

다층 신경망을 이용한 비선형 예측 알고리즘에 관한 연구

박형근*, 김선엽**

*남서울대학교 전자공학과

**남서울대학교 정보통신공학과

e-mail:phk315@nsu.ac.kr

A Study on the Non-linear Prediction Algorithm using Multi-layer Neural Network

Hyoung-Keun Park*, Sun-Youb Kim**

*Dept of Electronic Eng., Namseoul University

**Dept of Information Communication Eng., Namseoul University

요 약

현대사회의 발전으로 인해 생성되는 수많은 정보는 시간과 공간의 제약이 없는 다차원적인 특성을 갖고 있으며, 사회 전반에 걸쳐 보다 나은 결과를 위한 의사결정에 활용되고 있다. 또한 우리생활에서 발생하는 많은 현상을 보다 합리적이고 과학적인 방법으로 분석하여 정확한 예측이 이루어진다면 미래의 불확실성에 대한 불안을 해소하고, 현재의 의사 결정을 하는데 큰 도움이 될 수 있다. 따라서 본 논문에서는 다층 신경망을 이용하여 비선형적 관계를 표현할 수 있는 적응 능력을 갖을 뿐만 아니라 비선형 통계예측에 적용이 가능한 알고리즘을 제안하고 분석하였다.

1. 서론

IT의 발전으로 인하여 생성되는 수많은 정보는 시간과 공간의 제약이 없는 다차원적인 특성을 갖고 있으며, 사회 전반에 걸쳐 보다 나은 결과를 위한 의사결정에 활용되고 있다.[1] 올바른 정보의 사용은 효율적인 정책의 결정, 집단의 구조적 성질의 이해, 최적 생산 조건의 결정 등을 가능하게 하여 사회 발전에 큰 도움을 준다. 그러나 잘못된 정보는 예상하지 못한 오류를 발생시키게 되며, 올바른 정보라 할지라도 분석 방법이 나쁘면 전혀 비현실적인 결과를 가져오게 된다. 그러므로 우리생활에서 발생하는 많은 현상을 보다 합리적이고 과학적인 방법으로 분석하여 정확한 예측이 이루어진다면 미래의 불확실성에 대한 불안을 해소하고, 현재의 의사 결정을 하는데 큰 도움이 될 수 있다.[2]

따라서 본 논문에서는 다층 신경망을 이용하여 비선형적 관계를 표현할 수 있는 적응 능력을 갖을 뿐만 아니라 비선형 통계예측에 적용이 가능한 알고리즘을 제안하고 분석하였다.

2. 다층 신경망을 이용한 적응 및 통계예측

신경망 모델을 예측 목적으로 사용하기 위해서는 모델의 연결강도(Weight)들이 추정되어야 한다. 이를 추정하는 훈련(Training)과정은 분류 문제에서 지도 학습(Supervised Learning) 또는 비지도 학습(Unsupervised Learning)에 의해 수행된다. 지도 학습은 입력과 해당 출력이 사전에 알려져 있어 입력 데이터가 해당 출력을 생성하도록 연결강도를 반복적으로 조정하여 생성된 분류값의 오차를 최소화하는 알고리즘에 의해 수행되며 비지도 학습은 클래스(Class)에 대한 사전 정보가 없이 자기조직화(Self-organizing) 알고리즘에 의하여 군집을 구분한다.[3] 신경망을 사용한 일반적인 예측 시스템은 그림 1과 같다.

과거의 데이터베이스는 전처리되어 신경망 S/W 시뮬레이터를 통해 학습된 후 현재의 데이터를 바탕으로 산출된 출력값을 후처리하여 예측값을 계산한다. 실제 사용되는 전처리 기법의 종류에는 이동평균(Moving Average), 변화율(Rate of Change), 모멘텀(Momentum) 등이 있다. 또한 이동평균은 단순 이동

평균(Simple Moving Average)과 지수이동평균(Exponential Moving Average)으로 분류되며, 식(1), 식(2), 식(3)과 같다.

