1999년도 대한전기학회 하계학술대회 논문집 1999. 7.19-21

# HCM을 이용한 퍼지 모델의 On-Line 동정

박호성, 박병준, 오성권
원광대학교 전기전자공학부, 전라북도 익산시 신룡동 344-2 ㊞570-749

# On-line Identification of fuzzy model using HCM algorithm

Hosung Park, Byoungjun Park, Sungkwun Oh
Division of electrical & electronic Engineering, Wonkwang Univ., Iksan, KOREA

Abstract - In this paper, an adaptive fuzzy inference and HCM(Hard C-Means) clustering method are used for on-line fuzzy modeling of nonlinear and complex system. Here HCM clustering method is utilized for determining the initial parameter of membership function of fuzzy premise rules and also avoiding overflow phenomenon during the identification of consequence parameters. To obtain the on-line model structure of fuzzy systems, we use the recursive least square method for the consequent parameter identification. And the proposed on-line identification algorithm is carried out and is evaluated for sewage treatment process system.

## 1. 서 론

기존 퍼지 모델들의 대부분은 수동제어와 전문가의 경험으로부터 얻어진 퍼지 언어 명제로 표현되어 진다[1]. 그러나 다중입출력 시스템의 퍼지 모델링에 있어서는 퍼지 언어 명제의 수가 많은 문제와 퍼지관계가 다차원적이라는 어려운 문제들이 있다. 따라서 퍼지 모델의 동정[2,3,4,5,7]이 필연적이며 이것은 퍼지 모델의 성능을 결정한다.

본 논문에서는 클러스터링 알고리즘을 이용하여 퍼지 규칙을 동정한다. 또한 시스템이 동적으로 변하거나 주위 환경에 대해 민감한 시스템 또는 노이즈가 있는 시스템을 동정하고 공정의 실시간 적용을 위해 온라인[8]으로 동정한다. 퍼지 규칙의 동정을 위해 주어진 공정의 입출력 데이터를 이용한다. 입출력 데이터의 유기된 분포를 통해 전반부의 입력변수를 결정하고, 클러스터링 알고리즘을 이용하여 공정의 특성을 파악한다. 이것은 전반부 멤버쉽함수를 정의하는데 적용되며 멤버쉽함수는 벨(bell)모양을 사용한다. 클러스터링 알고리즘으로는 데이터들간의 거리를 기준으로 하여 근접한 정도를 측정하고 이를 바탕으로 데이터를 분류하는 HCM(Hard C-Mean)방법을 이용한다. 제안된 퍼지 모델은 후반부가 일차 선형식으로 표현된 구조로 선형추론에 의해 시행되며 온라인 동정을 위해 순환최소자승법(Recursive Least square Method)을 사용하여 퍼지 규칙의 후반부 파라미터를 실시간으로 조정한다. 하수처리 공정 데이터를 이용하여 제안된 모델의 타당성과 정확성을 확인한다.

## 2. 본 론

### 2.1 Hard C-Means clustering 방법

기존 퍼지 모델의 구조는 주어진 데이터에 대하여 최소값과 최대값사이를 임의의 개수로 균등하게 분할하여 일률적으로 멤버쉽함수를 정의했다. 그러므로 주어진 데이터의 특성을 살리지 못하고 모델의 성능을 저하시켰다. 따라서 그림 1과 같이 계통의 특성에 맞게 멤버쉽함수를 정의하는 것이 보다 성능을 향상시키는 방법이라 할 수 있겠다. 클러스터링 알고리즘이란 데이터의 분류를 위해서 사용하는 것으로 데이터의 내부가 비슷한 패턴, 속성, 형태 등의 기준을 통해 데이터를 분류하여 내부의 구조를 찾아내는 것이다.

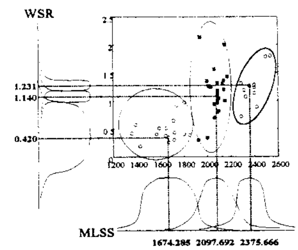


그림 1. HCM을 이용한 멤버쉽함수의 정의

본 논문에서는 HCM을 이용하여 입력 데이터의 특성을 파악한다. HCM방법은 데이터들간의 거리를 기준으로 근접한 정도를 측정하고, 이를 바탕으로 데이터를 분류하여 이용하는 방법이다. HCM방법을 통해 데이터를 특성별로 분류하고, 분류된 데이터들의 중심을 각 멤버쉽함수의 중심값으로 사용한다. HCM 클러스터링 알고리즘의 수행과정은 다음과 같다.

