

러프 집합 이론을 이용한 뉴로-퍼지 모델의 최적화 A Neuro-Fuzzy Model Optimization Using Rough Set Theory

연정흠 · 서재용 · 김용택 · 조현찬 · 전흥태

Jung-Heum Yon, Jae-Yong Seo, Yong-Taek Kim, Hyun-Chan Cho and Hong-Tae Jeon

중앙대학교 전자전기공학부

요 약

본 논문에서는 플랜트를 위한 규칙수가 줄어든 뉴로-퍼지 모델을 얻기 위한 접근을 제안한다. 뉴로-퍼지 네트워크는 가우시안 소속함수를 가진 RBF(Radial Basis Function) 네트워크들로 구성되고 오차 역전파 학습 알고리즘에 의해 학습된다. 러프 집합 이론에서 의존도는 규칙들을 줄이는데 사용된다. 모델에서 각 규칙의 조건 소속함수 값과 플랜트의 출력 값 사이의 의존도는 플랜트를 동정하기 위하여 규칙이 얼마나 많은 공헌을 하는가를 알 수 있도록 한다. 줄어든 모델은 원래의 것으로써 동일한 성능을 유지하는 동안 선택 알고리즘은 복잡성과 구조의 잉여성을 최소화할 수 있다.

ABSTRACT

This paper presents an approach to obtain a reduced neuro-fuzzy model for a plant. The Neuro-Fuzzy Networks are composed of the Radial Basis Function Networks with Gaussian membership and learned by using temporal back propagation. The dependency in rough set theory is used to eliminate rules. Dependency between the condition membership value of each rule in a model and the output of the plant can allow us to see how much contributive the rule is to identify the plant. While the reduced model maintains the same performance as the original one, the selection algorithm can minimize its complexity and redundancy of the structure.

1. 서 론

퍼지 로직은 인간의 추론능력을 이용하여 복잡한 비선형 시스템을 정량적(quantitative)분석을 하는 대신 정성적(qualitative)분석을 할 수 있다. 그러나 이러한 시스템에 대한 인간의 지식과 경험을 충분히 얻기가 어렵기 때문에 학습과 적응 능력이 있는 신경망 구조와 융합하여 시스템의 특성에 더 적합한 모델을 얻으려고 하였다. 이러한 시스템은 일반적으로 두 가지 장점을 보여준다. 첫째는 퍼지 로직의 특징인 시스템의 변수들이 물리적 의미를 갖는다. 둘째는 신경망을 사용함으로써 자동적으로 퍼지 로직 규칙들을 학습하여 퍼지 소속함수들을 조정할 수 있다.

그러나, 뉴로-퍼지 시스템은 퍼지 추론과 신경망의 학습을 융합했음에도 사용된 망구조의 잉여성(redundance)과 학습조건들에 대한 일반적인 원칙을 구하기가 어렵고 초기조건에 따라서 전역 최소점(global minimum)에 도달하지 않을 수도 있다. 이러한 단점을 극복하기 위해 뉴로-퍼지 시스템의 망구조를 최소화하기 위해 임의의 초기값 및 뉴로-퍼지망의 구조로부터 시스템의 잉여용량(redundent capacity)을

줄여가는 방법이 제안되기도 했다. 본 논문에서는 임의의 초기조건 및 Back Propagation 알고리즘[2]에 의해 학습된 망구조로부터 뉴로-퍼지 시스템의 잉여성을 줄이는데, 러프 집합(rough set) 이론의 의존도(dependency)를 사용하여 줄여야할 규칙을 결정한다. 모의 실험을 통하여 초기 분할 및 ϵ -충족 조건이 불필요하며 초기 규칙수가 충분히 크면 초기조건의 제약이 훨씬 적어짐을 확인할 수 있었다.

