

유전자 알고리즘과 하중값을 이용한 퍼지 시스템의 최적화

論 文

48A-6-17

Optimization of Fuzzy Systems by Means of GA and Weighting Factor

朴炳俊* · 吳聖權** · 安泰天*** · 金賢起§

(Byoung-Jun Park · Sung-Kwun Oh · Tae-Chon Ahn · Hyun-Ki Kim)

Abstract - In this paper, the optimization of fuzzy inference systems is proposed for fuzzy model of nonlinear systems. A fuzzy model needs to be identified and optimized by means of the definite and systematic methods, because a fuzzy model is primarily acquired by expert's experience. The proposed rule-based fuzzy model implements system structure and parameter identification using the HCM(Hard C-mean) clustering method, genetic algorithms and fuzzy inference method. Two types of inference methods of a fuzzy model are the simplified inference and linear inference. In this paper, nonlinear systems are expressed using the identification of structure such as input variables and the division of fuzzy input subspaces, and the identification of parameters of a fuzzy model. To identify premise parameters of fuzzy model, the genetic algorithms is used and the standard least square method with the gaussian elimination method is utilized for the identification of optimum consequence parameters of fuzzy model. Also, the performance index with weighting factor is proposed to achieve a balance between the performance results of fuzzy model produced for the training and testing data set, and it leads to enhance approximation and predictive performance of fuzzy system. Time series data for gas furnace and sewage treatment process are used to evaluate the performance of the proposed model.

Key Words : Fuzzy inference system(퍼지추론 시스템), Identification(동정), Genetic algorithms(유전자 알고리즘), Weighting factor(하중값), Clustering Method(클러스터링 방법)

1. 서 론

많은 연구자들은 자연 현상을 이해하고, 그것을 이용하기 위해 연구가 이루어지고 있다. 그 중 1965년에 Zadeh에 의해 창안된 "퍼지집합" 이론은 비선형적이고 복잡한 실 시스템의 특성을 해석하는데 적용함으로써 수학적 모델보다 좋은 결과를 가져왔다. 이들은 if-then 규칙, 퍼지그래픽, 퍼지보간법, 퍼지위상, 퍼지추론, 퍼지추론 시스템, 퍼지모델링 등의 계산법과 같은 다양한 응용영역을 형성하며 많은 분야로 발전되었다. 수리적인 모델 방법을 대신하여 Zadeh가 제안한 방법은 다음과 같은 특징을 가지고 있다. 첫 번째, 수치가 아니라 언어적으로 애매한 변수를 이용하여 시스템을 표현하였고, 두 번째는 조건부명제에 의해 애매한 변수간의 관계를 나타냈다는 것이다. 세 번째는 시스템을 퍼지 알고리즘방식을 이용하여 표현했다. 퍼지 알고리즘과 복잡한 시스템의 언어 모델링을 이용하여 1974년 Mamdani가 제어에 응용한 것이 퍼지제어의 시초이며, 퍼지제어를 발달시킨 것

은 퍼지 알고리즘과 언어 모델링이다. 퍼지 알고리즘이란 crisp알고리즘을 퍼지화한 것으로 모호한 지시를 포함한다. 또한 언어 모델링이란, 복잡한 시스템의 모델링을 위해 제안된 것으로 Zadeh의 부적합성 원리에 근거를 두고 있다. 비선형적이고 다변수인 시스템을 대상으로한 퍼지 제어기와 퍼지 모델링 기법의 유용성은 이미 알려진 사실이며 이들은 퍼지 추론 시스템에 기초하고 있다. 퍼지 제어기와 퍼지 모델의 성능은 퍼지 규칙의 구성 방법에 의존하며 보다 좋은 성능을 위해서는 퍼지 규칙의 동정이 필연적이다. 퍼지시스템 이론의 발전으로 퍼지모델 동정 알고리즘의 접근 방식도 향상되었다. 초기 퍼지 모델의 동정연구로는 언어적 접근 방식[1,2]과 퍼지 관계 방식식에 기초한 접근방식[3,4]이 제안되었다. 언어적 접근방식에서, Tong은 논리적 조사 방법에 의해 가스로 공정을 동정하였고[7], B.Li는 이 방법의 수정으로 더 좋은 결과를 얻는 방법[5]과 결정 테이블에 기초한 자기 학습 알고리즘을 제안하였다. 이 알고리즘은 필요한 컴퓨터 용량 및 계산시간 때문에 고계다변수 시스템의 적용에 문제점을 발생시켰다[8]. Pedrycz는 언어적 모델링에서 기준 퍼지집합 개념에 의해 해석하였다[2]. 퍼지관계 방식식에 기초한 접근방식에서, Pedrycz는 퍼지관계 방식식에 의해 새로운 합성규칙인 기준 퍼지집합과 Zadeh의 제한적 가능성 분포의 개념에 기초해서 동정하였고[3], Xu는 이 두 가지 방법에 기초한 기준 퍼지집합 이론과 자기학습 알고리즘에 의해 퍼지관계 모델을 구성하여 동정하였으나[5,8], 직접추론방식을 사용했기 때문에 혼합(선형)추론방식 보다 그

* 準 會 員 : 圓光大 電氣電子工學部 碩士課程

** 正 會 員 : 圓光大 電氣電子工學部 助教授 · 工博

*** 正 會 員 : 圓光大 電氣電子工學部 教授 · 工博

§ 正 會 員 : 水原大 電氣電子情報通信工學部 副教授 · 工博

接受日字 : 1998年 11月 16日

最終完了 : 1999年 4月 26日

성능이 떨어지는 단점이 있다. Sugeno는 자신이 제안한 선형 퍼지추론식에 경험적 방법을 통한 구조동정으로 컴플렉스법과 최소자승법에 의해 시스템을 동정하였으나[10], 전반부 구조동정에서 반복적인 퍼지분할에 의한 경험에 의존하는 문제점이 있다. 또한 퍼지 클러스터링 방법에 의해 가스로 공정을 동정하였으나[11], 직접 추론법에 의한 것으로 동정성능이 떨어졌다. 이들 모든 동정알고리즘은 주어진 공정의 학습 데이터에 기초함으로써 보다 좋은 성능 즉 근사화 능력에 초점이 맞추어졌다.

본 논문에서 제안된 규칙베이스 퍼지모델은 클러스터링 알고리즘과 유전자 알고리즘 및 퍼지규칙의 이론을 이용함으로써 시스템 구조와 파라미터 동정을 시행한다. 또한 기존의 학습 데이터뿐만 아니라 테스트 데이터를 고려한 퍼지모델 성능 즉 근사화 능력과 예측능력 모두를 고려하여 실 공정 적용의 유용성에 그 방향을 맞추었고, 특히 하중값을 가진 성능지수에 의해 설계자의 의도, 성능 결과의 상호균형, 시스템의 비선형 정도 등에 따른 최적 모델 동정을 시도하였다. 두 형태의 퍼지모델 방법은 간략추론 및 선형추론에 의해 시행되며, 멤버쉽함수의 형태로는 삼각형 형태의 멤버쉽함수를 사용한다. 규칙 전반부 멤버쉽함수의 초기 파라미터는 클러스터링 기법 중 HCM 방법을 사용하여 정의한다. 퍼지추론 시스템의 전반부 파라미터의 동정은 유전자 알고리즘을 통해 퍼지 입력공간을 정의함으로써 비선형 시스템을 보다 최적화 하였고, 후반부는 표준 최소자승법을 사용하여 동정한다. 유전자 알고리즘은 자연 도태와 유전적 성질에 기초한 탐색 알고리즘으로 복잡한 문제에서 뛰어난 강인성을 보이며, 기존의 방법들과는 달리 적용에 별다른 제한이 없다는 특성 때문에 다양한 분야에 응용되고 있다. 본 논문에서는 유전자 알고리즘을 사용하여 우리가 원하는 최적의 파라미터 값을 얻음으로써 제안된 모델을 최적화 한다. 이의 평가를 위해 Box와 Jenkins가 사용한 가스로 시계열 입출력 데이터[9]와 하수처리공정 데이터[17]를 모델링함으로써 기존의 경험적 방법에 의한(데이터의 최대, 최소값을 이용한) 퍼지모델링 방법과 성능지수를 비교한다.

