Distributed Gaussian Mixture Model for Monitoring Multimode Plant-wide Process

Jinlin Zhu[1], Zhiqiang Ge[1], Zhihuan Song[1]

1. State Key Laboratory of Industrial Control Technology, Institute of Industrial Process Control, College of Control Science and Engineering, Zhejiang University, Hangzhou, 310027, P. R. China
E-mail: gezhiqiang@zju.edu.cn

Abstract: For large-scale plant-wide processes with multiple operating conditions, a distributed Gaussian mixture modeling and monitoring mechanism is proposed. To overcome the deficient prior modeling knowledge for complex process, a two-dimensional probabilistic topic model based technique named Bayesian co-clustering method is developed to simultaneously conduct the sub-block division and operating mode recognition. With the obtained sub-blocks and operating modes, a global Gaussian mixture model is first built and then several local Gaussian mixture models are extracted and applied for distributed monitoring of plant-wide processes. By conducting distributed modeling and monitoring, both global and local process changes can be reflected and the fault region can be localized more easily for further analyses. The feasibility and effectiveness of the proposed method is confirmed through the Tennessee Eastman benchmark process.

Key Words: Plant-wide process monitoring; Distributed data modeling; Gaussian mixture model; Bayesian co-clustering; Variational Bayesian inference.

1 Introductions

Due to the ever increasing development of modern industry, the plant-wide process monitoring has become more and more popular in the academic research field [1]. Generally, the plant-wide process is usually characterized with multiple operating units, workshops and even plants; moreover, there may also include multiple operating conditions in order to produce various products with different values. To ensure the safe operation and stable production quality, automatic modeling and monitoring of plant-wide process are especially necessary [2].

Traditional process modeling and monitoring methods have been designed based on mechanistic model such as state-space method. However, the mechanistic analysis and kinematic equation inductions can be sophisticated, time-consuming and sometimes even unavailable. As a class of alternatives, the data-based process monitoring methods have been widely focused and applied [3]. For example, MacGregor proposed a multiblock partial least square (PLS) method and monitoring charts are defined for both sub-blocks and the entire process so as to enhance the monitoring performance [4]. Qin analyzed several multi-block methods and the decentralized monitoring and diagnosis framework was proposed[5]. However, a critical issue for these methods is that the decomposition for local blocks should largely rely on a priori process knowledge. In practice, however, a priori process knowledge is not always available for plant-wide process.

There have been attempts to block plant-wide data with different automatic mechanisms. For example, Ge proposed a distributed principal component analysis method for plant-wide process monitoring [2]. PCA sub-blocks are constructed through different directions of principal components. Recently, the Generalized Dice's coefficient is used for measuring similarity among vectors in latent space, and then PCA sub-blocks are constructed according to latent vectors with high similarity [6]. However, both methods are constructed based on PCA and seldom consider the multiple operating characteristic of plant-wide process

In this paper, a probabilistic method called Bayesian co-clustering is developed for analyzing the plant-wide conditions of multiple modes and sub-blocks. Technically, the Bayesian co-clustering is based on the latent Dirichlet allocation (LDA) and can simultaneously conduct the sub-block division and operating mode recognition from the two-dimension aspects through topics [7]. The merits of Bayesian co-clustering for plant-wide monitoring can be found from the following three aspects [8]. First, topic models can discover the potential topics and then organize data according to the discovered theme. Notice that in industrial process, a topic on variable-wise can be viewed as a unit like a reactor, a condenser or even a workshop with a combination of coupled units, etc. While for sample-wise, a topic can be presented with a certain operating mode. Therefore, this merit is very useful for plant-wide data blocking since the blocking progress can be regarded as a way to discover the topic of each block and then reorganize data with various blocks from different topics. Second, recent advance show that topic model can be applied to massive collection of industrial data. With the development of digitized store technique, the volume of data is growing, especially for plant-wide situation. Therefore, probabilistic topic model provides a desirable statistic solution for data mining and information discovery for plant-wide process issue. Third, topic model can handle different kinds of data. For example, in applications out of process control, the topic model has been adapted to other process system engineering fields like analysis of genetics data, images and social networks. Actually, with the coming of big data era, the advancement of process monitoring methods show

This work was supported in part by the National Natural Science Foundation of China (NSFC) (61273167), National 973 Project (2012CB720500), and the Alexander von Humboldt Foundation.

similar trends that should also combine information from different kinds of data types, for example, data from spectra measurements, images and sound, etc [9]. In this sense, the topic model provides a good alternative. Once the blocking information and mode allocation have been obtained, distributed Gaussian mixture models can be constructed for distributed monitoring. Specifically, a global GMM is first built for capturing the global operating mode of the current process. Based on that, sub-block GMMs are extracted from the global model so as to monitor the local behavior of each process theme. In contrast to traditional GMM, the distributed GMM (DGMM) can be more feasible for reflecting local process changes. Meanwhile, the fully probabilistic treatment of blocking, mode allocation, and distributed modeling provides a feasible and effective solution for the general plant-wide process monitoring.

