Chemometrics and Intelligent Laboratory Systems 155 (2016) 170–182

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemolab

Tutorial article

Soft sensor development for online quality prediction of industrial batch rubber mixing process using ensemble just-in-time Gaussian process regression models

Kai Yang [a,b], Huaiping Jin [a,*], Xiangguang Chen [a,*], Jiayu Dai [a], Li Wang [a], Dongxiang Zhang [a]

[a] Department of Chemical Engineering, Beijing Institute of Technology, Beijing 100081, People's Republic of China
[b] Beijing Research & Design Institute of Rubber Industry, Beijing 100143, People's Republic of China

ARTICLE INFO

Article history:
Received 14 January 2016
Received in revised form 13 April 2016
Accepted 16 April 2016
Available online 20 April 2016

Keywords:
Soft sensor
Ensemble learning
Just-in-time learning
Gaussian process regression
Mooney viscosity
Rubber mixing

ABSTRACT

Rubber mixing is a nonlinear batch process that lasts for very a short time (*ca.* 2–5 min). However, the lack of online sensors for quality variable (*e.g.*, the Mooney viscosity) has become a main obstacle of controlling rubber mixing accurately, automatically and optimally. This paper proposes a novel soft sensing method based on Gaussian process regression (GPR) models fortified with both ensemble learning and just-in-time (JIT) learning, which ensures precision and robustness at the same time. More specifically, this method first builds multiple input variable sets from random local datasets, then uses the obtained input variable sets to establish local models and send them to ensemble learning with Bayesian inference and finite mixture mechanism before making the final prediction output. The superiority of the proposed method is demonstrated using an industrial rubber mixing process.

1. Introduction

Rubber mixing is the first and very crucial process in the rubber and tire manufacturing industry. It is a typical nonlinear and time-varying batch process, for which the accurate and reliable online measurements of quality variables are critical for controlling, monitoring, optimizing and stabilizing production. Mooney viscosity can indicate the molecular weight of an elastomer and its viscoelastic behavior, and it is one of the key end-use quality variables in rubber mixing [1,2]. However, it can only be determined through manual analysis carried out offline in laboratories, which cannot be done until 4–6 h after rubber mixing, whereas the mixing process of a batch run lasts merely for about 2–5 min. This means large amounts of rubber will be manufactured more or less blindly, *i.e.*, without timely and accurate knowledge and control over the ongoing production. Therefore, real-time estimation of the Mooney viscosity is highly desired for optimal and uniform rubber product quality.

Nowadays, soft sensor technology has been widely used to determine quality variables that are difficult to measure online [3–5]. Generally, soft sensors are either model-driven or data-driven [6,7]. The former are based on the chemical and physical principles underlying the process, which understandably are normally unavailable for complex industrial applications. In contrast, the latter are preferable because they mainly rely on the operation data and require minimal process knowledge. Thus we here focus on data-driven soft sensors. Multivariate statistical techniques such as principle component regression (PCR) [8] and partial least squares (PLS) [9] are well-established data-driven methods for soft sensors, but they are essentially linear modeling techniques that cannot handle process nonlinearity. Consequently, various machine learning methods have been introduced to soft sensor applications, including artificial neural networks (ANN) [10,11], neuro-fuzzy systems [12], support vector regression (SVR) [13], Gaussian process regression (GPR) [1,14], as well as other kernel-based methods [15].

Data selection plays a crucial role in developing data-driven soft sensors. In industrial applications, however, historical data collected for soft sensor modeling cannot cover all possible states and conditions of the process, and new process states may emerge when new data are sampled. Moreover, the process characteristics vary constantly due to equipment aging, raw material changes, catalyst deactivation, seasonal changes, *etc.* As a result, a single rigidly defined model is unsuitable. To account for such evolving behavior, various recursive strategies have been proposed to build adaptive soft sensor models [16,17]. A discounted-measurement recursive partial least squares modeling method has been proposed in [2] to predict the Mooney viscosity in rubber mixing, which can overcome the noise and multi-collinearity in the original data. Nevertheless, recursive methods can only deal with gradual changes in operation conditions and fail upon abrupt changes in process characteristics. In addition, recursive strategies tend to excessively adapt the model when the process is operated within narrow operating conditions. Another popular strategy is the moving

* Corresponding authors.
E-mail addresses: jinhuaiping@gmail.com (H. Jin), xgc1@bit.edu.cn (X. Chen).

http://dx.doi.org/10.1016/j.chemolab.2016.04.009
0169-7439/

window method [18,19], which uses the latest samples within a sliding window to adapt the model. The drawback, however, is that the old samples that may contain useful information are continuously discarded as time elapses.

To address the nonlinear and time-varying behavior of process simultaneously, local learning methods, inspired by the "divide and rule" philosophy, have attracted growing interests in soft sensor applications. The most popular local learning methods for soft sensor development include the just-in-time (JIT) learning and the ensemble learning. The JIT learning framework has been proposed to achieve local optimization and estimation, based on the principle that similar inputs should lead to similar outputs [20]. When estimation is requested, the JIT method builds a local model by selecting samples most similar to the query data, and the model is discarded after it makes the estimation. In this way, the JIT method can handle both nonlinearity and time-varying characteristics. Currently, many similarity criteria have been developed for the sample extraction in JIT learning [21–23]. Regarding JIT, locally weighted partial least squares (LWPLS) and local kernel learning have been proven useful in dealing with multimode and time-varying problems in industrial process estimation [24]. Recently, Liu proposed a novel auto-switch GPR (AGPR) soft sensing algorithm for multigrade processes in polymer and fine industries, in which the posterior probabilities of the query data with respect to the current state is evaluated and then the most suitable model is selected for online prediction [25]. In this approach, JIT-GPR modeling method was used to achieve the online estimation in the transient states while the offline built models are used during the steady states. Nevertheless, a single local model by itself may fall short in capturing all relevant characteristics of the process. Furthermore, because the size and the similarity measure of the optimal local model are so diverse, the JIT predictions are at times unstable. On the other hand, ensemble learning has been used widely to improve the robustness of online predictions [1,14,26]. Under the ensemble learning framework, a series of local datasets are built from the historical database, on which multiple local models are built. The local datasets can be constructed in a number of ways, such as bagging, boosting, random subspaces, fuzzy *c*-means clustering, *etc.* [27,28]. For rubber mixing, a probabilistic ensemble Gaussian process regression (GPR) modeling method has been developed to allow accurate prediction of the Mooney viscosity [1].

