

유전자신경망을 이용한 시계열예측

윤여창

우석대학교 정보보안학과
e-mail : yoonyc@cyber.ws.ac.kr

Time Series Forecasting Based On Genetic Neural Network

YeoChang Yoon

Dept. of Information Security, Woosuk University

요 약

이 연구에서는 유전자알고리즘과 인공신경망의 특성을 결합한 유전자신경망모형에 대하여 논의한다. 이 모형을 이용하여 단기 시계열자료를 예측한다. 그 예측 결과는 유전자신경망모형이 역전파 신경망모형에서 보다 더 작은 예측오차를 보였다. 역전파 신경망보다 더 효과적임을 보임으로써 유전자신경망모형을 이용한 시계열자료 예측이 보다 효율적인 방법임을 제시한다.

1. 서론

시계열자료의 예측은 여러 분야에서 중요하게 다루어지고 있다. 최적의 예측방법을 선택하는 문제는 특히 중요하다. 전통적인 신경망 분석기법은 그 학습의 수렴 정도가 매우 느리고 지역 최소값에 쉽게 수렴되는 경향이 있다. 유전자알고리즘은 1970 년대에 생물학적 진화론을 모방한 메커니즘으로 개발되었고 이제는 폭넓게 적용되고 있다.

이 연구에서는 신경망과 유전자 알고리즘의 특성을 결합하여 유전자신경망모형(genetic neural network)을 구축하고자 한다. 그리고 자동차생산량과 같은 시계열자료에 유전자신경망모형을 적용하고 그 결과를 살펴본다.

2. 유전자신경망모형

2.1 전통적 역전파 신경망의 단점

역전파 신경망은 경사(gradient) 알고리즘을 주로 이용한다. 신경망학습은 아주 작은 학습을 값으로 시작하며 이는 학습과정의 수렴을 이 매우 늦어지는 이유중의 하나다. 비록 학습율을 증가시킬 수 있는 모멘텀 방법을 사용할지라도 학습속도가 충분히 개선되지 않는 문제에 직면하게 된다. 학습과정에서 역전파 신경망은 이에 대한 명확한 해결책을 찾을 수 없었다. 몇몇 경우에서 역전파 신경망은 지역최소값에 수렴되는 학습결과를 낳기도 한다. 이때는 전역 최소값을 얻기 위하여 초기 가중값의 변화를 통한 신경망 학습 그리고 학습량의 증가를 통한 시행착오적인 해결방법을 모색하기도 한다.

2.2 유전자 알고리즘의 기초 이론

유전자 알고리즘은 다윈의 자연선택(natural selection)에 근거한 생물학적 진화론을 이용한 일종의 계산모형으로써 최적의 성능을 보이는 학습 알고리즘 중의

하나다. 다윈의 진화론에 따르면 각 종들은 외부 환경의 적응을 통하여 진화 과정 중에 계속 개선된다. 환경에 적응할 수 있는 각자의 특질은 여기서 가장 알맞게 보존되고 살아남게 된다. 이들의 특질은 선택(selection), 유전교차(crossover), 그리고 돌연변이(mutation) 과정을 통하여 형성되며 그 과정은 단계적이면서 점차적으로 최적의 해에 접근한다. 본 연구에서 다루는 유전자 알고리즘의 처리과정에는 부호화(coding), 모집단의 초기화, 적합성 계산, 선택, 유전교차, 돌연변이 등과 같은 과정들이 포함된다.

2.2.1 부호화 문제

실제적인 문제를 해결하기 위하여, 유전자 알고리즘의 연산에 적합시키기 위한 이진 코드로 문제의 해를 변환할 필요가 있다. 먼저, 변화 범위와 같은 모수들을 결정할 필요가 있다. 만약 변화 범위와 같은 모수들이 $[a_{min}, a_{max}]$ 라면, a 는 이진수 b 처럼 표시될 수 있다. 그리고 a 와 b 는 다음과 같은 관계를 갖는다.

$$a = a_{min} + \frac{b}{2^m - 1} (a_{max} - a_{min})$$

여기서 모수들의 모든 이진 문자열을 하나의 문자열로 연결할 수 있다. 각각의 문자열과 문자열의 모든 비트들은 0 과 1 의 두 종류 값 들이다. 이와 같은 방법으로 유전자 알고리즘은 문자열을 연산한다.

