

러프집합을 이용한 다층 신경망의 구조최적화에 관한 연구

論文

48A - 1 - 12

A Study on the Structure Optimization of Multilayer Neural Networks using Rough Set Theory

丁榮俊* 全孝炳* 沈貴寶**

(Young-June Chung · Hyo-Byung Jun · Kwee-Bo Sim)

Abstract - In this paper, we propose a new structure optimization method of multilayer neural networks which begin and carry out learning from a bigger network. This method reduces redundant links and neurons according to the rough set theory. In order to find redundant links, we analyze the variations of all weights and output errors in every step of the learning process, and then make the decision table from their variation of weights and output errors. We can find the redundant links from the initial structure by analyzing the decision table using the rough set theory. This enables us to build a structure as compact as possible, and also enables mapping between input and output. We show the validity and effectiveness of the proposed algorithm by applying it to the XOR problem.

Key Words : Neural networks, Back-propagation learning, Hyperplane, Rough set

1. 서 론

일반적으로 신경망의 정보처리능력은 그 신경망의 구조에 의해서 좌우되는 것이 보통이다. 우리가 신경망을 이용하여 어떤 문제를 해결하고자 할 때 최적의 신경망 구조(은닉층 수와 은닉층 뉴런의 개수)를 결정하는 것은 매우 중요한 문제이다. 만약 이들의 수가 너무 많으면 자유도가 증가하여 학습에 많은 시간이 걸리고, 학습시 초기값에 따라 가중치가 전역 최소값에 도달하지 않으며, 계산량이 증가하는 등의 문제가 발생한다. 반대로 그 수가 너무 적으면 입출력 패턴을 기억하지 못하고 학습이 최적해에 수렴하지 않는다. 일반적으로 입·출력층의 뉴런수는 문제에 따라 산출할 수 있으나 은닉층에 대한 정보는 알 수 없기 때문에 은닉층을 몇 개로 할 것인지, 각 은닉층의 뉴런 수를 몇 개로 할 것인지 결정하기가 어렵다. 따라서 대부분의 신경망의 구조 설계는 전문가의 경험적 지식에 의존하고 있는 실정이다. 최근 최적의 신경망을 얻기 위한 연구가 여러 방법으로 활발히 진행되고 있다. 그 접근방법으로 최소 크기에서부터 은닉층 뉴런과 링크 등을 하나씩 추가해 가면서 최적의 신경망을 구성하는 방법[1], 퍼지추론 등을 이용하여 최대 크기의 신경망으로부터 불필요한 구성요소들을 하나 하나씩 제거해 가면서 최적의 신경망을 구성하는 방법[2], 진화 알고리즘 등을 이용하여 최적의 신경망을 구성하는 방법[3][4][5] 등이 보고되고 있다. 그러나 이들의 연구가 제각기 우수한 결과를 제시하고 있

음에도 불구하고 여전히 다음과 같은 문제점을 안고 있다. 진화 알고리듬을 이용한 방법에서는 주어진 문제가 복잡해 질수록 유전자의 길이가 길어져야 하기 때문에 연산속도가 많이 걸리며, 진화에 필요한 또 다른 변수들을 설정자가 결정해 주어야하는 전문가의 또 다른 개입이 필요하게 된다. 또한 퍼지추론 등을 이용한 방법에서는 기본적으로 연결강도의 값에 따라 연결강도의 중요도를 평가하기 때문에 이때 사용되는 소속함수를 전문가가 정해 주어야 하며, 연결강도의 중요도를 연결강도 자체의 값으로 판단하기 때문에 다른 변수와의 연관성을 보장 할 수 없는 등의 단점이 있다.

한편 초평면(hyperplane)의 관점에서 본 최적의 신경망은 최소의 초평면과 최소의 판단규칙으로 주어진 문제를 해결 할 수 있는 신경망으로 볼 수 있다. 따라서 신경망의 구조가 최적구조보다 작을 경우 주어진 문제를 해결 할 수 있으며, 신경망의 구조가 최적구조보다 클 경우 주어진 문제를 해결할 수는 있지만 불필요한 신경망 요소가 내포되어 있어 신경망의 해석을 어렵게 할뿐만 아니라 시뮬레이션이나 하드웨어 구성에 있어서 자원을 낭비하고 계산량을 증가시켜 전체적인 성능을 저하시키는 요인이 된다.

본 논문에서는 신경망의 구조를 최적화하는 새로운 방법으로 러프집합 이론을 이용한 정보처리방법을 제안한다. 러프집합 이론은 1982년 Pawlak에 의해서 제안된 이론으로, 불충분하고 모순된 정보로부터 자동적이고 체계적인 방법으로 중요한 정보를 추출해 낼 수 있다[6][7]. 이 러프집합은 어떤 방대한 정보자료로부터 유효정보의 손실 없이 최소의 판단규칙을 추출해 낼 수 있으며, 이를 자료사이에 숨어있는 정보의 패턴을 발견해 낼 수 있다.

