Research Article

Semi-Supervised Hybrid Local Kernel Regression for Soft Sensor Modelling of Rubber-Mixing Process

Haiqing Yu,1 Jun Ji,2,3 Ping Li,1 Fengjing Shao,2 Shunyao Wu,2 Yi Sui,2 Shujing Li,2 Fengjiao He,4 and Jinming Liu2,3

1State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou, China
2College of Computer Science and Technology, Qingdao University, Qingdao, China
3Beijing Wanling Pangu Science and Technology Ltd., Beijing, China
4XiangYa School of Medicine, Central South University, Changsha, China

Correspondence should be addressed to Jun Ji; junji@qdu.edu.cn

Received 11 July 2019; Accepted 26 September 2019; Published 31 January 2020

Guest Editor: Yuan Yao

Copyright © 2020 Haiqing Yu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Soft sensor techniques have been widely adopted in chemical industry to estimate important indices that cannot be online measured by hardware sensors. Unfortunately, due to the instinct time-variation, the small-sample condition and the uncertainty caused by the drifting of raw materials, it is exceedingly difficult to model the fed-batch processes, for instance, rubber internal mixing processing. Meanwhile, traditional global learning algorithms suffer from the outdated samples while online learning algorithms lack practicality since too many labelled samples of current batch are required to build the soft sensor. In this paper, semi-supervised hybrid local kernel regression (SHLKR) is presented to leverage both historical and online samples to semi-supervised model the soft sensor. Moreover, the recursive formulas are deduced to improve its adaptability and feasibility. Additionally, the rubber Mooney soft sensor of internal mixing processing is implemented using real onsite data to validate proposed method. Compared with classical algorithms, the performance of SHLKR is evaluated and the contribution of unlabelled samples is discussed.

1. Introduction

Fed-batch processes play an important role in chemical and biochemical industry. They are widely adopted in the production of a vast range of fermentation-derived products such as fine-chemical industry, pharmaceuticals and food products. Rubber internal mixing [1] is a classical fed-batch process performed in an internal mixer to achieve an optimal Mooney viscosity for further processing. Since Mooney viscosity cannot be online measured while its laboratory assay is labour-intensive and time-consuming, soft-sensing approaches are investigated to establish a real-time evaluation of it. Furthermore, data-driven but not mechanism-modelling methods are commonly used for its soft sensor modelling because it is a complex nonlinear process without well-developed mechanism. Additionally, its instinctive time-variation, varying properties of natural rubber and additives accompanied with process drifting caused by field conditions, e.g., equipment aging, introduce a great amount of complexity to the process. Moreover, in order to avoid affecting the regular productions, small sample condition always occurred, which further reinforces the difficulty of rubber internal mixing modelling.

In the past decades, many data-driven techniques have been proposed. Extensive reviews can be found in work of Kadlec [2]. Among these methods, multivariate static techniques [3–6] have been widely used. However, these algorithms are relatively sensitive to measurement noise and commonly require a large number of samples to build the promising soft sensor as well. Meanwhile, various artificial neural network (ANN) algorithms [7] have been proposed and successfully applied to polymerization processes, but how to effectively construct the network topology is still an open question. To overcome these shortcomings, kernel-based methods, such as support vector regression [8], least squares support vector regression [9] are presented. These kernel
techniques can attain a better performance under small-sample condition owing to the structural risk minimization criterion.

Note that all the aforementioned algorithms are offline approaches, which can achieve a universal generalization performance but lack the mechanisms to leverage the time-variation characteristics such as drifting of the processes. So, kernel based online modelling algorithms [10–13] were presented. However, too many labelled samples of current batch are required to online build the model, while in most cases in industry field, those samples also have to be predicted instead of lab assay.

