

# 상시감시기술에서 SVR과 PLSR을 이용한 Auto-association 모델링 및 성능비교

## Modeling and Comparison for Auto-association using Support Vector Regression (SVR) and Partial Least Square Regression (PLSR) in Online Monitoring Techniques

김성준 · 서인용

Seong-Jun Kim and In-Yong Seo

강릉원주대학교 산업공학과, 한국전력공사 전력연구원

### 요 약

센서시스템을 이용한 상시감시는 발전소의 효율적인 운전과 안전을 담보하는 데 필수적이다. 상시감시기술을 구현하기 위해서는 우선 센서로부터 전송된 신호로부터 발전소 운전과 파라미터의 참값을 예측하는 모델 즉 Auto-association (AA) 모델을 확보하는 것이 중요하다. 이를 위해 본 논문에서는 Support Vector Regression (SVR)과 Partial Least Square Regression (PLSR)을 이용하는 방안을 각각 제시한다. 이렇게 해서 구축된 모델은 모니터해야 할 파라미터가 많을 때에도 쉽게 적용할 수 있다. 실제 발전소에서 수집된 데이터셋을 이용하여 AA 모델링의 정확도 및 민감도를 비교한 결과, 정확도 면에서는 SVR이 우수한 반면 민감도 면에서는 PLSR이 다소 나은 것으로 나타났다.

**키워드 :** 상시감시, Auto-association, Support Vector Regression, Partial Least Square Regression, 정확도, 민감도

### Abstract

An online monitoring based upon sensor system is essential to assure both efficient operation and safety in the power plant. Of great importance is modeling for auto-association (AA) in online monitoring technique. The objective of auto-associative models lies in predicting true values of plant operation parameters from sensor signals transmitted. This paper presents two AA models using Support Vector Regression (SVR) and Partial Least Square Regression (PLSR). The presented models are useful, in particular, when there are many parameters to monitor in the power plant. Illustrative examples are given by using a real-world plant dataset. AA performances of SVR and PLSR are finally summarized in terms of accuracy and sensitivity. According to our results, SVR shows much higher accuracy and, however, its sensitivity is relatively degraded.

**Key Words :** Online Monitoring, Auto-association, Support Vector Regression, Partial Least Square Regression, Accuracy, Sensitivity

## 1. 서 론

센서기술의 발전에 힘입어 원자력발전소에서도 발전설비에 대한 상태기반정비 (Condition based Maintenance)가 보편화되어가고 있다. 따라서, 플랜트운전 중에도 기기 및 센서의 상태를 모니터링하면서 결함이 검출된 센서에 대해서만 선별적으로 교정하는 것이 가능해진다. 센서교정을 목적으로 하는 상시감시 (On-line Monitoring, OLM) 기술은 바로 이러한 배경을 두고 도입되었다[1].

원자력발전소에서 상시감시기술의 가장 큰 장점은 불필요한 교정을 줄일 수 있을 뿐 아니라 설비운전의 중지를 요

하지도 않기 때문에 많은 시간과 비용을 절약할 수 있다는 데서 찾을 수 있다. 흔히 정기적인 교정을 수행하는 경우, 결함 센서에 대해서는 대책이 너무 늦어지는 문제가 있고 또한 정상 센서에 대해서는 새로운 결함이 개입될 수도 있는 여지를 남겨두게 된다. 하지만 상시감시기술을 이용하게 되면 센서의 결함을 실시간으로 탐지할 수 있고 많은 센서들 중에서도 교정이 필요한 센서를 효과적으로 식별할 수 있다. 아울러 정비비용 절감, 교정오류 감소, 측정기기신뢰도 유지, 운전중지시간 축소 등에도 도움이 되며 특히 방사능에 노출되는 횟수도 줄이는 효과가 있다[1].

OLM을 통한 센서교정이 가능하려면 우선 주어진 입력을 그대로 모사하여 출력으로 내보낼 수 있는 모델이 필요하다. 이러한 모델을 Auto-association (AA) Model이라고 부른다[2]. 이는 곧 입력이 무엇이었는가를 추정하는 것과 같은 문제인데, 예를 들어 센서들로부터 얻은 전류, 수위, 온도, 압력 등 설비파라미터의 측정값  $(x_1, x_2, \dots, x_m)$ 으로

접수일자 : 2010년 2월 17일

완료일자 : 2010년 7월 31일

본 논문은 2009년도 정부재원으로 한국연구재단의 지원을 받아 연구되었음 (NRF-2009-0069111).

부터 그 참값 (true value) 에 대한 예측치 ( $\hat{x}_1, \hat{x}_2, \dots, \hat{x}_m$ ) 을 어떻게 출력할 것인가로 요약된다. 이와 같은 AA 문제는 그림 1로 표현된다.

