IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received June 1, 2020, accepted June 21, 2020, date of publication June 29, 2020, date of current version July 20, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3005716

Just-in-Time Learning-Based Soft Sensor for Mechanical Properties of Strip Steel via Multi-Block Weighted Semisupervised Models

JIE DONG[1], YINGZE TIAN[1], AND KAIXIANG PENG[1,2], (Member, IEEE)

[1]Key Laboratory of Knowledge Automation for Industrial Processes of Ministry of Education, School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China
[2]Institute of Artificial Intelligence, University of Science and Technology Beijing, Beijing 100083, China

Corresponding author: Kaixiang Peng (kaixiang@ustb.edu.cn)

This work was supported in part by the Natural Science Foundation of China (NSFC) under Grant 61773053 and Grant 61873024, in part by the Fundamental Research Funds for the China Central Universities of USTB, China, under Grant FRF-TP-19-049A1Z, and in part by the National Key Research and Development Program of China under Grant 2017YFB0306403.

ABSTRACT Mechanical properties are important indexes to evaluate the quality of hot rolling strips. It is a research hotspot in the field of hot rolling that realizing timely and accurate soft sensing of mechanical properties. Traditional soft sensing methods have poor performance in the application of strong nonlinearity and multiple working conditions. Moreover, the utilization rate of data is relatively low, which limit the improvement of prediction accuracy. To solve the problems above, a just-in-time learning (JITL) based multi-block weighted semisupervised Gaussian mixture regression (JMWSSGMR) soft sensor is proposed in the paper. There are two stages in the soft sensor: off-line variable blocking and on-line local modeling. In the off-line phase, process variables are divided into different sub-blocks by partial least square (PLS) according to distinct principal component directions. In each sub-block, original variables with high contribution rate are retained. In the on-line phase, optimized Mahalanobis distance is constructed to select the most similar historical samples to the query sample. Next, various real-time semisupervised sub-models are built to estimate the output of the query sample. Finally, predicted values of sub-models are fused and ultimate prediction of mechanical properties is obtained. Case studies are carried out on a numerical example and a hot rolling process. The feasibility and effectiveness of proposed soft sensor are verified by the predicted results.

INDEX TERMS Soft sensor, just-in-time learning (JITL), semi-supervised learning, hot-rolling process.

I. INTRODUCTION

Mechanical properties refer to the mechanical characteristics of materials under various external loads (tension, compression, bending etc.) in different environments, which are important indexes to evaluate the quality of materials. Mechanical properties of hot rolled strips mainly include tensile strength (TS), yield strength (YS) and elongation (EL). Accurate measuring and monitoring of mechanical properties is critical to realizing optimal control of hot rolling. It is also the key to ensuring safety of the process, assuring production efficiency and improving quality of products. However, due to the characteristics of high nonlinearity, complex interactions among variables and multiple working conditions in hot rolling process, the estimation of mechanical properties is difficult and unstable, which restricts the optimization and development of hot rolling equipment and controlling system. At present, there are three main ways to measure mechanical properties of strips: sampling analysis, online analysis instruments and soft sensors [1]. In the first method, pieces of samples need to be cut from strips and transferred to the laboratory for destructive testing. The results are meaningless because of its long time delay. For the second method, we obtain parameter information of strips through the devices installed on the production lines, which can detect mechanical properties of strips in time and greatly shortens measuring

The associate editor coordinating the review of this manuscript and approving it for publication was Ming Luo.

time. Nevertheless, on-line measuring devices are expensive and the maintenance of them is difficult, which increases running costs of steel plants. Different from the methods above, soft sensors do not rely on any professional testing equipment, but build virtual models to predict mechanical properties. There are two kinds of models in the soft sensor: metallurgical mechanism models and data-driven models [2]. The former establish theoretical formulas for the evolution of strip structure and process parameters, and predict mechanical properties by revealing microstructure evolution of steel. The structure of the mechanism model is generally complex including many critical dependencies. Moreover, we often need to update model parameters and structures based on the type of steel, resulting in limited application of metallurgical mechanism models. By contrast, data-driven soft sensors no longer focus on process mechanism, but attempt to extract useful information from process data to accomplish prediction tasks. Compared with complex mechanism models, data-driven ones have simpler structure, wider applicability and stronger learning ability. These models can extract effective parts from numerous process information and provide precise prediction of mechanical properties. In recent years, how to build responsive and accurate data-driven soft sensors is a research hotspot in the field of steel quality prediction and process control.

Due to the improvement of computing capability, some common data-driven models, such as artificial neural network (ANN) [3], [4], support vector machine (SVM) [5] and partial least square (PLS) [6], have been successfully applied in various industrial processes. Among them, ANN is the most widely used model for its strong nonlinear approximation and learning ability. As early as the 1990s, Liu *et al.* [7] have applied ANN to predict the mechanical properties of hot rolled C-Mn steel, showing great learning and generation ability compared with traditional regression models. In the early 2000s, there were more and more applications of ANN. Kim *et al.* [8] proposed a new integrated BP neural network, which screened variables by correlation analysis and integrated multiple neural networks to predict mechanical properties of hot rolled strips. Lalam *et al.* [9] combined principal component analysis (PCA) with BP neural network, which reduced the influence of redundant variables and variable collinearity. Wu *et al.* [10] constructed Bayesian neural network (BNN) to establish a reliable prediction model for mechanical properties of C-Mn steel. Combined with multi-objective optimization algorithm and expert knowledge, it was also successfully applied in the prediction of mechanical properties of Q235B steel. Except for ANN, other data-driven models were rarely used in the monitoring of mechanical properties. Due to the multiple working conditions in hot rolling process, Wang *et al.* [11] established a multi support vector regression model to distinguish data from different working conditions, and used grey correlation degree for weighted fusion. In recent years, because of its excellent feature extraction ability, deep learning is widely concerned in regression and classification problems. Yan *et al.* [12] combined stacked denoising autoencoders with neural networks. The proposed model is able to capture the essential information of input data through deep architecture. The model has better prediction performance than traditional soft sensors. Xu *et al.* [13] transformed process data into a two-dimensional data matrix, and used convolutional neural network (CNN) to predict mechanical properties.

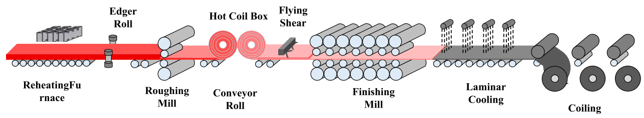

FIGURE 1. Process of hot strip rolling.

At present, most soft sensors applied in the prediction of mechanical properties of steel are single networks. Although ANN showed great performance in dealing with nonlinearity, its way of handling data is relatively simple. In addition, ANN models had poor interpretability and the results were unstable. In addition to inherent characteristics of neural networks, they also ignored the characteristics of industrial process in practical applications. Still, traditional ANN discarded a large number of unlabeled samples, which resulted in the lack of process information and limited the improvement of accuracy.

Aiming at the problems above, taking the mechanical properties of steel as targets, this paper presents a JITL-based multi-block weighted semisupervised Gaussian mixture regression soft sensing method. On the premise of analyzing various characteristics of hot rolling process, we achieve real-time and accurate soft sensing of mechanical properties of steel strips pertinently.

The rest of this paper is organized as follows. Section 2 introduces hot rolling process. Section 3 proposed the theory of variable blocking based on PLS, just-in-time learning, semisupervised GMR and basic modeling strategy. Next, each part of the soft sensor is proposed in detail in section 4. Then, case studies on a numerical example and an actual hot rolling process are given in section 5 to verify the effectiveness of soft sensor. Finally, section 6 gives some conclusions about this research.

II. DESCRIPTION OF HOT ROLLING PROCESS

Steel is widely used in many industrial fields, such as construction industry, bridges, mechanical manufacturing, aerospace, etc. The quality of steel is directly related to the hot rolling process. Hot rolling process is mainly divided into reheating, rough rolling, finish rolling, laminar cooling and coiling [13]. The layout of hot rolling process equipment is shown in Figure. 1.

