

개선된 퍼지 단층 퍼셉트론

이재언^{*} · 허주용^{*} · 김광백^{**}

^{*}신라대학교 컴퓨터정보공학부

^{**}신라대학교 컴퓨터공학과

Enhanced Fuzzy Single Layer Perceptron

Jae-Eon Lee^{*} · Joo-Yong Her^{*} · Kwang-Baek Kim^{**}

^{*}Division of Computer Information and Engineering, Silla University

^{**}Dept. of Computer Engineering, Silla University

E-mail : ljapc@naver.com, wndyd2000@hanmail.net, gbkim@silla.ac.kr

요 약

기존의 단층 퍼셉트론은 출력 노드가 선형 분리 가능한 패턴들만을 분류할 수 있고 Exclusive OR와 같은 비선형 문제에 대해서는 분류할 수 없는 단점이 있다. 그러나 퍼지 단층 퍼셉트론은 퍼지 소속 함수(fuzzy membership function)를 적용하여 단층 구조로 Exclusive OR 문제와 같은 고전적인 문제를 개선하였다. 그러나 퍼지 단층 퍼셉트론은 기존의 단층 퍼셉트론과 마찬가지로 결정 경계선이 진동하는 경우가 생기며 초기 가중치의 범위와 학습률에 따라 수렴성이 매우 낮아지는 단점이 있다. 따라서 본 논문에서는 바이어스항을 도입하여 결정 경계선이 진동하는 것을 방지하여 수렴성을 개선시키고 선형 활성화 함수를 제안하고 학습률과 모멘텀 개념을 도입하여 학습 시간을 단축시키는 개선된 퍼지 단층 퍼셉트론 알고리즘을 제안한다.

제안된 방법과 퍼지 단층 퍼셉트론간의 학습 성능을 분석하기 위하여 인공 신경망에서 벤치마크로 사용되는 exclusive OR 문제와 문자 패턴 분류에 적용하여 epoch 수와 수렴성을 비교한 결과, 제안된 방법이 기존의 퍼지 단층 퍼셉트론보다 학습 시간이 적게 소요되고 수렴성이 개선된 것을 확인하였다.

키워드

Exclusive OR, 퍼지 단층 퍼셉트론, 퍼지 소속 함수, 바이어스항, 학습률, 모멘텀

1. 서 론

인간의 뇌를 모방하는 접근 방법에는 두 가지 접근 방법이 있다. 인간의 뇌에 들어 있는 지식을 그 기능에 의거하여 프로그램화하려는 기호 처리 인공지능(symbolic processing artificial intelligence)과 뇌를 분석함으로써 얻은 뇌의 구조를 이용하여 지능을 구현하려는 인공지능으로 나눌 수 있다. 이 방법들 중에서 기호 처리 인공지능은 기능적 수준에서의 접근 방법인데, 이 접근 방법을 퍼지 논리(Fuzzy Logic)라고 한다. 그리고 생물학적 수준에서의 접근 방법을 인공신경망(Artificial Neural Network)이라고 한다[1,2].

본 논문에서는 퍼지 논리와 인공 신경망을 융

합한 기존의 퍼지 단층 퍼셉트론 알고리즘을 개선한다. 인공신경망의 지도 학습 모델인 단층 퍼셉트론(single layer perceptron)은 1950년대에 처음 등장할 당시 간단한 패턴을 스스로 학습할 수 있다는 능력 때문에 많은 사람들이 관심을 가졌다. 그러나 기존의 단층 퍼셉트론 경우는 입력 패턴에 대한 분류 결정 경계선이 초기하평면(hyperplane)으로만 구성되어 있지 않을 경우에는 부적합하다. 그 대표적인 예가 Exclusive OR 문제인데 여기서는 초기하 평면만으로는 입력 패턴들을 두 부류로 분리하는 것이 불가능하므로 단층구조를 가진 퍼셉트론은 이 문제를 해결할 수 없다[3]. 그러나 Wang이 제안한 퍼지 단층 퍼셉트론은 퍼지 소속 함수(fuzzy membership

