

부분최소제곱법 모델의 파라미터 추정을 이용한 화학공정의 이상진단 모델 개발

이광오* · 이창준**†

The Development of a Fault Diagnosis Model based on the Parameter Estimations of Partial Least Square Models

Kwang Oh Lee* · Chang Jun Lee**†

†Corresponding Author

Chang Jun Lee

Tel : +82-51-629-6465

E-mail : changjunlee@pknu.ac.kr

Received : May 17, 2019

Revised : July 2, 2019

Accepted : August 2, 2019

Abstract : Since it is really hard to construct process models based on prior process knowledges, various statistical approaches have been employed to build fault diagnosis models. However, the crucial drawback of these approaches is that the solutions may vary according to the fault magnitude, even if the same fault occurs. In this study, the parameter monitoring approach is suggested. When a fault occurs in a chemical process, this leads to trigger the change of a process model and the monitoring parameters of process models is able to provide the efficient fault diagnosis model. A few important variables are selected and their predictive models are constructed by partial least square (PLS) method. The Euclidean norms of parameters of PLS models are estimated and a fault diagnosis can be performed as comparing with parameters of PLS models based on normal operational conditions. To improve the monitoring performance, cumulative summation (CUSUM) control chart is employed and the changes of model parameters are recorded to identify the type of an unknown fault. To verify the efficacy of the proposed model, Tennessee Eastman (TE) process is tested and this model can be easily applied to other complex processes.

Copyright©2019 by The Korean Society
of Safety All right reserved.

Key Words : fault diagnosis, partial least square method, parameter estimation, tennessee eastman process

1. 서론

화학 공정에서 원하고자 하는 최종 제품을 얻기 위하여 공정의 환경적 영향, 생산량, 안전, 제품의 질 등 다양한 부분을 고려해야 한다. 그중에서, 화학 공정이 연속 공정인 특성을 고려할 때, 안전과 관련된 문제로 공정의 운전이 문제가 생기는 경우 막대한 피해가 발생하기 때문에 공정 내의 안전성을 확보하는 부분은 매우 중요하며, 공정에서 발생하는 이상의 원인을 조기에 감지하여, 이상이 전파되기 전에 해결하는 것이 매우 중요하다¹⁾.

공정에서 발생한 이상을 센서로부터 얻어지는 데이터를 바탕으로 감지하여 그 원인을 찾는 것을 이상 진

단이라 한다. 이상진단 방법론은 크게 두 가지로 나눌 수 있다. 첫 번째는 공정과 관련된 모델을 설계하고 이를 바탕으로 공정의 이상을 진단하는 모델기반 이상진단이다. 모델기반 알고리즘은 에너지, 물질 수지 식을 이용하여 모델링하고, 실제 센서로부터 얻어지는 데이터와 비교하여 이상의 원인을 진단하는 방법이다²⁾. 상당히 정확한 방법으로써, 정확하게 공정을 표현할 수 있는 모델을 얻는 경우, 공정에서 발생하는 이상을 정확하게 진단할 수 있다는 장점을 갖고 있다. 하지만, 이 방법론은 공정 모델링에 매우 큰 노력과 시간이 필요하며, 공정이 복잡해질수록 공정 내에서 발생하는 현상을 수지 식으로 표현하는데 많은 한계점을 가지고 있다는 단점을 갖고 있다.

*부경대학교 안전공학과 박사과정 (Department of Safety Engineering, Pukyong National University)

**부경대학교 안전공학과 교수 (Department of Safety Engineering, Pukyong National University)

두 번째는 센서로부터 얻어지는 데이터를 바탕으로 통계기법을 적용하여 이상의 원인을 진단하는 방법이다. 공정이 정상인 경우와 이상이 발생하는 경우의 데이터를 이용하여 미리 고유의 패턴을 추출하거나 공정 변수를 예측하는 모델을 만든다. 패턴을 추출하는 방법의 경우 실시간으로 얻어지는 데이터의 패턴을 분석하여 이상진단을 수행한다. 공정 변수를 예측 모델을 이용하는 경우, 센서로 측정하는 실제 데이터와 예측 모델로 계산된 변수의 값의 차이를 모니터링하며, 이상이 발생하는 경우 이 차이에 중요한 변화가 생기기 때문에, 이러한 특성을 이용하여 공정 내의 이상을 진단한다.

대표적인 방법론으로는 주성분분석(Principal Component Analysis), 독립성분분석(Independent Component Analysis), 부분최소제곱법(PLS (Partial Least Square)), Artificial Immune System, Hidden Markov Model, Neural Network 등이 이 방법론에서 사용되는 대표적인 통계 기법이다⁶⁾.

하지만, 데이터에 기반을 둔 방법론들의 큰 단점은 이상이 발생한 경우 데이터의 패턴은 같은 종류의 이상이 발생하는 경우라도 세기가 달라짐에 따라 그 패턴이 바뀌게 되어 오 진단을 할 가능성이 매우 커지게 된다⁴⁾. 정확하게 이상을 진단하기 위해서는 같은 종류의 이상이라도, 그 이상의 세기가 다양한 이상데이터를 수집해야 하지만 현실적으로 이러한 데이터를 얻는 것은 거의 불가능하다⁷⁾.

