Quality prediction for polypropylene production process based on CLGPR model

Zhiqiang Ge^{1*}, Tao Chen², Zhihuan Song¹

¹*State Key Laboratory of Industrial Control Technology, Institute of Industrial Process Control, Zhejiang University, Hangzhou 310027, Zhejiang, China*

²*Division of Civil, Chemical and Environmental Engineering, University of Surrey, Guildford GU2 7XH, UK*

Abstract

Online measurement of the melt index is typically unavailable in industrial polypropylene production processes, soft sensing models are therefore required for estimation and prediction of this important quality variable. Polymerization is a highly nonlinear process, which usually produces products with multiple quality grades. In the present paper, an effective soft sensor, named Combined Local Gaussian Process Regression (CLGPR), is developed for prediction of the melt index. While the introduced Gaussian process regression model can well address the high nonlinearity of the process data in each operation mode, the local modeling structure can be effectively extended to processes with multiple operation modes. Feasibility and efficiency of the proposed soft sensor are demonstrated through the application to an industrial polypropylene production process.

Keywords: Melt index; Quality prediction; Gaussian process regression; Principal component analysis; Multiple operation modes.

1. Introduction

As an important material, polypropylene has been widely used in many different fields including chemical, optical and medical sectors. The manufacture of polypropylene is a billion-dollar business, which has seen about 5% annual growth rate in consumption in recent years (Shi et al., 2006). The quality of polypropylene is conventionally assessed by the melt index in practical industrial processes (Kiparissides et al., 1993). However, due to the challenged engineering activity and the complexity of the process, the melt index is usually obtained through an offline analytical procedure, which may take up to two hours. Therefore, this will cause a time delay to the quality control system, since the process is without any quality indicator during this period of time. An alternative is to install an online analyzer, such as those based on near infrared spectroscopy or ultrasound, for measuring the melt index (Coates et al., 2003). However, current online analyzers are very expensive and require considerable maintenance effort, resulting in limited adoption in practical plants.

Recently, with the wide utilization of the distributed control system (DCS) in industrial processes, a large amount of process data can be routinely recorded. Among these recorded process data, some process variables that are highly correlated with the process quality can be used to estimate and predict the quality variable. The method of inferring difficult-to-measure

* To whom all correspondence should be addressed.

Email: zqge@iipc.zju.edu.cn; gezhiqiang@gmail.com

quantities by easy-to-measure variables is known as soft sensor, inferential sensor or virtual sensor (Kadlec et al., 2009). Particularly, in the polypropylene production process, it was shown that the melt index can be predicted by some related process variables that significantly reflect the process condition and the product quality, such as hydrogen concentration of the reactor, propylene feed rate, reaction temperature, among others (Shi et al., 2006; Kiparissides et al., 1993; Coates et al., 2003; Zhang et al., 1998; Ohshima & Tanigaki, 2000; Liu, 2007).

To date, different data-based soft sensors have been developed for quality prediction purpose, including principal component regression (PCR) and partial least squares (PLS) for linear processes, artificial neural network (ANN) and support vector machine (SVM) for nonlinear processes and etc (Kadlec et al., 2009; Gonzaga et al., 2009; Kano et al., 2008; Yang & Gao, 2006; Gao & Ren, 2010). The polypropylene production process is a well-known highly nonlinear process as evidenced by mechanistic analysis of the reactions and plants (Liu, 2007). Therefore, nonlinear soft sensors should be considered. Besides, practical industrial processes are always contaminated by noises, and thus those measured process variables are inherently random variables. In this case, it is more appropriate to make statistical inference and prediction decisions based on probabilistic models. Unfortunately, most traditional soft sensor modeling methods were constructed in a deterministic manner.

Recently, a new probabilistic modeling method namely Gaussian process regression (GPR) has gained much attention in both statistical and engineering areas, which is initially proposed by (O'Hagan, 1978). It is demonstrated that a large class of ANN based Bayesian regression models will finally converge to an approximate Gaussian process. Therefore, GPR model has been considered as an alternative method for nonlinear system modeling. Along last several decades, lots of comparative studies have shown that the GPR model performs better than other nonlinear modeling approaches (Csato & Opper, 2002; Chu & Ghahramani, 2005; Rasmussen & Williams, 2006; Likar & Kocijan, 2007; Shi et al., 2007; Chen & Ren, 2009; Tang et al., 2010). Another advantage of the GPR model is due to its probabilistic model structure, which can successfully incorporate the noise information and provide an uncertainty prediction result for the process. To our best knowledge, GPR has rarely been reported for soft sensor modeling in the process system engineering area. Due to its efficiency for nonlinear system modeling, the GPR model is employed for soft sensor construction in the present paper. For dimension compression of process data, and also to address high correlations between different variables, the traditional PCA method can be initially performed, which means the GPR model will be constructed upon score variables of the PCA model. In this sense, the PCA-GPR method can be considered as a probabilistic form of nonlinear PCA based regression model.

