I&EC research
Industrial & Engineering Chemistry Research

Process Systems Engineering

Enhancing Quality of Multivariate Process Monitoring Based on Vine Copula and Active Learning Strategy

Yang Zhou, Xiang Ren, and Shaojun Li

Ind. Eng. Chem. Res., **Just Accepted Manuscript** • DOI: 10.1021/acs.iecr.7b05128 • Publication Date (Web): 22 May 2018

ACS Publications

is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

Enhancing Quality of Multivariate Process Monitoring Based on Vine Copula and Active Learning Strategy

Yang Zhou[1], Xiang Ren[2], Shaojun Li[1*]

([1]Key Laboratory of Advanced Control and Optimization for Chemical Processes, East China University of Science and Technology, Ministry of Education, Shanghai 200237, China
[2]Department of Chemical and Biochemical Engineering, Rutgers University, Piscataway, New Jersey 08854, United States)

Abstract: This paper proposes a new process monitoring method based on vine copula and active learning strategy under a limited number of labeled samples. The proposed method uses active learning strategy and the generalized Bayesian inference-based probability (GBIP) index to choose samples that can provide the most significant information for the process monitoring model. An adaptive strategy is used to select the number of training samples in every active learning loop. Then, the vine copula-based dependence description (VCDD) is used to fulfill fault detection for complex chemical processes. The validity and effectiveness of the proposed approach are illustrated using a numerical example and the Tennessee Eastman(TE) benchmark process. The results show that the proposed method can maximize the process monitoring performance while minimizing the number of samples labeled.

Keywords: Active learning; Dependence structures; C-vine copula; Fault detection

* Corresponding author. Tel.: 86-21-64253820
E-mail addresses: lishaojun@ecust.edu.cn (S.J. Li)

1. Introduction

Process monitoring is one of the key techniques that improves the competitiveness in industries[1]. On the basis of previous work, Venkatasubramanian et al.[2-3] divided the main methods and techniques used in process monitoring into three categories: methods based on the analytical model, methods based on knowledge, and methods based on data. Usually, the variables involved in the analytical model need clear definition, and the causal relationships between variables must be clarified. At present, the scale of chemical production processes is increasing, and the process is becoming increasingly complicated. Some variables are highly nonlinear, and display a high degree of coupling with each other. In addition, the mechanisms of many physical and chemical processes are not yet clear. These factors limit the practical application of the process monitoring methods based on an analytical model. Fault detection methods based on knowledge need to simulate experienced operators or specialists. The difficulty is translating qualitative or quantitative experience, reasoning, and rules into mathematical expressions, and then implementing them in software. For nonlinear, and time-varying characteristics of chemical process, data-driven methods are widely used. Data-driven methods use statistics, pattern recognition, machine learning, and technology from other fields to find and describe the change rules of the system from historical process data. The widespread application of distributed control systems (DCSes) and real time data acquisition systems in the process industry have greatly enriched process data, which enables data-based process monitoring technology based on data to develop rapidly[4].

The data-driven method considers detected outliers as faults by analyzing the process data. One of the most common ideas is to convert the high-dimensional data to low dimensional data. With this approach, a set of independent variables in different dimensions is obtained. The decoupling process is implemented, and then outliers can be detected by calculating the Hotelling's T^2 statistic and the Q statistic[5-7]. Principal component analysis (PCA)[8-10] and partial least squares (PLS)[11-15] regression analysis are the most widely used data-driven methods in chemical processes. For a non-Gaussian distribution estimation, the general idea is to use several Gaussian distributions to characterize the non-Gaussian properties of the data. The most typical method is the Gaussian mixture model (GMM)[16-18]. In addition, independent component analysis

(ICA)[19-21] is another typical projection technique for multivariate non-Gaussian data. In the field of process monitoring, fully understanding the nonlinear relationship of variables is of great significance. Gnanadesikian[22] proposed a generalized principal component analysis method in 1977 (Generalized PCA, GPCA), in which the basic idea was to map a low-dimensional nonlinear space to a high-dimensional linear space, and then apply the traditional PCA method to the high-dimensional linear space. In 1998, Scholkopf[23] proposed the kernel PCA (KPCA) method to construct the inner product function of mapping data as the inner product matrix of the elements in the feature space (a high-dimensional space). The inner product matrix of the feature space is then replaced by the corresponding kernel matrix of the sample in the original space. There is another paper[24] to solve the nonlinear problem of PCA. Yu et.al proposed a nonparametric PCA model based on nonlinear correlation measures, including Spearman's and Kendall's tau rank correlation. Yu et.al. also proposed a two-stage statistical fault detection technique to screen the important process variables[25], first, the authors identified the faulty monitored variable, and used the modified independent component analysis for fault detection; second, a Bayesian Network model is constructed to screen the important process variables. Zadakbar et.al proposed a methodology to calculate process risk in combination with a data-based fault detection method[26]. There is another paper[27] developing a methodology to combine diagnostic information from various fault detection and isolation tools to diagnose the true root cause of an abnormal event in industrial processes. To display the coupling, Ren et.al. proposed a VCDD method[28]. The VCDD method is able to monitor nonlinear and non-Gaussian process via a C-vine copula model to describe the distribution characteristics of different modes. This method constructs a kind of generalized Bayesian inference-based probability monitoring index that is suitable for arbitrary distribution. To identify and fully understand the complex dependence patterns in multivariate data, a mixture of D-vine copulas (MDVC) was proposed[29]. By using the expectation-maximization (EM) algorithm and stepwise semiparametric (SSP) estimation for parameter estimation, the proposed model can capture the complex dependence structures in multivariate data. For fault diagnosis, Yu et.al.[30] determined the reference dependence structures of the process variables from normal process data. The reference structures are compared with those obtained from the faulty data samples, then the root-cause variable(s) can be identified. Recently, Ren et.al. proposed a copula subspace division method for fault detection and diagnosis[31], margin distribution subspace (MDS)

modeled by joint margin distribution, and dependence structure subspace (DSS) modeled by copula. This method is more robust than the VCDD method for the incompleteness of training data, and achieves better monitoring performance for nonlinear and non-Gaussian processes in comparison to the traditional methods such as ICA and KPCA.

It is well known that the quality of process data seriously affect the process monitoring performance, and the potential factors probably result in incomplete data sets easily. There are many different blocks in chemical process causing missing data, such as human interface, fault detection systems, process, data acquisition, control. Commonly, the input data source of a chemical system includes off-line analyses, sensors reading and records of actuators status constitute. In addition, different types of medias transmit data, for instance fibre-optic, wireless and wire. And various software, hardware administrative troubles will result in the incomplete datasets of any situation. There are a lot of factors causing different occasional failures of sensors and actuators, such as short of energy supply, abrasion or exposure to physical damage[32]. Some parts of data are inaccurate because of maintenance of the sensors and actuators. Besides, the inapposite calibration of control values or sensors causes unreliable or out of range data[33]. Therefore, the preprocessing block of the monitoring system can detect the records voluntarily, such as the outlier detection. Various kinds of sampling rates and measurements will lead to sparse datasets[34], such as the offline analysis and the online sensing, and the administrative issues perhaps cause the morra of information[35]. Because of disruption or malfunction of the data acquisition system the data will be unavailable momentarily, for instance, induced current, higher bit rate error, break of circuit, short circuit of cables, or even harsh weather[36]. Such as the flown data, it contains noise itself, and if the DCS devices do not have a filter, this kind of data will seriously affect the accuracy of the model. For many industrial processes, some monitoring variables are of key value, because they are highly related to the faults, and the faults may cause serious changes in the related variables. If the incomplete or affected variables are ignored, it will result in a poor monitoring of the related faults. Thus, this kind of data must be revised. The adaptive active learning (AL) VCDD method we proposed mainly focus on the situations that the variables of the data contain noise, have offset or are missing, like the above mentioned cases. Generally, the process engineer can revise the variables of the data by analyzing the process and

the whole data, or use the filtering methods to filter the noise, such as the slow feature analysis[37], finally they can filter noise and revise the data one by one. However, revising the data is labor-intensive and time-consuming, and in most situations, the data is redundant, so we just need to label the data that contain the greater information. In this paper, active learning strategy is applied to the VCDD method for multivariate, multimode process monitoring to determine the number of revised data. Unlike passive learning, active learning simulates the human learning process. During the training process, the data with the most information is selected for the purpose of minimizing the labeling cost as much as possible. Active learning was originally proposed by Angluin of Yale University in the paper "Queries and concept learning"[38]. Different from previous algorithms, it uses unlabeled samples to assist the training process: unlabeled samples are selected and labeled before being added to the training dataset. The strategy is then used to choose additional unlabeled samples. Selective extension of labeled sample sets and cyclical training give the model better generalization. At present, active learning is applied in many other fields, such as soft sensing, classification, etc. Zhiqiang Ge[39,40] applied active learning strategy to the PCA method and Gaussian process regression. Ref 41 introduced applying the active learning to the SVM method, in which the author used three kinds of indexes to select the label samples. In literature[42], Li et al. used active learning support vector machines in natural language processing, selecting samples according to the distances to the hyperplane.

This study uses a joint probability density function (PDF) to evaluate information content. The joint PDF describes the probability that the sample belongs to a model. The GBIP index is used to evaluate the quality of each sample in the unlabeled dataset. Based on the probabilistic structure of a VCDD model, the reliability of a new sample can be measured by its GBIP index. The GBIP index is a distance measure. It represents the distance between test sample and the modal center. The test sample with a high GBIP index are far away from the center of the mode, meaning that this kind of sample cannot be well explained by the existing models, and thus delivers greater unknown information in active learning. If a sample follows the VCDD model, its GBIP index is small, otherwise, the GBIP index is high. Compared with samples within a VCDD model, those far away from a VCDD model, or having a relatively large GBIP index, can provide additional information for the process monitoring model. The greatest advantage of the method

proposed in this paper is to maximize the process monitoring performance while minimizing the number of samples used, and thus to reduce the costs related to human efforts. The rest of this paper is organized as follows. Section 2 introduces the preliminaries, which include the introduction of vine copula and the procedure for the VCDD method. The detailed procedure for applying active learning to a vine copula-based dependence description for multivariate multimode process monitoring is described in Section 3. Section 4 illustrates the effectiveness of the active learning VCDD (AL-VCDD) method through two examples: a numerical example, and the Tennessee Eastman benchmark process. Finally, conclusions are presented in Section 5.

