

퍼지 추론 방법을 이용한 퍼지 동정과 유전자 알고리즘에 의한 이의 최적화

박병준, 박춘성, 안태천, 오성권

원광대학교 제어계측공학과, 전라북도 익산시 신룡동 344-2 @570-749

Fuzzy Identification by means of Fuzzy Inference Method and its Optimization by GA

Byoungjun Park, Chunseong Park, Taechon Ahn, Sungkwun Oh

Dept. of Control & Instrumentation Engineering, Wonkwang Univ., KOREA

Abstract - In this paper, we are proposed optimization method of fuzzy model in order to complex and nonlinear system. In the fuzzy modeling, a premise identification is very important to describe the characteristics of a given unknown system. Then, the proposed fuzzy model implements system structure and parameter identification, using the fuzzy inference method and genetic algorithms. Inference method for fuzzy model presented in our paper include the simplified inference and linear inference. Time series data for gas furance and sewage treatment process are used to evaluate the performance of the proposed model. Also, the performance index with weighted value is proposed to achieve a balance between the results of performance for the training and testing data.

1. 서 론

퍼지시스템 이론의 발전으로 퍼지모델 동정 알고리즘의 접근 방식도 향상되었다. 초기 퍼지모델의 동정연구로 언어적 접근 방식[1,2]과 퍼지관계 방정식에 기초한 접근방식[3,4]이 제안되었다. 언어적 접근방식에서, Tong은 논리적 조사 방법에 의해 가스로 공정을 동정하였고[5], Pedrycz는 언어적 모델링에서 기준 퍼지집합 개념에 의해 해석하였다[2]. 퍼지관계 방정식에 기초한 접근방식에서, Pedrycz는 퍼지관계 방정식에 의해 새로운 합성규칙인 기준 퍼지집합과 Zadeh의 제한적 가능성 분포의 개념에 기초해서 동정하였고[3], Sugeno는 자신이 제안한 선형 퍼지추론식에 경험적 방법을 통한 구조동정으로 컴플렉스법과 최소자승법에 의해 시스템을 동정하였으나[8], 전반부 구조동정에서 반복적인 퍼지분할에 의한 경험에 의존하는 문제점이 있다. 또한 퍼지클러스터링 방법에 의해 가스로 공정을 동정하였으나[6], 직접 추론법에 의한 것으로 동정성능이 떨어졌다.

본 논문에서는 간략추론과 선형추론의 두가지 퍼지추론 방식에 유전자 알고리즘을 이용하여 퍼지모델을 동정한다. 잠재적으로 좋은 결과를 가져다주는 과정을 추적하고, 반면에 지역 극소에 빠질 가능성을 제거하기 위해 임의의 지점을 검사함으로써, 해 공간 전체를 탐색할 수 있는 능력을 가지고 있는(Goldberg,1989) 유전자 알고리즘을 사용하여 전반부 파라미터를 동정하며, 후반부 파라미터는 최소자승법을 사용하여 동정한다. 두가지 추론 방법에는 삼각형 형태의 소속함수를 사용한다. 제안된 모델의 성능 평가를 위해 가스로 시계열 입출력 데이터[7]와 하수처리공정 데이터[11]를 모델링하고, 학습 및 테스트 데이터의 성능결과에 대한 상호균형을 얻기 위해 하중값을 가진 성능지수가 제시된다.

2. 본 론

2.1 퍼지 추론 방법에 의한 모델링

퍼지 모델링이란 if-then 형식으로 플랫폼을 기술하는 것으로, 구체적으로 입출력 데이터를 주고 여러 변수를 설정하는 것이다. 전반부 동정은 입력 정보, 즉 퍼지 규칙의 전반부 변수의 선택과 입력변수 공간의 퍼지분할 결정, 그리고 입력공간의 파라미터 결정이 필요하며, 후반부 동정은 후반부 변수의 선택과 후반부 변수의 파라미터를 결정하는 것이다. 본 논문에서는 전반부 구조로 결정된 입력 수와 규칙 수를 가지고, 유전자 알고리즘을 이용하여 전반부 파라미터를 동정한다. 후반부 구조는 2가지 구조로 표현하며, 후반부 파라미터 동정은 최대 피벗팅(maximal pivoting) 알고리즘을 가지는 가우스 소거법에 의한 최소자승법을 사용한다.

