

차별학습에 의한 시계열 예측에 대한 신경망 접근

조태호, 서정현
연구개발정보센터
정보시스템 개발실

e-mail : {tcjo, jeery}@ns.kordic.re.kr

Neural-based Approach to Time Series Prediction with Discriminant Learning

Taeho Charles Jo and Jerry Seo
KORDIC (Korea Research & Development Information Center)
GIIS (Group of Intelligent Information System)

요 약

시계열 예측에 있어서 과거의 측정치 보다 최근의 측정치가 미래의 측정치 예측에 중요한 영향을 미친다. 시계열 예측에 있어서 최근의 측정치와 과거의 측정치가 미래의 값을 예측하는 인자로서 차별화 되어 학습해야 할 것이다. 기존의 시계열에 대한 신경망 접근에서는 최근의 측정치에 대한 학습 패턴과 과거의 측정치에 대한 학습 패턴을 동일하게 학습 하였다. 이 논문에서는 과거의 학습 패턴과 최근의 학습 패턴을 학습 횟수 면에서 차별화 하였다. 이러한 학습을 이 논문에서는 차별학습 이라 한다. 차별학습에서는 주어진 학습 패턴을 시간 순으로 나열하고 일정 개수로 분할 한다. 시간의 역순에 의해 등차 또는 등비의 형태로 학습 횟수를 설정한다. 각 학습 패턴의 분할집단을 시간의 역순으로 일정 횟수를 감소시켜 학습 횟수를 설정하는 등차차별학습과 일정 비율로 감소시켜 학습횟수를 설정하는 등비차별학습을 소개한다. 기존의 신경망 접근 방법과 이 논문에서 제안한 신경망 접근방법을 비교하기 위해 Mackay-Galss 공식에 의해 인공적으로 생성된 시계열 데이터를 예로 사용하였다.

1. 서론

시계열 예측이란 과거의 측정치들의 관계를 분석하여 미래의 측정값을 예측하는 과정을 말한다. 과거의 측정값들과 미래의 측정값은 비선형적인 관계를 이룬다. 시계열 예측에 대한 통계적 모델, AR(Auto Regressive), MA (Moving Average), IMA (Integrated Moving Average), 그리고 ARIMA (Auto Regressive Integrated Moving Average)등은 과거의 측정값을 선형적으로 분석하여 미래의 값을 예측하는 통계적 모델이다 [1]. 실제로 시계열 예측은 비선형적이고 노이즈가 존재하므로 이러한 선형적 모델로는 한계가 있다.

시계열 예측에 대하여 통계적 모델 대신 신경망을 접근하는 많은 연구가 진행되어 왔다. 신경망은 기존의 통계적 모델에 비해 예측 오차가 적고 사전에 시

계열의 경향을 분석하지 않아도 되는 장점을 지니고 있다. 주로 사용되는 신경망 모델은 역전파 (Back Propagation)로 시계열 예측 뿐만 아니라 거의 모든 분야에서 가장 많이 사용되는 신경망 모델이다. 역전파에 대한 학습 및 일반화 알고리즘은 문헌 [2]를 참고하기 바라며, 이 논문에서는 언급하지 않겠다. 시계열 예측에 대한 신경망 접근의 연구에 있어서, 1991년 A. S. Weigend 와 D. E. Rumelhart 는 최초로 시계열 예측에 신경망 모델인 역전파를 적용하였다 [3]. 그 적용 과정을 2 절에서 자세히 언급하겠다 [3]. 1992년 K. Chakroborty, K. Mehrotra, C.K. Mohan, 그리고 S. Rana 는 다변수 시계열 예측에 대해 신경망 모델 중의 하나인 역전파를 적용하였다 [4]. 1995년 M. Cortell, B. Girard, Y. Girard, M. Mangeas, 그리고 C. Muller 는 시계열 예측에 대해 역전파의 불필요한 연결 강도를 제거