First of all, cast slab is sent to the reheating furnace for heating to around 1200°C. Hot rolling process is a heat treatment process with two main purpose. First, at room temperature, the deformation resistance of steel is high. As temperature rises, steel gradually softens. Its deformation resistance also becomes lower, making it easy to be rolled. Second, morphology and distribution of internal microstructure of the steel changes gradually, and single-phase austenite with high plasticity and workability is obtained. Insufficient reheating will lead to insufficient dissolution of carbides and nitrates, resulting in series of cumulative effects in the subsequent rolling process, thus affecting mechanical properties of steel.

Before the strip is transferred to the roughing mill, some oxide impurities are generated from the reaction between strip steel surface and oxygen under high temperature. In order to avoid these impurities being rolled into steel and affecting surface quality of the steel, dephosphorization with high-pressure water is required before rolling. Rough rolling is the first stage of rolling, where the strip is rolled several times. Thickness is significantly reduced while the width is controlled by the vertical roll here. Austenite recrystalization mainly occurs in rough rolling. The size of austenite grain is greatly reduced. As a result, the mechanical properties of steel are significantly improved.

After rough rolling, the strip is sent to hot coil box through conveying roller for coiling temporarily. The main purpose of coiling here is to save space. In addition, when the strip is coiled together, temperature difference between the head and the tail of the strip is reduced, so as to achieve overall temperature balance and facilitate the control of the mechanical properties. Overall temperature of the strip is about 1050°C when uncoiling. Then, the head and the tail of the strip are cut by flying shear at the end of conveying roller to ensure flatness, and dephosphorization with high-pressure water is carried out again.

Finish rolling is a fine rolling stage for significantly im. Finishing mill usually consists of several rolling mills, which jointly control the tension, speed and other parameters of the strip to ensure that it meets specified requirements. At this stage, austenite is further refined to improve the strength and plasticity of the steel.

The stage after finish rolling is called laminar cooling. It is an on-line cooling process after rolling with laminar water flow. According to different steel grades, the strategy of cooling is also different, which is an important technological system to determine the performance of strip steel. In the laminar cooling process, when the temperature of the strip decreases to a certain extent, austenite transforms into other carbide structures with different mechanical properties, such as ferrite, pearlite, martensite, etc. Through the combination of various structures, mechanical properties of the strip meet final requirements.

Coiling is the last stage of hot rolling process, and there is still some phase transformation during coiling. After coiling and subsequent cooling, the qualified strip can be sold as mature product or further processed as raw material. Main works of each stage of hot rolling and the changes of strip structure are shown in Table 1.

TABLE 1. Main works and microstructure evolution in each stage of hot rolling.

Stage	Main works	Microstructure evolution
Reheating	Reduce deformation resistance & facilitate rolling	Austenite formation and growth
Rough rolling	Controll plate shape & provide advanced deformation for finishing rolling	Austenite recrystallization and grain refinement
Finish rolling	Make strip shape and flatness meet requirements	Austenite recrystallization and grain refinement
Laminar cooling	Reduce strip temperature	Austenite transformation
Coiling	Coil strips	Austenite transformation

III. PRELIMINARIES AND MODELING STRATEGY

A. PARTIAL LEAST SQUARE (PLS)

PLS is described as bilinear decomposition of both input space $X \in R^{N\times a}$ and output space $Y \in R^{N\times b}$, which are presented as follows [14]:

$$X = TP^T + E \tag{1}$$

$$Y = UQ^T + F \tag{2}$$

where N is the number of samples. $T = [t_1, t_2, \cdots, t_k] \in R^{N\times k}$ and $U = [u_1, u_2, \cdots, u_k] \in R^{N\times k}$ are matrices composed of score vectors of input space and output space, respectively. $P = [p_1, p_2, \cdots, p_k] \in R^{a\times k}$ and $Q = [q_1, q_2, \cdots, q_k] \in R^{b\times k}$ are loading matrices of input space and output space, respectively. $E(N \times a)$ and $F(N \times b)$ are residual matrices of input space and output space, respectively. Appropriate number of latent variables k is important for prediction accuracy. If there are too many latent variables, the model residual is small, but at the same time, the noise will be introduced, which leads to poor generalization performance. If too few, the relationship between X and Y cannot be exactly described, and the residual is large.

PLS algorithm is accomplished by nonlinear iteration. Its purpose is to find the main eigenvector of the following problem [15].

$$X^T yy^T Xw = \lambda w \tag{3}$$

The matrix formed by its eigenvector $W = [\omega_1, w_2, \cdots, w_k] \in R^{a\times k}$ is weighted matrix of PLS. In order to get relationship between latent variables and original variables, it is derived as:

$$XW = TP^T W \Rightarrow T = XW^* \tag{4}$$

where $W^* = W(P^T W)^{-1} \in R^{a\times k}$ is the mapping matrix from input space to potential space, in which each column is a mapping vector from original space to potential space. We present U by T, which is $U = TV + H$, where V is a diagonal matrix, and H is residual matrix. Eq.(2) can be expressed as:

$$Y = UQ^T + F = TVQ^T + HQ^T + F = TQ^{*T} + F^* \tag{5}$$

Therefore, the score matrix of X is also the predictor of Y. Eq.(5) can be further expressed as:

$$Y = XW^*Q^{*T} + F^* = XB + F^* \tag{6}$$

where B is the regression coefficient of PLS algorithm.

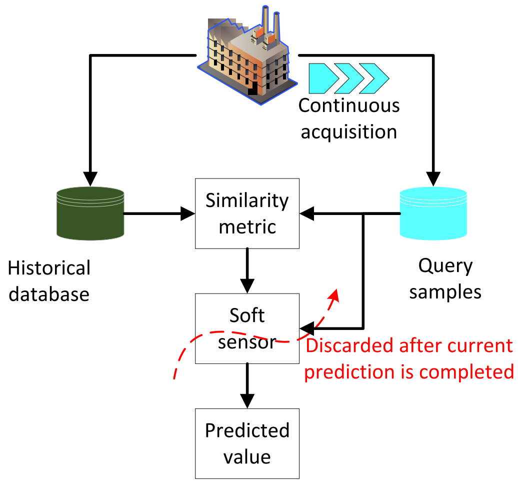

FIGURE 2. Just-in-time learning framework.

B. JUST-IN-TIME LEARNING (JITL)

When monitoring the industrial process with obvious batch properties such as hot rolling one, it is necessary for the soft sensor to adjust model structure and parameters adaptively due to the type of products. As a result, it is challenging to build an appropriate global model suitable for products under diverse working conditions. As an alternative, an adaptive on-line local modeling strategy, just-in-time learning, is proposed [16]. JITL based models are built on-line. According to different characteristics of query samples, local models are built by searching samples with similar characteristics in the database, which reduces the interference of redundant samples. Such local models can well adapt to the nonlinearity and multiple working conditions in the industrial process. JITL is generally divided into four steps [17], [18], and the overall framework is shown in Figure.2.

1) Establish the historical dataset by collecting data from an industrial process.
2) Calculate the distance or similarity between historical samples and the query sample by some means, and the historical samples with the closest distance or the highest similarity are selected as modeling samples.
3) A local model is trained with modeling samples and used to give an estimation of the query sample.
4) The local model is discarded immediately after the predicted value is obtained.

There are two concerns when implementing JITL, which are "relevance" and "accuracy". The first means that the selected samples should be relevant to the query sample. The second refers that the local models should be accurate enough [19]. In the industrial process, different batches of products are usually produced under different working conditions, which results in the difference of data characteristics. If the distance or similarity metrics are not appropriate, it will be likely to select unsuitable samples as modeling samples and reduce prediction accuracy of the model. As a result, the core issue of JITL is how to build an appropriate distance or similarity metric to select the most relevant samples.

C. SEMISUPERVISED GAUSSIAN MIXTURE REGRESSION (SSGMR)

It is generally difficult for single global model to estimate the output of samples under multiple working conditions. Moreover, because of the inevitable noises and measuring errors in industrial applications, there are inherent randomness and uncertainty in samples. In this case, probabilistic models are more suitable than non-probabilistic ones [20]. Gaussian mixture regression is an extended form of Gaussian mixture model, which can be used in regression problems. The regression is based on the Gaussian conditioning and linear combination properties of Gaussian distributions. By dividing samples into input and output part, the joint probability distribution of input and output is obtained in GMM. Then, the conditional probability distribution of output to input is estimated by parameters in GMM. After training, GMR model can offer estimations of the query samples [21]. In addition, in order to avoid the waste of unlabeled samples, this paper adopts SSGMR as the prediction model.

