

# Multidimensional Visualization에 기반한 화학공정의 이상감지

신봉수, 이창준, 이기백\*, 윤인섭  
서울대학교 화학생명공학부  
충주대학교 화공생명공학과\*

## Chemical Process Fault Detection Using Multidimensional Visualization

BongSu Shin, Chang Jun Lee,  
Gibaek Lee\*, En Sup Yoon

School of Chemical & Biological Engineering, Seoul National University  
Department of Chemical & Biological Engineering, Chungju National University\*

### 1. 서론

가스/화학공정은 고성능의 복잡한 장비와 여러 공정으로 구성되며 재순환 흐름(recycle flow)과 반응으로 인한 비선형성(nonlinearity)을 가지고 있다. 장치가 집약화되어 있는 화학 및 가스공정은 실시간으로 적절한 모니터링이 이루어지지 않을 경우 정상조업범위를 벗어날 가능성이 커지며 이는 생산품의 품질저하와 공정사고로 이어져 막대한 물질적 피해를 유발할 수 있다. 따라서 실시간 모니터링과 실시간 제어를 위한 실시간 변수추론이 중요한 의미를 가지며, 화학 공정 데이터에 대한 정확한 분석을 요하는 많은 이상감지 기법이 연구되고 있다.

효율적인 이상감지 기법이 도출되기 위해서는 우선적으로 화학공정 상태를 나타내는 공정 데이터를 효과적으로 처리할 수 있어야 한다. 이를 위해 본 연구는 효율적 이상감지의 기법 중 하나로서 '다차원 시각화(Multidimensional Visualization)'에 초점을 맞추었다. 일반적으로 기존의 유클리드 공간에서는 서로 직교하는 세 개의 축으로 이루어진 3차원의 공간이 표현가능한데, 유클리드 공간에서는 네 개 이상의 변수 값을 표현하는데 한계가 있다. 그러나 오늘날의 다양하고 복잡한 공정에서 나오는 공정 데이터는 수백 개의 변수 정보를 담고 있는데 이를 3차원 공간에 시각화 하기는 불가능하다. 그러나 공정 데이터에 내재한 변수 정보를 공간에 시각화하기 위한 방법으로 '평행좌표계(Parallel Coordinates)<sup>1)</sup>'를 이용하면 유클리드 공간의 한계를 벗어날 수 있다. 평행좌표계는 여러 개의 변수축을 서로 평행하게 평면에 나열해서 그 축을 차례로 지나가는 꺾인 선들의 집합으로 표현된다. 평행좌표표를 통한 시각화는 그래프, 그림, 색

상 등 다양한 방법을 통해 사용자에게 공정 데이터에 대한 통찰을 제공한다.

이와 같이 제안된 다차원 시각화 모델은 공정 데이터를 분석하고 이상을 감지하는 보다 정확한 모델이 될 수 있으며, 기존의 이상감지 기법들과는 다른 차별화된 방법론을 제시할 수 있을 것이다. 다차원 시각화 모델의 성능을 확인하기 위해 본 모델을 TE (Tennessee Eastman) 공정에 적용하여 보았다. 기존의 이상감지 기법인 PLS (Partial Least Squares), FDA (Fisher Discriminant Analysis), SVM (Support Vector Machine) 모델 등은 TE 공정의 이상을 감지하는데 효율적인 방법론이 될 수 있지만 몇몇 영역에 있어서 복수의 해를 찾아내는 한계가 있다. 본 연구의 목표는 PLS, FDA, SVM 등 기존의 이상감지 모델을 통해서 구분되지 않는 복수의 해집단을 세분화 하는 것이다. 이를 위해 TE 공정에 대하여 SVM 모델을 통해 분석하고 난 후, 몇가지 영역의 복수의 해집단에 대한 다차원 시각화를 통해 최종적으로 분류하는 계층적 방법론을 제시하고자 한다.

