*Article*

# Soft Sensing of Silicon Content via Bagging Local Semi-Supervised Models

**Xing He [1], Jun Ji [2], Kaixin Liu [1], Zengliang Gao [1]⬤ and Yi Liu [1,*]⬤**

[1]   Institute of Process Equipment and Control Engineering, Zhejiang University of Technology,
   Hangzhou 310023, China
[2]   College of Computer Science and Technology, Qingdao University, Qingdao 266071, China
*   Correspondence: yliuzju@zjut.edu.cn; Tel./Fax: +86-571-85290402

**Abstract:** The silicon content in industrial blast furnaces is difficult to measure directly online. Traditional soft sensors do not efficiently utilize useful information hidden in process variables. In this work, bagging local semi-supervised models (BLSM) for online silicon content prediction are proposed. They integrate the bagging strategy, the just-in-time-learning manner, and the semi-supervised extreme learning machine into a unified soft sensing framework. With the online semi-supervised learning method, the valuable information hidden in unlabeled data can be explored and absorbed into the prediction model. The application results to an industrial blast furnace show that BLSM has better prediction performance compared with other supervised soft sensors.

**Keywords:** soft sensor; silicon content; semi-supervised learning; extreme learning machine; just-in-time-learning

## 1. Introduction

As a type of metallurgical furnace, blast furnaces are used for smelting to produce industrial metals. The silicon content in hot metal, both as a quality factor and as a chief indicator of the thermal level of the blast furnace, is of central importance. However, it is difficult to measure it online. Additionally, chemical reactions and transfer phenomena in blast furnaces are very complex. Till now, a reliable first-principles model for industrial practice is not available [1–5]. Because of the complexity of the task, the prediction of the silicon content is tricky and it recently attracted much attention. In past two decades, several data-driven soft sensors were proposed to predict the silicon content [6–24]. For example, some neural networks (NNs) [6–10] and support vector regression (SVR) [17,18] were built as a black-box predictor. The partial least squares regression [11], fuzzy logic approach [12], nonlinear time series analysis [14–16], chaos-based iterated multistep predictor [19], multiscale modeling methods [20], and multiple models [21], were also applied to predict the silicon content. Saxén et al. gave a review on data-driven discrete time models for the hot metal silicon content prediction in the blast furnace [22]. These empirical data-driven soft sensor models can be built quickly using the available measured variables [25–32].

In industrial processes, a large amount of sensor variables that are available can be used as input to the soft sensor model. The quality-relevant variable to be predicted using a soft sensor can be regarded as "labeled" data. However, the amount of quality-relevant variable ("labeled" data) is often limited mainly because it is difficult to measure online. Till now, most soft sensors in industrial ironmaking processes act in a supervised manner. That is, for construction of a soft sensor, both of inputs (sensor variables) and outputs (quality-relevant variables) are required for the task of supervised modeling. The labeled dataset contains both input and output data, while the unlabeled one consists of only input data (i.e., large amount of sensor variables). Actually, the labeled data are much fewer

than the unlabeled data mainly because the assaying process of silicon contents is infrequent and time-consuming. In contrast, the process input variables are measured frequently. Using a limited set of labeled data, the soft sensors are often inaccurate. To enhance the prediction performance, with large amounts of unlabeled data available, some semi-supervised soft sensors have been applied to chemical processes [33–35]. Therefore, the information hidden in unlabeled data is explored to develop a semi-supervised soft sensor for the silicon content prediction.

Most soft sensors have their fixed prediction domains. The predictive accuracy of soft sensors gradually decreases due to changes in the state of chemical plants [36]. Consequently, flexible models with adaptive structure, e.g., just-in-time-learning (JITL) soft sensors [23,24,37] are more attractive than only using a fixed one in practice use. Unfortunately, most conventional JITL-based soft sensors were constructed only with the labeled data. Only the labeled data are considered in the process of selection and modeling of similar samples. Consequently, without integration of useful information in the unlabeled data, the prediction performance of JITL-based models may still not be sufficient for some applications.

In this work, bagging local semi-supervised models (BLSM) for online silicon content prediction are proposed. It integrates the bagging strategy, the JITL modeling manner [37] and semi-supervised extreme learning machine (SELM) [34,38,39] into a unified soft sensing framework. For online prediction of a test sample, the useful information in both of similar labeled and unlabeled samples is taken into its special JITL model. Additionally, a simple bagging strategy is adopted to online construct the model. Compared with conventional JITL models only with the labeled data, the prediction performance of BLSM is improved by utilizing the useful information in unlabeled data.

This work is organized in the following way. The extreme learning machine (ELM) and SELM soft sensors are described in Section 2. Additionally, the BLSM online modeling method and its detailed implementation are proposed in this section. In Section 3, BLSM is applied to online silicon content prediction and compared with other approaches. Finally, a conclusion is given in Section 4.

## 2. Soft Sensor Modeling Methods

In this section, three soft sensing methods for the silicon content prediction are presented. First, the ELM-based supervised regression algorithm is briefly described. Second, the SELM-based semi-supervised regression algorithm is presented. Finally, the BLSM online local modeling method is proposed.