The rest of this article is organized as follows. Conventional GMM based monitoring method is first revisited. Then Distributed modeling and monitoring with Bayesian co-clustering approach are introduced and developed. In the simulation part, the proposed method is validated on the challenging Tennessee Eastman process. Finally, conclusions and future outlooks are made.

2 Process monitoring with GMM

In this part, we give a brief revisit on GMM based process monitoring. Assume data set $\mathbf{X}=\left\{\mathbf{x}_n \in R^D\right\}_{n=1}^N$ is generated from a multimode process with K local operating conditions, D denotes the dimensionality of data and N is the total number of sampling data. The probability density function of GMM can be defined as follows

$$p\left(\mathbf{X}\mid\boldsymbol{\pi},\Theta\right)=\prod_{n=1}^{N}\sum_{c=1}^{K}p\left(z_n=c\mid\boldsymbol{\pi}\right)p\left(\mathbf{x}_n\mid\theta_c\right) \tag{1}$$

where z_n is the latent indicator, $\Theta=\{\theta_c\}_{c=1}^K$ is the parameter set, and $\theta_c=\{\mu_c,\Sigma_c\}$, $\boldsymbol{\pi}=\{\pi_c\}_{c=1}^K$ is the weight set for all local Gaussian components and $\sum_{c=1}^{K}\pi_c=1$. Due to the fixed weights, $p\left(z_n=c\mid\pi\right)$ is written with π_c in some cases for simplicity. For parameter estimation, expectation maximization (EM) algorithm is commonly used based on the derivative of log-likelihood function, in the E-step, expectations for latent variables are computed and in the M-step, parameters are updated with the expectations derived in the former step. The explicit steps can be found in [10]. Yu has also proposed the Bayesian inference-based probability (BIP) index [10]. The BIP index is a desirable global probabilistic index designed for fault detection with GMM. Specifically, for a new coming data $\mathbf{x}_{new}$, the BIP index can be calculated as:

$$\mathrm{BIP}=\sum_{c=1}^{K}p\left(c\mid\mathbf{x}_{new},\Theta\right)P_L^c\left(\mathbf{x}_{new}\right) \tag{2}$$

where $P_L^c\left(\mathbf{x}_{new}\right)$ is defined as the Mahalanobis distance based probability of $\mathbf{x}_{new}$ with respect to component c, and can be computed by

$$P_L^c\left(\mathbf{x}_{new}\right)=\Pr\left\{\mathrm{MD}_c\left(\mathbf{x}\right)\le\mathrm{MD}_c\left(\mathbf{x}_{new}\right)\right\} \tag{3}$$

where $\mathrm{MD}_c\left(\mathbf{x}_{new}\right)=\left(\mathbf{x}_{new}-\mu_c\right)^T\left(\Sigma_k+\varepsilon I\right)^{-1}\left(\mathbf{x}_{new}-\mu_c\right)$ is the squared Mahalanobis distance. The Mahalanobis distance from the center of cluster c follows the χ_D^2 distribution with degree of freedom D. Hence the above equation can be easily solved by a simple integration. Besides, one has $0\le\mathrm{BIP}_{new}\le1$. The BIP upper control limit can be given a priori, such as 0.95 or 0.99.

3 Distributed GMM method

3.1 Bayesian co-clustering structure

Latent Dirichlet allocation or LDA is originally used for discrete text modeling, and the goal is to extract topics from a series of documents so as to provide statistic evidences for further analysis such as classification, summarization, novelty detection, etc. It is worth to emphasize that although the samples in the context are assumed to be Gaussian distributed which are not discrete data with tokens, we can still regard that the model as a general form of LDA. One can easily see from Eq.(1) that GMM takes the assumption that mixing weights are fixed for all data items. This is actually a kind of 'reduced description' on a fixed set of components. In other words, all data are constrained to be generated from component set with fixed priors. As a Bayesian method, latent Dirichlet allocation relaxes this limitation by assuming that a separate weight vector $\boldsymbol{\pi}$ should be employed for each sampling data. In addition, the weights are further assumed to be sampled from a Dirichlet distribution with prior $\boldsymbol{\alpha}$, denoted as $\mathrm{Dir}\left(\boldsymbol{\pi}\mid\boldsymbol{\alpha}\right)$. Therefore, the probability density function of LDA can be defined as follows[11]:

$$p\left(\mathbf{X}\mid\boldsymbol{\alpha},\Theta\right)=\int_{\pi}\mathrm{Dir}\left(\boldsymbol{\pi}\mid\boldsymbol{\alpha}\right)\prod_{n=1}^{N}\sum_{c=1}^{K}p\left(z_n=c\mid\boldsymbol{\pi}\right)p\left(\mathbf{x}_n\mid\theta_c\right)d\boldsymbol{\pi} \tag{4}$$

The basic idea here is that process data are represented as random mixtures over latent components, where each component is characterized by a distribution over data samples. By doing so, LDA assumes that each data sample should be sampled with the following steps: First, component weights are sampled from Dirichlet distribution; second, a specific component is chosen from the multinomial distribution of $p\left(z\mid\boldsymbol{\pi}\right)$; in the last step, process data is generated from the empirical Gaussian distribution. Compared with GMM, such latent allocation puts more flexibility for modeling. In addition, due to the similar allocation mechanism, data sampled from the same operating mode could be gathered again after statistical modeling has been completed.

The above allocation is based on the topic of mode similarity which is performed on the dimension of sample wise. From the variable wise, the same mechanism can also be introduced. Similar to the sample wise data clustering with mode topic, the variable wise clustering can be viewed as following the topic of block division. In other words, variables with same implicated topic are gathered in the same cluster or block. As aforementioned, the topic here can

refer to a unit like a reactor, a condenser or even a workshop with a combination of coupled units. Therefore, the entire clustering progress is named as Bayesian co-clustering. For modeling, in order to distinguish between both dimensionalities, we assume two Dirichlet distributions $\mathrm{Dir}(\boldsymbol{\alpha}_1)$ and $\mathrm{Dir}(\boldsymbol{\alpha}_2)$ for rows and columns. Besides, mixing weights are defined for each row u and column v and denoted as $\boldsymbol{\pi}_{1u}$ and $\boldsymbol{\pi}_{2v}$, $u \in \{1,2,...,D\}$ and $v \in \{1,2,...,N\}$. Latent indicators are also separately annotated as z_{1uv} and z_{2uv} representing latent row and column cluster respectively. Notice that $z_{1uv} \in \{1,2,...,K_1\}$, $z_{2uv} \in \{1,2,...,K_2\}$ where K_1 and K_2 are total number of row clusters and column clusters. The probability density function of Bayesian co-clustering model can be defined as follows [7]

$$\begin{aligned} & p(\mathbf{X} \mid \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \Theta) \\ & = \int_{\pi_1} \int_{\pi_2} \left(\prod_{u=1}^{D} \mathrm{Dir}(\boldsymbol{\pi}_{1u} \mid \boldsymbol{\alpha}_1) \right) \left(\prod_{v=1}^{N} \mathrm{Dir}(\boldsymbol{\pi}_{2v} \mid \boldsymbol{\alpha}_2) \right) \\ & \left(\prod_{u=1}^{D} \prod_{v=1}^{N} \sum_{z_{1uv}=1}^{K_1} \sum_{z_{2uv}=1}^{K_2} p(z_{1uv} \mid \boldsymbol{\pi}_{1u}) \, p(z_{2uv} \mid \boldsymbol{\pi}_{2v}) \, p\left(x_{uv} \mid \theta_{z_{1uv}, z_{2uv}}\right) \right) d\boldsymbol{\pi}_1 d\boldsymbol{\pi}_2 \end{aligned} \tag{5}$$

3.2 Bayesian inference

Now turn to the computational problem where the goal is to estimate the parameters of implicated distributions. However, learning such a hierarchical model with maximum likelihood is computationally intractable due to the integral of complex latent structure. In this work, the variational inference is used. The core idea of variational inference is to approximate the complex distribution over latent structure with a simpler distribution. Define $\mathbf{T} \triangleq \{\mathbf{Z}_1, \mathbf{Z}_2, \boldsymbol{\pi}_1, \boldsymbol{\pi}_2\}$ for latent variables; the marginal log-likelihood of the model can be given as

$$\begin{aligned} & \log p(\mathbf{X} \mid \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \Theta) \\ & = \log \int p(\mathbf{X}, \mathbf{T} \mid \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \Theta) d\mathbf{T} \\ & \geq \int q(\mathbf{T}) \log \frac{p(\mathbf{X}, \mathbf{T} \mid \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \Theta)}{q(\mathbf{T})} d\mathbf{T} \\ & = F(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \Theta, q(\mathbf{T})) \end{aligned} \tag{6}$$

