WILEY | Hindawi

Research Article

Ensemble Just-In-Time Learning-Based Soft Sensor for Mooney Viscosity Prediction in an Industrial Rubber Mixing Process

Huaiping Jin [ID], Jiangang Li, Meng Wang, Bin Qian [ID], Biao Yang, Zheng Li, and Lixian Shi

Department of Automation, Faculty of Information Engineering and Automation,
Kunming University of Science and Technology, Kunming 650500, China

Correspondence should be addressed to Huaiping Jin; jinhuaiping@gmail.com

The lack of online sensors for Mooney viscosity measurement has posed significant challenges for enabling efficient monitoring, control, and optimization of industrial rubber mixing process. To obtain real-time and accurate estimations of Mooney viscosity, a novel soft sensor method, referred to as multimodal perturbation- (MP-) based ensemble just-in-time learning Gaussian process regression (MP-EJITGPR), is proposed by exploiting ensemble JIT learning. This method employs perturbations on similarity measure and input variables for generating the diversity of JIT learners. Furthermore, a set of accurate and diverse JIT learners are built through an evolutionary multiobjective optimization by balancing the accuracy and diversity objectives explicitly. Moreover, all base JIT learners are combined adaptively using a finite mixture mechanism. The proposed method is applied to an industrial rubber mixing process for Mooney viscosity prediction, and the experimental results demonstrate its effectiveness and superiority over traditional soft sensor methods.

1. Introduction

Rubber mixing is a crucial step in rubber and tire industry. The quality of rubber products highly depends on the exact mixing of raw materials and additives. The Mooney viscosity, indicating the molecular weight and viscoelastic behavior of an elastomer, has been recognized as an important quality index for producing nonvulcanized rubbery materials [1, 2]. In most rubber and tire factories, however, the Mooney viscosity can only be determined through manual analysis, which often takes 4~6 h after a batch has been discharged, while the duration of a batch run of mixing process is only about 2~5 min. Therefore, in recent years, soft sensor methods have been widely applied to provide real-time estimations of the Mooney viscosity to obtain the optimal and uniform rubber product quality [3–12].

In general, it is time-consuming and even impossible to build accurate first-principles soft sensors for Mooney viscosity due to the lack of in-depth chemical and physical knowledge of rubber mixing process. Alternatively, data-driven soft sensors have attracted much attention because of the availability of large amounts of data and advanced data mining and analytics tools [13–18]. The early attempts to data-driven soft sensors for quality estimation mainly focus on global modeling techniques, such as multivariate statistical techniques [19, 20], artificial neural networks [3, 21], support vector regression [22], and Gaussian process regression [4, 5]. Recently, deep learning methods have also been introduced to soft sensor applications [23].

However, global soft sensor methods cannot always function well because they may fail to handle local process characteristics effectively and perform model adaptation efficiently. Thus, in practical applications, local learning-based soft sensor methods are more appealing for providing accurate predictions. Compared to global modeling, local learning exhibits two outstanding advantages. First, global soft sensors are usually based on the underlying assumption of a constant operating phase and conditions throughout the entire duration of a production process, whereas industrial processes are often characterized by strong nonlinearity, and multiple operation phases or modes. Thus, from the process industry viewpoint, local learning is more suitable for

handling complex process characteristics. Second, since local learning is completely localized, the local models can be built and updated independently of each other, which greatly simplifies the incremental adaptation, inclusion, or removal of local models and receptive fields.

Generally, there are two categories of local learning methods: ensemble methods [24–27] and just-in-time learning (JIT) methods [6, 7, 9, 12]. These methods employ the divide-and-conquer philosophy to model the relationships between the inputs and output by building a set of locally valid models. In particular, JIT learning, known as a representative local learning paradigm, has gained growing attention in soft sensor applications for Mooney viscosity estimation due to its strong capability of handling nonlinearity, time-varying behavior, multiphase and multimode process characteristics, etc. [9, 11, 12]. However, traditional JIT soft sensors attempt to build a globally optimal encapsulation of local modeling techniques, similarity measures, input variables, model hyper-parameters, etc., while the diversity of JIT learning is ignored. To tackle this problem, various ensemble JIT learning (EJIT) soft sensors have been developed [6, 8, 25, 28–31].

The basic idea of EJIT modeling is to build multiple component JIT learners and then combine their predictions. For instance, Liu et al. (2012) [28] employed heterogeneous predictive models to build base JIT models and then combined them through a simple averaging rule. Liu and Gao (2015) [6] developed an EJIT soft sensor by using diverse similar data sets, which are obtained by assigning diverse hyperparameters to the support vector clustering for outlier detection. Kaneko and Funatsu (2016) [25] developed an ensemble locally weighted partial least squares (LWPLS) soft sensor, where diverse subsets are first built using moving window method and then multiple of the most relevant ones to the query state are selected to build diverse LWPLS models, which are finally integrated via Bayes' theorem. Liu et al. (2015) [29] built an EJIT kernel learning framework through perturbing the hyperparameters of local learning methods. Yuan et al. (2018) [30] developed an EJIT soft sensor by using different similarity measures. Besides, we proposed an EJIT soft sensor by perturbing the input features for building diverse input subspaces [8]. Recently, we developed an EJIT soft sensor by employing multiple weighted Euclidean distance- (WED-) based similarity measures, which are optimized through an EMO approach [31]. These studies show that it is feasible and effective to enhance the prediction accuracy of JIT soft sensors by introducing ensemble learning.

However, it remains challenging to build high-performance EJIT soft sensors due to the following issues. First, many current EJIT soft sensor methods only consider single perturbation, such as perturbing training data [25], similarity measure [30, 31], perturbing input variables [8], perturbing local modeling technique [28], or perturbing model parameters [29]. In practice, the diversity of JIT learning is often originated from multiple factors. Second, most of the current EJIT methods construct base JIT learners in a heuristic way. In such situations, the accuracy and diversity objectives of JIT learners are difficult to achieve a good tradeoff. Finally, most methods employ nonadaptive weightings for the combination of base JIT learners, which will limit the prediction performance of EJIT soft sensors.

To address the aforementioned issues, a novel EJIT soft sensor method, referred to as multimodal perturbation-based EJITGPR (MP-EJITGPR), is proposed for enabling accurate predictions of Mooney viscosity. This method works through integrating perturbation on similarity measures and perturbation on input variables together. With the multimodal perturbation mechanism, a set of accurate and diverse JIT learners are built by balancing the accuracy and diversity objectives explicitly through an evolutionary multiobjective optimization (EMO) approach. Then, a finite mixture mechanism- (FMM-) based weighting stagey is used to achieve an adaptive combination of base learners. In summary, the main contributions of this study are as follows:

(1) A multimodal perturbation mechanism is proposed by utilizing heterogeneous similarity measures and building diverse input subspaces, which allows enhancing the diversity of base JIT learners efficiently

(2) The generation of accurate and diverse JIT learners is formulated as a multiobjective optimization problem and then solved by an EMO approach

(3) The combination of base JIT learners is achieved through the finite mixture mechanism, which enables adaptive assignments of weights

(4) A novel EJIT soft sensor modeling framework is built by integrating the multimodal perturbation mechanism-based diversity creation, the EMO-based generation of base JIT learners, and the FMM-based adaptive combination of base JIT learners

The rest of the paper proceeds as follows. Section 2 briefly introduces JIT learning and Gaussian process regression. Section 3 details the proposed MP-EJITGPR soft sensor method and its implementation procedure. The application of MP-EJITGPR for Mooney viscosity prediction in an industrial rubber mixing process is reported in Section 4. Finally, conclusions are drawn in Section 5.

2. Preliminaries

2.1. Just-In-Time Learning. JIT learning [32], also known as lazy learning [33] and locally weighted learning [34], refers to a family of algorithms in which all historical data are stored in a database and local models are built dynamically by retrieving the most similar data to the query state. Compared to conventional global modeling methods, JIT learning has the following features:

(1) All available modeling data are stored at a database. And only those samples most similar to the query point are used for modeling for each run of prediction.

(2) Only when an estimation is required, a local model is built dynamically based on samples with high similarities to the query point.

(3) The constructed local model is discarded after the estimation is given.

2.2. Gaussian Process Regression.

A Gaussian process is a collection of random variables, any finite number of which follows joint Gaussian distributions [35]. Considering a data set $D = \{\mathbf{X}, \mathbf{y}\} = \{\mathbf{x}_i, y_i\}_{i=1}^n$, 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.

