마지막으로 함수 근사(function approximation) 문제, 혼돈 시계열 예측(chaotic time series prediction) 문제 등에 제안된 방법을 적용하여 그 타당성을 살펴 본다.

II. 비선형 사상의 식별 모델

임의의 비선형 함수, $f: X \in R^n \rightarrow R$, 관계를 갖는 모델로부터 발생된 입출력 데이터 쌍이 주어 졌을 때, 이 데이터로부터 원래 모델의 입출력 사상관계를 근사적으로 잘 식별해 낼 수 있는 범용 근사자(universal approximator)를 생각해 보자. 이 범용

근사자는 매개변수화된 요소함수를 가지고 있고, 이 요소함수들의 조합(combination)과 매개변수 값의 조정에 의해 임의의 비선형 사상 관계를 식별해 내게된다. 따라서 비선형 사상 식별 모델은 요소함수의 형태, 요소함수들의 조합 방법, 그리고 매개변수의 추정방법 등에 의해 특성지워진다.

본 논문에서는 다음의 유사사상을 요소함수로 사용한다.

$$\phi_i(x) = w_i^T x + b_i, \quad (1)$$

여기서 w_i 는 n차원의 매개변수 벡터이고, $(\cdot)^T$ 는 (\cdot) 의 이항(transpose) 벡터를 의미한다. 이 유사사상은 선형사상에 상수가 더해진 형태로 이는 비선형사상의 1차 근사를 의미한다. 따라서 임의의 비선형 사상이 주어 졌을 때 상기의 유사사상은 국부영역의 사상을 근사적으로 표현할 수 있다고 하면 다음의 퍼지규칙을 설정할 수 있다:

$$\text{Rule}(i): \text{If } x \in N(c_i, \delta_i), \quad (2)$$

$$\text{Then } f(x) = \phi_i(x), \quad i = 1, \dots, M, \quad (3)$$

여기서 $N(c, \delta)$ 는 중심벡터 c_i 를 갖고 반지름이 δ_i 인 근역(neighborhood)을 의미하며 M 은 퍼지규칙의 수를 나타낸다. 상기 퍼지규칙에서 조건부(premise part)는 다음의 방사형기저함수를 이용하여 나타낼 수 있다:

$$\mu_i(x) = \exp(-\|x - c_i\|^2 / \gamma_i), \quad i = 1, \dots, M. \quad (4)$$

여기서 γ_i 는 근역의 크기를 결정하는 경쟁변수(vigilance parameter)이다. 각 규칙의 소속함수(membership function)는 단위분할(partition of unity)^[12]를 만족하기 위해 다음과 같이 정규화된다:

$$\bar{\mu}_i(x) = \frac{\mu_i(x)}{\sum_{j=1}^M \mu_j(x)}. \quad (5)$$

각 규칙의 유사사상을 통합하여 전체 식별함수 F 를 구하기 위해 다음과 같이 소속함수를 이용한다.

$$F(\theta, x) = \sum_{i=1}^M \bar{\mu}_i(x) \phi_i(x), \quad x \in X \quad (6)$$

여기서 θ 는 다음과 같이 정의되는 매개변수의 집합이다:

$$\theta = \{(c_i, w_i, b_i) | c_i \in R^n, w_i \in R^n, b_i \in R, i = 1, 2, \dots, M\} \quad (7)$$

식별함수 (6)이 임의로 주어진 미분가능 함수 f 를

근사화 시킬 수 있음을 다음 정리에서 보인다

정리 1 미분가능한 함수, $f: X \in R^n \rightarrow R$, 가 주어 졌을 때, 임의로 주어진 $\epsilon > 0$ 에 대해, 다음을 만족하는 근역 $N(c_i, \delta_i)$ 가 존재한다:

$$|f(x) - f(c_i) - \nabla f(c_i)(x - c_i)| \leq \epsilon, \quad \forall x \in N(c_i, \delta_i) \quad (8)$$

여기서 $\nabla f(c_i)$ 는 c_i 에서 f 의 경사도(gradient)이다

증명: 부록 1 참조.

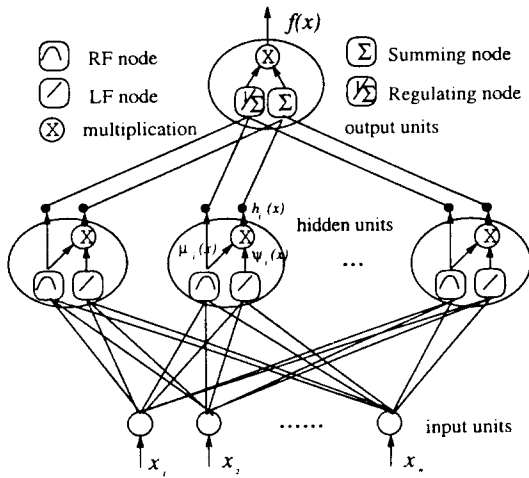


그림 1. 유사사상의 퍼지통합에 기반한 연결형 모델 구조

Fig. 1. Structure of connectionist model based on fuzzy integration of affine mappings.

