

시간대에 따른 국부모델을 이용한 회분식 공정의 실시간 감시 및 진단

이영학 · 조현우 · 한종훈[†]

공정산업의 지능자동화 연구센터, 포항공과대학교 화학공학과
(1998년 10월 16일 접수, 1999년 3월 22일 채택)

Real-Time Monitoring and Diagnosis for Batch Processes Using Local Models by Time Intervals

Young-Hak Lee, Hyun-Woo Cho and Chonghun Han[†]

Automation Research Center, Department of Chemical Engineering,
Pohang University of Science and Technology
(Received 16 October 1998; accepted 22 March 1999)

요 약

회분식 공정은 산업계의 정밀화학제품 등의 고부가가치 제품생산에 중요한 역할을 담당해 왔다. 이러한 회분식 공정의 안전 운전과 양질의 제품을 생산하기 위해서 이상의 감시 및 진단 기술은 필수적이다. Multiway principal component analysis(MPCA)를 이용한 통계적인 방법은 회분식 공정의 강한 비선형성이나 unsteady-state를 제거해줌으로써 수월한 모니터링을 수행할 수 있게 해준다. 그러나 이러한 방법은 파라미터들이 너무 많으므로 계산 속도나 미래의 동태를 예측할 때 어려움이 있다. 본 연구에서는 이를 해결하기 위해 시간대별로 회분시간을 나누어서 구성한 국부 모델을 이용하여 이상을 감시하는 방법을 제안하였다. 또한 각 변수의 실시간 기여도(contribution)를 통해 진단을 수행하였다. 본 연구에서는 이러한 방법들을 기반으로 회분식 공정의 이상을 감시하고 진단하는 방법을 PVC 중합반응기에 적용하였으며, 그 결과 좀더 빠르고 정확하게 이상을 감시하고 원인을 진단할 수 있음을 확인하였다.

Abstract – Batch and semi-batch processes have played an important role in the chemical industry, because of their capabilities to produce high-value products such as fine chemicals. Monitoring batch processes is very important to ensure their safe operation and to assure the production of consistent high-quality products. Multivariate statistical analysis method using multiway principal component analysis(MPCA) provides efficient monitoring scheme by removing constraints in the batch process such as the presence of significant nonlinearity, the absence of steady state. But this method has too many parameters, delays computing speed, makes it inaccurate to predict future observations. In this study we suggested more rapid, accurate monitoring and diagnosis methods by constructing local models based on time intervals. In special we could select faulty sensors by checking the real-time contribution of variables on the abnormal events. The suggested monitoring and diagnosis schemes were tested to PVC polymerization reactor, capable of monitoring faults and diagnosing the assignable causes in the good performance.

Key words: Multiway PCA, Monitoring and Diagnosis, Local Model, PVC Polymerization

1. 서 론

회분식 공정은 다품종 소량생산에 많이 이용되고 있는 공정으로 고분자제품, 의약품, 생화학 재료 등의 고부가가치 제품을 생산하는데 중요한 역할을 담당해 왔다. 회분식 공정을 통한 제품들은 종종 여러 단계를 거쳐 생산되는데 임의의 어떤 단계에서 불량 제품이 생산된다는 것을 초기에 감지하고 적절한 조치를 취함으로써 불량률을 줄일 수 있고, 이는 공장의 생산성 증가에 크게 기여한다. 즉, 공정의 이상을 감시할 수 있는 모니터링 체계를 갖추으로써 공정 이상의 원인을 분석하고 그 원인을 제거할 수 있으며 이로 인해 공정의 불안

정한 상태를 개선시킨다. 또한 이러한 모니터링 기술을 온라인으로 구현함으로써 적절한 제어전략을 이용하여 조업 초반부에 이상을 제거하고 공정을 정상으로 복귀시킬 수 있는 기회가 생기게 된다.

회분식 공정은 다음 단계별로 운전이 진행되는데 먼저 원료 물질이 회분용기로 주입되고, 온도, 압력, 유량 등의 공정 변수가 시간에 따라 변하므로 이를 제어하면서 생성물의 전환율이 일정한 값까지 도달했을 때 조업을 멈춘다. 마지막으로 용기로부터 제품을 뽑아낸다. 일반적으로 이런 과정 중에 공정이 흔들리는 요인은 평균추이를 크게 벗어나는 변수들의 움직임, 원료 주입상의 예리, 불순물로 인해 생기는 외란 등인 것으로 알려져 있다. 성공적인 회분 조업이라 함은 최적 원료의 혼합 및 적합한 시간에서의 반응 매체들의 주입, 그리고 변수들의 추이가 재현성을 갖도록 조업하는 것이다. 유량, 온도나 압

[†]E-mail : chan@postech.ac.kr

력 profile은 set-rate tracking을 위해 servo-controllers를 통해 제어되고, 시간에 따른 각각의 조업은 PLC(Programmable Logic Controller) 같은 도구를 이용해 실제 현장에서 구현된다[5].

회분식 공정의 이상 감시 및 원인의 진단은 공정상에 강한 비선형성이 존재하고 제품의 질을 측정하는 온라인 센서가 부족하며 공정상에 존재하는 복잡한 반응, 혼합, 열전달 현상들을 수식화하기가 어렵다는 회분식 공정상의 특성으로 인해 많은 어려움이 있다. 이런 어려움을 해결하기 위해 제안된 것이 다량의 데이터베이스에 기반한 다변량 통계 방법이다.

MacGregor와 Nomikos는 측정된 공정 변수들을 이용한 회분식 공정 감시 방법을 제안하였다[10-12]. 복잡한 변수들과 시간의 상관관계를 고려한 Multiway Principal Component Analysis(MPCA)와 Multiway Projection to Latent Structures(MPLS) 등의 다변량 투영 방법을 도입하였다. MPLS는 회분 조업 진행시 수집되는 공정변수를 이용해 최종 품질을 예측하고 제품의 품질을 감시하기 위해 이용한다. Kourti 등은 공정변수와 품질변수에 초기 조건(set-up 데이터)들을 포함시켜 이상을 감시하는 Multiblock Multiway Projection 방법을 발표하였다[4]. MPCA 모델은 또한 회분식 공정 데이터를 분석하여 조업진행상의 주요 흔들림의 원인을 확인하는데 이용할 수 있는데, 그 흔들림의 원인을 파악하고 흔들림의 정도를 줄임으로써 일정한 양질의 제품을 생산할 수 있다[5]. 회분식 공정 데이터는 일반적으로 각 회분마다의 외란과 물리적인 제약 조건들로 인해 일정한 전환율에 대해서 동일한 양의 데이터를 갖지 않는다. Kassidas 등은 Dynamic Time Warping(DTW)을 이용해 측정값의 개수가 같지 않은 두 데이터를 동조화(synchronization)시킴으로써 회분식 공정의 분석, 감시 및 진단을 위한 데이터 전처리 방법을 발표했다[3]. 이러한 통계적인 공정감시 방법은 각 측정값들에 따른 변수의 contribution을 이용하여 진단을 좀 더 수월하게 할 수 있는 장점을 가지지만, contribution값들에 대한 또 한 번의 해석을 요하며 변수개수가 많을 때는 그 해석 또한 쉽지가 않다. 이러한 일련의 과정을 자동화하려는 연구가 기존의 진단방법과 결합하여 활발하게 일고 있는데 Venkatasubramanian과 Vedam 은 연속공정에서 PCA를 적용하고 발생한 이상의 contribution값들의 해석을 signed digraphs를 이용한 이상 진단 기법과 결합하여 공정 이상 감시 및 진단을 자동화한 논문을 발표하였다[13].

기본의 MPCA를 이용한 방법은 모델의 해석을 위해 이용되는 parameter/loading coefficient) 수가 너무 많고, 이로 인해 감시할 때 가장 중요한 과정이라 할 수 있는 미래 조업값의 예측 어려가 크다.