근본적으로 이러한 문제를 해결하기 위해서 본 연구에서는 정상상태 데이터와 부분최소제곱법을 이용하여 중요변수를 예측하는 모델을 만들고, 각 모델의 파라미터를 구하게 된다. 그리고 센서로부터 측정하는 실시간 데이터를 이용해 모델의 파라미터를 모니터링하여 이상을 진단하는 알고리즘을 개발하고자 한다. 공정에서 발생하는 이상은 실제로 공정의 모델 자체를 변형하기 때문에 모델의 파라미터 변화를 모니터링하면 신속하고 정확하게 이상을 진단할 수 있다.

2장과 3장에서는 본 연구에서 사용한 공정분할과 부분최소제곱법에 대해 소개하고, 4장과 5장에서는 이를 이용한 이상진단 방법을 제안하고 사례연구를 통해 본 연구에서 제안한 모델을 검증하였다.

2. 공정 분할 (Process Decomposition)

공정 내에 서로 상관관계가 있는 많은 변수가 존재할 때, 공정에서 이상이 발생한 경우 어디에서 발생했는지 진단하기는 매우 어렵다. 또한, 공정 내에 존재하는 모든 변수를 측정하기 위해 센서를 설치하는 것은

비용적이나 기술적으로 어려운 일이다. 따라서, 센서의 측정치를 이용한 이상진단 모델을 개발하는 것은 매우 어려운 일이다. 이를 해결하기 위해 G. Lee 등⁸⁾은 공정 분할을 통해 모든 변수를 관찰할 필요 없는 효과적인 이상진단 모델을 개발한 바 있다.

본 연구에서도 이러한 공정 분할을 이용해 이상진단 모델을 개발하고자 한다. 예상 가능한 이상의 원인과 직접 연결된 특정 변수를 사전에 선정하고 이 특정 변수에 영향을 주는 변수들을 입력값으로 설정하여 특정 변수를 예측하는 모델을 만들고자 한다. 이상이 발생하는 경우 이러한 모델에 변형이 발생하며, 이는 모델과 관련된 파라미터에도 변화가 생기게 된다. 따라서, 모델과 관련된 파라미터를 모니터링하면 이상의 감지와 진단이 가능해진다.

예를 들어, Fig. 1에서 보는 바와 같이 간단한 두 물탱크가 존재하는 경우 첫 번째와 두 번째 물탱크의 수위와 출구량은 Fig. 2에서 정의된 변수의 관계를 이용하여 다음의 식으로 정리할 수 있다.

$$\begin{aligned} L_1(t+1) &= h_1(F_0(t)) \\ F_1(t+1) &= h_2(L_1(t)) \\ L_2(t+1) &= h_3(F_1(t)) \\ F_2(t+1) &= h_4(L_2(t)) \end{aligned} \quad (1)$$

위 식에서 출구로 나가는 물의 양은 입구에 물이 들어오는 양 F_0 와 물탱크의 수위 L_1 에 의해 결정된다. 만약에 이 탱크에 누출이 발생하여 탱크 측면으로 물이 새는 경우, 식 (1)로 예측하는 탱크의 출구량은 달라진다.

만약에 첫 번째 탱크 측면에서 누출이 발생하는 경우, 탱크로 들어오는 물의 양과 수위로 출구량을 예측하는 모델을 식 (2)와 같이 만들면 식 (1)에서 구한 함수 h_1 , h_2 와 h_1^* , h_2^* 사이에는 차이가 발생하게 된다. 누출

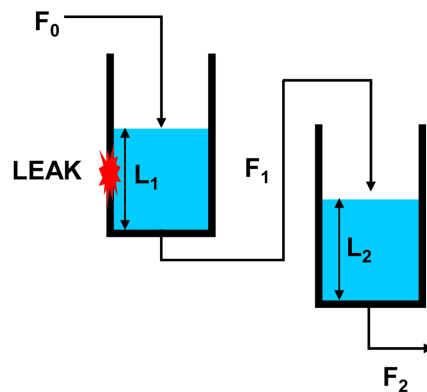


Fig. 1. Two stage tank systems.

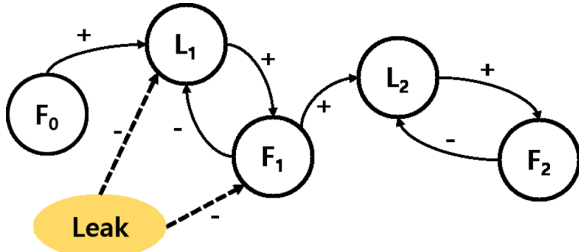


Fig. 2. The relationships between process variables of Fig. 1.

