



(51) International Patent Classification:

G06F 17/18 (2006.01)

(21) International Application Number:

PCT/CN2018/086128

(22) International Filing Date:

09 May 2018 (09.05.2018)

(25) Filing Language:

English

(26) Publication Language:

English

(71) Applicant: JIANGNAN UNIVERSITY [CN/CN]; Xi-

aoyue Geng, No.1800 Lihu Avenue, Wuxi, Jiangsu 214122 (CN).

(72) Inventors: XIONG, Weili; Xiaoyue Geng, No.1800 Lihu

Avenue, Wuxi, Jiangsu 214122 (CN). SHI, Xudong; Xi-

aoyue Geng, No.1800 Lihu Avenue, Wuxi, Jiangsu 214122

(CN). GU, Bingbin; Xiaoyue Geng, No.1800 Lihu Avenue, Wuxi, Jiangsu 214122 (CN). CHE, Xiaoqing; Xiaoyue Geng, No.1800 Lihu Avenue, Wuxi, Jiangsu 214122 (CN). SHENG, Xiaochen; Xiaoyue Geng, No.1800 Lihu Avenue, Wuxi, Jiangsu 214122 (CN).

(74) Agent: HARBIN SHINEIP INTELLECTUAL PROP-

ERTY AGENCY CO., LTD.; Xiaoyue Geng, Room E501, Building 18, Science and Technology 1st Street No.99, District Songbei, Harbin, Heilongjiang 150028 (CN).

(81) Designated States (unless otherwise indicated, for every

kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ,

(54) Title: ADVANCED ENSEMBLE LEARNING STRATEGY BASED SEMI-SUPERVISED SOFT SENSING METHOD

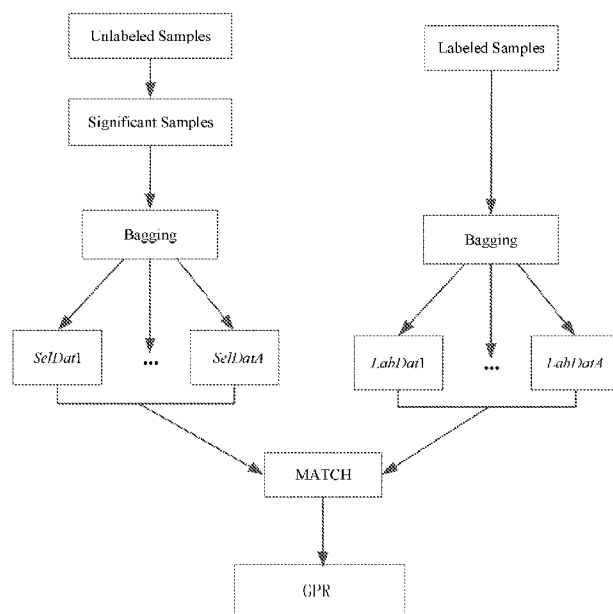


Fig. 1

(57) Abstract: An advanced ensemble learning strategy for soft sensor development with semi-supervised model is provided. The main target of the soft sensor is to improve the prediction performance with a limited number of labeled data samples, under the ensemble learning framework. Firstly, in order to improve the prediction accuracy of sub-models for ensemble modeling, a sample selection mechanism is established to select the most significantly estimated data samples. Secondly, the Bagging method is employed to both of the labeled and selected data set, and the two different kinds of data sets are matched based on the Dissimilarity (DISSIM) algorithm. As a result, the proposed method guarantees the diversity and accuracy of the sub-models which are two significant issues of the ensemble learning. In this work, the soft sensor is constructed upon the Gaussian Process Regression (GPR) model. An industrial case is carried out to demonstrate the effectiveness of the proposed method in dealing with semi-supervised regression problems in the soft sensing area.

OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA,
SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN,
TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

- (84) Designated States** (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

Advanced Ensemble Learning Strategy Based Semi-supervised Soft Sensing Method

The present disclosure relates to industrial modeling and soft sensing. More particularly, the present disclosure relates to the advanced ensemble learning strategy based semi-supervised soft sensing method.

BACKGROUND OF THE DISCLOSURE

Nowadays, with the growing demand of the high-quality chemical/biological products, soft sensor has become more and more important to obtain reliable and real-time predictions of the key quality variables in industrial process. However, it is hard to install hardware sensors to measure those key quality variables, because of the badly circumstance in the industrial environment.

In order to solve the issue that it is hard to measure key quality variables via hardware sensors, soft sensing technology has been paid more and more attentions. With the wide installation of the distributed control system (DCS) which is capable to record a large amount of process data in industrial processes, a variety of data-based soft sensors have been developed. To date, the widely used soft sensing methods include: principal component regression (PCR), partial least square (PLS), artificial neural network (ANN), support vector machine (SVM), Gaussian process regression (GPR), etc.

Nevertheless, the data samples with the output variables are difficult to be acquired, since analyzing quality variables is time-consuming and involves human resource and expensive instruments in the laboratory. As a result, we may need to incorporate the soft sensor modeling problem under a very limited number of data samples for both process and quality variables, and a large amount of data samples for only process variables.

SUMMARY OF THE DISCLOSURE

In order to deal with the unbalanced numbers of the training data, this invention is motivated to build an ensemble learning based semi-supervised regression model which incorporates both labeled and unlabeled data samples.

Firstly, in order to improve the diversity performance, not only the Bagging algorithm is implemented upon both of the labeled and estimated dataset to generate sub-models, but also the way to match the labeled sub-datasets and estimated sub-datasets is designed to improve the diversity performance.

The Dissimilarity (DISSIM) algorithm is employed to measure the dissimilarity index of each labeled sub-dataset and estimated sub-dataset, and the most dissimilar pair of sub-datasets has been matched regularly.

Another important issue is how to guarantee the accuracy of the sub-models, which means that significant estimated data samples should be selected carefully into the next ensemble learning.

Therefore, we define an evaluation index to evaluate the important degree of each estimated data samples based on the influence that each estimated data sample worked on the labeled dataset. In this way, the estimated data samples can be selected according to the evaluation index, which means that the significant estimated data samples with additional information will be selected into the selected dataset repairing for the next ensemble step.

In summary, the missing quality variables are estimated through the K-nearest neighborhood method. Then, the important estimated samples are picked out, while the useless samples are discarded, according to the defined evaluation index. The selected samples and labeled samples are divided into several sub data sets by the Bagging algorithm. These two kinds of the sub data sets are combined based on the DISSIM algorithm. Finally, construct sub models for each combined sub data set to make a prediction for the quarry sample.