2. 퍼지추론에 의한 모델링

퍼지모델링이란 if-then 형식으로 플랜트를 기술하는 것으로, 구체적으로 입출력 데이터의 상호관계에 의해 설정된 입출력 변수로부터 확립되는 것이다. 규칙베이스 퍼지모델의 동정은 전반부와 후반부의 동정으로 나누어진다. 전반부 동정은 입력 정보, 즉 퍼지규칙의 전반부 변수의 선택과 입력변수 공간의 퍼지분할 결정, 그리고 입력공간의 파라미터 결정이 필요하다. 후반부 동정은 후반부 변수의 선택과 후반부 변수의 파라미터를 결정하는 것이다. 전반부의 입력변수와 퍼지집합의 멤버쉽함수는 입력변수의 공간을 규칙 수만큼의 퍼지분할된 부분공간으로 나누는 것과 관련되고, 후반부의 파라미터는 각각 퍼지분할된 부분공간에서 입출력 관계를 묘사한다. 퍼지모델링에서 전반부 동정, 즉 규칙의 구조동정 및 파라미터의 동정은 비선형 시스템을 표현하는데 있어서 중요하다(후반부의 동정은 보통 선형 시스템 동정문제와 본질적으로 같다). 본 논문에서는 전반부 구조로 결정된 입력 변수와 규칙을 가지고, 유전자 알고리즘을 이용

하여 전반부 파라미터를 동정한다. 후반부 구조는 두형태의 구조로 표현하며, 후반부 파라미터 동정으로는 최대 피벗팅(maximal pivoting) 알고리즘을 가지는 가우스 소거법에 의한 표준 최소자승법을 사용한다. 퍼지모델링의 전체구성도는 그림 1과 같다.

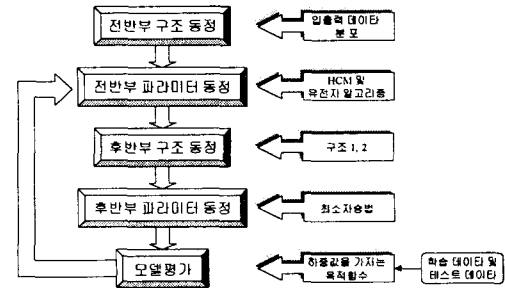
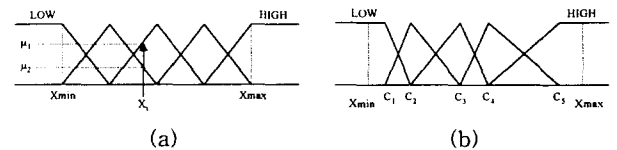


그림 1 퍼지모델링의 전체 구성도
Fig. 1 Overall scheme of fuzzy modeling.

3. 퍼지 동정 알고리즘

3.1 전반부 동정

기존의 방법은 멤버쉽함수를 그림 2의 (a)처럼 입력 변수의 최소값과 최대값 사이를 임의의 개수로 등분해서 일률적으로 정의하였다. 그러나 계통은 나름대로의 특성을 가지고 있기 때문에, 그 특성에 따라 멤버쉽함수를 정의하는 것이 보다 합리적일 수 있다.



(a) 기존 방법에 의한 소속함수의 균일한 정의
(b)클러스터링 알고리즘에 의한 소속함수의 정의
그림 2 멤버쉽함수의 정의

Fig. 2 The definition of the membership function

본 논문에서는 입출력 관계의 특성을 이용하여 모델링을 수행한다. 따라서 데이터의 특성을 파악하는 것이 매우 중요하다. 입출력 데이터의 특성을 찾아내기 위해 논문에서는 클러스터링 알고리즘을 이용한다. 클러스터링 기법을 통해 데이터를 특성별로 분류하고, 분류된 클러스터들의 중심을 사용하여 그림 2의 (b)처럼 멤버쉽함수를 정의한다.

클러스터링 알고리즘이란 데이터 내부의 비슷한 패턴, 속성, 형태 등의 기준을 통해 데이터를 분류하여 내부의 구조를 찾아내는 것이다. 본 논문에서는 데이터들간의 거리를 기준으로 하여 근접한 정도를 측정하고, 이를 바탕으로 데이터를 분류하는 HCM 방법을 사용한다. HCM 방법은 n개의 데이터를 c개의 그룹으로 분류하고 데이터의 거리가 최소인 각 그룹의 중심을 찾는다. 또한 클러스터의 소속을 "0", "1"로 나타내는 이치논리를 사용한다. 클러스터링 기법에서 가장 먼저 결정해야 될 것이 클러스터의 개수이다. 일반적으로 n개의 데이터로 구성된 집합 안의 클러스터 개수는 클러

스터가 2개 이상 n개 미만이어야 한다. 만약 클러스터가 1 개라면 모든 데이터가 하나의 클러스터에 포함되기 때문에 결국 주어진 데이터 자체가 되고, 클러스터의 개수가 n개라면 주어진 데이터 개수만큼 클러스터 개수가 정의되는 것이기 때문에 데이터 하나 하나가 하나의 클러스터를 형성하는 것과 같게되므로 클러스터링을 하는 의미가 없다. 클러스터의 개수는 데이터에 따라 적절한 개수를 선택해 주어야 한다. 이렇게 결정된 클러스터의 중심값을 모델의 전반부 멤버쉽함수의 정점으로 사용한다. HCM 알고리즘의 수행과정은 다음과 같다.

step 1. 클러스터의 개수 ($2 \leq c \leq n$)를 결정하고 중심 c_i ($i=1, \dots, c$)값을 랜덤하게 선택하여 초기화한다.

step 2. 소속행렬 U 를 결정한다.

$$u_{ij} = \begin{cases} 1 & \text{if } \|x_i - c_j\|^2 \leq \|x_i - c_k\|^2, \text{ for } k \neq j \\ 0 & \text{otherwise} \end{cases}$$

step 3. 아래 식을 계산한다. 중심 근처에 충분히 다가갔는지 확인한다

$$J = \sum_{i=1}^n J_i = \sum_{i=1}^n \left(\sum_{k, x_i \in G_k} d(x_i - c_i) \right) \quad (1)$$

step 4. 새로운 클러스터 중심 c_i 을 생성하고 step 2로 간다.

$$c_i = \frac{i}{|G_i|} \sum_{k, x_i \in G_i} x_i \quad (2)$$

이렇게 구해진 클러스터의 중심을 이용하여 전반부 멤버쉽함수를 정의한다. 전반부 구조는 입력이 두 개일 때 그리고 각각의 입력에 대해 멤버쉽함수를 정의하였으며, 전반부 파라미터를 유전자 알고리즘을 이용하여 동정한다. 삼각형 형태의 멤버쉽함수의 파라미터는 유전자 알고리즘에 의해 최적화 되며, 최적화에 따른 퍼지 모델의 성능을 평가한다.

3.2 후반부 동정

퍼지모델의 후반부 동정은 구조의 선택과 그들의 파라미터를 동정하는 문제이다. 후반부 구조로는 퍼지추론에 의해 구별되는 구조 1, 2를 사용한다. 구조 1은 후반부가 단일 상수항만을 가지고, 구조 2는 후반부가 일차선형식 형태를 가진다. 각각의 후반부 구조로부터 시스템 입출력 데이터를 가지고 가우스 소거법에 의한 최소자승법을 사용하여 파라미터를 동정한다. 두 가지 구조에 의한 퍼지규칙추론과 파라미터 동정을 위한 매개변수 추정은 다음과 같다.

구조 1(후반부 : 상수)

후반부가 단일 상수항만을 가지는 것으로, 이와 같은 추론법을 간략추론법이라 한다. 이 퍼지모델은 아래와 같은 형태를 가지는 구현규칙들로 구성된다.

$$\begin{aligned} R^1 & \text{ IF } x_1 \text{ is } A_{11} \text{ and } \dots \text{ and } x_k \text{ is } A_{1k} \text{ Then } a_{10} \\ & \vdots \\ R^j & \text{ IF } x_1 \text{ is } A_{j1} \text{ and } \dots \text{ and } x_k \text{ is } A_{jk} \text{ Then } a_{j0} \\ & \vdots \\ R^n & \text{ IF } x_1 \text{ is } A_{n1} \text{ and } \dots \text{ and } x_k \text{ is } A_{nk} \text{ Then } a_{n0} \end{aligned} \quad (3)$$

여기서 R^j 는 $j(j=1, \dots, n)$ 번째 규칙, $x_i(i=1, \dots, k)$ 는 입력변수,

$A_{jl}(j=1, \dots, n; l=1, \dots, k)$ 는 퍼지집합의 멤버쉽함수, a_{j0} 는 상수이고 n 은 규칙 수이다. 추론된 값 y^* 는 다음과 같다.

$$y^* = \frac{\sum_{j=1}^n w_{ji} y_j}{\sum_{j=1}^n w_{ji}} = \frac{\sum_{j=1}^n w_{ji} a_{j0}}{\sum_{j=1}^n w_{ji}} \quad (4)$$

$$w_{ji} = A_{j1}(x_{i1}) \wedge \dots \wedge A_{jk}(x_{ki}) \quad (5)$$

후반부 파라미터 동정에서 전반부 입력변수 및 파라미터가 주어지면, PI(Performance Index)를 최소화하는 최적 후반부 파라미터를 결정할 수 있다. PI는 원 시스템의 출력 데이터와 모델의 데이터간의 차이를 의미하는 성능지수로 식(6)으로 정의된다.

$$PI = \frac{1}{m} \sum_{i=1}^m (y_i - y_i^*)^2 \quad (6)$$

후반부의 파라미터는 a_{j0} 로써 입출력 데이터가 주어졌을 때 최소자승법에 의해 구해진다. 최소자승법에 의한 후반부 파라미터의 동정은 식(7)에 의해 구해진다.

