

새로운 클러스터링 알고리즘을 적용한 향상된 뉴로-퍼지 모델링

論 文

53D-7-10

Advance Neuro-Fuzzy Modeling Using a New Clustering Algorithm

金 承 石* · 金 聖 洙** · 柳 正 雄***

(Sung-Suk Kim, Sung-Soo Kim, Jeong-Woong Ryu)

Abstract - In this paper, we proposed a new method of modeling a neuro-fuzzy system using a hybrid clustering algorithm. The initial parameters and the number of clusters of the proposed system are optimally chosen simultaneously with respect to the process of regression, which is a unique characteristics of the proposed system. The proposed algorithm presented in this work improves the overall performance of the proposed a neuro-fuzzy system by choosing a proper number of clusters adaptively according the characteristics of given data. The process of clustering is performed by deciding on the number of classes, which yields the property of convergence of the system. In experiments, the superiority of the proposed neuro-fuzzy system is demonstrated, especially the process of optimizing parameters and clustering of learning speed.

Key Words : Gaussian Mixture Model, Chen Clustering, TSK Fuzzy model, Neuro-Fuzzy System

1. 서 론

시스템 모델링 방법으로는 선형 시스템을 기반으로 하는 미분방정식 등과 같은 수학적 지식을 기반으로 하는 다양한 방법들이 일반적으로 연구되고 있다[1][2]. 모델링에서의 많은 문제점은 실제의 시스템이 비선형성이 강하거나, 실제의 시스템 자체가 모델링이 어렵거나 불가능한 경우이다[2]. 이러한 여러 문제점들을 해결하기 위하여 최근에 인간의 사고나 의사판단 등을 모사하여 모델링을 구성하는 신경회로망 등이 많은 연구자들의 관심이 되었다. 이러한 인공지능적인 방법은 수학적 제약에서 자유로우며, 실제 시스템의 입출력 데이터 등을 기반으로 모델링을 하므로 다양한 분야로의 적용이 가능한 특성을 지니고 있다. 이러한 특성은 패턴인식, 영상처리, 자동제어 등 다양한 분야에 적용되어 왔다 [1-3].

새로운 알고리즘의 제안이나 기존 알고리즘의 성능개선에 대한 연구는 현재 신경회로망에서 활발하게 진행되고 있다 [4][5]. 이와 더불어 각 알고리즘이 가지는 장점을 유지하면서도 기존의 알고리즘들의 문제점을 개선하거나 또 다른 알고리즘들을 통합하여 전체 성능을 개선하고자 하는 시도 역시 다양하게 진행되어 왔다. 특히, 통합된 모델은 특정 알고리즘이 가지는 제약이나 문제점을 해소하면서도 원하는 성능 또는 모델을 설계할 수 있다는 장점을 지니고 있다[2].

조건부적인 Fuzzy C-means (Conditional Fuzzy C-means : CFCM) 등은 한 알고리즘 내에 같은 알고리즘을 추가하여 비교사 학습(unsupervised learning)에 교사 학습(supervised learning)기능을 추가하여 퍼지 분할 성능의 개선을 시도하였고, 다항식 신경망 (Polynomial Neural Network : PNN) 등과 같은 기법은 퍼지 논리와 신경회로망, 유전 알고리즘을 이용하여 최적의 모델링을 시도하였다[6][7]. 또한 뉴로-퍼지 모델링은 신경회로망의 학습기능(역전파 알고리즘)과 퍼지 시스템의 의사 결정 문제를 통합하여 양방향 학습을 가능하게 하였다[2].

일반적인 모델링에서 최종적으로 얻어진 모델의 성능은 파라미터의 초기치에 대하여 민감한 국부적 최적에 수렴하는 특성을 가진다. 입출력 데이터를 기반으로 모델을 최적화하는 경우 이를 표현할 수 있는 초기 파라미터의 결정 역시 데이터에 의존하게 된다[1-3]. 여기서, 여러 알고리즘들의 통합은 모델의 성능에 대하여 좋은 결과를 보이는 반면 전체 구조가 복잡해지는 경우가 발생할 수 있고, 이러한 복잡한 구조는 모델의 학습과 구현에 문제를 가질 수 있다. 알고리즘들의 통합과 더불어 모델의 복잡성 해소는 성능과 더불어 모델의 이해와 실제 구현에 있어 또 하나의 문제로 제기되고 있다.

본 논문에서는 초기치 문제의 해결 방안으로서, 기존의 Gaussian Mixture Model (GMM) [8-11]을 기본 개념으로 자율적인 파라미터 구조를 결정하는 새로운 클러스터링 알고리즘을 제안하였다. 일반적으로 사전에 주어진 조건에 의해서 자율적인 클러스터 구조를 결정하며 최적화하는데[12], 제안된 알고리즘은 초기 클러스터 파라미터 문제의 해결 방안을 제안하고, 제안된 알고리즘이 진행되는 동안 일정한 구조(클러스터의 수)로 시스템이 수렴함을 보였다. 본 연구의 장점은 시스템이 Gaussian 함수가 가지는 장점을 그대로 활

* 正 會 員 : 忠 北 大 전전컴 공학부 電氣工學 博士課程

** 正 會 員 : 忠 北 大 전전컴 공학부 電氣工學 副教授 · 工博

*** 正 會 員 : 忠 北 大 전전컴 공학부 電氣工學 教授 · 工博

接受日字 : 2004年 3月 15日

最終完了 : 2004年 5月 20日

용하면서도 초기치 결정 문제를 해결하였고, 최적화된 클러스터 파라미터를 뉴로-퍼지 모델의 입력으로 직접적으로 이용함으로써 그리드 분할에 의한 뉴로-퍼지 모델의 전체부에서 발생하는 퍼지 규칙 증가 문제를 해결할 수 있게 하였다.

