

진화 신경트리의 결합에 의한 시계열 예측

정 제 균^o 장 병 탁
E-mail: {jgjoung, btzhang}@scail.snu.ac.kr
서울대학교 컴퓨터공학과

Time Series Prediction by Combining Evolutionary Neural Trees

Je-Gun Joung^o Byoung-Tak Zhang
Dept. of Computer Engineering, Seoul National University

요 약

신경트리(evolutionary neural trees)는 트리 구조의 신경망 모델로서 진화 알고리즘으로 학습하기에 적합한 구조이다. 본 연구에서는 진화 신경트리를 시계열 예측에 적용하였다. 시계열 데이터는 대개 잡음이 포함되어 있으며 동역학적인 특성을 지닌다. 본 논문에서는 견고한 예측 결과를 획득하기 위해 한 개의 신경트리가 아닌 여러개의 신경트리를 결합하여 예측 모델을 구성하는 committee machine을 소개한다. 출력 패턴간에 correlation이 최소가 되도록 상이한 신경트리를 선택하여 결합함으로써 모델 결합 효과를 최대화하는 방법을 사용하였다. 인공적인 잡음을 포함한 시계열 예측 문제와 실세계 데이터에 대한 실험에서 예측에 대한 정확도가 단일 모델을 사용한 경우 보다 향상되었다.

1. 서 론

시계열 예측 문제는 전형적으로 동역학적인 특성을 지니고 있어서 견고한 예측이 중요하다. 예측을 하는 모델들의 견고성을 증가시키기 위한 한가지 방법은 모델들의 앙상블(ensemble) 또는 committee를 사용하는 것이다.

committee machine은 다수 모델들이 평균에 의하여 분산(variance)을 감소시키게 함으로써 예측의 정확성을 증가시킨다. 모델들의 생성에는 많은 국소 최소치(local minima)를 가기기 때문에 평균에 의한 방법이 효과적일 것이다. 많은 국소 최소치들이 존재함에도 불구하고 모든 전역적 최적화(global optimization) 방법들은 서로 다른 초기 점으로부터 발생하는 무수한 랜덤을 나타내는 다음 상태까지 최적의 파라미터(parameters)들을 생산한다. 이러한 랜덤은 모델들의 에러(error)가 서로 다르게 되도록 하는 경향이 있다. 따라서 모델들은 학습하지 않은 입력에 대하여 서로 다른 일반화(generalization) 에러를 만들어 낸다. 그리고, 이러한 모델들이 결합된 committee에 의하여 만들어진 협조적 의사결정은 어느 단일 모델이 만들어 낸 결정 보다 다소 에러가 적을 것이다.

진화 계산적 관점에서 committee machine은 물론 자연스럽다. 모든 진화 알고리즘은 그들의 결과로 다수의 개체를 만들어 낸다. 사실, 이전에 연구된 결과로 진화 알고리즘에 의해 진화된 다수 개체들을 결합한 대표적인 두 방법들이 있다. 첫번째 방법으로 EPNNet은 다층 퍼셉트론(perceptron)을 진화하는 진화 프로

그래밍을 사용하였다[1]. 그리고, 두번째 방법인 MGP는 진화 신경트리(ENTs)를 만들어 내는 유전자 프로그래밍을 사용한다[2]. 이들 두 가지 방법은 다수 모델들을 결합함으로써 진화 알고리즘에 의해 만들어진 모델들의 일반화 성능을 향상시켰다.

본 논문에서는 회귀(regress)문제를 위하여 committee를 형성하기 위한 모델들의 새로운 가중치 학습 방법을 통해서 혼합 유전자 프로그램(mixing genetic programming)의 개념을 확장한다.

논문의 구성은 다음과 같다. 2장에서는 시계열 예측 데이터의 모델링을 위한 신경트리의 구조와 신경트리를 진화하기 위한 알고리즘에 대하여 기술한다. 3장에서는 진화된 각각의 신경트리로부터 committee의 형성 방법에 대하여 언급한다. 그리고 4장에서는 잡음이 포함된 사인 웨이브(sine-wave)와 레이저 데이터에 대한 실험 결과를 보인다. 마지막 5장에서는 결론을 제시한다.

2. 진화 신경트리

2.1 신경트리

$NT(d, b)$ 최대 깊이(depth)가 d 이고 최대 가지(branch)가 b 인 모든 가능한 트리의 집합이라고 하자. 그러면 비종단(nonterminal) 노드들은 신경 유닛으로 표현되고, 뉴런 형(neuron type)은 기본적인 함수집합 $F = \{\text{뉴런 형들}\}$ 의 요소들이 된다. 각각의 종단 노드는 종단집합 $T = \{x_1, x_1, \dots, x_n\}$ 로 명시된다. 여기서 x_i 는

i 번째 외부입력 x 이다. 고차원 망(higher-order networks)을 진화시키기 위해서는 다음과 같은 두 형태의 유닛을 고려할 수 있다.