This invention is implemented through the following steps:

Step 1: Collect training data set which can be divided as two counterparts: 1) the labeled training data set $D^L = \{\mathbf{X}^L \in R^{n_l \times m}, \mathbf{y}^L \in R^{n_l \times 1}\}$, 2) the unlabeled training data set $D^U = \{\mathbf{X}^U \in R^{n_u \times m}, \mathbf{y}^U \in R^{n_u \times 1}\}$;

Then use self-training method to estimate the missing quality variable in the unlabeled training data set, and denote the estimated and selected data set as $D^E = \{\mathbf{X}^E \in R^{n_e \times m}, \mathbf{y}^E \in R^{n_e \times 1}\}$ and $D^S = \{\mathbf{X}^S \in R^{n_s \times m}, \mathbf{y}^S \in R^{n_s \times 1}\}$, respectively.

where the superscript L , U , E and S , present the labeled, unlabeled, estimated and selected data samples, respectively; and $\{\mathbf{x}\}$ and $\{\mathbf{y}\}$ are the input (process) and output (quality) variables, respectively; n_l , n_u , and n_s are the number of the labeled, unlabeled, and selected data samples, respectively; while m is the number of the process variables;

Step 2: Form the selected dataset $D^S = \{\mathbf{X}^S \in R^{n_s \times m}, \mathbf{y}^S \in R^{n_s \times 1}\}$ according to the labeled training data set and the unlabeled training data set;

where the superscript S means the selected data samples, and n_s is the number of the selected data samples;

Step 3: Employ Bagging algorithm to D^L and D^S respectively, then A labeled sub-datasets $\{D_i^L\}_{i=1}^A$ and A selected sub-datasets $\{D_j^S\}_{j=1}^A$ have been obtained; A is the number of the sub datasets;

Step 4: Match all the sub-datasets, then A matched sub-datasets have been obtained;

Step 5: When a quarry sample comes, predict its quality variable with the matched sub-datasets via the Gaussian Process Regression (GPR) model;

Optionally, between the step 1 and step 2, said method comprises the step:

Step 1.1: Use Eqs. (1) to (3) to calculate the estimated values of the missing output variables to form the estimated data set; said Eqs. (1) to (3) are termed as self-training algorithm;

$$S(i, j) = \gamma \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|) + (1 - \gamma) \cos \langle \mathbf{x}_i, \mathbf{x}_j \rangle \quad (1)$$

$$RS(i, \cdot) = \text{descend}\{S(i, \cdot)\} \quad (2)$$

$$\hat{y}_i = \frac{\sum_{k=1}^K RS(i, \cdot)_k y_{RS(i, \cdot)_k}^L}{\sum_{k=1}^K RS(i, \cdot)_k} \quad (3)$$

where $\|\cdot\|$ denotes the Euclidean distance, and γ is the parameter for the computing the similarity, and $RS(i, \cdot)_k$ represents the k -th element in the rearranged similarity vector $RS(i, \cdot)$, $S(i, \cdot)$ is the similarity vector whose elements present the similarity between the $\mathbf{x}_i$ and all the labeled samples, $RS(i, \cdot)$ is a vector that rearranged in descend for the similarity vector $S(i, \cdot)$ and $y_{RS(i, \cdot)_k}^L$ is the quality variable which has the similarity $RS(i, \cdot)_k$ with the unlabeled data sample $\mathbf{x}_i$;

Optionally, said step 2 comprises the steps:

Step 2.1: Construct temporary datasets which can be denoted as $D_{\text{temp}}(i) = \{[\mathbf{x}^L; \mathbf{x}_i^U], [\mathbf{y}^L; \mathbf{y}_i^E]\}$.

Therefore, we have n_u temporary datasets and they are needed for calculating the evaluation index Q ; The definition of the evaluation index Q is given as follow:

$$Q(i)=Q(\mathbf{x}_i^U, \mathbf{y}_i^E)=\sum_{\mathbf{x}_j \in \mathbf{X}^L} (\tilde{\mathbf{y}}_j^L - \mathbf{y}_j^L)^2 \quad (4)$$

where $\mathbf{y}_j^L$ is the corresponding quality variable of the j -th data sample $\mathbf{x}_j$ in the labeled dataset $\mathbf{X}^L \in R^{n_l \times m}$, and $\tilde{\mathbf{y}}_j^L$ is the prediction of $\mathbf{x}_j$ specified by the GPR model with the i -th temporary dataset $D_{\text{temp}}(i) = \{[\mathbf{x}^L; \mathbf{x}_i^U], [\mathbf{y}^L; \hat{\mathbf{y}}_i^E]\}$;

Step 2.2: Re-arranged the index Q in a descend order;

$$RQ = \text{descend}(Q) = RQ_1 > \dots > RQ_i > \dots > RQ_{n_u} \quad (5)$$

where RQ_i represents the i -th largest value of the element in Q ;

Step 2.3: Select the first n_s data samples with the largest Q values to form the selected dataset $D^S = \{\mathbf{X}^S \in R^{n_s \times m}, \mathbf{y}^S \in R^{n_s \times 1}\}$.

Optionally, said step 3 comprises the following step:

Step 3.1: Calculate the dissimilarity index matrix d whose ij -th element $d(i, j)$ represents the dissimilarity between D_i^E and D_j^S ;

In Bagging algorithm, the m samples of the l -th sub-dataset are all randomly drawn from the original dataset D , which can be represented as $D_l = \{(\mathbf{x}_j, \mathbf{y}_j)\}_{j=1}^m$.

Optionally, said step 4 comprises the following steps:

Step 4.1: Search the largest value $d(i_{\max}, j_{\max})$ of the matrix d and match the $i_{\max}$ -th labeled sub-dataset and the $j_{\max}$ -th selected sub-dataset;

Step 4.2: Update dissimilarity index matrix d by removing the row $d(i_{\max}, \cdot)$ and the column $d(\cdot, j_{\max})$ from the original matrix;

Step 4.3: Where matrix d is the dissimilarity index matrix whose ij -th represents the dissimilarity between the i -th labeled sub-dataset and the j -th selected sub-dataset, based on the dissimilarity (DISSIM) algorithm.

Optionally, the procedure of the DISSIM algorithm of said step 4.3 is given as follows:

Consider the following two datasets, $\mathbf{X}_1 \in R^{n_1 \times m}$ and $\mathbf{X}_2 \in R^{n_2 \times m}$, where n_i is the number of

samples in dataset $\mathbf{X}_i$ and m is the number of variables. The covariance matrices are given by

$$\mathbf{R}_i = \frac{1}{N_i} \mathbf{X}_i^T \mathbf{X}_i \quad (6)$$

The covariance matrix of the mixture of two data sets is given by

$$\mathbf{R} = \frac{N_1}{N_1 + N_2} \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}^T \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} = \frac{N_1}{N_1 + N_2} \mathbf{R}_1 + \frac{N_2}{N_1 + N_2} \mathbf{R}_2 \quad (7)$$

Then $\mathbf{R}$ can be diagonalized by an orthogonal matrix $\mathbf{P}_0$ using eigenvalue decomposition

$$\mathbf{P}_0^T \mathbf{R} \mathbf{P}_0 = \Lambda \quad (8)$$

where Λ is a diagonal matrix whose diagonal elements are eigenvalues of $\mathbf{R}$.

Then the original data $\mathbf{X}_i$ are transformed into $\mathbf{Y}_i$

$$\mathbf{Y}_i = \sqrt{\frac{N_i}{N_1 + N_2}} \mathbf{X}_i \mathbf{P}_0 \Lambda^{-\frac{1}{2}} = \sqrt{\frac{N_i}{N_1 + N_2}} \mathbf{X}_i \mathbf{P} \quad (9)$$

where $\mathbf{P}$ is transformation matrix defined as $\mathbf{P} = \mathbf{P}_0 \Lambda^{-\frac{1}{2}}$

The covariance matrices of the transformed data are given as

$$\begin{aligned} \mathbf{S}_i &= \frac{1}{N_i} \mathbf{Y}_i^T \mathbf{Y}_i = \frac{N_i}{N_1 + N_2} \mathbf{P}^T \frac{\mathbf{X}_i^T \mathbf{X}_i}{N_i} \mathbf{P} \\ &= \frac{N_i}{N_1 + N_2} \mathbf{P}^T \mathbf{R}_i \mathbf{P} \end{aligned} \quad (10)$$

Accordingly, the index D is defined for evaluating the dissimilarity of data sets

$$D = \text{diss}(\mathbf{X}_1, \mathbf{X}_2) = \frac{4}{m} \sum_{j=1}^m (\lambda^j - 0.5)^2 \quad (11)$$

where λ^j denotes the eigenvalues of the covariance matrix $\mathbf{S}_i$;

In this work, only on those eigenvalues whose values are far away from 0.5 are focused on, and the dissimilarity index $\tilde{D}$ is re-defined as follow

$$\tilde{D} = \text{diss}(\mathbf{X}_1, \mathbf{X}_2) = \frac{4}{\tilde{m}} \sum_{j=1}^{\tilde{m}} (\tilde{\lambda}^j - 0.5)^2 \quad (12)$$

where $\tilde{\lambda}^j$ is the j -th element of the set $\tilde{\lambda}^j = \left\{ \text{descend}(|\lambda^j - 0.5|) \right\}$, and $\tilde{m}$ is the number of selected eigenvalues; Unlike the conventional index D measures the whole dissimilar information of the datasets, the new index $\tilde{D}$ relies on the most dissimilar information between two datasets, thus $\tilde{D}$ is more sensitive to different datasets.