[단계 1] 클러스터의 개수 ($2 \le c \le n$)를 결정하고 중심 $c_i(i=1,\cdots,c)$값을 랜덤하게 선택하여 초기화 한다.

[단계 2] 소속행렬 $U$를 결정한다.

$$u_{ij}=\begin{cases}1 & \text{if } \| x_i - c_i \|^2 \le \| x_1 - c_k \|^2, \text{ for } k \ne 1 \\ 0 & \textit{otherwise}\end{cases} \quad (1)$$

[단계 3] 아래 식을 계산한다. 중심 근처에 충분히 다가갔는지 확인한다

$$J = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c} \Big( \sum_{k, x_k \in G_i} d(x_k - c_i) \Big) \quad (2)$$

[단계 4] 새로운 클러스터 중심 $c_i$을 생성하고 [단계 2]로 간다.

$$c_i = \frac{i}{|G_i|} \sum_{k, x_i \in G_i} x_i \quad (3)$$

### 2.2 퍼지 모델의 동정

퍼지 모델의 동정은 구조의 선택과 그들의 파라미터를 동정하는 문제이다. 또한 이들은 전반부와 후반부로 나누어진다. 퍼지 모델의 전반부 동정은 앞 절에서 제시한 입출력 데이터의 특성을 이용하여 동정한다. 본 논문에서 퍼지 모델의 후반부 구조는 일차선형식 형태의 구조이며, 시스템 입출력 데이터를 가지고 순환 최소자승법을 사용하여 파라미터를 동정한다.

후반부가 일차 선형식으로 표현된 것을 선형 추론법 또는 혼합 추론법이라 하며 이 퍼지 모델은 식 (4)의 형태를 가지는 구현규칙들로 구성된다.

$$
\begin{aligned}
&R^1\ IF\ \ x_1\ is\ A_{11}\ \text{and}\ \cdots\ \text{and}\ x_k\ is\ A_{1k}\ Then\\
&\qquad\qquad y=a_{10}+a_{11}x_1+\cdots+a_{1k}x_k\\
&\qquad\vdots\\
&R^j\ IF\ \ x_1\ is\ A_{j1}\ \text{and}\ \cdots\ \text{and}\ x_k\ is\ A_{jk}\ Then\\
&\qquad\qquad y=a_{j0}+a_{j1}x_1+\cdots+a_{jk}x_k\\
&\qquad\vdots\\
&R^n\ IF\ \ x_1\ is\ A_{n1}\ \text{and}\ \cdots\ \text{and}\ x_k\ is\ A_{nk}\ Then\\
&\qquad\qquad y=a_{n0}+a_{n1}x_1+\cdots+a_{nk}x_k
\end{aligned}
\tag{4}
$$

여기서, $R^j$는 $j(j=1,\cdots,n)$번째 규칙, $x_l(l=1,\cdots,k)$는 입력변수, $A_{jl}$은 퍼지 집합의 멤버쉽 함수, $a_{jl}(j=1,\cdots,n;\ l=0,\cdots,k)$는 후반부의 파라미터이고, $n$은 규칙 수이다. 추론된 값 $y^*$는 다음과 같다.

$$y^*=\frac{\sum_{j=1}^{n} w_{ji}\,y_j}{\sum_{j=1}^{n} w_{ji}}=\frac{\sum_{j=1}^{n} w_{ji}\,(a_{j0}+a_{j1}x_{1i}+\cdots+a_{jk}x_{ki})}{\sum_{j=1}^{n} w_{ji}} \tag{5}$$

$$w_{ji}=A_{j1}(x_{1i})\wedge\cdots\wedge A_{jk}(x_{ki}) \tag{6}$$

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(7)으로 정의된다.

$$PI=\frac{1}{m}\sum_{i=1}^{m}(y_i-y_i^*)^2 \tag{7}$$

후반부의 파라미터는 $a_{j0},\cdots,a_{jk}$로써 입출력 데이터가 주어졌을 때 오프라인의 경우 최소자승법에 의해 구해진다. 최소자승법에 의한 후반부 파라미터의 동정은 식(8)에 의해 구해진다.

$$\hat{a}=(X^TX)^{-1}X^TY \tag{8}$$

$$
\begin{aligned}
&x_i^T=[w_{1i},\cdots,w_{ni},\cdots,\ x_{ki}w_{1i},\cdots,x_{ki}w_{ni}]\\
&a^T=[a_{10},\cdots,\cdots,\ a_{1k},\cdots,a_{nk}]\\
&Y=[y_1,y_2,\cdots,y_m]^T,\quad X=[x_1,x_2,\cdots,x_m]^T
\end{aligned}
\tag{9}
$$

