Reliability Engineering and System Safety 119 (2013) 150–157

Contents lists available at SciVerse ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

Online probabilistic operational safety assessment of multi-mode engineering systems using Bayesian methods

Yufei Lin, Maoyin Chen, Donghua Zhou *

Department of Automation, Tsinghua University, Beijing, PR China

ARTICLE INFO

Article history:
Received 29 November 2012
Received in revised form
6 May 2013
Accepted 13 May 2013
Available online 3 June 2013

Keywords:
Operational safety assessment
Multi-mode engineering systems
Bayesian method
Sample dependency

ABSTRACT

In the past decades, engineering systems become more and more complex, and generally work at different operational modes. Since incipient fault can lead to dangerous accidents, it is crucial to develop strategies for online operational safety assessment. However, the existing online assessment methods for multi-mode engineering systems commonly assume that samples are independent, which do not hold for practical cases. This paper proposes a probabilistic framework of online operational safety assessment of multi-mode engineering systems with sample dependency. To begin with, a Gaussian mixture model (GMM) is used to characterize multiple operating modes. Then, based on the definition of safety index (SI), the SI for one single mode is calculated. At last, the Bayesian method is presented to calculate the posterior probabilities belonging to each operating mode with sample dependency. The proposed assessment strategy is applied in two examples: one is the aircraft gas turbine, another is an industrial dryer. Both examples illustrate the efficiency of the proposed method.

1. Introduction

Operational safety is a crucial issue in industry that has attracted tremendous attention from both academia and industry in the past several decades. It is well-known that engineering systems' operational safety may deteriorate as the process characteristics drift with time, which results in a dangerous condition. Therefore, the assessment of operational safety is crucial for engineering systems.

Conventional safety assessment of engineering systems consists of qualitative methods like process hazards analysis (PHA), hazard and operability analysis (HAZOP), failure mode and effects analysis (FMEA) [1–3] and quantitative methods like fault tree analysis (FTA) [4–6], Markov models [7] and reliability block diagram [8]. Recently, some interdisciplinary methods including Bayesian approach [9–11], complex networks [12], fuzzy logic or clustering [13,14] and data driven approach were adopted to assist the safety assessment. The data driven approach is usually termed as multivariate statistical process control (MSPC), which can avoid the extremely difficult and time-consuming work of constructing complex mechanistic models for engineering systems. However, those methods commonly work off-line and do not consider multi-mode operating scenarios, which are very likely to happen in practical engineering systems. For instance, there are multi-modes in a turbine engine corresponding to the different gears of the aircraft. Furthermore, since there are several operating modes, sample dependency between modes is very likely to happen. As far as we know, there exists no work on the online probabilistic assessment of operational safety of multi-mode engineering systems with sample dependency.

In this paper, based on the definition of safety index (SI) [15], we evaluate the operational safety of multi-mode engineering systems with consideration of sample dependency by probabilistic methods. We follow the idea of Bayesian methods on control loop diagnosis [16], and calculate the posterior probabilities of sampled data belonging to each operating mode. The sample dependency here follows a Markov process. This consideration allows us to make use of all time information captured by sensors. Subsequently, to get a definite assessment of the current operational safety, an online SI belonging to each operating mode is obtained.

The remainder of this paper is organized as follows. In Section 2, a brief introduction of Gaussian mixture model (GMM) is given. The specific operational safety assessment strategy including definitions and calculations of safety indices belonging to each operating mode with sample dependency are presented in Section 3. Section 4 gives two examples to show how to calculate the online SI for a multi-mode engineering system and illustrate the efficiency of the proposed method. Section 5 concludes this paper.

2. Gaussian mixture model

Most real complex engineering systems are running at multiple operating conditions and a multimodal distribution can be used to describe this kind of system. It was stated that an arbitrary probability density can be approximated by a mixture model [17,18]. So a mixture model is chosen here to characterize the

* Corresponding author. Tel.: +86 010 62783125; fax: +86 010 62786911.
E-mail address: zdh@mail.tsinghua.edu.cn (D. Zhou).

0951-8320/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ress.2013.05.018

operating data distribution of multi-mode engineering systems. According to central limit theorem [19], the mean of Nrandom variables asymptotically follows a Gaussian distribution. In each single operating mode, multivariate Gaussian distribution is a reasonable assumption for the process data. Therefore, GMM is adopted in this paper for characterizing the sampled data from multiple operation modes.

The Gaussian distribution probability density of one single operating mode is expressed by

$$p(x|\theta) = \frac{1}{(2\pi)^{m/2}\det(\sum)^{1/2}} \exp\left[-\frac{1}{2}(x-\mu)^T \sum^{-1}(x-\mu)\right] \quad (1)$$

where $x \in R^m$ denotes the m-dimensional vector of sampled process variable, $\mu \in R^m$ and $\sum \in R^{m\times m}$ represent the mean and covariance of the components.

Note that the safety assessment is online. Based on the information of sampled variables, we here use the expectation-maximization (EM) algorithm to estimate the parameters μ and $\sum$, which has two iterative steps including expectation calculation and maximization as follows [19,20]:

1. E-step: The posterior probabilities of the ith training sample x_i in the sth iteration are calculated by Bayesian law:

$$P^{(s)}(M|x_i) = \frac{P(M)^{(s)}P(x_i|\mu^{(s)}, \Sigma^{(s)})}{\sum_{k=1}^{K} P(M_k)^{(s)}P(x_i|\mu_k{}^{(s)}, \Sigma_k{}^{(s)})} \quad (2)$$

where M denotes one single operating mode, $P(M)$ represents the prior probability of operating mode M.