Beside of the nonlinear behavior, the polypropylene production process also exhibits multiple production grades, which is probably driven by different market requirements (Liu, 2007). Therefore, this process always has multiple operation modes. A straightforward idea is to build multiple GPR models under different operation conditions. However, an important issue is how to select the appropriate GPR model for the current data sample. If the plant engineer knows the operation mode of the process, he/she can simply select the corresponding local GPR model for prediction. However, automatic process operation with minimal human intervention is usually desired in modern processing plants. Besides, if the current data belong

to the transition between different operation modes, automatic weighting of multiple local models should also be considered. To address this issue, different forms of mixture Gaussian process models have been developed, such as infinite mixtures of Gaussian process experts and hierarchical Gaussian process mixture model (Rasmussen & Ghahramani, 2002; Shi et al., 2005; Ou & Martin, 2008). However, most of these mixture Gaussian process models involve computationally expensive Monte Carlo methods, such as Gibbs sampling and hybrid Markov chain Monte Carlo (MCMC) sampling. The algorithmic complexity of the mixture Gaussian process models is the major reason to limit their application to large datasets and/or high dimensional processes. Besides, when some new operation mode is identified, the traditional mixture Gaussian process models need to be re-trained, demanding considerable computational resource and maintenance effort. Other similar techniques for multimode modeling include the Gaussian mixture model (GMM) approach and the fuzzy modeling method (Choi et al., 2004; Yu & Qin, 2008; Yu & Qin, 2009; Rong et al., 2006; Huang & Hahn, 2009). The GMM method can also give a probabilistic model structure for different operation modes. However, most of GMM approaches are limited in the linear case. Although the fuzzy modeling method can address the nonlinear behavior of the process data, its performance depends on the modeling structure of nonlinear models. Besides, the fuzzy modeling method may have some user-defined parameters, which are difficult to determine.

In this paper, we intend to build multiple local PCA-GPR model based soft sensors for different operation modes in the first step. Then, a new soft assignment and combination strategy is proposed for result fusion in different operation modes for the new data sample. Compared to traditional mixture Gaussian process models, the implementation of our method is much easier, thus, it is more useful for practical application. It is noted that this new assignment and combination strategy can perform automatically without requirement of additional information of the process. Besides, when some new operation mode is available for modeling, we can simply build a new local PCA-GPR model for this new operation mode, and put it into the model pool. Therefore, compared to the traditional mixture Gaussian process model, model updating according to the change of process conditions is much easier in our proposed method.

The rest of this paper is organized as follows. In section 2, some preliminaries of the traditional PCA and GPR models are introduced, which is followed by the detailed description of the proposed soft sensor for quality prediction in the next section. In section 4, an industrial application case study of the polypropylene production process is provided performance evaluation of the proposed method. Finally, some conclusions are made.

2. Some preliminaries

2.1. Principal component analysis (PCA)

Given a dataset $\mathbf{X} \in R^{n \times m}$, where m is the number of process variables, and n is the sample number for each variable, PCA is carried out upon the covariance matrix of $\mathbf{X}$. Traditionally, the singular value decomposition (SVD) method can be employed for construction of the PCA model. Suppose k principal components have been selected in the PCA model, $\mathbf{X}$ can be decomposed as (Qin, 2003)

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T + \tilde{\mathbf{T}}\tilde{\mathbf{P}}^T = \mathbf{T}\mathbf{P}^T + \mathbf{E} \quad (1)$$

where $\mathbf{T} \in R^{n \times k}$ and $\tilde{\mathbf{T}} \in R^{n \times (m-k)}$ are score matrices in the principal component subspace (PCS) and the residual subspace (RS), $\mathbf{P} \in R^{m \times k}$ and $\tilde{\mathbf{P}} \in R^{m \times (m-k)}$ correspond to loading matrices in PCS and RS. $\mathbf{E} \in R^{n \times m}$ is the residual matrix.

2.2. Gaussian process regression (GPR)

Consider a training dataset $\mathbf{X} \in R^{n \times m}$ and $\mathbf{y} \in R^n$, where $\mathbf{X} = \{\mathbf{x}_i \in R^m\}_{i=1,2,\dots,n}$ is the input data samples with m dimensions, and $\mathbf{y} = \{y_i \in R\}_{i=1,2,\dots,n}$ is the output data sample, the aim of the regression model is to build a functional relationship between $\mathbf{x}$ and y . Particularly, a Gaussian process regression model is defined such that the regression function $y = f(\mathbf{x})$ has a Gaussian prior distribution with zero mean, which is given as

$$\mathbf{y} = [f(\mathbf{x}_1), f(\mathbf{x}_2), \dots, f(\mathbf{x}_n)] \sim GP(0, \mathbf{C}) \quad (2)$$

where $\mathbf{C}$ is an $n \times n$ covariance matrix, with its ij -th element defined as $\mathbf{C}_{ij} = \mathbf{C}(\mathbf{x}_i, \mathbf{x}_j)$. To calculate the GPR model, different covariance functions can be selected. A commonly used covariance function is the squared-exponential covariance function, which is given as (Rasmussen & Williams, 2006)