### 2.1. Extreme Learning Machine (ELM) Regression Method

The labeled dataset is denoted as $\{\mathbf{S}\} = \{\mathbf{X}, \mathbf{Y}\}$, where $\left\{\mathbf{X}^l\right\} = \left\{\mathbf{x}_i^l\right\}_{i=1}^{L}$ and $\left\{\mathbf{Y}^l\right\} = \left\{y_i^l\right\}_{i=1}^{L}$ are $L$ input and output data, respectively. ELM works for generalized single-hidden layer feedforward networks (SLFNs) [38]. The ELM model has an input layer, a single-hidden layer, and an output layer. With $N$ hidden nodes, ELM approximates the training data, i.e., $\sum_{i=1}^{L} \|y_i^l - \hat{y}_i^l\| = 0$, where $y_i^l$ and $\hat{y}_i^l$ denote the actual output and predicted one, respectively. Compactly, the ELM-based regression formulation [38] is described as:

$$\mathbf{P}\boldsymbol{\alpha} = \mathbf{Y}^l \tag{1}$$

where the output matrix of hidden-layer $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_N]_{L \times N}$ with $\mathbf{p}_i = \begin{bmatrix} v\left(\left\langle \mathbf{a}_i, \mathbf{x}_1^l \right\rangle + b_i\right) \\ \vdots \\ v\left(\left\langle \mathbf{a}_i, \mathbf{x}_L^l \right\rangle + b_i\right) \end{bmatrix}_{L \times 1}$, $i = 1, \ldots, N$; $v\left(\left\langle \mathbf{a}_i, \mathbf{x}_j^l \right\rangle + b_i\right)$ is the activation function output of the $i$th hidden node related to the $j$th input $\mathbf{x}_j^l$. For the $i$th hidden node, $a_i$ and $b_i$ are its input weight and bias, respectively; and $\left\langle \mathbf{a}_i, \mathbf{x}_j^l \right\rangle$ is the

inner product of $\mathbf{a}_i$ and $\mathbf{x}_j^l$. Here, the commonly used nonlinear sigmoidal function $v(x) = \frac{1}{1+\exp(-x)}$ is utilized. The output weights are $\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix}_{N \times 1}$.

Different from the gradient-descent based training algorithms (e.g. backpropagation method) for many NNs and the optimization method based for support vector machines, the essence of ELM is that the hidden layer of SLFNs need not be tuned. Without resorting to some complex training algorithms, the weights of the hidden neurons in ELM can be efficiently computed [38]. For many regression cases, the number of hidden nodes is much less than the number of training samples, i.e., $N << L$. In such a situation, the output weights $\boldsymbol{\alpha}$ [38] are determined as:

$$\boldsymbol{\alpha} = \left(\mathbf{P}^T\mathbf{P}\right)^{-1}\mathbf{P}^T\mathbf{Y}^l \tag{2}$$

Using the Moore–Penrose generalized inverse of matrix $\mathbf{P}$ to solve $\boldsymbol{\alpha}$ in ELM is feasible, i.e., $\boldsymbol{\alpha} = \mathbf{P}^+ Y^l$ [38]. Additionally, to avoid the problem of $\mathbf{P}^T\mathbf{P}$ being noninvertible, a regularized ELM (RELM) model was formulated [34]

$$\boldsymbol{\alpha} = \left(\mathbf{P}^T\mathbf{P} + \gamma\mathbf{I}\right)^{-1}\mathbf{P}^T\mathbf{Y}^l \tag{3}$$

where $\gamma > 0$ is the ridge parameter for the unit matrix $\mathbf{I}$.

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

$$\hat{y}_t = \mathbf{p}_t \boldsymbol{\alpha} = \mathbf{p}_t \left(\mathbf{P}^T\mathbf{P} + \gamma\mathbf{I}\right)^{-1}\mathbf{P}^T\mathbf{Y}^l \tag{4}$$

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

*2.2. Semi-supervised Extreme Learning Machine (SELM) Regression Method*

For the semi-supervised learning methods, the input and output samples are represented as $\{\mathbf{X}\} = \left\{\mathbf{X}^l \cup \mathbf{X}^u\right\}$ and $\mathbf{Y} = \begin{bmatrix} \mathbf{Y}^l \\ \mathbf{Y}^u \end{bmatrix} = \begin{bmatrix} y_1^l \\ \vdots \\ y_L^l \\ \hline 0 \\ \vdots \\ 0 \end{bmatrix}_{(L+U) \times 1}$, respectively. Additionally, the hidden layer output matrix $\mathbf{P}$ can be defined as $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \cdots, \mathbf{p}_N]_{(L+U) \times N}$ as aforementioned. The manifold regularization framework is utilized to learn the matrix $\mathbf{W}$ of an SELM model [39].

$$\min_{\mathbf{W}} \frac{1}{2}\left\{\|\mathbf{JPW} - \mathbf{Y}\|^2 + \lambda(\mathbf{PW})^T\mathbf{LPW}\right\} \tag{5}$$

where $\|\mathbf{JPW} - \mathbf{Y}\|^2$ is the approximation errors of labeled training data (i.e., for the empirical risk) while $\lambda(\mathbf{PW})^T\mathbf{LPW}$ is the penalty term utilizing the graph Laplacian $\mathbf{L}$ with a parameter $\lambda \geq 0$ (i.e., for the complexity of learnt function). All the unlabeled data are integrated into the matrix $\mathbf{P}$. The graph Laplacian $\mathbf{L}$ can be designed using a basic identity in the spectral graph theory [39]. Additionally, for the convenience of calculation, $\mathbf{J} = \begin{bmatrix} \mathbf{I}_L & 0 \\ 0 & 0 \end{bmatrix}_{(L+U)\times(L+U)}$ is defined [39].

