어닐링에 의한 Hierarchical Mixtures of Experts를 이용한 시계열 예측

유정수*, 이원돈**

*전주교육대학교 실과교육과
**충남대학교 컴퓨터 과학과

Prediction of Time Series Using Hierarchical Mixtures of Experts Through an Annealing

Jeong Su Yu and Won Don Lee
Dept. of Practical Arts Education, Chonju National University of Education
Dept. of Computer Science, Chungnam National University

Abstract

In the original mixtures of experts framework, the parameters of the network are determined by gradient descent, which is naturally slow. In [2], the Expectation-Maximization(EM) algorithm is used instead, to obtain the network parameters, resulting in substantially reduced training times. This paper presents the new EM algorithm for prediction. We show that an efficient training algorithm may be derived for the HME networks. To verify the utility of the algorithm we look at specific examples in time series prediction. The application of the new EM algorithm to time series prediction has been quite successful.

1. 서 론

사람들은 미래에 일어날 일들을 예측하기를 원한다. 미래에 일어날 일을 미리 예측하여 대비하기 위해서는 어떤 현상의 과거 결과를 측정해야한다. 측정된 현상을 설명할 수 있는 수학적 모델을 미리 알고 있다면 예측은 쉬운 일이 될 것이다. 그러나, 현상을 설명할 수 있는 모델이 알려지지 않았거나 모델 그 자체가 불완전하다면 현상의 과거 결과들을 가지고 미래를 예측하기란 쉬운 일이 아니다.

일정한 시간 간격에서 측정된 관찰 변수들의 집합을 가지고 미래의 행위를 예측 할 수 없는 것을 시계열(time series)이라 하는데 이러한 시계열(time series) 분석과 관련된 문제들은 매우 오랫동안 존재해왔다. 예를 들어, 흑점 발생, 생산라인에서의 불량률, 전력 소비량, 통신 기술 및 제어 이론 등과 같은 과학과 공학 분야 등에서의 많은 시계열 문제들은 어느 정도까지는 예측이 가능하나 언제나 예측의 정확도를 향상시키는 것이 중요한 문제가 되었다. 따라서, 정확하게 예측을 하기 위한 효율적인 방법들이 요구된다. 시계열의 행위는 종종 확률 법칙을 통해서 계열을 설명함으로써 예상한다.

일반적으로 시계열 예측 문제들은 통계적 방법으로 접근되었으며 최근에는 신경회로망으로 접근하여 좋은 예측률을 보이고 있다.

신경회로망(Neural network)에 관한 연구가 활발히 이루어지면서 상황에 따라 변화는 환경에 잘 적응하는 모델로 널리 알려졌다. 신경회로망은 시간에 따라 변하는 사상을 실현하는 문제를 일반적으로 순차 학습 문제와 같이 다루고 있다.

다층 퍼셉트론(MLP)과 같은 신경회로망은 대변동의 간섭을 보여주는 것으로 잘 알려져 있다. 즉, MLP는 새로운 사상에 접하게 됐을 때 전에 학습된 사상은 빠르게 잊어버리는 경향이 있다. 대변동의 간섭은 주로 MLP에 의해서 보여진다. 왜냐하면 이와 같은 회로망들은 학습 과정에서 매우 높게 연결된 전역 모수들(global parameters)을 변경한다. 여러 연구에서 다층 퍼셉트론 회로망의 연결 또는 비학습을 감소시킴으로써 또는 국부적(local)인 반응들을 가진 셀 들의 망들을 사용하여 대변동의 간섭을 완화시키려고 노력하였다. 최근에는 mixtures of experts(ME) 네트워크가 대변동의 간섭을 완화시키기 위한 대체 접근 방법으로 제안되었다. ME 모델은 시계열을 예측하는데 적용하여 좋은 예측 결과를 얻고 있다. ME 모델을 일반화한 것이 HME(hierarchical mixtures of experts)로 각 expert는 반복 형태의 ME로 이루어졌다.

본 논문에서는 기존의 방법들 보다 더 시계열 예측의 정확도를 향상시키기 위하여 Jacobs, Jordan, Nowlan과 Hinton[2]에 의해서 개발된 HME 모델을 확장하여 새로운 모델을 개발하고, 어닐링 기법을 적용한 새로운 EM 학습 알고리즘을 제안하였다. 제안된 네트워크를 비선형 시계열을 예측하는데 적용하였다.