한편 다층 신경망을 오차역전파법으로 학습시킬 때, 연결강도의 증·감량(Δw)은 이전 단계의 출력오차(e_i)와 각 뉴런의 출력함수(y_i)에 의해서 결정되고, 그 다음 단계의 출력

* 準會員：中央大 制御計測學科 碩士課程

** 正會員：中央大 電子電氣工學部 副教授 · 工博

接受日字：1998年 9月 11日

最終完了：1998年 11月 3日

오차는 입력과 변경된 연결강도 그리고 뉴런의 출력함수에 의해 결정된다. 그런데 입력과 뉴런의 출력함수는 일정하기 때문에 결과적으로 신경망의 연결강도와 출력오차만이 서로 어떤 관계를 가지면서 오차를 감소시키는 방향으로 학습하게 되는 것이다. 따라서 이들 두 변수는 신경망의 학습에 결정적인 영향을 미치게 된다. 만약 학습이 완료된 어떤 신경망에서 학습이 진행되는 동안에 이들 두 변수의 관계를 관찰할 수 있다면 학습에 관계되지 않는, 또는 학습에 덜 관계되는 링크(우리는 이것을 redundant link라고 한다)를 발견할 가능성이 있다.

하지만 이 두 변수의 관계를 정형화하는 것은 상당히 어려운 문제이다. 신경망을 초평면의 관점에서 해석할 때, 출력오차의 변화에 영향을 주지 않는 연결강도는 출력오차를 감소시키는데 영향을 주지 않는 연결강도이거나 아니면 이미 최적의 값을 가지게 되어서 더 이상 출력 오차를 감소시키는데 영향을 주지 못하는 연결강도로 생각 할 수 있다. 만약 신경망의 출력오차를 감소시키는데 영향을 주지 않는 연결강도를 찾을 수 있다면, 이 연결강도를 신경망에 불필요한 요소로 간주하고 제거해도 신경망의 성능에는 전혀 영향을 주지 않으며 오히려 계산속도나 출력에 있어서 더 좋은 성능을 보일 수 있을 것이다. 본 논문에서는 이러한 문제를 러프집합이론을 이용하여 해결하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서 본 연구에 사용된 러프집합 이론의 개념과 정보시스템에 대한 소개, 3장에서 본 연구에서 도입한 초평면 기반의 최적화 신경망에 대한 정의, 4장에서 러프집합을 이용하여 최적의 신경망을 얻는 방법, 5장에서 본 연구의 방법을 적용한 예, 그리고 6장에서 결론을 맺는다.

2. 정보시스템과 러프집합 이론

러프집합 이론은 1982년 Pawlak에 의해서 제안된 정보분석방법으로, 정보시스템의 개념을 기반으로 하고 있다. 본 절에서는 이러한 내용을 간단히 기술한다[6][7].

정보시스템(Information System) I 는 개체(object)로 이루어진 유한 전체집합 U , 속성(Attribute)으로 이루어진 유한 집합 A 에 의한 관계로 $I = (U, A)$ 로 나타낸다. 즉, U 의 원소인 개체가 A 의 원소인 a 의 어떤 속성 값과 연결되는 관계를 말한다. 전전부와 후전부로 나뉘는 판단테이블(decision table)도 이러한 정보시스템의 하나이다. 판단테이블의 어떤 두 개체가 전전부는 동일하고 후전부가 다른 경우, 우리는 이 자료를 가지고 판단을 내릴 수 없는 모순된 상황에 이르게 된다. 이러한 경우 어떤 측정치들을 통한 제약을 적용한다면 판단을 내릴 수 있게 된다.

러프집합은 근사화(Approximation)의 개념을 이용하여 이러한 모순 상황에 대한 측정치를 정의하고 있으며 이를 통해서 어떤 범위 안에서 주어진 정보에 대한 자동적인 규칙을 찾을 수 있게 해준다. 이렇게 정보시스템의 판단테이블에서 얻은 규칙인 최소규칙은 판단에 관여하지 않는 요소를 제거한 것이며 가장 적은 정보로 전체 시스템을 대표하며 정보손실 없이 모든 정보를 포함하게 된다.

다음은 본 논문에서 사용한 러프집합 이론에 관련된 용어들에 대한 정의를 간단히 나타낸다.

- 근사(하한, 상한)와 경계영역 :

X 가 전체영역 U 의 부분집합일 때 B -하한근사, B -상한근사, B -경계영역은 각각 다음과 같이 정의된다.

X 의 B -하한근사(B -lower approximation)는 조건속성집합 B 에 의해서 X 로 확실하게 분류될 수 있는 모든 개체들의 집합이며 다음과 같이 나타낸다.

$$B_{\cdot}(X) = \{x \in U : B(x) \subseteq X\} \quad (1)$$

X 의 B -상한근사(B -upper approximation)는 조건속성집합 B 에 의해서 X 로 분류될 가능성이 있는 모든 개체들의 집합이며 다음과 같이 나타낸다.

$$B^{*}(X) = \{x \in U : B(x) \cap X \neq \emptyset\} \quad (2)$$

X 의 B -경계영역(B -boundary region)은 조건속성집합 B 에 의해서 X 로 분류하기 모호한 모든 개체들의 집합이며 다음과 같이 나타낸다.

$$BN_B(N) = B^{*}(X) - B_{\cdot}(X) \quad (3)$$

- 러프집합(Rough set)

집합의 경계영역이 공집합이 아닌 경우, 그 집합을 러프집합이라 한다.

