

클러스터링과 퍼지 규칙을 이용한 뉴로-퍼지 시스템 학습 및 모델링

김승석*, 박근창**, 김주식***, 유정웅*

*충북대학교 전기공학과, **한국전자통신연구원, ***특허청

Learning and Modeling of Neuro-Fuzzy modeling using Clustering and Fuzzy rules

Sung-Suk Kim*, Keun-Chang Kwak**, Ju-Sik Kim***, Jeong-Woong Ryu*

*Chungbuk Nat'l Univ, **ETRI, ***KIPO

Abstract - 본 논문에서는 뉴로-퍼지 모델의 전체부 소속함수의 새로운 학습방법을 통한 모델링 기법을 제안한다. 모델의 크기와 학습시간을 줄이는 기법으로 클러스터링 기법을 이용한 모델의 초기 파라미터 결정 방법이 있다. 이는 클러스터링 후 이들 파라미터를 다시 모델에 적용하여 모델을 학습하는 순차적 방법으로써 모델의 학습이 끝난 후의 전체부 파라미터가 클러스터링 파라미터와 연관성을 가지지 못하는 경우가 발생하였다. 또한 오차미분 기반 학습에서는 전체부 초기치가 국부적 최적해에서 벗어나지 못하는 문제점을 가지고 있다. 본 논문에서는 자율적으로 클러스터의 수를 추정하며 이들 파라미터를 최적화하며 이를 이용하여 뉴로-퍼지 모델의 학습을 실시하는 학습기법을 제안하였다. 제안된 방법에서는 기존의 오차미분 기반 학습을 클러스터링 기반 학습으로 확장하였으며 이를 이용한 모델의 성능을 기존의 연구결과와 비교하여 우수성을 보인다.

모델에 대하여 해석이 가능하도록 하였다. 시뮬레이션을 통하여 제안된 방법의 유용성을 보인다.

2. 클러스터링 기반 뉴로-퍼지 모델링

2.1 클러스터링

클러스터링의 기본 개념은 주어진 입력공간에 대하여 각 데이터간의 유사도를 측정하여 높은 유사도를 가지는 클러스터에 속하게 하고 그렇지 않은 데이터는 다른 클러스터에 속하게 하는 것이다. 이때 적용하는 유사도 측정함수에 따라 클러스터 추정 성능이 차이가 난다. 본 논문에서는 클러스터의 중심뿐만 아니라 공분산까지 고려한 유사도 측정방식을 다음과 같이 이용하였다[7].

$$r_{ij} = \exp\left(-\frac{1}{2}(x_i - c_j)^T(\zeta \cdot \Sigma_j)^{-1}(x_i - c_j)\right) \quad (1)$$

여기서 c_j 는 해당 클러스터의 중심이며 Σ_j 는 이에 대응하는 공분산 행렬이고 ζ 는 알고리즘이 시작하기 전에 지정해주는 임계치이다. 사전 임계치에 의하여 자율적으로 클러스터의 수를 수렴시키기 위하여 다음과 같이 유사도를 제한하였다.

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

식(1)과 식(2)의 연산 후의 유사도 범위는 ζ 급격한 변화에도 강인한 성능을 가진다. 유사도를 이용한 클러스터의 중심은 가중 평균법에 의하여 다음과 같이 추정하였다[8].

$$c_j = \frac{\sum_{i=1}^N r_{ij} x_i}{\sum_{i=1}^N r_{ij}} \quad (3)$$

클러스터 중심을 이용하여 Maximum likelihood 최대화 시키는 방향으로 이에 관련된 공분산 행렬과 사전확률을 추정하여 파라미터를 최적화 시켰다. 이러한 기법의 문제점으로 각 클러스터 파라미터는 모든 데이터와 연산을 실시하는데 초기 클러스터 중심의 수가 큰 경우 알고리즘의 구현에서 연산량에 대한 부담이 존재하게 된다. 이러한 문제를 해결하기 위하여 데이터의 밀집여부를 조사하여 대표 클러스터를 추정하는 변형된 Subtractive 클러스터링을 다음과 같이 도입하였다.