회분식 공정 특성상 시간대에 따라 covariance structure가 나누어지는 경향에 착안하여 전체 조업 시간을 임의의 몇 개 구간으로 나누어 국부 모델들을 구성함으로써 이러한 문제점들을 해결한 이상 감시 및 진단 방법을 제안하였다.

본 논문에서는 MPCA 국부 모델링을 이용한 이상감시 및 진단의 이론을 제시하고 공정 이상의 예측방법을 제안하였다. 국부 MPCA 모델들이 공정의 이상 감시 및 진단에 얼마나 효과적인지를 검증하기 위해 실제 산업계의 PVC 중합 반응기로부터 얻은 데이터를 이용하였다.

2. 이론 및 제안한 방법론

2-1. 다변량 통계 분석에 의한 이상 감시 및 진단의 이론적 배경

변수들 사이에 강한 상호 관계가 존재하는 다변수 데이터는 일반적으로 수 개의 주성분(PC)으로 데이터 내의 variation을 나타낼 수 있으며, 고차원의 데이터를 저차원으로 가시화함으로써 공정감시 및 진단, 제어기 디자인, 데이터 reconciliation 등으로 활용 가능하다. 그러나 PCA는 데이터의 stationarity가 보장되어야 하고 시간에 따른 auto-

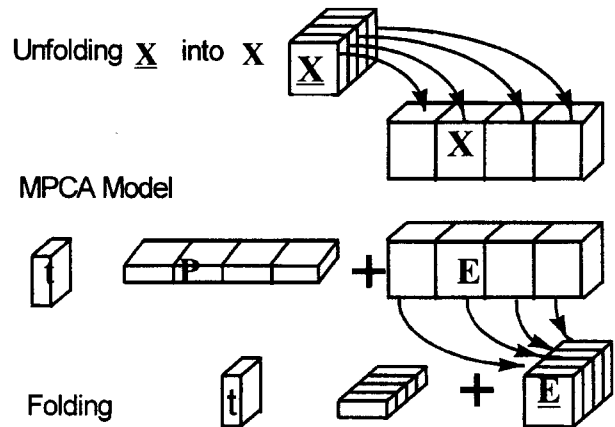


Fig. 1. MPCA modeling procedures.

correlation이 존재하지 않아야 한다는 한계점을 가지고 있다. 회분식 공정의 경우는 비정상 상태(unsteady state) 조업이 일어나므로 일반적인 PCA는 이용할 수 없다. 이러한 문제점을 해결하기 위해 Wold 등에 의해 도입된 것이 unfolding이다[14], 즉 회분들의 측과 변수, 그리고 시간에 대한 축으로 구성된 3차원의 데이터를 회분과 변수×시간으로 2차원화함으로써 autocorrelation을 제거하였다. Unfolding된 데이터 행렬에 PCA(Multiway PCA)를 적용함으로써 회분들의 variation을 모델링하고 통계적 공정감시 및 진단에 적용할 수 있다. 이러한 일련의 데이터 처리과정을 Fig. 1에 나타내었다. MPCA 모델링 결과 데이터는 회분들의 주요 분포를 설명하는 signal 부분과 모델에 의해 설명되지 못하는 noise 부분으로 나누어진다. 식 (1)에서 signal 부분은 score 벡터(t_r)와 loading 행렬(P)의 tensor product에 의해 표현되며, t_r 은 회분들에 대한 정보를, P 은 변수와 시간에 대한 정보를 포함하고 있다.

$$\mathbf{X} = \sum_{r=1}^R t_r \otimes P_r + \mathbf{E} \quad (1)$$

임의의 회분조업이 진행될 때 공정의 감시는 모델 데이터가 공정조업의 전형적인 정보를 모두 포함하고("comparable") 있어야 하고 공정상에 변화들(예를 들면, 새로운 촉매의 도입)이 생기면 새로운 데이터를 수집해야 하고, 각 센서에서 측정된 값들에 공정의 이상에 대한 정보가 포함되어 있다("observable"). 다변량 통계적 방법에 의한 공정 감시는 SPC(Statistical Process Control)의 기존적인 개념에 근간을 두고 있으며, 공정의 미래 동태는 과거에 관찰된, 공정이 정상적으로 진행했을 때의 동태와 비교함으로써 모니터링이 이루어진다. 감시 차트들의 제어한계선은 과거의 "good" 회분들을 수집하고 이들의 분포에 기반한 통계적 성질들로부터 구해진다. 이는 미래의 good 회분들은 과거의 좋은 것들과 유사한 동태를 가져야 한다는 생각에 기반한다.

공정감시는 signal과 noise 부분의 상호 보완적 관계를 통해 이루어지는데, 모델 데이터의 signal 부분(모델이 설명하는 부분)을 통해 현재 진행되고 있는 회분조업이 정상적인 signal로부터 어느 시간에 얼마나 벗어나고 있는지를, noise 부분을 통해 signal이 설명하지 못하는 부분에서 어떤 변화가 일어났는지를 알 수가 있다.

회분식 공정의 이상 감시는 회분 조업이 진행되고 있는 임의의 시점에서 이 회분이 조업이 끝난 후 좋은 제품을 생산할 수 있는지를 예측하여 비정상 제품이 생성될 경우라 판단했을 때 회분 조업을 정지시키던가 혹은 통계적 제어를 통해 정상제품이 생성되도록 하는 행동을 재빨리 취할 수 있도록 도와주는 기능을 한다. 먼저 이상감시를 위한 지표들을 알아보면, 모델의 t-score값과 Hotelling's Statistic을 통

해 이상 감시가 이루어지고 SPE값을 통해 모델의 신뢰성 정도나 공정의 NOC(Normal Operating Conditions)의 변화를 감지할 수 있다. 이때 Hotelling's Statistic은 회분들의 mean trajectory에서 얼마나 떨어진 회분인지를 t-score값을 기준으로 계산한 metric이고, SPE는 모델에 의해 계산된 데이터, $\sum_{i=1}^R t_i \otimes P_i$ 와 원래의 데이터, \mathbf{X} 의 차이를 제곱하여 합한 값이다. 이러한 index들은 sampling time에 따라 한 점씩 나타나는데, 현재의 시간까지 측정된 데이터를 중심으로 회분이 끝날 때까지 앞으로 측정될 값을 예측하여 한 회분 전체데이터를 구성한 후 위의 index들을 이용해 현재의 시간에서 이 회분이 정상으로 끝날 것인지 혹은 비정상적으로 끝날 것인지를 모니터링한다. 모델 데이터의 각 회분을 이러한 과정에 의해 off-line으로 모니터링하면서 데이터의 모든 회분들에 대해 t-score값(PC 개수/sampling time), Hotelling's statistics, SPE값을 구하고 이들의 확률분포를 계산하여 일정한 유의수준(α)으로 제어 한계값(control limit)을 정하게 된다. 완성된 제어 차트들의 기본적인 개념은 귀무가설(null hypothesis, H_0)과 대립가설(alternative hypothesis, H_1)로 구성된 가설검정(hypothesis testing)에 기반한다. 즉, 다음과 같은 가설을 검정한다.

