

Efficient Plant Modeling Methods Using Probabilistic Models

Akio Nakabayashi *1

Shouta Ukai *1

Hidehiko Wada *1

In the process industries, applications based on plant models such as real-time estimation of product properties, advanced process controls, and optimization of operating conditions are becoming more important for improving plant operations. Although the solutions based on plant models are very effective, development and maintenance of plant models are not easily achieved. Engineers need to resolve many issues, for example, removal of outliers and selection of model structures, explanatory variables and the order of dynamics. In addition, because plant characteristics change over time, it is necessary to validate and reconstruct current plant models. To overcome these difficulties, Yokogawa has developed a new efficient plant modeling method. Its key idea is that of formulating a modeling procedure applying probabilistic models, and of dealing with parts of the procedure as numerical problems. This paper introduces an overview of the method and gives the results of its numerical experiments.

INTRODUCTION

In recent years, the importance of applications based on plant models has been increasing at operation sites across the process industries. For example, soft sensors that predict product quality in real time, advanced process controls that can operate a plant efficiently and stably, and real-time optimizations that determine appropriate operating conditions for reducing material costs, environmental loads and others, are all based on plant models. To manage competitive plant operations in a rapidly changing economic environment, plant models can be effectively used as tools for various applications.

Although applications based on plant models are very effective when used properly, development and maintenance of plant models is not easy. One of the typical problems when developing a plant model is the removal of outliers. An outlier is an abnormal value measured in a plant. Its causes range widely, including disturbances in a sampling line, or a mistake during data input to a database. Figure 1 shows an unsuccessful example applying a neural network type nonlinear modeling technique to data containing outliers. It is known that chemical processes including reaction systems and so forth generally indicate nonlinear characteristics, and applying a nonlinear modeling technique to them is a natural approach. However, if data contains outliers and modeling is performed without appropriately removing them, the model

may not be properly identified as affected by outliers. During nonlinear modeling, although determining a model structure is a major concern, it is necessary to carefully examine whether each sample needs to be explained by a model. In addition, engineers must resolve many issues when constructing a plant model, such as for selecting explanatory variables and determining the order of dynamics.

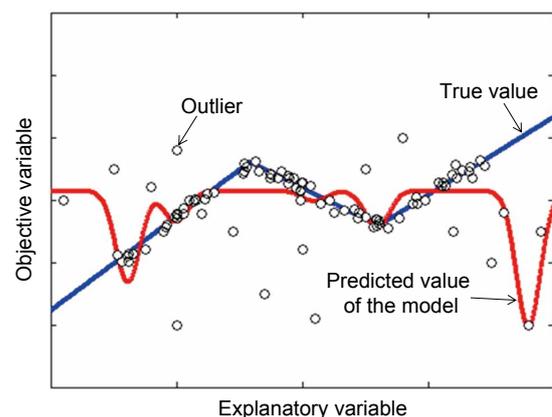


Figure 1 An unsuccessful example of nonlinear modeling

Furthermore, as discussed in the reference (1), degradation of a plant model appearing after its introduction is a possible operational problem. The degradation is caused by changes in a plant such as those in operating conditions due to deterioration and replacement of a catalyst or devices and changes of grades to be produced, and it is a problem arising all the time. Promptly detecting the degradation is a

*1 Research & Development Division, Innovation Headquarters

technical issue, and as well, reconstruction of the plant model is required after the detection. This again causes a burden for plant modeling.

Yokogawa has been working on developing technologies to effectively and simply develop and maintain applications using plant models. This paper introduces an attempt to make plant modeling more efficient by using probabilistic models.

The concept of the probabilistic model is opposite to that of the conventional deterministic model. The model parameters conventionally assumed to be constant are treated as random variables, and each model parameter is estimated by calculating a posterior probability distribution of each random variable for the observation data. That the model parameters are random variables means that various possible models are simultaneously considered based on the probability of each model, that is, overall possible models are being evaluated.