Therefore, both online and offline algorithms cannot effectively achieve the promising model [14–17]. On the other hand, taking advantage of the development of both information technology and industrial automation, there are lots of historical productive process data saved in the database of manufacturing execution system [18]. To leverage those data, local learning modelling algorithms [19, 20] were proposed. Nevertheless, those models are not stable owing to the outdated data, which would be used for training. Meanwhile, the unlabelled data are abundant, which contain the production data without indices to be predicted. According to the semi-supervised learning theory, those unlabelled data can be potentially used to improve the predictive model. Therefore, how to effectively leverage both existing historical and online productive process data to create the robust soft sensing model still need to be solved.

In our work, we explore the potential of the hybrid local semi-supervised mechanism to leverage both unlabelled and labelled data via the proposed time window mixed with both historical and online samples. To enhance its feasibility, corresponding recursive calculation formulas are deducted. Furthermore, the soft sensors using proposed and comparative algorithm are implemented to evaluate its performance. To the best of our knowledge, there is no such hybrid local semi-supervised algorithm presented in any article so far.

The remainder of this paper is organized as follows. In Section 2, the detail of proposed SHLKR method, including its recursive calculation derivation is presented. In Section 3, soft sensor modelling experiments of rubber internal mixing process using SHLKR method and comparative algorithms with real industrial field data are presented. Finally, in Section 4, the main contribution of this paper is summarized.

2. Materials and Methods

The thinking of local learning is to create the predictive model dedicated to the prediction of targeted unlabelled sample instead of building the global model using all samples. Since the model will only be created when the prediction is needed, it is also called "Just-in-time learning" or lazy learning [21]. Theoretically it can get more precise model under the condition that similar inputs lead to similar outputs.

Basically, there are three steps of the local learning modelling:

(1) Similar sample set selection: select similar samples from historical data based on one or some similarity calculation algorithms according to the features of the samples to be predicted.

(2) Local modelling: build the local learning model using selected samples with corresponding algorithm.

(3) Prediction: make the prediction and desert the predictive model.

Obviously, the key points of local learning are the algorithms to evaluate the similarity of samples and build the local model. Currently there are two categories that correlation based [19] and distance/angle based [10] similarity calculation algorithms. In this work, distance-based kernel is used because simply algorithm prone to be adopted under industrial application circumstances.

There are two major disadvantages of aforementioned local learning algorithm:

(1) In many cases the online time variation and drifting characteristics cannot be tracked since only similar historical data will be used for the modelling.

(2) Many unlabelled historical and online samples are orderly existed between labelled samples. Those time-series sequence data theoretically can be used to improve the model based on the manifold hypothesis [22] but currently leave unused.

In order to leverage those unused widely existed unlabelled data, we proposed recursive weighted kernel regression (RWKR) [23] before, which has already been validated in penicillin production process soft sensor modelling. But it behaves not promising for some other fed-batch processes, such as rubber internal mixing, since it behaves much more drifting and the time-based weighting mechanism does not work since the Mooney viscosity of rubber is not monotonic increased as the penicillin concentration in penicillin fermentation process. Therefore, in this paper, semi-supervised hybrid local kernel regression (SHLKR) is proposed to fully leverage both labelled and unlabelled data selected from historical and online data.

Different from traditional local kernel learning algorithms:

(1) Besides of labelled samples, combined with labelled samples, unlabelled samples are also used as time window during the training of SHLKR.

(2) Both historical data and online manufactural data are used during training. According to the current run’s index of batch, hybrid training data set is formed by selecting corresponding historical samples joined with online manufactural samples, which can potentially improve the practicability and precision of the soft sensor.

2.1. SHLKR Flow. As is shown in Figure 1, the time window is defined as run’s labelled sample \( \{x_{t-1}, y_{t-1}\} \) with \( X_{u_t} = \{x_{t-u_t}, y_{t-u_t}\} \) which is the \( u_t \) unlabelled sequence samples between \( t-1 \) and \( t \) of current batch. In this way, each labeled sample associated
with its \( u_t \) unlabeled samples is formed as an ordered sequence, which will be entirely used to semi-supervised model the soft sensor. According to the manifold hypothesis of semi-supervised learning theory [24–27], samples tend to be similar within a small local space, unlabelled samples make the data space denser to more precisely describe the characteristic of data samples. So theoretically proposed semi-supervised data combination mechanism can more effectively model the soft sensor than only using labelled samples.