본 논문의 구성은 다음과 같다. 2장에서는 러프 집합 이론을 소개하며 의존도에 대한 의미를 살펴본다. 3장에서는 뉴로-퍼지 시스템의 구조, 방사함수 네트워크의 간단한 소개와 서로의 연관성, 그리고 학습알고리즘과 의존도에 의한 규칙수를 줄이는 알고리즘에 관하여 살펴본다. 4장은 초기 규칙수가 다른 3번의 모의 실험을 행한 결과 및 검토를 했으며, 다른 방법들과의 성능을 비교하였다. 마지막으로 5장은 본 논문의 결론을 기술한다.

2. 러프 집합 이론

두 개의 집합, 논의영역 U 와 속성집합 A 를 가정하

본 연구는 과학기술부의 뇌과학연구 프로그램의 지원 하에 수행되었습니다.

자. 논의영역내의 모든 성분 $x, y \in U, a \in B$ 이고 속성 집합 $B \subset A$ 일 때 $x \text{IND}_B(B)y$ 를 속성집합 B 에 대해 x, y 는 서로 불구분 관계(indiscernibility relation) 혹은 동가관계(equivalence relation)라고 한다.

$x \text{IND}_B(B)y$ iff 모든 a 에 대해 $a(x) \Rightarrow a(y) \Rightarrow x$ 와 y 는 B -불구분관계.

따라서 논의영역 U 는 임의의 속성 B 에 대한 불구분 관계에 의하여 분류될 수 있다. B -불구분 관계인 모든 성분들로 이루어진 집합을 동가 클래스(equivalence class)라고 하고, 모든 동가 클래스들의 집합을 B -등가 클래스라고 하며 $U/\text{IND}(B)$ 로 표시한다.

$$\text{논의영역 } U \xrightarrow[\text{불구분관계}]{B\text{속성값}} U/\text{IND}_B(B) : B\text{-등가 클래스} \quad (1)$$

이때 임의의 부분집합 $X(C \cup U)$ 에 대하여 상한 근사와 하한 근사는

$$B_*(X) = \{x \in U \mid \text{IND}_B(x) \subseteq X\}, \quad (2)$$

$$B^*(X) = \{x \in U \mid \text{IND}_B(x) \cap X \neq \emptyset\} \quad (3)$$

로 정의한다. 임의의 집합 X 를 논의영역 U 까지 고려했을 때 속성집합 B 에 의해 일관성 있게 정의할 수 있는 최소의 성분집합이 하한 근사이다. 반면에 상한 근사는 집합 X 를 논의영역 U 까지 고려했을 때 일관성이 없는 것까지 포함한 최대의 성분집합을 의미한다.

데이터 분석에서 중요한 또 다른 개념이 속성들 사이의 의존도(dependency)이다. 일반적으로 두 개의 속성집합, C, D 가 있을 때 “ D 가 C 에 의존한다” 함은 D 의 모든 속성값들이 C 의 속성값들에 의하여 유일하게 분류된다는 것을 의미한다. 예를 들어 “속성집합 D 가 속성집합 C 에 완전히 의존한다” 함은 C 에 의해 분류된 C -등가 클래스가 D 에 의해 분류된 D -등가 클래스보다 더 조밀하게 분류되었다는 것을 의미한다.

위의 개념을 체계화하면 다음과 같다. 임의의 두 속성 집합 $C, D(C \subset A)$ 가 있을 때 C 에 대한 D 의 의존도 $k(0 \leq k \leq 1)$ 는

$$k = \frac{\text{card}\{POS_C(D)\}}{\text{card}\{U\}} \quad (4)$$

이며 $k=1$ 일 때는 완전 의존, $k < 1$ 일 때는 부분 의존이라고 한다.

여기서,

$$POS_C(D) = \bigcup_{x \in U/\text{IND}(C)} \{C \cdot (X)\} \quad (5)$$

$POS_C(D)$ 는 속성집합 C 에 대한 D -등가 클래스의 양의 영역(positive region)이라 한다. 즉 양의 영역은 논의영역 U 를 C 에 의하여 D -등가 클래스로 유일하게 분류할 수 있는 모든 성분들의 집합을 의미한다.

