

# 모듈화된 웨이블릿 신경망의 적응 구조

## Adaptive Structure of Modular Wavelet Neural Network

서재용\*, 김용택\*\*, 김성현\*\*\*, 조현찬\*, 전홍태\*\*

\*한국기술교육대학교 정보기술공학부

\*\*중앙대 전자전기공학부

\*\*\*동원대학 전자과

Jae-Yong Seo\*, Yong-Taek Kim\*\*, Seong Hyun Kim\*\*\*,

Hyun-Chan Cho\*, Hong-Tae Jeon\*\*

\*School of Info. & Tech. Eng. at Korea University of Tec. & Edu.

\*\*School of Electrical and Electronics Engineering, Chun-Ang University

\*\*\*Dept. of Electronics, Tongwon College

E-mail : sjyong@kut.ac.kr

### ABSTRACT

In this paper, we propose an growing and pruning algorithm to design the adaptive structure of modular wavelet neural network(MWNN) with F-projection and geometric growing criterion. Geometric growing criterion consists of estimated error criterion considering local error and angle criterion which attempts to assign wavelet function that is nearly orthogonal to all other existing wavelet functions. These criteria provide a methodology that a network designer can constructs wavelet neural network according to one's intention. The proposed growing algorithm grows the module and the size of modules. Also, the pruning algorithm eliminates unnecessary node of module or module from constructed MWNN to overcome the problem due to localized characteristic of wavelet neural network which is used to modules of MWNN. We apply the proposed constructing algorithm of the adaptive structure of MWNN to approximation problems of 1-D function and 2-D function, and evaluate the effectiveness of the proposed algorithm.

**Keywords** : Modular Wavelet Neural Network, Growing and Pruning Algorithm, Adaptive Structure

### 1. 서론

신경망에 대한 모델 선택과 망의 성장 및 소거 알고리즘에 대한 연구가 활발하게 진행되고 있다[1]. 또한 모듈화된 설계방식의 장점이 인식되면서 Jordan et al.이 제안한 모듈화된 신경망(Modular Neural Network; MNN)의 성장 및 소거 알고리즘에 대한 연구결과가 보고되고 있다[2][3].

제안한 방식들의 대부분은 소거 알고리즘을

적용하기 위해 초기에 구성하는 MNN의 적합한 크기를 결정하는 문제와 다분히 시행착오적인 방법을 사용하였다.

따라서 본 논문에서는 시행착오적인 방법보다는 대상 문제에 적응적인 모듈화된 웨이블릿 신경망(MWNN)의 성장과 소거 알고리즘을 제안한다. 제안한 알고리즘은 F-투영 방법과 성장기준을 바탕으로 전역오차가 감소하도록 엑스퍼트 네트워크(EN)의 팽창과 이동계수를 결

정하여 노드를 추가한다[4]. 전역오차가 원하는 문턱치에 도달하여 성장 알고리즘이 종료되면 전역오차의 감소에 기여하지 못하는 각 EN의 노드를 제거하는 소거 알고리즘을 수행하여 최적의 MWNN의 구조를 얻는다.

MWNN의 게이팅 네트워크(GN)은 비선형성이 강한 다층신경망으로 구성하고, 학습 알고리즘은 기울기 학습 방법과 Expectation-Maximization(EM) 알고리즘을 결합하여 사용한다. EM 알고리즘은 수렴 속도가 우수하다는 장점이 있다[5][6].

## II. EM 알고리즘을 이용한 MWNN의 학습 방법

EM 알고리즘의 반복(iteration)은 추정단계(E-step)와 극대화 단계(M-step)인 두 단계로 구성된다. M-단계는 E-단계의 각 반복 과정에서 재 정의된 유사 함수를 극대화하는 과정이다. EM 알고리즘은 먼저 주어진 데이터나 현재 모델에서 완전한 데이터 유사 함수의 기댓값을 정의한다. 이것이 E-단계이다.

$$C_F(\theta, \theta^i) = E[l_c(\theta, D) | X] \quad (1)$$

여기서  $\theta^i$ 는  $i$ 번째 반복에서 매개변수 값,  $E[\cdot]$ 는  $\theta^i$ 에 대한 기댓값,  $X$ 는 불완전 데이터,  $l_c(\theta, D)$ 는 완전 데이터 유사 함수,  $D$ 는 완전 데이터이다.