A Gaussian process is completely specified by its mean function $m(\mathbf{x})$ and covariance function $C(\mathbf{x}, \mathbf{x}')$:

$$f(\mathbf{x}) \sim \mathrm{GP}\big(m(\mathbf{x}), C(\mathbf{x}, \mathbf{x}')\big). \tag{2}$$

Since the modeling data is usually normalized to be zero mean, the output observations follow a Gaussian distribution as

$$\mathbf{y} \sim \mathrm{GP}(0, \mathbf{C}), \tag{3}$$

where $\mathbf{C}$ is an $n \times n$ covariance matrix with $\mathbf{C}_{ij} = C(\mathbf{x}_i, \mathbf{x}_j)$ representing the ijth element. In this study, a Matérn covariance function with noise term is adopted:

$$C(\mathbf{x}_i, \mathbf{x}_j) = \sigma_f^2 \left(1 + \frac{\sqrt{3}\, \|\mathbf{x}_i - \mathbf{x}_j\|}{l} \right) \exp\left(-\frac{\sqrt{3}\, \|\mathbf{x}_i - \mathbf{x}_j\|}{l} \right) + \sigma_n^2 \delta_{ij}, \tag{4}$$

where $\Theta = \{\sigma_f^2, l, \sigma_n^2\}$ are the hyperparameters, l is the input scale, σ_n^2 is the noise variance, σ_f^2 is the output scale, and $\delta_{ij} = \begin{cases} 1; i = j \\ 0; i \neq j \end{cases}$. The hyperparameters Θ are determined by Bayesian inference.

Given a query data $\mathbf{x}_{\mathrm{new}}$, the training outputs $\mathbf{y}$ and the test output y_{new} follow a joint Gaussian distribution as follows:

$$\begin{bmatrix} \mathbf{y} \\ y_{\mathrm{new}} \end{bmatrix} \sim N\left(0, \begin{bmatrix} \mathbf{C} & \mathbf{k}_{\mathrm{new}} \\ \mathbf{k}_{\mathrm{new}}^{\mathrm{T}} & C(\mathbf{x}_{\mathrm{new}}, \mathbf{x}_{\mathrm{new}}) \end{bmatrix} \right), \tag{5}$$

where $\mathbf{k}_{\mathrm{new}} = [C(\mathbf{x}_{\mathrm{new}}, \mathbf{x}_1), \dots, C(\mathbf{x}_{\mathrm{new}}, \mathbf{x}_n)]^{\mathrm{T}}$. Then, the prediction output $\hat{y}_{\mathrm{new}}$ and variance σ_{new}^2 can be calculated as

$$\begin{cases} \hat{y}_{\mathrm{new}} = \mathbf{k}_{\mathrm{new}}^{\mathrm{T}} \mathbf{C}^{-1} y, \\ \sigma_{\mathrm{new}}^2 = C(\mathbf{x}_{\mathrm{new}}, \mathbf{x}_{\mathrm{new}}) - \mathbf{k}_{\mathrm{new}}^{\mathrm{T}} \mathbf{C}^{-1} \mathbf{k}_{\mathrm{new}}, \end{cases} \tag{6}$$

3. Proposed MP-EJITGPR Soft Sensor

In this section, a multimodal perturbation-based ensemble just-in-time learning Gaussian process regression (MP-EJITGPR) is presented. First, data preprocessing is conducted on the three-way data matrices of industrial rubber mixing process. Then, heterogeneous similarity measures are defined. Furthermore, by introducing the multimodal perturbation mechanism, a set of accurate and diverse JIT learners, which are equipped with heterogeneous similarity measures and diverse subspaces, are generated through an EMO approach. Next, a finite mixture mechanism is employed to achieve an adaptive combination of base JIT learners. Finally, the implementation procedure of MP-EJITGPR is provided.

3.1. Data Preprocessing.

Typically, the online data of those easy-to-measure variables in industrial rubber mixing process can be arranged in a three-way matrix $\underline{X}(I \times J \times K)$ consisting of J process variables measured at K points for I batches. Meanwhile, the quality variable (i.e., Mooney viscosity), which is only available at the end of the batch, i.e., time point K, can be expressed as $\mathbf{y}_K(I \times 1)$. Before soft sensor modeling, it is desirable to perform essential processing. First, the process data is first preprocessed using a simple 3σ rule for outlier detection. Then, the three-way data matrix $\underline{X}(I \times J \times K)$ is unfolded into a two-way matrix, which allows utilizing the standard regression techniques for building the predictive model between online measured variables and the end-use quality variable. Generally, $\underline{X}$ can be unfolded in six different ways [36], among which batchwise unfolding is employed in this study, as illustrated in Figure 1. In practice, this way of unfolding has been actually recognized to be the most meaningful one for analysis and monitoring of batch processes. With this way of unfolding, all potential input variables at different time instants can be obtained for predicting the final quality variable. Moreover, since the dimensions and magnitudes of various process variables are significantly different from each other, another crucial step to guarantee the reliability and accuracy of soft sensors is data normalization, which is achieved by scaling the unfolded data matrices to zero mean and unit variance in this study.

3.2. Definition of Heterogeneous Similarity Measures.

Similarity measure plays a central role in JIT modeling. In contrast to traditional modeling methods utilizing all available data, the JIT method is to construct a local model based on a small data set with high similarities to the query data. Thus, the key to building highly accurate JIT soft sensors is to define appropriate similarity metrics. Despite the availability of various similarity measures, it is hopeless to pursue a similarity evaluation criterion, which is consistently better than other metrics across different application scenarios. Consequently, it is a common practice to select one of the best similarity measures for a given task, which is usually time-consuming and even impossible. In practice, different similarity measures can provide different insights into similarity evaluation between data points. Thus, in this work, heterogeneous similarity measures will be combined for JIT learning, including Euclidean distance-(ED-) based similarity, cosine similarity (cosine), covariance weighted distance- (CWD-) based similarity, and correlation coefficient- (CC-) based similarity.

The ED similarity measure is the most commonly used metric for JIT learning due to its simplicity and efficiency. It is defined based on the Euclidean distance between points in space; that is,
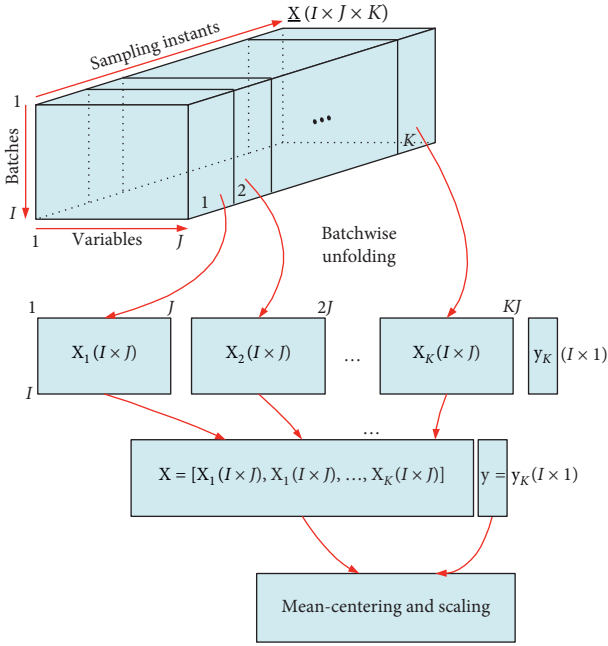
FIGURE 1: Batchwise unfolding of three-way data matrices.

$$\omega_i = \exp\left(-\frac{d_i}{\sigma_d \varphi}\right),$$

$$d_i = \|\mathbf{x}_i - \mathbf{x}_{\text{new}}\|,$$

(7)

where σ_d is the standard deviation of d_i $(i = 1, 2, \ldots, n)$ and φ is a localization parameter.

One disadvantage of ED similarity is that the differences among input variables are ignored. To address this issue, various weighted distance-based similarity measures have been proposed, among which the CWD similarity measure is defined by considering the relationships among input variables and among input and output variables. [37] That is, CWD similarity is defined by using the weighted distance metric:

$$d_i = \sqrt{(\mathbf{x}_i - \mathbf{x}_{\text{new}})^{\text{T}} \text{H} (\mathbf{x}_i - \mathbf{x}_{\text{new}})},$$

$$\text{H} = \frac{(\mathbf{X}^{\text{T}}\mathbf{y})^{\text{T}} (\mathbf{X}^{\text{T}}\mathbf{y})}{\|\mathbf{X}^{\text{T}}\mathbf{y}^2\|},$$

(8)

where H denotes a weighting matrix, and $\mathbf{X}$ and $\mathbf{y}$ are the input and output matrices, respectively.

Alternatively, by exploiting the angle between two vectors in space, the cosine similarity measure can be defined as

$$\omega_i = \exp\left(\frac{\cos_i}{\sigma_{cos} \varphi}\right),$$

$$\cos_i = \frac{\mathbf{x}_i^T \mathbf{x}_{\text{new}}}{\|\mathbf{x}_i\|\|\mathbf{x}_{\text{new}}\|},$$

(9)

where $\cos_i$ denotes the cosine value of the angle between two vectors.

