Preprints of the 21st IFAC World Congress (Virtual)
Berlin, Germany, July 12-17, 2020

Bayesian-based anomaly detection in the industrial processes ⋆

Yijun Pan, Zeyu Zheng, Dianzheng Fu

Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, 110016, Liaoning, China
Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang, 110016, Liaoning, China
Key Laboratory of Network Control System, Chinese Academy of Sciences, Shenyang, 110016, Liaoning, China
(e-mail: zhengzeyu@sia.cn).

Abstract: In general, the industrial processes are semi-automatic, and are controlled by the operators. Since the operation principles of the industrial processes are complicated, it is difficult to label observations. The disturbances may be contained in the observations. Therefore, the unsupervised anomaly detection method is promising for research in the industrial processes. In the paper, a multivariate anomaly detection method is proposed, which is unsupervised and online. The priori probability of anomaly occurrence is necessary, and a hazard function selection method is defined at first. Secondly, Bayesian-based method is adopted for anomaly detection. In final, the Dempster-Shafer theory is introduced for fusing the univariate anomaly detection results. The numerical simulation is used for illustrating the anomaly detection power of the proposed method, and the TE process is implemented for testing the fault detection effectiveness. A real data set collected from a bathyscaphe is applied for demonstrating the power of leakage detection.

Keywords: Anomaly detection; Bayesian; Dempster-Shafer theory; Bathyscaphe; TE process; Change point detection; Leakage detection

1. INTRODUCTION

In the industrial processes, the operating principles are more and more complicated, and there are many uncontrollable factors (Zhou et al. (2018)). With the development of distributed control system, a large amount of unlabeled observations are collected. The labels contain normal data, various fault types data. Labeling the observations requires abundant expert knowledge, which is time-consuming and expensive (Ge et al. (2013)). Moreover, the industrial processes are semi-automatic in general. The guidance from experts and operators are various. The normal condition observations may contain some disturbances, which will lead to poor anomaly detection results based on these observations. Therefore, it is promising for research the unsupervised anomaly detection methods. The unsupervised methods could ignore the influences of the operator habits, and only detect the changes of observation distributions.

The anomaly detection is widely researched in decades (Pan et al. (2018), An et al. (2015)). A novel distributed process monitoring framework for quality-related fault detection based on distributed modified principal component regression was proposed by Rong et al. (Rong et al. (2019)). Wang et al. proposed a novel three-stage intelligent fault diagnosis approach for practical industrial process monitoring (Wang et al. (2019)). Chen et al. proposed an unsupervised change point detection based on the graph method (Chen et al. (2015)). An et al. proposed a weight graph-based change point detection method for fault detection in a blast furnace process (An et al. (2019)). Jove et al. proposed a method including a visual tool for the detection of faults, and its final aim is to optimize system performance and consequently obtain increased economic savings, in terms of energy, material, and maintenance (Jove et al. (2019)).

Recently, the Bayesian change point detection method was proposed by Adams et al., which is an online and unsupervised method (Adams et al. (2007)). The change point detection methods are widely used in process control, river flow analysis and DNA sequence prediction (Rosenbaum (2005)). Aminilhanghahi et al. reviewed many of the methods that have been proposed to detect change points in time series (Aminikhanghahi et al. (2017)). Most of the change point detection methods are off-line (Eriksson et al. (2019)), and some online change point detection methods have been researched (Aminikhanghahi et al. (2019),Wilson et al. (2010)).

Compared with the above methods, the Bayesian online change point detection method calculates the posterior distribution of the current run length since the last change point occurs, and a recursive message-passing method is used for computing the joint distribution. As far as the authors know, the Bayesian online change point detection

⋆ This work was supported by the National Key Research and Development Project (Grant No. 2018YFF0214704).

method is barely used for anomaly detection in the industrial processes. The online Bayesian change point detection method needs to calculate the predictive distribution, and it requires the distributions of observations belong to the exponential family. The bernoulli, the gamma and the univariate Gaussian are some members of a broader class of distributions known as the exponential family (Murphy (2012)).