Let $X = [x_1, x_2, \cdots, x_N]^T \in R^{N\times d}$ and $Y = [y_1, y_2, \cdots, y_N]^T \in R^{N\times 1}$ be $d - dimension$ input matrix and $1 - dimension$ output matrix of N samples, respectively. Labeled samples are presented as $\{X_l, Y_l\} = \{x_i, y_i\}_{i=1}^{n_l}$ and unlabeled ones are denoted as $\{X_u\} = \{x_j\}_{j=1}^{n_u}$. Assuming the marginal distribution of x follows Gaussian distribution, and the relationship between x and y of the $k - th$ Gaussian component is linear.

$$p_k(x) = \mathcal{N}(x|\mu_k^x, \Sigma_k^x) \tag{7}$$
$$p_k(y|x) = \mathcal{N}(y|x^T\omega_k + \psi_k, \sigma_k^2) \tag{8}$$

where ω_k and ψ_k are regression coefficient, σ_k is the variance of predictions.

In order to get the parameters of SSGMR, which are denoted as $\Theta = \{\alpha_k, \mu_k^x, \Sigma_k^x, \omega_k, \psi_k, \sigma_k^2\}$, we use expectation maximization algorithm (EM) to solve the problem iteratively. EM algorithm is divided into expectation (E-step) and maximization (M-step). In E-step, the posteriori probability density functions of labeled and unlabeled samples belonging to the $k - th$ component are:

$$p(z_i = k|x_i, y_i) = \frac{\alpha_k \mathcal{N}(x_i, y_i|\mu_k^{xy}, \Sigma_k^{xy})}{\sum_{k=1}^K \alpha_k \mathcal{N}(x_i, y_i|\mu_k^{xy}, \Sigma_k^{xy})} \tag{9}$$
$$p(z_j = k|x_j) = \frac{\alpha_k \mathcal{N}(x_j|\mu_k^x, \Sigma_k^x)}{\sum_{k=1}^K \alpha_k \mathcal{N}(x_j|\mu_k^x, \Sigma_k^x)} \tag{10}$$

where the probability that the sample belongs to the $k - th$ component is expressed as $p(z_i = k) = p(z_j = k) = \alpha_k$. To reduce the complexity of subsequent derivation, we denote $p(z_i = k|x_i, y_i)$ and $p(z_j = k|x_j)$ as γ_k^i and γ_k^j, respectively.

TABLE 2. **Flow diagrams of the algorithm.**

Algorithm 1: Learning procedure of parameters
Set hyper-parameters of SSGMR
While 1
E-step:
For i=1:ncomp(number of components)
Calculate the probability density function of the $i-th$ Gaussian component $p(z=i\|x), p(z=i\|x,y)$
End for
Using (9) and (10) to calculate the posterior probabilities of the $i-th$ Gaussian component for labeled and unlabeled samples, respectively.
M-step:
For i=1:ncomp
Update parameter set.
End for
Stop criterion:
Calculate the log-likelihood function using Eq.(12)
If the criterion in Eq.(13) is satisfied
End algorithm
End if
End while

In M-step, it is assumed that samples are independent from each other. The log-likelihood function is defined as [22]:

$$\begin{aligned}\mathcal{L}(\Theta) = &\sum_{i=1}^{n_l}\sum_{k=1}^{K}\gamma_k^i \ln p_k(y_i|x_i) + \sum_{i=1}^{n_l}\sum_{k=1}^{K}\gamma_k^i \ln p_k(x_i) \\ &+ \sum_{i=1}^{n_l}\sum_{k=1}^{K}\gamma_k^i \ln \alpha_k + \sum_{j=n_l+1}^{n_l+n_u}\sum_{k=1}^{K}\gamma_k^j \ln p_k(x_j) \\ &+ \sum_{j=n_l+1}^{n_l+n_u}\sum_{k=1}^{K}\gamma_k^j \ln \alpha_k) + \beta(\sum_{k=1}^{K}\alpha_k - 1)\end{aligned} \tag{11}$$

Θ can be obtained by setting derivatives of $\mathcal{L}(\Theta)$ with respect to each parameter to 0, which can be found in [22].

After several iterations, the convergence can be diagnosed by comparing log-likelihood function defined as:

$$\ln p(D|\Theta) = \sum_{i=1}^{n_l}\ln(p(x_i, y_i)) + \sum_{j=n_l+1}^{n_l+n_u}\ln(p(x_j)) \tag{12}$$

where D is the dataset, and the convergence criterion can be:

$$\left|\frac{\ln p(D|\Theta^{(t+1)}) - \ln p(D|\Theta^{(t)})}{\ln p(D|\Theta^{(t)})}\right| < \varepsilon \tag{13}$$

where $\Theta^{(t)}$ denote the parameters set obtained in the $t-th$ iteration, and ε represents the pre-defined threshold. The learning algorithm above are summarized in Table 2.

Then the probability density function of y_q conditioned on x_q is presented as:

$$\begin{aligned}p(y_q|x_q) &= \sum_{k=1}^{K} p(z_q = k|x_q)p(y_q|x_q, z_q = k) \\ &= \sum_{k=1}^{K}\gamma_k^q \mathcal{N}(y_q|x_q^T\omega_k + \psi_k, \sigma_k^2)\end{aligned} \tag{14}$$

Finally, the predicted value of query sample is computed as:

$$\hat{y}_q = E\left[y_q|x_q\right] = \sum_{k=1}^{K}\gamma_k^q(x_q^T\omega_k + \psi_k) \tag{15}$$

There is an advantage of SSGMR that it can provide the uncertainty of prediction, which is in the form of variance of predicted values [22].

$$\begin{aligned}\sigma_q^2 &= \int p(y_q|x_q)y_q^2 dy_q - (E\left[y_q|x_q\right])^2 \\ &= \sum_{k=1}^{K}\gamma_k^q(\sigma_k^2 + (x_q^T\omega_k + \psi_k)^2) - \hat{y}_q^2\end{aligned} \tag{16}$$

The quality of each estimation can be evaluated by the uncertainty, which offers the reference for model fusion.

D. MODELING STRATEGY

In order to adapt to strong nonlinearity, multiple working conditions and low utilization of samples, this paper proposed a JITL based multi-block weighted SSGMR soft sensor, which is briefly shown in Figure.3. First, the available dataset is obtained by preprocessing original data from the hot rolling process. Then, PLS algorithm is used to divide the original variable set into sub-blocks, which contain different process variables, and variables with high contribution rate in each sub-block are selected to form the auxiliary dataset. Next, under JITL framework, several local SSGMR models are built in each sub-block. Finally, predicted values of sub-blocks are fused according to their uncertainty and the accurate predictions of mechanical properties are obtained.

IV. JITL BASED REAL TIME SEMISUPERVISED SOFT SENSOR

A. VARIABLE BLOCKING AND ALLOCATION BASED ON PLS

Hot rolling process is a complex industrial process with various variables, which include temperature, specification and element composition etc. Their effects on mechanical properties of hot strip steel are generally different, and there are often complicated correlations among them. Ordinarily, two principles should be satisfied for a blocking method [23]: diversity and accuracy. First of all, it is necessary to distinguish variables with different influence on mechanical properties. These variables are divided into different sub-blocks, reducing correlations among sub-blocks. Meanwhile, we also make full use of process variables, which reflects the principle of diversity. For the principle of accuracy, we establish auxiliary variable set in each sub-block by selecting important variables. By removing trivial variables according to specific rules, we not only reduce the overlapping parts of sub-blocks and the interference of redundant variables, but also reduce the dimension of variables and improve training speed and model performance.

PLS is able to decompose original variables into low dimensional variables with different principal component directions, and ensure a maximum correlation between them and output. Therefore, it is a proper algorithm for variables blocking in our task. The implementation of PLS is described as follows.