function)를 적용하여 어느 정도의 안정성을 보장하여 단층 구조로 Exclusive OR 문제와 같은 고전적인 문제를 개선하였다[4]. 그러나 퍼지 단층 퍼셉트론은 기존의 단층 퍼셉트론과 마찬가지로 결정 경계선이 진동하는 경우가 생기며 초기 가중치의 범위와 학습률에 따라 수렴성이 낮아진다. 그러므로 문자 인식과 같은 분야에 적용할 경우에는 계산량이 증가하게 되고 수렴성이 낮아진다. 기존의 단층 퍼셉트론과 퍼지 단층 퍼셉트론은 역치값을 고정시켜서 각 패턴을 학습시키므로 부적절한 역치에 의해서 각 패턴들의 결정 경계선이 부정확하게 분류되기 때문에 수렴성이 낮아지는 원인이 될 수 있다. 그러므로 본 논문에서는 바이어스항을 도입하여 결정 경계선이 진동하는 것을 방지하여 수렴성을 개선시키고 선형 활성화 함수를 제안하고 학습률과 모멘트 개념을 도입하여 학습 시간을 단축시키는 개선된 퍼지 단층 퍼셉트론 알고리즘을 제안한다.

II. 관련 연구

기존의 퍼지 단층 퍼셉트론 알고리즘으로는 Wang이 제안한 퍼지 단층 퍼셉트론과 퍼지 논리 연산자를 이용한 퍼지 단층 퍼셉트론이 있다.

2.1 Max_Min을 이용한 퍼지 단층 퍼셉트론

Max_Min을 이용한 퍼지 단층 퍼셉트론은 신경망과 퍼지 이론의 유사한 특징을 중첩시킨 학습 알고리즘이다. 퍼지 이론과 인공 신경망의 외형적인 유사성은 다음과 같다. 첫째, 퍼지 추론 규칙의 조건부(IF부분)의 퍼지 변수와 입력간의 Min연산은, 뉴런의 입력과 연결선에 부여된 가중치의 곱셈 연산에 대응된다. 둘째, 여러 추론 규칙의 결론부분(THEN 부분)에 대한 Max연산은 뉴런의 입력과 가중치를 곱한 것을 합하는 연산과 대응된다[5]. 따라서 Max_Min을 이용한 퍼지 단층 퍼셉트론은 퍼지 논리 연산자를 적용하며 입력 패턴을 0과 1사의 값으로 구성하여 학습하는 알고리즘이다. 이 알고리즘의 장점은 기존의 단층 알고리즘이 이진 데이터를 학습하는데 중점을 둔 반면에 퍼지 학습 0과 1사이의 값을 학습하는데 효과적으로 학습 할 수 있도록 퍼지 논리 연산자를 이용한 것이 큰 장점이다. 그러나 이 알고리즘은 패턴 분류 문제에 있어서 패턴의 수가 많아지는 경우에는 수렴성이 떨어지는 단점이 있고 n bit parity와 같은 복잡한 비선형 문제에 대해서는 학습이 되지 않는 경우가 발생한다. Max_Min을 이용한 퍼지 단층 퍼셉트론은 다음과 같다.

I_1, I_2, \dots, I_m 은 입력 노드이고 O_1, \dots, O_n 은

출력 노드이다. $t_k (1 \leq k \leq K)$ 은 목표 값을 나타낸다.

단계 1 : 입력값 $v_k = (v_{k1}, \dots, v_{km})$ 과 목표값 $t_k = (t_{k1}, \dots, t_{km})$ 을 제시한다.