$$\mathbf{C}(\mathbf{x}_i, \mathbf{x}_j) = \sigma_f^2 \exp\left\{-\frac{1}{2}(\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{M}(\mathbf{x}_i - \mathbf{x}_j)\right\} + \delta_{ij} \sigma_n^2 \quad (3)$$

where $\delta_{ij} = 1$ if $i = j$, otherwise $\delta_{ij} = 0$, $\mathbf{M} = \ell^{-2} \mathbf{I}$, $\mathbf{I}$ is an identify matrix with appropriate dimension. The exponential term is similar to the form of a radial basis function, with its length-scale ℓ , and the terms σ_f^2 and σ_n^2 correspond to signal variance and noise variance. To identify the GPR model, or precisely, to determine the value of the hyperparameter set $\theta = (\ell, \sigma_f, \sigma_n)$, the following log-likelihood function can be maximized

$$L = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log(\det(\mathbf{C})) - \frac{1}{2} \mathbf{y}^T \mathbf{C}^{-1} \mathbf{y} \quad (4)$$

As a result, the hyperparameter set $\theta = (\ell, \sigma_f, \sigma_n)$ can be obtained. An alternative way to determine the value of the hyperparameter is to use the sampling methods such as Markov chain Monte Carlo (MCMC) sampling and Gibbs sampling, which generate samples for approximation of the posterior distribution of the hyperparameter.

For a new input data sample $\mathbf{x}_{new}$, the predictive distribution of its corresponding output

y_{new} is also Gaussian, the mean and variance values of which are calculated as follows

$$\bar{y}_{new} = \mathbf{k}^T(\mathbf{x}_{new}) \mathbf{C}^{-1} \mathbf{y} \quad (5)$$

$$\sigma_{y_{new}}^2 = C(\mathbf{x}_{new}, \mathbf{x}_{new}) - \mathbf{k}^T(\mathbf{x}_{new}) \mathbf{C}^{-1} \mathbf{k}(\mathbf{x}_{new}) \quad (6)$$

where $\mathbf{k}(\mathbf{x}_{new}) = [C(\mathbf{x}_{new}, \mathbf{x}_1), C(\mathbf{x}_{new}, \mathbf{x}_2), \dots, C(\mathbf{x}_{new}, \mathbf{x}_n)]^T$.

3. Quality prediction based on local GPR model

In this section, the detailed description of the proposed method is provided. First, multiple local GPR model based soft sensors are construction in different operation modes, depending on which the online quality prediction strategy is formulated. Finally, some discussions are made.

3.1. Construction of local GPR model based soft sensor

Suppose the while process consist of Q operation modes, we then represent the dataset as $\mathbf{X} = [\mathbf{X}_1^T, \mathbf{X}_2^T, \dots, \mathbf{X}_Q^T]^T \in R^{n \times m}$ and $\mathbf{y} = [\mathbf{y}_1^T, \mathbf{y}_2^T, \dots, \mathbf{y}_Q^T]^T$, where $\mathbf{X}_q = \{\mathbf{x}_i \in R^m\}_{i=1,2,\dots,n_q}$ is the input dataset in the q -th operation modes, with its data sample number as n_q , $\mathbf{y}_q = \{y_i \in R\}_{i=1,2,\dots,n_q}$ is the corresponding output dataset of the q -th operation modes, and $\sum_{q=1}^Q n_q = n$. Before the implementation of the GPR modeling procedure, an initial PCA pre-processing step can be used for reduce the dimensionality of the input variables. Therefore, a total of Q PCA models can be built, which are given as

$$\begin{aligned} \mathbf{X}_1 &= \mathbf{T}_1 \mathbf{P}_1^T + \mathbf{E}_1 \\ \mathbf{X}_2 &= \mathbf{T}_2 \mathbf{P}_2^T + \mathbf{E}_2 \\ &\vdots \\ \mathbf{X}_Q &= \mathbf{T}_Q \mathbf{P}_Q^T + \mathbf{E}_Q \end{aligned} \quad (7)$$

where $\mathbf{T}_q \in R^{n_q \times k_q}$ and $\mathbf{P}_q \in R^{m \times k_q}$ are score and loading matrices of the q -th operation mode, $\mathbf{E}_q \in R^{n_q \times m}$ is the residual matrix, k_q is the selected number of principal components in the q -th operation mode, which can be easily determined by the cumulative percentage variance method. After the PCA information extraction step, the dimensionality of the input variables can be greatly reduced. In the following GPR modeling step, we can only focus on the score matrices $\mathbf{T}_q \in R^{n_q \times k_q}$, $q=1,2,\dots,Q$ in different operation modes.