2.1.1 전반부 동정

전반부 구조로는 입력이 두 개일 때, 그리고 각각의 입력에 대해 소속함수를 정의하였으며, 전반부 파라미터는 유전자 알고리즘을 이용하여 동정한다.

2.1.2 후반부 동정

후반부 구조로는 단일 상수항(구조1)만을 가지거나, 후반부가 일차 선형식 형태(구조2)를 가지는 두 가지 구조를 사용한다. 두 가지 구조에 의한 퍼지규칙추론과 파라미터 동정은 다음과 같다.

$$\begin{aligned}
\text{구조 1(후반부 : 상수)} \\
\text{후반부가 단일 상수항만을 가지는 것으로, 간략추론법이라 한다. 이 퍼지모델은 식(1)과 같은 형태를 가지는 구현규칙들로 구성된다.} \\
\begin{matrix}
R^1 \text{ IF } x_1 \text{ is } A_{11} \text{ and } \dots \text{ and } x_n \text{ is } A_{1n} \text{ Then } a_{10} \\
\vdots \\
R^j \text{ IF } x_1 \text{ is } A_{j1} \text{ and } \dots \text{ and } x_n \text{ is } A_{jn} \text{ Then } a_{j0} \\
\vdots \\
R^n \text{ IF } x_1 \text{ is } A_{n1} \text{ and } \dots \text{ and } x_n \text{ is } A_{nn} \text{ Then } a_{n0}
\end{matrix} \quad (1)
\end{aligned}$$

추론된 값 y' 는 식(2)와 같다.

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

$$\begin{aligned}
(i : i(i=1, \dots, m)\text{번째 데이터}) \\
w_{ji} = \frac{(A_{j1}(x_{i1}) \wedge \dots \wedge A_{jn}(x_{in}))}{\sum_{j=1}^n (A_{j1}(x_{i1}) \wedge \dots \wedge A_{jn}(x_{in}))} \quad (3)
\end{aligned}$$

후반부 파라미터 동정에서 전반부 입력변수 및 파라미터가 주어지면, 성능지수를 최소화하는 최적 후반부 파라미터를 결정할 수 있다. 시스템의 출력 데이터와 모델 데이터간의 차이를 의미하는 성능지수(PI: Performance Index)는 식(4)로 정의된다.

$$PI = \frac{1}{m} \sum_{i=1}^m (y_i - y'_i)^2 \quad (4)$$

(m : 데이터의 수)

최소자승법에 의한 후반부 파라미터의 동정은 다음 식

에 의해 구해진다.

$$\hat{a} = (X^T X)^{-1} X^T Y \quad (5)$$

구조 2(후반부 : 일차 선형식)

후반부가 일차 선형식으로 표현되며, 선형추론법 또는 혼합추론법이라 한다.

$$\begin{aligned} R^1 \text{ IF } x_1 \text{ is } A_{j1} \text{ and } \dots \text{ and } x_k \text{ is } A_{jk} \\ \text{Then } y = a_{j0} + a_{j1}x_1 + \dots + a_{jk}x_k \\ \vdots \\ R^j \text{ IF } x_1 \text{ is } A_{jn} \text{ and } \dots \text{ and } x_k \text{ is } A_{jk} \\ \text{Then } y = a_{j0} + a_{j1}x_1 + \dots + a_{jk}x_k \\ \vdots \\ R^n \text{ IF } x_1 \text{ is } A_{n1} \text{ and } \dots \text{ and } x_k \text{ is } A_{nk} \\ \text{Then } y = a_{n0} + a_{n1}x_1 + \dots + a_{nk}x_k \end{aligned} \quad (6)$$

추론된 값 y^* 는 다음과 같다.

$$y^* = \frac{\sum_{j=1}^n w_{ji} y_j}{\sum_{j=1}^n w_{ji}} = \frac{\sum_{j=1}^n w_{ji} (a_{j0} + a_{j1}x_1 + \dots + a_{jk}x_k)}{\sum_{j=1}^n w_{ji}} \quad (7)$$

최소자승법에 의한 후반부 파라미터의 동정은 구조 1과 같이 식(5)에 의해 구해진다

2.2 모델의 최적화

최적화 문제는 많은 영역에서 발생되고, 기술적인 환경에 지배된다. 몇 가지 해석적이고 수치적인 최적화 방법이 있지만, 이들 방법에 의해 풀리지 않는 함수들의 부류들이 아직 까지 있다. 이러한 함수들을 해결하기 위해서 유전자 알고리즘은 고안되었고, 함수 최적화기법으로써 현재 굉장한 주목을 받고 있다.