H_0 : The process is in a state of statistical control

H_1 : The process is out of state of statistical control

t-score값의 경우 검정할 가설은 $H_0: t_i=0$, $H_1: t_i \neq 0$ 이다. Normality 가정에 Student t-분포의 검정통계량을 따른다. $\alpha=0.05$ 의 경우 귀무가설($t_i=0$)이 채택됐을 때 95%의 신뢰구간을 가지며 이 영역 안에 존재하면 in control을 의미한다. 유의수준(α)은 H_0 이 옳을 때 H_0 를 기각할 확률(제1종 오류)이며 허용 가능한 제1종 오류의 최대값이 유의수준이다. 즉 $\alpha=0.05$ 라 하면 100개의 측정값의 검정시 최대 5개 정도는 in control 상태인데 out of control이라고 판단할 수 있음을 의미한다. T^2 는 PC공간 중심($T^2=0$)에서 t-score 벡터까지의 거리제곱을 나타내는 값, 즉 $(t_1/s_1)^2 + (t_2/s_2)^2 + \dots + (t_R/s_R)^2$ 로 계산되며 t-분포의 제곱의 합의 형태이므로 이는 F-분포를 따르며 R(number of PC), I(number of batches)-R의 자유도를 갖는다. 다변수(t_1, \dots, t_R)의 평균에 대한 가설검정은 이 통계량을 이용한다. 따라서 T^2 statistic은 여러 개의 t-score값에 대한 SPC 차트들을 하나로 묶어서 표현하는 수단으로 이용된다. 신뢰구간은 t-score값의 경우와 동일한 방법으로 계산한다. SPE값의 경우는 식 (1)에서 E가 multinormal distribution을 따른다고 가정하면 SPE는 EE^T 로 표현되므로 이는 chi-squared 분포를 따르게 된다.

다변량 통계적 방법을 이용한 공정감시(Fault Id.)는 진단할 수 있는 능력을 갖는다는 점이 하나의 큰 장점이다. 진단의 과정은 다음과 같은 네 가지 단계로 이루어진다. Fault Identification → Fault Signature Identification → Fault Library Construction → On-line Fault Diagnosis가 그 과정이다. 진단으로의 연결을 위해 도입된 것이 signal 부분에서 감지된 이상에 대한 각 변수들의 contribution이다. 임의의 시간에서 contribution은 loading coefficients에 절대값을 취한 후 도입되는 변수(unfolding 상태)의 측정값과 곱하고 원래 변수별로 합하여 나타낸다. 이것 또한 모델 데이터를 이용해 제어 한계값을 구한다. 각 시간과 변수별로 위 가설을 이용하여 동일한 방법으로 구하면 회분 조업이 진행되면서 비정상적으로 감지되는 변수를 선택할 수 있다. 이 변수의 추이와 다른 변수로의 fault propagation을 통해 assignable cause를 찾아낸다(Fault Signature Id.). 일반적으로 회분식 공정의 경우는 발생하는 이상이 그리 많지 않으므로 fault library는 운전원들의 경험적 지식에 기반하여 contribution 차트에서 알아낸 변수들의 추이나 다른 변수로의 영향을 통해 fault의 원인을 알아낸다.

2-2. 제안된 방법론

화학공학 제37권 제3호 1999년 6월

국부 모델을 구성하고 이상을 진단하는 과정은 다음과 같다.

(1) 전체 모델을 구성한다. 회분 전시간에 해당하는 모델을 구성하면 몇 개의 PC가 생성되는데 각 PC에 의미를 부여하고 이는 이상 진단시 중요한 정보로 이용된다.

(2) 표준이 되는, 즉 가장 mean trajectory를 따르는 회분(변수×시간의 데이터 행렬)에 PCA를 적용하여 축소된 공간에서 score값들의 추이를 통해 시간대의 개수와 길이를 정한다.

(3) (2)에서 정해진 개수와 길이로 각 국부 모델 데이터를 이용해 국부 모델들을 구성한다.

(4) 전체 모델과 국부 모델에 적합한 future value prediction 방법을 선정한다. Future value prediction 방법으로는 세 가지가 제안되어 왔다. Future value를 모델에 이용한 데이터들의 mean trajectory와 일치하게 하는 방법, 공정에 한 번 발생한 외란이 조업 내내 지속적으로 존재하게 되면 외란이 발생한 시점에서 mean trajectory로부터 차이가 future value에 그대로 지속된다고 가정하고 예측하는 방법, PCA가 missing value를 해결할 수 있다는 점을 이용하여 아직 측정되지 않은 공정 변수를 missing value로 취급하여 예측하는 방법이 있다.

(5) 선정된 예측방법을 이용하여 실시간 공정감시를 위한 scheme을 전체 모델과 국부 모델들에 대해 2-1절의 가설검정을 이용하여 수립한다. 전체 모델에서 발생한 이상은 국부 모델을 통해 검증되고 어느 시간대에서 이상이 발생하였는지에 대한 정보를 얻게 된다.

(6) 전체 모델과 국부 모델에 대한 contribution값들을 이용해 이상의 원인이 되는 변수를 파악하고 국부 모델을 통해 이상변수를 검증하고 이상의 원인규명을 위한 상세정보를 얻는다. 운전원의 경험에 의한 fault library가 구축되어 있다면 임의의 fault와 library에 구축된 fault들과의 pattern 인식을 통해 진단이 가능한데 회분식 공정의 진단시 문제점 중에 하나가 회분 초기 비정상 signal이 회분 중반이나 말기 때의 fault와 pattern matching되는 경우가 있다. 각 국부모델을 이용하면 이를 해결할 수 있다.

3. 사례연구

3-1. 대상 공정

대상 공정은 PVC(Polyvinyl Chloride) resins을 제조하는 회분식 중합 반응기이다. 일반적으로 PVC는 현탁중합(suspension polymerization) 공정이며 물을 매질로, VCM(Vinyl Chloride Monomer)과 유용성 개시제 용액을 현탁안정제와 교반력에 의해 미세한 입자로 만들어 개개의 입자가 한 개의 과상중합(VCM과 개시제만의 중합) 반응계로서 독립적으로 중합을 진행시키는 회분식 공정을 일컫는다. VCM중합 반응은 자유 라디칼 중합에 의한 발열 반응이다. 물은 VCM을 분산시키는 매체일 뿐만 아니라 중합도중에 발생하는 중합열을 외부로 전달하는 역할을 한다. 전체 공정은 중합 반응기(Fig. 2)와 stripping column, 탈수 및 건조부, VCM 회수부로 나뉜다. 중합 반응기에서 반응이 종료되면 slurry(PVC resin 입자들이 물에 분산되어 있는 것)를 blow down tank를 거쳐 미반응 VCM을 회수하고 stripping column으로 이동시킨다. 수지에 미량 남아있는 VCM을 제거하고 탈수 및 건조시킨다. PVC 현탁 중합 반응기의 조업 사이클은 다음과 같다. 중합기에 물과 부원료를 넣고 특정한 온도까지 가열한 후 VCM을 함께 중합기로 투입한다. 원료를 넣은 후에 교반을 시작하고 중합기 jacket으로 특정온도의 냉각수를 넣어 냉각하여 일정한 온도를 유지하며 중합을 진행한다. 중합이 전환율에 의해 끝났음을 확인한 후 반응기로부터 slurry를 뽑아내고 cleaning 및 antifouling 작업을 행한다[1,9].

위 공정은 스티프, 냉각수, 급냉수의 공급량의 3개의 조작변수를 갖는다. 스티프는 반응이 일어나기 전에 반응 온도를 최소의 시간에 원하는

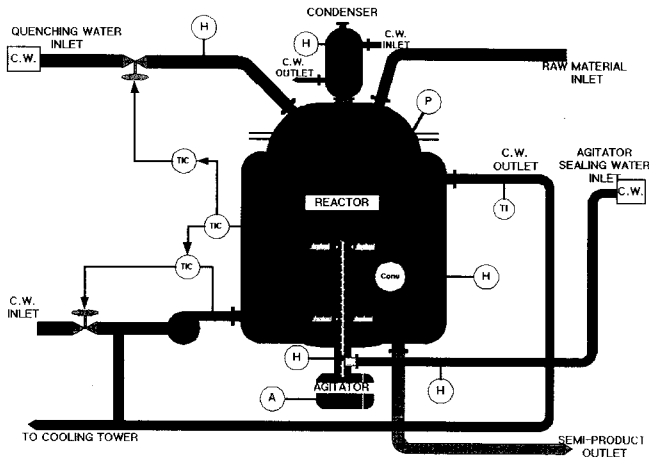


Fig. 2. PVC polymerization reactor.

