Chemometrics and Intelligent Laboratory Systems 180 (2018) 36–41

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

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

Just-in-time semi-supervised soft sensor for quality prediction in industrial rubber mixers

Wenjian Zheng, Yi Liu *, Zengliang Gao, Jianguo Yang

Institute of Process Equipment and Control Engineering, Zhejiang University of Technology, Hangzhou, 310014, China

ARTICLE INFO

Keywords:
Soft sensor
Just-in-time learning
Semi-supervised learning
Extreme learning machine
Rubber mixing process

ABSTRACT

Increasing data-driven soft sensors have been adopted to online predict the quality indices in polymerization processes to improve the availability of measurements and efficiency. However, in industrial rubber mixing processes, most existing soft sensors for online prediction of the Mooney viscosity only utilized the limited labeled data. By exploring the unlabeled data, a novel soft sensor, namely just-in-time semi-supervised extreme learning machine (JSELM), is proposed to online predict the Mooney viscosity with multiple recipes. It integrates the just-in-time learning, extreme learning machine (ELM), and the graph Laplacian regularization into a unified online modeling framework. When a test sample is inquired online, the useful information in both of similar labeled and unlabeled data is absorbed into its prediction model. Unlike traditional just-in-time learning models only utilizing labeled data (e.g., just-in-time ELM and just-in-time support vector regression), the prediction performance of JSELM can be enhanced by taking advantage of the information in lots of unlabeled data. Moreover, an efficient model selection strategy is formulated for online construction of the JSELM prediction model. Compared with traditional soft sensor methods, the superiority of JSELM is validated via the Mooney viscosity prediction in an industrial rubber mixer.

1. Introduction

Accurate and reliable measurements of process variables and quality indices in chemical processes can ensure the success in their products. The industrial rubber mixing process is a fast (only 2–5 min), nonlinear, and time-varying batch process performed in an internal mixer. The Mooney viscosity is a key quality index which represents the viscoelastic behavior of an elastomer. Up to date, an accurate first-principles Mooney viscosity model is still not available. On the other hand, in most rubber/tire factories, the Mooney viscosity can be only obtained offline using laboratory analysis several hours later after a batch has been discharged [1,2]. In the absence of an economical or effective online measurement, soft sensors (or inferential sensors) could serve as an alternative solution [3–7]. Additionally, with the wide availability of process data in rubber/tire factories, increasing data-driven soft sensors have been adopted to predict the Mooney viscosity information [8–16].

Nowadays, common data-driven soft sensors for the Mooney viscosity prediction include neural networks (NN) [8–10], multivariable regression (e.g., partial least squares regression), and kernel learning-based regression approaches [11–16]. However, up to now, most of them are constructed in a supervised learning manner. This means that complete data samples including both input and output variables are need. Here, the labeled dataset denotes the one containing both of input and output data. And the unlabeled dataset represents the one only has input data. In the machine learning area, training a model with both of labeled and unlabeled data is known as semi-supervised learning. By suitably incorporating the information of unlabeled data into the supervised regression model, more accurate prediction of semi-supervised models can often be obtained than related supervised methods [17,18]. However, in chemical processes, semi-supervised soft sensor applications are still much fewer than those with supervised models [19–22]. From a practical modeling viewpoint, with lots of unlabeled data available, the development of novel semi-supervised soft sensors is promising.

As shown in Fig. 1, due to the time-consuming and costly lab assaying process in industrial rubber/tire plants, the assayed Mooney viscosity are delayed and limited. On the other hand, the mixing pressure, the mixing temperature, and other mixing variables are online measured continually during every production batch in an industrial mixer. This means lots of unlabeled data are available, while labeled data are limited. To cope with the gap, recently, a semi-supervised extreme learning machine (SELM) [23] and its improved version were applied to predict the Mooney viscosity with one recipe and showed better prediction performance than

* Corresponding author.
E-mail address: yliuzju@zjut.edu.cn (Y. Liu).

https://doi.org/10.1016/j.chemolab.2018.07.002
Received 30 January 2018; Received in revised form 20 June 2018; Accepted 2 July 2018
Available online 3 July 2018

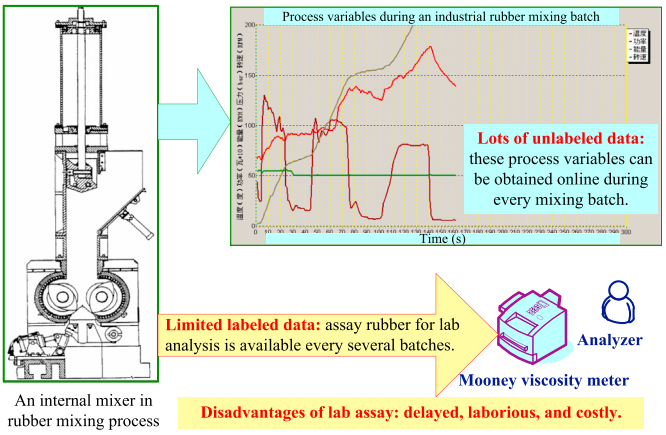

Fig. 1. Lots of unlabeled data during every batch and limited labeled data only available offline every several batches.

only using the soft sensor with labeled data [24]. However, for multiple mixing recipes with different characteristics, only using a single model is still not enough. Additionally, the varying properties of raw materials and mixing operating conditions indeed introduce batch-to-batch variations to both of the quality index and process data. As pointed out by Lu and Chiang [7], the life spans of most data-driven soft sensors are limited in practice. Consequently, a flexible model with adaptive structure is more suitable and attractive in industrial processes.

