

First Principle을 결합한 최소제곱 Support Vector Machine의 예측 능력

(Prediction Performance of Hybrid Least Square Support
Vector Machine with First Principle Knowledge)

김 병 주 [†] 심 주 용 ^{††} 황 창 하 ^{††} 김 일 곤 ^{†††}

(Byung-Joo Kim) (Joo-Yong Shim) (Chang-Ha Hwang) (Il-Kon Kim)

요 약 본 논문에서는 최근 뛰어난 예측력으로 각광받는 최소제곱 Support Vector Machine(Least Square Support Vector Machine:LS-SVM)과 First Principle(FP)을 결합한 하이브리드 최소제곱-Support Vector Machine 모델, HLS-SVM(Hybrid Least Square-Support Vector Machine)을 제안한다. 제안한 모델인 하이브리드 최소제곱 Support Vector Machine을 기준의 방법인 하이브리드 신경망(Hybrid Neural Network:HNN), 비선형 칼만필터와 하이브리드 신경망을 결합한 HNN-EKF (Hybrid Neural Network with Extended Kalman Filter) 모델과 비교해 보았다. HLS-SVM 모델은 학습 및 validation 과정에서는 HNN-EKF와 근사한 성능을 보였고, HNN 보다는 우수한 결과를 보였고, 일반화 성능에서는 HNN-EKF에 비해 3배, HNN보다 100배정도 우수한 결과를 보였다.

키워드 : 하이브리드 최소제곱 Support Vector Machine, first principle, 신경망, 비선형 칼만필터

Abstract A hybrid least square Support Vector Machine combined with First Principle(FP) knowledge is proposed. We compare hybrid least square Support Vector Machine(HLS-SVM) with early proposed models such as Hybrid Neural Network(HNN) and HNN with Extended Kalman Filter(HNN-EKF). In the training and validation stage HLS-SVM shows similar performance with HNN-EKF but better than HNN, whereas, in the testing stage, it shows three times better than HNN-EKF, hundred times better than HNN model.

Key words : hybrid least squares support vector machine, first principle, neural network, extended Kalman filter

1. 서 론

학습 시스템(learning system)은 관측된 데이터를 이용하여 시스템을 학습시킨 후, 학습에 이용되지 않은 새로운 데이터가 입력될 때 올바른 답을 도출해 낼 수 있도록 구성된 시스템을 말한다. 이러한 학습 시스템은 데이터에 대한 사전지식(prior knowledge)이나 정보의 유무에 따라 Black-Box, White-Box, Gray-Box로 나눌 수 있다. Black-Box는 학습 데이터에 대한 사전 정보

는 없고 오직 관측 혹은 수집된 자료들만 존재하는 것으로 신경망이나 Support Vector 기계(Support Vector Machine:SVM)[1][2]가 이에 해당한다. White-box는 학습 자료에 대해 경험이나 전문가의 지식, 혹은 휴리스틱/heuristic과 같은 구조화된 지식이 있는 경우를 말하며, 이러한 지식은 일반적으로 if-then 형태로 표현되며 Fuzzy Logic Model이 여기에 해당된다. Gray-Box는 학습 데이터에 대해 어느 정도의 정보와 자료를 가지고 있는 경우로 많은 현실적인 데이터들이 여기에 해당하며 Neuro-Fuzzy 혹은 Fuzzy-Neuro Model이 여기에 해당한다[3].

학습시키고자 하는 데이터에 대해 사전 지식이 없거나 관측 혹은 수집이 어려운 데이터가 존재하는 경우 그러한 학습 자료에 의해 발생하는 오차는 학습 시스템의 성능을 저하시킨다. 학습 데이터의 관측이 어렵거나

[†] 정 회 원 : 영산대학교 컴퓨터 정보공학부 교수
bjkim@ysu.ac.kr

^{††} 비 회 원 : 대구가톨릭대학교 정보통계학과 교수
ds1631@hanmail.net
chhwang@cataegu.ac.kr

^{†††} 종신회원 : 경북대학교 컴퓨터과학과 교수
ikkim@knu.ac.kr

논문접수 : 2002년 10월 19일
심사완료 : 2003년 4월 18일

고비용인 경우의 대표적인 예는 화학 반응이다. 화학 반응은 경우에 따라 반응 중에 반응의 상태 관측이 어려운 경우가 종종 존재한다. 이러한 경우의 해결책으로 사전 지식을 바탕으로 추론에 의해 제어(inference control)하는 방법이 널리 사용되고 있으며 이러한 하이브리드 시스템을 통해 제어 모델의 정밀도를 높일 수 있다[4][5][6].

