

정보 입자 기반 퍼지 모델의 최적 동정을 위한 유전론적 접근

박건준*, 오성권*, 이동윤**
 * 수원대학교 전기공학과
 ** 중부대학교 경찰경호학과

Genetic Approach for Optimal Identification of IG-based Fuzzy Model

Keon-Jun Park*, Sung-Kwon Oh*, and Dong-Yoon Lee**
 * Dept. of Electrical Engineering, The University of Suwon
 ** Dept. Police & Protection, JoongBu University

Abstract - 본 논문에서는 복잡하고 비선형적인 시스템에 대하여 구체적이고 체계적인 방법에 의한 퍼지 모델을 동정하기 위해 유전자알고리즘을 이용하여 전반부 및 후반부의 구조와 파라미터 동정하기 위한 유전론적 접근을 소개한다. 정보 입자 기반 퍼지 모델의 구조를 동정하기 위하여 유전자 알고리즘을 이용하여 입력 변수의 수, 선택된 입력 변수, 멤버십함수의 수, 그리고 후반부 형태를 결정하고, 파라미터를 동정하기 위하여 전반부 멤버십 파라미터를 동조하여 최적의 퍼지 모델을 설계한다. 또한 구조 동정 및 파라미터 동정에 있어서 개별적인 방법과 동시적인 방법으로 접근하여 정보 입자 기반 퍼지 모델의 최적 동정을 도모한다. 마지막으로 제안된 퍼지 모델은 표준 모델로서 널리 사용되는 수치적인 예를 통하여 평가한다.

1. 서 론

퍼지집합 이론은 비선형적이고 복잡한 실 시스템의 특성을 해석하는데 적용함으로써 수학적 모델보다 좋은 결과를 가져왔으며, 설계할 시스템의 성능 및 기능의 요구조건에 따라 퍼지 모델은 애매 모호한 언어적 변수를 수치적으로 표시할 수 있어서 응용성 있는 시스템 설계를 가능하게 하고 시스템의 기능을 향상시키며 설계를 간단하게 해주는 장점이 있다. 그러나 대상 시스템을 모델링하기 위해서는 구조 동정 및 파라미터 동정에 있어서 전문가와 시행착오에 의존해야 하는 어려움이 아직까지 남아있으며, 특히 데이터 특성에 맞는 퍼지 모델을 구축하는데 있어서는 동적으로 변화하는 환경에서 적용적으로 대처할 수 있는 퍼지 모델을 구축하기가 힘들다. 이러한 문제를 해결하기 위한 방법으로 본 논문에서는 정보 입자[1, 2]의 특성에 맞는 퍼지 모델을 설계 하고, 최적의 퍼지 모델을 동정하기 위하여 구조 동정 및 파라미터 동정에 있어서 유전자 알고리즘을 이용하여 개별적 방법과 연속적 방법의 두 가지 측면에 대해 고려한다. 퍼지 모델의 구조를 결정하기 위하여 유전자 알고리즘[4]에 이용하여 시스템의 입력 수, 선택된 입력 변수, 멤버십함수의 수, 그리고 후반부 추론 형태를 결정한다. 파라미터 동정에서는 전반부 멤버십 파라미터를 유전자 알고리즘을 이용하여 최적으로 동정한다.

2. 정보 입자 기반 퍼지 모델

2.1 정보 입자

정보 입자[1, 2]는 근접성, 유사성 또는 기능성의 기준에 의해 서로 결합된 물체(특히, 데이터 점)의 연결된 모임으로 간주된다. 정보 입자화는 어떤 문제를 쉽게 이해하기 위해 수행되는 인간의 고유 활동이며, 특히, 어떤 문제를 다루기 쉬운 몇 개의 큰 덩어리로 나누는 것을 목적으로 한다. 이러한 방법으로 이 문제들은 기존보다 더 적은 계산적인 복잡성을 위해 잘 정의된 하위문제(모듈)들로 분할된다. 정보 입자는 클러스터링, 퍼지클러스터링, Granular Computing 및 진화점퓨팅의 융합, 결합 및 확장을 통해 데이터의 전처리, 지능모델의 초기 구조 또는 파라미터를 결정함으로써 정보 데이터의 특성을 효과적으로 반영하게 된다. 본 논문에서는 데이터들간의 거리를 기준으로 근접한 정도를 측정하여 데이터를 특성별로 분류하는 HCM 클러스터링 알고리즘[3]을 이용한다. 본 논문에서는 HCM 클러스터링을 통해 입력력 데이터의 중심값을 이용하여 퍼지 모델의 전반부 멤버십 함수의 초기 정점을 동정하고, 후반부 입력력 데이터의 중심값을 적용하여 정보 입자 기반 퍼지 모델을 구축한다.