2. Preliminaries

In this section, we briefly introduce copula and vine copula, and then give a brief review of the vine copula-based dependence description for multivariate multimode process monitoring.

2.1 Vine Copula

In probability theory and statistics, a copula is a multivariate probability distribution for which the marginal probability distribution of each variable is uniform. Copulas are used to describe the dependence between random variables. Sklar's theorem[43] states that any multivariate joint distribution can be written in terms of univariate marginal distribution functions multiplied by a copula function that describes the dependence structure between the variables. Copulas are popular in high-dimensional statistical applications, as they allow one to easily model and estimate the distribution of random vectors by estimating marginal distributions independently of the copula. That is to say, there is always a copula function C for an *m*-dimensional random vector of the joint cumulative distribution function (CDF) satisfying the following relationship:

$$F(\boldsymbol{x}) = F(x_1, x_2, \cdots, x_m) = C(F_1(x_1), F_2(x_2), \cdots, F_m(x_m)) \tag{1}$$

Where, $F_i(x_i)$ represents the marginal distribution function of the random variable X_i, and $F_i(x_i)$ satisfies Eq. (2)

$$F_i(x_i) = u_i = \int_{-\infty}^{x_i} f_i(\overline{x_i})\, \mathrm{d}\overline{x} \in [0,1] \tag{2}$$

$F(\square)$ is the cumulative probability density distribution function, $f(\square)$ is the probability density function. When $F(\square)$ is continuous on $\Re^m$, the C is unique in Eq. (1). When C is differentiable on the $\Re^m$, we take the derivative of both sides of Eq. (1) with respect to $\boldsymbol{x} \in \Re^m$. This way, the corresponding joint PDF can be expressed as the product of marginal PDF and copula density function C as Eq. (3)

$$f(\boldsymbol{x}) = f(x_1, x_2, \cdots, x_m) = c(u_1, u_2, \cdots, u_m) \prod_{i=1}^{m} f_i(x_i) \tag{3}$$

Where the density function c is defined as Eq. (4)

$$c(u_1, u_2, \cdots, u_m) = \frac{\partial^m C}{\partial u_1 \partial u_2 \cdots \partial u_m} \tag{4}$$

Schweizer and Wolff[43] discovered the invariance property of copula in 1981, i.e., a corresponding copula function remains invariant after the linear (or nonlinear) increasing transformation of each random variable. Copula can be regarded as a special kind of distribution function. The modeling process is actually used to estimate the corresponding copula parameters by using sample with the relevant optimization criteria. A bivariate copula optimization process is easy to implement. However, when the data dimensionality is too high, the "curse of dimensionality" occurs. In order to solve this problem, Joe[44] proposed a new type of copula, the vine copula.

The basic idea of the vine copula is to decompose multiple copulas into a number of bivariate copulas, and the optimization of a high-dimensional variable correlation structure is correspondingly transformed into the optimization of a series of bivariate copulas. This way, the number of parameters that need to be optimized each time is reduced to one or two. Because the vine copula makes full use of a copula's own specific correlation, the method has greater flexibility in capturing the conditional correlations, asymmetry, and tail dependence of high-dimensional data. Bedford and Cook[45,46] have done an in-depth study on the properties of the vine copula. Kurowicka and Cook[47] further discussed the construction of different vine structures. Aas et al.[48] gave an analytical formula for solving the conditional distribution function of copula pairs, and proposed a basic strategy for constructing vine copulas. A detailed description of vine copula theory and the correlation modeling process of vine copulas can be found in the

reference[49-52].

For a random vector $\mathbf{X}=(X_1, X_2, \cdots, X_m)$, suppose that the random variables $X_i (i = 1,2,\cdots,m)$ are exchangeable (the relative positions of the variables don't affect the shape of the distribution). The joint PDF of $\boldsymbol{x}$ can be decomposed into the following forms:

$$f(\boldsymbol{x}) = f(x_1)\prod_{t=2}^{m} f(x_t \mid x_1, x_2, \cdots, x_{t-1}) \tag{5}$$

According to the variety of vine copula decomposition strategies, the higher the dimensionality, the more the decomposition form increases. In this paper, only the most classic decomposition, the C-vine, is discussed. The joint PDF of the random variables can be expressed as Eq. (6):

$$f(\boldsymbol{x}) = \prod_{t=1}^{m} f_t(x_t) \times \prod_{i=1}^{m-1}\prod_{j=1}^{m-i} c_{i,i+j|1:i-1}(F(x_i \mid x_1, \cdots, x_{i-1}), F(x_{i+j} \mid x_1, \cdots, x_{i-1}); \theta_{i,i+j|1:i-1}) \tag{6}$$

Where, $f(\boldsymbol{x})$ is the expression of the sample $\boldsymbol{x}$'s joint PDF. The estimation process of the vine copula model has the following three steps: (1) the iterative calculation of the conditional distribution function; (2) the optimal selection of root nodes[39] in the vine copula and (3) the optimization of the bivariate copula's structure and parameters. A detailed description of C-vine copula and parameter estimation in C-vine copula can be found in the references[39-40, 42-43, 45, 53-54].

2.2 Overview of Vine Copula-Based Dependence Description for Multivariate Multimode Process Monitoring

The VCDD method shows good performance in describing the highly nonlinear and tail characteristics of complex variables, and it is more flexible in constructing multivariate correlation structures. The model abandons the general dimension reduction of high-dimensional data; instead, it analyses the complex process by capturing the relationship among the multivariate data. In addition, this method achieves real-time monitoring of the non-Gaussian nonlinear process with high density region (HDR) and density quantile approach (DQA) theory. The fault detection process based on the C-vine copula model is divided into two stages: offline modeling, and online monitoring. As for the VCDD multimode process monitoring method, the specific steps are

summarized as follows:

Offline modeling:

(1) According to expert knowledge or the clustering method, we get training samples of normal operational data under different modes.

(2) The joint PDF of different modes can be obtained by the correlation modeling of C-vine copula, including the determination of root nodes, the calculation of the conditional distribution function, the structure and parameter optimization of pair copula, the copula fit test, and the procedure for C-vine copula modeling.

(3) The joint PDF value of the sample can be calculated by the markov chain monte carlo (MCMC) method that is used to sample the joint distribution of different modes. However, if the training sample is sufficient, this step can be omitted.

(4) The discrete step size L can be determined according to the control limit (CL). Following this, the joint PDF value of the sample under different modes should be calculated to construct the static density quantile table.

Online monitoring:

(1) Calculate the posterior probability $P(C_k \mid \mathbf{x}_t^{\text{monitor}})$ that the current monitoring data $\mathbf{x}_{\text{t}}^{\text{monitor}}$ belongs to mode $C_k (k = 1,2,\cdots K)$ using Eq. (7):

$$P(C_k \mid \mathbf{x}_t) = \frac{P(C_k) \cdot p(\mathbf{x}_t \mid C_k)}{\sum_{i=1}^{K} P(C_i) \cdot p(\mathbf{x}_t \mid C_i)} \tag{7}$$

(2) Estimate the generalized local probability index of the current monitoring data under mode k by using the look-up table.

(3) Calculate the GBIP index, and then judge whether the index is out of order so as to complete the real-time process monitoring.

VCDD method has a lot of advantages. First, The VCDD method constructs the complex

correlation structure between different variables, and builds accurate distribution of each mode. Compared with conventional dimension reduction methods, such as PCA or ICA, the VCDD method doesn't have to reduce the dimension of data, thus preserveing more useful data information. It is a promising method for monitoring high dimensional nonlinear non-Gaussian systems. Second, the VCDD method handles very well for the data with tail bias characteristics, which the conventional dimension reduction methods hardly fulfill. The VCDD method can determine the internal correlation of the data, capture the correlation directly instead of reducing the dimension, thus there is no information loss. The most prominent feature of the VCDD method is to decompose the complex relationship into a series of bivariate copulas in a sparse framewotk. This kind of model can show the correlation and essential feature of different variables. Thirdly, the performance of the VCDD method is very prominent, and the performance of the VCDD method that is compared with the PCA, ICA and KPCA method has been listed in the reference 31, it can be seen that for all 21 kinds of faults of TE process, the performance of the VCDD method is better than the conventional dimension reduction methods. For the faults whose nonlinear and non-Gaussian character are very significant, the performance of VCDD method will become far better.

3. Applying Active Learning to Vine Copula-Based Dependence Description for Multivariate Multimode Process Monitoring

3.1 Active learning strategy

In industrial processes, some instruments or sensors, such as temperature or flow meters, often break down. In many cases these data have an important effect on fault detection in the monitoring process. Therefore, it is necessary to label these data. Labeling all the missing data is labor-intensive and time-consuming. Active learning strategy solves this problem by minimizing the number of labeled examples required to achieve a given level of performance. Unlike the conventional statistical modeling framework, which utilizes all the information to train the model at one time, the active learning framework, does not have a fixed number of training datasets. Instead, the number will be increased properly and sequentially. Only new favorable samples are included to update a model.

Generally, according to the sampling strategy of the unlabeled sample, the active learning algorithm can be divided into three categories: membership query synthesis[38], stream-based active learning, and pool-based active learning. Membership query synthesis was first proposed with queries criterion[38], which assumes that the learning system has a certain control over the environment and can ask for help from the human tagging. The algorithm identifies the labels of the samples, and learns the unknown concepts by asking questions[55]. Actually, the criterion for the sampling selection strategy is equal to the evaluation of the sample information, in other words, it is the degree of the influence on the generalization ability of the model when the unlabeled sample is added to the training dataset. The evaluation of the information can be either a set function or a fixed threshold. In this study, an active learning strategy with a membership query synthesis algorithm is used for fault detection.

