



Robust Soft Sensor with Deep Kernel Learning for Quality Prediction in Rubber Mixing Processes

Shuihua Zheng ¹, Kaixin Liu ¹, Yili Xu ², Hao Chen ³, Xuelei Zhang ² and Yi Liu ¹,*

- ¹ Institute of Process Equipment and Control Engineering, Zhejiang University of Technology, Hangzhou 310023, China; zneu@zjut.edu.cn (S.Z.); kxliu@zjut.edu.cn (K.L.)
- ² Shanghai Customs, Shanghai 200120, China; lily_xyl@163.com (Y.X.); zhangxueleippt@aliyun.com (X.Z.)
- ³ Quanzhou Institute of Equipment Manufacturing, Haixi Institutes, Chinese Academy of Sciences, Jinjiang 362200, China; chenhao@fjirsm.ac.cn
- * Correspondence: yliuzju@zjut.edu.cn; Tel./Fax: +86-571-85290402

Received: 23 December 2019; Accepted: 25 January 2020; Published: 27 January 2020



Abstract: Although several data-driven soft sensors are available, online reliable prediction of the Mooney viscosity in industrial rubber mixing processes is still a challenging task. A robust semi-supervised soft sensor, called ensemble deep correntropy kernel regression (EDCKR), is proposed. It integrates the ensemble strategy, deep brief network (DBN), and correntropy kernel regression (CKR) into a unified soft sensing framework. The multilevel DBN-based unsupervised learning stage extracts useful information from all secondary variables. Sequentially, a supervised CKR model is built to explore the relationship between the extracted features and the Mooney viscosity values. Without cumbersome preprocessing steps, the negative effects of outliers are reduced using the CKR-based robust nonlinear estimator. With the help of ensemble strategy, more reliable prediction results are further obtained. An industrial case validates the practicality and reliability of EDCKR.

Keywords: soft sensor; deep learning; semi-supervised learning; robust estimator; ensemble strategy; rubber mixing process; Mooney viscosity

1. Introduction

The rubber mixing process is the first and important phase in tire and rubber manufacturing. During the process, natural rubber, synthetic raw materials, and additives are put into the internal mixer. After two to five minutes of mixing, the mixture is discharged to an extruder. In summary, the rubber mixing process is a complex nonlinear process performed in batches. The Mooney viscosity is one of the key quantities concerning end product quality. Despite the commercial importance, no comprehensive analysis of the rubber mixing process is currently available in practice. Additionally, the Mooney viscosity cannot be measured online, and instead it is only assayed offline in the lab with a large delay [1,2]. In such a situation, soft sensors (or inferential sensors) for quality modeling and prediction become very necessary in practice [3–9].

Current data-driven soft sensors for the Mooney viscosity information are generally divided into two categories, supervised and semi-supervised, according to the training datasets being labeled or semi-labeled. Most of the existing Mooney viscosity soft sensors belong to the first category, such as shallow neural networks (NNs) [10,11], partial least squares (PLS) [12,13], Gaussian process regression (GPR) [12–15], and extreme learning machine (ELM) [16]. Generally, they learn a labeled dataset $\mathbf{S}^{l} = \{\mathbf{X}^{l}, \mathbf{Y}\}$ with *N* pairs of input and output samples, denoted as $\mathbf{X}^{l} = \{\mathbf{x}_{i}^{l}\}_{i=1}^{N}$ and $\mathbf{Y} = \{y_{i}\}_{i=1}^{N}$, respectively. One main disadvantage of these supervised prediction models is that the information hidden in *U* unlabeled samples (*U* >> *N*), denoted as $\mathbf{X}^{u} = \{\mathbf{x}_{i}^{u}\}_{i=1}^{U}$, is omitted and not utilized. Alternatively, the semi-supervised soft sensors, such as semi-supervised ELM (SELM), enhance the



2 of 10

prediction results (e.g., compared with ELM) by suitably modeling of both the labeled dataset S^l and the unlabeled dataset $S^u = \{X^u\}$ [17]. To further improve the prediction accuracy, both supervised and semi-supervised soft sensors are further combined with the ensemble learning or just-in-time learning strategies in different scenarios [5,18–20].