Input space of samples is presented as $X \in R^{N\times a}$, output space of samples is $Y \in R^{N\times b}$. With Eq.(1), (4) and (5),

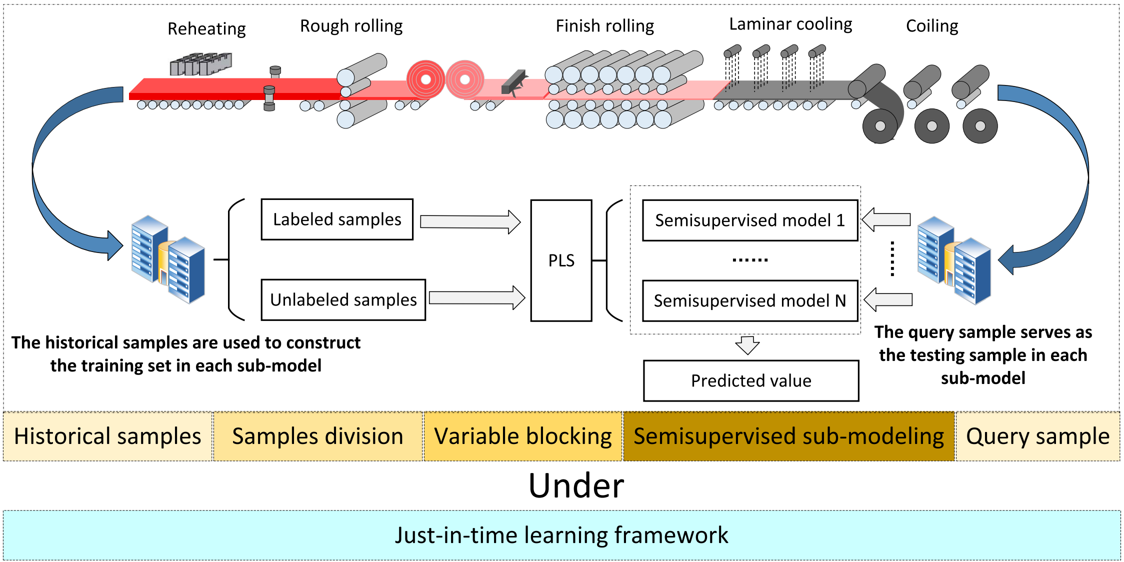

FIGURE 3. Semisupervised soft sensing strategy based on JITL.

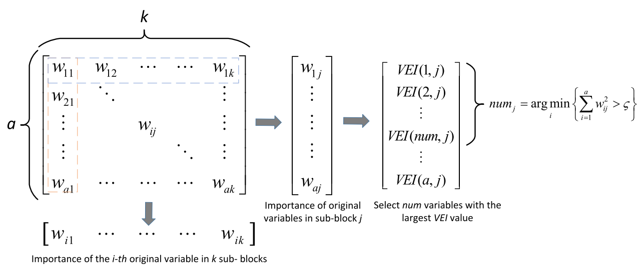

FIGURE 4. Diagram of blocking based on PLS.

score matrix T is a latent matrix for mapping, which is also the predictor of both input and output. Before calculating, each column of X is standardized to reflect the influence of original variables on respective sub-block. For the purpose of removing redundant variables, variable evaluation index (VEI) is constructed based on W^*:

$$VEI(i,j) = \frac{w_{ij}^2}{w_{1j}^2 + w_{2j}^2 + \cdots + w_{ij}^2 + \cdots + w_{aj}^2} \quad (17)$$

where $i = 1, 2, \cdots, a$, $j = 1, 2, \cdots, k$. w_{ij}^2 indicates the importance of the $i - th$ original variables to the $j - th$ sub-block. The variables with higher VEI are more important in sub-blocks. We predefine a threshold value $\varsigma \in (0, 1)$. Generally, it is bigger than 0.85 for sufficient process information. The number of variables retained in each sub-block is:

$$num_j = \arg\min_i \{\sum_{i=1}^{a} w_{ij}^2 > \varsigma\} \quad (18)$$

After blocking through PLS, several variable sets with low correlations construct the auxiliary variable set. The overall structure of PLS blocking is shown in Figure.4.

B. JITL BASED REAL TIME SEMISUPERVISED MODELING METHOD

When the query sample is available, N_1 labeled samples and N_2 unlabeled samples with the most similar characteristics to the query sample are selected to construct dataset $L_q = \{x_l, y_l\}_{l=1,2,\cdots,N_1}$ and $U_q = \{x_u\}_{u=1,2,\cdots,N_2}$, respectively. The core issue of JITL is how to construct an appropriate distance or similarity metric. Generally, Euclidean distance (ED) is the most widely used distance metric, which regards the importance of all variables as the same, and do not consider the scale of variables. When using ED as the distance metric, some variables with large value such as width of slabs and temperatures play a significant role in modeling. Meanwhile, variables with small value such as elements content are neglected. It obviously disagrees with the facts that the content of elements in steel have great influence on the mechanical properties. As a result, ED is not acceptable in our task. In contrast, Mahalanobis distance (MD) is not affected by the dimension of variables, which is better than ED. Based on MD, we construct variable-related Mahalanobis distance (VRMD) between the query sample and labeled ones, unlabeled ones, respectively, which is defined as:

$$d_{q,l,j} = \sqrt{(x_{q,j} - x_{l,j})^T \Sigma_j (x_{q,j} - x_{l,j})}, \quad l = 1, 2, \cdots, N_1 \quad (19)$$

$$d_{q,u,j} = \sqrt{(x_{q,j} - x_{u,j})^T \Sigma_j (x_{q,j} - x_{u,j})}, \quad u = 1, 2, \cdots, N_2 \quad (20)$$

where j represents the $j - th$ sub-block, Σ_j is the weighting matrix for variables, which is:

$$\Sigma_j = diag(\frac{VEI(1,j)}{\sigma_1^2}, \frac{VEI(2,j)}{\sigma_2^2}, \cdots, \frac{VEI(n,j)}{\sigma_n^2}) \quad (21)$$

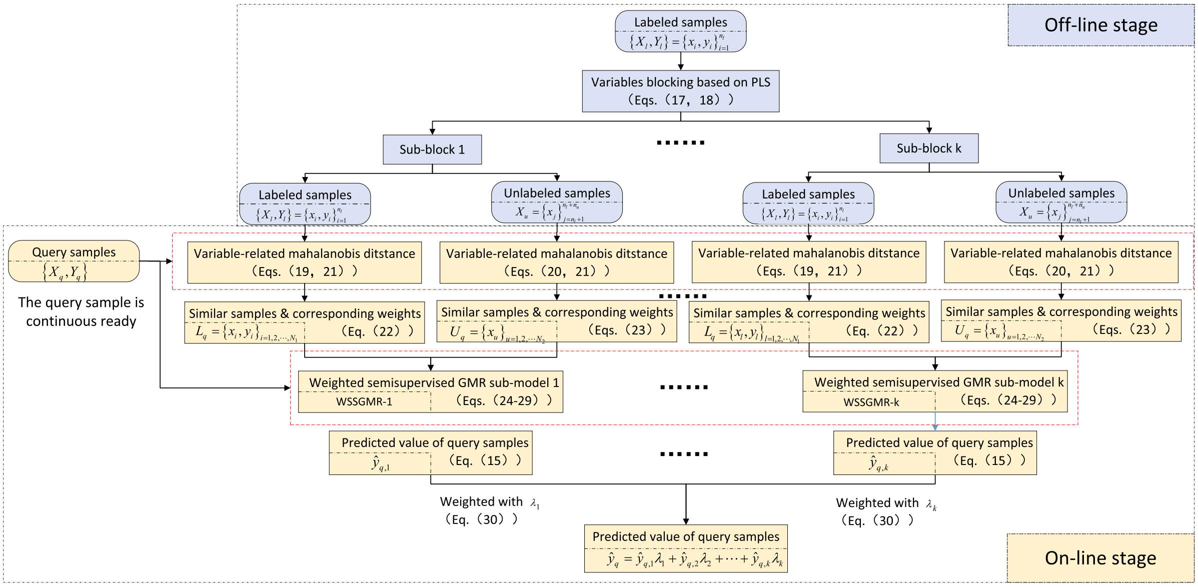

FIGURE 5. Overall procedure of proposed soft sensor.