$$c_{new} = c_j, \text{ if } D_j = \max\left(\sum_{i=1}^N U_{ij}\right) \quad (4)$$

여기서 D_j 는 클러스터 중심에서의 누적 밀도이다. 이를 이용하여 다음과 같은 제거 유사도 s_{ij} 를 추정한다.

$$s_{i,new} = \exp\left(-\frac{1}{2}(c_i - c_{new})^T \Sigma_{new}^{-1}(c_i - c_{new})\right) \quad (5)$$

여기서 Σ_{new} 는 c_{new} 에 대응하는 오차에 대한 공분산 행렬을 의미한다. 또한 제거 범위를 제한하기 위해 다음과

1. 서 론

비선형 시스템 모델의 최적화에서 뉴로-퍼지 시스템은 기존의 선형 시스템을 기반으로 하는 모델에 비하여 좋은 성능을 보여 왔다[1-3]. 수학적인 모델링 기법에서 벗어나 사람의 학습이나 의사 판단 등의 능력을 모사한 뉴로-퍼지 모델이 경우 학습 모델의 형태나 학습방법에 따라 전체 모델의 성능이 달라지는 경우가 발생한다[2]. 또한 격자분할에 의한 모델의 전체부 소속함수 생성은 입력의 차원이 증가하거나 소속함수의 수가 증가할 경우 뉴로-퍼지 모델의 규칙이 지수함수적으로 증가하는 문제점을 가진다. 이러한 문제들을 해결하기 위하여 다양한 연구들이 진행되어 왔는데 주어진 입력력 공간을 각각의 의미를 가지는 부분 공간으로 분할하여 전체부 소속함수를 부여하는 클러스터링 기법이 대표적인 예이다[4][5]. 클러스터링 기법은 하나의 비선형 데이터 공간에 여러 개의 패턴이 존재한다고 가정할 때 각 패턴과의 유사도를 측정하여 주어진 조건에 만족하는 클러스터에 속하도록 하는 방법이다. 클러스터링은 단독으로 학습하여 성능을 평가하거나 다른 기법과 순차적으로 융합하여 전체 모델에 대하여 평가를 하기도 한다[4]. 순차적 기법에 의한 전체 시스템의 모델링의 경우 클러스터링에 의하여 학습된 파라미터가 최종 모델의 초기치로 하여 다시 학습을 실시하는 경우 클러스터에서 추정된 파라미터가 무의미하게 될 수 있다[4][6].

이러한 문제점들을 개선하기 위하여 제안된 방법에서는 기존의 오차미분 기반 뉴로-퍼지 모델링을 클러스터링 학습과 연관시켜 모델을 구성하는 클러스터링 기반 뉴로-퍼지 모델을 제안하였다. 클러스터의 학습이 모델의 학습과 직접적인 연관이 되도록 구성하여 두 개의 학습을 하나의 모델로 고려할 수 있도록 하면서도 각각의

같이 $s_{i,new}$ 의 범위를 제한한다.

$$s_{ij} = \begin{cases} 0, & \text{if } s_{ij} < \epsilon \\ s_{ij}, & \text{otherwise} \end{cases} \quad (6)$$

클러스터링이 진행되는 동안 클러스터 후보 중심들은 특정 지역으로 수렴하는데 수렴하면서 밀집하게 되는 클러스터 파라미터를 대표 파라미터를 제외한 나머지를 모두 소거함으로써 연산량을 줄일 수 있다. c_{new} 를 이용하여 인근의 클러스터 중심을 제거한 뒤 새로운 클러스터 중심 $c_{new'}$ 을 다음과 같이 추정하여 알고리즘을 반복 수행한다[2].

$$c_{new'} = c_j, \text{ if } D_j' = \max \left(D_{new} - \sum_{i=1}^c \left(\sum_{j=1}^c s_{i,j} \right) \right) \quad (7)$$

추정된 클러스터 정보는 뉴로-퍼지 모델의 구성에 직접 적용되어 모델을 구성하거나 평가하는데 쓰여 진다.