$$\hat{a} = (X^T X)^{-1} X^T Y \quad (7)$$

$$\begin{aligned} x_i^T &= [w_{i1}, \dots, w_{in}], & a^T &= [a_{10}, \dots, a_{n0}] \\ Y &= [y_1, y_2, \dots, y_m]^T, & E &= [\epsilon_1, \dots, \epsilon_m]^T \\ X &= [x_1, x_2, \dots, x_m]^T \end{aligned} \quad (8)$$

구조 2(후반부 : 일차 선형식)

후반부가 일차 선형식으로 표현된 것으로 선형추론법 또는 혼합추론법이라 한다. 이퍼지 모델은 식(9)의 형태를 가지는 구현규칙들로 구성된다.

$$\begin{aligned} R^1 & \text{ IF } x_1 \text{ is } A_{11} \text{ and } \dots \text{ and } x_k \text{ is } A_{1k} \text{ Then} \\ & \qquad \qquad \qquad y = a_{10} + a_{11}x_1 + \dots + a_{1k}x_k \\ & \qquad \qquad \qquad \vdots \\ R^j & \text{ IF } x_1 \text{ is } A_{j1} \text{ and } \dots \text{ and } x_k \text{ is } A_{jk} \text{ Then} \\ & \qquad \qquad \qquad y = a_{j0} + a_{j1}x_1 + \dots + a_{jk}x_k \\ & \qquad \qquad \qquad \vdots \\ R^n & \text{ IF } x_1 \text{ is } A_{n1} \text{ and } \dots \text{ and } x_k \text{ is } A_{nk} \text{ Then} \\ & \qquad \qquad \qquad y = a_{n0} + a_{n1}x_1 + \dots + a_{nk}x_k \end{aligned} \quad (9)$$

여기서 R_j 는 j 번째 규칙, x_l 는 입력변수, A_{jl} 는 퍼지집합의 멤버쉽함수, $a_{jl}(j=1, \dots, n; l=0, \dots, k)$ 는 후반부의 파라미터이고, n 은 규칙 수이다. 추론된 값 y^* 는 다음과 같다.

$$y^* = \frac{\sum_{j=1}^n w_{ji} y_j}{\sum_{j=1}^n w_{ji}} = \frac{\sum_{j=1}^n w_{ji} (a_{j0} + a_{j1}x_{i1} + \dots + a_{jk}x_{ki})}{\sum_{j=1}^n w_{ji}} \quad (10)$$

여기서 i 는 $i(i=1, \dots, m)$ 번째 데이터이고 w_{ji} 는 식(5)와 같다. 최소자승법에 의한 후반부 파라미터의 동정은 구조 1과 같이 식(11)에 의해 구해진다.

$$\hat{a} = (X^T X)^{-1} X^T Y \quad (11)$$

$$\begin{aligned} x_i^T &= [w_{i1}, \dots, w_{ni}, x_{i1}w_{i1}, \dots, x_{i1}w_{ni}, \\ & \qquad \qquad \qquad \dots, x_{ki}w_{i1}, \dots, x_{ki}w_{ni}] \\ a^T &= [a_{10}, \dots, a_{n0}, a_{11}, \dots, a_{n1}, \dots, a_{1k}, \dots, a_{nk}] \\ Y &= [y_1, y_2, \dots, y_m]^T, & X &= [x_1, x_2, \dots, x_m]^T \end{aligned} \quad (12)$$

4. 최적화를 위한 유전자 알고리즘

4.1 최적화 문제

최적화 문제는 많은 영역에서 발생되고, 기술적인 환경에 지배된다. 몇 가지 해석적이고 수치적인 최적화 방법이 있지만, 그들 방법에 의해 풀리지 않는 함수들의 부류들이 아직 까지 많다. 예를 들어 현재 가장 많이 사용되고 있는 수학적 탐색방법은 미분 등의 보조 수단이 필요하며, 탐색 가능한 문제는 연속인 점이 유일하다. 또한 적용에 있어서도 미분이 가능한 경우로 제한되어 실제로 적용 가능한 경우는 매우 적다. 이러한 단점을 극복하기 위해서 임의의 탐색방법이 사용된다. 그러나 이 방법도 모든 가능한 영역을 탐색하는 방법으로 탐색영역이 넓으면 효율은 극히 나빠진다. 이 두 방법에서 야기되는 문제를 해결하기 위해서 유전자 알고리즘이 고안되었다. 기존의 방법에 의해 풀기 어려운 함수들에는 비컨벡스(non-convex), 다수극점(multi-modal), 그리고 노이즈(noisy) 함수들이 있으며[14], 이러한 함수들을 해결하기 위해서 유전자 알고리즘 고안되었고, 함수 최적화 기법으로써 유전자 알고리즘은 현재 굉장한 주목을 받고 있다. 유전자 알고리즘은 최적화 문제에 있어서 향상된 성능을 가진 새로운 해들을 얻기 위해 사실적 정보를 효과적으로 이용하는 능력을 가지고 있기 때문에 매우 유용하며, 또한 복잡한 탐색 공간에서 강한 탐색 능력을 제공한다. 이는 이론상 그리고 경험적으로 증명되었다. 뿐만 아니라, 그들은 경사 하강법과는 달리 지역 극소에 빠지지 않는다. 유전자 알고리즘은 기존의 방법들과 다음과 같은 차이가 있다. 첫째, 유전자 알고리즘은 최적화 문제의 파라미터 집합을 유한 길이의 문자열로 코딩한다. 그 결과로 파라미터 영역의 연속성이나 미분치의 존재와 같은 제약은 받지 않는다. 둘째, 많은 최적화 방법에서 사용되는 점에서 점으로의 이동 방법은 다수극점(multi-modal)을 갖는 탐색 영역에서는 국부적인 극점을 찾을 위험이 있으나, 유전자 알고리즘은 점들의 집단(population of string)으로부터 동작하여 다수의 최고점을 병렬로 올라가기 때문에 잘못된 최고점을 찾을 확률이 줄어든다. 셋째, 많은 탐색 기법이 효과적인 수행을 위해 보조적인 정보를 필요로 하는데 비해 유전자 알고리즘은 이러한 보조적인 정보를 필요치 않으며 단지 개별적인 문자열과 관련된 목적 함수값만을 필요로 한다. 넷째, 유전자 알고리즘은 탐색영역에 대한 보다 효율적인 탐색 방향을 유도하기 위해 확률적 천이 규칙을 사용한다. 이상의 4가지 차이점으로 인해 유전자 알고리즘은 보다 강인하게 된다.

4.2 유전자 알고리즘

Holland(1992)는 파라미터 공간의 동정에 있어서 높은 성능의 영역을 위해 지역 탐색 루틴에 의해 수행되는 유전자 알고리즘을 제안했다. 후에 지역 탐색 능력은 형식이 없는 돌연변이에 의해 개선되었다(Michalewicz 1992).

유전자 알고리즘은 생물학적 진화, 자연선택, 유전자 재조합의 생물학 원리에 바탕을 둔 확률적 탐색 방법으로, 잠재적으로 좋은 결과를 가져다주는 과정을 추적하고, 반면에 지역 극소에 빠질 가능성을 제거하기 위해 임의의 지점을 검

사함으로써, 해 공간 전체를 탐색할 수 있는 능력을 가진다 (Goldberg, 1989). 유전자 알고리즘의 특징은 주어진 값을 2진 코딩하여 코딩된 문자열을 하나의 개체로 그리고 개개의 비트를 유전자처럼 취급하며, 하나의 지점부터 최적지점으로 가 아닌 동시에 여러 지점에서 최적지점을 찾아 나간다. 그리고 미분과 같은 수학적 연산이 아닌 결과의 적합도를 기준(목적함수)으로 수행되며, 결정적인 방법이 아닌 확률적인 방법이다.

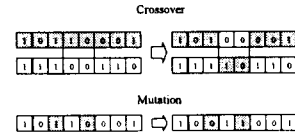


그림 3 유전자 알고리즘 연산자

Fig. 3 Genetic operators

유전자 알고리즘은 3개의 기본 연산자-생산, 교배, 돌연변이(그림 3)-를 가지고 있다. 생산(reproduction)은 다음 세대를 위해 교배 유전자가 선택되어지는 과정으로, 적합도에 따라 다음세대를 생산할 수 있는 기회를 차등적으로 부여한다. 각각의 문자열들은 그들의 적합도를 가지고 다음세대에 전달된다. 교배(crossover)는 두 단계의 과정을 가지는데 첫 번째 단계는 교배 유전자로부터 염색체들은 짝짓는다. 두 번째 단계로 짝지어진 문자열 쌍은 문자열의 k위치가 [1, l-1] 사이에서 확률적으로 균일하게 선택되어진다. 여기서 l은 문자열의 길이이다. 선택된 k와 l사이의 모든 특성들이 교차함으로써 두 개의 새로운 문자열이 생성되는 과정으로 교배가 이루어진다. 돌연변이(mutation)는 문자열 위치의 값을 임의로 변경한다. 2진 코딩에서 돌연변이는 0을 1로 또는 1을 0으로 바꾸는 것을 의미한다. 적절하게 정의된 적합도 함수와 함께 조합된 연산자는 유전자 알고리즘의 주요 부분이 된다.

