능동적인 데이터 선택에 의한 Committee Machine의 진화적 학습

정제균, 조동연, 장병탁
서울대학교 컴퓨터공학과

Evolutionary Learning of Committee Machine by Adaptive Data Selection

Je-Gun Joung, Dong-Yeon Cho, Byoung-Tak Zhang
Dept. of Computer Engineering, Seoul National University

요 약

Committee machine의 목적은 서로 다른 다수 모델을 결합함으로써 시스템의 일반화 성능을 향상시키는데 있다. 이러한 committee machine이 최대의 효과를 발휘하기 위해서는 각 모델들간의 에러가 uncorrelated 되어야 한다. 본 논문에서는 모델들을 전문화시키도록 진화를 유도하여 correlation을 최소화하는 방법을 제시한다. 각각의 모델들은 서로 다른 데이터 집합을 학습하고, 이러한 데이터를 선택적으로 유전시키는 방식으로 진화를 함으로써 committee를 형성하도록 한다.

1. 개 요

일반적으로 여러 모델들의 결합은 예측의 정확도를 향상시킨다고 알려져 있다. 이러한 목적을 위하여 여러 가지 committee machine 기법들이 개발되었다. 대부분의 경우, 모델들은 독립적으로 훈련시켜서 예측할 때 결합한다. 그러한 이유는 committee machine이 최대의 효과를 발휘하기 위해서 각 모델들간의 에러가 uncorrelated 되어야 하기 때문이다. 진화 계산적 관점에서 진화 알고리즘이 다수의 개체를 만들기 때문에 committee machine을 만들기에 자연스럽지만, 기본적인 알고리즘은 각 개체들이 uncorrelated 되도록 진화되지 않는다.

Committee machine을 진화 알고리즘에 도입한 이전의 연구들에 대하여 대표적인 두 방법들이 있다. 첫번째 방법으로 EPNet은 다층 퍼셉트론(perceptron)을 진화하는 진화 프로그래밍을 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.

하지만, 이들 방법들은 committee machine을 형성하는데 있어서 두 가지 단점이 있다. 첫번째, 진화 알고리즘은 기본적으로 같은 데이터를 학습하기 때문에 개체들간의 다양성이 유지되기 어렵다. 두 번째는 많은 데이터로 학습하기 때문에 진화의 속도가 느리다. 개체들간의 다양성을 유지하기 위해서는 서로 상이한 구조를 가져야 하고, 개체의 파라미터들 역시 다양한 값을 가져야 한다. 학습의 속도에 있어서는 데이터를 활용하는 방법에 따라 가속화시킬 수 있음을 보여주는 연구가 있었다[3].

본 논문에서는 committee를 형성하기 위하여 데이터의 선택적 학습 방법을 통해서 committee 구성원을 전문화시킴으로써 일반화 성능과 진화 속도를 동시에 향상시키기 위한 방법을 제시한다.

논문의 구성은 다음과 같다. 2장에서는 committee를 형성하기 위한 전체적인 알고리즘을 기술하고, committee 구성원을 전문화시키기 위한 훈련데이터의 활용 방법과 committee의 구성 방법을 기술한다. 3장에서는 MacKey-Glass시계열 예측 문제를 통하여 committee의 예측에 대한 평균 bias와 variance를 분석해 보고, sunspot 시계열 예측 문제를 통해서 일반화 성능과 진화 속도를 기존의 방법과 비교해 본다. 마지막 4장에서는 결론을 제시한다.

2. Committee machine의 진화적 학습

2.1 학습 알고리즘

Committee를 형성하기 위하여 능동적인 데이터 선택 기법을 이용하게 된다. 본 논문에서의 능동적인 데이터 선택 기법의 기본 개념은 각각의 개체에 적합한 훈련 데이터를 학습하게 것이다. 각 개체들은 자신의 훈련 데이터 집합을 유지하면서 유전연산자에 의해 자식을 생성할 때 동시에 자신이 학습한 데이터들을 유전연산자에 의해 자식의 데이터도 생성한다.

그림 1에 committee machine을 위한 진화의 전체적인 흐름을 보여주고 있다. 개체 집단의 크기는 고정적인 크기로 진화하는 반면, 개체들의 데이터는 증가치 만큼 단조 증가하면서 진화한다. 각 진화횟수에서 committee $C(g)$는 적합도가 우수한 개체 M개를 이용하여 구성한다.

