Adaptive Weighting Just-in-Time-Learning Quality Prediction Model for an Industrial Blast Furnace

Kun CHEN1) and Yi LIU2)*

1) Department of Electrical and Information Engineering, Shaoxing University, Shaoxing, 312000 People’s Republic of China.
2) Institute of Process Equipment and Control Engineering, Zhejiang University of Technology, Hangzhou, 310014 People’s Republic of China.

(Received on May 15, 2016; accepted on September 8, 2016)

Development of accurate soft sensors for online quality prediction (e.g., silicon content) in an industrial blast furnace is a difficult task. A novel just-in-time-learning (JITL) prediction approach using adaptive feature-weighting for similar samples is developed. First, a dual-objective joint-optimization framework is proposed to introduce both input and output information into the model. Then, a suitable similarity criterion with feature weighting strategy is formulated, which is not considered in conventional JITL methods. Moreover, the trade-off parameter in the joint-optimization problem can be chosen automatically, without the time-consuming cross-validation procedure. The proposed method is applied to online predict the silicon content in an industrial blast furnace in China. Compared with other JITL-based soft sensors, better prediction performance has been obtained.

KEY WORDS: industrial ironmaking process; silicon content prediction; soft sensor; just-in-time learning model; feature weighting.

1. Introduction

The blast furnace ironmaking process is an important unit operation in the manufacturing of iron and steel. It consumes the main energy input to the integrated route of steel production and emits much carbon oxide which is the main cause of the greenhouse effect. As one of the most energy intensive and complicated industrial process, there has been a growing awareness of modeling and controlling the blast furnace ironmaking process for increasing efficiency and reducing cost. The silicon content, indicating the thermal state in blast furnace, is the most important index of pig iron quality. It must be kept at an appropriate level to facilitate the production and stable running of the ironmaking process. Therefore, accurate online prediction of the silicon content in hot metal is very critical.

Without substantial understanding of the complicated phenomenology, the data-driven soft sensor models can be built in a quick manner. Among them, SVR and LSSVR have shown promising prediction performance, especially when the training data are insufficient.

One main disadvantage of most existing data-driven modeling approaches for the silicon content is that a single global model is built. However, it is not enough to describe all the process characteristics only using a single model, especially for some complicated regions with insufficient information. To improve the prediction performance, Nurkkala et al. presented multiple autoregressive vector models to describe complex systems. On the other hand, although moving-window-based recursive soft sensors can gradually be adapted to new operational conditions, how to choose a suitable moving-window size for complex blast furnace ironmaking processes is difficult. Addition-
assigned with the same feature weight. This is inconsistent with many practical applications mainly because each input variable has its special impact on the final quality. Second, for JLSSVR and most current just-in-time modeling methods, similarity measurements are only considered on the input variables. However, the information of output variables is not utilized.

This work mainly develops an adaptive just-in-time-learning (JITL)-based local model for better prediction of the silicon content. First, a dual-objective optimization framework is proposed to preserve the local structure of both input and output variables. Then, according to the importance, weights are assigned to the variables in the projection space. Using the new similarity, the relevant samples can be selected and weighted. Moreover, the trade-off parameter in the optimization problem can be chosen in an efficient manner, without the time-consuming cross-validation. All these improvements can enhance the performance of JITL-based models.

The remainder of this paper is organized as follows. The JLSSVR soft sensor modeling method is described in Section 2. In Section 3, the detailed implementations of the adaptive weighting JLSSVR model are developed. It is evaluated by the silicon content online prediction in an industrial process in Section 4. Comparison studies with other methods are also investigated. Finally, a conclusion is made in Section 5.

2. Local Soft Sensor Model for Online Quality Prediction

2.1. Basic LSSVR Method

Using the kernel learning framework, the SVR/LSSVR soft sensor model is to learn a mapping \( f: \mathbf{X} \rightarrow \mathbf{Y} \) with a modeling set \( \mathbf{S} = \{ \mathbf{X}, \mathbf{Y} \} = \{ \mathbf{x}_i, \mathbf{y}_i \}_{i=1}^N \). A general form of the kernelized nonlinear model for process modeling can be formulated as\(^{38}\)

\[
y_i = f(\mathbf{x}_i; \theta) + e_i = f(\mathbf{x}_i; \mathbf{w}, \mathbf{b}) + e_i = \mathbf{w}^T \phi(\mathbf{x}_i) + b + e_i, i = 1, \ldots, N \quad \text{.............. (1)}
\]

where \( y_i \) and \( e_i \) denote the output measurement and the process noise at instance \( i \), respectively; \( \mathbf{x}_i \) is usually composed of several measured variables, and \( f \) is the wanted model; \( \phi \) is a feature map; \( \theta = [\mathbf{w}, b]^T \); that is, the symbols \( \mathbf{w} \) and \( b \) are the model parameter vector and the bias term, respectively.\(^{38}\) Applying the LSSVR framework to Eq. (7),\(^{33}\) the following optimization problem is formulated

