

# 병렬유전자 알고리즘을 기반으로한 퍼지 시스템의 동정

## Identification of Fuzzy System Driven to Parallel Genetic Algorithm

최정내\*, 오성권\*\*

(Jeoung-Nae Choi, Sung-Kwun Oh)

**Abstract** - The paper concerns the successive optimization for structure and parameters of fuzzy inference systems that is based on Parallel Genetic Algorithms (PGA) and information data granulation (IG). PGA is multi-population based genetic algorithms, and it is used to optimize structure and parameters of fuzzy model simultaneously. The granulation is realized with the aid of the C-means clustering. The concept of information granulation was applied to the fuzzy model in order to enhance the abilities of structural optimization. By doing that, we divide the input space to form the premise part of the fuzzy rules and the consequence part of each fuzzy rule is newly organized based on center points of data group extracted by the C-Means clustering. It concerns the fuzzy model-related parameters such as the number of input variables to be used in fuzzy model, a collection of specific subset of input variables, the number of membership functions according to used variables, and the polynomial type of the consequence part of fuzzy rules. The simultaneous optimization mechanism is explored. It can find optimal values related to structure and parameter of fuzzy model via PGA, the C-means clustering and standard least square method at once. A comparative analysis demonstrates that the proposed algorithm is superior to the conventional methods.

**Key Words** : Parallel genetic algorithm, fuzzy model, information granulation, successive tuning

### 1. 서론

비선형적이고 다변수인 시스템을 대상으로 한 퍼지 모델링 기법의 유용성은 이미 잘 알려져 있으며 이들은 퍼지추론 시스템에 기초하고 있다. 퍼지모델의 성능은 퍼지규칙의 구성 방법에 의존하며 보다 좋은 성능을 위해서는 퍼지규칙의 동정이 필연적이다. 1980년대 초에 언어적 접근 방식과 퍼지 관계식에 기초한 접근 방식이 제안되었다. 후반부에 선형식을 사용하는 Sugeno-Takagi 모델은 정교한 규칙 기반 모델의 초석이 되었으며, 콤플렉스법과 최소자승법에 의해 동정하는 방법이 제안되었다[2].

Pedrycz 와 Oh는 진보된 퍼지모델들을 제안하고, 유전자 알고리즘을 이용하여 구조와 전반부 파라미터를 최적화 하는 방법을 제시하였다[3-6]. 많은 연구에도 불구하고 퍼지모델의 동정에 관한 문제는 여전히 큰 관심 대상으로 남아있다.

본 논문은 퍼지모델의 최적화에 중점을 두었으며, 최적화를 위해서 계층적 경쟁 기반 병렬유전자 알고리즘(HFCGA)을 개발하고 적용하였다. 일반적으로 유전자 알고리즘은 전역해를 찾을 수 있는 최적화 알고리즘으로 잘 알려져 있으며, 많은 응용에 성공적으로 사용되고 있다. 그렇지만 탐색공간이 크거나, 파라미터 수가 많은 경우 세대가 증가해도 더 좋은 최적해를 찾지 못하고 지역해에 빠지는 조기수렴 문제를 내재하고 있다. 조기 수렴을 억제하고 전역해를 찾기

위한 해결방안으로써 병렬유전자 알고리즘이 개발되었으며 [6], 특히 유전자 프로그래밍에서 이주모델중의 하나인 계층적 경쟁 경쟁 (HFC) 구조는 큰 탐색 공간을 갖는 문제에 효과적이다[7].

퍼지모델의 설계는 구조와 파라미터의 최적화를 통하여 수행된다. 구조/파라미터의 동정 방법론 또한 하나의 큰 관심 대상이며, 이에 대한 많은 연구가 이루어졌다. 대부분의 연구는 구조 또는 파라미터만을 동정하거나 단계적으로 수행된다. 본 논문에서는 HFCGA를 이용하여 정보입자 기반 퍼지 모델의 구조와 파라미터를 동시에 동정할 수 있는 연속동정 방법론을 제시하고 기존의 단계별로 수행했던 개별동정 방법과 비교를 통하여 제안된 동정방법론뿐 아니라 HFCGA가 우수한 성능을 가짐을 보였다.

