변형된 혼합 밀도 네트워크를 이용한 비선형 근사

Nonlinear Approximations Using Modified Mixture Density Networks

조원희*, 박주영**
* Jatco Korea Engineering Co., ** 고려대학교 제어계측공학과

Won-Hee Cho*, Jooyoung Park**
* Jatco Korea Engineering Co.
** Dept. of Control and Instrumentation Engineering, Korea University

E-mail : loadneo@hanmail.net, parkj@korea.ac.kr

요 약

Bishop과 Nabney에 의해 소개된 기존의 혼합 밀도 네트워크(Mixture Density Network)에서는 조건부 확률밀도 함수의 매개변수들(parameters)이 하나의 MLP(multi-layer perceptron)의 출력 벡터로 주어진다. 최근에는 변형된 혼합 밀도 네트워크(Modified Mixture Density Network)라고 하는 이름으로 조건부 확률밀도 함수의 선분포(priors), 조건부 평균(conditional means), 그리고 공분산(covariances) 등이 각각 독립적인 MLP의 출력벡터로 주어지는 경우를 다룬 연구가 보고된 바 있다. 본 논문에서는 조건부 평균이 입력에 관해 선형인 경우를 위한 버전에 대한 이론과 매트랩 프로그램 개발 및 적용을 다룬다.

본 논문에서는 우선 일반적인 혼합 밀도 네트워크에 대해 간단히 설명하고, 혼합 밀도 네트워크의 출력인 다층 퍼셉트론의 매개변수를 각각 다른 다층 퍼셉트론에서 학습시키는 변형된 혼합 밀도 네트워크를 설명한 후, 각각 다른 다층 퍼셉트론을 통해 매개변수를 얻는 것은 동일하나 평균값은 선형함수를 통해 얻는 혼합 밀도 네트워크 버전을 소개한다. 그리고, 모의실험을 통하여 이러한 혼합 밀도 네트워크들의 적용가능성에 대해 알아본다.

키워드 : 가우시안 혼합 모델, 혼합 밀도 네트워크

1. 서론

다층 퍼셉트론(MLP, multi-layer perceptron)과 같은 일반적인 전방향 신경망은 비선형 근사(nonlinear approximation)에 광범위하게 사용되고 있으며 이러한 근사는 일반적으로 목표값(target)의 분산이 상수이고, 목표 값의 조건부 확률밀도함수(conditional pdf)가 일정한 크기의 분산을 갖는 단일 가우시안 형태 (the single gaussian of constant variance)라는 점을 가정한다. 근래에는, Bishop 등에 의하여 일반적인 전방향 신경망과는 달리 입출력 데이터의 연관성을 보다 세밀하게 표현할 수 있는 새로운 형태의 신경망인 혼합 밀도 네트워크(MDN, mixture density network)가 제시된 바 있다.[1-1] 혼합 밀도 네트워크는 비상수(non-constant) 분산이나 가우시안 형태가 아닌 임의의 분포를 사용해 조 모델링할 수 있는 특징이 있다. 그리고, 조건부 확률밀도 함수의 매개변수들(parameters)은 단일형태의 다층 퍼셉트론(multi-layer perceptron)의 출력 벡터 형태로 주어진다. 한편, 최근에는 원래의 MDN을 약간 변형한 형태의 혼합 밀도 네트워크 형태도 제시되었다.[5] 새롭게 제시된 변형 혼합 밀도 네트워크가 갖는 주요 특징은 Bishop 등이 하나의 다층 퍼셉트론을 사용하여 조건부 확률밀도 함수의 변수들을 예측하는데 반하여, 조건부 확률밀도 함수의 선분포(priors), 조건부 평균, 그리고 공분산등을 독립적인 각각의 MLP를 이용하여 예측한다는 점이다. 본 논문에서는 변형된 혼합 밀도 네트워크에서 특별히 조건부 평균이 입력에 관해 선형인 경우를 고려하고 이를 위한 매트랩 프로그램을 작성하였다. 본 논문의 구성은 다음과 같다. 2장에서는 일반적인 혼합 밀도 네트워크에 대해 설명하고 3장에서는 혼합 밀도 네트워크의 출력인 다층 퍼셉트론의 매개변수를 각각

다른 다층 퍼셉트론에서 학습시키는 변형된 혼합 밀도 네트워크와 각각 다른 다층 퍼셉트론을 통해 매개변수를 얻는 것은 동일하나 평균값은 선형함수를 통해 얻는 본 논문의 혼합 밀도 네트워크를 소개한다. 4장에서는 모의실험을 통해 3장에서 소개한 혼합밀도 네트워크를 적용한 결과를 살펴본다. 그리고, 마지막으로 5장에서는 결론을 제시하고 향후 과제에 대해 논의한다.

