

기호코딩기반 유전자 알고리즘을 이용한 개선된 정보입자 기반 퍼지시스템의 구조 개발

최정내, 오성권, 김현기
수원대학교 전기공학과

Structural Developments of Improved IG-based Fuzzy Systems Using Symbolic Coded Genetic Algorithms

Jeoung-Nae Choi, Sung-Kwun Oh, Hyun-Ki Kim
Department of Electrical Engineering, The University of Suwon

Abstract - 본 논문에서는 개선된 정보입자 기반 퍼지모델과 기호코딩 기반 유전자 알고리즘을 이용한 퍼지모델의 구조동정을 다룬다. 클러스터링 방법을 이용하여 초기 데이터를 분할하고 각 클러스터에 대한 중심값과 소속정도에 대한 정보가 취득되며 이 취득된 정보입자는 퍼지 모델에 적용된다. 또한 많은 입력변수를 갖는 시스템에 대하여 발생하는 고차원성 문제를 해결하기 위하여 기호코딩 기반 유전자 알고리즘을 이용하여 적절한 입력변수, 멤버십 함수의 수, 후반부 다항식의 차수를 효율적으로 선택할 수 있는 구조동정방법을 제시한다.

1. 서 론

비선형적이고 다변수인 시스템을 대상으로 한 퍼지 모델링 기법의 유용성은 이미 잘 알려져 있으며 이들은 퍼지추론 시스템에 기초하고 있다. 퍼지 모델의 성능은 퍼지규칙의 구성 방법에 의존하며 보다 좋은 성능을 위해서는 퍼지규칙의 동정이 필수적이다. 1980년대 초에 언어적 접근 방식과 퍼지 관계식에 기초한 접근 방식이 제안되었고 후반부에 선행식을 사용하는 Sugeno-T모형 모델은 정교한 규칙 기반 모델의 초석이 되었으며, Complex 방법과 Least Square Method에 의해 동정하는 방법이 제안되었다.

Pedrycz와 Oh는 진보된 퍼지모델들을 제안하고, Simple Genetic Algorithms를 이용하여 Topology과 전반부 파라미터를 최적화 하는 방법을 제시하였다. 많은 연구에도 불구하고 퍼지모델의 동정에 관한 문제는 여전히 큰 관심 대상으로 남아있다.

퍼지모델의 최적화는 크게 구조동정과 파라미터동정에 대하여 수행되며 구조동정에서는 입력변수의 선택, 각 입력변수에 대한 멤버십함수의 수 그리고 후반부 다항식의 형태를 결정하게 되고, 파라미터 동정에서는 전반부 멤버십함수의 정점을 결정하게 된다. 기존의 이진코딩, 실수코딩 기반 유전자 알고리즘에서는 구조동정시에 교배, 돌연변이 연산자의 특성으로 인하여 입력변수가 중복선택되는 문제점이 발생하여 많은 입력 변수를 갖는 시스템에서는 탐색되는 구조에 한계가 있다. 본 논문에서는 구조동정시 기호코딩과 실수코딩의 혼합 사용을 통하여 입력변수가 중복선택되는 문제점을 해결하였으며, 시뮬레이션을 통하여 제안된 방법의 우수성을 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 데이터를 간의 거리를 기준으로 근접한 정도를 측정하여 데이터를 특성별로 분류하는 C-means 클러스터링 알고리즘[3]을 이용한다. 본 논문에서는 C-means 클러스터링을 통해 입출력 데이터의 중심값을 이용하여 퍼지모델의 전반부 멤버십함수의 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 멤버십함수의 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 $x_k=[x_{k1}, \dots, x_{km}]^T$, $y=[y_1, \dots, y_m]^T$, 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 C-Means 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.

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

[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 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.

