

클러스터링 및 하이브리드 알고리즘을 이용한 퍼지모델의 최적화

박병준, 윤기찬, 오성권, 장성환
 원광대학교 전기전자공학부, 전라북도 익산시 신동동 344-2 ☎670-749

Optimization of the fuzzy model using the clustering and hybrid algorithms

Byoungjun Park, Ki chan Yoon, Sungkwun Oh, and Seongwhan Jang
 Division of Electrical and Electronic Engineering, Wonkwang Univ., Iksan, KOREA

Abstract - In this paper, a fuzzy model is identified and optimized using the hybrid algorithm and HCM clustering method. Here, the hybrid algorithm is carried out as the structure combined with both a genetic algorithm and the improved complex method. The one is utilized for determining the initial parameters of membership function, the other for obtaining the fine parameters of membership function. HCM clustering algorithm is used to determine the confined region of initial parameters and also to avoid overflow phenomenon during auto-tuning of hybrid algorithm. And the standard least square method is used for the identification of optimum consequence parameters of fuzzy model. Two numerical examples are shown to evaluate the performance of the proposed model.

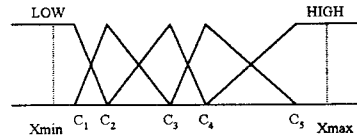
파라미터의 동정은 비선형 시스템을 표현하는데 있어서 매우 중요하다. 다시 말하면 전반부 입력 변수의 선택과 선택된 입력 변수의 공간 분할 그리고 파라미터는 비선형 시스템의 성능을 결정하는데 많은 영향을 미친다. 따라서 본 논문에서는 퍼지모델의 최적화를 위해 입출력에 의해 유기된 데이터 분포 구조를 통해 전반부 입력 변수를 결정하고 전반부 파라미터 동정은 HCM 및 하이브리드형의 최적화 알고리즘을 이용하여 동정한다. 전반부 파라미터를 동정함에 있어서 기존의 방법은 그림 1(a)처럼 입력 변수의 최소값과 최대값 사이를 임의의 개수로 등분해서 일률적으로 멤버쉽함수를 정의하였다. 그러나, 계통은 나름대로의 특성을 가지고 있기 때문에, 그 특성에 따라 멤버쉽함수를 정의하는 것이 보다 합리적인 방법이다.

1. 서 론

비선형적이고 다변수인 시스템을 대상으로 한 퍼지 모델의 성능은 퍼지 규칙의 구성 방법에 의존하며 보다 좋은 성능을 위해서는 퍼지 규칙의 동정이 필연적이다. 초기 퍼지 모델의 동정연구로 언어적 접근방식에서, Tong은 논리적 조사 방법에 의해 가스로 공정을 동정하였고 [2], 퍼지관계 방정식에 기초한 접근방식에서, Pedrycz는 퍼지관계 방정식에 의해 새로운 합성규칙인 기준 퍼지집합과 Zadeh의 제한적 가능성 분포의 개념에 기초해서 동정하였다[1]. 본 논문에서 제안된 규칙베이스 퍼지 모델은 클러스터링 알고리즘과 하이브리드 알고리즘 및 퍼지규칙의 이론을 이용함으로써 시스템 구조와 파라미터 동정을 시행한다. 또한 설계자의 의도, 성능 결과의 상호관형, 시스템의 비선형 정도 등에 따른 최적 모델 동정을 위해 하중값을 가진 목적함수[5,8]를 이용한다. 두 형태의 퍼지모델 방법은 간략추론 및 선형추론에 의해 시행되며, 삼각형 형태의 멤버쉽함수를 사용한다. 멤버쉽함수는 클러스터링 기법 중 HCM 방법을 사용하여 정의하며, HCM에 의한 파라미터를 초기값으로 하이브리드 알고리즘을 이용하여 퍼지추론 시스템의 전반부 파라미터를 최적화한다. 하이브리드 알고리즘은 유전자 알고리즘이 가지는 한계를 극복하기 위해 유전자 알고리즘과 개선된 콤플렉스 방법이 결합한 형태이다. 제안된 모델의 평가를 위해 두 개의 수치 데이터를 이용하여 기존의 경험적 방법에 의한 퍼지모델링 방법과 성능지수를 비교한다.