2.2 뉴로-퍼지 모델의 학습방법

클러스터링에 의하여 생성된 클러스터의 형태가 뉴로-퍼지 모델의 전체부 소속함수와 유사하거나 같을 때 이를 직접적으로 이용할 수 있다. 제안된 클러스터링은 뉴로-퍼지 모델에서의 Gaussian 소속함수와 같으므로 변환없이 직접 모델을 구성하고 평가하는데 이용한다. 또한 클러스터링의 개념을 유지하기 위하여 비선형 공간을 선형 부분공간으로 변형하기 쉬운 Takagi-Sugeno-Kang (TSK) 퍼지 모델을 이용하여 성능을 평가하였다 [2]. 이 경우 모델의 전체부 소속함수 부분은 비선형 공간으로 고려할 수 있으며 결론부는 1차 선형방정식 형태로 구성되어 선형 변환으로 고려할 수 있다. 그림 1에서 제안된 방법의 개념을 나타내었다.

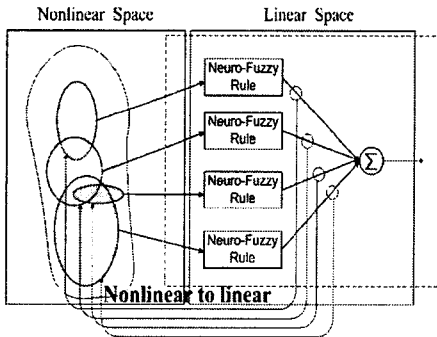


Fig. 1 Proposed scheme of the learning

모델의 추론 형태는 다음과 같이 언어적 입력(전제부)에 대하여 1차 선형방정식 형태의 출력(결론부)로 다음과 같이 구성되어 있다.

$$\text{Rule } j: \text{ IF } x \text{ is } A_i \text{ and } y \text{ is } B_j \quad (8)$$

$$\text{THEN } f_i = p_i x + q_i y + r_i$$

이를 그림으로 보면 언어적 입력을 가지는 전제부와 선형방정식의 출력을 가지는 결론부로 다음과 같다.

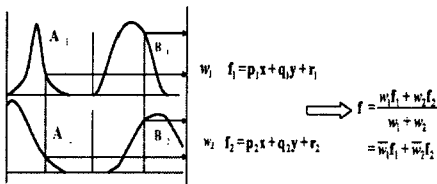


Fig. 2 Concept of the TSK model

또한 클러스터링의 성능을 개선하기 위하여 뉴로-퍼지 모델에서 추정된 정보를 다시 클러스터링 학습에 이용하도록 하였다. 전형적인 비교사학습(Unsupervised Learning)인 클러스터링 학습을 전체 모델의 출력을 이용하여 입력력 인과 관계의 정보를 포함하는 학습으로 구성하여 교사학습(Supervised Learning)으로 확장하여 모델의 성능을 개선하고자 하였다[2][5].

각 클러스터에서 추정된 정보는 뉴로-퍼지 모델의 각 소속함수로 연결되며 입력의 차원에 대해 각 규칙에 일대일 대응이 된다. TSK 퍼지 모델의 5층의 구성 중에 4층에서 각 규칙에 의한 부분 출력이 발생하고 5층에서 가장 평균법에 의한 최종출력을 추정하게 된다. 즉 4층에서의 출력이 각 클러스터에 의하여 추정된 결과를 말한다. 입력이 x, y 의 두 개의 차원을 가질 때 4층의 i 번째 클러스터에 연결된 규칙에 의한 출력은 다음과 같이 추정할 수 있다.