In addition to the distance and angle criteria, the relevance between two vectors can also be used to evaluate the similarity between samples. For the sake of simplicity, and without loss of generality, the frequently used correlation coefficient (CC) criterion is used to define similarity measure as follows:

$$\omega_i = \exp\left(\frac{r_i}{\sigma_r \varphi}\right),$$

$$r_i = \frac{\text{Cov}(\mathbf{x}_i, \mathbf{x}_{\text{new}})}{\sqrt{\text{Var}(\mathbf{x}_i)\text{Var}(\mathbf{x}_{\text{new}})}},$$

(10)

where $\text{Cov}(\cdot, \cdot)$ and $\text{Var}(\cdot)$ are used to compute the covariance and variance, respectively.

It is noteworthy that the above similarity measures are defined from different points of view, thus behaving differently in different applications. So, one promising solution towards further improving the prediction performance of JIT learning soft sensors is to use heterogeneous similarity measures together, which has not been well investigated and will be discussed in the following section.

3.3. Generation of Base JIT Learners through Evolutionary Multiobjective Optimization. When considering JIT learning as a base learner for ensemble learning, EJIT modeling is essentially one ensemble method. It has been proven both theoretically and empirically that both accuracy and diversity of base learners are crucial to guarantee high ensemble performance [38]. According to the famous bias-variance decomposition [39] and error-ambiguity decomposition [40] theories, the more accurate and the more diverse base learners are, the better the ensemble is. Hence, the success of developing high-performance EJIT soft sensors lies in generating accurate and diverse base JIT learners.

Among various ensemble learning soft sensors, perturbing training data remains dominant for creating diversity, such as clustering [41], moving window [24, 27], bootstrapping sampling [42, 43], and sequential sampling [44]. However, such data manipulation strategies do not always function well for EJIT modeling because JIT learning only relies on a small subset of relevant samples for each run of prediction and is less sensitive to the randomness injected to the database. Moreover, the perturbation on input variables is often ignored in developing ensemble soft sensors. In addition, many of the current methods for base learner generation are based on heuristic mechanisms without measuring or ensuring diversity among base learners explicitly.

Thus, in this study, the diversity generation is achieved through the multimodal perturbation mechanism, i.e., perturbing similarity measure and input variables together. Then, the generation of accurate and diverse base JIT learners is formulated as a multiobjective optimization problem (MOP). Finally, the MOP problem is solved by using an EMO approach, which leads to a tradeoff between accuracy and diversity objectives. In the following, the

decision variables, optimization objectives, and the adopted EMO approach are detailed.

Suppose M JIT learners, each of which is characterized by one similarity measure and one input subspace, are required for constructing an EJIT model. Since M heterogeneous similarity measures have been defined in advance, the decision variables $\mathbf{z}$ only include the selection variables, which indicate whether an input variable is selected or not in order to form a subspace. That is, the decision vector $\mathbf{z}$ can be expressed as follows:

$$\mathbf{z} = \left[\overbrace{\theta_1^1, \theta_2^1, \cdots, \theta_D^1}^{\text{Simi 1, JIT learner 1}}, \ldots, \overbrace{\theta_1^m, \theta_2^m, \cdots, \theta_D^m}^{\text{Simi } m, \text{ JIT learner } m}, \ldots, \overbrace{\theta_1^M, \theta_2^M, \cdots, \theta_D^M}^{\text{Simi } M, \text{ JIT learner } M} \right], \tag{11}$$

where M and D denote the number of base JIT learners and that of potential input variables for building subspaces, respectively. θ is a binary variable, which indicates inclusion and exclusion of an input variable by using "1" and "0", respectively.

Furthermore, the accuracy and diversity objectives are defined. In this study, the ensemble accuracy is given as the average of individual accuracies. Given the training data set $D_{\text{trn}} = \{\mathbf{X}_{\text{trn}}, \mathbf{y}_{\text{trn}}\}$ and an independent validation set $D_{\text{val}} = \{\mathbf{X}_{\text{val}}, \mathbf{y}_{\text{val}}\}$, the accuracy objective is estimated as

$$\text{RMSE}_{\text{avg,val}} = \frac{\sum_{m=1}^{M} \text{RMSE}_{\text{val}}^m}{M}, \tag{12}$$

where $\text{RMSE}_{\text{val}}^m$ denotes the root-mean-squared error (RMSE) on the validation set for the m th JIT learner.

In comparison with the accuracy objective, measuring ensemble diversity is not straightforward. Up to now, there is no generally accepted formal formulation and measures for ensemble diversity. Thus, in this study, the standard deviation of individual prediction outputs is used to evaluate the ensemble diversity. By applying the base JIT learners, $\{f_{\text{JIT}}^m\}_{m=1}^M$, the prediction outputs of individual JIT learners can be given as

$$\widehat{\mathbf{Y}}_{\text{val}} = \left[\widehat{\mathbf{y}}_{\text{val}}^1, \cdots, \widehat{\mathbf{y}}_{\text{val}}^m, \cdots, \widehat{\mathbf{y}}_{\text{val}}^M \right], \tag{13}$$

where $\widehat{\mathbf{y}}_{\text{val}}^m$ is a column vector denoting the prediction output vector on the validation data using the mth JIT learner.

Let italic R be a row vector denoting the prediction outputs from various base JIT learners for the i th query data in validation set. Then, the ensemble diversity can be defined as follows:

$$\sigma_{\text{avg,val}} = \frac{\sum_{i=1}^{N_{\text{val}}} \sigma_{\text{val},i}}{N_{\text{val}}}, \tag{14}$$

where N_{val} is the number of validation samples, and $\sigma_{\text{val},i}$ is the standard deviation of $\widehat{\mathbf{y}}_{\text{val},i}$.

To build accurate and diverse base JIT learners, small $\text{RMSE}_{\text{avg,val}}$ and large $\sigma_{\text{avg,val}}$ are desirable. Thus, generating accurate and diverse JIT learners can be formulated as a biobjective optimization problem:

$$\min \left\{ \text{RMSE}_{\text{avg,val}}, 1/\sigma_{\text{avg,val}} \right\}. \tag{15}$$

To solve the MOP problem in equation (15), one of the most famous EMO algorithms, i.e., NSGA-II (nondominated sorting genetic algorithm II), is employed.

Details about the NSGA-II algorithm can be found in [45]. First, the decision vector $\mathbf{z}$ is coded as a binary string and an initial population is created. Then, the Pareto-optimal solutions can be obtained by performing the following procedure:

(i) Step 1: generate an initial population with N_{pop} individuals.

(ii) Step 2: evaluate each individual in the population by calculating the accuracy and diversity objectives, i.e., $\text{RMSE}_{\text{avg,val}}$ and $1/\sigma_{\text{avg,val}}$.

(iii) Step 3: repeat the following steps until the stopping condition is reached.

(a) Assign a front number to all solutions by using nondominated sorting and calculated the crowding distance.

(b) Generate an offspring population by using the binary tournament selection, recombination, and mutation operations.

(c) Evaluate each solution as that in Step (2).

(d) Merge the current and offspring populations such that elitism is ensured.

(e) Sort the solutions from the merged population according to the nondominated sorting method.

(f) Choose the first N_{pop} solutions from the merged population and increase the generation counter.

(iv) Step 4: find the Pareto-optimal solutions from the combined population in the last generation by applying the nondominated sorting method.

The outcome of this step is a set of Pareto-optimal solutions, one of which is selected for the ensemble construction of MP-EJITGPR modeling.

3.4. Adaptive Combination of Base JIT Learners by Finite Mixture Mechanism.

With the utilization of multimodal perturbation and EMO optimization, a group of accurate and diverse JIT learners $\{f_{\text{JIT}}^m\}_{m=1}^M$ can be obtained, where f_{JIT} is built by the GPR method in this paper. When a query data is requested to be predicted, each base JIT learner makes a prediction for the output variable. To get the final predictions, these individual predictions have to be combined.

Generally, there are two classes of combination methods: nonadaptive and adaptive weightings. In the former, weights

assigned to base learners remain unchanged once deployed into the real-life operation, whereas in the latter weights are assigned adaptively to accommodate the query process state. The simplest nonadaptive weighting method is the simple averaging rule, which provides the average of individual predictions as the final prediction. Another popular non-adaptive weighting is to determine weights according to their prediction capability on training set or validation set. For instance, weights can be determined by using linear regression methods, such as PCR and PLS. Besides, the combination of base learners can be achieved by learning, i.e., stacking, which usually leads to a nonadaptive combination model. However, a deficiency of nonadaptive combination methods is that they tend to assign larger weights to the models that exhibit excellent prediction on the training set or validation set, which may lead to overestimation or underestimation of weights and thus deteriorate the generalization capability of ensemble models. Therefore, adaptive combination strategies are highly appealing.