단계 2 : $o_k = (o_{k1}, \dots, o_{kn})$ 은 출력값을 나타내고 r_{ij} 은 가중치로서 임의의 값으로 초기화한다. θ_i 은 바이어스항으로서 임의의 값으로 초기화한다. w_{ki} 은 이 학습 알고리즘의 실제적인 초기 가중치이고 계산하는 식은 다음과 같다.

$$w_{ki} = \bigvee \{ \bigwedge \{ v_{ki}, r_{1i} \}, \dots, \bigwedge \{ v_{km}, r_{mi} \} \}$$

여기서 \bigvee 은 퍼지 논리 합 연산자이고 \bigwedge 은 퍼지 논리 곱 연산자이다.

$$o_{ki} = w_{ki} \bigvee \theta_i$$

단계 3 : 목표값과 출력값 사이의 오류 값을 계산한다. 여기서 Q은 Total Sum of Square이다

$$Q = \frac{1}{2} \sum_{k=1}^K \sum_{i=1}^n (o_{ki} - t_{ki})^2$$

단계 4 : 가중치와 바이어스항을 조정한다.

$$r_{ij}(l+1) = r_{ij}(l) + n_1 \Delta r_{ij}(l+1) + a_1 \Delta r_{ij}(l),$$

$$\theta_i(l+1) = \theta_i(l) + n_2 \Delta \theta_i(l+1) + a_2 \Delta \theta_i(l),$$

$$\Delta r_{ij} = 1, \text{ when } o_{ki} = r_{ij} \\ = 0, \text{ otherwise.}$$

$$\Delta \theta_i = 1, \text{ if } o_{ki} = \theta_i \\ = 0, \text{ otherwise.}$$

단계 5: 오류 한계 값을 만족할 때 까지 단계 2로 가서 반복한다.

2.2 퍼지 단층 퍼셉트론

일반화된 델타 규칙에 기반을 둔 알고리즘을 사용한 퍼지 단층 퍼셉트론은 Wang 등이 제안하였다[6]. 그러나 이 알고리즘은 데이터의 적용에 대해 어느 정도의 수렴성을 보장하지만 단층 퍼셉트론과 마찬가지로 결정 경계선이 진동하는 경우가 생기며 초기 가중치의 범위와 학습률에 따라서 수렴성이 낮아진다. Wang이 제안한 퍼지 퍼셉트론 알고리즘은 다음과 같다.

단계 1: 가중치의 초기화 및 입력 값과 목표 값을

제시한다.

가중치 : $w_i(i=1, \dots, n)$,

입력값 : $x_i(i=1, \dots, n)$

단계 2: 입력된 x_i 을 소속도 m_i 의 오름차순으로 정렬하고 정렬된 차이를 구한다.

$$0 = m_{j_0} \leq m_{j_1} \leq \dots \leq m_{j_n} \leq 1$$

$$p_i = m_{j_i} - m_{j_{i-1}}$$

단계 3: 정렬된 노드의 소속도에 대응하는 b_i 값을 계산한다.

$$b_i = \sum_{j=1}^n w_j$$

단계 4: b_i 값을 정규화한다.

$$F_i = F(b_i)$$

$$F(b_i) = 1 \quad b_i \leq 1$$

$$F(b_i) = 0 \quad b_i \leq (1 - \frac{1}{n})$$

$$F(b_i) = n \times (b_i + \frac{1}{n-1}) \text{ otherwise}$$

단계 5: 출력값과 목표값의 차이인 오류를 계산하여 가중치를 조정한다.

$$\Delta w_i = error \times p_i \times F(b_i)$$

$$w_i(n+1) = a \times w_i(n) + \Delta w_i$$

단계 6: 오류 한계를 만족할 때까지 단계 3으로 가서 반복 한다.

성화 함수를 제안하기 위하여 오류 역전파 알고리즘에서 델타를 구하는 식을 살펴보면 다음과 같다.

$$\delta = error \times f(net) \times (1-f(net))$$

여기서 사용하는 $f(net) \times (1-f(net))$ 은 활성화 함수를 미분한 값이다. $f(net)$ 이 취할 수 있는 값의 범위는 0과 1사이의 실수이다. $f(net)$ 과 $f(net) \times (1-f(net))$ 의 관계를 살펴보면 $f(net)$ 의 값이 0.0과 0.25미만에서는 $f(net) \times (1-f(net))$ 의 값이 $f(net)$ 의 값과 유사하며, 0.25와 0.70 사이에서는 0.25에 근사하고, 0.70과 1.0사이에서는 $(1-f(net))$ 의 값과 유사하다 이러한 분석에 의하여 식 (1)과 같이 델타를 구하는 식으로 수정할 수 있다.

