

PCA-기반 고장 진단 시스템 설계에 관한 연구

이 영 삼\*, 김 성 호\*, 이 기 상\*\*

\* 군산대학교 전자정보공학부, \*\* 단국대학교 전기공학과

A study on the design of fault diagnostic system based on PCA

Young-Sam Lee\*, Sung-Ho Kim\*, Kee-Sang Lee\*\*

\*School of Electronic and Information Eng., Kunsan National, \*\*Department of Electrical Eng., DanKook Univ.

**Abstract** - PCA(Principle Component Analysis) has emerged as a useful tool for process monitoring and fault diagnosis. The general approach requires the user to identify the root cause by interpreting the residual or principle components. This could be tedious and often impossible for a large process. In this paper, PCA scheme is combined with the FCM-based fault diagnostic algorithm to enhance the diagnostic results. The implementation of the PCA-FCM based fault diagnostic system is done and its application is illustrated on the two-tank system.

1. 서 론

장치고장 및 시스템 파라미터의 변화 등으로 인해 발생된 계통의 고장은 일반적으로 센서측정값의 변화로 나타나게 되며 이러한 고장은 자주는 아니지만 경제적 또는 안전도 측면에 상당한 영향을 미친다. 특히 미국의 대규모 화공플랜트의 경우, 발생된 고장의 적절치 못한 처리로 인해 약 200억 달러의 손실을 입고 있다는 점을 고려한다면 효율적인 ASM(Abnormal Situation Management)의 개발이 절실히 요구되고 있다. 효율적인 ASM의 설계를 위해서는 고장 원인을 신속히 찾기 위한 많은 정보들이 요구되고 있으며 현재 이들 정보를 기반으로 하는 조기 고장검출 및 진단에 관한 연구가 활발히 진행되고 있다.

일반적으로 화공프로세스의 경우, 정상동작과 관련된 상당량의 데이터 취득이 가능하며 따라서 이들 데이터를 기반으로 하는 통계학적 기법이 조기 고장검출 및 진단에 널리 사용되고 있다. 이러한 통계학적 기법은 크게 단변량 기법과 다변량 기법으로 분류된다. 단변량 기법은 구현하기 쉽다는 장점을 갖지만 PCA(Principle Component Analysis) 및 PLS(Partial Least Square) 등과 같은 다변량 통계학적 기법이 변수들간에 존재하는 상관관계를 사용한다는 점에서 보다 효율적이다. PCA 및 PLS에 기반한 고장검출 시스템은 PCA/PLS 모델의 예측값과 실제 측정값간의 잔차에 대한 통계학적 변이를 기반으로 구현된다. 최근 Gregor 및 Dunia 등은 PCA/PLS에 기반한 고장 진단 시스템을 제안하고 그 유용성을 입증한 바 있다[1-2]. 그러나 이들 연구 대부분은 고장의 검출은 통계학적 기법에 의해 수행하지만 고장 원인의 판별은 수계산에 의해 수행되고 있어 대규모 계통에의 직접적인 적용은 어렵다는 적용상의 한계를 갖는다. 따라서 대규모 계통의 고장진단을 위한 자동화된 진단 기법의 도입이 요구된다.

고장진단 기법은 크게 수학적 모델에 기반한 정량적 고장진단 기법과 수학적 모델을 얻기 힘든 계통에 적용가능한 정성적 고장진단 기법으로 구분된다. 이중 정성적 고장진단 기법은 80년대 중반 이후 활발히 연구되기 시작하였으며, 그 대표적인 예로 고장트리(Fault tree)기법, 유향부호 그래프(Signed Directed Graph: SDG)기법 등을 들 수 있다[3]. 또한 Lee 등은 정성적 모델에 기반한

진단기법의 하나로 Kosko에 의해 제안된 FCM(Fuzzy Cognitive Map)이 SDG의 일반화된 표현이라는 점에 착안하여 간단한 행렬 연산에 의해 효율적인 고장진단이 가능한 FCM-기반 고장진단 알고리즘을 제안하고 이의 유용성을 확인한 바 있다. 또한 FCM-기반 고장진단 알고리즘은 FCM의 연결행렬로부터 도출되는 고장 FCM 모델의 사용으로 인해 실시간 패턴매칭을 통한 조기 고장진단이 가능하다는 특성을 갖는다[4-5].