---

1. 개체 집단 $A(0)$을 M의 크기로 초기화
2. 개체들에 대한 데이터 집합 $D(0)$을 n_0로 초기화
   (각각의 데이터 $D_i(0)$는 전체 데이터 집합에서 무작위로 샘플링)
3. 진화 횟수 $g \leftarrow 1$로 설정
4. 최대 진화횟수 g_{max}만큼 반복
   4.1. $D_i(g)$에 대해서 개체 $A_i(g)$의 적합도를 측정
   4.2. M개의 개체가 만들어 질 때 까지 반복
      4.2.1. 두 부모 $A_i(g)$와 $A_j(g)$에 대해서 교차와 복제연산에 의해서 두 자식 $A_i(g+1)$와 $A_j(g+1)$을 생성
      4.2.2. 두 부모의 데이터 집합 $D_i(g)$와$D_j(g)$에 대해서 교차연산에 의해서 두 자식의 데이터 $D_i(g+1)$와 $D_j(g+1)$를 생성
   4.3. committee개체집단 $C(g)$를 M개 크기로 구성
   4.4. committee개체 $C_i(g)$의 가중치 v_i 생성
   4.5. committee $C(g)$의 적합도 측정
   4.6. $n_{g+1} = n_g + \lambda$
   4.7. $g \leftarrow g + 1$
5. 전체 진화횟수 g_{max} 중 적합도가 가장 우수한 committee 선택

그림 1. 능동적인 데이터 선택에 의한 committee machine의 진화적 학습의 개요.

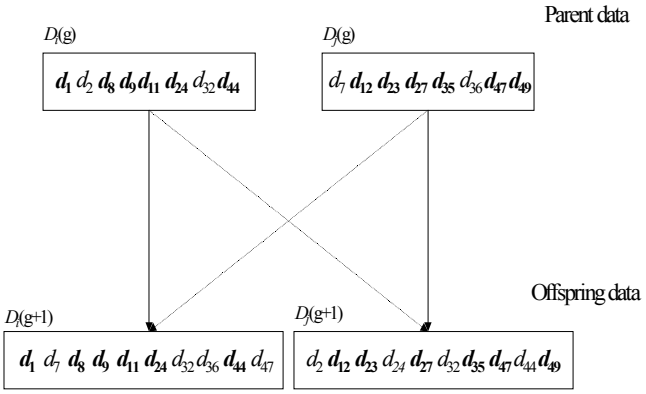

그림 2. 능동적인 데이터 선택의 개념도.

2.2 개체들의 전문화를 위한 훈련 데이터 선택

그림 2는 능동적 데이터 선택의 개념을 나타내고 있다. 두 부모의 데이터 집합을 각각 $D_i(g)$와 $D_j(g)$라고 하면 두 자식의 데이터 집합 $D_i(g+1)$과 $D_j(g+1)$이 교차연산에 의해서 생성된다.

$$D_i(g+1) = D_i^e(g) \bigcup D_j^d(g) \quad (1)$$

$$D_j(g+1) = D_j^e(g) \bigcup D_i^d(g) \quad (2)$$

여기서 $D_i^e(g)$는 자식 i의 데이터 집합 중에서 자식 i에 대한 에러가 적은 데이터를 선택한 집합을 의미하고, $D_j^d(g)$는 자식 j의 데이터 집합 중에서 자식 j에 대한 에러가 많은 데이터를 선택한 집합을 의미한다. 그림 2에서는 $D_i^e(g) = \{d_1, d_8, d_9, d_{11}, d_{24}, d_{44}\}$다. 자식의 데이터 집합의 크기는 $n_{g+1} = n_g + \lambda$로 부모의 데이터 크기에 λ만큼 증가한다. 그래서 데이터 집합 D^d의 크기는 진화 횟수가 증가함에 따라 단조 증가하게 된다. D^e와 D^d의 데이터 집합의 크기에 대한 비율은 9:1로 에러가 적은 데이터를 많이 선택한다.

