

주성분 분석과 서포트 벡터 머신을 이용한 폴리스티렌 중합 반응기 이상 진단 모델 개발

정연수 · 이창준[†]

부경대학교 안전공학과
48513 부산시 남구 용소로 45
(2021년 10월 19일 접수, 2021년 12월 6일 수정본 접수, 2021년 12월 14일 채택)

The Development of a Fault Diagnosis Model Based on Principal Component Analysis and Support Vector Machine for a Polystyrene Reactor

Yeonsu Jeong and Chang Jun Lee[†]

Department of Safety Engineering, Pukyong National University, 45, Yongso-ro, Nam-gu, Busan, 48513, Korea
(Received 19 October 2021; Received in revised form 6 December 2021; Accepted 14 December 2021)

요 약

화학공정에서 의도되지 않게 발생하는 이상은 큰 사고를 유발할 수 있다. 이러한 문제를 해결하기 위해, 신속하게 이상의 원인을 감지하고 판별하는 이상 진단 모델이 필요하다. 하지만, 이상 진단을 연구하는 대부분 연구의 경우, 상용프로그램에서 공정 시뮬레이션을 이용하여 이상 데이터를 생성하고 이를 이용하여 연구한 방법론을 적용하고 있다. 이는 실제 공정상에서 이상을 포함하는 실제 데이터를 얻는 데 많은 제약이 있음을 의미한다. 본 연구에서는 실제 폴리스티렌 반응기에서 얻은 이상 데이터와 정상 데이터를 분석하여 적절한 이상 진단 모델을 설계하고자 하였다. 먼저, 정상 데이터를 분석하여 세 가지의 조업 모드가 존재함을 확인하였으며, 모드 판별을 위한 모델을 SVM (Support Vector Machine)을 이용하여 만들었다. 각 조업 모드 별로 PCA (Principal Component Analysis)를 이용하여 이상 진단 모델을 만들었으며, 실제 이상 데이터를 이용하여 계산한 결과 신속하게 이상을 진단할 수 있음을 확인하였다. 본 연구에서 제안한 모델을 통해, 실제 사고가 발생하는 경우 신속한 대처가 가능하며, 이는 잠재적인 손실의 감소에 기여할 수 있음을 의미한다.

Abstract – In chemical processes, unintended faults can make serious accidents. To tackle them, proper fault diagnosis models should be designed to identify the root cause of faults. To design a fault diagnosis model, a process and its data should be analyzed. However, most previous researches in the field of fault diagnosis just handle the data set of benchmark processes simulated on commercial programs. It indicates that it is really hard to get fresh data sets on real processes. In this study, real faulty conditions of an industrial polystyrene process are tested. In this process, a runaway reaction occurred and this caused a large loss since operators were late aware of the occurrence of this accident. To design a proper fault diagnosis model, we analyzed this process and a real accident data set. At first, a mode classification model based on support vector machine (SVM) was trained and principal component analysis (PCA) model for each mode was constructed under normal operation conditions. The results show that a proposed model can quickly diagnose the occurrence of a fault and they indicate that this model is able to reduce the potential loss.

Key words: Fault diagnosis, Principal component analysis, Support vector machine

1. 서 론

화학 공정은 다양한 장치들로 구성되어 있으며, 고온과 고압 하에서 운전되고 있다. 따라서, 공정에 문제가 발생하는 경우 그 원인을

신속하게 파악하기 어려우며, 이러한 문제가 공정 전반으로 확산하여 사고가 발생하는 경우 막대한 피해를 유발하게 된다. 공정에서 발생하는 물리적 고장, 교란, 작업자의 오류 등의 원인으로 공정의 운전 조건을 정상 상태에서 벗어나는 경우를 이상이라고 하며, 다양한 공정 지식을 이용하여 이상을 감지하는 것을 이상 감지라 하며, 이상의 원인을 파악하는 것을 이상 진단이라 한다[1-2]. 화학 공정의 경우, 공정 내에 다양한 순환 사이클과 제어 루프로 인해 이상을 감지하는 일은 매우 어렵다[1-2].

기존 이상 진단 방법론은 공정 내에 존재하는 다양한 에너지 수

[†]To whom correspondence should be addressed.

E-mail: changjunlee@pknu.ac.kr

‡이 논문은 광운대학교 고재욱 교수의 정년을 기념하여 투고되었음.