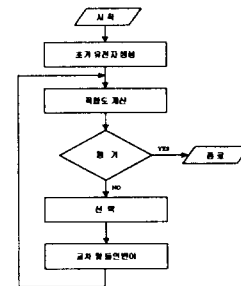


그림 4 유전자 알고리즘의 수행과정

Fig. 4 The operation procedure of the basic genetic algorithms

유전자 알고리즘은 확률적이거나 알려진 사양의 집합으로부터 만들어진 문자열의 군집을 가지고 3단계-평가(evaluation), 선택(selection), 재생산(reproduction)-를 통해 반복 수행된다. 그림 4는 3단계에 의한 유전자 알고리즘의 수행과정을 보여준다.

본 논문에서는 퍼지추론의 방법, 삼각형 멤버십함수의 특성, 그리고 하중값을 바탕으로 퍼지모델을 최적화한다. 동정 대상은 앞 절에서 제시한 것처럼 모델의 정도를 높이기 위해 유전자 알고리즘을 이용하여 퍼지추론의 전반부 파라미

터를 동정한다.

변수1	변수2	변수n
1101010101	0110101000		1110101011
1000011010	0110111011		0101011101
⋮	⋮		⋮
0110101001	1110101111		1101010111

그림 5 다변수 동정을 위한 직렬방식
Fig. 5 Serial method to identify multi-variables.

제안된 모델의 최적화를 위하여 사용된 유전자 알고리즘은 2진(binary)형의 직렬방식(그림 5)을 사용하며, 선택(selection)연산으로는 roulette-wheel방식, 교배(crossover)연산은 one-point crossover 연산을 사용한다. 그리고 돌연변이 연산은 선택된 비트를 반전시키는 invert방식을 사용한다. 세대는 100세대, 군집은 60개체까지 진행시켰으며, 각 문자열의 길이는 10 bit를, 교배율은 0.6, 돌연변이 발생율은 0.35를 사용한다. 그림 6은 모델의 최적화를 위해 사용된 유전자 알고리즘의 데이터 구조를 보여준다.

염색체	유전자 또는 개체	군 집
0' 또는 '1'	1001110111	1001110111 0101100010 0111010111

(a) 유전자 알고리즘의 데이터 구조

C ₁₁	C ₁₂	C ₁₃	C ₂₁	C ₂₂
10 bit	10 bit	10 bit	10 bit	10 bit

(b) 5변수의 동정을 위한 유전자 구조

그림 6 모델 최적화를 위한 유전자 알고리즘의 데이터 구조

Fig. 6 Data structure of genetic algorithms for optimization of fuzzy model.

4.3 하중값을 가진 목적함수

입출력 데이터로 시스템을 모델링할 때 가장 문제가 되는 것은 구축된 모델이 학습용 데이터에만 적합한 모델이 되고 시스템을 올바르게 표현하지 않는 경우이다. 즉 학습 데이터만을 사용하여 대상 공정의 모델을 구축하여 주어진 데이터에 의해서만 성능향상을 피하였기 때문에 실 공정에 적용하면 많은 오차를 발생시키고 정확한 예측능력을 가지지 못하였다. 이러한 현상은 비선형 공정인 경우 더욱 문제가 된다. 그것을 해결하기 위하여 여러 방법이 제안되고 있는데 가장 간단한 방법은 테스트 데이터를 만들어 그 데이터에 모델을 적용시켜보는 것이다. 그러나 이것 또한 테스트 데이터에 대한 평가를 반영하지 못하기 때문에 기존의 방법과 다를 바가 없다. 따라서 본 연구에서는 주어진 데이터를 이용하여 퍼지 모델링하고 퍼지 규칙을 추출한 다음, 그 생성된 모델에 평가용으로 사용될 테스트 데이터를 사용하여 모델을 평가하고 최적모델을 추출한다. 즉 성능 지수에 하중값을 가지는 목적함수를 제시한다.

학습 데이터와 테스트 데이터에 대해 하중값을 가진 목적함수를 다음과 같이 정의한다.

$$f(PI, E_PI) = \theta \times PI + (1 - \theta) \times E_PI \quad (13)$$

θ 는 PI 와 E_PI에 대한 하중값이다. PI는 학습데이터에

대한 성능지수를, E_PI는 테스트 데이터에 대한 성능지수이다. 이 목적함수가 최소값이 되도록 하기 위해 전반부 멤버쉽함수, 즉 삼각형 멤버쉽함수의 모든 파라미터들은 유전자 알고리즘에 의해 변화한다. 위에서 정의한 목적함수는 하중값의 설정에 따라 다음과 같은 특징을 가진다.

첫째, 만약 $\theta=1$ 이면 목적함수는 $f(PI, E_PI) = PI$ 와 같고, 그 모델은 학습을 바탕으로 최적화된다. 테스트는 고려하지 않는다. 모델의 근사화 능력은 뛰어나나, 일반화 능력(예측 능력)은 떨어진다.

둘째, 반대로, $\theta=0$ 일 경우에 목적함수는 $f(PI, E_PI) = E_PI$ 이며, 퍼지모델은 테스트를 바탕으로 전반부 파라미터는 최적화된다. 이것은 근사화 능력이 첫 번째보다 월등히 떨어지나 일반화 능력은 더욱 향상된다.

셋째, $\theta=0.5$ 는 학습과 테스트 모두가 같은 비중을 가지고 평가되는 경우이다. 이 퍼지모델은 첫 번째의 특징보다 퍼지모델 자체의 근사화 능력은 떨어질지 모르나, 일반화 능력은 향상된다. 또한 두 번째 특징보다 근사화 능력은 높으나 일반화 능력은 떨어진다.

넷째, $\alpha \in [0, 1]$ 에 대해서 $\theta = \alpha$ 이면 학습과 테스트 모두를 포함하고, α 의 선택은 퍼지모델의 근사화와 일반화 사이에서 최적화에 대한 방향을 설정한다. 이것의 의미는 학습 데이터에 의한 퍼지모델의 성능지수 PI를 구하고, 이 학습 데이터에 의해 생성된 퍼지모델에 테스트 데이터를 가하여 평가용 성능지수 E_PI를 구한다. 구해진 PI와 E_PI에 하중값 $\theta, 1-\theta$ 를 곱한 다음 함에 의한 목적함수 $f(\cdot)$ 를 최적화시키는 방향으로 모델의 전, 후반부 파라미터를 자동 동정함으로써 최적 퍼지모델을 구축한다. 이것은 모델 자체의 근사화 능력과 일반화 능력 양쪽을 다 고려한 목적함수의 형태로 적절한 하중값 선정에 따라 최적퍼지모델을 구축할 수 있다.

본 논문에서 사용되는 성능지수로 식(14)을 사용한다.

$$PI = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (14)$$

5. 시뮬레이션 및 결과고찰

5.1 가스로 공정

제안된 규칙 베이스 퍼지모델링을 Box와 Jenkins가 사용한 가스로 시계열 데이터[9]를 이용하여, 입출력 데이터인 가스 흐름율과 연소된 이산화탄소 농도의 가스로 공정을 퍼지모델링한다. 메탄 가스 흐름율($X(t)$)은 실험에서 -2.5에서 2.5로 변하는 것을 사용하나 실제 공정에서 메탄 가스 흐름율($X_m(t)$)은 식(15)에 의해 0.5에서 0.7로 변한다[9].

$$X_m(t) = 0.060 - 0.048 \times X(t) \quad (15)$$

입력이 가스 흐름율이고 출력이 이산화탄소 농도인 1입력 1출력의 가스로 시계열 입출력 데이터 296쌍을 시뮬레이션을 위해 입력으로 $u(t-3)$ 과 $y(t-1)$ 을, 출력으로 $y(t)$ 를 사용하는 2입력 1출력으로 설정하여 사용한다. 또한 데이터 집합은 학습과 테스트 데이터로 나누어 퍼지추론에 의한 모델링을 하고, 유전자 알고리즘으로 모델을 최적화한다.

그림 7는 입력 $u(t-3)$, $y(t-1)$ 과 출력 $y(t)$ 에 대한 데이터

의 분포를 보여준다. 데이터 집합 $(u(t-3), y(t-1), y(t))$ 이외의 $(u(t-1), y(t-1), y(t)), \dots, (u(t-5), y(t-1), y(t))$ 및 $(u(t-3), y(t-2), y(t)), (u(t-3), y(t-3), y(t))$ 등에 의한 데이터 분포는 덜 균일하고 흩어진 분포구조로 제안된 입출력 구조인 $(u(t-3), y(t-1), y(t))$ 보다 성능이 현저히 떨어진다.

