

다차원 독립변수 분석을 통한 SBR 공정 모니터링 및 제어 Part. II

(Integration of process monitoring and supervisory control of SBR process)

유창규

본 연구 기본 아이디어는 Approach B에 기반 한 방법을 택했으며 Approach A의 장점을 약간 접목시키고 여기에 주성분분석보다 독립요소분석(Independent Component Analysis)을 적용하는 모니터링 방법을 제시하였다.

대부분의 공정에서 측정된 데이터들은 차원이 크고 각 변수간의 상관성이 크기 때문에 데이터 축약을 위해선 잠재해 있는 독립 요소를 찾아내야 한다. 독립요소분석은 이런 독립 요소를 찾아내는데 유용한 방법이 된다. 독립요소분석은 주성분분석의 확장된 개념으로 볼 수 있다. 주성분분석이 평균과 분산에만 기반하여 변수들의 상호 관계를 없앤 축을 찾아내어 분석하는 방법에 비하여 독립요소분석은 평균 및 분산뿐만 아니라 첨도(kurtosis), 비대칭도(skewness) 등 고차원의 통계학적 성질에 기반 하여 변수들의 상호 관계를 없애기 때문에 다변량 데이터에서 독립적인 요소만을 따로 뽑아낼 수 있다. 더구나 주성분 분석에 기반하여 모니터링을 하게 되는 경우 차트의 제한선을 설정할 때 주성분분석 후 나오는 스코어 값들이 정규분포를 따른다는 가정 하에 구하기 때문에 문제점이 있게 된다. 한편, 독립요소분석은 처음부터 다변량 데이터 속의 독립 요소들이 정규분포를 따르지 않는다는 가정을 한 채 찾기 때문에 실제 데이터를 처리하는 데 있어서 많은 장점들을 가질 수 있다. 그림 2은 주성분분석과 독립성분분석을 이용할 경우 각 방법의 복원신호의 차이를 보여 준다.

독립요소분석은 아래 식과 같이 다변량 데이터 \mathbf{X} (변수 \times 샘플) 만을 가지고 분해 행렬 \mathbf{W} 를 예측하여 독립요소 \mathbf{S} 를 구하게 된다.

$$\mathbf{S} = \mathbf{W}\mathbf{X} \quad (1)$$

물론 독립요소분석에서도 데이터 차원을 줄일 수 있는데 주성분분석처럼 각 스코어 벡터의 분산의 크기로 축의 중요성을 따지기가 힘들다. 독립요소분석에서 나오는 \mathbf{S} 의 행들은 모두 분산의 크기가 1이 되며 중요한 요소를 찾아내기가 힘들다. 본 연구에선 \mathbf{X} 의 변화가 \mathbf{W} 를 통해 \mathbf{S} 에 전파됨을 이용하여 \mathbf{W} 의 L2 norm이 큰 행에 해당하는 \mathbf{S} 의 행을 구하여 데이터를 축약하였다. 독립요소분석을 이용한 온라인 모니터링 방법은 다음과 같다. 우선 정상 상태의 데이터 \mathbf{X}_{normal} 에서 독립요소분석 알고리즘을 이용하여 $\mathbf{S}_{normal} = \mathbf{W}\mathbf{X}_{normal}$ 의 식으로부터 \mathbf{S}_{normal} 과 \mathbf{W} 를 찾아낸다. 그런 다음 테스트 데이터를 정상 상태일 때 구한 \mathbf{W} 에 투영함으로서 새로운 \mathbf{S}_{new} 를 구하게 된다. 주성분분석을 이용한 모니터링에서 사용하는 T^2 와 SPE 차트처럼 독립요소분석을 이용한 모니터링에서는 다음과 같이 I^2 와 SPE를 구할 수 있다.

$$I^2(k) = \mathbf{s}_{new}(k)^T \mathbf{s}_{new}(k) \quad (2)$$

$$SPE(k) = \sum_{j=1}^d (x_j(k) - \hat{x}_j(k))^2 \quad (3)$$

식(2)와 (3)처럼 두가지 statistics를 구할 수 있다.