2.2.1 유전자 알고리즘

유전자 알고리즘은 생물학적 진화, 자연선택, 유전자 재조합의 생물학 원리에 바탕을 둔 확률적 탐색 방법이다. 유전자 알고리즘은 3개의 기본 연산자-생산, 교배, 돌연변이-를 가지고, 평가, 선택, 생산을 통해 반복 수행된다(그림1). 유전자 알고리즘은 최적화 문제에 있어서 향상된 성능을 가진 새로운 해들을 얻기 위해 실질적 정보를 효과적으로 이용하는 능력을 가지며, 복잡한 탐색 공간에서 강한 탐색능력을 발휘한다는 것이 이론상 그리고 경험적으로 증명되었다.

논문에서는 퍼지모델을 최적화하기 위해 앞에서 제시한 전반부 파라미터를 유전자 알고리즘을 이용하여 동정한다. 퍼지추론의 방법, 삼각형 소속함수의 특성, 그리고 하중값을 바탕으로 퍼지모델을 최적화한다.

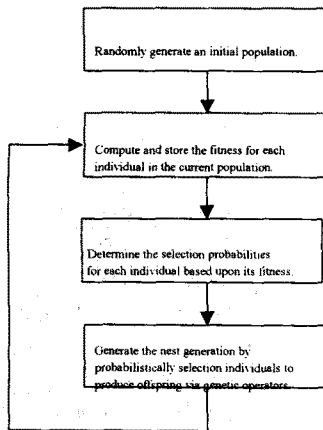


그림 1. 유전자 알고리즘

2.2.2 하중값을 가진 목적함수

학습과 테스트 데이터에 대해 하중값을 가진 목적함수를 다음과 같이 정의한다.

$$f = \theta \times PI + (1 - \theta) \times E_PI \quad (8)$$

θ 는 PI와 E_PI에 대한 하중값이고, PI는 학습에 대

한 성능지수를, E_PI는 테스트에 대한 성능지수 값을 나타낸다. 위에서 정의한 목적함수는 θ 값의 선택에 따라 퍼지모델의 근사화와 일반화 사이에서 최적화에 대한 방향을 설정할 수 있는 특징을 가진다.

3 시뮬레이션

3.1 가스로 공정

가스로 시계열 입출력 데이터를 학습과 테스트 데이터로 나누어 사용한다. 학습과 테스트 데이터의 성능지수를 유전자 알고리즘의 목적함수(식8)로 사용하여 모델을 최적화한다. 표1은 기존의 경험적 방법에 의해 퍼지 입력공간의 파라미터값을 동정한 성능지수와 유전자 알고리즘에 의한 성능지수이다. θ 의 선택에 따른 성능지수는 표2에 보였다.

표1. 모델의 성능지수

구조	모델	PI	E_PI
구조 1	기본모델	0.022855	0.335439
	최적화모델 ($\theta=0.5$)	0.023456	0.344973
구조 2	기본모델	0.024739	0.358243
	최적화모델 ($\theta=0.5$)	0.018809	0.264046

표2. 하중값을 가진 목적함수에 의한 성능지수

구조	Weight (θ)	PI	E_PI
구조 1	1.0	0.022855	0.335407
	0.25	0.056886	0.316511
	0.4	0.025553	0.328181
	0.6	0.023995	0.329839
	0.75	0.022856	0.335137
구조 2	0.0	0.117397	0.306032
	1.0	0.018360	0.350625
	0.25	0.029384	0.283629
	0.4	0.019560	0.278573
	0.6	0.018944	0.279414
	0.75	0.018859	0.279530
	0.0	0.031077	0.278938

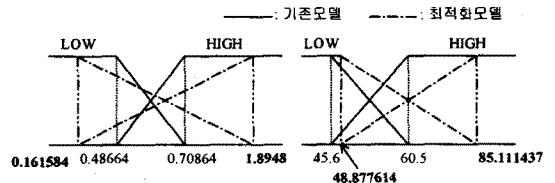


그림 2. 구조 1의 동정된 파라미터 ($\theta=0.0$)