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

지, 물질 수지, 반응식 등을 이용하여 특정 변수를 예측하는 모델을 만드는 방법과 센서로부터 얻어지는 데이터를 이용하여 통계기법을 적용하여 공정의 상태를 판별하는 모델을 만드는 방법이 있다[1-5]. 이 중에서 통계기법을 이용한 방법론은 모델을 개발하는데 많은 시간과 노력을 절감할 수 있고, 공정 지식이 없고 쉽게 공정 데이터들의 특징을 추출하여 만들 수 있다는 장점이 있으며, 대표적으로 주성분분석(PCA(Principal Component Analysis)), 부분 최소제곱법(Partial Least Square), 인공신경망(Artificial Neural Network) 등의 기법이 이용되고 있다[5]. 기본적으로 통계기법을 이용한 방법론은 공정 내의 특정 변수를 예측하는 모델을 만들고 실제 센서로부터 얻는 데이터와의 잔차를 비교하여 이상을 감지한다. Park 과 Lee[2]은 주성분 분석을 이용하여 공정의 이상을 진단하는 모델을 개발하였으며, Lee와 Lee[1]는 부분 최소제곱법으로 공정 내의 주요 변수를 예측하는 모델을 센서로부터 들어오는 데이터를 이용하여 계속 업데이트하여 모델 내에 존재하는 파라미터의 변화를 이용한 이상 진단 모델을 개발하였다. 또한, Lee 등[6]은 SVM(Support Vector Machine)을 이용하여 이상 진단 모델을 연구하였다. 선행연구를 살펴보면, 실제 공정 내에서 발생하는 이상 데이터를 얻기 어려운 문제점으로 인하여 Tennessee Eastman 공정과 같이 상용프로그램에 기반한 가상 공정에서 의도적으로 이상 데이터를 생성하여, 이를 바탕으로 연구를 진행하고 있다.

또한, 시장 상황이나 공정 조건 등의 변화로 인하여 대상 공정의 조업 모드를 변경하는 경우 이상 진단 모델은 해당 모드에 특화된 통계적 모델을 이용해야 한다. 만약, 특정 조업 모드에 대해 학습된 통계적 모델을 다른 모드에 적용한다면, 계속해서 잘못된 진단을 수행할 수 있다. 따라서, 조업 모드를 판별하는 모델 구축도 매우 중요하다. Lee 등[7]은 Euclidean distance, Fisher's discriminant analysis, 주성분 분석을 이용하여 조업 모드 판별법을 연구한 바가 있다.

본 연구에서는 실제 공정에서 고장이 발생한 시점을 기준으로 전후 공정 데이터를 활용하여 이상 진단 모델을 만들고자 한다. 먼저 SVM(Support Vector Machine)을 이용하여 조업 모드를 판별하는 모델을 제안하고 각 모드에 맞는 이상 진단 모델을 주성분 분석을 이용하여 학습하고자 한다. 실제 사고가 발생했을 때 저장된 사고 데이터를 통해 본 연구에서 제안한 모델이 얼마나 신속하게 이상의 유무를 판별하는지 보이고자 한다.

2장과 3장에서는 본 연구에서 이용할 SVM과 PCA를 소개하고, 4장에서는 대상 공정에 대해 설명하고자 한다. 5장에서 대상 공정의 실제 데이터를 이용하여 SVM을 이용한 조업 기능을 판별하는 모델과 PCA를 이용한 이상 진단 모델에 대해 설명하고자 한다.

2. SVM(Support Vector Machine)

SVM(Support Vector Machine)은 Vapnik에 의해 고안된 기계학습 이론으로, 데이터가 비선형이며, 고차원일 경우에 적합한 기계학습 방법론이다[8]. 분류 문제(Classification Problem) 해결을 위해 개발된 방법론인 SVM은 커널함수를 이용하여 기존 데이터를 특징 공간(Feature Space)로 변환하며, 이러한 공간상에서 하나의 집단과 다른 집단을 분류하는 최적의 결정 경계면인 초평면(Hyper-plane)을 구하게 된다. 초평면과 가장 가까운 데이터를 SV(Support Vector)라고 하며[8], 초평면과 SV와의 거리를 마진이라고 한다. 분류를 위

해서는, 이 마진이 최대가 되는 최적의 초평면을 구하게 되면, 이는 데이터들간의 분류가 최대 성능을 가지고 있다고 볼 수 있다.

n 개의 데이터(x_i, y_i)로 주어진 SVM 문제는, 모든 데이터를 다음의 제약조건($w^T x_i + b \geq 1$), for $y_i = 1$ 와 ($w^T x_i + b \leq -1$), for $y_i = -1$)을 갖도록 설계한다[8]. 여기서 w 와 b 는 분류방정식을 나타내는 매개 변수이다. 이는 데이터를 $w^T x_i + b$ 으로 계산하면 두 개의 그룹($y_i = 1, y_i = -1$)에 대하여 1 이상 혹은 -1 이하의 값을 갖게 되므로 데이터를 두 개의 클래스로 분류할 수 있다. 이를 하나의 조건식으로 나누면 다음과 같다[8].

$$y_i(w^T x_i + b) - 1 \geq 0, \text{ for } i = 1, \dots, N \quad (1)$$

SVM에서는 모델의 구조를 나타내는 $\frac{1}{2}\|w\|^2$ 를 최소화하면서 데이터의 분류를 최대로 하는 모델을 탐색한다. 이러한 문제를 종합적으로 기술하면 다음과 같다.