In the paper, a multivariate anomaly detection method based on Bayesian is proposed. The prior probability of the anomaly occurrence is defined based on the false alarm rate. The graph-based off-line change point detection method is used for obtaining some observations with the same distributions at first (Chen et al. (2015)). Half of the observations are used as a training matrix constructing the PCA model, and the others are used as a testing matrix to calculate the false alarm rate. The false alarm rate is defined as a prior probability $P_{anomaly}$. The online Bayesian change point detection method is applied for detecting anomaly event. The posterior probability of run length is calculated by the observed data. Since the TE process and the bathyscaphe operation in the paper could be regarded as the Gaussian processes approximately, the conjugate prior of the univariate Gaussian is used for calculating the predictive distribution. The Dempster-Shafer (D-S) theory will be adopted for fusing the anomaly detection results of single variable.

The organization of this paper is described as follows. Section 2 is the problem formation. The Bayesian online change point detection method is introduced in Section 3. The steps for anomaly detection are summarized in Section 4. Section 5 is the simulations. Finally, a conclusion is presented in Section 6.

2. PROBLEM FORMULATION

Suppose that the observations $x_i, i = 1, 2, ...$ are collected by some sensors. The purpose of the proposed method in this paper is to detect whether an anomaly occurs in the observations. The null hypothesis H_0 is formulated as follows.

$$H_0 : x_i \sim F_0, i = 1, 2, ... \tag{1}$$

where F_0 is the distribution of observations.

The alternative hypothesis H_1 is

$$H_1 : x_i \sim \begin{cases} F_0, i = 1, 2, ...t \\ F_1, i = t+1, t+2, ... \end{cases} \tag{2}$$

where t is the time of anomaly occurrence. F_0 and F_1 are two different distributions of observations. The different distribution of observations indicates a change in the time series, and the anomaly occurs. In this part, only the situation that one change point occurs is listed. The online Bayesian method can also handle the multi-change points with different data distributions $F_0, F_1, F_2, ...$

3. BAYESIAN ONLINE CHANGE POINT DETECTION

The unsupervised Bayesian online change point detection method needs to calculate the predictive distribution, which requires the observation distributions belong to the exponential family. In the section, the univariate Gaussian data is used for introducing the principal of the method. The observations $x_i, i = 1, 2, ..., t$ are used for estimating the operation condition at current sampling point t. Some symbols are defined at first. $x_{a:b}$ indicates the observations between sampling point a and b. r_t represents the run length since the last change point occurs, and the observations collected in this period are defined as $x_t^{(r)}$. The discrete a priori probability of anomaly occurrence is $P_{anomaly}$. Initially, the $r_1 = 0$, and there is only x_1 in the $x_1^{(r)}$. The purpose of the Bayesian change point detection method is to calculate the posterior distribution $P(r_t|x_{1:t})$. The above introductions is illustrated in Fig.1.

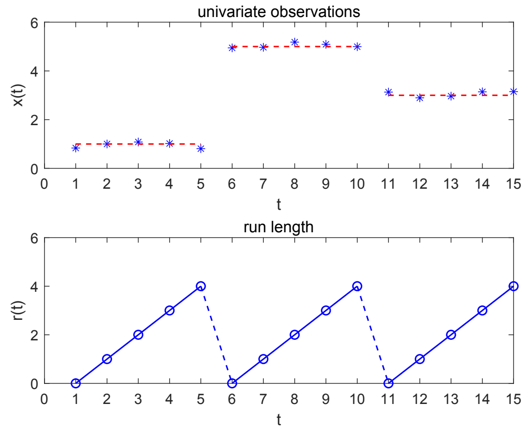

Fig. 1. This is a simple example to illustrate the Bayesian online change point detection method. There are two change points included in the univariate Gaussian observations. The run length adds to 1 if there is no change point, and otherwise, the run length drops to 0. The goal is to calculate the posterior distribution of the run length based on the observations, $P(r_t|x_{1:t})$.

