

유전 알고리즘을 이용한 퍼지모델의 자동 동정

손유석, 장옥*, 박진배*, 주영훈**

* 연세대 전기공학과, ** 군산대 제어계측공학과

Automatic Fuzzy Model Identification Using Genetic Algorithm

You Seok Son, Wook Chang*, Jin Bee Park*, Young Hoon Joo**

* Dept. of Electrical Engineering, Yonsei Univ.

** Dept. of Control & Instrumentation Engineering, Kunsan Univ.

Abstract : This paper presents an approach to building multi-input and single-output fuzzy models for nonlinear data-based systems. Such a model is composed of fuzzy rules, and its output is inferred by simplified reasoning. Optimal structure and membership parameters for a fuzzy model are automatically and simultaneously identified by GA(Genetic Algorithm). Numerical examples are provided to evaluate the feasibility of the proposed approach. Comparison shows that the suggested approach can produce a fuzzy model with higher accuracy and a smaller number of fuzzy rules than the ones achieved previously in other methods.

1. 서론

복잡하고 불확실한 비선형 시스템을 수학적으로 모델링하는 것은 어려운 일이며 만족할만한 결과를 얻기 또한 어려운 것이 사실이다. 이러한 시스템을 다루기 위하여 퍼지 모델 동정이 다각도로 연구되어 지고 있다.

퍼지 모델 동정은 Tong [1]에 의해 처음 제시되었고, Pedrycz [2]는 상대 퍼지 집합과 Zadeh의 조건 확률 분포에 기초하여 퍼지 시스템의 새로운 동정 알고리즘을 제안하였고, Xu [3]는 퍼지 규칙의 구조 동정 및 파라미터 추정과 연관 자기 학습 알고리즘을 포함하는 일반적 퍼지 시스템 동정 기법을 제안하였다. 또한 Sugeno [4]는 질적(qualitative) 모델링 기법과 위치-경사 형의 퍼지모델을 제안하였다. Karr [5]의 경우에는 유전 알고리즘을 미리 정해진 규칙 기반위에서 먼저 규칙들의 수를 결정하고 다음으로 멤버십 함수의 파라미터를 조정하는데 이용하였다. Melsheimer [6]의 경우는 규칙과 파라미터를 동시에 조정하는 방법을 이용하였다. 퍼지 클러스터링으로 구조를 결정하고 그 전체부와 결론부 멤버십 함수를 GA hybrid scheme을 이용하여 동시에 동정하는 방법도 있다 [7].

본 논문에서는 입력력 데이터에 의한 퍼지 모델 동정을 규칙 동정과 전제부 및 결론부 파라미터 동정을 동시에 하고 후보되는 멤버십 파라미터에는 최소한의 제한인 전체 규칙상에서 멤버십 함수를 고려하는 방법으로 모델 결정에서의 자유도를 최대한 높임으로써 GA의 장점인 전범위 검색을 효율적으로 이용하는 방법을 제안한다. 여기서는 이러한 방법론을 다중 입력과 단일

출력시스템의 입력력 데이터에 대한 퍼지 모델링에 구체적으로 적용하여 컴퓨터 시뮬레이션을 통해 그 유용성을 검증한다.

2. 퍼지 모델과 유전 알고리즘

2.1 퍼지 모델

본 논문에서 다루게 될 퍼지모델의 각 규칙은 식 (1)과 같이 나타낼 수 있다.

$$\text{Rule } i: \text{ if } x_1 \text{ is } A_{i1} \dots x_n \text{ is } A_{in}, \text{ then } y_i \text{ is } w_i \quad (1)$$

(i = 1, \dots, c)

여기서 Rule i는 i번째 규칙(1 ≤ i ≤ c)을 말하며, x_j(1 ≤ j ≤ n)는 입력 변수, y_i는 i번째 규칙의 출력력을 나타내는 것이다. 본 논문에서는 입력부의 멤버십 함수를 식 (2)와 같은 이등변 삼각형으로 정의 하였고 출력부는 실수 w_i로 정의하였다.

$$A_{ij}(x_j) = \begin{cases} \text{if } (a_{ij} - b_{ij}/2) \leq x_j \leq (a_{ij} + b_{ij}/2) \\ \text{then } 1 - 2|x_j - a_{ij}|/b_{ij} \\ \text{if } x_j < (a_{ij} - b_{ij}/2) \text{ or } x_j > (a_{ij} + b_{ij}/2) \\ \text{then } 0 \end{cases} \quad (2)$$

여기서 a_{ij}는 중심값을 나타내고 b_{ij}는 이등변 삼각형의 넓이를 의미한다. 이상의 규칙과 멤버십 함수에 의한 퍼지 추론 과정은 아래와 같다.