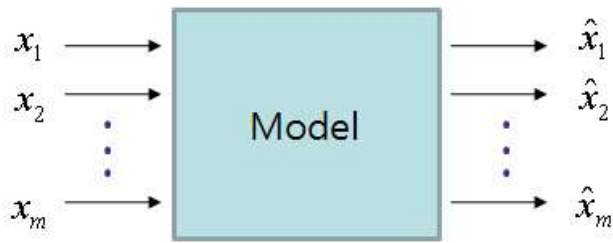


그림 1. 상시감시를 위한 Auto-association 모델  
Fig. 1. Auto-association Model for OLM

이러한 AA 모델링을 위해서, Artificial Neural Network (ANN), Kernel Regression (KR), Multi-variate State Estimation Technique (MSET) 등과 같은 기법이 제시되었고 그 중에서 ANN이 광범위하게 적용되고 있다[1, 3]. ANN 구조로서는 다층퍼셉트론이 주로 쓰이고 있으며 만족할만한 AA 성능을 얻기 위해서는 뉴런계층이 5개 이상 필요한 것으로 판단된다[1]. 한편 Support Vector Regression (SVR)은 다양한 영역에서 ANN의 대안으로 채택되고 있으나 AA 문제에 관련해서는 그 연구가 매우 미흡한 실정이다. 더 적은 학습데이터로도 더 나은 일반화능력을 제공한다는 점에서 SVR은 AA 모델링에 유익한 특징을 갖고 있다. 본 논문의 목적은 SVR을 이용하여 상시감시를 위한 Auto-association 모델링 방안을 제안하는 데 있다. 제안된 방법은 실제 발전소 데이터를 이용하여 정확도 (Accuracy) 와 민감도 (Sensitivity) 관점에서 검토가 이루어진다. 또한 변수변환의 효과를 살펴보기 위해 Partial Least Square Regression (PLSR)을 이용했을 때의 모델링 결과도 함께 비교하고자 한다.

## 2. 정확도 및 민감도에 의한 Auto-association 평가

$m$  개의 센서가 있고 그 중  $i$  번째 센서로부터 시점  $t$  에서 얻은 측정값을  $x_i(t)$  라고 할 때 이는 다음과 같이 가정할 수 있다.

$$x_i(t) = \mu_i(t) + e_i(t) \quad (1)$$

식 (1)에서,  $\mu_i(t)$  는 시점  $t$  에서  $i$  번째 운전과라미터의 참값을 의미하고  $e_i(t)$  는 시간에 독립인 가우시안 노이즈로 가정한다. 앞서서도 언급하였듯이, OLM의 기본 목적은 일단  $m$  개의 관측값  $x_1(t), x_2(t), \dots, x_m(t)$  가 주어질 때 예측값  $\hat{x}_1(t), \hat{x}_2(t), \dots, \hat{x}_m(t)$  를 구하는 데 있다. 엄밀히 말하면,  $\hat{x}_i(t)$  는  $x_i(t)$  가 아니라  $\mu_i(t)$  의 추정치라고 볼 수 있다. 잔차 (Residual)는 측정값과 예측값의 차이로서 식 (2)와 같이 정의된다.

$$r_i(t) = x_i(t) - \hat{x}_i(t) \quad (2)$$

따라서 잔차분석을 통해 해당 센서채널에 문제가 있는지

여부를 판단할 수 있게 된다.

상시감시가 효과적으로 이루어지려면 드리프트 (Drift)가 발생한 센서채널을 즉각 탐지할 수 있어야 한다. 이것은 모델에 자기수정 (Auto-correction) 기능이 내제되어야 함을 의미한다[1]. 다시 말해, 드리프트가 있는 입력  $x_i^d(t)$  가 주어졌을 때 모델이 제시한 출력  $\hat{x}_i^d(t)$  에도 드리프트가 포함된다면 이는 바람직하지 않다. 이렇게 되면 잔차 쪽에 드리프트가 나타나지 않게 되어 센서의 결함을 탐지할 수 없기 때문이다. AA 모델이란 주어진 입력을 그대로 모사하여 출력으로 내보낼 수 있는 모델로서 드리프트가 포함된 입력 (관측값)에 대해 학습된 메모리를 활용하여 드리프트가 수정된 출력 (예측값)을 제공할 수 있다. 그러면 예측값을 관측값과 비교함으로써 센서에 드리프트가 발생했는지를 판단할 수 있다. AA 모델의 이러한 특징은 각 센서의 검교정 상태를 효율적으로 모니터링하는 주요 기반이 된다.

AA 모델의 성능기준은 크게 정확도 (Accuracy)와 민감도 (Sensitivity)로 나누어 생각할 수 있다[1]. 정확도란 모델이 얼마나 정확하게 참값을 예측할 수 있는가를 측정하는 것으로 대개는 예측값과 참값 간의 Mean Squared Deviation (MSD)로서 나타낼 수 있다. 한편 민감도는 다시 자기민감도 (Auto-sensitivity)와 교차민감도 (Cross-sensitivity)로 구분된다. 자기민감도란 결함이나 드리프트로 인해 어떤 한 센서의 측정값에 바이어스가 걸렸을 때 출력값이 얼마나 민감하게 반응하는가를 측정하는 것으로 모델의 결함탐지력을 평가하는 측도이다. 또한 교차민감도는 어떤 한 결함센서의 입력이 인접한 다른 센서의 예측값에 미치는 영향을 말하는 것으로 Robustness를 알 수 있게 한다. 온라인센서교정을 효과적으로 수행하기 위해서는, MSD가 작고 자기수정능력이 있으며 인접센서의 드리프트에는 덜 민감한 모델이 바람직하다[3].