때문에 두 번째 탱크의 수위와 출구량은 급격하게 변하지만, 두 번째 탱크의 함수 자체에는 문제가 발생하지 않기 때문에 모델을 비교하면, 첫 번째 모델에서 변화가 생겼음을 감지할 수 있고, 이를 통해 첫 번째 탱크와 관련된 두 모델에서 변화가 발생하기 때문에 첫 번째 탱크에서 이상이 발생하였음을 진단할 수 있다.

$$\begin{aligned} L_1(t+1) &= h_1^*(F_0(t)) \\ F_1(t+1) &= h_2^*(L_1(t)) \\ L_2(t+1) &= h_3(F_1(t)) \\ F_2(t+1) &= h_4(L_2(t)) \end{aligned} \quad (2)$$

본 연구에서는 이상이 발생하지 않은 정상 데이터를 이용해 부분최소제곱법을 이용해 모델을 구축하고 모델의 파라미터를 계산한다. 실제 데이터가 들어오면 바로 모델의 파라미터를 계산하여 정상 데이터를 이용해 만든 모델의 파라미터와 비교하여 공정을 모니터링하고자 한다.

3. 배경이론

3.1 부분최소제곱법 (Partial Least Square)

부분최소제곱법은 다변량 통계 분석 기법의 하나로 입력값과 출력값에 대해서 각 값에 존재하는 변수들의 선형결합으로 표시되는 새로운 주성분(Principal Component (PC))을 탐색하여 새로운 Score 벡터를 만들고 입력값과 출력값의 Score 벡터의 관계를 선형으로 탐색하는 기법이다⁸⁾.

부분최소제곱법은 주성분분석을 이용하여 데이터들의 선형 조합을 통해 주성분으로 불리는 새로운 Score 벡터를 생성한다. 이때, Score 벡터는 변수 간의 분산(variance)이 최대가 되도록 설정이 된다⁸⁾. 이를 통해 원래 데이터를 정보를 최대한 많이 가지고 있으면서도 차수가 적은 데이터로 변환할 수 있다⁸⁾. 이를 통하여 복잡하고 방대한 데이터 간의 회귀모델을 만들 수 있다¹⁾.

데이터의 입력값 $X = [X_1, X_2, \dots, X_p]$ 과 출력값

$Y = [Y_1, Y_2, \dots, Y_a]$ 에 대해 부분최소제곱법은 다음과 같은 식으로 표현된다.

$$\begin{aligned} X &= TP^T + e_x \\ Y &= UQ^T + e_f \end{aligned} \quad (3)$$

P 와 Q 는 X 와 Y 의 Loading 행렬이며, T 와 U 는 X 와 Y 의 Score 벡터이다. e_x 와 e_y 는 각각 데이터가 Loading 행렬에 의해 새로운 PC로 선형으로 프로젝션 (projection)되는 과정에서 발생하는 에러이다. 부분최소제곱법에서는 Score 벡터의 관계를 이용해 회귀분석을 수행한다.

$$\begin{aligned} u_k &= b_s t_s + l \\ k &= 1, \dots, \alpha, s = 1, \dots, \beta \end{aligned} \quad (4)$$

위 식에서 b_s 는 U 의 k 번째 행렬을 예측하는데 필요한 상관계수이며, l 은 에러항이다. 식 (3)에서 X 와 Y 가 원래 차수보다 더 작은 α 와 β 로 축약되고 NIPALS (nonlinear iterative partial least squares)에 의해 b_s 를 탐색한다. 회귀분석 모델을 부분최소제곱법에 의해 만드는 경우 b_s 가 모델파라미터이며 본 연구에서는 b_s 의 변화를 식 (5)에 나오는 Euclidean Norm에 의해 모니터링하고자 한다.

$$\|b_s\|^2 = \sqrt{b_{s,1}^2 + b_{s,2}^2 + \dots + b_{s,\alpha}^2} \quad (5)$$

3.2 누적합관리도 (Cumulative Sum Control Chart)

본 연구에서는 부분최소제곱법을 이용하여 공정상의 중요한 변수를 예측하는 모델을 만들고 그 파라미터를 지속적으로 모니터링해야 한다. 이를 위해 누적합관리도를 사용하여 모니터링의 효율성을 높이고자 하였다.

Shewhart chart와 같은 일반적인 관리도가 데이터에 큰 변화가 발생하는 경우 그 이상을 빨리 감지할 수 있지만, 작은 변화가 발생할 때는 그 감지속도가 늦어진다. 누적합관리도를 사용하는 경우 데이터에 작은 변화가 발생하더라도 다른 관리도에 비해 더욱 빨리 이상의 발생을 감지할 수 있다는 장점이 있다⁹⁾.

