

# 계층적 클러스터링과 Gaussian Mixture Model을 이용한 뉴로-퍼지 모델링

## A Neuro-Fuzzy Modeling using the Hierarchical Clustering and Gaussian Mixture Model

김승석, 박근창, 유정웅, 전명근

Sung-Suk Kim, Keun-Chang Kwak, Jeong-Woong Ryu, and Myung-Geun Chun\*\*

충북대학교 전기전자컴퓨터 공학부 컴퓨터 정보통신 연구소

### 요 약

본 논문에서는 계층적 클러스터링과 GMM을 순차적으로 이용하여 최적의 파라미터를 추정하고 이를 뉴로-퍼지 모델의 초기 파라미터로 사용하여 모델의 성능 개선을 제안한다. 반복적인 시도 중 가장 좋은 파라미터를 선택하는 기존의 알고리즘과 달리 계층적 클러스터링은 데이터들 간의 유클리디언 거리를 이용하여 클러스터를 생성하므로 반복적인 시도가 불필요하다. 또한 클러스터링 방법에 의해 퍼지 모델링을 행하므로 클러스터와 동일한 갯수의 적은 규칙을 갖는다. 제안된 방법의 유용함을 비선형 데이터인 Box-Jenkins의 가스로 예측 문제와 Sugeno의 비선형 시스템에 적용하여 이전의 연구보다 적은 규칙으로도 성능이 개선되는 것을 보였다.

### Abstract

In this paper, we propose a neuro-fuzzy modeling to improve the performance using the hierarchical clustering and Gaussian Mixture Model(GMM). The hierarchical clustering algorithm has a property of producing unique parameters for the given data because it does not use the object function to perform the clustering. After optimizing the obtained parameters using the GMM, we apply them as initial parameters for Adaptive Network-based Fuzzy Inference System. Here, the number of fuzzy rules becomes to the cluster numbers. From this, we can improve the performance index and reduce the number of rules simultaneously. The proposed method is verified by applying to a neuro-fuzzy modeling for Box-Jenkins's gas furnace data and Sugeno's nonlinear system, which yields better results than previous ones.

**Key Words :** Hierarchical clustering, Dendrogram, Gaussian Mixture Model, Expectation-Maximization algorithm, Adaptive Network-based Fuzzy Inference System

### 1. 서 론

뉴로-퍼지 시스템은 비선형 시스템에 대하여 기존의 미분을 기반으로 하는 수학적인 모델링 방법에서 접근하기 어려운 부분에 대하여 좋은 성능을 보여 왔다[1]. 이러한 지능시스템의 성능은 패턴 인식, 영상, 음성 처리, 지능 제어 등 다양한 분야에서 응용되어 좋은 결과를 보여 왔으며, 보다 향상된 성능이나 구조를 가지기 위한 연구들이 이어져 신경회로망의 학습기능과 퍼지 시스템의 추론기능을 조합한 뉴로-퍼지 시스템 등이 제안되었다[2][3]. 그러나 일반적인 뉴로-퍼지 시스템에서는 여러 가지 문제점들이 지적되고 있다. 대표적으로 입력의 차원과 각 소속함수의 수가 증가할수록 지수함수적으로 증가하는 그리드 분할에 의한 규칙은 시스템의

학습 시간을 증가시키며 전체 시스템의 크기 역시 증가한다[4]. 이러한 문제점을 해결하기 위하여 연구된 방법 중 하나가 클러스터링에 의한 규칙 생성 및 초기 파라미터 최적화이다. Lee[5]는 자기-적응 뉴로-퍼지 추론시스템을 통하여 자기 적응적이며 구성적인 내부 구조를 가지는 뉴로 퍼지 시스템을 제안하였다. 이 방법은 자기 적응적인 클러스터링 방법을 통하여 입력의 차원과 무관한 규칙을 가지는 뉴로-퍼지 시스템을 구성하였다. 또한 Zhang 등은 B-spline을 통한 매핑으로, 최적의 공간 분할 능력과 언어적 함수의 표현을 확장하려 하였다[6]. 그리고 Shapiro는 신경회로망과 퍼지 논리, 유전알고리즘 등을 이용하여 클러스터링과 모델 성능개선에 관련된 문제들을 해결하려 하였다[7]. 그러나 위에 언급된 대부분의 방법들은 목적함수를 기반으로 한 파라미터 최적화 문제로써 초기 파라미터 결정에 따라 전체의 모델링 성능에 영향을 받는 문제점이 있었다.

본 논문에서 제안한 방법은 클러스터링 초기 파라미터 선택과 최적화 방법으로 계층적 클러스터링[8]과 Gaussian Mixture Model (GMM)[9]을 이용하였다. 즉, 클러스터링 초기 파라미터 선택에서 각 데이터의 유클리디언 거리 정보를

접수일자 : 2003년 3월 17일

완료일자 : 2003년 8월 18일

감사의 글 : 본 연구는 한국과학재단 목적기초연구(R01-2002-000-00315-0) 지원으로 수행되었음.