$$\delta = error \times f(net) \quad \text{where } 0.0 \leq f(net) < 0.25$$

$$\delta = error \times 0.25 \quad \text{where } 0.25 \leq f(net) \leq 0.70$$

$$\delta = error \times (1 - f(net)) \quad (1)$$

식 (1)에서 델타를 구하는데 있어 활성화 함수의 이차 미분값을 사용하였던 과정을 생략할 수 있게 해준다. 그리고 식 (1)을 사용하면 굳이 활성화 함수가 미분 가능하고 연속이어야 한다는 조건이 필요 없다. 그러므로 제안된 선형 활성화 함수의 식은 다음과 같다.

$$f(net) = 1.0 \quad \text{where } net > 5.0$$

$$f(net) = \lambda \times net + 0.5$$

$$\text{where } -5.0 \leq net \leq 5.0, 0 < \lambda < 0.5$$

$$f(net) = 0.0$$

$$\text{where } net < -0.5$$

이러한 선형 활성화 함수의 그래프는 다음과 같다.

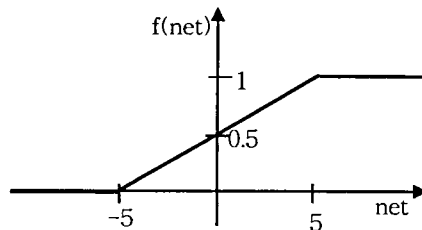


그림 1. 제안된 선형 활성화 함수

제안된 선형 활성화 함수는 다음과 같이 일반화하여 나타낼 수 있다.

$$f(net) = (\frac{1}{\mu_{range} \times 2}) \times net + 0.5 \quad (2)$$

III. 개선된 퍼지 단층 퍼셉트론

기존의 단층 퍼셉트론과 퍼지 단층 퍼셉트론은 역치값을 고정시켜서 각 패턴을 학습시키므로 부적절한 역치에 의해서 각 패턴들의 결정 경계선이 부정확하게 분류되기 때문에 수렴성이 낮아지는 원인이 된다. 그러므로 본 논문에서는 바이어스항을 도입하여 결정 경계선이 진동하는 것을 방지하여 수렴성을 개선시키고 학습률과 모멘텀 개념을 도입하여 학습 시간을 단축시키는 개선된 퍼지 단층 퍼셉트론을 제안한다.

신경망에서 일반적으로 사용하는 활성화 함수는 다음의 두 가지 조건을 만족해야 한다. 첫째는 연속이어야 하며, 둘째는 미분이 가능해야 한다. 신경망에서 일반적으로 활성화 함수를 시그모이드 함수(sigmoidal function)를 사용하고 있다. 시그모이드 함수를 사용할 경우 net값이 거의 0에 가까운 경우에는 입력 및 출력 곡선의 기울기가 급하여 높은 gain을 제공하며, net값이 커질수록 gain은 작은 값을 제공하는 이점도 있다. 선형 활

식 (2)에서 μ_{range} 는 $f(\text{net})$ 이 0과 1의 값을 갖는 범위를 제외한 단조 증가하는 구간을 의미한다. 제안된 퍼지 단층 퍼셉트론 알고리즘에서 활성화 함수는 식 (2)를 사용한다.

제안된 퍼지 단층 퍼셉트론에서 전방향 단계인 출력값을 계산하는 과정은 입력값을 퍼지 퍼셉트론과 같이 정렬하여 정렬된 입력과 가중치에 의해서 계산한다. 그리고 제안된 선형 활성화 함수를 사용하여 수렴성을 향상시키고 부적절한 역치에 의해서 각 패턴들의 결정 경계선이 진동하는 것을 방지하기 위해 바이어스항을 추가하고 학습 속도가 느려지는 경우를 방지하기 위해서 학습률을 오류값에 따라 동적으로 변화시킨다. 제안된 퍼지 단층 학습알고리즘의 학습 모형은 그림 2로 나타낼 수 있다.