By solving Equation (5), the coefficient matrix $\mathbf{W}$ [39] is obtained as:

$$\mathbf{W} = \left[\left(\mathbf{J} + \lambda\mathbf{L}^T\right)\mathbf{P}\right]^+ \mathbf{JY} \tag{6}$$

Generally, for semi-supervised learning methods, there is an assumption that the input patterns from both labeled and unlabeled data are from the same distribution. In such a situation, the data samples in the local region should have similar labels [33,34,39]. Useful information hidden in the unlabeled data can be explored from the above modeling framework. The graph Laplacian $\mathbf{L}$ of SELM contains the information in both of labeled and unlabeled data. Once the unlabeled data are ignored (i.e., $\lambda = 0$), $\mathbf{W}$ is the same as $\boldsymbol{\alpha}$ in Equation (3). The prediction performance improvement can be obtained by suitably choosing $\lambda$ as the penalty of model complexity. Finally, for a query sample $\mathbf{x}_t = [x_{t1}, x_{t2}, \ldots, x_{tn}]^T \in R^n$, its prediction $\hat{y}_t$ is obtained below:

$$\hat{y}_t = \mathbf{p}_t \mathbf{W} = \mathbf{p}_t \left[ \left( \mathbf{J} + \lambda \mathbf{L}^T \right) \mathbf{P} \right]^+ \mathbf{J} \mathbf{Y} \tag{7}$$

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

### 2.3. Bagging Local Semi-supervised Models (BLSM) Online Modeling Method

In industrial processes, JITL-based local soft sensors are more flexible than only using a fixed one for the relatively long-term utilization [23,24]. Nevertheless, most conventional JITL approaches only use limited labeled data, regardless of the useful information in lots of unlabeled data samples. As can be expected, using the unlabeled data, the prediction accuracy of JITL models can be improved.

Online inquiry of $\mathbf{x}_t$ contains three main steps. First, select a similar set $\{\mathbf{S}_t\} = \left\{ \mathbf{S}_t^l \cup \mathbf{S}_t^u \right\}$, including both of $L_t$ labeled data and $U_t$ unlabeled data (i.e., $\left\{ \mathbf{S}_t^l \right\} = \left\{ \mathbf{X}_t^l, \mathbf{Y}_t^l \right\}$ and $\left\{ \mathbf{S}_t^u \right\} = \left\{ \mathbf{X}_t^u \right\}$), from the historical database $\{\mathbf{S}\}$ via some defined similarity criteria [37]. The common Euclidean distance-based similarity is adopted here. Other similarity criteria available [23,24,37] can also be combined with local SELM models. Second, construct a local SELM model $f(\mathbf{x}_t)$ using the selected similar dataset $\{\mathbf{S}_t\}$. Third, online predict and then repeat the same procedure for another query sample.

For a selected $\{\mathbf{S}_t\}$, two parameters, i.e., the number of hidden nodes $N$ and the balance parameter $\lambda \geq 0$, are necessary to train a local SELM model. To avoid the overfitting problem, a simple bagging strategy is adopted to generate multiple local candidate models with diversities and then aggregate them as a new predictor. With the bootstrapping re-sampled strategy, several candidate regression models are ensembling to achieve an improved prediction [40].

For the similar labeled dataset $\left\{ \mathbf{S}_t^l \right\} = \left\{ \mathbf{X}_t^l, \mathbf{Y}_t^l \right\}$, $L_t$ pairs of samples are randomly selected to replace $\left\{ \mathbf{S}_t^l \right\}$ where the probability of each pair being chosen is $\frac{1}{L_t}$ [40]. These $L_t$ pairs of data are a re-sampled training set $\left\{ \mathbf{S}_t^l \right\}$. Sequentially, the procedure is repeated for $K$ times and to obtain $K$ re-sampled datasets, i.e., $\left\{ \mathbf{S}_{t1}^l, \cdots, \mathbf{S}_{tK}^l \right\}$. Similarly, the bagging strategy is applied to the unlabeled dataset $\left\{ \mathbf{S}_t^u \right\} = \left\{ \mathbf{X}_t^u \right\}$ to get $K$ re-sampled datasets $\left\{ \mathbf{S}_{t1}^u, \ldots, \mathbf{S}_{tK}^u \right\}$.

For the $k$th dataset $\{\mathbf{S}_{tk}\} = \left\{ \mathbf{S}_{tk}^l \cup \mathbf{S}_{tk}^u \right\}$, $\mathbf{W}_k$ of the $k$th local SELM model is obtained (similar with Equations (5) and (6)). Consequently, for a test input $\mathbf{x}_t = [x_{t1}, x_{t2}, \ldots, x_{tn}]^T \in R^n$, the prediction value of the $k$th local SELM model, i.e., $\hat{y}_{k,t}$, is formulated:

$$\hat{y}_{k,t} = \mathbf{p}_t \mathbf{W}_k \tag{8}$$

where $\mathbf{p}_t$ is the output matrix of the hidden-layer associated with $\mathbf{x}_t$.