First Principle(FP)은 관측 혹은 수집할 수 있는 학습 데이터를 바탕으로 구성된다. 반면에 학습 시스템의 성능에는 영향을 미치지만 관측이 어려운 학습 데이터는 기계학습에서 사용되는 학습 방법중 하나를 택하여 간접적으로 추론한다. 이러한 하이브리드 형태의 제어 시스템의 성능은 단일 형태의 시스템 보다 학습과정에서 적은 수의 학습 자료만으로도 기계학습 기법만을 사용한 시스템보다 일반화 성능이 뛰어난 것으로 알려져 있다[7]. FP와 결합되는 학습 모델은 신경망, 퍼지 등이 사용되며 다양한 형태로 구성 될 수 있다. 신경망과 결합한 형태는 Ungar[8]가 제안하였는데 여기서는 다중 신경망과 FP가 결합하여 측정이 불가능한 프로세스 상의 변수 값을 추정하였다. Lindskog[9]는 Fuzzy 추론 모델과 FP를 결합한 형태를 사용하였다.

기계학습에서 사용되는 학습 방법중 신경망의 통계적 특성은 비모수(nonparametric) 적이면서도 비선형의 회귀모델과 비슷하여 사전 지식이 거의 알려지지 않은 시스템의 비선형 속성을 모델화 하는데 아주 우수한 특성을 가지고 있다. 하지만 신경망은 학습시 학습 속도의 느림, 지역최소화(local minima)에 빠질 가능성, 과도한 학습에 의한 과적합(overfitting)에 의해 일반화 성능이 떨어지는 단점뿐만 아니라 학습 시스템의 해석도 어렵다[10].

이러한 신경망이 가지는 제약을 해결하기 위해 SVM이 제안되었으며 패턴인식과 합수 추정에 성공적으로 적용되었다[1][2]. 유일한 전역해(global minima)를 보장하는 SVM이 많은 분야에서 성공적으로 적용되고 있으나 몇 가지 단점이 존재한다. SVM을 학습시키는데 이용되는 기법인 QP(Quadratic Programming) 과정에 복잡한 계산이 요구되며 시스템 구현에도 어려움이 따르게 된다. 또한 학습에 요구되는 메모리는 데이터 수의 제곱에 해당하며 학습 속도 또한 느린다.

이러한 문제를 해결하기 위해 Suykens[11]은 SVM의 분류기(classifier)를 구하는데 있어 SVM처럼 QP 문제를 해결하는 것이 아닌 선형 방정식을 푸는 방법인 최소제곱 SVM(Least Squares Support Vector Machine: LS-SVM)을 제안하였다. SVM이 전통적인 분류(classification)나 합수 추정에만 사용되는데 반해 LS-SVM은 순환(recurrent)

모델이나 제어 시스템에서 최적의 제어를 할 수 있는 문제에까지 적용 할 수 있다. 본 논문에서는 예측 능력이 우수한 LS-SVM과 FP를 결합한 하이브리드 모델, 하이브리드 최소제곱 Support Vector Machine을 제시한 모델의 성능을 평가하기 위해 기존의 논문[8]에서 실험한 자료인 생물학적 폐수 처리 과정에 적용시키고, 기존의 하이브리드 모델과의 성능 비교를 통해 제안된 시스템의 우수성을 나타내고자 한다.

논문의 구성은 다음과 같다. 2절에서는 생물학적 폐수 처리 시스템에서 미생물 성장의 동적 특성에 대해 알아보고 3절에서는 LS-SVM, 본 논문에서 제시하는 하이브리드 최소제곱 Support Vector Machine에 대해 설명한다. 4절에서는 제안된 방법에 의한 실험 및 기존의 다른 방법들과의 성능을 비교하며, 마지막으로 5절에서는 제안된 모델의 고찰 사항에 대해 기술한다.

2. 미생물 성장의 동적 특성

생물학적 폐수 처리과정에서 생물반응기(bioreactor)는 일반적으로 미생물(biomass)과 하나 이상의 기질(substrate)과 이를 포함하는 용기로 구성되어 있는 시스템이다. 이러한 생물 반응기는 용기 안에 최초의 미생물과 기질을 넣고 일정한 간격으로 하나 혹은 그 이상의 기질을 시간당 일정 비율로 첨가하는 fed-batch방식으로 작동된다. 처리 과정에서 알 수 있는 미생물, 기질, 시스템의 volume에 대한 변화를 FP로 나타내면 다음과 같다.

$$\frac{dB}{dt} = \mu(t)B(t) - \frac{Q(t)}{V(t)}B(t) \quad (1)$$

$$\frac{dS}{dt} = -k\mu(t)B(t) + \frac{Q(t)}{V(t)}[S_i(t) - S(t)] \quad (2)$$

$$\frac{dV}{dt} = Q(t) \quad (3)$$

여기서 $B(t)$ 는 시간 t 에서 미생물 농도를 나타내며 $S(t)$ 는 기질의 농도, $Q(t)$ 는 단위 시간당 시스템에 유입되는 기질의 비율을 나타낸다. $V(t)$ 는 시간 t 에서 용기의 양을 나타내며 $\mu(t)$ 는 미생물의 성장비율을 나타낸다. $\mu(t)$ 는 측정이 불가능한 파라메터로 이것이 처리 과정을 복잡하게 하는 주 요인이다. 많은 논문에서 측정이 불가능한 $\mu(t)$ 를 추론하기 위해 학습 단계에서 필요한 목표값(target value) $\mu(t)$ 에 대해 다양한 모델을 제시하였는데, 가장 널리 사용되는 모델은 Monod와 Haldane 제안한 것으로 식 (4)와 (5)에서처럼 두 가지로 표시한다.

