ART1 기반 퍼지 지도 학습 알고리즘

김광백

신라대학교 컴퓨터공학과

ART1-based Fuzzy Supervised Learning Algorithm

Kwang-Baek Kim

Dept. of Computer Engineering, Silla University

E-mail : gbkim@silla.ac.kr

요 약

본 논문에서는 오류 역전파 알고리즘에서 은닉층의 노드 수를 설정하는 문제와 ART1의 경계 변수의 설정에 따른 인식률이 저하되는 문제점을 개선하기 위해 ART1 알고리즘과 퍼지 단층 지도 학습 알고리즘을 결합한 ART1 기반 퍼지 지도 학습 알고리즘을 제안한다. 제안된 알고리즘은 가중치 조정에 승자 뉴런 방식을 도입하여 은닉층에 해당하는 클래스에 영향을 끼친 패턴들의 정보만 저장하게 하여 은닉층 노드로의 책임 분담에 의한 정체 현상이 일어날 가능성을 줄인다. 그리고 학습 시간과 학습의 수렴성도 개선한다. 제안된 알고리즘의 학습 성능을 분석하기 위하여 주민등록번호 분류를 대상으로 실험한 결과, 제안된 방법이 기존의 신경망보다 경계 변수나 모멘트에 민감하지 않으며 학습 시간도 적게 소요되고 수렴성도 우수한 성능이 있음을 확인하였다.

키워드

오류 역전파, 은닉층의 노드 수, ART1, 경계 변수, 퍼지 단층 지도 학습 알고리즘, ART1 기반 퍼지 지도 학습 알고리즘

Ⅰ. 서 론

인공 신경망의 지도 학습(supervised learning) 모델 중의 하나인 오류 역전파(error backpropagation) 알고리즘은 단층 퍼셉트론으로 해결하지 못했던 비선형 문제를 해결하기 위해 제시되었던 다층 구조 신경망(multilayer neural network)의 대표적인 학습 알고리즘이다[1]. 이 학습 알고리즘은 적용되는 입력 패턴에 대해 파라미터 값들을 어떤 값으로 설정하는가에 따라 학습 속도가 결정되어지고, 은닉층의 노드 수의 결정에 따라 지역 최소화(local minima)문제가 발생한다[2]. 오류 역전파 알고리즘에서 은닉층(hidden layer)의 노드 수가 필요 이상으로 많으면 판별 성능에 영향을 거의 미치지 않는 부가적인 노드가 존재하여 학습 시간이 많이 소요되고 은닉층의 노드 수가 적으면 연결 가중치들이 지역 최소화에 위치할 가능성이 높아진다[3,4]. 그러므로 간단한 선형 분리 문제나 Exclusive OR 문제는 필요한 은닉층의 노드 수를 알기 쉽지만 영상 인식 등과 같은 복잡한 문제의 경우는 적당한 은닉층의 노드 수를 찾아내기가 어렵다. Crossberg와 Capenter에 의해 제안된 ART(Adaptive Resonance Theory) 신경망 모델은 뉴런들 간에 경쟁 학습(competitive learning)에 의하여 자율적으로 패턴들을 분류하는 네트워크 구조이다. 임의의 패턴을 학습할 수있는 안정성(stability)과 적응성(plasticity)을 갖는 신경망의 학습 모델이다[5]. 또한 저속 및 고속 학습이 가능한 장점을 가지고 있고, 지역 최소화(local minima) 문제가 발생하지 않는다[6]. ART 신경망 모델 중에서 ART1 알고리즘은 자동 학습 시스템(autonomous learning system)구축에 널리 사용되고 있다. ART1 알고리즘은 첫 번째 입력을 첫 번째 클러스터의 대표 패턴으로 선택하고 다음 입력이 들어오면 첫 번째 대표 패턴과의 유사성을 비교하여 경계 변수 보다 크거나 같으면 첫 번째 클러스터로 분류하고, 그렇지 않으면 새로운

클러스터를 생성하게 된다. 이러한 과정이 모든 입력 패턴들에 적용되면 클러스터의 개수는 시간에 따라 점점 증가하게 된다. 그러나 ART1 알고리즘은 경계 변수 설정에 따라 유사한 패턴이 다른 클러스터로 생성되는 경우와 서로 다른 패턴이 같은 클러스터로 분류되는 경우가 발생한다[7]. 따라서 ART1 알고리즘을 영상 인식에 적용할 경우, 입력 벡터와 클러스터의 대표 벡터간의 유사도 측정 방법과 경계 변수에 따라 인식률이 낮아지는 문제점이 있다.