## 3. Auto-association을 위한 SVR 및 PLSR

### 3.1 Support Vector Regression (SVR)

통계적 학습이론에 근거한 Support Vector Machine (SVM)은 Vladimir Vapnik[4]에 의해 이론적 체계를 갖춘 이래, 최근 패턴인식을 포함한 다양한 분야에서 분류, 회귀, 추정의 문제를 다루는 데 널리 활용되고 있다. 신경회로망 등 많은 방법이 Empirical Risk Minimization에 기초하고 있는 반면 SVM은 Structural Risk Minimization에 기초한다. 이런 연유로, SVM 쪽이 보다 적은 학습데이터로도 더 나은 일반화능력을 보여줄 수 있으며 유망한 대안으로 주목 받고 있다. 또한 SVM은 OLM 이외에도 발전소 현장의 다양한 문제를 해결하는 데에도 활용되고 있다. 예를 들어, 발전소의 NOx 감소를 위한 최적운전조건 결정[5]이나 터빈이나 베어링의 진동신호 프로파일분류[6] 등을 들 수 있다.

SVM은 클래스 간의 최소거리를 최대화하는 초평면 (Hyperplane)을 찾아냄으로써 클래스를 분류하는 방법이다. 만약 클래스를 최적으로 분리하는 초평면 (Optimal Separating Hyperplane, OSH)이 존재하지 않을 경우에는 원 데이터를 보다 높은 차원의 공간으로 사상시키고 새로운 영역에서 OSH를 찾게 된다. 그림 2는 이 개념을 그림으로 보여주고 있다. 원래 차원에서는 OSH 대신 복잡한 비선형 판별함수를 이용하여 클래스를 구분하고 있다. 판별함수가 복잡한 구조를 가지면 학습데이터에 대한 의존도가 높아져

서 과적합 (Over-fitting)이 발생하기 쉽고 따라서 분류의 일반화 능력도 저하된다. SVM은 이 문제를 해결하기 위해 비선형적인 변수변환( $\phi(x)$ )을 이용한다. 보다 높은 차원에서서는 그림과 같이 선형적인 판별함수 즉 OSH를 쉽게 찾을 수 있기 때문이다.

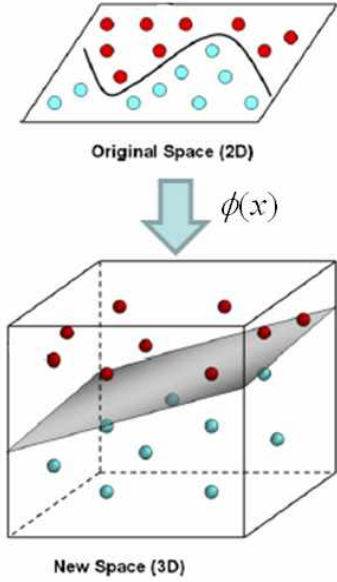


그림 2. 고차변환을 통한 OSH의 발견  
Fig. 2. An Example of OSH Found in the Higher Dimensional Feature Space[7]

바로 이 아이디어를 함수관계식을 추정하는 회귀문제로 확장한 것이 바로 Support Vector Regression (SVR)이다. SVR은 회귀를 포함한 다양한 문제를 다룰 수 있으며 최근에는 이상치 데이터분석에도 활용된 바 있다[8]. 물론 상시 감시를 위한 AA 문제도 역시 회귀문제에 해당되므로 SVR로 다룰 수 있다. 단 SVR은 단일 출력변수에 대한 모델이므로 AA를 다루기 위해서는 다음 그림 3과 같이 출력변수 별로 다수의 SVR이 필요하다.

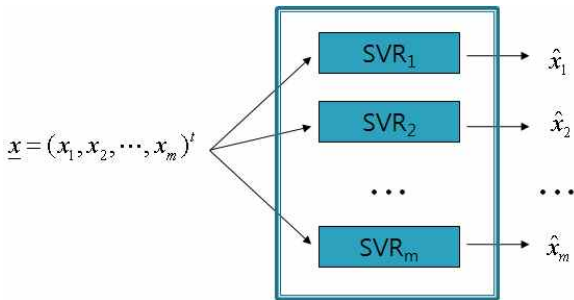


그림 3. AA를 위한 SVR 구조  
Fig. 3. Support Vector Regression for Auto-association

우선 학습데이터셋을 다음과 같이 정의하기로 한다.

$$D = \{(x_i, y_i) | x_i, y_i \in R^m \text{ for } i = 1, 2, \dots, n\} \quad (3)$$