$$\mu(t) = \frac{\mu^* S(t)}{K_m + S(t)} \quad (4)$$

$$\mu(t) = \frac{\mu^* S(t)}{K_m + S(t) + \frac{S(t)^2}{K_n}} \quad (5)$$

미분 방정식으로 표시된 식 (1), (2), (3)을 이산형의 형태로 표시하면 다음과 같다.

$$B_{t+1} = B_t + (\mu - \frac{Q_t}{V_t}) B_t \quad (6)$$

$$S_{t+1} = -k \mu_t B_t + S_t (1 - \frac{Q_t}{V_t}) + \frac{Q_t}{V_t} S_t(t) \quad (7)$$

$$V_{t+1} = V_t + Q_t \quad (8)$$

B 와 S 의 관측시 발생하는 관측 오차와 모델 자체의 불확실성을 반영하기 위해 오차항 ε , ν 를 추가하면 시스템의 형태는 이산 시간상에서 비선형 동적 시스템의 형태를 취하게 되며 다음과 같은 식으로 표현한다.

$$x_{t+1} = f_t(x_t, \Psi_t, \mu_t(x_t)) + \varepsilon_t \quad (9)$$

$$z_{t+1} = H(x_t) + \nu_t \quad (10)$$

앞에서 언급한 것처럼 식 (9)에서 $\mu(t)$ 는 알려지지도 않고 측정이 불가능한 값이다. 만약 $\mu(t)$ 를 알 수만 있다면 비선형 Kalman Filter(EKF)[12]를 사용하여 현재의 추정치를 바탕으로 다음 상태를 예측 할 수 있다. 따라서 생물학적 폐수 처리와 같이 데이터의 사전 지식이 없거나 측정이 불가능한 변수가 존재하는 제어 시스템에는 학습에 의해서 측정 불가능한 변수를 추정하고, 알려진 정보를 바탕으로 구성한 FP에 의해 다음 상태를 추정하는 하이브리드 시스템이 필요하다.

3. 하이브리드 최소제곱 Support Vector Machine

3.1 LS-SVM(Least Squares Support Vector Machine)

SVM은 신경망의 단점인 과적합과 지역최소화에 빠지는 단점을 해결하기 위해 Vapnik과 공동연구자들에 의해 개발되었다. SVM이 신경망에서 발생하는 과적합을 효과적으로 막아주는 것은 Vapnik-Chervonenkis(VC) 이론으로 설명 할 수 있으며 SVM의 학습은 볼록함수(convex function)를 최대화함으로써 수행하는 것으로 이는 polynomial 시간 내에 발견될 수 있는 유일해가 존재한다는 것을 의미한다. SVM에서의 가중치는 QP 문제를 해결함으로써 찾을 수 있다. 유일한 전역 해를 보장하는 SVM이 많은 분야에서 성공적으로 적용되고 있으나 몇 가지 단점이 존재한다. SVM을 학습시키는데 이용되는 기법인 QP 과정에 복잡한 계산이 요구되며 시스템 구현에도 어려움이 따르게 된다. 또한 학습에 요구되는 메모리는 데이터 수의 제곱에 해당하며 학습 속도 또한 느린다. 이러한 문제를 해결하기 위해 Suykens[11]은 SVM의 분류기(classifier)를 구하는데

있어 SVM처럼 QP 문제를 해결하는 것이 아닌 선형방정식을 푸는 형태인 LS-SVM모델을 제안하였는데 계산상의 단순함 및 대용량 학습 데이터에도 적용할 수 있는 장점이 있다.

본 논문에서 적용하고자 하는 문제는 함수 추정에 관한 것이므로 비선형 함수 추정을 위한 LS-SVM 모델의 추정 함수식에 대해서 기술한다. LS-SVM모델은 특징 공간(feature space)상에서 식 (11)과 같이 표현한다.

$$y(x) = w^T \varphi(x) + b \quad (11)$$

식 (13)와 같은 제약 조건하에서 비용함수(cost function) (12)를 최소화하는 문제를 해결하기 위해

$$\min_{w, b, e} L(w, e) = \frac{1}{2} w^T w + \gamma \frac{1}{2} \sum_{k=1}^N e_k^2 \quad (12)$$

$$\text{subject to } y_k = w^T \varphi(x_k) + b + e_k, \quad k=1, \dots, N \quad (13)$$

다음과 같은 라그랑즈 함수를 정의한다.

$$L(w, b, e; a) = J(w, e) - \sum_{k=1}^N a_k [w^T \varphi(x_k) + b + e_k - y_k] \quad (14)$$

