

정보 입자 기반 퍼지 모델의 하이브리드 동정

박건준*, 이동윤**, 오성권*
 *수원대학교 공과대학 전기공학과
 **중부대학교 정보통신학과

Hybrid Identification of IG based Fuzzy Model

Keon-Jun Park*, Dong-Yoon Lee** and Sung-Kwon Oh*
 *Dep. of Electrical Engineering, The University of Suwon.
 **Dep. of Information and Communications, Joongbu University

Abstract - We introduce a hybrid identification of information granulation(IG)-based fuzzy model to carry out the model identification of complex and nonlinear systems. To optimally design the IG-based fuzzy model we exploit a hybrid identification through genetic algorithms(GAs) and Hard C-Means (HCM) clustering. An initial structure of fuzzy model is identified by determining the number of input, the selected input variables, the number of membership function, and the conclusion inference type by means of GAs. Granulation of information data with the aid of HCM clustering help determine the initial paramters of fuzzy model such as the initial apexes of the membership functions and the initial values of polyminial functions being used in the premise and consequence part of the fuzzy rules. And the inital parameters are tuned effectively with the aid of the GAs and the least square method. Numerical example is included to evaluate the performance of the proposed model.

1. 서 론

많은 연구자들은 자연 현상을 이해하고 그것을 이용하기 위한 연구가 이루어지고 있다. 그 중 1965년에 Zadeh[1]에 의해 창안된 "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 있어 동적으로 변화하는 환경에서 적용적으로 대처할 수 있는 퍼지 모델을 구축하기가 힘들다. 이러한 문제를 해결하기 위한 방법으로 본 논문에서는 정보 입자 기반 퍼지 모델의 하이브리드 동정을 제안한다.

본 논문에서는 정보 입자[2, 3]의 특성에 맞는 퍼지 모델을 설계 하고 최적의 퍼지 모델을 동정하기 위하여 하이브리드 동정을 시행한다. 우선, 유전자 알고리즘[5]에 의하여 시스템의 입력 수, 선택된 입력 변수, 멤버십함수의 수, 그리고 후반부 추론 형태를 결정하여 초기 퍼지 모델을 동정한다. 다음으로, 데이터들간의 근접성을 기준으로 한 HCM 클러스터링[4]을 이용하여 입출력 데이터의 중심값을 이용하여 전반부 멤버십 함수의 초기 정점과 후반부 다항식 함수의 중심값을 결정한다. 마지막으로, 앞서 정의된 전반부 멤버십 파라미터를 유전자 알고리즘을 이용하여 최적으로 동정한다. 제안된 퍼지 모델의 성능을 평가하기 위해서 Box와 Jenkins가 사용한 가스로 공정 데이터[6]를 이용한다.

2. 정보 입자

정보 입자[2, 3]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 진화컴퓨팅의 융합, 결합 및 확장을 통해 데이터의 전처리, 지능모델의 초기 구조 또는 파라미터를 결정함으로써 정보 Data 특성을 효과적으로 반영하게 된다.

3. 하이브리드 동정

3.1 HCM 클러스터링

클러스터링이란 데이터의 분류를 위해서 사용되는 것으로 데이터의 내부가 비슷한 패턴, 속성, 형태 등의 기준을 통해 데이터를 분류하여 내부의 구조를 찾아내는 것이다. HCM 클러스터링[4]은 데이터들간의 거리를 기준으로 근접한 정도를 측정하여 데이터를 특성별로 분류하여 데이터들의 특성을 파악하는데 사용된다. HCM 클러스터링에 의한 데이터 분류는 다음과 같다.

[단계 1] 클러스터 개수 ($2 \leq c \leq n$)를 결정하고, 소속행렬 U를 초기화한다.

[단계 2] 각각의 클러스터에 대한 중심벡터를 구한다.

[단계 3] 각각의 클러스터 중심과 데이터와의 거리를 계산하여 새로운 소속행렬을 생성한다.

[단계 4] 종료 조건을 만족한다면 종료하고, 그렇지 않으면 [단계 2]로 간다.

본 논문에서는 HCM 클러스터링을 통해 입출력 데이터의 중심값을 이용하여 퍼지 모델의 전반부 멤버십 함수의 초기 정점을 동정하고, 후반부 입출력 데이터의 중심값을 적용하여 정보 입자 기반 퍼지 모델을 구축한다.