정리2: 미분가능한 함수, $f: X \in R^n \rightarrow R$, 가 주어 졌을 때, 임의로 주어진 $\epsilon > 0$ 에 대해, 다음을 만족하는 함수 F 가 존재한다.

$$|f(x) - F(\theta, x)| \leq \epsilon, \quad \forall x \in X. \quad (9)$$

증명: 유사사상 (1)의 매개변수를

$$w_i = \nabla f(c_i) \quad (10)$$

$$b_i = f(c_i) - \nabla f(c_i)c_i, \quad (11)$$

와 같이 놓고 정리 1을 이용하면 증명된다. 자세한 내용은 부록 2 참조.

증명 2로부터 (6)으로 정의된 식별함수 F 가 임의의 비선형 사상을 근사적으로 식별할 수 있음을 알 수 있다. 여기서 중요한 관심은 첫째로, 식별함수 F 를 구

현할 수 있는 연결형 모델(connectionist model)을 구성하는 문제이고, 둘째로는 주어진 입출력 데이터로부터 식별함수 F 를 어떻게 찾아 내느냐 하는 문제. 즉 정리 2를 만족하는 매개변수 θ 및 퍼지규칙 수 N 을 학습에 의해 찾아내는 것이다.

본 논문에서 제안된 연결형 모델은 그림 1에서와 같이 3층망(입력층, 은닉층, 출력층)으로 구성된다. 은닉층의 각 유니트는 2개의 노드(Node), 즉 RF(Receptive Field) 노드와 LF(Linear Fitting) 노드로 구성된다. RF 노드는 특정 국부영역의 중심점을 매개변수로 갖고 있으며, 식 (4)에 따라 방사형기저함수 형태의 출력을 발생한다. LF 노드는 식 (1)에 따라 유사사상을 담당하며 상수항을 포함한 매개변수는 선형 알고리즘에 의해 고속으로 학습될 수 있다. 이 RF와 LF 노드의 출력이 곱해져서 은닉층 유니트의 출력이 된다. 즉, i 번째 은닉층 유니트의 출력은 다음과 같이 주어진다

$$h_i(x) = \mu_i(x)\psi_i(x). \quad (12)$$

또한 RF 노드의 출력은 그대로 출력 유니트로 전달된다. 출력 유니트도 두개의 노드로 구성되는데, 하나는 조절노드(Regulating Node)이고 또 하나는 적분노드(Integrating Node)이다. 조절노드는 RF 노드들의 출력을 조절하는데, 이는 RF 노드들의 합이 단위 분할 조건을 만족하도록 정규화한다. 이 조절노드는 입력합의 역을 출력한다. 한편 적분노드는 은닉층 유니트들에 의해 발생된 용기함수들을 단순히 합산하는 역할을 한다. 출력 유니트의 최종 출력은 조절노드의 출력과 적분노드의 출력이 곱해져 출력된다. 이와같이 구성하였을 때 이 연결형 모델은 앞절 식 (6)의 식별함수 F 를 구현할 수 있다.

III. 식별 모델의 학습방법

제안된 식별모델의 학습은 RF 노드를 학습하기 위한 비지도학습(unsupervised learning)과 LF 노드의 학습을 위한 지도학습(supervised learning)을 동시에 사용한다. 제안된 학습방법은 주어진 입력패턴에 따라 필요시 은닉층 유니트를 자동적으로 생성한다. 개괄적인 학습전략은 다음과 같다.

- 새로이 인가된 입력패턴이 현재 생성되어 있는 RF 노드의 수신 영역(receptive field)에 속하면, RF 노드의 출력값이 가장 큰 유니트, 즉 승리 유니트(winning unit)의 매개변수를 다음에 제시되는 학습법칙에 따라 갱신한다.

그렇지 않고 인가된 입력패턴이 속할 RF 노드가 없는 경우 새로운 은닉층 유닛을 생성한다. RF 노드의 중심은 입력벡터를 가지고 초기화하며 LF 노드의 매개변수는 임의로 설정한다.