2.3 Committee 형성

훈련 데이터에 대해서 committee에 의한 출력과 실제값과의 MSE(mean squared error)를 최소화하도록 각각의 committee 구성원에 대한 가중치를 발견하기 위하여 학습을 한다. 여기서 가중치는 일반화된 앙상블 방법(GEM: generalized ensemble method)에 의해서 결정된다[4]. 일반화된 앙상블 방법은 다음 식과 같이 선형적 결합에 근거를 하고 있다.

$$f_{GEM}(\mathrm{x}) \equiv \sum_{i=1}^{N} v_i f_i(\mathrm{x}), \quad (3)$$

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

$$e_i(\mathrm{x}) = y(\mathrm{x}) - f_i(\mathrm{x}) \tag{4}$$

이고, 상관 행렬(correlation matrix)은

$$C_{ij} = E[e_i(\mathrm{x})\, e_j(\mathrm{x})], \tag{5}$$

으로 표현되면, 다음 식

$$MSE[\bar{f}] = \sum_{i,j} \alpha_i \alpha_j C_{ij}. \tag{6}$$

을 최소화해야 한다.
각각의 v_i는

$$v_i = \frac{\sum_{j=1}^{M} C_{ij}^{-1}}{\sum_{k=1}^{M}\sum_{j=1}^{M} C_{ij}^{-1}} \tag{7}$$

로 주어진다.
여기서 C_{ij}는 committee member들 f_i와 f_j로 부터 생성된 에러들의 상관 행렬의 인자들이다.

3. 실험결과와 분석

3.1 MacKey-Glass 시계열 예측 문제

자주 사용되는 시계열 예측에 대한 문제는 아래에 있는 MacKey-Glass 미분식이다.

$$\frac{dx(t)}{dt} = \frac{ax(t-\tau)}{1+x^{10}(t-\tau)} - bx(t) \tag{8}$$

여기서 a = 0.2, b= 0.1 그리고 τ = 17이다[6,7]. x(0)에서 x(17)까지의 초기값은 다음과 같다.

x_0 = 1.000, x_1 = 1.002, x_2 = 1.000, x_3 = 0.992
x_4 = 0.983, x_5 = 1.973, x_6 = 1.966, x_7 = 1.963
x_8 = 1.972, x_9 = 1.987, x_{10} = 1.004, x_{11} = 1.025
x_{12} = 1.049, x_{13} = 1.077, x_{14} = 1.101, x_{15} = 1.123
x_{16} = 1.139, x_{17} = 1.133

그림 3에서 시간 t 에 대한 각각의 값들을 보여주고 있다. 입력은 $x(t-10)$, $x(t-9)$, ..., $x(t-1)$의 10개 데이터 점으로 구성되고, 출력은 $x(t)$이다. 훈련 데이터 점은 100개이고 테스트 점은 400개이다. 실험을 위한 파라메터에 대하여 개체집단의 크기는 100, 진화 횟수는 50,

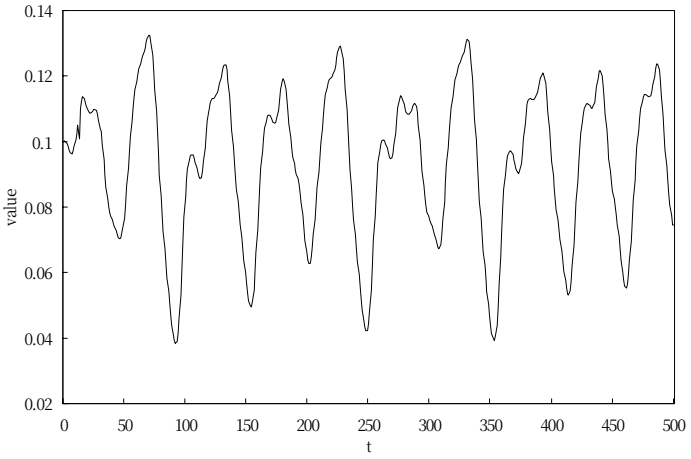

그림 3. Mackey-Glass 시계열 예측.

초기 개체들의 데이터 집합크기 n_0은 40으로 설정하였다. 그리고, 신경트리(neural trees)를 진화 모델로써 이용하였다[5].