Notice that $F(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \Theta, q(\mathbf{T}))$ is usually called lower bound while $q(\mathbf{T})$ is called variational posterior. Therefore, maximizing the likelihood turns to the problem of maximization of lower bound. In order to make the calculation more appropriate, variational posterior is further factorized by mean field approximation as

$$q(\mathbf{T}) = q(\mathbf{Z}_1 \mid \phi_1) q(\mathbf{Z}_2 \mid \phi_2) q(\boldsymbol{\pi}_1 \mid \boldsymbol{\gamma}_1) q(\boldsymbol{\pi}_2 \mid \boldsymbol{\gamma}_2) \tag{7}$$

where each variational posterior factor is defined by the same distribution assignation as the true corresponding posterior distribution. Therefore, the lower bound can be further written as:

$$\begin{aligned} & F(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \Theta, q(\mathbf{T})) \\ & = E_q(\log(\boldsymbol{\pi}_1 \mid \boldsymbol{\alpha}_1)) + E_q(\log(\boldsymbol{\pi}_2 \mid \boldsymbol{\alpha}_2)) \\ & + E_q(\log(\mathbf{Z}_1 \mid \boldsymbol{\pi}_1)) + E_q(\log(\mathbf{Z}_2 \mid \boldsymbol{\pi}_2)) \\ & + E_q(\log(\mathbf{X} \mid \mathbf{Z}_1, \mathbf{Z}_2, \Theta)) - E_q(\log q(\mathbf{Z}_1 \mid \phi_1)) - E_q(\log q(\mathbf{Z}_2 \mid \phi_2)) \\ & - E_q(\log q(\boldsymbol{\pi}_1 \mid \boldsymbol{\gamma}_1)) - E_q(\log q(\boldsymbol{\pi}_2 \mid \boldsymbol{\gamma}_2)) \end{aligned} \tag{8}$$

For maximization, take the derivative of $F(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \Theta, q(\mathbf{T}))$ with respect to variational parameters ϕ_1, ϕ_2, $\boldsymbol{\gamma}_1$, $\boldsymbol{\gamma}_2$ and parameters $\Theta = \{\theta_{ij} = \{\mu_{ij}, \sigma_{ij}^2\}, i = 1,...,K_1, j = 1,...,K_2\}$ and set to zeros. With some manipulations, one can derive the following updating rules

$$\phi_{1ui} \propto \exp\left(\Psi(\gamma_{1ui}) - \Psi\left(\sum_{l=1}^{K_1} \gamma_{1ul} \right) + \frac{\sum_{v=1}^{N} \sum_{j=1}^{K_2} \phi_{2vj} \log p\left(x_{uv} \mid \theta_{ij}\right)}{m_u} \right) \tag{9}$$

$$\phi_{2vj} \propto \exp\left(\Psi(\gamma_{2vj}) - \Psi\left(\sum_{l=1}^{K_2} \gamma_{2vl} \right) + \frac{\sum_{u=1}^{D} \sum_{i=1}^{K_1} \phi_{1ui} \log p\left(x_{uv} \mid \theta_{ij}\right)}{m_v} \right) \tag{10}$$

$$\gamma_{1ui} = \alpha_{1i} + m_u \phi_{1ui} \tag{11}$$

$$\gamma_{2vj} = \alpha_{1j} + m_v \phi_{2vj} \tag{12}$$

$$\mu_{ij} = \frac{\sum_{u=1}^{D} \sum_{v=1}^{N} \phi_{1ui} \phi_{2vj} x_{uv}}{\sum_{u=1}^{D} \sum_{v=1}^{N} \phi_{1ui} \phi_{2vj}} \tag{13}$$

$$\sigma_{ij}^2 = \frac{\sum_{u=1}^{D} \sum_{v=1}^{N} \left(x_{uv} - \mu_{ij}\right)^2 \phi_{1ui} \phi_{2vj}}{\sum_{u=1}^{D} \sum_{v=1}^{N} \phi_{1ui} \phi_{2vj}} \tag{14}$$