함으로써 예측오차를 감소시키는 기법을 제안하였다 [5]. 1996년에 N. Kohzadi, M. S. Boyd, B. Kermanshahi, 그리고 I. Kaastra는 월 평균 소와 밀값의 예측에 선형적 시계열 예측의 통계적 모델인 ARIMA (Auto Regressive Integrated Model)와 신경망 모델인 역전파를 비교하여 역전파의 성능이 우수함을 입증하였다 [6]. 동일 연도에 M. Malliaris와 L. Salchenberger는 S&P 주가 지수 예측에 대해 가장 보편적으로 사용되는 통계적 모델, Black-Scholes model과 신경망 모델인 역전파를 비교하여 역시 신경망 모델인 역전파가 그의 성능이 우수함을 입증하였다 [7]. D. Brownstone은 동일 연도에 주식시장의 주가예측에 있어서 다중회귀모델보다 신경망 모델인 역전파가 성능이 우수함을 입증하였다 [8]. 1997년에 F.W. Op't Landt는 그의 석사 학위 논문에서 주가 예측에 대한 신경망 모델인 역전파의 적용방법을 체계화하여 정리하였다 [9]. 1998년에는 B. F. Hobbs, U. Helman, S. Jiprapraikulsarn, S. Konda, 그리고 D. Maratukulam는 일일 전기 과부하 예측에 대해 신경망 모델인 역전파를 적용하였다 [10]. 이들의 문헌을 통해, 시계열 예측에 있어서 신경망 모델인 역전파가 전통적으로 사용되는 통계적 모델보다 성능이 우수하다는 것을 알 수 있고 [6][7][8], 모두 역전파를 신경망 모델로서 사용하였다 [3][4][5][6][7][8][9][10].

시계열 데이터는 시간에 대해 종속적이므로 측정 시점을 고려할 필요가 있다. 즉 미래의 측정치를 예측하는 데 있어서 과거의 측정치보다 최근의 측정치가 상대적으로 더 중요하게 고려되어야 한다 [1]. 즉 최근의 측정치가 미래의 측정치를 예측하는 것에 대해 과거의 측정치와 차별화 되어야 한다는 것이다. 신경망 모델에서 역전파보다는 데이터의 순서를 고려한 recurrent network를 시계열 예측에 적용한 연구가 진행되어 왔다. 1994년에 J.T. Connor, R.D. Martin와 L.E. Atlas는 신경망 모델 중에서도 recurrent network이 역전파 알고리즘보다 성능이 우수함을 나타내었다 [11]. 1998년에 B. Kermanshahi는 전력 과부하 예측에 있어서 역전파 알고리즘보다 recurrent network의 성능이 우수함을 입증하였다 [12]. 1999년에 R. J. Duro와 J.S. Reyes는 recurrent network을 사용하기보다 기존의 역전파 알고리즘을 개량한 이산시간역전파 (Discrete Time Backpropagation)를 시계열 예측에 적용하였다 [13].

신경망 모델 중에서 역전파 알고리즘이 다른 신경망 모델에 비해 구현하기가 용이하기 때문에 시계열 예측에 있어서 recurrent network보다는 역전파 알고리즘이 선호되고 있고 그의 개량형이 제안되고 있다 [13]. 시계열로부터 생성된 학습 데이터는 역전파, 이산시간역전파[13], 그리고 recurrent network[11][12] 모두 동일하게 학습하고 있다. 과거의 데이터로부터 추출된 학습 데이터는 필요 이상의 학습이 되어 효율성이 저하될 수 있다. 반면에 최근의 데이터로부터 추출된 학습데이터는 불충분하게 학습될 수 있다. 이 논문에서는 과거의 학습 패턴과 최근의 학습 패턴을 학습 횟수에서 차별화 하여 학습하는 것을 제안한다.

이 논문의 내용이 있어서 2 절에서는 시계열 예측

에 대한 역전파의 적용방법에 대해 기술 하며 3 절에는 차별학습 방법으로 등차차별학습과 등비차별학습에 대해 논의하겠다. 4 절에는 Mackay-Glass 공식에서 생성된 인공시계열을 대상으로 역전파의 차별학습과 일정학습의 성능 비교를 하겠다. 그리고 5 절에서는 본 연구의 의의와 앞으로의 연구방향에 대해 결론으로서 논의하겠다.

2. 시계열 예측에 대한 신경망의 접근

이 절에서는 시계열 예측에 대하여 신경망을 적용하는 과정에 대하여 기술하겠다. 적용할 신경망 모델은 역전파이고, 적용 과정은 문헌 [3]을 참조하였다. 1절에서 언급하였듯이 역전파의 학습 및 일반화 알고리즘은 [2]을 참조하고 이 절에서는 입출력 정의 및 학습 패턴과 테스트 패턴 추출 과정에 대해 기술하겠다.