그림 4는 능동적 데이터 선택에 의해 형성된 committee의 예측된 결과를 보여주고 있다. 표 1 은 전체 데이터에 의해 진화한 committee와 능동적 데이터 선택에 의해 진화한 committee에 대한 integrated bias, integrated variance, integrated covariance를 비교하고 있다. 주어진 테스트 집합의 패턴 n에 대하여 committee와 각 개체 i 에 대한 평균의 출력을 각각 $\overline{F}(n)$과 $\overline{F}_i(n)$ 로 나타낼 수 있다.

$$\overline{F}(n) = \frac{1}{K}\sum_{k=1}^{K} F^{(k)}(n), \tag{9}$$

$$\overline{F}_i(n) = \frac{1}{K}\sum_{k=1}^{K} F_i^{(k)}(n) \tag{10}$$

여기서 $F^{(k)}(n)$와 $F_i^{(k)}(n)$은 각각 k번째 시뮬레이션에 대한 committee와 각 개체의 출력이다. 그리고 K는 시뮬레이션 수이다. integrated bias, integrated variance, integrated covariance는 각각 다음과 같은 식으로 나타낼 수 있다.

$$E_{bias} \equiv \frac{1}{N}\sum_{n=1}^{N}(\overline{F}(n) - d(n))^2 \tag{11}$$

$$E_{var} \equiv \sum_{i=1}^{M}\frac{1}{N}\sum_{n=1}^{N}\frac{1}{K}\sum_{k=1}^{K}\frac{1}{M^2}(F_i^{(k)}(n) - \overline{F}_i(n))^2 \tag{12}$$

$$E_{cov} \equiv \sum_{i=1}^{M}\sum_{j=1,\, i\neq j}^{M}\frac{1}{N}\sum_{n=1}^{N}\frac{1}{K}\sum_{k=1}^{K}\frac{1}{M^2}(F_i^{(k)}(n) - \overline{F}_i(n))(F_j^{(k)}(n) - \overline{F}_j(n)) \tag{13}$$

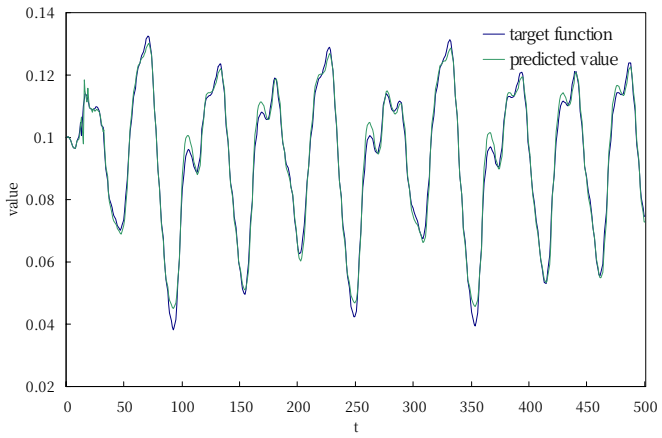

그림 4. Mackey-Glass 시계열 예측에 대한 능동적 데이터 선택에 의해 진화한 committee의 훈련집합과 테스트 집합의 출력.

능동적 데이터 선택에 의해 진화한 committee에 대한 bias, variance는 전체 데이터에 의해 진화한 committee에 비해서 낮은 값을 나타내었다. 그리고, covariance가 상대적으로 낮은 값을 나타내기 때문에 각각의 개체들은 상이하다고 볼 수 있다.

	committee (base data set)	committee (active data selection)
E_{bias}	1.12×10^{-5}	8.97×10^{-6}
E_{var}	2.39×10^{-5}	7.75×10^{-6}
E_{cov}	5.93×10^{-3}	4.29×10^{-5}

표 1. Bias와 variance비교: 전체 데이터에 의해 진화한 committee, 능동적 데이터 선택에 의해 진화한 committee

3.2 Sunspots 시계열 예측 문제

그림 5에 있는 Wolfe sunspot 데이터는 1700년부터 최근까지의 연도에 대하여 sunspot의 수로부터 만들어진 것이다. 이 데이터는 몇몇 신경망의 예측에 대한 연구에 사용된 적이 있다[8,9,10]. 이 문제에 대한 실험에서는 앞에서 제시한 알고리즘의 일반화 성능과 진화 속도를 평가하기로 한다.

본 실험에서 1700에서 1920년까지는 훈련 데이터로 사용하였고, 1921에서 1950년까지는 테스트 데이터로 사용하였다. $x(t-5)$, $x(t-4)$, ..., $x(t-1)$의 5개 입력에 대하여 $x(t)$의 출력을 예측한다. 실험을 위한 파라메터에 대하여 개체집단의 크기는 100, 진화 횟수는 100, 초기 개체들의 데이터 집합크기 n_0는 40으로 설정하였다.