2. M-step:

$$\mu_k{}^{(s+1)} = \frac{\sum_{i=1}^{n} P^{(s)}(M|x_i)x_i}{\sum_{i=1}^{n} P^{(s)}(M|x_i)} \quad (3)$$

$$\Sigma_k{}^{(s+1)} = \frac{\sum_{i=1}^{n} P^{(s)}(M|x_i)(x_i-\mu_k^{(s+1)})(x_i-\mu_k^{(s+1)})^T}{\sum_{i=1}^{n} P^{(s)}(M|x_i)} \quad (4)$$

$$P(M)^{(s+1)} = \frac{\sum_{i=1}^{n} P^{(s)}(M|x_i)}{n} \quad (5)$$

These two steps are iterated until the likelihood converges.

3. Online probabilistic operational safety assessment

Assume that a system under assessment has K operating modes. When the process is running at an operating mode, the probability distribution is characterized by the *ith* Gaussian component of the previously obtained GMM as

$$x \sim N(\mu_i, \textstyle\sum_i) \quad (6)$$

where $x \in R^m$ is the vector of sampled variables with $\mu_i \in R^m$ and $\sum_i \in R^{m\times m}$ representing the mean and covariance, respectively.

3.1. Safety index of single operating mode

Now we introduce the SI in detail, which originates from the work [15]. In practical cases, most engineering systems have the safety limits or process variables' security border, within which the system's operation is considered to be safe. Without loss of generality, the safety limits here can be obtained according to the process constraints and represented by equations

$$g_r(x) = 0 \quad (7)$$

where $r = 1, 2, \ldots l$, l is the total number of safety limits. The expressions $g_r(x)$ can be linear or nonlinear functions. And $g_r(x) \le 0$ means the variable x being within the safety limits.

For one single operating mode, contours of probability density are distributed around the operating mean point and can be described by the annular hyperellipsoids in the high-dimensional space, as shown in Fig. 1. The first contour that reaches the safety limits is considered as the critical contour. Let's denote the critical contour by B. The system is thought to be safe if the variables are kept inside B. Thus the safety index of one single operating mode is defined by [15]

$$SI = \Pr(x \text{ is inside } B) \quad (8)$$

It can be inferred from Eq. (1) that in one single operating mode, the probability density is determined only by the exponential term $(x-\mu)^T(\sum)^{-1}(x-\mu)$, which is equal to the Mahalanobis distance from the center to the points on the hyperellipsoid as

$$d_M(x,\mu) = \sqrt{(x-\mu)^T \Sigma^{-1}(x-\mu)}$$

Using δ to denote the Mahalanobis distance between the critical contour and the mean point. Then the safety index can be rewritten as;

$$SI = \Pr(x \text{ is inside } B) = \Pr(d_M(x,\mu) \le \delta) \quad (9)$$

To calculate the safety index, we first need to obtain the value of δ. However, in the high-dimensional space, it is very difficult to find the critical contour. But as can be seen from Fig. 1, this task is equivalent to finding the minimum Mahalanobis distance on the safety limits. Therefore, this work is transformed to a nonlinear optimization problem with its objective function formulated as

$$\begin{aligned} &\min J = (x-\mu)^T \textstyle\sum^{-1}(x-\mu) \\ &s.t.\ g_r(x) = 0 \end{aligned} \quad (10)$$

Let x^* denote the solution of this problem, and δ can then be obtained by;

$$\delta = \sqrt{(x^*-\mu)^T \textstyle\sum^{-1}(x^*-\mu)} \quad (11)$$

Therefore, the probability in Eq. (8) can be calculated by the following method. First, the principal component analysis of the process variables is obtained after centering and scaling as;

$$t = P^T x \quad (12)$$

where t is the score vector and P is the loading matrix. So the process variable vector can then be reconstructed by $x = Pt$, and the expression of Mahalanobis distance can be rewritten as

$$d_M^2(x,\mu) = t^T(\Lambda)^{-1}t \quad (13)$$

where Λ is the covariance of$t$. Let $u = \Lambda^{-0.5}t$, then the Mahalanobis distance can be rewritten as

$$d_M^2(x,\mu) = u^T u = u_1^2 + u_2^2 + \ldots + u_m^2 \quad (14)$$

where $u_1, u_2, \ldots, u_m$ are the components of vector u. And these components are independent and follow the same standard Gaussian distribution. Therefore, the Mahalanobis distance follows a χ^2 distribution with m degrees of freedom. So the safety index in

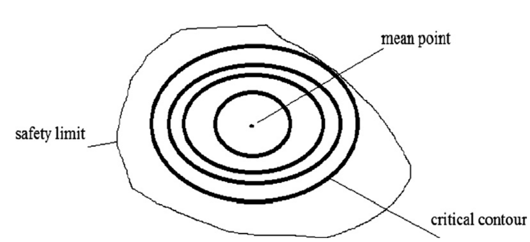

Fig. 1. Contours of probability density (solid lines) and safety limits.

Eq. (8) can be calculated as [15]

$$SI = \Pr(d_M(x,\mu) \le \delta) = \Pr(d_M^2(x,\mu) \le \delta^2) = \Pr(u^T u \le \delta^2) = \Pr(\chi^2(m) \le \delta^2) \tag{15}$$

3.2. Online safety indices

To obtain the online safety index, we now consider the issue that which kind of operating modes the samples come from, i.e. the posterior probabilities of an arbitrary monitored sample x coming from the running operating mode. Bayesian methods have a great potential to solve this problem by providing a probabilistic information synthesizing framework. Applications of the Bayesian method have been reported in medical science, image processing, target recognition, pattern matching, information retrieval, reliability analysis, and engineering diagnosis [21–24]. It is one of the most widely applied techniques in probabilistic inferencing.

Based on the previous work in Bayesian fault diagnosis, Huang [16] developed a data-driven Bayesian method for control loop diagnosis with consideration of sample dependency. The key idea is to introduce transition sample and the sample transition probability is estimated from the historical sample data. Inspired by this kind of Bayesian method, here we propose an approach to solve the problem in deciding which kind of operating modes the current sample comes from and calculating the SI online.

