LOF를 이용한 ICA 기반 통계적 공정관리의 성능 개선 방법론*

이재신** · 강복영**† · 강석호**

The Use of Local Outlier Factor(LOF) for Improving Performance of Independent Component Analysis(ICA) based Statistical Process Control(SPC)

Jaeshin Lee** · Bokyoung Kang** · Suk-Ho Kang**

■ Abstract ■

Process monitoring has been emphasized for the monitoring of complex system such as chemical processing industries to achieve the efficiency enhancement, quality management, safety improvement. Recently, ICA (Independent Component Analysis) based MSPC (Multivariate Statistical Process Control) was widely used in process monitoring approaches. Moreover, DICA (Dynamic ICA) has been introduced to consider the system dynamics. However, the existing approaches show the limitation that their performances are strongly dependent on the statistical distributions of control variables. To improve the limitation, we propose a novel approach for process monitoring by integrating DICA and LOF (Local Outlier Factor). In this paper, we aim to improve the fault detection rate with the proposed method. LOF detects local outliers by using density of surrounding space so that its performance is regardless of data distribution. Therefore, the proposed method not only can consider the system dynamics but can also assure robust performance regardless of the statistical distributions of control variables. Comparison experiments were conducted on the widely used benchmark dataset, Tennessee Eastman process (TE process), and showed the improved performance than existing approaches.

Keyword : Statistical Process Control(SPC), Local Outlier Factor(LOF), Independent Component Analysis(ICA), Process Monitoring, Fault Detection, Tennessee Eastman process

논문접수일 : 2010년 11월 04일 논문게재확정일 : 2011년 02월 14일
논문수정일(1차 : 2011년 02월 10일)
* 본 논문은 2010년도 한국경영과학회 추계학술대회 경쟁부문에 제출하여 우수논문상을 수상한 논문임. 논문의 내용을 일부 보완, 확장한 논문이며, 소정의 심사과정을 거쳐 게재 추천되었음.
이 논문은 2009년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2009-0080628).
** 서울대학교 산업공학과
† 교신저자

1. 서 론

최근 30년 동안 안전성, 안정성, 지속가능성, 경쟁력에 대한 요구가 증가되어 왔다. 특히 항공, 원자력발전, 화학공정 산업 등의 안전에 치명적인 산업에서는 작은 이상 현상(fault)이 전체 시스템의 붕괴를 야기할 수 있다[34]. 하지만 공정의 대형화 및 복잡화에 따른 데이터의 증가로, 사람이 실시간으로 이상 현상을 검출하기에는 여러 문제점이 발생한다[30]. Venkatasubramanian의 연구에 따르면 70%의 산업 사고가 사람의 오류에 의해 발생하였다[30]. 그리고 Wang et al.에 따르면 미국 석유 생산 산업에서 이상 현상으로 인해 연 3%~8% 생산 감소가 야기되었고, 이는 200억 달러의 손실로 이어졌다[31]. 이런 문제점들을 해결하기 위해 실시간 이상 현상 검출을 수행하는 공정 모니터링 시스템의 필요성이 제기되었다. 공정 모니터링을 통해 공장 규모의(plant-wide) 효율성, 품질, 안전성 개선 등을 성취할 수 있다[12]. 그러한 이유로, 이상 현상 검출을 위한 공정 모니터링에 대한 연구가 활발히 진행되어 왔다.

MSPC(Multivariate Statistical Process Control)는 이상 현상 검출을 위한 프로세스 모니터링에서 널리 사용되는 통계적 방법론이다[31]. MSPC는 모니터링 대상 공정에 대한 관리 변수들의 관측치들을 얻어서, 그것을 이용해 각 모니터링 시점에 모니터링 통계량(statistic)의 값을 계산한다. 만약 계산된 통계량의 값이 관리 한계보다 클 경우, 공정은 현 시점에 'out of control', 즉 이상 현상으로 판별된다. 그렇지 않을 경우, 공정은 'in control'로 판별된다. 대용량의 공정 변수 데이터의 차원을 줄이고 의미있는 변수들만을 추출하기 위하여, MSPC는 PCA(Principal Component Analysis)나 ICA(Independent Component Analysis)와 통합되어 수행된다. PCA 기반 MSPC는 모니터링 관측치들을 주성분들로 변환한 다음 T^2, SPE 등의 통계량들을 이용해 공정을 모니터링 한다. Kresta et al.은 MSPC의 도구로써의 PCA와 PLS(Partial Least Squares)의 효율성을 보였다[20]. 최근, 비선형 최적화가 필요 없는 KPCA (Kernel PCA) 방법론이 제안되었다[6, 23, 32]. 그러나 PCA 기반 MSPC는 데이터 분포에 대한 가정의 부적합성 때문에 불가피하게 제한된 성능을 보인다. PCA는 도출 변수들이 정규분포를 따른다고 가정하지만[15], 현업의 산업 공정에서 PCA 도출 변수들은 거의 정규분포를 따르지 않는다[27].

이를 개선하기 위해 ICA 기반 MSPC 방법론들이 제안되었다[17, 31]. ICA는 선형적으로 혼합된 신호들을 독립적인 신호들로 변환한다[16]. ICA 기반 MSPC는 도출된 독립요소들을 이용해서 I^2, SPE, I^2_e 등의 통계량을 계산한다. Kano et al.은 ICA를 MSPC 공정 모니터링에 도입하였다[19]. 최근 dynamic ICA [24], modified ICA[22] 등의 ICA를 이용한 공정 모니터링 방법론들이 기본 ICA 모델의 추출 독립요소들의 수 및 순서 결정의 모호함 등의 문제점을 해결하기 위해 제안되었다. Kernel ICA는 비선형 공정들을 다루기 위해 제안되었다[26]. 비가우시안 공정(non-Gaussian process)의 경우, ICA 기반 MSPC 방법론들이 PCA 기반 MSPC 방법론들에 비해 더 의미있는 정보를 얻어낼 수 있다[25]. ICA 기반 방법론들은 관측 변수들이 시간의 흐름에 독립적이라고 가정한다[16]. 하지만 이 가정은 실제 공정의 자기상관(autocorrelation)을 가지는 역동적 특성 때문에 유효하지 못하다[13]. 예를 들면 어떤 시점의 공정 상태의 관측치는 바로 이전의 관측치와 관계를 가지는 값으로 정해진다는 것이다. 실제 공정의 자기상관을 고려하기 위해서 Lee et al.에 의해 DICA (Dynamic ICA)가 제안되었다[24]. DICA는 시간지연 변수 l만큼의 관측치를 서로 연결하여 훈련 데이터셋을 구성하는 방법이다. 이 행렬에 ICA를 수행하게 되면 역동성이 제거된 독립요소 값을 도출할 수 있다[24]. 하지만 DICA 방법론 또한, ICA 도출 변수들의 통계적 분포에 따라 성능이 좌우된다. 정리하면, 기존 ICA 기반 방법론들은 다음과 같은 한계점을 갖는다. 첫째, 자기상관을 고려하지 못한다. 둘째, 도출변수의 분포에 대한 가정이 현실적이지 않아, 통계적 분포에 의존적인 성능을 보인다.

이러한 한계점들을 해결하여 이상 현상 검출 성

능을 개선하기 위해, 본 논문에서는 DICA와 LOF (Local Outlier Factor)를 통합한 새로운 모니터링 방법론을 제안한다. LOF는 밀도 기반 이상치 검출 기법으로 변수들의 분포에 상관없이 이상도를 정량적 수치로 제공한다. 따라서 잠재 변수들이 가우시안 혼합 분포나 감마 분포 등 어떤 분포를 따르는가에 상관없이 비선형 이상 현상 판별 경계를 정확하게 구할 수 있고, 안정적인 성능을 제공한다. 또한 실제 공정의 성질에 맞게 자기상관을 고려하여 기존 방법론들에 비해 이상 현상 검출 능력이 높다. 자기상관을 고려하지 않은 방법론들은 현 시점의 공정 관리 변수에 자기상관이 남아있기 때문에 ICA의 시간에 따른 독립성에 대한 가정이 위배되어, 도출된 독립요소들이 부정확해지고 공정에 대한 설명력이 감소된다. 그에 반해 제안 방법론은 자기상관이 제거되기 때문에 도출된 독립요소들의 설명력이 증가되어 이상 현상 검출의 정확도를 높일 수 있다.

본 논문에서는 제안된 방법론의 성능을 입증하기 위하여 관련 연구들에서 널리 사용된 Tennessee Eastman process(TE process) 데이터를 이용하여 기존 방법론들과 비교실험을 수행하였다. 실험결과 제안된 방법론은 기존 방법론들에 대해 주목할 만한 성능 개선을 이루었음을 확인할 수 있었다. 또한 LOF의 실시간 모니터링에의 적용을 위해 근사 LOF 계산 알고리즘을 제안하고 사용하였다.