For complex rubber mixing processes, without enough prior knowledge, the suitable selection and exaction of input variables is not easy. Although the traditional principal component analysis (PCA) and PLS preprocessing approaches can be used to extract latent variables, they are both linear [4]. Additionally, most PCA-related analysis methods process multivariate data in their raw forms. Alternatively, the representation of data at a deeper level reveals inherent features and becomes more attractive. Recently, increasing applications of deep neural networks (DNNs) have been reported, especially in the speech recognition and computer vision fields [21–29]. As a popular DNN, the deep brief network (DBN) comprises multiple layers for representing data with multilevel abstraction [22]. To describe the important trends in a combustion process, a multilayer DBN was constructed to obtain the nonlinear relationship between the flame images and the outlet oxygen content [25]. An ensemble deep kernel learning model was proposed for the melt index prediction and exhibited good predictions in an industrial polymerization process [26]. The process modeling results indicate that DNNs characterize nonlinear features better and enhance the automation level of industrial manufacturing processes, especially for the Mooney viscosity modeling and prediction.

Another common challenge for a practice soft sensor development is its reliability. This is mainly because the modeling dataset often contains various outliers caused by instrument degradation, process disturbances, transmission problems, etc. [4,30–32]. Robust data mining approaches are necessary and more attractive for development of a reliable soft sensor in industrial processes [33,34]. A soft sensor corrupted by fitting those unwanted outliers inevitably results in erroneous predictions of the output variables. Even with some outlier detection methods as preprocessing, those inconspicuous outliers are difficult to be detected because they are masked by adjacent outliers [31]. In practice, it is more promising to develop a unified soft sensor integrated with a definite reduction of the negative effects of outliers.

To address the two above-mentioned issues simultaneously, this work aims to develop a robust DNN soft sensor for the modeling of nonlinear processes with outliers. Specially, the proposed ensemble deep correntropy kernel regression (EDCKR) framework integrates the ensemble learning [35], DBN structure [22], and correntropy kernel regression (CKR) [31,32]. The DBN-based unsupervised learning is adopted as a multilevel nonlinear feature extractor to absorb the information in related input variables. Sequentially, a supervised CKR-based prediction model is built to capture the relationship between the extracted features and the Mooney viscosity values. Without cumbersome preprocessing steps, the negative effects of outliers is reduced straightforwardly using the CKR-based robust nonlinear estimator [31]. Furthermore, with the help of ensemble learning, more reliable prediction results are obtained.

The remainder of this paper is structured as follows: In Section 2, the EDCKR-based soft sensing method with its algorithmic implemented steps is described in detail. In Section 3, its application to the Mooney viscosity prediction in an industrial rubber mixing process is presented. Finally, in Section 4 the conclusions are summarized.

2. Ensemble Deep Correntropy Kernel Regression Method

2.1. Restricted Boltzmann Machine Construction

Traditionally, using a labeled dataset $S^l = \{X^l, Y\}$, supervised soft sensors are built. Different from traditional supervised learning methods, deep learning methods can integrate unsupervised and supervised learning tasks into a semi-supervised framework [21,22]. When DBN is applied to regression problems, higher-level features are learnt in the unsupervised learning stage to absorb

useful information in all input data, i.e., $\{\mathbf{X}^{u} \cup \mathbf{X}^{l}\}$. For soft sensors, the input data are often considered as those secondary variables which can be measured online during the corresponding process. Using the extracted features, a supervised regression model is then established [25,26].

A brief construction of the main DBN structure with multiple layers is shown in Figure 1. With *L* individual restricted Boltzmann machine (RBM) modules represented as RBM_l, l = 1, ..., L, DBN can extract nonlinear features of the input data hierarchically in the unsupervised learning stage [22]. Each RBM module has a visible layer, V, related to the input data and a hidden layer, H, denoting the outputs, respectively. $\mathbf{V} \in \mathbb{R}^{n \times 1}$ and $\mathbf{H} \in \mathbb{R}^{m \times 1}$ are both vectors with binary values (one or zero). Utilizing the input data as the first visible layer \mathbf{V}_1 , the first RBM module (i.e., RBM₁) is trained using the parameters $\theta_1 = {\mathbf{W}_1, \mathbf{b}_1, \mathbf{c}_1}$ to obtain \mathbf{H}_1 . With a built RBM₁, let $\mathbf{V}_2 = \mathbf{H}_1$, and RBM₂ can be trained similarly. Sequentially, the RBM_l module with \mathbf{H}_l and \mathbf{V}_l is trained and finally a series of RBMs are obtained [22].



