

부호유향그래프와 동적 부분최소자승법에 기반한 화학공정의 다중이상진단

Multiple-Fault Diagnosis for Chemical Processes Based on Signed Digraph and Dynamic Partial Least Squares

이기백, 신동일, 윤인섭
(Gibaek Lee, Dongil Shin and En Sup Yoon)

Abstract : This study suggests the hybrid fault diagnosis method of signed digraph (SDG) and partial least squares (PLS). SDG offers a simple and graphical representation for the causal relationships between process variables. The proposed method is based on SDG to utilize the advantage that the model building needs less information than other methods and can be performed automatically. PLS model is built on local cause-effect relationships of each variable in SDG. In addition to the current values of cause variables, the past values of cause and effect variables are inputted to PLS model to represent the process dynamics. The measured value and predicted one by dynamic PLS are compared to diagnose the fault. The diagnosis example of CSTR shows the proposed method improves diagnosis resolution, and facilitates diagnosis of masked multiple-fault.

Keywords : fault diagnosis, multiple-fault, signed digraph, principal component analysis, partial least squares

1. 서론

화학공장에서 대규모의 인명피해를 동반한 사고가 자주 발생하면서 공정의 안전에 대한 관심이 크게 높아지고 있다. 특히 미국에서만 사고로 인해 평균적으로 년 200 억불의 손실이 발생하는 것으로 알려져 있다. 미국의 화학 및 석유화학업체에서는 1985 년에서 1990 년사이에 거의 30 건의 대형 사고가 발생하였다[1]. 이것들은 수백의 인명피해를 일으켰고 환경을 오염시켰으며 사고정리, 법적처리비용, 시장점유율 하락, 조업정지비용 등 추가적으로 수십억달러의 금전적피해를 가져왔다. 따라서, 안전하고 신뢰성있는 조업은 화학공정이 수많은 국제시장의 경쟁구조에서 생존하기 위한 가장 중요한 요소 중의 하나가 되었다.

이와 같은 상황에서 조업자가 올바른 판단을 내릴 수 있도록 도와주는 자동화된 이상진단시스템이 필요하게 되었다. 이상진단 시스템은 공정데이터를 온라인으로 분석하고, 공정상태를 감시하다가 비정상적인 상황에서 이상을 진단하여 조업자가 올바른 의사결정을 하는데 충분한 정보를 제공하고 궁극적으로는 조업자가 공정을 더 안전하게 운전하도록 도와주어야 한다.

화학공정이 가진 특유의 비선형성과 불확실성을 고려하여 이상진단방법론의 성능지표들이 다음과 같이 제시되었다[2].

(1) 속도 : 이상진단시스템은 실시간으로 사용되어야 하므로 실시간의 감지와 진단에 대한 시간적인 조건을 만족시켜야 한다.

(2) 정확성 : 진단결과는 하나나 그 이상의 이상후보로서 제시되는데 이때 제시되는 결과는 반드시 진짜 해를 포함하고 있어야 한다.

(3) 선명성 : 제시되는 이상후보의 수는 주어진 증상하에서 최소의 것이어야 한다.

(4) 강건성 : 화학공정이 근본적으로 가지고 있는 불확실하고 불완전한 정보 하에서도 진단할 수 있어야 하고 잘못 감지된 증상, 감지순서의 변화 등에도 불구하고 강건한 진단을 수행해야 한다.

(5) 분산형진단 : 화학공정은 기본적으로 분산된 형태의 제어기를 통해 제어되고 있으며 이상진단시스템은 최종적으로 제어시스템과 통합되어 운영되어야 하므로 이상진단시스템도 분산된 구조로 구축될 수 있어야 한다. 분산형 진단시스템을 통해 지식베이스의 단계적 구축이 가능하고 지식베이스의 관리 및 개선에 유리하다[3].

(6) 이식성 : 개발된 시스템을 다른 공정에 적용하는데 드는 노력이 최소화되어야 한다. 이를 위해 추론기구는 이해 및 구축이 쉽고 지식베이스의 자동합성이 가능해야 한다. 이를 통해 규모가 큰 화학공장에서의 적용이 용이해지며 얻어진 지식베이스의 유지/관리 및 개선이 용이해진다.

(7) 신뢰성 : 이상진단시스템은 예상할 수 있는 모든 이상을 진단할 수 있어야 한다. 특히 기존에 발생하였던 이상뿐만 아니라 공정특유의 이상과 예상가능한 다중이상을 진단하여야 한다. 이상진단에 관한 대부분의 연구에서는 다중이상은 거의 발생하지 않는다고 생각하였다. 그러나, 화학공정의 조업중에는 수요변화나 제품규격의 변화같은 외부외란(external disturbance)이 종종 발생한다. 이런 외란은 공정상태의 변화를 야기하고 이어서 이상이 발생할 수 있다. 결과적으로 이상진단방법론은 외란과 함께 발생한 다중이상을 다룰 수 있어야 한다.

논문접수 : 2002. 9. 19., 채택확정 : 2003. 1. 14.

이기백 : 충주대학교 공업화학과(glee@chungju.ac.kr)

신동일 : 명지대학교 화학공학과(dongil@mju.ac.kr)

윤인섭 : 서울대학교 응용화학부(esyoon@pslab.snu.ac.kr)

* 본 연구는 한국과학재단 목적기초연구(R05-2002-000-00057-0)지원으로 수행되었음.

다중이상은 두개 이상의 이상이 동시 또는 연속적으로 발생하는 것으로 Lee et al.은 다중이상을 유발된 이상, 독립적인 다중이상, 가려진 다중이상, 상호연관된 다중이상의 4 가지 종류로 구분하였다[4]. 이 중에서 특히 가려진 다중이상은 한 개의 이상이 다른 이상으로 인한 모든 정상적 증상을 설명할 수 있는 경우를 말하며 정성적방법만 사용하는 경우에 결코 진단될 수 없다.

화학공정의 이상진단 방법론은 공정모델을 구축하여 진단하는 모델기반 방법론과 공정의 과거데이터를 이용하여 진단하는 데이터기반 방법론으로 구분된다. 또한, 두 방법은 각각 정성적방법과 정량적방법으로 나뉘어 정성적 모델기반, 정량적 모델기반, 정성적 데이터기반, 정량적 데이터기반의 4 가지로 구분된다[5]. 본 연구는 정성적 모델기반 방법인 부호유향그래프(signed digraph, SDG)와 정량적 데이터기반 방법인 부분최소자승법(partial least squares or projection to latent structures, PLS)을 혼합하여 연속화학공정의 이상진단을 위한 새로운 방법을 제시한다. 이를 통해 진단의 정확성과 선명성을 높이고 특히 가려진 다중이상의 진단이 가능하도록 하여 신뢰성을 향상시킨다.