1) 1번째 입력력 데이터 {x₁, x₂, ..., x_n, y₁}에 대해 i번째 규칙의 적합도 μ_i를 식 (3)과 같이 구한다.

$$\mu_i = A_{i1}(x_1) \times A_{i2}(x_2) \times \dots \times A_{in}(x_n) \quad (3)$$

2) 간략화 퍼지 추론의 결과 y_i*를 식 (4)에 의해 얻을 수 있다.

$$y_i^* = \frac{\sum_{i=1}^c \mu_i w_i}{\sum_{i=1}^c \mu_i} \quad (4)$$

퍼지 모델의 입력력 데이터의 갯수가 N일 때 성능 평가는 목표 출력 y_i과 추론 결과 출력 y_i*와의 평균 제곱 오차에 의하여 행한다.

$$E = \frac{1}{N} \sum_{i=1}^N (y_i - y_i^*)^2 \quad (5)$$

2.2 유전 알고리즘

유전 알고리즘은 자연선택을 흉내낸 최적화 기법으로서 복잡한 최적화 문제에서 뛰어난 강인성을 발휘하며 일반적으로 다음의 3개의 연산자들로 구성된다.

- 1) 재생 연산자 : 재생은 적합도 함수값에 의해서 개별적인 스트링이 복제되는 과정이다. 이 연산자를 통해 보다 높은 적합도를 갖는 스트링이 이후의 세대에서 더 많은 수의 자손을 갖게 된다.
- 2) 교차 연산자 : 재생에 이어서 교차가 진행된다. 교배풀에서 새로이 재생된 스트링이 임의로 짝을 이루면, 각 쌍의 스트링은 임의로 정해지는 정수 위치 $K(1 < K < L, L$ 은 스트링의 길이)에 대해 $K+1$ 과 L 사이의 모든 문자열이 교환된 새로운 스트링으로 형성된다.
- 3) 돌연변이 연산자 : 스트링내의 임의의 위치에서 돌연변이를 형성하는 연산자이다. 이 연산자는 교차 연산자와 함께 변형 개체 생성에 중요한 역할을 한다.

본 논문에서는 최초의 개체군 P_0 를 형성시 입출력부 파라미터 스트링들과 규칙 구조 결정 스트링들을 포함하는 통합검색체로 표시되는 M 개의 개체를 생성한다. 이러한 스트링들은 정수 코딩 스트링들과 실수 코딩 스트링들로 나뉘게되며, 실수 코딩의 경우 정규화 과정을 통해 0과 1사이의 실수 값을 가지게 하며, 정수코딩은 각 스트링에 주어지는 한계값 내의 정수값을 가지게 된다. 적합도 평가는 일반적으로 퍼지모델에 의한 출력과 실제 출력 사이의 평균 제곱오차를 식 (5)에 의해 구하여 행하나 본 논문에서는 전진부에서의 null set 생성 여부에 따라 penalty를 부과하는 방법을 추가한 식 (6)에 따라 행한다.

$$fitness(a) = \begin{cases} \text{if no null set exists :} \\ \frac{1}{E} = \frac{N}{\sum_{i=1}^N (y_i - \hat{y}_i)^2} \\ \text{if null set exists :} \\ \frac{1}{E} \times penalty = \frac{N}{\sum_{i=1}^N (y_i - \hat{y}_i)^2} \times penalty \end{cases} \quad (6)$$

재생은 적합도 비례법을 선택하고 부모개체의 변형에는 교차와 돌연변이 연산자만을 이용한다. 새로운 개체군의 생성에 있어서 Elitist 이론을 적용하여 현세대의 최고의 적합도를 갖는 개체는 항상 다음 개체군에 전해지게 한다. 이러한 Elitist 이론은 GA 탐색 결과의 수렴성을 보장하기 위한 확실한 방법으로 알려져 있다[5].

3. 퍼지 모델의 자동 동정

3.1 통합 유전 검색체의 구성

유전 알고리즘 적용은 검색체의 구성에서부터 시작한다. 각 개체의 특성들은 관련 스트링들을 하나로 통합한 유전 검색체로 나타낼 수 있으며 유전 검색체내의 각 스트링들은 다음과 같다.

1) 입력부 파라미터 스트링

이는 입력 변수에 대한 멤버십 함수의 파라미터들을 정의하는 스트링이다. 선택될 후보 멤버십 함수의 파라미터에 관한 정보를 갖는 스트링 구조는 그림 1과 같다.

