

Development of a Multi-Model Strategy Based Soft Sensor Using Gaussian Process Regression and Principal Component Analysis in Fermentation Processes

Congli Mei*, Yao Chen, Haiyang Zhang, Xu Chen, Guohai Liu

School of electrical and information engineering, Jiangsu University, Zhenjiang 212013, China
 clmei@ujs.edu.cn.

In fermentation processes, single model based soft sensors cannot guarantee prediction performance owing to process characteristics of non-linearity, shifting operating modes, dynamics and uncertainty. In this paper, a novel multi-model based modeling method using Gaussian process regression (GPR) and principal component analysis (PCA) was proposed to construct a soft sensor for biomass concentration estimation in fermentation processes. In the method, principal components (PCs) extracted from original process data are firstly used to build GPR based sub-models. Then, to obtain final predictions, posteriori probabilities of the GPR based sub-models are used to combine outputs of sub-models. The proposed soft sensor was validated on simulation data of a Penicillin fermentation process. For comparisons, several other soft sensor models, e.g. GPR, back-propagation neural network (BP-NN) and least square support vector machine (LSSVM), were also studied. Results show that the proposed soft sensor has better prediction accuracy and smaller confidence intervals.

1. Introduction

In fermentation processes, it is difficult to measure some key variables, e.g. biomass concentration, in real time by existing online analysis sensors. Therefore, advanced control algorithms always cannot be used for process control and optimization. To overcome the problem, soft sensor technology was proposed as a good alternative. In the study of Kadlec et al. (2009), many soft sensor methods have been developed for estimation and prediction of key variables by using other easy-to-measure process variables. Nowadays, soft sensor technology has become a hot research area in the field of process control.

Two kinds of soft sensors, i.e. model-driven soft sensors and data-driven soft sensors, are widely used in bioprocesses. For the difficulty in constructing exact mathematical models of complex biological fermentation processes, most of soft sensors are of data-driven models, such as artificial neural network (ANN) (Liu et al., 2014) and support vector machine (SVM) (Liu et al., 2010). However, these conventional methods cannot give accuracy or confidence intervals of predictions, which limits applications of soft sensors in practical processes. In other words, conventional soft sensors cannot guarantee reliability of predictions. Recently, Gaussian process regression (GPR) was introduced to the field of soft sensors in Mei et al. (2015), which is a relatively new kind of machine learning methods based on probabilistic kernel function as an alternative approach to artificial neural network approach. This method assumes that the input-output relations are in concordance with the prior Gaussian distribution and obtains prediction results by model parameters which are determined by the maximum likelihood, while simultaneously the uncertainty of outputs are given. Compared with ANN and SVM, GPR has the advantages of integrating priori probability with less parameters and the measurement uncertainty of the outputs. According to Williams and Rasmussen (2006), it has a good prospect in soft sensor modeling.

Fermentation processes usually show characteristics of strong nonlinearity and multiphase/multimode. It results in poor prediction performance of soft sensors by using single regression models (Yuan et al., 2014). Recently, local learning methods have become popular in the machine learning area, e.g. JIT learning (Saporo, 2014), multi-model learning (Yu, 2012), and ensemble learning (Ge and Song, 2014). It was

demonstrated that the ensemble local model has better performance than a single regression model. JIT learning can cope with process nonlinearity by building local models repeatedly during online operation. However, it encounters heavy online computational load and difficulties in defining appropriate similarity measures. An alternative local learning method is the multi-model approach, where a model library is defined as a collection of local models. Then the output variable can be predicted using the corresponding local model relevant to the query state. However, process division information on process patterns and process stages is not often available for constructing local models. This problem can be solved using ensemble learning methods, in which local models are first trained and then combined to provide an overall output estimate. The keys to the success of ensemble learning are diversity and the ensemble strategy, which are still under research (Soares et al., 2011).

In this paper, we proposed a novel multi-model soft sensor method using GPR and PCA based on the basic idea of local learning. In the method, multi-scale features on variation, i.e. PCs sorted by the order of primary and secondary, are firstly extracted from process data by PCA. Then, GPR is used to construct sub-models in the main directions of the variation by using PCs. Finally, posteriori probabilities of the GPR based sub-models are used to design weights to combine sub-models. we give comparisons of different soft sensors in a simulated Penicillin fermentation process.