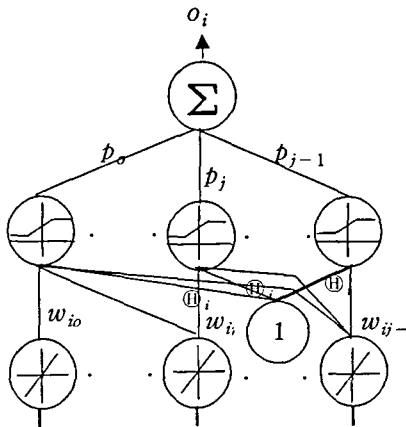


그림 2. 개선된 퍼지 단층 퍼셉트론 학습 모형

본 논문에서 제안된 퍼지 단층 학습 알고리즘은 다음과 같다.

입력 : $x_i = (x_0, \dots, x_{n-1})$
 출력 : $o_j = (o_0, \dots, o_{m-1})$

단계 1: 가중치와 바이어스항을 초기화 한다.

$w_{ij}(0 \leq i \leq I-1)$ 는 입력층 i 로부터 출력층 j 로의 연결 가중치이고 θ_j 은 출력층 노드의 바이어스항(bias term)이다.

단계 2: 입력값 (x_i)과 목표값 (t_j)을 제시한다.

단계 3: 입력값 (x_i)을 귀속도의 오름차순으로 정렬하고 정렬된 차이를 구한다.

$$0.0 = m_0 \leq m_1 \leq \dots \leq m_{I-2} \leq m_{I-1} \leq 1$$

$$p_k = m_i - m_{i-1} \text{ where } 0 \leq k \leq n-1$$

단계 4: 출력층 노드의 출력(o_j)값을 계산한다.

여기서 $f(\sum_{i=k}^{I-1} w_{ij} + \theta_j)$ 은 제안된 선형활성화함수이다.

$$o_j = (\frac{1}{\mu_{range} \times 2}) \times f(\sum_{i=k}^{I-1} w_{ij} + \theta_j) + 0.5$$

단계 5: 출력값과 목표값의 차이인 오류값을 계산하여 가중치를 조정하고 학습률(α_j)을 동적으로 변화시킨다.

$$e_j = t_j - o_j$$

$$\Delta w_{ij} = \alpha_j \times e_j \times \sum_{k=0}^m p_k \times f(\sum_{i=k}^{I-1} w_{ij} + \theta_j) + \beta \times \Delta w_{ij}(t)$$

$$w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij}$$

$$\Delta \theta_j = \alpha_j \times e_j \times f(\theta_j) + \beta \times \Delta \theta_j(t)$$

$$\theta_j(t+1) = \theta_j(t) + \Delta \theta_j$$

$$\text{if } |t_j - o_j| < \epsilon \text{ then } \Delta \alpha_j(t+1) = e^2$$

$$\alpha_j(t+1) = \alpha_j(t) + \Delta \alpha_j$$

여기서 α_j 는 학습률이고 β 는 모멘텀이다

단계 6: 총 오차 자승합이 오류 한계 크면 단계3으로 가고 오류 한계보다 적거나 같으면 학습을 종료한다.