The applicable scope of the probabilistic model includes not only determining simple model parameters such as regression coefficients, but also determining structural parameters of a model such as the order of dynamics, selecting explanatory variables, and judging whether each sample is an outlier or not. If a wide-ranging investigation required for plant modeling is formulated as a probabilistic model and transformed into numerical problems, automatic or semi-automatic plant modeling, assisted by interaction with engineers, becomes possible.

This paper introduces an overview of developed technologies, taking examples of applications of the probabilistic model to outlier and nonlinear modeling problems among various applications.

PLANT MODELING BASED ON PROBABLISTIC MODELS

Assumed Plant Model

First, this section defines the structure of a plant model assumed in this paper. A piecewise linear model is supposed as a plant model. The piecewise linear model is defined as follows:

$$y = \begin{cases} [x^T \ 1] w_1 + \epsilon & \text{if } [x^T \ 1] \in V_1 \\ \vdots & \\ [x^T \ 1] w_K + \epsilon & \text{if } [x^T \ 1] \in V_K \end{cases}$$

where y is an objective variable, x is an explanatory variable, and ϵ is an observation noise. V_k ($k = 1, \dots, K$) are subspaces defined in the explanatory variable space.

As shown in Figure 2, this model supposes that the explanatory variable space is divided into several subspaces, and that a different linear regression model is applied to each subspace. The parameters w_k ($k = 1, \dots, K$) are the regression coefficients of each liner regression model. For example, the piecewise linear model assumes such a situation in which a plant is operated under multiple operational conditions according to grades, loads, etc., and has different characteristics depending on the range of those conditions.

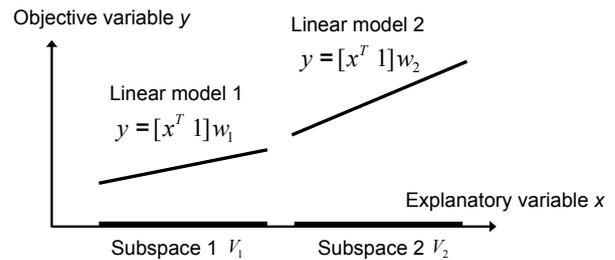


Figure 2 Piecewise linear model

Even if a process characteristic is nonlinear, the piecewise linear model can express it by subdividing its subspaces. From a practical point of view, it has enough representation capability as do general plant models.

In recent years, a technique called the Just-In-Time (JIT) model⁽²⁾ has been used together with its successful results in process industries. The piecewise linear and the JIT models are intrinsically equivalent because an explicit expression of the JIT model is the piecewise liner model.

Applying the piecewise linear model to plant modeling has two advantages. The first is the advantage when applying it to control. Because the operating conditions of a plant generally remain within a certain range for a certain period of time, if a characteristic within the range is expressed with a linear model, the mature linear system theory can be easily applied for designing a control system. In addition, even when considering control covering multiple operating conditions, the results of hybrid systems theory can be made use of.

The second one is extensibility to a dynamic model by using its history of measurements. In that case, the above-mentioned piecewise linear model is referred to as the piecewise autoregressive exogenous (PWARX) model.

Although the piecewise linear model is convenient as described above, it is very difficult to identify the model. Because subspaces into which the explanatory variable space is divided are usually unknown, they must be estimated simultaneously with regression models. Moreover, the appropriate number of divided subspaces is also usually unknown; and it must also be simultaneously estimated. These problems cannot be analytically solved, unlike an inverse problem in the case of linear regression problems, but require many trials and errors. Furthermore, as described previously, plant data often include outliers, which must be appropriately removed when applying the piecewise linear model as well.

To overcome these difficulties, we have developed a technique to stably estimate the piecewise linear model from data including outliers by using probabilistic models.

Estimating a Piecewise Linear Model from Data Containing Outliers

This paper introduces only an overview of the developed technique due to limited space. For more details, the reference (3) can be referred.

