Batch Process Monitoring and Fault Diagnosis Based on MKMFDA

Xiao Yingwang[1, a]

[1] School of Automation, Guangdong Polytechnic Normal University, Guangzhou, China 510665

[a]ymy19701030@163.com

Keywords: MKMFDA; batch process; monitoring and fault diagnosis; fed-batch penicillin fermentation

Abstract: In view of the characteristics of batch process and the defect of batch process monitoring method based on multiway principal component analysis (MPCA), using the advantage of kernel mapping in dealing with nonlinear process and the advantage of fisher discriminant analysis (FDA) in the ability of fault diagnosis, a novel batch performance monitoring and fault diagnosis method based on multi-model kernel multi-way FDA (MKMFDA) was proposed. The key to the proposed approach was to calculate the distance of block data which were projected to the optimal kernel Fisher discriminant vector between new batch and reference batch. Similar degree between the present discriminant vector and the optimal discriminant vector of fault in historical data set was used to perform fault diagnosis. The proposed method was applied to monitoring fed-batch penicillin production, and the results clearly showed that, in comparison to the moving window MPCA method, the proposed method was more accurate and efficient.

Introduction

Currently, the most widely applied of multivariate statistical method for batch process is based on multiway principal component analysis (MPCA) method [1, 2]. However, it has three main defects. Firstly, the method based on MPCA needs to estimate future measurements for online monitoring, then the inevitable prediction error make the monitoring result unreliable. Secondly, the method based on MPCA is essentially a linear modeling technique, hence it usually has bad monitoring effects for highly nonlinear batch processes. Although the neural network has good nonlinear fitting ability, it is difficult to solve the nonlinear optimization problem while training neural network. To overcome the deficiency of neural network, Lee [3] proposed the kernel principal component analysis method for batch process monitoring, and it has a better effect. Lastly, Lee [3] also pointed out that when a fault is detected using KPCA method, it can not take advantage of the contribution plots of PCA to identify the fault type, because we cannot find one inverse projection from high-dimensional feature space to low-dimensional input space. Fortunately, Fisher discriminant analysis (FDA) find optimal Fisher discriminant vector through linear transformation. This makes the data maximum degree of separation for different classes, and its fault diagnosis ability is significantly better than PCA method [4].

MKMFDA Batch Process Monitoring Method

The Division of Multi-way and Multi-model. The data in statistic monitoring method are the historical data collected from the normal batch operating conditions. Because the trajectories of process variables will be influenced to generate some fluctuation by random perturbations in the operations of different batches, this paper will expand the three-dimensional matrix $X_{I \times K \times J}$ (I represents number of production batch, K represents time series, J represents measured variable) of batch process along the batches to obtain the two-dimensional data matrix $X_{I \times KJ}$ of batch process reference trajectory, that is batches with multi-way. We use the modeling method of multi-model structure to overcome the first defect based on MPCA method in section 1. Divide the two-dimensional data matrix $X_{I \times KJ}$ of batch process reference trajectory into a series of sub data space collection, then establish sub-statistical models for each sub data space, that is multi-model structure. The procedure of division is as follows:

(1) Construction of the sub data space:

Divide $X_{I \times KJ}$ into a series of sub data space $X_k^d(I, dJ)$ by moving data window method, $k = d, d+1, \mathbf{L}, k, \mathbf{L} \, K$ represents the data collection on sampling points $[k - d + 1, \mathbf{L}, k]$ with I batch and J process variables used to predict. In which time-lagged variable d is determined by the algorithm in the paper [5]. In this way, the two-dimensional data matrix $X_{I \times KJ}$ can be represented by a series of sub data space collection:

$$X_{I \times KJ} = \{X_d^d(I, dJ) X_{d+1}^d(I, dJ) \mathbf{L} \, X_k^d(I, dJ) \, \mathbf{L} \, X_{K-1}^d(I, dJ) X_K^d(I, dJ)\} \tag{1}$$

(2) Decomposition of the sub data space:

Quantize $X_k^d(I, dJ)$ to a standard data matrix with mean 0 and variance 1, then establish sub statistical model respectively with kernel Fisher discriminant analysis (KFDA) which is described in the paper [3].

Multi-model structure divide the original data space into smaller sub data space, then establish a series of KFDA models to overcome the deficiencies of MPCA global linear modeling. Meanwhile expanding the single model structure to multi-model structure could implement online monitoring.

