2004년도 대한전기학회 하계학술대회 논문집 2004.7.14-16

자기-구성 클러스터링에 의한 퍼지 모델링

김승석, 전병석, 김주식, 유정웅, 김성수
충북대학교 전기공학과

Fuzzy Modeling using Self-Organizing Clustering

Sung-Suk Kim, Byung-Suk Jeon, Ju-Sik Kim, Jeong-Woong Ryu, Sung-Soo Kim
Dept of Electrical Engineering, Chungbuk National University

Abstract - 본 논문에서는 주어진 데이터를 나누어 부분공간으로 구성하여 특성을 구분하거나 또다른 모델의 입력 파라미터로 제공하는 방법 중 하나의 클러스터링의 성능 개선과 이를 이용하여 퍼지 모델링을 실시하였다. 일반적인 클러스터링에서 볼 수 있는 초기 파라미터 결정 문제와 알고리즘의 수렴 문제에 대하여 문제점을 개선하였으며 클러스터링에 의하여 추정된 파라미터를 퍼지 모델에 적용하였다. 또한 일반적인 퍼지 모델의 경우 각 입력의 차원이 서로 독립적으로 구성되어 있어 데이터에서 존재하는 입력간의 상관관계를 고려하지 않았다. 제안된 퍼지 모델에서는 클러스터링에서 추정된 입력간의 상관관계(공분산)까지 고려하여 모델의 성능을 개선하였다. 제안된 논문의 유용성을 시뮬레이션을 통하여 보이고자 한다.

1. 서 론

주어진 데이터를 기반으로 모델을 구성하는 인공지능 기법은 모델의 학습과 구성에 대하여 다양한 시도들이 연구되어 왔다. 모델의 초기치에 의한 모델의 학습 성능이나 수렴 등에 대한 문제들은 해당 시스템에 대하여 결정적인 성능 지표가 되며 이에 대한 연구는 현재 활발하게 이루어지고 있다. 또한 모델의 구조에 대한 연구는 실제 알고리즘의 적용이 가능하도록 복잡한 구조에서 좀 더 간단한 구조로의 시도가 이루어져 왔다[1][2].

데이터를 여러 개의 부분 공간으로 나누는 방법으로는 선형 분리 방법을 이용하는 방법과 클러스터 파라미터를 이용하여 데이터를 군집형태로 나누는 방법이 있는데 데이터를 각 데이터가 가지는 정보량을 이용하거나 다른 모델로의 계층적 연결을 고려할 때 클러스터링에 의한 데이터의 특성 분리가 더 효율적인 경우가 있다[3-5]. 이 경우 선형 분리 방법과 마찬가지로 데이터를 정확하게 표현할 수 있는 클러스터의 수 결정과 각 클러스터 파라미터 최적화 문제가 발생한다[3]. 클러스터링 방식으로는 크게 사전 임계값을 이용하여 클러스터의 수를 추정하는 방식과 클러스터의 수가 주어졌을 때 파라미터를 최적화하는 방식이 있다. 특정한 임계값을 이용하여 클러스터의 수를 추정하는 경우 임계값의 변화에 따라 클러스터의 수가 심하게 변동할 수 있으며 클러스터의 수를 최적화하는 경우 초기 학습 파라미터 결정 문제가 제기될 수 있다[1]. 클러스터로부터 추정된 파라미터를 또 다른 모델의 초기 파라미터로 이용하는 경우 파라미터의 변환이나 손실이 발생할 수 있다. 일반적으로 클러스터 파라미터가 퍼지 모델의 전제부 초기 파라미터로 이용되는 경우 데이터 입력간에 존재하는 상관관계에 대한 정보는 전달되지 않는다. 퍼지 모델에서의 각 입력은 서로 영향을 주지 않는다고 가정하고 모델링을 실시함으로써 상관관계 정보를 가지고 있는 공분산행렬이 모두 사용되지 않는다[6][7].

본 논문에서는 클러스터 파라미터의 수를 자율적으로 결정하며 동시에 파라미터 최적화를 실시하는 클러스터 알고리즘을 이용하여 클러스터링을 실시하였다. 사전에 주어진 임계값에 대하여 알고리즘이 진행되는 동안 특정한 클러스터의 수로 수렴을 하면서 동시에 최적화 과정을 실시한다. 또한 추정된 파라미터는 데이터의 상관 관계까지 고려하는 퍼지 모델의 입력으로 주어져 모델을 구성하게 된다. 제안된 방법의 우수성을 시뮬레이션을 통하여 기존 모델과의 비교를 통하여 보이고자 한다.

