

EM알고리즘을 기반으로 한 뉴로-퍼지 모델링

김승석, 전병석, 김주식, 유정웅  
 충북대학교 전기 전자 및 컴퓨터 공학부

EM Algorithm based Neuro-Fuzzy Modeling

Seoung Suk Kim, Beung Suk Jun, Ju Sik Kim, Jeoung Woong Ryu  
 Department of Electric and Computer Engineering, Chung Buk National University

**Abstract** - 본 논문은 뉴로-퍼지 시스템에서의 규칙 선택 및 모델 학습에 대하여 EM 알고리즘을 기반으로 하는 구조 동정을 제안한다. 뉴로-퍼지 모델링에서의 초기 파라미터가 학습과정에서의 모델 성능에 큰 영향을 주고 있다. 주어진 데이터에 근거한 파라미터 추정에는 다양한 방법들이 소개되고 응용되어져 왔는데 이전 연구들에서 볼 수 있는 HCM, FCM 등은 데이터와의 유클리디언 거리를 최소화하는 중심점을 파라미터로 선택하는 등의 방법과 퍼지 균등화 등은 데이터의 확률 밀도함수를 이용하여 파라미터를 추정하였다. 제안된 방법에서는 데이터에서의 Maximum Likelihood Estimator를 기반으로 하는 방법으로 EM 알고리즘을 이용하였다. 초기 파라미터의 결정에서 EM 알고리즘을 이용하여 뉴로-퍼지 모델의 전체부 소속함수 파라미터 추정을 실시한다. EM 알고리즘을 이용한 퍼지 모델의 특징으로는 전체부가 클러스터링에 의하여 생성되므로 입력의 차원이나 소속함수의 수가 증가하여도 규칙의 수는 증가하지 않는다. 이를 자동차 MPG 예제를 통하여 제안된 방법의 유용성을 보이고자 한다.

나인 자동차 MPG 문제[9]에 적용하여 유용성을 보이고자 한다.

2. EM 알고리즘을 기반으로 모델링

2.1 EM 알고리즘

Gaussian mixture 모델에서 확률밀도 함수는 식(1) 같이 주어진다.

$$p(x|\Phi) = \sum_{j=1}^n p(x|\omega_j, \Phi)P(\omega_j) \tag{1}$$

여기서  $\omega_j$ 는 각각의 요소들이고,  $P(\omega_j)$ 는 사전 확률이고  $p(x|\omega_j, \Phi)$ 는  $\Phi$ 의 파라미터 벡터이다.  $\Phi$ 는 또한  $P(\omega_j)$ 를 포함한다. 샘플  $x = \{x_k\}_k$ 는 다음 식(2)과 같다.

$$L(\Phi|x) = \sum_k \log p(x_k|\Phi) \\ = \sum_k \log \sum_j p(x_k|\omega_j, \Phi)P(\omega_j) \tag{2}$$

이 식은 직접적으로 문제 해결이 될 수 없다. 이 경우 다대일(many to one) 사상을 통하여 다음과 같이 계산된다. EM 알고리즘은 두 단계를 포함한다. 먼저 E-step에서 이후 확률을 식(3)과 같이 계산한다.

$$P(\omega_j|x_k, \Phi) = \frac{p(x_k|\omega_j, \Phi)P(\omega_j)}{\sum_l p(x_k|\omega_l, \Phi)P(\omega_l)} \equiv h_{kj} \tag{3}$$

만약 요소 밀도가  $d$ 차원 Gaussian 확률이라면  $p(x|\omega_j, \Phi) \sim N_d(\mu_j, \Sigma_j)$ 라고 하면 식(4)와 같다.

$$p(x|\omega_j, \Phi) = \frac{1}{(2\pi)^{d/2}|\Sigma_j|^{1/2}} \\ \times \exp\left[-\frac{1}{2}(x-\mu_j)^T \Sigma_j^{-1}(x-\mu_j)\right] \tag{4}$$

일 때  $g_j \equiv P(\omega_j)$ 이라 하면 이후 확률은 식(5)와 같다.

$$h_{kj} = \frac{g_j |\Sigma_j|^{-1/2} e^{-(1/2)(x_k - \mu_j)^T \Sigma_j^{-1} (x_k - \mu_j)}}{\sum_l g_l |\Sigma_l|^{-1/2} e^{-(1/2)(x_k - \mu_l)^T \Sigma_l^{-1} (x_k - \mu_l)}} \tag{5}$$

다음 M-step에서 요소 파라미터  $\Phi$ 를 식(6)에서 식(8)과 같이 갱신한다.