2. 혼합 밀도 네트워크

일반적으로 다층 퍼셉트론을 학습하는 목적을 데이터 발생기의 통계적인 특성을 모델링하는 관점에서 볼 수 있다. 즉, 조건 밀도 함수 $p(t|x)$의 형태로 표현하는 것을 의미한다. 그러나 data가 다음과 같이 복잡한 구조를 갖게 되면 단일형태로의 분포로는 표현하는데 한계가 있다.[2]

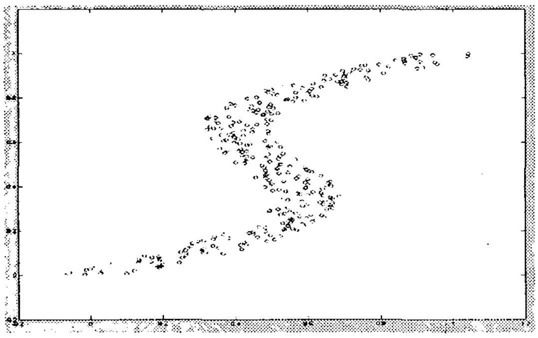

그림 2.1 전형적인 inverse problem [2]

그림 2.1에서 범위 [0.4 0.6]부분에서는 입력이 형태가 다중이다. 따라서, 우리가 이전에 사용했던 단일 가우시안 형태의 분포로는 조건 밀도 함수 $p(t|x)$-에러함수에 해당-를 표현할 수가 없다. 이런 종류의 문제는 inverse problem의 범주에 속하는 문제들에서 자주 발생하며 일반적인 조건 확률밀도(general conditional probability density)의 모델링에 대한 구조가 필요하게 된다. 만약 우리가 혼합 모델 매개변수 Z를 입력 벡터 x의 함수로 설정하면, 우리는 x에 대한 분포 조건에 따라 모델링할 수 있다. 일반적으로, x를 Z(x)에 매핑시키는 함수는 복잡하므로 이러한 함수를 모델링하는데 다층 퍼셉트론을 사용한다. 본 논문에서 사용할 모델의 형태는 다음 식과 같다.

$$p(t|x) = \sum_{j=1}^{M} \alpha_j(x)\,\phi_j(t|x) \qquad (2.1)$$

그리고 커널 함수 식 (2.1)와 같이 가우시안 형태만을 사용하는 것으로 한다.[2]

$$\phi_j(t|x) = \frac{1}{(2\pi)^{c/2}\sigma(x)^c} \exp\left\{ -\frac{\| t-\mu_j(x) \|^2}{2\sigma_j^2} \right\} \qquad (2.2)$$

식 (2.1)에서 M은 혼합(mixture) 안에서의 구성요소의 개수이며 $\alpha_j(x)$는 혼합 계수(mixing coefficient)이고, prior probabilities로 간주된다. 식 (2.2)에서 벡터 $\mu_j(x)$는 j번째 커널의 중심을 의미한다. 이러한 신경망의 조합을 혼합 밀도 네트워크라고 하며 그림 2.2와 같은 구조를 갖게 된다.[2]

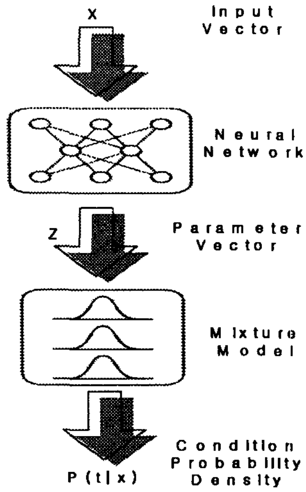

그림 2.2 혼합 밀도 네트워크의 구조

본 논문에서는 혼합밀도 네트워크의 신경망 부분을 몇 가지 형태로 변화시켜 각각의 용도에 적합하도록 성능을 개선하고자 하는 점에 중점을 두었다. 가장 일반적인 형태의 혼합 밀도 네트워크의 신경망부분은 다음의 그림 2.3와 같다.

