

베이지안 진화 학습 알고리즘을 사용한 시계열 예측

조동연, 장병탁
서울대학교 컴퓨터공학부
{dycho, btzhang}@scail.snu.ac.kr

Times Series Prediction by Using Bayesian Evolutionary Algorithms

Dong-Yeon Cho and Byoung-Tak Zhang
School of Computer Science and Engineering
Seoul National University

요 약

본 논문에서는 대개 잡음이 포함되어 있고 불규칙적인 특성을 갖고 있는 시계열 자료에 대해 신경 트리 모델을 사용하여 시계열 예측 문제를 해결하고자 한다. 주어진 시계열 자료에 적합한 구조와 가중치를 갖는 신경 트리를 찾기 위해 베이지안 진화 알고리즘을 적용한 결과, 자료의 개수가 적어 과적합될 우려가 있는 경우 제안된 방법은 모델의 복잡도가 커지는 것을 억제하여 일반화 성능이 급격하게 나빠지는 않는다는 것을 확인하였다.

1. 서론

시계열 자료는 시간에 따라 순차적으로 관찰된 값을 나열한 것으로, 경제학, 공학, 자연과학 등 여러 분야에 매우 다양한 형태로 존재한다[1]. 예를 들면 주가나 환율의 변동, 매시간 관측한 특정 시스템의 출력값, 매일 평균 기온 등이 있다. 이러한 시계열 자료의 내재적인 특징은 인접한 관측값이 서로 의존적이라는 것이다. 시계열 예측은 자료간의 의존성을 분석하여 과거의 값들을 바탕으로 미래의 값을 예측하는 모델을 찾는 과정으로 볼 수 있다.

일반적으로 시계열 자료는 대개 잡음(noise)이 포함되어 있고 불규칙적인 특성을 가지고 있어 과거와 같이 선형 모델로는 시계열 예측을 수행하기 어렵다. 따라서 근래에는 신경 회로망이나[2] 비선형 모델을 사용하여 시계열 예측을 수행하는 경향이 있다. 또한 이러한 모델의 구조와 매개 변수를 결정하기 위하여 진화 연산 기법이 널리 사용되고 있다[3,4].

본 연구에서는 신경 트리(neural tree)를 모델로 설정하였다. 신경 트리는 신경망을 트리 구조로 표현한 것으로 진화 연산 기법 중 하나인 유전자 프로그래밍에 의하여 자연스럽게 진화될 수 있다[5]. 여기서는 베이지안 진화 알고리즘(Bayesian Evolutionary Algorithm, BEA)[6]을 사용하여 주어진 시계열 자료에 알맞은 구조와 가중치를 갖는 신경 트리를 찾는 방법을 제안한다.

논문의 구성은 다음과 같다. 2절에서는 시계열 예측을 위해 사용되는 모델인 신경 트리의 구조를 살펴보고 3절에서는 베지

안 진화 알고리즘을 이용한 신경 트리의 학습 방법을 설명한다. 4절에서는 제안된 방법을 통해 수행한 시계열 예측의 성능을 평가하고 5절에서 결론을 맺는다.

2. 신경 트리의 구조

신경 트리는 다층 신경망을 트리 구조로 표현한 것으로 비단말 노드(nonterminal node)와 단말 노드(terminal node) 그리고 각 노드간의 연결 가중치로 구성된다. 비단말 노드는 신경 유닛(neural unit)을 나타내고 각 신경 유닛의 형태는 기본 함수 집합(function set) F 에서 선택된다. 단말 노드는 단말 집합(terminal node) $T=\{x_1, x_2, \dots, x_n\}$ 에서 선택되는데 여기서 x_i 는 입력 \mathbf{x} 의 i 번째 성분이다. 링크(link) (j,i) 는 노드 j 로부터 노드 i 로의 연결을 나타내며, 각 링크에는 연결 가중치 w_{ij} 가 존재한다. 이러한 구조의 신경 트리에서 루트(root) 노드가 출력 노드가 되고 각 단말 노드들이 입력 노드가 된다.

비단말 노드는 최대 b_{max} 개의 자식 노드로부터 입력을 받아 하나의 출력을 내보내게 되는데, 신경 유닛의 형태에 따라 그 입력을 받아들이는 방법이 다르다. 가장 대표적인 유닛은 시그마(Σ) 유닛과 파이(Π) 유닛으로 다음과 같이 모든 자식 노드로부터의 입력의 가중 합과 가중 곱을 구한다.

$$net_i = \sum_j w_{ij} y_j, \quad net_i = \prod_j w_{ij} y_j \quad (1)$$

여기서 y_j 는 자식 노드 j 로부터의 입력이다. 신경 유닛의 출력은 식 (2)의 시그모이드(sigmoid) 함수에 의하여 계산된다.

$$y_i = f(\text{net}_i) = \frac{1}{1 + e^{-\text{net}_i}} \quad (2)$$

여기서 net_i 는 식 (1)에 의하여 계산된 값이다.