According to the Bayesian inference, the posterior distribution can be calculated as follows.

$$P(r_t|x_{1:t}) = \frac{P(r_t, x_{1:t})}{P(x_{1:t})} \tag{3}$$

where

$$\begin{aligned} P(r_t, x_{1:t}) &= \sum_{r_{t-1}} P(r_t, r_{t-1}, x_{1:t}) \\ &= \sum_{r_{t-1}} P(r_t, x_t|r_{t-1}, x_{1:t-1})P(r_{t-1}, x_{1:t-1}) \\ &= \sum_{r_{t-1}} P(r_t|r_{t-1})P(x_t|r_{t-1}, x_t^{(r)})P(r_{t-1}, x_{1:t-1}) \end{aligned} \tag{4}$$

The joint distribution $P(r_t, x_{1:t})$ shown in the formula (4) is solved by a recursive algorithm. The conditional prior $P(r_t|r_{t-1})$ is calculated by a hazard function. The hazard function is a conditional probability that a process will occur fault during the time t and dt under the condition that the process is normal until time t. In the paper, the proposed method is used for detecting the faults in the TE process and the leakage detection in the bathyscaphe. Since these two processes are memoryless, the hazard function is set to be a constant in the paper, and is calculated on basis of the probability of anomaly occurrence. The $P(x_t|r_{t-1}, x_t^{(r)})$ is a predictive distribution as a new observation arrives, and is calculated by the conjugate prior of the exponential family likelihoods.

4. ANOMALY DETECTION BASED ON THE BAYESIAN METHOD

The Bayesian online change point detection method needs to calculate a predictive distribution, and the distribution of the exponential family is the only one with finite-sized sufficient statistics. For realizing the online change point detection, the observations should follow the distributions of the exponential family. Since the TE process and the bathyscaphe operation could be regarded as the Gaussian processes approximately, the conjugate prior of the univariate Gaussian is used for calculating the predictive distribution. The D-S theory is used for fusing the change point detection results of single variable (Dempster (1967)). For convenience, an observation with m variables is presented for anomaly detection based on the proposed method in the section. If there are n testing observations, repeat the following steps for n times to get the anomaly detection result at each sampling point. $x_i, i = 1, 2, ...$ is the univariate observations, and $y_i, i = 1, 2, ..., m$ is the anomaly detection results of m variables. The Bayesian method is used for detecting the change points for each variable at first, and then the D-S theory fuses the detection results. Since the D-S theory is a common fusion method, the theory is ignored in the paper. The steps for anomaly detection are described.

Step 1: Parameter initialization

For convenience, the testing observations start with a change point, and the run length is zero, that is $P(r_1 = 0) = 1$. The hyperparameters ν, χ are used for calculating $P(x_t|r_{t-1}, x_t^{(r)})$, and are set based on the actual data characteristics, i.e. mean and variance. The prior distribution $P_{anomaly}$ of the anomaly occurrence is selected based on the false alarm rate. Firstly, the graph-based off-line change point detection method is used for obtaining some observations with the same distributions. Half of the observations are regarded as a training matrix constructing the PCA model, and the others are used as a testing matrix for calculating the false alarm rate. The false alarm rate is defined as a prior distribution $P_{anomaly}$.

$$\begin{aligned} &P(r_1 = 0) = 1 \\ &\nu_1^{(1)} = \nu_{prior} \\ &\chi_1^{(1)} = \chi_{prior} \end{aligned} \tag{5}$$

Step 2: Predictive probability calculation

To realize the online anomaly detection, the distribution of the testing observations should belong to the exponential family. Exponential family likelihoods have the following form

$$P(x|\eta) = h(x)exp(\eta^T U(x) - A(\eta)) \tag{6}$$

where

$$A(\eta) = \log \int d\eta h(x) exp(\eta^T U(x)) \tag{7}$$

The predictive distribution $P(x_t|r_{t-1}, x_t^{(r)})$ denoted as $\pi_t^{(r)}$ is described as follows. The conjugate-exponential represents that the prior takes the same form with posterior of an exponential-family distribution over η.

$$\begin{aligned} &\pi_t^{(r)} = P(x_t|\nu_t^{(r)}, \chi_t^{(r)}) \\ &P(\eta|\chi, \nu) = \tilde{h}(\eta) exp(\eta^T \chi - \nu A(\eta) - \tilde{A}(\chi, \nu)) \end{aligned} \tag{8}$$

where χ, ν are two hyperparameters.