식 (3)에서  $x_i$ 와  $y_i$ 는 각각 입력과 출력을 나타내는  $m$  차원의 벡터이다. 그림 3에서는 출력이 입력의 예측치이므

로  $y_i = \hat{x}_i$ 로 나타내기로 한다.  $k$ 번째 출력에 대한 SVR로 구하게 될 최적회귀식 (Optimum Regression Line, ORL)은 다음 식 (4)와 같이 나타낼 수 있다.

$$f_k(x) = w_k^T x + b_k \quad (4)$$

단  $k = 1, 2, \dots, m$ . 여기서  $k$ 번째 출력변수  $y^{(k)}$ 에 대한 손실함수를 다음과 같이 정의한다.

$$L_k(y_k) = \begin{cases} 0, & |f_k(x) - y^{(k)}| < \epsilon_k \\ |f_k(x) - y^{(k)}| - \epsilon_k, & \text{elsewhere} \end{cases} \quad (5)$$

식 (5)는  $\epsilon$ -insensitive Loss Function이라고 부르는 데 SVR에서 가장 많이 사용되는 손실함수이다. 그러면,  $y^{(k)}$ 에 대한 ORL을 구하기 위한 최적화문제는 다음과 같이 나타낼 수 있다.

$$\begin{aligned} \text{Minimize } \Phi(w_k, \xi_k) &= \frac{1}{2} w_k^T w_k + C_k \sum_{i=1}^n (\xi_{k,i} + \xi_{k,i}^*) \\ \text{s.t. } & y_i^{(k)} - w_k^T x_i - b \leq \epsilon_k + \xi_{k,i} \\ & w_k^T x_i + b - y_i^{(k)} \leq \epsilon_k + \xi_{k,i}^* \\ & \epsilon_k, \xi_{k,i}, \xi_{k,i}^* \geq 0 \text{ for } i = 1, 2, \dots, n \end{aligned} \quad (6)$$

역시  $k = 1, 2, \dots, m$  이고  $\xi_{k,i}$ 와  $\xi_{k,i}^*$ 는 그림 5에서 보는 바와 같이  $x_i$ 가  $\epsilon$ -insensitive zone (이를  $\epsilon$ -tube라고 함)을 벗어난 경우를 위한 여유변수 (Slack Variable)를 의미한다.

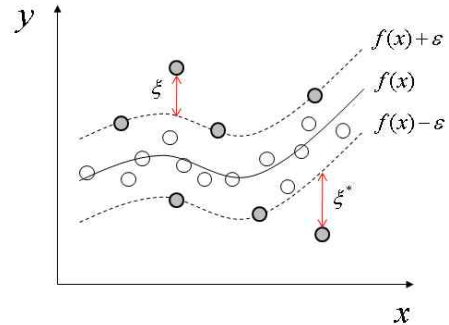


그림 4. SVR에서 손실이 발생하지 않는  $\epsilon$ -tube  
Fig. 4. Support Vector Regression with  $\epsilon$ -tube

식 (6)의 최적화문제는 다음과 같은 쌍대문제 (Dual Problem)로 바꾸어 나타낼 수 있다.

$$\begin{aligned} \max_{\lambda_k, \lambda_k^*} & \left\{ -\frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\lambda_{k,i} - \lambda_{k,i}^*) (\lambda_{k,j} - \lambda_{k,j}^*) x_i^T x_j \right. \\ & \left. + \sum_{i=1}^n [\lambda_{k,i} (y_i^{(k)} - \epsilon_k) - \lambda_{k,i}^* (y_i^{(k)} + \epsilon_k)] \right\} \\ \text{s.t. } & 0 \leq \lambda_{k,i}, \lambda_{k,i}^* \leq C_k \text{ for } i = 1, 2, \dots, n \\ & \sum_{i=1}^n (\lambda_{k,i} - \lambda_{k,i}^*) = 0 \end{aligned} \quad (7)$$

단  $k = 1, 2, \dots, m$ . 최적화문제 (7)을 풀어 라그랑지승수  $\lambda_{k,i}$ 와  $\lambda_{k,i}^*$ 를 구하면 다음과 같이 SVR의  $k$ 번째 출력변수에 대한 ORL을 식 (8)과 같이 결정할 수 있다.

$$f_k(x) = w_k^{*T} x + b_k^* = \sum_{i=1}^n (\lambda_{k,i} - \lambda_{k,i}^*) x_i^T x + b_k^* \quad (8)$$

지금까지 살펴본 내용은 최적의 선형회귀식을 찾는 과정이었고 비선형적인 고차변환  $\phi(x)$ 을 이용할 경우는  $K(x_i, x) = \phi(x_i)^T \phi(x)$ 를 식 (8)에 대신 넣어주면 된다.

$$f_k(x) = \sum_{i=1}^n (\lambda_{k,i} - \lambda_{k,i}^*) K(x_i, x) + b_k^* \quad (9)$$

여기서  $K(x_i, x) = \phi(x_i)^T \phi(x)$ 를 커널함수라고 부른다. 한편 식 (9)의 바이어스항은 다음과 같이 임의의 Support Vector인  $x_r$ 과  $x_s$ 를 이용하여 다음과 같이 계산된다.

$$b_k^* = -\frac{1}{2} \sum_{i=1}^n (\lambda_{k,i} - \lambda_{k,i}^*) [K(x_i, x_r) + K(x_i, x_s)] \quad (10)$$

이 과정을 반복하여 각각의 출력에 대한 총  $m$ 개의 SVR을 얻고 그림 3의 구조를 구축함으로써 원자력발전소의 센서기반 상시감시문제를 다룰 수 있게 된다. Auto-association을 다루기 위해 지금까지 설명한 SVR을 본 논문에서는 AASVR (Auto-associative Support Vector Regression)이라 칭하기로 한다.