where σ_n^2 is the variance of the $n-th$ variable in the $j-th$ sub-block. Different from traditional MD, newly constructed VRMD is not only independent with the scale of variables, but also more suitable for the data from industrial process by introducing variable evaluation index. In addition, because the historical samples closer to the query sample are more important, more attention should be paid to these samples during training. Thus, these samples should be given corresponding weights, which are defined as follows:

$$\omega_l^j = \exp(-d_{q,l,j}^2/\tau^2), \quad l = 1, 2, \cdots, N_1 \tag{22}$$

$$\omega_u^j = \exp(-d_{q,u,j}^2/\tau^2), \quad u = 1, 2, \cdots, N_2 \tag{23}$$

where τ represents weight attenuation coefficient. The samples with smaller VRMD are more similar to the query sample. In practice, input of samples are weighted when updating parameters. After locally weighted, a new log-likelihood function is defined by adding weights of samples to SSGMR:

$$\begin{aligned}\mathcal{L}(\Theta) = &\sum_{i=1}^{n_l}\sum_{k=1}^{K}\gamma_k^i\omega_i(\ln p_k(y_i|x_i) + \ln p_k(x_i) + \ln\alpha_k)\\ &+\sum_{j=n_l+1}^{n_l+n_u}\sum_{k=1}^{K}\gamma_k^j\omega_j(\ln p_k(x_j) + \ln\alpha_k)\\ &+\beta(\sum_{k=1}^{K}\alpha_k - 1)\end{aligned} \tag{24}$$

Correspondingly, parameters of the model are calculated as:

$$\alpha_k = (\eta_k^l + \eta_k^u)/(\sum\nolimits_{i=1}^{n_l}\omega_i + \sum\nolimits_{j=n_l+1}^{n_l+n_u}\omega_j) \tag{25}$$

$$\mu_k^x = (\sum_{i=1}^{n_l}\gamma_k^i\omega_i x_i + \sum_{j=n_l+1}^{n_l+n_u}\gamma_k^j\omega_j x_j)/(\eta_k^l + \eta_k^u) \tag{26}$$

$$\Sigma_k^x = (\sum_{i=1}^{n_l}\gamma_k^i\omega_i\bar{x}_k^i(\bar{x}_k^i)^T + \sum_{j=n_l+1}^{n_l+n_u}\gamma_k^j\omega_j\bar{x}_k^j(\bar{x}_k^j)^T)/(\eta_k^l+\eta_k^u) \tag{27}$$

$$\tilde{\omega}_k = (\tilde{X}_l^T\Gamma_k\tilde{X}_l)^T\tilde{X}_l^T\Gamma_k Y_l \tag{28}$$

$$\sigma_k^2 = \sum_{i=1}^{n_l}\gamma_k^i\omega_i(y_i - \tilde{x}_i^T\tilde{\omega}_k)^2/\eta_k^l \tag{29}$$

where $\Gamma_k = diag[\gamma_k^1\omega_1, \cdots, \gamma_k^{n_l}\omega_{n_l}]$, $\eta_k^l = \sum_{i=1}^{n_l}\gamma_k^i\omega_i$, $\eta_k^u = \sum_{j=n_l+1}^{n_l+n_u}\gamma_k^j\omega_j$, while others are the same as those in SSGMR.

By analyzing the equations above, input-related parameters such as α_k, μ_k^x and Σ_k^x are updated by all samples. Output-related ones such as $\hat{\omega}_k$ and $\hat{\sigma}_k^2$ are updated by labeled samples only. Thus, in semisupervised models, unlabeled samples affect the model by providing prior information of data distribution. Predicted values are available depending on labeled samples.

Finally, the results of sub-blocks are fused by the uncertainty $\{\sigma_{q,n}^2|n = 1, 2, \cdots, j\}$, the weight λ_n of each part is calculated as:

$$\lambda_n = \frac{(\sigma_{q,n}^2)^{-1}}{\sum_{n=1}^{j}\sigma_{q,n}^2)^{-1}} \tag{30}$$

The final predicted value of the query sample is $\hat{y}_q = \hat{y}_{q,1}\lambda_1 + \hat{y}_{q,2}\lambda_2 + \cdots + \hat{y}_{q,j}\lambda_j$. The overall procedure of the soft sensor is shown in Figure.5.

V. CASE STUDY

In order to verify the feasibility and effectiveness of proposed soft sensor, a numerical example is carried out for comparison with other algorithms. Then we apply the soft sensor in the real hot rolling process. The configuration of the machine

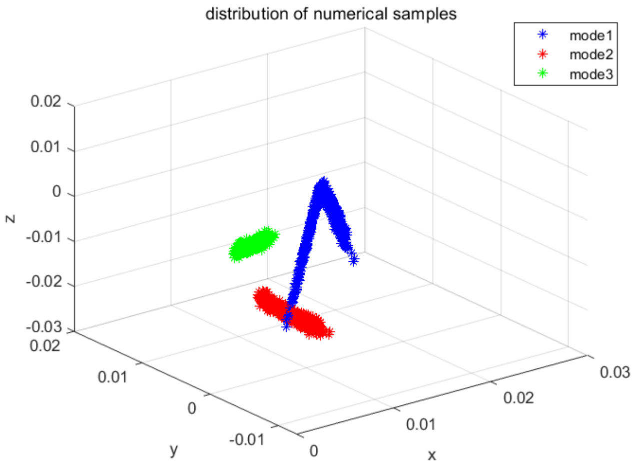

FIGURE 6. Distribution of virtual dataset.

used in this paper is: CPU: core i7-6567u (3.3GHz, 3.2GHz); RAM: 8.00GB; MATLAB (r2018a).

A. NUMERICAL EXAMPLE

The numerical example follows the idea in reference [24]. It is assumed that there are five virtual signal sources. Through the linear combination of these five signal sources, 10 virtual input variables and 1 output variable are constructed. The function expression of the signal source is as follows:

$$\begin{cases} s_1(k) = \sin((k+15)/10) - \cos((k+15)/10) \\ s_2(k) = \sin(k) + \cos(0.5k) \\ s_3(k) = \ln(k+1) \\ s_4(k) = \textit{numbers of Gaussian distribution} \\ \qquad\quad \textit{in } (0,1) \\ s_5(k) = (\exp(k) - 1)/e \end{cases} \tag{31}$$

For the simulation of the characteristics of multiple working conditions, 3 different kinds of samples are obtained by linear combination of different signal sources.

1000 samples of mode 1 are:

$$\begin{cases} x_1 = As + e_1 \\ y_1 = 0.55s_1 + 1.6s_2 + 2.5s_3 + e_2 \end{cases} \tag{32}$$

1000 samples of mode 2 are:

$$\begin{cases} x_2 = ABs + e_1 \\ y_2 = 2.4s_1 + 0.84s_2 + 1.3s_3 + e_2 \end{cases} \tag{33}$$

1000 samples of mode 3 are:

$$\begin{cases} x_3 = AB^2s + e_1 \\ y_3 = 1.8s_1 + 2.1s_2 + 0.45s_3 + e_2 \end{cases} \tag{34}$$

where A and B are the mapping matrices. $e_1 \sim \mathcal{N}(0, 0.01)$ and $e_2 \sim \mathcal{N}(0, 0.1)$ are Gaussian noises. After dimensionality reduction, the dataset constructed is visualized as 3-dimensional graphical representation in Figure 6.

TABLE 3. Description of variables in hot rolling process.

Variable(1-14)	Range	Variable(15-27)	Range
Furnace temp(°C)	1202-1315	ALS(Wt%)	0.004-0.05
Rough rolling temp(°C)	1037-1200	Cu(Wt%)	0.011-0.35
Finish rolling temp(°C)	791-931	Ni(Wt%)	0.008-0.127
Coiling temperature(°C)	555-696	Cr(Wt%)	0.022-0.43
Thickness(mm)	1.6-8	Mo(Wt%)	0-0.01
Width(mm)	928-1500	Nb(Wt%)	0-0.002
Actual weight(kg)	12470-28100	V(Wt%)	0-0.0109
Slab weight(kg)	16133-28166	Ti(Wt%)	0-0.0108
Rolled weight(kg)	15000-28100	B(Wt%)	0-0.00112
C(Wt%)	0.027-0.2	Ca(Wt%)	0.0004-0.0031
Si(Wt%)	0.014-0.49	As(Wt%)	0.00188-0.006
Mn(Wt%)	0.12-0.48	Sn(Wt%)	0-0.00462
P(Wt%)	0.005-0.106	Reheating time(s)	66-233
S(Wt%)	0.001-0.021	-	-

950 samples of each mode are selected to join the training set, and the remaining samples are added to the testing set. Superiority of the proposed soft sensor will be proved from two aspects: the effective use of unlabeled samples and the applicability under multiple working conditions.