3. 베이지안 진화 알고리즘

3.1 확률 분포의 정의

어떤 신경 트리 A 가 주어졌을 때 그 사후 확률은 다음과 같이 정의된다.

$$P(A|D) \propto P(D|A)P(A) = P(D|\mathbf{w}, k)P(\mathbf{w}, k) \quad (3)$$

$$= P(D|\mathbf{w}, k)P(\mathbf{w}|k)P(k)$$

여기서 k 는 신경 트리의 노드의 개수이고 \mathbf{w} 는 가중치의 벡터이다. 훈련 데이터가 다음과 같이 주어지고,

$$D = \{(\mathbf{x}_c, y_c)\}_{c=1}^N \quad (4)$$

신경 트리 A 에 의하여 구현되는 함수가 f_A 라 하면,

$$y_c = f_A(\mathbf{x}_c) + \varepsilon \quad (5)$$

가 성립한다. 여기서 ε 은 주어진 신경 트리의 예러로 평균이 0이고 표준편차가 σ 인 정규 분포를 따른다고 가정한다. 이 경우 각 데이터가 서로 독립이라고 가정하면 신경 트리의 가능도는 다음과 같다.

$$P(D|\mathbf{w}, k) = \prod_{c=1}^N \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_c - f_{(\mathbf{w}, k)}(\mathbf{x}_c))^2}{2\sigma^2}\right)$$

$$= \left\{\frac{1}{\sqrt{2\pi}\sigma}\right\}^N \exp\left(-\frac{\sum_{c=1}^N (y_c - f_{(\mathbf{w}, k)}(\mathbf{x}_c))^2}{2\sigma^2}\right) \quad (6)$$

가중치 벡터를 이루는 각 성분이 평균이 0이고 표준편차가 1인 정규 분포를 따르며 서로 독립이라고 가정하면 가중치의 사전 확률은 다음과 같다.

$$P(\mathbf{w}|k) = \prod_{j=1}^{k-1} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{w_j^2}{2}\right) \quad (7)$$

$$= \left\{\frac{1}{\sqrt{2\pi}}\right\}^{k-1} \exp\left(-\frac{\sum_{j=1}^{k-1} w_j^2}{2}\right)$$

노드 개수는 평균이 λ 인 포아송(Poisson) 분포를 따른다고 가정하면 노드 개수의 사전 확률은 다음과 같다.

$$P(k-3) = \frac{\lambda^{k-3} \exp(-\lambda)}{(k-3)!}, \quad k=3, 4, \dots \quad (8)$$

여기서 k 의 최소값이 3인 이유는 단말 노드만으로 이루어진 신경 트리는 존재하지 않는다고 가정했기 때문이다. 식 (6), (7), (8)를 식 (3)에 대입하여 정리하면 어떤 신경 트리 A 의 사후 확률은 다음과 같이 정의된다.

$$P(A|D) \propto P(D|\mathbf{w}, k)P(\mathbf{w}|k)P(k)$$

$$= \left\{\frac{1}{\sqrt{2\pi}\sigma}\right\}^N \exp\left(-\frac{\sum_{c=1}^N (y_c - f_{(\mathbf{w}, k)}(\mathbf{x}_c))^2}{2\sigma^2}\right) \quad (9)$$

$$\times \left\{\frac{1}{\sqrt{2\pi}}\right\}^{k-1} \exp\left(-\frac{\sum_{j=1}^{k-1} w_j^2}{2}\right) \frac{\lambda^{k-3} \exp(-\lambda)}{(k-3)!}$$

3.2 베이지안 진화 알고리즘에 의한 신경 트리의 진화

주어진 데이터에 적합한 신경 트리의 구조와 연결 가중치를 찾기 위해 다음과 같이 신경 트리 A_i 로 구성된 개체군 $A(g)$ 를 유지한다.

$$A(g) = \{A_1, A_2, \dots, A_M\} \quad (10)$$

여기서 M 은 개체군의 크기를 나타낸다. 초기 개체군 $A(0)$ 는 모델의 사전 확률에 따라 생성된다. 즉, 각 개체 A_i 의 노드 수는 식 (8)과 같은 포아송 분포에 의하여 결정되고, k 가 주어졌을 때 신경 트리의 가중치는 평균이 0이고 표준 편차가 1인 정규 분포에 의하여 생성된다. 각 세대에서 개체군 내의 모든 신경 트리에 대하여 에러 $E_i(g)$ 가 다음과 같이 계산되고,

$$E_i(g) = \sum_{c=1}^N (y_c - f_{A_i}(\mathbf{x}_c)) \quad (11)$$

이 값을 사용하여 식 (6)의 가능도를 계산한다. 끝으로 식 (9)에 따라 개체군의 각 신경 트리에 대한 사후 확률을 계산한다.