3.2.1. Data-driven Bayesian method [16]

Given current sample x_t, historical samples set D, the posterior probability of each possible operating mode can be calculated according to Bayes' rule as

$$P(M|x_t,D) = \frac{P(x_t|M,D)P(M|D)}{P(x_t|D)} \tag{16}$$

where $P(M|x_t,D)$ is the posterior probability; $P(x_t|M,D)$ is the likelihood probability; $P(M|D)$ is the prior probability of operating mode M; and $P(x_t|D)$ is a scaling factor. Note that historical data provide no information of prior probabilities of operating mode M, we have $P(M|D) = P(M)$. Therefore,

$$P(M|x_t,D) \propto P(x_t|M,D)P(M) \tag{17}$$

Among all the possible modes, the one with the largest posterior probability is supposed as the operating mode from which the samples come from according to the maximum a posteriori (MAP) principle, and the safety index associated with this operating mode is determined by

$$SI(x_t) = P(M_{max}|x_t,D) SI_{M_{max}} \tag{18}$$

Therefore, the main task for calculating the online safety index is equivalent to the estimation of the likelihood probability with historical samples set D, i.e. $P(x_t|M,D)$.

The likelihood probability can be computed by marginalization over all possible sample transition probability parameters as

$$\begin{aligned} P(x_t|M,D) &= \int_{\Psi_1,\dots,\Psi_K} p(x_t|\Phi_1,\dots,\Phi_K,M,D) \\ &\quad \times f(\Phi_1,\dots,\Phi_K|M,D) d\Phi_1 \dots \Phi_K \\ &= \int_{\Psi_1,\dots,\Psi_K} \phi_{t-1,t} \bullet f(\Phi_1,\dots,\Phi_K|M,D) d\Phi_1 \dots \Phi_K \end{aligned} \tag{19}$$

where $\Phi_i = \{\phi_{i,1}, \phi_{i,2}, \dots \phi_{i,K}\}$ represents the probability parameter set for all possible sample transitions from sample x_i to x_j where $j = 1, 2, \dots K$, under mode M, and K is the total number of possible sample values. From the definition of ϕ, it is clear that $\Sigma_{j=1}^K \phi_{i,j} = 1$. ψ_i is the space of all the probability parameter sets Φ_i.

The calculation of the term $f(\Phi_1,\dots,\Phi_K|M,D)$ is also tackled from a Bayesian perspective. Since it can be regarded as a posterior probability of parameter sets $\{\Phi_1,\dots,\Phi_K\}$, according to Bayes' rule,

$$f(\Phi_1,\dots,\Phi_K|M,D) = \frac{p(D|M,\Phi_1,\dots,\Phi_K) f(\Phi_1,\dots,\Phi_K|M)}{p(D|M)} \tag{20}$$

where

$$P(D|M) = \int_{\Psi_1,\dots,\Psi_K} p(D|\Phi_1,\dots,\Phi_K,M) \times f(\Phi_1,\dots,\Phi_K|M) d\Phi_1 \dots \Phi_K$$

Note that the first term in the numerator of the above equation, $p(D|\Phi_1,\dots,\Phi_K,M)$ is the likelihood of transition data set given parameter sets $\{\Phi_1,\dots,\Phi_K\}$, and is determined only by the operating mode and parameter sets $\{\Phi_1,\dots,\Phi_K\}$, we have

$$p(D|\Phi_1,\dots,\Phi_K,M) = \prod_{i=1}^K \prod_{j=1}^K \phi_{i,j}^{n_{i,j}} \tag{21}$$

where $n_{i,j}$ is the number of sample transitions from x_i to x_j.

The prior probability of transition parameter sets $\{\Phi_1,\dots,\Phi_K\}$ is solely determined by the current operating mode M. With the common assumption that the priors for different parameter sets are independent [22], we have

$$f(\Phi_1,\dots,\Phi_K|M) = f(\Phi_1|M) \cdots f(\Phi_K|M) \tag{22}$$

Dirichlet model is often used to model priors of the likelihood parameters [25], so here we have

$$f(\Phi_i|M) = \frac{\Gamma(\sum_{j=1}^K b_{i,j})}{\prod_{j=1}^K \Gamma(b_{i,j})} \prod_{j=1}^K \phi_{i,j}^{b_{i,j}-1} \tag{23}$$

where $b_{i,j}$ is the number of prior samples for sample transition from x_i to x_j. Γ is the gamma function.

From Eqs. (18)–(21), we have

$$f(\Phi_1,\dots,\Phi_K|M,D) = \frac{\prod_{i=1}^K \frac{\Gamma(\sum_{j=1}^K b_{i,j})}{\prod_{j=1}^K \Gamma(b_{i,j})} \prod_{j=1}^K \phi_{i,j}^{b_{i,j}-1} \prod_{i=1}^K \prod_{j=1}^K \phi_{i,j}^{n_{i,j}}}{p(D|M)} \tag{24}$$

So

$$\begin{aligned} P(x_t|M,D) &= \int_{\Psi_1,\dots,\Psi_K} \phi_{t-1,t} \frac{\prod_{i=1}^K \frac{\Gamma\left(\sum_{j=1}^K b_{i,j}\right)}{\prod_{j=1}^K \Gamma(b_{i,j})} \prod_{j=1}^K \phi_{i,j}^{b_{i,j}-1} \prod_{i=1}^K \prod_{j=1}^K \phi_{i,j}^{n_{i,j}}}{p(D|M)} d\Phi_1 \dots \Phi_K \\ &= \frac{\prod_{i=1}^K \frac{\Gamma\left(\sum_{j=1}^K b_{i,j}\right)}{\prod_{j=1}^K \Gamma(b_{i,j})}}{p(D|M)} \int_{\Psi_1} \prod_{j=1}^K \phi_{1,j}^{b_{1,j}-1+n_{1,j}} d\Phi_1 \\ &\quad \cdots \int_{\Psi_{t-1}} \phi_{t-1,t}^{n_{t-1,t}+b_{t-1,1}} \prod_{j \ne t} \phi_{t-1,j}^{n_{t-1,j}+b_{t-1,j}-1} d\Phi_{t-1} \cdots \int_{\Psi_K} \prod_{j=1}^K \phi_{K,j}^{b_{K,j}-1+n_{K,j}} d\Phi_K \end{aligned} \tag{25}$$