In this study, a finite mixture mechanism-based adaptive weighting method is proposed to achieve the combination of base learners. For a new query data $\mathbf{x}_{\text{new}}$, the predictive distribution of the mth output $y_{m,\text{new}}$ of the target variable is estimated from the mth JITGPR model and $y_{m,\text{new}}$ follows a Gaussian distribution as follows:

$$y_{m,\text{new}} \sim \mathcal{N}\left(\mathbb{E}\left(y_{m,\text{new}}\right), \text{Var}\left(y_{m,\text{new}}\right)\right), \ m = 1, 2, \cdots, M, \quad (16)$$

where $\mathbb{E}\left(y_{m,\text{new}}\right)$ and $\text{Var}\left(y_{m,\text{new}}\right)$ are the prediction mean and variance, respectively.

Assume the local predictions $y_{1,\text{new}}, \ldots, y_{m,\text{new}}, \ldots, y_{M,\text{new}}$ are independent realizations of the overall output variable y_{new}. That is to say, y_{new} arises from a finite mixture distribution of $y_{1,\text{new}}, \ldots, y_{m,\text{new}}, \ldots, y_{M,\text{new}}$. Thus, by applying the finite mixture mechanism [26, 46], the mean and variance of the predictive distribution of the target variable can be estimated by combining all local predictions:

$$\begin{cases} \widehat{y}_{\text{new}} = \sum_{m=1}^{M} \omega_{m,new} \widehat{y}_{m,\text{new}}, \\ \sigma_{\text{new}}^2 = \sum_{m=1}^{M} \omega_{m,new} \left\{ \sigma_{m,\text{new}}^2 + \left(\widehat{y}_{m,\text{new}} - \widehat{y}_{\text{new}}\right)^2 \right\}, \end{cases} \quad (17)$$

where $\widehat{y}_{m,\text{new}}$ and $\sigma_{m,\text{new}}^2$ represent the prediction output and variance using the m th JITGPR model, respectively; and ω_m denotes the mixture weights satisfying the following constraints:

$$0 \leq \omega_{m,\text{new}} \leq 1, \ \sum_{m=1}^{M} \omega_{m,\text{new}} = 1. \quad (18)$$

Since the prediction uncertainty can effectively indicate the confidence level of the output predictions, we assume the mixture weights are inversely proportional to the prediction variances from individual JITGPR models. Thus, $\omega_{m,\text{new}}$ can be estimated as follows:

$$\omega_{m,\text{new}} = \frac{\left(\widehat{y}_{m,\text{new}}/\sigma_{m,\text{new}}\right)^p}{\sum_{m=1}^{M}\left(\widehat{y}_{m,\text{new}}/\sigma_{m,\text{new}}\right)^p}, \quad (19)$$

where p is an adjustable parameter.

The proposed FMM-based combination strategy allows predictions from individual JITGPR models to be combined adaptively at each run of prediction.

3.5. Implementation Procedure.
The step-by-step procedure of the proposed MP-EJITGPR soft sensor method for Mooney viscosity prediction is summarized below and the schematic diagram of this approach is illustrated in Figure 2.

3.5.1. Offline Optimization Phase

(a) Collect the process data of the batch process for model training and validation

(b) Data processing is performed, including outlier detection, data unfolding, mean-centering, and scaling

(c) Formulate the generation of accurate and diverse JITGPR models as a multiobjective optimization problem (MOP)

(d) Solve the MOP problem using the EMO approach, i.e., NSGA-II

(e) By using one of the best-performing Pareto-optimal solutions, a set of JITGPR models characterized by heterogeneous similarity measures and diverse input subspaces are constructed

3.5.2. Online Prediction Phase

(a) For any query data, M relevant subsets are obtained by using the heterogeneous similarity measures and diverse input subspaces.

(b) M JITGPR models are constructed and further provide the prediction outputs and variances of the target variable.

(c) The final prediction output and variance are produced using the proposed FMM-based combination method and then those built JITGPR models are discarded. When new query data comes, go to Step (a).

It is worth noting that the computational load of the proposed MP-EJITGPR method is mainly focused on the offline optimization phase, especially the NSGA-II-based EMO optimization. However, once the learning configurations for generating diverse JIT models have been obtained, the online prediction for new test samples can be conducted fast. This is because, on one hand, if the similarity measures are defined appropriately, only a small number of samples are selected for online local modeling, which enables fast training of diverse JIT models for each query data. On the other hand, the finite mixture mechanism-based adaptive combination is very efficient because only simple calculations are involved. Therefore, the proposed approach can be applied for providing real-time estimations of Mooney viscosity in an industrial rubber mixing process.
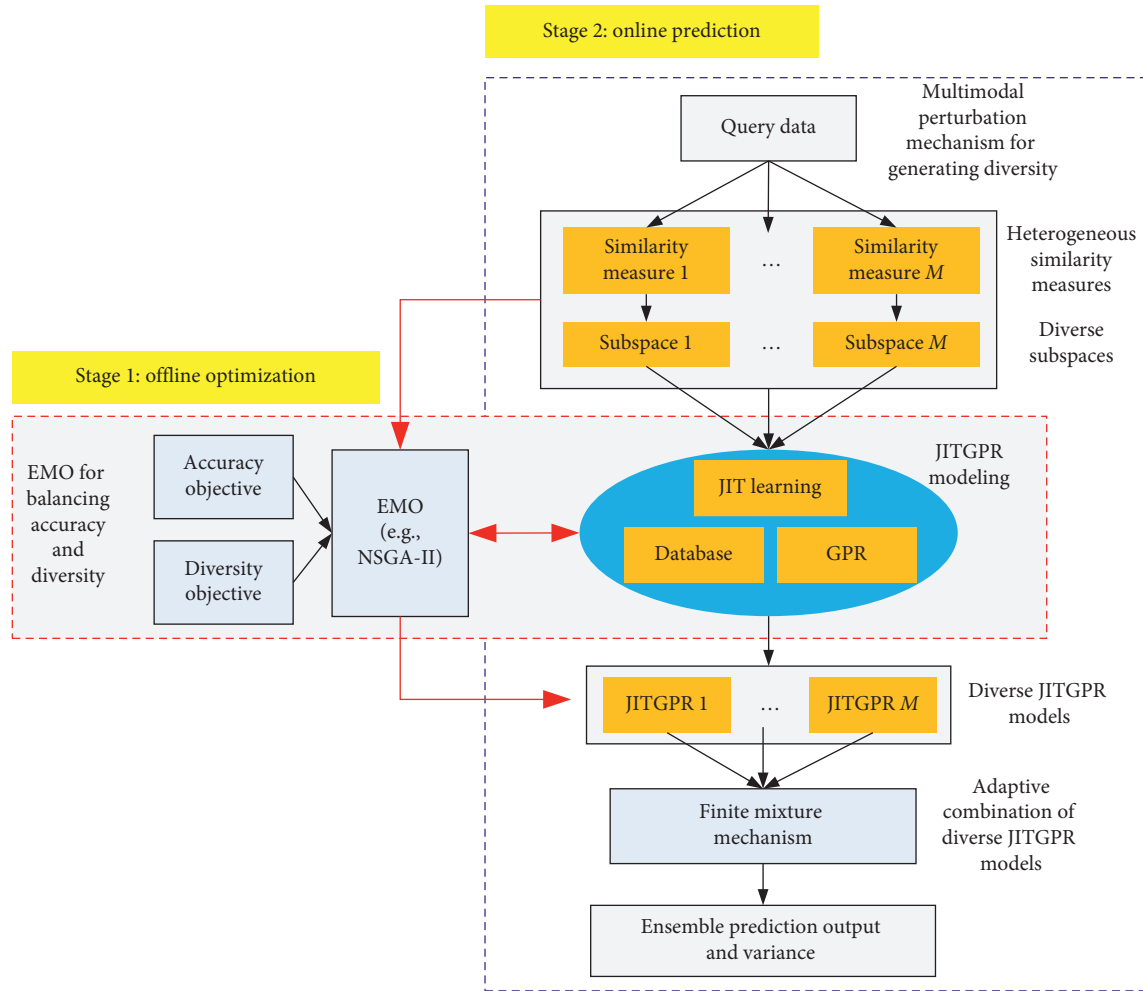
Figure 2: Schematic diagram of the proposed MP-EJITGPR soft sensing approach.