Information required for constructing a piecewise linear model from data including outliers is expressed as one vector

parameter θ . We refer θ as the engineering parameter, which includes such a parameter representing necessity of removal of each sample as an outlier, and a parameter indicating the subspace to which each sample belongs. Although θ also includes other parameters, the details of them are omitted here.

When θ is determined to be a certain value, the information required to construct the piecewise linear model is decided upon, which includes which sample is judged as an outlier, which samples are contained in which subspaces, and what is the number of the subspaces, so that parameters included in the model formula can be determined. In other words, one piecewise linear model is obtained for one instance of θ . To determine θ , we regard θ as a random variable. The appropriate θ can be obtained by estimating the posterior distribution $p(\theta | Y, X)$, where Y denotes the time series data of the objective variable and X denotes the time series data of the explanatory variable. Thus, engineering for constructing the piecewise linear model can be easily performed by numerical calculation.

Figure 3 shows a probabilistic model for associating θ with X and Y . $Z(z_i \in Z)$, $S(s_i \in S)$, $w, \beta, \mu, \Lambda, \pi, C_{out,y}$ and $C_{out,x}$ in the figure are random variables that comprise θ . This figure denotes the relation of cause and effect among each objective, explanatory and random variable using a directed graph, and the arrows finally reach objective variables y_i . That is, the graph is a model representing the process where objective variables y_i are generated. By using the Bayes' theorem, we can estimate the posterior probability distributions of each unobserved random variable from observed data. The details of each random variable and the shape of its probability distribution are described in the reference (3). Meanwhile, in the case where removal of outliers is not taken into consideration, formulation based on this probabilistic model is generally used for dealing with a mixture model into which multiple models are combined⁽⁴⁾. The feature of the technique developed by us is that the removal of outliers, which is important when handling plant data, has been added to the formulation.

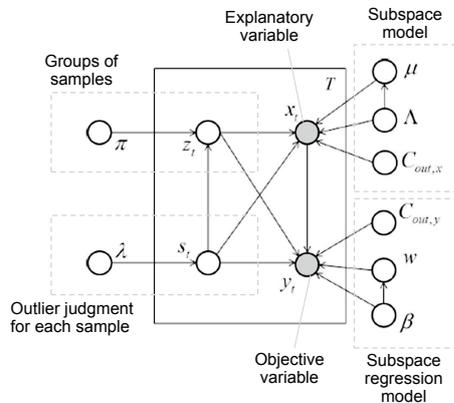


Figure 3 A probabilistic model for constructing a piecewise linear model under the condition where data contains outliers

Although it is necessary to calculate the posterior probability distribution of θ based on this probabilistic model, this calculation is not tractable. Thus, we have decided to estimate the posterior probability distribution on the basis of the Variational Bayes method, one of the approximation methods for posterior probability distribution calculation, in which research has been conducted in the machine learning area. The formulation based on the Variational Bayes method is shown below:

$$p(\theta|Y,X) \approx \tilde{q}(\theta) = \arg \max_{q(\theta)} F[q(\theta)]$$

$$F[q(\theta)] \equiv E_q[\ln p(Y,X,\theta)] + H[q(\theta)]$$

where q is an arbitrary probability distribution of the random variable θ , E_q is an operator for calculating an expected value under the probability distribution q , and the expected value to be calculated is a log likelihood under the supposed probabilistic model corresponding to θ and the data Y and X . H is the entropy of the probability distribution q . The functional F that takes q as an argument function provides the lower bound of the log marginal likelihood $\ln p(Y, X)$.

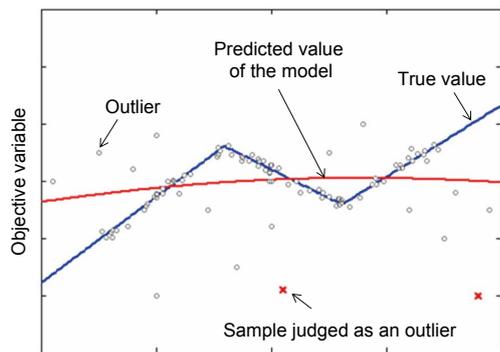
The Variational Bayes method is formulated as an optimization problem that maximizes the functional F for the probability distribution q , and can be solved by iterative calculation.