본 아이디어에서 제안한 회분 공정 모니터링 방법은 그림 3와 같다. 우선 3차원 배열의 회분 데이터 $\underline{X}(I \times J \times K)$ 를 Approach A로 재배열한 다음 열의 평균을 빼서 회분 데이터의 평균 추이를 빼어 비선형성을 어느 정도 제거한다. 그런 다음 이 데이터($I \times KJ$)에서 Approach B로 다시 재배열하여 독립요소분석을 적용한다. 이럴 경우 Approach A 사용 시 평균추이를 제거하는 장점을 살린 채 Approach B의 장점을 적용할 수 있다. 물론 이 아이디어는 MPCA에도 적용할 수 있다. 따라서 온라인 모니터링 적용 시 각 회분의 미래 값들을 예측할 필요가 없고 회분 시간의 길이가 달라도 된다. 더욱이 주성분분석보다 독립요소 분석을 사용하기 때문에 다변량 데이터에 내재되어 있는 요소를 보다 효율적으로 찾아냄으로써 모니터링에 적용 시

효과적으로 이상현상이 일어나는 경우들을 쉽게 감지할 수 있다. 또한 MICA모델 구축을 위해서 사용된 독립성분들의 f^2 메트릭 값뿐만 아니라 제외된 독립성분들에 대해 f^2 를 적용하면 f_e^2 메트릭 값을 계산할 수 있다. SPE 값이 에러의 평균값만의 변화를 탐지할 수 있는데 비해 f_e^2 메트릭 값은 평균화 분산의 변화를 탐지할 수 있는 능력이 있다. 한편 모니터링 차트의 제한선들은 kernel density estimation을 이용하여 보다 정확하게 찾아내어 잘못된 alarm을 줄일 수 있다. 표 1은 다차원 독립성분 분석을 이용한 실시간 공정모니터링 방법의 순서도가 정리되어있다.

