

시계열 자료의 예측을 위한 베이지안 순환 신경망에 관한 연구

A Study on the Bayesian Recurrent Neural Network for Time Series Prediction

홍찬영, 박정훈, 윤태성, 박진배*

(Chan-Young Hong, Jung-Hoon Park, Tae-Sung Yoon, and Jin-Bae Park)

Abstract : In this paper, the Bayesian recurrent neural network is proposed to predict time series data. A neural network predictor requests proper learning strategy to adjust the network weights, and one needs to prepare for non-linear and non-stationary evolution of network weights. The Bayesian neural network in this paper estimates not the single set of weights but the probability distributions of weights. In other words, the weights vector is set as a state vector of state space method, and its probability distributions are estimated in accordance with the particle filtering process. This approach makes it possible to obtain more exact estimation of the weights. In the aspect of network architecture, it is known that the recurrent feedback structure is superior to the feedforward structure for the problem of time series prediction. Therefore, the recurrent neural network with Bayesian inference, what we call Bayesian recurrent neural network (BRNN), is expected to show higher performance than the normal neural network. To verify the proposed method, the time series data are numerically generated and various kinds of neural network predictor are applied on it in order to be compared. As a result, feedback structure and Bayesian learning are better than feedforward structure and backpropagation learning, respectively. Consequently, it is verified that the Bayesian recurrent neural network shows better a prediction result than the common Bayesian neural network.

Keywords : time series prediction, Bayesian inference, recurrent neural network, Bayesian neural network, Bayesian recurrent neural network

I. 서론

시계열(time series)이란 시간에 따라 순차적으로 관측된 값들의 배열이다. 즉, 주식 시세의 추이와 같은 경제 현상, 인구 증감과 같은 사회 현상, 기상 자료의 변화와 같은 자연 현상, 그리고 전력 부하의 변동과 같은 공학적 현상까지 서로 다른 시점에서 관측된 자료의 집합은 모두 시계열 자료(time series data)의 범주에 포함시킬 수 있다.

엄밀한 의미에서 시계열은 시간에 따라 변하는 확률 변수의 계열인 확률 과정(stochastic process)이다. 그러므로 시간에 따른 관측 자료의 확률적, 통계적 성질들을 분석하여, 이를 바탕으로 해당 확률 현상의 미래 행태를 예측하는 것이 가능하다. 이렇게 시계열의 특성을 파악하여 미래의 패턴(pattern)을 예측하는 방법론이 바로 시계열 분석(time series analysis)이다[1]. 그러나 시계열 자료의 관측값 사이에는 보통 복잡한 상관성과 종속성이 존재하기 때문에 일반적인 통계적 분석법이 잘 적용되지 않는다. 따라서 시계열 자료를 효과적으로 분석하여 보다 정확한 미래 값을 예측하기 위해 수학자 및 경제학자 그리고 공학자들 사이에서 오랜 기간 다양한 연구가 이루어졌다.

시계열 분석을 위한 전통적인 방법은 대상 자료에 대한 통계적인 분석을 기반으로 하는 방법으로, 통계적 기법으로는 자기회귀법, 이동평균법, 자기회귀 이동평균법을 거쳐 이

들을 일반화 시킨 Box-Jenkins법이 대표적인 분석 방법으로서 제안되었다. 이러한 방법들은 비교적 단순하여 널리 쓰이지만, 복잡도가 높은 시계열의 경우에는 보다 진일보한 분석법을 요구한다[2-4]. 시계열 분석을 위한 또 다른 방법 중 하나는 상태 공간 이론의 관점에서 필터링(filtering) 기법으로 접근하는 것이다. 필터링은 매번 새로운 자료 값을 얻을 때마다 시계열을 표현하는 모델에 대한 기존의 지식을 갱신하는 과정을 일컫는다[5].

특히 시계열이 선형성(linearity)을 가진 경우, 칼만 필터(Kalman filter)는 알고리듬이 간단하면서도 최적의 성능을 보이는 것으로 알려져 있다. 최근에는 확률론적 방법에 입각한 베이지안 필터(Bayesian filter)가 비선형적 특성을 가진 시계열의 분석에 적용될 수 있음을 보여주는 연구가 이루어졌다[6-8]. 시계열 분석의 방법론에 있어서 획기적인 변화는 인공지능(artificial intelligence, AI)형 기법이 도입되면서 시작되었다. 퍼지(fuzzy), 인공신경망(artificial neural network), 전문가 시스템(expert system)과 같은 인공지능형 기법은 통계적 수식에 의존하지 않고 주어진 자료를 표현하는 일종의 규칙을 찾아내므로, 비선형성을 가진 시계열을 표현하는 데에도 유리하다. 특히 최근 수년 사이에 이루어진 신경망 이론의 연구 결과들은 시계열 분석에 있어 효과적인 예측 성능을 보여주었다[9-11]. 이러한 대표적인 방법들 이외에도 주파수 영역 분석, 웨이블릿(wavelet) 변환을 이용한 방법 등 다양한 분석법이 제안되고 있다. 그러나 각각의 방법은 적용하려는 시계열 자료의 특성에 따라 장점을 가지기도 하고 단점을 가지기도 한다.

본 논문에서는 베이지안 필터와 신경망 이론의 결합을

* 책임저자(Corresponding Author)

논문접수 : 2004. 10. 25., 채택확정 : 2004. 11. 11.

홍찬영 : 삼성전자(cyong@control.yonsei.ac.kr)

박정훈, 박진배 : 연세대학교 전기전자공학과

(skyzami@control.yonsei.ac.kr/jbpark@control.yonsei.ac.kr)

윤태성 : 창원대학교 전기공학과(tsyoon@sarim.changwon.ac.kr)

통해 시계열 분석 성능의 향상을 모색한다. 베이지안 신경망이라는 이름으로 연구되어 온 이러한 기법에서는, 신경망의 가중치를 베이지안 필터의 상태 변수로 설정하고 이를 추정해 나간다. 이 기법에서는 가중치의 개별 값이 아니라 가중치의 확률 분포(probability distribution)을 구하기 때문에 대상 자료에 대한 확률론적 접근이 가능하다. 또한 기존의 자료를 훈련 집합(training set)과 검증 집합(validation set)으로 나눌 필요 없이 전체 자료를 모두 학습에 이용할 수 있기 때문에 결과적으로 보다 나은 모델을 제시할 수 있다 [12]. 본 논문에서는 이러한 베이지안 신경망 기법을 엘만 순환구조의 신경망에 적용한 베이지안 순환신경망을 새롭게 제안하기로 한다. 순환 구조의 신경망은 시계열 분석에 용이한 구조로 알려져 있으므로 일반적인 구조의 신경망에 비해 예측 성능을 향상시킬 수 있다. 따라서 제안한 베이지안 순환 신경망을 기존의 신경망 및 필터링 기법의 결과와 비교하여 그 효과를 확인하고자 한다.

II. 본론

1. 신경망 이론을 이용한 시계열 자료의 예측

시계열 분석을 위한 가장 최근의 동향은 인공지능적 기법의 도입이다. 퍼지이론, 신경망이론, 전문가 시스템과 같은 인공지능적 기법에서는 시계열 자료에 대한 직접적인 수식적 접근보다는, 해당 시계열 자료를 구성하는 법칙(rule)을 학습하는데 중점을 둔다. 특히 신경망(neural networks) 이론은 시계열 자료의 분석 및 예측에 효과적으로 적용될 수 있음을 보여주는 상당수의 연구가 진행되었다[13-15]. 본 절에서는 신경망이론의 기본 및 시계열 예측에의 적용법에 대해 다룬다.