누적합관리도는 현재와 과거의 변수 변화를 동시에 사용하여 모니터링의 효율을 증가시키는 방법으로써, 최소변동값(minimal jump size, v_m)과 데이터를 비교하여 데이터의 값이 최소변동값 보다 큰 경우 그 차이만큼 계속 누적한다⁹⁾. 이 누적값이 관리통계량(Control Threshold,

T^*)을 넘어서게 되면 공정에 이상을 탐지하게 된다. 누적합관리도는 다음의 식으로 표현할 수 있다.

$$\begin{aligned}
 U(0) &= 0 \\
 L(t) &= \max[0, (r(t) - u) - v_m] \\
 S(t) &= \min[0, (r(t) - u) + v_m] \\
 U_+(t) &= U(t-1) + L(t) \\
 U_-(t) &= U(t-1) + S(t) \\
 \text{If } U_+(t) &> T^*, (+) \\
 \text{If } U_-(t) &< T^*, (-) \\
 \text{Otherwise, } &0
 \end{aligned}
 \tag{6}$$

식 (6)에서 $r(t)$ 는 모니터링을 하는 실제 변수의 값이며, u 는 변수의 평균값이다. 데이터가 들어올 때 마다 평균값의 차이가 최소변동값 ($\pm v_m$)을 넘는지를 $L(t)$ 와 $S(t)$ 를 이용해 측정한다. 누적값 ($U(t)$)가 관리통계량인 T^* 를 넘어서게 되면 이상이 있다 (+/-)고 가정하며, 그렇지 않은 경우에는 이상이 없다 (0)고 가정한다.

본 연구에서는 정상 데이터를 이용하여 미리 모델파라미터의 평균값과 최소변동값, 관리통계량을 정의하고, 실제 데이터를 이용하여 모델의 파라미터를 누적합관리도에 적용하여 모니터링하여 공정 내의 이상을 진단하였다.

4. 사례연구

4.1 TE(Tennessee Eastman) 공정

공정제어, 이상진단, 최적화 연구를 위해 Downs와 Vogel⁷⁾에 의하여 개발된 벤치마킹 시뮬레이터로 Eastman 사의 공정을 모사하였으며, 포트란이나 매트랩에서 구현이 가능하다. 이를 이용하여 그동안 수많은 이상진단 연구가 수행되어 왔다¹⁰⁻¹²⁾. 또한, TE 공정을 이용하여 다양한 통계적 기법에 관한 비교 연구가 S. Yin 등¹²⁾에 의해서 수행된 바가 있다.

Fig. 3에서 TE 공정을 보여주고 있다. 반응기, 분리기, 스트리퍼, 압축기가 존재하며, 총 22개의 측정변수, 19개의 조성변수, 11개의 조작변수가 있으며 4가지 반응물(A, B, C, D)을 이용하여 2가지 생성물(G, H)이 생성된다. 반응기에서는 촉매에 의해 다음의 4가지 비가역 발열반응이 일어난다¹⁰⁻¹¹⁾. TE 공정과 관련된 자세한 내용은 Downs와 Vogel¹⁰⁾의 연구를 참고하면 된다.

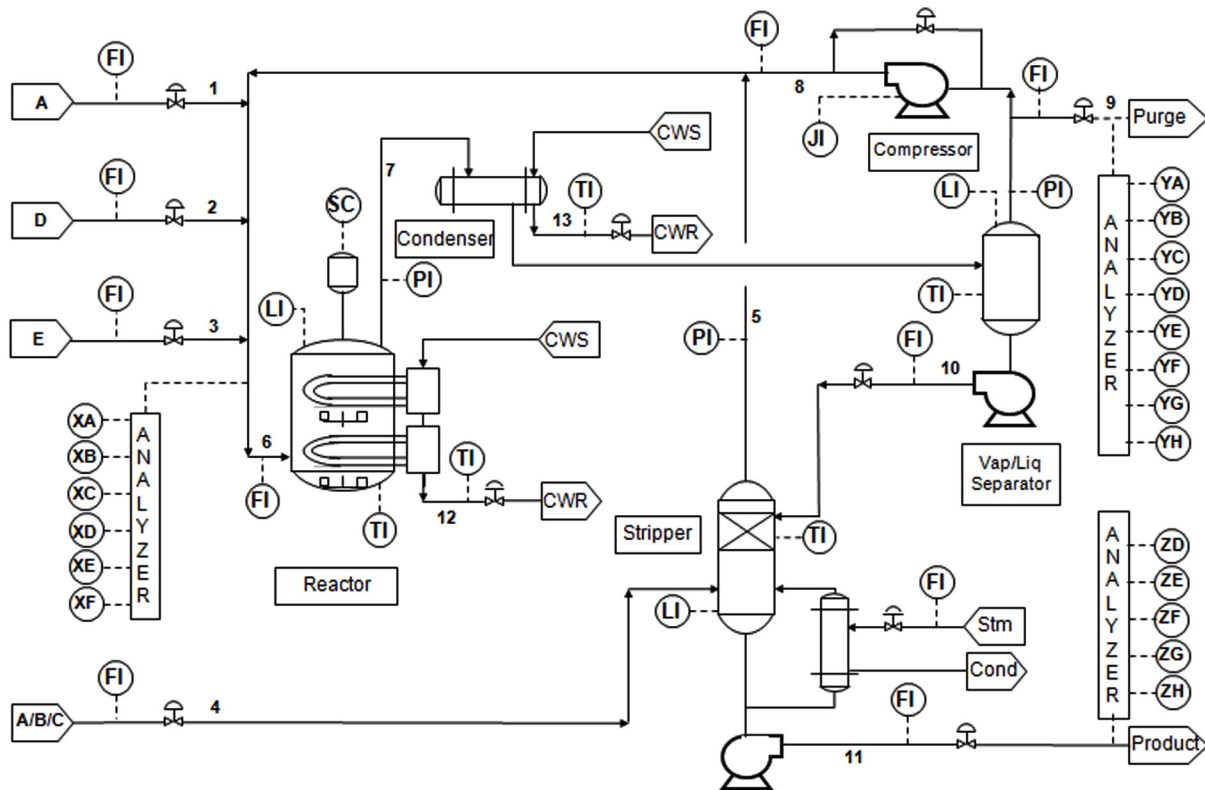
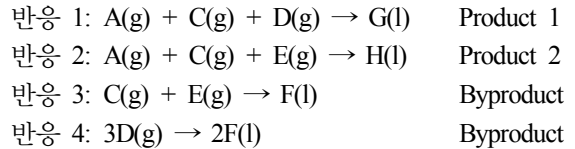