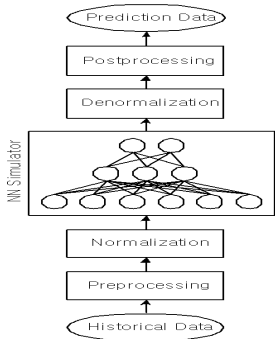


그림 1. 신경망 예측 시스템

$$SMA_t = \frac{P_t + P_{t-1} + \dots + P_{t-n}}{n} \quad (1)$$

n : 이전 시간주기의 수
 P_{t-n} : n 이전 시간주기의 가격

$$EMA_t = (1 - \alpha)P_t + \alpha(EMA_{t-1}) \quad (2)$$

α : 상수, P_t : 시간 t 에서 가격

$$M_t = P_t - P_{t-\alpha}, \quad ROC = \frac{P_t}{P_{t-\alpha}} \quad (3)$$

3. 신경망 모델 및 오차 역전파 학습

인간 두뇌의 구조와 기능을 모방한 신경망은 최근 패턴인식, 음성인식, 문자인식, 로봇공학 등에 이용되고 있으며, 연산단위(Processing Element : PE)가 서로 연결되어 임의의 입력과 출력에 대해 매핑을 수행한다.[4]

3.1. 신경망 모델

신경세포(Neuron)를 수학적인 모델로 추상화하면, 시냅스는 임의의 입력에 대한 연결강도, 수상돌기와 세포체는 시냅스의 입력을 축적하는 덧셈기, 축색은 어떠한 비선형 함수에 해당한다. 이에 따른 신경 세포의 모델인 연산단위는 그림 2와 같다.

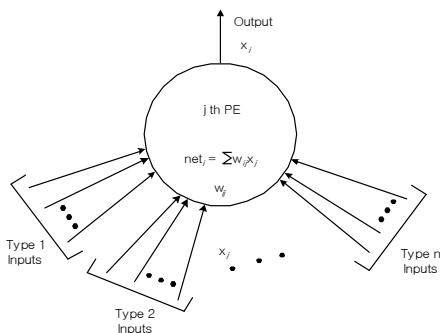


그림 2. 연산단위의 구조

연산단위는 임의의 입력 x_i 에 대해 연결강도 w_{ji} 을 곱하고 이를 합한 후 비선형 함수를 통해 출력한다. 비록 생물학적인 신경 세포의 세부 동작과는 거리가 있지만 가장 기본적인 특성만을 수학적으로 추상화한 모델이다.[7] 신경망은 연산단위가 서로 연결된 구조로 되어 있으며, 본 논문에서는 여러 개의 연산단위가 층을 이루며, 각 층에서 다음 층으로만 연결된 구조로 되어 있는 순방향 신경망을 기준으로 한다.

3.2. 오차 역전파 학습 알고리즘

신경망의 학습 방법은 목표값을 지정하고 출력값과의 오차를 이용하여 신경망의 연결강도를 조절하는 지도 학습, 입력되는 데이터의 구조와 분포에 따라 내부적인 범주를 형성하는 비지도 학습으로 분류된다. 이러한 비지도 학습이 생물학적인 학습 과정에 더 가깝지만 실제 응용 분야에서는 지도 학습 방법이 더 좋은 결과를 얻고 있으며, 어느 정도 성공적으로 학습됨이 입증된 알고리즘이 오차 역전파(Error Backpropagation) 학습 알고리즘이다.[5][6]

이 알고리즘은 오차의 제곱합(MSE)을 최소화하도록 반복적으로 기울기를 따라가는(Gradient Descent) 방법을 적용하여 연결강도를 조정해 나간다. 이를 적용하기 위해서는 연산단위의 비선형 함수가 미분 가능한 단순 증가함수이어야 한다. 여기에 응용되는 시그모이드(Sigmoid) 함수는 로지스틱(Logistic), 하이퍼볼릭 탄젠트(Hyperbolic Tangent) 함수 형태가 사용된다.