## 2. 이론

### SVM(Support Vector Machine)<sup>2)</sup>

SVM(Support Vector Machine)은 통계적 학습 이론에서 유도한 학습 바이어스를 이용한 학습알고리즘으로 훈련되는 고차원의 feature space 상의 선형식의 가상 공간을 이용한다. 이 학습 전략은 Vapnik에 의해 소개된 강력한 통계방법으로써 이미 폭 넓은 분야에서 다른 시스템보다 뛰어난 능력을 발휘하고 있다. 이 시스템은 'Sparse dual representation'을 만들어 내는데 이러한 특징은 시스템의 문제 해결 능력에 있어서 중요한 역할을 한다. 또 하나의 특징은 Mercer's condition으로 인하여 최적화 문제가 볼록한 기하학적인 형태로 산출되어 Local Minima를 만들지 않는다는데 있다. SVM을 통해 변수간 상관관계를 분석하여 공정상의 여러 모델을 만든 후, 이를 토대로 각 모델에서 나온 예측값과 실제 공정상의 데이터의 차이로 얻어지는 잔차(Residual)를 이용하면 공정상의 모델과 이상간의 상관관계를 구축할 수 있다.

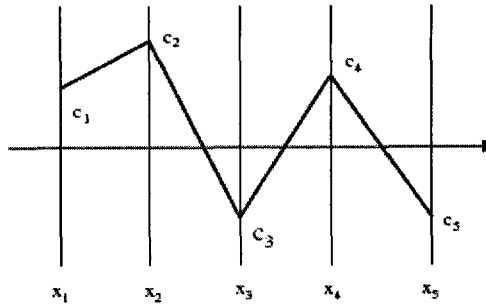
### 다차원 시각화(Multidimensional Visualization)<sup>3)</sup>

다차원 시각화란 다변량 통계분석 문제에서 다차원의 공간을 2차 평면 또는 3차원 공간으로 변환시켜, 어떠한 정보의 손실도 없이 많은 변수를 포함한 공정 데이터를 다루는 방법이다. 다차원 시각화를 통해 복잡하고 거대한 가스 또는 화학공정에서, 공정 변수와 데이터에 대한 더욱 강력한 통찰을 할 수 있게 된다.

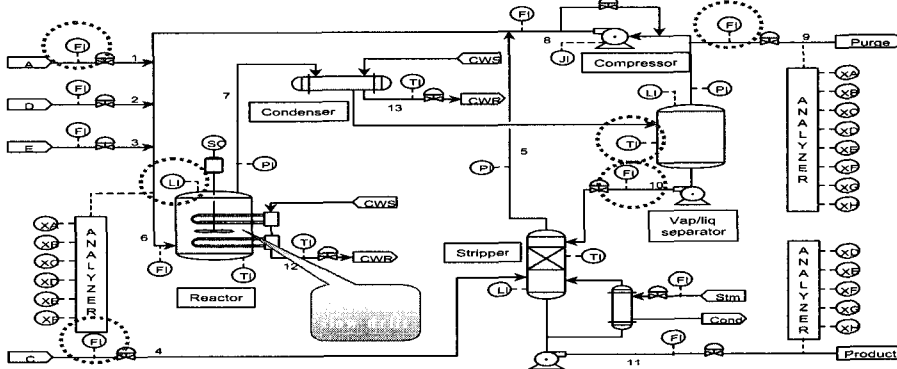
예를 들면, 5차원 데이터인  $C=(c_1, c_2, c_3, c_4, c_5)$ 를 생각해 보자. 다섯 개의 변수 정보를 포함하는  $C$ 는 [그림1]과 같이 평행좌표평면에 표시된다. 여기서,  $x_1, x_2, x_3, x_4, x_5$  축은 다섯 가지의 변수를 나타내고 데이터  $C$ 가 포함하는 다섯 개의 값  $c_1 \sim c_5$ 는 각각의 변수축에서 해당하는 점을 지나며 꺾은 선으로 나타나게 된다.

### 3. 예제

SVM을 통한 이상의 분석  
 본 연구의 성능을 평가하기 위하여 Tennessee Eastman 공정을 대상으로 이상진단을 수행하였다.