본 논문의 구성은 2장에서 GMM의 기본 개념과 제안된 클러스터링의 구성을 설명하고 이를 적용할 뉴로-퍼지 시스템을 설명하였으며, 3장에서는 시뮬레이션을 통하여 제안된 방법의 유용성을 보였으며 4장에서 정리하였다.

2. 제안된 모델

제안된 클러스터링 모델은 초기 알고리즘의 파라미터가 클러스터의 수가 아닌 임계값 ξ 로 주어지며 알고리즘이 진행되는 동안 특정한 클러스터의 수로 수렴하는 특성을 Gaussian Mixture Model (GMM) 개념에 적용하였다. 이 경우 GMM에서 발생하는 초기 파라미터 결정문제에 대하여 전체 학습데이터를 초기 클러스터 파라미터로 주어 일반적인 클러스터링 알고리즘에서 발생하는 초기 파라미터에 의한 최종 파라미터의 변화 문제를 해결하였다.

2.1 Gaussian Mixture Model (GMM)

클러스터링 파라미터 추정에서 GMM은 통계적 데이터 추정 방식으로 다양한 분야에서 좋은 성능을 보여 왔다. Expectation-Maximization (EM) 알고리즘을 기반으로 하는 Gaussian Mixture Model은 확률분포함수의 추정을 통하여 클러스터 파라미터 μ, Σ 를 최적화하는 알고리즘이다[9][10]. 먼저 데이터가 독립, 동형 분포 (independent and identically distributed : iid)라고 가정할 때 확률 분식 (1)과 같이 표현할 수 있다.

$$p(X|\theta) = \prod_{i=1}^N p(x_i|\theta) = L(\theta|X) \quad (1)$$

여기서 $L(\theta|X)$ 를 주어진 데이터에서 파라미터의 likelihood 함수라 한다. Maximum likelihood (ML) 문제에서는 함수 L 을 최대화하는 파라미터 θ^* 를 추정하는 것으로 식 (2)와 같다[8].

$$\theta^* = \arg \max_{\theta} L(\theta|X) \quad (2)$$

EM 알고리즘은 주어진 데이터가 불완전할 때 이들 분포를 이용하여 파라미터의 ML 추정을 찾는 일반적인 방법이다. 즉 클러스터링 문제에서 잡음 등으로 데이터가 잘못된 값을 가지거나, likelihood 함수의 최적화가 해석적으로 쉽지 않으나 이 함수가 추가된 잡음 또는 은닉한 파라미터에 대한 값이 존재한다는 가정에 의하여 간략화시킬 때 이용한다.

주어진 데이터 집합 X 를 관측된 데이터라고하고 이를 불완전하다고 가정한다. 또한 완전한 데이터 집합 $Z=(X, Y)$ 를 가정하면 다음과 같은 결합 밀도함수 고려할 수 있다[9].

$$p(z|\theta) = p(x, y|\theta) = p(y|x, \theta)p(x|\theta) \quad (3)$$

여기서 Y 는 주어진 분포에서 사전에 더해진 임의의 알 수 없는 정보로 고려한다.

이때 위의 식 (3)을 만족하는 새로운 likelihood 함수를 다음과 같이 정의할 수 있다.

$$L(\theta|Z) = L(\theta|X, Y) = p(X, Y|\theta) \quad (4)$$

이 함수는 잘못된 정보를 가지는 Y 를 포함한 완전한 데이터 likelihood이다. 또한 데이터 X 와 파라미터 θ 가 고정되어 있으며 Y 가 랜덤 변수로 가정한다. 즉 원래의 likelihood $L(\theta|X)$ 는 불완전 데이터 likelihood이다.

EM 알고리즘의 Expectation 단계는 미지의 Y 를 고려한 데이터 X 와 현재 파라미터 θ 를 이용하여 완전한 데이터 log likelihood를 다음과 같이 계산하는 것이다[9][10].

$$Q(\theta, \theta^{s-1}) = E[\log p(X, Y|\theta)|X, \theta^{s-1}] \quad (5)$$

여기서 θ^{s-1} 은 기대값을 평가하는 현재의 파라미터이며 θ 는 Q 를 증가하여 최적화 시키려고 하는 새로운 파라미터이다. 즉 θ 는 likelihood를 최대화 시켜 최종적으로 최적화 되는 파라미터이다.

또한 EM 알고리즘의 Maximization 단계는 Expectation 단계에서 추정된 기대값 중 최대값을 찾는다[9].

$$\theta^s = \arg \max_{\theta} Q(\theta, \theta^{s-1}) \quad (6)$$

이 두 개의 과정은 종결조건을 만족할 때까지 반복한다. 각 반복은 log likelihood가 증가되는 것이 보장되는 방향으로 국부적 최적해로 수렴한다.