1.1 신경망과 다층구조 퍼셉트론

McCulloch와 Pitts(1943)는 생물 신경세포(neuron)의 정보 처리 과정을 단순화시켜 수학적인 모형으로 제시하였다. McCulloch-Pitts 뉴론은 (1)과 같이 입력과 가중치 벡터를 곱해서 더한 결과를 하나의 활성화 함수(activation function)에 적용하는 방식으로 계산되며, 그림 1과 같다. 활성화 함수 $g(x)$ 가 특히 (2)와 같을 경우는 임계 요소(threshold unit)라고도 불리며, 이외에도 다양한 형태의 함수를 가질 수 있다.

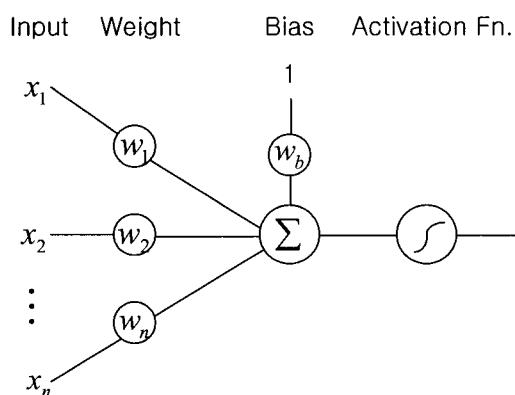


그림 1. 뉴론의 구현.

Fig. 1. Implementation of neuron.

$$y = g(\sum_{i=1}^n w_i x_i - \delta) \quad (1)$$

여기서, w : weight, δ : threshold or offset

$$g(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases} \quad (2)$$

한편 Rosenblatt(1958)가 제안한 퍼셉트론(perceptron)은 최초의 학습기(learning machine)로서, (2)의 임계 함수를 사용하는 단층 뉴런의 망으로 되어있다. 퍼셉트론의 학습규칙은 보통 델타 규칙(delta rule)이라 불리는 단순한 가중치 조정 전략이다. 즉 퍼셉트론이 정확한 결과를 도출하면 가중치를 그대로 두고, 틀린 결과를 도출하면 가중치를 바꾸어주는 것이다. 그러나 이후 Rumelhart(1986) 등에 의해 다층 퍼셉트론(multi layer perceptron, MLP)과 역전파(backpropagation, BP) 학습의 개념이 확립되면서 신경망 이론의 폭넓은 활용이 가능하게 되었다. MLP는 그림 2에서와 같이 입력층과 출력층 사이에 하나 이상의 은닉층을 가지는 신경망이며, 활성화 함수로서 $1/(1 + e^{-x})$ 로 계산되는 시그모이드(sigmoid) 함수나 선형(ramp) 함수를 사용한다. MLP에서와 같이 다수의 뉴런이 전향(feedforward) 다층구조를 이루게 되면 범용 근사자(approximator)로 쓰일 수 있음이 확인되었다.

BP는 평균제곱오차(mean squared error, MSE)와 같은 성능 함수(performance function) P 를 설정하고 그 성능 함수를 최대화하는 출력값을 찾기 위해 기울기 감소법(gradient descent algorithm)으로 가중치를 조정하는 방법론이다. 이는 (3)과 같이 표현되며, 이때 포함되는 학습률 β 는 기울기에 의해 지시된 방향으로 가중값들이 얼마나 조정되느냐를 결정한다.

$$w_i \leftarrow w_i + \beta \frac{\partial P}{\partial w_i} \quad (3)$$

MLP는 최적의 구조를 설정하는 일이 어렵고, BP는 느린 수렴속도 및 오차 함수의 지역 최소점(local minima)에 빠질 수 있다는 점이 각각 단점으로 지적되지만, MLP와 BP는 현재까지 신경망 이론의 기본 바탕이 되고 있다.

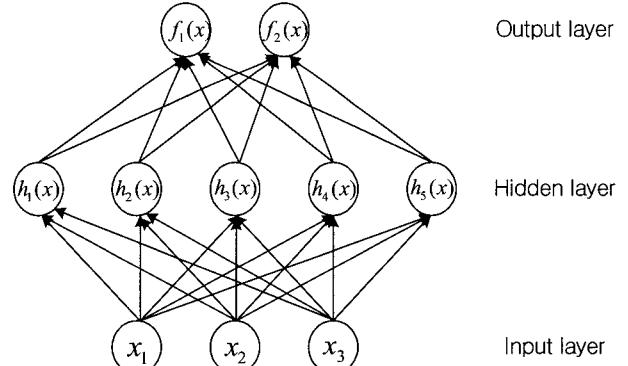


그림 2. MLP의 구조.

Fig. 2. Architecture of MLP.

이러한 신경망 이론을 시계열 분석에 적용하는 대표적인 방법은 $AR(p)$ 모형식을 신경망을 통해 (4)의 비선형식으로 일반화시키는 것이다.

$$y_t = f(y_{t-1}, \dots, y_{t-p}; w) + a_t \quad (4)$$

신경망의 구조를 결정한 후에는, 가장 적절한 가중치의 집합을 구하기 위한 학습과정을 거친다. 이때 사용한 데이터를 훈련 집합과 검증 집합으로 나눈다. 신경망의 학습은 훈련 데이터들만을 사용하여 진행하나, 진행 도중 계속해서 검증 집합에 대한 오차를 측정하여 그 오차가 최소가 되는 지점에서 학습을 멈춤으로써, 신경망이 과다적합되는 것을 방지한다.

1.2 순환구조 퍼셉트론

순환 신경망(recurrent neural network, RNN)은 신경망의 요소(unit)들 사이에 피드백 연결이 있다는 점에서 일반적인 전향구조의 신경망과 다르다. 이러한 특징은 '시간'이라는 인자를 다루기에 적절한 특성이 된다. 그리고 그 외 다른 응용에서의 유용함 때문에 RNN은 1990년대 후반에 들어 널리 연구되었다. RNN의 효시가 되는 기본적인 두개의 모델은 엘만 신경망(Elman network)과 조단 신경망(Jordan network)이다[16]. 이 두 모델은 그림 5에서 보는 바와 같이 기본적인 2층 퍼셉트론 구조에 각각 이전 단계의 은닉층 요소와 출력층 요소를 추가적인 입력으로 첨가한 것이다.

Elman(1990)은 언어에 있어서 구문이나 의미적 특징들의 연속적인 패턴을 학습시키는 것에 관심이 있었고, 이를 위해서는 신경망에서 내부 단계의 기억, 즉 은닉층의 활성화가 필요했다. 그런 이유로 그림 3 (a)에서 오른쪽의 context units는 시간 지연(time delay)을 나타낸다. 다시 말하자면, 시점 $t-1$ 에서의 은닉층 요소가 활성화되어 시간 t 에서의 계산에 추가적인 입력으로써 이용되었다. 엘만 신경망은 (5)와 같이 표현될 수 있다.

$$\hat{y}_t = f(y_{t-1}, \dots, y_{t-p}; u_1(t-1), \dots, u_h(t-1); w) \quad (5)$$

반면 Jordan(1986, 1992)은 이동 로봇의 제어와 학습에 관심을 가지고 있었다. 그 분야에서는 신경망이 로봇의 이전 위치를 기억하는 것이 필요했으므로, 그는 이전 위치들을 신경망의 추가 입력으로 사용하는 구조를 제안하였다. 보다 일반적으로 말하면, 조단 신경망은 일련의 연속적인 출력이 중요한 경우에 특히 적당하다. 조단 신경망은 (6)과 같이 표현된다.

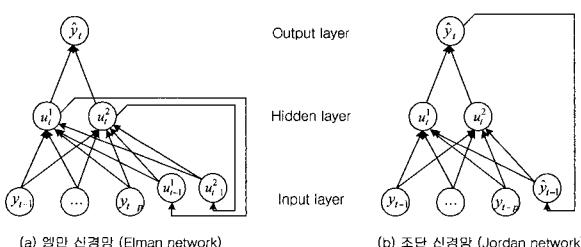


그림 3. 순환 신경망의 예시.