From the first run of first batch, the number of current labelled sample \( N_t \) is 0. If productive process data of current run \( x_t \) will only be collected for modelling in future, it will be added into the unlabelled sample set \( X_u \) of current batch, otherwise, since at this time only historical data can be used for modeling, evaluated by the similarity with \( x_t \), \( k \) most similar historical labelled samples \( (x_t', y_t') \) associated with the unlabelled samples \( X_{t,u} \) within corresponding time windows are selected to semi-supervised train the model. On the other hand, if there are labelled samples \( (x_t, y_t) \) existing, they and associated unlabeled samples \( X_u \) will be both leveraged for training, in this case, if \( N_t \geq k \), only online productive process data will be used, otherwise, \( k - N_t \) most similar historical labelled samples and corresponding unlabeled samples will also be used to train the model.

2.2. SHLKR Recursive Calculation Derivation. Harmonic function is adapted to semi-supervised train the model. Its effectiveness and recursion have been validated before [23]. Although the historical data of training set cannot be recursively adopted since they depend on the \( x_t \) remaining online productive process data can be recursively added because all of them will be used for training. The larger \( t \) becomes, the more reduction it will have from following recursive calculation derivation.

Here we referred to the approach presented by Zhu et al. [28], in which the regularization framework is defined as follows:

\[
Y^* = \arg \min_Y \left\{ \sum_{i=1}^{n} \omega_i \| Y_i - y_i \|^2 + \sum_{i=1}^{j} \| Y_i - y_i \|^2 \right\},
\]

(1)

where \( y_i \) is the real label of sample \( i \), and \( \omega_i \) can be treated as the similarity between sample \( i \) and \( j \), since Gaussian kernel is usually used to calculate the similarity, \( \omega_{ij} \) is typically defined as

\[
\omega_{ij} = K(x_i, x_j) = \exp \left( -\frac{\|x_i - x_j\|^2}{2\sigma^2} \right).
\]

(2)

Gram matrix \( W \) can be partitioned into 4 blocks for labelled samples \( L \) and unlabelled samples \( U \):

\[
W = \begin{bmatrix} W_{LL} & W_{LU} \\ W_{UL} & W_{UU} \end{bmatrix}.
\]

(3)
Then the solution of Equation (1) is formulated as:

\[ Y_U = (D_{UU} - W_{UU})^{-1}W_{LU}Y_L, \]

\[ \Delta^{-1} = (D_{UU}, -W_{UU})^{-1}, \]

\[ \Delta^{-1} = \begin{bmatrix} \sum_j w(t+1)j & \cdots & \sum_j w(t+1)(t+n') \\ \vdots & \ddots & \vdots \\ w(t+1)(t+n') & \cdots & w(t+1)(t+n') \end{bmatrix}, \]

\[ D_{UU}, = \begin{bmatrix} w(t+1) & \cdots & w(t+1)(t+n') \\ \vdots & \ddots & \vdots \\ w(t+1)(t+n') & \cdots & w(t+1)(t+n') \end{bmatrix}, \]

\[ W_{UU}, = \begin{bmatrix} w(t+1) & \cdots & w(t+1)(t+n') \\ \vdots & \ddots & \vdots \\ w(t+1)(t+n') & \cdots & w(t+1)(t+n') \end{bmatrix}, \]

here \( \Delta^{-1} \) can also be divided into four parts:

\[ \Delta^{-1} = \begin{bmatrix} \Delta_{OO} & \Delta_{OH} \\ \Delta_{HO} & \Delta_{HH} \end{bmatrix}^{-1}, \]

where \( \Delta_{OH} \) is the kernel matrix between onlinemanufactual and historical data of time \( t \). \( \Delta_{HO} \) is its transpose. \( \Delta_{OO} \) and \( \Delta_{HH} \) are the kernel matrices of online manufactural data and historical data respectively. First the \( \Delta_{OO} \) is considered as follows:

\[ \Delta_{OO} = \begin{bmatrix} \Delta_{OO} \cdot p^T \\ p \cdot b \end{bmatrix}. \]

