FAULT DETECTION IN A BATCH PROCESS USING A BAYESIAN MODEL

TOSHIYUKI NONAKA, YOSHIYUKI YAMASHITA AND
MUTSUMI SUZUKI

Department of Biochemistry and Engineering, Tohoku University,
Sendai 980

Application of Bayesian dynamic modeling to fault detection is developed for a nonstationary batch process. In the modeling, the observed time series are expressed in several specific components such as local polynomial trend, observation noise and globally stationary autoregressive component.

To illustrate the method, detection of a fault in an operation of a stirred vessel with a heater is presented.From the sequential probability ratio test of the model estimation error, the fault can be detected successfully with high sensitivity.

Introduction

To ensure safe operations in chemical plants, early detection and diagnosis of faults are required. For that purpose, numerous kinds of computer-aided systems have been developed and applied to plants with varying success. Most of the systems are for continuous-process plants, and are not easily applicable to batch processes. There are several approaches to solving this problem. For the sake of fault detection and diagnosis by computer, a dynamic model of a process should structurally represent the system as accurately as possible. In traditional chemical engineering practice, a physical model of the process is derived from balance equations and the like.

Park and Himmelblau[11] proposed a method for fault detection and diagnosis for a continuous stirred-tank reactor. Dalle Molle and Himmelblau[6] described a system for a heat exchanger. In their works, the extended Kalman filtering algorithm was used to estimate state variables and system parameters.

But for some batch processes it is difficult to construct a physical dynamic model. Thus an experimental approach such as the black-box model or pattern recognition is needed.

Kutsuwa *et al.*[9,10] used a pattern recognition approach to the problem of fault diagnosis of a batch process.

In our work, a method of real-time fault detection was developed by use of the technique in a time series analysis. Since most measured variables of a batch process have trends, analysis of nonstationary time series is necessary to build models of the system.

Here, a Bayesian model developed by Kitagawa and Gersch[8], is applied to the modeling of normal oper-

ation of processes. Statistical testing for the residual sequence of the model estimation error can indicate the occurrence of a fault. As an experimental example to confirm the method, fault detection is illustrated for a stirred vessel with a heater is described.

1.   Theory

1.1   Bayesian model

Bayesian modeling has applied mainly in the analysis of time series in the field of economics. Bayesian modeling is based on Bayesian statistics[4], which uses Bayes' theorem for analysis or inference. When the probability distribution for a parameter prior to the data and the likelihood of it are given, its probability distribution posterior to this data is proportional to their product. That is,

posterior distribution

$\propto$ likelihood $\times$ prior distribution

The Bayesian approach that we adopt in the nonstationary time series analysis has been used for solving the smoothing problem.

The original paper about this problem was written by Whittaker[12]. He suggested that smoothed time series data were obtained by balancing a tradeoff between infidelity to the observed data and infidelity to a difference-equation constraint of the smoothed data, but an investigator had to choose the value of the trade-off parameter. Akaike[2, 3] suggested that the choice of trade-off parameter be made by maximizing the posterior density of this parameter of a Bayesian model, and proposed a method of trend estimation and seasonal adjustment of nonstationary time series by this model. Kitagawa and Gersch[8] developed a modified model by use of the state space

representation and Kalman filtering.

In this paper we use their model, and will describe that model here. In their Bayesian model, an observed nonstationary time series is assumed to be decomposed into local polynomial trend, seasonal, globally stationary autoregressive, trading day effect, and observation noise components. Their model is intended to apply to the economic situation and including the trading-day effect component, which is used to adjust for the different number of trading days (for example Sunday) of the week per month. But we do not need this component. So the form of the model is

$$y(n) = t(n) + v(n) + s(n) + \varepsilon(n) \tag{1}$$

where $y(n)$ denotes the observed time series, $t(n)$ the local polynomial trend, $s(n)$ the seasonal, $v(n)$ the globally stationary autoregressive and $\varepsilon(n)$ the observation noise components at time n.

Each component has its respective prior information.

1) Local polynomial trend component This component represents the trend and satisfies the kth-order stochastically perturbed difference equation

$$\nabla^k t(n) = w_1(n) \tag{2}$$

where $w_1(n)$ is an independently and normally distributed noise with zero mean and variance τ_1^2, and ∇ is the difference operator, defined as $\nabla t(n) = t(n) - t(n-1)$. For $k = 2$ this component is given by the following equation.

$$t(n) = 2t(n-1) - t(n-2) + w_1(n) \tag{3}$$

2) Globally stationary autoregressive component This component represents the stochastic trend and satisfies an autoregressive (AR) model of order p. That is,

$$v(n) = \alpha_1 v(n-1) + \alpha_2 v(n-2)... $$
$$+ \alpha_p v(n-p) + w_2(n) \tag{4}$$

where $w_2(n)$ is an independently and normally distributed noise with zero mean and variance τ_2^2.