Fig. 3. Tennessee eastman process⁷⁾.

본 연구에서는 포트란을 이용하여 생성한 총 16개의 데이터 Set을 이용하였다. 22개의 측정변수는 1분 간격으로 측정된 것이며, 19개의 성분변수는 6분 혹은 9분에 한 번씩 측정된 것이다. 16개 데이터 Set 중 한 개는 이상이 전혀 발생하지 않은 정상상태 데이터이며, 나머지 15개의 Set에서는 공정 시작 후 30분이 지나서 이상이 발생한 경우이다.

4.2 부분최소제곱법을 이용한 모델 수립

TE 공정은 매우 복잡하고 변수도 많아서 Fig. 2에서 정리한 공정 간의 관계를 그림으로 간단하게 정리하기는 불가능하다. 따라서, G. Lee 등⁸⁾의 연구를 참고하여 15개 이상과 먼밀한 변수를 선정하고, 이들 변수에 직접적인 영향을 주는 입력값을 정리하여 부분최소제곱법을 이용하여 모델을 수립하였다.

TE 공정에서 변수는 Table 1에 정리하였으며, Table 2는 본 연구에서 가정한 15가지 이상을 정리하였다.

Table 3은 본 연구에서 이상진단을 위해 정리한 모델의 입력값과 출력값을 보여주고 있으며, 이러한 관계는 G. Lee 등⁸⁾의 연구를 참조하였다.

정상상태 데이터를 이용해 부분최소제곱법으로 모델을 만드는 경우 계산되는 상관계수(b_s)의 Euclidean Norm을 계산하고 이 값의 평균값(μ)과 표준편차(σ)를 이용하여 각 모델의 누적합관리도를 만들 수 있다. 실제 이상이 있는 데이터가 들어오는 경우 누적합관리도를 통해 문제가 발생한 부분을 감지하고 그 원인을 진단하게 된다. 본 연구에서는 최소변동값(minimal jump size, v_m)은 평균값을 중심으로 $\pm 2\sigma$, 관리통계량은 $\pm 6\sigma$ 로 설정하였다.

5. 진단 결과

Table 2에서 정리한 15가지 이상과 부분최소제곱법을 이용한 19가지 모델 간의 관계를 Table 4에 정리하였다. 이러한 관계를 바탕으로 TE 공정에서 15가지 이

Table 1. Manipulated and Measured Variables of the TE Process¹⁰⁾

Variable	Description	Variable	Description
MV1	D feed flow (stream 2)	P16	Stripper pressure
MV2	E feed flow (stream 3)	F17	Stripper underflow (stream 11)
MV3	A feed flow (stream 1)	T18	Stripper temperature
MV4	Total feed flow (stream 4)	F19	Stripper steam flow
MV5	Compressor recycle value	J20	Compressor work
MV6	Purge valve (stream 9)	T21	Reactor cooling water outlet temp
MV7	Separator pot liquid flow (stream 10)	T22	Condenser cooling water outlet temp
MV8	Stripper liquid product flow	XA	Composition of A (stream6)
MV9	Stripper steam valve	XB	Composition of B (stream6)
MV10	Reactor cooling water flow	XC	Composition of C (stream6)
MV11	Condenser cooling water flow	XD	Composition of D (stream6)
MV12	Agitator speed	XE	Composition of E (stream6)
F1	A feed (stream 1)	XF	Composition of F (stream6)
F2	D feed (stream 2)	YA	Composition of A (stream6)
F3	E feed (stream 3)	YB	Composition of B (stream6)
F4	Total feed (stream 4)	YC	Composition of C (stream6)
F5	Recycle flow (stream 8)	YD	Composition of D (stream6)
F6	Reactor feed rate (stream 6)	YE	Composition of E (stream6)
P7	Reactor pressure	YF	Composition of F (stream6)
L8	Reactor level	YG	Composition of G (stream6)
T9	Reactor temp	YH	Composition of H (stream6)
F10	Purge rate (stream 9)	ZD	Composition of D (stream6)
T11	Separator temp	ZE	Composition of E (stream6)
L12	Separator level	ZF	Composition of F (stream6)
P13	Separator pressure	ZG	Composition of G (stream6)
F14	Separator underflow (stream 10)	ZH	Composition of H (stream6)
L15	Stripper level		