Fig. 3. Examples of recurrent neural networks.

$$\hat{y}_t = f(y_{t-1}, \dots, y_{t-p}; \hat{y}_{t-1}; w) \quad (6)$$

엘만과 조단 신경망을 포함한 순환 구조의 신경망은 적용 환경에 따라 일반적인 전향 구조의 신경망보다 더 좋은 성능을 발휘할 수 있다. 특히 시계열 데이터의 분석과 같은 경우에는 이러한 RNN이 일반적으로 더 유용하다고 알려져 있다. 현재 RNN은 신경망의 여러 분야 중에서 하나의 큰 범주로서 활발히 연구되고 있다.

2. 베이지안 순환 신경망을 통한 시계열 예측

베이지안 신경망은 신경망의 가중치를 상태 공간법의 상태 벡터로 설정하고, 가중치의 추정에 있어 베이지안 추론법을 적용하는 방법이다. 그러나 베이지안 추론의 풀이 과정에는 다차원 적분이 존재하므로 계산이 어렵다는 조건이 붙게 되는데, 여기서는 몬테카를로 적분의 일종인 파티클 필터의 기법을 써서 가중치 벡터를 추정한다. 이는 결과적으로 신경망이론과 파티클 필터의 결합이라 할 수 있다. 본 장에서는 BNN의 이론에 대해 세부적으로 논하고, BNN 모델을 시계열 예측에 효율적으로 응용하기 위해 순환구조의 신경망에 적용한 형태인 베이지안 순환 신경망 모델을 제안한다.

2.1 상태 공간법을 이용한 신경망의 모델링

신경망 이론에서는 새로운 데이터가 들어올 때마다 기존의 가중치(weight)를 갱신함으로써 최적의 모형에 접근해간다. 기존의 일반적인 가중치 갱신 알고리듬으로는 BP가 이용되었지만, 상태 공간법을 이용하여 가중치를 구할 수도 있음이 연구되었다[17]. 즉, 신경망의 가중치 벡터 w_k 를 상태 벡터로 설정하여 (7)과 같은 상태 공간 방정식을 세울 수 있다. 여기서 공정 방정식은 가중치의 변천(evolution)을 표현하며, 측정 방정식은 입력값과 출력값의 비선형 관계를 표현한다.

$$\begin{array}{ll} \text{process equation : } & w_{k+1} = w_k + d_k \\ \text{measurement equation: } & y_k = g(w_k, x_k) + v_k \end{array} \quad (7)$$

여기서 x_k 와 y_k 는 각각 입력과 출력을, w_k 는 신경망의 가중치를 나타내며, 함수 $g(\cdot)$ 는 MLP와 같은 신경망의 구조에 따른 비선형적인 사상(mapping)을 표시한다. 공정 잡음 d_k 는 가중치의 변화 추이에 대한 불확실성을 잡음의 형태로 모델링한 것이며, 측정 잡음 v_k 는 측정시의 오차를 모델링한 것이다. 이 두 잡음항은 평균이 0이고 공분산이 각각 Q_k 와 R_k 인 가우시안 과정이라 가정하며, 가중치나 초기조건과는 상관관계가 없다고 가정한다.

위와 같이 신경망의 가중치 결정을 상태ベkt의 추정 문제로 설정하였을 때 해결해야 할 목표는, 관측된 측정치 $Y_k = \{y_1, y_2, \dots, y_k\}$ 와 상태 공간 방정식 (7)을 통하여 가중치의 추정값 \hat{w}_k 을 찾는 것으로 귀결된다.

이때 함수 $g(\cdot)$ 는 비선형 특성을 가지므로, 베이지안 학습(Bayesian learning), 혹은 베이지안 추론(Bayesian inference)이라 불리는 비선형 상태 방정식의 해법을 적용함으로써 \hat{w}_k 를 찾을 수 있다.

2.2 베이지안 추론을 이용한 신경망 가중치의 추정

Mackay(1992)는 (7)과 같이 신경망의 가중치를 상태방정식의 상태변수로 모델링하고 이를 베이지안 추론법에 의해 해결하는 방법을 제안하여 베이지안 신경망(Bayesian neural network, BNN)이라 칭하였다[18]. 예측과 갱신 과정을 반복하여 상태변수의 cpdf(conditional probability density function)을 구하는 베이지안 추론의 원리를 단계별로 자세히 논하면 다음과 같다. 문제의 설정은 2.1절에서와 같으며, 상태벡터의 초기 pdf $p(w_0|Y_{k-1}) = p(w_0)$ 가 주어졌을 때 현재 시점의 pdf $p(w_k|Y_k)$ 를 구하는 두 단계 과정이다.

2.2.1 예측 단계

예측 단계는 시점 $k-1$ 에서의 pdf $p(w_{k-1}|Y_{k-1})$ 가 주어졌을 때 공정 방정식을 통해 시점 k 에서의 사전 확률분포(prior pdf) $p(w_k|Y_{k-1})$ 를 구하는 과정이며, 유도과정은 다음과 같다.

$$\begin{aligned} p(w_k|Y_{k-1}) &= \frac{p(w_k|Y_{k-1})}{p(Y_{k-1})} \\ &= \int \frac{p(w_k|w_{k-1}, Y_{k-1})}{p(Y_{k-1})} dw_{k-1} \\ &= \int \frac{p(w_k|w_{k-1}, Y_{k-1})}{p(w_{k-1}, Y_{k-1})} \cdot \frac{p(w_{k-1}, Y_{k-1})}{p(Y_{k-1})} dw_{k-1} \\ &= \int p(w_k|w_{k-1})p(w_{k-1}|Y_{k-1}) dw_{k-1} \end{aligned} \quad (8)$$

(8)에서 transition density인 $p(w_k|w_{k-1})$ 에 의해, $p(w_{k-1}|Y_{k-1})$ 가 $p(w_k|Y_{k-1})$ 로 전개되어나감을 알 수 있으며, $p(w_k|w_{k-1})$ 는 다음과 같이 정의된다.

$$\begin{aligned} p(w_k|w_{k-1}) &= \frac{p(w_k, w_{k-1})}{p(w_{k-1})} \\ &= \int \frac{p(w_k, d_{k-1}, w_{k-1})}{p(w_{k-1})} dd_{k-1} \\ &= \int \frac{p(w_k, d_{k-1}, w_{k-1})}{p(d_{k-1}, w_{k-1})} \cdot \frac{p(d_{k-1}, w_{k-1})}{p(w_{k-1})} dd_{k-1} \\ &= \int p(w_k|d_{k-1}, w_{k-1})p(d_{k-1}|w_{k-1}) dd_{k-1} \\ &= \int \delta(w_k - w_{k-1} - d_{k-1})p(d_{k-1}) dd_{k-1} \\ (\because p(d_{k-1}|w_{k-1}) &= p(d_{k-1})) \end{aligned} \quad (9)$$

여기서 디랙 델타 함수 $\delta(\cdot)$ 는 w_{k-1} 과 d_{k-1} 을 알고 있을 때에 공정방정식을 통해 w_k 가 구해짐을 나타내며, $p(d_{k-1}|w_{k-1}) = p(d_{k-1})$ 인 것은 두 잡음항 d_k 와 v_k 가 과거와 현재 상태에 독립적이라는 가정 때문이다.