1. 서 론

신경회로망[1]이나 퍼지 모델[2][3] 등에서 주된 연구 중에 최적의 초기 파라미터 추정을 통하여 모델이 최적의 성능으로 도달하거나 적은 학습으로도 충분한 성능 개선을 이루고자 하는 등의 연구들이 수행되어 왔다. 이러한 인간의 두뇌나 판단 능력을 모사한 모델들에서 좋은 성능은 초기 파라미터에 크게 좌우된다. 또한 퍼지 모델에서 그리드 분할에 의한 규칙의 증가 등은 모델을 생성하거나 응용할 때 문제를 가지고 있었다. 이러한 문제점을 해결하기 위하여 HCM, FCM, Fuzzy Equalization[4] 등 다양한 클러스터링 알고리즘들이 제안되어 파라미터의 초기치 추정에 이용되어 왔다. 이러한 클러스터링 알고리즘 등은 기존의 그리드 분할 형태의 모델에서 볼 수 있는 입력 차원 및 규칙의 증가에 따른 모델 크기의 증가로 인한 문제점들을 해결할 수 있었다. 여기서 제안된 방법은 클러스터링에 의한 방법 중의 하나인 EM 알고리즘[5][6]을 통하여 좀더 확률에 근거한 모델의 초기 파라미터 추정을 통하여 원하는 모델의 성능을 개선하고자 한다. 제안된 EM 알고리즘[7]은 데이터의 기본 분포에서 관측을 통제하는 분포로 다대일(many to one) 사상이 있을 때 파라미터의 Maximum Likelihood Estimation을 통하여 파라미터를 추정한다. 이 알고리즘은 예측 단계와 최대화 단계로 나누어 지며 예측 단계에서는 현재의 파라미터를 추정하고 관측의 필요조건이 되므로 미지의 기본 변수에 관계한다. 또한 최대화 단계에서는 예측 단계의 파라미터를 이용하여 새로운 파라미터를 추정한다. 이 두 단계는 정해진 조건에 만족할 때 까지 수렴한다. 제안된 방법은 초기 뉴로-퍼지 모델[8]에서의 전체부 소속함수의 파라미터와 결론부 파라미터를 추정하여 더 빠른 수렴과 성능을 보여준다. 이를 비선형 예제 중 하

$$\mu_j^{t+1} = \frac{\sum_k h_{kj} x_k}{\sum_k h_{kj}} \quad (6)$$

$$\Sigma_j^{t+1} = \frac{\sum_k (x_k - \mu_j)(x_k - \mu_j)^T}{\sum_k h_{kj}} \quad (7)$$

$$g_j = \frac{1}{n} \sum_k h_{kj} \quad (8)$$

다음 그림1에서 볼 수 있듯이 임의로 생성된 데이터에 대하여 EM 알고리즘을 이용한 파라미터 추정은 각각의 중심과 분산에 대하여 효과적으로 구하는 것을 볼 수 있다.

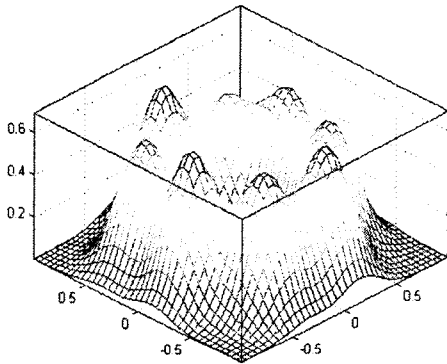


그림 1. EM 알고리즘의 예

반복적인 알고리즘 수행을 그림 2으로 보면 다음과 같다.