X	width (1)	center (1)	width (f)	center (f)
---	-----------	------------	------	-----------	------------

그림 1. 입력 변수 멤버십 함수의 스트링 구조

입력 변수의 각 파라미터들에 대해 r 개의 후보 파라미터값이 실수로 주어진다. 본 논문에서는 입력 변수를 0에서 1까지의 값으로 정규화 하여서 사용하므로 각 후보값 또한 0에서 1까지의 임의의 실수 값으로 주어진다. 이러한 스트링 구조는 다른 시스템 입출력 데이터들에도 쉽게 적용할 수 있는 범용성을 가진다. 본 논문에서는 후보 파라미터 값들에 대해 중심값들을 순서대로 배열하는 등의 일체의 재배열을 주지 않으므로 유전 알고리즘 검색의 효율성을 높게 된다.

2) 출력부 스트링

출력 변수는 정규화하여 사용하고 출력부는 실수치이므로 w_i 의 후보값들은 0에서 1사이의 실수 값을 가지며 그림 2와 같은 구조를 갖는다.

Y	position (1)	position (2)	position (s)
---	--------------	--------------	------	--------------

그림 2. 출력 변수 스트링

3) 입력 구조 결정 스트링

i 번째 규칙의 입력 구조 결정은 그림 3과 같은 정수 스트링 구조에 의해서 결정된다. 스트링내의 각 값은 0에서 r 까지의 정수로 된다. 이 값들은 입력 변수들에 대한 r 개의 후보 파라미터 중 하나를 선택하는 것을 의미한다. 그리고 0의 의미는 그 입력 변수는 고려하지 않겠다는 의미로 해석되며 입력 구조 결정 스트링내의 값들이 모두 0이되면 그 규칙은 제외된다.

Rule i	XMF i1	XMF i2	XMF in
--------	--------	--------	------	--------

그림 3. 입력 구조 결정 스트링

4) 출력 구조 결정 스트링

출력 구조의 결정은 규칙의 수를 결정하는 것과 직접 연관을 가지게 된다. 최대 규칙 수 max_c 만큼의 스트링 요소를 가지는 출력 구조 결정 스트링은 그림 4와 같이 나타낼 수 있으며, 각 스트링 요소는 0에서 s 까지의 값을 갖게되어 출력부 스트링 요소 중 하나의 값을 지정하게 된다. 단 0의 값이 선택된 경우는 그 규칙은 제외된다.

wi	w1	w2	w max_c
----	----	----	------	---------

그림 4. 출력 구조 스트링

3.2 유전 알고리즘에 의한 모델 동정

유전 알고리즘에 의한 모델 동정은 먼저 최대규칙수 max_c 와 입출력부 스트링에서 후보 파라미터의 갯수 r, s 의 결

정, 개체수 m의 결정으로 시작된다. 이러한 골격 아래에서 특정 입출력 데이터 집합을 퍼지 모델로 동정하게 된다. 입출력 데이터 쌍을 입력시키고 이들을 정규화 시켜 범용으로 구성된 염색체 스트링 구조에 일치시킨다.

정규화된 입출력 데이터를 얻은 후에는 모든 개체들의 염색체의 스트링들을 초기화시키고 첫번째 세대를 시작하게 된다. 염색체의 초기화는 실수 스트링에서는 0에서 1까지의 임의의 실수를 난수 발생기로 발생시켜서 행하고 정수 스트링은 한계값내의 정수를 난수 발생기로 발생시켜 행한다.

모든 개체들의 초기화가 완료되면 각개체의 염색체내의 정보를 추출하여 퍼지모델을 평가하게 된다. 추출한 모델을 기반으로 식 (4)에 의한 간략화 퍼지추론을 통하여 y_j 를 구하여낸다. 그 개체의 적합도값을 식 (6)에 의해 결정하게 되는데 입력 변수에 대한 null set을 유발하는 개체를 방지하기 위한 penalty의 적용은 개체에 주어지는 유일한 제한조건으로 작용하게 된다. 그리고 새로운 자식 개체들의 생성이 재생, 교차, 돌연변이 연산자에 의해 행하여진다. 재생 방법은 적합도 비례법을 사용하여 식 (7)과 같은 확률로 행한다.

$$p(a) = \text{fitness}(a) / \sum_{i=1}^m \text{fitness}(a) \quad (7)$$

이 과정을 반복하여 M개의 새로운 개체들을 만들어내고 이러한 자식 개체군은 부모 개체군을 대체하게 된다. 이 과정에서 Elitist 이론이 적용되어 max_fitness를 지닌 개체는 다음 개체에서도 보존 되는 것이다. 이후 새로 생성된 개체군에 대해 퍼지 모델 평가를 통하여 새로운 적합도 값을 부여하고 다시 반복하여서 세대 교합을 하게되면 점차 발전된 퍼지 모델을 얻을 수 있게된다.