이용하여 순차적으로 클러스터를 줄여나가는 계층적인 클러스터링을 이용하였다. 제안된 방법의 특징으로는, 목적함수를 사용하지 않음으로써 목적함수를 최적화하는 방향으로 클러스터링을 실시하는 통상의 알고리즘과는 달리 초기치 결정 문제나 목적함수 최적화 문제로부터 생기는 문제점이 없다. 이렇게 계층적 클러스터링을 통하여 추정된 파라미터를 GMM을 이용하여 최적화 하여 뉴로-퍼지 모델의 초기 파라미터로 사용하였다. GMM에서 가지는 초기 파라미터에 의한 국부적 수렴 문제에서, 초기 파라미터에 데이터 클러스터 정보를 사전에 제공함으로써 빠른 수렴과 좋은 성능을 기대할 수 있다. 이러한 순차적인 클러스터 파라미터 최적화는 각 알고리즘의 장점을 이용함으로써 전체 시스템의 성능을 개선하였다. 이를 전형적인 비선형 모델인 Box-Jenkins 가스로 데이터와 Sugeno의 비선형 시스템에 적용하여 제안된 방법의 개선된 성능과 유용성을 보인다. 본 논문의 구성은 다음과 같다. 2장에서는 계층적 클러스터링과 GMM의 클러스터 추정 방법을 소개하였고, 3장에서는 클러스터링에 의한 뉴로-퍼지 시스템에 대하여 설명하였으며, 4장에서 시뮬레이션 및 결과를 통하여 제안된 방법의 결과와 기존 연구결과와의 차이를 나타내었다. 끝으로 5장에서 결론을 통하여 제안된 방법의 우수성과 향후 개선 방향 및 연구 방향에 대하여 제시하였다.

## 2. 계층적 클러스터링과 GMM

### 2.1 계층적 클러스터링

계층적 클러스터링에서, 데이터 샘플들은 분할된 수열로 묶여진다[8].  $m$ 개의 클러스터에서  $n$ 개의 샘플들의 분할된 수열에 대하여 고려하면 다음과 같이 생각할 수 있다. 각각의 분할에서 임의의 하나는  $n$ 개의 클러스터를 가지고 있고, 각각은 정확히 하나의 샘플을 가지고 있다. 다음 단계에서 각 샘플들은 유사도 측정에 의하여  $n-1$ 개의 클러스터로 나누어진다. 이러한 과정이 클러스터의 수가 1이 될 때까지 반복적으로 수행되고 이러한 수열을 통하여 클러스터를 생성하는 방법을 계층적 클러스터링이라 한다. 계층적 클러스터링에서 각각의 클러스터링은 그림 1과 같이 Dendrogram으로 표현할 수 있다.

Dendrogram에서 각 계층에서의 클러스터들의 유사성을 쉽게 확인할 수 있다. 계층적 클러스터링은 흡수(agglomerative) 과정과 분리(divisive) 과정으로 크게 두 개의 처리 방법으로 나누어진다. 흡수 과정에서는 Dendrogram에서 볼 수 있듯이 아래에서 위로의 처리과정을 통하여 클러스터를 흡수한다.  $n$ 개의 각각의 클러스터들과 수열의 형태가 연속적인 흡수 클러스터링 과정을 통하여 이루어진다. 분리 과정은 위에서 아래로 또는 분류과정으로, 하나의 클러스터가  $n$ 개의 샘플을 가지고 있으며 연속적인 분리 과정에 의하여 수행된다. 구별과정과 흡수과정은 최적화 방법과 연관되어 있다. 분리 과정에서, 초기에 모든 샘플은 하나의 클러스터에 속해 있다고 가정한다. 초기 클러스터는 계층간의 거리를 최대화하는 방법에 의하여 두 개의 클러스터로 나누어진다. 이러한 과정을 통하여 샘플들은 각각의 새로운 클러스터로 다시 할당된다. 이를 반복적으로 새로운 클러스터를 생성하는 방법으로 계속 수행되며 이러한 구조는 이진 트리 구조로 처리된다. 클래스 내부의 거리가 임계치 이하로 떨어질 경우 정지조건으로 하여 클러스터가 계속 나누어지는 것을 방지한다.

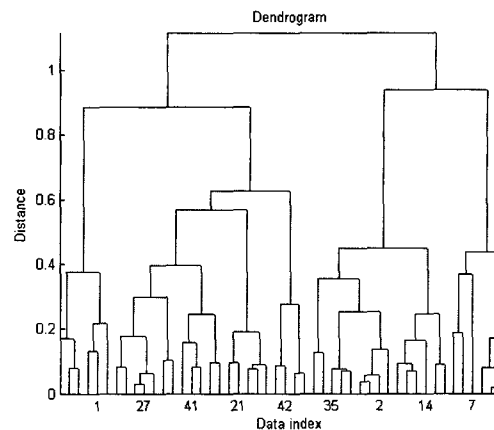


그림 1. Dendrogram  
Fig. 1 Dendrogram

목적함수를 이용하지 않고 단지 각 데이터 또는 클러스터 간의 유클리디언 거리만을 이용하여 클래스를 생성하므로써 계층적 클러스터링은 초기 파라미터가 고정되어 있다고 볼 수 있다. 즉, 초기 전체 데이터가 각각 하나의 샘플 또는 초기 파라미터가 되므로 초기값 결정문제에 대하여 고려할 필요가 없다. 초기 파라미터를 임의로 설정하는 K-means, Fuzzy C-mean (FCM) 등과 같은 클러스터링 방법의 경우 초기 파라미터의 설정에 따라 매 시도 때마다 추정된 클러스터의 정보가 달라지는 문제를 가지는 반면 계층적 클러스터링은 항상 같은 초기 파라미터 정보를 유지하는 특징을 가지고 있다. 이러한 성질에 의하여, 기존의 알고리즘의 경우 클러스터의 수와 데이터의 특징의 수가 차이나거나 편중된 데이터 또는 복잡한 형태를 가지는 데이터의 경우 반복 실험마다 다른 파라미터를 추정하는 경우가 발생하였다. 하지만 계층적 클러스터링의 경우 기존의 알고리즘들에서 볼 수 있었던 데이터의 편중이나 복잡성이 존재하는 데이터에 대하여 매 시도마다 파라미터의 값이 변동하는 문제점을 해결할 수 있다.