Just-in-time learning (JITL) methods, as an alternative solution, have been utilized in the development of data-driven soft sensors in chemical processes, especially for those with multiple modes/grades [25–29]. However, most traditional JITL-based soft sensors were built in a supervised learning manner. For online prediction of a query sample, only the labeled data are considered as the similar data. The information hidden in lots of unlabeled data (e.g., those online measured process variables) is omitted in selection and modeling of similar samples. As a result, the performance of JITL-based soft sensors may be restricted to some extent, especially when the labeled data are limited in industrial practice. To our best knowledge, the JITL-based semi-supervised soft sensor has not yet been developed, especially for rubber mixing process applications.

In this work, a JITL-based SELM (denoted as JSELM) soft sensor is proposed to predict the Mooney viscosity. The JSELM method integrates the JITL modeling manner, extreme learning machine (ELM) [30,31], and the graph Laplacian regularization into a unified framework. For online inquiry of a test sample, the useful information in both of similar labeled and unlabeled data is absorbed into its prediction model. Additionally, a fast cross-validation strategy is formulated for online construction of the JSELM model efficiently. Consequently, compared with traditional JITL models only using labeled data (e.g., JITL-based ELM [29]), the prediction performance of JSELM can be enhanced by taking advantage of the information in unlabeled data.

The remainder is so organized. In Section 2, after a brief introduction of ELM, the SELM soft sensing method is formulated. In Section 3, the proposed JSELM method with its model selection strategy is developed. The JSELM approach is validated via the industrial Mooney viscosity prediction in Section 4. Finally, the conclusion is drawn in Section 5.

2. SELM soft sensor modeling method

2.1. Brief overview of ELM model

In chemical processes, NN-based soft sensors have been widely utilized in past two decades. With a single layer feed-forward NN structure, the weights of the hidden neurons in ELM can be obtained very fast [30,31]. Due to its easy-to-use and nonlinear modeling properties, increasing ELM-based soft sensor applications have recently caught more attention [29,32–34]. Here, the ELM-based supervised modeling algorithm is briefly described. The labeled set is denoted as $\{\mathbf{S}^l\} = \{\mathbf{X}^l, \mathbf{Y}^l\}$, where $\{\mathbf{X}^l\} = \{\mathbf{x}_i^l\}_{i=1}^L$ and $\{\mathbf{Y}^l\} = \{y_i^l\}_{i=1}^L$ are the input and output datasets with L samples, respectively. ELM with N hidden nodes and the activation function $g(\cdot)$ can approximate the training data with zero error, which means $\sum_{i=1}^L \left\| y_i^l - \hat{y}_i^l \right\| = 0$, where y_i^l and $\hat{y}_i^l$ denote the measured output and predicted output, respectively. Compactly, the regression formulation of ELM is represented as [30,31]:

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{Y}^l \tag{1}$$

where the output matrix of hidden-layer $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \cdots, \mathbf{h}_N]_{L\times N}$ with $\mathbf{h}_i = \begin{bmatrix} g(\langle \mathbf{a}_i, \mathbf{x}_1^l \rangle + b_i) \\ \vdots \\ g(\langle \mathbf{a}_i, \mathbf{x}_L^l \rangle + b_i) \end{bmatrix}_{L\times 1}, i = 1, \cdots, N$; $g(\langle \mathbf{a}_i, \mathbf{x}_j^l \rangle + b_i)$ is the output of the ith hidden node related to the jth input $\mathbf{x}_j^l$. The terms $\mathbf{a}_i$ and b_i denote the input weight and the bias of the ith hidden node, respectively; and $\langle \mathbf{a}_i, \mathbf{x}_j^l \rangle$ denotes the inner product of $\mathbf{a}_i$ and $\mathbf{x}_j^l$. Here, the common sigmoidal function $g(v) = \frac{1}{1+\exp(-v)}$ is adopted for its nonlinear modeling ability. And $\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_N \end{bmatrix}_{N\times 1}$ is the output weight parameter.

The weights of ELM are not necessarily retuned using some complex training algorithms [30,31]. For many regression cases, the number of training samples is much more than the number of hidden nodes, i.e., $L >> N$. As a result, the output weights $\hat{\boldsymbol{\beta}}$ are formulated as [30,31]:

$$\hat{\boldsymbol{\beta}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{Y}^l \tag{2}$$

Furthermore, to avoid the problem of $\mathbf{H}^T\mathbf{H}$ being noninvertible, a regularized ELM (RELM) model was constructed by simply adding a small value of ridge parameter $\gamma > 0$ to obtain $\hat{\boldsymbol{\beta}}$ [29,32].

$$\hat{\boldsymbol{\beta}} = (\mathbf{H}^T\mathbf{H} + \gamma\mathbf{I})^{-1}\mathbf{H}^T\mathbf{Y}^l \tag{3}$$

where $\mathbf{I}$ is a unit matrix.

Finally, for a test sample $\mathbf{x}_t = [x_{t1}, x_{t2}, \cdots, x_{tn}]^T \in R^n$, its prediction $\hat{y}_t$ is obtained below:

$$\hat{y}_t = \mathbf{h}_t\hat{\boldsymbol{\beta}} = \mathbf{h}_t(\mathbf{H}^T\mathbf{H} + \gamma\mathbf{I})^{-1}\mathbf{H}^T\mathbf{Y}^l \tag{4}$$

where $\mathbf{h}_t$ is the output hidden-layer vector related to $\mathbf{x}_t$.