본 연구에서는 다변수 계통에 직접 적용이 가능한 고장검출 및 진단 시스템을 제안하고자 한다. 제안된 시스템은 고장 검출 메카니즘으로 PCA 기법을 사용하며 고장진단 기법으로는 FCM-기반 진단 기법을 사용한다. 또한 제안된 고장진단 시스템의 유용성 확인을 위해 two-tank 시스템에 대한 시뮬레이션 고찰을 수행하고자 한다. 본 논문의 구성은 다음과 같다. 2장에서는 PCA의 기본 개념에 대하여 기술하며 3장에서는 FCM-기반 고장검출 및 진단 알고리즘에 대해 설명한 후, 4장에서는 본 연구에서 제안된 고장 진단시스템 및 이의 Two-tank 시스템에 대한 적용에 대해 고찰하며 5장에서는 결론을 기술하고자 한다.

2. 주성분 분석(Principle Component Analysis)

주성분 분석은 변수들의 선형 결합을 통해 변수들 간에 존재하는 중요한 변량의 방향을 결정하는 통계학적 모델링 기법으로 상관관계가 있는 변수들 사이의 복잡한 관계를 단순화하기 위하여 사용된다. 즉, 상관관계가 있는 변수들을 결합하여 서로 독립적인 새로운 저차의 새로운 변수들을 생성하는 것으로 생각할 수 있으며 이때 새로이 생성된 변수를 주성분이라 부른다. PCA의 기본 개념에 대해 설명하면 다음과 같다.

$n$  개의 변수들에 대해  $m$ -회 측정으로 얻어진 데이터 행렬을  $X(m \times n)$ 라고 정의하고  $X$ 는 0 인 평균과 단위 분산을 갖는다고 가정한다. 이 경우  $X$ 의 공분산 행렬은 다음과 같이 정의된다.

$$cov(X) = \frac{X^T X}{m-1} \quad (1)$$

PCA는 행렬  $X$ 를 다음과 같이 분해할 수 있으며

$$X = \hat{X} + E \quad (2)$$

여기서

$$\hat{X} = \sum_{i=1}^r t_i p_i \quad r \leq n \quad (3)$$

식(2)에서  $E$  는 PCA 모델의 예측 편차를 나타낸다. 식(3)의 scores로 불리는 벡터  $t_i$  는 프로세스에 존재하는 잡음에 대한 정보를 포함하며 벡터  $p_i$  는 변수들이 어떻게 관계되어 있는지에 대한 정보를 나타낸다. 벡터  $p_i$  는 다음과 같이 공분산 행렬의 고유벡터이기도 하다.

$$\text{cov}(X) P_i = \lambda_i P_i \quad (4)$$

식(3)에서 사용되는  $r$ -개의 고유벡터( $p_i$ )는 공분산 행렬에 대한  $n$ -개 고유값중 가장 큰 고유값에서부터  $r$ -개까지의 고유값에 해당하는 고유벡터를 의미한다. 또한 각 고유벡터에 대한 가중치 벡터  $t_i$  는 다음과 같이 구할 수 있다.

$$t_i = X p_i \quad (5)$$

식(3)으로 표현되는 PCA 모델을 이용한다면 새로운 측정 벡터  $x(1 \times n)$ 에 대한 PCA 모델링 편차  $e$ 는 다음과 같이 구할 수 있으며

$$e = (x - \hat{x}) \quad (6)$$

새로운 측정 벡터  $x$  에 대한 잔차  $R$ 은 다음과 같이 구할 수 있다.

$$R = e e^T \quad (7)$$

상기 식들을 이용한다면 프로세스의 정상동작에 대한 PCA 모델을 구하는 것이 가능하다. 일반적으로 프로세스에 고장이 발생할 경우 식(7)로 표현되는 잔차  $R$ 은 고장발생 정보를 포함하며 따라서 식(7)에 대한 임계치 검사를 통해 효율적인 고장 검출이 가능하게 된다.

