Adaptive Fuzzy Modeling For A Large-Scale Nonlinear System

Jialin Liu[1]

[1]Department of Information Management, Fortune Institute of Technology,
1-10, Nwongchang Rd., Neighborhood 28, Lyouciyou Village, Daliao Township,
Kaohsiung Country, Taiwan, Republic of China
E-Mail: jialin@center.fotech.edu.tw

Abstract

A data-driven Takagi-Sugeno (TS) fuzzy model is developed for modeling a real plant with the dependent inputs, the nonlinear and the time-varying input-output relation. The collinearity of inputs can be eliminated through the principal component analysis (PCA). The TS model split the operating region into a collection of *IF-THEN* rules. For each rule, the premise is generated from clustering the compressed input data and the consequence is represented as a linear model. A post-update algorithm for model parameters is also proposed to accommodate the time-varying nature. Effectiveness of the proposed model is demonstrated using real plant data from a polyethylene process.

Keywords: Nonlinear system modeling, principal component analysis, Bayesian classification, Takagi-Sugeno fuzzy method.

1. Introduction

Measurement noises and redundancy often exist in the process operating data, the former corrupts the data quality that reduce the model reliability, the latter induces collinearity of the variables that results in over-fits. Principal component analysis (PCA) is a popular multivariate statistical method of compressing high-dimensional inputs into a few orthogonal variables. It is also advantageous to use compressed data for process modeling is that the alarms would be triggered when the on-line data are not consistent with the reference dataset that was used to construct the model. Thus unreasonable extrapolation will be avoided.

The mixture model is a linear combination of priori probabilities and conditional probability density functions. It is used for evaluating the unknown probabilities in Bayesian classification. Generally, the maximum likelihood (ML) method is used to estimate the parameters of mixture model, the means and covariances of each density function and the priori probabilities. Before the new observations are classified into the clustered groups, the conditional probabilities can be used to evaluate the goodness of the data in belonging to the mixture. An outlier would have a low value for any of conditional probability density functions [1]. On the contrary, a good data at least has a high conditional probability.

In this paper, historical data are used for building a piecewise linear virtual sensor model for inferring the final quality using process inputs. Before building inferential model, PCA is applied to remove collinearity within inputs and filter out outliers from plant data. The compressed data are then clustered using Bayesian classification into antecedents of *IF-THEN* rules in a TS fuzzy model to represent different operating regions. A linear prediction model is trained as the consequent of each *IF-THEN* rule. In on-line prediction phase, the process inputs should be within the PCA subspace, and belong to the at least one known operation regions. Otherwise, alarms are triggered to inform operator about abnormalities instead of generating unreliable predictions. When the PCA subspace is reconstructed in order to cover the new events, the clusters and the inferential models are updated to the new PCA subspace using the proposed method.

2. Basic Theory

2.1. Principal Component Analysis

Consider the data matrix $\mathbf{W} \in R^{m?n}$ with m rows of observations and n columns of variables. Each column is normalized to zero mean and unit variance: $\mathbf{X} = \left(\mathbf{W} - \mathbf{1}\overline{\mathbf{W}} \right)\mathbf{S}^{-1}$ where $\overline{\mathbf{W}}$ is a mean vector, $\mathbf{1}$ is a column vector that elements are one, and $\mathbf{S}$ is a diagonal matrix of standard deviations. The eigenvectors ($\mathbf{P}$) of the covariance matrix can be obtained from the normalized data. The score vectors

(**t**) are the projection of the data matrix **X** to each eigenvector. The data matrix **X** can be decomposed as:

$$\mathbf{X} = \sum_{i=1}^{K} \mathbf{t}_i \mathbf{p}_i^{\mathrm{T}} + \sum_{i=K+1}^{n} \mathbf{t}_i \mathbf{p}_i^{\mathrm{T}} = \hat{\mathbf{X}} + \mathbf{E} \qquad (1)$$

with $\hat{\mathbf{X}}$ being the projection of the data matrix **X** in the subspace formed by the first K eigenvectors and **E** being remainder of **X** that is orthogonal to the subspace.