2.2.2 갱신 단계

갱신 단계는 시점 k 에서 관측치 y_k 가 들어왔을 때 (10)의 베이즈 공식을 통해 사전 확률분포를 갱신하여 사후 확률분포(posterior pdf) $p(w_k|Y_k)$ 를 구하는 과정이다.

$$p(w_k|Y_k) = \frac{p(y_k|w_k)p(w_k|Y_{k-1})}{p(y_k|Y_{k-1})} \quad (10)$$

(10)에서 우도(likelihood) $p(y_k|w_k)$ 는 추정방정식에 의해 다음과 같이 정의된다.

$$\begin{aligned} p(y_k|w_k) &= \frac{p(y_k, w_k)}{p(w_k)} = \int \frac{p(y_k, v_k, w_k)}{p(w_k)} dv_k \\ &= \int \frac{p(y_k, v_k, w_k)}{p(v_k, w_k)} \cdot \frac{p(v_k, w_k)}{p(w_k)} dv_k \\ &= \int p(y_k|v_k, w_k)p(v_k|w_k) dv_k \\ &= \int \delta(y_k - g(w_k, x_k) - v_k)p(v_k)dv_k \end{aligned} \quad (11)$$

또한 (10)의 분모 $p(y_k|Y_{k-1})$ 은 정규화를 위한 것으로 evidence density라고도 불리며, 그 계산은 다음과 같다.

$$\begin{aligned} p(y_k|Y_{k-1}) &= \frac{p(y_k|Y_{k-1})}{p(Y_{k-1})} = \int \frac{p(y_k, w_k, Y_{k-1})}{p(Y_{k-1})} dw_k \\ &= \int \frac{p(y_k, w_k, Y_{k-1})}{p(w_k, Y_{k-1})} \cdot \frac{p(w_k, Y_{k-1})}{p(Y_{k-1})} dw_k \\ &= \int p(y_k|w_k)p(w_k|Y_{k-1}) dw_k \end{aligned} \quad (12)$$

BNN의 가중치 추론 과정은 (8)과 (10)의 예측 및 갱신 과정을 반복적으로 되풀이함으로써 이루어진다. 간단히 정리하면, BNN의 파라미터인 가중치를 결정함에 있어서 가중치의 사전 확률분포를 가지고 있을 때, 관측된 데이터에 의한 우도 함수를 곱하여 사후 확률분포를 계산하고, 그 사후 확률분포의 기대값(mean)으로써 가중치를 추정한다. 이 과정을 그림으로 도시하면 그림 4와 같다. 그러나 BNN의 추론 과정에서 문제가 되는 점은 (8)과 (10)의 예측 및 갱신 과정에서 다차원 적분의 계산이 필요하다는 것이다. 이러한 적분들은 계산하기가 매우 어렵기 때문에 효율적인 적분 계산을 위한 실제적 계산법이 필요하다.

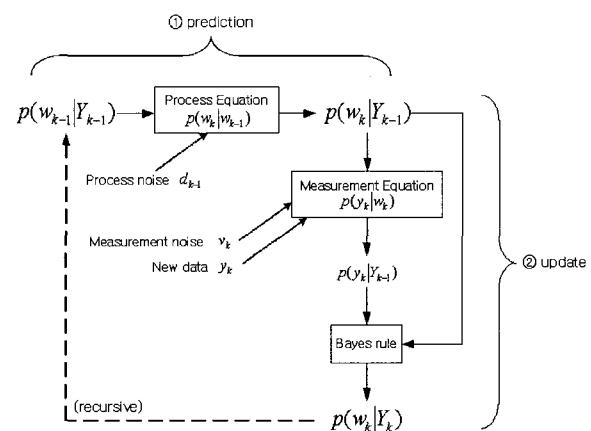


그림 4. 순환 베이지안 추론 : 예측과 갱신.

Fig. 4. Recursive Bayesian inference : prediction and update.

2.3 적분을 위한 실제적인 계산법

다차원 적분의 문제는 베이지안 접근법의 실제 적용에 있어 가장 큰 난점이 되는 부분이다. 따라서 적분의 수치 해석적인 풀이 방법이 필요하며, 이를 위해 직접적 수치적 분법(numerical integration)과 몬테카를로 적분법(Monte Carlo integration) 등과 같은 적분 계산법을 적용할 수 있다[19].

수치적분법은 대상 분포곡선을 유한한 개수의 격자 좌표(grid) 위에 이산화된 분포로 근사시키는 것이다. 각 격자점에서 분포값이 계산되면 나머지 공간에 대해서는 보간법(interpolation)을 이용하여 계산한다. 이러한 방법은 Kitagawa (1987) 등에 의해 이용되었으며, 컴퓨터의 발전에 힘입어 많은 분야에 적용 가능하다. 그러나 다차원 적분의 경우에는 적분 차원이 증가함에 따라 계산량과 복잡도가 급격히 증가 하므로, BNN과 같은 신경망 모델의 계산에는 적합하지 않다. 몬테카를로 적분법은 부정적분을 구하기가 어려운 복잡한 함수의 적분 값을 확률론을 이용하여 구하는 방법으로, 분포 함수로부터 난수(random number)를 생성하여 샘플들을 얻은 후에 확률법칙을 적용함으로써 원하는 정적분 값으로 근사시키는 방법이다. 예를 들어 (13)과 같은 적분 문제를 생각해보자.

$$I = \int_{R^n} g(x) dx \quad (13)$$

(13)의 피적분함수를 $g(x) = f(x)\pi(x)$ 로 적절히 인수 분해하여 (14)와 같은 형태로 표현할 수 있다.

$$I = \int_{R^n} f(x)\pi(x) dx \quad (14)$$

여기서 $i(x) \geq 0$, $\int_{R^n} \pi(x) dx = 1$ (14)에서 $\pi(x)$ 에 대한 가정으로 말미암아 $\pi(x)$ 를 pdf로 해석할 수 있다. 이때 확률분포 $\pi(x)$ 에 따라 분포된 $N(\gg 1)$ 개의 샘플들 $x_{1:N}$ 을 얻어낼 수 있다는 가정 하에서, 그 샘플들을 $f(x)$ 에 대입한 후 그 평균을 취한 것으로서 (13)의 적분값을 근사적으로 대신할 수 있다는 것이 몬테카를로 적분법이다.(15)

$$I \approx \frac{1}{N} \sum_{i=1}^N f(x_i) \quad (15)$$

결국 몬테카를로 적분법은 적분을 대수적인 덧셈의 문제로 변환시키는 기법이다. 이러한 근사법의 정확도는 뽑아낸 샘플들의 수에 의존한다. 샘플의 개수가 무한대에 가까워질 수록 그 샘플들의 정규화된 히스토그램은 정규화된 확률분포, 즉 pdf에 가까워질 것이다.

이때 $\pi(x)$ 의 서포트(support)로부터 $N(\gg 1)$ 개의 샘플들을 뽑아내는 샘플링 전략(sampling strategy)이 필요하며, 대표적으로 acceptance rejection sampling, Markov chain Monte Carlo, importance sampling 등의 방법이 있다. 여기서, 베이지안 기법의 예측 및 갱신 과정을 표현하는 관계식인 (8)과 (10)에 포함되어 있는 적분을 몬테카를로 적분법으로 해결

하는 과정을 통틀어서 순환 몬테카를로 기법(recursive Monte Carlo method)이라고 한다. 이러한 기법에서는, 확률 분포 함수로부터 샘플들을 뽑아내고 각 샘플마다 가중치를 주어 ‘파티클’의 집합을 형성한다. 그리고 새로운 관측치가 들어온에 따라 반복적으로 파티클을 진화시켜 새로운 분포를 가지도록 한다. 이러한 과정상의 특징으로 인해 순환 몬테카를로 기법을 파티클 필터라고도 부른다[20].