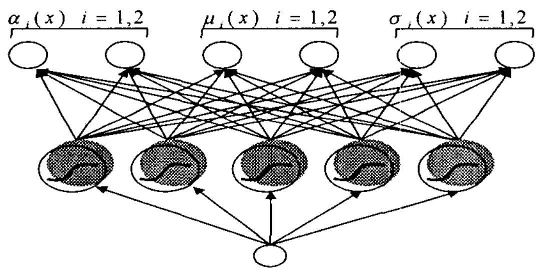

그림 2.3 일반적인 혼합 밀도 네트워크의 신경망 부분

3. 변형된 혼합 밀도 네트워크

3.1 Schittenkopf 등이 제시한 혼합 밀도 네트워크[5]

[5]에서 제시된 변형된 혼합 밀도 네트워크(Modified Mixture Density Network)는, 그림 3.1에서와 같이 혼합 밀도 네트워크의 선분포(priors), 조건부 평균(conditional means), 그리고 공분산(covariances) 등이 각각 독립적인 다층 퍼셉트론의 출력벡터로 주어진다. 이러한 혼합 밀도 네트워크에서는 조건 확률밀도 함수가 다음과 같이 모델링된다.[5]

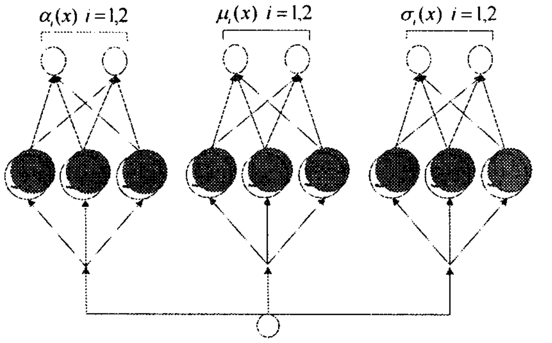

그림 3.1 변형된 혼합 밀도 네트워크의 신경망 부분

$$p(x_t|x_{t-1},\ldots,x_{t-m}) = \sum_{i=1}^{n} \alpha_{i,t}\, g(\mu_{i,t},\ \sigma^2{}_{i,t}) \quad (3.1)$$

$$g(\mu_{i,t}, \sigma_{i,t}^2) = \frac{1}{\sqrt{2\pi\sigma_{i,t}^2}} \exp\left\{ -\frac{(x_t-\mu_{i,t})^2}{2\sigma_{i,t}^2} \right\} \quad (3.2)$$

그리고, 식 (3.1), (3.2)에서 사용된 매개변수 $\alpha_{i,t}$, $\mu_{i,t}$과 $\sigma_{i,t}^2$ 등은 다음과 같이 MLP 신경망을 이용하여 구해진다:

$$\alpha_{i,t} = s(\tilde{\alpha}_{i,t}) = \frac{\exp(\tilde{\alpha}_{i,t})}{\sum_{j=1}^{n}\exp(\tilde{\alpha}_{j,t})} \quad (3.3)$$

$$\tilde{\alpha}_{i,t} = MLP_{1,i}(x_{t-1},\ldots,x_{t-m}) \quad (3.4)$$

$$\mu_{i,t} = MLP_{2,i}(x_{t-1},\ldots,x_{t-m}) \quad (3.5)$$

$$\sigma_{i,t}^2 = (MLP_{3,i}(x_{t-1},\ldots,x_{t-m}))^2 \quad (3.6)$$

여기에서, 식 (3.3)의 소프트맥스(softmax) 함수 $s(\tilde{\alpha}_{i,t})$는 하중(weight) $\alpha_{i,t}$가 양이 되고 그 합이 1이 되도록 한다. 그리고, 식 (3.6)의 $\sigma_{i,t}^2$는 분산값이 항상 양수가 됨을 보장하게 된다 결과적으로 각각의 다층 퍼셉트론은 동일한 m차원의 입력 $x_{t-1},\ldots,x_{t-m}$을 받고 서로 다른 n차원의 출력을 한다. n의 개수는 가우시안 구성요소의 개수와 동일하다. 각각의 MLP는 다음과 같은 입출력 관계식을 갖는다:

$$\begin{aligned} & MLP_i(x_{t-1},\ldots,x_{t-m}) \\ &= \sum_{j=1}^{h} v_{ij}\tanh\Big(\sum_{k=1}^{m} w_{jk}\, x_{t-k} + c_j\Big) + b_i \end{aligned} \quad (3.7)$$