현재의 개체군으로부터 다음 세대의 개체군을 생성하기 위해서 먼저 현재 개체군 내의 부모 트리 A_i 부터 후보 모델 A'_i 를 생성한다. 이러한 후보 모델은 다음의 수용 확률(accept probability)에 의하여 받아들여진다.

$$\alpha(A_i, A'_i) = \min\left\{1, \frac{P(A'_i|D)}{P(A_i|D)}\right\} \quad (12)$$

새로운 개체군을 생성하기 위해 사용되는 유전 연산자에는 크게 교차(crossover) 연산과 돌연변이(mutation) 연산이 있다. 교차 연산은 부모 트리 A_i 와 개체군 내에서 임의로 선택된 또 다른 개체 A_j 에 대하여 임의의 부트리(subtree)를 서로 교환함으로써 새로운 후보 A'_i 을 생성한다. 또한 돌연변이 연산은 임의로 선택된 부트리에 대하여 각 신경 유닛의 형태를 바꾸거나 단말 노드에서 입력의 성분을 다른 것으로 교체한다. 각 개체에 적용할 연산은 교차 확률 p_c 와 돌연변이 확률 p_m 에 의하여 결정된다. 이러한 과정을 후보들의 개체군 $A'(g)$ 의 크기가 L 이 될 때까지 반복한다.

신경 트리의 가중치는 확률적 언덕 오르기(hill-climbing)에 의하여 조정된다. 가중치의 변경은 가중치 벡터 \mathbf{w} 의 모든 성분에 대하여 임의의 순서로 이루어지는데 이 때 사용되는 변경식은 다음과 같다.

$$w'_j = w_j + N(0, 1) \quad j=1, 2, \dots, k_i-1 \quad (13)$$

여기서 k_i 는 개체 A_i 의 노드의 개수이며 $N(0, 1)$ 은 평균이 0이고 표준편차가 1인 정규 분포를 따르는 값이다. 이렇게 하나의 가중치가 변경될 때마다 역시 식 (12)의 수용 확률에 의하여 새로운 후보 모델을 받아들일지 결정하게 된다.

이와 같은 과정을 거쳐서 후보 개체군 $A'(g)$ 의 생성이 끝나면 이 중에서 가장 높은 사후 확률을 갖는 개체 M 개를 선택하여 다음 세대 $A(g+1)$ 을 구성한다.

4. 실험 및 결과

4.1 레이저 데이터

이 데이터는 원격적외선(far-infrared) NH_3 레이저를 통해 만들어진 것으로, 본 실험에서는 1992년 Santa Fe 시계열 예측 시

합[7]을 위해 제공된 데이터 중 짝수번째의 데이터 1000개만을 사용하였다. 훈련 데이터와 검사 데이터의 개수는 각각 500개이다. b_{max} 는 입력의 개수와 같은 3으로 하였으며, 에러의 표준 편차 σ 는 0.05, 트리의 평균 노드 수 λ 는 30, 최대 계산 회수 E_{max} 는 10^6 으로 설정하였다. 또한 모든 실험에서 부모 개체군의 크기와 자식 개체군의 크기는 동일하게 설정하였다.

평균 제곱 에러와 모델의 복잡도에 대한 벌점으로 이루어진 적합도 함수를 사용하는 유전자 프로그래밍 (genetic programming) 방법과 베이지안 진화 알고리즘을 통해 얻은 최적해를 레이저 데이터에 적용한 결과에 표 1에 제시되어 있다 (10번 실험의 평균). 두 방법에 대하여 거의 비슷한 결과를 얻을 수 있었다.

표 1. 레이저 데이터에 대한 NMSE 값

방법	훈련 집합			검사 집합		
	평균±표준편차	최소	최대	평균±표준편차	최소	최대
GP	0.0687±0.0047	0.0610	0.0768	0.1167±0.0095	0.0994	0.1308
BEA	0.0646±0.0066	0.0571	0.0774	0.1185±0.0069	0.1107	0.1345

4.2 태양 흑점 데이터

이 데이터는 1700년부터 1998년까지 연평균 태양 흑점 (sunspot)의 수를 나타낸 것이다¹⁾. 훈련 데이터는 1700년부터 1920년까지의 221개를 사용하였으며 나머지 78개를 검사 데이터로 사용하였다. $b_{max}=4$, $\lambda=40$ 인 것을 제외하면 나머지 설정은 레이저 데이터에 대한 실험과 같다.