2. 관련 연구

본 장에서는 ICA의 개념과 ICA 기반 공정 모니터링 연구 사례들을 설명한다.

2.1 ICA

ICA는 선형적으로 혼합된 실제 관측 변수들에서 독립 잠재 변수들을 추출한다. [그림 1]은 ICA의 적용 예시를 보여주고 있다.

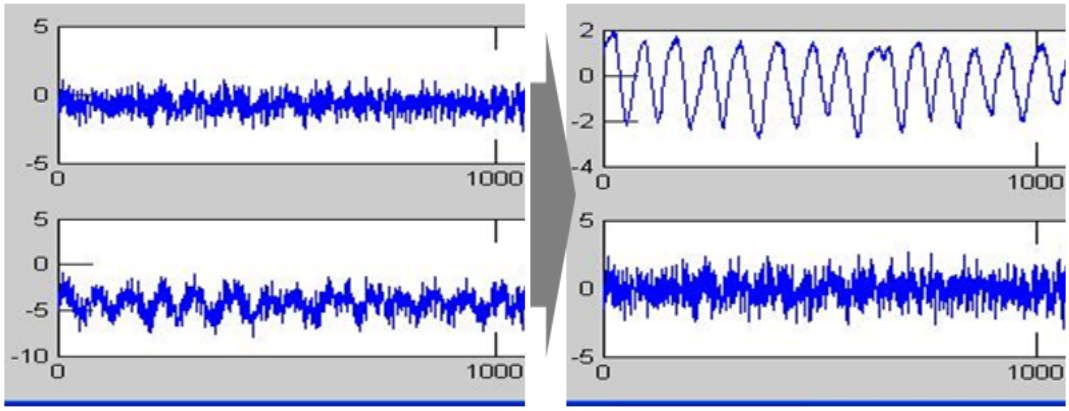
혼합된 변수들과 본래의 잠재 변수들을 각각 $\mathbf{X}$와 $\mathbf{S}$로 표현하면, 벡터-행렬 기호정의에 따라 다음의 식으로 나타낼 수 있다.

$$\mathbf{X} = \mathbf{AS} = \mathbf{A}\begin{bmatrix} \mathbf{s}^T(1) \\ \mathbf{s}^T(N) \end{bmatrix} \tag{1}$$

본래의 독립 변수들 $\mathbf{s}^T(1)$, …, $\mathbf{s}^T(N)$은 de-mixing 행렬 $\mathbf{W}$를 계산함으로써 추정될 수 있고, $\mathbf{W}$는 $\mathbf{A}$의 역행렬이다.

$$\hat{\mathbf{S}} = \mathbf{WX} = \begin{bmatrix} \hat{\mathbf{s}}^T(1) \\ \hat{\mathbf{s}}^T(N) \end{bmatrix} \approx \mathbf{S} \tag{2}$$

ICA를 수행하기 위해서 비정규성(non-Gaussianity)의 최대화, 상호정보(mutual information)의 최



[그림 1] ICA를 이용해 혼재된 관측 신호 데이터에서 독립성분 신호를 추출해 낸 결과

소화, 최대우도추정 등에 기반한 알고리즘들이 제안되었다[16]. 본 논문에서 사용된 FastICA 알고리즘은 비정규성의 최대화에 기반한 알고리즘으로 neg-entropy의 최대화를 통해 비정규성을 최대화한다. FastICA 알고리즘은 빠른 계산 대비 우수한 정확도의 특징을 가지기 때문에 관련 분야에서 가장 널리 사용되고 있다. 알고리즘의 세부내용은 Hyvarinen의 연구에서 찾아볼 수 있다[16].

2.2 ICA 기반 공정 모니터링

ICA는 공장 규모의 공정 모니터링 분야에 차원감축을 위해 도입되었다. Kano et al.은 ICA를 MSPC 공정 모니터링에 처음 도입하였고[19], 뒤이어 Lee et al.은 ICA 기반 MSPC 방법론을 위한 세 가지 통계량 (I^2, SPE, I^2_e)을 제안하였다[25]. $\mathbf{W}$의 각 행에 대해 계수의 합의 내림차순에 따라 몇 개의 행을 선택함에 따라서, $\mathbf{W}$는 지배적 부분 $\mathbf{W}_d$ 와 비지배적 부분 $\mathbf{W}_e$이 분리된다. I^2는 공정 변동성의 시스템적 부분을 설명하고, $\hat{\mathbf{s}}_d = \mathbf{W}_d\mathbf{x}$ 일 때 아래와 같이 정의된다.

$$I^2 = \hat{\mathbf{s}}_d^T\hat{\mathbf{s}}_d \tag{3}$$

SPE(squared prediction error)는 공정 변동성의 비시스템적 부분을 모니터링 하는데 사용되고, $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}$일 때 아래와 같이 정의된다.

$$SPE = \mathbf{e}^T\mathbf{e} \tag{4}$$

I^2_e는 비지배적 부분에 의한 공정 변동성에 해당하며 부정확한 숫자의 독립요소 선택의 손실을 보상해 주는 역할을 한다. $\hat{\mathbf{s}}_e = \mathbf{W}_e\mathbf{x}$ 일 때 아래와 같이 정의된다.

$$I_e^2 = \hat{\mathbf{s}}_e^T\hat{\mathbf{s}}_e \tag{5}$$

<표 1>은 ICA 기반 MSPC 모니터링을 위해 제안된 통계량들을 보여준다[2, 10, 11, 19, 26, 33]. 각 독립 성분들이 모니터링에 직접적으로 이용되거나, I^2 통계량으로 통합되어 사용하는 것을 확인할 수 있다. I^2 통계량은 ICA 도출 변수들이 정규분포를 따른다는 가정에 기반하고 있고, 이상 현상 검출 경계를 타원형으로 결정한다[15]. Hsu et al.은 AO (Adjusted Outlyingness)를 도입한 모니터링 방법론을 제안하였다[14]. ICA 도출 변수들이 비정규분포를 따른다는 가정에 근거한 사각형 타입 거리의 모니터링 통계량의 사용을 주장하였다. 하지만 Ge

〈표 1〉 ICA를 이용한 기존연구에서의 모니터링 통계량 정리

저자(연도)	차원감축기법	모니터링 통계량	통계량 타입	설명
Kano et al.(2003)	ICA	독립요소들	직접	각각의 독립요소들을 직접 모니터링
Ge and Song(2007)	ICA-PCA	I^2	분산 기반	기존 통계량
Albazzaz et al.(2007)	Dynamic ICA	독립요소들	직접	영향력 있는 독립 요소들을 직접 모니터링
Lee et al.(2007)	Kernel ICA	I^2, I^2_e, SPE	분산 기반	기존 통계량
Zhang et al.(2009)	KPCA+KICA	T^2, SPE	분산 기반	기존 통계량
Ge and Song(2009)	ICA-PCA	Bayesian monitoring statistic(BMS)	확률적	I^2를 이용해 계산됨
Yu(2010)	PCA, ICA	Mahalanobis distance negative log likelihood probability(MDNLLP)	확률적	I^2를 입력으로 이용해 계산됨

and Song에 따르면 실제 산업 공정은 항상 정규분포와 비정규분포를 따르는 변수들의 혼합이다[10]. 따라서 AO 또한 분포에 대한 가정의 부적합성 때문에 높은 성능을 보장할 수 없다. 사각형 타입 거리의 통계량으로는 다양한 분포의 다차원 ICA 도출 데이터에 대한 비선형 경계를 정확히 구할 수 없기 때문이다.

앞서 설명한 ICA 기반 공정 모니터링 방법론들은 실제 공정의 역동적 성질은 고려하지 못하고 있다. 이를 해결하기 위해 Lee et al.는 DICA를 제안하였다[24]. DICA는 관측 트레이닝 데이터셋에 시간지연 관측치들을 더함으로써 연결된 데이터 행렬을 구성하여 ICA 변환을 수행한다. 모니터링 과정에서는 관측치에 이전 관측치들을 연결하여 독립요소를 추출함으로써 자기상관을 제거한다. 이후, Lee et al.의 DICA를 기반으로 하는 다양한 방법론들이 제안되었다[2, 3, 13, 28]. 하지만 DICA를 이용한 접근법들 또한 여전히 도출 변수 분포에 대한 가정의 부적합성을 해결하지 못하였다.