먼저, 2 절에서 SDG, PCA(principal component analysis), PLS 에 대한 간략한 소개를 하고 3 절에서 진단방법을 제시하였다. 4 절에서는 예제를 통해 진단성능을 확인하였는데 대상공정은 비가역 1 차반응, A→B 이 발생하는 CSTR 과 열교환기로 구성되어 있다(그림 1). 이 공정은 단순하지만 화학공정에서 나타나는 모든 일반 특성을 나타내고 있다. 이 공정은 Kramer 가 사용하기 시작했으며 이 연구에서는 Sorsa 의 모델로 모사한 결과를 학습 및 진단데이터로 사용하였다[6, 7]. 데이터의 측정간격은 5 초이며 모사시간은 2,000 초이다. 첫번째 또는 단일이상의 발생시간은 100 초로 하였고 두 번째 이상은 100 초 또는 200 초에 발생되도록 하였다.

II. 관련된 연구

1. SDG

SDG 는 이상진단을 위해 공정변수 간의 미시적인 인과관계를 그림으로 표현하는 간단하면서도 효율적인 방법이다[8]. SDG 는 정점(node)과 정점사이를 연결하는 간선(arc)으로 구성되는데 정점은 공정변수를 의미하고 간선은 정점사이의 즉각적인 인과관계를 나타낸다. 영향의 방향은 간선의 부호로 표시한다. 양부호(+)는 원인-결과정점이 같은 방향으로 움직임을 나타내며 음부호(-)는 반대방향으로 움직임을 나타낸다.

SDG 는 공정의 수학적모델이나 공정지식 및 데이터로부터 얻어질 수 있어 다른 방법에 비해 상대적으로 적은 정보로 모델을 구성할 수 있고 특히 각 장치에 대한 SDG 로부터 공정 전체에 대한 SDG 를 자동합성할 수 있다는 장점이 있다. 또한, 공정 물리량간의 인과관계나 공정장치들간의 이상전파경로를 이해하기 쉽게 시각적으로 볼 수 있다. 반면에 진단의 선명성이 떨어진다는 단점이 있고 다른 정성적 모델 마찬가지로 다중이상 진단에 어려움이 있다[4].

2. PCA, DPCA 와 PLS

PCA 는 대표적인 차원감소기법으로 모델링, 잡음감소 등의 진단의 선명성이 떨어진다는 단점이 있고 다른 정성적 모델

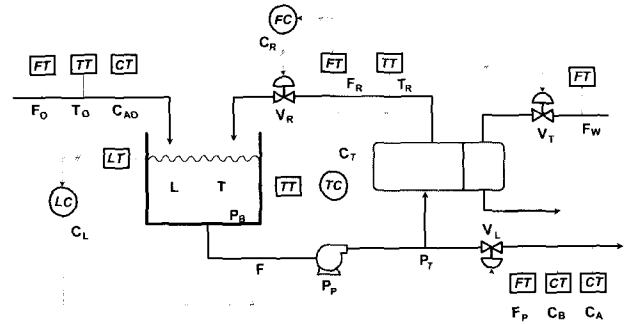


그림 1. CSTR 공정의 공정흐름도[6].
Fig. 1. Process flow diagram of the CSTR process [6].

마찬가지로 다중이상 진단에 어려움이 있다[4].

2. PCA, DPCA 와 PLS

PCA 는 대표적인 차원감소기법으로 모델링, 잡음감소 등의 목적으로 사용된다. PCA 는 정보의 손실을 최소화하면서 원래 변수들의 선형조합을 통해 주성분(PC)이라 불리는 새로운 특성변수를 만들어 고차원의 변수를 저차원의 변수로 표현한다[9]. 데이터행렬 X 는 score 행렬 T 와 loading 행렬 P 의 곱으로 분할된다.

$$X = \bar{X} + E = \sum_{i=1}^k t_i p_i^T + E = TP^T + E \quad (1)$$

여기서, E 는 원래 값과 모델에 의한 값의 차이, 잔차라고 하고 k 는 주성분의 수이다. t 는 원래변수를 PC 공간으로 투영한 좌표이며 p 는 선형 변환의 계수로서 원래 변수와 PC 의 관계를 의미한다.

DPCA(dynamic PCA)는 공정변수의 정적특성뿐만 아니라 동특성도 모델에 포함시켜 모델의 정확성을 높이기 위해 PCA 와 ARMAX(auto-regressive moving average exogenous)시계열모델을 결합시킨 것이다[10]. 즉, 변수의 현재값이 과거값에 의해 영향을 받는 것을 표현하기 위해 DPCA 의 데이터행렬은 각 변수의 현재값 X 에 더하여 l 개의 과거값을 포함하는 다음 형태가 된다.

$$X_d = [X(t) \quad X(t-1) \quad \dots \quad X(t-l)] \quad (2)$$

PLS 는 입력변수 X 를 이용하여 출력변수 Y 를 예측하는 방법이다[9]. PLS 는 입력 X 와 출력 Y 를 각각 loading 벡터 P 와 C 를 사용하여 PCA 를 실행하고 입력 X 와 출력 Y 에 대한 두 PC 사이의 모델을 구성한다.

$$X = TP^T + E \quad (3)$$

$$Y = UC^T + F \quad (4)$$

이때, 출력 Y 에 대한 PC 의 잔차, F 를 가능한 한 작게 하여 Y 의 예측성능을 극대화하면서 동시에 X 와 Y 의 관계를 잘 나타낼 수 있도록 한다. PLS 의 모델계수를 결정하기 위한 여러 방법 중에서 NIPALS(Non-iterative Partial Least Squares)알고리즘이 가장 많이 사용된다[9].

표 1. CSTR 공정의 각 장치에 대해 정의된 이상.

Table 1. Fault classification according to equipment in CSTR.

Equipment	Fault	Abbreviation letter
Sensor	Bias high (including stuck)	-BH
	Bias low (including stuck)	-BL
Valve	Bias high (including failure)	-BH
	Bias low (including failure)	-BL
Pipe	Blockage	-BK
Reactor	Leak	-LK
Pump	Equipment failure	-EF
Heat exchanger	Fouling	-FL
External disturbance	Flowrate change high	-FCH
	Flowrate change low	-FCL
	Temperature change high	-TCH
	Temperature change low	-TCL
	Composition change high	-CCH
	Composition change low	-CCL
Operator disturbance	Setpoint change high	-SVCH
	Setpoint change low	-SVCL

III. 진단방법

제안된 방법은 SDG 를 수정하고 SDG 의 부분적 인과관계를 이용하여 DPLS 를 구축하는 오프라인 해석과정과 온라인으로 수집되는 데이터를 이용하여 진단을 수행하는 과정으로 구분된다.