The statistic Q is defined in order to examine the new data can be explained by PCA subspace or not.

$$Q = \mathbf{e}\mathbf{e}^{\mathrm{T}} = (\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^{\mathrm{T}} = \mathbf{x}(\mathbf{I} - \mathbf{P}_K \mathbf{P}_K^{\mathrm{T}})\mathbf{x}^{\mathrm{T}} \quad (2)$$

The confidence limit of Q is defined as follows [2]:

$$Q_\alpha = \Theta_1 \left[\frac{c_\alpha \sqrt{2\Theta_2 h_0^2}}{\Theta_1} + 1 + \frac{\Theta_2 h_0 (h_0 - 1)}{\Theta_1^2} \right]^{1/h_0} \qquad (3)$$

$$h_0 = 1 - \frac{2\Theta_1 \Theta_3}{3\Theta_2^2}, \quad \Theta_i = \sum_{j=K+1}^{n} \lambda_j^i, \qquad i = 1,2,3$$

The percentile α is the probability of type I error in hypothesis testing and c_α is the standard normal deviate corresponding to the upper $(1-\alpha)$ percentile. Another measure of the difference between new data and the PCA subspace is the statistic T^2.

$$T^2 = \mathbf{x}\mathbf{P}_K \ddot{\mathbf{E}}^{-1} \mathbf{P}_K^{\mathrm{T}} \mathbf{x}^{\mathrm{T}} \qquad (4)$$

The diagonal matrix $\ddot{\mathbf{E}}$ is first K terms of eigenvalues. The confidence limit is defined as:

$$T_\alpha^2 = \frac{K(m-1)}{m-K} F_{K,m-1,\alpha} \qquad (5)$$

where $F_{K,m-1,\alpha}$ is an F distribution with the degrees of freedom K and $m-1$. The new data belong to PCA subspace within $(1-\alpha)$ confidence limits only when $Q < Q_\alpha$ and $T^2 < T_\alpha^2$.

2.2. Bayesian Classification

The priori probabilities that c classes exist are $P_j, j = 1...c$. Given the conditional probability density functions of $\mathbf{x}_i$ in class j, $p(\mathbf{x}_i | j; \grave{\mathbf{e}})$, the joint probabilities of $\mathbf{x}_i$ belonging to class j are:

$$p(\mathbf{x}_i, j; \grave{\mathbf{e}}, P_j) = p(\mathbf{x}_i | j; \grave{\mathbf{e}}) P_j \qquad (6)$$

where $\grave{\mathbf{e}}$ is the parameter vector of conditional probability density function. The posteriori probabilities can be gotten from the total probability theorem and Bayes rule.

$$P(j | \mathbf{x}_i; \grave{\mathbf{e}}, \mathbf{P}) = \frac{p(\mathbf{x}_i | j; \grave{\mathbf{e}}) P_j}{p(\mathbf{x}_i; \grave{\mathbf{e}})} = \frac{p(\mathbf{x}_i | j; \grave{\mathbf{e}}) P_j}{\sum_{k=1}^{c} p(\mathbf{x}_i | k; \grave{\mathbf{e}}) P_k} \quad (7)$$

Here $\mathbf{P} = (P_1, P_2, ... P_c)$ is the priori probability vector. Those parameters can be iterated from Expectation-Maximization (EM) algorithm.

2.3. Fuzzy TS Modeling

Takagi and Sugeno [3] proposed a fuzzy rule-based model, which decompose a nonlinear system into c fuzzy rules:

$$R_i: \quad \textbf{If } \mathbf{t}_j \textbf{ is } A_i \textbf{ then } \mathbf{y}_{ij} = f_i(\mathbf{x}_j) \qquad (8)$$

where $\mathbf{t}_j$ are the jth inputs, which are the score vectors from PCA, A_i and f_i respectively are the antecedent part and the consequence inferential function of the ith rule (R_i), the $\mathbf{y}_{ij}$ are the outputs estimated from the ith rule and the jth inputs. The antecedent parts of TS model are generated from FCM-like algorithm in most of fuzzy modeling applications. However, since that the memberships of an observation across all of classes must sum to 1. It forces the new observation classifying into groups, even though the outliers were observed. In this paper, the antecedents of TS model are generated from Bayesian classification, which posteriori probabilities are instead of memberships. The advantage is the outliers of new observations can be identified from the conditional probabilities of the new data, which outliers should be grouped into unknown events for next model updating, instead of estimating outputs from the current model. Therefore, the antecedent part for each rule can be written as:

$$A_i = \left\{ \mathbf{t}_j \middle| p(\mathbf{t}_j, i) \geq p_{thr} \quad and \quad P(i | \mathbf{t}_j) \geq P_{thr} \right\} (9)$$

where p_{thr} and P_{thr} respectively are the predefined thresholds for the conditional and the posteriori probabilities, the former determines the tolerance of outliers and the latter represents the share-ability for each observation.