레이저 데이터에서와 같은 방법으로 얻은 결과가 표 2에 정리되어 있다. 훈련 집합의 경우에는 두 방법이 크게 차이가 없으나 검사 집합에 대해서는 베이지안 진화 알고리즘을 사용하는 경우가 더 좋은 성능을 보이며 안정된 결과를 얻을 수 있었다.

표 2. 태양 흑점 데이터에 대한 NMSE 값

방법	훈련 집합			검사 집합		
	평균±표준편차	최소	최대	평균±표준편차	최소	최대
GP	0.1434±0.0050	0.1308	0.1489	0.1895±0.0464	0.1267	0.2493
BEA	0.1482±0.0031	0.1415	0.1522	0.1789±0.0121	0.1637	0.1982

4.3 결과 분석

일반적인 진화 연산에 언덕 오르거나 엘리트 주의(elitism)와 같은 지역적 탐색을 추가하여 수행하였을 경우 훈련 집합에 대해서만 과도하게 학습되는(overfitting) 현상이 발생하게 된다. 즉 진화가 이루어짐에 따라 모델의 복잡도가 커지는 경향을 보이며 결국 검사 집합에 대한 일반화 성능은 떨어지게 된다. 그러나 베이지안 진화 알고리즘에서는 비록 지역적 탐색을 수행한다 할지라도 사전 분포가 모델의 복잡도가 커지는 것을 억제하는 역할을 하기 때문에 일반화 성능이 급격하게 나빠지는 현상을 방지한다. 그림 1과 2는 각 실험에서 얻는 최적 모델의 평균 복잡도를 보여주고 있다.

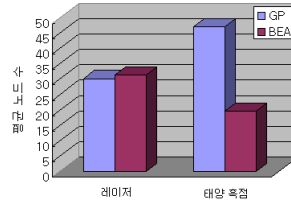


그림 1. 노드 수

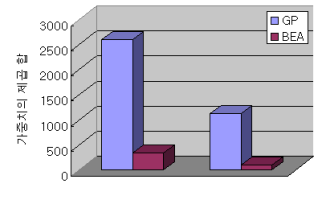


그림 2. 가중치

5. 결론

본 논문에서는 시계열 예측을 위한 모델로 신경 트리를 사용하였고, 주어진 시계열 자료에 적합한 구조와 매개 변수를 갖는 신경 트리를 찾기 위해 베이지안 진화 알고리즘을 사용하였다. 실험 결과 데이터의 개수가 적어 과적합될 우려가 있는 경우에 제안된 방법은 모델의 복잡도가 커지는 것을 억제하여 일반화 성능이 급격하게 나빠지는 현상을 방지한다는 사실을 확인하였다.

감사의 글

본 연구는 과학기술부 뇌연구개발사업(BR-2-1-G-06)에 의하여 일부 지원되었음.

참고문헌

- [1] Box, G.E.P, Jenkins, G.M, and Reinsel, G.C., *Time Series Analysis: Forecasting and Control*, 3rd Ed, Prentice-Hall, 1994.
- [2] Wan, E.A., "Time series prediction by using a connectionist network with interanl delay lines," *Time Series Predictions: Forecasting the Future and Understanding the Past*, pp. 195-217, Addison-Wesley, 1993.
- [3] Rao, S.S., Hemanth, K.B., and Chellapilla, K., "Evolving nonlinear time-series models using evolutionary programming," *Proceedings of the 1999 Congress on Evolutionary Computation*, Vol 1, pp. 236-243, 1999.
- [4] Gomez-Ramirez, E., Poznyak, A., Gonzalez-Yunes, A., and Avila-Alvarez, M., "Adaptive architecture of polynomial artificial neural network to forecast nonlinear time series," *Proceedings of the 1999 Congress on Evolutionary Computation*, Vol 1, pp. 317-324, 1999.
- [5] Zhang, B.-T., Ohm, P., and Mühlenbein H., "Evolutionary Induction of Sparse Neural Trees," *Evolutionary Computation*, Vol. 5, No. 2, pp. 213-236, 1997.
- [6] Zhang, B.-T., "A Bayesian framework for evolutionary computation," *Proceedings of the 1999 Congress on Evolutionary Computation*, Vol 1, pp. 722-728, 1999.
- [7] Hübner, H., Weiss, C.O., Abraham, N.B., and Tang, D., "Lorenz-like chaos in NH₃-FIR laser," *Time Series Prediction: Forecasting the Future and Understanding the Past*, pp. 73-104, Addison-Wesley, 1993.

1. 이 자료는 Sunspot Index Data Center (SIDC)로부터 얻을 수 있다.
<http://www.astro.oma.be/SIDC/>