2.2 전반부 동정

퍼지 모델링에서 전반부 동정, 즉 구조 동정 및 파라미터 동정은 비선형 시스템을 표현하는데 있어서 매우 중요하다. 기존의 방법은 멤버십 함수를 입력 변수의 최소값과 최대값 사이에서 임의의 개수로 등분하여 일률적으로 정의하였으나 이는 데이터들이 가지고 있는 특성을 제대로 반영하지 못하는 단점이 있다. 그래서 전반부 파라미터 동정을 위해 HCM 클러스터링에 의해 초기 멤버십함수의 정점을 동정한다.

HCM 클러스터링을 통한 정보 입자에 의한 전반부 동정은 다음과 같다.

전체 데이터 집합 $U = \{x_1, x_2, \dots, x_l; y\}$ 이고, 여기서 $x_k = [x_{k1}, \dots, x_{km}]^T$, $y = [y_1, \dots, y_m]^T$, l 은 변수의 수 그리고 m 은 데이터의 수로 가정한다.

[단계 1] 전체 데이터 집합 U 를 각각의 입력 데이터와 출력 데이터의 데이터 집합 X_k 로 배열한다.

$$X_k = [x_k; y] \quad (1)$$

여기서, X_k 는 k 번째 입력 데이터와 출력 데이터의 데이터 집합이고, $k=1, 2, \dots, l$ 이다.

[단계 2] 데이터 집합 X_k 로부터 중심 벡터 v_{kg} 를 구하기 위해 HCM 클러스터링을 시행한다.

[단계 2-1] 데이터 집합 X_k 를 c 개의 클러스터(정보 입자)로 분류한다.

[단계 2-2] 각 클러스터의 중심 벡터 v_{kg} 를 계산한다.

$$v_{kg} = \{v_{k1}, v_{k2}, \dots, v_{kc}\} \quad (2)$$

[단계 3] 중심 벡터 v_{kg} 로 해당하는 입력의 개별적인 퍼지 공간을 분할하고 각 클러스터에 Small, Big과 같은 언어적 변수를 할당한다.

[단계 4] 중심 벡터 v_{kg} 를 멤버십함수의 초기 정점으로 설정한다.

2.3 후반부 동정

퍼지 모델의 후반부 동정도 전반부와 마찬가지로 구조 동정과 파라미터 동정으로 나뉘어진다. 정보 입자에 따른 다항식 함수의 초기 값을 가지고 구조를 동정한다.

[단계 1] j 번째 규칙의 퍼지 공간에 속한 데이터 집합을 찾는다.

[단계 2] 각 규칙에서 산출 평균에 의한 데이터 집합의 중심 벡터 V_j 를 계산한다.

$$V_j = \{V_{j1}, V_{j2}, \dots, V_{jn}; M_j\} \quad (3)$$

여기서, $j=1, 2, \dots, n$. V_{jk} 와 M_j 는 각각 입력 데이터와 출력 데이터의 중심값이다.

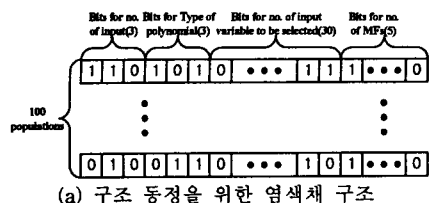
[단계 3] 중심 벡터 V_j 를 후반부 다항식 함수의 초기값으로 설정한다.

3. 최적 동정을 위한 유전론적 접근

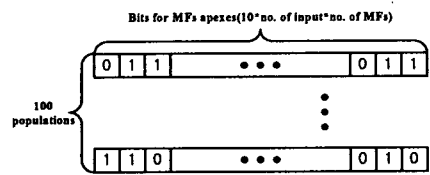
유전자 알고리즘[4]은 자연 선택과 유전학에 기반으로 하는 확률적인 탐색방법으로써 탐색과 해의 가능영역들을 균형 있게 이용하여 위하여 생산, 교배, 돌연변이의 과정을 수행하는 일련성 있는 탐색법으로 비선형 최적화 이론에 탁월한 성능을 발휘하고 있다. 기존의 다른 탐색방법들과 달리 유전자 알고리즘은 해가 될 가능성이 있는 개체집단을 유지하면서 그들 모두가 동시에 최적값을 찾아나가기 때문에 지역 극소에 빠질 위험을 어느 정도 해결할 수 있다는 점과 더불어 모델의 성능지수가 최소가 되는 전역 극소 영역을 찾는 능력을 갖고 있으며, 또한 기존의 방법들과는 달리 선형, 연속, 미분가능 등의 제한이 없기 때문에 다양한 분야에 별다른 제한 없이 적용할 수 있다는 장점을 가진다는 것이 중요한 특징이다.