[그림152] 5차원 데이터 C의 표현



[그림2] Tennessee Eastman Process

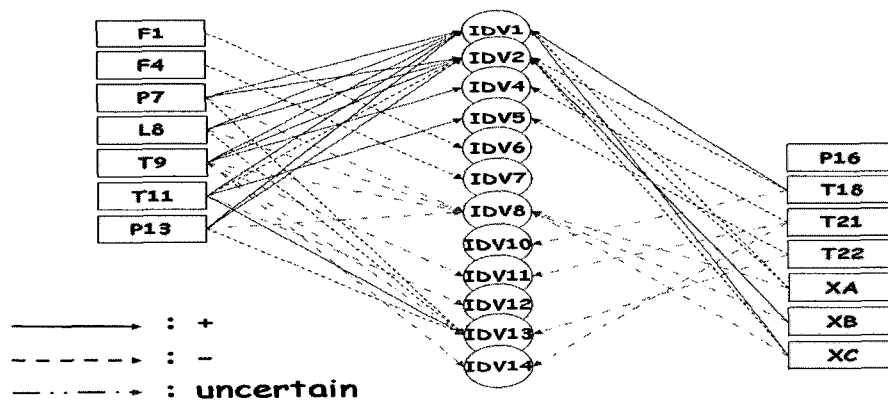
공정제어기법이나 최적화, 이상진단 등의 평가를 위하여 Downs와 Vogel(1995)에 의하여 제시된 것으로, 성분과 반응식 및 운전 조건들을 수정하였으나 각종 제어 기술을 실제 공정에 대하여 적용한 것과 동일한 효과를 나타내도록 Eastman사의 TE 공정을 실제로 모사한 것이다. McAvoy와 Ricker 등은 본 공정을 대상으로 제어와 최적화 등에 대한 연구를 수행하였다.

TE 공정은 반응기, 스트리퍼(stripper), 기액 분리기(vapor/liquid separator), 압축기 및 환류가 있는 공정으로 41개의 측정변수와 12개의 조작변수가 있다. TE에서 시뮬레이션이 가능한 15개의 이상을 가지고 진단을 수행하였다. [표1]은 TE 공정에서 발생 가능한 이상 15가지를 보여주고 있다. [그림3]은 이상 발생 시 TE 공정내부에서 만들 수 있는 Local Model의 잔차와 이상과의 관계를 나타낸다.

Fault ID	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) and IDV(10)	D and C, feed temperature (stream 2 and 4)	random variation
IDV(11)andIDV(12)	Reactor and condenser cooling water inlet temperature	random variation
IDV(13)	Reaction kinetics	slow drift
IDV(14)andIDV(15)	Reactor and condenser cooling water valve	sticking

[표56] TE 공정에서의 이상의 종류와 유형

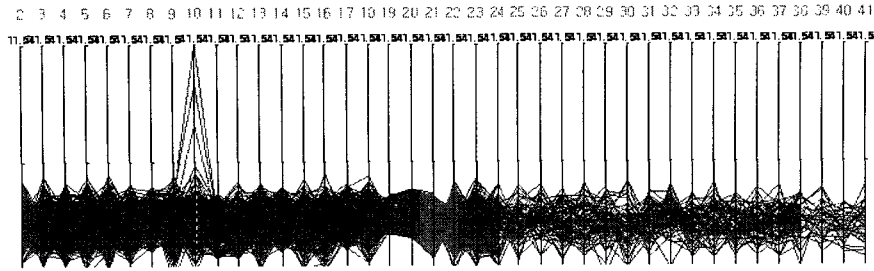
SVM 모델을 통해 TE 공정을 분석한 결과 15가지의 이상을 각각의 해집단으로 분류할 수 있다. 이 가운데 복수의 해를 가지는 해집단으로서, IDV(1), IDV(2), IDV(8)과 IDV(4), IDV(11), IDV(14)이 있다. 이 중 IDV(4), IDV(11), IDV(14)를 포함하는 해집단에 대해 알아보자.