Finally, using a simple ensemble strategy, $K$ candidate SELM models are equally weighted to generate the final prediction.

$$\hat{y}_t = \frac{1}{K} \sum_{k=1}^{K} \hat{y}_{k,t} \tag{9}$$

The main modeling flowchart of BLSM is given in Figure 1. In summary, BLSM has two main characteristics. First, the useful information hidden in unlabeled data is explored and absorbed. Second,

using the bagging strategy [40], the BLSM model can be aggregated using multiple local candidates with diversities.
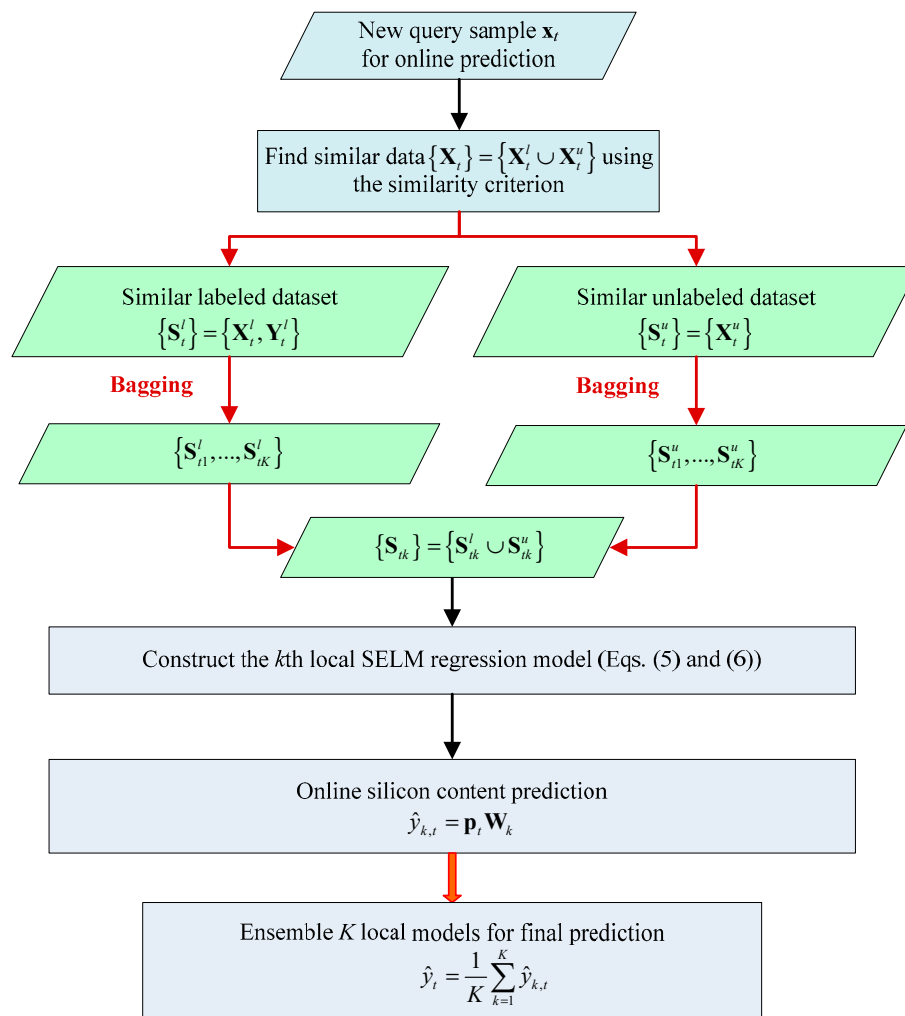


**Figure 1.** Bagging local semi-supervised models (BLSM)-based online soft sensing flowchart for the silicon content prediction.

## 3. Industrial Silicon Content Online Prediction

### 3.1. Data Sets and Pretreatment

The BLSM method is applied to the silicon content prediction in an industrial blast furnace in China. For construction of soft sensors, the related 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 rate [22–24]. After preprocessing the data set with 3-sigma criterion, most of obvious outliers were removed out. A set of about 260 labeled samples was investigated. Half of labeled samples are considered as the historical samples. The remaining part is used for testing the models. Additionally, 500 unlabeled data were obtained as historical samples in the same furnace. The labeled and unlabeled data are from the same industrial blast furnace, indicating that they share with similar characteristics in a production process. Consequently, the semi-supervised learning methods can be applied.

As a recent supervised method with good nonlinear regression performance, the just-in-time least squares SVR (JLSSVR) soft sensor [23] is adopted for comparison. Additionally, as a semi-supervised model, the SELM model [39] is also combined with JITL to construct a local SELM soft sensor here.

Two common performance indices, including the root-mean-square error (RMSE), the relative RMSE (simply denoted as RE), and the hit rate (HR), are adopted and defined, respectively.

$$\text{RMSE} = \sqrt{\sum_{t=1}^{N_{tst}} \left(\frac{y_t - \hat{y}_t}{N_{tst}}\right)^2} \tag{10}$$

$$\text{RE} = \sqrt{\frac{1}{N_{tst}}\sum_{t=1}^{N_{tst}} \left(\frac{y_t - \hat{y}_t}{y_t}\right)^2} \tag{11}$$

$$\text{HR} = \frac{\sum_{t=1}^{N_{tst}} H_t}{N_{tst}} \times 100\% \tag{12}$$

where $N_{tst}$ is the number of test samples. $H_t$ is defined as:

$$H_t = \begin{cases} 1, & \left|\hat{y}_t - y_t\right| < 0.1 \\ 0, & \text{else} \end{cases} \tag{13}$$