2.2. SELM modeling approach

Semi-supervised learning algorithms assume that the input patterns from both labeled and unlabeled data are drawn from the same marginal distribution. Additionally, with the smoothness assumption [17,18], the data in the local region should have similar labels. In such a situation, the unlabeled data naturally provide useful information for exploring the data structure in the input space. By assuming that the input data follow the same manifold in the input space, semi-supervised learning algorithms can incorporate both labeled and unlabeled data into the learning process. Consequently, benefiting from lots of unlabeled data, the SELM model could provide more accurate prediction performance than ELM [23,24].

The input and output data samples are denoted as $\{\mathbf{X}\} = \{\mathbf{X}^l \cup \mathbf{X}^u\}$ and $\mathbf{Y} = \begin{bmatrix} \mathbf{Y}^l \\ \mathbf{Y}^u \end{bmatrix} = \begin{bmatrix} y_1^l \\ \vdots \\ y_L^l \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{(L+U)\times 1}$, respectively. Correspondingly, the hid-

den layer output matrix **H** can be described as below.

$$\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \cdots, \mathbf{h}_N]_{(L+U)\times N} \tag{5}$$

To build an SELM model, the manifold regularization framework is adopted to learn its coefficient matrix **W** [23].

$$\min_{\mathbf{W}} \quad \frac{1}{2}\left\{\|\mathbf{JHW} - \mathbf{Y}\|^2 + \lambda(\mathbf{HW})^T\mathbf{LHW}\right\} \tag{6}$$

where $\|\mathbf{JHW} - \mathbf{Y}\|^2$ is the empirical fitting errors of labeled training data while $\lambda(\mathbf{HW})^T\mathbf{LHW}$ denotes the penalty of the model complexity using the graph Laplacian **L** with a balance parameter $\lambda \geq 0$ [23]. And define the matrix $\mathbf{J} = \begin{bmatrix} \mathbf{I}_L & 0 \\ 0 & 0 \end{bmatrix}_{(L+U)\times(L+U)}$ for the convenience of calculation.

The coefficient matrix **W** can be obtained by solving the optimization problem in Eq. (6) [23]:

$$\mathbf{W} = [(\mathbf{J} + \lambda\mathbf{L}^T)\mathbf{H}]^+\mathbf{JY} \tag{7}$$

Briefly, SELM is an extension of ELM by incorporating the unlabeled data into the modeling framework. The graph Laplacian **L** contains the information in both of labeled and unlabeled data. Once the unlabeled data are ignored (i.e., $\lambda = 0$), **W** can be considered as the same as $\widehat{\boldsymbol{\beta}}$ in Eq. (3). With a suitable choice of λ, the prediction performance can be enhanced. Finally, for a test sample $\mathbf{x}_t = [x_{t1}, x_{t2}, \cdots, x_{tn}]^T \in R^n$, its prediction $\widehat{y}_t$ is obtained below:

$$\widehat{y}_t = \mathbf{h}_t\mathbf{W} = \mathbf{h}_t[(\mathbf{J} + \lambda\mathbf{L}^T)\mathbf{H}]^+\mathbf{JY} \tag{8}$$

where $\mathbf{h}_t$ is the output hidden-layer vector associated with $\mathbf{x}_t$.

3. JSELM and its implementations

3.1. JITL-based semi-supervised modeling framework

For nonlinear processes, JITL-based local modeling methods are more flexible than only using a global/fixed model [25–29]. First, it is not easy to construct a global and accurate data-driven model and always sustain the high performance when it is commissioned online for industrial processes. Second, for some processes with complex characteristics and changing dynamics, a global/fixed model may not be reliable for the long-term utilization. Recently, the JITL-based RELM (simply denoted as JRELM) approach was proposed to predict the Mooney viscosity [29]. For online modeling and prediction of a test sample $\mathbf{x}_t$, three main steps are need. First, choose a similar set from the historical labeled data using some defined similarity criteria. Second, build a JRELM model $f_{\text{JRELM}}(\mathbf{x}_t)$ using the selected similar set. Third, online obtain the prediction $\widehat{y}_t$ for $\mathbf{x}_t$ and then discard the current JRELM model $f_{\text{JRELM}}(\mathbf{x}_t)$.

Most of the traditional JITL-based modeling methods only utilize the limited labeled data. However, the useful process information hidden in large amount of unlabeled data samples is ignored. With the help of unlabeled data, JITL-based soft sensors can be constructed using more informative data and thus the prediction accuracy can be improved.

Here, a simplified example is discussed to exhibit the effect of JITL-based soft sensors using both of labeled and unlabeled data. As an illustrated case shown in Fig. 2, for the 1st test data, the online prediction is obtained using the JITL-based regression with the selected similar labeled data. Nevertheless, as shown in Fig. 3, for the 1st test data, its prediction with the JITL-based regression model only using the labeled data is not accurate mainly because the modeling information is very limited. It should be noticed that blue cross points in Fig. 3 are historical "unlabeled" data without knowing the output values. With the help of semi-supervised learning, the JITL-based regression model using both of the labeled and unlabeled data could become more accurate for this query sample. Additionally, as shown in Figs. 2 and 3, only using the labeled data, the main characteristics of a complex process may not be easily learnt. For example, for the 2nd test data, the prediction with the JITL-based regression model using the labeled data is accurate. However, the process tendency in some operating areas may not be captured accurately. Therefore, to overcome these problems, the JITL-based semi-supervised modeling framework is investigated in this work to exploit both labeled and unlabeled data for local modeling. Specially, the proposed JSELM model has good modeling ability by balancing the empirical fitting errors and the complexity of a learnt local model.

