

러프셋 이론을 이용한 신경망의 구조 최적화

Structure Optimization of Neural Networks using Rough Set Theory

정 영 준*, 이 동 욱, 심 귀 보

로보틱스 및 지능정보시스템 연구실
중앙대학교 공과대학 제어계측공학과

Tel: 02)820-5319, Fax: 02)817-0553, E-mail:kbsim@cau.ac.kr

Young-June Chung, Dong-Wook Lee, and Kwee-Bo Sim

Robotics and Intelligent Information System Laboratory
Dept. of Control and Instrumentation Engineering, Chung-Ang University
221, Huksuk-Dong, Dongjak-Ku, Seoul 156-756, Korea
Tel:+82-2-820-5319, Fax:+82-2817-0553,
E-mail:kbsim@cau.ac.kr, URL : http://rics.cie.cau.ac.kr

ABSTRACT

Neural Network has good performance in pattern classification, control and many other fields by learning ability. However, there is no effective rule or systematic approach to determine optimal structure. In this paper, we propose a new method to find optimal structure of feed-forward multi-layer neural network as a kind of pruning method. That is eliminating redundant elements of neural network. To find redundant elements we analysis error and weight changing with Rough Set Theory, in the condition of executing back-propagation learning algorithm.

I. 서론

주어진 문제에 대한 신경망의 최적 구조를 찾는 문제는 신경망 자체에 대한 확실한 해석을 내리기 어려운 점 때문에 일반화된 규칙을 찾기 힘들다. 어떤 주어진 문제에 대한 신경회로망의 구조 설계는 주로 설계자의 경험과 직관에 의존해서 구성하고 있다. 신경망을 초평면 기반의 대응(hyperplane - based mapping)으로 볼 때 최소의 초평면과 최소의 판단규칙(decision rule)으로 구성되는 신경망 구조가 최적의 신경망 구조가 된다.

학습을 수행하기 이전의 신경망 구조는 다음과 같이 구분 할 수 있다.

- 1) 초기 네트워크 구조가 최적의 구조보다 작다.
 - a. 주어진 문제가 필요로 하는 최소한의 대응을 할 수 없다.
 - b. 모든 네트워크의 요소(element)들이 빠짐없이

필요로 하다.

- 2) 초기 네트워크 구조가 최적의 구조보다 크다.

- a. 학습에 따라서 주어진 문제가 필요로 하는 최소한의 대응을 할 수 있다.
- b. 대응에 관여하지 않는 요소는 불필요한 요소이다.
- c. 불필요한 요소를 제거해서 최적의 구조를 만든다.
- d. 시뮬레이션 또는 하드웨어 구현에서의 자원(resource)낭비, 계산시간의 증가, 또한 무엇보다도 신경망의 구조를 해석하기가 더욱 어려워지는 단점이 있다.

본 논문에서는 초기 네트워크 구조가 최적의 구조보다 큰 구조에서 시작해서 러프셋(rough set)을 이용해서 불필요한 요소를 찾아 제거함으로써 최적의 네트워크 구조를 찾는 일종의 차원화(pruning) 방법에 대한 내용이다.

II. Information System & Rough Set

정보시스템(Information System) I 는 개체(object)로 이루어진 유한 전체집합 U , 속성(Attribute)으로 이루어진 유한집합 A 에 의한 관계로 $I = (U, A)$ 이다. 즉, U 의 원소인 개체가 A 의 원소인 a 의 어떤 속성 값과 연결되는 관계를 말한다^{[1][2]}. 전건부와 후건부로 나뉘는 판단 표(decision table)도 이러한 정보시스템의 일부이다. 판단표의 어떤 두 개체가 전건부는 동일하고 후건부가 다른 경우, 우리는 이 자료를 가지고 판단을 내릴 수 없는 모순된(inconsistent) 상황에 이르게 된다. 이러한 경우 어떤 측정치(measure)들을 통한 제약을 적용한다면 판단을 내릴 수 있게 된다.

러프셋은 근사화(Aproximation)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 판단 표에서 얻은 규칙인 최소규칙(reduct)은 판단에 관여하지 않는 요소를 제거한 것이며 가장 적은 정보로 전체 시스템을 대표하며 기존 측정치의 범위 안에서 정보손실 없이 모든 정보를 포함하게 된다. 하나의 정보시스템은 여러 개의 최소규칙을 가질 수도 있으며 여러 개의 최소규칙들 중 공통되는 요소들은 핵심규칙(core)으로서 정보시스템을 해석하는데 있어서 없어서는 안 되는 중요한 정보이다.