From Dirichlet Integrals [26], we have

$$\begin{aligned} &\int_{\Psi_1} \prod_{j=1}^K \phi_{1,j}^{b_{1,j}-1+n_{1,j}} d\Phi_1 \cdots \int_{\Psi_{t-1}} \phi_{t-1,t}^{n_{t-1,t}+b_{t-1,1}} \prod_{j \ne t} \phi_{t-1,j}^{n_{t-1,j}+b_{t-1,j}-1} d\Phi_{t-1} \\ &\quad \cdots \int_{\Psi_K} \prod_{j=1}^K \phi_{K,j}^{b_{K,j}-1+n_{K,j}} d\Phi_K \\ &\quad = \frac{\Gamma(n_{t-1,t} + b_{t-1,t} + 1)}{\Gamma(N_{t-1} + B_{t-1} + 1)} \frac{\prod_{i,j \ne t-1,t} \Gamma(n_{i,j} + b_{i,j})}{\prod_{i \ne t-1} \Gamma(N_i + B_i)} \end{aligned} \tag{26}$$

where $N_i = \sum_j n_{i,j}$ is the total number of transition samples, $B_i = \sum_j b_{i,j}$ is the corresponding total number of prior transition samples under mode M.

From Eqs. (18)–(21), we calculate

$$p(D|M) = \prod_{i=1}^K \frac{\Gamma\left(\sum_{j=1}^K b_{i,j}\right)}{\prod_{j=1}^K \Gamma(b_{i,j})} \prod_{i=1}^K \frac{\prod_{j=1}^K \Gamma(n_{i,j} + b_{i,j})}{\Gamma(N_i + B_i)} \tag{27}$$

So

$$\begin{aligned} &p(x_t|M,D) \\ &\quad = \prod_{i=1}^K \frac{(\Gamma \sum_{j=1}^K b_{i,j})}{\prod_{j=1}^K \Gamma(b_{i,j})} \frac{\Gamma(n_{t-1,t} + b_{t-1,t} + 1)}{\Gamma(N_{t-1} + B_{t-1} + 1)} \frac{\prod_{i,j \ne t-1,t} \Gamma(n_{i,j} + b_{i,j})}{\prod_{i \ne t-1} \Gamma(N_i + B_i)} \end{aligned}$$

$$\frac{\prod_{i=1}^{K}\Gamma(N_i+B_i)}{\prod_{i=1}^{K}\frac{(\Gamma\sum_{j=1}^{k}b_{i,j})}{\prod_{j=1}^{K}\Gamma(b_{i,j})}\prod_{i=1}^{K}\prod_{j=1}^{K}\Gamma(n_{i,j}+b_{i,j})}$$
$$=\frac{n_{t-1,t}+b_{t-1,t}}{N_{t-1}+B_{t-1}} \quad (28)$$

We observe that the likelihood probability is determined by both prior samples and historical samples. And the prior and historical samples refer to the amount of the transition samples under the current operating mode. As the number of historical transition samples increases, the likelihood probability will converge to the relative frequency determined by the historical samples. Compared with the conventional Bayesian method, our method reduces the risk of misclassification and the subsequent biased safety assessment results.

Thus, from Eqs. (15), (16) and (26), we have

$$SI(x_t)\propto P(x_t|M,D)P(M)SI_M=\frac{n_{t-1,t}+b_{t-1,t}}{N_{t-1}+B_{t-1}}P(M)SI_M \quad (29)$$

According to the above proposed method, the schematic diagram of operating safety assessment is shown in Fig. 2.

4. Example

4.1. Aircraft gas turbine engine

The first example was developed by Abhinav Saxena and Kai Goebel [27], based on a model of aircraft gas turbine engines, which has been widely used for the test of process control design, monitoring, optimization, etc. The engine diagram and modules connections are shown in Fig. 3. There are 5 rotating components in the engine (Fan, Low-Pressure Compressor (LPC), High-Pressure Compressor (HPC), High-Pressure Turbine (HPT), and Low-Pressure Turbine (LPT)). And the atmosphere of engine operation includes (a) altitudes ranging from sea level to 40,000 ft, (b) Mach numbers from 0–0.90, and (c) sea-level temperatures from −60–103 °F. The system outputs include various sensors corresponding to surfaces and operating margins and have 21 variables. For the purpose of this paper, we here select 6 variables from the 21 different variables. All the data are contaminated with sensor noise. There are six operating modes in the system, and under each operating mode there are at least 30 samples collected as the historical data set. The selected variables are shown in Table 1.

First, in the off-line part, we use GMM to characterize the operating conditions of each individual operating mode. To ensure that the historical data can be described by GMM, the normality testing is applied. Although it is not expected that the data would exactly match the Gauss distribution, it is expected that the distribution of historical data should be similar to the Gauss distribution. The results are shown in Fig. 4, from which we can observe that all the 6 variables in each mode are in concordance with the Gaussian distribution. Therefore, GMM is fit to characterize the data distribution here. Afterward, for each component of the GMM, i.e. each operating mode of the engineering system, we use EM algorithm to estimate the parameters.