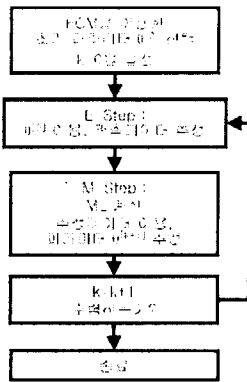


그림 2. EM 알고리즘 순서도

각 알고리즘에서 구하고자 하는 파라미터  $\theta$ 는 식(9)와 같이 추정할 수 있다.

$$Q(\theta|\theta^t) = E[L_c(\theta|x, \theta^t)] \quad (9)$$

알고리즘은 다음과 같은 단계를 반복하여 수행한다.

$$E\text{-step} : Q(\theta|\theta^t) = E[L_c(\theta|x, Z)|x, \theta^t] \quad (10)$$

$$M\text{-step} : \theta^{t+1} = \arg \max_{\theta} Q(\theta|\theta^t) \quad (11)$$

반복적인 알고리즘의 수행에 의하여 Likelihood 함수는 감소하지는 않는다. 이러한 조건에 의하여 감소하지 않고 수렴할 때 까지 증가한다. 그러나 이 경우 국부적인 최적해를 보장할 수 있어도 전역적인 최적해를 보장할 수 없다. 또한 초기치에 의하여 알고리즘의 성능이 좌우된다.

## 2.2 뉴로-퍼지 시스템

제안된 논문에서 사용된 뉴로-퍼지 모델은 언어적 형태의 입력과 1차 선형 방정식 형태를 가지는 TSK 퍼지 모델로 Jang이 제안한 ANFIS(Adaptive Network-Based Fuzzy Inference System)을 이용하였다. 두 개의 입력을 가지는 추론 시스템에 대하여 고려해 보면 다음과 같다.

$$\begin{aligned} R^1 : & \text{IF } x \text{ is } A_1 \text{ and } y \text{ is } B_1, \text{ then } f_1 = p_1 x q_1 y + r_1 \\ R^2 : & \text{IF } x \text{ is } A_2 \text{ and } y \text{ is } B_2, \text{ then } f_2 = p_2 x q_2 y + r_2 \end{aligned} \quad (12)$$

여기서  $x, y$ 는 입력이고  $A_i, B_i$ 는 언어적 형태의 소속 함수이고  $\{p_i, q_i, r_i\}$ 는 결론부 파라미터이다. 이를 각각의 단계로 나누면 다음과 같다.

$$\begin{aligned} \text{단계 1} : & O_{1,i} = \mu_{A_i}(x), \text{ for } i=1,2 \\ & O_{2,i} = \mu_{B_i}(y), \text{ for } i=3,4 \end{aligned}$$

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

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

$$\text{단계 3} : O_{3,i} = \bar{w}_i = \frac{w_i}{w_1 + w_2}, \quad i=1,2$$

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

$$\text{단계 5} : O_{5,1} = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i}$$

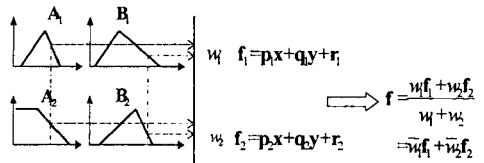


그림 3. TSK 퍼지 모델 추론

이와 같은 적응 네트워크 기반 퍼지 시스템의 결론부 파라미터는 다음과 같은 최소자승법으로 추정한다.

$$X^* = (A^T A)^{-1} A^T B \quad (13)$$

이를 EM 알고리즘에 의하여 추정된 파라미터를 초기 소속함수의 파라미터로 이용하는 뉴로-퍼지 모델의 형태는 다음과 같다.

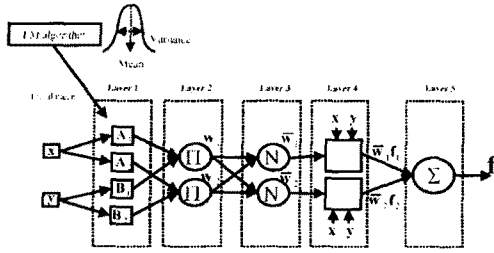


그림 4. 적용된 ANFIS 모델

그림 4에서 볼 수 있듯이 EM 알고리즘으로 추정된 파라미터는 단계 1에서의 초기 소속함수 생성에 주어진다.