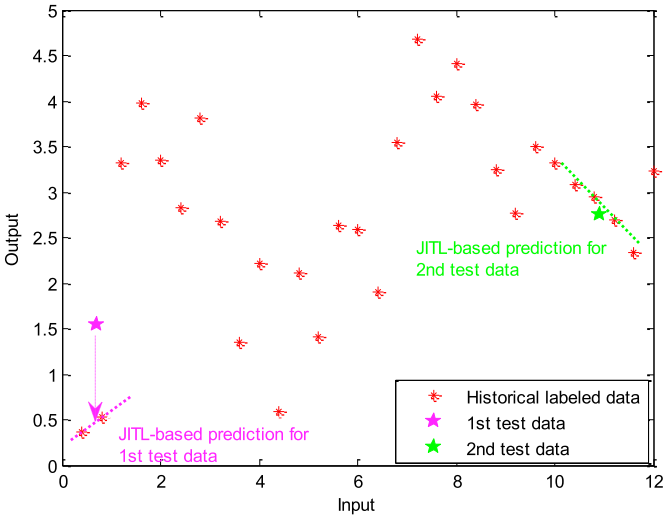

Fig. 2. Traditional supervised JITL-based soft sensor modeling method only using historical labeled data (take two test data as illustrated cases).

For online modeling and prediction of a test sample $\mathbf{x}_t$, the main procedure of the JSELM method is described as follows. First, choose a similar set $\{\mathbf{S}_t\} = \{\mathbf{S}_t^l \cup \mathbf{S}_t^u\}$, including both of labeled and unlabeled data (i.e., $\{\mathbf{S}_t^l\} = \{\mathbf{X}_t^l, \mathbf{Y}_t^l\}$ and $\{\mathbf{S}_t^u\} = \{\mathbf{X}_t^u\}$), from the historical database $\{\mathbf{S}\}$ using some defined similarity criteria. Second, build a JSELM model $f_{\text{JSELM}}(\mathbf{x}_t)$ using the selected similar dataset $\{\mathbf{S}_t\}$. Third, online obtain the prediction $\widehat{y}_t$ for $\mathbf{x}_t$ and then repeat the same steps for another query sample.

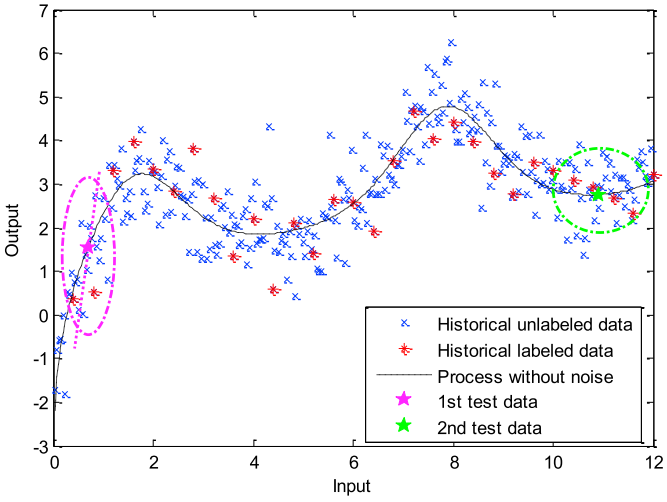

Fig. 3. Proposed semi-supervised JITL-based soft sensor modeling method by incorporating both of historical labeled and unlabeled data (take two test data as illustrated cases). It should be noticed that blue cross points are historical "unlabeled" data without knowing the output values. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

3.2. JSELM model selection and its implementations

For a selected similar dataset $\{\mathbf{S}_t\}$, two parameters should be determined to train a JSELM regression model. They are the number of hidden nodes N and the balance parameter $\lambda \geq 0$ in Eq. (6). For many regression modeling methods, the common leave-one-out (LOO) strategy was utilized as a suitable model selection criterion [35]. However, the traditional LOO method is time-consuming, especially for large data sets. Recently, a fast LOO (FLOO) method was proposed for training of the RELM regression model [29]. For an RELM model with N hidden nodes shown in Eq. (4), its FLOO-based error index, denoted as $E_{N,\gamma}^{\mathrm{FLOO}}$, can be formulated as [29].

$$E_{N,\gamma}^{\mathrm{FLOO}} = \frac{1}{L}\sum_{i=1}^{L}\left(\frac{y_i - \mathbf{h}_i(\mathbf{H}^T\mathbf{H}+\gamma\mathbf{I})^{-1}\mathbf{H}^T\mathbf{Y}^l}{1-\left[\mathbf{h}_i(\mathbf{H}^T\mathbf{H}+\gamma\mathbf{I})^{-1}\mathbf{H}^T\right]_i}\right)^2 \tag{9}$$

where $[\mathbf{h}_i(\mathbf{H}^T\mathbf{H}+\gamma\mathbf{I})^{-1}\mathbf{H}^T]_i$ denotes the ith element of the vector $\mathbf{h}_i(\mathbf{H}^T\mathbf{H}+\gamma\mathbf{I})^{-1}\mathbf{H}^T$.