Following the modeling procedure of the GPR algorithm, the regression model between the score matrix $\mathbf{T}_q = \{\mathbf{t}_{i,q} \in R^{k_q}\}_{i=1,2,\dots,n_q}$ and the output data vector $\mathbf{y}_q = \{y_{i,q} \in R\}_{i=1,2,\dots,n_q}$

can be formulated as follows

$$\mathbf{y}_q = [y_{1,q}, y_{2,q}, \dots, y_{n_q,q}] = [f^q(\mathbf{t}_{1,q}), f^q(\mathbf{t}_{2,q}), \dots, f^q(\mathbf{t}_{n_q,q})] \sim GP(0, \mathbf{C}_q) \quad (8)$$

where $q=1, 2, \dots, Q$, $\mathbf{C}_q$ is the covariance matrix of the q -th GPR model. The general form of the covariance matrices for different operation modes can be selected as eq. (3). However, the hyperparameter values for different operation modes $\theta_q = (\ell_q, \sigma_{f,q}, \sigma_{n,q})$, $q=1, 2, \dots, Q$ are differentiated from each other, depending on the GPR model optimization.

3.2. Online quality prediction through soft combination strategy

After the local GPR model has been constructed in each operation mode, they can be used for online quality prediction of the new input data sample $\mathbf{x}_{new}$. However, an important issue is how to select appropriate PCA and GPR models for dimensionality reduction and quality prediction. Without additional process information, we do not know which operation mode the new data sample $\mathbf{x}_{new}$ belongs to. In the present paper, we intend to propose a new method to select the GPR model automatically, which can softly assign the new data sample $\mathbf{x}_{new}$ to different operation modes with corresponding probabilities.

Based on the PCA model, a normal operation region can be built for each operation mode. Particularly, the T^2 statistic which is conventionally used for process monitoring purpose can be constructed as (Qin, 2003)

$$T_{i,q}^2 = \mathbf{t}_{i,q}^T \mathbf{\Lambda}_q^{-1} \mathbf{t}_{i,q} \quad (9)$$

where $q=1, 2, \dots, Q$, $i=1, 2, \dots, n_q$, $\mathbf{\Lambda}_q = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_{k_q}\}$ is a diagonal matrix with its elements as the eigenvalues of the corresponding PCA model. To determine the operation region of the q -th grade, the control limit of the T^2 statistic can be calculated as

$$T_{\text{lim},q}^2 = \frac{k_q(n_q-1)}{n_q-k_q} F_{k_q, (n_q-k_q), \alpha} \quad (10)$$

where $F_{k_q, (n_q-k_q), \alpha}$ represents F-distribution, with its two parameters k_q and n_q-k_q , α is significance level. Therefore, by checking if the $T_{i,q}^2 \leq T_{\text{lim},q}^2$ holds, we can easily determine the operation mode of the data sample.

For the new input data sample $\mathbf{x}_{new}$, we first calculate the score vector by each PCA model, and then the value of the T^2 statistic can be determined in each operation mode, thus

$$\mathbf{t}_{\text{new},q} = \mathbf{P}_q^T \mathbf{x}_{\text{new}}, q=1, 2, \dots, Q \quad (11)$$

$$T_{\text{new},q}^2 = \mathbf{t}_{\text{new},q}^T \mathbf{\Lambda}_q^{-1} \mathbf{t}_{\text{new},q} \quad (12)$$

To determine the probability of the new data sample $\mathbf{x}_{new}$ in each operation mode, a Bayesian inference can be incorporated, which is given as follows

$$P(q | \mathbf{x}_{new}) = \frac{P(q, \mathbf{x}_{new})}{P(\mathbf{x}_{new})} = \frac{P(\mathbf{x}_{new} | q)P(q)}{\sum_{j=1}^Q \{P(\mathbf{x}_{new} | j)P(j)\}} \quad (13)$$

where $q = 1, 2, \dots, Q$. To calculate the posterior probability value, two terms in the right side of eq. (13) should be defined, which are known as prior probability and conditional probability. Without any process or expert knowledge, the prior probability for each operation mode can be simply defined as

$$P(q) = n_q / n \quad (14)$$

The conditional probability can be defined based on the T^2 statistic, which is given as

$$P(\mathbf{x}_{new} | q) = \exp\left\{-\frac{T_{new,q}^2}{T_{lim,q}^2}\right\} \quad (15)$$

This is an exponential function, based on which the value of the conditional probability is restricted between 0 and 1. Although the distribution of the conditional probability does not exactly expect to be an exponential distribution, the exponential form of the function seems to be effective for modeling of such distribution. Therefore, eq. (13) becomes

$$P(q | \mathbf{x}_{new}) = \frac{n_q \exp\left\{-\frac{T_{new,q}^2}{T_{lim,q}^2}\right\}}{\sum_{j=1}^Q \{n_j \exp\left\{-\frac{T_{new,j}^2}{T_{lim,j}^2}\right\}\}} \quad (16)$$