Notice that $\Psi(.)$ is the digamma function, ϕ_{1ui} denotes the i th component of ϕ_1 for row u and $\sum_{i=1}^{K_1} \phi_{1ui} = 1$, γ_{1ui} is the corresponding variational prior, m_u and m_v are the number of entries in row u and column v respectively. Finally, the Dirichlet parameters $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ should be updated. In order to estimate such parameters efficiently, the Newton-Rapson optimization method is utilized [11]. Basically, the formulation is

$$\boldsymbol{\alpha}' = \boldsymbol{\alpha} - H(\boldsymbol{\alpha})^{-1} g(\boldsymbol{\alpha}) \tag{15}$$

where $H(\boldsymbol{\alpha})$ and $g(\boldsymbol{\alpha})$ are Hessian matrix and gradient term respectively. Therefore, for each element α_{1i} in $\boldsymbol{\alpha}_1$, we have

$$\alpha_{1i}^{'} = \alpha_{1i} - \frac{g_{1i} - u}{h_{1i}} \tag{16}$$

$$g_{1i} = D\left(\Psi\left(\sum_{l=1}^{K_1} \alpha_{1l} \right) - \Psi(\alpha_{1i}) \right) + \sum_{u=1}^{D} \left(\Psi(\gamma_{1ui}) - \Psi\left(\sum_{l=1}^{K_1} \gamma_{1ul} \right) \right) \tag{17}$$

$$h_{1i} = -D\Psi'(\alpha_{1i}) \tag{18}$$

$$u = \frac{\sum_{i=1}^{K_1} g_{1i} / h_{1i}}{w^{-1} + \sum_{l=1}^{K_1} h_{1l}^{-1}} \tag{19}$$

$$w = D\Psi\left(\sum_{l=1}^{K_1} \alpha_{1i}\right) \tag{20}$$

For $\boldsymbol{\alpha}_2$, one can obtain the similar derivations. In the E-step, the optimizations of variational parameters are conducted for Eq. (9)-(12), while in the M-step, the goal is to maximize the lower bound on the log-likelihood with respect to model prarameters Eq.(13) -(20) (and also with equations for $\boldsymbol{\alpha}_2$) based on the sufficient statistics computed in E-step. The two steps iterates until the lower bound converges.

3.3 Distributed process monitoring

Once the Bayesian co-clustering has been implemented, the block division and mode recognition have also been simultaneously accomplished. For variable blocking, variables with the same latent topic are clustered together and can be viewed as generated from the same block. Therefore, the entire plant-wide process is divided into several sub-block processes. In the meantime, samples have also been automatically clustered with respect to the corresponding mode topic. It should be noticed that in this work, the actual number of variable blocks and operating modes are both assumed to be unknown a priori. In that case, some 'over-estimated' initializations could be assigned. In fact, as can be seen in the following case study, the Bayesian co-cluster method with probabilistic assignation is self-driven and those redundant clusters can be removed once the co-clustering model has been established. With all the above well-paved steps, the last step which is to build the distributed Gaussian mixture models can be intuitional results. Since the local blocks may not always indicate the global multimode characteristics, a global GMM is first constructed according to the obtained mode information. The global GMM model captures and tracks the entire operating conditions of the plant-wide process. Moreover, it can be used as an inference mode indicator for monitoring the current status of the entire plant. As mentioned above, such monitoring results can only be served as the reflection of global process operating condition. In order to localize the monitoring results, the distributed method is utilized. On the basis of co-clustered variables, all sub-models of distributed GMM can be directly extracted from the global mixture model.

In practical, it is worth to notice that as a Bayesian method which is based on probabilistic mechanism, the variable blocking results may also show some uncertainties. Still use the above five variable process as an example. In some cases, variables may be divided into more blocks than expected. Moreover, the divided blocks may show inconsistent results by different running. In order to deal with such logical but not so accurate results, a reasonable alternative is to conduct Monte Carlo methods and conduct statistical analysis from multiple random Bayesian clustering results. Specifically, assume that a total of N_t multiple experiments have been conducted. Furthermore, let N_b denotes the initial total number of blocks. A block matrix $BM_i \in \mathrm{R}^{D\times N_b}$ is defined for experiment $i, i = 1,2,...,N_t$. Each element b_{pq} in BM_i denotes the ownership probability of variable p belongs to block cluster q . For simplicity, in this work we assume that each variable belongs to the cluster with the highest assigning probability, that is:

$$\begin{cases} b_{pq} = 1, q = \arg\max_k b_{pk} \\ b_{pq} = 0, \quad \text{others} \end{cases} \tag{21}$$