1. 모델구성

1.1 SDG의 수정: 먼저 SDG 를 작성하고 측정정점만으로 구성된 축소된 유향그래프를 작성한다[6]. 또한, 물리적으로 의미없는 이상만을 다루기 위해서 각 장치에 대해 물리적으로 발생가능한 이상을 정의하고 이것들을 축소된 유향그래프의 각 측정정점에 추가한다. 기존 SDG 기반 방법론에서 이상은 물리적으로 의미없는 정점과 동일시되어 SDG 에서 완벽하게 표현될 수 없었다. 예로서, CSTR 의 각 장치에 대해 정의된 이상은 표 1 과 같고 축소된 유향그래프는 그림 2 와 같다. 또한, 독립센서(independent sensor)이상을 제외한 각 이상을 측정정점에 할당된 결과는 표 2 에 있다. 이상의 이름은 장치

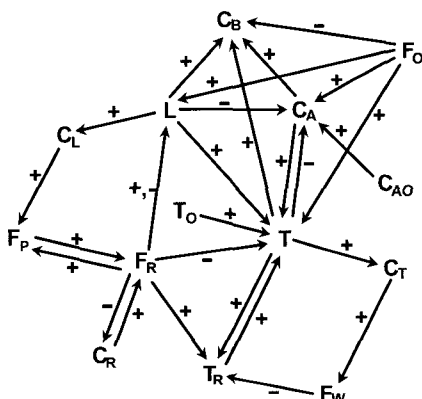


그림 2. CSTR 공정의 축소된 유향그래프.
Fig. 2. Reduced digraph of the CSTR process.

표 2. CSTR 공정의 측정정점에 추가된 이상.

Table 2. Fault added to the measured node in CSTR.

Measured variable	Arc sign of positive (+)	Arc sign of negative (-)
CA0	FEED-CCH	FEED-CCL
F0	FEED-FCH	FEED-FCL, FP-BK
T0	FEED-TCH	FEED-TCL
CL	LC-SVCL	LC-SVCH
CR	FC-SVCH	FC-SVCL
CT	TC-SVCL	TC-SVCH
FP	VL-BH	VL-BL, PP-BK, PUMP-EF
FR	VR-BH, FS-BH	VR-BL, FS-BL, PUMP-EF, RP-BK
FW	VT-BH	VT-BL, WP-BK
L	LS-BH, PUMP-EF	LS-BL, RX-LK
T	TS-BH	TS-BL
TR	CW-TCH, HX-PL	CW-TCL

의 이름에 이상의 종류를 추가시켜 만들었다. 예를 들어 VL-BH 는 수위제어밸브의 높은 편향을 의미한다.

1.2 DPLS 모델구성: 유향그래프는 공정변수의 즉각적인 영향을 표현하고 있으므로 간선의 결과정점은 원인정점에 의해 직접적인 영향을 받으며, 원인정점의 앞쪽에 위치한 이상 또는 정점이 결과정점에 미치는 영향은 원인정점을 통과하여 간접적으로 나타나게 된다.

따라서, 각 측정정점의 변수를 출력 Y 로 하고, 출력변수를 결과정점으로 하는 간선의 원인정점을 입력 X 로 하는 PLS 모델을 구성하여 출력변수를 예측할 수 있다. 또한, 동특성을 정확히 다루기 위해 PLS 에 ARMAX 모델을 결합시켜 원인정점의 측정값뿐만 아니라 원인정점의 과거값, 출력변수의 과거값을 입력으로 사용하는 DPLS(Dynamic PLS)모델을 구성하였다. 예를 들어, CSTR 에서 TR 은 측정정점 FR, FW, T 의 DPLS 모델에 의해 추정될 수 있다. 측정값과 추정값 사이에 오차를 만드는 원인 이상은 두 종류로 나뉘어지는데, 각 결과정점에 추가된 이상과 DPLS 모델을 구성하는 측정원인정점의 센서이상이다. TR 의 예에서는 TR 에 추가된 이상인 CW-TCH, CW-TCL(냉각수 온도상승 및 하락), HX-PL(열교환기 파울링)과 TR, FR, FW, T 의 센서이상이 원인 이상이 될 수 있다. 다른 이상들, 예를 들어 VR-BH 는 FR 을 거쳐 TR 에 나타나므로 TR 의 측정값과 추정값 사이에 오차를 만들지 않는다. 예제에서 FP 를 결과정점으로 하는 원인정점은 CL, FR 로, 1 개의 과거값을 사용할 때 DPLS 모델의 입력 X 는 다음과 같다.

$$X = \begin{bmatrix} CL(1) & FR(1) & FP(0) & CL(0) & FR(0) \\ CL(2) & FR(2) & FP(1) & CL(1) & FR(1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ CL(n) & FR(n) & FP(n-1) & CL(n-1) & FR(n-1) \end{bmatrix} \quad (5)$$

1.3 학습데이터 : DPLS 모델을 얻기 위한 학습용 조업데이터는 운전과정에서 쉽게 얻을 수 있는 데이터, 예를 들어 설정치변경 등에 의한 데이터를 이용할 수 있다. 원인정점에서

결과정점으로서의 영향, 즉 각 DPLS 의 입력력변수 사이의 모든 관계를 확인할 수 있는 데이터만으로 DPLS 모델을 얻을 수 있으므로 기존 데이터기반 방법과는 달리 실제로 얻기 어려운 이상발생시의 데이터를 필요로 하지 않는다[10]. 이때, 각 데이터의 원인이 되는 이상에 의해 직접 영향을 받는 측정정점의 모델을 만드는 경우에는 그 데이터를 학습데이터로 사용하지 않는다. 이것은 이상에 의해 그 변수의 예측모델이 부정확해지기 때문이다. 예를 들어, CSTR 공정에서 CW-TCH 와 CW-TCL 에 의한 학습데이터는 TR 의 DPLS 모델을 구할 때 사용할 수 없다.

1.4 시간지연과 DPLS 모델계수의 결정 : DPLS 모델에서 사용할 과거값의 수인 시간지연 l 을 학습데이터로부터 구한다. 시간지연의 결정은 DPCA 에서 사용한 것과 유사한 방법으로 상호상관도 그래프를 기준으로 주성분의 수를 결정하는 방법을 이용한다[10]. 예제공정 CSTR 에서 결정된 각 DPLS 의 시간지연과 주성분의 수는 표 3 과 같다. 결정된 시간지연에 따라 (5)와 같은 입력 X 를 정하고 다변량 통계 소프트웨어인 PlantAnalyst[®] 를 이용하여 DPLS 모델의 계수를 계산한다. 그림 3 은 예제공정에서 반응기온도 T 의 추정값(점선)과 측정값(실선)을 시간지연을 다르게 하여 비교하는 그래프이며 동특성의 포함으로 예측결과가 매우 좋아짐을 보여준다.