### 2. 정보 입자 기반 퍼지모델

#### 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. 정보 입자는 클러스터링, Granular Computing 및 진화컴퓨팅의 융합, 결합 및 확장을 통해 데이터의 전처리, 지능모델의 초기 구조 또는 파라미터를 결정함으로써 정보 데이터의 특성을 효과적으로 반영하게 된다. 본 논문에서는 데이터들 간의 거리를 기준으로 근접한 정도를 측정하여

저자 소개

\* 최정내 : 水原大學 電氣工學科 博士

\*\* 오성권 : 水原大學 電氣工學科 教授

데이터를 특성별로 분류하는 C-means 클러스터링 알고리즘[3]을 이용한다. 본 논문에서는 C-means 클러스터링을 통해 입출력 데이터의 중심값을 이용하여 퍼지모델의 전반부 멤버쉽함수의 초기 정점을 동정하고, 후반부 입출력 데이터의 중심값을 적용하여 정보 입자 기반 퍼지모델을 구축한다.

### 2.2 전반부 동정

퍼지 모델링에서 전반부 동정, 즉 구조 동정 및 파라미터 동정은 비선형 시스템을 표현하는데 있어서 매우 중요하다.

전반부 파라미터 동정을 위해 C-means 클러스터링에 의해 초기 멤버쉽함수의 정점을 동정한다.

C-means 클러스터링을 통한 정보 입자에 의한 전반부 동정은 다음과 같다.

전체 데이터 집합  $U = \{x_1, x_2, \dots, x_l; y\}$ 이고, 여기서  $x_k = [x_{1k}, \dots, x_{mk}]^T$ ,  $y = [y_1, \dots, y_m]^T$ ,  $l$ 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,  $k=1, 2, \dots, l$ 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.

$$v_{kG} = \{U_{k1}, U_{k1}, \dots, U_{kc}\} \quad (2)$$

[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 멤버쉽함수의 초기 정점으로 설정한다.

### 2.3 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 2] 각 규칙에서 산술 평균에 의한 데이터 집합의 중심 벡터  $V_j$ 를 계산한다.

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

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

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

정보입자를 기반으로 퍼지집합 모델은 후반부 다항식의 형태에 따라 다음처럼 4가지가 있다.

$$f_j = 1 + c_1 u_{1k} + c_2 u_{2k} + \dots + c_m u_{jk} + f_j(1 + c_k) \quad (4)$$

Typ 1 : 간략식  
 $f_j = 1 + c_j u_{jk}$

Type 2 : 선형식

$$f_j = 1 + c_j u_{jk} + 0 + c_j(1 + c_j u_{jk}) \quad (5)$$

Type 3 : 2차식

$$f_j = 1 + c_j u_{jk} + 0 + c_j(1 + c_j u_{jk}) + c_j(2k + 1)u_{jk} + c_j(2k)u_{jk}^2 \quad (6)$$

Type 4 : 변형된 2차식

$$f_j = 1 + c_j(1 + c_j u_{jk}) + c_j(2k + 1)u_{jk} + c_j(2k)u_{jk}^2 + c_j(2k + 1)u_{jk}^2 + c_j(2k + 2)u_{jk}^2 / (2k - 1) + c_j(1 + c_j u_{jk}) \quad (7)$$

### 3. HFCGA를 이용한 정보입자 기반퍼지모델의 동정

퍼지모델의 최적화는 구조동정과 파라미터 동정을 통하여 이루어진다. 퍼지모델의 구조에 관해서는 전체 입력 중 몇 개의 입력을 사용할 것인가? 어떤 입력변수를 사용할 것인가?, 각 선택된 입력에 대하여 몇 개의 멤버쉽함수를 사용할 것인가? 그리고 퍼지규칙의 후반부 다항식은 어떤 타입을 사용할 것인지를 결정해야한다. 퍼지모델의 파라미터에 관해서는 퍼지규칙의 후반부 다항식의 계수와 전반부의 멤버쉽함수의 정점이 있다. 후반부 다항식의 계수는 LSE를 통하여 구해지며, 구조와 멤버쉽함수의 정점(파라미터)은 HFCGA를 통하여 구해진다.