### 3.2 Partial Least Square Regression (PLSR)

한편 입력변수의 상관관계 (Correlation)를 수용하기 위해 그림 3의 구조에 특징추출단계로서 주성분분석 (Principal Component Analysis, PCA)의 적용을 생각할 수 있다[9, 10]. 하지만 PCA는 변수압축에 입력변수들만 고려하기 때문에 모델링 및 민감도개선 효과가 제한적일 수 있으므로 대신 본 논문에서는 입력변수와 출력변수를 동시에 고려하는 기법인 PLSR을 적용해 보기로 한다. PLSR은 다변량 입력에서 흔히 나타나는 상관관계가 존재할 때 혹은 입력변수가 매우 많을 때 유용한 회귀분석기법의 하나로 측정 분야에서 성분분석 등의 용도로 주로 활용되고 있다[11].

입력변수와 출력변수를 모두 포함한 데이터셋을 행렬  $(X, Y)$ 로 표현하기로 한다. 즉 행렬  $X \in R^{n \times m}$ 는 입력변수  $m$ 개이고 관측표본이  $n$ 개로 구성되는 모델입력 데이터가 되고 행렬  $Y \in R^{n \times p}$ 는 출력변수가  $p$ 개이고 관측표본이  $n$ 개로 구성되는 모델출력이 된다. 여기서  $A$ 개의 모델기저변수를  $T = (t_1, t_2, \dots, t_A)$ 라고 하면, 차원축소를 위해 PCA는  $X = TP^T$ 의 관계를 가정하지만 PLSR은 동시에  $Y$ 를 포함하는 것이 다르다. 즉 PLSR에서의 변수변환은 다음 관계식을 만족하도록 이루어진다.

$$\begin{aligned} X &= TP^T + E \\ Y &= TQ^T + F \end{aligned} \quad (11)$$

식 (11)에서  $E$ 와  $F$ 는 각각  $X$ 와  $Y$ 에 대한 모델잔차이다. 식 (11)을 이용하여  $A$ 개의 기저변수 값을 순차적으로 구하는 절차는 NIPALS 알고리즘[12]이 가장 대표적이다. PLSR에 관한 보다 자세한 설명과 이론은 역시 참고문헌 [12]를 참조하기 바란다.

## 4. 사례연구

실제 발전소 데이터셋을 통해 제안된 AASVR을 적용하여 보기로 한다. 원전에는 많은 센서들이 동작하고 있어 센서기능을 실시간으로 감시하고 또한 문제가 발생한 센서를

효율적으로 탐지하는 것은 대단히 중요하다. 사례연구를 통해 제안된 기법이 센서 입력력을 제대로 모사할 수 있는지 검토하고, 아울러 수치실험을 통해 센서의 고장을 효과적으로 검출할 수 있는지 살펴보고자 한다. 데이터는 고리원전 3호기의 1차 및 2차 측에서 11개 센서를 이용하였다. 대상이 된 센서는 각각 원자로 출력, 가압기 수위, 증기발생기 증기유량 및 압력, 증기발생기 광역 및 협역수위, 터빈출력, 충전유량, 잔열제거계통유량, 원자로상부 냉각수온도, 증기발생기 주급수유량 등이다. 1분 간격으로 관측된 2290개의 데이터를 동일한 크기를 갖는 5개의 하부데이터셋으로 분할한 후 그 중 하나를 훈련용으로 삼았다. 다음 그림 6은 훈련데이터셋을 0과 1 사이로 변환한 결과를 한꺼번에 도시한 것이다.