- 의존도(Dependency) :

의존도는 얼마나 많은 수의 하한근사 개체가 존재하는가의 척도이다. 의존도의 척도 k 는 다음과 같이 정의된다.

$$k = \frac{|POS_C(D)|}{|U|} = \sigma(C, D) \quad (4)$$

$$POS_C(D) = \bigcup_{X \in U \setminus IND(D)} C_{\cdot}(X) \quad (5)$$

- 리덕트(Reduct) :

조건속성의 한 부분집합 B 에 의한 판단이 전제 조건속성집합에 의한 판단과 같다면, 판단테이블의 정보는 집합 B 에 의해서 보존된다. 이러한 집합 B 중 가장 적은 조건속성을 가지는 집합을 리덕트라 한다.

- 규칙(Rules) :

규칙은 조건과 판단의 관계를 말한다. 하한근사 규칙은 하한근사 개체에 의해서만 생성되며, 베스트규칙(best rules)은 하한근사 개체와 상한근사 개체 모두에 의해서 생성된다. 경계영역에 속한 개체들의 규칙 생성은 많은 개체가 속한 판단으로 규칙을 정한다.

3. 초평면에 기반한 최적의 신경망 구조

본 연구에서 고려하는 최적의 신경망 구조는 주어진 문제에 대해서 출력오차 허용범위를 만족하는 가장 작은 구조의 신경망으로 정의한다. 이 장에서는 신경망을 초평면의 관점에서 해석하여 최적인 신경망의 의미를 명확히 한다. 두 개의 은닉층을 갖는 다층 신경망이 있을 때, 첫 번째 은닉층은 초평면층에, 두 번째 은닉층은 규칙층에, 출력층은 규칙의 조합층에 각각 해당한다. 모든 신경망의 입력신호는 신경망의 첫 번째 은닉층에 보내진다.

첫 번째 은닉층의 연결강도는 초평면을 형성하게 되며 이

초평면들은 분류의 경계를 이루게 된다. 첫 번째 은닉층 뉴런의 출력함수가 계단형(step) 함수일 경우 뉴런의 출력은 각 초평면에 의해서 나누어지는 영역을 나타내고, 시그모이드(sigmoid)함수 일 경우 뉴런의 출력은 입력이 각 초평면에 의해서 나누어지는 영역에 속한 정도를 나타낸다. 본 논문의 내용은 뉴런의 출력함수가 계단함수인 전제아래 기술하고 있지만 시그모이드 함수나 기타 다른 잘 알려진 출력함수의 경우에도 이진판단이 아닌 것 이외에는 같은 방식으로 해석 할 수 있다.

두 번째 은닉층의 연결강도는 초평면의 합성 역할을 하며 규칙을 생성시킨다. 각 연결강도는 초평면에 의해서 나누어진 영역이 각 규칙에 관계하는지의 여부를 결정하게 된다. 이 은닉층의 뉴런은 판단규칙에 해당되면 뉴런의 출력은 입력이 판단규칙에 의해서 판단된 결과를 나타낸다.

출력층의 연결강도는 규칙의 합성 역할을 한다. 두 번째 은닉층에서 발생된 규칙 중 일부는 같은 판단을 내리는 규칙이다. 즉, 같은 분류를 하는 영역이 두 영역 이상 나타날 수 있다. 한편 출력층 연결강도는 같은 판단을 내리는 영역을 묶어주는 역할을 한다. 출력뉴런의 출력은 최종적으로 입력 정보가 어느 영역으로 분류되는지를 나타낸다.

3.1 신경망의 불필요한 요소

신경망의 최적 구조를 모르는 상태에서 전문가에 의해서 설계된 신경망의 구조는 주어진 문제 해결에 필요 이상의 링크나 뉴런을 포함할 수 있다. 첫 번째 은닉층에서는 불필요한 링크에 의해서 필요 이상의 초평면이 나타날 가능성이 있다. 만약 두 개 이상의 초평면이 하나의 분류 경계를 나타낸다면 신경망의 성능은 그 중 하나만의 초평면으로 구성된 신경망에서 얻을 수 있는 성능과 차이가 없으므로 하나의 초평면을 제외한 나머지 초평면은 불필요한 초평면이 된다. 이 불필요한 초평면은 첫 번째 은닉층에서 발생하기 때문에 이 영향으로 불필요한 규칙이 두 번째 은닉층에서, 그리고 불필요한 규칙의 합성이 출력층에서 계속해서 발생하게 된다.

한편 다중 신경망의 뉴런들은 통상 입력층 뉴런을 제외한 모든 뉴런이 바이어스 입력을 갖는다. 뉴런은 입력층 연결강도로부터 입력된 정보로 출력을 계산하기 때문에 뉴런의 입력층 연결이 없다면 뉴런의 출력은 다음 층의 뉴런에 또 하나의 바이어스 입력으로 작용한다. 모든 뉴런이 바이어스 입력을 포함하고 있기 때문에 이러한 뉴런 또한 불필요한 신경망의 요소이다.

앞에서 언급한 바와 같이 최적의 신경망 구조는 주어진 문제에 대해서 출력오차 허용범위를 만족하는 가장 작은 구조의 신경망으로 불필요한 요소를 포함하지 않는 구조를 의미한다.