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

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.

a) Type 1 (간략 퍼지추론)

후반부가 단일 상수항만을 가지는 것으로, 이와 같은 추론법을 간략 퍼지추론법이라 한다. 퍼지모델은 식 (4)와 같은 형태를 가지는 구현 규칙들로 구성되며, 퍼지추론에 의해 추론된 값 y_j^* 는 식 (5)과 같다.

$$R^j: IF x_1 \text{ is } A_{1c} \text{ and } \dots \text{ and } x_k \text{ is } A_{kc} \text{ then } y_j = a_{j0} \quad (4)$$

$$y_j^* = \frac{\sum_{i=1}^n w_{ji} y_j}{\sum_{j=1}^n w_{ji}} = \sum_{j=1}^n \hat{w}_{ji} a_{j0} \quad (5)$$

후반부 파라미터 a_{j0} 은 최소자승법에 의해 결정된다.

$$\hat{A} = (X^T X)^{-1} X^T Y \quad (6)$$

b) Type 2 (선형 퍼지추론)

$$R^j: IF x_1 \text{ is } A_{1c} \text{ and } \dots \text{ and } x_{kc} \text{ is } A_{kc} \text{ then } y_j = f_j(x_1, \dots, x_k) \quad (7)$$

$$f_j(x_1, \dots, x_k) = a_{j0} + a_{j1}x_1 + \dots + a_{jk}x_k \quad (8)$$

$$y_j^* = \frac{\sum_{i=1}^n w_{ji} y_j}{\sum_{j=1}^n w_{ji}} = \sum_{j=1}^n \hat{w}_{ji} f_j(x_1, \dots, x_k) \quad (9)$$

c) Type 3 (2차식 퍼지추론)

$$R^j: IF x_1 \text{ is } A_{1c} \text{ and } \dots \text{ and } x_{kc} \text{ is } A_{kc} \text{ then } y_j = f_j(x_1, \dots, x_k) \quad (10)$$

$$f_j(x_1, \dots, x_k) = a_{j0} + a_{j1}x_1 + \dots + a_{jk}x_k + a_{j(k+1)}x_1^2 + \dots + a_{j(2k)}x_k^2 + a_{j(2k+1)}x_1x_2 + \dots + a_{j((k+2)(k+1)/2)}x_{k-1}x_k$$

모델의 추론된 값 y_j^* 는 식 (9)와 같다.

d) Type 4 (변형된 2차식 퍼지추론)

$$R \text{ IF } x_1 \text{ is } A_{1c} \text{ and } \dots \text{ and } x_k \text{ is } A_{kc} \text{ then } y_j = f_j(x_1, \dots, x_k)$$

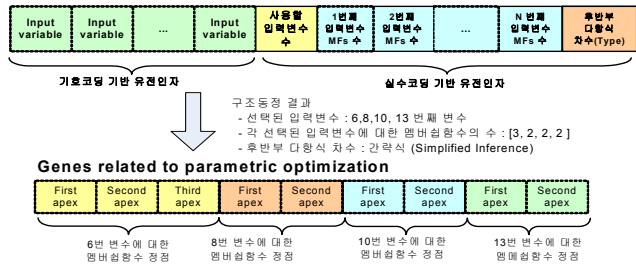
$$f_j(x_1, \dots, x_k) = a_{j0} + a_{j1}x_1 + \dots + a_{jk}x_k + a_{j(k+1)}x_1^2 + \dots + a_{j(2k)}x_k^2$$

모델의 추론된 값 y_j^* 는 식 (9)와 같다.

3. 기호코딩 기반 정보입자 기반 퍼지모델의 구조동정

퍼지모델의 최적화는 크게 구조동정과 파라미터동정에 대하여 수행된다. 구조동정에서는 주어진 전체입력을 사용하지 않고 출력에 많은 영향을 미치는 특정 입력변수만을 선택함으로써 다차원성문제를 해결한다. 구조동정에서는 모델구축에 사용될 입력변수의 선택, 각 입력변수에 대한 멤버쉽함수의 수 그리고 후반부 다항식의 형태를 결정하게 되고, 구조동정으로부터 얻어진 결과에 대하여 파라미터 동정이 수행된다. 그림1에는 구조동정과 파라미터 동정에 대한 염색체의 구성을 보인다. 구조동정에 대한 염색체는 선택된 입력변수, 입력변수의 수, 각 입력에 대한 멤버쉽함수의 수 그리고 후반부다항식의 차수에 정보를 갖는 서브염색체로 구성이 되어있으며 염색체의 길이는 모델링시에 주어진 전체입력수×2+2이다. 여기에서 선택된 입력변수에 대한 부분은 정수로 표현되는 기호코딩방식을 사용하 사용하고 나머지 세 개의 서브염색체는 실수로 표현되는 실수코딩 방식을 사용한다. 교배와 돌연변이 연산시 기호코딩에 관련된 염색체부분의 값들은 서로 중복이 되지 않도록 순회관매원문제(TSP)의 해결을 위해 제시된 부분사상교배(PMX)연산자와 상호돌연변이(RM)방법을 사용하였고, 실수코딩관련 염색체들은 산술교배(AX)와 균등돌연변이(UM)을 사용하였다.