Table 1. Process measurement variables

No.	Measurement	No.	Measurement
v1	Reactor temp.	v6	Quenching water heat
v2	Jacket inlet temp.	v7	Jacket cooling water heat
v3	Jacket outlet temp.	v8	Agitator heat
v4	Reactor pressure	v9	Agitator sealing water heat
v5	Conversion	v10	Reflux condenser cooling water heat
		v11	Agitator ampere

수준까지 올려서 반응 물질을 가열시켜주고, 냉각수는 중합열을 제거 해주며, 급냉수는 반응열을 제거시켜줌과 동시에 단량체(monomer)와 고분자와의 밀도차에 의한 부피감소를 보충해 줘서 반응물질의 부피를 일정하게 유지시켜준다. 또한 스팀에 의한 가열 작업 후에 반응 온도와 공정의 외란을 제거해 주기 위해 캐스케이드 제어가 이용된다.

데이터는 현장에 구축된 PIS(Plant Information System)로부터 11개의 공정변수에 대해 360개의 측정값(sampling time은 1.5분)들을 얻었다. 측정 변수들은 Table 1에 나타내었다. 이 측정값 중 batch 조업 초반(원료의 투입 및 가열기간)과 후반(blow down)의 데이터는 불규칙한 센서값의 분포로 인해 제거하고 중합 반응이 일어나는 시간동안 측정된 241개의 데이터를 사용하였다. 중합 반응 동안의 측정값들은 모든 회분에 대해 일정한 반응의 정도를 나타내기 때문에 각 회분의 synchronizing은 필요가 없다[3]. Fig. 3은 이렇게 얻어진 21개 batch에 대한 반응기의 온도, 압력, agitator ampere를 나타냈다. 이 데이터

Table 2. Model overview

No. of PC	Explained variance	Eigenvalue
1	0.365	7.669
2	0.581	4.537
3	0.710	2.707

는 모두 조업이 끝난 후 실험실에서의 분석을 통하여 최종 제품의 물성을 만족시킨다. 18번 batch의 경우는 최종 제품의 물성을 만족시킴에도 불구하고 온도와 압력의 경향이 다른 정상 batch들과는 좀 다를 수 있다. 이 회분을 outlier라 판단하여 모델 데이터에서 제거시킬 수도 있지만 이 회분 또한 조업자들의 조작이 반영된 것이므로 모델에 포함시켰다.

3-2. 모델 구성 및 분석

3-2-1. 전체 모델 구성 및 각 PC의 의미

위에서 설명한 회분식 반응기에 대한 공정 감시 및 진단 시스템을 MPCA 모델링 기법을 이용하여 구축했다. 모델링에 이용한 데이터는 21개의 정상 회분이며, 각 회분은 11개의 변수(열)와 241번의 측정값(행)으로 이루어진 데이터 행렬이다. 3차원 배열의 데이터는 비선형성을 제거하고 PCA 알고리즘을 사용하기 위해 각 시간에 따라 unfolding 시켜서 각 행렬의 열들을 centering, scaling한다. 이 행렬에 NIPALS (Nonlinear Iterative Partial Least Square) 알고리즘을 적용하여 3개의 PC(고유벡터값을 통한 cross validation)를 구했으며, 각 PC가 모델을 설명하는 정도(explained variance)와 고유값(eigenvalue)은 Table 2에 정리하였다.

구성한 모델의 각 PC에 대한 loading 벡터들은 공정의 중요 variance를 설명하는 가중치역할을 하므로 각각에 물리적 의미를 부여할 수 있다. Fig. 4는 각 PC에 해당하는 loading coefficients를 변수별로 분류한 것이다.

먼저 첫번째 PC(a)의 경우 가장 영향을 크게 미치는 요인은 jacket 으로 도입 및 배출되는 온도가 낮을수록 그 열량은 클수록 t1-score 값이 커지므로 jacket heat load(JHL이라 표기함)라는 물리적 의미를 부여하면 타당하다. 두 번째 PC(b)의 경우는 quenching water heat과 sealing water heat이 가장 큰 영향력을 행사한다. 각 열량이 작을수록 t2-score값이 커지므로 reactor 내의 heat load(RHL이라 함)로 정의한다. t3-score는 중반부의 시간대에서 반응기 온도와 압력이 커질수록 중반부에서는 작아질수록 커지며 특히 중반부의 전환율이 커질수록(중반부에 비해 상대적으로 더 크다) score값이 커지므로 반응속도로 물리적 의미를 정의할 수 있다. 그러나 각 loading coefficients

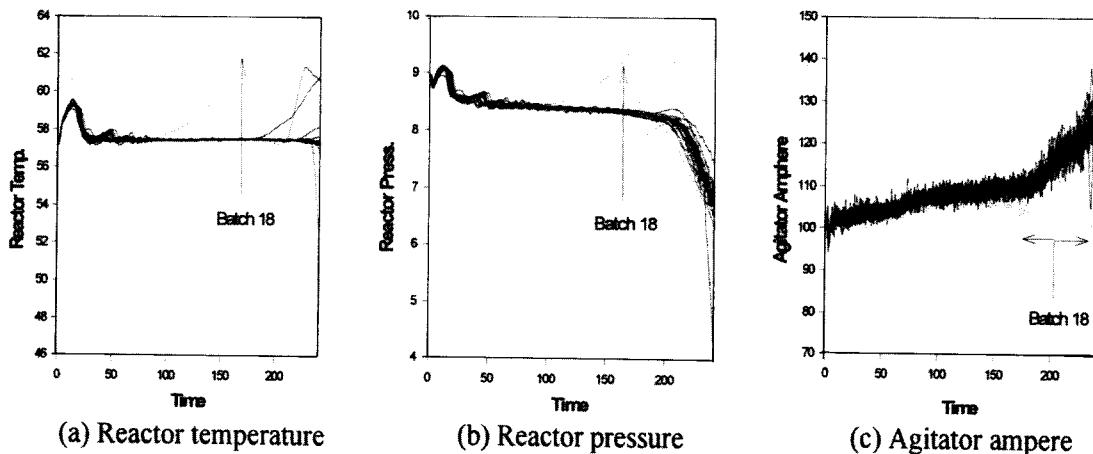


Fig. 3. Process measurement trajectory of reference batch.