IV. 실험 및 결과 분석

실험은 IBM 호환 기종의 펜티엄 PC 상에서 VC로 구현하였다. 제안된 방법과 퍼지 단층 퍼셉트론 및 Max_Min 기반 퍼지 단층 퍼셉트론간의 학습 성능을 분석하기 위하여 인공 신경망에서 벤치마크로 사용되는 exclusive OR 문제와 한글 자음 패턴 분류에 적용하여 epoch 수와 수렴성을 비교하였다. 제안된 방법에서 학습률을 동적으로 조정하기 위해 α 을 0.5로 설정하였다. 인공 신경망에서 exclusive OR 문제가 벤치마크로 사용되는 이유는 비선형 문제를 해결하는 기본 문제로 사용되기 때문에 기존의 퍼지 단층 퍼셉트론과 제안된 방법간의 epoch 수를 표 1로 나타내었다. 표 1에서 알 수 있듯이 제안된 방법이 기존의 퍼지 단층 퍼셉트론보다 학습 시간이 적게 소요되었고 학습률에 따라 민감하게 반응하지 않으므로 수렴성이 개선되었다. 즉 제안된 학습 방법은 학습률의 설정에 따라 학습의 수렴성이 민감하지 않았다. 따라서 제안된 학습 방법 학습 성능이 개선된 것을 확인할 수 있다.

표 1. XOR 문제에서 Epoch수 비교

XOR	학습률	Epoch 수
퍼지단층 퍼셉트론	0.2	10
	0.5	9
	0.7	7
Max_Min 퍼지 단층 퍼셉트론	0.2	40
	0.5	40
	0.7	29
제안된방법	0.2	7
	0.5	6
	0.7	6

기존의 학습 방법들과 제안된 방법간의 epoch 수에 따른 TSS 변화 과정을 그림 3으로 나타내었다.

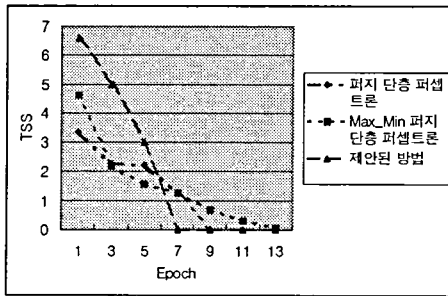


그림 3. XOR에서 epoch 수에 따른 TSS 변화

10개의 한글 자음 패턴 분류는 그림 4와 같은 패턴을 학습 패턴으로 사용하여, 7×8 크기의 행렬을 56×1의 열 벡터로 변환하여 입력으로 사용하였다.

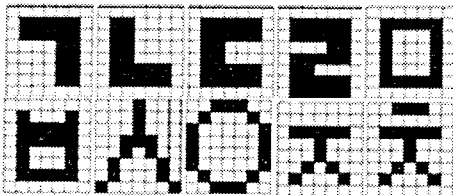


그림 4. 한글 자음 학습 패턴

한글 자음 패턴에 대한 기존의 학습 방법들과 제안된 방법간의 epoch 수와 TSS를 표 2로 나타내었다.

표 2. 한글 자음 패턴 분류 문제에서 Epoch수 비교

XOR	학습률	Epoch 수
퍼지단층 퍼셉트론	0.2	17
	0.5	75
	0.7	x
Max_Min 퍼지 단층 퍼셉트론	0.2	x
	0.5	x
	0.7	x
제안된방법	0.2	18
	0.5	18
	0.7	19

표 2의 각 학습 알고리즘에서 모멘텀은 0.5로 설정하였다. 기존의 퍼지 단층 퍼셉트론 알고리즘은 초기 학습률의 설정에 따라 학습의 수렴성이 민감하게 반응하는 것을 확인할 수 있다. 학습률을 0.2로 설정한 경우에는 학습의 수렴성이 보장되었으나 학습률을 0.5로 설정하는 경우에는 목표값과 실제 출력값이 상치되어 조기 포화 현상이 발생하여 학습 시간이 매우 많이 소요되었다. 그리고 학습률을 0.7로 설정하는 경우에는 지역 최소화에 빠져 학습이 되지 않았다. Max_Min 퍼지 단층 퍼셉트론은 퍼지 기본 논리 연산자인 Max와 Min 연산이 적용되기 때문에 학습 패턴이 0과 1로 구성되어 있는 자음 학습 패턴 분류 문제에서는 전혀 학습이 되지 않았다. 반면에 제안된 학습 알고리즘은 기존의 두 학습 알고리즘보다 학습의 수렴성이 개선되었고 학습 시간이 적게 소요되는 것을 알 수 있다. 또한 학습률에 따라 학습 성능이 민감하게 반응하지 않는 것을 확인할 수 있다.