## 2.3 온라인(on-line) 동정

대부분의 선형 퍼지 추론 모델링의 경우 $N$개의 샘플링 데이터를 가지고서 최소자승법을 이용한 오프라인으로 후반부 파라미터를 결정하였다. 그러나 오프라인 모델링은 컴퓨터 시뮬레이션을 통해 모델을 획득하기 때문에 실제 플랜트에 적용하기에는 많은 문제점이 따른다. 이에 비해 온라인 모델링은 시스템이 동적으로 변하거나 주위 환경에 대해 민감한 시스템을 동정하는데 필요하다. 주위 환경에 민감한 시스템이나 노이즈가 있는 시스템은 오프라인 보다는 온라인에 의한 모델링이 필요하고 온라인 모델링을 위한 전체 시스템 구축이 중요하다. 특히, 실제 플랜트에 모델을 적용하기 위해서는 온라인 모델링에 의한 모델 구축이 필수적이다. 그림 2는 기본적인 온라인 모델링에 의한 파라미터 동정의 구조이다. 대상 시스템의 출력과 모델의 출력의 오차를 근거로 모델의 파라미터는 순환 최소자승법에 의해서 실시간으로 조정된다.

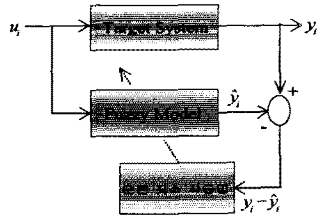


그림 2 온라인 모델링에 의한 파라미터 동정

$N$개의 샘플링 데이터에 대해서는 최소자승법을 이용하여 식 (10)과 같이 후반부 파라미터를 결정한다

$$a_N=(X_N^TX_N)^{-1}X_N{}^TY_N \tag{10}$$

여기서,

$$X_N=\begin{bmatrix} w_{11} & \cdots & w_{j1} & \cdots & x_{k1}w_{11} & \cdots & x_{k1}w_{j1}\\ w_{12} & \cdots & w_{j2} & \cdots & x_{k2}w_{12} & \cdots & x_{k2}w_{j2}\\ \vdots & & \vdots & \vdots & & & \vdots\\ w_{1N} & \cdots & w_{jN} & \cdots & x_{kN}w_{1N} & \cdots & x_{kN}w_{jN}\end{bmatrix}$$

$$a_N=[\,a_{10}\cdots a_{j0}\ \ a_{11}\cdots a_{j1}\cdots a_{1k}\cdots a_{jk}]$$

$$Y_N=\begin{bmatrix} y_1\\ y_2\\ \vdots\\ y_N\end{bmatrix},\quad \varepsilon_N=\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \vdots\\ \varepsilon_N\end{bmatrix}$$

새로운 입력 데이터가 들어오는 경우 순환 최소자승법(Recursive Least Square Method)을 적용하여 새로운 파라미터를 계산한다. $N+1$번째부터 입력 데이터에 대해서 파라미터 추정은 식 (11)과 같이 표현할 수 있다.

$$a_{N+1}=\left(\begin{bmatrix}X_N\\ x_{N+1}\end{bmatrix}^T\begin{bmatrix}X_N\\ x_{N+1}\end{bmatrix}\right)^{-1}\begin{bmatrix}X_N\\ x_{N+1}\end{bmatrix}^T\begin{bmatrix}Y_N\\ y_{N+1}\end{bmatrix} \tag{11}$$

여기서

$$x_{N+1}=[\,w_{1N+1}\cdots w_{jN+1}\vdots\cdots\vdots\ x_{kN+1}w_{1N+1}\cdots x_{kN+1}w_{jN+1}]$$

이것은 식 (14)과 같이 간략하게 표현하여 선형 퍼지 추론의 후반부 파라미터를 결정한다. 그러므로 다음과 같이 순환 최소자승법으로 후반부 파라미터를 새로운 입력 데이터에 대해서 조정할 수 있다.

$$P_N=[X_N^T X_N]^{-1} \tag{12}$$

$$K_{N+1}=\frac{P_N\,x^T{}_{N+1}}{1+x_{N+1}P_N\,x^T{}_{N+1}} \tag{13}$$

$$a=a_N+K_{N+1}[y_{N+1}-x_{N+1}a_N] \tag{14}$$

$$P_{N+1}=P_N-\frac{P_N\,x^T{}_{N+1}x_{N+1}P_N}{1+x_{N+1}P_N\,x^T{}_{N+1}} \tag{15}$$

최소자승법을 사용하지 않고 처음 $N$개의 데이터부터 순환 최소자승법 알고리즘을 사용할 경우에 초기 파라미터는 다음과 같이 초기화한다.

$$a(0)=0 \tag{16}$$

$$P(0)=\alpha I \tag{17}$$

여기서 $\alpha$는 임의의 큰 양수이며 보통 100보다 큰 양수를 사용하고 $I$는 단위행렬이다.