본 논문에서는 오류 역전파 알고리즘에서 은닉층의 노드 수를 설정하는 문제와 ART1의 경계 변수의 설정에 따른 인식률이 저하되는 문제점을 개선하기 위해 ART1 알고리즘과 퍼지 단층 지도 학습 알고리즘을 결합한 ART1 기반 퍼지 지도 학습 알고리즘을 제안한다.

II. 관련 연구

2.1 퍼지 단층 지도 학습 알고리즘

퍼지 논리 연산자를 이용한 퍼지 단층 지도 학습 알고리즘은 0과 1사이의 입력 패턴을 학습하는 알고리즘이다. 기존의 단층 퍼셉트론은 이진 패턴의 학습에 효율적인 반면에 퍼지 단층 지도 학습 알고리즘은 이진 패턴뿐만 아니라 퍼지 데이터도 효율적으로 학습 할 수 있는 장점이 있다[8]. 그러나 이 알고리즘은 패턴 인식과 같은 실세계 문제에 적용할 경우에는 지역 최소화 문제가 발생하여 수렴이 되지 않는 경우가 발생한다.

퍼지 단층 지도 학습 알고리즘에서 NET는 입력 벡터(x_i)와 연결 가중치(w_i)의 최대-최소 합성(max-min composition) 연산에 의해 다음과 같이 계산된다.

$$NET = \bigvee\{\bigwedge\{x_i, w_{0j}\}, \ldots, \bigwedge\{x_i, w_{n-1j}\}\} \quad (1)$$

여기서 i는 입력층이고 $j(j=0, \ldots, n)$는 출력층이다.

식 2와 같이 출력 벡터(o_j)는 NET와 바이어스항(Θ_j)간의 퍼지 논리합 연산에 의해 계산된다.

$$o_j = NET \vee \Theta_j \quad (2)$$

연결 가중치(w_{ji})는 출력 벡터(o_j)와 목표 벡터(t_j)가 동일한 경우에는 연결 가중치(w_{ji})와 바이어스항(Θ_j)를 변경하지 않고, 출력 벡터가 목표 벡터와 다른 경우에는 다음 식을 이용하여 가중치와 바이어스항을 조정한다.

$$w_{ji}(n+1) = w_{ji}(n) + \alpha \Delta w_{ji}(n+1) + \beta \Delta w_{ji}(n+1)$$
$$\Theta_j(n+1) = \Theta_j(n) + \alpha \Delta \Theta_j(n+1) + \beta \Delta \Theta_j(n+1)$$

$$\begin{aligned} \Delta w_{ji} &= 1, \ when \ \ o_j = w_{ij}, \\ &= 0 \ \ otherwise. \end{aligned} \quad (3)$$
$$\begin{aligned} \Delta \Theta_j &= 1, \ when \ \ o_j = \Theta_j, \\ &= 0 \ \ otherwise. \end{aligned}$$

여기서 α는 학습률이고 β는 모멘텀이다. 퍼지 단층 학습 알고리즘에서 학습 종료는 총 오차 자승합(TSS : total sum of squares)이 오류 한계보다 적을 때까지 학습한다.