3. Update Inferential Model

Consider the addition of data of new events to the original data matrix. The numbers of observations are increasing from m to m^*. The data matrix $\mathbf{W}^{*\mathrm{T}} = \begin{bmatrix} \mathbf{W}^{\mathrm{T}} & \mathbf{W}_{\mathrm{new}}^{\mathrm{T}} \end{bmatrix}$ is normalized to $\mathbf{X}^*$, which each column is with zero mean and unit variance. The mean vector and the standard deviation matrix are $\overline{\mathbf{W}}^*$ and $\mathbf{S}^*$. The loading vectors $\mathbf{P}^*$ and correspondingly score vectors $\mathbf{T}^*$ can be determined by assuming $\mathbf{E}^* \approx 0$ in the new subspace. The center ($\grave{\mathbf{i}}_j$) and the covariance ($\acute{\mathbf{O}}_j$) of any class j in the original subspace can be transferred to the new subspace:

$$\grave{\mathbf{i}}_j^* \approx \grave{\mathbf{i}}_j \mathbf{A} + \Delta \overline{\mathbf{W}} \mathbf{S}^{*-1} \mathbf{P}^*, \quad \acute{\mathbf{O}}_j^* \approx \mathbf{A}^T \acute{\mathbf{O}}_j \mathbf{A} \qquad (10)$$

$$\mathbf{A} \equiv \mathbf{P}^{\mathrm{T}}\mathbf{S}\,\mathbf{S}^{*-1}\mathbf{P}^{*}, \ \Delta\overline{\mathbf{W}} \equiv \overline{\mathbf{W}} - \overline{\mathbf{W}}^{*}$$

The jth center ($\grave{\mathbf{i}}_{j}^{*}$) and covariance ($\acute{\mathbf{O}}_{j}^{*}$) on the new subspace can be obtained through linear operation with coordinate rotation ($\mathbf{A}$) and shifting ($\Delta\overline{\mathbf{W}}\mathbf{S}^{*-1}\mathbf{P}^{*}$), without re-clustering the trained data on the new subspace.

There are two steps for adapting the TS model. Firstly, the regression matrix $\mathbf{B}_{i}$ of the ith rule is updated to the new subspace. After that, the recursive least square algorithm and the additional (m^{*}-m) observations are used to correct the regression matrices.

Classify the new quality data $\mathbf{Y}_{new}$ into each class; let $\mathbf{Y}_{i}^{*}$ be the quality data of the ith class, i.e., $\mathbf{Y}_{i}^{*\mathrm{T}} = \begin{bmatrix} \mathbf{Y}_{i}^{\mathrm{T}} & \mathbf{Y}_{i,new}^{\mathrm{T}} \end{bmatrix}$. Assuming that there are m_i observations in the ith class before adding the new data, the regression function is:

$$\mathbf{Y}_{i,(m_i)} = \left(\mathbf{T}_{i,(m_i)}^{*} - \mathbf{1}\grave{\mathbf{i}}_{i}^{*}\right)\mathbf{B}_{i,(m_i)}^{*} + \mathbf{E}_{i,(m_i)} \qquad (11)$$

where $\mathbf{E}_{i,(m_i)}$ is the regression error. From above equation, the regression matrix in the new subspace ($\mathbf{B}_{i,(m_i)}^{*}$) can be derived.

$$\mathbf{B}_{i,(m_i)}^{*} = \left(\mathbf{C}_{i,(m_i)}^{*\mathrm{T}}\mathbf{C}_{i,(m_i)}^{*}\right)^{-1}\mathbf{C}_{i,(m_i)}^{*\mathrm{T}}\mathbf{Y}_{i,(m_i)} \qquad (12)$$

$$\mathbf{C}_{i,(m_i)}^{*} \equiv \mathbf{T}_{i,(m_i)}^{*} - \mathbf{1}\grave{\mathbf{i}}_{i}^{*}$$

where $\mathbf{C}_{i,(m_i)}^{*}$ can be obtained from the previous subspace scores and Eq. 10. Therefore, the regression matrix in the new subspace before adding the new information can be directly transferred from the previous subspace:

$$\mathbf{B}_{i,(m_i)}^{*} = \mathbf{A}^{-1}\mathbf{B}_{i,(m_i)} \qquad (13)$$

After updating the TS model to the new subspace, the regression matrices need to be corrected using the additional data. Assuming that there are Δm_i observations are added into class i, $\Delta m_i \equiv m_i^{*} - m_i$. The matrix is corrected using the new data and the recursive least square algorithm [4].

