

다중 퍼지 추론 모델에 의한 비선형 시스템의 최적 동정

정희열, 오성권  
원광대학교 전기·전자·정보공학부 제어계측 공학과

The optimal identification of nonlinear systems by means of Multi-Fuzzy Inference model

Hoe-Yeol Jeong and Sung-Kwun Oh  
Department of Electrical, Electronic and information Engineering, Wonkwang Univ.

**Abstract** - In this paper, we propose design a Multi-Fuzzy Inference model structure. In order to determine structure of the proposed Multi-Fuzzy Inference model, HCM clustering method is used. The parameters of membership function of the Multi-Fuzzy are identified by genetic algorithms. A aggregate performance index with a weighting factor is used to achieve a sound balance between approximation and generalization abilities of the model. We use simplified inference and linear inference as inference method of the proposed Multi-Fuzzy model and the standard least square method for estimating consequence parameters of the Multi-Fuzzy. Finally, we use some of numerical data to evaluate the proposed Multi-Fuzzy model and discuss about the usefulness.

1. 서 론

실제할 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 적응적으로 대처할 수 있는 퍼지 모델을 구축하기가 힘들다. 이러한 문제를 해결하기 위한 방법으로 본 논문에서는 다중 퍼지 추론 모델을 제안한다.

제안된 다중 퍼지 추론 모델은 시스템의 데이터 특성을 분석하기 위해 HCM 클러스터링 방법[1]을 이용하여 입·출력 데이터를 서로 유사한 특성을 가진 그룹으로 분류하고, 분류된 데이터들은 다중 퍼지 추론 모델의 설계에 사용 된다. 또한 퍼지 모델의 최적화를 위해 유전자 알고리즘[2,3]을 이용하여 멤버쉽 함수를 최적으로 동조한다. 더 나아가 학습 데이터와 테스트 데이터의 퍼지 모델 성능을 고려한 성능 결과를 위해 하중 값을 가진 합성 목적함수[4]를 사용한다. 즉 근사화 및 일반화 능력을 고려한 모델을 설계하는 것이다. 이는 모델의 성능지수가 데이터의 개수, 비선형성의 정도 등에 의해 근사화 및 예측의 정확도가 차이가 있으므로 상호 균형과 의존성을 고려하여 합성 목적함수의 하중값을 선택하므로 시스템의 최적모델을 구축하고자 한 것이다.

본 논문에서 제안된 다중 퍼지 추론 모델은 후반부를 간략 추론과 선형 추론 두 가지 형태를 사용하였으며, 멤버쉽 함수의 형태로는 삼각형 형태를 사용한다. 제안된 다중 퍼지 모델의 성능을 평가하기 위해서 비선형 시스템의 데이터를 사용한다.

2. 다중 퍼지 추론 모델

2.1 퍼지 모델

퍼지 모델링이란 If - then 형식으로 플랜트를 기술하는 것으로, 입·출력 데이터의 상호관계에 의해 설정된 입·출력변수로부터 확립되는 것이다. 규칙 베이스 퍼지 모델의 동정은 전반부와 후반부의 동정으로 나누어진다. 전반부 동정은 입력 정보, 즉 퍼지 규칙의 전반부 변수의 선택과 입력 변수 공간의 퍼지 분할 결정, 그리고 입력공간의 파라미터를 결정하는 과정이다. 이것은 비선형 시스템의 성능을 결정하는데 많은 영향을 준다. 본 논문에서 전반부 멤버쉽함수는 삼각형 형태를 이용한다. 기존의 멤버쉽함수는 입력 변수의 최소값과 최대값 사이를 임의의 개수로 등분해서 일률적으로 정의하였으나 이는 데이터들이 가지고 있는 특성을 제대로 반영하지 못하는 단점이 있다. 그래서 전반부 파라미터 동정을 위해 유전자 알고리즘을 이용한다. 즉, 유전자 알고리즘을 이용 멤버쉽 함수의 정점을 최적으로 동정함으로 위에서 언급한 문제점을 해결할 수 있다.