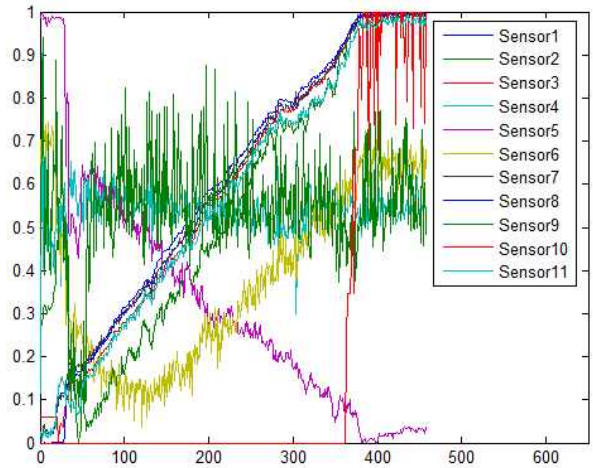


그림 5. 발전소 센서데이터셋  
Fig. 5. Normalized Sensor Data for Training

센서에 따라 관측값의 차이가 크기 때문에 모든 센서 데이터를 일정구간으로 변환해서 다루는 것은 필수적이었고 판단된다. 비교실험은 2단계로 이루어진다. 먼저 SVR과 PLSR을 이용하여 각각 AA 모델을 구축하고 정확도를 평가한다. 다음 특정센서에 드리프트를 가한 후 여기에 각 AA 모델이 어떻게 반응하는지 민감도를 이용하여 평가한다. 이 모든 과정은 MATLAB 2009b 기반의 코딩을 통해 이루어진다.

우선, SVR과 PLSR에 의한 AA 모델링을 각각 수행한다. 특히 SVR 회귀식을 추정하기 위한 툴박스는 많은 연구자들이 제공하고 있으나 본 논문에서는 사용이 비교적 간단한 Libsvm[13]을 이용하기로 한다. 또한 PLSR 회귀식은 MATLAB 2009b에 내장된 plsregress 함수를 그대로 이용하여 추정한다. 예비적인 분석을 위해 먼저 주성분분석을 적용한 결과 11개 주성분에 대한 분산값은 그림 6과 같이 나타났다. 이처럼 주성분의 분산값을 크기 순으로 도시한 것을 Scree Plot이라고 부르는 데 유의한 주성분은 상위 4개임을 알 수 있다. 이들 4개는 분산값을 누적했을 때 전체의 99.2%를 차지하는 것으로 나타났다. 따라서 본 실험에서는 PLSR을 위해 주성분을 상위 4개만 고려하기로 한다.

한편 AASVR 적용을 위해서는 3개의 파라미터  $\sigma$ ,  $\nu$ ,  $C$ 가 결정되어야 한다.  $\sigma$ 는 커널함수를 Radial Basis Function (RBF)으로 택했을 때 그 대역폭을 의미한다.

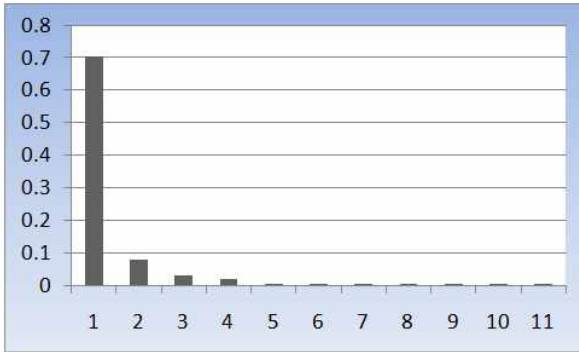


그림 6. Scree Plot: 주성분의 분산크기 비교

Fig. 6. Scree Plot: Variances of Principal Components

$\epsilon$ -tube의 크기와 관계가 있는 파라미터  $\nu$ 는 Support Vector의 비율을 뜻한다. C는 식 (6)의 목적함수  $\Phi(w_k, \xi_k)$ 에 필요한 페널티계수이다. 이들 값은 개별 센서의 SVR 모델마다 별도로 정할 수 있지만 본 사례에서는 모델에 공통인 것으로 가정한다. 특히 AASVR의 모델링 성능은 이들 파라미터의 미세조정에 의존할 수 있으므로 별도의 예비실험이 필요하다. 이러한 목적으로 활용되는 기법으로는 유전자알고리즘[14], Grid Search[5], 반응표면분석법[10] 등이 있다. 일단 본 논문에서는 Libsvm에서 제공하는 초기값을 그대로 사용하여 AASVR을 구축하기로 한다. 그림 5의 발전소 데이터를 이용해 SVR과 PLSR을 적용한 결과 MSD로 측정되는 모델정확도 (Model Accuracy)는 11개 센서 별로 볼 때 다음 그림 7과 같이 나타났다.

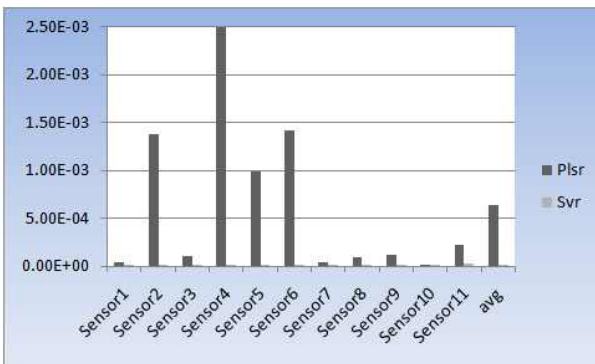


그림 7. SVR과 PLSR의 모델 정확도

Fig. 7. Model Accuracy of SVR and PLSR

모든 센서에 대해 SVR 쪽이 정확도가 높은 것을 알 수 있다. 평균적으로는 SVR과 PLSR의 정확도는 0.00001과 0.00063이었으므로 이를 백분율로 환산하면 각각 0.3%와 2.5%에 해당된다. 즉 정상적인 운전환경 하에서 신호의 참값을 재생할 때의 오류율은 SVR쪽이 훨씬 낮다고 말할 수 있다.