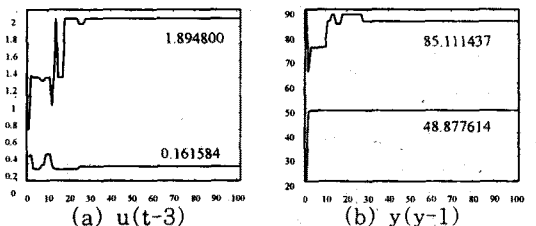


그림 3. 유전자 알고리즘에 의한 구조 1의 파라미터 탐색과정 ($\theta=0.0$)

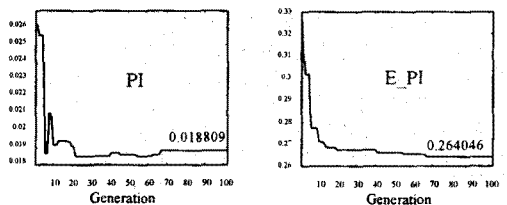


그림 4. 유전자 알고리즘에 의한 구조2 모델의 최적화 과정 ($\theta=0.5$)

그림2에는 구조1의 동정된 파라미터를 보였으며, 파라미터의 탐색과정은 그림3에 보였다. 또한 그림4에는 유전자 알고리즘에 의한 구조2의 최적화 과정을 보여준다.

3.2 하수처리 공정

하수처리공정을 앞 절에서 제안한 퍼지추론에 의해 모델링을 하고, 유전자 알고리즘으로 모델을 최적화한다. 그림5는 하수처리공정 모델링을 위한 입력력 구조를 보여주며, MLSS는 혼합액 부유물, WSR은 잉여오니 흐름율, RRSP는 반송을 설치치, DOSP는 용존산소 설치치, ESS는 부유물의 농도를 가지는 4입력 1출력 구조

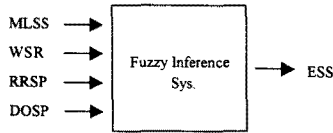


그림 5. 하수처리 공정 모델링을 위한 입력력 구조

이다. 여기서는 시뮬레이션을 위해 2입력 1출력 구조를 사용한다. 표3은 기존모델과 유전자 알고리즘에 의한 성능지수이고, 6 값에 따른 모델의 성능지수는 표4와 같다.

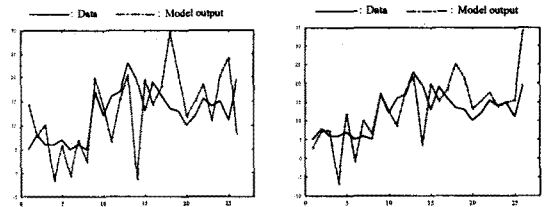
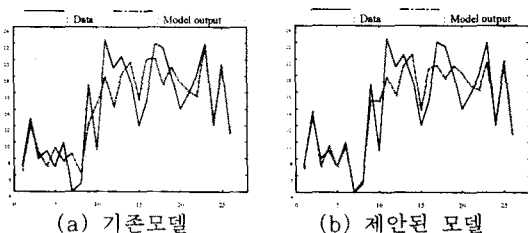
표3. 모델의 성능지수

모델	입력1	입력2	구조1		구조2	
			PI	E PI	PI	E PI
기존 모델	MLSS	WSR	15.875622	17.371275	6.396533	54.529442
	MLSS	RRSP	15.492023	16.407095	4.328988	222.968689
	MLSS	DOSP	15.177470	21.739044	8.332672	476.234711
	WSR	RRSP	19.117628	17.004622	7.798279	225.496414
	WSR	DOSP	17.725309	20.548513	8.622758	57.188206
	RRSP	DOSP	19.472698	22.089949	7.425736	99.412163
최적화 모델 ($\theta=0.5$)	MLSS	WSR	12.847437	14.155079	7.087610	32.970261
	MLSS	RRSP	12.873838	13.531213	13.319167	65.584373
	MLSS	DOSP	14.609385	17.124805	7.556438	26.155487
	WSR	RRSP	15.221871	13.473765	8.657568	26.561729
	WSR	DOSP	15.659515	14.344687	6.871958	28.599873
	RRSP	DOSP	17.854191	15.889150	4.135979	66.810814

