유전 알고리즘을 이용한 웨이브릿 모듈라 신경망의 최적 구조 설계

Optimal Structure of Wavelet Modular Wavelet Network Systems Using Genetic Algorithm

최영준, 서재용, 연정흠, 전홍태

Young-Jun Choi, Jae-Yong Seo, Jung-Heum Yon, Hong-Tae Jeon

중앙대학교 전자·전기공학부 지능시스템 및 휴먼로봇연구실

E-Mail: yj0263@ms.cau.ac.kr, sjyong@ms.cau.ac.kr

ABSTRACT

In order to approximate a nonlinear function, modular wavelet networks combining wavelet theory and modular concept based on single layer neural network have been proposed as an alternative to conventional wavelet neural networks and kind of modular network. Modular wavelet networks provide better approximating performance than conventional one. In this paper, we propose an effective method to construct an optimal modualr wavelet network using genetic algorithm. This is verified through experimental results.

1. 서론

일반적인 신경망은 서로 다른 은닉층을 학습시킬 때 가중치 갱신으로 이전에 학습한 매핑 결과를 잊어버리는 일시적 혼선(Temporal crosstalk)현상이 발생한다. 따라서 지역적, 부분적으로 중요한 변화를 가지는 대상함수의 학습이 어렵게 된다. 이러한 현상을 방지하기 위해서 신경망 노드의 상호 연관성과 중복 학습을 줄임으로서 지역화된 응답 특성을 갖는 모듈화된 신경망을 사용한다.[1][2]

일반적으로 단일 신경망은 전역적인 매핑 구조로 수행능력이 우수하지만 느린 학습속도와 응용범위가 한정되는 단점이 있고, 방사기저함수망(RBF)은 지역적인 매핑 구조로 학습속도가 빠른 반면 기저함수들이 일반적으로 직교하지 않기 때문에 필요 이상의 기저함수의 사용에 따른 고유한 망구성의 방해 요인이 된다. 이러한 각각의 지역적·전역적인 매핑방법의 장점을 수용하고 전체 하나의 복잡한 문제(complex problem)를 단순한 여러개의 부분적인-문제(sub-problem)로 나누어 해결하는 방식을 수용한 모듈화 설계방법이 관심의 대상이 되고있다.[1][2]

한편, 신경회로망과 웨이블릿 이론을 결합한 직교성을 갖는 기저함수를 갖는 웨이블릿 신경회로망(Wavelet Neural Network)은 RBF 신경회로망의 대부분의 장점을 유지하고, 미지의 함수에 대하여 효율적이고 고유한 망구성을 제공한다. 그러나 웨이블릿 함수들은 강한 직교 조건을 만족해야하며 이러한 직교 조건은 망의 유연성을 저해하는 요소로 작용될 수 있다.[3]

이에 본 논문에서는 직교성을 갖는 웨이블릿 함수 대신에 망에 유연성을 제공하고 직교 웨이블릿 함수와 비슷한 알고리즘을 제공하는 웨이블릿 프레임(Frame)함수을 가지고 Jacobs와 Jordan의 혼합 모듈라 신경망(Mixtured Modular Neural Network)으로 망을 구성한 뒤, 망의 각 파라미터들을 적절히 구성하고 학습시키는 최적 구조 설계 알고리즘을 제안한다. 제

안한 모듈라 웨이블릿 신경회로망의 최적 구조 설계 알고리즘은 최적화 알고리즘으로 알려진 유전 알고리즘을 사용하여 문제 해결에 적합한 최적의 기저 함수를 이용하여 모듈라 웨이블릿 신경망을 구성하고, 확률적인 강화 학습으로 각 모듈 은닉층의 가중치를 조절한다. 제안한 모듈라 웨이브릿 신경망의 최적구조설계 알고리즘의 우수성을 검증하기 위해 함수 근사화 문제에 적용하고 모의 실험을 통하여 그 우수성을 검증하고자 한다.

2. 가우스 혼합 모듈라 신경망

Jacobs와 Jordan이 제안한 가우스 혼합 모듈라 신경망의 구조는 그림 1과 같다. 하나의 문제에 대해 두 개 또는 그 이상의 독립적인 서로 다른 모듈로 분할되고, 경쟁하는 N개의 엑스퍼트 네트워크와 각각 그것들의 중재역할을 담당하는 1개의 게이팅 네트워크로 구성되어 있다.