$Q(\theta|\theta^{s-1})$ 을 최대화 시키는 것을 대신하여 변형된 Maximization 단계에서 $Q(\theta^s, \theta^{s-1}) > Q(\theta, \theta^{s-1})$ 으로 조건을 바꿀 때 이를 일반화된 EM 알고리즘이라 하며 이러한 방법 역시 수렴이 보장된다.

Gaussian Mixture Model 에서의 EM 알고리즘

혼합 밀도 파라미터 추정문제는 EM 알고리즘에서 가장 일반적으로 이용되고 있으며 다음과 같은 확률 모델을 가정한다.

$$p(x|\theta) = \sum_{i=1}^M \alpha_i p_i(x|\theta_i) \quad (7)$$

여기서 파라미터 $\theta=(\alpha_1, \alpha_2, \dots, \alpha_M, \theta_1, \theta_2, \dots, \theta_M)$ 는 $\sum_{i=1}^M \alpha_i=1$ 이고 확률함수 p_i 는 θ_i 에 의하여 추정되는 확률밀도함수이다. 즉 M 개의 혼합된 계수 α_i 로 구성된 M 개의 구성 밀도

로 가정할 수 있다. 이때 확률 함수 $p_i(x|\theta_i)$ 를 Gaussian 확률로 나타내면 다음과 같다.

$$p(x|\mu_i, \Sigma_i) = \frac{1}{(2\pi)^{d/2} |\Sigma_i|^{1/2}} \times \exp\left(-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1}(x-\mu_i)\right) \quad (8)$$

이를 Bayes's의 규칙과 Lagrange 곱수를 이용하여 각 파라미터를 최적화 하면, 반복적인 알고리즘의 수행 과정에서, 각 반복(epoch) g 에서 파라미터 가중치 α_i 과 중심 μ_i , 분산 Σ_i 은 식(9)에서 식(11)와 같이 추정되어진다[9][10].

$$\alpha_i^{new} = \frac{1}{N} \sum_{i=1}^N p(l|x_i, \theta^g) \quad (9)$$

$$\mu_i^{new} = \frac{\sum_{i=1}^N x_i p(l|x_i, \theta^g)}{\sum_{i=1}^N p(l|x_i, \theta^g)} \quad (10)$$

$$\Sigma_i^{new} = \frac{\sum_{i=1}^N p(l|x_i, \theta^g) N_{l,i}}{\sum_{i=1}^N p(l|x_i, \theta^g)} \quad (11)$$

$$= \frac{\sum_{i=1}^N p(l|x_i, \theta^g) (x_i - \mu_i)(x_i - \mu_i)^T}{\sum_{i=1}^N p(l|x_i, \theta^g)}$$

여기서 위첨자 g 는 각 반복의 횟수이며 알고리즘이 진행되는 동안 θ^g 는 ML에서 log-likelihood를 최대화 시키는 방향으로 최적화된다.

임의로 주어진 데이터에서 GMM 추정을 통한 확률 분포를 그림 1에 표현하였다. GMM이 우수한 성능을 가지는 반면 몇 가지 문제점을 가진다[4][13]. 먼저 파라미터 최적화를 하기 위하여 파라미터의 크기 또는 구조를 미리 지정하여야 한다. 사전에 정해진 파라미터의 구조에서 최적화 과정을 가지므로 미지의 데이터나 사람의 주관적 지식으로 구성된 파라미터의 구조가 모델의 최적화에 도움을 주지 못하는 경우가 발생할 수 있다. 두 번째로 초기 파라미터 결정 문제이다. GMM은 초기 파라미터에 의하여 국부적 최적해로 수렴하는 알고리즘이다. 초기 파라미터를 임의로 생성하는 일반적인 방법에서는 여러 번의 시도를 통하여 최적의 값을 찾는 시행 착오법을 수행하여야 한다.

이러한 문제점들을 해결하기 위하여 제안된 방법에서는 파라미터의 수와 초기치 결정을 자율적으로 결정하는 알고리즘을 제안하였다. Gaussian 함수에서 분산 Σ_i 는 소속된 중심 μ_i 의 분포를 표현한다. 동일한 중심에서 서로 다른 분산에 대한 표현을 그림 2에 나타내었다.