Step 3: Growth probability calculation

On basis of the formula (4), the growth probability of normal condition can be calculated as follows.

$$P(r_t = r_{t-1} + 1, x_{1:t}) = P(r_{t-1}, x_{1:t-1}) \pi_t^{(r)} (1 - P_{anomaly}) \tag{9}$$

Step 4: Change point occurrence probability calculation

On basis of the formula (4), the change point occurrence probability can be calculated as follows.

$$P(r_t = 0, x_{1:t}) = \sum_{r_{t-1}} P(r_{t-1}, x_{1:t-1}) \pi_t^{(r)} P_{anomaly} \tag{10}$$

Step 5: Run length distribution calculation

The run length distribution at current sampling point is calculated.

$$P(r_t|x_{1:t}) = \frac{P(r_t, x_{1:t})}{P(x_{1:t})} \tag{11}$$

where

$$P(x_{1:t}) = \sum_{r_t} P(r_t, x_{1:t}). \tag{12}$$

Step 6: Parameters update

The iterative algorithm is used for calculating the joint distribution, and the parameters should update as follows.

$$\begin{aligned} &\nu_{t+1}^{(1)} = \nu_{prior} \\ &\chi_{t+1}^{(1)} = \chi_{prior} \\ &\nu_{t+1}^{(r+1)} = \nu_t^{(r)} + 1 \\ &\chi_{t+1}^{(r+1)} = \chi_t^{(r)} + u(x_t) \end{aligned} \tag{13}$$

where $u(x_t)$ derives from the sufficient statistics of the exponential family likelihoods.

Step 7: Repeat

Return to Step 1 for m times, and m is the number of the variables. The run length distributions of variables are collected in $y_i, i = 1, 2, ..., m$.

Step 8: Result fusion

The run length distributions of variables are fused based on the D-S theory, and a integrated run length distribution for m variables is calculated. The run length probability distribution of two variables are fused at first, and then is fused with the other variable by the D-S theory.

5. SIMULATIONS

In the section, a simple Gaussian data is used for illustrating the power of the proposed method at first. Moreover, the method is adopted for fault detection in the TE process, and the goal is to test the process monitoring effectiveness. In final, the proposed method is implemented for leakage detection in a bathyscaphe.

5.1 Numerical simulation

In the section, a simple Gaussian example is applied for illustrating the power of the proposed anomaly detection

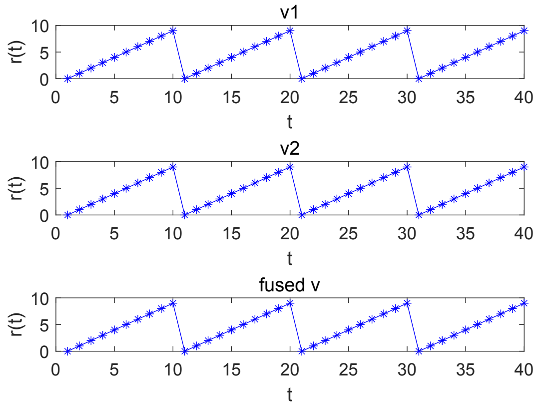

Fig. 2. This is a two simulated variables Gaussian example to illustrate the power of the proposed anomaly detection method. There are three change points at 11st, 21st, 31st sampling point, and the observations are constructed with different means. Based on the variable 1, variable 2 and fused variables, the anomaly could be detected exactly.

method. The testing data consist of 40 observations, and follow the distributions $N_1 \sim (0, 0.5 * I_2)$, $N_2 \sim (5, 0.5 * I_2)$, $N_3 \sim (0, 0.5 * I_2)$ and $N_4 \sim (3, 0.5 * I_2)$ respectively. I_2 is a identity matrix with 2 dimensions. There are three change points in 40 observations at 11st sampling point, 21st sampling point and 31st sampling point. The prior probability of the anomaly occurrence in the Gaussian observations is $P_{anomaly} = 0.05$, and it is chosen according to the confidence $1 - \alpha = 0.95$. The simulation result is shown in Fig.2.