예제공정에서 측정정점 F_0, T_0, CA_0 는 영향을 주는 측정정점이 없으므로 DPLS 모델을 구성할 수 없으며 별도로 감시되어야 한다. 또한, 제어출력 CR, CL, CR 에 대한 감시는 설정치변화를 진단하기 위한 것으로 대부분의 공정에서는 설정치가 진단시스템으로 입력되므로 감시의 필요성이 없다. 그러나, 예제공정에서는 설정치가 입력되지 않으므로 설정치변화를 감지하기 위해서 제어출력을 감시해야 한다. 이를 위해 제어기의 오차와 적분오차를 입력으로 제어출력을 예측하는 PLS 모델을 구성하였다. 이 모델을 위한 주성분의 수는 각각 2 개였다.

각 DPLS 의 입력변수 중에서 통계적으로 결과에 거의 영향을 미치지 않는 변수를 확인할 수 있다. 즉, SDG 작성에서의 오류를 공정데이터를 이용하여 통계적으로 확인하게 되며 이 간선을 제거하여 진단의 선명성을 향상시킬 수 있다. 이 연구에서는 스케일된 회귀계수를 이용하여 각 입력변수가 출력변수의 예측에 미치는 영향의 정도를 얻었다. 가장 낮은 스케일된 회귀계수를 가진 변수는 L 을 추정하는 DPLS

표 3. CSTR 공정의 주성분수와 시간지연.
Table 3. Number of principal components (PCs) and time delay of the CSTR process.

Variable	Number of PCs	Time delay
FP	2	1
FR	2	1
FW	1	1
L	4	2
T	7	2
TR	3	1
CA	5	2
CB	5	2

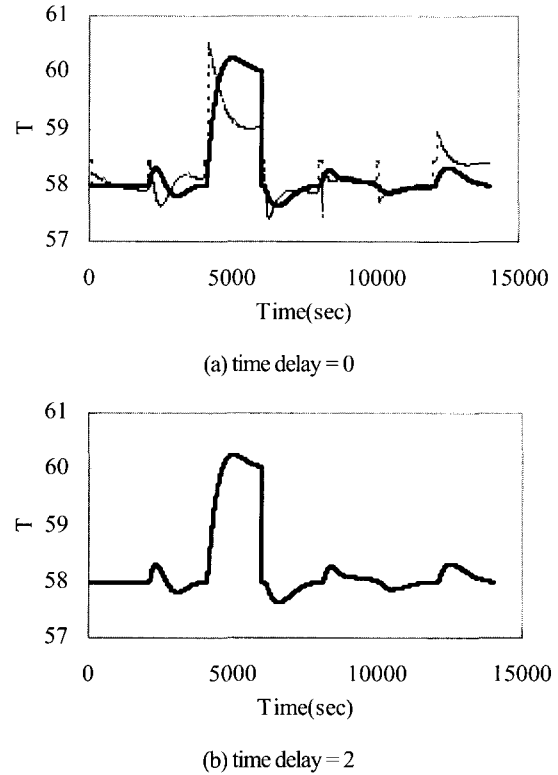


그림 3. 반응기온도의 측정값과 추정값의 비교.
Fig. 3. The comparison of measured reactor temperature with estimated one.

의 입력변수 FR 이었으며 스케일된 회귀계수는 0.012%였다. 간선 $FR \rightarrow L$ 이 +와 -의 두 부호를 가지기 때문에 예측모델에서 영향력이 적어졌음을 확인하고 L 의 추정모델에서 FR 을 제거하였다. 그러나, 실제로 의미있는 간선을 제거할 경우 진단에 실패할 수가 있다. 따라서, SDG 를 수정하는 것은 진단의 선명성과 정확성을 고려하여 조심스럽게 결정하여야 한다.

1.5 이상집합(fault set)의 작성: 각 측정변수의 측정값과 DPLS 에 의한 추정값의 차이, 잔차를 감시하여 진단을 수행한다.

$$\text{잔차}_i = \text{측정값}_i - \text{추정값}_i \quad (6)$$

각 변수의 잔차는 이상에 따라 증가(+) 또는 감소(-)하게 되고 각 잔차의 (+) 또는 (-)의 정성적상태는 증상이 된다. 각 잔차의 증상에 따라 이상을 구분하여 이상집합으로 나타낸다. 어떤 측정변수의 잔차에 변이를 일으키는 이상은 측정값에 변이를 일으키는 이상과 추정값에 변이를 일으키는 이상으로 각각 구분된다. 측정값에 변이를 주는 그 측정정점에 부가된 이상들이며 증상의 부호는 이상과 측정변수를 연결하는 간선의 부호와 같다. 따라서, 측정정점에 부가된 이상집합이 그대로 이상집합에 포함된다. 추정값에 변이를 주는 것은 DPLS 모델에 포함된 변수 중 출력변수를 제외한 측정변수의 센서이상이며 증상의 부호는 센서이상의 부호와 입력/출력변수사이의 간선에 할당된 부호의 곱과 반대이다. 이 연구에서는 제어와 관련된 측정변수의 편차이상만이 이상집합

표 4. CSTR 공정의 이상집합.

Table 4. Fault sets of the CSTR process.

Symptom	Fault set
CA0(+)	FEED-CCH
CA0(-)	FEED-CCL
F0(+)	FEED-FCH
F0(-)	FEED-FCL, FP-BK
T0(+)	FEED-TCH
T0(-)	FEED-TCL
CL(+)	LC-SVCL
CL(-)	LC-SVCH
CR(+)	FC-SVCH
CR(-)	FC-SVCL
CT(+)	TC-SVCL
CT(-)	TC-SVCH
FP(+)	VL-BH, FS-BH
FP(-)	VL-BL, PP-BK, PUMP-EF, FS-BL
FR(+)	VR-BH, FS-BH
FR(-)	VR-BL, FS-BL, PUMP-EF, RP-BK
FW(+)	VT-BH
FW(-)	VT-BL, WP-BK
L(+)	LS-BH, PUMP-EF
L(-)	LS-BL, RX-LK
T(+)	TS-BH, FS-BH, LS-BL
T(-)	TS-BL, FS-BL, LS-BH
TR(+)	CW-TCH, HX-PL, FS-BL, TS-BL
TR(-)	CW-TCL, FS-BH, TS-BH
CA(+)	LS-BH, TS-BH
CA(-)	LS-BL, TS-BL
CB(+)	LS-BL, TS-BL
CB(-)	LS-BH, TS-BH

에 포함된다. 예제공정에 대한 얻어진 이상집합은 표 4와 같다.