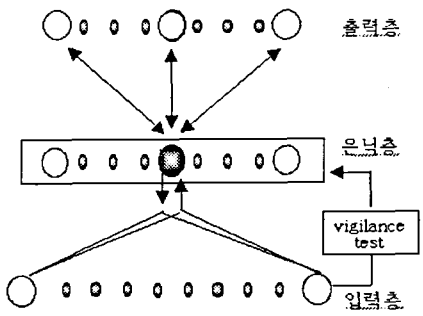
2.2 ART1 알고리즘

ART1은 1976년 보스턴 대학의 Stephen Grossberg가 경쟁학습(competitive learning)의 약점인 안정성을 보강하여 제안한 자율 학습 알고리즘이다. 이 알고리즘은 기존에 학습되었던 것이 새로운 학습에 의해 지워지지 않도록 새로운 지식을 자동적으로 전체 지식 베이스에 일관성 있는(self-consistent)방법으로 통합한다. 즉 적절하게 매치(match)되는 새로운 정보를 이용하여 이미 배운 내용들을 정제하며(refine), 새로운 인식 카테고리의 학습을 위하여 새로운 노드를 선택하고, 기억용량을 넘어서는 과다한 새로운 입력에 의해 기존에 취득한 내용이 지워지는 것을 방지한다. 따라서 끊임없이 변하는 환경에서 자신의 메모리 용량을 전부 소모할 때까지는 제한없는 입력에 대해 실시간으로 빠르고 안정되게 배울 수 있는 구조이다. ART1 알고리즘은 임의의 패턴을 학습할 수 있는 안정성(stability)과 적응성(plasticity)을 갖는 신경망의 학습 알고리즘이다. 또한 저속 및 고속 학습이 가능한 장점을 가지고 있고, 지역 최소화(local minima) 문제가 발생하지 않는다. ART1 알고리즘에서 경계 변수(vigilance parameter)는 패턴들을 클러스터링 하는데 있어서 반지름 값이 되며 임의의 패턴과 저장된 패턴과의 불일치(mismatch)허용도를 결정한다. 이 경계 변수가 크면 입력 패턴과 기대 패턴 사이에 약산의 차이만 있어도 새로운 카테고리(category)로 분류하게 된다. 반대로 경계 변수가 작으면 입력 패턴과 기대 패턴 사이에 많은 차이가 있더라도 허용하므로 입력 패턴들을 대략적으로 분류한다[9]. 그러므로 실제 영상 인식과 같은 분야에 적용하기 위해서는 경험적으로 경계 변수를 설정해야 한다. 그리고 서로 다른 유사한 특징을 갖는 패턴들이 많을 경우에는 경계 변수를 설정하는 것은 매우 어렵다.

III. ART1 기반 퍼지 지도 학습 알고리즘

은닉층의 노드를 자가 생성하도록 구성된 제안된 구조는 그림 1과 같다.

그림 1. 제안된 ART1 기반 퍼지 지도 학습 모형



제안된 학습 알고리즘에서 입력층과 은닉층간의 연결 구조는 ART1의 구조와 같고, ART1의 출력층이 제안된 구조의 은닉층이다. 은닉층의 노드는 각 클래스를 나타낸다. 따라서 전체적으로는 완전 연결 구조이나 출력층에서 목표 벡터와 실제 출력 벡터를 비교하여 역전파시킬때는 대표 클래스와 연결된 가중치만을 역전파하는 승자 뉴런 방식(winner-take all method)을 채택한다. 출력층에서 은닉층으로의 연결 가중치 조정 방식은 퍼지 단층 학습 방법을 사용한다.

3. 1 은닉층의 동적인 생성과 대표 패턴 선정

은닉층 노드의 생성은 제시된 패턴들로 결정될 클래스 수에 기반 하여 본 논문에서는 ART1을 기반한 방법에 은닉층의 초기 최대 노드 수를 클래스 수로 생성되도록 한다. 하나의 노드로 시작하여 그 노드를 입력층에서 처음 제시된 패턴에 해당하는 클래스로 할당한다. 그 다음부터 제시될 패턴들은 현 상태에서 존재하는 노드들에 대해서 승자를 채택하고 승자 노드 선정에 실패할 경우에는 노드를 하나 증가하여 해당 패턴에 대한 클래스로 할당한다. 이런 방식으로 모든 패턴들이 제시되면 클래스에 따른 노드가 동적으로 생성되게 한다. 그러나 입력 패턴이 제시되어 선정된 승자 노드가 그 패턴에 대한 대표 클래스가 되는 것은 아니다. 여기서 승자 노드의 저장 패턴과 유사성을 측정하여 유사성이 인정되면 승자 노드로 채택되고, 대표 클래스가 된다. 만약 유사성이 인정되지 않으면 새로운 노드를 할당한다.

3. 2 가중치 조정 문제

제안된 구조는 가중치 조정을 위해 기존의 퍼지 단층 지도 학습에 승자 뉴런 방식을 채택한다. 입력층과 은닉층간의 연결과 은닉층과 출력층간의 연결로 구분해서 살펴보면 은닉층에서 채택된 승자 노드는 제시된 패턴의 대표 클래스가 된다. 따라서 대표 클래스의 저장 패턴에 제시된 패턴을 반영하기 위해서는 은닉층에서 입력층으로의 승자 노드와 연결된 가중치를 조정한다. 또한 제시된 패턴에 대한 목표 벡터를 대표 클래스에 의한 실제 출력 벡터에 반영하기 위해 출력층 노드와 그 대표 클래스와 관련된 가중치만을 조정한다.