Here, the FLOO criterion is extended to train a JSELM model with its parameters of N and λ. Similarly, the formula of FLOO-based error index $E_{N,\lambda}^{\mathrm{FLOO}}$ can be deduced as follows.

$$E_{N,\lambda}^{\mathrm{FLOO}} = \frac{1}{L}\sum_{i=1}^{L}\left(\frac{y_i - \mathbf{h}_i[(\mathbf{J}+\lambda\mathbf{L}^T)\mathbf{H}]^{+}\mathbf{J}\mathbf{Y}}{1-\left[\mathbf{h}_i[(\mathbf{J}+\lambda\mathbf{L}^T)\mathbf{H}]^{+}\mathbf{J}\right]_i}\right)^2 \tag{10}$$

Note that most of the items in Eq. (10) are similar with Eq. (9). Applying the FLOO strategy to obtain the parameters of a JSELM model is straightforward. This is mainly because the related terms in Eq. (10) are available and thus additionally computational loads are trivial. Moreover, using the FLOO criterion rather than determination of the model parameters in a trial-and-error manner, online training of a JSELM model is more efficiently.

Here, as one of the simplest unsupervised learning algorithms, the k-means clustering method [36] is utilized to search the similar samples for JSELM. For a query sample $\mathbf{x}_t$, the samples in the same cluster can be considered as its similar neighbors for semi-supervised learning of a local regression model. Consequently, the similar set can be obtained using the simple k-means clustering method. In summary, for a query sample $\mathbf{x}_t$, the step-by-step JSELM-based online modeling procedures are described as follows.

Step 1: Collect the labeled and unlabeled data samples as the historical dataset $\{\mathbf{S}\}$.
Step 2: For a new input sample $\mathbf{x}_t$, first add it into all input data $\{\mathbf{X}\} = \{\mathbf{X}^l \cup \mathbf{X}^u\}$. Then, the k-means clustering is unitized to partition $\{\mathbf{X}, \mathbf{x}_t\}$ into p clusters. Find out the samples with the same cluster of $\mathbf{x}_t$. These labeled and unlabeled data samples are considered as the similar set $\{\mathbf{S}_t\} = \{\mathbf{S}_t^l \cup \mathbf{S}_t^u\}$, i.e., $\{\mathbf{S}_t^l\} = \{\mathbf{X}_t^l, \mathbf{Y}_t^l\}$ and $\{\mathbf{S}_t^u\} = \{\mathbf{X}_t^u\}$.
Step 3: Construct a JSELM model $f_{\mathrm{JSELM}}(\mathbf{x}_t)$ using $\{\mathbf{S}_t\}$. Its parameters of N and λ are determined using the FLOO-based criterion using Eq. (10). For a candidate parameter p of the k-means clustering algorithm, an evaluated value of $E_{N,\lambda,p}^{\mathrm{FLOO}}$ and the related JSELM model can be obtained. Consequently, for several candidates, find out the JSELM model with the smallest value of $E_{N,\lambda,p}^{\mathrm{FLOO}}$. Finally, obtain the online prediction $\hat{y}_t$ using Eq. (8) and add $\mathbf{x}_t$ to the historical database.
Step 4: Go to **Step 2** and **Step 3** and repeat the same procedures for online prediction of another new input measurement.

The flowchart of the JSELM modeling and prediction method is shown in Fig. 4. For online prediction of a query sample, the similar labeled and unlabeled data samples are first searched using the k-means clustering approach. Additionally, three parameters of p, N and λ can be determined efficiently to select a suitable prediction model from candidates. Consequently, by utilizing the local semi-supervised learning strategy, JSELM is more flexible and suitable than only using the fixed SELM regression model.

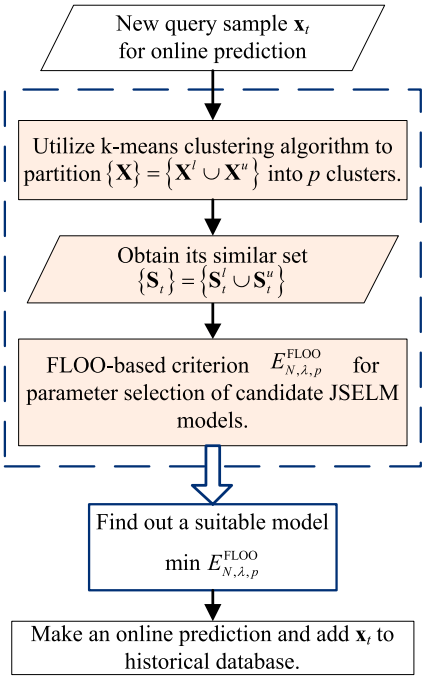

Fig. 4. Main modeling flowchart of clustering-based semi-supervised JITL soft sensing approach for online quality prediction.

4. Industrial application results and discussion

In this section, the JSELM method is applied to online predict the Mooney viscosity in an industrial rubber mixing process located in China. In industrial practice, for simply implementation, not all process variables during the whole batch duration are considered as the input. Instead, using the accumulated process knowledge, the online process measured variables near the end of the batch, including mixing pressure, mixing temperature, mixing power, mixing energy, and mixing duration in the chamber of internal mixer, were selected as the input variables of the prediction model [11–16]. From the same internal mixer, the unlabeled and labeled data samples of three recipes have been continually collected from daily process records and the corresponding laboratory analysis, respectively. In such a situation, the input patterns from both labeled and unlabeled data can be considered from the same marginal distribution. And some obvious outliers of both input and output data were removed out using a simple 3-sigma criterion. Finally, the training labeled set had about 210 pairs of data. Additionally, about 660 unlabeled training data were obtained in the same internal mixer during the same production period.