From Fig.2, it can be concluded that the proposed method is powerful for detecting the anomaly. The selection of prior probability is appropriate and the D-S theory is suitable for fusing the detection results.

5.2 TE process

The TE process is constructed by the Tenessee Eastman company, and it aims at evaluating the process monitoring methods. The TE process is derived from a real industrial process, and the process flow diagram is presented as Fig.3. The TE process includes 41 measured variables and 12 manipulated variables. There are 21 faults designed by Downs et al. in the TE process. More details of the TE process could found in the literature (Chiang et al. (2001)). In the part, three typical faults are used for illustrating the fault detection power of the proposed method. The three faults are a step of the D feed temperature (Fault 3), C header pressure loss-reduced availability (Fault 7) and an unknown fault (Fault 18). Most of the traditional fault detection methods exhibit a poor effectiveness for fault 3 and fault 18, and are powerful for fault 7 (Pan et al. (2018)).

Among 53 variables, reactor agitator speed stays constant, which is ignored for constructing models. Based on the experience and expert knowledge, the variable 18 is used for detecting the fault 3. The variable 45 is used for detecting the fault 7. The variables 18, 19 are used for

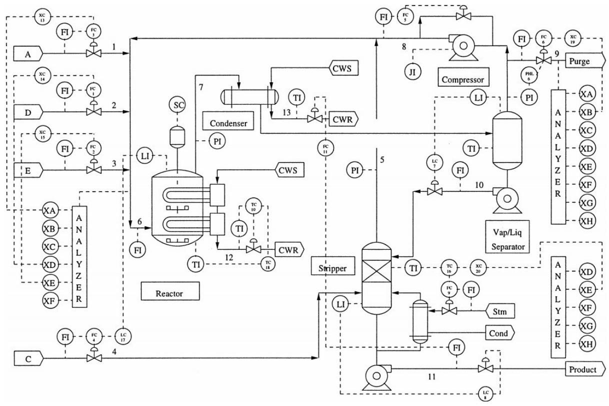

Fig. 3. This is the process flow diagram of the TE process. There are five main units and eight components.

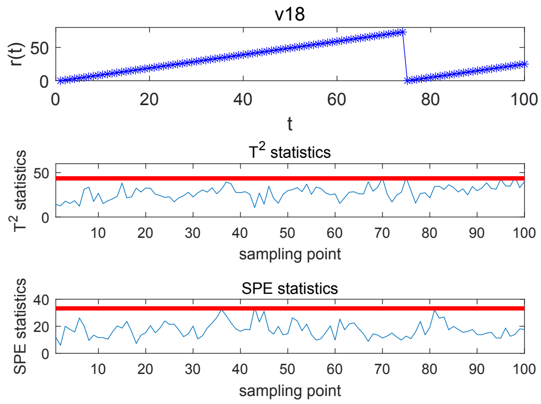

Fig. 4. This is the simulation results of fault 3 in the TE process. The fault is introduced at 61st sampling point, and is detected at 75th sampling point. Although there are 14 sampling points delay, the fault can be detected. The PCA method can not detect the fault 3.

detecting the fault 18, and the D-S theory is used for fusing the fault detection results for two variables of the fault 18. The sampling time is 3 minutes, and each matrix includes 100 observations. The faults are introduced after 3 hours corresponding to 61st sampling point. The means and variances of variables are unknown, and the standard conjugate prior of a normal-inverse-gamma is used for calculating the predictive distribution. Although the means and variances of variables are unknown, the parameters could be set based on the first observation and experience. The unsupervised online fault detection methods for the TE process are barely researched. In the part, the traditional method PCA is applied for comparing with the proposed anomaly detection method. The fault detection results are presented in Fig.4-Fig.7. The prior probability of fault occurrence is 0.028 in the part. The parameters in the PCA method are omitted in the part, which are easy to find in the literature (Chiang et al. (2001)).