\[
\begin{align*}
\min J(\mathbf{w}, b) &= \frac{1}{2} \left\| \mathbf{w} \right\|^2 + \frac{\gamma}{2} \left\| e \right\|^2 \\
\text{s.t.} \quad y_i - \mathbf{w}^T \phi(\mathbf{x}_i) - b - e_i = 0, i = 1, \ldots, N
\end{align*}
\quad \text{.............. (2)}
\]

where \( e = [e_1, \ldots, e_N]^T \) is the approximation error. The user-defined regularization parameter \( \gamma (\gamma > 0) \) determines the trade-off between the model’s complexity and approximation accuracy, and a suitable choice of \( \gamma \) can prevent over-fitting.\(^{33}\) After solving the optimization problem in Eq. (2),\(^{33}\) the solution can be expressed as

\[
\begin{align*}
\alpha &= \mathbf{P} \left[ \mathbf{y} - \mathbf{1}^T \mathbf{P} y \right] \\
b &= \mathbf{1}^T \mathbf{P} y \\
\end{align*}
\quad \text{.............. (3)}
\]

where \( \alpha = [\alpha_1, \ldots, \alpha_N]^T \) are Lagrange multipliers; \( \mathbf{y} = [y_1, \ldots, y_N]^T \); \( \mathbf{I} \in \mathbb{R}^{N \times N} \) is a unit matrix and \( \mathbf{1} \) is a vector of ones; \( \mathbf{P} = \mathbf{\Omega} + \gamma \mathbf{I} \). The kernel matrix \( \mathbf{\Omega} \) is composed of elements \( \Omega(i,j) = \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \), \( \forall i, j = 1, \ldots, N \). Briefly, the development of an LSSVR-based soft sensor model amounts to solving a set of linear equations using the kernel transform. Compared with solving a convex quadratic programming problem of the SVR method, LSSVR can be implemented more efficient and easily.\(^{33}\)

Finally, the LSSVR model for prediction of a test sample \( \mathbf{x}_q \), i.e., \( \hat{y}_q \), can be obtained\(^{33}\)

\[
\hat{y}_q = f(\theta, \mathbf{x}_q) = \sum_{i=1}^{N} \alpha_i \phi(\mathbf{x}_i), \phi(\mathbf{x}_q)) + b \quad \text{.............. (4)}
\]

where \( \alpha_i = [\alpha_i, \ldots, \alpha_i]^T \) are Lagrange multipliers related to the model parameter vector \( \mathbf{w} \), and \( \mathbf{k}_i(i) = \phi(\mathbf{x}_i), \phi(\mathbf{x}_q)) \), \( \forall i = 1, \ldots, N \) is a kernel vector.

2.2. JLSSVR-based Local Model

For some industrial processes, the direct application of a global/fixed model with complicated structure is often difficult. Another limitation of the global methods is that it is difficult for them to be updated in a quick manner when the process dynamics are changing.\(^{37}\) Additionally, for some situations with more complex characteristics, the training data samples are not sufficient. Therefore, only using an LSSVR/SVR model for industrial ironmaking processes is still not enough.

To alleviate these problems and construct the local models automatically, the JITL method inspired by the ideas of local modeling has been developed as an alternative to nonlinear process modeling and control.\(^{34-40}\) As illustrated in Fig. 1, for a query sample \( \mathbf{x}_q \), there are three steps to construct a JITL model (take JLSSVR for example). First, select similar samples as a similar set \( \mathbf{S}_q \) in the database \( \mathbf{S} \) based on some defined similarity criterions. Second, construct a JLSSVR model \( f_{\text{LSSVR}}(\mathbf{x}_q) \) using the relevant dataset

![Fig. 1. The common flowchart of JITL-based online soft sensor modeling method.](image-url)
S_i. Third, online predict the output ŷ_i for the current query sample x_q and then discard the JLSSVR model \text{JLSSVR}(x_q).^{40}

With the same three-step implements, a new JLSSVR model can be constructed for the next query sample. The Euclidean distance-based similarity is commonly utilized.\textsuperscript{14,40} The similarity index (SI) between the query sample x_q and the sample x_i in the dataset is defined below.\textsuperscript{25}

\[
S_{x,q} = \exp(-d_{x,q}) = \exp(-[|x_i - x_q|]) \text{, } i = 1, \cdots, N \text{ .... (5)}
\]

where \(d_{x,q}\) is the distance similarity between \(x_i\) and \(x_q\) in the dataset. The value of \(S_{x,q}\) is bounded between 0 and 1 and when \(S_{x,q}\) approaches to 1, \(x_q\) resembles \(x_i\) closely. With the distance-and-angle-based similarity factor, a little better prediction performance can be obtained. However, another parameter for balance of the distance and the angle should be chosen.\textsuperscript{36} Additionally, some correlation-based similarity criteria have been proposed recently.\textsuperscript{37} Based on these similarity criteria, a relevant dataset \(S_p\) with \(n_p\) similar samples can be adopted to build an JLSSVR model.