표4. 하중값을 가진 목적함수에 의한 성능지수

구조	입력1	입력2	Weight (θ)	PI	E PI
구조 1	MLSS	RRSP	1.0	9.917910	117.707550
			0.25	12.913282	13.499081
			0.4	12.913285	13.499073
			0.5	12.873838	13.531213
			0.6	12.210629	14.308299
			0.75	11.730998	15.339379
			0.0	13.137547	13.478786
구조 2	WSR	DOSP	1.0	3.214219	822.611267
			0.25	6.108916	26.913485
			0.4	6.103705	27.942829
			0.6	6.269360	26.920250
			0.75	6.105598	27.116993
			0.0	6.436031	30.676256

그림6은 구조2의 WSR과 DOSP의 입력에 대해 (a)와 (b)는 학습출력을, (c)와 (d)는 테스트출력을 보여준다.



(c) 기존모델 (d) 제안된 모델
그림 6. WSR, DOSP 입력에 대한 출력 ($\theta=0.25$)

4. 결 론

체계적이고 효율적인 모델을 얻기 위하여 유전자 알고리즘을 사용하여 퍼지 추론 시스템의 전반부 파라미터를 동정함으로써 퍼지 입력 공간을 정의하였고, 최적화된 모델은 시뮬레이션에서 볼 수 있듯이 복잡하고 비선형이 강한 공정에서 보다 향상된 모델을 설계할 수 있었다. 또한, 하중값을 가진 목적함수에 하중값을 설정함으로써 모델의 근사화와 일반화 사이에 상호 연계를 통한 최적화 향상을 위한 방향을 제시하였다.

(참 고 문 헌)

- [1] R.M. Tong, "Synthesis of fuzzy models for industrial processes", *Int. J. Gen. Syst.*, Vol. 4, pp.143-162, 1978.
- [2] W. Pedrycz, "An identification algorithm in fuzzy relational system", *Fuzzy Sets Syst.*, Vol. 13, pp.153-167, 1984.
- [3] W. Pedrycz, "Numerical and application aspects of fuzzy relational equations", *Fuzzy Sets Syst.*, Vol. 11, pp.1-18, 1983.
- [4] E. Czogola and W. Pedrycz, "On identification in fuzzy systems and its applications in control problems", *Fuzzy Sets Syst.*, Vol. 6, pp.73-83, 1981.
- [5] R.M. Tong, "The evaluation of fuzzy models derived from experimental data", *Fuzzy Sets Syst.*, Vol.13, pp.1-12, 1980.
- [6] M. Sugeno and T. Yasukawa, "Linguistic modeling based on numerical data", *IFSA 91 Brussels, Computer, Management & System Science*, PP.264-267, 1991.
- [7] Box and Jenkins, "Time Series Analysis, Forecasting and Control", Holden Day, San Francisco, CA.
- [8] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control", *IEEE Trans. Syst. Cybern.*, Vol.SMC-15, No. 1, pp.116-132, 1985.
- [9] David E. Goldberg, "Genetic Algorithms in search, Optimization & Machine Learning", Addison-wesley.
- [10] Zbigniew Michalewicz, "Genetic Algorithms + Data Structure = Evolution Programs", Springer-Verlag.
- [11] 오성권, 김현기, 우광방, "활성오니공정의 퍼지 모델링과 제어시스템", 대한 전기학회 논문집 40권 10호 PP.1072-1083, 1990년 10월.
- [12] 오성권, 우광방, "퍼지추론 방법에 의한 퍼지동정과 하수처리공정시스템 응용", 대한전자공학회 논문집 제 31권 B편 제 6호, PP.43-52, 1994년 6월
- [13] 최재호, 오성권, 안태천, 황형수, "유전자 알고리즘을 사용한 퍼지-뉴럴네트워크 구조의 최적모델과 비선형공정시스템으로의 응용", 한국퍼지 및 지능 시스템학회 '96 추계학술대회 논문집 Vol.6, No.2, PP.302-305, 1996
- [14] S.K. Oh and W. Pedrycz, "Identification of Fuzzy Systems by means of an Auto-tuning Algorithm and Its Application to Nonlinear System", *Fuzzy Sets Syst.*, 1998(accepted)