3. 뉴로-퍼지 (Neuro-Fuzzy) 구조와 최적화

3.1 뉴로-퍼지 구조

본 논문에서 사용한 뉴로-퍼지 모델은 일반적인 3 가지 구조[1]중 첫 번째를 사용하였으며 그림 1과 같다. 이 구조의 특징으로는 서로 다른 퍼지 규칙들은 고려하지 않기 때문에 입력 공간에 대한 일반적인 퍼지 분할(fuzzy partition)을 할 필요가 없다. 또한 규칙마디(rule node)의 개수가 일반적인 구조에 비하여 상대적으로 적은 반면에 언어마디(term node)의 수는 많다고 할 수 있다. 또한 입력공간의 퍼지분할을 하지 않기 때문에 잉여의 규칙마디를 줄이면 언어마디도 함께 줄일 수 있는 장점이 있다. 반면에 퍼지분할을 하지 않기 때문에 전문가의 지식으로부터 초기조건들에 대한 정보를 곧바로 적용하기가 어렵다는 단점이 있다.

뉴로-퍼지 모델에서 입력 벡터 $x = [x_1, x_2, \dots, x_n]$ 와 출력 벡터 $y = [y_1, y_2, \dots, y_m]$ 가 주어질 때 단일(singleton) 퍼지화, 곱하기 추론과 중간평균 디퍼지화(defuzzification)을 사용한다고 가정하자. 이 퍼지 모델은 4층 피드-포드(feed-forward) 네트워크 구조이다. 그림 1에서 보듯이 첫째 층은 입력층으로 n 개의 입력값이 아무런 계산 과정 없이 바로 다음 층으로 주어진다.

두 번째 층은 퍼지 변수층으로써 입력 벡터 x 에 대한 소속정도를 계산한다. 이 층에는 입력 개수만큼의 n 개의 그룹으로 이루어져 있으며 각 그룹은 m 개의 퍼지 규칙 수만큼의 언어 마디(term node)로 이루어진다.

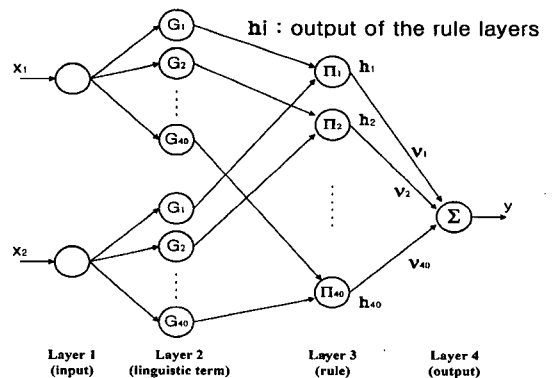


그림 1. 뉴로-퍼지 네트워크의 구조

어진다. 입력 벡터 중에서 i, j 번째 규칙의 마디출력은 가우스(Gaussian) 소속 함수의 출력력으로써

$$\mu_{ij} = \exp(-net_{ij}) \quad (6)$$

이며, $net = -(x_i^2 - \omega_{ij})^2 / (\sigma_{ij}^2)$ 는 네트 입력이다. 여기서 마디 사이의 가중치 ω_{ij} 와 σ_{ij} 는 퍼지 소속함수 μ_{ij} 의 중앙값과 폭이다.

세 번째 층은 곱셈연산에 의한 추론이 이루어지는 규칙층이다. 이 층의 마디수는 퍼지 규칙의 수와 같으며 j 번째 뉴런의 출력은

$$\mu_j(x) = \prod_{i \in L_I} \mu_{ij}(x_i) \quad (7)$$

이다.