The purpose and essence of the AL strategy is an optimization problem that solves the number of samples, and the optimal number of selected samples is a decision problem. Thus, the best settlement of this kind of decision problem is based on a probability model. The certain samples are chosen with certain probability. The probability model will be very suitable to the AL strategy. Note that the VCDD method is a probability model, and the GBIP index we used is a probability index, the method this paper proposed is based on the probability view to select different samples. In reference 39, the author used the T^2 and SPE indexes to select the samples, but the PCA method cannot solve the nonlinear problem. Besides, there are some other probability model, such as the Gaussian mixture model (GMM), the GMM method can't solve the nonlinear problem. However, the VCDD method can solve the nonlinear and non-Gaussian problem very well, and the monitoring performance of VCDD method is also better than the PCA, ICA and FGMM method. The method this paper proposed attempts to ensure that the scale of the samples to be labeled as small as possible, while the performance of process monitoring maintains or even becomes better. The key idea of this paper is based on the joint probability density function of different samples as shown in Eq. (6), which represents the probability that a sample belongs to a model. In other words, the smaller the joint probability density function of the sample is, the smaller the probability that the sample belongs to the model. However, when the training samples are normal samples, the sample with the smallest joint probability density function contains the

greatest amount of information that the existing model doesn't have. Therefore, this kind of sample should be added to the training dataset of the existing model.

3.2 Methodology

This study applies active learning strategy to fault detection with the membership query synthesis approach. The GBIP index is used to describe the distribution characteristics of different modes, and to realize real-time multimode process monitoring. The GBIP index is an extension of the Bayesian inference probability (BIP) index. The BIP index was proposed by Yu and Qin[56] to describe a mixture of Gaussian processes, shown as Eq. (8):

$$BIP = \sum_{k=1}^{K} P(C_k \mid \mathbf{x}_t) P_L^{(k)}(\mathbf{x}_t) \tag{8}$$

Where, $\boldsymbol{x}_t$ is the process data of time t, $P(C_k \mid \mathbf{x}_t)$ represents the posterior probability that the sample $\boldsymbol{x}_t$ belongs to mode C_k, and satisfies the following Bayesian equation:

$$P(C_k \mid \mathbf{x}_t) = \frac{P(C_k) \cdot p(\mathbf{x}_t \mid C_k)}{\sum_{i=1}^{K} P(C_i) \cdot p(\mathbf{x}_t \mid C_i)} \tag{9}$$

Where, $P(C_k)$ represents the priori probability that $\mathbf{x}_t$ belongs to mode C_k, $p(\mathbf{x}_t \mid C_k)$ is the likelihood function of the data $\mathbf{x}_t$. In Eq. (10), $P_L^{(k)}(\mathbf{x}_t)$ is a probability index based on the local Mahalanobis distance, and is used to measure the distance from the real data $\mathbf{x}_t$ to the mode C_k.

$$P_L^{(k)}(\mathbf{x}_t) = \Pr\left\{ D((\mathbf{x},\mathbf{C}_k) \mid \mathbf{x} \in \mathbf{C}_k) \le D((\mathbf{x},\mathbf{C}_k) \mid \mathbf{x}_t \in \mathbf{C}_k) \right\} \tag{10}$$

Where $D((\mathbf{x},\mathbf{C}_k) \mid \mathbf{x} \in \mathbf{C}_k)$ represents the Mahalanobis distance from variable $\mathbf{x}$ to mode C_k, assuming that there are K modes $C_k (k = 1,2,\cdots,K)$, and there is a complex nonlinear correlation between the process variables in each mode. All in all, first $P(x_t \mid C_k)$ is used to determine the probability that the test sample belongs to different modes. And then we defaults that this test sample belongs to the most probabilistic mode, under this premise, we will calculate the $P_L^{(k)}(\mathrm{X}_\mathrm{t}^{\mathrm{monitor}})$, which is the distance between the test sample and the center of that mode. The joint

PDF of the first k mode is defined as $f^{(k)}(\mathbf{x})$, where $x \in \Re^m$ represents the process vector. This paper uses a GBIP index, which is applicable to nonlinear, non-Gaussian multimode processes:

$$GBIP = \sum_{k=1}^{K} P(C_k \mid \mathbf{x}_t^{\text{monitor}}) P_L^{(k)}(\mathbf{x}_t^{\text{monitor}}) \tag{11}$$

Where $P(C_k \mid \mathbf{x}_{\text{t}}^{\text{monitor}})$ represents the posterior probability that the current monitoring sample $\boldsymbol{x}_t^{\text{monitor}}$ belongs to mode C_k. Different from Eq. (10), $P_L^{(k)}(\mathbf{x}_{\text{t}}^{\text{monitor}})$ represents the generalized local probability (GLP) index from $\boldsymbol{x}_t^{\text{monitor}}$ to non-Gaussian mode C_k. Because the local probability index based on the Mahalanobis distance is no longer applicable to the non-Gaussian mode, the HDR[40] is introduced to describe the GLP. According to the HDR, the GLP index of mode C_k about $\boldsymbol{x}_t^{\text{monitor}}$ is defined as:

$$P_L^{(k)}(\mathbf{x}_t^{\text{monitor}}) = \Pr(f^{(k)}(\mathrm{X}) \geq f^{(k)}(\mathbf{x}_t^{\text{monitor}})) \tag{12}$$

Where, $f^{(k)}(\mathbf{X})$ represents a one-dimensional random vector that is mapped from m random variables passed through its own joint PDF (the kth mode). Similarly, the n PDF values y_i are obtained from the n sample $\boldsymbol{x}_i$, and defined as $\boldsymbol{y} = (y_1, y_2, \cdots, y_n)$. As for the current monitoring data $\boldsymbol{x}_{\text{t}}^{\text{monitor}}$, if it satisfies:

$$q_\alpha^y = f^{(k)}(\mathbf{x}_{\text{t}}^{\text{monitor}}) \tag{13}$$

Then

$$P_L^{(k)}(\mathbf{x}_{\text{t}}^{\text{monitor}}) = 1 - \alpha \tag{14}$$

Where α represents the estimation value of PDF based on the corresponding confidence level of a quantile q_α^y. Eq. (12) – Eq. (14) show the constructing GLP index and the density quantile of a confidence level are concordant. The GLP index can be regarded as the extended form of the local probability index in Eq. (10) based on the Mahalanobis distance metric.

In this paper, the GBIP index is used to evaluate the quality of each unlabeled sample. If only

the samples containing greater information are labeled and added to the training dataset, the model quality can be improved significantly. As mentioned above, the GBIP index is a distance measure, It represents the distance between the test sample and the modal center. The test sample that has a high GBIP index is far away from the center of the mode, it means that this kind of sample cannot be well explained by the existing models, and it delivers greater unknown information in active learning. Based on the existing VCDD model, the GBIP index can estimate the information content of the sample very well. If a sample stays out of the VCDD model, its GBIP index will remain high, otherwise, it will become a low value. Compared to the samples that stay within the VCDD model, those that are found far from the VCDD model or that have relatively high GBIP index values can provide more additional unknown information for process monitoring. The algorithm flowchart of adaptive AL-VCDD method for multivariate multimode process monitoring is shown in Figure 1.

Figure 1 shows the whole steps of the adaptive AL-VCDD method, including the offline modeling and the online monitoring. For offline modeling, the GBIP index is used to select the unlabeled samples and the new model is built with the updated training samples. The built model is then used for online monitoring.

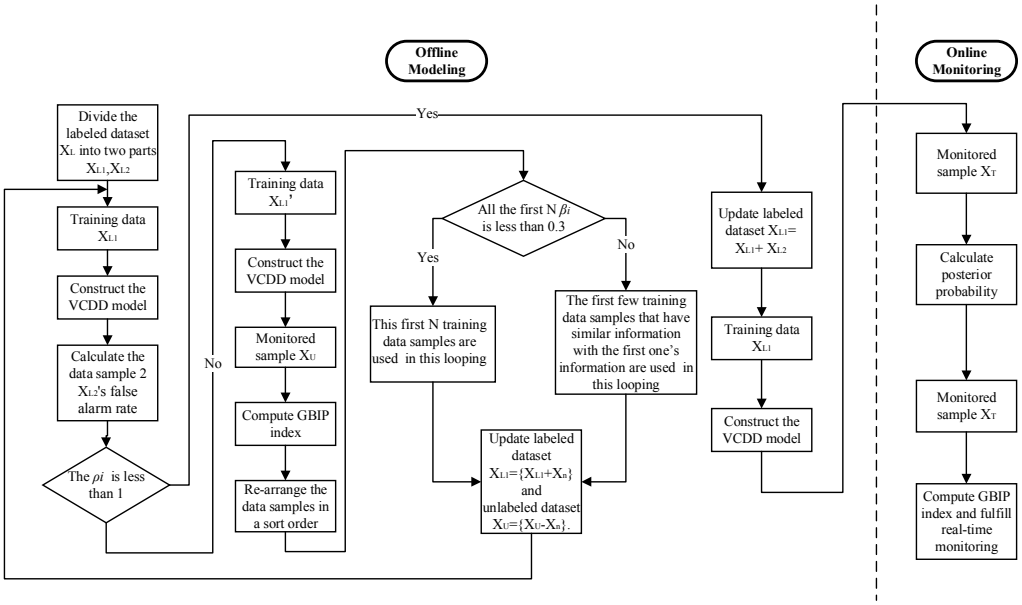


Fig. 1 The algorithm flowchart of the adaptive AL-VCDD for multivariate multimode process monitoring.

In the offline modeling process, active learning is an iterative process. An adaptive method is used to select the number of unlabeled samples in every active learning loop. In every loop, the maximum number of samples to be labeled is N. In order to choose the best samples, all the unlabeled samples are taken into a VCDD model to calculate their GBIP index. According to the GBIP indexes sorted from large to small, if the first N unlabeled data training samples have similar information, all N samples will be chosen in this loop. Otherwise, only samples with the more information will be selected. An adaptive factor β is defined to select the number of training samples in every active learning loop.

$$\beta_i = \mathrm{GBIP}(\mathrm{X}_{Ui}) - \mathrm{GBIP}(\mathrm{X}_{U(i+1)}) \quad i=2,\cdots,N \tag{15}$$

Where X_{Ui} is the i^{th} unlabeled sample, if β_i is more than 0.3, that is to say that unlabeled data samples I and i+1 don't contain similar information, the first I unlabeled data sample will be chosen for labeling. The ρ_i index is the ratio of the i^{th} X_{L2}'s false alarm rate and the $(i+1)^{th}$ X_{L2}'s false alarm rate, and the definition of ρ_i is as follow:

$$\rho_i = \frac{\mathrm{FAR}(\mathrm{X}_{L2(i)})}{\mathrm{FAR}(\mathrm{X}_{L2(i+1)})} \tag{16}$$

In order to choose the most applicable unlabeled data, after every active learning loop, a new VCDD model that uses the training samples and the labeled samples is built. If the false alarm rate (FAR) index of the normal examine samples in the new model is bigger than that of the last active learning loop (in other word, the ρ_i is less than 1), this means the model quality has declined in the new model, and the active learning process should be stopped. Otherwise, the selection of samples to be labeled should continue.