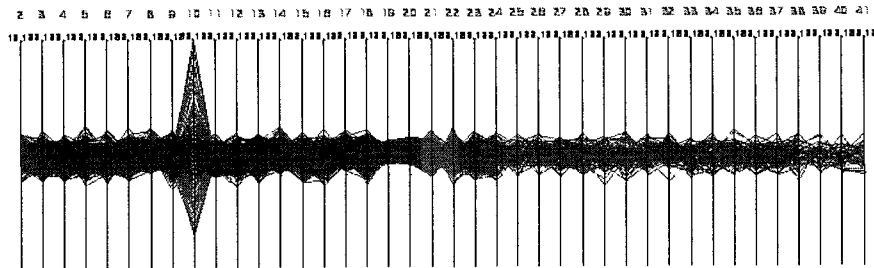
[그림154] SVM을 이용한 TE공정에서의 이상과 잔차와의 관계

IDV(4), IDV(11), IDV(14)의 도시

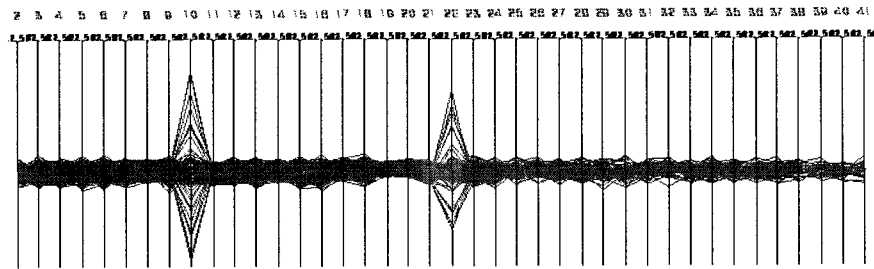
해집단 IDV(4), IDV(11), IDV(14)에 대한 경우는 다차원시각화를 통한 구분이 더욱 자명하다. [표1]에서 알 수 있듯, 세 가지 이상의 경우 반응기에서 cooling water의 온도와 유량에 관한 이상인데, 평행좌표의 변수 10번(T9, reactor temp.)과 22번(T21, reactor cooling water outlet temp.)에 밀접하게 관련되며 이는 아래 이상별 평행좌표의 도시를 통해 명확한 구분이 가능하다.



[그림155] IDV(4)



[그림156] IDV(11)



[그림157] IDV(14)

#### 4. 결론

본 연구에서는 보다 빠르고 정확한 가스화학공장 이상진단을 위해 기존의 SVM 모델을 적용하여 일차적으로 해를 구하였다. 그 다음, 구분이 안되는 복수의 해를 가진 해집단에 대해 다차원시각화를 이용해 최종적인 해를 판별할 수 있다는 방법론을 제시하였다. 이러한 방법론을 TE공정에 적용시켜 본 결과 우수한 성능을 지님을 확인할 수 있었다. 결론적으로 본 연구에서 제시한 방법론을 기반으로 임의의 공정에 대한 이상데이터를 분석하는 작업을 통해서, 공정이상에 대한 명확한 감지가 가능할 것이다.

## 5. 참고문헌

- 1) A. Inselberg, Visual data mining with parallel coordinates, Comput. Stat. 13(1998)
- 2) Chiang, L. H., Kotanchek, M. E., and Kordon, A. K. ; Fault Diagnosis Based on Fisher Discriminant Analysis and Support Vector Machines, Computers Chem. Eng., 28, 1389-1401 (2004)
- 3) Xue Z. Wang, Multidimensional visualisation for process historical data analysis : a comparative study with multivariate statistical process control, (2005)
- 4) G. Lee, C-H. Han, and E. S. Yoon; Mutiple-Fault Diagnosis of the Tennessee Eastman process Based on System Decomposition and Dynamics PLS, Ind. Eng. Chem., Res. 43, 8037-8048 (2004)