이러한 알고리즘을 적용하여 H/W를 구현할 경우에 초기 연결강도나 바이어스 값은 가능한 작은 값으로 해야 한다. 또한, 학습률(Learning Rate)을 크게 하면 학습속도가 향상되지만 국부최소치(Local Minima)에 빠지거나 Overshoot가 발생하여 정확도가 떨어진다. 학습률을 작게할 경우 속도는 느리지만 국부최소치 및 Overshoot를 방지할 수 있다. 따라서 적당한 학습률의 사용이 시스템의 성능을 좌우할 뿐만 아니라 신경망의 크기, 즉 은닉층의 뉴런수가 적절해야 한다.

4. MNN 모델의 선택 방법

다양한 신경망 모델은 각기 고유의 특성을 가지고 있어서 특정 분야의 현상에만 선별적으로 적용할 수 있다. 특히, 분석가가 주어진 문제를 해결하기 위하여 적절한 신경망 모델을 찾아내고 선택하는 것에 대하여는 아직까지 정립된 일반적인 규칙은 없다. 따라서 신경망 모델을 사용할 경우에 주어진 상황에 따라 모델을 구축하는 사람의 자의적 판단에 의하여

최선의 모델을 선택해야 한다는 취약점이 있다.

신경망의 특정 모델의 결정을 위해서는 먼저 네트워크의 형태를 결정해야 하며, 가장 흔히 사용되는 다층 퍼셉트론(Multi-layer Perceptron : MLP)과 RBF(Redial Basis Function) 네트워크 모델이 있다.

4.1. 신경망의 구조

다층 퍼셉트론 네트워크에는 하나의 입력층과 하나의 출력층이 있어야하므로 계층의 수를 결정하는 것은 은닉층의 수를 결정하는 것이다. 이러한 다층 퍼셉트론은 단층 퍼셉트론과 유사한 구조를 가지고 있지만 중간층과 각 유닛의 입출력 특성을 비선형으로 함으로써 네트워크의 능력을 향상시켜 단층 퍼셉트론의 여러 가지 단점들을 극복했다.

하나의 은닉층을 가진 모델은 덜 복잡하기 때문에, 대부분의 연구자들은 두 개의 은닉층을 가진 신경망의 경우 함수를 정확하게 학습할 수 있다고는 하지만 데이터 통신망 예측이나 금융관련 데이터 분석의 경우에는 데이터들이 두 개의 은닉층을 적용할 경우에 과적합(Overfitting)이 일어날 가능성이 많아지게 되어 학습시간도 길어지게 된다. 이러한 특성으로 인해 한 개의 은닉층을 가진 모델을 사용하는 것이 적당하다. 입력층의 노드 수는 입력 변수의 수와 같다. 그러나 최소한의 노드로 구성된 모델이 더 복잡한 모델보다 그 결과를 일반화하는데 유리하다.

은닉층의 노드 수도 일정한 규칙은 없으나 과도한 수의 은닉층 노드를 갖는 신경망 모델은 훈련 표본 외의 데이터에 대한 분류 능력을 저하시키며, 은닉층의 노드 수는 입력층의 노드 수보다 작아야 한다. 또한, 너무 적은 개수의 노드를 사용할 경우 선형 모델에 가까워지게 되어 데이터의 분포가 비선형일 경우 정확한 예측을 하기 어렵다. 따라서 입력 노드의 수가 많을 경우에는 50%~70%의 정도의 입력 노드를 사용해야 하고 입력 노드의 수가 적을 경우에는 $2N \sim 2N+1$ (N : 입력 노드 수 + 출력 노드 수) 개의 노드를 은닉 노드로 가져야 한다.

출력층의 노드 수를 결정하는 방법에 확실한 방법이 있는 것은 아니다. 실수 데이터를 입력 변수 및 출력 변수로 할 때 이것을 구간 변수로 구분하여 적용하는 것은 변환하는 과정이나 방법에 따라 값에 오차가 생기게 된다. 한편, 데이터 분석 및 예측 문제에서 실수 데이터를 그대로 사용할 경우 각각의 값별로 출력 노드를 만들어 둔다는 것은 불가능하다. 따라서 데이터 분석 분야에서 실수 데이터를 입력 변

수로 사용할 경우에는 하나의 출력 노드를 사용하여야 정확한 값을 예측할 수 있다.