4. 러프집합을 이용한 불필요한 요소 발견과 신경망의 최적구조 구성

본 장에서는 러프집합을 이용하여 불필요한 신경망 요소를 제거하여 최적의 신경망 구조를 얻는 방법을 체계적으로 나타낸다.

4.1 러프집합을 이용한 불필요한 요소의 발견

오차역전과 학습법에서의 출력오차는 입력패턴, 연결강도, 그리고 뉴런의 출력함수에 의해서 결정된다. 따라서 다음 단계에서 개선될 연결강도의 변화량은 출력오차, 현재의 연결강도, 뉴런의 출력함수에 의존하게 된다. 그런데 연결강도와 출력오차를 제외한 나머지 변수들은 전체 학습구간에서 이미 정해진 요소이기 때문에, 실제로 학습에 영향을 미치는 것은 출력오차와 연결강도의 변화량뿐이다. 따라서 이 두 요소의 관계를 명확하게 밝힐 수 있다면 신경망의 불필요한 요소를 쉽게 발견할 수다. 하지만 이 두 요소 사이의 관계는 아주 비선형적이므로 정량적으로 해석하기가 매우 어렵다. 우리는 수차례 결친 실험 결과로부터 어떤 연결강도는 출력오차에 거의 비례적으로 변하고, 어떤 연결강도는 거의 변하지 않는 것 등이 존재하고 있는 것을 발견했다. 실제로 신경망의 불필요한 요소를 찾기 위해서는 이들 연결강도가 얼마나 많은 양이, 얼마나 많은 학습구간에서 변하는지를 분석해야 한다. 본 연구에서는 신경망의 불필요한 요소를 찾기 위해서 러프집합 이론을 적용한다.

러프집합 이론을 적용하기 위해서는 우선 판단 테이블을 만들어야 한다. 표 1은 본 연구에서 적용할 판단테이블의 구성 예를 나타낸 것이다. 표에서 조건속성은 연결강도의 변화량이며, 판단속성은 출력오차의 변화량이다. 오차역전과 학습에서 얻을 수 있는 이들 속성들은 실수 값으로 러프집합 이론을 적용하기 위해서는 양자화가 필요하다.

표 1 판단테이블

Table 1 Decision Table

학습단계	조건속성 (연결강도 변화량)						판단속성 (출력오차 변화량)
	W1	W2	W3	W4	W5	W6	
S							E
1	4	4	1	1	2	2	4
2	3	4	1	1	3	2	4
3	2	3	1	1	3	3	3
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
100	2	2	1	3	1	1	2
101	1	1	1	2	1	2	1

표 1의 판단테이블에서 유효한 정보만을 추출하기 위해서는 두 가지의 접근방법이 있다. 하나는 리덕트(reduct)를 찾는 것인데 이 방법은 판단속성에 영향을 주는 조건속성만을 찾는 방법으로 의존도가 1이 되는 정보시스템에서는 유용한 방법이다. 만약 판단테이블의 의존도가 1이 되지 않는 경우에 설계자가 의존도를 미리 정하여 적용할 수 도 있지만 추출된 정보에는 여전히 불확실한 데이터를 포함하게 된다. 또 다른 하나는 규칙을 찾는 방법이다. 이 방법은 앞의 방법에 비해 전문가의 지식을 필요로 하지 않으며, 본 연구에서는 판단테이블에서 얻은 규칙에 의하여 불필요한 요소를 찾는 방법을 제안한다. 생성된 규칙은 판단테이블의 정보를 유지하기 때문에 그 규칙에 속하는 조건속성들의 집합이 유익한 정보가 된다. 결과적으로 이 조건속성집합에 포함되지 않는 조건속성은 불필요한 조건속성이 되고, 이를 이용하여 불필요한 요소를 발견할 수 있게 된다.

4.2 신경망의 최적구조를 찾는 방법

본 연구에서 제안한 러프집합을 이용해서 불필요한 요소를 찾는 방법은 그림 1과 같은 과정에 의해서 진행되며, 각 단계별 세부 사항은 다음과 같다.

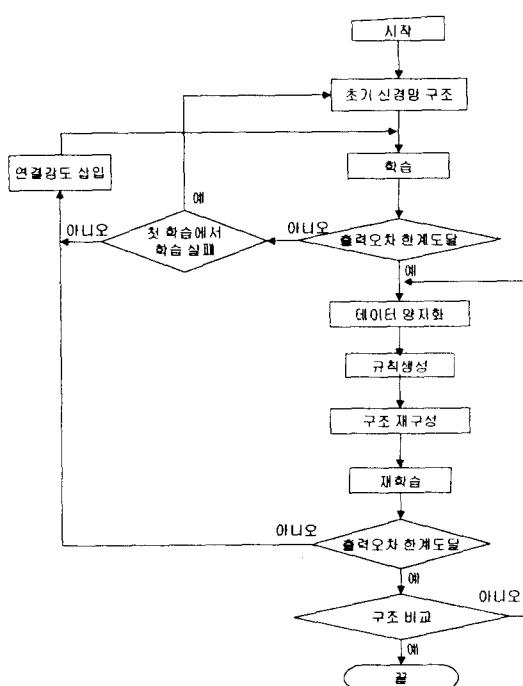


그림 1 불필요한 요소의 발견과 최적의 신경망을 얻는 순서도

Fig. 1 Flowchart for finding redundant links and construct optimal neural networks