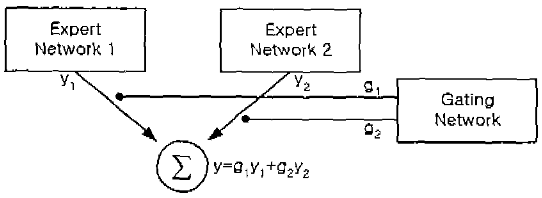

그림 1. 모듈라 웨이브릿 네트워크

여기서 엑스퍼트 네트워크의 수와 게이팅 네트워크의 출력수가 일치해야 하며, 이러한 제약조건을 만족시키고 단조증가 함수인,

$$g_i = \frac{e^{\frac{s_i}{T}}}{\sum_{j=1}^{n} e^{\frac{s_j}{T}}} \tag{1}$$

와 같은 softmax 함수를 사용한다. 여기서 s_i는 i번째 가중치와 입력의 곱의합을 나타내고 T는 “Temperature” 파라미터를 나타낸다. 그리고 g_i는 모든 i에 대해서 $0 \le g_i \le 1$과 $\sum_{j=1}^{M} g_i = 1$ 를 만족해야 한다.

최종 출력은 다음과 같이 나타낼수 있다.

$$\boldsymbol{y} = \sum_{i=1}^{M} g_i \boldsymbol{y}_i \tag{2}$$

여기서 i는 모듈의 개수이고 $\boldsymbol{y}_i$는 i번째 엑스퍼트 네트워크의 출력 벡터이다.

엑스퍼트, 게이팅 네트워크의 가중치는 확률적인 강화학습법과 체인룰을 적용하여 다음과 같은 단조증가 형태의 로그 비용함수(Cost function)가 최대가 되도록 동시에 조절되어 진다.

$$\ln L = \ln \sum_{i=1}^{n} \frac{g_i}{\sigma_i} e^{-\frac{1}{2\sigma_i^2}\|\boldsymbol{d}-\boldsymbol{y}_i\|^2} \tag{3}$$

$\boldsymbol{d}$는 원하는 목적 출력벡터, σ_i 는 엑스퍼트와 관계된 스케일링 파라미터이다.

응답벡터 $\boldsymbol{d}$는 회귀과정(regressive progress)에 따라 선택된 i번째 규칙에 의해서 발생된다.

$$\boldsymbol{d} = \boldsymbol{F}_i(\boldsymbol{x}) + \varepsilon_i, \quad i = 1, 2, \cdots, M \tag{4}$$

여기서 ε_i는 평균이 0이고 분산 행렬이 $\sigma^2 \boldsymbol{I}$ 인가우스 분포이고, $\boldsymbol{F}_i(\boldsymbol{x})$는 어느 표본함수의 미래값을 과거의 값으로 정확하게 예측할 수 있는 결정론적(deterministic)인 함수이다. 위와 같은 구조를 학습하는데 사용되어지는 알고리즘의 궁극적인 목적은 학습패턴 $\{x, d\}$의 확률 분포 (Probability distribution)를 모델링(Modeling) 하는 것이다. 위(1)식과 체인룰을 통해서 사후확률(Posteriori probability)을정의할 수 있는데, i번째 엑스퍼트 네트워크의 출력은 아래의 식 (5)과 같이 다변수 가우스 분포(multivariate Gaussian distribution)의 조건부 평균(conditional mean)으로 볼 수

$$\begin{aligned} \boldsymbol{y}_i &= \boldsymbol{\mu}_i \\ &= E[\boldsymbol{d}|\boldsymbol{x}, i] = \boldsymbol{F}_i(\boldsymbol{x}), \ (i = 1, 2, \cdots, M) \end{aligned} \tag{5}$$

있고, 결국 게이팅 네트워크 출력을 뜻하는 사전확률(prior probability) g_i 는 가중치 갱신으로 사후확률을 추종하게 된다.

3. 웨이블릿 이론과 웨이블릿 신경회로망

웨이블릿 이론에 의하면, $L^2(R)$ 공간을 직교분해할 수 있는 웨이블릿 함수 $\phi(t)$가 존재한다[3].

$$L^2(R) = \cdots \oplus W_{-1} \oplus W_0 \oplus W_1 \oplus \cdots \quad (6)$$

$$\psi_{m,n}(t) = 2^{m/2}\psi(2^m t - n) \quad (7)$$

여기서 W_m은 $[2^{m/2}\psi(2^m t-n)]_{n=-\infty}^{n=+\infty}$으로 표시할 수 있는 부분공간이며, m,n은 정수이다.웨이블릿 $\psi(t)$는 스케일링 함수라 알려진 $\varphi(t)$로부터 구할 수 있으며 스케일링 함수의 신축항과 이동항은 $L^2(R)$의 다중 분해 분석(multiresoulution analysis, MRA)을 가능케하고 다음과 같은 관계를 성립케한다[3].

$$\cdots \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \cdots \subset L^2 \quad (8)$$

여기서 V_m은 $[2^{m/2}\varphi(2^m t-n)]_{n=-\infty}^{n=+\infty}$으로 표시할 수 있는 부분공간이다.