여기서 i 번째 RF 노드의 수신영역은 다음의 근역으로 정의된다

$$N(c_i, \delta_i) = \{x \mid \|x - c_i\| \leq \delta_i, x \in X\}, \quad (13)$$

여기서 δ_i 는 중심 c_i 에서 부터 수신영역의 경계 (boundary)까지의 거리이다. 입력패턴이 수신영역에 속했는지의 판별은 RF 노드의 출력값을 이용한다. 식 (4)의 $\mu_i(x)$ 가

$$\mu_i(x) \geq \exp(-\delta_i^2/\gamma_i) \quad (14)$$

을 만족하면, i 번째 RF 노드에 속한 것으로 판정한다. 본 논문에서는 편의상 각 RF 노드의 출력을 검사하는 대신에 조절노드의 출력을 이용한다. 조절노드의 출력이 미리 설정된 문턱값 T 이상이면 입력패턴이 속할 RF 노드가 없는 것으로 판정한다. 여기서 γ_i 는 수신영역의 크기를 결정하는 중요한 변수이므로 이의 선택에 따라 식별 오차가 결정된다. γ_i 가 작을 수록 식별오차가 적게 되나 학습패턴 수가 한정되어 있으므로 무작정 작게 할 수는 없다. 따라서 초기에는 γ_i 를 크게 하고 성능을 향상시킬 필요가 있으면 γ_i 를 점진적으로 줄여가는 방법을 쓸 수 있다. 만족스런 성능을 얻을 때까지 이 과정을 연속적으로 반복하면 성공적으로 학습할 수 있다. 또한 γ_i 를 일률적으로 줄이지 않고 유니트 별로 차등을 둘 수도 있다. 즉 가장 큰 오차를 유발하는 유니트의 γ_i 만을 줄이는 방법이다. 이 경우 유니트 별 수신영역이 차이가 나서 경제적인 모델을 구성할 수도 있으나 학습에 소요되는 시간이 길어 질 수가 있다.

RF 노드의 중심은 소속패턴의 평균으로 결정하며 다음의 순환식에 의해 갱신된다.

$$c_i(k+1) = c_i(k) + \frac{1}{k+1}(x - c_i(k)), \quad (15)$$

여기서 k 는 i 번째 RF 노드의 중심 c_i 의 갱신회수이다.

LF 노드의 학습은 다음의 목적함수를 최소화하는 매개변수 $\bar{w}_i = [w_i, b_i]$, $i=1, \dots, M$ 을 구하는 것이다. 각 유니트 별 목적함수는

$$E_i = \sum_{p=1}^{P_i} (y_i(p) - \bar{x}_i^T(p) \bar{w}_i)^2 \quad (16)$$

로 주어지며, P_i 는 근역 $N(c_i, \delta_i)$ 에 속한 입력패턴의 수이고, $y_i(p)$ 는 입력 $x_i(p)$ 에 대한 원래 시스템의 출력이다. $\bar{x}_i(p)$ 는 $(n+1)$ 차원 벡터 $[x_i(p), 1]$ 로 정의된다.

식 (16)를 벡터 형태로 바꾸면

$$E_i = (Y_i - X_i \bar{w}_i)^T (Y_i - X_i \bar{w}_i) \quad (17)$$

이 되는데, 여기서 Y_i, X_i 는

$$Y_i = [y_i(1), y_i(2), \dots, y_i(P_i)]^T \in R^{P_i \times 1} \quad (18)$$

$$X_i = [\bar{x}_i(1), \bar{x}_i(2), \dots, \bar{x}_i(P_i)]^T \in R^{P_i \times (n+1)} \quad (19)$$

로 정의된다. 식 (17)을 \bar{w}_i 로 미분하여 최소점을 구하면

$$(X_i^T X_i) \bar{w}_i = X_i^T Y_i \quad (20)$$

을 만족해야 한다. 자기상관 행렬 (autocorrelation matrix) $X_i^T X_i$ 가 비특이 (nonsingular)하면 역행렬이 존재하여 다음과 같이 유일해 (unique solution)를 구할 수 있다.

$$\bar{w}_i = (X_i^T X_i)^{-1} X_i^T Y_i \quad (21)$$

그렇지 않고 $X_i^T X_i$ 이 특이 (singular)하면 식 (20)은 유일해를 갖지 않는다. 참고로 근역 $N(c_i, \delta_i)$ 가 최소한 $(n+1)$ 개의 선형독립벡터 (linearly independent vector) $\bar{x}_i(p)$ 를 포함해야 $X_i^T X_i$ 가 비특이하게 된다^[4].