The PCA method can not detect the fault 3 and fault 18, and it can detect the fault 7 exactly. Fault 3 is a step fault, and it is difficult to detect the fault 3 based on traditional methods. The related variable is chosen,

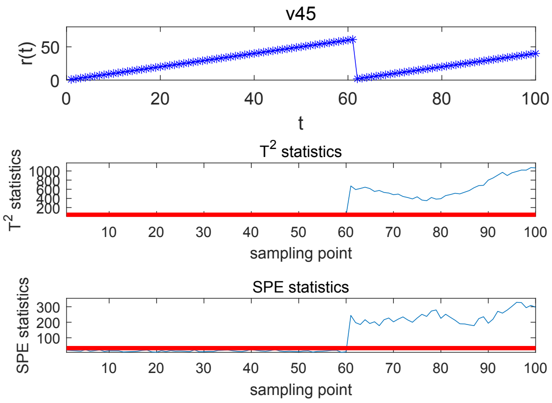

Fig. 5. This is the simulation result of fault 7 in the TE process. The fault is introduced at 61st sampling point, and is detected at 61st sampling point exactly.

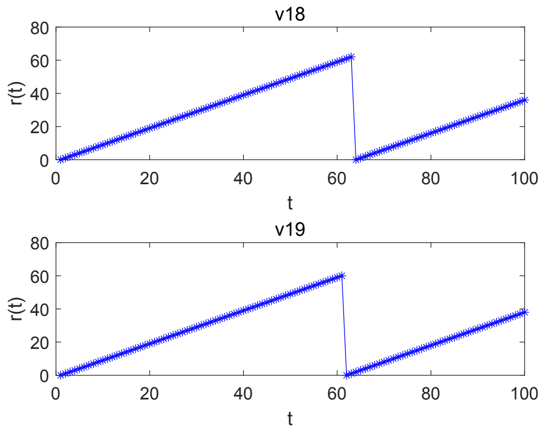

Fig. 6. This is the simulation result of fault 18 in the TE process. The fault is introduced at 61st sampling point. It is detected at 64th sampling point based on variable 18, and is detected at 62nd sampling point based on variable 19.

and the Bayesian-based method is used for detecting it. The Bayesian-based method is unsupervised, and it only uses the current observations to obtain the probability of fault occurrence as a new observation arrives. Although there are 14 sampling points delay, it can detect the fault without any training observations. The unsupervised fault detection is necessary in the industrial processes. Fault 7 is used for testing the power of the Bayesian-based method. Fault 7 is easy to detect by traditional methods, and it also can be detected exactly based on the proposed method. Fault 18 is an unknown fault, and the PCA method exhibits a poor fault detection result. The proposed method could detect the fault based on two variables. The D-S theory is used for fusing the fault detection results, and it can detect the fault at 62nd sampling point. From the run length distribution, it can be concluded that compared with the single variable, the fusion result gives a greater probability value, and it obtains a more powerful evidence for confirming the fault occurrence. In real production processes, it may be unreasonable to know the fault-related