Fault Diagnosis. The statistical monitoring fault diagnosis based on MPCA is achieved by the method of contribution plots. When the fault occurs, calculating the contribution rate of each variable to fault could roughly determine the cause of the fault. However, it's difficult to find an inverse mapping to calculate the contribution rate of variables in original space to fault after the monitoring methods based on kernel map the data to high-dimensional feature space, so the contribution plots method is not suitable for the monitoring based on kernel. Based on MKMFDA monitoring proposed in this paper, the optimal kernel Fisher discriminant vector can be calculated. The optimal kernel Fisher discriminant vector can be considered as kernel Fisher fault direction vector under the fault conditions. The kernel Fisher fault direction vectors could be different due to different fault conditions. By KFDA analysis for different fault conditions, we can set up a library of optimal kernel Fisher discriminant vectors for fault data set, which contains k types of faults:

$$F = \{a_{opt,1}, a_{opt,2}, \mathbf{L}, a_{opt,k}\} \tag{2}$$

where $a_{opt,i}$ is the optimal kernel Fisher discriminant vector corresponding to i-th type of fault. When diagnosing the faults online, extracting the optimal kernel Fisher discriminant vector a_{cur} of current fault data, then the type of fault could be judged by comparison of the similarity between a_{cur} and $a_{opt,i}$.

$$S_i = \frac{(a_{opt,i})(a_{cur})^T}{\|a_{opt,i}\| \bullet \|a_{cur}\|} (i = 1, 2, \mathbf{L}, k) \tag{3}$$

From the equation (3) above, the similarity S_i is, in fact, the cosine of the angles between two kernel Fisher discriminant vectors. When two vectors are similar, the similarity S_i is close to 1, thus determined as i-th type of fault. In real application, we can set a diagnostic threshold t, it can be determined as i-th type of fault when $S_i \geq t$. If each similarity of a_{cur} of current data and any $a_{opt,i}$ of fault data set is less than the threshold, it is probably that a new fault which was not recognized before occurs. The fault type can be determined by combining with process knowledge, and add it to the optimal kernel Fisher discriminant vector library of fault data set.

The Application for Fault Monitoring of MKMFDA in Fed-batch Penicillin Fermentation Process

Penicillin is a secondary metabolite. In industrial production, the production of such secondary metabolite is produced by filamentous microorganisms. The target product penicillin began to

produce after cell growth. Therefore, we must first allow the growth of microorganisms in a batch culture process, and then promote the synthesis of penicillin through fed-batch filling in glucose. In a typical penicillin fermentation process, most of the essential cells are produced during the initial culture phase. Then the penicillin began to grow exponentially and continued until the cell was stopped. In order to maintain a high yield of penicillin, the rate of cell growth must exceed a certain minimum value. So the glucose should be continuously supplemented to the fermentation tank, rather than filling in at once at the beginning.

Here is applied MKMFDA method to online monitoring for fed-batch penicillin fermentation process. Sampled data is from PenSim v2.0 Emulator [6] designed by researchers of Illinois Institute of Technology in USA. The emulator is capable of simulation for each detection variable during penicillin fermentation process under different operating conditions. The improved MPCA which proposed in recent years, pays more attention to how to implement online monitoring, in which the most typical is Moving Window MPCA (MWMPCA) method [7]. Some other methods have the same idea like it, but with different window moving size.

In this paper, we use the MKMFDA method proposed and MWMPCA [7] method to monitor penicillin batch process to compare the performance of this two methods. In order to get the reference trajectory of monitoring variables and establish normal operating condition models of MKMFDA and MWMPCA, we use PenSim 2.0 model to produce 100 batches of data under standard initial conditions as the normal condition data. The independent Gaussian white noise with a mean of 0 and standard deviation of 0.1 is added to process variables in order to simulate the process noise under real industrial conditions. In addition, the initial conditions of each batch are slightly changed and distributed uniformly within the range of variation. The 10 monitoring variables and the range of their initial values which selected are referred to paper [7]. The duration of each batch of penicillin fermentation process is 400 hours, including about 50 hours of bacteria culture stage and 350 hours of fed-batch fermentation stage, in which the sampling interval is selected as 0.05 hours.

The reduction of feed flow (fault 4) will lead to reducing the concentration of culture medium in fermentation tank. As the nutrient substance of growth and metabolism for microbes, the reduction of medium concentration would affect the concentration of metabolites of penicillin. Figure 1, Figure 2 and Figure 3 are the results of using MWMPCAand MKMFDA methods for monitoring, and the dotted line is the 99% statistical control limit. From the monitoring results, the statistical distance of MKMFDA is immediately exceeding the control limit after the fault occurred. While the faults of statistic T^2 and statistic SPE in MWMPCA method were detected after 70 hours and 45 hours from the fault occurred. Compared with MWMPCA method, the MKMFDA method in this paper can timely monitor the occurrence of faults, thereby facilitating the operator as soon as possible to take measures to ensure the final product quality and reduce loss. Fault diagnosis is achieved by comparing the similarity of optimal kernel Fisher discriminant vectors between current data and historical fault data. As shown in Figure 4, the similarity between optimal kernel Fisher discriminant vector and fault 4 of current data is greatest, with the similar value of 0.91, so we can judge the fault is caused by the reduction of feed flow rate of culture medium.
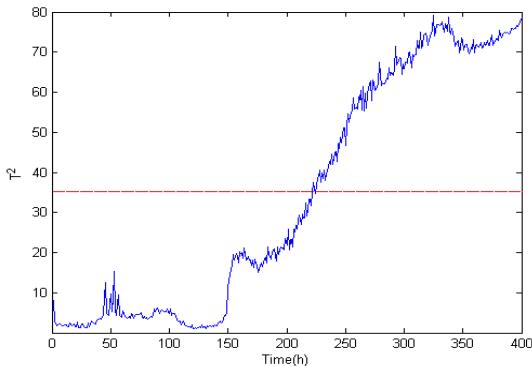

