

# 학습률 및 모멘텀의 자동 조정 방법과 ART2를 이용한 개선된 RBF네트워크

## Enhanced RBF Network by Using Auto-Turning Method of Learning Rate, Momentum and ART2

주영호\*, 김태경\*\*, 김광백\*\*\*

Young-Ho Ju\*, Tae-Kyoung Kim\*\*, Kwang-Baek Kim\*\*\*

\*Division of Computer and Information Engineering, Silla University

\*\*Dept. of Computer and Information Engineering, Silla University

\*\*\*Dept. of Computer Engineering, Silla University

E-mail : salladinpl@hanmail.net, gbkim@silla.ac.kr

### 요 약

본 논문에서는 RBF 네트워크의 중간층과 출력층 사이의 연결강도를 효율적으로 조정하기 위해 퍼지 논리 시스템을 이용하여 학습률과 모멘텀을 동적으로 조정하는 개선된 RBF 네트워크를 제안한다. 입력층과 중간층 사이의 학습 구조로 ART2를 적용하고 중간층과 출력층 사이의 연결 강도 조정 방법으로는 제안된 학습률 자동 조정 방식을 적용한다. 제안된 방법의 학습 성능을 평가하기 위해 기존의 delta-bar-delta 알고리즘, 기존의 ART2 기반의 RBF 네트워크와 비교 분석한 결과, 제안된 방법이 학습 속도와 수렴성에서 개선된 것을 확인하였다.

### 1. 서 론

오늘날 신경망 학습 알고리즘의 학습 시간과 일반화 능력을 향상시키기 위한 연구가 진행 되어왔다. 그 결과, 통계학의 다변량 분석(Multivariate Analysis) 및 보간(interpolation)문제 해결에 이용되었던 RBF(Radial Basis Function)를 Brodmann과 Low의 신경망 모델을 구성하는데 이용함으로써 RBF 네트워크가 제안되었다[1]. RBF 네트워크는 빠른 학습 시간, 일반화, 단순화의 특징으로 학습 데이터를 분류하는 작업과 비선형 시스템 모델링 등에 적용되고 있다. RBF 네트워크는 입력, 중간, 출력 등의 3개의 층으로 구성된 전방향 신경망(feedforward neural network)이다. 층간에 각기 다른 일을 수행하기 때문에 각 층간에 서로 다른 알고리즘을 적용할 수 있으며 층간의 최적화(optimization)를 분리하여 구성할 수 있다[2]. 층간의 구성은 크게 3가지의 분류로 나눌 수 있다.

첫째로, Fixed Centers Selected at Random으로 중간층의 노드가 학습 데이터 집합으로부터 임의적으로 선택되는 방식과 둘째, Self-Organized Selection of Centers으로 자기 조직화 형식에 따라 중간층을 결정하고, 출력층에서는 지도(supervised) 학습을 적용하는 방식이다. 그리고 마지막으로 Supervised Selection of Centers로 중간층과 출력층을 지도 학습으로 학습을 시키는 방식으로 구분할 수 있다. 따라서 RBF 네트워크의 중간층은 클러스터링하는 층이다. 즉, 이 층의 목적은 주어진 자료 집합을 유사한 클러스터들(homogeneous clusters)로 분류하는 것이다. 여기서 유사하다는 것은 입력 데이터들에 대한 특징 벡터 공간사이에서 한 클러스터내의 벡터들간에 거리를 측정하여 정해진 반경내에 존재하면 같은 클러스터로 분류하고 정해진 반경내에 존재하지 않으면 다른 클러스터로 분류한다. 그러나 정해진 반경 내에서 클러스터링하는 것은 잘못된 클러스터를 선택하는 단점을 가지게 된다. 그러므로 중간층을 결정하는 것은 RBF 네트워크의 전반적인 효율성에 큰 영향을 준다[3]. 따라서 본 논문에서는 중간층을 효율적으로 구성하기 위해 ART2

\* 신라대학교 컴퓨터정보공학과

\*\* 신라대학교 컴퓨터정보공학과부

\*\*\* 신라대학교 컴퓨터공학과

를 적용하고 중간층과 출력층 사이의 연결 강도 조정에 퍼지 논리 시스템을 이용한 학습률과 모멘텀 자동 조정 방법을 적용하는 개선된 RBF 네트워크를 제안한다.

