

Time Series Prediction using Virtual Term Generation Scheme

Taeho Jo and Sungzoon Cho

Department of Computer Science and Engineering, POSTECH

Abstract

The values measured at different time and enumerated sequentially by homogenous interval is called time series. Its goal is to predict values in future by analysing the measured values in past. The stastical approach to time series prediction tend to be by a neural approach with difficulties in expressing the relationship among past data. In neural approach, the preblem is the acquisition of the enough training data in advance. The goal of this paper is that such problem is solved by generating another term as virtual term between terms in time series.

1. 서 론

시계열(time series)이란 일정 간격으로 서로 다른 시간에 측정된 값들을 시간적 순서로 나열한 집합을 말한다. 이의 목적은 주어진 항(term)의 시간에 따른 변화를 분석하여 미래의 시간에 나타날 측정값을 예측하는 데 있다. 이러한 예측을 시계열 예측(time series prediction)이라 한다.[2]

현재 주어진 과거의 데이터를 이용하여 미래의 값을 예측하는 것은 과거의 값들간에 비선형적(nonlinear)이고, 복잡도(complexity)가 높기 때문에 기존의 통계적 접근으로는 실제 문제의 적용에 있어서 곤란하다. 그리하여 최근에는 시계열 예측에 있어서 신경망을 이용하는 방법이 통계적인 방법을 대체하고 있는 추세이다.[2]

신경망으로 대체할 경우 가장 문제가 되는 것은 충분한 학습 데이터(training data)의 확보이다. 특히 실제의 문제에 있어서는 학습을 하기위한 충분한 데이터를 구하기 어려운 경우가 많다. 학습 데이터가 충분하지 않다면 학습 속도(learning speed)와 예측 성능(prediction performance)이 저하되기 때문이다. 그리하여 이러한 문제를 해결하기 위해 주어진 시계열의 항사이에 또 다른 항, 즉 가상항을 생성함으로써 새로운 시계열을 형성하여 예측 성능을 향상하는 방법을 이 논문에서 제시하겠다.

이를 검증하기 위하여 실험 데이터로는 연도 별 1700년 부터 1979까지의 흑점(sunspot)의 개수를 0에서 1사이로 표준화한 것을 사용한다.[1] 원래의 시계열을 사용한 표준 역전파(standard backpropagation), 연결강도를 제거한 역전파(backpropagation with weight elimination)[1], 그리고 생성된 가상항을 포함한 시계열을 이용한 표준역전파 등을 비교 분석 함으로서 가상항을 사용한 역전파가 연결강도 제거를 이용한 역전파보다 예측 성능이 우수함을 나타내겠다.

2. 가상 항 생성(Virtual Term Generation):전 학습 단계

주어진 시계열의 항 사이에 또 다른 가상항을 생성하는 단계로 주어진 학습 데이터의 개수를 확장하는 것을 목적으로 한다. 이들의 생성을 위해 신경망을 이용한다.

2.1 학습(Training)

우선 시계열이 다음과 같이 주어져 있다.

$$x(1), x(2), x(3), \dots, x(n)$$

학습 패턴으로서 다음과 같이 입출력이 정의 된다.

*입력 벡터: $[x(t-1), x(t+1)]$

*출력 벡터: $[x(t)]$

위의 입출력 벡터를 이용하여 교사학습 (supervised learning)을 시킨다. 이때 교사 학습으로서 다층 퍼셉트론(Multilayer Perceptron)을 이용한다.

위의 입력 벡터를 날개 벡터(wing vector)라 하며 이때 날개 길이(wing length)는 1이다.

2.2 일반화(Generalization)

이제 학습된 다층 퍼셉트론으로 가상 항 (virtual term)을 생성하는 단계이다. 가상 항은 $v(t)$ 로 표시하며 실제 항 $x(t)$ 와 $x(t+1)$ 사이에 위치하게 된다. 즉 $x(t+0.5)$ 에 대한 추정치이다.

가상 항 $v(t)$ 를 생성하기 위한 입력 벡터는 다음과 같이 시간 지연 벡터(time delay vector)를 사용 한다.

*입력 벡터: $[x(t), x(t+1)]$

*출력 벡터: $[v(t)]$

그리하여 가상 항을 생성하여 형성된 새로운 시계열은 다음과 같이 되며 전체 항의 개수는 $2n-1$ 이 된다.

$$x(1), v(1), x(2), v(2), \dots, v(n-1), x(n)$$

3. 학습 및 예측: 후 학습 단계

여기서는 전학습 단계에서 형성한 시계열을 이용하여 그들을 학습 시킨후 미래의 값을 예측하는 단계이다. 먼저 시계열 예측에 대한 신경망의 접근에 대해서 언급하겠다. 시계열이 다음과 같이 주어져 있다면,

$$x(1), x(2), x(3), \dots, x(t)$$

$x(t+1)$ 을 예측하는 문제는 $[x(t-d), x(t-d+1), \dots, x(t)]$ 를 입력으로, $x(t+1)$ 을 출력으로 하여 비선형 회귀(nonlinear regression)으로 변환 시킨다.[4]

이를 구현하기 위한 방법으로는 다단계 직접 예측 (multi step direct prediction)과 단단계 반복 예측(single step iterated prediction)이 있다.