2.2.2 초기 모집단 생성

초기 모집단을 생성시키는 간단한 방법에는 동전 던지기 와 같은 랜덤한 방법이 있다. 또한 구간 -1 과 1 사이의 균등분포로부터 확률난수 발생을 통하여 얻을 수 있다.

2.2.3 적합도 함수

유전자알고리즘에서, 환경에 개별적으로 적응하는 정도를 적합도라고 한다. 적합도함수는 환경에 개별적

으로 적응하는 능력을 표현한다. 즉 적합도가 클수록 개별적 유전능력은 더욱 강하다. 다음 세대에 선택되는 개별적 특질의 확률이 크면 클수록 우리가 얻게 되는 해는 더욱 더 좋은 결과를 얻는다. 만약 목적함수가 최소 오차를 구하는 것이라면 개별 적합도는 다음과 같이 오차제곱합의 역으로 정의할 수 있다.

$$E = \frac{1}{n} \sum_{k=1}^n (t_k - y_k)^2.$$

여기서 t_k 는 모형의 출력값이고 y_k 는 목표값, n 은 전체 자료 개수이다.

2.2.4 유전자 연산자

1) 선택

개별적 적합도를 따르는 선택은 복사될 개별 확률을 결정한다. 적합도가 높은 개체는 유전적으로 다음 세대에 나타난다. 그룹의 개체번호 i 에 대한 A_i 에 대하여 복사될 기대값 M 은 다음과 같다.

$$M = n \frac{E(A_i)}{\sum_{i=1}^n E(A_i)}.$$

여기서 n 은 모집단의 크기이고 $E(A_i)$ 는 개체 A_i 의 적합도이다.

복사 연산을 실행하기 위하여 먼저 현재 모집단 내의 개체 적합도 $E(A_i)$ 를 계산한다. 그러면 적합도 F 의 전체 모집단은 다음과 같다.

$$F = \sum_{i=1}^n E(A_i).$$

이때 각 개체의 확률은 다음과 같다.

$$P_i(A_i) = \frac{E(A_i)}{F}.$$

2) 유전교차

유전교차는 모집단으로부터 어떤 확률 P_c 에 따르는 개체 수를 선택하는 것이다. 유전교차율은 다음 식과 같이 자동 적응조정 알고리즘을 이용한다.

$$P_c = \begin{cases} k_1 (f_{max} - f'_c) / (f_{max} - \bar{f}), & f'_c \geq \bar{f} \\ k_3, & f'_c < \bar{f} \end{cases}$$

여기서 f'_c 은 유전자 교차가 이루어지기 전에 부모가 갖고 있는 값이다. \bar{f} 는 모집단의 평균값이고 f_{max} 는 모집단의 최대값이다. 그리고 $k_1 = 1$, $k_3 = 5$ 이다. 돌연변이율은 1 점 교차와 2 점 교차 후에 얻어지는 고정된 상수값이다.

개체 코드 문자열은 몇몇 비트를 임의의 위치로부터 시작하면서 교환한다. 교환이 기대되는 개수는 다음과 같다.

$$n_c = nP_c.$$

여기서 n 은 모집단 크기이고 P_c 는 유전교차 확률이다.

3) 돌연변이

돌연변이는 변종을 어떤 장소의 개체코드로 특별한 확률에 따라 전이시킨다. 즉 코드를 1에서 0으로 또는 0에서 1로 변화시킨다. 돌연변이를 하도록 기대

되는 부호화 비트는 다음과 같다.

$$B_m = P_m \cdot L \cdot n.$$

여기서 n 은 모집단의 크기이고 L 은 부호화 문자열의 길이 그리고 P_m 은 돌연변이 확률이다.