후반부 동정은 후반부 변수의 선택과 후반부 변수의 파라미터를 결정하는 것이다. 후반부 동정의 구조로는 구조 1(간략)과 구조 2(선형)를 사용하며, 각 구조에 대한 후반부 파라미터 동정은 다음과 같다.

구조 1(간략 추론)

퍼지 모델은 식 (1)과 같은 형태를 가지는 구현 규칙들로 구성되며, 퍼지 추론에 의해 추론 값  $y^*$ 는 식 (2)와 같다.

$$R^n: IF x_1 \text{ is } A_{n1} \text{ and } \dots \text{ and } x_k \text{ is } A_{nk} \text{ Then } a_{n0} \quad (1)$$

$$y^* = \frac{\sum_{j=1}^n w_{ji} y_i}{\sum_{j=1}^n w_{ji}} = \frac{\sum_{j=1}^n w_{ji} a_{j0}}{\sum_{j=1}^n w_{ji}} \quad (2)$$

후반부 파라미터는 최소 자승법 식 (3)에 의해 구해진다.

$$\hat{a} = (X^T X)^{-1} X^T Y \quad (3)$$

구조 2(선형추론)

퍼지 모델은 식 (4)의 형태를 가지는 구현 규칙들로 구성된다.

$$R^n: IF x_1 \text{ is } A_{n1} \text{ and } \dots \text{ and } x_k \text{ is } A_{nk} \text{ Then } a_{j0} + a_{n1}x_1 + \dots + a_{nk}x_k \quad (4)$$

선형 추론법에 의해 추론 값  $y^*$ 는 다음과 같다.

$$y^* = \frac{\sum_{j=1}^n w_{ji} y_i}{\sum_{j=1}^n w_{ji}} = \frac{\sum_{j=1}^n w_{ji} (a_{j0} + a_{n1}x_1 + \dots + a_{nk}x_k)}{\sum_{j=1}^n w_{ji}} \quad (5)$$

후반부 파라미터 동정은 구조 1의 식 (3)같이 최소 자승법에 의해 구해진다.

2.2 다중 퍼지 추론 모델

본 논문에서는 HCM 클러스터링을 이용하여 유사한 특성을 가진 그룹으로 데이터를 분류하고 그 분류된 데이터의 특성에 맞는 각각의 퍼지 모델을 그림 1과 같이 구축한다

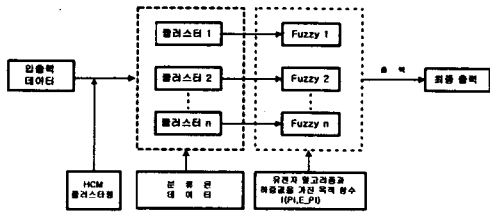


그림 1. 다중 퍼지 추론 모델의 구조  
Fig. 1. Multi-Fuzzy Inference Model structure

HCM 클러스터링에 의해 분류된 클러스터 수는 다중 퍼지 모델을 구성하는 단일 모델의 수가 되고, 분류된 데이터 집합들은 각각 기본 모델의 입·출력이 되어 단일 퍼지 모델의 기능을 수행하게 된다. 여기서 최적의 전반부 파라미터를 동정하기 위해 유전자 알고리즘을 이용한다. 또한 구축된 모델의 근사화와 일반화의 상호 균형을 위하여 하중값을 가진 합성 목적 함수를 사용한다.

### 3. HCM클러스터링 및 최적화

본 논문에서는 최적의 다중 퍼지 추론 모델을 생성하기 위해서, HCM 클러스터링 [2] 방법에 의해 분할된 데이터 그룹으로부터 각 퍼지 모델을 구축하고 모델의 멤버십 함수의 정점을 유전자 알고리즘에 의해 최적화한다. 또한 하중값을 가진 합성 성능 지수를 이용하여 퍼지 모델의 예측성능을 향상시킨다.

#### 3.1 HCM을 이용한 클러스터링

본 논문에서는 데이터들간의 거리를 기준으로 근접한 정도를 측정하여 데이터를 특성별로 분류하는 HCM 클러스터링 알고리즘을 이용하였다. 주어진 데이터 분류는 먼저 학습 데이터를 분류하고, 분류된 학습 데이터의 중심에 의해 테스트 데이터를 분류한다. HCM클러스터링에 의한 학습 데이터 분류는 다음과 같다.