### 3. FCM-기반 고장진단 알고리즘의 개요

Lee 등은 FCM이 SDG와 구조적으로 동일하다는 점이 외에도 FCM만이 갖는 장점 즉, TAM(Temporal Associative Memory) recall 특성을 갖는다는 점을 이용하여 Iri 등이 제안한 SDG 기반 진단기법을 FCM 관점에서 재구성한 FCM-기반 고장진단 알고리즘을 제안하고 이의 유용성을 확인한 바 있다[5]. 본 절에서는 FCM-기반 진단 알고리즘에서 사용되는 정성적 관측 벡터를 생성하는 방법과 FCM-기반 진단 알고리즘의 개요에 대해 설명한다.

#### 3.1 정성적 관측 벡터의 생성

진단의 첫 단계는 관측패턴의 생성이다. 관측패턴 벡터  $W$  내의 요소  $w_i$  는  $\{-1, 0, 1\}$  중의 한 값을 가지며 이때 "1"은 정상치로부터의 증가, "0"은 정상, "-1"은 정상치로부터의 감소를 의미한다.  $w_i$  의 결정을 위해 먼저 프로세스 상태가 정상동작대역(Normal Operation Band : NOB)에서 벗어난 정도를 나타내는 편향 지표(Deviation Index : DI)를 식(8)와 같이 정의한다.

$$DI_i = \frac{\text{Measured Value} - \text{NOB}}{0.5 \cdot \text{NOB}} \quad (8)$$

여기에서 첨자  $i$  는  $i$ -번째 요소를 나타내며 시스템이 정

상일 경우  $DI_i$  는  $[-1, 1]$ 인 값을 갖는다. 고장패턴의 요소  $w_i$  는  $DI_i$ 에 대한 다음의 검사를 통해 얻어진다.

$$w_i = \begin{cases} 1, & DI_i \geq 1 \\ 0, & |DI_i| < 1 \\ -1, & DI_i \leq -1 \end{cases} \quad (9)$$

### 3.2 FCM-기반 고장진단 알고리즘

FCM-기반 진단기법은 계통으로부터의 관측패턴 벡터와 계통에 대한 FCM모델을 이용하여 고장의 영향을 받는 상태간의 인과 관계를 추출하는 과정 및 그 인과관계의 출발점, 즉 계통에 발생된 고장의 원인을 식별하는 과정으로 구성된다.

계통에 대한 FCM 행렬,  $E$  및 고장의 전파가 완료된 후의 관측패턴 벡터  $W$ 가 주어진다면 기본 FCM-기반 진단 알고리즘은 다음과 같다.

Step 1. CR 매트릭스의 계산

고장의 영향을 받는 상태간의 인과관계 행렬은 다음의 두가지 연산에 의해 구해진다.

$$WE = \text{Diag}(W) \cdot E \cdot \text{Diag}(W) \quad (10)$$

$$CR(i, j) = T(WE(i, j)) \quad (11)$$

여기서  $T$ 는 임계 함수를 의미하며 inconsistent한 branch를 제거하는 역할을 하며, 임계치(Th)는 "0"으로 선정한다.

인과관계 행렬  $CR$ 은 단지 고장 요소의 식별에만 관계하므로  $CR(i, j)$ 는 다음과 같이 정의될 수 있다.

$$CR(i, j) = \begin{cases} 1, & WE(i, j) \geq 0 \\ 0, & WE(i, j) < 0 \end{cases} \quad (12)$$

Step 2. 고장원인의 판별

Iri 등에 의해 제안된 consistent rooted tree 기법을 식(11)로 표현되는  $CR$ 행렬에 적용한다면 고장 원인의 판별은 간단히 수행될 수 있다. 고장원인의 판별과정은 다음과 같다.

$CR(i, j)$  행렬의 각 변수에 대해

- ① 영향을 받는 변수의 수(Indegree)를 계산
- ② 영향을 주는 변수의 수(Outdegree)를 계산
- ③  $Indegree(j)=0$  이고  $Outdegree(i) \geq 1$  인  $I$  번째절점에 대응하는 변수가 고장 원인

위 과정에 대한 알고리즘은 다음과 같다.