본 논문의 구성은 다음과 같다. 2장에서는 확장된 HME의 계산 구조와 새로운 EM 알고리즘에 관해서 설명하고 3장에서는 2장에서 제안한 모델과 학습 알고리즘을 비선형 시계열 문제에 적용하여 기존의 방법들(BP, MLP,TAR)과 비교한다. 그리고 4장에서는 마지막으로 결론을 언급한다.

2. 확장된 HME 구조와 새로운 EM 알고리즘

ME는 컴퓨터 과학 분야에서 일반적으로 사용되는 분리에 의한 해결(divide and conquer)의 원리로 동작하는 강력한 모듈 구조이다[1,2,3]. Expert는 선형 회귀기이며 gate는 비선형성을 예측된 가능한 함수에 첨가한다.
ME 모델에서 비선형 함수들을 표현 할 수 있으나 gate가 입력 공간에서 인접한 expert 영역들간에 선형 경계를 형성하므로 다소 제한적이다. 따라서, 해결 방법은 expert를 더욱 더 비선형적으로 하거나 또는 gate를 더 비선형적으로 하는 것이다. 이 같이 하는 방법 중의 하나는 MLP를 사용하여 비선형성을 experts와 gate에 첨가하는 것이다. 이를 보완한 방법 그들 자신의 ME모델들을 *experts*로 사용한 것이다. 이를 HME(hierarchical mixtures-of-experts) 모델이라 하며 트리 구조를 갖는다. 단말 노드(terminal nodes) 또는 잎들(leaves)을 *experts* 네트워크를 갖고 있고, 비단말 노드(non-terminal nodes)들은 gating 네트워크들을 갖고 있다. Experts 모델은 프로세스의 가장 하위 레벨이며 gate 모델은 experts를 가지고 분해 가능한 프로세스들을 선택한다.

비선형 예측에서는 풍부한 함수적 사상 도메인을 허용한다. 기본적인 예측 구성은 $y(k)=\hat{y}(k)+e(k)$를 유지한다. 모델은 다음과 같은 비선형 autoregression을 근거로 한다.

$$y(k)=g(y(k-1),y(k-2),\ldots,y(k-M))+e(k) \quad (1)$$

여기서 g는 비선형 함수이다. 이같은 모델은 스칼라와 벡터 결과에 동일하게 적용할 수 있다. 단일 단계 예측은 다음과 같다.

$$\hat{y}(k)=g(y(k-1),y(k-2),\ldots,y(k-M)) \\ =y(k)-e(k) \quad (2)$$

다른 모델(즉, 상태-공간)상에서의 비선형 autoregression 사용은 chaos 이론에서 수행된 작업으로부터 기인된다. 비선형 예측에서 다른 방법은 어떻게 함수 g를 실현하느냐에 따라 특성화된다. DBF(Radial basis function)은 국부적인 선형을 k-nearest 이웃들을 가지고 적합(fit)한다. 다층 네트워크는 시간 윈도우 $y(k-1)$에서 $y(k-M)$까지 대응되는 네트워크에 입력을 하는데 사용된다. 신경회로망을 사용한 예측은 점차로 일반적이 되어가고 있다. 본 논문에서는 g를 그림1 네트워크를 가지고 실현하였다.

$$\hat{y}(k)=N_M(y(k-1)) \quad (3)$$

여기서 N_M은 전체 메모리 길이가 M인 네트워크이다.

그림 1은 확장된 HME 네트워크에 대한 기본적인 예측자 훈련 형태를 설명한 것이다. 각 시간 단계에서 확장된 HME 네트워크의 입력은 알려진 값 y(k-1)이고 출력 $\hat{y}(k)=N_M(y(k-1))$은 실제 계열 값 y(k)의 단일 단계 추정이다. 훈련하는 동안 제곱 오류 $e(k)^2=(y(k)-\hat{y}(k))^2$는 네트워크를 적응하기 위한 one-variable 시뮬레이티드 어닐링 알고리즘을 사용하여 최소화한다. 제안된 네트워크는 open-loop 적응을 수행한다. 즉, 입력과 출력은 알려진 훈련 계열로부터 제공됨을 의미한다. 네트워크의 실제 출력은 훈련하는 동안 입력으로 피드백 되지 않는다.