2. 온라인 진단

2.1 감지법과 계수의 설정: 제안된 방법은 잔차를 감시하면서 잔차의 변이가 감지되어 증상이 얻어지면 진단을 수행한다. 이와 같이 정량적 데이터(잔차)를 정성적인 정보(증상)로 바꾸는 방법은 Shewhart chart, 이동평균법, CUSUM 등의 통계적인 방법들이 사용된다. 이 연구에서는 실시간용으로 적합한 반복적 계산형태를 가지고 있고 잡음에 강하면서 필터링이 필요없는 높은 감지력을 가진 장점이 있는 CUSUM을 사용하였다. CUSUM에서는 2개의 매개변수, 최소급등크기와 경계치의 크기를 정해주어야 하는데, 이 연구에서는 잔차의 분포에 대한 6σ 를 최소급등크기로 하였고 CUSUM의 분포에 대한 3σ 를 경계치로 사용하였다[11]. 이때 발생할 수 있는 문제점으로는 거짓정보와 시간지연이 있을 수 있는데 거짓정보는 경계치를 크게 함으로써 막을 수 있고 시간지연은 경계치를 작게 함으로써 막을 수 있다. 따라서 경계치는 두 가지 문제점의 중요성을 고려하여 결정하여야 한다.

2.2 이상집합에 의한 진단 : 기본적인 진단전략은 감지된 모든 증상을 설명할 수 있는 이상의 최소집합을 최종이상후보

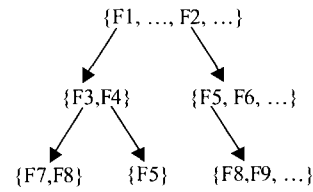


그림 4. 이상진단 사례

Fig. 4. An example of fault diagnosis procedure.

로 하는 것이다. 즉, 한 개의 이상으로 모든 증상이 설명되지 않으면 다중이상을 이상후보로 제시한다. 이상진단순서는 다음과 같다.

① 감지된 각 변수의 잔차는 감지된 증상집합(set of detected symptoms)의 원소가 되며 이 집합의 각 원소에 대응되는 이상집합을 리스트로 작성하고 기본이상집합 리스트라 한다. 이 리스트의 원소인 이상집합들의 합집합을 구성하면 기본이상후보집합이 된다. 이 때 어떤 측정변수의 센서이상 이 기본이상후보가 되면, 그 센서 이상이 부가된 측정변수의 잔차가 정해진 부호로 감지되었는지 확인하고 감지되지 않았으면 기본이상후보에서 제외하여 진단의 선명성을 높인다. 측정변수의 센서이상은 그 측정변수의 측정값에 우선적으로 영향을 주게 되므로 그 측정변수의 잔차가 반드시 감지되어야 하기 때문이다. 기본이상후보에서 제외된 이상은 기본이상집합리스트에서도 제외한다.

② 각 기본이상후보가 몇 개의 증상을 설명하는지 알기 위해 그 이상이 기본이상집합리스트의 이상집합 중 몇 개에 포함되었는지 계산하고 이 수를 nES(no. of explained symptoms)라 한다. nES가 가장 큰 이상후보를 일차이상후보 집합에 포함시킨다. 일차이상후보의 nES가 감지된 증상의 수와 같으면 일차이상후보가 모든 증상을 설명할 수 있음을 뜻하므로 일차이상후보집합이 최종이상후보집합이 된다. 일차이상후보의 단일 이상이 감지된 모든 증상을 설명할 수 없으면 다음 단계에서 이중이상을 탐색한다.

③ 일차이상후보집합의 첫번째 원소를 선택한다. 그 일차 이상후보가 포함된 이상집합을 기본이상집합리스트에서 제외하여 얻어진 이상집합리스트를 그 일차이상후보에 할당한다. 이 리스트의 원소인 이상집합들의 합집합을 구성하면 일차이상후보의 이중이상후보집합이 된다. 얻어진 이중이상후보집합의 각 원소, F가 몇 개의 증상을 설명할 수 있는지 계산하면 nES(F)가 얻어진다. nES가 가장 큰 이상후보를 일차이상후보에 연결된 일차이상후보집합에 포함시킨다. 예를 들어 일차이상후보가 F1이고 F1에 대한 일차이상후보로 F3, F4가 얻어졌으면 그림 4와 같은 트리구조로 연결시킨다. 일차이상후보의 nES가 일차이상후보에 할당된 이상집합리스트에 포함된 이상집합의 수와 같으면, 모든 증상이 설명되었음을 뜻하므로 일차이상후보와 일차이상후보의 조합은 최종이상후보가 된다. 모든 일차이상후보에 대해 같은 과정을 수행한다. 모든 증상이 설명되지 않았으면 다음 단계에서 모든 일차이상후보에 대해 삼차이상후보를 탐색한다.

④ 첫번째 일차이상후보에 연결된 일차이상후보집합의

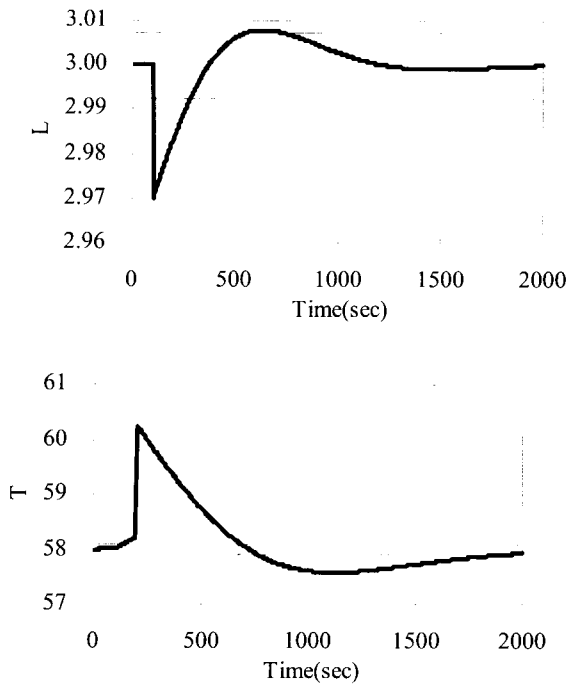


그림 5. 측정변수들의 동특성.
Fig. 5. The dynamics of measured variables.