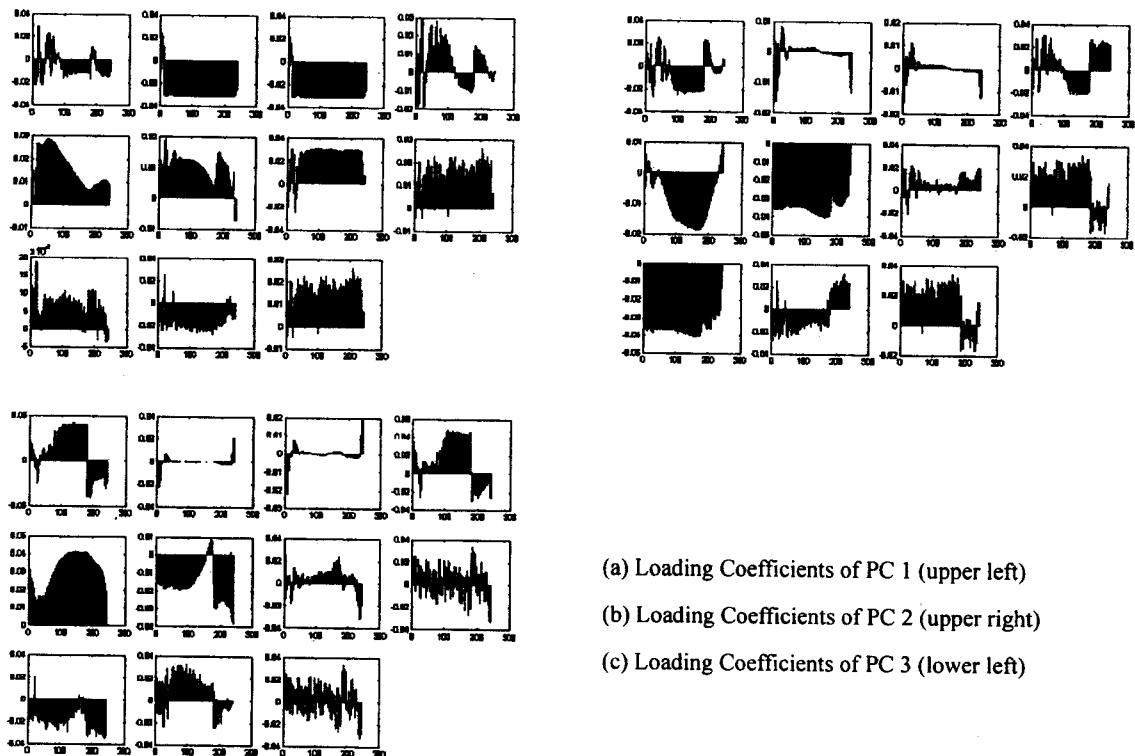


Fig. 4. Loading coefficients of each PC classified by variable(v1-v11: from left to right).

는 어떤 시간대에서 각 PC에 따른 물리적 값이 우리가 정의한 방향과 반대방향으로 진행된다(Fig. 4의 (a)에서 v2, 3의 경우 초기 시간에서 정의한 방향과 반대방향으로 진행시키는 시간대가 존재함). 이는 시간대에 따른 모델을 구성함으로써 해결할 수 있다.

3-2-2. 국부 모델 구성 및 PC의 의미

국부 모델을 구성하기 위해서는 일단 몇 개의 국부 모델이 필요한지와 각 모델의 시간대를 알아야 한다. 이를 위해 도입된 것이 3차원 데이터의 또 다른 unfolding이다. 각 시간의 추이를 알아내기 위해 $\text{sampling time} \times [\text{변수} \times \text{회분}]$ 형태의 행렬을 도입해 변수들의 추이를 score값으로 압축하여 나타낸 후, 시간대 별로 분류한다. Fig. 5는 score 벡터들을 시간에 따라 나타내었다. 조업 초반부에 보이는 심한 비선형성의 궤적을 1-50시간에서 표현하였고 조업 중반부에 일정한 궤적을 보이는 부분을 51-200시간에, 조업 후반부의 변화 부분을 201-241시간에서 표현하였다. 이 plot에서 알 수 있는 또 하나의 정보는 각 변수의 추이와 각 t-score값의 추이를 비교하면 t1-score는 jacket의 도입 및 출력온도의 trajectory와 거의 유사하고 이는 PC의 의미를 고려할 때 주요 영향을 끼쳤던 변수와 일치한다. t2-score와 t3-score는 각각 sealing water heat와 반응기 온도 및 압력에 대응하며 이 또한 PC의 의미에서 생각했던 변수들과 일치한다. 결국 이 모델을 이용해 위의 PC의 의미를 검증할 수 있다.

Fig. 5에서 정해진 시간대에 따른 모델을 Table 2에 서술한 형식으로 구성하면 Table 3과 같다. 각 모델들을 비교해 보면 중반부에서의 모델이 가장 큰 explained variance값을 갖는다. Fig. 3의 18회분에 대해서 보는 것처럼 대부분의 공정의 흔들림은 중반부에서 잘 일어나기 때문에 국부 모델로 나누었을 때 모델의 공정감시가 좀더 강건해질 수 있다.

각 time interval에서 각 PC의 의미를 전체 모델에서 예측했던 것처럼 부여하고 전체 모델에서 밝혔던 의미들을 정리하면 Table 4와 같다. 각 t-score값이 양의 방향으로 진행할 때 어떤 물리적인 현상이

어떻게 진행되는지를 보여준다. 특히 국부 모델들의 경우는 각 PC에 따라 공정에 어떤 물리적 현상이 일어나는지 비교적 정확히 보여준다. 각 경우에 대해 JHL은 jacket의 heat load를, RHL은 반응기 내의 heat load를 의미한다. ReR은 반응속도의 축약형이다. 압축공간에서 PC축을 중심으로 움직이는 축값들이 어떠한 물리적 의미를 갖고서 움직이는지를 국부 모델을 이용해 각 time에 따라 정확히 파악하면 그 만큼 공정이상 감시 및 진단이 수월해지게 된다.

3-3. 적절한 Future Value 예측방법의 선정

MPCA모델을 이용한 공정감시 및 진단 scheme을 구성할 때 필요한 구체적인 알고리즘으로써 가장 핵심적인 부분은 future value의 예측과 관련된 것이다. 일반적으로 변수들의 time unfolding을 거쳐 데이터 행렬을 구하게 되는데 이와 같은 작업은 이미 회분의 조업이 끝난 측정 데이터를 off-line으로 test할 때는 문제가 없으나 실제 감시에서는 on-line으로 각 시간마다 데이터가 들어오게 되므로 이 시각 이후의 측정되지 못한 데이터에 대한 적절한 예측값이 필요하게 된다.

일반적으로 회분식 공정의 감시에 사용하는 future value의 예측 방법에는 세 가지가 제안되어 왔다. 첫째, k시간 이후부터 조업 종료 시까지의 미래값들이 모델로 구성된 데이터들의 mean trajectory와 완벽하게 일치한다고 가정하여 mean-centering된 데이터에 대하여 아직 측정되지 않고 있는 미래값들을 '0'으로 대체하는 방법이 있을 수 있다. 두 번째는 공정에 발생되는 외란이 조업 내내 지속적으로 존재하게 되면 모델의 mean trajectory로부터 벗어나는 정도가 유사하다는 점에 착안하여 아직 측정되지 않은 미래값들에도 현재의 편차가 그대로 지속된다고 가정하는 것이다. 세 번째는 PCA가 missing value를 해결할 수 있다는 점을 이용하여 아직 측정되지 않은 공정 변수를 missing value로 취급하여 예측하는 방법이 있다. 위의 세 가지 방법들은 각각 장점과 단점을 가지고 있어서 어느 한 방법만을 사용하는 것보다는 여러 방법들을 적절하게 혼합하여 사용하는 것이 보