4. Application to an Industrial Rubber Mixing Process

The effectiveness and superiority of the proposed MP-EJITGPR soft sensor for Mooney viscosity prediction are demonstrated through an industrial rubber mixing process in China. The methods for comparison are as follows:

(1) PLS (partial least squares regression): global PLS model

(2) GPR (Gaussian process regression): global GPR model

(3) GMMGPR: Gaussian mixture model- (GMM-) based ensemble GPR models

(4) JITGPR (ED similarity): JIT learning GPR using ED similarity measure

(5) JITGPR (cosine similarity): JIT learning GPR using cosine similarity measure

(6) JITGPR (CWD similarity): JIT learning GPR using CWD similarity measure

(7) JITGPR (CC similarity): JIT learning GPR using CC similarity measure

(8) SP-EJITGPR (SAR): similarity perturbation-based EJITGPR with simple averaging rule (SAR) for combination

(9) SP-EJITGPR (PLS stacking): similarity perturbation-based EJITGPR with PLS stacking for combination

(10) SP-EJITGPR (GPR stacking): similarity perturbation-based EJITGPR with GPR stacking for combination

(11) SP-EJITGPR (FMM): similarity perturbation-based EJITGPR with the finite mixture mechanism for combination

(12) MP-EJITGPR (SAR): multimodal perturbation-based EJITGPR with a simple averaging rule for combination

(13) MP-EJITGPR (PLS stacking): multimodal perturbation-based EJITGPR with PLS stacking for combination

(14) MP-EJITGPR (GPR stacking): multimodal perturbation-based EJITGPR with GPR stacking for combination

(15) MP-EJITGPR (FMM) (the proposed method): multimodal perturbation-based EJITGPR with the finite mixture mechanism for combination

A total of 15 soft sensor methods are involved, and their characteristics are compared in Table 1. These methods can be roughly categorized into five classes: (1) global methods, i.e., PLS and GPR; (2) ensemble methods through training data perturbation, i.e., GMMGPR; (3) JITGPR methods using single similarity measure; (4) EJITGPR methods using similarity perturbation and ensemble learning, i.e., SP-EJITGPR; (5) EJITGPR methods using multimodal perturbation and ensemble learning, i.e., MP-EJITGPR. For those EJIT methods, four types of combination methods are investigated, i.e., simple averaging rule (SAR), PLS stacking, GPR stacking, and finite mixture mechanism (FMM).

The modeling data are split into three parts: training set for modeling learning, validation set for parameter tuning and EMO optimization, and testing set for model evaluation. Moreover, some critical parameters should be predetermined. In detail, the number of principal components for PLS is selected based on the prediction accuracy on the validation set. The local modeling size l and the adjustable parameter p in equation (19) are determined by trial and error. In addition, the optimization settings for NSGA-II are given as follows: population size $N_{pop} = 100$, and maximum generation size $N_{gen}^{max} = 100$.

To assess the prediction performance of soft sensors, three indices, namely, root-mean-square error (RMSE), relative RMSE (RRMSE), and coefficient of determination (R^2), are used:

$$RMSE = \sqrt{\frac{1}{n_{test}} \sum_{i=1}^{n_{test}} (\hat{y}_i - y_i)^2},$$

$$RRMSE = \sqrt{\frac{1}{n_{test}} \sum_{i=1}^{n_{test}} \left(\frac{\hat{y}_i - y_i}{y_i}\right)^2} \times 100\%, \qquad (20)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n_{test}} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n_{test}} (y_i - \overline{y})^2},$$

where y_i and $\hat{y}_i$ denote the actual and predicted outputs, respectively; $\overline{y}$ represents the mean value; and n_{test} is the number of testing samples.

The computer configurations for experiments are as follows. OS : Windows 10 (64 bit); CPU : Intel (R) Core(TM) i7-6700 (3.4 GHz × 2); RAM : 8G byte; and the simulation software is MATLAB R2016a. The MATLAB codes for running GPR regression can be downloaded from the website: http://www.gaussianprocess.org/gpml/code/matlab/doc/.

4.1. Process Description. Rubber mixing is a crucial step in the rubber and tire industry. [7, 8] The industrial rubber mixing process in this study is practiced in a Chinese tire company. The industrial production site is illustrated in Figure 3. The rubber mixing process lasts for 2 min, during which various raw materials are fed into the raw rubber to produce synthetic rubber according to the technical formula and then various complex chemical reactions take place in an internal mixer. Generally, Mooney viscosity is a crucial index for monitoring the product quality of the rubber mixing process. However, in a practical production process, Mooney viscosity is only measured through a viscometer with a large delay of 4~6 h after one batch has discharged. Consequently, it becomes challenging to assure the optimal and uniform quality of mixed rubber. Fortunately, soft sensor technology provides the possibility of estimating Mooney viscosity in real-time. Thus, we attempt to build a high-performing soft sensor for Mooney viscosity prediction in this study. The process variables used for soft sensor modeling include temperature in the mixer chamber, motor power, ram pressure, motor speed, and energy.

4.2. Prediction Results of Mooney Viscosity. The modeling data have been collected from DCS system and laboratory analysis and are then preprocessed by using a simple 3σ rule for outlier detection, batchwise unfolding, zero mean centering, and one variance scaling. With a sampling interval of 2 s, a total of 1172 batches are selected from three internal mixers and are further divided into three sets: 822 batches as the training set, 175 batches as the validation set, and 175 batches as the testing set. By considering the time instants 0 s, 14 s, 18 s, 22 s, . . ., 118 s, a total of 140 delayed and nondelayed variables are obtained as potential input variables and the Mooney viscosity is chosen as the output variable.

The prediction results of Mooney viscosity from different soft sensor methods are presented in Table 2. It is readily observed that PLS leads to the poorest prediction performance among those methods in terms of RMSE, RRMSE, and R^2. This is mainly because PLS cannot effectively handle the nonlinearity of rubber mixing process. In comparison, other nonlinear soft sensor methods achieve much better prediction accuracy than PLS. Though GPR obtains significant accuracy improvement, it still produces high prediction errors due to its failure in dealing with local process characteristics. Instead of relying on a global model, GMMGPR and various JITGPR methods employ local learning philosophy, thus obtaining much better performance than global GPR. Although GMMGPR performs well in this case study, JITGPR methods are more appealing to provide better prediction performance.

However, the prediction accuracy of JITGPR methods is highly related to the similarity measure definition. As can be seen in Table 2, different prediction performance is obtained by using different similarity measures. In real applications, it is difficult to determine which similarity measure performs best in advance. Thus, a promising idea is to fully exploit the advantages of multiple similarity measures for JIT learning by using ensemble methods. As expected, by introducing ensemble learning, SP-EJITGPR can deliver better prediction results than single similarity measure-based JITGPR methods. It is noteworthy that, however, inappropriate

TABLE 1: Characteristics of soft sensor methods for comparison.

No.	Method	Model structure	Learning type	Diversity generation mechanism
1	PLS	Single	Global	—
2	GPR	Single	Global	—
3	GMMGPR	Ensemble	Local	Training data perturbation
4	JITGPR (ED similarity)	Single	Local	—
5	JITGPR (cosine similarity)	Single	Local	—
6	JITGPR (CWD similarity)	Single	Local	—
7	JITGPR (CC similarity)	Single	Local	—
8	SP-EJITGPR (SAR)	Ensemble	Local	Similarity perturbation
9	SP-EJITGPR (PLS stacking)	Ensemble	Local	Similarity perturbation
10	SP-EJITGPR (GPR stacking)	Ensemble	Local	Similarity perturbation
11	SP-EJITGPR (FMM)	Ensemble	Local	Similarity perturbation
12	MP-EJITGPR (SAR)	Ensemble	Local	Multimodal perturbation
13	MP-EJITGPR (PLS stacking)	Ensemble	Local	Multimodal perturbation
14	MP-EJITGPR (GPR stacking)	Ensemble	Local	Multimodal perturbation
15	MP-EJITGPR (FMM)	Ensemble	Local	Multimodal perturbation

FIGURE 3: Industrial rubber mixing process.

TABLE 2: Comparison of Mooney viscosity prediction results using different soft sensors ($l = 10$ for JIT learning).

No.	Method	RMSE	RRMSE (%)	R^2
1	PLS	7.3298	11.7026	0.8002
2	GPR	4.2628	5.9354	0.9324
3	GMMGPR	3.4606	4.8077	0.9555
4	JITGPR (ED similarity)	3.1561	4.2700	0.9630
5	JITGPR (cosine similarity)	3.2053	4.3494	0.9618
6	JITGPR (CWD similarity)	3.6552	5.2370	0.9503
7	JITGPR (CC similarity)	3.2029	4.3313	0.9618
8	SP-EJITGPR (SAR)	3.2127	4.4105	0.9616
9	SP-EJITGPR (PLS stacking)	3.9916	5.4067	0.9407
10	SP-EJITGPR (GPR stacking)	3.7792	5.0508	0.9469
11	SP-EJITGPR (FMM)	3.0670	4.2073	0.9650
12	MP-EJITGPR (SAR)	4.3769	6.0332	0.9288
13	MP-EJITGPR (PLS stacking)	4.1966	5.3943	0.9345
14	MP-EJITGPR (GPR stacking)	3.7819	4.9735	0.9468
15	MP-EJITGPR (FMM)	**2.9202**	**3.9085**	**0.9683**

ED: Euclidean distance; CWD: covariance weighted distance; CC: correlation coefficient; SAR: simple averaging rule; FMM: finite mixture mechanism; SP: similarity perturbation; and MP: multimodal perturbation.

combination methods can lead to performance degradation instead of improvements. Among the compared combination methods, the proposed FMM-based combination successfully obtains significant performance enhancement, while simple averaging rule, PLS stacking, and GPR stacking lead to performance degradation. These results reveal that the integration of heterogeneous similarity measures and the FMM-based adaptive combination method significantly allows improving the prediction accuracy of JIT learning soft sensors.