네트워크 구조는 노드들 간의 상호연결에 의해서도 특징 지워지며, 기본적으로 네트워크에서 노드 사이의 연결은 네트워크의 행동을 결정한다. 입력 노드에서 출력 노드로 직접 연결하게 되면 은닉층에서 계산된 연결 가중치보다 더 큰 영향력을 미치게 되므로 예측오차가 커질 가능성이 높아진다. 한편, 완전히 연결하지 않고 신경망 모델을 사용할 경우에도 노드간에 영향력에 문제가 생기게 된다. 따라서 한 층에서 출력층을 제외한 다음 계층으로 완전히 연결한 모델이 적합하다.

앞에서 살펴본 전이함수 가운데 하나를 선택하는 것은 입력 데이터의 성격과 네트워크에 무엇을 학습시키려는가에 달려 있다.

4.2. 신경망의 학습

적절한 수의 은닉층과 은닉마디의 수가 정해지면 오차함수(Error Function)라고 하는 목적함수(Object Function)가 최소가 되도록 데이터로부터 연결강도 계수(Coefficient)들을 식(4)와 같이 나타낸다.

$$e_i = \sum_{i=1}^n (Y_i - P_i)^2 \tag{4}$$

Y_i : i 번째 개체의 실제값
 P_i : i 번째 개체의 예측값

오차 함수는 SSE(Sum-of-Square-Error)를 많이 사용하며 데이터 분석 및 예측의 경우에도 SSE가 가장 적합하다.

학습규칙(Learning Rule)은 서로 다른 계층 노드간의 연결강도를 조정하는 규칙이며, 일반적으로 무작위성이 포함된 데이터에 대해서는 델타규칙이 좋은 성과를 나타낸다. 델타규칙을 적용할 경우 학습에서 데이터를 무작위로 입력하는 것이 중요하다. 따라서 대부분의 신경망 소프트웨어는 데이터를 난수(Random Number)에 의해서 무작위로 입력하는 기능을 제공한다. 여기에서 데이터의 무작위성에 따른 문제점을 보완하기 위하여 도입된 학습규칙이 표준화 누적 델타 규칙이다. 이 규칙은 무작위성의 문제를 경감시키기 위하여 연결강도의 변화를 일정 케이스 수만큼 누적시켰다가 수정하는 것이다. 이렇게 함으로써 수정치를 평균화하여 학습에 입력되는 데이터의 순서에 따른 영향을 경감할 수 있다. 실수값을 그대로 사용하여 예측하는 경우에는 데이터 무작위성이 높으므로 델타규칙이 표준화 누적 델타 규칙보다 더 적합하다.

또한, 초기 연결강도는 훈련 이전에 결정되어야 한다. 일반적으로 이러한 연결강도들은 먼저 지정한 범위 안에서 랜덤하게 할당한다. 실수 주가 데이터를 이용한 예측의 경우 무작위성이 높으므로 랜덤하게 초기 연결강도 값을 정해야 한다.

4.3. 데이터의 준비

신경망에 대한 입력 데이터는 전이함수에 사용 가능하도록 변환시켜야 한다. 일반적으로 사용하는 시그모이드 함수에서 0에서 1사이의 입력 데이터 범위를 사용하는 것은 한가운데 데이터 비율을 정하는 것이고, 함수 선형 가까운 부분의 초기 반복에서 계산을 제한할 것이다.

입력 데이터의 범위를 더 넓게 함에 따라 가장 먼저 훈련 표본들은 전이함수에 더 높거나 더 낮은 범위에서 잘려질 것이다. 결국, 이끌어낸 가중치들은 전이함수의 입력범위와 출력범위 사이의 어떤 차이를 보상해야 한다. 그러나 입력 데이터의 범위의 올바른 지정은 신경망의 훈련을 촉진하는 기능을 한다.