M-단계에서는 결정론적인 함수  $Q$ 를 고려하여 새로운 매개변수  $\theta^{i+1}$ 을 추정하기 위해  $\theta$ 와 관련된 함수를 극대화한다.

$$\theta^{i+1} = \arg \max_{\theta} Q(\theta, \theta^i) \quad (2)$$

다시 E-단계에서 완전한 데이터 유사 함수를 개선하기 위한 추종이 이루어지고 이 과정을 반복하게 된다. 이러한 일반적인 EM 알고리즘을 Xu et al.가 MNN에 적용하여 학습 알고리즘을 개발하였다.

### 2.1 EN의 가중치 갱신 규칙

MWNN의 구조에서  $t$ 번째 학습 데이터에 대한 MWNN의 출력은 다음과 같다[7].

$$y_k^j(\mathbf{x}^t, \mathbf{w}_k^j) = \sum_{m=1}^M w_{km}^j \psi'_{k, mn}(\mathbf{x}^t) \quad (3)$$

다음은  $t$ 번째 학습 데이터가 주어질 때 비용

함수에 대한  $k$ 번째 모듈에서  $j$ 번째 가중치 벡터  $\mathbf{w}_k^j = [w_{k1}^j \ w_{k2}^j \ \dots \ w_{kM}^j]^T$ 의 변화량은 체인룰을 적용하면 다음과 같이 표현할 수 있다.

$$\frac{\partial C_F}{\partial w_k^j} = \frac{\partial C_F}{\partial y_k^j(\mathbf{x}^t, \mathbf{w}_k^j)} \frac{\partial y_k^j(\mathbf{x}^t, \mathbf{w}_k^j)}{\partial w_k^j} \quad (4)$$

결국 M-단계에서  $k$ 번째 모듈인 WNN의 가중치 벡터는 다음 식과 같이 미세조정 값에 의해서 매 반복마다 갱신된다.

$$w_k^j(i+1) = w_k^j(i) + \eta \sum_{t=1}^N h_k^t \frac{e_k^j}{\sum_k} \cdot \phi_{k, mn}(\mathbf{x}^t) \quad (5)$$

여기서  $i$ 는 반복을 나타낸다.

### 2.2 GN의 가중치 갱신 규칙

$t$ 번째 학습 데이터가 주어질 때 비용함수에 대한 GN의 출력층 가중치  $a_k$ 의 변화량은 체인룰을 적용하여 다음과 같은 식을 얻을 수 있다.

$$\frac{\partial C_F}{\partial a_k} = \sum_{t=1}^N \sum_{k=1}^K h_k^t (1 - g_k(\mathbf{x}^t, a_k)) \frac{\partial u_k}{\partial a_k} \quad (6)$$

여기서  $g_k(\mathbf{x}^t, a_k)$ 는  $t$ 번째 학습 데이터에 대한 GN의  $k$ 번째 사전확률을 의미한다. 확률적인 기울기 상승법에 의해 매 반복마다 다음과 같은 식에 의해 갱신된다.

$$a_k(i+1) = a_k(i) + \eta \sum_{t=1}^N \sum_{k=1}^K h_k^t (1 - g_k(\mathbf{x}^t, a_k)) f'(net_k^t) v_m \quad (7)$$

위와 유사한 방법으로 유사함수에 대한  $m$ 번째 입력층과  $k$ 번째 은닉층의 가중치 갱신규칙은 다음과 같다.

$$a_k^m(i+1) = a_k^m(i) + \eta \sum_{t=1}^N \sum_{k=1}^K \{h_k^t (1 - g_k(\mathbf{x}^t, a_k^m)) f'(net_k^t) a_k^m\} v'_m x_i^t \quad (8)$$

위의 가중치 갱신규칙은 MWNN의 확률적인 기울기 상승 학습[7]과 유사한 형태이지만 모든 학습 데이터에 대해서 사후확률과 오차를 고려한다는 것이 상이하다. 즉 E-단계에서 사후확률에 의해 지역오차가 고려되고, M-단계에서 가중치가 극대화된다.

## III. MWNN의 성장과 소거 알고리즘

### 3.1 MWNN의 성장 알고리즘

MWNN의 성장 알고리즘은 모듈과 모듈을

구성하는 노드를 성장시키데 직관적인 방법을 사용한다. 추가되는 EN(WNN)은 입력공간을 포괄하고 전역 오차를 감소시키게 된다. 또한 기하학적인 성장 기준을 사용하기 때문에 추가되는 EN은 준 직교성을 보장하는 프레임으로 구성되며 최적의 노드 구성이 가능하게 된다.

· 초기화 과정

초기화 과정은 WNN의 성장 알고리즘과 동일하다. MWNN의 성장알고리즘은 초기에 하나의 모듈에서 시작한다. 따라서 초기화 과정에서는 EN 모듈의 초기 크기를 결정하고 관련 매개변수를 설정하는 과정이다.