$$\mathbf{B}_{i,(m_i*)}^{*} = \mathbf{B}_{i,(m_i)}^{*} + \mathbf{D}\left(\mathbf{Y}_{(\Delta m_i)} - \left(\mathbf{T}_{(\Delta m_i)}^{*} - \mathbf{1}\grave{\mathbf{i}}_{i}^{*}\right)\mathbf{B}_{i,(m_i)}^{*}\right) \quad (14)$$

$$\mathbf{D} \equiv \left(\mathbf{C}_{i,(m_i)}^{*\mathrm{T}}\mathbf{C}_{i,(m_i)}^{*}\right)^{-1}\mathbf{T}_{(\Delta m_i)}^{*}\left(\mathbf{I} + \mathbf{T}_{(\Delta m_i)}^{*}\left(\mathbf{C}_{i,(m_i)}^{*\mathrm{T}}\mathbf{C}_{i,(m_i)}^{*}\right)^{-1}\mathbf{T}_{(\Delta m_i)}^{*\mathrm{T}}\right)^{-1}$$

where $\mathbf{B}_{i,(m_i*)}^{*}$ and $\mathbf{B}_{i,(m_i)}^{*}$ are the regression matrices respectively with m_i^{*} and m_i observations.

4. Illustrative Example

In this paper, the polyethylene plant is located in Kaohsiung, Taiwan. In Fig. 1, the process flow diagram has been shown; the detailed description can be found in the previous work [5]. In this paper, 14 inputs of the process model, which have been chosen according to the plant expert's advices. The process final quality is the laboratory MI that is measured every 4 hours. The data of 14 inputs and 1 output were collected in one month for the training dataset. There are 150 observations in the dataset, including three different production grades, labeled as #1, #2 and #3. There are 145 observations can be explained by the PCA subspace, which the captured variances were 91.68% with 4 PCs, within 99% confidence limits. The clustering results of projection of first two score vectors are shown in Fig. 2, the solid line for each group represents the conditional probabilities equal 10^{-4}, which the conditional probability threshold for outlier was set in this study. The regression matrix for each linear model was generated according to the data from the group that the posteriori probabilities of observations are greater than 0.1. The relative root-mean-square errors (RRMSEs) are 9.16% for the training dataset using the proposed method, the comparisons of regression results of this method and partial least squares (PLS) have been shown in Fig. 3. It is obvious that only one set of parameters in the PLS cannot properly explain outputs with three different grades of products, which the RRMSEs are 31.81%.
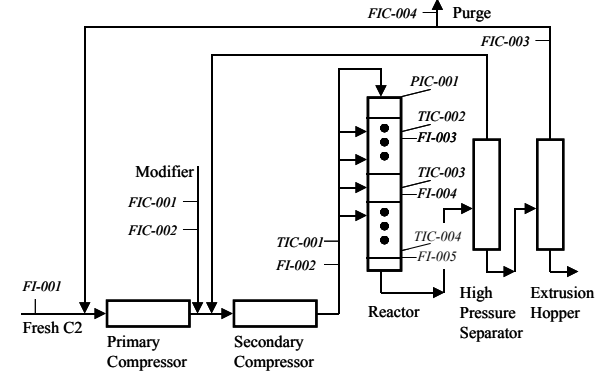

Fig. 1: Process flow diagram of polyethylene process.

Additional 153 observations in one month following the period of the above training data were collected for the test dataset. The statistic Q and T^2 of new data are shown in Fig. 4, only in the case of the $Q<Q_\alpha$ and $T^2<T^2_\alpha$, the new data belong to the PCA subspace within $(1-\alpha)$ confidence limits. In this dataset, there are only 88 observations can be properly explained by the PCA subspace. It is because of that a new grade had emerged after the observation number 113. The RRMSEs of predictions within the PCA subspace are 12.84% for the proposed method and 53.12% for the PLS. The new PCA subspace was constructed using the training and the test datasets in order to explain the data of all events. The captured variances were 90.53% with 4 PCs; there were 290 observations within the new subspace. Fig. 5 shows

the comparisons of regression results with the proposed method and the recursive PLS algorithm [6]; the RRMSEs respectively are 10.07% and 44.10%. The RPLS has reduced the RRMSEs from 53.12% to 44.10%, showing the model was adapted with the new information, but it still cannot tackle the input-output nonlinearity.
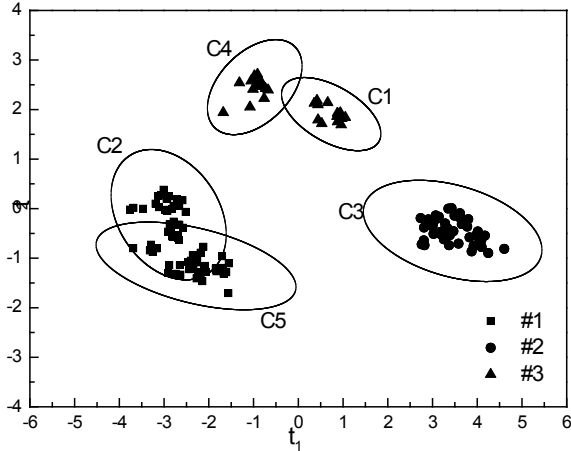