Table 2. The descriptions of fault types⁷⁻⁸⁾

Fault	Description
IDV1	A/C Feed ratio, B Composition constant (Stream 4)
IDV2	B Composition, A/C Ratio constant (Stream 4)
IDV3	D Feed temperature (Stream 2)
IDV4	Reactor cooling water inlet temperature
IDV5	Condenser cooling water inlet temperature
IDV6	A feed loss (Stream 1)
IDV7	C Header pressure loss-reduced availability (Stream 4)
IDV8	A, B, C Feed composition (Stream 4)
IDV9	D Feed temperature (Stream 2)
IDV10	C Feed temperature (Stream 4)
IDV11	Reactor cooling water inlet temperature
IDV12	Condenser cooling water inlet temperature
IDV13	Reaction kinetics
IDV14	Reactor cooling water valve
IDV15	Condenser cooling water valve

상 중 하나가 발생하면 어떤 이상이 발생하였는지 진단할 수 있다.

Fig. 4는 첫 번째 이상인 F1이 일어나는 경우 4개의 부분최소제곱법을 이용해 만든 모델(P7, L8, T9, T11) 파라미터를 측정하여 누적합관리도를 이용하여 모니터링한 결과이다. 본 연구에서 제안한 모델을 이용하면, 15개의 이상 중에서 9개의 이상(IDV3, IDV4, IDV5, IDV8, IDV9, IDV10, IDV11, IDV12, IDV15)을 정확하게 진단할 수 있다. 나머지 6개의 이상은 정확한 진단

은 어렵지만, 두 개의 그룹으로 분리가 가능하다. 첫 번째 그룹은 IDV1, IDV2, IDV6, IDV7이며 두 번째 그룹은 IDV3, IDV14이다. 이 두 개의 그룹은 정확하게 부분최소제곱법의 파라미터의 변화가 일치하며, 이 이상들을 정확하게 분리하기 위해서는 다른 알고리즘을 이용해야 한다. TE를 이용한 기존의 이상진단 관련 연구에서도 본 연구의 결과와 동일하게 증상이 일치하는 그룹이 존재하며, 이를 해결하기 위해서 추가적인 통계적 기법을 이용하였다⁷⁻⁸⁾.

Table 3. Target variables and their source variables⁷⁻⁸⁾

Target variable	Source variables connected to the target variable
P7	F1, F2, F3, F4, F5, L8, T9, P13, F17, YA, YB, YC, YD, YE, YF, YG, YH
L8	F1, F2, F3, F4, F5, P7, T9, P13, F14, F17, YA, YC, YD, YE, YF, YG, YH
T9	F1, F2, F3, F4, F5, P7, L8, T11, P13, F17, T18, YA, YB, YC, YD, YE, YF, YG, YH, MV10
T11	F1, F2, F3, F5, P7, L8, T9, F10, L12, P13, F14, F17, T18, YA, YB, YC, YD, YE, YF, YG, YH, MV11
P13	F1, F2, F3, F4, F5, P7, L8, T9, F10, T11, L12, F14, F17, YA, YB, YC, YD, YE, YF, YG
T18	F4, T11, F14, L15, F17, F19, ZD, ZE, ZF, ZG, ZH
T21	L8, T9, MV10
T22	T9, MV11, P7, P13
XA	F1, F2, F3, F4, F5, F17, YA
XB	F1, F2, F3, F4, F5, F17, YB
XC	F1, F2, F3, F4, F5, F17, YC
YA	F5, F6, P7, L8, T9, F10, T11, L12, XA, XC, XD, XE
YC	F5, F6, P7, L8, T9, F10, T11, L12, XA, XC, XD, XE
YD	F5, F6, P7, L8, T9, F10, T11, L12, F14, XA, XC, XD
YE	F5, F6, P7, L8, T9, F10, T11, L12, F14, XA, XC, XE
YF	F5, F6, P7, L8, T9, F10, T11, L12, F14, XA, XC, XE, XF
YG	F5, F6, P7, L8, T9, F10, T11, L12, F14, XA, XC, XD
YH	F5, F6, P7, L8, T9, F10, T11, L12, F14, XA, XC, XD

본 연구에서 제시한 이상진단 모델의 검증을 위해 4가지의 이상의 원인을 모르고 그 이상의 세기도 모델을 만들 때와 다른 임의의 데이터를 생성하여 테스트

를 수행하였다. Table 5는 임의의 데이터를 진단한 결과이다. Time delay는 이상이 발생한 후 진단하기까지 걸리는 시간을 의미한다.