After the definition of the single block matrix, we have to make a statistic analysis on the overall blocking results. In other words, the goal is to calculate the total times that certain variables are clustered in the same block. In this paper, a confusion cluster matrix $C \in R^{D\times D}$ is further defined for statistical analysis of multiple results as

$$C = BM_1 \oplus BM_2 \oplus \cdots BM_{N_t} \tag{22}$$

where the general operator $\oplus$ denotes the accumulation procedure of the same allocated cluster variables. After the accumulation of all block matrix, each element c_{mn} in the confusion matrix denotes the total times of variable n which is clustered together with variable m , while the diagonal element c_{mm} in C denotes the times that variable m should be allocated alone with a single block. Assume that all variables have been labeled with a positive integer number. Then $C = A \oplus B$ can be calculated with the following algorithm:

Algorithm 1

Input: Block matrix A, B .

Output: Confusion matrix C .

1. Find the non-zero columns (e.g. non-zero variable entries in a column can be viewed as a single variable block) of both matrixes;
2. Initialize a token set Ts with zero entries, this vector indicates the manipulated variables;
3. For each non-zero column, find the variable with the smallest label;
   - 3.1 Check if the current variable with smallest label has already been manipulated in the token set, if so, go back to step 3 for the next non-zero column;
   - 3.2 If the block contains more than one variable then the entry with respect to the smallest label should be neglected and set as zero;
   - 3.3 Add the transpose of each column to the corresponding row of C (the row number equals to the smallest label for each block);
   - 3.4 Add the variable token of non-zero entries to the token set Ts .

In order to make the multiple results more reasonable with uncertainties, block allocation probability (BAP) index is further introduced by normalizing each element of C with

N_t and $C \leftarrow C/N_t$. With the defined BAP index, X_i is allocated in the same block with X_j if

$$j = \max\left(\arg\max_j C_{ji}, \arg\max_k C_{ik}\right) \tag{23}$$

Based on the Eq.(23), blocks can be easily obtained. Assume one has finally derive N_f blocks, then all blocks can be denoted as $B_1, B_2, \cdots B_{N_f}$. With the final block information, distributed GMM can be constructed. The entire distributed modeling process is summarized as follows.

Algorithm 2

Input: Process data $\mathbf{X}$ and Labeled variables $1, 2, \cdots, D$.

Output: Blocks $B_1, B_2, \cdots B_{N_f}$ and the corresponding distributed GMM.

1. Conduct multiple Bayesian co-clustering, derive mode information and multiple block matrixes $BM_1, BM_2, \cdots, BM_{N_t}$;
2. Transform each block matrix into 0-1 entries with(21);
3. Calculate the confusion matrix C with (23) and then normalize each element;
4. Separate blocks $B_1, B_2, \cdots B_{N_f}$ from the confusion matrix C;
5. Construct the general GMM modeling based on process data $\mathbf{X}$ and the derived multimode information;
6. Extract distributed GMMs based on the separated variable blocks.

The extracted distributed GMMs can be readily applied for plant-wide process monitoring. Specifically, for each new coming sample $\mathbf{x}_{new}$, first calculate the potential mode. Then compute the corresponding BIP index defined in Eq.(2) for each GMM block. Notice that the BIP trends in each block can finally serve as a meaningful reference for distributed monitoring. In some cases, one can determine the implicated topic for each block. Although this work is focused on fault detection, the topic information by various blocks can provide important guidance for further fault identification and diagnosis. Therefore, if any sub-blocks trigger alarms, the operators can first determine the corresponding topics and then make further measurements. Finally, the monitoring performance of the entire distributed model can be denoted by an integrated index with 0-1 entries (0 for normal and 1 for fault). Basically, it is assumed that the process is in faulty condition if any sub-blocks alarm faults. For a new sample $\mathbf{x}_{new}$, let $\mathrm{BIP}_1, \mathrm{BIP}_2, \cdots, \mathrm{BIP}_{N_f}$ denote the monitoring results from all blocks and CL is the defined BIP upper control limit. The integrated index is then defined as 1 if any BIP index exceeds CL.