After the probability of the new data sample $\mathbf{x}_{new}$ in each operation mode has been determined, the predictive distribution of its corresponding output $P(y_{new})$ in each operation mode can then be calculated, which is also Gaussian. The mean and variance values of the predictive distribution for $\mathbf{x}_{new}$ in different operation modes are given as follows

$$\begin{aligned} \bar{y}_{new,q} &= \mathbf{k}_q^T(\mathbf{t}_{new,q})\mathbf{C}_q^{-1}\mathbf{y}_q \\ \mathbf{k}_q(\mathbf{t}_{new,q}) &= [C_q(\mathbf{t}_{new,q}, \mathbf{t}_{1,q}), C_q(\mathbf{t}_{new,q}, \mathbf{t}_{2,q}), \dots, C_q(\mathbf{t}_{new,q}, \mathbf{t}_{n_q,q})]^T \\ \mathbf{y}_q &= [y_{1,q}, y_{2,q}, \dots, y_{n_q,q}]^T \end{aligned} \quad (17)$$

$$\sigma_{y_{new,q}}^2 = C_q(\mathbf{t}_{new,q}, \mathbf{t}_{new,q}) - \mathbf{k}_q^T(\mathbf{t}_{new,q})\mathbf{C}_q^{-1}\mathbf{k}_q(\mathbf{t}_{new,q}) \quad (18)$$

where $q = 1, 2, \dots, Q$.

Finally, the overall predictive distribution of the new data sample can be formulated

through the weighted combination strategy, given as

$$P(y_{new}) = \sum_{q=1}^Q P(y_{new} | q) P(q | \mathbf{x}_{new}) \quad (19)$$

The mean and variance values of the final predictive distribution can be calculated as

$$\bar{y}_{new} = \sum_{q=1}^Q \bar{y}_{new,q} P(q | \mathbf{x}_{new}) = \sum_{q=1}^Q \mathbf{k}_q^T(\mathbf{t}_{new,q}) \mathbf{C}_q^{-1} \mathbf{y}_q P(q | \mathbf{x}_{new}) \quad (20)$$

$$\sigma_{y_{new}}^2 = \sum_{q=1}^Q \sigma_{y_{new,q}}^2 P^2(q | \mathbf{x}_{new}) = \sum_{q=1}^Q \{ [C_q(\mathbf{t}_{new,q}, \mathbf{t}_{new,q}) - \mathbf{k}_q^T(\mathbf{t}_{new,q}) \mathbf{C}_q^{-1} \mathbf{k}_q(\mathbf{t}_{new,q})] P^2(q | \mathbf{x}_{new}) \} \quad (21)$$

It is noted that in eq. (21), we have assumed that different operation modes are independent from each other, thus the variance of the final prediction result is a simple weighted summation of the variances in different operation modes.

3.3. Discussions

So far, the local GPR model based soft sensor has been developed. Compared to existing soft sensors, such as PCR, PLS and SVM, the new soft sensor can not only address the nonlinear behavior of the process, but also it can provide a probabilistic prediction result for the quality variable. Besides, the new soft sensor is specially designed for quality prediction of processes which may have several different operation conditions. From an engineering standpoint, this new proposed method is easy for practical implementation, and the interpretation of the prediction result is also straightforward. While the single GPR model is only efficient for quality prediction in its specific region, the combined local GPR model can handle overlapping operation modes problem, which may exist in many industrial processes. By softly assigning the data sample to different operation modes with corresponding weights, a probabilistic prediction result can be generated. Different from the GMM approach which always limited in the linear case, the local GPR model can model the nonlinear relationship in each operation mode of the process. Therefore, the proposed method is more appropriate for multimode modeling in nonlinear processes.

Before the modeling procedure of the new soft sensor, we have assumed that the whole process dataset has been partitioned into sub-datasets according to different operation modes. However, this may be not available in practice, since the operation mode information is not always feasible. To address this problem, the clustering method can be employed, traditionally, such as the K-mean method, finite Gaussian mixture model based method, and etc. Compared to the K-mean method, the finite Gaussian mixture model is more appropriate for mode clustering, since it does not require the prior process knowledge on the total number of operation modes. Therefore, the finite Gaussian mixture model is used for mode clustering in the present work. Another important issue of the PCA-GPR modeling strategy is that the performance of the PCA information extraction step may be influenced by some outliers or disturbances. This can be solved by employing the robust PCA method, data screening and filtering, or data reconciliation methods. However, in the present work, it is assumed that any outlier or data disturbance has already been removed.

Based on the proposed soft assignment and combination strategy, the final quality prediction result can be generated automatically, which means we do not need to know the

exact mode information of the new input data sample. However, the mode information (posterior probability) of the new data sample can be obtained simultaneously within the new approach, which can be calculated through eq. (16). In our opinion, although the mode information is not required for quality prediction, it is important for mode localization and identification of the new process data, which may play a significant role in process understanding, monitoring, design and improvement.

Finally, it can be noted in eq. (21) that the predictive variance of the quality variable can be compressed after the soft combination step. Since the posterior probability of the new process data obeys $0 \leq P(q | \mathbf{x}_{new}) \leq 1, q = 1, 2, \dots, Q$, and $\sum_{q=1}^Q P(q | \mathbf{x}_{new}) = 1$, the following

result can be easily obtained

$$\sigma_{y_{new}}^2 = \sum_{q=1}^Q \sigma_{y_{new,q}}^2 P^2(q | \mathbf{x}_{new}) \leq [\sigma_{y_{new,q}}^2]_{\forall q=1,2,\dots,Q} \quad (22)$$

It can be seen that the variance of the new prediction is less than the variance of the predictions based on any single prediction model. In other words, compared to a single local soft sensor, the prediction uncertainty of the combined soft sensor has been improved.