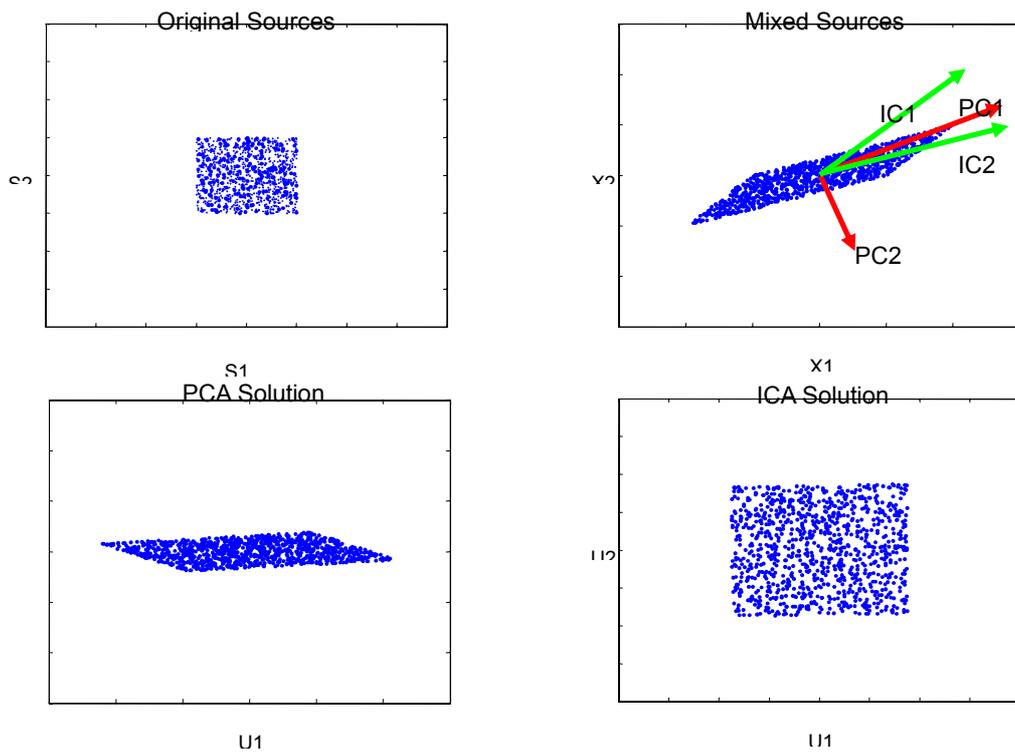


그림 2. 주성분 분석과 독립성분분석의 복원 신호의 차이점

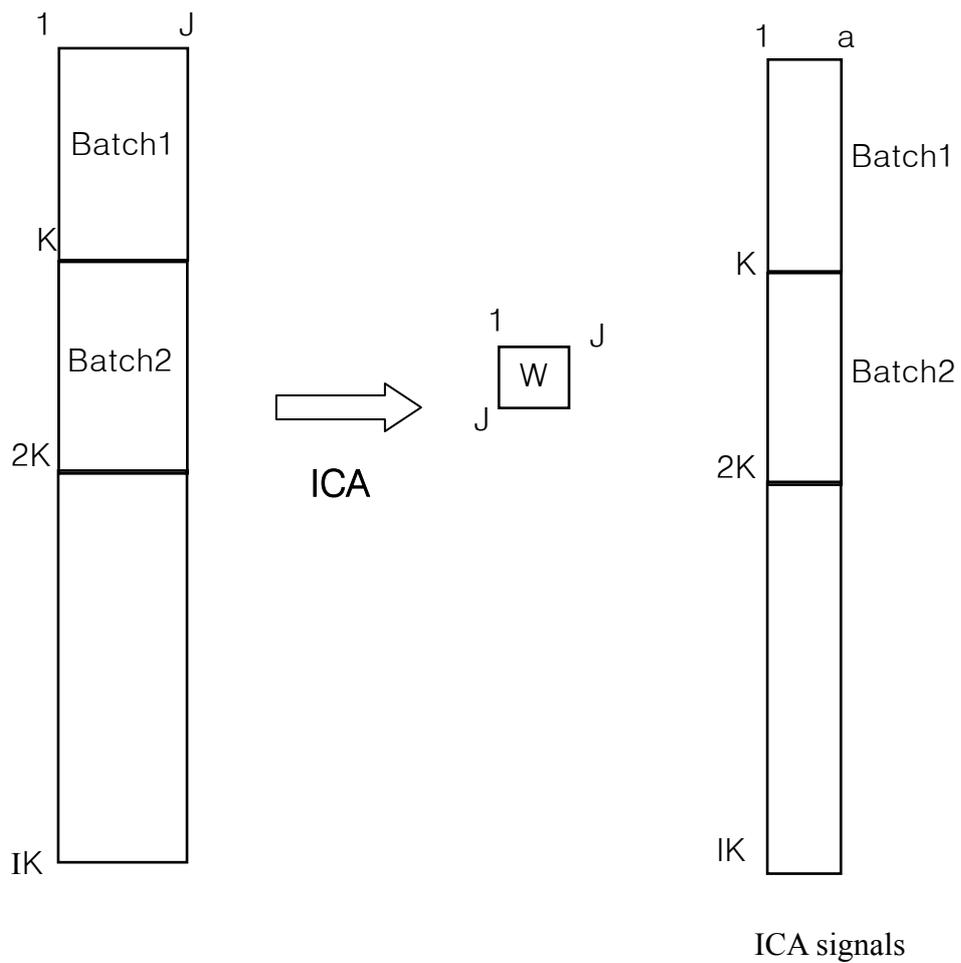
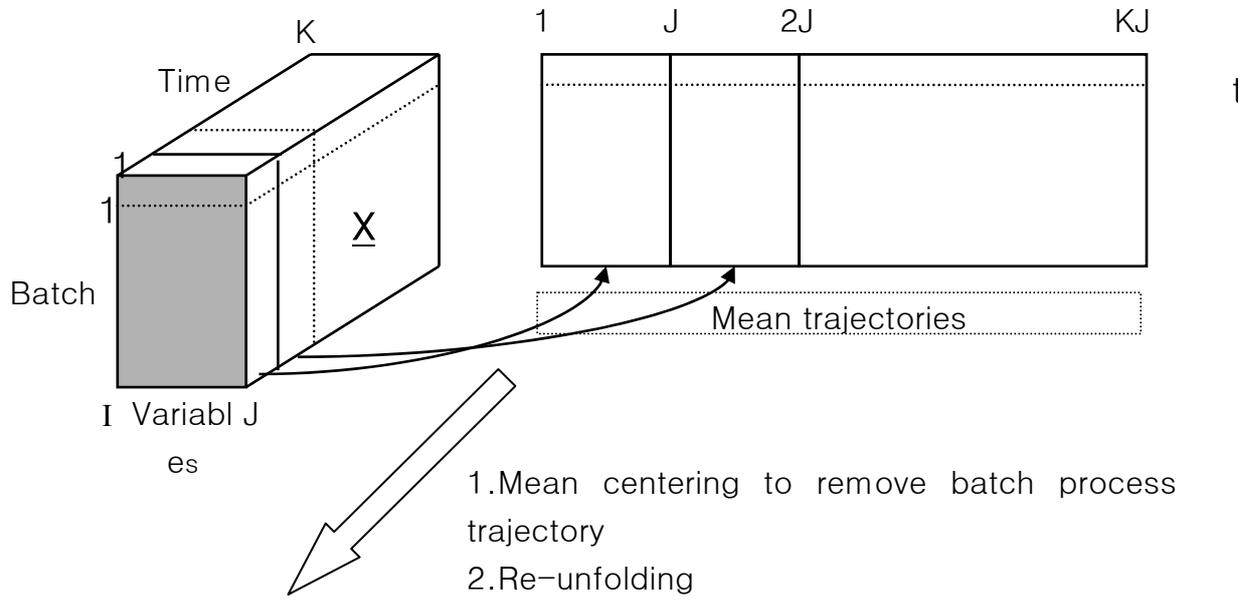