3. 연구 방법론

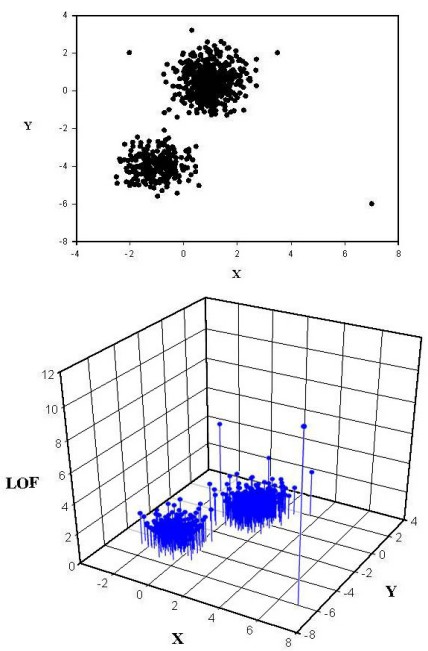
본 장에서는 LOF의 개념과 LOF를 이용한 모니터링 통계량을 설명한다.

3.1 LOF

LOF는 Breunig et al.에 의해 제안된 이상치 검출 기법이다[4]. LOF는 각각의 전체 데이터 개체에 대해 개별적인 개체 마다 이상치 정도를 나타내는 측정치를 계산하는 것이다. 이상치인 정도를 밀도에 근거한 정량적 수치로 나타냄으로써 지역적인 소규모의 이상치도 탐지할 수 있다. 데이터셋 **D**의 한 개체 **p**의 LOF 값을 계산하기 위한 식은 아래와 같이 정의된다.

$$\mathrm{LOF}(\mathbf{p}) = \{1/k * \sum_{\mathbf{o}\in k\mathrm{NN}(\mathbf{p})} lrd_k(\mathbf{o})\} / lrd_k(\mathbf{p}) \quad (6)$$

위의 식에서 **o**는 **p**의 k-nearest neighbor(kNN) 안의 각각의 개체, kNN(**p**)는 **p**에서 가장 가까운 개체부터 k번째로 가까운 개체까지 포함하는 집합, lrd는 local reachability density의 약자이다. k는 LOF 계산에 사용되는 nearest neighbor의 숫자를 의미하고, 사용자에 의해 정의되는 유일한 파라미터이다. LOF(**p**)는 **p**의 밀도 대비 kNN(**p**)의 평균 밀도 의 비율이다. 만약 **p**가 이상치가 아니라면, **p**와 kNN(**p**)의 밀도가 비슷하므로 LOF 값은 1에 가까워진다. 만약 **p**가 이상치라면, **p**가 kNN(**p**)에 비해 상대적으로 밀도가 작기 때문에 LOF 값은 1보다 커지게 된다. 그것은 그 개체가 보통 샘플들로부터 멀리 위치해서 LOF 값이 커지는 것을 의미한다. 정리해보면 LOF 값은 대상 개체가 kNN(**p**)로부터 멀리 떨어지고 다를수록 점점 커지고, 가깝고 비슷할수록 1에 수렴한다. [그림 2]는 계산된 LOF 값의 예를 보여주고 있다. 위의 그림은 인공 생성된 이차원 데이터의 산포도이고, 아래 그림은 각 데이터의 LOF 값이 z-축에 더해진 삼차원 산포도를 나타내



[그림 2] LOF 값의 개념적 설명을 위한 그림

고 있다. 위의 그림에서 우측 하단에 위치한 샘플은 실제 글로벌 이상치고 가장 높은 LOF 값을 가지는 것을 확인할 수 있다. 세 개의 로컬 이상치가 그림 상단에 주변 클러스터에서 조금 떨어진 곳에 위치하고 있는데, 그들의 LOF 값도 주변 클러스터 내 샘플들의 LOF 값보다 높아서 이상치로 검출된다.

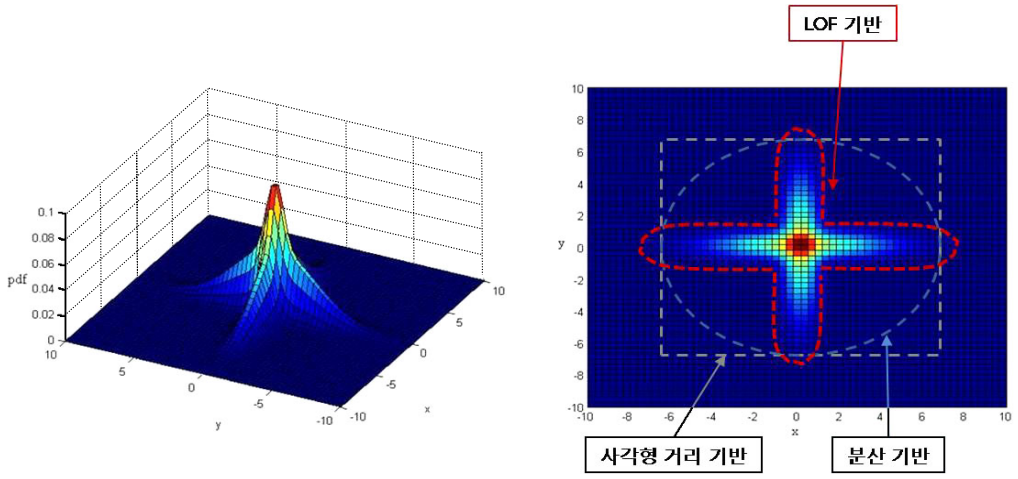
Lazarevic et al.은 LOF, NN 접근법, 마할라노비스 접근법, 비지도(unsupervised) SVM(Support Vector Machine) 등을 포함한 이상치 검출 알고리즘들의 성능비교 실험을 수행하였고, LOF의 성능이 가장 좋은 것을 보였다[21]. 그리고 데이터 스트림을 위한 incremental LOF 계산 알고리즘이 개발되었고, 이는 비디오 영상 모니터링에 적용되었다[29]. Chenetal et al.은 LOF를 교통 데이터 모니터링에 사용하였고, 다른 이상치 검출 기법들과 비교하였다[5]. Ganeriwal et al.은 LOF를 센서 네트워크의 감시 모듈을 구현하기 위해 사용하였다[9]. Duan et al.은 LOF를 이용한 위치 데이터 클러스터링 어플리케이션을 제안하였다[8].

3.2 LOF를 이용한 모니터링 통계량

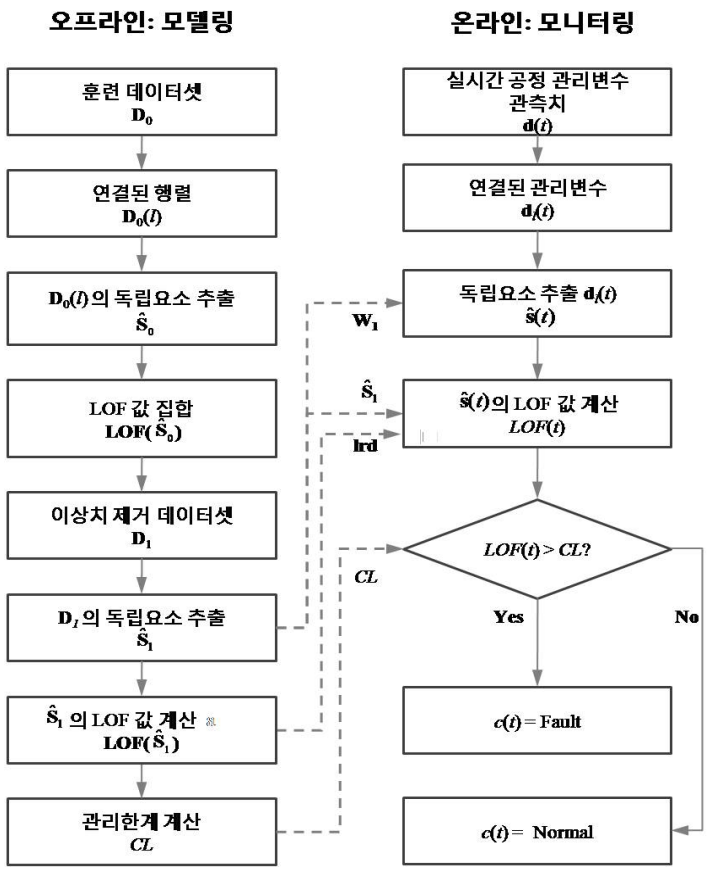
분산 기반 거리나 사각형 타입 거리 등을 모니터링 통계량으로 사용할 경우, 변수의 분포에 따라 이상 현상 검출의 정확도가 의존적이다. [그림 3]은 가우시안 혼합 분포를 따르는 두 변수의 pdf의 3차원 분포도이다. 두 개의 변수는 각각 좁은 폭과 넓은 폭을 가지는 두 개의 정규분포의 혼합으로 이루어진다. 이를 2차원 평면상에서 도시했을 때는 그림과 같고 분산 기반 거리나 사각형 타입 거리 기반 통계량은 이상 현상 판별 경계를 부정확하게 잡게 된다. 하지만 분포에 상관없는 통계량을 이용했을 때는 이상 현상 경계를 정확하게 잡을 수 있다. LOF를 이용하여 99% 경계를 잡았을 때는 이와 같다.