Figure 1. Construction of the main deep brief network (DBN) structure with multiple restricted Boltzmann machine (RBM) layers.

The energy function $E(\mathbf{V}, \mathbf{H})$ in Equation (1) with its parameters $\theta = {\mathbf{W}, \mathbf{b}, \mathbf{c}}$ is utilized to describe the energy level of RBM with the available information [22].

$$E(\mathbf{V}, \mathbf{H}) = -\mathbf{b}^{\mathrm{T}}\mathbf{V} - \mathbf{c}^{\mathrm{T}}\mathbf{H} - \mathbf{H}^{\mathrm{T}}\mathbf{W}\mathbf{V}$$
(1)

Specially, to construct an RBM module, the hidden layer H needs to be estimated. To achieve this aim, the probability distribution of the visual layer $P(\mathbf{V})$ in Equaiton (2) is required to be maximized [22]:

$$P(\mathbf{V}) = \frac{\sum_{\mathbf{H}} \exp[-E(\mathbf{V}, \mathbf{H})]}{\sum_{\mathbf{V}} \sum_{\mathbf{H}} \exp[-E(\mathbf{V}, \mathbf{H})]}$$
(2)

Using Equation (2), the log-likelihood function of all visible variables $\log M(\theta)$ is formulated as follows:

$$\log M(\theta) = \sum_{\mathbf{V}} \log P(\mathbf{V}) = \sum_{\mathbf{V}} \{\log \sum_{\mathbf{H}} \exp[-E(\mathbf{V}, \mathbf{H})] - \log\{\sum_{\mathbf{V}} \sum_{\mathbf{H}} \exp[-E(\mathbf{V}, \mathbf{H})]\}\}$$
(3)

The contrastive divergence algorithm is an effective solution to obtain the RBM structure with its parameters $\theta = \{W, b, c\}$. The algorithmic details can be found in [22]. Several trained RBMs are stacked sequentially to form the DBN architecture. Using the layer-by-layer feature extraction, more

useful information with high-level representations is learnt from all available unlabeled data. This is helpful to further model soft sensors for quality prediction.

2.2. Deep Correntropy Kernel Regression Model

As aforementioned, the constructed multilayer unsupervised DBN model characterizes the input data layer-by-layer. To further train a regression model with the output data $\mathbf{Y} = \{y_i\}_{i=1}^N$, supervised learning methods are implemented to fine tune the weights of DBN. Therefore, the extracted features ($\mathbf{\Phi}$) using DBN can be suitably related to the values of the Mooney viscosity (\mathbf{Y}).

Recently, the kernel learning regression method and DBN were combined to construct a prediction model [26]. Compared with the traditional back propagation NN, the kernel learning regression model can be trained more easily. Additionally, it has good prediction performance, especially with limited labeled data [26]. However, the negative effects of outliers degrade the prediction performance and affect the explanation abilities. To solve this problem, using the correntropy concept [36], a supervised deep CKR (DCKR) prediction model is built to explore the relationship between the extracted features ($\mathbf{\Phi} = {\{\boldsymbol{\varphi}_i\}}_{i=1}^N$) and related Mooney viscosity values ($\mathbf{Y} = {\{y_i\}}_{i=1}^N$). Basically, the DCKR-based soft sensor model is described below [31,32].

$$y_i = f(\boldsymbol{\varphi}_i; \boldsymbol{\beta}, b) + e_i$$

= $\boldsymbol{\beta}^T \boldsymbol{\varphi}_i + b + e_i, i = 1, \dots, N$ (4)

where y_i and e_i are the process output and noise for *i*th sample, respectively; *f* is the DCKR model with its parameters β , and bias *b*, respectively.

The following optimization problem is formulated to solve the DCKR model [31,32]:

$$\begin{cases} \min J(\boldsymbol{\beta}, b, \rho) = \frac{\gamma}{2} \sum_{i=1}^{N} \rho(e_i) e_i^2 + \frac{1}{2} \|\boldsymbol{\beta}\|^2 \\ \text{s.t. } y_i - \boldsymbol{\beta}^T \phi(\boldsymbol{\varphi}_i) - b - e_i = 0, \ i = 1, \cdots, N \end{cases}$$
(5)

where the positive regularization parameter γ balances the model's accuracy and complexity. Here, a

simple method is adopted to select σ of the correntropy item $\rho(e_i) = \frac{\exp\left(-\frac{e_i^2}{2\sigma^2}\right)}{\sigma^3 \sqrt{2\pi}}$, i.e., $\sigma = \frac{\max|e_i|}{2\sqrt{2}}$, i = 1, ..., N [31].