시계열이 다음과 같이 주어져 있음을 가정하자.

$$X(1), X(2), X(3), \dots, X(T)$$

위의 시계열로부터 입출력 정의는 다음 그림 1과 같이 나타낸다. 그림 1에서 d 는 파라미터로서 지정되고 미래의 측정치를 예측하기 위한 과거 측정치의 범위를 나타낸다. 범위 내의 과거의 측정치와 현재의 측정치가 입력 벡터로 지정되고 미래의 측정치 $X(t+1)$ 이 출력값이 된다. 시계열 예측은 과거의 측정값을 입력으로 하여 미래의 측정값을 예측하는 다변수 비선형 회귀라고 할 수 있다.

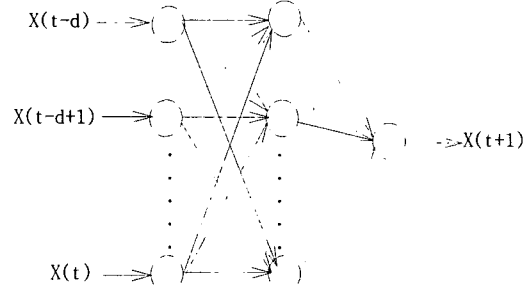


그림 1. 시계열 예측에 대한 역전파의 입출력 정의

위의 시계열로부터 학습 패턴을 추출하기 위한 학습 기간과 테스트 패턴을 추출하기 위한 테스트 기간으로 다음과 같이 분할 한다.

학습 기간: $X(1), X(2), X(3), \dots, X(T_0)$

테스트 기간: $X(T_0 + 1), X(2), X(3), \dots, X(T)$

학습기간으로부터 추출된 학습패턴은 다음과 같이 나타낼 수 있다.

$$\{X(1), \dots, X(d) | X(d+1)\}$$

$$\{X(2), \dots, X(d+1) | X(d+2)\}$$

.....

$$\{X(T_0 - d - 1), \dots, X(T_0 - 1) | X(T_0)\}$$

중괄호 내에서 수직막대의 좌측은 입력 벡터이고 우측은 목표 출력치가 된다.

잔여 학습 기간과 테스트 기간에서 추출되는

테스트패턴은 다음과 같이 나타낼 수 있다.

$$\{X(T_0 - d), \dots, X(T_0) | X(T_0 + 1)\}$$

$$\{X(T_0 - d + 1), \dots, X(T_0 + 1) | X(T_0 + 2)\}$$

.....

$$\{X(T - d - 1), \dots, X(T - 1) | X(T)\}$$

위의 과정에 의해 $T_0 - d$ 개의 학습 패턴과 $T_0 - d$ 개의 테스트 패턴이 생성된다. 위와 같이 생성된 학습 패턴과 테스트 패턴의 입력 벡터를 시간 지연 벡터라 하고 d 를 시간 지연 벡터의 크기라 한다.

3. 차별 학습

이 절에서는 주어진 시계열로부터 2절에서 기술한 과정에 의해 학습 패턴은 시간에 대해 순차적으로 생성된다. 이들을 일정개수로 분할하고 분할된 각 그룹을 학습패턴의 서브 그룹이라 하자. 서브그룹에 대해 최근에서 과거로 순으로 그 학습 횟수가 감소한다. 이 논문에서 두가지 차별 학습 방법을 제안하였는데, 등차차별학습과 등비차별학습이 제안된다. 등차차별 학습이란 학습 패턴의 서브그룹에 있어서 최근에서 과거의 순으로 학습 횟수가 일정 횟수에 의해 감소되는 경우를 말한다. 여기서 주어지는 파라미터는 분할하고자 하는 서브그룹의 개수 m 과 최근에서 과거의 순으로 학습 횟수를 감소시키는 고정된 값 r , 그리고 최근 학습 패턴의 서브 그룹의 초기 학습 횟수 N 이다. 2절에서 주어진 학습 패턴이 m 개의 서브그룹으로 분할된다고 가정하면 학습 패턴의 전체 집합 L 은 다음과 같이 표현할 수 있다.