4. An industrial case study

A typical polypropylene production device always contains a catalytic body system, which consists of TiCl_4 , triethylaluminum (TEAL), and diphenyldimethoxysilane (DONOR), a series of three reactors are connected. The flowchart of the polypropylene production process is shown in Figure 1. To record the data characteristic of this process, over 40 variables are measured online. In this study, the all data samples are collected from the process daily records and the corresponding laboratory analysis of one polypropylene production company in China. For prediction of the melt index in this process, a total of 14 process variables have been selected, which are highly correlated with the quality variable. These 14 selected input process variables are listed in Table 1.

Figure 1: Flowchart of the polypropylene production process

Table 1. Selected variables in polypropylene production process for quality prediction

No.	Measured variables	No.	Measured variables
1	Hydrogen concentration of the first reactor	8	Propylene feed of the first reactor
2	Hydrogen concentration of the second reactor	9	Propylene feed of the second reactor
3	Density of the first reactor	10	Power for the first reactor
4	Density of the second reactor	11	Power for the second reactor
5	TEAL flow	12	Lever of the second reactor
6	DONOR flow	13	Temperature of the first reactor
7	Atmer-163 flow	14	Temperature of the second reactor

In this process, three operation modes have been carried out. 100 data samples of each operation mode have been selected for modeling training and 20 of each are used for performance evaluation. Therefore, a total of 300 data samples are used for construction of the new soft sensor, and 60 data samples are used for testing purpose. Since a lot of comparative studies have shown that the GPR model performs better than other nonlinear modeling approaches, this case study is mainly focused on the multiple operation mode behavior of the process data, and compares the prediction performance of the combined strategy based soft sensor with that of different local GPR model based soft sensors and the global GPR model based soft sensor. Besides, detailed illustrations and interpretations of the process data behavior, operation mode information, and the prediction uncertainty are also provided.

To examine the data behavior of the training dataset, the scatter plot of two input variables is shown in Figure 2 (a). It can be seen that three clusters have clearly exhibited, which are highlighted in ellipses. Correspondingly, the characteristic of the quality variable is given in Figure 2 (b), in which three different data behaviors can also be identified. Before the GPR model construction under each operation mode of the process, an initial PCA pre-processing step is carried out. To determine the number of retained principal components in each PCA model, the CPV rule has been used, which ensures that these retained principal components can explain over 85% information of the process data. Detailed explanation percentages of retained principal component in each local PCA model are shown in Figure 3. As can be seen, 7 principal components have been retained in the first and second local PCA models, while 8 principal components should be selected to explain over 85% of the data information in the third operation mode. However, it should be noted that with the increase of the model complexity, the overfitting problem may happened. Therefore, attentions should be paid on this issue if too many principal components are selected.

(a) (b)
Figure 2: Data characteristic of the training dataset, (a) Input data; (b) Quality data.

Figure 3: Explanation rates of principal components in each local PCA model, (a) First model; (b) Second model; (c) Third model.

Next, depending on the score matrices obtained by the three PCA models, local GPR models can be constructed in each operation mode. Three parameters in each local GPR model are optimized through the traditional gradient based optimization method. After about 30 steps, the optimal parameters values can be obtained. For comparison, a global GPR model has also been developed, which incorporates the information of all 300 data samples. To evaluate the prediction performance of the developed soft sensors, the root mean squared error (RMSE) criterion is used, which is defined as follows

$$RMSE = \sqrt{\frac{\sum_{j=1}^{n_{te}} \|\mathbf{y}_j - \hat{\mathbf{y}}_j\|^2}{n_{te}}} \quad (23)$$

where $j=1,2,\dots,n_{te}$, $\mathbf{y}_j$ and $\hat{\mathbf{y}}_j$ are real and predicted values, respectively, n_{te} is the total number of test data samples. Detailed prediction results of the soft sensors based on combined local GPR model (CLGPR), single local GPR model (SLGPR), and global GPR model (GGPR) are tabulated together in Table 2. Compared to GGPR and SLGPR based soft

sensors, the new CLGPR based soft sensor performs much better, since the RMSE value is much smaller. It can be seen that the GGPR model based soft sensor has better performance than that of the three SLGPR model based soft sensors. This is because the GGPR model has used all information of the training dataset, while the SLGPR model has only incorporated a portion of the training data information. Detailed prediction results of these three different types of soft sensor are given in Figure 4. Different from the single local model based approach, the global and combined local model based soft sensors can both track the grade change of the process, and thus perform much better.