### 2.2 Gaussian Mixture Model

계층적 클러스터링에서 추정된 파라미터는 다시 GMM의 초기 파라미터로 주어져 최적화를 실시한다[9][10]. 주어진 파라미터는 각 클러스터의 추정된 중심값으로써 이를 초기치로 GMM 알고리즘을 다음과 같이 수행한다. 주어진 데이터를 불완전한 관측데이터라 가정하고 구하고자 하는 파라미터를 완전한 데이터의 파라미터라고 가정하면 다음과 같이 설명할 수 있다. GMM의 기본 알고리즘은 초기 파라미터  $\theta$ 로부터 시작하여 관측 데이터의 likelihood가 증가하도록  $\theta$ 를 갱신하는 반복적인 파라미터 최적화의 반복이다. 이를 최적화 하기 위하여 Expectation Maximization (EM) 단계를 통하여 반복 수행한다.

EM 단계의 기본 구성은 다음과 같다. 먼저 데이터  $x$ 를 이용하여 파라미터  $\theta$ 를 추정하는데 Maximum Likelihood (ML)을 이용한다.

$$\arg \max_{\theta} \ln p(D_x | \theta) \quad (1)$$

즉 주어진 관측 데이터  $y$ 와 현재의 추정된 파라미터  $\theta$ 에서  $p(x | \theta)$ 의 기대값을 최대화 하는 것이다. 이를 단계별로

보면 다음과 같다.

단계 1 : 먼저 파라미터  $\theta^0$ 를 초기화한다.

$$p(x|\omega_j, \theta) = \frac{1}{(2\pi)^{(d/2)}|\Sigma_j|^{1/2}} \times \exp\left[-\frac{1}{2}(x-\mu_j)^T \Sigma_j^{-1}(x-\mu_j)\right] \quad (6)$$

여기서  $g_j$ 가  $P(\omega_j)$ 라고 한다면 다음과 같이 표현할 수 있다.

$$h_{kj} = \frac{g_j |\Sigma_j|^{1/2} \exp\left[-\frac{1}{2}(x_k - \mu_j)^T \Sigma_j^{-1}(x_k - \mu_j)\right]}{\sum_l g_l |\Sigma_l|^{1/2} \exp\left[-\frac{1}{2}(x_k - \mu_l)^T \Sigma_l^{-1}(x_k - \mu_l)\right]} \quad (7)$$

이를 Maximization 단계에서 파라미터  $\theta$ 를 다음과 같이 갱신한다.

$$\mu_j^{t+1} = \frac{\sum_k h_{kj} x_k}{\sum_k h_{kj}} \quad (8)$$

$$\Sigma_j^{t+1} = \frac{\sum_k h_{kj} (x_k - \mu_j)(x_k - \mu_j)^T}{\sum_k h_{kj}} \quad (9)$$

$$g_j^{t+1} = \frac{1}{n} \sum_k h_{kj} \quad (10)$$

여기서  $t$ 는 알고리즘의 반복 횟수를 말한다.

EM 알고리즘의 처리과정을 그림 2에 나타내었다. GMM에서의 EM 알고리즘은 주어진 데이터의 파라미터 추정에서 통계적 방법을 통하여 Gaussian 분포의 파라미터 최적화 추정이다. 반복 횟수가 증가하면서 GMM의 중심과 분산에 대한 파라미터는 로그 함수가 최대화되는 방향으로 최적화 된다. 여기서 최적화된 파라미터 중 중심값을 뉴로-퍼지 모델의 전제부 파라미터 중 Gaussian 소속함수 중심의 초기치로 설정한다. 순차적으로 초기화되고 최적화된 중심값은 뉴로-퍼지 모델의 초기 학습부터 좋은 성능을 나타낸다. 또한 학습에서의 초기 파라미터에 의한 국부적 최적화로 빠져 모델의 성능 개선에 장애가 되는 것에 대하여 이를 해결할 수 있는 하나의 해결책이 된다.

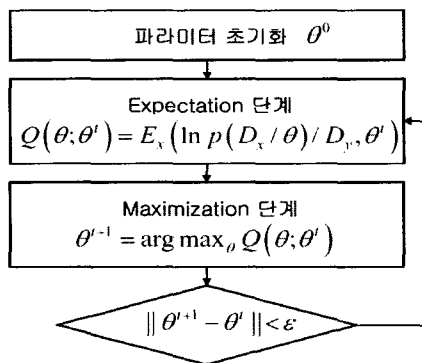


그림 2. EM 알고리즘  
Fig. 2 EM algorithm

### 3. 뉴로-퍼지 시스템

신경회로망의 학습기능과 퍼지 시스템의 추론 기능으로 구성된 뉴로-퍼지 시스템은 다양한 분야에서 좋은 성능을 보여왔다. 퍼지 시스템은 언어적 입력과 언어적 출력을 가지는 Mamdani 퍼지 모델과 언어적 입력과 1차 선형 방정식 형태의 Takagi-Sugeno-Kang (TSK) 퍼지 모델이 있는데 본 논문에서는 Jang이 제안한 TSK 퍼지 모델인 적응 뉴로-퍼지 시스템(Adaptive Network-base Fuzzy Inference System : ANFIS)을 이용하였다[2]. 두 개의 입력을 가지는 모델에 대하여 고려하면 다음과 같다.

$$R^1 : \text{IF } x \text{ is } A_1 \text{ and } y \text{ is } B_1 \\ \text{THEN } f_1 = p_1 x + q_1 y + r_1 \quad (11)$$

$$R^2 : \text{IF } x \text{ is } A_2 \text{ and } y \text{ is } B_2 \\ \text{THEN } f_2 = p_2 x + q_2 y + r_2$$

여기서  $x, y$ 는 입력이고  $A_i, B_i$ 는 언어적 형태의 소속함수이며  $p_i, q_i, r_i$ 는 결론부 파라미터이다. 이를 각각의 층으로 나누어 보면 다음과 같다.