The online monitoring is the same as in the VCDD method introduced in section 2.2. The detailed process for offline modeling is as follows.

Offline Modeling:

Step 1: Divide the dataset X_L containing all variables into two parts. Part one X_{L1} is used to construct the VCDD model, and the other part X_{L2} is used to judge the terminal condition of the active learning strategy.

Step 2: Build a VCDD model with the dataset X_{L1}. The procedure of the modeling process is in the Section 2.2 offline modeling, which includes the determination of root nodes, the

calculation of the conditional distribution function, the model structure, and the parameter optimization of pair copula.

Step 3: Take X_{L2} into the VCDD model to calculate the FAR index according to the online modeling process as introduced in Section 2.2. If the FAR index is bigger than that of the last loop (in other words, ρ_i is less than 1), go to Step 8.

Step 4: Delete the variables that are lost in X_U from X_{L1} to form a new sample X_{L1}'. Build a VCDD model with the dataset X_{L1}', and then take the unlabeled samples X_U into the model to calculate the GBIP index.

Step 5: Sort the GBIP index from large to small. If the first N (maximum sample selection) training samples have similar information (in other words, β_i is less than 0.3), all the first N unlabeled samples are chosen for labeling in this looping; otherwise, the first i unlabeled samples are chosen.

Step 6: Label the chosen unlabeled samples, named X_n, by the process engineer.

Step 7: Update labeled dataset $X_{L1}= X_{L1}+X_n$ and unlabeled dataset $X_U= X_U - X_n$. Go to step 2, and re-build the VCDD model for the new labeled dataset X_{L1}.

Step 8: Update labeled dataset $X_{L1}= X_{L1}+ X_{L2}$.

Step 9: Build the VCDD model for new current labeled dataset X_{L1}. Update all parameters in the VCDD model.

4. A Numeral Example

In this section, a numeral example is used to demonstrate the effectiveness of the AL-VCDD method in dealing with non-Gaussian, nonlinearly correlated and least label data. In order to simulate the case of limited labeled samples, two kinds of data that were collected under normal operational conditions are designed as follows. The first kind of data has three modes, each of which has 100 groups, and there are three dimensions (X_1, X_2, X_3) in every mode. The second kind of data also has three modes, each of which has 1000 groups, but only two dimensions, X_1 and X_2, in every mode. The two kinds of data are used to construct the offline model. In this case, 60 groups (20 groups from each mode) of the first kind of data are randomly selected, and the other groups are used for judging the terminal condition of the active learning strategy.

Analogously, 500 groups of monitoring (testing) data are generated as follows. The process initially runs in the normal state of the third mode, and then 4 offsets are added to X_1 between the 51st and the 150th sample (fault 1). Afterward, the system returns to the normal operating condition of the third mode until the 200th sample is reached. The process then changes to the normal state of the second mode and the first mode, from the 201st to the 300th sample and the 301st to the 400th sample, respectively. In the last 100 moments, the variable X_2 is assigned to a 0.5 offset and a nonlinear drift (fault 2). Figure 2 shows the sequence diagram of 500 test data. Compared with fault 1, the deviation range of fault 2 is relatively small, especially during the period from the 401st to the 440th moments. The range of each variable under the effect of fault 2 is still in the normal range, but in fact, the correlation structure between X_1 and X_2, and between X_2 and X_3 have changed.

In order to illustrate the necessity of active learning strategy, two VCDD models are built according to different data. The first model is built by the 3000 training samples with three variables and the third variable with the Gaussian noise with the expectation and variance of 0 and 0.1, and the second model is also built by the 3000 training samples with three variables, and the third variable is added 0.1 offset, the third model is built by the 3000 training samples with the first two variables, and the fourth model is only built by the 60 training samples with three variables. Tested with 500 groups of monitoring data, thefault detection rate (FDR) of fault 1, fault 2, and the FAR of the four models are 100%, 77%, 9.3%; 100%, 85%, 14.2%; 100%, 73%, 11.1% and 100%, 75% 19.5% respectively. We can see that when a variable contains noise, has offset or are missing, the performance of the model is bad. However, there are only 60 groups of samples with complete variables, the result of process monitoring is also poor. Therefore, it is necessary to add active learning strategy to the VCDD method.

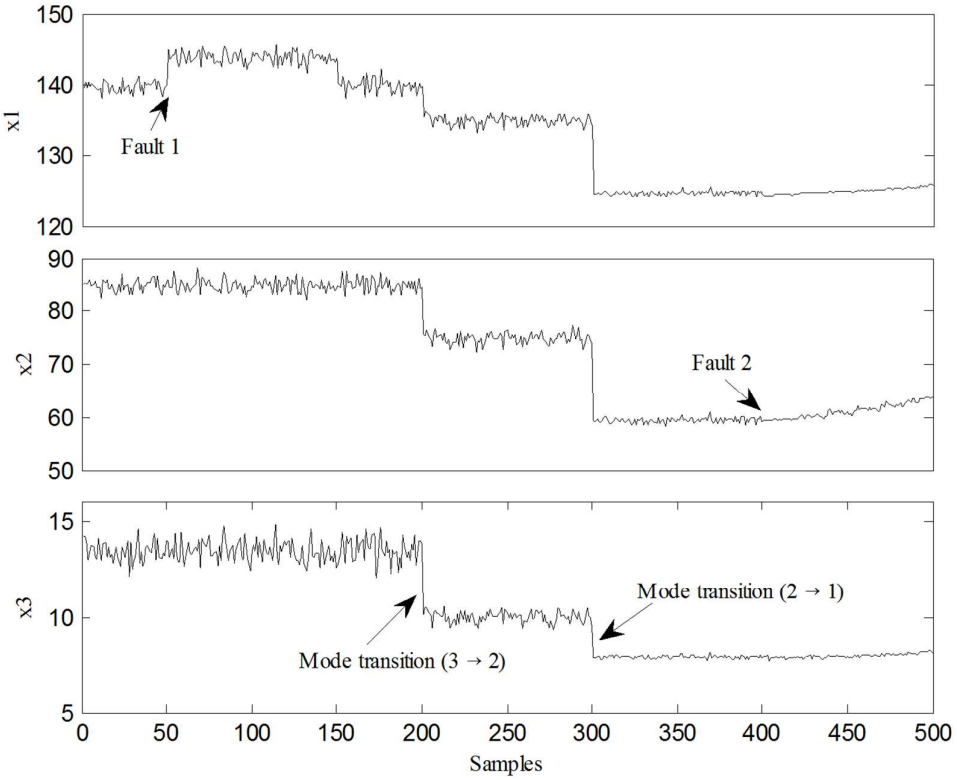
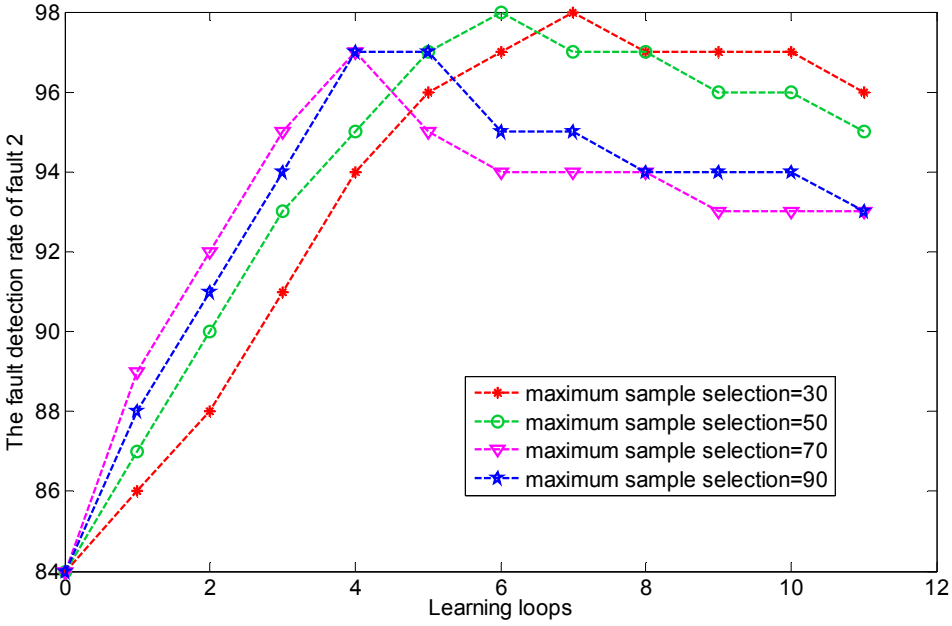


Fig. 2 Sequence diagram of the testing data



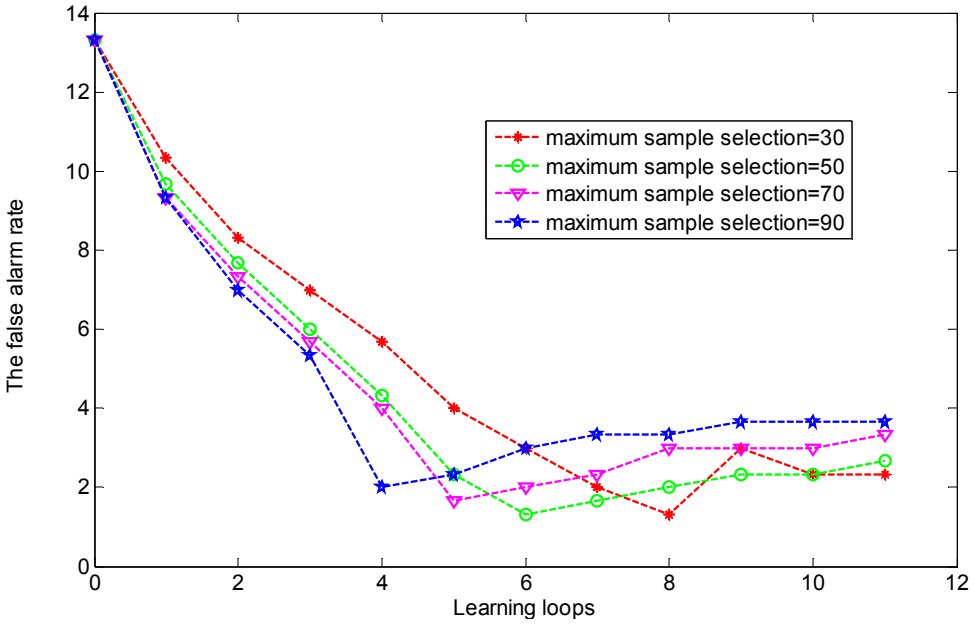


Fig. 3 The FDR of fault 2 and the FAR under different maximum sample selection values