• 초기 신경망 구조

초기 신경망은 완전 연결(fully-connected)된 다층(입력층, 두 개의 은닉층, 출력층) 구조이다. 입력층을 제외한 모든 뉴런은 독립적인 바이어스 입력을 가지며, 연결강도의 초기값은 임의의 값을 가진다. 뉴런의 출력함수는 설계자가 임의로 결정한다. 입력층과 출력층의 뉴런 수는 문제에 따라서 정해지기 때문에 신경망 구조의 정해지지 않은 요소는 각 은닉층의 뉴런의 개수이다. 첫 번째 은닉층의 뉴런 개수는 초평면의 개수와 같다. 초기 신경망 구조를 정할 때 초평면의 개수를 알 수 없다. 하지만, 발생할 수 있는 초평면의 최대 개수는 하나의 초평면이 하나의 입력패턴을 분류하는 경우로 입력 패턴의 개수 - 1 개이다. 두 번째 은닉층의 뉴런 개수는 초평면에 의해서 나누어지는 판단영역의 개수와 같다. 이 역시 초기구조를 정할 때 알 수 없다. 하지만, 발생할 수 있는 판단영역의 최대 개수는 초평면의 모든 조합으로 2^n 개이며, n 은 초기 신경망 구조의 초평면 개수이다.

• 학습 :

학습은 오차역전파 학습법을 사용한다. 오차역전파 학습법은 연결강도와 출력오차의 관계에 대한 정보를 제공하며, 러프집합에 의한 분석을 위해서 모든 학습단계에서의 연결

강도와 출력오차의 변화량을 저장한다.

• 데이터 양자화 :

러프집합 이론을 적용하기 위해서는 양자화 된 데이터를 갖는 판단테이블이 필요하므로 실수 값인 데이터들을 양자화해야 한다. 본 연구에서의 판단테이블에 쓰이는 데이터는 연결강도와 출력오차의 각 학습단계별 변화량이다. 변화량의 최대값은 초기구조와 문제에 따라서 어떤 값을 갖는지 알 수 없으므로 양자화 간격을 정해서 수행하는 방법이나 양자화 단계를 정해서 수행하는 방법은 효과적이지 못하다. 따라서 본 연구에서는 각 속성의 각 학습단계에서의 변화량을 각 속성의 최대 변화량으로 나눈 비율을 양자화 하여 사용한다.

• 규칙 생성 :

러프집합 이론에 의한 규칙 생성은 모순된 정보를 어떻게 처리하느냐에 따라 여러 가지 종류의 규칙이 있을 수 있다. 여기에서는 판단테이블의 정보를 최대한 포함할 수 있는 베스트 규칙(best rule)을 사용한다.

• 구조 재구성 :

규칙들의 조건속성에 포함되지 않는 요소를 제거해서 새로운 구조를 구성한다. 이때 사용하는 규칙은 생성된 규칙 중 판단값이 가장 작은 값인 규칙은 제외된다. 가장 작은 판단값은 학습과정 중 출력오차의 변화량이 거의 없는 학습 단계에서 만들어진 데이터들이다. 이러한 단계에서의 자료는 학습을 수행하는데 필요한 연결강도에 대한 정보가 없으므로 무시하며, 이 과정이 불필요한 연결강도를 제거하는 과정이다. 규칙생성단계에서 이러한 데이터들을 제외하지 않은 이유는 이러한 데이터와 같은 조건값 들을 갖는 다른 데이터와의 관계를 규칙생성 과정에서 고려하도록 하기 위해서이다. 그리고 입력이 없는 뉴런은 다음 층의 바이어스 입력과 같은 역할을 하므로 제거하며, 이 과정이 불필요한 뉴런을 제거하는 과정이다.

• 재학습 :

재구성 된 구조는 불필요한 요소가 제거되었다. 이로 인해서 각 입력 패턴의 초평면에 의한 분할은 뉴런의 출력함수에 따라 변화하게 된다. 뉴런의 출력함수를 시그모이드 함수를 사용한 경우 뉴런의 출력은 입력패턴이 첫 번째 은닉층에서 정해진 초평면에 의한 분할 영역에 포함되는 가중치로 나타나게 되므로 불필요한 신경망의 제거로 인해서 가중치는 변하게 된다. 이 변화는 규칙생성을 담당하는 두 번째 은닉층에 영향을 주게되며, 두 번째 은닉층에서도 첫 번째 은닉층에서 발생한 변화가 발생한다. 이를 보완하기 위해서 재학습을 수행하게된다. 재학습에 실패 한 경우는 불필요한 요소를 찾는 과정에서 생기는 오류로 인해서 요소의 불필요성을 잘못 판단했기 때문이다.

• 구조 비교 :

재학습에 성공한 경우에는 재구성한 신경망의 구조와 재구성 이전의 신경망의 구조를 비교함으로서 재구성된 구조가 최적의 구조인지를 확인한다. 재구성된 구조가 재구성 이전의 구조와 같은 경우는 최적구조를 찾은 경우이므로 이

구조는 최적구조가 된다. 그렇지 않은 경우는 다시 불필요한 요소를 찾는 과정을 반복한다.