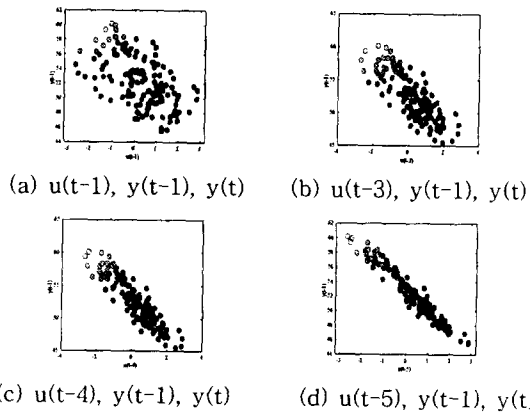


그림 7 학습데이터 집합에 의해 유겨된 데이터 점들
Fig. 7 Training data points induced by I/O data set

전반부 멤버쉽함수의 파라미터를 결정하기 위해서 식(14)의 PI를 최소화시키는 전반부 파라미터를 클러스터링 알고리즘과 유전자 알고리즘에 의해 구한다. 제안된 모델은 그림 8의 (a), (b)가 입력 퍼지변수의 멤버쉽함수로 작용하며 6개의 규칙을 가진다. 그림 8의 (a)는 입력변수 $u(t-3)$ 이 셋으로 퍼지분할된 경우이고 (b)는 입력변수 $y(t-1)$ 이 둘로 퍼지분할된 경우로, 클러스터링 알고리즘에 의해 입력공간을 퍼지분할 하였으며, 이것을 초기값으로 유전자 알고리즘에 의해 동정될 멤버쉽함수의 파라미터를 가지는 전반부 구조를 나타낸다.

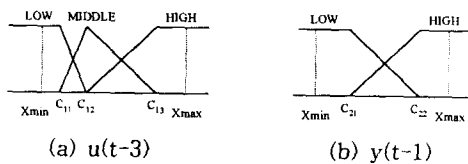


그림 8 입력 변수의 멤버쉽함수
Fig. 8 Membership function of input variables

후반부 구조는 앞 절에서 제시한 구조 1, 2를 사용하며, 후반부 파라미터 동정으로 최대 피벗팅(maximal pivoting) 알고리즘을 가지는 가우스 소거법에 의한 최소자승법을 사용한다.

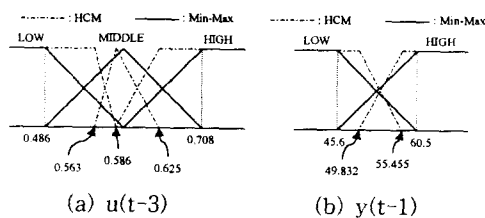


그림 9 퍼지 입력공간의 분할
Fig. 9 Fuzzy subspace using HCM

그림 9는 클러스터링 알고리즘(HCM)에 의한 퍼지 입력

공간의 분할을 기존의 Min-Max에 의한 분할과 비교하여 보여주고 있다. 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 유전자 알고리즘에 의해 파라미터를 동정한 성능지수이다. 하중값을 가진 목적함수에서 θ 의 선택에 따른 각 모델의 성능지수는 표 2에 보였다.

표 1 퍼지모델의 성능지수

Table 1 The performance index of fuzzy model

모델	구조 1		구조 2	
	PI	E-PI	PI	E-PI
Min-Max(rule4)[20]	0.022	0.335	0.024	0.358
Min-Max(rule6)[20]	0.022	0.336	0.020	0.362
HCM(rule6)	0.755	1.439	0.018	0.286
HCM+GA(rule6)	0.022	0.333	0.020	0.264

표 2 하중값을 가진 목적함수에 의한 성능지수

Table 2 The performance index with weighting factor

구조	Weight(θ)	PI	E-PI
구조 1	0.0	0.1769	0.3050
	0.25	0.0227	0.3332
	0.4	0.0227	0.3333
	0.5	0.0227	0.3333
	0.6	0.0227	0.3333
	0.75	0.0227	0.3333
	1.0	0.0225	0.3383
구조 2	0.0	0.019	0.272
	0.25	0.018	0.267
	0.4	0.023	0.271
	0.5	0.020	0.264
	0.6	0.019	0.272
	0.75	0.018	0.268
	1.0	0.016	0.271

구조 1은 후반부 구조가 간략추론으로 입출력 변수간의 상호관계가 상수항으로 표현되기 때문에 학습데이터 및 테스트 데이터에 의한 최적화에 있어서 뚜렷한 성능향상을 볼 수 없지만 구조 2는 선형추론으로 입출력 관계가 1차 선형식이기 때문에 최적화에 있어서 뚜렷한 성능향상을 볼 수 있다. 또한 표 1과 표 2에서 볼 수 있듯이 하중값 선택에 의해 설계자의 의도, 각 성능결과의 상호균형 및 시스템의 비선형 정도에 따라 최적화를 이룰 수 있고 향상된 모델을 구축할 수 있다.

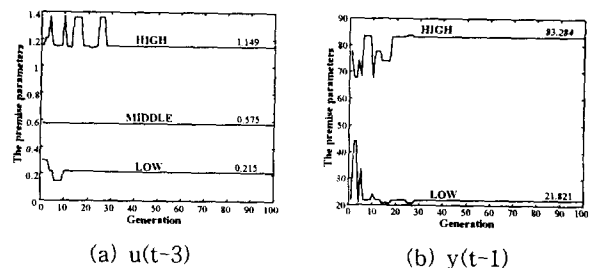


그림 10 유전자 알고리즘에 의한 구조1의 파라미터 탐색 과정($\theta=0.75$)

Fig. 10 Search process of parameters of structure 1 by GA.

구조 1에서는 유전자 알고리즘에 의한 최적 모델로 $\theta = 0.75$ 를 선택하였으며, 이 때 학습 성능지수는 0.0227이고 테스트 성능지수는 0.3333이다. 그림 10, 11은 선택된 구조 1 모델의 유전자 알고리즘에 의한 각 입력변수의 최적 파라미터 탐색과정과 동정된 파라미터를 보여준다. 그림 12는 파라미터 탐색과정에 따른 모델의 최적화 과정이다.

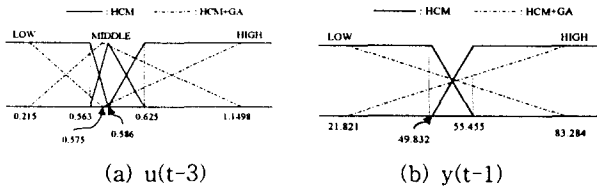


그림 11 구조1의 동정된 파라미터($\theta = 0.75$)
Fig. 11 Identified parameters of structure 1.

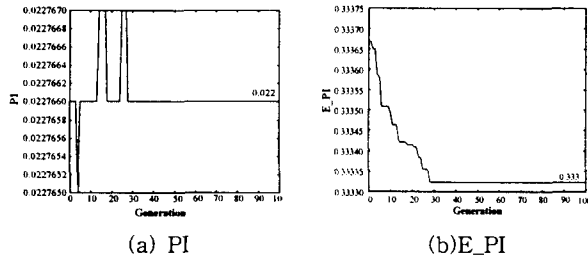


그림 12 유전자 알고리즘에 의한 구조1 모델의 최적화 탐색과정($\theta = 0.75$)
Fig. 12 Optimal search process of structure 1 by GA.

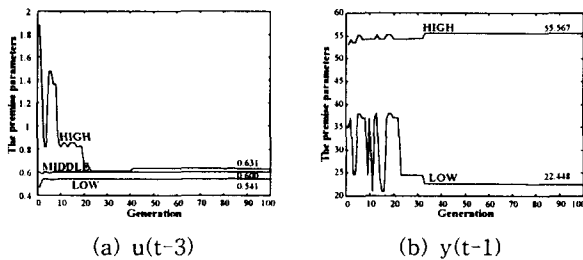


그림 13 유전자 알고리즘에 의한 구조2의 파라미터 탐색 과정($\theta = 1.0$)
Fig. 13 Search process of parameters of structure 2 by GA.

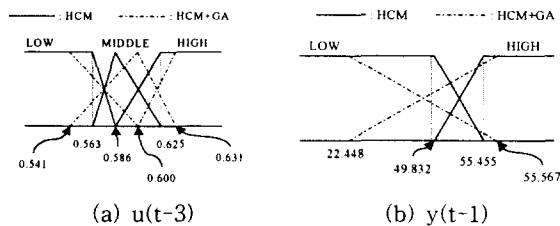


그림 14 구조2의 동정된 파라미터($\theta = 1.0$)
Fig. 14 Identified parameters of structure 2.

구조 2에서는 유전자 알고리즘에 의한 최적 모델로 $\theta = 1.0$ 을 선택하였으며, 이 때 학습 성능지수는 0.016이고 테스트

성능지수는 0.271이다. 그림13은 각 입력 변수에 대한 파라미터 탐색과정이며, 그에 따른 유전자 알고리즘에 의한 동정된 파라미터는 그림 14에서 보여준다. 그림 15는 파라미터 탐색과정에 따른 모델의 최적화 과정이다. 그림 16은 학습과 테스트 데이터에 대해 선택된 구조 2 모델의 출력과 비교하여 보여준다.