[단계 1] 클러스터 개수 ( $2 \leq c \leq n$ )를 결정하고, 소속행렬  $U$ 를  $U^{(0)} \in M_c$ 으로 초기화 한다.

$$M_c = \left\{ U \mid u_{ij} \in \{0, 1\}, \sum_{k=1}^c u_{ik} = 1, 0 < \sum_{i=1}^n u_{ik} < n \right\} \quad (6)$$

여기서,  $u_{ij}$ 는 소속행렬의 파라미터

[단계 2] 각각의 클러스터에 대한 중심벡터  $V_i$ 를 구함.

$$V_i = \{v_{i1}, v_{i2}, \dots, v_{ij}, \dots, v_{im}\} \quad (7)$$

$$v_{ij} = \frac{\sum_{k=1}^n u_{ik} \cdot x_{kj}}{\sum_{k=1}^n u_{ik}} \quad (8)$$

여기서,  $i$ 는  $i$ 번째 클러스터,  $j$ 는  $j$ 번째 입력 변수 ( $j=1, \dots, m$ ),  $n$ 은 data 수.

[단계 3] 각각의 클러스터 중심과 데이터와의 거리를 계산하여 새로운 소속행렬  $U^{(r)}$ 을 생성한다.

$$d_{ik} = d(x_k - V_i) = \|x_k - V_i\| = \left[ \sum_{j=1}^m (x_{kj} - v_{ij})^2 \right]^{1/2} \quad (9)$$

$$u_{ik}^{(r+1)} = \begin{cases} 1 & d_{ik}^{(r)} = \min\{d_{ik}^{(r)}\} \text{ for all } j \in c \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

여기서,  $d_{ik}$ 는  $k$ 번째 데이터  $x_k$ 와  $i$ 번째 클러스터 중심  $V_i$ 의 기하학적 거리

[단계 4] 만일 식 (11)을 만족한다면 종료하고, 그렇지 않으면  $r=r+1$ 로 놓고 [단계 2]로 간다.

$$\|U^{(r+1)} - U^{(r)}\| \leq \epsilon \text{ (tolerance level)} \quad (11)$$

### 3.2 유전자 알고리즘 통한 모델 구조의 최적화

기존의 다른 탐색방법들은 탐색공간에서 지역극소 (local minimum)에 빠질 위험이 있지만 유전자 알고리즘은 해가 될 가능성이 있는 개체집단을 유지하면서 그들 모두가 동시에 최적값을 찾아나가기 때문에 지역극소에 빠질 위험을 어느 정도 해결할 수 있다는 것이다. 또한 선형, 연속, 미분 가능 등의 제한이 없기 때문에 다양한 분야에 별다른 제한 없이 적용할 수 있다는 장점이 있다. 본 논문에서는 이런 장점을 이용 제한된 다중 퍼지 추론 모델의 전반부 파라미터를 최적으로 동정한다.

### 3.3 하중값을 가지는 합성 목적 함수

시스템을 모델링 할 때 가장 문제가 되는 것은 구축된 모델의 근사화와 일반화이다. 따라서 본 논문에서는 근사화 및 일반화의 상호 균형과 의존 능력을 가진 합성 목적 함수 [5]를 이용 상호 연계된 최적의 모델 구조 생성에 의한 성능지수의 향상을 꾀하였다. 이를 위해 학습 및 테스트 데이터의 성능 결과에 대한 하중값을 가진 합성 목적 함수를 다음과 같이 정의한다.

$$f(PI, E\_PI) = \theta \times PI + (1 - \theta) \times E\_PI \quad (12)$$

$PI$ 는 학습 데이터,  $E\_PI$ 는 테스트 데이터에 대한 성능지수, 그리고  $\theta$ 는  $PI$ 와  $E\_PI$ 에 대한 하중값을 나타낸다. 식 (12)에서 하중값  $\theta$ 의 설정에 따라 학습데이터 및 테스트 데이터의 비중이 결정되고 이것을 기초로 모델을 최적화하는 특징을 가진다.