• 요소의 삽입 :

재구성된 신경망의 구조가 재학습에 실패했을 때 제거했던 요소를 연결하고 학습과 불필요한 요소를 찾는 과정을 반복 수행한다. 이 방법을 통해서 불필요한 요소로 잘못 판단된 요소를 신경망의 구조에 복귀시킬 수 있다.

5. 실험결과

제안한 방법의 유효성을 검증하기 위해서 본 연구에서는 XOR 연산의 예를 든다. XOR 문제는 입력과 출력변수가 적고 신경망의 해석이 용이하며, 이미 최적구조가 알려져 있고 때문에 제안한 방법의 유효성을 검증할 수 있는 좋은 예제이기 때문이다. 이 문제를 해결하기 위한 초기 신경망은 그림 2와 같이 2개의 입력뉴런, 1개의 출력뉴런, 입력패턴의 개수와 같은 4개의 뉴런을 갖는 1개의 은닉층으로 구성된 다층 신경망이다. XOR문제는 출력 패턴이 두 가지뿐이므로 출력층의 뉴런은 하나만으로 충분하다. 따라서 규칙의 합성이 필요치 않기 때문에 두 번째 은닉층을 생략한다. 또 입력 뉴런을 제외한 모든 뉴런은 바이어스 입력을 가지며 뉴런의 출력함수는 시그모이드 함수를 사용한다. 초기의 연결강도는 ± 0.01 범위 내에서 임의의 값을 사용하고 학습 방법은 오차역전파 학습법을 사용한다.

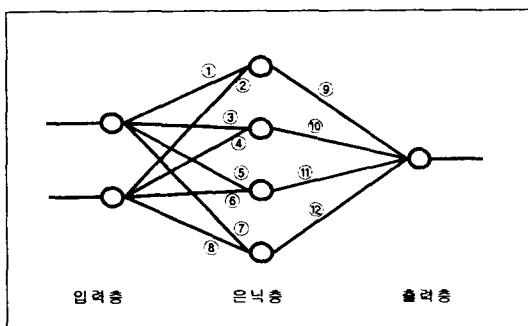


그림 2 XOR 문제를 위한 신경망의 초기 구조
Fig. 2 Initial structure of neural network for XOR problem

판단테이블을 만들기 위한 양자화는 각 연결강도와 출력오차의 변화값을 각각의 최대값으로 나눈 값을 사용해서 균등하게 4등분하였다. 표 2는 전체 학습구간 동안의 연결강도 변화량으로 만든 판단테이블이며, 표 3은 규칙생성에 의해서 표 2로부터 생성된 규칙을 나타내었다. 규칙의 판단속성값이 1인 규칙을 제외한 나머지 규칙들로부터 규칙을 구성하는 조건속성들의 집합을 W_{opt} 라 하면 $W_{opt} = \{W_3, W_4, W_7, W_8, W_9, W_{10}, W_{12}\}$ 가 된다.

그림 3은 XOR 문제를 해결하기 위한 신경망의 최적화 구조이다. 러프집합의 적용 결과 얻어진 조건속성들의 집합 W_{opt} 의 연결강도로 신경망을 구성할 경우 연결강도 W_9 (그림 3의 점선)가 구조에 포함되지만 W_9 의 입력측 뉴런에는 어떠한 입력도 가해지지 않기 때문에 무시될 수 있다. 결과

표 2 XOR 학습결과에 의한 판단테이블

Table 2 Decision Table made by XOR learning result

학습 단계	조건속성 (연결강도 변화량)											판단속성 (출력오차 변화량)
	S	W1	W2	W3	W4	W5	...	W10	W11	W12	E	
1	4	4	4	2	4	...	4	4	4	4	4	4
2	4	4	4	2	3	...	4	4	4	4	3	3
3	4	4	4	2	3	...	4	4	4	4	2	2
4	4	4	4	2	3	...	4	4	4	4	2	2
:	:	:	:	:	:	...	:	:	:	:	:	:
364	1	1	3	4	2	...	1	1	3	4	4	4
365	1	1	3	4	1	...	1	1	2	4	4	4
366	1	1	3	4	1	...	1	1	2	4	4	4
367	1	1	2	4	1	...	1	1	2	3	3	3
368	1	1	2	4	1	...	1	1	2	3	3	3
:	:	:	:	:	:	...	:	:	:	:	:	:
594	1	1	1	2	1	...	1	1	1	1	1	1
595	1	1	1	1	1	...	1	1	1	1	1	1

표 3 생성한 규칙

(Wn : 연결강도의 변화량, E : 출력오차의 변화량)

Table 3 Generated rule

(Wn : variation of weight, E : variation of output error)

규칙	Wn	E	규칙	Wn	E
R1	W4=3,W8=2	4	R14	W3=3,W9=3	2
R2	W4=4,W8=2	4	R15	W3=4,W4=1	2
R3	W4=4,W12=4	4	R16	W4=3,W7=1	2
R4	W9=2,W12=2	4	R17	W3=2,W12=1	2
R5	W4=2,W12=4	3	R18	W4=2,W12=3	1
R6	W4=3,W9=3	3	R19	W3=1,W6=1	1
R7	W4=4,W12=2	3	R20	W1=3,W10=3	1
R8	W4=3,W12=2	3	R21	W5=3,W12=1	1
R9	W9=2,W12=4	2	R22	W1=1,W6=2	1
R10	W9=1,W10=4	2	R23	W8=2,W12=2	1
R11	W8=4,W10=3	2	R24	W3=2,W12=3	1
R12	W8=4,W9=1	2	R25	W4=2,W12=1	1
R13	W3=2,W12=3	2			