Apart from the similarity measure, input variable selection is also critical to guarantee the performance of JIT learning. Thus, it is interesting to explore whether EJITGPR model using only similarity perturbation can be further improved or not by performing perturbations on similarity measure and input variables simultaneously. As we have expected, when the FMM-based adaptive combination is employed, EJITGPR model using multimodal perturbation, i.e., MP-EJITGPR (FMM), performs better than SP-EJITGPR methods. Once again, the simple averaging rule, PLS stacking, and GPR stacking does not function well in this study because they are nonadaptive. The above observations show that the proposed MP-EJITGPR (FMM) soft sensor method is the best among the compared methods. In addition, as illustrated in Figure 4, the superior prediction performance of MP-EJITGPR (FMM) is further verified by a good agreement between the predicted and actual trend plots of Mooney viscosity.

To further investigate the estimation performance of MP-EJITGPR (FMM) soft sensor, the prediction RMSE values of JITGPR, SP-EJITGPR, and MP-EJITGPR under different local modeling sizes are compared in Figure 5. It can be found that the increase of local modeling samples can lead to prediction accuracy reduction in most cases for all compared methods. In particular, for this case study, the prediction accuracy of JITGPR methods and SP-EJITGPR using small local modeling sizes is significantly better than that using large local modeling sizes. In comparison, the proposed MP-EJITGPR (FMM) is much less sensitive to local modeling size than other methods. Therefore, the proposed approach is more desirable than other traditional JIT methods in providing accurate and reliable predictions.

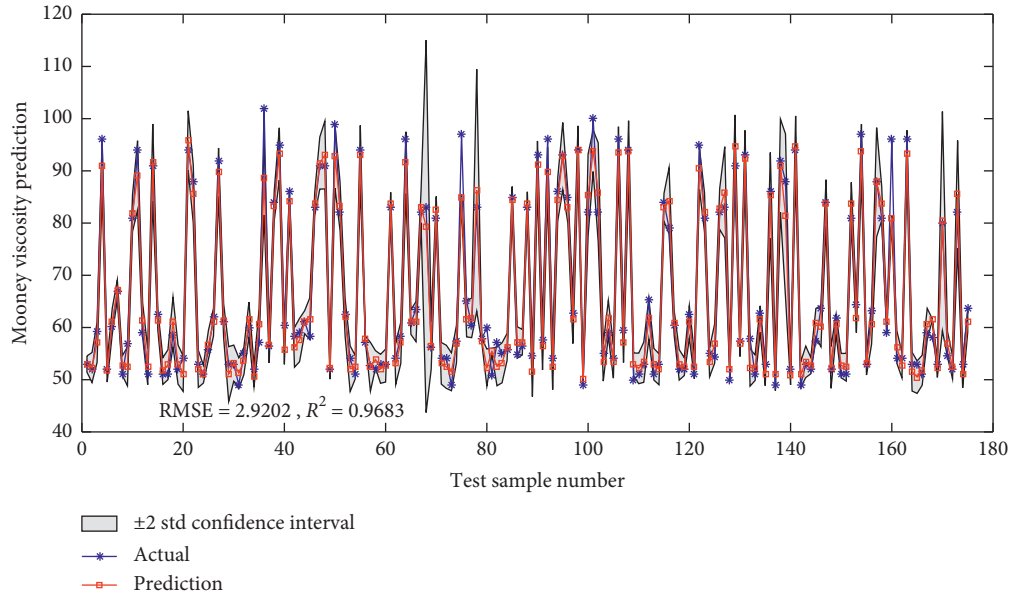Compared to the traditional JIT learning soft sensors, the outstanding prediction performance of MP-EJITGPR

FIGURE 4: Trend plots of Mooney viscosity predictions using the proposed MP-EJITGPR (FMM) approach (l=10).
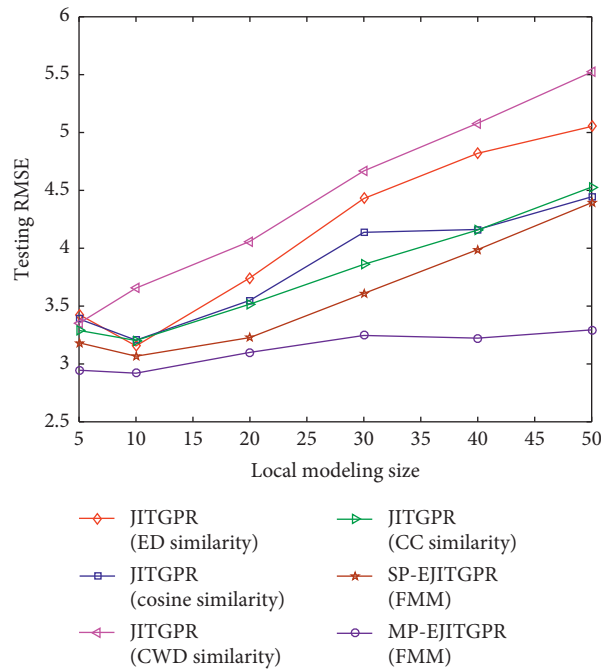

FIGURE 5: Comparison of prediction performance using single and ensemble JITGPR models.

(FMM) is mainly due to the effective cooperation of multimodal perturbation, EMO optimization, and adaptive combination for ensemble construction. On the one hand, the utilization of heterogeneous similarity measures, i.e., ED, cosine, CWD, and CC similarity metrics, and input variable selection for constructing subspaces shown in Figure 6, can be helpful to generate accurate and diverse base JITGPR models. On the other hand, the accuracy and diversity objectives of base JITGPR models can be well balanced by using an EMO approach. Additionally, the FMM-based

adaptive combination scheme allows the proposed MP-EJITGPR method to accommodate the query process state by dynamically assigned weights to base JITGPR models, as illustrated in Figure 7.