$$\min J(w) = \frac{1}{2}\|w\|^2 \quad (2)$$

$$\text{subject to } y_i(w^T x_i + b) - 1 \geq 0, \text{ for } i = 1, \dots, N \quad (3)$$

이 최적화 문제를 풀기 위해서는 라그랑지안(Lagrangian) 최적화 기법을 이용하며, 아래의 식과 같다 [8].

$$L(w, b, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^N \alpha_i (y_i(w^T x_i + b) - 1) \quad (4)$$

여기서 C 는 라그랑지안 multiplier이다. 이를 식 (5), (6)를 만족하는 KKT (Karush-Kuhn-Tucker) 조건을 이용하면 가장 높은 성능을 갖는 분류 모델을 얻을 수 있다.

$$\min \tilde{L}(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (5)$$

$$\text{subject to } \sum_{i=1}^N \alpha_i y_i = 0, 0 \leq \alpha_i \leq C \quad (6)$$

여기서, C 는 성능과 오 분류 간의 균형(trade-off)을 조절해주는 비 용변수이며, $K(x_i, x_j)$ 는 Kernel Function이다[8]. Kernel Function은 Linear Function, Polynomial Function, RBF (Radial Basis Function) 등 다양한 함수가 존재하며(Table 1), 본 연구에서는 $K(x_i, x_j) = (x_i^T x_j)^d$ (d : polynomial degree)의 Polynomial Function을 이용하여 데이터를 초평면 상의 데이터로 변환하였다. 위에서 설명한 SVM을 이용하여 조업 모드를 판별하는 모델을 생성하였으며, SVM의 다양한 파라미터는 Particle Swarm Optimization으로 최적의 파라미터를 탐색하였다[9]. 각 조업 모드에 따라 다음 장에서 설명할 PCA 모델을 구축하였다.

3. PCA(Principal Component Analysis)

PCA는 다양한 공정 변수를 몇 개의 주성분(Principal Component

Table 1. Kernel Functions

Kernel Function	Formula
Linear	$K(x_i, x_j) = x_i^T x_j$
Polynomial	$K(x_i, x_j) = (1 + x_i^T x_j)^d$
RBF	$K(x_i, x_j) = \exp(-a \ x_i - x_j\ ^2)$

(PC))으로 나타내는 데이터 축약 기법이다[2]. 이를 통해, 고차원의 복잡한 데이터를 저차원의 데이터로 축약함과 동시에 변수 간 상관관계에 대한 정보를 최대한 많이 포함되도록 유도할 수 있다.

공정 변수의 분산 분포 중 가장 큰 분산을 갖는 방향에 대해 첫 번째 주성분(PC1)을 정의한다[2]. 그리고, PC1과 직교하는 방향 중에서 가장 큰 분산을 갖는 두 번째 주성분(PC2)를 정의한다 [2]. 이를 반복하면, 기존 공정 변수의 차수보다 적은 차수의 PC만으로 변수 간 상관관계를 대부분 포함하는 새로운 주성분을 생성할 수 있으며, 기존 데이터는 이러한 PC에 직교하여 새로운 데이터가 생성된다[2]. n 개의 데이터(x_i, y_i)의 행렬인 데이터 X 와 Y 는 다음의 관계를 갖는다[10].

$$X = TP^T + E \quad (7)$$

식 (7)에서 T 는 스코어 벡터, P 는 가중치>Loading) 벡터이며 E 는 오차 벡터이다. PCA를 이용해 구한 가중치 벡터를 이용해 새로운 스코어 벡터로 투영하여 저차원의 데이터를 생성하게 된다.

PCA의 결과를 이용해 데이터를 모니터링하는 방법은 스코어 벡터를 이용하는 T^2 (Hotelling's T^2 statistic)과 잔차 벡터를 이용한 SPE(Squared Prediction Error)가 있다[11-12].

$$S = (n-1)^{-1} \sum_{i=1}^n (x_i - m)(x_i - m)^T \quad (8)$$

$$T^2 = (x_i - m)^T S^{-1} (x_i - m) \quad (9)$$

$$SPE = x_i^T (1 - PP^T) x_i \quad (10)$$

여기서 m 은 정상 상태 데이터들의 평균값이며, S 는 공분산이다. T^2 를 모니터링하기 위해서는 상위제어선(UCL, Upper Control Limit)을 설정해야 한다[11-12].

$$T_{UCL}^2 = \frac{(n-1)(n+1)a}{n(n-a)} F_{\alpha}(a, n-a) \quad (11)$$

n 은 데이터 개수, a 는 PC의 수, α 는 신뢰도이며, $F_{\alpha}(a, n-a)$ 는 a 와 $n-a$ 의 자유도를 갖는 F 분포이다. 본 연구에서는 99% 신뢰도($1-\alpha=0.99$)를 갖도록 T^2 의 UCL을 설정하였다. SPE의 UCL은 식 (12)의 오른쪽 항과 같다[10-11].

$$SPE \leq \theta_1 \left\{ \frac{h_0 C_{\alpha} \sqrt{2\theta_2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right\}^{1/h_0} \quad (12)$$

$$\theta_i = \sum_{j=a+1}^m \lambda_j^i \quad (13)$$

$$h_0 = 1 - \frac{2\theta_1 \theta_3}{3\theta_2^2} \quad (14)$$

여기서 m 은 원래 데이터의 차수이며, C_{α} 는 신뢰도가 $1-\alpha$ 인 경우 정규분포 값, λ_j 는 j 번째 PC의 고유값(eigen value), θ_1 는 고유값의 합을 의미한다[11-12]. 본 연구에서는 T^2 와 SPE를 이용해서 실제 공정의 데이터를 모니터링하고 각각 값의 상위 제어선을 벗어나는 지점이 이상이 발생한 것으로 간주할 수 있다.