여기서 a_k 는 라그랑즈 승수이며 이 함수가 최적이 되기 위한 조건은 식 (15)와 같으며

$$\left. \begin{aligned} \frac{\partial L}{\partial w} &= 0 \Rightarrow w = \sum_{k=1}^N a_k \varphi(x_k) \\ \frac{\partial L}{\partial b} &= 0 \Rightarrow \sum_{k=1}^N a_k = 0 \\ \frac{\partial L}{\partial e_k} &= 0 \Rightarrow a_k = \gamma e_k, \quad k=1, \dots, N \\ \frac{\partial L}{\partial a_k} &= 0 \Rightarrow w^T \varphi(x_k) + b + e_k - y_k = 0, \\ k &= 1, \dots, N \end{aligned} \right\} \quad (15)$$

이는 다음과 같이 선형 방정식을 푸는 문제로 바꾼다.

$$\begin{bmatrix} 0 & 1 \\ 1 & Q + \gamma^{-1} I \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \quad (16)$$

여기서 $y = [y_1; \dots; y_N]$, $1 = [1; \dots; 1]$, $a = [a_1; \dots; a_N]$ 을 나타낸다.

Mercer의 조건을 식 (16)에 적용하면

$$Q_{kl} = \varphi(x_k)^T \varphi(x_l) = K(x_k, x_l) \quad (17)$$

식 (16) (17)에 의해서 LS-SVM의 함수 추정식은 다음과 같다.

$$y(x) = \sum_{k=1}^N a_k K(x, x_k) + b \quad (18)$$

3.2 하이브리드 최소제곱 Support Vector Machine

생물학적 폐수 처리 시스템에서 가장 어려운 부분은 미생물의 성장을 $\mu(t)$ 를 측정하기 어렵다는 사실이다. 이러한 경우의 해결책으로 사전 지식을 바탕으로 추론에 의해 프로세스(process)를 제어하는 추론에 의한 제어 방법이 널리 사용되고 있으며 이러한 하이브리드 시

스템은 시스템의 정밀도를 높일 수 있다.

그림 1에서 First Principle(FP) Model은 폐수 처리 과정에서 알려진 식 (6), (7), (8)이다. 관측이 어려운 학습 데이터 $\mu(t)$ 는 LS-SVM을 이용하여 간접적으로 예측한다. 이러한 하이브리드 형태의 제어 시스템의 성능은 단일 형태의 시스템 보다 학습과정에서 적은 수의 학습 자료만으로도 기계학습 기법만을 사용한 시스템 보다 일반화 성능이 뛰어난 것으로 알려져 있다[7]. 시스템의 동작 방식은 측정이 불가능한 미생물의 성장을 $\mu(t)$ 는 LS-SVM에 의해 추론하며 다음 단계의 B와 S는 식 (6)과 (7)에 의해 학습하지 않은 데이터에 대해 one step ahead 방식으로 예측하여 제안된 모델의 일반화 성능을 평가 한다.

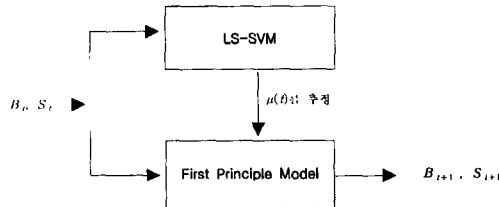


그림 1 LS-SVM과 FP의 결합에 의한 B_t , S_t 의 예측 과정

LS-SVM에서 수행에 필요한 파라미터 σ 와 λ 는 cross validation을 통해 구해진다. 또한 사용된 커널 함수(kernel function)는 Radial Basis Function이며 형태는 다음과 같다.

$$K(x, y) = \exp\left(-\frac{(x-y)^2}{2\sigma^2}\right) \quad (19)$$

4. 실험 및 결과

실험내용은 다음과 같은 3가지 모델에 대해 참고문헌 [8]에서 실현한 자료를 바탕으로 학습, validation, 테스트를 수행하였다. 첫째는 Ungar에 의해서 제안된 FP와 신경망을 결합한 모델, HNN(Hybrid Neural Network), 두 번째는 HNN모델에서 가중치 수정 부분은 EKF를 결합한 모델, HNN-EKF(Hybrid Neural Network with EKF), 세 번째는 FP와 LS-SVM을 결합한 모델, 하이브리드 최소제곱 Support Vector Machine(HLS-SVM:Hybrid LS-SVM)이다. 세 가지 모델의 일반화 능력을 평가하기 위해 학습, validation, 테스트 데이터를 통해 이들의 성능을 비교한다.

각 시스템의 동작 방식은 측정이 불가능한 미생물의 성장을 $\mu(t)$ 는 Black-Box 모델에 해당하는 신경망, 신경망과 EKF를 결합한 방식, LS-SVM에 의해서 추론

하며 다음 단계의 B와 S는 식 (6)과 (7)에 의해 one step ahead 방식으로 예측한다.