(a) Max-Min에 의한 균일한 정의



(b) HCM에 의한 정의
 그림 1. 멤버쉽함수의 정의

본 논문에서는 입출력 데이터의 특성을 이용하여 모델링을 수행한다. 데이터의 특성을 찾아내기 위해 논문에서는 클러스터링 알고리즘을 이용한다. 클러스터링 기법을 통해 데이터를 특성별로 분류하고, 분류된 클러스터들의 중심을 사용하여 그림 1(b)처럼 멤버쉽함수의 중심으로 정의한다.

클러스터링 알고리즘으로는 데이터들간의 거리를 기준으로 하여 근접한 정도를 측정하고, 이를 바탕으로 데이터를 분류하는 HCM(Hard C-means) 방법을 사용한다.

2.1.2 후반부 동정

퍼지 모델의 후반부 동정은 구조의 선택과 그들의 파라미터를 동정하는 문제이다. 후반부 구조로는 퍼지추론에 의해 구별되는 구조 1, 2를 사용하며 각 구조에 대한 후반부 파라미터의 동정은 아래와 같다.

구조 1(후반부 : 상수)

후반부가 단일 상수항만을 가지는 것으로, 이와 같은 추론법을 간략추론법이라 한다. 이 퍼지모델은 식(1)과 같은 형태를 가지는 구형규칙들로 구성되며, 퍼지 추론

2. 본 론

2.1 퍼지 모델의 동정

2.1.1 전반부 동정

퍼지모델링에서 전반부 동정, 즉 규칙의 구조 동정 및

에 의해 추론된 값 y^* 는 식(2)와 같다.

R^n IF x_1 is A_{n1} and ... and x_k is A_{nk} Then a_{n0} (1)

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

후반부의 파라미터는 a_{n0} 로써 입출력 데이터가 주어졌을 때 최소자승법에 의해 구해진다. 최소자승법에 의한 후반부 파라미터의 동정은 식(3)에 의해 구해진다.

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

구조 2(후반부 : 일차 선형식)

후반부가 일차 선형식으로 표현된 것으로 선형추론법 또는 혼합추론법이라 한다. 이 퍼지 모델은 식(4)의 형태를 가지는 구현규칙들로 구성된다.

R^n IF x_1 is A_{n1} and ... and x_k is A_{nk} Then $y = a_{n0} + a_{n1}x_1 + \dots + a_{nk}x_k$ (4)

선형 추론법에 의한 추론된 값 y^* 는 다음과 같다.

$$y^* = \frac{\sum_{j=1}^n w_{ji} y_j}{\sum_{j=1}^n w_{ji}} = \frac{\sum_{j=1}^n w_{ji} (a_{n0} + a_{n1}x_{1j} + \dots + a_{nk}x_{kj})}{\sum_{j=1}^n w_{ji}} \quad (5)$$

최소자승법에 의한 후반부 파라미터의 동정은 구조1과 같이 식(3)에 의해 구해진다

2.2 퍼지 모델의 최적화

2.2.1 하이브리드 알고리즘

유전자 알고리즘[6,7]은 다른 알고리즘에 비해 전역 극소를 찾을 확률이 높기 때문에 모델의 최적화 기법 중에서 뛰어난 성능을 발휘하고 있다. 그러나 유전자 알고리즘은 많은 최소점의 지역 극소를 제거하고 모델의 성능지수가 최소가 되는 전역 극소 영역을 찾지만 탐색 능력의 한계성 때문에 전역 극소점에 도달하는데 어려움을 겪는다.

또한 기하학적 개념, 즉 반사, 확장, 축소 기본개념을 바탕으로 하는 콤플렉스 방법[5,8,10]은 초기값의 영향을 많이 받는다. 즉, 잘못된 초기값 설정은 결과적으로 지역 극소에 빠지게 된다.

유전자 알고리즘과 콤플렉스 방법의 장점을 이용하여 상호 결합된 하이브리드 알고리즘은 아래 기술된 각 알고리즘의 단점 및 한계성을 극복하여 보다 최적화된 모델 구조의 생성 효과를 얻을 수 있다.

