

## SVM에 기반한 비선형 부분최소자승법을 이용한 화학공정의 이상감지

이창준, 이동언, 최승준, 송상옥, 김구회\*, 윤인섭  
서울대학교 응용화학부  
서울대학교 화학공정신기술연구소\*

### Chemical Process Fault Detection Using Non-Linear Partial Least Square Method Based on the Support Vector Machine

Chang Jun Lee, Dong Eon Lee, Seung Jun Choi, Sang Ok Song,  
Ku Hwi Kim, En Sup Yoon  
School of Chemical Engineering, Seoul National University  
Institute of Chemical Process, Seoul National University\*

#### 서론

화학공장은 고성능의 복잡한 장비와 여러 공정으로 구성되며 재순환 흐름(recycle flow)과 반응(reaction)으로 인한 비선형성(nonlinearity)을 가지고 있다.

장치가 집약화되어 있는 화학공정은 실시간으로 적절한 제어나 모니터링이 이루어지지 않을 경우 정상조업범위를 벗어날 가능성이 커지며 이는 생산품의 품질저하와 공정사고로 이어져 막대한 물질적 피해를 유발할 수 있다. 따라서 실시간 모니터링과 실시간 제어, 실시간 제어나 모니터링을 위한 실시간 변수추론이 중요한 의미를 가진다.

효율적인 이상감지 기법이 도출되기 위해서는 우선적으로 화학공정 상태를 나타내는 공정 데이터를 효과적으로 처리할 수 있어야 한다. 수많은 변수가 존재하는 화학공정 시스템은 개별적 감지를 수행하는 단변량 통계분석방법을 사용하기 보다는 다변량 통계분석방법을 사용하여야 한다.

다변량 통계적 분석기법들은 화학뿐만 아니라 제약, 의약, 고분자, 생명과학과 같은 공학분야에서 널리 사용되고 있으며, 화공분야에서는 공정의 이상감지, 실시간 추정이 어려운 물성추정을 위한 추론모델의 개발 등에 응용되고 있다. 그러나 이를 사용함에 있어서도 변수들간의 높은 상관성에서 기인하는 multi-collinearity문제, 데이터 측정회수 보다 데이터를 이루는 변수가 많은 상황에서 기인하는 dimensionality 문제, 변수들간의 비선형에서 기인하는 non-linearity 문제 등의 난점이 존재한다. 이러한 난점을 극복하고자 보다 효율적인 data mining 기법을 이용하여 데이터들간의 의존성(dependency)이나 상관구조(correlation structure)와 같은 유용한 정보를 찾아 내어야 한다. 그리고 나아가 공정에 대한 해석을 쉽게 하고 획득한 데이터로부터 미지의 값을 추정함에 있어 보다 향상된 예측성능을 가지게 하여야 한다.

본 연구에서는 위에서 제시된 여러 난점들을 극복하고 보다 빠르고 정확한 화학공장 이상진단을 위해 선형 PCA와 SVM(Support Vector Machine)을 적용하여 다변량 통계기법의 하나인 부분최소자승법(non-linear partial least square)을 이용하였다.

이와 같이 구성된 부분최소자승법 모델을 가지고 Tennessee Eastman 공정에 적용하여 보았다.

#### 본론

##### PCA(Principal Component Analysis)

1901년에 Pearson에 의해 개발된 PCA는 Hotelling이 다변량 문제들에 있어서 변수들의 상호 연관 구조들을 분석하기 위하여 많이 사용하였다.

PCA를 이해하는 방법에는 기하학적인 방법과 수학적 방법이 있는데, 기하학적으로 PCA는 데이터 공간에서 데이터분포가 넓은 축부터 차례대로 서로 직교하도록 새로운 축들을 하나씩 정의하고 이들 축으로 정사영 되는 값을 그 축에 대한 새로운 좌표값으로 가지는 것을 의미한다. 이때 데이터의 상관관계를 더 쉽게 분석하기 위해 새롭게 정의된 축을 PC(Principal Component)라고 하며 데이터를 이 축에 투영시켜 얻은 정사영 값들을 그 축에 대한 Score vector라고 한다.

**NLPLS(non-linear partial least square)모델**

PLS(partial least squares)는 데이터의 노이즈가 심하고 변수간의 상관관계가 강하며, 제한된 수의 데이터만이 존재하는 문제들에 대해 강력한 선형회귀성능을 보이는 다변량 통계분석법의 하나이다.

이는 X블록에 대해 PCA를 수행한 후 얻어지는 PC score 행렬에 Y블록을 다중선형회귀 시키는 PCR(principal component regression)방법을 확장시켜 Y블록에 대해서도 PCA를 수행하여 score 벡터를 구하고 이것을 X블록의 score벡터와 회귀시키는 방법이다.