After the estimation of parameters, we can calculate SI for each operating mode. And the online SI is calculated according to Eq. (28). We here use 20 groups of sample data as newly sampled data to show how to calculate the SI online. Since we have already got 200 groups of sample data as the historical data, for the newly sampled data-group 201, we should first know which operating mode it belongs to. According to Eq. (27), if the current operating mode is mode 1, then the total number of transition samples until group 201 is 36, the corresponding total number of prior transition samples is 26, the number of prior samples for sample transition from x_{200} to x_{201} is 9, and the number of sample transitions from x_{200} to x_{201} is 8. The likelihood probability $P(x_t|M,D)$ is 0.76. If the current operating mode is other mode, we can similarly calculate the likelihood probability. After this calculation, we compare the results and choose the one with the largest posterior probability as the current operating mode from which the newly sampled data comes from. Then we can calculate the Mahalanobis distance

Table 1
Variables selected to calculate safety index.

Variable	Variable description	Units
1	Total temperature at Fan inlet	°R
2	Total temperature at LPC outlet	°R
3	Total temperature at HPC outlet	°R
4	Total temperature at LPT outlet	°R
5	Total temperature at HPT outlet	°R
6	Total pressure at HPC outlet	psia

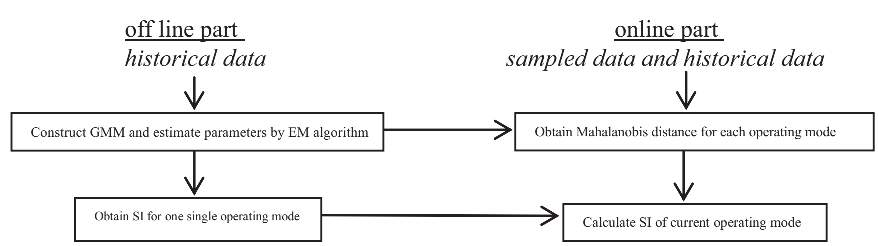

Fig. 2. Schematic diagram of the proposed framework for operational safety assessment.

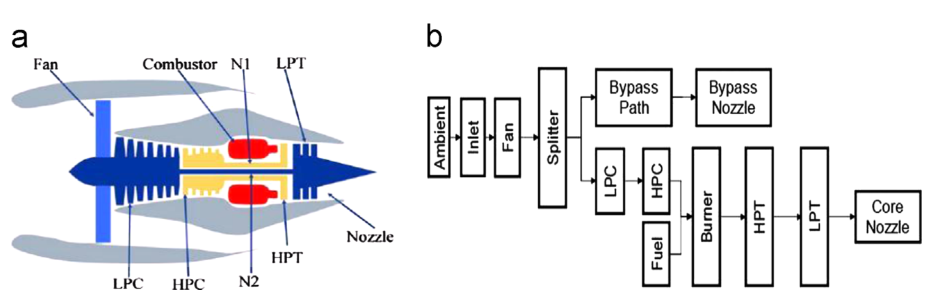

Fig. 3. Simplified diagram of engine and modules connections [28].

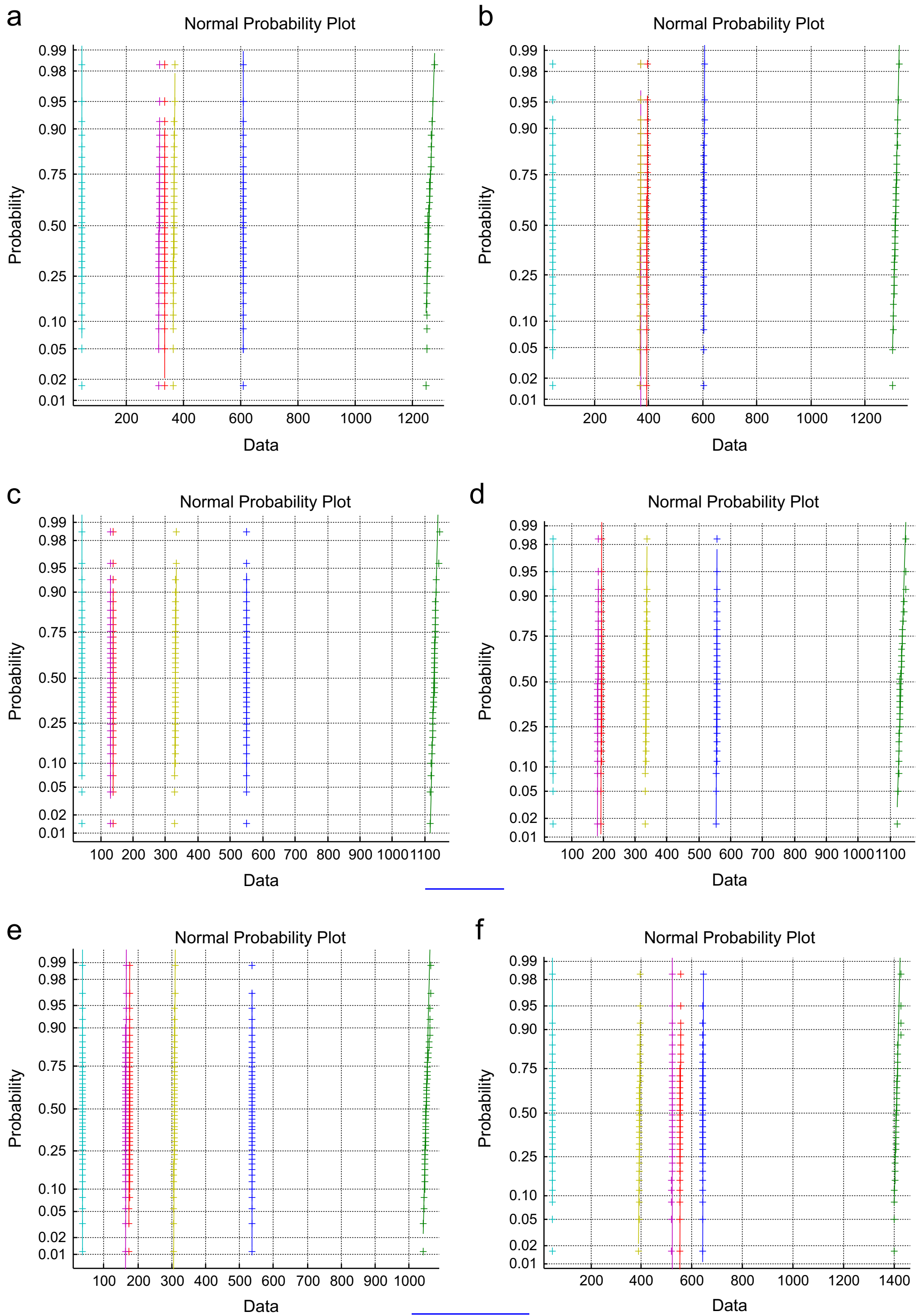

Fig. 4. Normality testing results. (a) Mode 1, (b) Mode 2, (c) Mode 3, (d) Mode 4, (e) Mode 5 and (f) Mode 6.

