

SVM에 기반한 비선형 부분최소자승법과 클러스터링 기법을 이용한 화학공정감시에 관한 연구

이창준*, 송상옥, 박송원¹, 윤인섭
서울대학교 응용화학부
상파울로 대학교 화학공학부¹
(jun96@pslab.snu.ac.kr*)

Chemical Process Fault Detection and Diagnosis Using Non-Linear Partial Least Square Method Based on the Support Vector Machine And Clustering Analysis

Chang Jun Lee*, Sang Ok Song, Song Won Park¹, En Sup Yoon
School of Chemical Engineering, Seoul National University
Department of Chemical Engineering, University Sao Paulo¹
(jun96@pslab.snu.ac.kr*)

서론

장치가 집약 화되어 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 및 진단기법이 도출되기 위해서는 우선적으로 화학공정 상태를 나타내는 공정 데이터를 효과적으로 처리할 수 있어야 한다. 수많은 변수가 존재하는 화학공정 시스템은 개별적 감지를 수행하는 다변량 통계분석법을 사용하여야 한다.

다변량 통계적 분석기법은 화공분야에서는 공정의 이상감지, 실시간 추정이 어려운 물성추정을 위한 추론모델의 개발 등에 응용되고 있다. 그러나 이를 사용함에 있어서도 변수들 간의 Multi-collinearity 문제, Dimensionality 문제, Non-linearity 문제 등의 난점이 존재한다. 이러한 난점을 극복하고자 보다 효율적인 Data mining 기법을 이용하여 데이터들 간의 의존성(Dependency)이나 상관구조(Correlation structure)와 같은 유용한 정보를 찾아내어야 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 화학공정 감시를 위하여 SVM(Support Vector Machine)을 적용한 다변량 통계기법의 하나인 부분최소자승법(non-linear partial least square)과 클러스터링 기법(Clustering Analysis)을 이용한 알고리즘을 제안하였다.

이와 같이 구성된 모델을 가지고 Tennessee Eastman 공정에 적용하여 보았다.

본론

1. PLS(Partial Least Square)

PLS는 두 개의 변수데이터 블록간의 예측성능을 최대로 하는 잠재변수공간을 구하는데 초점을 두는 것이다.

각각의 X, Y 블록에 대해서 독립적으로 PCA를 적용하지 않고, 회귀성능을 높이기 위해 두 개의 score 공간간의 회귀성능이 최대가 되는 score 공간을 얻도록 외적관계를 수정한

것이다. 즉, X와 Y블록간의 PCA를 통해 score행렬을 구하는 과정에서 X블록의 각 열을 Y블록의 score 벡터 u 에 회귀시켜 w 를 얻고 Y블록의 각 열을 X블록의 score 벡터인 t 에 회귀시켜 q 를 구하는 방법을 통하여 서로에 대한 정보를 공유하게 함으로 X, Y블록간의 상관관계가 최대가 되게 한다.

Linear PLS는 다음의 외적, 내적 관계를 갖는다.

- 1) 외적관계: X와 Y블록에 대해서 PCA를 수행하여 각 score 행렬인 T, U를 얻는다.
- 2) 내적관계: T, U 간에 회귀모델인 MLR을 수행한다.

선형이나 근사 선형(nearly linear)문제에 있어서 PLS는 안정적인 예측값을 나타내고, 독립적인 데이터 셋에 대한 예측능력이 뛰어나고, 구해진 score를 해석하기도 용이하다는 것이 알려져 왔다. 그러나 화학공정의 데이터 비선형성을 가지고 있다. 이러한 비선형성에 대해서는 기존의 선형 PLS는 제대로 된 모델을 구성하지 못한다는 치명적인 단점이 있다. 본 연구에서는 내적변환에 비선형구조인 SVM을 이용한 Nonlinear PLS를 이용하였다.