## 3. 하수처리 공정의 온라인 모델링

현재 대부분의 하수처리 플랜트는 제어 공정에서 조절 데이터를 얻기 위해 수학적 모델을 이용하고 있다. 그러나 이러한 수학적 모델이 하수처리공정 변수들 사이의 관계를 정확하고 효과적으로 설정하지 못하므로, 제안된

방법을 통해 수도권 하수처리장 중의 하나를 모델로 선정하여 1년간의 수질 데이터[4]를 가지고 모델링한다. 그림 3은 하수처리시스템의 개요도이다.

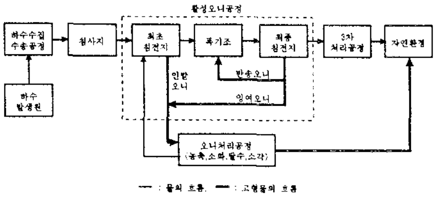


그림 3. 활성오니공정을 이용한 하수처리시스템의 개요도

활성오니공정은 하수처리를 위한 다양한 방법중 가장 많이 사용되는 방법으로, 침사지, 최초침전지, 폭기조 및 최종침전지로 구성된다. 침사지와 최초침전지에서는 하수중에 포함된 이물질들이 중력에 의해 침강되고, 폭기조에서는 하수에 공기를 흡입시킨다. 활성오니란 유기물 제거의 중심을 이루는 미생물로 폭기조에서 자연 발생되며, 응집한 미생물의 덩어리를 플록이라 한다. 활성오니는 폭기조내에서 공급되는 산소와 결합하여 유기물을 물과 탄산가스로 분해한다. 최종침전지에서는 플록을 침강, 제거하고 맑은 물을 외부로 흘려보낸다. 한편 최종침전지에서 침강한 미생물은 폭기조 입구에서 재순환되어 다시 유기물 제거에 사용된다. 본 논문에서는 활성오니공정을 폭기조 및 최종침전지를 하나의 단일 모델로 설정하여, 실제 최초침전지로 유입되는 생물학적 산소요구량 및 유입수의 부유물 농도를 측정하고, 최종침전지에서 유출되는 생물학적 산소요구량 및 부유물 농도를 통하여 하수처리 정도를 비교한다. 하수처리 공정은 혼합액 부유물(MLSS : Mixed Liquid Suspended), 잉여오니 흐름율(WSR : Waste Sludge Ratio), 반송율 설정치(RRSP : Recycled Return Sludge) , 용존산소 설정치(DOSP : Dissolved Oxygen Set Point)를 입력으로 하고, 부유물의 농도(ESS : Effluent Suspended Solids)를 출력으로 하는 4입력 1출력 구조를 가진다.

본 논문에서는 하수처리 공정의 입력 중 2개를 선택하여 모델링을 수행하였으며 성능지수(PI)는 식 (7)을 이용하였다.

표 1. 수처리 공정의 모델링 결과

| 입력변수 | 규칙수 | 학습 | 테스트 |
|---|---|---|---|
| MLSS, WSR | 9 | 0.1146 | 10.0445 |
| MLSS, RRSP | 9 | 0.1173 | 2.8038 |
| MLSS, DOSP | 9 | 40.0787 | 84.3933 |
| WSR, RRSP | 9 | 0.3424 | 7.9254 |
| WSR, DOSP | 9 | 0.5831 | 7.8464 |
| RRSP, DOSP | 9 | 0.2695 | 3.8724 |