2. 제안된 클러스터링 알고리즘

클러스터링의 기본 개념은 유사성을 가지는 데이터를 같은 클러스터에 속하게 하면 그렇지 않은 데이터를 다른 클러스터에 속하게 하는 것이다. 이 때 데이터 집합을 몇 개의 유사성을 가지는 부분 공간으로 나누어야 하는 문제와 학습을 통한 클러스터 파라미터 최적화 문제를 고려할 수 있다. 각 클러스터 파라미터를 중심과 분산, 공분산이라 하고 알고리즘에서 적절한 클러스터의 수 결정과 최적화는 다음과 같이 할 수 있다.

먼저 Chen 알고리즘에서의 추정 방법을 보면 한 데이터가 특정 클러스터에 속하는지 여부를 결정하기 위하여 유사도(similarity)를 다음과 같이 측정한다.

$$s_{ij} = \exp\left(-\frac{1}{2}\frac{\| x_i - x_j \|^2}{\sigma^2}\right) \quad (1)$$

이를 사전에 지정한 임계값 ζ를 이용하여 일정한 범위를 벗어나는 유사도를 제거하여 일정한 범위 안의 유사도만을 가지는 데이터를 같은 클러스터에 속하게 한다[8].

$$r_{ij} = \begin{cases} 0 & \text{if } s_{ij} < \zeta \\ r_{ij} & \text{otherwise} \end{cases} \quad (2)$$

이 경우 각 데이터가 특정 클러스터 파라미터를 추정하기 위하여 영향을 주기 위한 조건은 식(1)에서의 유사도가 ζ보다 큰 값을 가져야 한다.

Chen 알고리즘에서 σ의 값은 ζ와 함께 사전에 지정되어져 있어야 하며 이들 값에 의하여 클러스터의 범위와 수가 결정된다. 이 알고리즘의 문제점은 클러스터의 형태가 타원형을 가질 때 클러스터의 추정 결과가 이를 반영하지 못하며 경우에 따라 클러스터의 크기와 형태가 다를 경우 문제점을 발생한다. 따라서 제안된 알고리즘은 클러스터 형태 정보를 가지는 σ를 각 입력간의 상관관계까지 고려하는 공분산 행렬 Σ를 이용하여 추정하는 형태를 취하였다. 또한 Σ의 결정은 알고리즘 시작 전에 결정하는 것이 아니라 알고리즘 진행 동안 추정된 파라미터의 결과에 따라 자율적으로 변화하도록 하였다. 식(1)을 제안된 방법에서는 다음과 같이 표현하였다[9].

$$s_{ij} = \exp\left(-\frac{1}{2}\| x_i - x_j \|_{\Sigma_i^{-1}} \| x_i - x_j \|\right) \quad (3)$$

공분산 행렬을 통하여 각 입력의 차원에서 존재하는 상관관계를 유사도 측정에 이용하였다. 이 경우 임계값 ζ의 변화에 따라 클러스터의 수가 급변하는 것을 제한하고 일정한 ζ 범위에서 클러스터 수가 일정하게 추정될 수 있도록 s_{ij}의 값을 제한할 필요가 있다. 명확하지 않은 분포를 데이터의 경우 임계값의 변화에 따른 클러스터의 수 변동이 크다. 따라서 제안된 알고리즘에서는 ζ의 변동에 역으로 제한을 가하는 방법으로 s_{ij}를 다음과 같이 제약을 두었다.

$$s_{ij}=\exp\left(-\frac{1}{2}\,\|x_i-x_j\|\,(\Sigma_i\times\zeta^{-1})\,\|x_i-x_j\|\right) \quad (4)$$

이 경우 식(4)에 의한 유사도와 식(2)의 클러스터 추정 조건 결정 조건이 상호 영향을 주어 ζ의 변화에 유사도의 변화를 비선형적인 역증가 형태로 구성되어 진다.

이를 알고리즘으로 표현하면 다음과 같다.

단계 1 : 모든 데이터를 클러스터 파라미터 중심으로 정한다. 이 경우 데이터의 수는 클러스터의 수가 된다. 즉 $v_i=x_i$가 된다. 또한 유사도 임계값 ζ를 정한다[8].

단계 2 : 각 클러스터와 클러스터 또는 데이터 간의 유사도를 식(4)와 같이 추정한다. 이때 공분산 행렬 Σ_i는 다음과 같이 추정한다.

$$\Sigma_j=\frac{1}{n}\sum_{i=1}^{n}(x_i-v_j)(x_i-v_j)^T \quad (5)$$

단계 3 : 유사도를 측정하여 임계값 이하인 데이터의 유사도를 클러스터 추정에서 식(2)와 같이 제외한다.