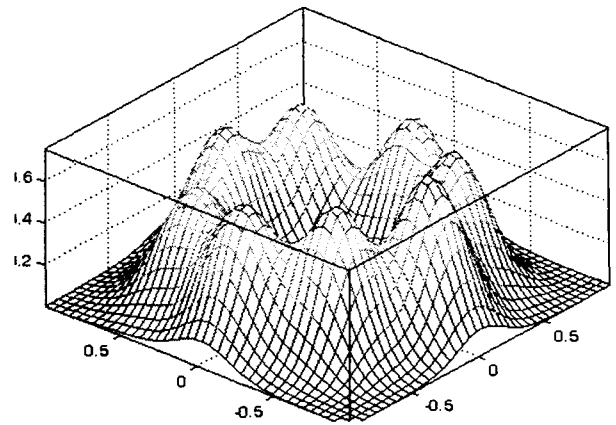


그림 1 Gaussian Mixture Model의 예
Fig. 1 The Sample of Gaussian Mixture Model

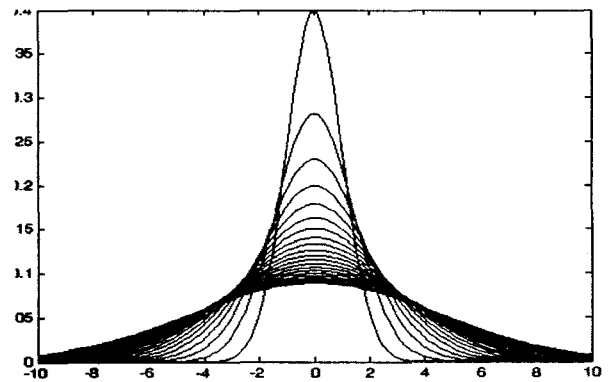


그림 2 분산에 의한 Gaussian 함수의 특성
Fig. 2 Property of Gaussian function with variance

그림 2에서 볼 수 있듯이 각기 다른 분산은 Gaussian 함수에서 동일한 중심에 대한 다양한 분산에 의한 결과를 나타내었다. 클러스터의 중심과 비교하여 데이터는 분산에 의하여 멀어질수록 영향이 크게 변화하며 특히 상대적으로 중심과 거리가 멀리 떨어진 데이터들과의 확률 평가에 크게 영향을 준다. 즉 클러스터의 수가 다르거나 초기치가 다를 경우 반복적인 알고리즘에 의한 중심 평가의 경우 최종적으로 전혀 다른 결과를 추정할 수 있다.

2.2 제안된 클러스터링 알고리즘을 적용한 향상된 뉴로-퍼지 모델링

제안된 방법의 기본 개념은 상대적으로 유사성이 높은 중심 근처의 데이터 x_i 가 중심 μ_k 평가에 더욱 영향을 주며 그렇지 않은 데이터 x_j 는 현재의 중심 μ_k 의 평가에 영향을 되도록 미치지 못하도록 하는 것이다. 추정된 분산은 사전에 정해진 일정한 제한을 가하며 이를 이용하여 전체 확률을 추정한다.

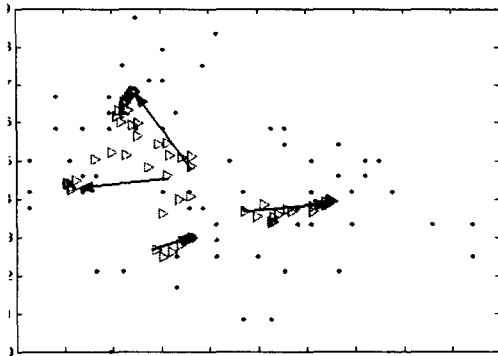


그림 3 파라미터들의 수렴
Fig. 3 Convergence of parameters

제안된 방법에서는 초기 클러스터를 전체 데이터와 동일하게 설정을 한다. 즉 전체 데이터가 초기 파라미터의 크기와 값으로 결정이 되며 반복적인 연산 과정을 통하여 특정한 크기의 파라미터로 수렴한다. 그림 3에서 알 수 있듯이 초기 파라미터는 Gaussian 확률 분포에 의하여 특정한 파라미터로 수렴하며 수렴 과정동안 같은 값을 가지는 클러스터는 통합되어 전체 연산시간을 감소시킬 수 있다[12].

단계 1 : 각 파라미터를 초기화 시킨다. 먼저 공분산행렬의 제한 값을 결정한 후, 전체 학습 데이터를 초기 중심으로 설정하며 이를 이용하여 각 중심의 공분산 행렬을 추정한다. 초기화 되는 파라미터는 $\zeta = \text{predetermined}$, $\mu^{\text{Initial}} = \text{training parameter}$, $\mu_{\text{ref}} = \mu^{\text{Initial}}$, $\Sigma_i = \frac{1}{N} (x - \mu_i)(x - \mu_i)^T$ 이다.

단계 2 : 각 데이터와 혼합 요소와의 Gaussian 확률을 계산한다.

$$p(x|\mu_i, \Sigma_i) = \exp\left(-\frac{1}{2}(\mu_{\text{ref}} - \mu_i)^T (\Sigma_i \times \zeta)^{-1} (\mu_{\text{ref}} - \mu_i)\right) \quad (12)$$

$$p_{li} = \frac{p(x_i|\mu_i, \Sigma_i)}{\sum_j p(x_i|\mu_j, \Sigma_j)} \quad (13)$$

단계 3 : 각 혼합 요소의 파라미터를 추정한다.

$$\mu_i^{\text{new}} = \frac{\sum_{l=1}^N x_l p_{li}}{\sum_{l=1}^N p_{li}} \quad (14)$$

$$\Sigma_i^{\text{new}} = \frac{1}{N} \sum_{l=1}^N (x_l - \mu_i^{\text{new}})(x_l - \mu_i^{\text{new}})^T \quad (15)$$

단계 4 : 종결 조건을 검사, 조건에 만족할 경우 알고리즘을 종결하고 그렇지 않으면 단계 2로 돌아가 반복수행을 한다.