4 Case studies

In this section, the TE benchmark process is considered as a simulated plant-wide process. For closed loop simulation, the TE process is equipped with decentralized control strategy as described in [12]. There are 41 measured process variables collected for monitoring. Data cases are employed by 1000 samples and all faults in the testing dataset are introduced in the TE process at sample 501. As a plant-wide process, the TE benchmark consists of six different operating modes with different G/H mass ratios (stream 11). In this work, mode 1 and mode 3 are used for simulation. For Bayesian modeling, a total of 1000 samples are generated from each mode. After that, the initial guesses for modes and variable blocks are randomly selected as 4 and 12 respectively. The Bayesian co-clustering is then applied for bi-clustering of both modes and blocks. As above, the derived variable blocks through the normalized confusion matrix for variable blocks with BAP index after 50 experiments of the TE process are given in Table 1. For samples-wise clustering, a typical derived result for all samples is depicted in Figure 1.

One can infer from Table 1 that, five blocks can be finally derived according to the clustering results. It should be noticed that the variables are clustered according to the potential topics. Therefore, variables which belong to different units are clustered in the same block due to the implicated coupled units. In other words, topic can be focused on multiple units and productions. For example, it can be speculated that the topic of block 1 should be focused on components B and G; the second block mainly pays attention to implicated variables with regard to the input and production procedure of component A and C; the topic of block 3 should be relevant with levels of reactor/separator/stripper as well as the production of component D; the potential theme of block 4 should be mainly about temperature of reactor/separator/stripper/separator water, the underflow of separator/stripper and the production of component E, F and H; and the last block is focused on component C. Besides topic related variables, some blocks are also involved with attached miscellaneous variables after probabilistic dividing. For example, block 1 is attached with separator pressure and component B of reactor feed analysis; block 2 is attached with three miscellaneous variables including purge rate, reactor water temperature and component F; block 3 has three miscellaneous variables including reactor pressure, stripper steam flow and compressor work; block 4 is also attached with three miscellaneous variables containing reactor feed rate, stripper pressure and recycle flow. One can clearly see that the divided blocks can effectively reflect the essential parts of the plant-wide process. Hence, it is validated that the clustering result is consistent with the actual modeling data.

Table 1: Obtained variable blocks of the TE process

Blocks	1	2	3	4	5
Variable	2,13,24, 26,30,35,40	1,4,10,21, 23,29,31,39	7,8,12,15, 19,20,32,37	3,5,6,9,11,14,16,17, 18,22,27,28,33,34,36, 38,41	25
Topic	Components B and G	Component A and C	Levels of reactor/separator/stripper; Component D	Temperature of reactor/separator/stripper/separator water; the underflow of separator/stripper; Component E, F and H	Component C

Once the Bayesian co-clustering has been completed, distributed GMM can be easily constructed. Follow the diagram given in algorithm 2, a global GMM is first established based on the mode recognition result. Next, five separate GMMs can be extracted from the global model on the basis of variable blocking result. After then, the distributed GMM is applied for monitoring the multimode

plant-wide process. Actually, one can effectively monitor the plant-wide process by monitoring multiple distributed models with different topics. As an illustration, we take Fault 10 (IDV 10 in mode 1) as a study case. To evaluate the effectiveness of the proposed method, the distributed GMM is compared with traditional GMM. IDV10 is activated with random variations in C feed temperature (stream 4). The monitoring results of GMM and distributed GMM are given in Figure 2. The control limits denoted with red lines are set as 0.95. It can be seen that the Mode 1 is correctly recognized by the global GMM. In the meanwhile, one can infer that the global GMM can only tell the occurrence of the faulty event. On the contrast, it can be seen that distributed GMM can further indicate that only block 4 is responsible for this fault.