본 논문에서 제안한 확장된 HME 구조의 학습을 위해서 어닐링 EM 알고리즘을 제안하였다. 제안된 어닐링 EM 알고리즘은 다음과 같다.

어닐링 EM 알고리즘
모든 Example에 대해서
단계 1 모든 단말 노드들의 segmentation을 평균값으로 초기화한다. 온도 T를 설정한다.
단계 2 각 단말 노드에 대하여 루트 노드에서 아래로 recursion 계산하여 트리를 통하여 사후(posterior)를 전파한다.
단계 3 각 expert와 gate의 기울기를 계산한다.
단계 4 expert와 gate 네트워크의 모수들을 가지고 최대화된 우도(likelihood) 구할 때까지 다음 단계들을 반복한다.
① 우도 계산
② expert와 gate 네트워크의 모수들을 perturb
③ $\Delta L=L_{새로운값}-L_{과거값}$ 계산
④ $\Delta L\geq 0$ 이면 P(accept)=1
$\Delta L>0$ 이면 P(accept)= $e^{(-\Delta L/T)}$
⑤ accept된 값들의 평균을 구하여 새로운 값으로 받아들임.
단계 5 만일 온도 T가 최종 온도에 도달했으면 학습을 끝내고 그렇지 않으면 온도를 감소하여 단계 4를 수행한다.

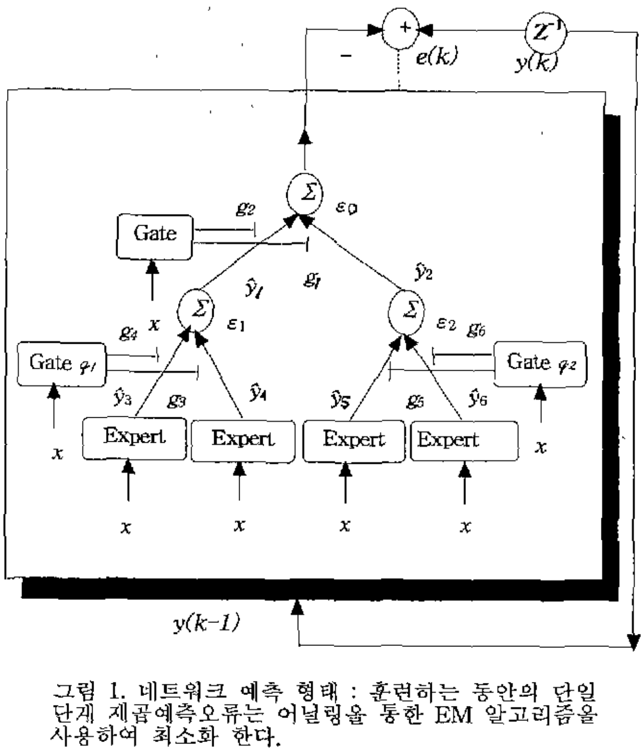

그림 1. 네트워크 예측 형태 : 훈련하는 동안의 단일 단계 제곱예측오류는 어닐링을 통한 EM 알고리즘을 사용하여 최소화 한다.

그림 1은 확장된 HME 네트워크에 대한 기본적인 예측자 훈련 형태를 설명한 것이다. 각 시간 단계에서 확장된 HME 네트워크의 입력은 알려진 값 y(k-1)이고 출력 $\hat{y}(k)=N_M(y(k-1))$은 실제 계열 값 $y(k)$의 단일 단계 추정이다. 훈련하는 동안 제곱 오류 $e(k)^2=(y(k)-\hat{y}(k))^2$는 네트워크를 적응하기 위한

한다. 즉, 입력과 출력은 알려진 훈련 계열로부터 제공됨을 의미한다. 네트워크의 실제 출력은 훈련하는 동안 입력으로 피드백 되지 않는다.