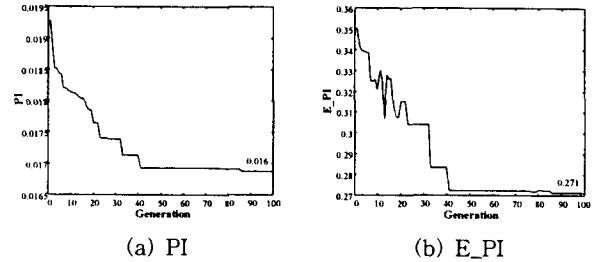


그림 15 유전자 알고리즘에 의한 구조2 모델의 최적화 탐색과정($\theta = 1.0$)
Fig. 15 Optimal search process of structure 2 by GA.

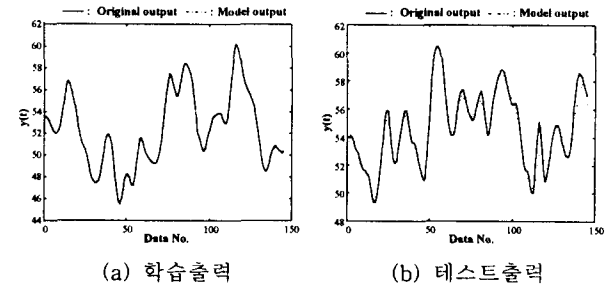


그림 16 구조2의 출력($\theta = 1.0$)
Fig. 16 Output of structure 2.

표 3 기존 퍼지모델과 동정 에러 비교
Table 3 Comparison of identification error with previous fuzzy models

모델		PI	E_PI	Total No. of rules
Neural Network		0.034	4.997	
Oh's model [15, 16]	Simplified Triangular	0.024	0.328	4
	Linear Triangular	0.022	0.326	4
Min-Max +GA [20]	Simplified Triangular	0.021	0.364	6
	Linear Triangular	0.023	0.344	4
Min-Max [20]	Simplified Triangular	0.018	0.264	4
	Linear Triangular	0.022	0.335	4
HCM	Simplified Triangular	0.022	0.336	6
	Linear Triangular	0.024	0.358	4
Our model (HCM+GA)	Simplified Triangular	0.020	0.362	6
	Linear Triangular	0.020	0.264	6
HCM	Simplified Triangular	0.755	1.439	6
	Linear Triangular	0.018	0.286	6
Our model (HCM+GA)	Simplified Triangular	0.035	0.289	4
	Linear Triangular	0.022	0.333	6
Our model (HCM+GA)	Simplified Triangular	0.026	0.272	4
	Linear Triangular	0.020	0.264	6

표 3은 Min-Max[15, 16, 20]를 기준으로 초기값을 설정한 기존의 방법들과 성능지수를 비교하여 보여주고 있다. 여기서 Oh의 모델[15, 16]은 최적화 알고리즘인 개선된 컴플렉스만을 이용한 것이다. Min-Max를 초기값으로 한 유전

자 알고리즘의 동정은 규칙이 증가하였을 때, 즉 규칙이 6개 일 때 데이터의 특징을 나타내지 못함으로써 시뮬레이션에 있어서 에러를 발생시킨다. 그러나 제안된 모델은 데이터의 성질을 반영함으로써 좋은 결과를 얻을 수 있다.

5.2. 하수처리 공정

본 논문에서는 수도권 하수처리장 중의 하나를 모델로 선정하여 이 처리장의 1년치 수질 데이터를 토대로 활성오니 공정을 앞 절에서 제안한 퍼지 추론에 의해 퍼지모델링을 하고 HCM을 기반으로 한 유전자 알고리즘으로 모델을 최적화한다. 하수처리시스템의 개요도는 그림 17과 같다.

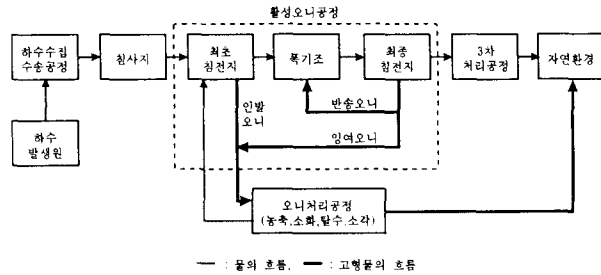


그림 17 하수처리시스템의 개요도
Fig. 17 Configuration of sewage treatment system.

하수처리에는 활성오니공정이 일반적으로 사용되고 있다. 하수처리 공정에서 침사지와 최초 침전지를 통과한 처리수에 잔존 용해 물질과 고형물은 활성오니처리(Activated Sludge Process : ASP)를 함으로써 유출수 수질의 질적 변화를 향상시킬 수 있다. 활성오니공정은 침사지, 최초 침전지, 폭기조 및 최종 침전지를 포함한다. 침사지와 최초 침전지에서는 하수중에 포함된 현탁물질의 입자가 중력에 의해 침강된다. 폭기조에서는 수 시간 연속적으로 하수중에 공기가 흡입된다. 유기물 제거의 중심을 이루는 미생물은 폭기조의 과정에서 자연 발생하는 다 종류의 미생물 군이고 이를 활성오니라 한다. 응집한 미생물의 덩어리를 플록(floc)이라 한다. 활성오니는 폭기조내 공기로부터 용해한 산소에 의해 유기물을 생물학적으로 산화해서 증식하고 대사과정에서 유기물을 물과 탄산가스로 분해한다. 최종 침전지에서 플록(floc)은 침강 제거되고 맑은 물이 3차 처리 공정으로 유입된다. 한편 최종 침전지에서 침강한 미생물은 폭기조 입구에서 재순환 되어 다시 유기물 제거에 사용된다.

활성오니공정은 폭기조 및 최종 침전지를 하나의 단일 모델로 선정하였다. 실제 최초 침전지로 유입되는 생물학적 산소요구량(Biological Oxygen Demand : BOD) 및 유입수의 부유물(Suspended Solid : SS) 농도를 측정하고, 최종 침전지에서 유출되는 생물학적 산소요구량(Effluent Biological Oxygen Demand : EBOD) 및 부유물 농도(Effluent Suspended Solid : ESS)가 주요 측정 대상이 된다. 활성오니공정의 용존산소 설정치(Dissolved Oxygen Set-Point : DOSP) 및 반송율 설정치(Recycle Sludge Ratio Set-Point : RRSF)는 유입수의 생물학적 산소요구량 및 부유물 농도에 따라 유출수의 생물학적 산소요구량(EBOD), 부유물(ESS) 농도가 영향을 받으므로 조작자가 이 유출수의 농도를 소정치 이하

로 낮추기 위해 사용되는 데이터이다. 유출수의 생물학적 산소요구량(EBOD) 및 부유물(ESS) 농도는 측정된 데이터의 혼합액 부유물(Mixed Liquid Suspended Solid : MLSS) 농도, 잉여오니 흐름율(Waste Sludge Ratio : WSR), 반송율 설정치(RRSP), 용존산소 설정치(DOSP)에 의존하며 생물학적 산소요구량(BOD), 부유물(SS)농도는 서로 일정한 상관관계를 가지고 있다.

본 논문에서는 활성오니공정을 폭기조 및 최종 침전지를 하나의 단일 모델로 설정하여, 실제 최초 침전지로 유입되는 생물학적 산소요구량 및 유입수의 부유물 농도를 측정하고, 최종침전지에서 유출되는 생물학적 산소요구량 및 부유물 농도를 통하여 하수처리 정도를 비교한다. 모델은 MLSS, WSR, RRSP, DOSP를 입력으로 하고, ESS를 출력으로 하는 4입력 1출력 구조를 가진다. 데이터의 개수는 52개로 전반부 26개는 학습 데이터로, 후반부 26개는 테스트 데이터로 이용한다. 클러스터링 알고리즘(HCM)을 이용하여 전반부 멤버쉽함수를 결정하고 결정된 멤버쉽함수의 파라미터를 초기값으로 식(14)의 PI를 최소화시키는 파라미터를 유전자 알고리즘에 의해 구한다. 제안된 모델은 그림 18의 (a), (b)가 입력 퍼지변수의 멤버쉽함수로 작용하며 6개의 규칙을 가진다. 그림 18의 (a)는 입력변수 MLSS가 HCM에 의해 셋으로 퍼지분할 된 경우이고 (b)는 입력변수 WSR이 퍼지분할된 경우로 Min-Max에 의해 균일하게 퍼지 분할된 것과 HCM에 의해 분할된 멤버쉽함수를 비교하여 보여주며, 동정할 멤버쉽함수의 파라미터를 가지는 전반부 구조를 나타낸다.