다음은 민감도를 평가하기로 한다. 이를 위해 센서 7에 드리프트가 있다고 가정한다. 드리프트의 크기는 시간당 5%로 가해졌는데 원신호와 드리프트가 가미된 신호를 그림으로 나타내면 다음과 같다. 시간이 흐를수록 드리프트 영향이 커짐을 알 수 있는데, 민감도분석은 결국 센서 7의 드리프트에 AA 모델의 출력이 얼마나 민감한가를 알아보는 것이다.

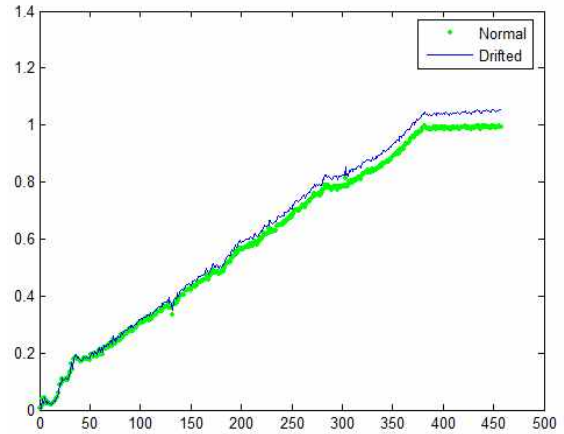


그림 8. 센서 7의 정상신호와 드리프트를 가미한 신호

Fig. 8. Normal and Drifted Signals of Sensor 7

민감도는  $i$ 번째 센서입력의 변화량으로 인해 발생한  $j$ 번째 센서출력의 변화량으로 다음과 같이 정의할 수 있다.

$$S_{ij} = \frac{1}{n} \sum_{t=1}^n \left| \frac{\hat{x}_j^{drift}(t) - \hat{x}_j(t)}{\hat{x}_i^{drift}(t) - \hat{x}_i(t)} \right|$$

단  $j=1,2,\dots,m$ 이고  $i=7$ 이다. 위의 식에서,  $x_i(t)$ 와  $x_i^{drift}(t)$ 는 시점  $t$ 에서 센서  $i$ 의 정상적인 입력과 드리프트된 입력을 각각 의미하며  $\hat{x}_i(t)$ 와  $\hat{x}_i^{drift}(t)$ 는 각각의 해당출력을 나타낸다. 여기서  $S_{ij}$ 는  $i=j$ 일 때 자기민감도를 의미하고  $i \neq j$ 이면 교차민감도를 나타냄을 알 수 있다. AA 모델링 결과를 이용하여 민감도를 계산한 결과는 다음과 같다.

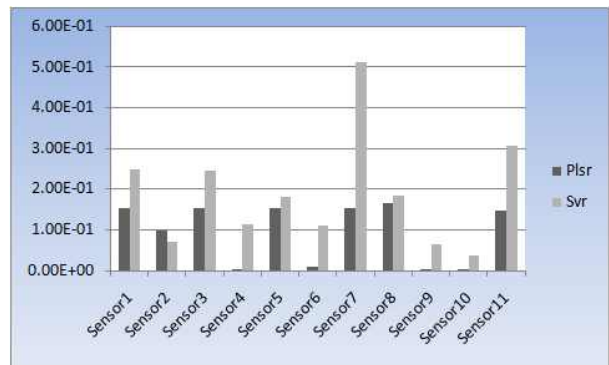


그림 9. 센서 7에 드리프트가 있을 경우 SVR과 PLSR의 민감도

Fig. 9. Sensitivities of SVR and PLSR When There is a Drift in Sensor 7

센서 2를 제외한 모든 센서에서 PLSR가 다소 나은 민감도를 갖는 것으로 나타났다. 평균적으로는 PLSR이 0.09474이고 AASVR이 0.18883이었다. 특히 자기민감도에 해당되는 센서 7의 경우도 역시 PLSR이 0.0154이고 AASVR이 0.0513이어서 PLSR 쪽이 자기수정능력이 우수한 것으로 확인되었다. 즉 PLSR은 드리프트 탐지능력이 3배 정도 높은 것으로 나타났다.

## 5. 결론

본 논문에서는 원자력발전소의 상시감시를 다루기 위한 Auto-association 문제를 소개하고 SVR 및 PLSR을 이용한 모델링에 대해 연구하였다. Auto-association를 위해 Neural Network, Multivariate State Estimation Technique, Kernel Regression 등을 이용한 모델이 개발된 바 있으나 대부분 특허로 보호받고 있어 접근이 제한되는 현실을 감안할 때 범용성이 있는 SVR과 PLSR의 활용방안을 제시했다는 데서 본 연구의 의의가 있다고 판단되며 국내 발전소에서 독자적 상시감시기술을 확보하는 데에도 유용한 자료가 될 것으로 기대된다.