$CR(i, j)$ ,  $1 \leq (i, j) \leq m$ 인 모든 요소에 대해

Initial:

$OUT(i)=0, IN(j)=0;$

Loop:

for( $i=1$  to  $n$ ) {for( $j=1$  to  $n$ ) {

if( $CR(i, j)=1$ ) then  $OUT(i) = OUT(i)+1;$

}}

for( $j=1$  to  $n$ ) {for( $i=1$  to  $n$ ) {

if( $CR(i, j)=1$ ) then  $IN(j)=IN(j)+1;$

}}

Diagnosis:

for( $k=1$  to  $n$ ) {

if( $OUT(k) \geq 1$  and  $IN(k) = 0$ )  
 then  $k$ 'th node is the origin of the fault;  
 if(Number of the failure source  $\geq 2$ )  
 then Basic FDI fails;

#### 4. 제안된 PCA-기반 고장진단 시스템의 구조

본 연구에서 제안된 PCA를 이용한 FCM-기반 고장검출 및 진단 시스템의 구성도를 나타내면 그림 1과 같다. 그림에서 고장검출부는 식(7)로 표현되는 PCA 모델링 잔차( $R$ )를 출력하며 잔차  $R$ 에 대한 임계치 검사를 통해 고장 검출을 수행한다. 일단 고장이 검출되면 다음 단의 FCM-기반 진단 알고리즘의 관측벡터발생부에 측정 데이터  $x$ 가 입력되고 식(8),(9)를 이용하여 관측 패턴 벡터를 생성하게 된다. 이를 기반으로 하여 3.2절에서 설명한 FCM-기반 고장진단 알고리즘이 수행된다.

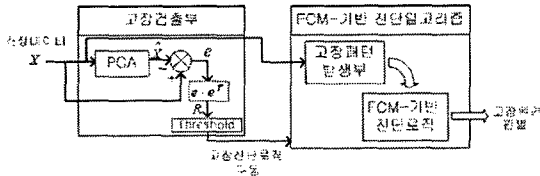


그림 1. PCA를 이용한 FCM-기반 고장검출 및 진단 시스템의 구성도

Fig. 1. FCM-base fault detection and schematic diagram of diagnosis system that use PCA

#### 4.1 PCA-기반 고장진단 시스템의 two-tank 시스템에의 적용

본 연구에서 제안한 고장검출 및 진단 시스템의 유용성을 확인하기 위해 그림 2와 같은 Two-tank 시스템에 적용하고자 한다. 그림 2(b)는 two-tank 시스템에 대한 SDG이며 이로부터 FCM 연결행렬  $E$ 를 구하면 다음과 같다.

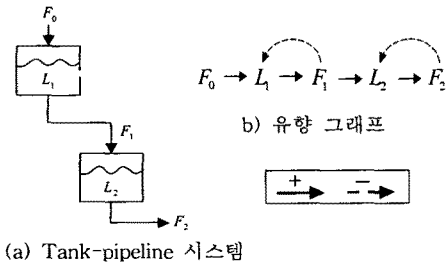


그림 2. Two-tank 시스템  
 Fig. 2. Two-tank system

$$E = \begin{matrix} & F_0 & L_1 & F_1 & L_2 & F_2 \\ \begin{matrix} F_0 \\ L_1 \\ F_1 \\ L_2 \\ F_2 \end{matrix} & \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -1 & 0 \end{bmatrix} \end{matrix}$$

상기 시스템에서 발생할 수 있는 이상은 다음과 같이 4가지로 가정하였다.

- Fault 1 : Blockage of the upper tank
- Fault 2 : Leakage of the upper tank
- Fault 3 : Blockage of the lower tank
- Fault 4 : Leakage of the lower tank

그리고 고장검출을 위한 PCA를 위해 그림 3과 같은 약 500개의 정상동작상태 데이터를 취득하여 데이터의 주성분을 파악하기 위해 식(1)을 이용하여 공분산을 계산하였다.

$$cov(X) = \begin{bmatrix} 3.34e-5 & -2.40e-6 & 6.06e-7 & 7.73e-6 & 1.75e-6 \\ -2.40e-6 & 0.02446 & 0.00367 & 0.0249 & 0.00350 \\ 6.06e-7 & 0.00367 & 0.00064 & 0.00468 & 0.0007 \\ 7.73e-6 & 0.02490 & 0.00468 & 0.0355 & 0.00540 \\ 1.75e-6 & 0.00350 & 0.00070 & 0.0054 & 0.00086 \end{bmatrix}$$