2. Gaussian process regression

In this section, GPR is briefly introduced (Williams and Rasmussen, 2006). For a regression problem, it can be assumed that n -dimensional input vector \mathbf{x} and observed target value y are generated according the following mechanism,

$$y = f(\mathbf{x}) + \varepsilon, \varepsilon \sim N(0, \sigma_n^2) \quad (1)$$

where ε is the noise, σ_n^2 is the variance of the nose, and $N(\cdot)$ represents Gaussian distribution. Further, the prior distribution of the observed target value y is got, which has been added the noise.

$$y \sim N(0, K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathbf{I}) \quad (2)$$

where $K(\mathbf{X}, \mathbf{X}) = (k(\mathbf{x}_i, \mathbf{x}_j))$ is the $n \times n$ dimensional symmetric and positive definite covariance matrix, $k(\mathbf{x}_i, \mathbf{x}_j)$ is used to measure the correlation between \mathbf{x}_i and \mathbf{x}_j , \mathbf{I} is the $n \times n$ dimensional identity matrix. Then we can write the joint distribution of the observed target values and the function values at the test point under the prior as:

$$\begin{bmatrix} y \\ f_* \end{bmatrix} \sim N \left(0, \begin{pmatrix} K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathbf{I} & K(\mathbf{X}, \mathbf{x}_*) \\ K(\mathbf{x}_*, \mathbf{X}) & k(\mathbf{x}_*, \mathbf{x}_*) \end{pmatrix} \right) \quad (3)$$

where f_* is the predicted values at the test point, $K(\mathbf{X}, \mathbf{x}_*)$ is the $n \times 1$ covariance matrix between training and test data set, $K(\mathbf{x}_*, \mathbf{X})$ is the transposition of $K(\mathbf{X}, \mathbf{x}_*)$, $k(\mathbf{x}_*, \mathbf{x}_*)$ is the covariance function evaluated at test data \mathbf{x}_* and itself. Then, the predicted distribution of GPR model can be obtained as

$$f_* / \mathbf{X}, y, \mathbf{x}_* \sim N(\bar{f}_*, \text{cov}(f_*)) \quad (4)$$

Where

$$\bar{f}_* = K(\mathbf{x}_*, \mathbf{X}) [K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathbf{I}]^{-1} y \quad (5)$$

$$\text{cov}(f_*) = k(\mathbf{x}_*, \mathbf{x}_*) - K(\mathbf{x}_*, \mathbf{X}) [K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathbf{I}]^{-1} K(\mathbf{X}, \mathbf{x}_*) \quad (6)$$

To establish the GPR model, we must firstly determine the covariance function. In consideration of the continuity and the requirement of the covariance function, the radial basis function is a common choice and given by

$$k(\mathbf{x}_i, \mathbf{x}_j) = v_0 \exp \left[-\frac{1}{2} \sum_{t=1}^n w_t (\mathbf{x}_i^t - \mathbf{x}_j^t)^2 \right] + v_1 \delta_{ij} \quad (7)$$

where $\Theta = [w_1, w_2, \dots, w_n, v_0, v_1]^T$ are the hyperparameters of the covariance function, and adopting the maximum likelihood method to obtain the optimal value. v_0 represents an overall measure of the prior knowledge, in other words, it controls the typical amplitude of covariation. v_1 represents the variation of noise and it follows Gaussian distribution. δ_{ij} is Kronecker delta. w_t can be seen as the weight of the t th dimension. GPR is in the framework of Bayesian rule described by Williams and Rasmussen (2006). The marginal likelihood function of Bayesian formalism, typically negative log marginal likelihood function, was chosen as an optimization objective function of hyperparameters in this study.

3. Proposed multi-model strategy based soft sensor

PCA is widely used to extract data features in processes. The basic idea is to determine the primary and secondary change directions, i.e. PCs, according to the variance of the data. The main PCs are obtained by the order of the primary and secondary, which are independent of each other. In this work, we proposed a multi-model based modeling approach using former k PCs. In the method, the i th sub-model is constructed on the i th PC and output. That is a multi-scale modeling method because PCs are sorted in descending order. It means that the modeling accuracy of the final ensemble model depends on the number of PCs used in modeling. More PCs used in modeling means better accuracy of the multi-model based soft sensor. The final ensemble model can be written as Eq(8).

$$\bar{f}(\mathbf{x}) = \sum_{i=1}^k w_i \bar{f}_i(\mathbf{x}) \quad (8)$$

Where $w_i, i = 1, \dots, k$ is weighting factor, and k refers to the number of sub-models. $\bar{f}_i(\mathbf{x})$ is the estimate of the i th sub-model. For normalization Eq(9) should be satisfied:

$$\sum_{i=1}^n w_i = 1 \quad (9)$$

In this work, we construct w_i using variances of estimates because variances can reflect the stability of models. The weight of the k sub-model can be designed as

$$w_i = \frac{1/\delta_i^2}{\sum_{j=1}^K 1/\delta_j^2} = \frac{1}{\delta_i^2 \sum_{j=1}^K 1/\delta_j^2} \quad (10)$$