For construction of the JLSSVR model, the user-defined parameters include the regularization parameter \(\gamma\) and the kernel parameter \((e.g., the Gaussian kernel K(x, x_\gamma) = \exp(-|x_i - x_j|/\sigma) \text{ with the width parameter } \sigma > 0\). With a pair of parameters \([\gamma; \sigma]\), the JLSSVR model can be constructed. The leave-one-out (LOO) cross-validation has been shown to provide a sensible criterion for model selection.\textsuperscript{40} However, the LOO criterion for LSSVR is computationally expensive and, therefore, is not suitable for online performance. Fortunately, a fast LOO (FLOO) cross-validation criterion was proposed by Cawley and Talbot.\textsuperscript{41} Based on the FLOO criterion,\textsuperscript{40,41} the JLSSVR model can be optimized by solving the minimum FLOO-based error with \(n_p\) samples.

\[
E_{\text{FLOO}} = \sum_{i=1}^{n_p} \|x_i - \hat{x}_i\|^2 \text{ , } \hat{x}_i = \sum_{j=1}^{n_p} \frac{\alpha_i}{P_i + s_i^2/\|x_i\|^2} \text{ .... (6)}
\]

where \(P_i\) is the item at the \(i\)th row and \(i\)th column of \(P\) in Eq. (3), \(s = P1 = [s_1, \cdots, s_{n_p}]^T\) and \(o = -1P1\). The related terms \((i.e., P \text{ and } \alpha)\) are available. Additionally, the computational load of \(s\) and \(o\) is small. The complexity of FLOO can be reduced to about \(O(n_p^2)\) operations, compared to the basic LOO with about \(O(n_p^3)\) operations.\textsuperscript{41} Consequently, the computation is much more efficient and online selection of the parameters of JLSSVR is feasible.

3. Adaptive Weighting Local Soft Sensor Model

Although the criterion in Eq. (5) has been widely adopted in JITL-based methods for process modeling and monitoring, the similarity still may not be described adequately in two folds. First, different input variables are equally weighting, which is inconsistent with many practical situations. Additionally, similarity measurements are only built on the input variables \((i.e., \text{ in Eq. (5)})\). No information in output quality variables is used.

Recently, a supervised locality preserving projection (LPP) strategy\textsuperscript{42} was proposed to construct the similarity measurement using both of input and output information.\textsuperscript{35} With the similarity defined in Eq. (7), an LPP approach is employed to seek the mapping direction, which serves as the weights for variables. Then, relevant samples are selected according to the Euclidean distance in the projection space. Unfortunately, there are still two drawbacks in the LPP-based similarity method. First, the parameter \(\eta\) is selected through the cross-validation method, which is really time-consuming. Second, in the low-dimensional space, the internal variables are assigned with the equal weights, which may cause the same problem as the traditional criterion in.\textsuperscript{36}

The proposed relevant sample selection strategy is to construct an adaptive weighted distance as the similarity criterion by adequately utilizing the information in both of input and output variables. Consequently, the LPP algorithm\textsuperscript{42} is adopted to keep the local structure of both input and output information. And it is solved in a dual-objective optimization form as a general eigen problem. Then, according to the eigenvalue of projection direction, adaptive weights are assigned to the internal variables to construct the similarity criterion to select relevant samples.

3.1. Feature Extraction Using LPP

One challenge to establish the suitable similarity criterion is to select representative features from various input variables while preserving the local structure as much as possible. To achieve this goal, the LPP approach is employed. LPP\textsuperscript{42} is a recent dimensionality reduction method which has been successfully applied in information retrieval and pattern recognition areas. LPP aims at preserving the neighborhood structure of the data set, while principal component analysis (PCA) only retains most of the original variance.\textsuperscript{42} Additionally, LPP shares many of the properties of nonlinear methods such as locally linear embedding\textsuperscript{43} and Laplacian eigenmaps.\textsuperscript{44} Therefore, compared with PCA, LPP can reveal the intrinsic geometrical structure of the observed data and find more meaningful low-dimensional information hidden in the high-dimensional observations. Moreover, the linear property of LPP makes it efficient in computation and suitable for practical applications.\textsuperscript{42}