여기에서 i는 출력 뉴런의 색인이며, h는 은닉 뉴런의 개수, w_{jk}와 v_{ij}는 첫 번째와 두 번째 층의 하중(weight)이며, c_j와 b_i는 첫 번째와 두 번째 층의 바이어스이다. 혼합 밀도 네트웍의 매개변수 값들은 식 (3.8)과 같이 우도(likelihood) 함수에 음의 로그를 취한 성능지수를 scaled conjugate gradient 기법이나 conjugate gradient 기법 등의 비선형 최적화 풀이법을 사용하여 최소화하여 최종해를 얻을 수 있다.

$$L = -\frac{1}{N}\log \Pi_{t=m+1}^{m+N}\ p(x_t|x_{t-1},\ldots,x_{t-m}) \quad (3.8)$$

3.2 조건부 평균이 선형 모델인 혼합 밀도 네트워크

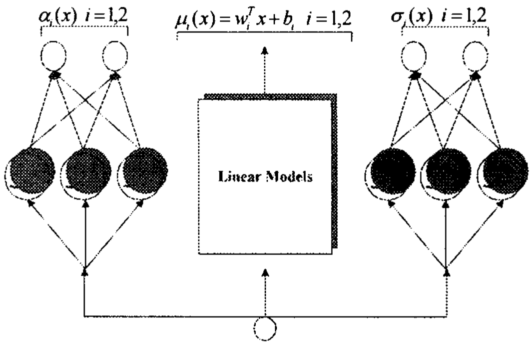

그림 3.2 본 논문에서 고려하는 혼합 밀도 네트워크의 신경망 부분

본 논문에서 고려하는 혼합 밀도 네트워크(improved mixture density network)는 선분포(priors), 조건부 평균(conditional means), 그리고 공분산(covariances) 등이 각각 독립적인 다층 퍼셉트론의 출력벡터로 주어진다는 점에서 변형된 혼합 밀도 네트워크와 유사한 성질을 가지고 있다. [5]의 MDN과 다른 점은 조건부 평균이 입력에 관하여 선형인 경우를 고려한다는 점이다. 이것은 Takagi-Sugeno 퍼지모델처럼 국소적 선형 모델의 혼합으로 전체 매핑을 표현하는 종류의 문제 해결에 유용할 것으로 예상된다. 고려된 혼합 밀도 네트워크에서는 평균값이 다음과 같이 선형 표현을 취하게 됨에 유의하자: $\mu_{i,t} = w_i^T x_t + b$ (3.9)

4. 모의 실험

모의실험에서는 다음의 두 식에 의해 발생된 데이터가 섞여있는 상황을 고려한다.

$$x = 0.015t + 0.27 + error_1 \quad (4.1)$$

$$x = -0.02t + 0.285 + error_2 \quad (4.2)$$

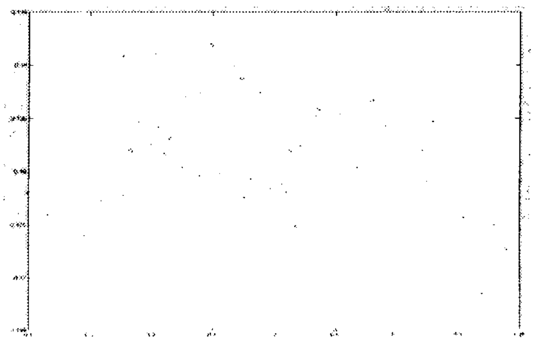

그림 4.1 모의실험에서 사용된 데이터

그림 4.1의 데이터를 본 논문에서 고려한 혼합 밀도 네트워크를 이용하여 발견한 각 모드에 대한 정보가 그림 4.2에서 보여졌다.

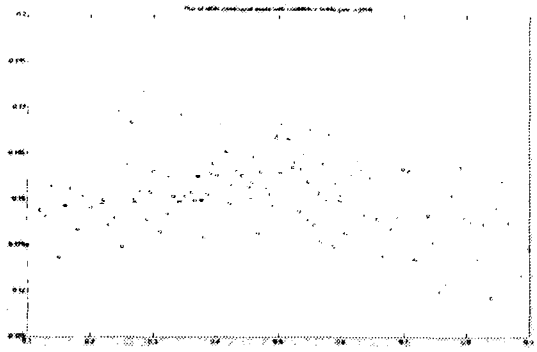

그림 4.2 본 논문의 혼합 밀도 네트워크를 구현한 매트랩 프로그램이 생성한 결과

위의 그림에 따르면 본 논문의 방법론은 모의 실험의 데이터의 내재적 특성을 잘 파악할 수 있는 결과를 제시함을 알 수 있다. 참고로, 전체적인 조건밀도함수와 각 component likelihood가 각각 그림 4.3과 4.4에서 보여졌다.