Previously, four soft sensor methods were applied to predict the Mooney viscosity information in rubber mixing processes. They are RELM [29–32], SELM [24], JRELM [29], and just-in-time support vector regression (JSVR) [29]. The main characteristics of JSELM and other four soft sensor modeling methods are described in Table 1. It should be mentioned that JSELM and SELM are semi-supervised learning methods. The useful information in unlabeled data could be utilized to learn a better regression model. In contrast, RELM, JRELM, and JSVR are supervised learning methods. Additionally, JSELM, JRELM, and JSVR are local modeling methods with a JITL manner. SELM and RELM are global modeling approaches with a fixed model structure. The relative root mean square error (simply denoted as RE) is adopted to evaluate the prediction performance of soft sensors.

Table 1
Brief description of main characteristics of JSELM and other four soft sensor models.

Soft sensor models	Brief description
JSELM (proposed)	Online build local semi-supervised learning models with a JITL manner
JRELM [29]	Online build local supervised RELM models with a JITL manner
JSVR [29]	Online build local supervised SVR models with a JITL manner
SELM [24]	Offline build a global/fixed semi-supervised model
RELM [29–32]	Offline build a global/fixed supervised RELM model

$$\mathrm{RE} = \sqrt{\sum_{t=1}^{L_t}\left(\frac{\hat{y}_t - y_t}{y_t}\right)^2 \Big/ L_t} \times 100\% \tag{11}$$

where $\hat{y}_t$ and y_t are the predicted and actual values of the Mooney viscosity, respectively. And L_t is the number of test data.

Details on the performance comparisons of JSELM, JRELM, JSVR, SELM, and RELM prediction methods are listed in Table 2. For all recipes, the RE indices indicate that JSELM can achieve the best prediction performance among all five methods. Three local learning prediction models generally perform better than only using the global learning models. This is mainly because industrial multiple rubber recipes show different characteristics. Additionally, even in the same recipe, batch-to-batch variations may introduce uncertainties to labeled and unlabeled data. Consequently, local models with flexible structure are more suitable in practice.

For three ELM-based models (i.e., JSELM, JRELM, and SELM), their detailed Mooney viscosity prediction results using the parity plot are shown in Fig. 5. The prediction results show that JSELM is better than JRELM and SELM approaches. Taking the 3rd recipe as an example, the prediction values of JSELM, JRELM, and SELM methods are shown in Fig. 6. And the absolute prediction errors are shown in Fig. 7. The prediction results for most of the test data exhibit that JSELM is more accurate than both of JRELM and SELM.

For online learning a JSELM model, the number of similar samples can be first determined using the number of cluster p. The candidate number of clusters for this application is set as 3–6 mainly because there are only three recipes. Additionally, two model parameters (i.e., the number of hidden nodes N and the balance parameter $\lambda \geq 0$) should be chosen. With the help of FLOO-based criterion, three model parameters can be online selected efficiently. The online prediction time for a test data only needs less than 2 s (with CPU main frequency 2.3 GHz and 4 GB memory), which is much less than the JSVR method needing about 50 s. Consequently, all the obtained results show that the JSELM model is more suitable for the industrial Mooney viscosity prediction with multiple recipes.

Table 2
Online prediction performance comparison of five soft sensor models for different recipes.

Soft sensor models	RE (%)			
	1st recipe	2nd recipe	3rd recipe	All recipes
JSELM (proposed)	2.94	2.78	3.32	3.04
JRELM [29]	3.46	4.22	5.45	4.48
JSVR [29]	5.16	5.09	5.27	5.18
SELM [24]	4.15	6.50	6.29	5.71
RELM [29–32]	5.34	7.56	8.14	7.10

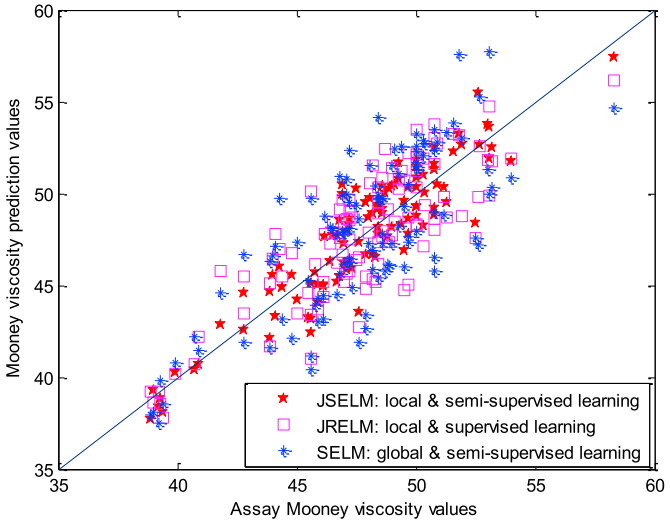

Fig. 5. Assay and predicted values of Mooney viscosity using JSELM, JRELM, and SELM soft sensors.

5. Conclusion

A just-in-time semi-supervised soft sensing method (denoted as JSELM) is developed to online predict the Mooney viscosity in industrial internal mixers. With an efficient model selection strategy, JITL and the semi-supervised learning are integrated into a modeling framework to

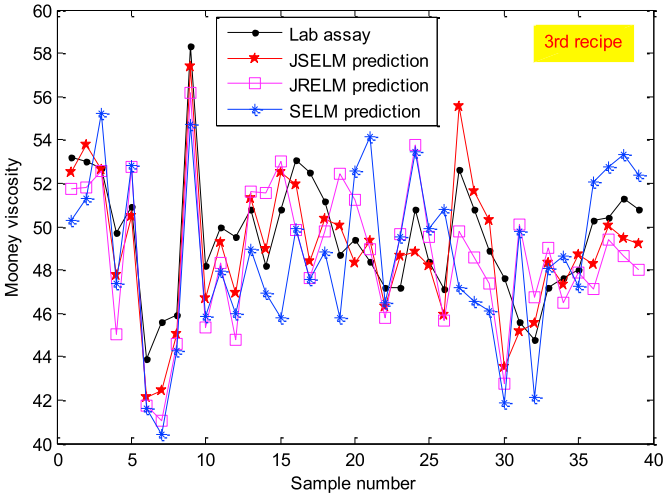

Fig. 6. Mooney viscosity prediction comparison results for the 3rd recipe using JSELM, JRELM, and SELM soft sensors.