### 3.2. Results and Discussion

First, with different sizes of unlabeled data, the comparison results of three performance indices between two semi-supervised models, i.e., BLSM and local SELM, are shown in Figures 2–4, respectively. For both BLSM and local SELM models, the prediction performance is enhanced gradually with the increase in the size of the unlabeled data. Due to the ensemble local modeling ability, BLSM exhibits superior prediction performance to a single local SELM one. In this case, the prediction performance is not further enhanced when the number of unlabeled samples is more than about 400. This is mainly because most of useful information in unlabeled dataset is absorbed from the first 400 data.
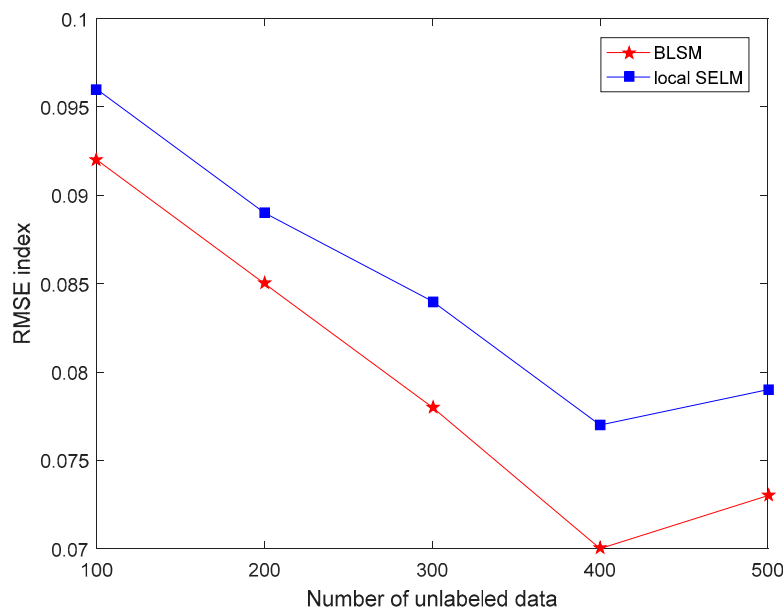


**Figure 2.** Root mean square error (RMSE) comparison of the silicon content prediction between bagging local semi-supervised models (BLSM) and local semi-supervised extreme learning machine (SELM) models with different numbers of unlabeled data.
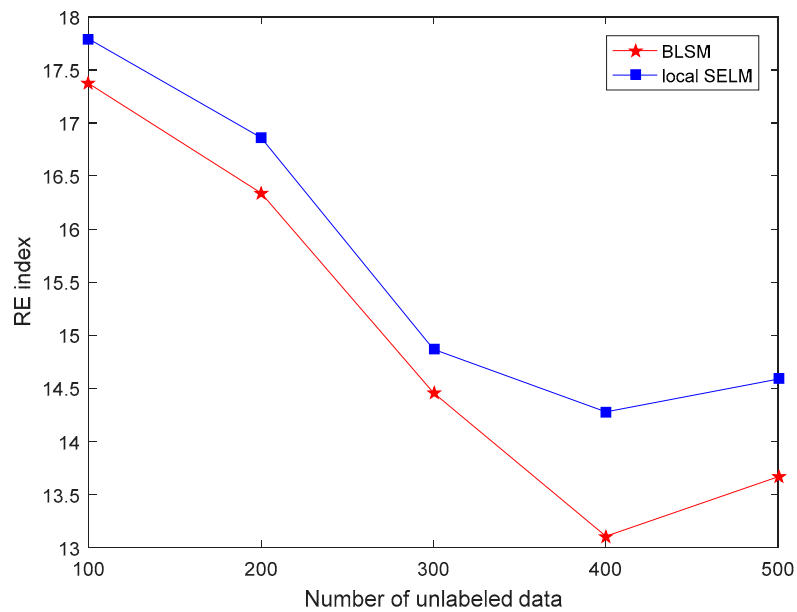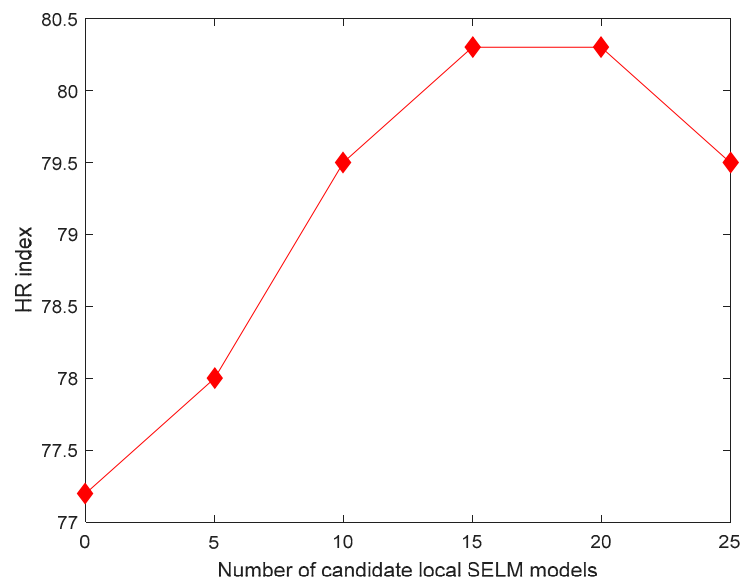
**Figure 3.** Relative RMSE (RE) comparison of the silicon content prediction between bagging local semi-supervised models (BLSM) and local semi-supervised extreme learning machine (SELM) models with different numbers of unlabeled data.
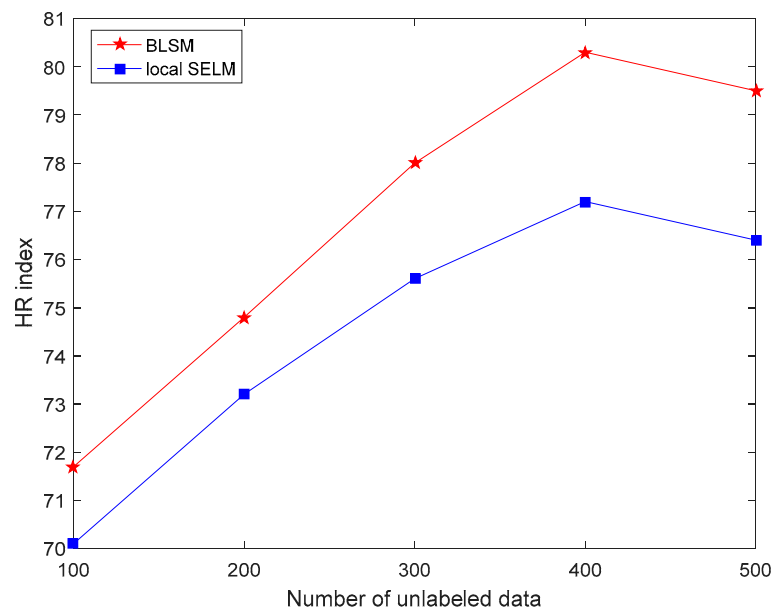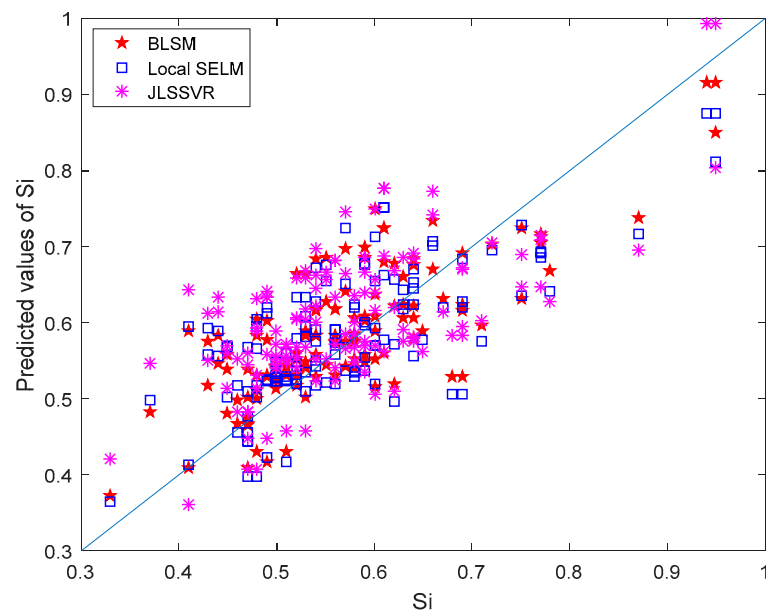