Table 2. Quality prediction results (RMSE) of different soft sensors

Soft sensors		RMSE
SLGPR	First SLGPR	0.9492
	Second SLGPR	1.3126
	Third SLGPR	0.4866
GGPR		0.3502
CLGPR		0.2849

(a)

(b)

(c)

(d)

(e)

Figure 4: Detailed prediction results of different soft sensors, (a) First SLGPR; (b) Second SLGPR; (c) Third SLGPR; (d) GGPR; (e) CLGPR.

Although the single local model based soft sensor has much worse performance, in specific operation mode, it can perform very well. For example, when testing data samples are generated from the first operation mode, the prediction performance of the first SLGPR model based soft sensor will be very high. However, when it is used for prediction of other data samples that belongs to the second or third operation grades, the performance will be significantly deteriorated. Among the testing dataset used in this study, the first 20 data samples are from the operation mode one, while data samples 21-40 and 41-60 belong to the second and third operation modes, respectively. RMSE values of the three local model based soft sensors in these specific operation modes are tabulated in Table 3. Through this table, it can be seen that the local GPR model can perform well in its corresponding operation mode based on which the model has been constructed. If we combined the prediction results in different operation modes which are obtained by their corresponding local GPR model, the quality prediction result is the same as the CLGPR model. This is because there is no overlapping data sample in the testing dataset. However, compared to the multiple SLGPR models based soft sensor, the CLGPR method does not need to switch the prediction model if the operation condition has been changed, which means its automation level is higher than that of the SLGPR model. In order to evaluate the advantage of the GPR model, the compared results of the CLGPR model with other methods, such as the GMM model, fuzzy-learning based model, multiple local PLS, ANN, and SVR model based soft sensors are given in Table 4. It can be seen that the best prediction result has been obtained by the CLGPR model based soft sensor. This result is in accordance with previous research studies on the GPR model.

Table 3. Quality prediction results (RMSE) of SLGPR for different operation modes

Sample number	First SLGPR	Second SLGPR	Third SLGPR
1-20	0.3930	1.5906	0.1972
21-40	2.2545	0.2240	0.1900
41-60	0.6686	0.4738	0.1351

Table 4. Quality prediction results (RMSE) comparisons of different soft sensors

Soft sensor models	CLGPR	GMM	Fuzzy-PLS	Multiple PLS	Multiple ANN	Multiple SVR
RMSE	0.2849	0.3160	0.3161	0.3159	0.3008	0.2986

To examine the mode information of testing data samples, monitoring results of the T^2 statistic by the three local PCA models are given in Figure 5. Compared to data samples 21-60, the first 20 data samples have much smaller T^2 statistic values in Figure 5 (a), which means that these data samples have high probabilities in the first operation mode. Similarly, it can also be inferred that data samples 21-40 and 41-60 have high probabilities in the second and third operation modes. Precisely, the posterior probability value of each testing data sample under the three different operation modes can be examined, which are shown in Figure 6. As can be seen, the results are consistent with the monitoring results presented in Figure 5.

Figure 5: Monitoring results of the T^2 statistic, (a) First PCA model; (b) Second PCA model; (c) Third PCA model.

Figure 6: Posterior probability value of each testing data sample in different operation modes, (a) First grade; (b) Second grade; (c) Third grade.

Finally, the uncertainty information of the prediction result is examined. Predictive variances of both local and combined soft sensors are demonstrated in Figure 7 (a-c) and Figure 7 (d). Compared to the result of single local model based soft sensor, the predictive variance of the combined local model based soft sensor has been greatly reduced. Actually, it can be inferred from eq. (21) that the combined predictive variance can be significantly reduced if different weights have been taken by local models. The most ideal case is that an equal weight is taken by each of the local model, thus the predictive variance can be reduced to the smallest value.

Figure 7: Predictive variance of local and combined soft sensors, (a) First SLGPR; (b) Second SLGPR; (c) Third SLGPR; (d) CLGPR.

5. Conclusions

In the present paper, a combination form of the local Gaussian process regression model based soft sensor has been developed for quality prediction of the polypropylene production process. Different from existing soft sensors, the new soft sensor can simultaneously address the nonlinear and multiple operation mode characteristics in the process. Besides, based on the structure of the Gaussian process model, the new soft sensor can also provide a predictive distribution for the quality variable, which can exhibit the uncertainty information of the soft sensor. Through a real industrial application case study, the feasibility and efficiency of the proposed soft sensor have both been confirmed. In our opinion, further researches for soft sensor modeling on this topic may be focused on the dynamic and time-varying extensions of the Gaussian process regression model. Besides, incorporation of the developed soft sensor into the feedback control system is also an interesting research topic in the future.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (61004134), the National 863 High Technology Research and Development Program of China (2009AA04Z154), and China Postdoctoral Science Foundation (20090461370).

References

- Chen, T., & Ren, J. (2009). Bagging for Gaussian process regression. *Neurocomputing*, 72, 1605-1610.
- Choi, S. W., Park, J. H., & Lee, I. B. (2004). Process monitoring using a Gaussian mixture model via principal component analysis and discriminant analysis. *Computers and Chemical Engineering*, 28, 1377-1387.
- Chu, W., & Ghahramani, Z. (2005). Gaussian processes for ordinal regression. *Journal of Machine Learning Research*, 6, 1019-1041.