## 4. 실험 데이터를 통한 결과 고찰

### 4.1 가스로 공정

제안된 다중 퍼지 추론 모델에 Box와 Jenkins가 사용한 가스로 시계열 데이터 [6]를 이용하여, 입·출력 데이터인 가스 흐름율과 연소된 이산화 탄소 농도로 가스로 공정을 모델링한다. 실험을 위해 1입력 1출력의 가스로 시계열 입·출력 데이터 296쌍을 입력으로  $u(t-3), u(t-1)$ , 출력으로  $y(t)$ 를 사용하는 2입력 1출력으로 설정한다. 성능지수는 식 (13)을 이용한다.

$$PI = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2 \quad (13)$$

표 1은 단일 퍼지 모델과 본 논문에서 제안한 다중 퍼지 추론 모델의 성능지수를 비교하여 나타내고 있다.

표 1. 클러스터 개수에 따른 성능지수 (구조 2)  
Table 1. Identification errors according to the change of number of clusters

$\theta$	단일 Fuzzy		다중 Fuzzy							
			클러스터 2		클러스터 3		클러스터 4		클러스터 5	
	PI	E\_PI	PI	E\_PI	PI	E\_PI	PI	E\_PI	PI	E\_PI
0.0	0.031	0.278	0.030	0.261	0.040	0.256	0.050	0.237	0.040	0.229
0.25	0.029	0.283	0.024	0.259	0.029	0.264	0.020	0.246	0.021	0.238
0.5	0.018	0.264	0.020	0.264	0.028	0.272	0.018	0.240	0.019	0.239
0.75	0.018	0.279	0.017	0.265	0.018	0.269	0.016	0.256	0.012	0.235
1.0	0.018	0.350	0.016	0.312	0.013	0.316	0.012	0.291	0.009	0.320

표 1.에서 알 수 있듯이 본 논문에서 제안한 HCM 클러스터링을 이용한 다중 퍼지 추론 모델이 단일 퍼지 모델보다 학습 및 테스트 성능에서 우수함을 알 수 있고, 이것은 본 논문에서 제안한 방법이 데이터 특성을 반영함으로써 좀 더 우수한 모델을 구축할 수 있음을 보인 것이다. 표 2.는 유전자 알고리즘과 하중값을 가진 합성 목적 함수를 고려한 다중 퍼지 추론 모델 통해 얻은 결과

가 기존 모델에 비해 학습 및 테스트 성능이 향상됨을 보인다.

표 2. 기존 모델과 성능지수 비교

Table2. Comparison of identification errors with conventional models

모 델	추론법	규칙	PI	E.PI
Oh's model [4,7]	간 락	4	0.024	0.328
	선 형	4	0.022	0.326
		6	0.021	0.364
HCM+Hybrid[9]	간 락	6	0.022	0.332
	선 형	6	0.017	0.269
Max-Min[11]	간 락	6	0.022	0.336
	선 형	6	0.021	0.362
Our model	간 락	4	0.017	0.309
	선 형	4	0.012	0.235

#### 4.2 발전소 NOx 배출 공정

제안한 다중 퍼지 추론 모델의 성능을 평가하기 위해서 가스 터빈 발전소의 데이터[7]를 이용한다. 모델의 입력은 Tamb, COM, LPT, Pcd, Texh이고 출력은 NOx이다. NOx의 성능 평가를 위해서 식 (14)을 이용한다. 표 3.은 후반부로 구조 1을 사용한 성능지수 값이다. 또한 단일 퍼지 모델과 다중 퍼지 모델을 비교하고 있다. 표 3.에서도 알 수 있듯이, HCM 클러스터링을 이용하여 데이터의 특성에 따른 다중 퍼지 모델이 우수함을 알 수 있다.

표 3. 클러스터 개수에 따른 성능지수(구조 1)

Table3. Identification errors according to the change of number of clusters

모 델	클러스터	PI
단일 Fuzzy	1	0.8284
	2	0.2984
다중 Fuzzy	3	0.0765
	4	0.0043

표 4.는 기존의 모델과 제안한 다중 퍼지 추론 모델과의 성능지수 비교한 것이다.