Gersch and Kitagawa[7] showed that good performance was achieved by addition of this component in state estimation and prediction.

3) Local polynomial seasonal component This component represents periodical data and satisfies the following equation.

$$s(n) + s(n-1) + ... + s(n-L+1) = w_3(n) \tag{5}$$

where L is the number of period and $w_3(n)$ is an independently and normally distributed noise with zero mean and variance τ_3^2.

4) Observation noise component This component represents measured error, is independently and normally distributed with zero mean and variance σ^2, and is always included in the model.

These components and the observed time series are also expressed in a state space model. For example, when an observed time series has a second order-local polynomial trend ($k = 2$), a first-order autoregressive component ($p = 1$) and a seasonal one with period 2 ($L = 2$), the space model is expressed in the following form.

$$
\begin{bmatrix} t(n) \\ t(n-1) \\ v(n) \\ s(n) \end{bmatrix} =
\begin{bmatrix} 2 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & \alpha_1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}
\begin{bmatrix} t(n-1) \\ t(n-2) \\ v(n-1) \\ s(n-1) \end{bmatrix}
$$
$$
+ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}
\begin{bmatrix} w_1(n) \\ w_2(n) \\ w_3(n) \end{bmatrix}
$$
$$
y(n) = \begin{bmatrix} 1 & 0 & 1 & 1 \end{bmatrix}
\begin{bmatrix} t(n) \\ t(n-1) \\ v(n) \\ s(n) \end{bmatrix} + \varepsilon(n)
$$

Generally, the space model for an observed time series is

$$x(n) = Fx(n-1) + Gw(n)$$
$$y(n) = Hx(n) + \varepsilon(n) \tag{6}$$

where F, G, H are matrices whose number of row and column is M × M, M × L and 1 × M respectively. x(n) is the state vector, and w(n) the process noise.

When the dimensionality of state space (k, p, L) are defined, and the initial state and the values of parameters $(\tau_1^2, \tau_2^2, \tau_3^2, \sigma^2, \alpha_1,..., \alpha_p)$ are given, the components are calculated by using the Kalman filter recursively in real time. The best model is selected by calculation of AIC[1] for normal data according to Kitagawa and Gersch's method.

$$\text{AIC} = (-2)log \text{ (maximum likelihood)}$$
$$+ 2 \text{ (dimension of F}$$
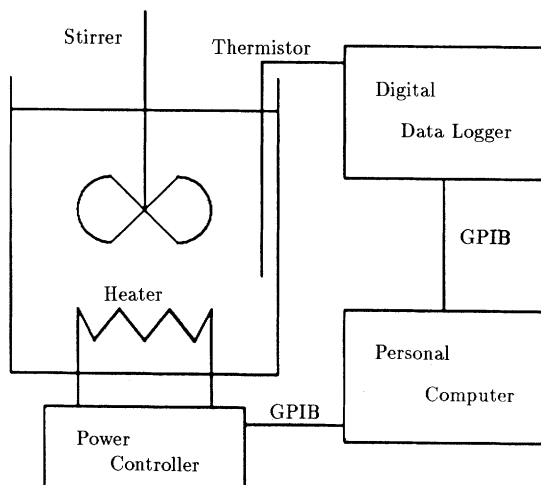$$+ \text{ number of estimated parameters)} \tag{7}$$

where the likelihood of a model is calculated by the following equation.

$$\text{Likelihood} = \prod_{n=1}^{N} (2\pi r (n|n-1))^{-1/2} \exp\left(\frac{-r(n)^2}{2r(n|n-1)} \right) \tag{8}$$

where $r(n)$ is the innovation and $r(n|n-1)$ is the conditional variance of $r(n)$ at time n. The selected model for the normal situation has a minimum value of AIC.