**Figure 4.** Hit Rate (HR) comparison of the silicon content prediction between bagging local semi-supervised models (BLSM) and local semi-supervised extreme learning machine (SELM) models with different numbers of unlabeled data.

With 400 unlabeled data, taking the HR index as an example, different numbers (i.e., $K$) of candidate local SELM models for construction of a BLSM one is shown in Figure 5. With the ensemble learning strategy, the efforts on parameter selection of BLSM can be reduced. The HR index indicates that the ensemble learning can enhance the prediction performance to some extent (the HR value increases from 77.2% to 80.3%). And BLSM achieves the best prediction performance when $K = 15$ for this application.

**Figure 5.** HR comparison of bagging local semi-supervised model (BLSM) different numbers of candidate local semi-supervised extreme learning machine (SELM) models.

For the three soft sensors (i.e., BLSM, local SELM, and JLSSVR [23]), the silicon content prediction results are shown in Figure 6. This parity plot shows that BLSM is better than local SELM and JLSSVR methods. The prediction performance comparison of three modeling methods is listed in Table 1. Their main characteristics are also described briefly. Generally, BLSM is a local semi-supervised learning model and therefore it can better capture nonlinear characteristics in local regions, especially with the help of unlabeled data. For JLSSVR [23] only with a few labeled data, the prediction domain may be limited. Different from JLSSVR [23], BLSM explores and utilizes the hidden information in lots of unlabeled data to improve the local modeling ability. Moreover, using the simple bagging ensemble strategy, the prediction performance of a semi-supervised local model (e.g., a local SELM) can be enhanced.



**Figure 6.** The silicon content assay values against prediction results using bagging local semi-supervised model (BLSM), semi-supervised extreme learning machine (SELM), and just-in-time least squares support vector regression (JLSSVR) soft sensors.

**Table 1.** Detailed prediction performance comparison of semi-supervised and supervised learning models (best results are bold and underlined).

| Soft Sensor Models | Brief Description | RMSE | RE (%) | HR (%) |
|---|---|---|---|---|
| BLSM | Bagging local semi-supervised learning method with ensemble learning strategy | 0.070 | 13.11 | 80.3 |
| Local SELM | Local semi-supervised learning method without ensemble learning strategy | 0.077 | 14.28 | 77.2 |
| JLSSVR [23] | Local supervised learning method | 0.091 | 17.43 | 70.9 |

The computational complexity of BLSM is about $K$ times of a local SELM model. Based on the experiences, $K$ is often much less than 100. The online prediction time of BLSM for a test sample is about 1 s (with CPU main frequency 2.3 GHz and 4 GB memory). Compared with the interval time of lab assay, the computational load is accepted. With more historical data (especially unlabeled data), the computational load of online modeling becomes larger. To alleviate this problem, it is suggested that the online and offline models are integrated using the Bayesian analysis [37]. Alternatively, development of the recursive version of BLSM may be a choice. In summary, all the obtained results show that BLSM is a promising prediction method of the silicon content in hot metal produced in blast furnaces.