Moreover, the real-time performance of MP-EJITGPR (FMM) for online prediction is analyzed. The average CPU time for each run of prediction under different local modeling sizes is shown in Figure 8. Clearly, the online computational load becomes large with the increase of local modeling samples. However, only a small relevant subset is
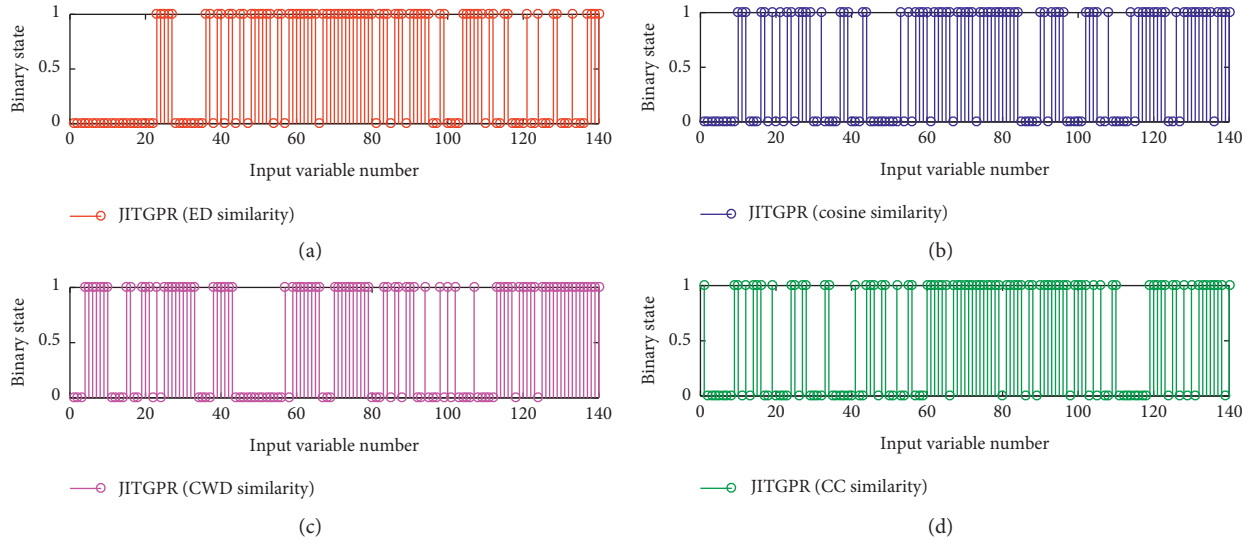
FIGURE 6: Input variable selection results for various similarity measures using EMO optimization. (a) JITGPR (ED similarity). (b) JITGPR (cosine similarity). (c) JITGPR (CWD similarity). (d) JITGPR (CC similarity).
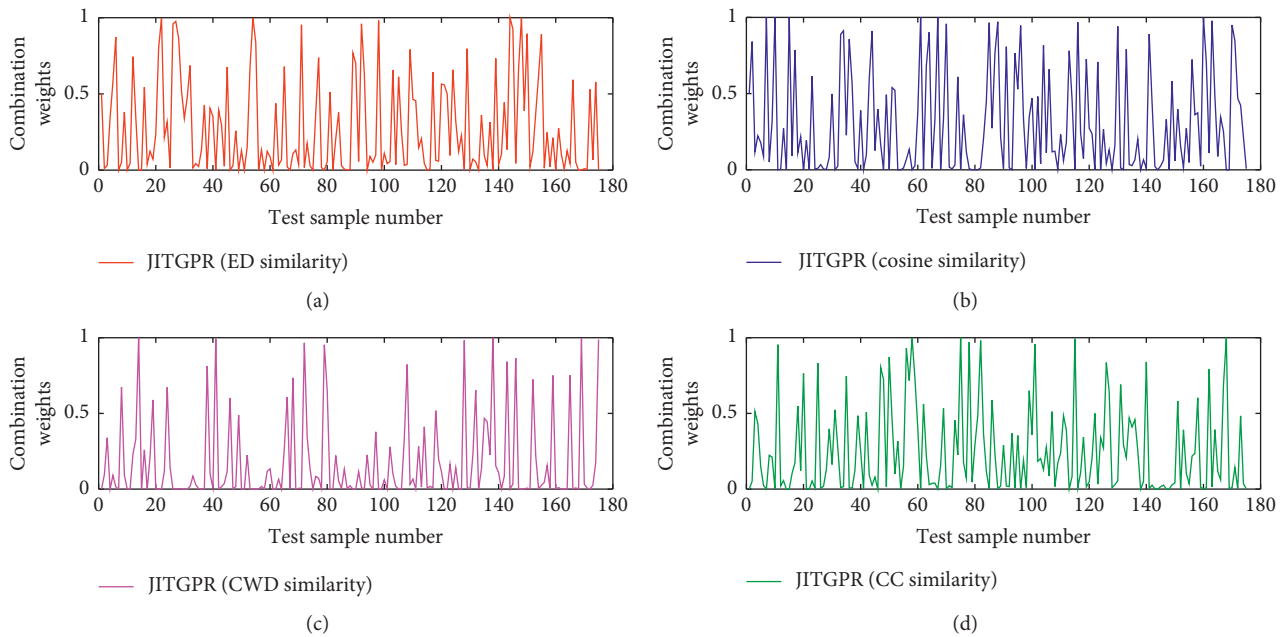

FIGURE 7: Combination weights assigned to different JITGPR models in MP-EJITGPR (FMM) method. (a) JITGPR (ED similarity). (b) JITGPR (cosine similarity). (c) JITGPR (CWD similarity). (d) JITGPR (CC similarity).
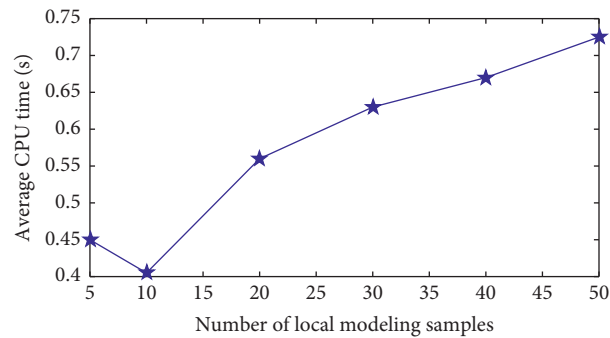

FIGURE 8: Average CPU time for online prediction using the proposed MP-EJITGPR (FMM) method.

required for local modeling, and the prediction taking less than 1 s is completely acceptable in practical application.

The obtained application results confirm that the proposed MP-EJITGPR (FMM) soft sensor method outperforms the other traditional JIT soft sensors, implying that it is more suitable for providing accurate predictions of Mooney viscosity in an industrial rubber mixing process.

5. Conclusions

In this paper, a new soft sensor method MP-EJITGPR is proposed for facilitating accurate estimations of Mooney viscosity in an industrial rubber mixing process. This method enables to enhance the diversity of base JIT learners through the multimodal perturbation mechanism, i.e., perturbing similarity measure and input variables. Moreover, a group of accurate and diverse base JIT learners is generated by employing an EMO approach to achieve a tradeoff between the accuracy and diversity objectives explicitly. In addition, a finite mixture mechanism is exploited to achieve an adaptive combination of base JIT learners. By integrating the multimodal perturbation-based diversity generation, the EMO optimization-based generation of base JIT learners, and the FMM-based adaptive combination of base learners for EJIT modeling, the proposed MP-EJITGPR method allows providing marked improvement of prediction performance over its conventional counterparts in nonlinear process modeling. The superiority and effectiveness of the proposed approach are demonstrated through the Mooney viscosity prediction of an industrial rubber mixing process.

Besides the presented case study, the proposed method has the potential of addressing other nonlinear modeling issues in process industry. In future research, more efforts are encouraged to extend the library of heterogeneous similarity measures and improve the diversity generation mechanism for building high-performance JIT soft sensors. Moreover, although this paper mainly focuses on manipulating input variables for building diverse input spaces based on evolutional multiobjective optimization approach, exploiting feature extraction by deep learning and making use of unlabeled data by semisupervised learning for improving the prediction performance of soft sensors are also interesting [47]. These will be investigated in the future.

Nomenclature

CC: Correlation coefficient
CWD: Covariance weighted distance
ED: Euclidean distance
EJIT: Ensemble just-in-time learning
EJITGPR: Ensemble just-in-time learning Gaussian process regression
EMO: Evolutionary multiobjective optimization
FMM: Finite mixture mechanism
GPR: Gaussian process regression
JIT: Just-in-time learning
JITGPR: Just-in-time learning Gaussian process regression

MOP: Multiobjective optimization problem
MP-   Multimodal perturbation-based ensemble just-
EJITGPR: in-time learning Gaussian process regression
NSGA-II: Nondominated sorting genetic algorithm II
SAR: Simple averaging rule
SP-   Similarity perturbation-based ensemble just-
EJITGPR: in-time learning Gaussian process regression.

Data Availability

The industrial data set of rubber mixing process involved in the present study is not allowed to be disclosed for reasons of commercial confidentiality.

Conflicts of Interest

The authors declare no conflicts of financial interest.

Acknowledgments

References

[1] E. Ehabé, F. Bonfils, C. Aymard, A. K. Akinlabi, and J. Sainte Beuve, "Modelling of Mooney viscosity relaxation in natural rubber," *Polymer Testing*, vol. 24, no. 5, pp. 620–627, 2005.

[2] J. E. Mark, B. Erman, and M. Roland, *The Science and Technology of Rubber. The Science and Technology of Rubber*, Academic Press, Roland, IA, USA, 2013.

[3] V. Vijayabaskar, R. Gupta, P. P. Chakrabarti, and A. K. Bhowmick, "Prediction of properties of rubber by using artificial neural networks," *Journal of Applied Polymer Science*, vol. 100, no. 3, pp. 2227–2237, 2006.

[4] Z. Zhang, K. Song, T.-P. Tong, and F. Wu, "A novel nonlinear adaptive Mooney-viscosity model based on DRPLS-GP algorithm for rubber mixing process," *Chemometrics and Intelligent Laboratory Systems*, vol. 112, pp. 17–23, 2012.

[5] K. Song, F. Wu, T.-P. Tong, and X.-J. Wang, "A real-time Mooney-viscosity prediction model of the mixed rubber based on the independent component regression-Gaussian process algorithm," *Journal of Chemometrics*, vol. 26, no. 11-12, pp. 557–564, 2012.

[6] Y. Liu and Z. Gao, "Real-time property prediction for an industrial rubber-mixing process with probabilistic ensemble Gaussian process regression models," *Journal of Applied Polymer Science*, vol. 132, 2015.

[7] Y. Liu, Y. Fan, L. Zhou, F. Jin, and Z. Gao, "Ensemble correntropy-based Mooney viscosity prediction model for an industrial rubber mixing process," *Chemical Engineering & Technology*, vol. 39, no. 10, pp. 1804–1812, 2016.