1.2 Fault detection

According to the Kalman filtering theory, the nor-

Fig. 1 Experimental apparatus

Fig. 2 Observed time series $y(n)$ (normal)

Table 1. Comparison of AIC

TREND (k)	AR (p)	SEASONAL (L)	AIC
2	0	0	-2098
3	0	0	-2049
4	0	0	-1997
2	1	0	-2175

malized residual of one-step-ahead prediction has an independent and normal distribution with zero mean and variance one. The error residual does not have a not standard normal distribution; it is considered that the batch process is in abnormal situation. In this work, the fault detection is calculated by SPRT[5] (sequential probability ratio test) based on the error residual. The sequential probability ratio $\lambda(n)$ is computed recursively by use of the following equation.

$$\lambda(n) = \lambda(n-1) + log\frac{p(\eta(n)|H_1)}{P(\eta(n)|H_0)} \qquad (9)$$

where $\eta(n)$ is the error residual at time n, and $P(\eta(n)|H_0)$ and $P(\eta(n)|H_1)$ are the probability density functions of normal (H_0) and abnormal (H_1) situation respectively. In the case where the value of λ is greater than a given upper threshold, the process is regarded as an abnormal situation. When there is an abnormal situation, the mean must be changed, and Eq.(9) is expressed by the following equation.

$$\lambda(n) = \lambda(n-1) + m_a\left\{\eta(n) - \frac{m_a}{2}\right\} \qquad (10)$$

m_a is the mean under an abnormal situation. In that situation a change of variance occurs and Eq.(9) is expressed by the following equation.
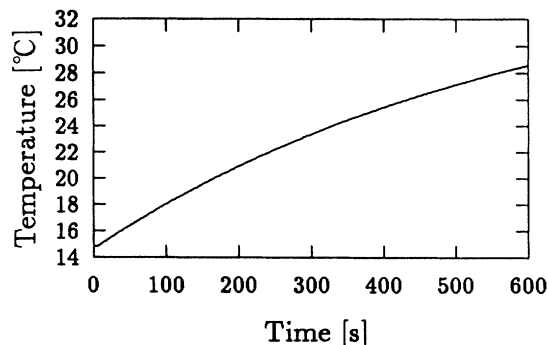
$$\lambda(n) = \lambda(n-1) - \frac{\eta(n)^2}{2}(\frac{1}{v_a} - 1) - \frac{1}{2}log(v_a) \qquad (11)$$

v_a is the variance in an abnormal situation.

2. Experiment

2.1 Modeling

As an example of the procedure's application of the procedure, a simple temperature control process was examined. The experimental apparatus is shown in

Fig.1. Three liters of water in a Dewar vessel was heated by a 500-watt heater, the power of which was controlled to keep the temperature following the scheduled rise. The temperature in the vessel was measured by a thermistor and sampled by a digital data logger and a personal computer every two seconds. Typical operation data of temperature are shown in **Fig.2.** For this process, the best Bayesian model was selected as follows.

The nonstationary nature of batch process data is caused by the existence of trends. In this experiment it was thus assumed that the observed time series $y(n)$ consisted of three components: local polynomial trend $t(n)$, globally stationary autoregressive $v(n)$ and observation noise components $\varepsilon(n)$.

To evaluate some models, the values of AIC for each models were calculated by Eqs. (7) and (8) as shown in **Table1.** The numbers in the table are values of order (k, p) of the local polynomial trend $t(n)$ and the globally stationary autoregressive component $v(n)$, the period (L) of the local polynomial seasonal component $s(n)$, and the AIC for each model. The values of AIC for the models, which had local polynomial trend $k = 2, 3, 4$ and observation noise components $p, L = 0$, indicated that the optimal order of the trend component $t(n)$ was second one $(k = 2)$. Besides, by comparing the values of AIC for the models with and without autoregressive component $v(n)$, the best model was estimated in the case of second-order local polynomial trend $(k = 2)$ and first-order autoregressive component $(p = 1)$.

The trend $t(n)$ and autoregressive $v(n)$ component of the model are shown in **Figs.3** and **4** respectively. The calculated trend component (**Fig.3**) was in good agreement with the observed time series (**Fig.2**). The error residual $\eta(n)$is shown in **Fig.5** rather than the observa-
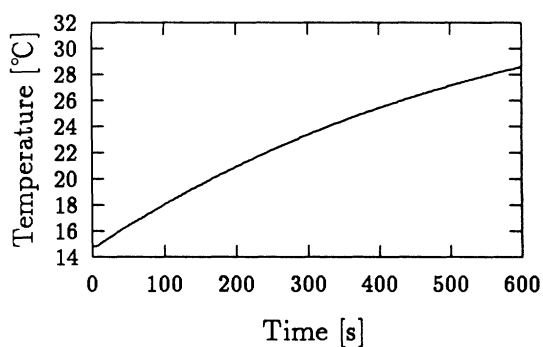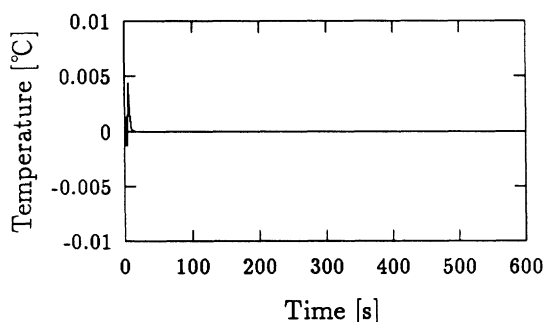
Fig. 3 Trend component $t(n)$ (normal)
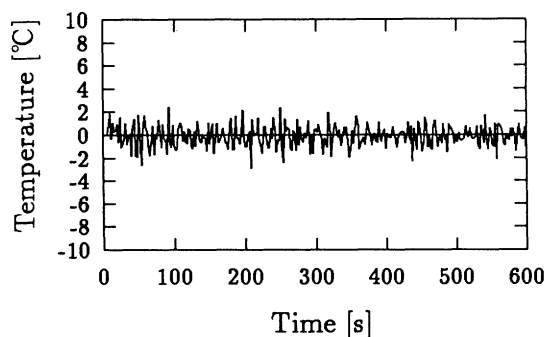