PLS는 외적관계(outer relation)와 내적관계(inner relation)를 독립적으로 수행하는 회귀모델로 볼 수 있다.

그러나 이러한 PLS모델은 X, Y 블록간에 비선형성이 있을 경우는 적절치 않다. NLPLS방법은 크게 내적관계에 비선형 회귀모델을 이용하는 경우와 외적관계와 내적관계에 모두 비선형 모델을 사용하는 경우로 나누어 볼 수 있다. 일반적인 모양은 다음 그림과 같다. 본 연구에서는 내적관계에 비선형 모델인 SVM Regression을 적용하였다. 일반적인 모양은 다음 그림과 같다.

본 연구에서 외적변수변환은 기존의 선형 PCA를 이용하였고, 내적회귀모델로는 SVM(Support

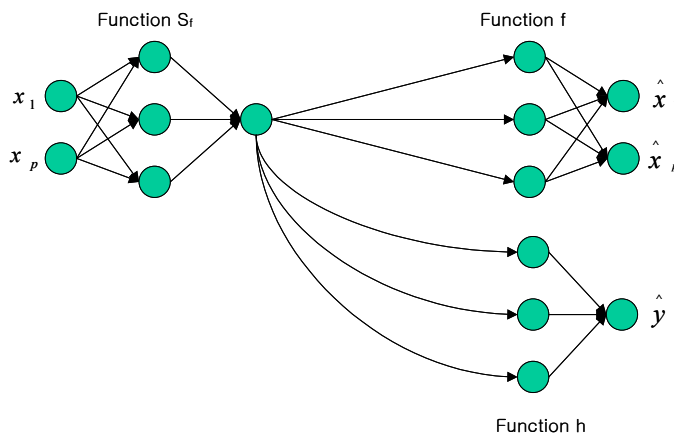


Fig.1 Non-linear partial least square model

Vector Machine)을 사용하였다.

즉 요약하자면 PCA를 이용한 외적변환을 통해 입력, 출력데이터는 잠재변수로 사영되며, 이렇게 얻어진 잠재변수간의 회귀는 SVM Regression을 통해서 이루어진다.

**SVM(Support Vector Machine)**

본 연구에서는 SVM Regression을 내적변환에 사용하였다.

Fig 2. 는  $\epsilon$ -intensive band를 이용한 비선형 Regression Function을 보여주고 있다. 본 연구에서는 SVM의 Loss function중  $\epsilon$ -intensive loss function을 이용하였다.

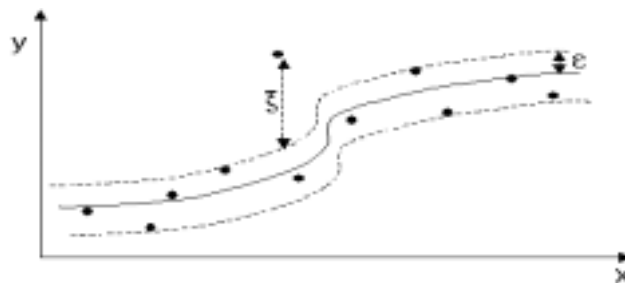


Fig 2. The intensive band for a non-linear regression function

$\epsilon$ -intensive loss function은 다음과 같이 정의한다.

$$L^\epsilon(x, y, f) = |y - f(x)|_\epsilon = \max(0, |y - f(x)| - \epsilon) \quad \text{Eq. (1)}$$

이때 Margin Slack Variable은 다음과 같이 주어진다.

$$\xi((x_i, y_i), f, \theta, \gamma) = L^{\theta-\gamma}(x_i, y_i, f) \quad \text{Eq. (2)}$$

위의 Loss Function을 가지고 다음을 최소화해 가는 과정을 거친다.

$$\text{minimise } \|w\|^2 + C \sum_{i=1}^{\ell} (\xi_i^2 + \hat{\xi}_i^2) \quad \text{Eq. (3)}$$

$$\text{subject to } (\langle w \cdot x_i \rangle + b) - y_i \leq \epsilon + \xi_i, \quad i = 1, \dots, \ell$$

$$y_i - (\langle w \cdot x_i \rangle + b) \leq \epsilon + \hat{\xi}_i, \quad i = 1, \dots, \ell$$

$$\xi_i, \hat{\xi}_i \geq 0, \quad i = 1, \dots, \ell$$

**예제**

개선된 PLS의 성능을 평가하기 위하여 Tennessee Eastman 공정을 대상으로 이상진단을 수행하였다. 공정 반응기내의 촉매의 활성이 정상조업모드의 80%로 저하되는 slow kinetic drift 이상 상황에 대하여 이상진단을 수행하였다.