1층 : 이 층의 모든 노드는 다음과 같은 소속 값을 가진 출력을 가진다.

$$O_{1,i} = \mu_{A_i}(x), \text{ for } 1, 2 \\ O_{2,i} = \mu_{B_i}(y), \text{ for } 3, 4 \quad (12)$$

여기서  $\mu_{A_i}(x) = \exp\left\{-\left(\frac{x-c_i}{a_i}\right)^2\right\}$

2층 : 1층에서 구한 소속 값을 각 규칙의 곱으로 가중치 (Weight)를 나타낸다.

$$O_{2,i} = w_i = \mu_{A_i}(x) \times \mu_{B_i}(y), \quad i = 1, 2 \quad (13)$$

3층 : 모든 노드는 다음과 같이 정규화 된 가중치 값으로 출력하게 된다.

$$O_{3,i} = \bar{w}_i = \frac{w_i}{w_1 + w_2}, \quad i = 1, 2 \quad (14)$$

4층 : 정규화된 가중치 값과 결론부의 곱으로 출력을 나타낸다.

$$O_{4,i} = \bar{w}_i f_i = \bar{w}_i (p_i x + q_i y + r_i) \quad (15)$$

5층 : 가중 평균법(Weighted average method)에 의한 최종 출력을 구한다.

$$O_{5,i} = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \quad (16)$$

결론부 파라미터 추정은 전제부 파라미터와  $P$ 개의 학습 데이터가 주어지면 다음과 같은  $P$ 개의 선형 방정식을 행렬로 표현하면 다음과 같다.

$$AX = B \quad (17)$$

여기서  $A$ 가 정방행렬이고 정칙일 경우는 위의 수식으로 구하고자 하는 파라미터  $X$ 를 구하지만 일반적으로 파라미터의 수보다 데이터의 수가 더 많음으로 의사역행렬을 이용하여 다음과 같이 나타낼 수 있다.

$$X^* = (A^T A)^{-1} A^T B \quad (18)$$

여기서  $A^T$ 는  $A$ 의 전치 행렬이다. 이를 이용하여 전체부 소속함수의 파라미터와 결론부 1차 선형 방정식의 파라미터를 모두 구할 수 있다.

이를 개념적으로 표현하면 그림 3과 같이 구성된다. 즉 언어적 입력형태의 전체부와 1차 선형 방정식 형태의 결론부로 구성되며, 결론부가 1차 선형 방정식 형태이므로 일반적인 퍼지 시스템에서 필요로 하는 비퍼지화 단계가 생략된다.

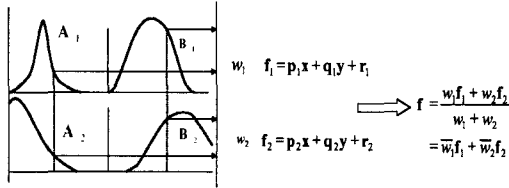
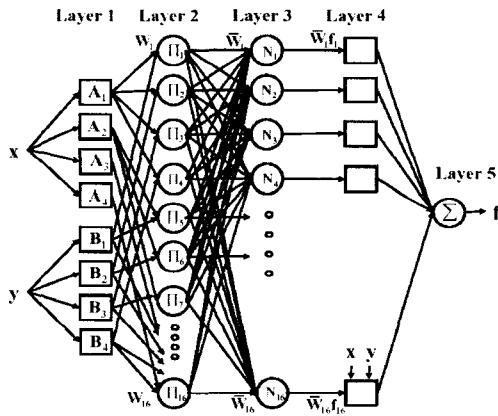
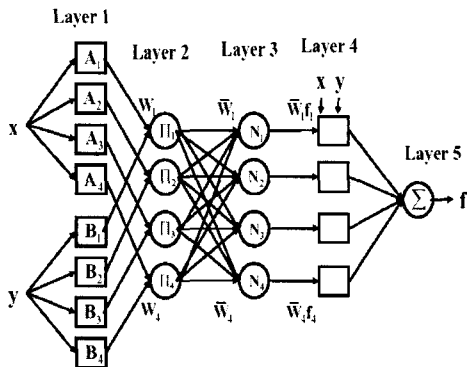


그림 3. TSK 퍼지 추론  
Fig. 3 TSK Fuzzy Inference

입력이 2개이며 각 소속함수의 수를 4개로 하였을 경우 전체적인 구성을 일반적인 ANFIS와 비교하면 그림 4와 같이 표현할 수 있다.



