

연속 동조 방법을 이용한 퍼지 집합 퍼지 모델의 유전자적 최적화

Genetic Optimization of Fuzzy Set-Fuzzy Model Using Successive Tuning Method

박건준*, 오성권**, 김현기***

(Keon-Jun Park, Sung-Kwun Oh, and Hyun-Ki Kim)

Abstract - In this paper, we introduce a genetic optimization of fuzzy set-fuzzy model using successive tuning method to carry out the model identification of complex and nonlinear systems. To identify we use genetic algorithm (GA) and C-Means clustering. GA is used for determination the number of input, the selected input variables, the number of membership function, and the conclusion inference type. Information Granules (IG) with the aid of C-Means clustering algorithm help determine the initial parameters of fuzzy model such as the initial apexes of the membership functions in the premise part and the initial values of polynomial functions in the consequence part of the fuzzy rules. The overall design arises as a hybrid structural and parametric optimization. Genetic algorithms and C-Means clustering are used to generate the structurally as well as parametrically optimized fuzzy model. To identify the structure and estimate parameters of the fuzzy model we introduce the successive tuning method with variant generation-based evolution by means of GA. Numerical example is included to evaluate the performance of the proposed model.

Key Words : Information Granules, Fuzzy Set-Fuzzy Model, Successive Tuning Method, Optimal Identification, GAs

1. 서론

많은 연구자들은 복잡하고 비선형적인 시스템을 모델링 하기 위해 여러 분야에서 다양한 형태로 연구가 이루어지고 있다. 그 중 퍼지집합 이론은 비선형적이고 복잡한 실 시스템의 특성을 해석하는데 적용함으로써 수학적 모델보다 좋은 결과를 가져왔다. 설계할 시스템의 성능 및 기능의 요구조건에 따라 퍼지 모델은 애매 모호한 언어적 변수를 수치적으로 표시할 수 있어서 융통성 있는 시스템 설계를 가능하게 하고 시스템의 기능을 향상시키며 설계를 간단하게 해주는 장점이 있다. 그러나 데이터 특성에 맞는 퍼지 모델을 구축하는데 있어서는 전문가와 시행착오에 의존해야 하는 어려움이 있으며 동적으로 변화하는 환경에서 적응적으로 대처할 수 있는 퍼지 모델을 구축하기가 힘들다. 이러한 문제를 해결하기 위한 방법으로 정보 입자(Information Granules:IG)[1,2]에 대한 연구가 행해지고 있다.

본 논문에서는 정보 입자의 특성에 맞는 퍼지 모델을 설계하고 최적의 퍼지 모델을 동정하기 위하여 유전자 알고리즘에 의한 연속 동조 방법을 제안한다. 이를 위해, 데이터들간의 근접성을 기준으로 한 C-Means 클러스터링 알고리즘[3]에 의한 정보 입자를 이용하여 공간 분할 및 초기 멤버쉽함수의 정점을 동정하고, 유전자 알고리즘[4]에 의한 연속 동조

방법을 이용하여 시스템의 입력 수, 선택된 입력 변수, 멤버쉽함수의 수, 후반부 추론 형태와, 초기 멤버쉽함수의 정점과 같은 구조/파라미터를 동시에 동정한다. 또한, 연속 동조 방법에 의해 동조된 멤버쉽함수의 정점은 퍼지 모델의 규칙 후반부 중심에 적응적으로 대체되며 그에 따른 정보 입자에 의한 퍼지 모델을 설계한다.

2. 퍼지 집합 퍼지 모델

2.1 정보 입자 기반 퍼지 집합

정보 입자[1,2]는 근접성, 유사성 또는 기능성 등의 기준에 의해 서로 결합된 대상(특히, 데이터 점)의 연결된 모임으로 간주되며, 기존보다 계산적인 복잡성 줄이기 위해 잘 정의된 하위문제(모듈)들로 분할된다. 따라서, 데이터들간의 근접성을 기준으로 데이터들을 분류하고 각 분류된 클러스터들의 중심점을 이용하여 정보 입자 기반 퍼지 집합을 형성한다. 이를 위해 근접성을 기준으로 한 C-Means 클러스터링 알고리즘[3]을 이용한다.