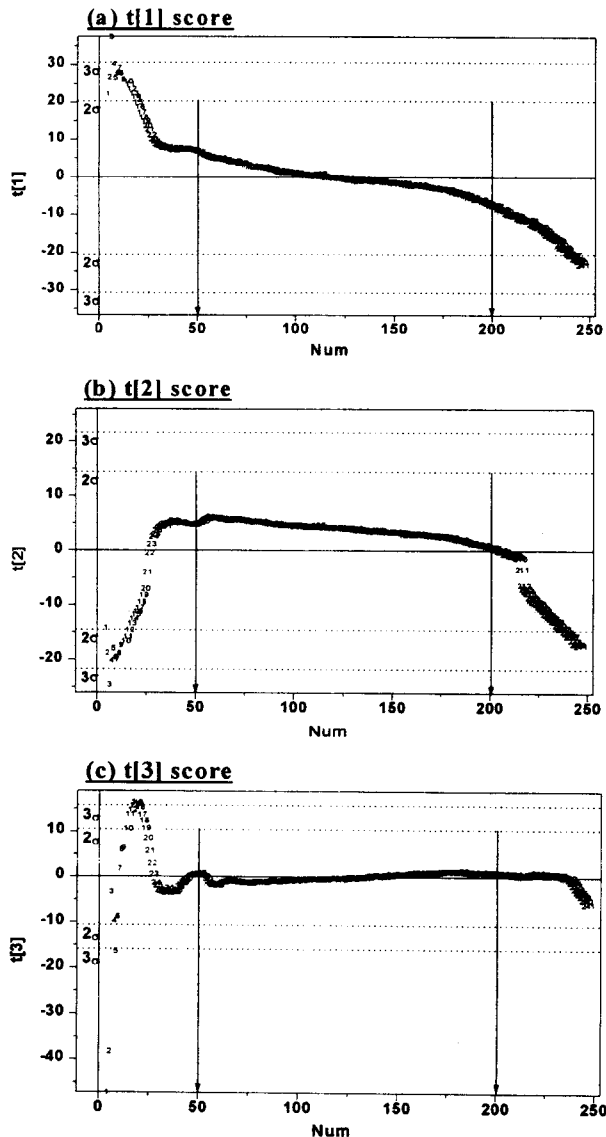


Fig. 5. Determination of three time intervals for local models.

다 정확한 예측값을 주는 것으로 알려졌다. 그러나 일반적으로 미래 값들을 '0'으로 대체하는 방법은 모니터링 목적상 적합하지가 않다. '0'

을 대체한다는 의미는 모든 회분의 미래값을 평균값으로 대체하는 것을 의미하므로 비정상 회분이 발생했을 때도 미래값의 예측은 평균으로 대체된다. 즉, 비정상회분의 미래값을 평균, 즉 정상회분으로 여기면서 예측하기 때문에 이상의 감시를 위해서는 부적합하다.

전체 조업 시간을 각 영역별로 나눴을 때의 가장 큰 장점은 미래 값의 예측과 관련이 크다. Fig. 6은 전체 모델과 국부 모델을 사용했을 때 세 가지 예측방법을 이용했을 때의 에러값의 차이를 37개-good batch 24개(batches 1-24), 기타 5개(batches 25-29), bad batch 8개(batches 30-37)의 회분들에 대해서 보여주는 그림이다. Fig. 6의 (a)와 (b)를 비교하면, (b)의 전체 에러값이 차수의 크기(magnitude of order)가 한 자리 더 큼을 알 수 있다. 다시 말해서 전체 조업 시간에 대하여 하나의 모델을 구성하게 되면 조업 초기에는 측정되어 있는 데이터보다 더 많은 데이터를 예측하여야 하므로 위의 세 방법 중 어떤 것을 사용하더라도 큰 오차를 보이는 반면 각 조업 시간 영역에 대한 모델을 사용하게 되면 그만큼 오차의 크기를 줄이면서 정확한 예측이 가능하기 때문이다.

세 가지 방법들을 비교해 보면 good batch일 경우는 방법 3(PCA)이 방법 2(current deviation)에 비해서 오차가 더 작지만 bad batch일 경우에는 방법 2가 3보다 오차가 훨씬 작음을 알 수 있다. 본 연구에서는 공정의 이상 감시를 수행할 때 bad batches들을 분별하는 것이 주 목적이므로 방법 2를 이용할 수 행하였고 이를 국부 모델에 적용함으로써 좀더 정확한 감시 scheme을 구성할 수 있었다.

3-4. 공정감시 및 진단기술의 적용

이상 감시 및 진단 차트들의 통계적 의미를 볼 때 각 차트들은 0.05와 0.01의 유의수준에 의해 이상과 정상의 판단을 하게 되는데 0.05 유의수준이라 하면 100개의 도입된 측정값에 대해 5개 정도는 정상인데 이상이라고 판단할 가능성을 전제한 기준이다. 따라서 어떤 공정이 제어영역 밖으로 벗어났음을 확인하는 규칙으로는 일반적으로 다음과 같은 세 가지가 있다.

- (1) 95 % 신뢰도 구간(i.e. 0.05 유의수준; warning limits)을 벗어나는 값이 두 포인트 이상이 발생했을 때
- (2) 99 % 신뢰도 구간(i.e. 0.01 유의수준; action limits)을 벗어나는 값이 한 포인트 이상이 발생했을 때
- (3) 일곱 포인트가 연속으로 증가하거나 감소했을 때

위의 세 가지 기준에 의해 이상이라고 판단한 후 진단하려면 이상이라고 판단한 점(측소 공간상의 score값)을 원래의 공정변수로 환원

Table 3. Local models overview

Section No. PC	Time interval 1(1-50)		Time interval 2(51-200)		Time interval 3(200-241)	
	Exp. var.	Eigenvalue	Exp. var.	Eigenvalue	Exp. var.	Eigenvalue
1	0.358	7.512	0.385	8.090	0.326	6.845
2	0.593	4.946	0.628	5.107	0.525	4.179
3	0.694	2.120	0.776	3.092	0.699	3.661

Table 4. Physical meaning of each PC in the all models

(JHL : Jacket Heat Load, RHL : Heat Load within the Reactor, ReR : Reaction Rate)

Models PC score	Total model	Local model 1(1-50)	Local model 2(51-200)	Local model 3(201-241)
t1-score(+)	JHL(+)	" 1-16 "; JHL(-) " 17-50 "; JHL(+)	JHL(+)	JHL(-)
t2-score(+)	RHL(-)	RHL(-)	RHL(+)	RHL(-)
t3-score(+)	Reaction rate(+)	" 1-16 "; ReR(+) " 17-35 "; ReR(-)	ReR(-)	ReR(+)

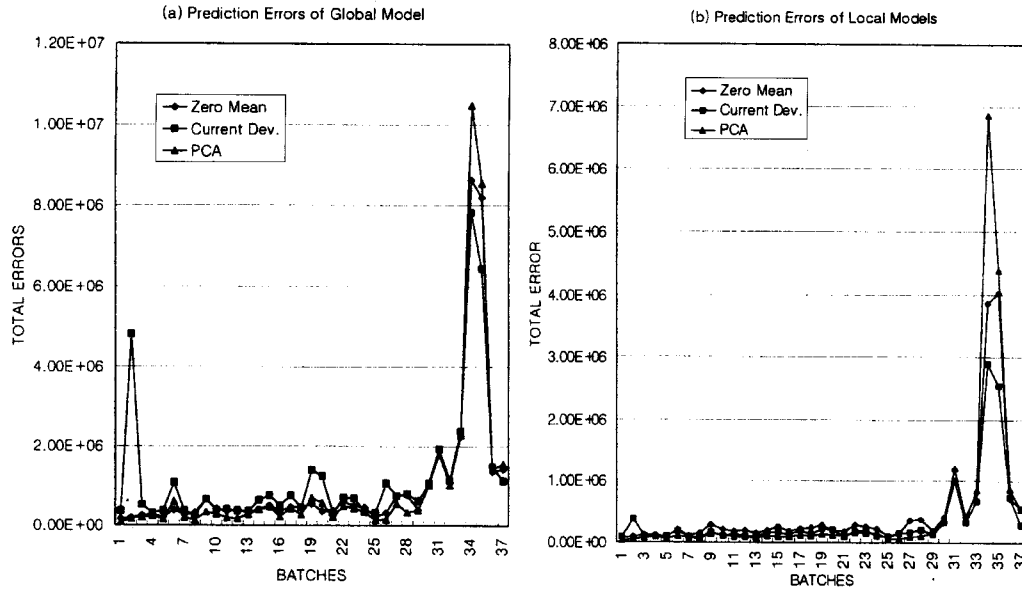


Fig. 6. Comparison of prediction errors of global model with that of local models.

