

다차원 독립성분분석을 이용한 회분 공정의 온라인 모니터링 기술
(On-Line Monitoring of Batch Processes Using Multiway Independent
Component Analysis)

유창규

이번 연구 보고서는 Nomikos and MacGregor의 연구에서 제안된 unfolding method를 ICA에 적용하여 두 가지 MPCA/MICA 방법의 비교와 공정 모니터링기술에 대한 정보를 제공한다. 앞 연구 보고서에서 제시된 배치 공정에서 독립 성분 분석(independent component analysis, ICA)을 수정하여 다차원 독립성분분석(multiway independent component analysis, MICA)을 이용한 새로운 배치공정모니터링 방법을 제시한다 (ICA의 기본개념은 앞 보고서를 참조).

본 연구의 기본 아이디어는 MPCA에서 사용된 방법과 유사하다. MICA models은 normal batch operation하에서 수행된 historical database를 기초로 만들어진다. 먼저 MICA 방법은 그림 1의 3차원 \underline{X} 데이터에서 unfolding 방법을 사용하여 2차원 데이터에 대해 독립 성분 분석(independent component analysis, ICA)을 수행한다. MICA방법은 three-way array \underline{X} 를 구해진 독립성분들(ICs, \mathbf{s})이 가능한한 서로 통계적으로 독립이 되도록 하기 위해 아래식과 같이 the product of independent vectors \mathbf{s}_r 와 loading matrices \mathbf{W}_r 과 a residual array \mathbf{E} 의 합으로 분해한다.

$$\underline{\mathbf{X}} = \sum_{r=1}^R \mathbf{s}_r \otimes \mathbf{A}_r + \mathbf{E} = \sum_{r=1}^R s_r \mathbf{a}_r^T + \mathbf{E} = \hat{\underline{\mathbf{X}}} + \mathbf{E}$$

여기서 \otimes the Kronecker product ($\underline{\mathbf{X}} = \mathbf{s} \otimes \mathbf{A}$ is $\underline{X}(i, j, k) = s(i)A(j, k)$)를 나타내고 R 은 선택된 the number of IC를 나타낸다. \mathbf{S} 와 \mathbf{A} matrices는 MPCA의 the loading matrix와 score matrices와 같은 기능을 한다 (*i.e.* \mathbf{S}

can be regarded as the score matrix T , and A can be treated as the loading matrix P .) independent vector s 의 i 번째 요소는 i 번째 배치에 대응되고 다른 배치에 비교하여 i 번째 배치의 overall variations을 나타낸다. The mixing matrix, A 는 각각의 평균 trajectories에 대한 측정변수들의 time variations을 나타내면 A 의 각 요소는 그 배치안에서 각각의 시간에서 각 변수에 적용되었을 때 가중치(weights)를 나타낸다.

MICA 모니터링의 performance는 ICA model의 정확한 choice of the ordering과 dimension에 많은 영향을 받는다. 몇 개의 중요한 IC의 선택은 robust performance와 reduction of analysis complexity라는 측면에서 고려된다. 주요한 성분의 선택은 다음과 같은 두가지 스텝으로 이루어 진다.

Step 1. List and order all the ICs in the appropriate order.

Step 2. Select the first few ICs in the list as the dominant ones.

PCA와 달리 ICA ordering에 대한 일반적인 방법이 제시되어 있지 않다. 현재 연구에서는 demixing matrix, W 의 행을 각각의 Euclidean norms (L_2)에 따라 ordering을 정하는 방법을 사용하였다.

$$\arg \text{Max}_i \|w_i\|_2$$

즉, ICs는 공정에서 커다란 변화를 일으키는 것으로 알려진 몇 개의 ICs를 L_2 norm을 사용하여 ordering을 한다. 만일 ICA ordering이 이루어 졌으면 모니터링에 사용될 최적의 IC개수를 선택하는 것이 두번째 중요한 과정이다. (The data dimension can be reduced by selecting the first few rows of the ordered W based upon the assumption that the rows with the largest Euclidean norm have the greatest effect on the variation of S . This approach is based on the idea that the dominant variation in a process can be monitored by considering the cumulative sums of only the first few

dominant ICs.) 본 연구에서는 최적 ICA 갯수를 구하기 위해 PCA의 SCREE test와 같은 graphical technique을 이용했다.