$$\text{Rule } i: \bar{w}_i f_i(x, y) = \bar{w}_i (p_i x + q_i y + r_i) \quad (9)$$

뉴로-퍼지 모델의 부분추정 결과를 주어진 입출력 결과와 비교하여 다음과 같이 각데이터에 대하여 오차를 발생시킨다.

$$e_{j, \text{neurofuzzy}} = z - \bar{w}_i f_i \quad (10)$$

이를 이용하여 결론부 오차 소속함수를 다음과 같이 생성한다.

$$ef_{ij} = \exp \left(-\frac{1}{2} \frac{(e_{j, \text{neurofuzzy}} - \mu_{ei})^2}{\sigma_{ei}^2} \right) \quad (11)$$

이를 클러스터링 유사도 평가에 다음과 같이 적용한다.

$$r_{ij} = r_{ij} \times (ef_{ij} \cdot (1-t) + t) \quad (12)$$

유사도 평가에 출력 정보가 포함됨으로써 입출력을 동시에 고려하는 모델의 학습을 실시할 수 있다. 뉴로-퍼지 모델의 전체부 소속함수는 클러스터링의 결과에 따라 학습을 진행하며 클러스터의 수렴과 함께 학습을 종결하게 된다.

3. 시뮬레이션 및 결과

제안된 기법의 유용성을 대표적인 비선형 시계열 데이터인 Box-Jenkins의 가스르 예제에 대한 결과로써 비교자 한다. 먼저 Jang의 입력선택방법에 의하여 두 개의 입력력을 다음과 같이 선택하였다[2].

$$y(k+1) = f(y(k), u(k-3)) \quad (14)$$

2차원의 입력과 1차원의 출력데이터를 이용하여 임계치와 결론부 변수를 가변하였을 때 규칙의 변화를 그림 3에 나타내었다. 임계치가 0.05에서 0.25까지 급격히 변화하지만 규칙은 일정한 범위 내에서 추정되는 것을 볼 수 있다.

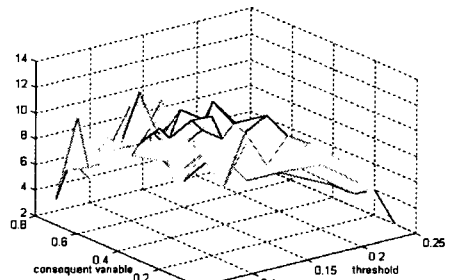


Fig. 3 The distribution of the rules with ξ and t

클러스터링이 학습이 진행되는 동안 뉴로-퍼지 모델은 이들 파라미터를 이용하여 모델을 구성하고 성능을 평가한다. 이러한 학습방법은 오차미분 기반 학습에서 발생하는 소속함수의 국부적 최적화로 수렴하는 문제가 발생

하지 않는다. 클러스터링의 성능이 뉴로-퍼지 모델의 성능과 직접적인 연관이 있으며 클러스터링의 성능을 개선하기 위하여 뉴로-퍼지 모델의 부분출력 정보를 다시 클러스터링 학습에 반영하여 전체 성능을 개선한다. 모델의 학습에 이용된 데이터의 오차뿐만 아니라 검증데이터에 대한 오차 역시 모델링의 성능을 평가하는데 중요한 요소가 된다. 표 1에서 볼 수 있듯이 제안된 방법에 의한 모델링이 학습오차뿐만 아니라 검증오차까지 동시에 감소하는 것을 볼 수 있다.