Using a two-level training procedure to solve the optimization problem in Equation (5) [31], the DCKR model is established in a straightforward manner. For a test input \mathbf{x}_t , its DBN-based feature is denoted as φ_t . Then, the prediction \hat{y}_t can be obtained.

$$\hat{y}_t = f(\boldsymbol{\beta}, b; \boldsymbol{\varphi}_t) = \sum_{i=1}^N \beta_i K(\boldsymbol{\varphi}_i, \boldsymbol{\varphi}_t) + b$$
(6)

where $K(\boldsymbol{\varphi}_i, \boldsymbol{\varphi}_t)$ is the kernel function of the ith sample.

The weights of samples in a trained DCKR model are $\rho(e_i) = \frac{\exp\left(-\frac{e_i^2}{2\sigma^2}\right)}{\sigma^3 \sqrt{2\pi}}$, i = 1, ..., N. The outliers are not expected to be fitted into the regression model. In such a situation, their fitting errors are relatively larger, and thus they have smaller weights automatically [31]. A sample is assigned with a smaller weight if it is more likely to be an outlier. Meanwhile, using a simple criterion, e.g., $\rho(e_i) < \overline{\rho}$ (0.5 $\leq \overline{\rho} < 1$ is a cutoff value after normalizing all the weights $\rho(e_i)$, i = 1, ..., N into [0,1]), the candidate outliers can be identified and removed out [32]. Interestingly, although the candidate outliers are kept in the DCKR model, they cannot degrade the prediction performance mainly because of their negligible effects. Consequently, compared with the deep kernel learning model [26], the correntropy metric-based DCKR model is more robust for outliers because it cannot amplify the outliers' negative effects.

It should be noticed that, in contrast to correntropy metric-based criterion, most traditional soft sensor and identification methods adopt the mean squared error loss function, which is suitable when the underlying noises obey Gaussian distribution [31,36]. However, they are sensitive to outliers. Additionally, although different weighting strategies to reduce the effect of outliers are available, most of them are not easily designed and implemented for complicated industrial data beforehand.

2.3. Reliability Enhancement Using Bagging-Based Ensemble Strategy

Both the quality and quantity of training data play an important role in the soft sensor model development. Unfortunately, due to the costly assaying process of the Mooney viscosity in industrial rubber mixers, the number of labeled samples is often limited. To improve the model reliability in practice, a simple bagging-based ensemble strategy [37] is integrated with the DCKR model to form EDCKR. The proposed EDCKR model generates multiple predictors and achieves an aggregated prediction.

By bootstrapped resampling the original training dataset, the bagging-based ensemble strategy generates a diversity of regression models [37]. Sequentially, the outputs are aggregated in different weighting ways [35,37–39]. A resampled training dataset $\mathbf{S}_1^l = \{\mathbf{X}_1^l, \mathbf{Y}_1\}$ with *N*-pair samples are randomly selected from $\mathbf{S}^l = \{\mathbf{X}_1^l, \mathbf{Y}\}$, with the probability of each pair being selected as $\frac{1}{N}$. Then *M* resampled datasets denoted as $\mathbf{S}_1^l, \dots, \mathbf{S}_M^l$ can be obtained by repeating the procedure *M* times. Similarly, *M* resampled unlabeled datasets denoted as $\mathbf{X}_1^u, \dots, \mathbf{X}_M^u$ are obtained.

For $\mathbf{S}_m = \{\mathbf{S}_m^l \cup \mathbf{X}_m^u\}$, train a DBN model to extract features $\mathbf{\Phi}_m = \{\mathbf{\varphi}_{m,i}\}_{i=1}^N$. Sequentially, the corresponding DCKR model $f(\mathbf{\beta}_m, b_m)$ is built using $\{\mathbf{\Phi}_m, \mathbf{Y}_m\}$. For online prediction of \mathbf{x}_t , its new features are denoted as $\mathbf{\varphi}_{m,t}$. Accordingly, the DCKR-based prediction $\hat{y}_{m,t}$ is calculated below:

$$\hat{y}_{m,t} = f(\boldsymbol{\beta}_m, \boldsymbol{b}_m, \, \boldsymbol{\varphi}_{m,t}) = \sum_{i=1}^N \beta_{m,i} K(\boldsymbol{\varphi}_{m,i}, \boldsymbol{\varphi}_{m,t}) + b_m \tag{7}$$

where the meanings of parameters are similar with Equation (6).