4. 대상 공정

4-1. 폴리스티렌 중합 공정

폴리스티렌은 과상 중합 공정(Bulk Polymerization)을 통해 제조되며, 중합 공정, 탈 휘발 공정, 압출 공정, 저장 및 제품 포장공정으로 구성되어 있다. 그리고, 중합 공정과 탈 휘발 공정에서 필요한 진공상태를 유지하기 위한 진공 시스템과 스타트 업과 셧다운 시 필요한 승온 및 공정 온도 조절에 필요한 보일러 시스템도 존재한다. 반응기 위치에 따라 개시반응, 성장반응, 종결반응이 일어나게 된다.

Fig. 1에서 보는 바와 같이, 첫 단계의 중합은 수직형의 예비 중합 반응기(R-0)에서 수행되며 다음 과정의 성장반응은 2개의 연속 반응기(R-1, R-2)에서 일어난다. 마지막 단계에 들어가기 전에 반응하지 않고 존재하는 모노머는 두 개의 탈 휘발기에서 제거되며, 마지막 단계에서는 종결반응이 일어나게 된다.

본 중합 반응기(R-1)는 보온재로 둘러싸인 수직형의 반응기로서 교반기가 설치되어 있다. 반응온도는 원하는 분자량에 따라 결정되며 원하는 수준의 반응물을 얻기 위해서 체류시간을 일정하게 유지해 주어야 한다. 중합 반응기의 수위는 원료 공급과 재순환 사이클에 존재하는 조절밸브에 의하여 제어된다. 교반기는 축 방향 흐름 형태로 고점도 하에서도 반응 혼합물 온도를 균일하게 하는 데 사용되며 아주 높은 고점도에서도 교반속도 조절이 가능한 가변속 장치가 붙어있다. 중합 반응기의 보온재는 공장의 스타트업에 반응물의 온도를 유지하는 데 사용된다. 정상 운전 시에는 중합 반응기에서 발생하는 반응은 발열반응이며, 온도 조절은 진공 시스템으로 들어가는 증기 라인 상에 있는 압력조절밸브에 의해 이루어진다. 중합 반응기 내의 반응물은 가 변속 폴리머 펌프에 의해 다음 공정

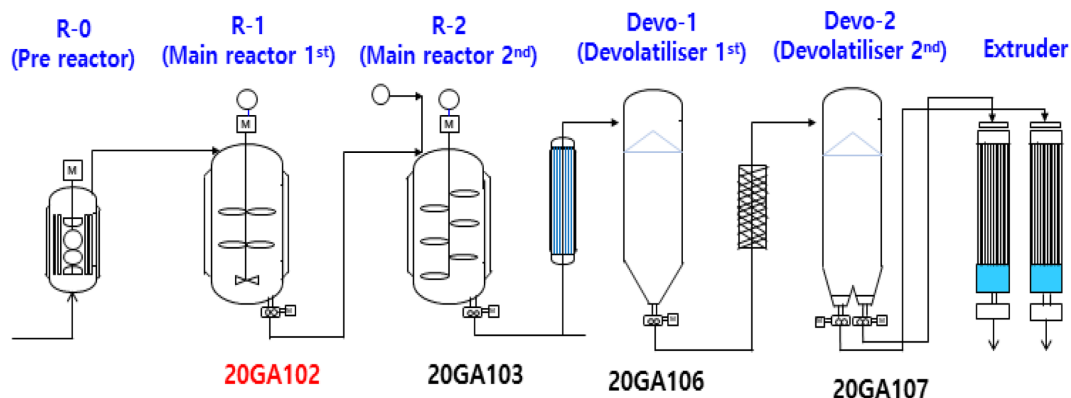


Fig. 1. Bulk polymerization reactors for producing polystyrene.

Table 2. Three operational modes on polystyrene reactors

Operational mode	R-0 (°C)	R-1 (°C)	R-2 (°C)	Melt Flow Index Target	Purpose
1st	132.0	133.0	134.0	1.4~2.0	Heat resistant polystyrene
2nd	132.0	142.5	143.0	2.5~2.9	Double flow polystyrene
3rd	110.0	148.0	152.0	7.0~8.0	High transparency polystyrene

으로 이송되며 이 펌프는 생산량을 조절하는 데 사용된다.