3.2 유전자 알고리즘

유전자 알고리즘[5]은 자연 선택과 유전학에 기반으로 하는 확률적인 탐색방법으로써 탐색과 해의 가능영역을 균형 있게 이용하기 위하여 생산, 교배, 돌연변이의 과정을 수행하는 일반성 있는 탐색법으로 비선형 최적화 이론에 탁월한 성능을 발휘하고 있다. 기존의 다른 탐색 방법들과 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 초기 구조를 찾아낸다. 또한 초기 구조에서 앞서 정의된 전반부 멤버십 파라미터를 최적으로 동정한다.

4. 정보 입자 기반 퍼지 모델

퍼지 동정이란 if-then 형식으로 플랜트를 기술하는 것으로, 구체적으로 입출력 데이터의 상호 관계에 의해 설정된 입출력 변수로부터 확립되는 것이다. 제안된 퍼지 모델은 클러스터에 의한 입출력 데이터의 중심값으로 퍼지 공간을 분할하고 입출력 관계에 의해 퍼지 규칙을 형성한다. 그리고 각각의 퍼지 공간(규칙수)에 속한 입출력 데이터 쌍으로부터 후반부 입출력 중심값을 결정하게 된다.

4.1 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 임률적으로 정의하였으나 이는 데이터들이 가지고 있는 특성을 제대로 반영하지 못하는 단점이 있다. 그래서 전반부 파라미터 동정을 위해 HCM 클러스터링에 의해 초기 멤버십 함수의 정점을 동정한다.

HCM 클러스터링을 통한 데이터 입자에 의한 전반부 동정은 다음과 같다.

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.

$$X_k = \{x_k; y\} \quad (1)$$

여기서, X_k 는 k 번째 입력 데이터와 출력 데이터의 데이터 집합이고, $x_k = [x_{1k}, \dots, x_{mk}]^T$, $y = [y_1, \dots, y_m]^T$ 이며, $k=1, 2, \dots, l$ 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.

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

여기서, $k=1, 2, \dots, l$, $g = 1, 2, \dots, c$.

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

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

4.2 후반부 동정

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

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

[단계 1-1] 각 입력 변수의 분할 매트릭스로부터 각 클러스터(정보 입자)에 포함된 입력 데이터를 찾는다.

[단계 1-2] j 번째 규칙의 퍼지 공간에 속한 입력 데이터 쌍을 찾는다.

[단계 1-3] 앞서 찾은 입력 데이터의 쌍으로부터 그에 대응하는 출력 데이터를 찾는다.

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

$$V_j = \{V_{1j}, V_{2j}, \dots, V_{kj}; M_j\} \quad (3)$$

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

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

후반부 구조로는 구조 1(간략 퍼지추론), 구조 2(선형 퍼지추론)와 구조 3(2차식 퍼지추론), 그리고 구조 4(변형된 2차식 퍼지추론)를 이용한다. 본 논문에서는 구조 4만을 설명한다.

변형된 2차식 퍼지추론은 입력 변수의 2차항이 생략된 구조 3의 변형된 형태로써, 정보 입자 기반 퍼지 모델은 식 (4)와 같은 구형 규칙으로 구성된다.

$$R^j: \text{If } x_1 \text{ is } A_{1j} \text{ and } \dots \text{ and } x_k \text{ is } A_{kj} \text{ then } y_j - M_j = f_j(x_1, \dots, x_k) \quad (4)$$

$$f_j(x_1, \dots, x_k) = a_{j0} + a_{j1}(x_1 - V_{1j}) + \dots + a_{jk}(x_k - V_{kj}) + a_{j(k+1)}(x_1 - V_{1j})(x_2 - V_{2j}) + \dots + a_{j(k(k+1)/2)}(x_{k-1} - V_{(k-1)j})(x_k - V_{kj}) \quad (5)$$

퍼지 추론에 의해 추론된 값 y^* 은 식 (6)과 같다.

$$y^* = \frac{\sum_{j=1}^n w_j y_j}{\sum_{j=1}^n w_j} = \frac{\sum_{j=1}^n \hat{w}_j (f_j(x_1, \dots, x_k) + M_j)}{\sum_{j=1}^n \hat{w}_j} \quad (6)$$

후반부 파라미터는 최소자승법에 의해 결정된다.