$$L = \{G_1, G_2, \dots, G_m\}$$

학습 패턴의 서브 그룹 G_i 는 다음과 같이 시계열에서 생성된 학습 패턴으로 구성된다.

$$G_i = \left\{ \left\{ X\left(\frac{(T_0 - d)(i - 1)}{m} + 1\right), \dots, X\left(\frac{(T_0 - d)(i - 1)}{m} + d\right), \left| X\left(\frac{(T_0 - d)(i - 1)}{m} + d + 1\right), \dots, X\left(\frac{(T_0 - d)i}{m}\right), \dots, X\left(\frac{(T_0 - d)i}{m} + d - 1\right), \left| X\left(\frac{(T_0 - d)i}{m} + d\right) \right\} \right\}$$

각 학습 패턴의 서브그룹 G_i 에 대한 학습 횟수 K_i 는 다음과 같이 할당 된다.

$$K_i = N - (m - i)r$$

등비차별학습이란 학습 패턴의 서브그룹을 최근에서

과거로 1보다 큰 일정 비율로 학습 횟수를 감소시키는 것을 말한다. 여기서 주어지는 파라미터는 학습 횟수에 역수를 취하여 곱하는 상수 r 를 제외하고 등차차별학습의 경우와 동일하다. 각 학습 패턴의 서브그룹 G_i 에 대한 학습 횟수 K_i 는 다음과 같이 할당 된다.

$$K_i = \frac{N}{s^{m-i}} (s > 1.0)$$

4. 실험 및 결과

차별학습과 기존의 동일 학습의 성능 비교를 위해 동적 시스템인 Mackay-Glass 공식에 의한 시계열을 데이터를 사용하였다. Mackay-Glass 공식은 다음과 같이 주어진다.

$$X(t + 1) - X(t) = \frac{aX(t - \Delta T)}{1 + X^{10}(t - \Delta T)} - bX(t)$$

위의 공식에서 임의로 a 는 0.1로 하고 b 는 0.2로 하였으며, ΔT 는 75 및 100으로 설정하였다. 500개의 시계열이 위의 공식으로부터 추출되고 1-300은 학습 기간으로 301-500은 테스트 기간으로 설정하였다.

위의 공식에 의해 생성된 시계열로부터 시간 지연 벡터인 d 는 10으로 설정하여 학습 패턴 및 테스트 패턴을 형성하였다. 290개의 학습 패턴과 200개의 테스트 패턴이 생성된다.

역전파를 신경망 모델로서 사용하였으며, 그의 구조로는 입력층, 은닉층 그리고 출력층의 3개의 층으로 구성된다. 입력 층에서는 10개의 입력 노드, 은닉층에서는 10개의 은닉노드, 그리고 출력층에는 1개의 출력 노드가 사용되었다. 학습률은 0.1로 설정하였고, 학습 횟수는 2500, 5000, 7500, 그리고 10000으로 하여 이들의 성능을 비교하였다.

ΔT 를 75로 설정한 경우 동일 학습과 차별학습의 시계열 예측의 오차를 다음 표 1과 같이 요약하였다.

표 1. $\Delta T = 75$ 인 경우 예측 오차 (10^4)

	동일학습	등차차별학습	등비차별학습
2500	4.085	5.690	4.409
5000	3.545	2.434	3.296
7500	2.344	1.904	2.630
10000	1.917	1.458	1.879

표1에서 학습 횟수가 2500회인 경우 차별학습이 동일 학습보다 예측 성능이 향상되지 않았으나, 학습횟수가 5000회와 10000회인 경우 차별학습이 동일학습보다 예측성능이 향상되었다. 7500회인 경우는 등차차별학습이 동일학습보다 성능이 향상되었으나, 등비 차별학습이 향상되지 않음을 나타내고 있다.

ΔT 를 100으로 설정한 경우 동일 학습과 차별학습의 시계열 예측의 오차를 다음 표 2와 같이 요약하였다.