네 번째 층은 최종 출력값이 COA(center of area) 방법으로 계산되는 출력층이다. 입력 벡터 x 에 대하여 k 번째 뉴런의 출력은

$$y_k(x) = \frac{\sum_{j \in L_R} \mu_j(x) v_{ik}}{\sum_{j \in L_R} \mu_j(x)} \quad (8)$$

이며 v_{ik} 는 j 번째 규칙 마디와 k 번째 출력 마디를 연결하는 가중치이다. j 번째 퍼지 규칙은

$$\begin{aligned} \text{IF } (x_1 \text{ is } \omega_{1j}) \wedge \dots \wedge (x_n \text{ is } \omega_{nj}) \\ \text{THEN } (y_1 \text{ is } v_{1j}) \wedge \dots \wedge (y_m \text{ is } \omega_{mj}) \end{aligned} \quad (9)$$

이다.

본 논문에서 뉴로-퍼지 모델 구조는 가우스(gaussian) 소속함수를 사용하였기 때문에 함수 근사화 기법중의 하나인 방사함수망(radial basis function networks)[11,12] 구조와 같다.

입력 벡터 x 와 출력 벡터 y 가 주어졌을 때, 함수 f 를 다음과 같은 방사함수를 사용하여 근사시킬 수 있다.

$$f(x) = \sum_{i=1}^m n_i h_i(|x - \omega|) + e \quad (10)$$

i 번째 방사함수의 출력은

$$h_i = \prod_k (G_k(x_k)) \quad (k = 1, 2, \dots, n) \quad (11)$$

이며 $G_k(x_k)$ 는 가우스 소속함수, $v = [v_1, v_2, \dots, v_m]$ 는 상수계수벡터이다.

즉,

$$y = [h_1 v_1 + h_2 v_2 + \dots + h_m v_m] + e \quad (12)$$

로써 m 개의 벡터들의 합으로 벡터 y 를 근사시킨다.

방사함수망에서는 미지의 입출력 함수 관계를 기지

의 함수인 가우스 함수들과 가중치들의 선형연산(Linear Operation)에 의하여 근사시키는 방법이다.

방사함수망 같은 구조를 갖는 뉴로-퍼지 모델에서 가우스 함수의 파라미터인 ω_{ij} , σ_{ij} 와 후건부 가중치이며 선형 파라미터인 v_{ik} 는 Back propagation 알고리즘 [2]에 의하여 학습된다. 이때 에러 벡터 e 가 충분히 작을 때 어떤 방사함수(radial basis function)의 출력 벡터가 y 의 근사에 가장 적은 영향을 미치는지 계산하는 것이 중요한 문제이다. 즉, 뉴로-퍼지 모델의 구조에서 출력에 대한 의존도를 구하기 위하여 고려해야 할 데이터집합은 방사함수망에서의 방사함수들의 출력이다. 기저함수(basis functions)들 중에서 가장 모델링에 영향을 적게 주는 것들을 선택하는 것이다.

3.2 뉴로-퍼지 모델의 최적화

본 논문에서는 임의의 초기 규칙수와 초기 조건들로부터 순차적으로 규칙수를 줄인다. 200번의 학습후 에러값이 특정값 이하로 수렴할 때 의존도의 계산으로부터 규칙을 줄이고 다시 200번의 학습을 하는 방법을 되풀이한다. 의존도가 0/100인 경우는 시스템의 모델링에 전혀 기여하지 못하고 있는 것으로 생각할 수 있기 때문에 모두 제거한다. 만약 의존도가 0/100인 경우가 없을 때는 의존도가 가장 작은 규칙을 제거한다.

전체적인 알고리즘은 다음과 같다.

- 1) 등간격으로 초기화된 뉴로-퍼지 모델을 BP를 사용하여 200번 학습.
- 2) 의존도를 계산하고 작은 값의 규칙마디를 선택하여 삭제.
- 3) 줄여진 모델을 200번 학습.
- 4) RMS 에러값이 0.01이상일 때까지 2)-3)과정을 반복.