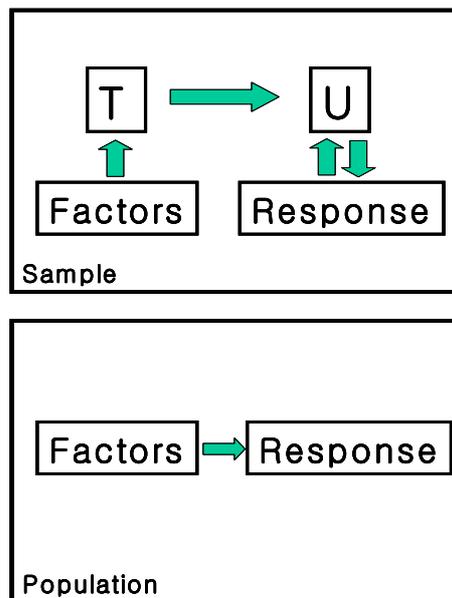


Fig. 1 Indirect modeling by PLS

2. SVM(Support Vector Machine)

SVM(Support Vector Machine)은 통계적 학습 이론에서 유도한 학습 바이어스를 이용한 학습알고리즘으로 훈련되는 고차원의 feature space상의 선형식의 가상 공간을 이용한다. 이 학습 전략은 Vapnik에 의해 소개된 강력한 통계방법으로써 이미 폭 넓은 분야에서 다른 시스템보다 뛰어난 능력을 발휘하고 있다.

본 연구에서는 SVM Regression을 내적변환에 사용하였다.

3. Clustering Analysis

비선형 PLS를 이용하여 상한값과 하한값을 설정한 후 이상이 발생한 경우의 데이터를 다음 식과 같이 변수가 정상상태일 때는 0, 넘는 경우에는 +1을, 하한값 아래로 감소한 경우

에는 -1을 할당하여 경향(tendency)을 표현한다. 이를 바탕으로 K-means Clustering을 수행하였으며 Unknown Data를 이용하여 이상진단을 수행하였다.

4. 사례 연구

본 연구에서는 기존의 PLS의 내적변환을 비선형회귀분석법인 SVM으로 개선한 비선형 PLS를 가지고 이상감지를 수행하였으며 이를 바탕으로 Clustering 기법을 이용하여 이상진단을 수행하였다.

4.1. Tennessee Eastman 공정 이상진단

TE 공정은 반응기, 스트리퍼(stripper), 기액 분리기(vapor/liquid separator), 압축기 및 환류가 있는 공정으로 41개의 측정변수와 12개의 조작변수가 있으며 4가지 반응물 A, B, C, D, E로부터 2가지 G, H가 생성된다. 반응기에서는 액상에 녹아있는 촉매에 의하여 다음과 같은 비가역 발열반응들이 일어난다.

4.2. 이상감지 수행

정상상태의 데이터로 먼저 모델을 만들었다. 이를 이용하여 정상 상태의 범위도 결정하였으며 이상상태의 데이터로 감지를 수행하였다. 52가지 변수 중에서 22가지의 측정변수를 X블록으로 설정하였고 19가지의 성분변수를 Y블록으로 설정하였다.

500×52로 이루어진 정상상태의 데이터를 이용하여 학습하였고 960×52의 정상상태 데이터는 Validation에 이용하였다. 이를 비선형 PLS알고리즘을 적용하여 모델을 만들었으며, 전체 X와 각각의 Y변수와의 관계와의 모델, Y변수의 수인 19개의 모델을 만들었다. 그런 다음 이상상태 데이터를 가지고 이상감지를 수행하였다.

Table 1. Result comparison with other models

	Linear PCA	NNPLS	Proposed NLPLS
MSEP	2.1573	0.1084	0.0759
Infinite Norm	3.0128	1.4668	0.0891

4.3. 이상진단 수행

위에서 언급한 이상감지모델을 이용하여 다음 이상에 대한 이상감지를 수행하였다. 각각의 이상감지시점에서의 변수의 움직임을 정상상태일 때는 0, 상한값을 넘는 경우에는 +1, 하한값 아래로 감소한 경우에는 -1로 할당하여 경향(tendency)을 표현하였다. 하나의 이상에 대하여 새로운 19개의 데이터가 생성되며 15개의 이상에 대한 감지를 수행하였다. 기존의 이상에 포함되는 unknown fault가 들어오면 clustering model을 통하여 이상진단이 가능하게 하였다.