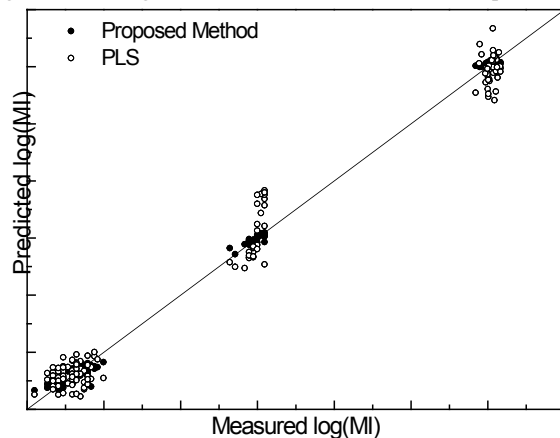Fig. 2: Clustering results on the t_1-t_2 of the PCA subspace.
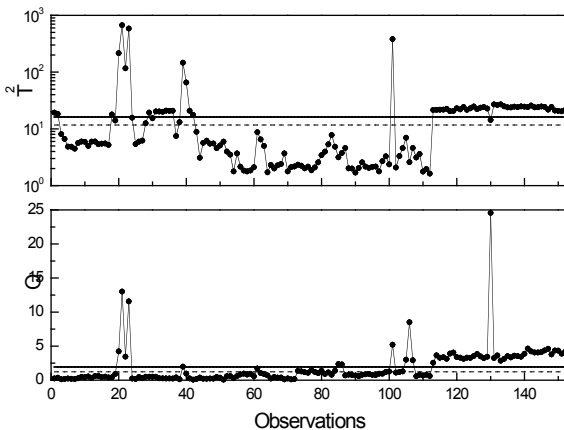

Fig. 3: Comparisons of the proposed method and PLS.

Fig. 4: Statistic Q and T^2 for the test dataset.
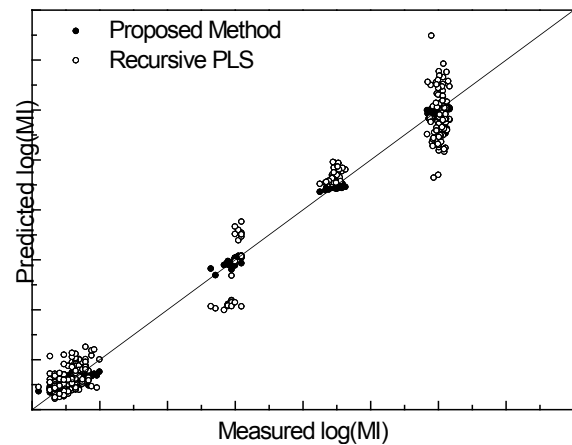

Fig. 5: Comparisons of the proposed method and RPLS.

5. Conclusions

In this study, the collinearity of process inputs is eliminated through the PCA; meanwhile outliers of input data can be filtered out. The nonlinear system is split into a collection of local linear subsystems that were clustered from the compressed data. Due to the time-varying nature of the real plant, the PCA subspace has to be reconstructed in order to cover the data of new events. The Bayesian classification and the TS model are updated to the new subspace; and the consequence functions are corrected with additional information. The proposed method is applied to the MI prediction of a polyethylene plant. Results show that nonlinearity and time varying characteristics of the polyethylene process plant can be dealt with effectively.

6. References

[1] S. Medasani, and R. Krishnapuram, "Categorization of image databases for efficient retrieval using robust mixture decomposition." *Computer Vision and Image Understanding*, pp. 216-235, 2001.

[2] J. E. Jackson, "A User's Guide To Principal Components", Wiley, New York. 1991.

[3] T. Takagi, and M. Sugeno, "Fuzzy identification of systems and its application to modeling and control." *IEEE Trans. Systems, Man and Cybernetics*, pp. 116-132, 1985.

[4] L. Ljung, "System Identification Theory For The User", Prentice-Hall Inc., New Jersey. 1999.

[5] J. Liu, "Process Monitoring Using Bayesian Classification on PCA Subspace.", *Ind. Eng. Chem. Res.*, pp. 7815-7825, 2004.

[6] S. J. Qin, "Recursive PLS algorithms for adaptive data modeling." *Comp. Chem. Engng.*, pp. 503-514, 1998.