Given a set of \(m\)-dimensional input variables \(X = \{x_1, \cdots, x_m\}\) with corresponding output variables \(Y = \{y_1, \cdots, y_m\}\) where \(x_i \in \mathbb{R}^m\), LPP aims to find a transformation matrix \(B = [b_1, \cdots, b_k]\) to project these samples to a low-dimensional sample set \(Z = \{z_1, \cdots, z_n\}\) in \(\mathbb{R}^d (d \leq m)\) based on following objective function:\textsuperscript{42}

\[
J_{\text{LPP}} = \min \sum_{i,j=1,\cdots,N} \left(z_i - z_j\right)^2 S_{ij} \text{ .... (8)}
\]

where \(S_{ij}\) indicates the similarity between the \(i\)th and \(j\)th samples, and \(z_i\) is the one-dimensional representation of \(x_i\) with a projection vector \(\beta_i\), i.e., \(z_i = \beta_i x_i\), with \(z_i\) as the vector of most important \(d\) representation of \(x_i\).

To better utilizing both the secondary and primary information, a dual-objective optimization scheme is proposed:
\[
J_x = \min \sum_{i,j=1,\ldots,N} \|z_i - \eta_i S_{x,ij}\|^2_S \quad \text{subject to} \quad \sum_{i,j=1,\ldots,N} \|z_i - \eta_i S_{x,ij}\|^2_S = 1 \quad \text{subject to} \quad \sum_{i,j=1,\ldots,N} \|z_i - \eta_i S_{x,ij}\|^2_S = 1 \quad \text{(9)}
\]

\[
J_y = \min \sum_{i,j=1,\ldots,N} \|z_i - \eta_i S_{y,ij}\|^2_S \quad \text{subject to} \quad \sum_{i,j=1,\ldots,N} \|z_i - \eta_i S_{y,ij}\|^2_S = 1 \quad \text{(10)}
\]

It is expected that, after projection, the nearby points, with both similar input and output variables, are still close in the low-dimensional projection space. However, these two objectives are generally difficult to be obtained at the same time, as the projection directions are usually different. To solve this problem, a trade-off parameter \(\eta_1\) is introduced to balance these two objectives. Then, the objective function can be described as follows:

\[
J = \min \eta_1 J_x + (1 - \eta_1) J_y = \min \sum_{i,j=1,\ldots,N} \|z_i - \eta_i S_{x,ij} + (1 - \eta_i) S_{y,ij}\|^2_S \quad \text{subject to} \quad \sum_{i,j=1,\ldots,N} \|z_i - \eta_i S_{x,ij} + (1 - \eta_i) S_{y,ij}\|^2_S = 1 \quad \text{(11)}
\]

where \(0 \leq \eta_1 \leq 1\). Details on the selection of parameter \(\eta_1\) will be presented in Section 3.3.

To solve this optimization problem, an orthogonal constraint \(\beta^T \beta = 1\) is introduced to avoid singularity problem. Then, the optimization problem can be solved through an eigen problem:

\[
XLX^T \beta = \lambda \beta \quad \text{(12)}
\]

where \(L = D - SI\), with \(SI_{ij} = \eta_i S_{x,ij} + (1 - \eta_1) S_{y,ij}\) and \(D\) a diagonal matrix defined as \(D_{ii} = \sum_{j=1}^N SI_{ij}\). The items \(\beta_1, \ldots, \beta_d\) are the eigenvectors corresponding to the smallest \(d\) eigenvalues \(\lambda_1, \ldots, \lambda_d\). And the internal variables \(z\) can be obtained as \(z_i = B^T x_i\), with \(B = [\beta_1, \ldots, \beta_d]\).

### 3.2. Adaptive Weighting Similarity Criterion

Instead of using the Euclidean distance directly (e.g., Eq. \(5\)), a weighted distance is adopted here:

\[
d_{w_{ij}} = \sqrt{\sum_{k=1}^d v_k (z_{ik} - z_{jk})^2} \quad \text{(13)}
\]

where \(v_k\) is the feature weight parameter to balance the components of \(z_i\) and \(z_k\) indicates the \(k^{th}\) element in \(z\).

Inspired by the principle component analysis, an eigenvalue based weighting strategy is formulated to calculate the importance of the latent variables:

\[
v_k = \frac{\lambda_k^{-1}}{\sum_{i=1}^d \lambda_i^{-1}} \quad \text{(14)}
\]

With Eq. \(13\), the element with a larger eigenvalue will be assigned a relative smaller importance (≈0) and can be ignored to reduce the projection dimension automatically. Then, according to the weighted distance between the training data and the query sample, \(\eta_1\) samples with the largest weighted similarity measurements are selected as relevant samples.

A simple illustration of sample selection in two-dimensional projection space is shown in Fig. \(2\) to distinguish the proposed strategy and the criterion in. \(39\) Suppose the values of \(\beta_1, \beta_2\) are the mapping direction corresponding to the \(\lambda_1, \lambda_2\), respectively. The shape of samples with the same similarity indices would be an ellipse in with the long-axis-to-short-axis ratio \(a : b = \lambda_1^{-1} : \lambda_2^{-1}\), rather than a circle in. \(39\) It is straightforward to observe that the more relevant feature would be assigned a larger weight and play a more important role to determine the similarity measurement, which can make the criterion more accurate.