단계 4 : 클러스터 중심을 다음과 같이 추정한다[2].

$$v_j'=\frac{\sum_{i=1}^{n}r_{ij}v_j}{\sum_{i=1}^{n}r_{ij}} \quad (6)$$

단계 5 : 클러스터 파라미터가 $v_i=v_i'$이면 알고리즘을 종료하고 그렇지 않으면 새로 추정된 v_i'을 v_i에 대입하고 단계 2로 돌아가 반복 수행한다.

3. 입력의 상관관계를 고려한 퍼지 모델링

본 논문에서 이용된 퍼지 모델은 언어적 입력 형태를 가지는 전제부와 1차 선형 방정식 형태를 가지는 Takagi-Sugeno-Kang (TSK) 퍼지 모델이다[2-4]. 이 모델의 특징은 다음과 같이 결론부가 선형 방정식으로 되어 있어 비 퍼지화 과정이 필요 없으며 결론부 파라미터 추정을 최소자승법을 통하여 쉽게 할 수 있다.

$$\begin{aligned} R^1&: \text{IF } x \text{ is } A_1 \text{ and } y \text{ is } B_1 \\ &\quad \text{THEN } f_1=p_1x+q_1y+r_1 \\ R^2&: \text{IF } x \text{ is } A_2 \text{ and } y \text{ is } B_2 \\ &\quad \text{THEN } f_2=p_2x+q_2y+r_2 \end{aligned} \quad (7)$$

Jang이 제안한 적응 네트웍 기반 뉴로-퍼지 시스템에서의 TSK 퍼지 모델의 구성은 모두 5층으로 구성이 되는데 1층과 2층에서 각 입력 차원에 대하여 소속함수를 생성하고 이를 융합하는 과정이 다음과 같이 포함되어 있다.

단계 1 : 각 입력차원에 대한 퍼지 소속함수 생성

$$\begin{aligned} &O_{1,i}=\mu_{A_i}(x), \quad \text{for } 1,2 \\ &O_{2,i}=\mu_{B_i}(y), \quad \text{for } 3,4 \\ &\text{여기서 } \mu_{A_i}(x)=\exp\left(-\frac{1}{2}\left(\frac{x-c_i}{\sigma^2}\right)^2\right) \end{aligned} \quad (8)$$

단계 2 : 각 입력 차원 간의 소속함수 융합

$$O_{2,i}=w_i=\mu_{A_i}(x)\times\mu_{B_i}(y), \quad i=1,2 \quad (9)$$

이 경우 각 차원별로 소속함수를 생성한 후에 이를 다시 융합하는 형태를 취하므로 각 입력간의 상관관계가 고려되지 않았다. 따라서 제안된 퍼지 모델은 다음과 같이 기존 모델의 1층과 2층을 하나의 층으로 구성하는 다음과 같은 퍼지 모델을 구성하였다.

$$\mu_i=\exp\left(-\frac{1}{2}(x_j-\mu_i)^T\Sigma_i^T(x_j-\mu_i)\right) \quad (10)$$

즉 본 논문에서는 1차원의 데이터 분포 정보를 나타내는 σ^2에서 입력간의 상관관계까지 포함하는 Σ_i를 이용하여 퍼지 모델의 성능을 개선하고자 하였다.

제안된 방법에 의한 TSK 퍼지 모델의 각 층은 다음과 같다.

1층 : 각 입력에서 소속도를 통하여 퍼지화 과정에 상호 상관관계를 고려된 변환이 식(10)으로 이루어진다. 즉 기존의 TSK 퍼지 모델의 1층과 2층이 하나의 층으로 구성되어 진다.

2층 : 각 퍼지 규칙들은 정규화 과정을 실시한다.

$$\overline{w_i}=\frac{\mu_i}{\sum_{j=1}^{c}\mu_j} \quad (11)$$

3층 : 정규화된 가중치 값과 결론부의 곱으로 출력을 나타낸다.

$$O_{4,i}=\overline{w_i}f_i=\overline{w_i}(p_ix+q_iy+r_i) \quad (9)$$

4층 : 가중 평균법에 의한 최종 출력을 구한다.

$$O_{5,i}=\sum_i\overline{w_i}f_i=\frac{\sum_i w_if_i}{\sum_i w_i} \quad (10)$$

이를 그림으로 나타내면 기존의 퍼지 모델을 그림 1에, 제안된 퍼지 모델을 그림 2에 나타내었다.