원자력발전소 현장의 감시계통 데이터를 이용하여 Auto-association 성능을 검토한 결과, 정상적인 운전조건 하에서 신호를 재생하는 능력은 SVR이 훨씬 우수한 반면 드리프트가 발생했을 때의 탐지력은 PLSR이 다소 나은 것으로 나타났다. 정확도 면에서 SVR이 훨씬 우수한 것은 SVR이 고차의 비선형변환을 이용하는 반면 PLSR은 선형 변환에 의존하기 때문인 것으로 판단된다. 반면 민감도 관점에서 PLSR이 더 나은 결과를 보여준 것은 다수의 SVR을 결합하더라도 AASVR로는 여러 센서간의 상관관계를 충분히 수용하는 데 한계가 있음을 시사한다. 또한 SVR을 이용할 때에는 세 가지의 모델 파라미터의 선택에 따라 Auto-association 성능이 영향을 받을 수 있음을 유의해야 하고, 향후 SVR과 PLSR의 장점을 동시에 확보할 수 있는 모델링 기법에 대한 연구가 유용할 것으로 전망된다.

## 참고 문헌

[1] J. W. Hines, D. Garvey, R. Seibert and A. Usynin, *Technical Review of On-line Monitoring Techniques for Performance Assessment*, Vol. 2, Technical Report, U. S. Nuclear Regulatory Commission, 2007.

[2] D. Wrest, J. W. Hines and R. Uhrig, "Instrument Surveillance and Calibration Verification Through Plant Wide Monitoring Using Autoassociative Neural Networks," *Proceedings of The 1996 American Nuclear Society International Meeting on Nuclear Plant Instrumentation, Control and Human Machine Interface Technologies*, University Park, PA, 1996.

[3] J. W. Hines and A. Usynin, "MSET Performance Optimization Through Regularization," *Nuclear Engineering and Technology*, Vol. 38, No. 2, pp. 177-184, 2005.

[4] V. Vapnik, *Statistical Learning Theory*, Wiley, New York, 1998.

[5] L. Zheng, H. Zhou, K. Cen and C. Wang, "A Comparative Study of Optimization Algorithms for Low NOx Combustion Modification at a Coal-fired Utility Boiler," *Expert Systems with Applications*, Vol. 36, pp. 2780-2793, 2009.

[6] 김성준, "SVM을 이용한 웨이블릿기반 프로파일 분류에 관한 연구," *한국지능시스템학회논문지*, 18권, 5호, pp. 718-723, 2008.

[7] F. J. Martinez-de-Pison, C. Barreto, A. Pernia and F. Alba, "Modelling of an elastomer profile extrusion process using support vector machines," *Journal of Materials Processing Technology*, Vol. 197, pp. 161-169, 2008.

[8] 전성해, "Support Vector Regression을 이용한 이상치 데이터분석," *한국지능시스템학회논문지*, 18권, 6호, pp. 876-880, 2008.

[9] R. Rosipal, M. Girolami, L. Trejo and A. Cichocki, "Kernel PCA for Feature Extraction and Denoising in Nonlinear Regression," *Neural Computing & Applications*, Vol. 10, pp. 231-243, 2001.

[10] 김성준, 서인용, "센서기반 상시감시를 위한 Auto-association 모델," *대한산업공학회 추계학술회의논문집*, 한양대학교, A07-3, 2008.

[11] 이대중, 박상영, 정남정, 이혜근, 박진일, 전명근, "PLS기반 c-퍼지 모델트리를 이용한 클로로필-a 농도 예측," *한국지능시스템학회논문지*, 16권, 6호, pp. 777-784, 2006.

[12] B. S. Dayal and J. F. MacGregor, "Improved PLS Algorithm," *Journal of Chemometrics*, Vol. 11, pp. 73-85, 1997.

[13] C. Chang and C. Lin, <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>, 2000.

[14] M. G. Na, H. Y. Yang and D. H. Lim, "A Soft-Sensing Model for Feedwater Flow Rate Using Fuzzy Support Vector Regression," *Nuclear Engineering and Technology*, Vol. 40, No. 1, pp. 69-76, 2007.

## 저자 소개



**김성준 (Kim, Seong-Jun)**

1989년 연세대학교 응용통계학과 학사.  
1991년 카이스트 산업공학과 석사.  
1995년 카이스트 산업공학과 박사.  
현재 강릉원주대학교 산업공학과 교수.

관심분야 : 데이터마이닝, 품질관리  
Phone : 033-640-2375  
E-mail : sjkim@kangnung.ac.kr



**서인용 (Seo, In-Yong)**

1984년 성균관대학교 전기공학과 학사.  
1989년 부산대학교 전기공학과 석사.  
2003년 브라운대학교 전기공학과 박사.  
현재 한전 전력연구원 책임연구원.

관심분야 : 모델링, 시뮬레이션, 송배전  
Phone : 042-865-5932  
E-mail : iyseo@kepri.re.kr