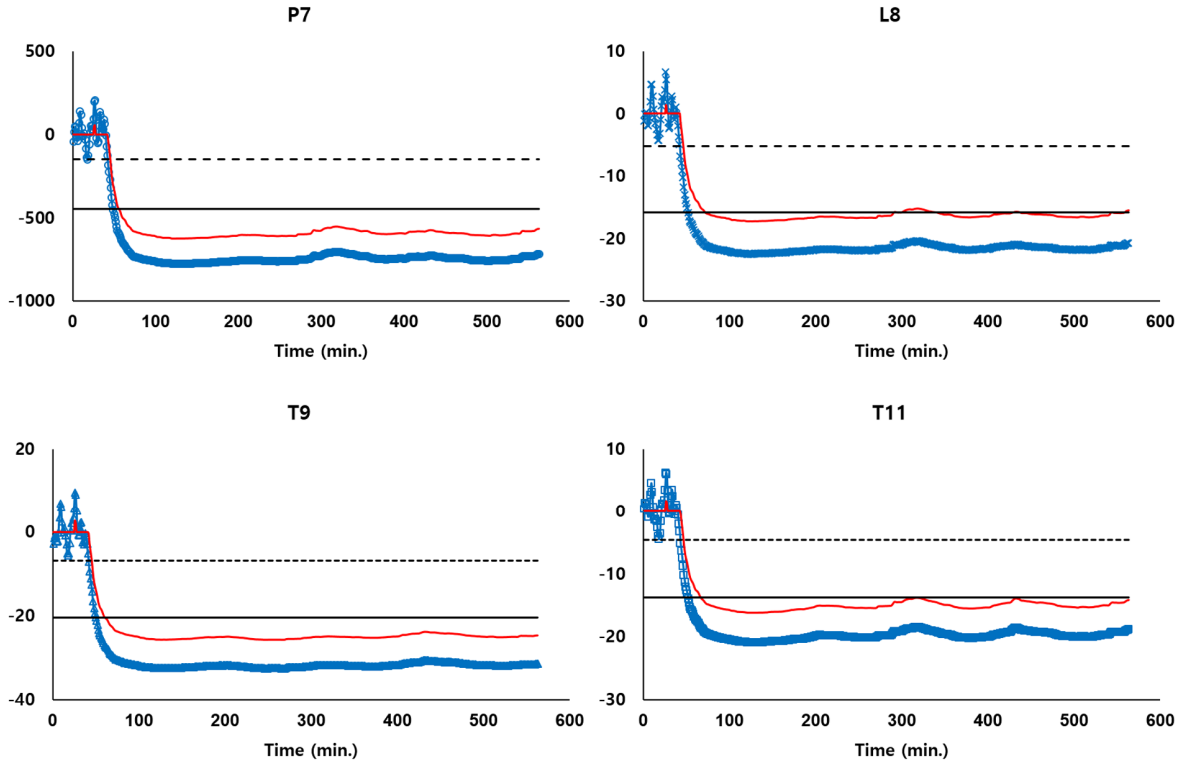


Fig. 4. Monitoring parameters of PLS models for predicting P7, L8, T9 and T11. The blue and red lines are Euclidean norms and their CUSUM charts. The black solid lines are the threshold and the dash ones are the minimum jump sizes.

Table 4. The relationships between assumed faults and PLS models

	IDV1	IDV2	IDV3	IDV4	IDV5	IDV6	IDV7	IDV8	IDV9	IDV10	IDV11	IDV12	IDV13	IDV14	IDV15
F1	0	0	+	+	+	0	0	0	+	0	+	+	0	0	0
P7	-	-	0	0	+	-	-	-	0	0	0	-	-	-	0
L8	-	-	0	0	0	-	-	-	0	0	0	-	-	-	0
T9	-	-	0	0	-	-	-	-	0	0	-	-	-	-	0
T11	-	-	0	0	-	-	-	-	0	0	0	-	-	-	0
P13	-	-	0	0	-	-	-	-	0	0	0	-	-	-	0
T18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
T21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
T22	0	0	+	+	0	0	0	0	0	+	0	0	0	0	0
XA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
XB	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
XC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
YA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
YC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
YD	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
YE	-	-	+	0	0	-	-	0	+	+	0	0	+	+	+
YF	-	-	+	+	-	-	-	-	+	+	-	+	+	+	+
YG	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
YH	-	-	+	0	-	-	-	-	+	+	-	+	+	+	+

Table 5. The results of unknown data

The fault type of unknown data	Candidate faults	Time delay
IDV1	IDV1, IDV2, IDV6, IDV7	22
IDV5	IDV5	18
IDV12	IDV12	42
IDV6	IDV1, IDV2, IDV6, IDV7	32

6. 결론

본 연구에서는 화학 공정에서 발생할 수 있는 이상이 공정 내 모델에 변형을 가져온다는 점을 이용하여 부분최소제곱법의 모델파라미터를 추정하고 모니터링하여 이상의 원인을 효과적으로 진단하는 알고리즘을 제안하였다.

많은 변수가 존재하고 매우 복잡한 화학 공정의 특성상 내부에 문제가 발생하였을 때 센서값을 통해 이상의 원인을 파악하기는 매우 어렵다. 기존의 통계적 기법을 이용한 일반적인 정량적인 기법은 공정데이터의 패턴을 분석하여 정상상태와 이상 상태 패턴으로 분류한 후 실시간 데이터를 분석하여 어느 패턴과 일치하는지를 판단하는 것이 일반적인 방법이다. 그러나, 같은 이상이 발생하더라도 그 이상의 크기에 따라 이러한 패턴은 달라질 수 있어서 오 진단할 가능성이 커지게 된다.