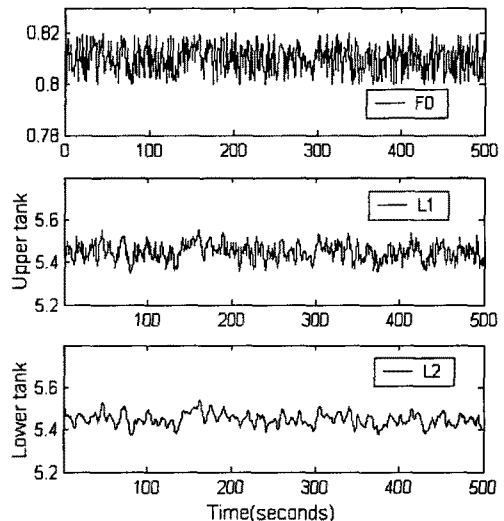
여기서 얻어진 공분산의 고유벡터  $V$ 와 고유값  $D$ 를 나타내면 다음과 같다.

$$V = \begin{bmatrix} 0.002393 & 0.20227 & 0.97933 & 0.0015033 & 8.351e-5 \\ -0.065317 & -0.021406 & 0.0057315 & -0.7839 & 0.61704 \\ 0.94416 & -0.30569 & 0.0608 & 0.013549 & 0.10599 \\ -0.029793 & 0.19402 & -0.040994 & 0.60403 & 0.77132 \\ -0.32156 & -0.90969 & 0.18844 & 0.14306 & 0.1144 \end{bmatrix}$$

$$D = \begin{bmatrix} 6.753e-6 & 0 & 0 & 0 & 0 \\ 0 & 3.232e-5 & 0 & 0 & 0 \\ 0 & 0 & 3.34e-5 & 0 & 0 \\ 0 & 0 & 0 & 0.0045714 & 0 \\ 0 & 0 & 0 & 0 & 0.056876 \end{bmatrix}$$

식(3)에서  $p_i$ 는 공분산의 고유값  $D$ 에 나타난 데이터 사이의 상관관계가 큰 즉,  $r$ 개의 가장 큰 고유값에 일치하는 고유벡터로부터 구해지며 시뮬레이션을 위해 다음과 같은  $r=4$ 개의 주성분을 사용하였다.

$$P = \begin{bmatrix} 0.20227 & 0.97933 & 0.0015033 & 8.351e-5 \\ -0.021406 & 0.0057315 & -0.7839 & 0.61704 \\ -0.30569 & 0.0608 & 0.013549 & 0.10599 \\ 0.19402 & -0.040994 & 0.60403 & 0.77132 \\ -0.90969 & 0.18844 & 0.14306 & 0.1144 \end{bmatrix}$$



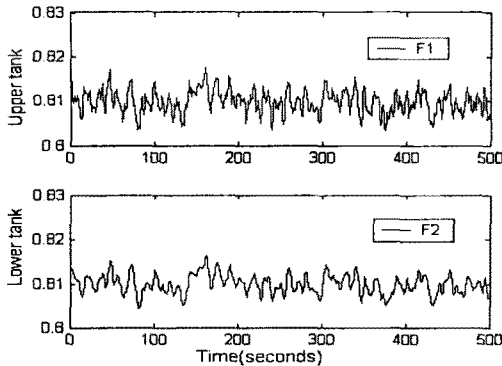


그림 3. 정상운전 데이터  
Fig. 3. Normal operating data

#### 4.2 시뮬레이션 고찰

제안된 고장진단 시스템의 유용성 확인을 위해 Matlab의 simulink를 사용하였으며 구성된 시뮬링크 블록도를 나타내면 그림 4와 같다.

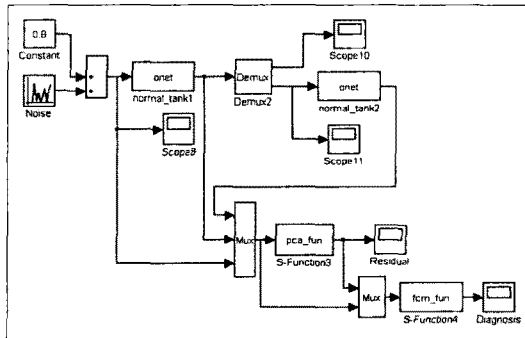


그림 4. 시뮬레이션을 위한 Simulink block diagram  
Fig. 4. Simulink block diagram for simulation