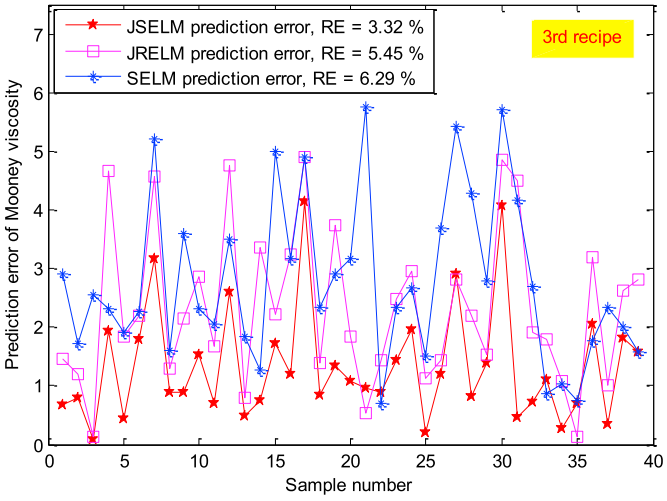

Fig. 7. Mooney viscosity prediction error comparison results for the 3rd recipe using JSELM, JRELM, and SELM soft sensors.

online construct more accurate local prediction models. By utilizing the information in lots of unlabeled data, the prediction performance of JITL-based local soft sensors can be enhanced. The application results to an industrial internal mixer validate that JSELM is more accurate than several JITL-based and semi-supervised soft sensors. In this study, only part of process variables during a whole batch are considered as input. How to utilize deep learning methods (e.g., deep brief networks [37] and convolutional neural networks [38]) to extract more intrinsic process features during the whole batch is an interesting research direction.

Acknowledgment

The authors would like to gratefully acknowledge Zhejiang Provincial Natural Science Foundation of China (grant no. LY18F030024), and Foundation of Key Laboratory of Advanced Process Control for Light Industry (Jiangnan University), Ministry of Education of China (grant no. APCLI1603) for their financial support.

Nomenclature

ELM	extreme learning machine
FLOO	fast leave-one-out
JITL	just-in-time learning
JSELM	just-in-time semi-supervised extreme learning machine
LOO	leave-one-out
NN	neural networks
RE	relative root mean square error
RELM	regularized extreme learning machine
SELM	semi-supervised extreme learning machine