대상 시스템을 퍼지 모델링하기 위해서는 전문가와 시행착오에 의존하는 어려움이 있다. 이를 해결하기 위해 주어진 시스템의 입력 변수 및 멤버십 함수 그리고 후반부 구조를 유전자 알고리즘을 통해 구조를 찾아낸다. 또한 전반부 멤버십 파라미터를 최적으로 동정한다. 구조 동정 및 파라미터 동정에 있어서 개별적인 방법과 연속적인 방법을 고려하여 최적의 퍼지 모델을 설계한다.

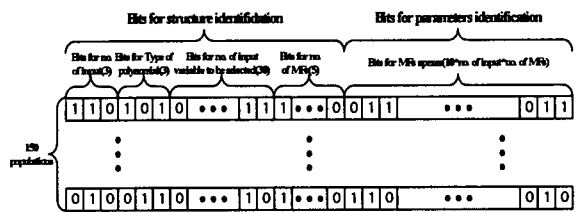
첫 번째로 개별적인 방법은 구조 동정과 파라미터 동정에 있어서 시스템의 구조 먼저 결정하고 그 구조에 대해 파라미터를 동정한다. 두 번째로, 연속적인 방법은 시스템의 구조와 파라미터를 연속적인 탐색체 구조를 가지며 구조와 파라미터를 동시에 동정한다. 그림 1은 개별적인 방법에 의한 유전자 알고리즘의 탐색체 구조를 보여 주며, 그림 2는 연속적 방법에 의한 탐색체 구조를 보여준다.



(a) 구조 동정을 위한 검색채 구조



(b) 파라미터 동정을 위한 검색채 구조
〈그림 1〉 개별적인 방법에 의한 검색채 구조



〈그림 2〉 연속적 방법에 의한 검색채 구조

4. 실험 데이터를 통한 결과 고찰

제안된 퍼지 모델의 평가를 위해 다른 지능모델에 널리 사용된 비선형 공정에 대한 성능 평가의 척도로 사용되고 있는 가스로 공정[6]을 사용한다. 모델의 평가 기준인 성능지수는 수치 데이터인 가스로 공정에 대해서 MSE(Mean Squared Error), 식 (4)를 이용한다.

$$PI = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2 \quad (4)$$

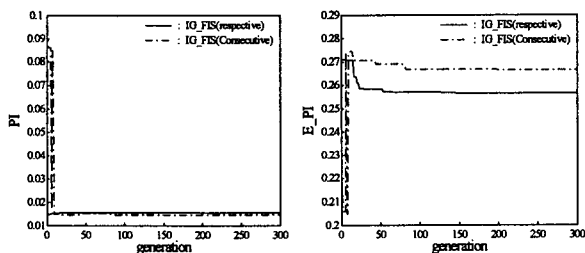
가스로 시계열 데이터는 입출력 데이터인 가스 흐름을 $u(t)$ 과 연소된 이산화탄소 농도 $y(t)$ 의 가스로 공정으로부터 추출된다. 시스템의 입력 변수를 동정하기 위해서 $u(t-3), u(t-2), u(t-1), y(t-3), y(t-2), y(t-1)$ 의 6입력을 적용한다. 출력은 $y(t)$ 이다. 또한 전체 시스템 입출력 데이터 쌍 296개 중 처음 148개의 데이터는 학습 데이터로 이용하고, 나머지 148개의 데이터는 테스트 데이터로 나누어 모델의 근사화와 일반화 능력을 평가한다

구조와 파라미터를 개별적으로 동정한 퍼지 모델의 경우 입력 변수는 $u(t-3), y(t-1)$ 이 선택되었고, 멤버쉽함수의 수는 각 입력 변수에 대해 각 3개이고, 후반부 구조는 구조 4가 결정되었으며, 연속적으로 동정한 퍼지 모델 경우에는 입력 변수가 개별적으로 동정한 퍼지 모델과 같이 선택되었고, 멤버쉽함수의 수는 각 입력 변수에 대해 2, 3개이고, 후반부 구조는 구조 3이 결정되었다. 표 1은 두 가지 동정에 의한 정보 입자 기반 퍼지 모델의 구조와 그에 따른 성능지수를 보여준다. 그림 3은 연속적 방법에 의한 동정된 멤버쉽 파라미터를 보여준다.

〈표 1〉 제안된 모델의 성능지수

Identification	input variable	No. of MFs	Type	PI	E_PI	
respective	S	$u(t-3)$	3x3	Type 4	0.015	0.281
	P	$y(t-1)$			0.015	0.256
consecutive	S+P	$u(t-3)$ $y(t-1)$	2x3	Type 3	0.014	0.266

*S : Structure, P : Parameters



(a) PI (b) E_PI
〈그림 4〉 제안된 모델의 성능지수

그림 3은 두 가지 방법론적 최적화인 개별적인 동정과 연속적인 동정에 의한 제안된 정보 입자 기반 퍼지 모델의 성능지수를 비교하여 보여준다.