Optionally, said step5 comprises the steps:

Step 5.1: Make the final prediction of the quality variable by combining the sub prediction

results, whose calculating method is given as follow:

$$\hat{y}_{final} = \sum_{i=1}^A \frac{1}{A} \hat{y}_i \quad (13).$$

Optionally, the Gaussian Process Regression model of said step 5 is described as follows:

Consider a training dataset $\mathbf{X} \in R^{n \times m}$ and $\mathbf{y} \in R^{n \times 1}$, where $\{\mathbf{x}_i \in R^m\}_{i=1}^n$ is the input data with m dimensions, and $\{y_i \in R\}_{i=1}^n$ is the output data; The aim of a regression model is to find a functional relationship $y = f(\mathbf{x}|\theta) + \varepsilon$ between input and output; Particularly, the Gaussian process regression model assumes that the regression function with zero-mean Gaussian prior distribution; The description of the function is given as follow

$$y = \{f(\mathbf{x}_1), f(\mathbf{x}_2), \dots, f(\mathbf{x}_n)\} \sim \text{GP}(\mathbf{0}, \mathbf{K}) \quad (14)$$

where $\mathbf{K}$ is the covariance matrix whose ij -th element is $k_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$, and in this paper, the kernel function $k(\cdot)$ is the commonly used squared-exponential covariance function, which is defined as follow

$$k(\mathbf{x}_i, \mathbf{x}_j) = \sigma_f^2 \exp\left\{\frac{(\mathbf{x}_i - \mathbf{x}_j)^T (\mathbf{x}_i - \mathbf{x}_j)}{2l^2}\right\} + \delta_{ij} \sigma_n^2 \quad (15)$$

where $\delta_{ij} = 1$ only when $i = j$, otherwise $\delta_{ij} = 0$; l is length-scale; and σ_f^2 and σ_n^2 are terms as signal and noise variance, respectively; An appropriate hyper-parameter set $\Theta = [\sigma_n^2, \sigma_f^2, l]$ is vital for a GPR model to make an accurate prediction about the output variable; To determine the optimal value of the hyper-parameter set, the following likelihood function should be maximized in the training process

$$L = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log(|\mathbf{K}|) - \frac{1}{2} \mathbf{y}^T \mathbf{K}^{-1} \mathbf{y} \quad (16)$$

Conjugate gradient method is used to obtain the closed solution for the above maximizing problem; An alternative way to determine the optimal value is to employ the sampling methods such as Markov chain Monte Carlo (MCMC) method, which draws samples to approximate the poster distribution of the hyper-parameters;

Once we get the optimal hyper-parameter set, the GPR model is available to make a prediction about the distribution of y_{new} corresponding for $\mathbf{x}_{\text{new}}$; When a quarry data sample $\mathbf{x}_{\text{new}}$ comes, according to the property of the multivariate Gaussian distribution, the posterior distribution of the

output, namely $(y_{\text{new}} | \mathbf{X}, \mathbf{y}, \mathbf{x}_{\text{new}}) \sim N(\mu(y_{\text{new}}), \sigma^2(y_{\text{new}}))$ can be identified, where $\mu(y_{\text{new}})$ and $\sigma^2(y_{\text{new}})$ is the mean and variance of the poster distribution, respectively;

They can be calculated as follows

$$\mu(y_{\text{new}}) = \mathbf{k}^T(\mathbf{x}_{\text{new}}, \mathbf{X}) \mathbf{K}^{-1} \mathbf{y} \quad (17)$$

$$\sigma^2(y_{\text{new}}) = k(\mathbf{x}_{\text{new}}, \mathbf{x}_{\text{new}}) - \mathbf{k}^T(\mathbf{x}_{\text{new}}, \mathbf{X}) \mathbf{K}^{-1} \mathbf{k}(\mathbf{x}_{\text{new}}, \mathbf{X}) \quad (18)$$

where $\mathbf{k}(\mathbf{x}_{\text{new}}, \mathbf{X}) = [k(\mathbf{x}_{\text{new}}, \mathbf{x}_1), k(\mathbf{x}_{\text{new}}, \mathbf{x}_2), \dots, k(\mathbf{x}_{\text{new}}, \mathbf{x}_n)]^T$; Finally, take the expectation $\mu(y_{\text{new}})$ of the posterior distribution as the prediction result about the output variable of the quarry data sample $\mathbf{x}_{\text{new}}$.

Optionally, the above advanced ensemble learning strategy based semi-supervised soft sensing method is characterized by improving the prediction performance with a limited number of labeled data samples, under the ensemble learning framework, and the improved prediction performance can be obtained.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig.1 shows the modeling steps of the advanced ensemble learning strategy based semi-supervised soft sensing method in example1;

Fig.2 shows the modeling steps of the advanced ensemble learning strategy based semi-supervised soft sensing method;

Fig.3 shows the prediction errors under different modeling strategies.

DETAILED DESCRIPTION

Example 1

Based on Fig. 1, the detailed illustration of the advanced ensemble learning strategy based semi-supervised soft sensing method is given.

Step1: Collect training data set which can be divided as two counterparts: 1) the labeled training data set $D^L = \{\mathbf{X}^L \in R^{n_l \times m}, \mathbf{y}^L \in R^{n_l \times 1}\}$, 2) the unlabeled training data set $D^U = \{\mathbf{X}^U \in R^{n_u \times m}, \mathbf{y}^U \in R^{n_u \times 1}\}$;

Then use self-training method to estimate the missing quality variable in the unlabeled training data set, and denote the estimated and selected data set as $D^E = \{\mathbf{X}^U \in R^{n_l \times m}, \mathbf{y}^E \in R^{n_l \times 1}\}$ and $D^S = \{\mathbf{X}^S \in R^{n_s \times m}, \mathbf{y}^S \in R^{n_s \times 1}\}$, respectively.

where the superscript L , U , E and S , present the labeled, unlabeled, estimated and selected data samples, respectively; and $\{\mathbf{x}\}$ and $\{\mathbf{y}\}$ are the input (process) and output (quality) variables, respectively; n_l , n_u and n_s are the number of the labeled, unlabeled, and selected data samples, respectively; while m is the number of the process variables;

Step2: Form the selected dataset $D^S = \{\mathbf{X}^S \in R^{n_s \times m}, \mathbf{y}^S \in R^{n_s \times 1}\}$ according to the labeled training data set and the unlabeled training data set;

where the superscript S means the selected data samples, and n_s is the number of the selected data samples;

Step3: Employ Bagging algorithm to D^L and D^S respectively, then A labeled sub-datasets $\{D_i^L\}_{i=1}^A$ and A selected sub-datasets $\{D_j^S\}_{j=1}^A$ have been obtained; A is the number of the sub datasets;

Step4: Match all the sub-datasets, then A matched sub-datasets have been obtained;

Step5: When a quarry sample comes, predict its quality variable with the matched sub-datasets via the Gaussian Process Regression (GPR) model.