표 4. 기존 모델과 성능지수 비교

Table4. Comparison of identification errors with conventional models

모 델	FNNs[11]	GMDH[11]	퍼지추론[11]	Our model
PI	0.0520	0.2484	2.2172	0.0043

표에서 나타난 바와 같이 다중 퍼지 모델이 기존의 다른 모델들 보다 성능이 우수함을 보인다.

### 3. 결 론

본 논문에서는 비선형적이고 복잡한 실 시스템의 데이터 특성을 해석하고 데이터의 특성에 따른 체계적이고 효율적인 모델을 구축하기 위하여 HCM 클러스터링, 유전자 알고리즘과 하중값을 가진 합성목적함수를 사용한 다중 퍼지 모델을 제안한다. 제안된 다중 퍼지 추론 모델은 기존 모델들의 단점인 데이터의 특성을 고려하지 않고 모델을 설계하는 것에서 벗어나 HCM 클러스터링 알고리즘을 이용하여 데이터의 특성에 맞는 모델을 설계한다. 또한 모델의 최적화를 위하여 유전자 알고리즘과 하중값을 가진 합성 목적함수를 사용한다.

제안된 모델의 실험 데이터를 결과 고찰에서 볼 수 있듯이 인간의 경험적 방법이 아닌 HCM 클러스터링을 이용한 입·출력 데이터의 분포 특성에 따른 클러스터 개수는 다중 퍼지 추론 모델을 이루는 기본 퍼지 모델의 개수를 결정한다. 이것은 복잡하고 비선형성이 강한 공 정에서 기존의 모델들 보다 성능이 향상됨을 알 수 있었다. 또한 유전자 알고리즘을 이용하여 보다 체계적인 방법에 의한 객관적인 모델을 획득할 수 있었으며, 하중값을 가진 합성 목적함수에 의해 모델의 근사화 및 일반화를 고려한 최적의 다중 퍼지 추론 모델을 생성할 수 있음을 보였다.

#### 감사의 글

이 논문은 2001년 에너지 기술 학술 진흥 사업에서 지원하여 연구하였습니다.

#### [참 고 문 헌]

- [1] 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.
- [2] David E. Goldberg. "Genetic Algorithms in search, Optimization & Machine Learning". Addison-wesley.
- [3] Zbigniew Michalewicz. "Genetic Algorithms + Data Structure = Evolution Programs", Springer-Verlag.
- [4] 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.
- [5] Box and Jenkins. "Time Series Analysis, Forecasting and Control", Holden Day, San Francisco, CA.
- [6] G. Vachtsevanos, V. Ramani and T. W. Hwang. "Prediction of Gas Turbine NOx Emissions using Polynomial Neural Network" Technical Report, Georgia Institute of Technology, Atlanta, 1995.
- [7] Sung Kwun Oh, "Fuzzy Identification by Means of an Auto-tuning Algorithm and a Weighted Performance Index", Journal of Fuzzy Logic and Intelligent Systems, Vol. 8, No. 6, pp.106-118, 1998
- [8] 오성권, 윤기찬, 김현기. "유전자 알고리즘과 합성 성능지수에 의한 퍼지-뉴럴 네트워크 구조의 최적 설계". 제어·자동화·시스템공학회, 제 6권, 제 3호, pp. 273-283, 2000년 3월.
- [9] 박병준, 윤기찬, 오성권, 장성환, "클러스터링 및 하이브리드 알고리즘을 이용한 퍼지모델의 최적화", 대한전기학회 추계 학술대회 논문지, 제 6권, pp.2908-2910, 1999년 7월.
- [10] 박병준, 오성권, 안태천, 김현기. "유전자 알고리즘과 하중값을 이용한 퍼지시스템의 최적화", 대한전기학회 논문지, 제 48권, 제 6호, pp.789-799, 1999년 6월.
- [11] 안태천, 오성권, "발전소의 대기오염물질 배출패턴 모델정립(최종보고서)", 기초전력공학 공동연구소, 과제관리번호 : 96-지-08. 1997년 8월.