It is first compared with PLS, GPR, GMR to verify the effectiveness of unlabeled samples. The number of components in PLS is 10. Matern kernel is used in GPR. In JMWSSGMR, because the dataset is not complicated, the number of sub-block is not as important as one in industrial process. In order to reduce prediction time, the number of sub-blocks is set to 3. The detailed influence of the number of sub-blocks is shown in next part. We select 70 neighbors of the query sample as modeling samples. The number of samples in labeled and unlabeled set is the same. The results of prediction are shown in Figure.7.

In addition, the comparison between proposed soft sensor and SSGMR is shown in Figure.8. The number of Gaussian components in SSGMR is 3. After JITL framework is introduced, performance of the model is higher than that of SSGMR. At the same time, we can see that SSGMR performs better than GMR.

The numerical example verified the effectiveness of proposed soft sensor. It also proved that JMWSSGMR is applicability to the data with varieties of characteristics.

B. SIMULATION IN HOT ROLLING PROCESS

1) DESCRIPTION OF HOT ROLLING VARIABLES

Hot rolling is an important heat treatment in the process of steel rolling. Mechanical properties are critical quality indicators. During hot rolling process, they are affected by many factors, which lead to the difficulty of monitoring. Therefore, it is important to realize the real time soft sensing of mechanical properties. In this paper, from 17 February 2019 to 7 March 2019, a total of 1095 samples were collected from the hot rolling process in Anling iron and steel Ltd, Liaoning, China. Each sample contains 27 input variables and 3 mechanical properties which are TS, YS and EL as output. Input variables are mainly divided into process temperature, product specification and element mass fraction. The kind and range of process variables are shown in Table 3.

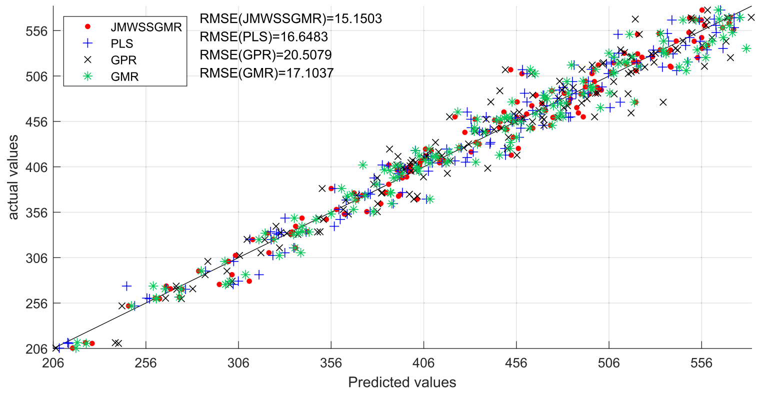

FIGURE 7. Comparison with supervised models.

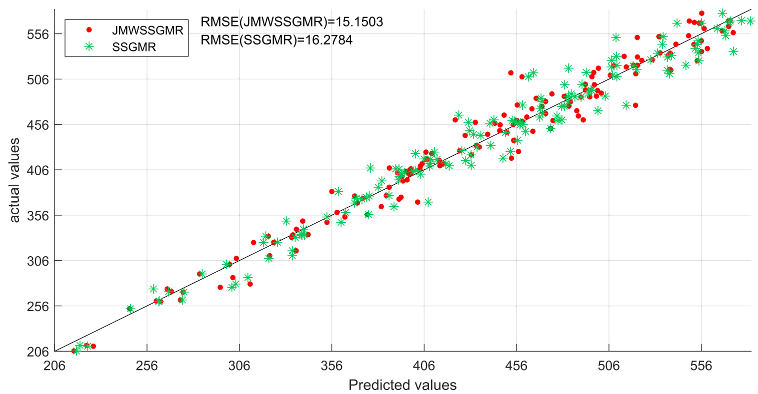

FIGURE 8. Comparison between proposed soft sensor and SSGMR.

TABLE 4. Distributions of mechanical properties of the steel.

Type of steel	TS(MPa)	(lower)YS(MPa	EL(%)
SPHC	336-393	237-310(lower)	42.2-54.3
Q235B	364-515	267-409	26.6-42.6
SPA-H	517-594	404-486	27.8-40.1

There are three kinds of steel in our dataset, namely, SPHC (quality carbon structural steel), Q235B (plain carbon constructional steel) and SPA-H (Climate resistance low-alloy structural steel). The distributions of mechanical properties of the steel are listed in Table 4.

Due to the high content of Si, Mn, Cu, Ni, Cr and other alloy elements, the strength of SPA-H is higher than that of carbon steel, but it has worse plasticity. Among carbon steel, SPHC has lower carbon content, which brings it lower strength and higher plasticity.

After the dataset is available, outliers in dataset are removed first according to prior knowledge. Then, normal samples are divided into training set and testing one, including 1006 and 85 samples, respectively. The training set is divided into labeled set and unlabeled set with equal number of samples.

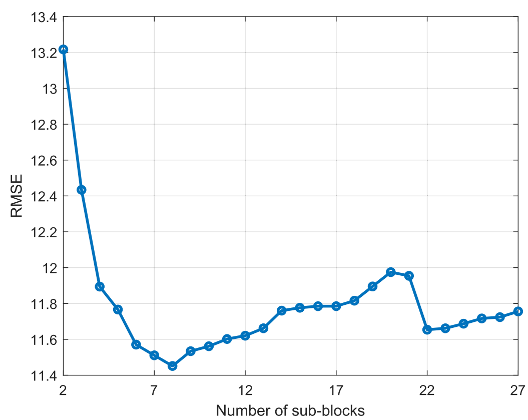

FIGURE 9. RMSE of predictions under diverse numbers of sub-blocks.

2) OFF-LINE VARIABLE BLOCKING

There are two stages in the soft sensor: off-line variable blocking and on-line modeling. In the off-line phase, the relationship between the number of sub-blocks and prediction accuracy is explored first, as shown in Figure.9. For the balance of diversity and prediction time, we build 8 sub-blocks in practical operation.

TABLE 5. **Auxiliary variables in sub-blocks.**

No.	Auxiliary variables							
	sub-block1	sub-block2	sub-block3	sub-block4	sub-block5	sub-block6	sub-block7	sub-block8
1	**11**	**10**	**10**	**10**	1	**10**	**10**	23
2	**10**	23	21	25	15	12	12	**11**
3	23	6	7	23	**10**	27	21	**10**
4	12	8	23	14	12	15	1	18
5	18	9	9	12	27	2	23	12
6	6	7	8	20	**11**	5	**11**	13
7	13	25	**11**	21	14	**11**	4	5
8	17	19	22	**11**	25	3	14	24
9	16	5	6	19	-	1	13	21
10	8	26	19	5	-	4	5	19
11	21	4	18	27	-	20	18	-
12	9	22	25	8	-	-	27	-
13	7	21	-	4	-	-	-	-
14	26	20	-	-	-	-	-	-
15	22	15	-	-	-	-	-	-
16	-	13	-	-	-	-	-	-

Taking TS as example, after determining the number of sub-blocks, we retain variables whose cumulative contribution rate is more than 90% in each sub-block. The variables kept in sub-blocks are listed in Table 5.

From Table 5, it can be seen that C (10) and Si (11) as the main components of steel have a major influence on TS. Appropriate amount of C and Si can improve the strength of steel and have little impact on plasticity. Therefore, the variables that have greatest impact on almost all sub-blocks are C and Si. B (23) only exists in the steel with high carbon content, which can indirectly reflect the carbon content and play an important role in some variable blocks. Mn (12), as the most important metal element in steel, can effectively improve the strength of steel without affecting plasticity, which also eliminate the adverse effects of S (14) and oxygen in steel. The simulation results agree with the facts, which indicates that the variable evaluation index used in the paper is appropriate. The most favorable variables in each sub-block are retained and establish the auxiliary dataset.

3) ON-LINE LOCAL MODELING

On-line prediction is under the framework of JITL. There are two key indicators to be determined in JITL, which are the number of modeling samples and similarity metric for samples. The number of modeling samples has a certain influence on prediction accuracy. If it is too small, the model cannot fully learn the characteristics of data. If it is too large, the model might be misled by some redundant samples. In either case, performance of the model will be reduced. Different distance metrics may select different historical samples for modeling. If the characteristics of the selected samples and the query sample differ greatly, structure and parameters of the model will also be different. Therefore, the proposed VRMD is compared with ED and MD under different number of samples. The result is shown in Figure.10.