## 2. 관련 연구

### 2.1 delta-bar-delta 알고리즘

역전파 알고리즘의 성능을 개선한 delta-bar-delta 알고리즘[4]은 각각의 연결강도에 대해 학습률을 동적으로 조정하여 학습성능을 향상시킨 알고리즘이다. 학습률을 동적으로 조정하기 위해 delta와 delta-bar를 형성한다.

delta를 형성하는 식은 다음과 같다. 여기서 i는 입력층, j는 중간층, k는 출력층이다.

$$\Delta_{ji} = \frac{\sigma E}{\sigma w_{ji}} = -\delta_j x_i \quad (1)$$

$$\Delta_{kj} = \frac{\sigma E}{\sigma w_{kj}} = -\delta_k z_j \quad (2)$$

delta-bar를 형성하는 식은 다음과 같다.

$$\bar{\Delta}_{ji}(t) = (1 - \beta) \Delta_{ji}(t) + \beta \bar{\Delta}_{ji}(t-1) \quad (3)$$

$$\bar{\Delta}_{kj}(t) = (1 - \beta) \Delta_{kj}(t) + \beta \bar{\Delta}_{kj}(t-1) \quad (4)$$

식 (4)에서 파라미터  $\beta$  값은 0과 1사이의 값으로 고정된 상수이다. delta와 delta-bar 변화에 대한 학습률의 변화는 다음과 같다. 만약 연결강도의 변화가 연속된 학습과정에서 같은 방향으로 이루어진다면 학습률은 증가한다. 이것은 delta와 delta-bar가 같은 부호에서 발생하며, 이 때 학습률은 증가한다. 또한 만약 delta와 delta-bar의 부호가 반대 부호를 가지면 학습률은 현재값의  $1 - \gamma$ 의 비율로 감소한다. 이러한 가변 학습률은 각 층에 대하여 다음과 같은 식으로 표현된다.

$$\begin{aligned} \alpha_{ji}(t+1) &= \alpha_{ji}(t) + k && \text{if } \bar{\Delta}_{ji}(t-1) \cdot \Delta_{ji}(t) > 0 \\ &= (1 - \gamma)\alpha_{ji}(t) && \text{if } \bar{\Delta}_{ji}(t-1) \cdot \Delta_{ji}(t) < 0 \\ &= \alpha_{ji}(t) && \text{if } \bar{\Delta}_{ji}(t-1) \cdot \Delta_{ji}(t) = 0 \end{aligned} \quad (5)$$

$$\begin{aligned} \alpha_{kj}(t+1) &= \alpha_{kj}(t) + k && \text{if } \bar{\Delta}_{kj}(t-1) \cdot \Delta_{kj}(t) > 0 \\ &= (1 - \gamma)\alpha_{kj}(t) && \text{if } \bar{\Delta}_{kj}(t-1) \cdot \Delta_{kj}(t) < 0 \\ &= \alpha_{kj}(t) && \text{if } \bar{\Delta}_{kj}(t-1) \cdot \Delta_{kj}(t) = 0 \end{aligned} \quad (6)$$

### 2.2 ART2기반 RBF 네트워크

ART2 알고리즘을 기반으로 한 RBF 네트워크의 학습은 크게 두 단계로 구분된다. 첫 번째 단계에서는 입력층과 중간층 사이의 학습 구조로 경쟁 학습이 적용되고, 중간층과 출력층 사이에는 지도 학습으로 수행된다[5]. ART2 기반 RBF 네트워크에서 중간층

의 출력 벡터는 식(7)과 같이 계산하고, 식(8)와 같이 가장 작은 출력 벡터를 가진 노드가 승자 노드가 된다.

$$O_j = \sum_{i=1}^N (|x_i - w_{ji}(t)|) \quad (7)$$

$$O_j^* = \text{Min}\{O_j\} \quad (8)$$

여기서  $w_{ji}(t)$ 는 입력층과 중간층 사이의 연결 가중치이다.