첫째, 유전자 알고리즘에서는 탐색영역이 넓어 광범위한 영역에서 최적값 추출이 용이한 반면, 콤플렉스 방법은 제한된 영역 및 경계조건에서만 멤버십함수의 최적 파라미터를 추출할 수 있다.

둘째, 유전자 알고리즘은 전역 극소 영역에서 최적값을 가지는 최적점 선택에 어려움이 있는 반면에 콤플렉스 방법은 제한된 영역에서 최적점 추출이 용이하다.

따라서 콤플렉스 알고리즘에서 사용되는 초기값을 유전자 알고리즘으로 구하여 콤플렉스 알고리즘의 초기값 문제에 따른 어려움과 유전자 알고리즘의 탐색 능력의 한계성을 극복한다.

2.2.2 하중값을 가진 목적함수

학습 데이터와 테스트 데이터를 고려한 퍼지 모델 성능은 즉 근사화 능력과 예측능력 모두를 고려하기 위해 목적함수를 다음과 같이 정의한다.

$$f(PI, E_PI) = \theta \times PI + (1 - \theta) \times E_PI \quad (6)$$

θ 는 PI와 E_PI에 대한 하중값이다. PI는 학습데이터에 대한 성능지수, E_PI는 테스트 데이터에 대한 성능지수이다. 이 목적함수는 θ 의 선택에 따라 퍼지모델의 근사화와 일반화 사이에서 최적화에 대한 방향을 설정할 수 있는 특징을 가진다[8].

본 논문에서는 수치적 실험에 이용되는 성능지수로 식(7)을 사용한다.

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

3. 시뮬레이션 및 결과 고찰

3.1 가스로공정

제안된 규칙 베이스 퍼지모델링을 Box와 Jenkins가 사용한 가스로 시계열 데이터[3]를 이용하여, 입출력 데이터인 가스 흐름율과 연소된 이산화탄소 농도의 가스로 공정을 퍼지 모델링한다.

표 1은 후반부 구조 1, 2로 구성된 각 모델에 대하여 데이터의 Min-Max를 이용해 퍼지 입력공간의 파라미터 값을 동정한 성능지수와 HCM을 이용하여 퍼지 입력 공간을 나누고, 그것을 초기 값으로 하이브리드 알고리즘에 의해 파라미터를 동조한 성능지수이다. 하중값을 가진 목적함수에서 θ 의 선택에 따른 모델의 성능지수는 표 2에 보였다.

표 1. 퍼지모델의 성능지수

| 모델 | 간략 | | 선형 | |
|-------------------|-------|-------|-------|-------|
| | PI | E_PI | PI | E_PI |
| Min-Max(rule6)[9] | 0.022 | 0.336 | 0.021 | 0.362 |
| HCM(rule6)[9] | 0.755 | 1.439 | 0.018 | 0.286 |
| HCM+Hybrid(rule6) | 0.022 | 0.332 | 0.017 | 0.275 |

그림 2는 HCM에 의한 퍼지 입력공간의 분할을 기존의 Min-Max에 의한 분할과 비교하여 보여주고 있으며, 그림 3은 HCM을 초기값으로 한 모델의 최적화된 파라미터를 보여준다.

표 2. 하중값에 따른 성능지수(구조 2)

| 입력1 | 입력2 | θ | Min-Max+GA[9] | | Hybrid(GA+Complex) | |
|--------|--------|----------|---------------|-------|--------------------|-------|
| | | | PI | E_PI | PI | E_PI |
| u(t-3) | y(t-1) | 0.0 | 0.031 | 0.278 | 0.0174 | 0.276 |
| | | 0.25 | 0.029 | 0.283 | 0.0173 | 0.276 |
| | | 0.5 | 0.018 | 0.264 | 0.0174 | 0.275 |
| | | 0.75 | 0.018 | 0.279 | 0.0172 | 0.276 |
| | | 1.0 | 0.018 | 0.350 | 0.0174 | 0.269 |