## 4. Conclusions

This work has presented an online semi-supervised soft sensor model, i.e., BLSM, for blast furnace hot metal silicon content prediction. Two main advantages distinguish BLSM from most current hot metal silicon prediction soft sensors. First, the useful information in unlabeled data is absorbed into the online modeling and prediction framework efficiently. Second, a bagging-based ensemble strategy is integrated into the online semi-supervised model to improve its prediction reliability. The application results show that BLSM has better prediction performance than traditional soft sensors. This is the first application of semi-supervised learning methods to industrial blast furnaces. How to efficiently select the more informative unlabeled data in an error-in-variables environment for construction of a more robust semi-supervised model will be tackled in our future work.

**Author Contributions:** Data curation, X.H. and K.L.; Funding acquisition, Y.L.; Investigation, X.H., J.J. and K.L.; Methodology, X.H., J.J. and Y.L.; Project administration, Z.G.; Writing—original draft, Y.L.; Writing—review & editing, Y.L.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Nomenclature

| | |
|---|---|
| BLSM | bagging local semi-supervised model |
| ELM | extreme learning machine |
| JITL | just-in-time-learning |
| JLSSVR | just-in-time least squares support vector regression |
| NNs | neural networks |
| RE | relative root-mean-square error |
| RMSE | root-mean-square error |
| RELM | regularized extreme learning machine |
| SELM | semi-supervised extreme learning machine |
| SLFNs | single-hidden layer feedforward networks |
| SVR | support vector regression |

## References

1. Sugawara, K.; Morimoto, K.; Sugawara, T.; Dranoff, J.S. Dynamic behavior of iron forms in rapid reduction of carbon-coated iron ore. *AIChE J.* **1999**, *45*, 574–580. [CrossRef]
2. Radhakrishnan, V.R.; Ram, K.M. Mathematical model for predictive control of the bell-less top charging system of a blast furnace. *J. Process Control* **2001**, *11*, 565–586. [CrossRef]
3. Nogami, H.; Chu, M.; Yagi, J.I. Multi-dimensional transient mathematical simulator of blast furnace process based on multi-fluid and kinetic theories. *Comput. Chem. Eng.* **2005**, *29*, 2438–2448. [CrossRef]
4. Nishioka, K.; Maeda, T.; Shimizu, M. A three-dimensional mathematical modelling of drainage behavior in blast furnace hearth. *ISIJ Int.* **2005**, *45*, 669–676. [CrossRef]
5. Ueda, S.; Natsui, S.; Nogami, H.; Jun-Ichiro, Y.; Ariyama, T. Recent progress and future perspective on mathematical modeling of blast furnace. *ISIJ Int.* **2010**, *50*, 914–923. [CrossRef]
6. Radhakrishnan, V.R.; Mohamed, A.R. Neural networks for the identification and control of blast furnace hot metal quality. *J. Process Control* **2000**, *10*, 509–524. [CrossRef]
7. Jimenez, J.; Mochon, J.; Sainz, D.A.J.; Obeso, F. Blast furnace hot metal temperature prediction through neural networks-based models. *ISIJ Int.* **2007**, *44*, 573–580. [CrossRef]
8. Pettersson, F.; Chakraborti, N.; Saxén, H. A genetic algorithms based multi-objective neural net applied to noisy blast furnace data. *Appl. Soft Comput.* **2007**, *7*, 387–397. [CrossRef]
9. Nurkkala, A.; Pettersson, F.; Saxén, H. Nonlinear modeling method applied to prediction of hot metal silicon in the ironmaking blast furnace. *Ind. Eng. Chem. Res.* **2011**, *50*, 9236–9248. [CrossRef]
10. Hao, X.; Shen, F.; Du, G.; Shen, Y.; Xie, Z. A blast furnace prediction model combining neural network with partial least square regression. *Steel Res. Int.* **2005**, *76*, 694–699. [CrossRef]
11. Bhattacharya, T. Prediction of silicon content in blast furnace hot metal using partial least squares (PLS). *ISIJ Int.* **2005**, *45*, 1943–1945. [CrossRef]
12. Martin, R.D.; Obeso, F.; Mochon, J.; Barea, R.; Jimenez, J. Hot metal temperature prediction in blast furnace using advanced model based on fuzzy logic tools. *Ironmak. Steelmak.* **2007**, *34*, 241–247. [CrossRef]
13. Waller, M.; Saxen, H. On the development of predictive models with applications to a metallurgical process. *Ind. Eng. Chem. Res.* **2000**, *39*, 982–988. [CrossRef]
14. Gao, C.; Zhou, Z.; Chen, J. Assessing the predictability for blast furnace system through nonlinear time series analysis. *Ind. Eng. Chem. Res.* **2008**, *47*, 3037–3045. [CrossRef]
15. Waller, M.; Saxen, H. Application of nonlinear time series analysis to the prediction of silicon content of pig iron. *ISIJ Int.* **2002**, *42*, 316–318. [CrossRef]
16. Miyano, T.; Kimoto, S.; Shibuta, H.; Nakashima, K.; Ikenaga, Y.; Aihara, K. Time series analysis and prediction on complex dynamical behavior observed in a blast furnace. *Physica D* **2000**, *135*, 305–330. [CrossRef]
17. Jian, L.; Gao, C.; Xia, Z. A sliding-window smooth support vector regression model for nonlinear blast furnace system. *Steel Res. Int.* **2011**, *82*, 169–179. [CrossRef]
18. Gao, C.; Jian, L.; Luo, S. Modeling of the thermal state change of blast furnace hearth with support vector machines. *IEEE Trans. Ind. Electron.* **2012**, *59*, 1134–1145. [CrossRef]
19. Gao, C.; Chen, J.; Zeng, J.; Liu, X.; Sun, Y. A chaos-based iterated multistep predictor for blast furnace ironmaking process. *AIChE J.* **2009**, *55*, 947–962. [CrossRef]
20. Chu, Y.; Gao, C. Data-based multiscale modeling for blast furnace system. *AIChE J.* **2014**, *60*, 2197–2210. [CrossRef]
21. Nurkkala, A.; Pettersson, F.; Saxén, H. A study of blast furnace dynamics using multiple autoregressive vector models. *ISIJ Int.* **2012**, *52*, 1763–1770. [CrossRef]
22. Saxén, H.; Gao, C.; Gao, Z. Data-driven time discrete models for dynamic prediction of the hot metal silicon content in the blast furnace—A review. *IEEE Trans. Ind. Electron.* **2013**, *9*, 2213–2225. [CrossRef]
23. Chen, K.; Liu, Y. Adaptive weighting just-in-time-learning quality prediction model for an industrial blast furnace. *ISIJ Int.* **2017**, *57*, 107–113. [CrossRef]
24. Chen, K.; Liang, Y.; Gao, Z.; Liu, Y. Just-in-time correntropy soft sensor with noisy data for industrial silicon content prediction. *Sensors* **2017**, *17*, 1830. [CrossRef] [PubMed]
25. Kano, M.; Nakagawa, Y. Data-based process monitoring, process control, and quality improvement: Recent developments and applications in steel industry. *Comput. Chem. Eng.* **2008**, *32*, 12–24. [CrossRef]