본 논문에서 이용되는 뉴로-퍼지 시스템은 적응 네트워크 기반 퍼지 추론 시스템(Adaptive Network-Based Fuzzy Inference System : ANFIS)이다[2][3]. 이 모델의 특징은 언어적 형태를 가지는 전제부와 1차 선형 방정식 형태를 가지는 결론부로 구성되어진 Takagi - Sugeno - Kang (TSK) 퍼지 모델로 그림 4와 같이 되어 있다. 이 모델의 장점은 결론부의 구성이 1차 선형 방정식 형태로 구성되어 있어 비퍼지화 과정이 생략된다. 또한 퍼지 규칙의 수가 선형 방정식의 수로 되며 전체적인 구성이 간략화 되어진다. 일반적인 뉴로-퍼지 모델의 규칙 생성은 전체 입력 데이터 공간을 격자 형태로 분할하는 그리드 (grid) 분할이다. 이 경우 각 데이터 공간에서 소속함수의 수가 증가하거나 입력 공간(차원)의 증가에 따라 지수함수 형태와 같이 증가하는 문제점을 가지고 있다. 이러한 문제점은 규칙의 증가에 따라 모델의 크기 또한 커지며 또한 학습 시간의 증가와 실제 시스템 구현에도 어려움을 가진다. 그림 5에서 볼 수 있듯이 그리드 분할의 경우 입력차원의 차승 × 소속함수의 수에 비례하여 퍼지 규칙의 수가 증가하는 반면 클러스터링에 의한 scatter 분할은 퍼지 규칙의 수가 단지 소속함수의 수에 비례하므로 입력의 조건이 복잡한 경우에도 시스템의 구조를 간단하게 할 수 있는 장점이 있다.

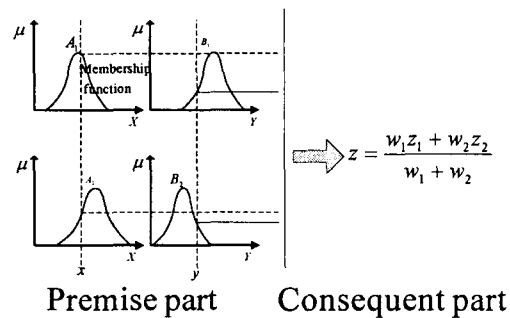


그림 4 TSK 퍼지 모델

Fig. 4 TSK fuzzy model

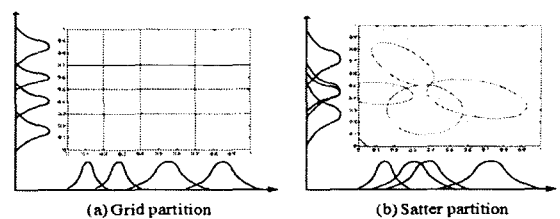


그림 5 그리드 분할과 Scatter 분할 비교

Fig. 5 Compare with grid partition and scatter partition

클러스터링의 기본 개념은 데이터 공간상에서 데이터가 밀집한 영역과 그렇지 않은 영역에 대하여 서로 다른 분할을 통하여 그림과 같이 적절한 퍼지 규칙을 생성하는 것이다. 따라서 그리드 분할과 같은 규칙 증가에 대한 문제점을 해결할 수 있다. 그림 5과 6에서와 같이 전체 데이터 공간을 이용하는 것이 아니라 필요한 공간에 대한 소속도만을 부여함으로써 불필요한 규칙의 증가를 억제하면서 클러스터링을 효과적으로 표현할 수 있다.

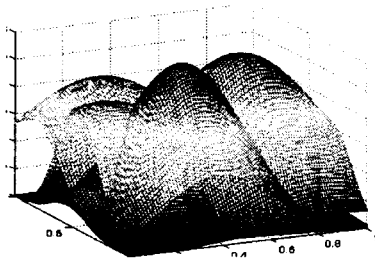


그림 6 클러스터링에 의한 소속도
Fig. 6 Membership degree by clustering

데이터 클러스터링은 뉴로-퍼지 모델의 초기 구조 및 파라미터 동정(Identification)을 실시한다. 클러스터링 파라미터는 뉴로-퍼지 시스템의 전제부 소속함수의 수 및 파라미터, 규칙을 결정한다. 뉴로-퍼지 모델은 이를 기반으로 학습을 실시함으로써 좀 더 빠른 학습 속도와 성능을 보장한다. 이러한 TSK 퍼지 모델의 구조를 보면 다음과 같다.

$$\begin{aligned}
 R^1: & \text{ IF } x \text{ is } A_1 \text{ and } y \text{ is } B_1 \\
 & \text{ THEN } f_1 = p_1x + q_1y + r_1 \\
 R^2: & \text{ IF } x \text{ is } A_2 \text{ and } y \text{ is } B_2 \\
 & \text{ THEN } f_2 = p_2x + q_2y + r_2
 \end{aligned} \tag{16}$$

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.

$$\begin{aligned}
 O_{1,i} &= \mu_{A_i}(x), \text{ for } 1,2 \\
 O_{2,i} &= \mu_{B_i}(y), \text{ for } 3,4 \\
 \text{여기서 } \mu_{A_i}(x) &= \exp\left\{-\frac{1}{2}(x-c_i)^T \Sigma_i^{-1}(x-c_i)\right\}
 \end{aligned} \tag{17}$$

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

$$O_{2,i} = w_i = \mu_{A_i}(x) \times \mu_{B_i}(y), \quad i=1,2 \tag{18}$$

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

$$O_{3,i} = \overline{w}_i = \frac{w_i}{w_1 + w_2}, \quad i=1,2 \tag{19}$$

4층 : 정규화된 가중치 값과 결론부의 곱으로 출력을 나타낸다.