첫번째 원소를 선택한다. 일차이상후보의 이상집합리스트에서 그 일차이상후보가 포함된 이상집합을 제외하여 얻어진 이상집합리스트를 그 일차이상후보에 할당한다. 이 리스트의 원소인 이상집합들의 합집합을 구성하면 일차이상후보의 삼중이상후보집합이 된다. 얻어진 삼중이상후보집합의 각 원소에 대해 nES 를 계산한다. nES 가 가장 큰 이상후보를 일차이상후보를 위한 삼차이상후보집합에 포함시킨다. 예를 들어 일차이상후보가 F1 이고, F1 에 대한 일차이상후보가 F3 과 F4 이며, F3 에 대한 삼차이상후보가 F7, F8 이 얻어졌으면 그림 4 와 같은 트리구조가 된다. 삼차이상후보의 nES 가 일차이상후보의 이상집합리스트에 포함된 이상집합의 수와 같으면, 연결된 일차, 이차, 삼차이상후보의 조합이 최종이상후보가 된다. 그림 4 의 예에서는 {F1, F3, F7}, {F1, F3, F8} 등이 최종이상후보집합의 원소가 된다. 모든 일차이상후보에 대해 같은 과정을 수행한다. 모든 증상이 설명되지 않으면 유사한 방법으로 삼중이상후보가 탐색되어야 한다.

예를 들어, 수위센서가 100 초에 낮게 편향되었고(LS-BL) 200 초에 온도센서가 높게 편향된(TS-BH) 다중이상예에서 215 초에 대한 진단과정을 보자(그림 5). 105 초에 L(-)의 증상이 감지되어 최종이상후보집합은 두 개의 단일이상으로서 {{LS-BL}, {RX-LK}}가 제시된다. 175 초에는 CA(-)가 감지되어 최종이상후보는 LS-BL 의 단일이상이다. 200 초에 CB(+), 205 초에 T(+)가 감지되지만 LS-BL 이 설명할 수 있으므로 최종이상후보는 175 초에서의 진단과 같다.

① 215 초에는 TR(-)의 잔차가 감지되어 감지된 증상집합은 {L(-), CA(-), CB(+), T(+), TR(-)}가 된다. 기본이상집합리스

표 4. 생성된 다중이상 사례.

Table 4. The selected multiple fault cases.

No.	1 st fault	2 nd fault	2 nd fault time
1	FEED-FCH	FEED-CCL	100
2	FEED-FCL	CW-TCL	200
3	FC-SVCH	TC-SVCH	100
4	LC-SVCL	TC-SVCH	200
5	RX-LK	PP-BK	100
6	RX-LK	WP-BK	200
7	LS-BL	FS-BL	100
8	LS-BL	VL-BH	100
9	LS-BL	TS-BH	200
10	VT-BH	FS-BL	200
11	VR-BH	TS-BL	200
12	VL-BL	TS-BL	200
13	FC-SVCH	VR-BL	200
14	LC-SVCH	VR-BL	100
15	VR-BH	TC-SVCL	100
16	FC-SVCH	VT-BH	100
17	TC-SVCH	VR-BL	200
18	LC-SVCH	VT-BH	100
19	FS-BL	RX-LK	200
20	VR-BL	WP-BK	100
21	LS-BL	RX-LK	100
22	FEED-TCL	FC-SVCH	100
23	LC-SVCH	FEED-CCL	100
24	LC-SVCL	RX-LK	200
25	TC-SVCH	FP-BK	200
26	CW-TCH	LS-BL	100
27	CW-TCH	TS-BH	100
28	FEED-FCH	LC-BL	100
29	FEED-CCH	TC-BH	100
30	FEED-TCL	VL-BL	200
31	FS-BL	FEED-TCL	200
32	FEED-CCL	VR-BH	100
33	FEED-FCL	FS-BH	100
34	FEED-FCH	RX-LK	100
35	WP-BK	FEED-CCL	200
36	CW-TCH	RX-LK	200
37	FP-BK	FEED-TCL	100

트는 다음과 같다.

- {LS-BL, RX-LK}
- {LS-BL, TS-BL}
- {LS-BL, TS-BL}
- {TS-BH, FS-BH, LS-BL}
- {CW-TCL, FS-BH, TS-BH}

이 리스트에 포함된 이상집합들의 합집합을 구하면 기본이상후보집합은 다음과 같다.

- {LS-BL, RX-LK, TS-BL, TS-BH, FS-BH, CW-TCL}

이 중에서 TS-BL 은 T(-)의 증상이, FS-BH 은 FR(+)의 증상이 감지되지 않아 제외된다. 따라서, 기본이상후보집합은 {LS-BL, RX-LK, TS-BH, CW-TCL}이 되며 기본이상집합리스트는

표 5. 단일이상사례의 선명성 비교.

Table 5. The comparison of resolution for the single faults.

Fault	Suggested method	Fault-effect tree
LC-SVCH	1(1)	1(1)
FEED-FCH	1(1)	1(1)
FEED-CCL	1(1)	1(1)
FEED-TCL	1(1)	1(1)
RX-LK	1.16(2)	1(1)
WP-BK	2(2)	3(3)
FC-SVCH	1(1)	1(1)
VR-BL	4(4)	4(4)
VL-BH	1(1)	2.04(4)
TC-SVCH	1(1)	1(1)
VT-BH	1(1)	2(2)
CW-TCH	2(2)	2(2)
LS-BL	1.03(2)	1(1)
TS-BH	1(1)	1(1)
FS-BL	1.03(4)	4(4)

다음과 같다.

- {LS-BL, RX-LK}
- {LS-BL}
- {LS-BL}
- {TS-BH, LS-BL}
- {CW-TCL, TS-BH}

② 각 기본이상후보의 nES 를 구하면 다음과 같다. LS-BL 의 nES 가 4 로서 가장 커서 일차이상후보가 된다. 그러나, 감지된 증상의 수(기본이상집합리스트의 이상집합수)가 5 이기 때문에 이중이상을 탐색해야 한다.

- nES(LS-BL)=4
- nES(RX-LK)=1
- nES(TS-BH)=2
- nES(CW-TCL)=1

③ 일차이상후보가 LS-BL 의 1 개이므로 LS-BL 을 선택하고 기본이상집합리스트에서 LS-BL 를 포함하고 있는 이상집합을 제외하여 얻어진 이상집합리스트를 LS-BL 에 할당한다.

- {CW-TCL, TS-BH}

이 두개의 이상이 이중이상후보집합의 원소가 되며 각이상이 설명할 수 있는 증상의 수, nES 는 1 로서 같다. 따라서, 두 이상은 모두 이차이상후보집합이 되며 nES 가 LS-BL 에 할당된 이상집합리스트에 포함된 이상집합의 수와 같으므로 모든 증상이 설명되었다. 따라서, 최종이상후보는 {{LS-BL, TS-BH}, {LS-BL, CW-TCL}}로 두개의 이중이상이다. 이 예에 대해 정성적 진단법이 적용되었을 때 가려진 다중이상이 되어 진단에 실패하였다. 그러나, 제안된 방법은 정성적 진단법에 비해 정확한 진단성능을 보였으며 최종이상후보도 대부분의 시간에서 두 개의 이중이상으로 제시되어 높은 선명성을 보인다[4].