그림 6은 능동적 데이터 선택에 의해 진화한 신경트리의 committee에 대한 일반화 성능을 보이고 있다. 피크(peak) 부분을 제외한 모든 부분에서 비교적 좋은 예측값을 나타내었다. 표 2는 10번의 시뮬레이션에 대하여 세 가지 방법에 대한 예측 에러를 비교한 것이다. 능동적 데이터 선택에 의해 진화한 committee는 진화의 속도의 개선과 동시에 일반화 성능도 우수함을 보여주고 있다. 그림 7은 우수한 하나의 개체, 전체 데이터에 의해 진화한 committee, 능동적 데이터 선택에 의해 진화한 committee에 대한 진화 횟수에 따른 적합도를 비교한 것이다. 능동적 데이터 선택에 의해 진화한 committee은 비교적 초기에 수렴하는 특성을 보이고 있다. 그림 8은 그림 7의 결과를 진화 횟수대신 진화시간에 따른 적합도의 변화를 보여주고 있다. 진화가 초기에 빨리 가속화됨을 알 수 있다.

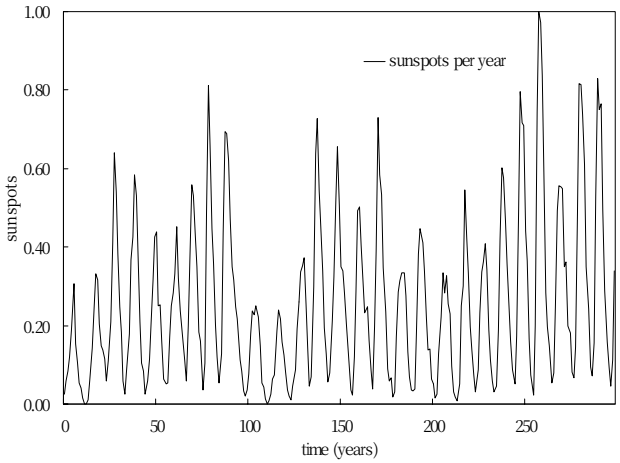

그림 5. 1700년 이후 관찰된 sunsopts의 평균 수.

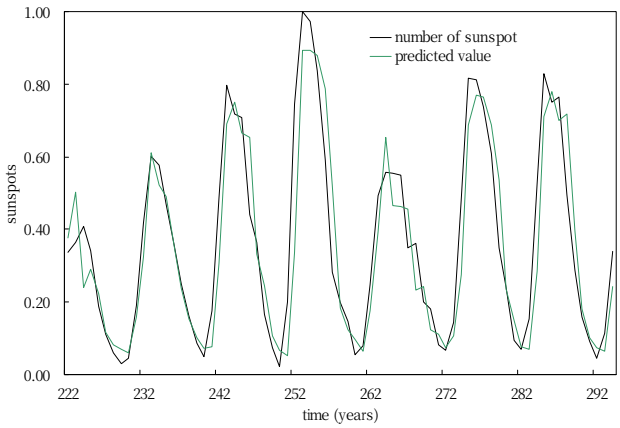

그림 6. Sunspot 시계열 예측에 대한 능동적 데이터 선택에 의해 진화한 committee.

	prediction error			
	MEAN	SD	MIN	MAX
Best NT	0.0252	0.0088	0.0113	0.0359
Committee (base data set)	0.0155	0.0060	0.0103	0.0253
Committee (active data selection)	0.0148	0.0029	0.0099	0.0195

표 2. Sunspot 시계열 예측에 대한 테스트 에러 비교: 단일 개체, 전체 데이터에 의해 진화한 committee, 능동적 데이터 선택에 의해 진화한 committee.