As for active learning strategy, the number of training samples selected in one active learning loop is quite important. Training samples that are too small will require more iterations and will increase the computational burden in achieving a given performance goal. Alternatively, it would be inaccurate and would require more human effort to label samples. Hence, deciding how to select the number of unlabeled samples in each active learning loop is actually a trade-off problem between process monitoring effect and resource cost. In order to validate the influence of the number of samples selected in each active learning loop, selections with different sample maximums between 30 to 90 were tested in active learning processes in Figure 3. The FDR and the FAR of fault 2 are shown in Figure 3. It can be found that the convergence rate is accelerated with an increase in the number of learning samples in each loop. When different maximum samples are selected from the unlabeled dataset, the estimation performance is similar, but the performance of the selection with a maximum of 50 samples is a little better. It is worth noting that we initially set 50 as the maximum number of samples n selected in each active learning loop.

In order to show the effectiveness of the AL-VCDD method, two VCDD models, the AL-VCDD and the random selection VCDD (RS-VCDD), are built. The AL-VCDD and RS-VCDD methods label 50 samples in every iterative loop. For RS-VCDD, a total of 50 simulations were conducted to avoid randomness. The two models added the same number of

samples, but the samples were different.

Table 1 gives the FAR1 and FAR2 indexes of twelve active learning loops. Here, FAR1 is the FAR index of the 240 normal samples which have not been used to build the AL-VCDD model, and FAR2 is the FAR index of the 500 test samples.

Table 1. The FDR and the FAR values of the fixed-step AL-VCDD model and the RS-VCDD model

Number of training samples	The fixed-step AL-VCDD model				The RS-VCDD model		
	FAR1	FDR of fault 1	FDR of fault 2	FAR2	FDR of fault 1	FDR of fault 2	FAR2
60	15	100	84	13.33	100	84	13.33
110	13	100	87	9.67	100	86.3	12
160	11	100	90	7.67	100	87.1	10.67
210	9.33	100	92	6.33	100	86.6	11
260	7.67	100	94	4.67	100	87.5	10.33
310	6	100	96	2.67	100	88.3	10
360	**4.67**	**100**	**97**	**1.67**	100	89.8	9.67
410	5.33	100	97	2	100	90	9.67
460	5.33	100	96	2	100	90.4	10
510	5.67	100	96	2.33	100	90.8	9.33
560	5.67	100	96	2.33	100	90.5	9.33
610	6	100	96	2.67	100	91.2	8.67

As shown in Table 1, according to the modelling steps introduced in Section 3, only 6 iterative loops and 300 groups of samples are needed for labeling. In order to contrast the difference between the two models, Table 1 gives 12 iterative loops results. The FAR2 indexes of AL-VCDD are better than those of RS-VCDD model, and up to 610 samples are labeled. We reach the conclusion that the AL-VCDD model has a better fault detection effect than the RS-VCDD. In addition as can be seen after six active learning loops, fault detection performance becomes convergent. The fault detection performance of the RS-VCDD model developed may not become convergent by adding more labeled samples. In particular cases, when randomly selected samples are incorporated into the model, the performance may even deteriorate. The results show that the AL-VCDD model needs fewer samples to be labeled in order to achieve satisfactory fault detection results.

The number of training samples selected in one active learning loop is quite important. If each iteration only selects a small group of training samples to label, the result will be more

accurate. On the contrary, when a large group of training samples is chosen to label in each iteration, the modeling process may require fewer iteration loops to achieve a given performance goal. In order to examine the detailed information of the selected samples in the unlabeled dataset, the GBIP values of the first to fourth iteration loops are plotted in Figure 4. Table 2 lists the FDRs and the FARs for the model (adaptive AL-VCDD) that is developed with an adaptive active learning strategy. In order to contrast with AL-VCDD and RS-VCDD, Table 2 lists the results of the first 12 iteration learning loops.

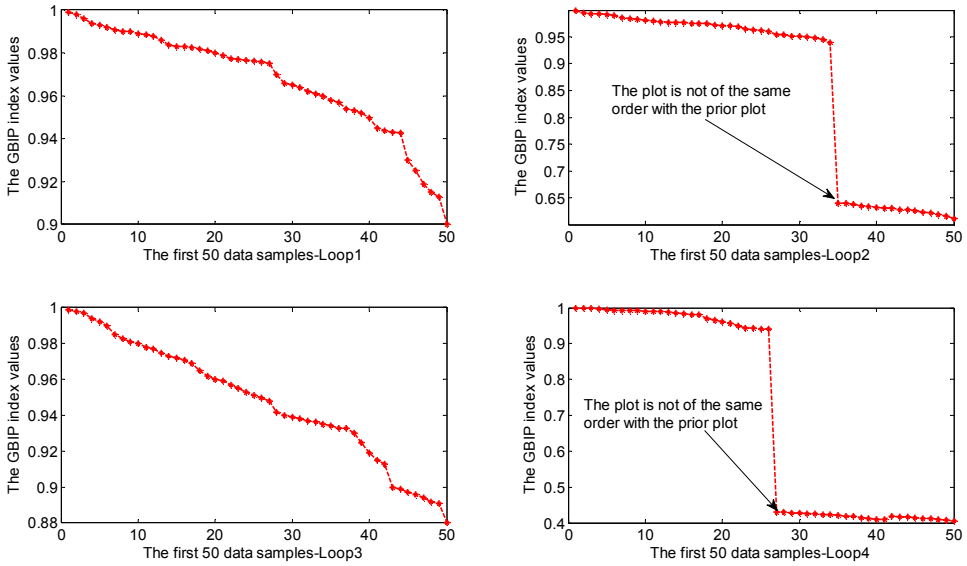


Fig. 4 The first 50 samples' GBIP values of the first to fourth iterations

It can be seen from Figure 4 that when the model is developed by using 60 training samples in the first and third iteration, most of the unlabeled sample's GBIP index values contain similar information, and all β_i is less than 0.3. Thus, all the first 50 samples are labeled and added to the AL-VCDD model of the next iteration. But for the second iteration and fourth iteration, the first 50 samples contain different information, and we need to add only the samples that contain the maximal information to the current model. We can see that the 34th sample's GBIP index value and the 35th sample's GBIP index value in the third iteration are obviously different, as the adaptive factor of the 34th sample is greater than 0.3. Therefore, we only need the first 34 samples. The fourth iteration is the same. The Figure 4 shows that the first loop model is poor. And because of the samples selected by the adaptive AL-VCDD, the update model caused the jump of index. This indicates that the update of the model is very effective. This way, the redundant samples can

be avoided. Because fewer samples need to be labeled, this prevents labor-intensive and time-consuming process. Furthermore, from Table 2 we find that the optimal FDR and FAR with adaptive active learning are better than those with the fixed active learning steps.

Table 2. The FDR and the FAR values of the adaptive AL-VCDD model

Iteration	Number of training samples	FAR1	FDR of fault 1	FDR of fault 2	FAR2
1	60	15	100	84	13.33
2	110	13	100	87	9.67
3	144	11	100	90	7.67
4	194	9	100	93	6
5	220	7.33	100	95	4.33
6	235	5.67	100	97	2.33
7	**285**	**4.33**	**100**	**98**	**1.33**
8	335	5	100	97	1.67
9	370	5.33	100	97	2
10	420	5.33	100	96	2.33
11	470	5.67	100	96	2.33
12	520	5.67	100	95	2.67

As for active learning strategy, selecting a suitable number of labeled training samples for each case is especially important. Too few training samples will lead to an unstable, worse performance; on the contrary, too many training samples will seriously affect time consumption and cost. This study used the normal samples (X_{L2} introduced in offline process) to judge the terminal condition of the active learning strategy. From the column FAR1 of Table 2, it can be seen that the specific trend of that column's values decline rapidly, followed by a slow rise. Thus, the optimal active learning loop can be determined according to this trend. In this case, we only need 225 samples to be labeled, which is far less than that of AL-VCDD and RS-VCDD. At the same time, the detection performances of the adaptive AL-VCDD method is the best of the three VCDD methods.

Table 3. The performance index of AL-PCA method

Number of training samples	T^2 index			SPE index		
	FDR of fault 1	FDR of fault 2	FAR2	FDR of fault 1	FDR of fault 2	FAR2
60	1	1	36.67	85	75	40
110	3	1.5	33.33	88	77	34.67
160	6	1.2	33	88	76	34.33

210	8	3.1	30	91	80	32
260	10	3.4	29.33	91	81	31.33
310	11	2.7	29.33	91	82	30
360	12	4	29.33	92	82	29
410	12	4	29	92	83	28.33
460	12	4	28.67	92	83	27.66
510	12	4.2	28.67	93	83	27.66
560	13	5	28	93	84	27.66
610	13	6	27	94	84	27

Table 4. The performance index of AL-GMM method

Number of training samples	FDR of fault 1	FDR of fault 2	FAR2
60	100	75	40
110	100	77	33.33
160	100	74	34
210	100	81	28.33
260	100	82	26.67
310	100	83	26
360	100	81	25.67
410	100	85	23.33
460	100	85	23.33
510	100	85	23.33
560	100	86	23
610	100	86	23

Table 3 and Table 4 show the performances of the other two methods, AL-PCA[39] and AL-GMM[40], for comparison. Figure 5 shows the indexes of adaptive AL-VCDD, RS-VCDD, AL-GMM, and AL-PCA. The results show that both the FDR and the FAR of adaptive AL-VCDD have better effects in every loop than the other three methods. The optimal results of adaptive AL-VCDD (the FDR of fault 2 is 98% and the FAR is 1.33%) are also far better than the RS-VCDD method (91.2%, 8.67%), the AL-GMM method (86%, 23%), and the AL-PCA method (84%, 27%).