IV. 진단결과

1. 모델구성

다중이상 진단이 가능한 정성적 모델기반 진단법인 이상-

결과트리모델에 의한 진단결과와 비교하기 위해 동일한 이상상황으로 15 개의 단일이상(표 5)과 37 개의 이중이상(표 4)을 선택하였다[12]. 앞서 서술한 바와 같이 8 개 측정변수의 DPLS 모델과 3 개 제어출력에 대한 PLS 모델을 구성하였다. F0, T0, CA0 는 SDG 기반 진단법처럼 CUSUM 을 사용하여 정상적상태를 결정하도록 하였다. 모델계수를 위한 학습데이터로는 단일이상 중에서 조업 중 자주 발생할 수 있는 상황으로 표 5 에서 설정치변화와 외부교란에 의한 데이터 7 개를 함께 사용하였다. 같은 데이터를 이용하여 CUSUM 의 매개변수, 최소급등크기와 경계치의 크기를 결정하였다.

이상-결과트리를 이용한 진단 연구에서 제안된 제약변수를 이 연구에서도 적용하여 진단의 선명성을 높였다[4]. 제약변수는 수지식과 밸브관계식같은 정량적 지배방정식을 변수 형태로 표현한 것이다. 앞서의 연구에서는 1 개의 물질수지식과 2 개의 제어밸브관계식을 사용하였으나 밸브관계식은 DPLS 에 의해 표현되므로 물질수지식만을 적용하였다. 반응기누출(RX-LK)은 제약변수 DF 의 (+)변이에 대한 원인이상이 된다.

$$DF_1 = F_o - F_p \tag{7}$$

2. 단일이상의 진단결과

진단의 성능을 평가하기 위해 세 개의 지수를 사용한다. 정확성(accuracy)은 진단이 정확할 때, 즉 진짜 해가 최종 이상후보집합의 원소이면 1, 그렇지 않으면 0 이 된다. 어떤 단일이상이 발생한 경우에 그 단일이상이 포함된 다중이상이 해로 제시된 경우에도 정확하다고 한다. 강건성(robustness)은 오감지된 증상의 수를 뜻한다. 정확성이 1 일 때 1 이상의 강건성은 진짜 해와 진짜 해 이외의 이상을 포함하는 다중이상이 해로 제시됨을 의미한다. 일반적으로 진단의 강건성은 불완전하고 부정확한 데이터, 계수의 오류, 잡음 등 증상의 왜곡에도 불구하고 실패하지 않는 진단을 의미한다. 이상-결과트리모델에 의한 진단법에서는 오감지의 문제가 발생하지 않으므로 강건성은 성능지수가 되지 않아 비교될 수 없다. 선명성(resolution)은 정확성이 1, 강건성이 0 일 때 최종 이상후보집합의 원소 수를 나타낸다. 진단이 정확한 경우에도 최종이상후보로 제시된 후보집합의 수가 많으면 진단성능이 낮음을 나타낸다.

모든 경우에서 정확성은 1, 강건성은 0로 가장 좋은 결과가 얻어졌다. 정성적 진단법은 VL-BH의 예에서 45초동안 이중이상을 해로 제시하였으나, 제안된 방법은 더 강건한 결과를 보였다. 표 5는 선명성을 이상-결과트리모델과 비교한 것으로 최초감지시간부터 최종진단시간까지의 구간에 대한 지수의 평균을 보이고 있다. 괄호안은 전체 진단구간에서 가장 나쁜 진단성능을 보이는 시간에서의 지수이다. RX-LK와 LS-BL의 두 예에서는 정성적 진단방법에 비해 선명성이 좋지 않으나 WP-BK, VL-BH, VT-BH, FS-BL의 네 경우에는 정성적 진단방법에 비해 선명성이 향상되었다[4].

3. 다중이상의 진단결과

이상-결과트리모델에 의한 진단법에서는 9개 사례(5, 8, 9, 13, 21, 24, 27, 29, 33)가 가려진 다중이상으로 진단에 실패하였

표 6. 이중이상사례의 선명성 비교.

Table 6. The comparison of resolution for the selected double faults.

No.	Suggested method	Fault-effect tree
6	2.23(4)	5.38(6)
7	1.09(4)	4.87(10)
10	1.03(4)	8.42(10)
13	3.84(4)	failed
14	4.0(4)	5(5)
17	3.82(4)	3.86(4)
19	1.17(4)	6.75(7)
20	8(8)	10(10)
31	4(4)	4(4)

으나, 제안된 진단방법은 37개의 이중이상에 대한 진단에서 한 개의 사례(5)만 진단에 실패하여 정확성이 0으로 나왔고 나머지 36개 경우는 전체 진단구간에서 해를 놓치지 않아 정확성이 1이었다. 따라서, 정성적방법에 비해 훨씬 우수한 정확성을 보여주었다. RX-LK와 PP-BK의 5번 사례에서는 RX-LK만 진단하였고 PP-BK는 공정에 대한 영향이 약하여 증상이 감지되지 않아 진단에 실패하였다.

36개 사례 중에서 대부분의 경우에 강건성이 0이었으나 3개 사례(1, 15, 28)는 증상의 왜곡에 의해 일부 구간에서 각각 1개의 변수를 잘못 감지하였다. 표 6은 36개 사례 중에서 강건성이 1이거나 최악의 선명성이 2이하인 27개 사례를 제외한 9개 사례에 대해 이상-결과트리모델에 의한 진단과의 선명성을 비교하였다. 선명성도 제안된 방법이 이상-결과트리 모델에 비해 우수함을 보여준다. 선명성이 8로 가장 나쁜 사례 중 하나인 20번 사례에서는 FW(-)와 FR(-)의 잔차가 감지되었는데 각각 {VT-BL, WP-BK}와 {VR-BL, FS-BL, PUMP-EF, RP-BK}의 일차이상후보를 제시하면서 8개의 이중이상후보를 최종이상후보집합으로 제시하였기 때문이다.

15번 사례는 VR-BH와 TC-SVCL의 이중이상 사례로 CT(+), FR(+), TR(+)의 감지 외에 TR(+)가 110초부터 오감지되어 각 이상테이블에서 각각 {TC-SVCL}, {VR-BH, FS-BH}, {CW-TCH, HX-PL}이 제시되어 4개의 삼중이상이 최종이상후보로 제시된다. TR(+)은 이중이상의 복합적인 작용으로 측정값에 비해 추정값이 더 낮게 계산되어 감지되었다.