그림 3. 독립요소분석을 이용하는 새로운 회분식 모니터링 과정

표 1. MICA on-line monitoring procedures

A. 정상상태 공정 모델링

1. 배치 공정데이터 $\mathbf{X}(I \times J \times K)$ 를 2차원 배열로($\mathbf{X}(I \times JK)$) 펼침
2. 공정 데이터 $\mathbf{X}(I \times JK)$ 는 모든 배치에서 각시간에서 각 변수의 평균과 표준편차를 이용하여 표준화

3. 표준화된 2차원 배열을 $\mathbf{X}(I \times JK)$ 제안된 각 변수에 대한 배열로 변환
 $\mathbf{X}_{normal}(J \times IK)$

4. Whitening 공정

$$\mathbf{Z}_{normal} = \mathbf{Q}\mathbf{X}_{normal}$$

5. 독립성분분석(ICA) 실행

$$\text{Obtain } \mathbf{W}, \mathbf{B}, \text{ and } \mathbf{S}_{normal} \text{ from } \mathbf{S}_{normal} = \mathbf{W}\mathbf{X}_{normal} = \mathbf{B}^T \mathbf{Z}_{normal}.$$

6. \mathbf{W} 각 행의 크기를 계산하고 크기에 따라 구해진 독립성분의 순서를 정한 후 \mathbf{W} 를 중요부분과 노이즈 부분으로 분리, \mathbf{B} and \mathbf{S}_{normal} 도 같은 방법으로 분리 가능