$$
\begin{aligned}
&R^1 : \text{If } WSR \text{ is } NA_1 \text{ and } DOSP \text{ is } NA_2, \text{ then } ESS = \boldsymbol{c_1} \cdot \boldsymbol{X}\\
&R^2 : \text{If } WSR \text{ is } NA_1 \text{ and } DOSP \text{ is } ZE_2, \text{ then } ESS = \boldsymbol{c_2} \cdot \boldsymbol{X}\\
&R^3 : \text{If } WSR \text{ is } NA_1 \text{ and } DOSP \text{ is } PO_2, \text{ then } ESS = \boldsymbol{c_3} \cdot \boldsymbol{X}\\
&R^4 : \text{If } WSR \text{ is } ZE_1 \text{ and } DOSP \text{ is } NA_2, \text{ then } ESS = \boldsymbol{c_4} \cdot \boldsymbol{X}\\
&R^5 : \text{If } WSR \text{ is } ZE_1 \text{ and } DOSP \text{ is } ZE_2, \text{ then } ESS = \boldsymbol{c_5} \cdot \boldsymbol{X} \qquad (18)\\
&R^6 : \text{If } WSR \text{ is } ZE_1 \text{ and } DOSP \text{ is } PO_2, \text{ then } ESS = \boldsymbol{c_6} \cdot \boldsymbol{X}\\
&R^7 : \text{If } WSR \text{ is } PO_1 \text{ and } DOSP \text{ is } NA_2, \text{ then } ESS = \boldsymbol{c_7} \cdot \boldsymbol{X}\\
&R^8 : \text{If } WSR \text{ is } PO_1 \text{ and } DOSP \text{ is } ZE_2, \text{ then } ESS = \boldsymbol{c_8} \cdot \boldsymbol{X}\\
&R^9 : \text{If } WSR \text{ is } PO_1 \text{ and } DOSP \text{ is } PO_2, \text{ then } ESS = \boldsymbol{c_8} \cdot \boldsymbol{X}
\end{aligned}
$$

수처리 공정의 입력 변수가 $[WSR,\ DOSP]$에 대하여 최종적으로 동정된 9개의 규칙은 식 (18)과 같다. 여기서, $\boldsymbol{X} = [1,\ MLSS,\ WSR,\ RRSP,\ DOSP]$이고 $\boldsymbol{c}_i$는 다음 행렬의 $i$번째 열 벡터이다.

$$
\boldsymbol{C} = \begin{bmatrix}
0.445664 & -0.000496 & 9.348637 & 0.925011 & -7.336737\\
0.594517 & -0.213888 & 5.979132 & 8.761591 & 19.632444\\
0.236534 & 0.074335 & 0.872586 & 1.903394 & 2.459841\\
-0.212140 & -0.179357 & -8.813239 & 13.089543 & 5.249285\\
1.317529 & 0.029557 & -2.232782 & -1.342177 & -0.033137\\
0.933700 & -0.094575 & 5.954466 & 4.836252 & 6.403017\\
0.088376 & 0.205534 & -0.744187 & -2.247697 & -2.976803\\
-0.024555 & 0.106215 & -0.554416 & -5.141068 & -15.110278\\
0.621924 & 0.152472 & -7.144414 & -8.083989 & -12.663297
\end{bmatrix}
$$

멤버쉽함수 $NA_i,\ ZE_i,\ PO_i\ (i=1,2)$은 식 (19)과 같이 파라미터 a, b, c를 가진 벨 모양 함수에 의해서 정의된다.

$$
\mu_{A_i}(x) = \frac{1}{1+\left[\left(\frac{x-c_i}{a_i}\right)^2\right]^{b_i}} \qquad (19)
$$

## 3. 결 론

본 논문은 퍼지 모델을 이용하여 비선형성이 강하고 복잡한 시스템에 대해서 온라인 모델링을 수행하였다. 또한 클러스터링 알고리즘(HCM)을 사용하여 퍼지 규칙의 전반부 멤버쉽함수를 정의함으로써 공정의 특성을 퍼지 모델에 반영하였다. 온라인 모델링을 위해 순환 최소자승법을 사용하였으며 하수처리공정에 적용함으로써 실시간으로 모델을 획득하였다. 본 논문에서는 온라인 모델링 방법에 의해 체계적이고 효율적인 퍼지 모델을 구축하였고 환경 시스템으로써 하수처리 공정, 폐수처리 공정 및 쓰레기 소각로 등에 응용할 수 있을 것으로 기대된다.


**감사의 글**

이 논문은 1998년 한국학술진흥재단의 공모과제 연구비에 의하여 연구되었습니다.


**[참 고 문 헌]**


[1] E.H. Mamdani, "Application of fuzzy algorithm for control of a simple dynamic plant", *Proc. IEE* 121(12), pp.1585-1588, 1974.

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

[3] W. Pedrycz, "An identification algorithm in fuzzy relational system", *Fuzzy Sets Syst.*, Vol. 13, pp.153-167, 1984.

[4] 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.

[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.*, 1999.(To appear)

[6] 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.

[7] 박병준, 오성권, 안태천, 김현기,"유전자 알고리즘과 하중값을 이용한 퍼지 시스템의 최적화", 대한전기학회 논문집, 1999,6월 개재예정.

[8] 오성권, "퍼지모델 및 제어이론과 프로그램", 기다리출판사, 1999년 3월.