The functional F has the meaning as follows: the first term indicates how suitably the observed data is described by the piecewise linear model defined by a certain engineering parameter θ as the result in which the expected value is calculated over the entire θ . The second term represents the coverage of θ used for the calculation, that is, the coverage of the model. Therefore, this optimization problem can be interpreted as a problem as a whole; a search for the engineering parameters defining the model which can suitably describe the observed data as exhaustively as possible.

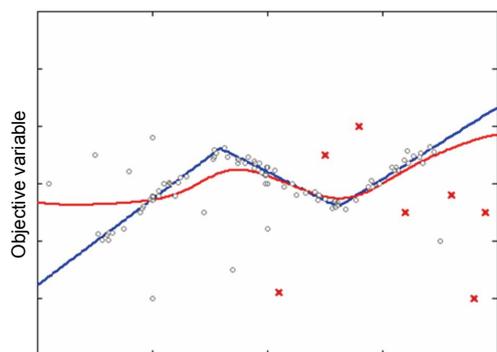
Once the posterior probability distribution of θ is obtained, it is possible to determine the unique model by using the mode value of θ . In addition, various possibilities of θ can be probabilistically combined into one model by taking the expectation.

Numerical Experiments

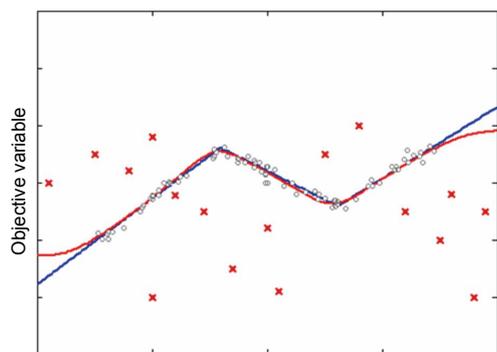
Figure 4 shows the obtained results by applying the developed technique to the same data as that in Figure 1. Although the posterior probability distribution of each random variable has been calculated, it is difficult to visualize the probability distributions. Therefore, the figure provides, as the estimated result, only markings on samples whose expected value of the probability to be an outlier exceeds 0.5 and the predicted values combined by calculation of the expectation. Figure 4 (a), (b), (c) shows the iterative calculation process for the estimation of the posterior distribution by the Variational Bayes method. As the iteration proceeds, outliers are more properly judged, and the predicted values of the model are becoming close to the true values.



(a) Initial stage of calculation process



(b) Middle stage of calculation process



(c) At the completion of calculation

Figure 4 Results of the determination of outliers (expected value of the probability is 0.5 or more) and the predicted values of the piecewise linear model

CONCLUSION

This paper has introduced plant modeling based on the probabilistic model as one of the attempts for developing technologies at Yokogawa. We have indicated that the newly developed method makes it possible to automatically obtain the piecewise linear model from data containing outliers. This means that the problems of outliers and nonlinear modeling in plant modeling can be dealt with mathematically, suggesting possibilities for significantly reducing loads imposed on the development of plant models.

In the future, Yokogawa will conduct its field tests in the actual environment, and prepare peripheral technologies required for usage on site.

REFERENCE

- (1) M. Kano, M. Ogawa, "The state of the art in chemical process control in Japan: Good practice and questionnaire survey," *Journal of Process Control*, Vol. 20, Issue. 9, 2010, pp. 969-982
- (2) S. Ukai, A. Nakabayashi, H. Wada, et al., "A plant simulator based on hybrid model of physical model and Just-In-Time model using statistical approach," *SICE Annual Conference 2011*, September 2011
- (3) A. Nakabayashi, S. Ukai, H. Wada, et al., "A Bayesian Robust Identification Method for Piecewise Affine Autoregressive Exogenous Model from Outlier Contaminated Data," *SICE Annual Conference 2013*, September 2013
- (4) C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006