또한, V_m은 다음과 같이 W_m과 연관되어진다.

$$V_{m+1} = V_m \oplus W_m \quad (9)$$

이로부터 $L^2(R)$공간의 임의의 함수 f(t)가 다음과 같이 분해됨을 알 수 있다[3].

$$f(t) = \sum_n \langle f, \varphi_{m_0,n}\rangle \varphi_{m_0,n}(t) + \sum_{m \ge m_0, n} \langle f, \psi_{m,n}\rangle \psi_{m,n}(t) \quad (10)$$

또는 $f(x) = \sum_{m,n} \langle f, \psi_{m,n}\rangle \psi_{m,n}(t)$ (11)여기서 $\langle \cdot,\cdot \rangle$는 내적을 표시하고, m_0는 가장 낮은 분해능(resolution)을 표시하는 임의의 정수이다.

상기의 논의로부터 다음과 같이 웨이블릿 신경회로망을 구성할 수 있다.

$$f(\boldsymbol{x}) = \sum_i w_i \psi_d(\boldsymbol{a}_i \boldsymbol{x} - \boldsymbol{t}_i) \quad (12)$$

여기서, $\boldsymbol{a} \in \boldsymbol{R}^d$, $\boldsymbol{t} \in \boldsymbol{R}^d$는 각각 신축과 이동 변수이고, $\boldsymbol{x} \in \boldsymbol{R}^d$는 웨이블릿 함수의 입력값이며, $\boldsymbol{w} = (w_1, w_2, \cdots, w_m)^T$는 가중치 벡터를 나타낸다. 한편, 직교 웨이블릿 함수는 앞서 서술한 바와 같이 실제 응용에 있어서 부적절한 면이 있으므로, 웨이블릿 프레임 함수를 사용한다.

4. 유전 알고리듬을 이용한 모듈라 웨이블릿 신경망의 설계

본 논문에서 사용된 유전 알고리듬의 목표는 근사화 대상함수에 대하여 모듈라 웨이블릿 신경망의 각 모듈 은닉층의 신축항과 이동항의 최적값을 찾아내는 것이다. 이를 위하여, 다음과 같이 유전알고리듬을 수행한다.

[단계 1] 각 모듈의 신축항과 이동항을 모두 표현할수 있도록 스트링의 길이를 결정한 후 이를 2진수로 표현하여 초기 스트링을 구성한다.

[단계 2] 단계 1의 초기 스트링을 사용하여 초기 집단을 구성한다.

[단계 3] 집단에 있는 각각의 스트링을 10진수로 변환하여 모듈라 웨이블릿 신경회로망에 대입하여 각 집단의 적합도 값을 계산한다. 변환시, 신축항의 최소값은 적은 수의 웨이블릿 기저 함수를 사용하여도 근사화 대상 함수의 영역을 충분히 포함할 수 있도록 설정하며 최대값은 대상 함수의 입력 데이타 분포에 맞게 적절히 설정한다. 또한, 이동항은 해당되는 신축항 값에 따라서 웨이블릿 기저 함수가 이동하여도 대상함수의 영역을 벗어나지 않도록 최대값, 최소값을 설정하여 변환되도록 한다. 각 집단의 적합도 값은 변환된 신축항과 이동항으로 웨이블릿 신경회로망을 고정시킨 후에 은닉층과 출력층 사이의 가중치 값을 확률적인 강화학습법을 이용하여 학습하여 구한다. 강화학습시 너무 많은 반복 횟수는 많은 시간이 소요되므로 각 집단의 적합도 차이를 알 수 있는 선에서 학습을 중단한다. 적합도는 다음 식에 의하여 구한다.

$$\text{fitness} = \frac{1}{\sqrt{\sum_{i=1}^{n}(f(x_i)-g(x_i))^2/n+1}} \quad (13)$$

여기서 $f(x_i)$와 $g(x_i)$는 각각 i번째 입력의 근사화 대상함수와 모듈라 웨이블렛 신경회로망의 출력이다.

[단계 4] 유전자 집단에서 높은 적합도 값을 갖는 스트링에 높은 선택 확률을 주고 유전자들 사이에서 교배나 돌연변이의 진화과정을 통해 새로운 집단을 얻는다.

[단계 5] 새로운 최고의 스트링이나 원하는 적합도 값이 얻어질때까지 상기 단계를 반복한다.

[단계 6] 최종적으로 얻어진 스트링으로 망을 고정시킨 후 각 모듈의 은닉층과 출력층 사이의 가중치 값을 확률적인 강화학습법을 이용하여 충분히 학습시켜 모듈라 웨이블릿 신경망의 최종 형태를 완성한다.