Example 2

Based on Fig. 1, the detailed illustration of the advanced ensemble learning strategy based semi-supervised soft sensing method is given.

The debutanizer column process is utilized, and the soft sensors are designed for the butane content.

Step①: Collect training data set which can be divided as two counterparts: 1) the labeled training data set $D^L = \{\mathbf{X}^L \in R^{n_l \times m}, \mathbf{y}^L \in R^{n_l \times 1}\}$, 2) the unlabeled training data set $D^U = \{\mathbf{X}^U \in R^{n_u \times m}, \mathbf{y}^U \in R^{n_u \times 1}\}$.

Then use self-training method to estimate the missing quality variable in the unlabeled training data set, and denote the estimated and selected data set as $D^E = \{\mathbf{X}^U \in R^{n_l \times m}, \mathbf{y}^E \in R^{n_l \times 1}\}$ and $D^S = \{\mathbf{X}^S \in R^{n_s \times m}, \mathbf{y}^S \in R^{n_s \times 1}\}$, respectively.

where the superscript L, U, E and S present the labeled, unlabeled, estimated and selected data samples, respectively; and $\{\mathbf{X}\}$ and $\{\mathbf{y}\}$ are the input (process) and output (quality) variables, respectively; n_l and n_u are the number of the labeled and unlabeled data samples, respectively; while m is the number of the process variables.

Then use Eqs. (1) to (3) to calculate the estimated values of the missing output variables to form the estimated data set, Eqs. (1) to (3) are termed as self-training algorithm.

$$S(i, j) = \gamma \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|) + (1 - \gamma) \cos \langle \mathbf{x}_i, \mathbf{x}_j \rangle \quad (1)$$

$$RS(i, \cdot) = \text{descend}\{S(i, \cdot)\} \quad (2)$$

$$\hat{y}_i = \frac{\sum_{k=1}^K RS(i, \cdot)_k y_{RS(i, \cdot)_k}^L}{\sum_{k=1}^K RS(i, \cdot)_k} \quad (3)$$

where $\|\cdot\|$ denotes the Euclidean distance, and γ is the parameter for the computing the similarity, and $RS(i, \cdot)_k$ represents the k -th element in the rearranged similarity vector $RS(i, \cdot)$, and $y_{RS(i, \cdot)_k}^L$ is the quality variable which has the similarity $RS(i, \cdot)_k$ with the unlabeled data sample $\mathbf{x}_i$.

Step ② : Then construct temporary datasets which can be denoted as $D_{\text{temp}}(i) = \{[\mathbf{x}^L; \mathbf{x}_i^U], [\mathbf{y}^L; \mathbf{y}_i^E]\}$. Therefore, we have n_u temporary datasets and they are needed for calculating the evaluation index Q . The definition of the evaluation index Q is given as follow

$$Q(i) = Q(\mathbf{x}_i^U, \mathbf{y}_i^E) = \sum_{\mathbf{x}_j \in \mathbf{X}^L} (\tilde{y}_j^L - y_j^L)^2 \quad (4)$$

where y_j^L is the corresponding quality variable of the j -th data sample $\mathbf{x}_j$ in the labeled dataset $\mathbf{X}^L \in R^{n_l \times m}$, and $\tilde{y}_j^L$ is the prediction of $\mathbf{x}_j$ specified by the GPR model with the i -th temporary dataset $D_{\text{temp}}(i) = \{[\mathbf{x}_L; \mathbf{x}_i^U], [\mathbf{y}_L; \hat{\mathbf{y}}_i^E]\}$.

Re-arranged the index Q in a descend order

$$RQ = \text{descend}(Q) = RQ_1 > \dots > RQ_i > \dots > RQ_{n_u} \quad (5)$$

where RQ_i represents the i -th largest value of the element in Q .

Select the first n_s data samples with the largest Q values to form the selected dataset $D^S = \{\mathbf{X}^S \in R^{n_s \times m}, \mathbf{y}^S \in R^{n_s \times 1}\}$.

Step ③ : Employ Bagging algorithm to D^L and D^S respectively, then A labeled sub-datasets $\{D_i^L\}_{i=1}^A$ and A selected sub-datasets $\{D_j^S\}_{j=1}^A$ have been obtained. Calculate the dissimilarity index matrix d whose ij -th element $d(i, j)$ represents the dissimilarity between D_i^L and D_j^S .

In Bagging algorithm, the m samples of the l -th sub-dataset are all randomly drawn from the original dataset D , which can be represented as $D_l = \{(\mathbf{x}_j, y_j)\}_{j=1}^m$.

Step ④ : Repeat this step until all the sub-datasets are matched. Search the largest value $d(i_{\max}, j_{\max})$ of the matrix d and match the $i_{\max}$ -th labeled sub-dataset and the $j_{\max}$ -th selected sub-dataset. Then update dissimilarity index matrix d by removing the row $d(i_{\max}, \cdot)$ and the column $d(\cdot, j_{\max})$ from the original matrix.

where matrix d is the dissimilarity index matrix whose ij -th represents the dissimilarity between the i -th labeled sub-dataset and the j -th selected sub-dataset, based on the dissimilarity (DISSIM) algorithm.

The procedure of the DISSIM algorithm is given as follows:

Consider the following two datasets, $\mathbf{X}_1 \in R^{n_1 \times m}$ and $\mathbf{X}_2 \in R^{n_2 \times m}$, where n_i is the number of samples in dataset $\mathbf{X}_i$ and m is the number of variables. The covariance matrices are given by

$$\mathbf{R}_i = \frac{1}{N_i} \mathbf{X}_i^T \mathbf{X}_i \quad (6)$$

The covariance matrix of the mixture of two data sets is given by

$$\mathbf{R} = \frac{N_1}{N_1 + N_2} \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}^T \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} = \frac{N_1}{N_1 + N_2} \mathbf{R}_1 + \frac{N_2}{N_1 + N_2} \mathbf{R}_2 \quad (7)$$

Then $\mathbf{R}$ can be diagonalized by an orthogonal matrix $\mathbf{P}_0$ using eigenvalue decomposition

$$\mathbf{P}_0^T \mathbf{R} \mathbf{P}_0 = \Lambda \quad (8)$$

where Λ is a diagonal matrix whose diagonal elements are eigenvalues of $\mathbf{R}$.

Then the original data $\mathbf{X}_i$ are transformed into $\mathbf{Y}_i$

$$\mathbf{Y}_i = \sqrt{\frac{N_i}{N_1 + N_2}} \mathbf{X}_i \mathbf{P}_0 \Lambda^{\frac{1}{2}} = \sqrt{\frac{N_i}{N_1 + N_2}} \mathbf{X}_i \mathbf{P} \quad (9)$$

where $\mathbf{P}$ is transformation matrix defined as $\mathbf{P} = \mathbf{P}_0 \Lambda^{\frac{1}{2}}$

The covariance matrices of the transformed data are given as

$$\begin{aligned} \mathbf{S}_i &= \frac{1}{N_i} \mathbf{Y}_i^T \mathbf{Y}_i = \frac{N_i}{N_1 + N_2} \mathbf{P}^T \frac{\mathbf{X}_i^T \mathbf{X}_i}{N_i} \mathbf{P} \\ &= \frac{N_i}{N_1 + N_2} \mathbf{P}^T \mathbf{R}_i \mathbf{P} \end{aligned} \quad (10)$$

Accordingly, the index D is defined for evaluating the dissimilarity of data sets

$$D = \text{diss}(\mathbf{X}_1, \mathbf{X}_2) = \frac{4}{m} \sum_{j=1}^m (\lambda^j - 0.5)^2 \quad (11)$$

where λ^j denotes the eigenvalues of the covariance matrix $\mathbf{S}_i$.