26. Abonyi, J.; Farsang, B.; Kulcsar, T. Data-driven development and maintenance of soft-sensors. In Proceedings of the IEEE 12th International Symposium on Applied Machine Intelligence and Informatics (SAMI), Herlany, Slovakia, 23–25 January 2014; pp. 239–244.

27. Liu, Y.; Yang, C.; Liu, K.; Chen, B.; Yao, Y. Domain adaptation transfer learning soft sensor for product quality prediction. *Chemom. Intell. Lab. Syst.* **2019**, 192. [CrossRef]

28. Ge, Z.; Song, Z.; Ding, S.; Huang, B. Data mining and analytics in the process industry: The role of machine learning. *IEEE Access* **2017**, *5*, 20590–20616. [CrossRef]

29. Liu, Y.; Fan, Y.; Chen, J. Flame images for oxygen content prediction of combustion systems using DBN. *Energy Fuels* **2017**, *31*, 8776–8783. [CrossRef]

30. Xuan, Q.; Fang, B.; Liu, Y.; Wang, J.; Zhang, J.; Zheng, Y.; Bao, G. Automatic pearl classification machine based on a multistream convolutional neural network. *IEEE Trans. Ind. Electron.* **2018**, *65*, 6538–6547. [CrossRef]

31. Xuan, Q.; Chen, Z.; Liu, Y.; Huang, H.; Bao, G.; Zhang, D. Multiview generative adversarial network and its application in pearl classification. *IEEE Trans. Ind. Electron.* **2019**, *66*, 8244–8252. [CrossRef]

32. Zheng, W.; Liu, Y.; Gao, Z.; Yang, J. Just-in-time semi-supervised soft sensor for quality prediction in industrial rubber mixers. *Chemom. Intell. Lab. Syst.* **2018**, *180*, 36–41. [CrossRef]

33. Ge, Z.; Huang, B.; Song, Z. Mixture semisupervised principal component regression model and soft sensor application. *AIChE J.* **2014**, *60*, 533–545. [CrossRef]

34. Zheng, W.; Gao, X.; Liu, Y.; Wang, L.; Yang, J.; Gao, Z. Industrial Mooney viscosity prediction using fast semi-supervised empirical model. *Chemom. Intell. Lab. Syst.* **2017**, *171*, 86–92. [CrossRef]

35. Liu, Y.; Yang, C.; Gao, Z.; Yao, Y. Ensemble deep kernel learning with application to quality prediction in industrial polymerization processes. *Chemom. Intell. Lab. Syst.* **2018**, *174*, 15–21. [CrossRef]

36. Kaneko, H.; Arakawa, M.; Funatsu, K. Applicability domains and accuracy of prediction of soft sensor models. *AIChE J.* **2011**, *57*, 1506–1513. [CrossRef]

37. Liu, Y.; Chen, J. Integrated soft sensor using just-in-time support vector regression and probabilistic analysis for quality prediction of multi-grade processes. *J. Process Control* **2013**, *23*, 793–804. [CrossRef]

38. Huang, G. An insight into extreme learning machines: Random neurons, random features and kernels. *Cogn. Comput.* **2014**, *6*, 376–390. [CrossRef]

39. Liu, J.; Liu, M.; Chen, Y.; Zhao, Z. SELM: Semi-supervised ELM with application in sparse calibrated location estimation. *Neurocomputing* **2011**, *74*, 2566–2572. [CrossRef]

40. Chen, T.; Ren, J.H. Bagging for Gaussian process regression. *Neurocomputing* **2009**, *72*, 1605–1610. [CrossRef]