- Csato, L., & Oppper, M. (2002). Sparse on-line Gaussian process. *Neural Computation*, 14, 641-668.
- Coates, P. D., Barnes, S. E., Sibley, M. G., Brown, E. C., Edwards, H. G. M., & Scowen I. J. (2003). In-process vibrational spectroscopy and ultrasound measurements in polymer melt extrusion. *Polymer*, 44, 5937-5949.
- Gao, L., & Ren, S. X. (2010). Combing orthogonal signal correction and wavelet packet transform with radial basis function neural networks for multicomponent determination. *Chem. Intell. Lab. Syst.*, 100, 57-65.
- Gonzaga, J. C. B., Meleiro, L. A. C., Kiang, C., & Filho, R. M. (2009). ANN-based soft-sensor for real-time process monitoring and control of an industrial polymerization process. *Computers and Chemical Engineering*, 33, 43-49.
- Huang, Z. Y., & Hahn, J. (2009). Fuzzy modeling of signal transduction networks. *Chemical Engineering Science*, 64, 2044-2056.
- Kadlec, P., Gabrys, B., Strandt, S. (2009). Data-driven soft sensors in the process industry. *Computers and Chemical Engineering*, 33, 795-814.
- Kano, M., & Nakagawa, Y. (2008). Data-based process monitoring, process control and quality improvement: recent developments and applications in steel industry. *Computers and Chemical Engineering*, 32, 12-24.
- Kiparissides, C., Verros, G., & MacGregor, J. F. (1993). Mathematical modeling, optimization, and quality control of high pressure polymerization reactors. *J. Macromolecular Science-Reviews in Macromolecular Chemistry and Physics*, C33 (4), 437-527.
- Likar, B., & Kocijan, J. (2007). Predictive control of a gas-liquid separation plant based on a Gaussian process model. *Computers and Chemical Engineering*, 31, 142-152.
- Liu, J. L. (2007). On-line soft sensor for polyethylene process with multiple production grades. *Control Engineering Practice*, 15, 769-778.
- O'Hagan, A. (1978). Curve fitting and optimal design for prediction. *Journal of Roy. Stat. Soc. B*, 40, 1-42.
- Ohshima, M., Tanigaki, M. (2000). Quality control of polymer production processes. *Journal of Process Control*, 10, 135-148.
- Ou, X. L., & Martin, E. (2008). Batch process modeling with mixtures of Gaussian processes. *Neural Computing and Application*, 17, 471-479.
- Qin, S. J. (2003). Statistical process monitoring: basics and beyond. *J. Chemometrics*, 17, 480-502.
- Rasmussen, C. E., Ghahramani, Z. (2002). Infinite mixtures of Gaussian process experts. In: Dietterich T., Becker S., Ghahramani Z. *Advances in Neural Information Processing Systems*, 14, MIT Press.
- Rasmussen, C. E., & Williams, C. K. I. (2006). Gaussian processes for machine learning. The MIT Press, 2006.
- Rong, H. J., Sundararajan, N., Huang, G. B., & Saratchandran, P. (2006). Sequential adaptive fuzzy inference system (SAFIS) for nonlinear system identification and prediction. *Fuzzy Sets and Systems*, 157, 1260-1275.
- Shi, J., Liu, X. G., & Sun, Y. X. (2006). Melt index prediction by neural networks based on independent component analysis and multi-scale analysis. *Neurocomputing*, 70, 280-287.

- Shi, J. Q., Murray-Smith, R., & Titterington, D. M. (2005). Hierarchical Gaussian process mixtures for regression. *Statistics and Computing*, 15, 31-41.
- Shi, J. Q., Wang, B., Murray-Smith, R., & Titterington, D. M. (2007). Gaussian process functional regression modeling for batch data. *Biometrics*, 63, 714-723.
- Tang, Q., Lau, Y., Hu, S., Yan, W., Yang, Y., & Chen, T. (2010). Response surface methodology using Gaussian processes: towards optimizing the trans-stilbene epoxidation over Co^{2+} -NaX catalysts. *Chemical Engineering Journal*, 156, 423-431.
- Yang, Y., Gao, F. R. (2006). Injection molding product weight: Online prediction and control based on a nonlinear principal component regression model. *Polymer Engineering and Science*, 46, 540-548.
- Yu, J., & Qin, S. J. (2008). Multimode process monitoring with Bayesian inference-based finite Gaussian mixture models. *AIChE Journal*, 54, 1811-1829.
- Yu, J., & Qin, S. J. (2009). Multimode process monitoring with Bayesian inference-based finite Gaussian mixture models. *Ind. Eng. Chem. Res.*, 48, 8585-8594.
- Zhang, J., Morris, A. J., Martin, E. B., & Kiparissides, C. (1998). Prediction of polymer quality in batch polymerization reactors using robust neural networks. *Chemical Engineering Journal*, 69, 135-143.