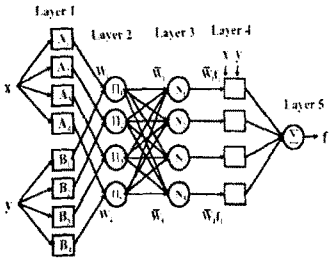

그림 1. TSK 퍼지 모델

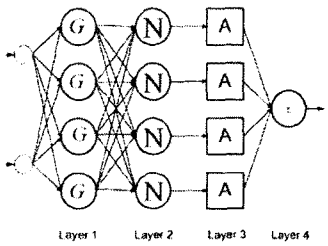

그림 2 제안된 TSK 퍼지 모델

4. 시뮬레이션 및 결과

먼저 제안된 클러스터링 알고리즘을 이용하여 클러스터를 추정하였을 때 결과를 보면 다음과 같다.

각 입력 데이터를 0과 1사이에서 정규화 하여 클러스터 추정에 이용하였다. 시뮬레이션에 이용된 데이터는 전형적인 비선형 시계열 데이터인 Box-Jenkins의 가스로 데이터를 이용하였다. ζ를 0.1로 하였을 때 제안된 클러스터링 알고리즘 추정 결과가 그림 3에 표현하였으며 퍼지 모델을 통하여 최종 출력이 그림 4와 같다. 이를 일반적인 TSK 퍼지 모델을 이용하여 추정하였을 때와 비교하면 표 1과 같다.

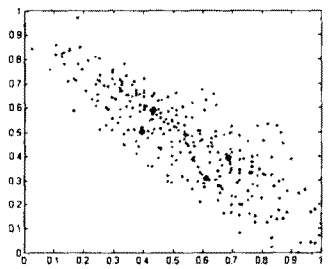

그림 3. 클러스터 추정 결과

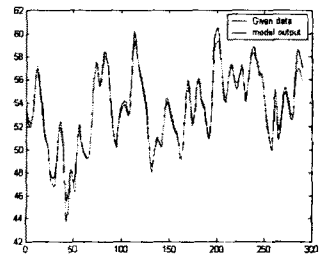

그림 4. 제안된 모델의 추정 결과

표 1. 성능지표

	기존의 TSK 모델	제안된 모델
Cluster : 4	0.508	0.434
Cluster : 6	0.554	0.445

5. 결 론

본 논문에서는 자율적으로 클러스터를 생성하며 최적화 시키는 클러스터 알고리즘과 각 입력의 차원을 고려한 퍼지 모델링을 제안하였다. 클러스터 추정에서 발생하는 클러스터의 수 결정과 최적화 문제를 제안된 클러스터링을 이용하여 해결하였으며 기존의 알고리즘과는 달리 사전에 지정하는 파라미터에 의한 클러스터 결과의 급격한 변화를 억제하였으며 초기 클러스터 파라미터 지정 문제 또한 개선하였다. 또한 퍼지 모델에서의 입력간의 상관관계를 고려한 모델링을 실시하여 성능을 개선하였다.

향후 연구과제로는 클러스터링 알고리즘의 연산량 축소 문제와 다양한 최적화 방법의 도입이 있으며 퍼지 모델에 대하여는 양방향 학습이 가능하도록 뉴로-퍼지 모델로 전환하여 양방향 학습이 가능한 알고리즘의 구현 등이 있다.

[참 고 문 헌]

[1] Simon Haykin, Neural Networks A Comprehenshive Foundation, Macmillan Publishing Company, 1994.

[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] 김승석, 곽근창, 유정웅, 전명근, "GMM과 클러스터링 기법에 의한 뉴로-퍼지 시스템 모델링", 한국퍼지및지능시스템학회 논문지, Vol. 12, No. 6, pp. 571-576, 2002.

[4] 김승석, 곽근창, 유정웅, 전명근, "계층적 클러스터링과 Gaussian Mixture Model을 이용한 뉴로-퍼지 모델링", 한국퍼지및지능시스템학회 논문지, Vol. 13, No. 5, pp. 512-519, 2003.

[5] Richard O. Duda, Peter E. Hart, David G. Stork, Pattern Classification, John Wiley & Sons, Inc. 2001.

[6] Timothy J. Ross, "Fuzzy Logic with Engineering Applications", McGraw-Hill, Inc, 1995.

[7] Todd. K. Moon, "The Expectation- Maximization Algorithm" , IEEE Signal Processing Magazine, 1996.

[8] 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.

[9] Ethem Alpayd, "Soft vector quantization and the EM Algorithm", neural Network, Vol. 11, Issue. 3, pp. 467- 477 1998.