### 3.3. Auto Selection of Parameter \(\eta_1\)

For the dual-objective optimization problem in Eq. \(9\), it is difficult to obtain an optimal solution mainly due to the conflict between the two sub-objectives. To obtain a relatively good solution, the scales and convergence speed of the sub-objectives should be considered carefully to select the parameter \(\eta_1\).

The solution of Eq. \(10\) is obtained by solving the eigen problem rather than using an iterative manner. As a result, the convergence speed problem can be avoided. Thus, the parameter \(\eta_1\) should be selected to balance the scale issue. Inspired by, \(46\) the scale of \(J_x\) and \(J_y\) can be defined as:

\[
S_{\text{in}} = \rho(XL_x X^T) \quad \text{subject to} \quad \sum_{i,j=1,\ldots,N} \|z_i - \eta_i S_{x,ij} + (1 - \eta_i) S_{y,ij}\|^2_S = 1 \quad \text{(15)}
\]

where \(L_x\) and \(L_y\) are defined based on \(S_{x,ij}\) and \(S_{y,ij}\), respectively, \(\rho(\cdot)\) indicates the largest eigenvalue of the matrix \(\cdot\). Thus, \(\eta_1\) can be obtained by:

\[
\eta_1 = \frac{S_{\text{in}}}{S_{\text{out}} + S_{\text{in}}} \quad \text{(16)}
\]

In summary, the steps to select the relevant samples for construction of a JLSSVR model can be described as follows:

Step 1: Establish the matrices \(S_x, S_y\) (The elements in \(S_x\) and \(S_y\) are \(S_{x,ij}\) and \(S_{y,ij}\), respectively.), \(L_x\) and \(L_y\).

Step 2: Set \(\eta_1\) automatically according to Eqs. \((14)\) and \((15)\).

Step 3: Solve the optimization problem in Eq. \((10)\) to obtain the corresponding eigenvalues and eigenvectors.

Step 4: Assign adaptive weights to the latent variables and calculate the weighted distance based on Eqs. \((12)\) and \((13)\).

Step 5: Select \(n_0\) relevant samples on the basis of the weighted similarity measurement.

Step 6: The JLSSVR soft sensor model can be online.
It should be noted that, by introducing the dual-objective optimization scheme, the local structure of both input and output information can be better preserved. Additionally, using the LPP method, the new variables will be independent with each other by removing the correlation among the process variables. Moreover, the adaptive weighted similarity criterion keeping the local information can help estimate the similarity and select the relevant samples. Consequently, with the properly selected relevant samples, the built local model in a JITL manner would be more accurate to predict the output variables.

4. Industrial Silicon Content Prediction

In this section, the proposed improved JITL-based soft sensor modeling methods are applied to online predict the silicon content of an industrial blast furnace ironmaking process in China. All the data samples have been collected from daily process records and the corresponding laboratory analysis. The process input variables correlated with the product quality (i.e., the silicon content) have been selected.\(^\text{21,22,24}\) These input variables include the blast volume, the blast temperature, the top pressure, the gas permeability, the top temperature, the ore/coke ratio, and the pulverized coal injection. The sampling time of most of these input variables is 1 minute. According to expert experience and correlation analysis,\(^\text{30}\) the time difference between the silicon content and input variables can be selected. For example, the time difference between the silicon content and the top pressure is about 2 h; and the time difference between the silicon content and the gas permeability is about 1 h.

After simply preprocessing the modeling set with 3-sigma criterion, most of obvious outlier samples and missing values have been removed out. After preprocessing, a set of about 260 samples is investigated. The first 150 samples are treated as the historical samples. The rest set of about 110 samples is for testing. The simulation environment in this case is MatLab V2009b with CPU main frequency 2.3 GHz and 4 GB memory.

As discussed in previous research, only a single global model is not enough for description of all the process characteristics in industrial ironmaking processes.\(^\text{40}\) Additionally, the blast furnace dynamics may change, gradually or abruptly, and a fixed soft sensor model validated on earlier data may not perform well on future data.\(^\text{29}\) Here, JLSSVR is considered as a JITL-based local modeling method. To better illustrate the effect of the proposed method, the adaptive weighted relevant sample selection strategy is applied to acquire the relevant data set for each query sample. Two cases are investigated here. The first one in Section 4.1 is to study the performance of eigenvalue based adaptive weighting approach only using the information in input variables, i.e. \(\eta = 1\) in Eq. (7). The other one in Section 4.2 focuses on the utilizing both of the information in input and output variables and how to auto-select the trade-off parameter \(\eta_1\) in Eq. (10).