4. 실험 및 결과 고찰

퍼지 모델 동정의 실험으로서는 Box와 Jenkins[8]에 의해 주어지는 가스로(gas furnace)의 데이터를 이용하였다. 이 데이터들은 입력으로 가스 유입률(u)과 출력으로 CO₂ 밀도(y)로 나타내는 296개의 데이터 쌍으로 구성되어있다. 이 프로세서는 동적이므로 현재의 출력 y(l)에 영향을 미치는 입력들로서는 u(l-4)와 y(l-1)의 두개를 택하였다. 이러한 프로세서의 동정을 위하여 100개의 유전 개체를 택하고 최대 규칙수 max_c는 6개로하고 입력부 후보 멤버십 파라미터의 갯수는 10개로 하고 출력부 후보 실수 값도 10개로 정하였다. 유전 알고리즘에 의한 진화 과정에서 교차율은 0.5로두고 돌연변이율은 0.05로 두었다. 적합도 결정에서의 null set고려에 의한 penalty는 0.0001로 두었다. 유전 알고리즘에 의한 자동 동정으로 규칙 수는 5로 결정 되었고 얻어진 퍼지 모델은 그림 5와 같다.

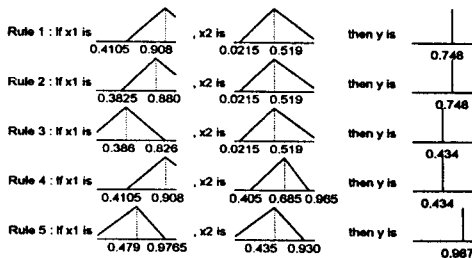


그림 5. 가스로에 대해 동정된 퍼지 모델

본 논문에서 제안한 기법에 의한 성능과 기존 모델들의 성능은 표 1에서 비교되었다. 표 1에서 알 수 있듯이 본 논문에서 제안한 기법에 의해 동정된 퍼지모델이 적은 규칙을 사용하여 가장 정확하게 동정되었음을 알 수 있다.

표 1. 가스로에대한 기존 모델들과의 성능 비교

모 델	규칙의 수	오 차
Tong [1]	19	0.469
Pedrycz [2]	81	0.320
Xu [3]	25	0.328
Sugeno [4]	6	0.190
본 논문의 결과	5	0.162

5. 결론

본 논문에서는 수학적으로 모델링하기 어려운 복잡하고 불확실한 비선형 시스템을 입출력 데이터를 사용하여 모델링할 수 있는 퍼지 시스템 모델링 기법을 제안하였다. 제안한 기법은 유전알고리즘을 이용하여 시스템의 퍼지 제어 규칙수와 규칙의 전 반부 및 후반부 파라미터를 동시에 효율적으로 동정할 수 있는 새로운 퍼지 시스템 모델링 기법이다. 이 기법은 퍼지 시스템 모델을 구축하는데 있어 자유도를 높은 통합 유전 염색체를 구성하여 유전 알고리즘을 보다 효율적으로 퍼지 모델의 자동 동정에 이용하는 것이다. 마지막으로 제안한 기법을 가스로 시스템에 응용하여 기존의 기법보다 적은 수의 퍼지 규칙을 가지고 우수한 성능을 갖음을 보였다.

6. 참고문헌

- [1] Tong R. M., "The evaluation of fuzzy models derived from experimental data", *Fuzzy Sets and Systems*, vol. 4, pp. 1-12, 1980.
- [2] Pedrycz W., "An identification algorithm in fuzzy relational systems", *Fuzzy Sets and Systems*, vol. 13, pp. 153-167, 1984.
- [3] Xu C. W., "Fuzzy model identification and self-learning for dynamic systems", *IEEE Trans. Syst. Man, Cybern.*, vol. 17, no. 4, pp. 683-689, 1987.
- [4] Sugeno M. and Yasukawa T., "A fuzzy-logic-based approach to qualitative modeling", *IEEE Trans. Fuzzy Sys.*, vol. 1, no. 1, 1993.
- [5] Karr C., "Genetic algorithms for fuzzy controllers". *AI Expert*, vol. 6, pp. 26-33, 1991.
- [6] Liska J. and Melsheimer S. S., "Complete design of fuzzy logic systems using genetic algorithms", *Proc. of the third IEEE confer. Fuzzy systems*, pp. 1377-1382, 1994.
- [7] Y.H. Joo, H.S. Hwang, K.B. Kim and K.B. Woo, "Fuzzy System Modeling by Fuzzy Partition and GA Hybrid Schemes", *International Journal for Fuzzy Sets and Systems*, 1996, (To be printed)
- [8] Box G. E. P and Jenkins G. M., *Time series analysis, forecasting and control*, Holden Day, San Francisco, 1970.