결국 순환 몬테카를로 기법은 베이지안 접근법에 가장 성공적으로 적용되어온 importance sampling 방법을 샘플링 전략으로 채택한 몬테카를로 기법이다. 순환 몬테카를로법이 일반적인 몬테카를로법과 다른 점은, 밀도 $\pi(x)$ 가 시간에 따라 변하므로 일정한 형태로 표현되지 않는다는 점이다. 순환 몬테카를로법을 구현함에 있어 본 논문에서는 대표적인 importance sampling 방법 중 하나인 sampling importance resampling(SIR)을 반복적으로 적용한다. 이것을 recursive SIR, 또는 베이지안 부트스트랩(bootstrap) 필터라고도 부른다. 본 절에서는 다음과 같이 recursive SIR의 과정을 따라 적분을 수행한다. 베이지안 과정의 두 단계를 (16)과 (17)에 다시 정리하였다.

갱신 :

$$\pi(w_k | Y_k) = \frac{\pi(y_k | w_k) \pi(w_k | Y_{k-1})}{\pi(y_k | Y_{k-1})} \propto \pi(y_k | w_k) \pi(w_k | Y_{k-1}) \quad (16)$$

예측 :

$$\pi(w_{k+1} | Y_k) = \int \pi(w_{k+1} | w_k) \pi(w_k | Y_k) dw_k \quad (17)$$

Recursive SIR은 우선 $\pi(w_k | Y_{k-1})$ 에서 뽑은 $M (= r \times N)$, r 은 정수)개의 파티클 $w_k^{i=1:M}$ 이 있다는 가정 하에서 시작한다. 즉 $w_k^{i=1:M} \sim \pi(w_k | Y_{k-1})$ 이며, 따라서 몬테카를로 적분에 의해 $\pi(w_k | Y_{k-1}) \approx \frac{1}{M} \sum_{i=1}^M \delta(w_k - w_k^i)$ 이다.

1) 이를 갱신단계인 (16)에 적용하면 SIR 절차에 따라, 우도 $\pi(y_k | w_k)$ 에 비례하도록 w_k 가 N 번 복원 재추출(resample with replacement)되어 새로운 N 개 파티클의 집합 $w_k^{i=1:N}$ 을 얻는다. 새로 얻은 파티클의 집합은 근사적으로 $\pi(w_k | Y_k)$ 을 따른다. : $w_k^{i=1:N} \sim \pi(w_k | Y_k)$

2) 새로 얻어진 $w_k^{i=1:N}$ 를 예측단계인 (17)에 적용하면 몬테카를로 적분에 의해 다음과 같은 결과를 얻는다.

$$\begin{aligned} \pi(w_{k+1} | Y_k) &\approx \frac{1}{N} \sum_{i=1}^N \pi(w_{k+1} | w_k^{i=1:N}) \\ &= \frac{1}{N} \pi(w_{k+1} | w_k^{1:N}) + \pi(w_{k+1} | w_k^{2:N}) + \dots + \pi(w_{k+1} | w_k^{N:N}) \end{aligned}$$

시간 k 를 증가시켜 반복 루프를 계속하기 위해서는 이러한 확률 밀도의 결합으로부터 새로운 샘플들을 뽑아내어야 하며, 이를 위한 직접적이며 간단한 방법은 이 덧셈식의 각 항에서 한 개씩, 또는 N 개씩의 샘플을 생성시키는 것이다.

따라서 새로운 집합 w_{k+1}^M 은 $p(w_{k+1}|w_k^*)$ 로부터 독립적으로 $M (= r \times N)$ 번 추출하여 얻어진 것이며, 그 결과 $p(w_{k+1}|Y_k)$ 을 따른다. : $w_{k+1}^M \sim p(w_{k+1}|Y_k)$

과정 (2) 이후에 과정 (1)로 돌아가 같은 과정을 반복함으로써 recursive SIR의 알고리듬이 계속된다. recursive SIR 알고리듬을 따라 파티클들의 분포가 점차적으로 변해가는 과정을 도시하면 그림 5와 같다[21]. 이때 Q_k 와 R_k 는 분산을 나타낸다. 결과적으로, 2.2절에서 기술한 신경망 가중치 벡터를 베이지안 추론법으로 결정하는 과정 중에 생기는 다차원 적분 문제를, 파티클 필터의 방법을 적용함으로써 해결하였다. 이러한 방법을 통하여 BRNN의 가중치 추정에 대한 수학적인 풀이가 가능해졌으므로, 이제 시계열 예측의 문제에 BRNN을 적용할 수 있다.

2.4 시계열 예측을 위한 베이지안 순환 신경망

신경망 모델은 비선형적 특성이 강한 시계열의 경우에도 다른 방법에 비해 상대적으로 높은 예측 성능을 기대할 수 있다. 또한 신경망 중에서도 학습 모형으로 베이지안 추론을 이용하는 BNN을 이용한다. 최근 BNN을 시계열 자료의 분석에 응용하여 그 성능을 확인하는 연구가 이루어졌다[22-25]. BP를 이용하는 신경망에 대한 BNN의 일반적인 장점은, 가중치의 개별 값이 아니라 가중치의 확률 분포를 구하는 확률론적 방법을 사용함으로써 예측 결과에 신뢰 구간(confidence interval)을 할당할 수 있다는 점이다. 그리고 기존의 자료를 훈련 집합과 검증 집합으로 나눌 필요 없이 전체 자료를 모두 학습에 이용할 수 있기 때문에 결과적으로 보다 나은 모델을 제시할 수 있다.

특히, BNN을 사용함에 있어 본 논문에서는 순환구조(recurrent architecture), 즉 RNN의 구조를 적용한다. 시계열의 예측이라는 목적을 위해서는 일반적인 MLP보다 RNN 구조가 더 우수한 성능을 보이는 것으로 알려져 있기 때문이다. 이러한 베이지안 순환 신경망(Bayesian-recurrent neural network, BRNN)을 그림 6에서와 같이 구성한다.

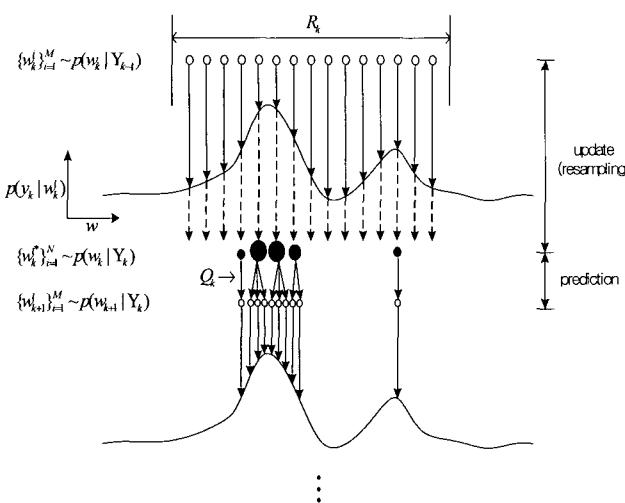


그림 5. Recursive SIR 알고리듬에 따른 파티클 변천 과정.
Fig. 5. Evolution of particles through recursive SIR algorithm.

그림에서 보는 바와 같이 본 논문에서는 은닉층으로부터 추가적인 입력을 취하는 엘만 신경망의 형태로 BRNN을 구성하였다.

제안된 BRNN의 입력층은 과거 시점 p 까지의 관측치 $y_k, y_{k-1}, \dots, y_{k-p}$ 로 구성된다. 또한 본 신경망은 기본적인 한 개의 은닉층을 갖도록 한다. RNN의 구조 설정에 따라 은닉층의 노드(node)는 입력층에 추가적인 입력 요소로 이용된다. 성능 비교를 위하여 입력층과 은닉층을 구성하는 노드의 개수는 실험 상황에 따라 동일하게 설정한다. 신경망의 출력은 미래값 y_{k+1} 을 결과값으로 추정하는 것이므로, 출력층은 한 개의 노드로 되어있다. 이러한 BRNN 시스템의 모든 가중치를 포함하고 있는 가중치 벡터 w 를 각 요소별로 추정함으로써 정확한 출력값을 기대할 수 있다.