4. DICA-LOF 기반 모니터링

본 논문에서는 기존의 접근법들이 가진 한계점을 해결하기 위해 DICA-LOF 기반 모니터링 방법론을 제안하였다. 이를 통해 관리 변수 분포와 무관하게 높은 이상 현상 검출 능력을 보임과 동시에 시스템의 역동성을 고려할 수 있도록 하였다. [그림 4]는 DICA-LOF 기반 모니터링 방법론의 순서도를 보여준다. 방법론은 크게 모델링 단계와 모니터링 단계로 구성된다. 점선은 모델링 단계에서 구해진 계산 결과가 모니터링 단계에서 사용되는 흐름을 나타내고 있다. 이 장에서 사용되는 기호는 벡터-행렬 기호 정의에 따라 행렬에는 대문자 굵은 글씨체, 벡터는



[그림 3] 통계량의 성질에 따른 이상 현상 경계 결정 차이점 설명을 위한 그림



[그림 4] DICA-LOF의 순서도

소문자 굵은 글씨체, 스칼라에는 기울임꼴을 사용하였다. 모델링 단계는 훈련 데이터셋 내의 이상치를 제거, de-mixing 행렬 $\mathbf{W}$를 계산, 관리한계 CL의 계산이 목적이다. 그리고 모니터링 단계에서는 이상현상을 검출하기 위해 공정상태가 전 시점에 걸쳐 실시간으로 모니터링된다. 모니터링 방법론의 목표는 시점 t에 관측된 관측 변수들로부터 공정상태 $c(t)$를 판별하는 것이다. 모델링 단계와 모니터링 단계의 세부내용 설명이 각각 제 4.1절과 제 4.2절에 제공된다.

4.1 오프라인 훈련단계

모델링 단계는 오프라인으로 진행된다. 먼저 m개의 관리변수와 n개의 샘플로 구성된 데이터셋 $\mathbf{D}_0 = [\mathbf{d}_0(1), \cdots, \mathbf{d}_0(n)] \in \mathbf{R}^{m \times n}$ 가 훈련단계을 위해 구축된다. 데이터셋은 정상 공정상태를 설명하는 관측치들로 이루어진다. 그 다음 연결된 데이터셋 $\mathbf{D}_0(l)$을 구성한다. l은 시간지연을 의미한다. 연결된 행렬은 다음과 같다.

$$\mathbf{D}_0(l) = \begin{bmatrix} \mathbf{d}_0(l+1) & \mathbf{d}_0(l+2) & \cdots & \mathbf{d}_0(n) \\ \mathbf{d}_0(l) & \mathbf{d}_0(l+1) & \cdots & \mathbf{d}_0(n-1) \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{d}_0(1) & \mathbf{d}_0(2) & \cdots & \mathbf{d}_0(n-l) \end{bmatrix} \begin{matrix} lag0 \\ lag1 \\ \vdots \\ lagl \end{matrix} \quad (7)$$

DICA 모델에 의해 각 샘플에 이전의 l개의 연속된 샘플을 연결시켜 구성한다. 다음으로 FastICA 알

고리즘의 수행을 통해 독립요소 변환행렬인 $\hat{\mathbf{S}}_0$가 구해지고, $\hat{\mathbf{S}}_0 = \mathbf{W}_0\mathbf{D}_0$와 같다. de-mixing 행렬 $\mathbf{W}_0$에 의해 $\mathbf{D}_0(l)$의 각 열들이 추정 독립요소들로 변환되고 $[\hat{\mathbf{s}}_0(l+1), \cdots, \hat{\mathbf{s}}_0(n)] = \hat{\mathbf{S}}_0$이다. 그 후에 각 독립변수들의 LOF 값 집합 $\mathbf{LOF}(\hat{\mathbf{S}}_0)$가 다음 식에 의해 계산된다.

$$\mathbf{LOF}(\hat{\mathbf{S}}_0) = [LOF(\hat{s}_0(l+1), \cdots, LOF(\hat{s}_0(n))]$$

where

$$LOF(\hat{\mathbf{s}}_0(i)) = \{1/k * \sum_{\mathbf{o}\in k\mathrm{NN}(\hat{\mathbf{s}}_0(i))} lrd_k(\mathbf{o})\}/lrd_k(\hat{\mathbf{s}}_0(i)),$$
$$\text{for i = } l+1, \cdots, \text{n} \qquad (8)$$

그리고 데이터셋을 무결 상태로 만들기 이상치들이 제거된다. 만약 데이터셋이 이상치들에 의해 오염되어 있다면, 관리한계가 정확한 값보다 높게 추정(overestimated)된다. 이 경우 실제 이상 현상인 관측치가 입력으로 들어와도 정상으로 판별할 가능성이 높아진다. 따라서 $\mathbf{D}_0(l)$의 각 열에 대해, $LOF(\hat{s}_0(i))$가 $\mathbf{LOF}(\hat{\mathbf{S}}_0)$에 대해 핵밀도추정(Kernel Density Estimation, KDE)에 의해 정해지는 99.3% 한계보다 높은 경우 이상치로 판단하고 제거를 수행한다. 그 결과로 $\mathbf{D}_0(l)$에서 이상치가 제거된 n'개의 샘플로 구성된 $\mathbf{D}_1$을 얻게 된다. 핵밀도추정은 샘플들에서 비선형적 확률분포함수를 구하기 위해 사용된다. 핵밀도추정의 공식은 식 (9)에 의해 정의되며, x는 변수, $\hat{f}(x)$는 추정 확률밀도함수, x_i는 샘플값, n은 샘플의 수, K는 커널 함수를 의미한다. 가우시안 커널 함수가 가장 널리 사용되고 있다.

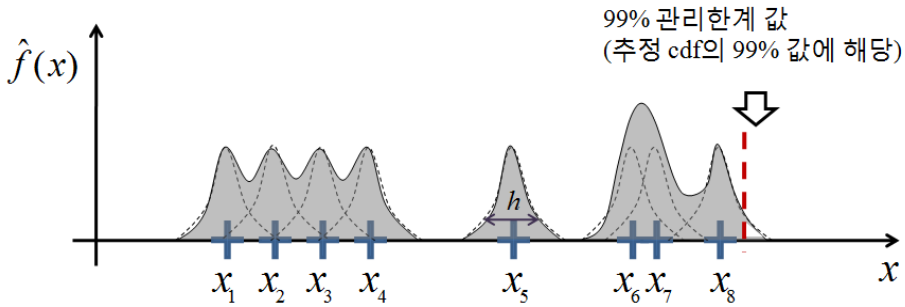
$$\hat{f}(x) = (\sum_i K[(x - x_i)/h])/(nh) \qquad (9)$$

핵밀도추정 계산의 설명을 위한 예시가 [그림 5]에 제공된다. 핵밀도추정에 의한 99% 한계값은 추정 누적분포함수의 99% 값에 해당한다. 이 예시에서 8개의 샘플 포인트가 있고, 각 샘플의 확률분포함수는 샘플값 x_i에 평균이 위치하도록 가우시안 커널 함수에 의해 계산된다. 점선으로 표시된 곡선은 각 샘플 확률분포함수, 직선 곡선은 추정 확률밀도함수를 나타낸다. 추정 확률밀도함수는 8개의 점선의 합과 같다. 설명한 방법으로 $\mathbf{LOF}(\hat{\mathbf{S}}_0)$의 99% 한계값이 $LOF(\hat{s}_0(i))$의 값들을 이용해 핵밀도추정을 통해 계산될 수 있다.

다음으로 $\mathbf{W}_1$이 계산되고, 이상치 안정 추정 독립요소 셋 $\hat{\mathbf{S}}_1$이 FastICA 알고리즘 수행에 의해 구해진다.

$$\hat{\mathbf{S}}_1 = \mathbf{W}_1\mathbf{D}_1 = [\hat{s}_1(1), \cdots, \hat{s}_1(n')] \qquad (10)$$

그리고 $\hat{\mathbf{S}}_1$에 대한 LOF 값들이 앞선 식과 같은 방식으로 계산된다. 마지막으로 모니터링 단계에서



[그림 5] 핵밀도추정을 통한 99% 관리한계를 구하는 과정 설명을 위한 그림

사용될 관리한계 *CL*을 정하기 위해 핵밀도추정에 의해 $\hat{\mathbf{S}}_1$의 LOF 값들에 대해 99% 관리한계가 계산된다.