### 3. 시뮬레이션 및 결과

제안된 방법의 유용성을 보이기 위하여 자동차 연료 예측 문제(MPG) 데이터에 대하여 제안된 시스템을 적용하였다. 자동차 연료 예측 문제는 6개의 입력과 1개의 출력을 가지며 모두 392개의 데이터 쌍으로 다음과 같이 구성되어 있다.

- No. of cylinders
- Displacement
- Horsepower
- Weight
- Acceleration
- Model year
- Mile per gallon

입력을 모두 사용하는 경우 그리드 분할에 의한 뉴로-퍼지 시스템의 규칙은 지수함수적으로 증가하여 모델의 크기가 증가하여 모델의 생성 및 응용에 문제점을 발생하였다.

제안된 방법은 FCM 알고리즘을 이용하여 주어진 데이터에 대하여 초기 중심값을 추정한다. 추정된 중심값은 EM 알고리즘의 초기 중심값으로 주어지며 각각 E-step과 M-step을 반복 수행한다. 정해진수렴 조건에 의하여 수렴할 때 까지 알고리즘을 반복한다. 또한 EM 알고리즘에 의하여 추정된 중심값과 분산을 뉴로-퍼지 시스템(ANFIS)의 전체부 초기 입력으로 선택하여 모델의 학습에 이용한다. 클러스터링 기법을 이용하여 입력의 차원이나 소속함수의 수가 증가하여도 규칙은 증가하지 않으므로 모델의 크기 및 학습 속도를 줄일 수 있는 장점을 가지고 있다.

시뮬레이션에서는 입력 6개를 모두 사용하여 모델을 생성하였으며 전체 데이터 쌍에 대하여 실시한 결과 및 학습 및 검증 데이터로 나누어 실험한 결과를 비교하였다. 전체 데이터에 대한 추론의 경우 그림 5에서 볼 수 있듯이 제안된 모델의 출력이 실제 데이터의 출력과 근사하게 일치하는 것을 볼 수 있다.

또한 학습 데이터와 검증 데이터로 나누어 시뮬레이션을 실시한 결과 제안된 방법에 의한 결과가 학습 및 검증 데이터에 대하여 같이 근사화 되는 것을 볼 수 있다. 그림 6에서 학습 데이터에 의한 모델의 출력을 나타내었으며, 그림 7에서 검증데이터에 대한 제안된 모델의 출력을 나타내었다. 학습에 편중된 데이터의 경우 검증 데이터에 대하여 적합하지 않는 경우가 발생하는데 반하여 제안된 방법에 의한 모델의 경우 학습 데이터에 대한 오차 및 검증 데이터에 대한 오차가 근사한 크기로 같이 감소함을 볼 수 있다. 즉 제안된 모델이 실제 시스템에 대하여 근사하게 일치함을 볼 수 있다.