학습, validation, 테스트 데이터 생성을 위해 설정한 각 파라메타의 값은 참고문헌[8]과 같이 $\mu^* = 5$, $K_m = 10$, $K_u = 0.1$, $k = 1.0$, $S_i = 3.5$ 그리고 시스템에 유입되는 미생물의 양 Q는 0.1로 설정하였다. 이때 B_1 , S_1 , μ_1 과 같은 다음 상태의 B, S, μ 값들은 식 (5), (6), (7)을 이용하여 데이터 생성식에 random noise를 추가하였는데, 이때 random noise는 평균이 0, 표준편차가 0.01인 정규분포로부터 생성된다. 학습, validation, 테스트를 위한 데이터 생성 방법에서 사용된 각 변수의 초기 값을 표 1에 나타내었다.

표 1 학습, validation, 테스트 데이터 생성시
파라메타의 초기값

	B_0	S_0	V_0	Noise	데이터 갯수
Train	0.1	0.5	10	$N(0, 0.01^2)$	20
Validation	0.3	0.7	4	$N(0, 0.01^2)$	80
Test	0.001	1.5	4	$N(0, 0.01^2)$	80

학습 데이터의 개수가 validation, 테스트 데이터의 25% 밖에 되지 않는 것은 하이브리드 모델의 특성이 순수한 기계학습 방법보다 적은 수의 학습 데이터를 사용하여도 일반화 성능이 우수하다는 점을 확인하기 위함이다. 그리고 테스트 데이터의 B_0 , S_0 를 각각 0.001과 1.5로 한 이유는 각각의 하이브리드 모델이 학습 및 validation에 사용하지 않은 데이터 범위에서 얼마만큼 일반화 성능이 우수한가를 보기 위해 설정하였다. 용기의 불륨 V의 값은 단위 시간당 0.1씩 증가하여 최종 값이 12가 되도록 설정되어 있으므로 학습 데이터의 초기 값은 10, validation과 테스트의 값은 4로 설정하였다.

첫 번째 방법인 HNN에서 신경망의 구조는 입력층에 변수 B와 S를 위한 2개의 입력노드, 은닉층에 3개의 노드, 출력층에 하나의 노드로 구성된다. 이때 사용된 활성화 함수(activation function)는 각각 시그모이드 함수와 항등함수이며 가중치 수정 방법은 신경망 학습기법 중 널리 사용되는 오류역전파 학습 방법(Backpropagation)을 사용하였다.

HNN에서 신경망의 학습은 validation 데이터의 예측 SSE(Sum of Square Error)를 최소로 하는 시점에서 학습을 멈추는 cross-validation 방법을 사용한다. 여기서 구해진 연결 가중치의 값과 신경망의 학습에 의해 추정된 $\mu(t)$ 에 의해 B와 S의 값을 식 (6)과 (7)에 의해 one step ahead 예측 방식으로 테스트 데이터에 적용하

여 일반화 성능을 검정한다. 두 번째 방법인 HNN-EKF에서 신경망 연결가중치의 초기값은 모두 1로 설정하였으며, EKF의 공정 단계에서 공분산 행렬, 오차의 초기 값 행렬은 단위행렬로, 관측 단계에서의 오차의 분산은 0.01로 설정하였다. 세 번째 방법인 HLS-SVM에서 LS-SVM 수행에 필요한 파라메터 $\sigma = 1.96$ 과 $\lambda = 0.479$ 는 cross-validation을 통해 구한 커널 함수의 모수(parameter) 값이다.

그림 2, 3, 4에서 *로 표시된 그래프는 각 Black-Box 모델에 의해 추정된 $\mu(t)$ 의 값이다. Ungar에 의해서 제안된 HNN모델의 경우 $\mu(t)$ 의 추정이 올바르게 되었다고 보기는 힘들다. 반면에 HNN-EKF와 HLS-SVM 모델은 HNN에 비해 $\mu(t)$ 를 잘 추정하고 있다.

하이브리드 신경망의 경우 초기 연결 가중치의 값은 난수 발생에 의해 정해지는데 신경망의 초기 연결 가중치의 값에 따라 하이브리드 신경망의 학습 및 일반화