$$O_{4,i} = \overline{w}_i f_i = \overline{w}_i (p_i x + q_i y + r_i) \tag{20}$$

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

$$O_{5,i} = \sum_i \overline{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{21}$$

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

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

$$\begin{bmatrix} \overline{w}_1^{(1)} & \overline{w}_1^{(1)} x^{(1)} & \overline{w}_1^{(1)} y^{(1)} & \dots \\ \overline{w}_1^{(2)} & \overline{w}_1^{(2)} x^{(2)} & \overline{w}_1^{(2)} y^{(2)} & \dots \\ \dots & \dots & \dots & \dots \\ \overline{w}_1^{(p)} & \overline{w}_1^{(p)} x^{(p)} & \overline{w}_1^{(p)} y^{(p)} & \dots \end{bmatrix} \begin{bmatrix} r_1 \\ p_1 \\ q_1 \\ \vdots \end{bmatrix} = \begin{bmatrix} d^{(1)} \\ d^{(2)} \\ d^{(3)} \\ \vdots \end{bmatrix} \tag{22}$$

이를 간략이 표현하면 다음과 같다.

$$AX = B \tag{23}$$

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

$$X^* = (A^T A)^{-1} A^T B \tag{24}$$

여기서 A^T 는 A 의 전치 행렬이다.

전방향(forward) 학습이 위와 같이 끝나면 오차를 기반으로 하는 역전파 (Back-propagation) 알고리즘을 통하여 전제부 파라미터를 학습하게 된다. 이 경우, 주어진 데이터 출력 y_d 와 모델 출력 y^* 에 대하여 Mean Square Error (MSE)를 최소화하기 위하여 다음과 같은 오차 함수를 정의한다.

$$E = \frac{1}{2} \sum_{i=1}^n (y_d - y^*)^2 \tag{25}$$

여기서 y_d 는 입력 벡터 $x = [x, y]^T$ 이다. 소속함수의 모양이 전제부 파라미터에 의해 정의되므로 오차함수 E 는 결론부 파라미터는 고정되며, E 를 감소시키기 위해서 전체

부 파라미터 Σ, c 를 다음과 같은 방법에 의하여 학습을 실시한다.

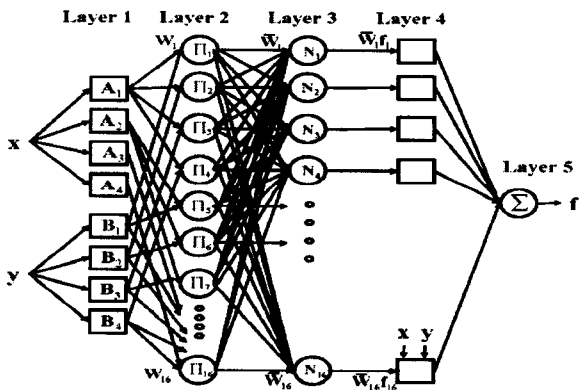
$$\Sigma(t+1) = \Sigma_j^i(t) - \eta \frac{\partial E}{\partial \Sigma_j^i} \quad (26)$$

$$c_j^i(t+1) = c_j^i(t) - \eta \frac{\partial E}{\partial c_j^i} \quad (27)$$

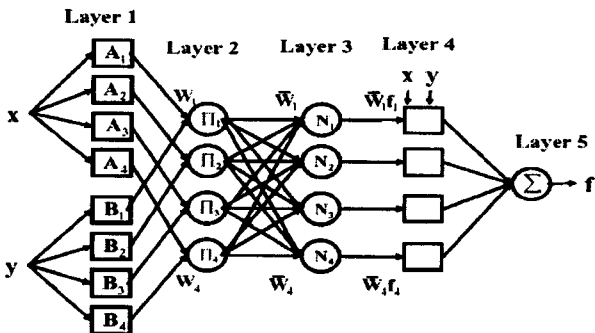
$$\eta = \frac{k}{\sqrt{\sum_a \left(\frac{\partial E}{\partial a}\right)^2}} \quad (28)$$

여기서 η 는 학습률이며 k 는 파라미터 공간에서 경사벡터의 이동한 거리로 수렴속도를 변화하기 위한 스텝 크기를 변화할 수 있다.

이를 이용하여 전체부 소속함수의 파라미터와 결론부 1차 선형 방정식의 파라미터를 모두 구할 수 있다. 입력이 2개이며 각 소속함수의 수를 4개로 하였을 경우 전체적인 구성을 일반적인 ANFIS와 비교하면 그림 7과 같이 표현할 수 있다.