4.2 온라인 모니터링 단계

모니터링 단계에서는 각 시점 t마다 새로운 관측치 $\mathbf{d}(t)$가 실시간으로 얻어진다. 그리고 모델링 단계에서와 같이 시간지연 l에 따라 연결된 관측치 $\mathbf{d}_l(t)$가 구축된다. $\mathbf{d}_l(t)$는 모델링 단계에서 구해진 $\mathbf{W}_1$에 의해 독립요소들 $\hat{\mathbf{s}}(t) = \mathbf{W}_1\mathbf{d}_l(t)$로 변환된다. 다음으로, $\hat{\mathbf{s}}(t)$의 LOF 값, *LOF*(t)가 계산된다. 마지막으로 현재 공정 상태 $c(t)$가 다음과 같이 판별된다. 만약 *LOF*(t)가 *CL*을 초과할 경우, $c(t)$는 이상 현상으로 판정된다. 아닐 경우 $c(t)$는 정상으로 판정되고 모니터링이 계속된다. 본 연구에서는 *LOF*(t)의 계산 복잡도를 줄이기 위해 근사 알고리즘을 사용하였고 수도 코드는 아래와 같다.

각 시점 t에, *LOF*(t)가 $\{\hat{\mathbf{s}}(t) \cup \hat{\mathbf{S}}_1\}$의 셋에 대해 계산된다. 정확한 LOF 알고리즘은 각 시점 t마다 반복적으로 셋 내의 모든 개체에 대해 LOF 값 계산을 수행해야 한다. 만약 $\hat{\mathbf{S}}_1$내의 개체 수와 모니터링 관측치의 수가 각각 N_1, N_2라면, $(N_1+1)*N_2$번의 계산이 요구된다. 본 논문에서 제안된 근사 LOF 계산 알고리즘은 훈련 데이터셋에 대해 모델링 단계에서 LOF와 lrd 값들을 미리 계산해 두고, 모니터링 단계에서 각 시점에 새로운 관측치에 대한 LOF 값을 계산할 때 그 값들을 다시 계산하지 않는다. 그 결과 N_2번의 계산만 요구된다. 만약 N_1이 충분히 크다면, 훈련 데이터셋의 lrd 값들은 거의 바뀌지 않는다. 결과적으로 조금의 정보손실로 계산부담은 크게 감소된다. *LOF*(t)는 각 모니터링 시점의 관측치의 셋 $\{\hat{\mathbf{s}}(t) \cup \hat{\mathbf{S}}_1\}$에 대한 이상도에 해당하고, 그것은 $\hat{\mathbf{s}}(t)$의 $\hat{\mathbf{S}}_1$에 대한 비유사도와 같다.

5. 실험 및 결과

본 연구에서는 최근 공정 모니터링 분야에서 방법론의 성능비교를 위해 널리 사용되어 온 TE process의 데이터를 사용하여 비교실험을 수행하였다. 제 5.1절에서는 TE process 데이터와 그를 이용한 비교실험 설계가 설명된다. 그리고 실험결과 및 토의는 제 5.2절에서 제공된다.

5.1 실험 데이터셋

TE process는 Downs and Vogel[7]에 의해 제안된 공장 규모의 공정 관리 문제로, 최적화, 예측적 관리, 공정 진단, 관리 교육 등에 광범위하게 사용되어 왔다[18]. 실험 데이터는 Chiang et al.(2001)의

근사 **LOF**_계산(비교 데이터셋 $\hat{\mathbf{S}}_1$, 훈련 *lrd* 벡터 **lrd**, 훈련 *k*NN 색인 행렬 **NN**, 변환된 모니터링 관측값 벡터 $\hat{\mathbf{s}}(t)$)

- 주어진 정보 : $\hat{\mathbf{S}}_1\{\hat{\mathbf{s}}_1(1), \cdots, \hat{\mathbf{s}}_1(n')\}$, $\hat{\mathbf{s}}_1(i) \in \mathrm{Rm}'$, $\mathbf{lrd} = [lrd_k(\hat{\mathbf{s}}_1(1)), \cdots, lrd_k(\hat{\mathbf{s}}_1(n'))]$, $\mathbf{NN} = \{\mathbf{nn}(\hat{\mathbf{s}}_1(1)), \cdots, \mathbf{nn}(\hat{\mathbf{s}}_1(n'))\}$, ($m'$은 모니터링 IC 벡터의 차원, n'은 훈련 IC 데이터셋의 샘플 수에 해당하고 $\mathbf{nn}(\hat{\mathbf{s}}_1(i)) \in \mathbf{R}^k$일 때)
- 온라인 모니터링 단계에서, 각각의 새로운 관측치 $\hat{\mathbf{s}}(t)$에 대해
  - $(\forall \hat{\mathbf{s}}_1(i) \in \hat{\mathbf{s}}_1)$
    $\hat{\mathbf{s}}(t)$와 $\hat{\mathbf{s}}_1(i)$간의 거리를 계산
  - $\mathbf{nn}(\hat{\mathbf{s}}(t))$를 구축하기 위해, $\hat{\mathbf{S}}_1$에서 $\hat{\mathbf{s}}(t)$의 kNN을 탐색
  - $(\forall \hat{\mathbf{s}}_1(i) \in (\mathbf{nn}(\hat{\mathbf{s}}(t)))$
    $\hat{\mathbf{s}}(t)$의 *lrd*인 lrdk($\hat{\mathbf{s}}(t)$)를 계산하기 위해, $\hat{\mathbf{s}}(t)$에 대한 $\hat{\mathbf{s}}_1(i)$들의 거리를 이용함.
  - 식 (6)을 이용해 아래 식과 같이 *LOF*(t)를 계산,
    $$LOF(t) = \{1/k^* \sum_{\mathbf{o} \in k\mathrm{NN}(\hat{\mathbf{s}}(t))} lrd_k(\mathbf{o})\} / lrd_k(\hat{\mathbf{s}}(t))$$
- 종료 // 반복

연구에서 생성되었던 데이터를 사용한다[35]. [그림 6]은 Downs and Vogel[7]에서 사용된 TE process의 공정도이다. 이 프로세스는 stirred tank reactor, 콘덴서, 증기-액체 분리기, 증류기로 구성된다. TE process는 53개의 변수로 구성되어 있지만, 본 실험에서는 Lee et al.[22]에 의해 제안된 <표 2>의 33개의 변수를 사용한다.

Fault 0는 500개의 정상 상태 관측치들로 구성된다. 이것을 오프라인 트레이닝 과정에 트레이닝 데이터로 사용한다. 그리고 Fault 1~Fault 21은 각기 다른 종류의 이상 현상 상태 관측치들을 가진 데이터이며 각각 960개의 관측치들로 구성되어 있고, 이상 현상은 161번째 데이터부터 도입된다. 따라서 161~960번 관측치들은 모두 이상 현상으로 구성된다. 이상 현상들에 대한 상세한 설명은 <표 3>에 정리되어 있다. DICA를 위한 시간지연변수 l로는 실험을 통해 가장 좋은 성능을 보인 2를 사용하였다.

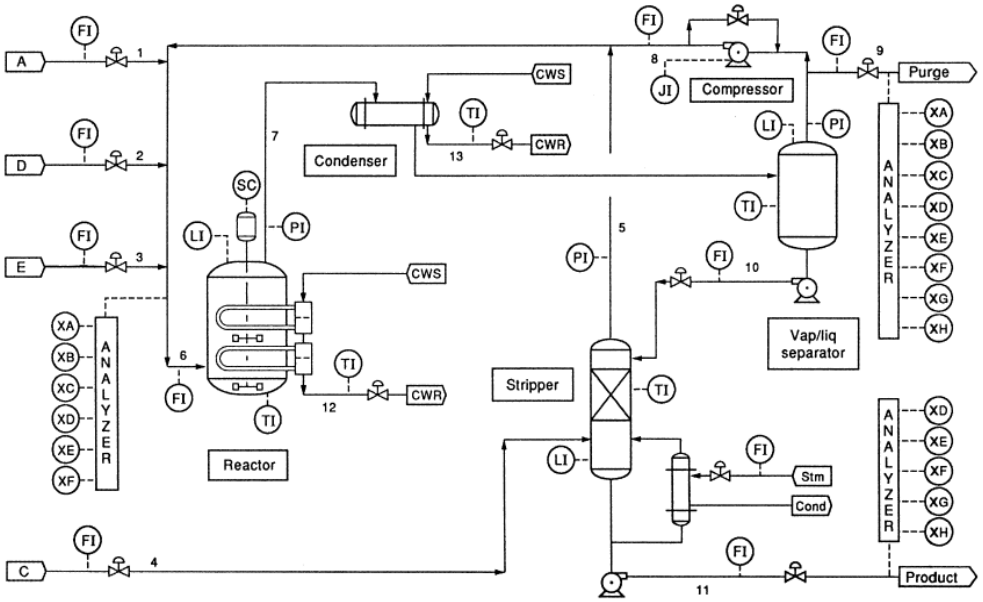
시점 별 관측치들 사이에 자기상관이 존재하는지 확인하기 위한 그래프가 [그림 7]에 제공된다. 그림은 각 모니터링 데이터들의 LOF 값의 자기상관함수 값을 시간지연에 대해 그린 결과이다. 시간지연 h에 대한 자기상관 대한 식은 아래와 같다. N은 샘플의 수, Y_i는 샘플의 값, $\overline{Y}$는 샘플 데이터의 평균값을 의미한다.