between the newly sampled data and the mean points of current operating mode. According to Eq. (15), the Mahalanobis distance is 1.6. Finally, we obtain the online SI for the current operating mode is 0.96.

Likewise, we can calculate the online SI from sample data-group 202 to 220. For the sake of operational safety assessment, some constraints for process variables are provided here as the safety limits. The safety limits are determined via operability margins—like stall and temperature limits. The curve of SI change and the safety limits for operating modes are shown in Fig. 5.

From Fig. 5 we can see that nearly all the sample data from the engineering system under consideration are beyond the safety limits except the sample data 213. The SI of sample data 213 is under the safety limit of operating mode 1, but the mode which sample data 213 comes from is mode 3 according to the calculation of posterior probability. So the system is safe under current operating mode.

Compared with the conventional Bayesian method in which the sample dependency is simply ignored, the proposed method significantly reduces the risk of wrong judgment for operating mode. Here we use 1000 groups of sampled data to show the performance of two different methods. The performance results are illustrated in Fig. 6.

It can be seen that there are some differences between the above 2 figures. Owing to the ignorance of sample dependency, the conventional Bayesian method has some wrong judgments of operating mode, its overall correct judgment rate is only 51.45%, and is much lower in comparison to the rate of the proposed method, which is 73.86%.

Due to the difference between two Bayesian methods, namely, considering and ignoring the sample dependency, the online operational safety assessment results in Fig. 7 are very different.

In Fig. 7 we can see that the SI of sample data 351 is under the safety limit of operating mode 1 and 2 according to the conventional Bayesian method, and since the judgment of operating mode is mode 2, the system is in dangerous condition. But in fact, the current operating mode is mode 3, so the system is actually in safe condition. The Bayesian method with considering of the sample dependency provides right information. We can also observe that the change of SI can be caused either by some turbulence or the change of operating mode.

4.2. Industrial dryer

The second example concerns with an industrial drying process which has 6 process variables. The data is contributed by Jan Maciejowski [29] from engineering department of Cambridge University, and all the data for this example consists of measurements taken on a real process. The process inputs include the fuel flow rate, the hot gas exhaust fan speed, and the rate of flow of raw material. The outputs include dry bulb temperature, wet bulb temperature, and the moisture content of the raw material when it leaves the dryer. There are 4 operating modes in the process, and under each operating mode there are at least 60 samples collected with sampling period 10 sec. The process variables are shown in Table 2.

In the off-line part, we use the first 240 samples to obtain GMM of operating modes. The rest of the sample data is used for online safety assessment. After that, we use EM algorithm to estimate the parameters of each operating mode of the engineering system.

To compare with the conventional Bayesian method, we here use 256 groups of sampled data to show the performance. The performance results of two different methods are shown in Fig. 8.

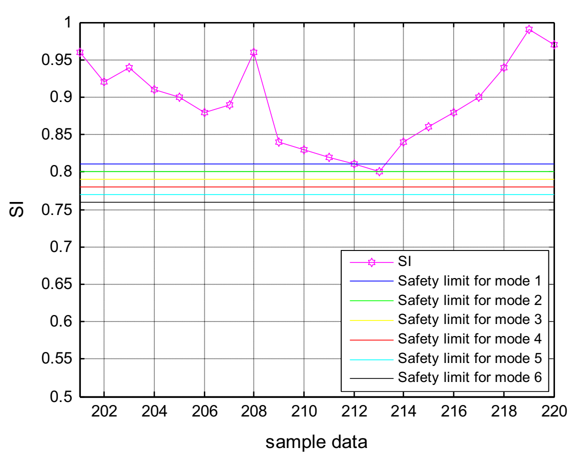

Fig. 5. Online SI for the system from the sample data group 201 to 220 and the safety limit for operating modes.

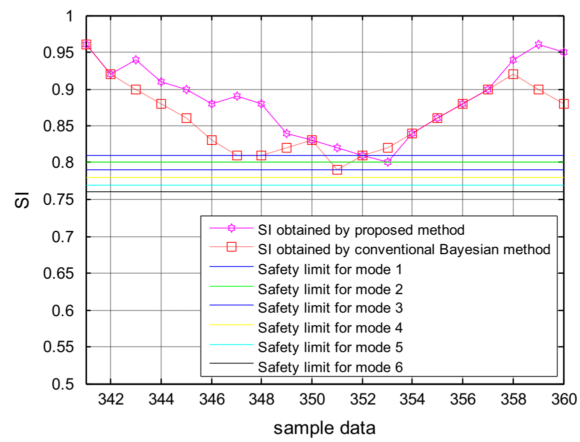

Fig. 7. Online SI obtained by two methods.

Table 2
Process variables of an industrial drying process.