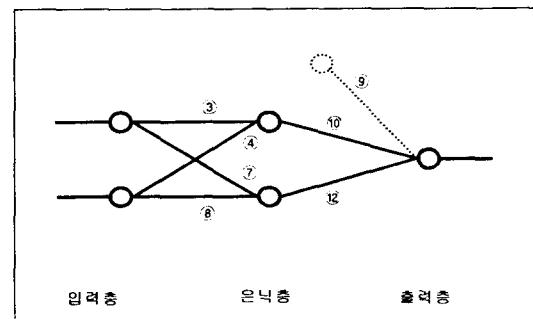


그림 3 XOR 문제를 위한 신경망의 최적화 구조

Fig. 3 Optimal structure of neural network for XOR problem

직으로 본 연구에서 얻어진 최적화 후의 신경망 구조는 잘 알려진 XOR 연산의 최적구조와 동일함을 알 수 있다. 한편 그림 4는 신경망의 초평면을 나타낸다. 그림 4(a)는 초기 신경망의 초평면으로 어떠한 초평면도 패턴의 경계를 나누지 못하기 때문에 패턴분류에 관여하는 초평면은 없다. 그림 4(b)는 학습에 성공한 후의 초평면을 나타낸다. 3개의 초평면이 하나의 패턴경계에 모여있음을 볼 수 있다. 이 3개의 초평면 중 2개는 불필요한 초평면으로 이에 해당하는 첫 번째 은닉층의 두 뉴런은 불필요한 신경망 요소이다. XOR 연산과 같이 입력변수가 적은 간단한 문제에서는 계산과 그래프 해석에 의해서 불필요한 요소들을 간단히 찾을 수도 있지만, 입력변수가 증가함에 따라서 다차원 문제가 되기 때문에 계산과 그래프 해석으로는 불필요한 요소를 찾기 어렵다. 그러나 본 논문에서 제안한 방법은 이러한 문제를 간단히 해결할 수 있다. 그림 4(c)는 최적화 이후의 초평면을 나타낸다. 2개의 초평면이 서로 다른 패턴의 경계를 이루고 있음을 볼 수 있다.

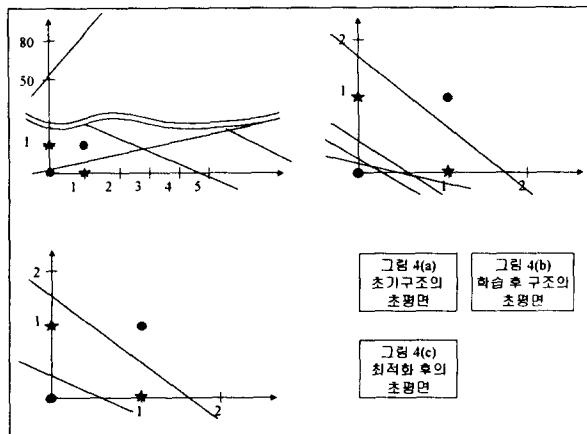


그림 4 각 단계에서의 신경망의 초평면 위치
 (a) 학습 전, (b) 학습 후, (c) 최적화 구조
 Fig. 4 Position of Hyperplanes in each step
 (a) Before learning, (b) After learning,
 (c) Optimal structure

참고적으로 그림 5와 6는 각 학습단계에서의 연결강도의 일부분(W1, W3, W6, W8, W12)과 출력오차와의 관계를 나타낸 그림이다. 그림 5(b)로부터 알 수 있는 바와 같이 출력오차가 학습에 수반하여 급격히 감소하는 구간이 세 군데 존재한다. 첫 번째 구간은 모든 연결강도가 초평면과 규칙들을 해에 가까운 공간으로 이동시키는 것을 의미한다. 두 번째 구간에서는 그림의 쇄선으로 표시된 연결강도 W8이 급격히 변화함에 따른 결과이다. 이는 연결강도 W8과 출력오차 사이에 중요한 상호작용이 일어나고 있음을 의미한다. 세 번째 구간에서는 그림의 굵은 선으로 표시된 연결강도 W3이 급격히 변화함에 따른 결과이다. 이 구간 역시 연결강도 W3과 출력오차 사이에 중요한 상호작용이 일어나고 있음을 의미한다. 본 논문에서는 이러한 출력 오차와 연결강도의 관계를 러프집합으로 분석함으로써 학습에 관계되는 연결강도만을 찾아내어 최적의 신경망 구조(그림 3)을 구성할 수 있음을 보인 것이다.