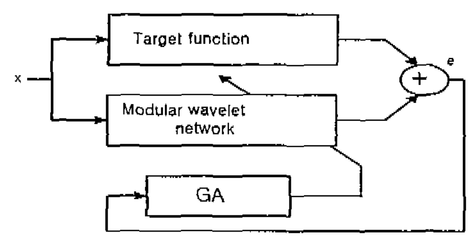

그림 2. 전체 블록도

5. 모의실험

모듈라 웨이블릿 신경망에 사용된 웨이블릿 프레임 함수는 가우시안 함수의 미분형태로 알려진 $\psi(x) = x\,e^{-\frac{1}{2}x^2}$ 이고, 이 함수는 기술한 웨이브릿 프레임 조건 및 함수 근사 이론을 만족한다. 구간별 정의된 함수는 다음과 같다. 그림 2에서 다음 함수의 근사 결과를 보였다.

$$f(x) = \begin{cases} -2.186x - 12.864 & [-10 \le x < -2] \\ 4.246x & [-2 \le x < 0] \\ 10\,e^{-0.05x-0.5} \cdot \sin[(0.03x+0.7)x] & [0 \le x \le 10] \end{cases} \tag{14}$$

유전 알고리즘에 사용된 제어 파라미터들은 표 1과 같다.

표 1. GA parameter

개체군 크기	교배율	돌연변이율
30	100%	70%

표 2. 제안한 알고리즘에 의해 얻어진 팽창항과 이동항

모듈 1			모듈 2			모듈 3		
뉴런	팽창	이동	뉴런	팽창	이동	뉴런	팽창	이동
1	2^{-1}	-4	1	2^{-2}	-3	1	2^{0}	-3
2	2^{-2}	-3	2	2^{-1}	-3	2	2^{-1}	5
3	2^{-2}	2	3	2^{-1}	-4	3	2^{-1}	3
4	2^{-1}	0	4	2^{-1}	2	4	2^{-2}	-1
5	2^{0}	-4	5	2^{-2}	5	5	2^{-1}	3

각 객체의 이진 정보를 적절히 팽창, 이동으로 변환시킨 다음 모듈라 웨이브릿 신경망에 대입, 100번 정도의 확률적인 강화학습을 통해서 표 2. 와 같이 최적의 팽창항과 이동항을 얻었다. 각 모듈 5개의 뉴런을 사용하였고, 신축항은 함수의 범위(-10~10)에 맞춰 (-2~3)까지의 범위를 주었고, 수행조건은 학습률 $\eta = 0.001$, 이다. 최적 구조로 구성된 모듈라 웨이블렛 신경망을 이용하여 2000번 반복학습이 끝난 후 근사화를 수행한 결과를 나타낸 것이다.

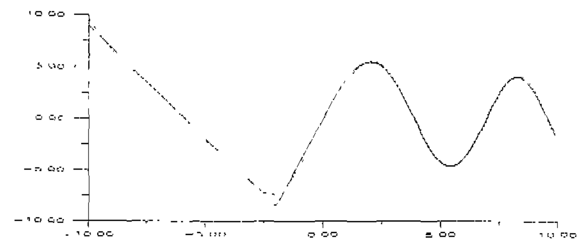

그림 3. 유전 알고리듬으로 구성된 모듈라 웨이블릿 신경망의 근사 결과

6. 결론

본 논문에서 제안한 모듈라 웨이블렛 신경망은 혼합 모듈라 신경망과 유사한 구조로 엑스퍼트 네트워크에 해당하는 각 모듈을 웨이블렛 신경망으로 구성하고 게이팅 네트워크는 다층 신경망으로 구성하였다. 그리고 확률론적인 강화학습 알고리즘을 통해 최적구조를 설계할 수 있는 방법론을 제시 하였다. 모의실험 결과 새로운 구조인 모듈라 웨이블릿 신경망과 최적구조 설계 알고리즘에 의하여 적은수의 뉴런으로도 좋은 근사 결과를 나타냄을 알 수 있다. 끝으로 개선할 점은 앞으로 좀더 다양한 문제에 적용하여 제안한 구조와 최적화 알고리즘의 우수성을 검증하는 것이다.

7. 참고문헌

[1] Jacobs R. A. and Jordan M. I., "Learning Piecewise Control Strategies in a Modular Neural Network Architecture,", *IEEE Trans. Sys. Man and Cybernetics*, vol. 23, no. 2, pp. 337-345, 1993.

[2] Gasser Auda and Mohamed Kamel, "Modular Neural Network Classifiers: A Comparative Study," *Journal of Intelligent and Robotic Systems*, Vol. 21, pp, 117-129, 1998.

[3] Q. Zhang and A. Benveniste, "Wavelet networks," *IEEE Trans. Neural Networks*, vol. 3, pp. 889-898, Nov. 1992.

[4] David E.Goldberg, Genetic Algorithms in Search, Optimization,and Machine Learning, Addison-Wesley Publishing,1989.