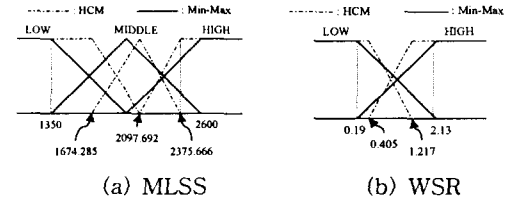


그림 18 입력 변수의 소속함수
Fig. 18 Membership function of input variables

표 4 모델의 성능지수

Table 4 The performance index of fuzzy model

모델	입력1	입력2	구조1		구조2	
			PI	E.PI	PI	E.PI
Min-Max [20]	MLSS	WSR	12.487	32.647	0.972	2480124.25
	MLSS	RRSP	12.081	17.825	0.863	526124.187
	MLSS	DOSP	12.435	20.614	2.155	131549.625
	WSR	RRSP	12.884	28.922	0.041	1978133.25
	WSR	DOSP	13.048	18.484	1.024	1409.257
	RRSP	DOSP	13.530	14.487	1.922	50.801
HCM-GA ($\theta=0.5$)	MLSS	WSR	12.352	11.179	0.001	126.912
	MLSS	RRSP	12.259	13.174	0.507	246.725
	MLSS	DOSP	14.453	14.801	52.682	276.072
	WSR	RRSP	12.530	10.983	5.212	210.923
	WSR	DOSP	10.600	11.231	0.003	284.192
	RRSP	DOSP	11.140	10.480	2.774	98.579

후반부는 구조 1, 2를 사용하며, 후반부 파라미터 동정으로 표준 가우스 소거법에 의한 최소자승법을 사용한다. 표 4는 후반부 구조 1, 2로 구성된 각 모델에 대하여 경험에 의해(데이터의 최대, 최소값을 이용해) 퍼지 입력공간의 파라

미터 값을 동정한 성능지수와 유전자 알고리즘에 의해 얻어진 성능지수이다. θ 의 선택에 따른 각 모델의 성능지수는 표 5에 보였다.

표 5 하중값을 가진 목적함수에 의한 모델의 성능지수
Table 5 The performance index with weighting factor

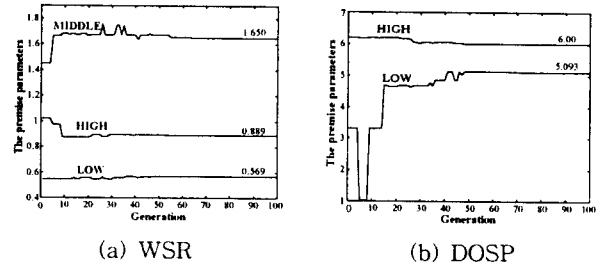
입력1	입력2	하중값 (θ)	구조1		구조2	
			PI	E_PI	PI	E_PI
MLSS	WSR	0.0	14.718	11.084	0.184	94.616
		0.25	12.852	10.917	0.045	87.053
		0.4	12.367	11.444	0.094	146.484
		0.5	12.352	11.179	0.001	126.912
		0.6	12.347	11.302	0.780	166.815
		0.75	11.865	12.962	2.989	179.884
		1.0	9.347	116.427	0.003	1637.820
MLSS	RRSP	0.0	12.649	12.972	0.678	189.964
		0.25	12.654	12.994	0.711	175.097
		0.4	12.544	13.019	0.018	1207.820
		0.5	12.259	13.174	0.507	246.725
		0.6	11.955	13.608	0.538	302.129
		0.75	11.885	13.966	0.256	319.538
		1.0	9.899	35.287	0.002	32022.923
MLSS	DOSP	0.0	14.731	14.598	52.682	276.072
		0.25	14.725	14.599	6.625	1241.469
		0.4	14.767	14.901	11.971	140.216
		0.5	14.453	14.801	52.682	276.072
		0.6	14.153	15.135	52.682	276.072
		0.75	10.266	24.711	6.025	117.097
		1.0	10.058	26.547	0.023	26352.402
WSR	RRSP	0.0	12.872	11.389	8.832	211.311
		0.25	12.679	11.039	5.605	208.121
		0.4	12.585	11.088	5.871	176.693
		0.5	12.530	10.983	5.212	210.923
		0.6	12.309	11.221	0.194	106.363
		0.75	10.592	14.254	1.918	113.615
		1.0	8.287	25.768	0.0009	2304.115
WSR	DOSP	0.0	10.929	11.255	1.687	307.951
		0.25	10.737	11.278	0.557	346.433
		0.4	10.581	11.185	0.473	300.433
		0.5	10.600	11.231	0.003	284.192
		0.6	10.597	11.196	0.557	346.433
		0.75	9.363	14.924	0.212	359.757
		1.0	9.085	16.708	0.017	397.101
RRSP	DOSP	0.0	13.971	9.143	8.375	65.438
		0.25	13.761	9.138	3.446	57.667
		0.4	11.085	10.301	0.395	71.897
		0.5	11.140	10.480	2.774	98.579
		0.6	11.140	10.480	2.804	104.844
		0.75	10.534	12.321	0.100	51.206
		1.0	8.568	18.140	0.007	1773.415

표 4, 5에서 볼 수 있듯이 제안된 모델은 비선형이 강하고 복잡한 모델에서 뚜렷한 성능향상을 볼 수 있으며, 특히 테스트에 대한 성능지수의 향상은 뚜렷하게 나타난다. 또한, 하중값의 선택에 의해 주어진 조건에 따라 좀 더 향상된 최적화 모델을 구축할 수 있다.

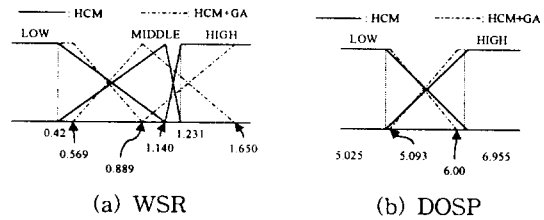
구조 1에서는 유전자 알고리즘에 의한 최적 모델로 $\theta = 0.5$ 일 때 WSR, DOSP 입력을 선택하였으며, 이 때 학습 성능지수는 10.600이고 테스트 성능지수는 11.231이다. 그림 19는 유전자 알고리즘에 의한 각 입력 변수의 파라미터 탐색과정을 보여주며, 그에 따른 입력 변수의 동정한 파라미터는 그림 20에서 보여준다. 또한 그림 21에서는 파라미터 탐색과정에 따른 모델의 최적화 과정을 보여준다.

구조 2에서는 유전자 알고리즘에 의한 최적 모델로 입력 RRSP, DOSP에 대해 $\theta = 0.75$ 를 선택하였으며, 이 때 학습

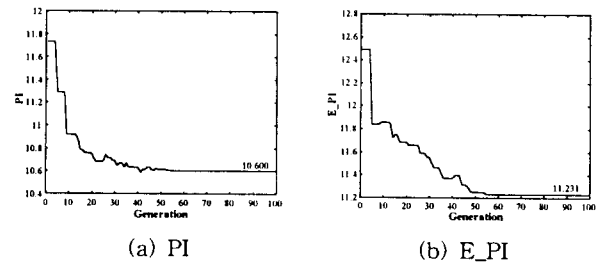
성능지수는 0.100이고 테스트 성능지수는 51.206이다. 그림 22, 23은 유전자 알고리즘에 의한 각 입력 변수에 대한 파라미터 탐색과정과 동정한 파라미터를 보여주며, 파라미터 탐색과정에 따른 모델의 최적화 과정은 그림 24에서 보여준다. 그림 25는 학습과 테스트 데이터에 대해 선택된 구조 2 모델의 출력을 원 출력과 비교하여 보여준다.



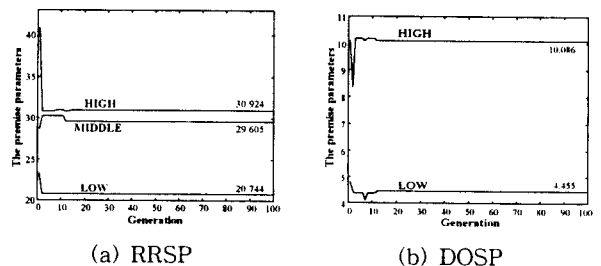
(a) WSR (b) DOSP
그림 19 유전자 알고리즘에 의한 구조 1의 파라미터 탐색과정 ($\theta = 0.5$)
Fig. 19 Search process of parameters of structure 1 by GA.



(a) WSR (b) DOSP
그림 20 구조 1의 동정한 파라미터 ($\theta = 0.5$)
Fig. 20 Identified parameters of structure 1.



(a) PI (b) E_PI
그림 21 유전자 알고리즘에 의한 구조 1의 모델 최적화 탐색과정 ($\theta = 0.5$)
Fig. 21 Optimal search process of structure 1 by GA.



(a) RRSP (b) DOSP
그림 22 유전자 알고리즘에 의한 구조 2의 파라미터 탐색과정 ($\theta = 0.75$)
Fig. 22 Search process of parameters of structure 2.

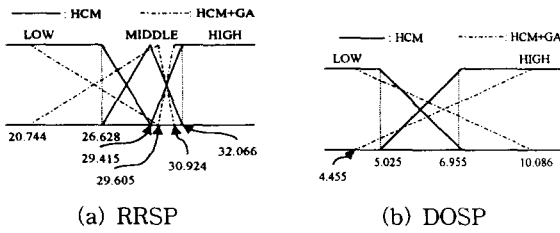


그림 23 구조 2의 동정된 파라미터 ($\theta=0.75$)
 Fig. 23 Identified parameters of structure 2.