$$\mathbf{W} \rightarrow \mathbf{W}_d, \mathbf{W}_e$$

$$\mathbf{B} \rightarrow \mathbf{B}_d, \mathbf{B}_e$$

$$\mathbf{S}_{normal} \rightarrow \mathbf{S}_d, \mathbf{S}_e$$

7. 3가지 I^2 , I_e^2 , and SPE metrics을 계산

$$I^2(n) = \mathbf{s}_d(n)^T \mathbf{s}_d(n)$$

$$I_e^2(n) = \mathbf{s}_e(n)^T \mathbf{s}_e(n)$$

$$SPE(n) = \sum_{j=1}^d (x_j(n) - \hat{x}_j(n))^2$$

여기서 n 은 1 부터 IK 값을 가지고 $\hat{\mathbf{X}} = \mathbf{Q}^{-1} \mathbf{B}_d \mathbf{S}_d = \mathbf{Q}^{-1} \mathbf{B}_d \mathbf{W}_d \mathbf{X}_{normal}$ 로 복원됨

8. $I^2(1 \times IK)$, $I_e^2(1 \times IK)$ and $SPE(1 \times IK)$ 벡터를 각각 $I^2(I \times K)$, $I_e^2(I \times K)$, $SPE(I \times K)$ 로 재배열
9. 커널밀도추정을 통한 I^2 , I_e^2 and SPE metrics 의 제어 한계치를 계산

B. 실시간 공정 모니터링

1. 새로운 배치 데이터의 시간 k 까지의 데이터 $\mathbf{X}_{test}(k \times J)$ 를 $\mathbf{x}_{test}^T(1 \times Jk)$ 로 펼친 후 모델링에 사용된 같은 스케일로 표준화
2. 표준화된 $\mathbf{x}_{new}^T(1 \times Jk)$ 을 $\mathbf{X}_{new}(J \times k)$ 로 재배열..
3. $\mathbf{S}_{newd} = \mathbf{W}_d \mathbf{X}_{new}$, $\mathbf{S}_{newe} = \mathbf{W}_e \mathbf{X}_{new}$ 로부터 \mathbf{S}_{newd} 와 \mathbf{S}_{newe} 계산
5. 계산된 \mathbf{S}_{newd} 와 \mathbf{S}_{newe} 값으로부터 각 시간의 $I_{newd}^2(k)$, $I_{newe}^2(k)$, $SPE(k)$ 계산

$$I_{newd}^2(k) = \mathbf{s}_{newd}(k)^T \mathbf{s}_{newd}(k), \quad I_{newe}^2(k) = \mathbf{s}_{newe}(k)^T \mathbf{s}_{newe}(k)$$

$$SPE(k) = \sum_{j=1}^d (x_{newj}(k) - \hat{x}_{newj}(k))^2$$

여기서 $\hat{\mathbf{X}} = \mathbf{Q}^{-1} \mathbf{B}_d \mathbf{S}_{newd} = \mathbf{Q}^{-1} \mathbf{B}_d \mathbf{W}_d \mathbf{X}_{new}$

6. 계산된 $I_{newd}^2(k)$, $I_{newe}^2(k)$, $SPE(k)$ 과 정상상태 모델과정에서 계산된 제어 한계치와 비교하여 공정의 이상여부 판단

C. 기여도 분석 (Contribution Plot)

1. Variable contribution for $I_{newd}^2(k)$

$$\mathbf{x}_{cd}(k) = \frac{\mathbf{Q}^{-1} \mathbf{B}_d \mathbf{s}_{newd}(k)}{\|\mathbf{Q}^{-1} \mathbf{B}_d \mathbf{s}_{newd}(k)\|} \|\mathbf{s}_{newd}(k)\|$$

2. Variable contribution for $I_{newe}^2(k)$

$$\mathbf{x}_{ce}(k) = \frac{\mathbf{Q}^{-1} \mathbf{B}_e \mathbf{s}_{newe}(k)}{\|\mathbf{Q}^{-1} \mathbf{B}_e \mathbf{s}_{newe}(k)\|} \|\mathbf{s}_{newe}(k)\|$$

3. Variable contribution for $SPE(k)$

$$\mathbf{x}_{cspe}(k) = \mathbf{x}(k) - \hat{\mathbf{x}}(k)$$