[단계 1] 초기 팽창 계수 결정

[단계 2] 노드 사이의 거리 값 초기화

[단계 3] 초기 노드의 이동계수 결정

[단계 4] 초기 가중치 결정

· 적응 과정

적응과정에서는 초기화 과정을 통해 단일 모듈을 구성하고, 대상문제에 적용하여 학습을 수행한 다음 다중 모듈을 추가하는 과정이다. WNN의 성장 알고리즘에서 제안한 방법[4]을 사용하여 전역 오차를 감소시키도록 새로운 모듈을 추가한다. 추가되는 모듈을 구성하는 WNN의 매개변수는 발생하는 지역 오차와 입력시간을 반영하여 결정한다. 전역 오차가 원하는 문턱치에 도달할 때까지 모듈의 성장을 반복한다. 수행 목표치가 원하는 문턱치( $\epsilon_{MAE}$ )에 도달하면 성장 알고리즘을 종료한다.

[단계 1] 새로운 모듈 추가(WNN의 매개변수 결정)

- 새롭게 추가되는 모듈의 매개변수 결정
- GN의 추가된 노드의 가중치 결정

[단계 2] 새롭게 추가된 모듈 학습

[단계 3] 전체 모듈 학습

3.2 MWNN의 소거 알고리즘

제안한 알고리즘은 수행능력이 부족한 모듈을 제거하는 방식을 사용하지 않고, 각 모듈의 노드를 제거하도록 설계하였다. WNN의 지역화 특성을 근거로 GN에 의해 입력시간의 분할된 상태를 고려하여 지역 오차 감소에 영향을 미치지 못하는 각 모듈의 노드를 제거한다. 제안한 MWNN의 성장 알고리즘은 새로운 모듈을 추가할 때 전체 입력시간을 고려하기 때문에 GN에 의해 부분적으로 입력시간이 분할되어

특정 입력시간에서만 각 GN이 동작한다.

[단계 1] 모듈의 불필요한 노드 제거

[단계 2] 학습단계

학습단계에서는 GN의 매개변수를 조정할 필요가 없기 때문에 고정시키고 모듈의 가중치를 조정하는 학습단계를 수행한다. 오차의 변화량이 E-6 이하일 경우 학습단계를 종료한다.

IV. 모의실험 및 검토

본 장에서 제안한 MWNN의 성장과 소거 알고리즘을 함수근사화 문제에 적용하여 수행능력을 검증하였다. WNN의 기저 함수는 Mexicanhat 웨이블릿 프레임을 사용하였고, 대상 함수는 비주기 함수와 주기 함수인 1-D 함수를 사용하였다.

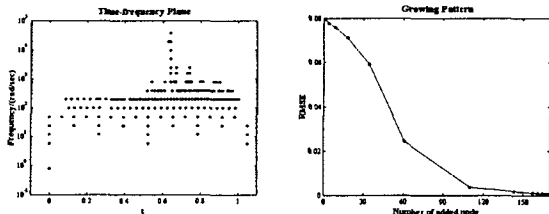
근사화 대상 함수는 식 (9)와 같은 비주기 함수이며 시간축 상에서 [0, 1]을 고려했고, 100개의 학습 데이터를 사용하였다.

$$f(t) = t \sin t \cos 5t \sin 10t \cos 30t \sin 50t \quad (9)$$

그림 1의 (a)는 최대 절대치 에러(Maximum Absolute Error; MAE)를 0.005로 설정하고 성장 알고리즘을 적용한 결과를 보인 것이다. 제시한 근사화 수행결과와 같이 MWNN의 성장 알고리즘이 성공적으로 적용되었음을 확인할 수 있다. 모의실험 결과에서 MWNN의 성장 알고리즘은 전체 입력시간을 고려하여 MAE를 근거로 모듈의 노드를 추가함을 알 수 있었다. 즉, 추가되는 모듈에 의해 일부 입력시간에서는 지역 오차를 감소시킬 수 있지만 GN에 의해 영역이 분할되어 특정 영역에서는 해당 모듈이 동작하지 못하기 때문에 지역 오차를 감소하는데 기여하지 못하게 된다. 따라서 불필요한 잉여 노드를 추가되어 WNN의 성장 알고리즘에 비해 추가되는 노드의 개수가 증가되었다. 성장과정에서 15개의 모듈이 사용되었다. 그림 1의 (b)는 MWNN의 성장 알고리즘이 적용되는 성장 패턴을 나타낸 것이다. 가로축은 추가되는 노드의 개수를 나타내고, 세로 축은 새로운 노드가 추가되기 전의 RMSE를 나타낸 것이다. 초기에 각 모듈이 추가되면서 RMSE가 급격하게 감소하고, 후반부에서는 미세조정 단계로 동작하여 RMSE의 감소량이 작다.