2.2 전반부 동정

퍼지 모델링에서 전반부 동정, 즉 구조 동정 및 파라미터 동정은 비선형 시스템을 표현하는데 있어서 매우 중요하다. 기존의 방법은 멤버쉽 함수를 입력 변수의 최소값과 최대값 사이에서 임의의 개수로 등분하여 일률적으로 정의하였으나 이는 데이터들이 가지고 있는 특성을 제대로 반영하지 못하는 단점이 있다. 따라서, 전반부 동정을 위해 정보 입자에 의한 클러스터 중심을 이용하여 입력 공간을 분할하고 삼각형 멤버쉽함수의 정점을 동정한다.

저자 소개

* 박건준: 水原大學 電氣工學科 博士課程

** 오성권: 水原大學 電氣工學科 教授 · 工博

*** 김현기: 水原大學 電氣工學科 教授 · 工博

전체 데이터 집합 $U = \{x_1, x_2, \dots, x_l; y\}$ 이고, 여기서 $x_k = [x_{1k}, \dots, x_{mk}]^T$, $y = [y_1, \dots, y_m]^T$, 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.

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

[단계 2] 데이터 집합 X_k 로부터 중심 벡터 v_{kg} 를 구한다.

[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.

2.3 후반부 동정

입출력 데이터간의 중심값을 이용하여 전반부뿐만 아니라 후반부에도 입출력 데이터 관계의 특성을 고려한다. 후반부 다항식 함수에 입출력 데이터의 중심값을 적용하여 데이터 입자에 의한 퍼지 모델을 구축한다.

각 규칙에 속한 입력 데이터의 중심값과 출력 데이터의 중심값은 다음과 같다.

[단계 1] 입력 데이터의 중심 벡터 V_{kj} 는 각 규칙의 전반부 멤버쉽 함수의 정점으로 설정한다. 즉

$$V_{kj} = v_{kj}, j=1, 2, \dots, n \quad (3)$$

[단계 2] 출력 데이터의 중심 벡터 M_j 출력 데이터와 각 규칙의 활성화도의 가중 평균에 의해 구한다. 즉,

$$M_j = \frac{\sum_{i=1}^m w_{ji} y_j}{\sum_{i=1}^m w_{ji}} \quad (4)$$

후반부 구조로는 구조 1(간략 퍼지추론), 구조 2(선형 퍼지추론)와 구조 3(2차식 퍼지추론), 그리고 구조 4(변형된 2차식 퍼지추론)를 이용하였다. 2차식 퍼지추론의 경우 후반부가 2차식 함수의 다항식 형태를 가지며, 퍼지 집합 퍼지 모델은 다음과 같은 구현 규칙으로 구성된다.

$$R^j: \text{If } x_k \text{ is } A_{kc} \text{ then } y_j - M_j = f_j(x_1, \dots, x_k) \quad (5)$$

$$f_j(x_1, \dots, x_k) = a_{j0} + a_{j1}(x_1 - V_{1j}) + \dots + a_{jk}(x_k - V_{kj}) + a_{j(k+1)}(x_1 - V_{1j})^2 + \dots + a_{j(2k)}(x_k - V_{kj})^2 + a_{j(2k+1)}(x_1 - V_{1j})(x_2 - V_{2j}) + \dots + a_{j((k+2)(k+1)/2)}(x_{k-1} - V_{(k-1)j})(x_k - V_{kj}) \quad (6)$$

퍼지 추론에 의해 추론된 값 y^* 은 식 (7)과 같다.

$$y^* = \frac{\sum_{j=1}^n w_{ji} y_j}{\sum_{j=1}^n w_{ji}} = \sum_{j=1}^n \hat{w}_{ji} (f_j(x_1, \dots, x_k) + M_j) \quad (7)$$

후반부 파라미터는 최소자승법에 의해 결정된다.

3. 연속 동조 방법을 이용한 유전자적 최적화

유전자 알고리즘은 자연 선택과 유전학에 기반으로 하는 확률적인 탐색방법으로써 탐색과 해의 가능영역들을 균형 있게 이용하기 위하여 생산, 교배, 돌연변이의 과정을 수행하는 일반성 있는 탐색법으로 비선형 최적화 이론에 탁월한 성능을 발휘하고 있다[4].