$$r_h = \sum_{t=1}^{N-h}(Y_t-\overline{Y})(Y_{t+h}-\overline{Y})/\sum_{t=1}^{N}(Y_t-\overline{Y})^2 \quad (11)$$

신뢰한계 B에 대한 식은 아래와 같다.

$$B = \pm z_{1-\alpha/2}/N^{1/2} \quad (12)$$

그래프의 가로 방향으로 그어진 선은 각각 유의수준에 대한 신뢰 상한, 하한 값을 의미하는데, 자기상관 값이 이 한계값을 넘어갈 경우 데이터에 자기상관이 없다는 귀무가설이 기각되고, 자기상관이 존재한다는 것을 의미한다. [그림 7]에서는 자기상관이 큰 값으로 유지되므로, TE process의 관측치들 사이에 자기상관이 강하게 존재한다는 것을 확



[그림 6] TE process의 플로시트

〈표 2〉 모니터링 변수에 대한 설명

변 수	설명
XMEAS(1)	A feed(stream 1)
XMEAS(2)	D feed(stream 2)
XMEAS(3)	E feed(stream 3)
XMEAS(4)	A and C feed(stream 4)
XMEAS(5)	Recycle flow(stream 8)
XMEAS(6)	Reactor feed rate(stream 6)
XMEAS(7)	Reactor pressure
XMEAS(8)	Reactor level
XMEAS(9)	Reactor temperature
XMEAS(10)	Purge rate(stream 9)
XMEAS(11)	Product sep temp
XMEAS(12)	Product separator level
XMEAS(13)	Product separator pressure
XMEAS(14)	Product separator underflow (stream 10)
XMEAS(15)	Stripper level
XMEAS(16)	Stripper pressure
XMEAS(17)	Stripper underflow(stream 11)
XMEAS(18)	Stripper temperature
XMEAS(19)	Stripper steam flow
XMEAS(20)	Compressor work
XMEAS(21)	Reactor cooling water outlet temperature
XMEAS(22)	Separator cooling water outlet temperature
XMV(1)	D feed flow(stream 2)
XMV(2)	E feed flow(stream 3)
XMV(3)	A feed flow(stream 1)
XMV(4)	A and C feed flow(stream 4)
XMV(5)	Compressor recycle valve
XMV(6)	Purge valve(stream 9)
XMV(7)	Separator pot liquid flow (stream 10)
XMV(8)	Stripper liquid product flow (stream 11)
XMV(9)	Stripper steam valve
XMV(10)	Reactor cooling water valve
XMV(11)	Condenser cooling water flow

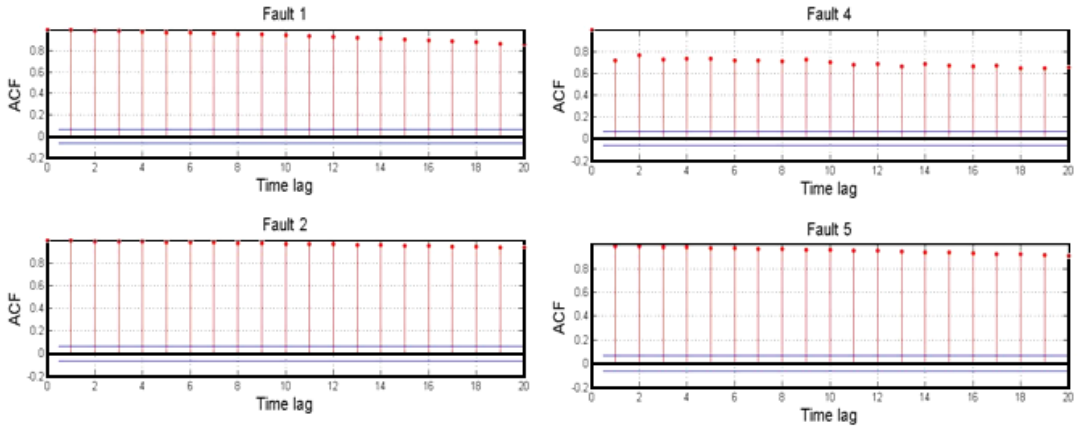
〈표 3〉 이상 현상들에 대한 설명

Fault 넘버	상태	이상 현상 타입
0	No fault	No
1	A/C feed ratio, B composition constant(stream 4)	Step
2	B composition, A/C ratio constant (stream 4)	Step
3	D feed temperature(stream 2)	Step
4	Reactor cooling water inlet temperature	Step
5	Condenser cooling water inlet temperature	Step
6	A feed loss(stream 1)	Step
7	C header pressure loss-reduced availability(stream 4)	Step
8	A, B, C feed composition (stream 4)	Random variation
9	D feed temperature (stream 2)	Random variation
10	C feed temperature (stream 4)	Random variation
11	Reactor cooling water inlet temperature	Random variation
12	Condenser cooling water inlet temperature	Random variation
13	Reaction kinetics	Slow drift
14	Reactor cooling water valve	Sticking
15	Condenser cooling water valve	Sticking
16	Unknown	Unknown
17	Unknown	Unknown
18	Unknown	Unknown
19	Unknown	Unknown
20	Unknown	Unknown
21	Valve position constant (stream 4)	Constant position

인할 수 있다.

5.2 실험결과 및 토의

본 논문에서 제안하는 방법론의 성능 개선을 입증하기 위해, 기존 ICA 기반 접근법들과의 비교실험

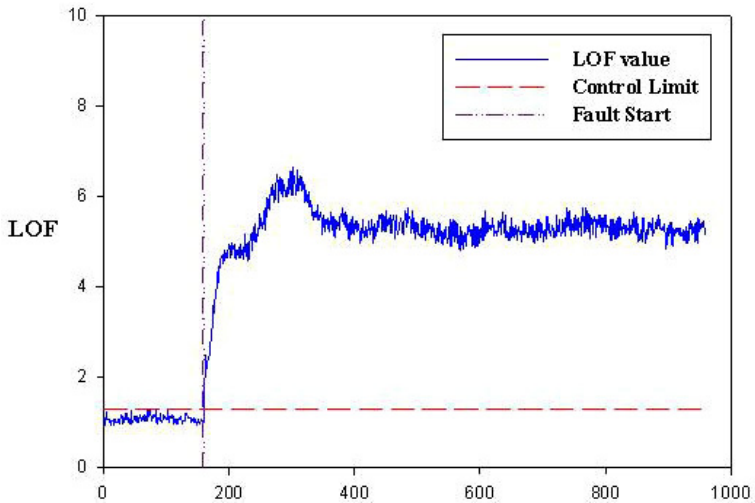


[그림 7] 모니터링 데이터의 LOF 값에 대한 자기상관 그래프

을 수행하였다. [그림 8]은 제안한 방법론으로 이상 현상을 검출하는 MSPC 차트의 예시를 보여준다. MSPC는 각 데이터에 대해서 시점별 관측치로부터 통계량을 구하여 공정 상태를 판별한다.

본 논문의 실험에서는 각 모니터링 방법론의 성능을 이상 현상 검출 정확도로 비교하였다. <표 4>는 비교 실험에 활용된 각 방법론들을 요약하였다. 방법론 이름은 괄호 안은 사용되는 모니터링 통계량, 밖은 차원감축기법을 의미한다. No(*LOF*)는 독립요소 추출을 하지 않고, LOF를 모니터링 통계량으로 이용한다. ICA(I^2), ICA(*AO*), ICA(*LOF*)는 독립요소 추출을 수행하고, 각각 I^2, AO, LOF의 통계량을 사용한다. 그리고 DICA(I^2), DICA(*AO*), DICA(*LOF*)은 자기상관을 고려하기 위해 DICA를 이용하고, 각각 I^2, AO, LOF의 통계량을 사용한다. 제안된 방법론인 DICA-LOF는 DICA(*LOF*)와 같다.