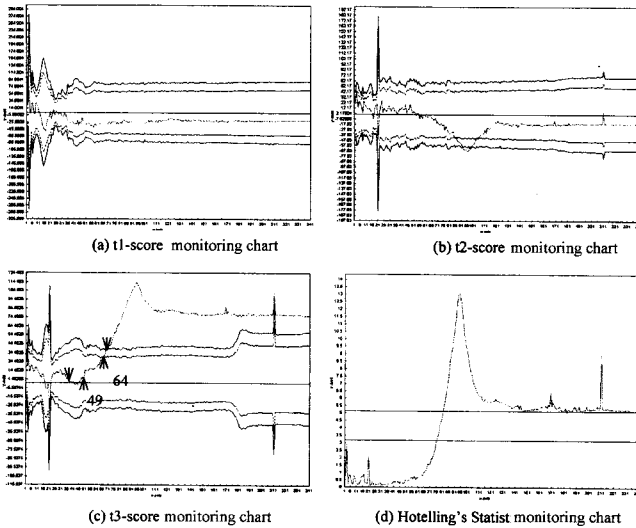


Fig. 7. Monitoring charts for an abnormal batch.

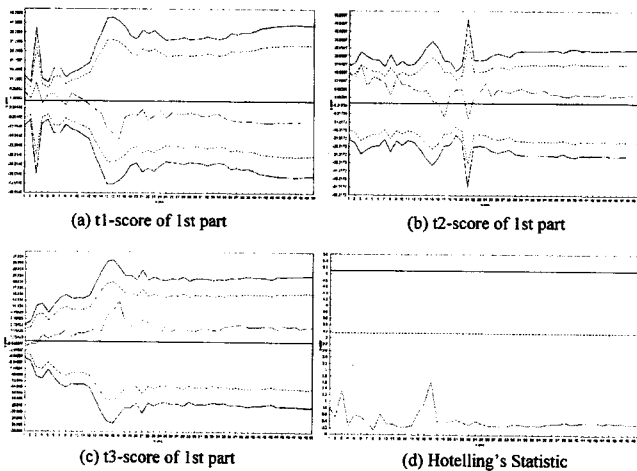


Fig. 8-1. Monitoring charts of an abnormal batch on the 1st time interval model.

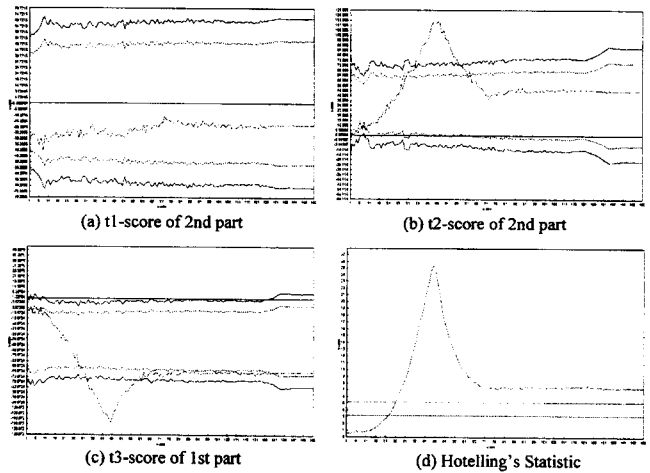


Fig. 8-2. Monitoring charts of an abnormal batch on the 2nd time interval model.

시켜줘야 하는데 이때 사용하는 것이 바로 각 변수의 contribution이다. Contribution값들을 통해 어떤 변수들이 제어 한계선을 벗어나게 하는 주요 원인에 영향을 미쳤는지를 알 수 있고, 이상이 발생한 시점부터 이상이 전파된 경위를 공정의 비정상 상태 중의 하나로 간주하고 fault library에 저장한다. 물론 fault library에는 조업자의 지식이나 경험을 바탕으로 각 fault에 대한 적절한 조치에 대한 정보가 포함되어 있어야 한다.

이제 어떤 비정상 회분의 경우에 대해 국부 모델이 얼마나 강건한지를 모니터링 차트들을 보고서 감지 및 진단하는 과정을 알아보면, Fig. 7은 전체 모델에 대한 비정상회분(비교적 조업초반에 이상이 일어남)의 감시차트를 Fig. 8은 국부 모델(전반부와 중반부)을 이용한 이 회분의 감시차트를 나타내고 있다. 3-3절의 Fig. 6에서 보인 것처럼 future value prediction상에서 에러값의 차이가 확실히 드러나므로 국부 모델을 이용하였을 때가 모니터링을 좀더 정확하게 수행하게 된다. 공정의 감시는 초반부에는 전체 감시차트와 첫 번째 국부 모델의 감시차트를 통해, 중반부에서는 두 번째 국부 모델 감시차트를 이용한다. Fig. 9와 10에는 전체 감시차트와 국부 감시차트의 t3-score에 대한

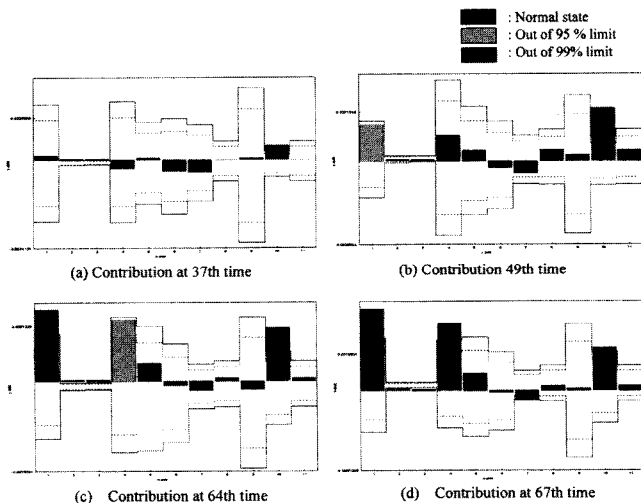


Fig. 9. Contribution charts of total model.

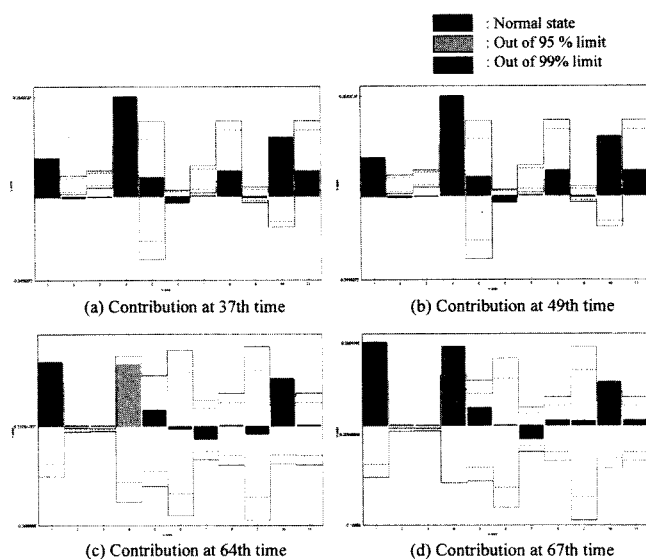


Fig. 10. Contribution charts of local model.

contribution값을 몇 가지 시간에 따라 나타내었다.

여러 차트들을 이용해서 적절한 진단을 내리는 과정은 다음과 같다.