Fig. 4 Autoregressive component $v(n)$ (normal)

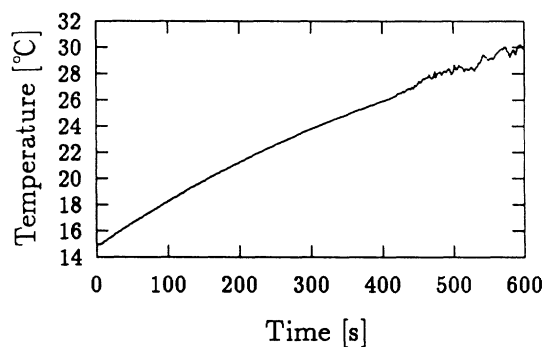**Fig. 5** Error residual $\eta(n)$ (normal)

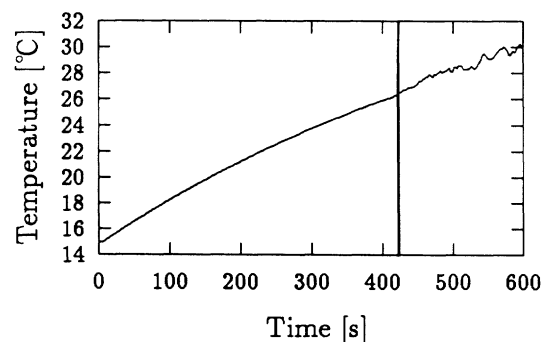**Fig. 6** Observed time series $y(n)$ (abnormal)

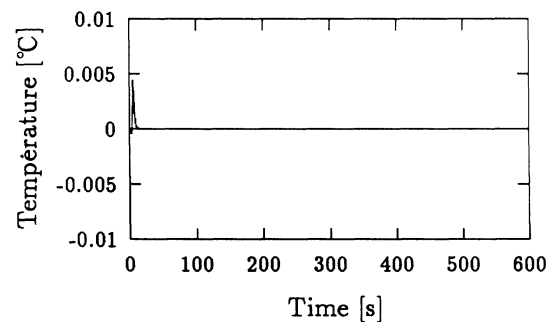
Fig. 7 Trend component $t(n)$ (abnormal)

Fig. 8 Autoregressive component $v(n)$ (abnormal)

tion error component $\varepsilon(n)$.

2.2 Fault detection

To demonstrate the application of Bayesian-model-based fault detection for the residual $\eta(n)$, an abnormal situation was simulated (**Fig.6**) with the same experimental apparatus. In this experiment the stirrer was stopped at 400 seconds ($n = 200$) after the start of the experiment. These data have been decomposed into trend $t(n)$, autoregressive $v(n)$ and observation error components $\varepsilon(n)$ by the above-mentioned Bayesian model.

The trend $t(n)$ and autoregressive $v(n)$ components of the model are shown in **Figs.7** and **8**. The trend component (**Fig.7**) was smoother than the observed time series (**Fig.6**), but the autoregressive component (**Fig.8**) in the abnormal situation did not differ much from normal one (**Fig.4**).

The error residual $\eta(n)$ is shown in **Fig.9**. In the case of the abnormal error residual a rapid change appeared immediately after the occurrence of the fault. SPRT was able to detect the fault at 422 seconds ($n =$

211) by Eq.(10) when the probabilities of false alarms and miss alarms were assumed to be 10^{-8} and the variance in the abnormal situation was 3.

The vertical line in **Fig.7** shows the detection time.