신경망은 기본적으로 초평면에 의해서 나누어지는 영역에 따른 대응 네트워크로 해석할 수 있다. 단순한 전방향 다층 신경망의 경우 2개의 중간층(hidden layer)만으로 대응 네트워크를 구성할 수 있다. 대응의 관계를 정보시스템의 판단규칙에 대응한다고 볼 수 있고 특히 패턴분류(pattern classification)의 문제에서는 첫 번째 중간층을 초평면에 상응하는 요소로, 두 번째 중간층을 판단규칙에 해당하는 요소로 볼 수 있다. 이 경우 판단규칙은 러프셋 또는 통계적 방법에 의해서 정할 수 있지만 초평면은 시행착오를 거쳐서 구할 수밖에 없다^{[1][2]}.

III. 신경망의 구조 최적화를 위한 Rough Set의 적용방법

3.1. 에러변화와 연결강도변화의 관계

오차 역전파 알고리즘(Back Propagation Algorithm)은 기본적으로 최급강하법(steepest descent method)을 사용하고 있다. 이 방법은 에러를 줄이는 방향으로 연결강도(weight)를 변화시키게 된다^{[3][4]} 어떤 학습단계(step)에서 에러와 연결강도의 관계는 다음의 표1과 같다.

표 1. 한 학습단계에서 에러변화와 연결강도의 관계

case	에러변화	weight변화	mapping에 관여여부
1-1	변화	변화	관여>불관여
1-2	변화	불변	관여?불관여
1-3	불변	변화	관여<불관여
1-4	불변	불변	관여?불관여

case1-1) 에러가 변화했을 때 어떤 연결강도가 변했다면 그 연결강도는 에러를 변화시키는 방향으로 학습을 한 것이며 결국 대응에 영향을 준 것이다. 간섭현상(cross-talk)을 감안하면 대응에 관여하지 않을 수도 있지만 대응에 관여할 가능성이 더 크다.

case1-2) 에러가 감소했는데도 연결강도가 변하지 않았다면 그 연결강도는 이미 대응이 끝났거나 대응에 관여하지 않는 연결강도이다. 그러므로 현 학습단계에서는 연결강도의 대응에 대한 관여 여부를 알 수 없다.

case1-3) 에러가 변하지 않았는데도 연결강도가 변했다면 그 연결강도는 대응에 관여하지 못하는 연결강도이거나 연결강도의 변화에 의해서 초평면이 분할하는 영역이 그 전의 초평면 영역과 같은 대응을 하는 경우이다. 이 경우 연결강도가 대응에 관여하지 않을 가능성이 크다.

case1-4) 에러가 변하지 않았고 연결강도도 변하지 않았다면 그 연결강도는 대응에 관여하는지 그렇지 않은지 알 수 없다.

이처럼 한 학습단계에서의 에러와 연결강도의 관계를 정할 수는 있지만 이 정보만으로 어떤 연결강도가 불필요한 요소인지 알 수 없게 된다. 하지만 학습 과정의 에러와 연결강도의 변화를 계속해서 관찰하게 되면 표2, 표3과 같은 규칙을 찾을 수 있다.

표 2. 에러와 연결강도의 연속적 관계 I

case	에러 변화구간에서 weight의 변화	해석
2-1	모든 구간에서 변화	모든 분류경계에 대한 학습에 관여함
2-2	일부 구간에서는 변화	몇몇 분류경계에 대한 학습에 관여함
2-3	모든 구간에서 변하지 않음	신경망의 mapping에 관여하지 않거나 초기에 이미 패턴분류의 초평면을 형성함

에러의 변화구간은 그 구간에서의 전체 또는 일부 연결강도의 변화로 인해서 초평면 또는 초평면의 조합이 변화하고 변화된 초평면 또는

초평면의 조합에 의한 분류 공간이 어떤 패턴들의 분류경계를 넘나들거나 초평면의 조합이 변하는 구간이며, 에러 불변구간은 연결강도의 변화가 없어서 초평면 또는 초평면 조합이 변하지 않고 따라서 패턴분류의 경계를 넘나드는 초평면이 없거나, 연결강도가 변해서 초평면이 변하지만 패턴의 경계영역 안쪽에서만 변하기 때문에 에러의 변화가 없는 구간이다.

case2-1) 어떤 연결강도가 에러의 모든 변화구간에서 모두 변한다면 이 연결강도는 패턴의 경계를 모두 거치므로 모든 분류경계에 대한 학습에 관여한다고 본다.

case2-2) 어떤 연결강도가 몇몇 에러의 변화구간에서 변한다면 연결강도가 변하는 구간에서는 패턴의 경계를 거치므로 학습에 관여한다.

case2-3) 어떤 연결강도가 모든 에러변화구간에서 변하지 않는다면 이 연결강도는 패턴의 경계와 무관하거나 초기값이 이미 패턴의 경계를 구분하는 연결강도이다. 하지만 초기에 초평면 값이 패턴분류의 경계를 이룬다 해도 초평면의 조합에 의해서 판단을 내리는 두 번째 중간층의 연결강도가 정확히 주어진 문제에 일치할 가능성이 없으므로 이 경우는 초평면에 상응하는 연결강도가 재학습하게 된다.