3. 실험 및 결과

본 장에서는 제안한 BRNN 모델을 이용하여 시계열 자료를 예측하는 모의실험을 보인다. 제안한 모델의 유용성을 확인하기 위해 수학적 모델을 세워 실험 대상 데이터를 생성하고, 각 신경망 모델의 예측 성능을 비교 검증한다.

3.1 수학적 데이터의 생성과 실험 조건

모의실험을 위해 다음의 (18)과 같은 모델식에 의한 데이터 y_k 를 생성한다. (18)은 비선형성을 내포하고 있으며, 한 주기의 사각파(square wave) 성분을 포함시켜 비정상성도 가지도록 구성하였다.

$$\begin{aligned} x_k &= 0.5x_{k-1} + \frac{25x_{k-1}}{(1+x_{k-1}^2)} + 8 \cos(1.2(k-1)) \\ y_k &= (x_k + 6\text{square}(0.05(k-1)) + 3)/10 \quad (18) \end{aligned}$$

목표 데이터 y_k 의 초기값은 $y_0=1$ 로, 샘플링 간격은 $\Delta k=1$ 로 설정하고, 총 시간은 $k=120$ 까지로 하여 나타내면 그림 7과 같다.

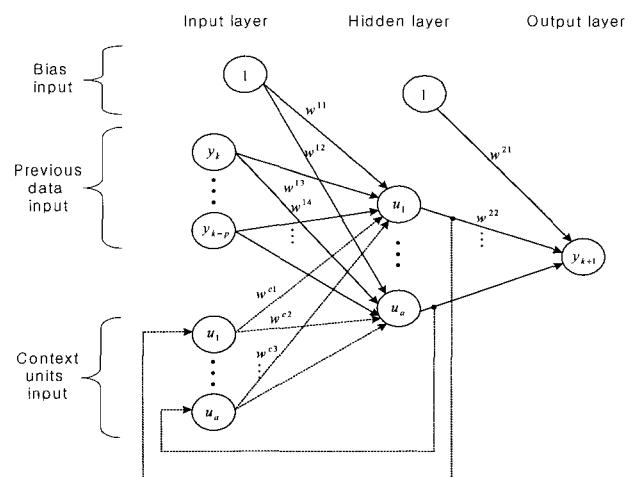


그림 6. 시계열 예측을 위한 베이지안 순환 신경망(BRNN).
Fig. 6. Bayesian recurrent neural network for time series prediction.

주어진 시계열 자료에 대한 예측을 위해 다양한 신경망 모형을 구성한다. 신경망은 과거의 관측치들을 입력으로 하여 그 다음 시점의 값을 예측하는 한 단계 이전 예측 (one-step-ahead prediction)을 수행한다. 신경망의 학습은 관측치를 받아들이면서 바로 신경망 가중치를 수정하는 온라인(online) 학습이다. 신경망의 기본적인 조건은 동일하며, 각 신경망의 예측 정확성을 비교하는 척도로 평균제곱오차 (MSE)를 계산한다. 또한 순환 신경망의 구조는 엘만 신경망의 형태로 하며, 따라서 RNN 및 BRNN의 경우 은닉층 노드의 개수와 같은 수의 context unit가 입력으로 추가된다. 그리고 BNN 및 BRNN의 경우에는 몬테카를로 계산법에서 가중치의 분포를 나타내기 위해 추출하는 샘플의 수를 오십 개로 설정한다. 우선 이러한 구조를 바탕으로 한 BRNN을 이용하여 대상 시계열을 예측하였다. 이때 한 가중치 w_{11} 의 사후 확률밀도함수 $f(w_{11} | Y_k)$ 가 시간에 따라 변천하는 과정을 그림 8에서 확인할 수 있다.

3.2 결과 분석

- 신경망 구조에 따른 비교 : feedforward 신경망과 feedback 신경망

일반적인 MLP 구조(feedforward)의 신경망과 순환 구조(feedback)의 신경망인 RNN일 경우 예측 성능을 비교하였다. 학습방법은 동일하게 BP를 이용하였으며, 비교실험 결과는 그림 9와 10에서 보는 바와 같다. MLP일 경우와 RNN인 경우 MSE는 각각 0.2544와 0.2341로 계산되었으며, 따라서 일반적인 전향 구조의 신경망보다 순환 구조를 가진 신경망이 보다 좋은 예측 성능을 가진다.

- 학습법에 따른 비교 : 가중치에 대한 BP 학습법과 베이지안 학습법

BP 학습법과 베이지안 학습법을 이용하였을 때의 예측 성능을 비교하였다. 이때 신경망의 구조는 동일하게 일반적인 전향 구조의 MLP를 이용하였으며, 비교 실험 결과는 그림 11과 12에서 보는 바와 같다. BP 학습법인 경우와 베이지안 학습법인 경우 MSE는 각각 0.2544와 0.1907로 계산되

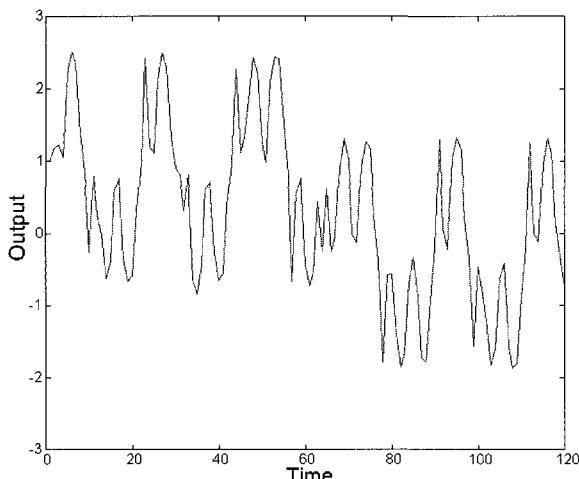


그림 7. 实验数据生成。

Fig. 7. Realization of target data.

었으며, 따라서 일반적인 BP 신경망보다 베이지안 신경망이 보다 좋은 예측 성능을 가진다.

• BNN과 BRNN의 비교

위에서 시행한 두 가지 경우의 실험에서 확인한 바와 같이 구조에서는 RNN이, 학습법에서는 BNN이 각각 우위를 보이는 것을 알 수 있다. 따라서 RNN과 BNN을 결합한 BRNN이 일반적인 BNN보다 예측 성능이 높을 것으로 기대되며, 이를 그림 13과 14에서 확인하였다.

BNN일 경우와 BRNN인 경우 MSE는 각각 0.1907과 0.0828로 계산되었다. 따라서 BNN에 순환 구조를 적용한 BRNN이 보다 좋은 예측성능을 가진다.

일련의 모의실험 결과, 본 논문에서 제안한 BRNN 모델은 비선형, 비정상 특성을 가진 시계열의 경우에도 상대적으로 정확한 예측을 할 수 있음을 확인할 수 있다.

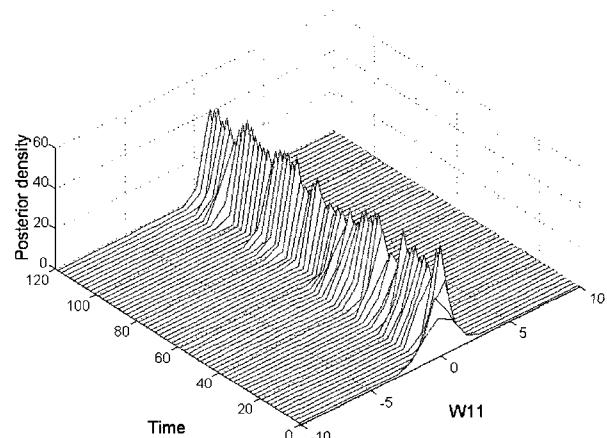


그림 8. 가중치 w_{11} 의 사후 확률밀도함수 변천.