Table. 1 Comparison of results with previous one using RMSE

Method	Number of rules	Training error	Checking error	Remark
Pedrycz	81	0.320	None	Training only
Xu	25	0.328	None	Training only
Sugeno	2	0.359	None	Training only
Abonyi	16	0.154	None	Training only
Babuska	23	0.124	None	Training only
Jang	4	0.135	0.530	
Oh[9]	4	0.026	0.272	
Proposed $\zeta = 0.1, t = 0.4$	9	0.1282	0.1863	50 epoch
Proposed $\zeta = 0.2, t = 0.4$	5	0.1267	0.1523	100 epoch
Proposed $\zeta = 0.1, t = 0.2$	14	0.1037	0.1697	50 epoch

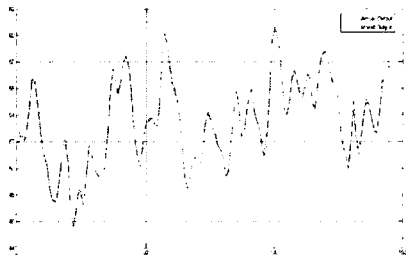


Fig. 4 Comparison with actual output and model output of training data

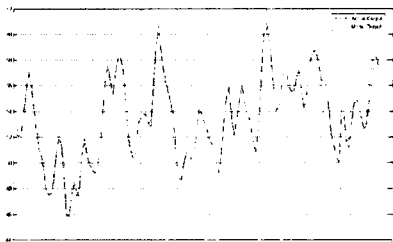


Fig. 5 Comparison with actual output and model output of checking data

그림 4와 그림 5에서 학습데이터에 대한 출력과 검증데이터에 대한 출력을 나타내었다.

3. 결 론

본 논문에서는 클러스터링 학습을 뉴로-퍼지 모델의 학습에 직접 적용시켜 학습을 진행하는 모델을 제안하였다. 모델의 학습과 클러스터링의 학습이 동시에 이루어지면서도 기존 연구들에서 제기된 입력차원에 대한 규칙 증가 문제나 학습방법에 따른 최적해 문제가 가지는 제약들을 해소하면서도 좋은 성능을 가지는 모델을 구성하였다. 사전임계값을 이용하여 자율적인 클러스터링과 모델의 구성을 가지며 클러스터링과 뉴로-퍼지 모델의 학습이나 구조를 유지하면서 상호간의 문제점을 개선할 수 있었다.

향후 연구과제로는 제안된 모델이 가지는 연산량을 더욱 줄일 수 있는 기법을 도입하거나 개발하고 이를 실제 응용에 적용해 성능을 확인하는 것 등이 있다.

[참 고 문 헌]

- [1] Ya Lei Sun and Meng Joo Er, "Hybrid Fuzzy Control of Robotics Systems," IEEE Trans on Fuzzy Systems, Vol. 12, No. 6, pp. 755-765, 2002.
- [2] J-S. R. Jang, C. T. Sun, and E. Mizutani, Neuro-Fuzzy and Soft Computing : A Computational Approach to Learning and Machine Intelligence, Prentice Hall, 1997.
- [3] Kazuo Tanaka and Hua O. Wang, Fuzzy Control Systems Design and Analysis : A Linear Matrix Inequality Approach, John Wiley and Sons Inc, 2001.
- [4] 김승석, 박근창, 유정웅, 전명근, "GMM과 클러스터링 기법에 의한 뉴로-퍼지 시스템 모델링," 한국퍼지및지능시스템학회 논문지, Vol. 12, No. 6, pp. 571-576, 2002.
- [5] 김승석, 김성수, 유정웅, "새로운 클러스터링 알고리즘을 적용한 향상된 뉴로-퍼지 모델링," 대한전기학회 논문지, Vol. 53D, No. 7, pp. 536-543, 2004.
- [6] 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.
- [7] Ching-Chang Wong and Chia-Chong Chen, "A Hybrid Clustering and Gradient Descent Approach for Fuzzy Modeling," IEEE Trans on Systems, Man, and Cybernetics-Part B : Cybernetics, Vol. 29, No.6, pp. 686-693, 1999.
- [8] Jeff Bilms, "A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models," ICSI TR-97-021, April 1998.
- [9] S. R. Jang, "Input Selection for ANFIS Learning," Proceeding of Fifth IEEE International Conference on Fuzzy Systems, Vol. 2, pp. 8-11, 1996.