개별동정은 구조동정과 파라미터 동정이 순서적으로 수행된다. 이때 구조동정 시에 멤버쉽함수의 정점으로써 C-Means 클러스터링에 의하여 얻어진 중심값을 사용한다. 파라미터 동정에서는 구조동정으로부터 얻어진 모델의 멤버쉽함수의 정점을 동정한다. 개별동정은 고정된 하나의 파라미터만을 고려하여 구조동정이 수행되고, 고정된 하나의 구조에 대하여 파라미터 동정이 수행되기 때문에 최적화과정에서 탐색되는 구조/파라미터의 영역은 매우 한정적이며 더 좋은 모델을 구할 수 있음에도 불구하고 방법론적인 제약으로 인하여 탐색조차 되지 않은 경우가 있다.

연속동정에서는 구조와 파라미터에 관련된 정보의 하나의 염색체에 포함함으로써 다양한 구조에 대하여 다양한 파라미터를 고려해 볼 수 있다. 즉 구조/파라미터에 대한 탐색범위가 훨씬 많아지게 되어 최적의 모델을 찾을 수 있는 확률이 많아진다.

그림1에 HFCGA에 기반을 둔 연속동정법에서 사용되는 염색체의 구조를 보인다.

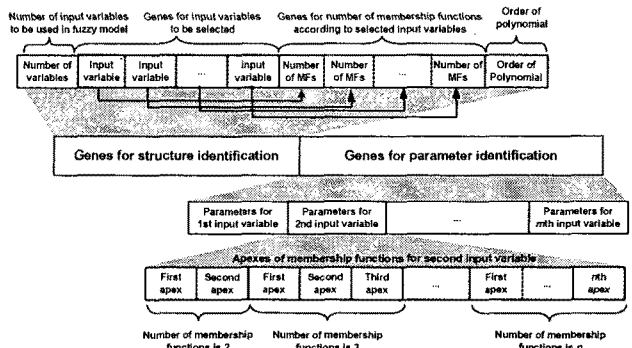


그림 1. 연속동정법에서 사용되는 염색체의 구성

#### 4. 시뮬레이션

다음과 같은 2개의 입력변수를 갖는 비선형 시스템을 고려한다.

$$y = 1.3 + x_1 - x_2 \sin(x_1 - 0.2) \sin(2x_2)$$

퍼지모형을 구축하기 위하여 랜덤하게 60개의 데이터를 취득하였으며, 40개의 데이터를 모형을 구축하기 위한 Training data로 사용하였고, 나머지 20개의 데이터는 모형을 평가하기 위한 Testing data로 사용하였다. PGA에서 Training data는 20개씩 나누어 각 염색체에 대한 모델의 학습과 타당성을 검토하기 위해 사용되었다.

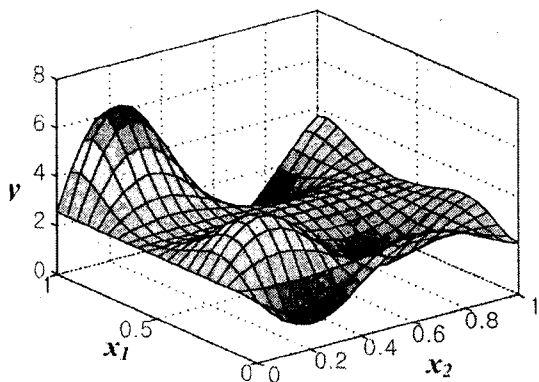
모델의 성능지수는 MSE(Mean square error)를 사용하였고, PI는 학습데이터에 대한 성능지수, VPI는 각 개체에서 구해진 모델에 대한 성능평가 지수이며, EPI는 최종적으로 구해진 정보입자 기반 모델에 대한 성능평가 지수이다.

HFCGA에서 목적함수의 값은 overfitting의 방지를 위해  $(PI+VPI)/2$ 를 사용하였다.