본 논문에서는 퍼지 모델의 구조/파라미터 동정에 있어서

연속 동조 방법을 고려하여 최적의 퍼지 모델을 설계한다. 연속 동조 방법은 그림 1과 같이 퍼지 모델의 구조 동정과 파라미터 동정의 연속적인 검색체 구조를 가지며 구조와 파라미터를 동시에 동정한다. 구조 동정에서는 시스템의 입력 변수의 수, 선택된 입력 변수, 입력 변수당 멤버쉽함수의 수, 후반부 구조의 형태를 결정하며, 파라미터 동정에서는 멤버쉽함수의 정점을 동정한다. 여기서, 괄호안의 수는 검색체 수를 나타낸다.

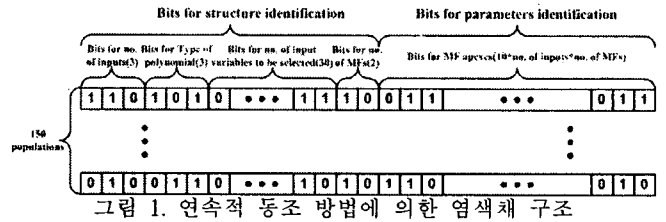


그림 1. 연속적 동조 방법에 의한 검색체 구조

연속 동조 방법은 구조/파라미터 동시 진화 문제에 있어서 구조 연산과 파라미터 연산을 분리해서 각각의 연산 선택과 선택된 해당 연산 세대를 조정해 주는 방법이다. 구조/파라미터의 연산 구분은 교배와 돌연변이 연산에서 선택되며 교배 연산시 교배점으로 구조 부분이 선택되면 구조 연산이 수행되며, 파라미터 부분이 선택되면 파라미터 연산이 수행된다. 돌연변이 연산도 마찬가지로 수행된다. 구조/파라미터의 연산의 세대 할당에 대해서는 가변 세대 할당을 수행한다. 즉, 모든 개체에 대해서 세대가 지나감에 따라 파라미터 연산의 비중을 높이면서 진화한다.

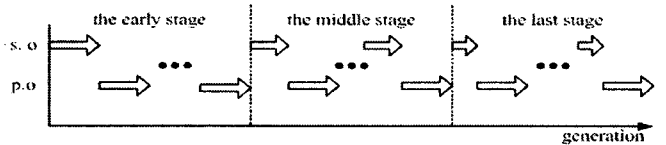


그림 2. 가변 세대 기반 구조/파라미터 진화 연산 (s.o : structure operation, p.o : parameter operation)

가변 세대 기반 연속 동조 방법을 이용한 퍼지 모델의 최적화 과정에서 각 세대마다 구조에 따른 동정된 멤버쉽함수의 정점은 정보 입자에 의한 초기 파라미터를 최적적으로 동조하게 되며 퍼지 규칙의 후반부 중심값은 그에 따라 적응적으로 변화하게 되어 최적의 퍼지 모델을 설계한다.

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)를 이용한다.

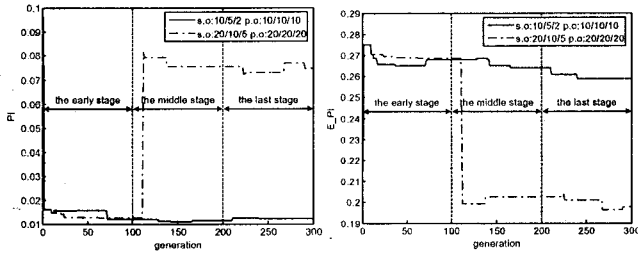
가스로 시계열 데이터는 입출력 데이터인 가스 흐름을 $u(t)$ 과 연소된 이산화탄소 농도 $y(t)$ 의 가스로 공정으로부터 추출된다. 시스템의 입력 변수를 동정하기 위해서 $u(t-3)$, $u(t-2)$, $u(t-1)$, $y(t-3)$, $y(t-2)$, $y(t-1)$ 의 6입력을 적용한다. 출력은 $y(t)$ 이다. 또한 전체 시스템 입출력 데이터 쌍 296개 중 처음 148개의 데이터는 학습 데이터로 이용하고, 나머지 148개의 데이터는 테스트 데이터로 나누어 모델을 평가한다.