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

또한, 뉴런의 출력은 시그모이드(sigmoid) 함수

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

로 전환된다.

2.2 신경트리의 진화

진화 모델을 명시하기 위하여 A 개체로 구성된 M 개체집단을 유지한다. 그리고, 각각의 개체는 신경망을 표현한 신경트리이다.

$$A(g) = \{A_1, A_2, \dots, A_M\} \quad (3)$$

초기 개체집단 $A(0)$ 는 무작위로 생성된다. 각각의 진화횟수 g 에서 신경트리의 적합도 $F_i(g)$ 은 평가되고 상위 $\tau\%$ 가 교배(mating)를 위하여 선택된다. 다음 진화횟수 $A(g+1)$ 은 부트스트랩의 교환에 의하여 다시 M 개체를 생성한다.

신경트리의 가중치는 확률적인 언덕오르기 탐색 (stochastic hill-climbing search)로 조정된다[3]. 스텝 크기 Δw 는 임의의 값 $\epsilon \in [0, 1]$ 은 다음 식에 의해 결정된다.

$$\Delta w = R \cdot 2^{-\epsilon \cdot K} \quad (4)$$

그리고, 각각의 개체 A_i 의 적합도 F_i 는 다음과 같이 정의된다.

$$F_i(g) = F(DA_i^g) = E(DA_i^g) + \alpha(g)C(A_i^g) \quad (5)$$

여기서 $\alpha(g)$ 는 에러와 복잡도의 균형을 유지하기 위한 오감 인자(oaam factor)이다[4].

3. 진화 신경트리의 Committee의 형성

먼저, 훈련 데이터에 대해서 committee에 의한 출력과 실제값과의 MSE(mean squared error)를 최소화 하도록 각각의 committee member에 대한 가중치를 발견하기 위하여 학습을 한다. 여기서 가중치는 일반화된 앙상블 방법(GEM: generalized ensemble method)에 의해서 결정된다[5]. 일반화된 앙상블 방법은 다음 식과 같이 선형적 결합에 근거를 하고 있다.

$$f_{GEM}(x) \equiv \sum_{i=1}^K v_i f_i(x), \quad (5)$$

여기서 v_i 는 $\sum v_i = 1$ 의 제약사항을 따른다. 목표함수 $y(x)$ 에 관한 MSE를 최소화하는 v_i 를 선택해야 할 것이다. 만약 i 번째 member의 에러가

$$e_i(x) = y(x) - f_i(x) \quad (6)$$

이고, 상관 행렬(correlation matrix)은

$$C_{ij} = E[e_i(x) e_j(x)], \quad (7)$$

으로 표현되면, 다음 식

$$MSE[\bar{f}] = \sum_{i,j} \alpha_i \alpha_j C_{ij} \quad (8)$$

을 최소화해야 한다.

각각의 v_i 는

$$v_i = \frac{\sum_{j=1}^m C_{ij}^{-1}}{\sum_{k=1}^m \sum_{j=1}^m C_{ij}^{-1}} \quad (9)$$

로 주어진다.

여기서 C_{ij} 는 committee member들 f_i 와 f_j 로부터 생성된 에러들의 상관 행렬의 인자들이다.

이러한 접근의 이론적인 배경은 bias-variance 분석에 의해 설명되어질 수 있다. 가중치가 부여된 앙상블 평균(weighted ensemble average)은 committee의 최종 출력값

$$\bar{f}(x) = \sum_k v_k f_k(x) \quad (10)$$

에 의해 정의될 수 있다. 여기서 v_i 는 결합에 대한 가중치이다.

입력값 x 에 대해서 앙상블 에러는 $\epsilon(x)$, k 번째 member는 $\epsilon_k(x)$, 그리고 k 번째 member의 모호성(ambiguity)은 $a_k(x)$ 로 다음 식과 같이 정의된다.

$$\epsilon(x) = (y(x) - \bar{f}(x))^2 \quad (11)$$

$$\epsilon_k(x) = (y(x) - f_k(x))^2 \quad (12)$$

$$a_k(x) = (f_k(x) - \bar{f}(x))^2. \quad (13)$$

앙상블 에러는 다음 식

$$\epsilon(x) = \bar{\epsilon}(x) - \bar{a}(x) \quad (14)$$

쓰여질 수 있다.