In this work, only on those eigenvalues whose value is far away from 0.5 is focused on, and the dissimilarity index $\tilde{D}$ is re-defined as follow

$$\tilde{D} = \text{diss}(\mathbf{X}_1, \mathbf{X}_2) = \frac{4}{\tilde{m}} \sum_{j=1}^{\tilde{m}} (\tilde{\lambda}^j - 0.5)^2 \quad (12)$$

where $\tilde{\lambda}^j$ is the j -th element of the set $\tilde{\lambda}^j = \left\{ \text{descend}(|\lambda^j - 0.5|) \right\}$, and $\tilde{m}$ is the number of selected eigenvalues. Unlike the conventional index D measures the whole dissimilar information of the datasets, the new index $\tilde{D}$ relies on the most dissimilar information between two datasets, thus $\tilde{D}$ is more sensitive to differentiate datasets.

Step⑤: Based on the above steps, the A sub-datasets have been obtained. When a quarry sample comes, predict its quality variable with the sub-datasets via the Gaussian Process Regression (GPR) model. Make the final prediction of the quality variable by combining the sub prediction results, whose calculating method is given as follow

$$\hat{y}_{final} = \sum_{i=1}^A \frac{1}{A} \hat{y}_i \quad (13)$$

The Gaussian Process Regression model is described as follows:

Consider a training dataset $\mathbf{X} \in R^{n \times m}$ and $\mathbf{y} \in R^{n \times 1}$, where $\{\mathbf{x}_i \in R^m\}_{i=1}^n$ is the input data with m dimensions, and $\{y_i \in R\}_{i=1}^n$ is the output data. The aim of a regression model is to find a functional relationship $y = f(\mathbf{x} | \theta) + \varepsilon$ between input and output. Particularly, the Gaussian process regression model assumes that the regression function with zero-mean Gaussian prior distribution. The description of the function is given as follows

$$y = \{f(\mathbf{x}_1), f(\mathbf{x}_2), \dots, f(\mathbf{x}_n)\} \sim \text{GP}(\mathbf{0}, \mathbf{K}) \quad (14)$$

where $\mathbf{K}$ is the covariance matrix whose ij -th element is $k_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$, and in this paper, the kernel function $k(\cdot)$ is the commonly used squared-exponential covariance function, which is defined as follow

$$k(\mathbf{x}_i, \mathbf{x}_j) = \sigma_f^2 \exp\left\{\frac{(\mathbf{x}_i - \mathbf{x}_j)^T (\mathbf{x}_i - \mathbf{x}_j)}{2l^2}\right\} + \delta_{ij} \sigma_n^2 \quad (15)$$

where $\delta_{ij} = 1$ only when $i = j$, otherwise $\delta_{ij} = 0$; l is length-scale; and σ_f^2 and σ_n^2 are terms as signal and noise variance, respectively. An appropriate hyper-parameter set $\Theta = [\sigma_n^2, \sigma_f^2, l]$ is vital for a GPR model to make an accurate prediction about the output variable. To determine the optimal value of the hyper-parameter set, the following likelihood function should be maximized in the training process

$$L = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log(|\mathbf{K}|) - \frac{1}{2} \mathbf{y}^T \mathbf{K}^{-1} \mathbf{y} \quad (16)$$

Conjugate gradient method is used to obtain the closed solution for the above maximizing problem. An alternative way to determine the optimal value is to employ the sampling methods such as Markov chain Monte Carlo (MCMC) method, which draws samples to approximate the poster distribution of the hyper-parameters.

Once we get the optimal hyper-parameter set, the GPR model is available to make a prediction about the distribution of $\mathbf{y}_{new}$ corresponding for $\mathbf{x}_{new}$. When a quarry data sample $\mathbf{x}_{new}$ comes, according to the property of the multivariate Gaussian distribution, the posterior distribution

of the output, namely $(y_{\text{new}} | \mathbf{X}, \mathbf{y}, \mathbf{x}_{\text{new}}) \sim N(\mu(y_{\text{new}}), \sigma^2(y_{\text{new}}))$ can be identified, where $\mu(y_{\text{new}})$ and $\sigma^2(y_{\text{new}})$ is the mean and variance of the poster distribution, respectively, which can be calculated as follows

$$\mu(y_{\text{new}}) = \mathbf{k}^T(\mathbf{x}_{\text{new}}, \mathbf{X}) \mathbf{K}^{-1} \mathbf{y} \quad (17)$$

$$\sigma^2(y_{\text{new}}) = k(\mathbf{x}_{\text{new}}, \mathbf{x}_{\text{new}}) - \mathbf{k}^T(\mathbf{x}_{\text{new}}, \mathbf{X}) \mathbf{K}^{-1} \mathbf{k}(\mathbf{x}_{\text{new}}, \mathbf{X}) \quad (18)$$

where $\mathbf{k}(\mathbf{x}_{\text{new}}, \mathbf{X}) = [k(\mathbf{x}_{\text{new}}, \mathbf{x}_1), k(\mathbf{x}_{\text{new}}, \mathbf{x}_2), \dots, k(\mathbf{x}_{\text{new}}, \mathbf{x}_n)]^T$. Finally, take the expectation $\mu(y_{\text{new}})$ of the posterior distribution as the prediction result about the output variable of the quarry data sample $\mathbf{x}_{\text{new}}$.

Fig. 3 shows the prediction errors under different modeling strategies, it can be found that the advanced ensemble learning strategy based semi-supervised soft sensing method is superior to other strategies.

1. An advanced ensemble learning strategy based semi-supervised soft sensing method, includes the following steps:

Step 1: Collect training data set which can be divided as two counterparts: 1) the labeled training data set $D^L = \{\mathbf{X}^L \in R^{n_l \times m}, \mathbf{y}^L \in R^{n_l \times 1}\}$, 2) the unlabeled training data set $D^U = \{\mathbf{X}^U \in R^{n_u \times m}, \mathbf{y}^U \in R^{n_u \times 1}\}$;

Then use self-training method to estimate the missing quality variable in the unlabeled training data set, and denote the estimated and selected data set as $D^E = \{\mathbf{X}^E \in R^{n_l \times m}, \mathbf{y}^E \in R^{n_l \times 1}\}$ and $D^S = \{\mathbf{X}^S \in R^{n_s \times m}, \mathbf{y}^S \in R^{n_s \times 1}\}$, respectively.

where the superscript L , U , E and S , present the labeled, unlabeled, estimated and selected data samples, respectively; and $\{\mathbf{x}\}$ and $\{\mathbf{y}\}$ are the input (process) and output (quality) variables, respectively; n_l , n_u , and n_s are the number of the labeled, unlabeled, and selected data samples, respectively; while m is the number of the process variables;

Step 2: Form the selected dataset $D^S = \{\mathbf{X}^S \in R^{n_s \times m}, \mathbf{y}^S \in R^{n_s \times 1}\}$ according to the labeled training data set and the unlabeled training data set;

where the superscript S means the selected data samples, and n_s is the number of the selected data samples;

Step 3: Employ Bagging algorithm to D^L and D^S respectively, then A labeled sub-datasets $\{D_i^L\}_{i=1}^A$ and A selected sub-datasets $\{D_j^S\}_{j=1}^A$ have been obtained; A is the number of the sub datasets;

Step 4: Match all the sub-datasets, then A matched sub-datasets have been obtained;

Step 5: When a quarry sample comes, predict its quality variable with the matched sub-datasets via the Gaussian Process Regression (GPR) model.