With *M* resampled datasets, altogether *M* DCKR candidate models are trained. Each DCKR candidate exhibits its individual prediction ability. Generally, a DCKR candidate with fewer outliers is more reliable. Consequently, these candidates are aggregated to a final prediction according to their reliabilities. A simple index R_m is defined to evaluate the reliabilities.

$$R_m = \frac{\operatorname{num}(\rho(e_{mi}) \ge \overline{\rho})}{N} \times 100\%$$
(8)

where num($\rho(e_{mi}) \ge \overline{\rho}$) indicates how many samples with larger weights than $\overline{\rho}$ for the *m*th DCKR candidate.

The DCKR candidate with a larger value of R_m tends to be relatively more reliable because it is trained with fewer outliers. Consequently, the final EDCKR model for prediction is simply formulated below.

$$\hat{y}_{t} = \frac{1}{M} \sum_{m=1}^{M} \frac{R_{m}}{\sum_{m=1}^{M} R_{m}} \hat{y}_{m,t}$$
(9)

The main modeling flowchart of EDCKR is shown in Figure 2. Notice that all input data (i.e., those online measured secondary variables during the process) are utilized. Compared with current soft sensors for the Mooney viscosity, the EDCKR model extracts more intrinsic features using DBN and it is relatively insensitive to outliers in the modeling stage. Moreover, it is expected that, resorting to ensemble strategies, more reliable predictions can be obtained.



Figure 2. Main modeling flowchart of ensemble deep correntropy kernel regression (EDCKR) for soft sensing of the Mooney viscosity.

3. Industrial Mooney Viscosity Prediction

The EDCKR soft modeling approach is applied to an industrial internal mixer. Several measured variables during a short period before the discharge are chosen as secondary variables. These variables include temperature, pressure, energy, power, and duration in the mixer chamber, and they are obtainable in all batches [12,13]. They can reflect important information according to long-term accumulated process knowledge, and thus they are considered as the input data $\{X^u \cup X^l\}$. In contrast to this, the Mooney viscosity can only be assayed about every 10 batches in this manufacturing process. In such a situation, for the investigated recipe during about one month, the labeled dataset $S^l = \{X^l, Y\}$ has only 140 pair of samples. Half of the labeled samples (i.e., 70 pairs) are utilized for training a model. The rest, 70 pairs, are adopted to test the prediction performance. Additionally, the unlabeled training dataset $S^u = \{X^u\}$ has about 680 input variables during the same production period in the same mixer. That is to say, for training a DCKR model, the semi-supervised training data include 680 unlabeled input samples and 70 pair labeled samples. Although obvious sampling and systematic errors can be deleted easily, the modeling dataset still has uncertainties, including process noise and those inconspicuous outliers. In this work, complex outlier defection methods are not utilized. Consequently, robust data-driven process modeling approaches are required in industrial practice.

The relative root mean squares error (RRMSE) is utilized to quantitatively evaluate the prediction performance of different soft-sensor models.

$$\text{RRMSE} = \sqrt{\sum_{t=1}^{N_{\text{tst}}} \left(\frac{y_t - \hat{y}_t}{y_t}\right)^2 / N_{\text{tst}}} \times 100(\%)$$
(10)

where y_t and \hat{y}_t are the assayed and predicted values of the Mooney viscosity, respectively, for N_{tst} test samples.

For comparison studies, four robust soft sensors, including CKR [31,32], PCA-CKR, DCKR, and EDCKR, are investigated. Their main characteristics are described briefly in Table 1. As a supervised method, CKR shows more robustness to outliers than GPR [32]. Additionally, PCA-CKR is designed as a two-step approach by PCA-based feature extraction as preprocessing. DCKR and EDCKR are two proposed robust semi-supervised soft sensors with deep structure. The CKR, PCA-CKR, and DCKR models were trained using the cross-validation method. In particular, the value of $\overline{\rho}$ was selected as 0.8 for this case. The developed DCKR model has a five-layer structure, i.e., 14-20-10-5-1. No further constraints have been adopted in the parameter estimation stage because this is not our main aim.