ART2 기반 RBF 네트워크에서 중간층의 승자 노드는 입력 벡터와 중간층의 출력 벡터의 차이가 최소인 값을 계산하여 승자노드로 선정하고 선정된 승자노드에 대한 유사성의 검증은 식(9)와 같다.

$$O_j^* < \rho \quad (9)$$

여기서  $\rho$ 는 경계 변수이고, 승자노드의 출력 벡터가 경계 변수보다 적으면 같은 패턴으로 분류하고 경계 변수보다 크면 다른 패턴으로 분류한다. 같은 패턴으로 분류되면, 입력 패턴의 유사한 특징을 연결 강도에 반영하기 위해 연결 강도를 조정한다. ART2 알고리즘의 연결강도의 조정은 다음과 같다.

$$w_{ji}(t+1) = \frac{w_{ji}(t) \times u_n + x_i}{u_n + 1} \quad (10)$$

여기서  $u_n$ 은 생성된 클러스터에 갱신된 패턴들의 개수이다. 중간층의 출력벡터는 식 (11)에 의해 정규화가 되고 출력층의 입력 벡터로 적용된다.

$$z_j = 1 - \frac{O_j}{N} \quad (11)$$

출력층의 출력벡터는 식(12)와 같이 계산한다.

$$O_k = f\left(\sum_{j=1}^M w_{kj} \times z_j\right) \quad (12)$$

$$f(x) = \frac{1}{1 + e^{-x}} \quad (13)$$

출력 벡터는 목표 벡터와 비교하여 오차 값과 오차 신호를 구하고 연결강도를 조정한다.

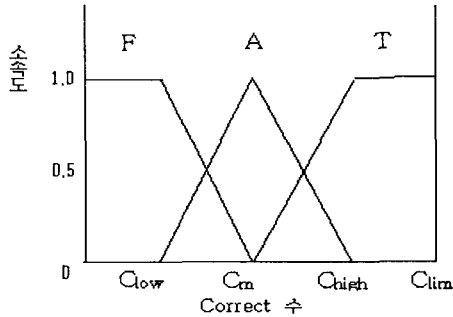
$$\delta_k = (T_k - O_k) O_k (1 - O_k) \quad (14)$$

$$w_{kj}(t+1) = w_{kj}(t) + \alpha \delta_k z_j \quad (15)$$

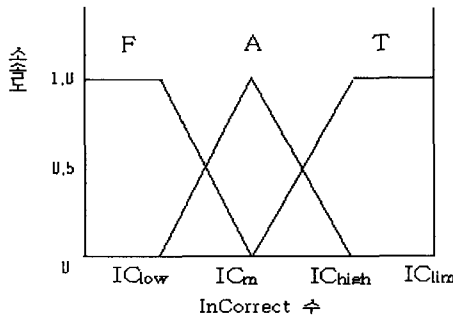
## 3. 개선된 RBF 네트워크

개선된 RBF 네트워크는 입력층과 중간층 사이의 학습 구조로 ART2를 적용하고 중간층과 출력층 사이의 연결 강도 조정 방법으로는 제안된 학습률 자동 조정 방식을 적용한다. 각 패턴에 대해서 출력 벡터와 목표 벡터간의 차이의 절대값이 0.1 이하이면

정확성으로 분류하고, 그렇지 않을 때는 부정확성으로 분류한다. 그리고 정확성과 부정확성의 수를 퍼지 논리 시스템의 입력으로 적용하여 학습률과 모멘텀을 동적으로 조정한다. 정확성의 소속 함수는 그림 1과 같고 부정확성의 소속 함수는 그림 2와 같다.



[그림 1] 정확성에 대한 소속함수



[그림 2] 부정확성에 대한 소속함수

여기서  $C_{low}$  값과  $C_{high}$  값은 식 (14), (15)와 같이 계산한다.

$$C_{low} = \log_2(\text{입력 노드수} + \text{패턴수}) \quad (14)$$

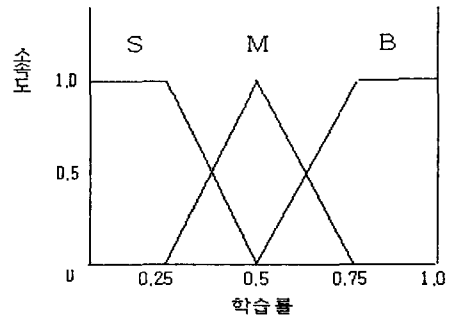
$$C_{high} = C_{lim} - C_{low} \quad (15)$$

그림 2와 3에서 F는 false, A는 average, T는 true를 의미하는 소속함수이다. 학습률을 조정하기 위한 퍼지 제어 규칙을 if ~ then 형식의 퍼지 관계로 표현하면 다음과 같다.