2. The method of claim 1, between the step 1 and step 2, said method comprises the step:

Step 1.1: Use Eqs. (1) to (3) to calculate the estimated values of the missing output variables to form the estimated data set; said Eqs. (1) to (3) are termed as self-training algorithm;

$$S(i, j) = \gamma \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|) + (1 - \gamma) \cos \langle \mathbf{x}_i, \mathbf{x}_j \rangle \quad (1)$$

$$RS(i, \cdot) = \text{descend}\{S(i, \cdot)\} \quad (2)$$

$$y_i^E = \frac{\sum_{k=1}^K RS(i, \cdot)_k y_{RS(i, \cdot)_k}^L}{\sum_{k=1}^K RS(i, \cdot)_k} \quad (3)$$

where $\|\cdot\|$ denotes the Euclidean distance, and γ is the parameter for the computing the similarity, and $RS(i, \cdot)_k$ represents the k -th element in the rearranged similarity vector $RS(i, \cdot)$, $S(i, \cdot)$ is the similarity vector whose elements present the similarity between the $\mathbf{x}_i$ and all the labeled samples, $RS(i, \cdot)$ is a vector that rearranged in descend for the similarity vector $S(i, \cdot)$ and $y_{RS(i, \cdot)_k}^L$ is the quality variable which has the similarity $RS(i, \cdot)_k$ with the unlabeled data sample $\mathbf{x}_i$;

3. The method of claim 2, said step 2 comprises the steps:

Step 2.1: Construct temporary datasets which can be denoted as $D_{\text{temp}}(i) = \{[\mathbf{x}^L; \mathbf{x}_i^U], [\mathbf{y}^L; y_i^E]\}$.

Therefore, we have n_u temporary datasets and they are needed for calculating the evaluation index Q ; The definition of the evaluation index Q is given as follow:

$$Q(i) = Q(\mathbf{x}_i^U, y_i^E) = \sum_{\mathbf{x}_j \in \mathbf{X}^L} (\tilde{y}_j^L - y_j^L)^2 \quad (4)$$

where y_j^L is the corresponding quality variable of the j -th data sample $\mathbf{x}_j$ in the labeled dataset $\mathbf{X}^L \in R^{n_l \times m}$, and $\tilde{y}_j^L$ is the prediction of $\mathbf{x}_j$ specified by the GPR model with the i -th temporary dataset $D_{\text{temp}}(i) = \{[\mathbf{x}^L; \mathbf{x}_i^U], [\mathbf{y}^L; \hat{y}_i^E]\}$;

Step 2.2: Re-arranged the index Q in a descend order;

$$RQ = \text{descend}(Q) = RQ_1 > \dots > RQ_i > \dots > RQ_{n_u} \quad (5)$$

where RQ_i represents the i -th largest value of the element in Q ;

Step 2.3: Select the first n_s data samples with the largest Q values to form the selected dataset $D^S = \{\mathbf{X}^S \in R^{n_s \times m}, \mathbf{y}^S \in R^{n_s \times 1}\}$.

4. The method of claim 3, said step 3 comprises the following step:

Step 3.1: Calculate the dissimilarity index matrix d whose ij -th element $d(i, j)$ represents the dissimilarity between D_i^L and D_j^S ;

In Bagging algorithm, the m samples of the l -th sub-dataset are all randomly drawn from

the original dataset D , which can be represented as $D_l = \{(\mathbf{x}_j, y_j)\}_{j=1}^m$.

5. The method of claim 4, said step 4 comprises the following steps:

Step 4.1: Search the largest value $d(i_{\max}, j_{\max})$ of the matrix d and match the $i_{\max}$ -th labeled sub-dataset and the $j_{\max}$ -th selected sub-dataset;

Step 4.2: Update dissimilarity index matrix d by removing the row $d(i_{\max}, \cdot)$ and the column $d(\cdot, j_{\max})$ from the original matrix;

Step 4.3: Where matrix d is the dissimilarity index matrix whose ij -th represents the dissimilarity between the i -th labeled sub-dataset and the j -th selected sub-dataset, based on the dissimilarity (DISSIM) algorithm.

6. The method of claim 5, the procedure of the DISSIM algorithm of said step 4.3 is given as follows:

Consider the following two datasets, $\mathbf{X}_1 \in R^{n_1 \times m}$ and $\mathbf{X}_2 \in R^{n_2 \times m}$, where n_i is the number of samples in dataset $\mathbf{X}_i$ and m is the number of variables. The covariance matrices are given by

$$\mathbf{R}_i = \frac{1}{N_i} \mathbf{X}_i^T \mathbf{X}_i \quad (6)$$

The covariance matrix of the mixture of two data sets is given by

$$\mathbf{R} = \frac{N_1}{N_1 + N_2} \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}^T \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix} = \frac{N_1}{N_1 + N_2} \mathbf{R}_1 + \frac{N_2}{N_1 + N_2} \mathbf{R}_2 \quad (7)$$

Then $\mathbf{R}$ can be diagonalized by an orthogonal matrix $\mathbf{P}_0$ using eigenvalue decomposition

$$\mathbf{P}_0^T \mathbf{R} \mathbf{P}_0 = \Lambda \quad (8)$$

where Λ is a diagonal matrix whose diagonal elements are eigenvalues of $\mathbf{R}$.

Then the original data $\mathbf{X}_i$ are transformed into $\mathbf{Y}_i$

$$\mathbf{Y}_i = \sqrt{\frac{N_i}{N_1 + N_2}} \mathbf{X}_i \mathbf{P}_0 \Lambda^{-\frac{1}{2}} = \sqrt{\frac{N_i}{N_1 + N_2}} \mathbf{X}_i \mathbf{P} \quad (9)$$

where $\mathbf{P}$ is transformation matrix defined as $\mathbf{P} = \mathbf{P}_0 \Lambda^{-\frac{1}{2}}$

The covariance matrices of the transformed data are given as

$$\begin{aligned} \mathbf{S}_i &= \frac{1}{N_i} \mathbf{Y}_i^T \mathbf{Y}_i = \frac{N_i}{N_1 + N_2} \mathbf{P}^T \frac{\mathbf{X}_i^T \mathbf{X}_i}{N_i} \mathbf{P} \\ &= \frac{N_i}{N_1 + N_2} \mathbf{P}^T \mathbf{R}_i \mathbf{P} \end{aligned} \quad (10)$$

Accordingly, the index D is defined for evaluating the dissimilarity of data sets

$$D = \text{diss}(\mathbf{X}_1, \mathbf{X}_2) = \frac{4}{m} \sum_{j=1}^m (\lambda^j - 0.5)^2 \quad (11)$$

where λ^j denotes the eigenvalues of the covariance matrix $\mathbf{S}_i$;

In this work, only on those eigenvalues whose values are far away from 0.5 are focused on, and the dissimilarity index $\tilde{D}$ is re-defined as follow

$$\tilde{D} = \text{diss}(\mathbf{X}_1, \mathbf{X}_2) = \frac{4}{\tilde{m}} \sum_{j=1}^{\tilde{m}} (\tilde{\lambda}^j - 0.5)^2 \quad (12)$$

where $\tilde{\lambda}^j$ is the j -th element of the set $\tilde{\lambda}^j = \{\text{descend}(|\lambda^j - 0.5|)\}$, and $\tilde{m}$ is the number of selected eigenvalues; Unlike the conventional index D measures the whole dissimilar information of the datasets, the new index $\tilde{D}$ relies on the most dissimilar information between two datasets, thus $\tilde{D}$ is more sensitive to different datasets.