(a) Conventional ANFIS



(b) Proposed ANFIS

그림 7 기존의 ANFIS와 제안된 ANFIS의 비교

Fig. 7 Comparison between the conventional ANFIS and the proposed ANFIS

그림 7(a)는 일반적인 ANFIS 구조로서 입력의 차원이나 소속함수의 수가 증가할 경우 Layer 2에서 Layer 4까지의 구조(노드)는 지수함수 형태로 증가한다. 이 경우 방대한 구조로 ANFIS의 학습 및 실제 구현에 심각한 문제가 발생할 수 있다. 그림 7(b)의 경우 같은 조건의 입력 및 소속함수의 증가가 발생하여도 Layer 2에서 Layer 4는 단지 소속함수의 수에 비례하므로 전체 구조는 단지 소속함수의 수에만 비례하여 그리드 분할에서 발생하는 문제점을 해결할 수 있다. 즉, 입력으로 표현되는 Layer 1과는 무관하게 Layer 2~4는 클러스터의 수와 일치하는 구조를 가지게 된다. 입력공간에서의 클러스터링이 규칙공간에서의 규칙 생성에 이용됨으로서 제안된 모델의 성능 및 규칙 증가 억제를 동시에 만족할 수 있다. 즉 제안된 알고리즘의 전체 구성은 클러스터의 초기치 및 클러스터의 구조를 자율적으로 실시하는 하나의 알고리즘으로 구현하며 이를 뉴로-퍼지 시스템의 초기 파라미터로 설정한다. 이는 자체 알고리즘의 개선 및 순차적 최적화를 통하여 전체 모델의 성능을 개선하는 것이다.

3. 실험 및 결과

제안된 시스템에 적용된 데이터로는 전형적인 비선형 시계열 시스템인 Box-Jenkins의 가스 시스템 데이터이다[2]. 데이터의 구성은 입력으로 가스 흐름률 $u(t)$ 와 출력 CO_2 밀도 $y(t)$ 를 가지며 총 296개의 데이터 집합 중 유효한 290개를 사용하였으며 기존 논문들에서 시뮬레이션한 조건과 같이 홀수열의 145개 집합을 학습 데이터로 나머지 데이터를 검증 데이터로 사용하여 시스템을 학습 및 검증에 사용하였다[15].

먼저 입력 데이터를 0과 1 사이에서 정규화 시킨 후, 시뮬레이션을 실시하였다. 여기서 ζ 는 0.01로 하였다. 그림 10에서는 제안된 알고리즘에 의하여 최적화된 초기 뉴로-퍼지 모델의 소속함수 및 이를 50회 학습한 후의 최적화된 소속함수를 나타내었다. 그림 8에서 볼 수 있듯이 뉴로-퍼지 모델의 초기 소속함수와 최적화된 파라미터의 형태가 어느 정도 유사한 형태를 가지는 것을 알 수 있다. 이를 표 1에서 Root Mean Square Error (RMSE) 오차값으로 초기학습 오차와 최종 학습 오차를 나타내었다.

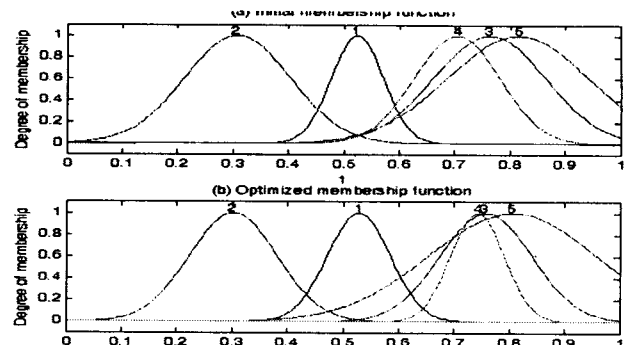


그림 8 초기 소속함수와 최종 소속함수 비교

Fig. 8 Compare with initial and final membership function

표 1 성능평가

Table. 1 Performance Index (RMSE)

	초기 (학습 전) 오차	최종 (학습 후) 오차
학습데이터 오차	0.0733	0.0563
검증데이터 오차	0.0835	0.0899

또한 이를 이전의 연구 결과와 비교하면 표 2와 같다. 표에서 볼 수 있듯이 학습 데이터에 의한 성능뿐만 아니라 검증 데이터의 성능까지 동시에 개선되는 것을 알 수 있다. 이는 학습 데이터에 의한 과학습(overfitting) 문제가 제안된 알고리즘에서는 발생하지 않는 것을 알 수 있다.

표 2 성능비교

Table. 2 Compare with performance index (RMSE)

	Number of Input	Number of MF	Number of rule	Training error	Checking error
Jang[2]	2	4	16	0.135	0.530
Oh[5]	2	4	4	0.079	0.219
	2	6	6	0.024	0.330
Park[10]	6	6	6	0.020	0.264
Kim[14]	2	10	10	0.019	0.340
Ours	2	5	5	0.056	0.089

4. 결 론

본 논문에서는 뉴로-퍼지 모델링을 위한 데이터의 클러스터링에 관련된 최적의 파라미터들의 초기치 결정과 클러스터링의 구조 개선을 통하여 시스템의 성능을 향상시켰다. 파라미터 초기치의 선택에 의한 클러스터링 갯수의 수렴 및 시스템의 성능 개선을 통하여 규칙의 개수를 줄임으로써, 일반적으로 전체 모델의 구조가 복잡해짐으로써 발생하는 학습 속도 저하와 실제 구현에 대한 문제점에 대하여 다음과 같이 성능을 개선하였다.