Since the values of parameters were not set to ring false alarms, the required time until detection became somewhat long. But this time seems to be shorter than in the case of operators' monitoring, as shown in **Fig.9**. Thresholds of parameters between normal and abnormal situation must be set. These values were set on the basis of SPRT calculation in the normal situation, and had to be adjusted by experiment. Although the detection of a fault by SPRT is the fastest of all sequential tests in the case when the parameters under the abnormal situation are known, another test may be better to use.

But we think this result indicates the usefulness of the Bayesian modeling.
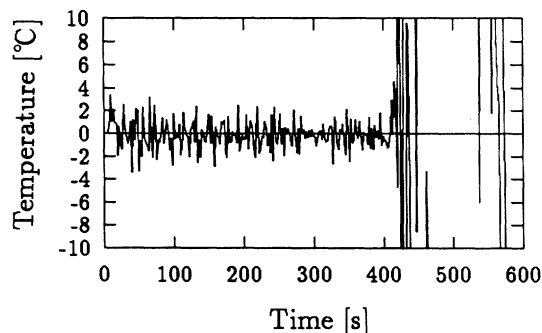
Fig. 9 Error residual $\eta(n)$ (abnormal)

Conclusion

A nonstationary batch process is modeled by a Bayesian dynamic equation, where the observed data are decomposed into three components. The model order of each component is determined by means of AIC. Based on the model, a fault can be detected by a sequential probability ratio test of the error residual. An example of a stirred vessel with a heater is is shown to confirm successfully the fault detection method.

The model adopted in this work is expressed by a linear and Gaussian state space model. Modeling with nonlinear filters may be needed for strongly nonlinear processes. Detailed evaluation of practical processes will clarify the usefulness and limitions of this method.

Acknowledgment

Nomenclature

F, G, H	= matrices of state space	
k	= order of difference equation	[-]
L	= number of period	[-]
m_a	= mean in an abnormal situation	
n	= time	
N	= number of data of a time series	[-]
p	= order of autoregressive model	[-]
$r(n)$	= innovations at time n	
$r(n|n-1)$	= conditional variance of $r(n)$	
$s(n)$	= seasonal component at time n	
$t(n)$	= local polynomial trend component at time n	
v_a	= variance in an abnormal situation	
$v(n)$	= globally stationary autoregressive component at time n	
w_i	= stochastic variable ($i = 1, 2, 3$)	
$w(n)$	= process noise vector at time n	
$x(n)$	= state vector at time n	
$y(n)$	= observed time series at time n	
α_n	= n-th constant of autoregressive model	[-]
$\varepsilon(n)$	= observation error component at time n	
$\eta(n)$	= one-step-ahead output prediction error	
$\lambda(n)$	= probability ratio at time n	[-]
σ^2	= variance of ε	
τ_i^2	= variance of w_i ($i = 1, 2, 3$)	

\<Subscripts\>

1	= local polynomial trend component
2	= globally stationary autoregressive component
3	= seasonal component

Literature Cited

1) Akaike, H.: *IEEE Trans. Auto. Contr.*, **AC-19**, 716-723 (1974).
2) Akaike, H.: in "Bayesian Statistics" eds. Bernado, J.M., M.H. DeGroot, D.V. Lindley, A.F.M. Smith, University Press, Valencia, Spain (1980).
3) Akaike, H.: *J. Time Series Analysis*, **1**, 1-13 (1980).
4) Box, G. E. P. and G.C.Tiao: "Bayesian Inference in Statistical Analysis", Addison-Wesley, Massachusetts, U.S.A. (1973).
5) Chien, T.T. and M.B.Adams: *IEEE Trans. Auto. Contr.*, **AC-21**, 750-757 (1976).
6) Dalle Molle, D.T. and D.M.Himmelblau: *Ind. Eng. Chem. Res.*, **26**, 2482-2489 (1987).
7) Gersch, W. and G. Kitagawa: *J. Business Econ. Stat.*, **1**, 253-264 (1983).
8) Kitagawa, G. and W. Gersch: *J. Am. Statist. Assoc.*, **79**, 378-389 (1984).
9) Kutsuwa, Y., K. Kojima and H. Matsuyama: *Kagaku Kogaku Ronbunshu*, **14**, 20-25 (1988).
10) Kutsuwa, Y. and H. Matsuyama: *Kagaku Kogaku Ronbunshu*, **16**, 76-82 (1990).
11) Park, S. and D.M.Himmelblau: *Ind. Eng. Chem. Process Des. Dev.*, **22**, 482-487 (1983).
12) Whittaker, E.T.: *Proc. Edinborough Math. Soc.*, **41**, 63-75 (1923).