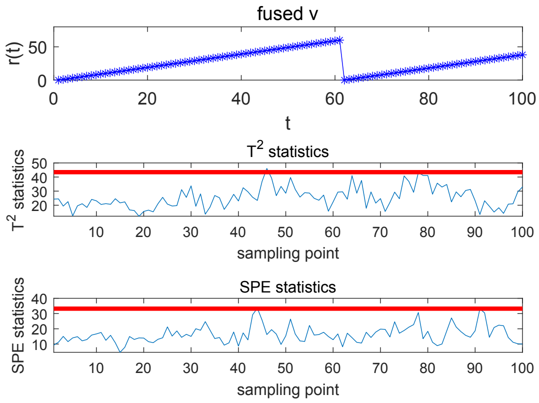

Fig. 7. This is the simulation result of fault 18 in the TE process. There are two related variables for fault 18, and the D-S theory is used for fusing the simulation results. The fault is introduced at 61st sampling point. It is detected at 62nd sampling point based on two variables. The PCA method can not detect the fault 18.

variables. Therefore, the D-S theory fusing the fault detection results of variables is necessary.

5.3 Leakage detection for a bathyscaphe

In the section, the power for leakage detection based on the proposed method is tested in a bathyscaphe equipment. The data is collected on July, 29, 2011, and there is a leakage event occurrence. There are five variables related to the occurrence of leakage including stem trim level detection mercury level detection, stern trim level detection mercury level detection, tank temperature, cabin humidity and ocean temperature. In normal condition, the value of the underwater acoustic communication apparatus leakage detection fluctuating around 9. The less value corresponds to the serious leakage condition. The testing matrix contains 100 observations, and the leakage occurs at 51st sampling point. The prior probability of leakage occurrence is 0.004. The sampling time is 1s. Two PCs are retained based on the PCA method. The control limit of the T^2 statistic is 3.45, and the control limit of the SPE statistic is 3.55. In the section, the leakage detection results based on two variables are presented in Fig.8. Five variables are used for fusing the simulation results, which is shown in Fig.9.

The effectiveness of the proposed method for leakage detection is proved in a bathyscaphe. There is a four sampling points delay based on the ocean temperature variable. The reason may be the ocean temperature is not a standard Gaussian variable. The stern trim level detection mercury level detection variable could detect the leakage exactly. The D-S theory is powerful for fusing the simulation results, and the leakage can be detected at 51st sampling point exactly.

6. CONCLUSION

In the paper, a multivariate Bayesian change point detection method is proposed for anomaly detection in the

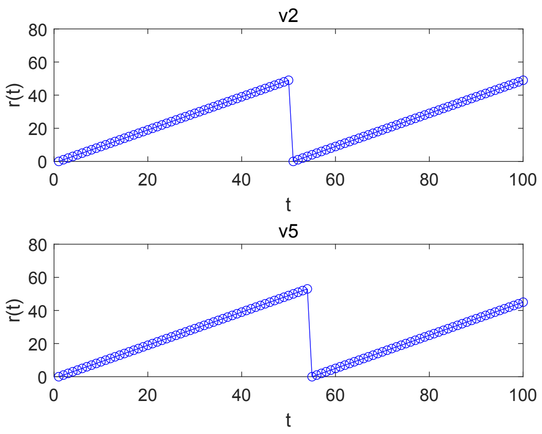

Fig. 8. This is the leakage detection for a bathyscaphe based on two variables: stern trim level detection mercury level detection (v2) and ocean temperature (v5). The leakage occurs at 51st sampling point, and it can be detected exactly based on stern trim level detection mercury level detection variable. The detect time is at 55th sampling time based on the ocean temperature variable.

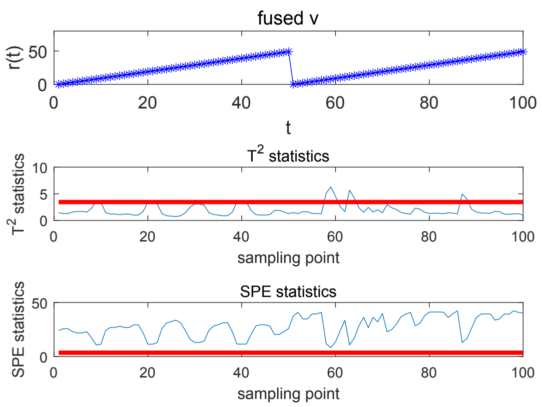

Fig. 9. This is the leakage detection for a bathyscaphe based on five variables and the D-S theory is used. The leakage occurs at 51st sampling point, and it can be detected at 51st sampling point. The PCA method can not detect the leakage fault.

industrial process. A prior probability of change point occurrence is defined, and the D-S theory is used for fusing the simulation results. This is the first attempt for anomaly detection in the industrial processes based on Bayesian change point detection method. The proposed method requires the distributions of observations derived from the exponential family. However, in most cases, it is difficult to satisfy the condition. In the future, some related works should be researched to relax the restriction, i.e. independent principal component analysis solving the non-Gaussian issue.