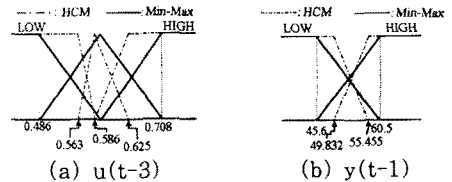


그림 2. 퍼지 입력공간의 분할

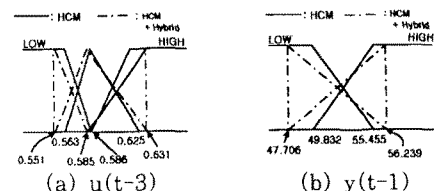


그림 3. 구조2의 동정한 파라미터 ($\theta = 1.0$)

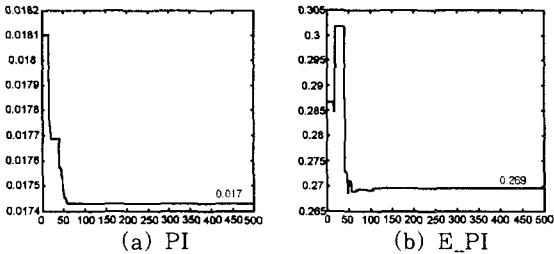


그림 4. 하이브리드 알고리즘에 의한 퍼지 모델의 최적화 과정

최적화된 모델로 구조 2의 $\theta=1.0$ 을 선택하였을 때 하이브리드 알고리즘에 의한 PI와 E_PI의 최적화 과정을 그림 4에 보였다. 표 3에서는 기존의 퍼지 모델과 동정 에러를 비교하였다.

표 3. 기존 퍼지모델과 동정 에러 비교

| 모델 | | PI | E_PI | 규칙 |
|---------------------------|----|--------|-------|----|
| Oh's model (5,8) | 간략 | 0.024 | 0.328 | 4 |
| | 선형 | 0.022 | 0.326 | 4 |
| | | 0.0211 | 0.364 | 6 |
| HCM+GA (9) | 간략 | 0.027 | 0.333 | 6 |
| | 선형 | 0.020 | 0.264 | 6 |
| Our model (HCM+Hybrid) | 간략 | 0.022 | 0.332 | 6 |
| | 선형 | 0.0174 | 0.269 | 6 |

3.2 하수처리공정

수도권 하수처리장 중의 하나를 모델로 선정하여 이 처리장의 1년치 수질 데이터를 토대로 앞 절에서 제안한 퍼지 추론에 의해 퍼지모델링을 하고 하이브리드 알고리즘으로 모델을 최적화한다. 모델의 입력은 MLSS, WSR, RRSP, DOSP이며 ESS를 출력으로 하는 구조이다. HCM을 이용하여 전반부 멤버쉽함수를 결정하고 이를 초기값으로 식(6)의 목적함수를 최소화 하는 전반부 파라미터를 하이브리드 알고리즘에 의해 구한다.

표 4. 모델의 성능지수

| 모델 | 입력1 | 입력2 | 구조1 | | 구조2 | |
|--------|------|------|--------|--------|---------|----------|
| | | | PI | E_PI | PI | E_PI |
| HCM+GA | MLSS | WSR | 12.324 | 11.106 | 0.0002 | 121.801 |
| | MLSS | RRSP | 12.259 | 13.149 | 0.354 | 237.728 |
| | MLSS | DOSP | 14.127 | 14.757 | 4.356 | 1360.946 |
| | WSR | RRSP | 12.102 | 9.902 | 0.00003 | 159.685 |
| | WSR | DOSP | 10.582 | 11.178 | 0.00003 | 222.313 |
| | RRSP | DOSP | 11.054 | 10.326 | 0.017 | 69.853 |

표 5. 하중값을 가진 목적함수에 의한 모델의 성능지수

| 입력1 | 입력2 | θ | 구조1 | |
|-----|------|----------|--------|--------|
| | | | PI | E_PI |
| WSR | DOSP | 1.0 | 10.587 | 11.176 |
| | | 0.25 | 10.585 | 11.177 |
| | | 0.5 | 10.582 | 11.178 |
| | | 0.75 | 10.576 | 11.192 |
| | | 0.0 | 9.068 | 16.711 |