Additionally, for this case, relative good prediction performance of DCKR can be obtained when the number of extracted features is in the range of four to six. More features do not improve the prediction performance. Therefore, the network structure is selected by cross-validation on several candidates and the optimal is not guaranteed.

Mooney Viscosity	Main Characteristics		RRMSE	Maximum
Soft Sensor	Model Structure	Feature Extraction	(%)	Absolute Error
EDCKR (proposed)	deep (multiple)	nonlinear	4.55	3.28
DCKR (proposed)	deep	nonlinear	5.53	4.16
PCA-CKR	shallow	linear	7.71	5.86
CKR [32]	shallow	no	8.10	5.99

Table 1. Comparison of the Mooney viscosity soft-sensor models: Main characteristics and prediction results.

The comparisons of the Mooney viscosity prediction results are listed in Table 1. The RRMSE index indicates that EDCKR achieves the smallest prediction errors. The prediction results and their assayed values of the test data using the CKR, PCA-CKR, DCKR, and EDCKR models are shown in Figure 3. This parity plot exhibits that EDCKR and DCKR are more accurate mainly because they absorb the information of unlabeled data into a deeper structure. As shown in Table 1 and Figure 3, for feature extraction, the designed PCA-CKR model improves the prediction accuracy inapparently, inferior to DCKR. This is mainly because the two-step PCA-CKR method extracts linear features while they are not very related to sequential quality predictions.



Figure 3. Assayed values and their prediction results of the Mooney viscosity using EDCKR, deep correntropy kernel regression (DCKR), principal component analysis and correntropy kernel regression (PCA-CKR), and correntropy kernel regression (CKR) models.

The Mooney viscosity prediction comparison results between a single DCKR model and an EDCKR one using multiple candidates are plotted in Figure 4, with different candidate numbers. Compared with a single DCKR model, the maximum improvement of EDCKR on the RRMSE index is about 1% (from 5.53% to 4.55%). As listed in Table 1, the values of maximum absolute error (i.e., $\max |y_t - \hat{y}_t|, t = 1, ..., N_{tst}$) of EDCKR and DCKR methods are 3.28 and 4.16, respectively. It indicates that, compared with other methods, the reduction of maximum absolute error is obvious. Additionally,

for this applied recipe, about 15 to 20 DCKR candidates are enough and more candidates are not helpful, while all the training processes are implemented offline.



Figure 4. Relative root mean squares error (RRMSE) comparisons of Mooney viscosity between a single DCKR model and an EDCKR model with different candidates.

The training time required by the EDCKR is about several hours on a personal computer with a CPU main frequency of 2.5 GHz and 8 GB RAM. This is much more than that of CKR and PCA-CKR models (both of which only need several minutes). However, the model training step can be implemented offline. Using the constructed EDCKR model, the online prediction time for a test sample needs about one second. Additionally, recent deep learning training modules are available to make the training process more efficient. In practice, more importantly, the prediction performance of EDCKR is much better than that of both CKR and PCA-CKR.

In summary, the Mooney viscosity prediction results indicate that both EDCKR and DCKR are robust semi-supervised modeling approaches, while the former is more reliable in practice. One main advantage of the recommended EDCKR method is that it can provide more accurate prediction results while the training dataset still contains noises and outliers.

4. Conclusions

A correntropy-based robust semi-supervised soft sensing method has been developed to predict the rubber-mixing Mooney viscosity. The proposed EDCKR-based soft sensor extracts informative features and sequentially constructs a robust prediction model without cumbersome preprocessing steps. The application results indicate that robust deep learning models are alternative tools for industrial data analytics. When new labeled and unlabeled samples are available, how to update the EDCKR model efficiently rather than training from scratch is interesting and needs to be investigated. Additionally, modeling of multiple recipes with uneven datasets, especially for those recipes with extremely limited labeled data, is a practical topic.

Author Contributions: Data curation, S.Z. and H.C.; funding acquisition, S.Z., H.C., and Y.L.; investigation, S.Z., K.L., Y.X., and X.Z.; methodology, S.Z., H.C., and Y.L.; project administration, Y.L.; writing—original draft, S.Z. and Y.L.; writing—review and editing, Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: The National Natural Science Foundation of China (grant nos. 61873241, 51476144, and 61603369) and Zhejiang Provincial Natural Science Foundation of China (grant no. LY18F030024).