이러한 문제를 해결하기 위해, 본 연구에서는 공정 내 주요 변수를 예측하는 부분최소제곱법 모델을 생성하고 모델 내의 파라미터를 모니터링하는 알고리즘을 제안하였다. 이상의 세기가 달라지더라도 이상의 종류가 같으면 공정 모델에 변화가 생기는 방향성은 일치하기 때문에, 본 연구에서 제시된 알고리즘으로 효과적인 모니터링이 가능하다.

본 연구의 경우 TE 공정 내에서 발생 가능한 15개의 데이터에 대해 진단을 수행하였으며, 이상의 원인을 모르는 4개의 데이터를 이용하여 본 연구에서 제시한 알고리즘을 검증하였다. 15개 이상 중에서 9개의 이상(IDV3, IDV4, IDV5, IDV8, IDV9, IDV10, IDV11, IDV12, IDV15)을 정확하게 진단하였으며, 나머지 6개의 이상(IDV1, IDV2, IDV6, IDV7, IDV13, IDV14)이 발생한 경우는 이상의 발생 여부는 감지할 수 있지만, 파라미터 변화가 일치하기 때문에 정확한 진단은 불가능하였다. 이 6개 이상의 경우, 정확한 진단은 어렵더라도 두 개의 그룹까지는 분류할 수 있는데 첫 번째 그룹은 IDV1, IDV2, IDV6, IDV7이며 두 번째 그룹은 IDV13, IDV14이다. 예를 들어, IDV1이 발생하는 경우,

제안한 모델은 IDV1, IDV2, IDV6, IDV7 중 한 개의 이상이 발생한 것으로 진단할 수 있다.

기존의 다른 정량적인 방법과는 다르게 모델 자체의 파라미터 변화만을 이용하여 공정 내에서 발생하는 이상 진단이 가능함을 확인할 수 있었으며, 기존에 제안된 다른 알고리즘과의 융합을 통해 정확한 이상 진단 모델의 개발이 가능할 것으로 기대된다.

References

- 1) V. Venkatasubramanian, R. Rengaswamy, K. Yin and S. N. Kavuri, "A Review of Process Fault Detection and Diagnosis: Part I : Quantitative Model-based Methods", *Computers & Chemical Engineering*, Vol. 27, No. 3, pp. 293-311, 2003.
- 2) V. Venkatasubramanian, R. Rengaswamy, K. Yin and S. N. Kavuri, "A Review of Process Fault Detection and Diagnosis: Part II: Qualitative Models and Search Strategies", *Computers & Chemical Engineering*, Vol. 27, No. 3, pp. 311-326, 2003.
- 3) V. Venkatasubramanian, R. Rengaswamy, K. Yin and S. N. Kavuri, "A Review of Process Fault Detection and Diagnosis: Part III: Process History Based Methods", *Computers and Chemical Engineering*, Vol. 27, No. 3, pp. 327-346, 2003.
- 4) T. Rato, M. Reis, E. Schmitt, M. Hubert, B. De Ketelaere, "A Systematic Comparison of PCA-based Statistical Process Monitoring Methods for High-dimensional, Time-dependent Processes", *AIChE Journal*, Vol. 62, pp. 1478-1493, 2016.
- 5) J. Fan and Y. Wang, "Fault Detection and Diagnosis of Non-linear Non-gaussian Dynamic Processes using Kernel Dynamic Independent Component Analysis", *Information Sciences*, Vol. 259, pp. 369-379, 2014.
- 6) Y. Shu and J. Zhao, "Fault Diagnosis of Chemical Processes using Artificial Immune System with Vaccine Transplant", *Industrial and Engineering Chemistry Research*, Vol. 55, pp. 3360-3371, 2016.
- 7) C. J. Lee, G. Lee and J. M. Lee, "A Fault Magnitude Based Strategy for Effective Fault Classification", *Chemical Engineering Research and Design*, Vol. 91, No. 3, pp. 530-541, 2012.
- 8) G. Lee, C. Han and E. S. Yoon, "Multiple-Fault Diagnosis of the Tennessee Eastman Process Based on System Decomposition and Dynamic PLS", *Industrial and*

- Engineering Chemistry Research, Vol. 43, No. 25, pp. 8037-8048, 2004.
- 9) D. C. Montgomery, "Introduction to Statistical Quality Control", John Wiley, New York, 2001.
 - 10) J. J. Downs and E. F. Vogel, "A Plant-Wide Industrial Process Control", Computers and Chemical Engineering, Vol. 17, No. 3, pp. 245-255, 1993.
 - 11) J. Y. Park and C. J. Lee, "Principal Component Analysis Based Method for Effective Fault Diagnosis", J. Korean Soc. Saf., Vol. 29, No. 4, pp. 73-77, 2014.
 - 12) S. Yin, S. X. Ding, A. Haghani, H. Hao and P. Zhang, "A Comparison Study of Basic Data-driven Fault Diagnosis and Process Monitoring Methods on the Bench-mark Tennessee Eastman Process", Journal of Process Control, Vol. 22, pp. 1567-1581, 2012.