목표값들의 범위는 출력층에서 전이함수의 출력 범위 안에 있도록 조정되어야 하며, 오차 함수는 전역 최소화 외에도 지역 최소화를 가지고 있다. 범위 근처 또는 범위에서 목표값들은 네트워크의 포화상태를 야기할 것이다. 비선형 전이함수는 비대칭적인 제한을 가지기 때문에 아주 큰 입력값은 범위에 있는 출력값을 생산하기 위해서 전이함수를 야기하도록 요구된다. 알고리즘은 큰 입력값에 도달하기 위해 큰 입력값들에서 가중치를 조정하려고 할 것이다. 더 나은 방법은 목표값이 전이함수의 출력범위의 90%에 정해지는 것이다. 시그모이드 전이함수를 위해서는 목표값들은 0.05에서 0.95 사이의 범위의 값을 가져야 한다.

특히, 데이터 분석 및 예측의 경우에 데이터의 성격에 따라 값의 규모의 차이가 크게 나타나기 때문에 값의 규모를 비슷하게 통일하기 위해서 정규화가 반드시 필요하다.

훈련 및 테스트 데이터는 신경망을 구축할 때 일반적으로 요구되는 부분이다. 훈련 데이터(Training Sample)는 신경망 모델을 개발할 때 사용되고 테스트 데이터(Test Sample)는 모델의 예측 능력을 평가하는데 사용된다. 또한, 평가 데이터(Validation Sample)는 훈련을 멈춰야 하는 시점을 결정하기 위해 사용되거나 과적합 문제를 피하기 위해 사용한다. 특히, 적은 데이터 집합의 경우 평가와 테스트 목적으로 하나의 테스트 집합을 사용하는 것은 일반

적이다. 이러한 관점에서 훈련 데이터와 테스트 데이터를 선택하는 것이 신경망의 성능을 결정한다.

따라서 훈련 데이터와 테스트 데이터의 분할을 위해 문제의 특성, 데이터 형태, 이용할 수 있는 데이터의 크기 등과 같은 몇 가지 요소들이 고려되어야 한다. 훈련 데이터와 테스트 데이터의 적절하지 못한 분리가 신경망 모델의 예측 능력을 평가하고 최적의 신경망 구조를 선택하는데 영향을 미친다.

대부분의 연구들은 90% : 10%, 80% : 20%, 70% : 30%으로 데이터를 나누었다. 금융관련 데이터 분석 및 예측 분야에 이용할 수 있는 데이터의 크기가 충분히 크다고 할 때 훈련 데이터와 테스트 데이터의 비는 70% : 30%가 적당하다. 90% : 10%, 80% : 20%일 경우에는 테스트 데이터 비가 낮아 훈련 결과에 대한 평가가 충분하지 못하게 된다.

V. 결론

비선형 통계 예측 분야의 경우 기본적으로 대부분의 신경망은 컴퓨터 상에서 운영되는 S/W에 의해 구현되어 사용하고 있다. 하지만 직렬 컴퓨터 상에서의 S/W 시뮬레이션은 병렬 연산이 직렬 연산으로 전환되기 때문에 요구되는 방대한 계산량을 제공할 수가 없다는 단점을 가지고 있어서 신경망을 실시간 처리가 요구되는 응용분야에 적용하기 위해서는 다층 신경망을 이용한 하드웨어 구현이 매우 필수적이다. 따라서 본 논문에서는 비선형 예측기를 하드웨어로 구현하기 위한 선행 연구로서 특화된 다층 신경망 모델의 구조와 학습 알고리즘의 적용방법을 제안하였다.

참고 문헌

- [1] 정재영, "정보화 시대의 대학교육", 서울경제신문 15판 제8820호.
- [2] 김혜중, "시계열 분석", 동국대학교 출판부, 1989.
- [3] 김원경, "시계열 분석", 경문사, 1993.
- [4] Casdagli, M. and Eubank, S. "Nonlinear Modeling and Forecasting", Addison, New York, 1992.
- [5] T. Poggio and F. Girosi, "Network for Approximation and Learning", Proc. IEEE, vol. 78, no.9, pp.1481-1497, Sep. 1990.
- [6] H. Hecht-Nielsen, "Neurocomputing : Picking the Human Brain", IEEE Spectrum, pp.36-41, March 1988.