여기서 $\bar{\epsilon}(x) = \sum_k v_k \epsilon_k(x)$ 는 각각 예측자의 에러들의 평균이고, $\bar{a}(x) = \sum_k v_k a_k(x)$ 는 앙상블에 있어서 출력들의 분산인 모호성의 평균이다. 이러한 모든 공식들은 입력분포 $P(x)$ 에 관해서 평균으로 나타낼 수 있으며, GEM을 얻을 수 있다.

$$\epsilon = \bar{\epsilon} - \bar{a} \quad (15)$$

여기서 $\epsilon(x)$ 의 $P(x)$ 에 관한 평균은 간단하게 ϵ 으로 나타내어질 수 있다. 첫 번째 항은 각각의 예측자의 일반화 에러의 가중치가 부여된 평균이고, 두 번째 항은 모호성의 평균이다. 식 (15)에서의 특징은 member들의 일반화 에러에 근거한 항과 member간에 모든 상관관계를 포함한 항으로 나누어졌다는 것이다. 여기서 알 수 있는 것은 예측자들의 에러는 고정되어 있는 반면 서로 예측자의 의견이 서로 다르면 더 낮은 에러를 유도할 수 있다.

1. 모델들의 진화
2. 반복횟수 N_c 만큼 반복
 - 2.1 committee member들의 후보 집합을 형성
 - 2.2 후보 집합으로부터 member들을 선택하여 committee를 형성
 - 2.3 효과적인 의사결정을 위하여 GEM을 적용
 - 2.4 우수한 committee를 형성
3. 최종적으로 우수한 committee를 이용하여 문제 해결

그림 1. 진화 신경트리에 대하여 최적의 committee의 구축

진화 신경트리의 최적 committee의 구축에 대한 알고리즘이 그림 1에 보여지고 있다. 이 알고리즘은 진화가 끝난 후에 적용

된다. 결합을 위한 조합의 수가 너무 크기 때문에 작은 크기로 선택공간을 줄이기 위해서 후보 집합을 생성한다. committee member들의 선택은 지수 비례적으로 적합도에 근거하여 확률적으로 선택된다.

4. 실험 결과

두 가지 실험에 사용된 매개 변수는 다음과 같다. 개체집합의 수는 200, 총 진화횟수는 50, 반복횟수(iteration)는 50, committee의 후보 집합 크기는 10, committee member들의 수는 3이다.

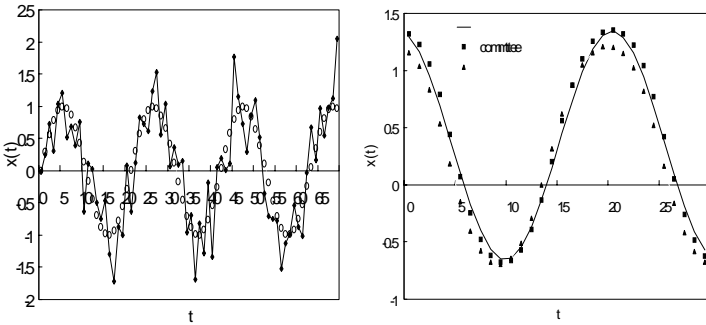


그림 2. 훈련 데이터로 사용된 30% white noise에 의한 사인 웨이브(왼쪽)와 사인 웨이브에 대해서 단일 신경트리와 committee에 의한 신경트리의 예측된 값(오른쪽).

잡음을 포함한 시계열 예측에 관한 문제로 그림2에서 보이고 있다. 30%의 white noise에 의해 생성시켜서 훈련한 다음, 테스트 데이터에 대해서 단일 신경트리와 committee에 의한 신경트리의 예측된 값을 비교하였다. 여기서는, $x(t-10)$, $x(t-9)$, ..., $x(t-1)$ 의 계속되는 10개의 변수가 입력으로 사용되고 $x(t)$ 를 예측한다. committee에 의한 신경트리의 예측된 값이 더 사인 웨이브에 근접함을 알 수 있다.

그림 3은 물리학 실험실의 far-infrared NH_3 [6] 레이저로부터 생성된 chaotic intensity fluctuations에 대해서 2000개의 측정값과 신경트리의 committees에 의한 예측값을 보이고 있다. 훈련데이터와 테스트데이터에 대해서 $x(t-3)$, $x(t-2)$, $x(t-1)$ 의 계속되는 3변수가 입력으로 사용되고 $x(t)$ 를 예측한다.