표 2. $\Delta T = 100$ 인 경우 예측 오차 (10^4)

	동일학습	등차차별학습	등비차별학습
2500	10.352	11.132	9.661
5000	7.950	5.804	6.034
7500	8.937	5.549	4.918
10000	6.284	4.506	3.596

표 2에서 학습 횟수가 2500 회인 경우 등비차별학습은 동일학습에 비해 예측성능이 향상되었으나, 등차차별학습은 성능이 향상되지 않았다. 학습횟수가 5000 회, 7500 회 그리고 10000 회인 경우 30%정도 동일 학습에 비해 차별학습의 예측성능이 향상되었다.

5. 결론

시계열 예측에 대한 신경망의 차별학습은 예측성능을 향상시킬 뿐만 아니라 학습시간을 단축시킨다. 예를 들어 5000회 정도 주어진 신경망을 학습시킨다면 동일 학습은 모든 학습 패턴이 5000회 학습하지만 차별 학습에서는 과거의 학습 패턴일수록 학습횟수가 감소하기 때문에 처리시간이 감소하므로 학습시간이 단축된다.

이 논문에서는 Mackay-Glass 공식에서 생성된 인공 데이터를 이용하여 차별학습의 성능을 검증하였다. 강우량, 온도, 주가 지수, 그리고 환율등의 실제 데이터를 이용하여 차별학습의 성능을 검증하는 것이 중요하다. 차별학습은 학습 횟수가 작을수록 과거의 데이터가 과도하게 불충분한 학습이 이루어져 동일 학습보다 성능이 저하될 수 있다. 최근학습 패턴에 대한 학습횟수를 동일학습에 비해 높게 설정해야 할 필요가 있다. 학습횟수, 학습데이터의 서브그룹의 개수, 그리고 차별학습 강도를 나타내는 등차차별학습의 r 과 등비차별학습의 s 의 값의 설정이 중요하다.

참고문헌

[1] G. E. P. Box, C. M. Jenkins, and G. C. Reinsel, *Time Series Analysis*, Prentice Hall, 1994.
 [2] J. A. Freeman and D. M. Skapura, *Neural Network: Algorithms, Applications, and Programming Techniques*, Addison-Wesley Publishing Company, 1991.
 [3] A.S. Weigend and D.E. Rumelhart, "Generalization through Minimal Networks with Application to Forecasting", pp362-370, *Interface* 91, 1991.
 [4] K. Chakroborty, K. Mehrotra, C.K. Mohan, and S. Ranka, "Forecasting the Behavior of Multivariate Time Series using Neural Networks", pp961-970, *Neural Network Vol 5*, 1992.
 [5] M. Cortell, B. Girard, Y. Girard, M. Mangeas, and C. Muller, "Neural Modeling for Time Series: A Statistical Stepwise Method for Weight Elimination", pp1355-1363, *IEEE Transaction on Neural Network Vol 6*, No 6, 1995.
 [6] N. Kohzadi, M. S. Boyd, B. Kermanshahi, and I. Kaastra, "A comparison of artificial neural network and time series models for forecasting commodity prices", pp169-181, *Neurocomputing No 10*, 1996.
 [7] M. Malliaris and L. Salchenberger, "Using neural networks to forecast the S&P 100 implied volatility", pp183-195, *Neurocomputing No 10*, 1996.
 [8] D. Brownstone, "Using percentage accuracy to measure neural network prediction on Stock Market movements", pp237-250, *Neurocomputing No 10*, 1996.
 [9] F.W. Op't Landt, "Stock Price Prediction using Neural Networks", Master Thesis of Leiden University, 1997.
 [10] B. F. Hobbs, U. Helman, S. Jiprapraikulsam, S. Konda,

and D. Maratukulam, "Artificial Neural Network for short-term energy forecasting: Accuracy and Economic Value", pp71-74, *Neurocomputing 23*, 1998.
 [11] J.T. Connor, R.D. Martin and L.E. Atlas, "Recurrent Neural Network and Robust Time Series Prediction", pp240-254, *IEEE Transaction on Neural Network vol 5*, 1994.
 [12] B. Kermanshahi, "Recurrent Neural Network for forecasting next 10 years loads of nine Japanese Utilities", pp125-133, *Neurocomputing 23*, 1998.
 [13] R. J. Duro, and J.S. Reyes, "Discrete Time Backpropagation for Training Synaptic Delay-Based Artificial Neural Network", pp779-789, *IEEE Transaction on Neural Network Vol 10 No 4*, 1999.