임의의 입력패턴 $x \in X$ 에 대해 \bar{w}_i 를 갱신하는 순환식은 다음과 같이 유도될 수 있다.

$$G_i(k) = \frac{R_i(k-1) \bar{x}}{1 + \bar{x}^T R_i(k-1) \bar{x}} \quad (22)$$

$$R_i(k) = R_i(k-1) - G_i(k) \bar{x}^T R_i(k-1), \quad R_i(0) = \alpha I \quad (23)$$

$$\bar{w}_i(k) = \bar{w}_i(k-1) + G_i(k) [f(x) - \bar{x}^T \bar{w}_i(k-1)], \quad (24)$$

여기서 $\bar{x} = [x, 1] \in R^{(n+1)}$ 이고, I 는 단위행렬 (identity matrix)이며, α 는 $\alpha \gg 1$ 을 만족하는 상수이다.

상기 학습의 주요 목적은 최적의 단위분할을 형성하고 각 근역에서 원래의 사상을 최적으로 유사근사 (affine approximation) 하는 매개변수 \bar{w}_i 를 구하

는 것이다. 단위분할 과정이 완료 되었다고 가정하면 최적의 매개변수는 정리 2의 증명에서 알 수 있듯이

$$w_i^* = \nabla f(c_i), \tag{25}$$

$$b_i^* = f(c_i) - \nabla f(c_i)c_i \tag{26}$$

이 된다. 근역 $N(c_i, \delta)$ 에서 학습 데이터가 충분하면, 위 순환식의 \bar{w}_i 는 최적 매개변수 (26) 및 (26)로 수렴함을 쉽게 증명할 수 있다. 그러나 충분하지 않더라도 \bar{w}_i 는 임의의 상수로 수렴한다. 이 경우 시험패턴에 대해 만족할 만한 성능이 보장되지 않는다. 이러한 경우는 학습패턴의 수에 비해 너무 많은 근역으로 분할 했을 경우에 발생할 수 있는데, 이는 망(network)의 규모가 학습패턴의 수에 비해 지나치게 클 경우 학습의 정확도는 높아지나 시험패턴에 대해서는 일반화 성능 (generalization performance)은 나빠지는 것을 의미한다. 위에서 설명한 학습 방법을 정리한 학습 절차는 다음과 같다.

학습 절차

단계 1: 초기화:

- 초기 경쟁변수 γ , 문턱값 T , 및 α 설정한다.

단계 2: 새로운 입력에 대한 모델의 출력계산:

- 생성되어 있는 RF 노드가 없으면, 즉 $M=0$ 이면 단계 4로 간다.
- 그렇지 않고 생성된 RF 노드가 있으면 식별 모델의 출력을 계산한다.

단계 3: 매개변수 갱신:

- 조절노드의 출력이 T 보다 크면 단계 4로 간다.
- 그렇지 않고 T 보다 작으면 승리유니트의 매개변수를 다음과 같이 갱신한다.
 - RF 노드의 중심을 (15)에 따라 갱신한다.
 - LF 노드의 중심을 (24)-(24)에 따라 갱신한다.

단계 4: 새로운 유니트 생성:

- $M=M+1$, 처음엔 $M=0$.
- 새로운 RF 노드의 중심, c_s 는 입력 x 로 초기화한다. 즉 $c_s=x$.
- 새로운 LF 노드의 매개변수, \bar{w}_M 는 임의의 벡터로 초기화한다.

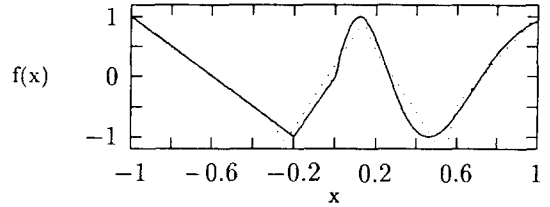
단계 5: 학습성능 평가:

- 입력패턴이 마지막이 아니거나, 규정된 반복횟수보다 적게 학습했을 경우 단계 2로 간다.
- 그렇지 않고, 평균자승오차의 제공근 E_{rms} 이 미리 설정된 값 ϵ 보다 작으면 학습을 완료한다.
- 그렇지 않고, ϵ 보다 크면 모델을 초기화 ($M=0$) 하고 경쟁변수를 $\gamma=\beta\gamma$, $\beta<1$ 와 같이 줄인 후 단

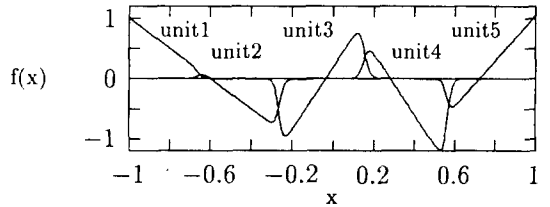
계 2부터 다시 시작한다.