The root-mean-square error (RMSE) and relative RMSE (simply noted as RE) are two common performance indices to quantitatively evaluate the prediction performance of different soft sensor models.\(^\text{40}\) Additionally, the hit rate (HR) index is often adopted in industrial blast furnace ironmaking processes.\(^\text{21-28}\) Three indices of RMSE, RE, and HR are defined, respectively.

\[
\text{RMSE} = \sqrt{\frac{1}{l} \sum_{q=1}^{l} (y_q - \hat{y}_q)^2} \quad \text{(16)}
\]

\[
\text{RE} = \frac{100}{l} \sum_{q=1}^{l} \left(1 - \frac{\hat{y}_q}{y_q}\right)^2 \quad \text{(17)}
\]

\[
\text{HR} = \frac{\sum_{q=1}^{l} H_q}{l} \times 100\% \quad \text{(18)}
\]

where \(\hat{y}_q\) and \(y_q\) are the predicted value and the actual value, respectively. And \(l\) is the number of test samples.

4.1. Performance of Weighting Strategy

As aforementioned, the comparison of the feature weighting strategies is investigated only using the information in input variables, i.e. \(\eta_1 = 1\) in Eq. (7). Two following weighting strategies are considered for the variables in the projected space: (1) equal weighting, where the values of \(v_k\) are assigned as \(v_1 = \cdots = v_q\) in Eq. (12); (2) eigenvalue based weighting, with its weights assigned as Eqs. (12) and (13).

The prediction results of these two weighting strategies are listed in Table 1 with the number of relevant samples \(n_q = 10, 20, 30, 40\), respectively. Additionally, the online prediction results of the silicon content with \(n_q = 30\) are shown in Fig. 3. To show the result more clear, only the first 50 testing samples are plotted in Fig. 3. From Fig. 3 and Table 1, it can be shown that, regardless of the value of \(n_q\) (\(n_q = 10, 20, 30, 40\)), the eigenvalue based weighting criterion outperforms the equal weighting strategy, with the smaller RMSE and RE indices, and the larger HR value. For this case, the number of similar samples \(n_q = 30\) is a suitable choice.

The prediction results indicate that the eigenvalue based weighting criterion can help determine the feature weights.

| Similar samples | Feature weighting strategy | RMSE | RE (%) | HR (%) |
|-----------------|---------------------------|------|--------|--------|
| 10              | Eigenvalue based weighting| 0.099| 19.1   | 68.2   |
|                 | Equal weight              | 0.108| 20.3   | 65.4   |
| 20              | Eigenvalue based weighting| 0.097| 18.2   | 70.1   |
|                 | Equal weight              | 0.107| 20.1   | 67.3   |
| 30              | Eigenvalue based weighting| 0.095| 18.1   | 71.0   |
|                 | Equal weight              | 0.105| 19.9   | 69.2   |
| 40              | Eigenvalue based weighting| 0.101| 18.9   | 69.2   |
|                 | Equal weight              | 0.109| 20.6   | 66.4   |
and construct the similarity measurement more suitably. This is mainly because the feature weights are assigned according to their relative importance. Consequently, the local structures of initial data samples can be kept and the similarity criterion can be estimated more properly.

### 4.2. Performance of Utilizing the Output Information

To show the effect of utilizing the output variable information of the training dataset, five LPP-based similarity criteria are utilized to search the mapping direction, followed by the eigenvalue based weighting strategy to calculate the weighted distance in projection space. They are listed as: (1) $S_x$, only using the input variable information of the training samples; (2) $S_y$, only using the output information of the training samples; (3) $S_xS_y$, that is the criterion $S_{x,y} = S_y^2/S_{y,x}^2$ defined in Eq. (7), with the trade-off parameter $\eta$ determined by the cross-validation approach; (4) the proposed criterion $J = \min \eta \lambda_x + (1-\eta) \lambda_y$ in Eq. (10) with $\eta_1$ selected by cross-validation, noted as PC$_{cv}$; (5) the proposed criterion $J = \min \eta \lambda_x + (1-\eta) \lambda_y$ in Eq. (10) with auto parameter setting for $\eta_1$, noted as PC$_{auto}$ for simplicity.

The comparison results of the aforementioned criteria with $n_q = 30$ are tabulated in Table 2. From Table 2, one can see that a better prediction performance can be obtained by utilizing the information in both of input and output variables, compared with only using $S_x$ or $S_y$. The output variables contain some important information and should be adopted. However, it was not explored in most traditional JITL-based soft sensors in industrial processes. The prediction results of traditional JLSSVR without LPP-based feature weighting for searching the similar set are also listed in Table 2. It is inferior to other methods with LPP-based similarity criteria. The adaptive weighting JLSSVR (i.e., PC$_{auto}$) and traditional JLSSVR methods are compared using a parity plot shown in Fig. 4. It can be indicated that the adaptive weighting JLSSVR method is more accurate than JLSSVR (the HR index increases from 66.4% to 73.8%).