(1) 각 t-score값들 중 이상에 민감하게 반응하는 것을 선택한다. Fig. 7과 8에서 세 번째 t-score 차트가 가장 먼저 제어 한계선을 넘어섰다. 다음, 두 번째 score 차트가 벗어났음을 보고서 일단 세 번째 PC의 전파에 의해 두 번째 PC가 영향을 받았음을 알 수 있다. 그러나 바로 두 번째 차트에서 감지된 이상들은 두 번째 PC와 밀접한 변수들이 적절하게 잘 조작되어 바로 제어한계선 안으로 들어가지만, 세 번째 PC 감시차트에서는 계속해서 이상조업이 감지되고 있다. Table 4의 각 PC의 의미로 생각해 보면 7(c)에서 조업 초반부에 반응속도가 불규칙적으로 움직이고 있음을 알 수 있고 8-1(c)에서는 반응이 계속해서 활발하게 일어나고 있음을 알 수 있다. 8-2(b)와 (c)에서 갑자기 t2-score는 양의 방향(반응기 내 제거열량의 증가)으로 조업이 진행되면서 t3-score는 음의 방향(반응속도 증가)으로 급증하였다. 후에 증명하겠지만 이는 응축기에 과부하를 초래하게 된다.

(2) 감시차트들 중 t3-score와 그것의 contribution을 나타내는 Fig. 9와 10을 비교해 봄으로써 이상을 일으킨 근본적인 변수를 선정할 수

있고, 시간에 따른 contribution을 검토해 봄으로써 비정상 회분의 trend까지 파악할 수 있다. Figure에서 얻은 회색은 유의수준 0.05를 진한 회색은 유의수준 0.01을 넘었을 때 나타난다. 이상의 trend는 정상적인 정보(명암의 변화)로써 nominal 변수로 저장하고 하나의 fault library로 정해진다. Fig. 9에서 보면 37번째 시간에서 환류 응축기의 냉각수 열량이 이상, 즉 큰 variation을 나타냄으로 인해 반응기 온도와 압력에 영향을 미쳤으며(b, c) 이는 공정 전체에 영향을 끼치게 된다. Fig. 10에서 보면 같은 시간(a)에 국부 모델 감시차트는 9(a)에서 응축기의 이상을 감지하는 시간에 벌써 온도나 압력으로의 영향을 행사하고 있음을 보여준다. 국부 모델의 성능을 전적으로 보여주는 예라 하겠다. 즉, 반응기의 열을 제거시켜 주는 장치는 가장 중요한 것이 jacket이고 급냉수와 교반기의 sealing water, 응축기가 제열을 도와준다. 응축기의 load가 갑자기 줄어들거나 늘어나면[(1)에서 늘어났음을 의미] jacket의 냉각수 유량과 급냉수의 유량 제어를 위해 제어기가 활발한 작동을 하게 되고 stability 영역을 벗어나면 공정은 굉장히 불안정하게 된다.

(3) 근본적인 변수와 관계되어 있는 장치들이 잘 발생하는 failure를 점검함으로써 이상에 대한 corrective action을 취한다. 응축기의 경우는 fouling을 피하는 것이 급선무라 하겠다. 왜냐하면 응축기는 cleaning이 쉽지 않을 뿐더러 반응기의 효율을 높이는데 큰 역할을 하므로 cleaning동안 응축기 사용이 자유롭지 않으면 큰 경제적 손실을 가져올 수 있다. 이러한 진단은 fault library가 완전히 구축되면 운전원의 공정지식베이스와 결합하여 전문가 써시스템화할 수 있고 적절한 지침방향까지 정리되면 완전 자동화할 수 있게 된다.

4. 결 론

본 연구에서는 데이터를 축소된 공간으로 압축하고 투영시킴으로써 유용한 정보를 추출할 수 있는 다변량 통계 투영 방법을 기반으로 공정의 이상 감시 및 진단을 수행하였다. 좀더 정확하고 신속한 감시 및 진단을 위해 전체 회분시간을 시간대별로 나누어서 각각의 시간대에 대해 모델을 구성하였다. 이 방법은 조업상 큰 비선형성을 갖는 초반부와 중반부를 따로 모델링함으로써 PC의 의미를 시간에 따라 좀더 구체적으로 부여할 수 있고, future value를 예측할 때 오차를 줄여줌으로써 정확한 공정의 감시 및 진단을 가능하게 한다. 사례 연구를 통해 전체 모델에서는 이상으로 감지되지 않은 것이 국부 모델에서는 심각한 이상으로 나타난 회분을 예로 들어 국부 모델의 유용성을 설명하였다. 또한 이러한 다변량 통계 방법을 통해 얻은 t-score 값에 대한 각 변수의 기여도를 패턴화함으로써 회분식 공정의 진단에 효과적으로 활용할 수 있는데, 국부 모델을 이용하면 같은 패턴이라도 시간에 따라 달라질 수 있는 각 패턴을 시간대의 분리에 의해 매칭함으로써 이상 진단을 좀더 강건하게 할 수 있다. 즉, 초기의 이상 패턴이 말기와 매칭되는 것을 방지시켜준다.

정확성을 높인 국부 모델에서 감지된 contribution의 추이는 기존에 발표되어 왔던 여러 진단 방법들과 결합하여 운전원들에게 쉽고 정확하며 신뢰성있는 운전 가이드를 제공할 수 있고, 이는 완전한 이상 진단 자동화를 위한 전문가 써시스템의 형태로 구현될 수 있다.

감 사

본 연구를 위해 포항공대 공정산업의 지능자동화 연구센터를 통해 재정적 지원을 해주신 한국과학재단에 감사를 드립니다.

사용기호

E : residual matrix of a new batch

\underline{E} : residual matrix for historical database
 $F_{R,I,R}$: F-distribution with R and I_R degrees of freedom
 h : degree of freedom for chi-squared distribution
 H_0 : null hypothesis
 H_1 : alternative hypothesis
 i : index for batches
 I : total number of batches
 J : index for measurement variables
 J : total number of variable measurements
 k : index for time intervals
 K : total number of time intervals
 p : loading vector
 P : loading matrix
 r : index for principal components
 R : total number of principal components
 s : diagonal element of covariance matrix of t scores
 SPE : squared prediction error
 t : score for a new batch
 \underline{t} : score vector
 T^2 : Hotelling's statistic
 \underline{X} : historical database

그리스 문자

α : level of significance

참고문헌

- Burgess, R. H.: "Manufacture and Processing of PVC," Applied Science publishers(1980).
- Johnson, R. A. and Wichern, D. W.: "Applied Multivariate Statistical Analysis," 3rd ed., Prentice Hall(1992).
- Kassidas, A., MacGregor, J. F. and Taylor, P. A.: *AIChE J.*, **44**(4), 864 (1998).
- Kourtí, T., Nomikos, P. and MacGregor, J. F.: *J. Proc. Cont.*, **5**(4), 277(1995).
- Kosanovich, K. A., Dahl, K. S. and Pivoso, M. J.: *Ind. Eng. Chem. Res.*, **35**, 138(1996).
- Lopes, J. A. and Menezes, J. C.: "Faster Development of Fermentation Processes. Early Stages Process Diagnosis," FOCAPO, Identification/Diagnosis Part(1998).
- MacGregor, J. F., Jackle, C., Kiparissidas, C. and Koutoudi, M.: *AIChE J.*, **40**, 826(1994).
- Marlin, E. B., Morris, A. J. and Zhang, J.: *IEE Proc-Control Theory Anal.*, **143**(2), 132(1996).
- Nass, L. I. and Heiberger, C. A.: "Encyclopedia of PVC," 2nd ed., Marcel Dekker(1986).
- Nomikos, P. and MacGregor, J. F.: *Technometrics*, **37**(1), 41(1995).
- Nomikos, P. and MacGregor, J. F.: *AIChE J.*, **40**(8), 1361(1994).
- Nomikos, P. and MacGregor, J. F.: *Chem. and Int. Lab. Sys.*, **30**, 97 (1995).
- Vedam, H. and Venkatasubramanian, V.: "Automated Interpretation of PCA-based Process Monitoring and Fault Diagnosis using Signed Digraphs," IFAC Conf., Session 4A(1998).
- Wold, S., Geladi, P., Esbensen, K. and Ohman, J.: *J. of Chemometrics*, **1**, 41(1987).