In Figure.10, no matter which distance metric it is, with increasing of the number of modeling samples, prediction

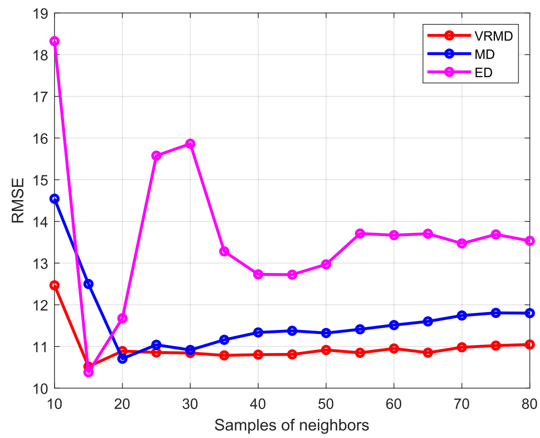

FIGURE 10. **Model performance of three similarity metrics under different sampling numbers.**

performance of the soft sensor increases first and decreases gradually. The main reason for the condition is that when the number of historical samples is small, the model cannot extract enough process information from those samples to predict mechanical properties. The learning ability of the model is restricted by the number of samples. As more and more samples are provided, although the information available for learning becomes sufficient, some redundant samples participate in the prediction, bringing some disturbing information, which result in slow or even sharp degradation of model performance. For example, the second peak of ED curve in Figure. 10 is the result of redundant information. When ED is used as the similarity metric, the model error is always bigger than those when using other similarity metrics. This is mainly because ED neglects variables with smaller values, and always tends to pay more attention to variables with larger values, even they have little to do with mechanical properties. VRMD considers the scale of variables and correlation between variables and output. Therefore, it is more stable than ED and MD when selecting samples. With the

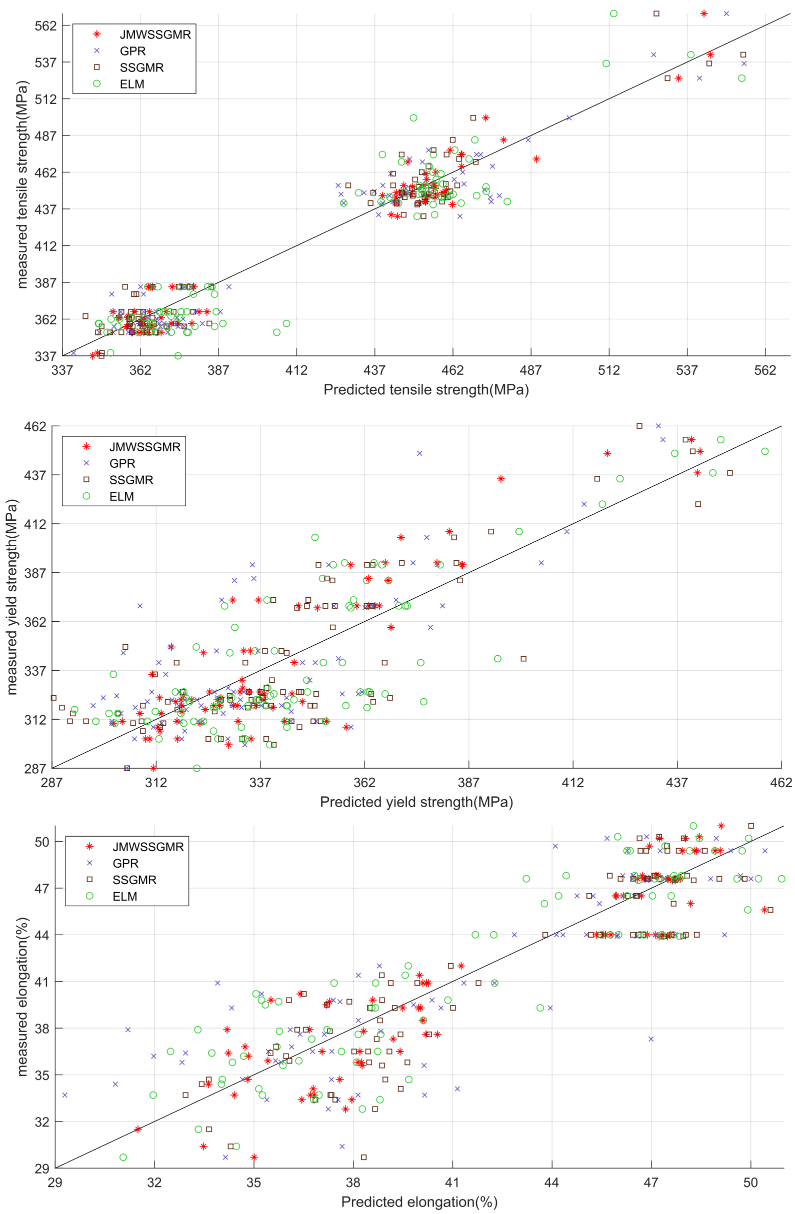

FIGURE 11. **Predictions of mechanical properties using different models.**

increasing of neighbors, the downward trend of performance is not obvious, which is better than other two similarity metrics.

In JITL framework, referring to Figure.10, we select 15 samples forming labeled set and unlabeled set as neighbors of the query sample, respectively. In order to verify the superiority of the proposed soft sensor in the prediction of mechanical properties of hot rolled strip, it is compared with GPR, SSGMR and ELM, and the comparison of different models are shown in Figure.11. The detailed numerical comparison is showed in Table 6. The regression evaluation indexes we use are:

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(y_i - \hat{y}_i\right)^2} \tag{35}$$

$$MAE = \frac{1}{m}\sum_{i=1}^{m}\left|\left(y_i - \hat{y}_i\right)\right| \tag{36}$$

$$R^2 = 1 - \frac{\sum_i \left(y_i - \hat{y}_i\right)^2}{\sum_i \left(y_i - \bar{y}_i\right)^2} \tag{37}$$

GPR and ELM are supervised models, and they discard unlabeled samples when training, which do not make use of process information existing in unlabeled samples. By contrast, SSGMR adds unlabeled samples in the training stage by introducing semisupervised learning, which improve utilization rate of sampling. However, it is a kind of global modeling method, and does not fully consider the characteristics of variables in hot rolling process. Moreover, the number of Gaussian components is hard to be determined, which may lead to the introduction of polluted samples with

TABLE 6. Prediction performance of different models.

Mechanical properties	Model	Regression evaluation index			
		RMSE	MAE	R^2	Accuracy
TS	GPR	13.61	10.51	0.93	72.62
	ELM	18.55	13.72	0.88	64.71
	SSGMR	12.86	10.19	0.94	80.00
	JMWSSGMR	**10.51**	**8.59**	**0.96**	**83.95**
YS	GPR	24.96	19.43	0.63	50.00
	ELM	24.83	20.35	0.63	47.06
	SSGMR	23.37	19.77	0.67	38.82
	JMWSSGMR	**18.83**	**14.75**	**0.77**	**62.16**
EL	GPR	3.30	2.51	0.67	63.10
	ELM	2.60	2.14	0.79	69.41
	SSGMR	2.58	2.05	0.80	74.12
	JMWSSGMR	**2.20**	**1.77**	**0.85**	**80.00**

unexpected characteristics. Compared with models above, the proposed JMWSSGMR adopts variable blocking strategy based on correlation, and constructs the auxiliary set based on improved VRMD distance metric under JITL framework, which effectively solves the problem that the modeling sample set is easy to be polluted. Then, we improve sampling utilization rate by introducing semisupervised learning in training stage. Finally, prediction stability is increased by fusing predicted values in each sub-blocks with uncertainty. It can be seen from Table 6 that proposed soft sensor has better performance than other soft sensors. Especially, at the tolerance level of 15 MPa with TS, YS and 3% with EL, the prediction accuracy of JMWSSGMR is higher, which verifies its superiority. These results clearly show that the proposed soft sensor can effectively deal with various problems in hot rolling, and it can be successfully applied in the prediction of mechanical properties of strips.