Where $\delta_i^2 = \text{cov}(f_i)_i, i = 1, \dots, k$, f_i is the predicted values and k represents the number of sub-models. For uncorrelated PCs, the variance of the ensemble output can be written as

$$\delta_{final} = \sqrt{\sum_{i=1}^k w_i^2 \delta_i^2} \quad (11)$$

The modelling procedures of the proposed soft sensor are depicted as follows:

Step 1: Collect and standardize training data and test data.

Step 2: Use PCA method to obtain PCs of training data. Determine k value, the number of selected main PCs, according to 85 % cumulative contribution rate.

Step 3: Construct local models using GPR to reflect correlation between PCs and output respectively.

Step 4: Calculate predicted means and posterior probabilities corresponding to every test data with all GPR based sub-models.

Step 5: Calculate the weights by the posterior probabilities with respect to each sub-model using Eq(10)

Step 6: Combine predictions of local models as the final predictions by Eq(8).

4. Case study

This case is based on the simulation software of penicillin fermentation process (Pensim) (Birol et al. 2002) to produce data, and to carry out the research work of soft sensor modeling. The simulator contains a fermenter where the biological reaction takes place. The detailed description of the process is given in Birol et al. (2002). For different demands, the simulator provides several settings including the controller, process duration,

sampling rates, etc. In this study, we set the sampling intervals is 1 h and the fermentation period is set to 400 hours. Then, the Pensim software randomly run 10 batches under normal conditions. 7 batch data are used as training data sets, and the remaining 3 batch data are used as test data sets. There are totally 16 measure and variables in the simulation plant. The penicillin concentration is difficult to measure online and chosen as output variable. 11 variables among them are highly related to it are selected as input variables (Yuan et al., 2014).

After PCA on training data, former 5 PCs, the cumulative contribution rate of which exceeds 85 %, are selected to construct sub-models. For performance comparison, the GPR based soft sensor, the BPNN based soft sensor and the LSSVM based soft sensor were also studied. In the BPNN based soft sensor, the number of input-layer nodes is set to 5 and the number of hidden-layer nodes is set to 8. In the LSSVM based soft sensor, the optimized regularization parameter is equal to 613.96 and optimal kernel parameter is equal to 0.21.

Figure 1, 2 and 3 give predictions of different soft sensors, i.e. the proposed soft sensor, the GPR base soft sensor, the BPNN based soft sensor and the LSSVM based soft sensor, on 3 test batches. For the proposed soft sensor and the GPR based soft sensor, the predictions are expressed as mean with 2*std (std, standard deviation) error bars (dotted lines). The regions between the two dotted lines depict the confidence intervals. Compared with the other 3 soft sensors, it can be observed that predictions of the proposed soft sensor follow the actual values more closely. Also from Figure 1, 2 and 3, it can be found that the proposed soft sensor has smaller confidence intervals compared to the GPR based soft sensor. It reflects the modelling ability of multi-model approach integrated in the proposed soft sensor.

For comparison in detail, the root-mean-square error (RMSE) criterion is adopted. Table 1 gives quantitative comparisons of different soft sensors in the PFP. It can be observed that the proposed soft sensor outperforms other soft sensors because the RMSE values of the former are smaller than those of the others. Also from Table 1, it can be found that the average confidence intervals of the proposed soft sensor are smaller than those of the GPR based soft sensors. It means that the proposed soft sensor has higher reliability (smaller uncertainty) than that of the GPR based soft sensor.