- $R_1 : \text{If correct is F, incorrect F Then } \alpha \text{ is B}$
- $R_2 : \text{If correct is F, incorrect A Then } \alpha \text{ is B}$
- $R_3 : \text{If correct is F, incorrect T Then } \alpha \text{ is B}$
- $R_4 : \text{If correct is A, incorrect F Then } \alpha \text{ is M}$
- $R_5 : \text{If correct is A, incorrect A Then } \alpha \text{ is M}$
- $R_6 : \text{If correct is A, incorrect T Then } \alpha \text{ is M}$
- $R_7 : \text{If correct is T, incorrect F Then } \alpha \text{ is S}$
- $R_8 : \text{If correct is T, incorrect A Then } \alpha \text{ is S}$
- $R_9 : \text{If correct is T, incorrect T Then } \alpha \text{ is S}$

학습에 적용될 학습률을 계산하는 출력 소속함수는

그림 3과 같다.



[그림 3] 학습률에 대한 소속함수

그림 3에서 S는 small, M은 medium, B는 big을 의미하는 소속함수이다. 정확성과 부정확성이 퍼지 논리 시스템의 입력값으로 정해지면 각 소속 함수에 대한 정확성과 부정확성의 소속도를 구한다. 각 소속 함수에 대하여 소속도가 구해지면 퍼지 제어 규칙을 적용하고 Max\_Min 방법으로 추론한다. 퍼지 추론 후에는 비퍼지화 하여 학습에 사용될 학습률을 구한다. 비퍼지화는 무게 중심법을 사용하는데, 식(16)와 같다[10].

$$\alpha = \frac{\sum \mu(y)y}{\sum y} \quad (16)$$

모멘텀은 식 (17)와 같이 계산한다.

$$\mu = \zeta - \alpha \quad (17)$$

여기서  $\zeta$ 는 경험적으로 주어지는 1과 1.5사이의 파라미터이다.

#### 4. 실험 및 결과 분석

실험환경은 Intel Pentium-IV CPU와 256MB RAM이 장착된 IBM 호환 PC상에서 C++ Builder 6.0으로 구현하였다.

주민등록증 영상에서 추출한 10×10크기를 가진 136개의 숫자 패턴을 기존의 delta-bar-delta 방법, ART2 기반 RBF 네트워크 그리고 본 논문에서 제안한 학습 알고리즘에 적용하여 학습 반복 횟수와 수렴성을 비교 및 분석하였다. 학습에 사용된 숫자패턴은 그림 4와 같고, 목표 벡터는 표 1과 같다.



[그림 4] 숫자 패턴

[표 1] 학습에 사용된 목표 벡터

	0	1	2	3	4	5	6	7	8	9
목 표 값	0	0	0	0	0	0	0	0	1	1
	0	0	0	0	1	1	1	1	0	0
	0	0	1	1	0	0	1	1	0	0
	0	1	0	1	0	1	0	1	0	1

표 2는 실험에 사용된 각 알고리즘의 파라미터이고 표 3은 학습 결과이다.

[표 2] 학습에 사용된 파라미터

학습방법	파라미터	$\alpha$	$\rho$	$\xi$	$\beta$	$\kappa$	$\gamma$
delta-bar-delta		0.5			0.7	0.005	0.2
ART2기반 RBF네트워크		0.5	10				
제안된 방법			10	1.0			

표 2에서  $\alpha$ 는 학습률,  $\rho$ 는 ART2의 경계변수,  $\xi$ 는 모멘텀을 구하기 위한 파라미터,  $\beta, \kappa, \gamma$ 는 delta-bar-delta 알고리즘에서 설정되는 파라미터이다.

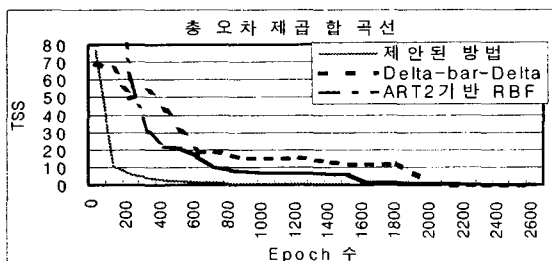
학습 판정은 10000번의 epoch의 수행에서 모든 패턴들의 실제 출력 벡터와 목표 벡터의 차이의 절대값이  $\epsilon(\epsilon \leq 0.1)$ 보다 적거나 같으면 정확성으로 분류되는 것으로 하여 10회의 실험을 수행하였다.