그림 2는 소거 알고리즘을 적용하여 해당 모듈의 노드를 제거한 후 재학습을 수행한 결과를 나타낸 것이다. 그림 1의 (a)와 비교해 불

때 특정 모듈의 노드가 삭제된 후에도 수행결과가 크게 악화되지 않았음을 확인할 수 있다. 이것은 모듈의 구성하는 WNN의 지역화 특성에 기인한 것이다.



(a) 시간-주파수 분포 (b) MWNN의 성장 패턴  
그림 1. 성장 알고리즘 적용(MAE : 0.005)

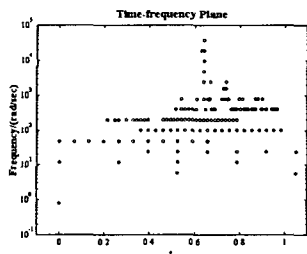


그림 2. 소거 알고리즘 적용(시간-주파수 분포)  
(MAE : 0.005)

표 1은 MAE를 0.005와 0.002로 설정하고 MWNN의 성장과 소거 알고리즘을 적용한 수치결과를 나타낸 것이다.

표 1. 비주기 함수에서 MWNN의 성장과 소거

MAE : 0.005	전체 노드의 개수	RMSE
성장 알고리즘의 적용	169	0.008364
소거 알고리즘의 적용	126	0.011672
MAE : 0.002	전체 노드의 개수	RMSE
성장 알고리즘의 적용	226	0.000368
소거 알고리즘의 적용	197	0.000648

소거 알고리즘을 적용한 후에 전체 노드의 개수가 감소되었음을 확인할 수 있다. 성장과 소거 알고리즘의 수행결과를 비교하면 소거 알고리즘을 수행한 후의 RMSE가 증가되었음을 확인할 수 있다. 이것은 삭제된 모듈의 노드가 팽창계수의 폭 만큼 주변 노드에 영향을 미치기 때문이다.

### III. 결론

제안한 MWNN의 성장과 소거 알고리즘을 함수근사화 문제에 적용하여 최적구조 설계 알고리즘의 우수성과 근사화 능력을 검증하였다. 또한 제안한 알고리즘은 단일차원의 함수근사화 문제에 성공적으로 적용될 수 있음을 검증하였다. 이러한 결과를 통해서 모듈화식 설계가 갖는 장점을 적용하고, 적합한 망의 크기를 결정할 수 있는 가능성을 확인하였다.

감사의 글 : 본 연구는 과학기술부의 뇌과학 연구 프로그램에 의해 지원 받았습니다.

### IV. 참고문헌

- [1] R. Reed, "Pruning Algorithm - A Survey," *IEEE Trans. on Neural Networks*, Vol. 6, pp. 610-622, 1995.
- [2] J. Ghosh and K. Tumer, "Structural Adaptation and Generalization in Supervised Feedforward Networks," *Journal of Artificial Neural Networks*, Vol. 1, No. 4, pp. 431-458, 1994.
- [3] J. Fristch, M. Finke, and A. Waibel, "Adaptively Growing Hierarchical Mixtures of Experts," *Advances in Neural Information Processing Systems-9*, M. I. Jordan, M. C. Mozer, and T. Petsche, Eds. Cambridge, MA: MIT Press, pp. 459-465, 1997.
- [4] 서재용, 김용택, 조현찬, 김용민, 전홍태, "F-투영법을 이용한 웨이블릿 신경망의 성장 알고리즘," 대한전자공학회 2001 하계종합학술대회 논문집, 제24권 제1호, Vol. 3, pp. 165-168, 2001.
- [5] S. R. Waterhouse and A. J. Robinson, "Constructive Algorithms for Hierarchical Mixture of Experts," *Advances in Neural Information Processing Systems-8*, D. Touretzky, M. Mozer, and M. Hasselmo, Eds. Cambridge, MA:MIT Press, pp. 584-590, 1996.
- [6] R. A. Jacobs, M. I. Jordan, s. J. Nowlan, and G. E. Hinton, "Adaptive Mixtures of Local Experts," *Neural Computation*, Vol. 6., pp. 181-214, 1994.
- [7] 서재용, 김용택, 조현찬, 전홍태, "시간-주파수 분석을 이용한 모듈라 웨이블릿 신경망의 최적 구조 설계," 대한전자공학회 논문지, 제38권, SC편, 제2호, pp. 12-19, 2001.