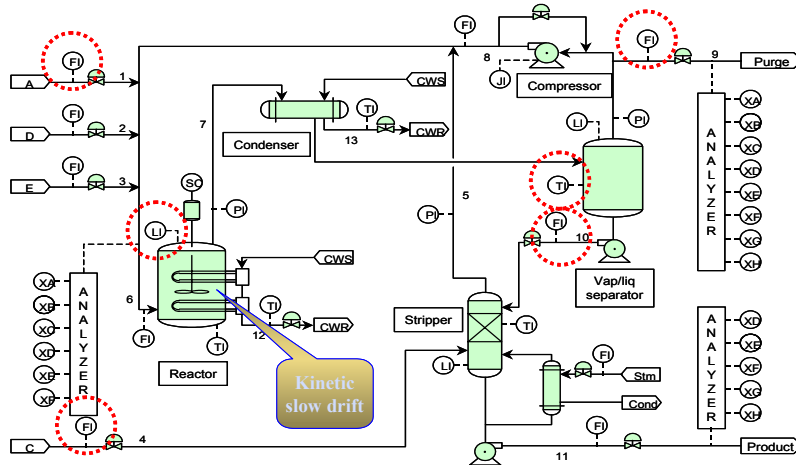


Fig. 3 Tennessee Eastman Process

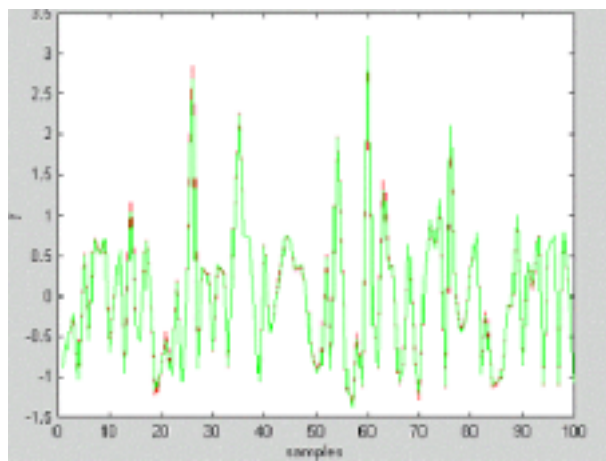


Fig 4. Proposed PLS prediction result

Linear PLS, Neural Network PLS, Proposed PLS의 결과를 비교해 본 결과 다음표와 같았다.

Table 1. Results comparison using one PC

	Linear PLS	Neural Network PLS	Proposed PLS
MSEP	0.6556	0.1980	0.0189
Infinite Norm	2.3400	0.8554	0.3017

### 결론

본 연구에서는 보다 빠르고 정확한 화학공장 이상진단을 위해 기존의 PLS(partial least square)에 SVM(Support Vector Machine)을 적용하여 다변량 통계기법의 하나인 부분최소자승법(non-linear partial least square)을 이용한 모델을 제안하였고 그 모델을 이용하여 Tennessee 공정에 적용시켜 본 결과 우수한 성능을 지님을 확인할 수 있었다.

### 감사

본 연구는 교육부를 통한 두뇌한국 21 사업과 한국산업기술평가원을 통한 한-브라질 협력 청정생산을 위한 통합공정 모니터링·이상진단 시스템 개발사업 지원금에 의한 것입니다.

### 참고문헌

1. G. Baffi, E. B. Martin, A. J. Morris, "Non-linear projection to latent structures revisited(the neural network PLS algorithm)", Computers and Chemical Engineering, 23, pp. 1293-1307(1999)
2. T. Kohonen, Self-Organization and Associative Memory, 2nd Ed. Berlin: Springer-Verlag, (1987)
3. Nello Cristianini. "An Introduction to Support Vector Machines", Cambridge, (2000)
4. Geladi, P., B. R. Kowalski, "Partial Least-Squares Regression :A Tutorial,' Anal. Chim. Acta, 185, 1(1986a)
5. L.H. Chiang, E.L. Rusell, R.D. Braatz "Fault Detection and Diagnosis in Industrial Systems", Berlin: Springer-Verlag, (2000)
5. 정신호, "비선형 주성분 해석을 이용한 화학공정의 이상감지에 관한 연구", 석사학위논문, 서울대학교 응용화학부(2000)
6. 손정현, "비선형 부분최소자승법을 이용한 화학공정의 추론모델개발에 관한 연구", 석사학위논문, 서울대학교 응용화학부(2001)