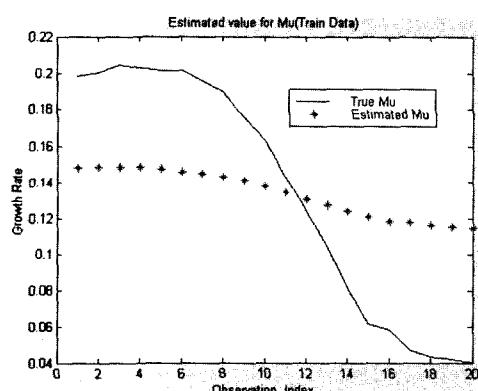


그림 2 HNN모델에서 $\mu(t)$ 의 예측값

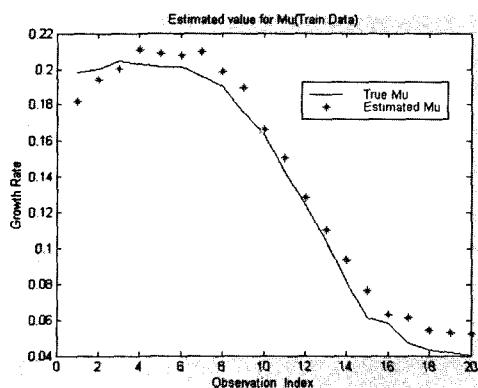


그림 3 HNN-EKF모델에서 $\mu(t)$ 의 예측값

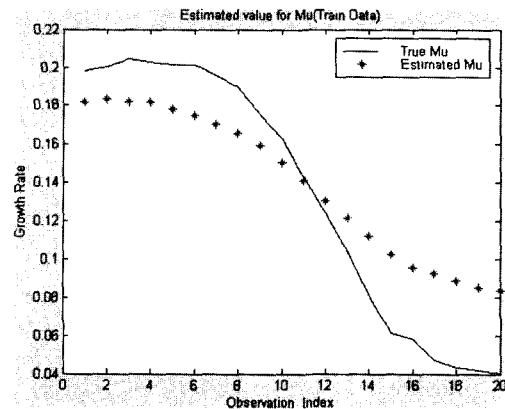


그림 4 HLS-SVM모델에서 $\mu(t)$ 의 예측값

성능에 많은 영향을 미치므로 10번의 실험을 통하여 평균값을 구하고 이를 B와 S의 SSE로 사용하였으며 표2에 결과값이 나타나 있다. 표 2의 결과에서 보면 학습시의 SSE에 비해 validation, 테스트의 SSE는 약 100배, 115배 증가하였음을 보인다.

표 2 신경회로망의 10번의 시도에 대한 B와 S의 SSE 평균값

	Train		Validation		Test	
	잘못된 계산식	S	B	S	B	S
SSE	0.21	0.21	19.57	19.51	23.51	23.29

HNN-EKF의 경우 전방향 단계(forward pass)에서는 신경망의 계산 과정을 따르고 연결 가중치를 수정하는 부분은 EKF를 이용하는데, EKF 자체의 특징인 초기 값의 영향이 오래 가지 않으므로 HNN과 달리 학습시에 SSE의 편차 없이 안정적으로 학습이 이루어진다. 학습시의 과적합을 방지하기 위해 validation 과정을 거치는데 이 과정에서의 SSE는 학습시의 순서와 동일하게 HNN-EKF, HLS-SVM, HNN의 순이다. HNN-EKF 와 HLS-SVM의 경우 두 방법은 근소한 차이를 보이는데 비해 HNN의 경우는 validation 데이터의 SSE 값이 학습시에 비해 100배정도 증가하였는데 이는 HNN이 과적합에 따른 일반화 성능이 떨어짐을 보여준다.

그림 5, 6, 7은 B에 대해 세가지 모델에서 대한 테스트 데이터의 일반화 성능을 나타내고, 그림 8, 9, 10은 S에 대한 일반화 성능을 나타낸다.

HLS-SVM의 경우 SVM이 신경망에서 발생하는 과적합을 효과적으로 막아 주고, 또한 유일한 전역해를 보장하며 예측력이 뛰어난 점으로 인해 일반화 성능에서