3. 3 ART1 기반 퍼지 지도 학습 알고리즘

단계 1: 입력층과 은닉층 사이의 상향 가중치인 w_{ji}와 하향 가중치인 t_{ji}를 초기화하고, 은닉층과 출력층 사이의 가중치인 w_{kj}와 바이어스항 Θ_k를 1로 초기화한다. 여기서 i ($i=1...m$)는 입력층, j($j=1...l$)는 은닉층, k($k=1...p$)는 출력의 노드이다.

$$t_{ji}(0)=1,\ w_{ji}(0)=\frac{1}{m+1}$$

단계 2: 입력 벡터 x_i와 목표 벡터(t_k)를 제시한다.

단계 3: 은닉층의 출력 벡터(o_j)를 계산한다.

$$o_j=\sum_{j=1}^{n} w_{ji}\times x_i$$

단계 4: 승자 노드(o^*_j)를 선택한다.

$$o^*_j=Max[o_j]$$

단계 5: 승자 노드의 하향(t_{j^*i}) 가중치와 입력 패턴간의 유사성을 계산한다.

$$||X||=\sum_{i=1}^{m} x_i$$

$$||T\odot X||=\sum_{i=1}^{m} t_{j^*i}\times x_i$$

$$\frac{||T\odot X||}{||X||}$$

단계 6: 만약 계산된 유사도가 $\frac{||T\odot X||}{||X||}\geq\rho$이면 현재의 승자 노드를 대표 클래스로 선택하고 단계 8로, 그렇지 않으면 단계 7로 간다. 여기서 ρ는 경계 변수로서 0과 1사의 값이다.

단계 7: 현재 승자 노드 (o^*_j)의 출력 벡터를 0으로 하고 새로운 노드를 할당하여 대표 클래스로 선택하고 단계 8로 간다.

단계 8: 입력 패턴과 승자 노드간의 유사성이 인정된 단계 6의 대표 클래스와 단계 7에서 새로 생성된 대표 클래스의 상향과 하향 가중치를 조정한다.

$$t_{j^*i}(n+1) = t_{j^*i}(n) \times x_i$$
$$w_{j^*i}(n+1) = \frac{t_{j^*i}(n+1) \times x_i}{0.5 + \sum_{i=1}^{m} w_{j^*i} \times x_i}$$

단계 9: 은닉층의 대표 클래스(o^*_j), 은닉층의 대표 클래스와 출력층의 연결 가중치(w_{kj^*})를 이용하여 출력층의 NET를 계산하고 NET와 바이어스항(Θ_k)를 이용하여 출력층의 출력 벡터(o_k)를 계산한다.

$$NET = \bigvee\{\bigwedge\{o^*_j, w_{1j^*}\}, \ldots, \bigwedge\{o^*_j, w_{pj^*}\}\}$$
$$o_k = NET \bigvee \Theta_k$$

단계 10: 연결 가중치(w_{kj^*})와 바이어스항(Θ_k) 및 학습률(α_k)을 조정한다. 여기서 β는 모멘텀이다.

$$w_{kj^*}(n+1) = w_{kj^*}(n) + \alpha_k \times \Delta w_{kj^*}(n+1) + \beta \times \Delta w_{kj^*}(n)$$

$$\Theta_k(n+1) = \Theta_k(n) + \alpha_k \times \Delta\Theta_k(n+1) + \beta\Delta\Theta_k(n)$$

$$\alpha_k = (t_k - o_k)^2$$

$$\Delta w_{kj^*} = 1, \ when \ o_k = w_{kj^*},$$
$$= 0 \ \ otherwise.$$

$$\Delta\Theta_k = 1, \ when \ o_k = \Theta_k,$$
$$= 0 \ \ otherwise.$$

$$\text{if } |t_k - o_k| < \varepsilon \ then \ \ \Delta\alpha_k(n+1) = e^2$$

$$\alpha_k(n+1) = \alpha_k(n) + \Delta\alpha_k$$

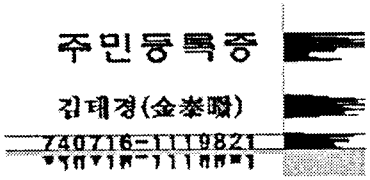
단계 10: 총오차 자승합(TSS)이 오류 한계(ε)보다 크면 단계 3으로 가고 오류 한계보다 적거나 같으면 학습을 종료한다.