Here \( p = -K(x_{u_{0}}, x_{u_{1}}), \ldots, -K(x_{u_{m-1}}, x_{u_{n}}) \), \( b = K(x_{u_{0}}, x_{u_{1}}) + \cdots + K(x_{u_{m-1}}, x_{u_{n}}) \) and:

\[ \Delta_{OO}^{-1} = \Delta_{OO}^{-1} + \begin{bmatrix} K(x_{u_{0}}, x_{u_{0}}) \\ \vdots \\ K(x_{u_{m-1}}, x_{u_{0}}) \end{bmatrix}. \]

Apply Sherman–Morrison–Woodbury to formula, then we get:

\[ \Delta_{OO}^{-1} = \begin{bmatrix} \Delta_{OO}^{-1} + \frac{1}{p^T b - p \Delta_{OO}^{-1} p^{T}} \\ \Delta_{OO}^{-1} + \frac{1}{p^T b - p \Delta_{OO}^{-1} p^{T}} \end{bmatrix} \]

\[ \Delta_{OO}^{-1} = \begin{bmatrix} \Delta_{OO}^{-1} + \frac{1}{p^T b - p \Delta_{OO}^{-1} p^{T}} \\ \Delta_{OO}^{-1} + \frac{1}{p^T b - p \Delta_{OO}^{-1} p^{T}} \end{bmatrix}. \]

Then the \( \Delta_{OO}^{-1} \) can be recursively calculated by \( \Delta_{OO}^{-1} \).

2.3. Application System. Smart Internal Mixing system is a product of MESNAC Co., Ltd., which is widely used in many rubber factories in China. It is mainly formed by four parts: internal mixing modelling, Mooney viscosity prediction, internal mixing process optimization and internal mixing expert system. As is shown in following Figure 2, Smart Internal Mixing system is embedded in the manufacturing execution system, which can monitor the online manufactural data and retrieve the historical manufactural data.

2.4. Experimental Data. Authorized by one rubber manufactory, 222 batches containing 19,148 runs historical samples were retrieved from the system. 2,140 of them were labelled and 17,008 runs are unlabelled which only contain manufactural information without Mooney viscosity value. All samples are from one rubber internal mixing formula to get rid of the formula variation impact. In the industrial application environment, to get the better performance, it also works to model the soft sensor respectively according to different rubber internal mixing formulas. Each sample includes:

(1) Index of current run.
(2) Density.
(3) Hardness.
(4) Minimum torque.
(5) Maximum torque.
(6) Elapsed time to reach 30% maximum torque.
(7) Elapsed time to reach 60% maximum torque.
(8) Elapsed time to increase 2 units after reaching minimum torque.

For labelled samples, all Mooney viscosity values were manually lab assayed. The Mooney viscosity values of first 10 batches are shown in Figure 3, the Mooney viscosity value of unlabelled samples are 0, the dash lines are used to separate different batches. Obviously, the run number of each batch changes a lot owing to its industrial manufactural requirement and the lab assay is performed generally every 8 runs. Besides of that, although the Mooney viscosity is required to be consistent, but the truth is it varies a lot within and between different batches under no obvious rules. It verified our hypothesis that data driven algorithms work in this situation to train the soft sensor.

3. Result and Discussion

To validate the performance of SHLKR, support vector machine (SVM) and Harmonic Functions based soft sensors are also implemented respectively to make the comparison, in which only labelled samples are used. To be fairied, all these three algorithms are using the same labelled samples and only the unlabelled samples respective to those labelled samples are additionally used in SHLKR.