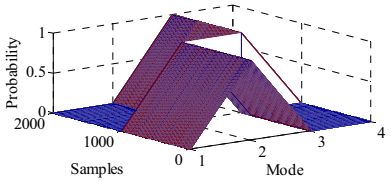

Figure 1: Mode clustering result for TE samples

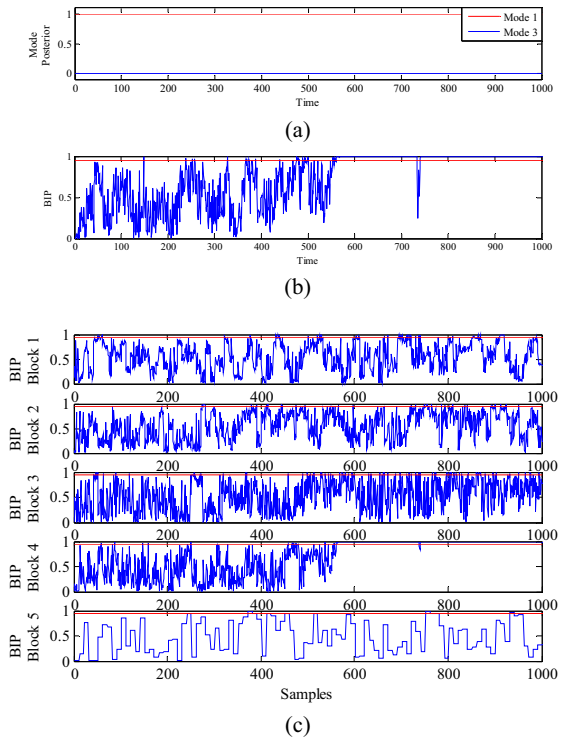

Figure 2: Monitoring results of Fault 10 with (a) Mode recognition by global GMM (b) BIP trends of global GMM (c) BIP trends of distributed GMM

Actually, as can be inferred from the TE working diagram, C feed flows into the stripper and the random variations in C feed can also result in corresponding variations of stripper temperature. However, under the control mechanism, although the variations in temperature keep on existing, impacts from IDV10 have been rejected by the slave temperature control loops. Therefore, as can be validated from Table 1 and Figure 2, only the 4th distributed GMM block alarms faults while other models indicate not being affected. On the basis of distributed modeling, operators can localize the root cause of fault according to the topic of block 4. Monitoring results by traditional GMM can hardly localize the fault area which may result in shutdown of the entire plant-wide process. However, by blocking and distributed monitoring, one may directly determine the operating condition (mode 1) and also derive the general topic of fault once abnormal event happens. In other words, the Bayesian co-clustering method provides an effective alternate for blocking and mode recognition of large scale plant-wide process in spite of the little process knowledge that can be obtained a priori.

5 Conclusions

In this work, a distributed Gaussian mixture model has been proposed for modeling and monitoring multimode pant-wide processes. Given the fact that process knowledge is not always available, the Bayesian co-clustering approach is developed to extract topic information simultaneously from both sample-wise and variable-wise directions. Based on the co-clustering mechanism, variables with same potential topics are gathered together in the same block. Meanwhile, samples are also clustered according to the topic of underlying operating modes. In order to deal with the modeling uncertainties with probabilistic allocation, multiple experiment and analysis diagram have been constructed. Simulation results on a numerical study case and the TE benchmark have demonstrated the effectiveness of the developed approach.

Inferences

[1] S. J. Qin, "Survey on data-driven industrial process monitoring and diagnosis," *Annual Reviews in Control,* vol. 36, pp. 220-234, 2012.

[2] Z. Ge and Z. Song, "Distributed PCA Model for Plant-Wide Process Monitoring," *Industrial & Engineering Chemistry Research,* vol. 52, pp. 1947-1957, 2013.

[3] Z. Ge, Z. Song, and F. Gao, "Review of Recent Research on Data-Based Process Monitoring," *Industrial & Engineering Chemistry Research,* vol. 52, pp. 3543-3562, 2013.

[4] J. F. MacGregor, C. Jaeckle, C. Kiparissides, and M. Koutoudi, "Process monitoring and diagnosis by multiblock PLS methods," *AIChE Journal,* vol. 40, pp. 826-838, 1994.

[5] S. J. Qin, S. Valle, and M. J. Piovoso, "On unifying multiblock analysis with application to decentralized process monitoring," *Journal of chemometrics,* vol. 15, pp. 715-742, 2001.

[6] B. Wang, X. Yan, Q. Jiang, and Z. Lv, "Generalized Dice's coefficient - based multi - block principal component analysis with Bayesian inference for plant - wide process monitoring," *Journal of Chemometrics,* vol. 29, pp. 165-178, 2015.

[7] H. Shan and A. Banerjee, "Bayesian co-clustering," in *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on*, 2008, pp. 530-539.

[8] D. M. Blei, "Probabilistic topic models," *Communications of the ACM,* vol. 55, pp. 77-84, 2012.

[9] S. J. Qin, "Process data analytics in the era of big data," *AIChE Journal,* vol. 60, pp. 3092-3100, 2014.

[10] J. Yu and S. J. Qin, "Multimode process monitoring with Bayesian inference - based finite Gaussian mixture models," *AIChE Journal,* vol. 54, pp. 1811-1829, 2008.

[11] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *the Journal of machine Learning research,* vol. 3, pp. 993-1022, 2003.

[12] N. Lawrence Ricker, "Decentralized control of the Tennessee Eastman challenge process," *Journal of Process Control,* vol. 6, pp. 205-221, 1996.