7. The method of claim 5, said step5 comprises the step:

Step 5.1: Make the final prediction of the quality variable by combining the sub prediction results, whose calculating method is given as follow:

$$\hat{y}_{final} = \sum_{i=1}^A \frac{1}{A} \hat{y}_i \quad (13)$$

8. The method of claim 5, the Gaussian Process Regression model of said step 5 is described as follows:

Consider a training dataset $\mathbf{X} \in R^{n \times m}$ and $\mathbf{y} \in R^{n \times 1}$, where $\{\mathbf{x}_i \in R^m\}_{i=1}^n$ is the input data with m dimensions, and $\{y_i \in R\}_{i=1}^n$ is the output data; The aim of a regression model is to find a functional relationship $y = f(\mathbf{x} | \theta) + \varepsilon$ between input and output; Particularly, the Gaussian process regression model assumes that the regression function with zero-mean Gaussian prior distribution; The description of the function is given as follow

$$y = \{f(\mathbf{x}_1), f(\mathbf{x}_2), \dots, f(\mathbf{x}_n)\} \sim \text{GP}(\mathbf{0}, \mathbf{K}) \quad (14)$$

where $\mathbf{K}$ is the covariance matrix whose ij -th element is $k_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$, and in this paper, the kernel function $k(\cdot)$ is the commonly used squared-exponential covariance function, which is defined as follow

$$k(\mathbf{x}_i, \mathbf{x}_j) = \sigma_f^2 \exp\left\{\frac{(\mathbf{x}_i - \mathbf{x}_j)^T (\mathbf{x}_i - \mathbf{x}_j)}{-2l^2}\right\} + \delta_{ij} \sigma_n^2 \quad (15)$$

where $\delta_{ij} = 1$ only when $i = j$, otherwise $\delta_{ij} = 0$; l is length-scale; and σ_f^2 and σ_n^2 are terms as signal and noise variance, respectively; An appropriate hyper-parameter set $\Theta = [\sigma_n^2, \sigma_f^2, l]$ is vital for a GPR model to make an accurate prediction about the output variable; To determine the optimal value of the hyper-parameter set, the following likelihood function should be maximized in the training process

$$L = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log(|\mathbf{K}|) - \frac{1}{2} \mathbf{y}^T \mathbf{K}^{-1} \mathbf{y} \quad (16)$$

Conjugate gradient method is used to obtain the closed solution for the above maximizing problem; An alternative way to determine the optimal value is to employ the sampling methods such as Markov chain Monte Carlo (MCMC) method, which draws samples to approximate the poster distribution of the hyper-parameters;

Once get the optimal hyper-parameter set, the GPR model is available to make a prediction about the distribution of y_{new} corresponding for $\mathbf{x}_{\text{new}}$; When a quarry data sample $\mathbf{x}_{\text{new}}$ comes, according to the property of the multivariate Gaussian distribution, the posterior distribution of the output, namely $(y_{\text{new}} | \mathbf{X}, \mathbf{y}, \mathbf{x}_{\text{new}}) \sim N(\mu(y_{\text{new}}), \sigma^2(y_{\text{new}}))$ can be identified, where $\mu(y_{\text{new}})$ and $\sigma^2(y_{\text{new}})$ is the mean and variance of the poster distribution, respectively;

They can be calculated as follows

$$\mu(y_{\text{new}}) = \mathbf{k}^T(\mathbf{x}_{\text{new}}, \mathbf{X}) \mathbf{K}^{-1} \mathbf{y} \quad (17)$$

$$\sigma^2(y_{\text{new}}) = k(\mathbf{x}_{\text{new}}, \mathbf{x}_{\text{new}}) - \mathbf{k}^T(\mathbf{x}_{\text{new}}, \mathbf{X}) \mathbf{K}^{-1} \mathbf{k}(\mathbf{x}_{\text{new}}, \mathbf{X}) \quad (18)$$

where $\mathbf{k}(\mathbf{x}_{\text{new}}, \mathbf{X}) = [k(\mathbf{x}_{\text{new}}, \mathbf{x}_1), k(\mathbf{x}_{\text{new}}, \mathbf{x}_2), \dots, k(\mathbf{x}_{\text{new}}, \mathbf{x}_n)]^T$; Finally, take the expectation $\mu(y_{\text{new}})$ of the posterior distribution as the prediction result about the output variable of the quarry data sample $\mathbf{x}_{\text{new}}$.

9. The method of claim 5, the above advanced ensemble learning strategy based semi-supervised soft sensing method is characterized by improving the prediction performance with a limited number of labeled data samples, under the ensemble learning framework, and the improved prediction performance can be obtained.