Moreover, it is shown that the cross-validation based criterion (PC$_{cv}$) can also make the model predict well. However, for a query sample with $n_q = 30$, the total computational time for online modeling and prediction is about 11 s, which is much larger than using the PC$_{auto}$ method (less than 1 s). It is rather time-consuming for the PC$_{cv}$ method mainly because all the possible values of $\eta_1$ need to be computed for the proper value by cross-validation. Meanwhile, for the proposed auto parameter setting strategy defined in Eqs. (14) and (15) (PC$_{auto}$), the computation is more efficient because only a generalized eigen problem is solved. As illustrated in Fig. 5, the predictive performance of the RE index with possible trade-off parameters $\eta_1$ is presented. It is shown that the auto-setting parameter can be a good and suitable choice in practice although it is suboptimal. This is mainly because the optimal value of $\eta_1$ shown in Fig. 5 is not known beforehand. In a word, the proposed similarity PC$_{auto}$ can achieve almost the best performance as the traditional cross-validation method, while saving a lot of computational time.

Therefore, from all the obtained results and comparison analysis, the proposed PC$_{auto}$ approach by utilizing both of the input and output information with automatic trade-off

| JLSSVR-based models | Brief description | RMSE | RE (%) | HR (%) |
|----------------------|------------------|------|--------|--------|
| $S_x$                | LPP-based input variable information | 0.095 | 18.2 | 71.0 |
| $S_y$                | LPP-based output variable information | 0.097 | 18.3 | 70.1 |
| $S_xS_y$             | LPP-based input and output variable information | 0.093 | 18.0 | 72.0 |
| PC$_{cv}$            | LPP-based feature weighting for input and output variable information (with $\eta$ selected by cross-validation) | 0.087 | 17.3 | 73.8 |
| PC$_{auto}$          | LPP-based feature weighting for input and output variable information (with auto parameter setting for $\eta_1$) | **0.086** | **17.4** | **73.8** |

Table 2. Comparisons of the online prediction errors using six different JLSSVR soft sensor models (best results are bold and underlined).
and the auto-selection of the trade-off parameter determination and the eigenvalue based weighting strategy, can achieve a promising prediction performance. Moreover, compared with the traditional cross-validation approach, it can be implemented in an efficient manner.

5. Conclusion

This paper has proposed a novel JITL-based local soft sensor model with adaptive relevant sample selection for better prediction of the silicon content. The main contributions are three folds: (1) the eigenvalue based adaptive weighting strategy, (2) using a dual-objective optimization framework for employment of the output information, (3) and the auto-selection of the trade-off parameter $\eta_1$ in an efficient manner. The superiority of the proposed method is demonstrated and compared with several JLSSVR soft sensors in terms of online prediction of the silicon content in an industrial blast furnace. Note that other JITL-based modeling methods can also be integrated with the proposed relevant sample selection strategy. Additionally, some advanced outlier detection methods can be applied as a preprocessing method to enhance the reliability of quality prediction. Therefore, there are several interesting research directions worth investigating to further enhance the accuracy and transparency of a silicon content prediction model.

Acknowledgment

The authors would like to gratefully acknowledge the National Natural Science Foundation of China (Grant No. 61004136) and Jiangsu Key Laboratory of Process Enhancement & New Energy Equipment Technology (Nanjing University of Technology) for their financial support.

Abbreviations

FLOO: fast leave-one-out
HR: hit rate
JITL: just-in-time learning
JLSSVR: just-in-time least squares support vector regression

LPP: locality preserving projection
LSSVR: least squares support vector regression
RE: relative root-mean-square error
RMSE: root-mean-square error
SI: similarity index
SVR: support vector regression