본 연구에서 제안된 회분 공정 모니터링 방법은 다음과 같다. MPCA와 마찬가지로, the MICA model은 preprocessed된 normal data로부터 구축된다. 첫째, 3차원 배열의 회분 데이터 $\underline{X}(I \times J \times K)$ 를 two-dimensional matrix ($\mathbf{X}(I \times JK)$)로 재배열한다. 둘째, 열의 평균을 빼고 분산으로 나누어 standardization을 시킨다 (회분 데이터의 평균 추이를 빼어 비선형성을 어느 정도 제거한다). 셋째, normalized $\mathbf{X}(I \times JK)$ 는 transpose되어 transposed matrix $\mathbf{X}_{normal}(JK \times I)$ 를 구한다. 넷째, uncorrelated whitened matrix을 이루기 위해 whitening과정이 $\mathbf{X}_{normal}(JK \times I)$ 에 수행된다 ($\mathbf{Z}_{normal} = \mathbf{Q}\mathbf{X}_{normal}$). 여기서 $\mathbf{X}_{normal}(JK \times I)$ 의 covariance matrix 랭크만큼의 칼럼이 구해진다. 따라서 whitening matrix \mathbf{Q} 는 $(r \times I)$ 차원을 가진다. 다섯째, $\mathbf{s}(n) = \mathbf{B}^T \mathbf{z}(n)$ 변환이 최대한 non-Gaussianity를 지니도록 \mathbf{B} 를 구한다 (FastICA method). 이로부터 \mathbf{A} 와 \mathbf{W} 가 다음 식으로부터 $\mathbf{A} = (\mathbf{Q}^T \mathbf{Q})^{-1} \mathbf{Q}^T \mathbf{B}$ and $\mathbf{W} = \mathbf{B}^T \mathbf{Q}$ 계산된다. 여섯째, ICA ordering과 dimension reduction과정이 수행된다. \mathbf{W} 의 행중 중요한 몇 개의 m 가 선택되고(\mathbf{W}_d , deterministic part of \mathbf{W}), \mathbf{W} 의 나머지 행은 \mathbf{W}_e (excluded part of \mathbf{W})로 나누어 진다. 마지막으로 \mathbf{W} 의 선택된 행에 대응하는 \mathbf{A} 의 열이 선택되고 (\mathbf{A}_d) 나머지는 \mathbf{A}_e 로 분리된다. 또 \mathbf{S}_d 와 \mathbf{S}_e 는 각각 \mathbf{W}_d and \mathbf{W}_e 을 사용하여 계산된다.

제시된 방법은 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 ℓ^2 메트릭 값뿐만 아니라 제외된 독립성분들에 대해 ℓ^2 를 적용하면 ℓ_e^2 메트릭 값을 계산할 수 있다. SPE 값이 에러의 평균값만의 변화를 탐지할 수 있는데 비해 ℓ_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. 밑의 프로시저는 다차원 독립성분분석을 이용한 실시간 공정모니터링 방법의 순서가 정리되어있다.

Procedure for On-line Batch Monitoring using MICA

A. 정상상태 공정 모델링

1. 정상상태의 배치공정 데이터를 수집
2. 배치 공정데이터 $\mathbf{X}(I \times J \times K)$ 를 2차원 배열로($\mathbf{X}(I \times JK)$) 펼침
3. 공정 데이터 $\mathbf{X}(I \times JK)$ 는 모든 배치에서 각시간에서 각 변수의 평균과 표준편차를 이용하여 표준화
4. 표준화된 2차원 배열을 $\mathbf{X}(I \times JK)$ 을 $\mathbf{X}_{normal}(JK \times I)$ 로 transpose
5. Whitening 공정 $\mathbf{Z}_{normal} = \mathbf{Q}\mathbf{X}_{normal}$