<표 5>는 비교실험의 결과를 보여준다. 데이터는 변수 간의 스케일 차이로 발생할 수 있는 부정확성을 방지하기 위해 모두 정규화되었다. LOF 계산에 쓰이는 파라미터인 k는 실험을 통하여 가장



[그림 8] Fault 1 데이터에 대해 계산된 LOF 값을 MSPC에 사용하는 실제 예시

〈표 4〉 비교실험에 사용되는 공정 모니터링 방법론의 요약

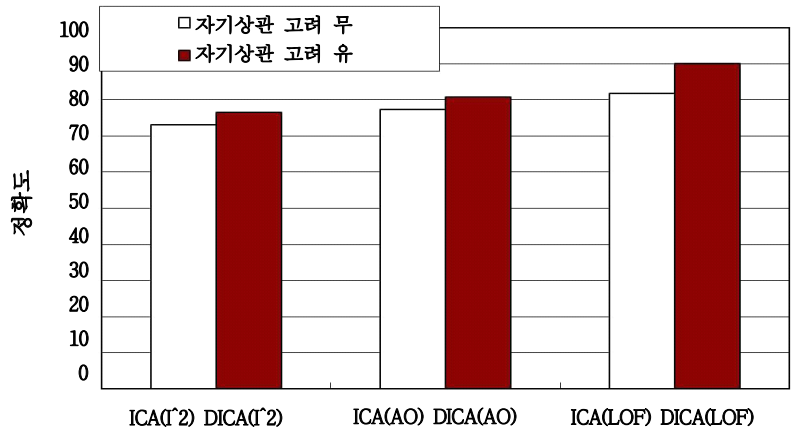
방법론	독립요소 추출	행렬 연결	이상치 제거	통계량	관리 한계
No(*LOF*)	No	No	*LOF*(99.3% KDE)	*LOF*	99%(KDE)
ICA(I^2)	ICA	No	I^2(99.3% KDE)	I^2	99%(KDE)
ICA(*AO*)	ICA	No	*AO*(AO rejection rule)	*AO*	99%(KDE)
ICA(*LOF*)	ICA	No	*LOF*(99.3% KDE)	*LOF*	99%(KDE)
DICA(I^2)	DICA	Yes	I^2(99.3% KDE)	I^2	99%(KDE)
DICA(*AO*)	DICA	Yes	*AO*(AO rejection rule)	*AO*	99%(KDE)
DICA(*LOF*) {DICA-LOF}	DICA	Yes	*LOF*(99.3% KDE)	*LOF*	99%(KDE)

〈표 5〉 이상 현상 검출 정확도로 제시된 비교실험 결과(각 Fault에 대해, 제일 좋은 결과는 고딕체로 표시)

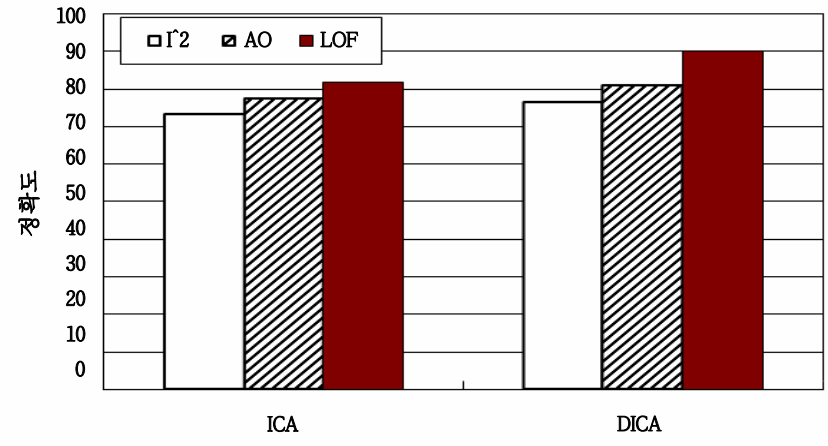
Faults	독립요소 추출 미수행	독립요소 추출 수행					
		자기상관 고려 무			자기상관 고려 유		
	No(*LOF*)	ICA(I^2)	ICA (*AO*)	ICA (*LOF*)	DICA(I^2)	DICA (*AO*)	DICA (*LOF*)
1	100	100	100	100	100	100	100
2	99	98	98	99	99	99	99
3	19	1	2	19	2	2	43
4	100	61	84	100	97	100	100
5	39	100	100	100	100	100	100
6	100	100	100	100	100	100	100
7	100	99	100	100	100	100	100
8	99	97	97	99	98	98	99
9	15	1	1	14	1	1	37
10	70	78	82	90	82	90	96
11	69	52	70	74	54	83	95
12	100	99	100	100	100	100	100
13	96	94	95	96	100	96	97
14	100	100	100	100	95	100	100
15	23	2	2	21	2	2	40
16	71	71	78	94	82	91	99
17	90	93	94	96	90	96	98
18	91	90	90	91	90	90	94
19	35	69	80	87	81	95	100
20	68	87	91	83	88	92	92
21	42	45	62	54	46	62	100
평균	72.7	73.2	77.4	81.8	76.5	80.8	90

좋은 성과를 보인 20으로 사용하였다. 결과를 살펴보면 제안된 방법론인 DICA-LOF가 비교 방법론 중 가장 좋은 정확도를 보였다. 이 중 Fault 3, 9, 15는 Lee et al.의 연구[22]에 따르면 도입된 이상 현상의 크기가 너무 작아 모니터링 중 검출하기 어려움에도 불구하고 제안된 방법론은 기존 방법론들보다 월등한 성능을 보여 작은 크기의 이상 현상도 민감하게 검출할 수 있는 장점을 보였다. 반면에 독립요소 추출이 선행되지 않은 상태에서 LOF를 모니터링에 사용했을 때는(No(*LOF*)), 기존 방법론들에 비해 개선된 결과를 도출하지 않았다. 그 결과는 ICA 수행이 비가우시안 공정에서 중요한 의미를 가지는 정보를 추출해 내고, 따라서 독립요소를 추출하는 과정이 효과적인 이상 현상 검출을 위해서 핵심적인 역할을 수행한다는 것을 의미한다.

[그림 9]에서는 자기상관을 고려하지 않았던 이상 현상 검출 방법론에 자기상관을 고려했을 때 생기는 정확도 개선을 보여주고 있다. 역동성(dynamic)이 고려된 모든 DICA 기반 방법론들은 그렇지 않은 일반적인 ICA 기반 방법론들보다 높은 성능을 보였



[그림 9] 자기상관 고려에 따른 정확도 개선



[그림 10] LOF 도입에 따른 정확도 개선

다. LOF를 사용하는 방법론인 DICA-LOF에서 8.2%로 큰 개선이 있었다. 이는 자기상관을 고려하는 효과가 LOF 사용 방법론에서 가장 큰 것을 의미한다.

[그림 10]은 LOF 도입에 의한 모니터링 방법론의 성능 개선을 보여주고 있다. ICA 기반 방법론 및 DICA 기반 방법론들에서 통계량에 따른 검출 정확도를 비교하였다. ICA 및 DICA 기반 방법론 모두에서 I^2와 AO보다 LOF를 통계량으로 이용한 경우 성능의 개선이 나타났다. 실제 관리변수에서 추출된 독립요소들은 정규분포와 다른 분포들의 혼합으로 나타나기 때문에, 특정 분포에 대한 가정이 필요한 I^2나 AO는 제한적인 성능을 보였다. 하지만 LOF를 통계량으로 이용할 경우에는 변수들의 분포에 무관한 이상도가 계산되기 때문에 높은 성능을 보였다.

5. 결 론

본 논문에서는 DICA와 LOF를 통합한 공정 모니터링 방법론을 제안하였다. 기존 ICA 기반 모니터링 방법론의 공정의 역동성을 고려하지 못하는 한계점을 극복하기 위하여 DICA 기반 방법론이 제안되었다. 하지만 DICA에 기반한 모니터링 방법론도 ICA 도출 변수의 분포에 성능이 좌우되는 한계점이 있다. 이를 해결하기 위하여 본 논문에서는 DICA 기반 방법론의 모니터링 통계량으로 활용하기 위하여 LOF를 도입하였다. LOF 값은 모니터링에 사용되는 변수들의 특정 분포를 가정할 필요가 없기 때문에, DICA 기반 공정 모니터링 방법론에 기존 통계량보다 더 적합한 통계량이다. 그 결과 제안된 방법론은 모니터링에 사용되는 변수들의 분포에 무관하게 안정적인 성능을 제공함으로써 기존 방법론들보다 개선된 성능을 보일 수 있다. 또한 LOF 값 계산에 따르는 계산 부담을 줄이기 위해 근사 알고리즘이 사용되었고, 이로 인해 제안된 방법론을 실시간 모니터링에 더욱 실용적으로 적용할 수 있었다. 제안된 방법론은 TE process에 기존 방법론들과 성능 비교실험이 수행되었다. LOF를 사용한 방법론이 기존 ICA 기반 모니터링과 DICA 기반 모니터링에서 큰 폭의 성능 향상을 가져옴을 확인할 수 있었다. 실험결과는 실제 공정은 여러 가지 분포를 따르는 잠재변수들의 혼합으로 구성되기 때문에, LOF가 기존 방법론들의 통계량보다 공정 모니터링에 더 적합함을 의미한다. 제안된 방법론을 현업에 적용함으로써, 사고 위험의 감소와 공정의 효율성 향상을 통한 지속적 공정 개선이 이루어질 수 있을 것이다.