Besides data selection, variable selection is yet another key consideration in constructing soft sensor models. Commonly known variable selection methods include absolute shrinkage and selection operator (LASSO) [29], PLS with variable importance in projection (PLS-VIP), genetic algorithm with PLS (GA-PLS), *etc.* It has been shown that the performance of soft sensors can improve greatly when nonessential variables are excluded [30,31]. For online prediction of the end-use quality variables in batch processes, all samplings at different time instants can be used as inputs of soft sensor models to estimate the target variable at the end of a batch run. This means that lots of potential input variables are available to predict a single output variable. Thus variable selection is essential to achieve dimension reduction and redundant information removal, which aims to improve the estimation accuracy as well as computational efficiency.

Regrettably, diversity of variable selection within local learning framework has rarely been explored. In fact, most local learning based soft sensors simply employ one single input variable set and merely manipulate samples by dividing historical data into local subsets. However, these soft sensor methods cannot effectively capture the process characteristics. In particular, because rubber mixing is a multimode batch process that has constantly varying operation conditions (*e.g.*, feed material, recipe, *etc.*) and process characteristics, a fixed set of input variables may neglect important features that are unique only to a subset of data and generate redundancy from other data when they capture overlapping information. Consequently, it is often difficult and even impossible to choose a fixed and optimal set of process variables to provide the satisfactory prediction results under all possible operation conditions. To address this issue, the selection and combination of input variables for local learning should be diverse enough to cover the process characteristics and contribute to the accuracy of the eventual global output.

To address the above-mentioned issues, a novel soft sensor, referred to as the ensemble just-in-time GPR (EJITGPR) model, is proposed in this work to improve the performance of quality prediction in rubber mixing process. This method is based on the integration of ensemble learning and JIT learning. Different from traditional local learning methods, diverse input variable sets are first customized by using the partial mutual information (PMI) criterion and random resampling of historical data. Then a PLS analysis mechanism is applied to remove unimportant or redundant local variable sets, which protects the diversity of input variable sets without sacrificing the online prediction performance. Furthermore, based on the selected input variable sets, multiple local GPR models are built using JIT learning to provide local prediction results, which include the local output estimation and prediction variance. Finally, the Bayesian inference and the finite mixture mechanism are applied to consolidate the best local prediction results into the final output. This novel EJITGPR method can handle both process nonlinearity and time-varying behavior thanks to JIT learning, and it can also improve the prediction robustness and accuracy through ensemble learning.

The remainder of this paper is organized as follows. Section 2 briefly reviews Gaussian process regression (GPR), partial mutual information (PMI), and the JIT learning framework. In Section 3, the proposed EJITGPR soft sensing algorithm is described in detail. Section 4 demonstrates the effectiveness of the EJITGPR method and its superiority over traditional methods in an industrial batch rubber mixing case study. Concluding remarks are drawn in Section 5.

2. Preliminaries

The GPR model, PMI criterion and JIT learning method are revisited in this section.

2.1. Gaussian process regression

Over the past decade, Gaussian process regression has attracted much attention in machine learning because it provides a principled, practical and probabilistic approach for kernel learning machines [32]. Normally, the Gaussian process includes a collection of random variables, any finite number of which follows joint Gaussian distributions. With the dataset $\mathrm{D}=\{\mathbf{X},\mathbf{y}\}=\{\mathbf{x}_i,y_i\}_{i=1}^{n}$, the regression model can be formulated as:

$$y = f(\mathbf{x}) + \varepsilon \tag{1}$$

where $f(\cdot)$ represents an unknown regression function and ε denotes the Gaussian noise with zero mean and variance σ_n^2. The Gaussian process is completely specified by its mean function $m(\mathbf{x})$ and covariance function $C(\mathbf{x},\mathbf{x}')$ from the function space view, which are defined as follows:

$$m(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})] \tag{2}$$

$$C(\mathbf{x},\mathbf{x}') = \mathbb{E}[(\mathrm{f}(\mathbf{x})-m(\mathbf{x}))(\mathrm{f}(\mathbf{x}')-m(\mathbf{x}'))] \tag{3}$$

Then the Gaussian process can be denoted as:

$$f(\mathbf{x}) \sim \mathrm{GP}(m(\mathbf{x}), C(\mathbf{x},\mathbf{x}')) \tag{4}$$

Usually, the data is normalized for notation simplicity, and then the output observations follow a Gaussian distribution as:

$$\mathbf{y} \sim GP(0, C(\mathbf{x},\mathbf{x}')) \tag{5}$$

When a query input $\mathbf{x}_*$ comes in, the joint distribution of the training outputs $\mathbf{y}$ and the test output y_* according to the prior is

$$\begin{bmatrix}\mathbf{y}\\ y_*\end{bmatrix} \sim N\left(0, \begin{bmatrix}\mathbf{C} & \mathbf{k}_*\\ \mathbf{k}_*^{\mathrm{T}} & C(\mathbf{x}_*,\mathbf{x}_*)\end{bmatrix}\right) \tag{6}$$