VI. CONCLUSIONS

To monitoring the mechanical properties of hot rolled strips accurately, a JITL based multi-block weighted semisupervised soft sensor is proposed in this paper. The characteristics of the soft sensor are:

1) In view of the complex correlations among variables in hot rolling process, we construct diverse multiple sub-blocks with low correlation by PLS, and fuse sub-blocks with a specific rule.
2) In order to distinguish samples from different working conditions, overall procedure is under the JITL framework. More concretely, an improved Mahalanobis distance metric (VRMD) is proposed. The most similar historical samples to the query sample are selected as modeling samples, so as to avoid redundant samples polluting dataset.
3) Semisupervised learning is introduced into modeling, which greatly improves the model performance when labeled samples are insufficient.

Similar to semisupervised learning, transfer learning is another hotspot in the field of soft sensing. In hot rolling process, we sometimes do not have enough samples in face of new kinds of steel. At this time, it is expected to learn process information from other kinds of steel and transfer the knowledge to the new condition. Transfer learning is especially applicable in this case, and it is also the research direction of us.

REFERENCES

[1] J. C. B. Gonzaga, L. A. C. Meleiro, C. Kiang, and R. Maciel Filho, "ANN-based soft-sensor for real-time process monitoring and control of an industrial polymerization process," *Comput. Chem. Eng.*, vol. 33, no. 1, pp. 43–49, Jan. 2009.

[2] M. Ren, Y. Song, and W. Chu, "An improved locally weighted PLS based on particle swarm optimization for industrial soft sensor modeling," *Sensors*, vol. 19, no. 19, p. 4099, Sep. 2019.

[3] I. Pisa, I. Santín, J. Vicario, A. Morell, and R. Vilanova, "ANN-based soft sensor to predict effluent violations in wastewater treatment plants," *Sensors*, vol. 19, no. 6, p. 1280, Mar. 2019.

[4] W. Zheng, Y. Liu, Z. Gao, and J. Yang, "Just-in-time semi-supervised soft sensor for quality prediction in industrial rubber mixers," *Chemometric Intell. Lab. Syst.*, vol. 180, pp. 36–41, Sep. 2018.

[5] Y. A. W. Shardt, S. Mehrkanoon, K. Zhang, X. Yang, J. Suykens, S. X. Ding, and K. Peng, "Modelling the strip thickness in hot steel rolling mills using least-squares support vector machines," *Can. J. Chem. Eng.*, vol. 96, no. 1, pp. 171–178, Jan. 2018.

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

[7] Z. Y. Liu, W. D. Wang, and W. Gao, "Prediction of the mechanical properties of hot-rolled C-Mn steels using artificial neural networks," *J. Mater. Process. Technol.*, vol. 57, nos. 3–4, pp. 332–336, Feb. 1996.

[8] H. J. Kim, M. Mahfouf, and Y. Y. Yang, "Modelling of hot strip rolling process using a hybrid neural network approach," *J. Mater. Process. Technol.*, vol. 201, nos. 1–3, pp. 101–105, May 2008.

[9] S. Lalam, P. K. Tiwari, S. Sahoo, and A. K. Dalal, "Online prediction and monitoring of mechanical properties of industrial galvanised steel coils using neural networks," *Ironmaking Steelmaking*, vol. 46, no. 1, pp. 89–96, Jan. 2019.

[10] S.-W. Wu, X.-G. Zhou, J.-K. Ren, G.-M. Cao, Z.-Y. Liu, and N.-A. Shi, "Optimal design of hot rolling process for C-Mn steel by combining industrial data-driven model and multi-objective optimization algorithm," *J. Iron Steel Res. Int.*, vol. 25, no. 7, pp. 700–705, Jul. 2018.

[11] L. Wang, D.-M. Fu, and W.-D. Yang, "A multiple SVR modeling of hot rolling process combined with kernel clustering and grey relational grade," in *Proc. Int. Conf. Mach. Learn. Cybern.*, Jul. 2012, pp. 387–393.

[12] W. Yan, D. Tang, and Y. Lin, "A data-driven soft sensor modeling method based on deep learning and its application," *IEEE Trans. Ind. Electron.*, vol. 64, no. 5, pp. 4237–4245, May 2017.

[13] Z.-W. Xu, X.-M. Liu, and K. Zhang, "Mechanical properties prediction for hot rolled alloy steel using convolutional neural network," *IEEE Access*, vol. 7, pp. 47068–47078, 2019.

[14] S. Wold, M. Sjöström, and L. Eriksson, "PLS-regression: A basic tool of chemometrics," *Chemometric Intell. Lab. Syst.*, vol. 58, no. 2, pp. 109–130, Oct. 2001.

[15] D.-Y. Huang, Z. Zhang, and S. S. Ge, "Speaker state classification based on fusion of asymmetric simple partial least squares (SIMPLS) and support vector machines," *Comput. Speech Lang.*, vol. 28, no. 2, pp. 392–419, Mar. 2014.

[16] M. Fan, Z. Ge, and Z. Song, "Adaptive Gaussian mixture model-based relevant sample selection for JITL soft sensor development," *Ind. Eng. Chem. Res.*, vol. 53, no. 51, pp. 19979–19986, Dec. 2014.

[17] X. Yuan, J. Zhou, and Y. Wang, "A spatial-temporal LWPLS for adaptive soft sensor modeling and its application for an industrial hydrocracking process," *Chemometric Intell. Lab. Syst.*, vol. 197, Feb. 2020, Art. no. 103921.

[18] X. Yuan, Z. Ge, B. Huang, Z. Song, and Y. Wang, "Semisupervised JITL framework for nonlinear industrial soft sensing based on locally semisupervised weighted PCR," *IEEE Trans. Ind. Informat.*, vol. 13, no. 2, pp. 532–541, Apr. 2017.

[19] W. Shao, Z. Ge, and Z. Song, "Bayesian Just-in-Time learning and its application to industrial soft sensing," *IEEE Trans. Ind. Informat.*, vol. 16, no. 4, pp. 2787–2798, Apr. 2020.

[20] X. Yuan, Z. Ge, B. Huang, and Z. Song, "A probabilistic Just-in-Time learning framework for soft sensor development with missing data," *IEEE Trans. Control Syst. Technol.*, vol. 25, no. 3, pp. 1124–1132, May 2017.

[21] X. Yuan, Z. Ge, and Z. Song, ‘‘Soft sensor model development in multiphase/multimode processes based on Gaussian mixture regression,’’ *Chemometric Intell. Lab. Syst.*, vol. 138, pp. 97–109, Nov. 2014.
[22] W. Shao, Z. Ge, and Z. Song, ‘‘Soft-sensor development for processes with multiple operating modes based on semisupervised Gaussian mixture regression,’’ *IEEE Trans. Control Syst. Technol.*, vol. 27, no. 5, pp. 2169–2181, Sep. 2019.
[23] Z. Ge, ‘‘Quality prediction and analysis for large-scale processes based on multi-level principal component modeling strategy,’’ *Control Eng. Pract.*, vol. 31, pp. 9–23, Oct. 2014.
[24] L. Xie, J. Zeng, and C. Gao, ‘‘Novel Just-In-Time learning-based soft sensor utilizing non-Gaussian information,’’ *IEEE Trans. Control Syst. Technol.*, vol. 22, no. 1, pp. 360–368, Jan. 2014.

JIE DONG received the B.E., M.E., and Ph.D. degrees from the University of Science and Technology Beijing, in 1995, 1997, and 2007, respectively. She is currently a Professor with the School of Automation and Electrical Engineering, University of Science and Technology Beijing. In 2004, she visited The University of Manchester as a Visiting Scholar. Her research interests include intelligent control theory and application, process monitoring and fault diagnosis, and complex system modeling and control.

YINGZE TIAN received the B.E. degree from the School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing, China, in 2008, where he is currently pursuing the M.E. degree in control science and engineering. His research interests include data-driven soft sensors and machine learning.

KAIXIANG PENG (Member, IEEE) received the B.E. degree in automation and the M.E. and Ph.D. degrees from the Research Institute of Automatic Control, University of Science and Technology Beijing, Beijing, China, in 1995, 2002, and 2007, respectively. He is currently a Professor with the School of Automation and Electrical Engineering, University of Science and Technology Beijing. His research interests include fault diagnosis, prognosis, and maintenance of complex industrial processes, modeling and control for complex industrial processes, and control system design for the rolling process.