[표 3] 각 알고리즘간의 수렴성 비교

학습방법	학습결과	실험 횟수	성공 횟수	평균 Epoch 수
delta-bar-delta 방법		10	2	2793
ART2 기반 RBF 네트워크		10	10	2710
제안된 방법		10	10	1464

표 3에서와 같이 기존의 방법들보다 제안된 방법이 학습 속도나 수렴성에서 개선된 것을 확인할 수 있다. 그리고 기존의 방법들은 실험 횟수에 따라 학습 횟수와 수렴성에 민감하게 반응하였으나 제안된 방법에서는 학습 횟수와 수렴성에 민감하게 반응하지 않아서 학습 성능이 개선되었다.

그림 5는 epoch수에 따른 총 오차 제곱합의 변화율을 그래프로 나타내었다. 총 오차 제곱합 곡선에서도 제안된 방법이 기존의 방법들 보다 초기의 수렴 속도가 빠르고 총 오차 제곱합도 적은 것을 확인할 수 있다.



[그림 5] 총 오차 제곱합 곡선

## 5. 결론 및 향후 연구 과제

ART2 알고리즘을 기반으로 한 RBF 네트워크의 학습은 크게 두 단계로 구분된다. 첫 번째 단계는 입력층과 은닉층 사이에 경쟁 학습이 수행되고, 두 번째 단계는 은닉층과 출력층 사이에 지도학습이 수행된다. 본 논문에서는 기존의 ART2 기반 RBF 네트워크의 학습 성능을 개선하기 위하여 입력층과 중간층에 ART2 알고리즘을 적용하고 중간층과 출력층 사이의 효율적인 가중치 조절을 위해 퍼지 논리 시스템 이용하여 학습률과 모멘텀을 자동으로 조정하는 방법을 적용하는 개선된 RBF 네트워크를 제안하였다. 제안된 학습률 및 모멘텀을 동적으로 조정하는 방법은 출력 벡터와 목표 벡터의 차이의 절대값이  $\epsilon$  보다 적거나 같으면 정확성으로 분류하고, 그렇지 않은 경우에는 부정확성으로 분류하여 정확성과 부정확성의 수를 퍼지 논리 시스템에 적용하여 학습률을 동적으로 조정하였고 동적으로 조정된 학습률을 이용하여 모멘텀을 조정하여 학습 성능을 개선하였다.

제안된 방법의 학습 성능을 평가하기 위하여 주민 등록증 영상에서 추출한 숫자 패턴 분류 문제에 적용한 결과, 기존의 학습 방법들보다 학습 횟수와 수렴성에 민감하게 반응하지 않았고, 총 오차 제곱합도 기존의 학습 방법에 비해 현저히 감소하여 학습 성능이 개선된 것을 확인하였다.

향후 연구 과제는 ART2 알고리즘의 성능을 개선시켜 최적의 중간층 노드를 생성하는 방법에 대해 연구 할 것이다.

## 6. 참고 문헌

- [1] M. Watanabe, K. Kuwata and R. Katayma, "Adaptive Tree-Structured Self Generating Radial Basis Function and its Application to Nonlinear Identification Problem," Proceedings of IIZUKA, pp.167-170, 1994.
- [2] 황영섭, 방승양, "RBF 신경망 분류기의 효율적 구성 방법," 한국정보과학회논문지, Vol. 21, No. 5, pp.451-460, 1997.
- [3] C. Panchapakesan, D. Ralph and M. Palaniswami, "Effects of Moving the Centers in an RBF Network," Proceedings of IJCNN, Vol. 2, pp.1256-1260, 1998.
- [4] R. A. Jacobs, "Increased rates of convergence through learning rate adaptation," IEEE Transactions on Neural Networks, Vol. 1, No. 4, pp. 295-308, 1988.
- [5] 류재욱, 김태경, 김광백, "개선된 RBF 신경망을 이용한 여권 인식", 한국지능정보시스템학회 발표논문집, pp.529-534, 2002.
- [6] M. Jamshidi, N. Vadiiee, T. J. Ross, Fuzzy Logic and Control, Prentice-Hall, 1993.