(a) Basic ANFIS



(b) Proposed ANFIS

그림 4. ANFIS 비교  
Fig. 4 Compare with ANFIS

그림 5에서 볼 수 있듯이 제안된 방법에 의한 ANFIS의 경우 규칙으로 생각할 수 있는 Layer 2에서 Layer 4까지의 구성을 입력공간 클러스터링에 의한 규칙 조합으로 구성, 기본 ANFIS의 구조보다 단순하다. 즉, 입력으로 표현되는 Layer 1과는 무관하게 Layer 2~4는 클러스터의 수와 일치하는 구조를 가지게 된다. 또한 그림 6에서 소속함수에 의한 규칙 생성을 나타내었다. 입력공간에서의 클러스터링이 규칙 공간에서의 규칙 생성에 이용됨으로서 제안된 모델의 성능 및 규칙 증가 억제에 동시에 만족할 수 있다.

일반적으로 전방향 처리는 최소자승법(Least Square Estimation)로 추정하며 역방향 처리는 파라미터의 기울기를 이용하여 미분에 근거한 역전파 알고리즘 (Back-propagation)을 이용하였다.

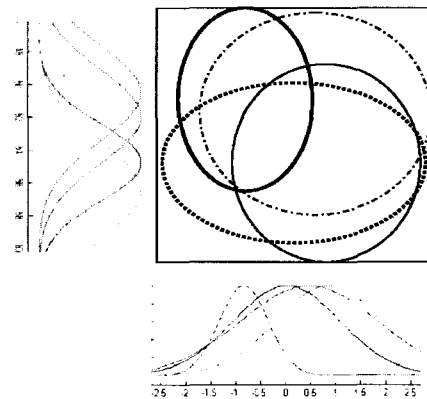


그림 5. 소속함수에 의한 규칙공간  
Fig 5. Rule space with membership function

일반적인 ANFIS의 경우 구조 동정과 파라미터 동정으로 나누어져 시스템을 구성하는데 전체부 소속함수의 수와 파라미터에 의하여 국부적으로 수립하는 최적 시스템이 만들어진다. 제안된 방법에서는 초기치에 영향을 받는 전체부 소속함수의 수와 초기 파라미터를 계층적 클러스터링과 GMM을 이용하여 추정한 뒤, 파라미터에서 중심의 값을 Gaussian 소속함수의 중심으로 하고 Yager의 방법을 이용하여 분산을 추정하여[11], 이를 ANFIS의 입력으로 다음 그림 6과 같이 주어 전체적인 시스템의 성능을 향상시키고자 하였다.

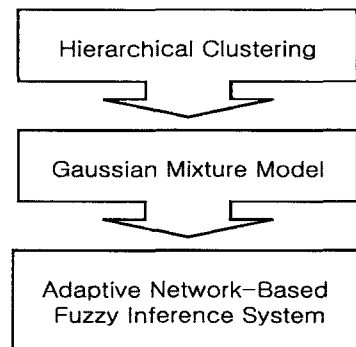


그림 6. 순서도  
Fig. 6 Flowchart

일반적인 ANFIS의 경우, 그리드 분할 형태의 규칙 구성을 가진다. 그리드 분할 형태의 규칙 생성 방법은 소속함수의 수가 증가하거나 입력 차원의 증가에 대하여 규칙이 지수 함수적으로 증가하는 문제가 발생하여 모델의 학습 속도에 영향을 주거나 실제 시스템의 구성에서 필요 이상의 많은 부분을 차지하여 시스템의 크기를 증가시킨다. 제안된 방법에서는 계층적 클러스터링을 통하여 소속함수의 수나 입력 차원의 증가와 무관한 클러스터링에 의한 규칙 생성을 실시하였고 이를 GMM을 통하여 최적화된 파라미터로 추정하여 ANFIS의 초기 파라미터로 사용하였다.

#### 4. 시뮬레이션 및 결과

제안된 시스템의 유용성을 보이기 위하여 두 가지 예제에 대하여 시뮬레이션을 적용하였다. 먼저 전형적인 비선형 모델인 Box-Jenkins의 가스로 데이터에 적용하여 성능을 비교하였으며 두 번째로 Sugeno의 비선형 시스템[12]에 대하여 적용하여 제안된 시스템의 성능을 비교하였다.

##### 1) Box-Jenkins 가스로 모델

이 모델은 입력이 10개이며 하나의 출력을 가지는 총 290개의 데이터 쌍을 가지는 시계열 예측 모델로서 제안된 시스템은 Jang의 입력 선택 방법에 의하여 두 개의 입력  $y(t-1)$ 와  $u(t-4)$ 을 선택하여 모델을 구성하였다. 먼저 계층적 클러스터링을 통하여 생성된 파라미터를 GMM의 초기치로 주어 파라미터를 최적화 한다. GMM에서 발생하는 초기 파라미터에 의한 국부적 수렴 문제에 대하여 주어진 데이터에 근거한 계층적 클러스터링에서의 파라미터를 이용함으로써 해결하고자 하였다. 시뮬레이션을 통하여 결과를 보면 다음과 같다.

먼저 전체 데이터를 홀수열과 짝수열로 균등하게 학습 데이터와 검증 데이터로 나누며 학습데이터를 이용하여 시스템을 생성한 뒤 이를 검증 데이터를 이용하여 시스템을 평가하였다. 계층적 클러스터링을 이용하여 Dendrogram을 생성한 후 실행 착오 방법에 의하여 선택된 4개의 클러스터를 선택하여 각 클러스터 중심을 GMM의 초기 파라미터로 설정한다. GMM 알고리즘은 최대 반복횟수 100회나 개선오차  $10^{-6}$  이하 일 때 알고리즘이 종료하도록 하였다.

그림 7에서 학습이 진행되는 동안 학습 데이터에 의한 오차와 검증 데이터에 대한 오차를 표시하였다. 학습 데이터에 대한 오차와 검증 데이터의 오차는 표 1에서와 같이 기존의 연구 결과와 달리 일정한 범위 안에서 동시에 감소하는 것을 볼 수 있다.