Fig. 8. Evolution of posterior density for weight w_{11} .

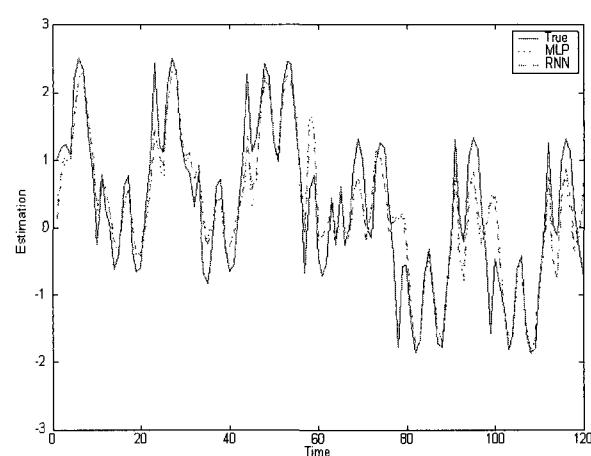


그림 9. Feedforward 신경망과 feedback 신경망의 예측 성능 비교.

Fig. 9. Prediction performance of feedforward and feedback network.

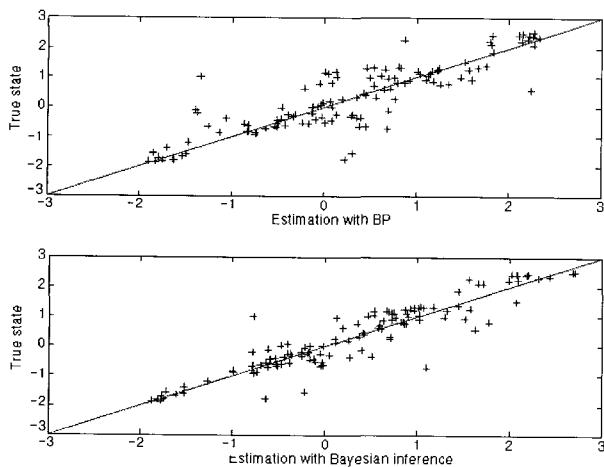


그림 10. MLP와 RNN으로부터 얻은 추정값과 참값의 비교.
Fig. 10. True value versus estimates from MLP and RNN.

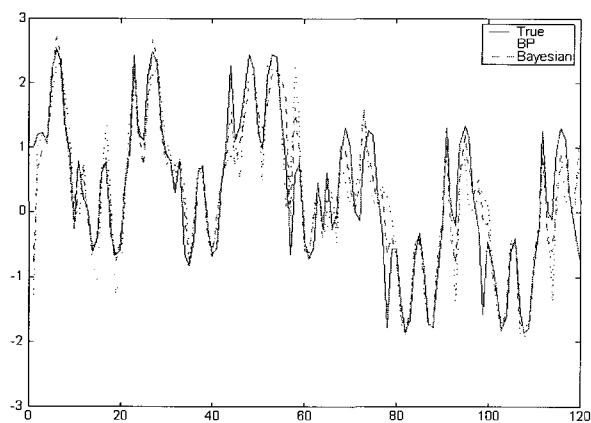


그림 11. BP 학습법과 베이지안 학습법의 예측 성능 비교.
Fig. 11. Prediction performance of BP and Bayesian learning.

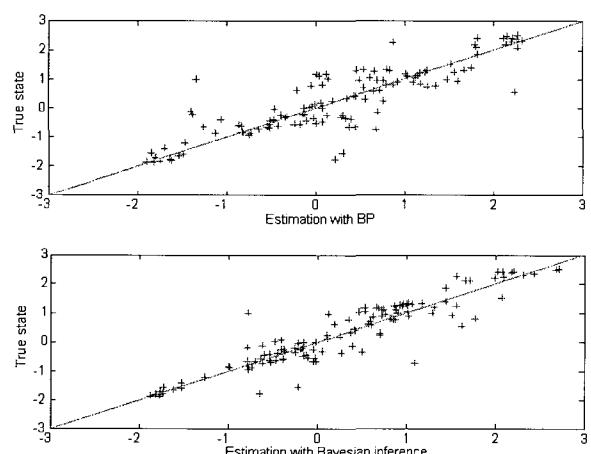


그림 12. BP 학습법과 베이지안 학습법으로부터 얻은 추정값과 참값의 비교.
Fig. 12. True value versus estimates from BP and Bayesian learning.

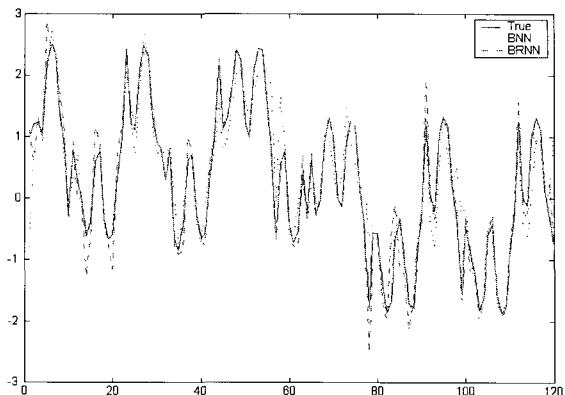


그림 13. BNN과 BRNN의 예측 성능 비교.
Fig. 13. Prediction performance of BNN and BRNN.

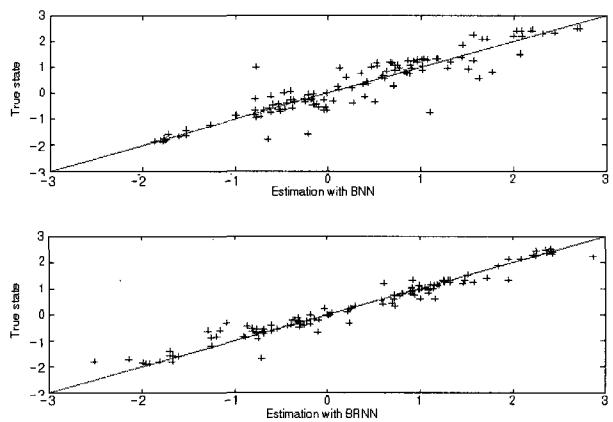


그림 14. BNN과 BRNN으로부터 얻은 추정값과 참값의 비교.
Fig. 14. True value versus estimates from BNN and BRNN.

III. 결론

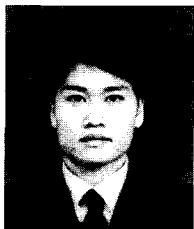
본 논문에서는 시계열 자료의 예측을 위한 베이지안 순환 신경망에 관하여 연구하였다. 통계 이론, 상태 공간법, 신경망 이론과 같은 시계열 예측법 중에서 신경망을 이용한 방법은 복잡한 특성을 가지는 시계열 자료에 보다 효과적이다. 기존의 신경망과는 달리 베이지안 신경망은 가중치의 학습전략으로써 베이지안 추론법을 이용하는 것으로, 이 과정에서 생기는 적분의 계산은 순환 몬테카를로 기법을 통해 해결한다. 이러한 방법을 통해 시간에 따른 가중치의 확률 분포를 얻게 됨으로써 가중치의 값을 보다 정확히 추정할 수 있다. 특히 시계열 자료의 예측 문제에 있어서는 순환 형태의 신경망이 더 효율적이므로, 순환 구조를 가지는 베이지안 신경망을 제안하여 이를 시계열 예측에 적용하였다.