Table 2. Process Faults

Variable	Description	Type
IDV(1)	A/C Feed Ratio, B Composition Constant(Stream 4)	Step
IDV(2)	B Composition, A/C Ratio Constant(Stream 4)	Step
IDV(3)	D Feed Temperature(Stream 2)	Step
IDV(4)	Reactor Cooling Water Inlet Temperature	Step
IDV(5)	Condenser Cooling Water Inlet Temperature	Step
IDV(6)	A Feed Loss(Stream 1)	Step
IDV(7)	C Header Pressure Loss-Reduced Availability(Stream 4)	Step
IDV(8)	A, B, C Feed Composition(Stream 4)	Random Variation
IDV(9)	D Feed Temperature(Stream 2)	Random Variation
IDV(10)	C Feed Temperature(Stream 4)	Random Variation
IDV(11)	Reactor Cooling Water Inlet Temperature	Random Variation
IDV(12)	Condenser Cooling Water Inlet Temperature	Random Variation
IDV(13)	Reaction Kinetics	Slow Drift
IDV(14)	Reactor Cooling Water Valve	Sticking
IDV(15)	Condenser Cooling Water Valve	Sticking

예를 들면 **Fault1**과 같은 결과가 나올 시에는 예측 값이 $[1,0,0,0,0,\dots,0]$ 이 나오도록 하였으며 단일변수 이상이 있는 **Unknown Data**를 이용하여 이를 진단하였다.

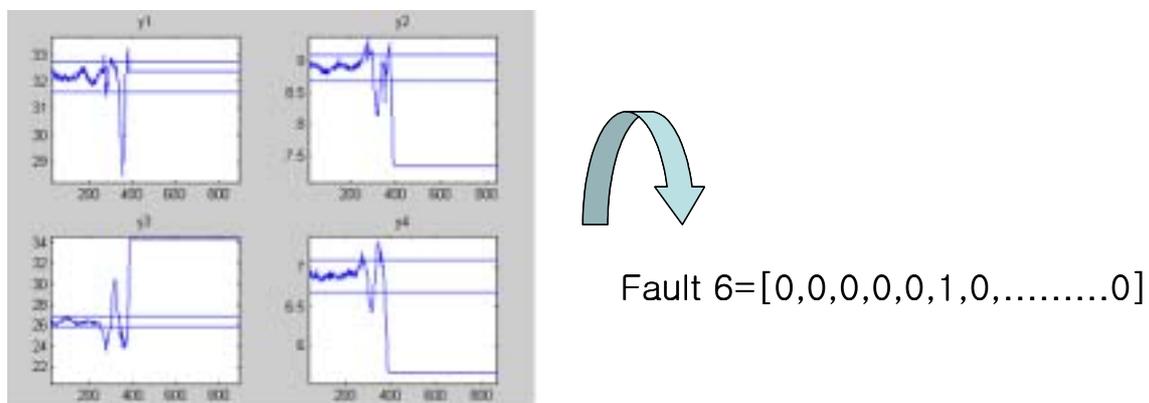


Fig. 2. Fault Diagnosis Using Proposed Model

결론

본 연구에서는 높은 상관성, 비선형성이 포함된 데이터들로 구성되는 공정의 모델링이나 제어에 있어서 아주 우수한 성능을 나타내고 있는 비선형 **PLS** 방법을 이용하여 보다 정확하고 빠른 이상감지를 가능케 하는 모델을 제안하였으며 이를 토대로 **Clustering**을 이용하여 이상진단을 수행하였다. 사전에 미리 파악한 이상을 토대로 보다 신속하며 정확한 화학공정 감시를 수행하였다. 앞으로 위의 이상감지, 진단법과 더불어 **SDG**등의 방법을 이용하여 사전에 파악하지 못한 이상을 진단할 수 있는 모델을 개발할 예정이며 이를 통해 공정의 전반적인 능률 향상을 도모할 예정이다.