전체적인 구성도는 그림 2와 같다. 플랜트를 올라

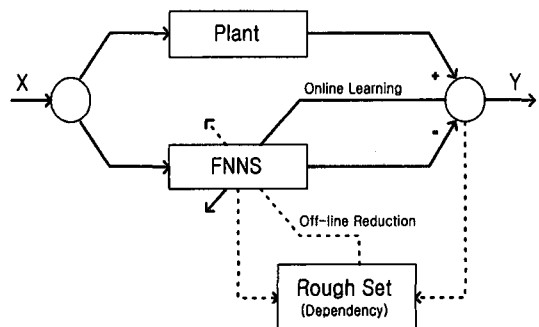


그림 2. 전체 구성도

인 학습에 의하여 200번 학습을 하고 학습 후 에러가 특정 성능지수 범위 내에 있을 경우 오프라인으로 의존도를 계산한다. 의존도에 따라 각 규칙들의 출력에 대한 기여도를 알 수 있으며 기여도가 낮은 규칙을 제거한다.

기여도가 낮은 규칙들은 출력에 미치는 영향이 적기 때문에 학습함에 따라서 더욱더 0으로 수렴하게 된다. BP의 특성상 적당한 규칙수의 변수들만 모델링에 기여하며 나머지는 의존도가 0임을 알 수 있다.

또한, 줄여진 규칙들도 모두 같은 의존도를 나타내지는 않기 때문에 가장 작은 의존도를 갖는 규칙을 줄였을 때 성능에 가장 작은 영향을 준다는 것을 쉽게 생각할 수 있다.

이런 식으로 줄여갈 때 각 규칙들의 의존도를 모두 합한 전체 의존도는 일정한 크기를 유지하게 된다. 이것은 규칙의 수는 줄어들지만 각 규칙들의 의존도가 커지기 때문이다. 결국 에러가 특정 성능지수 범위를 벗어날 때 각 규칙들의 의존도와 전체 의존도는 크게 떨어짐을 짐작할 수 있다.

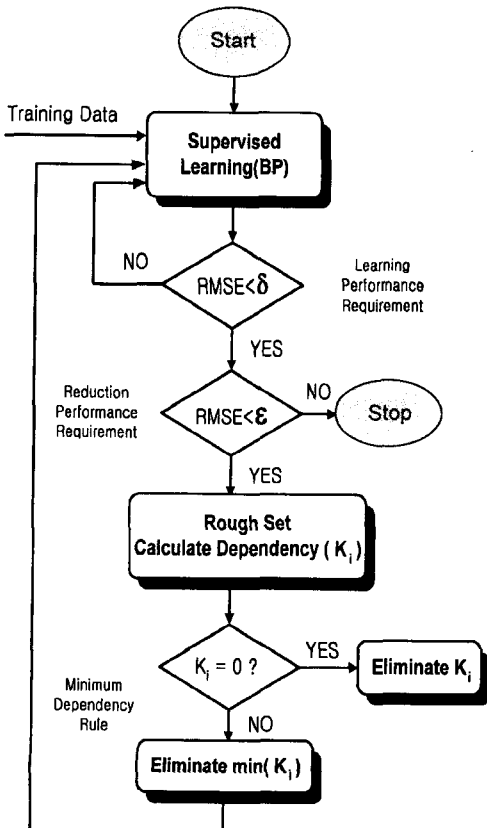


그림 3. 전체 순서도

그림 3에서 δ 는 BP학습의 성능 지수이며 RMS 에러가 δ 이하를 유지할 때까지 학습을 계속한다. ϵ 은 의존도의 성능지수이고, ϵ 보다 RMS 에러가 클 때 규칙수를 줄이는 것을 중지한다.