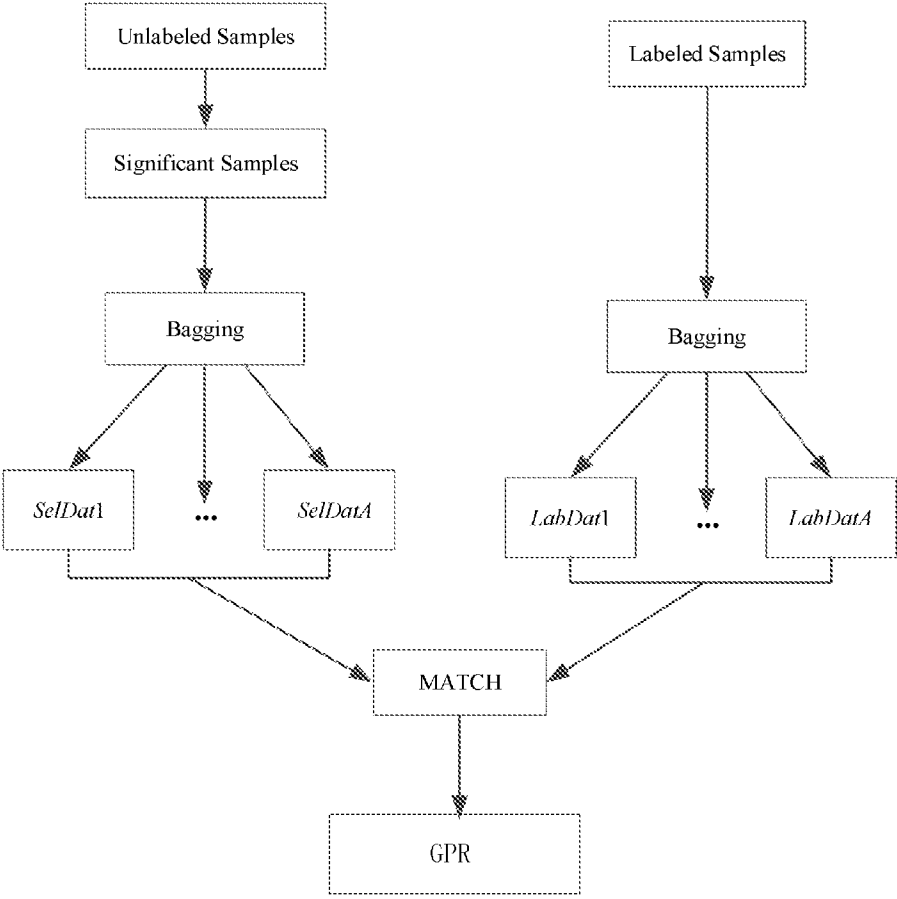


Fig. 1

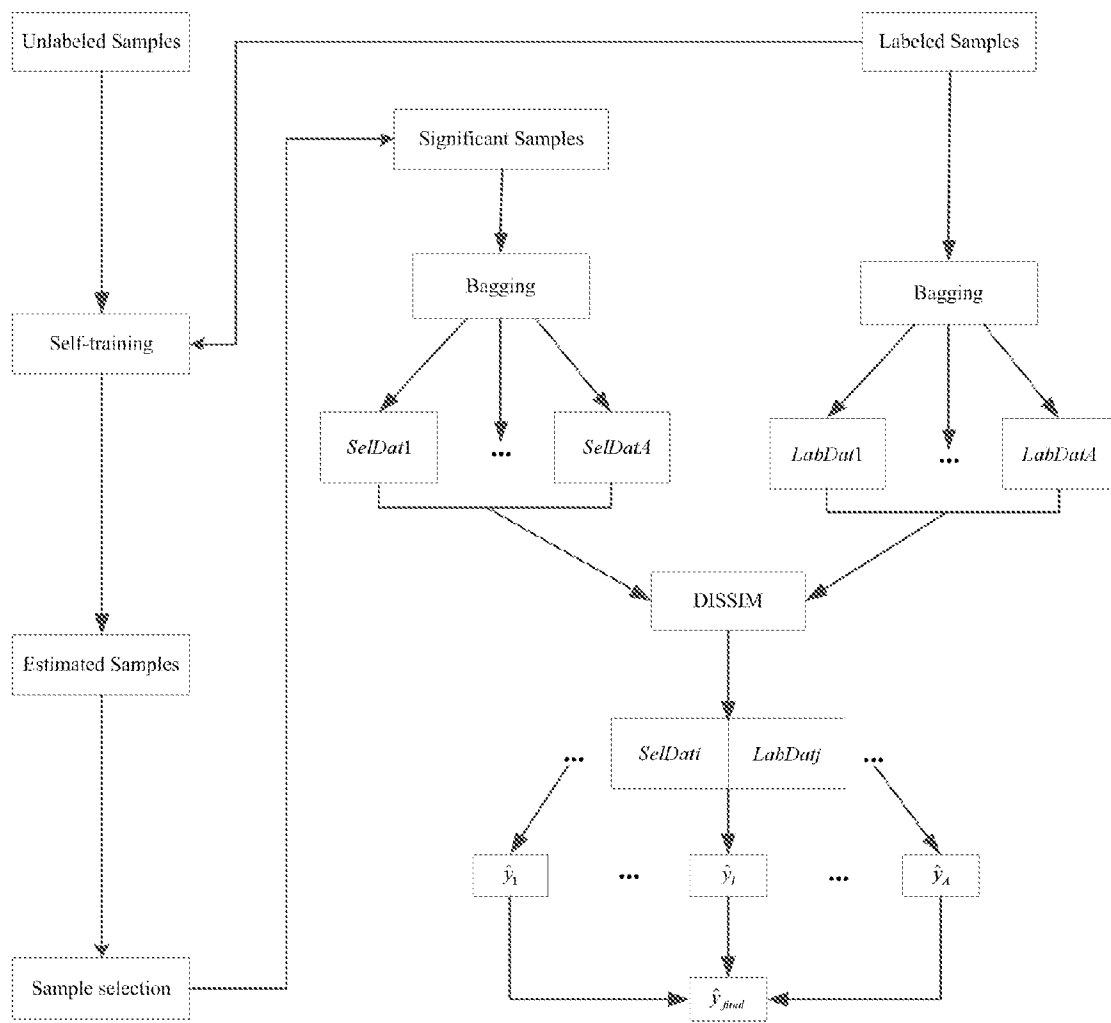


Fig. 2

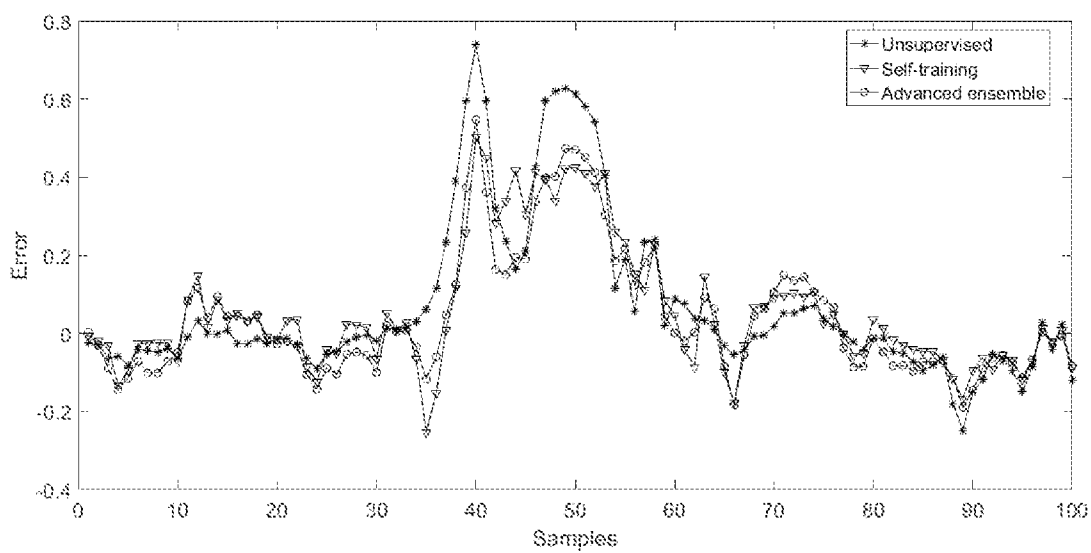


Fig. 3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2018/086128

A. CLASSIFICATION OF SUBJECT MATTER

G06F 17/18(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F; G06K; G06N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, CNKI, EPODOC, WPI: semi w supervised, soft w (sens+ or measur+), label+, unlabel+, self w training, estimate?, select+, Bagging, match+, GPR, Gaussian w process w regression, predict+

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 107451102 A (JIANGNAN UNIVERSITY) 08 December 2017 (2017-12-08) description, paragraphs [0016] to [0028]	1-4
A	CN 102693452 A (SHANGHAI JIAO TONG UNIVERSITY ET AL.) 26 September 2012 (2012-09-26) the whole document	1-9
A	CN 102542126 A (SHANGHAI JIAO TONG UNIVERSITY) 04 July 2012 (2012-07-04) the whole document	1-9
A	CN 102708294 A (SHANGHAI JIAO TONG UNIVERSITY ET AL.) 03 October 2012 (2012-10-03) the whole document	1-9
A	CN 107451101 A (JIANGNAN UNIVERSITY) 08 December 2017 (2017-12-08) the whole document	1-9
A	US 7979363 B1 (MINTER, THOMAS CECIL) 12 July 2011 (2011-07-12) the whole document	1-9

☐ Further documents are listed in the continuation of Box C.☒ See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

16 January 2019

Date of mailing of the international search report

30 January 2019

Name and mailing address of the ISA/CN

National Intellectual Property Administration, PRC
6, Xitucheng Rd., Jimen Bridge, Haidian District, Beijing
100088
China

Authorized officer

ZHU,Xiaolin

Facsimile No. (86-10)62019451

Telephone No. 86-(10)-53962507

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2018/086128

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	107451102	A	08 December 2017	None	
CN	102693452	A	26 September 2012	None	
CN	102542126	A	04 July 2012	CN 102542126	B 22 April 2015
CN	102708294	A	03 October 2012	CN 102708294	B 01 April 2015
CN	107451101	A	08 December 2017	None	
US	7979363	B1	12 July 2011	None	