여기서 간단한 유전자 알고리즘과 유전교차율 그리고 돌연변이율은 일반적으로 연구자들의 경험적인 학습 결과에 의하여 주어지게 된다. 이 연구에서 적용시킨 확률구간은 $P_c = 0.5 \sim 1$, $P_m = 0.001 \sim 0.5$ 구간의 상수이다

2.3 유전자신경망모형

역전과 신경망의 단점은 수렴 속도가 늦고 지역 최소값에 수렴가능성이 높다는 것이다. 반면에 유전자 알고리즘은 발견적 탐색이 가능하고 전역 최소값에 수렴할 수 있다. 그러므로 우리는 신경망에 유전자모형을 결합한 유전자신경망모형을 구축한다: 연결가중값을 부호화하고 신경망의 한계값을 이진코드화한다. 랜덤하게 초기 모집단을 발생시킨 후 유전자 알고리즘을 최적화 계산에 이용한다. 이와 같은 과정을 수식화 하면 다음 단계와 같다.

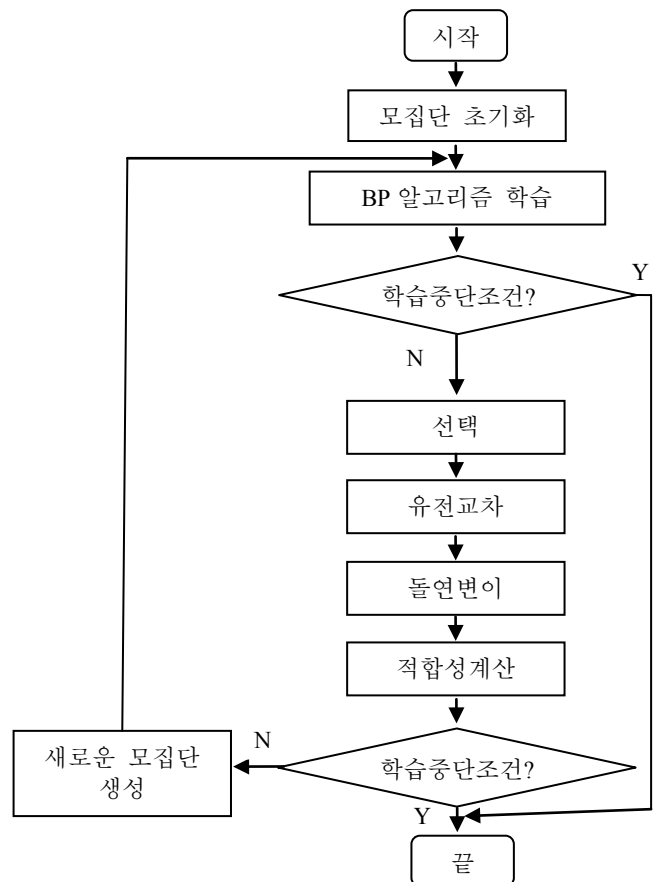


그림 1. 유전자신경망모형의 학습 처리절차

단계 1. 각각의 암호화된 가중값에 대한 이진 암호화 체계를 사용하면서 랜덤하게 초기 모집단을 생성한다.

단계 2. 오차함수를 계산하여 그들의 적합도를 결정한다. 만약 계산된 오차의 한계가 크면 클수록 그 적

합도는 낮아진다.

단계 3. 높은 적합도를 갖는 개체가 유전적으로 다음세대로 된다.

단계 4. 교배와 돌연변이 연산은 모집단의 다음 세대를 만드는 현재 모집단으로 처리한다.

단계 5. 새로운 모집단을 원래의 코드로 바꾸고 네트워크의 오차를 계산한다.

단계 6. 단계 2 에서 단계 5 을 반복한다. 모집단의 적합도가 일정하거나, 오차의 허용한계와 또는 최대 반복회수에 도달하면 학습을 중단한다.

이와 같은 유전자신경망모형의 학습 처리절차는 다음 그림 1 과 같다.

3. 응용 예

시계열예측을 위한 자료는 1966 년부터 2007 년 까지의 한국의 자동차등록 현황 자료(z_k)이다. 이 자료는 자동차를 용도별로 구분하였을 경우의 관용, 자가용 및 영업용에서 각각 승용차, 화물차, 승합차, 이륜차 및 특수차를 포함하는 모든 자동차 등록 현황 자료이다. 학습자료는 1966 년부터 2003 년까지 38 년간의 자료이고 검정자료는 2004 년부터 2007 년까지의 4 년간의 자료를 이용한다.