구조동정 관련 염색체 구성



〈그림 1〉 염색체의 구성

4. 시물레이션

BostonHousing dataset을 이용하여 시물레이션을 수행하였다. 이 데이터는 Boston에 있는 거주지에 대한 범죄 발생률, 집의 연도, 크기, 방의 수, 세금, 인구밀집도, 고용센터로부터의 거리 등의 13개의 특징과 집값에 대한 데이터를 정보로 구성되어 있으며 이 데이터를 이용하여 예측 모델을 구성하고 예측모델로부터 집의 가격을 추정할 수 있다. (www.ics.uci.edu/~mllearn/MLRepository.html). 이 데이터는 13 입력 1출력의 506개 데이터로 구성이 되었으며 총 데이터의 50%는 학습에 사용되었고, 나머지 50%로는 구축된 모델의 평가에 사용되었다. 본 논문에서는 병렬 유전자 알고리즘을 사용하였으며 파라미터는 표1과 같다.

〈표 2〉 유전자 알고리즘에서의 파라미터설정

GA parameter	구조동정	파라미터동정
세대수	300	300
집단수	5	5
개체수	[50 50 50 50 50]	[80 80 80 80 80]
교배율	[0.85 0.75 0.65 0.65 0.65]	
돌연변이율	[0.1 0.1 0.1 0.1 0.1]	
이주 주기	20 세대	
선택 연산자	선형 순위기반선택	
교배연산자	PMX+ AX	MSX
돌연변이 연산자	RM + Uniform	Uniform
이주 방법	HFC method	

표 2에서는 제안된 방법에 의하여 구축된 최적화 퍼지모델의 구조 및 성능지수를 보인다. 성능지수는 RMSE를 사용하여 구하였다. 결과에서 보면 선택된 입력변수는 RM, DIS, TAX, LSTAT이다. 이는 Boston 지역에서 방의 개수, 고용센터와의 거리, 세금 그리고 인구밀집도가 주택의 가격에 많은 영향을 미침을 알 수 있고, 구축된 모델로부터 임의의 주택에 대한 가격을 예측할 수 있다.

〈표 1〉 최적화 퍼지모델의 구조 및 성능지수

	항목	동정 결과	PI	EPI
구조	선택된 입력	RM, DIS, TAX, LSTAT	3.862	3.876
	멤버쉽함수의 수	3×2×2×2		
동정	후반부 다항식 차수	Simplified (Type 1)	3.460	3.476
	파라미터동정			

5. 결 론

본 논문에서는 비선형 공정에 대한 정보입자 기반 퍼지모델의 구축 방법을 제시한다. 퍼지모델의 구조에 관련된 입력변수의 선택, 멤버쉽함수의 수, 후반부다항식의 형태는 기호코딩기반 유전자 알고리즘에 의하여 구해지며, 멤버쉽함수의 정점은 실수코딩기반 유전자 알고리즘을 통하여 동정된다. 제안된 방법은 구조동정시에 입력변수의 중복선택으로 인하여 많은 입력을 갖는 모델이 탐색되지 못하는 문제점을 해결하여 입력변수 선택시 고려되는 탐색범위를 확장하였다. 시물레이션에서 Boston Housing Data에 적용하여 모델을 구축하였다. 모델구축시 많은 입력변수가 선택되면 퍼지규칙이 기하학적으로 증가하는 고차원성문제가 발생할지라도 최적화 과정에서 이러한 경우도 고려할 수 있어 Global 최적인 퍼지모델을 구축할 수 있다.

감사의 글

이 논문은 2007년도 과학재단의 국제협력연구지원을 받아 수행된 연구임 (F01-2007-000-10058-0)

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. Pderycy 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.