5. 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 (7)을 이용한다.

$$PI = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (7)$$

가스로 시계열 데이터는 입출력 데이터인 가스 흐름을 $u(t)$ 와 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.

입력데이터의 Min_Max값을 이용한 퍼지 모델과 제안된 정보 입자 기반 퍼지 모델을 설계하기 위해 두 가지 퍼지 모델에 대해 6입력 시스템을 2입력 시스템으로 설정하였다. 주어진 시스템으로부터 입력 변수와 멤버십함수의 수 그리고 후반부 구조를 결정하였다.

Max_Min 기반 퍼지 모델의 경우 입력 변수는 $y(t-2)$ 와 $y(t-1)$ 이, 정보 입자 기반 퍼지 모델 경우에는 입력 변수가 $u(t-3)$ 과 $y(t-1)$ 이 각각 선택되었다. 따라서, 같은 구조의 퍼지 모델을 비교하기 위해 입력 변수가 $y(t-2)$ 와 $y(t-1)$ 인 경우와 $u(t-3)$ 과 $y(t-1)$ 인 경우에 대해 각각 퍼지 모델을 설계하였다. 입력 변수가 $y(t-2)$ 와 $y(t-1)$ 인 경우에는 멤버십함수의 수가 3개, 2개이고, 후반부 구조는 구조 2가 결정되었으며, 입력 변수가 $u(t-3)$ 과 $y(t-1)$ 인 경우에는 멤버십함수의 수가 각 3개이고, 후반부 구조는 구조 4가 결정되었다. 구조 동정된 퍼지 모델에 대해 전반부 멤버십 파라미터를 최적으로 동정하였다. 표 1은 Max_Min 기반 퍼지 모델과 정보 입자 기반 퍼지 모델의 성능지수를 보여준다.

표 1을 보면 구조 동정에서 Max_Min 기반 퍼지 모델의 성능보다 정보 입자 기반 퍼지 모델의 성능이 나은 것을 알 수 있으며, 전반부 파라미터를 최적으로 동정한 후에도 향상된 것을 알 수 있다.

그림 1은 입력 변수가 $u(t-3)$ 과 $y(t-1)$ 에 대해 정보 입자 기반 퍼지 모델의 전반부 멤버십 파라미터가 최적으로 동정된 것을 보여준다.

그림 2는 입력 변수가 $u(t-3)$ 과 $y(t-1)$ 에 대해

Max_Min 기반 퍼지 모델과 정보 입자 기반 퍼지 모델의 최적화 탐색과정을 보여주고 있다. 정보 입자 기반 퍼지 모델의 경우 입출력 데이터의 초기 중심값에 의한 영향으로 세대 초기부터 좋은 성능을 나타내는 것을 알 수 있다.

표 1. Max_Min 기반 퍼지 모델과 정보 입자 기반 퍼지 모델의 성능지수

Model	Identification	input variable	No. of MFs	Type	M_PI	PI	E_PI
Max/Min_FIS	Structure*	y(t-2)	3x2	Type 2	0.152	0.092	0.212
	Parameters*				0.145	0.086	0.203
IG_FIS	Structure	y(t-1)	3x2	Type 2	0.151	0.085	0.218
	Parameters	0.143			0.087	0.199	
Max/Min_FIS	Structure	u(t-3)	3x3	Type 4	0.157	0.017	0.297
	Parameters				0.143	0.016	0.270
IG_FIS	Structure*	y(t-1)	3x3	Type 4	0.148	0.015	0.281
	Parameters*	0.139			0.015	0.263	

*유전자 알고리즘에 의한 구조 동정

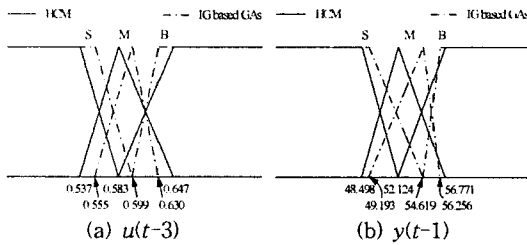


그림 1. 정보 입자 기반 퍼지 모델의 동정된 멤버십 파라미터(u(t-3), y(t-1))