IV. 모의실험 및 성능평가

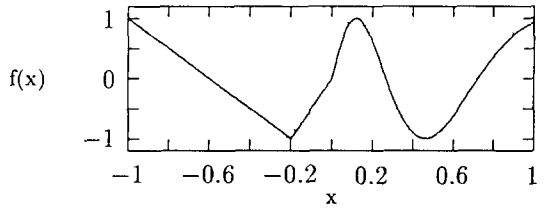
제안된 모델의 은닉층 유니트의 역할을 살펴보기 위해 다음의 1차원 함수를 근사화하는 모델을 구성하였다.



(a)



(b)



(c)

그림 2. 제안된 모델의 1차원함수의 근사능력: (a) 5개의 유니트를 갖는 모델에 의해 근사된 함수, (b) (a)의 모델에서 5개의 유니트에 의해 생성된 각 함수의 형태, (c) 20개의 유니트를 갖는 모델에 의해 근사된 함수

Fig. 2. Approximation capability of the proposed model for a function of one variable: (a) the approximated function by the model with 5 hidden units, (b) the shape of the function generated by each hidden unit in the model of (a), (c) the approximated function by the model with 20 hidden units.

$$f(x) = \begin{cases} -2.5x-1.5 & -1.0 \leq x < -0.2 \\ 5x & -0.2 \leq x < 0.0 \\ \sin [15x/(x+1)] & 0.0 \leq x < 1.0 \end{cases} \quad (27)$$

그림 2(a)에서 5개의 은닉층 유니트에 의해 근사화된 결과를 보여준다. 이때 각 유니트에 의해 발생된 함수들은 그림 2 (b)에서는 보여주는 바와 같이 다양한 형태를 취한다. 그림 2 (c)는 20개의 유니트에 의해 잘 근사된 결과를 보여주고 있다. 이로서 은닉층 유니트를 증가시킴으로써 비선형함수를 임의의 정확도로 근사시킬 수 있음을 보여준다.

또 실제 응용분야에 대한 성능을 알아보기 위해, 임의성(randomness)과 비선형성(nonlinearity)이 강한 Mackey-Glass(M-G)혼돈 시계열 예측문제에 적용하였다. M-G 시계열은 다음과 같다.

$$x(t+1) = (1-a)x(t) + \frac{bx(t-\gamma)}{1+x^{10}(t-\gamma)} \quad (28)$$

여기서 $a=0.1, b=0.2, \gamma=17$ 로 했을 때 임의성과 비선형성이 강한 시계열이 발생된다. (그림 3 참조). 제안된 모델의 입력으로는 다른 방법들^{16, 10, 9, 11}에서와 같이 $x(t), x(t-6), x(t-12), x(t-18)$ 의 4개 과거 데이터가 사용된다. 예측하고자 하는 미래 데이터는 85 스텝뒤의 $x(t+85)$ 가 된다. 제안된 모델의 구성 및 매개변수를 혼란시키기 위하여 식 (28)에서 초기치 $x(0)=0.9$ 에 대해서 $t=2000$ 부터 생성된 500개의 데이터를 사용하였으며 (그림 3), 예측성능을 시험하기 위하여 그 다음 500개의 데이터를 사용하였다(그림 4). 모의실험에서는 학습법의 변수를 $T=100, \alpha=10,000$ 로 설정하였다. γ 는 0.0019부터 0.0001까지 0.0001 간격으로 감소시키며 19개의 모델을 구성하였다. γ 에 따라서, 각 모델은 67~383개의 은닉층 유니트를 생성하였다. 이 중에서 322/290/261/243개 유니트를 갖는 모델 각각은 시험패턴에 대해 4.6%/4.7%/4.8%/5.7%의 E_{rms} (평균자승오차의 제곱근을 출력값의 표준편차로 나눈 값)의 예측성능을 얻었다. 4.8%의 경우 그림 4에서 보듯이 식별오차는 육안으로 거의 식별이 불가능하다. 또한 이의 학습을 위해서 SUN 4/330(16Mips)에서 1분 정도밖에 소요되지 않았다.

표 1에서 이 결과와 다른 여러가지 방법과 성능을 비교하였다. 비교된 방법들은 다층인식자 모델에 속하는 Three-layer Neural Network (TNN)¹⁶, 방사형기저함수망 형태인 Locally-Tuned Processing Units (LTPU)¹⁰, 계층적 자료탐색(table look-up)에 기반한 Multi-Resolution Hierarchies

(MRH)¹⁹ 그리고 다층인식자의 변형 형태인 Cascade-Correlation Network (CCN)¹¹ 방법등이다.