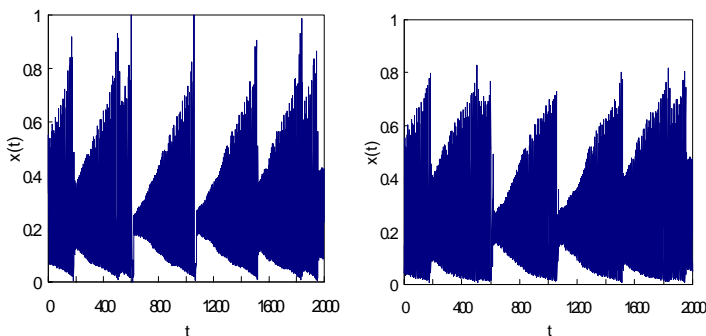


그림 3. 실제 레이저 데이터(왼쪽)와 신경트리의 committees에 의한 성능: 처음 1000데이터 포인트들은 훈련에 대해 사용되었고 나머지 1000은 예측에 대해서 사용함.

그림 4는 단일 신경트리의 각 포인트에 대한 에러에 대하여 전체 진화횟수를 통해 축적된 값에서 같은 방법으로 committee에

의한 것에 대한 차를 보이고 있다. 그림에서 보여주는 바와 양수 부분이 많다는 것은 단일 신경트리에 의해 예측된 에러가 committee에 의한 것보다 크다는 것은 의미한다.

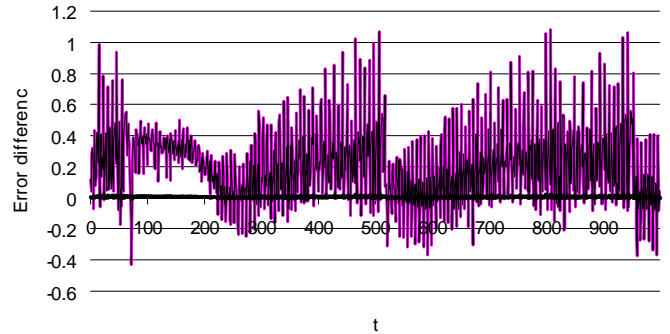


그림 4. 단일 신경트리의 각 포인트에 대한 에러와 committee에 의한 신경트리의 각 포인트에 대한 에러에 대해서 전체 진화횟수에 대해 축적된 값들의 차.

5. 결론

본 논문에서는 시계열 예측에 대해서 ENTs(진화 신경트리의 committee)가 단일 ENTs보다 예측의 정확도를 향상시킬 수 있음을 보였다. 특히, committee member들간의 상관관계를 고려하여 새로운 가중치 부여 방법을 제시하였다. 이러한 방법은 혼합 유전자 프로그램의 개념을 회귀 문제로 확장한 것이다.

bias-variance분석은 committee에 있어서 그들의 bias에러가 줄어들어 따라 상관관계가 없는 많은 변이를 가진 예측자들이 결합되도록 유도함을 보인다. 이러한 관점에서 진화 신경트리는 그들의 구조와 가중치가 다양하기 때문에 committee machines를 형성하기 위한 자연스러운 조건을 제공한다.

감사의 글: 본 연구는 한국과학재단 핵심전문연구(과제번호 981-0920-107-2)와 과학기술부 뇌연구개발사업(BR-2-1-G-06)에 의하여 일부 지원되었음.

참고문헌

- [1] X. Yao and Y. Liu, Making use of population informatin in evolutionary artificial neural networks, *IEEE Transactions on Systems, Man and Cybernetics*, 28B(3): 417-425, 1998.
- [2] B. T. Zhang, J. G. Joung, Enhancing robustness of genetic programming at the species level, *Genetic Programming Conference (GP-97)*, Morgan Kaufmann, pp. 336-342, 1997.
- [3] H. Mühlenbein, and D. Schlierkamp-Voosen, The science of breeding and its application to the breeder genetic algorithm, *Evolutionary Computation*, 1(4): 335-360, 1994.
- [4] B. T. Zhang, P. Ohm and H. Muehlenbein, Evolutionary Induction of sparse neural trees, *Evolutionary Computation*, 5(2): 213-236, 1997.
- [5] M. P. Perrone and L. N. Cooper, When networks disagree: Ensemble methods for hybrid neural networks, in *Artificial Neural Networks for Speech and Vision*, Chapman & Hall, pp. 126-142, 1994.
- [6] U. Hübner, C. O. Weiss, N. B. Abraham, and D. Tang, Lorenz-like chaos in NH_3 -FIR laser, A. S. Weigend, N. A. Gershenfeld, (eds.) *Time Series Prediction: Forecasting the Future and Understanding the Past*, Addison-Wesley, pp. 73-104, 1993.