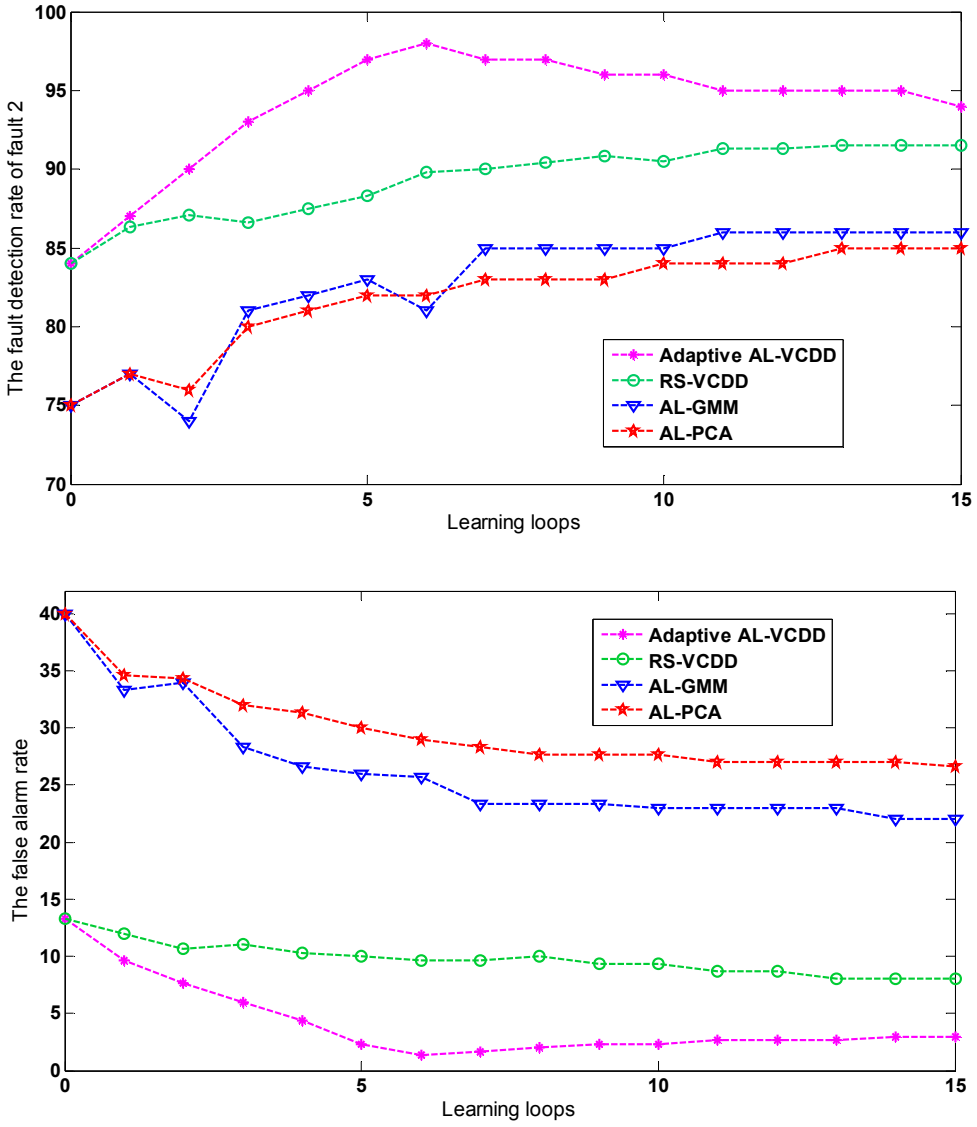


Fig. 5 The FDR of fault 2 and the FAR of the adaptive AL-VCDD, random VCDD, AL-GMM, and AL-PCA methods

The real-time monitoring chart of testing samples by the initial VCDD model and the adaptive AL-VCDD model are shown in Figure 6(a) and (b), respectively. It can be seen that process monitoring performance has been improved through the application of an adaptive active learning strategy. From the 400th sample to the 430th sample in particular, the process monitoring performances have improved significantly. As shown in Figure 6, after six iterative loops, both the

FDR and the FAR index have achieved good results. And it can be seen that the result of the adaptive AL-VCDD method is a little better than the result of putting all the samples into the model, and the number of labeled samples is very small. The adaptive AL-VCDD method achieves maximization of the process monitoring performance while minimizing the number of samples labeled.

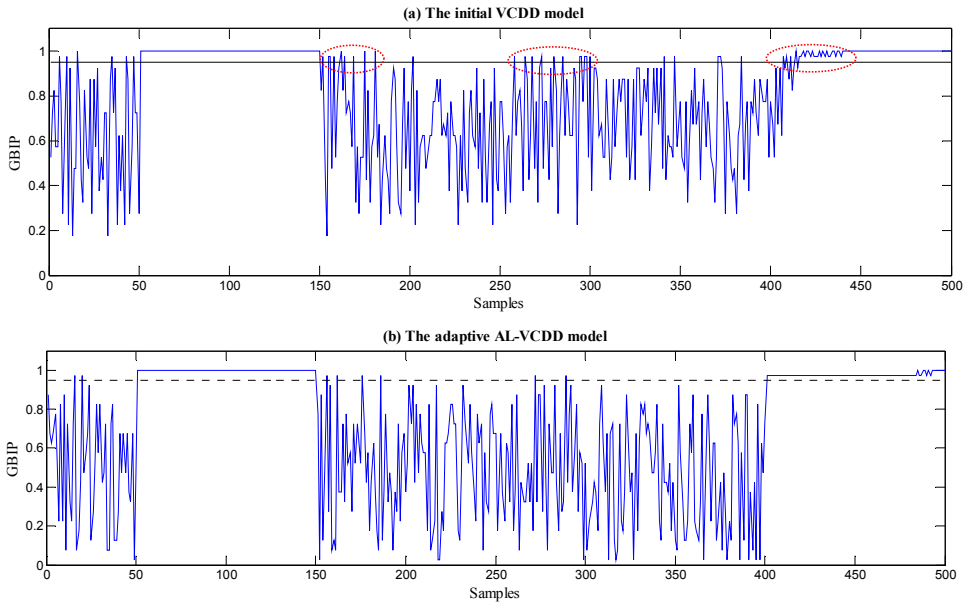


Fig. 6 The real-time monitoring chart of the initial VCDD model and the adaptive AL-VCDD model

5. Application to the TE process

Tennessee Eastman, as a typical case of a complex process, is widely used in the effect evaluation of process monitoring methods. This process mainly consists of five basic units: reactor, condenser, stripper, compressor, and separation tower. Each unit builds a series of algebraic equations and differential equations according to the material balance, energy balance, and vapor-liquid equilibrium[57]. Figure 7 shows the flow chart of the TE process.

The TE process consists of 22 continuous process variables, 19 component variables, 12 operating variables, and 21 kinds of faults. According to the varying mass ratio G/H of stream 11, the TE process can be further divided into 6 kinds of operating modes. Because the initial design

of the TE process is an open-loop unstable system, this section only discusses the TE simulation platform based on the decentralized control strategy designed by Ricker[58, 59]. All the sample time of the samples is preset as 0.05 hours.

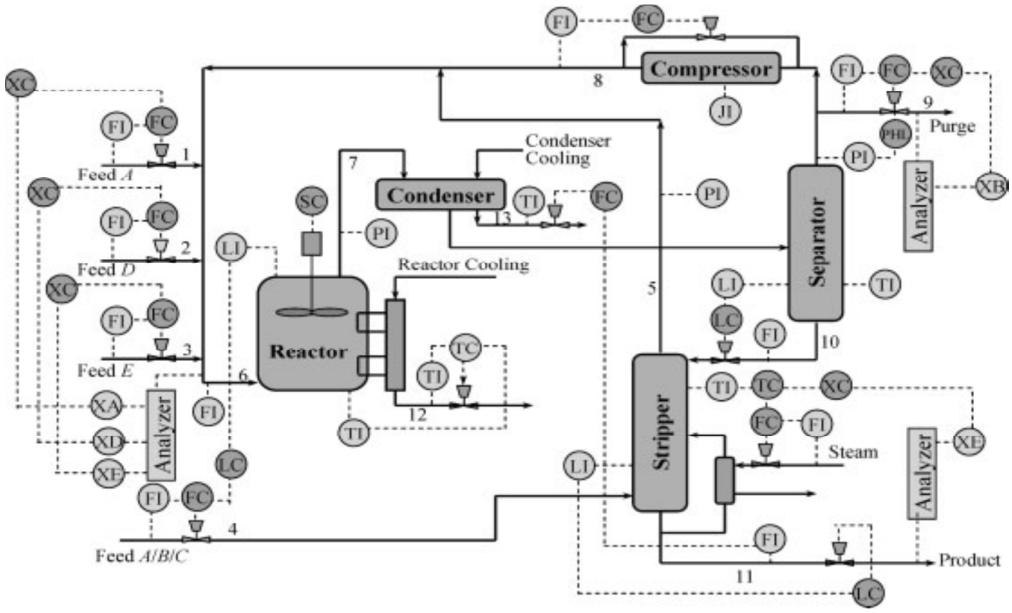


Fig. 7 Flowchart of TE process.

In this TE process, there are two groups of training samples: The first part has two modes (mode 1 and mode 3), each of which has 100 sets of training data, and there are 22 dimensions in every mode; The other group also has the two modes, each of which has 1000 sets of training data, but there are only 21 dimensions (the XME AS10 (Purge rate (stream 9) is missing) in every mode. There are 400 sets of test monitoring data defined as follows: The first 100 sets are of the normal operation in mode 3, and then from the 101st to the 200th moments, fault 13 (the slow drift change of reaction kinetics) happens, and for the subsequent 100 moments, the process switches to mode 1, and finally, for the 301st to the 400th moments, fault 6 (the step change of A feed loss in stream 1) occurs.

The two indexes FDR and FAR of the adaptive AL-VCDD ,the RS-VCDD, the AL-PCA and the AL-GMM are given in Figure 8.