시장 상황에 따라 다양한 제품을 생산하기 위해 3개의 조업 모드 (Table 2)가 존재하며, 반응기가 진공으로 운전됨에 따라 온도가 가장 중요한 요소로 관리되고 있다.

4-2. 실제 사고 상황

본 연구에서 적용한 대상 반응기는 R-1 반응기(Fig. 1)이며, 아래와 같이 온도, 압력, 액위, 펌프의 입력과 출력 유량, 총 5개의 변수가 존재한다.

- 1) R-1 액위
- 2) R-1 압력
- 3) R-1 온도
- 4) R-1 PUMP 유량
- 5) R-2 PUMP 유량

2018년에 R-1 반응기에서 폭주반응이 발생하였다. 오후 12시 59분쯤 반응기 내부에서 오랜 시간 동안 반응을 한 폴리머 덩어리가 R-1 반응기에서 R-2 반응기를 연결하는 배관에 연결된 펌프를 막아 R-2 반응기로의 물질 이송 불량이 발생하였고, R-1 반응기 내 액위가 계속 상승하면서 R-1 반응기의 중합물 온도와 압력이 증가하였다. 제어실에서는 액위 데이터가 급격하게 상승하는 것을 토대로 1시 15분경에 사고가 발생하였음을 인지하였다. 이 사고에 대응하기 위해 절차서에 따른 조치(반응억제제 투입)를 시행하였음에도 불구하고, 폭주반응이 발생하여 사고가 발생하였다. 다행히 섯다운을 통해 폭발 사고 없이 사고는 마무리되었다. 하지만, 반응기와 열교환기는 재가동 불능상태가 되어서 모두 교체하였다. 이 반응기와 관련하여 제대로 된 모니터링 시스템이 존재하지 않았기 때문에 사고가 발생한 후 15분이 지나서 반응기 액위가 정상 상태의 30%를 초과하고 나서야 제어실에서 인지하였으며, 적절한 이상 진단 시스템의 부재는 유사한 사고가 다시 발생하는 경우 큰 폭발 사고와 함께 막대한 인명 및 재산 손실을 불러올 수 있다. 실제 공정 운전에서 유량, 압력등의 센서값을 이용하여 이상을 모니터링하는 경우 이상 이외에도 공정 운전 조건 변화에 따라 센서값의 변화가 생길 수 있으므로 작업자가 센서값만을 이용하여 이상을 모니터링하고 적절한 조치를 취하기는 매우 어렵다. 본 연구에서는 실시간으로 얻을 수 있는 공정 데이터를 활용하여 2장 및 3장에서 설명한 이론을 바탕으로 사고 발생 후 신속하게 사고를 감지하고 이를 작업자들에게 알려줄 수 있는 이상 진단 시스템을 구축하고자 한다.

5. 모델 적용 결과

본 연구를 위해 3가지 조업 모드의 데이터를 확보하여 모델을 검증하였다. 1번, 2번, 3번 모드 데이터를 각각 2,000개, 4,000개, 1,000개씩 확보하였으며, 데이터는 1분마다 센서를 통해 획득한 데이터를 이용하였다. 이를 가지고 SVM을 이용하여 모드를 판별하는 모델을 만들었다. 각 모드의 데이터의 10%인 총 700개의 데이

터를 이용하여 판별 모델을 훈련하였으며, 나머지 90%의 데이터로 테스트를 하였다. 테스트 결과, SVM을 이용하여 조업 모드를 판별하는 것은 완벽하게 가능함을 확인하였다.

Fig. 2는 위에서 언급한 총 7,000개의 데이터를 이용하여 PCA를 통해 5개의 변수를 2개의 변수로 축약하였을 때 패턴을 보여주고 있다. 여기서 확인하는 바와 같이 선형적인 방법을 이용하면 2번, 3번 모드의 주성분(PC)들이 서로 겹치는 것을 볼 수 있으며, 이는 단순한 선형적인 방법을 이용하여 조업 모드를 판별하는 모델을 만드는 경우 2번과 3번 모드가 명확하게 구분이 안 됨을 의미한다. 이러한 이유로 본 연구에서는 비선형 데이터를 판별하는데 효과적인 SVM을 이용하였다.

Fig. 3은 총 7,000개의 데이터를 이용하여 SVM을 통해 조업 모드를 예측한 결과이다. 여기서 X축은 데이터의 개수이며, Y축은 SVM을 이용하여 만든 판별 모델이 제시한 공정의 조업 모드이다. 이를 통해, SVM을 이용하여 1번, 2번, 3번 조업 모드 모두 서로 겹치는 부분 없이 완벽하게 판별할 수 있음을 확인하였다.