표 2는 기존의 퍼지모델과 동정 오차를 비교하여 보여준다. 제안된 정보 입자 기반의 퍼지 모델이 두 가지 방법론적 접근에서 기존의 퍼지 모델보다 성능이 향상된 것을 알 수 있다.

〈표 2〉 기존 퍼지모델과 동정 오차 비교

Model	PIt	PI	E_PI	규칙수	
Tong's model[5]	0.469			19	
Pedrycz's model[6]	0.776			20	
Xu's model[7]	0.328			25	
Sugeno's model[8]	0.355			6	
Oh et al.'s model[9,10]	구조 1		0.024	0.328	4
	구조 2		0.022	0.326	4
HCM+GA [11]	구조 1		0.021	0.364	6
	구조 2		0.035	0.289	4
Our model	구조 1		0.022	0.333	6
	구조 2		0.026	0.272	4
Our model	respective		0.015	0.256	9
	consecutive		0.014	0.266	6

5. 결 론

본 논문에서는 정보 입자를 이용한 퍼지 모델의 유전론적 최적화를 위해 두 가지의 방법으로 접근을 시도하였다. 유전자 알고리즘을 이용하여 최적화 모델을 동정하는 데 있어서 시스템의 구조를 먼저 결정하고 파라미터를 동정하는 개별적 방법과 구조 동정과 파라미터 동정을 동시에 시행하는 연속적 방법을 제시하였다. 또한, HCM 클러스터링 알고리즘에 의한 정보 입자는 퍼지규칙의 전반부 및 후반부에 각각 사용될 멤버쉽함수의 초기 정점 및 다항식 함수의 초기값과 같은 퍼지 모델의 초기 파라미터를 결정하여 데이터 특성을 적용하였고, 초기 파라미터는 유전자 알고리즘에 의해 효과적으로 동조하였다. 제안된 모델은 입출력 특징을 이용함으로써 복잡하고 비선형이 강한 공정에 기존의 퍼지 모델들 보다 성능이 향상된 퍼지 모델을 설계할 수 있었고, 객관적인 모델링을 할 수 있었다.

감사의 글

본 연구는 산업자원부의 지원에 의하여 기초전력연구원 (R-2003-B-274) 주관으로 수행된 과제임.

[참 고 문 헌]

- [1] L. A Zadeh, 'Fuzzy logic = Computing with words,' *IEEE Tr ans. Fuzzy Syst.*, Vol. 4, No. 2, pp. 103-111, 1996.
- [2] W.Pedryca and G. Vukovich, 'Granular neural networks,' *Ne urocomputing*, Vol. 36, pp. 205-224, 2001.
- [3] P. R. Krishnaiah and L. N. Kanal, editors. Classification, patte rn recognition, and reduction of dimensionality, volume 2 of Handbook of Statistics. North-Holland, Amsterdam, 1982.
- [4] D. E. Goldberg, 'Genetic Algorithm in search, Optimization & Machine Learning,' Addison wesley, 1989.
- [5] R. M. Tong, "The evaluation of fuzzy models derived from e xperimental data," *Fuzzy Sets Syst.*, Vol. 13, pp.1-12, 1980.
- [6] W. Pedrycz, "An identification algorithm in fuzzy relational s ystem," *Fuzzy Sets Syst.*, Vol. 13, pp.153-167, 1984.
- [7] C. W. Xu and Y. Zailu, "Fuzzy model identification self-learnin g for dynamic system," *IEEE Trans. on Syst. Man, Cybern.*, Vol. SMC-17, No. 4, pp.683-689, 1987.
- [8] M. Sugeno, T. Yasukawa, "Linguistic modeling based on nu merical data," *IFSA'91 Brussels, Copmputer, Management & S ystem Science*, pp.264-267, 1991.
- [9] S.-K. Oh and W. Pedrycz, "Identification of Fuzzy Systems b y means of an Auto- Tuning Algorithm and Its Application t o Nonlinear Systems," *Fuzzy Sets and Syst.*, Vol. 115, No. 2, pp. 205-230, 2000.
- [10] C.-S. Park, S.-K. Oh, and W. Pedrycz, "Fuzzy Identification by means of Auto-Tuning Algorithm and Weighting Factor," *The Third Asian Fuzzy Systems Symposium(AFSS)*, PP.701-706, 1998.
- [11] B.-J. Park, W. Pedrycz and S.-K. Oh, "Identification of Fuzz y Models with the Aid of Evolutionary Data Granulation," *IE E Proc.-Control Theory and Applications*, Vol. 148, Issue 05, pp. 406-418, Sept. 2001.