IV. 실험 및 결과 분석

본 논문에서 제안된 알고리즘의 학습 성능을 분석하기 위하여 Intel Pentium-550MHz CPU와 256 MB RAM이 장착된 IBM 호환 PC상에서 비주얼 C로 구현하였다. 실험에 사용된 데이터는 HP Scanjet 4200C 스캐너를 이용하여 500 × 300 크기를 가진 5개의 주민등록증 영상에 추출한 10개의 개별 주민등록번호(0,1,2,3,4,5,6,7,8,9)이다.

주민등록증 영상에서 주민등록 번호를 추출하기 위해 주민등록증 영상을 평균 밝기 값을 적용하여 이진화하였다. 이진화된 영상에서 주민등록번호의 위치가 고정되어 있다는 사전의 정보를 이용하여 주민등록번호 영역이 포함된 일부 영역에 대해서만 수평 방향으로 히스토그램을 수행하여 주민등록 번호 영역을 추출하였다. 수평 방향으로 히스토그램을 적용한 후의 특징 영역은 그림 2의 b와 같이 3개영역으로 구분되고 주민등록증 영역은 3번째 부분에 위치해 있으므로 3번째 영역을 주민등록번호 영역으로 추출하였다. 주민등록번호 영역의 추출 단계를 나타내면 그림 3과 같다.

(a) 원 영상



(b) 주민등록번호 영역 추출 과정

그림 2. 주민등록번호 영역 추출 과정

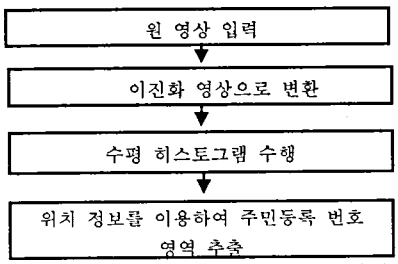


그림 3. 주민등록번호 영역 추출 단계

추출된 주민등록번호 영역 대해서 수직 방향 히스토그램을 이용하여 개별 문자를 추출하였다. 추출된 개별 문자를 10×15 크기의 패턴으로 정규화하였다. 개별 문자를 추출하는 단계는 그림 4와 같다.

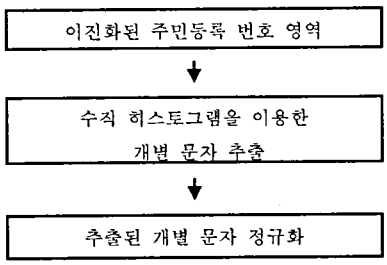


그림 4. 개별 문자 추출 단계

추출된 개별 문자는 그림 5와 같다.

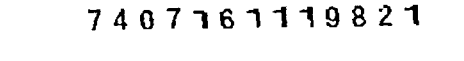


그림 5. 추출된 개별 문자

5개의 주민등록증 영상에서 추출된 65개의 개별 주민등록번호 중에서 그림 6과 같은 10개의 개별 문자를 제안된 방법에 적용하여 모멘텀에 따른 epoch 수와 TSS를 표 1로 나타내었다.

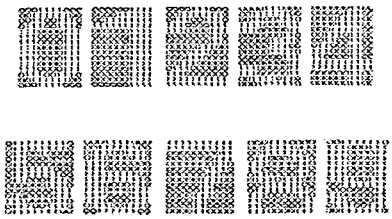


그림 6. 주민등록번호 학습패턴

표 1. 모멘텀에 따른 Epoch수와 TSS

	모멘트	epoch 수	TSS
제안된 학습알고리즘	0.5	4	0.043961
	0.7	4	0.005470
	0.8	4	0.019126

표 1의 결과는 오류 한계를 0.05로 설정하여 학습한 결과이다. 제안된 방법에서 학습률은 동적으로 조정되므로 모멘트를 0.5로 설정했을 때 최종 학습율은 0.000008로 조정되었고 모멘텀이 0.7일 때 최종 학습률은 0.000193, 모멘텀이 0.8일 때 최종 학습률은 0.003839로 조정되었다. 그리고 은닉층에서 생성된 노드 수는 각각 10개가 생성되었다. 표 1에서와 같이 제안된 알고리즘은 모멘텀에 민감하게 반응하지 않고 학습 시간이 적게 소요되었다. 제안된 알고리즘에서 모멘텀에 따른 epoch 수와 TSS 변화 과정은 그림 7과 같다.