28번 사례는 FEED-FCH와 LS-BL이 동시에 발생한 사례로 L(-), FO(+), CA(-), T(+), CB(+), DF(+)와 감지에 의해 진짜 해들이 제시되다가 FEED-FCH에 의해 FO가 양의 방향으로, LS-BL에 의해 FP가 음의 방향으로 움직여 225초부터 735초까지 DF(+)가 감지되었다. 최종이상후보는 {LS-BL, FEED-FCH, RX-LK}의 삼중이상으로 이중이상의 진짜 해와 반응기누출의 가짜 해가 함께 제시되었다.

강건성이 0이 아닌 세 사례는 각각 1개의 잔차에 대한 CUSUM의 매개변수가 낮게 설정되어 거짓정보가 발생한 것이다. CUSUM의 각 매개변수를 2배로 하여 테스트해보았다. 그 결과, 오감지가 발생한 세 사례 모두에서 오감지의 문제를 해결하였다. CA의 감지가 500초 이상 늦어지는 경우가 있었고 CB와 T가 감지되지 않는 경우가 많아 대부분의 사례에서 이상진단이 5-30초 지연되었다. 그러나, 36개 사례에서 정확성과 선명성에 전혀 영향을 주지 않았다.

V. 결론

비선형성이 심하지 않은 화학공정을 위한 새로운 이상진단법을 제시하였다. 정성적 모델기반 진단방법과 다성분 통계모델을 혼합한 진단방법으로서, SDG의 부분적 인과관계에 구축된 DPLS 모델을 이용하여 진단을 수행한다. 제안된 방법을 이용하여 재순환이 있는 CSTR 공정에 대한 진단을 수행하여 그 성능을 확인하였다. 제안된 진단방법이 가진 장점은 다음과 같다.

- (1) SDG에 기반함으로써 시스템개발이 쉬우면서도 정성적 진단방법에 비해 선명성과 강건성이 향상되었다.
- (2) 정성적 방법에 비해 다중이상의 진단성능이 크게 개선되어 진단에 대한 신뢰성이 크게 향상되었다.
- (3) SDG의 부분적 인과관계를 이용함으로써 더욱 유연한 분산형진단이 가능하도록 하였다.
- (4) 정량적모델인 DPLS모델을 얻기 위해 필요한 학습용 데이터로서 이상데이터를 사용할 필요가 없어 실제공정에 대한 적용가능성을 높였다.

다중이상에 대한 진단사례 중에서 CUSUM의 매개변수에 따라 오진단이 발생하여 진단의 강건성이 저하되는 사례가 있었다. 오진단은 CUSUM의 매개변수, 즉 최소급등크기와 경계치를 크게 하여 막을 수 있으나 이 경우 이상진단이 늦어지는 시간지연의 문제점이 생길 수 있다. 따라서, 오진단과 시간지연의 두 문제점의 중요성에 따라 매개변수를 결정하여야 한다. 앞으로 실제 공정에 대한 적용을 통해 비선형성이 심한 공정에 대한 적용가능성을 높이고 CUSUM의 매개변수를 결정하는 최적의 방법을 개발할 예정이다.

참고문헌

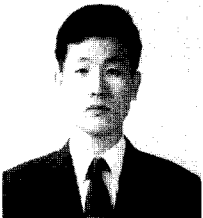
- [1] D. K. Lorenzo, *A Manager's Guide to Reducing Human Error*, Chemical Manufacturers Association, Washington. DC., 1990.
- [2] F. E. Finch, *Automated Fault Diagnosis of Chemical Process Plants Using Model-based Reasoning*, Sc. D. Thesis, Massachusetts Institute of Technology, 1989.
- [3] G. Lee and E. S. Yoon, "A process decomposition strategy for qualitative fault diagnosis of large-scale processes," *Ind. Eng. Chem. Res.*, Vol. 40, No. 11, pp. 2474-2484, May, 2001.
- [4] G. Lee, B. Lee, E. S. Yoon and C. Han, "Multiple fault diagnosis under uncertain conditions by the quantification of qualitative relation," *Ind. Eng. Chem. Res.*, Vol. 38, No. 3, pp. 988-998, March, 1999.
- [5] V. Venkatasubramanian, "Managing abnormal situations in process plants: the next control frontier," *Proc. of the PSE Asia 2000*, pp. 597-616, Dec., 2000.
- [6] M. A. Kramer and B. L. Palowitch, "A rule-based approach to fault diagnosis using the signed directed Graph," *AIChE J.*, Vol. 33, No. 7, pp. 1067-1078, Jul., 1987.
- [7] T. Sorsa and H. N. Koivo, "Neural networks in process fault diagnosis," *IEEE Trans. Systems. Man. Cybern.*, Vol. 21, No. 4, pp. 815-825, Jul., 1991.
- [8] M. Iri, K. Aoki and E. O'Shima, "An algorithm for diagnosis of system failures in the chemical process," *Comput. Chem. Eng.*, Vol. 3, No. 4, pp. 489-493, 1979.
- [9] P. Geladi and B. R. Kowalski, "Partial least-squares regression: a tutorial," *Anal. Chim. Acta*, Vol. 195, pp. 1-17, 1986.
- [10] W. Ku, R. H. Storer and C. Georgakis, "Disturbance detection

and isolation by dynamic principal component analysis," *Chemometrics Intell. Lab. Syst.*, Vol. 30, pp. 179-196, 1995.

[11] 최용진, 실시간 공정데이터의 정성적 해석을 위한 Scale-space 필터링기법의 응용과 구현에 관한 연구, 박사학위

논문, 서울대학교 화학공학과, 1995.

[12] 이기백, 이상-결과트리모델을 이용한 공정이상진단시스템에 관한 연구, 박사학위논문, 서울대학교 화학공학과, 1997.



이 기 백

1968년 7월 4일생. 1991년 서울대학교 화학공학과(공학사). 1993년 서울대학교 화학공학과(공학석사). 1997년 서울대학교 화학공학과(공학박사). 1998년~현재 충주대학교 공업화학과 조교수. 관심분야는 공정이상진단, 공정시뮬레이션.



윤 인 섭

1949년 6월 4일생. 1971년 서울대학교 화학공학과(공학사). 1982년 MIT 화학공학과(공학박사). 1983년~현재 서울대학교 응용화학부 교수. 관심분야는 공정설계 및 제어, 공정최적화, 이상진단



신 동 일

1965년 3월 24일생. 1987년 서울대학교 화학공학과(공학사). 1989년 서울대학교 화학공학과(공학석사). 1997년 Purdue University 화학공학과(공학박사). 2002년~현재 명지대학교 화학공학과 조교수. 관심분야는 지능시스템.