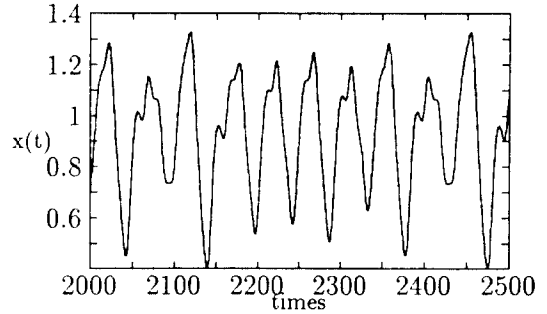


그림 3. 학습에 사용된 혼돈 시계열의 예
Fig. 3. Plot of the chaotic time series for the training data.

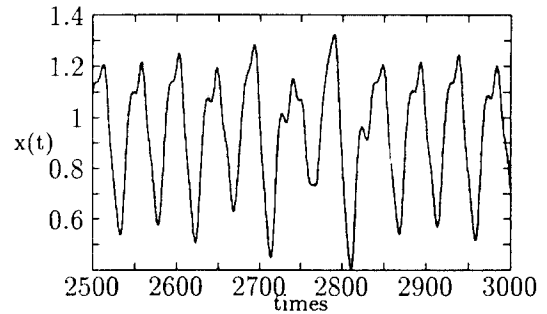


그림 4. 예측성능시험에 사용된 Chaotic 시계열의 예(실선은 실제의 시계열이며, 점선은 제안된 모델에 의해 예측된 시계열로서 육안으로 식별이 거의 불가능함)
Fig. 4. Plot of the chaotic time series for the training data (the solid/dotted line indicates the desired/estimated time series: the difference is almost indiscernible to the naked eye.)

표 1에서 보면, 첫째, MRH 및 LTPU의 학습속도가 빠르나, 많은 조정 매개변수를 필요로 하며 원하는 성능을 얻기 위해 많은 양의 데이터를 필요로 하고; 둘째, TNN은 적은 매개변수를 요하나 계산시간이 매우 많이 요구되며; 셋째, CCN은 예측성능이 가장 나쁘다. 이에 비해 제안된 모델은 LTPU와 MRH방법에 비해 현저히 적은 매개변수를 필요로 하면서 비슷한 학습속도를 보이며, TNN에 비해 800 - 1600 배

의 계산속도 향상을 얻었고, 좋은 예측성능을 보여준다. 일반적으로 500개의 데이터로 학습하고 그 다음 500개 데이터로 시험할 경우 구간에 따라 예측성능에 차이를 보일 수 있다. 표 1의 예측성능 결과는 해당 논문에서 발췌 인용한 것이며, CCN을 제외하고는 최고로 성능이 좋은 구간에 대한 결과이고, CCN 인 경우는 30 여 구간에 대한 평균치이다. 제안된 모델에서는 20 여 구간의 평균치로는 8% E_{rms} 정도의 성능을 나타냈다.

표 1. 제안된 방법과 다른 방법과의 시계열 예측 (85 스텝 뒤) 성능 비교표

Table.1. Comparison of experimental results for the prediction of M-G Chaotic time series.

모델	학습데이터 수	매개변수 개수	계산 시간	Enrms (시험)
MRH	500	17009	3.5분 (Sun 3/60 :4Mips)	0.05
LTPU	10000	5000	31분 (Sun 3/60)	0.061
TNN	500	541	5-10분 (Cray X/MP :90Mflops)	0.05
제안된 모델	500	1450	1분 16초2 (Sun 4/330 :16Mips)	0.048
CCN	500	693	265 백만번 곱셈	0.32

V. 결론

본 논문에서 제안된 모델은 다음과 같은 장점을 가지고 있다. 첫째는 다층인식자 모델에 비해 현저히 적은 계산시간을 요하고, 학습속도가 빠르다. 둘째는 원하는 성능을 얻기위해 자동으로 망을 구성함으로써 국부최소점 문제가 해결된다. 셋째는 제안된 모델에서는 한번 학습시킨 유니트만 학습에 참가함으로써 온칩 학습을 위한 하드웨어가 현격히 줄어 온칩 학습 (On-chip Learning)이 가능한 VLSI로 구현이 용이하다. 넷째는 각 은닉층 유니트가 국부영역에 대해 유사사상 즉 상수항을 포함한 선형사상을 표현하므로 선형 시스템이론을 용이하게 도입할 수 있어 시스템 분석 및 설가 용이하다. 다섯째는 전문가 규칙이 주어지면 용이하게 초기망 구성에 활용할 수 있다.