표 6. 기존 퍼지모델과 동정 에러 비교 (입력변수 : MLSS, WSR)

| 모델 | | PI | E_PI | 규칙 |
|---------------------------|----|--------|-----------|----|
| Oh's model (5,8) | 간략 | 13.726 | 16.206 | 4 |
| | | 14.107 | 16.563 | 6 |
| | 선형 | 6.396 | 54.233 | 4 |
| | | 1.461 | 80618.742 | 6 |
| Min-Max+GA (9) | 간략 | 12.847 | 14.155 | 4 |
| | | 12.318 | 13.117 | 6 |
| | 선형 | 7.087 | 32.970 | 4 |
| | | 0.0002 | 187.911 | 6 |
| Our model (HCM+Hybrid) | 간략 | 12.324 | 11.106 | 6 |
| | 선형 | 0.0002 | 121.801 | 6 |

표 4는 후반부 구조 1, 2로 구성된 각 모델의 성능지수이다. 입력이 WSR, DOSP일 때 θ 의 선택에 따른 모델의 성능지수는 표 5에 보였다.

표 6은 기존의 Min-Max를 이용한 퍼지모델의 성능지수와 본 논문에서 제안된 모델의 성능지수를 비교하여 보여주고 있다. 여기서 Oh의 모델은 개선된 컴플렉스를 이용한 모델의 성능지수이다.

3. 결론

본 논문에서는 HCM과 하이브리드 알고리즘을 이용하여 보다 효과적으로 최적의 퍼지 모델을 동정하였다. HCM을 이용하여 임출력 데이터의 특징을 찾아 모델 설계에 반영하였고, 하이브리드 알고리즘을 이용하여 유전자 및 컴플렉스 알고리즘이 가진 각 단점과 한계성을 극복함으로써 퍼지규칙 전반부 멤버쉽함수의 파라미터 값을 자동 동조하여 최적 퍼지 모델을 구현하였다. 시뮬레이션에서 볼 수 있듯이 제안된 모델은 복잡하고 비선형이 강한 공정에서 기존의 퍼지 모델들 보다 성능이 향상된 퍼지 모델을 설계할 수 있었고, 인간의 경험적 방법이 아닌 보다 체계적인 방법에 의한 객관적 모델을 획득할 수 있었다. 또한 하중값을 가지는 목적함수에 의해 퍼지모델의 근사화와 일반화 사이에서 최적화에 대한 방향을 제시함으로써 모델의 근사화와 일반화 사이에 상호연계를 통한 최적화 향상을 위한 방향을 제시하였다.

감사의 글

이 논문은 1998년 한국학술진흥재단의 공모과제 연구비에 의하여 연구되었습니다.

[참고 문헌]

- W. Pedrycz, "Numerical and application aspects of fuzzy relational equations", *Fuzzy Sets Syst.*, Vol. 11, pp.1-18, 1983.
- R.M. Tong, "The evaluation of fuzzy models derived from experimental data", *Fuzzy Sets Syst.*, Vol. 13, pp.1-12, 1980.
- Box and Jenkins, "Time Series Analysis, Forecasting and Control", Holden Day, San Francisco, CA.
- M.A. Ismail, "Soft Clustering Algorithm and Validity of Solutions", *Fuzzy Computing Theory, Hardware and Applications*, edited by M.M. Gupta, North Holland, pp.445-471, 1988.
- 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.*, 1999 (To appear)
- David E. Goldberg, "Genetic Algorithms in search, Optimization & Machine Learning", Addison-wesley.
- Zbigniew Michalewicz, "Genetic Algorithms + Data Structure = Evolution Programs", Springer-Verlag.
- 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.
- 박병준, 오성권, 안태권, 김현기 "유전자 알고리즘과 하중값을 이용한 퍼지시스템의 최적화", 대한적기학회 논문지, 1999년 6월호 게재예정.
- 오성권, "퍼지모델 및 제어이론과 프로그램", 기다리출판사, 1999년 3월.