3.1. 다단계 직접 예측

다단계 직접 예측(multi step direct prediction)이란 가상항을 생성하기 이전의 일 단계 전도 예측(one step ahead prediction)을 가상항을 생성함으로써 이단계 전도 예측(two step ahead prediction)으로 변환하여 직접 미래의 값을 예측하는 것을 말한다.

2절에서 언급한 전학습에 의해 $x(1), x(2), \dots, x(t)$ 에는 $v(1), v(2), \dots, v(t-1)$ 가 사 이사이에 첨가된 시계열을 다음과 같이 전환 한다.

$$y(1), y(2), \dots, y(2m-1)$$

(이때 $y(2m)=v(m), y(2m-1)=x(m)$)

위의 시계열로 부터 다음과 같이 학습 패턴을 형성하여 미래의 값을 예측한다..

*입력벡터 -- $[y(t-k-1), \dots, y(t-2)]$

*출력벡터 -- $[y(t)]$

그리하여 위의 입력 벡터를 이용하여 여러단계를 전도하여 직접 $y(t)$ 를 예측하기 때문에 다단계 직접 예측(multi step direct prediction)이라 한다.

3.2. 일단계 반복 예측

3.1에서 언급 했듯이, 전 학습 단계에서 생성된 시계열이 다음과 같음을 가정한다.

$$y(1), y(2), \dots, y(2m-1)$$

(이때 $y(2m)=v(m), y(2m-1)=x(m)$)

일단계 반복 예측을 위하여 다음과 같이 학습 패턴을 나타낸다.

*입력벡터 -- $[y(t-k-1), \dots, y(t-2)]$

*출력벡터 -- $[y(t-1)]$

위의 학습 패턴을 이용하여 학습 시킨 후, $y(t)$ 를 예측하기 위해 먼저 $y(t-1)$ 을 예측하고, 예측된 $y(t-1)$ 을 다음과 같이 입력 벡터로 피드백하여 $y(t)$ 를 예측한다.

*입력벡터 -- $[y(t-k), \dots, y(t-1)]$

*출력벡터 -- $[y(t)]$

그리하여 위의 입력벡터를 이용하여 일 단계를 전도하고 반복하여 $y(t)$ 를 예측하기 때문에 일단계 반복 예측(single step iterated prediction)이라 한다.

4. 실험 및 결과

실험에서 사용한 데이터는 흑점의 연도별 개수를 0에서 1사이로 표준화 한 것을 사용하였다. 학습 데이터는 1700년-1920년의 표준화된 태양의 흑점의 개수로 하였고 테스트 데이터로는 2개의 부분으로서 1921년-1955년의 태양의 흑점의 개수와 1956년-1979년의 태양의 흑점의 개수로 하였다.

4. 1. 실험 과정

가상 항을 생성하기 위한 전학습 단계에서는 주어진 시계열 $x(1), x(2), \dots, x(n)$ 을 이용하여 다층 퍼셉트론으로 다음과 같은 입출력을 이용하여 학습 시킨다.

*입력 벡터 ---- $[x(k-1), x(k+1)]$

*출력 벡터 ---- $[x(k)] \quad (k=2, \dots, n-1)$

그리하여 가상항을 생성하기 위한 네트워크의 구조와 그를 구성하는 모수(parameter)는 다음과 같이 된다.

*입력층의 노드수----2개

*은닉층의 노드수----3개

*출력층의 노드수----1개

*학습율(learning rate)----0.03

그리고 가상 항, $v(k)$ 를 생성하기 위하여 다음과 같이 입력 벡터를 형성한다.

*입력 벡터 ---- $[x(k), x(k+1)]$
($k=1, \dots, n-1$)

그리하여 다음과 같이 새로운 시계열이 생성된다.

$x(1), v(1), x(2), v(2), \dots, v(n-1), x(n)$

위의 시계열을 다음과 같이 나타내고

$y(1), y(2), y(3), \dots, y(m) \quad (m=2n-1)$

다단계 직접 예측을 위한 학습 패턴은 학습 데이터로부터 입력 벡터를 23차원으로서 다음과 같이 나타낸다.