실제 사고 데이터를 이용하여 모드 판별 모델의 성능을 검증하였다. 여기서 사용한 데이터는 1번 조업 모드에서 공정이 운전되는

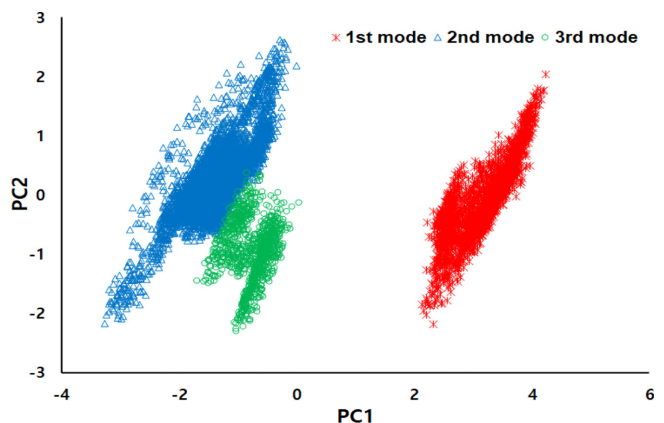


Fig. 2. The results of PCA based on 7000 data sets.

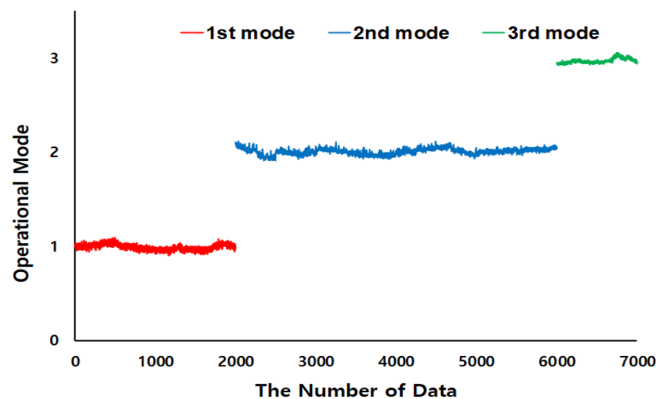


Fig. 3. The results of a mode classification model based on SVM.

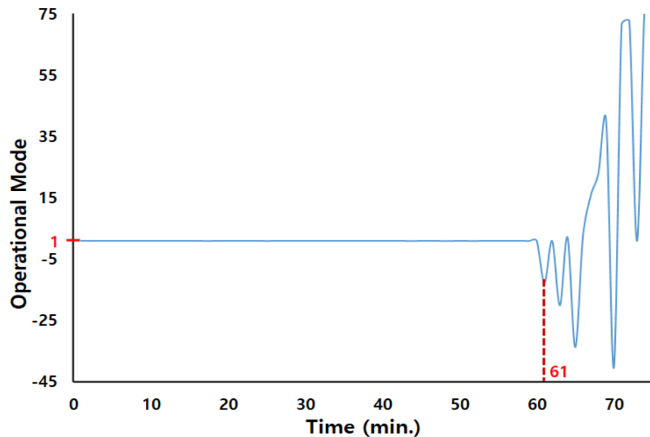


Fig. 4. The results on a real data set monitored by a SVM model classification model.

중 사고가 발생하기 약 1시간 전(오후 12시) 부터 확보한 실제 데이터이며, 실제 사고는 오후 12시 59분에 발생하였다. Fig. 4는 실제 사고가 발생한 이후 조업 모드 판별 모델을 통해 예측한 결과를 보여주고 있으며 그래프에서 원점은 오후 12시를 의미한다. 1번 모드에서 61분 후(오후 1시 1분)에 다른 모드로 벗어남을 알 수 있으며, 이후에는 2번이나 3번 모드가 아닌 다른 모드로 불안정하게 변하는 것을 확인할 수 있다. 이 모델을 이용하더라도 사고 발생 후 약 2분 후부터는 반응기에서 이상이 발생함을 확인할 수 있다.

조업 모드를 판별하기 위해 수집한 총 7,000개의 데이터를 이용하여 PCA 모델을 만들었다. 각 조업 모드에 맞는 모델을 구축하였으며, 각각의 조업 모드의 T^2 와 SPE의 UCL을 식 (11)과 (12)를 통해 계산하였다. 실제 사고가 발생한 시점에서는, 1번 모드로 운전되다가 사고가 발생하였기 때문에 1번 모드에 맞는 PCA 모델을 이용하여 T^2 와 SPE를 계산하였다. 이때, T^2 와 SPE의 UCL은 각각 9.37, 6.84이다.

Fig. 5는 1번 모드로 운전되는 데이터를 이용하여 PCA 모델을 구축하고 실제 사고 데이터를 이용하여 계산한 결과를 보여주고 있다. 정상 상태의 경우 Fig. 5 가운데 점선으로 표시된 영역에서 PC가 존재하다가 사고가 발생하면 이 영역을 급격하게 벗어나는 것을 확인할 수 있다.