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

References

- 1. Mark, J.E.; Erman, B.; Eirich, F.R. *The Science and Technology of Rubber*, 3rd ed.; Elsevier Academic Press: San Diego, CA, USA, 2005; p. 1395.
- 2. Ehabe, E.; Bonfils, F.; Aymard, C.; Akinlabi, A.K. Sainte Beuve, Modelling of Mooney viscosity relaxation in natural rubber. *Polym. Test.* 2005, 24, 620–627. [CrossRef]
- 3. Fortuna, L.; Graziani, S.; Rizzo, A.; Xibilia, M.G. *Soft Sensors for Monitoring and Control of Industrial Processes*; Springer Science & Business Media: Berlin, Germany, 2007.
- 4. Kadlec, P.; Gabrys, B.; Strandt, S. Data-driven soft sensors in the process industry. *Comput. Chem. Eng.* **2009**, 33, 795–814. [CrossRef]
- 5. Ge, Z.; Song, Z.; Ding, S.X.; Huang, B. Data mining and analytics in the process industry: The role of machine learning. *IEEE Access* **2017**, *5*, 20590–20616. [CrossRef]
- 6. Lu, B.; Chiang, L. Semi-supervised online soft sensor maintenance experiences in the chemical industry. *J. Process Control* **2018**, *67*, 23–34. [CrossRef]
- 7. Fortuna, L.; Graziani, S.; Xibilia, M.G. Comparison of soft-sensor design methods for industrial plants using small data sets. *IEEE Trans. Instrum. Meas.* **2009**, *58*, 2444–2451. [CrossRef]
- 8. Han, H.G.; Qiao, J.F. Prediction of activated sludge bulking based on a self-organizing RBF neural network. *J. Process Control* **2012**, *22*, 1103–1112. [CrossRef]
- 9. 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*, 103813. [CrossRef]
- 10. Padmavathi, G.; Mandan, M.G.; Mitra, S.P.; Chaudhuri, K.K. Neural modelling of Mooney viscosity of polybutadiene rubber. *Comput. Chem. Eng.* **2005**, *29*, 1677–1685. [CrossRef]
- 11. Marcos, A.G.; Espinoza, A.V.; Elías, F.A.; Forcada, A.G. A neural network-based approach for optimising rubber extrusion lines. *Int. J. Comput. Integ. Manuf.* **2007**, *20*, 828–837. [CrossRef]
- 12. Zhang, Z.; Song, K.; Tong, T.P.; Wu, F. A novel nonlinear adaptive Mooney-viscosity model based on DRPLS-GP algorithm for rubber mixing process. *Chemom. Intell. Lab. Syst.* **2012**, *112*, 17–23. [CrossRef]
- Song, K.; Wu, F.; Tong, T.; Wang, X. A real-time Mooney-viscosity prediction model of the mixed rubber based on the independent component regression-Gaussian process algorithm. *J. Chemom.* 2012, 11, 557–564. [CrossRef]
- 14. Jin, H.P.; Chen, X.G.; Wang, L.; Yang, K.; Wu, L. Adaptive soft sensor development based on online ensemble Gaussian process regression for nonlinear time-varying batch processes. *Ind. Eng. Chem. Res.* **2015**, *54*, 7320–7345. [CrossRef]
- 15. Yang, K.; Jin, H.P.; Chen, X.G.; Dai, J.Y.; Wang, L.; Zhang, D.X. Soft sensor development for online quality prediction of industrial batch rubber mixing process using ensemble just-in-time Gaussian process regression models. *Chemom. Intell. Lab. Syst.* **2016**, *155*, 170–182. [CrossRef]
- 16. Jin, W.Y.; Liu, Y.; Gao, Z.L. Fast property prediction in an industrial rubber mixing process with local ELM model. *J. Appl. Polym. Sci.* **2017**, *134*, 45391. [CrossRef]
- 17. Zheng, W.J.; Gao, X.J.; Liu, Y.; Wang, L.M.; Yang, J.G.; Gao, Z.L. Industrial Mooney viscosity prediction using fast semi-supervised empirical model. *Chemom. Intell. Lab. Syst.* **2017**, *171*, 86–92. [CrossRef]
- 18. Zheng, W.J.; Liu, Y.; Gao, Z.L.; Yang, J.G. Just-in-time semi-supervised soft sensor for quality prediction in industrial rubber mixers. *Chemom. Intell. Lab. Syst.* **2018**, *180*, 36–41. [CrossRef]
- 19. 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]
- 20. Liu, Y.; Chen, T.; Chen, J. Auto-switch Gaussian process regression-based probabilistic soft sensors for industrial multigrade processes with transitions. *Ind. Eng. Chem. Res.* **2015**, *54*, 5037–5047. [CrossRef]
- 21. Lecun, Y.; Bengio, Y.; Hinton, G. Deep learning. Nature 2015, 521, 436–444. [CrossRef]
- 22. Hinton, G.E. A practical guide to training restricted Boltzmann machines. *Momentum* **2010**, *9*, 599–619.
- 23. Badar, M.; Haris, M.; Fatima, A. Application of deep learning for retinal image analysis: A review. *Comput. Sci. Rev.* **2020**, *35*, 100203. [CrossRef]