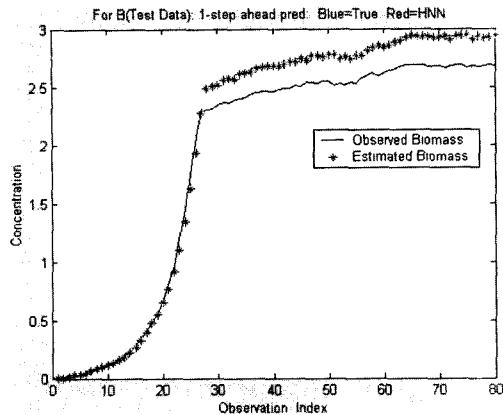


그림 5 HNN모델에서 B의 예측값

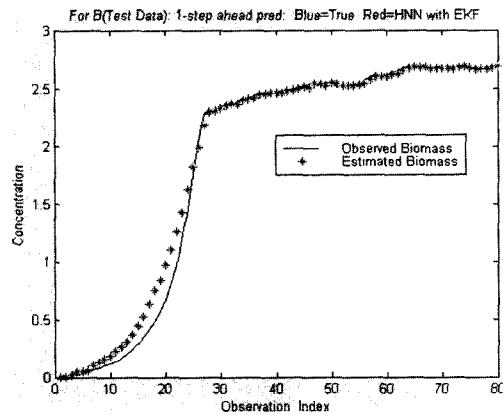


그림 6 HNN-EKF 모델에서 B의 예측값

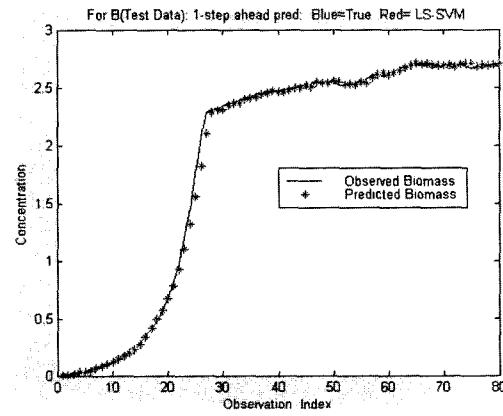


그림 7 HLS-SVM 모델에서 B의 예측값

HNN에 비해 약 100배정도, HNN-EKF에 비해 3배정도 우수함을 보이고 있다. HNN-EKF가 학습, validation 과정에서 HLS-SVM 보다 근소한 차이로 우수 하나 일반화 성능이 떨어지는 것은 HNN-EKF 역시 어느 정도의 과적합이 발생 했다고 볼 수 있다.

HNN의 경우는 10번의 실험에서 다양한 형태로 나타났는데 B와 S의 SSE가 각각 2.78862, 2.69960 인 경우의 것을 나타내었다.

HNN, HNN-EKF, HLS_SVM 세 가지 방법에 의한 실험 결과가 표 3에 나타나 있다. 먼저 학습데이터에 대한 학습 능력을 보면 HNN-EKF, HLS-SVM 순이다.

HNN의 경우는 학습이 잘 되는 best case의 경우 HNN-EKF와 비슷하나 그러한 경우는 10번의 실험에서 2번 정도 그러한 결과를 보이고, worst case의 경우

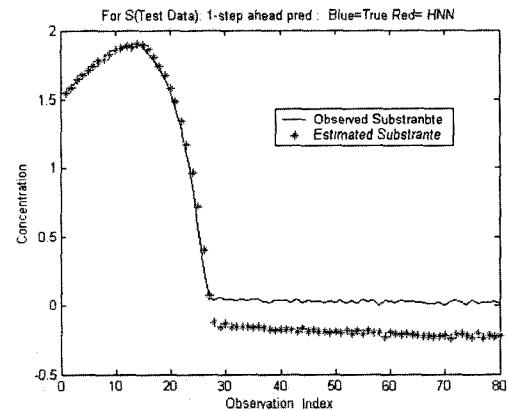


그림 8 HNN모델에서 S의 예측값

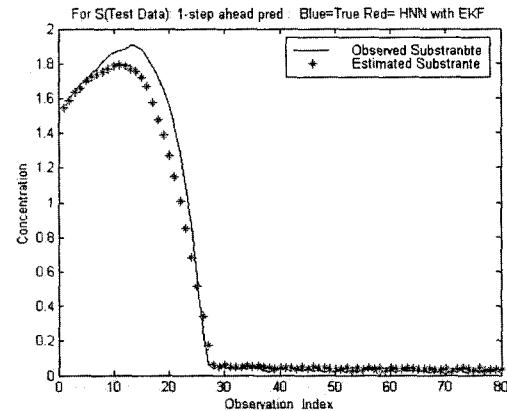


그림 9 HNN-EKF 모델에서 S의 예측값

표 3 세 가지 방법에 대한 B와 S의 SSE 비교

방법	Train		Validation		Test	
	B	S	B	S	B	S
HNN	0.22	0.22	19.57	19.52	23.52	23.30
HNN-EKF	0.00310	0.02375	0.01846	0.01584	0.67861	0.67558
HLS-SVM	0.1206	0.03356	0.03759	0.03433	0.18469	0.19361

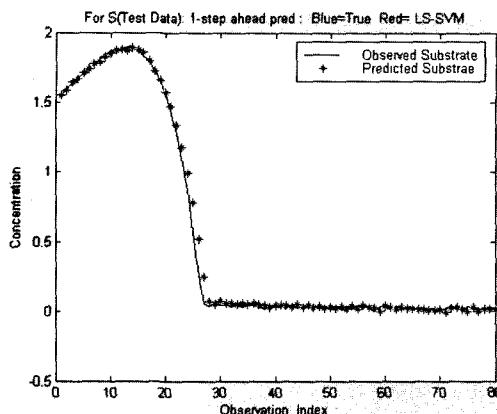


그림 10 HLS-SVM 모델에서 S의 예측값

best case에 비해 100배 정도의 차이를 보인다. 테스트 데이터에 대한 일반화 성능은 제안된 모델인 HLS-SVM이 HNN-EKF에 비해 3배, HNN에 비해 100배 정도의 우수함을 실험을 통해 알 수 있다.

5. 결 론

본 논문에서는 우수한 예측력을 보여준 최근에 각광을 받고 있는 LS-SVM과 FP를 결합한 하이브리드 최소제곱 Support Vector Machine (HLS-SVM)을 제안하고 이를 참고문헌 [8]에서 실증한 자료인 생물학적 폐수 처리 시스템에 적용 시켜 보았다. 논문에서 제안한 HLS-SVM과 Ungar[8]가 제안한 HNN, HNN-EKF를 비교해 본 결과 HLS-SVM은 학습 및 validation 과정에서 HNN-EKF와 근사한 성능을 보이며, HNN보다는 과적합이 발생하지 않으면서 학습이 안정적으로 이루어졌다. 실험결과에 의하면 일반화 성능에서는 논문에서 제안하는 HLS-SVM이 HNN-EKF에 비해 3배, HNN에 비해 100배 정도 우수함을 보여 준다.