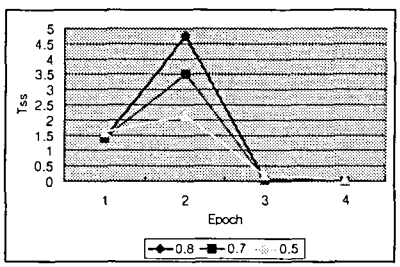


그림 7. 모멘텀에 따른 Epoch 수와 TSS 변화 과정

그리고 제안된 알고리즘에서 은닉층의 노드를 생성 또는 갱신하는데 사용되는 경계 변수를 각각 0.93에서 0.99로 설정하여 학습한 결과는 표 2와 같다. 표 2에서 모멘텀은 0.5로 설정하였다.

표 2. 경계 변수에 따른 epoch 수와 생성된 노드 수

	경계 변수	epoch 수	생성된 은닉층의 노드수
제안된 학습 알고리즘	0.93	4	10
	0.94	4	10
	0.95	4	10
	0.96	4	10
	0.97	4	10
	0.98	4	10
	0.99	4	10

V. 결 론

본 논문에서는 오류 역전파 알고리즘에서 은닉층의 노드 수를 설정하는 문제와 ART1의 경계변수의 설정에 따른 인식률이 저하되는 문제점을 개선하기 위해 ART1 알고리즘과 퍼지 단층 지도 학습 알고리즘을 결합한 ART1 기반 퍼지 지도 학습 알고리즘을 제안하였다. 제안된 알고리즘은 가중치 조정에 승자 뉴런 방식을 도입하여 은닉층에 해당하는 클래스에 영향을 끼친 패턴들의 정보만 저장하게 하여 은닉층 노드로의 책임 분담에 의한 정체 현상이 일어날 가능성을 줄였고, 학습 시간과 학습의 수렴성을 개선하였다. 제안된 알고리즘의 학습 성능을 분석하기 위하여 주민등록증 영상의 주민등록 번호를 분류하는데 적용한 결과, 제안된 방법이 다른 신경망 알고리즘보다 경계 변수나 모멘텀에 민감하게 반응하지 않으며 학습 시간도 적게 소요되고 수렴성도 개선된 것을 확인하였다. 그러나 제안된 알고리즘은 입력 패턴의 크기에는 민감하지 않으나 패턴 크기에 상관없이 패턴의 수가 많아질수록 은닉층에서 생성되는 노드가 증가하였다.

향후 연구과제는 패턴의 수에 민감하지 않게 은닉층의 노드를 생성하는 방법에 대해 연구할 것이고 더 나아가서 제안된 방법을 기반으로 하여 주민등록증 인식 시스템을 개발할 것이다.

참고문헌

[1] R. Hecht-Nielsen, "Theory of Backpropagation Neural Networks," Proceedings of IJCNN, Vol.1, pp.593-605, 1989.

[2] A. James and Freeman, Neural Networks : Algorithm, Application and Programming

Techniques, Addison-Wesley, 1991.
[3] R. A. Jacobs, "Increased rates of convergence through learning rate adaptation," Neural Networks, Vol.1, No. 4, pp.295-308, 1988.
[4] K. B. Kim, M. H. Kang and E. Y. Cha, "Fuzzy Competitive Backpropagation using Nervous System," Proceedings of WCSS, pp.188-193, 1997.
[5] M. Georipoulos, G. L. Heileman and J. Huang, Properties of Learning Related to Pattern Diversity in ART1," Neural Networks, Vol.4, pp.751-757, 1991.
[6] K. B. Kim and K. C. Kim "A Study on Face Recognition using New Fuzzy ART," Proceedings of ITC-CSCC, Vol.2, pp.1057 ~1060, 1998.
[7] K. B. Kim, S. W. Jang and C. K. Kim,"Recognition of Car License Plate by Using Dynamical Thresholding Method and Enhanced Neural Networks," Lecture Notes in Computer Science, LNCS 2756, pp.309-319, 2003.
[8] Y. Hayashi, E. Czogala and J. J. Buckley," Fuuzy Neural Controller," Proceedings of IEEE Fuzzy Systems, pp.197-202, Vol.1, 1992.
[9] K. B. Kim, Y. H. Joo, J. H. Cho, "An Enhanced Fuzzy Neural Network," Lecture Notes in Computer Science, LNCS 3320, pp.176-179, 2004.