[8] K. Yang, H. Jin, X. Chen, J. Dai, L. Wang, and D. Zhang, "Soft sensor development for online quality prediction of industrial batch rubber mixing process using ensemble just-in-time Gaussian process regression models," *Chemometrics and Intelligent Laboratory Systems*, vol. 155, pp. 170–182, 2016.

[9] W. Jin, Y. Liu, and Z. Gao, "Fast property prediction in an industrial rubber mixing process with local ELM model," *Journal of Applied Polymer Science*, vol. 134, no. 41, p. 45391, 2017.

[10] W. Zheng, X. Gao, Y. Liu, L. Wang, J. Yang, and Z. Gao, "Industrial Mooney viscosity prediction using fast semi-supervised empirical model," *Chemometrics and Intelligent Laboratory Systems*, vol. 171, pp. 86–92, 2017.

[11] W. Zheng, Y. Liu, Z. Gao, and J. Yang, "Just-in-time semi-supervised soft sensor for quality prediction in industrial rubber mixers," *Chemometrics and Intelligent Laboratory Systems*, vol. 180, pp. 36–41, 2018.

[12] B. Pan, H. Jin, L. Wang et al., "Just-in-time learning based soft sensor with variable selection and weighting optimized by evolutionary optimization for quality prediction of nonlinear processes," *Chemical Engineering Research and Design*, vol. 144, pp. 285–299, 2019.

[13] L. Fortuna, S. Graziani, A. Rizzo, and M. G. Xibilia, *Soft Sensors for Monitoring and Control of Industrial Processes*, Springer Science & Business Media, Heidelberg, Germany, 2007.

[14] M. Kano and Y. Nakagawa, "Data-based process monitoring, process control, and quality improvement: recent developments and applications in steel industry," *Computers & Chemical Engineering*, vol. 32, no. 1-2, pp. 12–24, 2008.

[15] P. Kadlec, B. Gabrys, and S. Strandt, "Data-driven soft sensors in the process industry," *Computers & Chemical Engineering*, vol. 33, no. 4, pp. 795–814, 2009.

[16] Y. Shen, X. Li, H. Gao, and O. Kaynak, "Data-based techniques focused on modern industry: an overview," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 1, pp. 657–667, 2015.

[17] Z. Ge, Z. Song, S. X. Ding, and B. Huang, "Data mining and analytics in the process industry: the role of machine learning," *IEEE Access*, vol. 5, pp. 20590–20616, 2017.

[18] Y. Jiang and S. Yin, "Recent advances in key-performance-indicator oriented prognosis and diagnosis with a matlab toolbox: Db-kit," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 5, pp. 2849–2858, 2018.

[19] H. Kaneko, M. Arakawa, and K. Funatsu, "Development of a new soft sensor method using independent component analysis and partial least squares," *AIChE Journal*, vol. 55, no. 1, pp. 87–98, 2010.

[20] L. Zhou, Y. Wang, Z. Ge, and Z. Song, "Multirate factor analysis models for fault detection in multirate processes," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 7, pp. 4076–4085, 2019.

[21] G. Padmavathi, M. G. Mandan, S. P. Mitra, and K. K. Chaudhuri, "Neural modelling of Mooney viscosity of polybutadiene rubber," *Computers & Chemical Engineering*, vol. 29, no. 7, pp. 1677–1685, 2005.

[22] K. Desai, Y. Badhe, S. S. Tambe, and B. D. Kulkarni, "Soft-sensor development for fed-batch bioreactors using support vector regression," *Biochemical Engineering Journal*, vol. 27, no. 3, pp. 225–239, 2006.

[23] X. Yuan, L. Li, and Y. Wang, "Nonlinear dynamic soft sensor modeling with supervised long short-term memory network," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 5, pp. 3168–3176, 2020.

[24] P. Kadlec and B. Gabrys, "Local learning-based adaptive soft sensor for catalyst activation prediction," *AIChE Journal*, vol. 57, no. 5, pp. 1288–1301, 2011.

[25] H. Kaneko and K. Funatsu, "Ensemble locally weighted partial least squares as a just-in-time modeling method," *AIChE Journal*, vol. 62, no. 3, pp. 717–725, 2016.

[26] H. Jin, X. Chen, W. Li, Y. Kai, and W. Lei, "Adaptive soft sensor development based on online ensemble Gaussian process regression for nonlinear time-varying batch processes," *Industrial & Engineering Chemistry Research*, vol. 54, no. 30, pp. 7320–7345, 2015.

[27] H. Jin, X. Chen, J. Yang, H. Zhang, L. Wang, and L. Wu, "Multi-model adaptive soft sensor modeling method using local learning and online support vector regression for nonlinear time-variant batch processes," *Chemical Engineering Science*, vol. 131, pp. 282–303, 2015.

[28] Y. Liu, D. Huang, and Y. Li, "Development of interval soft sensors using enhanced just-in-time learning and inductive confidence predictor," *Industrial & Engineering Chemistry Research*, vol. 51, no. 8, pp. 3356–3367, 2012.

[29] Y. Liu, Z. Zhang, and J. Chen, "Ensemble local kernel learning for online prediction of distributed product outputs in chemical processes," *Chemical Engineering Science*, vol. 137, pp. 140–151, 2015.

[30] X. Yuan, J. Zhou, Y. Wang, and C. Yang, "Multi-similarity measurement driven ensemble just-in-time learning for soft sensing of industrial processes," *Journal of Chemometrics*, vol. 32, no. 9, p. e3040, 2018.

[31] H. Jin, B. Pan, X. Chen, and B. Qian, "Ensemble just-in-time learning framework through evolutionary multi-objective optimization for soft sensor development of nonlinear industrial processes," *Chemometrics and Intelligent Laboratory Systems*, vol. 184, pp. 153–166, 2019.

[32] G. Cybenko, "Just-in-time learning and estimation," in *Identification, Adaptation, Learning: The Science of Learning Models from Data*, S. Bittanti and G. Picci, Eds., Springer-Verlag, Berlin, Germany, vol. 153 of Nato ASI Series F Computer and Systems Sciences, pp. 423–434, 1996.

[33] D. W. Aha, *Lazy Learning*, Springer Science & Business Media, Berlin, Germany, 2013.

[34] C. G. Atkeson, A. W. Moore, and S. Schaal, "Locally weighted learning," in *Artificial Intelligence Review*, pp. 11–73, Springer, Berlin, Germany, 1997.

[35] C. K. Williams and C. E. Rasmussen, *Gaussian Processes for Machine Learning*, MIT Press Cambridge, Cambridge, MA, USA, 2006.

[36] J. A. Westerhuis, T. Kourti, and J. F. MacGregor, "Comparing alternative approaches for multivariate statistical analysis of batch process data," *Journal of Chemometrics*, vol. 13, no. 3-4, pp. 397–413, 1999.

[37] K. Hazama and M. Kano, "Covariance-based locally weighted partial least squares for high-performance adaptive modeling," *Chemometrics and Intelligent Laboratory Systems*, vol. 146, pp. 55–62, 2015.

[38] Z.-H. Zhou, *Ensemble Methods: Foundations and Algorithms*, Chapman and Hall/CRC, Boca Raton, FL, USA, 2012.

[39] S. Geman, E. Bienenstock, and R. Doursat, "Neural networks and the bias/variance dilemma," *Neural Computation*, vol. 4, no. 1, pp. 1–58, 1992.

[40] A. Krogh and J. Vedelsby, "Neural network ensembles, cross validation and active learning," *Advances in Neural Information Processing Systems*, vol. 5, pp. 231–238, 1995.

[41] J. Yu, "Online quality prediction of nonlinear and non-Gaussian chemical processes with shifting dynamics using finite mixture model based Gaussian process regression approach," *Chemical Engineering Science*, vol. 82, pp. 22–30, 2012.

[42] T. Chen and J. Ren, "Bagging for Gaussian process regression," *Neurocomputing*, vol. 72, no. 7–9, pp. 1605–1610, 2009.

[43] Z. Ge and Z. Song, "Ensemble independent component regression models and soft sensing application," *Chemometrics and Intelligent Laboratory Systems*, vol. 130, pp. 115–122, 2014.

[44] H.-X. Tian, Y.-D. Liu, K. Li, R.-R. Yang, and B. Meng, "A new AdaBoost.IR soft sensor method for robust operation optimization of ladle furnace refining," *ISIJ International*, vol. 57, no. 5, pp. 841–850, 2017.

[45] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: nsga-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.

[46] S. Frühwirth-Schnatter, *Finite Mixture and Markov Switching Models*, Springer Science & Business Media, Berlin, Germany, 2006.

[47] S. Zheng, K. Liu, Y. Xu, H. Chen, X. Zhang, and Y. Liu, "Robust soft sensor with deep kernel learning for quality prediction in rubber mixing processes," *Sensors*, vol. 20, no. 3, p. 695, 2020.