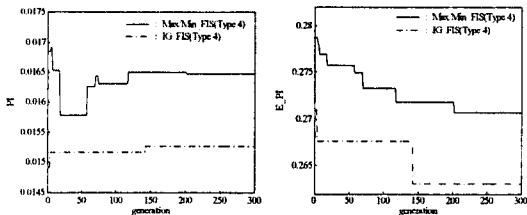


그림 2. Max_Min 기반 퍼지 모델과 정보 입자 기반 퍼지 모델의 최적화 탐색과정(u(t-3), y(t-1))

표 2는 기존의 퍼지모델과 제안된 퍼지 모델과의 동정 오차를 보여준다. 본 논문에서 제안된 퍼지 모델은 표에서 보는 바와 같이 기존의 퍼지 모델보다 향상된 것을 알 수 있다.

표 2. 기존 퍼지모델과 동정 오차 비교

Model	PI _i	PI	E_PI	규칙수
Tong's model[7]	0.469			19
Pedrycz's model[8]	0.776			20
Xu's model[9]	0.328			25
Sugeno's model[10]	0.355			6
Oh et al.'s model[11,12]	구조 1	0.024	0.328	4
	구조 2	0.022	0.326	4
HCM+GA [13]	구조 1	0.021	0.364	6
		0.035	0.289	4
	구조 2	0.022	0.333	6
		0.026	0.272	4
	0.020	0.264	6	
Our model		0.015	0.263	9

5. 결 론

비선형 공정에 대해 체계적이고 효율적인 모델을 얻기 위한 퍼지 모델을 제안하였다. 제안된 퍼지 모델은 비선형 공정을 최적으로 모델링하기 위하여 하이브리드 동정을 시행하였다. 입출력 데이터의 특징을 모델에 반영하기 위해 클러스터링 알고리즘에 의한 중심값으로 퍼지 입력 공간을 정의하였고, 또한 후반부 다항식 함수에 의한 정보 입자 기반 구조 동정을 하였다. 체계적이고 효율적인 퍼지 모델을 설계하기 위해 유전자 알고리즘을 이용하여 구조 및 파라미터 동정을 하였다. 제안된 퍼지 모델은 입출력 특징을 이용함으로써 복잡하고 비선형이 강한 공정에 기존의 퍼지 모델을 보다 성능이 향상된 퍼지 모델을 설계할 수 있었고, 인간의 경험적 방법이 아닌 보다 체계적인 방법에 의한 객관적 모델을 획득할 수 있었다.

감사의 글

본 연구는 산업자원부의 지원에 의하여 기초전력연구원 (R-2003-0-285) 주관으로 수행된 과제임.

[참 고 문 헌]

- [1] L. A. Zadeh, "Fuzzy sets," Inf. Control 8, pp.338-353, 1965.
- [2] ———, "Fuzzy logic = Computing with words," IEEE Trans. Fuzzy Syst., Vol. 4, No. 2, pp. 103-111, 1996.
- [3] W. Pedrycz and G. Vukovich, "Granular neural networks," Neurocomputing, Vol. 36, pp. 205-224, 2001.
- [4] 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.
- [5] D. E. Goldberg, "Genetic Algorithm in search, Optimization & Machine Learning," Addison wesley, 1989.
- [6] G. E. P. Box and G. M. Jenkins, Time Series Analysis, Forecasting, and Control, 2nd edition Holden-Day, SanFrancisco, 1976.
- [7] R. M. Tong, "The evaluation of fuzzy models derived from experimental data," Fuzzy Sets Syst., Vol. 13, pp.1-12, 1980.
- [8] W. Pedrycz, "An identification algorithm in fuzzy relational system," Fuzzy Sets Syst., Vol. 13, pp.153-167, 1984.
- [9] 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.
- [10] M. Sugeno, T. Yasukawa, "Linguistic modeling based on numerical data," IFSA'91 Brussels, Copmuter, Management & System Science, pp.264-267, 1991.
- [11] 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.
- [12] 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.
- [13] B.-J. Park, W. Pedrycz and S.-K. Oh, "Identification of Fuzzy Models with the Aid of Evolutionary Data Granulation," IEE Proc.-Control Theory and Applications, Vol. 148, Issue 05, pp. 406-418, Sept. 2001.