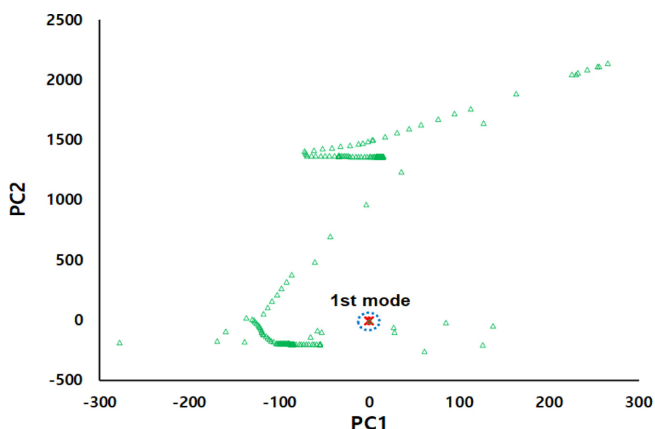


Fig. 5. The principal components based on a real data set. The blue dash line indicates the area of a normal operation mode 1.

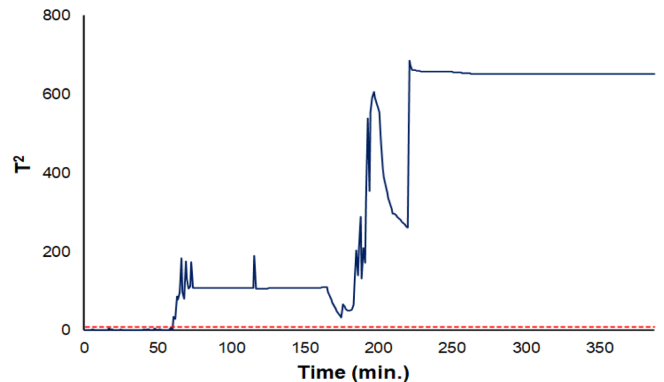


Fig. 6. The monitoring results based on T^2 . The red dash line is the UCL of T^2 .

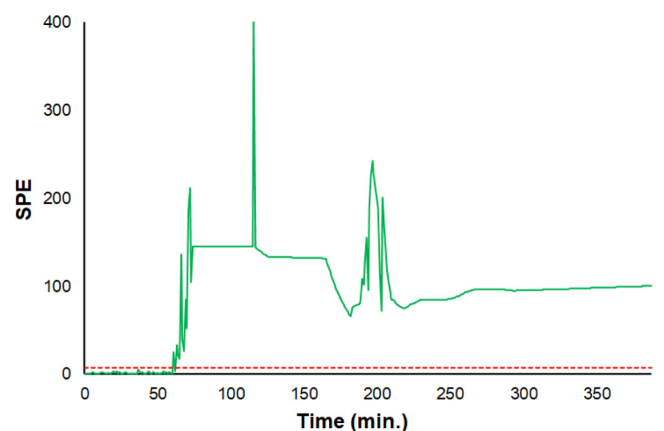


Fig. 7. The monitoring results based on SPE. The red dash line is the UCL of SPE.

T^2 의 경우도 61분부터 정상 상태를 벗어남을 확인할 수 있다 (Fig. 6). 실제 사고의 경우 사고가 발생한 후 15분이 지나서 제어실에서 인지하였으나, 본 연구에서 제안한 모델의 경우를 이용해서 이상을 지속적으로 모니터링 했다면 사고 발생 직후, 그리고 실제 사고를 인지한 시점보다 14분 먼저 인지가 가능했으며, 중대 사고의 발생 가능성을 크게 낮출 수 있음을 확인하였다.

이를 통해, 본 연구에서 제안한 SVM과 PCA를 이용하여 공정의 이상을 진단하는 모델이 단순히 제어실에서 모니터링 하는 것보다 신속하고 정확하게 이상의 발생 여부를 확인할 수 있다고 결론 낼 수 있다.

Fig. 7에서 보는 바와 같이, SPE로 모니터링을 하면 정확하게 데이터를 수집한 지 61분부터(사고가 발생한 지 2분 후) 정상 상태를 벗어남을 확인할 수 있었다.

6. 결 론

본 연구에서는 실제 폴리스티렌 중합 공정에서 발생한 사고 데이터를 이용하여 SVM과 PCA에 기반한 이상 진단 모델을 제안하였다. 실제로 발생한 사고를 늦게 인지하여 대처가 늦었으며, 이로 인해 설비의 막대한 손실이 발생하였다. 본 연구에서 제안한 모델을 적용한 결과 신속하고 정확하게 이상의 발생 여부를 확인할 수 있었다. 많은 변수가 존재하는 화학 공정의 복잡성으로 인해, 단순히

몇 개의 변수만을 모니터링해서 효과적으로 이상의 발생에 대응하는 것은 많은 어려움이 있다.

또한, 다양한 운전 모드가 존재하는 공정인 경우는 각 모드에 적합한 모델을 이용해야 한다. 본 연구에서는 먼저 공정이 어떤 운전 모드에서 운영되는지 판별하는 모델을 SVM에 기반하여 구축하였다. 그리고, 각 운전 모드에 적용할 PCA 모델을 만들었으며, T^2 와 SPE의 UCL을 이용하여 이상의 발생 여부를 진단하였다. 결과를 통해 공정상에 문제가 발생하자마자 바로 이상의 발생 여부를 확인할 수 있었으며, 이는 본 연구에서 제안한 모델을 이용하면 작업자가 신속하게 이상에 대응하여 공정의 큰 손실을 방지하는 데 큰 도움을 줄 수 있음을 의미한다.