제안된 모델에서 방사형기저함수망 대신에 타원형기저함수망을 도입하고 국부영역을 최적으로 학습하기 위한 학습방법등을 사용할 시 더 효율적인 식별모델을 구성할 수 있다. 또한 학습속도가 빠르고 계산의 복잡도가 적어 실시간 적응학습을 요하는 분야, 즉, 비선형성, 시변성, 불안정성이 강한 이동통신채널에서의

적용 이퀄라이저, 비선형 시스템의 적응제어 등에 효율적으로 적용될 수 있으리라 기대된다.

부 록

부록 1: 정리 1의 증명

f 가 미분가능하므로 그의 도함수 (derivative)는 연속이다. 이는 주어진 $\epsilon > 0$ 에 대해

$$|\frac{\sigma f(x)}{\alpha_j} - \frac{\sigma f(c_j)}{\alpha_j}| < \frac{\epsilon}{n\delta_j}, \forall x \in N(c_j, \delta_j), j = 1, 2, \dots, n \quad (29)$$

을 만족하는 근역 $N(c_j, \delta_j), \delta_j > 0$ 가 존재함을 의미한다.

벡터 h 를 $h = x - c_j = \sum_{j=1}^n h_j e_j$ 로 정의하자. 여기서 e_j 는 단위벡터 (unit vector), 즉 $e_j = (0, \dots, 1, \dots, 0)$ 이고, h_j 는 h 의 j 번째 요소이다. 또 q_k 를

$$q_k = \sum_{j=1}^k h_j e_j, \quad k = 1, 2, \dots, n, \quad (30)$$

로, q_0 를 영벡터 (zero vector)로 정의하자. 그러면 모든 k 에 대해 $\|q_k\| \leq \|h\|$ 이 되고 따라서

$$\begin{aligned} |f(x) - f(c_j) - \sum_{k=1}^n \frac{\partial f(c_j)}{\partial x_k} h_k| &= | \sum_{k=1}^n (f(c_j + q_k) - f(c_j + q_{k-1})) - \frac{\partial f(c_j)}{\partial x_k} h_k | \\ &\leq \sum_{k=1}^n |f(c_j + q_k) - f(c_j + q_{k-1}) - \frac{\partial f(c_j)}{\partial x_k} h_k| \end{aligned} \quad (31)$$

이 된다.

식 (31)의 k 번째 항을 살펴 보면, 벡터 $c_j + q_k$ 와 $c_j + q_{k-1}$ 는 k 번째 요소만 다르다. 즉, $c_j + q_k = c_j + q_{k-1} + h_k e_k$ 이 된다. 따라서 단변수함수의 평균치의 정리 (mean value theorem)를 적용하면, 임의의 $a, 0 \leq a \leq h_k$ 에 대해

$$f(c_j + q_k) - f(c_j + q_{k-1}) = \frac{\partial f(c_j + q_{k-1} + a e_k)}{\partial x_k} h_k \quad (32)$$

이 성립한다.

또한 $\|h\| \leq \delta_j$ 이므로 $c_j + q_{k-1} + a e_k \in N(c_j, \delta_j)$ 이 된다. 그러므로 (29)와 (32)은 모든 $h, \|h\| < \delta_j$ 에 대해

$$|f(c_j + q_k) - f(c_j + q_{k-1}) - \frac{\partial f(c_j)}{\partial x_k} h_k| \leq \frac{\epsilon}{n\delta_j} \|h\| \leq \frac{\epsilon}{n}, \quad (33)$$

를 의미한다. 따라서 (31) 및 (33)로부터

$$|f(x) - f(c_j) - \sum_{j=1}^n \frac{\partial f(c_j)}{\partial x_j} h_j| \leq \epsilon \quad (34)$$

이 증명된다. Q.E.D.

부록 2: 정리 2의 증명

정리 1로 부터, 임의로 주어진 $\epsilon > 0$ 에 대하여 다음을 만족하는 근역 $N(c, \delta)$ 가 존재한다:

$$|f(x) - f(c) - \nabla f(c)(x - c)| \leq \epsilon/2, \quad \forall x \in N(c, \delta). \quad (35)$$

이는 $\bigcup_{i=1}^M N(c_i, \delta)$ 가 정의역 X 의 덮개(covering)가 되고 각 근역은 식 (35)을 만족하는 근역의 집합 $\{N(c_i, \delta) \mid i=1, \dots, M\}$ 을 구성할 수 있음을 의미한다. 여기서 식 (1)의 매개변수를

$$w_i = \nabla f(c_i) \quad (36)$$

$$b_i = f(c_i) - \nabla f(c_i)c_i \quad (37)$$

와 같이 놓으면 식 (35)는

$$|f(x) - \phi_j(x)| \leq \epsilon/2, \quad \forall x \in N(c_i, \delta_i) \quad (38)$$