표 3. 에러와 연결강도의 연속적 관계 II

case	에러 불변구간에서 weight의 변화	해석
3-1	변화	패턴의 classification에는 영향을 주지 않지만 hyperplane, 또는 decision rule은 변화
3-2	변하지 않음	신경망에 영향을 주지 않음

case3-1) 연결강도가 변하기는 하지만 그 변화가 패턴분류 경계를 넘어가지 못하거나 초평면의 합성을 변화시키지 못한다.

case3-2) 연결강도가 변하지 않으므로 초평면이나 초평면의 합성이 바뀌지 않는다. 따라서 에러도 변하지 않는다.

표2, 표3으로부터 표4와 같은 결과를 얻는다.

3.2. 에러 변화량과 연결강도 변화량의 관계

위에서 보았듯이 에러와 연결강도의 연속적 관계에 의해 신경망에 불필요한 요소를 찾을 수 있다. 하지만 신경망에는 여러 개의 연결강도가 있고 그 연결강도들이 독립적으로 변하지 않고 에러의 변화에 의해서 모든 연결강도가 영

표 4. 에러와 연결강도의 연속적 관계

case		신경망에의 필요성
2-1	3-1	필요
	3-2	필요
2-2	3-1	필요
	3-2	필요
2-3	3-1	불필요
	3-2	불필요

향을 받기 때문에 하나의 학습과정에서 에러가 변할 때 초평면이 변해서 패턴분류영역이 바뀌거나 초평면의 조합이 바뀌는 학습에 중요한 역할을 하는 연결강도의 변화뿐 만 아니라 학습에 영향을 주지 않는 연결강도도 패턴분류영역이 바뀌지 않는 초평면으로의 또는 초평면의 조합이 바뀌지 않는 조합으로의 연결강도 변화가 있을 수 있다. 이러한 현상은 표4에 의한 판단만으로 신경망에 불필요한 요소를 찾는 데 한계가 있음을 나타낸다. 이 경우 가장 문제가 되는 연결강도는 신경망에 불필요한 연결강도임에도 표4의 2-3:3-1, 2-3:3-2 case에 포함되지 않는 즉, 에러 변화구간에서 연결강도가 변하는 경우이다. 이로부터 표4의 나머지 경우에 해당하는 관계가 결론을 내릴 수 없는 모호한 상황에 이르게 된다. 이러한 상황을 해결하기 위해서 러프셋 이론을 에러변화와 연결강도 변화의 분석에 도입한다.

한 학습 단계를 정보 시스템의 개체로 보고 각 연결강도의 변화량과 에러 변화량을 속성으로 보았을 때 초기 신경망 구조에서부터 원하는 성능 또는 에러 허용범위까지 학습하는 과정을 정보시스템으로 만들 수 있다. 이 정보시스템으로부터 신경망의 최적구조를 찾는 방법은 다음과 같다.

- ① 연결강도 변화 속성들을 전건부로, 에러 변화 속성을 후건부로 하는 판단 표를 작성한다. 모호성 판단을 위해서 에러변화가 없는 개체도 판단 표에 넣는다.
- ② 판단 표를 러프셋으로 분석하여 최소규칙을 얻는다.
- ③ 구해진 최소 규칙 중 후건부가 에러 변화 없음인 규칙은 신경망의 학습에 영향을 주지 않는 규칙이므로 제거한다.
- ④ 이렇게 해서 얻은 규칙에 포함되어있는 연결강도는 신경망의 최적구조를 이루는 요소이므로 이 요소들을 제외한 나머지 요소는 제거한다.

러프셋으로 최소규칙을 만들 때 위에서 제시한 표4의 case들은 표5와 같이 처리되어야 한다.