- 24. Gao, X.; Shang, C.; Jiang, Y.; Huang, D.; Chen, T. Refinery scheduling with varying crude: A deep belief network classification and multimodel approach. *AIChE J.* **2014**, *60*, 2525–2532.
- 25. Liu, Y.; Fan, Y.; Chen, J. Flame images for oxygen content prediction of combustion systems using DBN. *Energy Fuels* **2017**, *31*, 8776–8783. [CrossRef]
- 26. 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]
- 27. Wu, H.; Zheng, K.; Sfarra, S.; Liu, Y.; Yao, Y. Multi-view learning for subsurface defect detection in composite products: A challenge on thermographic data analysis. *IEEE Trans. Ind. Inform.* 2020. [CrossRef]
- 28. Liu, Y.; Liu, K.; Yang, J.; Yao, Y. Spatial-neighborhood manifold learning for nondestructive testing of defects in polymer composites. *IEEE Trans. Ind. Inform.* **2020**. [CrossRef]
- 29. 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]
- 30. Xu, S.; Lu, B.; Baldea, M.; Edgar, T.F.; Wojsznis, W.; Blevins, T.; Nixon, M. Data cleaning in the process industries. *Rev. Chem. Eng.* 2015, *31*, 453–490. [CrossRef]
- 31. Liu, Y.; Chen, J. Correntropy kernel learning for nonlinear system identification with outliers. *Ind. Eng. Chem. Res.* **2014**, *53*, 5248–5260. [CrossRef]
- 32. Liu, Y.; Fan, Y.; Zhou, L.C.; Jin, F.J.; Gao, Z.L. Ensemble correntropy-based Mooney viscosity prediction model for an industrial rubber mixing process. *Chem. Eng. Technol.* **2016**, *39*, 1804–1812. [CrossRef]
- Zhu, J.; Ge, Z.; Song, Z.; Gao, F. Review and big data perspectives on robust data mining approaches for industrial process modeling with outliers and missing data. *Annu. Rev. Control* 2018, 46, 107–133. [CrossRef]
- 34. Shao, W.; Ge, Z.; Song, Z.; Wang, J. Semi-supervised robust modeling of multimode industrial processes for quality variable prediction based on Student's t mixture model. *IEEE Trans. Ind. Inform.* **2020**. [CrossRef]
- 35. Zhou, Z.H.; Wu, J.X.; Tang, W. Ensembling neural networks: Many could be better than all. *Artif. Intell.* 2002, 137, 239–263. [CrossRef]
- 36. Liu, W.; Pokharel, P.P.; Principe, J.C. Correntropy: Properties and applications in non-gaussian signal processing. *IEEE Trans. Signal Process.* **2007**, *55*, 5286–5298. [CrossRef]
- 37. Chen, T.; Ren, J.H. Bagging for Gaussian process regression. Neurocomputing 2009, 72, 1605–1610. [CrossRef]
- 38. Tang, J.; Chai, T.; Yu, W.; Zhao, L. Modeling load parameters of ball mill in grinding process based on selective ensemble multisensor information. *IEEE Trans. Autom. Sci. Eng.* **2013**, *10*, 726–740. [CrossRef]
- Jin, H.; Pan, B.; Chen, X.; Qian, B. Ensemble just-in-time learning framework through evolutionary multi-objective optimization for soft sensor development of nonlinear industrial processes. *Chemom. Intell. Lab. Syst.* 2019, 184, 153–166. [CrossRef]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).