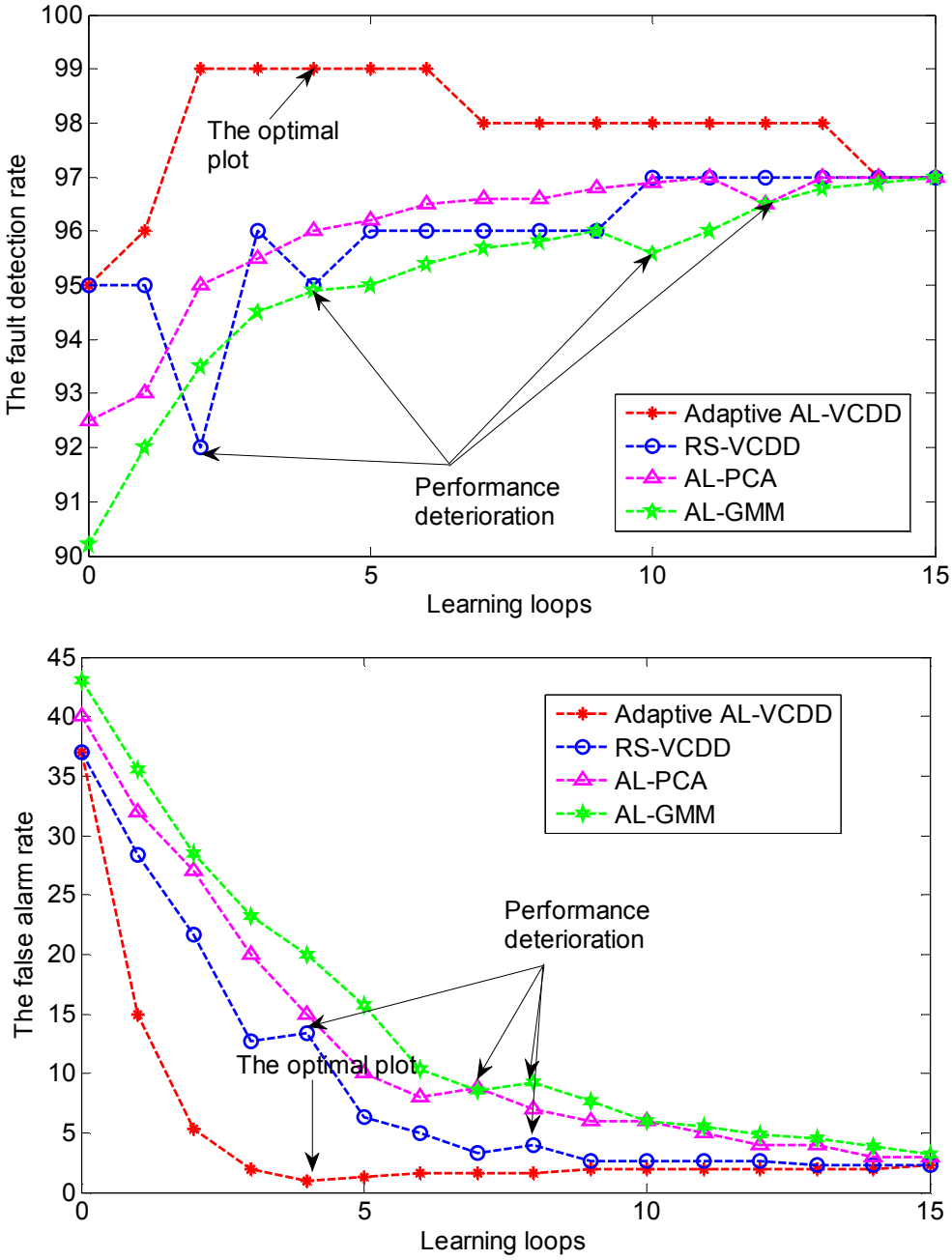


Fig. 8 The dramatic contrast of the adaptive AL-VCDD, RS-VCDD, AL-PCA and AL-GMM

We can conclude from the comparison of Figure 8 with the RS-VCDD that the adaptive AL-VCDD model has a higher FDR and a lower FAR in each loop. In the adaptive AL-VCDD method, most information samples are selected for model construction, but this cannot be guaranteed in the random selection strategy. As highlighted in Figure 8, it can be seen that after the fourth iteration, the monitoring performance of the adaptive AL-VCDD model no longer improves. However, the situation is quite different in the RS-VCDD model, which requires more loops. Thus,

compared to the random selection strategy, fewer samples need to be labeled in the active learning strategy in order to construct a satisfactory VCDD model. Compared with the AL-PCA and the AL-GMM method, the adaptive AL-VCDD method is also the best. The two indexes FDR and the FAR have a better fault detection results than the AL-PCA and AL-GMM method, and the number of samples that need to be labeled is also the smallest.

To test the process monitoring effect of the adaptive AL-VCDD model, Figure 9 shows the real-time monitoring chart of the initial VCDD model and the adaptive AL-VCDD model.

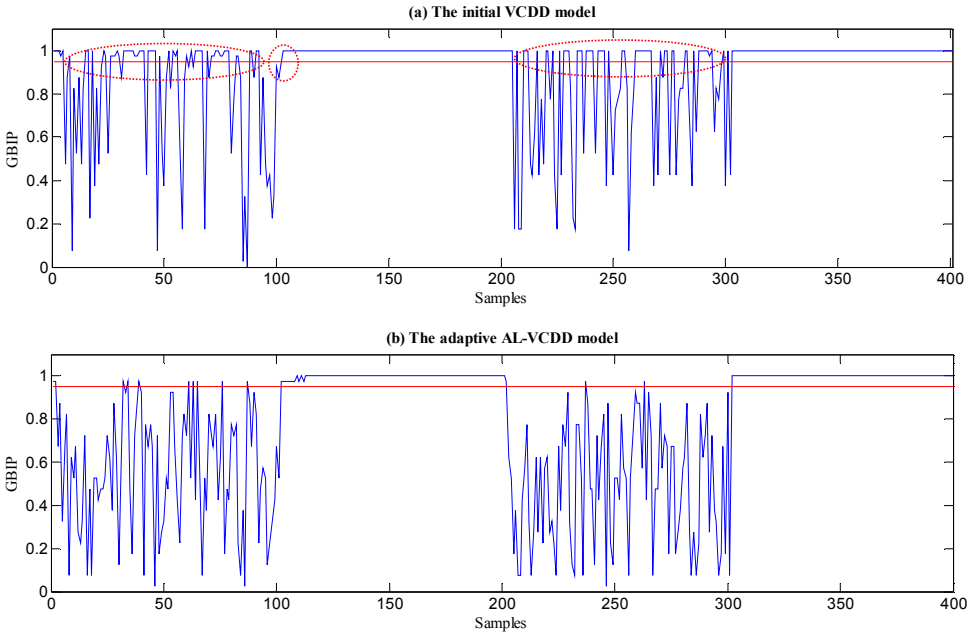


Fig. 9 The real-time monitoring chart of the initial VCDD model and the adaptive AL-VCDD model

Generally, as the learning loop increases, the process monitoring effect improves. According to the results of the comparison, we can see that the process monitoring effect of the adaptive AL-VCDD model is better than that of the initial VCDD model. Besides, the results show that the final result of the adaptive AL-VCDD method is a little better than the RS-VCDD model, and the number of labeled samples is very small. Thus, it can be seen that the adaptive AL-VCDD method achieves maximization of the process monitoring performance while minimizing the number of samples labeled.

6. Conclusions

In this study, active learning strategy is applied to the VCDD method to solve the problem of samples that is affected, contain noise, have offset or are missing and so on. Applying active learning strategy to the VCDD method can minimize the number of labeled samples to achieve a given level of performance. The study uses the membership query synthesis active learning strategy and the GBIP index to label the samples that can provide the most significant information for the VCDD model. The method uses adaptive strategy to select the number of training samples in every active learning loop, and uses some of the non-training samples to judge the terminal condition of the offline model. A numerical example and the TE benchmark process are used to illustrate the effectiveness and benefits of the adaptive AL-VCDD method. The performance of the adaptive AL-VCDD method can be significantly improved with the use of active selection strategy for unlabeled data samples. All in all, the proposed method can maximize the process monitoring performance while minimizing the number of samples used, and thus to reduce the costs related to human efforts. In the future study, we may focus on selecting the initial samples and the maximum number of samples.

Acknowledgments

The authors of this paper appreciate the National Natural Science Foundation of China (under Project No. 21676086 and No. 21406064) and the Fundamental Research Funds for the Central Universities under Grant 222201717006 for their financial support.

References

(1) Venkatasubramanian, V.; Rengaswamy, R.; Yin, K.; Kavuri, S. N. A review of process fault detection and diagnosis Part I: Quantitative model-based methods. *Comput. Chem. Eng.* **2003**, 27, 293−311.

(2) Venkatasubramanian, V.; Rengaswamy, R.; Kavuri, S. N. A review of process fault detection and diagnosis Part II: Quantitative model and search strategies. *Comput. Chem. Eng.* **2003**, 27, 313−326.

(3) Venkatasubramanian, V.; Rengaswamy, R.; Kavuri, S. N.; Yin, K. A Review of Process Fault Detection and Diagnosis, Part III: Process History Based-Methods. *Comput. Chem. Eng.* **2003**, 27, 327−346.

(4) Qin, S. J. Statistical process monitoring: basics and beyond. *J. Chemom*. **2003**, 17, 480-502.

(5) Hotelling H. Multivariate quality control, illustrated by the air testing of sample bombsights; New York: McGraw-Hill, 1947.

(6) Ge, Z.; Yang, C.; Song, Z. Improved kernel PCA-based monitoring approach for nonlinear processes. *Chem. Eng. Sci.* **2009**, 64, 2245-2255.

(7) Zhang, Y.; Hu, Z. On-line batch process monitoring using hierarchical kernel partial least squares. *Chem. Eng. Res. Des*. **2011**, 89, 2078-2084.

(8) Jung, D.; Wang, H.; Kim, J.; Lee, S. Fault detection method with PCA and LDA and its application to induction motor. *J. Central South University of Technology*. **2010**, 17, 1238-1242.

(9) Xie, Y.; Wang, J.; Liang, Y.; Sun, L. et al. Robust principal component analysis by projection pursuit. *J. Chemom*. **1993**, 7, 527-541.