Variable	Variable description	Units
1	The fuel flow rate	Kg/h
2	The hot gas exhaust fan speed	rpm
3	The rate of flow of raw material	Kg/h
4	Dry bulb temperature	°C
5	Wet bulb temperature	°C
6	The moisture content of the raw material	%

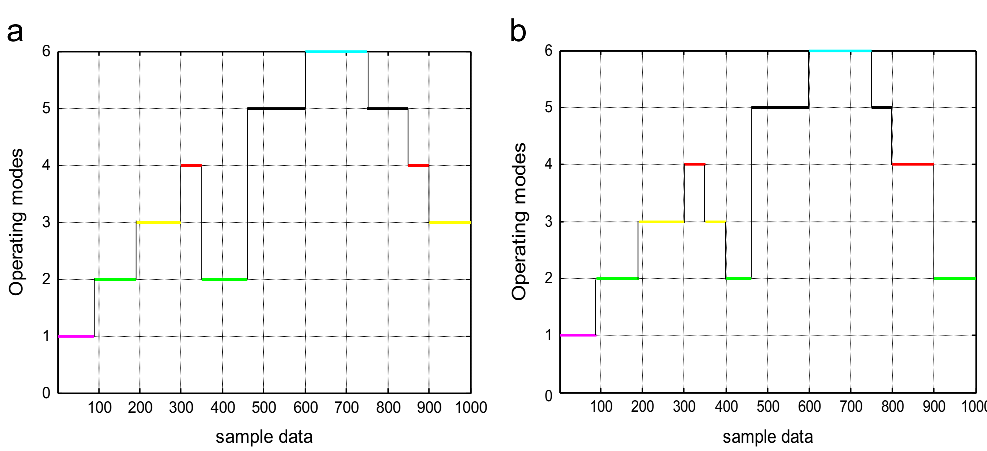

Fig. 6. Online judgment of operating mode. (a) Conventional Bayesian method and (b) Proposed Bayesian method.

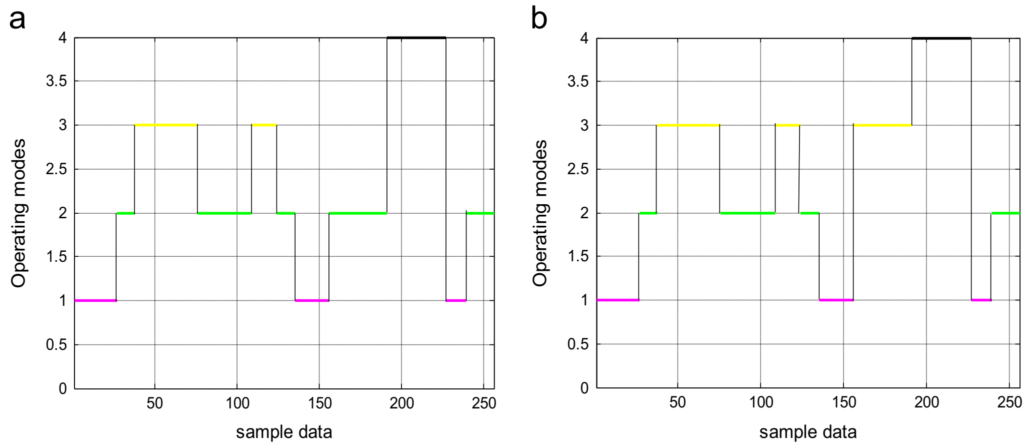

Fig. 8. Online judgment of operating mode. (a) Conventional Bayesian method (b) Proposed Bayesian method.

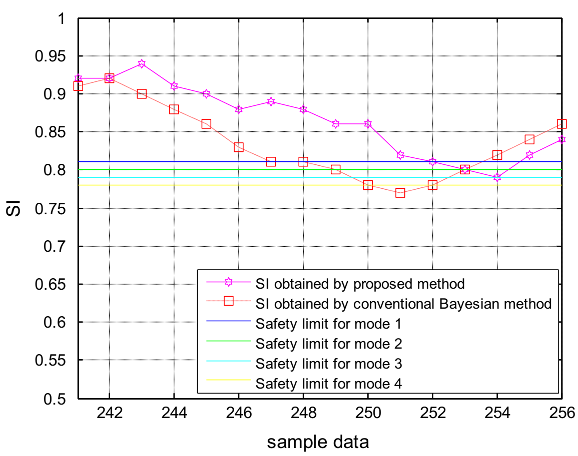

Fig. 9. Online SI obtained by two methods.

It can be seen from Fig. 8 that the process is first running at operating mode 1 and then climbs to mode 2. Till sample data 160 there is nearly no difference between two methods. But as the number of historical transition samples increases, the proposed Bayesian method becomes more and more accurate in judgment of operating mode. The overall correct judgment rate of conventional Bayesian method is only 61.75%, and is much lower than the correct judgment rate of the proposed method, which is 83.66%.

After the estimation of parameters, we calculate SI for each operating mode. The online operational safety assessment results are shown in Fig. 9.

From Fig. 9 we notice that on sample 247, the SI of the process first touches the safety limit for mode 1 according to the conventional Bayesian method. But since the process is running at operating mode 2, it is still safe on sample 247. Then on sample 250, the SI of the process is beyond the safety limit for mode 2 according to the conventional Bayesian method, which means that the process is in dangerous condition. But in fact the process variable on sample 250 is within the safety limit for mode 2. The reason why the SI calculated by the conventional Bayesian method is inaccurate is that the likelihood probability obtained by the conventional Bayesian method is too small. And this relates to the setting of initial value in the conventional Bayesian method. We can also see that the SI calculated by the proposed method is in accordance with the fact. Then on sample 254, the SI obtained by the proposed method exceeds the safety limit for mode 2, which means that the process is in dangerous condition now. The fact is that the process variable on sample 254 is outside the safety limit for mode 2, so the assessment of the proposed method is accurate compared to the conventional Bayesian method in this sense.

5. Conclusions

For multi-mode engineering systems with sampling dependency, a Bayesian method for online assessment of operational safety is proposed. Gaussian mixture model is constructed on the basis of historical data with the parameters estimated by EM algorithm. The operational safety assessments are formulated on the basis of safety index, which has been developed by Ye et al.