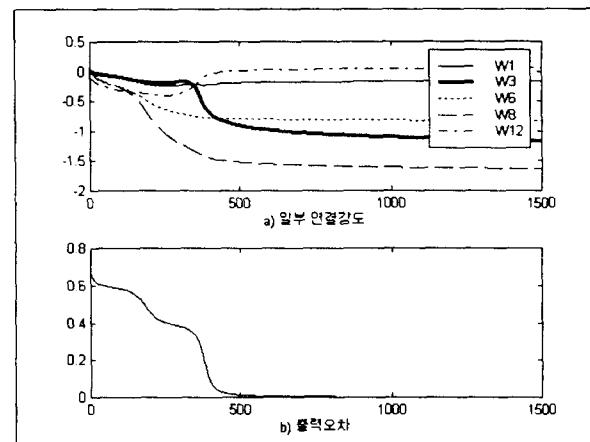


그림 5 학습단계에서의 연결강도와 출력오차
 Fig. 5 Output error and weight in learning step

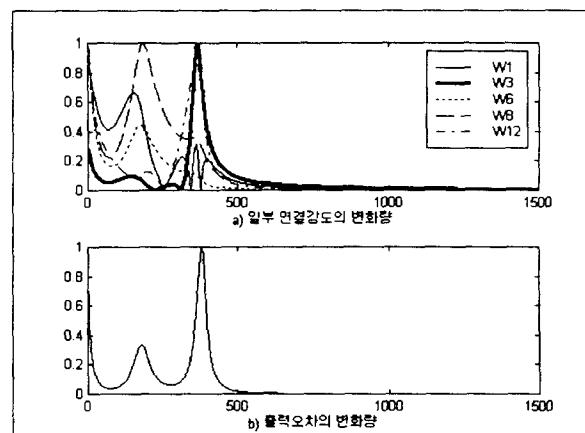


그림 6 학습단계에서의 연결강도와 출력오차의 변화량
 Fig. 6 Variation of Output error and weight in learning step

6. 결론

본 논문에서는 초평면에 기반한 최적 구조의 신경망을 얻기 위해서 러프집합 이론을 이용한 새로운 정보처리방법을 제안했다. 본 연구에서 제안한 신경망의 구조최적화 방법은 오차역전파 학습과정에서 관측한 신경망의 모든 연결강도와 출력오차의 변화량을 이용하여 조건속성과 판단속성으로 이루어진 판단테이블을 만들고 이 테이블로부터 러프집합 이론을 적용하여 분석함으로써 학습에 영향을 미치지 않는 연결강도와 뉴런을 제거하는 방법이다. 이 방법은 오차역전파 학습을 통하여 형성되는 네트워크 구조를 보다 명확하게 할 수 있을 뿐만 아니라, 학습과정에서 최적인 신경망의 구조를 효율적으로 발견할 수 있는 특징을 가지고 있다. 제안한 방법에 의하여 얻어진 신경망의 구조가 최적인지 아닌지를 확인하기 위하여 본 논문에서는 이미 최적구조를 알고 있는 XOR 연산 문제에 적용하여 그 유효성을 확인했다.

감사의 글

본 연구는 한국과학재단 특정기초 연구비 지원 (96-01-02-13-01-3)으로 수행되었으며 지원에 감사드립니다.

참 고 문 헌

- [1] Jacek M. Zurada., Introduction to Artificial Neural Systems, West Publishing Co., 1992.
- [2] 有澤正樹, 和多田淳三, ファジィ推論を用いたニューラルネットワークの構造學習アルゴリズム, 計測自動制御學會論文集, Vol.33, No.11, pp. 1087-1092, 1997.
- [3] G. F. Miller., P. M. Todd., and S. U. Hegde, Designing neural networks using genetic algorithms, In Proc. of the third Int'l Conf. on Genetic Algorithms and Their Applications, San Matco, CA, 1989.
- [4] S. A. Harp., T. Samad., and A. Guba, Toward the genetic synthesis of neural networks, In Proc. of the third Int'l Conf. on Genetic Algorithms and Their Applications, 1989.
- [5] X. Yao, The evolution of connectionist networks, Artificial Intelligence and Creativity, Kluwer Academic Publishers, 1994.
- [6] Pawlak. Z., Rough Sets, International Journal of Computer and Information Science, 11, pp. 341-356, 1982.
- [7] Pawlak. Z., Rough Sets : Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, 1991.

저 자 소 개



정 영 준(丁榮俊)

1973년 5월 20일 생. 1998년 중앙대 제어계측공학과 졸업. 1998년~현재 동 대학교 제어계측학과 석사과정



전 효 봉(全孝炳)

1969년 2월 23일 생. 1997년 중앙대 제어계측공학과 졸업. 1997년~현재 동 대학교 제어계측학과 석사과정



심 귀 보(沈貴寶)

1956년 9월 20일 생. 1984년 중앙대 전자공학과 졸업(공학사). 1986년 동 대학교 전자공학과 대학원 졸업(공학석사). 1990년 동경대학 전자공학과 졸업(공학박사). 1990년 동경대학 생산기술연구소 연구원. 1991년~1994년 중앙대학교 제어계측공학과 조교수. 1995년~현재 중앙대학 전자전기공학부 부교수. 1992년~현재 대한전자공학회 인공지능·신경망 및 퍼지시스템 연구회 협동전문위원. 1995년~현재 대한전자공학회 논문편집위원. 1998~현재 한국퍼지 및 지능시스템학회 논문 편집위원 및 이사. 1999년~현재 한국뇌학회 학술위원