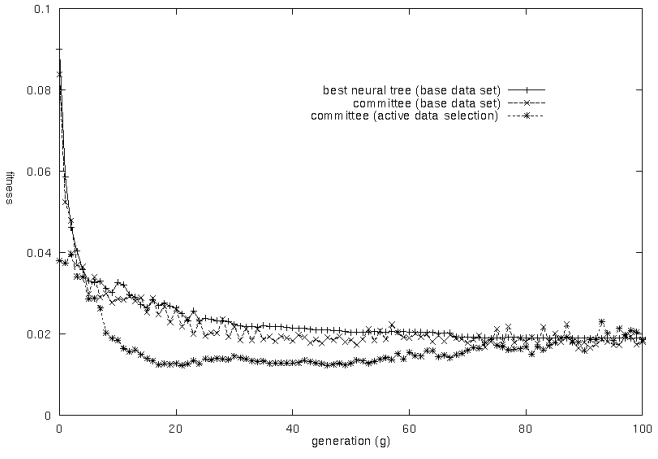

그림 7. Sunspot 시계열 예측에 대한 진화 횟수에 따른 적합도 비교: 단일 개체, 전체 데이터에 의해 진화한 committee, 능동적 데이터 선택에 의해 진화한 committee.

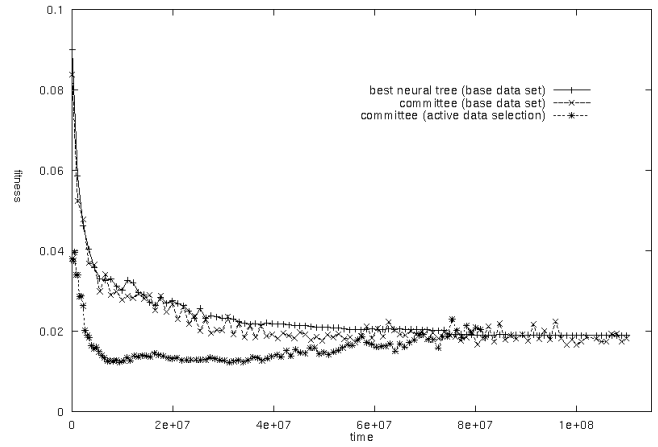

그림 8. Sunspot 시계열 예측에 대한 시간에 따른 적합도 비교: 단일 개체, 전체 데이터에 의해 진화한 committee, 능동적 데이터 선택에 의해 진화한 committee

4. 결론

본 논문에서는 데이터의 선택적 학습 방법을 통해서 committee 구성원을 전문화시킴으로써 committee의 일반화 성능과 진화 속도를 동시에 향상시킬 수 있음을 확인하였다. 특히, committee 구성원들간의 상관관계를 실험을 통하여 분석함으로써 committee 구성원의 다양성을 검증하였다. 제시된 방법은 근본적으로 개체를이 서로 다른 훈련 데이터의 학습을 통한 진화로써 committee를 형성하기 때문에 같은 데이터를 통한 진화보다 committee 구성원의 다양성을 높여준다고 할 수 있다. 이러한 기법은 빠른 적응 속도를 요구하는 동적인 환경에서의 학습 시스템이나 방대한 학습 데이터에 대한 부담을 가지고 있는 시스템에 유용하게 적용될 수 있을 것이다.

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 nerual 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] B.-T. Zhang, Accelerated learning by active example selection, *International Journal of Neural Systems*, 5(1): 67-75, 1994.

[4] M. P. Perrone and L. N. Cooper, When networks disagree: Ensemble methods for hybrid neural networks, in *Artificial Neual Networks for Speech and Vision*, Chapman & Hall, pp. 126-142, 1994.

[5] B.-T. Zhang, P. Ohm and H. Müehlenbein, Evolutionary induction of sparse neural trees, *Evolutionary Computation*, 1(4): 335-360, 1994.

[6] J. D. Farmer and J. J. Sidorowich, Predicting chaotic time series, *Physical Review Letters*, 59: 845-152, 1993.

[7] M. Mackey and L. Glass, Oscillation and chaos in physiological control systems, *Science,* 197: 287, 1977.

[8] N, Aerrabotu, A. J. Owens, and M. J. Walsh, Ensemble encoding for time series forecasting with MLP networks, *Applications and Science of Artificial Neural Networks: Proceedings of the SPIE Volume 3007*, S. Rogers (ed.), SPIE, Bellingham, Washington, USA, pp. 84-89, 1997.

[9] C. Svarer, L. K. Hansen, and J. Larsen, On design and evaluation of tapped-delay neural architectures, *IEEE International Conf. on Neural Networks*, pp. 46-51, 1992.

[10] A. S. Weigend, D. E. Rummelhart, and B. A. Huberman, back-propagation, weight elimination and time series prediction, *Proc. of the 1990 Connectionist Models Summer School*, pp. 105-116, 1990.