여기서 $\mathbf{Z}_{normal} = [\mathbf{z}(1), \mathbf{z}(2), \dots, \mathbf{z}(I)] \in R^{r \times I}$ 이고 r 은 the covariance matrix of \mathbf{X}_{normal} 의 rank

6. 독립성분분석(ICA) 실행

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

$\mathbf{s}(n) = \mathbf{B}^T \mathbf{z}(n)$ 이 최대의 non-Gaussianity를 갖도록하는 \mathbf{W} , \mathbf{B} , and \mathbf{S}_{normal}

계산

$$\mathbf{S}_{normal} = \mathbf{W}\mathbf{X}_{normal} = \mathbf{B}^T\mathbf{Z}_{normal}$$

여기서 $\mathbf{A} = (\mathbf{Q}^T\mathbf{Q})^{-1}\mathbf{Q}^T\mathbf{B}$ and $\mathbf{W} = \mathbf{B}^T\mathbf{Q}$.

7. \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

$$\mathbf{S}_{normal} = \begin{bmatrix} \mathbf{S}_d \\ \mathbf{S}_e \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} \mathbf{A}_d \\ \mathbf{A}_e \end{bmatrix}, \quad \mathbf{W} = \begin{bmatrix} \mathbf{W}_d \\ \mathbf{W}_e \end{bmatrix}$$

8. $\mathbf{X}(I \times JK)$ 의 각 배치 데이터 ($\mathbf{x}'(1 \times JK)$)에 대해 MICA모델의 reduced space로 투영. 모든 / batches에 대해 ICs와 residuals은 \mathbf{S}_d 와 \mathbf{S}_e 로부터 계산.

9. I^2 와 I_e^2 statistic을 계산

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

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

여기서 $1 \leq i \leq I$.

10. I^2 와 I_e^2 statistics의 control limits을 kernel density estimation을 이용하여 계산

11. 각 배치의 시간에 대해 SPE를 $SPE(i, k) = \sum_{j=(k-1)J+1}^{kJ} (\mathbf{x}_{predj} - \hat{\mathbf{x}}_{predj})^2$ 을 이용하여

계산. 여기서 $\hat{\mathbf{x}}_{pred} = \mathbf{A}_d \mathbf{W}_d \mathbf{x}_{pred}$, \mathbf{x}_{predj} 는 j번째 샘플. Reference database의

전체 배치 공정에서 계산된 MICA모델의 \mathbf{A}_d 와 \mathbf{W}_d 을 사용. 배치 i 와

시간 k 에서 현재 $\mathbf{x}_{normal}(kJ \times 1)$ 까지의 데이터만 존재하므로 $\mathbf{x}_{normal}(kJ \times 1)$ 의

미래값을 예측

12. kernel density estimation을 이용하여 각 시간에서 SPE값의 control limit를 계산

B. 실시간 공정 모니터링

1. 새로운 배치 데이터의 시간 k 까지의 데이터 $\mathbf{X}_{test} (k \times J)$ 를 $\mathbf{x}_{test} (kJ \times 1)$ 로 펼친 후 모델링에 사용된 같은 스케일로 표준화하고 나머지 the missing values를 예측하기 위해서 Nomikos and MacGregor (1995a)의 세가지 방법중 하나를 선택하여 $\mathbf{x}_{pred_test} (JK \times 1)$ 를 예측

2. 표준화된 $\mathbf{x}_{pred_test} (JK \times 1)$ 에 대해 $\mathbf{S}_{newd} = \mathbf{W}_d \mathbf{X}_{new}$, $\mathbf{S}_{newe} = \mathbf{W}_e \mathbf{X}_{new}$ 로부터 \mathbf{s}_{testd} and \mathbf{s}_{teste} 의 ICs을 계산

$$\mathbf{s}_{testd}(k) = \mathbf{W}_d \mathbf{x}_{pred_test}, \quad \mathbf{s}_{teste}(k) = \mathbf{W}_e \mathbf{x}_{pred_test}$$