*입력벡터--- $[y(t-24), y(t-23)\dots y(t-2)]$

*출력벡터--- $[y(t)]$

일단계 반복 예측인 경우 다음과 같은 입출력 벡터로 주어진 네트워크를 학습 시킨다.

*입력벡터--- $[y(t-23), y(t-22) \dots y(t-1)]$

*출력벡터--- $[y(t)]$

그리하여 학습과 예측을 위한 네트워크 구조와 이를 구성하는 모수(parameter)는 다단계 직접 예측과 일단계 반복 예측 모두 다음과 같다.

*입력층의 노드수---23개

*은닉층의 노드수--- 8개

*출력층의 노드수--- 1개

*학습율-----0.03

4. 2. 예측 성능 평가 방법

예측 성능을 평가하기 위해서는 평균 상대 분산(average relative variace)를 이용하며 그의 공식은 다음과 같다.[1]

$$\frac{\sum_{k \in S} (\text{target}_k - \text{prediction}_k)^2}{\sum_{k \in S} (\text{target}_k - \text{mean})^2} \text{-----(1)}$$

S는 주어진 테스트 패턴의 집합이며 target은 관측치이다. mean은 테스트 패턴의 집합에서 관측치에 대한 평균이다.

4. 3. 각 방법의 예측 성능 결과

위에서 언급한 바와 같이 표준 역전파, 연결강도 제거 역전파, 가상항을 이용한 다단계 직접 예측, 그리고 그를 이용한 일단계 반복 예측의 성능 결과를 비교하였다. 그리고 연결강도 제거 역전파인 경우 [1]을 참고하기 바란다.

그리고 가상항을 사용하지 않는 경우의 표준 역전파와 연결강도 제거 역전파인 경우의 네트워크 구조와 모수는 다음과 같다.[5]

*입력층의 노드수----12개

*은닉층의 노드수----8개

*출력층의 노드수----1개

*학습율(learning rate)----0.03

각 방법의 결과는 다음 표1과 같다.

방 법	1921-1955	1956-1979
표준 역전파	0.0984	0.3678
연결강도 제거 역전파	0.0860	0.3511
다단계 직접 예측	0.0750	0.1226
일단계 반복 예측	0.0802	0.1171

표 1. 각 방법의 예측 성능

1956-1979인 경우 1700-1920까지의 데이터만 학습에 이용하였다.

그리하여 가상항을 생성하는 경우가 그렇지 않는 경우보다 예측성능이 우수함을 알 수 있다. 단 연결 강도 제거 역전파에서 1921-1955인 경우 [1]로 부터 인용된 수치이고, 1956-1979인 경우 [1]의 그래프에서 최저치를 목적으로 구한 값이므로 정확하지 않을 수 있다.

5.결 론

가상 항을 생성하여 예측한 결과가 그렇지 않은 결과 보다 우수함을 나타내었다. 이 방법은 학습 데이터의 부족을 해결해줄 뿐만 아니라 학습 데이터가 충분한 경우에도 예측 성능을 향상하기 위해 이 논문에서 제안한 방법이 이용될 수 있다.

이러한 방법이 예측성능을 향상 시킬 뿐만 아니라 또한 표준 역전파 보다 적은 학습횟수에도 충분한 학습 에러를 감소시킬 수 있음을 알 수 있다. 연결강도 제거 역전파인 경우 표준 역전파보다 예측 성능이 우수하기는 하지만[1], 학습 에러를 충분히 감소시키기 위해 더 많은 학습 횟수가 필요하며, 가상항을 생성하여 예측한 표준 역전파 보다 그의 성능이 우수하지 못하다.

여기서 남아 있는 과제는 날개 길이를 설정하는 방법으로서 주어진 날개길이에 따른 예측성능의 변화를 관찰하는 것이다. 현재 논문에서는 실험의 편의를 위해 날개길이 1을 사용하였지만, 그 길이에 따른 예측성능의 변화가 현재의 과제로 남아 있다.

참고 문헌

[1]Andreas S. Weigend and David E. Rumelheart "Generalization through Minimal Networks with Application to Forecasting" p362-370 In E.M. Keramidas editor, 23rd Symposium on the interface: Computing Science and Statistics (Interface 91) Seattle Interface Foundation of North America 1991

[2]Andreas S. Weigend and Neil A. Gershenfeld "Time Series Prediction" Santa Fe Institute 1992

[3]S. Cho and K. Cha, "Evolution of Neural Network Training Set through Addition of Virtual Sample," to appear in 1996 IEEE International Conference in Evolutionary Computation

[4]Andreas S. Weigend, Bernado A Huberman and David E. Rumelheart "Predicting the future: A Connectionist

approach 1990

[5]Chris M. Bishop "Training with Noise is Equivalent to Thikhonov Regularization" Neural Computation, Volume 7, Number1, 1995