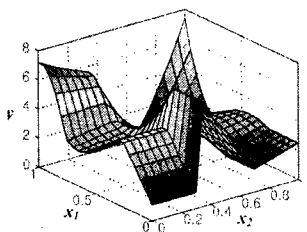
HFCGA에 기반을 두어 개별동정법과 연속동정법을 비교하여 결과를 보였다. 구축된 모델에서 사용하는 입력변수는  $x_1, x_2$ 이며 각각에 대하여 3개의 멤버쉽함수가 얻어졌으며, 후반부 다항식은 개별동정에서는 Type 3, 연속동정에서는 Type 4가 선택되었다. 표 1은 개별동정과 연속동정에서의 성능지수를 보이며 그림 2는 실제시스템에 대한 입출력, 개별동정에 의해 구해진 퍼지모형의 입출력, 그리고 연속동정법에 의하여 구해진 퍼지모형의 입출력을 보인다.

표 1. 개별동정법과 연속동정법에서의 성능지수

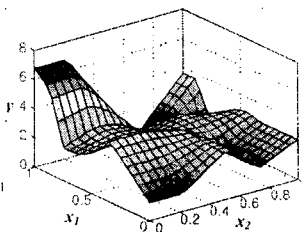
동정방법	Training		Testing
	PI	VPI	EPI
개별동정	2.17E-26	0.16	1.90
연속동정	1.23E-28	0.17	0.54



(a) 실제 시스템



(b) 개별동정법에 의한 모델



(c) 연속동정법에 의한 모델

그림 2. 실제시스템과 퍼지모형의 입출력

#### 3. 결론

본 논문에서는 비선형 공정에 대한 정보입자 기반 퍼지모델의 구축 방법을 제시한다. 퍼지모델의 구조에 관련된 입력변수의 수, 사용할 입력변수, 각 입력변수에 대한 멤버쉽함수의 수와 파라미터에 관련된 멤버쉽함수의 정점은 HFCGA에 기반을 두어 연속동정법에 의하여 동시에 동정된다. 제안된 연속동정법은 기존의 개별동정법에서 구조와 파라미터의 탐색영역이 제한되는 문제점을 해결하여 탐색영역을 확장하였고, 또한 기존의 GA를 개선한 HFCGA를 적용함으로써 더 우수한 정보입자 기반 퍼지모델을 구할 수 있다. 이 방법의 우수성은 시뮬레이션을 통하여 보였고, 입력변수가 많은 경우에는 더욱더 효과적이다.

#### 감사의 글

이 논문은 산업자원부에서 시행하는 대학전력연구센터 육성 지원사업(I-2004-0-074-0-00)에 의해 작성됨.

#### 참고 문헌

- [1] L. A Zadeh, "Fuzzy sets," *Inf. Control* 8, pp.338-353, 1965.
- [2] T. Tagaki and M. sugeno, "Fuzzy identification of system and its applications to modeling and control", *IEEE Trans. Syst. Cybern.*, Vol. SMC-15, No. 1, pp. 116-132, 1985
- [3] W. Pderycya 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*, Vol. 2 of Handbook of Statistics. North-Holland, Amsterdam, 1982.
- [5] 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.
- [6] Lin, S.C., Goodman, E., Punch, W.: Coarse-Grain Parallel Genetic Algorithms: Categorization and New Approach. *IEEE Conf. on Parallel and Distrib. Processing*. Nov. 1994
- [7] Hu, J.J., Goodman, E.: The Hierarchical Fair Competition (HFC) Model for Parallel Evolutionary Algorithms. *Proceedings of the 2002 Congress on Evolutionary Computation: CEC2002*. IEEE. Honolulu. Hawaii. (2002)
- [8] L. X. Wang, J. M. Mendel, "Generating fuzzy rules from numerical data with applications," *IEEE Trans. Systems, Man, Cybern.*, Vol. 22, No. 6, pp. 1414-1427, 1992.
- [9] J. S. R. Jang, "ANFIS: Adaptive-Network-Based Fuzzy Inference System," *IEEE Trans. System, Man, and Cybern.*, Vol. 23, No. 3, pp. 665-685, 1993.
- [10] L. P. Maguire, B. Roche, T. M. McGinnity, L. J. McDaid, "Predicting a chaotic time series using a fuzzy neural network," *Information Sciences*, Vol. 112, pp. 125-136, 1998.
- [11] C. James Li, T. -Y. Huang, "Automatic structure and parameter training methods for modeling of mechanical systems by recurrent neural networks," *Applied Mathematical Modeling*, Vol. 23, pp. 933-944, 1999.