The key aspect of the proposed approach is the ability to online evaluate the SI by capturing all information that can be possibly sampled by engineering systems. Conventional Bayesian methods commonly ignore the dependencies between the temporal sampled data and historical data, which do not hold for practical multi-mode engineering systems. With the consideration of sample transition probability, the temporal sampled data and historical data can be synthesized by the Bayesian method, and therefore allows the posterior probability of each possible operating mode to be correctly calculated and obtain the SI of the current operating mode.

The proposed Bayesian method is applied in two examples, from which we can conclude that the proposed method based on safety index is able to give an efficient assessment of operational safety for an engineering system or an industrial process.

Acknowledgments

This work was partially supported by the national 973 project under Grants 2010CB731800 and 2009CB32602, and the NSFC under Grants 61210012, 61021063, 61290324.

References

[1] Crowl DA, Louvar JF. Chemical Process Safety: Fundamentals with Applications. 2nd edition. Upper Saddle River, NJ: Prentice Hall; 2002.

[2] Venkatasubramanian V, Zhao J, Viswanathan S. Intelligent systems for HAZOP analysis of complex process plants. Computers and Chemical Engineering 2000;24:2291–302.

[3] Rouvroye JL, Van den Bliek. Comparing safety analysis techniques. Reliability Engineering and System Safety 2002;75:289–94.

[4] Khan FI, Husain T, Abbasi SA. Design and evaluation of safety measures using a newly proposed methodology SCAP. Journal of Loss Prevention in the Process Industries 2002;15:129–46.

[5] Arena U, Romeo E, Mastellone ML. Recursive operability analysis of a pilot plant gasifier. Journal of Loss Prevention in the Process Industries 2008;21:50–65.

[6] Durga Raoa K, Gopikaa V, Sanyasi Raoa VVS, Kushwahaa HS, Vermab AK, Srividya A. Dynamic fault tree analysis using Monte Carlo simulation in probabilistic safety assessment. Reliability Engineering and System Safety 2009(94):872–83.
[7] Knegteringa B, Brombacher AC. Application of micro Markov models for quantitative safety assessment to determine safety integrity levels as defined by the IEC 61508 standard for functional safety. Reliability Engineering and System Safety 1999;66:171–5.
[8] Lisnianski A, Levitin G. Multi-state system reliability: assessment, optimization, and application. Singapore: World Scientific; 2003.
[9] Meel A, Seider WD. Plant-specific dynamic failure assessment using Bayesian theory. Chemical Engineering Science 2006;61:7036–56.
[10] Meel A, Seider WD. Real-time risk analysis of safety systems. Computers and Chemical Engineering 2008;32:827–40.
[11] Khakzada Nima, Khana Faisal, Amyotte Paul. Safety analysis in process facilities: comparison of fault tree and Bayesian network approaches. Reliability Engineering and System Safety 2011;96:925–32.
[12] Jiang, H, Gao, J, Gao, Z, Li, G. In safety analysis of process industry system based on complex networks theory. In: Proceedings of the international conference on mechatronics and automation, Harbin, China: IEEE; 2007., p. 480.
[13] Zhou J, Chen G, Chen Q. Real-time data-based risk assessment for hazard installations storing flammable gas. Process Safety Progress 2008;27:205–11.
[14] Podofillinia L, Ziob E, Mercurioa D, Dang VN. Dynamic safety assessment: scenario identification via a possibilistic clustering approach. Reliability Engineering and System Safety 2010;95:534–49.
[15] Lubin Ye, et al. Online probabilistic assessment of operating performance based on safety and optimality indices for multimode industrial processes. Industry and Engineering Chemistry Research 2009;48(24):10912–23.
[16] Qi Fei, Huang Biao. Bayesian methods for control loop diagnosis in the presence of temporal dependent evidences. Automatica 2011;47:1349–56.
[17] Choi SW, Park JH, Lee I. Process monitoring using a Gaussian mixture model via principal component analysis and discriminant analysis. Computers and Chemical Engineering 2004;28:1377–87.
[18] Bishop CM. Neural networks for pattern recognition. New York: Oxford University Press; 1995.
[19] Paalanen P, Kamarainen JK, Ilonen J, Kalviainen H. Feature representation and discrimination based on Gaussian mixture model probability densities practices and algorithms. Pattern Recognition 2006;39:1346.
[20] Yu J, Qin SJ. Multimode process monitoring with Bayesian inference-based finite Gaussian mixture models. AIChE Journal 2008;54:1811.
[21] Chien CF, Chen SL, Lin YS. Using Bayesian network for fault location on distribution feeder. IEEE Transactions on Power Delivery 2002;17(13):785–93.
[22] Dey S, Stori JA. A Bayesian network approach to root cause diagnosis of process variations. International Journal of Machine Tools 2005;45(1):75–91.
[23] Mehranbod N, Soroush M, Panjapornpon C. A method of sensor fault detection and identification. Journal of Process Control 2005;15(3):321–39.
[24] Steinder M, Sethi AS. Probabilistic fault localization in communication systems using belief networks. IEEE/ACM Transactions on Networking 2004;12 (5):809–22.
[25] Pernestal, A.A. Bayesian approach to fault isolation with application to diesel engine diagnose [Ph.D. thesis]. KTH School of Electrical Engineering; 2007.
[26] DeGroot Morris H. Optimal statistical decisions. New York: McGraw-Hill; 1970.
[27] Abhinav Saxena, Kai Goebel. "C-MAPSS Data Set", NASA Ames Prognostics Data Repository, [⟨http://ti.arc.nasa.gov/project/prognostic-data-repository⟩]. NASA Ames, Moffett Field, CA, 2008.
[28] D. Frederick, J. DeCastro, J. Litt User's guide for the commercial modular aero-propulsion system simulation (CMAPSS). NASA/ARL, Technical Manual TM 2007-215026, 2007.
[29] Chou CT, Maciejowski JM. System identification using balanced parametrizations. IEEE Transactions on Automatic Control 1997;42:956–74.