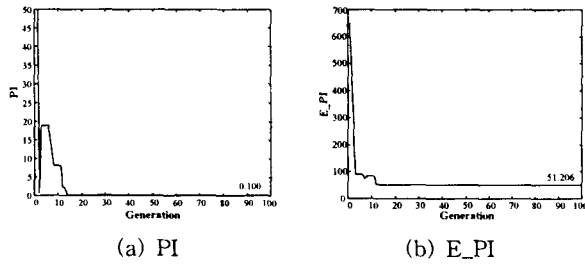


그림 24 유전자 알고리즘에 의한 구조 2의 모델 최적화 탐색과정 ($\theta=0.75$)
 Fig. 24 Optimal search process of structure 2 by GA.

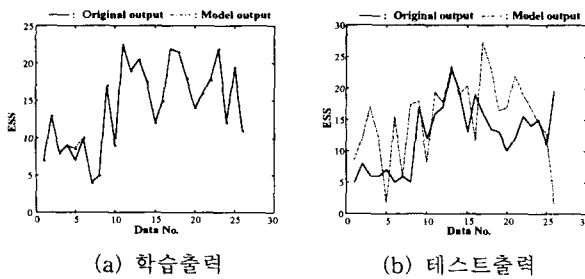


그림 25 구조2의 출력 ($\theta=0.75$)
 Fig. 25 Output of structure 2.

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

Table 6 Comparison of identification error with previous fuzzy models

모델		PI	E_PI	Total No. of rules
Neural Network		1.801	39.294	
Oh's model [15, 16]	Simplified Triangular	13.726	16.206	4
		14.107	16.563	6
	Linear Triangular	6.396	54.233	4
		1.461	80618.742	6
Min-Max +GA [20]	Simplified Triangular	12.847	14.155	4
		12.318	13.117	6
	Linear Triangular	7.087	32.970	4
		0.0002	187.911	6
Min-Max [20]	Simplified Triangular	15.875	17.371	4
		12.487	32.647	6
	Linear Triangular	6.396	54.529	4
		0.972	2480124.2	6
HCM	Simplified Triangular	12.649	21.025	6
	Linear Triangular	0.001	4557.270	6
Our model (HCM+GA)	Simplified Triangular	12.352	11.179	6
	Linear Triangular	0.001	126.912	6

표 6은 Min-Max[15, 16, 20]를 기준으로 초기값을 설정한 기존의 방법들과 성능지수를 비교하여 보여주고 있다. 여기서 Oh의 모델[15, 16]은 최적화 알고리즘인 개선된 컴플렉스만을 이용한 것이다. 표에서 알 수 있듯이 제안된 모델은 데이터의 특징을 모델에 반영함으로써 기존의 퍼지 모델들 보다 훨씬 좋은 결과를 얻었다. 기존의 모델들은 선형 추론에 있어서 규칙을 증가시키면 학습 성능은 향상되나 테스트 성능 면에서는 큰 에러를 발생시킨다. 물론 제안된 퍼지 모델 또한 선형 추론의 테스트 성능에서 좋은 결과를 보여 주지는 못했지만 기존의 방법들 보다 테스트 에러를 많이 줄였으며, 학습 및 테스트 성능을 향상시킨 퍼지 모델을 구축할 수 있었다.

6. 결론

비선형 공정에 대해 체계적이고 효율적인 모델을 얻기 위한 퍼지 모델을 제안하였다. 제안된 퍼지 모델은 입출력 데이터의 특징을 모델에 반영하기 위해 클러스터링 알고리즘 (HCM)을 사용하여 퍼지 입력 공간을 정의하였고, 또한 유전자 알고리즘과 하중값을 가진 목적함수를 이용하여 퍼지 모델을 최적화 하였다. 제안한 모델은 시뮬레이션에서 볼 수 있듯이 입력력 특징을 이용함으로써 복잡하고 비선형이 강한 공정에서 기존의 퍼지 모델들 보다 성능이 향상된 퍼지 모델을 설계할 수 있었고, 인간의 경험적 방법이 아닌 HCM 및 유전자 알고리즘을 이용하여 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 상호 연계를 통한 최적화 향상을 위한 방향을 제시하였다.

감사의 글

이 논문은 1998년도 원광대학교의 교비지원에 의해서 연구되었습니다.

참고 문헌

- [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] C.W. Xu, "Fuzzy system identification", IEEE Proceeding Vol. 126, No. 4, pp.146-150, 1989.

[6] M. Sugeno and T. Yasukawa, "Linguistic modeling based on numerical data", IFSA'91 Brussels, Computer, Management & System Science, PP.264-267, 1991.

[7] R.M. Tong, "The evaluation of fuzzy models derived from experimental data", Fuzzy Sets Syst., Vol. 13, pp.1-12, 1980.

[8] C.W. Xu and Y. Zailu, "Fuzzy model identification self-learning for dynamic system", IEEE Trans. on Syst. Man, Cybern., Vol. SMC-17, No. 4, pp.683-689, 1987.

[9] Box and Jenkins, "Time Series Analysis, Forecasting and Control", Holden Day, SanFrancisco, CA.

[10] 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.

[11] 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.

[12] David E. Goldberg, "Genetic Algorithms in search, Optimization & Machine Learning", Addison-wesley.

[13] Zbigniew Michalewicz, "Genetic Algorithms + Data Structure = Evolution Programs", Springer-Verlag.

[14] De Jong, K. A. "Are Genetic Algorithms Function Optimizers?", Parallel Problem Solving from Nature 2, Manner, R. and Manderick, B. eds., North-Holland, Amsterdam.

[15] 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)

[16] C.S. Park, S.K. Oh, and W. Pedrycz, "Fuzzy Identification by means of Auto-Tuning Algorithm and Weighting Factor", The Third Asian Fuzzy Systems Symposium(AFSS), PP.701-706, 1998

[17] 오성권, 김현기, 우광방, "활성오니공정의 퍼지 모델링과 제어시스템", 대한 전기학회 논문집 40권 10호 PP.1072-1083, 1990년 10월.

[18] 오성권, 우광방, "퍼지추론 방법에 의한 퍼지동정과 하수처리공정시스템 응용", 대한전자공학회 논문집 제 31권 B편 제 6호, PP.43-52, 1994년 6월

[19] 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.

[20] 박병준, 박춘성, 오성권, 김현기, "유전자 알고리즘을 이용한 퍼지 시스템의 최적화", 한국퍼지 및 지능시스템학회 '98 춘계학술대회 학술발표논문집 Vol 8, No.1, PP.112-115. 1998.

[21] 오성권, "퍼지모델 및 제어이론과 프로그램", 기다리출판사, 1993년 3월.

저 자 소 개

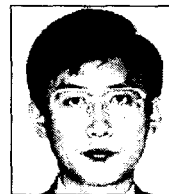


박 병 준(朴炳俊)

1972년 12월 1일 생. 1998년 원광대 제어계측공학과 졸업. 1998년~현재 동 대학교 제어계측공학과 석사과정. 주관심분야는 시스템자동화, 퍼지이론 및 신경회로망 응용, 지능형 모델링 및 제어 등.

Tel : (0653) 850-6342

E-mail : lcap@autosys.wonkwang.ac.kr



오 성 권(吳聖權)

1957년 2월 13日生. 1981년 연세대학교 전기공학과 졸업. 1983-89년 금성산전연구소(선임연구원). 1993년 연세대 대학원 전기공학과 졸업(공박). 1996-97 캐나다 Manitoba대학 전기 및 컴퓨터공학과 Post-Doc. 1993-현재 원광대 제어계측공학과 조교수. 주관심분야는 시스템자동화, 퍼지이론 및 신경회로망 응용, 지능형 모델링 및 제어 등.

Tel : (0653) 850-6342

E-mail : ohsk@wonnms.wonkwang.ac.kr



안 태 천(安泰天)

1955년 10월 11日生. 1978년 연세대학교 전기공학과 졸업. 1980년 동 대학원 전기공학과 졸업(석사). 1986년 동 대학원 졸업(공박). 1987-88년 스웨덴 Uppsala 대학 제어 및 시스템 공학과 Post-Doc. 1994-1995 미국 Georgia 공대 전기 및 컴퓨터공학부 교환교수. 1981-현재 원광대 제어계측공학과 교수. 주관심분야는 적용제어, 신호처리, 지능형 모델링 및 제어 등.

Tel : (0653) 850-6344

E-mail : tcahn@wonnms.wonkwang.ac.kr



김 현 기(金賢起)

1955년 2월 26日生. 1977년 연세대학교 전기공학과 졸업. 1985년 동 대학원 전기공학과 졸업(석사). 1991년 동 대학원 졸업(공박). 1997-현재 경기지역 산학연 기술지원단 단장. 1998-현재 산업기술정책연구소 심의위원. 1998-현재 사단법인 산학연 컨소시엄 전국협의회 부회장. 1989-현재 수원대학교 전기전자정보통신공학부 교수. 주관심분야는 시스템자동화 및 감시제어, 지능형 모델링 및 제어 등.

Tel : (0331) 220-2300

E-mail : hkkim@mail.suwon.ac.kr