References

[1] J.E. Mark, B. Erman, F.R. Eirich, The Science and Technology of Rubber, Elsevier Academic, San Diego, CA, 2005.
[2] E. Ehabe, F. Bonfils, C. Aymard, A.K. Akinlabi, J. Sainte Beuve, Modelling of Mooney viscosity relaxation in natural rubber, Polym. Test. 24 (2005) 620–627.
[3] L. Fortuna, S. Graziani, A. Rizzo, M.G. Xibilia, Soft Sensors for Monitoring and Control of Industrial Processes, Springer-Verlag, New York, 2007.
[4] P. Kadlec, B. Gabrys, S. Strandt, Data-driven soft sensors in the process industry, Comput. Chem. Eng. 33 (4) (2009) 795–814.
[5] Z.Q. Ge, Z.H. Song, S.X. Ding, B. Huang, Data mining and analytics in the process industry: the role of machine learning, IEEE Acc. 5 (2017) 20590–20616.
[6] W. Yan, S. Hu, Y. Yang, F. Gao, T. Chen, Bayesian migration of Gaussian process regression for rapid process modeling and optimization, Chem. Eng. J. 166 (2011) 1095–1103.
[7] B. Lu, L. Chiang, Semi-supervised online soft sensor maintenance experiences in the chemical industry, J. Process Contr. 67 (July 2018) 23–34, in: https://doi.org/10.1016/j.jprocont.2017.03.013.
[8] G. Padmavathi, M.G. Mandan, S.P. Mitra, K.K. Chaudhuri, Neural modelling of Mooney viscosity of polybutadiene rubber, Comput. Chem. Eng. 29 (2005) 1677–1685.
[9] V. Vijayabaskar, R. Gupta, P.P. Chakrabarti, A.K. Bhowmick, Prediction of properties of rubber by using artificial neural networks, J. Appl. Polym. Sci. 100 (2006) 2227–2237.
[10] A.G. Marcos, A.V.P. Espinoza, F.A. Elías, A.G. Forcada, A neural network-based approach for optimising rubber extrusion lines, Int. J. Comput. Integrated Manuf. 20 (2007) 828–837.
[11] Z. Zhang, K. Song, T.P. Tong, F. Wu, A novel nonlinear adaptive Mooney-viscosity model based on DRPLS-GP algorithm for rubber mixing process, Chemometr. Intell. Lab. Syst. 112 (2012) 17–23.
[12] K. Song, F. Wu, T. Tong, X. Wang, A real-time Mooney-viscosity prediction model of the mixed rubber based on the independent component regression-Gaussian process algorithm, J. Chemom. 11 (2012) 557–564.
[13] Y. Liu, Z.L. Gao, Real-time property prediction for an industrial rubber-mixing process with probabilistic ensemble Gaussian process regression models, J. Appl. Polym. Sci. 132 (2015) 1905–1913.
[14] H.P. Jin, X.G. Chen, L. Wang, K. Yang, L. Wu, Adaptive soft sensor development based on online ensemble Gaussian process regression for nonlinear time-varying batch processes, Ind. Eng. Chem. Res. 54 (2015) 7320–7345.
[15] K. Yang, H.P. Jin, X.G. Chen, J.Y. Dai, L. Wang, D.X. Zhang, Soft sensor development for online quality prediction of industrial batch rubber mixing process using ensemble just-in-time Gaussian process regression models, Chemometr. Intell. Lab. Syst. 155 (2016) 170–182.
[16] Y. Liu, Y. Fan, L.C. Zhou, F.J. Jin, Z.L. Gao, Ensemble correntropy-based Mooney viscosity prediction model for an industrial rubber mixing process, Chem. Eng. Technol. 39 (2016) 1804–1812.
[17] O. Chapelle, A. Zien, B. Scholkopf, Semi-supervised Learning, MIT Press, London, 2006.
[18] X. Zhu, Semi-supervised learning literature survey, Comput. Sci. 37 (1) (2008) 63–77.
[19] Z.Q. Ge, B. Huang, Z.H. Song, Mixture semisupervised principal component regression model and soft sensor application, AIChE J. 60 (2) (2014) 533–545.
[20] J. Ji, H.Q. Wang, K. Chen, Y. Liu, N. Zhang, J.J. Yan, Recursive weighted kernel regression for semi-supervised soft-sensing modeling of fed-batch processes, J. Taiwan Inst. Chem. Eng. 43 (2012) 67–76.
[21] W.W. Yan, P.J. Guo, Y. Tian, J.J. Gao, A framework and modeling method of data-driven soft sensors based on semi-supervised Gaussian regression, Ind. Eng. Chem. Res. 55 (27) (2016) 7394–7401.
[22] Z.B. Yan, C.C. Huang, Y. Yao, Semi-supervised mixture discriminant monitoring for chemical batch processes, Chemometr. Intell. Lab. Syst. 134 (5) (2014) 10–22.
[23] J.F. Liu, Liu MJ. ChenYQ, Z.T. Zhao, SELM: semi-supervised ELM with application in sparse calibrated location estimation, Neurocomputing 74 (2011) 2566–2572.
[24] W.J. Zheng, X.J. Gao, Y. Liu, L.M. Wang, J.G. Yang, Z.L. Gao, Industrial Mooney viscosity prediction using fast semi-supervised empirical model, Chemometr. Intell. Lab. Syst. 171 (2017) 86–92.
[25] C. Cheng, M.S. Chiu, A new data-based methodology for nonlinear process modeling, Chem. Eng. Sci. 59 (2004) 2801–2810.
[26] K. Fujiwara, M. Kano, S. Hasebe, A. Takinami, Soft-sensor development using correlation-based just-in-time modeling, AIChE J. 55 (2009) 1754–1765.
[27] Y. Liu, J. Chen, Integrated soft sensor using just-in-time support vector regression and probabilistic analysis for quality prediction of multi-grade processes, J. Process Contr. 23 (2013) 793–804.
[28] Y. Liu, T. Chen, J. Chen, Auto-switch Gaussian process regression-based probabilistic soft sensors for industrial multigrade processes with transitions, Ind. Eng. Chem. Res. 54 (2015) 5037–5047.
[29] W.Y. Jin, Y. Liu, Z.L. Gao, Fast property prediction in an industrial rubber mixing process with local ELM model, J. Appl. Polym. Sci. 134 (41) (2017), https://doi.org/10.1002/app.45391.
[30] G.B. Huang, An insight into extreme learning machines: random neurons, random features and kernels, Cognit. Comput. 6 (2014) 376–390.
[31] G.B. Huang, Z. Bai, L.L.C. Kasun, C.M. Vong, Local receptive fields based extreme learning machine, IEEE Comput. Intell. Mag. 10 (2015) 18–29.
[32] Q. Yu, Y. Miche, E. Eirola, M. van Heeswijk, E. Séverin, A. Lendasse, Regularized extreme learning machine for regression with missing data, Neurocomputing 102 (2013) 45–51.
[33] M. Zhang, X.G. Liu, Z.Y. Zhang, A soft sensor for industrial melt index prediction based on evolutionary extreme learning machine, Chin. J. Chem. Eng. 24 (2016) 1013–1019.
[34] Y.S. Ye, J. Ren, X.H. Wu, G.F. Ou, H.Z. Jin, Data-driven soft-sensor modelling for air cooler system pH values based on a fast search pruned-extreme learning machine, Asia Pac. J. Chem. Eng. 12 (2017) 186–195.
[35] G.H. Golub, M. Heath, G. Wahha, Generalized cross-validation as a method for choosing a good ridge parameter, Technometrics 21 (1979) 215–223.
[36] S. Miyamoto, H. Ichihashi, K. Honda, Algorithms for Fuzzy Clustering, Methods in C-Means Clustering with Applications, Springer, New York, 2008.
[37] Y. Liu, Y. Fan, J. Chen, Flame images for oxygen content prediction of combustion systems using DBN, Energy Fuel. 31 (2017) 8776–8783.
[38] Q. Xuan, B. Fang, Y. Liu, J. Wang, J. Zhang, Y. Zheng, G. Bao, Automatic pearl classification machine based on multistream convolutional neural network, IEEE Trans. Ind. Electron. 65 (2018) 6538–6547.