4. 모의 실험

규칙을 줄이는데 의존도를 사용하는 것의 타당성을 검증하기 위하여 2차 차등방정식[1]으로 표현되는 시스템을 모델링하는 모의 실험을 행한다. 시스템 방정식은 다음과 같다.

$$y(k+1) = f[y(k), y(k-1)] + u(k) \quad (13)$$

여기서 미지의 함수 f 는

$$f[y(k), y(k-1)] = \frac{y(k)y(k-1)[y(k)+2.5]}{1+y^2(k)+y^2(k-1)} \quad (14)$$

이며 $u(k) = \sin(2\pi k/25)$ 이다. 시스템의 출력을 학습하기 위하여 뉴로-퍼지 네트워크의 두 개의 입력과 하나의 출력에 대해서 100개의 샘플값들을 사용한다. 규칙을 줄이기 전에 200번을 학습하여 의존도를 구한다. 초기의 규칙수를 40개로부터 시작하는 모의 실험을 통하여 의존도에 의한 규칙의 줄어짐을 확인한다.

결론부 속성의 이산화 간격의 결정은 모든 샘플들을 구분해 줄 수 있을 정도로 작게 한다. 조건부 속성의 이산화 간격은 5배 가량 더 작게 잡아준다.

그림 4는 등간격으로 초기화된 40개의 규칙일 때 소속함수들의 출력 모양을 나타낸다.

초기 40개의 규칙수로부터 RMS 에러가 0.01을 최초로 벗어나기 전까지 규칙수를 줄인다.

표 1은 각 규칙들에 대한 의존도를 계산한 것으로

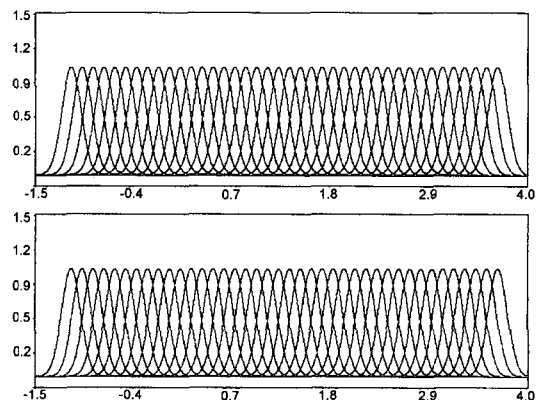


그림 4. 등간격으로 초기화된 퍼지 소속함수

표 1. 11, 6, 5개의 규칙일 때 의존도

규칙수 11		규칙수 6		규칙수 5	
규칙번호	의존도	규칙번호	의존도	규칙번호	의존도
1	23/100	1	36/100	1	36/100
2	13/100	2	19/100	2	39/100
3	14/100	3	46/100	3	32/100
4	12/100	4	35/100	4	45/100
5	32/100	5	62/100	5	51/100
6	20/100	6	16/100		
7	30/100				
8	52/100				
9	15/100				
10	10/100				
11	30/100				