와 같이 다시 쓸 수 있다. 한편 임의의 근역 $N(c_i, \delta_i)$ 내의 벡터 x 에 대하여,

$$\begin{aligned} & |f(x) - F(\theta, x)| \\ & \leq |f(x) - \phi_j(x)| + |\phi_j(x) - \sum_{j=1}^M \bar{\mu}_j(x) \phi_j(x)| \\ & \leq \epsilon/2 + (|1 - \bar{\mu}_j(x)| + \sum_{j \neq i}^M \bar{\mu}_j(x)) \max \{|\phi_k(x)|\} \end{aligned} \quad (39)$$

여기서 $\phi_j(x), j=1, \dots, M$ 은 근역 $N(c_i, \delta_i)$ 내에서 한정(bound)되어 있고 $\bar{\mu}_j, j \neq i$ 은 γ_j 에 의해 임의로 작게 할 수 있다. 또한 $\bar{\mu}_i$ 는 충분히 1에 가깝게 할 수 있으므로, 주어진 ϵ 에 대해

$$\begin{aligned} & |1 - \bar{\mu}_j(x)| \max \{|\phi_k(x)|\} \leq \epsilon/4 \\ & \left| \sum_{j \neq i}^M \bar{\mu}_j(x) \right| \max \{|\phi_k(x)|\} \leq \epsilon/4 \end{aligned} \quad (40)$$

을 만족하는 소속함수 $\mu_j(x), j=1, \dots, M$ 이 존재한다. 따라서 식 (39)와 (40)으로 부터

$$|f(x) - F(\theta, x)| \leq \epsilon, \quad \forall x \in X. \quad (41)$$

임이 증명된다. Q.E.D.

참 고 문 헌

[1] R.S. Crowder, "Predicting the mackey-glass time series with cascade-cor

relation learning," Proceedings of the 1990 Connectionist Models Summer School, pages 117-123, 1990.

[2] K.Funahashi, "On the approximate realization of continuous mappings by neural networks," Neural Networks, 2:183-192, 1989.

[3] N.Manukian G.D.Wilensky, "The projection neural network," IEEE International Joint Conference on Neural Networks, 2, 1992.

[4] G.C. Goodwin and R.L. Payne, Dynamic System Identification: Experiment Design and Data Analysis, Academic Press, 1977.

[5] K.Hornik, "Approximation capabilities of multilayer feedforward networks," Neural Networks, 4:251-257, 1991.

[6] A.S. Lapedes and R.Farber, "Nonlinear signal processing using neural networks: Prediction and system modeling, Technical Report LA-UR-87-2662, Los Alamos National Laboratory, 1987.

[7] S.Lee and R.M. Kil, "A gaussian potential function network with hierarchically self-organizing learning," Neural Networks, 4(2):207-224, 1991.

[8] J.E. Marsden, Elementary Classical Analysis, Freeman and Company, 1974.

[9] J.Moody, "Fast learning in multi-resolution hierarchies," Neural Information Processing Systems, pages 29-39, 1989.

[10] J.Moody and C.J. Darken, "Fast learning in networks of locally-tuned processing units," Neural Computation 1, pages 281-283, 1989.

[11] D.F. Specht, "Probabilistic neural networks and the polynomial adaline as complementary techniques for classification," IEEE Trans. on Neural Networks, 1(1):111-121, 1990.

[12] M.Spivak, "Calculus on Manifold," W. A.Benjamin, New York, 1965.

저 자 소 개



崔 鎮 榮(正會員)

1982년 2월 서울대학교 제어계측공학과 학사. 1984년 2월 서울대학교 제어계측공학과 석사. 1993년 2월 서울대학교 제어계측공학과 박사. 1984년 3월 - 1994년 2월 한국전자통신연구소 연구원 및 선임연구원. 1994년 3월 - 현재 서울대학교 제어계측공학과 전임강사. 1994년 4월 현재 서울대학교 자동화시스템공동연구소 겸임교수. 1994년 - 현재 서울대학교 제어계측신기술연구센터 참여교수. 1994년 12월 - 현재 서울대학교 자동제어특화연구센터 참여교수. 관심분야는 Neuro-Fuzzy System, Adaptive & Learning System, Adaptive Control of Autopilot, Detection & Recognition System.

崔 悰 鎬(正會員) 제 27권 B편 제 2호 참조.

현재 서울대학교 제어계측공학과 교수