REFERENCES

1) K. Sugawara, K. Morimoto, T. Sugawara and J. S. Dranoff: *AIChE J.*, 45 (1999), 574.
2) X. G. Bi, K. Torrssel and O. Wijk: *ISIJ Int.*, 32 (1992), 481.
3) H. Nogami, M. S. Chu and J. Yaji: *Comput. Chem. Eng.*, 29 (2005), 2438.
4) K. Nishioika, T. Mueda and M. Shimizu: *ISIJ Int.*, 45 (2005), 669.
5) V. R. Radhakrishnan and K. M. Ram: *Process Control*, 11 (2001), 565.
6) S. Ueda, S. Natsui, H. Nogami, J. I. Yagi and A. Tatsuro: *ISIJ Int.*, 50 (2010), 914.
7) H. Singh, N. V. Sridhar and B. Deo: *Steel Res. Int.*, 67 (1996), 521.
8) V. R. Radhakrishnan and A. R. Mohamed: *Process Control*, 10 (2000), 509.
9) J. Chen: *Eng. Appl. Artif. Intell.*, 14 (2001), 77.
10) J. Jimenez, J. Mochon, J. S. de Ayalaí and F. Obeso: *ISIJ Int.*, 44 (2004), 573.
11) H. Saxon and F. Petterson: *ISIJ Int.*, 47 (2007), 1732.
12) F. Petterson, N. Chakraborti and H. Saxon: *Appl. Soft Comput.*, 7 (2007), 387.
13) A. Nurkkala, F. Petterson and H. Saxon: *Ind. Eng. Chem. Res.*, 50 (2011), 9236.
14) X. J. Hao, F. M. Shen, G. Du, Y. S. Shen and Z. Xie: *Steel Res. Int.*, 76 (2005), 694.
15) T. Bhattacharya: *ISIJ Int.*, 45 (2005), 1943.
16) R. D. Martin, F. Obeso, J. Mochon, R. Barea and J. Jimenez: *Iron-making Steelmaking*, 34 (2007), 241.
17) M. Waller and H. Saxon: *Indust. Eng. Chem. Res.*, 39 (2000), 982.
18) C. H. Gao, Z. M. Zhou and J. M. Chen: *Ind. Eng. Chem. Res.*, 10 (2008), 3037.
19) M. Waller and H. Saxon: *ISIJ Int.*, 42 (2002), 316.
20) S. H. Luo, C. H. Gao, J. S. Zeng and S. H. Luo: *IEEE Trans. Ind. Electron.*, 59 (2012), 1134.
21) L. Jian, S. Q. Shen and Y. Q. Song: *J. Appl. Math.*, (2012), Article ID 949654.
22) C. H. Gao, J. S. Zeng, X. Y. Liu and X. Y. Sun: *AIChE J.*, 55 (2009), 947.
23) C. H. Gao, J. S. Zeng and Z. M. Zhou: *AIChE J.*, 57 (2011), 3448.
24) Y. X. Chu and C. H. Gao: *AIChE J.*, 60 (2014), 2197.
25) S. H. Luo, C. H. Gao, J. S. Zeng and J. Huang: *Asian J. Control*, 15 (2013), 553.
26) A. Nurkkala, F. Petterson and H. Saxon: *ISIJ Int.*, 52 (2012), 1763.
27) H. Saxon, C. H. Gao and Z. W. Gao: *IEEE Trans. Ind. Inform.*, 9 (2013), 2213.
28) M. Kano and Y. Nakagawa: *Comput. Chem. Eng.*, 32 (2008), 12.
29) P. Kadlec, B. Gabrys and S. Strandt: *Comput. Chem. Eng.*, 33 (2009), 795.
30) J. A. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor and J. Vandewalle: *Least Squares Support Vector Machines, World Scientific, Singapore*, (2002).
31) Y. Liu and J. Chen: *J. Process Control*, 23 (2013), 793.
32) G. Bontempi, M. Birattari and H. Bersini: *Int. J. Control*, 72 (1999), 643.
33) C. Cheng and M. S. Chiu: *Chem. Eng. Sci.*, 59 (2004), 2801.
34) K. Fujiwara, M. Kano, S. Hasebe and A. Takinami: *AIChE J.*, 55 (2009), 1754.
35) Y. Liu, Z. L. Gao, P. Li and H. Q. Wang: *Ind. Eng. Chem. Res.*, 51 (2012), 4313.
36) K. Chen, J. Ji, H. Q. Wang, Y. Liu and Z. H. Song: *Chem. Eng. Res. Des.*, 89 (2011), 2117.
37) Y. Liu and Z. L. Gao: *Ironmaking Steelmaking*, 42 (2015), 321.
38) G. C. Cawley and N. L. C. Talbot: *Neural Netw.*, 17 (2004), 1467.
39) X. F. He and P. Niyogi: *Proc. Conf. on Advances in Neural Information Processing Systems, MIT Press, Massachusetts*, (2004).
40) S. T. Roweis and L. K. Saul: *Science*, 290 (2000), No. 5500, 2323.
41) M. Belkin and P. Niyogi: *Neural Comput.*, 15 (2003), No. 6, 1373.
42) X. Q. Liu, U. Kruger, T. Littler, L. Xie and S. Q. Wang: *Chemom. Intell. Lab. Syst.*, 96 (2009), 132.
43) M. G. Zhang, Z. Q. Ge, Z. H. Song and R. W. Fu: *Ind. Eng. Chem. Res.*, 50 (2011), 6837.

Fig. 5. Parity plot based of assay values against the prediction values of the silicon content in the test set using the adaptive weighting JLSSVR and traditional JLSSVR soft sensor models.