5. 계산된 \mathbf{S}_{newd} 와 \mathbf{S}_{newe} 값으로부터 현재의 $I_{newd}^2(k)$, $I_{newe}^2(k)$, $SPE(k)$ 계산

$$I_{testd}^2(k) = \mathbf{s}_{testd}(k)^T \mathbf{s}_{testd}(k)$$

$$I_{teste}^2(k) = \mathbf{s}_{teste}(k)^T \mathbf{s}_{teste}(k)$$

$$SPE(k) = \sum_{j=(k-1)J+1}^{kJ} (\mathbf{x}_{predj} - \hat{\mathbf{x}}_{predj})^2$$

여기서 $\hat{\mathbf{x}}_{pred} = \mathbf{A}_d \mathbf{W}_d \mathbf{x}_{pred}$ 이고 \mathbf{x}_{predj} 는 \mathbf{x}_{predj} 의 j 번째 요소

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

C. 기여도 분석 (Contribution Plot)

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

$$\mathbf{x}_{cd}(k) = \frac{\mathbf{A}_d \mathbf{s}_{newd}(k)}{\|\mathbf{A}_d \mathbf{s}_{newd}(k)\|} \|\mathbf{s}_{newd}(k)\|$$

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

$$\mathbf{x}_{ce}(k) = \frac{\mathbf{A}_e \mathbf{s}_{newe}(k)}{\|\mathbf{A}_e \mathbf{s}_{newe}(k)\|} \|\mathbf{s}_{newe}(k)\|$$

3. Variable contribution for $SPE(k)$

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

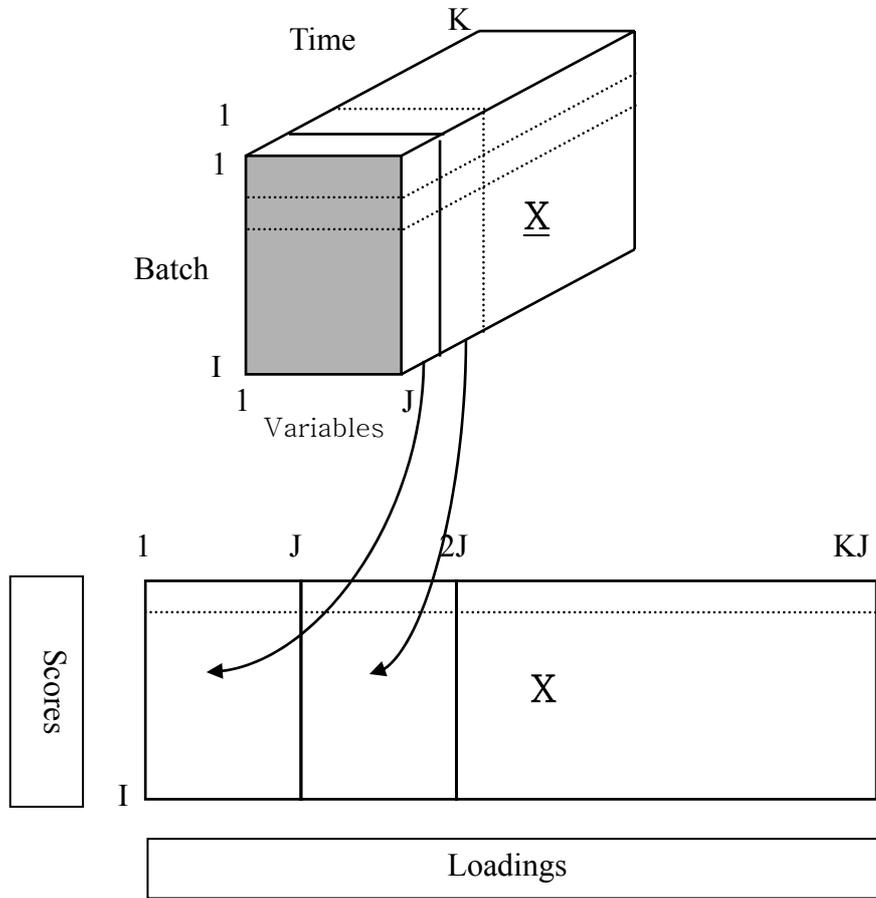


그림 1. 다차원 독립요소분석을 이용하는 새로운 회분식 모니터링 과정.