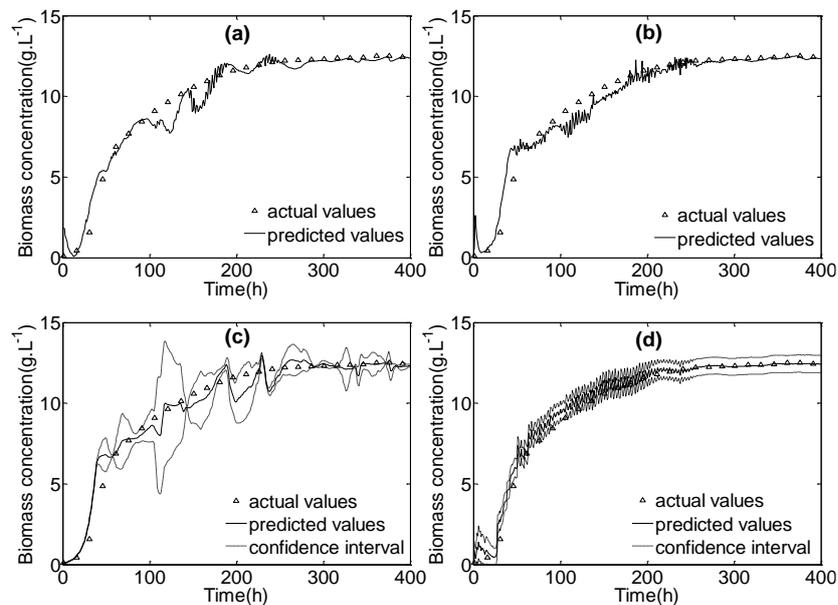


Figure 1: Predictions on the first test batch with different soft sensors in the PFP. (a) The BPNN based soft sensor; (b) The LSSVM based soft sensor; (c) The GPR based soft sensor; (d) The proposed soft sensor

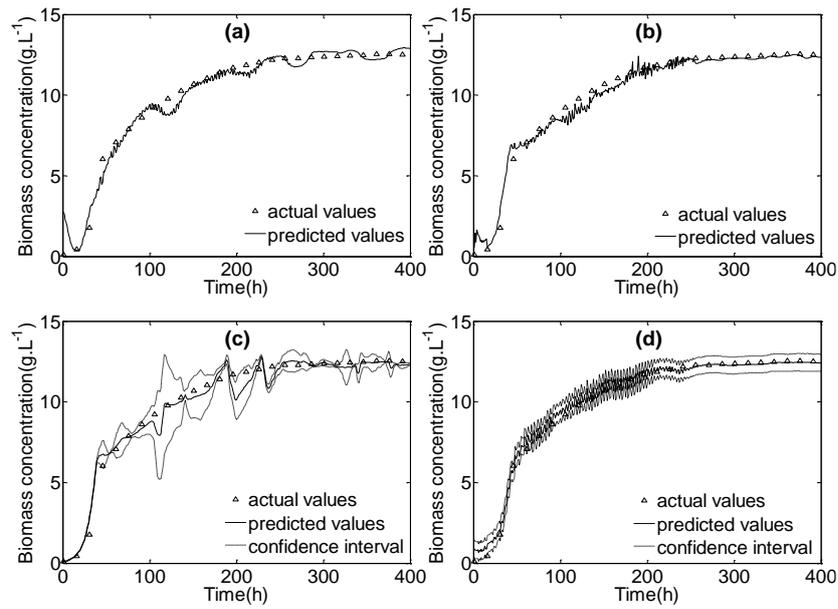


Figure 2: Predictions on second test batch with different soft sensors in the PFP. (a) The BPNN based soft sensor; (b) The LSSVM based soft sensor; (c) The GPR based soft sensor; (d) The proposed soft sensor.

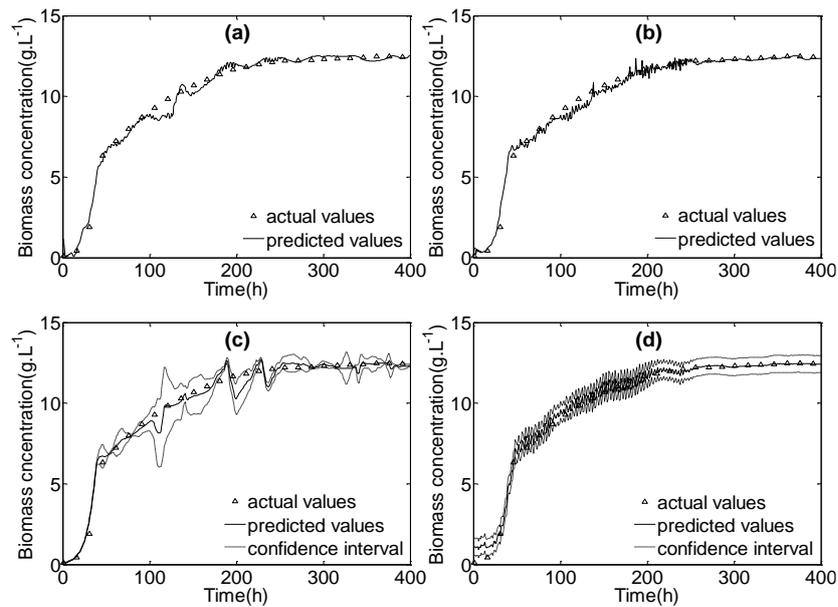


Figure 3: Predictions on the third test batch with different soft sensors in the PFP. (a) The BPNN based soft sensor; (b) The LSSVM based soft sensor; (c) The GPR based soft sensor; (d) The proposed soft sensor.