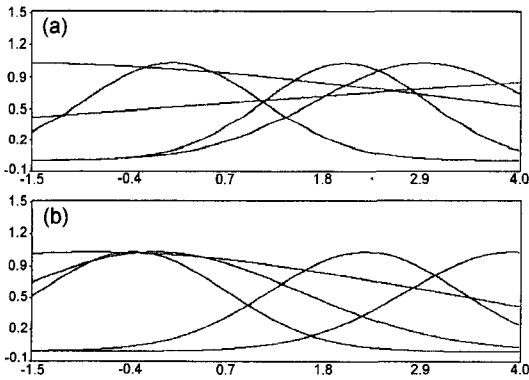


그림 5. 마지막 5개의 소속함수

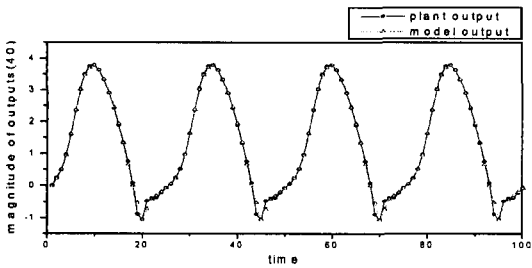


그림 6. 5개 규칙의 모델과 플랜트의 출력모양

마지막 5개의 규칙일 때 의존도를 알 수 있다. 그림 5는 5개로 줄여진 모델의 소속함수의 모양을 나타낸다. 또한 그림 6은 5개의 규칙일 때 모델의 출력과 플랜트의 출력을 나타낸다. 5개의 규칙만으로도 충분히 근사화시켰음을 알 수 있다. 그림 7은 40개로부터 규칙수가 줄어들 때마다 RMS 에러값과 규칙수를 나

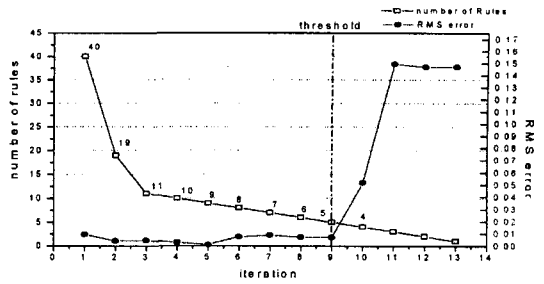


그림 7. 규칙수와 RMS 에러값

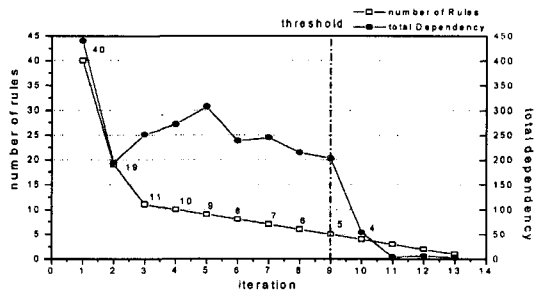


그림 8. 규칙수와 총 의존도

타낸 것이다. 그림 8은 각 단계별로 총 의존도와 규칙수를 나타낸다.

5. 결 론

BP를 사용하여 뉴로-퍼지 모델을 학습하면 BP는 발생하는 에러만을 보면서 학습을 하기 때문에 각각의 선형성분들의 중요도는 고려하지 않는다. 따라서 학습이 끝난 후에는 모든 선형성분들이 같은 중요도를 가지고 시스템에 기여하지 않을 수도 있다. 여기서 러프 집합 이론의 의존도를 사용하여 실제로 모델링에 기여하지 못하는 선형성분들을 결정하여 시스템의 잉여도를 줄이는 것을 본 논문은 제안한다. 또한 전체 RMS 에러값의 문턱값을 관찰하여 뉴로-퍼지 모델을 결정하는 것과 플랜트의 출력에 대한 뉴로 퍼지 모델의 모든 규칙들의 전체 의존도를 관찰하는 것으로부터 최소로 간략화된 뉴로-퍼지 모델을 얻을 수 있음을 보였다. 뉴로-퍼지 모델의 전체 의존도가 규칙수를 줄임에 따라 급격히 감소하는 것은 뉴로-퍼지 모델이 시스템을 증가시키지 못하고 있음을 의미한다. 또한 의존도를 사용함으로써 모델링에 전혀 기여를 못하는 많은 규칙들을 초기에 쉽게 제거함으로써 규칙수의 빠른 수렴성을 확인하였다.

의존도를 사용한 뉴로-퍼지 모델의 최적화는 순수

데이터로부터의 초기 퍼지 분할을 할 필요가 없으며 ϵ -충족(completeness)조건을 고려할 필요가 없다는 장점이 있다. 또한 초기값 문제와 관련하여 초기 규칙수를 충분히 잡아주면 초기값에 대한 모델링의 제약이 적어짐을 확인하였다.

참고문헌

[1] S. Horikawa, T. Furuhashi, and Y. Uchikawa, "On fuzzy modeling using fuzzy neural networks with the back-propagation algorithm," *IEEE Trans. Neural Networks*, vol. 3, no. 5, pp. 801-806, Sep. 1992.