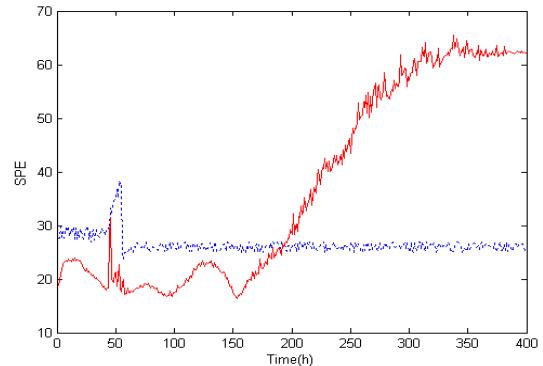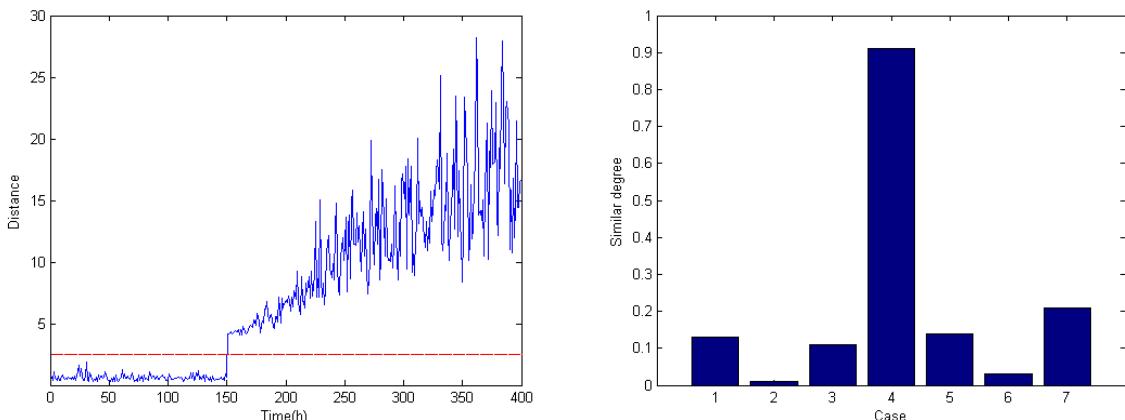Fig.1 T^2 chart using MWMPCA for fault 4   Fig.2   SPE chart using MWMPCA for fault 4

Fig.3 Monitoring chart for fault 4 using MKMFDA   Fig.4 Similar degree of fault 4 in historical database

Conclusions

In this paper, a new method based on MKMFDA is proposed for the main characteristics of limiting production and no realistically steady-state. In order to improve the real-time and sensitivity of batch process monitoring, the running trajectory of process variables is divided into several modules in this paper and each module is considered as a monitoring unit, therefore, it is just need the data of current moment and the time before that moment when using for monitoring. Thus can overcome the deficiency of requirement of estimating process future value when using traditional MPCA method.

Then map the process data from low-dimensional input space into high-dimensional feature space by nonlinear mapping to realize the linear transformation of nonlinear correlation between variables. Monitoring and fault diagnosis is achieved by comparing extracted optimal kernel Fisher feature vector and discriminant vector. So the accuracy and reliability of diagnosis is high.

Using this method for simulation monitoring of penicillin fermentation process and comparing with typical MWMPCA method, it is verified that the proposed method is effective and reasonable.

References

[1] Jie Xu, Shousong Hu: Chinese Journal of Scientific Instrument Vol. 31(11) (2014), p. 2428-2433

[2] Yingwei Zhang, Shuai Li and Yongdong Teng: Chemical Engineering Science Vol. 72(1) (2013), p. 78-86

[3] Lee, J. M., Lee, I. B: Chemical Engineering Science Vol. 59(1) (2006), p. 223-234

[4] Q. P. He, S. J. Qin and J Wang: AIChE Journal Vol. 51(2) (2005), p. 555-571

[5] Jie Yu: Chemical Engineering Science Vol. 68(1) (2012), p. 506-519

[6] Lee, J. M., Yoo, C. K. and Lee, L. B: Journal of Process Control Vol. 14(5) (2004), p. 467-485

[7] Liqing Di, Jie Zhang and Xianhui Yang: Journal of Jilin University Vol. 22 (4) (2004), p. 397-400