Table 1: Comparisons of different soft sensors in the PFP

Method	RMSE			Average confidence intervals		
	Batch1	Batch2	Batch3	Batch1	Batch2	Batch3
BPNN	0.7244	0.5538	0.4472	-	-	-
LSSVM	0.6393	0.4370	0.3655	-	-	-
GPR	0.7232	0.5947	0.4995	1.5608	1.1880	0.9604
proposed method	0.4621	0.3050	0.2939	1.0868	1.0592	0.9580

5. Conclusions

Soft sensor modelling is essentially data-driven modelling. Single model based soft sensors and multi-model based soft sensor are both common in fermentation processes. Especially, multi-model based models have advantages in modelling complex fermentation processes with characteristics of multiphase and multistage. However, it is difficult to determine the structures and parameters of a multi-model, e.g. the number and initial parameters of sub-models. In this paper, a novel multi-model soft sensor using GPR and PCA in fermentation processes was proposed. In the method, PCA is firstly used to extract features of the data of fermentation processes for building sub-model. Then, GPR based sub-models are constructed in the directions of main PCs. Finally, probabilistic outputs of sub-models are used to design weights to ensemble sub-models. Unlike conventional multi-model modelling approach, data features are used for modelling in the proposed method. Validations on process data from a Penicillin fermentation simulation case show that the proposed soft sensor outperforms the GPR based soft sensor, the BPNN based soft sensor and the LSSVM based soft sensor and has great potential in soft sensor modelling for key variables in fermentation processes. In further study, we will prove and improve the proposed method theoretically.

Acknowledgments

The authors gratefully acknowledge the financial support provided by Natural Science Foundation of Jiangsu Province of China (grant no. BK20130531, BK20140538) and the Priority Academic Program Development of Jiangsu Higher Education Institutions (grant no. PAPD 6), the Graduate practical innovation Foundation of Jiangsu province (Grant No. SJLX16_0441).

References

- Birol G., Undey C., Cinar A., 2002, A modular simulation package for fed-batch fermentation: penicillin production, *Computers & Chemical Engineering*, 26, 1553-1565.
- Ge Z., Song Z., 2014, Ensemble independent component regression models and soft sensing application, *Chemometrics & Intelligent Laboratory Systems*, 130(130), 115-122.
- Kadlec P., Gabrys B., Strandt S., 2009, Data-driven soft sensors in the process industry, *Computers & Chemical Engineering*, 33(4), 795-814.
- Liu G., Yu S., Mei C., Ding Y., 2014, A novel soft sensor model based on artificial neural network in the fermentation process, *African Journal of Biotechnology*, 10(85), 19780-19787.
- Liu G., Zhou D., Xu H., Mei C., 2010, Model optimization of svm for a fermentation soft sensor, *Expert Systems with Applications*, 37(4), 2708-2713.
- Mei C., Yang M., Shu D., Jiang H., Liu G., Liao Z., 2015, Soft sensor based on Gaussian process regression and its application in erythromycin fermentation process, *Chemical Industry & Chemical Engineering Quarterly*, 22 (00), 26-26.
- Saporo A., 2014, State of the art in the development of adaptive soft sensors based on just-in-time models, *Procedia Chemistry*, 9, 226-234.
- Soares S., Araújo R., Sousa P., Souza F., 2011, Design and application of Soft Sensor using Ensemble Methods, *Emerging Technologies & Factory Automation (Vol.19, pp.1-8)*, IEEE.
- Williams C.K., Rasmussen C. E., 2006, Gaussian processes for machine learning, 14 (481), 69-106.
- Yu J., 2012, Multiway gaussian mixture model based adaptive kernel partial least squares regression method for soft sensor estimation and reliable quality prediction of nonlinear multiphase batch processes, *Industrial & Engineering Chemistry Research*, 51(40), 13227-13237.
- Yuan X., Ge Z., Song Z., 2014, Soft sensor model development in multiphase/multimode processes based on gaussian mixture regression, *Chemometrics & Intelligent Laboratory Systems*, 138, 97-109.