추후 연구과제로는, 제안된 방법론은 모델링 단계에서 결정된 관리한계를 정적으로 사용하기 때문에, 관리한계가 모니터링 과정에서 실시간으로 조정된다면 추가적인 개선이 있을 것이다. 또한 SVDD (Support Vector Description) 등의 다른 비선형 경계 추정 기법을 이용하는 모니터링 방법론들과의 성능 비교도 이루어질 수 있을 것이다. LOF 통계량의 PCA 기반 MSPC와 ICA 기반 MSPC에서의 성능 비교도 의미있는 주제가 될 것이다.

참 고 문 헌

[1] 이재신, 강복영, 강석호, “LOF와 dynamic ICA를 이용한 프로세스 모니터링 방법론”, 2010 한국경영과학회 추계학술대회논문집, pp.9-24.

[2] Albazzaz, H. and X. Wang, “Introduction of dynamics to an approach for batch process monitoring using independent component analysis,” *Chemical Engineering Communications*, Vol.194, No.2(2007), pp.218-233.

[3] Albazzaz, H. and X. Wang, “Multivariate statistical batch process monitoring using dynamic independent component analysis,” *Computer Aided Chemical Engineering*, Vol.21, No.(2006), pp.1341-1346.

[4] Breunig, M.M., H.-P. Kriegel, R.T. Ng, and J. Sander, “LOF : identifying density-based local outliers,” *SIGMOD Rec.*, Vol.29, No.2(2000), pp.93-104.

[5] Chen, S., W. Wang, and H. van Zuylen, "A comparison of outlier detection algorithms for ITS data," *Expert Systems with Applications,* Vol.37, No.2(2010), pp.1169-1178.

[6] Cho, H.-W., K.-J. Kim, and M.K. Jeong, "Multivariate statistical diagnosis using triangular representation of fault patterns in principal component space," *International Journal of Production Research,* Vol.43, No.24(2005), pp.5181-5198.

[7] Downs, J.J. and E.F. Vogel, "A plant-wide industrial process control problem," *Computers and Chemical Engineering,* Vol.17, No.3 (1993), pp.245-255.

[8] Duan, L., L. Xu, F. Guo, J. Lee, and B. Yan, "A local-density based spatial clustering algorithm with noise," *Information Systems,* Vol.32, No.7(2007), pp.978-986.

[9] Ganeriwal, S., L.K. Balzano, and M.B. Srivastava, "Reputation-based framework for high integrity sensor networks," *ACM Trans. Sen. Netw.,* Vol.4, No.3(2008), pp.1-37.

[10] Ge, Z. and Z. Song, "Multimode process monitoring based on Bayesian method," *Journal of Chemometrics,* Vol.23, No.12(2009), pp.636-650.

[11] Ge, Z. and Z. Song, "Process Monitoring Based on Independent Component Analysis-Principal Component Analysis (ICA-PCA) and Similarity Factors," *Industrial and Engineering Chemistry Research,* Vol.46, No.7 (2007), pp.2054-2063.

[12] Harris, T.J., C.T. Seppala, and L.D. Desborough, "A review of performance monitoring and assessment techniques for univariate and multivariate control systems," *Journal of Process Control,* Vol.9, No.1(1999), pp.1-17.

[13] Hsu, C., M. Chen, and L. Chen, "A novel process monitoring approach with dynamic independent component analysis," *Control Engineering Practice,* Vol.18, No.3(2010), pp.242-253.

[14] Hsu, C.C., L.S. Chen, and C.H. Liu, "A process monitoring scheme based on independent component analysis and adjusted outliers," *International Journal of Production Research,* Vol.48, No.6(2010), pp.1727-1743.

[15] Hubert, M. and S. Van der Veeken, "Outlier detection for skewed data," *Journal of Chemometrics,* Vol.22, No.3-4(2008), pp.235-246.

[16] Hyvärinen, A. and E. Oja, "Independent component analysis : algorithms and applications," *Neural Networks,* Vol.13, No.4-5(2000), pp. 411-430.

[17] Kano, M. and Y. Nakagawa, "Data-based process monitoring, process control, and quality improvement : Recent developments and applications in steel industry," *Computers and Chemical Engineering,* Vol.32, No.1-2(2008), pp.12-24.

[18] Jockenhövel, T., L.T. Biegler, and A. Wächter, "Dynamic optimization of the Tennessee Eastman process using the Opt Control Centre," *Computers and Chemical Engineering,* Vol.27, No.11(2003), pp.1513-1531.

[19] Kano, M., S. Tanaka, S. Hasebe, I. Hashimoto, and H. Ohno, "Monitoring independent components for fault detection," *AIChE Journal,* Vol.49, No.4(2003), pp.969-976.

[20] Kresta, J.V., J.F. Macgregor, and T.E. Marlin, "Multivariate statistical monitoring of process operating performance," *The Canadian Journal of Chemical Engineering,* Vol.69, No.1(1991), pp.35-47.

[21] Lazarevic, A., L. Ertoz, V. Kumar, A. Ozgur, and J. Srivastava, "A comparative study of

anomaly detection schemes in network intrusion detection," *Proceedings of Third SIAM Conference on Data Mining,* (2003), pp.25-36.

[22] Lee, J.-M., S.J. Qin, and I.-B. Lee, "Fault detection and diagnosis based on modified independent component analysis," *AIChE Journal,* Vol.52, No.10(2006), pp.3501-3514.

[23] Lee, J.-M., C. Yoo, S.W. Choi, P.A. Vanrolleghem, and I.-B. Lee, "Nonlinear process monitoring using kernel principal component analysis," *Chemical Engineering Science,* Vol.59, No.1(2004), pp.223-234.

[24] Lee, J.-M., C. Yoo, and I.-B. Lee, "Statistical monitoring of dynamic processes based on dynamic independent component analysis," *Chemical Engineering Science,* Vol.59, No.14 (2004), pp.2995-3006.

[25] Lee, J.-M., C. Yoo, and I.-B. Lee, "Statistical process monitoring with independent component analysis," *Journal of Process Control,* Vol.14, No.5(2004), pp.467-485.

[26] Lee, J., S. Qin, and I. Lee, "Fault detection of non-linear processes using kernel independent component analysis," *The Canadian Journal of Chemical Engineering,* Vol.85, No.4 (2007), pp.526-536.

[27] Martin, E.B. and A.J. Morris, "Non-parametric confidence bounds for process performance monitoring charts," *Journal of Process Control,* Vol.6, No.6(1996), pp.349-358.

[28] Monroy, I., R. Benitez, G. Escudero, and M. Graells, "DICA enhanced SVM classification approach to fault diagnosis for chemical processes," *Computer Aided Chemical Engineering,* Vol.26, No.(2009), pp.267-272.

[29] Pokrajac, D., A. Lazarevic, and L.J. Latecki, "Incremental local outlier detection for data streams," *In Proceedings of IEEE Symposium on Computational Intelligence and Data Mining,* Vol.769(2007), pp.504-515.

[30] Venkatasubramanian, V., R. Rengaswamy, K. Yin, and S.N. Kavuri, "A review of process fault detection and diagnosis : Part I : Quantitative model-based methods," *Computers and Chemical Engineering,* Vol.27, No.3(2003), pp.293-311.

[31] Wang, H., T.-Y. Chai, J.-L. Ding, M. Brown, "Data Driven Fault Diagnosis and Fault Tolerant Control : Some Advances and Possible New Directions," *Acta Automatica Sinica,* Vol.35, No.6(2009), pp.739-747.

[32] Wang, K. and F. Tsung, "Monitoring feedback-controlled processes using adaptive T2 schemes," *International Journal of Production Research,* Vol.45, No.23(2007), pp.5601-5619.

[33] Yu, J., "Hidden Markov models combining local and global information for nonlinear and multimodal process monitoring," *Journal of Process Control,* Vol.20, No.3(2010), pp. 344-359.

[34] Zhang, Y. and J. Jiang, "Bibliographical review on reconfigurable fault-tolerant control systems," *Annual Reviews in Control,* Vol. 32, No.2(2008), pp.229-252.

[35] http://brahms.scs.uiuc.edu.