이 연구에서 적용하는 자료는 다음 식을 이용하여 표준화 과정을 거친 새로운 시계열자료 y_k 이다.

$$y_k = g\left(\frac{z_k - \bar{z}}{s.d}\right)$$

여기서 \bar{z} 와 $s.d$ 는 각각 z_k 의 평균과 표준편차이다. 그리고 $g(u)$ 는 다음과 같은 변환함수이다.

$$g(u) = \frac{1}{1 + e^{-u}}, \quad 0 < g(u) < 1,$$

여기서 변환함수 식은 원자료를 범위 (0,1)로 변환시킨다.

역전과 신경망에서 네트워크의 구조는 3-5-1 구조이고 입력과 출력노드는 선형이며 은닉층에는 시그모이드형 변환함수를 이용한다. 초기 네트워크 연결가중값은 균등분포 $U(0,1)$ 을 따르는 확률난수(RND)를 이용한 식 $20 \times (RND - 0.5)$ 에서 주어진 구간 안에서 발생시킨다. 학습의 중단조건으로 설정한 최종 학습 허용오차의 한계는 0.001 이다.

유전자신경망모형에서 유전교차 확률 P_c 은 0.8 이다. 돌연변이 확률 P_m 은 0.03 이다. 최대 반복회수는 1000 이다. 이 결과에서 알 수 있듯이 유전자신경망을 이용한 예측이 BP 신경망에서 보다 단기 예측에서 상대적으로 더 좋은 예측효율을 보이고 있다

<표 1> 예측결과(RMSE)

기간	유전자신경망	BP신경망
2005	0.00995	0.010102
2006	0.01021	0.010874
2007	0.01024	0.011545
2008	0.05214	0.012571

4. 결론

이 연구에서 우리는 유전자 알고리즘과 신경망을 결

합시킨 유전자신경망모형을 논의하였다. 신경망의 가중값과 학습량을 최적화시키기 위한 유전자 알고리즘의 사용은 전역값 수렴능력 뿐만 아니라 신경망 학습의 효율에서 역전과 알고리즘의 결과와 비교할 때 상대적 효율을 유지하였다. 제시한 예제를 통하여 유전자신경망모형은 시계열자료의 성공적인 예측력을 확인할 수 있었다. 결론적으로 이 연구에서 결합한 유전자신경망모형의 예측결과는 단기 예측인 경우 전통적인 역전과 신경망의 결과보다 더 작은 예측오차를 보였다.

참고문헌

[1]. Chen Ming, Yao Zhengwei, "Classification Techniques of Neural Networks using Improved Genetic Algorithms," Proceedings of Second International Conference on Genetic and Evolutionary Computing, IEEE Computer Society, 115-119, 2008.

[2]. Na-Na Li, Jie Song, Rui-Ying Zhou, Jun-Hua Gu, "A People-Counting System Based on BP Neural Network," Fourth International Conference on Fuzzy Systems and Knowledge Discovery, 24-27 Aug. 2007.

[3]. Wen Jin, Zhao Jia Li, Luo Si Wei, Han Zhen, "The improvements of BP neural network learning algorithm," 5th International Conference on Signal Processing Proceedings, 21-25 Aug. 2000.

[4]. Bo Yang, Xiao-Hong Su, Ya-Dong Wang, "BP neural network optimization based on an improved genetic algorithm," Proceedings of International Conference on Machine Learning and Cybernetics, 4-5 Nov. 2002.

[5]. Wang, Huiqiang Lai, Ji-bao Liu, Xiaowu, "A Quantitative Forecast Method of Network Security Situation Based on BP Neural Network with Genetic Algorithm," Second International Multi-Symposiums on Computer and Computational Sciences, 13-15 Aug. 2007.

[6]. Cao, Liting, "Environment Air Quality Evaluation System Based on Genetic Arithmetic and BP Neural Network," International Conference on Intelligent Computation Technology and Automation (ICICTA), 20-22 Oct. 2008.

[7]. C.X. Yang and Y.F. Zhu, "Time Series Analysis Using GA Optimized Neural Networks," Third International Conference on Natural Computation, pp. 270-276, 2007.

[8]. http://kosis.kr/html/InterStatic/staticCount.jsp?org_id=116&tbl_id=DT_1I29_FRM101&olapAt=Y&vwcd=MT_ZTITLE&path=국토해양부>자동차등록현황조사&keyword=자동차등록현황보고.