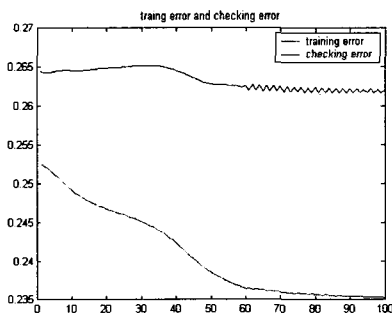
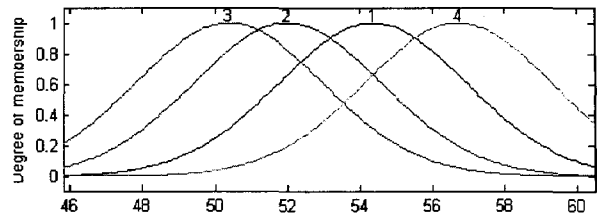
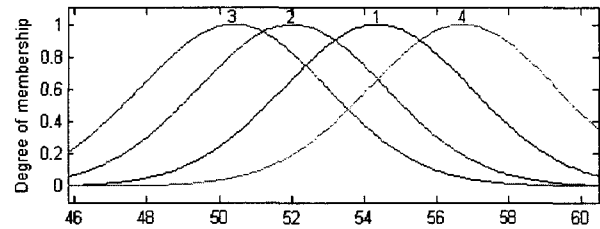


그림 7. 학습 오차와 검증 오차  
Fig. 7 Training Error and Checking Error



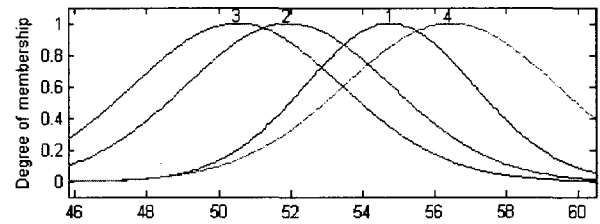
(a) Initial membership function of  $y(t-1)$



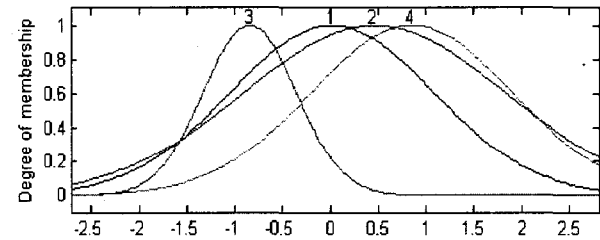
(b) Initial membership function of  $u(t-4)$

그림 8. 초기 소속함수

Fig. 8 Initial Membership Function



(a) Final membership function of  $y(t-1)$



(b) Final membership function of  $u(t-4)$

그림 9. 최종 소속함수

Fig. 9 Final Membership Function

그림 7에서는 계층적 알고리즘과 GMM을 이용하여 추정된 파라미터로 만들어진 소속함수를 나타내었으며 그림 8에서 ANFIS의 학습 후에 생성된 소속함수를 나타내었다. 또한 평균자승오차(Mean Square Error)를 이전 연구 결과와 비교하여 표로 나타내면 다음과 같다.

	입력의 수	규칙의 수	학습오차	검증오차
Jang[4]	2	4	0.135	0.530
Oh[13]	2	4	0.020	0.272
	2	6	0.020	0.264
Kim[14]	2	6	0.179	0.188
제안된 방법	2	4	<b>0.055</b>	<b>0.068</b>

표 1. 모델링 결과  
Table. 1 Modeling results

표 1에서 볼 수 있듯이 제안된 방법에 의하여 시뮬레이션 결과가 이전의 연구 결과 보다 더 적은 오차를 가지는 것을 볼 수 있으며, 학습 데이터에 의한 오차 뿐만 아니라 검증 데이터에 의한 오차도 학습 데이터에 의한 오차와 일정한 비율로 동시에 감소하는 것을 볼 수 있다. 학습 데이터에 의한 뉴로-퍼지 시스템의 일방적인 성능개선이 검증 데이터에 의한 검증 오차와 차이가 심하다면 과학습(overfitting)에 의한 모델로서 좋은 시스템이라 할 수 없는데, 제안된 방법에 의한 모델은 학습 및 검증 데이터에 의한 오차가 일정한 범위 안에서 동시에 감소함을 볼 수 있다.

또한 각 단계별 추정된 클러스터의 중심은 다음 표 2와 같다.

표 2. 단계별 파라미터  
Table 2. Parameters with each step

	계층적 클러스터링		GMM	
	입력 $y(t-1)$	입력 $u(t-4)$	입력 $y(t-1)$	입력 $u(t-4)$
소속함수 1	55.12	-0.36	54.40	-0.29
소속함수 2	51.00	0.80	52.01	0.43
소속함수 3	47.45	-1.71	50.40	-1.04
소속함수 4	58.91	1.91	56.73	0.97

여기서 학습된 ANFIS의 파라미터는 다음과 같다. 입력은 2개이며 1개의 출력을 가지고 각각의 입력은 4개의 소속함수를 가지며 결론부 1차 선형 방정식은 4개로 구성되어 있다. 또한 각각의 소속함수는 Gaussian 소속함수로 구성되어 있다. 입력 파라미터는 분산과 중심으로 다음과 같이 표 3에 표시되어 있다.