하지만 HLS-SVM 모델은 커널 함수의 모수인 λ , σ 를 구하기 위해 cross-validation 과정에서 많은 시간을 필요로 한다. 본 논문을 준비하는 동안 이를 해결하기 위해 많은 연구가 이루어 지고 있는데 Gavin[16]은 선형

방정식을 푸는데 있어 Gauss-Jordan 소거법에 의한 빠른 leave-one-out cross-validation 기법을 제안하였다.

앞으로의 연구 과제는 현재 제안되고 있는 방법을 바탕으로 HLS-SVM 모델에서 최적의 λ , σ 를 빠르게 구하는 방법에 대해 연구할 계획이다.

참 고 문 헌

- [1] S. Gunn, "Support Vector Machines for Classification and Regression," ISIS Technical Report, U. of Southampton, 1998.
- [2] E. Osuna, R. Freund, and F. Girosi, "Support Vector Machines: Training and Applications," Technical Report, MIT AI Laboratory, 1997.
- [3] <http://www.support-vector.ws/html/downloads.html>
- [4] C. Brosilow, and M. Tong, "Inference Control of Processes, Part II The Structure and Dynamics of Inferential Control Systems," AIChE J., Vol. 24, No. 3, 1978.
- [5] M. HADJISKI, Control of Indirect Measurable and Time-Varying Technological Plants, Ph. D. Thesis, HIChT, Sofia, Bulgaria, 1979.
- [6] B. JOSEPH, and C. BROSILOU, "Inferential Control of Processes, Part I Steady State Analysis and Design," AIChE J., Vol. 24, No. 3, 1978.
- [7] http://www.ici.ro/ici/revista/sic99_1/art04.html
- [8] De Veaux, R. Bain and L.H. Ungar, "Hybrid Neural Network models for environmental process control," Journal of ENVIRONMETRICS10, 225-236, 1999.
- [9] P. LINDSKOG, and L. LJUNG, "Ensuring Certain Physical Properties in Black Box Model by Applying Fuzzy Techniques," Technical Report, 1996.
- [10] R. Caruana, S. Lawrence and L. Giles, "Overfitting in Neural Nets: Backpropagation, Conjugate Gradient, and Early Stopping," Neural Information Processing Systems, Denver, Colorado, November 28-30, 2000.
- [11] J.A.K. Suykens, "Nonlinear Modeling and Support Vector Machines," Accessible at <http://www.kdiss.or.kr/kdiss/>
- [12] G. Welch, and G. Bishop, "An Introduction to the Kalman Filter," Siggraph Course8, 2001.
- [13] X. Shao, Model Selection Using Statistical

- Learning Theory, Ph. D. Thesis, U. of Minnesota,
1999.
- [14] V. Vapnik, The Nature of Statistical Learning
Theory, Springer, 1995.
- [15] V. Vapnik, Statistical Learning Theory, Springer,
1998.
- [16] S. Gavin, C. Cawley and L.C. Talbot, "Fast Exact
Leave One Out Cross Validation of Least Squares
Support Vector Machines," European Symposium
on Artificial Neural Networks, Bruges, Belgium,
April 24~26, 2002.



김명주

1990년 부산대학교 전자계산학과 졸업
(이학사). 1992년 부산대학교 대학원 전
자계산학 전공 (이학석사). 1998년 경북
대학교 대학원 전자계산학 전공 박사
수료. 1992~1996년 서라벌 대학 전자계
산과 교수. 1996~2002년 성심외국어대
학 정보통신학부 교수. 2003~현재 영산대학교 컴퓨터 정보
공학부 교수. 관심분야는 기계학습(machine learning)



심주용

1983년 서강대학교 화학공학과 졸업
1988년 Northeastern Illinois University 대학원 수학과 졸업(이학석사).
1992년 University of Illinois at Urbana-Champaign 대학원 통계학과
박사 수료. 1996년 경북대학교 대학원
통계학과 (이학박사). 2003~현재 대구가톨릭대학교 정보통
계학과 겸임교수. 관심분야는 dynamic programming,
machine learning



황창하

1982년 경북대학교 사범대학 수학교육
과(이학사). 1984년 서울대학교 대학원
계산통계학과(이학석사). 1985~1987년
한국통신 전임연구원. 1991년 미국 미시
간대학교 통계학과(이학박사). 1992년~
1994년 경성대학교 전산통계학과 조교
수. 1995년~현재 대구가톨릭대학교 정보통계학과 부교수
관심분야는 기계학습, 정보과학

김일곤

정보과학회논문지 : 소프트웨어 및 응용
제 30 권 제 6 호 참조