As is shown in Figure 4, the predictive results of all three different algorithms are plotted. The result is for last 27 of 222 batches as well as 1,777 of 19,148 runs including 1,577 unlabelled runs and 200 runs to be predicted. In order to predict those 200 samples, both 1,940 labelled and 15,431 unlabelled samples are used to train the soft sensor.

At the first step of training is to choose the parameter \( k \). After the kernel width 1.1 is determined by leave-one-out cross validation [29], from 2 to 20, the results of using different \( k \) are shown in Figures 5(a)–5(c).

Because SVM cannot be resolved when \( k < 8 \), only SHLKR and Harmonic Functions have results shown in those figures. Obviously when \( k = 5 \), both of them have the best performance, when \( k < 5 \) they both behave unstably and when \( k > 5 \) they all trend to worse but stably. It means that: since \( k \)
only controls the number of historical samples but not the online sample number, besides of too small sample size condition, the model suffers from too many historical samples, as well as that there will be an optimized $k$ existing to trade-off between underfitting and overfitting. Because of that, theoretically $k$ can be automatically selected by traversing from smaller $k$ to larger ones. Besides of algorithms, also depends on the scale of the historical data and the varieties of noise and
corresponding batch. Among all algorithms, SVM behaves the worst since both Harmonic Functions and SHLKR algorithms formula. Here the optimized values are $k = 5$ for SHLKR, $k = 6$ for Harmonic Functions and $k = 11$ for SVM, which are also determined by leave-one-out cross validation.

Some researches indicate that many indices have their own virtues to validate the soft-sensor model. In order to fully investigate the model performance, 3 commonly used criterions: Root-Mean-Square Error (RMSE), Relative root-mean-square Error (RE) and Mean Absolute Error (MAE) [30] are adopted. As is shown in Figures 5(a)–5(c) and Tables 1–4, $N_b$ denotes batch number and $N_r$ represents the run amount of ordering batch.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>RMSE</th>
<th>RE</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>1.772</td>
<td>0.0479</td>
<td>1.363</td>
</tr>
<tr>
<td>HF</td>
<td>1.751</td>
<td>0.0473</td>
<td>1.344</td>
</tr>
<tr>
<td>SHLKR</td>
<td>1.718</td>
<td>0.0466</td>
<td>1.321</td>
</tr>
</tbody>
</table>

Figure 5: Error varying with $k$.  

Table 1: Performance comparison between different algorithms.
Advances in Polymer Technology

are smooth hypothesis based. By leveraging unlabelled samples, SHLKR performs best, which has a 2.7% smaller RMSE than SVM, 1.9% smaller RMSE than Harmonic Functions, 1.5% smaller RE than the others, 3.9% smaller MAE than SVM and 1.7% smaller MAE than Harmonic Functions.

4. Conclusion

In this paper, we propose a new semi-supervised hybrid local kernel regression model for soft sensor modelling of internal rubber mixing processing. Distinguished from traditionally supervised models, it leverages unlabelled samples associated with labelled ones to benefit from widely existed supervised data. And the hybrid mechanism is proposed to effectively use both historical and online manufactural data to improve its practicability. Moreover the recursive formula is deduced to enhance its feasibility. With on-site data, soft sensors using proposed and comparative algorithms are implemented to make the evaluation. Experimental results demonstrate that it has a better performance than classical ones. In our future work, SHLKR will be applied to various rubber manufactories and more features will be added into your model, such as raw rubber information, energy cost of each rubber internal mixing phase etc., which will further increase the precision of proposed model.

Data Availability

The rubber mixing processing data used to support the findings of this study were supplied by Haiqing Yu under license and so cannot be made freely available. Requests for access to these data should be made to Haiqing Yu, 10130207@qq.com.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors’ Contributions

Haiqing Yu and Jun Ji are contributed equally to this work.

Funding

This work is partially supported by the National Natural Science of China (No. 61503208), the National Science Foundation of Shandong Province (No. ZR2015PF002) and the Ministry of Education of Humanities and Social Science Project (No. 15YJC860001).

References