- (1) 초기 파라미터의 자율적인 구조 결정 및 최적화 : 제안된 알고리즘에서는 초기 파라미터를 전체 학습데이터로 설정함으로써 임의로 생성하는 일반적인 기법과 달리 매 시도마다 동일한 결과를 추정하며 사전에 정해진 ζ 에 의하여 일정한 클러스터의 수로 수렴을 한다.
- (2) 뉴로-퍼지 모델의 규칙 결정 : 최적화 알고리즘 진행 중에서도 클러스터링 기법을 유지하여 실제 뉴로-퍼지 모델의 전체 규칙은 클러스터의 수와 같게 되어 소속함

수의 수나 입력의 차원이 증가하여도 규칙은 단지 소속함수의 수와 같게 하였다.

- (3) 모델의 성능 : 뉴로-퍼지 모델의 초기치를 최적화함으로써 초기 학습성능 및 최종 학습성능을 개선하였으며 학습 데이터의 성능뿐만 아니라 검증 데이터의 성능까지 개선하였다.

이러한 뉴로-퍼지 모델의 향후 연구과제로는 제안된 클러스터링의 연산 과정 단순화에 의한 연산량 감소와 뉴로-퍼지 시스템의 구조 개선을 통한 전체 모델의 성능 개선 등이 있다.

참 고 문 헌

- [1] C. T. Lin, C. S. G. LEE, Neural Fuzzy Systems : A Neuro-Fuzzy Synergism to Intelligent Systems, Prentice Hall, 1996.
- [2] J. S. R. Jang, C. T. Sun, E. M. Mizutani, Neuro-Fuzzy and Soft Computing : A Computational Approach to Learning and Machine Intelligence, Prentice Hall, 1997.
- [3] J. S. R. Jang, "ANFIS : Adaptive Network-based Fuzzy Inference System", IEEE Trans on System, Man, and Cybernetics, Vol.23, No. 3, pp. 665-685, 1993.
- [4] 김승석, 박근창, 유정웅, 전명근, "GMM과 클러스터링 기법에 의한 뉴로-퍼지 시스템 모델링", 한국 퍼지 및 지능시스템 학회 논문지, Vol. 12, No. 6, pp. 571-576. 2002.
- [5] S. K. Oh, Witold Pedrycz, "Identification of Fuzzy system by means of an auto-tuning algorithm and its application to nonlinear systems", Fuzzy Sets and Systems, Vol. 115, pp. 205-230, 2000.
- [6] Witold Pedrycz, "Conditional Fuzzy C-Means", Pattern Recognition Letters, Vol. 17, Issue. 6, pp. 625-631, 1996.
- [7] B. J. Park, W. Pedrycz, S. K. Oh, "Identification of fuzzy models with the aid of evolutionary data granulation", Control Theory and Applications, IEE Proceedings, Vol. 148, Issue. 5, pp. 406-418, 2001.
- [8] Roy L. Streit, Tod E. Luginbuhl, "Maximum Likelihood training of Probabilistic Neural Networks", IEEE Trans on Neural Networks, Vol. 5, No. 5, pp. 764-782, 1994.
- [9] Guorong Xuan, Wei Zhang, Peiqi Chai, "EM algorithm of Gaussian Mixture Model and Hidden Markov Model", Image Processing Proceedings, International Conference on, Vol. 1, pp. 145-148. 2001.
- [10] Ethem Alpayd, "Soft Vector Quantization and the EM algorithm", Neural Network, Vol. 11, Issue. 3, pp. 467-477, 1998.
- [11] Xiangyu Yang, Jun Liu, "Unsupervised Learning of

Finite Mixture Models”, Pattern Recognition Letters, Vol. 23, Issue. 5, pp. 501-512, 2002.

[12] Ching-Chang Wong, Chia-Chong Chen, Mu-Chun Su, “A novel algorithm for data clustering”, Pattern Recognition”, Vol. 34, Issue. 2, pp. 425-442, 2001

[13] 김승석, 박근창, 유정웅, 전명근, “계층적 클러스터링과 Gaussian Mixture Model을 이용한 뉴로-퍼지 모델링”, 한국 퍼지 및 지능시스템 학회 논문지, Vol. 13, No. 5, pp. 512-519, 2003.

[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 Model”, ICCAS 2002, 2002.

[15] J. S. R. Jang, “Input selection for ANFIS learning”, Fuzzy Systems, Proceedings of the Fifth IEEE International Conference on, Vol. 2, pp. 1493-1499, 1996.

저 자 소 개



김 승 석 (金 承 石)

1972년 12월 19일생. 1998년 충주대학교 전기공학과 졸업. 2002년 충북대학교 전기공학과 졸업(석사). 2002년~현재 동 대학원 박사과정

Tel : 043-261-2422
Fax : 043-263-2419
E-mail : powerkimss@hotmail.com



유 정 웅 (柳 正 雄)

1941년 1월 20일생. 1965년 한양대학교 졸업. 1976년 단국대학교 전기공학과 대학원 졸업(석사). 1988년 단국대학교 전자공학과 대학원 졸업(공학박). 1979년~현재 충북대학교 전기공학과 교수. 충북대학교 정보통신연구소 교수

Tel : 043-261-2422
Fax : 043-263-2419
E-mail : jwryu@vod.chungbuk.ac.kr



김 성 수 (金 聖 洙)

1959년 5월 17일생. 1983년 충북대학교 전기공학과 졸업. 1989년 University of Arkansas-Fayetteville(석사). 1997년 University of Central Florida 졸업(공학박). 2002년~현재 충북대학교 전기공학과 부교수. 충북대 정보통신연구소 교수.