ACKNOWLEDGEMENTS

This work was supported by the National Key Research and Development Project (Grant No. 2018YFF0214704).

REFERENCES

Adams, R.P. and Mackay, D.J.C. (2007). Bayesian online change-point detection. arXiv Preprint arXiv:0710.3742.

Aminikhanghahi, S. Wang, T.H. and Cook, D.J. (2019). Real-time change point detection with application to smart home time series data. *IEEE Transactions on Knowledge and Data Engineering*, 31(5), 1010-1022.

Aminikhanghahi, S. Wang, T.H. and Cook, D.J. (2019). Real-time change point detection with application to smart home time series data. *IEEE Transactions on Knowledge and Data Engineering*, 31(5), 1010-1022.

An, R.Q. Yang, C.J. and et al. (2015). Comparison of different optimization methods with support vecto machine for blast furnace multi fault classification. *Ifac Papersonline*, 48(21), 1204-1209.

An, R.Q. Yang, C.J. and Pan, Y.J. (2019). Unsupervised change point detection using a weight graph method for process monitoring. *Industrial and Engineering Chemistry Research*, 58(4), 1624-1634.

Chiang, L. Russell, E. and Braatz, R. (2001). *Fault detection and diagnosis in industrial systems*. Springer Verlag.

Chen, H. and Zhang, N.(2015). Graph-based change-point detection. *The Annals of Statistics*, 43(1), 139-176.

Dempster, A.P.(1967). Upper and lower probabilities induced by a multi-valued mapping. *Annals of Mathematical Statistics*, 38(2), 325-339.

Eriksson, M. and Olofsson, T.(2019). Computationally efficient off-Line joint change point detection in multiple time series. *IEEE Transactions on Signal Processing*, 67(1), 149-163.

Ge, Z.Q. Song, Z.H. and Gao, F.R.(2013). Review of recent research on data-based process monitoring. *Industrial and Engineering Chemistry Research*, 52(10), 3543-3562.

Jove, E. Casteleiro-Roca, J.L. and et al.(2019). A fault detection system based on unsupervised techniques for industrial control loops. *Expert Systems*, 36(4).

Murphy, K.P. (2012). *Machine learning A probabilistic perspective*. MIT Press.

Pan, Y.J. Yang, C.J. and et al. (2018). Robust principal component pursuit for fault detection in a blast furnace process. *Industrial and Engineering Chemistry Research*, 57(1), 283-291.

Rong, M.Y. Shi, H.B. and Tan, S.(2019). Large-scale supervised process monitoring based on distributed modified principal component regression. *Industrial and Engineering Chemistry Research*, 58(2), 18223-18240.

Rosenbaum, P. R.(2005). An exact distribution-free test comparing two multivariate distributions based on adjacency. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(4), 515-530.

Wang, Y.X. Wei, Z.X. and Yang, J.W. (2019). Feature trend extraction and adaptive density peaks search for intelligent fault diagnosis of machines. *IEEE Transactions on Industrial Informatics*, 15(1), 105-115.

Wilson, R. C. Nassar, M. R. and Gold, J. I. (2010). Bayesian online learning of the hazard rate in change-point problems. *Neural Computation*, 22(9), 2452-2476.

Zhou, P. Wang, C. and et al. (2018). Modeling error PDF optimization based wavelet neural network modeling of dynamic system and its application in blast furnace ironmaking. *Neurocomputing*, 285, 167-175.