표 5. 최소규칙 생성시 에러와 연결강도의 변화에 따라서 원하는 처리

case		원하는 처리
2-1	3-1	최소규칙에 포함
	3-2	최소규칙에 포함
2-2	3-1	최소규칙에 포함
	3-2	최소규칙에 포함
2-3	3-1	최소규칙에서 제외
	3-2	최소규칙에서 제외

다음은 러프셋을 사용해서 위에서 제시한 방법을 수행했을 때 표5와 같은 결과가 나타나는지를 설명하고 있다.

case2-1:3-1)에러와 연결강도의 변화관계가 일치하므로 핵심규칙으로 최소규칙에 존재한다.

case2-1:3-2)전 구간에서 연결강도가 변하기 때문에 정보로서의 가치가 없으며 최소규칙에 포함되지 않는다. 이것은 표5에서의 원하는 처리가 아니다. 다음은 이런 경우의 연결강도가 나타날 수 있는지에 대한 분석이다.

- ① 연결강도가 초평면에 해당할 경우 : 학습된 연결강도는 하나의 초평면으로서 패턴의 경계를 나타낸다. 하나의 초평면은 모든 패턴을 분류할 수는 없다. 분할 할 수 있는 모든 패턴을 분류한 후에 다시 연결강도가 변하게 되면 에러가 증가하게되므로 오차 역전과 알고리즘이 내포하고있는 최급강하법의 방법으로 나타날 수 없다.
- ② 연결강도가 초평면의 조합에 해당할 경우 : 패턴경계를 나타내는 여러 초평면의 조합을 나타내는 연결강도는 학습이 수행된 후 다시 변하게 되면 에러가 증가하게되므로 오차 역전과 알고리즘이 내포하고있는 최급강하법의 방법으로는 나타날 수 없다.

위에서처럼 이러한 연결강도는 오차 역전과 알고리즘으로 학습을 수행하는 경우 나타날 수 없다.

case2-2:3-1)러프셋의 분석에 의해 모호성이 없어지고 규칙에 포함된다.

case2-2:3-2)러프셋의 분석에 의해 모호성이 없어지고 규칙에 포함된다.

case2-3:3-1)에러 변화구간에서 항상 연결강도가 변하지 않기 때문에 러프셋에 의한 최소규칙의 후건부가 변화 있음인 규칙에 포함될 수 없다.

case2-3:3-2)전 구간에서 연결강도가 변하지 않으므로 정보로서의 가치가 없으며 러프셋에 의한 최소규칙의 어느 규칙에도 포함되지 않는다.

이와 같이 러프셋에 의한 판단표의 분석은 신경망의 최적구조를 찾는 방법으로 쓰일 수 있다.

IV. 결론

에러 역전과 학습의 최급강하 성질, 연결강도 변화와 에러변화의 관계, 러프셋의 정보분석방법을 통해서 주어진 문제에 대한 최적의 신경망 구조를 찾을 수 있다. 하지만 아래와 같은 단점과 앞으로 연구되어야 할 부분이 남아있다.

본 논문의 방법은 신경망의 학습에 관여하지 않는 요소들을 찾아서 제거하는 방법이기 때문에 원하는 성능의 최적 신경망을 얻기 위해서는 적어도 원하는 성능에 도달할 때까지 학습을 마쳐야한다. 원하는 성능에 도달하기 전에 적용하게 되면 학습이 수행된 이상의 성능을 구현할 수 없으며, 매 학습마다의 에러와 연결강도의 정보를 분석에 이용하므로 원하는 성능에 도달할 때까지 학습한 정보를 모두 저장해야하므로 많은 데이터 저장공간이 필요하다.

자료분석에 있어서 러프셋의 자료분석 방법은 연속적인 자료를 다루기에는 부적합하다. 신경망의 학습과정에서 취할 수 있는 에러와 연결강도의 변화량은 연속적인 실수자료로서 러프셋에 적용하기 위한 자료의 정규화 및 양자화 과정이 필요하다. 현재 연결강도 (변화량/최대 연결강도 변화량), (연결강도 변화량/에러 변화량)의 정규화 방법 등이 연구되고 있다.

감사의 글

본 연구는 한국과학재단 특정기초 연구비(96-01-02-13-01-3) 지원으로 수행되었으며 지원에 감사를 드립니다.

참고문헌

- [1] Marcin S. Szczuka, Rough Sets and Artificial Neural Networks, 1998
- [2] Nguyen Hung Son, Marcin S. Szczuka, Dominik Slezak, Neural Networks Design : Rough Set Approach to Continuous Data, 1998
- [3] James A. Freeman et al. Neural Networks : Algorithms, Applications, and Programming Techniques, Addison-Wesley Publishing Company, 1991
- [4] Martin Y. Hagan, Howard B. Demuth, Mark Beale, Neural Network Design, PWS Publishing Company, 1996