베이지안 순환 신경망의 유용성을 검증하기 위한 모의실험에서는 비선형, 비정상성을 지닌 시계열 자료를 수학적으로 생성하여 이 대상 자료에 대해 한 단계 전향 예측을 시도하였으며, 신경망의 순환 형태로는 은닉층 요소를 추가적인 입력으로 침가하는 엘만 신경망의 구조를 채택하였다. 신경망의 구조에 따른 비교 실험 결과 전향 구조의 MLP보다 순환 구조의 RNN이 더 우수한 예측 성능을 보였다. 또

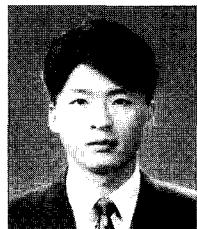
한 학습 방법에 대한 비교 실험 결과 일반적인 BP 학습의 경우보다 베이지안 학습의 경우가 더 우수한 결과를 나타내었다. 이러한 결과들을 바탕으로, 전향 구조인 BNN과 순환 구조인 BRNN의 예측 성능을 각각 비교하여 BRNN이 가장 우수한 예측 성능을 가진다는 것을 확인하였다. 베이지안 순환 신경망은 그 특성상 상대적으로 많은 계산량을 요구하므로 연산 속도가 다소 느리다는 단점이 있으나, 일반적인 신경망에 비해 구조와 학습에 있어서 각각 우수성을 보이므로 정확한 예측 결과가 필요한 경우에 유용하다. 따라서 전력 부하나 환율 변동과 같이 다소 긴 시간의 샘플링 간격으로 얻어지는 시계열의 예측에 응용될 수 있다. 또한 베이지안 신경망의 학습법은 일반성을 지니므로, 본 논문에서 다룬 시계열 자료의 예측 문제에 뿐만 아니라 신경망 이론이 적용될 수 있는 다른 여러 분야에도 적용 가능하다. 이때 해당 베이지안 신경망의 구조를 본 논문에서와 같이 적절한 형태로 변형한다면 보다 향상된 성능을 기대할 수 있다.

참고문헌

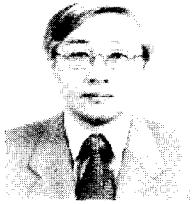
- [1] 김해경, 김태수, 시계열 분석과 예측 이론, 경문사, 2003.
- [2] 이덕기, 예측 방법의 이해, 고려정보산업, 1999.
- [3] C. Chatfield, Time-series forecasting, Chapman & Hall/CRC, 2000.
- [4] P. J. Brockwell and R. A. Davis, Introduction to time series and forecasting, Springer, New York, 2002.
- [5] R. G. Brown and P. Y. C. Hwang, Introduction to random signals and applied Kalman filter, John Wiley & Sons, 1997.
- [6] G. Kitagawa, "Non-gaussian state-space modeling of nonstationary time series", *Journal of the American Statistical Association*, vol. 82, no. 400, pp. 1032-1036, 1987.
- [7] G. C. Franco and R. C. Souza, "A comparison of methods for bootstrapping in the local level model", *Journal of Forecasting*, 21, pp. 27-38, 2002.
- [8] O. A. Alsayegh, "Annual energy consumption prediction using particle filters", *Seventh International Symposium on Signal Processing and Its Applications*, vol. 2, pp. 571-574, 2003.
- [9] A. C. Tsakouris, S. S. Vladov and V. M. Mladenov, "Electric load forecasting with multilayer perceptron and elman neural network", *2002 6th Seminar on Neural Network Applications in Electrical Engineering*, pp. 87-90, 2002.
- [10] J. Zhang, K. S. Tang and K. F. Man, "Recurrent NN model for chaotic time series prediction", *23rd International Conference on Industrial Electronics, Control and Instrumentation*, vol. 3, pp. 1108-1112, 1997.
- [11] M. Crucianu, "Bayesian learning for recurrent neural networks", *Neurocomputing*, 36, pp. 235-242, 2001.
- [12] C. M. Bishop, Neural Networks for Pattern Recognition, Oxford University Press, 1995.
- [13] J. Durbin and S. J. Koopman, Time Series Analysis by State Space Methods, Oxford University Press, 2001.
- [14] N. J. Gordon, D. J. Salmond and A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian bayesian state estimation", *Radar and Signal Processing*, vol. 140, Issue 2, pp. 107-113, 1993.
- [15] T. L. Song, "Filtering theory", *Journal of Control, Automation, and Systems Engineering*, vol. 9, no. 6, pp. 413-419, 2003.
- [16] D. Pena, G. C. Tiao and R. S. Tsay, A course in time series analysis, John Wiley & Sons, 2001.
- [17] V. Kadirkamanathan and M. Niranjan, "A function estimation approach to sequential learning with neural networks", *Neural Computation*, vol. 5, pp. 954-975, 1993.
- [18] D. J. C. Mackay, "A practical bayesian framework for backpropagation networks", *Neural Computation*, vol. 4, no. 3, pp. 448-472, 1992.
- [19] R. M. Neal, Bayesian learning for neural network, Lecture Notes in Statistics No. 118, Springer-Verlag, New York, 1996.
- [20] N. Bergman, Recursive Bayesian estimation : navigation and tracking applications, Linköping University, 1999.
- [21] J. F. D. Freitas, M. Niranjan and A. H. Gee, "Hybrid sequential Monte carlo/kalman methods to train neural networks in non-stationary environments", *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 1057-1060, 1999.
- [22] T. Zhang and A. Fukushige, "Forecasting time series by bayesian neural networks", *Proceedings of the 2002 International Joint Conference on Neural Networks*, vol. 1, pp. 382-387, 2002.
- [23] E. H. Tito, G. Zaverucha, M. Vellasco and M. Pacheco, "Bayesian neural networks for electric load forecasting", *6th International Conference on Neural Information Processing*, vol. 1, pp. 407 - 411, 1999.
- [24] A. Vehtari and J. Lampinen, "Bayesian neural networks for industrial applications", *Proceedings of the 1999 IEEE Midnight-Sun Workshop on Soft Computing Methods in Industrial Applications*, pp. 63-68, 1999.
- [25] M. Crucianu, R. Bone and J.-P. A. De Beauville, "Bayesian learning for time series prediction with exogenous variables", *International Joint Conference on Neural Networks*, vol. 4, pp. 2594-2599, 1999.

**홍찬영**

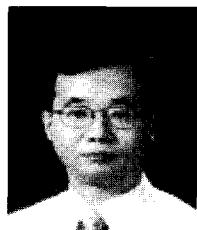
2002년 연세대 기계전자공학부 졸업.
동대학원 석사(2004). 2004년~현재 삼
성전자 근무. 관심분야는 Time-Series
Prediction, Bayesian Recurrent Neural
Network.

**박정훈**

1996년 아주대 전자공학과 졸업. 삼성
전자 근무(1996-1999). 연세대학원
석사(2003). 2003년~현재 동대학원 박
사과정. 관심분야는 Filtering Theory
및 Visual Tracking.

**윤태성**

1978년 연세대 전기공학과 졸업. 동대
학원 석사(1980). 동대학원 박사(1988).
1989년~현재 창원대학교 전기공학과 교
수. 관심분야는 DSP 신호처리 및 응용.

**박진배**

1977년 연세대 전기공학과 졸업. Kansas
State Univ 대학원 석사(1985). 동대학
원 박사(1990). 1992년~현재 연세대 전
기전자공학과 교수. 관심분야는 로보틱
스, 지능제어, 로버스트 제어 및 응용.