기존의 많은 이상 진단 관련 연구에서는 벤치마킹 공정을 이용하여 시뮬레이션을 통해 데이터를 얻어서 이를 이상 진단에 활용하였다. 본 연구에서는 실제 사고가 발생한 공정의 데이터를 이용하여 이상 진단을 수행하였으며, 공정의 특성을 분석하여 그에 적합한 통계기법을 선정하고 이를 적용하였다. 데이터를 분석한 결과 총 3 개의 조업 모드가 있음을 확인하였고, 이는 공정 운영 매뉴얼과 일치함을 확인하였다. 각 조업 모드의 정상 상태 데이터를 이용하여 PCA 모델을 수립하였으며, 각 조업 모드의 T^2 와 SPE의 UCL을 계산하였다. 실제 사고가 발생한 데이터의 경우 첫 번째 조업 모드에서 운영되는 도중에 사고가 발생함을 확인하였으며, 조업자가 센서를 통해 사고를 인지한 시점보다 14분 먼저 사고 발생 여부를 확인할 수 있었다.

본 논문에서 제안한 모델을 통해 이상이 발생하는 경우 신속하게 이상 발생 여부를 판별할 수 있으며, 이를 통해 공정상의 이상을 대응하는데 충분한 시간 확보가 가능하며, 사고로 인해 발생 가능한 막대한 사고 비용을 줄일 수 있다. 이를 통해 공정의 경쟁력 향상에 큰 도움을 줄 수 있을 것으로 기대된다.

감 사

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2020R1F1A105544212).

References

1. Lee, K. O. and Lee, C. J., "The Development of a Fault Diagnosis Model based on the Parameter Estimations of Partial Least Square Models," *J. Korean Soc. Saf.*, **34**, 59-67(2019).
2. Park, J. Y. and Lee, C. J., "Principal Component Analysis Based Method for Effective Fault Diagnosis," *J. Korean Soc. Saf.*, **29**, 73-77(2014).
3. Venkatasubramanian, V., Rengaswamy, R., Yin, K. and Kavuri, S. N., "A Review of Process Fault Detection and Diagnosis: Part I: Quantitative Model-based Methods," *Com. Chem. Eng.*, **27**, 293-311(2003).
4. Venkatasubramanian, V., Rengaswamy, R., Yin, K. and Kavuri, S. N., "A Review of Process Fault Detection and Diagnosis: Part II: Qualitative Models and Search Strategies," *Com. Chem. Eng.*, **27**, 311-326(2003).
5. Venkatasubramanian, V., Rengaswamy, R., Yin, K. and Kavuri, S. N., "A Review of Process Fault Detection and Diagnosis: Part III: Process History Based Methods," *Com. Chem. Eng.*, **27**, 327-346(2003).
6. Lee, C. J., Lee, G. and Lee, J. M., "A Fault Magnitude Based Strategy for Effective Fault Classification," *Chem. Eng. Res. Des.*, **91**, 530-541(2013).
7. Lee, C. J., Ko, J. W. and Lee, G., "Operation Modes Classification of Chemical Processes for History Data-Based Fault Diagnosis Methods," *Korean Chem. Eng. Res.*, **46**, 383-388(2008).
8. Vapnik, V. N., *The nature of statistical learning theory*, Springer Verlag, New York, U.S.A., 17-68(1995).
9. Almeida, G. M., Park, S. W., Kim, S. C. and Lee, C. J., "A Smoothing-based Stochastic Optimization for a Boiler Process in a Pulp Industry," *Korean J. Chem. Eng.*, **37**, 588-596(2020).
10. Wold, S., Esbensen, K. and Geladi, P., "Principal Component Analysis," *Chemom. Intell. Lab. Syst.*, **2**, 37-52(1987).
11. Hotelling, H., *Multivariate Quality Control—Illustrated by the Air Testing of Sample Bombsights*, in *Techniques of Statistical Analysis*, Eisenhart C, Hastay M.W. and Wallis W.A., (eds) McGraw-Hill, New York, 111-184(1947).
12. Pyun, H., Lee, C.-J. and Lee, W. B., "Development of Monitoring System for the LNG Plant Fractionation Process based on Multi-mode Principal Component Analysis," *J. Korean Inst. Gas*, **23**, 32-38(2019).

Authors

Yeonsu Jeong: Graduate student (Master degree), Department of Safety Engineering, Pukyong National University, Busan 48513, Korea; shine0718@naver.com

Chang Jun Lee: Ph.D, professor, Department of Safety Engineering, Pukyong National University, Busan 48513, Korea; changjunlee@pknu.ac.kr