3. 비선형 시계열에 적용

첫 번째 예로 다음과 같은 Mackey-Glass 지연-미분 방정식을 가지고 실험하였다.

$$\frac{dx(t)}{dt} = -1.0x(t) + \frac{0.2x(t-\Delta)}{1+x(t-\Delta)^{10}}$$

이는 호흡기 질환과 조혈 병의 모델이다. 지연(delay)은 Δ=17과 30으로 하였다. 초기 조건 $x(t)=0.9(0 \le t \le \Delta)$이고 표본률 $\tau=6$이다. 이 같은 모수들은 이전의 연구 결과들과 비교가 용이하게 하기 위해서 선택하였다. Δ=30에 대한 시계열은 그림 2(4차 Runge-Kutta 기법이 시스템을 시뮬레이션하기 위해서 사용)와 같다.

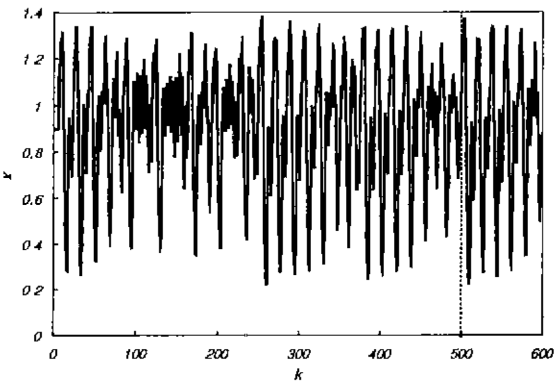

그림 2. Mackey-Glass(30) 시계열

실험은 시계열에서 4개의 자료점, t-18,t-12,t-6 그리고 t가 주어지면 계열에서 다음에 발생할 점의 값을 예측한다. 다른 예측 점들의 수는 t+85, t+84 그리고 t+6이다. 훈련자료는 t=200에서 t=3200이고 예측자료는 t=5000에서 5500이다. 표 1은 기존의 방법과 본 논문에서 제안한 알고리즘을 비교한 것이다. 표의 값들은 5번 실행한 결과의 평균이다.

표 1. Mackey-Glass 결과

신경망	중간층 노드수	훈련률	훈련 NMSE 평균	예측 NMSE 평균
	1,6,12,18 지연된 입력 윈도우 사용			
BP	20	0.005	0.00870	0.08800
BP	25	0.005	0.02200	0.02250
BP	30	0.005	0.01540	0.01580
HME	깊이=3		0.00500	0.00500

두 번째 실험은 시계열 예측의 benchmark 중의 하나인 평균 연 흑점자료(흑점 수를 근사하게 센다)를 사용하였다. 훈련 집합은 1700년도에서 1920년도 자료이고 예측 집합은 1921년부터 1979년 자료를 사용하였다. 예측 기간 전체에 대한 오류를 계산하기 위해서 편의상 초기 기간(1921년부터 1955년까지)과 후기 기간(1956년부터 1979년까지)으로 구분하였다. 입력과 출력 자료에 대해서 모두 0.1에서 0.9사이로 척도화하였다. 예측 결과는 표2와 같다.

표 2. Sunspots 결과

방법	NMSE	
	1700-1920	1960-1979
MLP	0.082	0.35
TAR	0.097	0.28
HME(기존의 EM)	0.061	0.27
확장된 HME (one-variable 시뮬레이티드 어닐링)	0.0345	0.12

4. 결론

본 논문에서는 새로운 예측 모델과 one-variable 시뮬레이티드 어닐링을 가진 EM 학습 알고리즘을 제안하였다. 제안된 모델을 가지고 비선형 시계열 문제에 적용한 결과 기존의 다른 방법들과 비교한 결과 예측의 정확도를 향상시켰음이 입증되었다.

참고문헌

[1] Bengio, S., Fessant, F. and collobert, D,"Use of modular architecture for time series prediction," Neural Processing Letters 3(2), pp101-106, 1996.
[2] M. I. Jordan and R. A. Jacobs, "Hierarchical Mixtures of Experts and the EM algorithm," Neural Computation, 6:181-214, 1994.
[3] S. R. Woterhouse and A. J. Robinson, "Non-linear Prediction of Acoustic Vectors Using Hierarchical Mixtures of Experts," NIPS 7, pp835-842, 1995.