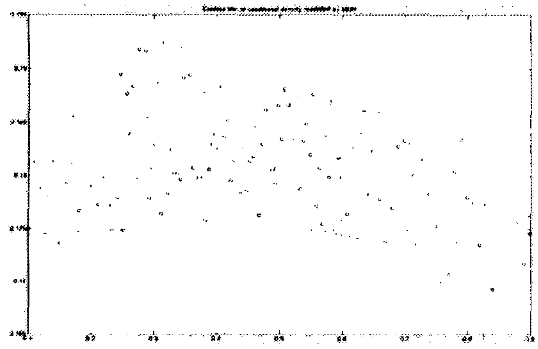

그림 4.3 MDN에 의해 구해진 조건 밀도

그리고, 성능 비교를 위하여 변형 혼합 밀도 네트워크를 적용한 결과도 그림 4.5에서 보여졌다. 그림 4.2와 4.5의 비교로부터 본 논문에서 고려한 방법론이 의미 있는 결과를 생성하였음을 관찰할 수 있다.

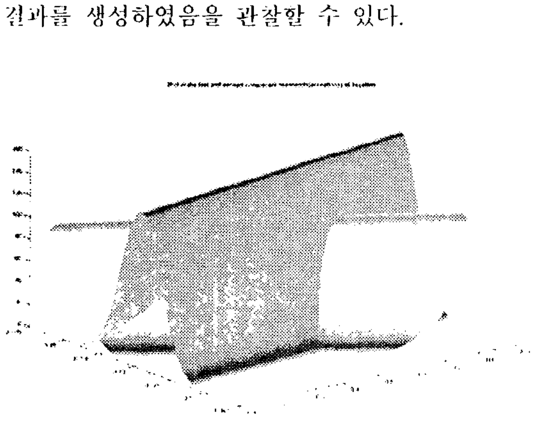

그림 4.4 각 component likelihood의 도식

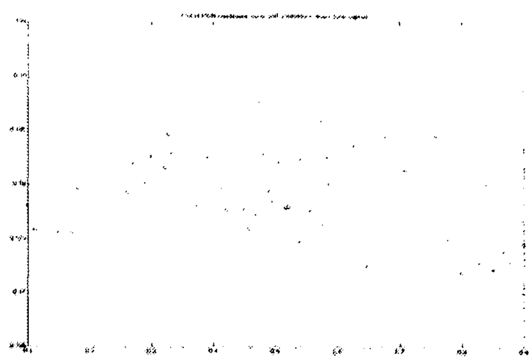

그림 4.5 변형 혼합 밀도 네트워크를 구현한 매트랩 프로그램이 생성한 결과

5. 결론 및 향후 과제

Bishop의 혼합 밀도 네트워크는 모든 파라메터의 추정을 하나의 다층 퍼셉트론만을 사용하여 학습을 하는 방법을 취하였다. 본 논문에서 고려한 혼합 밀도 네트워크는 [5]에서와 같이 매개변수 각각에 대해 분리된 다층 퍼셉트론을 사용하는 방법을 채용하고 출력이 평균인 곳의 다층 퍼셉트론 부분을 선형모델로 대치하는 방안을 고려하였다. 이것은 주어진 데이터가 선형 모드의 혼합인 경우에 데이터의 원형복원에 더 우세하고 본래 함수의 함수식을 찾는 것이 가능하다는 장점이 있기 때문이다. 또한, 이러한 성질은 Takagi-Sugeno 퍼지모델의 identification과 같은 종류의 문제 해결에도 활용할 여지가 있을 것으로 예상된다. 향후과제로는 현재의 학습 데이터보다 더 많은 개수와 더 복잡한 양상의 띠는 데이터에 대한 분석 등을 들 수 있다.

참고문헌

[1] Christopher M. Bishop, Mixture Density Network, Neural Computing Research Group Report, 1994.

[2] Ian T. Nabney, NETLAB Algorithms for Pattern Recognition, Springer, 2001.

[3] Christopher M. Bishop, Neural Network for Pattern Recognition, Oxford University Press, 1999.

[4] Perry Moerland, Mixture Models for Unsupervised and Supervised Learning, Ph.D. thesis, Computer Science Department of the Swiss *Federal Institute of Technology Lausanne*, 2000.

[5] Christian Schittenkopf, Georg Dorffner, Engelbert J. Dockner, Identifying Stochastic Processes with Mixture Density Networks, "Adaptive Information Systems and Modelling in Economics and Management Science" Working Paper No.11, 1998.