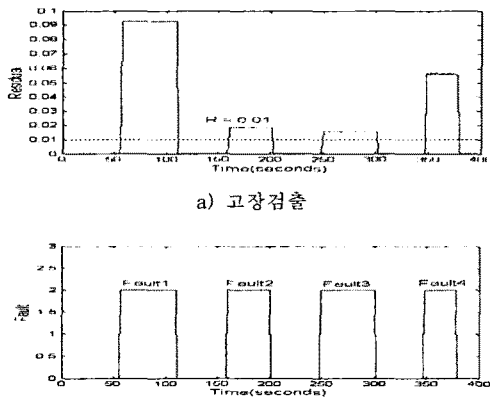


그림 5. 고장 검출 및 진단 그래프  
Fig. 5. Fault detection and diagnostic graph

Two-tank 시스템에 대한 고장 검출 및 진단관련 응답 특성을 나타내면 그림 5와 같다. 그림 5(a)는 가정된 4개의 고장이 각각 발생한 경우의 PCA 모델과 실제 센서 간의 잔차를 나타낸다. 그림으로부터 알 수 있듯이 식(7)로부터 얻어지는 잔차는 0이 아닌 값을 갖게 되며 따라서 문턱값(R)에 의해 효과적으로 고장이 검출됨을 알 수 있다. 본 연구에서는 고장 검출의 신뢰도에 큰 영향을 미치는 요소인 R을 0.01로 선정하였다. 따라서 PCA와 실제 계측데이터간의 잔차가 0.01을 초과할 때 고장이 발생되었다고 판정하게 된다. 또한 그림 5(b)는 PCA에 의해 고장검출이 완료된 후 수행되는 FCM-기반 진단 알고리즘의 응답특성으로 실제 계통으로부터 얻어진 관측패턴 벡터와 FCM 모델을 이용하여 고장 원인의 판별을 수행하며 그림으로부터 정확한 고장의 진단이 가능함을 알 수 있다.

#### 5. 결 론

본 연구에서는 다변량 통계학적 처리기법 중 하나인 PCA 기법과 Lcc 등에 의해 제안된 FCM-기반 기법을 병합한 PCA를 이용한 FCM-기반 고장 검출 및 진단 시스템을 제안하고 two-tank 시스템에 대한 시뮬레이션 고찰을 통해 제안된 진단 기법의 유용성을 확인하였다. 시뮬레이션 결과 가정된 각각의 고장에 대한 정확한 고장의 검출 및 진단이 가능함을 확인할 수 있었다. 그러나 본 연구에서 고찰한 two-tank 시스템은 측정변수가 한정되어 있어 제안된 진단기법의 유용성을 확인하기에는 적합하지 않았으며 따라서 보다 많은 측정 변수를 갖는 대규모 시스템에의 적용을 통한 제안된 시스템의 유용성 검증이 예상된다. 또한 PCA와 실제 계측데이터간의 잔차 검사에 사용되는 임계치값은 시뮬레이션시 사용한 노이즈 정보를 기반으로 선정하였으나 실제 시스템에의 적용을 위해서는 miss alarm 및 false alarm과 밀접한 관계가 있는 임계값의 체계적인 선정 방법에 대한 연구가 요구된다.

#### [참 고 문 헌]

- [1] MacGregor, K. Costas, "Process monitoring and diagnosis by multiblock PLS methods", *AIChE J.*, Vol.40, No.5, pp. 826-838, 1994
- [2] Dunia, R., S.J. Qin, "Identification of faulty sensors using principle component analysis", *AIChE J.*, Vol.42, No.10, pp. 2797-2812, 1996
- [3] J. Shiozaki, and H. Matsuyama, "An Improved Algorithm for Diagnosis of System Failure in Chemical Processes", *Comput. Chem. Eng.*, Vol. 9, 1985
- [4] B. Kosko, "Neural Networks and Fuzzy Systems", A Dynamical Systems Approach to Machine Intelligence, Prentice Hall, 1992
- [5] K.S. Lee, S.H. Kim and N. Sakawa, "On Line Fault Diagnosis by using Fuzzy Cognitive Map", *IEICE Trans. on Fundamentals*, Vol. E-79-A, no. 6, 1996
- [6] H. Vedam, and V. Venkatasubramanian, "Automated Interpretation of PCA-based Process Monitoring and Fault Diagnosis using Digraphs",