표 3. 전제부 파라미터  
Table 3 Parameters of premise part

	입력 $y(t-1)$	입력 $u(t-4)$
소속함수 1	2.351, 54.76	1.063, 0.01738
소속함수 2	2.856, 51.94	1.363, 0.4624
소속함수 3	2.91, 50.51	0.4876, -0.8424
소속함수 4	2.963, 56.45	1.05, 0.8503

여기서 각 파라미터의 첫 번째는 분산을, 두 번째는 중심을 나타내었다.

출력 (결론부) 파라미터는 1차 선형 방정식(15)의 계수는 다음과 같다.

표 4. 결론부 파라미터  
Table 4 Parameters of consequent part

	p	q	r
규칙 1	-1.256	-3.527	63.6
규칙 2	-0.8189	-0.3819	40.84
규칙 3	0.6314	3.932	-30.54
규칙 4	-1.268	0.7793	79

## 2) Sugeno의 비선형 시스템

Sugeno의 비선형 시스템은 다음과 같은 두 개의 입력과 하나의 출력을 가진다[12].

$$y = (1 + x_1^{-2} + x_2^{-1.5})^2, 1 \leq x_1, x_2 \leq 5$$

이를 다음과 같이 그림 10으로 표현할 수 있다.

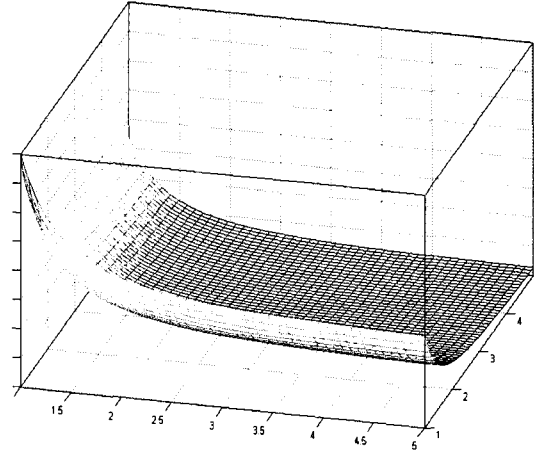


그림 10. 입력력 관계  
Fig. 10 Input-Output Relation

입출력 데이터는 각 50쌍으로 구성되어 있으며 제안된 방법에서는 이전 논문에서의 부가(dummy) 데이터를 검증 데이터로 이용하였다. 클러스터의 수를 7개로 하여 제안된 방법을 통하여 초기 파라미터를 구한 후 ANFIS를 이용하여 150회 학습을 실시하였다.

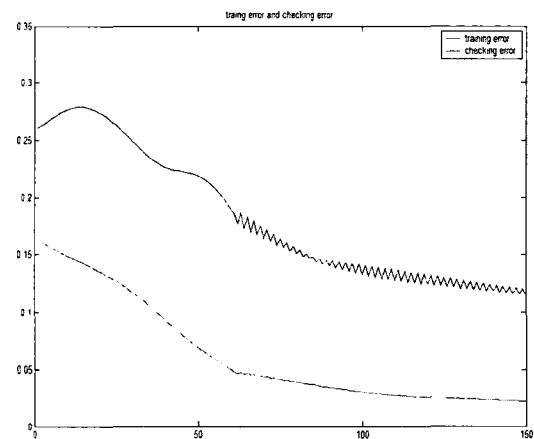


그림 11. 학습 오차와 검증 오차  
Fig. 11 Training Error and Checking Error

그림 11에서 보는 것과 같이 학습 데이터에 의한 오차 및 검증 데이터에 의한 오차가 일정한 비율로 같이 감소하는 것을 볼 수 있다. 이를 Sugeno의 결과와 비교할 때 성능지수 MSE 오차가 0.010에서 0.00481로 제안된 방법의 성능이 우수하였으며 검증데이터 역시 0.0133으로 더 나은 성능을 보였다. 각 단계별 클러스터의 중심은 다음 표 5과 같다.

표 5. 단계별 파라미터  
Table 5. Parameters with each step

	계층적 클러스터링		GMM		ANFIS	
	입력 $x_1$	입력 $x_2$	입력 $x_1$	입력 $x_2$	입력 $x_1$	입력 $x_2$
소속함수 1	1.28	1.78	1.28	1.95	1.38	2.44
소속함수 2	4.22	4.81	4.38	4.81	4.39	4.73
소속함수 3	1.89	4.33	1.83	4.33	2.02	4.22
소속함수 4	3.66	1.37	3.56	1.38	2.97	1.45
소속함수 5	4.63	2.65	4.84	2.54	4.61	2.14
소속함수 6	3.26	3.78	2.83	3.78	2.62	3.61
소속함수 7	2.70	3.18	2.38	2.72	2.12	2.21

### 5. 결론

일반적인 뉴로-퍼지 시스템은 입력 차원과 소속함수의 증가에 대하여 규칙이 증가되는 문제점과 뉴로-퍼지 시스템의 초기 파라미터 최적화를 통하여 학습속도 및 성능 개선을 논문에서는 클러스터링 방법을 통하여 해결하고자 하였다.

또한 목적함수를 이용하지 않고 단지 각 데이터들의 유클리디언 거리만을 이용하는 계층적 클러스터링을 이용하여 GMM의 초기 파라미터(중심)를 결정하므로써 매 시도마다 추정된 클러스터가 일정하게 되는 특징을 가지고 있다. 이는 목적함수에 의하여 초기 파라미터가 변동하여 여러 번 실험을 통하여 가장 좋은 파라미터를 찾는 기존의 방법과는 달리 한번의 실험으로 시스템의 성능을 평가할 수 있는 장점이 있다.