(10) Chen, J.; Bandoni, J. A.; Romagnoli, J. A. Robust PCA and normal region in multivariate statistical process monitoring. *AIChE J.* **1996**, 42, 3563-3566.

(11) Kaneko, H.; Funatsu, K. Development of soft sensor models based on time difference of process variables with accounting for nonlinear relationship. *Ind. Eng. Chem. Res.* **2011**,50, 10643-10651.

(12) Kaneko, H.; Funatsu, K. Adaptive soft sensor model using online support vector regression with time variable and discussion of appropriate hyperparameter settings and window size. *Comput. Chem. Eng.* **2013**, 58, 288-297.

(13) Kaneko, H.; Funatsu, K. Application of online support vector regression for soft sensors. *AIChE J.* **2014**, 60, 600-612.

(14) Geladi, P.; Kowalski, B. R. Partial least-squares regression: a tutorial. *Anal. Chim. Acta*. **1986**, 185(86), 1-17.

(15) Höskuldsson, A. PLS regression methods. *J. Chemom*. **1988**, 2, 211-228.

(16) Bishop, C. M. Neural networks for pattern recognition. New York: Oxford University Press. 1995.

(17) Duda, R. O.; Hart, P. E.; Stork, D. G. Pattern classification, 2nd edition. New York: Wiley. 2001.

(18) Figueiredo, M. A. T.; Jain, A. K. Unsupervised learning of finite mixture models. *IEEE Transactions on Pattern Analysis & Machine Intelligence*. **2002**, 24, 381-396.

(19) Hyvärinen, A.; Oja, E. Independent component analysis: algorithms and applications. *Neural Networks*. **2000**, 13, 411-430.

(20) Bingham, E.; Hyvärinen, A. A fast fixed-point algorithm for independent component analysis of complex valued signals. *Int. J. Neural Syst*. **2000**,10, 1-8.

(21) Hyvärinen, A. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*. **1999**, 10, 626-634.

(22) Gnanadesikian, R. Methods for statistical data analysis of multivariate observations. New York: Wiley. 1977.

(23) Scholkopf, B.; Smola, A.; Mguller, K. R. Nonlinear component analysis as kernel eigenvalue problem. *Neural Computation*. **1998**, 10, 1299-1319.

(24) Yu, H.; Khan, F.; Garaniya, V. A sparse PCA for nonlinear fault diagnosis and robust feature discovery of industrial processes. *AIChE J.* **2016**, 62,1494-1513.

(25) Yu, H.; Khan, F.; Garaniya, V. Modified Independent Component Analysis and Bayesian Network based Two-stage Fault Diagnosis of Process Operations. *Ind. Eng. Chem. Res.* **2015**, 54, 1-19.

(26) Zadakbar, O.; Imtiaz, S.; Khan, F. Dynamic Risk Assessment and Fault Detection Using Principal Component Analysis. *Ind. Eng. Chem. Res.* **2013**, 52:809-816.

(27) Gharahbagheri, H.; Imtiaz, S.; Khan, F. Root Cause Diagnosis of Process Fault Using KPCA and Bayesian Network. *Ind. Eng. Chem. Res.* **2017**, 56: 2054-2070.

(28) Ren, X.; Tian, Y.; Li, S. Vine Copula-Based Dependence Description for Multivariate Multimode Process Monitoring. *Ind. Eng. Chem. Res.* **2015**, 54, 10001-10019.

(29) Zheng, W.; Ren, X.; Zhou, N. et al, Mixture of D-Vine copulas for chemical process monitoring. *Chemom. Intell. Lab. Syst*. **2017**, 169, 19-34, .

(30) Yu, H.; Khan, F.; Garaniya, V. A probabilistic multivariate method for fault diagnosis of industrial processes. *Chem. Eng. Res. Des*. **2015**, 104,306-318.

(31) Ren, X.; Zhu, K.; Cai, T. Fault Detection and Diagnosis for Nonlinear and Non-Gaussian

Processes Based on Copula Subspace Division. *Ind. Eng. Chem. Res.* **2017**, 56, 11545-11564.

(32) Zhao, C.; Fu, Y. Statistical analysis based online sensor failure detection for continuous glucose monitoring in type I diabetes. *Chemom. Intell. Lab. Syst*. **2015**,144,128–37.

(33) Wang, Y.; Yang, Y.; Zhou, D.; Gao, F. Active fault-tolerant control of nonlinear batch processes with sensor faults. *Ind. Eng. Chem. Res.* **2007**, 46, 9158–9169.

(34) Kadlec, P.; Grbi´c, R.; Gabrys, B. Review of adaptation mechanisms for data-driven soft sensors. *Comput. Chem. Eng.* **2011**, 35, 1–24.

(35) Warne, K.; Prasad, G.; Rezvani, S.; Maguire, L. Statistical and computational intelligence techniques for inferential model development: a comparative evaluation and a novel proposition for fusion. *Engineering Applications of Artificial Intelligence*. **2004**,17, 871–885.

(36) Ji, C.; Elwalid, A. Measurement-based network monitoring and inference: scalability and missing information. *IEEE Journal On Selected Areas In Communications*. **2002**, 20, 714–725.

(37) Wiskott, L.; Sejnowski, T. Slow feature analysis: unsupervised learning of invariances. *Neural Computation*. **2002**, 14: 715- 770.

(38) Angluin, D. Queries and concept learning. *Machine Learning*. **1988**, 2, 319-342.

(39) Ge, Z. Active learning strategy for smart soft sensor development under a small number of labeled data samples. *J. Process Control*. **2014**, 24, 1454-1461.

(40) Ge, Z. Active probabilistic sample selection for intelligent soft sensing of industrial processes. *Chemom. Intell. Lab. Syst*. **2016**, 151, 181-189.

(41) Chen, Y.; Cao, H.; Mei, Q. Applying active learning to supervised word sense disambiguation in MEDLINE. *Journal of the American Medical Informatics Association*. **2013**, 20, 1001.

(42) Li, Y.; Bontcheva, K.; Cunningham, H. Using uneven margins SVM and perceptron for information extraction. *Conference on Computational Natural Language Learning*. **2005**, 72-79.

(43) Schweizer, B.; Wolff, E. F. On nonparametric measures of dependence of random variables. *Ann. Stat.* **1981**, 9, 879-885.

(44) Joe, H. Families of m-Variate distributions with given margins and m(m-1)/2 bivariate dependence parameters. *Distributions with Fixed Marginals & Related Topics Lecture Notesmonograph*. **1996**, 28, 120-141.

(45) Bedford, T.; Cooke, R. M. Probability density decomposition for conditionally dependent random variables modeled by Vines. *Annals of Mathematics & Artificial Intelligence*. **2001,** 32, 245-268.

(46) Bedford, T.; Cooke, R. M. Vines: a new graphical model for dependent random variables. *Ann. Stat.* **1999**, 30, 1031-1068.

(47) Kurowicka, D.; Cooke, R. M. Uncertainty analysis with high dimensional dependence modeling. Chichester: Wiley. 2006.

(48) Aas, K.; Czado, C.; Frigessi, A. et al. Pair-Copula constructions of multiple dependence. *Insurance Mathematics & Economics*. **2009**, 44, 182-198.

(49) Kurowicka, D.; Joe, H. Dependence modeling: Vine Copula handbook. New Jersey: World Scientific. 2011.

(50) Genest, C.; Favre, A. C. Everything you always wanted to know about Copula modeling but were afraid to ask. *J. Hydrologic Engineering*. **2007**, 12, 347-368.

(51) Kim, D.; Kim, J. M.; Liao, S. M. et al. Mixture of D-Vine Copulas for modeling dependence. *Computational Statistics & Data Analysis*. **2013**, 64, 1-19.

(52) Smith, M.; Min, A.; Czado, C. et al. Modelling longitudinal data using a Pair-Copula decomposition of serial dependence. *J. Am. Stat. Assoc.* **2010**, 105, 1467-1479.

(53) Schweizer, B.; Sklar, A. Probabilistic metric spaces. Mineola: Dover Publications. 2005.

(54) Alexander, J. M. Sampling nested Archimedean copulas. *J. Statistical Computation & Simulation*. **2008**, 78, 567-581.

(55) Baum, E. B.; Lang, K. Query learning can work poorly when a human oracle is used. *Proc of IEEE IJCNN*. **1992**, 335-340.

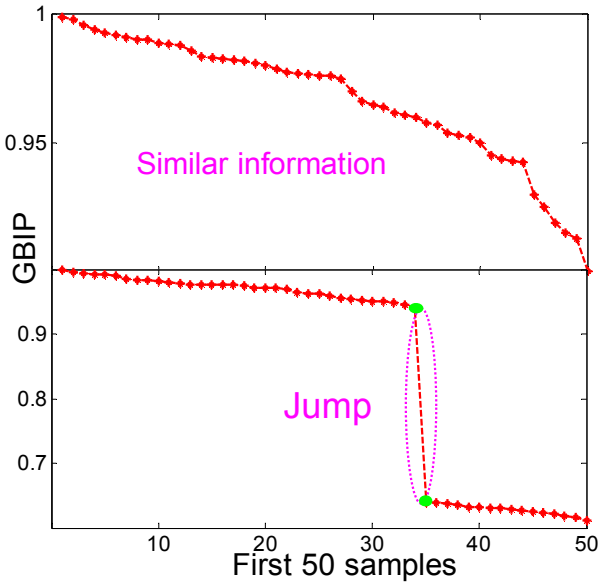
(56) Yu, J.; Qin, S. J. Multimode process monitoring with Bayesian inference-based finite Gaussian mixture models. *AIChE J.* **2008**, 54, 1811-1829.

(57) Downs, J. J.; Vogel, E. F. A plant-wide industrial process control problem. *Comput. Chem. Eng.* **1993**, 17, 245-255.

(58) Ricker, N. L. Decentralized control of The Tennessee Eastman challenge process. *J. Process Control*. **1996**, 6, 205-221.

(59) Ricker N. L. Tennessee Eastman challenge archive [EB/OL]. http://depts.washington.edu/control/LARRY/TE/download.html, 2002.

TOC graphic:


1
0.95
0.9
0.8
0.7
GBIP
Similar information
Jump
10
20
30
40
50
First 50 samples