한글 자음 패턴에 대한 제안된 방법의 epoch 수에 따른 TSS 변화 과정을 그림 5로 나타내었다.

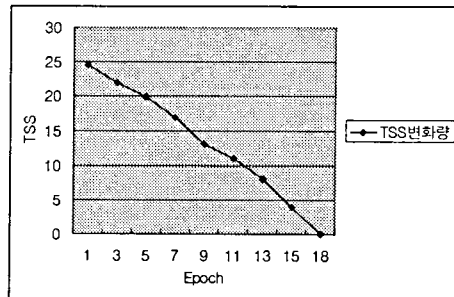


그림 5. 숫자 패턴에서 제안된 방법의 epoch수에 따른 TSS 변화

V. 결 론

기존의 단층 퍼셉트론은 출력 노드가 선형 분리 가능한 패턴들만을 분류할 수 있고 Exclusive OR 와 같은 비선형 문제에 대해서는 분류할 수 없는 단점이 있다. 그러나 퍼지 단층 퍼셉트론은 퍼지 소속 함수(fuzzy membership function)를 사용하여 단층 구조로 Exclusive OR 문제와 같은 고전적인 문제를 개선하였다. 그러나 퍼지 단층 퍼셉트론은 기존의 단층 퍼셉트론과 마찬가지로 결정 경계선이 진동하는 경우가 생기며 초기 가중치의 범위와 학습률에 따라 수렴성이 매우 낮아지는 단점이 있다. 따라서 본 논문에서는 바이어스항을 도입하여 결정 경계선이 진동하는 것을 방지하여 수렴성을 개선시키고 학습률과 모멘트 개념을 도입하여 학습 시간을 단축시키는 개선된 퍼지 단층 퍼셉트론 알고리즘을 제안하였다.

제안된 방법과 기존의 퍼지 단층 퍼셉트론들 간의 학습 성능을 분석하기 위하여 인공 신경망에서 벤치마크로 사용되는 exclusive OR 문제와 한글 자음 패턴 분류에 적용하여 epoch 수와 수렴성을 비교한 결과, XOR 문제에서는 제안된 방법이 기존의 퍼지 단층 퍼셉트론들보다 학습 시간이 적게 소요되었고 한글 자음 패턴 분류 문제에 있어서는 제안된 방법은 수렴하였으나 기존의 퍼지 단층 퍼셉트론은 지역 최소화에 위치하여 학습되지 않았다. 따라서 제안된 방법이 기존의 퍼지 단층 퍼셉트론들보다 학습 시간이 적게 소요되고 수렴성이 개선된 것을 확인하였다.

향후 연구 방향은 제안된 방법을 복잡한 비선형 문제에 적용할 수 있도록 하기 위해 다층 구조로 확장하는 방법을 연구할 것이다.

참고문헌

- [1] J. B. Kiszaka, M. M. Gupta, "Fuzzy Logic Neural Network," BUSFAL, No.4, 1990.
- [2] M. M. Gupta and J. Qi, "On Fuzzy Neuron Models," proceedings of IJCNN, Vol.2, pp.431-435, 1991.
- [3] Marin T. Hagan, Howard B. Demuth, Mark Beale, Neural Network Design, PWS Publishing, Boston, 1995.
- [4] T. H. Goh, P. Z. Wang and H. C. Lui, "Learning Algorithm for Enhanced Fuzzy Perceptron," proceedings of IJCNN, Vol.2, pp.435-440, 1992.
- [5] K. B. Kim, M. H. Kang and E. Y. Cha, "A Fuzzy Competitive Backpropagation using Nervous System," in Proc. WCSS, pp.188-193, 1997.
- [6] R. Hecht-Nielsen, "Theory of Back-propagation Neural Networks," proceedings of IJCNN, Vol.1, pp.593-605, 1989.