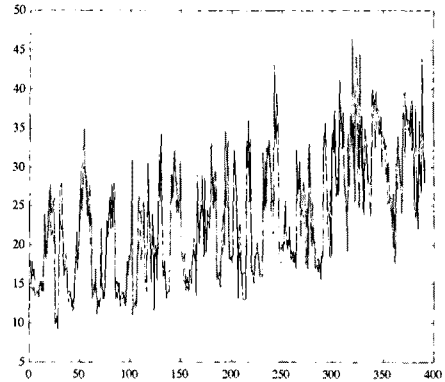


그림 5. 전체 데이터에 대한 모델의 추론

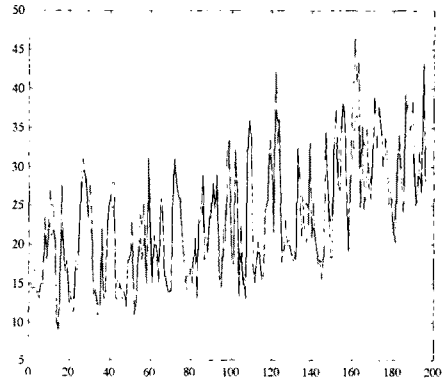


그림 6. 학습데이터에 대한 모델의 추론

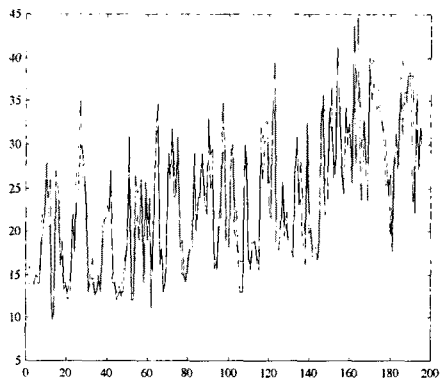


그림 7. 검증 데이터에 대한 모델의 추론  
검증 데이터에 대한 추론

표 1. 제안된 모델의 오차 비교

	학습데이터 오차	검증데이터 오차	비고
Pedrycz[9]	6.16	8.24	FCM
Pedrycz[9]	5.67	6.70	CFCM
제안된 방법	2.82		전체 데이터이용
제안된 방법	2.70	2.93	

표에서 볼 수 있듯이 제안된 방법에 의한 모델의 출력 오차가 이전의 연구들에 비하여 향상됨을 볼 수 있다.

#### 4. 결 론 및 향후 연구과제

제안된 방법에서는 뉴로-퍼지 시스템의 초기치 추정 문제에 대하여 EM 알고리즘을 이용하여 실시하였다. 결과에서 볼 수 있듯이 제안된 방법에 의한 모델 생성에서 모델의 학습 오차 감소와 함께 유사하게 감소하는 검증 오차이다. 일반적으로 학습 데이터에 대한 오차를 감소하는 경우 검증 데이터에 대한 오차는 증가하여 모델의 생성에 대하여 문제를 가지고 있었으나 제안된 방법의 경우 학습 오차 및 검증 오차가 동시에 감소하면서도 오차의 크기가 어느 정도 일정한 비율인 것을 볼 수 있다.

향후 연구과제로서 EM 알고리즘의 국부적 최적점 해결 문제이다. EM 알고리즘은 국부적 최적점은 보장하는 반면에 전역적 최적점에 대한 보장이 없으므로 원하는 최적의 성능을 위하여 전역적 최적점 탐색에 대한 연구가 필요하다. 또한 EM 알고리즘에서의 초기값에 의하여 성능이 크게 좌우되므로 적절한 초기치 선택 방법을 통하여 EM 알고리즘이 성능 개선을 연구 하여야 한다.

#### (참 고 문 헌)

- [1] Simon Haykin, "Neural Networks", Macmillan, 1994
- [2] Li-Xin Wang, "A Course in Fuzzy Systems and Control", Prentice Hall International, Inc, 1997.
- [3] J. S. R. Jang, C. T. Sun, E. Mizutani, "Neuro-Fuzzy and Softing : A Computational Approach to Learning and Machine Intelligence", Prentice Hall, 1997.
- [4] 김승식, 박근창, 유정웅, 전명근, "퍼지군동화와 언어적 Hedge를 이용한 GA기반 순차적 모델링", 한국 퍼지 및 지능 시스템 학회 논문지, Vol. 11, No. 9, pp. 827-832, 2001.
- [5] Richard O. Duda, Peter E. Hart, David G. Stork, "Pattern Classification", John Wiley & Sons Inc, 2001.
- [6] Todd K. Moon, "IEEE Signal Processing Magazine", IEEE, 1996.
- [7] Ethem Alpayd, "Soft vector quantization and the EM algorithm", Neural Networks, Vol. 11, pp. 467-477, Issue 3, April 1998.
- [8] J. S. Jang, "ANFIS : Adaptive-Networks-based Fuzzy Inference System", IEEE Trans. on System, Man and Cybernetics, Vol. 23, No. 3, pp. 665-685, 1993.
- [9] Witold Pedrycz, "Conditional fuzzy clustering in the design of radial basis function neural networks", IEEE Trans on Neural Networks, Vol. 9, No. 4, July, 1998.