[2] C. T. Chao, Y. J. Chen, and C. C. Teng, "Simplification of Fuzzy-Neural System Using Similarity Analysis," *IEEE Trans. Syst., Man, Cybern.*, pp. 344-354, vol. 26, NO. 2, April 1996.

[3] G. Castellano, A. M. Fanelli, "A neuro-fuzzy model reduction strategy," *IEEE ICNN*, vol. 4, NO. 4, August. 1997.

[4] C. C. Lee, "Fuzzy logic in control systems: Fuzzy logic controller-Part I&II," *IEEE Trans. Syst., Man, Cybern.*, vol. 20, no. 2, pp. 404-435, 1990.

[5] Kumpati S. Narendra. K. Parthasarathy, "Identification and Control of Dynamical Systems Using Neural Networks," *IEEE Trans. Neural Networks*, pp. 4-27, vol. 1, no. 1, Mar. 1990.

[6] Zdzislaw Pawlak, "Why Rough Sets?," *IEEE International Conference on Systems*, pp. 738-743, Mar. 1996.

[7] F. B. Maria, G. M. Angel and P. L. Concepcion, "Rough Sets as a Tool in Data Engineering," *IEEE International Conference on Systems*, pp. 435-440, vol. 1 of 5, Jan. 1995.

[8] Pawlak, Z., "Rough sets and Fuzzy Sets," *Fuzzy sets and Systems* 17 pp. 99-102, 1985.

[9] Pawlak, Z., "Rough sets: Theoretical Aspects of Reasoning about Data," Kluwer Academic Press, pp 51-159, 1991.

[10] C. Young wan, "Extraction of Fuzzy Rules from Data using Rough Set," *Pro. of KFIS Fall conf. '96*, pp. 327-332, vol. 6, no. 2, 1996.

[11] S. Chen, C. N. Cowan, and P. M. Grant, "Orthogonal Least Squares Learning Algorithm for Radial Basis

Function Networks," *IEEE Trans. Neural Networks*, pp. 302-309, vol. 2, No. 2, Mar. 1991.

[12] A. Cichocki, R. Unbehauen. *Neural Networks for Optimization and Signal Processing*, Wiley, pp. 434-461. 1993.

[13] Qinghua Zhang, "Using wavelet network in nonparametric estimation," *IEEE Trans. Neural Networks*, vol. 8, no. 2, pp. 227-236, Mar. 1997.

연 정 흠 (Jung-Heum Yon) 정회원

1994년 2월 : 중앙대학교 전자공학과 졸업(학사)
 1996년 8월 : 중앙대학교 대학원 전자공학과 졸업(석사)
 1996년 9월~현재 : 중앙대학교 대학원 전자공학과 박사과정 재학중
 주관심분야 : 퍼지, 유전알고리즘, 신경망 등임

서 재 용 (Jae-Yong Seo) 정회원

1996년 2월 : 중앙대학교 전자공학과 졸업(학사)
 1998년 2월 : 중앙대학교 대학원 전자공학과 졸업(석사)
 1998년 8월~현재 : 중앙대학교 대학원 전자공학과 박사과정 재학중
 주관심분야 : 퍼지, 유전알고리즘, 신경망, FNN 등임

김 용 택 (Yong-Taek Kim) 정회원

1992년 2월 : 중앙대학교 전자공학과 졸업(학사)
 1994년 8월 : 중앙대학교 대학원 전자공학과 졸업(석사)
 1996년 9월~현재 : 중앙대학교 대학원 전자공학과 박사과정 재학중
 주관심분야 : 신경망, 로보틱스, 퍼지 등임

조 현 찬 (Hyun-Chan Cho)

현재 : 한국기술교육대학교 전자공학과 교수

전 흥 태 (Hong-Tae Jeon) 정회원

현재 : 중앙대학교 전자공학과 교수