퍼지 모델의 구조/파라미터 동정을 위한 진화 연산에서 가

변 세대 할당 비율이 s.o:10/5/2, p.o:10/10/10인 경우, 입력 변수는 $u(t-3)$, $y(t-1)$ 이 선택되었고 멤버쉽함수의 수는 각 입력 변수에 대해 각 5개이고, 후반부 구조는 구조 3이 결정되었으며, 가변 세대 할당 비율이 s.o:20/10/5, p.o:20/20/20인 경우에는 입력 변수가 $y(t-2)$, $y(t-1)$ 이 선택되었고 멤버쉽함수의 수는 각 입력 변수에 대해 5개이고, 후반부 구조는 구조 4가 결정되었다. 표 1은 연속 동정 방법에 의한 제안된 퍼지 모델의 구조와 그에 따른 성능지수를 보여주며, 그림 3은 제안된 퍼지 모델의 최적화 과정을 보여준다.

표 1. 제안된 퍼지 모델의 성능지수 ($\theta=0.5$)

세대 할당	input variable	No. of MFs	Type	PI	E-PI
s.o:10/5/2 p.o:10/10/10	$u(t-3)$ $y(t-1)$	5+5	3	0.012	0.258
s.o:20/10/5 p.o:20/20/20	$y(t-2)$ $y(t-1)$	5+5	4	0.074	0.197



(a) PI (b) E-PI
그림 3. 제안된 퍼지 모델의 최적화 탐색과정

표 2는 기존의 퍼지모델과 동정 오차를 비교하여 보여준다. 제안된 퍼지 모델이 기존의 퍼지 모델보다 성능이 향상된 것을 알 수 있다.

표 2 기존 퍼지모델과 동정 에러 비교

Model	PI _t	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		0.020	0.264	6
Our model		0.012	0.258	10

5. 결론

본 논문에서는 비선형 공정에 대해 체계적이고 효율적인 모델을 얻기 위하여 연속 동조 방법에 의한 퍼지 집합 퍼지 모델의 유전자적 최적화를 제안하였다. 퍼지 모델의 초기 구조를 동정하기 위해 유전자 알고리즘을 이용하였고, 각 입력에 대해 퍼지 공간을 분할하여 퍼지 모델을 구축하였으며, C-Means 클러스터링 알고리즘에 의한 정보 입자는 퍼지규

칙의 전,후반부에 각각 사용될 멤버쉽함수의 초기 정점 및 방향식 함수의 초기값과 같은 퍼지모델의 초기 파라미터를 결정하였다. 유전자 알고리즘을 이용하여 최적화 모델을 동정하는 데 있어서 시스템의 구조/파라미터 동정을 위해 연속 동조 방법을 제시하였고 가변 세대 할당을 기반으로 진화연산을 하였다. 제안된 모델은 입력력 특징을 이용함으로써 복잡하고 비선형이 강한 공정에 기존의 퍼지 모델들 보다 성능이 향상된 퍼지 모델을 설계할 수 있었고, 객관적인 모델링을 할 수 있었다.

감사의 글

본연구는 산업자원부의 지원에 의하여 기초전력연구원(R-2007-2-044) 주관으로 수행된 과제임.

참고 문헌

- [1] _____, "Fuzzy logic = Computing with words," IEEE Trans. Fuzzy Syst., Vol. 4, No. 2, pp. 103-111, 1996.
- [2] W. Pedrycz and G. Vukovich, "Granular neural networks," Neurocomputing, Vol. 36, pp. 205-224, 2001.
- [3] P. R. Krishnaiah and L. N. Kanal, editors. Classification, pattern 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 experimental data," Fuzzy Sets Syst., Vol. 13, pp.1-12, 1980.
- [6] W. Pedrycz, "An identification algorithm in fuzzy relational system," Fuzzy Sets Syst., Vol. 13, pp.153-167, 1984.
- [7] C. W. Xu and Y. Zailu, "Fuzzy model identification self-learning 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 numerical data," IFSA'91 Brussels, Copmputer, Management & System Science, pp.264-267, 1991.
- [9] S.-K. Oh and W. Pedrycz, "Identification of Fuzzy Systems by means of an Auto-Tuning Algorithm and Its Application to 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 Fuzzy Models with the Aid of Evolutionary Data Granulation," IEE Proc.-Control Theory & Applications, Vol. 148, Issue 05, pp. 406-418, Sept. 2001.