시뮬레이션 결과에서 볼 수 있듯이 좋은 성능에도 불구하고 초기 파라미터에 의하여 결과가 달라지는 GMM의 문제를 계층적 클러스터링을 이용하여 해결하고자 하였다. 결과에서 볼 수 있듯이 제안된 순차적 방법에 의한 뉴로-퍼지 시스템의 결과가 기존의 연구 결과보다 좋은 성능을 보임을 알 수 있다.

추후 지속적인 연구과제로는 계층적 클러스터링에서의 적합한 계층 결정 문제와 매우 큰 데이터 집합의 대한 처리속도 저하 문제, 뉴로-퍼지 시스템의 전제부 및 결론부 파라미터 최적화의 다양한 방법 등이 연구되어야 할 것이다.

### 참 고 문 헌

[1] Chin-Teng Lin, C. S. George. Lee, "Neural Fuzzy Systems : A Neuro-Fuzzy Synergism to Intelligent Systems", Prentice Hall, 1996.  
 [2] J-S. R. Jang, C. T. Sun, E. Mizutani, "Neuro-Fuzzy and Soft Computing : A Computational Approach to Learning and Machine Intelligence", Prentice Hall, 1997.  
 [3] Timothy J. Ross, "Fuzzy Logic with Engineering Applications", McGraw-Hill, Inc, 1995.  
 [4] J-S. R. Jang, Input Selection for ANFIS Learning, Proceedings of the Fifth IEEE International

Conference on, Fuzzy Systems, Vol. 2, pp. 1493-1499, 1996.  
 [5] Jeon-Shing Wang, C.S. George Lee, "Self Adaptive Neuro-Fuzzy Inference Systems for Classification Applications", IEEE Trans on Fuzzy Systems, Vol. 10, No. 6, pp790-802, 2003  
 [6] J. Zhang, S. Koper, A. Knoll, "Extracting compact fuzzy rules based on adaptive data approximation using B-splines", Information Sciences, Volume 142, Issues 1-4, pp 227-248, 2002  
 [7] Arnold F. Shapiro, "The merging of neural network, fuzzy logic, and genetic algorithms", : Mathematics and Economics, Vol. 31, Issue 1, pp. 115-131, 2002  
 [8] Arun D. Kulkarni, "Computer Vision and Fuzzy-Neural Systems", Prentice Hall, 2001.  
 [9] Ethem Alpayd, "Soft vector quantization and the EM Algorithm", neural Network, Vol. 11, Issue. 3, pp. 467- 477 1998.  
 [10] Todd. K. Moon, The Expectation-Maximization Algorithm, IEEE Signal Processing Magazine, 1996.  
 [11] R. R. Yager, D. P. Filev, Generation of Fuzzy Rules by Mountain Clustering, Journal of Intelligent and Fuzzy System, Vol.2, pp. 209-219, 1994.  
 [12] Michio Sugeno, Takahiro Yasukawa, "A Fuzzy-Logic-Based Approach to Qualitative Modeling", IEEE Trans on Fuzzy Systems, Vol. 1, No. 1, pp. 7-31, 1993.  
 [13] S. K. Oh, Witold Pedrycz, "Identification of fuzzy systems by means of an auto-tuning algorithm and its application to nonlinear systems", Fuzzy Sets and Systems, Vol. 115, pp. 205-230, 2000.  
 [14] S.S. Kim, K. C. Kwak, S. S. Kim, J. W. Ryu, M. G. Chun, "A Novel Neuro Fuzzy Modeling using the Gaussian Mixture Models", ICCAS2002, 2002.

### 저 자 소 개



김승석  
 1998. 2 : 충주대학교 전기공학과(공학사)  
 2002. 2 : 충북대학교 전기공학과(공학석사)  
 2002. 3~ 현재 : 충북대학교 전기공학과 박사과정

주관심분야 : 퍼지시스템, 신경회로망, 지능 제어

E-mail : powerkim@vod.chungbuk.ac.kr



**곽근창**

1996. 2 : 충북대학교 전기공학과(공학사)  
1998. 2 : 충북대학교 전기공학과(공학석사)  
2002. 2 : 충북대학교 전기공학과(공학박사)  
2002. 3~현재 충북대학교 BK사업단  
박사후연구원  
2003.3~현재 Dept. of Electrical and  
Computer Engineering , University of  
Alberta, Postdoc.

주관심분야 : 계산지능, 패턴분류, 생체인식  
E-mail : kkc@vod.chungbuk.ac.kr



**전명근**

1987. 2 : 부산대학교 전자공학과(공학사)  
1989. 2 : KAIST 전기 및 전자공학과  
공학석사)  
1993. 2 : KAIST 전기 및 전자공학과  
(공학박사)  
1993~1996 : 삼성전자 자동화연구소  
선임연구원

2000~2001 : University of Alberta 방문교수  
1996~현재 : 충북대학교 제어계측공학과 부교수

주관심분야 : 생체인식, 감정인식, 지능제어  
E-mail : mgchun@cbucc.chungbuk.ac.kr



**유정웅**

1965. 2 : 한양대학교 전기공학과(공학사)  
1976. 2 : 단국대학교 전자공학과(공학석사)  
1987. 2 : 단국대학교 전자공학과(공학박사)  
1979. 3~현재 : 충북대학교 전기공학과 교수

주관심분야 : 가변구조제어, 지능제어, QFT  
E-mail : jwryu@vod.chungbuk.ac.kr